Machine learning of accurate energy-conserving molecular force fields

Stefan Chmiela,[1] Alexandre Tkatchenko,[2,3]* Huziel E. Sauceda,[3] Igor Poltavsky,[2] Kristof T. Schütt,[1] Klaus-Robert Müller[1,4,5]*

Using conservation of energy—a fundamental property of closed classical and quantum mechanical systems—we develop an efficient gradient-domain machine learning (GDML) approach to construct accurate molecular force fields using a restricted number of samples from ab initio molecular dynamics (AIMD) trajectories. The GDML implementation is able to reproduce global potential energy surfaces of intermediate-sized molecules with an accuracy of 0.3 kcal mol^{-1} for energies and 1 kcal mol^{-1} Å^{-1} for atomic forces using only 1000 conformational geometries for training. We demonstrate this accuracy for AIMD trajectories of molecules, including benzene, toluene, naphthalene, ethanol, uracil, and aspirin. The challenge of constructing conservative force fields is accomplished in our work by learning in a Hilbert space of vector-valued functions that obey the law of energy conservation. The GDML approach enables quantitative molecular dynamics simulations for molecules at a fraction of cost of explicit AIMD calculations, thereby allowing the construction of efficient force fields with the accuracy and transferability of high-level ab initio methods.

INTRODUCTION

Within the Born-Oppenheimer (BO) approximation, predictive simulations of properties and functions of molecular systems require an accurate description of the global potential energy hypersurface $V_{BO}(\vec{r_1}, \vec{r_2}, \ldots, \vec{r_N})$, where $\vec{r_i}$ indicates the nuclear Cartesian coordinates. Although V_{BO} could, in principle, be obtained on the fly using explicit ab initio calculations, more efficient approaches that can access the long time scales are required to understand relevant phenomena in large molecular systems. A plethora of classical mechanistic approximations to V_{BO} have been constructed, in which the parameters are typically fitted to a small set of ab initio calculations or experimental data. Unfortunately, these classical approximations may suffer from the lack of transferability and can yield accurate results only close to the conditions (geometries) they have been fitted to. Alternatively, sophisticated machine learning (ML) approaches that can accurately reproduce the global potential energy surface (PES) for elemental materials (1–9) and small molecules (10–16) have been recently developed (see Fig. 1, A and B) (17). Although potentially very promising, one particular challenge for direct ML fitting of molecular PES is the large amount of data necessary to obtain an accurate model. Often, many thousands or even millions of atomic configurations are used as training data for ML models. This results in nontransparent models, which are difficult to analyze and may break consistency (18) between energies and forces.

A fundamental property that any force field $\mathbf{F}_i(\vec{r_1}, \vec{r_2}, \ldots, \vec{r_N})$ must satisfy is the conservation of total energy, which implies that $\mathbf{F}_i(\vec{r_1}, \vec{r_2}, \ldots, \vec{r_N}) = -\nabla_{\vec{r_i}} V(\vec{r_1}, \vec{r_2}, \ldots, \vec{r_N})$. Any classical mechanistic expressions for the potential energy (also denoted as classical force field) or analytically derivable ML approaches trained on energies satisfy energy conservation by construction. However, even if conserva-

tion of energy is satisfied implicitly within an approximation, this does not imply that the model will be able to accurately follow the trajectory of the true ab initio potential, which was used to fit the force field. In particular, small energy/force inconsistencies between the force field model and ab initio calculations can lead to unforeseen artifacts in the PES topology, such as spurious critical points that can give rise to incorrect molecular dynamics (MD) trajectories. Another fundamental problem is that classical and ML force fields focusing on energy as the main observable have to assume atomic energy additivity—an approximation that is hard to justify from quantum mechanics.

Here, we present a robust solution to these challenges by constructing an explicitly conservative ML force field, which uses exclusively atomic gradient information in lieu of atomic (or total) energies. In this manner, with any number of data samples, the proposed model fulfills energy conservation by construction. Obviously, the developed ML force field can be coupled to a heat bath, making the full system (molecule and bath) non–energy-conserving.

