npj | computational materials

Article

Published in partnership with the Shanghai Institute of Ceramics of the Chinese Academy of Sciences

https://doi.org/10.1038/s41524-024-01225-6

Incorporating long-range electrostatics in neural network potentials via variational charge equilibration from shortsighted ingredients

Check for updates

Yusuf Shaidu [1,2,3] ✉, Franco Pellegrini[3], Emine Küçükbenli[3,4] ✉, Ruggero Lot[3] & Stefano de Gironcoli [3,5] ✉

We present a new approach to construct machine-learned interatomic potentials including long-range electrostatic interactions based on a charge equilibration scheme. This new approach can accurately describe the potential energy surface of systems with ionic and covalent interactions as well as systems with multiple charge states. Moreover, it can either be regressed against known atomic charge decompositions or trained without charge targets, without compromising the accuracy of energy and forces. We benchmark our approach against other state-of-the-art models and prove it to have equivalent performances on a set of simple reference systems while being less computationally expensive. Finally, we demonstrate the accuracy of our approach on complex systems: solid and liquid state sodium chloride. We attain accuracy in energy and forces better than the model based on local descriptors and show that our electrostatic approach can capture the density functional theory tail of the potential energy surface of the isolated Na-Cl dimer, which the local descriptor-based model fails to describe.

Most modern machine-learned interatomic potentials exploit the idea of *nearsightedness* of interatomic interactions, i.e., they assume that the contribution of an atom to the total energy depends mostly on its local chemical environment, while the influence of atoms that are far away is negligibly small. Following this idea, Behler and Parrinello[1] proposed in 2007 the decomposition of the total energy of an extended system into atomic contributions, each computed from a descriptor of the local chemical environment. In that work, the atomic energy was modeled using neural networks (NN), but the same idea can also be combined with kernel regression to construct, for example, Gaussian Approximation Potentials [2]. Numerous high-accuracy machine-learned interatomic potentials have been developed based on local descriptors and have been shown to be transferable enough to successfully capture the energetics of systems beyond the conditions they were trained on refs. 2–10, including ionic systems such as sodium chloride in solid and liquid phases[11].

Despite these successes, it is known that these short-range models are limited by design, and as the non-local interactions such as electrostatics or van der Waals (vdW) effects become more significant in a system, the applicability of a short-range model becomes questionable. For van der Waals effects, one strategy is to include them via ad hoc corrections[7], namely, first subtracting this contribution from the reference total energy, then using the remainder as the target of the machine learning training of a short-range potential, and then adding back the long-range contribution to predict the total energy. This approach is suitable if the reference energy is already obtained with an ad hoc description of the vdW effects, as in the Grimme-D2[12] approach. This range separation scheme has been proposed to account for the long-range part of the electrostatic interactions in a similar way: in refs. 13,14 a two-part neural network interatomic potential was constructed for the water dimer. One neural network was trained to predict the local atomic charge on each atom, and its parameters were optimized to reproduce ab initio charges obtained from the Hirshfeld decomposition

[1]Department of Physics, University of California Berkeley, Berkeley, CA 94720, USA. [2]Materials Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA. [3]Scuola Internazionale Superiore di Studi Avanzati, Trieste, Italy. [4]John A. Paulson School of Engineering and Applied Sciences, Harvard University, Cambridge, MA 02138, USA. [5]CNR-IOM DEMOCRITOS, Istituto Officina dei Materiali, Trieste, Italy. ✉e-mail: yshaidu@berkeley.edu; kucukben@g.harvard.edu; degironc@sissa.it

scheme[15]. Using the charges predicted by this model, the electrostatic energy was computed and subtracted from the total energy. A second network, a short-ranged model, was then tasked to learn the remainder. One limitation of this approach is its inability to capture global charge redistribution as atomic charges are only given as functions of the local environment descriptors.

An alternative approach to include long-range electrostatics in NN potentials (NNPs) is the charge equilibration scheme (CENT) proposed by Ghasemi et al.[16]. In CENT, the total energy is written as the sum of atomic contributions plus a Hartree energy of Gaussian charge distributions centered at ionic positions. A neural network is used to predict atomic electronegativity whose inputs are atomic descriptors. The atomic charges are obtained through a charge equilibration scheme[17] where total energy is minimized with respect to atomic charges subject to charge neutrality. This, by construction, results in coupled linear equations for atomic charges, and allows to capture of non-local charge redistribution. It also allows for a seamless way to conserve the total charge (see Results Section for a more detailed description). The CENT scheme has been successfully applied to construct NNPs for ionic systems[16,18,19]. However, it is not expected to work well for systems with competing covalent and ionic interactions, as the functional form chosen for the total energy lacks any description of covalent bonding.

A combination of the CENT scheme and the standard local atomic energy approach has been proposed independently by Ko et al.[20] and Jacobson et al.[21]. In both 4G-HDNNP (fourth generation high-dimensional neural network potential)[20] and QRNN (charge recursive neural network)[21], the total energy is obtained in two steps, requiring the construction of two separate NN models: one for the atomic electronegativities and the other for the atomic energies. The predictions of the first model are used for charge equilibration, and optimized against the reference atomic charges. The resulting charges are then used to compute the Coulomb energies, and supplied to the second NN in addition to the local environments, to predict the remainder of the total energy (see Results Section for more details). This approach is shown to accurately predict the energy and forces of systems with competing covalent and ionic interactions. However, it comes at the cost of training and evaluating two separate NNs, and computing forces require the solution of a second linear system.

Several other approaches to include electrostatics in machine learning interatomic potentials have been proposed[22–26]; we review here the main ones. PhysNet[22] is a message-passing high-dimensional neural network that predicts energy and atomic charges to incorporate electrostatics; it requires only total charge as input but it misses a non-local charge redistribution other than the average. CENT2[23] introduces a short-range term that is independent of the atomic charge and retains the general functional form of CENT for the long-range description while proposing a new trial electrostatic potential energy, which however requires a post-processing of the density functional theory (DFT) self-consistent charge density in addition to the decomposition of the total charge density into nuclear and electronic densities. kQeq[24] implements the CENT model with the electronegativity predicted using a kernel model and uses dipole moments as a target rather than atomic charges used in 4G-HDNNP. The long-distance equivariant (LODE)[25] framework is a kernel model that implements a long-range kernel symmetrized for long-distance equivariance. Another approach for incorporation of electrostatics in NNP is based on the position of maximally localized Wannier functions[26,27]. Self-consistent determination of long-range electrostatics in neural network potentials (SCFNN)[26], for example, incorporates electrostatics in neural network potentials via two modules, each consisting of two networks: one for short range and the other for long range. While the first module uses the position of the maximally localized Wannier function centers to predict electronic response, the second module predicts nuclear forces, requiring a total of four neural networks.

In this work, we propose an alternative approach that requires the solution of a single linear system and the training and evaluation of only one NN. To do this, we use the total energy for charge equilibration like in the CENT approach, but we retain the higher expressiveness of 4G-HDNNP through learned atomic energy terms. As detailed in Results Section, our method can be seen as a second-order Taylor expansion (in the charge variable) of generic machine-learned energy, depending on local environments and charges. By adopting this approximation we are able to replace a computationally challenging self-consistent optimization that would arise from the first derivative of the short-range energy term with respect to atomic charges, with a single NN predicting all the expansion coefficients. The neural network architecture proposed in this work enables the prediction of the environment-dependent expansion coefficients without a significant increase in the number of parameters of the model. Moreover, the model is highly flexible due to its variational nature and because all expansion coefficients are constructed to be functions of the atomic chemical environments. Thanks to this flexibility, the model can accurately describe electrostatic interactions in ionic, covalent, and mixed ionic-covalent systems. More importantly, this accuracy is reached without requiring reference atomic charges, nor any complicated approach for computing the derivative of energy with respect to atomic positions. Finally, multiple targets can be used for network optimization: while we can train the NN to minimize the error on total energy, forces, and reference atomic charges, the procedure is equally well-defined in the absence of charge targets. Besides allowing training in situations where atomic charges are not available, this additional degree of freedom can lead to the discovery of different charge decompositions, leading to similar or even better force and energy predictions, as we show in the results section.

Results

Existing approaches to including electrostatics in neural network potentials

In this section, we introduce the problem by summarizing the common aspects of the existing approaches. We consider the task of predicting the total energy E_{tot} of a system of N atoms at positions $\mathbf{R}_i$, and relative distances, $R_{ij} = |\mathbf{R}_{ij}| = |\mathbf{R}_j - \mathbf{R}_i|$. We assume that a charge q_i can be attributed to each atom i, by enforcing the constraint, $\sum_i q_i = Q$, where Q is the total charge of the system. In this work, G_i is used to describe a generic descriptor of the environment around atom i, containing information about the neighboring atomic species and their relative positions, and it is an input for the NN.

A common idea of CENT[16], 4G-HDNNP[20], and the present approach is the division of the energy into a short range and Coulomb terms:

$$E_{\text{tot}} = E_{\text{SR}} + E_{\text{Coul}}. \tag{1}$$

The first term has a different expression in each approach but generally depends on the atomic positions $\{\mathbf{R}_i\}$ and charges $\{q_i\}$. The second term is the Coulomb energy defined in terms of the total charge density $\rho(\mathbf{r})$ as:

$$E_{\text{Coul}} = \frac{1}{2}\frac{e^2}{4\pi\epsilon_0}\int \frac{\rho(\mathbf{r})\rho(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|}\mathrm{d}^3r\mathrm{d}^3r'. \tag{2}$$

Considering $\rho(\mathbf{r})$ to be the sum of atomic contributions normally distributed around each atom with a fixed spread controlled by α_i

$$\rho_i(\mathbf{r}) = \frac{q_i}{\alpha_i^3\pi^{\frac{3}{2}}}\exp\left(-\frac{|\mathbf{r}-\mathbf{R}_i|^2}{\alpha_i^2}\right), \tag{3}$$

the Coulomb energy can be written as $E_{\text{Coul}} = \frac{1}{4\pi\epsilon_0}\sum_{i<j} V_{ij}q_iq_j$ with

$$V_{ij} = \frac{\mathrm{erf}(\gamma_{ij}R_{ij})}{R_{ij}}, \tag{4}$$

where $\gamma_{ij} = (\alpha_i^2 + \alpha_j^2)^{-1/2}$.

