

# Extensive Attention Mechanisms in Graph Neural Networks for Materials Discovery

Guojing Cong\*, Talia Ben-Naim\*, Victor Fung\*, Anshul Gupta†, Rodrigo Neumann †, Mathias Steiner †

\* Oakridge National Laboratory, Oak Ridge, TN, 37830

† IBM Research, Yorktown Heights, NY, 10598

**Abstract**—We present our research where attention mechanism is extensively applied to various aspects of graph neural networks for predicting materials properties. As a result, surrogate models can not only replace costly simulations for materials screening but also formulate hypotheses and insights to guide further design exploration. We predict formation energy of the Materials Project and gas adsorption of crystalline adsorbents, and demonstrate the superior performance of our graph neural networks. Moreover, attention reveals important substructures that the machine learning models deem important for a material to achieve desired target properties. Our model is based solely on standard structural input files containing atomistic descriptions of the adsorbent material candidates. We construct novel methodological extensions to match the prediction accuracy of state-of-the-art models some of which were built with hundreds of features at much higher computational cost. We show that sophisticated neural networks can obviate the need for elaborate feature engineering. Our approach can be more broadly applied to optimize gas capture processes at industrial scale.

## INTRODUCTION

We explore applying AI techniques for accelerated discovery in material science. Designing new materials is a major research challenge for its technical complexity and potential global impact. For each application area, e.g., better alloys for transportation vehicles and new porous carbon forms for warm-house gas capturing, there are millions of candidates. Identifying the most suitable material for a given scenario requires screening for an inordinate number of material instances. While experimental screening is unviable by its high cost, pure computational screening based on physics-based simulations presents significant challenges due to the computational complexity. An AI-augmented framework where machine learning technologies are leveraged to learn from prior data and propose insights can help accelerate this process.

Let us first consider a concrete example of screening and designing materials for CO<sub>2</sub> capturing. Among the classes of solid-state CO<sub>2</sub> adsorbents, Metal-Organic Frameworks (MOF) stand out for their chemical diversity and customisable porous structure [20]. Since it is not feasible to perform molecular-level simulations for millions of materials candidates, novel hierarchical screening methods for scientific inference in materials research are required. Existing materials screening approaches contain a top layer in which rapid geometric and topological characterizations of the materials are deployed to eliminate samples having less favorable properties. Previous studies have employed a variety of geometric, topological and chemical descriptors to analyse the carbon dioxide

adsorption performance of nano-porous materials. Geometric features can include the accessible pore surface area and volume, pore diameter metrics and crystal density [1, 7, 8, 18, 10, 30, 9, 13, 16]. Topological representations of the materials can be generated using concepts extracted from persistent homology and clustering of high performance materials with similar topologies can be performed [17, 18, 30, 6, 16]. These topological methods can also be combined with chemical information to enhance correlations between the descriptors and the material adsorption properties [23]. Recently, in a comprehensive article by Moosavi *et al.* [20], a set of 165 geometric and chemical material descriptors were built and combined with supervised machine learning techniques to train algorithms for predicting the adsorption performance of samples from several large MOF databases. By using this approach, only the properties of the most promising material candidates are subsequently calculated using molecular-level physics simulation, significantly reducing discovery time and computational cost. This approach is promising; however, these top layer topological and geometric descriptors can only classify samples for elimination or further study. Furthermore, they neglect the intricate chemical interactions between various atomic species present in the nanopore framework and gas phase, thus limiting the effectiveness of such descriptors as a screening tool. Also there is very little insight revealed by such approaches as to what sub-structures in the materials that lend to desirable adsorption properties.

Graph neural network (GNN) represent the topological structures of materials by graphs. GNN has been shown to learn complex features from raw structure data with only atomic numbers and the position of each atom. The topological structure of a material is naturally incorporated (e.g., as an adjacency matrix) in the graph for material property predictions. GNN potentially frees the material scientist from having to calculate a multitude of features if it is powerful enough. GNNs have been developed for predicting electronic properties of crystalline materials, such as formation energy, absolute energy and band gap. They learn to predict material properties from the atoms in the crystal and their nearest-neighbor connections. The Crystal Graph Convolutional Neural Network (CGCNN) [29] chose the distance between atoms to represent the edges in the crystal graph. In a derived work [22], the original CGCNN formulation was shown to outperform the accuracy of several other other descriptor-based regression models in a high-throughput material screening

study of electronic properties. Other GNN models such as improved CGCNN (iCGCNN) model [21] further improved the predictive performance. Recent efforts increasingly incorporate physics-informed information such as spatial distance and direction into features fed to a GNN. The intuition is that the relative spatial position of atoms in the microstructure is closely related to the charge interaction between them and thus affects macroscopic properties of the material. The Materials Graph Network (MEGNet) [2] introduced manual features that included topological distance and spatial distance. The directional message passing neural network (DimeNet) [15] encoded the directional information into GNN models for molecular materials. It is argued in [4] that many GNN based works have provided incomplete spatial geometrical information, such as just the distance, or one additional angle, which makes the models unable to learn the complete local geometrical relationship between atoms have significantly improved prediction accuracy.

