MACHINE
LEARNING
Science and Technology

OPEN ACCESS

RECEIVED
27 January 2026

REVISED
15 March 2026

ACCEPTED FOR PUBLICATION
23 March 2026

PUBLISHED
17 April 2026

PAPER

Regularity priors for the linear atomic cluster expansion

James P Darby[1], **Joe D Morrow**[2], **Albert P Bartók**[3], **Volker L Deringer**[2], **Gábor Csányi**[1]
and Christoph Ortner[4,*]

[1] Engineering Laboratory, University of Cambridge, Cambridge, CB2 1PZ, United Kingdom
[2] Inorganic Chemistry Laboratory, Department of Chemistry, University of Oxford, Oxford OX1 3QR, United Kingdom
[3] Department of Physics and Warwick Centre for Predictive Modelling, School of Engineering, University of Warwick, Coventry, CV4 7AL, United Kingdom
[4] Department of Mathematics, University of British Columbia, 1984 Mathematics Road, Vancouver, BC V6T 1Z2, Canada
[*] Author to whom any correspondence should be addressed.

E-mail: christoph.ortner@ubc.ca

Keywords: machine learning interatomic potentials, over-regularization, regularity priors

Supplementary material for this article is available online

Abstract

machine-learned interatomic potentials enable large systems to be simulated for long time scales at near *ab-initio* accuracy. this accuracy is achieved by fitting extremely flexible model architectures to high quality reference data. in practice, this flexibility can cause unwanted behavior such as jagged predicted potential energy surfaces and generally poor out-of-distribution behavior. we investigate a general strategy for incorporating prior beliefs on the regularity of the target energy into linear atomic cluster expansion (ACE) models and explore to what extent this approach improves the quality of the fitted models. our main focus is an over-regularization that replicates the Gaussian broadening used in smooth overlap of atomic positions descriptors within the ACE framework. numerical tests indicate that the exact form of the prior is non-critical but that including such a prior leads to significant improvement in test errors, consistent repulsion at close-approach, eliminates spurious false minima in the potential energy and enhances stability during molecular dynamics simulations.

1. Introduction

For the past few decades density-functional theory (DFT) [1, 2] has been the workhorse of computational chemistry, however, the cubic scaling with the number of electrons places significant restrictions on the accessible simulation sizes and timescales. In contrast, classical force fields [3] offer linear scaling and are fast to evaluate, but, the simple functional forms used limit their accuracy and typically lead to a poor description of complex processes such as fracture [4], phase transitions [5, 6], or chemical reactions [7, 8]. Machine-learned interatomic potentials (MLIPs) also approximate the total energy as a sum of atomic energies [9], but do so using highly flexible architectures often involving millions of learned parameters. As such, provided they are trained on a sufficient quantity of high fidelity data, they can provide near *ab-initio* accuracy across a much wider range of simulation scenarios.

However, the extreme flexibility of ML architectures has drawbacks. In particular, MLIPs are notoriously poor at extrapolating outside the domain of the training dataset [10, 11], which often leads to catastrophic failure during molecular dynamics (MD) simulations e.g. a ‘hole’ in the potential energy surface (PES) could be encountered causing the predicted energy to decrease by $\sim 100\,\mathrm{s} - 1000\,\mathrm{s}$ of $\mathrm{eV\,atom^{-1}}$ and the simulated system to explode [12, 13]. Avoiding such behavior tends to require careful construction of the training dataset, either using labor intensive manually chosen configurations, or via automated approaches such as active learning [14, 15]. The recent wave of foundation potentials [16–24] have come a long way towards fixing this issue, primarily by using enormous training datasets in combination with careful architectural decisions. These potentials show unprecedented stability across a diverse range of materials and simulations [16] whilst retaining meaningful accuracy. However, they are

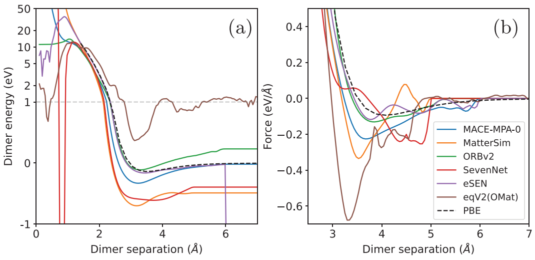

Figure 1. The energy (a) and force (b) predictions on the Mg–Mg dimer are shown for various foundation models [16–21] as a function of dimer separation. The MLIP Arena python package [26] was used to load all models and obtain the reference PBE result, originally computed with VASP [29]. For the avoidance of doubt, model versions are MACE-MPA-0, MatterSim-v1.0.0-5M, ORB-v2, 7net-0, eSEN-30M-OAM and eqV2_86M_omat_mp_salex.

not perfect, with the occasional hole still encountered for particularly unusual configurations or, more commonly, for high pressure configurations under extreme compression where there is relatively little training data available. Furthermore, some architectures are capable of achieving excellent point wise errors on energies, forces and stresses [25] whilst interpolating between the training data in a surprisingly 'rough' fashion, with the predicted PES showing small amplitude high frequency oscillations, which can be sufficiently sharp to appear almost as kinks. This behavior is seen in the homo-nuclear diatomic energy curves presented in the MLIP Arena benchmark [26] and can lead to many practical concerns, such as failure of geometry optimization or transition path search [27] or encountering false minima during crystal structure prediction [28].

The issues described above are 'obvious' failings in the sense that we can identify them almost at a glance for two key reasons. Firstly, empirically, the reference PES in most low-energy regions of configuration space for chemical and materials systems is smooth in the sense that it varies over a characteristic length scale that depends on the size of the atoms in questions, for example see figure 1. Secondly, the energy should increase sharply at short interatomic distances r due to a combination of Pauli and Coulomb repulsion, with the $1/r$ Coulomb repulsion between nuclei dominating at very close approach [30, 31]. In this work we investigate how this chemical intuition can be incorporated into linear atomic cluster expansion (ACE) models by adjusting the form of the regularization term. We focus on ACE for two reasons: (i) it remains an attractive light-weight alternative [32] to large-scale message passing neural networks; and (ii) it provides a prototype study that can be extended to many related architectures, such as MACE [33] and GRACE [24], that build on the ACE framework.

When fitting linear models the standard choice is to employ Tikhonov regularization. This leads to a single parameter, the regularization strength, which balances model accuracy (on the training set) against stability, by penalizing models with large coefficients. From a Bayesian perspective, such a choice implies a uniform prior belief for the coefficients of all basis functions, irrespective of their polynomial degree. Here, we investigate different 'shapes' for the regularization which aim to incorporate our prior belief on the smoothness of the PES. By smoothness, or more formally regularity, we not only mean that the energy should be continuous and differentiable, but also that it should not contain small amplitude high-frequency oscillations. This is analogous to the length-scale parameter within Gaussian process regression kernels [34, 35] and we show how the σ parameter in smooth overlap of atomic positions (SOAPs) descriptors [36] corresponds to a particular choice of regularization within linear ACE. In practice, we find that the exact form of the regularization term is not critical, but that picking one which incorporates a notion of regularity leads to significant improvement in model performance without incurring any additional computational cost. We validate this with a variety of numerical tests including error estimates, MD stability, and random structure searching (RSS) as well as by comparing various cuts through the predicted PES with the true reference.

2. Methods

2.1. Regularized ACE

Within the basic ACE framework [37, 38] an atom is represented as $(\boldsymbol{r}_i, Z_i)$ where $\boldsymbol{r}_i$ is the position and Z_i a categorical variable specifying the chemical element. Relative positions are denoted by $\boldsymbol{r}_{ij} = \boldsymbol{r}_i - \boldsymbol{r}_j$ and 'bonds' between atoms i, j by $x_{ij} = (\boldsymbol{r}_{ij}, Z_i, Z_j)$. The ACE site energy is defined in terms of a self-interacting many-body expansion. Let $\mathcal{N}(i)$ be the atomic environment of atom i, i.e. all atoms within distance $r_{ij} \leqslant r_{\mathrm{cut}}$, then

$$\begin{aligned} E_i &= E_i\left(\left\{x_{ij}\right\}_{j\in\mathcal{N}(i)}\right) \\ &= \sum_{N=0}^{N_{\max}} \sum_{j_1,\dots,j_N\in\mathcal{N}(i)} U_N\left(x_{ij_1},\dots,x_{i,j_N}\right). \end{aligned} \tag{1}$$

Each $(N+1)$-body potential U_N is expanded in terms of tensor products of the one-particle basis, resulting in

$$U_N\left(x_{ij_1},\dots,x_{ij_N}\right) = \sum_{\boldsymbol{nlm}} c_{\boldsymbol{nlm}} \prod_{t=1}^{N} \phi_{n_t l_t m_t}\left(x_{ij_t}\right), \tag{2}$$

where the one-particle basis ϕ_{nlm} acts on bonds x_{ij} and is defined by

$$\phi_{nlm}\left(\boldsymbol{r}_{ij}, Z_i, Z_j\right) = R_{nl}\left(r_{ij}, Z_i, Z_j\right) Y_l^m\left(\hat{\boldsymbol{r}}_{ij}\right),$$

where R_{nl} is called the *radial basis* and Y_l^m are the (real) spherical harmonics. For notational convenience there is not a specific index for the chemical element, instead the element dependence is absorbed into the n index on R_{nl}, see equation (15). The tensor product structure of the summation over j_t-tuples can be exploited to rewrite E_i in a form that is highly efficient computationally [36, 37],

$$\begin{aligned} A_{i,nlm} &= \sum_{j\in\mathcal{N}(i)} \phi_{nlm}\left(x_{ij}\right), \\ \boldsymbol{A}_{i,\boldsymbol{nlm}} &= \prod_{t=1}^{N} A_{i,n_t l_t m_t} \\ E_i &= \boldsymbol{c}\cdot\boldsymbol{A} = \sum_{\boldsymbol{nlm}} c_{\boldsymbol{nlm}} \boldsymbol{A}_{i,\boldsymbol{nlm}}, \end{aligned} \tag{3}$$

so that the evaluation cost is now approximately linear in the number of neighboring atoms and linear in the number of parameters.

Rotational, or $O(3)$, invariance is imposed by forcing the coefficients to obey a linear constraint,

$$\mathcal{L}\boldsymbol{c} = 0. \tag{4}$$

This constraint is most commonly implemented through a transformation $\boldsymbol{c} = \mathcal{C}^{\mathrm{T}}\bar{\boldsymbol{c}}$, which is equivalent to a basis transformation $\boldsymbol{B}_i = \mathcal{C}\boldsymbol{A}_i$ with $\mathcal{L} = I - \mathcal{C}^{\mathrm{T}}\mathcal{C}$; see appendix, section A for details. For our purposes it will be clearer to work with the $\boldsymbol{A}$ basis and simply keep the constraint (4) in mind.

Being a linear model, the problem of finding the coefficients using regularized least squares regression is of the form,

$$\min_{\boldsymbol{c}} ||X\boldsymbol{c} - \boldsymbol{y}||^2 + \lambda||\Gamma\boldsymbol{c}||^2, \tag{5}$$

subject to (4), where $X, \boldsymbol{y}$ is given by the fitting data, λ is a tuning parameter and the Tikhonov operator Γ will be chosen diagonal and of the form

$$\Gamma_{\boldsymbol{nlm},\boldsymbol{nlm}} = \gamma_{\boldsymbol{nl}}. \tag{6}$$

We are foremost concerned with different choices for Γ, how they can be interpreted and how they affect model performance and predicted properties. Specifically, we will explain how certain choices of γ can be interpreted as regularity priors for the potentials U_N and, by investigating the effect of each prior on the implied atomic neighbor density, draw a connection to the Gaussian broadening used when combining Gaussian approximation potentials (GAPs) with SOAP descriptors [39].