We remark that atomic forces are true quantum-mechanical observables within the BO approximation by virtue of the Hellmann-Feynman theorem. The energy of a molecular system is recovered by analytic integration of the force-field kernel (see Fig. 1C). We demonstrate that our gradient-domain machine learning (GDML) approach is able to accurately reproduce global PESs of intermediate-sized molecules within 0.3 kcal mol^{-1} for energies and 1 kcal mol^{-1} Å^{-1} for atomic forces relative to the reference data. This accuracy is achieved when using less than 1000 training geometries to construct the GDML model and using energy conservation to avoid overfitting and artifacts. Hence, the GDML approach paves the way for efficient and precise MD simulations with PESs that are obtained with arbitrary high-level quantum-chemical approaches. We demonstrate the accuracy of GDML by computing AIMD-quality thermodynamic observables using path-integral MD (PIMD) for eight organic molecules with up to 21 atoms and four chemical elements. Although we use density functional theory (DFT) calculations as reference in this development work, it is possible to use any higher-level quantum-chemical reference data. With state-of-the-art quantum chemistry codes running on current high-performance computers, it is possible to generate accurate reference data for molecules with a few dozen atoms. Here, we focus on intramolecular

[1]Machine Learning Group, Technische Universität Berlin, 10587 Berlin, Germany. [2]Physics and Materials Science Research Unit, University of Luxembourg, L-1511 Luxembourg, Luxembourg. [3]Fritz-Haber-Institut der Max-Planck-Gesellschaft, 14195 Berlin, Germany. [4]Department of Brain and Cognitive Engineering, Korea University, Anam-dong, Seongbuk-gu, Seoul 136-713, Korea. [5]Max Planck Institute for Informatics, Stuhlsatzenhausweg, 66123 Saarbrücken, Germany.
*Corresponding author. Email: alexandre.tkatchenko@uni.lu (A.T.); klaus-robert.mueller@tu-berlin.de (K.-R.M.)
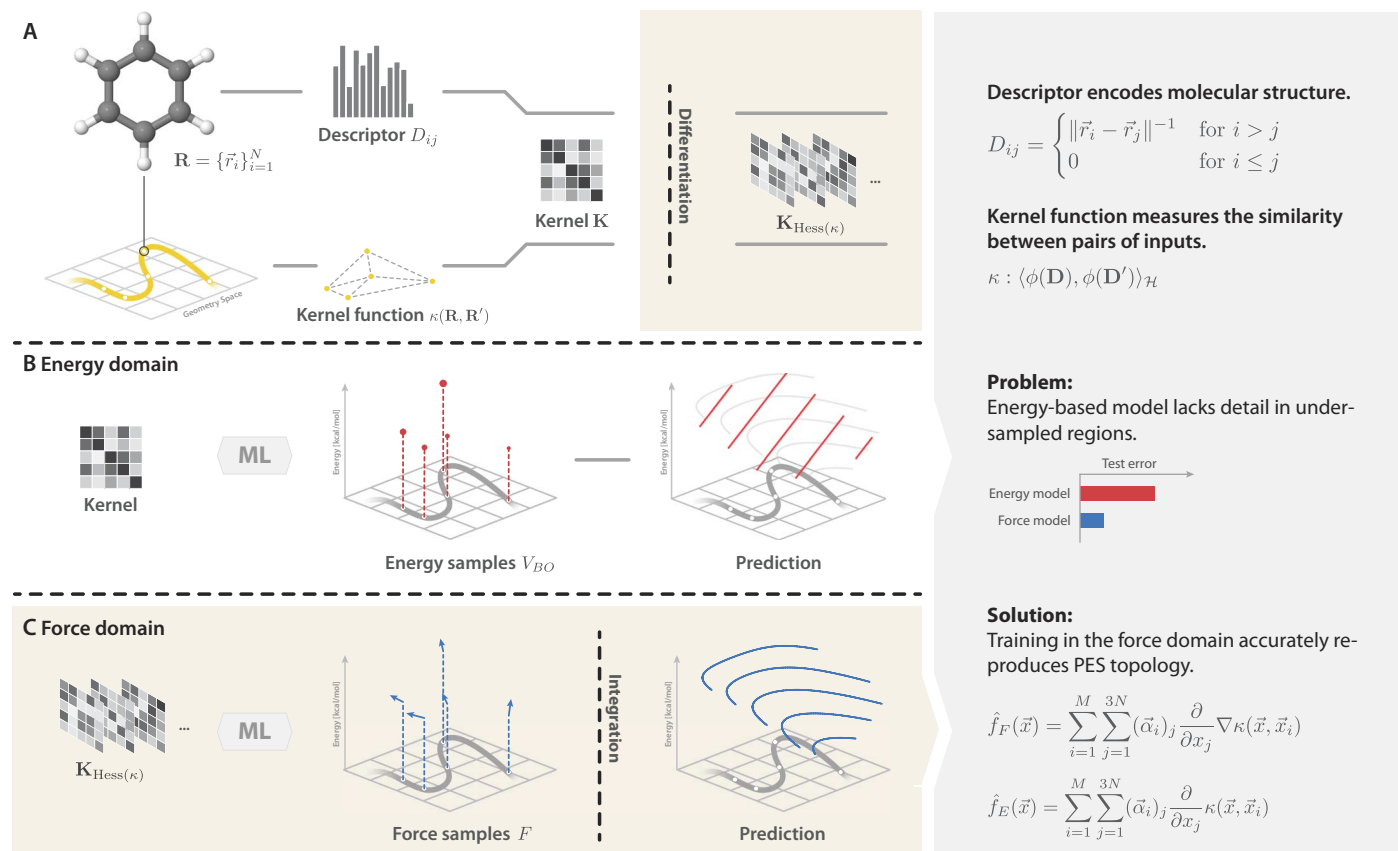
Fig. 1. The construction of ML models: First, reference data from an MD trajectory are sampled. (**A**) The geometry of each molecule is encoded in a descriptor. This representation introduces elementary transformational invariances of energy and constitutes the first part of the prior. A kernel function then relates all descriptors to form the kernel matrix—the second part of the prior. The kernel function encodes similarity between data points. Our particular choice makes only weak assumptions: It limits the frequency spectrum of the resulting model and adds the energy conservation constraint. Hess, Hessian. (**C**) These general priors are sufficient to reproduce good estimates from a restricted number of force samples. (**B**) A comparable energy model is not able to reproduce the PES to the same level of detail.
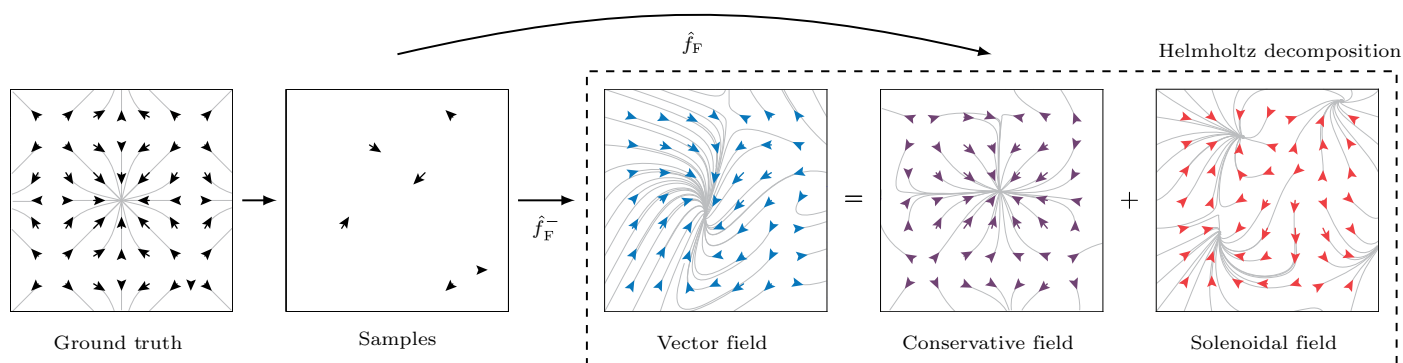
Fig. 2. Modeling the true vector field (leftmost subfigure) based on a small number of vector samples With GDML, a conservative vector field estimate $\hat{f}_F$ is obtained directly. A naïve estimator $\hat{f}_F^-$ with independent predictions for each element of the output vector is not capable of imposing energy conservation constraints. We perform a Helmholtz decomposition of this nonconservative vector field to show the error component that violates the law of energy conservation. This is the portion of the overall prediction error that was avoided with GDML because of the addition of the energy conservation constraint.

