



# Building robust machine learning force fields by composite Gaussian approximation potentials<sup>☆</sup>

Diego Milardovich<sup>\*</sup>, Dominic Waldhoer, Markus Jech, Al-Moatasem Bellah El-Sayed, Tibor Grasser

Institute for Microelectronics, Technische Universität Wien, Gusshausstraße 27–29, 1040 Vienna, Austria

## ARTICLE INFO

### Keywords:

Machine learning  
ML  
DFT  
Silicon dioxide  
Force field  
Gaussian approximation potential  
GAP

## ABSTRACT

Machine learning (ML) interatomic potentials have received a lot of interest in recent years, motivated by their high accuracy at low computational costs. However, these potentials tend to overfit, which threatens their reliability. This work proposes a systematic solution to this problem, by augmenting ML potentials with simpler auxiliary potentials, which aim at ensuring that the physics behind interatomic interactions are respected. The versatility of the proposed solution is demonstrated by developing a ML force field for amorphous silicon dioxide (a-SiO<sub>2</sub>), in which a main potential is augmented with a set of simpler pairwise short-range auxiliary potentials. The resulting potential exhibits a significant improvement in transferability and scalability, at only a moderate increase in computational costs.

## 1. Introduction

The interest in machine learning (ML) interatomic potentials has intensified over the past decade. These interatomic potentials serve as an alternative to more expensive *ab initio* methods like density functional theory (DFT), for which computational costs severely narrow their application to relatively small systems (100s of atoms) and short simulation times (10s of ps). The progression of these alternative potentials is further aided by readily available software packages [1–4]. However, the development of transferable ML interatomic potentials remains a topic of active research in this field. In this context, we refer to *transferability* as the ability of a ML potential to retain its accuracy when employed in tasks not trained for (e.g., to accurately reproduce forces, despite only training on energies). One of the main obstacles which weakens the transferability of ML potentials is overfitting, an undesired statistical phenomena in which ML models present a high accuracy on training data, but perform poorly in previously unseen testing data and, therefore, in real applications.

We implement our potentials as instances of the established Gaussian approximation potential (GAP) [5], a type of kernel-based model successfully used in previous studies, to develop highly accurate and computationally efficient interatomic potentials for various materials and applications [6–8]. This model assumes that similar local atomic configurations give similar contributions to the total potential energy.

It estimates the potential energy of a given atomic configuration by comparing it to those provided in a training dataset. This comparison is done by means of a similarity measure function or *kernel*, which gives this type of ML models its name. In order for a GAP, or any other ML model used in an interatomic potential, to be able to compute the potential energy of an atomic structure, a mathematical representation, or *descriptor*, of such atomic structure must be provided first. This model and its implementation in ML interatomic potentials are depicted in Fig. 1, where a descriptor of the input atomic structure is computed and used as input to the ML model, which computes the potential energy.

In this work, a systematic solution to mitigate the effects of overfitting is proposed, by augmenting a GAP with a set of independent repulsive-only potentials, in contrast to previous solutions found in the literature [6,9]. We will demonstrate our proposed solution by building a potential for amorphous silicon dioxide (a-SiO<sub>2</sub>). Our potential will be validated against DFT calculations and compared to the results obtained with a traditional ML potential.

## 2. Methodology

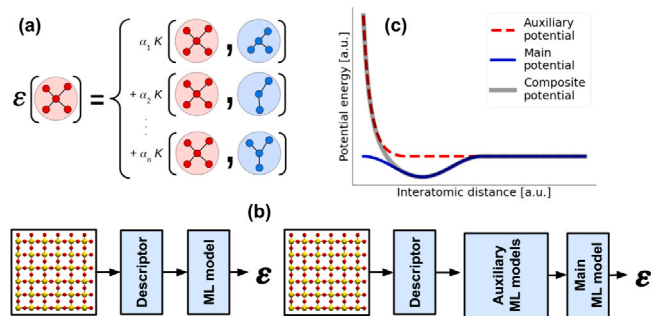
As previously mentioned, a ML interatomic potential is composed of a descriptor and a ML model. A descriptor computes a mathematical representation (i.e., a vector or matrix) of a given atomic structure

<sup>☆</sup> The review of this paper was arranged by Francisco Gamiz.

