Supporting Information

Machine Learning Reveals Factors that Control Ion Mobility in Anti-Perovskite Solid Electrolytes

Kwangnam Kim[†] and Donald J. Siegel[†, ‡, §, ||, *]

[†]Mechanical Engineering Department, [‡]Materials Science & Engineering, [§]Applied Physics Program, [||]University of Michigan Energy Institute, University of Michigan, Ann Arbor, Michigan 48109-2125, United States

*Donald.Siegel@austin.utexas.edu

Descriptors

Descriptors in the 'elemental' category include the ionic radius,[1] atomic mass,[2] Pauling electronegativity,[3,4] polarizability, Bader charge, and Born effective charge. The polarizability, α, of alkali metal ions was estimated by averaging reported values in solid-state chalcogenides and halides (e.g., α_K in K_3OCl was determined by averaging α_K in K_2O and in KCl). Similarly, the polarizabilities of the anions were adopted from chalcogenides and halides (e.g., α_O and α_{Cl} in K_3OCl were adopted from their values in K_2O and KCl, respectively).[5,6] Bader charges and Born effective charges were predicted using the authors' prior DFT calculations on the anti-perovskites.[7–13] Since two cation sites are involved in a migration path as the end points of that path, the Bader charge and the Born effective charge corresponding to each path were estimated by averaging the values of the cations at each end point. The Bader and Born effective charges for a given anion species were taken as the average over all instances of that species within the unit cell.

Descriptors in the 'lattice' category include those related to volume, atomic packing, coordination, phonon frequencies, and atomic distances. The packing fraction of the anions (PF_A) is defined as the ratio of the unit cell volume occupied by the anions to the total cell volume. This fraction was obtained by overlaying the cell volume onto a grid, and defining the packing fraction as the ratio of grid points that overlap with atomic spheres to the total number of grid points. This procedure is iterated using increasingly finer grids until the packing fraction converges.[14] The tolerance factor, *t*, measures the degree of lattice distortion in the (anti)perovskite structure,[15] and is calculated by $t = (R_X + R_B)/\left[\sqrt{2}(R_X + R_A)\right]$, where R_X, R_B, and R_A are the atomic radii of the cation (X site), the anion at the octahedron center (A site), and the anion on cubic framework sites (B site). The lattice tends to be cubic for compositions with $t \sim 1$, and becomes orthorhombic for smaller t (usually when $t <\sim 0.8$). The channel size is defined as the largest diameter of a sphere that can pass through the anion sublattice; it was calculated using the *Zeo++* code.[16] The channel size was reported as an important factor for ion mobility within a series of LGPS-related electrolytes.[17] The coordination number of a cation was determined by evaluating the number of neighboring ions within a distance of 1.2 times of the sum of the cation's and anion's ionic radii. The coordination numbers of two cations at the ends of path were averaged (as was done for the Bader charge).

Atomic distances used as features include: the distance between neighboring cations, CCD, equivalent to the length of an octahedron edge); the distance between cation and octahedral/framework anions, COD and CFD, respectively; and the between anions, AAD.[14] COD and CFD were evaluated as the distance from a cation to the closest octahedral/framework anion. COD/CFD of cations at the ends of a migration path were averaged (as done for the Bader charge). AAD was obtained by identifying the closest neighboring anion to each anion, and averaging these distances over all anions in the cell.

The path width was evaluated by identifying ions near to the migration path, and calculating the perpendicular distance from the path to the nearest ion(s) (see Ref.[14] for a detailed explanation). Three categories of path width are considered: a width determined by the distance to closest ion (PW_c), a width determined by the distance to the second-closest ion (PW_{2c}), and the total path width (PW), which is the sum of PW_c and PW_{2c}.

Lattice dynamics (i.e., phonon frequencies) have been reported to be import for ion mobility.[18–20] We employed the frequency of lowest-energy optical phonon (ω_{LEO}) to account for the effect of lattice dynamics.[13,19] Bulk modulus was included as a descriptor in the 'mechanical' category. This modulus was calculated in the authors' prior study.[11,12]

Descriptors in the 'electronic' category include the band gap, dielectric constant, and polarizability. The total dielectric constant (ε) was obtained by the sum of electronic (ε_e) and ionic (ε_i) contributions. ε_e and ε_i were calculated in the authors' prior DFT calculations.[13] High polarizability has been linked to high ion mobility through a 'lattice softening' effect.[21] The polarizability of compound was calculated using the Clausius-Mossotti relation:[22]

$$\bar{\alpha} = \frac{3}{4\pi}\left(\frac{\varepsilon_e - 1}{\varepsilon_e + 2}\right) v_{\text{atom}},$$

where v_{atom} is the lattice volume per atom.

