Accelerated discovery of inorganic materials with the Bayesian optimisation-driven exploration of combinatorial space

A. Vasylenko

Department of Chemistry, University of Liverpool, Crown Street L6 97ZD, UK

Introduction

Discovery of materials requires systematic exploration of the materials' combinatorial space[1,2]. Theoretically vast, the combinatorial space of stable materials is reduced by limitations of the synthetically accessible formation energies. Formation energies of hypothetical compositions in the combinatorial space shape its energy profile, which can be explored by sampling. Computationally, the energy of a composition can be estimated by first, predicting its crystal structure[3,4] and then, calculating its formation energy with respect to the reported compounds along possible decomposition routes. Such calculations performed at the density-functional theory (DFT)[5] level of accuracy prove a reliable guide to synthetic chemists[6,7]. At this level of precision, the structures can be differentiated with a resolution of 1 meV/atom, however extensive combinatorial exploration with a dense sampling of the combinatorial space bares computationally intractable cost. The faster approaches for energy estimation via, e.g., interatomic force fields, are not accessible for the majority of the prospective materials[8–10].

Recent developments of machine learning models for prediction of energy for the material's compositions without knowledge of their structures offer a potential solution to the combinatorial challenge[11–13], but the present state-of-the-art methods allow resolution between different compositions with an error of up to 100 meV/atom in comparison to the

DFT methods, disabling their reliable incorporation into exploratory experimental workflows.

In this work, we demonstrate that exploration of the compositional space can be presented as a global optimisation problem. In this approach, on-the-convex-hull composition with 0 eV/atom formation energy represents a synthetically accessible stable phase and is an optimum. Hence, an energy profile of a material's combinatorial space can be approximated with Gaussian process (GP) and optimised via Bayesian optimisation (BO)[14,15]. BO has been proven an effective algorithm for the exploration of costly-to-evaluate and the black-box functions, it has been increasingly used in materials science to design experiments and optimise sampling[16–19]. Here, we demonstrate that by treating the energy profile of the combinatorial space as a function, evaluation of which at any point of combinatorial space with DFT-based crystal structure prediction (CSP) is indeed a costly exercise, one can employ the BO-based algorithm PhaseBO to update the information about the combinatorial space, suggest candidate compositions for CSP and discover the global minimum corresponding to new stable material. In the example of the previously studied Li-Sn-S-Cl combinatorial space[20], PhaseBO discovers the experimentally stable composition more consistently and faster in comparison to the conventionally used random sampling. This approach offers a significant acceleration of the computationally-driven materials discovery without compromising the accuracy of energy evaluation.

Results and discussion

Approximation of the energy profile of compositional space with a function has proved effective in discovery of materials within the probe structure calculation approach[6,20]. The approach stems from the empirical observations that stoichiometrically close compounds have similar enthalpies of formation. This enables the computationally guided exploration of

compositional space by first, modelling the structures of the compositions sampled across the phase field, and then synthetically zooming into the regions of low enthalpies of formation. Experimental findings discovered with this approach in the range of chemistries suggest the energetics of a compositional space can be modelled as a function of stoichiometric ratios. However, exhaustive evaluation of this function is an unsurmountable challenge: its assessment of any point involves expensive CSP calculations, which due to the steep DFT scaling with the number of atoms, are tractable only for a limited number of constituent elements and their stoichiometries.

Bayesian optimisation is a class of machine learning methods aiming to find the global optimum in the problem

$$\max_{x \in \mathbb{R}^d} f(x), \tag{1}$$

where the analytical form of the objective function f in d-dimensional space is unknown and its evaluation for any point x is expensive. Bayesian optimisation has efficiently found the optima for the range of problems with different dimensionality and has been demonstrated effective for $d{\sim}100$ [21]. Inspired by the success of the probe structure approach, we represent the search for stable compositions in the materials phase fields as the problem (1). In this formulation, we search for the minima of formation enthalpy, presented as a function of stoichiometry in multi-element space, in which feasible values of x are stoichiometric coefficients that form a d-dimensional simplex $\{x \in \mathbb{R}^d : \sum_i x_i = 1\}$. To find the minima, we employ Bayesian optimisation, illustrated in the example quaternary field in Figure 1.

