On the role of long-range electrostatics in machine-learned interatomic potentials for complex battery materials

Carsten G. Staacke,[1,2] Hendrik H. Heenen,[2] Christoph Scheurer,[1,2] Gábor Csányi,[3] Karsten Reuter,[1,2] and Johannes T. Margraf[1,2]

[1] *Chair for Theoretical Chemistry and Catalysis Research Center, Technische Universität München, Lichtenbergstraße 4, D-85747 Garching, Germany*

[2] *Fritz-Haber-Institut der Max-Planck-Gesellschaft, Faradayweg 4-6, D-14195 Berlin, Germany*

[3] *Engineering Laboratory, University of Cambridge, Cambridge CB2 1PZ, United Kingdom*

Modeling complex energy materials such as solid-state electrolytes (SSEs) realistically at the atomistic level strains the capabilities of state-of-the-art theoretical approaches. On one hand, the system sizes and simulation time scales required are prohibitive for first-principles methods like density functional theory (DFT). On the other hand, parameterizations for empirical potentials are often not available and these potentials may ultimately lack the desired predictive accuracy. Fortunately, modern machine learning (ML) potentials are increasingly able to bridge this gap, promising first-principles accuracy at a much reduced computational cost. However, the local nature of these ML potentials typically means that long-range contributions arising, e.g., from electrostatic interactions are neglected. Clearly, such interactions can be large in polar materials like electrolytes, however. Herein, we investigate the effect that the locality assumption of ML potentials has on lithium mobility and defect formation energies in the SSE $Li_7P_3S_{11}$. We find that neglecting long-range electrostatics is unproblematic for the description of lithium transport in the isotropic bulk. In contrast (field-dependent) defect formation energies are only adequately captured by a hybrid potential combining ML and a physical model of electrostatic interactions. Broader implications for ML based modelling of energy materials are discussed.

I. INTRODUCTION

The development of new analytical approximation frameworks is currently leading to an unparalleled surge of machine learning (ML) approaches in all areas of chemistry and materials science.[1–6] Here, ML is typically considered a universal approach for learning (fitting) a complex relationship $y = f(x)$ without explicitly knowing the physical (analytic) form of f.[7,8] In the context of interatomic potentials, this means establishing the relationship between a system's atomistic structure and its total energy $E = f(\{\mathbf{Z}, \mathbf{R}\})$, where $\mathbf{Z}$ are the atomic numbers and $\mathbf{R}$ the position vectors of the constituting atoms. The expectation here is that flexible ML potentials can overcome long-standing limitations of empirical potentials that use simple fixed functional forms.[8]

Such limitations are especially acute when covalent bonds are formed or broken, when atoms vary their hybridization or charge state and generally when large changes in chemical environments occur. All these aspects apply prominently to the simulation of *operando* energy conversion systems in general and battery materials in particular.[3,9–16] With the structural and compositional complexity of contemporary battery materials severely limiting direct first-principles based simulations, there is thus considerable hope that ML potentials trained with first-principles data will enable simulations at unprecedented length and time scales and a predictive quality matching electronic structure methods.[17–24]

To achieve size-extensivity and create a general ML potential that can be employed for systems of varying size and composition, just as with many empirical potentials (e.g. EAM, Tersoff) a locality assumption is typically made.[7,25–28] The system's total energy is thus approximated as a sum of local (atomic) contributions:

$$E = \sum_{i}^{N} \epsilon(Z_i, \chi_i) \tag{1}$$

where the sum runs over the N atoms in the system and each atom i contributes with an energy ϵ that only depends on its atomic number Z_i and its local chemical environment χ_i (which is itself a function of $\{\mathbf{Z}, \mathbf{R}\}$). This local environment is suitably encoded into a representation that (like the total energy) obeys general symmetries like invariance to translation, rotation and permutation of atoms of the same element.[7,27,29–31] Importantly, to allow for efficient training and scalability to large simulation cells, these representations are almost always short-ranged, i.e. they only describe the environment within a few Ångstrom

around each atom.

In recent years, a large variety of ML potentials using kernel or neural network based regression has been developed for molecular and condensed systems, heralding the great potential of this new data-driven approach.[4,7,27,32] Notably, such short-ranged ML potentials have also been applied to polar systems like water.[33–36] Though naively one would expect long-range electrostatic interactions to play a significant role here, these potentials provide a remarkably accurate description of the structural and dynamic properties of bulk liquid water and different ice phases.[37] This is presumably because these systems are highly isotropic, so that long-range interactions average out.[38] Consequently, short-ranged ML potentials are now commonly applied to study polar and even ionic systems.[34,39–41]

However, the importance of long-range effects will clearly depend on the material and property of interest and thus demands more systematic scrutiny.[42] For example, ionic diffusion in electrolytes may lead to the transient local accumulation and depletion of charges, which can break the isotropy of the electrostatic environment. Even more critically, grain-boundaries, interfaces, and defects may lead to a permanent localized polarization of materials. Finally, the effect of applied electric fields, e.g. in batteries, can obviously only be studied if an electrostatic description is part of the model.

Notably, several groups have recently proposed ML models that explicitly include long-range electrostatics.[43–50] This ranges from simple point-charge models, to polarizable models to full self-consistent approximations of the charge density. While these approaches offer a route to overcome the locality constraints of current ML potentials, they also lead to an increased computational cost, both in terms of training and evaluation. In particular, they break the favourable linear scaling of computational cost with system size. This makes it crucial to understand when such explicit treatments of electrostatics are necessary and when a local ML potential can be used instead.

The goal of this paper is to analyze the effect of electrostatic contributions to ML based simulations of battery materials, using the Li mobility and interface stability of the crystalline phase of the solid-state electrolyte (SSE) $Li_7P_3S_{11}$ as an example.[51,52] $Li_7P_3S_{11}$ exhibits an exceptionally high Li ion conductivity and has been suggested as a promising candidate for all-solid-state lithium batteries[51,52]. Large scale simulations of this material, possible e.g. through Gaussian Approximation Potentials (GAPs), can provide critical insight towards solving open challenges in connection with this material.

