Exploring the robust extrapolation of high-dimensional machine learning potentials

Claudio Zeni [*]
Physics Area, International School for Advanced Studies, Trieste, Italy

Andrea Anelli
Roche Pharma Research and Early Development, Therapeutic Modalities, Roche Innovation Center Basel, F. Hoffmann-La Roche Ltd., Grenzacherstrasse 124, 4070 Basel, Switzerland

Aldo Glielmo
Physics Area, International School for Advanced Studies, Trieste, Italy and Banca d'Italia, Italy

Kevin Rossi [†]
Laboratory of Nanochemistry, Institute of Chemistry and Chemical Engineering, Ecole Polytechnique Fédérale de Lausanne, Lausanne, CH-1950 Sion, Switzerland

(Received 20 December 2021; revised 5 April 2022; accepted 8 April 2022; published 21 April 2022; corrected 9 May 2022)

We show that, contrary to popular assumptions, predictions from machine learning potentials built upon high-dimensional atom-density representations almost exclusively occur in regions of the representation space which lie outside the convex hull defined by the training set points. We then propose a perspective to rationalize the domain of robust extrapolation and accurate prediction of atomistic machine learning potentials in terms of the probability density induced by training points in the representation space.

DOI: 10.1103/PhysRevB.105.165141

I. INTRODUCTION

Machine learning (ML) potentials for atomistic systems infer the mapping between configurations and a target objective function, i.e., the total energy of the system and the forces acting on each atom. These potentials are trained on a database of configurations whose objective function has been calculated via a computationally expensive, yet accurate, reference, e.g., density functional theory methods. Following the training procedure, ML potentials offer predictions that are accurate with respect to, and much faster to compute than, the reference method [1–5]. A key aspect toward the making of accurate and efficient ML potentials lies in the choice of the representation function, which maps atomic coordinates to a set of numerical features. Among the most successful ones, we find expansions of local densities around atoms in the systems. In a nutshell, these representations are built upon the description of an atomic environment in terms of atom-centered distributions (encoding N-body correlations up to a desired order of N), which are approximated via a truncated expansion in radial and angular basis sets [6–11].

One notable characteristic of these representations is that they originate a high-dimensional feature space. The accuracy of the ML predictions is, moreover, generally observed to correlate with the representation's dimensionality, when the other free parameters in the expansion are fixed [1,2,12]. For this reason, there is an incentive to employ high-dimensional atomic-density representations when training ML potentials. Recent publications by Balestriero *et al.* [13] and Yousefzadeh [14], built upon the theoretical results presented by Bárány and Füredi [15], showcased that the predictions made by image recognition models whose inputs are high-dimensional, happen in an *extrapolation regime*, where interpolation and extrapolation regimes are formally defined according to a geometric criterion, in particular [13]: Interpolation occurs for a point $\mathbf{x}^*$ whenever the latter belongs to the convex hull (CH) of a set of training points $\mathbf{X} \triangleq \mathbf{x}_1, \mathbf{x}_2, \ldots \mathbf{x}_M$, if not, extrapolation occurs.

The above definition has also been employed in the community of scientists applying ML methods to atomistic systems, and a common assumption is that the accuracy of ML potentials is strongly dependent on the fact that their predictions take place in an interpolation regime. When a ML potential accurately predicts the objective function for a structure outside the training database, this result is often interpreted as a sign that the atomic environment representations in the out-of-sample structures are contained in, covered by, or interpolated between points in the training set [16–24].

Atom-density representations are built to naturally encode physical symmetries [6–11]. Further, protocols underlying the generation of atomistic configuration databases hinge on the sampling of trajectories and of physically relevant phases. These properties reduce the effective degrees of freedom possessed by such representations when compared, e.g., to images. It is therefore not trivial to predict *a priori* whether ML potentials exploiting high-dimensional atom-density

[*]claudio.zeni@sissa.it
[†]kevin.rossi@epfl.ch

representations also suffer from a curse of dimensionality and whether they carry on predictions in an *extrapolative regime*, defined according to the previously mentioned geometric criterion.