Note that we do not allow $\boldsymbol{m}$-dependent γ as this would imply a preferred orientation; see section 2.4 for more details.

2.2. Regularity priors

The Tikhonov operator Γ can also be interpreted within the Bayesian framework as a prior on the parameters; $\boldsymbol{c}\sim\mathcal{N}(\boldsymbol{0},\Gamma^{-2})$. Since the parameters $\boldsymbol{c}$ are directly related to the expansion of U_N, the operator Γ therefore encodes our 'prior belief' about those potentials. A natural class of priors are regularity priors. The idea is that, if ϕ_{nlm} is an orthogonal polynomial basis (in a suitably chosen coordinate system), then the $(N+1)$-body potentials U_N can be expanded *exactly* as

$$U_N = \sum_{\boldsymbol{nlm}} c_{\boldsymbol{nlm}} \bigotimes_{t=1}^{N} \phi_{n_t l_t m_t}. \tag{7}$$

The characteristic decay of the coefficients $c_{\boldsymbol{nlm}}$ is directly related to the regularity class to which U_N belongs. Roughly speaking, the more regular ('smooth') a function is the faster the coefficients decay. Sharp characterizations of this decay in dimension greater than one are technically subtle and depend on fine details of the interplay between the choice of basis R_{nl} and precise characterizations of multi-variate regularity of U_N which goes beyond the scope of this work. Roughly speaking, one can expect the following behavior on two prototypical cases:

$$c_{\boldsymbol{nlm}} \lesssim \begin{cases} \prod_t \left(1+\sigma_n n_t + \sigma_l l_t\right)^{-p}, & \text{if } p \text{ times diff,} \\ \exp\left(-\sum_t \sigma_n n_t + \sigma_l l_t\right), & \text{if analytic,} \end{cases} \tag{8}$$

where σ_n and σ_l are parameters. By U_N being p times differentiable, we mean that U_N has partial derivatives of the form $\partial_{\boldsymbol{r}_1}^{p_1}\cdots\partial_{\boldsymbol{r}_N}^{p_N} U_N$ up to $\max_t p_t \leqslant p$. By U_N being analytic we mean jointly analytic in all variables. We are again ignoring the m indices since we always have $|m_t| \leqslant l_t$.

If our prior belief (or knowledge) is that the U_N belong to one of those classes, then we can encode this into the prior distribution on the parameters by choosing

$$\gamma\left(\boldsymbol{n},\boldsymbol{l}\right) = \begin{cases} \prod_t \left(1+\sigma_n n_t + \sigma_l l_t\right)^{p}, & \text{if } p \text{ times diff,} \\ \exp\left(\sum_t \sigma_n n_t + \sigma_l l_t\right), & \text{if analytic.} \end{cases} \tag{9}$$

In Bayesian modeling, one generally prefers a 'cautious' prior that under-regularizes, rather than an over-confident prior that over-regularizes. In our context, this would result in choosing a lower regularity class. This rule-of-thumb is based on the assumption that there is sufficient data to specify the model. In the setting of MLIPs this is almost never the case. Even the most diverse datasets only cover a relatively small part of configuration space that is 'likely to be visited often' during inference (e.g. MD simulation). To ensure maximal robustness of the fitted models when extensive sampling of configuration space is required, it is therefore interesting to explore the effect of over-regularization, i.e. selecting regularity priors that may be theoretically too strong. In that spirit we will also consider the Gaussian prior,

$$\gamma\left(\boldsymbol{n},\boldsymbol{l}\right) = \exp\left(\sum_t \sigma_n^2 n_t^2 + \sigma_l^2 l_t^2\right). \tag{10}$$

The regularity class of functions having the analogous decay of coefficients is extremely small (the algebra of polynomials and Gaussians), hence this choice should be thought of as over-regularization and not as describing the regularity of the underlying target function. A detailed discussion on the interplay between regularity priors and symmetrization over the rotation group is given in section A.

2.3. Regularization vs rescaling

It is always possible to transform $\tilde{\boldsymbol{c}} = \Gamma\boldsymbol{c}$ so that the effective prior for $\tilde{\boldsymbol{c}}$ will be the standard Gaussian $\mathcal{N}(\boldsymbol{0},I)$. This has the advantage that standard parameter estimation methods that implicitly assume the 'canonical' Tikhonov regularization can be used. For many choices of γ the prior can also be interpreted as a rescaling of the basis functions. The priors considered above all have the form

$$\gamma\left(\boldsymbol{nlm}\right) = \prod_t \gamma_{n_t l_t}.$$

For such priors, the ACE model for the site energy transforms as

$$\begin{aligned}E_i &= \boldsymbol{c}\cdot\boldsymbol{A} = \tilde{\boldsymbol{c}}\cdot\left(\Gamma^{-1}\boldsymbol{A}\right)\\ &= \sum_{\boldsymbol{nlm}}\tilde{c}_{\boldsymbol{nlm}}\prod_t \gamma^{-1}_{n_t,l_t}A_{i,n_tl_tm_t}\\ &=: \sum_{\boldsymbol{nlm}}\tilde{c}_{\boldsymbol{nlm}}\prod_t \tilde{A}_{i,n_tl_tm_t},\end{aligned}$$

where $\tilde{A}_{i,n_tl_tm_t}$ is the rescaled atomic basis,

$$\tilde{A}_{i,n_tl_tm_t} = \gamma^{-1}_{n_tl_t}A_{i\,n_tl_tm_t}. \qquad (11)$$

If the radial basis has an l-channel then one can take this further and implement the regularity prior as a scaling of the radial basis alone via

$$\tilde{R}_{nl}\left(r_{ij},Z_i,Z_j\right) = \gamma_{nl}^{-1}R_{nl}\left(r_{ij},Z_i,Z_j\right).$$

In other words, the regularity prior can equivalently be interpreted as a specific scaling of either the one-particle basis or the radial basis R_{nl}. These observations are of practical interest for implementing ACE models, and will also be important in the next step where we draw a connection with Gaussian broadening of the atomic density in the SOAP descriptor. Furthermore, we note that rescaling the basis functions provides a natural way to include such priors in non-linear models where there is no explicit regularized least squares problem.

2.4. Implied neighbor density

ACE features can be understood in terms of symmetrized tensor products of a fictitious atomic neighbor density formed by representing each neighboring atom as a delta function. The neighbor density can be written as

$$\begin{aligned}\rho_i^{(0)}(\boldsymbol{r}) &= \sum_{j\in\mathcal{N}(i)}\delta\left(\boldsymbol{r}-\boldsymbol{r}_j\right) \qquad (12)\\ &= \sum_{nlm}A_{i,nlm}\phi_{nlm}(\boldsymbol{r}) \qquad (13)\end{aligned}$$

where here the $A_{i,nlm}$ are expansion coefficients. As such, scaling the expansion coefficients according to a specific regularity prior $\tilde{A}_{i,n_tl_tm_t} = \gamma^{-1}_{n_tl_t}A_{i\,n_tl_tm_t}$ can be interpreted as modifying the corresponding neighbor density, $\tilde{\rho}$. This connection is particularly interesting since GAP potentials, which use $N=2$ ACE features where each atom is represented by a Gaussian of width σ, have empirically shown better stability than ACE potentials fit to the same dataset [40]. To examine this relationship further, let us assume that ϕ_{nlm} are Laplacian eigenfunctions (LEs), i.e. are L^2-orthonormal and satisfy $-\Delta\phi_{nlm} = \lambda_{nl}\phi_{nlm}$ as well as $\phi_{nlm}(|\boldsymbol{r}| = r_{\mathrm{cut}}) = 0$. Note that λ_{nl} has no m dependence as Δ is isotropic. Upon writing $\gamma_{nl}^{-1} = f(\lambda_{nl})$, we then expand a general modified density

$$\begin{aligned}\tilde{\rho}_i &= \sum_{nlm}\tilde{A}_{i,nlm}\phi_{nlm}\\ &= \sum_{nlm}A_{i,nlm}f(\lambda_{nlm})\,\phi_{nlm}\\ &= \sum_{nlm}A_{i,nlm}f(-\Delta)\,\phi_{nlm}\\ &= f(-\Delta)\,\rho_i^{(0)}.\end{aligned}$$

In principle, any operator could be used in place of Δ but, as noted by Bigi *et al* [41], the Laplacian is a natural choice a.s. Specificially, its eigenstates are constructed to minimize their average squared gradient, subject to orthonormality constraints, with λ_{nl} quantifying this measure of smoothness. Importantly for our context, powers of Δ provide the means of specifying regularity classes. For example, one could minimize curvature by considering eigenfunctions of the Bi-Laplacian Δ^2 or one could measure smoothness with respect to a scaled radial coordinate. Here, we choose to use Δ as (i) λ_{nl} scales approximately as $\sim n^2 + l^2$, which allows for a simple correspondence with our regularity priors, and (ii) it enables approximate expressions for $\tilde{\rho}$ to be derived. This allows us to better understand the effects of the different priors and make a connection to the Gaussian neighbor density used in SOAP.

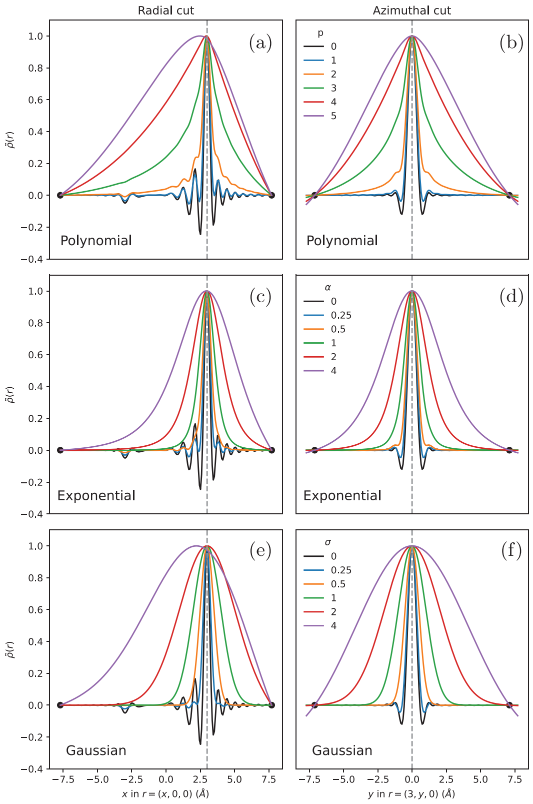

Figure 2. The neighbor density $\tilde{\rho}$ corresponding to different choices of regularity prior are shown for a single neighbor atom at $r = (3, 0, 0)$. The left-hand column shows a radial cut along $(r, 0, 0)$ while the right-hand column shows a perpendicular cut along $(3, r, 0)$. The black line shows the density with no regularity prior (a delta function represented using $n_{\mathrm{max}} = l_{\mathrm{max}} = 20$) whilst the legend indicates the regularity parameter in $\lambda_{nl}^{-p/2}$, $\exp\left(-\alpha\sqrt{\lambda_{nl}}\right)$ and $\exp\left(-\sigma^2/2\lambda_{nl}\right)$ for the polynomial, exponential and Gaussian priors respectively, where λ_{nl} are the eigenvalues of Δ. To aid visual comparison the maximum value of $\tilde{\rho}$ has been scaled to 1 for all curves. The black circles indicate the cutoff radius. The oscillatory behavior of the densities for low smoothing occur naturally in high-accuracy approximations of a delta function (cf Dirichlet kernel and analogy with Methfessel–Paxton smearing [43]) and may be beneficial in the context of very dense datasets.