We develop two GAP models where one is strictly short-ranged (GAP) while the other includes a simple electrostatic baseline (ES-GAP). We find that the inclusion of long-range ES interactions only benefits the description of non-isotropic chemical environments, while diffusion properties in the homogeneous bulk material are well captured by both potentials. In contrast, the stability of Frenkel defects in the presence of electric fields (which influence the material's stability at the inhomogeneous electrode/electrolyte interface) can only be captured with a model that includes long-range interactions.

II. METHODS

Computational Details Reference DFT calculations are performed with the PBE functional, default 'light' integration grids and a 'tier 1' basis set of numerical atomic orbitals, as implemented in FHI-aims.[53,54] The Brillouin zone is sampled with a 2x2x2 k-grid. Initial training configurations are generated with ab initio molecular dynamics (MD) using the Γ-point approximation for the k-grid. GAP-based MD and Nudged-Elastic-Band (NEB) simulations are performed using the LAMMPS[55] code and the corresponding interface to QUIP.[56] Pairwise electrostatic interactions in the ES-GAP model are included via a fixed charge model. To avoid the divergence of point-charge Coulomb interaction at short distances, atomic charge densities are modeled by s-type Slater orbitals as, e.g., in the QEq charge equilibration model.[57,58] Further details on the ES-GAP are noted in section A of the SI. For training set construction and data analysis, the atomic simulation environment ASE, SciPy and scikit-learn are used.[59–61]

GAP training: To train the GAP models, a simple iterative procedure is used. Briefly, an initial model is trained on a set of 80 crystalline $Li_7P_3S_{11}$ configurations, taken from a short DFT-based MD simulation and Monte Carlo sampled Li-ion distributions on crystal and interstitial sites. This potential is then used to generate new configurations via molecular dynamics (MD) simulations at 800 K, which are added to the training set. This procedure is repeated for several iterations (termed "generations") until the force and energy errors on new configurations no longer improve. The thus obtained models provide an increasingly accurate description of high-temperature crystalline $Li_7P_3S_{11}$. GAP and ES-GAP models are trained on identical configurations and the ES-GAP model was used to generate new configurations in the iterative procedure. Further Details can be found in the SI section C.

III. RESULTS

A. GAP and ES-GAP Models

As discussed above, a hallmark of many-body ML potentials is the assumption that the total energy can be described as a sum of local atomic contributions, which corresponds to a complete neglect of long-range interactions. The locality of these interactions is thus often tacitly assumed when ML potentials are generated. However, it can also be quantified more rigorously by analyzing the force induced on a reference atom by perturbations of other atoms in the distance beyond a given cutoff radius.[62] This "locality" analysis is shown for $Li_7P_3S_{11}$ in Fig. 1A (see SI for details).

As can be seen, the induced forces are quite large (between 0.1 and 0.5 eV/Å) and decay slowly with the cutoff radius. This is particularly evident for the phosphorous atoms, which bear the largest formal charge in this system (+5). Since these forces originate from perturbations outside of a given radius, they cannot be described by an ML potential with the corresponding cutoff. The locality test therefore provides a lower bound for the residual force errors that a ML potential can achieve.

In Fig. 1B, the same locality test is performed after subtracting the fixed-charge ES baseline model from the DFT forces (see SI for details). For cutoffs larger than 5 Å, this significantly lowers the locality error, most prominently for phosphorous. Counterintuitively, the phosphorous and sulfur errors are actually increased at shorter distances. On one hand, this is because the charges of this model were parameterized to minimize the locality error at 6 Å, and are thus not ideal for shorter cutoffs. On the other hand, a fixed-charge ES model is generally inaccurate for short-range interactions, where polarization, charge-transfer and induction effects become important. To capture such effects with a baseline model one would require the use of more complex polarizable models. Nonetheless, cutoffs of 6 Å are commonly used in state-of-the-art ML potentials so that a fixed charge ES baseline can be used here, though residual errors remain and are discussed below.

