Active Learning of Atomic Size Gas/Solid Potential Energy Surfaces via Physics Aware Models

Nikolaos Patsalidis ,[†] Mohsen Doust Mohammadi,[†] Somnath Bhowmick,[†] George Biskos,^{†,‡} and Vagelis Harmandaris^{*,†,‡,¶}

†Computation-based Science and Technology Research Center, The Cyprus Institute, 2121, Cyprus

Department of Mathematics and Applied Mathematics, University of Crete, Heraklion, GR-71110, Greece

¶Institute of Applied and Computational Mathematics, Foundation for Research and Technology -Hellas, Heraklion, GR-71110 Crete, Greece

E-mail: v.harmandaris@cyi.ac.cy

Abstract

We propose an active learning (AL) framework to develop force fields (FFs) that accurately model the potential energy surfaces (PES) of gas/solid atomic-scale complexes. A central challenge is integrating AL with flexible, physics-aware potentials to achieve quantum-level accuracy for complex interfacial systems. Our approach trains physics-aware potentials, with incorporated flexibility and smoothness, on actively sampled Density Functional Theory (DFT) data to describe interactions between undercoordinated atomic silver (Ag) clusters and gaseous pollutants (CO₂, CO, SO₂), relevant for environmental applications like sensing. The AL process follows three stages: (1) FFs are trained using adaptable physics aware potentials

of semi-empirical descriptors, optimized via a Pareto analysis scheme; (2) refined FFs generate candidate structures through Metropolis Hastings Monte Carlo (MHMC) or stochastic molecular dynamics (sMD); (3) a subset of candidates is selected for DFT labeling based on an outlier score (OS), which utilizes the existing data descriptor distributions, ensuring diverse PES exploration. This framework produces FFs capable of capturing cohesive, physisorption, and chemisorption interactions with accuracy comparable to *ab initio* methods and advanced machine learning models, while retaining the efficiency of semi-empirical potentials. Our methodology is highly versatile, easily accommodating various choices of descriptors, model basis sets, and sampling techniques.

1 Introduction

The holy grail in the development of atomistic models, necessary for performing atomistic molecular dynamics (MD) or Monte Carlo (MC) simulations, is to obtain classical force fields (FFs) of high accuracy and computational efficiency. A promising route to address this challenge is based on active learning (AL) approaches. These involve strategic selection of quantum chemical data, often obtained through computationally expensive density functional theory (DFT) calculations, and the active training and refining of machine-learned FFs on diverse data, generated via efficient exploration schemes of the underlying potential energy surface (PES).^{1–8} Such machine-learned FFs, although highly accurate, present high computational costs compared to semi-empirical potentials that rely on physically justified potential energy terms when applied to long MD simulations.^{9–13} At the same time, for MD simulations to have a profound impact on material properties prediction and design, long simulations, nowadays of the order of microseconds (μ s), are required. Therefore, semi-empirical potentials are practically still used in the majority of simulation studies for complex molecular systems. An important question is whether AL approaches can be effectively combined with low-cost, physics-informed models.

Any atomistic model that describes the PES of an underlying molecular system via predefined descriptors of the atomic environment should have certain properties. These include invariances to

Euclidean transformations (translation, rotation, reflection) and permutation of chemically equivalent atoms, FF smoothness (implies continuity), and (the most challenging) faithfulness; that is, each configuration should be described by a unique descriptor.^{14–16} In practice, besides faithfulness, the above properties are inherently satisfied by the choice of the descriptor and the potential form, i.e., the functional basis set used to describe the underlying PES. For example, the descriptors in popular semi-empirical FFs used in soft matter, like OPLS,¹⁰ and COMPASS,¹² are the bond lengths, the bond angles, the dihedral angles, and pairwise distances between non-bonded atoms. Models such as the embedded atom model (EAM) and modified EAM,^{11,13} applied in solid-state of matter, such as metals, include many-body interactions via a scalar density descriptor, which counts the effective number of neighbors of each atom by expanding on pairwise distances and, hence, preserving pairwise-like efficiency.

FFs developed by Machine Learning (ML) methods are practically extended functional basis sets describing n-body interactions, which are of quantum origin, at the classical level.^{14–26} Functional many-body representations, such as the Smooth Overlap of Atomic Positions (SOAP), are commonly used in kernel-based ML methods to predict by interpolation energies and forces, using the similarity of the new descriptor vector with the reference data.^{15,23,27–29} In a similar way, neural network (NN) models are used to represent many-body interactions, usually using complex non-linear mappings of atomic descriptors to accurately compute the interaction potential.^{5,14,17,22,24,25} Such ML models demonstrate high accuracy (1-10 meV/atom), but their computational cost scales linearly with the size of the reference data or of the NN. They are also prone to overfitting and may predict unrealistic energy values when extrapolating beyond their training domain.^{1,17,21,27,29–31} Finally, significant improvement in the extrapolation ability of the model could be achieved by considering physics-aware 2- and 3-body descriptors, like pairwise distances, in combination with extended ML basis sets.²¹

Extrapolation and overfitting issues show that adequately sampling the PES is crucial. To address this, AL algorithms are often used to ensure diversity in training data.^{1,2,4–7,31–33} For instance, Vandermause et al.³¹ used the inherent uncertainty in a Gaussian process regression model to decide whether to accept its prediction or perform a DFT calculation during temperature-annealing MD simulations of crystalline materials. Kulichenko et al.⁴ used metadynamics to sample different local minima of the PES and estimated model uncertainty via the disagreement among multiple NN models capturing proton transfer in glycerine. Wilson et al.³² developed a batch AL scheme that enhances the training set with various structures selected from a pool of data generated by *ab initio* MD simulations of germanium selenide (GeSe). In each AL iteration they selected the most dissimilar structures with existing data, based on the Euclidean distance of the feature vectors. Moreover, their model was based on 2- and 3-body terms resembling semi-empirical forms. In another recent work, Bernstein et al.⁶ developed a batched AL scheme that samples new data based on structural dissimilarity, using energy minimization trajectories to explore different PES local minima and, hence, stable structures of crystalline materials.

The studies referred above are very promising, but they consider either crystalline solid-state phases, small isolated molecules, or water clusters. In contrast, to the best of our knowledge, complex interfaces and/or undercoordinated hybrid materials have not yet been addressed. Examples include systems associated with important technological applications that rely on molecules adsorbed on surfaces³⁴ and gas/metal hybrids,³⁵ relevant to emerging environmental challenges such as pollutant detection, clean energy production and decarbonization. For example, atomic silver nanoclusters are very promising for gas sensing applications due to the different gas adsorption characteristics, which could be either of physisorption, or chemisorption character and of different interaction strength.³⁵ Moreover, they can be readily produced by innovative gas-phase synthesis methods, such as atmospheric-pressure spark ablation.³⁶These systems are extremely challenging in terms of modeling, as they exhibit an extremely rough PES and are characterized by a variety of complex quantum chemical interactions within scales of a few Å, Moreover, adsorption, desorption or nucleation around solid nanoclusters phenomena occurs on a wide range of time scales, from a few *fs* that corresponds to adsorption of single molecules to atomic clusters, up to several μs , relevant to cluster nucleation.^{37,38}

To address the above challenges, the development of accurate and preeminently efficient atom-

istic FFs is essential. Moreover, the FF basis set should be tunable to different characteristic interactions and, ideally, transferable over a range of systems with different cluster sizes and temperatures. Additionally, the potential form should be generic and systematically improvable through a suitable AL framework that efficiently explores the PES.