For the algebraic regularity prior

$$\gamma_{nl}^{-1} \approx (1 + n + l)^{-p} \approx \lambda_{nl}^{-p/2}$$

which results in

$$\tilde{\rho} = (-\Delta)^{-p/2} \rho_i^{(0)}. \tag{14}$$

For $p = 2$ this is Poisson's equation, so that for p even $\tilde{\rho}$ can be written as $p/2$ convolutions of the associated Green's function G with the initial density $\rho_i^{(0)}$

$$\underbrace{G * \cdots * G *}_{p/2 \text{ times}} \rho_i^{(0)}.$$

Each successive convolution leads to further broadening of the density, with $\phi \sim r^{p-3}$ for $p > 0$ as shown in the top row of figure 2.

One way to explore the limit as $p \to \infty$ is to consider regularity operators that decay super-algebraically. A seemingly natural choice is an exponential, $f_\tau(\lambda_{nl}) = \exp(-\tau\lambda_{nl})$, or equivalently

$$f_\tau(-\Delta) = \exp(\tau\Delta).$$

This choice is interesting if we note that the formal solution to the heat equation $\partial_t u = \Delta u$ can be written as $u(x,t) = \exp(t\Delta)\, u(x,0)$, which then corresponds directly to $\tilde{\rho}_i = \exp(\tau\Delta)\rho_i^{(0)}$. The heat equation can also be solved by convoluting the initial density with the heat kernel [42] which results in

$$\tilde{\rho}_i = \Phi(\tau) * \rho_i^{(0)},$$

where, ignoring the previous boundary condition $\phi_{nlm}(|\mathbf{r}| = r_{\mathrm{cut}}) = 0$, $\Phi(\tau)$ is a normalized Gaussian of width $\sigma = \sqrt{\tau/2}$. As such, we see that applying the Gaussian regularity prior introduced in equation (10) $f_\tau(\lambda_{nl}) = \exp(-\tau\lambda_{nl}) \sim \exp(-\frac{1}{2}\sigma^2(n+l)^2)$ approximately corresponds to forming ACE features from a SOAP-like Gaussian broadened neighbor density. This connection can also be made in a more explicit fashion, see appendix B, by directly expanding a Gaussian neighbor density in terms of eigenfunctions of the Laplacian and then extracting $f_\sigma(\lambda_{nl})$ from the expansion coefficients.

Finally, the effect of the exponential prior on the neighbor density can be understood in a similar manner by noting that $f(\lambda_{nl}) = \exp(-\alpha\sqrt{\lambda_{nl}})$ corresponds to a fractional heat equation,

$$\tilde{\rho}_i = \exp\left(-\alpha\left(-\Delta\right)^{1/2}\right)\rho_i^{(0)}.$$

Interestingly this heat equation has an analytic kernel

$$H_\alpha = \frac{\alpha}{\pi^2\left(r^2+\alpha^2\right)^2},$$

which implies the associated neighbor density $\tilde{\rho} = H_\alpha * \rho_i^{(0)}$ has a Lorentzian shape. A comparison between the smoothing effects of all three priors, i.e. polynomial, exponential and Gaussian, on $\tilde{\rho}$ is shown in figure 2 for $n_{\max} = l_{\max} = 20$. Interestingly, despite the differences between the forms of the priors, the actual neighbor densities are strikingly similar once a moderate level of smoothing is applied.

2.5. Potential fitting

The `ACEPotentials.jl` package [44] was used to fit all ACE models shown in the subsequent section. This package uses a radial basis of the form

$$\begin{aligned} R_n\left(r_{ij}, Z_i, Z_j\right) &= R_{\zeta n'}\left(r_{ij}, Z_i, Z_j\right) \\ &= \delta_{\zeta Z_j} P_{n'}\left(y_{ij}\right) f_{\mathrm{env}}\left(r_{ij}, y_{ij}\right), \\ y_{ij} &= y\left(Z_i, Z_j, r_{ij}\right), \end{aligned} \tag{15}$$

where $y_{ij} = y(Z_i, Z_j, r_{ij})$ denotes a coordinate transform, f_{env} is an envelope function enforcing pre-defined behavior as $r \to 0$ and $r \to r_{\mathrm{cut}}$, P_n are shifted and scaled Legendre polynomials, while the chemical element dependence is folded into the composite index $n = (n', \zeta)$. Explicit functional forms are provided in [44]. The regularity priors are therefore to be understood as acting in the transformed y-space and *after* factoring out the envelope function. The priors were applied using the definitions in (8) and (10) with $\sigma_l = 1.5\sigma_n$ used for both the exponential and algebraic priors, since this corresponds to the default w_L parameter for determining basis set truncation [44]. For the Gaussian prior a different ratio, $\sigma_n = \sqrt{5}\sigma_{\mathrm{l}}$, was used; determined by the expansion of a Gaussian in the LEs basis of [41], see appendix B for details. To better facilitate comparison to typical value used with SOAP descriptors, all values of σ for the Gaussian prior will be quoted in Å where, $\sigma_n = \sqrt{5}\sigma/r_{\mathrm{cut}}$ and $\sigma_l = \sigma/r_{\mathrm{cut}}$.

To promote repulsion at very short inter-atomic distances, where Coulomb repulsion between nuclei dominates [30, 31], the training data was augmented with a single synthetic dimer configuration labeled with energy $E_{\mathrm{syn}} = 100$ eV and with separation r_{syn}, defined by $f_{\mathrm{env}}(r_{\mathrm{syn}}) = E_{\mathrm{syn}}$ for each pair of elements. Observations of energies, forces and stresses were given relative weights of $w_{\mathrm{E}} = 30$, $w_{\mathrm{F}} = 1$ and $w_{\mathrm{S}} = 1$ respectively while much smaller weights of $w_{\mathrm{syn}} = 0.01$ were used for the synthetic dimer data. Furthermore, a purified version of the ACE basis was used, where the two-body contributions were removed from all higher order terms, see [44, 45] for a detailed discussion.

3. Results

Having examined different possible forms for the regularity prior and how they can be interpreted, we now present a diverse set of numerical tests devised to assess the effect of the priors in realistic use cases. We include a combination of quick metrics, such as pointwise error estimates and 1D cuts through the PES, and more involved tests which probe stability and phase transitions during MD simulations. Silicon is used as the primary test system with additional experiments carried out on the aspirin molecule to ensure the results are not specific to single-element or condensed phase systems.

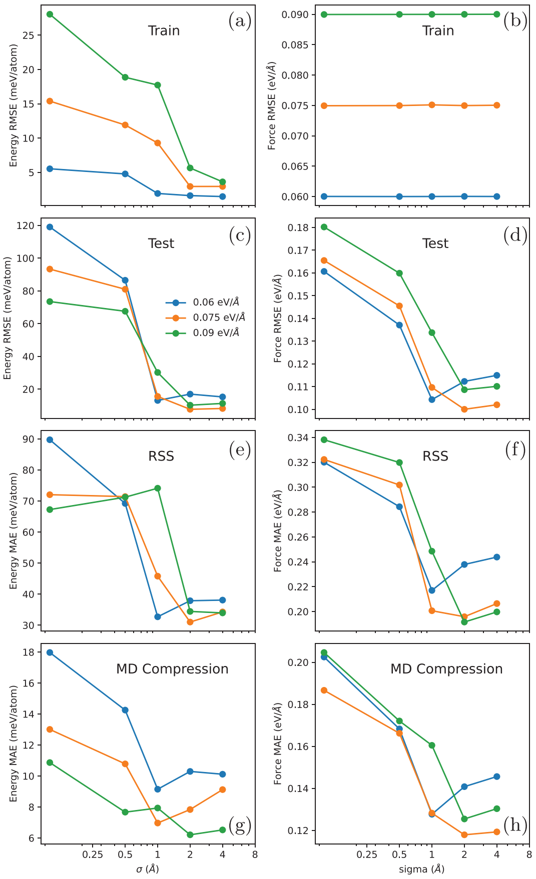

Figure 3. Force and energy errors are shown for ACE models fit using Gaussian regularity priors on configurations in the training and test sets, as well as two different configuration types from [40] that are not present in the training data, see main text for details. Larger values of σ correspond to a more regularized prior. The force error on the training set, shown in the legend, is constant as this was used to select the strength of the regularization.

3.1. Silicon

The Silicon-GAP-18 dataset from [46] was used to investigate the practical consequences of fitting ACE potentials using various regularity priors. This database is very structurally diverse, containing a mixture of crystalline, amorphous and liquid configurations as well as various crystalline surfaces, defects and dislocations. This diversity ensures good coverage of the silicon PES and made it possible to fit a general-purpose potential [46] using the GAP [34, 35, 39] framework. Specifically we fit ACE potentials to the original dataset and a smaller subset, referred to as Si10pc, containing ∼10% of the original data. We focus on results obtained on Si10pc, where the differences between different choices of priors should be more pronounced. Recent work has shown that a carefully selected 5% of the original dataset is sufficient to fit high-quality ACE potentials [14]; by contrast, Si10pc was constructed by applying a random 10:90 train:test split to the full dataset where the split was stratified by both configuration size and type. Additionally, the most and least dense configuration of each type was included in the training set to avoid excessive extrapolation in density.

The ACE potentials were constructed using a $r_{\text{cut}} = 6$ Å local cutoff and correlation orders up to $N_{\text{max}} = 4$ (5-body) features, with a maximum polynomial degree of 21 for all correlation orders. The potentials were then fit using simple Tikhonov regularization applied after the regularity prior was used to scale the basis functions appropriately as outlined in section 2.3. To distinguish between the strength and the 'shape' of the regularization, i.e. different form of regularity prior, the regularization strength was tuned such that the force root mean square error (RMSE) on the training set was equal for all models. This was repeated for three different target training force RMSEs with the resulting energy and force errors on the training and test sets shown in figure 3 for models fit using the Gaussian regularity prior. Errors were also evaluated for two further configurations types: 100 configurations from RSS [47] and 100 configurations from a MD simulation where amorphous Si was compressed from 0 to 20 GPa. These configurations were taken from ref [40] and are used to illustrate the effect of regularity priors in a highly extrapolative regime. For all configuration types, using $\sigma = 2$ Å lead to ∼40% reduction in force

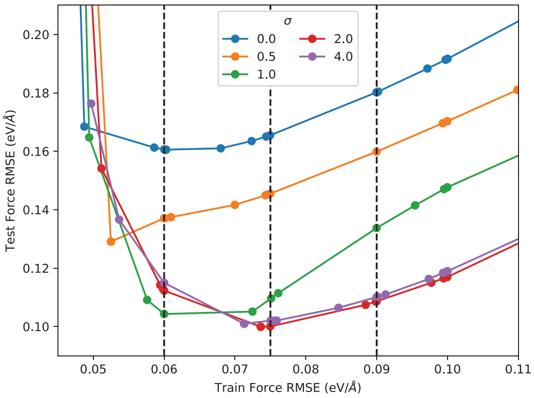

Figure 4. Force RMSE on the training and test sets is shown for potential fit to Si10pc using a Gaussian regularity prior with varying σ. Each marker corresponds to a particular choice of Tikhonov regularization strength. The dashed black lines correspond to the three different training force RMSEs shown in figure 3.