First, we show that, for the example case of data sets and benchmarks widely quoted in literature, ML potentials predictions generally occur at a point outside the training set CH. We then propose a protocol to measure the sampling density induced by the training data on points belonging to the test data set. This quantity is computed on a test point's features as the log probability density of training points estimated via an adaptive k-nearest-neighbors algorithm. We show that such measure strongly correlates with the prediction errors incurred by ridge-regression potentials on test sets, thus providing an effective tool to identify low-accuracy regions in the representation space and to rationalize the accuracy of ML potentials according to a rigorous geometric criterion. We thus clarify the difference between the CH-derived definition of interpolation and the alternative concept of a well-sampled region in the representation space. While the knowledge of whether a test point lies within the training set CH yields no information on the test accuracy of a high-dimensional ML potential, we show that one can establish a relationship between the test error incurred by ML potentials and the probability density function induced by the training set and computed on a test point.

II. METHOD

To draw general conclusions, we consider three data sets in our investigation. These comprise periodic and finite-size systems with different chemistries:

(1) The ice-water data set by Monserrat *et al.* [19], which was employed to test the transferability of a ML force field trained on water configurations to the case of ice crystals. It encompasses a training set of forces and energies in 1000 liquid water configurations—corresponding to 192 000 atomic environments—and a test set containing structures corresponding to 54 known ice phases—comprising 2847 atomic environments—which also includes all the experimentally verified ice structures.

(2) The Li, Mo, Ge, Si, Ni, Cu data set by Zuo *et al.* [12], which was used to benchmark cost and accuracy of several ML force-field flavors. It gathers energies and forces in systems of the six different elements for their ground-state crystalline bulk configuration, strained crystals, low Miller index surfaces, bulk structures sampled during *ab initio* molecular dynamics (MD) trajectories at different temperatures, and bulk structures with a vacancy, also sampled during *ab initio* MD trajectories. Configurations are then organized, according to a random 90:10 split, into a training and testing set.

(3) The Au_{13} database, which was custom built to probe the likelihood of extrapolation during a MD trajectory. It comprises five subsets of 1000 configurations of planar Au_{13} nanoclusters, with energies and forces labels, sampled every 3 ps during finite-temperature (50 K, 100 K, 200 K, 300 K, and 400 K) MD runs where no structural rearrangements were observed; for further details, we refer the reader to Sec. C of the Supplemental Material [25].

To associate features to atomic environments, we employ the atom-centered symmetry functions (ACSFs) [6,26], the smooth overlap of atomic positions (SOAP) [7,27], or the atomic cluster expansion (ACE) [11,28] representation (see Supplemental Material [25], Secs. A and B and Tables S1–S6 for further details), and adopt previously reported setups when these are available in the literature [12,19,29,30]. We then transform the high-dimensional representations in a linear fashion via principal components analysis (PCA), and construct sets of P-dimensional representations which employ the first P principal components, using the same procedure as reported in Zeni *et al.* [30]. We then systematically investigate whether test points are contained within the CH of the training set, also as a function of the number of employed PCA components. This approach parallels emerging protocols in the literature where low-dimensional embedding of atom-density representations are employed as a tool to probe the similarity among structures [31,32] and rationalize ML accuracy and transferability [1,12,19–23]. Rather than computing the CH—a time-consuming task in high dimensions [33]—we verify whether a test point $\mathbf{x}^*$ can be expressed as a linear combination of the points in the training set $\{\mathbf{x}_i\}_{i=1}^{M}$ constrained to non-negative coefficients λ_i summing up to one:

$$\mathbf{x}^* = \sum_{i=1}^{M} \lambda_i \mathbf{x}_i$$

$$\text{with} \quad \sum_{i=1}^{M} \lambda_i = 1 \wedge \lambda_i \geqslant 0 \, \forall \, i = 1, \ldots, M. \tag{1}$$

The test point $\mathbf{x}^*$ is in the CH of the training set if and only if the above can be satisfied; this can be verified efficiently via a linear programming approach [34,35].

III. RESULTS AND DISCUSSION

Figure 1 reports the number of test atomic environments which fall within the CH induced by the training set as a function of the dimensionality P of the PCA representations, for the three databases described. For reference, we show in Fig. S1 the cumulative variance explained by 2 to 25 principal components. The test points are completely contained in the CH determined by the training set, when considering a projection in the space of the first two PCA components. Increasing the dimensionality of the embedding, the number of test points enclosed within the CH diminishes rapidly. Regardless of the database design, chemical nature of the system, and choice of representation (see also Fig. S2 for additional benchmarks), embedding on a low yet sizable ($P \sim$ 10–20) number of PCA components results in an almost complete separation between each of the test points and the CH associated to the training points.

