Editors' Suggestion

Experimentally Driven Automated Machine-Learned Interatomic Potential for a Refractory Oxide

Ganesh Sivaraman[*]
Leadership Computing Facility, Argonne National Laboratory, Lemont, Illinois 60439, USA

Leighanne Gallington
X-ray Science Division, Argonne National Laboratory, Lemont, Illinois 60439, USA

Anand Narayanan Krishnamoorthy
Helmholtz-Institute Munster: Ionics in Energy Storage (IEK-12), Forschungszentrum Julich GmbH,
Corrensstrasse 46, 48149 Munster, Germany

Marius Stan
Applied Materials Division, Argonne National Laboratory, Lemont, Illinois 60439, USA

Gábor Csányi
Department of Engineering, University of Cambridge, Trumpington Street, Cambridge CB2 1PZ, United Kingdom

Álvaro Vázquez-Mayagoitia
Computational Science Division, Argonne National Laboratory, Lemont, Illinois 60439, USA

Chris J. Benmore[†]
X-ray Science Division, Argonne National Laboratory, Lemont, Illinois 60439, USA

(Received 9 September 2020; accepted 17 February 2021; published 14 April 2021)

Understanding the structure and properties of refractory oxides is critical for high temperature applications. In this work, a combined experimental and simulation approach uses an automated closed loop via an active learner, which is initialized by x-ray and neutron diffraction measurements, and sequentially improves a machine-learning model until the experimentally predetermined phase space is covered. A multiphase potential is generated for a canonical example of the archetypal refractory oxide, HfO_2, by drawing a minimum number of training configurations from room temperature to the liquid state at ~2900 °C. The method significantly reduces model development time and human effort.

DOI: 10.1103/PhysRevLett.126.156002

Refractory oxides are essential components in the development of high temperature ceramic materials [1], thermal barrier coatings [2], and nuclear applications [3,4]. Their high melting temperatures, $T_m > 1500\,°\mathrm{C}$, make refractories suitable for applications in harsh environments, in addition to their insulating properties and ability to prevent oxidation. It is therefore important to identify phase transformations and structural rearrangements close to the melting point. Diffraction plays an important role in the computing of phase diagrams and thermochemistry using the CALPHAD method, which has been the foundation for providing a consistent picture of the stable structures and thermodynamic properties of materials through the calculation of the Gibbs free energy. X-ray powder diffraction in particular is a workhorse for materials characterization, providing data on crystallographic phases, thermal expansion, and volume changes associated with phase transitions in different atmospheres. Neutron powder diffraction also provides valuable structural information, especially on lighter elements such as oxygen, but generally requires larger samples and longer count times. However, there are few suitable containers for x-ray and neutron diffraction experiments at temperatures >2000 °C. In the last decade, advances in aerodynamic levitation and laser heating techniques combined with high-energy x-ray and neutron diffraction have pushed crystallographic measurements above 1500 °C [5,6] providing accurate structural data over a wide range of phase space.

On the computational modeling front, *ab initio* molecular dynamics simulations (AIMD) provide atomic scale resolution with quantum mechanical accuracy, but are restricted to short simulation times, and small system sizes. Empirical interatomic potentials based on fixed analytical functional forms are derived from physical or chemical intuitions and parametrized to experimental properties or from *ab initio* calculations, but lack the sophistication to capture the many-body interactions required to arrive at *ab initio* accuracies. In recent years, advances in combining quantum-mechanical calculations with machine learning has resulted in a new class of interatomic potentials that learns the potential energy surface landscape directly from reference *ab initio* datasets [7–11]. Machine learning interatomic potentials (ML-IP) can maintain near *ab initio* accuracy while affording atomic resolution at larger system sizes (through linear scaling) and time scales comparable to classical interatomic potentials [12,13]. In particular, ML-IP based on the Gaussian approximation potential (GAP) [14] have been successfully applied to model liquids [15,16], crystals [17], defects [16], amorphous [18], multicomponent materials [19], and molecules [20]. Training ML-IP requires efficiently drawing configurations from a wide chemical space of interest and finding the best hyperparameters. Active learning is a subdomain of machine learning where an unsupervised machine learning arrives at an optimal supervised machine learning model (i.e., ML-IP) with a minimum number of training configurations [21]. Smith *et al.* proposed a “query by committee” strategy, which is an active learning strategy that exploits disagreement in ensemble of ML-IP model by sampling regions of chemical space where the ML-IP fails to predict the potential energy accurately [22]. Podryabinkin *et al.* used an active learning strategy based on a “D-optimality” criterion for selecting atomic configurations [23]. Zhang *et al.* employed a deep potential generator to efficiently sample configuration space, and generate an accurate reference dataset from the configuration with low prediction accuracy, and perform iterative training [24]. Active learning strategies based on Bayesian inference have also been reported [19,25,26]. We recently reported an active learner that relied on exploiting the cluster structure embedded in a given unlabeled atomic configuration so as to arrive at a minimum number of training configurations [15,27]. Here we propose to bring together the advances in experiments at extreme conditions and theoretical modeling through a closed loop active learning scheme as shown in Fig. 1. Our scheme consists of three components: (1) Experimental measurements are performed up to the melting temperature on a refractory oxide sample. Model structures are fitted to the neutron and x-ray diffraction measurements of each of the phases at different reference temperatures. *In situ* high energy x-ray diffraction is used to obtain unit cell volume as a function of temperature. (2) An active learning scheme initialized by the model structures drives the phase space exploration over the experimental measurement region. (3) A ML-IP is generated that can be iteratively improved by the active learning scheme. To illustrate this approach, we consider an archetypal refractory oxide, hafnium dioxide, HfO_2 (which is isostructural with the most studied ceramic ZrO_2). Upon heating, HfO_2 undergoes transformations from monoclinic (m-HfO_2) to tetragonal (t-HfO_2) to cubic (c-HfO_2) phases before melting at ∼2800 °C [28,29].