The other common step in all the approaches is charge equilibration, i.e., the computation of the atomic charges through the minimization of an

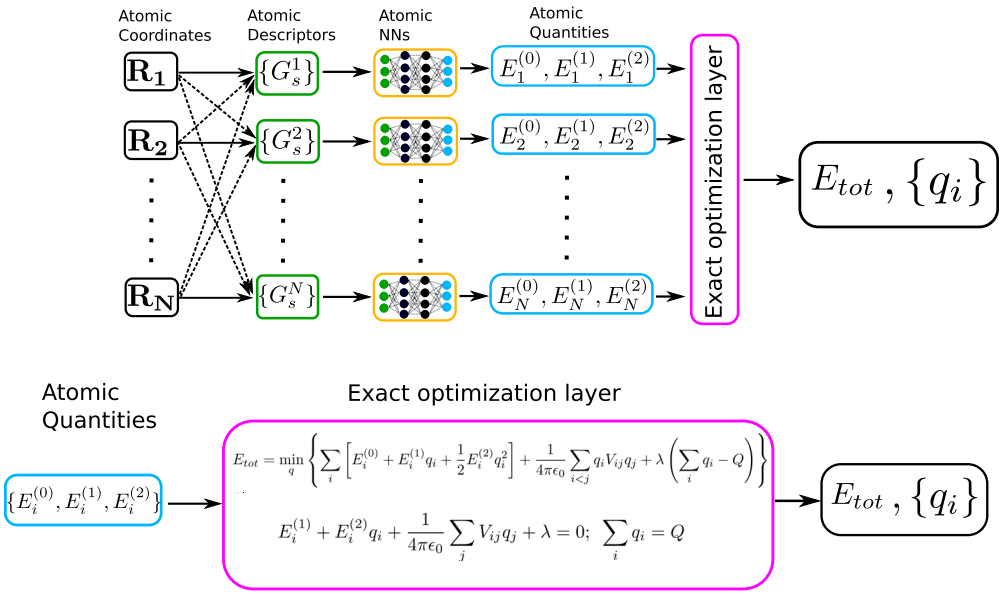

Fig. 1 | Workflow of our approach for a system of *N* atoms. For each atom i with Cartesian coordinate R_i, its descriptor, G_i is constructed and serves as the input to the atomic NN. Each atomic network produces three atomic quantities, $E_i^{(0)}$, $E_i^{(1)}$ and $E_i^{(2)}$. These outputs are then used to compute the atomic charges via the charge equilibration scheme of Eq. (7). The short-range energy contribution E_{SR} can then be obtained from Eq. (6), and the Coulomb energy E_{Coul} can be computed from the atomic charges. Their sum represents the prediction for the total energy.

energy term of the form

$$E_{\mathrm{Qeq}} = \sum_i \left(\chi_i q_i + \frac{1}{2} J_i q_i^2\right) + E_{\mathrm{Coul}}, \tag{5}$$

with some learned parameters, χ_i, representing electronegativities, and some fixed parameters, J_i, representing atomic hardnesses. Including the constraint of total charge conservation $\sum_i q_i = Q$ through a Lagrange multiplier, the $N+1$ equations system can be solved to obtain the equilibrium atomic charges.

Here, we refer to all contributions other than the Hartree energy defined in Eq. (2) as "short range" as they only depend on local atomic charges. While we use this notation for the sake of simplicity, it has to be noted that, since our charges are obtained from a global charge equilibration process, as discussed later, this term is also dependent, although indirectly, on non-local information.

In the original CENT approach, following the energy decomposition scheme introduced in Eq. (1), the short-range energy is defined as $E_{\mathrm{SR}} = \sum_i \left(E_i^0 + \chi_i(G_i) q_i + \frac{1}{2} J_i q_i^2\right)$, where only the susceptibilities, χ_i are environment dependent, while E_i^0 and J_i are fixed quantities for each species. While this allows the energy to be computed with a single NN evaluation per atom and the solution of a $N+1$-dimensional system, it lacks the flexibility to capture non-ionic interactions.

In the 4G-HDNNP scheme[20] the problem is split into two stages: a first network predicts electronegativities from the local descriptors, to be used only in Eq. (5) for charge equilibration. These charges are then used in combination with the local atomic descriptors as inputs for a second network that produces atomic energy contributions, summing to the total short-range energy $E_{\mathrm{SR}} = \sum_i E_i(G_i, q_i)$. This means that, in contrast with CENT, the charge is equilibrated with respect to a fictitious energy, and the targets of this first network must be some precomputed local charges, e.g., the Hirshfeld atomic charges. The second network has the advantage of improving the expressiveness of the local potential, however, this comes at the cost of two separate NN evaluations per atom. Moreover, computing forces in this scheme require the derivatives of the charges with respect to the positions, which can be obtained by solving another linear system.

Our approach to including electrostatics in neural network potentials

In the most general approach, we can write the short-range energy for each atom as the output of an NN depending on both the local environment and charge $E_{\mathrm{SR}} = \sum_i E_i(G_i, q_i)$, and then equilibrate the charges on the resulting total energy. This equilibration however requires the solution of a system of equations containing the derivatives of the NN with respect to the input, a complex optimization problem that could be solved self-consistently but would greatly increase the cost of this approach. To avoid this complication, we propose a simplified functional form in the charge variable, a second-order Taylor expansion around $q = 0$ local charge state:

$$E_{\mathrm{SR}} = \sum_i \left[E_i^{(0)}(G_i) + \left(E_i^{(1)}(G_i) + \chi_i^0\right) q_i + \frac{1}{2}\left(E_i^{(2)}(G_i) + J_i^0\right) q_i^2\right], \tag{6}$$

where all the $E_i^{(\alpha)}$ are obtained from an NN depending on local environments, while χ_i^0 and J_i^0 are fixed offsets for each species meant to simplify the training process. In this sense, our approach is an extension of CENT where all terms are learnable, thus increasing the potential expressiveness of the model. At the same time, the advantage with respect to 4G-HDNNP is that the three environment-dependent terms are independent, which can be exploited by making them three different outputs of the same NN. This way we can obtain the parameters in a single pass, and have a single network to optimize. Once these environment-independent parameters are obtained, the derivatives of the energy with respect to the charges can be computed and the equilibrium charges are solutions of the system of equations:

$$\begin{aligned} \sum_j \left[\left(E_i^{(2)}(G_i) + J_i^0\right)\delta_{ij} + V_{ij}\right] q_j + \\ + E_i^{(1)}(G_i) + \chi_i^0 + \lambda = 0, \end{aligned} \tag{7}$$

where δ_{ij} is the Kronecker delta and λ is the Lagrange multiplier needed to fix the total charge $\sum_i q_i = Q$.

The final energy can thus be obtained after a single network evaluation per atom and a matrix inversion. Due to the flexibility of $E_i^{(2)}(G_i)$, the matrix may be singular, especially at the beginning of the training. Although this is in part regularized by J_i^0, we have also implemented an algorithm that ensures that the diagonal term, $E_i^{(2)}(G_i) + J_i^0$, is greater than the negative of

the minimum eigenvalue of the matrix in the training dataset, hence ensuring that the matrix remains positive and definite.

The evaluation of forces is also simplified due to the variational characteristics of the model. For the short-range contribution, the derivative of the network with respect to the atomic positions becomes:

$$\mathbf{F}_i^{\mathrm{SR}} = -\sum_{j,\mu,\alpha} \frac{\partial E_j^{(\alpha)}}{\partial G_{j\mu}} q_j^{\alpha} \frac{\partial G_{j\mu}}{\partial \mathbf{R}_i}, \tag{8}$$

where j runs over atoms, α over the network outputs, and μ over the elements of the environment vectors. To obtain the total forces, we then need to add the Coulomb component $\mathbf{F}_i^{\mathrm{Coul}} = \sum_j \mathbf{F}_{ij} q_i q_j$ with

$$\mathbf{F}_{ij} = \left(\frac{\mathrm{erf}(\gamma_{ik} R_{ij})}{R_{ij}^3} - \frac{2\gamma_{ij}}{\sqrt{2} R_{ij}^2} \exp(-\gamma_{ij}^2 R_{ij}^2) \right) \mathbf{R}_{ij} \tag{9}$$

The complete workflow of this approach is summarized in Fig. 1. We train the NN through the PANNA package[8,28], employing the modified atom-centered symmetry functions[1,3,8] as implemented there.