Finally, descriptors in the 'Chemical' category include bond ionicity and thermodynamic stabilities. The bond ionicity between a cation and chalcogen anion ($I_{C\text{-}Ch}$), between cation and halogen anion ($I_{C\text{-}H}$), and between chalcogen and halogen anions ($I_{Ch\text{-}H}$) were calculated as the difference in electronegativities between the respective species.[14] The energy above the convex hull (i.e., decomposition energy) was included as a measure of stability of the anti-perovskite.[11,12,23] The formation energies of a neutral vacancy and of an interstitial dumbbell were calculated as:[24]

$$E_f = E_{\text{AP\_defect}} - \left(E_{\text{AP\_pristine}} + \sum_i n_i \mu_i\right),$$

where $E_{\text{AP\_defect}}$ and $E_{\text{AP\_pristine}}$ are the total energies of the anti-perovskite supercell with and without a cation defect, *i* is the defect species (Li, Na, or K), and n_i and μ_i are, respectively, the number of cations removed or added to the cell (n = -1 for a vacancy and +1 for an interstitial dumbbell) and the chemical potential of the cation. The defect formation energy corresponding to a given elementary path is assigned by averaging the formation energies of the two defects at the end points of that path.

Feature filtering

In a few cases correlated descriptors did not undergo filtering. This occurred in cases where the descriptors appeared to be unrelated (based on the authors' chemical intuition) or they conveyed information needed to differentiate between migration pathways. As an example of the former scenario, the feature pair 'Bader charge of cation (BC_C)' and 'Cation – octahedral anion distance (COD)' were not filtered despite having a correlation ρ = -0.81. For the latter scenario, 'Path distance (CCD)' and 'Volume per atom (V_a)' have ρ = +0.89, reflecting that a compound comprised of larger atoms will have greater distances between cations. Nevertheless, both descriptors were retained because CCD is needed to differentiate distinct cation migration paths in different compounds and within a compound. Also, for simplicity, secondary correlates were eliminated as redundant descriptors; e.g., E_g and ω_{LEO} were removed since they both correlate with B, yet B also correlates strongly with V_a. Hence, through the transitive property, all of E_g, ω_{LEO}, and B were dropped as redundant descriptors in favor of retaining V_a.

Machine-learning algorithms

Ridge[25] and *LASSO*[26] linear regression regularize the error function by adding a penalty (L2 or L1 norm), so as to reduce the absolute value of the regression coefficients and prevent overfitting. The *elastic net* uses both L1 and L2 norms with an additional hyperparameter that controls the ratio between contributions of two penalties.[27] This algorithm performs better than LASSO when some features are correlated.[27] *Bayesian ridge regression* is a iterative learning process performed by updating coefficients; similar to *Ridge* regression, it includes a regularization term (L2 norm).[28] *Bayesian automatic relevance determination (ARD) regression* is a modified version of the Bayesian ridge regression that uses separated regularization hyperparameters to model coefficients; this is particularly effective when that data contains irrelevant features.[29] The regularization hyperparameter(s) in the Bayesian models are optimized automatically during fitting. *Huber* regression is similar to ridge regression, but uses an absolute error for outliers so that the model is less sensitive to them (an additional hyperparameter controls the threshold to identify outliers).[30] *Passive-aggressive regression* (*PAR*) is a margin-based regression algorithm using the hinge loss (allowing errors when they are within a margin, otherwise the linear L1 error is used), which is suitable for large-scale data.[31] *Random sample consensus* (RANSAC) is an iterative algorithm to find inlier samples that will be used to train the final model (a hyperparameter controls the ratio of random subsamples to total samples). RANSAC was implemented with the ridge and LASSO regressors (namely *RANSAC + Ridge* and *RANSAC + LASSO*, respectively).