forces in small- and medium-sized molecules. However, in the future, the GDML model should be combined with an accurate model for intermolecular forces to enable predictive simulations of condensed molecular systems. Widely used classical mechanistic force fields are based on simple harmonic terms for intramolecular degrees of freedom. Our GDML model correctly treats anharmonicities by using no assumptions whatsoever on the analytic form on the interatomic potential energy functions within molecules.
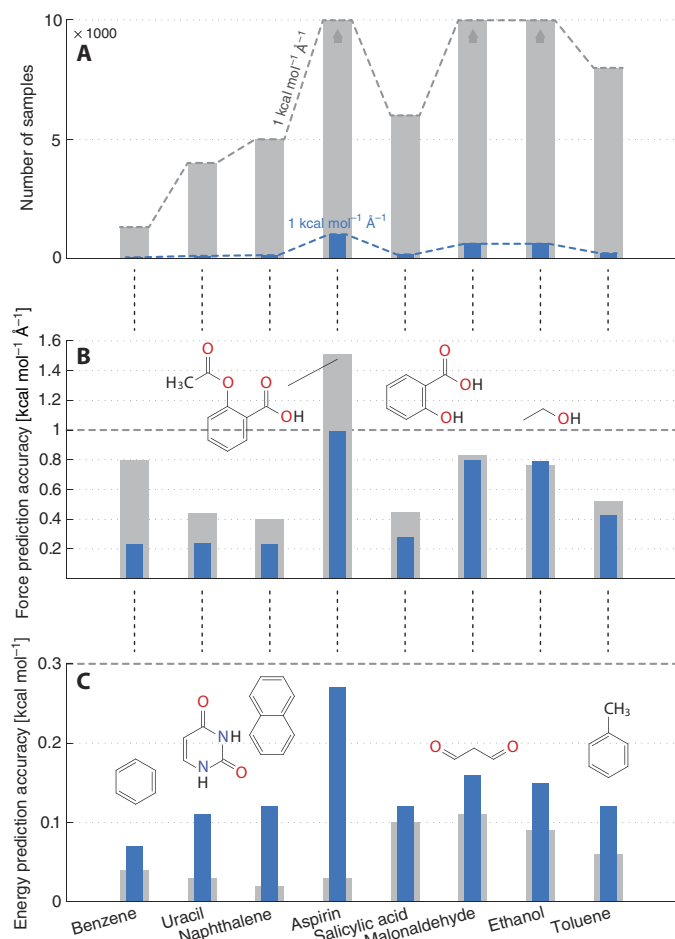
Fig. 3. Efficiency of GDML predictor versus a model that has been trained on energies. (**A**) Required number of samples for a force prediction performance of MAE (1 kcal mol^{-1} Å^{-1}) with the energy-based model (gray) and GDML (blue). The energy-based model was not able to achieve the targeted performance with the maximum number of 63,000 samples for aspirin. (**B**) Force prediction errors for the converged models (same number of partial derivative samples and energy samples). (**C**) Energy prediction errors for the converged models. All reported prediction errors have been estimated via cross-validation.