The loss function used for training the network is the sum of three contributions, coming from target energy, force, and charge terms. The ground truth reference charges may be obtained from Hirshfeld or other equally valid decomposition schemes. In addition to these terms, we add an optional L^2 regularization term to reduce the spread of the model weights and biases. The total loss function is then:

$$\mathcal{L} = \mathcal{L}^{\mathrm{E}} + \gamma_q \mathcal{L}^q + \gamma_{\mathrm{F}} \mathcal{L}^{\mathrm{F}} + \gamma_{\mathrm{R}} \mathcal{L}^{\mathrm{R}}, \tag{10}$$

and each loss term is described as:

$$\begin{cases} \mathcal{L}^{\mathrm{E}}(W) = \sum\limits_{c \in \mathrm{batch}} \left(E_c^{\mathrm{ref}} - E_{\mathrm{tot},c}(W) \right)^2 \\ \mathcal{L}^{q}(W) = \sum\limits_{c \in \mathrm{batch}} \frac{1}{N_c} \sum\limits_{i=1}^{N_c} \left(q_i^{\mathrm{ref}} - q_i(W) \right)^2 \\ \mathcal{L}^{\mathrm{F}}(W) = \sum\limits_{c \in \mathrm{batch}} \frac{1}{3N_c} \sum\limits_{i=1}^{N_c} \left| \mathbf{F}_i^{\mathrm{ref}} - \mathbf{F}_i(W) \right|^2 \\ \mathcal{L}^{\mathrm{R}}(W) = \frac{|W|^2}{2}. \end{cases} \tag{11}$$

where the γs are prefactors that modify the relative weights of each loss component, c runs over all the examples in a training mini-batch, and i over the N_c atoms in that example. The predictions depend on W, the vector of all the weight parameters of the network.

In particular, we note that the scheme presented here is still well-defined when setting $\gamma_q = 0$, i.e., when completely ignoring the reference information on the charge decomposition. In the following section, we present example cases where doing so does not degrade the predictive performance of the method.

In the following, we test the accuracy of our approach in describing long-range electrostatic energy and forces in different systems. We start with the four emblematic cases proposed in ref. 20: a carbon chain with different terminations, a metallic silver trimer, small sodium chloride clusters, and a gold dimer on a MgO (001) surface. The training dataset including atomic information, energy, forces, and Hirshfeld charges[15] is taken from ref. 20. For all systems, we train long-range NNPs with and without considering atomic charge information in the loss function, and in both cases, we impose charge neutrality. We compare these two potentials with our trained short-range model as well as 4G-HDNNP results and short-range model (referred to as 2G) as reported in ref. 20

In addition, we benchmark our approach on solid and liquid sodium chloride configurations, recently generated to study the properties of molten salt at elevated temperatures and pressures[11]. Furthermore, we present the test of the performance of our proposed approach on Li adsorption on a graphene system. This system has been studied for Li-ion battery applications[29–40]. For these examples, only the total charge is imposed during training.

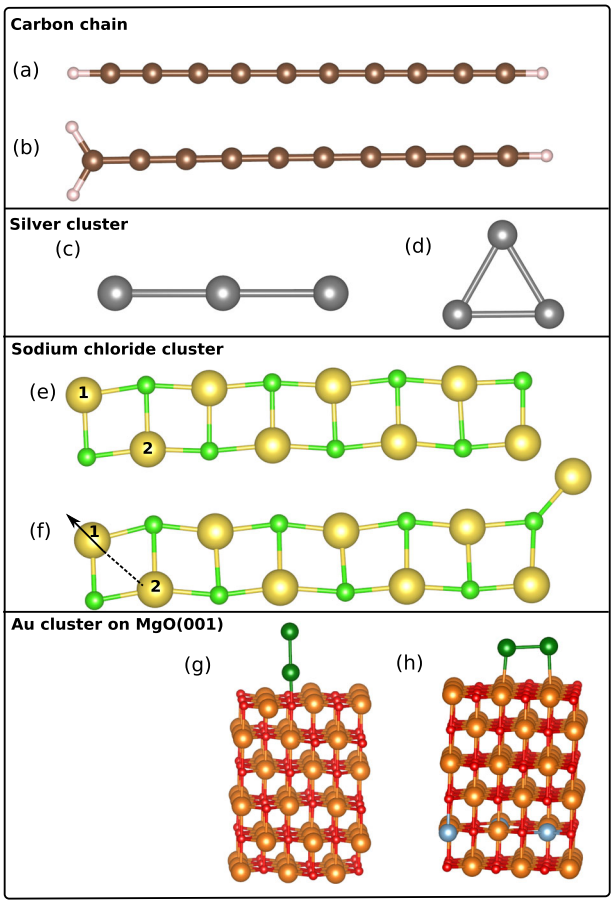

Fig. 2 | Crystal structures of example systems. a Linear carbon chain clusters $C_{10}H_2$ with total charge 0 and **b** a protonated carbon chain, $C_{10}H_3^+$ with a total charge of + 1. **c** Linear chain silver cluster with a net charge of − 1, Ag_3^- **d** triangular silver cluster with a + 1 total charge, Ag_3^+. Charged sodium chloride cluster with + 1 net charge **e** $Na_8Cl_8^+$ **f** $Na_9Cl_8^+$. Periodic structures of MgO (001) surface doped with Au dimers (**g**) undoped MgO surface and Au dimer adsorbed in the non-wetting configuration (**h**) doped MgO surface and Au dimer adsorbed in the wetting configuration.

Benchmarking: carbon chain

Here, we discuss the benchmark of our approach to the carbon chain. The two benchmark systems are as follows: 1) $C_{10}H_2$, a neutral linear chain of carbons with H-terminated ends (see Fig. 2a). 2) $C_{10}H_3^+$, which is generated by adding an extra H atom to one end of the chain and removing an electron (see Fig. 2b), resulting in a system with a total charge of +1. The additional hydrogen perturbs its local environment and causes a global charge redistribution. Further details about the dataset are presented in the Methods section IV.

Table 1 shows the root mean square error (RMSE) in charge, energy, and force on training and validation sets. When the long-range model is trained on the Hirshfeld target, we obtain an RMSE on the validation set of 1.167 meV per atom and 78.98 meV Å^{-1} for energy and forces, respectively, which are comparable to the ones obtained with 4G-HDNNP. Remarkably, lower RMSE is obtained for energy and forces when the long-range model is not regressed against Hirshfeld charges ($\gamma_q = 0$). The distribution of atomic charges and the deviation from the Hirshfeld charge decomposition[15], in this case, depends on the initial conditions and can be tuned via the parameters χ and J. However, the accuracy of the model is independent of the choice of

Table 1 | Training and validation RMSE for carbon chains

Model		Charge (me)	Energy (meV/atom)	Force (meV/Å)
SR				
	train	-	1.48	117
	test	-	1.47	117
LR				
$\gamma_q > 0$	train	10.3	1.17	78
	test	10.4	1.17	79
$\gamma_q = 0$	train	93.3	1.15	72
	test	93.6	1.15	73
2G	train	-	1.58	131
	test	-	1.62	130
4G	train	5.8	1.21	78
	test	6.6	1.19	78

SR represents a short-range model which has the same functional form as 2G in ref. 20 and the LR represents our long-range model with $\gamma_q > 0$ representing the LR model trained with charges and $\gamma_q = 0$ is trained without target charges.

Table 2 | Training and validation RMSE for silver trimers

Model		Charge (me)	Energy (meV/atom)	Force (meV/Å)
SR				
	train	-	320.3	1438
	test	-	303.1	1463
LR				
$\gamma_q > 0$	train	2.3	0.9	20
	test	2.2	0.8	20
$\gamma_q = 0$	train	73.5	0.8	19
	test	72.9	0.8	19
2G	train	-	355.0	1803
	test	-	352.0	1812
4G	train	10.6	1.3	32
	test	10.0	1.3	32

The RMSE on charge, energy, and force for the silver trimers. SR represents the short-range model whose functional form is the same as 2G. For the long-range model, $\gamma_q > 0$ represents the model trained using target charges, and $\gamma_q = 0$ without reference charges.

these parameters and the resulting charge decomposition (see Supplementary Table 2).

Next, we examine the charge distributions for the DFT-optimized geometries. In Fig. 3a, b, we depict the charge distribution on the DFT-optimized structures of $C_{10}H_2$ and $C_{10}H_3^+$ systems, respectively, comparing the results obtained when using or neglecting the Hirshfeld charges as a target. For the $C_{10}H_2$ configuration (see Fig. 3a), both models capture the symmetric nature of the charge distribution around the center of the linear chain. A sign difference can be seen in the second C atoms from either edge, to which the two models assign charges with opposite signs. We note that the model regressed against charges should have similar charges to the Hirshfeld ones, which it was trained to reproduce. In the case of $C_{10}H_3^+$ (see Fig. 3b), similar charges are assigned to atoms on the unprotonated edge. Negligibly small negative charges are assigned to the two hydrogens on the protonated edge by the model not trained on reference charges while the model trained on charges assigned larger positive charges. The difference in charge distribution does not influence the accuracy of energy and forces.

Benchmarking: silver clusters

We now consider an Ag cluster system. The training configurations consist of triangular and linear Ag trimers with +1 and −1 total charges, respectively. Representative structures for the linear and triangular configurations are shown in Fig. 2c, d, respectively.

Due to the size of the systems, a typical cutoff radius, usually chosen between 5–10 Å to describe atomic chemical environments contains all the interactions of the central atom with every other atom in the systems. However, the dependence of the energetically favorable geometry on the total charge of the system makes it an ill-posed problem for machine learning models that are built solely from atom-centered symmetry functions based on atomic coordinates. Indeed, as demonstrated in ref. 20, the 4G-HDNNP model was able to accurately reproduce the energy and forces of these systems, while the 2G model yields a very large root mean square error in both energy and forces. Here, we show that our proposed approach can accurately reproduce the DFT potential energy surface of the systems even when local charge information is not trained on. More interestingly, we attain a lower RMSE in energy, forces, and Hirshfeld charges (when used as a target) than 4G-HDNNP.

Table 2 shows the RMSE on charges, energy, and forces obtained in this work and compared with 2G- and 4G-HDNNP results. The accuracy attained with our short-range model is comparable to those of 2G-HDNNP. As shown in the table, the RMSEs on energy and forces are lower than those obtained with 4G-HDNNP. Similarly, when trained on Hirshfeld charges we obtain a smaller RMSE on charges than the 4G-HDNNP. The RMSE on energy and forces are similar whether trained or not trained on charges. In the case where the Hirshfeld charges are not used as a target ($\gamma_q = 0$), an RMSE of 0.07 e is observed. This deviation of the atomic charge from Hirshfeld is not surprising since there is no uniqueness in the decomposition of the total charge density into atomic contributions, hence the decomposition obtained by our model is as valid as the Hirshfeld DFT atomic charges, provided that physical observables such as energy and forces are accurately predicted.