<sup>\*</sup> Corresponding author.

E-mail addresses: [milardovich@iue.tuwien.ac.at](mailto:milardovich@iue.tuwien.ac.at) (D. Milardovich), [waldhoer@iue.tuwien.ac.at](mailto:waldhoer@iue.tuwien.ac.at) (D. Waldhoer), [jech@iue.tuwien.ac.at](mailto:jech@iue.tuwien.ac.at) (M. Jech), [el-sayed@iue.tuwien.ac.at](mailto:el-sayed@iue.tuwien.ac.at) (A.-M.B. El-Sayed), [grasser@iue.tuwien.ac.at](mailto:grasser@iue.tuwien.ac.at) (T. Grasser).

<https://doi.org/10.1016/j.sse.2022.108529>



**Fig. 1.** (a) Schematic of a GAP, a ML model which computes the potential energy of a given atomic configuration as the weighted sum over the similarity measurements between the given atomic configuration and those in a training dataset. The similarities are measured by a kernel. Training this model is equivalent to finding the optimal weight values  $(\alpha_1, \alpha_2, \dots, \alpha_n)$ . (b) Workflows for a traditional potential built with this ML model (left) and our proposed composite interatomic potential (right). (c) Schematic of the proposed composite potential, augmenting the main potential with an auxiliary potential.

and serves as an intermediate between the atomic structure and the ML model. However, not any possible mathematical representation is valid as a descriptor, as they must be invariant under translation and rotation of the atomic structures, as well as under permutation of identical atoms. Such properties aid in training a ML model to compute a property from atomic structures, since, if atomic coordinates were used instead of a descriptor, the ML model would have to learn that any possible translation, rotation or atomic permutation in an atomic structure corresponds to the same atomic property. Namely, the ML model would have to learn that an infinite number of different inputs correspond to the same single output, therefore rendering the training process unfeasible. Examples of frequently used descriptors are *smooth overlap of atomic positions* (SOAP) [10] and *atom-centered symmetry functions* (ACSF) [11]. Descriptors can be either global (they represent the entire atomic system) or local (they represent the local or near environment of a given atom in the system), with the latter being the most commonly used for ML interatomic potentials. The ML model is trained to find a functional relationship between a descriptor (highly-dimensional) and the corresponding potential energy (scalar-valued) of the atomic structure which it represents. In a traditional ML interatomic potential, the total potential energy of a given atomic system is approximated by a sum of local energy contributions from every atom, computed by a **single ML model** [12],

$$E_{\text{total}} = \sum_i^{\text{Atoms}} E_i(\mathbf{d}_i),$$

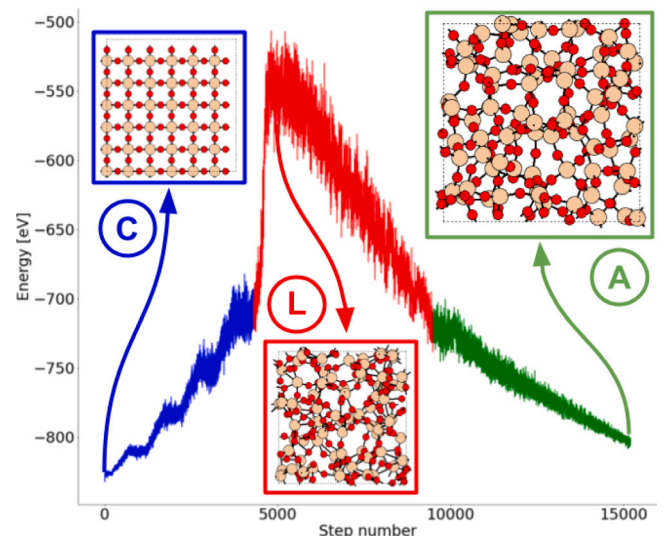
where  $\mathbf{d}_i$  is a local descriptor for the environment of the  $i$ th atom and  $E_i$  is its local contribution to the total potential energy, computed by the ML model from the local descriptor. We propose to build a potential which computes these local energies as a **composite set of multiple components**, i.e.,

$$E_i = \sum_j^{\text{Models}} E_i^j(\mathbf{d}_i^j),$$

where  $j$  runs over all components of the final potential. Each component can either be a non-parametric potential; or a ML model paired with different local descriptors, allowing for varying degrees of complexity.

