Supporting Information to: Accurate Description of Ion Migration in Solid-State Ion Conductors from Machine-Learning Molecular Dynamics

Takeru Miyagawa, Namita Krishnan, Manuel Grumet,
Christian Reverón Baecker, Waldemar Kaiser,* and David A. Egger†
Department of Physics, TUM School of Natural Sciences,
Technical University of Munich, 85748 Garching, Germany

CONTENTS

* waldemar.kaiser@tum.de
† david.egger@tum.de

S1. ADDITIONAL INFORMATION ON COMPUTATIONAL METHODS

A. VASP 'on-the-fly' machine learning method

In this work, we employ the 'on-the-fly' machine learning method, implemented in VASP; the underlying theory and concepts are described in the literature [1–3]. In brief, we first trained the machine-learning force fields (MLFFs) using the 'on-the-fly' machine learning (ML) method. After training, we employed the MLFFs in MD simulations directly turning off the 'on-the-fly' ML component. The potential energy of a system containing N atoms is decomposed into individual atomic energies:

$$U = \sum_{i=1}^{N} U_i. \tag{S1}$$

Local configurations determine this local energy, which can be expressed as a functional $U = F[\rho_i]$ of the probability distribution ρ_i of finding another atom j in the local atomic environments:

$$\rho_i(r) = \sum_{j=1}^{N} f_{cut}(r_{ij}) g(r - r_{ij}). \tag{S2}$$

f_{cut} is a cutoff function to ensure that only atoms within a certain cutoff radius are included and $g(r)$ is approximated as

$$g(r) = \frac{1}{\sqrt{2\sigma_{\text{atom}}\pi}} \exp\left(-\frac{r^2}{2\sigma_{\text{atom}}^2}\right), \tag{S3}$$

with the broadening of the atomic distributions, σ_{atom}. In the VASP 'on-the-fly' machine learning method, radial and angular descriptors construct the local environment in a way such that rotational invariance is satisfied. The radial descriptor ($\rho_i^{(2)}$) and angular descriptors ($\rho_i^{(3)}$) are given as

$$\rho_i^{(2)}(r) = \frac{1}{\sqrt{4\pi}} \sum_{n=1}^{N_R^0} c_n^i \chi_{n0}(r), \tag{S4}$$

$$\rho_i^{(3)}(r, s, \theta) = \sum_{l=0}^{L_{max}} \sum_{n=1}^{N_R^l} \sum_{\mu=1}^{N_R^l} \sqrt{\frac{2l+1}{2}} p_{n\mu l}^i \chi_{\mu l}(r) \chi_{\mu l}(s) P_l(\cos(\theta)). \tag{S5}$$

Here, the expansion coefficients ($P_{n\mu l}$) for angular descriptors are

$$P_{n\mu l}^i = \sqrt{\frac{8\pi^2}{2l+1}} \sum_{m=-l}^{l} [c_{nlm}^i c_{\mu lm}^{i*} - \sum_{j}^{N_a} c_{nlm}^{ij} c_{\mu lm}^{ij*}], \tag{S6}$$

where $\chi_{\mu l}$ and P_l describe normalized spherical Bessel function and Legendre polynomials of order l, respectively. c^j_{nlm}, N^0_R, and N^l_R denote the expansion coefficients, number of radial basis functions for radial descriptors, and number of radial basis functions for angular descriptors. The indices n, l, m denote the radial, angular, and magnetic quantum numbers. The descriptors are then expressed for a local configuration i as

$$x_i^{(2)^T} = (c^i_{100}, c^i_{100}, ...)\,, \tag{S7}$$

as well as

$$x_i^{(3)^T} = (p^i_{110}, p^i_{111}, ..., p^i_{120}, p^i_{121}, ...)\,, \tag{S8}$$

which contain the coefficients c^i_{n00} and $p^i_{n\mu l}$, respectively. The local potential energy of atom i can then be expressed using a weighted kernel function, $K(x_i, x_{i_\mathrm{B}})$, where x_i and x_{iB} are local instantaneous and local reference configurations as

$$U_i^\alpha = \sum_{i_\mathrm{B}=1}^{N_B} \omega_{i_\mathrm{B}} K(x_i, x_{i_\mathrm{B}})\,, \tag{S9}$$

with basis set size N_B and the linear weights ω_{i_B} obtained from the regression model.

B. Detailed numerical parameters and settings

In the following, we provide all parameters used in the *ab initio* molecular dynamics (AIMD) and ML molecular dynamics (MLMD) simulations along the list of solid-state ion conductors (SSICs) discussed in the main text.

α-AgI: A 128 atom supercell with cell parameters $a = b = c = 18.32\,\text{Å}$ and $\alpha = \beta = \gamma = 109.47°$ was simulated at 500 K using a plane-wave cutoff energy of 280 eV [4]. We used the PAW potentials for Ag with $4d^{10}$, $5s^1$ valence configuration and I with $5s^2$, $5p^5$ valence configuration. The system was equilibrated for 10 ps before carrying out the production run of 100 ps, using a time step of 10 fs. AgI has only very low frequency vibrations due to the large atomic masses of Ag^+ and I^- ions, for which we found the mentioned time step of 10 fs to be sufficient. The training of the MLFF involved 4000 equilibrated AIMD snapshots in order to obtain well-converged BEEF values. The number of Bessel functions was set to 8 for both the radial and the angular parts, and the cutoff radius for both was set to 5 Å. The atomic coordinates were smeared in the distributions by placing Gaussians with a width of

0.5 Å for both radial angular descriptors. From these, 1696 local atomic configurations were selected for iodine atoms, and 2744 were selected for Ag atoms.