To shed light on the origin of the RMSE in charges when atomic charge information is not used, we examine the charge distribution on the lowest energy structure in the training dataset for each charge state. In Fig. 3c, d, we compare the atomic charges predicted with our long-range models with the Hirshfeld charges results. As can be seen, when regressed against Hirshfeld charges ($\gamma_q > 0$), excellent agreement is observed for both linear and triangular clusters. When the long-range model is not trained on Hirshfeld charges ($\gamma_q = 0$), we obtain a charge distribution that is identical to the Hirshfeld charges for the triangular cluster with a total charge of + 1 but a different charge distribution is found for the linear cluster with total charge of − 1. The NNP-Hirshfeld charge distribution agreement observed for the + 1 charge state is a consequence of the atomic arrangement of the system: all atoms have an equivalent chemical environment, hence there is a unique distribution of charges to atoms, compatible with a total charge of + 1. On the other hand, for the case of the linear system of charge − 1, several distributions are possible, with the only constraint being that the atoms at the edges with similar chemical environments have similar magnitude of charges. Here, we obtain − 0.331 e for the two atoms at the edges and − 0.338 e for the central atom compared to the Hirshfeld charges of − 0.404 e and − 0.192 e, respectively.

Finally, we investigate the accuracy of our NNP models in describing the potential energy surface along a reaction coordinate such that a transformation from the triangular cluster into a linear one can be monitored. In particular, we use the angle between atoms 1, 2, and 3 as the reaction coordinate, denoted as θ_{132}. In Fig. 4a, we depict the relative energy at each charge state as a function of θ_{132}. Both long-range models, $\gamma_q = 0$ and $\gamma_q > 0$, are identical in the relevant region of the reaction coordinates and correctly describe the two energy wells. The equilibrium position for the system with a total charge of + 1 is predicted to be at $\pi/3$ while that with a total charge of − 1 is at π. The short-range model fails to capture the expected trends as it cannot distinguish between the two systems around the coexistence region between the Ag_3^+ and Ag_3^-.

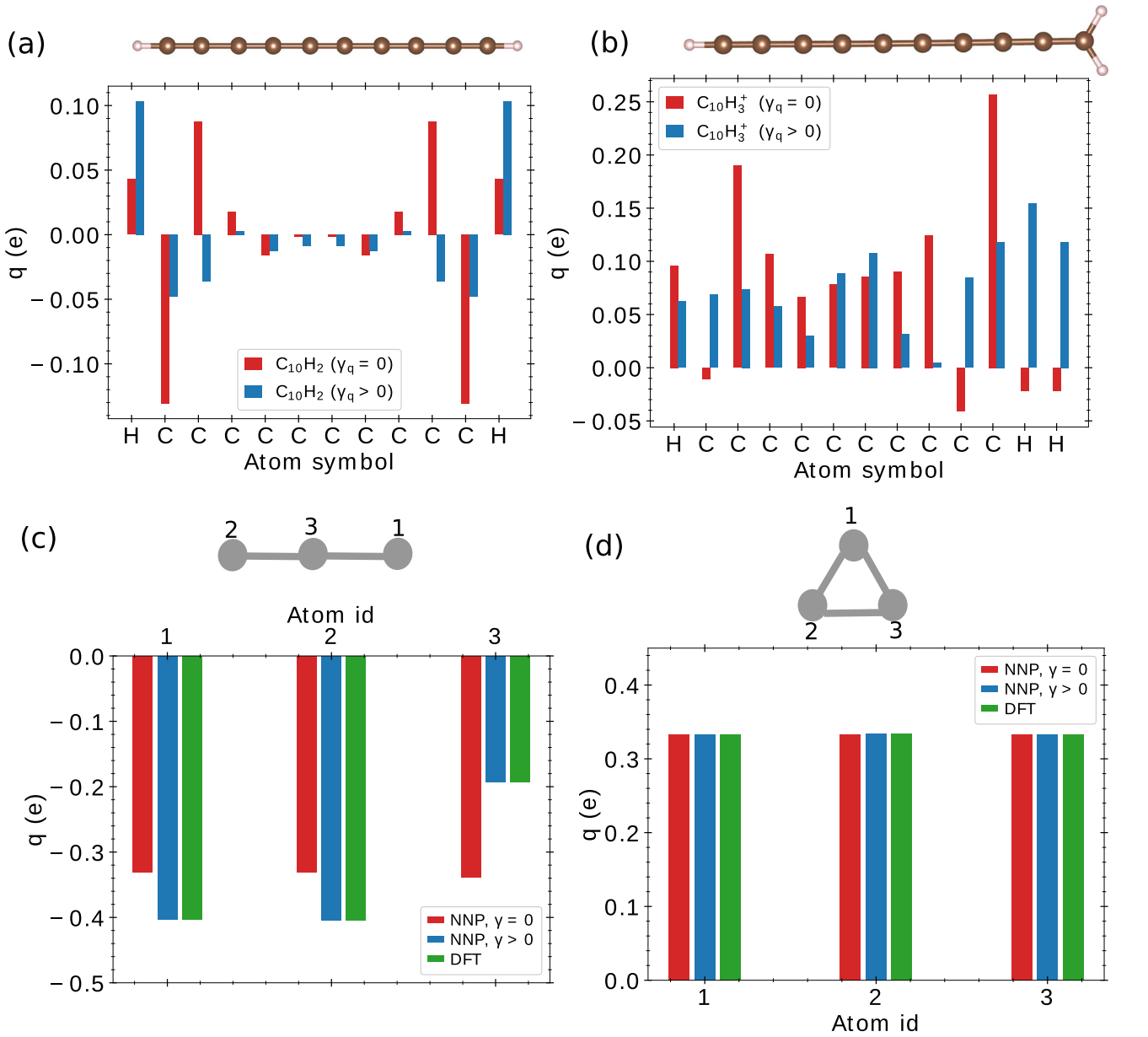

Fig. 3 | Comparison of charge distribution between long-range model trained and not trained on reference charges on DFT-optimized structures of carbon chains systems. a $C_{10}H_2$ and **b** $C_{10}H_3^+$. predicted atomic charges on the lowest energy structures in the training dataset of silver clusters systems, **c** Ag_3^-, and **d** Ag_3^+ compared with DFT values obtained with Hirshfeld decomposition.

Benchmarking: sodium chloride clusters

In this section, we test our model on sodium chloride clusters and compare its accuracy to that achieved by 4G-HDNNP. The systems considered here consist of $Na_8Cl_8^+$ (8 Na atoms and 8 Cl atoms) and $Na_9Cl_8^+$ (9 Na atoms and 8 Cl atoms) each having a total charge of + 1. A representation of these configurations is depicted in Fig. 2e, f.

The accuracy of the 4G-HDNNP was evaluated against the DFT potential energy surface as the Na atom is moved along the path indicated by the black arrow in Fig. 2f. The potential energy surface has two distinct minima along the path shown by the arrow for the two configurations. The 4G-HDNNP was demonstrated to accurately reproduce the DFT potential energy surfaces while the short range failed to distinguish these two minima. Here, we show that our approach not only gives lower RMSE in energy and forces than 4G-HDNNP both with and without regressing the model against Hirshfeld charges but also accurately reproduces the DFT potential energy surfaces in a similar manner to the 4G-HDNNP with about two times fewer parameters.

In Table 3, we show the RMSE in charges, energy, and forces for both short-range and long-range models and compare them with 2G- and 4G-HDNNP results. Our long-range and short-range models attain similar RMSE in energy to the results in ref. 20. Instead, we obtain lower RMSE in forces than 4G-HDNNP with our long-range models trained with and without charges.

To examine the performance of our approach on the potential energy surface of these systems, we compute the energy and force acting on the Na atom, indicated by 1 in Fig. 2f, when moving along the arrow. Figure 4b, c show the energy of the systems and force on the Na atom projected along the direction shown by the arrow in Fig. 2f. The DFT results are obtained from Fig. 7 of ref. 20. For the short-range model, we obtain similar trends to those reported in ref. 20, where the position of the minimum for $Na_8Cl_8^+$ and

