A Combined Machine Learning and High-Energy X-ray Diffraction Approach to Understanding Liquid and Amorphous Metal Oxides.

Ganesh Sivaraman[1], Gabor Csanyi[2], Alvaro Vazquez-Mayagoitia[3], Ian T. Foster[1], Stephen K. Wilke[4], Richard Weber[4] and Chris J. Benmore[5,*]

[1] Data Science and Learning Division, Argonne National Laboratory, Lemont, Illinois 60439, USA.

[2]University of Cambridge, Cambridge, CB2 1PZ, UK.

[3] Computational Science Division, Argonne National Laboratory, Lemont, Illinois 60439, USA

[4] Materials Development Inc., 3090 Daniels Court, Arlington Heights, IL 60004, USA

[5] X-ray Science Division, Argonne National Laboratory, Lemont, Illinois 60439, USA.

(Received 2021)

Determining the structure-property relations of liquid and amorphous metal oxides is challenging, due to their variable short-range order and polyhedral connectivity. To predict chemically realistic structures, we have developed a Machine Learned, Gaussian Approximation Potential (GAP) for HfO_2, with a focus on enhanced sampling of the training database and accurate density functional theory calculations. By using training datasets for the GAP model at the level of Density Functional Theory - Strongly Constrained and Appropriately Normed (DFT-SCAN) level of theory, our results show that the topology of both the low viscosity liquid and the amorphous form are dominated by edge-shared chains and small corner-shared rings of polyhedra. This topology is shown to be consistent with the structure of other liquid and amorphous transition metal oxides of variable ion size, such as TiO_2 and ZrO_2. Current limitations of the ML-GAP modeling method for obtaining glass structures and future perspectives are also discussed.

Keywords: Glass structure, Machine Learning, X-ray Diffraction, Density Functional Theory

1. Introduction

Amorphous transition metal oxides (TMOs) possess many desirable optical, chemical, thermal, and mechanical properties, making them useful in fuel cell, catalytic, high

temperature ceramic, and thermal barrier applications [1-3]. Using the classification of Angell [4], TMO liquids are considered to be extremely "fragile". That is to say, the temperature dependence of their viscosity deviates from Arrhenius behavior, making glass formation difficult. This phenomenon is illustrated in Fig. 1, where the "strong" glass former liquid SiO_2 follows an Arrhenius viscosity dependence, while the "fragile" liquids ZrO_2, Al_2O_3, and Si follow a highly non-Arrhenius pattern. In terms of structure, strong liquids such as SiO_2 quench to form covalently bonded network glasses, whereas fragile binary metal-oxide silicate liquids form a network that is often heterogeneously modified by the larger metal cations [5]. Less is known about the structure of fragile metal-oxide liquids in which edge-sharing octahedra dominate. In these cases, traditional bulk quenching methods used for glass formation usually result in the crystallization of TMOs. Consequently, amorphous TMOs are usually produced by other methods, such as rf-sputtering, laser ablation, or sol gel processes. Here, we give a brief overview of our combined x-ray diffraction and simulation methods aimed at developing machine learned interatomic potentials for understanding the structural topology associated with inviscid liquids and amorphous TMOs.

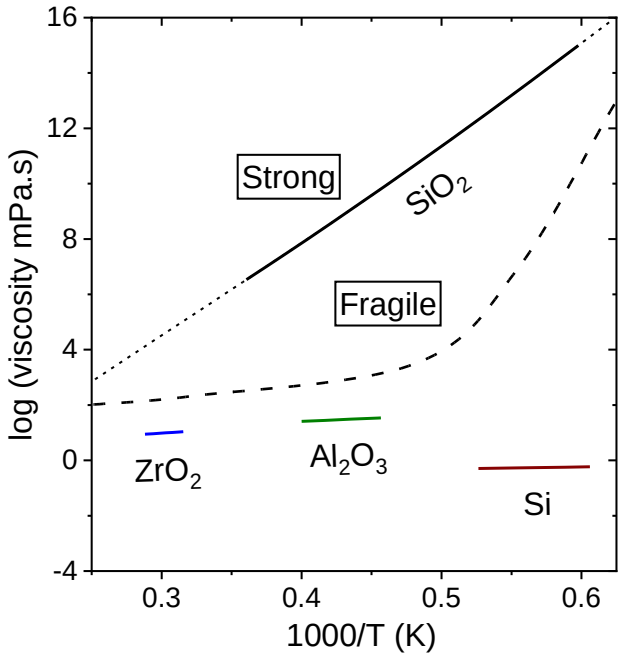


Fig. 1. (Color online) An Angell plot of a strong liquid represented by the SiO_2 data of Doremus [6] (shown by the solid line, with extrapolated data represented by the dotted line), which

exhibits Arrhenius behavior, compared to the fragile liquids ZrO_2 (data of Kondo [7]), Al_2O_3 (data of Paradis [8]) and Si (data of Sato [9]) and hypothetical dashed line.

Our underlying methodology uses a computational science approach to link the physics behind Pair Distribution Function (PDF) measurements [10], obtained from high energy x-ray or neutron diffraction, to the machine learning (ML) based Gaussian Approximation Potential (GAP) method for *ab initio* accurate scalable molecular dynamics simulations [11]. Although only a handful of studies based on this methodology have so far been published [12-15], the method holds promise for extracting not only atomic structures of liquids and glasses, but also electronic, dynamic, and property information on bonding, density, diffusion, and conductivity. We start at the level of classical Molecular Dynamics (MD) simulations, which were designed to solve Newton's classical equations of motion for hard spheres by Alder and Wainwright [16]. It was not until Rahman [17] that these equations were solved for a set of Lennard-Jones particles. Lennard-Jones potentials comprise of a short-range repulsive term between two interacting atoms and an attractive term at moderate distances that decays to zero at long distances. The repulsion term in the interatomic potential determines the distance of closest approach between two atoms, and the minimum in the pair potential is close to the first maximum in the PDF (and the two are coincident at low density) [18]. Many types of classical interatomic potential have been developed for different systems to capture the underlying physics of the interactions between atoms. However, the assumption of a fixed functional form to construct interatomic potential is often inadequate to explain over a range of experimental conditions. To accurately describe the electronic behavior a quantum mechanical method such as density function theory (DFT) is required. However, due to the computational cost, such *ab initio* molecular dynamics (AIMD) simulations are limited to smaller systems and shorter timescales. ML interatomic potentials bridge this gap by directly fitting to DFT reference datasets [19-29].

From an experimental standpoint, carefully analyzed high-energy photon diffraction and spallation neutron PDF measurements provide an ideal metric for discerning between different structural models of disordered materials [10]. Here, we emphasize the need for a critical analysis of liquid and amorphous experimental x-ray data to obtain accurate PDFs, an aspect that, as we have noted in previous work [30], is often overlooked in many modern-day software packages. This point is especially important in the work reported here, which builds on our recent efforts on driving the automated fitting of the GAP method by using experimental PDFs

and active learning [14]. Here the GAP is constructed through an active learning and fitting process that can be initialized by configuration sampling or experimental data, with total energy, E, written as a sum of local energies given by two-body and many body terms given by a nonlinear kernel function [31, 32],

$$E = \sum_{i<j} U^{(2)}(r_{ij}) + \sum_i \sum_s^M \alpha_s K(d_i, d_s) \qquad (1)$$

where $U^{(2)}$ is modeled as a sum of a squared exponential kernel, with r_{ij} the distance between atoms i and j, that describes the short-range interaction within a defined radius. The Smooth Overlap of the Atomic Position (SOAP) kernel, K, captures the many-body interactions. It provides the similarity measured between two atomic neighborhoods, d_i and d_s, where the chemical descriptor, d_i describes the local chemical environment surrounding the i^{th} atom which is defined by a neighbor density at each point in space r for each atom i. The α_s are determined by fitting to the training set. The active learner consists of an unsupervised clustering algorithm combined with Bayesian optimization for on-the-fly hyperparameter tuning of the GAP model. A detailed description of the methodologies can be read elsewhere [32, 33].

