

Cite this: *Digital Discovery*, 2025, 4, 1544

# Application of pretrained universal machine-learning interatomic potential for physicochemical simulation of liquid electrolytes in Li-ion batteries†

Suyeon Ju,  ‡<sup>a</sup> Jinmu You,  ‡<sup>a</sup> Gijin Kim,  <sup>a</sup> Yutack Park,  <sup>a</sup> Hyungmin An  <sup>a</sup> and Seungwu Han  \*<sup>ab</sup>

Achieving higher operational voltages, faster charging, and broader temperature ranges for Li-ion batteries necessitates advancements in electrolyte engineering. However, the complexity of optimizing combinations of solvents, salts, and additives has limited the effectiveness of both experimental and computational screening methods for liquid electrolytes. Recently, pretrained universal machine-learning interatomic potentials (MLIPs) have emerged as promising tools for computational exploration of complex chemical spaces with high accuracy and efficiency. In this study, we evaluated the performance of the state-of-the-art equivariant pretrained MLIP, SevenNet-0, in predicting key properties of liquid electrolytes, including solvation behavior, density, and ion transport. To assess its suitability for extensive material screening, we considered a dataset comprising 20 solvents. Although SevenNet-0 was predominantly trained on inorganic compounds, its predictions for the properties of liquid electrolytes showed good agreement with experimental and *ab initio* data. However, systematic errors were identified, particularly in the predicted density of liquid electrolytes. To address this limitation, we fine-tuned SevenNet-0, achieving improved accuracy at a significantly reduced computational cost compared to developing bespoke models. Analysis of the training set suggested that the model achieved its accuracy by generalizing across the chemical space rather than memorizing trained configurations. This work highlights the potential of SevenNet-0 as a powerful tool for future engineering of liquid electrolyte systems.

Received 20th January 2025  
Accepted 2nd May 2025

DOI: 10.1039/d5dd00025d

rsc.li/digitaldiscovery

## 1 Introduction

Li-ion batteries (LIBs) have revolutionized modern technology by powering a wide range of devices, from mobile phones to electric vehicles.<sup>1,2</sup> Among the various components comprising LIBs, the liquid electrolytes play a crucial role in facilitating ion transport between the anode and cathode, enabling the charging and discharging cycles.<sup>3-7</sup> Commercial formulations often incorporate lithium hexafluorophosphate (LiPF<sub>6</sub>) as the Li salt,<sup>8</sup> while ethylene carbonate (EC)-based solvents are established as the industry standard due to their ability to form a robust solid electrolyte interphase (SEI) on graphitic anodes.<sup>9</sup> When mixed with linear carbonates such as dimethyl carbonate (DMC), ethyl methyl carbonate (EMC), and diethyl carbonate (DEC), these electrolytes offer the complementary advantages of high salt dissociation from cyclic carbonates with high

permittivity, alongside the enhanced ion mobility and reduced viscosity provided by linear carbonates.<sup>10,11</sup> The introduction of vinylene carbonate (VC) as an additive has further enhanced SEI stability, with its highly reactive carbon-carbon double bond that promotes the formation of polymeric species.<sup>10</sup>

While the current recipes of liquid electrolytes satisfy various requirements, there is still a pressing demand for further optimizing the liquid electrolyte for LIBs to achieve enhanced energy density, safety, cycle life, and performance across various temperatures.<sup>12-14</sup> For example, to increase the operation voltage, it is necessary to lower the HOMO (Highest Occupied Molecular Orbital) level of the electrolyte to prevent degradation, or introduce additives to form a stable cathode-electrolyte interphase (CEI).<sup>15,16</sup> On the other hand, incorporating bulky anions can increase the charging speed by elevating the Li-ion transference number,<sup>17,18</sup> although this results in lower ionic conductivity by retarding Li-ion movements.<sup>19</sup> Lastly, commercial EC-based electrolytes are vulnerable in low-temperature environments, where the viscosity increases significantly and solidification occurs, lowering the ionic conductivity.<sup>20</sup> In addition, due to the sluggish desolvation of Li ions, charge transfer between the anode and electrolyte is hindered.<sup>21</sup>

<sup>a</sup>Department of Materials Science and Engineering and Research Institute of Advanced Materials, Seoul National University, Seoul 08826, Korea. E-mail: hansw@snu.ac.kr

<sup>b</sup>Korea Institute for Advanced Study, Seoul 02455, Korea

† Electronic supplementary information (ESI) available. See DOI: <https://doi.org/10.1039/d5dd00025d>

‡ These authors contributed equally to this work.



Switching to ether-based electrolytes<sup>22</sup> or utilizing (localized) high-concentration electrolytes<sup>23</sup> may resolve these problems.

In the above-mentioned cases, selecting optimal formulations for liquid electrolytes often requires a careful balance between correlated material properties, such as viscosity and solvation, which calls for testing of various materials and their combinations. Considering the vast space of organic molecules and the challenges in measuring physicochemical properties experimentally, atomistic simulations have become highly useful for material screening and understanding variations in properties at the atomic level.<sup>24–26</sup> In particular, molecular dynamics (MD) simulations using classical potentials or density functional theory (DFT) calculations, have been instrumental in investigating solvation structures and physicochemical properties (*i.e.*, diffusivity and viscosity) of electrolytes,<sup>27–29</sup> as well as interfacial reactions between electrodes and electrolytes.<sup>30–32</sup>

However, theoretical studies based on DFT and classical potentials face challenges in computational cost and transferability, respectively. For example, the high computational cost of DFT limits simulation size and time to a few hundred atoms and tens of picoseconds.<sup>28,29</sup> This limitation raises concerns that the simulation may not reach equilibration within the DFT time scale.<sup>33</sup> For instance, the residence time of Li and solvent molecules can extend up to a few tens of nanoseconds,<sup>33,34</sup> a time scale DFT cannot practically achieve. On the other hand, while classical potentials allow simulations of tens of thousands of atoms over hundreds of nanoseconds, they sacrifice transferability and general accuracy by fitting model parameters to DFT results or experimental data specific to particular systems. For example, the charges in OPLS-AA (Optimized Potentials for Liquid Simulations-All Atom) were scaled by 80% to fit to the experimental Li diffusivities in EC electrolytes.<sup>35,36</sup> However, this approach deteriorated the diffusivities of both the Li ion and the PF<sub>6</sub><sup>-</sup> anion in PC solvent.<sup>35</sup> Similar trade-offs on different properties have been reported in other classical potentials, such as in TraPPE (Transferable Potentials for Phase Equilibria) force fields,<sup>37</sup> which accurately predicted densities but showed more than 20% error in relative permittivities. Beyond these Class I force fields, more advanced force fields, such as Class II,<sup>38</sup> incorporating bond and angle anharmonicity, and APPLE&P<sup>39</sup> (Atomistic Polarizable Potential for Liquids, Electrolytes, and Polymers, Class-III), which is many-body polarizable, achieve higher accuracy through additional parameters, but they demand careful parameter tuning and still suffer from limited transferability.

Over the past decade, data-driven machine-learning interatomic potentials (MLIPs) have gained significant attention in materials simulation by extending both the length and time scales to those of classical potentials while maintaining accuracy close to that of DFT.<sup>40–44</sup> Therefore, employing MLIPs in the simulation of liquid electrolytes is poised to overcome the difficulties in DFT or classical force fields mentioned above. However, there are significant challenges in developing MLIPs for liquid electrolytes, particularly with traditional application-specific, bespoke-style MLIPs. First, generating training sets with DFT-based MD simulations incurs high computational costs. This is because adequate sampling of all possible configurations,

including different molecular conformers and achieving ergodicity, requires long-term simulations. Second, in order to computationally identify optimal formulations from material screening, it is necessary to develop MLIPs that can be applied to a wide range of organic molecules. This in turn requires the creation of a comprehensive training set incorporating various combinations of solvents and salt pairs and a careful sampling of both intramolecular and intermolecular interactions among different chemical moieties. Thus, previous studies have relied on many cycles of iterative learning to generate such training sets,<sup>45,46</sup> incorporating various strategies to improve the precision of intermolecular interactions. Consequently, most studies using MLIPs have been limited to investigating specific solvent-salt systems, such as glyme-based electrolytes<sup>47</sup> and carbonate electrolytes,<sup>45,48</sup> or solvent-only systems consisting of mixtures of EC and EMC.<sup>46</sup> In another example using the graph neural network interatomic potential (GNN-IP), ref. 49 analyzed the Li transport mechanisms in deep eutectic electrolytes and lithium bis(trifluoromethanesulfonyl)imide (LiTFSI). It is notable that ref. 50 has extended the chemical space by incorporating ester materials and fluorine doping at various sites in carbonates. However, for untrained fluorine-doped systems, the density error reached a maximum of 21%.

Recently, pretrained general-purpose GNN-IPs such as M3GNet,<sup>51</sup> CHGNet,<sup>52</sup> PFP,<sup>53</sup> GNoME,<sup>54</sup> MACE-MP-0,<sup>55</sup> SevenNet-0,<sup>56</sup> MatterSim,<sup>57</sup> eqV2 M,<sup>58</sup> and ORB<sup>59</sup> have emerged, providing generalizability across diverse chemical spaces. It has been also shown that fine-tuning the pretrained model can achieve the precision of bespoke models at a small cost.<sup>60–62</sup> The generalizability of these models largely stems from the architecture of GNN-IPs, such as NequIP<sup>63</sup> and MACE,<sup>64</sup> which automatically extracts important features from deep learning. In addition, atomic species are embedded with learnable parameters, allowing the model to learn chemical similarities between elements.<sup>65</sup> This enables the pretrained model to capture general trends in chemical bonding.

Most of the current pretrained models were trained using inorganic materials databases such as the Materials Project,<sup>66</sup> Alexandria,<sup>67</sup> and OMat24.<sup>58</sup> Nevertheless, MACE-MP-0 demonstrated reasonable accuracy and stability in simulating liquid electrolytes when applied to the EC/EMC LiPF<sub>6</sub> electrolyte and a complete battery system.<sup>55</sup> In another example, ref. 68 compared densities and diffusivities of 3 : 7 EC : EMC solvents between the bespoke MACE and MACE-MP-0 models and found reasonable agreements. However, for pretrained models extensively trained on inorganic compounds, liquid electrolytes fall into strongly out-of-distribution domains. First, as will be detailed in this work, relevant structural motifs of organic molecules were not fully sampled in the dataset. One may question whether training on an organic dataset such as SPICE<sup>69</sup> or using molecule-based pretrained models like MACE-OFF<sup>70</sup> would offer a better alternative. However, SPICE contains few Li-containing structures, and MACE-OFF lacks training on ionic species such as Li and PF<sub>6</sub><sup>-</sup>, making these models unsuitable for comprehensive simulations of liquid electrolytes. Furthermore, most pretrained models do not explicitly account for long-range electrostatic interactions, and research on



implicit treatment is rarely addressed.<sup>71</sup> As a result, it is unclear how well they can describe solvation structures and dielectric properties, which are primarily governed by Coulomb interactions. Therefore, a systematic analysis on the general accuracy of current pretrained models in liquid electrolytes is in demand.

In this work, motivated by the above discussions, we systematically investigate the accuracy of SevenNet-0 (ref. 72) (simply SevenNet or 7net henceforth) in applications involving liquid electrolytes. SevenNet has achieved high performance in the Matbench Discovery benchmark, which assesses the performance of pretrained universal force fields on inorganic crystal discovery.<sup>73</sup> To evaluate whether the pretrained model is suitable for extensive material screening, we tested its performance across a diverse range of electrolytes. Although SevenNet does not explicitly incorporate electrostatic interactions, its message-passing traits are expected to implicitly encode these effects up to the receptive field (see ref. 71), and an explicit van der Waals functional is employed with SevenNet. Key properties such as densities, solvation shell structures, and diffusivities were compared with experimental data or *ab initio* MD (AIMD) results. While the overall agreement with reference data was good, SevenNet exhibited force softening and overestimated solvent density. To address this, we also fine-tuned SevenNet, which significantly improved the model's accuracy. This work highlights the potential of SevenNet for future engineering of liquid electrolyte systems.

## 2 Results and discussion

To test SevenNet with various types of molecules employed in liquid electrolytes, we considered a total of 20 solvents and 2 salts, as listed in Table S1† and schematically summarized in Fig. 1. The test solvents were selected to encompass a broad range of liquid electrolytes used in commercial batteries or in advanced battery research. They represent four major chemical groups: cyclic carbonates, linear carbonates, ethers, and esters. As base molecules, we used EC for cyclic carbonates, DMC for linear carbonates, dimethoxyethane (DME) for ethers, and ethyl acetate (EA) for esters, as shown on the left of Fig. 1. To expand the chemical space, we varied the molecular structures by altering the bond order, elongating carbon chains, incorporating fluorine atoms, and exploring *cis-trans* isomerism, as depicted in the center of Fig. 1. For salts, LiPF<sub>6</sub> and lithium bis(fluorosulfonyl)imide (LiFSI) were considered (see the right of Fig. 1). For cations, we focused primarily on Li ions (Li<sup>+</sup>), although sodium (Na<sup>+</sup>) and potassium (K<sup>+</sup>) ions were also included in the solvation shell analysis.