Following Fig. 1, we note that low-dimensional projections of atom-density features fail in faithfully preserving the information about whether an atomic environment is contained within the CH determined by a set of other ones. At a more fundamental level, we highlight that ML potentials based on high-dimensional representations are very likely to carry out predictions at data points not contained in the CH enclosing the training set. This is true not only when testing the

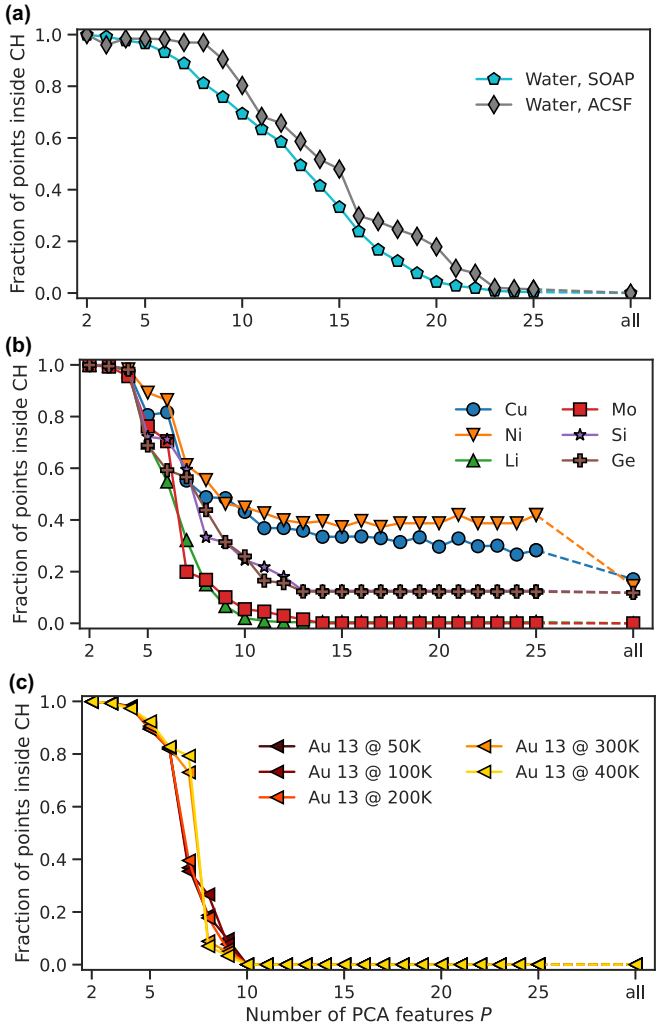

FIG. 1. Fraction of atomic environments in the test set which are contained in the CH enclosing the points in the training set, featurized according to their first P PCA components. (a) refers to results found for the water-ice database, using SOAP or ACSF representations. (b) corresponds to the case study of the six data sets described in Zuo *et al*. [12], using an ACE representation. (c) reports results found for structures extracted from an MD trajectory of an Au_{13} planar nanocluster at different temperatures, using an ACE representation, where the test and training set are constructed via a leave-one-out scheme.

transferability of the ML potential from one phase to another (i.e., the ice-water database) but also in apparently trivial MD trajectory where no structural rearrangements take place (i.e., the Zuo *et al*. [12] and the Au_{13} database). Notably, the non-negligible portion of atomic environments found within the training points CH in the Zuo *et al*. [12] data set [Cu, Ni, Ge and Si curves in Fig. 1(b)] actually consists of local atomic environments identical to those also present in the training set (see Supplemental Material [25], Sec. D).

After determining that predictions are likely to happen outside the training set CH when using a high-dimensional input representation, we test whether a ML potential's prediction effectively operates in such a high-dimensional space. To investigate if an accurate ML potential projects data into a low-dimensional space where test data are contained by the

