

# UC Berkeley

## UC Berkeley Previously Published Works

### Title

Incorporating Electronic Information into Machine Learning Potential Energy Surfaces via Approaching the Ground-State Electronic Energy as a Function of Atom-Based Electronic Populations.

### Permalink

<https://escholarship.org/uc/item/6ht4s9m6>

### Journal

Journal of Chemical Theory and Computation, 16(7)

### ISSN

1549-9618

### Authors

Xie, Xiaowei  
Persson, Kristin A  
Small, David W

### Publication Date

2020-07-14

### DOI

10.1021/acs.jctc.0c00217

Peer reviewed

# Incorporating Electronic Information into Machine Learning Potential Energy Surfaces via Approaching the Ground-State Electronic Energy as a Function of Atom-Based Electronic Populations

Xiaowei Xie,<sup>1,2</sup> Kristin A. Persson,<sup>3,2</sup> and David W. Small<sup>1,4</sup>

<sup>1</sup>*Department of Chemistry, University of California,  
Berkeley, California 94720, United States*

<sup>2</sup>*Energy Technologies Area, Lawrence Berkeley National Laboratory,  
Berkeley, California 94720, United States*

<sup>3</sup>*Department of Materials Science and Engineering,  
University of California, Berkeley, California 94720, United States*

<sup>4</sup>*Molecular Graphics and Computation Facility,  
College of Chemistry, University of California,  
Berkeley 94720, California United States\**

(Dated: May 27, 2020)

## Abstract

Machine Learning (ML) approximations to Density Functional Theory (DFT) potential energy surfaces (PESs) are showing great promise for reducing the computational cost of accurate molecular simulations, but at present they are not applicable to varying electronic states, and in particular, they are not well suited for molecular systems in which the local electronic structure is sensitive to the medium to long-range electronic environment. With this issue as the focal point, we present a new Machine Learning approach called “BpopNN” for obtaining efficient approximations to DFT PESs. Conceptually, the methodology is based on approaching the true DFT energy as a function of electron populations on atoms; in practice, this is realized with available density functionals and constrained DFT (CDFT). The new approach creates approximations to this function with neural networks. These approximations thereby incorporate electronic information naturally into a ML approach, and optimizing the model energy with respect to populations allows the electronic terms to self-consistently adapt to the environment, as in DFT. We confirm the effectiveness of this approach with a variety of calculations on  $\text{Li}_n\text{H}_n$  clusters.

## I. INTRODUCTION

DFT computations continue to be the predominant approach to first-principles modeling of chemical systems,<sup>1-3</sup> while empirical force-field based approximations continue to dominate large scale simulations.<sup>4,5</sup> Forming one of the categories of approaches intent on bridging this divide, ML-based methods for approximating first-principles PESs are drawing much attention. These methods largely reduce the computational cost while remaining remarkably near to the accuracy of quantum chemistry calculations, and practical applications in chemistry, physics and materials science have been demonstrated<sup>6-9</sup>. So far, predictions of the electronic energy for neutral, closed-shell molecules (equilibrium or off-equilibrium) have been successfully demonstrated. A ubiquitous theme in such methodologies is the development of translationally, rotationally, and permutationally invariant descriptors based on nuclear positions. Prominent examples include the atom-centered symmetry functions<sup>10,11</sup> of Behler and Parrinello and the smooth overlap of atomic positions (SOAP)<sup>12</sup> by Bartók et al. Some specific ML models that fit the general positions-to-energy characterization are RuNNer,<sup>13</sup> Tensormol,<sup>14-18</sup> SchNet,<sup>19,20</sup> ANI-1,<sup>21-23</sup> AMP,<sup>24</sup> DeepMD,<sup>25</sup> LASP,<sup>26</sup> QML models,<sup>27,28</sup> PhysNet,<sup>29</sup> GAP,<sup>30</sup> models based on atom-density representations by Ceriotti and coworkers,<sup>31-35</sup> the spectral neighbor analysis potential (SNAP),<sup>36</sup> graph kernel methods,<sup>37,38</sup> graph neural networks,<sup>39,40</sup> methods based upon eigenvalues of the Coulomb matrix,<sup>41,42</sup> Bag-of-Bonds,<sup>43</sup> permutation invariant polynomials,<sup>44</sup> and others. Impressive examples of large-scale applications of these methods include simulating proteins,<sup>14</sup> amorphous carbon,<sup>45-47</sup> constructing phase diagrams of amorphous  $\text{Li}_x\text{Si}$ ,<sup>48</sup> etc.

However, descriptors based on atom positions alone exhibit some important limitations. For example, they cannot be applied to more than one overall charge or spin state. Furthermore, far-away changes in the chemical environment may induce changes in the local charge or spin state of a region of a molecule even if the local geometry is not changed appreciably. We illustrate this with an example in section III B 3 below. This general problem applies to many molecules and materials. Examples relating to conjugation, like the emergence of polyradicality with increasing length in acenes,<sup>49-51</sup> armchair versus zigzag edge effects in graphene nanoribbons,<sup>52-54</sup> and the gradual ascension to the effective conjugation length in a variety of oligomers,<sup>55-57</sup> come to mind. Even more marked are examples involving explicit ionization and charge transfer (CT). These include outer-sphere electron transfer processes

such as those occurring in Ferredoxin protein cores and other proteins<sup>58,59</sup> and large-scale organic donor-acceptor complexes with appealing electronic properties.<sup>60-62</sup>

We stress the need for advanced ML-based methods to describe complicated heterogeneous systems such as found in electrode surfaces like the solid electrolyte interface (SEI) in lithium-ion batteries (LIB).<sup>63-67</sup> As a small glance into the complexities involved here, consider that the ground state of neutral lithium ethylene carbonate (LiEC), a key intermediate in the reductive decomposition process to form the LIB SEI, varies in character between neutral and CT, i.e. unpaired electron on Li or on the rest of the molecule. This variation is subtly dependent on where the Li atom locates around the ethylene carbonate (EC) molecule;<sup>68</sup> the transition between the two characters can be fairly rapid, and so for position-based descriptors, whose terms don't change rapidly in such transitions, this situation is clearly a practical difficulty. This is just one part of the SEI's multiplex of charged and uncharged species and the reactions between them, underscoring significant challenges for ML descriptors.

In principle, these general effects can be captured by a position based descriptor along with the total charge and spin multiplicity of the molecule, as after all these are the only inputs needed for a ground-state calculation. However, size extensibility, the crucial property that a ML model may be uniformly and consistently applied to systems of different sizes, favors the use of local, typically atom-centered, descriptors, which are difficult to reconcile with global parameters like total charge and spin. And, stretching these to include the (possibly very) long range information required for these cases is challenging in practice in terms of computational efficiency and numerical stability.

The above issues stem from a lack of explicit electronic information in the descriptors, which suggests a general remedial approach. To successfully incorporate electronic information, the basic scheme of the parent DFT model should be followed such that electronic terms can adapt according to changes in neighboring electronic information, preferably in a self-consistent or equilibrated way.

In this context, we would like to briefly reassess some of the models mentioned above. The SOAP based descriptors make use of smeared-atom proxy densities, and together these bear a rough resemblance to electron density. A recent variant<sup>33</sup> uses the electrostatic potential (ESP) associated with such proxy densities as the basis for the SOAP expansion. This approach provides enhanced long-range information, and we will comment more on

this important development at the end of this paper. Nevertheless, these proxy densities are fixed and do not adapt to the electronic environment.

Several of the above ML approaches include a separate long-range, pairwise electrostatic contribution to the energy. This inclusion has been shown to be very useful for describing large  $Zn_nO_n$  clusters,<sup>69</sup> water clusters,<sup>14,70–72</sup> proteins,<sup>14</sup> and other examples<sup>73,74</sup>. Popelier and coworkers first proposed to employ NNs to construct environment dependent multipoles.<sup>75</sup> Artrith et. al.,<sup>69</sup> and later on Yao et. al.<sup>14</sup> adopted a scheme where they pretrain a neural network to predict atomic point charges, using the same descriptors (i.e. symmetry functions) as they used to train energies. These charges were subsequently used to compute the electrostatic term through Ewald summation<sup>76</sup> for periodic systems, or Coulomb’s law for molecules. Another approach fits molecular dipoles to infer partial charges.<sup>77</sup> We note that there are also independent ML models specialized for predicting partial charges,<sup>78,79</sup> multipole moments,<sup>32,80,81</sup> and even the full electron density.<sup>31,82</sup> The latter method has been used e.g. to guide the inverse design of chemical materials based on electron density information, to serve as an initial guess for SCF convergence, and more. The above-mentioned methods all predict the electronic terms directly from local position-based descriptors, and thus are not properly adaptable to the electronic environment.

In another direction, the various charge equilibration schemes, such as EEM<sup>83</sup> and QEq<sup>84</sup> and several newer examples,<sup>85–91</sup> compute atomic charges self-consistently. These models therefore exhibit some basic flexibility in the context of the above issues. However, they also employ very simple forms; this is advantageous for application to very large molecular systems, but it limits the accuracy as compared to the above ML-based models. Ghasemi, et al<sup>92</sup> developed a charge equilibration scheme where the intra-atomic energy function comes from a NN that is parametrized by DFT data, and they used this to predict the energies for both neutral and ionized NaCl clusters. To our knowledge, this important development is the only extant positions-to-energy model that employs a high-level ML approximation and that is also able to accommodate varying electronic structure. However, the model is exclusively trained on self-consistent DFT energies and its charges do not enter the descriptor. Hence the charges are effectively auxiliary parameters as opposed to being more directly associated with the electron density. This precludes, for example, obtaining more than one SCF solution at a given geometry.

An open question remains as to how much electronic information should be incorporated.

At one extreme, it is possible to utilize the full electron density and create ML models that map this directly to the energy. Some of these efforts are concerned with finding new DFT approaches with unprecedented accuracy.<sup>93-96</sup> Others pertain more to the fact that commonly used functionals do not provide a direct non-iterative link between density and energy, so a ML approximation for this, as with other orbital-free approaches, can greatly reduce the associated burden.<sup>20,96-101</sup> Effective self-consistency is then obtained by optimizing the energy with respect to the density. This level of electronic information is ideal in terms of accuracy, and it easily addresses the above issues, but it entails a significant increase in computational requirements compared to the above models.

In the present paper, we attempt to identify the simplest amount of electronic information that is sufficient to solve the above issues, and how to incorporate it effectively. In a sense, we are seeking a good balance between the above position-only approaches and the full-density based approach. We focus on approximating regular DFT functionals, and we explore the incorporation of atomic electron populations and associated electrostatic interaction terms into the descriptor. Effectively, this divides the global charge and spin parameters into local terms that may be naturally incorporated into local descriptors. In the following sections, we exploit CDFT to realize a natural map from intrinsic populations (self-consistently optimal ones or otherwise) to energies, which may be approximated with neural networks, leading to a simple self-consistent approach that is qualitatively true to the parent DFT functional.

## II. THEORY

### A. Target Energy Function

#### 1. Formal viewpoint

In the Levy constrained search approach in DFT,<sup>102</sup> each density  $\rho$  maps to the lowest possible energy  $E[\rho]$  obtainable from a many-electron wave function associated with that density. Thinking of the density as being a reduction of the many-electron density matrix, the constrained-search idea may be generalized to any collection of reduced variables  $\{v_i\}$  (which may be functions, etc.):

$$E[\{v_i\}] = \min_{\Psi \rightarrow \{v_i\}} \langle \Psi | \mathbf{H} | \Psi \rangle, \quad (1)$$

where  $\mathbf{H}$  is the many-electron Hamiltonian. Note that both  $E$  and  $\mathbf{H}$  are dependent on the types of nuclei and their positions, which, for simplicity, have been omitted in this equation. Of course, in practice, the search over wave functions and the energy expectation value on the right hand side of eqn. (1) will be replaced by a search over orbitals and an approximate functional, respectively.

For  $\{v_i\}$  in this paper, we are primarily concerned with reducing the density to atom populations, in particular Becke populations.<sup>103</sup> For these, the density is multiplied by a weight function and integrated to produce a population value. For atom  $i$  and spin  $\sigma$  we have

$$p_{i,\sigma} = \int w_i(\mathbf{r}) \rho_\sigma(\mathbf{r}). \quad (2)$$

The weight functions  $w_i$  are localized to their respective atoms, they take on values between 0 and 1, maximizing on the pertinent atom, and they sum over all atoms to 1.

Our target function for machine learning is essentially  $E[\mathbf{p}]$ , where we have placed the populations into a vector  $\mathbf{p}$ . Technically, this function is defined only for population values that are non-negative and sum to appropriate values for the numbers of electrons of each spin, or equivalently, the total charge and spin multiplicity. We can extend this function to the larger domain of non-negative population vectors with no sum conditions by first applying  $P_{N_\alpha, N_\beta}$ , which “projects” the population vector to a vector obeying the sum conditions:

$$[P_{N_\alpha, N_\beta}(\mathbf{p})]_{i,\sigma} = p_{i,\sigma} + \frac{N_\sigma - \sum_j p_{j,\sigma}}{N_{\text{atom}}}, \quad (3)$$

where  $N_\alpha$  and  $N_\beta$  are the numbers of  $\alpha$  and  $\beta$  electrons, respectively, and  $N_{\text{atom}}$  is the number of atoms. Hence, our formal target function is  $E[P_{N_\alpha, N_\beta}(\mathbf{p})]$ .