Support vector regressor (SVR) is a regression version of the *support vector machine* (SVM) classifier that uses a separating hyperplane with a soft margin that allows some degree of misclassification.[28] *Epsilon-SVR* determines support vectors (i.e., misclassified samples) using the margin area and minimizes the regularized hinge loss with the support vectors.[28,32,33] *Linear-SVR* is similar to *epsilon-SVR* but uses a different solver suitable for large-scale data.[34] Rather than using margin area, *Nu-SVR* controls the number of support vectors (ratio of data outside the margin to entire data).[28,33,35]

Decision tree regression (*DTR*) constructs decision criteria (e.g., $x_1 \geq \theta_1$, $x_2 < \theta_2$, $x_2 \geq \theta_3$, etc., where x are features and θ are feature values for thresholds) so that samples in a node are divided into leaves of a tree-like structure.[28,36] A fully-grown decision tree is likely to suffer from overfitting, so adjusting a decision tree (i.e., 'pruning') can improve the predictive accuracy.[37] Adjustable hyperparameters are (1) the minimum number of samples to split, (2) the minimum number of samples at a leaf, (3) the maximum depth of the tree, and (4) the maximum number of features considered at each split. *Random forest regression* (*RFR*) averages tree models trained with random samples, a technique called 'bootstrap aggregating (bagging)' that improves the prediction (a hyperparameter controls the number of tree models).[38–40] *Extremely randomized tree regression* (*ERTR*) is similar to RFR, but further randomizes each tree model by arbitrarily selecting cut-point choices.[41] *Boosting* is a technique that trains multiple weak learners (models slightly better than random) during iterative learning. The learners are used as an ensemble classifier (or regressor); this approach can perform better than any single model.[28] The 'adaptive boosting (adaboost)' algorithm gives greater weights to training samples with higher errors shown in a previous weak model.[42,43] We implemented the adaboost algorithm with RFR and ERTR as weak models (namely *adaboost + RFR* and *adaboost + ERTR*, respectively). The *gradient boosting regression tree* (*GBRT*) successively fits a new model to residual errors shown in a previous step, and adds the predicted residuals to the previous prediction.[44]