$Li_{10}GeP_2S_{12}$ (LGPS): A $2 \times 2 \times 1$ supercell of 200 atoms was used containing 80 Li, 8 Ge, 16 P and 96 S atoms. Cell parameters were set to experimental values of $a = 17.18$ Å, $b = 17.76$ Å, $c = 12.97$ Å, and $\alpha = 91.98°$, $\beta = 90.64°$, $\gamma = 90.25°$ [5]. PAW potentials were used for Li (3 electrons; $1s^2$, $2s^1$) Ge (14 electrons; $3d^{10}$, $4s^2$, $4p^2$), P (5 electrons; $3s^2$, $3p^3$), and S (6 electrons; $3s^2$, $3p^4$). A plane-wave cutoff of 500 eV was used, and the temperature was set to 500 K in all simulations because the LGPS tetragonal phase is stable at that temperature [6, 7]. The equilibration time for the simulation was 9 ps followed by a production run of 20 ps, using a time step of 2 fs. MLFFs were generated from 1000 equilibrated AIMD snapshots. The number of Bessel functions was set to 4 for both the radial and the angular descriptors, and the cutoff radii for the radial and angular descriptors were set to 9 Å and 6 Å, respectively. The atomic coordinates were smeared in the distributions by placing Gaussians with a width of 0.5 Å for both radial angular descriptors. From these, 2852 local atomic configurations were selected for Li atoms, 537 for Ge atoms, 711 for P atoms, and 3601 for S atoms.

Na_3SbS_4: We constructed a $2 \times 2 \times 2$ supercell consisting of 128 atoms. The employed PAW potentials were Na (7 electrons; $2p^6$, $3s^1$), Sb (5 electrons; $5s^2$, $5p^3$), and S (6 electrons; $3s^2$, $3p^4$). The tetragonal phase was employed for pristine Na_3SbS_4, with cell parameters set as $a = b = 14.333$ Å and $c = 14.581$ Å following experimental results from X-ray powder diffraction analysis [8]. The plane wave cutoff was set to 400 eV and the temperature of the NVT runs to 300 K, using a time step of 2 fs. We equilibrated the system for 5 ps, followed by a 50 ps production run.

To explore the effect of aliovalent doping in Na_3SbS_4, tungsten (W) was used as a dopant at a concentration of 6 %, replacing one Sb atom and adding a Na vacancy resulting in a doped supercell with 127 atoms. Cell parameters were set to $a = b = 14.3996$ Å and $c = 14.4604$ Å according to results from experimental measurements [9]. The 6 electron PAW potential of W with electrons from $5d^5$, $6s^1$ was used in addition to the ones reported above. The simulation length was 80 ps after equilibration of the system for 5 ps with a time step of 2 fs. The temperature was kept at 300 K.

The MLFFs of pristine Na_3SbS_4 were generated using 10000 AIMD snapshots. Here, an increased amount of training data was chosen in order to describe both the low-frequency

vibrational modes of Na^+ ions and the high-frequency ones of the SbS_4 tetrahedra reasonably well. As we will discuss later, these are separated in frequency, which we found demanded enhanced MLFF training times. The number of Bessel functions was set to 12 for the radial and to 8 for the angular part and their cutoff radius was set to 5 Å and 8 Å, respectively. The atomic coordinates were smeared in the distributions by placing Gaussians with a width of 0.5 Å for both radial angular descriptors. For pristine Na_3SbS_4, 834 local atomic configurations were selected for Na atoms, 310 were selected for Sb atoms, and 1100 for S atoms. For tungsten-doped $Na_{2.94}Sb_{0.94}W_{0.06}S_4$, we retrained the MLFF starting from the one of the pristine system, using a sample of 5000 AIMD steps. All other parameters were kept fixed. Here, 1097 local atomic configurations were selected for Na atoms, 304 were selected for Sb atoms, 1609 for S atoms, and 136 for W atoms.

The computational cost of AIMD simulations are compared with the cost of the active learning for generation of the MLFFs and with the cost of performing MLMD simulations using the pre-trained MLFF, see Table S1. We observe a tremendous reduction in the amount of required core-hours per 1000 MD simulation steps for the MLMD simulations with respect to the AIMD ones, highlighting further the potential of MLMDs towards high-throughput materials simulations.

System	AIMD	On-the-fly learning	MLMD
AgI	476	31.2	4
LGPS	693	125	10.2
$Na_{2.94}Sb_{0.94}W_{0.06}S_4$	306	33.2	5.8

TABLE S1. Computational costs for AIMD simulations in comparison to the on-the fly learning of MLFFs, and MLMD simulations using pre-trained MLFFs for systems under investigation. All values are given in units of core-hours per 1000 MD simulation steps.

S2. ADDITIONAL FIGURES

A. AgI

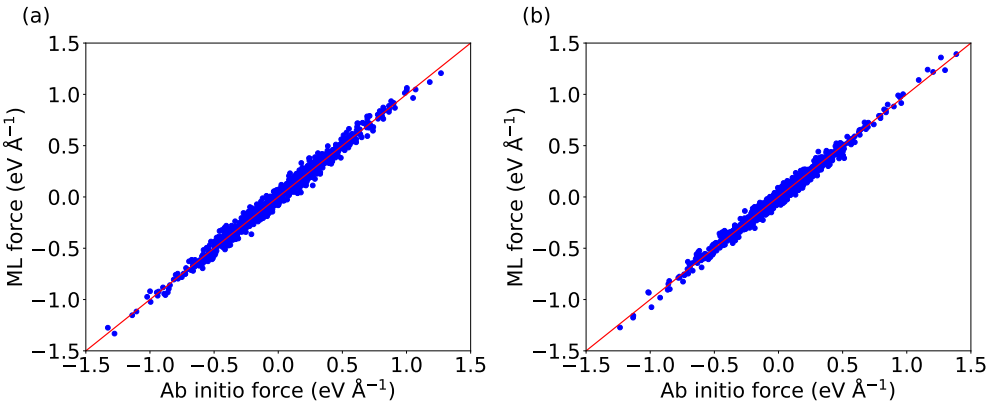

FIG. S1. Comparison of atomic forces calculated in MLFF and DFT calculations for (a) I and (b) Ag in AgI. Root mean square errors (RMSEs) are 45.5 meV Å^{-1} and 40.2 meV Å^{-1} for I and Ag, respectively. We randomly selected five snapshots from AIMD trajectories that serve as a test set.