In this work, we build an interatomic potential for a-SiO<sub>2</sub> composed of a complex main model responsible for giving accurate results for the atomic systems of interest and a simpler auxiliary potential, which represents the basic physics through pairwise short-range interactions, as shown in Fig. 1(c).

$$E_i = E_i^{\text{main}}(\mathbf{d}_i^{\text{main}}) + E_i^{\text{aux}}(\mathbf{d}_i^{\text{aux}})$$

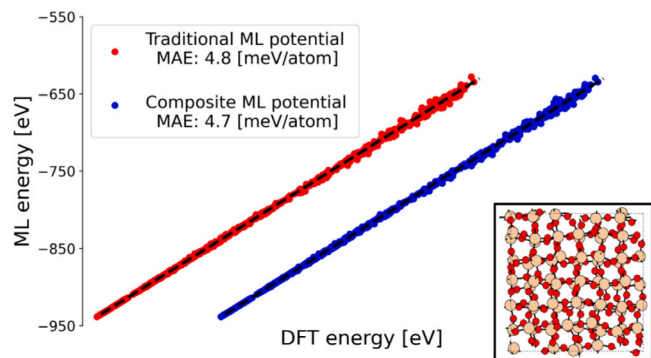


**Fig. 2.** Initial dataset for training and testing the *main* ML potential, created using the melt-and-quench technique and the ReaxFF force field. The MD begins with a crystalline SiO<sub>2</sub> structure, which is melted up to 5000 K and thereafter quenched back to 300 K. Once the dataset is created, the energies are recalculated on a subset using DFT. Blue: Crystalline phase (C). Red: Liquid phase (L). Green: Amorphous phase (A). The objective of the initial dataset is to provide a sampling of the PES of interest as comprehensive as possible. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

The main ML potential is responsible for the intricate details of the potential energy surface (PES) of the atomic systems of interest. On the other hand, the auxiliary potential is repulsive-only, meaning that it only produces positive energies. The closer two atoms are to one another, the higher the computed energy will be. This potential matches the positive part of the pairwise dissociation curve, as obtained with DFT, by using a simple two-body descriptor with a short and smooth cut-off, as implemented in QUIP (<https://github.com/libAtoms/QUIP>) [3,13]. The sum of both energies results in the energy computed by our proposed composite ML interatomic potential. In our implementation, both potentials are realized as GAPs.

The main potential employs the more sophisticated and highly-dimensional SOAP [10] descriptor. The training dataset for the main GAP was created by running molecular dynamics (MD), according to the melt-and-quench technique [14] within the LAMMPS engine [15]. A defect-free 216-atoms SiO<sub>2</sub> system was melted at 5000 K and subsequently quenched to 300 K, using the classical force-field ReaxFF [16] with a time-step of 0.25 fs. The process is depicted in Fig. 2. The resulting trajectory was sequentially sub-sampled to a training dataset of only 1500 atomic configurations, for which the energies were calculated with DFT, using the PBE functional [17] in the CP2K software package [18].

Ideally, the training dataset must be comprehensive, while containing only atomic configurations relevant to the intended application for which the ML interatomic potential will be used. In other words, the atomic environments contained in it must represent a faithful sampling of the PES subset relevant for a particular application. An under-sampling of the possible atomic environment space would undoubtedly result in poor ML prediction accuracy, while over-sampling results in an excessive computational cost. Furthermore, the sampling should not be biased towards any particular area of the PES, in order not to over-sample a sub-set of the possible atomic configurations space, while under-sampling the rest. It is for this reason that the training dataset is composed of a sub-set of the atomic configurations obtained by performing MD. The testing dataset is composed of 1000 randomly selected atomic configurations from the same melt-and-quench MD trajectory.



**Fig. 3.** Testing the potential energy accuracy for the traditional (red) and the composite (blue) ML potentials against DFT. The dashed black line represents the ideal results, in which all energy predictions would be equal to the DFT calculations. The closer the dots are to this line, the better the results are. This reference makes the  $X$ -axis immaterial. The results show a high accuracy for both potentials and virtually no difference between them. An example of the atomic structures used for testing is shown in the bottom-right sub-panel. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

It is important to notice that no atomic configuration belonging to the training dataset was used in the testing dataset.