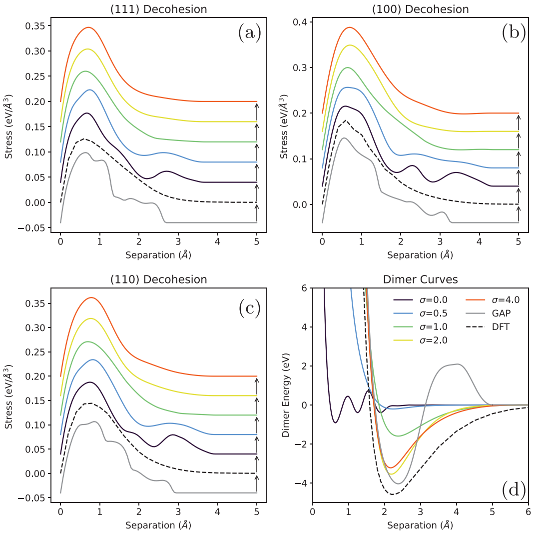

Figure 5. (a)–(c) The stress (derivative of energy) is plotted as a function of separation for decohesion of silicon diamond into the (100), (110) and (111) surfaces. The curves for different values of σ are offset (black arrows) to aid comparison. (d) The bottom right panel shows the energy of the Si–Si dimer as a function of separation between the atoms.

RMSE and up to 80% reduction in energy RMSE compared to $\sigma = 0$ (no regularity prior). As is evident from figure 4, which shows the error as a function of σ for three different choices that cover the minimum achievable test force RMSE, this is achieved regardless of the specific target force RMSE. It is notable that despite the equivalence between the Gaussian regularity priors and SOAP features, outlined in section 2.3, the optimal value for σ is much larger than would typically be used in SOAP-GAP potentials where $\sigma = 0.3$–0.5 Å is more typical [28]. This can be partially attributed to the distance transform applied to the radial coordinate, which increases the resolution near the typical nearest-neighbor position, effectively reducing σ where there is most data and increasing it elsewhere. We also note that the exact shape of the regularity prior appears to make little difference to the achievable errors, with very similar performance achieved using both the algebraic and exponential regularity priors as shown in figure S2. Given this, we only present results for the Gaussian prior (Force RMSE on the training set $= 0.075\,\mathrm{eV}\,\text{Å}^{-1}$) in the main text, with corresponding figures for the algebraic and exponential priors provided in figures S4 and S5 in the appendix.

Whilst reducing errors is highly desirable it is not the only important metric. For instance, many modern architectures achieve excellent errors but are unable to perform stable MD simulations [48], or do so at the cost of having very rough PESs with many spurious additional minima, see figure 1. Quantifying this is challenging but we can probe it by simply plotting various cuts through the PES. The simplest example of this is the Si–Si dimer interaction energy which is shown in the bottom right of figure 5. Despite all models having the same training force RMSE, the models with large values of σ

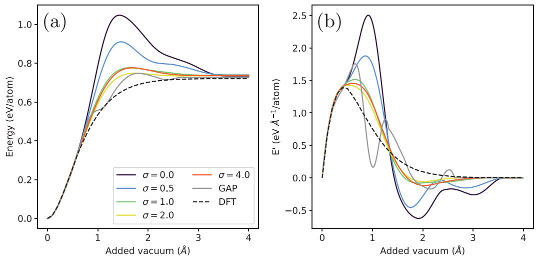

Figure 6. The energy (a) and its derivative with respect to added vacuum (b) are shown as the diamond structure silicon crystal is rigidly exfoliated into layers of silicene.

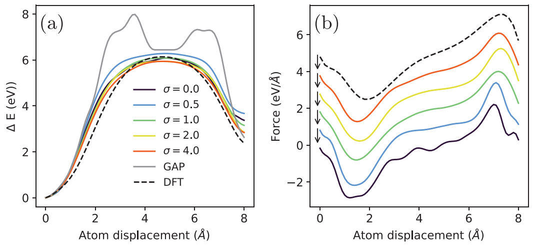

Figure 7. (a) The change in energy as an atom is move across the vacuum from the bottom (110) surface of a silicon slab to the top (110) surface. (b) The z component of the force on the atom being moved is shown. Curves for different values of σ are offset (black arrows) to aid comparison.

bear much closer resemblance to the reference DFT dimer curve. Crucially, they exhibit no false minima and are repulsive at close-approach, despite there being no dimer data in the training set.

Inspecting higher body-order terms individually is both impractical and challenging to visualize due to the increased number of parameters. As such, we instead analyze them by constructing a series of 1D cuts through the full PES that depend on all body-orders present in the potential. The first example was taken from [46] and involves computing the decohesion energy of the Si-diamond structure as it is rigidly separated to create a slab with two surfaces separated by vacuum. The derivative of the decohesion energy, which has the units of stress, is shown in figure 5 for trajectories that created (100), (110) and (111) surfaces . The reference DFT stress is notably smooth, with a single maximum and no further oscillations. In contrast, the $\sigma = 0$ ACE potentials show clear spurious oscillations which are gradually damped away as σ is increased. Smaller differences are observed in the decohesion energy itself, shown in figure S1, with variations in σ having little effect on the final surface energy; this is expected as these configurations are present in the training set. A similar test was performed whereby Si-diamond was rigidly exfoliated to individual layers of silicene. Here, clear differences in the energy are observed between the potentials with all ACE models having an additional energy barrier to exfoliation that is not present in the DFT reference; see the left panel of figure 6. The height of this barrier, and the size of associated oscillations in the derivative of the energy, decrease as σ is increased.

In the next test an atom in the top layer of a Si slab, a (110) surface, was moved in the z direction until the atom sat just below the bottom surface of the slab due to periodic boundary conditions. The energy and z-component of the force on the target atom are shown in figure 7; as before, increasing the regularity prior leading to better agreement with the DFT reference, particularly for the force. Notably, using larger σ smooths out the unphysical oscillations present from 2 Å and onwards but does not smooth out the relatively sharp feature around 0.5 Å which is present in the DFT reference. The same test was repeated for the (100) and (111) surfaces with the results shown in figure S3.

As a final test, the energy of various know Si phases, including diamond, bct5 and the simple-hexagonal (sh) polymorph, were computed as the volumes were varied isotropically. The resulting energy vs volume curves are provided in figure 8 with pressure vs volume in figure S3. Under compression, holes are present in the potentials with $\sigma = 0$, 0.5 and 1.0 Å, despite the dimer curves for $\sigma = 0.5$ Å and $\sigma = 1.0$ Å being repulsive at close-approach. Increasing σ removes this issue and is a clear sign that the regularity prior is improving the behavior (and extrapolation) of the higher body-order terms. All ACE potentials correctly reproduce the curvature near the equilibrium volume whilst those with larger regularity priors capture more of the an-harmonic behavior at larger volumes. It is worth noting that the GAP and $\sigma = 0$ potentials both exhibit additional minima in the energy-volume curves at large volumes.

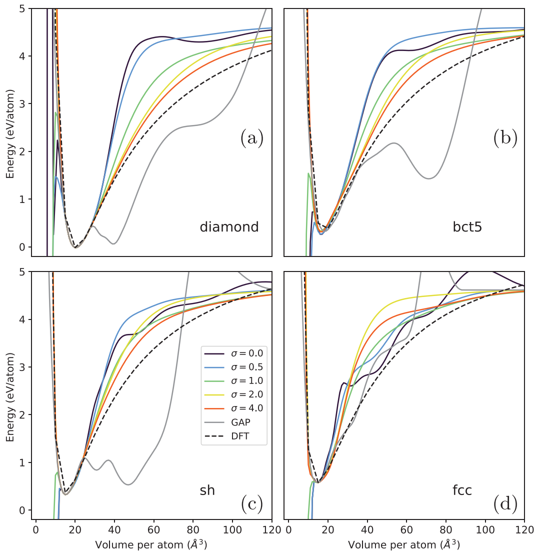

Figure 8. The energy of various silicon crystal structures is shown as the volume is varied with fractional positions of the atoms held constant. The corresponding pressure-volume curves (derivatives) are provided in figure S3.

The large energy barriers to these states means they would not be encountered during typical MD simulations initialized near equilibrium conditions. However, they could cause issues in e.g. random structure search (see figure 10) or nested sampling, which explore a much wider range of phase space.

Next, ACE potentials were fit to the full Silicon-GAP-18 dataset. The aforementioned tests were repeated with the new models, see figure S7, and similar improvements were seen when utilizing a regularity prior, suggesting that such priors still improve performance even in a much more data-rich setting. Secondly, compression simulations were performed, wherein the low-density amorphous (LDA) structure of Si from [49] was gradually compressed during a MD simulation by applying a pressure ramp from zero to 20 GPa. It is expected [50–52] that the material undergoes a sequence of phase transitions: LDA → very high density amorphous (VHDA) → polycrystalline simple-hexagonal (pc–sh). This sequence is known to be reproduced by GAP models trained on the full Silicon-GAP-18 dataset [49], and can serve as a challenging test for interatomic potentials [40, 53]. Indeed, it was previously shown that the nucleation of pc–sh was not consistently produced by moment tensor potentials (MTPs) trained directly on the Si-GAP-18 dataset [40]. Instead, consistent nucleation could only be achieved with MTP models if the training dataset was augmented with large, 1000 atom, configurations spanning the desired pressure range but labeled with a GAP model fit to the original data. This ‘student–teacher’ model relationship was exploited to transfer the desired behavior from the accurate, but slower, GAP model to the much faster MTP, enabling larger simulations to be performed. In the present work, we explore whether using a regularity prior might allow us to bypass the need for the addition of synthetic data and instead to achieve the desired sequence of phase transitions with ACE models fitted directly to the Silicon-GAP-18 dataset.

The results of such compression simulations, carried out as in [40], are shown in figure 9, where the SOAP similarity to the pressure adjusted pc–sh phase is shown as a function of pressure for ACE models fit with 11 different levels of Gaussian regularity prior. With the default prior, $\sigma = 0$, the dynamics become unstable during the transition from LDA → VHDA. Specifically, the model predicts spuriously large forces that accelerate atoms to velocities exceeding the integrator’s stability threshold, resulting in lost atoms and termination of the simulation. Prior to the crash, we observe the formation of chemically-unreasonable clusters of highly-coordinated Si, visible as blue patches in figures 9(B)–(C). These are false minima (‘holes’) in the PES accessed under the influence of compression. Increasing σ leads to a gradual increase in stability, with the $\sigma = 1.0$ Å model successfully completing the full simulation. With moderate levels of regularity prior, $\sigma = 1.0$–2.0 Å, nucleation of the pc-sh phase is successfully observed, indicating that a well chosen regularity prior has helped mitigate the need for the student-teacher relationship. At larger values of σ the dynamics becomes less stable, with no successful nucleation found to occur. This worsening performance correlates with the increasing errors shown in figure S6, and implies that a well chosen σ is one that approximately minimizes the force error on the test set.

