



Cite this: *Phys. Chem. Chem. Phys.*,  
2026, **28**, 8146

# Two-stage transfer learning for deep learning-based prediction of lattice thermal conductivity

Liudmyla Klochko \* and Mathieu d'Aquin

Machine learning promises to accelerate material discovery by enabling high-throughput prediction of desirable macro-properties from atomic-level descriptors or structures. However, the limited data available about precise values of these properties has been a barrier, leading to predictive models with limited precision or ability to generalize. This is particularly true of lattice thermal conductivity (LTC): existing datasets of precise (*ab initio*, DFT-based) computed values are limited to a few dozen materials with little variability. Based on such datasets, we study the impact of *transfer learning* on both the precision and robustness of a deep learning model (ParAlsite). We start from an existing model (MEGNet<sup>1</sup>) and show that significant improvements in predicting high-quality approximations of LTC are obtained through applying transfer learning twice: once on the basis of a pre-training of the model on a large number of materials for a different task (predicting formation energy), and a second time using a medium size dataset (a few thousand materials) of low-quality approximations of LTC (based on the AGL workflow). In other words, greater precision and robustness is obtained after a final training (fine-tuning) of the twice pre-trained model with our high-quality, smaller-scale dataset. We also analyze results obtained from using this higher-precision deep-learning model to scan large numbers of materials from the Material Project Database, in search of low-thermal-conductivity compounds.

Received 14th November 2025,  
Accepted 27th February 2026

DOI: 10.1039/d5cp04401d

rsc.li/pccp

## 1 Introduction

Machine learning models have advanced material research in several fields, including various domains of physics,<sup>2–5</sup> quantum chemistry,<sup>6,7</sup> drug discovery,<sup>8</sup> and cancer studies.<sup>9–11</sup> In particular, recent studies have proposed various models to predict the physical properties of materials.<sup>1,12,13</sup> These models utilize diverse datasets, input data, and tuned neural network designs for specific purposes. However, the applicability of those models, *i.e.* their efficacy when applied on large databases, remains limited by their precision and robustness, which are in turn dependent on the quality and size of the available training data. In other words, unsurprisingly, machine learning models trained on small databases of similar materials do not perform well on large databases of diverse materials.<sup>14</sup>

Predicting the low lattice thermal conductivity (LTC) of crystal compounds on the basis of their structure and physical properties is one of those tasks that are made challenging by the lack of quality data. However, it is critical, as the ability to identify low-LTC materials at a large scale could have profound implications for the design and optimization of materials in various applications, from electronics to energy storage,<sup>15,16</sup> including the integration of machine learning with molecular

dynamics simulations.<sup>5,17,18</sup> The difficulty of this problem lies in the complexity of the relationship between the structure of a material and its thermal properties. Although large datasets on material properties are available through databases such as AFLOW,<sup>19</sup> OQMD,<sup>20</sup> Materials Project,<sup>21</sup> and JARVIS,<sup>22</sup> available data on LTC is either based on approximate workflow (such as AGL<sup>23,24</sup>) and therefore not usable to build precise machine learning-based predictive models, or rely on *ab initio*, DFT-based computations which are too expensive to run at large scales and are therefore of very small sizes.

With the growing demand for high-throughput screening of materials with specific properties led to the development of various predictive AI models.<sup>25–27</sup> However, many of these approaches remain limited in scope, rely on narrowly defined application domains, or depend on training datasets that are not publicly available or accessible only upon reasonable request.<sup>28,29</sup> Despite recent advances, including corporate driven frameworks such as MatterSim<sup>30</sup> and EquiFlash,<sup>31</sup> these approaches remain constrained by limited openness, reproducibility, and generalizability, underscoring the need for alternative or complementary methodologies.

In this study, we apply a two-stage transfer learning methodology as a way to take advantage of both small and large datasets to achieve greater levels of precision and robustness in deep learning models for predicting LTC. Transfer learning as applied in this article is a process in which a model trained on a

Université de Lorraine, LORIA, Nancy F-54000, France.  
E-mail: liudmyla.klochko@loria.fr



first task is reused as a starting point for training on a second similar task. The idea is that initial patterns can be learned in larger, relevant data, which will bootstrap the learning on smaller, more targeted data. The main hypothesis of this work is therefore that an existing model that has demonstrated good performance in predicting a property other than LTC is effective in transfer learning, and that this approach can be further extended in a second stage of transfer learning by using larger, low-quality LTC datasets to pre-train models for predicting on smaller, high-quality datasets. While not solving the complex problem of predicting low LTC from limited data, our contribution shows a step in that direction, which could be replicated in different contexts. It also enables us to explore the predictions made by such models on a large database of materials, to better understand their applicability for the discovery of low LTC materials and the factors that affect those predictions and their robustness.

## 2 Methodology

Transfer learning<sup>32</sup> involves the use of a machine learning model that has been pre-trained on a large dataset and subsequently fine-tuning it on a smaller domain-specific dataset. This strategy is particularly effective when working with limited data, as it allows us to capitalize on the knowledge gained from broader datasets, as demonstrated in multiple applications for material properties.<sup>33,34</sup> Here, we describe the datasets, model, and process used to carry out our two-stage transfer learning process.

### 2.1 Data

As mentioned above, the accuracy and reliability of any machine learning model is highly dependent on the quality of the data used for training and validation. In this work, we rely on two datasets, each contributing to our analysis. The first one is derived from calculations of the first principle of anharmonic lattice dynamics,<sup>35,36</sup> looking at a small set of materials with specific structures. The latter can be seen to represent a larger dataset that contains low precision values for LTC.

**Togo15.** This dataset contains 96 materials used in a previous prediction study<sup>35</sup> in the rocksalt, zinc blende, and wurtzite structures that could be unambiguously identified in the Material Project Database. The LTC values are obtained using the phono3py software package<sup>37,38</sup> using the YAML files available through the PhononDB<sup>†</sup> repository. Obtaining predictions with low deviation from those values is the central motivation for this work.

**AFLOW AGL.** This “low precision” dataset contains 5578 materials obtained from the AFLOW-LIB repository<sup>19</sup> together with their corresponding thermal conductivity obtained with the use of a quasi-harmonic Debye-Grüneisen model.<sup>23,24</sup>

Fig. 1 shows the distributions of the LTC values in these two datasets. As can be seen, those LTC values were computed

using different techniques on different selections of materials, and therefore, their distributions are also very different.