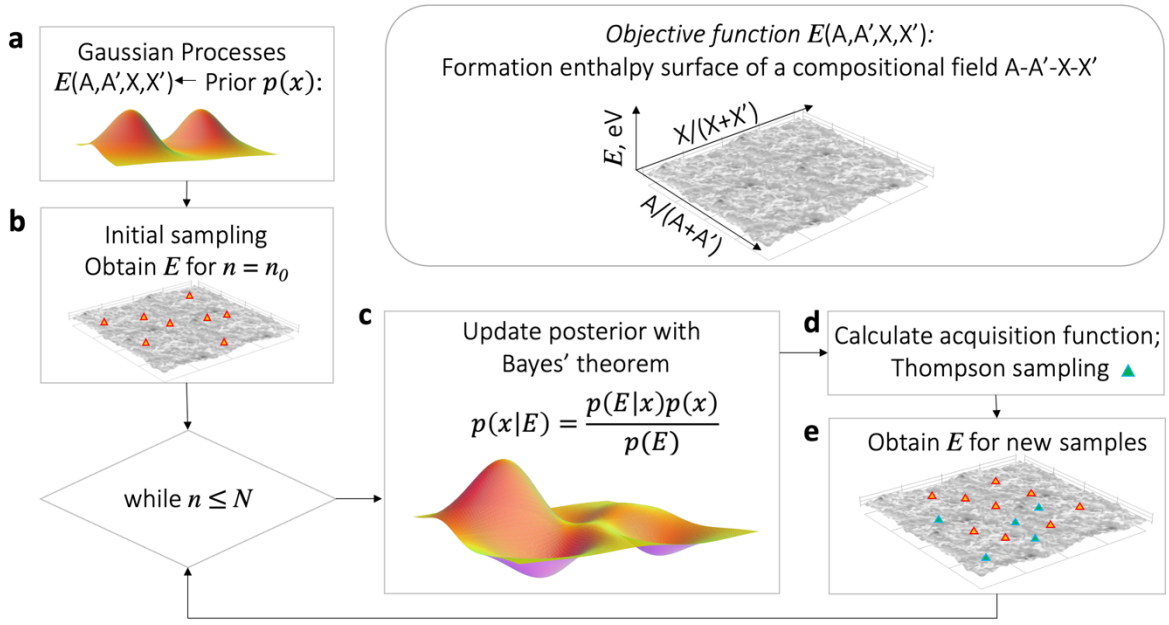


Figure 1. Schematics of the search for stable compositions – minima of formation enthalpy E – in a compositional field A-A'-X-X' with Bayesian optimisation. a The true dependency of formation enthalpy on a compositional content (stoichiometry) is the objective function E(A, A', X, X'), which is unknown and modelled with a Gaussian processes function $p(x)$ (prior). **b** For the initial sampling, DFT-based CSP is performed for several compositions, for which E is calculated with respect to the reference materials reported in the phase field. **c** The model function $p(x|E)$ (posterior) for formation energy is updated with the obtained data (E for sample compositions) according to the Bayes' theorem. **d** The surrogate function is built in the form of the Expected Improvement acquisition function[22]; it is sampled with Thompson sampling[23] to suggest the compositions for the next iteration assessment of E with DFT-based CSP (at stage **e**). Stages **c-e** are repeated while the computational budget allows.

The search for stable compositions and corresponding minima of enthalpy of formation in the combinatorial space, e.g., quaternary A – A' – X – X', starts with the approximation of the unknown function of enthalpy E(A, A', X, X') with a Gaussian process prior. In this approximation, the enthalpy dependence on a stoichiometric coefficient of each constituent chemical element is initialised with a random Gaussian distribution, and the distributions form a multivariate normal (Fig 1 a). In Bayesian optimisation approach, information about

the objective function is obtained iteratively: after each sampling – CSP for selected compositions, with subsequent evaluation of the energy of formation (Fig 1 b, e) – the posterior probability distribution of the function is updated via GP regression and according to the Bayes' theorem[14] (Fig 1 c). From GP regression and the posterior, one can estimate energy and uncertainties for the unexplored compositions, and construct a model surrogate function for selecting the best sampling points – the acquisition function (Fig 1 d). The latter can be derived in different forms from the posterior[24] and incorporate a strategy for exploration and exploitation during the search; in our approach, we employ the commonly used form of Expected Improvement[22]. The acquisition function is simple to optimise with, e.g., gradient-descent methods, to obtain the minimum suggesting the next material composition for evaluation with CSP. Alternatively, the acquisition function can be sampled with, e.g., Thompson sampling[23] to perform a batch evaluation. In our approach, we employ the latter to benefit from the high-throughput computational capabilities. The process of sampling the compositional space, evaluation of energy of candidates with traditional CSP and DFT methods, and posterior update is repeated until stopping criteria are satisfied. For the latter, one can choose a local or a global minimum or a number of costly CSP evaluations, defined by a computation budget.

To validate this approach we compare the explorations of a compositional space with a conventional random sampling and PhaseBO on the example quaternary Li-Sn-S-Cl, which was previously extensively studied with DFT-based CSP; this particular example is interesting because of the complexity of its formation energy landscape (See Fig. 2 a), where several compositions with 0 meV/atom were discovered, one of which was verified experimentally[20].