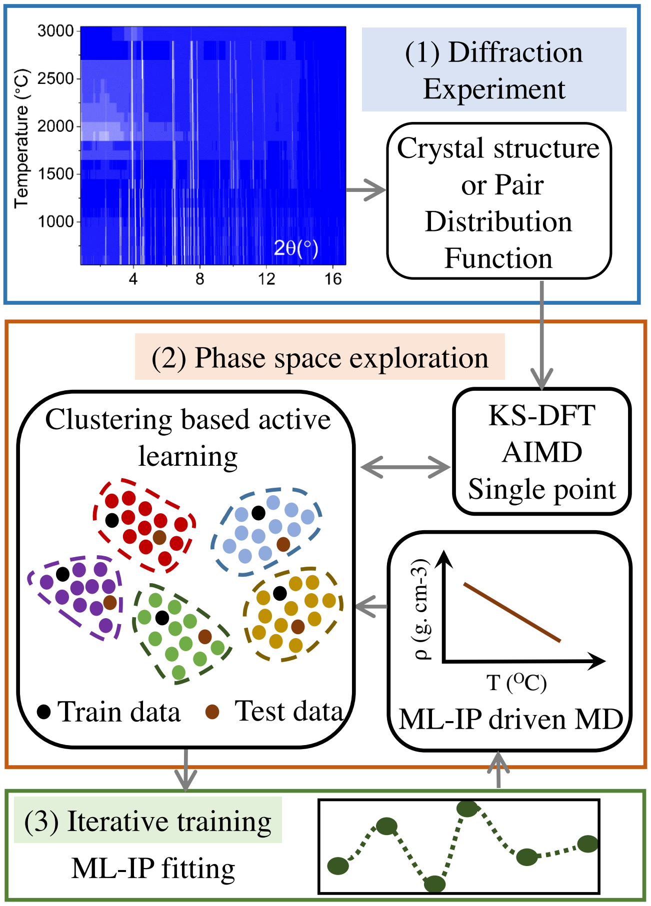

FIG. 1. The experiment driven workflow. (1) Experimental high energy x-ray and neutron diffraction patterns are measured over a wide temperature range using a uniaxial laser heating system on an aerodynamically levitated HfO_2 sample. (2) Cluster-based active learning enables exploration over a wide range of phase space. (3) Iterative training and fitting methods provide feedback into (2).

The experiment driven workflow is shown in Fig. 1. X-ray diffraction data were collected at beam line 6-ID-D at the Advanced Photon Source, Argonne National Laboratory on an amorphous silicon area detector (PE-XRD1621) using 60.07 keV ($\lambda = 0.2064$ Å) x rays. High purity samples (Aldrich, 99.995% trace metal purity) of ∼2 mm diam were levitated and heated up to ∼3000 °C in reducing (argon) and oxidizing (oxygen) atmospheres [29]. Calibration of the detector distance, beam center, detector tilt, and rotation were performed using the FIT2D software package based on the measurement of a CeO_2 NIST standard [30]. Reduction of the 2D images to 1D diffraction patterns yielded the x-ray intensities,$I_{\mathrm{XRAY}}(Q)$. Lattice parameters were obtained via LeBail whole pattern fitting of the previously reported monoclinic ($P2_1/c$), tetragonal ($P4_2/nmc$), and cubic ($Fm\text{-}3m$) crystal structure models to the diffraction data [31–33]. The volumes obtained were normalized to the number of HfO_2 formula units per unit cell to aid in the comparison of the cubic and