The current trend in AI methods (e.g., Random Forest and GNN) for studying materials increasingly incorporates elaborate physics-based features for property prediction. We argue and demonstrate that sophisticated architectures such as extensively attention mechanism in GNNs obviate the need to derive new intricate features for accurately predicting the properties of materials. Analogous to computer vision or machine translation, where feature engineering is rendered unnecessary by deep convolutional neural networks (CNNs), we show it is possible to achieve similar result with GNNs for material discovery. By introducing sophisticated attention mechanism, our GNNs surpass prior state-of-the-art prediction performance for materials while using fewer features. In addition, we show that attention mechanism can also reveal valuable insights for future material design.

#### FEATURES USED IN CURRENT GNNs FOR MATERIALS

GNNs like PhysNet [24] and DimeNet incorporate features calculated from Gaussian radial function and Fourier-Bessel basis function to encode spatial and chemical information. In DimeNet, it is shown that even utilizing the simplest wave function (the solution of Schrödinger equation under infinite sphere potential) can produce features that significantly improve the performance of the GNN model. GeoCGNN [4] encodes the geometric structure information, which is given by plane waves. They are the natural basis in the nearly-free-electron approximation. Local orbitals are used as a basis to carry out a full self-consistent solution of independent particle equations. GeoCGNN employs features of the form of mixed basis: the Gaussian radial basis and the plane wave. It is argued that since plane waves are eigenfunctions of the Schrödinger equation with constant potential, they are the natural basis in the nearly-free-electron approximation.

In one of the few attempts to merge the use of CGCNN with physics-based gas adsorption prediction [27], the authors applied high-throughput Grand Canonical Monte Carlo (GCMC) calculations to 10,000 entries in the CoRE-2019 materials database [5] to build a training dataset for a machine learning

classifier that distinguishes materials into two performance categories, i.e. high- and low-performance, relative to a given threshold. They latter applied it as a pre-screening step for selecting the most promising candidates for further analysis out of a much larger hypothetical MOF database [28] containing more than 320,000 materials. The CGCNN classifier was able to identify approximately 28,000 materials, a small fraction of the database, as deserving of further investigation via GCMC simulations.

#### ATTENTION-EVERYWHERE CRYSTAL GRAPH CONVOLUTIONS

We introduce our Attention-Everywhere Crystal Graph Convolutional Neural Network (AE-CGCNN) in comparison to some of the existing popular approaches, for example, CGCNN [29] and GeoCGNN [4]. In order to build a predictive model based on graph convolutions the material structure is represented by a graph that encodes both atomic information and bonding interactions between atoms. The vertices of the graph represent atoms, and the edges represent the atoms' nearest neighbors, determined by the Euclidean distance between the atoms in 3D Cartesian space; i.e., atoms closer than a threshold distance are considered neighbors and will have an edge connecting them. In fact, there are multiple edges between the same pair of neighbors to represent periodicity in the crystal. Feature vectors corresponding to vertices and edges encode atom and bond properties, respectively. A convolutional neural network is built atop the graph to automatically extract representations that are optimum for predicting target physical properties.

#### Notation

$G = (V, E)$  denotes an undirected multigraph defined by a vertex set  $V = 1, 2, \dots, n$  representing the atoms of a crystal with an  $n$ -atom unit cell and edges  $(i, j)_k$  representing connections between atoms  $i$  and  $j$ , where  $i, j \in V$ .  $G$  allows multiple edges between the same pair of vertices to represent periodicity of the crystal structure; therefore,  $k$  represents the  $k$ -th bond between  $i$  and  $j$ . A feature vector  $v_i$  is associated with each vertex  $i$  and a feature vector  $e_{i,j,k}$  is associated with each edge  $(i, j)_k$ .

#### Graph Convolution

Graph Neural Networks (GCN) apply convolutions to features according to the relationship defined by the graph. The vanilla GCN by [14] is a localized first-order approximation of spectral graph convolutions [14].

$$v_i^{l+1} = \sigma \left( b^l + \sum_{v_j \in N(v_i)} C_{v_i, v_j} v_j^l W^l \right) \quad (1)$$

Here  $v_i^l$  is the hidden feature at vertex  $v_i$  at convolution level  $l$ . For each vertex  $v_i$ , its hidden feature at next level  $v_i^{l+1}$  is computed by aggregating hidden states of its neighbors  $v_j \in N(v_i)$  and possibly of the incident edges as well.  $C_{v_i, v_j}$  is a constant,  $b^l$  is the bias vector, and  $W^l$  is the weight matrix;

the latter two are parameters to be learned. A prominent feature of GCN that is shared by its later extensions is the message passing mechanism that propagates computing through the connections among nodes in the graph.

Graph convolutions for crystal structures have a few major differences from typical GCNs. There can be multiple edges between a pair of nodes, and also the incident edges for a node may change depending on how large a neighborhood is considered according to the strength of the bonds.