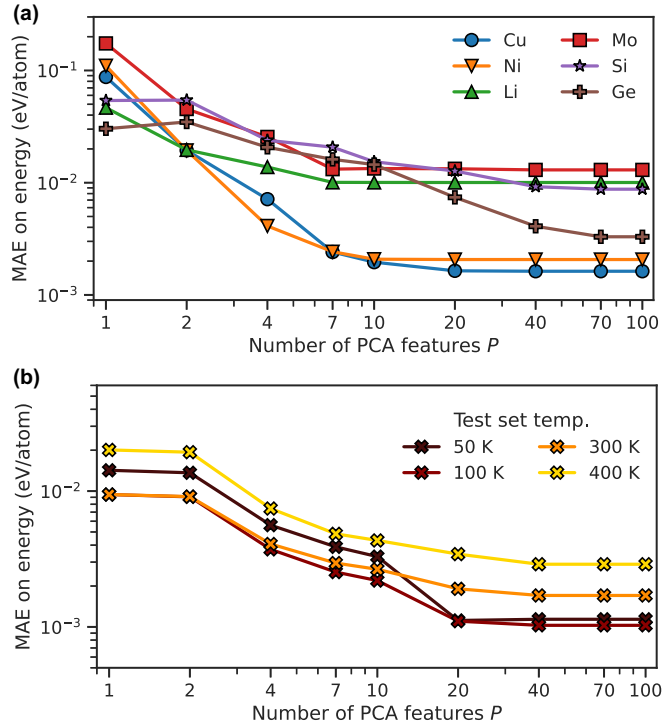

FIG. 2. Mean absolute error on energy prediction when accounting for the contribution of the first P PCA of the high-dimensional feature spaces deriving from an ACE representation for a ridge-regression potential trained on the Zuo *et al*. [12] data set (a) and on the Au_{13} database at 200 K (b).

ones in the training set, we train a regularized linear potential to predict the energy of structures for the example cases of the Zuo *et al*. [12] and Au_{13} data sets (see Supplemental Material [25], Secs. G and H for further info). We analyze the weights assigned to each feature by the regression algorithm following training, and map the relevance of each PCA feature toward diminishing the per-atom energy mean absolute error (MAE). We report the MAE of the potential as a function of the number of PCA components accounted for in the regression in Fig. 2. We refer the interested reader to Fig. S1 for a report on the amount of variance explained, and to Figs. 1(b) and 1(c) for the number of test points which are enclosed by the training points CH, as a function of the first P principal components. We confirm that ridge regression potentials do need to operate in a feature space where the majority of the test set atomic environments lie in a region of extrapolation to reach their best accuracy.

In light of the accuracy reported in the literature for predictions on train-set as well as test-set configurations [12,19,29,30] and our analysis, we demonstrate the need to revisit previous interpretations [1,12,19,21,22,24] relating ML potentials' accuracy and the geometrical relationship between test and training atomic environments, especially when these are carried out on low-dimensionality projections. Indeed, ML potentials exploiting a high-dimensional local density representation can generalize their predictions to atomic environments which lie in an extrapolation regime, as per defined according to a CH-based criterion. By the same token, and from the opposite perspective, we conclude that

accurate predictions do not imply that test points are contained within the high-dimensional CH enclosing the set of available training points. For the above reasons, a CH-based definition of interpolation and extrapolation is too weak and uninformative for the case of high-dimensional spaces. This finding further motivates the search for better suited definitions of interpolation and extrapolation, which align with generalization performances.

To rationalize the effectiveness of ML regressors in high dimensions, we hypothesize that the test points lie in regions of the representation space that are sufficiently sampled by the training set distribution. To test this hypothesis, we estimate the probability density generated by the training set points, which we call *sampling density*, on the locations of representation space where test points lie. We do so by using an adaptive k-nearest-neighbor density estimation, which works as follows. The test point $\mathbf{x}^*$ is virtually added to the training set and its k^* nearest training points are found. The density is then computed as

$$\rho(\mathbf{x}^*) = \frac{k^* - 1}{M V^*}, \tag{2}$$

where M is the size of the training set and V^* is the volume occupied by the first k^* training point neighbors [36].

The number k^* of training neighbors considered is chosen adaptively for each test point as to maximize the accuracy of the estimate [37]. The volume V^* is computed as the volume of a hypersphere in dimension d, where d is the intrinsic dimension of the training data manifold computed via the TwoNN estimator [38]. This way, we obtain a measure that indicates how much test set points are well-sampled by the training set density. We refer the interested reader to Sec. I of the Supplemental Material [25] for further details on the density estimation algorithm.

