Materials Graph Transformer predicts the outcomes of inorganic reactions with reliable uncertainties

Shresht A. Malik, Rhys E. A. Goodall, and Alpha A. Lee

Cavendish Laboratory, University of Cambridge, Cambridge CB3 0HE, UK.

A common bottleneck for materials discovery is synthesis. While recent methodological advances have resulted in major improvements in the ability to predicatively design novel materials, researchers often still rely on trial-and-error approaches for determining synthesis procedures. In this work, we develop a model that predicts the major product of solid-state reactions. The cardinal feature of this approach is the construction of fixed-length, learned representations of reactions. Precursors are represented as nodes on a ‘reaction graph’, and message-passing operations between nodes are used to embody the interactions between precursors in the reaction mixture. Through an ablation study, it is shown that this framework not only outperforms less physically-motivated baseline methods but also more reliably assesses the uncertainty in its predictions.

The discovery of new materials and control of their properties is fundamental to the advancement of technology. Adjusting the structure and composition of crystals allows fine-tuning of properties such as conductivity and band gap [1–3]. Advances in first-principles calculations and increased computational power have revolutionised the availability of materials information [4, 5]. Simultaneously, advances in computational materials design [6] have enabled high-throughput prediction and screening in silico [7].

However, due to the challenges associated with material synthesis, only a small proportion of candidates are produced in the laboratory and experimentally validated. The most common route to the formation of polycrystalline materials is solid-state synthesis [8], which typically involves calcining a mixture of solid reactants. Reactions occur through solid-state diffusion of ions, thus reactants are often milled and mixed to improve the reaction kinetics. However, the interplay between thermodynamically driven energy minimisation and kinetic factors results in products which are often difficult to predict; the field lacks the well-understood reaction mechanisms in organic chemistry. Particular procedures or reactants result in the formation of crystals with unique morphologies and compositions as they form through metastable states. Thus researchers frequently rely on a trial-and-error process of synthesis and evaluation. The development of tools to propose and evaluate the most promising synthetic pathways is, therefore, one of the biggest challenges facing materials discovery [9, 10].

A recently released inorganic synthesis dataset has now enabled the possibility of using a data-driven approach to predict the outcomes of inorganic synthesis [11]. To our knowledge, there exists only one study which explores this opportunity [12], which addresses the inverse problem of retrosynthesis – predicting precursors that can react to yield a target product – with an architecture that cannot be readily extended to solve the forward problem. However, save for large scale experimental validation, the accuracy of retrosynthesis models cannot be quantitatively benchmarked because there are almost infinitely many ways to synthesize a material, and reactions reported in the literature are not necessarily even the best synthetic route.

Here, we report an accurate forward reaction prediction model with reliable uncertainties. Methodologically, we leverage ideas from graph representation learning to learn the optimal representation of sets of inorganic reactants directly from data. Our work significantly outperforms baselines, and we show that model uncertainty can be robustly estimated using an ensemble. Our work is a building block towards an inorganic retrosynthesis planner that can interface directly with computational tools to close the materials design-make-test cycle.

DATA AND MODEL

Solid-State Syntheses Dataset

We can define a generalised solid-state synthesis procedure for a target material as a sequence of processes (actions) performed on a set of starting materials (precursors). A recent study has extracted detailed information for over 19,000 such synthesis procedures from academic literature [11]. Each synthesis procedure has four relevant fields for this problem,

1. Target: The stoichiometric formula of the target material for the synthesis.
2. Precursors: Defined as starting materials which explicitly share at least one element with the target material, excluding ‘abundant’ materials, i.e. those found in the air.
3. Processing Actions: The sequence of synthesis actions performed on the precursors, including the
relevant conditions for each action where available.

4. **Balanced Chemical Equation:** Balanced chemical equations for the formation of the target compound provide the relevant molar ratios of precursors in the reaction mixture.