METHODS

The GDML approach explicitly constructs an energy-conserving force field, avoiding the application of the noise-amplifying derivative operator to a parameterized potential energy model (see the Supplementary Materials for details). This can be achieved by directly learning the functional relationship

$$\hat{f}_F : (\vec{r_1}, \vec{r_2}, \ldots, \vec{r_N})_i \overset{\text{ML}}{\longrightarrow} \mathbf{F}_i \qquad (1)$$

between atomic coordinates and interatomic forces, instead of computing the gradient of the PES (see Fig. 1, C and B). This requires constraining the solution space of all arbitrary vector fields to the subset of energy-conserving gradient fields. The PES can be obtained through direct integration of $\hat{f}_F$ up to an additive constant.

To construct $\hat{f}_F$, we used a generalization of the commonly used kernel ridge regression technique for structured vector fields (see

the Supplementary Materials for details) (19–21). GDML solves the normal equation of the ridge estimator in the gradient domain using the Hessian matrix of a kernel as the covariance structure. It maps to all partial forces of a molecule simultaneously (see Fig. 1A)

$$\left(\mathbf{K}_{\text{Hess}(\kappa)} + \lambda \mathbb{I}\right)\vec{\alpha} = \nabla V_{BO} = -\mathbf{F} \qquad (2)$$

We resorted to the extensive body of research on suitable kernels and descriptors for the energy prediction task (10, 13, 17).

For our application, we considered a subclass from the parametric Matérn family (22–24) of (isotropic) kernel functions

$$\kappa : C_{v=n+\frac{1}{2}}(d) = \exp\left(-\frac{\sqrt{2v}d}{\sigma}\right)P_n(d),$$

$$P_n(d) = \sum_{k=0}^{n} \frac{(n+k)!}{(2n)!}\binom{n}{k}\left(\frac{2\sqrt{2v}d}{\sigma}\right)^{n-k} \qquad (3)$$

where $d = \|\vec{x} - \vec{x}'\|$ is the Euclidean distance between two molecule descriptors. It can be regarded as a generalization of the universal Gaussian kernel with an additional smoothness parameter n. Our parameterization $n = 2$ resembles the Laplacian kernel, as suggested by Hansen et al. (13), while being sufficiently differentiable.

To disambiguate Cartesian geometries that are physically equivalent, we use an input descriptor derived from the Coulomb matrix (see the Supplementary Materials for details) (10).