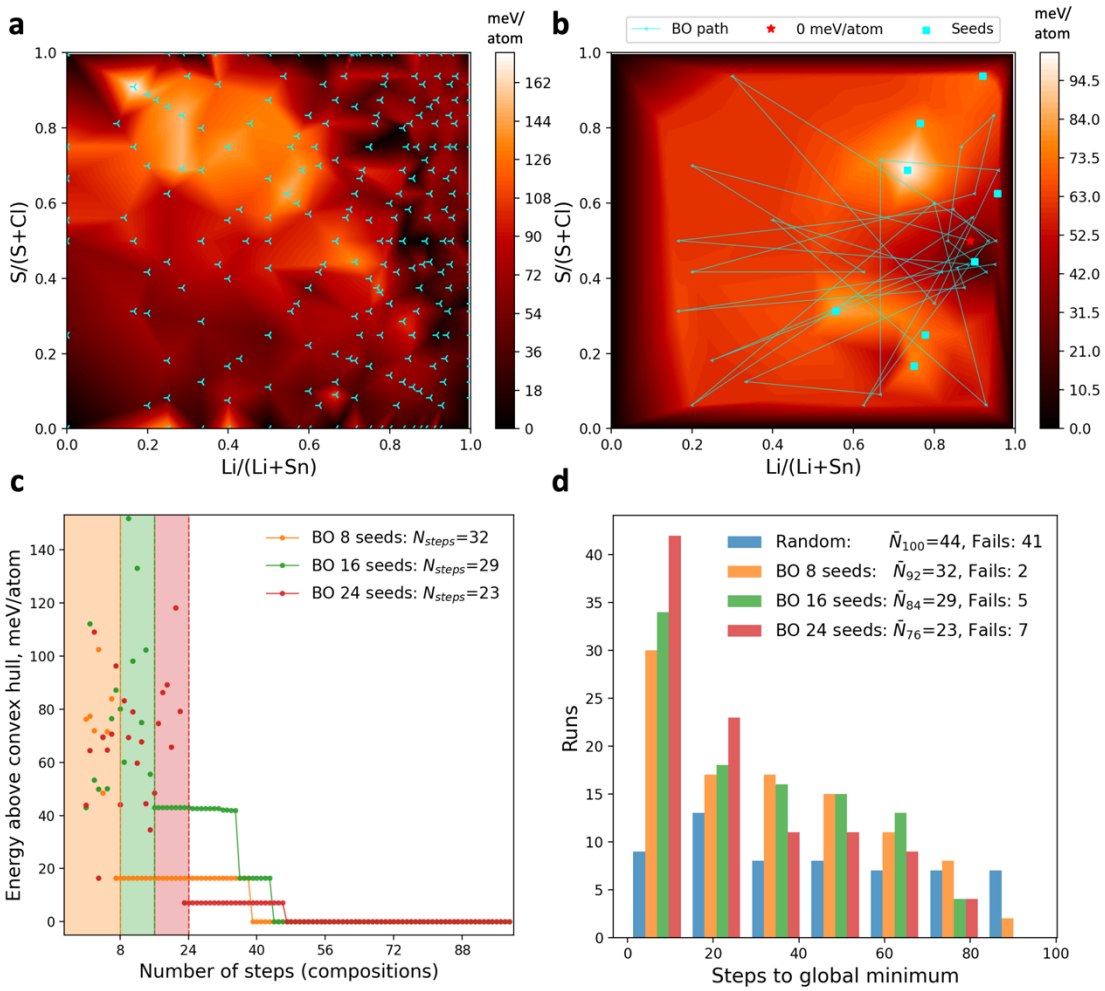


Figure 2. Exploration of Li-Sn-S-Cl phase field. a Energy of formation of 195 compositions $Li_xSn_{x'}S_yCl_{y'}$ presented as a surface in 2 dimensions: $\tilde{x}$ = x/(x+x') and $\tilde{y}$ = y/(y+y'); colours are linearly interpolated in between the points calculated with CSP and DFT in[20]. **b** Exploration of the phase field with PhaseBO: starting from 8 random seed compositions (cyan squares), posterior $p(\tilde{x}, \tilde{y}\,|E)$ is calculated (coloured surface); BO path (cyan lines) and next compositions for CSP are suggested by iterative Thompson sampling of the acquisition function built from $p(\tilde{x}, \tilde{y}\,|E)$. The posterior updates after each iteration are omitted for clarity. **c** Progression of discovery of compositions with low energy of formation on the BO paths starting from 8, 16 and 24 seeds illustrated by the paths with 32, 29 and 23 steps to the global minimum, that correspond to the average numbers of steps for such paths respectively. **d** Histogram of the number of explorations (runs) and the numbers of steps in which a global minimum was discovered out of 100 attempts with a random search (blue) and PhaseBO starting with 8, 16 and 24 seeds. With PhaseBO, the random uniform distribution is shifted strongly towards the small number of steps, and the number of failed runs is reduced dramatically.