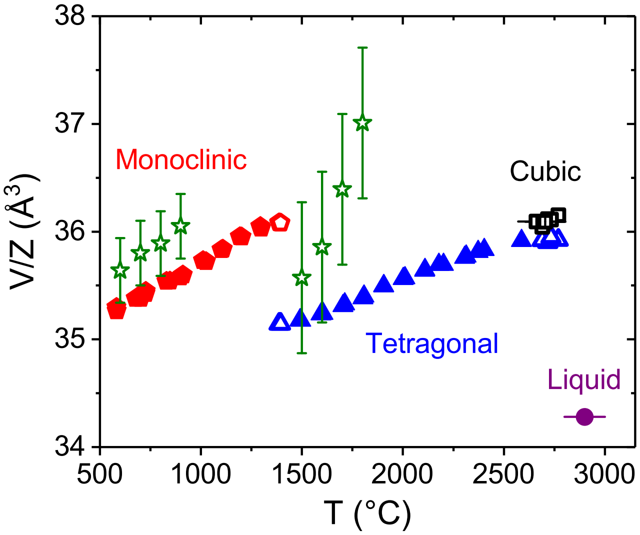

FIG. 2. Unit cell volume of the monoclinic (pentagon), tetragonal (triangles), and cubic (squares) forms of hafnia measured in an argon atmosphere. Open symbols represent mixed phases and solid symbols are single phase. The cubic form was only observed as a mixed phase with the tetragonal polymorph. The average cell volume estimates from GAP MD simulation for m-HfO_2 and t-HfO_2 with an isothermal-isobaric ensemble are shown as green stars with error bars.

monoclinic unit cell ($Z = 4$) volumes to that of the tetragonal phase ($Z = 2$). The phase transitions from monoclinic-tetragonal-cubic liquid in an argon atmosphere with increasing temperature are shown in Fig. 2. A deviation of ~0.2% to lower V is observed in an oxygen atmosphere for the monoclinic phase for temperatures >600 °C and the phase transition to tetragonal occurs at 1400–1500 °C depending on redox environment [34]. However, the cubic and tetragonal phase volumes are essentially the same in both Ar and O_2 [34]. Figures 3(a) and 3(b) shows the x-ray data for the high temperature crystalline phases and the liquid and amorphous forms. The latter $S_{XRAY}(Q)$ experimental data have previously been reported in Ref. [29] but are shown here to show the extent to which this multiphase potential has been trained.

Complementary neutron diffraction measurements were performed on the NOMAD beam line at the Spallation Neutron Source (Oak Ridge National Laboratory). Data were acquired for each of the crystalline phases of HfO_2, i.e., monoclinic at $T \sim 1000$ °C, tetragonal at ~1850 °C and cubic at ~2900 °C in argon and in a 80%Ar: 20%O_2 mixture using a laser-heated aerodynamic levitator [35]. The time-of-flight neutron data were reduced using in-house software [36] to extract the pair distribution functions,$G_{NEUTRON}(r)$. Neutron pair distribution functions for hafnia in the monoclinic, tetragonal, and cubic + tetragonal forms. Neutron levitation experiments are considerably more difficult that x rays due to the lower signal or background ratio and long count times required.