Existing GCNs for crystal structures emphasizes the importance of edge features in learning. At level  $l + 1$ , the feature vector  $v_i^{l+1}$  for vertex  $i$  is computed by a vertex convolution function  $\mathcal{C}_v$  using the previous level's features from  $i$ 's neighbors and from all edges incident on  $i$  as follows:

$$v_i^{l+1} = \mathcal{C}_V (v_i^l, v_j^l, \{e_{i,j,k}\}). \quad (2)$$

Typically  $\mathcal{C}_v$  is a linear combination of all features involved with some nonlinear activation function. As long as it aggregates the features from adjacent vertices and edges,  $\mathcal{C}_v$ 's definition can be chosen to suitably meet the requirements and the characteristics of the application at hand. In the original CGCNN paper, [29] used the following definition:

$$\mathcal{C}_v (v_i^l, v_j^l, \{e_{i,j,k}\}) = v_i^l + \sum_{j,k} \sigma(z_{i,j,k}^l W_c^l + b_c^l) \odot g(z_{i,j,k}^l W_s^l + b_s^l), \quad (3)$$

where  $\odot$  denotes element-wise multiplication,  $\sigma$  here is the Sigmoid function,  $g$  is a nonlinear activation function, and  $z_{i,j,k}$  is the concatenation of  $v_i$ ,  $v_j$ , and  $e_{i,j,k}$ .  $W_c$  and  $b_c$  are the shared convolution weight matrix and bias vector, and  $W_s$  and  $b_s$  are the self-weight matrix and self-bias vector, respectively.

Note the subtle differences between Equations 2 and 3, and Equation 1. Equations 2 and 3 compute  $v_i^l$  for a vertex  $i$  by first deriving  $|\{e_{i,j,k}\}|$  features using both edge features and node features, and then combine them.

### Edge Convolution

Edge features are obviously critical to learning the properties of a crystal structure. Notice that Equation 3 for CGCNN uses the edge features as input to the convolution, but these features themselves remain unchanged through the convolution steps. As shown in the Results, introducing parameters (weights) associated with the edges and evolving edge features through an edge-convolution function  $\mathcal{C}_e$  yielded a significant improvement in prediction accuracy.

We introduce convolution to the edges as follows:

$$e_{i,j,k}^{l+1} = \mathcal{C}_e (e_{i,j,k}^l, v_i^l, v_j^l), \quad (4)$$

where

$$\mathcal{C}_e (e_{i,j,k}^l, v_i^l, v_j^l) = \Gamma[e_{i,j,k}^l \oplus v_i^l \oplus v_j^l]. \quad (5)$$

Now a convolution layer produces two kinds of outputs, one feature for each vertex, and one feature for each edge.

### Attention

Attention mechanism has been shown to be a powerful mechanism in deep learning, especially in learning with sequences. In deep neural networks, attention mimics cognitive attention and can capture spacial or temporal interdependence between the most relevant parts of the input. In the graph neural networks, an attention function is learned simultaneously with the features [26].

An elementary form of attention (or more precisely, a gated mechanism) is implemented in CGCNN shown in Equation 3 by the Sigmoid function. Note the added  $v_i^l$  is similar to the concept of pass through connection first introduced in ResNet [12].

We introduce formally the attention mechanism as follows:

$$v_i^{l+1} = \mathbf{Softmax}_{j,k} \{z_{i,j,k}^l W_f^l + b_f^l\} \odot g(z_{i,j,k}^l W_s^l + b_s^l) \quad (6)$$

Here in producing a feature for  $v_i$ , the neural network attends to all its neighbors' features. Since the neighbors' features are computed using edge features as well, this mechanism also implicitly attends to the edge features.

### Multi-Head attention

Multi-head attention has been shown to improve the performance of graph convolution [26]. With multi-head attention, multiple attention mechanisms concurrently attend to the neighbors and incident edges of a vertex. We adapt multi-head attention (with  $M$  heads) to our graph convolution as follows

$$v_i^{l+1} = \frac{1}{M} \sum_{m=1}^M S \odot g(z_{i,j,k}^l W_s^{l,m} + b_s^{l,m}) \quad (7)$$

where  $S = \mathbf{Softmax}_{j,k,m} \{z_{i,j,k}^l W_f^{l,m} + b_f^{l,m}\}$

### Attention to Combine $M$ Heads

Instead of simply averaging the independent attention outputs as has been done in prior work and shown in Equation 7, we introduce another layer of attention mechanism to combine the computed hidden/output features. Intuitively, multiple attention heads allow for attending to different neighbors or bonds to compute the feature of an atom. Their combination also benefits from weighted contribution from each head.

$\frac{1}{M}$  becomes  $\mathbf{Softmax}(\sum_{m=1}^M S \odot g(z_{i,j,k}^l W_s^{l,m} + b_s^{l,m}))$  in Equation 7 if self attention is applied; otherwise we can introduce additional output feature to compute attention for  $v_i^{l+1}$ .