The convergence of the iterative training procedure can be seen in Fig. 1C-D. This shows that the energy and force errors of the ES-GAP model show no further improvement between the fifth and sixth generation. Importantly, the final force errors fall into the expected range estimated from the locality test (see Fig. 1B). This indicates convergence of the train-
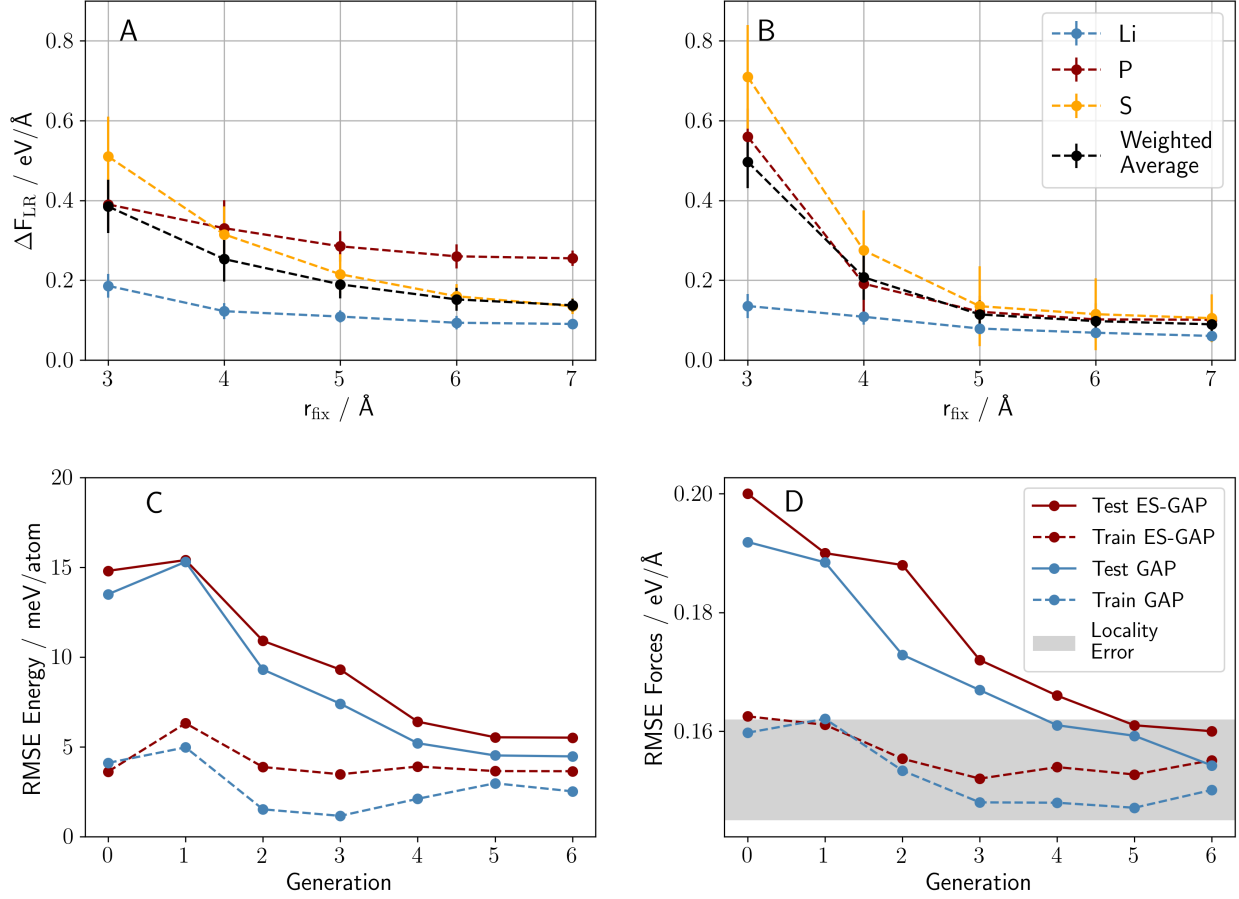
FIG. 1. A) Force locality in crystalline $Li_7P_3S_{11}$. B) Force locality in crystalline $Li_7P_3S_{11}$ after subtracting the electrostatic baseline model. Convergence of Energies (C) and Forces (D) throughout the generations for both models (GAP: blue and ES-GAP: red). Solid lines correspond to test errors, dashed lines to training errors. The grey shading corresponds to the expected force accuracy according to the locality test.

ing process, meaning that the remaining error will not be significantly reduced by further training, but is instead related to the locality of the model and/or potential inadequacies of the representation.

Interestingly, the root mean squared errors (RMSE) for predicted energies and forces are actually slightly lower for the short range GAP model. By analyzing the errors of the individual elements separately (see SI), we find that the ES-GAP displays somewhat higher errors for sulfur but lower errors for lithium and especially phosphorous. Since sulfur is the most abundant element in $Li_7P_3S_{11}$, this leads to the better average performance of the short range GAP.

6

This points to a disadvantage of the fixed charge approach used here: $Li_7P_3S_{11}$ features two distinct sulfur species, namely in bridging and terminal positions. Hirshfeld population analysis (see SI) indicates that these species correspond to different charge states, while they are treated equivalently by the ES baseline. This introduces an error, which needs to be compensated by the GAP potential. In principle, this could be mitigated by assigning different charges to these sulfur species. However, such atom-typing would run counter to one of the main advantage of ML potentials relative to classical force-fields, namely the fact that they can break and form bonds. A more satisfying solution would be the use of floating charge models and this will be explored in future work.[47,63]

From a different perspective, these unbalanced force errors also illustrate a weakness in average error metrics like the RMSE (or the least-squares loss function minimized by GAP) for multi-species systems: If the stoichiometry of a material is not balanced, more abundant species are implicitly weighted more strongly by the metric. In the present case, sulfur has the largest weight, although lithium is arguably more important. Nevertheless, the accuracy of both models is actually quite satisfying overall considering the magnitudes of the force components in the training and test sets, which range up to ca. 10 eV/Å. Note however, that in principle it would be possible to use different weightings in the loss function for forces on different elements.

B.   Lithium Ion Mobility

Both potentials introduced in the previous section are trained on the same data and have approximately the same force error, though small differences can be seen when looking at the description of individual elements. Do these differences affect the prediction of observables relevant to battery performance? As a first case in point, we investigate Li ion mobility in the (isotropic) bulk material. To this end we consider lithium diffusion barriers obtained via the nudged elastic band (NEB) method (see Fig. 2) and Li-ion mobilities obtained from MD simulation.

NEB calculations allow investigating minimum energy paths of individual Li hops between two equilibrium positions (an initial and final state).[64] Choosing these states is actually non-trivial, since $Li_7P_3S_{11}$ contains a large variety of possible Li interstitial positions, reflecting the highly dynamic nature of the Li sublattice. This was previously demonstrated by Chang

et al., who reported a number of Li configurations with nearly the same ground-state energy as the equilibrium crystal structure.[65,66] It is therefore important to focus on Li hopping events that actually contribute to conductivity and not just dynamic rearrangements of Li positions.

To obtain these relevant pathways, we therefore analyzed the training MD trajectories to isolate individual Li hopping events. In this manner, a variety of diffusion pathways were obtained. Further information on all pathways is given in the SI (section G). In the following, the lowest barrier pathway is discussed in more detail. Here, a Li ion diffuses in a channel formed by the PS_4^- and $P_2S_7^-$ anion complexes (positions tabulated in the SI G), along the b lattice vector, as shown in Fig. 2. It can be seen that the 1x2x1 supercell is traversed with two consecutive Li ion hops (obtained from two NEB calculations Fig. 2 I.+II.). These NEB calculations were performed based on both the GAP and ES-GAP potentials, leading to slightly different minimum energy paths.

