Boosting heterogeneous catalyst discovery by structurally constrained deep learning models

A.N. Korovin [a, *], I.S. Humonen [a], A.I. Samtsevich [a], R.A. Eremin [a], A.I. Vasilev [a], V.D. Lazarev [a], S.A. Budennyy [a, b]

[a] Artificial Intelligence Research Institute (AIRI), Moscow, 105064, Russia
[b] Sber AI Lab, Moscow 117312, Russia

ABSTRACT

The discovery of new catalysts is one of the significant topics of computational chemistry as it has the potential to accelerate the adoption of renewable energy sources. Recently developed deep learning approaches such as graph neural networks open new opportunity to significantly extend scope for modeling novel high-performance catalysts. Nevertheless, the graph representation of a particular crystal structure is not a straightforward task due to the ambiguous connectivity schemes and numerous embeddings of nodes and edges. Here, we present embedding improvement for graph neural networks that has been modified by Voronoi tessellation and is able to predict the energy of catalytic systems within the Open Catalyst Project dataset. The enrichment of the graph was calculated via Voronoi tessellation, and the corresponding contact solid angles and types (direct/indirect) were considered as edges' features, and Voronoi volumes were used as node characteristics. The auxiliary approach was enriching node representation by intrinsic atomic properties (electronegativity, period, and group position). The proposed modifications allowed us to improve the mean absolute error of the original model, and the final error equals to 651 meV on the Open Catalyst Project dataset and 6 meV/atom on the intermetallics dataset. Also, by the consideration of an additional dataset, we show that a sensible choice of data can decrease the error to values below a physically-based 20 meV/atom threshold.

1. Introduction

Catalysis is a key and emergent concept in renewable energy enabling high yield, strongly specific chemical processes. This concept is in high demand in the field of chemical engineering, renewable energy, and batteries. A general target of research is to develop more efficient catalysts, which stand for quite different aspects. Among others, this can be a higher activity, higher selectivity, longer lifetimes, better availability, or preferable materials [1,2]. Despite remarkable achievements in materials design and, particularly, heterogeneous catalysts development for various chemical processes such as water splitting and carbon dioxide reduction, etc., the experimental trial-and-error approach is still predominate. There are difficulties associated with the transition from empirical and experimental approaches towards a predictive catalyst design. The main ones lie in the fact that an ordinary heterogeneous catalysis process time span and spatial scale is of higher orders of magnitude than ab initio computational chemistry can cover [3]. Moreover, catalyst performance depends on many variables, such as the catalyst composition, morphology, support material, and reaction environment (for example, temperature, solvent, and external potential).

The development of a computational chemistry approach providing an understanding of individual elementary processes of catalytic reactions is the key to overcome the time span and spatial scale problem. However, such an approach significantly reduces the complexity of the problem and considers a single crystal of an actual catalyst material [4]. Modern computational methods for the investigation of catalytic processes mainly include: a) density functional theory (DFT) or high-throughput theoretical calculations to estimate the energetic parameters of the elementary steps in catalysis reactions [5,6]; b) reactive force field (ReaxFF)-based molecular dynamics simulations,prediction of real-time reaction dynamics [7,8]; c) kinetic (e.g. mean-field or Monte Carlo) modeling, which can simulate reaction dynamics at the time scales

where the experimental reaction conditions often take place [9−11]. Still, such a simplification for the model catalyst is a good starting point for the more efficient data-driven search for a new catalyst material with potentially outstanding properties. However, presented classes of ab initio methods are still too computationally expensive to cover an enormously large region of interest in the configurational and compositional spaces and are confined only to the specific spatial and time scales at which those methods were developed.

Machine learning is a robust tool that can be used to learn the relationship between the input features of an assumed catalyst and its predicted output performance [12,13]. Due to the ability to learn from raw data, deep learning approaches have boosted research and made it possible to discover new catalyst structures and extract the underlying correlated features for catalytic reactions [3,5,14−16].

In past few years, graph neural networks (GNNs) have been developed, most of which use spatial convolution operation to efficiently address properties of crystalline structures. For example, GNNs have been used to predict photosynthetic ammonia [17] or the catalytic behavior of materials in the competition named Open Catalyst Project (OCP) [18,19], which is driven by the deep learning methods materials design.

In the paper, we discuss the pros and cons of GNNs models referring to the OCP challenge and compare our solution with other models. We show that existing models still cannot accurately approximate the DFT calculation results within the OCP dataset. We also speculate about the origin of this. Finally, we show how a GNN-based model provided within the OCP challenge is able to predict energetics of intermetallics, which is a more specific task providing two orders of magnitude smaller errors of predictions.

2. Computational methods

2.1. Models

GNNs are able to solve a wide range of problems/tasks related to data in a graph format, including the prediction of a numeric property of an entire graph. In turn, molecules and crystal structures can be considered as graphs in a natural way, which allows one to apply GNNs for energy and force prediction tasks [20]. Nevertheless, GNNs have some principal restrictions: they operate with a graph as a formal object (a set of nodes and edges), while chemical compounds are complex spatial systems and thus, two isomorphic graphs can represent completely different structures [21]. To overcome this drawback, modern models utilize information about the relative atoms' positions as nodes' and edges' features and use invariant (and/or equivariant) convolution operators.

There are developed energy-rotation invariant (and force-rotation equivariant) GNN-based models of different architecture, size, complexity, and performance (*CGCNN* [22], *SchNet* [23], *DimeNet* [24], and *GemNet* [25], etc.). Trying to achieve a trade-off between the speed, quality, and flexibility of the model's architecture (the ease of encoding of angular information about node's neighbors), we chose the *SpinConv* [26] model.

All models mentioned above are capable to encode mutual atomic arrangements via edge-to-edge angles, dihedral angles, spherical projections, etc., and various spatial convolution operators. *SpinConv* uses Gaussian smearing and trainable linear transformation for distance encoding. Additionally, projection onto the unit sphere with one-dimensional convolution in the longitudinal direction of local spherical coordinate systems is applied for aggregating information about node neighbors to embedding. These distinctions determine such physical restrictions as a rotation equivariance and a difference in radii of atoms.