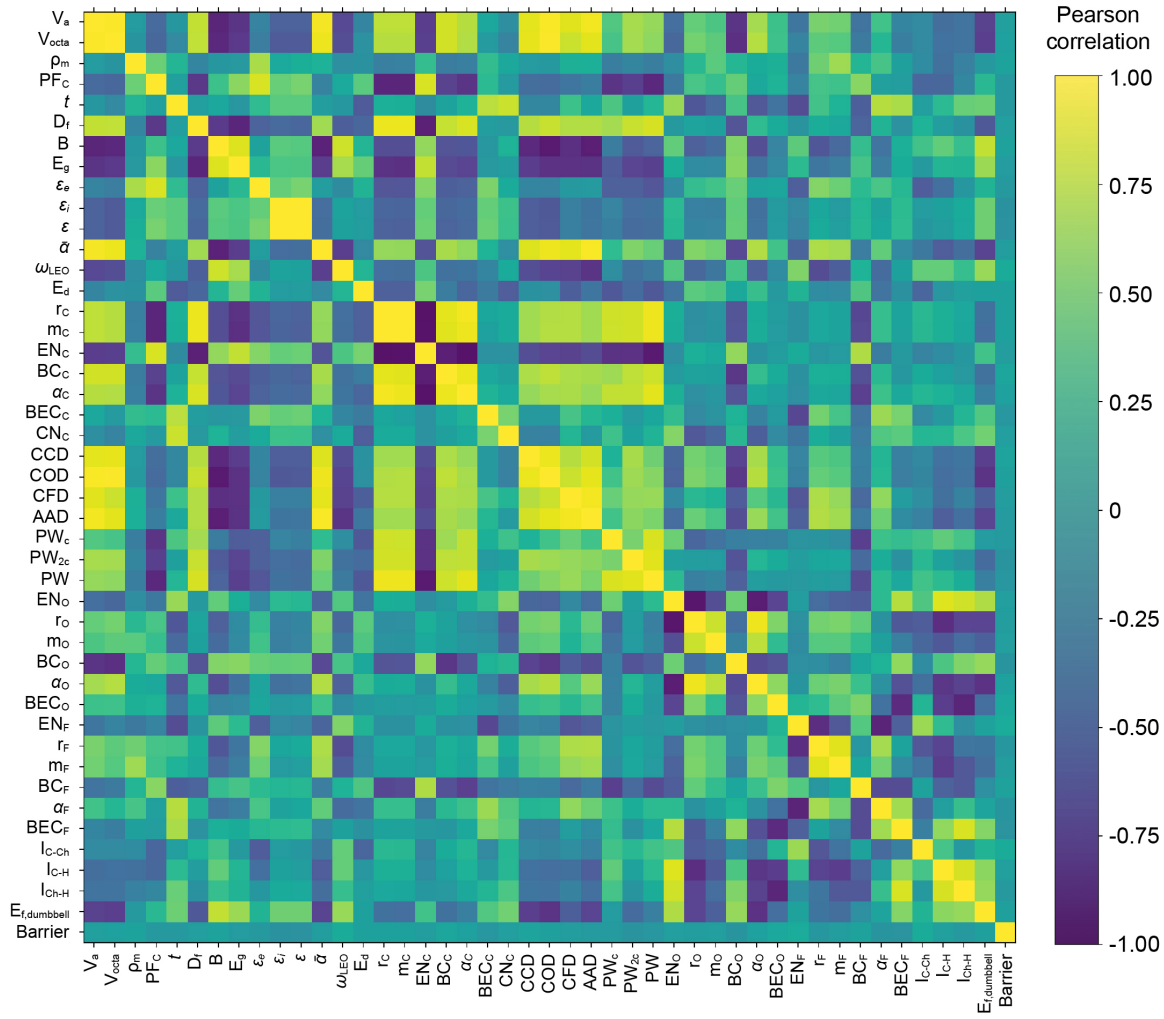

Figure S1. Pearson correlation analysis of descriptors and interstitial dumbbell migration barriers.

Table S1. Optimal feature combinations for a given number features for ML models for vacancy migration barriers. Values in parenthesis indicate feature importance.

No. of features in model	Features – vacancy migration							
	1st	2nd	3rd	4th	5th	6th	7th	8th
Single feature	V_a (100%)							
Two features	CCD (69%)	V_a (31%)						
Three features	CCD (46%)	PW (32%)	α_F (22%)					
Four features	CCD (45%)	PW (30%)	$I_{C\text{-}Ch}$ (14%)	$I_{C\text{-}H}$ (10%)				
Five features	CCD (42%)	PW (28%)	α_F (11%)	α_O (10%)	BC_O (9%)			
Six features	CCD (44%)	PW (28%)	$I_{C\text{-}Ch}$ (11%)	α_F (7%)	V_a (6%)	BEC_O (4%)		
Seven features	CCD (42%)	PW (29%)	$I_{C\text{-}Ch}$ (10%)	α_F (6%)	$I_{C\text{-}H}$ (5%)	E_d (5%)	ρ_m (4%)	
Eight features	CCD (42%)	PW (28%)	$I_{C\text{-}Ch}$ (8%)	α_F (6%)	BC_O (5%)	$I_{C\text{-}H}$ (5%)	V_a (4%)	BC_F (4%)

Table S2. Optimal feature combination for a given number features for ML models of interstitial migration barriers. Values in parenthesis indicate feature importance.

No. of features in model	Features – interstitial migration				
	1st	2nd	3rd	4th	5th
Single feature	E_f (100%)				
Two features	PW (53%)	BEC_C (47%)			
Three features	PW (49%)	E_f (35%)	$I_{C\text{-}Ch}$ (16%)		
Four features	PW (43%)	E_f (29%)	α_O (15%)	E_d (13%)	
Five features	PW (40%)	E_f (28%)	α_O (13%)	ρ_m (10%)	E_d (9%)