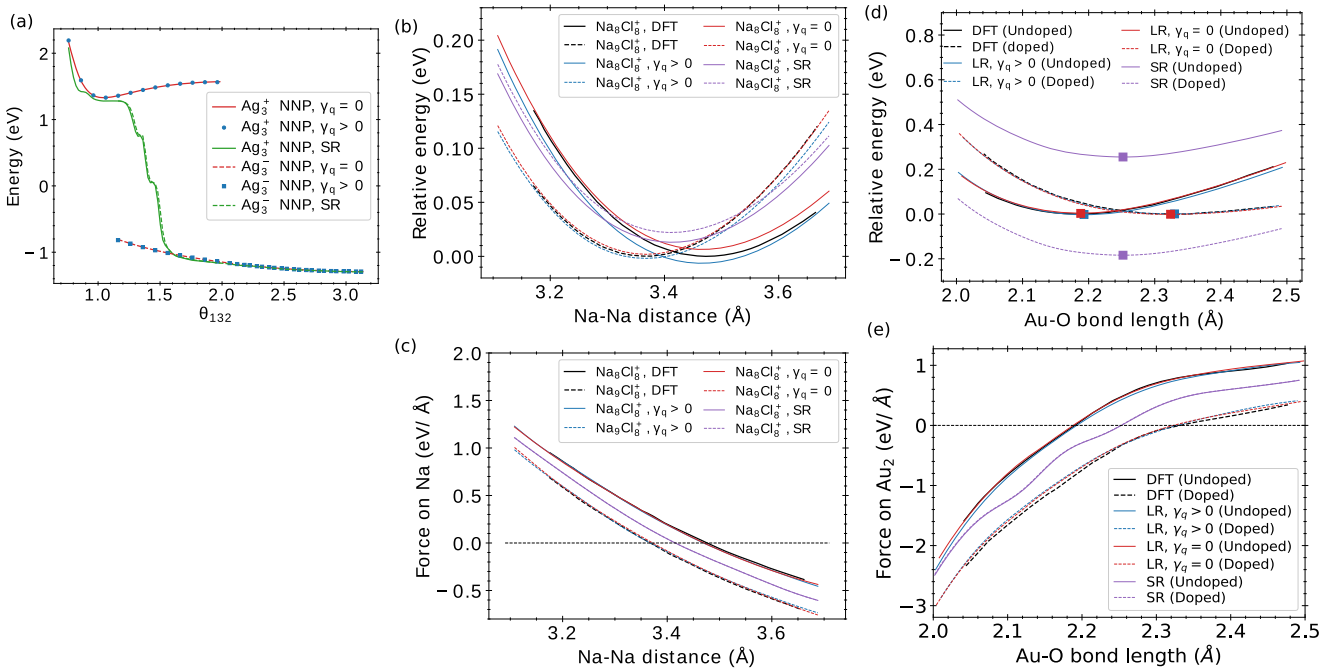

Fig. 4 | Comparison of energy and forces between NNPs and DFT. a Potential energy surface along θ_{132} for Ag_3^+ and Ag_3^-. For each charge state, energy is referenced to the atomic energies. The short-range model predicts the same energy for Ag_3^+ and Ag_3^-. Energy and force on Na atom along the arrow shown in Fig. 2f: **b** the energy as a function of the distance between Na atom 1 and 2. **c** projected force on Na atom 1 in the direction of the arrow shown, as a function of the distance between Na atom 1 and 2. Our NNP energies are in reference to the average of the minimum energy predicted by the two long-range models. Potential energy surface of Au cluster supported on MgO(001) surface: **d** energy versus Au-O bond length of our long-range model. **e** Force on the Au cluster projected along the perpendicular direction as a function of the bond length measured between the closest O atom to the cluster and the closest Au atom to the surface. DFT forces are obtained from ref. 20. The closed squares indicate the position of energy minima.

Table 3 | Training and validation RMSE for NaCl clusters

Model		Charge (me)	Energy (meV/atom)	Force (meV/Å)
SR				
	train	-	1.6	48
	test	-	1.6	48
LR				
$\gamma_q > 0$	train	13.0	0.4	19
	test	13.4	0.4	19
$\gamma_q = 0$	train	153.0	0.4	19
	test	154.8	0.4	19
2G	train	-	1.7	58
	test	-	1.7	57
4G	train	15.9	0.5	32
	test	15.8	0.5	33

The RMSE in charge, energy, and force obtained with the short-range (SR) model with the same functional form as 2G and the long-range (LR) model which can be compared to the 4G model. For the LR model, $\gamma_q > 0$ represents the model trained with atomic charge reference, and $\gamma_q = 0$ without charge target. The RMSE presented for 2G and 4G are obtained from ref. 20.

Table 4 | Training and validation RMSE for Au dimer on MgO(001)

Model		Charge (me)	Energy (meV/atom)	Force (meV/Å)
SR				
	train	-	2.3	125
	test	-	2.3	135
LR				
$\gamma_q > 0$	train	64.5	0.2	44
	test	65.8	0.2	52
$\gamma_q = 0$	train	475.1	0.2	39
	test	482.4	0.2	43
2G	train	-	2.3	155
	test	-	2.3	153
4G	train	5.6	0.2	81
	test	5.7	0.2	66

The RMSE in charge, energy, and force was obtained with the short-range (SR) model, long-range (LR) model trained with reference charge target ($\gamma_q > 0$) and without charge references ($\gamma_q = 0$). The RMSE reported for 2G and 4G are obtained from ref. 20.

$Na_9Cl_8^+$ are the same but separated by constant energy. Our long-range model accurately reproduces the DFT energy minima for both $Na_9Cl_8^+$ and $Na_8Cl_8^+$ with and without regressing against reference charges. Similarly, the forces on the Na atom computed with our long-range models are indistinguishable from the DFT results. We note that all energies in Fig. 4b, c are referenced to the average minimum energy predicted by our long-range models of the respective systems, while the DFT energy curves are referenced to DFT energy minima.

Benchmarking: gold dimer on MgO(001) surface

Here, we test our model on a periodic system: a gold dimer (Au_2) supported on a MgO(001) surface[20]. The dataset presented in ref. 20 consists of two configurations of Au dimer on the MgO(001) surface: the "wetting" configuration, in which the atoms of the Au dimer lie at a similar distance above two surface Mg atoms, and the "non-wetting" one, in which one Au atom binds to a surface Oxygen and the other Au atom is farther from the surface Oxygen. The relative stability of the wetting and non-wetting configurations

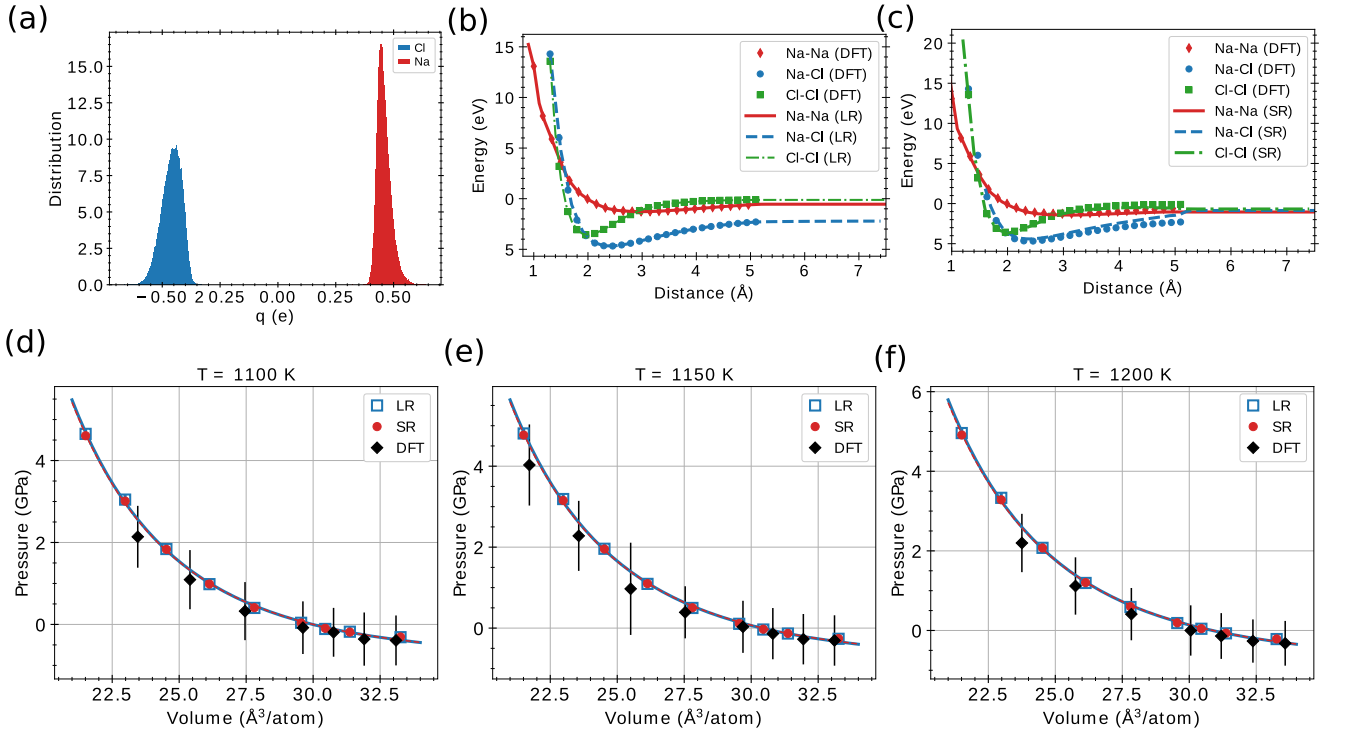

Fig. 5 | Charges distribution of liquid and solid states NaCl and potential energy surface Na-Cl dimers. a Atomic charges. the peak around -0.5 e corresponds to the charges for Cl and that around +0.5 e are for Na atoms. The potential energy of NaCl dimer as a function of atom separation. The comparison of DFT energies with **b** long-range model and **c** short-range model. **d**, **e**, and **f** are the pressure-volume plots at 1100, 1150, and 1200 K, respectively obtained with our long-range model (LR), short-range model (SR), and the DFT results as reported in ref. 11. The solid and dashed lines correspond to a fit to the Murnaghan equation of state.

is influenced by doping: when the MgO slab is doped with Al atoms, the wetting geometry is more energetically favorable, while the non-wetting geometry is preferred on the surface of the intrinsic MgO(001) surface. Here, we compare the performance of our long-range approach to the 4G-HDNNP model in describing the electronic structure of these systems.

Table 4 shows the RMSE in charges, energy, and forces, compared with those reported in ref. 20. With our long-range model, irrespective of whether it is regressed against reference charges or not, the RMSE in energy is similar to that reported for 4G-HDNNP. Remarkably, our long-range models attain lower RMSE in forces than 4G-HDNNP. The RMSE in charges obtained is larger than that of the 4G-HDNNP even when regressed against reference atomic charges. These results suggest that the Hirshfeld atomic charges as provided in the reference dataset are not necessarily the charge density decomposition that minimizes our long-range energy model.