However, except for *CGCNN*, all models (including *SpinConv*) do not consider chemical information for atoms, bounding themselves only with atomic numbers (or providing convolutional layers to learn all this information). Also, all models turn chemical compounds into graphs using a radius graph, which, as presented before, seems non-physical and compels the convolution operators to handle excess neighbors.

2.2. Crystal graph modifications

2.2.1. Graph connectivity and edge properties

As mentioned, the most commonly used approach (also implemented in *SpinConv*) to set graph connectivity is the radius graph, which considers atoms located closer than the predefined cut-off R_{cut} as connected, taking into account periodic boundary conditions. Thus, the intrinsic symmetry properties of a crystal structure as well as the anisotropic nature of crystalline materials are vanished by the reconstruction of graph connectivity. Furthermore, such an approach is not free of parameters, selection of which may influence the results obtained.

Inspired by the recent success of the *CGCNN* modifications [27] and the presence of particular GNN architectures [28], we have considered the Voronoi partition [29] of crystal space to modify the graph connectivity of the identical *SpinConv* model. This approach is based on the real space tessellation allowing to set a number of edges of the crystal structure graph concerning only the nearest neighbors of a particular node (atom) and, subsequently, accounting for a crystal symmetry in a parameter-free manner. Particularly, for the considered task, one can obtain a list of the crystal surface atoms that are most 'open' to the molecule. Comparison of the radius-graph connectivity and Voronoi partition is given in Fig. 1.

By using the Voronoi partition instead of the crystal graph method (providing a set of distances only), one can obtain additional properties for each edge and each node of the graph (see Fig. 1b). For instance, the edge obtained for the updated connectivity can belong to direct or indirect contacts of the Voronoi net. For the formerly mentioned group, the line corresponding to the contact between the central atom and its neighbor crosses the edge of the Voronoi polyhedron. This is not the case for indirect contacts.

Additionally, the number of the neighbors (faces of the corresponding Voronoi polyhedron) can be interpreted as a coordination number of a particular node of the graph (atom). For each of the neighbors (contacts), the solid angle provides information on the interaction intensity (see Fig. 1c). The higher the solid angle, the stronger the interaction between the central atom and particular neighbor. Thus, the solid angles become an additional property of the edges (contacts) of the crystal graph.

Expectedly, by using the Voronoi partition, we have met a problem of slab-to-slab connections through the vacuum region (see Fig. 1a). Such connections possess non-physically large distances (anisotropic Voronoi polyhedra of large volumes), although they satisfy Voronoi tessellation rules. For this reason, an additional limiting parameter R_{lim} was introduced to eliminate inter-slab connections due to the periodic boundary conditions. This limiting parameter R_{lim} determines the longest edge distance within the crystal graph and, in this sense, is similar to R_{cut} within the radius graph approach.

2.2.2. Node (atom) properties

Based on the Voronoi partition, each node of the graph can be characterized by the volume of the Voronoi polyhedron that represents the volume of the crystal structure 'occupied' by the corresponding atom. The identical *SpinConv* model describes each
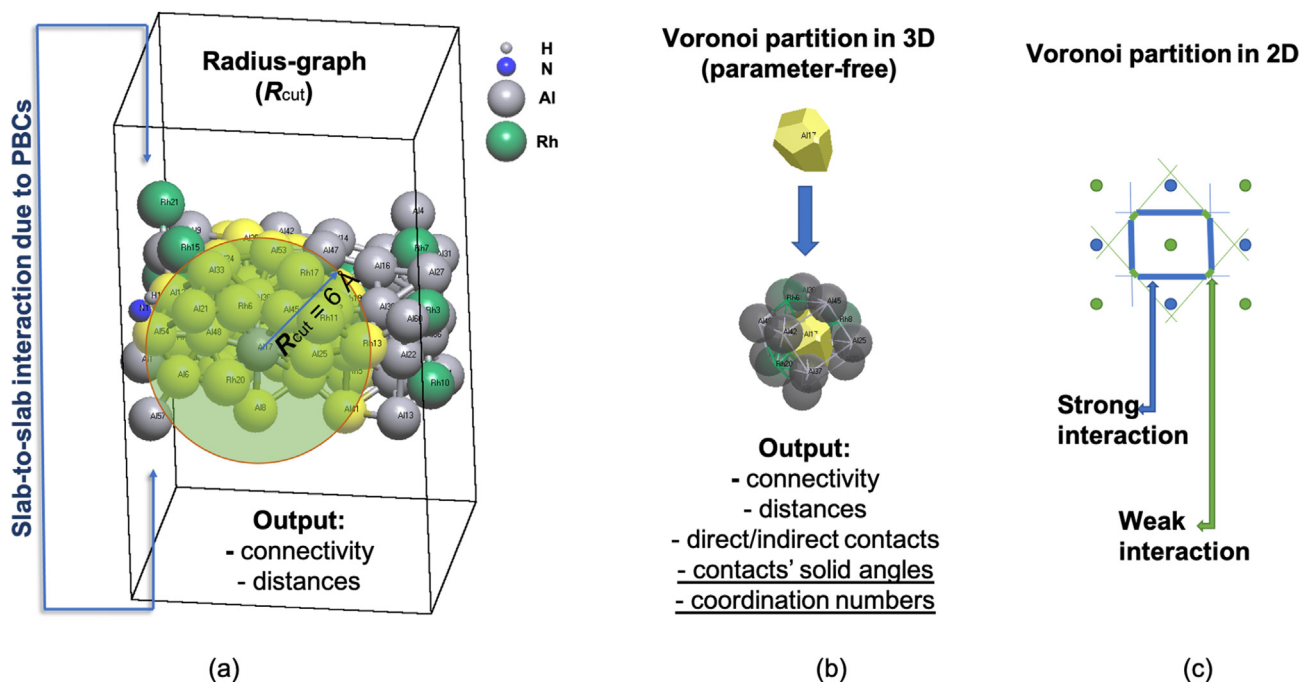
Fig. 1. (a) The sample crystal structure from those provided within the Open Catalyst Project dataset, and representation of neighbors of a certain node (atom) obtained using the radius graph method. $R_{cut} = 6$ Å. (b) For the same node of the graph, the Voronoi polyhedron and the corresponding nearest neighbors are determining its shape. The snapshots of the crystal structure were obtained by using TopasPro software [30] (c) The two-dimensional example of the Voronoi partition demonstrating the connection of interaction intensity differences in terms of Voronoi polygon edges. PBCs, periodic boundary conditions.

atomic species as a one-hot-encoded atomic number, which can potentially reduce the generalization ability of the trained model. For this reason, we have additionally checked a number of atomic (nodal) features.