In our past studies, we have utilized different configurations sampling methods. The first method considered for modeling bulk liquid structure in high temperature ionic liquids is initialized by classical MD simulations, which are then down-sampled by using the active learning algorithm [13]. DFT calculations are then iteratively performed, and a machine learned interatomic GAP potential developed until a specified energy threshold is reached. Models developed via this method have been shown to provide new physical insights into the temperature-dependent coordination environment of the molten state as well as densities, self-diffusion constants, and ionic conductivities [13, 15]. These results indicate that GAP models are able to capture the many-body interactions required to accurately model ionic systems. This significantly lowers the barrier to the understanding and design of molten salts for example, across the periodic table.

The second method, presented here on the high temperature refractory oxide HfO_2, uses an automated closed loop via an active learner [14, 33], which is initialized by x-ray and neutron diffraction measurements of crystalline phases, and sequentially improves an unsupervised machine-learning model until the experimentally predetermined phase space is covered. The resulting GAP MD simulations reproduce all the experimental phases with near *ab initio* precision and are capable of simulating to much slower quench rates (e.g., 1 K/ps). However,

it is not straightforward to directly incorporate the experimental PDFs of the disordered structures, such as liquids and glasses, into the GAP model training. We seek to address some of these challenges in this study by increasing the configuration sampling via enhanced sampling and by using more accurate reference calculations [34].

2. Machine Learning Interatomic Potentials (ML-IPs).

The underlying premise of this work is straightforward and seeks to address modeling limitations that have dominated the field of determining PDFs for glass, liquid, and amorphous materials for the past several decades. Typically, a Reverse Monte Carlo (RMC) [35] or Empirical Potential Structure Refinement (EPSR) [36] model is used to rearrange spheres of atoms (or constructed molecules) in a box until agreement within the error of the experimental structure factor (and PDF) is obtained. The usefulness of this approach is often limited by the number of experimental data sets (neutron, x-ray) and the chemical constraints that are imposed. Alternatively (or in addition) a MD simulation is performed to predict the underlying forces and bonding arrangements between atoms to predict the structure, dynamics (diffusion), and physical properties (density, viscosity) [37]. More sophisticated DFT calculations yielding both atomic and electronic structure information are usually in poorer agreement with the experimental data but do provide a realistic description of the interactions within the system.

Consequently, we ask three questions: (i) Can accurate (small box) *ab initio* level models be scaled up to larger MD box sizes, to capture structural, dynamic, and physical properties of a disordered material over a wide range of temperatures and pressures ?, as illustrated in Fig. 2 [19]. (ii) Can this process be driven by experimental diffraction data for the crystalline phases? (iii) Can we improve the data selection for disordered phases where experimental diffraction data cannot be infused directly into model generation process? Consider this thought experiment; if thousands of model simulations were carried out, then it is likely that one or more will agree with the measured data and be better placed to describe a particular phenomenon in a system. Here computer science, through machine learning, is used to address the structure of fragile liquids, using efficient active learning algorithms and supercomputing resources to reach that goal in the minimum number of configurations, i.e. with considerably less effort than performing thousands MD simulations.

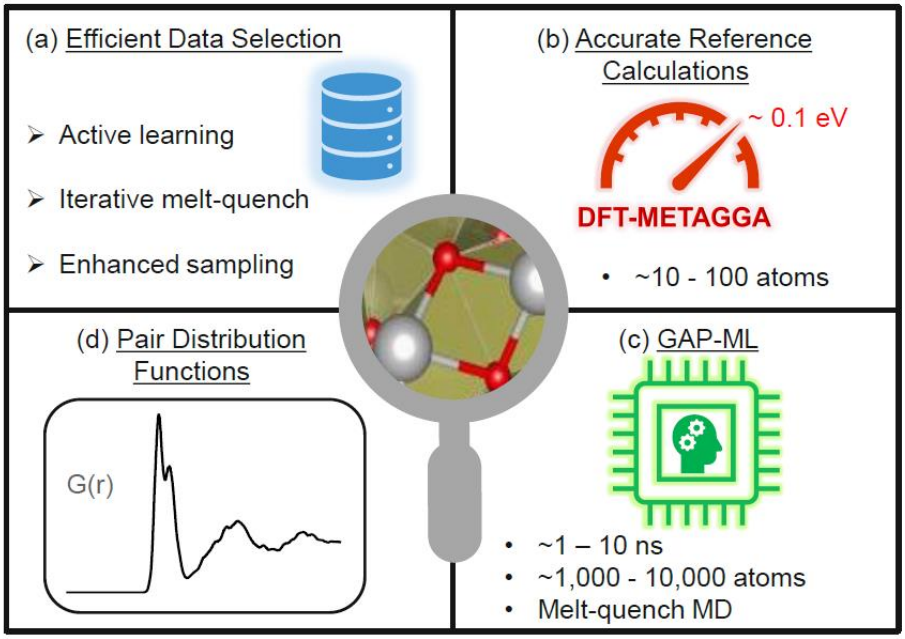


Fig. 2. (Color online) Overview of data-driven modeling of liquids and glasses. (a) Efficient data selection strategies. (b) Density functional theory (DFT) at the level of meta generalized gradient approximation (meta GGA [38]) on curated datasets with size of 10 – 100 atoms. (c) GAP model trained at DFT-METAGGA level of theory. GAP-ML driven large-scale melt-quench MD simulations by scaling to system sizes of 1,000 to 10,000 atoms. d) Rigorous validation of ML-generated PDFs with diffraction experiments.

Here we use an archetypical binary metal oxide, namely HfO_2, to showcase the modeling challenge by using a GAP model focused on the liquid and amorphous forms of HfO_2 with Strongly Constrained and Appropriately Normed (SCAN) level of DFT [39]. SCAN has shown improved predictive performance compared to standard GGA [38] exchange-correlation functionals [40, 41]. ML-IP trained at the level of DFT-SCAN has also shown improved performance on predicting structure, and properties such as melting point, thermal conductivity etc. [42, 43]. For this purpose, we have recomputed the active learned training dataset previously computed with PBE exchange-correlation functionals reported in previous work [33] with DFT-SCAN as implemented in the VASP software [44, 45]. Unless explicitly specified, the number of atoms and general computational settings used through this study are kept consistent with our previous work [14, 33]. Enriching the training database ensures the

model is trained on a diversity of local chemical environments. Strategies such as enhance sampling and random structure search have been shown to improve the diversity of training databases [15, 46, 47]. To further enrich the configuration space with enhanced sampling, we used the newly generated GAP model at DFT-SCAN level of theory to perform configuration sampling with near melting point as reference temperature (~2900^{o}C) using well-tempered variation of metadynamics [46, 48, 49]. The DFT-SCAN based GAP model training parameters are listed in Table 1. The disordered structure is influenced by the Hf-Hf and Hf-O pair coordination environments, which has been chosen intuitively as a collective variable (CV) to drive the exploration [50, 51]. The atom-atom pair coordination CV is parametrized with the values of first minima in the partial PDFs. A Gaussian was deposited every 250 fs with a bias factor equal to 50. The potential is subject to retraining with an actively learned configuration drawn from the metadynamics sampling. The liquid and amorphous HfO_2 results presented in the following sections are recomputed with this potential. The DFT-SCAN based GAP model potential, training data set and setting will be made available.