Here we employ a batch AL strategy that actively trains generalized forms of semi-empirical potentials by exploring and sampling the PES at the DFT level. We apply it to a variety of complex interfacial systems, *i.e.*, isolated silver (Ag) atomic clusters (of 7-16 atoms), and on CO_2/Ag , CO/Ag, SO₂/Ag undercoordinated atom sized hybrids. The FF basis set follows physically aware descriptors, e.g., bond distances and bond angles, pairwise distances between atoms and embedding densities that describe the local environment. Their dependency to the potential energy is direct, similar to the physically aware semi-empirical potential forms like Morse; moreover, our model also incorporates flexibility by considering potential curves of arbitrary shape, parameterized via, respectively, the flexible Bezier polynomials³⁹ (see Sections 2.1 and 2.2). Such parameterization allows for a smooth (continuously differentiable) and controllable potential that, with only a few extra training parameters, can capture complex underlying physicochemical interactions. Using Bezier polynomials, we include flexible pairwise corrections and flexible many-body potential terms. Part of the model captures the essential physics, while another part introduces the necessary flexibility to fit complex quantum deviations. In addition, this model maintains pairwise efficiency and, as we will show, it is transferable with a small trade-off in accuracy across different atomic cluster sizes. Even more importantly, the derived family of models is capable of describing very rough PES of the atom-sized gas/metal hybrids with quantum accuracy, via relatively simple physics-aware cost-efficient basis functions, without the need to consider more complex descriptors commonly used in ML FFs.¹⁷

2 Methods

The overall AL methodology is summarized in Figure 1. First, an initial dataset is used for training. Due to numerical noise of the DFT data (data uncertainty) and perhaps inadequacies of the model (model or epistemic uncertainty), minimizing the differences between improving the fit on the force labels may worsen the fit in energy labels and vice versa. Our training method minimizes the energies cost at descending values of forces cost, finding Pareto optimal solutions. We keep the solution that exhibits the best trade-off between the energy cost and the force cost (see Section 2.3). In the next step, the trained or refined FF is used to generate a large pool of new candidate gas/cluster hybrid structures (here 40,000) via either Metropolis Hastings Monte Carlo (MHMC) or stochastic Molecular Dynamics (sMD) simulations (see Section 2.4). Since they are generated using the model, these candidates are unlabeled (no ground truth values of energies and forces). From these, a small subset of structures (here a batch of 200) is selected with a probability proportional to each candidate's outlier score (OS). The latter is quantified based on the overlap of each new descriptor to the corresponding descriptor distribution of the existing data. This algorithm promotes selection of out-of-distribution or rarely occurring structures (see Section 2.5), and hence, the structures that have not yet been "seen" during training. In the final step, the selected data undergo DFT labeling and are used for performance evaluation and augmentation of the dataset for the next AL iterations. This is an effective way to actively sample and explore new regions of the PES, augmenting the dataset with diverse structures. The labeled dataset at each AL iteration is referred as existing (e).

The proposed AL-based methodology does not explicitly depend on the model itself, nor on the type of the descriptors. Any model, functional basis set, and any set of descriptors could be used that respect the permutational, translation, and rotational invariance, such as NNs that utilize symmetry functions as their input descriptors or kernel methods.^{14,15,25} Finally, our algorithm effectively generates the training data, since it can be initiated by just one point per chemical system, as we demonstrate in Section 3. In this context, exploiting and labeling out-of-distribution candidates becomes even more important, since ideally we would like to perform as fewer DFT

calculations as possible.



Figure 1: Schematic of the active learning (AL) methodology. Beginning from an initial labeled dataset, with as many as 1 structure per system, we train the model by finding Pareto optimal solutions and choosing the one that shows the best trade-off between the energies cost and forces cost. In the second step, using either Metropolis Hastings Monte Carlo (MHMC) or stochastic Molecular Dynamics (sMD) a large pool of candidate structures (here, 40,000) is generated. MHMC is performed on 100 parallel paths, starting from respectively 100 initial configurations, chosen randomly from the existing labeled dataset. A single longer trajectory is acquired using sMD, starting respectively from 1 random configuration from the existing dataset. In the next step, for each candidate, an outlier score (OS) is computed and a subset (batch) of these structures is selected (here 200), via a probability proportional to OS. DFT is performed on the selected structures to label their energies and forces. The newly labeled batch is used for performance evaluation and augmenting the existing dataset in the next AL iterations.

The subsections below describe in detail (a) the classical representation of the complex gas/cluster hybrid PES (classical model) and (b) its parameterization, each step of the AL methodology, including (c) the training procedure, (d) the candidate sampling methods, (e) the OS quantification method, and (f) the data labeling methods, DFT settings and performance metric. Finally, (g) we provide a small description of the AL initialization and the code.

2.1 Classical description of gas/cluster PES

The initial part of any classical model concerns the choice of the appropriate descriptors that describe the underlying many-body, of quantum origin, PES. Here a physics aware parametrization of the complex PES of the gas/atomic cluster hybrid is proposed as the sum of bonded $(U_B(\mathbf{r}))$, pairwise non-bonded $(U_{PW}(\mathbf{r}))$ and embedding $(U_E(\mathbf{r}))$ contributions, using the following functional basis set

$$U_{class}(\mathbf{r}) = U_B(\mathbf{r}) + U_{PW}(\mathbf{r}) + U_E(\mathbf{r}).$$
(1)

The bonded terms in the above expression include interactions describing chemical bonds, bond angles, and dihedral angles as

$$U_B(\mathbf{r}) = \sum_b f_b(r_b) + \sum_a f_a(\theta_a) + \sum_d f_d(\phi_d)$$
(2)

where r_b is the length of the bond indexed by b, θ_a the bond angle indexed by a, ϕ_d the dihedral angle indexed by d, and f_b , f_a , f_d are functions used to describe interactions between chemical bonds, angles and dihedrals respectively. The pairwise terms are of the general form

$$U_{PW}(\mathbf{r}) = \sum_{i < j} f_{pw}(r_{ij}) \tag{3}$$

where r_{ij} is the distance between non-bonded atoms *i* and *j* and f_{pw} the functions that describe the pairwise interactions. Last, the embedding contributions that typically describe the many-body terms that correspond to the local chemical environment

$$U_E(\mathbf{r}) = \sum_{i,\beta \in B_i} f_e\left(\boldsymbol{\rho}_{\alpha\beta}^i\right) \tag{4}$$

where the sum runs over all central atoms *i*, α is the i^{thst} atom chemical species, β runs over the neighboring chemical species of *i* denoted as the set B_i and f_e are properly chosen embedding functions. We should stress that in the latter model, despite that it accounts for many-body interactions, forces reduce to pairwise summations, and therefore the computational efficiency is preserved.^{40,41} f_b , f_a , f_c , f_{pw} , and f_e are the basis functions (potential curves) to be optimized given the current training set. The above representation is a generalized flexible form of a semi-empirical FF with an arbitrary shape of the potential curves f and direct dependence on the physics aware descriptors

(here, pairwise distances, embedding densities, bond distances and bending angles). We should also note that it is quite straightforward to introduce additional terms in the above functional basis set such as neural networks or Gaussian processes that are typically used in ML-based classical force fields.

The embedding part of the model, that is of great importance for the accurate description of many-body interactions, is further clarified below. Each density description of a central atom *i* of type α , $\rho_{\alpha\beta}^{i}$, is calculated by

$$\rho^{i}_{\alpha\beta} = \sum_{j\in\beta} \varphi(r_{ij}) \tag{5}$$

where *j* runs over all neighboring atoms of type β , r_{ij} is the distance of the atom *j* to the central atom and $\varphi(r)$ is the Lucy activation function (see Supporting Information (SI) Section S1 and Figure S1) Such symmetry function has been employed in coarse grained modeling and they have been implemented in Lammps MD software, ^{40,42,43} however, to the best of our knowledge it is the first time that they are used in atomistic modeling of interfacial systems. Taking the derivatives it is easy to see that the force on atom *i*, of type α from embedding contributions of atoms of type β is given by

$$\mathbf{F}_{i} = -\sum_{j \in \beta} \left[\frac{\partial f_{\alpha\beta}(\boldsymbol{\rho}_{\alpha\beta}^{i})}{\partial \boldsymbol{\rho}} + \frac{\partial f_{\beta\alpha}(\boldsymbol{\rho}_{\beta\alpha}^{j})}{\partial \boldsymbol{\rho}} \right] \frac{d\boldsymbol{\varphi}(r_{ij})}{dr} \frac{\mathbf{r}_{i} - \mathbf{r}_{j}}{r_{ij}}$$
(6)

The first term corresponds to embedding *i* into the atomic environment of β type atoms at positions \mathbf{r}_j . The second term is the contribution by atom *i* when embedding atom *j* of type β in the atomic environment of α type atoms. Preeminently, equation (6) demonstrates that the calculation of forces reduces to pairwise summations.