Exploration with PhaseBO starts with a random selection of seed compositions to construct the initial function of formation energy (See square markers for seeds and colour surface for posterior in Fig 2b). For the seed compositions as well as for all other compositions selected/suggested during the PhaseBO exploration, one has to perform energy evaluation, e.g., based on CSP and DFT calculations. In this case, the energies are known from the previous study. From the posterior, the acquisition function is constructed and can be optimised and sampled to suggest the next compositions for evaluation. Interestingly, the sequentially connected suggestions illustrate a competition between exploration and exploitation (See the lines between cyan markers in Fig. 2b). The posterior is recalculated, and the process is repeated until the global minimum is found. The example path to the global minimum is illustrated in Figure 2b. The paths and their lengths depend on the number of seed compositions and the energy of the seeds as illustrated in Fig. 2c. To consider the statistical significance of these differences and the stochastic nature of BO, we perform 100 attempts for different selections of 8, 16, 24 seeds and compare the optimisation paths with 100 random searches; the budget for all approaches is set to 100 evaluations. The latter has a close to uniform distribution of the number of steps required to find a global minimum, with an average of 44 steps; notably, no global minimum was found in 41 runs of random searches (Fig. 2d). Phase BO demonstrates distributions considerably shifted to the smaller number of steps for all 3 approaches. The average number of steps to a global minimum is decreased in comparison to random searches, and the number of failures is reduced significantly. The higher number of failures for the 24-seeds approach may be associated with the smaller number of iterations within the budget: 100 – 24 = 76 iterations (cf. 92 iterations for 8-seeds approach). At the same time, 24 seeds-based

calculation of the posterior results in the 10-step location of a global minimum in more than 40% of time. In probe structure approach, compositions with calculated formation energy below thermal energy kT per atom (25.9 meV/atom) are promising candidates for synthetic investigation. PhaseBO can also help to identify more of these promising low-energy compositions in comparison to a random search (Supplementary Fig. 1).

We apply PhaseBO to study the unexplored compositional field, Li-Si-S-Cl, which does not contain any reported quaternary phases[1], yet ranked as the first candidate to likely contain stable materials[20]. We repeat the process described above in Fig. 1,2, starting from 8 random compositions as initial seeds. We evaluate the energy of these compositions by first predicting their structures with a CSP code XtalOpt[25] coupled with energy calculations with VASP[26], and then comparing the energies of the predicted structures with the reference compositions reported in the phase field, thereby constructing a 2-dimensional convex hull surface. For the obtained formation energies, the posterior function is calculated (Figure 3a). We repeat this process by sampling the acquisition function with Thompson sampling of 4 compositions at each iteration (Figure 3b-h).

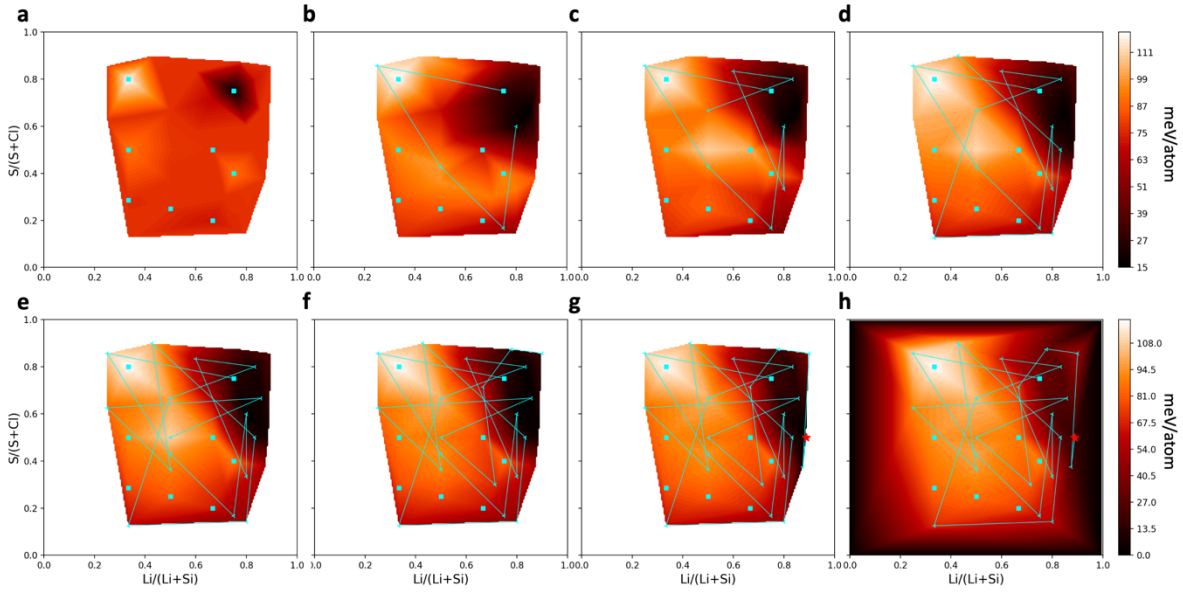


Figure 3. Exploration of Li-Si-S-Cl phase space with PhaseBO. a Formation energy estimation (posterior), calculated based on the 8 randomly sampled compositions (cyan squares). **b-g** Stages of BO process, in which at each stage, 4 compositions sampled with Thompson sampling (cyan dots) are assessed with CSP and posterior is updated. **h** Energy estimation with all compositions explored in the previous stages connected sequentially with a line; the last point of the search is the lowest energy composition found (red star).