Instead of atomic numbers, we have also considered the embedding of chemical properties previously used in a *CGCNN* model [22] including group and period numbers, electronegativity, covalent radius, valence electrons, and ionization energy. Additionally, electronegativity representing an intrinsic feature of a particular chemical element is a numerical type property which was checked. Within an individual model, the combination of a period and group (one-hot-encoded) was used instead of atomic numbers of the original *SpinConv*.

2.2.3. Graph modifications

The modifications tested within the current research are listed in Table 1, where RG and VG stands for the radius-graph and Voronoi graph, respectively.

2.3. Datasets

2.3.1. Open Catalyst Project

Predictions of catalysts are even more challenging as combinations of small molecules, and multiple crystalline surfaces should be considered. Open Catalyst Challenge [18], one of recent and aspiring attempts to tackle this problem, is an open dataset combining about half a million DFT models stable and small organic reactants (up to 225 atoms).

2.3.2. Additional dataset. Sc—Pd intermetallics

An auxiliary dataset was utilized to further test the proposed modifications of the graph representation and to study the dependence of the performance of the GNN-based solution on the provided data. We used a recently developed [31] dataset of structurally disordered Sc—Pd, Sc—Rh, Sc—Pt, and Sc—Ir complex intermetallics. Taking into account the structural similarity of the aforementioned series of compositions, we tested the proposed

Table 1
The detailed description of modifications of implemented models.

model	short names	graph connectivity	node properties		edge properties
			immanent	system-dependent	
VG$_{6Å}$ and Voronoi volumes	Vg + vol	VG	atomic numbers	Voronoi volumes	distances
VG$_{6Å}$ and solid angles	Vg + angle	VG	atomic numbers	−	distances, solid angles
VG$_{6Å}$	Vg	VG	atomic numbers	−	distances
VG$_{6Å}$ and electronegativity	Vg + e-neg	VG	electronegativity	−	distances
RG$_{6Å}$ and CGCNN embeds	Rg + embeds	RG	CGCNN embeds	−	distances
RG$_{6Å}$ and periods + groups	Rg + table	RG	periods and groups	−	distances
RG$_{6Å}$	Rg	RG	atomic numbers	−	distances

Abbreviations: RG, radius-graph; VG, Voronoi graph

modifications only for the thermodynamic properties of the Sc−Pd systems.

This data collection represents the composition/configuration space (model crystal structures) for searching for 1/1 Mackay-type [32,33] quasicrystal approximants. The obtained results for the OCP dataset for structures that contain Pd show the lowest average mean absolute error (MAE) among the transitional metals (Pd, Pt, Ir, and Rh) studied within the aforementioned work [31]. Each entry of the additional dataset represents a 3-periodic complex crystal structure with more than 100 metal atoms per unit cell and disordering within the region shown in Fig. 2. In contrast to the OCP dataset, the full DFT-based relaxation of such systems does not require restrictions neither for atomic positions (selective dynamics) nor for a unit cell size/shape.

The selected Sc positions of the crystal structure in Fig. 2 correspond to the disordering region between Mackay clusters, where vacancies and substitutions are added in different amounts during dataset generation and determining the structural and compositional diversity. Among the dataset comprising 2107 crystal structure realizations with different Sc/Pd contents, 1896 structures (*ca.* 90%) were used within the GNN training, whereas the rest of the structures were considered for validation (211 entries, *ca.* 10%). A test dataset comprising 63 energetically favored realizations of the studied crystal structures was additionally prepared and used.

We considered the DFT-relaxed energies of intermetallic crystals and the corresponding energies above the convex hull of the studied Sc−Pd system. The latter one is a good example of a thermodynamic property that is less independent of the calculation scheme. For more information on the dataset preparation and content, the readers are referred to the original work [31].

3. Results and discussion

3.1. Model performances

3.1.1. Open Catalyst Project dataset

For the OCP dataset, the results listed in Table 2 clearly show the influence of some of the proposed modifications of the graph connectivity and additional properties of nodes and edges (available from the Voronoi partitions) on the validation scores. It is worth noting that the proposed modifications were tested on the *val_ood_both* datasets comprising structures with substrates and molecular fragments out of the training set domain. The results listed in Table 2 correspond to the models trained on the full dataset (*ca.* 460k structures) and 10k structure subset; the Vg+vol model is omitted in Table 2 and excluded from further consideration because of its poor performance on the OCP (10k) dataset (MAE 1.1 eV).

For the OCP datasets, the best MAE value of 0.651 eV was obtained by using updated connectivity accompanied by the solid angles as properties of each edge of the crystal graphs of the modeled systems. Nevertheless, the difference between this improvement and the identical *SpinConv* model is *ca.* 0.02 eV/atom. In turn, the additionally considered modification of embeddings of nodes (atomic species) by period/group combinations as well as accounting for their tabular electronegativities result in an increase of the corresponding MAE values of the comparable magnitudes. An important observation at the current stage of the analysis is the fact of a non-uniform distribution of MAE with respect to the chemical compositions of the systems under consideration, schematically shown in Fig. 3.

As one can see, the lowest MAE obtained corresponds to the systems comprising transitional metals including Pt (mean MAE of 0.4953), Pd (mean MAE of 0.4317), Rh (mean MAE of 0.4628), etc. commonly used as catalysts.

