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, Kristin A. Persson, and David W. Small

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 LiHₙ clusters.

1. INTRODUCTION

DFT computations continue to be the predominant approach to first-principles modeling of chemical systems, while empirical force-field based approximations continue to dominate large scale simulations. 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. 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 of Behler and Parrinello and the smooth overlap of atomic positions (SOAP) by Bartók et al. Some specific ML models that fit the general positions-to-energy characterization are RuNNer, TensorMo, SchNet, ANI-1, AMP, DeepMD, LASP, QML models, PhysNet, GAP, models based on atom-density representations by Ceriotti and co-workers, the spectral neighbor analysis potential (SNAP), graph kernel methods, graph neural networks, methods based upon eigenvalues of the Coulomb matrix, Bag-of-Bonds, permutation invariant polynomials, and others. Impressive examples of large-scale applications of these methods include simulating proteins, amorphous carbon, constructing phase diagrams of amorphous Li,Si, 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 3.2.3 below. This general problem applies to many molecules and materials. Examples relating to conjugation, like the emergence...
of polyradicality with increasing length in acenes,\textsuperscript{49−51} armchair versus zigzag edge effects in graphene nanoribbons,\textsuperscript{52−54} and the gradual ascension to the effective conjugation length in a variety of oligomers,\textsuperscript{55−57} 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,\textsuperscript{58,59} and large-scale organic donor−acceptor complexes with appealing electronic properties.\textsuperscript{60−62}

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).\textsuperscript{63−67} 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, that is, unpaired electron on Li or on the rest of the molecule. This variation is subtly dependent on where the Li atom locates around the electron on Li or on the rest of the molecule.\textsuperscript{58,59} The transition between the two characters can be fairly rapid, and so for position-based descriptors, whose terms do not 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. 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\textsuperscript{78} 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 \(\text{Zn}_n\text{O}_m\) clusters,\textsuperscript{9} water clusters,\textsuperscript{14,70−72} proteins,\textsuperscript{14} and other examples.\textsuperscript{57,74} Popelier and co-workers first proposed to employ NNs to construct environment dependent multipoles.\textsuperscript{75} Artrith et al.\textsuperscript{69} and later on Yao et al.\textsuperscript{14} 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\textsuperscript{76} for periodic systems or Coulomb’s law for molecules. Another approach fits molecular dipoles to infer partial charges.\textsuperscript{77} We note that there are also independent ML models specialized for predicting partial charges,\textsuperscript{67,79} multipole moments,\textsuperscript{52,80,81} and even the full electron density.\textsuperscript{31,82} The latter method has been used, for example, 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\textsuperscript{83} and QEq\textsuperscript{84} and several newer examples,\textsuperscript{85−91} 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.\textsuperscript{92} 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.\textsuperscript{71−90} Others pertain more to the fact that commonly used functionals do not provide a direct noniterative link between density and energy, so a ML approximation for this, as with other orbital-free approaches, can greatly reduce the associated burden.\textsuperscript{20,96−101} 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 Article, 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.

2. THEORY

2.1. Target Energy Function. 2.1.1. Formal Viewpoint. In the Levy constrained search approach in DFT, each density \( \rho \) maps to the lowest possible energy \( E[\rho] \) obtainable from a many-electron wave function vector 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.)

\[
E[\{v_i\}] = \min_{\Psi \rightarrow \{v_i\}} \left\langle \Psi | \mathbf{H} | \Psi \right\rangle
\]

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 eq 1 will be replaced by a search over orbitals and an approximate functional, respectively.

For \( \{v_i\} \) in this Article, we are primarily concerned with reducing the density to atom populations, in particular Becke populations. 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

\[
P_{i,\sigma} = \int \omega_i(r) \rho_{i,\sigma}(r)
\]

The weight functions \( \omega_i \) 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 \( E[p] \), where we have placed the populations into a vector \( p \). 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 \( P_{N_e,N_\beta} \) which "projects" the population vector to a vector obeying the sum conditions:

\[
[P_{N_e,N_\beta}(p)]_{i,\sigma} = P_{i,\sigma} + \frac{N_{\beta} - \sum_{\mu} P_{\mu,\sigma}}{N_{\text{atom}}}
\]

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 \( E[P_{N_e,N_\beta}(p)] \).

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.

2.1.2. Target Function in Practice: CDFT. The above ideas may be applied to any of the functionals used in common practice, for example, 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 nonstandard 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) \( E_{\text{CDFT}} \) 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, 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, that is, it is desirable to train the model toward both the direct value of the target function and its first derivatives. The \( \alpha \beta \gamma \) position derivatives for CDFT have been previously derived. For this present Article, we also need a way to compute the derivatives of the CDFT energy with respect to populations. This is described in the Appendix A.

