Crystal Structure Search with Random Relaxations Using Graph Networks

Gowoon Cheon
Stanford University
gcheon@stanford.edu

Lusann Yang
Google Research
lusann@google.com

Kevin McCloskey
Google Research
mccloskey@google.com

Evan J. Reed
Stanford University
evanreed@stanford.edu

Ekin D. Cubuk
Google Research
cubuk@google.com

Abstract

Materials design enables technologies critical to humanity, including combating climate change with solar cells and batteries[1, 2, 3]. The properties of a material are determined by its atomic crystal structure. However, prediction of the atomic crystal structure for a given material is a long-standing grand challenge that remains a barrier in materials design. We investigate a data-driven approach to accelerating ab initio random structure search (AIRSS)[4, 5], a state-of-the-art method for crystal structure search. We build a novel dataset of random structure relaxations of Li-Si battery anode materials using high-throughput density functional theory calculations. We train graph neural networks to simulate relaxations of random structures. Our model is able to find an experimentally verified structure of $Li_{15}Si_4$ it was not trained on, and has potential for orders of magnitude speedup over AIRSS. Surprisingly, we find that data augmentation of adding Gaussian noise improves both the accuracy and robustness of our models.

1 Introduction

The atomic structure of a material determines its physical and chemical properties. For instance, graphite (a dark, soft, conductor) and diamond (a transparent, hard, insulator) both consist of carbon atoms that are arranged in different structures. In crystalline solids, atoms are arranged in patterns that repeat periodically, called unit cells. Crystal structure search - predictions of unit cells that may be experimentally observed - is a long-standing problem in materials science[6]. The number of possible configurations for a unit cell can be very large. A unit cell with volume 10 $Å^3$ and 10 atoms already has 10^{25} possible configurations (see appendix), and the number of possible structures increases exponentially with the number of atoms and atom types. Materials Project[7], one of the most widely used databases of materials, has 29.5 atoms per cell on average. But among the large number of possible structures, or local minima in the energy landscape, only the global minimum and a few lowest-energy local minima close in energies may be observed in nature.

Crystal structure search usually requires the use of computationally expensive density functional theory (DFT) calculations. Interatomic potentials that give fast approximations to these calculations often generalize poorly, meaning that potentials trained on a few structures fail to describe another structure that is experimentally observed[8, 9]. This is in contrast to molecule discovery, where the structure of most small molecules are determined by atomic bonds and interatomic potentials work well for structure search[10].

Machine Learning for Molecules Workshop at NeurIPS 2020. https://ml4molecules.github.io

In this paper, we build a novel dataset of structure relaxations from random initial states in the Li-Si system via high-throughput DFT calculations, and train graph networks to reproduce the relaxation trajectory. The Li-Si system has at least 7 crystal structures[11] with different stoichiometries that have been experimentally verified[12, 13, 14, 15, 11, 16, 17], and is widely studied for applications in lithium ion batteries[18, 19, 20]. Compared to our molecular dynamics simulations of crystalline Li-Si phases, the random initial configurations yield relaxations that cover a wide range of high-stress unit cells not observed in molecular dynamics even at high temperatures(see appendix).

Once trained, our graph networks enable random structure search with a fraction of the computational time used for ab initio random structure search methods, and benefit from better out-of-domain generalization to unseen stoichiometries. We improve generalization using data augmentation. There has been evidence of Gaussian noise increasing robustness [21] and interpretability[22] at the cost of accuracy in vision and particle-based simulations[23]; in vision, this behavior has been attributed to reduced sensitivity to high frequency noise[24, 25]. In contrast, we show that adding small perturbations to atomic positions help in calculating accurate forces and simulating relaxation trajectories.

2 Related Work

Crystal structure prediction is an optimization problem in the structure-energy landscape. Popular methods include random sampling[5, 4], evolutionary algorithms[26, 27, 28], simulated annealing[29] and basin hopping[30, 31]. These methods involve at least hundreds of computationally expensive ab initio calculations[32], which take hours on a supercomputer. Elemental substitutions from libraries of known crystal structures[33, 34] is another technique that requires only a few calculations. To address the problem of slow ab initio calculations, some researchers build faster approximations to energies and forces[35], including graph neural networks[36, 37, 38, 39, 40, 23, 41]. A combination of random structure searching and machine learned potentials has been proposed for phosphorus system in [9, 42]. The Li-Si system has been studied with machine learned potentials[43, 44, 45], genetic algorithm[46], embedded atom methods[47] and random sampling[48].

In computer vision, data augmentation[49, 25, 24] has been used for improving the accuracy of image classification models, and work on adversarial examples discuss model robustness against distribution shifts[50, 51, 52]. Recent work on adversarial examples in physical systems[53] showed that models approximating physical systems move atoms close to adversarial directions maximizing the error in energy predictions, which sheds light on why adding Gaussian noise was advantageous in this work.

3 Implementation

3.1 Dataset of Random Structure Relaxations

We use AIRSS[4, 5] to generate 5000 random unit cells for each of: LiSi, Li_2Si, Li_7Si_2, $Li_{13}Si_4$ and $Li_{15}Si_4$. AIRSS constructs random structures by choosing a set of random unit cell lengths and angles, and placing the atoms in random locations in the unit cell. Additional constraints on the cell volume, minimum separation between atoms and symmetry may also be imposed on the cell as described in [4, 5] and in the appendix. We use the Vienna Ab initio Simulation Package (VASP)[54] to perform DFT relaxations on the random unit cells. The total number of converged random structure relaxations for each stoichiometry were 4680(LiSi), 4365($Li_{13}Si_4$), 4596(Li_2Si), 4515(Li_7Si_2) and 4199($Li_{15}Si_4$). For all stoichiometries except Li_7Si_2, experimentally reported structures were recovered.