In the experiments training deep learning models described below, for all cases, logarithmic scaling is applied to all values of LTC, followed by standardization using the parameters of the corresponding dataset. For validation purposes, each dataset is split 9 times, keeping 80% for training. In other words, each dataset is associated with 9 different randomly selected validation sets that represent each 20% of the total dataset that is not used for training. Any result shown later in this article is measured as an average over those 9 validation sets and the corresponding models for a given dataset.

### 2.2 Model

Our model for predicting LTC from the properties and structure of materials (called ParAIsite) is based, as shown in Fig. 2, on the addition of a multilayer perceptron (MLP, a set of dense, feed-forward, fully connected layers) on top of an existing, pre-trained model. The additional MLP layers serves as a task-specific “head” that adapts the general structural representations learned by MEGNet to the specific task of thermal conductivity prediction without modifying MEGNet’s core architecture.

For more details regarding the model parameters, see Section S1 of the SI document. In line with our transfer learning approach, the idea here is to use an existing model (the pre-trained model) that has already shown its ability to predict properties of materials as a foundation to be adapted for the task of LTC prediction. More concretely, ParAIsite is based on connecting the last hidden layer of the pre-trained model to the input of the added MLP.

A first step, therefore, for establishing this model is the selection of the most appropriate pre-trained model from the top-performing models cataloged on MatBench.<sup>39</sup> We only considered models based on an unambiguous identification of the materials as input and on a representation of their structures (in the form of crystallographic information files, CIFs). An initial set of tests was carried out with Togo15 using the model of Fig. 2 with each candidate pre-trained model to validate the model’s performance and ascertain its suitability for the specific challenges associated with predicting the thermal conductivity in crystal compounds.

The results are shown in Fig. 3. We chose the model that combined the best performance (measured by the mean absolute percentage error, MAPE) and was most consistent with the features of our dataset. Despite the better average performance of the CrabNet model<sup>40</sup> on Togo15, results over multiple runs showed a lot of variability, demonstrating that this model was too unstable to be used. That result led us to use the graph-based neural network model MEGNet,<sup>1</sup> which was pre-trained on the formation energy data of 62 315 compounds from the Materials Project Database as of the 2018.6.1 version. In the next section, by fine-tuning MEGNet on our thermal conductivity data, we aim to improve the prediction accuracy and better understand the thermal properties of the crystal compounds in our dataset.

<sup>†</sup> <https://github.com/atztogo/phonondb>.



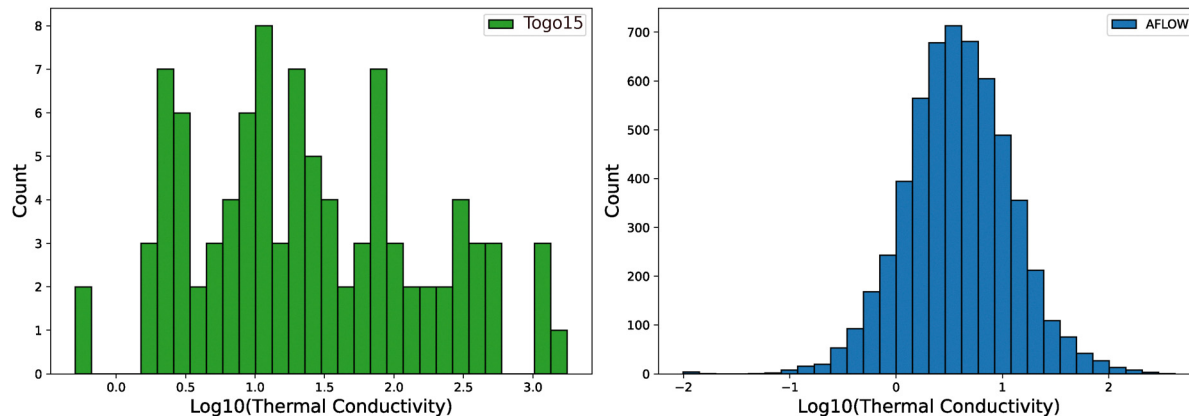


Fig. 1 Distribution of LTC values for each selected dataset. To provide a better insight into the datasets' chemical space coverage (see SI. Fig. S17–S28).

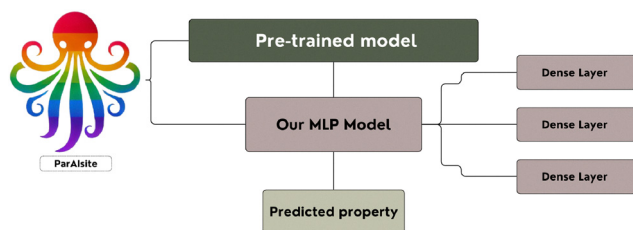


Fig. 2 The ParAIsite model architecture. Crystal structure (CIF file) containing atomic positions and unit cell parameters will serve as the input to the model. As the pre-trained model architecture, we selected MEGNet architecture that processes the structure. The output of MEGNet last hidden layer connects to the input layer of Multi-Layer Perceptron (MLP) with dense, fully-connected layers. The output property is the lattice thermal conductivity (LTC). Details of the model are shown in Section S1 of SI.

### 2.3 Model training

As mentioned earlier, we adopted a two-stage pre-training approach in order to improve the precision of LTC prediction from our low-size, “first-principles” dataset (see workflow in Fig. 4). To measure the impact of transfer learning, we compare the results at each stage with the results obtained from training MEGNet directly on Togo15. This leads to three different steps in our training of the ParAIsite model. In summary, those steps move from no transfer learning at all, to two levels of pre-training/transfer learning being applied, allowing us to evaluate how pre-trained models and fine-tuning affect the model's ability to precisely and robustly predict LTC.

Step 1 (no pretraining, pure training from scratch). In this step, we train and test ParAIsite on our two datasets using an uninitialized MEGNet (and MLP) model. The training is therefore done from scratch (with random initial weights), without any form of transfer learning.