2.2. 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 atomic descriptor \( D(Z, p, r) \) to the atomic energy. The model energy function can be written as

\[
E_{\text{ML}} = \sum_{\mu} \sum_{\nu \in A_\mu} c_{\mu}[D(Z, p, r)] + E_{\text{int}}[Z, p, r]
\]

where \( c_{\mu} \) is 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 \( \mu \); \( Z \) and \( r \) contain the atomic numbers and coordinates, respectively, for the entire molecule. \( E_{\text{int}} \) and \( E_{\beta} \) 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.

2.2.1. Atomic Descriptor. The descriptor we use in this paper is essentially a variant of SOAP, 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, 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 on 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, and we intend to explore this in future work (see Conclusions below). Our proxy functions take the form

\[
\rho_{\mu,t}(r) = \sum_{i \in A_\mu} \eta_{\mu,i}^t B \exp(-\gamma (r - r_i)^2)
\]

where \( t \) is either \( \nu \) (for nuclear), \( \alpha \), or \( \beta \) corresponding to the type of point population, and thus \( \eta_{\nu,i}^t = Z_i / \eta_{\nu,i}^t = p_{\alpha,i}^t \) and \( \eta_{\nu,i}^t = p_{\beta,i}^t \). 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 \( \rho_{\mu,t}(r) \) onto the pertinent set of spherical harmonics \( Y_{\ell m}(\hat{r}) \)

\[
\rho_{\mu,t}(r) = \sum_{i \in A_\mu} \sum_{\ell m} c_{\ell m,i}^t(r) Y_{\ell m}(\hat{r})
\]

where \( \hat{r} \) denotes the unit vector corresponding to \( r \). This is done separately for each proxy function, that is, 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 \( \eta_{\nu,i}^t \) does not depend directly on positions, the analytical form of the radial terms can be derived following previous works as

\[
c_{\ell m,i}^t(r) \equiv 4 \pi \eta_{\nu,i}^t B \exp(-\gamma (r^2 + r_i^2)) \ell_l(2\pi\eta_{\nu,i}^t) Y_{\ell m}(\hat{r})
\]

where the asterisk (*) denotes complex conjugation and \( \ell_l \) are the modified spherical Bessel functions of the first kind. Summing over pertinent atoms gives

\[
c_{\ell m}^t(r) = \sum_{i \in A_\mu} c_{\ell m,i}^t(r)
\]

and projecting these onto RBFs gives a sequence of expansion coefficients \( c_{\ell m}^n \) with an extra index \( n \). The projections are based on evaluation of the integrals \( \langle g, Y_{\ell m}(\hat{r}) \rangle \), with \( g \) 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 \( c_{\ell m}^n(r) \) on a grid of \( r \) values. So for us, \( c_{\ell m}^n = c_{\ell m}^n(r) \), 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 \( m \),

\[
\rho_{\mu,t}^{\ell m} = \sum_{m=-l}^l |c_{\ell m}^n|^2
\]

It is easy to prove the rotational invariance of \( \rho_{\mu,t}^{\ell m} \) as 3D rotations correspond to transformations of blocks of the \( c_{\ell m}^n \) 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, 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, that is, the power spectrum is split into different element channels. There is active research aimed at reducing the cost associated with adding more atomic species, 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, that is, when the proxy function is expanded relative to a given atom, the summation found in (this instance of) eq 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. 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 eq 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 Li₂Hₙ clusters than for some other ionic systems.¹¹¹

2.2.2. 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, 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 toward the DFT atomization energy, as opposed to total energy. This has been shown to accelerate and balance training.¹⁴ 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_{intra}[Z, p, 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

\[ E_{\text{intra,Li}} = 0.0901q_i^2 + 0.1088q_i - 7.4787 \]  
\[ E_{\text{intra,H}} = 0.2471q_i^2 + 0.2403q_i - 0.4805 \]  

where \( q_i \) is the atomic partial charge for atom \( i \), that is, \( q_i = Z_i - p_{i\alpha} - p_{i\beta} \). This resembles the intra-atomic term in charge equilibration models.⁹² 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 noninteger 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 model in this case will also exhibit this problem. As density functionals with such noninteger 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 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.

\[ E_q[Z, p, r] = \sum_{j>i} \frac{q_i q_j}{\kappa r_{ij}} \]  