49.5% of the random structures match the converged final structure by the time the maximum force acting on the atoms reaches 0.1 eV/Å. We split the data into high-force (maximum force on atoms > 0.1 eV/Å) and low-force (maximum force on atoms < 0.1 eV/Å) regions. Since our goal is to quickly screen random structures to find the few lowest energy ones, we focus our discussion on learning energy relaxations from the high-force dataset as computed using a steepest descent algorithm.

We construct our training set by sampling structures from the relaxation trajectory from regions where the maximum force acting on the atoms is larger than 0.1 eV/Å. We sample at most one structure out of every five ionic steps. We only sample structures that do not match the previous structure in the trajectory. We used pymatgen's[55] StructureMatcher class with stol=0.05 for structure comparison. StructureMatcher compares two structures by reducing them to primitive cells and evaluating whether

the maximum RMS displacement is less than stol, see [56]. After sampling, we use 80% of the trajectories for training, 10% each for validation and test. For trajectory predictions, we use all ionic steps until the maximum force in the crystal reaches 0.1 eV/Å.

3.2 Dataset of Molecular Dynamics Simulations

We take the experimental crystalline structures of LiSi, Li_2Si, Li_7Si_2, $Li_{13}Si_4$ and $Li_{15}Si_4$ from the Materials Project and perform molecular dynamics(MD) simulations with VASP. We created 64 to 76-atom supercells of the conventional structure in the Materials Project, and ran the simulations at 800K for at least 47 picoseconds. Again, we used StructureMatcher to discard matching structures, and end up with 17134(LiSi), 16088($Li_{13}Si_4$), 9613(Li_2Si), 13079(Li_7Si_2) and 12749($Li_{15}Si_4$) structures for each stoichiometry.

3.3 Graph Neural Networks (GNN)

We use graph neural nets to compute the forces on each atom and the stress tensor on the unit cell, with the architecture based on MEGNet[38]. Each node is initialized to states corresponding to the atomic number of the element, and a 16-dimensional embedding layer is applied. Edges are constructed for every pair of atoms within 4Å, and each edge is initialized to the displacement vector between the atoms, distance, and distance squared ($(\vec{r_{ij}}, |\vec{r_{ij}}|, |\vec{r_{ij}}|^2)$, where $\vec{r_{ij}} = \vec{r_i} - \vec{r_j}$). Global states are initialized to 0. Each state is passed through 7 message-passing layers as per in MEGNet, consisting of two MLP layers of size 64 and state updates. To get forces on each atom, we concatenate node features(with two MLP layers of size 64 and 32) with global features and follow with a output layer without activation.

For stress tensor predictions, we put fictitious non-interacting atoms on (100), (010) and (001) planes. We used the vector distances from the planes to each atom to compute edge features. Each fictitious atom outputs three stress tensor components - e.g. the atom on the x-axis outputs $\sigma_{xx}, \sigma_{yx}, \sigma_{zx}$.[1]

The force and stress inputs were scaled to have mean 0 and standard deviation 1. We used the sum of mean squared errors for forces and stresses for our loss function. For data augmentation, we rotate 50% of the structures at each epoch by 90 degrees, and perturb the atom positions and lattice vectors with Gaussian noise for 30% of the structures. The models were trained with ADAM optimizer with learning rate 10^{-3}, and after hyperparameter optimization the learning rate was halved every 15 steps. Using the force and stress outputs, we simulate relaxations using the velocity verlet equations as implemented in VASP with details in the appendix.

As a baseline, we train another model on predicting crystal energies with a similar architecture, and numerically differentiate the energies to get forces and stresses. After the message-passing layers, the node and edge states are aggregated with set2set[57] and concatenated with the global state, and passed through two MLP layers of sizes 64 and 32. The final predictions are produced by an output layer without activation. The energy inputs were also scaled to have mean 0 and standard deviation 1, and sum of mean squared errors was used as loss function. We used rotation augmentation for training the energy model.

4 Experiments

We train the models below on the Li-Si system, holding out $Li_{15}Si_4$ for out-of-domain generalization experiments.

Impact of data augmentation. We investigate the effects of adding Gaussian noise to atom positions and lattice vectors. When simulating relaxation trajectories, errors may accumulate over multiple time steps. Adding small perturbations during training may make the simulations more robust over longer trajectories.[23]

As a baseline, we use the second nearest-neighbor modified embedded atom method (2NN MEAM) from ref.[58]. We evaluate the force and stress prediction accuracy by mean absolute error on the test set sampled as described above. We simulate random structure relaxations with predicted force and

[1] we investigated three architectures and chose the one with best performance. Appendix section *Architectures for stress calculations*

stress outputs using the same algorithm as VASP and use pymatgen's StructureMatcher to compare the final relaxed structures. Additionally, we match structures at each step in the VASP and GNN trajectories and report the fraction of structures that match.

We evaluate model performance using metrics for force and stress prediction accuracy, and for producing accurate relaxation trajectory in Table 1. All GNN models outperform MEAM by an order of magnitude difference in both force and stress MAE. In contrast to ref.[23] where Gaussian noise helped simulate longer trajectories at the cost of force errors at each step (one-step MSE in their paper), in our system adding Gaussian noise leads to more accurate force and stress outputs and better relaxation trajectories. While the force and stress outputs are the most accurate with noise of 0.01 eV/Å, the trajectory prediction improves with a larger noise of 0.02 eV/Å. The advantages of Gaussian noise decrease with larger amounts of noise, suggesting that the energy landscape is locally smooth at the scale of 0.01 to 0.02 eV/Å.