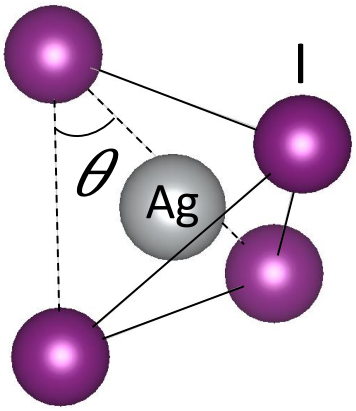

FIG. S2. Structural representation of AgI_4 and the tetrahedral angle, θ.

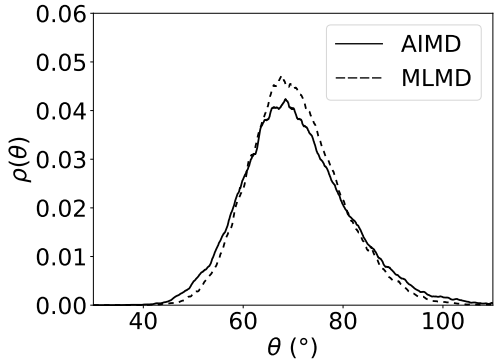

FIG. S3. Probability density, $\rho(\theta)$, as obtained from the 48 tetrahedra of the AgI supercell. The MLFF was trained first at 500 K and then retrained at 800 K to sample an extended phase space (see Main Text for discussion).

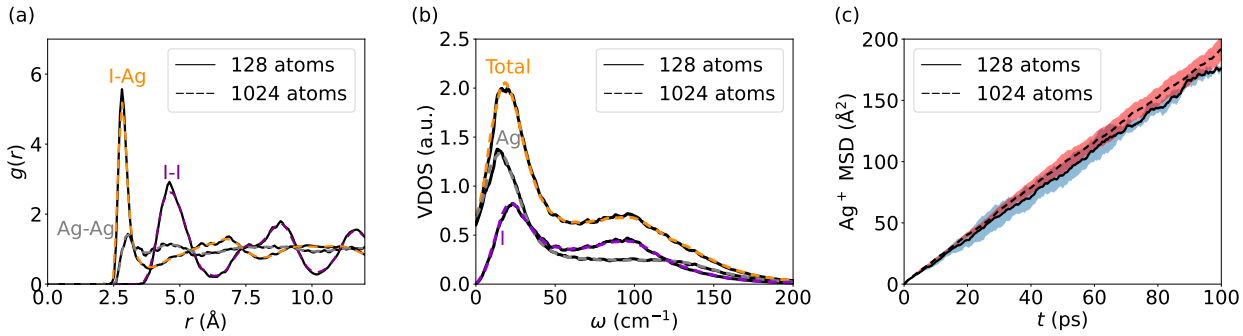

FIG. S4. Size transferability of machine-learning force fields in AgI from a 128 atom to a 1024 atom supercell: (a) Pair correlation function $g(r)$ as function of the distance r, (b) vibrational density of states, VDOS, as function of the frequency ω, and the Ag^+ mean square displacement, MSD, as function of time t.

B. LGPS

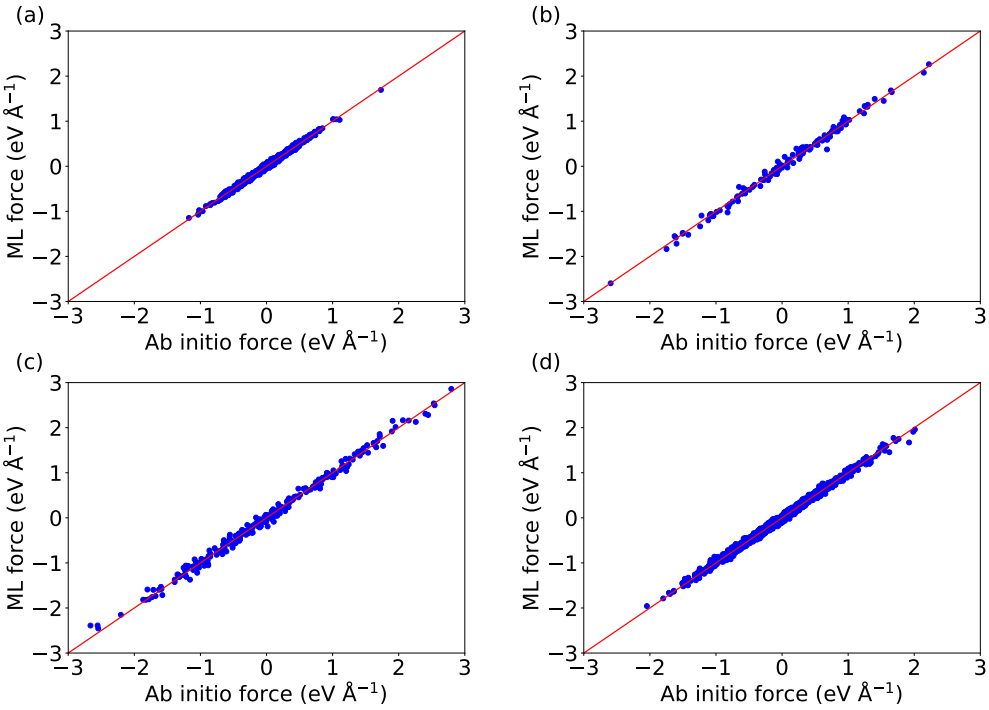

FIG. S5. Comparison of atomic forces calculated in MLFF and DFT calculations for (a) Li, (b) Ge, (c) P, and (d) S atoms in LGPS. RMSEs are 27.7 meV Å^{-1} (Li), 75.7 meV Å^{-1} (Ge), 85.0 meV Å^{-1} (P), and 45.4 meV Å^{-1} (S). The test set is constructed from five random snapshots from AIMD trajectories.