The full simulation of liquid electrolytes involves various types of bonding/nonbonding, intra-/intermolecular interactions. To systematically assess the accuracy of SevenNet, we apply the model to progressively more complex systems in the following subsections, starting from single molecules, moving to pure solvents, and finally to full electrolytes.

### 2.1 Single solvent molecules

For the simulation of organic systems, an accurate description of a single molecule is a basic requirement. We compared the

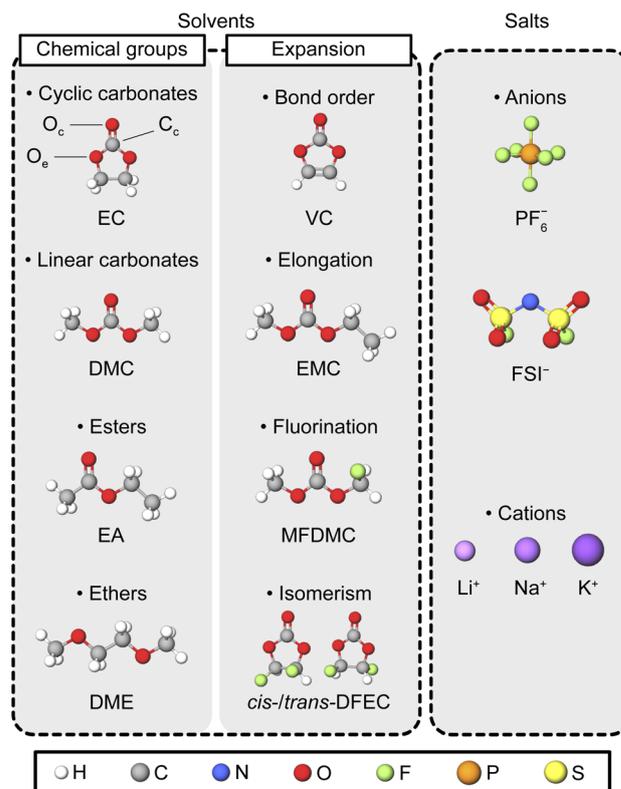


Fig. 1 Schematic description of the solvent molecules and ions investigated in this study. The solvents are categorized by chemical groups (carbonate, ester, and ether) and molecular structures (cyclic or linear) (left). Variations in the solvent set include changes in bond order, molecular elongation, partial fluorination, and isomerism (middle). Examples of the carbonyl oxygen (O<sub>c</sub>), carbonyl carbon (C<sub>c</sub>), and etheral oxygen (O<sub>e</sub>) are indicated on the EC molecule. The anions and cations used for salts in the electrolyte systems are shown in the right. For corresponding IUPAC names and formulas, refer to Table S1.†

single-molecule energies and structures of SevenNet and DFT for 20 solvent molecules. The molecule was initially placed in a cubic box with periodic boundary conditions, where the length of the box was set to 10 Å plus the maximum molecular length along each axis. Since the graph connectivity in SevenNet is truncated beyond the 5 Å cutoff, periodic images do not spuriously appear in the local atomic environment of the target molecule in the SevenNet calculations. In the DFT calculations, a dipole correction was applied in all directions to remove spurious dipole-dipole interactions between periodic images. Structural relaxation was performed until the magnitude of the atomic forces was reduced to within 0.02 eV Å<sup>-1</sup>. In Fig. 2a, we compare the per-atom energies obtained by SevenNet and DFT relaxation. The mean absolute error (MAE) is 23 meV per atom, where the largest deviation of 51 meV per atom found with the VC molecule. Although the energy error is larger than that of MLIPs trained solely on molecular systems (a few meV per atom),<sup>46,70</sup> it does not alter the relative energy ordering of the single molecules, offering basic evidence of accuracy of SevenNet. We also compared the bond lengths and angles of relaxed molecular structures to verify that SevenNet produced



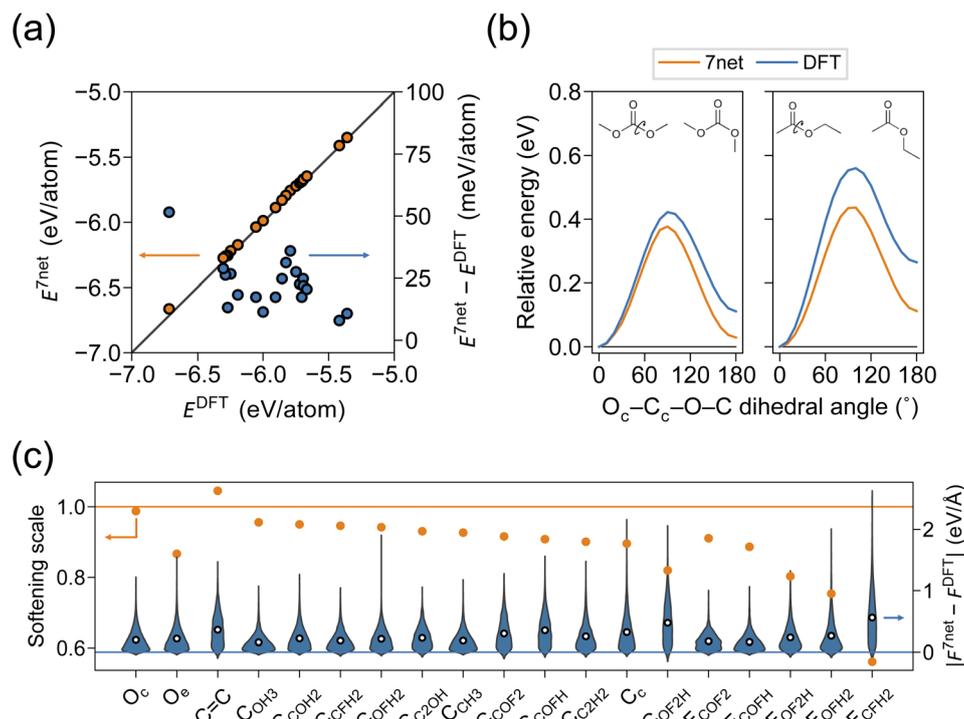


Fig. 2 (a) Comparison of per-atom energies (left axis) and corresponding errors (right axis) between DFT and SevenNet predictions for 20 single solvent molecules. (b) Relative energy profiles for DMC (left) and EA (right) molecules as a function of  $O_c-C_c-O-C$  dihedral angle. See the inset for the schematic images. (c) Force softening scales for each atom type (left axis) and the distribution of absolute errors in force components (right axis) obtained from single-molecule SevenNet MD trajectories. MAEs of the force components are indicated by white-filled circles.

molecular geometries correctly. The MAEs for bond lengths and angles are  $0.005 \text{ \AA}$  and  $0.7^\circ$ , respectively, indicating highly accurate structural predictions that underscore the robustness of SevenNet for short-range interactions.

Linear solvents exhibit *cis-trans* conformers, and structural differences affect dipole moments,<sup>74,75</sup> influencing the participation of solvent in Li-ion solvation.<sup>76</sup> We select two molecules, DMC and EA, since linear carbonates and esters generally exhibit larger energy barriers between conformers compared to other types of molecules. In Fig. 2b, by performing dihedral-angle-constrained ionic relaxations, we scan a total of  $180^\circ$  in  $10^\circ$  intervals for the  $O_c-C_c-O-C$  dihedral angle in DMC and EA (see the inset). It is seen that SevenNet underestimates the torsion barrier by 0.04 eV and 0.12 eV for DMC and EA, respectively. This underestimation of barriers may be related to the softening of potential energy surface (PES) in pretrained models that were primarily trained on low-energy structures.<sup>62</sup>

To investigate the accuracy of atomic forces, we conducted a 0.5 ns MD simulation of a single molecule in an NVT ensemble using SevenNet. A Nosé–Hoover thermostat<sup>77</sup> and a timestep of 0.5 fs were employed. The temperature was set to 600 K to sample high-energy structures, and 500 snapshots were extracted at intervals of 1 ps. Subsequently, single-point DFT calculations were performed on these snapshots. The parity plots for energy and force for each single-molecule MD are provided in Fig. S1 and S2.† Energy shifts observed in Fig. S1† are in agreement with Fig. 2a. The softening scales and the corresponding absolute error distributions of the force components are shown in Fig. 2c. The

softening scale is defined as the slope of the linear function fitted to the force parity plot. The ideal value is one, and those below one indicate that the forces predicted by SevenNet were systematically smaller than those calculated by DFT. To examine whether force error and softening depend on specific atom types, atoms were classified into 19 categories (Fig. S3†). To be specific, the oxygen atoms were divided into carbonyl ( $O_c$ ) and ethereal ( $O_e$ ) oxygens (see Fig. 1). The carbon atoms were categorized based on their bonding environments; for example,  $C_{COFH}$  represents a carbon atom bonded to one carbon, one oxygen, one fluorine, and one hydrogen atom. Double-bonded carbons in VC were classified separately. Fluorine atoms were labeled according to the type of carbon atom to which they are bonded; for instance,  $F_{COFH}$  refers to fluorine atoms bonded to a  $C_{COFH}$  carbon atom. In Fig. 2c, most atom types exhibit some degree of force softening; however, the extent of softening varies among atom types. Notably, pronounced softening is observed for fluorine atoms in  $F_{OFH_2}$  and  $F_{CFH_2}$  local structures. Additionally, molecules containing partially fluorinated carbon moieties, such as difluoromethyl fluoromethyl carbonate (TFDMC), show relatively larger force errors (Fig. S2†), which is consistent with the observations in Fig. 2c. As will be discussed below, this significant level of softening may stem from the limited representation of these chemical moieties in the training set.

## 2.2 Pure solvents

Next, we simulated pure solvents composed of a single type of organic molecule to obtain their theoretical densities. Liquid



density plays a critical role in determining the physicochemical properties of electrolytes. For example, a density decrease of just  $0.1 \text{ g cm}^{-3}$  can result in a twofold increase in diffusivity for acetonitrile at 298 K.<sup>78</sup> The initial configurations for the liquid simulations were generated using MolView<sup>79</sup> and PACKMOL.<sup>80</sup> The number of molecules was chosen so that the total number of atoms was closest to 1000. To determine the length of the cubic simulation box without relying on experimental data, we first obtained  $V_0$  by adding the van der Waals volumes<sup>81</sup> of all atoms in the simulation box. The initial box length was then set to  $1.1 \times V_0^{1/3}$ . This volume estimate ensures that neighboring molecules are close enough for inter-molecular connectivity within the graph neural network, while still providing sufficient space to prevent molecular overlap (the actual numbers of molecules and the sizes of the simulation boxes for each simulation are summarized in Table S2.†). For propylene carbonate (PC) and fluoroethylene carbonate (FEC), two chiral conformers were considered in equal proportions. For the *cis-trans* conformers of linear carbonates and the *syn/anti* conformers of esters, the Boltzmann distribution, with potential energies calculated using SevenNet, was used to determine the initial ratio (see Tables S3 and S4.†). This procedure was crucial during the equilibration step, as the conformer ratio could not equilibrate within a feasible time scale due to the large energy barrier ( $\sim 0.4 \text{ eV}$ ) between conformers, which is significantly higher than the thermal energy at room temperature.

The initial structures were relaxed under loose conditions ( $\max_i |F_i| < 2.0 \text{ eV \AA}^{-1}$ ), followed by MD simulations for 1 ns in the NPT ensemble with a timestep of 2 fs at the target temperature and a pressure of 1 atm as pre-equilibrations. The Nosé-Hoover thermostat and barostat<sup>82</sup> were applied as implemented in the LAMMPS package.<sup>83</sup> To ensure a fair comparison with experimental data, the simulation temperatures were matched to the experimental conditions. Specifically, all simulations were conducted at 298 K, except for EC (313 K), FEC (313 K), DME (293 K), and 1,2-diethoxyethane (DEE) (293 K). To maintain stable simulations and prevent large positional fluctuations with a timestep of 2 fs, the atomic mass of tritium ( $^3\text{H}$ , 3 a.u.) was assigned to hydrogen atoms. We emphasize that this modification was applied only during the pre-equilibration stage using 2 fs timestep throughout the study, except for the cases described in Section 2.3.1. Both the equilibration and the production run were performed with the original hydrogen mass and a 1 fs timestep. To determine the equilibrium density of the system, an additional 0.4 ns simulation was performed in the NPT ensemble with a reduced timestep of 1 fs. Following 0.2 ns equilibration, the density was calculated by averaging the instantaneous density values (recorded every 10 fs) during the last 0.2 ns production run. Throughout all equilibration and production runs, no spurious reactions were observed, confirming the stability of the simulations.