We assess the relationship existing between such measure and the error incurred by a ridge-regression potential based on ACE representations trained on the data that generates the density distribution when tested on points in the test set. We do so for the Zuo *et al.* [12] and for the Au_{13} database. In the latter case, each potential is trained on data from a MD trajectory at a single temperature, and tested on data coming from the other four MD trajectories. In Fig. 3, we report the test MAE on forces, averaged and binned over the negative log sampling density of local atomic environment representations in the test sets. We use 20 averaging windows equispaced between the lowest 5% and the highest 5% log density encountered across all test sets, and display only bins containing at least 1% of the data.

Figure 3 highlights how the log sampling density correlates with the binned test MAE on forces for points outside the training set, for all databases considered (see also Figs. S7–S9). We find that the metric we introduce offers an estimate of the degree by which an out-of-sample atomic environment lies within a well-sampled region of the representation space and, more importantly, correlates with the error incurred by trained regression potentials. We nevertheless find that the proposed metric is dependent on the choice of the representation (Figs. S8 amd S9) and that the precise relationship with the MAE is system- and model-dependent. Moreover, we observe a good correlation between

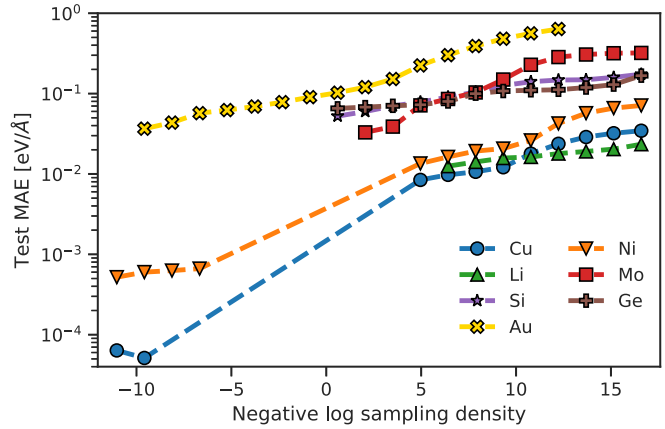

FIG. 3. Correlation between the test MAE on forces incurred by ridge regression potentials and the negative log probability density estimate for the training points in representation space, computed on test points (i.e., sampling density).

the proposed metric and a model-dependent error estimator, i.e., the prediction uncertainty drawn from a committee of models trained by sub-sampling a larger training set [39] (see Sec. K, Fig. S11 for further details). These results confirm the hypothesis that the training set sampling density provides a statistical/geometrical criterion to chart extrapolation robustness of high-dimensional ML models across the representation space.

IV. CONCLUSIONS

To conclude, we follow the definition spelled out by Balestriero *et al.* [13], where extrapolation occurs if the test structure lies outside the high-dimensional CH which encloses the set of training structure, and find that the large majority of the test predictions take place in an extrapolation regime when employing high-dimensional local atomic density representations (ACSF, SOAP, ACE). We thus demonstrate the need to revisit previous interpretations relating ML potentials' accuracies and the geometrical relationships between test and training atomic environments.

In a second instance, to understand why linear ML models exploiting atom-density representations are predictive for points outside their CH, we relate their accuracy to the probability density induced by training points in the representation space. This geometric measure of well-sampledness in the representation space is found to correlate with the error incurred by the ML model, and overcomes the limitations in the use of a CH construction to geometrically define interpolation and extrapolation regimes in high-dimensional spaces. The criterion suggested, in turn, enables us to verify whether a training set is well-suited to enable accurate predictions at a target point in representation space.

We envision that the density-sampling analysis will promote the rational development of novel database generation routines, adaptive sampling protocols, and data point selection algorithms. This area of research [40–43] is indeed critical toward a data-efficient route in the automatic construction of ML potential potentials. Future endeavors will be also directed to probe the mechanisms that rule the interpolation/extrapolation and the induced training densities inherent to

ML potentials exploiting learnable representations [44–49]. Finally, we highlight the generality of the training data-set sampling density analysis, which could be applied in other domains where high-dimensional ML models are widespread and successful, e.g., image recognition, diagnostics, therapeutic development and health-care delivery.

Scripts and inputs required to generate all the plots and results discussed in this paper are available on the Materials Cloud [50].

ACKNOWLEDGMENTS

C.Z. gratefully acknowledges support by the European Union's Horizon 2020 research and innovation program (Grant No. 824143, MaX, MAterials design at the eXascale Centre of Excellence). K.R. has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (Marie Curie Individual Fellowship Grant Agreement No. 890414). The authors thank Matteo Carli, Stefano De Gironcoli, Piero Gasparotto, Federico Grasselli, Andrea Grisafi, Giulio Imbalzano, and Nataliya Lopanitsyna for useful discussions.