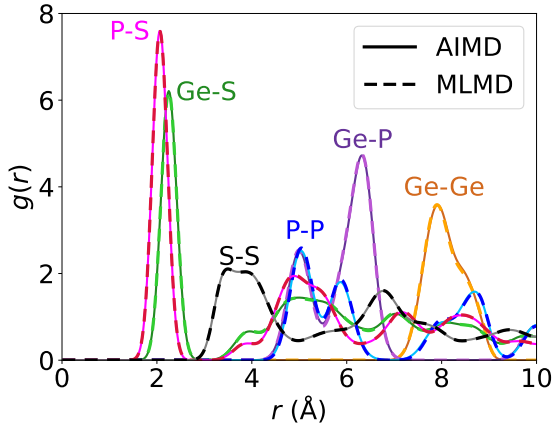

FIG. S6. Pair-correlation function, $g(r)$, for non-Li bonding ions in LGPS calculated from AIMD and MLMD.

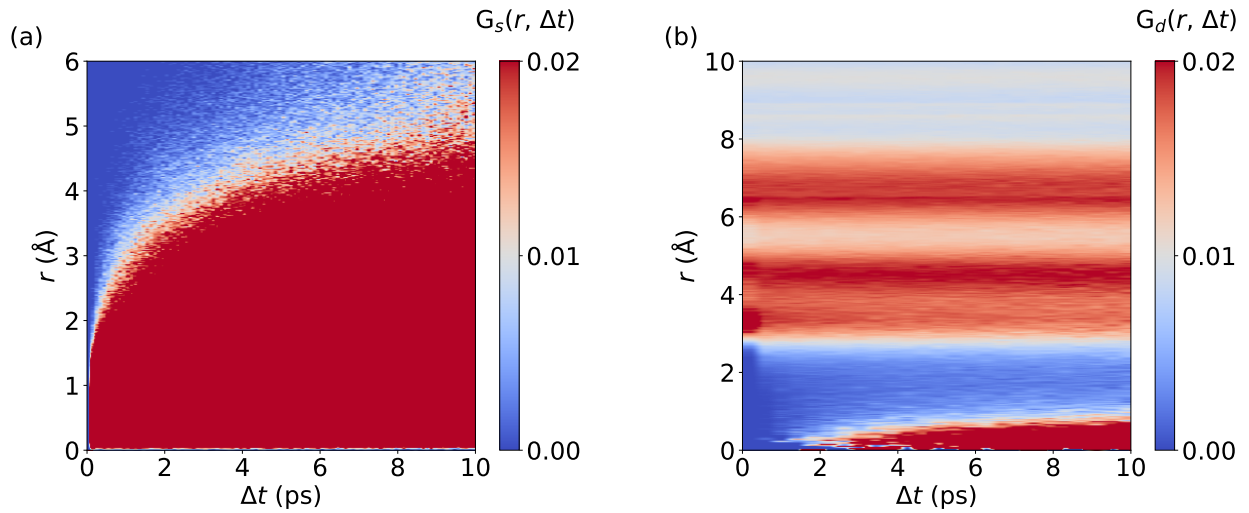

FIG. S7. Van Hove correlation function: (a) self part, $G_s(r, \Delta t)$, and (b) distinctive part, $G_d(r, \Delta t)$, of Li ions in LGPS calculated from AIMD.

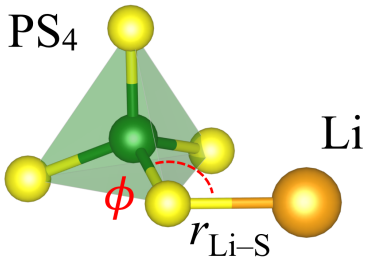

FIG. S8. Structural representation of a PS_4 tetrahedron bonded to a neighboring Li atom. Dynamic structural correlations are quantified by the dihedral angle, ϕ, between P, S, and Li, and the Li–S distance, $r_{Li–S}$.

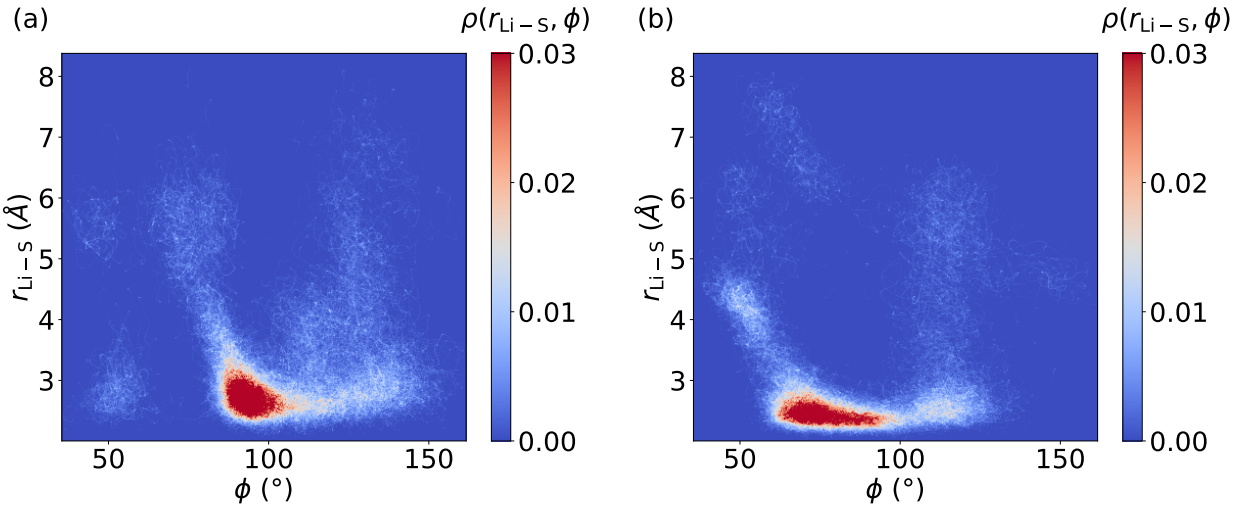

FIG. S9. 2D probability density, $\rho(r_{\mathrm{Li-S}}, \phi)$ of (a) P–S–Li and (b) Ge–S–Li angles, ϕ, and Li–S distances $r_{\mathrm{Li-S}}$, extracted from 32 PS_4 and 32 GeS_4 tetrahedra and their neighboring Li^+ ions. All results were computed with AIMD.