In 6 iterations, we find 4 compositions with the energy of formation below kT/atom; the resulting convex hull surface is illustrated in Figure 4. At the last step of the BO path, the lowest energy, 15.8 meV/atom, was found at $Li_8Si_1S_4Cl_4$ composition, which is predicted to form a modified fcc structure with 2 Li ions occupying tetrahedral interstitials (Figure 4b). The predicted structure could potentially be further improved (its energy reduced) via modification of the CSP approach, longer CSP runs and larger crystal models, which is beyond the scope of the present study. The full list of the discovered compositions and their formation energies as well as the reference materials are listed in Supplementary Table 1. Notably, the surface of formation energy in Figure 4a correlates well with the energy estimated from the posterior in Figure 3h, suggesting the promising areas of stoichiometries, where stable materials may be found. Considering the typical challenges for the CSP and DFT methods – characteristic disorder for the mixed-anions materials and temperature-associated and kinetic effects – rapid identification of the compositional regions of low energy (Figure 3h, Figure 4a) is imperative for computational guidance towards synthetically accessible materials.

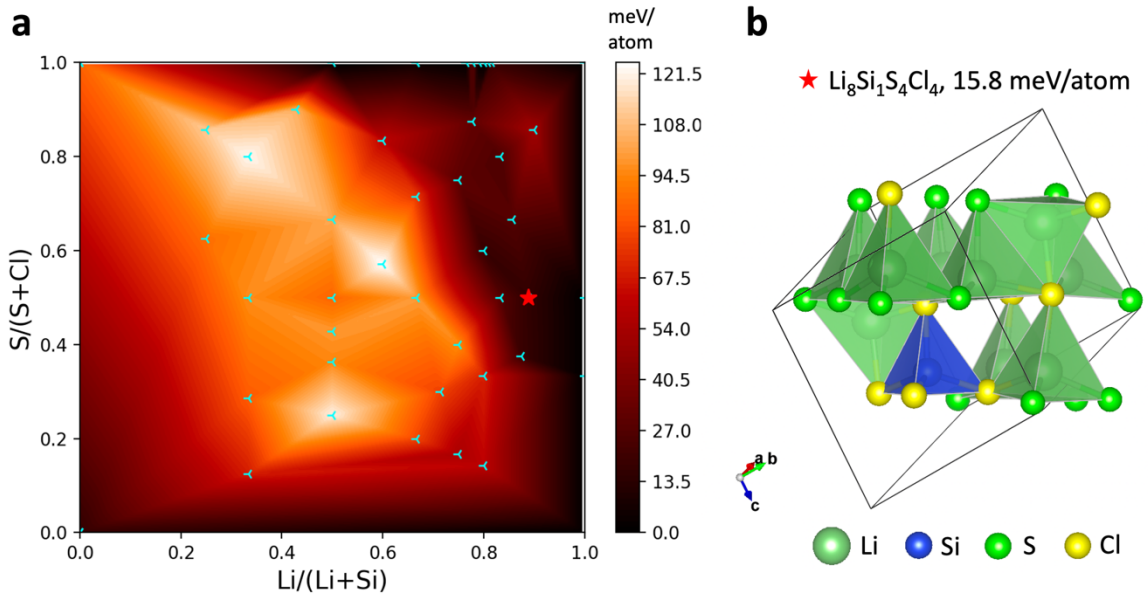


Figure 4. Exploration of Li-Si-S-Cl. a Formation energy surface in the compositional space, discovered in 6 iterations with PhaseBO. A red star highlights the lowest-energy composition. **b** Predicted structure of the lowest-energy composition discovered in **a**.

Methods

PhaseBO is built using the open-source libraries Numpy[27], Pandas[28], Pymatgen[29], GpyOpt[30] and Matplotlib[31].

In CSP with XtalOpt[25], each composition was initialized with a random structure and up to nine evolutional generations were considered with 50 mutated structures in each. The generations were created by mutations of a structure as well as by combining two-parent structures into a new structure. Mutations are direct transformations of the crystal structures—crossover, strain, nonlinear “ripple”, exchange (atomic swaps), and their combinations.

Calculations for structure prediction were based on energy calculations after geometry optimization for reference and probe structures that were performed in VASP-5.4.4[26] with PAW pseudopotentials[32], a 700 eV kinetic energy cutoff for plane waves, and adaptive k-points sampling with k-spacing 0.025. 1e–10 eV threshold for total energy convergence in self-consistent runs, and 0.001 eV/Å threshold for convergence of forces were used for all computations.

Acknowledgement

We thank the UK Engineering and Physical Sciences Research Council (EPSRC) for funding through grants number EP/N004884 and EP/V026887.