The trained force field estimator collects the contributions of the partial derivatives $3N$ of all training points M to compile the prediction. It takes the form

$$\hat{f}_F(\vec{x}) = \sum_{i=1}^{M}\sum_{j=1}^{3N}(\vec{\alpha_i})_j\frac{\partial}{\partial x_j}\nabla\kappa(\vec{x}, \vec{x_i}) \qquad (4)$$

and a corresponding energy predictor is obtained by integrating $\hat{f}_F(\vec{x})$ with respect to the Cartesian geometry. Because the trained model is a (fixed) linear combination of kernel functions, integration only affects the kernel function itself. The expression

$$\hat{f}_E(\vec{x}) = \sum_{i=1}^{M}\sum_{j=1}^{3N}(\vec{\alpha_i})_j\frac{\partial}{\partial x_j}\kappa(\vec{x}, \vec{x_i}) \qquad (5)$$

for the energy predictor is therefore neither problem-specific nor does it require retraining.

We remark that our PES model is global in the sense that each molecular descriptor is considered as a whole entity, bypassing the need for arbitrary partitioning of energy into atomic contributions. This allows the GDML framework to capture chemical and long-range interactions. Obviously, long-range electrostatic and van der Waals interactions that fall within the error of the GDML model will have to be incorporated with explicit physical models. Other approaches that use ML to fit PESs such as Gaussian approximation potentials (3, 8) have been proposed. However, these approaches consider an explicit localization of the contribution of individual atoms to the total energy. The total energy is expressed as a linear combination
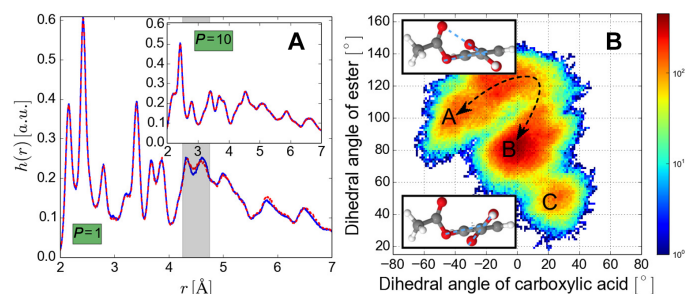
Fig. 4. Results of classical and PIMD simulations. The recently developed estimators based on perturbation theory were used to evaluate structural and electronic observables (30). (**A**) Comparison of the interatomic distance distributions, $h(r) = \langle \frac{2}{N(N-1)} \sum_{i<j}^{N} \delta(r - ||\vec{r_i} - \vec{r_j}||) \rangle_{P,t}$, obtained from GDML (blue line) and DFT (dashed red line) with classical MD (main frame), and PIMD (inset). a.u., arbitrary units. (**B**) Probability distribution of the dihedral angles (corresponding to carboxylic acid and ester functional groups) using a 20 ps time interval from a total PIMD trajectory of 200 ps.

of local environments characterized by a descriptor that acts as a nonunique partitioning function to the total energy. Training on force samples similarly requires the evaluation of kernel derivatives, but w.r.t. those local environments. Although any partitioning of the total energy is arbitrary, our molecular total energy is physically meaningful in that it is related to the atomic force, thus being a measure for the deflection of every atom from its ground state.

We first demonstrate the impact of the energy conservation constraint on a toy model that can be easily visualized. A nonconservative force model $\hat{f}_F^-$ was trained alongside our GDML model $\hat{f}_F$ on a synthetic potential defined by a two-dimensional harmonic oscillator using the same samples, descriptor, and kernel.

We were interested in a qualitative assessment of the prediction error that is introduced as a direct result of violating the law of energy conservation.

For this, we uniquely decomposed our naïve estimate

$$\hat{f}_F^- = -\nabla E + \nabla \times A \tag{6}$$

into a sum of a curl-free (conservative) and a divergence-free (solenoidal) vector field, according to the Helmholtz theorem (see Fig. 2) (25). This was achieved by subsampling $\hat{f}_F^-$ on a regular grid and numerically projecting it onto the closest conservative vector field by solving Poisson's equation (26)

$$-\nabla^2 E \overset{!}{=} \nabla \hat{f}_F^- \tag{7}$$

with Neumann boundary conditions. The remaining solenoidal field represents the systematic error made by the naïve estimator. Other than in this example, our GDML approach directly estimates the conservative vector field and does not require a costly numerical projection on a dense grid of regularly spaced samples.