This dataset is the most extensive and highest-fidelity available for inorganic syntheses. Chemical information (precursors, targets and balanced equations) has been extracted with high accuracy (93%).

However, there are some notable limitations that must be considered when utilising it for modelling. Firstly, some procedures have actions and conditions with missing or incorrect data. Over 10% of reactions have no, or only one recorded action. This is primarily because standard procedures are often referenced rather than explicitly stated. The accuracy of the complete procedures (where all actions, conditions, and chemical information are fully correct) is 51%. Further, there is no structural information for the products. This limits physical analysis of the materials and hence a purely stoichiometric approach must be used in this work.

Nevertheless, this dataset provides a wealth of structured information on successful solid-state syntheses. Machine learning (ML) frameworks provide a unique way to assimilate empirical patterns from data. One can thus envisage the target material as being an abstract but learnable function of the precursors and processing procedure.

**Representation of Inorganic Reactions**

ML models generally operate on fixed-length inputs whilst typically our understanding of inorganic reactions is expressed in terms of variable-sized sets of precursors and sequences of processing steps. Bridging this discrepancy is a key challenge that needs to be addressed in the construction of ML algorithms for inorganic reaction prediction.

Here this process is broken down into 2 stages: 1) obtaining fixed-length representations for the precursors and the processing sequence, and 2) performing set regression on this set of precursors to predict the product given the processing sequence.

**Representation of Precursors**

A naïve approach would consider representing materials as sparse vectors with components proportional to the relative amounts of its constituent elements (its stoichiometry). However, such a representation fails to capture critical correlations between different elements. In contrast, Magpie embeddings offer powerful general-purpose descriptors for bulk inorganic materials in the absence of structural information [13]. These are highly engineered, 145-dimensional vector representations which include as many scientific priors derived from its chemical formula as possible. These features fall into 4 groups: stoichiometric properties, elemental properties, electronic structure, and ionic compound features. We utilise these embeddings as initial precursor feature vectors.

**Representation of Processing Actions**

Sequence representation is a well studied problem in ML research [15–17]. Here we utilise single-layer Long Short-Term Memory Units (LSTMs) [18] in an autoencoder architecture to obtain fixed-length representations of the processing sequences.

**Reaction Graph Model**

In solid-state syntheses, precursors are inherently ‘mixed’ together to form a product. Here we make use of an architecture for set regression based on the Roost model introduced in [19]. The model treats this process through a series of message-passing stages in which precursors appear as nodes on a dense, weighted graph. A fixed-length representation for the reaction is derived from this graph which allows for generalisation to arbitrary numbers of precursors.

This learned reaction embedding is then used to predict the elements present in the target material. The same embedding is then used to separately predict the product stoichiometry given the elements in the product. If the stoichiometry was directly predicted, an arbitrary threshold stoichiometry would have to be used to select which precursor elements are actually present in the product. The two stage approach used here is therefore necessary to differentiate between products with and without trace elements.

**Reaction Representation Learning**

The crucial first step is to produce a fixed-length representation of the reaction mixture. Figure 1a. shows how precursors are represented as nodes on a dense, weighted graph and action sequences are transformed into a fixed-length representation using an LSTM encoder. The initial precursor feature vectors are transformed into a trainable embedding through a learnable affine transformation,

\[ \vec{v}_i = W \vec{v}_{0,i}, \]

where \( W \) is a learnable weight matrix, and \( \vec{v}_{0,i} \) is the initial precursor representation.
A series of message-passing operations (Figure 1b.) then update the precursor representations
\[
v_{t}^{i+1} = U^{t}(v_{t}^{i}, \sum_{j \neq i} v_{t}^{j}, u_{seq}),
\]
where \(v_{t}^{i}\) is the ith precursor feature vector at message-passing layer \(t\), \(v_{t}^{j}\) are the other precursor features, \(u_{seq}\) is the fixed-length representations of the action sequences and \(U^{t}\) is the update function.