Fig. 3a presents computed liquid densities. Since reference densities obtained with DFT are scarce, we compared with experimental data sourced from ref. 10, 50 and 84–90 (see Table S5.† for the actual values and errors). To identify systematic trends, we classified the solvents into four chemical groups;

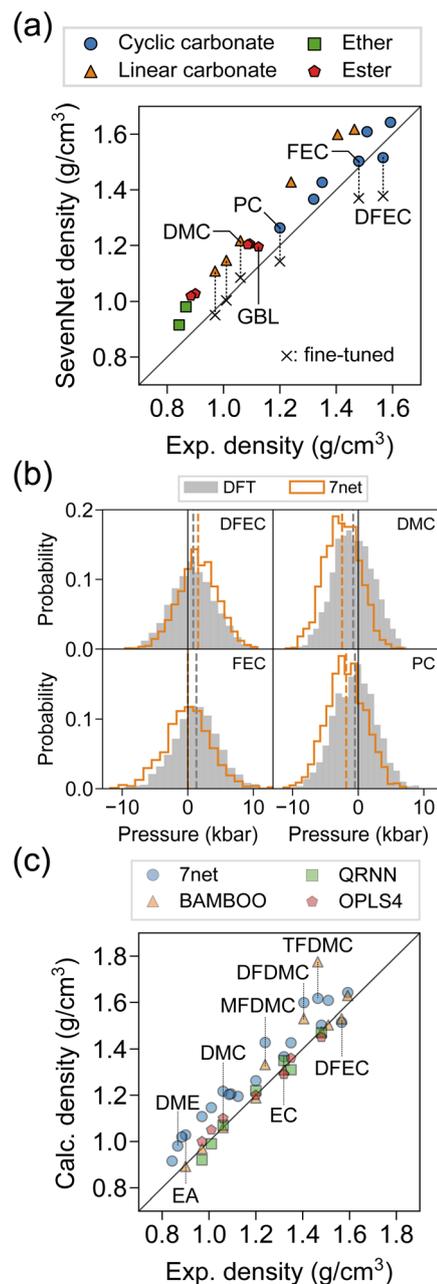


Fig. 3 (a) Parity plot comparing pure solvent densities calculated using SevenNet to experimental values, categorized by chemical groups. MD simulations were conducted at 298 K, except for EC (313 K), FEC (313 K), DME (293 K), and DEE (293 K). The blue circles, orange triangles, green squares, and red pentagons represent cyclic carbonates, linear carbonates, ether, and ester, respectively. The black 'x' symbols indicate the values obtained with fine-tuned SevenNet. (b) Pressure distributions for DFEFC, FEC, DMC, and PC from individual SevenNet-MD and AIMD simulations conducted in the NVT ensemble over 15 ps at experimental densities. Temperatures are set to match the experimental values in Table S5.† Vertical dashed lines indicate the mean values of each distribution. (c) Parity plot comparing pure solvent densities predicted by SevenNet (blue circles), BAMBOO (orange triangles), QRNN (green squares), and OPLS4 (red pentagons) versus experimental values.



cyclic carbonates, linear carbonates, ethers, and esters. Fig. 3a shows that the densities computed using SevenNet generally overestimates compared to experimental values. However, the degree of overestimation is not random but consistent within chemical groups. To be specific, cyclic solvents, including cyclic carbonates and  $\gamma$ -butyrolactone (GBL), exhibit smaller deviations, ranging  $-3$ – $7\%$ , while linear solvents show higher overestimation by  $9$ – $15\%$ .

To clarify whether the discrepancy in predicted densities is primarily due to limitations of the SevenNet model or inherent inaccuracies in the DFT functional (PBE-D3), we examined the pressure distributions (average of diagonal components of the virial stress) by SevenNet and DFT for FEC, difluoroethylene carbonate (DFEC), DMC, and PC liquids, when the volumes are adjusted to match with the experimental densities. The temperatures were set to 298 K (DFEC, DMC, and PC) or 313 K (FEC). For the feasibility with DFT calculations, we chose a smaller simulation cell containing 30 solvent molecules and the initial configurations (with equal chirality for PC and FEC) were equilibrated under SevenNet for 1 ns. Equilibrations are performed using a 2 fs timestep assigning tritium mass to hydrogen atoms. From these equilibrated structures, we ran 20 ps production simulations in NVT ensembles independently with either SevenNet or DFT with a 1 fs timestep and the original hydrogen mass, recording the pressure every 10 fs during the final 15 ps to gather 1500 data points.

Fig. 3b shows the pressure distributions of both DFT and SevenNet for each solvent, with vertical dashed lines marking the mean values of each distribution. The pressure values include dynamic contributions, meaning that the thermal effects at the reference temperature are fully accounted. For DFT, the average pressure remains close to zero in all solvents, suggesting that the experimental volume is close to the equilibrium density predicted by PBE-D3. The absolute pressures are 0.77 and 0.50 kbar for DMC and PC, respectively. Given the bulk moduli of 1.6 GPa for DMC<sup>91</sup> and 2.4 GPa for PC<sup>92</sup> at 298 K, the corresponding density errors for DFT are on the order of 5% and 2%, respectively. However, a more rigorous convergence study exploring larger simulation cells and longer time scales, would be needed for quantitative accuracy. A previous work demonstrated that bespoke MLIPs can predict densities of EMC-rich solvent mixtures to within about 5% of experiment at the PBE-D3 level, using a relatively small cutoff radius of 10 Å for the D3 term.<sup>46</sup>

In contrast, SevenNet predicts more negative pressures for DMC and PC in Fig. 3b, indicating a significant compressive stress that drives the system toward higher density. This shift is correlated with the degree of density overestimation observed for PC and DMC when using SevenNet (see Fig. 3a). Further single-point DFT calculations on SevenNet-generated snapshots also confirmed that these pressure discrepancies arise from inaccuracies in the SevenNet model rather than from fundamental errors in the DFT reference (Fig. S4†). The constant shift observed in the normal stress parity plots indicates a systematic error in pressure prediction, as the pressure is computed from the average of the normal stress components. The trends in both mean error in normal stress and average pressure aligned

closely. Consequently, while slight errors in the DFT reference cannot be completely ruled out, we conclude that the dominant source of density overestimation in SevenNet is an imperfect learning of stress. These systematic shifts in pressure and normal stress are likely attributable to insufficient sampling of intermolecular interactions in the predominantly inorganic training set. As discussed in the following section, we found that fine-tuned SevenNet reduced this discrepancy, suggesting that improved training strategies can bring the predicted pressures and densities closer to the DFT and also experimental values.

In Fig. S5 and S6,† we calculated potential energy curves for EC and DMC dimers across eight types of interactions (H–H, O<sub>c</sub>–H, O<sub>c</sub>–O<sub>e</sub>, O<sub>c</sub>–O<sub>e</sub>, O<sub>e</sub>–H, O<sub>e</sub>–O<sub>e</sub>, orthogonal, and planar orientations). It is found that SevenNet has a deeper potential well near equilibrium, with some equilibrium distances being shorter than those from DFT. Thus, the pressure deviations in the above can be attributed to the stronger intermolecular bonding in SevenNet compared to DFT.

We compared the performance of SevenNet in predicting liquid densities with other computational methods, namely QRNN,<sup>45</sup> OPLS4,<sup>45</sup> and BAMBOO<sup>50</sup> (Fig. 3c). QRNN and OPLS4 are quantitatively more accurate, showing errors of  $-5$ – $2\%$  and  $-2$ – $4\%$ , respectively. We note that a portion of the QRNN training set was generated at high pressures and another portion with  $\pm 20\%$  scaling in intermolecular distances, implying that QRNN may have learned the equilibrium volume from these training sets (this training method was also used in ref. 46 and 68.). The BAMBOO model generally provided density estimates closer to experimental data than SevenNet, with the exception of the fluorinated linear carbonate group. A density alignment method was employed for BAMBOO, wherein the model was trained to match experimental densities for systems such as EC, PC, FEC, DEC, DMC, and EA, which are plotted together in Fig. 3c. While this alignment method enabled accurate predictions for in-domain systems, it limited generalizability to less similar systems, as evidenced by the TFDMC case, where the model overestimates density by 20%. In contrast, SevenNet exhibited moderate and regularized accuracy across all solvent systems.

### 2.3 Full electrolytes

In this subsection, we applied SevenNet to simulate full electrolytes composed of solvents and ion salts. First, we investigated the solvation shell structures for dilute electrolytes. Next, at a conventional concentration of Li salt, we analyzed the effect of solvent type on Li solvation, with a focus on salt dissociation. Finally, diffusion coefficients were obtained and compared with experimental values. To avoid potential artefacts, fluorinated solvents were excluded from further analysis due to the force softening observed in Section 2.1. For all electrolytes, the densities were adjusted to match experimental values, thereby eliminating errors associated with density mismatch.

**2.3.1 Solvation shell structures in dilute solutions.** The solvation structure of ions plays a critical role in electrolyte systems, influencing both ion dissociation and the stability of



electrode interfaces, which directly impacts ionic conductivity and battery cycle life. In ref. 28, solvation shell structures around Li ions were investigated in detail by AIMD. In this subsection, we benchmark SevenNet against these reference results by adopting the same simulation protocol. In detail, the initial structure consisted of 63 EC (42 EMC) molecules and one LiPF<sub>6</sub> salt, corresponding to dilute conditions (0.2 M). Experimental densities of pure EC (1.32 g cm<sup>-3</sup> at 313 K) and EMC (1.01 g cm<sup>-3</sup> at 298 K) were used,<sup>10</sup> yielding cubic-cell lengths of 19.28 Å for EC and 19.52 Å for EMC. To obtain various solvation types, ten independent simulations were performed: five starting with a dissociated ion pair and the other five with an associated pair. Each structure was equilibrated for 7.5 ps at 330 K in the NVT ensemble, followed by a 30 ps production run with a timestep of 0.5 fs to ensure consistency with the simulation scheme in ref. 28 and enable a fair comparison. Here, the dispersion interaction was excluded like ref. 28. Snapshots were saved every 5 fs, generating 6001 snapshots for each production run. A total of 60 010 snapshots were collected for each solvent.

In liquid electrolytes, Li ions undergo various types of solvation environments since there can be multiple coordinating oxygens in a solvent molecule, and anion can also be introduced in the solvation shell. Each solvation shell type has distinct structural features, like radial and angular distributions related to Li–O, as studied by AIMD.<sup>28</sup> Following the reference, we classified the collected snapshots into five types of solvation shell: Li(EC<sub>c</sub>)<sub>4</sub>, Li(EC<sub>c</sub>)<sub>3</sub>(PF<sub>6</sub>), Li(EMC<sub>c</sub>)<sub>3</sub>(PF<sub>6</sub>), Li(EMC<sub>c</sub>)<sub>3</sub>(EMC<sub>c</sub>), and Li(EMC<sub>c</sub>)<sub>2</sub>(PF<sub>6</sub>). Such classification was based on the composition of the first solvation shell of a Li ion and the coordinating oxygen type. For example, if a Li ion is coordinated with four EMC molecules—three *via* O<sub>c</sub> and one *via* O<sub>e</sub>—then it is designated as Li(EMC<sub>c</sub>)<sub>3</sub>(EMC<sub>c</sub>). When both O<sub>c</sub> and O<sub>e</sub> of a single solvent molecule simultaneously coordinated a Li ion, the coordinating oxygen type was designated as O<sub>c</sub>, where coordination by two O<sub>e</sub> atoms was classified as O<sub>e</sub>-type coordination.

The radial distribution function (RDF)  $g(r)$  and the coordination number (CN) of an atom type B around an atom type A is defined as follows:<sup>93</sup>

$$g(r) = \frac{1}{\rho_B} \frac{1}{4\pi r^2} \frac{dN(r)}{dr} \quad (1)$$

$$\text{CN} = N(r_{\text{cut}}) \quad (2)$$

where  $N(r)$  is the average number of particle B within a sphere centered on particle A with a radius of  $r$ , and  $\rho_B$  is the number density of B atoms. The cutoff threshold  $r_{\text{cut}}$  of Li–O was set to 2.6 Å and 4.2 Å for Li–P, which were obtained from the first minimum of RDF.