2.2 Parameterization

All functions are parameterized partially or completely with Bezier polynomials,⁴⁴ instead of the commonly used cubic splines.^{40,45} Each embedding term is parameterized with a single Bezier curve, while pairwise terms are parameterized via the superposition of a Morse potential and a

flexible Bezier curve. Bonds and bond angles are also considered for the covalent interactions in CO, CO_2 and NO_2 molecules, using a superposition of a Morse and Bezier curves and a harmonic potential, respectively. The parameterization of the model's different energetic contributions is summarized below:

$$f_b(r) = f_{Morse}(r; D_e, r_e, \alpha) + f_{Bezier}(r; \mathbf{k})$$
$$= D_e \left[1 - e^{-a(r-r_e)} \right]^2 + f_{Bezier}(r; \mathbf{k})$$
(7)

$$f_a(\theta) = f_{harmonic}(\theta; \theta_0, k) = k(\theta - \theta_0)^2$$
(8)

$$f_{pw}(r) = f_{Morse}(r; D_e, r_e, \alpha) + f_{Bezier}(r; \mathbf{k})$$
$$= D_e \left[e^{-2a(r-r_e)} - 2e^{-a(r-r_0)} \right] + f_{Bezier}(r; \mathbf{k})$$
(9)

$$f_e(\rho_{\alpha\beta}) = f_{Bezier}(\rho_{\alpha\beta}; \mathbf{k}) \tag{10}$$

where terms on the right of ";" denote the training parameters to be optimized. Note that Morse potentials for bonded interactions possess a minimum at zero energy, since bonded interactions describe the internal deformation of the molecule, where the DFT energy is also set to zero at the energetically minimized configuration. Moreover, adding f_{Bezier} in the bonded part $f_b(r)$, offers more flexibility, which is especially useful when bonded interactions are modified by the presence of other materials in the environment such as Ag atomic clusters (*e.g.* molecules chemisorbed to the cluster).

Notice that a part of the potential preserves physical awareness, while Bezier curves add the necessary flexibility to capture complex physicochemical interactions. Bezier curves preserve some advantages such as smoothness, continuity, controllability of the shape of the curves with only a few more training parameters (8-13 per curve), compared to the commonly used cubic splines that preserve locality.³⁹ However, to our knowledge, this parameterization scheme has never been used before in the development of atomistic models. Detailed information on the Bezier curves and their current implementation is provided in the SI, while a demonstration of their construction is shown in Figure S2.

From here in, we denote the descriptors by a single vector \mathbf{d} per structure. Therefore, the total energy of the system might be written as a function of \mathbf{d} and a parameter set \mathbf{P} , which defines the shape of the curves.

2.3 Model Training

To optimize the classical gas/atomic cluster hybrid model, we consider both the energy loss (\mathscr{L}_E) and the force loss (\mathscr{L}_F) defined as the mean square error (MSE) over energy and force data derived via DFT calculations per each configuration as

$$\mathscr{L}_{E}(\mathbf{d}(\mathbf{r});\mathbf{P}) = \left(\overline{U}(\mathbf{d}(\mathbf{r});\mathbf{P}) - \overline{E}_{dft}[\mathbf{r}]\right)^{2}$$
$$\mathscr{L}_{F}(\mathbf{d}(\mathbf{r});\mathbf{P}) = \frac{1}{3n_{a}}\sum_{j}^{n_{a}}\sum_{k}^{3} \left(\overline{F}^{j,k}(\mathbf{d}(\mathbf{r});\mathbf{P}) - \overline{F}_{dft}^{j,k}[\mathbf{r}]\right)^{2}$$
(11)

where **d**, denotes the set of the classical descriptors (*e.g.*, pairwise distances and embedding densities) of the model, which are a function of the configuration **r**, **P** is the set of training parameters, *U* and *E* denote the classical and DFT energies respectively, F^{j} is the force on each atom *j*, *k* denotes the direction, n_{a} is the number of atoms, and overline symbols denote normalized values such that

$$\overline{U}(\mathbf{d}(\mathbf{r});\mathbf{P}) = \frac{U(\mathbf{d}(\mathbf{r});\mathbf{P}) - \mu_E}{\sigma_E}$$

$$\overline{E}_{dft}[\mathbf{r}] = \frac{E_{dft}[\mathbf{r}] - \mu_E}{\sigma_E}$$

$$\overline{F}(\mathbf{d}(\mathbf{r});\mathbf{P}) = \frac{F(\mathbf{d}(\mathbf{r});\mathbf{P}) - \mu_F}{\sigma_F}$$

$$\overline{F}_{dft}[\mathbf{r}] = \frac{F_{dft}[\mathbf{r}] - \mu_F}{\sigma_F}$$
(12)

where μ_E , σ_E is the mean and standard deviation of the DFT energy data and μ_F , σ_F is the mean and standard deviation of the DFT forces. Denoting \mathbb{D} the training set data, the energy cost (C_E), and the force $cost(C_F)$ are defined as

$$C_E(\mathbb{D}, \mathbf{P})) = \frac{1}{n_d} \sum_{i}^{n_d} (\overline{U}(\mathbf{d}(\mathbf{r}_i); \mathbf{P}) - \overline{E}_{dft}[\mathbf{r}_i])^2$$
(13)

$$C_F(\mathbb{D}, \mathbf{P})) = \frac{1}{3n_d n_a} \sum_{i}^{n_d} \sum_{j}^{n_a} \sum_{k}^{3} (\overline{F}^{j,k}(\mathbf{d}(\mathbf{r}_i); \mathbf{P}) - \overline{F}_{dft}^{j,k}[\mathbf{r}_i])^2$$
(14)

where n_d is the number of configurations or equivalently data points. To find the optimal set of parameters we initially solve the optimization problem

$$\mathbf{P}_{0} = \underset{\mathbf{P}}{\operatorname{argmin}} C_{E}(\mathbb{D}, \mathbf{P})) + \lambda ||\mathbf{P}||$$
(15)

where λ is a constant regularization parameter. Using the initially estimated solution \mathbf{P}_0 and eq. (15) we compute the $C_F^0(\mathbb{D}, \mathbf{P}_0)$. Then, we iteratively minimize the energy cost at successively lower values of the force cost, by setting an equality constraint for force cost at successively lower values. The overall training algorithm is shown below:

 Algorithm 1 Iterative Optimization of P

 $\mathbf{P}_0 \leftarrow \arg\min_{\mathbf{P}} (C_E(\mathbb{D}, \mathbf{P}) + \lambda ||\mathbf{P}||)$
 $C_F^0 \leftarrow C_F(\mathbb{D}, \mathbf{P}_0)$

 for k = 1 to n_S -1 do

 Compute constraint value: $c_k \leftarrow C_F^0 \times \frac{n_S - k}{n_S}$

 Solve the optimization problem:

 $\mathbf{P}_k \leftarrow \arg\min_{\mathbf{P}} (C_E(\mathbb{D}, \mathbf{P}) + \lambda ||\mathbf{P}||)$

 subject to: $C_F(\mathbb{D}, \mathbf{P}) - c_k = 0$

 end for

The successive constraint optimization problems are solved via the gradient-based Sequential Quadratic Programming (SQP) method,⁴⁶ starting from various randomized initial parameters. Solving for $k = 0, 1, ..., n_S - 1$ provides respectively, n_S different solutions, given that the equality constraints can be satisfied. Here we set $n_S = 15$. Note that during optimization C_E and C_F should in principle be non-conflicting objectives; however, this might not be the case due to noise in the data and the (relatively small) size of the dataset. Moreover, considering only energies may result in a solution in the training parameter space that does not necessarily minimize C_F , due to

overfitting of particular energetic contributions. This is an efficient framework not only to find Pareto optimal solutions but to evaluate different ones. We select the solution that has the best trade-off between the two using the selection metric

$$C_m = \sqrt{C_E^2 + C_F^2} \tag{16}$$

Finally, we split the existing data during each AL iteration into a training (train) and a development (dev) set in 80/20%, respectively. The training set is used in algorithm 1, while the development set is used to evaluate C_m and select the best solution.