At the heart of the update function is the soft-attention mechanism, which allows the model to capture the importance of the other precursors in the reaction to the precursor in question. Attention is in general a method for assigning the importance of certain features in a machine learning task. A soft-attention mechanism extends this idea by allowing the model to learn the attention coefficients itself [16]. This has been shown to improve graph representation learning [20]. Here, unnormalised attention coefficients are first calculated across pairs of precursors
\[
\epsilon_{ij} = f(v_{t}^{i} \oplus v_{t}^{j} \oplus u_{seq}),
\]
where \(f\) is a single hidden-layer neural network, \(\oplus\) is the concatenation operation, and \(u_{seq}\) is the action sequence embedding which is used as a ‘global’ state in all update steps. This gives procedural context to enable differentiation between products formed through different synthesis pathways. \(\epsilon_{ij}\) are then normalised by a weighted softmax function,
\[
\alpha_{ij} = \frac{w_{ij} \exp(\epsilon_{ij})}{\sum_{k} w_{ik} \exp(\epsilon_{ik})},
\]
where the weight \(w_{ij}\) is the molar amount of precursor \(j\) in the balanced chemical equation. This gives the model context of the quantities of each precursor in the reaction mixture. The node feature is then updated with the pairwise interactions between nodes, weighted by their attention coefficients
\[
v_{t}^{i+1} = v_{t}^{i} + \frac{1}{K} \sum_{k=1}^{K} \sum_{j \neq i} \alpha_{ij}^{k} g^{k}(v_{t}^{i} \oplus v_{t}^{j} \oplus u_{seq}).
\]
Here \(g\) is again a single hidden-layer neural network, and we average over \(K\) attention heads, which has been shown to stabilise performance [20].

A similar attention-mediated pooling operation is then used to create a fixed-length, learned representation of the reaction \(\vec{r}\) from the final graph. The model learns how much attention to pay to each precursor given its final representation and relative quantity.

**Product Prediction**

The learned reaction embedding is then fed through a feed-forward neural network trained to perform multi-label element classification for the target product of the reaction. In materials synthesis, the target material cannot contain elements that are not present in the precursors. We add this inductive bias in the form of a binary ‘alchemy mask’ on the output of the network. Oxygen is assumed to be abundant and is additionally included in all masks irrespective of its presence in the precursors.

Once the elements in the product have been predicted, the final task is to predict the relative amounts of each element.
element in the product. Here, we use Matscholar element embeddings, $\vec{e}_i$, to represent and differentiate between constituent elements [14]. These are concatenated with the reaction representation, $\vec{r}$, to provide a contextualised query and fed through another feed-forward neural network, $h$, to predict the relative amount of element $i$,

$$a_i = h (\vec{e}_i \oplus \vec{r}). \quad (6)$$

These are then normalised using a softmax function to arrive at a fractional stoichiometry

$$s_i = \frac{1}{N} \sum_{k=1}^{N} \text{softmax}_i(a_i^k), \quad (7)$$

where $s_i$ is the normalised stoichiometry of element $i$ in the product, and we average over $N$ predictions to increase stability and accuracy. In addition, using an ensemble of $N$ models also allow us to estimate the epistemic uncertainty by taking the variance in predictions across the models in the ensemble. We refer to this ensembled model as the reaction graph model.

### RESULTS AND DISCUSSION

#### Product Prediction

The reaction graph model was tested on solid-state reactions with up to ten precursors. Product elemental composition was predicted with a subset accuracy of 0.940. This can be compared to a null baseline accuracy of 0.338 which would be achieved if all elements in the precursors were assumed to be present in the product. There is evidently a bias towards positive element labelling; 83.9% of elements in the precursors are also present in the product. Thus another metric of interest is the F1 score. The average F1 score achieved was 0.9910, where the score for each element is weighted by the number of true occurrences to account for the elemental imbalance.