The bond length distributions of Li and oxygen of two representative solvation types, Li(EC<sub>c</sub>)<sub>4</sub> and Li(EMC<sub>c</sub>)<sub>3</sub>(PF<sub>6</sub>), are presented in Fig. 4a. (The corresponding results for other solvation types are shown in Fig. S7.†) Overall, the distributions of Li–O<sub>c</sub> and Li–O<sub>e</sub> distances and CNs (right axis) show good agreements with DFT results. In experiments, slightly longer

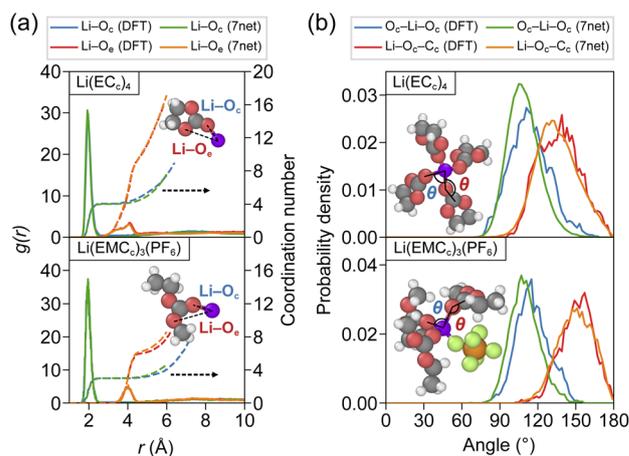


Fig. 4 Solvation structures of Li(EC<sub>c</sub>)<sub>4</sub> and Li(EMC<sub>c</sub>)<sub>3</sub>(PF<sub>6</sub>) solvation types obtained by SevenNet, compared with DFT results.<sup>28</sup> (a) RDFs (solid lines) of Li–O<sub>c</sub> and Li–O<sub>e</sub>, along with their CNs (dashed lines) for each solvation type. Examples of Li–O<sub>c</sub> and Li–O<sub>e</sub> distances are illustrated in the inset. (b) Angular distributions of O<sub>c</sub>–Li–O<sub>c</sub> and Li–O<sub>c</sub>–C<sub>c</sub> angles for each solvation type. Examples of angles are illustrated in the inset.

bond lengths of 2.04–2.08 Å were reported in concentrated solutions of LiPF<sub>6</sub> salt in PC and DMC.<sup>94,95</sup>

The distributions of O<sub>c</sub>–Li–O<sub>c</sub> and Li–O<sub>c</sub>–C<sub>c</sub> angles in Fig. 4b also agree well with DFT. The O<sub>c</sub>–Li–O<sub>c</sub> angle peaked at approximately 107°, minimizing steric hindrance within the near-tetrahedral coordination environment. This angle remains consistent regardless of the presence of PF<sub>6</sub><sup>-</sup> in the first solvation shell. Meanwhile, the Li–O<sub>c</sub>–C<sub>c</sub> angle deviates from 180°, likely due to the partial negative charges on the two O<sub>e</sub> atoms bonded to C<sub>c</sub>, which attract positively charged Li ions. This angle varies depending on the solvent: 132° in EC and ranging from 149° to 156° in EMC. This difference reflects the steric hindrance imposed by the bulky, linear structure of the EMC molecule. A similar trend has been observed experimentally, with PC exhibiting a Li–O<sub>c</sub>–C<sub>c</sub> angle of 138° and DMC showing an angle of 153°.<sup>94,95</sup> On the other hand, ReaxFF produced a sharper peak near 90°, suggesting a more rigid Li–O<sub>e</sub> interaction,<sup>28</sup> highlighting limitations in ReaxFF.

We extended our study to examine other alkali metal ions, Na<sup>+</sup> and K<sup>+</sup>. By replacing Li with Na or K in 63 EC + 1 LiPF<sub>6</sub> simulations, we conducted a 7 ps equilibration MD run followed by a 25 ps production run for each cation. The distributions of the cation–oxygen distance are shown in Fig. S8.† The Na–O and K–O bond lengths are 2.34 and 2.74 Å, respectively, aligning closely with AIMD results of 2.35 and 2.80 Å.<sup>29</sup> Compared to the Li–O distribution, the first peak is broadened and reduced in intensity, indicating weaker interactions between the cation and oxygen atoms as the ionic radius of the cation increases. This weaker solvation of Na<sup>+</sup> and K<sup>+</sup> relative to Li<sup>+</sup> arises from the delocalization of outer-shell electrons in larger cations, which hinders lone-pair sharing from oxygen atoms.<sup>29</sup>

These results demonstrate that SevenNet effectively recognizes the Li atom in the solvent as a cation and accurately



captures Coulomb interactions with negatively charged oxygen atoms, despite being trained without explicit charge information. Building on our previous analysis of both molecular structures and solvation shell configurations, we conclude that SevenNet demonstrates high accuracy in capturing the geometric properties of the system, which implies that short-range interactions are accurately represented.

**2.3.2 Effect of solvent type on Li solvation.** In this subsection, we investigated electrolytes with high Li salt concentrations used in commercial batteries. Depending on the solvents, the degree of ion dissociation between cations and anions varies, resulting in structures such as solvent-separated ion pairs (SSIPs), contact ion pairs (CIPs), and aggregates (AGGs), as illustrated in Fig. S9a.† Mixing cyclic solvents with high dielectric constants and linear solvents with lower dielectric constants influences ion dissociation and the composition of the first solvation shell.<sup>96,97</sup> Accurately simulating these phenomena is essential for identifying optimal electrolytes, which often balances multiple objectives, such as high ionic conductivity and robust SEI formation.<sup>3,10</sup>

As a concrete example, we selected the EC/DMC binary solvent system with 1 mol kg<sup>-1</sup> LiPF<sub>6</sub>, varying the EC molar fraction,  $x_{EC} = N_{EC}/(N_{EC} + N_{DMC})$ , where  $N_{\alpha}$  is the number of  $\alpha$  molecules. This system has been widely studied to explore the competition between two different solvent types within the Li solvation shell.<sup>39,96,98,99</sup> Recent experimental and computational studies have confirmed that EC is preferred for Li solvation over DMC, reducing cation–anion cross-correlation and enhancing ionic conductivity, particularly at low  $x_{EC}$  values.<sup>96,99</sup>

All simulations were performed in the NVT ensemble at experimental densities (like the case with pure solvents, SevenNet overestimated the density of solvent-salt systems, as shown in Fig. S10† for EC/LiFSI electrolyte). The conformer ratio of DMC for each composition was obtained from the experiment.<sup>96</sup> The temperature was set to 298 K to match the experimental conditions. The number of solvent and salt molecules was adjusted to correspond to a LiPF<sub>6</sub> concentration of 1 mol kg<sup>-1</sup>, ensuring the total number of atoms remained below 1000. Li ions were placed by dividing the simulation domain into distinct regions and randomly distributing the Li ions within these regions, as illustrated in Fig. S11.† This approach prevented the generation of initial structures where Li ions were clustered on one side. The specific numbers of solvent and salt molecules, along with the simulation box lengths, are summarized in Table S6.†

Equilibration was performed over 1.4 ns, starting with an initial 1 ns using a 2 fs timestep and a hydrogen atomic mass of 3 a.u., as described earlier, followed by an additional 0.4 ns of equilibration. A production run of 1 ns was conducted, with snapshots saved every 100 fs. The CNs for EC, DMC, and PF<sub>6</sub><sup>-</sup> anions were calculated and averaged across all snapshots for each simulation. To ensure statistical reliability, three independent runs were carried out, each starting from a distinct initial configuration.

Fig. 5a shows the computed compositions of the Li solvation shell as a function of  $x_{EC}$ . The sum of CNs of Li aligns well with experimental trends<sup>96</sup> and APPLE&P MD simulations

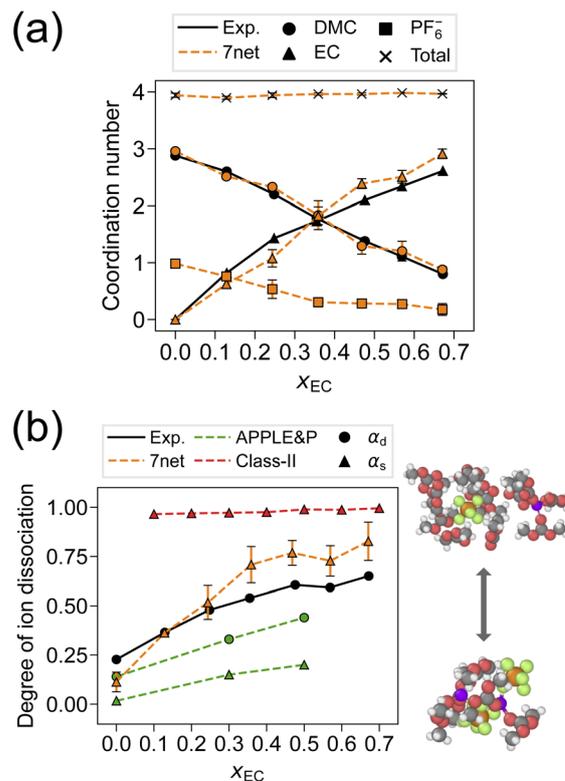


Fig. 5 Variations in solvation environment in the EC/DMC binary solvent system with 1 mol kg<sup>-1</sup> LiPF<sub>6</sub>. (a) Average CNs of Li contributed by EC and DMC solvents and (b) the degree of ion dissociation as a function of  $x_{EC}$ . Experimental values for 1 mol kg<sup>-1</sup> LiPF<sub>6</sub> (ref. 96) are shown alongside results from other force fields (APPLE&P<sup>39</sup> and Class-II force field<sup>100</sup>) for 1 M LiPF<sub>6</sub>. Circle markers represent the dynamic degree of dissociation  $\alpha_d$ , while triangle markers represent the static degree of dissociation  $\alpha_s$ . Error bars represent the standard deviation of 3 simulations. Representative solvation structures for fully dissociated and associated cases are displayed on the right side of the plot.

(Fig. S12†).<sup>39</sup> Notably, EC molecules were strongly favored in the solvation shell at low  $x_{EC}$ . For instance, at  $x_{EC} = 0.13$ , the local fraction of EC in the Li solvation shell is 0.25 (SevenNet) and 0.32 (experiment), significantly higher than the bulk EC ratio of 0.13. Over the entire range of  $0 < x_{EC} < 0.7$ , the total CN of Li remains approximately 4. A comparison of CNs obtained using different force fields is presented in Fig. S12.†

Next, we computed the degree of ion dissociation, whose definition subtly varies across the literature.<sup>39,101</sup> The static degree of ion dissociation,  $\alpha_s$ , is defined as the fraction of free ions,<sup>39</sup> which in turn is defined as having no counterions in its first solvation shell, corresponding to the SSIP solvation state. Some studies focused specifically on free cations,<sup>101</sup> which tends to yield slightly higher values of  $\alpha_s$  compared to free ions.<sup>39</sup> On the other hand, the dynamic degree of ion dissociation,  $\alpha_d$ , also referred to as the degree of uncorrelated motion, is defined as the ratio of ionic conductivity to the Nernst–Einstein conductivity.<sup>39,97</sup> Both measures are directly comparable to experiment:  $\alpha_s$  via Raman spectroscopy<sup>96</sup> and  $\alpha_d$  via NMR experiments combined with impedance measurements.<sup>97</sup> While  $\alpha_s$  and  $\alpha_d$  are not directly interchangeable, they exhibit similar trends



with varying solvent composition, as observed in both computational<sup>39</sup> and experimental<sup>96</sup> studies. For instance, in the GBL/DMC system with LiFSI salts,  $\alpha_s$  and  $\alpha_d$  were found to be comparable at  $x_{\text{GBL}} > 0.4$ .<sup>96</sup> However, as  $x_{\text{GBL}}$  approached zero,  $\alpha_s$  decayed more rapidly than  $\alpha_d$ . In our study, we employed the fraction of free cations,  $\alpha_s$ , to represent the degree of ion dissociation, as calculating  $\alpha_d$  requires long-term MD simulations.

Fig. 5b presents computational results alongside experimental data and results from other force fields. An increase in the degree of ion dissociation with higher  $x_{\text{EC}}$  is observed, consistent with experimental trends.<sup>96</sup> These results are not influenced by the initial ion pair configuration, whether dissociated or associated. APPLE&P predicted both  $\alpha_s$  and  $\alpha_d$  approaching zero at low  $x_{\text{EC}}$  but produced relatively lower values of  $\alpha_s$  (or  $\alpha_d$ ) at higher  $x_{\text{EC}}$ .<sup>39</sup> In contrast, the Class-II force field predicted largely dissociated Li ions even at  $x_{\text{EC}} = 0.1$ ,<sup>100</sup> which significantly disagrees with experimental data (to note, the results in ref. 39 and 100 were obtained at a salt concentration of 1 M, not 1 mol kg<sup>-1</sup>. Here, 1 mol kg<sup>-1</sup> of LiPF<sub>6</sub> corresponds to 1.02–1.16 M, depending on  $x_{\text{EC}}$ ). The observed dissociation trend is consistent with the decreasing coordination of PF<sub>6</sub><sup>-</sup> ions with increasing  $x_{\text{EC}}$  in Fig. 5a.

