Identification of Potential Solid-State Li-Ion Conductors with Semi-Supervised Learning

Forrest A. L. Laskowski‡ [a], Daniel B. McHaffie‡ [b], Kimberly A. See* [a]

[a]Division of Chemistry and Chemical Engineering

[b]Department of Applied Physics and Materials Science

California Institute of Technology, Pasadena, California 91125, United States

*Corresponding Author: ksee@caltech.edu

‡Co-authors with equal contribution

Abstract. Despite ongoing efforts to identify high-performance electrolytes for solid-state Li-ion batteries, thousands of prospective Li-containing structures remain unexplored. Here, we employ a semi-supervised learning approach to expedite identification of ionic conductors. We screen 180 unique descriptor representations and use agglomerative clustering to cluster ~26,000 Li-containing structures. The clusters are then labeled with experimental ionic conductivity data to assess the fitness of the descriptors. By inspecting clusters containing the highest conductivity labels, we identify 212 promising structures that are further screened using bond valence site energy and nudged elastic band calculations. Li_3BS_3 is identified as a potential high-conductivity material and selected for experimental characterization. With sufficient defect engineering, we show that Li_3BS_3 is a superionic conductor with room temperature ionic conductivity greater than 1 mS cm^{-1}. While the semi-supervised method shows promise for identification of superionic conductors, the results illustrate a continued need for descriptors that explicitly encode for defects.

Identifying new materials that could improve solid-state ion battery prospects is an ongoing challenge. The search for an ideal solid-state Li electrolyte is a prime example. Research has focused on eight classes of materials: LISICON-type structures, argyrodites, garnets, NASICON-type structures, Li-nitrides, Li-hydrides, perovskites, and Li-halides[1]. However, only three compounds with near-liquid-electrolyte conductivity (~10^{-2} S cm^{-1}) have been discovered: $Li_{10}GeP_2S_{12}$ (LGPS)[2], Li_6PS_5Br argyrodite[3], and $Li_7P_3S_{11}$ ceramic-glass[1,4]. Although promising discoveries, all three high-conductivity structures are unstable against the Li anode[5–10]. While investigations to limit instability are ongoing[11,12], identification of stable superionic structures is desirable. High-performing structures that enable new battery chemistries may exist outside of the eight classes. However, exploration under the traditional Edisonian approach prioritizes small perturbations to well-known variable spaces.

Machine learning (ML) is a promising tool for expediting the discovery of useful solid-state materials. By describing prospective materials with physically meaningful descriptors, ML models can identify high-dimensional patterns in large datasets that are not readily apparent[13–19]. Ongoing descriptor engineering[20–25] has enabled discovery of battery components[26,27], electrocatalysts[14,28], photovoltaic components[15,29], piezoelectrics[30], new metallic glasses[13] and new alloys[31]. However, application of ML for discovery of SSEs and other emerging technologies can be challenging. Supervised ML approaches require empirical data for use as “labels,” but relatively few SSEs have been experimentally characterized compared to the ~26,000 known Li-containing structures[18,32–34]. Characterized materials often exhibit ill-defined properties owing to the variety of synthetic approaches and non-standardized testing methods[35]. Well-performing materials often contain charge-carrying defects that are not explicitly characterized or reported[36]. Negative examples, i.e. materials with undesirable properties, are useful for ML models but are seldom reported.

Semi-supervised ML can guide synthetic prioritization of SSEs by overcoming the issues associated with label scarcity. Supervised ML requires labels because it infers correlation

functions by mapping the input descriptors to the labels[37]. Semi-supervised ML prioritizes comparison of descriptors to identify relationships between the descriptors in a dataset[34,37]. The input compositions are clustered (or grouped) by comparison of descriptors using a similarity metric. The clustering process does not consider labels, and thus circumvents the need for abundant labels. The resultant clusters can be labeled *ex post facto* to examine correlation between the descriptor and a physical property of interest. For semi-supervised ML, ideal descriptors result in a set of clusters where each cluster has similar labels and thus the label variance is minimized. Promising synthetic targets may then be identified by their membership in clusters that contain desirable labels.

Semi-supervised ML can help identify descriptors that are correlated to physical properties of interest. Descriptors are representations of the input materials that encode the chemistry, composition, structure, and/or other system properties. An ideal descriptor should be a unique representation, a continuous function of the structure, exhibit rotational/translational invariance, and be readily comparable across all structures in the dataset[23–25]. Recently, Zhang et al. demonstrated that a modified X-Ray diffraction (mXRD) descriptor lead to favorable clustering for Li SSEs[32]. By labeling the resultant clusters with experimental room-temperature Li-ion conductivities, they identified 16 prospective fast-ion conductors. Despite the promising results, no other descriptors have been explored using clustering approaches. Descriptor screening is desirable for identifying more predictive semi-supervised learning models. Identification of useful descriptors can improve chemical intuition by revealing which system properties are most correlated with a property of interest. Descriptor transformations for inorganic structures have been curated in a variety of software packages, including: Matminer[23], Dscribe[24], SchNet[38], and Aenet[39].

Herein, we employ hierarchical agglomerative clustering to screen many descriptors, without assuming correlation to ionic conductivity. The performance of 20 descriptors is assessed for semi-supervised identification of Li SSEs. Each descriptor is paired with 9 structural

simplification strategies, yielding a total of 180 unique representations per input structure. The approach is applied to a dataset of ~26,000 Li-containing phases, encompassing all Li-containing structures contained in the Inorganic Crystal Structure Database (ICSD - v.4.4.0) and the Materials Project (MP - v.2020.09.08) database (Fig. 1). A set of 220 experimental room temperature ionic conductivities ($\sigma_{25°C}$) are aggregated from literature reports and used as labels. Descriptors that encode the spatial environment are found to be most correlated with the ionic conductivity labels.

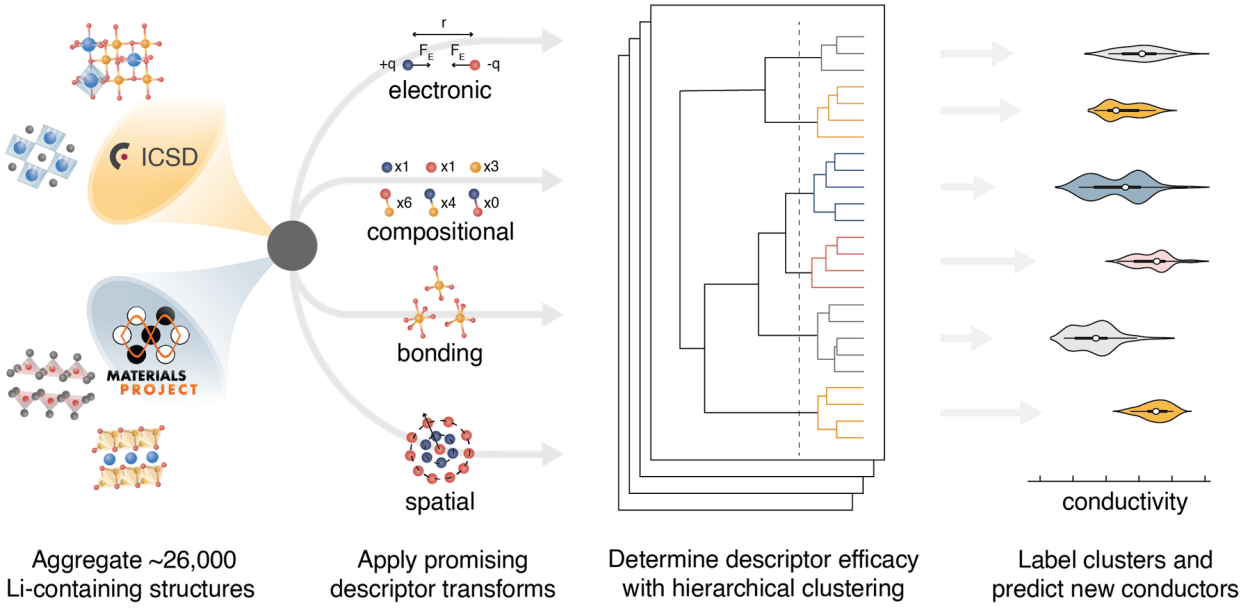

Figure 1. Schematic of the semi-supervised machine learning approach. *Li-containing structures are aggregated from the ICSD and MP database. Each input structure is simplified and transformed to yield a unique descriptor representation. The descriptor representations are clustered with hierarchical agglomerative clustering. Each cluster is then labeled with experimental $\sigma_{25°C}$ data and the intracluster conductivity variance is calculated. Comparison of the composite intracluster conductivity variance (intracluster conductivity variance summed across all clusters) enables identification of descriptors that are well correlated with ionic conductivity.*

Using the descriptors, the semi-supervised approach can identify potential fast solid-state Li-ion conductors. By selecting structures in clusters containing high conductivity labels, the

~26,000 input structures are down selected to just 212 promising structures. Practical considerations, a semi-empirical bond valence site energy (BVSE) method,[40] and the Nudged Elastic Band (NEB) method are employed to rank the structures. From the ten highest ranking structures, Li_3BS_3 is selected for model validation. Synthesis of pure Li_3BS_3 yields a poor conductor. However, by employing defect engineering strategies we demonstrate that Li_3BS_3 is a superionic conductor with an ionic conductivity greater than 10^{-3} S cm^{-1}.