The L1 and L2 distances between composition vectors can provide metrics for measuring material similarity. Here, the fractional stoichiometry vector of predicted products was accurate to a mean L1 distance of 0.1259, and L2 distance of 0.0729 to their true stoichiometries. An $R^2$ score of 0.9354 was achieved, which measures the correlation between actual and predicted stoichiometries.

#### Interpreting Model Predictions

##### Product Elements Prediction

In general, through the inclusion of the *alchemy mask* which prevents chemical inconsistencies, the prediction of which elements are present in the product is very accurate. The small number of erroneous predictions made by the model can typically be attributed to two areas:

**Carbon False Negatives**

| Example                          | Analysis                                                                 |
|---------------------------------|--------------------------------------------------------------------------|
| $0.19 \text{K}_2\text{CO}_3 + 0.31 \text{Li}_2\text{CO}_3 + 0.5 \text{CO}_2 + 0.25 \text{O}_2 \Rightarrow \text{Li}_{10.62}\text{K}_{0.38}\text{CO}_3$ | Products containing carbon are rare in the dataset and ones that do are mostly carbonates. The model has a structure-agnostic approach and treats each element as a simple ratio of presence in the product. Carbonates are covalently bonded polyatomic ions, and thus do not have the same characteristics as independent elements in a crystal. This could be why a lower accuracy is seen for these reactions. |

**Data Processing of Dopants**

| Example                          | Analysis                                                                 |
|---------------------------------|--------------------------------------------------------------------------|
| $3 \text{Mo}_3\text{O}_8 + \text{La}_2\text{O}_3 \Rightarrow \text{La}_2(\text{MoO}_4)_3$; with additives (Eu$^{3+}$) via \text{Et}_2\text{O}_3 | Some products in the dataset have additives/dopants which do not appear in the target stoichiometry. However, the relevant precursors are still present in the data (Eu$_2$O$_3$ in this case). The model in these cases accurately predicts the presence of these elements in the product, but the additive is not accounted for in the product stoichiometry data. |

##### Stoichiometry Prediction

The inclusion of precursor amounts as weights is seen to provide vital information in differentiating between product stoichiometries. For example, the model distinguished well between two otherwise equivalent ceramic syntheses:

**Utility of Precursor Weighting**

| Reaction                          | Analysis                                                                 |
|----------------------------------|--------------------------------------------------------------------------|
| PbO + 0.03 TiO$_2$ + 0.97 ZrO$_2$ $\Rightarrow$ Pb(Zr$_{0.97}$Ti$_{0.03}$)O$_3$ | Here, (A) was used to train the model but (B) was in the unseen test set and was predicted with an L1 distance of less than 0.0056 from its true value. The learned reaction representation has thus effectively encoded information about how the relative amounts of precursors affects stoichiometry. |

However, the model does have some failure modes. As many of these cases can be identified via estimation of the epistemic uncertainty we are primarily interested in explaining the worst predictions – those which have large error but low uncertainty.
Analysis of such erroneous predictions shows that they can generally be grouped into two main areas:

**Oxygen Over-prediction**

**Example**  
\[ 1.5 \text{Na}_2\text{CO}_3 + 0.333 \text{Co}_3\text{O}_4 \Rightarrow \text{Na}_3\text{CoO}_2 + 1.5\text{CO}_2 + 0.417\text{O}_2 \]

**Target**  
Na: 0.5000  Co: 0.1667  O: 0.3333

**Prediction**  
Na: 0.1985  Co: 0.2692  O: 0.5323

**Analysis**  
Many errors arise from the over-prediction of oxygen content. This may be due to the over-representation of oxygen in the dataset. A large subset of products are perovskite materials where oxygen typically is present in a fractional amount of 0.6 (e.g. BaTiO$_3$).