In Fig. S9b† and the accompanying text, we conducted a similar analysis of the degree of ion dissociation for EC, PC, DMC, and DEC with 1 M LiPF<sub>6</sub> salts, using solvents of similar donor numbers.<sup>102</sup> The dominance of LiPF<sub>6</sub> ion pairs in solvents with low dielectric constants, such as DMC and DEC, and their dissociation in high dielectric constant solvents, such as EC and PC, are consistent with the results for EC/DMC mixtures and align with infrared spectroscopy analysis.<sup>103</sup> Additionally, steric hindrance from bulkier alkyl groups leads to relatively weaker salt dissociation in PC and DEC compared to EC and DMC, as the Li–O interactions are hindered by the increased bulkiness of the solvent molecules.<sup>104,105</sup> These results demonstrate that the pretrained potential effectively captures the dynamic variations in dielectric shielding between cyclic and linear solvents, suggesting that it implicitly learns long-range interactions to a certain extent.

**2.3.3 Diffusivities.** The diffusivity of anions and cations in electrolytes is an important property that determines battery performance. We theoretically obtained the self-diffusion coefficient ( $D$ ) for Li<sup>+</sup> and PF<sub>6</sub><sup>-</sup> in pure solvents like PC and DMC as well as mixed solvents of EC/DMC, for which the experimental data are available. As in the previous subsection, we employed the experimental densities, while results with the theoretical densities are also discussed.

For single solvent electrolyte systems, simulations were performed in the NVT ensemble, following a procedure largely similar to that used for the EC/DMC 1 mol kg<sup>-1</sup> LiPF<sub>6</sub> system described in the previous subsection. The temperatures were set according to where the experimental diffusivities were measured: 293 K for PC or DEC<sup>106</sup> and 298 K for DMC.<sup>107</sup> Experimental densities were used to generate initial configurations,<sup>108–110</sup> and the numbers of solvent and salt molecules were adjusted to achieve a 1 M LiPF<sub>6</sub> concentration with total number of atoms close to 1000. Detailed information

on the number of molecules and the simulation box lengths are listed in Table S7.† An equilibration run of 1.4 ns was followed by a 1 ns production run, similar to the EC/DMC 1 mol kg<sup>-1</sup> LiPF<sub>6</sub> simulation. Snapshots were sampled every 100 fs for calculation of ion diffusivity. Five independent runs were conducted with different initial configurations for statistical average.

For the EC/DMC 1 mol kg<sup>-1</sup> LiPF<sub>6</sub> binary solvent electrolyte system, longer production runs were found to be essential for diffusion analysis, likely due to the more complex solvation nature of ions in binary solvents compared to single solvent electrolyte systems. Starting from the NVT MD simulations described in the former subsection, four compositions were selected, and the production runs for each composition were extended to 7 ns across three independent runs to obtain reliable diffusivity data.

From the MD trajectories during the production run, the self-diffusion coefficient was calculated by the mean squared displacement (MSD) for each ion type. In detail, with a given time window  $\tau$  and position vector  $\mathbf{r}_i(t)$  of a particle  $i$  at time  $t$ , the squared displacement was averaged over all particles with the same type and all available time origins:<sup>111</sup>

$$\text{MSD}(\tau) = \langle |\mathbf{r}_i(t + \tau) - \mathbf{r}_i(t)|^2 \rangle_{t,i} \quad (3)$$

$$= \frac{1}{N_\tau} \sum_{t=0}^{T-\tau} \frac{1}{N} \sum_{i=0}^N |\mathbf{r}_i(t + \tau) - \mathbf{r}_i(t)|^2 \quad (4)$$

where  $N_\tau$  is the number of available time windows and  $N$  is the number of particles of interest. We then obtained  $D$  using the Einstein relation.<sup>112</sup>

$$D = \lim_{t \rightarrow \infty} \frac{\text{MSD}(t)}{6t} \quad (5)$$

The linear regression of an MSD– $\tau$  curve using appropriate bounds efficiently captures the linear region of diffusion.<sup>111</sup> We set the lower and upper bound as 10% and 60% of the total production time, respectively, ensuring accurate identification of the linear region in the MSD– $\tau$  curve yielding an  $R^2$  value mostly above 0.98 (Fig. S13†). This approach filters out the short-time regime including ballistic motion and the noisy long-time tail due to fewer available time origins, ensuring a reliable diffusivity estimate.

In Fig. 6a and b, we present the calculated diffusivities of Li<sup>+</sup> and PF<sub>6</sub><sup>-</sup> ( $D_{\text{Li}}$  and  $D_{\text{PF}_6}$ , respectively) when 1 M LiPF<sub>6</sub> is desolved in PC, DMC, or DEC solvents. The corresponding MSD– $\tau$  curves with the  $R^2$  values obtained from the linear regression are presented in Fig. S13.† Relatively large standard deviations in diffusivity values, up to 50%, are attributed to statistical errors arising from the relatively short MD time scale compared to typical classical force field simulations. For comparison, other computational<sup>37,39,45</sup> and experimental<sup>108,113</sup> results are also provided. The actual diffusivity values, along with predicted and experimental densities, can be found in Table S8.† Among the computational approaches, SevenNet achieved reasonable accuracy in a consistent way. In comparison, other computational methods showed larger discrepancies with experiment



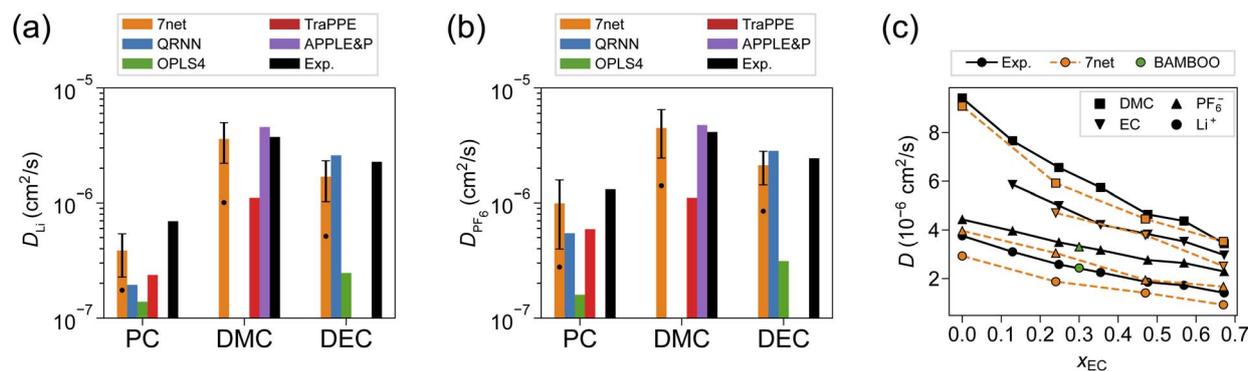


Fig. 6 Comparison of diffusivities of (a) Li and (b)  $\text{PF}_6^-$  obtained using SevenNet at experimental density (orange), QRNN<sup>45</sup> (blue), OPLS4<sup>45</sup> (green), TraPPE<sup>37</sup> (red), APPLE&P<sup>39</sup> (purple), and experiments<sup>108,113</sup> (black). Error bars indicate the standard deviation across five simulations. The black dots on the orange bar indicate the calculated diffusivity at the equilibrium density obtained from SevenNet-NPT simulations. (c) Diffusivity of solvents and ions in EC/DMC (1 mol  $\text{kg}^{-1}$   $\text{LiPF}_6$ ) electrolyte at 298 K. Values were averaged over 3 simulations. The actual diffusivity values and standard deviations can be found in Table S9.† Experimental values were adopted from ref. 96. The results by BAMBOO are also displayed.<sup>50</sup>

although they accurately predicted pure solvent density. This might be attributed to the general overestimation of viscosity in the QRNN and OPLS4 models and the overestimation of relative permittivity in the TraPPE model. The former slows molecular movement, while the latter affects solvent dynamics. In contrast, the APPLE&P model demonstrated less than 4% error in the density of 1 M  $\text{LiPF}_6$  in the DMC solvent and achieved accuracy comparable to SevenNet. SevenNet also successfully captured the general trend of higher anion diffusivities compared to cations, consistent with previous studies.<sup>28</sup> This phenomenon likely results from the strong solvation of Li ions by the solvents, which hinders their mobility relative to  $\text{PF}_6^-$  ions. Consistently, the higher permittivity of PC enhances solvation and suppresses Li-ion diffusivity relative to the DMC solvent. In Fig. 6a and b, the diffusivities obtained when the SevenNet equilibrium density was used are marked in black dots in the middle of the bar. The diffusivity was significantly reduced by 50–70%, underscoring the critical role of density. Fine-tuning with explicitly sampled intermolecular interactions, or developing a pretrained model incorporating both organic and inorganic datasets, are potential solutions to this problem.

In Fig. 6c, the diffusivity of ions in the EC/DMC (1 mol  $\text{kg}^{-1}$   $\text{LiPF}_6$ ) binary solvent system shows good agreement with experimental values, albeit with slight underestimation.<sup>96</sup> The overall decrease in diffusivity with increasing  $x_{\text{EC}}$  reflects reduced electrolyte mobility as the concentration of the highly viscous EC solvent increases, indicating that the variation in viscosity is also well described by SevenNet. Notably, the general trend of  $D_{\text{DMC}} > D_{\text{EC}} > D_{\text{PF}_6^-} > D_{\text{Li}^+}$  is consistent across all compositions. The lowest diffusivity of  $\text{Li}^+$  suggests its largest hydrodynamic size, even greater than that of the  $\text{PF}_6^-$  anion, due to strong coordination with surrounding solvent molecules.<sup>96</sup>

#### 2.4 Analysis on the training dataset

In the preceding subsections, the overall performance of SevenNet in simulating liquid electrolytes was found to be

satisfactory. To investigate whether this accuracy arises from adequate sampling in the training set, we analyzed the Materials Project Trajectory (MPtrj) dataset<sup>52</sup> for the presence of relevant solvent molecules and chemical moieties. We first identified molecular units embedded in the inorganic compounds that include O, C, and H atoms. The detailed procedure is described in the text accompanying Fig. S14.† Among the 20 solvent molecules studied in this work, only the DME molecule was found in four compounds (Fig. S14a†). To explore the presence of similar types of molecules, we manually inspected molecular units consisting of 6–24 atoms, including O, C, and H, and identified similar motifs, such as ether, ester, and five-membered ring groups (Fig. S14b†). However, no carbonate groups were found.

We further examined the presence of the local chemical moieties classified in Fig. S3.† As shown in Fig. S15,† structures containing species such as  $\text{O}_c/\text{C}_c$ ,  $\text{C}=\text{C}$ , and  $\text{C}_{\text{CH}_3}$  were prevalent in the training set. This abundance of these chemical moieties can be understood by general chemistry. However, several moieties were scarce in the training set, in particular fluorinated ones. For example, only 32 structures contained the  $\text{C}_{\text{OFH}_2}$  moiety, and the  $\text{C}_{\text{CFH}_2}$  moiety was absent. Such data scarcity may explain the pronounced softening for the fluorine atoms in Fig. 2c.

Extending the analysis to the Li solvation shell, we identified structures in the MPtrj training set containing  $\text{Li}-\text{O}_c$  or  $\text{Li}-\text{O}_e$ , as well as oxygen atoms bonded to  $\text{Na}^+$  and  $\text{K}^+$ , which share similar chemical properties with  $\text{Li}^+$ . Specifically, we found 183 and 63 structures containing  $\text{Li}-\text{O}_c$  and  $\text{Li}-\text{O}_e$  moieties, respectively (see Fig. S16† for representative structures). The  $\text{Li}-\text{O}_c$  structures predominantly included carbonate ions and their derivatives. Notably, a substantial number of these structures contained transition metals coexisting with  $\text{Li}^+$ ,  $\text{Na}^+$ , and  $\text{K}^+$ , which are commonly used in LIB cathodes. The exhaustive search for new cathode materials has likely contributed these structures to the databases. The Li ions are mostly charged in the corresponding structures, which allowed SevenNet to learn Coulomb interactions between  $\text{Li}^+$  and  $\text{O}_c$  or  $\text{O}_e$ .