Table 1: **Gaussian noise improves both force predictions and relaxation trajectories.** The units for Gaussian noise standard deviation, force and stress are Å, eV/Åand kBar. We averaged the results from three random seeds. Quantities in parentheses are the standard deviations from different random seeds. The last row uses MEAM from [47]. *: Fraction of relaxed final structures that match VASP **: Fraction of matching structures in each trajectory

Model	Gaussian noise std	Force MAE	Stress MAE	Relaxed structure match*	Matches in trajectory**
GNN	0	0.033 (2.1×10^{-4})	1.05 (0.021)	0.40 (0.015)	0.65 (0.015)
GNN	0.01	**0.031** (4.5×10^{-4})	**0.98** (0.047)	0.42 (0.007)	**0.67** (0.006)
GNN	0.02	0.032 (1.4×10^{-4})	1.01 (0.023)	**0.43** (0.005)	**0.67** (0.000)
GNN	0.03	0.034 (1.9×10^{-4})	1.06 (0.050)	0.42 (0.007)	0.66 (0.006)
MEAM	-	0.55	9.11	0.18	0.39

Impact of random structure data and architecture. Compared to the MEAM baseline, we have significant changes: a dataset rich in high stress structures, and a new GNN architecture. We evaluate the impact of the training data on model performance by training the same GNN on structures from molecular dynamics(MD) trajectories and evaluating it on the random structures data. To evaluate our architecture for predicting crystal forces and stresses, we note that many interatomic potentials are trained on energies, and forces and stresses are obtained from numerically differentiating the energies with respect to atom positions or lattice vectors. We establish a baseline comparison by training another GNN on energies of random structure relaxations, and use numerical differentiation to get forces and stresses. The model trained on random structure energies achieved a MAE of 15 meV/atom with more details in the appendix. The results are in Table 2.

Our architecture trained on random structure data shows an order of magnitude improvement against both the MD baseline and energy model baseline. The MD baseline suggests that the distribution of data generated by MD is quite different(see appendix) and training on random structures is crucial for crystal structure search. The poor performance of energy model baseline in our work is consistent with the findings from [53]. This also helps explain why adding Gaussian noise was effective in this work - models trained with Gaussian noise are less affected by the presence of adversarial directions[59, 60].

Table 2: **Using random structure relaxations with our architecture for directly predicting forces and stresses leads to an order of magnitude improvement in accuracy.**

Model	Trained on	Training data	Test data	Force MAE	Stress MAE
GNN	forces,stresses	MD	MD	0.067	1.63
GNN	forces,stresses	**MD**	**random relaxations**	0.61	32.3
GNN	**energies**	random relaxations	random relaxations	0.24	18.7
MEAM	-	-	random relaxations	0.55	9.11

Out-of-domain generalization and structure search of $Li_{15}Si_4$. We trained the models on four stoichiometries in the Li-Si system - LiSi, Li_2Si, Li_7Si_2, $Li_{13}Si_4$. We tested the model on 4199 random structure relaxations of $Li_{15}Si_4$. Again, we compare it to the 2NN MEAM baseline. Out of the 4199 random structures, three of the structures relaxed into the same experimental structure

reported in [13], and the structures relaxed to the force cutoff of 0.1 eV/Åmatches the final relaxed structures. We report how many out of these three trajectories each model predicted correctly in Table 3. All GNN models simulated two out of the three experimental structure trajectories, and MEAM did not find the experimental structure. This is surprising since $Li_{15}Si_4$ was included in the data used to fit MEAM. It suggests that for random structure relaxations, the behavior of the force field model at high-stress regions may relax the structure into the wrong basin in the energy landscape and training on this region may be crucial for random structure search.

Table 3: **Generalization of the model on $Li_{15}Si_4$**. *: number of trajectories that relaxed into the experimental structure, out of three found with VASP.

Model	Gaussian noise std	Relaxed structure match	Matches in trajectory	experimental*
GNN	0	0.35	0.64	2/3
GNN	0.01	0.38	0.66	2/3
GNN	0.02	**0.39**	**0.67**	2/3
GNN	0.03	**0.39**	0.66	2/3
MEAM	-	0.19	0.39	0/3

VASP simulations of $Li_{15}Si_4$ structures took 38 hours on average on Google Cloud's standard N1 machine with 8 cores. GNN simulations take less than a minute on a laptop (MacBook Pro (15-inch, 2017)), and is still able to find the experimental structure of $Li_{15}Si_4$.

5 Conclusion

We build a novel dataset of DFT random structure relaxations, and show that graph network models trained on this data has significant advantages for crystal structure search. Moreover, we show that using Gaussian noise in training force field models improve both accuracy in force predictions and robustness in simulating trajectories. It would be interesting to draw on the literature on adversarial defense and explore how it can be used to improve physical models.

6 Acknowledgements

We would like to thank Michael Brenner for helpful discussions.

References

[1] Daniel P. Tabor, Loïc M. Roch, Semion K. Saikin, Christoph Kreisbeck, Dennis Sheberla, Joseph H. Montoya, Shyam Dwaraknath, Muratahan Aykol, Carlos Ortiz, Hermann Tribukait, Carlos Amador-Bedolla, Christoph J. Brabec, Benji Maruyama, Kristin A. Persson, and Alán Aspuru-Guzik. Accelerating the discovery of materials for clean energy in the era of smart automation. *Nature Reviews Materials*, 3(5):5–20, May 2018. Number: 5 Publisher: Nature Publishing Group.