C. Na_3SbS_4

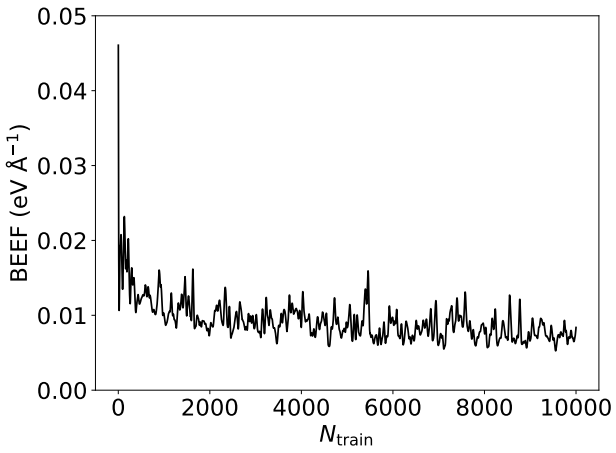

FIG. S10. Bayesian estimated error of the force (BEEF) during the training of the machine-learned force field for the pristine Na_3SbS_4.

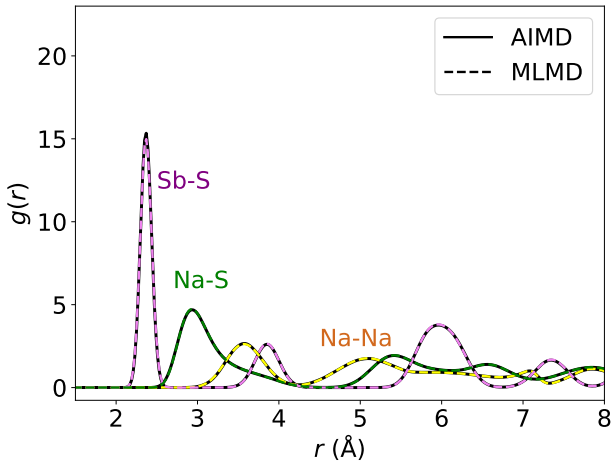

FIG. S11. Pair correlation function, $g(r)$, of pristine Na_3SbS_4 from AIMD and MLMD.

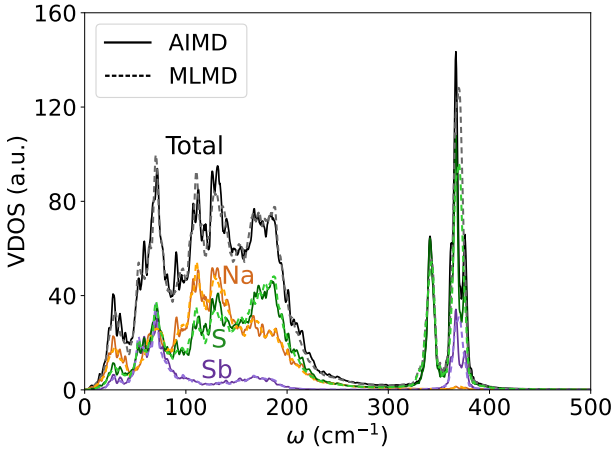

FIG. S12. Vibrational density of states, VDOS, of pristine Na_3SbS_4 from AIMD and MLMD.

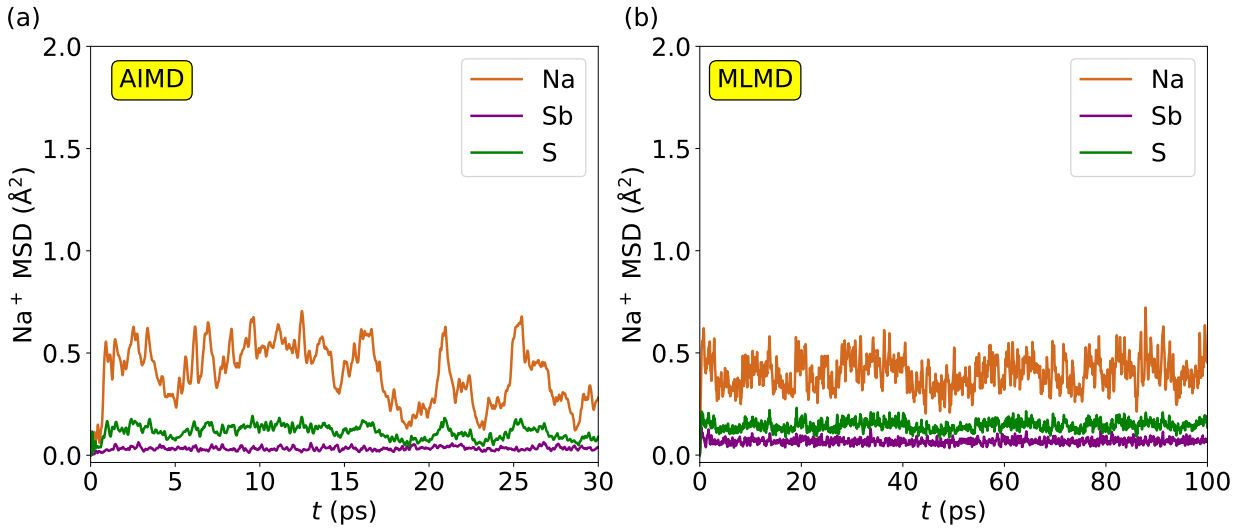

FIG. S13. Mean square displacement of all elements in pristine Na_3SbS_4 from (a) AIMD and (b) MLMD.