**Pure Precursors and Alloys**

**Example**  
\[ 8 \text{Ni} + 86 \text{Al} + 6 \text{La} \Rightarrow \text{Al}_{86}\text{Ni}_8\text{La}_6 \]

**Target**  
Ni: 0.0800  La: 0.0600  Al: 0.8600

**Prediction**  
Ni: 0.6859  La: 0.1497  Al: 0.1642

**Analysis**  
This reaction is for the formation of an amorphous aluminium alloy anode through arc melting. Although it looks as if product prediction should be simple, Magpie features are not designed to work with pure elements. This could be the reason for the large errors in these cases.

Overall, many of the incorrect predictions come from either misrepresentation in the dataset or from the lack of structural information. A material is defined not only by its stoichiometry but crucially by its crystal structure. The dataset and therefore this framework is completely structure-agnostic and thus is limited in its predictive abilities. The use of predicted crystal structures from stoichiometry could be an avenue of further investigation [21].

**Benchmarking and Ablation Study**

To place the accuracy of our model into context, we benchmark our reaction graph model against a baseline approach and conduct an ablation study on its key features.

The baseline model we use takes the Magpie precursor embeddings and concatenates them to be used as the input to a neural network with the same architecture as the output network of the reaction graph model. As with the reaction graph model, we make use of an ensemble of $N$ individual models for this baseline.

While the reaction graph model can handle variable numbers of precursors, this Magpie baseline approach is inherently restricted to a certain input size. Thus we limit our model to reactions with up to three precursors only, and zero-pad the input in the Magpie baseline model where there are less than three precursors to enable comparison.

Figure 2. Threshold plot for product stoichiometry prediction. The cumulative proportion of correctly predicted stoichiometries is plotted against an L1 error threshold. A clear improvement is seen through using the learned reaction graph embedding compared to fixed Magpie features. No apparent improvement is observed through the inclusion of the action sequences.

The ablation study compares results across reaction graph models with/without the processing actions and the alchemy mask to discern which parts of the model design are significant. A summary of the key accuracy metrics is shown in Table I for models in the benchmarking and ablation study. A null baseline result is also given for comparison where we assume all elements in the precursors are present in the product, and are present in their average relative amounts derived from the training data.

**Reaction Representation Learning**

For product element prediction, the reaction graph models are seen to perform significantly better than using Magpie features alone. For stoichiometry prediction, Figure 2 shows the proportion of predictions that are within given L1 error tolerances. A clear improvement is seen through incorporating the graph representation learning framework.

Figure 3 shows plots of actual against predicted fractional stoichiometry. This figure shows how predictions are distributed about the equivalence line, and their relative confidence as measured through ensemble variance. Similar to the null baseline, the Magpie baseline model predictions show a lack of parity about the equivalence line.

In contrast, the learned reaction graph embeddings show a significant improvement, both in absolute terms and in better prediction of the model uncertainty. In general, the most certain predictions are closest to the equivalence line and there is more parity. This indicates the model has assimilated some chemical knowl-
Table I. Ablation performance table for reactions with up to three precursors. The subset accuracy and weighted F1 score shown is for the prediction of product elements. The null baseline result assumes all elements in the precursors are present in the product, and assumes that elements are present in their average amounts from the training data. Providing inductive bias to the model in the form of the alchemy mask improves accuracies across the board. The reaction graph framework shows a clear improvement over the Magpie baseline. Removing the action context does not have a notable effect on model accuracy.