Notwithstanding, the optimized lowest energy paths display similar characteristics of a correlated ion migration where lithium ions diffuse, while neighboring ions are slightly displaced from the diffusion path (highlighted in the figure with solid and dashed arrows for the migrating and displaced atoms, respectively). For the first hop an almost identical path is found while the second hop yields a slightly different path when optimized with the two potentials (I. and II. in Fig. 2, respectively). The deviating paths should not be understood as different mechanisms favored by the respective potentials, however, but merely two feasible paths found by the NEB. We confirm this assumption by evaluating the energies along the short-range GAP path with the ES-GAP (and vice-versa). This analysis reveals almost identical barrier heights for a given potential on both its own NEB path and the one from the other potential. It is further noteworthy that the ES-GAP consistently predicts somewhat higher barriers than the short-range GAP. For both paths, reference DFT single-point values tend to lie between the GAP and ES-GAP values. In other words, GAP somewhat underestimates the barriers, while ES-GAP overestimates them to a similar degree. The deviation in energies is also not perfectly uniform along the paths, so that the agreement with DFT can be excellent for both potentials, at different points of the potential energy surface.

Despite the overall similar performance of the models, we thus find small systematic differences between the predictions of the two models. Interestingly, the short range GAP is
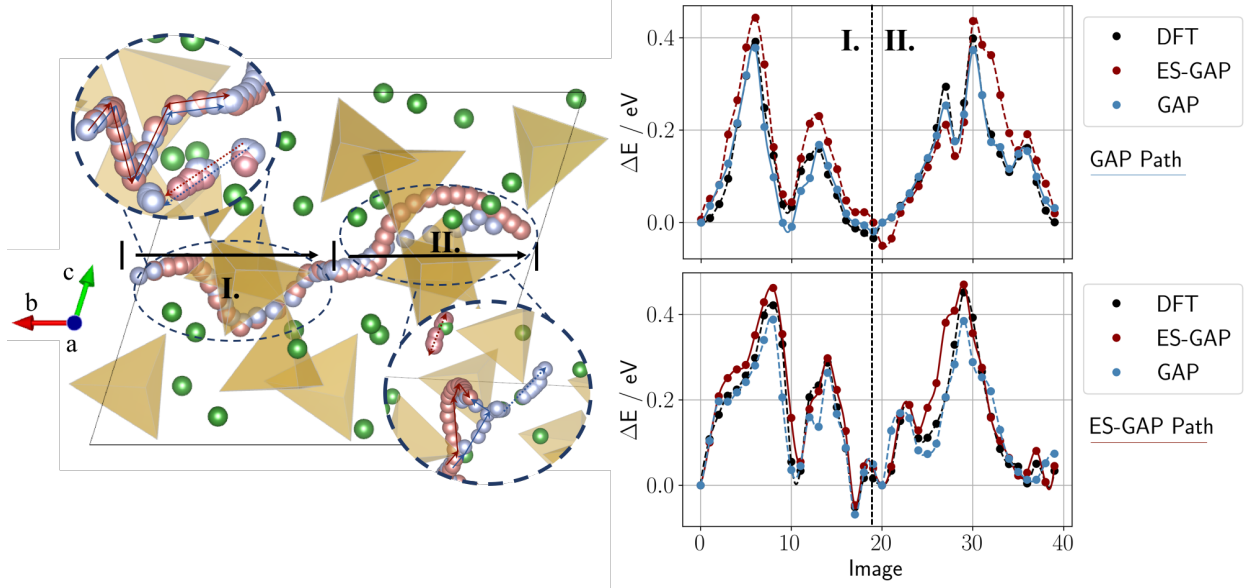
FIG. 2. Left: Illustration of minimum energy paths for lithium diffusion through $Li_7P_3S_{11}$ obtained with short range GAP (light blue) and ES-GAP (red) interatomic potentials (Li-green, Thiophosphates-orange tetrahedron). Solid and dashed arrows highlight the main migration path and displacements of neighboring lithium atoms, respectively. I. and II. correspond to two consecutive NEB calculations. Initial and final positions are obtained by analyzing hopping events from MD trajectories (see SI G). Right: Energies for DFT, GAP and ES-GAP potentials on the minimum energy paths calculated with the short range GAP (top) and the ES-GAP (bottom). Solid lines indicate the potential with which the path was obtained, dashed line to single point calculations.

actually slightly more accurate in predicting the lowest barrier heights. This indicates that the static charge model used for the ES-GAP does not faithfully reflect the electrostatics of the full DFT calculation, and the ES baseline represents an overcorrection: it correctly raises the barriers, but by too much. As a sidenote, we emphasize that both models are mainly trained on high-temperature MD data, while the NEB corresponds to the minimum energy path at 0 K. Presumably, even higher accuracy for NEB calculations could be achieved for both potentials by training on the corresponding data.

Next, we turn to the Li ion conductivity σ at finite temperature, predicted from MD simulations via the Nernst-Einstein equation (see SI for details).[67] Here we shift the focus from a microscopic property (the Li migration barrier height) to a macroscopic observable (Li-ion

9

conductivity). In principle, the two are closely related, since the barrier height determines the rate of the Li transport in transition state theory. However, MD simulations sample a multitude of different diffusion events. These are often dominated by the lowest barrier mechanism, but may also be influenced by higher barrier pathways (e.g. because they are entropically favoured). In this sense the NEB and MD simulations provide complementary information.