Step 2 (using a pre-trained MEGNet). In this step, we train and test ParAIsite on our datasets using a pre-trained MEGNet, that is, where MEGNet is initialized with the weights obtained through the training carried out by its authors on the task of predicting formation energy, while MLP model is initialized

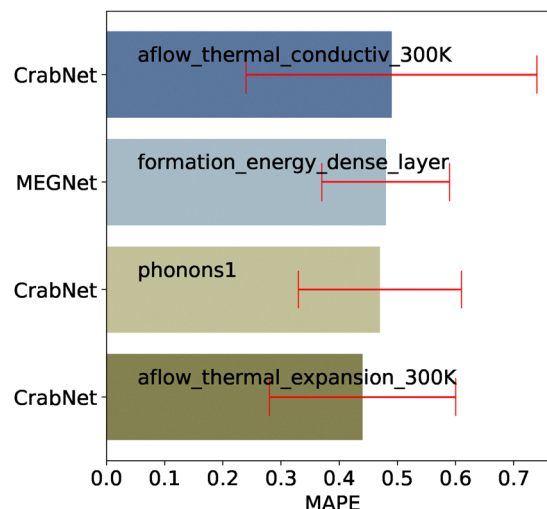


Fig. 3 Results of validation tests of ParAIsite when using top-performing models cataloged on MatBench<sup>39</sup> as pre-trained models. During our tests, we considered only models that take crystal structures, represented as CIF files, as input and provide an unambiguous identification of the materials. Initial tests on the Togo15 dataset, using the architecture shown in Fig. 2 with each candidate pretrained model, were conducted to evaluate their suitability for lattice thermal conductivity prediction. The name of the datasets on which model was pre-trained are indicated inside the box, and the error variability across 9 runs (9 division training/validation sets) are shown in red lines. As can be seen, the MEGNet model<sup>1</sup> trained on formation energy shows better stability compared to the CrabNet model<sup>40</sup> when adapted to Togo15.

randomly. It is important to note that there is no reuse of Step 1 weights here.

Step 3 (transfer-learning over model fine-tuned with AFLOW). In this step, we train and test ParAIsite on Togo15, taking as a starting point the best model obtained from fine-tuning the pretrained MEGNet on the AFLOW AGL dataset. In other words, we apply two rounds of transfer learning, first from energy formation to the AFLOW AGL dataset, and second from the AFLOW AGL dataset to Togo15. By pretraining the whole model further on a larger, but lower quality dataset, we



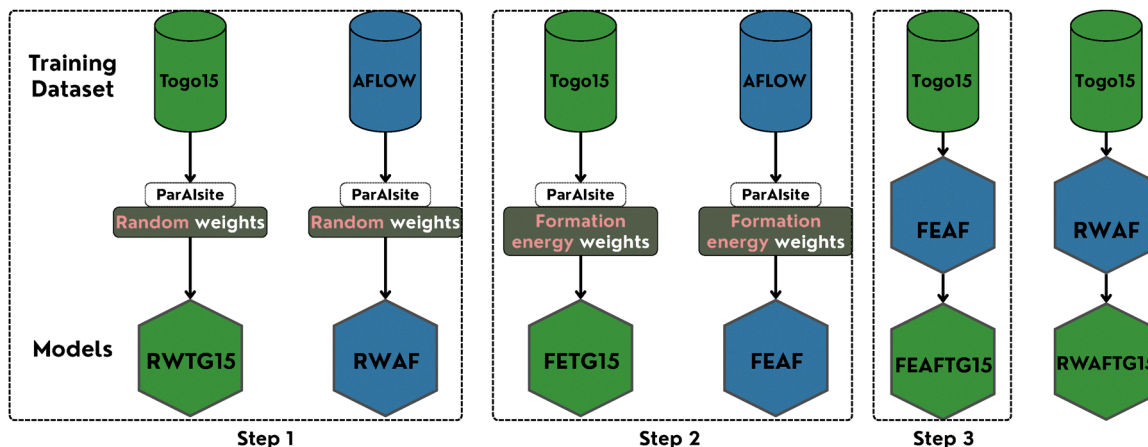


Fig. 4 Sketch representing the different model trained for comparison in our methodology. Training datasets are illustrated as cylinders, and the resulting models after training ParAlsite are represented as cubes. The models are labeled from left to right as follows: Step1 (no pre-training): Random Weights Togo15 (RWTG15), Random Weights AFLOW (RWF). Step 2 (using a pre-trained weights of MEGNet on the formation energy): Formation Energy Togo15 (FETG15), Formation Energy AFLOW (FEAF). Step 3 (transfer learning over model fine-tuned with AFLOW): Formation Energy AFLOW Togo15 (FEAFTG15), Random Weights AFLOW Togo15 (RWAFTG15).

anticipate that training on the other, more precise datasets will converge to better performance.

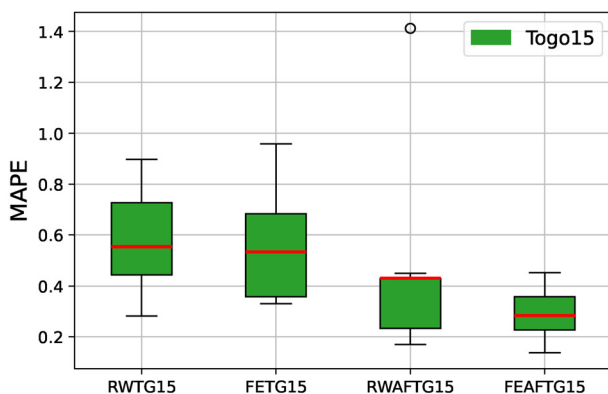
At each step, for each model, training is performed for 300 epochs with a fixed random seed of 42, ensuring reproducible results. The value 42 has no physical significance. It is a commonly used arbitrary seed value in machine learning to fix the random number generator for reproducibility. As already mentioned, each of these steps is repeated 9 times (*i.e.* for 9 runs) to ensure statistical robustness. The averaged results for the validation loss across all steps are presented in Fig. 6. For all training steps, we use MAPE as the loss function, having applied normalization and scaling. The main objective of this work being to achieve high-performance predictions on Togo15, to avoid confusion, in the rest of the article we refer as Step 1 the model RWTG15, Step 2 the model FETG15, and Step 3 the model FEAFTG15. An additional verification step is carried out with the model RWAFTG15 (1 stage transfer learning with pre-training only on AFLOW).