Parameter	**SOAP**	**Two-body**
Cut off radius (Å)	5.0	5.0
Smooth cut-off transition (Å)	1.0	1.0
Sparse method	Cur points	uniform
Sparse points	1500	32
(n_{max}; l_{max})	(8,6)	-
Kernel exponent	4	-

Table 1. DFT-SCAN based GAP model parameters

3. High Temperature Metal-Oxides.

Understanding the structure and properties of refractory TMOs is critical for high temperature applications. Their high melting temperatures, > 1500 °C, make TMOs suitable for applications in harsh environments, in addition to their insulating properties and ability to prevent oxidation. As a first step, it is therefore important to identify crystalline phase transformations and structural rearrangements close to the melting point. Using our ML methodology based on high energy x-ray data obtained from laser heating bulk samples, a multiphase potential is generated for a canonical example of the archetypal refractory oxide, HfO_2. This is achieved by drawing a small number of training configurations from room temperature to the liquid state at ~2900 °C. The method illustrated in Fig. 3 significantly reduces model development time and human effort [14].

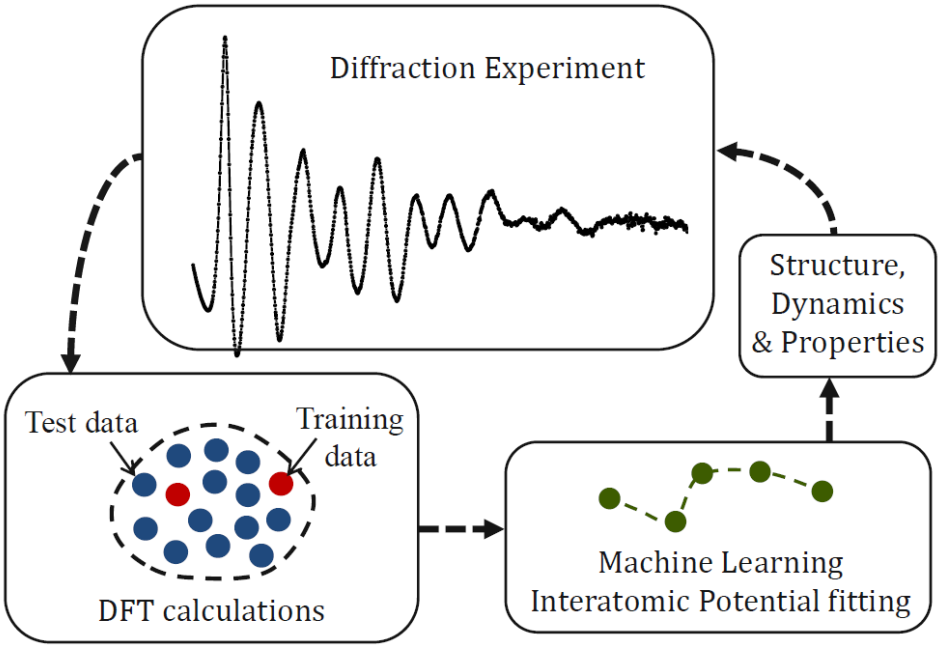


Fig. 3. (Color online) The ML workflow used to develop the Gaussian Approximation Potential (GAP) from diffraction data. The experimental data drives an active learning algorithm to test and train data using a DFT approach, to develop the GAP which is used in classical MD simulations to predict the structure, dynamics, and properties of the system.

X-ray diffraction data were collected at beam line 6-ID-D at the Advanced Photon Source using 60 keV x rays for the high-resolution crystalline measurements, 100 keV for the liquid,

and combining both 60 and 131 keV for the amorphous total scattering experiments [14]. High purity samples of ~2 mm in diameter were laser heated in an aerodynamic levitator to the melting point at in reducing (argon) and oxidizing (oxygen) atmospheres. Approximately 24% laser power from a 400W Synrad Firestar was needed to melt the HfO_2 samples. Diffraction measurements were made as the temperature was decreased in ~100°C increments to the limit of the pyrometer range ~600°C, see Fig. 4. The corresponding amorphous form was made by laser ablation [1].

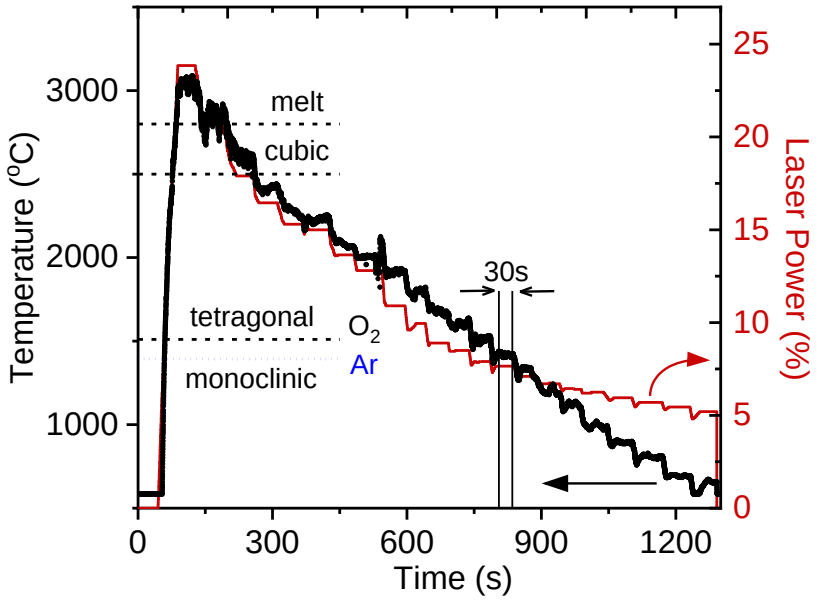


Fig. 4. (Color online) The abrupt temperature ramp and step-wise descent from the melt as a function of time, for a laser heated HfO_2 sample in the aerodynamic levitator. The associated laser power to achieve the temperatures measured is also shown. X-ray scans were taken for 30 seconds once the sample temperature had equilibrated.

The total scattering data analysis for the liquid and amorphous forms has been described previously [1]. Diffraction measurements on levitated polycrystalline materials at high temperatures often consist of Bragg spots from a few oriented crystals. In order to address this preferred orientation issue, a rotating nozzle is used [52] to provide enough instability in the gas flow to spin the sample in the x-ray beam and obtain a better powder average. Even then a full crystallographic refinement is not always possible (which is the case here), so lattice parameters are obtained via LeBail whole pattern fitting, rather than a full Rietveld refinement.