Next, we compare the charge decomposition predicted with our model to the Hirshfeld charges. In Supplementary Fig. 9, we depict the charge distribution as obtained with our long-range model. In all cases, the charges predicted for O, Mg, and Al follow the trends of the Hirshfeld charges, although the NNP predictions have wider spreads. Similarly, for the Au_2 molecules in the non-wetting case, the double peak Hirshfeld charges located in both negative and positive regions are captured. However, for the wetting Au_2 geometry, our long-range model predicts positive charges with and without doping while the Hirshfeld charge decomposition assigns positive charges for Au_2 without doping and negative charges when the system is doped. Previous DFT studies that use Bader decomposition techniques to assign charges for these systems[41,42] found that the Au_2 molecule is negatively charged with and without doping, the main effect of the doping being that the charge on Au_2 becomes more negative upon doping instead of the sign change observed with the Hirshfeld charge decomposition here. We note that when the long-range model is not regressed against reference charges, we observe negative charges for Au_2 irrespective of doping for the wetting geometry (see Supplementary Fig. 10).

From this analysis, we conclude that our long-range model trained on charges minimizes the total energy functional through an atomic charge decomposition different from the reference Hirshfeld charges. We will show in the following that these deviations have no negative consequences on the ability of our model to accurately describe the electronic structure of the systems.

Next, we compute the energy difference between the wetting and non-wetting geometry of the Au_2 molecule with and without doping. In the doped system, our long-range models with and without target charges predict an energy difference of −50 meV and −67 meV, respectively, comparable to − 41 meV obtained with the 4G-HDNNP and in qualitative agreement with DFT prediction of − 2.7 meV. For the undoped system, we obtain the energy difference between wetting and non-wetting geometry as 931 meV when our long-range model is trained with target charges and 924 meV when target charges are not used. These values are in excellent agreement with the DFT prediction of 929 meV and similar to the result obtained with 4G-HDNNP of 975 meV. The short-range model predicts the same energy differences for doped and undoped systems, similar to the results obtained with 2G-HDNNP[20].

Finally, we report the performance of our approach in describing the potential energy surface of the Au_2 molecule on MgO(001) surface. Fig. 4d, e show, respectively, the energy and force on Au_2 in the non-wetting geometry as a function of the distance between the closest surface O atom and Au_2. The energies obtained in this work are referenced to the average of the minima predicted by the two long-range models (i.e., models obtained with and without training on charges). Our long-range models correctly predict two distinct minima for the non-wetting geometry for doped and undoped configurations and compare well with the DFT predictions, while the short-range model fails to distinguish the adsorption of Au_2 on doped and undoped MgO(001) systems except for a constant shift in energy. We then compute the force on Au_2 projected along the direction perpendicular to the surface and compare it with DFT results from ref. 20. This force is equivalent to the derivative of the energy depicted in Fig. 4d. Both of our long-range

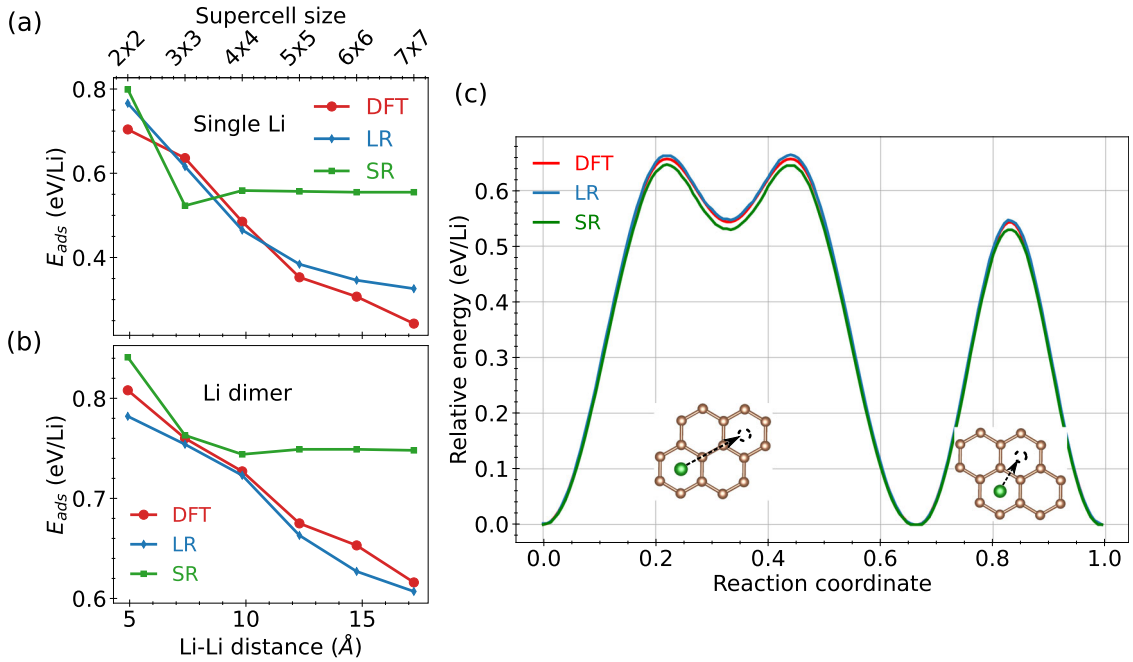

Fig. 6 | Lithium adsorption on graphene. Adsorption energy, E_{ads} of **a** isolated Li atoms and **b** Li dimers on graphene supercell. **c** Li diffusion on the graphene surface along the energy path for Li to migrate between two hexagon centers as indicated in the insets.

models accurately reproduce the DFT forces for doped and undoped systems. The short-range model gives the same force irrespective of doping.

Benchmarking: molten salt

We now benchmark our method on a dataset that was recently used to construct a neural network potential for molten sodium chloride[11] using a short-range model. The dataset comprises a wide variety of chemical structures spanning crystalline and high-temperature and pressure liquid phases of sodium chloride. The dataset does not contain atomic charge information, therefore, only the total charge of the system will be used to train the model. There are ~112,000 configurations including NaCl dimer configurations.

Here, we divide the dataset into 80:20 ratios for training and validation of the performance of our long-range model. As in the previous case, we use a modified Behler-Parrinello descriptor to describe the local atomic environment up to a cutoff of 10 Bohr. Details of the parameters used to describe atomic chemical environments are presented in Supplementary Note 3.

The RMSE on energy and forces obtained by the long-range model are 0.990 meV per atom and 30.22 meV Å^{-1}, respectively. This is a significant improvement over the short-range model that gives 1.349 meV per atom and 40.58 meV Å^{-1} for the energy and force RMSE, respectively. These results imply the presence of an electrostatic component in the energy and forces that are not well described by the short-range model but are captured by our long-range one, even without regressing against reference atomic charges.

Next, we examine the charge distribution predicted by our long-range model. Figure 5a shows the charge distribution predicted for the configurations in our validation set. The long-range model correctly assigned positive charges around +0.5 e to the Na atoms and negative charges around −0.5 e to the Cl atoms. The distribution rather than a delta-like distribution, expected for perfect crystalline NaCl systems, presents a substantial spread in the predicted charges, due to the presence of amorphous and liquid phases of NaCl in the test dataset, characterized by distorted octahedral chemical environments.

Next, we discuss the performance of our long-range model in describing the potential energy surfaces of isolated Na-Cl, Na-Na, and Cl-Cl dimers. In Fig. 5b, c, we show the energy of these dimers as a function of interatomic distances obtained with long-range and short-range models, respectively. The DFT potential energy surfaces are well described by both long-range and short-range models at short distances. However, as the interatomic distances approach 5.3 Å, the cutoff radius for the local atomic descriptors, the short-range models fail to distinguish the total energy between the different dimers (see Fig. 5c). On the other hand, our long-range model predicts the expected trends beyond 5.3 Å as shown in Fig. 5b. These results demonstrate the ability of our long-range model to accurately describe the physics of NaCl beyond the cutoff radius.

Furthermore, we investigate the accuracy of the long-range model on the equation of state of liquid phase of NaCl. We perform an NVT (constant number of atoms, volume, and temperature) molecular dynamics (MD) at different volumes using a supercell containing 512 atoms and a timestep of 0.5 fs. The pressure-volume plots at 1100, 1150, and 1200 K are shown in Fig. 5d–f, respectively. The errors estimated as the standard deviation of the MD results are smaller than the open squares for the long-range and filled circles for the short-range model. The DFT results reported in ref. 11 are also shown as black diamonds with their associated errors. The results obtained with short and long-range models are similar, suggesting a significant screening of the electrostatic field generated by the Na and Cl ions in molten salt, which is consistent with the properties of molten salt computed with a short model as reported in ref. 11. More interestingly, our short-range model developed with a maximum cutoff at 5.3 Å yields the radial distribution function of molten salt at 1150 K as accurate as DFT and our long-range electrostatic model (see Supplementary Fig. 11), again strengthening the observation that a local potential may be sufficient to describe the properties of molten salt.

The equilibrium average volume per atom predicted by long-range at 1100, 1150, and 1200 K are 29.99, 30.40, and 30.85 Å^3/atom, respectively, while the short-range model gives 30.01, 30.44, and 30.90 Å^3/atom, compared to the DFT results of 29.3, 29.7, and 30.1 Å^3/atom. Our consistent overestimation can be attributed to the simulation setup, considering the large error in pressure associated with the DFT results. The bulk moduli predicted by the long-range model are 4.89, 4.75, and 4.57 GPa at 1100, 1150, and 1200 K, respectively, while the short-range model predicts 4.85, 4.70 and 4.53 GPa at 1100, 1150, and 1200 K, respectively. The corresponding bulk moduli obtained with DFT are 4.8, 4.2, and 4.3 GPa respectively. Finally, the thermal expansion obtained with long range and short-range models at 1150 K are 2.85×10^{-4} and 2.92×10^{-4} K^{-1},

respectively, which are in agreement with the DFT prediction of 2.7×10^{-4} K^{-1} [11]. Although the effects of long-range electrostatics are seen to be minimal, overall, the long-range model performs better than the short-range model.