## 3 Results and discussion

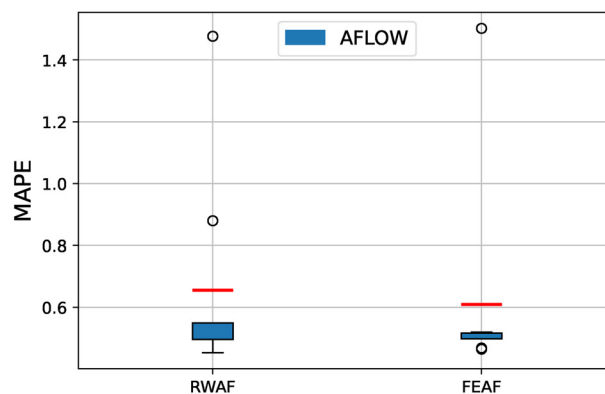
### 3.1 Hypothesis validation

A first straightforward conclusion that can be drawn from training and cross-validating in each step with the considered datasets is that the double-stage transfer learning process has a significant effect on the performance of the model on Togo15, as well as on the robustness of the prediction. In fact, in Fig. 5 we show the distribution of MAPE values (over the 9 runs) of the models trained on both datasets at each step.

As can be seen from this figure, a model trained on Togo15, which is considered particularly challenging, achieves low performance (55% error on average) without any pretraining (Step 1). Training and testing again using a pre-trained MEGNet for the task of energy formation (Step 2) falls to 53% on average. More significantly, at Step 3, the average MAPE falls to 28% when fine-tuning the model on Togo15, after having first fine-tuned the pre-trained MEGNet on the AFLOW AGL dataset



(a) Validation loss (MAPE) across training steps for models trained on Togo15.



(b) Validation loss (MAPE) across training steps for models trained on AFLOW dataset.

Fig. 5 Comparison of validation loss (MAPE) for different datasets.



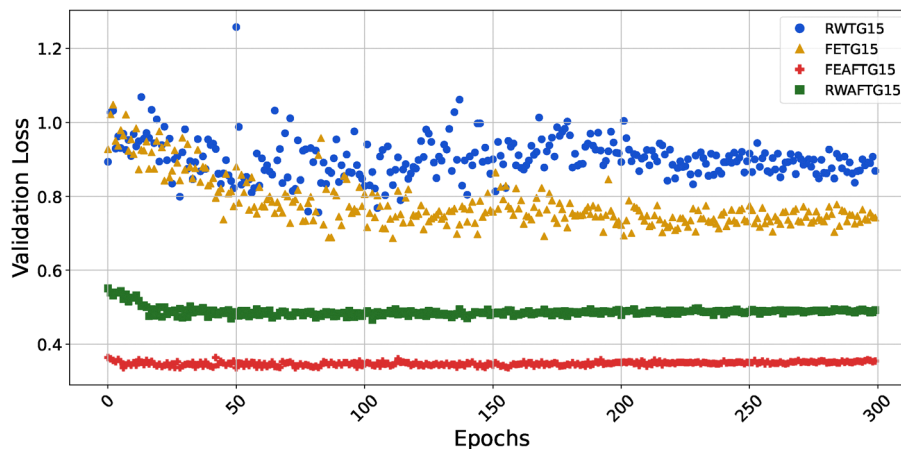


Fig. 6 Average validation loss for models trained and tested on Togo15 across training epochs.

(Step 3). In other words, transfer learning has had a significant effect, especially taking as starting point a model that has already been fine-tuned for a similar task (*i.e.* predicting approximated LTC through the AFLOW dataset). We can also see that the distribution of MAPE values on Togo15 at Step 3 is significantly narrower than at Steps 1 and 2, showing that models trained with double-stage transfer learning are also more robust, as their precision is less dependent on the initial conditions of the model. It could however be argued that this only shows that relying on a medium-size dataset of higher relevance (AFLOW) instead of one of larger scale, but lower relevance, is what leads to better performance. Indeed, as the results in Fig. 5a show, the performance of RWAFTG15, which was only pretrained on AFLOW, is significantly improved compared to that of FETG15 (Step 2), which was trained on formation energy, thus validating this intuitively valid hypothesis that AFLOW is a better candidate for transfer learning towards Togo15. However, the Step 3 model (FEFTG15) further improves the performance and robustness of the prediction over that of RWAFTG15, therefore demonstrating the contribution of our two-stage transfer learning approach. This effect of the two steps of transfer learning is particularly visible in Fig. 6, which shows the evolution of the average MAPE (over 9 runs) in the Togo15 validation subsets during training iterations (epochs). Separate plots for each 9 run *vs.* epoch for each model are shown in Section S3 of SI. Comparing this evolution between Step 1 and Step 2 shows that starting from a relevant pre-trained model, even if made for a different task, enabled the model to converge faster to slightly lower values of MAPE. We can also see that the MAPE on the validation subsets does not rise up in Step 2 as much as it did during the Step 1 training process, showing that the model was less prone to overfitting in this case. Looking at the chart for Step 3, we can see a significantly different behavior, with the MAPE value converging very quickly to much lower values. To estimate the impact of transfer learning on model performance, we also provide various metrics in Section S2 of SI.

### 3.2 Scan over material project database

To provide concrete validation of the best performing models, we applied them to obtain predictions for stable materials in the Material Project Database ( $\approx 34k$  materials). LTC for  $(\text{BaSbO}_3)_2$  (mp-9127) was then calculated using a robust *ab initio* calculation (with the same parameters as Togo15 dataset), as our models consistently found that it has a relatively low thermal conductivity. The result of the computation ( $7.1 \text{ W m}^{-1} \text{ K}^{-1}$ ) was on the same order of magnitude as the predictions of our models ( $1.23 \text{ W m}^{-1} \text{ K}^{-1}$ ). This agreement underscores the ability of the model to capture critical trends in LTC prediction, even for datasets it was not trained directly on. The complete set of screening results, including all predicted lattice thermal conductivity values, is provided in the associated GitHub repository.<sup>‡</sup>

In addition to the direct comparison of results for LTC, we performed the analysis of the variance in predicted thermal conductivities, pTC, and predicted mean thermal conductivities, mTC, for each stable material in the Materials Project database. Here, mTC is defined as the arithmetic mean of the predicted TC (pTC) values obtained from the 9 models independently trained on a given dataset at a given step. We scanned 34k stable materials from Materials Project with our 9 models. Each of 34k materials thus had 9 pTC values, over which we computed the variance (in units of TC,  $\text{W m}^{-1} \text{ K}^{-1}$ ),  $\text{var}(\text{pTC})$  (see Fig. 8). As shown in Fig. 8, for Togo15 our methodology results in a reduction in variance, indicating that the predictions became more stable and consistent in all steps. For the AFLOW dataset the nearly constant variance between Steps 1 and 2 reflects stable predictive performance, which was expected given the relatively large size of the training set.