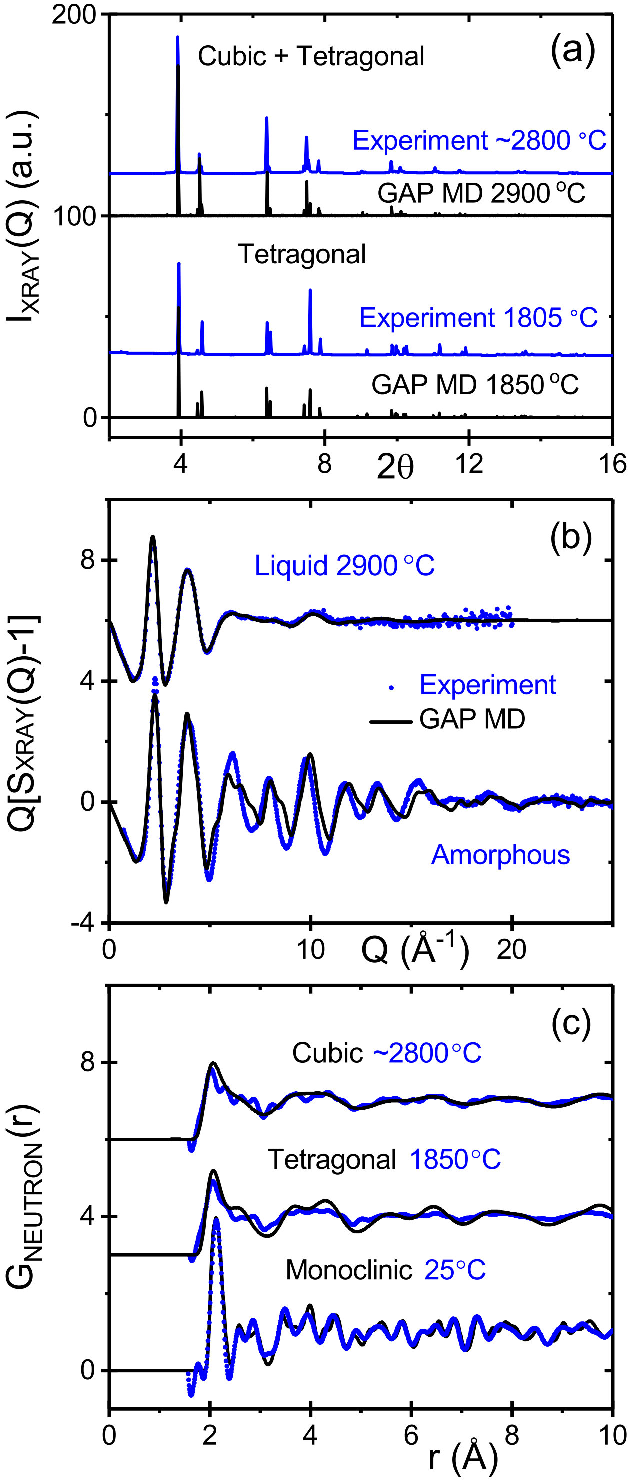

FIG. 3. Diffraction versus simulation data for (a) x-ray diffraction patterns compared to the GAP MD computed x-ray intensity ($\lambda = 0.123$–59 Å) for the two high temperature phases. (b) Experimental and simulated x-ray structure factors for amorphous and liquid HfO_2. (c) Experimental and simulated neutron pair distribution functions for the pure phases of HfO_2.

However, neutrons are more sensitive to oxygen correlations than x rays, which are important for understanding defects and diffusion; i.e., at $Q = 0$ Å^{-1} the O-O neutron partial weighting factor is 36% compared to 3% for x rays, and the Hf-O partial is 48% (neutrons) compared to 30% (x rays).

The second section in the workflow illustrated in Fig. 1 is the experimentally driven phase space exploration: Here we implement in the closed-loop, active learning, phase space exploration in two steps. In the initialization step, active learning starts from the model structures, and generates the ML-IP model from the ensemble of AIMD structures at the neutron diffraction reference temperatures. This initial ML-IP model corresponds to a poor approximation of the potential energy surface and is used to perform ensemble of isothermal-isobaric molecular dynamics corresponding to regions unexplored in the experiment. The active learning subsamples data from those MD simulation trajectories to perform *ab initio* single point calculation and iteratively retrain the ML-IP model. The advantage here is that the AIMD is no longer required beyond the initialization; ML-IP will give access to heating or cooling rates that are beyond the accessible limit of the AIMD [18]. The active learning ensures that only a small sample of DFT single points are required to arrive at the target accuracy. In practice, the closed loop can be initialized by a few AIMD or single point DFT configurations. The active learning exits the closed loop once the required phase space coverage is achieved. The active learning phase space exploration region corresponds to heating m-HfO_2 from 25 to 1500 °C, cooling t-HfO_2 from 1850 to 1400 °C, heating t-HfO_2 from 1850 to 2400 °C and cooling c-HfO_2 from 2900 to 2300 °C. A step-by-step explanation of the closed loop process with an example illustration is provided in the Supplemental Material D [34].