[2] Austin D. Sendek, Ekin D. Cubuk, Evan R. Antoniuk, Gowoon Cheon, Yi Cui, and Evan J. Reed. Machine Learning-Assisted Discovery of Solid Li-Ion Conducting Materials. *Chemistry of Materials*, 31(2):342–352, January 2019. Publisher: American Chemical Society.

[3] Ryan Jacobs, Guangfu Luo, and Dane Morgan. Materials Discovery of Stable and Nontoxic Halide Perovskite Materials for High-Efficiency Solar Cells. *Advanced Functional Materials*, 29(23):1804354, 2019. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/adfm.201804354.

[4] Chris J. Pickard and R. J. Needs. High-Pressure Phases of Silane. *Physical Review Letters*, 97(4), July 2006.

[5] Chris J. Pickard and R. J. Needs. Ab initio random structure searching. *Journal of Physics: Condensed Matter*, 23(5):053201, January 2011.

[6] Scott M. Woodley and Richard Catlow. Crystal structure prediction from first principles. *Nature Materials*, 7(12):937–946, December 2008. Number: 12 Publisher: Nature Publishing Group.

[7] A. Jain, S.P. Ong, G. Hautier, W. Chen, W.D. Richards, S. Dacek, S. Cholia, D. Gunter, D. Skinner, G. Ceder, and K.A. Persson. The Materials Project: A materials genome approach to accelerating materials innovation. *APL Materials*, 1(1, 011002), 2013.

[8] Szymon Winczewski, Mohamad Yousef Shaheen, and Jarosław Rybicki. Interatomic potential suitable for the modeling of penta-graphene: Molecular statics/molecular dynamics studies. *Carbon*, 126:165–175, January 2018.

[9] Volker L. Deringer, Davide M. Proserpio, Gábor Csányi, and Chris J. Pickard. Data-driven learning and prediction of inorganic crystal structures. *Faraday Discussions*, 211(0):45–59, 2018. Publisher: The Royal Society of Chemistry.

[10] Anna Theresa Cavasin, Alexander Hillisch, Felix Uellendahl, Sebastian Schneckener, and Andreas H. Göller. Reliable and Performant Identification of Low-Energy Conformers in the Gas Phase and Water. *Journal of Chemical Information and Modeling*, 58(5):1005–1020, May 2018.

[11] Michael Zeilinger, Iryna M. Kurylyshyn, Ulrich Häussermann, and Thomas F. Fässler. Revision of the Li–Si Phase Diagram: Discovery and Single-Crystal X-ray Structure Determination of the High-Temperature Phase Li4.11Si. *Chemistry of Materials*, 25(22):4623–4632, November 2013. Publisher: American Chemical Society.

[12] Linda A. Stearns, Jan Gryko, Jason Diefenbacher, Ganesh K. Ramachandran, and Paul F. McMillan. Lithium monosilicide (LiSi), a low-dimensional silicon-based material prepared by high pressure synthesis: NMR and vibrational spectroscopy and electrical properties characterization. *Journal of Solid State Chemistry*, 173(1):251–258, June 2003.

[13] Sven Dupke, Thorsten Langer, Rainer Pöttgen, Martin Winter, Stefano Passerini, and Hellmut Eckert. Structural characterization of the lithium silicides Li15Si4, Li13Si4, and Li7Si3 using solid state NMR. *Physical Chemistry Chemical Physics*, 14(18):6496–6508, April 2012. Publisher: The Royal Society of Chemistry.

[14] H. Axel, Herbert Schäfer, and Armin Weiss. Die Kristallstruktur von Lithiumsilicid Li2Si. *Angewandte Chemie*, 77(8):379–380, 1965. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/ange.19650770809.

[15] Herbert Schäfer, Hartmut Axel, and Armin Weiss. Die Kristallstruktur der Phase Li7Si2. *Zeitschrift für Naturforschung B*, 20(10):1010–1010, October 1965. Publisher: De Gruyter Section: Zeitschrift für Naturforschung B.

[16] Ursula Frank, Wiking Müller, and Herbert Schäfer. Zur Kenntnis der Phase Li13Si4 / On the Phase Li13Si4. *Zeitschrift für Naturforschung B*, 30(1-2):10–13, February 1975. Publisher: De Gruyter Section: Zeitschrift für Naturforschung B.

[17] J EVERS, G. OEHLINGER, and G. SEXTL. LiSi, a unique Zintl phase : although stable, it long evaded synthesis. *LiSi, a unique Zintl phase : although stable, it long evaded synthesis*, 34(7-8):773–784, 1997. Place: Paris Publisher: Elsevier.

[18] Xin Su, Qingliu Wu, Juchuan Li, Xingcheng Xiao, Amber Lott, Wenquan Lu, Brian W. Sheldon, and Ji Wu. Silicon-Based Nanomaterials for Lithium-Ion Batteries: A Review. *Advanced Energy Materials*, 4(1):1300882, 2014. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/aenm.201300882.

[19] Jeannine R. Szczech and Song Jin. Nanostructured silicon for high capacity lithium battery anodes. *Energy & Environmental Science*, 4(1):56–72, 2011. Publisher: Royal Society of Chemistry.

[20] Ekin D. Cubuk and Efthimios Kaxiras. Theory of Structural Transformation in Lithiated Amorphous Silicon. *Nano Letters*, 14(7):4065–4070, July 2014. Publisher: American Chemical Society.

[21] Justin Gilmer, Nicolas Ford, Nicholas Carlini, and Ekin Cubuk. Adversarial examples are a natural consequence of test error in noise. In *International Conference on Machine Learning*, pages 2280–2289, 2019.