The corresponding conductivities within the two potentials (at 800 K) are shown in Fig. 3, as a function of the MD trajectory length. After sufficient sampling, the conductivities converge to 190 mS/cm for the GAP and 120 mS/cm for the ES-GAP, respectively. This trend in conductivities perfectly reflects the the NEB barrier differences discussed above (i.e. slightly higher mobility and lower barriers for the short-range GAP model).
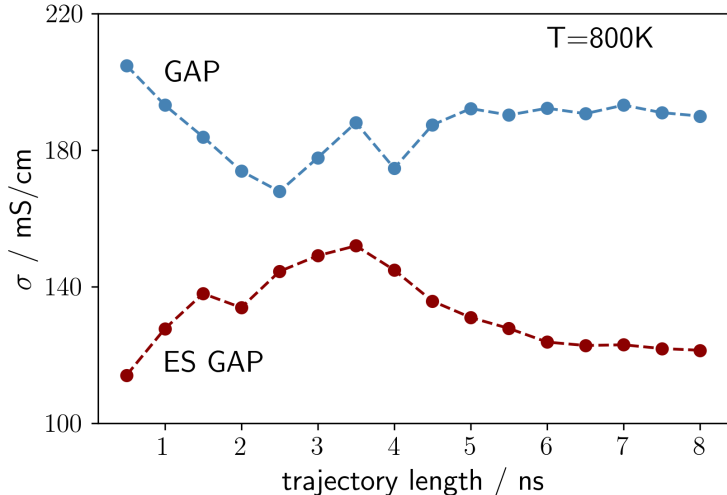

FIG. 3. Convergence of the Li ion conductivity σ at 800 K during an MD simulation. The red curve corresponds to the ES-GAP model, blue to the GAP model.

Figure 3 also highlights the benefit of using ML potentials for battery research more generally: To obtain fully converged conductivities from these simulations, MD trajectory lengths far beyond the tractability of typical AIMD simulations are necessary. Hence, AIMD simulations of Li ion conductivities should generally not be considered converged and yield, at best, a qualitative indicator of relative performance between closely related materials. Even at longer timescales in the low ns range (< 5 ns), conductivity differences predicted by both models vary from 10 - 80%.

While we do find some differences between the predicted Li mobilities of ES-GAP and

GAP, these should be put into perspective. Ion mobilities of potential SSE materials can vary by several orders of magnitude, and the small differences in observed barrier heights are certainly within the margin of the DFT error. From a practical perspective there is thus no significant difference between the two models.[68] At this point, one could conclude from a practitioner's point of view that a short-ranged ML potential would be fully sufficient to treat a complex battery material like $Li_7P_3S_{11}$.

C.   Applied Fields and Defects

Having established the similar behavior of GAP and ES-GAP for bulk Li mobilities, we now turn to their description of an anisotropic environment. Specifically, we consider a model system that mimics the effect of the potential drop in the interphase region at the solid/solid interface between an electrode and a SSE, as depicted in Fig. 4.[69] In a real battery, this potential drop reaches several nanometers into the bulk region before it is completely screened by displaced ions. To model the effect of this potential drop on the SSE near the interface, we apply electric fields to a $2\times3\times2$ $Li_7P_3S_{11}$ supercell and investigate how the field strength and direction affects the stability of a Frenkel defect (i.e. a Li vacancy/interstitial pair).

Defects have been argued to play an important role in the kinetics of the decomposition processes at SSE interfaces.[70] This kinetic stability is of high relevance to the applicability of SSEs.[71–73] Here we consider a Frenkel defect in particular, as it allows keeping the simulation cell overall charge neutral and lies within the phase space covered by the training set of the GAP potentials. Clearly, the effect of an electric potential drop can only be captured by the ES-GAP model, which contains charges that are able to respond to the applied field. In contrast, the short range GAP model can only model the "zero-field" scenario. We also compare the "zero-field" defect stability in both models with that in DFT.
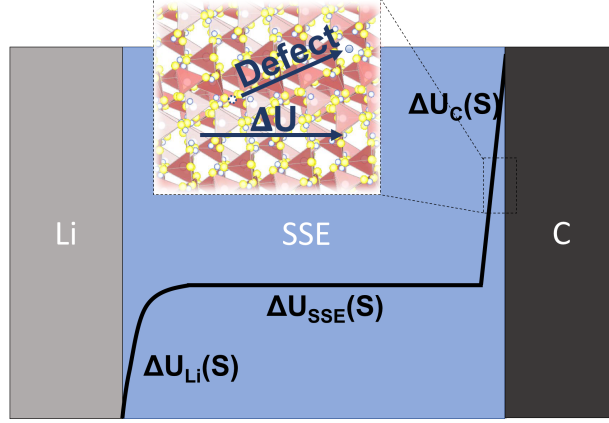
FIG. 4. Schematic of the spatially dependent potential variation U in a prototypical Li‖SSE‖C solid-state Li ion battery.[69] Since grain orientation varies throughout the SSE, the field dependent defect stability is studied by applying electric fields along different crystallographic axes of the simulation cell.

We construct the Frenkel defect by shifting a Li atom into an interstitial position and relaxing the resulting structure using the two GAP models (see Fig. 5, SI I). For comparison, single-point DFT calculations are also performed on both the GAP and ES-GAP geometries. Without an applied field, this leads to a predicted defect formation energy of ca. 0.8 eV with ES-GAP and DFT and ca. 0.6 eV with the short-range GAP model. Since the interstitial Li ion has a net positive charge and the ion vacancy a net negative charge, the defect forms a dipole. The observed 0.2 eV deviation of the defect formation energies between GAP and DFT can thus be attributed to the absence of long-range dipole-dipole interactions in the former. Nonetheless, all models predict defect formation to be highly unfavorable. When applying an electric field, this picture changes. We find an anisotropic response to the field where both destabilization and stabilization can occur (Fig. 5). This anisotropy reflects the fields' orientation relative to the defect dipole. Structural relaxation effects then lead to an unsymmetric stabilization/destabilization of the defect in either field direction.