We thank Dr Chris Collins and Dr Andy Zeng for testing the software and for suggestions on the user-friendly features of the PhaseBO.

Code availability

The software developed for this study is available at

https://www.github.com/lrcfmd/PhaseBO

Data availability

The data used in this study is available at https://www.github.com/lrcfmd/PhaseBO and available via the University of Liverpool data repository at

https://doi.org/yyy/datacat.liverpool.ac.uk/xxx

Competing Interests Statement

The authors declare there are no competing interests.

References

1. Zagorac, D., Müller, H., Ruehl, S., Zagorac, J. & Rehme, S. Recent developments in the Inorganic Crystal Structure Database: theoretical crystal structure data and related features. *J. Appl. Crystallogr.* **52**, 918–925 (2019).

2. Tabor, D. P. *et al.* Accelerating the discovery of materials for clean energy in the era of smart automation. *Nat. Rev. Mater.* **3**, 5–20 (2018).
3. Wang, Y. & Ma, Y. Perspective: Crystal structure prediction at high pressures. *J. Chem. Phys.* **140**, 040901 (2014).
4. Oganov, A. R., Pickard, C. J., Zhu, Q. & Needs, R. J. Structure prediction drives materials discovery. *Nat. Rev. Mater.* **4**, 331–348 (2019).
5. Hasnip, P. J. *et al.* Density functional theory in the solid state. *Philos. Trans. R. Soc. Math. Phys. Eng. Sci.* **372**, 20130270 (2014).
6. Collins, C. *et al.* Accelerated discovery of two crystal structure types in a complex inorganic phase field. *Nature* **546**, 280–284 (2017).
7. Vasylenko, A. *et al.* Electronic Structure Control of Sub-nanometer 1D SnTe via Nanostructuring within Single-Walled Carbon Nanotubes. *ACS Nano* **12**, 6023–6031 (2018).
8. Daw, M. S., Foiles, S. M. & Baskes, M. I. The embedded-atom method: a review of theory and applications. *Mater. Sci. Rep.* **9**, 251–310 (1993).
9. Behler, J. Perspective: Machine learning potentials for atomistic simulations. *J. Chem. Phys.* **145**, 170901 (2016).
10. Deringer, V. L. *et al.* Gaussian Process Regression for Materials and Molecules. *Chem. Rev.* **121**, 10073–10141 (2021).
11. Wang, A. Y.-T., Kauwe, S. K., Murdock, R. J. & Sparks, T. D. Compositionally restricted attention-based network for materials property predictions. *Npj Comput. Mater.* **7**, 1–10 (2021).
12. Jha, D. *et al.* ElemNet: Deep Learning the Chemistry of Materials From Only Elemental Composition. *Sci. Rep.* **8**, 17593 (2018).

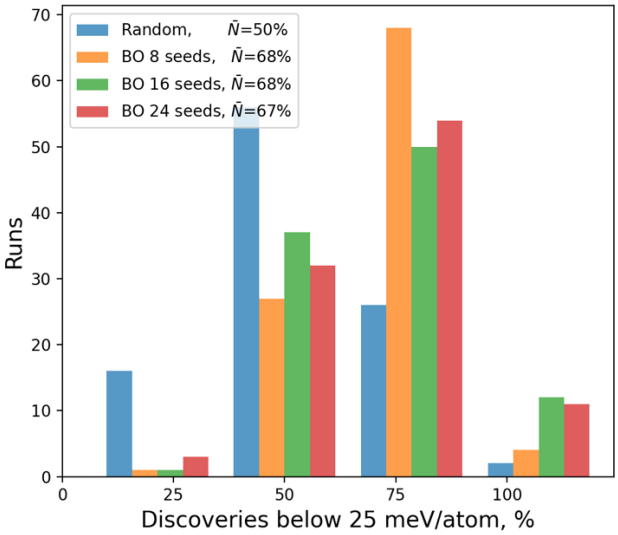
13. Goodall, R. E. A. & Lee, A. A. Predicting materials properties without crystal structure: deep representation learning from stoichiometry. *Nat. Commun.* **11**, 6280 (2020).
14. Rasmussen, C. E. Gaussian Processes in Machine Learning. in *Advanced Lectures on Machine Learning: ML Summer Schools 2003, Canberra, Australia, February 2 - 14, 2003, Tübingen, Germany, August 4 - 16, 2003, Revised Lectures* (eds. Bousquet, O., von Luxburg, U. & Rätsch, G.) 63–71 (Springer, 2004). doi:10.1007/978-3-540-28650-9_4.
15. Shahriari, B., Swersky, K., Wang, Z., Adams, R. P. & de Freitas, N. Taking the Human Out of the Loop: A Review of Bayesian Optimization. *Proc. IEEE* **104**, 148–175 (2016).
16. Lookman, T., Balachandran, P. V., Xue, D. & Yuan, R. Active learning in materials science with emphasis on adaptive sampling using uncertainties for targeted design. *Npj Comput. Mater.* **5**, 1–17 (2019).
17. Kusne, A. G. *et al.* On-the-fly closed-loop materials discovery via Bayesian active learning. *Nat. Commun.* **11**, 5966 (2020).
18. Solomou, A. *et al.* Multi-objective Bayesian materials discovery: Application on the discovery of precipitation strengthened NiTi shape memory alloys through micromechanical modeling. *Mater. Des.* **160**, 810–827 (2018).
19. Talapatra, A. *et al.* Experiment Design Frameworks for Accelerated Discovery of Targeted Materials Across Scales. *Front. Mater.* **6**, (2019).
20. Vasylenko, A. *et al.* Element selection for crystalline inorganic solid discovery guided by unsupervised machine learning of experimentally explored chemistry. *Nat. Commun.* **12**, 5561 (2021).
21. Wang, Z., Gehring, C., Kohli, P. & Jegelka, S. Batched Large-scale Bayesian Optimization in High-dimensional Spaces. *ArXiv170601445 Cs Math Stat* (2018).