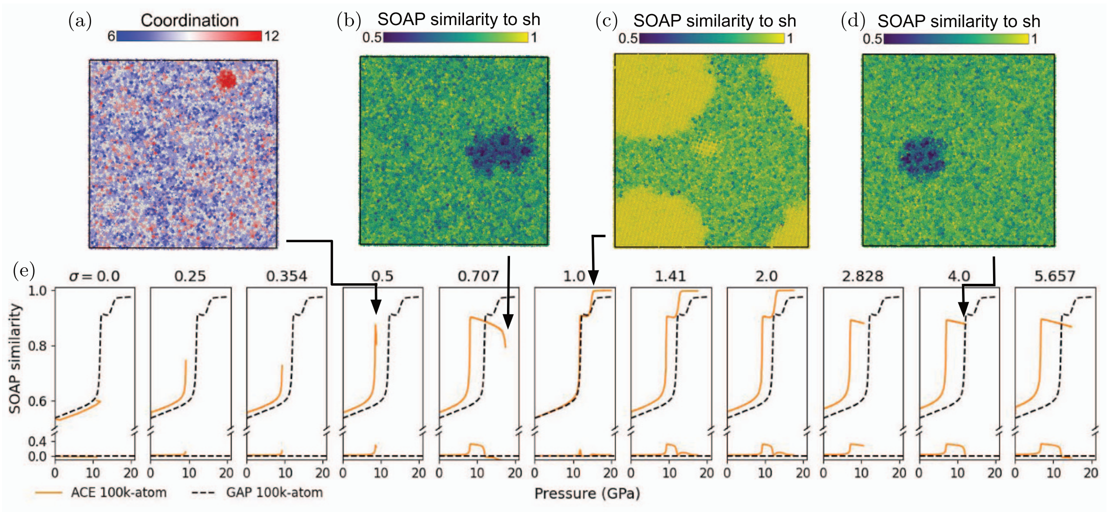

Figure 9. (e) Following [40], the average SOAP similarity to the pressure adjusted sh-Si phase is shown for 100 000 atom simulations of amorphous Si during compression from 0 to 20 GPa. The sharp increase around 10 GPa indicates the LDA → VHDA transition whilst the subsequent increase occurs on nucleation of the pc–sh phase. The difference in SOAP similarity compared to the GAP reference is shown below (note the break in the y-axis) whilst representative snapshots are highlighted above. (a) Atoms are colored according to local coordination number to highlight an erroneously highly coordinated cluster in the top right corner which occurs due to a hole in the potential. (b) and (d) No nucleation of the sh-Si phase is seen with some regions adopting an unusual local structure with hints of hexagonal symmetry. (c) Nucleation of the sh-crystal is seen with multiple different nucleation sites present. Similar structures are seen with $\sigma = 1.41$ and 2.0 Å, with simulations remaining stable as the growing grains impinge on each other, likely inducing large local stresses.

Following this, the Silicon ACE potentials were used to perform a random structure search. The search was restricted to 8 atom unit cells with 1000 putative structures generated without symmetry constraints and a further 2000 generated using 2–4 randomly chosen symmetry operations. Minimum distance constraints of 1.7 Å were enforced between neighboring atoms and the volume of the initial cell was drawn from a uniform distribution spanning 12–30 $\text{Å}^3\,\text{atom}^{-1}$. Structural relaxations were carried out with 0.01 GPa of external pressure applied in lieu of a dispersion-correction scheme and a force tolerance of 0.001 eV Å^{-1}. The distribution of relaxed structures is shown in figure 10 for searches performed using DFT (CASTEP [54]), GAP and representative ACE models. The GAP search correctly identified the diamond structure as lowest in energy but contains a much larger proportion of low density, high energy structures than the DFT reference. In contrast, the ACE models all incorrectly find at least one structure to be lower in energy than diamond, with such structures all comprised of locally dense, highly coordinated, Si clusters or nano-wires. Among the ACE potentials, using a non-zero σ significantly improves the marginal energy and volume distributions compared to the DFT reference and helps removes the low density, low energy structures. This result might have been expected from figure 8 where the spurious oscillations in the volume distribution are eliminated. However, we stress that a regularity prior will not remove all spurious minima in the predicted PES and indeed this is the case, as is seen for the search performed with $\sigma = 2.0$ Å. Perhaps the clearest benefit of applying the regularity prior is the fraction of structural relaxations that completed successfully. Structures were relaxed using the LBFGS algorithm [55] with a force tolerance of 0.05 eV Å^{-1} for DFT and 0.001 eV Å^{-1} for all potentials. DFT optimizations used a maximum of 200 steps, whilst with potentials a maximum of 20 batches of 100 steps was used with 0.001 Å of Gaussian noise added to the positions every 100 steps to aid convergence. With DFT and GAP the number of failures was <1% whilst with $\sigma = 0$ Å ∼11% failed. For $\sigma = 0.707$ Å this dropped back below 1% and remained below 2% for all larger values.

3.2. Aspirin

The effect of applying the Gaussian regularity prior was also tested on ACE potentials fit to Aspirin MD trajectories from the rMD17 dataset [56]. This data was chosen as it relates to an organic molecule and contains multiple chemical element, both in contrast with the single-element condensed-phased Silicon dataset. In SOAP-GAP potentials it is customary to specify a different value of σ for each element, with larger values used for larger elements. Here, γ_{nl} has no element dependence as the scaling of the basis functions γ_{nl} depends on σ/r_{cut}, see equation (34), and we choose to scale both σ and the cutoff r_{cut} in the same way for different element-pairs. A convenient consequence of this is that a single value of σ,

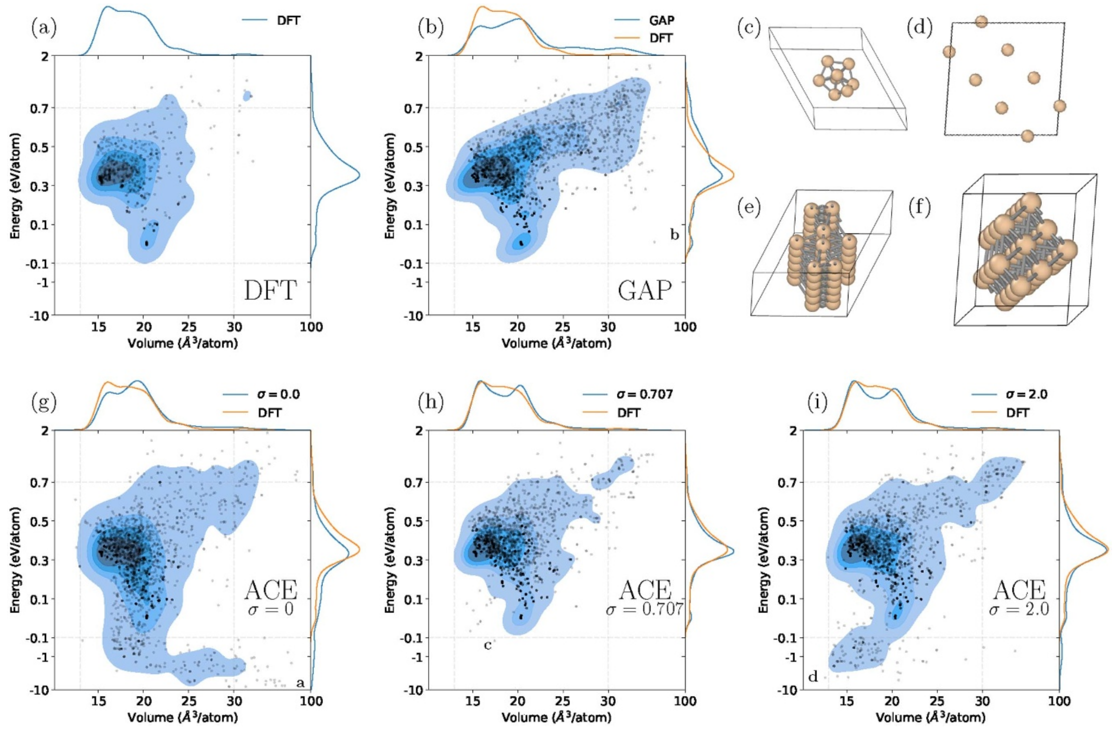

Figure 10. The distribution of energy, relative to the diamond structure, and volume of random structures relaxed with various energy methods are shown. The axes are linear within the central square bounded by dashed gray lines and logarithmic outside. The following four structures, which are all predicted to be overly stable, have also been illustrated: (c) a set of isolated clusters, (d) a 3D lattice where the closest distance between any pair of atoms is 3.7 Å as well as (e) and (f) 1D 'nano-wires' where some atoms have more than 4 immediate neighbors.

Table 1. Force RMSE on the Aspirin test set for ACE potentials fit using maximum correlation order $N = 3$ and 4 for Gaussian regularity priors with various σ. Bold number highlights the best fit.

	RMSE (meV Å^{-1})	
σ (Å)	$N = 3$	$N = 4$
0.0	50.6	50.4
0.25	48.8	41.2
0.5	48.0	**36.9**
1.0	51.9	38.2

combined with the default scaling based on tabulated covalent radii, specifies the regularity prior completely. In practice, γ_{nl} is computed using the largest cutoff present for any element-pair, so the specified value of σ corresponds to this cutoff and the effective σ for all other element-pairs is reduced accordingly.

The training and test sets used here contained 1000 configurations each and were obtained by drawing samples from the original trajectories at equi-spaced intervals, rather than using fully random splits. ACE potentials were then fit to this training set using smaller basis sets, with polynomial degree = (18, 18, 12) for the $N = 1, 2$ and 3 features respectively, and a larger basis set with $N = 4$ and polynomial degree = (18, 18, 18, 12). As before, an Agnesi transformation was applied to the radial coordinate [44] and the fits was performed using standard Tikhonov regularization with varying strengths. The best achievable force errors on the test set are tabulated in table 1 and are shown as a function of regularization strength for the $N = 4$ potentials in figure 11. With the smaller basis ($N \leqslant 3$), a very modest 4% reduction in the force RSME on the test set is observed when using $\sigma = 0.5$ Å compared to no regularity prior. However, on increasing the correlation order (and number of basis functions) we see a more significant difference, with $\sigma = 0.5$ Å reducing the test error by 27%. It is also notable that without the regularity prior, the extra basis functions give no reduction in the test error.

As with Silicon, we then looked at the effect of the regularity priors on various 1D cuts through the PES. The dimer curves are shown in figure S8 where it is clear that applying a modest regularity prior

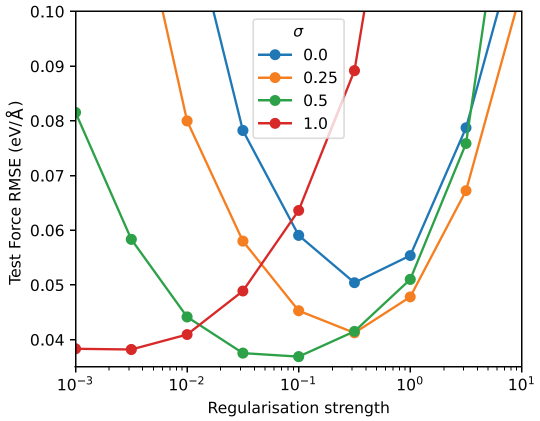

Figure 11. Force RMSE on the Aspirin test set as a function of regularization strength for the ACE potentials fit using $N = 4$ and a Gaussian regularity prior with various values for σ.