Main Text

Screening simplification-descriptor combinations

A set of 20 descriptors is selected for screening the semi-supervised learning approach (**Table 1**). The descriptors generally encode four types of information: the spatial environment, the chemical bonding environment, the electronic environment, and composition. All descriptors are implemented in Python using the Matminer[23] or Dscribe[24] libraries. The code is published to a github repository and is available for download (https://github.com/FALL-ML/materials-discovery). Zhang et al. illustrated that structure simplification prior to learning can produce lower variance outcomes[32]. Their mXRD descriptor was found to work best with removal of all cations, all the anions replaced by a single representative anion, and the structure volume scaled to 40 $Å^3$ per anion. Inspired by the previous success in using structure simplification, we screen eight structure simplifications in addition to the unperturbed structure. For simplifications the following categories of atoms are replaced with a representative specie: (1) **C**ations are represented as Al, (2) **A**nions are represented as S, (3) **M**obile ions are represented as Li, and (4) **N**eutral atoms are represented as Mg. Categories of atom are removed as to yield the four simplifications: CAMN (all atoms retained), CAN (mobile ions removed), AM (cations and neutral atoms removed), and A (only anions retained). Four additional simplifications are formed by scaling each lattice volume to 40 $Å^3$ per anion: CAMN-40, CAN-40, AM-40, and A-40.

Table 1. The descriptors used for agglomerative clustering. *Descriptor vectors are attained by simplifying the input structures and then applying the descriptor transformation. In total, 180 unique descriptor vectors are screened for each structure.*

Descriptor	Descriptor Description	Refs
Bond Fraction	"Bag of bonds" approach described in Hansen et. al. wherein pairwise nuclear charges and distances are encoded.	[41]
Band Center	Estimation of band center from constituent atoms' electronegativity values described by Butler et al.	[42]
Crystal Structure Analysis by Voronoi Decomposition (CAVD)	Calculation of the largest sphere that can pass through the lattice-sans-mobile-ion using Voronoi decomposition of structures.	[43]
Chemical Ordering	Warren-Cowley-like ordering method to determine how different the structure's ordering is from random.	[44]
Density Features	Calculates density, volume per atom, and the packing fraction.	[45]
Electronegativity Difference	Composition weighted calculation of the electronegativity difference between cations and anions.	[46]
Ewald Energy	Sum of coulomb interaction energies across all lattice sites described by Ewald et al.	[47]
Global Instability Index	Averaged square root of the sum of squared differences over the bond valence sums.	
Jarvis	Diverse set of descriptors from the Jarvis-ML library.	[48]
Maximum Packing Efficiency	A measure of the void space within the unit cell.	[44]
Meredig	Composite descriptor from Meredig et al.	[49]
Modified XRD (mXRD)	Powder diffraction pattern calculated using Bragg's law.	[45]
Orbital Field Matrix	Descriptor that encodes the distribution of valence shell electrons for each input structure.	[50]
Oxidation States	Concentration weighted oxidation state statistics.	[46]
Radial Distribution Function	Radial distribution function for each structure.	[45]
Sine Coulomb Matrix	Coulomb matrix for periodic lattices, developed by Faber et al.	[51,52]
Smooth Overlap of Atomic Positions (SOAP)	Geometric encoder that is rotationally/transitionally invariant through use of spherical harmonics and radial basis functions. Atoms are represented by a smeared gaussian.	[24]
Structural Complexity	The Shannon information entropy for a given structure.	[53]
Structure Variance	Bond length and atomic volume variance for each structure.	[44]
Valence Orbital	Structure averaged number of valence electrons in each orbital.	[46,54]
Control	A control descriptor is not explicitly used. Instead, clustering outcomes are randomly assigned. For composite intracluster variance calculations, 100 control iterations are averaged.	

Agglomerative clustering is performed on all Li-containing structures from the ICSD and MP repositories. Agglomerative clustering is a "bottom-up" approach to clustering where each structure starts in its own cluster of one. Clusters are merged according to Ward's Minimum Variance criterion in Euclidean space, which minimizes the global descriptor variance[55]:

$$W = \sum_{k=1}^{n_C} \sum_{i \in C_k} \left[d_i - \bar{d}_k \right]^2$$

where n_C is the number of clusters in a set, C_k is cluster k, d_i is a descriptor representation for structure i, and $\bar{d}_k$ is the average descriptor representation in cluster k. Each cluster merger results in the lowest variance set of clusters, relative to all other possible mergers. Other common linkage criteria (average, complete, and single linkages) and metrics (l1, l2, manhatten, cosine) were screened but are found to result in clustering outcomes with larger *W*. For each simplification-descriptor combination, all clustering sets from 2-300 are computed. Physically relevant labels are applied to the resultant clustering sets to assess how well each simplification-descriptor combination performs. To compare between the 180 different simplification-descriptions combinations, the data is labeled with 155 experimental room temperature conductivity (σ_{RT}) values aggregated from the literature reports (see supplementary info - sections I - IV). A secondary label set is also screened, comprised of 6845 activation energies (E_a) computationally generated using a bond valence energy approach (see supplementary info – section V).

An ideal simplification-descriptor combination results in clustering where each cluster contains labels with similar σ_{RT} values. Ward's minimum variance method is applied to the conductivity labels as a measure of clustering efficacy:[32]

$$W_\sigma = \sum_{k=1}^{n_C} \sum_{i \in C_k} \left[\log(\sigma_{RT})_i - \overline{\log(\sigma_{RT})_k} \right]^2$$

where n_C is the number of clusters in a set, C_k is cluster k, and $\overline{\log(\sigma_{RT})_k}$ denotes the mean for all labels in cluster k. Since clusters containing only one label effectively drop out of the W_σ calculation, a frozen-state strategy is employed when needed (see supplemental information – section IV). Each descriptor's W_σ results are shown in Figure 2 for the first 50 clustering outcomes

(i.e. the W_σ is shown for each set of 2, 3, …, 49, and 50 clusters). For simplicity, only the best-performing simplification-descriptor combination is shown for each descriptor.