2.4 Candidate Sampling via Stochastic Methods

The refined FF at each AL iteration is used to generate new candidate structures for labeling. When the initial dataset is extremely small (<10 points) 10 new candidates per point are generated by adding a small random Gaussian noise (~ N(0,0.02) in Å) to the initial structures . For early stage models, here between the 2nd up to the 10th AL iterations, candidates are generated via many (here 100) parallel Metropolis-Hastings Monte Carlo (MHMC) simulations, starting from different initial structures, chosen randomly from the existing labeled data. If the existing data structures, we choose all of them. On each step, a random Gaussian perturbation (~ $N(0,\sigma)$), where σ is autotuned (discussed below), on one randomly selected atom is applied. The proposed (new) structure is accepted as candidate one with probability

$$\min\left(1, e^{-b_a\left(U(\mathbf{d}^{new}; \mathbf{P}) - U(\mathbf{d}^{old}; \mathbf{P})\right)}\right)$$
(17)

where $b_a = 1/k_B T_a$, k_B is the Boltzmann constant and T_a is an annealing temperature that becomes higher as we accept new data, following the rule

$$T_a = T_s min\left(1, 2\frac{N_{acc}}{N_{tot}}\right) \tag{18}$$

where T_s is the maximum sampling temperature, N_{acc} is the number of accepted structures and N_{tot} is the total number of desired candidate structures, here set to 40,000. Hence, half of the structures are sampled at increasing temperature values between 0 and T_s and half at the maximum sampling temperature. Note that in such small systems temperature is ill-defined; therefore T_s is an effective hyperparameter that controls the extent of the energy distribution and energy fluctuations.

The reasoning behind the slow increase of T_a to T_s is twofold: (a) It allows for a smooth scan near PES local minima and (b) can effectively capture transition phenomena that may happen at unknown temperatures, below a hypothesized T_s , like desorption of the gas molecules to the atomic clusters. Such phenomena may be very fast at high effective temperatures and extremely slow at low ones, where in both cases information on the path of this structural rearrangement may be lost. Slowly increasing T_a allows for smoother transitions.

Finally, the step length σ is autotuned such that the acceptance ratio (AR) remains between 0.2 and 0.5. It is not allowed to exceed 0.2 Å and does not go below 0.001 Å. If the running mean of the AR is below 0.2 σ , it is multiplied by 0.99, while if it is above 0.5, it is divided by 0.99. The initial σ is set to 0.02 Å. Moreover, the first 100 MC steps are used for equilibration and therefore no structures are accepted. This MC sampling procedure ends when $N_{acc} \ge N_{tot}$.

For later stage models, after the 10th AL iteration, stochastic molecular dynamics, sMD, (Langevin dynamics) simulations are used starting from one initial configuration. The system runs for 2 ns, with a timestep of 0.1 fs to avoid plausible numerical instabilities, especially of early stage models. The damping factor used to calculate the friction forces as described by ref⁴⁷ was set to 1 fs. Structures are sampled every 50 fs, resulting in 40,000 candidate structures. In contrast to MC sampling, we chose just one initial configuration from the existing dataset, which is energetically minimized. In the first 1 ns, we anneal the temperature from 1 K to T_s , similarly to the MC sampling procedure. sMD is performed using the open source software LAMMPS.⁴²

2.5 Data Selection & Outlier Score

Next, our goal is to select, from the above pool of many (here 40,000) candidate structures, a relatively small subset (here we use a predefined batch of n_{batch} =200 structures) for DFT labeling. The selected configurations are used to evaluate the performance of the model and at the same time enrich our dataset for the next iteration of the AL-based optimization scheme. For this, we use a metric of the variability of the candidate structures with respect to the structures in the existing dataset. In more detail, the selection probability of each structure is proportional to its computed outlier score (OS) which is calculated as follows:

For each element *d* in the descriptor vector $\mathbf{d} = [d_1, d_2, \dots, d_m]^T$ of a given (candidate) structure, we compute a distribution overlap parameter:

$$\omega(d) = \int_{h_{low}}^{h_{up}} h(x) e^{-a(x-d)^2} dx$$
(19)

where h(x) represents the histogram density of the descriptor element *d* (which is also an estimate of the probability distribution function) in the existing data set, *x* is an integration parameter, h_{low} and h_{up} denote the lower and upper bounds of the distribution, respectively, and *a* is a scaling parameter set as $a = h_{up} - h_{low}$. For example, if *d* represents a specific pairwise atomic distance in a candidate structure, then h(x) denotes the probability distribution function (PDF) of this distance in the existing dataset. The overlap of the distribution $\omega(d)$ is effectively a smooth approximation of the density of the histogram at *d* and practically a measure of the variability of a new candidate configuration with respect to the underlying probability distributions of each element in the descriptor vector. This value is normalized as

$$\overline{\boldsymbol{\omega}}(d) = \frac{\boldsymbol{\omega}(d) - \boldsymbol{\omega}_{min}}{\boldsymbol{\omega}_{max} - \boldsymbol{\omega}_{min}}$$

where

$$\omega_{\min} = \min_{z} \left(\int_{h_{low}}^{h_{up}} h(x) e^{-a(x-z)^2} dx \right) \qquad \forall z \in [h_{low}, h_{up}]$$
(20)

$$\boldsymbol{\omega}_{max} = \max_{z} \left(\int_{h_{low}}^{h_{up}} h(x) e^{-a(x-z)^2} dx \right) \qquad \forall z \in [h_{low}, h_{up}]$$
(21)

Note that, here, *z* is just a minimization or maximization parameter and ω_{min} and ω_{max} depend only on the distribution h(x). $\overline{\omega}(d)$ takes values in [0, 1], when *d* is within the range of the distribution $[h_{low}, h_{up}]$ and can take negative values if $\omega(d) < \omega_{min}$, which denotes out of distribution data. Especially at the early stages of the AL iteration where h(x) is very narrow and $\omega_{max} - \omega_{min}$ is small $\overline{\omega}(d)$ may take highly negative values.

In the case of a one-dimensional descriptor vector, selecting data with a probability of $1 - \overline{\omega}(d)$ favors out-of-distribution or rare data points. Therefore, the outlier score OS is computed by

$$OS(d) = 1 - \overline{\omega}(d) \tag{22}$$

and each new candidate is selected via a probability

$$s \propto \frac{OS - OS_{min}}{OS_{max} - OS_{min}} \tag{23}$$

where OS_{min} and OS_{max} are the minimum and maximum OS values in the set of the current candidates. The latter normalization ensures that the selection probability is much higher for the relatively more deviating candidates. Extended discussion and example distributions of h(x) and their corresponding OS computation is provided in Figure S3 in the SI.

For a multi-dimensional descriptor vector $\mathbf{d} \in \mathbb{R}^m$, as is the case in this work,

$$OS(\mathbf{d}) = \left(\frac{1}{m}\sum_{i=1}^{m} \left(1 - \overline{\omega}(d_i)\right)^3\right)^{1/3}$$
(24)

where the power mean is used to emphasize larger values of $1 - \overline{\omega}(d)$, ensuring that rare or out-of-

distribution descriptor values contribute more to the OS.