Benchmarking: lithium adsorption on graphene

In this section, we further demonstrate the importance of long-range electrostatics on a system of general interest: Li on graphene, a prototypical example of Li-C interactions, and a system of general interest for rechargeable battery applications. Li interactions with graphene have been extensively studied, mostly within DFT[31,33–40], due to its being a potential standalone alternative anode for lithium-ion batteries (LIBs), and as a model to explain the high Li density observed in disordered, non-graphitic carbon materials[29,30]. Recently, we have also studied the effect of temperature on Li interaction with graphene using electrostatic energy-corrected cluster expansion techniques[32]. In that work, we highlighted the importance of incorporating an electrostatic term in the cluster expansion energy in order to accurately describe the binding energy of Li, especially in the dilute limit.

The details about data generation, DFT calculations, and the neural network parameters are in Supplementary Note 3. The RMSEs in energy for the short-range model on the training and validation set are 2.05 and 2.10 meV per atom, respectively. The corresponding RMSEs in forces are 45.40 and 45.75 meV $Å^{-1}$ for the training and validation dataset. The RMSEs in energy obtained with the long range model are 1.57 and 1.65 meV per atom for the training and validation set. The RMSEs in forces are 43.48 and 44.21 meV$Å^{-1}$, respectively.

While the RMSEs alone might seem to indicate that the long-range model is only slightly better than its short-range counterpart, we will now show that the addition of long-range interaction is essential to capture some physically relevant quantities. We will start by investigating the thermodynamics of Li adsorption on graphene, considering the adsorption energy as a function of Li concentration. This quantity is particularly important for the study of Li occupation with grand canonical Monte Carlo. The adsorption energy is defined as:

$$E_{\mathrm{ads}}(n) = \frac{E(n\mathrm{Li} + \text{ graphene}) - E(\text{graphene}) - nE(\mathrm{Li})}{n}, \qquad (12)$$

where $E(n\mathrm{Li} + \text{graphene})$ is the total energy of a Li-graphene system with n Li atoms, E(graphene) is the total energy of a bare graphene sheet and E(Li) is the total energy per atom of Li metal in the bcc phase.

We consider the adsorption of single Li atoms and Li dimers on 2×2, 3×3, 4×4, 5×5, 6×6 and 7×7 supercells. Specifically, Li atoms are adsorbed at the "hollow site" (i.e., the center of hexagons on the honeycomb lattice), and a Li dimer corresponds to two Li atoms adsorbed on neighboring hexagon centers. A single Li adsorbate on a 2×2 graphene supercell corresponds to 1:8 Li:C ratio, while on a 7×7 supercell, the ratio Li:C becomes 1:98. For the Li dimer adsorption, on a 2×2 graphene supercell, the ratio of Li to C is 1:4 while on a 7×7 supercell, the ratio is 1:49. Additionally, we compute the Li diffusion on the basal plane of graphene, a quantity that measures the kinetic properties of Li. For both Li adsorption energy and Li diffusion, we compare NNP predictions with the underlying PBE-based DFT results.

Figure 6a, b show the adsorption energy of single Li atoms and Li dimers for several supercell sizes, corresponding to different Li concentrations and varying minimum Li-Li distances. The DFT adsorption energy decreases with increased supercell size (i.e., increasing Li-Li distances) for both single Li and Li dimers adsorbates. Remarkably, the DFT trends are excellently captured by our long range model. The short-range model fails to reproduce the DFT adsorption energy, naturally saturating when the Li-Li distance exceeds the cutoff with which the descriptors are built. The inability of the short-range model to describe the binding energy of Li interaction with graphene, especially at low concentrations, makes it unsuitable to accurately describe the thermodynamics of Li-graphene systems and, by extension, Li-C systems. These results again point to the importance of correctly accounting for long-range electrostatics to describe thermodynamics in systems with significant electron transfer.

As a further example, in Fig. 6c, we show the relative energy against reaction coordinates of Li diffusion in a 6×6 graphene supercell along the high symmetry points, namely the center of a hexagon (H), the top of a carbon atom (T), and the bridge between two nearest neighbor C atoms (B). In this case, the Li height is fixed at 1.72 Å, the equilibrium height of Li adsorption at the H site. While the short-range model underestimates the energy barrier, due to its inability to capture interactions between periodic images of Li atoms, the long range is almost indistinguishable from the DFT results. To understand the implications of underestimating the activation barriers, we compute the enhancement of the diffusion constant at 300 K, as given by $\frac{D(T)}{D_0} = \bar{D} = \exp\left(-\frac{E_a}{k_B T}\right)$. The estimated diffusion coefficients from H to B sites, $\bar{D}_{H\to B}$ are 766, 650, and 1283 for DFT, long range, and short range, respectively. This implies a slight underestimation of 15% with respect to DFT for the long-range model, while the short-range model overestimates the DFT value by 68%. Similarly, the estimated diffusion coefficients from H to T sites, $\bar{D}_{H\to T}$ are 9.4, 7.0, and 13.8 for DFT, long range, and short range, corresponding to a 47% overestimate for the short-range model, much worse than the 25% of the long-range model.

Discussion

In this work, we propose a new scheme to incorporate electrostatic interactions in neural network interatomic potentials. The current scheme decomposes the total energy functional into short-range and Coulomb energies and consistently fixes the Coulomb energy from atomic charges obtained from the minimization of the total energy. The ingredients that enter the linear system which determines the atomic charges are obtained from atomic neural networks. Unlike 4G-HDNNP, only a single network potential for the atomic energy is required and the dependence of atomic forces on the variation of atomic charges with respect to ionic displacement is eliminated.

The approach introduced here is similar to the charge equilibration neural network (CENT) in that the atomic charges are obtained from total energy minimization with respect to the charges themselves. However, it is as general as the 4G-HDNNP, as it is applicable to all chemical systems including systems with competing covalent and ionic bonding, as well as systems with multiple charge states.

We benchmark our model on six examples. In most cases, our long range models perform as accurately as, or sometimes better than, the 4G-HDNNP. We demonstrate that the accurate representation of DFT energy and forces is independent of the details of the local charge distribution. Therefore, even in the case where Hirshfeld atomic charges are not accurately reproduced, the RMSEs in energy and forces are equivalent to or better than the results obtained with 4G-HDNNP, and the potential energy surfaces are in excellent agreement with DFT, even in cases that the short-range model fails to describe.

Our model is easy to incorporate into the existing architecture and, unlike the 4G-HDNNP, does not rely on the availability of any reference atomic charges. Therefore, datasets generated to train models that are derived from local environment descriptors are sufficient to describe electrostatics.

Methods

Computational details

In ref. 20, a neural network architecture with 2 hidden layers of 15 nodes each and a single-node output layer was used for both the short-range atomic energies and electronegativities. In order to compare our model with theirs, we adopt the same NN architecture for the first four examples except that for our long-range model, the output layer has 3 nodes rather than 1. We reiterate that only one NN is evaluated to compute the total energy in our approach while two NNs (one for electronegativity and one for the

short-range energy term) need to be evaluated in the case of 4G-HDNNP methods. For the last benchmark example, we used NN architecture with 2 hidden layers each of 64 nodes and either a single-node (for the short-range model) or three-node (for our long-range model) output layer.

For the linear carbon systems, the training dataset[20] was generated by performing ab initio molecular dynamics (MD) simulations for each system at 300 K. For each system, ~5000 MD trajectories were generated giving rise to ~ 10,000 configurations. Similarly, for the silver trimers, the training dataset is generated from ab initio MD simulations at 300 K as in the previous example. In the case of the sodium chloride clusters, the training dataset is generated from random displacement of atomic coordinates according to Gaussian distribution with 0 mean and 0.05 Å standard deviation.

In all cases, the atomic environments are described using the modified Behler-Parrinello descriptors as presented in ref. 8,28. The size of the input vectors for the carbon chain, silver cluster, NaCl cluster, and Ag dimer on the MgO surface are respectively, 60, 12, 45, and 256 compared to 60, 12, 45, and 151 used in ref. 20. The NNP for Li on the graphene problem is constructed with environment descriptors of size 112. The specific parameters for each system are highlighted in the Supplementary Note 3.

The relative contributions of each term in the loss function Eq. (10) are tuned with the coefficient γ also specified in each section (except for the constant $\gamma_R = 10^{-4}$). In particular, the case $\gamma_q = 0$ represents models that are not regressed against reference charges.

In all cases, the hidden layers are activated with Gaussian functions while the output layers are linear. While the choice of χ^0 and J^0 does not influence the accuracy of the predicted energy and forces, the distribution of the atomic charges can depend slightly on these parameters when the model is not regressed against reference atomic charges. Here, we discuss the results obtained with χ^0 and J^0 as tabulated in Supplementary Table 1. These parameters are computed from ionization potential and electron affinity obtained from the Mendeleev package[43].

Data availability

All data used for the benchmarking examples 1 to 5 can be found in ref. 20 and ref. 11, the data used, for example, 6 can be found in ref. 44.