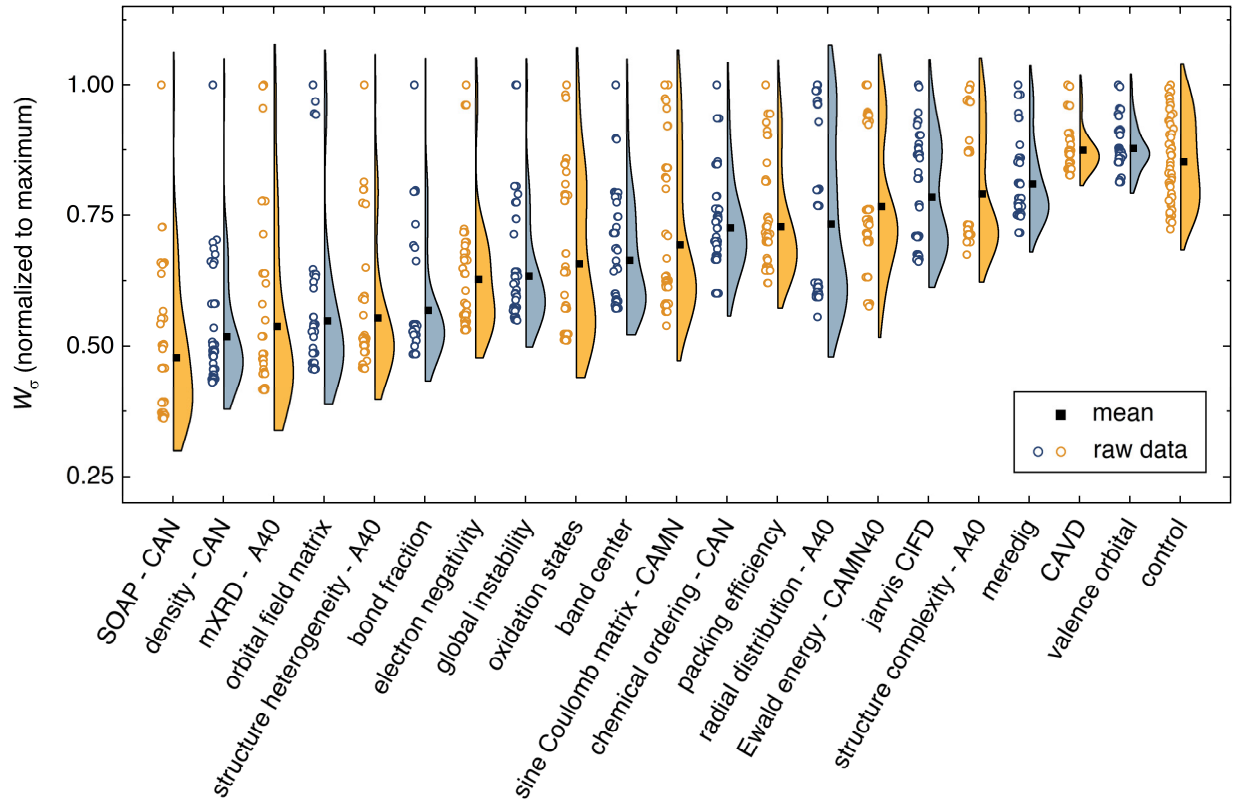

Figure 2. The composite intracluster conductivity variance (W_σ) for the first 50 clusters generated using each descriptor. *Half-violin plots show the raw W_σ score for each cluster as symbols next to the violin distribution. Simplification-descriptor combinations are sorted in order of ascending mean. The control is a random assignment of clusters, with W_σ values averaged over 100 randomly assigned sets. The smooth overlap of atomic positions (SOAP) descriptor outperforms all other descriptors. Although not shown here, SOAP continues to outperform for all depths of clustering through 300.*

Using $\sigma_{25°C}$ labels, the best semi-supervised ML performance is attained when using the SOAP descriptor. SOAP is a spatial descriptor that employs smeared gaussians to represent atomic positions for each crystal structure[24]. Predictions using the SOAP descriptor have exhibited similar performance to state-of-the-art graph neural networks (GCNs) on a variety of

materials science datasets[56]. Optimization of SOAP hyper-parameters (radial cutoff, number of radial basis functions, degree of spherical harmonics) is explored in section VI of the supplemental information. SOAP is found to perform best when combined with the CAN structure simplification. That is, the simplification where the mobile Li atoms are removed, and the remaining atoms are simplified into three representative species: cations, anions, and neutral atoms. SOAP outperforms all other descriptors for all depths of clustering. The SOAP descriptor can be modestly improved (2-3% decrease in W_σ) by mixing with other descriptors to make a 2nd order SOAP descriptor (see supplemental information – section VI).

Semi-supervised identification of prospective Li-ion conductors

Agglomerative clustering with the 2nd order SOAP descriptor is used to identify prospective ionic conductors. W_σ minimization is prioritized over W_{Ea} minimization because E_a alone is not necessarily a good predictor of conductivity; $\sigma_{25°C}$ may be affected by properties including the ionic carrier concentration, hopping attempt frequency, and the presence of concerted migration modes[57]. The agglomerative dendrogram for the 2nd order SOAP is shown in Figure 3, with the label densities plotted below. The agglomerative dendrogram is depicted to 241 clusters, after which the W_σ does not appreciably decrease. To facilitate discussion, an arbitrary cutoff is placed to yield 9 large clusters. The results show that although cluster #2 contains only 15% of the input structures, it accounts for over half of the high-conductivity ($\sigma_{25°C}$ >10^{-5} S cm^{-1}) labels. By the 17th clustering step, the densest cluster accounts for 6.2% of the structures while containing over half (52%) of the high-conductivity labels.

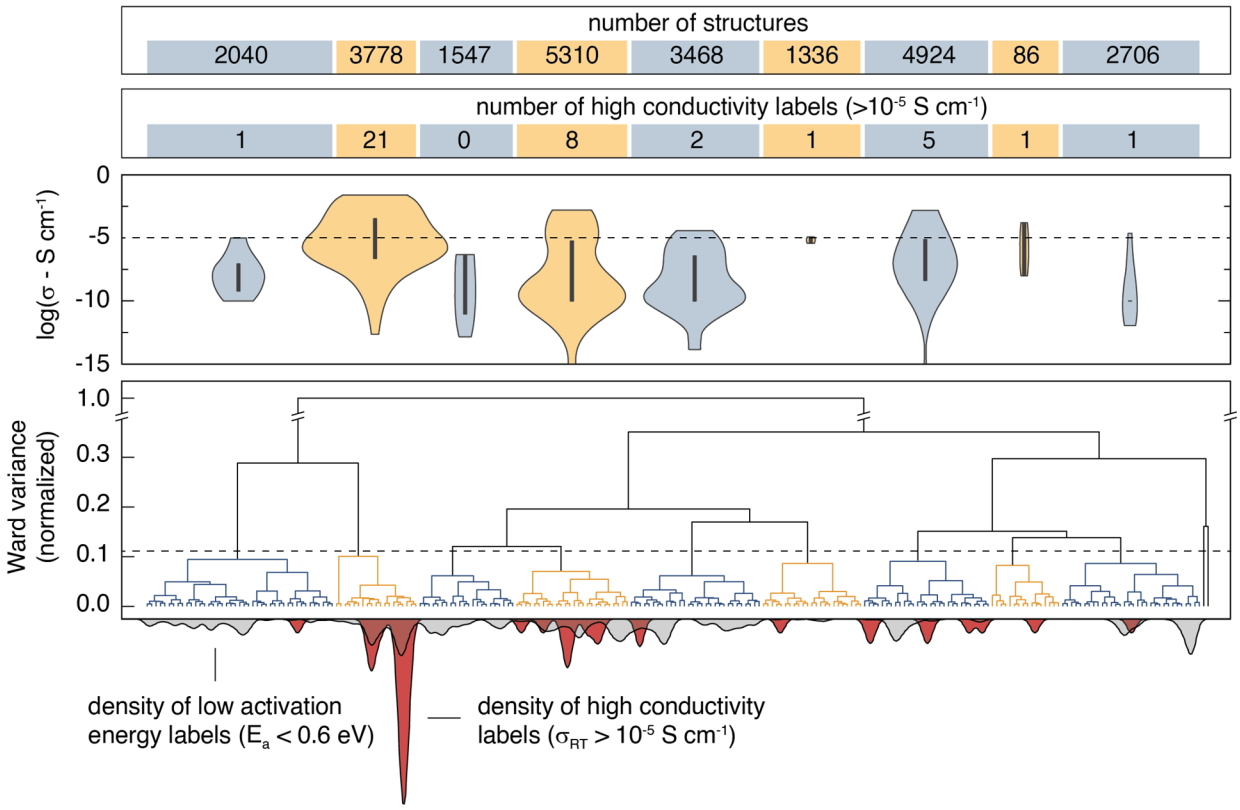

Figure 3. Agglomerative clustering dendrogram for the 2nd-order SOAP descriptor. *The hierarchical clustering representation is shown for the first 241 clusters. An arbitrary variance cutoff is placed such that 9 large clusters are produced to facilitate analysis. The violin plots show the $\sigma_{25°C}$ distribution for the labels within the 9 large clusters. Three outlier clusters are grouped into two additional clusters and are hereafter ignored. The density (per 241 clusters) of low E_a (<0.6 eV) and high conductivity ($\sigma_{25°C}$ >10^{-5} S cm^{-1}) labels is shown underneath the agglomerative dendrogram. The results illustrate that agglomerative clustering on the 2nd-order SOAP descriptor results in favorable aggregation of most high-conductivity labels.*

Candidates for next-generation SSEs can be identified by evaluating clusters that either contain or are near high conductivity labels. Clusters #2, #4, and #7 are promising because they account for 85% of the high $\sigma_{25°C}$ labels. However, targeting these clusters would necessitate screening thousands of structures. Instead, we search from the 241st cluster depth, targeting all clusters that contain or are directly adjacent (i.e. the nearest cluster in the Euclidean feature space) to high $\sigma_{25°C}$ labels. The promising structures are further screened using calculated stability

(E_{hull}) and band gap (E_g) properties from the Materials Project, and the BVSE E_a values. We select the structures that have **(1)** an E_{hull} of 70 meV or lower,[58] **(2)** an E_g of at least 1 eV, and **(3)** a BVSE-calculated E_a below a conservative 0.6 eV. The approach identifies 212 structures as prospective ionic conductors. Climbing image nudged elastic band (CI-NEB) is employed to calculate the E_a for Li-ion hopping on the ten materials with the lowest BVSE-calculated E_a and an E_{hull} of 0 eV. The CI-NEB functionals and parameters can be found in the supporting information section VII. The top 10 prospective structures are tabulated in Table 2.