3.1 Crystalline phases.

Traditionally, the interatomic potentials used to describe specific atom-atom interactions in classical MD simulations on glasses are developed by empirically fitting individual, pairwise, and three-body parameters to known crystal properties [53]. Here, the use of multiphase ML-IPs over specified temperature ranges helps to overcome the difficulty in trying to use a single empirical potential to describe multiple states. For HfO_2 the tetragonal to monoclinic transition temperature is often lower upon cooling when compared to heating (i.e. monoclinic to tetragonal) in analogous oxides such as ZrO_2 [54], and has been attributed to the pre-transformational increased cation disorder upon heating, compared to lattice instability as a whole upon cooling. In our experiments we observe that HfO_2 melted in an Ar atmosphere is shifted to lower temperatures, by up ~120^{o}C, when compared to the same samples in an oxygen atmosphere, see Fig. 5. This is significantly larger than our estimated pyrometery error or emissivity changes at this temperature [55], but may be influenced by temperature gradients across the spinning samples. The temperature shift (assuming it is real) is likely due to the presence of oxygen defects amplifying the lattice instability upon cooling, despite their presence being estimated to be <1%, which warrants further investigation. In contrast, the cubic to tetragonal transition is much less well defined because of the co-existence of both phases, even at the highest temperatures.

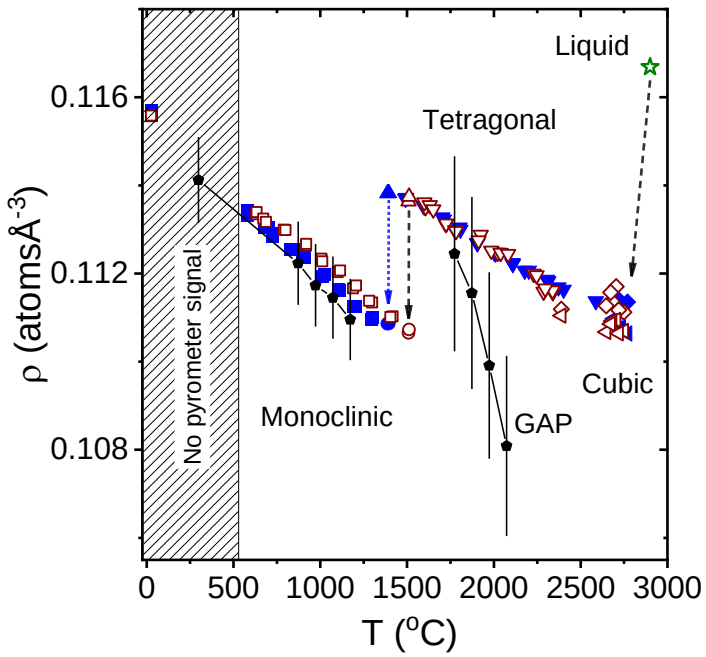


Fig. 5. (Color online) The atomic number density, ρ, of different phases of HfO_2 in Argon (solid blue symbols) and Oxygen atmospheres (open symbols). The pure monoclinic form is represented by squares and the tetragonal form (downward pointing triangles). The GAP densities are shown as solid black hexagons with error-bars. Other symbols are used for mixed phases including the tetragonal plus cubic forms at high temperatures.

To initialize the process, model crystalline structures are fitted to the PDFs and the densities obtained from the experimental unit cell volumes. The crystalline HfO_2 results reported in this study employ the DFT-PBE based potential, with settings reported in [14]. The details of the experiment-driven GAP model generation, are briefly discussed here. AIMD simulations for each phase are performed at the reference temperatures using VASP [56] and cluster based active learning methods are used to generate an initial ML-IP. For the machine learning of potential energy surface, we used the GAP model set up introduced in equation (1). Heating and cooling isothermal-isobaric ensemble MD simulations are used to map the different phases over their known temperature ranges. Cluster-based active learning is used to sub-sample configurations with an isothermal-isobaric ensemble MD trajectory as input. Single point DFT calculations on the actively learned configurations are obtained by iteratively training the ML-IP, until the experimental temperature range is covered. MD simulations were carried out with LAMMPS, compiled with QUIP pair style that uses the GAP model and experimental density

with a larger atom box size of ~6000 atoms [57]. MD simulations at high temperatures are compared to high energy x-ray diffraction structure factors, S(Q). We attribute the improved agreement between experiment and the GAP model versus the initial AIMD model to the improved configurational sampling and larger box size e.g. See [15].

3.2 Liquid Structures.

The total x-ray structure factors of liquid MO_2 as a function of metal ion size are shown in Fig. 6(a). A breakdown of the atom-atom partial structure factors for amorphous HfO_2 from the GAP model in Fig. 6(b) reveals that the M-M correlations dominate the contribution to the first diffraction peak. This result indicates that the changes in the magnitude of the first diffraction peak at Q~2.1Å^{-1} for different liquid TMOs are dominated by the scattering from M-M interactions and their x-ray weighted fractions.

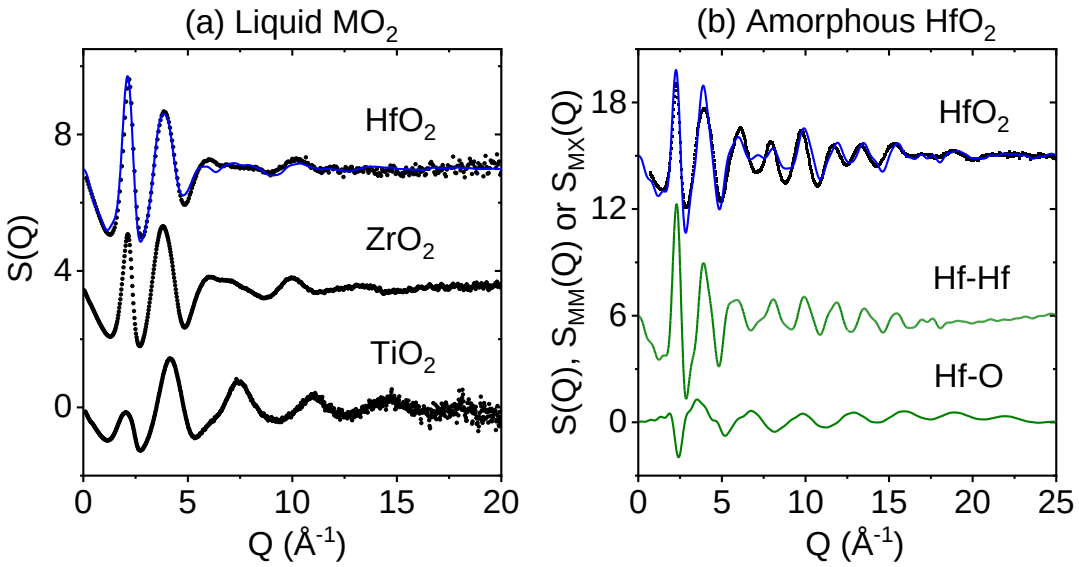


Fig. 6. (Color online) The measured x-ray structure factors of (a) molten MO_2, where M=Hf, Zr and Ti (circles). The line represents the GAP model for HfO_2. (b) Amorphous HfO_2 (circles), compared to the GAP model (line, top) and the contributions from the GAP Hf-Hf and Hf-O x-ray weighted partial structure factors, shifted for clarity.