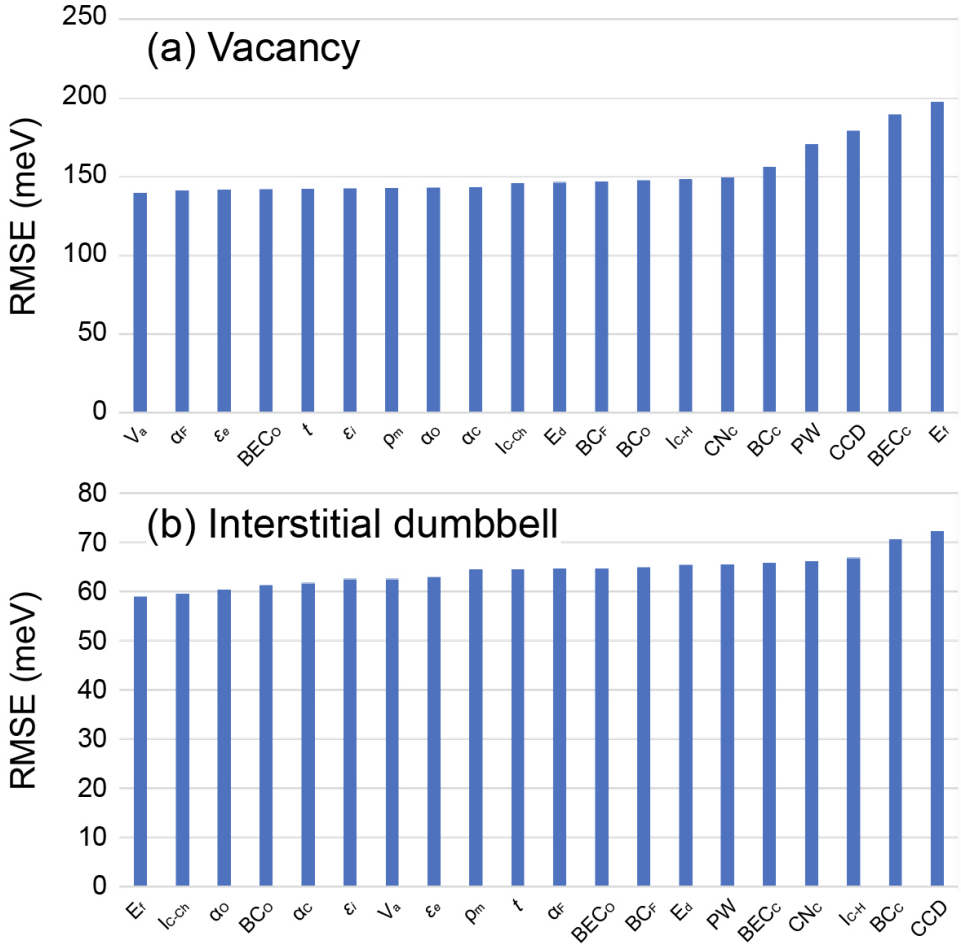

Figure S2. Test error (i.e., RMSE for the migration energy as evaluated within the test dataset) of 'adaboost + ERTR' models constructed using different single features as input for the (a) vacancy and (b) interstitial dumbbell migration mechanisms. The features are sorted by increasing RMSE from left to right.