The foregoing analysis indicates that significant portions of molecular units or chemical moieties were directly sampled in the MPtrj dataset, suggesting that relevant local chemical bondings were mostly captured. However, non-local bondings that contribute to the formation of whole molecular units and intermolecular interactions, in particular between organic molecules, were not sufficiently sampled. This suggests that the model learned large parts of the PES by generalizing across the chemical space, facilitated by deep learning and learnable atomic embeddings.

To understand model generalization in terms of latent space, we paid attention to  $C_{OF2H}$  and Li solvation shells. The  $C_{OF2H}$  moiety was absent in the training set but exhibited a notable degree of force accuracy (see Fig. 2c). This means that the model interpolate or extrapolate from other structures in the training set. For Li solvation shells, while the training set included structures containing Li–O<sub>c</sub> and Li–O<sub>e</sub> moieties, it remained uncertain how the model handled multiple Li–O interactions within solvation shells.

For both test cases, we analyzed the atomic descriptors in the latent space to examine relative proximity or similarity between structures. We first extracted 128-dimensional invariant atomic descriptors, which served as input vectors for the output block producing atomic energies. Dimensionality reduction techniques, including Principal Component Analysis (PCA) and Uniform Manifold Approximation and Projection (UMAP),<sup>144</sup> were employed. Using PCA, we reduced the dimensionality to 31 components, retaining 95% of the original variance, and applied a whitening scheme. As shown in Fig. S17a,† SevenNet positioned the  $C_{OF2H}$  moiety—absent in the training set—between the  $C_{OFH2}$  and  $C_{OF3}$  moieties. Furthermore, as the F : H ratio shifted from 3 : 0 to 0 : 3, the PCA data points for each moiety aligned linearly. That is to say, SevenNet interpolated untrained regions by leveraging knowledge derived from trained regions. This brings some evidence to why foundational models perform so unexpectedly well on untrained domain.

Next, we performed UMAP analysis<sup>145</sup> on both the training set and sampled structures from the simulation on the Li solvation shells. Euclidean distances between atomic descriptors were used to construct a high-dimensional graph, where similar descriptors were positioned closer together, and dissimilar descriptors were placed farther apart. In particular, we examined distinct solvation types identified in Section 2.3.1 by extracting MD trajectories (Fig. S17b†) for separate analysis. For each solvation type, the structure with the minimum potential energy among all snapshots was selected. The analysis in Fig. S17b† revealed that the Li environments within solvation shells formed distinct clusters in the latent space. The structures in the training set interpreted as being similar to these Li environments typically featured multiple oxygen atoms bonded to either Li or Na. While the training set did not include the same Li–solvent structures encountered during MD simulations, the model effectively learned first-neighbor interactions from these examples, where multiple oxygen atoms are bonded to a Li ion.

The above analysis on the latent space indicates that SevenNet might infer untrained regions by generalizing knowledge

from trained regions. However, full understanding of the generalization is beyond the current scope because of the black-box nature of deep learning models.

## 2.5 Fine-tuning the pretrained model

In the previous sections, we have demonstrated the capabilities and limitations of SevenNet on liquid electrolytes. Intermolecular interactions, particularly liquid density, are critical to the physicochemical properties of liquid electrolytes. However, SevenNet was less accurate for the intermolecular interactions, leading to the overestimation in the liquid density and underestimation of diffusivities. It has been shown that fine-tuning pretrained models can achieve accuracies comparable to bespoke models.<sup>60–62</sup> As an example of fine-tuning in the present applications, we selected the DMC solvent, which exhibited significant overestimations of density in Fig. 3a. In this simple fine-tuning for DMC, the model may lose some generalizability across different systems; however, our primary goal is to assess whether its density predictions can be improved with relatively modest efforts.

To construct the training set for fine-tuning, we conducted MD simulations using SevenNet for 100 ps in the 298 K NVT ensemble with 360 atoms and the experimental density. Subsequently, DFT single-point calculations were performed on 50 snapshots extracted from the last 50 ps at intervals of 1 ps. These 50 structures were further modified by scaling the lattice parameters, while maintaining fixed intramolecular distances, by factors of 0.9 and 1.1.<sup>46</sup> This procedure generated a total of 150 structures for the fine-tuning training set.

We fine-tuned the model using the same parameters as SevenNet, with adjustments to the learning rate and stress loss weight. The learning rate started at  $10^{-4}$  and decreased linearly to  $10^{-6}$  over 600 epochs. Furthermore, the weight of the stress loss was increased from 0.01 to 1.0 to enhance the fine-tuning effect on solvent density. For fine-tuning, we did not freeze any weights, allowing the model to adjust solely to the new fine-tuning training set. To monitor knowledge retention from the original SevenNet model, we evaluated the MAEs on a test set of 19 072 structures containing O, C, and H, filtered from the SevenNet training set (see Fig. S18†). The fine-tuned model (hereafter referred to as SevenNet-FT) after 50 epochs, achieved MAEs of 0.032 eV per atom, 0.086 eV Å<sup>-1</sup>, and 0.57 kbar for energy, forces, and stresses, respectively. Notably, the stress MAE decreased from 2.78 kbar (SevenNet) to 0.57 kbar (SevenNet-FT). In terms of computational cost, the entire fine-tuning procedure, including training set generation, required only a few hours on a moderate computing node.

Using SevenNet-FT, we obtained the liquid densities following the procedure described in Section 2.2, and the results are shown in Fig. 3a as 'x' markers. The liquid densities of linear carbonates align well with experimental results, whereas those of other solvents are underestimated. The normal stress parity plots for pure solvents (Fig. S19†) reveal significant improvements for linear carbonates, moderate improvements for cyclic carbonates, ethers, and esters, but only minor improvements for solvents containing fluorine atoms. This behavior may be attributed to the fine-tuning training set,



which includes only DMC molecules composed of O, C, and H atoms, thereby limiting its applicability to cyclic carbonates and fluorinated systems. This suggests that explicitly including more relevant intermolecular interactions in the fine-tuning dataset can significantly improve density predictions for the targeted molecular class. Additionally, forgetting of knowledge from SevenNet was not fully prevented during the fine-tuning process (Fig. S18†), which may have contributed to the underestimation of densities for PC, FEC, and DFEC. This forgetting could potentially be mitigated by replay methods<sup>116,117</sup> or elastic weight consolidation,<sup>118</sup> although implementing these strategies is beyond the scope of the present work.

### 3 Conclusions

In summary, we applied a pretrained universal interatomic potential, SevenNet, to the simulations of liquid electrolytes in LIBs. Even though SevenNet was mostly trained on the inorganic compounds, it demonstrated sound predictive capabilities for key properties such as solvation structures and diffusivities. However, it also exhibited limitations, particularly in predicting liquid density. These limitations stem from under-sampled intermolecular interactions in the training set; fine-tuning with relevant structures leads to improvements. However, the weak and long-range nature of these interactions makes them inherently difficult for MLIPs to select appropriate descriptors or to obtain effective learned representations, posing a challenge for predicting liquid densities. As such, accurate density prediction requires extensive sampling and careful force field representation,<sup>46</sup> underscoring the challenge of achieving high accuracy. Moreover, accurate description of density may require higher-level theories beyond DFT or exploring alternative machine learning architectures.

Despite these challenges, the model's ability to generalize across chemical spaces improved the accuracy in the strongly out-of-distribution domains. Analysis of latent space suggested that SevenNet leverages learnings from related chemical moieties and structural motifs to interpolate and predict untrained regions. Fine-tuning SevenNet for specific cases, as demonstrated with DMC solvents, significantly improved accuracy in density and stress predictions, with minimal computational costs. This approach paves the way for tailoring pretrained models to specialized applications, making them useful tools for material discovery and optimization in electrolyte engineering. Another potential approach would be to train a model using both an inorganic crystal database and a molecular database<sup>69</sup> through multi-fidelity training,<sup>119</sup> which will be explored in a future study. In conclusion, this work underscores the potential of SevenNet for advancing the engineering of liquid electrolyte systems, thereby accelerating the development of next-generation LIBs.

## 4 Methods

### 4.1 DFT calculation

All DFT calculations in this work were performed using the Vienna *ab initio* simulation package (VASP), employing the

projector-augmented wave (PAW) pseudopotentials.<sup>120</sup> The Perdew–Burke–Ernzerhof (PBE) exchange–correlation functional, based on the generalized gradient approximation (GGA), was used for electrons.<sup>121</sup> For condensed phases of organic molecules, the van der Waals dispersion interaction plays an important role in determining quantities such as density and viscosity. Since semilocal functionals such as PBE do not take into account the dispersion interactions, we added the Grimme's D3 dispersion correction with Becke–Johnson (BJ) damping.<sup>122,123</sup> The dispersion and coordination cutoff radii in the D3 correction term were set to 50.2 and 20.0 Å, respectively. The PBE-D3 functional shows similar accuracy to the PBE0-D3 (hybrid GGA) functional in predicting dimer interaction energies, while also achieving density predictions for organic crystals<sup>124</sup> and ionic liquids<sup>125</sup> that fall within 1% of experimental values. For molecules and dimers in vacuum, spin-polarized calculations were performed with a cutoff energy of 520 eV, while for bulk liquid configuration, a spin-unpolarized setting was used. However, we found that the spin-polarization was negligible in all isolated molecules (differences of <1 meV per atom and <3 meV Å<sup>-1</sup> for energies and atomic forces, respectively). The Brillouin zone was sampled only at the *Γ*-point. Concerning the specific PAW pseudopotentials, those without suffixes were used except for Li\_sv, in alignment with the calculation settings in Materials Project.

### 4.2 Pretrained model

In this work, we utilized a pretrained model SevenNet-0 (version 11July2024),<sup>56,72</sup> which is based on the architecture of NequIP.<sup>63</sup> SevenNet has achieved high performance in the Matbench Discovery benchmark, which assesses the performance of pretrained universal force fields on inorganic crystal discovery.<sup>73</sup> As a GNN-IP, SevenNet initializes node and edge features from atomic numbers and relative position vectors, respectively. An edge connects two nodes if their interatomic distance is less than a pre-defined cutoff radius. Starting from these features, multiple message-passing layers aggregate information from connected nodes and edges. After the last message-passing layer updates node features, they are used to predict total energy by the readout layer. Forces and stresses are derived from the energy gradient. Although the SevenNet model hyperparameters remain identical to its previous version in ref. 56, the training dataset has been updated to the MPtrj dataset without dataset splitting. The learning rate was initialized at 0.01 and decreased linearly to 0.0001 over 600 epochs. As a result, SevenNet achieves the MAEs of 0.011 eV per atom, 0.041 eV Å<sup>-1</sup>, and 2.78 kbar for energy, forces, and stresses, respectively.

We employed the Atomic Simulation Environment (ASE) interface<sup>126</sup> and the Large-scale Atomic/Molecular Massively Parallel Simulator (LAMMPS) package,<sup>83</sup> in conjunction with SevenNet, to compute the structural and dynamical properties of solvents and electrolytes. The ASE interface was used for static calculations such as (constrained) relaxation while the LAMMPS package was used for MD simulations. To account for dispersion interactions absent in the MPtrj dataset, we



integrated an in-house CUDA implementation of Grimme's D3 dispersion correction with BJ damping<sup>122,123</sup> into SevenNet, ensuring that the calculations were performed at the PBE-D3 level of theory, which is identical to the DFT calculation in the previous section. The parameter set for the dispersion interaction was consistent with that used in the DFT calculations (see above).

## Data availability

The data supporting this article have been included as part of the ESI† of this article. Data and processing scripts for this paper, including the MD trajectories and fine-tuning dataset, are available at Zenodo at <https://doi.org/10.5281/zenodo.15205477>. The SevenNet code utilized in this study is available in the project GitHub repository (<https://www.github.com/MDIL-SNU/SevenNet>).

## Author contributions

S. Ju and J. You conducted the DFT calculations, simulations using the pretrained model, data analyses, and script writing. G. Kim and H. An developed the CUDA-accelerated D3 implementation for LAMMPS. Y. Park generated SevenNet-0 and provided guidance on potential-related scripting. S. Han supervised all aspects of the work, including DFT calculations, simulations, and script development. All authors reviewed, provided feedback on, and approved the final manuscript.

## Conflicts of interest

There are no conflicts to declare.

## Acknowledgements

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (RS-2023-00247245). The computations were carried out at the Korea Institute of Science and Technology Information (KISTI) National Supercomputing Center (KSC-2024-CRE-0215) and at the Center for Advanced Computations (CAC) at Korea Institute for Advanced Study (KIAS).