The active learning process is built on a recently proposed scheme based on an unsupervised clustering method coupled to a Bayesian optimization (BO) [15]. The unsupervised clustering method uses the HDBSCAN algorithm to partition the input trajectory and sequentially samples sparse configurations for training the ML-IP [37,38]. The Bayesian optimization performs on-the-fly hyperparameter optimizations, to find the optimal ML-IP model by training on the sampled configurations and validating an independently sampled test dataset [39]. The advantage of this approach is that BO also provides the optimal hyperparameters on the fly. Previously [15], the active learning method has been applied to a very large AIMD “melt-quench” dataset of 33 000 configurations of liquid and amorphous hafnia. In our previous study [15], we demonstrated that for a single phase of hafnia the scheme is able to arrive at a near *ab initio* accurate training dataset with only 0.8%, i.e., 260 samples from this large dataset. For the ML-IP, we use the GAP model along with the many-body smooth overlap of atomic positions (SOAP) descriptor [14,40]. The details of the GAP model and the descriptor are further discussed in the Supplemental Material, Sec. B [34]. The active learning scheme is further discussed in the Supplemental Material, Sec. D [34] and the code implementation with examples usage with GAP model are available elsewhere [15].

The *ab initio* calculations were performed using density functional theory (DFT) as implemented in the VASP package [41,42]. The Perdew-Burke-Ernzerhof generalized gradient approximation and projector augmented plane wave methods were employed [43,44], with a 520 eV plane wave cutoff and $2 \times 2 \times 2$ K grid. A 1 fs time step and Nosé-Hoover thermostat were used for the AIMD [35,36]. For the iterative training of the GAP model, a system size of 96 atom was employed, except for t-HfO_2 where a 108-atom system size was used. An ensemble of AIMD simulations were performed for 12 ps starting from the pure phases based on the model structures at the neutron diffraction reference temperatures. The active learning was initialized with the last 6000 snapshots from each of the AIMD trajectories. Using the initial active learned GAP model, isothermal-isobaric ensemble sampling was performed with the LAMMPS simulation package compiled with the QUIP pair style support [45–47]. The single point DFT calculations employ the same DFT parameters as discussed above. The training dataset generation and isothermal-isobaric quench simulation setup for amorphous and liquid HfO_2 have been discussed in detail elsewhere [15]. In this work, the simulation for both the liquid and amorphous forms have been recalculated with the multiphase potential to show the entire phase space.

The training of the ML-IP based on the GAP model was iteratively mapped by the active learning, until a uniform coverage was achieved across experimental phase space. The active-learned multiphase potential provides a model that spans the entire phase space regions from the liquid to amorphous and crystalline states of HfO_2 with a meager 2053 configurations, the details of which are summarized in the Supplemental Material, Table C1 [34]. The parameters used for training the multiphase potential are summarized in the Supplemental Material, Table C2 [34]. A nonparametric two-body term was added to the SOAP descriptor to prevent nonphysical clustering of atoms at high temperatures [48]. The multiphase potential was validated on a randomly drawn DFT configuration (i.e., outside of training dataset) from the entire phase space region and gave a mean absolute error in energy of 2.4 meV/atom.

For the GAP MD based production simulation, a 6144-atom simulation cell was used for both m-HfO_2 and c-HfO_2. For t-HfO_2, a 6912-atom system was used. The trajectories were sampled for 1.1 ns at the reference neutron diffraction temperatures. The first 100 ps were omitted and the subsequent 1 ns trajectory was used for the analysis. The structural arrangements for all the phases of

HfO_2 are shown in Fig. 3. Figure 3(a) shows good agreement between the Hf-dominated experimental x-ray intensities and GAP MD simulations for the two high temperature crystalline phases [33,49,50]. Similarly, Fig. 3(c) compares the oxygen sensitive neutron diffraction patterns for the monoclinic, tetragonal, and cubic forms of HfO_2. The structure factors of the simulated liquid and amorphous form are shown in Fig. 3(b). Long-range ordering was found to be diminished considerably with increasing temperature in both the tetragonal and cubic forms, and this increased disorder at high temperatures is captured by the machine learned GAP model. The effect of both disorder and density is also seen in the $S(Q)$'s for the amorphous and liquid phases. Here strong oscillations in the lower density amorphous signal at high Q ($\sim$5–15 $Å^{-1}$) correspond to the edge or corner sharing ratio but these are washed out in the liquid signal. The computation of PDF, structure factor, and x-ray intensities are further discussed in the Supplemental Material [34], Secs. F and H.