22. Mockus, J. Application of Bayesian approach to numerical methods of global and stochastic optimization. *J. Glob. Optim.* **4**, 347–365 (1994).
23. Thompson, W. R. On the Likelihood that One Unknown Probability Exceeds Another in View of the Evidence of Two Samples. *Biometrika* **25**, 285–294 (1933).
24. Frazier, P. I. A Tutorial on Bayesian Optimization. *ArXiv180702811 Cs Math Stat* (2018).
25. Lonie, D. C. & Zurek, E. XtalOpt: An open-source evolutionary algorithm for crystal structure prediction. *Comput. Phys. Commun.* **182**, 372–387 (2011).
26. Kresse, G. & Hafner, J. Ab initio molecular dynamics for liquid metals. *Phys. Rev. B* **47**, 558–561 (1993).
27. Harris, C. R. *et al.* Array programming with NumPy. *Nature* **585**, 357–362 (2020).
28. Pandas developers. pandas-dev/pandas: Pandas. (2020) doi:10.5281/zenodo.3509134.
29. Ong, S. P., Wang, L., Kang, B. & Ceder, G. Li−Fe−P−O2 Phase Diagram from First Principles Calculations. *Chem. Mater.* **20**, 1798–1807 (2008).
30. GpyOpt developers. GPyOpt: A Bayesian Optimization framework in Python. (2016) http://github.com/SheffieldML/GPyOpt.
31. Hunter, J. D. Matplotlib: A 2D Graphics Environment. *Comput. Sci. Eng.* **9**, 90–95 (2007).
32. Kresse, G. & Joubert, D. From ultrasoft pseudopotentials to the projector augmented-wave method. *Phys Rev B* **59**, 1758–1775 (1999).

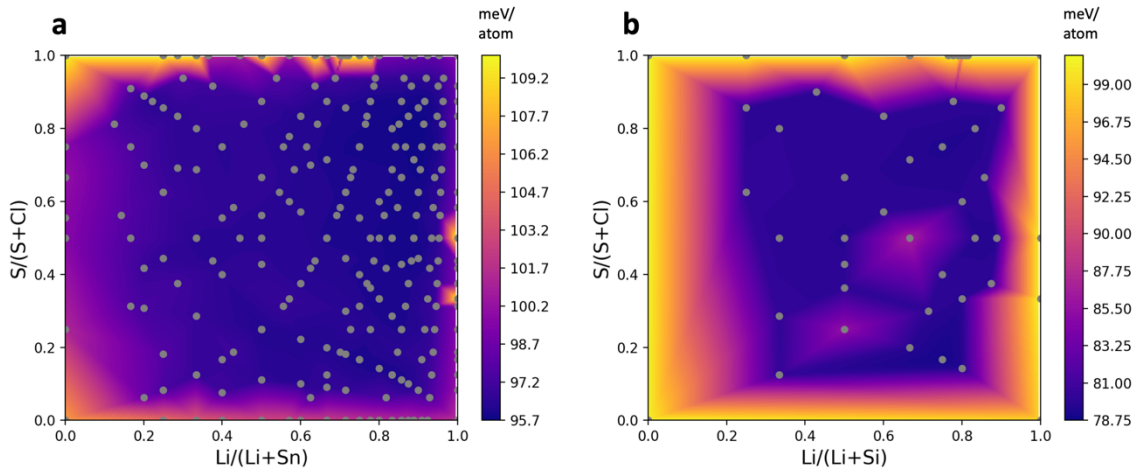
Supplementary Information. Accelerated discovery of inorganic materials with the Bayesian optimisation-driven exploration of combinatorial space

A. Vasylenko

Department of Chemistry, University of Liverpool, Crown Street L6 97ZD, UK



Supplementary Figure 1. Exploration of Li-Sn-S-Cl phase field. Percentage of compositions with formation energy below 25 meV/atom discovered with random sampling in 100 runs is 50% (blue columns); random search finds all 14 compositions discovered in the phase field[1] in only 2% of the runs. PhaseBO with 16 and 24 seeds (green and red columns) finds all low-energy compositions in ≥ 10% of runs and considerably increases the average number of discoveries in all 3 approaches (8, 16, 24 seeds).