The views and opinions expressed in this paper are those of the authors and do not necessarily reflect the official policy or position of Banca d'Italia.

A.A. is currently employed at Hoffman La Roche and has contributed to this paper outside of his Roche Post Doctoral Fellowship project.

[1] F. Musil, A. Grisafi, A. P. Bartók, C. Ortner, G. Csányi, and M. Ceriotti, Chem. Rev. **121**, 9759 (2021).
[2] V. L. Deringer, A. P. Bartók, N. Bernstein, D. M. Wilkins, M. Ceriotti, and G. Csányi, Chem. Rev. **121**, 10073 (2021).
[3] O. T. Unke, S. Chmiela, H. E. Sauceda, M. Gastegger, I. Poltavsky, K. T. Schütt, A. Tkatchenko, and K. R. Müller, Chem. Rev. **121**, 10142 (2021).
[4] J. A. Keith, V. Vassilev-Galindo, B. Cheng, S. Chmiela, M. Gastegger, K. R. Müller, and A. Tkatchenko, Chem. Rev. **121**, 9816 (2021).
[5] C. Zeni, K. Rossi, A. Glielmo, and F. Baletto, Adv. Phys.: X **4**, 1654919 (2019).
[6] J. Behler and M. Parrinello, Phys. Rev. Lett. **98**, 146401 (2007).
[7] A. P. Bartók, M. C. Payne, R. Kondor, and G. Csányi, Phys. Rev. Lett. **104**, 136403 (2010).
[8] A. P. Thompson, L. P. Swiler, C. R. Trott, S. M. Foiles, and G. J. Tucker, J. Comput. Phys. **285**, 316 (2015).
[9] A. V. Shapeev, Multiscale Model. Simul. **14**, 1153 (2016).
[10] M. J. Willatt, F. Musil, and M. Ceriotti, J. Chem. Phys. **150**, 154110 (2019).
[11] R. Drautz, Phys. Rev. B **99**, 014104 (2019).
[12] Y. Zuo, C. Chen, X. Li, Z. Deng, Y. Chen, J. Behler, G. Csányi, A. V. Shapeev, A. P. Thompson, M. A. Wood *et al.*, J. Phys. Chem. A **124**, 731 (2020).
[13] R. Balestriero, J. Pesenti, and Y. LeCun, arXiv:2110.09485.
[14] R. Yousefzadeh, arXiv:2101.09849.
[15] I. Bárány and Z. Füredi, Probab. Theory Relat. Fields **77**, 231 (1988).
[16] A. P. Bartók, J. Kermode, N. Bernstein, and G. Csányi, Phys. Rev. X **8**, 041048 (2018).
[17] K. Rossi, V. Jurásková, R. Wischert, L. Garel, C. Corminbœuf, and M. Ceriotti, J. Chem. Theory Comput. **16**, 5139 (2020).
[18] A. Tkatchenko, Nat. Commun. **11**, 4125 (2020).
[19] B. Monserrat, J. G. Brandenburg, E. A. Engel, and B. Cheng, Nat. Commun. **11**, 5757 (2020).
[20] T. T. Nguyen, E. Székely, G. Imbalzano, J. Behler, G. Csányi, M. Ceriotti, A. W. Götz, and F. Paesani, J. Chem. Phys. **148**, 241725 (2018).
[21] A. P. Bartók, S. De, C. Poelking, N. Bernstein, J. R. Kermode, G. Csányi, and M. Ceriotti, Sci. Adv. **3**, e1701816 (2017).
[22] C. Zeni, K. Rossi, T. Pavloudis, J. Kioseloglu, S. DeGironcoli, R. E. Palmer, and F. Baletto, Nat. Commun. **12**, 6056 (2021).
[23] E. Engel, A. Anelli, A. Hofstetter, F. Paruzzo, L. Emsley, and M. Ceriotti, Phys. Chem. Chem. Phys. **21**, 23385 (2019).
[24] Y. Shao, F. M. Dietrich, C. Nettelblad, and C. Zhang, J. Chem. Phys. **155**, 204108 (2021).
[25] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevB.105.165141 for additional details regarding the local atomic environment representations employed, the generation of the Au_{13} data set, additional CH benchmarks, the effect of identical local atomic environments on CH computation, ridge regression for energies and forces, errors as a function of PCA components in ridge regression for the Au_{13} database, the adaptive k-NN estimation algorithm, the correlation between the negative log density and the MAE incurred by ridge regression force fields, and the comparison of adaptive k-nn with other error estimation methods.
[26] A. Singraber, J. Behler, and C. Dellago, J. Chem. Theory Comput. **15**, 1827 (2019).
[27] L. Himanen, M. O. J. Jäger, E. V. Morooka, F. Federici Canova, Y. S. Ranawat, D. Z. Gao, P. Rinke, and A. S. Foster, Comput. Phys. Commun. **247**, 106949 (2020).
[28] G. Dusson, M. Bachmayr, G. Csanyi, R. Drautz, S. Etter, C. van der Oord, and C. Ortner, J. Comput. Phys. **454**, 110946 (2022).
[29] B. Cheng, E. Engel, A., J. Behler, C. Dellago, and M. Ceriotti, Proc. Natl Acad. Sci. USA **116**, 1110 (2019).
[30] C. Zeni, K. Rossi, A. Glielmo, and S. D. Gironcoli, J. Chem. Phys. **154**, 224112 (2021).
[31] E. A. Engel, A. Anelli, M. Ceriotti, C. J. Pickard, and R. J. Needs, Nat. Commun. **9**, 2173 (2018).
[32] A. Anelli, E. A. Engel, C. J. Pickard, and M. Ceriotti, Phys. Rev. Materials **2**, 103804 (2018).
[33] F. Nielsen, Inform. Proc. Lett. **59**, 255 (1996).
[34] P. Pardalos, Y. Li, and W. Hager, Comput. Math. Appl. **29**, 23 (1995).
[35] G. Sierksma and Y. Zwols, *Linear and Integer Optimization: Theory and Practice*, 3rd ed. (CRC Press, Boca Raton, Florida, 2015), p. 1.
[36] M. Carli and A. Laio, Mol. Phys. **119**, e1899323 (2021).