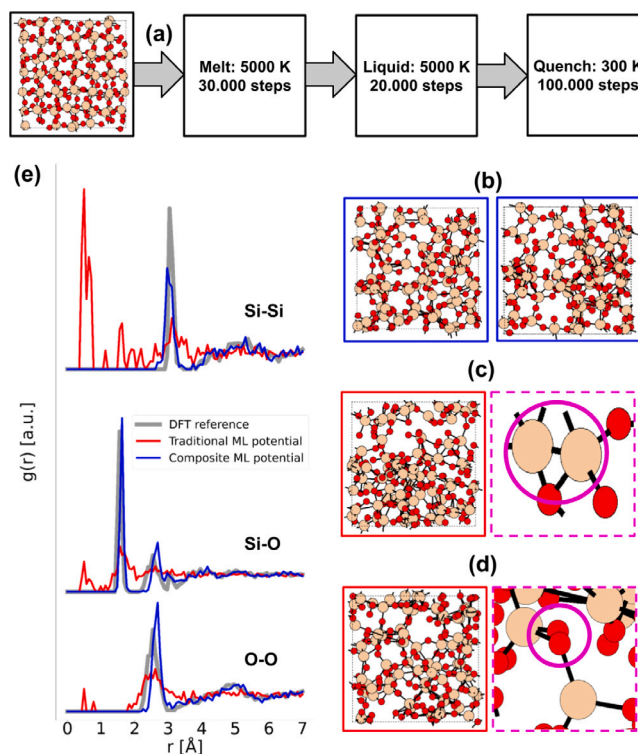
The main potential was trained on the residual between the DFT energies and the predictions of the auxiliary potential for the atomic configurations in the training dataset. The training was performed using the software package QUIP. The descriptor for the auxiliary potential was defined using smooth cut-offs:  $r_{\text{cut}}^{\text{SiSi}} = 1.60 \text{ \AA}$ ,  $r_{\text{cut}}^{\text{SiO}} = 1.10 \text{ \AA}$  and  $r_{\text{cut}}^{\text{OO}} = 0.80 \text{ \AA}$ . The SOAP descriptor for the main GAP was constructed with  $n_{\text{max}} = 6$ ,  $l_{\text{max}} = 6$  and  $r_{\text{cut}}^{\text{SOAP}} = 4.0 \text{ \AA}$ . Therefore, the main potential acts on the entire atomic neighborhood, while the auxiliary potential acts only if two atoms are unphysically close to one another.

### 3. Results and discussion

Following the traditional approach, a single GAP was trained on the dataset used to train the main ML model of our proposed composite potential. This potential was thereafter used to evaluate the performance of our proposed composite ML interatomic potential. In the following tests, both potentials were used to perform a number of tasks and their results were compared to one another, taking DFT calculations as reference.

The first test for the composite and traditional ML potentials was to compute the potential energies of the 1000 atomic configurations in the testing dataset. The results of each ML potential were compared to DFT calculations and they are shown in Fig. 3. Our proposed composite ML potential yielded an mean absolute error (MAE) of 4.7 meV/atom, therefore rendering it slightly more accurate than the traditional potential, which yielded a MAE of 4.8 meV/atom. In practical terms, this negligible difference in accuracy between both potentials means that they should produce near identical results when used in real applications. However, the following tests will demonstrate that this first impression is wrong and that the simple auxiliary potential plays an important role in the overall ML potential.