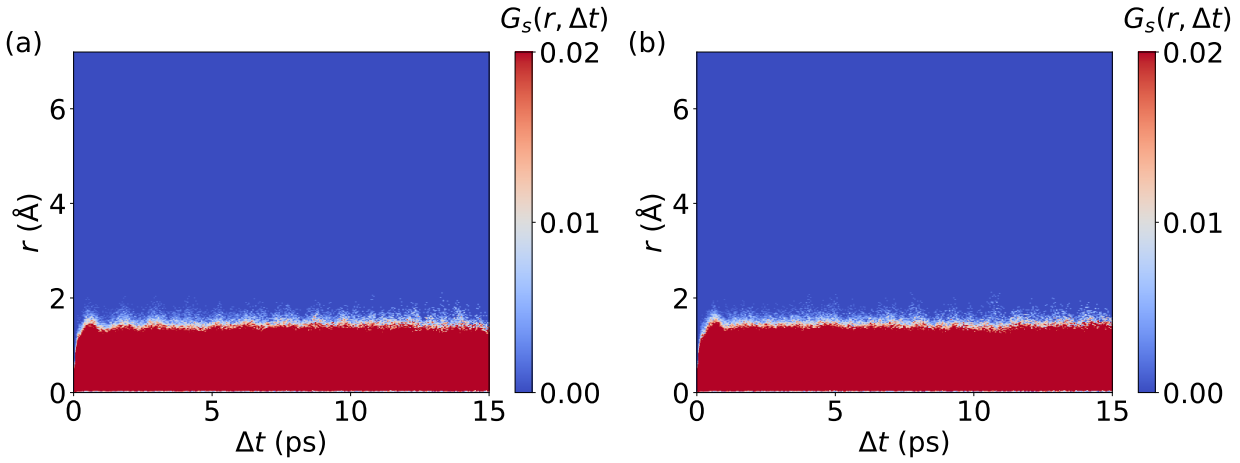

FIG. S14. Self part of the van Hove correlation, $G_s(r, \Delta t)$, for Na^+ ions in pristine Na_3SbS_4 from AIMD (left) and MLMD (right).

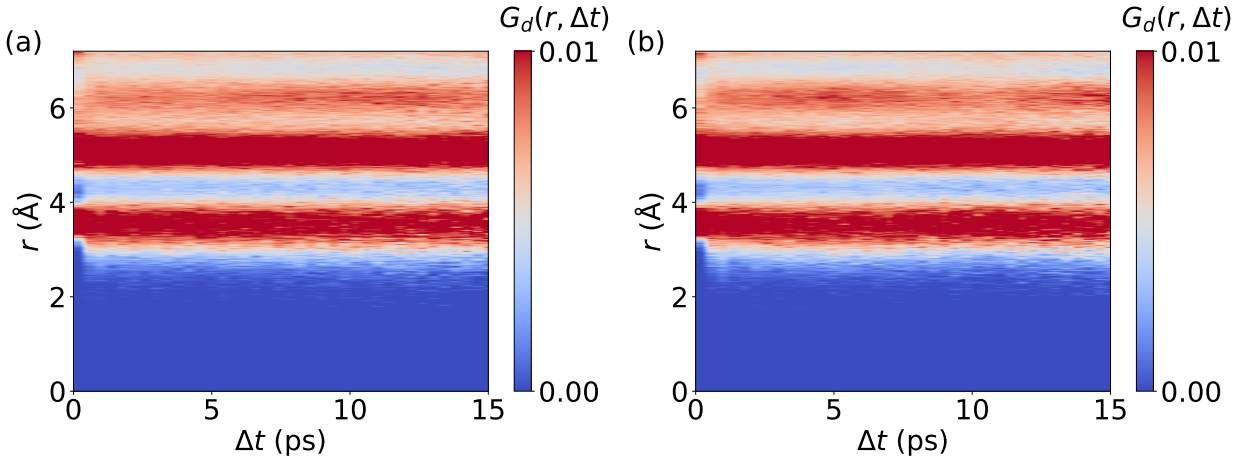

FIG. S15. Distinctive part of the van Hove correlation, $G_d(r, \Delta t)$, for Na^+ ions in pristine Na_3SbS_4 from AIMD (left) and MLMD (right).

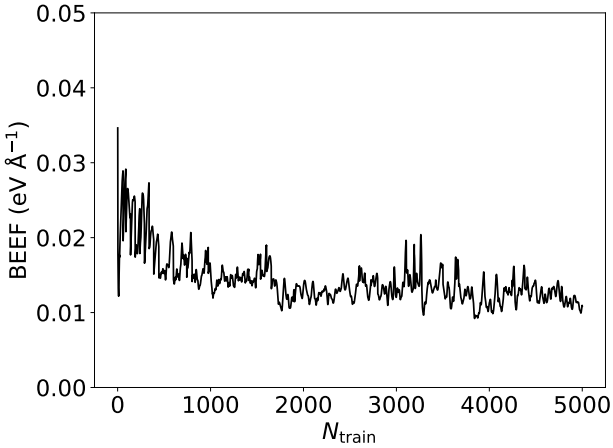

FIG. S16. Bayesian estimated error of the force (BEEF) during the training of the machine-learned force field for the W-doped $Na_{2.94}Sb_{0.94}W_{0.06}S_4$.