[22] Daniel Smilkov, Nikhil Thorat, Been Kim, Fernanda Viégas, and Martin Wattenberg. Smoothgrad: removing noise by adding noise. *arXiv preprint arXiv:1706.03825*, 2017.

[23] Alvaro Sanchez-Gonzalez, Jonathan Godwin, Tobias Pfaff, Rex Ying, Jure Leskovec, and Peter W. Battaglia. Learning to Simulate Complex Physics with Graph Networks. *arXiv:2002.09405 [physics, stat]*, February 2020. arXiv: 2002.09405.

[24] Raphael Gontijo Lopes, Dong Yin, Ben Poole, Justin Gilmer, and Ekin D. Cubuk. Improving Robustness Without Sacrificing Accuracy with Patch Gaussian Augmentation. June 2019.

[25] Dong Yin, Raphael Gontijo Lopes, Jonathon Shlens, Ekin D. Cubuk, and Justin Gilmer. A Fourier Perspective on Model Robustness in Computer Vision. *arXiv:1906.08988 [cs, stat]*, October 2019. arXiv: 1906.08988.

[26] Andriy O. Lyakhov, Artem R. Oganov, Harold T. Stokes, and Qiang Zhu. New developments in evolutionary structure prediction algorithm USPEX. *Computer Physics Communications*, 184(4):1172–1182, April 2013.

[27] Artem R. Oganov, Andriy O. Lyakhov, and Mario Valle. How Evolutionary Crystal Structure Prediction Works—and Why. *Accounts of Chemical Research*, 44(3):227–237, March 2011.

[28] William W. Tipton and Richard G. Hennig. A grand canonical genetic algorithm for the prediction of multi-component phase diagrams and testing of empirical potentials. *Journal of Physics: Condensed Matter*, 25(49):495401, 2013.

[29] J. Christian Schön and Martin Jansen. First Step Towards Planning of Syntheses in Solid-State Chemistry: Determination of Promising Structure Candidates by Global Optimization. *Angewandte Chemie International Edition in English*, 35(12):1286–1304, 1996. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/anie.199612861.

[30] David J. Wales and Harold A. Scheraga. Global Optimization of Clusters, Crystals, and Biomolecules. *Science*, 285(5432):1368–1372, August 1999. Publisher: American Association for the Advancement of Science Section: Review.

[31] Maximilian Amsler and Stefan Goedecker. Crystal structure prediction using the minima hopping method. *The Journal of Chemical Physics*, 133(22):224104, December 2010. Publisher: American Institute of Physics.

[32] Appendix: First Blind Test of Inorganic Crystal Structure Prediction Methods. In *Modern Methods of Crystal Structure Prediction*, pages 223–231. John Wiley & Sons, Ltd, 2010. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/9783527632831.app1.

[33] Lusann Yang and Gerbrand Ceder. Data-mined similarity function between material compositions. *Physical Review B*, 88(22):224107, December 2013. Publisher: American Physical Society.

[34] Geoffroy Hautier, Chris Fischer, Virginie Ehrlacher, Anubhav Jain, and Gerbrand Ceder. Data Mined Ionic Substitutions for the Discovery of New Compounds. *Inorganic Chemistry*, 50(2):656–663, January 2011. Publisher: American Chemical Society.

[35] Yunxing Zuo, Chi Chen, Xiangguo Li, Zhi Deng, Yiming Chen, Jörg Behler, Gábor Csányi, Alexander V. Shapeev, Aidan P. Thompson, Mitchell A. Wood, and Shyue Ping Ong. Performance and Cost Assessment of Machine Learning Interatomic Potentials. *The Journal of Physical Chemistry A*, 124(4):731–745, January 2020. Publisher: American Chemical Society.

[36] Tian Xie and Jeffrey C. Grossman. Crystal Graph Convolutional Neural Networks for an Accurate and Interpretable Prediction of Material Properties. *Physical Review Letters*, 120(14):145301, April 2018.

[37] K. T. Schütt, H. E. Sauceda, P.-J. Kindermans, A. Tkatchenko, and K.-R. Müller. SchNet – A deep learning architecture for molecules and materials. *The Journal of Chemical Physics*, 148(24):241722, March 2018. Publisher: American Institute of Physics.

[38] Chi Chen, Weike Ye, Yunxing Zuo, Chen Zheng, and Shyue Ping Ong. Graph Networks as a Universal Machine Learning Framework for Molecules and Crystals. *Chemistry of Materials*, 31(9):3564–3572, May 2019. Publisher: American Chemical Society.

[39] Justin Gilmer, Samuel S. Schoenholz, Patrick F. Riley, Oriol Vinyals, and George E. Dahl. Neural Message Passing for Quantum Chemistry. *arXiv:1704.01212 [cs]*, April 2017. arXiv: 1704.01212.

[40] Nathaniel Thomas, Tess Smidt, Steven Kearnes, Lusann Yang, Li Li, Kai Kohlhoff, and Patrick Riley. Tensor field networks: Rotation- and translation-equivariant neural networks for 3D point clouds. *arXiv:1802.08219 [cs]*, February 2018. arXiv: 1802.08219.

[41] Benjamin Ummenhofer, Lukas Prantl, Nils Thuerey, and Vladlen Koltun. Lagrangian Fluid Simulation with Continuous Convolutions. September 2019.

[42] Volker L. Deringer, Chris J. Pickard, and Davide M. Proserpio. Hierarchically Structured Allotropes of Phosphorus from Data-Driven Exploration. *Angewandte Chemie International Edition*, 59(37):15880–15885, 2020. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/anie.202005031.