In addition, Fig. 9 shows that the predictions are clustered and the trend in the maximum LTC prediction capability of the models is clearly evident. Looking at Fig. 1 one can clearly see that models trained on the AFLOW dataset do not produce as large values of LTC, which is consistent with the fact that the maximum

<sup>‡</sup> [https://github.com/liudmylaklochko/ParAIsite/tree/main/paper/main/src/results\\_scan](https://github.com/liudmylaklochko/ParAIsite/tree/main/paper/main/src/results_scan).



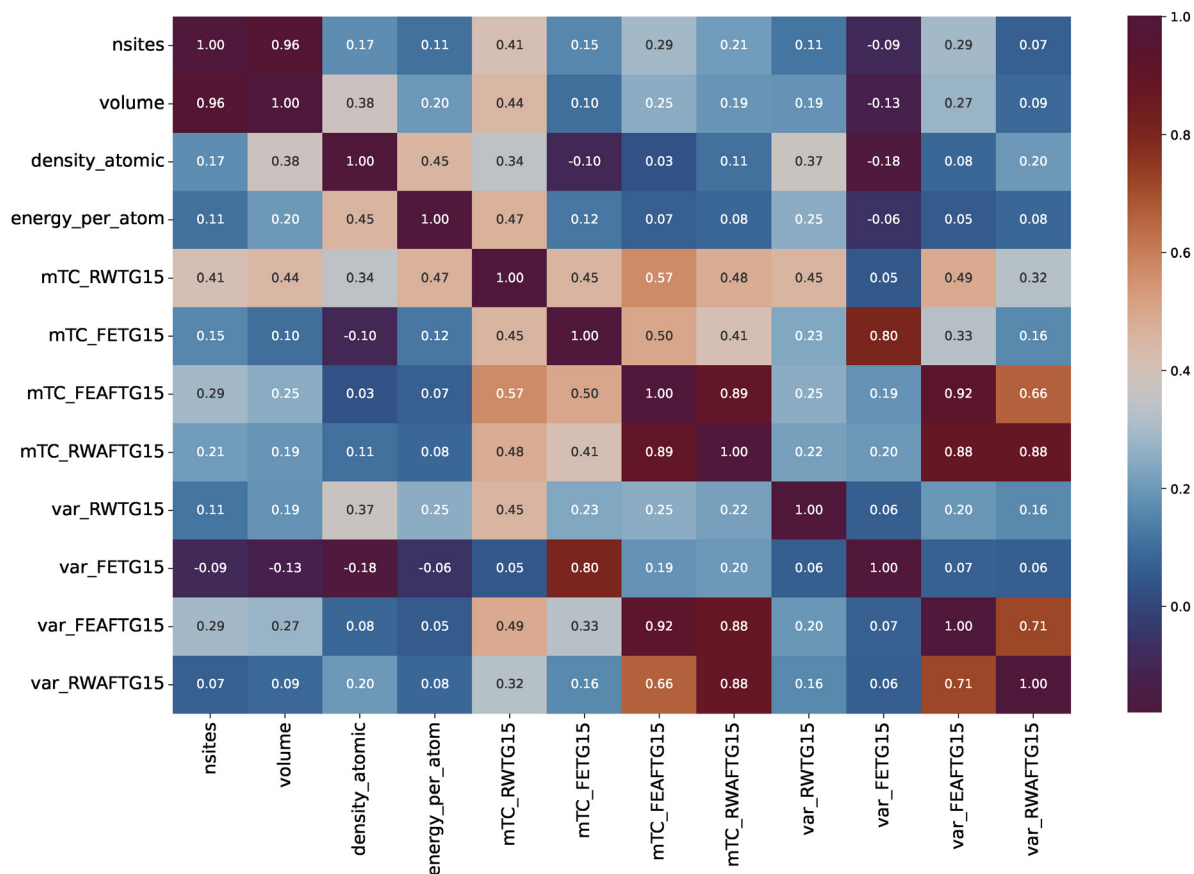


Fig. 7 Correlation (with threshold  $> 0.25$ ) of selected descriptors between steps and material properties. Correlation matrix with no threshold is shown in Fig. S12 in SI to save space in the main paper.

value of LTC for the AFLOW dataset is  $419.73 \text{ W m}^{-1} \text{ K}^{-1}$ . In other words, models trained on this dataset do not generalize beyond the original distribution of values in their training set. Considering this, it is interesting to see that, the fine-tuned models at Step 3 for Togo15 cover the broader distribution of LTC values available in that dataset.

In order to better understand the potential value of the created predictive models when exploring a large database such as the one of the Material Project, we look at identifying factors that appear to be related to predicted LTC values and to their variance. To achieve this, we calculated the Pearson correlation ( $r$ ) (see the heatmap representation in Fig. 7) and  $p$ -values (the probability of obtaining such an extreme correlation under the null hypothesis of no correlation) with a threshold of  $|r| > 0.10$  between mTC, variances of pTC (*i.e.*, variance per material over 9 model runs), and the selected features of the materials for the models trained on Togo15 (see Tables 1–4). It is important to note that the selected features correspond to the relevant descriptors available for the materials included in the training dataset.

What can be first seen from Fig. 7 is a strong correlation between the mTC values and the variances of pTC at each step. This is naturally expected, since the variance would tend to be higher in absolute value for larger values of pTC. Strong correlations can also be observed in the values of mTC across

different steps. This shows that, even though double-stage pretraining had a significant effect on the precision and robustness of the models, it improved over the capabilities of the models it fine-tuned, rather than completely changing them.

Let us start with Step 1 (RWTG15 models) by following results shown in Table 1. As mentioned above, in this step we use a random initialization of weights for ParAI site. The results show that the model captures fundamental geometric properties from the input CIF files, such as volume, number of sites (nsites), and atomic density (density\_atomic). The relationships between these features and lattice thermal conductivity (LTC) are well studied in the literature.<sup>41,42</sup> In particular, the strong correlations with volume ( $r = 0.44$ ) and nsites ( $r = 0.41$ ) indicate that the model identifies the basic thermal conductivity mechanisms, where larger unit cells and a higher number of atomic sites increase the complexity of thermal transport. It is important to note at this step that due to random initialization of the weights, the variance of the prediction reflects how differently the models converge from scratch, rather than the complexity of the material.