### Exploration with Deep GNNs

Vanilla graph convolution suffers from over-smoothing when the networks become deep. Recently proposed architectures address this deficiency with various techniques [3, 19]. For example, initial residual and identity mapping are introduced to Graph Convolution Network to overcome over-smoothing, and achieve good performance on standard benchmarks [3]. Unfortunately such techniques do not bring tangible improvement in predicting CO<sub>2</sub> adsorption in our experiments. Neither are they noticeably effective for predicting formation

energy. This could be in part due to the highly regular structure of crystal graph and the importance of the edge features.

Prior work in using graph convolutions to predict the material properties of crystal structures explore incorporating various physical atomic interaction expressions [2, 25, 4], including electronic state wave functions, directly into the neural network. Unfortunately, such techniques cannot be readily applied to our use case as all relevant pairwise atomic interactions in the adsorption simulation are between individual carbon dioxide gas molecules and individual framework atoms. While the latter are represented by nodes in our network, the former are not explicitly represented and thus their relative proximity and resultant attractive/repulsive interaction cannot be computed.

## EXPERIMENTAL RESULTS

We report the performance of AE-CGCNN for predicting two properties: CO<sub>2</sub> gas uptake for MOFs and the formation energy for crystalline materials.

### CO<sub>2</sub> Gas Uptake

The training dataset was built from existing Crystallographic Information File (CIF) databases [20]. The CoRE-19 database [5] contains approximately 10,000 MOF materials, while the BW-20K database is a subset of the original BW-DB database [1] containing only 20,000 MOFs.

The target adsorption properties of the materials were calculated using Physics-based GCMC simulations that also take the CIF as input. The GCMC simulation consists of a series of steps in which "Monte Carlo moves" are attempted. Suitable Monte Carlo moves include molecular translations, insertions, deletions, rotations and swaps. Moves that lower the energy are favored and those that increase the energy are accepted with a probability given by the Boltzmann distribution. The energy of the system is calculated after each move as the combination of the Lennard-Jones interaction and the Coulomb electrostatic energy. After a sufficiently high number of Monte Carlo steps, the number of molecules in the simulation cell equilibrates and the gas uptake at that temperature and pressure can be computed. The CO<sub>2</sub> gas uptake for each material was calculated at room temperature for high (16 bar) and low (0.15 bar) pressures, simulating adsorption and desorption conditions, respectively.

In all our training experiments, we use 4 layers of graph convolution. The original input features are one-hot encoding vectors for the atomic specie numbers. These vectors are first embedded into 64 dimensional vectors before graph convolution. All hidden features for the vertices and edges are 64 and 42 dimensional, respectively. We use three attention heads and ReLU for activation. Average pooling is used to pool from the atom features into a feature for the corresponding crystal. A two-layer multilayer perceptron (MLP) is used to predict the adsorption of the crystal. The fully connected layers are  $64 \times 128$ , and  $128 \times 1$ , respectively. In constructing the graph, the neighbors of each atom to be the nearest 12 atoms in the cutoff radius 8 Å. The optimizer used is Adam with

learning rate  $10^{-3}$  and weight decay  $5 \times 10^{-4}$ . We train for 50 epochs with the training set, and report the test accuracy on the test set with the model that achieved the best validation accuracy on the validation set. We randomly split the datasets into training, validation, and test sets. The reported results are averaged over 5 splits.

*Mean Absolute Error and Spearman Rank Correlation Coefficient:* The infusion of our enhancements into the CGCNN significantly improves the training performance of the network. Figure 1 shows the evolution of training (MSE) loss with the number of training epochs for the various GCNs for the BW-20K dataset at pressure 0.15 bar. In our experiment we start with CGCNN, and then introduce edge convolution, attention mechanism, atom charge information and geometric features, and then multi-head attention incrementally to the plain convolution used in CGCNN. It is clear from the plot that as the neural network becomes more complex and/or more new physics features are included, better training results (lower losses) are achieved. Figure 2 shows the evolution of validation MSE loss with the number of training epochs.

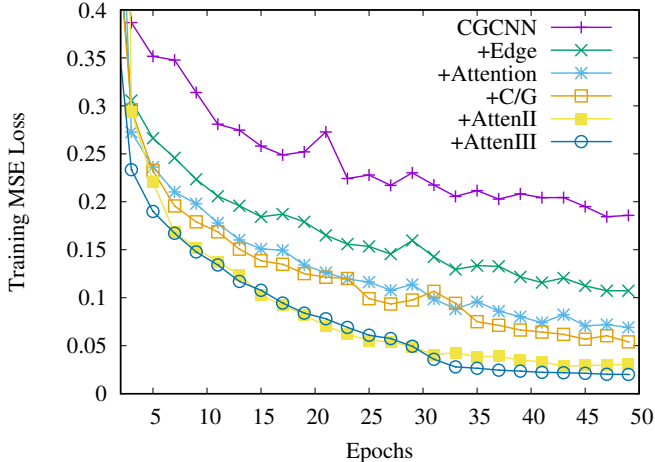


Fig. 1: A plot illustrating the corresponding reduction of training (MSE) loss with the addition of each enhancement encoded in the neural network. These include the original CGCNN, edge convolution, attention (+Attention), the atom partial charge and geometric information (lengths and angles), multi-head attention mechanism (+ATTENII), and attention to combine the heads (+ATTENIII). To show details, the plot starts at epoch 2 (with epoch 1 having a very high loss). The training losses after 50 epochs are recorded in Table I.