## References

- 1 A. Manthiram, *ACS Cent. Sci.*, 2017, **3**, 1063–1069.
- 2 Z. Zhu, T. Jiang, M. Ali, Y. Meng, Y. Jin, Y. Cui and W. Chen, *Chem. Rev.*, 2022, **122**, 16610–16751.
- 3 Y. S. Meng, V. Srinivasan and K. Xu, *Science*, 2022, **378**, eabq3750.
- 4 Z. Yu, H. Wang, X. Kong, W. Huang, Y. Tsao, D. G. Mackanic, K. Wang, X. Wang, W. Huang, S. Choudhury, Y. Zheng, C. V. Amanchukwu, S. T. Hung, Y. M. Ma, E. G. Lomeli, J. Qin, Y. Cui and Z. Bao, *Nat. Energy*, 2020, **5**, 526–533.
- 5 Z. Yu, P. E. Rudnicki, Z. Zhang, Z. Huang, H. Celik, S. T. Oyakhire, Y. Chen, X. Kong, S. C. Kim, X. Xiao, H. Wang, Y. Zheng, G. A. Kamat, M. S. Kim, S. F. Bent and J. Qin, *Nat. Energy*, 2022, **7**, 94–106.
- 6 M. Li, C. Wang, Z. Chen, K. Xu and J. Lu, *Chem. Rev.*, 2020, **120**, 6783–6819.
- 7 M. Li, R. P. Hicks, Z. Chen, C. Luo, J. Guo, C. Wang and Y. Xu, *Chem. Rev.*, 2023, **123**, 1712–1773.
- 8 I. G. J. Methlie, *US Pat.*, US3415687A, 1968.
- 9 R. Fong, U. Von Sacken and J. R. Dahn, *J. Electrochem. Soc.*, 1990, **137**, 2009.
- 10 K. Xu, *Chem. Rev.*, 2004, **104**, 4303–4418.
- 11 J. Tarascon and D. Guyomard, *Solid State Ionics*, 1994, **69**, 293–305.
- 12 S. Chen, F. Dai and M. Cai, *ACS Energy Lett.*, 2020, **5**, 3140–3151.
- 13 A. Masias, J. Marcicki and W. A. Paxton, *ACS Energy Lett.*, 2021, **6**, 621–630.
- 14 J. Neubauer, A. Pesaran, C. Bae, R. Elder and B. Cunningham, *J. Power Sources*, 2014, **271**, 614–621.
- 15 K. Guo, S. Qi, H. Wang, J. Huang, M. Wu, Y. Yang, X. Li, Y. Ren and J. Ma, *Small Sci.*, 2022, **2**, 2100107.
- 16 X. Fan and C. Wang, *Chem. Soc. Rev.*, 2021, **50**, 10486–10566.
- 17 C.-Y. Wang, T. Liu, X.-G. Yang, S. Ge, N. V. Stanley, E. S. Rountree, Y. Leng and B. D. McCarthy, *Nature*, 2022, **611**, 485–490.
- 18 Y. Liu, Y. Zhu and Y. Cui, *Nat. Energy*, 2019, **4**, 540–550.
- 19 E. Logan and J. Dahn, *Trends Chem.*, 2020, **2**, 354–366.
- 20 N. Zhang, T. Deng, S. Zhang, C. Wang, L. Chen, C. Wang and X. Fan, *Adv. Mater.*, 2022, **34**, 2107899.
- 21 D. Hubble, D. E. Brown, Y. Zhao, C. Fang, J. Lau, B. D. McCloskey and G. Liu, *Energy Environ. Sci.*, 2022, **15**, 550–578.
- 22 J. Holoubek, H. Liu, Z. Wu, Y. Yin, X. Xing, G. Cai, S. Yu, H. Zhou, T. A. Pascal, Z. Chen and P. Liu, *Nat. Energy*, 2021, **6**, 303–313.
- 23 G. Song, Z. Yi, F. Su, L. Xie, Z. Wang, X.-X. Wei, G. Xu and C.-M. Chen, *ACS Energy Lett.*, 2023, **8**, 1336–1343.
- 24 N. Yao, X. Chen, Z.-H. Fu and Q. Zhang, *Chem. Rev.*, 2022, **122**, 10970–11021.
- 25 T. A. Pham, *Int. J. Quantum Chem.*, 2019, **119**, e25795.
- 26 X. Chen and H. Yu, *ChemElectroChem*, 2024, **11**, e202400444.
- 27 P. Ganesh, D.-e. Jiang and P. R. C. Kent, *J. Phys. Chem. B*, 2011, **115**, 3085–3090.
- 28 M. T. Ong, O. Verners, E. W. Draeger, A. C. T. v. Duin, V. Lordi and J. E. Pask, *J. Phys. Chem. B*, 2015, **119**, 1535–1545.
- 29 T. A. Pham, K. E. Kweon, A. Samanta, V. Lordi and J. E. Pask, *J. Phys. Chem. C*, 2017, **121**, 21913–21920.
- 30 E. W. C. Spotte-Smith, R. L. Kam, D. Barter, X. Xie, T. Hou, S. Dwaraknath, S. M. Blau and K. A. Persson, *ACS Energy Lett.*, 2022, **7**, 1446–1453.
- 31 M. Gerasimov, F. A. Soto, J. Wagner, F. Baakes, N. Guo, F. Ospina-Acevedo, F. Roder, P. B. Balbuena and U. Kreuer, *J. Phys. Chem. C*, 2023, **127**, 4872–4886.



- 32 M. Bin Jassar, C. Michel, S. Abada, T. De Bruin, S. Tant, C. Nieto-Draghi and S. N. Steinmann, *ACS Appl. Energy Mater.*, 2023, **6**, 6934–6945.
- 33 E. Crabb, A. France-Lanord, G. Leverick, R. Stephens, Y. Shao-Horn and J. C. Grossman, *J. Chem. Theory Comput.*, 2020, **16**, 7255–7266.
- 34 E. Crabb, A. Aggarwal, R. Stephens, Y. Shao-Horn, G. Leverick and J. C. Grossman, *J. Phys. Chem. B*, 2024, **128**, 3427–3441.
- 35 M. I. Chaudhari, J. R. Nair, L. R. Pratt, F. A. Soto, P. B. Balbuena and S. B. Rempe, *J. Chem. Theory Comput.*, 2016, **12**, 5709–5718.
- 36 J. Self, K. D. Fong and K. A. Persson, *ACS Energy Lett.*, 2019, **4**, 2843–2849.
- 37 Z. Luo, S. A. Burrows, S. K. Smoukov, X. Fan and E. S. Boek, *J. Phys. Chem. B*, 2023, **127**, 2224–2236.
- 38 H. Sun, S. J. Mumby, J. R. Maple and A. T. Hagler, *J. Am. Chem. Soc.*, 1994, **116**, 2978–2987.
- 39 O. Borodin and G. D. Smith, *J. Phys. Chem. B*, 2009, **113**, 1763–1776.
- 40 V. L. Deringer, A. P. Bartók, N. Bernstein, D. M. Wilkins, M. Ceriotti and G. Csányi, *Chem. Rev.*, 2021, **121**, 10073–10141.
- 41 O. T. Unke, S. Chmiela, H. E. Sauceda, M. Gastegger, I. Poltavsky, K. T. Schütt, A. Tkatchenko and K.-R. Müller, *Chem. Rev.*, 2021, **121**, 10142–10186.
- 42 C. Hong, J. Kim, J. Kim, J. Jung, S. Ju, J. M. Choi and S. Han, *Sci. Technol. Adv. Mater.:Methods*, 2023, **3**, 2269948.
- 43 S. Hwang, J. Jung, C. Hong, W. Jeong, S. Kang and S. Han, *J. Am. Chem. Soc.*, 2023, **145**, 19378–19386.
- 44 J. Jung, S. Ju, P.-h. Kim, D. Hong, W. Jeong, J. Lee, S. Han and S. Kang, *ACS Catal.*, 2023, **13**, 16078–16087.
- 45 S. Dajnowicz, G. Agarwal, J. M. Stevenson, L. D. Jacobson, F. Ramezanghorbani, K. Leswing, R. A. Friesner, M. D. Halls and R. Abel, *J. Phys. Chem. B*, 2022, **126**, 6271–6280.
- 46 I.-B. Magdáu, D. J. Arismendi-Arrieta, H. E. Smith, C. P. Grey, K. Hermansson and G. Csányi, *npj Comput. Mater.*, 2023, **9**, 146.
- 47 F. Wang and J. Cheng, *Chin. J. Struct. Chem.*, 2023, **42**, 100061.
- 48 D. Zhu, L. Sheng, T. Hu, S. Chen, M. Shi, H. Hua, K. Yang, J. Wang, Y. Tang, X. He and H. Xu, *J. Phys. Chem. Lett.*, 2024, **15**, 4024–4030.
- 49 O. Shayestehpour and S. Zahn, *J. Chem. Phys.*, 2024, **161**, 134505.
- 50 S. Gong, Y. Zhang, Z. Mu, Z. Pu, H. Wang, X. Han, Z. Yu, M. Chen, T. Zheng, Z. Wang, L. Chen, Z. Yang, X. Wu, S. Shi, W. Gao, W. Yan and L. Xiang, *Nat. Mach. Intell.*, 2025, **7**, 543–552.
- 51 C. Chen and S. P. Ong, *Nat. Comput. Sci.*, 2022, **2**, 718–728.
- 52 B. Deng, P. Zhong, K. Jun, J. Riebesell, K. Han, C. J. Bartel and G. Ceder, *Nat. Mach. Intell.*, 2023, **5**, 1031–1041.
- 53 S. Takamoto, C. Shinagawa, D. Motoki, K. Nakago, W. Li, I. Kurata, T. Watanabe, Y. Yayama, H. Iriguchi, Y. Asano, T. Onodera, T. Ishii, T. Kudo, H. Ono, R. Sawada, R. Ishitani, M. Ong, T. Yamahuchi, T. Kataoka, A. Hayashi, N. Charoenphakdee and T. Ibuka, *Nat. Commun.*, 2022, **13**, 2991.
- 54 A. Merchant, S. Batzner, S. S. Schoenholz, M. Aykol, G. Cheon and E. D. Cubuk, *Nature*, 2023, **624**, 80–85.
- 55 I. Batatia, P. Benner, Y. Chiang, A. M. Elena, D. P. Kovács, J. Riebesell, X. R. Advincula, M. Asta, M. Avaylon, W. J. Baldwin, F. Berger, N. Bernstein, A. Bhowmik, S. M. Blau, V. Cárare, J. P. Darby, S. De, F. D. Pia, V. L. Deringer, R. Elijošius, Z. El-Machachi, F. Falcioni, E. Fako, A. C. Ferrari, A. Genreith-Schriever, J. George, R. E. A. Goodall, C. P. Grey, P. Grigorev, S. Han, W. Handley, H. H. Heenen, K. Hermansson, C. Holm, J. Jaafar, S. Hofmann, K. S. Jakob, H. Jung, V. Kapil, A. D. Kaplan, N. Karimitari, J. R. Kermode, N. Kroupa, J. Kullgren, M. C. Kuner, D. Kuryla, G. Liepuoniute, J. T. Margraf, I.-B. Magdáu, A. Michaelides, J. H. Moore, A. A. Naik, S. P. Niblett, S. W. Norwood, N. O'Neill, C. Ortner, K. A. Persson, K. Reuter, A. S. Rosen, L. L. Schaaf, C. Schran, B. X. Shi, E. Sivonxay, T. K. Stenczel, V. Svahn, C. Sutton, T. D. Swinburne, J. Tilly, C. van der Oord, E. Varga-Umbrich, T. Vegge, M. Vondrák, Y. Wang, W. C. Witt, F. Zills and G. Csányi, *arXiv*, 2024, preprint, arXiv:2401.00096, <https://arxiv.org/abs/2401.00096>.
- 56 Y. Park, J. Kim, S. Hwang and S. Han, *J. Chem. Theory Comput.*, 2024, **20**, 4857–4868.
- 57 H. Yang, C. Hu, Y. Zhou, X. Liu, Y. Shi, J. Li, G. Li, Z. Chen, S. Chen, C. Zeni, M. Horton, R. Pinsler, A. Fowler, D. Zügner, T. Xie, J. Smith, L. Sun, Q. Wang, L. Kong, C. Liu, H. Hao and Z. Lu, *arXiv*, 2024, preprint, arXiv:2405.04967, <https://arxiv.org/abs/2405.04967>.
- 58 L. Barroso-Luque, M. Shuaibi, X. Fu, B. M. Wood, M. Dzamba, M. Gao, A. Rizvi, C. L. Zitnick and Z. W. Ulissi, *arXiv*, 2024, preprint, arXiv:2410.12771, <https://arxiv.org/abs/2410.12771>.
- 59 M. Neumann, J. Gin, B. Rhodes, S. Bennett, Z. Li, H. Choubisa, A. Hussey and J. Godwin, *arXiv*, 2024, preprint, arXiv:2410.22570, <https://arxiv.org/abs/2410.22570>.
- 60 B. Focassio, L. P. M. Freitas and G. R. Schleder, *ACS Appl. Mater. Interfaces*, 2025, **17**, 13111–13121.
- 61 H. Kaur, F. Della Pia, I. Batatia, X. R. Advincula, B. X. Shi, J. Lan, G. Csányi, A. Michaelides and V. Kapil, *Faraday Discuss.*, 2025, **256**, 120–138.
- 62 B. Deng, Y. Choi, P. Zhong, J. Riebesell, S. Anand, Z. Li, K. Jun, K. A. Persson and G. Ceder, *npj Comput. Mater.*, 2025, **11**, 1–9.
- 63 S. Batzner, A. Musaelian, L. Sun, M. Geiger, J. P. Mailoa, M. Kornbluth, N. Molinari, T. E. Smidt and B. Kozinsky, *Nat. Commun.*, 2022, **13**, 2453.
- 64 I. Batatia, D. P. Kovács, G. N. C. Simm, C. Ortner and G. Csányi, MACE: Higher Order Equivariant Message Passing Neural Networks for Fast and Accurate Force Fields, in *Advances in Neural Information Processing Systems (NeurIPS2022)*, ed. S. Koyejo, S. Mohamed, A. Agarwal, D. Belgrave, K. Cho and A. Oh, Curran Associates, Inc., 2022, vol. 35, pp. 11423–11436.