However, these effects are less pronounced at Step 2 (FETG15 models) and Step 3 (FEAFTG15 models). The Step 2 models, which were pre-trained on formation energies, appear to capture more complex relationships between material properties and predicted mTC values. The moderate correlations observed with



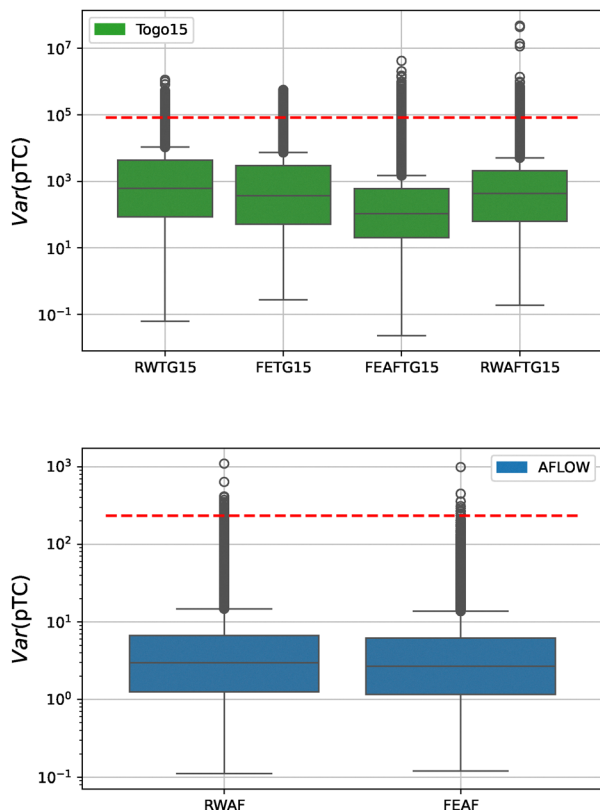


Fig. 8 The variance of the pTC values for models trained on AFLOW and Togo15. Here, the variance refers to the variance of the values of predicted thermal conductivities, pTC, obtained from 9 independent model runs for each of 34k stable material of Material. The dashed line above the box plots indicates the variance of the LTC values in the original dataset. The units of variance are the units of TC ( $\text{W m}^{-1} \text{K}^{-1}$ ).

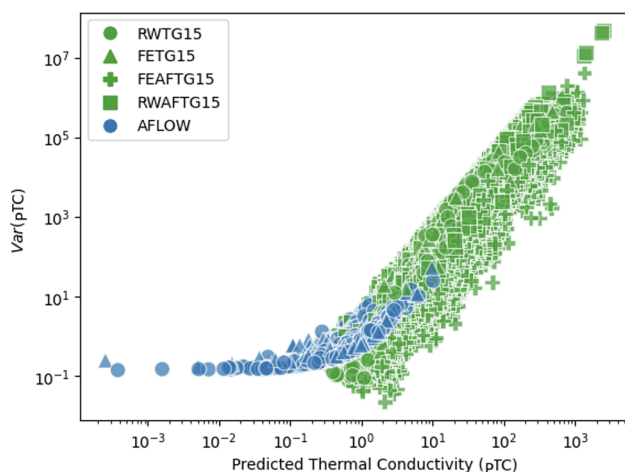


Fig. 9 Group distribution of various predicted thermal conductivities for each models at 3 steps.

certain descriptors are not indicative of direct physical dependencies but rather reflect patterns learned during the multi-stage pretraining process, where general chemical and structural trends were transferred from the broader MEGNet dataset.

Table 1 Correlation analysis at Step 1 (RWTG15 models)

(a) mTC vs. material features

Feature	<i>r</i> -Value	<i>p</i> -Value
Energy_per_atom	0.468	$1.38 \times 10^{-06}$
Volume	0.444	$5.27 \times 10^{-06}$
nsites	0.412	$2.70 \times 10^{-05}$
Density_atomic	0.343	$5.89 \times 10^{-04}$
Density	0.178	$8.03 \times 10^{-02}$
Theoretical	0.154	$1.31 \times 10^{-01}$

(b) Variance of pTC vs. material features

Feature	<i>r</i> -Value	<i>p</i> -Value
Density_atomic	0.370	$1.90 \times 10^{-04}$
Energy_per_atom	0.250	$1.34 \times 10^{-02}$
nelements	-0.216	$3.34 \times 10^{-02}$
Volume	0.187	$6.66 \times 10^{-02}$
Uncorrected_energy_per_atom	-0.177	$8.32 \times 10^{-02}$
Density	0.144	$1.59 \times 10^{-01}$
nsites	0.114	$2.66 \times 10^{-01}$

Table 2 Correlation analysis at Step 2 (FETG15 models)

(a) mTC vs. material features

Feature	<i>r</i> -Value	<i>p</i> -Value
Formation_energy_per_atom	-0.232	$2.20 \times 10^{-02}$
Uncorrected_energy_per_atom	0.190	$6.18 \times 10^{-02}$
nelements	0.163	$1.12 \times 10^{-01}$
nsites	0.155	$1.30 \times 10^{-01}$
Energy_per_atom	0.119	$2.46 \times 10^{-01}$
Is_stable	0.108	$2.90 \times 10^{-01}$
Density_atomic	-0.103	$3.15 \times 10^{-01}$
Volume	0.103	$3.15 \times 10^{-01}$

(b) variance of pTC vs. material features

Feature	<i>r</i> -Value	<i>p</i> -Value
Formation_energy_per_atom	-0.260	$1.02 \times 10^{-02}$
Density_atomic	-0.182	$7.38 \times 10^{-02}$
nelements	0.134	$1.92 \times 10^{-01}$
Volume	-0.127	$2.14 \times 10^{-01}$

In Step 3 (Table 3), it can be seen that the model returns to structural features, with fundamental properties such as volume and nsites continuing to influence thermal conductivity. This step illustrates the benefit of transfer learning: predictions are more robust, variance now reflects physical complexity rather than random initialization, and prior knowledge from Step 2 improves performance.