3.1.2. Sc−Pd intermetallics dataset

The obtained results for the DFT-relaxed energies of Sc−Pd intermetallics and the corresponding energies above the convex hull for this system are given in Table 2. It is worth noting that the MAE scores obtained using the identical *SpinConv* model are more than 10 times smaller for the studied intermetallic compounds, which can be associated with both a much poorer chemical/structural diversity resulting in a narrower target energy distribution for the studied dataset and the settings of the DFT-based modeling. Despite such inaccuracy of the GNN model, it was shown [31] to provide reliable thermodynamic property predictions and a number of stable phases in good agreement with available experimental data [35].
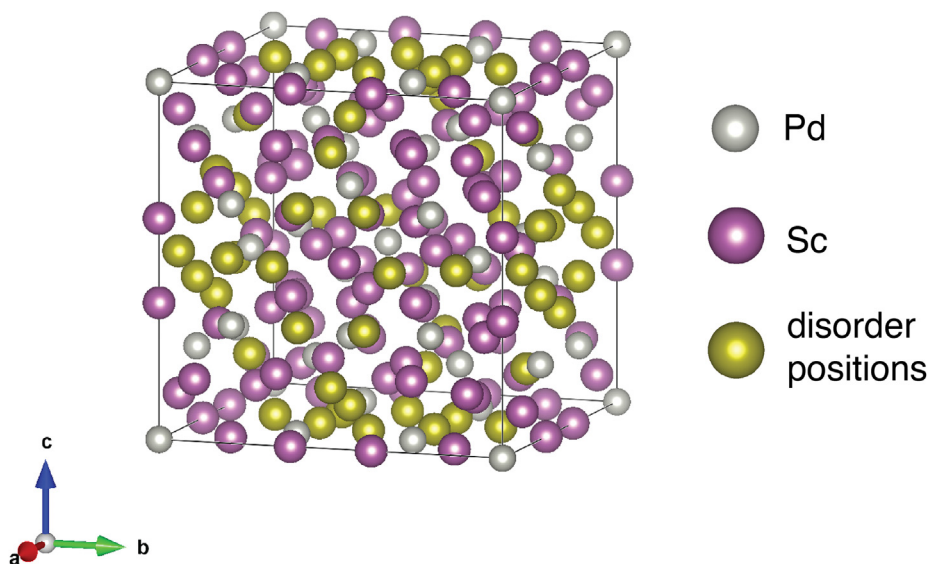

Fig. 2. The sample crystal structure from those provided within the additional dataset comprising Sc−Pd intermetallics. The selected (in yellow) atomic positions correspond to the disordered region variable within the dataset. The snapshot was prepared using the Vesta program [34].

Table 2

For the studied datasets, the obtained MAE scores with respect to the modification of the crystal graph connectivity and the features of its nodes and edges. See Table 1 for detailed information about the models.

model	all OCP	10k OCP	RE test intermet	EH test intermet	RE val intermet	EH val intermet
Vg + angle	**0.651**	**0.931**	**0.02013**	**0.00579**	**0.0182**	0.0151
Vg	–	1.002	0.0234	0.0194	0.0186	**0.0126**
Vg + e-neg	0.7096	0.9488	0.0409	0.01584	0.0259	0.01902
Rg	0.6686	0.954	0.0422	0.0169	0.0262	0.0193
Rg + embeds	–	1.009	0.046	0.017	0.0272	0.0188
Rg + table	0.6731	0.9348	0.0578	0.0189	0.0292	0.0196

Abbreviations: MAE, mean absolute error; OCP, Open Catalyst Project; RE, relaxed energy; EH, energy above the convex hull; Rg, radius graph; Vg, Voronoi graph
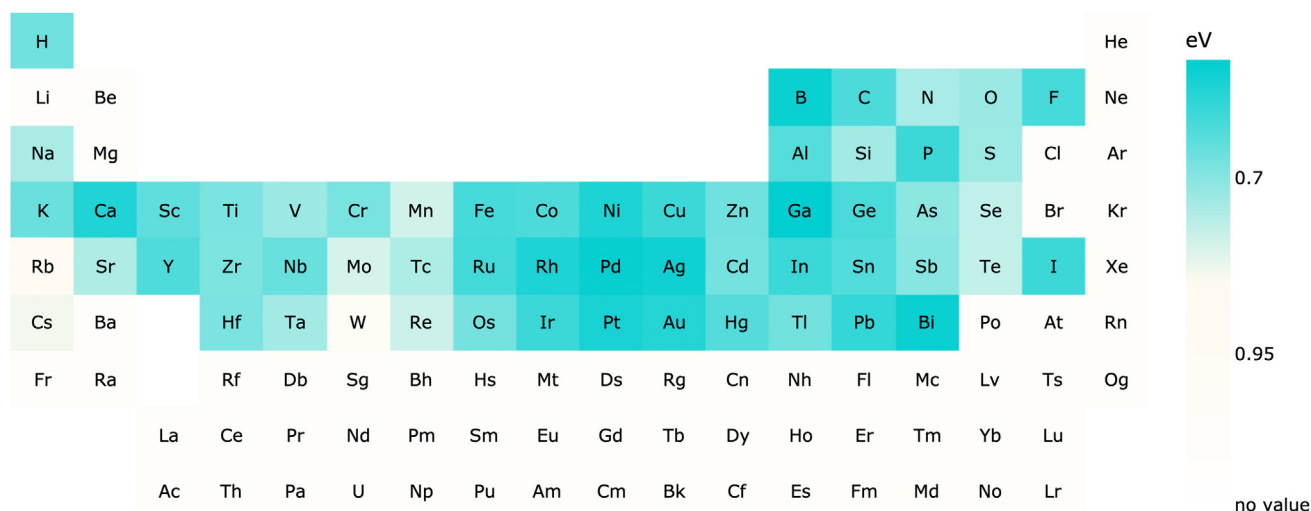

Fig. 3. The characteristic mean MAE distribution averaged over the systems comprising particular chemical species within the OCP dataset for the identical *SpinConv* model. MAE, mean absolute error; OCP, Open Catalyst Project

On the other hand, the obtained MAEs are more drastically dependent on solid angles for the both studied target energies. In the case of relaxed energy, the obtained results can also be improved by the graph connectivity only (see Table 2). For the studied Sc−Pd systems, the lowest obtained MAE of *ca.* 6 meV/atom for the energies above the convex hull is comparable to the metrics obtained by the modifications of the *SpinConv* architecture implemented and tested recently [31].