The second test employed both ML potentials to run the exact same melt-and-quench MD. This MD simulation begins with an initially defect-free crystalline structure, which is melted at 5000 K for a period of 30,000 steps (step-size = 1 fs). Once melted, the atomic system remains at a temperature of 5000 K for another 20,000 steps and it is then quenched back to 300 K within a period of 100,000 steps. The final result of this melt-and-quench MD is an a-SiO<sub>2</sub> structure, as shown in Fig. 4. The structures produced by the traditional model suffer from unphysical atomic cluster formations, as is apparent from the radial density functions (RDFs). In contrast, our composite potential shows

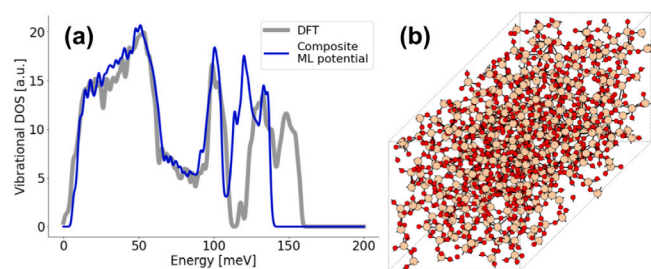


**Fig. 4.** Panel comparing our proposed framework results with those of a traditional ML interatomic potential, built as a single GAP. Both potentials were used to run the same MD, as specified in (a). Two of the resulting structures when using our proposed approach are presented in (b). Resulting structures from using the traditional ML potential are presented in (c) and (d), together with examples for unphysical behavior found in them. The RDFs for one of the structures built with our approach and one built with the traditional ML potential are presented in (e), together with the reference from the training dataset. As it can be seen, the composite potential performs much better when compared against the DFT reference.

excellent agreement with the DFT reference structure. The reason for this is that training datasets built by running MD rarely contain short-range interactions, as they are high in energy and therefore unlikely to be present in the MD trajectory. This prevents the traditional ML potential from learning that short interatomic distances correspond to high energies, resulting in the unphysical clustering seen in the MD results. Since this information is explicitly included in the auxiliary potential, our approach is much more robust while also retaining a high accuracy, with only little computational overhead (roughly 5%).

In order for a ML interatomic potential to be useful in practical applications, it must be scalable. This means that it must retain its accuracy for atomic systems noticeably larger than those used to train it. Ideally, the ML interatomic potential would also be able to yield accurate results when applied to tasks for which it was not specifically trained on, thus indicating a successful and comprehensive training process. As previously explained, accurate results for testing situations similar to those found in the training, paired with inaccurate results for different and new situations, would be a strong sign of overfitting.

To validate the transferability of our composite ML interatomic potential, the vibrational density of states (VDOS) for one of the resulting a-SiO<sub>2</sub> structures was computed and compared against DFT results, as shown in Fig. 5(a). The VDOS was determined by employing a finite-difference scheme in order to expand the system's PES to second order around the minimum configuration [19]. We choose an a-SiO<sub>2</sub> resulting from an MD run and relaxed it further using DFT, as well as our composite potential. The finite difference calculations were carried out by displacing a single atom by 0.01 Å along each coordinate axis. The resulting force constant was then diagonalized to obtain the



**Fig. 5.** Transferability tests for our proposed composite ML potential. (a) Computing the VDOS and comparing it to DFT results. The good match is remarkable, considering that the potential was not trained on forces. (b) Using the potential to create an a-SiO<sub>2</sub> structure 10 times the size of those used in the training dataset. No unphysical clustering is present.

system's normal modes and the corresponding phonon energies. As it can be seen, there is excellent agreement between DFT and our ML potential in predicting the VDOS for low-energy phonons. The fact that our ML potential is able to produce accurate results for a task for which it was not trained is a strong indication of a successful training process. The VDOS differ for higher energies; however, if higher accuracy was needed in that energy range, this could be improved by including forces in the training process. Including forces in the training process would provide the ML potential with the explicit value of the first derivative of the PES with respect to the atomic coordinates and a much more accurate representation of the PES curvature, drastically reducing the numerical errors when computing the VDOS.

The training and testing datasets for this application were composed of atomic configurations extracted from the same MD run, meaning that they all share certain distinctive properties, such as the number of atoms and the size of the cell in which they are contained. Moreover, as all frames belong to the same trajectory, a certain degree of correlation or similarity is therefore expected between the atomic environments of the structures in both datasets. It is for these reasons that, to validate the scalability of our composite ML potential, a new test was needed, in which we employed it to perform MD on an atomic system 10-times the size of the atomic structures found in the training dataset. The results are shown in Fig. 5(b), where no signs of unphysical clustering were found, therefore hinting at a successful training process and indicating the scalability of our proposed composite ML interatomic potential. It is important to highlight that these melt-and-quench MDs were performed using a relatively high quench rate, which allowed them to quench in barely 100,000 steps. A feat out of reach for several empirical potentials. This, paired with the small dataset needed for training, makes our ML potential highly computationally efficient.