RESULTS

We now proceed to evaluate the performance of the GDML approach by learning and then predicting AIMD trajectories for molecules, including benzene, uracil, naphthalene, aspirin, salicylic acid, malonaldehyde, ethanol, and toluene (see table S1 for details

of these molecular data sets). These data sets range in size from 150 k to nearly 1 M conformational geometries with a resolution of 0.5 fs, although only a drastically reduced subset is necessary to train our energy and GDML predictors. The molecules have different sizes, and the molecular PESs exhibit different levels of complexity. The energy range across all data points within a set spans from 20 to 48 kcal mol^{-1}. Force components range from 266 to 570 kcal mol^{-1} Å^{-1}. The total energy and force labels for each data set were computed using the PBE + vdW-TS electronic structure method (27, 28).

The GDML prediction results are contrasted with the output of a model that has been trained on energies. Both models use the same kernel and descriptor, but the hyperparameter search was performed individually to ensure optimal model selection. The GDML model for each data set was trained on ~1000 geometries, sampled uniformly according to the MD@DFT trajectory energy distribution. For the energy model, we multiplied this amount by the number of atoms in one molecule times its three spatial degrees of freedom. This configuration yields equal kernel sizes for both models and therefore equal levels of complexity in terms of the optimization problem. We compare the models on the basis of the required number of samples (Fig. 3A) to achieve a force prediction accuracy of 1 kcal mol^{-1} Å^{-1}. Furthermore, the prediction accuracy of the force and energy estimates for fully converged models (w.r.t. number of samples) (Fig. 3, B and C) are judged on the basis of the mean absolute error (MAE) and root mean square error performance measures.

It can be seen in Fig. 3A that the GDML model achieves a force accuracy of 1 kcal mol^{-1} Å^{-1} using only ~1000 samples from different data sets. Conversely, a pure energy-based model would require up to two orders of magnitude more samples to achieve a similar accuracy. The superior performance of the GDML model cannot be simply attributed to the greater information content of force samples. We compare our results to those of a naïve force model along the lines of the toy example shown in Fig. 2 (see tables S1 and S3 for details on the prediction accuracy of both models). The naïve force model is nonconservative but identical to the GDML model in all other aspects. Note that its performance deteriorates significantly on all data sets compared to the full GDML model (see the Supplementary Materials for details). We note here that we used DFT calculations, but any other high-level quantum chemistry approach could have been used to calculate forces for 1000 conformational geometries. This allows AIMD simulations to be carried out at the speed of ML models with the accuracy of correlated quantum chemistry calculations.

It is noticeable that the GDML model at convergence (w.r.t. number of samples) yields higher accuracy for forces than an equivalent energy-based model (see Fig. 3B). Here, we should remark that the energy-based model trained on a very large data set can reduce the energy error to below 0.1 kcal mol^{-1}, whereas the GDML energy error remains at 0.2 kcal mol^{-1} for ~1000 training samples (see Fig. 3C). However, these errors are already significantly below thermal fluctuations (k_BT) at room temperature (~0.6 kcal mol^{-1}), indicating that the GDML model provides an excellent description of both energies and forces, fully preserves their consistency, and reduces the complexity of the ML model. These are all desirable features of models that combine rigorous physical laws with the power of data-driven machines.

The ultimate test of any force field model is to establish its aptitude to predict statistical averages and fluctuations using MD simulations. The quantitative performance of the GDML model is demonstrated in

Fig. 4 for classical and quantum MD simulations of aspirin at $T = 300$ K. Figure 4A shows a comparison of interatomic distance distributions, $h(r)$, from MD@DFT and MD@GDML. Overall, we observe a quantitative agreement in $h(r)$ between DFT and GDML simulations. The small differences in the distance range between 4.3 and 4.7 Å result from slightly higher energy barriers of the GDML model in the pathway from A to B corresponding to the collective motions of the carboxylic acid and ester groups in aspirin. These differences vanish once the quantum nature of the nuclei is introduced in the PIMD simulations (29). In addition, long–time scale simulations are required to completely understand the dynamics of molecular systems. Figure 4B shows the probability distribution of the fluctuations of dihedral angles of carboxylic acid and ester groups in aspirin. This plot shows the existence of two main metastable configurations A and B and a short-lived configuration C, illustrating the nontrivial dynamics captured by the GDML model. Finally, we remark that a similarly good performance as for aspirin is also observed for the other seven molecules shown in Fig. 3. The efficiency of the GDML model (which is three orders of magnitude faster than DFT) should enable long–time scale PIMD simulations to obtain converged thermodynamic properties of intermediate-sized molecules with the accuracy and transferability of high-level ab initio methods.