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

\[ L = \sum_M \frac{(F_{M}^{\text{DFT}} - F_{M}^{\text{ML}})^2}{N_{\text{atom},M}} + \gamma_1 \sum_M \frac{||F_{M}^{\text{DFT}} - F_{M}^{\text{ML}}||^2}{N_{\text{atom},M}} + \gamma_2 \sum_M \frac{||F_{M}^{\text{DFT}} - F_{M}^{\text{ML}}||^2}{N_{\text{atom},M}} \]  

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 \( F \) is the position gradient, \( q \) is the population gradient, and \( N_{\text{atom},M} \) represents the number of atoms in molecule \( M \). Adaptive moment solver (Adam)¹¹³ was used to update the NN weights during training. Our implementation is built up from the open-source package TensorMol,¹⁴ which takes advantage of the automatic differentiation scheme of Tensorflow.¹⁴}

3. RESULTS AND DISCUSSION

We here test the performance of our method on Li₂Hₙ 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 theoretical¹¹⁵–¹¹⁷ and experimental¹¹¹–¹¹² works and have implications in various applications including hydrogen storage.¹²³–¹²⁷

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.

3.1. Training-Data Generation and Model Training. In the training data set, the cluster sizes vary from \( n = 1 \) to 24, with increments of 3 or 4. We use two basic types of structures from which we obtain more samples: (i) collections of relatively widely separated LiH molecules (2–5 Å between molecules), and (ii) denser Li₂Hₙ systems with roughly cubic structures with alternating Li and H. We denote the first data set 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 data set. For the nonsparse data, structures both from geometry-optimization trajectories and ab initio molecular dynamics (AIMD) trajectories were used. The AIMD simulations used a temperature of 500 K and a time step of 20 au (≈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 nonsparse data. All of these geometries exhibit alternating Li and H atoms, that is, no structures with Li clusters nor H₂ 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 data set are included in the SI.

All DFT computations were conducted using the QChem¹⁰⁷ program. The geometry-optimization trajectories for the sparse data used the B3LYP¹²⁹,¹³⁰ functional with D3(BJ) dispersion,¹³¹ the AIMD trajectories for the nonspase data used the ωB97X-D¹³² functional, and the geometry-optimization trajectories for the nonspase 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.133

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 A for the latter).

The training data (~37 000 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 ~500 epochs of training. The mean absolute error (MAE) on the independent validation set of the energy is 1.523 kcal/mol per molecule.

3.2. Testing the Model. 3.2.1. \(Li_{29}H_{29}\) and \(Li_{32}H_{32}\) Clusters. We generated two test sets with larger molecules to test the transferability of our model: nonsparse \(Li_{32}H_{32}\) (+1,0,−1)
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, that is, 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

### Table I. Energy and Charge Errors for Li$_{29}$H$_{29}$ Sparse and Li$_{32}$H$_{32}$ Non-sparse Test Sets

<table>
<thead>
<tr>
<th></th>
<th>MAE charge (e/atom)</th>
<th>MAE energy (kcal/mol/atom)</th>
<th>MAE charge (e/atom)</th>
<th>MAE energy (kcal/mol/atom)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Li$<em>{29}$H$</em>{29}$ neutral</td>
<td>0.012</td>
<td>0.260</td>
<td>Li$<em>{32}$H$</em>{32}$ neutral</td>
<td>0.008</td>
</tr>
<tr>
<td>Li$<em>{29}$H$</em>{29}$ cation</td>
<td>0.014</td>
<td>0.169</td>
<td>Li$<em>{32}$H$</em>{32}$ cation</td>
<td>0.011</td>
</tr>
<tr>
<td>Li$<em>{29}$H$</em>{29}$ anion</td>
<td>0.016</td>
<td>0.151</td>
<td>Li$<em>{32}$H$</em>{32}$ anion</td>
<td>0.014</td>
</tr>
</tbody>
</table>

![Figure 3. Charge optimization result for a neutral Li$_{32}$H$_{32}$.](image)