Table 2. The top 10 prospective structures from the semi-supervised learning model as ranked by BVSE-calculated E_a. *Structures in or directly adjacent to high-conductivity clusters were identified as promising. The list of promising structures was then further simplified by removing structures with E_{hull} values greater than 0 V and with E_g values less than 1 eV. To rank the remaining structures, the E_a was calculated using a BVSE approach.*

				E_{a_calc} (meV)	
compound	**space group**	**MP_ID**	**ICSD_ID**	**BVSE**	**NEB**
Li_3VS_4	$P\bar{4}3m$ (#215)	mp-760375		160	390
$Na_3Li_3Al_2F_{12}$	$Ia\bar{3}d$ (#230)	mp-6711	9923	230	340
Li_2Te	$Fm\bar{3}m$ (#225)	mp-2530	60434	260	320
$LiAlTe_2$	$I\bar{4}2d$ (#122)	mp-4586	280226	260	310
$LiInTe_2$	$I\bar{4}2d$ (#122)	mp-20782	658016	270	450
Li_6MnS_4	$P4_2/nmc$ (#137)	mp-756490		270	466
$LiGaTe_2$	$I\bar{4}2d$ (#122)	mp-5048	162555	270	340
Li_3BS_3	$Pnma$ (#62)	mp-5614	380104	280	260
KLi_6TaO_6	$R\bar{3}m$ (#166)	mp-9059	73159	300	300[59]
Li_3CuS_2	$Ibam$ (#72)	mp-1177695		310	440

The CI-NEB calculations generally agree with the BVSE calculated E_a values, suggesting favorable activation energies (< 500 meV). Discrepancies between the two values may arise because BVSE does not allow framework ions to relax during Li^+ migration and does not account for repulsive interactions between atoms of the mobile ion species.

Experimental validation of the semi-supervised learning model: Li_3BS_3

From the ten most promising candidates, Li_3BS_3 was selected for synthesis and characterization. Li_3BS_3 stands out because it has been explored experimentally and computationally before. Experimentally, Vinatier et al. previously determined that Li_3BS_3 has a total DC conductivity of $2.5 \cdot 10^{-7}$ S cm^{-1} with an activation energy of 700 meV[60]. The DC measurement was not included in our label set because DC measurements cannot differentiate between ionic and electronic conductivity, so they were categorically discounted from the label set (see supplemental information I for more details on label selection). Although the conductivity and activation energy reported by Vinatier et al. are underwhelming, there are promising theoretical reports. Density functional theory molecular dynamics (DFT-MD) simulations from Sendek et al.[61] suggest that Li_3BS_3 should have a room temperature conductivity between $3.1 \cdot 10^{-6}$ and $9.7 \cdot 10^{-3}$ S cm^{-1}. Our NEB-calculated activation energy for Li_3BS_3 is 260 meV, corroborating a previous NEB result from Bianchini et al.[62]. Additionally, Li_3BS_3 is practically attractive because: (1) Li_3BS_3 contains no redox-active metals, (2) band edge calculations have suggested stability against metallic Li[63], and (3) the synthesis is reported[64]. It is simpler to avoid redox active metals in the SSE as they may be reduced and oxidized at electrode interfaces. However, we note that $Li_{0.5}La_{0.5}TiO_3$ is a widely studied SSE that contains redox active Ti[65,66] so the compounds we report here that contain Mn, V, and Cu should not be categorically discounted. It is important to note that while studying Li_3BS_3 as a candidate Li-ion conductor for model validation, Kimura et al. reported that a so-called "Li_3BS_3 glass" exhibits an ionic conductivity of $3.6 \cdot 10^{-4}$ S/cm^{-1} at 25 °C[67].

Li_3BS_3 is prepared using solid-state synthesis from Li_2S, B, and S precursors. The diffraction and quantitative Rietveld refinement are shown in Figure 4a, suggesting a phase pure material. Electrochemical impedance spectroscopy (EIS) is employed at various temperatures and the resultant conductivity is plotted according to the Arrhenius-like relationship (Fig. 4b):

$$\sigma = \frac{\sigma_0}{T} e^{-\frac{E_a}{k_B T}}$$

where T is the temperature, k_B is the Boltzmann's constant, σ_0 is the conductivity prefactor and E_a is the activation energy. The room temperature ionic conductivity ($\sigma_{25°C}$) is 7.16(± 0.21)·10^{-7} S cm^{-1} and the activation energy is 400 ± 47 meV. The low conductivity and high activation energy may be due to lack of charge-carrying defects in the Li_3BS_3 lattice[68,69]. Although a sufficient carrier concentration is necessary for facile ionic conduction in most materials, the descriptors in the semi-supervised model do not explicitly encode for charge-carrying defects. In the label set, conductivity is likely influenced by the defect concentration but defects are typically not reported. Still, the semi-supervised model may infer a structure's capacity to support conductive defects via correlation with the descriptors. To test the hypothesis, we use two strategies to engineer vacancies: aliovalent substitution and amorphization via extended ball milling. Aliovalent substitution has been shown to improve conductivity in Li-argyrodites, -sulfides, and -garnets by introducing vacancies[68,69]. Similarly, amorphization can introduce defects and vacancies that enable Li^+ hopping[67,69–71].

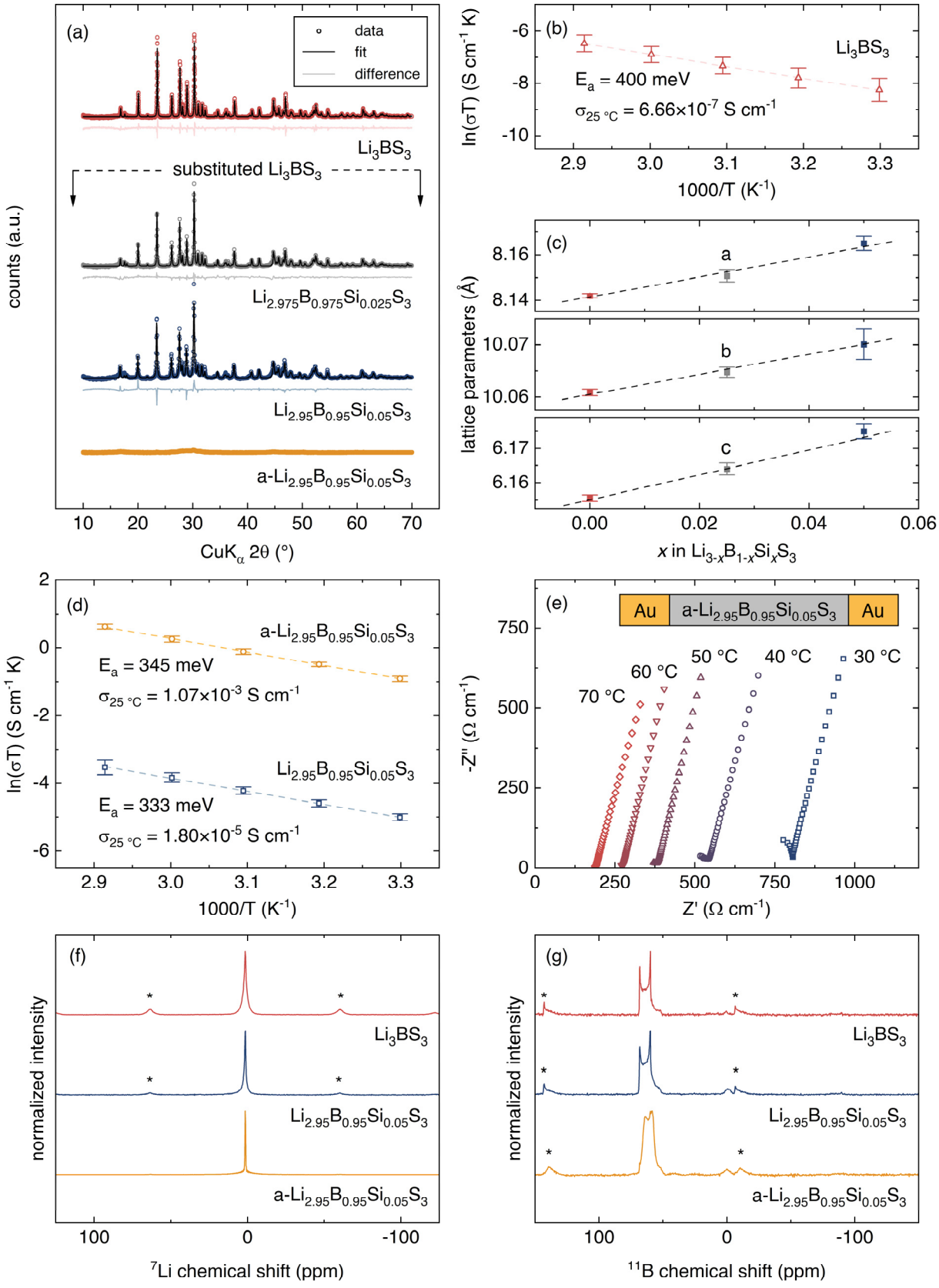
(a)
data
fit
difference
Li_3BS_3
substituted Li_3BS_3
$Li_{2.975}B_{0.975}Si_{0.025}S_3$
$Li_{2.95}B_{0.95}Si_{0.05}S_3$
a-$Li_{2.95}B_{0.95}Si_{0.05}S_3$
counts (a.u.)
CuK_α 2θ (°)
(b)
Li_3BS_3
E_a = 400 meV
$\sigma_{25\,°C}$ = 6.66×10^{-7} S cm^{-1}
ln(σT) (S cm^{-1} K)
1000/T (K^{-1})
(c)
a
b
c
lattice parameters (Å)
x in $Li_{3-x}B_{1-x}Si_xS_3$
(d)
a-$Li_{2.95}B_{0.95}Si_{0.05}S_3$
E_a = 345 meV
$\sigma_{25\,°C}$ = 1.07×10^{-3} S cm^{-1}
$Li_{2.95}B_{0.95}Si_{0.05}S_3$
E_a = 333 meV
$\sigma_{25\,°C}$ = 1.80×10^{-5} S cm^{-1}
ln(σT) (S cm^{-1} K)
1000/T (K^{-1})
(e)
Au
a-$Li_{2.95}B_{0.95}Si_{0.05}S_3$
Au
70 °C
60 °C
50 °C
40 °C
30 °C
-Z'' (Ω cm^{-1})
Z' (Ω cm^{-1})
(f)
Li_3BS_3
$Li_{2.95}B_{0.95}Si_{0.05}S_3$
a-$Li_{2.95}B_{0.95}Si_{0.05}S_3$
normalized intensity
^{7}Li chemical shift (ppm)
(g)
Li_3BS_3
$Li_{2.95}B_{0.95}Si_{0.05}S_3$
a-$Li_{2.95}B_{0.95}Si_{0.05}S_3$
normalized intensity
^{11}B chemical shift (ppm)