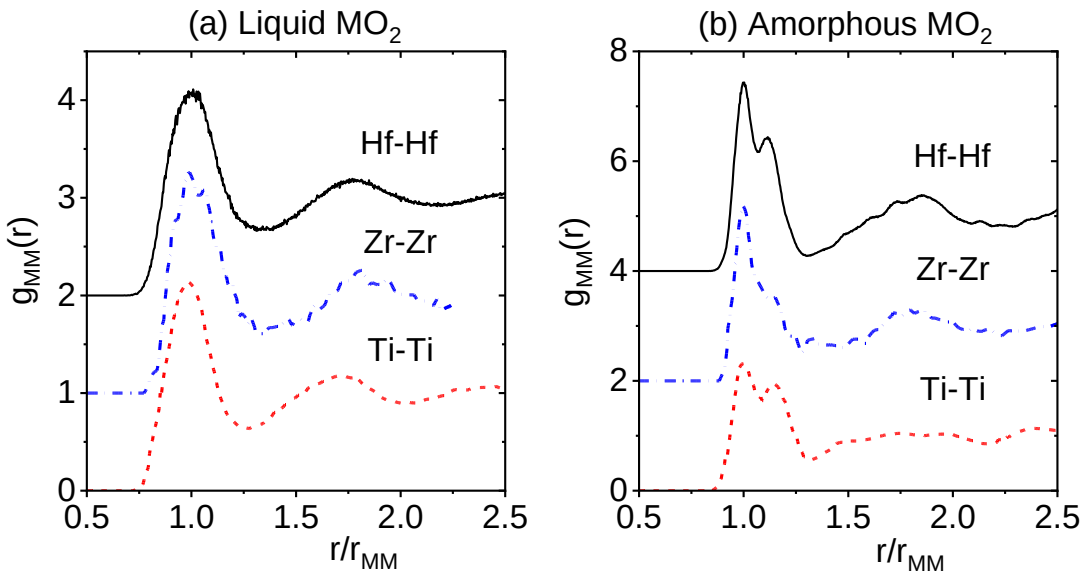


Fig. 7. (Color online) The M-M partial PDFs for (a) liquid MO_2 (Hf-Hf from the GAP model in this paper, Zr-Zr digitized from [58] and Ti-Ti from [59]), and (b) M-M for the amorphous MO_2 forms (Hf-Hf from the GAP model in this paper, Zr-Zr digitized from [2], and Ti-Ti digitized from [3]). The x-axis is scaled by the first metal-metal distance to compare the results from different ion sizes.

The metal-metal partial pair distribution functions for the different *liquid* MO_2 systems considered here have been extracted by various methods i.e. Hf-Hf using the GAP model from this study, Zr-Zr from a large scale density functional molecular dynamics simulation by Kohara *et al.* [58] and Ti-Ti from an Empirical Potential Structure Refinement fit by Alderman *et al.* [59]. These functions, shown in Fig. 7(a), show a high degree of similarity when plotted on a reduced distance scale to account for ion-size. In contrast, the metal-metal partial pair distribution functions for the corresponding *amorphous* forms, given in Fig. 7(b), show distinct peak splitting in the case of Hf-Hf partial (using the GAP model from this study) and the Ti-Ti partial (obtained using Reverse Monte Carlo fit by Petkov *et al.* [3]). The Zr-Zr partial shows a high-r shoulder rather than a distinct peak on the transient amorphous ZrO_2 form measured by Zhang *et al.* [2] (which is likely to be real because it was measured during a time resolved

experiment). Nonetheless, when ion size is accounted for by dividing by the first peak position r_{MM}, the amorphous forms also show a reasonably high degree of similarity. Previously, the first peak in the amorphous M-M partial pair distribution has been assigned with edge-sharing polyhedra and the second peak corresponds to corner-sharing polyhedra [1-3].

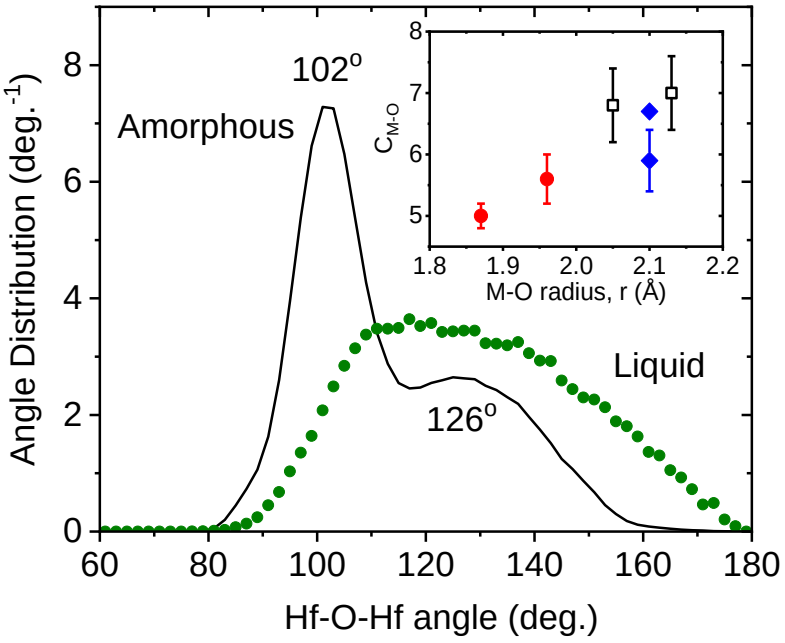


Fig. 8. (Color online) The Hf-O-Hf bond angle distribution for liquid (circles) and amorphous HfO_2 (line) obtained from the GAP model. The insert shows the metal-oxygen coordination number for liquid and amorphous TiO_2 (red circles), ZrO_2 (blue diamonds) and HfO_2 (open squares, this study) versus metal-ion radius.

The Hf-O-Hf bond angle distribution for liquid and amorphous HfO_2 obtained using the GAP model is shown in Fig. 8. The results are consistent with the Hf-Hf partial pair distribution functions shown in Fig. 7, that show two peaks corresponding to distinct edge- and corner-shared polyhedra in the amorphous forms, and a single (much broader) peak in the liquid state. While the average metal-oxygen coordination number increases (close to linearly) with metal ion size (see Fig. 8 insert), the edge-sharing M-O-M angles are found to be similar in all the

amorphous MO_2 cases considered here i.e. close to 100°. However, there is some variation in the M-O-M angles associated with the corner shared polyhedra between the amorphous TiO_2, ZrO_2 and HfO_2 forms, as well as the average angles of the very broad single peak in the liquid states. The liquid Hf-O-Hf bond angle distribution shows a much broader distribution than the amorphous, consistent with a large degree of thermal broadening and polyhedral disorder observed in the partial PDFs. It is also shifted to higher angles in the melt, consistent with a higher proportion of larger rings.

Kohara *et al.* have found the broad, average Zr-O-Zr bond angle distribution in liquid ZrO_2 to be centered on ~100°, which is narrower than the Hf-O-Hf angle of ~120° in liquid HfO_2 obtained using the GAP model. The bond angle distribution for liquid TiO_2 spans this range, showing two broad Ti-O-Ti peaks centered on ~97° and ~123°. TiO_2 is known to be strongly polymorphic with five known metastable forms at ambient conditions [59]. First-principles simulations for amorphous TiO_2 predict a Ti-O-Ti edge sharing bond angle of 98°, but a variation in corner-sharing bond angles between 120-140° depending on the model [60]. Variations are also observed in the experimental results of Petkov *et al.* [3] on amorphous TiO_2 materials synthesized using different processes, leading to a change in the edge/corner-sharing ratio. These authors [3] found that this ratio is directly linked to the crystallization behavior, with the amorphous/crystalline limit about 50%. Structurally, the amorphous form was found to comprise short, staggered chains of edge and corner-shared Ti-O octahedra.