While this is a rather simple model system, it already yields insights into the stability of $Li_7P_3S_{11}$ at the SSE/electrode interface. As shown in Fig. 5, field strengths typically occurring at this interface (which can reach up to 0.3 V/Å[69]) are sufficient to make the formation of this defect energetically favorable. Consequently, one would expect an accumulation of such defects towards the interface. As recently suggested[71], the kinetic processes

in electrolyte decomposition can be related to the delithiation of the solid state electrolyte. Intermediate to this delithiation process are local concentration gradients by Frenkel defects. We can therefore hypothesize from our findings that local fields play a crucial role in the evaluation of interphase stabilities. Further, we find that the defect stabilization is anisotropic to crystallographic orientation. This finding might explain previous observations that the SSE/anode interface stability of $Li_7P_3S_{11}$ was dependent on the crystallographic orientation of the latter.[74]
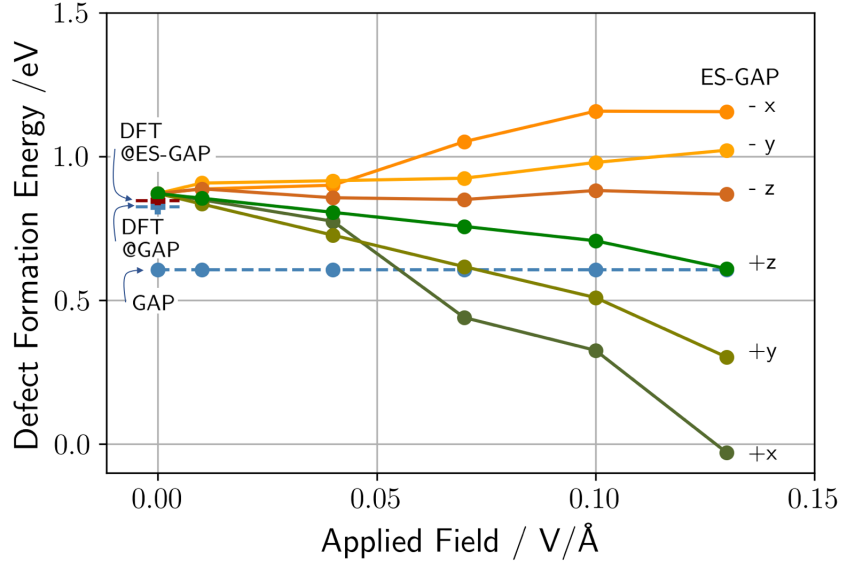
FIG. 5. Frenkel defect formation energies against applied field. The electric field is applied along the different crystallographic axes (x, y, z). Solid lines (green and orange) correspond to ES-GAP calculations with different applied field directions. Defect positions are given in the SI I.

From a methodological perspective, this example shows that the explicit inclusion of electrostatic interactions will be indispensable for the computational study of battery materials under operating conditions. Indeed, even the contact between two different materials will cause a potential drop across the interface, albeit at a smaller lengthscale.[14] The good performance of the short range GAP model in the previous section is thus not because long-range electrostatic interactions are small, but because they are reasonably isotropic in a periodic calculation. Breaking this symmetry with an interface or by applying an electric field clearly shows the importance of electrostatics which, by design, cannot be incorporated into a model with a short range cutoff.

13

We note that the response of the ES-GAP model to the applied field relies on the ionic partial charges of the baseline ES model. Hence, we exploited the corresponding *a priori* knowledge. Ideally, these charges could instead be determined in the training procedure including model structures of the full interface which, however, is beyond the scope of this conceptual study.

IV. CONCLUSION

In this paper, we have systematically explored the influence of explicitly including electrostatic interactions in ML potentials for battery materials. Using the same ML approach and training data, we find significant differences between a short range GAP model and the ES-GAP model that uses a electrostatic baseline when studying isotropic *vs.* non-isotropic systems. In standard isotropic simulation tasks, such as determining Li diffusion barriers and ionic conductivities, both models yield similar results. In contrast, simulations on non-isotropic systems show the importance of ES contributions and provide new insights into interphase stability of $Li_7P_3S_{11}$.

Specifically, we studied Frenkel defects in an applied field mimicking the potential drop at a solid/solid interface. In this setup we found that a stabilization of the defects can occur already at moderate fields. This would favor the accumulation of defects towards the interphase, which could influence the kinetic stability of $Li_7P_3S_{11}$/electrode interfaces. Additionally, such stabilizations are anisotropic to crystallographic orientation making grain shape and orientation an additional parameter to be considered in battery engineering and beyond.[75]

More generally, our results confirm that short-ranged ML potentials can be surprisingly accurate for polar and ionic materials in the absence of non-isotropic chemical environments like interfaces or electric fields. In contrast we found important qualitative deviation between our GAP models in non-isotropic systems. The further development of ML potentials with an explicit description of electrostatics therefore represents an important research goal, on the way to the computational study of battery materials in *operando* conditions.

REFERENCES

[1] O. A. von Lilienfeld, K.-R. Müller, and A. Tkatchenko, Nat. Rev. Chem. **4**, 347 (2020).

[2] B. Cheng, R.-R. Griffiths, S. Wengert, C. Kunkel, T. Stenczel, B. Zhu, V. L. Deringer, N. Bernstein, J. T. Margraf, K. Reuter, and G. Csanyi, Acc. Chem. Res. **53**, 1981 (2020).

[3] V. L. Deringer, J. Phys. Energy **2**, 041003 (2020).

[4] A. P. Bartók, S. De, C. Poelking, N. Bernstein, J. R. Kermode, G. Csányi, and M. Ceriotti, Sci. Adv. **3**, e1701816 (2017).

[5] S. Stocker, G. Csányi, K. Reuter, and J. T. Margraf, Nat. Commun. **11**, 1 (2020).

[6] F. Noé, A. Tkatchenko, K.-R. Müller, and C. Clementi, Annu. Rev. Phys. Chem. **71**, 361 (2020).