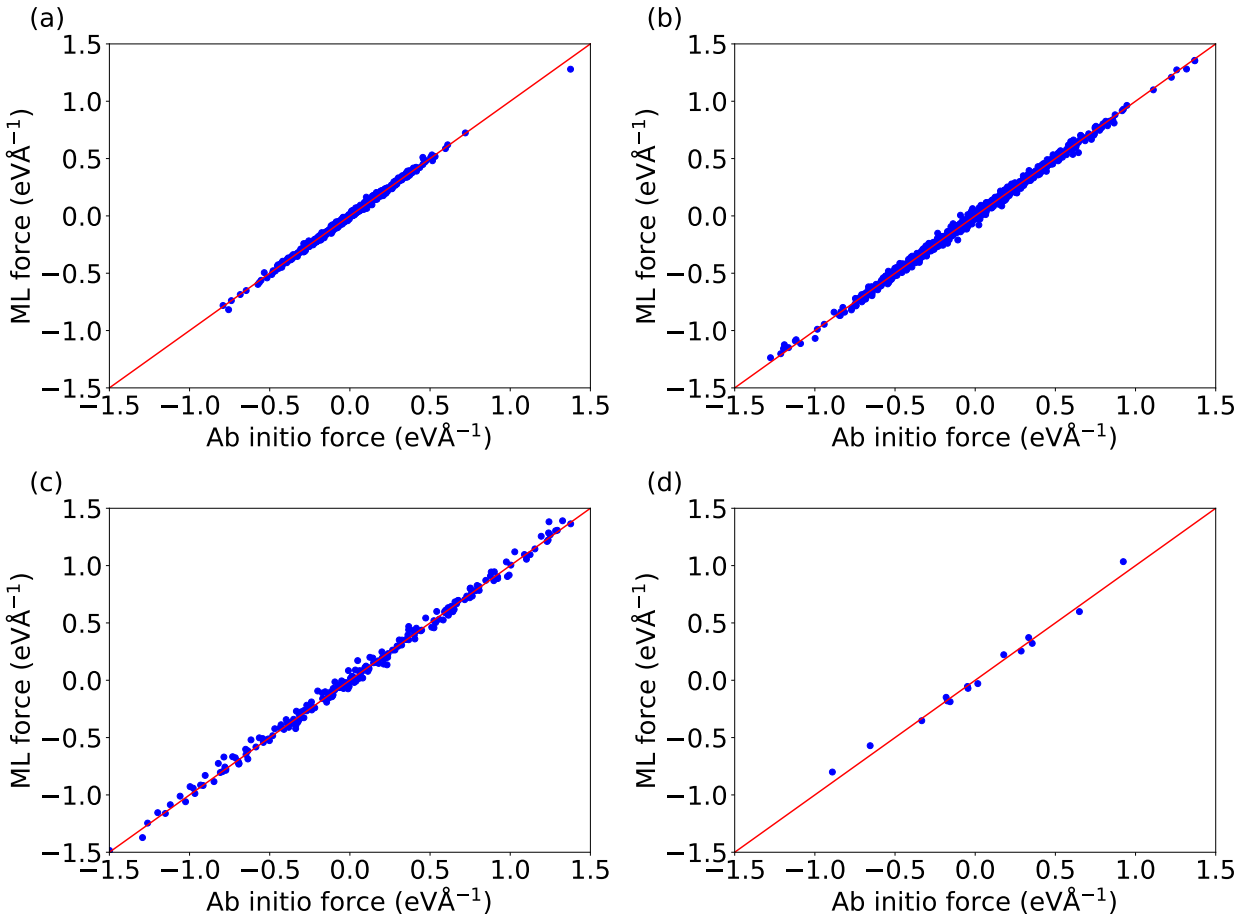

FIG. S17. Comparison of atomic forces calculated in MLFF and DFT calculations for (a) Na, (b) Sb, (c) S, and (d) W atoms in $Na_{2.94}Sb_{0.94}W_{0.06}S_4$. RMSEs are 14.0 meV Å^{-1} (Na), 41.4 meV Å^{-1} (Sb), 24.3 meV Å^{-1} (S), and 52.3 meV Å^{-1} (W). We randomly selected five snapshots from AIMD trajectories that serve as a test set.

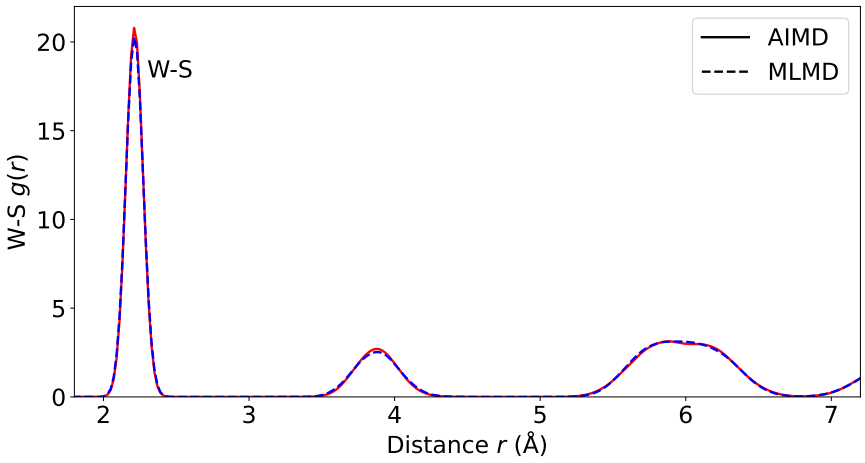

FIG. S18. Pair correlation function, $g(r)$, for W–S in the W-doped $Na_{2.94}Sb_{0.94}W_{0.06}S_4$ calculated from AIMD and MLMD.

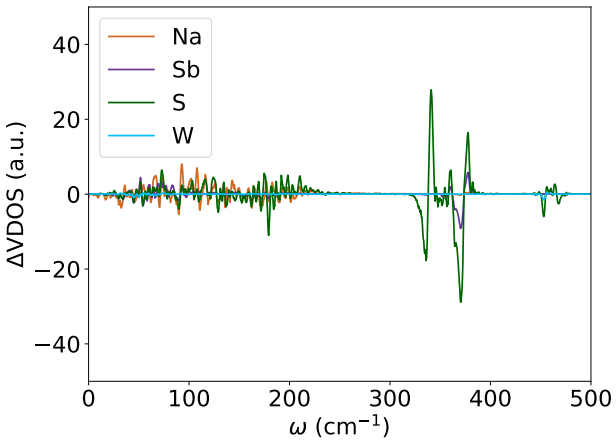

FIG. S19. Difference in the VDOS, ΔVDOS, between MLMD and AIMD for doped $Na_{2.94}Sb_{0.94}W_{0.06}S_4$.