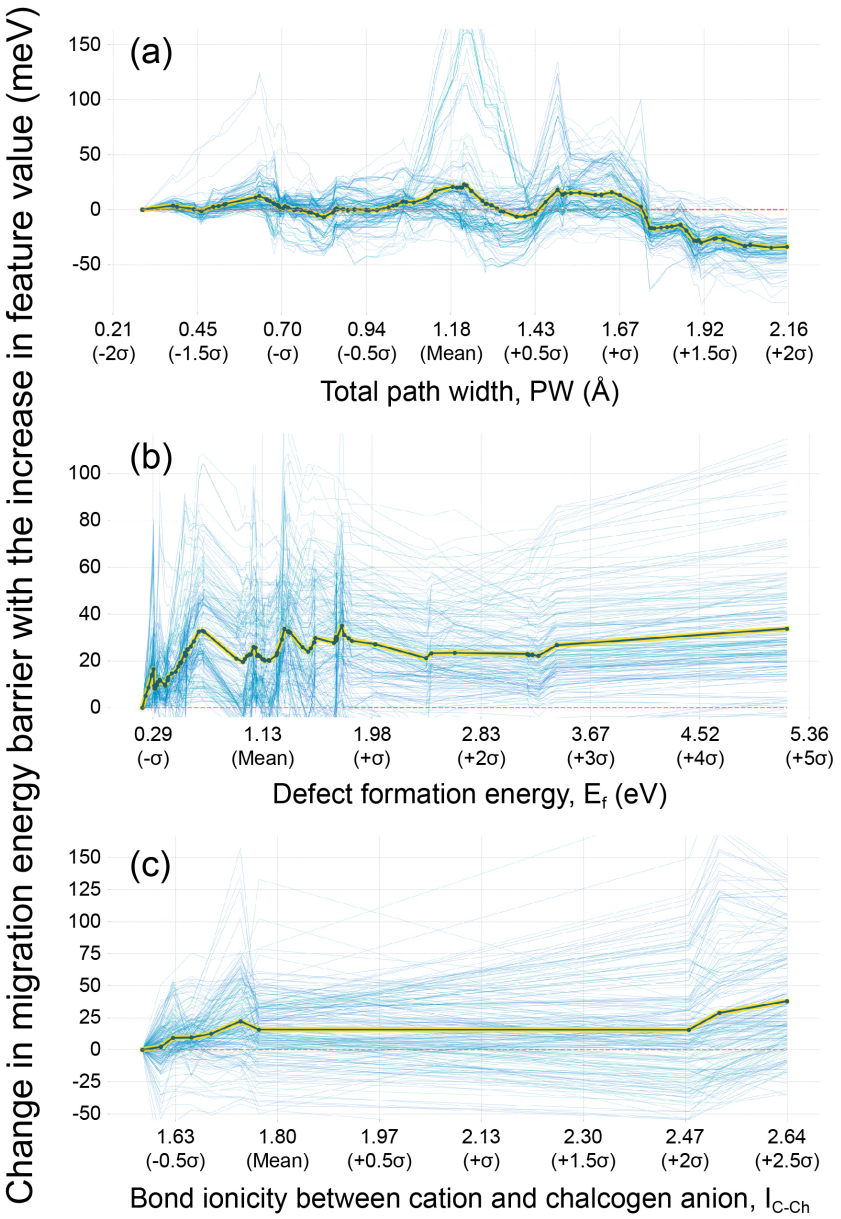

Figure S3. Individual conditional expectation (ICE) plots for features relevant for interstitial dumbbell migration: (a) total path width (PW), (b) defect formation energy (E_f), and (c) bond ionicity between cation and chalcogen anion ($I_{C\text{-}Ch}$). The bold line depicts the average of the migration barrier over all curves. σ denotes the standard deviation of a given feature within the training data. Blue lines represent the change in migration barrier for each migration path in the training dataset as the target feature (PW, E_f, $I_{C\text{-}Ch}$) is varied from its minimum to its maximum value.

REFERENCES

(1) Shannon, R. D. Revised Effective Ionic Radii and Systematic Studies of Interatomic Distances in Halides and Chalcogenides. *Acta Crystallogr. Sect. A* **1976**, *32* (5), 751–767. https://doi.org/10.1107/S0567739476001551.

(2) Coursey, J. S.; Schwab, D. J.; Tsai, J. J.; Dragoset, R. A. *Atomic Weights and Isotopic Compositions with Relative Atomic Masses*; National Institute of Standards and Technology (NIST) Physical Measurement Laboratory, 2015.

(3) Pauling, L. THE NATURE OF THE CHEMICAL BOND. IV. THE ENERGY OF SINGLE BONDS AND THE RELATIVE ELECTRONEGATIVITY OF ATOMS. *J. Am. Chem. Soc.* **1932**, *54* (9), 3570–3582. https://doi.org/10.1021/ja01348a011.

(4) Allred, A. L. Electronegativity Values from Thermochemical Data. *J. Inorg. Nucl. Chem.* **1961**, *17* (3–4), 215–221. https://doi.org/10.1016/0022-1902(61)80142-5.

(5) Shanker, J.; Agrawal, S. C. Electronic Polarizabilities and Sizes of Ions in Alkaline Earth Halides and Alkali Chalcogenides. *J. Phys. Chem. Solids* **1980**, *41* (3), 209–213. https://doi.org/10.1016/0022-3697(80)90187-0.

(6) Mahan, G. D.; Subbaswamy, K. R. *Local Density Theory of Polarizability*; Springer US: Boston, MA, 1990. https://doi.org/10.1007/978-1-4899-2486-5.

(7) Bader, R. F. W. *Atoms in Molecules: A Quantum Theory*; Oxford University Press: New York, 1990.

(8) Henkelman, G.; Arnaldsson, A.; Jónsson, H. A Fast and Robust Algorithm for Bader Decomposition of Charge Density. *Comput. Mater. Sci.* **2006**, *36* (3), 354–360. https://doi.org/10.1016/j.commatsci.2005.04.010.

(9) Sanville, E.; Kenny, S. D.; Smith, R.; Henkelman, G. Improved Grid-Based Algorithm for Bader Charge Allocation. *J. Comput. Chem.* **2007**, *28* (5), 899–908. https://doi.org/10.1002/jcc.20575.