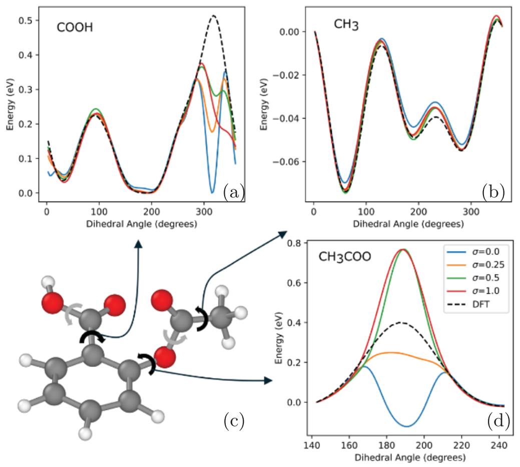

Figure 12. (c) A single aspirin molecule is shown with three bonds corresponding to the named functional groups marked in black. Clockwise from the top left the energy of rigidly rotating the COOH (a), CH_3 (b) and CH_3COO (d) groups are shown as a function of angle.

helps dampen the spurious high-frequency oscillations seen with $\sigma = 0$. Increasing σ further down-weights the higher-body order terms relative to the 2-body, and so forces the low polynomial degree 2-body terms to account for ever more of the binding energy. This leads to qualitatively improved behavior for $\sigma \geqslant 0.25$Å, with at most a single minimum seen in each dimer curve. However, due to the lack of dimer data in the training set, the depth of this minimum is not quantitatively correct. Next, we looked at the effect of the regularity prior on higher correlation order terms in the potentials by computing the energy as the COOH, CH_3COO and CH_3 groups in an aspirin molecule were rigidly rotated. The energy profiles are shown in figure 12, with the black dashed lines showing the true reference energy computed using ORCA [57] with the same settings as used for the original dataset. All potentials show good agreement with the reference for rotations of the methyl group, which is consistent with the low energy barriers for this rotational mode leading to it being well sampled in the training data. In contrast, there are much larger energy barriers for rotating the other two groups with significant differences between the predicted energies and the reference values. In particular, the $\sigma = 0$ potential has a deep minimum where there should be a maximum. Increasing σ gradually removes the minimum seen near 320° when rotating the COOH whilst for the CH_3COO it leads to an overestimation of the energy maximum at 190°.

The chronic underestimation of energy by the $\sigma = 0$ potentials seen in figure 12 suggests that using larger values of σ may lead to more stable MD as there appear to be fewer ‘holes’ in the PES. To test this we took 100 configurations in the test set and attempted to perform 1 ns of MD at both 300 and 500 K. The dynamics was performed using the `ase` python package [58] with a 0.5 fs time step

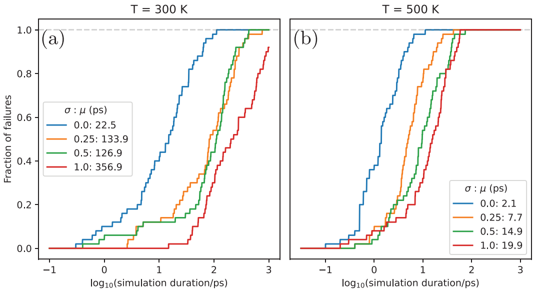

Figure 13. Duration of MD simulations of a single Aspirin molecule before encountering a 'hole' i.e. catastrophic failure. The simulations were performed at (a) 300 and (b) 500 K and were terminated after 1 ns if failure had not occurred.

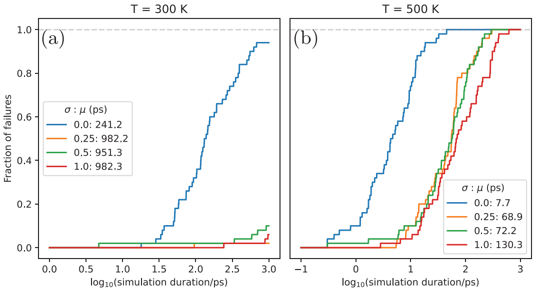

Figure 14. Duration of MD simulations of a single Aspirin molecule before encountering a 'hole' i.e. catastrophic failure, using the ACE potentials fit to the additional data generated via random dihedral rotations. The simulations were performed at (a) 300 and (b) 500 K and were terminated after 1 ns if failure had not occurred.

and a Langevin thermostat [59] with a friction coefficient of 3 ps^{-1}. The simulations were terminated whenever a covalent bond present in the original molecule broke or if a new bond was formed. Thresholds for bonds breaking/forming were determined by analyzing the minimum/maximum distances present between bonded elements in the training set and then adding/subtracting an additional 0.5 Å of tolerance. In practice the details of these criteria were found to be unimportant as any change in the number of bonds was typically accompanied by a dramatic failure of the potential where the molecule exploded. The distribution of the duration of these MD runs is shown in figure 13 and shows that using the regularity prior led to $\sim 10\times$ longer average simulation times. To verify that this increase was not caused exclusively by over-stabilization of certain configurations the joint distribution of dihedral angles was inspected for trajectories computed with each potential. The distributions are visualized in figure S12 and show no evidence that the trajectories with larger σ explored less of the configuration space. Furthermore, the minimum energy paths between minima were investigated using NEBs [27] and found to be in good agreement with DFT for all potentials, see figure S11. As such, we conclude that applying a regularity prior provides a significant improvement to MD stability.

We stress that the aim of this test was not to obtain a robust potential, but instead study the effects of the regularity priors on MD stability. As such, by design we have chosen to use a relatively small amount of training data so that the MD is not stable and the effect of the regularity priors can be easily seen. To demonstrate that the MD stability can easily be improved by adding more data, we generated 10 000 configurations by applying random rotations to all 5 dihedral angles marked in figure 12. These were then filtered down to ∼6500 configurations using minimum distance constraints before farthest-point-sampling was used to select 284 new configurations to be added to the training set. The entire process was repeated and used to generate an additional 144 configurations which were added to the test set. By generating the data in this way, as opposed to sampling from the MD trajectories generated using the ACE potentials, the new configurations are completely unbiased with respect to any of the potentials, see figure S9 for the energy and force distribution. This enlarged dataset was then refit using the same procedure as described previously. The force errors on the new test set are shown in figure S10 and, as before, applying a regularity prior leads to a significant reduction (40%) in test force RMSE. Repeating the MD simulations we see a clear improvement in MD stability, with the new durations shown in figure 14. Finally, we also note that the additional data leads to greatly improved dihedral energy profiles compared to DFT, as seen for all potentials in figure S13.

4. Discussion

In this work we analyzed the concept of regularity priors for linear ACE models. To summarize in non-technical terms, regularity priors can be understood as follows: ACE basis functions have an associated radial frequency (indexed by n) and angular frequency (indexed by l), with larger indices corresponding to higher frequencies. Large weights on the high frequency features result in non-smooth potentials. Generalized Tikhonov regularization enforces smaller weights on high frequency terms and thus results in smooth potentials that are significantly more robust in realistic simulation scenarios.

Regularity priors are interpretable from a Bayesian perspective and are implemented as a modified regularization term. Equivalently, the priors can be interpreted as a rescaling of the one-particle basis functions. This allows standard regression methods to be applied, and ensures that no additional complexity or computational cost arises during fitting.

We tested three different regularity priors: two which either assume that the predicted PES has p continuous derivatives or is analytical, as well as a third with an over-regularizing Gaussian form. By considering the effect of the Gaussian prior on the fictitious neighbor density it was shown be equivalent to the Gaussian broadening used in SOAP-GAP models.

Numerical tests indicated that the exact form of the prior made little difference but that using a regularity prior led to significant gains when compared to the standard uniform prior. These tests were performed on the diverse silicon-GAP-18 dataset, a small subset of the same data, and configurations of individual Aspirin molecules taken from MD trajectories. For all datasets using regularity priors improved the accuracy on the unseen test set, with an almost 30%–40% reduction in force RMSE. Additional testing also revealed improvements in smoothness along various 1D cuts through the PES, more consistent repulsion at close approach, the removal of many false minima encountered during RSS and enhanced stability during MD simulations.

We conclude by noting that regularity priors are not a substitute for data, and that using one clearly does not guarantee a potential will behave well in all circumstances. However, they offer consistently improved performance at essentially no additional computational cost or complexity. An interesting avenue for future work is to assess how similar ideas can be incorporated into more complex, non-linear architectures such as a neural-network-based potentials and message-passing architectures such as MACE [33], NequIP [60], PaiNN [61] etc

Acknowledgments

ABP acknowledges support from the CASTEP-USER Grant funded by UK Research and Innovation (EP/W030438/1). We used computational resources of the UK HPC service ARCHER2 via the UKCP consortium and funded by EPSRC Grant nos. EP/P022596/1 and EPSRC Grant nos. EP/X035891/1. CO acknowledges the support of the Natural Sciences and Engineering Research Council of Canada (NSERC). J D M acknowledges funding from the EPSRC Centre for Doctoral Training in Inorganic Chemistry for Future Manufacturing (OxICFM), Grant no. EP/S023828/1.

Data availability statement

Both the Silicon-GAP-18 dataset from [46] and the rMD17 dataset from [56] are publicly available. The additional Aspirin configurations generated by sampling random dihedral rotations, and evaluated using the same DFT settings as rMD17, are available on zenodo [64]. The `ACEPotentials.jl` code [44] is publicly available and contains example scripts and tutorials for fitting models.

The data that support the findings of this study will be openly available following an embargo at the following URL/DOI: https://doi.org/10.5281/zenodo.18834186 [64].

Supplementary Information available at: https://doi.org/10.1088/2632-2153/ae55fa/data1.

Conflicts of interest

GC, CO and JPD are partners in Symmetric Group LLP that licenses force fields commercially. GC also has equity interest in Ångström AI.

Author contributions

James P Darby 0000-0002-3365-599X
Conceptualization (equal), Data curation (equal), Formal analysis (equal), Methodology (equal), Project administration (equal), Software (equal), Validation (equal), Visualization (equal), Writing – original draft (equal), Writing – review & editing (equal)

Joe D Morrow 0000-0002-3441-8646
Data curation (equal), Formal analysis (equal), Validation (equal), Visualization (equal), Writing – original draft (equal), Writing – review & editing (equal)

Albert P Bartók 0000-0002-4347-8819
Conceptualization (equal), Formal analysis (equal), Methodology (equal), Writing – original draft (equal), Writing – review & editing (equal)

Volker L Deringer 0000-0001-6873-0278
Conceptualization (equal), Data curation (equal), Formal analysis (equal), Funding acquisition (equal), Resources (equal), Supervision (equal), Writing – review & editing (equal)

Gábor Csányi
Conceptualization (equal), Formal analysis (equal), Funding acquisition (equal), Resources (equal), Supervision (equal), Writing – review & editing (equal)

Christoph Ortner 0000-0003-1498-8120
Conceptualization (equal), Formal analysis (equal), Funding acquisition (equal), Methodology (equal), Project administration (equal), Software (equal), Supervision (equal), Writing – original draft (equal), Writing – review & editing (equal)

Appendix A. Priors for the symmetrized basis

In our simplified ACE formulation we wrote

$$E_i = \sum_{\boldsymbol{nlm}} c_{\boldsymbol{nlm}} \prod_t A_{i,n_t l_t m_t}, \tag{16}$$

with summation over all $\boldsymbol{nlm}$ tuples, not only ordered ones. This simplified the presentation of regularity priors. In practice, ACE is implemented by first noticing that only summation over ordered tuples is required,

$$\begin{aligned} E_i &= \sum_{\boldsymbol{nlm}\text{ ordered}} k_{\boldsymbol{nlm}} c_{\boldsymbol{nlm}} \prod_t A_{i,n_t l_t m_t}, \\ &= \sum_{\boldsymbol{nlm}\text{ ordered}} k_{\boldsymbol{nlm}} c_{\boldsymbol{nlm}} \boldsymbol{A}_{i,\boldsymbol{nlm}} \\ &= (\boldsymbol{k} \odot \boldsymbol{c}) \cdot \boldsymbol{A}_i, \end{aligned} \tag{17}$$

where $k_{\boldsymbol{nlm}}$ denotes the number of repeated basis functions (a multi-combination and explicitly computable). Next, one symmetrizes the basis $\boldsymbol{B}_i = \mathcal{C}\boldsymbol{A}_i$. Because $\boldsymbol{c}$, and hence $\boldsymbol{k} \odot \boldsymbol{c}$, result in an invariant model it follows that $\boldsymbol{k} \odot \boldsymbol{c} = \mathcal{C}^{\mathrm{T}}\boldsymbol{\theta}$ for some parameter vector $\boldsymbol{\theta}$ and hence

$$E_i = (\boldsymbol{k} \odot \boldsymbol{c}) \cdot \boldsymbol{A}_i = \boldsymbol{\theta} \cdot \boldsymbol{B}_i. \tag{18}$$

We call $\boldsymbol{B}_i$ the *ACE basis* or the *ACE features*.