## 4 Conclusion

In this work, we tested using multiple steps of transfer learning (pre-training and fine-tuning) a machine learning model that we developed to predict the thermal conductivity of crystal compounds. Despite the availability of large datasets for general material properties, databases specifically focused on thermal conductivity are limited, which introduced a challenge to our task. For this reason, the model was trained and



Table 3 Correlation analysis at Step 3 (FEAFTG15 models)

(a) mTC vs. material features		
Feature	r-Value	p-Value
nsites	0.290	$3.99 \times 10^{-03}$
Volume	0.249	$1.38 \times 10^{-02}$
Density	0.233	$2.14 \times 10^{-02}$
nelements	-0.160	$1.18 \times 10^{-01}$
Is_gap_direct	-0.151	$1.40 \times 10^{-01}$
Theoretical	0.144	$1.60 \times 10^{-01}$
Is_stable	0.134	$1.90 \times 10^{-01}$
Band_gap	-0.108	$2.94 \times 10^{-01}$
(b) variance of pTC vs. material features		
Feature	r-Value	p-Value
nsites	0.293	$3.57 \times 10^{-03}$
Volume	0.270	$7.38 \times 10^{-03}$
Density	0.243	$1.63 \times 10^{-02}$
nelements	-0.157	$1.23 \times 10^{-01}$
Is_gap_direct	-0.147	$1.52 \times 10^{-01}$
Theoretical	0.123	$2.30 \times 10^{-01}$
Is_stable	0.109	$2.88 \times 10^{-01}$
Band_gap	-0.105	$3.04 \times 10^{-01}$

Table 4 Correlation analysis at additional check (RWAFTG15 models)

(a) mTC vs. material features		
Feature	r-Value	p-Value
Density	0.228	$2.45 \times 10^{-02}$
nsites	0.209	$3.98 \times 10^{-02}$
nelements	-0.206	$4.31 \times 10^{-02}$
Is_gap_direct	-0.199	$5.13 \times 10^{-02}$
Volume	0.191	$6.12 \times 10^{-02}$
Band_gap	-0.172	$9.13 \times 10^{-02}$
Uncorrected_energy_per_atom	-0.136	$1.84 \times 10^{-01}$
Is_stable	0.108	$2.93 \times 10^{-01}$
Theoretical	0.106	$3.00 \times 10^{-01}$
Density_atomic	0.105	$3.05 \times 10^{-01}$
(b) variance of pTC vs. material features		
Feature	r-Value	p-Value
nelements	-0.255	$1.19 \times 10^{-02}$
Density	0.201	$4.79 \times 10^{-02}$
Density_atomic	0.197	$5.32 \times 10^{-02}$
Uncorrected_energy_per_atom	-0.190	$6.17 \times 10^{-02}$
Is_gap_direct	-0.177	$8.29 \times 10^{-02}$
Band_gap	-0.141	$1.69 \times 10^{-01}$

validated on three different “first-principles” datasets, one being a mixture of the other two. We also used a less precise, but larger dataset as part of one of the transfer learning steps.

Through a series of experiments involving different training configurations, we found that double stage transfer learning, which includes an additional phase of training on external data, proved to be effective in reaching not only better precision but also greater robustness and reduced overfitting. This is true for our dataset that includes a broader range of values for LTC, and some diversity in the type of material included. For this dataset, the error rate (MAPE) decreased consistently as we progressed through the steps, particularly in Step 3. This

indicates that transfer learning, when applied judiciously, can enhance model performance on small datasets.

The results presented in this article would need to be generalized to other datasets with possibly different characteristics. For example, we include as SI tests carried out with a dataset containing a much narrower interval of values for LTC, for a very specific family of materials. In this case, results were mixed. In other words, while double-stage transfer learning shows great promise in improving model accuracy and robustness, its success heavily depends on the dataset's diversity and scope. The results obtained are very promising but also demonstrate that the choice of dataset and training approach is crucial when predicting thermal conductivity. A greater availability of a broader range of datasets of LTC, whether of high precision for training or of approximate precision for pre-training, is therefore expected to enable us to reach better results in the future.

## Author contributions

All authors discussed the results and contributed equally to the manuscript. L. Klochko developed and carried out the implementation of the model and wrote the first versions of the manuscript. M. d'Aquin provided the domain expertise in deep learning and transfer learning necessary for this work and revised the manuscript.

## Conflicts of interest

There are no competing interests to declare.

## Data availability

Data for this article, including sets to train and test, codes, and scripts to visualise them are available at GitHub repository <https://github.com/liudmylaklochko/ParAIsite>.

The data supporting this article have been included as part of the supplementary information (SI). Supplementary information: detailed model parameters, performance metrics such as mean absolute error (MAE),  $R^2$  score, root mean squared error (RMSE), mean absolute percentage error (MAPE), and validation loss for each run. To test assumptions at the foundation of this work, we applied our methodology to 2 more datasets. The results for this verification are also included in SI. See DOI: <https://doi.org/10.1039/d5cp04401d>.

## Acknowledgements

The authors acknowledge Initiative d'Excellence Lorraine (<https://www.univ-lorraine.fr/lue/>) and AIREL center for the financial support of Liudmyla Klochko to work on this project.

## Notes and references

- 1 C. Chen, W. Ye, Y. Zuo, C. Zheng and S. P. Ong, *Chem. Mater.*, 2019, **31**, 3564–3572.