In summary, the developed GDML model allows the construction of complex multidimensional PES by combining rigorous physical laws with data-driven ML techniques. In addition to the presented successful applications to the model systems and intermediate-sized molecules, our work can be further developed in several directions, including scaling with system size and complexity, incorporating additional physical priors, describing reaction pathways, and enabling seamless coupling between GDML and ab initio calculations.

SUPPLEMENTARY MATERIALS

Supplementary material for this article is available at http://advances.sciencemag.org/cgi/content/full/3/5/e1603015/DC1

section S1. Noise amplification by differentiation
section S2. Vector-valued kernel learning
section S3. Descriptors
section S4. Model analysis
section S5. Details of the PIMD simulation
fig. S1. The accuracy of the GDML model (in terms of the MAE) as a function of training set size: Chemical accuracy of less than 1 kcal/mol is already achieved for small training sets.
fig. S2. Predicting energies and forces for consecutive time steps of an MD simulation of uracil at 500 K.
table S1. Properties of MD data sets that were used for numerical testing.
table S2. GDML prediction accuracy for interatomic forces and total energies for all data sets.
table S3. Accuracy of the naïve force predictor.
table S4. Accuracy of the converged energy-based predictor.
References (31–36)

REFERENCES AND NOTES

1. J. Behler, M. Parrinello, Generalized neural-network representation of high-dimensional potential-energy surfaces. *Phys. Rev. Lett.* **98**, 146401 (2007).
2. J. Behler, S. Lorenz, K. Reuter, Representing molecule-surface interactions with symmetry-adapted neural networks. *J. Chem. Phys.* **127**, 014705 (2007).
3. A. P. Bartók, M. C. Payne, R. Kondor, G. Csányi, Gaussian approximation potentials: The accuracy of quantum mechanics, without the electrons. *Phys. Rev. Lett.* **104**, 136403 (2010).
4. J. Behler, Atom-centered symmetry functions for constructing high-dimensional neural network potentials. *J. Chem. Phys.* **134**, 074106 (2011).
5. J. Behler, Neural network potential-energy surfaces in chemistry: A tool for large-scale simulations. *Phys. Chem. Chem. Phys.* **13**, 17930–17955 (2011).
6. K. V. J. Jose, N. Artrith, J. Behler, Construction of high-dimensional neural network potentials using environment-dependent atom pairs. *J. Chem. Phys.* **136**, 194111 (2011).
7. A. P. Bartók, R. Kondor, G. Csányi, On representing chemical environments. *Phys. Rev. B* **87**, 184115 (2013).
8. A. P. Bartók, G. Csányi, Gaussian approximation potentials: A brief tutorial introduction. *Int. J. Quantum Chem.* **115**, 1051–1057 (2015).
9. S. De, A. P. Bartók, G. Csányi, M. Ceriotti, Comparing molecules and solids across structural and alchemical space. *Phys. Chem. Chem. Phys.* **18**, 13754–13769 (2016).
10. M. Rupp, A. Tkatchenko, K.-R. Müller, O. A. von Lilienfeld, Fast and accurate modeling of molecular atomization energies with machine learning. *Phys. Rev. Lett.* **108**, 058301 (2012).
11. G. Montavon, M. Rupp, V. Gobre, A. Vazquez-Mayagoitia, K. Hansen, A. Tkatchenko, K.-R. Müller, O. A. von Lilienfeld, Machine learning of molecular electronic properties in chemical compound space. *New J. Phys.* **15**, 095003 (2013).
12. K. Hansen, G. Montavon, F. Biegler, S. Fazli, M. Rupp, M. Scheffler, O. A. von Lilienfeld, A. Tkatchenko, K.-R. Müller, Assessment and validation of machine learning methods for predicting molecular atomization energies. *J. Chem. Theory Comput.* **9**, 3404–3419 (2013).
13. K. Hansen, F. Biegler, R. Ramakrishnan, W. Pronobis, O. A. von Lilienfeld, K.-R. Müller, A. Tkatchenko, Machine learning predictions of molecular properties: Accurate many-body potentials and nonlocality in chemical space. *J. Phys. Chem. Lett.* **6**, 2326–2331 (2015).