The target function may be optimized with respect to the population values, with the latter considered as input variables. This is much the same as optimizing DFT energies with respect to the density or orbitals. Optimizing with respect to the populations thus defines an SCF procedure. As to the issue of representability here, we only need to ensure that the populations (after projection) are non-negative.



At this point, we need a way to obtain practical approximations to this target function, which is the subject of the next subsection.

## 2. *The target function in practice: CDFT*

The above ideas may be applied to any of the functionals used in common practice, e.g. B3LYP as we use below. That is, they each may be formally reduced to a functional of populations (and nuclear positions and types, as noted above).

Training data for the target function must be obtained with non-standard DFT calculations. Fully self-consistent DFT solutions provide data for certain population choices, however, by itself this would constitute a very limited training set. Energies for other population choices may be obtained by using constrained DFT (CDFT),<sup>104–106</sup> which uses Lagrange multipliers to optimize the DFT energy under various constraints. The use of constraints on Becke-weight based populations has been implemented in Q-Chem,<sup>107</sup> and allows us to, in principle, obtain the energy for any given set of populations. Much as how there are generally numerous wave functions (and thereby energies) associated with a given density, there will overall be many energies associated with a specific population vector. In fact, the energy spread is much wider for populations than for the density, because there will generally be many densities associated with a given population vector. As with standard DFT calculations, care must be exercised to attempt to obtain the lowest energy CDFT solution.

The accuracy of ML models of the DFT energy is significantly enhanced by incorporating energy derivatives into the loss function, i.e. it is desirable to train the model towards both the direct value of the target function and its first derivatives. The  $xyz$  position derivatives for CDFT have been previously derived.<sup>108</sup> For this present paper, we also need a way to compute the derivatives of the CDFT energy with respect to populations. This is described in the Appendix.

## **B. Model Energy Function**

We adopted a widely accepted scheme in the ML PES field, where the total energy is decomposed into atom-centered contributions to ensure transferability across systems with

different sizes and compositions. A neural network (NN) framework was used to map an atomistic descriptor  $D_i[\mathbf{Z}, \mathbf{p}, \mathbf{r}]$  to the atomic energy. The model energy function can be written as

$$E_{\text{ML}} = \sum_{\mu} \sum_{i \in A_{\mu}} \epsilon_{\mu}[D_i[\mathbf{Z}, \mathbf{p}, \mathbf{r}]] + E_{\text{intra}}[\mathbf{Z}, \mathbf{p}, \mathbf{r}] + E_{\text{el}}[\mathbf{Z}, \mathbf{p}, \mathbf{r}], \quad (4)$$

where  $\epsilon_{\mu}$  represents the output from the atomic neural network for the element  $\mu$ , and this function is applied to the descriptor for atom  $i$ ;  $A_{\mu}$  is the set of atom indices corresponding to element  $\mu$ ;  $\mathbf{Z}$  and  $\mathbf{r}$  contain the atomic numbers and coordinates, respectively, for the entire molecule.  $E_{\text{intra}}$  and  $E_{\text{el}}$  are simple intra-atomic and pairwise electrostatic terms, respectively, and will be described below. An overview of our model, which we will call ‘‘BpopNN’’ for ‘‘Becke Population Neural Network’’, is shown in Figure 1. More details are given in the following subsections.

### 1. Atomistic Descriptor

The descriptor we use in this paper is essentially a variant of SOAP,<sup>12</sup> modified to describe the electronic as well as nuclear environments. The regular SOAP descriptor is based on an atomic power spectrum of a proxy density that represents the nuclei: for each atom, the latter is expanded in a set of radial basis functions (RBF) and spherical harmonics that are symmetric about that atom, and the coefficients of this expansion directly produce the spectrum. In our case, we want to adapt the proxy function to represent the varying electrostatic environment.

The basic idea is to use the ESPs of the nuclei, and the ESPs of the  $\alpha$  populations treated as point charges, and likewise for the  $\beta$  populations. In fact, as in SOAP, we partition the proxy function into separate contributions for each element type, since this makes for a more refined descriptor. In our lithium hydride applications below, we simplified the descriptor to be based only on the nuclear charges and the total electron population on each atom. In the present subsection, we will continue the development using spin populations. This is more general, and reducing this descriptor to the total-population one is straightforward.

To simplify the computation of the descriptor elements, we approximate the  $1/r$  term in the ESPs with a single Gaussian function. This amounts to making a local approximation to a function with a significant long-range component. Incorporation of the latter into the descriptor is expected to be of significant benefit,<sup>33</sup> and we intend to explore this in future

work (see Conclusions below). Our proxy functions take the form

$$\rho_{\mu,t}(\mathbf{r}) = \sum_{i \in A_\mu} \eta_i^t B \exp(-\gamma |\mathbf{r} - \mathbf{r}_i|^2), \quad (5)$$

where  $t$  is either  $\nu$  (for nuclear),  $\alpha$ , or  $\beta$  corresponding to the type of point population, and thus  $\eta_i^\nu = Z_i$ ,  $\eta_i^\alpha = p_{i,\alpha}$ , and  $\eta_i^\beta = p_{i,\beta}$ . Here we used  $B = 1.128$  and  $\gamma = 0.171$ , and  $r$  is in the unit Bohr radius. Note that because these coefficients are chosen to approximate  $1/r$ , this Gaussian function is not normalized.

To build the descriptor for any atom in the molecule, we set the origin to that position and project the radial slices of  $\rho_{\mu,t}(\mathbf{r})$  onto the pertinent set of spherical harmonics  $\mathbf{Y}_{lm}(\hat{\mathbf{r}})$ :

$$\rho_{\mu,t}(\mathbf{r}) = \sum_{i \in A_\mu} \sum_{lm} c_{ilm}^t(r) \mathbf{Y}_{lm}(\hat{\mathbf{r}}), \quad (6)$$

where  $\hat{\mathbf{r}}$  denotes the unit vector corresponding to  $\mathbf{r}$ . This is done separately for each proxy function, i.e. for each element type  $\mu$ , so that the given atom's descriptor contains separate projection coefficients for each  $\mu$ . Note that the  $i$  sum includes the central atom (if it is of the element type  $\mu$ ). In practice, the  $l$  sum must be limited to a range. In the applications below, we use a maximum  $l$  value of 6. Since  $\eta_i^t$  does not depend directly on positions, the analytical form of the radial terms can be derived following previous works<sup>12,109</sup> as

$$c_{ilm}^t(r) \equiv 4\pi\eta_i^t B \exp[-\gamma(r^2 + r_i^2)] \iota_l(2\gamma r r_i) \mathbf{Y}_{lm}^*(\hat{\mathbf{r}}_i), \quad (7)$$

where  $*$  denotes complex conjugation and  $\iota_l$  are the modified spherical Bessel functions of the first kind. Summing over pertinent atoms gives

$$c_{lm}^{\mu,t}(r) = \sum_{i \in A_\mu} c_{ilm}^t(r), \quad (8)$$

and projecting these onto RBFs gives a sequence of expansion coefficients  $c_{nlm}^{\mu,t}$  with an extra index  $n$ . The projections are based on evaluation of the integrals  $\langle g_n \mathbf{Y}_{lm} | \rho \rangle$ , with  $g_n$  denoting an RBF. In general, this can be done by either analytical or numerical integration (e.g. Gauss-Legendre quadrature), but in our case, we bypass these complications by simply evaluating  $c_{lm}^{\mu,t}(r)$  on a grid of  $r$  values. So for us,  $c_{nlm}^{\mu,t} = c_{lm}^{\mu,t}(r_n)$ , with  $\{r_n\}$  being the set of grid points. The resulting coefficients are smooth with respect to nuclear positions, and technically, one could choose RBFs whose projection coefficients match the evaluated grid values. Hence the grid spacings can in principle be selected to produce a result that is

as satisfactory as the explicit RBF/integration approach. A simple evenly-spaced 32-point grid between 0.5 and 13.2 Å was used in this work; further refinement of the grid will be addressed in the future.

At this point, one can construct the power spectrum by contracting over  $m$ ,

$$p_{nl}^{\mu,t} = \sum_{m=-l}^l |c_{nlm}^{\mu,t}|^2 \quad (9)$$

It is easy to prove the rotational invariance of  $p_{nl}^{\mu,t}$ , as 3D rotations correspond to transformations of blocks of the  $c_{nlm}^{\mu,t}$  by Wigner matrices, which are unitary. Note that the  $m$  contractions entail some loss of angular information, particularly for larger  $l$  values. There exist various approaches for dealing with this problem, such as using the bispectrum,<sup>12</sup> and we intend to consider such enhancements in future work.

For each atom in the molecule, we employ a separate power spectrum for each element-type in the molecule, i.e. the power spectrum is split into different element channels. There is active research aimed at reducing the cost associated with adding more atomic species,<sup>18,28,35,110</sup> which is also of future interest for us. For now, the above treatment is sufficient to prove the principle.

A radial cutoff of 13.2 Å was used below in computing the descriptor, i.e. when the proxy function is expanded relative to a given atom, the summation found in (this instance of) eqn. (5) excludes atoms beyond this cutoff. The motivation for choosing this relatively long cutoff is that, as described below, in addition to the NN energies, we include a simple long-range pairwise electrostatic term, as has been shown to be very useful in previous papers.<sup>14,69</sup> The influence of the switch in dominance between short-range NN and explicit electrostatic terms is a general issue, but because this question is separate from the main theme of this work, we chose the long cutoff to minimize this influence.

As shown in the Supporting Information (SI), the use of shorter cutoff lengths, both for the one associated with eqn. (5) and the one defining the last radial grid point, appears to lead to overall worse performance for the test data described below. This might be connected to a subtle interplay between the various model parameters employed, such as the cutoffs and the maximum  $l$  value, and aspects of the training set such as its small size (see below). We would also like to note that it may have more physical origins, such as being connected to the observation of a longer intrinsic correlation length in  $\text{Li}_n\text{H}_n$  clusters than for some other ionic systems.<sup>111</sup>

## 2. Neural Network Architecture

We employ separate neural networks for different element types, and for the applications below, each element network contains two hidden layers with 30 neurons in each layer. This size of neural network, which is fairly shallow, was used to best accommodate the number of training data and prevent overfitting. Nonlinearities were introduced using a modified softplus activation function  $\ln(1 + \exp(100x))/100$ . This form resembles the RELU activation function, which itself is highly popular because it overcomes the vanishing gradient problem,<sup>112</sup> yet this softplus variant has the advantage of being twice continuously differentiable. Although the vanishing gradient problem is expected to be more significant for high NN depth and sigmoid (in the general sense) activation functions, we found that the latter produced higher errors for our data than the softplus variant.

In previous works, the ML model is usually trained towards the DFT atomization energy, as opposed to total energy. This has been shown to accelerate and balance training.<sup>14</sup> The definition of atomization energy is ambiguous in the case of charged molecules, hence we select a quadratic function of atomic charge that best fits the DFT energies for the pertinent charge states of the isolated atom.  $E_{\text{intra}}[\mathbf{Z}, \mathbf{p}, \mathbf{r}]$  is then the sum of these energies for each atom. For lithium and hydrogen atoms found in the molecular applications below, we used the following quadratic functions as baseline for atomic energies

$$E_{\text{intra,Li},i} = 0.0901q_i^2 + 0.1088q_i - 7.4787 \quad (10)$$

$$E_{\text{intra,H},i} = 0.2471q_i^2 + 0.2403q_i - 0.4805, \quad (11)$$

where  $q_i$  is the atomic partial charge for atom  $i$ , i.e.  $q_i = Z_i - p_{i,\alpha} - p_{i,\beta}$ . This resembles the intra-atomic term in charge equilibration models.<sup>92</sup> The quadratic form of these intra-atomic terms contrasts with the fact that the true energy has a piecewise linear dependence on the number of electrons. This is expected to lead to issues related to charge-transfer error, such as non-integer populations at atom dissociation in certain cases, which is also typical for common density functionals. But this is intentional: the density functional used below in our target energy function, B3LYP, generally exhibits significant charge transfer error, so we accept that our model in this case will also exhibit this problem. As density functionals with reduced charge-transfer error continue to be developed, the form of the intra-atomic energy function will need to be modified accordingly.

We also used a simple pairwise Coulomb term to incorporate long-range electrostatics, as shown in the following equation.

$$E_{\text{el}}[\mathbf{Z}, \mathbf{p}, \mathbf{r}] = \sum_{j>i} \tanh(\kappa r_{ij}) \frac{q_i q_j}{r_{ij}} \quad (12)$$

Parameters  $\kappa$  (one each for different element pair types) were trained together with other parameters in the neural network.