(10) Tang, W.; Sanville, E.; Henkelman, G. A Grid-Based Bader Analysis Algorithm without Lattice Bias. *J. Phys. Condens. Matter* **2009**, *21* (8), 084204. https://doi.org/10.1088/0953-8984/21/8/084204.

(11) Kim, K.; Siegel, D. J. Correlating Lattice Distortions, Ion Migration Barriers, and Stability in Solid Electrolytes. *J. Mater. Chem. A* **2019**, *7* (7), 3216–3227. https://doi.org/10.1039/C8TA10989C.

(12) Kim, K. Computational Discovery of Solid Electrolytes for Batteries: Interfacial Phenomena and Ion Mobility. PhD Thesis, University of Michigan, Ann Arbor, 2020.

(13) Kim, K.; Li, Y.; Tsai, P.-C.; Wang, F.; Son, S.-B.; Chiang, Y.-M.; Siegel, D. J. Exploring the Synthesis of Alkali Metal Anti-Perovskites. *Chem. Mater.* **2022**. https://doi.org/10.1021/acs.chemmater.1c02150.

(14) Sendek, A. D.; Yang, Q.; Cubuk, E. D.; Duerloo, K.-A. N.; Cui, Y.; Reed, E. J. Holistic Computational Structure Screening of More than 12 000 Candidates for Solid Lithium-Ion Conductor Materials. *Energy Environ. Sci.* **2017**, *10* (1), 306–320. https://doi.org/10.1039/C6EE02697D.

(15) Goldschmidt, V. M. Die Gesetze Der Krystallochemie. *Naturwissenschaften* **1926**, *14* (21), 477–485. https://doi.org/10.1007/BF01507527.

(16) Willems, T. F.; Rycroft, C. H.; Kazi, M.; Meza, J. C.; Haranczyk, M. Algorithms and Tools for High-Throughput Geometry-Based Analysis of Crystalline Porous Materials. *Microporous Mesoporous Mater.* **2012**, *149* (1), 134–141. https://doi.org/10.1016/j.micromeso.2011.08.020.

(17) Ong, S. P.; Mo, Y.; Richards, W. D.; Miara, L.; Lee, H. S.; Ceder, G. 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.* **2013**, *6* (1), 148–156. https://doi.org/10.1039/C2EE23355J.

(18) Köhler, U.; Herzig, C. On the Correlation between Self-Diffusion and the Low-Frequency LA ⅔ ⟨ 111 ⟩ Phonon Mode in b . c . c . Metals. *Philos. Mag. A* **1988**, *58* (5), 769–786. https://doi.org/10.1080/01418618808209952.

(19) Wakamura, K. Roles of Phonon Amplitude and Low-Energy Optical Phonons on Superionic Conduction. *Phys. Rev. B - Condens. Matter Mater. Phys.* **1997**, *56* (18), 11593–11599.

(20) Muy, S.; Bachman, J. C.; Giordano, L.; Chang, H.; Abernathy, D. L.; Bansal, D.; Delaire, O.; Hori, S.; Kanno, R.; Maglia, F.; et al. Tuning Mobility and Stability of Lithium Ion Conductors Based on Lattice Dynamics. *Energy Environ. Sci.* **2018**, *11* (4), 850–859. https://doi.org/10.1039/C7EE03364H.

(21) Kraft, M. A.; Culver, S. P.; Calderon, M.; Böcher, F.; Krauskopf, T.; Senyshyn, A.; Dietrich, C.; Zevalkink, A.; Janek, J.; Zeier, W. G. Influence of Lattice Polarizability on the Ionic Conductivity in the Lithium Superionic Argyrodites Li_6PS_5X (X = Cl, Br, I). *J. Am. Chem. Soc.* **2017**, *139* (31), 10909–10918. https://doi.org/10.1021/jacs.7b06327.

(22) Pearson, E. W.; Jackson, M. D.; Gordon, R. G. A Theoretical Model for the Index of Refraction of Simple Ionic Crystals. *J. Phys. Chem.* **1984**, *88* (1), 119–128. https://doi.org/10.1021/j150645a030.

(23) 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* **2016**, *4* (9), 3253–3266. https://doi.org/10.1039/C5TA08574H.

(24) Van de Walle, C. G.; Neugebauer, J. First-Principles Calculations for Defects and Impurities: Applications to III-Nitrides. *J. Appl. Phys.* **2004**, *95* (8), 3851–3879. https://doi.org/10.1063/1.1682673.