2.6 Data Labeling & Performance Evaluation

In the above framework E_{dft} and F_{dft} denote the label energy and forces calculated from DFT. The absolute value of the energy calculated at the DFT level should be modified according to the nature of the underlying interactions embodied in the form of $U(\mathbf{r})$. For the case of isolated atomic metal nanoclusters the energy is $E_{dft} := E_{coh}$, where *coh* denotes cohesive interactions, so that $U_{coh}(\mathbf{r}) \approx E_{coh}[\mathbf{r}]$ for all plausible configurations \mathbf{r} . For the notation below let the subscript denote the nature of the interaction and the superscript the structures involved. Let *cl* and *gas* denote the structure of the cluster and gas respectively. No subscript denotes the total energy of the system. Then, the cohesive energy for the cluster is given by

$$E_{coh}^{cl}[\mathbf{r}^{cl}] = E^{cl}[\mathbf{r}^{cl}] - \sum_{t} n_t E_t^{1atom}[\mathbf{r}^{1atom}]$$
(25)

where *t* denotes the different atom types, n_t the number of atoms of that type and $E_t^{1atom}[\mathbf{r}^{1atom}]$ is the energy of one atom of type *t*. The cluster deformation energy is given by

$$E_{def}^{cl}[\mathbf{r}^{cl}] = E^{cl}[\mathbf{r}^{cl}] - E_{min}^{cl}[\mathbf{r}_{min}^{cl}]$$
(26)

where *min* denotes the minimum total energy and the corresponding structure. Similarly for the gas,

$$E_{def}^{gas}[\mathbf{r}^{gas}] = E^{gas}[\mathbf{r}^{gas}] - E_{min}^{gas}[\mathbf{r}_{min}^{gas}]$$
(27)

Now, let the interaction energy in the cluster/gas complex (cg)

$$E_{int}^{cg}[\mathbf{r}^{cg}] = E_{ads}^{cg}[\mathbf{r}^{cg}] - E_{def}^{cg}[\mathbf{r}^{cg}]$$
(28)

where,

$$E_{ads}^{cg}[\mathbf{r}^{cg}] = E^{cg}[\mathbf{r}^{cg}] - E_{min}^{cl}[\mathbf{r}_{min}^{cl}] - E_{min}^{gas}[\mathbf{r}_{min}^{gas}]$$
(29)

and

$$E_{def}^{cg}[\mathbf{r}^{cg}] = E_{def}^{cl}[\mathbf{r}^{cl}] + E_{def}^{gas}[\mathbf{r}^{gas}]$$
(30)

Plugging equations (29), (30), (26) and (25) into (28) and after elimination of terms

$$E_{int}^{cg}[\mathbf{r}^{cg}] = E^{cg}[\mathbf{r}^{cg}] - E_{min}^{gas}[\mathbf{r}_{min}^{gas}] - E_{def}^{gas}[\mathbf{r}^{gas}] - E_{coh}^{cl}[\mathbf{r}^{cl}] - \sum_{t} n_{t} E_{t}^{1atom}[\mathbf{r}^{1atom}]$$
(31)

which may be rearranged as

$$E_{int}^{cg}[\mathbf{r}^{cg}] + E_{def}^{gas}[\mathbf{r}^{gas}] + E_{coh}^{cl}[\mathbf{r}^{cl}] = E^{cg}[\mathbf{r}^{cg}] - \left(E_{min}^{gas}[\mathbf{r}_{min}^{gas}][\mathbf{r}^{cg}] + \sum_{t} n_{t} E_{t}^{1atom}[\mathbf{r}^{1atom}]\right)$$
(32)

The latter form is particularly convenient. First, for each data point, only the total energy of the complex $E^{cg}[\mathbf{r}^{cg}]$ must be calculated since the remaining terms on the right-hand side of the equation are calculated only once and their sum is denoted by E_{ref} . Most importantly, it allows one to tackle all the interactions simultaneously by assuming that

$$U_{int}^{cg}(\mathbf{r}^{cg}) + U_{def}^{gas}(\mathbf{r}^{gas}) + U_{coh}^{cl}(\mathbf{r}^{cl}) \approx E^{cg}[\mathbf{r}^{cg}] - E_{ref} := E_{dft}$$
(33)

The above formulation assumes that the cluster/gas interaction, the gas deformation, and cluster cohesive interactions are decoupled. Hence, it allows tackling the cluster/gas system together with their isolated cluster and gas counterparts, augmenting the dataset in the sense that we effectively label the energy of the cluster/gas system at well-separated distances. This is particularly important for learning the different energetic contributions of eq. (33).

2.6.1 Density Functional Theory Settings

All DFT calculations were performed using the Gaussian software.⁴⁸ with ω B97XD hybrid functional and the def2SVP basis set.

2.6.2 Performance Metric

After we label the new structures with DFT energies, we evaluate the performance of the new model via the mean absolute error (MAE), defined as

$$MAE(\mathbb{D}, \mathbf{P}) = \frac{1}{n_d} \sum_{i}^{n_d} |U(\mathbf{d}(\mathbf{r}_i); \mathbf{P}) - E_{dft}[\mathbf{r}_i]|$$
(34)

Note that this metric is used for the evaluation of the model and not during the training; for the latter, the iterative optimization scheme, including energies and forces, is used.

2.7 Active Learning Initialization

The AL procedure may start by using one or more configurations from one system or a collection of systems. If a collection of systems is used, then the algorithm will assume different interactions based on the type of chemical elements. For example, including in the initial dataset Ag clusters of different sizes, the algorithm will train one model describing interactions between Ag atoms for all the different cluster sizes. All initial structures were obtained from the work of Mohammadi *et.* al..³⁵ and were re-optimized using the current DFT settings.

2.7.1 Code

The code, written mainly in Python 3 and bash, is a package of scripts that automatically communicates with the high-performance facility scheduler SLURM and, respectively, performs the DFT calculations in parallel using the Gaussian software. Similarly, it automatically communicates with LAMMPS. The code takes as input one single input file, which includes the settings and an initialization of the model parameters. All model parameters may be considered as fixed or training parameters, allowing for transfer learning options.

3 Results and Discussion

3.1 Application to the Ag₇CO₂ hybrid

We apply the AL algorithm to a diverse range of systems characterized by varying physicochemical interactions, demonstrating the broad applicability of our approach. These systems include undercoordinated Ag clusters governed by cohesive interactions, due to metallic bonds, as well as Ag/gas atomic cluster hybrids, where both strong physisorption and chemisorption interactions complicate the PES. Note also that models developed in the bulk phase, such as EAM,¹¹ are not suitable, exactly due to the heterogeneous coordination of these systems. In addition, we explore different cluster sizes to assess the robustness of the algorithm. To illustrate its effectiveness and behavior, we first employ the method to actively optimize the FF describing the interactions within the Ag_7CO_2 hybrid (comprising 7 Ag atoms and one CO_2 molecule). We then validate its general applicability and transferability in Section 3.2.

Throughout this work, we assume that the covalent interactions of the gas molecule, the cohesive forces among the metallic atoms, and the gas/metal interactions are decoupled (see eq. (33)). Consequently, the cohesive metallic interactions and the covalent interactions in the gas molecule can, in principle, be optimized independently of the metal/gas ones. Indeed, we use the AL algorithm to optimize the potential describing the covalent interactions of the isolated CO₂ molecule. Due to its simplicity, six AL iterations with $n_{batch} = 50$ were enough and the prediction error was about 1.0 meV/atom. The respective potential terms are kept fixed during the training of the gas/metal FF. In contrast, the cohesive forces among the Ag atoms are optimized simultaneously with the gas/metal interactions. We initialize the algorithm with both the isolated Ag₇ cluster and the Ag₇CO₂ hybrid. By generating and augmenting with data of the isolated cluster, and fixing the covalent potential terms optimized for isolated CO₂, we effectively "encourage" the model to learn the energies of well separated systems (the gas being far away from the metal cluster) and, hence, the true energetic contributions (e.g. cohesive, gas physisorption) to the total energy.

Note, that in this framework, we label the DFT energy as the sum of all the interactions in each system (see Section 2.6). Labeling the metal/gas interaction energy explicitly would demand, for each new structure of the hybrid, computing the DFT energies of the deformed Ag_7 cluster and CO_2 molecule in an isolated environment. Although in some interfaces this deformation energy was assumed to be negligible in the past,³⁴ here, the undercoordinated clusters deform significantly. Moreover, the problem could become intractable if more than one gas molecule were considered in the system.