The loss function for the training is

$$\mathcal{L} = \sum_M \frac{(E_M^{\text{CDFT}} - E_M^{\text{ML}})^2}{N_{\text{atom},M}} + \gamma_1 \sum_M \frac{\|\mathbf{F}_M^{\text{CDFT}} - \mathbf{F}_M^{\text{ML}}\|^2}{N_{\text{atom},M}} + \gamma_2 \sum_M \frac{\|\mathbf{f}_M^{\text{CDFT}} - \mathbf{f}_M^{\text{ML}}\|^2}{N_{\text{atom},M}}. \quad (13)$$

We use the Euclidean norm here, and for the model presented below, both  $\gamma_1$  and  $\gamma_2$  were set to 1.5. The above sums go over the training molecules, and  $\mathbf{F}$  is the position gradient,  $\mathbf{f}$  is the population gradient, and  $N_{\text{atom},M}$  represents the number of atoms in molecule  $M$ . Adaptive moment solver (Adam)<sup>113</sup> was used to update the NN weights during training. Our implementation is built up from the open-source package TensorMol,<sup>14</sup> which takes advantage of the automatic differentiation scheme of Tensorflow.<sup>114</sup>

### III. RESULTS AND DISCUSSION

We here test the performance of our method on  $\text{Li}_n\text{H}_n$  clusters of varying size and geometry, and overall charges of +1, 0, and -1. We chose this for the simplicity of having only two element types. Lithium hydride clusters have been studied in several theoretical<sup>115–120</sup> and experimental<sup>121,122</sup> works, and have implications in various applications including hydrogen storage.<sup>123–127</sup>

In the calculations below, we employ the descriptor described above, but, as mentioned above, we modified it so that the total electron population on each atom, rather than the 2 spin populations, is used for the electronic terms.

#### A. Training-data generation and model training

In the training dataset, the cluster sizes vary from from  $n = 1$  to 24, with increments of 3 or 4. We use two basic types of structures from which we obtain more samples: (1) collections of relatively widely separated LiH molecules (2-5 Å between molecules), and (2)

denser  $\text{Li}_n\text{H}_n$  systems with roughly cubic structures with alternating Li and H. We denote the first dataset as “sparse”, and the second as “non-sparse” from here on.

For the sparse data, structures from DFT geometry-optimization trajectories were used in the training dataset. For the non-sparse data, structures both from geometry-optimization trajectories and Ab Initio Molecular Dynamics (AIMD)<sup>128</sup> trajectories were used. The AIMD simulations used a temperature of 500 K and a time step of 20 a.u. ( $\approx 0.5$  fs).

All structures were generated using overall charge 0, and calculations for the ionized molecules used the same geometries. During the geometry optimizations, the LiH units in the sparse structures cluster together in various ways and become less sparse, although they remain less dense than the non-sparse data. All of these geometries exhibit alternating Li and H atoms, i.e. no structures with Li clusters nor  $\text{H}_2$  molecules were obtained. This is because Li clusters often entail spin polarization, and therefore we reserve the generation of such structures for future work in which spin populations are used. More details of the geometries used in the training dataset are included in the SI.

All DFT computations were conducted using the QChem<sup>107</sup> program. The geometry-optimization trajectories for the sparse data used the B3LYP<sup>129,130</sup> functional with D3(BJ) dispersion<sup>131</sup>, the AIMD trajectories for the non-sparse data used the  $\omega$ B97X-D<sup>132</sup> functional, and the geometry-optimization trajectories for the non-sparse data used B3LYP (without dispersion). The reason for the overall use of different functionals is that we began with  $\omega$ B97X-D, then switched to using B3LYP-D3(BJ) for computational efficiency, and then encountered some implementational problems with D3(BJ) for Li-Li interactions, so we finally switched to using B3LYP alone. These variations only occurred for the generation of structures; for all geometries used in the training and test sets (the latter is described below), we (re-)computed the DFT energy using B3LYP. All DFT calculations in this paper used the def2-SVPD basis set.<sup>133</sup>

For the final training set, we selected one per every several geometries from each trajectory, with an interval of 2 to 7 depending on the trajectory’s length. On each of these geometries, we computed unconstrained B3LYP solutions for the considered overall charges (+1,0,-1), and for each of these three, we performed 10 separate CDFT calculations. For each of these, we added a random (uniform distribution from -0.05 to 0.05) charge fluctuation to each atom’s charge value from the pertinent unconstrained SCF solution, and then computed the associated CDFT energy and its derivatives with respect to atomic positions

and atomic charges (see Appendix for the latter).

The training data ( $\sim 37000$  points) were randomly separated into training (95%) and validation subsets (5%). This was influenced by the general strategy of using the validation set primarily for tuning hyperparameters and for determining when to stop the training; sometimes this includes additional training phases that incorporate the validation subset back into the training set. Because we did not systematically explore hyperparameter tuning, we sought to increase the portion of data in the training set at the outset, hence our choice of a 95/5 split, in contrast to the more typical 80/20.

For the random data partitioning, all data points were approached equally, so because there were 30 calculations per geometry (10 charge perturbations on 3 charge states), it is more than likely that each geometry is found in the training subset. Additional testing on geometries not found in the training set is presented in the next subsection.

The validation error provides an estimate on how well the model is trained. Typically the model reaches its best performance on test sets after  $\sim 500$  epochs of training. The mean absolute error (MAE) on the independent validation set of the energy is 1.523 kcal/mol per molecule.

## B. Testing the model

### 1. $Li_{29}H_{29}$ and $Li_{32}H_{32}$ clusters

We generated two test sets with larger molecules to test the transferability of our model: non-sparse  $Li_{32}H_{32}$  (+1,0,-1 charge) from an AIMD trajectory and sparse  $Li_{29}H_{29}$  (+1,0,-1 charge) from a geometry-optimization trajectory. For those test sets, CDFT calculations for off-equilibrium charges were not conducted as we focus on comparing the self-consistent solutions from DFT and the BpopNN model.

As discussed earlier, one advantage of our model is that the populations (i.e. partial charges at present) can be optimized self-consistently. In other words, the ML-model energies are optimized with respect to the partial charges. This process is reminiscent of the SCF procedure in DFT calculations, and we denote this charge optimization procedure as “SCF-q” in the remaining text. We found the conjugate gradient (CG) method to be effective in optimizing the BpopNN energy, and this was used in all examples below. This procedure



requires computation of the BpopNN gradient with respect to the charges at each iteration, and repeatedly updates the charges until the gradient is sufficiently close to 0.

In general, both DFT SCF and SCF-q can produce more than one stationary point, i.e. solution. Since our objective is to find counterpart solutions between DFT and ML and then compare them, we used the actual DFT partial charges as initial guesses for SCF-q calculations for each geometry in the  $\text{Li}_{29}\text{H}_{29}$  and  $\text{Li}_{32}\text{H}_{32}$  test sets. Other initial guesses can be used but this increases the odds of obtaining an SCF-q solution that does not correspond to the SCF one even when a pair of corresponding solutions exists. Of course, general and effective initial guess schemes will eventually need to be developed for the partial charges. The Li partial charges in our systems usually exhibit values between 0.5 and 0.8 (although the actual range covered throughout the whole dataset is  $-0.31$  to  $0.99$ ). It turned out that we obtained the same SCF-q solutions when starting with uniform guesses of 0.4 for Li charges and  $-0.4$  for H charges as when starting from DFT partial charges, demonstrating the model’s capability of dealing with general reasonable initial guesses for partial charges. Subtleties regarding the model SCF procedure, such as the possibility for multiple solutions, or the possibility of structures with no solution, will be important to consider in future work. We suggest here that the intra-atomic energy component should help to secure the existence of solutions generally, and we note that for all structures considered below, the CG method was able to produce a SCF-q solution.

In Figure 2, we plot a comparison between the SCF-q BpopNN energies (converged to threshold  $10^{-7}$  Hartree/e) and the DFT energies for neutral and ionized molecules. The statistics for energy and charge errors are shown in Table I. For the sparse  $\text{Li}_{29}\text{H}_{29}$  test set (Figure 2 (a)), the BpopNN energies are showing close agreement with the DFT reference energies, with a charge MAE of  $\sim 0.014$  per atom, and an energy MAE of  $\sim 0.20$  kcal/mol per atom. Non-sparse  $\text{Li}_{32}\text{H}_{32}$  (Figure 2 (b)) provides a harder test set - there are more neighbors around each center atom and the Li-H bonds are actively forming and breaking in the AIMD trajectory, which makes it more challenging for the descriptor. We still observe a charge MAE of  $\sim 0.011$  and energy MAE of  $\sim 0.27$  kcal/mol per atom for those cases. There is no obvious deterioration in performance on ionized molecules compared with neutral molecules.

Figure 3a compares the partial charges obtained from DFT and the BpopNN SCF-q procedure for a random test neutral  $\text{Li}_{32}\text{H}_{32}$  structure; Figure 3b depicts the absolute

partial charge error for the same molecule. There is a good agreement between the two methods.

In addition to the SCF-q procedure, geometries and partial charges can be optimized simultaneously to a tight convergence ( $10^{-7}$  Hartree/Å for position gradients and  $10^{-7}$  Hartree/e for population gradients) using a CG algorithm for both neutral and ionized molecules. We adopted the geometry-optimization procedure that DFT uses, i.e. converging SCF-q at each geometry cycle. In theory there is no need to follow this order, but this method turned out to be more stable and efficient in practice. The cubic  $\text{Li}_{32}\text{H}_{32}$  structures were optimized using BpopNN separately for cation, neutral, anion, and compared with the DFT optimized geometries. Figure 4 shows the superposition of BpopNN and DFT optimized neutral  $\text{Li}_{32}\text{H}_{32}$  structures. The equilibrium geometries were successfully predicted with an root mean square displacement (RMSD) of 0.0522 Å, 0.0958 Å, 0.0516 Å for neutral, cation, and anion, respectively, compared to the true DFT geometries. The superpositions and RMSD values were obtained with Maestro.<sup>134</sup> The final absolute DFT and BpopNN energies are compared in Table II, along with the DFT and BpopNN predicted adiabatic ionization potential (IP) and electron affinity (EA) values. The predicted IP and EA are showing qualitative agreement with DFT (0.167 eV and 0.278 eV error for IP and EA respectively), although they are not highly accurate. These errors are fairly reasonable considering the size of the trial molecule (64 atoms) we are testing on, and the modest dataset size and NN depth associated with our trained model.

To make a preliminary comparison to existing ML methods, we used our training set and the QUIP software package<sup>135</sup> to obtain a few variations of Gaussian Approximation Potential (GAP) models,<sup>30</sup> differing e.g. in the cutoffs and auxiliary potentials employed (see the SI). It is necessary to train separate GAP models for cation, anion, and neutral. For this, only our training data for equilibrium (i.e. regular SCF) energies are applicable, which amounts to  $\sim 1100$  data points (for each overall charge). For neutral and anion, the GAP energy errors for the sparse ( $\text{Li}_{29}\text{H}_{29}$ ) and non-sparse ( $\text{Li}_{32}\text{H}_{32}$ ) geometries are comparable to those of our model, while for cation, the GAP energy errors are significantly larger. As to the latter result, we note that the cation-neutral energy gaps (from DFT or ML) are generally significantly larger than the corresponding neutral-anion gaps, and as shown in Fig. S2 in the SI, the distribution of DFT cation energies in the training set is noticeably less uniform than in the other cases. The GAP optimized geometries for the 3 charge states

of  $\text{Li}_{32}\text{H}_{32}$  are less accurate than those of BpopNN in both energy and structure. Details for the above results are given in the SI.

## 2. Reaction Pathways

As another practical application of our model, we sketched the energy landscape of various structural transformations for neutral and anionic  $\text{Li}_{12}\text{H}_{12}$ . Diverse stationary-point structures were found for this system, ranging from cuboid structures to hexagonal structures to structures with fused 6-4-6 membered rings, etc. Although some motifs can be seen in a couple of optimized geometries in the training, these structures are overall distinct from the structures found in the training set. This provides an assessment of how well the model can perform in mapping out energetics for realistic reaction pathways. The resulting energy profile for neutral  $\text{Li}_{12}\text{H}_{12}$  structures is shown in Figure 5 and for anion  $\text{Li}_{12}\text{H}_{12}$  in Figure 6. All structures shown were individually optimized, e.g. the DFT neutral and anionic structures are distinct.