- 2 G. E. Karniadakis, I. G. Kevrekidis, L. Lu, P. Perdikaris, S. Wang and L. Yang, *Nat. Rev. Phys.*, 2021, **3**, 422–440.
- 3 C. K. H. Borg, E. S. Muckley, C. Nyby, J. E. Saal, L. Ward, A. Mehta and B. Meredig, *Digital Discovery*, 2023, **2**, 327–338.
- 4 A. Asensio Ramos, M. C. M. Cheung, I. Chifu and R. Gafeira, *Living Rev. Solar Phys.*, 2023, **20**, 4.
- 5 W.-B. He, Y.-G. Ma, L.-G. Pang, H.-C. Song and K. Zhou, *Nucl. Sci. Tech.*, 2023, **34**, 88.
- 6 A. C. Mater and M. L. Coote, *J. Chem. Inf. Model.*, 2019, **59**, 2545–2559.
- 7 V. Vinod, U. Kleinekathöfer and P. Zaspel, *Mach. Learn.: Sci. Technol.*, 2024, **5**, 015054.
- 8 C. Sarkar, B. Das, V. S. Rawat, J. B. Wahlang, A. Nongpiur, I. Tiewsoh, N. M. Lyngdoh, D. Das, M. Bidarolli and H. T. Sony, *Int. J. Mol. Sci.*, 2023, **24**, 2026.
- 9 B. Zhang, H. Shi and H. Wang, *J. Multidiscip. Healthcare*, 2023, **16**, 1779–1791.
- 10 K. Swanson, E. Wu, A. Zhang, A. A. Alizadeh and J. Zou, *Cell*, 2023, **186**, 1772–1791.
- 11 A. Yaqoob, R. Musheer Aziz and N. K. verma, *Human-Centric Intell. Syst.*, 2023, **3**, 588–615.
- 12 A. Ihalage and Y. Hao, *Adv. Sci.*, 2022, **9**, 2200164.
- 13 Y. Suzuki, T. Taniai, K. Saito, Y. Ushiku and K. Ono, *Mach. Learn.: Sci. Technol.*, 2022, **3**, 045034.
- 14 Y. Zhang and C. Ling, *npj Comput. Mater.*, 2018, **4**, 25.
- 15 Y. Qian, Y. Luo, Y. Li, T. Xiong, L. Wang, W. Zhang, S. Gang, X. Li, Q. Jiang and J. Yang, *Chem. Eng. J.*, 2023, **467**, 143433.
- 16 T. Huang, T. Wang, J. Jin, M. Chen and L. Wu, *Chem. Eng. J.*, 2023, **469**, 143874.
- 17 Y. Ouyang, C. Yu, J. He, P. Jiang, W. Ren and J. Chen, *Phys. Rev. B*, 2022, **105**, 115202.
- 18 Y. Ouyang, C. Yu, G. Yan and J. Chen, *Front. Phys.*, 2021, **16**, 43200.
- 19 C. E. Calderon, J. J. Plata, C. Toher, C. Osos, O. Levy, M. Fornari, A. Natan, M. J. Mehl, G. Hart, M. Buongiorno Nardelli and S. Curtarolo, *Comput. Mater. Sci.*, 2015, **108**, 233–238.
- 20 S. Kirklin, J. E. Saal, B. Meredig, A. Thompson, J. W. Doak, M. Aykol, S. Rühl and C. Wolverton, *npj Comput. Mater.*, 2015, **1**, 15010.
- 21 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 Mater.*, 2013, **1**, 011002.
- 22 K. Choudhary, K. F. Garrity, A. C. E. Reid, B. DeCost, A. J. Biacchi, A. R. Hight Walker, Z. Trautt, J. Hattrick-Simpers, A. G. Kusne, A. Centrone, A. Davydov, J. Jiang, R. Pachter, G. Cheon, E. Reed, A. Agrawal, X. Qian, V. Sharma, H. Zhuang, S. V. Kalinin, B. G. Sumpter, G. Pilania, P. Acar, S. Mandal, K. Haule, D. Vanderbilt, K. Rabe and F. Tavazza, *npj Comput. Mater.*, 2020, **6**, 173.
- 23 M. Blanco, E. Francisco and V. Luaña, *Comput. Phys. Commun.*, 2004, **158**, 57–72.
- 24 C. Toher, J. J. Plata, O. Levy, M. de Jong, M. Asta, M. B. Nardelli and S. Curtarolo, *Phys. Rev. B: Condens. Matter Mater. Phys.*, 2014, **90**, 174107.
- 25 Z. Li, M. Li, Y. Luo, H. Cao, H. Liu and Y. Fang, *Digital Discovery*, 2025, **4**, 204–210.
- 26 G. Qin, Y. Wei, L. Yu, J. Xu, J. Ojih, A. D. Rodriguez, H. Wang, Z. Qin and M. Hu, *J. Mater. Chem. A*, 2023, **11**, 5801–5810.
- 27 R. Juneja, G. Yumnam, S. Satsangi and A. K. Singh, *Chem. Mater.*, 2019, **31**, 5145–5151.
- 28 Y. Luo, M. Li, H. Yuan, H. Liu and Y. Fang, *npj Comput. Mater.*, 2023, **9**, 1–11.
- 29 P. Paliwal and A. Alam, *Phys. Rev. Mater.*, 2025, **9**, 123804.
- 30 H. Yang, C. Hu, Y. Zhou, X. Liu, Y. Shi, J. Li, G. Li, Z. Chen, S. Chen, C. Zeni, M. Horton, R. Pinsler, A. Fowler, D. Zügner, T. Xie, J. Smith, L. Sun, Q. Wang, L. Kong, C. Liu, H. Hao and Z. Lu, *MatterSim: A Deep Learning Atomistic Model Across Elements, Temperatures and Pressures*, 2024, <https://arxiv.org/abs/2405.04967>.
- 31 S. Y. Lee, H. Kim, Y. Park, D. Jeong, S. Han, Y. Park and J. W. Lee, *Forty-second International Conference on Machine Learning*, 2025.
- 32 F. Zhuang, Z. Qi, K. Duan, D. Xi, Y. Zhu, H. Zhu, H. Xiong and Q. He, *A Comprehensive Survey on Transfer Learning*, 2019, <https://arxiv.org/abs/1911.02685>.
- 33 S. Kong, D. Guevarra, C. P. Gomes and J. M. Gregoire, *Appl. Phys. Rev.*, 2021, **8**, 021409.
- 34 N. Hoffmann, J. Schmidt, S. Botti and M. A. L. Marques, *Transfer learning on large datasets for the accurate prediction of material properties*, 2023, <https://arxiv.org/abs/2303.03000>.
- 35 A. Seko, A. Togo, H. Hayashi, K. Tsuda, L. Chaput and I. Tanaka, *Phys. Rev. Lett.*, 2015, **115**, 205901.
- 36 A. Togo, L. Chaput and I. Tanaka, *Phys. Rev. B: Condens. Matter Mater. Phys.*, 2015, **91**, 094306.
- 37 A. Togo, L. Chaput, T. Tadano and I. Tanaka, *J. Phys.: Condens. Matter*, 2023, **35**, 353001.
- 38 L. Chaput, *Phys. Rev. Lett.*, 2013, **110**, 265506.
- 39 A. Dunn, Q. Wang, A. Ganose, D. Dopp and A. Jain, *npj Comput. Mater.*, 2020, **6**, 138.
- 40 A. Y.-T. Wang, S. K. Kauwe, R. J. Muddock and T. D. Sparks, *npj Comput. Mater.*, 2021, **7**, 77.
- 41 L. H. Liang and B. Li, *Phys. Rev. B: Condens. Matter Mater. Phys.*, 2006, **73**, 153303.
- 42 L. Jiang, C. Liang, Z. Cheng, X. Wang, Y. Hao, X. Wang, Y. Liu, Y. Song and L. Wang, *J. Energy Storage*, 2025, **113**, 115633.