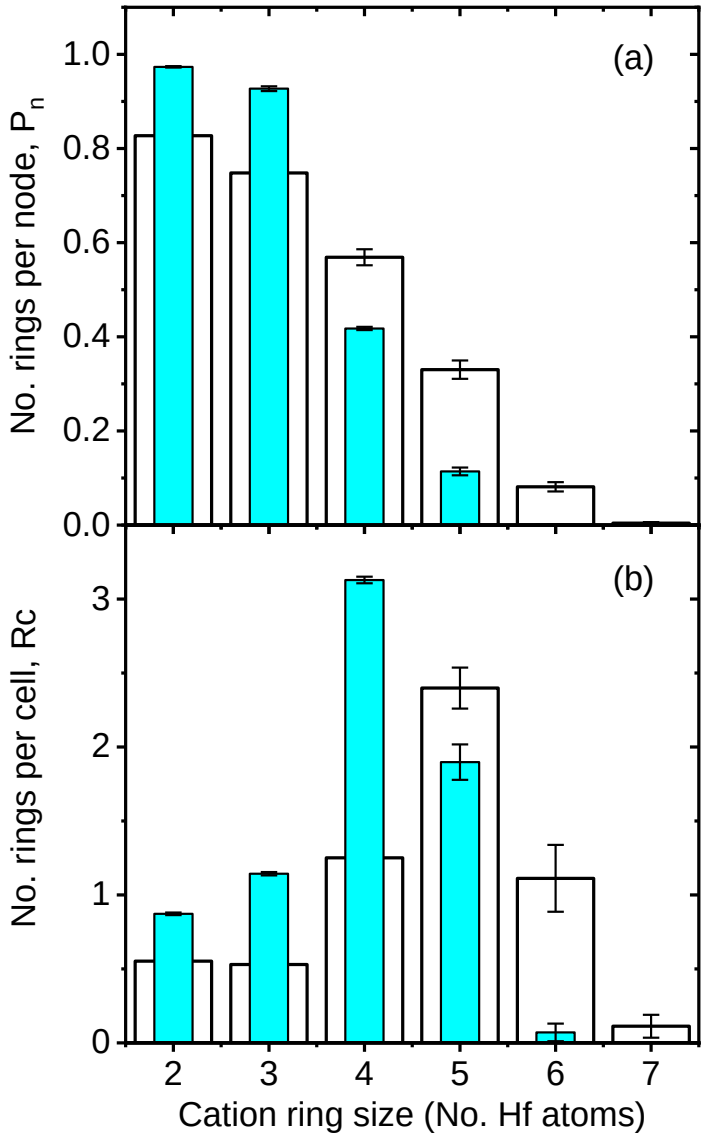


Fig. 9. (Color online) The cation ring size distributions corresponding to the number of Hf atoms in a -Hf-O-Hf- ring, using a cut-off distance corresponding to the first minimum in the Hf-O partial PDF. Liquid HfO_2 (wide open bars, r_{cut}=2.93Å) and amorphous HfO_2 (filled narrow bars, r_{cut}=2.82Å). Error bars indicate standard deviation of 5 model configurations [56].

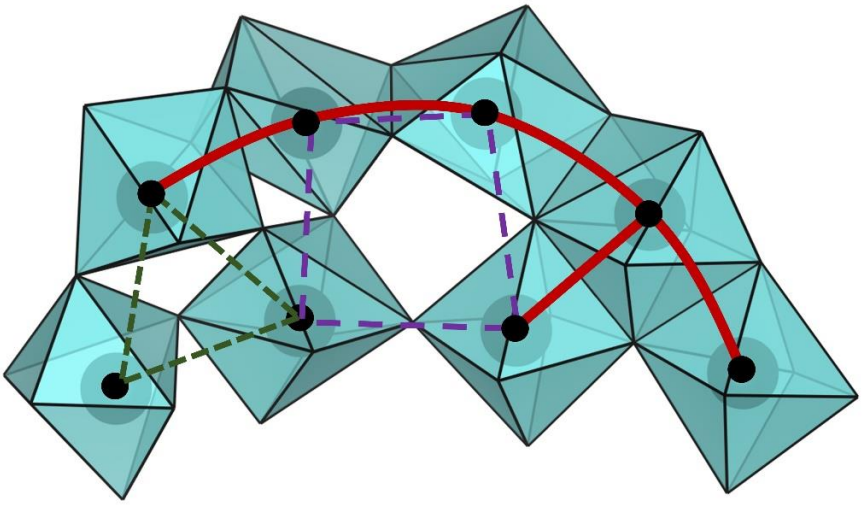


Fig. 10. (Color online) The graphic shows a section of the GAP training model, with edge-shared polyhedra linked by a thick solid line (cation ring size=2), three corner shared units (cation ring size=3) and a four-fold ring (cation ring size=4).

The ring size distributions shown in Fig. 9 correspond to (a) the proportion of atoms in rings of a given size, P_N, and (b) number of rings of a given size per cell, R_C [61]. The ring distributions were calculated using the King's shortest path criterion [62]. P_N provides a metric for how representative of the overall structure a given ring size is. For example, in networks with a significant amount of edge-sharing polyhedra, R_C may indicate a significant number of larger rings, while P_N shows the fraction of atoms that actually participate in those rings e.g., HfO_2 has a large R_C for 5-member rings, but only ~10% of atoms are part of those rings. The GAP model shows the general network connectivity P_N of both liquid and amorphous HfO_2 to be similar, with the highest proportion of Hf atoms involved in 2-fold rings corresponding primarily to edge-sharing (with a few face-sharing) polyhedra, and small sized corner shared rings (mainly 3, 4 and 5-fold rings). The dominant number of 2-fold rings in the liquid is surprising, given the substantial differences in the intensity of the 102° peak in the Hf-O-Hf bond angle distribution between the liquid and amorphous forms, but can be rationalized by the increased thermal disorder and polyhedral distortion at higher temperatures. The liquid HfO_2

GAP model contains 66% corner-, 30% edge and 4% face-sharing polyhedral, compared to 59% corner-, 37% edge- and 4% face-sharing in the amorphous GAP model. The R_C distribution for the liquid also indicates the preference for a modal number of 5-fold rings, compared to 4-fold rings in the amorphous form. The trends in the P_N and R_C ring distributions point toward a higher proportion of larger, corner-shared polyhedral rings in the liquid, which is consistent with the broader bond angle distribution shifted to higher angles in Fig 8. Overall, however, the findings suggest that the topology of both the liquid and amorphous forms of HfO_2 are similar, as are those of ZrO_2 and TiO_2. The structures of these very fragile liquids and amorphous materials comprise branched, winding, edge-shared chains, interlaced with primarily smaller, corner-shared rings as illustrated in Fig. 10.

4. Conclusions

In answer to our initial questions posed in section 2 we have demonstrated that (i) accurate *ab initio* level models can be scaled up to larger MD box sizes, to capture structural, dynamic, and physical properties of a disordered material over a wide range of conditions, and (ii) that the process can be driven by experimental diffraction data for the crystalline phases. This has been demonstrated for crystalline, liquid, and amorphous structures of the high temperature refractory transition metal oxide HfO_2, which have been investigated by using GAP machine learning models. Despite the enhanced thermal broadening at high temperatures, the topology of the fragile, inviscid liquid has been shown to have a similar structure to its amorphous counterpart: namely, the structures of both liquid and amorphous HfO_2 are dominated by winding edge-shared octahedral chains, and branching enables the formation of small rings of corner shared polyhedra. A comparison of HfO_2 to existing studies on liquid and amorphous ZrO_2 and TiO_2 in the literature indicates that this structural topology is likely to be common across other transition metal oxides, somewhat independent of metal ion size.