Figure 4. Characterization of Li_3BS_3 with vacancy engineering. ***(a)*** *XRD patterns for Li_3BS_3, 2.5% Si substituted Li_3BS_3 ($Li_{2.975}B_{0.975}Si_{0.025}S_3$), 5% Si substituted Li_3BS_3 ($Li_{2.95}B_{0.95}Si_{0.05}S_3$), and amorphized 5% Si substituted Li_3BS_3 (a-$Li_{2.95}B_{0.95}Si_{0.05}S_3$). No impurities are observed in any pattern.* ***(b)*** *Arrhenius fits for Li_3BS_3. (c) Lattice parameter comparison for Li_3BS_3, $Li_{2.975}B_{0.975}Si_{0.025}S_3$, and $Li_{2.95}B_{0.95}B_{0.05}S_3$.* ***(d)*** *Arrhenius fits for $Li_{2.95}B_{0.95}Si_{0.05}S_3$, and a-$Li_{2.95}B_{0.95}Si_{0.05}S_3$. (e) Electrochemical impedance spectroscopy for the a-$Li_{2.95}B_{0.95}Si_{0.05}S_3$ at various temperatures. (f) ^{7}Li NMR and (g) ^{11}B NMR of the Li_3BS_3, $Li_{2.95}B_{0.95}Si_{0.05}S_3$, and a-$Li_{2.95}B_{0.95}Si_{0.05}S_3$. Results show that combined aliovalent substitution and amorphization can improve the ionic conductivity of Li_3BS_3 by over four orders of magnitude.*

Aliovalent substitution of Li_3BS_3 is achieved by substituting Si for B. The XRD patterns and quantitative Rietveld refinements of $Li_{2.975}B_{0.975}Si_{0.025}S_3$ and $Li_{2.95}B_{0.95}Si_{0.05}S_3$ are shown in Figure 4a. The lattice parameters from the refinements are plotted vs. stoichiometry with the Li_3BS_3 end-member in Figure 4e. The linear trend shows that the materials obey Vegard's law and confirms that Si incorporates into the lattice as a solid-solution. Substitution to 7.5% Si continues the Vegard trend but unidentified impurities are present. With 5% Si substitution, the ionic conductivity is improved to $1.82(\pm0.21)\cdot10^{-5}$ S cm^{-1} and the activation energy is decreased to 333±47 meV (Fig. 4d). Kimura et al. demonstrated that extended ball milling of Li_3BS_3 causes amorphization and improves ionic conductivity, likely due to introduction of defects[60,67]. Extended ball milling is attempted on the 5%-substituted Li_3BS_3 to assess whether both defect engineering strategies are compatible. Planetary ball milling of the 5%-substituted Li_3BS_3 for 100 h achieves amorphization (a-$Li_{2.95}B_{0.95}Si_{0.05}S_3$), as verified by the lack of distinct peaks in the XRD pattern shown in Figure 4a.

We find that amorphization significantly improves Li-ion conductivity. EIS measurements of a-$Li_{2.95}B_{0.95}Si_{0.05}S_3$ are shown in Figure 4e. A high-frequency semicircle is partially resolved which may represent grain boundary or bulk ionic transport. A Warburg tail is evident at lower frequencies, indicating that electronic charge transfer is blocked. Although multiple high-

frequency semicircles may exist (see supplemental information – section VII), a conservative estimate of the ionic conductivity is determined by linear fit of the Warburg tail and extrapolation to the x-intercept. The $\sigma_{25°C}$ of a-$Li_{2.95}B_{0.95}Si_{0.05}S_3$ is $1.07(\pm0.08)\cdot10^{-3}$ S cm^{-1} with an activation energy of 345±2 meV (Fig. 4d). The electronic conductivity as measured by DC polarization at 0.5 V is $3.3\cdot10^{-10}$ S cm^{-1}.

To determine if the local structure in the crystalline material is maintained after amorphization, we turn to ^{7}Li and ^{11}B NMR. If the local structure is not altered by amorphization, then it is likely that the ion diffusion pathways are similar. Comparing the ion diffusion pathways is important because the machine learning points to the structure of the crystalline Li_3BS_3 phase. The ^{7}Li NMR spectra of Li_3BS_3, $Li_{2.95}B_{0.95}Si_{0.05}S_3$, and a-$Li_{2.95}B_{0.95}Si_{0.05}S_3$ are shown in Figure 4d. All materials show a single resonance at the same chemical shift, suggesting the Li local environment remains unchanged. The resonance width narrows significantly in the amorphous material due to the higher mobility. The ^{11}B NMR measurements are shown in Figure 4g. The ^{11}B NMR for Li_3BS_3 and $Li_{2.95}B_{0.95}Si_{0.05}S_3$ show a single, quadrupolar environment that can be assigned to the $[BS_3]^{3-}$ moieties[67,72]. The signal from the a-$Li_{2.95}B_{0.95}Si_{0.05}S_3$ shows a similar signal to that of the crystalline phases but the shape changes, similarly to the previous measurement for amorphous Li_3BS_3[67]. Li_3BS_3, $Li_{2.95}B_{0.95}Si_{0.05}S_3$, and a-$Li_{2.95}B_{0.95}Si_{0.05}S_3$ all exhibit a major peak at ~60 ppm and a relatively minor peak ~0 ppm. The major peak is assigned to trigonal planar $[BS_3]^{3-}$ while the minor peak likely indicates a minor impurity with tetrahedrally coordinated B[73–75]. The change in shape of the ^{11}B spectrum upon amorphization is likely due an averaging of the quadrupolar couplings due to the fast Li dynamics. Thus, Li_3BS_3 and a-$Li_{2.95}B_{0.95}Si_{0.05}S_3$ have similar local structures and we can attribute the faster Li dynamics to the introduction of charge-carrying defects.

In addition to our experimental model validation, another of the predicted materials, KLi_6TaO_6, was recently synthesized with aliovalent Sn-substitution by Suzuki et al[59]. With a reported ionic conductivity near 10^{-5} S cm^{-1}, KLi_6TaO_6 is better than 70% of the SSEs in the semi-

supervised labels. Further improvement may be possible via extended amorphization to introduce structural defects, as is observed for Li_3BS_3.

Conclusions

Identification of functional materials is critical for improving technologies. Here, we show the utility of using semi-supervised learning as a method for guiding next-generation materials discovery in emerging fields. The method's focus on identifying the relationships between descriptors, prior to labeling, enables understanding of compositional spaces where most inputs are unlabeled. We demonstrate how semi-supervised learning can be used to identify descriptors and prospective superionic Li SSEs. By analyzing all Li-containing structures from the ICSD and MP database, we identify 212 materials that show promise as SSEs. All 212 structures exhibit a BVSE-predicted E_a below 0.6 eV.