In order to assess the quality of the reported multiphase potential with two well-known parametrizations for HfO_2, a comparison of cohesive energy and diffusion coefficients are presented. Since the focus is on experiments, the theoretical validation of multiphase potential is restricted to this comparison. The cohesive energies of m-HfO_2, t-HfO_2, and c-HfO_2 computed by different methods are shown in Supplemental Material, Table I1 [34]. The DFT computed cohesive energies reproduce the correct phase order of the phases. The GAP predicted cohesive energy shows the closest agreement with DFT, followed by charge-optimized many-body potential (COMB) parametrized for the hafnium–hafnium-oxide system [51]. The well-known (classical MD) parametrization for HfO_2 by Broglia *et al.* [52], shows a large deviation with respect to DFT. To further test the quality of the multiphase potential, 50 random configurations from the m-HfO_2 AIMD trajectory were drawn and the forces are computed using our method, compared to COMB and Broglia *et al.* The resulting force validation plot with respect to DFT is shown in the Supplemental Material, Fig. E1 [34], and indicates that the GAP (0.09 eVÅ) gives the lowest root mean square error in predicted forces with *ab initio* accuracy, significantly outperforming COMB (4.23 eVÅ) and Broglia *et al.* (10.85 eVÅ) interatomic potentials.

The diffusion constants were also calculated from our molecular dynamics simulations via the mean square displacements of atoms, see Supplemental Material [34], Sec. J. Our simulation results for m-HfO_2 and t-HfO_2 structures show negligible diffusion at simulation temperatures (25, 1850 °C) for both Hf and O. Furthermore, we find the diffusion constants of c-HfO_2 to be $D_{\mathrm{Hf}} = 0.12 \pm 0.002 \times 10^{-6}\ \mathrm{cm}^2/\mathrm{sec}$ for Hf and $D_O = 1.53 \pm 0.005 \times 10^{-5}\ \mathrm{cm}^2/\mathrm{sec}$ for O. Similarly, for liquid HfO_2, the diffusion constants yield $D_{\mathrm{Hf}} = 3.3796 \pm 0.1 \times 10^{-5}\ \mathrm{cm}^2/\mathrm{sec}$ and $D_O = 6.2971 \pm 0.1 \times 10^{-5}\ \mathrm{cm}^2/\mathrm{sec}$, respectively. These values are in good agreement with previous simulation results reported by Hong *et al.* [53]. Furthermore, c-HfO_2 shows negligible diffusion for Hf compared to O. However, with increasing temperature we observe a strong diffusion of Hf atoms in liquid HfO_2 [15], comparable to that of O atoms. Our multiphase interatomic potential also accurately captures the melting point of HfO_2 as described in the Supplemental Material [34], Sec. K.

In conclusion, we show the proof of concept for an automated, experimentally driven scheme for generating a multiphase ML-IP for a canonical refractory oxide, namely, HfO_2. The approach offers the following distinct advantages: (1) The stable structures for initializing *ab initio* calculations can be obtained from convex hull construction or from evolutionary strategies such as USPEX [56,57]. However, the experimentally synthesized structures are not necessarily the lowest energy structure [58]. Our approach bypasses this problem by enabling experimental model structures to directly enter the ML-IP training process. (2) The process removes the ambiguity of sampling phase space required to train the ML-IPs by direct interfacing with experimental measurements. (3) It provides a direct validation of the model with experimental measurement. (4) Active learning ensures that sparse configurations are required to arrive at an ML-IP within *ab initio* accuracy. The results indicate the multiphase potential is able to reproduce both the structural and dynamical properties of HfO_2 from room temperature to the melt with *ab initio* accuracy. The accuracy of the simulated results are only limited by the choice of the *ab initio* method used for generating the training data and can be systematically improved by choosing more accurate quantum chemistry techniques. Although for this particular application the method involved the Gaussian approximation potential framework for generating the model, the proposed scheme could be generalized to other ML-IP methods. Finally, the automation scheme offers a systematic pathway for investigation of other refractory oxides and similar classes of materials.

This material is based upon work supported by Laboratory Directed Research and Development (LDRD) funding from Argonne National Laboratory, provided by the Director, Office of Science, of the U.S. Department of Energy under Contract No. DE-AC02-06CH11357. This research used resources of the Argonne Leadership Computing Facility, which is a DOE Office of Science User Facility supported under Contract No. DE-AC02-06CH11357. Argonne National Laboratory's work was supported by the U.S. Department of Energy, Office of Science, under Contract No. DE-AC02-06CH11357. We gratefully acknowledge the computing resources provided on Bebop; a high-performance computing cluster operated by the Laboratory Computing Resource Center at Argonne National Laboratory. This research used resources of the Advanced Photon Source, a U.S. Department of Energy