#### 4. Conclusions

ML models are a powerful tool for developing highly accurate and computationally efficient interatomic potentials. However, when underlying physical mechanisms are not explicitly taken into account, their transferability is in question. In our proposed approach, the ML potential is augmented by a set of simpler potentials. Results show that this significantly reduces the unphysical behavior when facing unexplored atomic environments, at a moderate increase in the computational costs. Future improvements to this work would include the incorporation of forces to the training process, as opposed to the use of only energies; and the use of a wider range and variety of ML

models combined with different descriptors with different degrees of complexity.

#### Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

#### Data availability

Data will be made available on request.

#### Acknowledgments

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 871813, within the framework of the project Modeling Unconventional Nanoscaled Device FABrication (MUNDFAB). The authors acknowledge support from the Vienna Scientific Cluster (VSC); and financial support from TU Wien Bibliothek, Austria, through its Open Access Funding Programme.

#### References

- [1] Khorshidi A, et al. Amp: A modular approach to machine learning in atomistic simulations. *Comput. Phys. Comm.* 2016;207:310–24.
- [2] Wang H, et al. DeePMD-kit: A deep learning package for many-body potential energy representation and molecular dynamics. *Comput. Phys. Comm.* 2018;228:178–84.
- [3] Kermode JR. F90wrap: an automated tool for constructing deep python interfaces to modern fortran codes. *J. Phys.: Condens. Matter* 2020;32:305901.
- [4] Lot R, et al. PANNA: Properties from artificial neural network architectures. *Comput. Phys. Comm.* 2020;256:107402.
- [5] Bartók AP, et al. Gaussian approximation potentials: The accuracy of quantum mechanics, without the electrons. *Phys Rev Lett* 2010;104:136403.
- [6] Sivaraman G, et al. Machine-learned interatomic potentials by active learning: amorphous and liquid hafnium dioxide. *Npj Comput Mater* 2020;6:104.
- [7] Bartók AP, et al. Machine learning a general-purpose interatomic potential for Silicon. *Phys. Rev. X* 2018;8:041048.
- [8] Deringer VL, et al. Machine learning based interatomic potential for amorphous carbon. *Phys. Rev. B* 2017;95:094203.
- [9] Mocanu FC, et al. Modeling the phase-change memory material,  $Ge_2Sb_2Te_5$ , with a machine-learned interatomic potential. *J. Phys. Chem. B* 2018;122:8998–9006.
- [10] Bartók AP, et al. On representing chemical environments. *Phys. Rev. B* 2013;87:184115.
- [11] Behler J. Atom-centered symmetry functions for constructing high-dimensional neural network potentials. *J Chem Phys* 2011;134.
- [12] Behler J, et al. Generalized neural-network representation of high-dimensional potential-energy surfaces. *Phys. Rev. Lett.* 2007;98:146401.
- [13] Gábor C, et al. Expressive programming for computational physics in fortran 95+. *IoP Comput Phys Newsletter* 2007. Spring 2007.
- [14] El-Sayed A-M, et al. Identification of intrinsic electron trapping sites in bulk amorphous silica from ab initio calculations. *Microelectron Eng* 2013;209:68–71.
- [15] Plimpton S. Fast parallel algorithms for short-range molecular dynamics. *J. Comput. Phys.* 1995;117:1–19.
- [16] Fogarty JC, et al. A reactive molecular dynamics simulation of the silica-water interface. *J. Chem. Phys.* 2010;132:174704.
- [17] Perdew JP, et al. Generalized gradient approximation made simple. *Phys. Rev. Lett.* 1996;77:3865.
- [18] VandeVondele J, et al. Quickstep: Fast and accurate density functional calculations using a mixed Gaussian and plane waves approach. *Comput. Phys. Comm.* 2005;167:103–28.
- [19] Baroni S, et al. Phonons and related crystal properties from density-functional perturbation theory. *Rev. Modern Phys.* 2001;73.