The energies of **Struct 1** from DFT and BpopNN were separately used as baselines, so that only the energy difference between structures is shown. The difference in energy between baseline structures for neutral relative to anion is 9.17 kcal/mol and 10.20 kcal/mol, respectively for DFT and BpopNN. All the minimum structures and the transition states (TSs) connecting them were obtained independently for DFT and BpopNN. Newton’s method was used in the BpopNN to converge the TSs (to geometry gradient threshold  $10^{-7}$  Hartree/Å and charge gradient threshold  $10^{-7}$  Hartree/e) and the presence of one imaginary frequency was confirmed for all cases. The pathways from TSs to minimum structures in BpopNN were confirmed by a steepest descent algorithm with a small step size. All the DFT TSs were confirmed with one imaginary mode as well and pathways confirmed by Intrinsic Reaction Coordinate (IRC) calculations. In general, the BpopNN geometries show close agreement with DFT geometries (RMSD values are shown in the table inside each figure). Energies for equilibrium structures are qualitatively good, although some deviations (2 to 3 kcal/mol) can be seen for **Struct 7** and anion **Struct 3**. Reaction barriers are showing qualitative agreement as well, although there are cases with noticeable quantitative error (neutral **Struct 3**  $\leftrightarrow$  **TS 3**  $\leftrightarrow$  **Struct 4**, neutral **Struct 3**  $\leftrightarrow$  **TS 6**  $\leftrightarrow$  **Struct 7**, anion **Struct 2**  $\leftrightarrow$  **TS 2**  $\leftrightarrow$  **Struct 3**). Some caution needs to be exercised if one were to predict the lowest en-

ergy structure from ML models, especially when the energy range is extremely small (See anion **Struct 1** vs **Struct 7**). At the very least, however, it is helpful to take advantage of the computational speed of the NN and map out the whole transformation picture prior to refinement with full electronic structure calculations.

### 3. Long-range environmental effects

Finally, we would like to demonstrate the importance of optimizing the electronic components self-consistently. Namely, we consider a model system that exhibits significant long-range charge effects. Consider a neutral  $\text{Li}_{15}\text{H}_{15}$  cluster, which is the side-by-side interaction of two sheets, one  $\text{Li}_8\text{H}_7$  (“A”) and one  $\text{Li}_7\text{H}_8$  (“B”) as shown in Figure 7. The two sheets are separated by 6.52 Å (“AB(6.5)”) or 14.02 Å (“AB(14)”). The first distance is beyond the cutoff used by other common local position-only descriptors and the second one is even beyond the long cutoff distance used in this work. For B3LYP, both inter-sheet distances exhibit nearly full charge separation: the left sheet (A) bearing +1 charge and the right sheet (B) bearing -1 charge. We note that a ML model based merely on local position-only descriptors will not succeed for this system. In that case, for AB(14), monomer A doesn’t “see” monomer B, so its predicted energy would mirror that for an isolated *neutral* A monomer, and likewise for B. The predicted atom charges and the intra-monomer energies would be incorrect, and the inter-monomer energy, which is largely electrostatic, would nearly be missed entirely.

As shown in Table III, this problem can be solved qualitatively by incorporating self-consistent electronic information: our model can predict the charge separation and qualitatively correct binding energies, although there were no direct scenarios of this kind in our training data.

At extremely large A-B separation, AB(1000), both DFT and BpopNN exhibit a partially ionized state. For DFT, this stems from the well-known “self interaction” error (also called “delocalization error”), and this kind of result is also expected of BpopNN. For AB(6.5) and AB(14), BpopNN shows increased A-B charge separation relative to infinite separation. The principal driving force for this, and the only one for AB(14), is the electrostatic attraction between  $\text{A}^+$  and  $\text{B}^-$ . In other words, the long-range environment is the sole reason for increased charge separation in BpopNN in this case, which explicitly establishes that the

model is adaptable to the environment.

The sum of BpopNN energies for  $A^+$  and  $B^-$  is -121.4718 a.u., which actually is slightly above the sum of BpopNN energies for  $A$  and  $B$ , -121.4785 a.u. The situation is opposite for DFT, which may partially explain the increased A-B charge separation for DFT for AB(6.5) and AB(14).

#### IV. CONCLUSIONS

In this paper, we have highlighted the importance of incorporating adaptable electronic information into ML models of PESs. For this, we presented a new general approach for building ML models whose target function is a DFT populations-to-energy map, afforded in practice by CDFT. Such models are functions of atom positions and atom-based electron populations; optimization with respect to the latter variables allows for self-consistent adjustment to the electronic environment, akin to optimizing the density in DFT. This leads to features analogous to regular orbital or density based SCF: solutions for different overall charge states; solutions with different character (e.g. CT and non CT); spin polarized versus not, etc. As a byproduct, this model can be used as a high-quality charge equilibration scheme, i.e. partial charges of this model can also be used by themselves to infer molecular properties, e.g. determining nucleophilic and electrophilic sites.

As a proof of concept for this approach, we trained a BpopNN model for  $Li_nH_n$  systems. For this, we simplified the target-energy map to use only total electron populations on each atom, as opposed to the two spin populations. Subsequent testing on a variety of  $Li_nH_n$  examples confirmed that the model exhibits qualitative accuracy for the energy of different overall charge states, a flexible reduced electronic structure whose populations are generally close to those of the parent DFT functional and that can adapt to the (particularly medium to long-range) environment, and transferability to clusters of different sizes. We conclude that BpopNN is a positive first step towards addressing some outstanding challenges for ML models, namely the handling of systems with electronic structure that is sensitive to the surrounding environment.

This first application was designed for qualitative testing and it has several aspects that can explain various inaccuracies and that are ready for refinement in future work. Although partial charges are in principle adequate as a basis for the BpopNN approach, spin popu-

lations will need to be incorporated in practice to effectively model the intricacies of the CDFT map. In this context, it should then be noted that the reduction of the density to atom-based populations is the furthest one can reasonably go; clearly less drastic reductions should also be explored, and perhaps a useful hierarchy for this can be established.

Improvements in the descriptor can be made. One idea would be to incorporate radial scaling functions into the above proxy functions.<sup>28,35</sup> We could alternatively modify the descriptor to be explicitly non-local. Our current descriptor is based conceptually on the ESP generated by the model’s populations as point charges, and it uses simple Gaussian functions to approximate the ESP for each point. ESPs are inherently non-local, while Gaussians are not. It would therefore be worthwhile to better approximate these ESPs or incorporate them exactly. It would be particularly interesting to combine the ideas of this paper with those of LODE:<sup>33</sup> instead of using point charges to generate the ESP, proxy densities, now integrating to the current populations on each atom, could be used to generate ESPs.

The fact that BpopNN models approximate the partial charges of the target density functional is advantageous for the incorporation of long-range electrostatic interactions. The simple approach used here, based on attenuated pairwise point-charge Coulomb interactions, would clearly benefit from refinements to more reasonably approximate the local morphology of the interacting density pieces.

Experimenting with the choice of ML architecture would be worthwhile, and especially experimenting with the NN depth. Of course, this would require a larger training dataset. The dataset used here is on the smaller side, so enlarging this is a focal point for future work. There is also the plausibility that our dataset is insufficiently balanced. More specifically, it only contains alternating Li-H clusters and cannot be reasonably applied to clusters with distinguishable  $\text{Li}_n$  and  $(\text{H}_2)_n$  substructures, its geometries were obtained from trajectories using only the neutral charge state, and generally our sampling method of particular AIMD and optimization trajectories might not be varied enough. In the future we will explore more efficient and balanced sampling methods, such as normal mode sampling.

In the future, we are primarily interested in moving beyond the LiH-cluster scope to more general Li-organic systems. This significant increase in complexity should entail many adjustments to the general approach presented above, and will serve as a firm test of its efficacy.

## V. ACKNOWLEDGEMENT

This work was supported as part of the Joint Center for Energy Storage Research (JCESR), an Energy Innovation Hub funded by the U.S. Department of Energy, Office of Science, Basic Energy Sciences, under Contract No. DE-AC02-06CH11357. Calculations were performed at Lawrence computational cluster, as well as the UC Berkeley Molecular Graphics and Computation Facility (MGCF). MGCF is supported by grant NIH S10OD023532.