**Li$_{32}$H$_{32}$ non-sparse neutral**

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

(b) Charge error for a neutral Li$_{32}$H$_{32}$
Li$_{29}$H$_{29}$ and Li$_{32}$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 data set 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 Li$_{29}$H$_{29}$ test set (Figure 2a), the BpopNN energies are showing close agreement with the DFT reference energies, with a charge MAE of $\sim0.014$ per atom, and an energy MAE of $\sim0.20$ kcal/mol per atom. Nonspare Li$_{32}$H$_{32}$ (Figure 2b) 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 $\sim0.011$ and energy MAE of $\sim0.27$ 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 Li$_{12}$H$_{12}$ 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, that is 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 Li$_{12}$H$_{12}$ structures were optimized using BpopNN separately for cation, neutral, and anion, and compared with the DFT optimized geometries. Figure 4 shows the superposition of BpopNN and DFT optimized neutral Li$_{12}$H$_{12}$ structures. The equilibrium geometries were successfully predicted with a root-mean-square displacement (RMSD) of 0.0522, 0.0958, and 0.0516 Å for neutral, cation, and anion, respectively, compared to the true DFT geometries. The superpositions and RMSD values were obtained with Maestro.$^{134}$ 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 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 data set 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$^{135}$ to obtain a few variations of Gaussian approximation potential (GAP) models,$^{30}$ 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 $\sim1100$ data points (for each overall charge). For neutral and anion, the GAP energy errors for the sparse (Li$_{29}$H$_{29}$) and nonspare (Li$_{32}$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 Figure S2, 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 Li$_{12}$H$_{12}$ are less accurate than those of BpopNN in both energy and structure. Details for the above results are given in the SI.

### 3.2.2. Reaction Pathways.

As another practical application of our model, we sketched the energy landscape of various structural transformations for neutral and anionic Li$_{12}$H$_{12}$. Diverse stationary-point structures were found for this system, ranging from cuboid structures to hexagonal structures to structures with fused 6–4–6–4–6 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 Li$_{12}$H$_{12}$ structures is shown in Figure S and for anion Li$_{12}$H$_{12}$ in Figure 6. All structures shown were

**Table II. Energy Statistics for a Cubic Li$_{32}$H$_{32}$ after Geometry Optimization**

<table>
<thead>
<tr>
<th></th>
<th>DFT energy (Hartree)</th>
<th>BpopNN energy (Hartree)</th>
<th>energy error (kcal/mol/atom)</th>
</tr>
</thead>
<tbody>
<tr>
<td>neutral</td>
<td>$-260.956$</td>
<td>$-260.967$</td>
<td>$0.106$</td>
</tr>
<tr>
<td>cation</td>
<td>$-260.711$</td>
<td>$-260.716$</td>
<td>$0.046$</td>
</tr>
<tr>
<td>anion</td>
<td>$-260.962$</td>
<td>$-260.983$</td>
<td>$0.206$</td>
</tr>
<tr>
<td>IP</td>
<td>6.656</td>
<td>6.823</td>
<td>0.172</td>
</tr>
<tr>
<td>EA</td>
<td>0.172</td>
<td>0.450</td>
<td>0.056</td>
</tr>
</tbody>
</table>

Figure 4. Geometry optimization on a cubic Li$_{32}$H$_{32}$ structure.
individually optimized, for example, 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 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 ↔ TS 3 ↔ Struct 4, neutral Struct 3 ↔ TS 6 ↔ Struct 7, anion Struct 2 ↔ TS 2 ↔ Struct 3). Some caution needs to be exercised if one were to predict the lowest energy structure from ML.

Figure 5. Energy profile for various structural transformations for neutral Li$_{12}$H$_{12}$. Superpositions and RMSD values were obtained with Maestro.$^{134}$

Figure 6. Energy profile for various structural transformations for anion Li$_{12}$H$_{12}$. Superpositions and RMSD values were obtained with Maestro.$^{134}$
models, especially when the energy range is extremely small (see anion Structure 1 vs Structure 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.

3.2.3. 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 Li15H15 cluster, which is the side-by-side interaction of two sheets, one Li8H7 (A) and one Li7H8 (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 intersheet 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 does not “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 intramonomer energies would be incorrect, and the intermonomer 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 A+ and 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 au, which actually is slightly above the sum of BpopNN energies for A and B, −121.4785 au. 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).

4. CONCLUSIONS

In this Article, 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, that is, partial charges of this model can also be used by themselves to infer molecular properties, for example, determining nucleophilic and electrophilic sites.

As a proof of concept for this approach, we trained a BpopNN model for LiH6 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 LiHn 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 toward 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 populations 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. We could alternatively modify the descriptor to be explicitly nonlocal. 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 nonlocal, 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: 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 data set. The data set used here is on the smaller side, so enlarging this is a focal point for future work. There is also the plausibility that our data set is insufficiently balanced. More specifically, it only contains alternating Li–H clusters and cannot be reasonably applied to clusters with distinguishable Liₙ and (H₂)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.