[37] A. Rodriguez, M. d'Errico, E. Facco, and A. Laio, J. Chem. Theory Comput. **14**, 1206 (2018).
[38] E. Facco, M. D'Errico, A. Rodriguez, and A. Laio, Sci. Rep. **7**, 12140 (2017).
[39] G. Imbalzano, Y. Zhuang, V. Kapil, K. Rossi, E. A. Engel, F. Grasselli, and M. Ceriotti, J. Chem. Phys. **154**, 074102 (2021).
[40] J. Vandermause, S. B. Torrisi, S. Batzner, Y. Xie, L. Sun, A. M. Kolpak, and B. Kozinsky, npj Comput. Mater. **6**, 20 (2020).
[41] M. Karabin and D. Perez, J. Chem. Phys. **153**, 094110 (2020).
[42] N. Bernstein, G. Csányi, and V. L. Deringer, npj Comput. Mater. **5**, 99 (2019).
[43] E. V. Podryabinkin and A. V. Shapeev, Comput. Mater. Sci. **140**, 171 (2017).
[44] M. Rupp, A. Tkatchenko, K.-R. Müller, and O. A. von Lilienfeld, Phys. Rev. Lett. **108**, 058301 (2012).
[45] K. T. Schütt, H. E. Sauceda, P. Kindermans, A. Tkatchenko, and K.-R. Müller, J. Chem. Phys **148**, 241722 (2018).
[46] S. Chmiela, H. E. Sauceda, I. Poltavksy, K.-R. Müller, and A. Tkatchenko, Comput. Phys. Commun. **240**, 38 (2019).
[47] T. Xie and J. C. Grossman, Phys. Rev. Lett. **120**, 145301 (2018).
[48] J. Gilmer, S. S. Schoenholz, P. F. Riley, O. Vinyals, and G. E. Dahl, in *Proceedings of the 34th International Conference on Machine Learning*, ICML'17 (PMLR, 2017), Vol. 70, pp. 1263–1272.
[49] J. P. Mailoa, M. Kornbluth, S. Batzner, G. Samsonidze, S. T. Lam, J. Vandermause, C. Ablitt, N. Molinari, and B. Kozinsky, Nat. Mach. Intell. **1**, 471 (2019).
[50] DOI:10.24435/materialscloud:8w-a7.

Correction: The second affiliation for the third author was corrected for consistency.