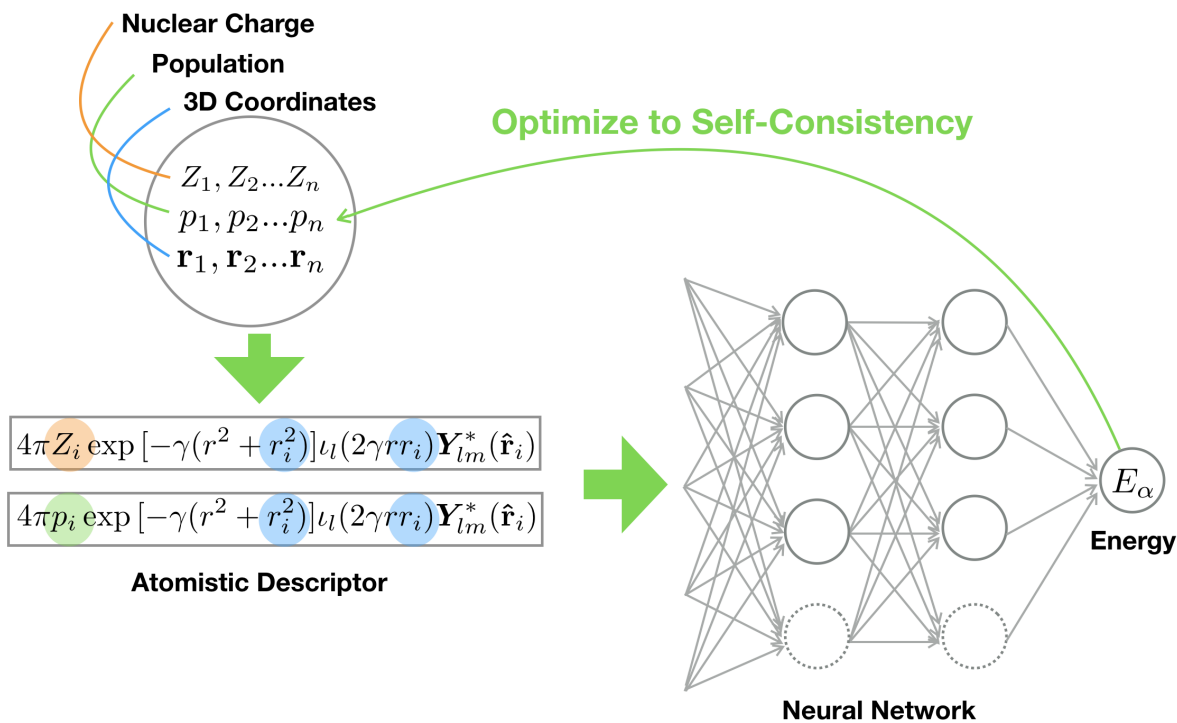


FIG. 1: General concept for the ML potential

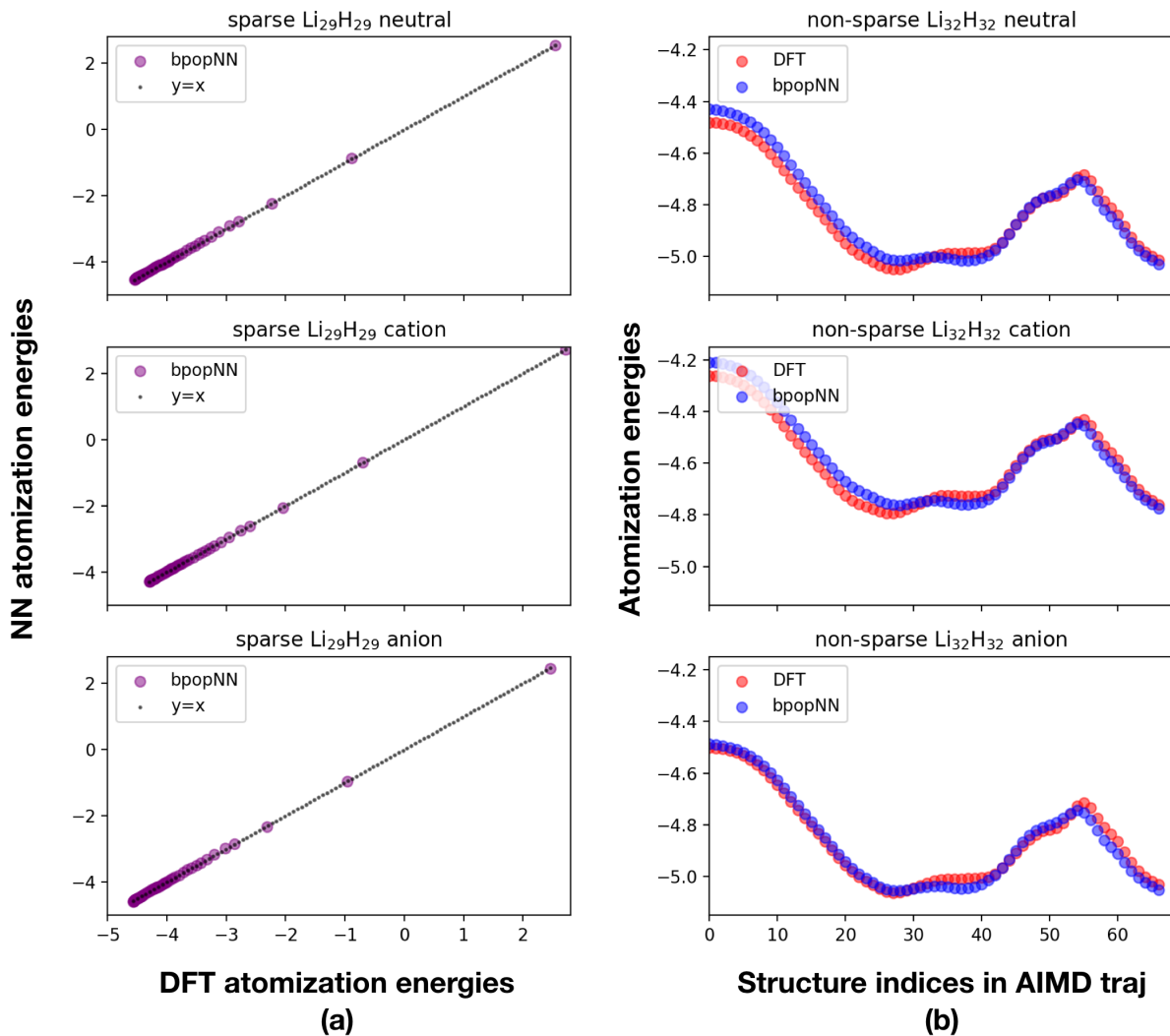
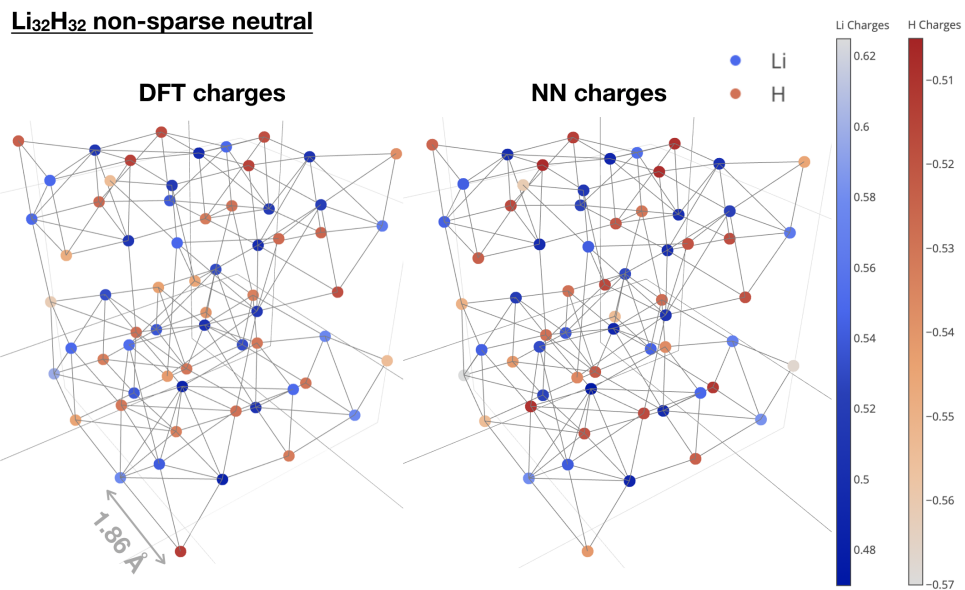


FIG. 2: Performance on sparse  $\text{Li}_{29}\text{H}_{29}$  and non-sparse  $\text{Li}_{32}\text{H}_{32}$  test sets. All energies are in a.u.

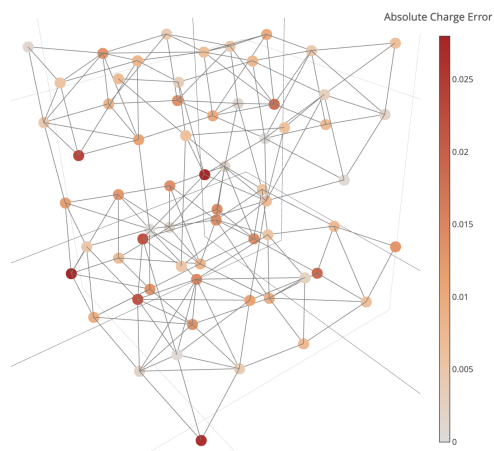
TABLE I: Energy and charge errors for  $\text{Li}_{29}\text{H}_{29}$  sparse and  $\text{Li}_{32}\text{H}_{32}$  non-sparse test sets

	MAE charge (e/atom)	MAE energy (kcal/mol/atom)		MAE charge (e/atom)	MAE energy (kcal/mol/atom)
$\text{Li}_{29}\text{H}_{29}$ neutral	0.012	0.260	$\text{Li}_{32}\text{H}_{32}$ neutral	0.008	0.293
$\text{Li}_{29}\text{H}_{29}$ cation	0.014	0.169	$\text{Li}_{32}\text{H}_{32}$ cation	0.011	0.317
$\text{Li}_{29}\text{H}_{29}$ anion	0.016	0.151	$\text{Li}_{32}\text{H}_{32}$ anion	0.014	0.189





(a) DFT and BpopNN partial charges for a neutral  $\text{Li}_{32}\text{H}_{32}$



(b) Charge error for a neutral  $\text{Li}_{32}\text{H}_{32}$

FIG. 3: Charge optimization result for a neutral  $\text{Li}_{32}\text{H}_{32}$

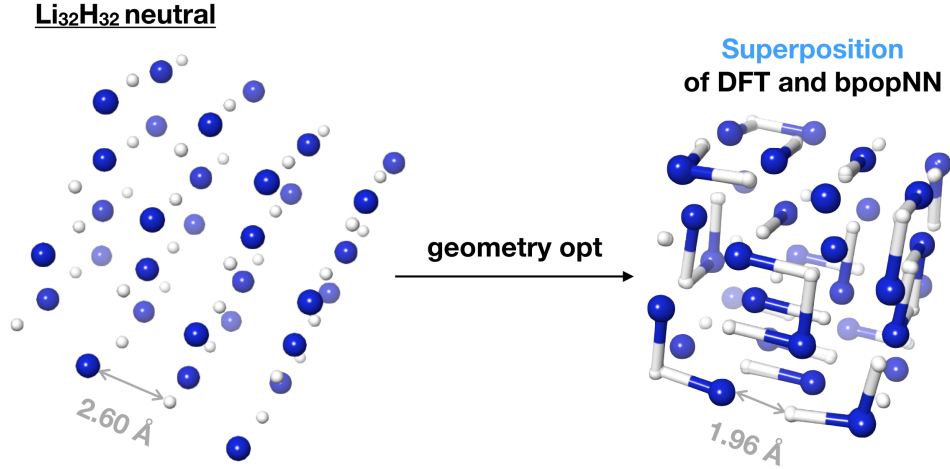


FIG. 4: Geometry optimization on a cubic Li<sub>32</sub>H<sub>32</sub> structure

TABLE II: Energy statistics for a cubic Li<sub>32</sub>H<sub>32</sub> after geometry optimization

	DFT energy (Hartree)	BpopNN energy (Hartree)	Energy Error (kcal/mol/atom)
Neutral	-260.956	-260.967	0.106
Cation	-206.711	-260.716	0.046
Anion	-260.962	-260.983	0.206

	DFT (eV)	BpopNN (eV)
IP	6.656	6.823
EA	0.172	0.450

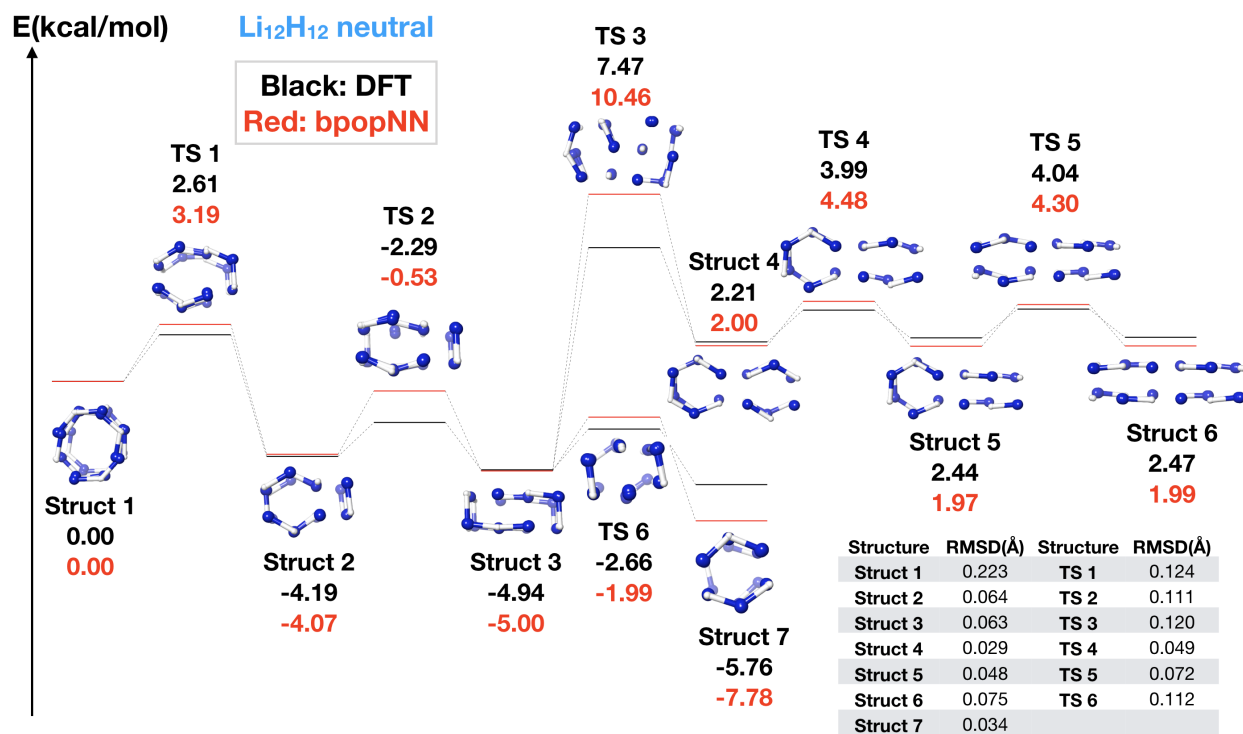


FIG. 5: Energy profile for various structural transformations for neutral Li<sub>12</sub>H<sub>12</sub>. Superpositions and RMSD values were obtained with Maestro.<sup>134</sup>

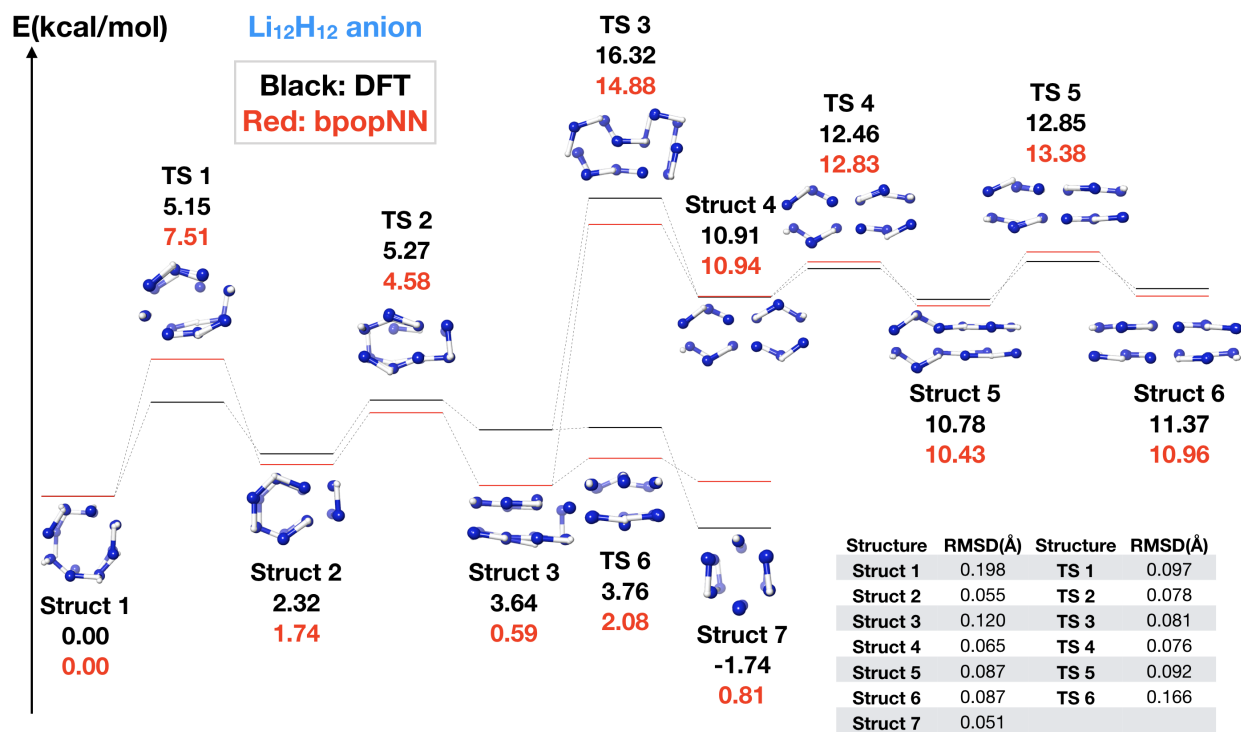


FIG. 6: Energy profile for various structural transformations for anion  $\text{Li}_{12}\text{H}_{12}$ . Superpositions and RMSD values were obtained with Maestro.<sup>134</sup>