The performance of CGCNN on this data set, and of the network with each incremental additional enhancement, is detailed in Table I. The networks show a monotonic increase in performance in the training phase, as dictated by an overall reduction of 88% in the training MSE (mean squared error) and 63% in training MAE (mean average error). This trend is replicated with a reductions of 21% in the test MAE and a 2% rise in the test Spearman ranking (SRCC). Note that we deployed a train-test-validation split of 80:10:10 to produce

Metric	CGCNN	+Edge	+Attention	+C/G	+ AttenII	+AttentionIII/AE-CGCNN
train MSE	0.197	0.110	0.066	0.058	0.023	0.020
train MAE	0.304	0.231	0.184	0.175	0.112	0.105
test MAE	0.304	0.260	0.256	0.256	0.239	0.235
test SRCC	0.923	0.931	0.932	0.932	0.943	0.948

TABLE I: Table assessing the performance of various crystal graph convolutional neural networks on the BW20K 0.15 bar data set [20]. A 80-10-10 train-validation-test split is deployed throughout. Note that C/G represents the addition of atomic partial charges and geometric information while Attention III denotes the final multi-head attention enhancement.

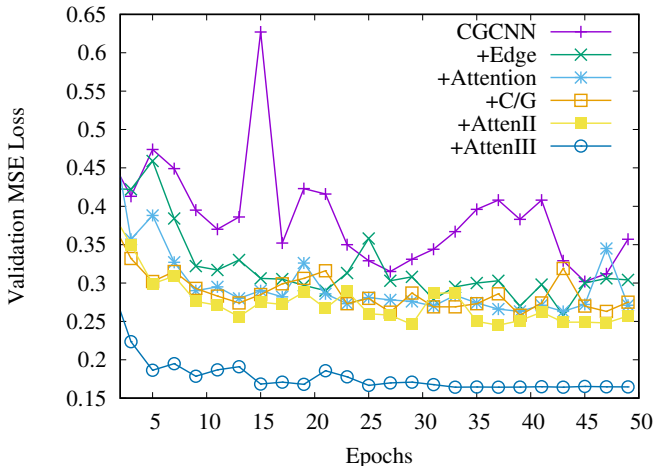


Fig. 2: A plot illustrating the corresponding reduction of validation (MSE) loss with the addition of each enhancement encoded in the neural network.

these results recorded in the first five columns of Table I.

In Table II, we contrast the test performance of our most enhanced crystal graph convolutional neural network (AE-CGCNN) to the previous best results obtained for this data set using the application of classical machine learning techniques to geometric and chemical material properties [20], for all combinations of temperatures and pressures available in this reference. The two middle columns employ the approximate 70-30 test-train split used by [20] while the rightmost column corresponds to the 80-10-10 test-train-validation split used earlier in Table I. Across the four distinct data sets corresponding to two sets of materials simulated under low (0.15 Bar) and high (16 Bar) pressure conditions, one can observe similar performance between our AE-CGCNN model and the framework outlined by [20] for the first three data sets. This observation holds for both absolute (test MAE) and relative (SRCC) metrics. Both of these performance metrics also improve significantly when one deploys the 80-10-10 test-train-validation split compared to the 70-30 test-train split used in the referenced study.

#### Formation Energy Prediction without Elaborate Features

Several GNNs have been developed and evaluated for formation energy prediction. We compare with one of the most recent GNNs, GeoCGNN [4], using different amounts of data from the Material Project (MP). GeoCGNN has been shown

Metric	Moosavi ( $s_m$ )	AE-CGCNN ( $s_m$ )	AE-CGCNN
BW-20K, 0.15bar			
test MAE	0.3	0.293	0.239
test SRCC	0.92	0.920	0.943
BW-20K, 16bar			
test MAE	0.74	0.73	0.62
test SRCC	0.99	0.99	0.993
CoRE-2019 0.15Bar			
test MAE	0.54	0.54	0.49
test SRCC	0.84	0.83	0.85
CoRE-19, 16bar			
test MAE	0.58	1.0	0.93
test SRCC	0.98	0.92	0.93

TABLE II: Table assessing the performance of our attention-everywhere crystal graph convolutional neural network (AE-CGCNN) for four adsorption data sets and two test-train splits.

to out perform other networks such as CGCNN, MEGNet [2], and iCGCNN [21]. For the comparison, we use the published source code of GeoCGNN and test on the latest MP dataset.<sup>1</sup>