The total training, development and prediction MAEs of the derived models, trained on data from both isolated Ag₇ and Ag₇CO₂ structures, as a function of the AL iterations are shown in Figure 2(a). During the first AL iteration the FF is trained on just one data point per system. Due to that very initial small dataset, in the first iteration 20 new candidates (10 per system) are generated by adding a small random Gaussian noise ($\sim N(0, 0.02)$ in Å) to the 2 initial structures. Consequently, the prediction error seems small. Next, during the 2nd up to the 10th AL iterations, new candidate structures are generated by MHMC, while for the consequent AL iterations (from the 11th up to the last one), sMD is performed, as described in the Methods section 2.4. Initial steps show very low training and development MAE; however, this is a result of overfitting to the small initial dataset and the restricted exploration of the PES, giving rise to high prediction MAE. Later, the training/development set error is approximately constant at about 30 meV, indicating the model's capacity. Switching to sMD at the 11th iteration results in a higher prediction error, especially for the hybrid Ag₇CO₂ system, as shown in Figure 2(b). As the AL algorithm explores new regions of the PES, it is likely to show increased prediction error, such as cases where the molecule desorbs or reorients in a different way near the Ag cluster. This increase indicates that sMD explores new regions of the PES more effectively.

Evidently, the selected structures, from the pool of candidates generated from sMD (after the 10th AL iteration), exhibit higher outlier scores (*OS*, defined in Eq. (24)) as depicted in Figure 3. This Figure depicts the mean, standard deviation, maximum, and minimum values of the *OS*

for the selected batch at each AL iteration, defined respectively as \overline{OS} , OS_{std} , OS_{max} , and OS_{min} , for the sake of discussion. By definition, $OS(\mathbf{d}) > 1$ denotes that, with respect to the specific descriptor \mathbf{d} , the candidate structure is outside of the distribution of \mathbf{d} in the existing dataset, while values close to one denote a structure that is dissimilar to the majority of the structures in the existing dataset. In the first few (two-three) iterations, especially for Ag₇CO₂, the existing data distributions for each descriptor element are quite narrow, resulting in elevated \overline{OS} and OS_{max} . In later iterations \overline{OS} is near 0.4. Between the 4th and the 10th AL iteration, OS_{max} stays below 1, while when using sMD, it takes values above 1 and the \overline{OS} is higher, indicating exploration of new PES regions. Towards the last AL iterations, the \overline{OS} converges to a value of about 0.45 indicating that the underlying descriptor distributions in the existing data are converging. Moreover, the oscillations in \overline{OS} seem to correlate well with the oscillations in the MAE prediction of Figure 2(b). In addition, the differences in \overline{OS} between the isolated cluster systems and the gas hybrid demonstrate the much more complex PES of the hybrid system. Despite that, after 21 iterations, the AL seems to have explored well the PES, with prediction MAE below 6 meV/atom. This is a significant result, considering that a superposition of flexible physics-aware models is used as the functional basis set, that preserves pairwise additive computational efficiency (computational cost scales with *NlogN*, using cell link lists, where *N* is the number of atoms).

Training the FF for such systems is also a challenging task. Finding the parameter set that globally minimizes the cost function (13) is not a trivial task, especially for multi-parametric models with strong non-linear dependencies and interfacial PES.³⁴ Here, we select the solution that ensures the best trade-off between the energy cost (C_E , eq. (13)) and the force cost (C_F , eq. (14)) among the Pareto optimal solutions. Pareto optimal solutions are defined as the ones for which there is no other solution that is better for both objectives. As described in the Methods section, these are found by initially minimizing the C_E without considering C_F , and then minimizing C_E while constraining the C_F at subsequently descending values. Here, indicatively, we choose to show the set of calculated solutions during the 13th AL iteration in Figure 4. Not all solutions are Pareto optimal, including the initial one, due to the various different parameter sets that locally



Figure 2: (a) Training (train), development (dev) and prediction (pred) MAE (meV) of the derived models, of the Ag_7CO_2 hybrid, as a function of the AL iterations. (b) Prediction MAE in energies (meV/atom) for each system were the FF is actively trained simultaneously, as a function of AL iteration. Indicative structures are depicted. At the 11th iteration sMD resulted in desorption of the CO₂ molecule, while in the 12th sMD drove the system in unexplored regions of the PES, increasing the error.

minimize C_E . Statistical differences between the training and the development set might also affect this behavior. Preeminently, this training approach provides a way to evaluate different model solutions and choose the solution among them exhibiting the best trade-off between C_E and C_F . In the inset, we demonstrate the agreement of the "best" solution with DFT forces. The positive effect of this methodology is even more pronounced in premature models in the early stages of AL, as discussed in the SI Section S4 and shown in Figure S4. Finally, the model parameters as a function of AL iteration is shown in Figure S5. Towards the last AL iterations, these exhibit convergence, and therefore the model converges to a similar functional form.

3.2 General Applicability & Transferability

Besides the above demonstration case, we applied the algorithm in several additional systems; here we present results obtained from three more gas/atomic cluster hybrids as summarized in Table 1. The second system (Case 2) concerns a hybrid of Ag clusters with different sizes, compared to



Figure 3: Mean outlier score, \overline{OS} , for the selected batch at each AL iteration, per system. Thick caps denote standard deviation, while thin caps denote the maximum and minimum values. In the inset the data are shown in a linear plot.



Figure 4: Set of calculated solutions (blue squares) on the normalized Force (C_F) and Energy Cost (C_E) for the development set during the 13th AL iteration. Note, that not all solutions are Pareto optimal, including the initial solution (black star), where we do not include the forces in the training. The solution that exhibits the best trade-off (red circle) of these two error metrics in the development set is kept. In the inset, we show the classical model's agreement with DFT forces. Since C_F and C_E are normalized quantities they are dimensionless.

the first one, and the CO gas. CO exhibits stronger physisorption interactions with Ag, compared to the CO₂. The hybrid in Case 3 involves SO₂ and the planar atomic cluster Ag₄. The planarity of the cluster, the exchange of electrons between SO₂ and Ag (chemisorption characteristics) and the strong SO₂/Ag interactions result in an even more complex PES of the hybrid. To fit well the different energetic contributions, we optimize the FF simultaneously for all energy terms (gas covalent, cluster cohesive, and hybrid chemisorption) and explore via the AL procedure the PES for the Ag₄SO₂ hybrid and both the isolated Ag₄ and SO₂. Due to the PES roughness, we sampled for a larger number of AL iterations. Case 4 involves Ag clusters of multiple sizes, while the FF is common for all systems (Ag atoms are treated as the same type).

Figures S6–S8 present the predicted MAE and selected batch OS statistics as a function of AL iterations for Cases 2–4, respectively. A similar convergence trend is observed across these cases, with the OS correlating well with the predicted MAE. In Case 2, after 21 AL iterations, the model achieves excellent accuracy, with an MAE below 10 meV/atom. In Case 3, the model successfully captures the complex chemisorption interactions, with an MAE of approximately 30 meV/atom. This level of error is acceptable given the model's simplicity and computational efficiency. Case 4 demonstrates that a unified FF model for a range of cluster sizes (9–16 atoms) can be developed, balancing accuracy and generalizability with an MAE between 10 and 35 meV/atom. All relevant details on the derived models and their optimized parameters are provided in Tables S1–S4 for Cases 1–4, respectively.