| Reaction Representation | Alchemy Mask | Subset Accuracy | F1 Score | Mean L1 | Mean L2 | $R^2$ |
|------------------------|--------------|----------------|----------|---------|---------|-------|
| Magpie Baseline        | 0.625        | 0.902          |          |         |         |       |
| Reaction Graph without Actions | 0.773 | 0.950 |          |         |         |       |
| Complete Reaction Graph | 0.770 | 0.948 |          |         |         |       |
| Null Baseline          | ✓            | 0.416          | 0.988    | 0.2233  | 0.1333  | 0.8406|
| Magpie Baseline        | ✓            | 0.933          | 0.990    | 0.1568  | 0.0948  | 0.9050|
| Reaction Graph without Actions | ✓ | 0.960 | 0.994    | 0.1298  | 0.0790  | 0.9148|
| Complete Reaction Graph | ✓            | 0.960          | 0.994    | 0.1326  | 0.0805  | 0.9131|

edge. The improvement in uncertainty prediction can be seen more clearly in the confidence-error plot in Figure 4. Including the learned reaction graph embedding shows a clear decrease in the mean error of predicted products as the most uncertain predictions are removed. The almost flat line observed when using fixed Magpie embeddings indicates that there is little calibration between error and uncertainty for the baseline model.

Alchemy Mask

The results in the lower half of the table are achieved through the inclusion of the alchemy mask, which effectively prevents the model from making predictions which are chemically unfeasible. This inductive bias provides the most notable increase in accuracy across the board for element prediction. This highlights the utility of incorporating physical principles into ML models.

Processing Action Sequences

The synthesis actions are crucial in determining the major product of an inorganic reaction; different procedures on the same precursors can lead to differing products due to the interplay between thermodynamics and kinetics. However, Table I shows the addition of the action sequences in the reaction embedding has a detrimental effect – if any – on model performance.

The level of detail of the action data used in the model can provide an explanation. Different product compositions and morphologies can be formed from even slight changes in temperature or pressure. Thus one cannot expect that using the generic, coarse terms to represent action sequences can lead to significant improvements in performance. The inaccuracies of the processing sequences in the dataset only serve to compound this effect. Thus a model that includes more granular processing procedures is required for practical applications.

Despite this, when the models were extended to the full dataset by including reactions with up to ten precursors, there was a significant improvement in accuracy through incorporating the processing actions sequences. An L1 error of 0.1357 was achieved without the actions, and an L1 error of 0.1259 with the actions. This is likely due to the greater diversity of reactions in the extended dataset, meaning the actions are more important in providing contextual information.

CONCLUSION

In this work, a data-driven approach to inorganic reaction prediction has been investigated for the first time. A physically-motivated model has been developed which can predict major products of solid-state reactions from precursors and synthesis procedures. The key feature of this work is the generation of a learned reaction embedding which effectively encodes the reaction context for product prediction. Through an ablation study, this framework is shown to predict product stoichiometries more accurately than less physically-motivated models. Importantly, it also more reliably assesses the uncertainty in its predictions.

The reaction representation used here can also be transferred to further problems. One can envisage using it to predict the probability of success of a given synthesis procedure. This would further aid retrosynthesis; precursors and synthesis pathways which have the highest probability of success can be preferentially chosen. This would require supervision in the form of both successful and unsuccessful reactions [22].

This work provides a basis for evaluating a future retrosynthesis planner for inorganic materials. For practical use, however, further work is required to differentiate between morphologies of products and more granular synthesis action steps with relevant conditions are needed. Access to greater quality and quantity of synthesis data and a concerted effort between ML practitioners and experimentalists will allow these advances.
Figure 3. Parity plots for product stoichiometry prediction. The fractional amount of each element in the target product is plotted against its predicted value for each model in the ablation study. The data points are coloured according to their relative confidence, estimated from the variation in predictions from the ensemble. An improvement in both correlation and uncertainty estimation is seen through the addition of the graph representation learning framework.

Figure 4. Confidence-error plots for stoichiometry prediction showing the mean L1 error in product stoichiometry prediction as the most uncertain predictions are removed. Uncertainty estimation is greatly improved through using the reaction graph embedding – more erroneous predictions are in general more uncertain.

METHODS

Data Processing

Reactions with non-stoichiometric or organic precursors and targets, and reactions with pure products or only one precursor were removed from the dataset. This reduced the dataset size to 16,231 reactions with up to ten precursors, and 11,083 reactions for the ablation study with up to three precursors. This was randomly split into training and test sets with an 80:20 ratio.