GeoCGNN uses an encoder as an attention mask to transform the discrete distance vector to two sets of orthogonal basis functions in the solution space of Schrödinger’s equation. An important feature used in GeoCGNN is motivated by the Bloch theorem in the form of plane waves to describe periodic structure of crystals. Our hypothesis is that such elaborate features that only expert physicists can come up with are probably not necessary for accurate prediction performance if the GNNs are sophisticated enough. As we have seen that AE-CGCNN contains extensive attention mechanism and it is interesting to see if that compensates for the absence of elaborate features. We deliberately remove the plane wave features from the input to our AE-CGCNN. The results in comparison to GeoCGNN are shown in Table. III.

train/validation/test split	GeoCGNN	AE-CGCNN
300/38/38	0.389	0.4025
500/63/63	0.2582	0.2767
1000/125/125	0.2092	0.2091
3,478/434/434	0.1480	0.1570
6,956/869/869	0.079	0.087
13,912/1,739/1,739	0.0603	0.0616
27,824/9,274/9,274	0.046 ± 0.001	0.043 ± 0.001
111,496/13,936/13,936	0.0285 ± 0.001	0.0248 ± 0.001

TABLE III: Comparison of AE-CGCNN and GeoCGNN in predicting formation energy. MAE is reported.

<sup>1</sup>The performance of GeoCGNN in our experiment is slightly different than reported in [4] as the dataset and train, validation, and test split are different

In Table. III, for both configurations, AE-CGCNN outperforms GeoCGNN in test MAE, even when it uses fewer features (to be exact, AE-CGCNN uses half of the input features of GeoCGNN). This validates our hypothesis that sophisticated neural network architectures may obviate the need of elaborate feature engineering for learning a surrogate model for materials.

#### ATTENTION AS A MECHANISM TO IDENTIFY IMPORTANT SUB-STRUCTURES

Average Pooling is typically used in current GNNs (e.g., CGCNN and GeoCGNN) for material study to derive crystal properties from individual atom properties. We further introduce another attention mechanism to attend to different atoms for their relative contributions to the crystal property. We extend Equation 7 for computing  $v_i^{l+1}$  such that a pair of features are computed:  $(v_i^{l+1}, a_i^{l+1})$ , and  $a_i^{l+1}$  is used to compute the attention on  $v_i^{l+1}$  as follows

$$\mathbf{c} = \sum_{x=1}^N \text{Softmax}(a_i^L) v_i^L \quad (8)$$

The introduction of  $a_i^{l+1}$  is expected to increase the capacity of the GNN model for learning and thus improve the predictive performance. Perhaps more importantly, it provides important hints as to when deciding the target property of the materials, which atoms and sub-structures the model is paying attention to.

With attention pooling, we are able to inspect and derive important sub-structures when the surrogate model makes a prediction for a desired property. As an example, Figure 3 shows for a certain MOF structure, AE-CGCNN considers two atoms (high-lighted by red arrows) to be critical for achieving certain  $CO_2$  adsorption. Intuitively, one would imagine the important atoms in a MOF structure to be the metal ones, shown as the two large centrally placed balls in Figure 3. In practice, however, from our analysis of the CoRE-19 and BW-20K datasets, AE-CGCNN almost never considers the metal atoms as important in making the adsorption prediction. One explanation could be the metal atoms are usually buried inside the cell and thus does not interact much with the gas.

Statistics on the amount of importance placed on different types of atoms may bring further insight. We note that such a capability goes beyond simply providing an accurate surrogate model for fast screening. It is critical in understanding the prediction of GNNs and identifying important motifs for material design. The work in this area is still in its very early stage. Further progress demands combining graph learning and graph analysis to answer harder questions such as, is attention mostly placed on rings or external leaf nodes of the graph?

#### DISCUSSION AND CONCLUSION

We introduce Attention-Everywhere Crystal Graph Convolutional Neural Network which rivals the performance of the state-of-the-art machine learning approach in the literature for two different tasks [20, 4].

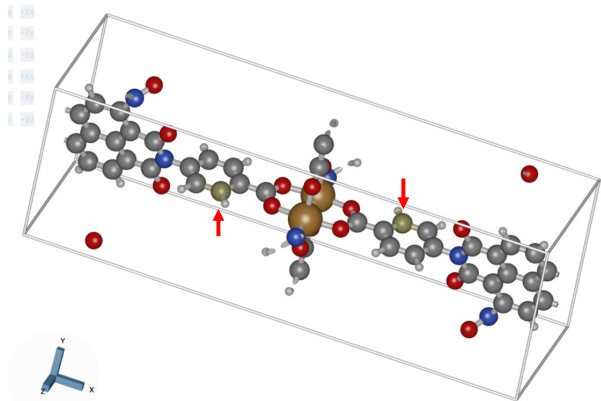


Fig. 3: A plot illustrating the importance of the atoms in predicting the MOF’s  $CO_2$  adsorption

In the gas absorption case, the authors construct a complex feature set of 165 chemical and geometric descriptors before deploying supervised machine learning regression techniques to train their model. Our simpler model does not require the computation of a feature set and relies purely on the information already encoded within the input file ingested by Grand Canonical Monte Carlo simulations.