As one can see in Fig. 4, the implemented modifications possess stability in terms of prediction quality with respect to the subsamples (validation and test subsets) of the intermetallic dataset as well as for different target properties studied. In contrast, for the OCP dataset, the obtained results do not follow any regular pattern. This observation can be associated with the intrinsic properties of the 2-periodic (slab-like) crystal structures, for which the Voronoi partition is not a straightforward (the aforementioned slab-to-slab connections) approach and requires a cut-off introduction for the graph edges.

3.2. Data preparation aspects

Obviously, the presented reduction in the chemical/structural diversity of model datasets providing better test scores is limited for any high-throughput screenings or general-purpose applications. For this reason, here we want to reflect on some additional aspects of data preparation and their possible effects on the developing data-driven approaches for searching for new functional materials including heterogeneous catalysts on the basis of 'synthetic' (e.g. DFT-derived) properties.

3.3. Target properties independent on the calculation scheme

The consideration of different target properties available from DFT calculations in connection with synthesizability of novel materials and their stability is widely discussed [36,37]. Obviously, there is a direct connection between the DFT-relaxed and formation energies (or energies above the convex hull) of a particular crystal structure. Nevertheless, from a computational point of view, the relaxed energies are dependent on the chosen calculation scheme (pseudopotentials, basis set size, sampling of the reciprocal space, etc.), whereas formation energies or any derivative thermodynamic properties obtained from these are free of any calculation biases. Moreover, the results obtained demonstrate possible differences in the model prediction quality for different target energies (see Table 2). In turn, the descriptor-based approaches may show differences between compositional and structural feature importance, when changing target values. For instance, structural descriptors became more important features (in comparison to the chemical composition) for the energies above the convex hull (in comparison to the DFT-relaxed energies) [31]. The presence of datasets that are widely diverse from chemical and structural points of view, such that of OCP, provides a great possibility for the pre-training of GNN-based models and their subsequent fine tuning for the purposes of any more particular problems [31].

To compare energy efficiency of models, we measured energy consumption with the help of an open-source python library, eco2AI [38]. A carbon footprint was also estimated using a regional specific carbon intensity coefficient equal to 240.56 g/kWh. As one can see from Table 3 models a pre-calculated Voronoi graph (Vg) were more than one order of magnitude more effective for the
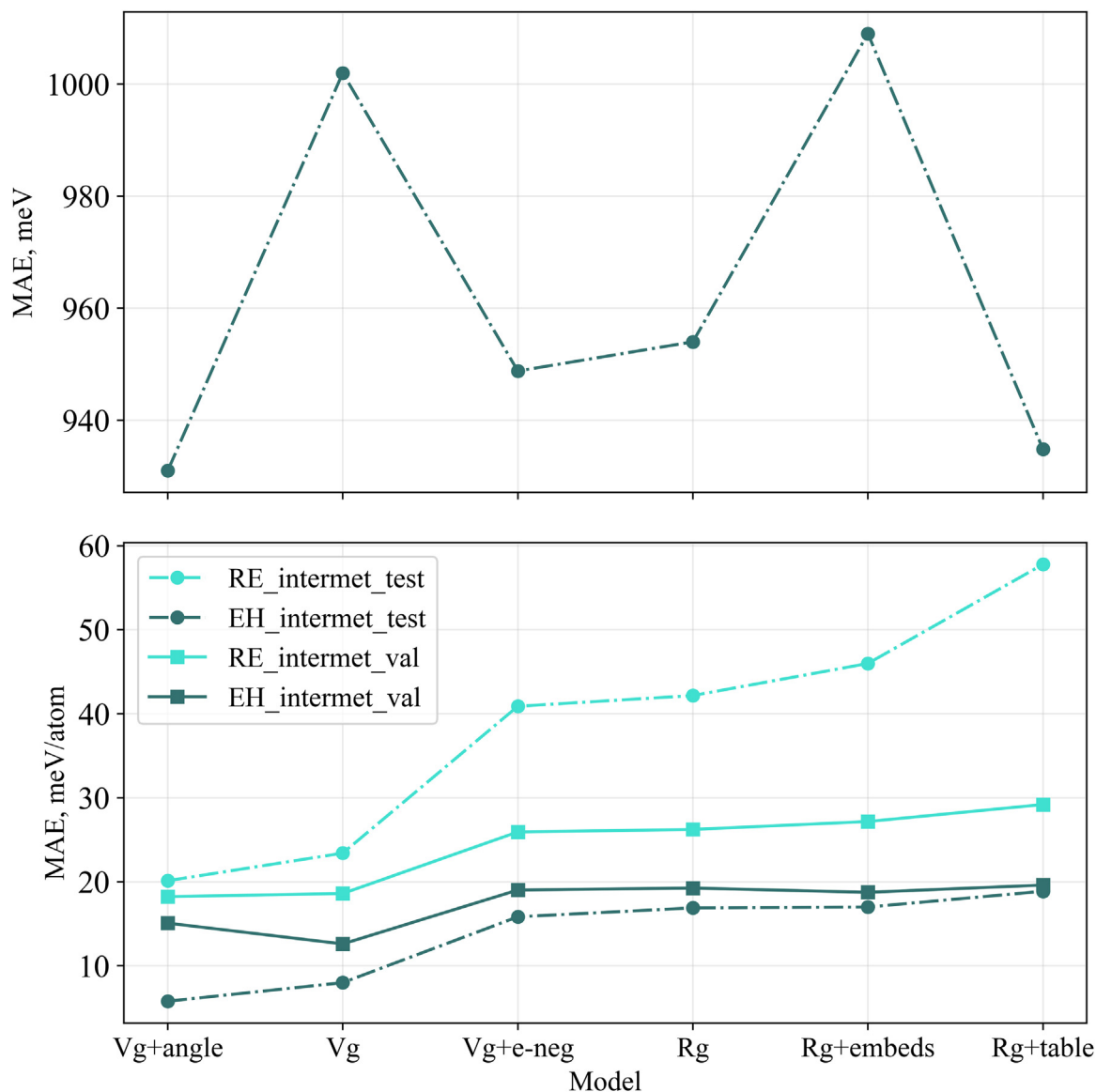
Fig. 4. performance of the models with various modifications on different samples. Top panel: the performance on the OCP *val_ood_both* dataset (models were trained on the OCP train 10k subset). Bottom panel: the performance on the Sc–Pd intermetallics validation and test datasets (models were trained on the relaxed energy and energy above the convex hull as target values). Please refer to Table 1 for detailed information about the models. OCP, Open Catalyst Project

estimation of relaxed energy (RE) comparing to models where the radius graph (Rg) was recalculated at each step.