Processing action sequences ranged from 0 to 16 steps in length, and were represented by 7 unique action types: {Dry Mixing, Mixing in solution, Quenching, Shaping, Heating, Drying, Liquid Grinding}.

Precursor stoichiometries were extracted along with their molar ratios from the balanced chemical equation for each reaction. Where precursors or targets were themselves a mixture of different materials/phases, their stoichiometries were added, weighted by their amounts.

Targets were extracted as 81-dimensional stoichiometric vectors. The set of elements present in the precursors was used to construct the alchemy mask.

Implementation Details

The model was implemented in PyTorch [23] and Matminer was used for generating Magpie features from precursor stoichiometries [24].

An embedding dimension of 8 and a hidden state dimension of 32 were used for the processing actions autoencoder. This lead to a sequence reconstruction accuracy of ~ 87%. Reconstruction accuracies of over 95% were achieved using a hidden dimension of 64 but did not significantly improve product prediction accuracy.

The learnable precursor embedding size was set to 128, and 5 message-passing layers were used, each with 3 attention heads. The hidden layer dimension was chosen to be 256 for both \( f \) and \( g \). The output network of the product element prediction model had hidden layer dimensions of 256, 512, 512 and 256 respectively with ReLU activation. The threshold output probability for elemental presence was chosen through maximising accuracy on a validation set. The stoichiometry prediction model had hidden layer dimensions of 256, 512, 512, 256, 128, 128 and 64 respectively with ReLU activation. These were averaged over 5 iterations. Skip connections were used to reduce the vanishing gradient effect [25].

Learning rates were chosen using a heuristic search where we choose a rate one magnitude lower than that which causes the training loss to diverge. A minibatch size of 256 and the Adam optimiser was used to train the model [26]. Early stopping using an 80:20
train:validation split was used in all models to avoid over-fitting.

The ensembles used to estimate the epistemic uncertainty in both the Magpie baseline and reaction graph models were constructed using 5 individual models trained on the same data with different initialisation.

Data availability

The exact data used in this work can be found at https://doi.org/10.6084/m9.figshare.9722159.v3. More recent versions of the synthesis dataset are released by the original authors at https://github.com/CederGroupHub/text-mined-synthesis_public.

Code availability

An open-source repository of the code used in this work is available at https://github.com/s-a-malik/inorg-synth-graph.

ACKNOWLEDGEMENTS

R.E.A.G. and A.A.L. acknowledge the support of the Winton Programme for the Physics of Sustainability.