A novel aspect of this study was our use of a ML algorithm driven by x-ray PDF measurements, with an active learning scheme used to test *ab initio* MD simulations by using a Gaussian Approximation Potential, as an alternative to conventional molecular simulations. This method allowed us to scale up from hundreds of atoms to many thousands of atoms, and

thus to obtain predictions for a wide range of medium range structures, dynamics, and material properties. The approach also enables the insertion of experimentally derived structures directly into the ML-IP training process, and overcomes the fact that many metastable amorphous structures are not always in the lowest energy state. It also provides a means of direct validation of the model via the active learning process to produce a ML-IP within *ab initio* accuracy. Furthermore, we have shown the development of the GAP model for a scenario when experiment PDF cannot directly enter GAP model generation. We have used amorphous and liquid HfO_2 to train a DFT-SCAN accurate GAP model as the phase space enriched with metadynamics. The accuracy of the results is limited only by the choice of quantum chemistry method, and the GAP approximation enables much larger box sizes and quench rates of ~1-3 K/ps not achievable by using AIMD.

To address question (iii) in section 2 and help improve the data selection for disordered phases where experimental diffraction data cannot be infused directly into model generation process, **r**ecent literature on ML-based structure prediction [32, 63-65] contains some lessons on training dataset creation that suggest avenues for further improvements in amorphous phase prediction. The effectiveness of an ML model in arriving at the correct amorphous structure depends on the diversity of the local chemical environments on which the model is trained. Yet a single melt-quench simulation from AIMD typically creates a single kind of amorphous structure, thereby lacking structural diversity. Hence, purely training on single melt-quench AIMD trajectory may not lead to ML predictions arriving at the correct amorphous structure. To mitigate this issue, Deringer *et al.* [32] proposed an iterative melt-quench scheme for generating a diversity of amorphous carbon structure so as to give the GAP model a variety of local chemical environments (e.g. sp/sp^2, sp^3 rich structures). The iterative protocol consists of rapid melt-quench of random packed structures at varying density. Short melt-quenches created a diversity of amorphous structures for single point DFT retraining. Another avenue that can lead to dramatic improvement in structure prediction is to build on the metadynamics sampling used in this study to incorporate more complex local chemical environment dependent CVs [46]. Random structure search can aid in the development of good transparent machine learning

potentials [47]. Further, modeling of non-isotropic chemical environments such as electric fields or interfaces will require explicit inclusion of long-range electrostatics [66]. Finally, the ML-GAP methodology and perspectives presented in this study offers a general scheme for the study of similar families of materials, which could be further adapted using other ML-IP methods such as neural networks.

Acknowledgment

This research was supported by the Exascale Computing Project (17-SC-20-SC), a collaborative effort of the U.S. Department of Energy Office of Science and the National Nuclear Security Administration. This research used resources of the Argonne Leadership Computing Facility and the Advanced Photon Source which are both DOE Office of Science User Facilities supported under Contract DE-AC02-06CH11357. We gratefully acknowledge the computing resources provided for Bebop; a high-performance computing cluster operated by the Laboratory Computing Resource Center at Argonne National Laboratory and the high energy x-ray beamline 6-ID-D at the APS.

*E-mail: benmore@anl.gov

References

1. L.C. Gallington, Y. Ghadar, L.B. Skinner, J.K.R. Weber, S.V. Ushakov, A. Navrotsky, A. Vazquez-Mayagoitia, J.C. Neuefeind, M. Stan, J.J. Low, and C.J. Benmore, Materials, **10**(11), 1290 (2017).
2. F. Zhang, P.J. Chupas, S.L.A. Lui, J.C. Hanson, W.A. Caliebe, P.L. Lee, and S.-W. Chan, Chemistry of Materials, **19**(13), 3118 (2007).
3. V. Petkov, G. Holzhüter, U. Tröge, T. Gerber, and B. Himmel, Journal of Non-Crystalline Solids, **231**(1-2), 17 (1998).
4. C.A. Angell, Science, **267**(5206), 1924 (1995).
5. G.N. Greaves, A. Fontaine, P. Lagarde, D. Raoux, and S.J. Gurman, Nature, **293**(5834), 611 (1981).
6. R.H. Doremus, Journal of Applied Physics, **92**(12), 7619 (2002).

7. T. Kondo, H. Muta, K. Kurosaki, F. Kargl, A. Yamaji, M. Furuya, and Y. Ohishi, Heliyon, **5**(7), e02049 (2019).
8. P.-F. Paradis and T. Ishikawa, Japanese Journal of Applied Physics, **44**(7A), 5082 (2005).
9. Y. Sato, Y. Kameda, T. Nagasawa, T. Sakamoto, S. Moriguchi, T. Yamamura, and Y. Waseda, Journal of Crystal Growth, **249**(3), 404 (2003).
10. C.J. Benmore, ISRN Materials Science, **2012**, 852905 (2012).
11. A.P. Bartók, M.C. Payne, R. Kondor, and G. Csányi, Physical Review Letters, **104**(13), 136403 (2010).
12. A.P. Bartók and G. Csányi, International Journal of Quantum Chemistry, **115**(16), 1051 (2015).
13. S. Tovey, A.N. Krishnamoorthy, G. Sivaraman, J. Guo, C. Benmore, A. Heuer, and C. Holm, Journal of Physical Chemistry C, **124**(47), 25760 (2020).
14. G. Sivaraman, L. Gallington, A.N. Krishnamoorthy, M. Stan, G. Csányi, Á. Vázquez-Mayagoitia, and C.J. Benmore, Physical Review Letters, **126**, 156002 (2021).
15. G. Sivaraman, J. Guo, L. Ward, N. Hoyt, M. Williamson, I. Foster, C. Benmore, and N. Jackson, Journal of Physical Chemistry Letters, (12), 4278 (2021).
16. B.J. Alder and T.E. Wainwright, J. Chem. Phys., **27**, 1208 (2004).
17. A. Rahman, Phys. Rev., **136**, A405 (2021).
18. P.A. Egelstaff, *An Introduction to the Liquid State*. 1994. Oxford University Press.
19. D.-Y.L. L. Zhang, H. Wang, R. Car, and E. Weinan, Physical Review Materials, **3**, 023804 (2019).
20. J. Behler and G. Csányi, The European Physical Journal B, **94**(7), 142 (2021).
21. A.P. Thompson, L.P. Swiler, C.R. Trott, S.M. Foiles, and G.J. Tucker, J. Comput. Phys., **285**(C), 316 (2015).
22. J.S. Smith, O. Isayev, and A.E. Roitberg, Chemical Science, **8**(4), 3192 (2017).
23. K.T. Schütt, H.E. Sauceda, P.J. Kindermans, A. Tkatchenko, and K.R. Müller, The Journal of Chemical Physics, **148**(24), 241722 (2018).
24. V. Botu, R. Batra, J. Chapman, and R. Ramprasad, The Journal of Physical Chemistry C, **121**(1), 511 (2017).
25. A.M. Miksch, T. Morawietz, J. Kästner, A. Urban, and N. Artrith, Machine Learning: Science and Technology, **2**(3), 031001 (2021).
26. S. Batzner, T. Smidt, L. Sun, J. Mailoa, M. Kornbluth, N. Molinari, and B. Kozinsky, *SE(3)-Equivariant Graph Neural Networks for Data-Efficient and Accurate Interatomic Potentials*. 2021.
27. I.S. Novikov, K. Gubaev, E.V. Podryabinkin, and A.V. Shapeev, Machine Learning: Science and Technology, **2**(2), 025002 (2021).
28. T.W. Ko, J.A. Finkler, S. Goedecker, and J. Behler, Accounts of Chemical Research, **54**(4), 808 (2021).