[7] A. P. Bartók, M. C. Payne, R. Kondor, and G. Csányi, Phys. Rev. Lett. **104**, 136403 (2010).

[8] J. Behler, J. Chem. Phys. **145**, 170901 (2016).

[9] N. Artrith, A. Urban, and G. Ceder, J. Chem. Phys. **148**, 241711 (2018).

[10] A. Urban, D.-H. Seo, and G. Ceder, Npj Comput. Mater. **2**, 16002 (2016).

[11] S. Stegmaier, R. Schierholz, I. Povstugar, J. Barthel, S. P. Rittmeyer, S. Yu, S. Wengert, S. Rostami, K. Hans, K. Reuter, R. Eichel, and C. Scheurer, Adv. Energy Mater , 2100707 (2021).

[12] H. H. Heenen, J. Voss, C. Scheurer, K. Reuter, and A. C. Luntz, J. Phys. Chem **10**, 2264 (2019).

[13] H. H. Heenen, C. Scheurer, and K. Reuter, Nano Lett. **17**, 3884 (2017).

[14] S. Stegmaier, J. Voss, K. Reuter, and A. C. Luntz, Chem. Mater. **29**, 4330 (2017).

[15] A. Bruix, J. T. Margraf, M. Andersen, and K. Reuter, Nat. Catal. **2**, 659 (2019).

[16] J. Timmermann, F. Kraushofer, N. Resch, P. Li, Y. Wang, Z. Mao, M. Riva, Y. Lee, C. Staacke, M. Schmid, C. Scheurer, G. S. Parkinson, U. Diebold, and K. Reuter, Phys. Rev. Lett. **125**, 206101 (2020).

[17] W. Li, Y. Ando, E. Minamitani, and S. Watanabe, J. Chem. Phys. **147**, 214106 (2017).

[18] B. Onat, E. D. Cubuk, B. D. Malone, and E. Kaxiras, Phys. Rev. B **97**, 094106 (2018).

[19] V. Lacivita, N. Artrith, and G. Ceder, Chem. Mater. **30**, 7077 (2018).

[20] Z. Deng, C. Chen, X.-G. Li, and S. P. Ong, Npj Comput. Mater. **5**, 75 (2019).

[21] C. Wang, K. Aoyagi, P. Wisesa, and T. Mueller, Chem. Mater. **32**, 3741 (2020).

[22]R. Jalem, K. Kanamori, I. Takeuchi, M. Nakayama, H. Yamasaki, and T. Saito, Sci. Rep. **8**, 5845 (2018).

[23]G. Houchins and V. Viswanathan, J. Chem. Phys. **153**, 054124 (2020).

[24]S. Fujikake, V. L. Deringer, T. H. Lee, M. Krynski, S. R. Elliott, and G. Csányi, J. Chem. Phys. **148**, 241714 (2018).

[25]M. S. Daw and M. I. Baskes, Phys. Rev. B **29**, 6443 (1984).

[26]J. Tersoff, Phys. Rev B **37**, 6991 (1988).

[27]J. Behler and M. Parrinello, Phys. Rev. Lett. **98**, 146401 (2007).

[28]H. Jung, S. Stocker, C. Kunkel, H. Oberhofer, B. Han, K. Reuter, and J. T. Margraf, ChemSystemsChem **1900052**, 201900052 (2020).

[29]R. Drautz, Phys. Rev. B **99**, 014104 (2019).

[30]F. A. Faber, A. S. Christensen, B. Huang, and O. A. von Lilienfeld, J. Chem. Phys. **148**, 241717 (2018).

[31]A. S. Christensen, L. A. Bratholm, F. A. Faber, and O. Anatole von Lilienfeld, J. Chem. Phys. **152**, 044107 (2020).

[32]A. V. Shapeev, Multiscale Model. Simul. **14**, 1153 (2016).

[33]B. Monserrat, J. G. Brandenburg, E. A. Engel, and B. Cheng, arXiv preprint arXiv:2006.13316 (2020).

[34]A. M. Cooper, J. Kästner, A. Urban, and N. Artrith, Npj Comput. Mater. **6**, 1 (2020).

[35]C. Schran, J. Behler, and D. Marx, J. Chem. Theory Comput. **16**, 88 (2020).

[36]C. Schran, F. Brieuc, and D. Marx, J. Chem. Phys. **154**, 051101 (2021).

[37]B. Cheng, E. A. Engel, J. Behler, C. Dellago, and M. Ceriotti, PNAS **116**, 1110 (2019).

[38]S. J. Cox, PNAS **117**, 19746 (2020).

[39]J. Vandermause, S. B. Torrisi, S. Batzner, Y. Xie, L. Sun, A. M. Kolpak, and B. Kozinsky, Npj Comput. Mater. **6**, 1 (2020).

[40]R. Jinnouchi, J. Lahnsteiner, F. Karsai, G. Kresse, and M. Bokdam, Phys. Rev. Lett. **122**, 225701 (2019).

[41]S. Tovey, A. N. Krishnamoorthy, G. Sivaraman, J. Guo, C. Benmore, A. Heuer, and C. Holm, J. Phys. Chem. C **124**, 25760 (2020).

[42]S. Yue, M. C. Muniz, M. F. Calegari Andrade, L. Zhang, R. Car, and A. Z. Panagiotopoulos, J. Chem. Phys. **154**, 034111 (2021).

[43]A. Fabrizio, A. Grisafi, B. Meyer, M. Ceriotti, and C. Corminboeuf, Chem. Sci. **10**, 9424

(2019).

[44]S. Faraji, S. A. Ghasemi, S. Rostami, R. Rasoulkhani, B. Schaefer, S. Goedecker, and M. Amsler, Phys. Rev. B **95**, 1 (2017).

[45]S. A. Ghasemi, A. Hofstetter, S. Saha, and S. Goedecker, Phys. Rev. B **92**, 045131 (2015).