- 65 Y.-W. Zhang, V. Sorkin, Z. H. Aitken, A. Politano, J. Behler, A. P. Thompson, T. W. Ko, S. P. Ong, O. Chalykh, D. Korogod, E. Podryabinkin, A. Shapeev, J. Li, Y. Mishin, Z. Pei, X. Liu, J. Kim, Y. Park, S. Hwang, S. Han, K. Sheriff, Y. Cao and R. Freitas, *Modell. Simul. Mater. Sci. Eng.*, 2025, **33**, 023301.
- 66 A. Jain, S. P. Ong, G. Hautier, W. Chen, W. D. Richards, S. Dacek, S. Cholia, D. Gunter, D. Skinner, G. Ceder and K. A. Persson, *APL Mater.*, 2013, **1**, 011002.
- 67 J. Schmidt, T. F. Cerqueira, A. H. Romero, A. Loew, F. Jäger, H.-C. Wang, S. Botti and M. A. Marques, *Mater. Today Phys.*, 2024, **48**, 101560.
- 68 S. P. Niblett, P. Kourtis, I.-B. Magdău, C. P. Grey and G. Csányi, *arXiv*, 2024, preprint, arXiv:2409.05590, <https://arxiv.org/abs/2409.05590>.
- 69 P. Eastman, P. K. Behara, D. L. Dotson, R. Galvelis, J. E. Herr, J. T. Horton, Y. Mao, J. D. Chodera, B. P. Pritchard, Y. Wang, G. De Fabritiis and T. E. Markland, *Sci. Data*, 2023, **10**, 11.
- 70 D. P. Kovács, J. H. Moore, N. J. Browning, I. Batatia, J. T. Horton, Y. Pu, V. Kapil, W. C. Witt, I.-B. Magdău, D. J. Cole and G. Csányi, *arXiv*, 2025, preprint, arXiv:arXiv:2312.15211, <https://arxiv.org/abs/2312.15211>.
- 71 S. Kang, *J. Chem. Phys.*, 2024, **161**, 244102.
- 72 *SevenNet-0*(11July2024), <https://www.github.com/MDIL-SNU/SevenNet.git>, (accessed August 2024).
- 73 *Matbench Discovery*, <https://www.matbench-discovery.materialsproject.org>, (accessed August 2024).
- 74 B. P. Kar, N. Ramanathan, K. Sundararajan and K. Viswanathan, *J. Mol. Struct.*, 2012, **1024**, 84–93.
- 75 Y. Wang and P. B. Balbuena, *Int. J. Quantum Chem.*, 2005, **102**, 724–733.
- 76 H. Lee, S. Hwang, M. Kim, K. Kwak, J. Lee, Y.-K. Han and H. Lee, *J. Phys. Chem. Lett.*, 2020, **11**, 10382–10387.
- 77 D. J. Evans and B. L. Holian, *J. Chem. Phys.*, 1985, **83**, 4069–4074.
- 78 R. L. Hurler and L. A. Woolf, *J. Chem. Soc., Faraday Trans. 1*, 1982, **78**, 2233–2238.
- 79 *Molview (v2.4)*, <https://www.molview.org>.
- 80 L. Martínez, R. Andrade, E. G. Birgin and J. M. Martínez, *J. Comput. Chem.*, 2009, **30**, 2157–2164.
- 81 M. Mantina, A. C. Chamberlin, R. Valero, C. J. Cramer and D. G. Truhlar, *J. Phys. Chem. A*, 2009, **113**, 5806–5812.
- 82 W. Shinoda, M. Shiga and M. Mikami, *Phys. Rev. B*, 2004, **69**, 134103.
- 83 A. P. Thompson, H. M. Aktulga, R. Berger, D. S. Bolintineanu, W. M. Brown, P. S. Crozier, P. J. in 't Veld, A. Kohlmeyer, S. G. Moore, T. D. Nguyen, R. Shan, M. J. Stevens, J. Tranchida, C. Trott and S. J. Plimpton, *Comput. Phys. Commun.*, 2022, **271**, 108171.
- 84 R. Väli, A. Jänes and E. Lust, *J. Electrochem. Soc.*, 2016, **163**, A851.
- 85 K. Hagiwara, K. Suzuki, M. Ohtake, M. Shimada, N. Nanbu, M. Takehara, M. Ue and Y. Sasaki, *Chem. Lett.*, 2008, **37**, 210–211.
- 86 P. Zheng, X. Meng, J. Wu and Z. Liu, *Int. J. Thermophys.*, 2008, **29**, 1244–1256.
- 87 M. Almasi, A. Hernández and J. Zarrini, *J. Chem. Eng. Data*, 2023, **68**, 1911–1920.
- 88 U.S. Coast Guard, *Chemical Hazards Response Information System (CHRIS) Hazardous Chemical Data*, Department of Transportation, Washington, DC, 1999.
- 89 J. M. França, C. A. Nieto de Castro, M. M. Lopes and V. M. Nunes, *J. Chem. Eng. Data*, 2009, **54**, 2569–2575.
- 90 S.-K. Yang, S.-J. Peng, J.-H. Huang, L.-Q. Fan and F.-X. Yang, *J. Chem. Thermodyn.*, 2007, **39**, 773–780.
- 91 F. Chen, Z. Yang, Z. Chen, J. Hu, C. Chen and J. Cai, *J. Mol. Liq.*, 2015, **209**, 683–692.
- 92 A. A. Wang, D. Persa, S. Helin, K. P. Smith, J. L. Raymond and C. W. Monroe, *J. Chem. Eng. Data*, 2023, **68**, 805–812.
- 93 N. Yao, X. Chen, Z.-H. Fu and Q. Zhang, *Chem. Rev.*, 2022, **122**, 10970–11021.
- 94 Y. Kameda, Y. Umebayashi, M. Takeuchi, M. A. Wahab, S. Fukuda, S.-i. Ishiguro, M. Sasaki, Y. Amo and T. Usuki, *J. Phys. Chem. B*, 2007, **111**, 6104–6109.
- 95 Y. Kameda, S. Saito, Y. Umebayashi, K. Fujii, Y. Amo and T. Usuki, *J. Mol. Liq.*, 2016, **217**, 17–22.
- 96 S. Uchida and T. Kiyobayashi, *Phys. Chem. Chem. Phys.*, 2021, **23**, 10875–10887.
- 97 K. Hayamizu, *Electrochim. Acta*, 2017, **254**, 101–111.
- 98 O. Borodin, M. Olguin, P. Ganesh, P. R. C. Kent, J. L. Allen and W. A. Henderson, *Phys. Chem. Chem. Phys.*, 2016, **18**, 164–175.
- 99 T. Kiyobayashi, S. Uchida, H. Ozaki and K. Kiyohara, *J. Chem. Phys.*, 2023, **159**, 074501.
- 100 B. Ravikumar, M. Mynam, S. Repaka and B. Rai, *J. Mol. Liq.*, 2021, **338**, 116613.
- 101 K.-i. Saitoh, Y. Takai, T. Sato, M. Takuma and Y. Takahashi, *Batteries*, 2022, **8**, 27.
- 102 V. Gutmann and V. Gutmann, *The donor-acceptor approach to molecular interactions*, Springer, 1978, vol. 228.
- 103 C. M. Burba and R. Frech, *J. Phys. Chem. B*, 2005, **109**, 15161–15164.
- 104 X. Li, Y. Pan, Y. Liu, Y. Jie, S. Chen, S. Wang, Z. He, X. Ren, T. Cheng, R. Cao and S. Jiao, *Carbon Neutrality*, 2023, **2**, 34.
- 105 X. Yuan, L. Du, J. Li, Z. Liu, D. Lu and X. Kong, *J. Phys.: Condens. Matter*, 2024, **36**, 235101.
- 106 K. Hayamizu, *J. Chem. Eng. Data*, 2012, **57**, 2012–2017.
- 107 S. Miyoshi, H. Nagano, T. Fukuda, T. Kurihara, M. Watanabe, S. Ida and T. Ishihara, *J. Electrochem. Soc.*, 2016, **163**, A1206.
- 108 M. Takeuchi, Y. Kameda, Y. Umebayashi, S. Ogawa, T. Sonoda, S. ichi Ishiguro, M. Fujita and M. Sano, *J. Mol. Liq.*, 2009, **148**, 99–108.
- 109 SigmaAldrich, <https://www.sigmaaldrich.com/KR/ko/product/aldrich/746754>, (accessed August 2024).
- 110 SigmaAldrich, <https://www.sigmaaldrich.com/KR/ko/product/aldrich/746770>, (accessed August 2024).
- 111 X. He, Y. Zhu, A. Epstein and Y. Mo, *npj Comput. Mater.*, 2018, **4**, 18.
- 112 M. P. Allen and D. J. Tildesley, *Computer Simulation of Liquids*, Oxford University Press, United Kingdom, 2017.



- 113 S. Miyoshi, H. Nagano, T. Fukuda, T. Kurihara, M. Watanabe, S. Ida and T. Ishihara, *J. Electrochem. Soc.*, 2016, **163**, A1206.
- 114 L. McInnes, J. Healy and J. Melville, *arXiv*, 2020, preprint, arXiv:1802.03426, <https://arxiv.org/abs/1802.03426>.
- 115 *SevenNet-UMAP*, DOI: [10.5281/zenodo.14613413](https://doi.org/10.5281/zenodo.14613413).
- 116 H. Shin, J. K. Lee, J. Kim and J. Kim, *Adv. Neural Inf. Process. Syst.*, 2017, vol. 30.
- 117 D. Rolnick, A. Ahuja, J. Schwarz, T. Lillicrap and G. Wayne, *Adv. Neural Inf. Process. Syst.*, 2019, vol. 32.
- 118 J. Kirkpatrick, R. Pascanu, N. Rabinowitz, J. Veness, G. Desjardins, A. A. Rusu, K. Milan, J. Quan, T. Ramalho, A. Grabska-Barwinska, D. Hassabis, C. Clopath, D. Kumaran and R. Handsell, *Proc. Natl. Acad. Sci. U. S. A.*, 2017, **114**, 3521–3526.
- 119 J. Kim, J. Kim, J. Kim, J. Lee, Y. Park, Y. Kang and S. Han, *J. Am. Chem. Soc.*, 2024, **147**, 1042–1054.
- 120 G. Kresse and D. Joubert, *Phys. Rev. B*, 1999, **59**, 1758.
- 121 J. P. Perdew, K. Burke and M. Ernzerhof, *Phys. Rev. Lett.*, 1996, **77**, 3865.
- 122 S. Grimme, J. Antony, S. Ehrlich and H. Krieg, *J. Chem. Phys.*, 2010, **132**, 154104.
- 123 S. Grimme, S. Ehrlich and L. Goerigk, *J. Comput. Chem.*, 2011, **32**, 1456–1465.
- 124 A. Zhugayevych, W. Sun, T. van der Heide, C. R. Lien-Medrano, T. Frauenheim and S. Tretiak, *J. Chem. Theory Comput.*, 2023, **19**, 8481–8490.
- 125 C. Červinka, *J. Comput. Chem.*, 2022, **43**, 448–456.
- 126 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. D. 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*, 2017, **29**, 273002.