[43] Berk Onat, Ekin D. Cubuk, Brad D. Malone, and Efthimios Kaxiras. Implanted neural network potentials: Application to Li-Si alloys. *Physical Review B*, 97(9):094106, March 2018. Publisher: American Physical Society.

[44] Nongnuch Artrith, Alexander Urban, and Gerbrand Ceder. Constructing first-principles phase diagrams of amorphous LixSi using machine-learning-assisted sampling with an evolutionary algorithm. *The Journal of Chemical Physics*, 148(24):241711, June 2018. arXiv: 1802.03548.

[45] Nan Xu, Yao Shi, Yi He, and Qing Shao. A Deep-Learning Potential for Crystalline and Amorphous Li–Si Alloys. *The Journal of Physical Chemistry C*, 124(30):16278–16288, July 2020.

[46] William W. Tipton, Clive R. Bealing, Kiran Mathew, and Richard G. Hennig. Structures, phase stabilities, and electrical potentials of Li-Si battery anode materials. *Physical Review B*, 87(18):184114, May 2013. Publisher: American Physical Society.

[47] Zhiwei Cui, Feng Gao, Zhihua Cui, and Jianmin Qu. A second nearest-neighbor embedded atom method interatomic potential for Li–Si alloys. *Journal of Power Sources*, 207:150–159, June 2012.

[48] Andrew J. Morris, C. P. Grey, and Chris J. Pickard. Thermodynamically stable lithium silicides and germanides from density functional theory calculations. *Physical Review B*, 90(5):054111, August 2014. Publisher: American Physical Society.

[49] Ekin D. Cubuk, Barret Zoph, Dandelion Mane, Vijay Vasudevan, and Quoc V. Le. AutoAugment: Learning Augmentation Policies from Data. *arXiv:1805.09501 [cs, stat]*, April 2019. arXiv: 1805.09501.

[50] Andrew Slavin Ross and Finale Doshi-Velez. Improving the Adversarial Robustness and Interpretability of Deep Neural Networks by Regularizing their Input Gradients. *arXiv:1711.09404 [cs]*, November 2017. arXiv: 1711.09404.

[51] Nicholas Carlini, Anish Athalye, Nicolas Papernot, Wieland Brendel, Jonas Rauber, Dimitris Tsipras, Ian Goodfellow, Aleksander Madry, and Alexey Kurakin. On Evaluating Adversarial Robustness. *arXiv:1902.06705 [cs, stat]*, February 2019. arXiv: 1902.06705.

[52] Aleksander Madry, Aleksandar Makelov, Ludwig Schmidt, Dimitris Tsipras, and Adrian Vladu. Towards Deep Learning Models Resistant to Adversarial Attacks. *arXiv:1706.06083 [cs, stat]*, September 2019. arXiv: 1706.06083.

[53] Ekin D Cubuk and Samuel S Schoenholz. Adversarial forces of physical models. In *NeurIPS Machine Learning for the Physical Sciences Workshop*, 2020.

[54] G Kresse and J Furhmuller. Software VASP, Vienna (1999); G. Kresse, J. Hafner. *Phys. Rev. B*, 47:R558, 1993.

[55] Shyue Ping Ong, William Davidson Richards, Anubhav Jain, Geoffroy Hautier, Michael Kocher, Shreyas Cholia, Dan Gunter, Vincent L. Chevrier, Kristin A. Persson, and Gerbrand Ceder. Python Materials Genomics (pymatgen): A robust, open-source python library for materials analysis. *Computational Materials Science*, 68:314–319, February 2013.

[56] `https://github.com/materialsproject/pymatgen/blob/v2020.9.14/pymatgen/analysis/structure_matcher.py#L291-L1158`. Accessed: 2020-09-23.

[57] Oriol Vinyals, Samy Bengio, and Manjunath Kudlur. Order matters: Sequence to sequence for sets. *arXiv preprint arXiv:1511.06391*, 2015.

[58] Zhiwei Cui, Feng Gao, Zhihua Cui, and Jianmin Qu. A second nearest-neighbor embedded atom method interatomic potential for Li–Si alloys. *Journal of Power Sources*, 207:150–159, June 2012.

[59] Nic Ford, Justin Gilmer, Nicolas Carlini, and Dogus Cubuk. Adversarial Examples Are a Natural Consequence of Test Error in Noise. *arXiv:1901.10513 [cs, stat]*, January 2019. arXiv: 1901.10513.

[60] Valentina Zantedeschi, Maria-Irina Nicolae, and Ambrish Rawat. Efficient Defenses Against Adversarial Attacks. In *Proceedings of the 10th ACM Workshop on Artificial Intelligence and Security*, AISec '17, pages 39–49, New York, NY, USA, November 2017. Association for Computing Machinery.

[61] Artem R. Oganov, Chris J. Pickard, Qiang Zhu, and Richard J. Needs. Structure prediction drives materials discovery. *Nature Reviews Materials*, 4(5):331–348, May 2019. Number: 5 Publisher: Nature Publishing Group.

[62] P. E. Blöchl. Projector augmented-wave method. *Physical Review B*, 50(24):17953–17979, December 1994.

[63] John P. Perdew, Kieron Burke, and Matthias Ernzerhof. Generalized Gradient Approximation Made Simple. *Physical Review Letters*, 77(18):3865–3868, October 1996.