The results illustrate that the utility of semi-supervised learning is contingent on careful screening of descriptors. While chemical intuition can be useful for descriptor selection, chemical intuition is often biased to favor previously investigated compositional spaces. For material discovery in emerging fields, use of handpicked descriptors may miss complex phenomena that more generally describe the dataset. Descriptor screening reveals which material properties are correlated to a property of interest to help enhance chemical intuition. In the case of Li SSEs, spatial descriptors excel over compositional, bonding, and electronic descriptors: the Smooth Overlap of Atomic Positions (SOAP), modified X-ray diffraction (mXRD), and general density descriptors are within the top four models. For spatial descriptors, simplification of the input structure tends to improve clustering outcomes. Removing the mobile ions from the structure and simplifying the remaining atoms, i.e. the "CAN" simplification, is most effective. Thus, the placement of framework atoms, but not their precise identity, is most correlated with ionic conductivity. Specifying the mobile ion positions hurts the model performance, suggesting a low correlation of mobile ion positions with ionic conductivity.

Predictions from the semi-supervised method are promising starting points for experimental identification of new superionic conductors but defects must be considered. The proposed materials are diverse, with the top thirty including halides, sulfides, tellurides, nitrides, oxides, and oxyhalides (see supplemental information – section IX). As a structure that falls outside of the eight routinely studied SSE classes, we demonstrate experimental characterization of Li_3BS_3 to confirm the utility of the approach. However, pure Li_3BS_3 exhibits poor ionic conductivity. Defects must be introduced into the material to achieve a superionic conductivity above 10^{-3} S cm^{-1}, a value that surpasses most reported SSEs. We note that the defects are introduced while maintaining the local structure of the crystalline material and thus the ionic conduction pathways are likely similar. The need to introduce defects highlights the paramount importance that defects play when measuring real materials. Many of the highest performing SSEs contain charge-carrying defects that are not explicitly encoded in their structure files. It is likely that some of the descriptors indirectly encode information about defects. By using experimental conductivity values as the evaluation metric, we may be prioritizing descriptors that encode information about a structures ability to support charge-carrying defects. To improve models and enhance chemical intuition, descriptors that explicitly encode defects are needed.

Now developed, the semi-supervised learning approach can serve as a template for material discovery beyond Li SSEs. The code is thoroughly documented following pythonic coding standards and made freely available on Github. Although the present effort focuses on Li SSEs, the approach is applicable to any material discovery space where labels are sparse. Discovery of new Li cathodes could be accomplished by using Li diffusivity, cathode capacity, and metal redox couple voltages as labels. Discovery of divalent SSEs (e.g. Mg^{2+}, Ca^{2+}, Zn^{2+}) could foreseeably be accomplished in a similar manner. The semi-supervised learning strategy may accelerate identification of fast ionic conductors for ion exchange membranes, solid oxide fuel cells, and various sensor applications.

Methods

Data Processing and Semi-supervised Learning

The ~26,000 input compositions are exported from the Inorganic Crystalline Structure Database (ICSD v.4.4.0) and Material's Project (MP - v.2020.09.08) as crystallographic information files (.cif). All structures containing Li are imported. Although transition metals could produce undesirable redox activity, transition metal containing structures are not screened out. Some of the best-performing SSEs contain transition metals (e.g. LLZO and LLTO). Entries that existed in both ICSD and MP are merged. Data manipulations and structure simplifications are performed using the Python libraries NumPy (v1.19.1), Pandas (v1.0.5), ASE (v3.19.1), and Pymatgen (v2020.8.3). Descriptor transformations are performed using the Python libraries Pymatgen (v2020.8.3), Matminer (v0.6.3), and Dscribe. Agglomerative hierarchical clustering is performed using the Python library scipy (v1.5.0). All code has been successfully executed on a custom-built CPU with an AMD Ryzen Threadripper 3990x Processor and 256 GB of RAM, in Ubuntu 20.04 running on Windows Subsystem for Linux 2. All code is made available on the github (https://github.com/FALL-ML/materials-discovery).

CI-NEB

Migration barriers for Li ion hopping are evaluated with the Climbing Image – Nudged Elastic Band (CI-NEB) method as implemented in the QuantumESPRESSO PWneb software package[76–79]. Density-functional theory (DFT) calculations are performed using the Perdew-Burke-Ernzerfof (PBE) generalized gradient approximation functional and projector-augmented wave (PAW) sets[80,81]. Convergence testing for the kinetic-energy cutoff of the plane-wave basis and the *k*-point sampling is performed for each structure to ensure an accuracy of 1 meV per atom. The lattice parameters and atomic positions of the as-retrieved structure are optimized. Supercells are created for each structure that are a minimum of 10 Å in each lattice direction to

minimize interactions between periodic images of the mobile ion. To study the migration barrier in the dilute limit, a single Li vacancy is created in the boundary endpoint structures of each studied pathway. A uniform background charge is used to balance excess charge. Each boundary configuration is relaxed until the force on each atom is less than $3x10^{-4}$ eV/Å. Images are created by linearly interpolating framework atomic positions between the initial and final boundary configurations. The initial pathway for the mobile ion is generated from the BVSE output minimum energy pathway to promote faster convergence of the NEB calculation. An NEB force convergence threshold of 0.05 eV/ Å is used. The calculation is first converged using the default NEB algorithm and then restarted with the CI scheme to allow for the maximum energy of the pathway to be determined.

Li_3BS_3 Synthesis

Li_3BS_3 is synthesized by reaction of Li_2S (Alfa Aesar, 99.9%), S_8 (Acros Organics, >99.5%), and elemental B (SkySpring Nanomaterials, Inc. 99.99%). The reactants are first mixed stoichiometrically (300 rpm for 1 h) using a planetary ball mill (MSE PMV1-0.4L) in 50 mL ZrO_2 jars with ZrO_2 balls. Two grams of reactants are always combined with 2 large balls (10 mm diameter), 34 medium balls (5 mm diameter), and 8 grams of small balls (3 mm diameter). Loading of ball mill jars occurs in an Ar-filled glovebox (Mbraun) and the jars are sealed before removal. After the 1 h of milling, the precursor mixture is pumped back into the glovebox and 330 – 340 mg of the powder is loaded into carbon coated vitreous silica ampoules (10 mm ID x 12 mm OD). The ampoules are evacuated (<10 mtorr) prior to sealing. Pure Li_3BS_3 is obtained via a four-step heating protocol in a Lindberg/Blue furnace: (1) ramp to 500 °C at 5 °C min^{-1}, (2) hold at 500 °C for 12 h, (3) ramp to 800 °C at 5 °C min^{-1}, and (4) hold at 800 °C for 6 h. The hot melt is then quenched from 800 °C into room temperature water. Recovered ingots are typically covered in a

C shell. The C shell is either sanded off or the ingot is ground into smaller pieces and the C is manually removed.

Substituted Li_3BS_3

Aliovalent substitution is accomplished by adding elemental Si (Acros, 99+%) into the precursor mixture prior to the 1 h mix. Si-substitution stoichiometry assumed that each Si atom replaces one Li and B: $Li_{3-x}B_{1-x}Si_xS_3$. Aside from the addition of Si, all steps are the same as for the synthesis of Li_3BS_3. Amorphization is accomplished via extended planetary ball milling in Ar of the 5% Si-substituted Li_3BS_3 ($Li_{2.95}B_{0.95}Si_{0.05}S_3$). Approximately 1 g of $Li_{2.95}B_{0.95}Si_{0.05}S_3$ is combined in a ZrO_2 ball mill jar with 3 large balls (10 mm diameter), 51 medium balls (5 mm diameter), and 12 g of small balls (3 mm diameter). The powder is ground in a planetary ball mill (MSE PMV1-0.4L), under Ar atmosphere, for 100 h.

Material Characterization

Li_3BS_3 materials are characterized using powder X-ray diffraction (XRD) and electrochemical impedance spectroscopy (EIS). XRD patterns are attained on a Rigaku Smartlab by scanning from 10° to 70° 2θ at 2 degrees per minute. The Smartlab employs a Cu-Kα source with a 20 kV accelerating voltage. For EIS measurements, 50-100 mg of powder is first hot-pressed (100 °C, 5 min) into a 1/4" diameter pellet. The pellet faces are polished using diamond lapping powder (Allied High Tech Products Inc.) in sequentially finer grits: 60, 30, 6, 0.5, and 0.1 micron. Au contacts are sputtered (90 s at 40 mA) onto the polished surfaces using a 108 Auto Sputter Coater (Cressington). Pellets are then assembled into a Swagelok 1/4" cell with stainless steel current collectors. After applying pressure with a hand vise (~100 MPa), EIS data is collected on a VSP-300 with a Biologic low-current channel. All EIS data is collected to an upper frequency

of 3 MHz. The lower frequency is case dependent, with a frequency cutoff selected such that the Warburg polarization feature is visible. ^{7}Li and ^{11}B MAS MAS NMR spectra were acquired using a Bruker DSX-500 spectrometer with a 4 mm ZrO_2 rotor. The operating frequencies for ^{7}Li and ^{11}B are 190.5 and 160.5 MHz, respectively. The ^{7}Li and ^{11}B spectra were referenced to a 1 M LiCl aq. solution and BF_3-OEt_2, respectively. A spinning speed of 12 kHz was used, and the spectra were gathered after applying a single 0.5 μs to 15° pulse for both ^{7}Li and ^{11}B.