Supplementary Figure 2. Final variance in formation energy estimation in BO posterior calculations. a Li-Sn-S-Cl phase field with 195 compositions sampled. b Li-Si-S-Cl phase field with 30 compositions sampled.

Table 1. Reported reference and predicted candidate compositions in Li-Si-S-Cl phase field.

Composition	Energy*, eV	Formation energy**, meV/atom	Composition	Energy*, eV	Formation energy**, meV/atom
References					
Li1	-1.90290	0.000	**S8 Cl8**	-50.42790	0.000
Li2 S1	-11.99520	0.000	**Si4 Cl8**	-48.24850	60.424
Li2 S8	-39.14548	174.389	**Li8 S4**	-45.87944	175.113
S1	-4.12770	0.038	**Si4 S4**	-41.59240	86.165
S8	-33.02160	0.038	**Li2 S8**	-39.14571	174.366
S32	-132.08763	0.000	**Li2 Si6**	-34.97082	276.110
Li88 Si20	-298.52873	17.817	**Si4 S8**	-31.42024	2620.364
Li84 Si20	-292.84133	0.000	**Cl16**	-28.83748	0.000
Si12 Cl32	-173.48595	16.126	**Li4 Si2**	-19.80073	17.627
Li30 Si8	-109.90141	0.000	**Si2**	-10.84942	0.000
Li28 Si8	-104.72411	30.657	**Li1 Cl1**	-7.38460	0.000
Li26 Si8	-101.75413	0.000	**Li4 S6 Si2**	-56.03187	4.399
Si4 Cl16	-76.24834	0.000	**Li8 S8 Si2**	-80.73700	0.000
S8 Cl16	-64.89583	0.000	**S8 Si4**	-62.86461	0.000
Li8 Si8	-61.92346	0.000			
Candidates					
Li8 Si1 S4 Cl4	-69.63889	15.765	**Li2 Si1 S1 Cl4**	-31.52575	79.071
Li4 Si1 S3 Cl2	-42.63356	17.797	**Li1 Si2 S2 Cl5**	-41.32504	83.78
Li7 Si1 S3 Cl5	-64.67214	18.324	**Li3 Si3 S5 Cl5**	-69.59203	86.449
Li5 Si1 S4 Cl1	-47.47466	25.312	**Li1 Si3 S5 Cl3**	-55.14009	88.828
Li5 Si1 S3 Cl3	-49.8703	27.152	**Li3 Si3 S4 Cl7**	-71.03397	94.954
Li6 Si1 S4 Cl2	-54.78154	27.397	**Li3 Si1 S2 Cl3**	-37.00661	95.927
Li3 Si1 S3 Cl1	-35.18889	29.754	**Li1 Si2 S3 Cl3**	-39.62215	96.414
Li7 Si2 S7 Cl1	-75.25513	31.782	**Li2 Si2 S3 Cl4**	-46.78837	98.737
Li9 Si1 S6 Cl1	-70.85402	52.323	**Li2 Si1 S2 Cl2**	-29.78746	99.699
Li3 Si1 S1 Cl5	-38.98812	55.48	**Li1 Si1 S2 Cl1**	-22.58756	102.637
Li4 Si1 S1 Cl6	-46.18979	61.477	**Li3 Si4 S9 Cl1**	-80.79248	104.877
Li3 Si2 S5 Cl1	-50.46359	61.772	**Li1 Si3 S6 Cl1**	-53.30746	111.418
Li1 Si2 S1 Cl7	-43.08071	68.645	**Li1 Si1 S1 Cl3**	-24.0497	120.67
Li4 Si1 S2 Cl4	-44.48871	69.622	**Li1 Si2 S4 Cl1**	-37.83637	122.567

Li4 Si2 S5 Cl2	-57.54086	75.909	**Li3 Si2 S4 Cl3**	-52.0907	124.617
Li5 Si2 S3 Cl7	-68.69978	78.146			

*Total energies are calculated with VASP[2]. **Structures of candidates are predicted with XtalOpt[3].

References

1. Vasylenko, A. *et al.* Element selection for crystalline inorganic solid discovery guided by unsupervised machine learning of experimentally explored chemistry. *Nat. Commun.* **12**, 5561 (2021).
2. Kresse, G. & Hafner, J. Ab initio molecular dynamics for liquid metals. *Phys Rev B* **47**, 558–561 (1993).
3. Lonie, D. C. & Zurek, E. XtalOpt: An open-source evolutionary algorithm for crystal structure prediction. *Comput. Phys. Commun.* **182**, 372–387 (2011).