This result is achieved despite the fact that our model does not account for atomic chemical specificity. We classify each atom as charged or not charged, but this is a very crude categorization criterion. Future work needs to be done to develop a technique to incorporate the inter-atomic potential between the carbon dioxide molecules and the framework atoms, into our network. In particular, this would require the network to provide an accurate prediction of the final positions and orientations of the adsorbed gas molecules in the solid crystalline materials. The accomplishment of this task would represent a significant step beyond existing graph convolution models for crystalline materials which only account for the constituent atoms of the solid crystals and neglect any interaction with unbound gas or liquid molecules. The set of pairwise inter-atomic interactions is the last significant piece of physical information encoded within the simulation input file that is not utilized in our current model.

The effect of this lack of chemical specificity inherent in our AE-CGCNN model may explain the variations in its overall performance on the data sets presented in Table II. In the benchmark paper, the relative importance of the 165 chemical and geometric descriptors to [20] model indicate that chemical specificity is more important for the CoRE-2019 data set relative to the BW-20K data set. These chemical features are especially influential at lower as opposed to higher pressures. The observations could explain why our AE-CGCNN metrics in Table II for  $CO_2$  adsorption are better for the BW-20K data set over the CoRE-2019 data set, and also why the network performance is superior at pressures of 0.15 bar as opposed to the 16 bar adsorption simulation results.

Future work beyond refining the chemical description of

constituent atoms within the network, could include the calibration of the model to additional adsorption target properties. In particular, investigating the evolution of these properties within varying pressure or temperature ranges would allow us to extract engineering metrics which are vital for process engineering optimisation [11].

Our experiment with using fewer input features illustrates the need for more in-depth study of sophisticated neural networks vs. elaborate features based on physics insights.

AE-CGCNN can identify important motifs in structures for desired target properties. At this stage this work reports simply the insights from the data and machine learning approaches. Obviously further validation with domain experts is necessary in advancing this field.

#### REFERENCES

- [1] Peter G Boyd and Tom K Woo. A generalized method for constructing hypothetical nanoporous materials of any net topology from graph theory. *CrystEngComm*, 18(21):3777–3792, 2016.
- [2] Chi Chen, Weike Ye, Yunxing Zuo, Chen Zheng, and Shyue Ping Ong. Graph networks as a universal machine learning framework for molecules and crystals. *Chemistry of Materials*, 31(9):3564–3572, 2019.
- [3] Ming Chen, Zhewei Wei, Zengfeng Huang, Bolin Ding, and Yaliang Li. Simple and deep graph convolutional networks. In Hal Daumé III and Aarti Singh, editors, *Proceedings of the 37th International Conference on Machine Learning*, volume 119 of *Proceedings of Machine Learning Research*, pages 1725–1735. PMLR, 13–18 Jul 2020.
- [4] Jiucheng Cheng, Chunkai Zhang, and Lifeng Dong. A geometric-information-enhanced crystal graph network for predicting properties of materials. *Communications Materials*, 2(1):1–11, 2021.
- [5] Yongchul G Chung, Emmanuel Haldoupis, Benjamin J Bucior, Maciej Haranczyk, Seulchan Lee, Hongda Zhang, Konstantinos D Vogiatzis, Marija Milisavljevic, Sanliang Ling, Jeffrey S Camp, et al. Advances, updates, and analytics for the computation-ready, experimental metal–organic framework database: Core mof 2019. *Journal of Chemical & Engineering Data*, 64(12):5985–5998, 2019.
- [6] Yu-Min Chung and Austin Lawson. Persistence curves: A canonical framework for summarizing persistence diagrams. *arXiv preprint arXiv:1904.07768*, 2019.
- [7] Francois-Xavier Coudert and Alain H Fuchs. Computational characterization and prediction of metal–organic framework properties. *Coordination Chemistry Reviews*, 307:211–236, 2016.
- [8] Hilal Daglar and Seda Keskin. Computational screening of metal–organic frameworks for membrane-based  $\text{CO}_2/\text{N}_2/\text{H}_2\text{O}$  separations: Best materials for flue gas separation. *The Journal of Physical Chemistry C*, 122(30):17347–17357, 2018.
- [9] Kathryn S Deeg, Daiane Damasceno Borges, Daniele Ongari, Nakul Rampal, Leopold Talirz, Aliaksandr V Yakutovich, Johanna M Huck, and Berend Smit. In silico discovery of covalent organic frameworks for carbon capture. *ACS applied materials & interfaces*, 12(19):21559–21568, 2020.
- [10] Hana Dureckova, Mykhaylo Krykunov, Mohammad Zein Aghaji, and Tom K Woo. Robust machine learning models for predicting high  $\text{CO}_2$  working capacity and  $\text{CO}_2/\text{H}_2$  selectivity of gas adsorption in metal organic frameworks for precombustion carbon capture. *The Journal of Physical Chemistry C*, 123(7):4133–4139, 2019.
- [11] Amir H Farmahini, Shreenath Krishnamurthy, Daniel Friedrich, Stefano Brandani, and Lev Sarkisov. Performance-based screening of porous materials for carbon capture. *Chemical Reviews*, 121(17):10666–10741, 2021.
- [12] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.
- [13] Kevin Maik Jablonka, Daniele Ongari, Seyed Mohamad Moosavi, and Berend Smit. Big-data science in porous materials: materials genomics and machine learning. *Chemical reviews*, 120(16):8066–8129, 2020.
- [14] Thomas N Kipf and Max Welling. Semi-supervised classification with graph convolutional networks. *arXiv preprint arXiv:1609.02907*, 2016.
- [15] Johannes Klicpera, Janek Groß, and Stephan Günnemann. Directional message passing for molecular graphs. *CoRR*, abs/2003.03123, 2020.
- [16] Aditi S Krishnapriyan, Maciej Haranczyk, and Dmitriy Morozov. Topological descriptors help predict guest adsorption in nanoporous materials. *The Journal of Physical Chemistry C*, 124(17):9360–9368, 2020.
- [17] Yongjin Lee, Senja D Barthel, Paweł Dłotko, S Mohamad Moosavi, Kathryn Hess, and Berend Smit. Quantifying similarity of pore-geometry in nanoporous materials. *Nature communications*, 8(1):1–8, 2017.
- [18] Yongjin Lee, Senja D Barthel, Paweł Dłotko, Seyed Mohamad Moosavi, Kathryn Hess, and Berend Smit. High-throughput screening approach for nanoporous materials genome using topological data analysis: application to zeolites. *Journal of chemical theory and computation*, 14(8):4427–4437, 2018.
- [19] Guohao Li, Matthias Mueller, Guocheng Qian, Itzel Carolina Delgadillo Perez, Abdulellah Abualshour, Ali Kassem Thabet, and Bernard Ghanem. Deepgcns: Making gcns go as deep as cnns. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pages 1–1, 2021.
- [20] Seyed Mohamad Moosavi, Aditya Nandy, Kevin Maik Jablonka, Daniele Ongari, Jon Paul Janet, Peter G Boyd, Yongjin Lee, Berend Smit, and Heather J Kulik. Understanding the diversity of the metal-organic framework