AUTHOR INFORMATION

Corresponding Author

Kimberly A. See – Division of Chemistry and Chemical Engineering, Department of Applied Physics and Materials, California Institute of Technology, Pasadena, CA 91125, USA. E-mail: ksee@caltech.edu

Author Contributions

The manuscript was written through contributions of all authors. All authors have given approval to the final version of the manuscript. The authors declare no competing financial interest.

Data Availability

The data that support the findings of this study are available from the corresponding author upon reasonable request.

ACKNOWLEDGEMENTS

F.A.L.L acknowledges the support of the Arnold and Mabel Beckman Foundation via a 2020 Arnold O. Beckman Postdoctoral Fellowship in Chemical Sciences. F.A.L.L would also like to thank Andrew J. Martinolich for his guidance and insightful scientific input. The NEB computations presented here were conducted in the Resnick High Performance Computing Center, a facility supported by Resnick Sustainability Institute at the California Institute of Technology.

References

1. Bachman, J. C. *et al.* Inorganic Solid-State Electrolytes for Lithium Batteries: Mechanisms and Properties Governing Ion Conduction. *Chem. Rev.* **116**, 140–162 (2016).
2. Kamaya, N. *et al.* A lithium superionic conductor. *Nat. Mater.* **10**, 682–686 (2011).
3. Adeli, P. *et al.* Boosting solid-state diffusivity and conductivity in lithium superionic argyrodites by halide substitution. *Angew. Chem.* **131**, 8773–8778 (2019).
4. Seino, Y., Ota, T., Takada, K., Hayashi, A. & Tatsumisago, M. A sulphide lithium super ion conductor is superior to liquid ion conductors for use in rechargeable batteries. *Energy Environ. Sci.* **7**, 627–631 (2014).
5. Zhu, Y., He, X. & Mo, Y. First principles study on electrochemical and chemical stability of solid electrolyte–electrode interfaces in all-solid-state Li-ion batteries. *J. Mater. Chem. A* **4**, 3253–3266 (2016).
6. Richards, W. D., Miara, L. J., Wang, Y., Kim, J. C. & Ceder, G. Interface stability in solid-state batteries. *Chem. Mater.* **28**, 266–273 (2016).
7. Kerman, K., Luntz, A., Viswanathan, V., Chiang, Y.-M. & Chen, Z. Review—practical challenges hindering the development of solid state li ion batteries. *J. Electrochem. Soc.* **164**, A1731 (2017).
8. Wenzel, S. *et al.* Direct observation of the interfacial instability of the fast ionic conductor $Li_{10}GeP_2S_{12}$ at the lithium metal anode. *Chem. Mater.* **28**, 2400–2407 (2016).
9. Zhu, Y., He, X. & Mo, Y. Origin of outstanding stability in the lithium solid electrolyte materials: insights from thermodynamic analyses based on first-principles calculations. *ACS Appl. Mater. Interfaces* **7**, 23685–23693 (2015).
10. Wenzel, S., Sedlmaier, S. J., Dietrich, C., Zeier, W. G. & Janek, J. Interfacial reactivity and interphase growth of argyrodite solid electrolytes at lithium metal electrodes. *Solid State Ion.* **318**, 102–112 (2018).

11. Ding, Z., Li, J., Li, J. & An, C. Review—interfaces: key issue to be solved for all solid-state lithium battery technologies. *J. Electrochem. Soc.* **167**, 070541 (2020).
12. Wang, S. *et al.* Interfacial challenges for all-solid-state batteries based on sulfide solid electrolytes. *J. Materiomics* **7**, 209–218 (2021).
13. Ren, F. *et al.* Accelerated discovery of metallic glasses through iteration of machine learning and high-throughput experiments. *Sci. Adv.* **4**, eaaq1566 (2018).
14. Tran, K. & Ulissi, Z. W. Active learning across intermetallics to guide discovery of electrocatalysts for CO_2 reduction and H_2 evolution. *Nat. Catal.* **1**, 696–703 (2018).
15. Lu, S. *et al.* Accelerated discovery of stable lead-free hybrid organic-inorganic perovskites via machine learning. *Nat. Commun.* **9**, 3405 (2018).
16. Xue, D. *et al.* Accelerated search for materials with targeted properties by adaptive design. *Nat. Commun.* **7**, 11241 (2016).
17. D. Sendek, A. *et al.* Holistic computational structure screening of more than 12000 candidates for solid lithium-ion conductor materials. *Energy Environ. Sci.* **10**, 306–320 (2017).
18. Butler, K. T., Davies, D. W., Cartwright, H., Isayev, O. & Walsh, A. Machine learning for molecular and materials science. *Nature* **559**, 547–555 (2018).
19. Liu, Y., Zhao, T., Ju, W. & Shi, S. Materials discovery and design using machine learning. *J. Materiomics* **3**, 159–177 (2017).
20. Ziletti, A., Kumar, D., Scheffler, M. & Ghiringhelli, L. M. Insightful classification of crystal structures using deep learning. *Nat. Commun.* **9**, 2775 (2018).
21. Isayev, O. *et al.* Universal fragment descriptors for predicting properties of inorganic crystals. *Nat. Commun.* **8**, 15679 (2017).
22. Schütt, K. T., Arbabzadah, F., Chmiela, S., Müller, K. R. & Tkatchenko, A. Quantum-chemical insights from deep tensor neural networks. *Nat. Commun.* **8**, 13890 (2017).

23. Ward, L. *et al.* Matminer: An open source toolkit for materials data mining. *Comput. Mater. Sci.* **152**, 60–69 (2018).
24. Himanen, L. *et al.* DScribe: Library of descriptors for machine learning in materials science. *Comput. Phys. Commun.* **247**, 106949 (2020).
25. Juan, Y., Dai, Y., Yang, Y. & Zhang, J. Accelerating materials discovery using machine learning. *J. Mater. Sci. Technol.* **79**, 178–190 (2021).
26. Suzuki, K. *et al.* Fast material search of lithium ion conducting oxides using a recommender system. *J. Mater. Chem. A* **8**, 11582–11588 (2020).
27. Wang, Z. *et al.* Harnessing artificial intelligence to holistic design and identification for solid electrolytes. *Nano Energy* **89**, 106337 (2021).
28. Zhong, M. *et al.* Accelerated discovery of CO_2 electrocatalysts using active machine learning. *Nature* **581**, 178–183 (2020).
29. Wang, Z., Zhang, H. & Li, J. Accelerated discovery of stable spinels in energy systems via machine learning. *Nano Energy* **81**, 105665 (2021).
30. Yuan, R. *et al.* Accelerated discovery of large electrostrains in $BaTiO_3$-based piezoelectrics using active learning. *Adv. Mater.* **30**, 1702884 (2018).
31. Li, J. *et al.* Accelerated discovery of high-strength aluminum alloys by machine learning. *Commun. Mater.* **1**, 1–10 (2020).
32. Zhang, Y. *et al.* Unsupervised discovery of solid-state lithium ion conductors. *Nat. Commun.* **10**, 1–7 (2019).
33. Liu, Y., Zhou, Q. & Cui, G. Machine learning boosting the development of advanced lithium batteries. *Small Methods* **5**, 2100442 (2021).
34. Forestier, G. & Wemmert, C. Semi-supervised learning using multiple clusterings with limited labeled data. *Inf. Sci.* **361–362**, 48–65 (2016).
35. Thangadurai, V., Narayanan, S. & Pinzaru, D. Garnet-type solid-state fast Li ion conductors for Li batteries: critical review. *Chem. Soc. Rev.* **43**, 4714–4727 (2014).