To assess the transferability across the cluster size, we use the FF optimized at the last AL iteration of case 4 to generate a batch of 200 DFT labeled data of the Ag_{17} and Ag_{18} atomic clusters, that were not included in the training set, following the steps of 1 AL iteration. The prediction MAE is 19.3 and 14.3 meV/atom, respectively, indicating that the FF is transferable for larger cluster sizes. In addition, we use the AL algorithm to refine part of the FF, specifically for these two cluster sizes. Without increasing the number of descriptors, we include a potential energy term, described via a Bezier polynomial, that depends on the embedding density, up to embedding densities of value 6 (undercoordinated atoms). The part of the potential that was optimized in case 4 is kept fixed, while we fine-tune this new potential term. We run the algorithm 6 AL iterations and the results are shown in Figure 5. The prediction MAE drops even below 10 meV/atom. This is a "transfer learning" procedure and, although, here, we have not increased the complexity of the model, for example using sophisticated many-body descriptors, such as the SOAP,¹⁵ we demonstrate that such smooth potentials could learn most of the essential physics, improving transferability and extrapolation of the model. The refined potential parameters are provided in Table S5.



Figure 5: Prediction MAE for Ag_{17} and Ag_{18} as a function of AL iteration. In the first iteration, 200 data points are generated and predicted via the FF optimized in case 4. Then, part of this FF is fined tuned to the specific data. On each iteration 200 new data are generated.

Case	Systems	Characteristic Interactions	Performance (meV/atom)	AL iterations
Case 1	Ag_7, Ag_7CO_2	Cohesive, Physisorption	5.3, 4.3	21
Case 2	Ag ₈ , Ag ₈ CO	Cohesive, Physisorption	6.6, 6.7	21
Case 3	Ag_4, Ag_4SO_2, SO_2	Cohesive, Chemisorption	12.3, 22.7, 2.2	66
Case 4	$Ag_{9}, Ag_{10}, Ag_{11}, Ag_{12},$	Cohesive	17.5, 34.3, 23.2, 31.3,	21
	$Ag_{13}, Ag_{14}, Ag_{15}, Ag_{16}$		27.0, 21.6, 11.1, 15.2	

Table 1: Different applications of the AL algorithm. The prediction performance is averaged over the last 5 AL iterations. It is shown in order respective to the naming order in the systems column.

4 Conclusions

In this work, a batch Active Learning (AL) framework has been developed and successfully applied to complex gas and atom-sized metal cluster hybrids. The AL algorithm iteratively trains and refines physics-aware FFs. On each AL step, the refined FF is used to generate a large pool of candidate structures via stochastic methods, *i.e.*, MHMC and sMD. The temperature is annealed from 1 K to a high value, T_s (set to 500 K here), to smoothly capture potential desorption or reorientation of gas molecules around the rough atom-sized silver (Ag) clusters, thereby exploring the PES. For each candidate structure, an OS is calculated that respects the permutational, rotational, and translational invariance, as well as the permutational invariance of the descriptor vector. The AL algorithm then selects a small batch from the candidates with a probability proportional to their relative OS, efficiently augmenting with new regions of the PES since the selected structures undergo Density Functional Theory (DFT) labeling, and augment the training dataset in the next AL iteration. We should emphasize that here, we iteratively generate the training dataset, in contrast to other AL approaches that strategically select data from a predefined labeled pool.³² This consists of a great challenge, since a poor model could result in poor augmentation of the dataset.

The FFs' basis set is composed of two-body and many-body terms (and covalent if necessary), similar to semi-empirical potentials, while maintaining the computational efficiency of pairwise additive interactions. The model relies on physically driven descriptors, each independently contributing to the total potential energy and predicts well the interactions of complex gas/metal atomic size hybrids. It can capture complex interactions with high accuracy, such as strong physisorption

on rough surfaces and cohesive forces in undercoordinated atomic clusters (Ag₇CO₂, Ag₈CO). With an acceptable trade-off in accuracy, a FF describing complex chemisorption interactions for the Ag₄SO₂ hybrid is also developed. Finally, with a similar trade-off in accuracy, a transferable FF across the cluster size (Ag₉₋₁₈) is produced.

In this work, we demonstrate that AL approaches can be effectively combined with lowcost models, by generalizing the form of semi-empirical potentials and combining them with the smooth, flexible, and controllable Bezier polynomials. Such models could be used either directly for quantitative MD or as a starting point for further optimization on extended functional basis sets, akin to the "transfer learning" procedure followed in Section 3.2. For example, it is highly plausible that a small NN, which takes as input intuitive symmetry functions ¹⁴ will effortlessly fit the deviation of such physics-aware models, instead of trying to directly predict the PES. The logic is similar to the work of Deringer *et. al.*, ²¹ who showed that ML models can better extrapolate when physics aware 2- and 3-body terms are used. However, we leave the examination of this hypothesis for future work.

Supporting Information

Embedding Density Activation Function ; Bezier Parameterization; Outlier Score Computation; Discussion on the Training Method; General Applicability; Optimized Force Field Parameter Tables

Acknowledgments

V.H. and N.P. acknowledge funding from the European Union's Horizon 2020 research and innovation program (grant agreement no. 810660, SimEA). N.P., S.B., G.B. and V.H. acknowledge the financial support of the European Regional Development Fund and the Republic of Cyprus through the Research and Innovation Foundation project ML-NANOCAT (CODEVELOP-GT/0322/0093). M.M. and G.B. acknowledge funding from the European Union's Horizon 2020 research and innovation program (grant agreement No. 856612, EMME-CARE) and the Cyprus Government. We acknowledge the support by computational time granted by the Greek Research & Technology Network (GRNET) in the National Hellenic HPC facility ARIS under projects POL-CONFINED and SPMDE.

References

- Xie, Y.; Vandermause, J.; Ramakers, S.; Protik, N. H.; Johansson, A.; Kozinsky, B. Uncertainty-aware molecular dynamics from Bayesian active learning for phase transformations and thermal transport in SiC. *npj Computational Materials* 2023, 9, 36.
- (2) Sivaraman, G.; Krishnamoorthy, A. N.; Baur, M.; Holm, C.; Stan, M.; Csányi, G.; Benmore, C.; Vázquez-Mayagoitia, Á. Machine-learned interatomic potentials by active learning: amorphous and liquid hafnium dioxide. *npj Computational Materials* **2020**, *6*, 104.
- (3) Huan, T. D.; Batra, R.; Chapman, J.; Krishnan, S.; Chen, L.; Ramprasad, R. A universal strategy for the creation of machine learning-based atomistic force fields. *NPJ Computational Materials* 2017, *3*, 37.
- (4) Kulichenko, M.; Barros, K.; Lubbers, N.; Li, Y. W.; Messerly, R.; Tretiak, S.; Smith, J. S.; Nebgen, B. Uncertainty-driven dynamics for active learning of interatomic potentials. *Nature Computational Science* 2023, *3*, 230–239.
- (5) Allotey, J.; Butler, K. T.; Thiyagalingam, J. Entropy-based active learning of graph neural network surrogate models for materials properties. *The Journal of Chemical Physics* 2021, 155.
- (6) Bernstein, N.; Csányi, G.; Deringer, V. L. De novo exploration and self-guided learning of potential-energy surfaces. *npj Computational Materials* 2019, *5*, 99.

- (7) Loeffler, T. D.; Patra, T. K.; Chan, H.; Cherukara, M.; Sankaranarayanan, S. K. Active learning the potential energy landscape for water clusters from sparse training data. *The Journal* of Physical Chemistry C 2020, 124, 4907–4916.
- (8) Thaler, S.; Mayr, F.; Thomas, S.; Gagliardi, A.; Zavadlav, J. Active learning graph neural networks for partial charge prediction of metal-organic frameworks via dropout Monte Carlo. *npj Computational Materials* 2024, *10*, 86.
- (9) Hollingsworth, S. A.; Dror, R. O. Molecular dynamics simulation for all. *Neuron* 2018, 99, 1129–1143.
- (10) Jorgensen, W. L.; Tirado-Rives, J. Potential energy functions for atomic-level simulations of water and organic and biomolecular systems. *Proceedings of the National Academy of Sciences* 2005, 102, 6665–6670.
- (11) Daw, M. S.; Foiles, S. M.; Baskes, M. I. The embedded-atom method: a review of theory and applications. *Materials Science Reports* **1993**, *9*, 251–310.
- (12) Sun, H. COMPASS: an ab initio force-field optimized for condensed-phase applications overview with details on alkane and benzene compounds. *The Journal of Physical Chemistry B* 1998, *102*, 7338–7364.
- (13) Baskes, M. I. Modified embedded-atom potentials for cubic materials and impurities. *Physical review B* **1992**, *46*, 2727.
- (14) Behler, J.; Parrinello, M. Generalized neural-network representation of high-dimensional potential-energy surfaces. *Physical review letters* **2007**, *98*, 146401.
- (15) Bartók, A. P.; Kondor, R.; Csányi, G. On representing chemical environments. *Physical Review B* 2013, 87, 184115.
- (16) Shapeev, A. V. Moment tensor potentials: A class of systematically improvable interatomic potentials. *Multiscale Modeling & Simulation* 2016, 14, 1153–1173.