3.4. Computational constrains applied

The differences between the validation/test scores obtained for the OCP heterogeneous structures and those for the intermetallics can also be associated with computational constraints, such as selective dynamics. To briefly discuss this point, we present the distribution of target energies available from the OCP datasets Fig. 5a. For a vast range of crystal structures, the target energies were found to be positive, which directly points at the consequences of preparing initial structure guesses comprising the crystal slab generation, placing molecular fragment, and, for sure, selective dynamics used by the authors [18].

In modern DFT-based approaches, the presence of systems with altered energetics (obtained within structurally constrained relaxations) is essential in many applications, such as the equation of

Table 3

For the studied datasets, the energy spent for model training and corresponding carbon emissions. See Table 1 for detailed information about the models.

model	RE		EH	
	E, kWh	CO_2, g	E, kWh	CO_2, g
Vg + angle	0.069	16.6	0.14	34.0
Vg	0.069	16.6	0.25	59.3
Rg + e-neg	1.30	312.6	1.30	313.4
Rg	1.31	315.5	1.31	315.6
Rg + embeds	1.32	317.6	2.22	320.9
Rg + table	1.33	319.0	1.30	312.9

Abbreviations: RE, relaxed energy; EH, energy above the convex hull; Rg, radius graph; Vg, Voronoi graph

state calculations, static and dynamic stability investigations, phonon spectra calculations, etc. Such 'negative' examples (for instance, those with positive energies) possess also high importance for data-driven approaches. From computational
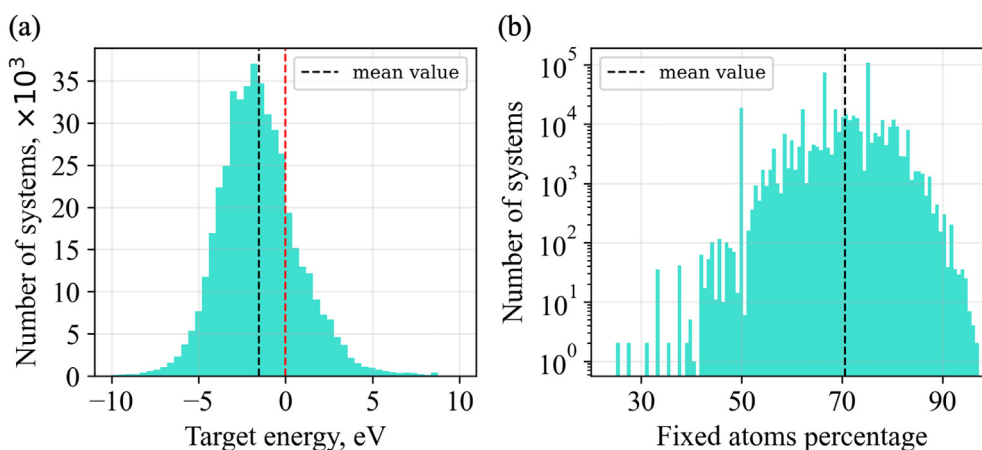
Fig. 5. (a) The distribution of the target energies available for the relaxed atomic systems within the full OCP dataset, comprising *ca.* 460k structures; red dotted line denotes zero energy. (b) The distribution of the percentage of fixed atomic positions within the slabs of the OCP structures.

perspectives, any selective dynamics can affect the way for a system going upon the energy landscape during the relaxation.

As one can see from Fig. 5b, the percentages of the fixed atomic positions within slabs possess wide distribution. Particularly for the Initial Structure to Relaxed Energy task, this points at the variable scale of the described effect for the dataset entries. Thus, selective dynamics might provide different biases of the target properties for different systems, although each system provided within the dataset remained structurally stable during the relaxation.

4. Conclusion and perspectives

Within the current research, we developed and tested modifications of the graph representation using the Voronoi partition in the case of heterogeneous systems comprising combinations of crystal slabs and molecular fragments. The comparison of the results of the developed approach for the studied 2-periodic entries of the OCP dataset and 3-periodic intermetallics showed advantages of the usage of proposed graph representations, perspectives of its further improvements and customization, and a set of the most promising features being considered further (especially, the solid angles of the contacts).

Additionally, the drastic decrease in the MAE scores accompanying the change of the systems modeled from 2-periodic (the OCP dataset) to 3-periodic (intermetallics) systems was demonstrated using the same GNN-based solutions. We associate such a behavior with the reduction of the chemical and structural diversity for the studied dataset and avoiding computational constraints, namely, selective dynamics used for the 2-periodic systems provided within the OCP dataset.

We want to emphasize that selective dynamics itself might be applicable in the case of a preset surface and the same amount/ disposition of the fixed atomic positions within this. In such a situation, the influence of the frozen degrees of freedom of the slab relaxation might have the same energetic effect for modifications of interest, such as adsorbate positioning as an initial guess or molecular ensembles on the adsorbent surface. The described approach, for sure, can hardly be aimed at any ambitious high-throughput applications and, again, represents a routine with a reduced structural/chemical diversity of the systems being investigated.