36. Gorai, P., Famprikis, T., Singh, B., Stevanović, V. & Canepa, P. Devil is in the defects: electronic conductivity in solid electrolytes. *Chem. Mater.* **33**, 7484–7498 (2021).
37. van Engelen, J. E. & Hoos, H. H. A survey on semi-supervised learning. *Mach. Learn.* **109**, 373–440 (2020).
38. Schütt, K. T. *et al.* SchNet: A continuous-filter convolutional neural network for modeling quantum interactions. *ArXiv170608566 Phys. Stat* (2017).
39. Artrith, N. & Urban, A. An implementation of artificial neural-network potentials for atomistic materials simulations: Performance for TiO_2. *Comput. Mater. Sci.* **114**, 135–150 (2016).
40. Adams, S. & Rao, R. P. *In structure and bonding, Vol. 158, Bond Valences, edited by ID Brown and KR Poeppelmeier*. (Berlin, Heidelberg: Springer, 2014).
41. Hansen, K. *et al.* Machine learning predictions of molecular properties: accurate many-body potentials and nonlocality in chemical space. *J. Phys. Chem. Lett.* **6**, 2326–2331 (2015).
42. Butler, M. A. & Ginley, D. S. Prediction of flatband potentials at semiconductor-electrolyte interfaces from atomic electronegativities. *J. Electrochem. Soc.* **125**, 228 (1978).
43. He, B. *et al.* CAVD, towards better characterization of void space for ionic transport analysis. *Sci. Data* **7**, 153 (2020).
44. Ward, L. *et al.* Including crystal structure attributes in machine learning models of formation energies via Voronoi tessellations. *Phys. Rev. B* **96**, 024104 (2017).
45. Ong, S. P. *et al.* Python Materials Genomics (pymatgen): A robust, open-source python library for materials analysis. *Comput. Mater. Sci.* **68**, 314–319 (2013).
46. Deml, A. M., O'Hayre, R., Wolverton, C. & Stevanović, V. Predicting density functional theory total energies and enthalpies of formation of metal-nonmetal compounds by linear regression. *Phys. Rev. B* **93**, 085142 (2016).
47. Ewald, P. P. Die Berechnung optischer und elektrostatischer Gitterpotentiale. *Ann. Phys.* **369**, 253–287 (1921).

48. Choudhary, K. *et al.* The joint automated repository for various integrated simulations (JARVIS) for data-driven materials design. *Npj Comput. Mater.* **6**, 1–13 (2020).
49. Meredig, B. *et al.* Combinatorial screening for new materials in unconstrained composition space with machine learning. *Phys. Rev. B* **89**, 094104 (2014).
50. Pham, T. L. *et al.* Machine learning reveals orbital interaction in materials. *Sci. Technol. Adv. Mater.* **18**, 756–765 (2017).
51. Rupp, M., Tkatchenko, A., Müller, K.-R. & von Lilienfeld, O. A. Fast and accurate modeling of molecular atomization energies with machine learning. *Phys. Rev. Lett.* **108**, 058301 (2012).
52. Faber, F., Lindmaa, A., Lilienfeld, O. A. von & Armiento, R. Crystal structure representations for machine learning models of formation energies. *Int. J. Quantum Chem.* **115**, 1094–1101 (2015).
53. Krivovichev, S. V. Structural complexity of minerals: information storage and processing in the mineral world. *Mineral. Mag.* **77**, 275–326 (2013).
54. Ward, L., Agrawal, A., Choudhary, A. & Wolverton, C. A general-purpose machine learning framework for predicting properties of inorganic materials. *Npj Comput. Mater.* **2**, 1–7 (2016).
55. Ward, J. H. Hierarchical grouping to optimize an objective function. *J. Am. Stat. Assoc.* **58**, 236–244 (1963).
56. Fung, V., Zhang, J., Juarez, E. & Sumpter, B. G. Benchmarking graph neural networks for materials chemistry. *Npj Comput. Mater.* **7**, 1–8 (2021).
57. He, X., Zhu, Y. & Mo, Y. Origin of fast ion diffusion in super-ionic conductors. *Nat. Commun.* **8**, 15893 (2017).
58. Sun, W. *et al.* The thermodynamic scale of inorganic crystalline metastability. *Sci. Adv.* **2**, e1600225 (2016).

59. Suzuki, N. *et al.* Theoretical and experimental studies of KLi_6TaO_6 as a Li-ion solid electrolyte. *Inorg. Chem.* **60**, 10371–10379 (2021).
60. Vinatier, P., Ménétrier, M. & Levasseur, A. Structure and ionic conduction in lithium thioborate glasses and crystals. *Phys. Chem. Glas.* **44**, 135–142 (2003).
61. Sendek, A. D. *et al.* Combining superionic conduction and favorable decomposition products in the crystalline lithium–boron–sulfur system: a new mechanism for stabilizing solid li-ion electrolytes. *ACS Appl. Mater. Interfaces* **12**, 37957–37966 (2020).
62. Bianchini, F., Fjellvåg, H. & Vajeeston, P. A first-principle investigation of the Li diffusion mechanism in the super-ionic conductor lithium orthothioborate Li_3BS_3 structure. *Mater. Lett.* **219**, 186–189 (2018).
63. Park, H., Yu, S. & Siegel, D. J. Predicting charge transfer stability between sulfide solid electrolytes and Li metal anodes. *ACS Energy Lett.* **6**, 150–157 (2021).
64. Vinatier, P., Gravereau, P., Ménétrier, M., Trut, L. & Levasseur, A. Li_3BS_3. *Acta Crystallogr. C* **50**, 1180–1183 (1994).
65. Zhang, L. *et al.* Lithium lanthanum titanate perovskite as an anode for lithium ion batteries. *Nat. Commun.* **11**, 3490 (2020).
66. Belous, A. G., Novitskaya, G. N., Polyanetskaya, S. V. & Gornikov, Y. I. Investigation into complex oxides of $La_{2/3-x}Li_{3x}TiO_3$ composition. *Izv Akad Nauk SSSR Neorg Mater* **23**, 470–472 (1987).
67. Kimura, T. *et al.* Characteristics of a Li_3BS_3 thioborate glass electrolyte obtained via a mechanochemical process. *ACS Appl. Energy Mater.* **5**, 1421–1426 (2022).
68. Zhou, L., Minafra, N., Zeier, W. G. & Nazar, L. F. Innovative approaches to Li-argyrodite solid electrolytes for all-solid-state lithium batteries. *Acc. Chem. Res.* **54**, 2717–2728 (2021).
69. Zhao, W., Yi, J., He, P. & Zhou, H. Solid-state electrolytes for lithium-ion batteries: fundamentals, challenges and perspectives. *Electrochem. Energy Rev.* **2**, 574–605 (2019).

70. Lacivita, V., Artrith, N. & Ceder, G. Structural and compositional factors that control the Li-ion conductivity in LIPON electrolytes. *Chem. Mater.* **30**, 7077–7090 (2018).
71. Knauth, P. Inorganic solid Li ion conductors: An overview. *Solid State Ion.* **180**, 911–916 (2009).
72. Larink, D., Eckert, H. & Martin, S. W. Structure and ionic conductivity in the mixed-network former chalcogenide glass system $[Na_2S]_{2/3}[(B_2S_3)_x(P_2S_5)_{1-x}]_{1/3}$. *J. Phys. Chem. C* **116**, 22698–22710 (2012).
73. Hwang, S.-J. *et al.* Structural study of $xNa_2S + (1 - x)B_2S_3$ glasses and polycrystals by multiple-quantum MAS NMR of ^{11}B and ^{23}Na. *J. Am. Chem. Soc.* **120**, 7337–7346 (1998).
74. Kaup, K. *et al.* A lithium oxythioborosilicate solid electrolyte glass with superionic conductivity. *Adv. Energy Mater.* **10**, 1902783 (2020).
75. Curtis, B., Francis, C., Kmiec, S. & Martin, S. W. Investigation of the short range order structures in sodium thioborosilicate mixed glass former glasses. *J. Non-Cryst. Solids* **521**, 119456 (2019).
76. Henkelman, G., Uberuaga, B. P. & Jónsson, H. A climbing image nudged elastic band method for finding saddle points and minimum energy paths. *J. Chem. Phys.* **113**, 9901–9904 (2000).
77. Henkelman, G. & Jónsson, H. Improved tangent estimate in the nudged elastic band method for finding minimum energy paths and saddle points. *J. Chem. Phys.* **113**, 9978–9985 (2000).
78. Giannozzi, P. *et al.* QUANTUM ESPRESSO: a modular and open-source software project for quantum simulations of materials. *J. Phys. Condens. Matter* **21**, 395502 (2009).
79. Giannozzi, P. *et al.* Advanced capabilities for materials modelling with Quantum ESPRESSO. *J. Phys. Condens. Matter* **29**, 465901 (2017).
80. Perdew, J. P., Burke, K. & Ernzerhof, M. Generalized gradient approximation made simple. *Phys. Rev. Lett.* **77**, 3865–3868 (1996).

81. Dal Corso, A. Pseudopotentials periodic table: From H to Pu. *Comput. Mater. Sci.* **95**, 337–350 (2014).