7 Appendix

7.1 Estimation of number of configurations in unit cells

The number of configurations for placing N atoms in a cell of volume V, discretizing the cell by length δ, is $\frac{1}{(V/\delta^3)}\frac{(V/\delta^3)!}{(V/\delta^3-N)!N!}$ [61]. For N=10, V=10Å^3 and δ=0.1Å, this number is 2.7×10^{25}

7.2 VASP simulation details

Random structures. We used AIRSS to generate random structures with up to 6 symmetry operations, and vary the number of formula units in the cell up to 38 atoms in the cell. We estimate the volume per element from the experimental data of Li-Si structures in the Materials Project, and set the target volume of AIRSS structure outputs to be within 80% 150% of the estimated volume. We set minimum separation between atoms to be 2.28 Å, as small values of atomic separation may cause forces to diverge. The random structures were relaxed using Kohn-Sham density functional theory with the projector augmented-wave method[62] as implemented in VASP. Relaxations were performed with steepest descent algorithm (IBRION=3, SMASS=2, POTIM=0.4) to relax both the ions and the unit cell(ISIF=3). The calculations were converged to energy convergence of 10^{-6} eV and force convergence of 10^{-2} eV/Åwith self-consistent, periodic DFT. A plane wave basis set with the kinetic energy cutoff of 520 eV was used, and Γ-centered k-point mesh with spacing 0.25 Å^{-1}

was used to sample the Brillouin zone. Generalized gradient approximation Perdue-Burke-Ernzerhof (PBE)[63] functional was used to treat the exchange-correlation energy.

Molecular Dynamics. We used the structures from Materials Project flagged as having an experimentally characterized structure, and created 64 to 78-atom supercells of the conventional cells of these structures to ensure the unit cells are sufficiently large for molecular dynamics. ($Li_{15}Si_4$: mp-569849, 76 atoms. $Li_{13}Si_4$: mp-672287, 68 atoms. Li_2Si: mp-27705, 72 atoms. Li_7Si_2: mp-27930, 72 atoms. LiSi: mp-795, 64 atoms) The simulations were run at a constant temperature of 800K with 2 femtosecond time steps with a single k-point at Gamma. All other details are the same as random structures.

7.3 Distribution of random structure relaxations data and MD data

In the experiments in the main text, models were trained on all stoichiometries except for $Li_{15}Si_4$, which was held out for generalization experiments. The distributions below are also plotted excluding $Li_{15}Si_4$. We see that the random structures have higher stresses than MD data, which is expected from random structures. The forces for random structures, however, are smaller than MD data. We enforced a moderately large minimum separation between atoms for random structure generation to ensure that none of the random structures diverged during relaxations, and moved atoms in the direction of forces during the relaxations. This led to a narrower distribution of forces in the random structure dataset.

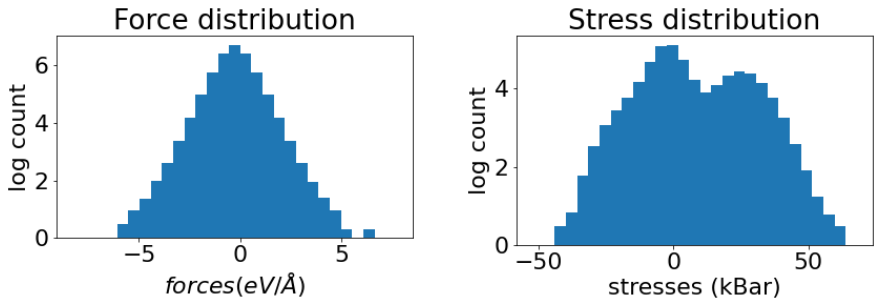

Figure 1: Force(left) and stress(right) distribution of MD data.

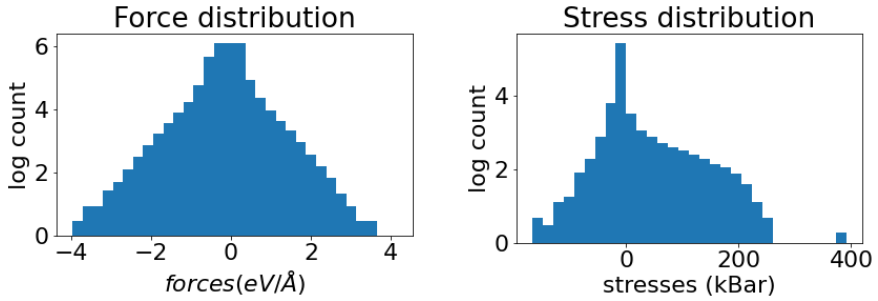

Figure 2: Force(left) and stress(right) distribution of random structure relaxations data.

7.4 Architectures for stress calculations

We investigated three different architectures below:

Stress tensor outputs as an global outputs. The 6 independent components of the stress tensor($\sigma_{xx}, \sigma_{yy}, \sigma_{zz}, \sigma_{xy}, \sigma_{yz}, \sigma_{zx}$) are predicted from the graph. The node and edge states are aggregated with set2set[57] and concatenated with the global state, and passed through two MLP layers of sizes 64 and 32. The final predictions are produced by an output layer without activation.

Table 4: **Details of the random structure dataset**. The first two rows correspond to the components of force vectors and stress tensors. Max |force| refers to the maximum force magnitude in each structure. Max stress component is the maximum value of the stress tensor components $(\sigma_{xx}, \sigma_{yy}, \sigma_{zz}, \sigma_{xy}, \sigma_{yz}, \sigma_{zx})$ in each structure.

	mean	standard deviation	min	max
Force (eV/Å)	0.000	0.208	-4.797	4.797
Stress (kBar)	0.733	12.318	-192.012	439.783
max |force| (eV/Å)	0.453	0.442	0.000	4.797
max stress component (kBar)	8.662	23.555	0.000	439.783

For this model, the distance from each atom to (100), (010) and (001) planes were concatenated to the node feature inputs.