Nevertheless, within the described modifications and through the application of data-driven solutions, it may become possible to comprehensively study a particular surface and catalytic processes

of interest. In such a case, thermodynamic properties of numerous states obtained computationally allow using Boltzmann statistics to take into account both the most energetically favorable states as well as altered/metastable ones that may have lower but non-negligible frequencies of occurrence [39,40]. From this perspective, the authors would like to mention that the implementation of the Voronoi partition instead of the conventional radius graph resulted in a reduction of the computational costs of the studied GNN-based solutions for the obvious reason of graph simplification.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

References

[1] Zhi Wei Seh, Jakob Kibsgaard, Colin F. Dickens, Ib Chorkendorff, Jens K. Nørskov, Thomas F. Jaramillo, Combining theory and experiment in electrocatalysis: insights into materials design, Science 355 (6321) (2017), eaad4998, https://doi.org/10.1126/science.aad4998.

[2] Ambarish Kulkarni, Samira Siahrostami, Anjli Patel, K. Jens, Nørskov, Understanding catalytic activity trends in the oxygen reduction reaction, Chem. Rev. 118 (55) (2018) 2302−2312, https://doi.org/10.1021/acs.chemrev.7b00488.

[3] Zachary W. Ulissi, Andrew J. Medford, Thomas Bligaard, Jens K. Nørskov, To address surface reaction network complexity using scaling relations machine learning and DFT calculations, Nat. Commun. 8 (1) (2017) 1−7, https://doi.org/10.1038/ncomms14621.

[4] Gerhard Ertl, Reactions at Solid Surfaces, John Wiley & Sons, 2009, https://doi.org/10.1002/9780470535295.

[5] Nongnuch Artrith, Alexie M. Kolpak, Understanding the composition and activity of electrocatalytic nanoalloys in aqueous solvents: a combination of DFT and accurate neural network potentials, Nano Lett. 14 (5) (2014) 2670−2676, https://doi.org/10.1021/nl5005674.

[6] Heine A. Hansen, Jan Rossmeisl, Jens K. Nørskov, Surface pourbaix diagrams and oxygen reduction activity of Pt, Ag and Ni (111) surfaces studied by DFT, Phys. Chem. Chem. Phys. 10 (25) (2008) 3722−3730, https://doi.org/10.1039/B803956A.

[7] Jacob R. Boes, Mitchell C. Groenenboom, John A. Keith, John R. Kitchin, Neural network and ReaxFF comparison for au properties, Int. J. Quant. Chem. 116 (13) (2016) 979−987, https://doi.org/10.1002/qua.25115.

[8] Thomas P. Senftle, Adri CT. Van Duin, Michael J. Janik, Methane activation at the Pd/CeO$_2$ interface, ACS Catal. 7 (1) (2017) 327−332, https://doi.org/10.1021/acscatal.6b02447.

[9] Grigoriy Kimaev, Luis A. Ricardez-Sandoval, Artificial neural network discrimination for parameter estimation and optimal product design of thin

films manufactured by chemical vapor deposition, J. Phys. Chem. C 124 (34) (2020) 18615−18627, https://doi.org/10.1021/acs.jpcc.0c05250.

[10] Baron P. Chapter 4 - Catalysis. In Reaction Rate Theory and Rare Events, Elsevier, 2017, pp. 79−128, https://doi.org/10.1016/B978-0-44-456349-1.00004-0.

[11] Yingzhe Yu, Jie Zhang, Xuanyu Sun, Minhua Zhang, Carbon chain growth mechanism of higher alcohols synthesis from syngas on CoCu (100) a combined DFT and kMC study, Surf. Sci. 691 (2020), 121513, https://doi.org/10.1016/j.susc.2019.121513.

[12] Li Zheng, Siwen Wang, Hongliang Xin, Toward artificial intelligence in catalysis, Nature Catalysis 1 (9) (2018) 641−642, https://doi.org/10.1038/s41929-018-0150-1.

[13] John R. Kitchin, Machine learning in catalysis, Nature Catalysis 1 (4) (2018) 230−232, https://doi.org/10.1038/s41929-018-0056-y.

[14] Ryosuke Jinnouchi, Ryoji Asahi, Predicting catalytic activity of nanoparticles by a DFT-aided machine-learning algorithm, J. Phys. Chem. Lett. 8 (17) (2017) 4279−4283, https://doi.org/10.1021/acs.jpclett.7b02010.

[15] Wenhong Yang, Timothy Tizhe Fidelis, Wen-Hua Sun, Machine learning in catalysis, from proposal to practicing, ACS Omega 5 (1) (2019) 83−88, https://doi.org/10.1021/acsomega.9b03673.

[16] Hao Li, Zhien Zhang, Zhijian Liu, Application of artificial neural networks for catalysis: a review, Catalysts 7 (1010) (2017) 306, https://doi.org/10.3390/catal7100306.

[17] Yuheng Zhou, Xiaohui Wang, Xubo Huang, Hui Deng, Yuntao Hu, Linfang Lu, Predicting the photosynthetic ammonia on nanoporous cobalt zirconate via graph convolutional neural networks, Mol. Catal. 529 (2022), 112565, https://doi.org/10.1016/j.mcat.2022.112565.

[18] Lowik Chanussot, Abhishek Das, Siddharth Goyal, Thibaut Lavril, Muhammed Shuaibi, Morgane Riviere, Kevin Tran, Javier Heras-Domingo, Caleb Ho, Weihua Hu, et al., Open catalyst 2020 (OC20) dataset and community challenges, ACS Catal. 11 (10) (2021) 6059−6072, https://doi.org/10.1021/acscatal.0c04525.

[19] C. Lawrence Zitnick, Lowik Chanussot, Abhishek Das, Siddharth Goyal, Javier Heras-Domingo, Caleb Ho, Weihua Hu, Thibaut Lavril, Aini Palizhati, Morgane Riviere, Muhammed Shuaibi, Anuroop Sriram, Kevin Tran, Brandon Wood, Junwoong Yoon, Devi Parikh, Zachary Ulissi, An Introduction to Electrocatalyst Design Using Machine Learning for Renewable Energy Storage, 2020 *preprint arXiv:2010.09435 [cond-mat]*, https://arxiv.org/abs/2010.09435.

[20] Michael M. Bronstein, Joan Bruna, Taco Cohen, Petar Velickovic, Geometric Deep Learning: Grids, Groups, Graphs, Geodesics, and Gauges, 2021 *preprint arXiv:2104.13478 [cs, stat]*, https://arxiv.org/abs/2104.13478.