Under the same transformation, the Tikhonov regularizer becomes,

$$\|\Gamma \boldsymbol{c}\|^2 = \left\|\Gamma_{\mathrm{ord}} (\boldsymbol{k} \odot \boldsymbol{c})\right\|^2 = \left\|\Gamma_{\mathrm{ord}} \mathcal{C}^{\mathrm{T}} \boldsymbol{\theta}\right\|^2, \tag{19}$$

where the diagonal entries of Γ_{ord} are simply those of Γ but restricted to ordered tuples. Equivalently, the prior $\boldsymbol{c} \sim \mathcal{N}(\boldsymbol{0}, \Gamma^{-2})$ transforms to a prior on $\boldsymbol{\theta}$,

$$\boldsymbol{\theta} \sim \mathcal{N}\left(\boldsymbol{0}, \left(\mathcal{C}\Gamma^2\mathcal{C}^{\mathrm{T}}\right)^{-1}\right). \tag{20}$$

Because of the block-structure of $\mathcal{C}$ one can in fact rewrite $\Gamma_{\mathrm{ord}}\mathcal{C}^{\mathrm{T}}\boldsymbol{\theta} = \bar{\Gamma}\boldsymbol{\theta}$, where $\bar{\Gamma}$ is diagonal.

Appendix B. SOAP-GAP connection

Both ACE and SOAP features can be understood as symmetrized tensor-products (N products for $N+1$ body-order features) of a neighbor density. In SOAP each neighbor atom is represented using a Gaussian of width σ,

$$c_{i,znlm} = \sum_{j\in\mathcal{N}(i)} \int g_\alpha\left(\boldsymbol{r}-\boldsymbol{r}_{ij}\right)\phi_{znlm}\left(\boldsymbol{r}\right)\mathrm{d}\boldsymbol{r}, \tag{21}$$

where $g_\alpha(\boldsymbol{r}) = \exp(-\alpha|\boldsymbol{r}|^2)$, whilst the ACE potentials effectively use a delta function.

$$A_{i,znlm} = \sum_{j\in\mathcal{N}(i)} \int \delta\left(\boldsymbol{r}-\boldsymbol{r}_{ij}\right)\phi_{znlm}\left(\boldsymbol{r}\right)\mathrm{d}\boldsymbol{r}. \tag{22}$$

It turns out, as we show next, that under the specific choice of the one-particle basis proposed in [41], applying the Gaussian prior (10) to the ACE features transforms them into SOAP features. Specifically, the self-interacting ACE features computed using theLE of [41] as the radial basis are equivalent to SOAP features computed without the central atom contributing to the neighbor density (i.e. `central_weight = 0` in the GAP code [39]).

To show this we start by noting that, due to the sifting property of $\delta(\boldsymbol{r}-\boldsymbol{r}_j)$, the $A_{i,znlm}$ in self-interacting ACE can viewed as the density expansion coefficients of the delta function neighbor density,

$$\begin{aligned} A_{i,znlm} &= \sum_{j\in\mathcal{N}(i)} \int \boldsymbol{d}\boldsymbol{r}\delta\left(\boldsymbol{r}_j-\boldsymbol{r}\right)R_{nl}\left(r\right)Y_l^m\left(\hat{\boldsymbol{r}}\right)\delta_{zZ_j} && (23)\\ &= \sum_{j\in\mathcal{N}(i)} \phi_{nlm}\left(\boldsymbol{r}_j\right) = \sum_{j\in\mathcal{N}(i)} R_{nl}\left(r\right)Y_l^m\left(\hat{\boldsymbol{r}}\right)\delta_{zZ_j}. && (24)\end{aligned}$$

Next we compute the density expansion coefficient $c_{i,nlm}$ for a single neighbor atom represented as a normalized Gaussian of width σ centered at $\boldsymbol{r}_i$ using the LE,

$$c_{i,nlm} = \int \phi_{nlm}\exp\left(-\alpha|\mathbf{r}-\mathbf{r}_i|^2\right)r^2\mathrm{d}r\mathrm{d}\hat{r} \tag{25}$$

The LE are defined as

$$R_{nl}\left(r\right) = N_{nl}j_l\left(z_{nl}a^{-1}r\right). \tag{26}$$

where $j_l(x)$ is a spherical Bessel function of the first kind, z_{nl} is the nth zero of $j_l(x)$, $a = r_{\mathrm{cut}}$ is the cutoff radius and

$$N_{nl} = \sqrt{\frac{a^3}{2}j_{l+1}\left(z_{nl}\right)} \tag{27}$$

is a normalization constant such that

$$\int_0^a drr^2R_{nl}\left(r\right)R_{n'l}\left(r\right) = \delta_{nn'}. \tag{28}$$

Using the same approach as in the original SOAP paper [36] we can then evaluate $c_{i,nlm}$ as

$$\begin{aligned} c_{i,nlm} &= \int \phi_{nlm}\exp\left(-\alpha|\mathbf{r}-\mathbf{r}_i|^2\right)r^2\mathrm{d}r\mathrm{d}\hat{r}\\ &= \int N_{nl}j_l\left(z_{nl}a^{-1}r\right)Y_{lm}\left(\hat{r}\right)\cdot 4\pi\exp\left(-\alpha\left(r^2+r_i^2\right)\right)\\ &\quad\times\sum_{l'm'} l_{l'}\left(2\alpha rr_i\right)Y_{l'm'}\left(\hat{r}\right)Y^*_{lm'}\left(\hat{r}_i\right)r^2\mathrm{d}r\mathrm{d}\hat{r}\\ &= 4\pi N_{nl}\exp\left(-\alpha r_i^2\right)Y_{l'm'}\left(\hat{r}_i\right)\sum_{l'm'}\int Y_{lm}\left(\hat{r}\right)Y^*_{l'm'}\left(\hat{r}\right)\mathrm{d}\hat{r}\\ &\quad\times\int j_l\left(z_{nl}a^{-1}r\right)\exp\left(-\alpha r^2\right)l_{l'}\left(2\alpha rr_i\right)r^2\mathrm{d}r \end{aligned}$$

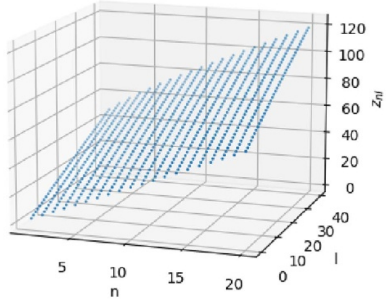

Figure 15. z_{nl} as function of n and l.

where $l_l(r)$ is a modified spherical Bessel function of the first kind that satisfies $l_l(x) = i^{-l} j_l(ix)$. Using the normalization of the spherical harmonics, $\int Y_{lm}(\hat{r}) Y^*_{l'm'}(\hat{r}) \mathrm{d}\hat{r} = \delta_{ll'} \delta_{mm'}$ this simplifies to

$$c^i_{nlm} = 4\pi N_{nl} \exp\left(-\alpha r_i^2\right) Y_{lm}\left(\hat{r}_i\right) \cdot I \tag{29}$$

where

$$I = \int_0^\infty j_l\left(z_{nl} a^{-1} r\right) \exp\left(-\alpha r^2\right) l_l\left(2\alpha r r_i\right) r^2 \mathrm{d}r. \tag{30}$$

Note that the upper integration limit is ∞, not $r_{\mathrm{cut}} = a$, as the cutoff function handles atoms entering and leaving the environment. This is a modified version of Weber's second exponential integral, see 13.31 of [62], which is used in equation (12) of [63] with complex u and v. Using these results we evaluate I as

$$I = \frac{1}{4}\sqrt{\frac{\pi}{\alpha^3}} j_l\left(\frac{z_{nl} r_i}{a}\right) \cdot \exp\left(-\frac{z_{nl}^2}{4\alpha a^2}\right) \cdot \exp\left(\alpha r_i^2\right). \tag{31}$$

Combining everything together we arrive at

$$c^i_{nlm} = N_{nl} \cdot j_l\left(\frac{z_{nl}}{a} r_i\right) Y_{lm}\left(\hat{r}_i\right) \cdot \exp\left(-\frac{\sigma^2}{2} E_{nl}\right) \tag{32}$$

$$= \phi_{nlm}\left(\boldsymbol{r}_i\right) \cdot \exp\left(-\frac{\sigma^2}{2} E_{nl}\right) \tag{33}$$

where $E_{nl} = z_{nl}^2 / a^2$. There is no analytic expression for z_{nl} but it can be crudely approximated as $z_{nl} \approx an + bl$ as shown in figure 15.

Under this approximation we see that broadening of the ACE features can be accomplished with a Gaussian prior of the form,

$$\gamma\left(\boldsymbol{nl}\right) = \prod_i \exp\left(\frac{\sigma^2}{a^2} \cdot \left[5n_i^2 + 4n_i l_i + l_i^2\right]\right) \tag{34}$$

where, given that we care about the scaling, the coefficients have been rounded to integer values for simplicity.