Received: 1 April 2023; Accepted: 9 February 2024;
Published online: 04 March 2024

References

1. Behler, J. örg & Parrinello, M. Generalized neural-network representation of high-dimensional potential-energy surfaces. *Phys. Rev. Lett.* **98**, 146401 (2007).
2. Bartók, A. P., Payne, M. C., Kondor, R. & Csányi, G. ábor Gaussian approximation potentials: the accuracy of quantum mechanics, without the electrons. *Phys. Rev. Lett.* **104**, 136403 (2010).
3. Smith, J. S., Isayev, O. & Roitberg, A. E. ANI-1: an extensible neural network potential with dft accuracy at force field computational cost. *Chem. Sci.* **8**, 3192–3203 (2017).
4. Artrith, N. & Behler, J. örg High-dimensional neural network potentials for metal surfaces: a prototype study for copper. *Phys. Rev. B* **85**, 045439 (2012).
5. Artrith, N., Urban, A. & Ceder, G. Constructing first-principles phase diagrams of amorphous Li x Si using machine-learning-assisted sampling with an evolutionary algorithm. *J. Chem. Phys.* **148**, 241711 (2018).
6. Bartók, A. P. & Csányi, G. ábor Gaussian approximation potentials: a brief tutorial introduction. *Int. J. Quant. Chem.* **115**, 1051–1057 (2015).
7. Rowe, P., Deringer, V. L., Gasparotto, P., Csányi, G. ábor & Michaelides, A. An accurate and transferable machine learning potential for carbon. *J. Chem. Phys.* **153**, 034702 (2020).
8. Lot, R., Pellegrini, F., Shaidu, Y. & Küçükbenli, E. Panna: Properties from artificial neural network architectures. *Comput. Phys. Commun.* **256**, 107402, (2020).
9. Shaidu, Y. et al. A systematic approach to generating accurate neural network potentials: the case of carbon. *npj Comput. Mater.* **7**, 52 (2021).
10. Shaidu, Y., Smith, A., Taw, E. & Neaton, J. B. Carbon capture phenomena in metal-organic frameworks with neural network potentials. *PRX Energy* **2**, 023005 (2023).
11. Li, Q-J et al. Development of robust neural-network interatomic potential for molten salt. *Cell Rep. Phys. Sci.* **2**, 100359 (2021).
12. Grimme, S. Semiempirical gga-type density functional constructed with a long-range dispersion correction. *J. Comput. Chem.* **27**, 1787–1799 (2006).
13. Artrith, N., Morawietz, T. & Behler, J. örg High-dimensional neural-network potentials for multicomponent systems: Applications to zinc oxide. *Phys. Rev. B* **83**, 153101, (2011).
14. Morawietz, T., Sharma, V. & Behler, J. örg A neural network potential-energy surface for the water dimer based on environment-dependent atomic energies and charges. *J. Chem. Phys.* **136**, 064103 (2012).
15. Hirshfeld, F. L. Bonded-atom fragments for describing molecular charge densities. *Theor. Chim. Acta* **44**, 129–138 (1977).
16. Ghasemi, S. A., Hofstetter, A., Saha, S. & Goedecker, S. Interatomic potentials for ionic systems with density functional accuracy based on charge densities obtained by a neural network. *Phys. Rev. B* **92**, 045131 (2015).
17. Rappe, A. K. & Goddard, W. A. Charge equilibration for molecular dynamics simulations. *J. Phys. Chem.* **95**, 3358–3363 (1991).
18. Hafizi, R., Ghasemi, S. A., Hashemifar, S. J. & Akbarzadeh, H. A neural-network potential through charge equilibration for ws2: from clusters to sheets. *J. Chem. Phys.* **147**, 234306 (2017).
19. Faraji, S. et al. High accuracy and transferability of a neural network potential through charge equilibration for calcium fluoride. *Phys. Rev. B* **95**, 104105 (2017).
20. Ko, TszWai, Finkler, J. A., Goedecker, S. & Behler, J. örg A fourth-generation high-dimensional neural network potential with accurate electrostatics including non-local charge transfer. *Nat. Commun.* **12**, 398 (2021).
21. Jacobson, L. D. et al. Transferable neural network potential energy surfaces for closed-shell organic molecules: Extension to ions. *J. Chem. Theory Comput.* **18**, 2354–2366 (2022).
22. Unke, O. T. & Meuwly, M. Physnet: A neural network for predicting energies, forces, dipole moments, and partial charges. *J. Chem. Theory Comput.* **15**, 3678–3693 (2019).
23. Khajehpasha, EhsanRahmatizad, Finkler, J. A., Kühne, T. D. & Ghasemi, S. A. Cent2: Improved charge equilibration via neural network technique. *Phys. Rev. B* **105**, 144106, (2022).
24. Staacke, C. G. et al. Kernel charge equilibration: efficient and accurate prediction of molecular dipole moments with a machine-learning enhanced electron density model. *Mach. Learn. Sci. Technol.* **3**, 015032 (2022).
25. Grisafi, A. & Ceriotti, M. Incorporating long-range physics in atomic-scale machine learning. *J. Chem. Phys.* **151**, 204105 (2019).
26. Gao, A. & Remsing, R. C. Self-consistent determination of long-range electrostatics in neural network potentials. *Nat. Commun.* **13**, 1572 (2022).
27. Zhang, L. et al. A deep potential model with long-range electrostatic interactions. *J. Chem. Phys.* **156**, 124107 (2022).
28. Pellegrini, F., Lot, R., Shaidu, Y. & Küçükbenli, E. PANNA 2.0: efficient neural network interatomic potentials and new architectures. *J. Chem. Phys.* **159**, 084117 (2023).
29. Yoo, EunJoo et al. Large reversible li storage of graphene nanosheet families for use in rechargeable lithium ion batteries. *Nano Lett.* **8**, 2277–2282 (2008).
30. Stevens, D. A. & Dahn, J. R. High capacity anode materials for rechargeable sodium ion batteries. *J. Electrochem. Soc.* **147**, 1271 (2000).

31. Yildirim, H., Kinaci, A., Zhao, Zhi-Jian, Chan, MariaK. Y. & Greeley, J. P. First-principles analysis of defect-mediated li adsorption on graphene. *ACS Appl. Mater. Interfaces* **6**, 21141–21150 (2014).
32. Shaidu, Y., Küçükbenli, E. & de Gironcoli, S. Lithium adsorption on graphene at finite temperature. *J. Phys. Chem. C* **122**, 20800–20808 (2018).
33. Fan, X., Zheng, W. T. & Kuo, Jer-Lai Adsorption and diffusion of li on pristine and defective graphene. *ACS Appl. Mater. Interfaces* **4**, 2432–2438 (2012).
34. Fan, X., Zheng, W. T., Kuo, Jer-Lai & Singh, D. J. Adsorption of single li and the formation of small li clusters on graphene for the anode of lithium-ion batteries. *ACS Appl. Mater. Interfaces* **5**, 7793–7797 (2013).
35. Lee, E. & Persson, K. A. Li absorption and intercalation in single layer graphene and few layer graphene by first principles. *Nano Lett.* **12**, 4624–4628 (2012).
36. Leggesse, ErmiasGirma, Chen, Chi-Liang & Jiang, Jyh-Chiang Lithium diffusion in graphene and graphite: Effect of edge morphology. *Carbon* **103**, 209–216 (2016).
37. Liu, M., Kutana, A., Liu, Y. & Yakobson, B. I. First-principles studies of li nucleation on graphene. *J. Phys. Chem. Lett.* **5**, 1225–1229 (2014).
38. Yang, G., Fan, X., Liang, Z., Xu, Q. & Zheng, W. Density functional theory study of li binding to graphene. *RSC Adv.* **6**, 26540–26545 (2016).
39. Okamoto, Y. Density functional theory calculations of lithium adsorption and insertion to defect-free and defective graphene. *J. Phys. Chem. C* **120**, 14009–14014 (2016).
40. Zhou, Liu-Jiang, Hou, Z. F. & Wu, Li-Ming First-principles study of lithium adsorption and diffusion on graphene with point defects. *J. Phys. Chem. C* **116**, 21780–21787 (2012).
41. Frondelius, P., Häkkinen, H. & Honkala, K. Adsorption of small Au clusters on MgO and MgO/Mo: the role of oxygen vacancies and the Mo-support. *New J. Phys.* **9**, 339 (2007).
42. Mammen, N. & Narasimhan, S. Inducing wetting morphologies and increased reactivities of small au clusters on doped oxide supports. *J. Chem. Phys.* **149**, 174701 (2018).
43. Mentel, Ł. Mendeleev—a Python resource for properties of chemical elements, ions and isotopes. https://github.com/lmmentel/mendeleev (2014).
44. Shaidu, Y. Li adsorption on graphene data from "Incorporating long-range electrostatics in neural network potentials via variational charge equilibration from shortsighted ingredients". *Zenodo* https://doi.org/10.5281/zenodo.10574375 (2024).

Acknowledgements

Y.S. work was supported by the European Commission through the MaX Centre of Excellence (grant number 824143). SdG work was supported by the European Commission through the MAX Centre of Excellence for supercomputing applications (grant numbers 10109337 and 824143) and by the Italian MUR, through the Italian National Centre from HPC, Big Data, and Quantum Computing (grant number CN00000013). Computational resources were provided by CINECA.

Author contributions

Y. Shaidu and S. de Gironcoli initiated and planned the project. E. Kucukbenli, F. Pellegrini, and S. de Gironcoli supervised all aspects of the project. Y. Shaidu, R. Lot, F. Pellegrini, and E. Kucukbenli implemented the methodology into the PANNA code. Y. Shaidu performed the NNP training and validation. Y. Shaidu, F. Pellegrini, R. Lot, E. Kucukbenli, and S. de Gironcoli analyzed the results. Y. Shaidu, F. Pellegrini, and S. de Gironcoli led the manuscript writing. All authors contributed to discussions throughout the study and commented on the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41524-024-01225-6.

Correspondence and requests for materials should be addressed to Yusuf Shaidu, Emine Küçükbenli or Stefano de Gironcoli.

Reprints and permissions information is available at http://www.nature.com/reprints

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

© The Author(s) 2024