14. M. Rupp, R. Ramakrishnan, O. A. von Lilienfeld, Machine learning for quantum mechanical properties of atoms in molecules. *J. Phys. Chem. Lett.* **6**, 3309–3313 (2015).
15. V. Botu, R. Ramprasad, Learning scheme to predict atomic forces and accelerate materials simulations. *Phys. Rev. B* **92**, 094306 (2015).
16. M. Hirn, N. Poilvert, S. Mallat, Quantum energy regression using scattering transforms. *CoRR* arXiv:1502.02077 (2015).
17. J. Behler, Perspective: Machine learning potentials for atomistic simulations. *J. Chem. Phys.* **145**, 170901 (2016).
18. Z. Li, J. R. Kermode, A. De Vita, Molecular dynamics with on-the-fly machine learning of quantum-mechanical forces. *Phys. Rev. Lett.* **114**, 096405 (2015).
19. C. A. Micchelli, M. A. Pontil, On learning vector-valued functions. *Neural Comput.* **17**, 177–204 (2005).
20. A. Caponnetto, C. A. Micchelli, M. Pontil, Y. Ying, Universal multi-task kernels. *J. Mach. Learn. Res.* **9**, 1615–1646 (2008).
21. V. Sindhwani, H. Q. Minh, A. C. Lozano, Scalable matrix-valued kernel learning for high-dimensional nonlinear multivariate regression and granger causality, in *Proceedings of the 29th Conference on Uncertainty in Artificial Intelligence (UAI'13)*, 12 to 14 July 2013.
22. B. Matérn, *Spatial Variation*, Lecture Notes in Statistics (Springer-Verlag, 1986).
23. I. S. Gradshteyn, I. M. Ryzhik, *Table of Integrals, Series, and Products*, A. Jeffrey, D. Zwillinger, Eds. (Academic Press, ed. 7, 2007).
24. T. Gneiting, W. Kleiber, M. Schlather, Matérn cross-covariance functions for multivariate random fields. *J. Am. Stat. Assoc.* **105**, 1167–1177 (2010).
25. H. Helmholtz, Über Integrale der hydrodynamischen Gleichungen, welche den Wirbelbewegungen entsprechen. *Angew. Math.* **1858**, 25–55 (2009).
26. W. H. Press, S. A. Teukolsky, W. T. Vetterling, B. P. Flannery, *Numerical Recipes: The Art of Scientific Computing* (Cambridge Univ. Press, ed. 3, 2007).
27. J. P. Perdew, K. Burke, M. Ernzerhof, Generalized gradient approximation made simple. *Phys. Rev. Lett.* **77**, 3865–3868 (1996).
28. A. Tkatchenko, M. Scheffler, Accurate molecular Van Der Waals interactions from ground-state electron density and free-atom reference data. *Phys. Rev. Lett.* **102**, 073005 (2009).
29. M. Ceriotti, J. More, D. E. Manolopoulos, i-PI: A Python interface for ab initio path integral molecular dynamics simulations. *Comput. Phys. Commun.* **185**, 1019–1026 (2014).
30. I. Poltavsky, A. Tkatchenko, Modeling quantum nuclei with perturbed path integral molecular dynamics. *Chem. Sci.* **7**, 1368–1372 (2016).
31. A. J. Smola, B. Schölkopf, *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond* (MIT Press, 2001).
32. J. C. Snyder, M. Rupp, K. Hansen, K.-R. Müller, K. Burke, Finding density functionals with machine learning. *Phys. Rev. Lett.* **108**, 253002 (2012).
33. J. C. Snyder, M. Rupp, K.-R. Müller, K. Burke, Nonlinear gradient denoising: Finding accurate extrema from inaccurate functional derivatives. *Int. J. Quantum Chem.* **115**, 1102–1114 (2015).
34. B. Schölkopf, A. Smola, K.-R. Müller, Nonlinear component analysis as a kernel eigenvalue problem. *Neural Comput.* **10**, 1299 (1998).
35. B. Schölkopf, S. Mika, C. J. C. Burges, P. Knirsch, K.-R. Müller, G. Ratsch, A. J. Smola, Input space versus feature space in kernel-based methods. *IEEE Trans. Neural Netw. Learn. Syst.* **10**, 1000–1017 (1999).
36. K.-R. Müller, S. Mika, G. Rätsch, K. Tsuda, B. Schölkopf, An introduction to kernel-based learning algorithms. *IEEE Trans. Neural Netw. Learn. Syst.* **12**, 181–201 (2001).

Citation: S. Chmiela, A. Tkatchenko, H. E. Sauceda, I. Poltavsky, K. T. Schütt, K.-R. Müller, Machine learning of accurate energy-conserving molecular force fields. *Sci. Adv.* **3**, e1603015 (2017).