- (17) Unke, O. T.; Chmiela, S.; Sauceda, H. E.; Gastegger, M.; Poltavsky, I.; Schütt, K. T.; Tkatchenko, A.; Müller, K.-R. Machine learning force fields. *Chemical Reviews* 2021, *121*, 10142–10186.
- (18) Park, C. W.; Kornbluth, M.; Vandermause, J.; Wolverton, C.; Kozinsky, B.; Mailoa, J. P. Accurate and scalable graph neural network force field and molecular dynamics with direct force architecture. *npj Computational Materials* **2021**, *7*, 73.
- (19) Huang, Y.; Kang, J.; Goddard III, W. A.; Wang, L.-W. Density functional theory based neural network force fields from energy decompositions. *Physical Review B* 2019, *99*, 064103.
- (20) Fujikake, S.; Deringer, V. L.; Lee, T. H.; Krynski, M.; Elliott, S. R.; Csányi, G. Gaussian approximation potential modeling of lithium intercalation in carbon nanostructures. *The Journal of chemical physics* **2018**, *148*.
- (21) Deringer, V. L.; Csányi, G. Machine learning based interatomic potential for amorphous carbon. *Physical Review B* 2017, 95, 094203.
- (22) Behler, J. Representing potential energy surfaces by high-dimensional neural network potentials. *Journal of Physics: Condensed Matter* **2014**, *26*, 183001.
- (23) Bartók, A. P.; Payne, M. C.; Kondor, R.; Csányi, G. Gaussian approximation potentials: The accuracy of quantum mechanics, without the electrons. *Physical review letters* 2010, *104*, 136403.
- (24) Behler, J. Atom-centered symmetry functions for constructing high-dimensional neural network potentials. *The Journal of chemical physics* **2011**, *134*.
- (25) Behler, J. Four generations of high-dimensional neural network potentials. *Chemical Reviews* 2021, *121*, 10037–10072.
- (26) Drautz, R. Atomic cluster expansion for accurate and transferable interatomic potentials. *Physical Review B* 2019, 99, 014104.

- (27) Parsaeifard, B.; Goedecker, S. Manifolds of quasi-constant SOAP and ACSF fingerprints and the resulting failure to machine learn four-body interactions. *The Journal of Chemical Physics* 2022, *156*.
- (28) Pozdnyakov, S. N.; Willatt, M. J.; Bartók, A. P.; Ortner, C.; Csányi, G.; Ceriotti, M. Comment on "Manifolds of quasi-constant SOAP and ACSF fingerprints and the resulting failure to machine learn four-body interactions"[J. Chem. Phys. 156, 034302 (2022)]. *The Journal of Chemical Physics* 2022, 157.
- (29) Podryabinkin, E. V.; Tikhonov, E. V.; Shapeev, A. V.; Oganov, A. R. Accelerating crystal structure prediction by machine-learning interatomic potentials with active learning. *Physical Review B* 2019, 99, 064114.
- (30) Deringer, V. L.; Bartók, A. P.; Bernstein, N.; Wilkins, D. M.; Ceriotti, M.; Csányi, G. Gaussian process regression for materials and molecules. *Chemical Reviews* 2021, *121*, 10073–10141.
- (31) Vandermause, J.; Torrisi, S. B.; Batzner, S.; Xie, Y.; Sun, L.; Kolpak, A. M.; Kozinsky, B. On-the-fly active learning of interpretable Bayesian force fields for atomistic rare events. *npj Computational Materials* **2020**, *6*, 20.
- (32) Wilson, N.; Willhelm, D.; Qian, X.; Arróyave, R.; Qian, X. Batch active learning for accelerating the development of interatomic potentials. *Computational Materials Science* 2022, 208, 111330.
- (33) Karabin, M.; Perez, D. An entropy-maximization approach to automated training set generation for interatomic potentials. *The Journal of Chemical Physics* **2020**, *153*.
- (34) Patsalidis, N.; Papamokos, G.; Floudas, G.; Harmandaris, V. Understanding the Interaction between Polybutadiene and Alumina via Density Functional Theory Calculations and Machine-Learned Atomistic Simulations. *The Journal of Physical Chemistry C* 2022, *126*, 16792–16803.

- (35) Mohammadi, M. D.; Patsalidis, N.; Bhowmick, S.; Harmandaris, V. A.; Biskos, G. Adsorption of air pollutants onto silver and gold atomic clusters: DFT and PNO-LCCSD-F12 calculations. *RSC advances* 2023, *13*, 18014–18024.
- (36) Biskos, G. Spark Ablation; Jenny Stanford Publishing, 2019; pp 335–353.
- (37) Hajianzadeh, M.; Mahmoudi, J.; Sadeghzadeh, S. Molecular dynamics simulations of methane adsorption and displacement from graphenylene shale reservoir nanochannels. *Scientific Reports* 2023, *13*, 15765.
- (38) Jiang, S.; Jelfs, K. E.; Holden, D.; Hasell, T.; Chong, S. Y.; Haranczyk, M.; Trewin, A.; Cooper, A. I. Molecular dynamics simulations of gas selectivity in amorphous porous molecular solids. *Journal of the American Chemical Society* **2013**, *135*, 17818–17830.
- (39) Piegl, L.; Tiller, W. The NURBS book; Springer Science & Business Media, 1996.
- (40) Sanyal, T.; Shell, M. S. Coarse-grained models using local-density potentials optimized with the relative entropy: Application to implicit solvation. *The Journal of chemical physics* 2016, 145.
- (41) Shahidi, N.; Chazirakis, A.; Harmandaris, V.; Doxastakis, M. Coarse-graining of polyisoprene melts using inverse Monte Carlo and local density potentials. *The Journal of Chemical Physics* 2020, 152.
- (42) Thompson, A. P.; Aktulga, H. M.; Berger, R.; Bolintineanu, D. S.; Brown, W. M.; Crozier, P. S.; in't Veld, P. J.; Kohlmeyer, A.; Moore, S. G.; Nguyen, T. D.; others LAMMPSa flexible simulation tool for particle-based materials modeling at the atomic, meso, and continuum scales. *Computer Physics Communications* **2022**, *271*, 108171.
- (43) Sanyal, T.; Shell, M. S. Transferable coarse-grained models of liquid–liquid equilibrium using local density potentials optimized with the relative entropy. *The Journal of Physical Chemistry B* 2018, *122*, 5678–5693.

- (44) Farin, G. Curves and surfaces for computer-aided geometric design: a practical guide; Elsevier, 2014.
- (45) DeLyser, M. R.; Noid, W. Analysis of local density potentials. *The Journal of Chemical Physics* **2019**, *151*.
- (46) Nocedal, J.; Wright, S. J. Numerical optimization; Springer, 1999.
- (47) Schneider, T.; Stoll, E. Molecular-dynamics study of a three-dimensional one-component model for distortive phase transitions. *Physical Review B* **1978**, *17*, 1302.
- (48) Frisch, M.; Trucks, G.; Schlegel, H.; Scuseria, G.; Robb, M.; Cheeseman, J.; Scalmani, G.; Barone, V.; Petersson, G.; Nakatsuji, H.; others Gaussian 16, 2016. *Gaussian Inc. Wallingford CT* 2016,