29. F. Musil, M. Veit, A. Goscinski, G. Fraux, M.J. Willatt, M. Stricker, T. Junge, and M. Ceriotti, The Journal of Chemical Physics, **154**(11), 114109 (2021).
30. L.B. Skinner, C.J. Benmore, and J.B. Parise, Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment, **662**(1), 61 (2012).
31. A.P. Bartók, R. Kondor, and G. Csányi, Physical Review B, **87**(18), 184115 (2013).
32. V.L. Deringer and G. Csányi, Physical Review B, **95**(9), 094203 (2017).
33. G. Sivaraman, A.N. Krishnamoorthy, M. Baur, C. Holm, M. Stan, G. Csányi, C. Benmore, and Á. Vázquez-Mayagoitia, npj Computational Materials, **6**(1), 104 (2020).
34. J.P. Perdew and K. Schmidt, AIP Conference Proceedings, **577**(1), 1 (2001).
35. R.L. McGreevy, Journal of Physics: Condensed Matter, **13**(46), R877 (2001).
36. A.K. Soper, Physical Review B, **72**, 104204 (2021).
37. M.P. Allen and D.J. Tildesley, *Computer Simulation of Liquids*. 2017. University Scholarship Online.
38. J.P. Perdew, J. Tao, V.N. Staroverov, and G.E. Scuseria, The Journal of Chemical Physics, **120**(15), 6898 (2004).
39. J. Sun, A. Ruzsinszky, and J.P. Perdew, Physical Review Letters, **115**(3), 036402 (2015).
40. M. Rang and G. Kresse, Physical Review B, **99**(18), 184103 (2019).
41. E.B. Isaacs and C. Wolverton, Physical Review Materials, **2**(6), 063801 (2018).
42. J. Guo, L. Ward, Y. Babuji, N. Hoyt, M. Williamson, I. Foster, N. Jackson, C. Benmore, and G. Sivaraman, *A Composition-Transferable Machine Learning Potential for LiCl-KCl Molten Salts Validated by HEXRD*. 2022, American Chemical Society (ACS).
43. D. Tisi, L. Zhang, R. Bertossa, H. Wang, R. Car, and S. Baroni, Physical Review B, **104**(22), 224202 (2021).
44. G. Kresse and J. Furthmüller, Computational Materials Science, **6**(1), 15 (1996).
45. G. Kresse and J. Furthmüller, Physical Review B, **54**(16), 11169 (1996).
46. L. Bonati and M. Parrinello, Physical Review Letters, **121**(26), 265701 (2018).
47. N. Bernstein, G. Csányi, and V.L. Deringer, npj Computational Materials, **5**(1), 99 (2019).
48. A. Laio and M. Parrinello, Proceedings of the National Academy of Sciences, **99**(20), 12562 (2002).
49. A. Barducci, G. Bussi, and M. Parrinello, Physical Review Letters, **100**(2), 020603 (2008).
50. G.A. Tribello, M. Bonomi, D. Branduardi, C. Camilloni, and G. Bussi, Computer Physics Communications, **185**(2), 604 (2014).
51. M. Bonomi, G. Bussi, C. Camilloni, G.A. Tribello, P. Banáš, A. Barducci, M. Bernetti, P.G. Bolhuis, S. Bottaro, D. Branduardi, R. Capelli, P. Carloni, M. Ceriotti, A. Cesari, H. Chen, W. Chen, F. Colizzi, S. De, M. De La Pierre, D. Donadio, V. Drobot, B. Ensing, A.L. Ferguson, M. Filizola, J.S. Fraser, H. Fu, P. Gasparotto, F.L. Gervasio, F. Giberti, A. Gil-Ley, T. Giorgino, G.T. Heller, G.M. Hocky, M. Iannuzzi, M. Invernizzi, K.E. Jelfs, A. Jussupow, E. Kirilin, A. Laio, V. Limongelli, K. Lindorff-Larsen, T. Löhr, F. Marinelli, L.

Martin-Samos, M. Masetti, R. Meyer, A. Michaelides, C. Molteni, T. Morishita, M. Nava, C. Paissoni, E. Papaleo, M. Parrinello, J. Pfaendtner, P. Piaggi, G. Piccini, A. Pietropaolo, F. Pietrucci, S. Pipolo, D. Provasi, D. Quigley, P. Raiteri, S. Raniolo, J. Rydzewski, M. Salvalaglio, G.C. Sosso, V. Spiwok, J. Šponer, D.W.H. Swenson, P. Tiwary, O. Valsson, M. Vendruscolo, G.A. Voth, A. White, and P.c. The, Nature Methods, **16**(8), 670 (2019).

52. C.J. Benmore and J.K.R. Weber, Advances in Physics: X, **2**(3), 717 (2017).
53. A. Cormack and Y. Cao, Molecular Engineering, **6**(1-2), 183 (1996).
54. R.N. Patil and E.C. Subbarao, Acta Crystallographica Section A, **A25** (2021).
55. J.K.R. Weber, A. Tamalonis, C.J. Benmore, O.L.G. Alderman, S. Sendelbach, A. Hebden, and M.A. Williamson, Review of Scientific Instruments, **87**(7), 073902 (2016).
56. G. Kresse and J. Hafner, Physical Review B, **47**(1), 558 (1993).
57. S. Plimpton, J. Comp. Phys., **117**, 1 (1995).
58. S. Kohara, J. Akola, L. Patrikeev, M. Ropo, K. Ohara, M. Itou, A. Fujiwara, J. Yahiro, J.T. Okada, T. Ishikawa, A. Mizuno, A. Masuno, Y. Watanabe, and T. Usuki, Nature Communications, **5**(1), 5892 (2014).
59. O.L.G. Alderman, L.B. Skinner, C.J. Benmore, A. Tamalonis, and J.K.R. Weber, Physical Review B, **90**(9), 094204 (2014).
60. B. Prasai, B. Cai, M.K. Underwood, J.P. Lewis, and D.A. Drabold, Journal of Materials Science, **47**(21), 7515 (2012).
61. S. Le Roux and P. Jund, Computational Materials Science, **49**(1), 70 (2010).
62. S.V. King, Nature, **213**(5081), 1112 (1967).
63. F.C. Mocanu, K. Konstantinou, T.H. Lee, N. Bernstein, V.L. Deringer, G. Csányi, and S.R. Elliott, The Journal of Physical Chemistry B, **122**(38), 8998 (2018).
64. V.L. Deringer, N. Bernstein, A.P. Bartók, M.J. Cliffe, R.N. Kerber, L.E. Marbella, C.P. Grey, S.R. Elliott, and G. Csányi, The Journal of Physical Chemistry Letters, **9**(11), 2879 (2018).
65. J. Mavračić, F.C. Mocanu, V.L. Deringer, G. Csányi, and S.R. Elliott, The Journal of Physical Chemistry Letters, **9**(11), 2985 (2018).
66. C.G. Staacke, H.H. Heenen, C. Scheurer, G. Csányi, K. Reuter, and J.T. Margraf, ACS Applied Energy Materials, **4**(11), 12562 (2021).