(DOE) Office of Science User Facility operated for the DOE Office of Science by Argonne National Laboratory under Contract No. DE-AC02-06CH11357 and the Spallation Neutron Source operated by Oak Ridge National Laboratory. Use of the Center for Nanoscale Materials, an Office of Science user facility, was supported by the U.S. Department of Energy, Office of Science, Office of Basic Energy Sciences, under Contract No. DE-AC02-06CH11357. G. S. would like to thank Dr. Felix Cosmin Mocanu for fruitful discussions on fitting the Gaussian Approximation Potential. A. N.K gratefully acknowledges useful discussions with Professor Christian Holm, Dr. Jens Smiatek, Dr. Frank Uhlig, and financial support from the German Funding Agency (Deutsche Forschungsgemeinschaft-DFG) under Germany's Excellence Strategy—EXC 2075—390740016. Dr. Rick Weber is thanked for providing the hafnia samples and Dr. Joerg Neuefeind is thanked for his help with the neutron experiments.

---

[*] Also at Data Science and Learning Division, Argonne National Laboratory, Lemont, Illinois 60439, USA.
[†] benmore@anl.gov

[1] R. C. Garvie, R. H. Hannink, and R. T. Pascoe, Nature (London) **258**, 703 (1975).
[2] X. Q. Cao, R. Vassen, and D. Stoever, J. Eur. Ceram. Soc. **24**, 1 (2004).
[3] J. Wang, H. P. Li, and R. Stevens, J. Mater. Sci. **27**, 5397 (1992).
[4] P. Piluso, M. Ferrier, C. Chaput, J. Claus, and J. P. Bonnet, J. Eur. Ceram. Soc. **29**, 961 (2009).
[5] S. V. Ushakov and A. Navrotsky, J. Am. Ceram. Soc. **95**, 1463 (2012).
[6] A. Pavlik, S. V. Ushakov, A. Navrotsky, C. J. Benmore, and R. J. K. Weber, J. Nucl. Mater. **495**, 385 (2017).
[7] J. Behler, J. Chem. Phys. **145**, 170901 (2016).
[8] V. L. Deringer, M. A. Caro, and G. Csányi, Adv. Mater. **31**, 1902765 (2019).
[9] A. P. Thompson, L. P. Swiler, C. R. Trott, S. M. Foiles, and G. J. Tucker, J. Comput. Phys. **285**, 316 (2015).
[10] A. P. Bartók, S. De, C. Poelking, N. Bernstein, J. R. Kermode, G. Csányi, and M. Ceriotti, Sci. Adv. **3**, e1701816 (2017).
[11] V. Botu, R. Batra, J. Chapman, and R. Ramprasad, J. Phys. Chem. C **121**, 511 (2017).
[12] S. Chmiela, H. E. Sauceda, K. R. Müller, and A. Tkatchenko, Nat. Commun. **9**, 3887 (2018).
[13] Y. Zuo, C. Chen, X. Li, Z. Deng, Y. Chen, J. Behler, G. Csányi, A. V. Shapeev, A. P. Thompson, M. A. Wood, and S. P. Ong, J. Phys. Chem. A **124**, 731 (2020).
[14] A. P. Bartók, M. C. Payne, R. Kondor, and G. Csányi, Phys. Rev. Lett. **104**, 136403 (2010).
[15] G. Sivaraman, A. N. Krishnamoorthy, M. Baur, C. Holm, M. Stan, G. Csányi, C. Benmore, and Á. Vázquez-Mayagoitia, npj Comput. Mater. **6**, 104 (2020).
[16] S. Tovey, A. N. Krishnamoorthy, J. Guo, C. Benmore, A. Heuer, and C. Holm, J. Phys. Chem. C **124**, 25760 (2020).
[17] A. P. Bartók, J. Kermode, N. Bernstein, and G. Csányi, Phys. Rev. X **8**, 041048 (2018).
[18] 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, J. Phys. Chem. Lett. **9**, 2879 (2018).
[19] R. Jinnouchi, J. Lahnsteiner, F. Karsai, G. Kresse, and M. Bokdam, Phys. Rev. Lett. **122**, 225701 (2019).
[20] D. J. Cole, L. Mones, and G. Csányi, Faraday Discuss. **224**, 247 (2020).
[21] B. Settles, Computer Sciences Department Active Learning Literature Survey (University of Wisconsin-Madison Department of Computer Sciences, 2009).
[22] J. S. Smith, B. Nebgen, N. Lubbers, O. Isayev, and A. E. Roitberg, J. Chem. Phys. **148**, 241733 (2018).
[23] E. V. Podryabinkin and A. V. Shapeev, Comput. Mater. Sci. **140**, 171 (2017).
[24] L. Zhang, D. Y. Lin, H. Wang, R. Car, and W. E., Phys. Rev. Mater. **3**, 023804 (2019).
[25] R. Jinnouchi, F. Karsai, and G. Kresse, Phys. Rev. B **100**, 014105 (2019).
[26] J. Vandermause, S. B. Torrisi, S. Batzner, Y. Xie, L. Sun, A. M. Kolpak, and B. Kozinsky, npj Comput. Mater. **6**, 20 (2020).
[27] S. Dasgupta and D. Hsu, in *Proceedings of the 25th International Conference on Machine learning* (Association for Computing Machinery (ACM), New York, New York, USA, 2008), pp. 208–215.
[28] J. F. Shackelford and R. H. Doremus, *Ceramic and Glass Materials: Structure, Properties and Processing* (Springer, New York, 2008).
[29] L. Gallington, Y. Ghadar, L. Skinner, J. Weber, S. Ushakov, A. Navrotsky, A. Vazquez-Mayagoitia, J. Neuefeind, M. Stan, J. Low, and C. Benmore, Materials (Basel) **10**, 1290 (2017).
[30] A. P. Hammersley, J. Appl. Crystallogr. **49**, 646 (2016).
[31] R. E. Hann, P. R. Suitch, and J. L. Pentecost, J. Am. Ceram. Soc. **68**, C-285 (1985).
[32] J. E. Jaffe, R. A. Bachorz, and M. Gutowski, Phys. Rev. B **72**, 144107 (2005).
[33] L. Passerini, Gazz. Chim. Ital. **60**, 762 (1930).
[34] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevLett.126.156002 for unit cell volume of pure phases.
[35] C. J. Benmore and J. K. R. Weber, Adv. Phys. X **2**, 717 (2017).
[36] J. Neuefeind, M. Feygenson, J. Carruth, R. Hoffmann, and K. K. Chipley, Nucl. Instrum. Methods Phys. Res., Sect. B **287**, 68 (2012).
[37] R. J. G. B. Campello, D. Moulavi, and J. Sander, in *Pacific-Asia Conference on Knowledge Discovery and Data Mining* (Springer, Berlin, Heidelberg, 2013), pp. 160–172.
[38] L. McInnes and J. Healy, in *2017 IEEE International Conference on Data Mining Workshops (ICDMW)* (IEEE Computer Society, New Orleans, 2017), pp. 33–42, http://dx.doi.org/10.1109/ICDMW.2017.12.
[39] J. Snoek, H. Larochelle, and R. P. Adams, Adv. Neural Inf. Process. Syst. (Curran Associates, Inc., 2012), https://proceedings.neurips.cc/paper/2012/file/05311655a15b75fab86956663e1819cd-Paper.pdf.
[40] A. P. Bartók, R. Kondor, and G. Csányi, Phys. Rev. B **87**, 184115 (2013).