[46]A. Grisafi, A. Fabrizio, B. Meyer, D. M. Wilkins, C. Corminboeuf, and M. Ceriotti, ACS Cent. Sci. **5**, 57 (2019).

[47]T. W. Ko, J. A. Finkler, S. Goedecker, and J. Behler, Nat. Comm. **12**, 1 (2021).

[48]O. T. Unke and M. Meuwly, J. Chem. Theory Comput. **15**, 3678 (2019).

[49]X. Xie, K. A. Persson, and D. W. Small, J. Chem. Theory Comput. **16**, 4256 (2020).

[50]K. Yao, J. E. Herr, D. W. Toth, R. McKintyre, and J. Parkhill, Chem. Sci. **9**, 2261 (2018).

[51]M. R. Busche, D. A. Weber, Y. Schneider, C. Dietrich, S. Wenzel, T. Leichtweiss, D. Schröder, W. Zhang, H. Weigand, D. Walter, S. J. Sedlmaier, D. Houtarde, L. F. Nazar, and J. Janek, Chem. Mater. **28**, 6152 (2016).

[52]C. Dietrich, D. A. Weber, S. J. Sedlmaier, S. Indris, S. P. Culver, D. Walter, J. Janek, and W. G. Zeier, J. Mater. Chem. A **5**, 18111 (2017).

[53]J. P. Perdew, K. Burke, and M. Ernzerhof, Phys. Rev. Lett. **77**, 3865 (1996).

[54]V. Blum, R. Gehrke, F. Hanke, P. Havu, V. Havu, X. Ren, K. Reuter, and M. Scheffler, Comput. Phys. Commun. **180**, 2175 (2009).

[55]S. Plimpton, J. Comput. Phys. **117**, 1 (1995).

[56]A. P. Bartók, R. Kondor, and G. Csányi, Phys. Rev. B **87**, 184115 (2013).

[57]A. K. Rappe and W. A. Goddard III, J. Phys. Chem. **95**, 3358 (1991).

[58]F. H. Streitz and J. W. Mintmire, Phys. Rev. B **50**, 16 (1994).

[59]A. H. Larsen, J. J. Mortensen, J. Blomqvist, I. E. Castelli, R. Christensen, M. Dułak, J. Friis, M. N. Groves, B. Hammer, C. Hargus, E. Hermes, P. C. Jennings, P. B. Jensen, J. Kermode, J. R. Kitchin, E. L. Kolsbjerg, J. Kubal, K. Kaasbjerg, S. Lysgaard, J. B. Maronsson, T. Maxson, T. Olsen, L. Pastewka, A. Peterson, C. Rostgaard, J. Schiøtz, O. Schütt, M. Strange, K. S. Thygesen, T. Vegge, L. Vilhelmsen, M. Walter, Z. Zeng, and K. W. Jacobsen, J. Phys. Condens. Matter **29**, 273002 (2017).

[60]P. Virtanen, R. Gommers, T. E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, E. Burovski, P. Peterson, W. Weckesser, J. Bright, S. J. van der Walt, M. Brett, J. Wilson, K. J. Millman, N. Mayorov, A. R. J. Nelson, E. Jones, R. Kern, E. Larson, C. J. Carey, İ. Polat, Y. Feng, E. W. Moore, J. VanderPlas, D. Laxalde, J. Perktold, R. Cimrman,

I. Henriksen, E. A. Quintero, C. R. Harris, A. M. Archibald, A. H. Ribeiro, F. Pedregosa, P. van Mulbregt, and SciPy 1.0 Contributors, Nat. Methods **17**, 261 (2020).

[61] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, J. Mach. Learn. Res. **12**, 2825 (2011).

[62] V. L. Deringer and G. Csányi, Phys. Rev. B **95**, 094203 (2017).

[63] O. T. Unke and M. Meuwly, J. Chem. Theory Comput. **15**, 3678 (2019).

[64] G. Henkelman, B. P. Uberuaga, and H. Jónsson, J. Chem. Phys. **113**, 9901 (2000).

[65] D. Chang, K. Oh, S. J. Kim, and K. Kang, Chem. Mater. **30**, 8764 (2018).

[66] H. Yamane, M. Shibata, Y. Shimane, T. Junke, Y. Seino, S. Adams, K. Minami, A. Hayashi, and M. Tatsumisago, Solid State Ion. **178**, 1163 (2007).

[67] D. Frenkel and B. Smit, *Understanding molecular simulation: from algorithms to applications*, Vol. 1 (Elsevier, 2001).

[68] S. Döpking, C. P. Plaisance, D. Strobusch, K. Reuter, C. Scheurer, and S. Matera, J. Chem. Phys. **148**, 034102 (2018).

[69] A. C. Luntz, J. Voss, and K. Reuter, J. Phys. Chem. Lett. **6**, 4599 (2015).

[70] L. Hong, K. Yang, and M. Tang, Npj Comput. Mater. **5**, 1 (2019).

[71] T. K. Schwietert, V. A. Arszelewska, C. Wang, C. Yu, A. Vasileiadis, N. J. de Klerk, J. Hageman, T. Hupfer, I. Kerkamm, Y. Xu, Y. Xu, E. van der Maas, E. M. Kelder, S. Ganapathy, and M. Wagemaker, Nat. Mater. **19**, 428 (2020).

[72] Y. Zhu, X. He, and Y. Mo, ACS Appl. Mater. Interfaces **7**, 23685 (2015).

[73] Y. Xiao, Y. Wang, S.-H. Bo, J. C. Kim, L. J. Miara, and G. Ceder, Nat. Rev. Mater. **5**, 105 (2020).

[74] L. R. Mangani and C. Villevieille, J. Mater. Chem. A **8**, 10150 (2020).

[75] D. Opalka, C. Scheurer, and K. Reuter, ACS Catal. **9**, 4944 (2019).