## APPENDIX A: POPULATION GRADIENT OF CDFT ENERGY

In the present context, the CDFT Lagrangian takes the form

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

where \( \gamma \) is the density matrix, the \( \lambda_{i,\sigma} \) are Lagrange multipliers, the \( p_{i,\sigma}(\rho) \) are the functions defined in eq 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 \( \lambda_{i,\sigma} \) and \( c_{i,\sigma} \) will adjust accordingly. This Lagrangian also formally applies to cases in which fewer population constraints are used. In that case, we simply set the \( c_{i,\sigma} \) for any technically unconstrained population to its relaxed value, and the associated \( \lambda_{i,\sigma} \) is 0.

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

$$\frac{\partial E_{CDFT}}{\partial c_{i,\sigma}} = \frac{\partial L_{CDFT}}{\partial c_{i,\sigma}} = -\lambda_{i,\sigma}$$  \hspace{1cm} (A2)

The target energy function \( E[p] \) is essentially the same as \( E_{CDFT}[\{c_{i,\sigma}\}] \) except that it is a function of all \( 2N_{atom} \) 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, \( c_{i,\sigma} \) and then vary \( c_{i,\sigma} \). Upon (re)optimizing \( L_{CDFT} \) for this new set of constraints, \( p_{i,\sigma} \) changes identically as \( c_{i,\sigma} \) does, and \( p_{N_{atom},\sigma} \) changes by the opposite amount, while all other populations are unchanged. Denoting the direction corresponding to this overall change in populations by \( e_{i,\sigma} \) we thus have

$$\Delta E[p] e_{i,\sigma} = \frac{\partial E_{CDFT}}{\partial c_{i,\sigma}} = -\lambda_{i,\sigma}$$  \hspace{1cm} (A3)

The vectors \( e_{i,\sigma} \) (most of) the columns of a block diagonal matrix \( W \) with two \( N_{atom} \) by \( N_{atom} \) blocks, one for each spin type. That is, \( W = W_{\uparrow} \oplus W_{\downarrow} \) with

$$W_{\uparrow} = \begin{bmatrix} 1 & 0 & \cdots & 0 & 1 \\ 0 & 1 & \cdots & 0 & 1 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & 1 \\ -1 & -1 & \cdots & -1 & -1 \end{bmatrix}$$  \hspace{1cm} (A4)

The last column in \( W_{\uparrow} \) 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 \( V \Delta E_{\downarrow} \) along this direction is 0. We, thus, have

$$V \Delta E_{\downarrow} W = \nu$$  \hspace{1cm} (A5)

where the latter vector contains the \( -\lambda_{i,\sigma} \) along with two 0s for the symmetric directions

$$\nu = \begin{cases} -\lambda_{i,\sigma} & \text{if } i < N_{atom} \\ 0 & \text{if } i = N_{atom} \\ -\lambda_{N_{atom},\sigma} & \text{if } N_{atom} < i < 2N_{atom} \\ 0 & \text{if } i = 2N_{atom} \end{cases}$$  \hspace{1cm} (A6)

By inverting \( W_{\downarrow} \) we readily obtain \( V \Delta E_{\downarrow} \) (in the regular population basis).

## ASSOCIATED CONTENT

**Supporting Information**

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jctc.0c00217.

Test-set results for the other BpopNN variants; details for training the GAP models with QUIP; test-set results for the GAP model variants (PDF).

All training-set data, including CDFT energies, charges, population gradients, and nuclear-position forces; Geometries used for all training and test data sets (ZIP).
AUTHOR INFORMATION

Corresponding Author
David W. Small — Department of Chemistry and Molecular Graphics and Computation Facility, College of Chemistry, University of California, Berkeley, California 94720, United States; orcid.org/0000-0002-3316-1885; Email: dsmallchem@gmail.com

Authors
Xiaowei Xie — Department of Chemistry, University of California, Berkeley, California 94720, United States; Energy Technologies Area, Lawrence Berkeley National Laboratory, Berkeley, California 94720, United States; orcid.org/0000-0001-5618-8768

Kristin A. Persson — Department of Materials Science and Engineering, University of California, Berkeley, California 94720, United States; Energy Technologies Area, Lawrence Berkeley National Laboratory, Berkeley, California 94720, United States; orcid.org/0000-0003-2495-5509

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.jctc.0c00217

Notes
The authors declare no competing financial interest.

ACKNOWLEDGMENTS

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 at the UC Berkeley Molecular Graphics and Computation Facility (MGCF). MGCF is supported by grant NIH S10OD023532.

REFERENCES