[21] Katsuhiko Ishiguro, Kenta Oono, Kohei Hayashi, Weisfeiler-lehman Embedding for Molecular Graph Neural Networks, 2020 *preprint arXiv:2006.06909 [cs.LG]*, https://arxiv.org/abs/2006.06909.

[22] Xie Tian, C. Jeffrey, Grossman, Crystal graph convolutional neural networks for an accurate and interpretable prediction of material properties, Phys. Rev. Lett. 120 (14) (2018), https://doi.org/10.1103/PhysRevLett.120.145301.

[23] Kristof Schütt, Pieter-Jan Kindermans, Huziel Enoc Sauceda Felix, Stefan Chmiela, Tkatchenko Alexandre, Klaus-Robert Müller, SchNet: A Continuous-Filter Convolutional Neural Network for Modeling Quantum Interactions, 2017. *Preprint arXiv:1706.08566 [physics, stat]*, http://arxiv.org/abs/1706.08566.

[24] Johannes Klicpera, Janek Gro, Günnemann Stephan, Directional Message Passing for Molecular Graphs, 2020 *preprint arXiv:2003.03123 [cs.LG]*, https://arxiv.org/abs/2003.03123.

[25] Johannes Klicpera, Florian Becker, Stephan Günnemann, GemNet: Universal Directional Graph Neural Networks for Molecules, 2021 *preprint arXiv: 2106.08903 [physics.comp-ph]*, https://arxiv.org/abs/2106.08903.

[26] Muhammed Shuaibi, Adeesh Kolluru, Abhishek Das, Aditya Grover, Anuroop Sriram, Zachary Ulissi, C. Lawrence Zitnick, Rotation Invariant Graph Neural Networks Using Spin Convolutions, 2021 *preprint arXiv:2106.09575 [cs.LG]*, http://arxiv.org/abs/2106.09575.

[27] Cheol Woo Park and Chris Wolverton, Developing an improved crystal graph convolutional neural network framework for accelerated materials discovery, Physical Review Materials 4 (6) (2020), 063801, https://doi.org/10.1103/PhysRevMaterials.4.063801.

[28] Ilia Igashov, Kliment Olechnovic, Maria Kadukova, Ceslovas Venclovas, Sergei Grudinin, VoroCNN: deep convolutional neural network built on 3D Voronoi tessellation of protein structures, Bioinformatics 37 (16) (2021) 2332−2339, https://doi.org/10.1093/bioinformatics/btab118.

[29] V.A. Blatov, Voronoi−dirichlet polyhedra in crystal chemistry: theory and applications, Crystallogr. Rev. 10 (4) (2004) 249−318, https://doi.org/10.1080/08893110412331323170.

[30] Vladislav A. Blatov, Alexander P. Shevchenko, Davide M. Proserpio, Applied topological analysis of crystal structures with the program package ToposPro, Cryst. Growth Des. 14 (7) (2014) 3576−3586, https://doi.org/10.1021/cg500498k.

[31] Roman A. Eremin, Innokentiy S. Humonen, Pavel N. Zolotarev, Inna V. Medrish, Leonid E. Zhukov, Semen A. Budennyy, Hybrid DFT/data-driven approach for searching for new quasicrystal approximants in Sc-X (X = Rh, Pd, Ir, Pt) systems, Cryst. Growth Des. 22 (7) (2022) 4570−4581, https://doi.org/10.1021/acs.cgd.2c00463.

[32] N. Tamura, The concept of crystalline approximants for decagonal and icosahedral quasicrystals, Philos. Mag. A 76 (2) (1997) 337−356, https://doi.org/10.1080/01418619708209979.

[33] T.G. Akhmetshina, V.A. Blatov, A fascinating building unit: Mackay cluster in intermetallics, Struct. Chem. 28 (2017) 133−140, https://doi.org/10.1007/s11224-016-0828-4.

[34] Koichi Momma, Fujio Izumi, Vesta: a three-dimensional visualization system for electronic and structural analysis, J. Appl. Crystallogr. 41 (3) (2008) 653−658, https://doi.org/10.1107/S0021889808012016.

[35] Pavlo Solokha, Roman A. Eremin, Tilmann Leisegang, Davide M. Proserpio, Tatiana Akhmetshina, Albina Gurskaya, Adriana Saccone, Serena De Negri, New quasicrystal approximant in the Sc−Pd system: from topological data mining to the bench, Chem. Mater. 32 (3) (2020) 1064−1079, https://doi.org/10.1021/acs.chemmater.9b03767.

[36] Christopher J. Bartel, Amalie Trewartha, Qi Wang, Alexander Dunn, Anubhav Jain, Gerbrand Ceder, A critical examination of compound stability predictions from machine-learned formation energies, npj Comput. Mater. 6 (97) (2020) 1−11, https://doi.org/10.1038/s41524-020-00362-y.

[37] Gordon Peterson, Jakoah Brgoch, Materials discovery through machine learning formation energy, J. Phys.: Energy 3 (2) (2021), 022002, https://doi.org/10.1088/2515-7655/abe425.

[38] S.A. Budennyy, V.D. Lazarev, N.N. Zakharenko, A.N. Korovin, O.A. Plosskaya, D.V. Dimitrov, V.S. Akhripkin, I.V. Pavlov, I.V. Oseledets, I.S. Barsola, I.V. Egorov, A.A. Kosterina, L.E. Zhukov, eco2ai: carbon emissions tracking of machine learning models as the first step towards sustainable AI, Dokl. Math. (2023), https://doi.org/10.1134/S1064562422060230.

[39] Raffaele Cheula, Matteo Maestri, Giannis Mpourmpakis, Modeling morphology and catalytic activity of nanoparticle ensembles under reaction conditions, ACS Catal. 10 (2020) 6149−6158, https://doi.org/10.1021/acscatal.0c01005.

[40] Zisheng Zhang, Borna Zandkarimi, Anastassia N. Alexandrova, Ensembles of metastable states govern heterogeneous catalysis on dynamic interfaces, Accounts Chem. Res. 53 (2) (2020) 447−458, https://doi.org/10.1021/acs.accounts.9b00531.