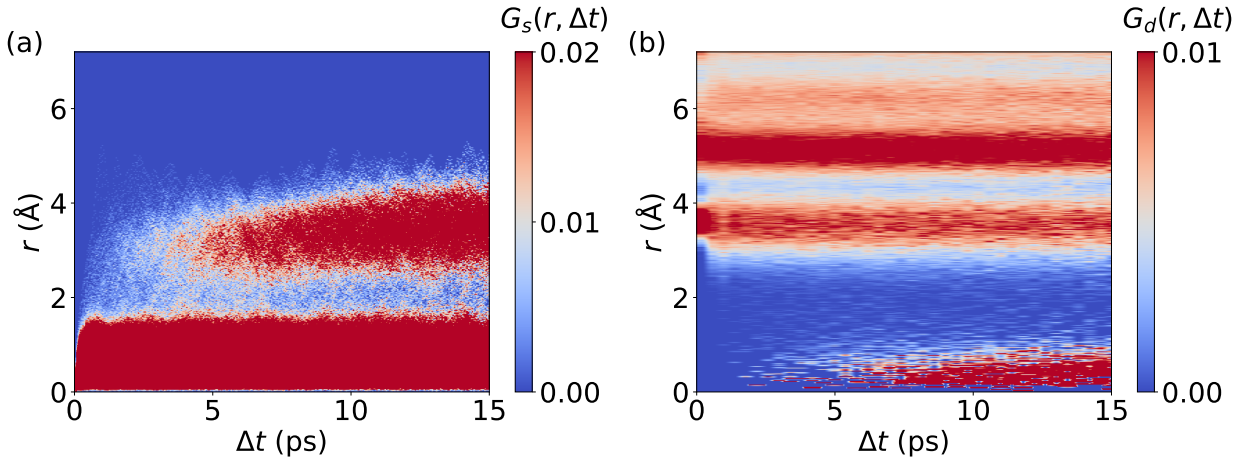

FIG. S20. Self (left) and distinctive part (right) of the van Hove correlation, $G_s(r, \Delta t)$ and $G_d(r, \Delta t)$, respectively, for Na^+ ions in W-doped $Na_{2.94}Sb_{0.94}W_{0.06}S_4$ calculated from AIMD.

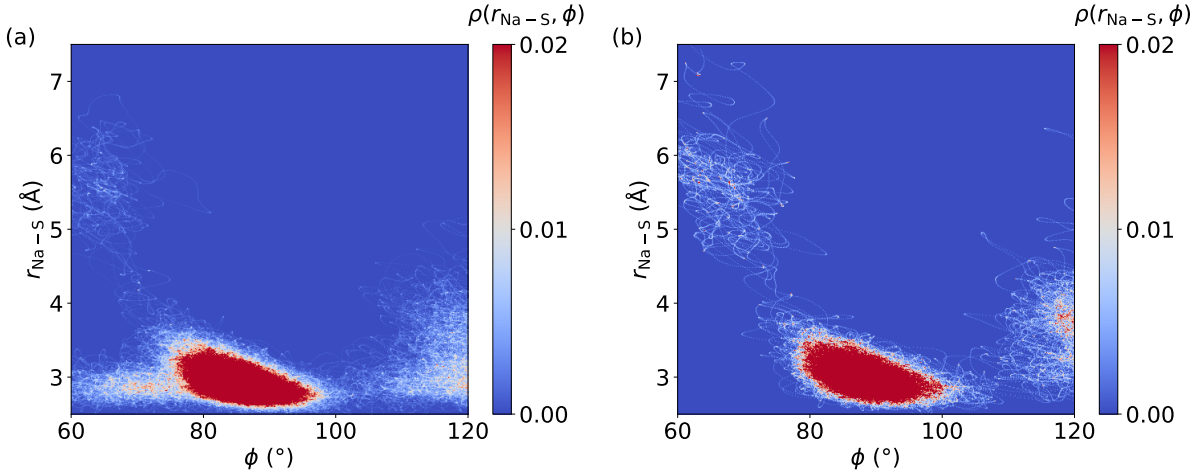

FIG. S21. 2D probability density, $\rho(r_{Na–S}, \phi)$, quantifying correlations between the Sb–S–Na (panel a) and W–S–Na angles (panel b), ϕ, respectively, with Na–S bond lengths, $r_{Na–S}$, as calculated from AIMD simulations.

[1] R. Jinnouchi, J. Lahnsteiner, F. Karsai, G. Kresse, and M. Bokdam, Phase transitions of hybrid perovskites simulated by machine-learning force fields trained on the fly with bayesian inference, Phys. Rev. Lett. **122**, 225701 (2019).

[2] R. Jinnouchi, F. Karsai, C. Verdi, R. Asahi, and G. Kresse, Descriptors representing two-and three-body atomic distributions and their effects on the accuracy of machine-learned interatomic potentials, J. Chem. Phys. **152** (2020).

[3] R. Jinnouchi, F. Karsai, and G. Kresse, On-the-fly machine learning force field generation: Application to melting points, Phys. Rev. B **100**, 014105 (2019).

[4] T. M. Brenner, C. Gehrmann, R. Korobko, T. Livneh, D. A. Egger, and O. Yaffe, Anharmonic host-lattice dynamics enable fast ion conduction in superionic AgI, Phys. Rev. Mater. **4**, 115402 (2020).

[5] S. P. Ong, Y. Mo, W. D. Richards, L. Miara, H. S. Lee, and G. Ceder, Phase stability, electrochemical stability and ionic conductivity of the $Li_{10\pm1}MP_2X_{12}$ (M = Ge, Si, Sn, Al or P, and X = O, S or Se) family of superionic conductors, Energy Environ. Sci. **6**, 148 (2013).

[6] S. Hori, M. Kato, K. Suzuki, M. Hirayama, Y. Kato, and R. Kanno, Phase diagram of the Li_4GeS_4–Li_3PS_4 quasi-binary system containing the superionic conductor $Li_{10}GeP_2S_{12}$, J. Am. Ceram. Soc. **98**, 3352 (2015).

[7] G. Winter and R. Gómez-Bombarelli, Simulations with machine learning potentials identify the ion conduction mechanism mediating non-arrhenius behavior in lgps, J. Phys.: Energy **5**, 024004 (2023).

[8] L. Zhang, D. Zhang, K. Yang, X. Yan, L. Wang, J. Mi, B. Xu, and Y. Li, Vacancy-contained tetragonal Na_3SbS_4 superionic conductor, Adv. Sci. **3**, 1600089 (2016).

[9] S.-I. Nishimura, A. Hayashi, A. Sakuda, and A. Yamada, Vacancy-stabilized superionic state in $Na_{3-x}Sb_{1-x}W_xS_4$, ACS Appl. Energy Mater. **5**, 14053 (2022).