* Correspondence email address: aal44@cam.ac.uk
[1] V. D’Innocenzo, A. R. Srimath Kandada, M. De Bastiani, M. Gandini, and A. Petrozza, Journal of the American Chemical Society, Journal of the American Chemical Society 136, 17730 (2014).
[2] K. Kalantari, I. Sterianou, S. Karimi, M. C. Ferrarelli, S. Miao, D. C. Sinclair, and I. M. Reaney, Advanced Functional Materials 21, 3737 (2011).
[3] Y. Gong, Z. Liu, A. R. Lupini, G. Shi, J. Lin, S. Najmæi, Z. Lin, A. L. Elias, A. Berkedemir, G. You, H. Terrones, M. Terrones, R. Vajtai, S. T. Pantelides, S. J. Pennycook, J. Lou, W. Zhou, and P. M. Ajayan, Nano Letters, Nano Letters 14, 442 (2014).
[4] 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, APL Materials 1, 011002 (2013), https://doi.org/10.1063/1.4812323.
[5] J. Saal, S. Kirklin, M. Aykol, B. Meredig, and C. Wolverton, JOM 65, 1501 (2013).
[6] M. Jansen, Advanced Materials, Advanced Materials 27, 3229 (2015).
[7] A. Ludwig, npj Computational Materials 5, 70 (2019).
[8] M. G. Kanatzidis, Inorganic Chemistry, Inorganic Chemistry 56, 3158 (2017).
[9] M. Aykol, V. I. Hegde, L. Hung, S. Suram, P. Herring, C. Wolverton, and J. S. Hummelshøj, Nature Communications 10, 2018 (2019).
[10] Y. Kim, E. Kim, E. Antono, B. Meredig, and J. Ling, “Machine-learned metrics for predicting the likelihood of success in materials discovery,” (2019), arXiv:1911.1201 [cond-mat.mtrl-sci].
[11] V. Kononova, H. Huo, T. He, Z. Rong, T. Botari, W. Sun, V. Tshitoyan, and G. Ceder, Scientific Data 6, 203 (2019).
[12] E. Kim, Z. Jensen, A. van Grootel, K. Huang, M. Staib, S. Mysore, H.-S. Chang, E. Strubell, A. McCulm, S. Jegelka, and E. Olivetti, Journal of Chemical Information and Modeling 60, 1194 (2020), pMID: 31909619, https://doi.org/10.1021/acs.jcim.9b00995.
[13] L. Ward, A. Agrawal, A. Choudhary, and C. Wolverton, npj Computational Materials 2, 16028 (2016).
[14] V. Tshitoyan, J. Dagdelen, L. Weston, A. Dunn, Z. Rong, O. Kononova, K. A. Persson, G. Ceder, and A. Jain, Nature 571, 95 (2019).
[15] I. Sutskever, O. Vinyals, and Q. V. Le, CoRR abs/1409.3215 (2014), arXiv:1409.3215.
[16] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, “Attention is all you need,” (2017), arXiv:1706.03762 [cs.CL].
[17] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, “Bert: Pre-training of deep bidirectional transformers for language understanding,” (2018), arXiv:1810.04805 [cs.CL].
[18] S. Hochreiter and J. Schmidhuber, Neural Comput. 9, 1735 (1997).
[19] R. E. A. Goodall and A. A. Lee, “Predicting materials properties without crystal structure: Deep representation learning from stoichiometry,” (2019), arXiv:1910.00617 [physics.comp-ph].
[20] P. Veličković, G. Cucurull, A. Casanova, A. Romero, P. Liò, and Y. Bengio, “Graph attention networks,” (2017), arXiv:1710.10903 [stat.ML].
[21] Y. Zhao, Y. Cui, Z. Xiong, J. Jin, Z. Liu, R. Dong, and J. Hu, ACS Omega, ACS Omega 5, 3596 (2020).
[22] P. Raccuglia, K. C. Elbert, P. D. F. Adler, C. Falk, M. B. Wenny, A. Mollo, M. Zeller, S. A. Friedler, J. Schrier, and A. J. Norquist, Nature 533, 73 (2016).
[23] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, A. Desmaison, A. Kopf, E. Yang, Z. DeVito, M. Raison, A. Tejani, S. Chilamkurthy, B. Steiner, L. Fang, J. Bai, and S. Chintala, in Advances in Neural Information Processing Systems 32, edited by H. Wallach, H. Larochelle, A. Beygelzimer, F. d’Alché Buc, E. Fox, and R. Garnett (Curran Associates, Inc., 2019) pp. 8024–8035.
[24] L. Ward, A. Dunn, A. Faghaninia, N. Zimmermann, S. Bajaj, Q. Wang, J. Montoya, J. Chen, K. Bystrom, M. Dylla, K. Chard, M. Asta, K. Persson, G. Snyder, I. Foster, and A. Jain, Computational Materials Science 152, 60 (2018).
[25] K. He, X. Zhang, S. Ren, and J. Sun, in 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2016) pp. 770–778.
[26] D. P. Kingma and J. Ba, “Adam: A method for stochastic optimization,” (2014), arXiv:1412.6980 [cs.LG].