(25) Hoerl, A. E.; Kennard, R. W. Ridge Regression: Biased Estimation for Nonorthogonal Problems. *Technometrics* **1970**, *12* (1), 55–67.

(26) Tibshirani, R. Regression Shrinkage and Selection via the Lasso. *J. R. Stat. Soc. Ser. B (Statistical Methodol.* **1996**, *58* (1), 267–288.

(27) Zou, H.; Hastie, T. Regression and Variable Selection via the Elastic Net. *J. R. Stat. Soc. Ser. B (Statistical Methodol.* **2005**, *67* (2), 301–320. https://doi.org/10.1111/j.1467-9868.2005.00503.x.

(28) Bishop, C. M. *Pattern Recognition and Machine Learning*; Springer: New York, NY, 2006.

(29) Mackay, D. J. C. Bayesian Methods for Backpropagation Networks. In *Models of Neural Networks III*; Domany, E., van Hemmen, J. L., Schulten, K., Eds.; Springer, 1994; pp 211–254.

(30) Huber, P. J. Robust Estimation of a Location Parameter. *Ann. Math. Stat.* **1964**, *35* (1), 73–101. https://doi.org/10.1214/aoms/1177703732.

(31) Crammer, K.; Dekel, O.; Keshet, J.; Shalev-Shwartz, S.; Singer, Y. Online Passive-Aggressive Algorithms. *J. Mach. Learn. Res.* **2006**, *7*, 551–585.

(32) Drucker, H.; Burges, C. J. C.; Kaufman, L.; Smola, A.; Vapnik, V. Support Vector Regression Machines. In *Neural Information Processing*

Systems; Mozer, M. C., Joradn, J. I., Petsche, T., Eds.; MIT Press: Cambridge, MA, 1997; Vol. 9, pp 155–161.
(33) Chang, C.-C.; Lin, C.-J. LIBSVM: A Library for Support Vector Machines. *ACM Trans. Intell. Syst. Technol.* **2011**, *2* (3), 27. https://doi.org/10.1145/1961189.1961199.
(34) Fan, R.-E.; Chang, K.-W.; Hsieh, C.-J.; Wang, X.-R.; Lin, C.-J. LIBLINEAR: A Library for Large Linear Classification. *J. Mach. Learn. Res.* **2008**, *9*, 1871–1874.
(35) Schölkopf, B.; Smola, A. J.; Williamson, R. C.; Bartlett, P. L. New Support Vector Algorithms. *Neural Comput.* **2000**, *12*, 1207–1245.
(36) Breiman, L.; Friedman, J. H.; Olshen, R. A.; Stone, C. J. *Classification and Regression Trees*; Chapman and Hall/CRC: Boca Raton, FL, 1984.
(37) Bramer, M. *Principles of Data Mining*, 2nd ed.; Springer-Verlag: London, 2007. https://doi.org/10.1007/978-1-4471-4884-5_1.
(38) Breiman, L. Bagging Predictors. *Mach. Learn.* **1996**, *24* (2), 123–140. https://doi.org/10.1007/BF00058655.
(39) Tin Kam Ho. Random Decision Forests. In *Proceedings of 3rd International Conference on Document Analysis and Recognition*; IEEE Comput. Soc. Press, 1995; Vol. 1, pp 278–282. https://doi.org/10.1109/ICDAR.1995.598994.
(40) Breiman, L. Random Forests. *Mach. Learn.* **2001**, *45* (1), 5–32. https://doi.org/10.1023/A:1010933404324.
(41) Geurts, P.; Ernst, D.; Wehenkel, L. Extremely Randomized Trees. *Mach. Learn.* **2006**, *63* (1), 3–42. https://doi.org/10.1007/s10994-006-6226-1.
(42) Freund, Y.; Schapire, R. E. A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting. *J. Comput. Syst. Sci.* **1997**, *55* (1), 119–139. https://doi.org/10.1006/jcss.1997.1504.
(43) Drucker, H. Improving Regressors Using Boosting Techniques. In *Proceedings of the 14th International Conference on Machine Learning*; 1997; pp 107–115.
(44) Friedman, J. H. . Greedy Function Approximation: A Gradient Boosting Machine. *Ann. Stat.* **2001**, *29* (5), 1189–1232.