- ecosystem. *Nature Communications*, 11(1):1–10, 2020.
- [21] Cheol Woo Park and Chris Wolverton. Developing an improved crystal graph convolutional neural network framework for accelerated materials discovery. *Physical Review Materials*, 4(6):063801, 2020.
- [22] Andrew S Rosen, Shaelyn M Iyer, Debmalya Ray, Zhenpeng Yao, Alán Aspuru-Guzik, Laura Gagliardi, Justin M Notestein, and Randall Q Snurr. Machine learning the quantum-chemical properties of metal–organic frameworks for accelerated materials discovery. *Matter*, 4(5):1578–1597, 2021.
- [23] Jacob Townsend, Cassie Putman Micucci, John H Hymel, Vasileios Maroulas, and Konstantinos D Vogiatzis. Representation of molecular structures with persistent homology for machine learning applications in chemistry. *Nature communications*, 11(1):1–9, 2020.
- [24] Oliver T. Unke and Markus Meuwly. Physnet: A neural network for predicting energies, forces, dipole moments, and partial charges. *Journal of Chemical Theory and Computation*, 15(6):3678–3693, 2019. PMID: 31042390.
- [25] Oliver T Unke and Markus Meuwly. Physnet: a neural network for predicting energies, forces, dipole moments, and partial charges. *Journal of chemical theory and computation*, 15(6):3678–3693, 2019.
- [26] Petar Velickovic, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Lio’, and Yoshua Bengio. Graph attention networks. *ArXiv*, abs/1710.10903, 2018.
- [27] Ruihan Wang, Yeshuang Zhong, Leming Bi, Mingli Yang, and Dingguo Xu. Accelerating discovery of metal–organic frameworks for methane adsorption with hierarchical screening and deep learning. *ACS Applied Materials & Interfaces*, 12(47):52797–52807, 2020.
- [28] Christopher E Wilmer, Michael Leaf, Chang Yeon Lee, Omar K Farha, Brad G Hauser, Joseph T Hupp, and Randall Q Snurr. Large-scale screening of hypothetical metal–organic frameworks. *Nature chemistry*, 4(2):83–89, 2012.
- [29] Tian Xie and Jeffrey C Grossman. Crystal graph convolutional neural networks for an accurate and interpretable prediction of material properties. *Physical Review Letters*, 120(14):145301, 2018.
- [30] Xiangyu Zhang, Jing Cui, Kexin Zhang, Jiasheng Wu, and Yongjin Lee. Machine learning prediction on properties of nanoporous materials utilizing pore geometry barcodes. *Journal of chemical information and modeling*, 59(11):4636–4644, 2019.