Stress tensor outputs from fictitious atoms on cell corners. We put fictitious non-interacting atoms on the corners of the cells, and predict the stress tensor components from the fictitious atoms. Each fictitious atom outputs three stress tensor components - e.g. the atom on the x-axis outputs $\sigma_{xx}, \sigma_{yx}, \sigma_{zx}$.

Stress tensor outputs from fictitious atoms on cell planes. We put fictitious non-interacting atoms on (100), (010) and (001) planes, and use distances from the planes to each atom for edge features.

Table 5 shows that stress predictions from fictitious atoms on cell planes and as global output perform similarly on force and stress MAE, but fictitious atoms on cell corners performs poorly on stress MAE. These two models outperform MEAM in force predictions by an order of magnitude difference and by a factor of three in stress MAE. Moreover, fictitious atoms on cell planes is the best at predicting the relaxation trajectory, with 77% improvement over the MEAM baseline. We used fictitious atoms on cell planes for all experiments in the main text.

Table 5: **Methods for predicting cell stress.** Corner and plane refer to stress outputs from putting fictitious atoms on cell corners and cell planes, and MEAM is the model from ref.[58]. Force and stress units are eV/Åand kBar. *: Fraction of relaxed final structures that match VASP **: Fraction of matching structures in each trajectory

Method	Force MAE	Stress MAE	Relaxed structure match*	Matches in trajectory**
corner	0.03	3.09	0.15	0.34
planes	**0.03**	1.05	**0.41**	**0.66**
global output	0.04	**0.96**	0.36	0.61
MEAM	0.55	9.11	0.23	0.39

7.5 Energy predictions on random structure data

The model trained on random structure energies achieved a MAE of 15 meV/atom. This is higher than other machine learning potentials on the Li-Si systems, such as [44, 43, 45]. As random structure data is quite different from what [44, 43, 45] are trained on, we trained and tested the same model on our MD data. The test MAE for MD energies is 3.8 eV/atom, similar to what's reported in these works.

7.6 Relaxation algorithm affects noise in relaxation data

We used the steepest descent algorithm as implemented in VASP (IBRION=3, SMASS=2, ISIF=3) and adjusted the time step parameter (POTIM) to a small value(0.4, default is 0.5) and checked all relaxations have a monotonic decrease in energy over ionic steps. But in some cases, we see large upward spikes in force or stress trajectory, as in Figure 3. To investigate whether these spikes introduce noise in our simulations, we quantified the amount of upward spikes in the relaxation trajectories that are successfully simulated by our model and those that fail. We calculated the maximum amount of increase in forces (maximum force applied to the atoms in the structure) and stresses (maximum component of the stress tensor) within 5 ionic steps for Li_2Si structures in the test set. We plot the distribution of those spikes in Figure 4. We see that for both forces and stresses, the

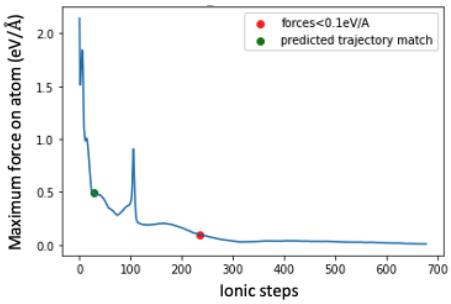

Figure 3: Sample trajectory of Li2Si in which GNN model did not reproduce the VASP relaxation trajectory. The red dot corresponds to where the maximum force on the atoms reach 0.1 eV/Åand the simulation terminates. The green dot is the ionic step until the GNN simulation matches VASP. There are large spikes in atomic forces in this trajectory.

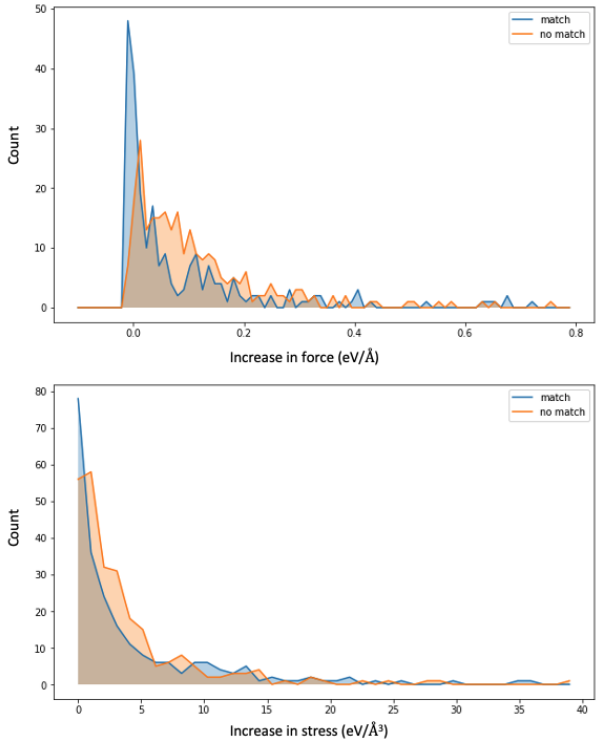

Figure 4: The distribution of upward spikes in relaxation trajectory for maximum forces on atoms (top) and maximum component of stress tensor (bottom). The trajectories where our graph network simulations match the VASP trajectory are plotted in blue(match), and the trajectories where our simulations do not match are plotted in yellow(no match).

distribution of spikes is generally larger for the trajectories that our model fails to simulate correctly. The distribution suggests that some of our model's failures may attributed to unphysical behavior in DFT simulations with VASP used to construct the ground-truth data.