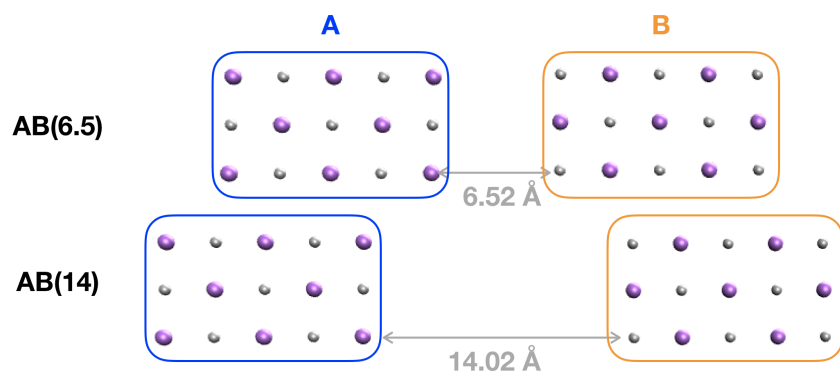


FIG. 7: A charge separation example for neutral  $\text{Li}_{15}\text{H}_{15}$

TABLE III: A charge separation example for neutral  $\text{Li}_{15}\text{H}_{15}$

	DFT energy (Hartree)	BpopNN energy (Hartree)	Absolute error (kcal/mol)
AB(6.5)	-121.524 <sup>a</sup>	-121.534	6.232
AB(14)	-121.506 <sup>a</sup>	-121.525	11.801
AB(1000)	-121.489 <sup>a</sup>	-121.515	16.286
A	-64.213	-64.224	7.213
B	-57.241	-57.254	8.403
A <sup>+</sup>	-64.110	-64.112	1.210
B <sup>-</sup>	-57.372	-57.360	7.584
DFT (kcal/mol)		BpopNN (kcal/mol)	
AB(6.5) BE <sup>b</sup>	-44.143	-34.760	
AB(14) BE	-33.168	-29.345	
	DFT (e)	BpopNN (e)	
$q(\text{A})^{\text{c}}$ in AB(6.5)	0.993	0.729	
$q(\text{A})$ in AB(14)	0.979	0.681	
$q(\text{A})$ in AB(1000)	0.683	0.609	

<sup>a</sup> For CT solutions. There exist DFT “neutral-neutral” solutions, with energies of -121.4542 and -121.4536 respectively for AB(6.5) and AB(14). This is very close to the sum of the energies of A and B.

<sup>b</sup> Binding Energy.

<sup>c</sup> sum of partial charges on A atoms.

## VI. SUPPORTING INFORMATION

Geometries used for all training and test datasets; all training-set data, including CDFT energies, charges, population gradients, and nuclear-position forces; test-set results for the other BpopNN variants; details for training the GAP models with QUIP; test-set results for

the GAP model variants.

---

\* dsmallchem@gmail.com

- <sup>1</sup> Parr, R. G. *Horizons of Quantum Chemistry*; Springer, 1980; pp 5–15.
- <sup>2</sup> Car, R.; Parrinello, M. Unified approach for molecular dynamics and density-functional theory. *Phys. Rev. Lett.* **1985**, *55*, 2471.
- <sup>3</sup> Marx, D.; Hutter, J. *Ab initio molecular dynamics: basic theory and advanced methods*; Cambridge University Press, 2009.
- <sup>4</sup> Hansson, T.; Oostenbrink, C.; van Gunsteren, W. Molecular dynamics simulations. *Curr. Opin. Struct. Biol.* **2002**, *12*, 190–196.
- <sup>5</sup> Karplus, M.; McCammon, J. A. Molecular dynamics simulations of biomolecules. *Nat. Struct. Mol. Biol.* **2002**, *9*, 646.
- <sup>6</sup> Behler, J. First principles neural network potentials for reactive simulations of large molecular and condensed systems. *Angew. Chem. Int. Ed.* **2017**, *56*, 12828–12840.
- <sup>7</sup> Schmidt, J.; Marques, M. R.; Botti, S.; Marques, M. A. Recent advances and applications of machine learning in solid-state materials science. *Npj Comput. Mater.* **2019**, *5*, 1–36.
- <sup>8</sup> Deringer, V. L.; Caro, M. A.; Csányi, G. Machine Learning Interatomic Potentials as Emerging Tools for Materials Science. *Adv. Mater.* **2019**, *31*, 1902765.
- <sup>9</sup> Schleder, G. R.; Padilha, A. C.; Acosta, C. M.; Costa, M.; Fazzio, A. From DFT to machine learning: recent approaches to materials science—a review. *J. Phys.: Mater.* **2019**, *2*, 032001.
- <sup>10</sup> Behler, J.; Parrinello, M. Generalized neural-network representation of high-dimensional potential-energy surfaces. *Phys. Rev. Lett.* **2007**, *98*, 146401.
- <sup>11</sup> Behler, J. Atom-centered symmetry functions for constructing high-dimensional neural network potentials. *J. Chem. Phys.* **2011**, *134*, 074106.
- <sup>12</sup> Bartók, A. P.; Kondor, R.; Csányi, G. On representing chemical environments. *Phys. Rev. B* **2013**, *87*, 184115.
- <sup>13</sup> Behler, J. RuNNer-A Neural Network Code for High-Dimensional Potential-Energy Surfaces; Universität Göttingen, 2018.
- <sup>14</sup> Yao, K.; Herr, J. E.; Toth, D. W.; Mckintyre, R.; Parkhill, J. The TensorMol-0.1 model chemistry: a neural network augmented with long-range physics. *Chem. Sci.* **2018**, *9*, 2261–2269.

- <sup>15</sup> Yao, K.; Herr, J. E.; Parkhill, J. The many-body expansion combined with neural networks. *J. Chem. Phys.* **2017**, *146*, 014106.
- <sup>16</sup> Herr, J. E.; Yao, K.; McIntyre, R.; Toth, D. W.; Parkhill, J. Metadynamics for training neural network model chemistries: A competitive assessment. *J. Chem. Phys.* **2018**, *148*, 241710.
- <sup>17</sup> Yao, K.; Herr, J. E.; Brown, S. N.; Parkhill, J. Intrinsic bond energies from a bonds-in-molecules neural network. *J. Phys. Chem. Lett.* **2017**, *8*, 2689–2694.
- <sup>18</sup> Herr, J. E.; Koh, K.; Yao, K.; Parkhill, J. Compressing physics with an autoencoder: Creating an atomic species representation to improve machine learning models in the chemical sciences. *J. Chem. Phys.* **2019**, *151*, 084103.
- <sup>19</sup> Schütt, K. T.; Sauceda, H. E.; Kindermans, P.-J.; Tkatchenko, A.; Müller, K.-R. SchNet—A deep learning architecture for molecules and materials. *J. Chem. Phys.* **2018**, *148*, 241722.
- <sup>20</sup> For an orbital-based ML approach whose goal is related, see Schütt, K.; Gastegger, M.; Tkatchenko, A.; Müller, K.-R.; Maurer, R. Unifying machine learning and quantum chemistry with a deep neural network for molecular wavefunctions. *Nat. Commun.* **2019**, *10*, 1–10.
- <sup>21</sup> 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.* **2017**, *8*, 3192–3203.
- <sup>22</sup> Smith, J. S.; Nebgen, B. T.; Zubatyuk, R.; Lubbers, N.; Devereux, C.; Barros, K.; Tretiak, S.; Isayev, O.; Roitberg, A. E. Approaching coupled cluster accuracy with a general-purpose neural network potential through transfer learning. *Nat. Commun.* **2019**, *10*, 1–8.
- <sup>23</sup> Smith, J. S.; Nebgen, B.; Lubbers, N.; Isayev, O.; Roitberg, A. E. Less is more: Sampling chemical space with active learning. *J. Chem. Phys.* **2018**, *148*, 241733.
- <sup>24</sup> Khorshidi, A.; Peterson, A. A. Amp: A modular approach to machine learning in atomistic simulations. *Comput. Phys. Commun.* **2016**, *207*, 310–324.
- <sup>25</sup> Wang, H.; Zhang, L.; Han, J.; Weinan, E. DeePMD-kit: A deep learning package for many-body potential energy representation and molecular dynamics. *Comput. Phys. Commun.* **2018**, *228*, 178–184.
- <sup>26</sup> Huang, S.-D.; Shang, C.; Kang, P.-L.; Zhang, X.-J.; Liu, Z.-P. LASP: Fast global potential energy surface exploration. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2019**, *9*, e1415.
- <sup>27</sup> Christensen, A.; Faber, F.; Huang, B.; Bratholm, L.; Tkatchenko, A.; Müller, K.; von Lilienfeld, O. QML: A Python toolkit for quantum machine learning. URL <https://github.com/qmlcode/qml> **2017**,

- <sup>28</sup> Faber, F. A.; Christensen, A. S.; Huang, B.; Von Lilienfeld, O. A. Alchemical and structural distribution based representation for universal quantum machine learning. *J. Chem. Phys.* **2018**, *148*, 241717.
- <sup>29</sup> Unke, O. T.; Meuwly, M. PhysNet: a neural network for predicting energies, forces, dipole moments and partial charges. *J. Chem. Theory Comput.* **2019**,
- <sup>30</sup> Bartók, A. P.; Csányi, G. Gaussian approximation potentials: A brief tutorial introduction. *Int. J. Quantum Chem.* **2015**, *115*, 1051–1057.
- <sup>31</sup> Grisafi, A.; Fabrizio, A.; Meyer, B.; Wilkins, D. M.; Corminboeuf, C.; Ceriotti, M. Transferable machine-learning model of the electron density. *ACS Cent. Sci.* **2018**, *5*, 57–64.
- <sup>32</sup> Grisafi, A.; Wilkins, D. M.; Csányi, G.; Ceriotti, M. Symmetry-adapted machine learning for tensorial properties of atomistic systems. *Phys. Rev. Lett.* **2018**, *120*, 036002.
- <sup>33</sup> Grisafi, A.; Ceriotti, M. Incorporating long-range physics in atomic-scale machine learning. *J. Chem. Phys.* **2019**, *151*, 204105.
- <sup>34</sup> De, S.; Bartók, A. P.; Csányi, G.; Ceriotti, M. Comparing molecules and solids across structural and alchemical space. *Phys. Chem. Chem. Phys.* **2016**, *18*, 13754–13769.
- <sup>35</sup> Willatt, M. J.; Musil, F.; Ceriotti, M. Feature optimization for atomistic machine learning yields a data-driven construction of the periodic table of the elements. *Phys. Chem. Chem. Phys.* **2018**, *20*, 29661–29668.
- <sup>36</sup> Thompson, A. P.; Swiler, L. P.; Trott, C. R.; Foiles, S. M.; Tucker, G. J. Spectral neighbor analysis method for automated generation of quantum-accurate interatomic potentials. *J. Comput. Phys.* **2015**, *285*, 316–330.
- <sup>37</sup> Tang, Y.-H.; de Jong, W. A. Prediction of atomization energy using graph kernel and active learning. *J. Chem. Phys.* **2019**, *150*, 044107.
- <sup>38</sup> Ferré, G.; Haut, T.; Barros, K. Learning molecular energies using localized graph kernels. *J. Chem. Phys.* **2017**, *146*, 114107.
- <sup>39</sup> Gilmer, J.; Schoenholz, S. S.; Riley, P. F.; Vinyals, O.; Dahl, G. E. Neural message passing for quantum chemistry. Proceedings of the 34th International Conference on Machine Learning-Volume 70. 2017; pp 1263–1272.
- <sup>40</sup> Chen, C.; Ye, W.; Zuo, Y.; Zheng, C.; Ong, S. P. Graph networks as a universal machine learning framework for molecules and crystals. *Chem. Mater.* **2019**, *31*, 3564–3572.