References

[1] Hohenberg P Kohn W 1964 Inhomogeneous electron gas *Phys. Rev.* **136** B864
[2] Kohn W Sham L J 1965 Self-consistent equations including exchange and correlation effects *Phys. Rev.* **140** A1133
[3] Müser M H, Sukhomlinov S V and Pastewka L 2023 Interatomic potentials: achievements and challenges *Adv. Phys. X* **8** 2093129
[4] Zhang L, Csányi G, Van Der Giessen E and Maresca F 2023 Atomistic fracture in BCC iron revealed by active learning of Gaussian approximation potential *npj Comput. Mater.* **9** 217
[5] Erhard L C, Rohrer J, Albe K and Deringer V L 2022 A machine-learned interatomic potential for silica and its relation to empirical models *npj Comput. Mater.* **8** 90
[6] Marchant G A, Caro M A, Karasulu B and Pártay L B 2023 Exploring the configuration space of elemental carbon with empirical and machine learned interatomic potentials *npj Comput. Mater.* **9** 131
[7] Mattsson T R, Lane J M D, Cochrane K R, Desjarlais M P, Thompson A P, Pierce F and Grest G S 2010 First-principles and classical molecular dynamics simulation of shocked polymers *Phys. Rev. B* **81** 054103

[8] Heijmans K, Tranca I C, Smeulders D M, Vlugt T J and Gaastra-Nedea S V 2020 Gibbs ensemble monte carlo for reactive force fields to determine the vapor–liquid equilibrium of CO_2 and H_2O *J. Chem. Theory Comput.* **17** 322
[9] Behler J Parrinello M 2007 Generalized neural-network representation of high-dimensional potential-energy surfaces *Phys. Rev. Lett.* **98** 146401
[10] Vita J A Schwalbe-Koda D 2023 Data efficiency and extrapolation trends in neural network interatomic potentials *Mach. Learn.: Sci. Technol.* **4** 035031
[11] Benoit M, Amodeo J, Combettes S, Khaled I, Roux A and Lam J 2020 Measuring transferability issues in machine-learning force fields: the example of gold–iron interactions with linearized potentials *Mach. Learn.: Sci. Technol.* **2** 025003
[12] Nandi A, Qu C and Bowman J M 2019 Using gradients in permutationally invariant polynomial potential fitting: a demonstration for CH_4 using as few as 100 configurations *J. Chem. Theory Comput.* **15** 2826
[13] van Der Oord C, Dusson G, Csányi G and Ortner C 2020 Regularised atomic body-ordered permutation-invariant polynomials for the construction of interatomic potentials *Mach. Learn.: Sci. Technol.* **1** 015004
[14] van der Oord C, Sachs M, Kovács D P, Ortner C and Csányi G 2023 Hyperactive learning for data-driven interatomic potentials *npj Comput. Mater.* **9** 168
[15] Sivaraman G, Krishnamoorthy A N, Baur M, Holm C, Stan M, Csányi G, Benmore C and Vázquez-Mayagoitia A 2020 Machine-learned interatomic potentials by active learning: amorphous and liquid hafnium dioxide *npj Comput. Mater.* **6** 104
[16] Batatia I *et al* 2025 A foundation model for atomistic materials chemistry *J. Chem. Phys.* **163** 184110
[17] Yang H *et al* 2024 MatterSim: a deep learning atomistic model across elements, temperatures and pressures (arXiv:2405.04967)
[18] Neumann M, Gin J, Rhodes B, Bennett S, Li Z, Choubisa H, Hussey A and Godwin J 2024 ORB: a fast, scalable neural network potential (arXiv:2410.22570)
[19] Park Y, Kim J, Hwang S and Han S 2024 Scalable parallel algorithm for graph neural network interatomic potentials in molecular dynamics simulations *J. Chem. Theory Comput.* **20** 4857
[20] Fu X, Wood B M, Barroso-Luque L, Levine D S, Gao M, Dzamba M and Zitnick C L 2025 Learning smooth and expressive interatomic potentials for physical property prediction *Proc. of the 42nd Int. Conf. on Machine Learning* (PMLR) pp 17875–93 (available at: https://proceedings.mlr.press/v267/fu25h.html)
[21] Barroso-Luque L, Shuaibi M, Fu X, Wood B M, Dzamba M, Gao M, Rizvi A, Zitnick C L and Ulissi Z W 2024 Open materials 2024 (OMAT24) inorganic materials dataset and models, (arXiv:2410.12771)
[22] Merchant A, Batzner S, Schoenholz S S, Aykol M, Cheon G and Cubuk E D 2023 Scaling deep learning for materials discovery *Nature* **624** 80
[23] Rhodes B, Vandenhaute S, Šimkus V, Gin J, Godwin J, Duignan T and Neumann M 2025 Orb-v3: atomistic simulation at scale (arXiv:2504.06231)
[24] Bochkarev A, Lysogorskiy Y and Drautz R 2024 Graph atomic cluster expansion for semilocal interactions beyond equivariant message passing *Phys. Rev. X* **14** 021036
[25] Riebesell J, Goodall R E, Benner P, Chiang Y, Deng B, Ceder G, Asta M, Lee A A, Jain A and Persson K A 2025 A framework to evaluate machine learning crystal stability predictions *Nat. Mach. Intell.* **7** 836
[26] Chiang Y *et al* 2025 MLIP arena: advancing fairness and transparency in machine learning interatomic potentials through an open and accessible benchmark platform *The Thirty-ninth Annual Conf. on Neural Information Processing Systems Datasets and Benchmarks Track* (available at: https://openreview.net/forum?id=SAT0KPA5UO)
[27] Henkelman G, Uberuaga B P and Jónsson H 2000 A climbing image nudged elastic band method for finding saddle points and minimum energy paths *J. Chem. Phys.* **113** 9901
[28] Deringer V L, Pickard C J and Csányi G 2018 Data-driven learning of total and local energies in elemental boron *Phys. Rev. Lett.* **120** 156001
[29] Kresse G Furthmüller J 1996 Efficiency of ab-initio total energy calculations for metals and semiconductors using a plane-wave basis set *Comput. Mater. Sci.* **6** 15
[30] Ziegler J F and Biersack J P 1985 The stopping and range of ions in matter *Treatise on Heavy-Ion Science* ed D A Bromley (Springer) pp 93–129
[31] Nordlund K, Lehtola S and Hobler G 2025 Repulsive interatomic potentials calculated at three levels of theory *Phys. Rev. A* **111** 032818
[32] Zhou Y, Thomas du Toit D F, Elliott S R, Zhang W and Deringer V L 2025 Full-cycle device-scale simulations of memory materials with a tailored atomic-cluster-expansion potential *Nat. Commun.* **16** 1
[33] Batatia I, Kovacs D P, Simm G, Ortner C and Csányi G 2022 MACE: higher order equivariant message passing neural networks for fast and accurate force fields *Advances in Neural Information Processing Systems* (Curran Associates, Inc) vol 35 pp 11423–36 (available at: https://proceedings.neurips.cc/paper_files/paper/2022/file/4a36c3c51af11ed9f34615b81edb5bbc-Paper-Conference.pdf)
[34] Bartók A P, Payne M C, Kondor R and Csányi G 2010 Gaussian approximation potentials: the accuracy of quantum mechanics, without the electrons *Phys. Rev. Lett.* **104** 136403
[35] Deringer V L, Bartók A P, Bernstein N, Wilkins D M, Ceriotti M and Csányi G 2021 Gaussian process regression for materials and molecules *Chem. Rev.* **121** 10073
[36] Bartók A P, Kondor R and Csányi G 2013 On representing chemical environments *Phys. Rev. B* **87** 184115
[37] Drautz R 2019 Atomic cluster expansion for accurate and transferable interatomic potentials *Phys. Rev. B* **99** 014104
[38] Dusson G, Bachmayr M, Csányi G, Drautz R, Etter S, van Der Oord C and Ortner C 2022 Atomic cluster expansion: completeness, efficiency and stability *J. Comput. Phys.* **454** 110946
[39] Klawohn S, Darby J P, Kermode J R, Csányi G, Caro M A and Bartók A P 2023 Gaussian approximation potentials: theory, software implementation and application examples *J. Chem. Phys.* **159** 174108
[40] Morrow J D Deringer V L 2022 Indirect learning and physically guided validation of interatomic potential models *J. Chem. Phys.* **157** 104105
[41] Bigi F, Huguenin-Dumittan K K, Ceriotti M and Manolopoulos D E 2022 A smooth basis for atomistic machine learning *J. Chem. Phys.* **157** 234101
[42] Evans L C 2010 *Partial Differential Equations* 2nd edn (*Graduate Studies in Mathematics*) vol 19 (American Mathematical Society)
[43] Methfessel M Paxton A T 1989 High-precision sampling for Brillouin-zone integration in metals *Phys. Rev. B* **40** 3616
[44] Witt W C *et al* 2023 Acepotentials. jl: a julia implementation of the atomic cluster expansion *J. Chem. Phys.* **159** 164101
[45] Ho C H, Gutleb T S and Ortner C 2024 Atomic cluster expansion without self-interaction *J. Comput. Phys.* **515** 113271

[46] Bartók A P, Kermode J, Bernstein N and Csányi G 2018 Machine learning a general-purpose interatomic potential for silicon *Phys. Rev. X* **8** 041048

[47] Pickard C J Needs R 2011 Ab initio random structure searching *J. Phys.: Condens. Matter* **23** 053201

[48] Fu X, Wu Z, Wang W, Xie T, Keten S, Gomez-Bombarelli R and Jaakkola T 2022 Forces are not enough: benchmark and critical evaluation for machine learning force fields with molecular simulations (arXiv:2210.07237)

[49] Deringer V L, Bernstein N, Csányi G, Ben Mahmoud C, Ceriotti M, Wilson M, Drabold D A and Elliott S R 2021 Origins of structural and electronic transitions in disordered silicon *Nature* **589** 59

[50] Deb S K, Wilding M, Somayazulu M and McMillan P F 2001 Pressure-induced amorphization and an amorphous–amorphous transition in densified porous silicon *Nature* **414** 528

[51] McMillan P F, Wilson M, Daisenberger D and Machon D 2005 A density-driven phase transition between semiconducting and metallic polyamorphs of silicon *Nat. Mater.* **4** 680

[52] Pandey K, Garg N, Shanavas K, Sharma S M and Sikka S 2011 Pressure induced crystallization in amorphous silicon *J. Appl. Phys.* **109** 113511

[53] Morrow J D, Gardner J L and Deringer V L 2023 How to validate machine-learned interatomic potentials *J. Chem. Phys.* **158** 121501

[54] Clark S J, Segall M D, Pickard C J, Hasnip P J, Probert M I, Refson K and Payne M C 2005 First principles methods using castep *Z. Kristallogr. Cryst. Mater.* **220** 567

[55] Liu D C Nocedal J 1989 On the limited memory BFGS method for large scale optimization *Math. Program.* **45** 503

[56] Christensen A S Von Lilienfeld O A 2020 On the role of gradients for machine learning of molecular energies and forces *Mach. Learn.: Sci. Technol.* **1** 045018

[57] Neese F 2012 The ORCA program system *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2** 73

[58] Larsen A H *et al* 2017 The atomic simulation environment-a python library for working with atoms *J. Phys.: Condens. Matter* **29** 273002

[59] Vanden-Eijnden E Ciccotti G 2006 Second-order integrators for Langevin equations with holonomic constraints *Chem. Phys. Lett.* **429** 310

[60] Batzner S, Musaelian A, Sun L, Geiger M, Mailoa J P, Kornbluth M, Molinari N, Smidt T E and Kozinsky B 2022 E(3)-equivariant graph neural networks for data-efficient and accurate interatomic potentials *Nat. Commun.* **13** 2453

[61] Schütt K, Unke O and Gastegger M 2021 Equivariant message passing for the prediction of tensorial properties and molecular spectra *Int. Conf. on Machine Learning* (PMLR) pp 9377–88

[62] Watson G N 1922 *A Treatise on the Theory of Bessel Functions* vol 3 (The University Press)

[63] Kaufmann K Baumeister W 1989 Single-centre expansion of Gaussian basis functions and the angular decomposition of their overlap integrals *J. Phys. B: At. Mol. Opt. Phys.* **22** 1

[64] Available at: https://doi.org/10.5281/zenodo.18834186