[41] G. Kresse and J. Furthmüller, Comput. Mater. Sci. **6**, 15 (1996).
[42] G. Kresse and J. Furthmüller, Phys. Rev. B **54**, 11169 (1996).
[43] J. P. Perdew, K. Burke, and M. Ernzerhof, Phys. Rev. Lett. **77**, 3865 (1996).
[44] G. Kresse and D. Joubert, Phys. Rev. B **59**, 1758 (1999).
[45] S. Plimpton, J. Comput. Phys. **117**, 1 (1995).
[46] G. J. Martyna, D. J. Tobias, and M. L. Klein, J. Chem. Phys. **101**, 4177 (1994).
[47] M. Parrinello and A. Rahman, J. Appl. Phys. **52**, 7182 (1981).
[48] V. L. Deringer and G. Csányi, Phys. Rev. B **95**, 094203 (2017).
[49] S. P. Coleman, D. E. Spearot, and L. Capolungo, Model. Simul. Mater. Sci. Eng. **21**, 055020 (2013).
[50] D. M. Adams, S. Leonard, D. R. Russell, and R. J. Cernik, J. Phys. Chem. Solids **52**, 1181 (1991).
[51] T. R. Shan, B. D. Devine, T. W. Kemper, S. B. Sinnott, and S. R. Phillpot, Phys. Rev. B **81**, 125328 (2010).
[52] G. Broglia, G. Ori, L. Larcher, and M. Montorsi, Model. Simul. Mater. Sci. Eng. **22**, 065006 (2014).
[53] Q. J. Hong, S. V. Ushakov, D. Kapush, C. J. Benmore, R. J. K. Weber, A. van de Walle, and A. Navrotsky, Sci. Rep. **8**, 14962 (2018).