- <sup>41</sup> Rupp, M.; Tkatchenko, A.; Müller, K.-R.; Von Lilienfeld, O. A. Fast and accurate modeling of molecular atomization energies with machine learning. *Phys. Rev. Lett.* **2012**, *108*, 058301.
- <sup>42</sup> Hansen, K.; Montavon, G.; Biegler, F.; Fazli, S.; Rupp, M.; Scheffler, M.; Von Lilienfeld, O. A.; Tkatchenko, A.; Müller, K.-R. Assessment and validation of machine learning methods for predicting molecular atomization energies. *J. Chem. Theory Comput.* **2013**, *9*, 3404–3419.
- <sup>43</sup> Hansen, K.; Biegler, F.; Ramakrishnan, R.; Pronobis, W.; Von Lilienfeld, O. A.; Müller, K.-R.; Tkatchenko, A. Machine learning predictions of molecular properties: Accurate many-body potentials and nonlocality in chemical space. *J. Phys. Chem. Lett.* **2015**, *6*, 2326–2331.
- <sup>44</sup> Shao, K.; Chen, J.; Zhao, Z.; Zhang, D. H. Communication: Fitting potential energy surfaces with fundamental invariant neural network. *J. Chem. Phys.* **2016**, *145*, 071101.
- <sup>45</sup> Deringer, V. L.; Csányi, G. Machine learning based interatomic potential for amorphous carbon. *Phys. Rev. B* **2017**, *95*, 094203.
- <sup>46</sup> Sosso, G. C.; Deringer, V. L.; Elliott, S. R.; Csányi, G. Understanding the thermal properties of amorphous solids using machine-learning-based interatomic potentials. *Mol. Simulat.* **2018**, *44*, 866–880.
- <sup>47</sup> Caro, M. A.; Deringer, V. L.; Koskinen, J.; Laurila, T.; Csányi, G. Growth Mechanism and Origin of High s p<sup>3</sup> Content in Tetrahedral Amorphous Carbon. *Phys. Rev. Lett.* **2018**, *120*, 166101.
- <sup>48</sup> 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.* **2018**, *148*, 241711.
- <sup>49</sup> Bendikov, M.; Duong, H. M.; Starkey, K.; Houk, K. N.; Carter, E. A.; Wudl, F. Oligoacenes: theoretical prediction of open-shell singlet diradical ground states. *J. Am. Chem. Soc.* **2004**, *126*, 7416–7417.
- <sup>50</sup> Hachmann, J.; Dorando, J. J.; Avilés, M.; Chan, G. K.-L. The radical character of the acenes: a density matrix renormalization group study. *J. Chem. Phys.* **2007**, *127*, 134309.
- <sup>51</sup> Jiang, D.-e.; Dai, S. Electronic Ground State of Higher Acenes. *J. Phys. Chem. A* **2008**, *112*, 332–335.
- <sup>52</sup> Dutta, S.; Pati, S. K. Novel properties of graphene nanoribbons: a review. *J. Mater. Chem.* **2010**, *20*, 8207–8223.

- <sup>53</sup> Cervantes-Sodi, F.; Csányi, G.; Piscanec, S.; Ferrari, A. C. Edge-functionalized and substitutionally doped graphene nanoribbons: Electronic and spin properties. *Phys. Rev. B* **2008**, *77*, 165427.
- <sup>54</sup> Barone, V.; Hod, O.; Scuseria, G. E. Electronic Structure and Stability of Semiconducting Graphene Nanoribbons. *Nano Lett.* **2006**, *6*, 2748–2754.
- <sup>55</sup> Meier, H.; Stalmach, U.; Kolshorn, H. Effective conjugation length and UV/vis spectra of oligomers. *Acta Polym.* **1997**, *48*, 379–384.
- <sup>56</sup> Izumi, T.; Kobashi, S.; Takimiya, K.; Aso, Y.; Otsubo, T. Synthesis and Spectroscopic Properties of a Series of  $\beta$ -Blocked Long Oligothiophenes up to the 96-mer: Reevaluation of Effective Conjugation Length. *J. Am. Chem. Soc.* **2003**, *125*, 5286–5287.
- <sup>57</sup> Kishino, S.; Ueno, Y.; Ochiai, K.; Rikukawa, M.; Sanui, K.; Kobayashi, T.; Kunugita, H.; Ema, K. Estimate of the effective conjugation length of polythiophene from its  $|\chi^{(3)}(\omega; \omega, \omega, -\omega)|$  spectrum at excitonic resonance. *Phys. Rev. B* **1998**, *58*, R13430–R13433.
- <sup>58</sup> Gray, H. B.; W.R. Ellis, Jr., In *Chapter 6: Electron Transfer*; Bertini, I., Gray, H. B., Lipard, S., Valentine, J., Eds.; University Science Books, Mill Valley, CA, 1994.
- <sup>59</sup> Marcus, R.; Sutin, N. Electron transfers in chemistry and biology. *Biochim. Biophys. Acta* **1985**, *811*, 265 – 322.
- <sup>60</sup> Jerome, D. Organic Conductors: From Charge Density Wave TTF-TCNQ to Superconducting (TMTSF)<sub>2</sub>PF<sub>6</sub>. *Chem. Rev.* **2004**, *104*, 5565–5592.
- <sup>61</sup> Forró, L.; Mihály, L. Electronic properties of doped fullerenes. *Rep. Prog. Phys.* **2001**, *64*, 649–699.
- <sup>62</sup> Takahashi, Y.; Hayakawa, K.; Naito, T.; Inabe, T. What Happens at the Interface between TTF and TCNQ Crystals (TTF = Tetrathiafulvalene and TCNQ = 7,7,8,8-Tetracyanoquinodimethane)? *J. Phys. Chem. C* **2012**, *116*, 700–703.
- <sup>63</sup> Peled, E. The electrochemical behavior of alkali and alkaline earth metals in nonaqueous battery systems—the solid electrolyte interphase model. *J. Electrochem. Soc.* **1979**, *126*, 2047.
- <sup>64</sup> Nie, M.; Chalasani, D.; Abraham, D. P.; Chen, Y.; Bose, A.; Lucht, B. L. Lithium ion battery graphite solid electrolyte interphase revealed by microscopy and spectroscopy. *J. Phys. Chem. C* **2013**, *117*, 1257–1267.
- <sup>65</sup> Winter, M. The solid electrolyte interphase—the most important and the least understood solid electrolyte in rechargeable Li batteries. *Z. Phys. Chem.* **2009**, *223*, 1395–1406.

- <sup>66</sup> Wang, A.; Kadam, S.; Li, H.; Shi, S.; Qi, Y. Review on modeling of the anode solid electrolyte interphase (SEI) for lithium-ion batteries. *Npj Comput. Mater.* **2018**, *4*, 1–26.
- <sup>67</sup> Peled, E.; Menkin, S. SEI: past, present and future. *J. Electrochem. Soc.* **2017**, *164*, A1703–A1719.
- <sup>68</sup> Wang, Y.; Nakamura, S.; Ue, M.; Balbuena, P. B. Theoretical studies to understand surface chemistry on carbon anodes for lithium-ion batteries: reduction mechanisms of ethylene carbonate. *J. Am. Chem. Soc.* **2001**, *123*, 11708–11718.
- <sup>69</sup> Artrith, N.; Morawietz, T.; Behler, J. High-dimensional neural-network potentials for multi-component systems: Applications to zinc oxide. *Phys. Rev. B* **2011**, *83*, 153101.
- <sup>70</sup> Wang, H.; Yang, W. Force Field for Water Based on Neural Network. *J. Phys. Chem. Lett.* **2018**, *9*, 3232–3240.
- <sup>71</sup> Morawietz, T.; Behler, J. A full-dimensional neural network potential-energy surface for water clusters up to the hexamer. *Z. Phys. Chem.* **2013**, *227*, 1559–1581.
- <sup>72</sup> Morawietz, T.; Behler, J. A density-functional theory-based neural network potential for water clusters including van der waals corrections. *J. Phys. Chem. A* **2013**, *117*, 7356–7366.
- <sup>73</sup> Deng, Z.; Chen, C.; Li, X.-G.; Ong, S. P. An electrostatic spectral neighbor analysis potential for lithium nitride. *Npj Comput. Mater.* **2019**, *5*, 1–8.
- <sup>74</sup> Shen, L.; Wu, J.; Yang, W. Multiscale quantum mechanics/molecular mechanics simulations with neural networks. *J. Chem. Theory Comput.* **2016**, *12*, 4934–4946.
- <sup>75</sup> Houlding, S.; Liem, S.; Popelier, P. A polarizable high-rank quantum topological electrostatic potential developed using neural networks: Molecular dynamics simulations on the hydrogen fluoride dimer. *Int. J. Quantum Chem.* **2007**, *107*, 2817–2827.
- <sup>76</sup> Ewald, P. P. Die Berechnung optischer und elektrostatischer Gitterpotentiale. *Ann. Phys. (Berl.)* **1921**, *369*, 253–287.
- <sup>77</sup> Gastegger, M.; Behler, J.; Marquetand, P. Machine learning molecular dynamics for the simulation of infrared spectra. *Chem. Sci.* **2017**, *8*, 6924–6935.
- <sup>78</sup> Bleiziffer, P.; Schaller, K.; Riniker, S. Machine learning of partial charges derived from high-quality quantum-mechanical calculations. *J. Chem. Inf. Model.* **2018**, *58*, 579–590.
- <sup>79</sup> Wang, Y.; Fass, J.; Stern, C. D.; Luo, K.; Chodera, J. Graph Nets for Partial Charge Prediction. *arXiv preprint arXiv:1909.07903* **2019**,

- <sup>80</sup> Bereau, T.; Andrienko, D.; Von Lilienfeld, O. A. Transferable atomic multipole machine learning models for small organic molecules. *J. Chem. Theory Comput.* **2015**, *11*, 3225–3233.
- <sup>81</sup> Bereau, T.; DiStasio Jr, R. A.; Tkatchenko, A.; Von Lilienfeld, O. A. Non-covalent interactions across organic and biological subsets of chemical space: Physics-based potentials parametrized from machine learning. *J. Chem. Phys.* **2018**, *148*, 241706.
- <sup>82</sup> Chandrasekaran, A.; Kamal, D.; Batra, R.; Kim, C.; Chen, L.; Ramprasad, R. Solving the electronic structure problem with machine learning. *Npj Comput. Mater.* **2019**, *5*, 1–7.
- <sup>83</sup> Mortier, W. J.; Van Genechten, K.; Gasteiger, J. Electronegativity equalization: application and parametrization. *J. Am. Chem. Soc.* **1985**, *107*, 829–835.
- <sup>84</sup> Rappe, A. K.; Goddard, W. A. Charge equilibration for molecular dynamics simulations. *J. Phys. Chem.* **1991**, *95*, 3358–3363.
- <sup>85</sup> Nistor, R. A.; Polihronov, J. G.; Müser, M. H.; Mosey, N. J. A generalization of the charge equilibration method for nonmetallic materials. *J. Chem. Phys.* **2006**, *125*, 094108.
- <sup>86</sup> Chen, J.; Martinez, T. J. QTPIE: Charge transfer with polarization current equalization. A fluctuating charge model with correct asymptotics. *Chem. Phys. Lett.* **2007**, *438*, 315–320.
- <sup>87</sup> Zhang, M.; Fournier, R. Self-Consistent Charge Equilibration Method and Its Application to Au<sub>13</sub>Na<sub>n</sub> (n = 1,10) Clusters. *J. Phys. Chem. A* **2009**, *113*, 3162–3170.
- <sup>88</sup> Wilmer, C. E.; Kim, K. C.; Snurr, R. Q. An Extended Charge Equilibration Method. *J. Phys. Chem. Lett.* **2012**, *3*, 2506–2511.
- <sup>89</sup> Verstraelen, T.; Ayers, P. W.; Van Speybroeck, V.; Waroquier, M. ACKS2: Atom-condensed Kohn-Sham DFT approximated to second order. *J. Chem. Phys.* **2013**, *138*, 074108.
- <sup>90</sup> Wells, B. A.; De Bruin-Dickason, C.; Chaffee, A. L. Charge Equilibration Based on Atomic Ionization in Metal-Organic Frameworks. *J. Phys. Chem. C* **2015**, *119*, 456–466.
- <sup>91</sup> Naserifar, S.; Brooks, D. J.; Goddard, W. A.; Cvacek, V. Polarizable charge equilibration model for predicting accurate electrostatic interactions in molecules and solids. *J. Chem. Phys.* **2017**, *146*, 124117.
- <sup>92</sup> 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* **2015**, *92*, 045131.
- <sup>93</sup> Dick, S.; Fernandez-Serra, M. Machine Learning Accurate Exchange and Correlation Functionals of the Electronic Density. *ChemRxiv e-prints* **2019**,

- <sup>94</sup> Schmidt, J.; Benavides-Riveros, C. L.; Marques, M. A. Machine Learning the Physical Nonlocal Exchange–Correlation Functional of Density-Functional Theory. *J. Phys. Chem. Lett.* **2019**, *10*, 6425–6431.
- <sup>95</sup> Nagai, R.; Akashi, R.; Sugino, O. Completing density functional theory by machine-learning hidden messages from molecules. *arXiv preprint arXiv:1903.00238* **2019**,
- <sup>96</sup> Kolb, B.; Lentz, L. C.; Kolpak, A. M. Discovering charge density functionals and structure-property relationships with PROPhet: A general framework for coupling machine learning and first-principles methods. *Sci. Rep.* **2017**, *7*, 1–9.
- <sup>97</sup> Snyder, J. C.; Rupp, M.; Hansen, K.; Blooston, L.; Müller, K.-R.; Burke, K. Orbital-free bond breaking via machine learning. *J. Chem. Phys.* **2013**, *139*, 224104.
- <sup>98</sup> Snyder, J. C.; Rupp, M.; Hansen, K.; Müller, K.-R.; Burke, K. Finding density functionals with machine learning. *Phys. Rev. Lett.* **2012**, *108*, 253002.
- <sup>99</sup> Li, L.; Snyder, J. C.; Pelaschier, I. M.; Huang, J.; Niranjani, U.-N.; Duncan, P. Understanding machine-learned density functionals. *Int. J. Quantum Chem.* **2016**, *116*, 819–833.
- <sup>100</sup> Yao, K.; Parkhill, J. Kinetic energy of hydrocarbons as a function of electron density and convolutional neural networks. *J. Chem. Theory. Comput.* **2016**, *12*, 1139–1147.
- <sup>101</sup> Brockherde, F.; Vogt, L.; Li, L.; Tuckerman, M. E.; Burke, K.; Müller, K.-R. Bypassing the Kohn-Sham equations with machine learning. *Nat. Commun.* **2017**, *8*, 1–10.
- <sup>102</sup> Levy, M. Universal variational functionals of electron densities, first-order density matrices, and natural spin-orbitals and solution of the v-representability problem. *Proc. Natl. Acad. Sci. U. S. A.* **1979**, *76*, 6062–6065.
- <sup>103</sup> Becke, A. A multicenter numerical integration scheme for polyatomic molecules. *J. Chem. Phys.* **1988**, *88*, 2547–2553.
- <sup>104</sup> Dederichs, P. H.; Blügel, S.; Zeller, R.; Akai, H. Ground States of Constrained Systems: Application to Cerium Impurities. *Phys. Rev. Lett.* **1984**, *53*, 2512–2515.
- <sup>105</sup> Wu, Q.; Van Voorhis, T. Direct optimization method to study constrained systems within density-functional theory. *Phys. Rev. A* **2005**, *72*, 024502.
- <sup>106</sup> Kaduk, B.; Kowalczyk, T.; Van Voorhis, T. Constrained Density Functional Theory. *Chem. Rev.* **2012**, *112*, 321–370.
- <sup>107</sup> Shao, Y. et al. Advances in molecular quantum chemistry contained in the Q-Chem 4 program package. *Mol. Phys.* **2015**, *113*, 184–215.

- <sup>108</sup> Wu, Q.; Van Voorhis, T. Direct Calculation of Electron Transfer Parameters through Constrained Density Functional Theory. *J. Phys. Chem. A* **2006**, *110*, 9212–9218.
- <sup>109</sup> Kaufmann, K.; Baumeister, W. Single-centre expansion of Gaussian basis functions and the angular decomposition of their overlap integrals. *J. Phys. B* **1989**, *22*, 1.
- <sup>110</sup> Artrith, N.; Urban, A.; Ceder, G. Efficient and accurate machine-learning interpolation of atomic energies in compositions with many species. *Phys. Rev. B* **2017**, *96*, 014112.
- <sup>111</sup> Ragot, S.; Gillet, J.-M.; Becker, P. J. A cluster partitioning method: determination of density matrices of solids and comparison with X-ray experiments. *arXiv e-prints* **2001**, arXiv:physics/0112078.
- <sup>112</sup> Glorot, X.; Bordes, A.; Bengio, Y. Deep sparse rectifier neural networks. Proceedings of the fourteenth international conference on artificial intelligence and statistics. 2011; pp 315–323.
- <sup>113</sup> Kingma, D. P.; Ba, J. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980* **2014**,
- <sup>114</sup> Abadi, M. et al. TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems. 2015; <http://tensorflow.org/>, Software available from tensorflow.org.
- <sup>115</sup> Rao, B.; Jena, P. Molecular cluster calculations of the electronic structure of lithium hydride. *J. Phys. C: Solid State Phys.* **1986**, *19*, 5167–5172.
- <sup>116</sup> Sapse, A.-M.; Jain, D. C.; Raghavachari, K. In *Lithium Chemistry: A Theoretical and Experimental Overview*; Sapse, A.-M., von R. Schleyer, P., Eds.; John Wiley & Sons, 1995; Chapter 2, p 45.
- <sup>117</sup> Chen, Y.-L.; Huang, C.-H.; Hu, W.-P. Theoretical Study on the Small Clusters of LiH, NaH, BeH<sub>2</sub>, and MgH<sub>2</sub>. *J. Phys. Chem. A* **2005**, *109*, 9627–9636.
- <sup>118</sup> Nolan, S. J.; Gillan, M. J.; Alfè, D.; Allan, N. L.; Manby, F. R. Calculation of properties of crystalline lithium hydride using correlated wave function theory. *Phys. Rev. B* **2009**, *80*, 165109.
- <sup>119</sup> Sato, T.; Ishikawa, K. L. Time-dependent multiconfiguration self-consistent-field method based on the occupation-restricted multiple-active-space model for multielectron dynamics in intense laser fields. *Phys. Rev. A* **2015**, *91*, 023417.
- <sup>120</sup> Huang, Z.-Q.; Chen, W.-C.; Chuang, F.-C.; Majzoub, E. H.; Ozolins, V. First-principles calculated decomposition pathways for LiBH<sub>4</sub> nanoclusters. *Sci. Rep.* **2016**, *6*, 26056.

- <sup>121</sup> Antoine, R.; Dugourd, P.; Rayane, D.; Broyer, M. Dissociation pathways and binding energies of  $(\text{LiH})_n\text{Li}^+$  and  $(\text{LiH})_n\text{Li}_3^+$  clusters. *J. Chem. Phys.* **1996**, *104*, 110–119.
- <sup>122</sup> Wang, X.; Andrews, L. Infrared Spectra and Theoretical Calculations of Lithium Hydride Clusters in Solid Hydrogen, Neon, and Argon. *J. Phys. Chem. A* **2007**, *111*, 6008–6019.
- <sup>123</sup> Wagner, L. K.; Majzoub, E. H.; Allendorf, M. D.; Grossman, J. C. Tuning metal hydride thermodynamics via size and composition: Li-H, Mg-H, Al-H, and Mg-Al-H nanoclusters for hydrogen storage. *Phys. Chem. Chem. Phys.* **2012**, *14*, 6611–6616.
- <sup>124</sup> Harder, S. Molecular early main group metal hydrides: synthetic challenge, structures and applications. *Chem. Commun.* **2012**, *48*, 11165–11177.
- <sup>125</sup> Stasch, A. Well-Defined, Nanometer-Sized LiH Cluster Compounds Stabilized by Pyrazolate Ligands. *Angew. Chem. Int. Ed.* **2014**, *53*, 1338–1341.
- <sup>126</sup> Huang, L.; Bonnet, J.-P.; Zlotea, C.; Bourgon, J.; Latroche, M.; Courty, M.; Aymard, L. Synthesis of destabilized nanostructured lithium hydride via hydrogenation of lithium electrochemically inserted into graphite. *Int. J. Hydrogen Energy*. **2015**, *40*, 13936–13941.
- <sup>127</sup> Wang, L.; Quadir, M. Z.; Aguey-Zinsou, K.-F. Direct and reversible hydrogen storage of lithium hydride (LiH) nanoconfined in high surface area graphite. *Int. J. Hydrogen Energy*. **2016**, *41*, 18088–18094.
- <sup>128</sup> Marx, D.; Hutter, J. *Ab Initio Molecular Dynamics: Basic Theory and Advanced Methods*; Cambridge University Press, 2009.
- <sup>129</sup> Becke, A. D. Density-functional thermochemistry. III. The role of exact exchange. *J. Chem. Phys.* **1993**, *98*, 5648–5652.
- <sup>130</sup> Stephens, P. J.; Devlin, F. J.; Chabalowski, C. F.; Frisch, M. J. Ab Initio Calculation of Vibrational Absorption and Circular Dichroism Spectra Using Density Functional Force Fields. *J. Phys. Chem.* **1994**, *98*, 11623–11627.
- <sup>131</sup> Grimme, S.; Ehrlich, S.; Goerigk, L. Effect of the damping function in dispersion corrected density functional theory. *J. Comput. Chem.* **2011**, *32*, 1456–1465.
- <sup>132</sup> Chai, J.-D.; Head-Gordon, M. Long-range corrected hybrid density functionals with damped atom–atom dispersion corrections. *Phys. Chem. Chem. Phys.* **2008**, *10*, 6615–6620.
- <sup>133</sup> Rappoport, D.; Furche, F. Property-optimized Gaussian basis sets for molecular response calculations. *J. Chem. Phys.* **2010**, *133*, 134105.
- <sup>134</sup> Schrödinger Release 2019-4: Maestro, Schrödinger, LLC, New York, NY, 2019.

### Appendix A: Population gradient of CDFT energy

In the present context, the CDFT Lagrangian takes the following form:

$$L_{\text{CDFT}} = E[\gamma] + \sum_{\sigma} \sum_{i=1}^{N_{\text{atom}}-1} \lambda_{i,\sigma} (p_{i,\sigma}(\rho) - c_{i,\sigma}), \quad (\text{A1})$$

where  $\gamma$  is the density matrix, the  $\lambda_{i,\sigma}$  are Lagrange multipliers, the  $p_{i,\sigma}(\rho)$  are the functions defined in eqn. (2), and the  $c_{i,\sigma}$  are the target values of the constraints. For each spin, the sum excludes one atom because the value of this last variable is implied by the usual condition that the density integrates to the total number of electrons of that spin. An equivalent result is obtained if a different atom is the one excluded from the sum, although of course the  $\lambda_i$  and  $c_i$  will adjust accordingly. This Lagrangian also formally applies to cases in which fewer population constraints are used. In that case, we simply set the  $c_i$  for any technically unconstrained population to its relaxed value, and the associated  $\lambda_i$  is 0.

Optimizing  $L_{\text{CDFT}}$  with respect to the orbitals (i.e.  $\gamma$ ) and the Lagrange multipliers gives the CDFT energy,  $E_{\text{CDFT}}[\{c_{i,\sigma}\}]$ . At a stationary point of  $L_{\text{CDFT}}$ , we may apply the “envelope” theorem to obtain

$$\frac{\partial E_{\text{CDFT}}}{\partial c_{i,\sigma}} = \frac{\partial L_{\text{CDFT}}}{\partial c_{i,\sigma}} = -\lambda_{i,\sigma}. \quad (\text{A2})$$

The target energy function  $E_t[\mathbf{p}]$  is essentially the same as  $E_{\text{CDFT}}[\{c_{i,\sigma}\}]$  except that it is a function of all  $2N_{\text{atom}}$  population variables. So we have to do a transformation between variables. Again excluding the last atom, we can fix all  $c_{i,\sigma}$  but one,  $c_{j,\sigma}$ , and then vary  $c_{j,\sigma}$ . Upon (re-)optimizing  $L_{\text{CDFT}}$  for this new set of constraints,  $p_{j,\sigma}$  changes identically as  $c_{j,\sigma}$  does, and  $p_{N_{\text{atom}},\sigma}$  changes by the opposite amount, while all other populations are unchanged. Denoting the direction corresponding to this overall change in populations by  $\mathbf{e}_{j,\sigma}$ , we thus have

$$\nabla E_t[\mathbf{p}] \cdot \mathbf{e}_{i,\sigma} = \frac{\partial E_{\text{CDFT}}}{\partial c_{i,\sigma}} = -\lambda_{i,\sigma}. \quad (\text{A3})$$

The vectors  $\mathbf{e}_{i,\sigma}$  are (most of) the columns of a block diagonal matrix  $W$  with two  $N_{\text{atom}}$



by  $N_{\text{atom}}$  blocks, one for each spin type. That is,  $W = W_s \oplus W_s$  with

$$W_s = \begin{bmatrix} 1 & 0 & \dots & 0 & 1 \\ 0 & 1 & \dots & 0 & 1 \\ \vdots & \vdots & \dots & \vdots & \vdots \\ 0 & 0 & \dots & 1 & 1 \\ -1 & -1 & \dots & -1 & 1 \end{bmatrix}. \quad (\text{A4})$$

The last column in  $W_s$  corresponds to the fully symmetric direction of changing all populations (of one spin type) by the same amount. Because the total numbers of each spin are constant, the component of  $\nabla E_t$  along this direction is 0. We thus have

$$\nabla E_t \cdot W = \boldsymbol{\nu}, \quad (\text{A5})$$

where the latter vector contains the  $-\lambda_i$  along with two 0's for the symmetric directions:

$$\nu_i = \begin{cases} -\lambda_{i,\alpha} & \text{if } i < N_{\text{atom}} \\ 0 & \text{if } i = N_{\text{atom}} \\ -\lambda_{i-N_{\text{atom}},\beta} & \text{if } N_{\text{atom}} < i < 2N_{\text{atom}} \\ 0 & \text{if } i = 2N_{\text{atom}} \end{cases}. \quad (\text{A6})$$

By inverting  $W$  we readily obtain  $\nabla E_t[\boldsymbol{p}]$  (in the regular population basis).