The Open Catalyst Challenge 2021: Competition Report

Abhishek Das[*1]**, Muhammed Shuaibi**[2]**, Aini Palizhati**[2]**, Siddharth Goyal**[1]**, Aditya Grover**[1→3]**, Adeesh Kolluru**[2]**, Janice Lan**[1]**, Ammar Rizvi**[1]**, Anuroop Sriram**[1]**, Brandon Wood**[4→1]**, Devi Parikh**[1,5]**, Zachary Ulissi**[2]**, C. Lawrence Zitnick**[1]

Organizing team
[1] *Meta AI Research*
[2] *Carnegie Mellon University*
[3] *University of California, Los Angeles*
[4] *National Energy Research Scientific Computing Center (NERSC)*
[5] *Georgia Institute of Technology*

Guolin Ke[6]**, Shuxin Zheng**[6]**, Yu Shi**[6]**, Di He**[6]**, Tie-Yan Liu**[6]**, Chengxuan Ying**[7]**, Jiacheng You**[8]**, Yihan He**[2]

Winning team participants
[2] *Carnegie Mellon University*
[6] *Microsoft Research Asia*
[7] *Dalian University of Technology*
[8] *Tsinghua University*

Rostislav Grigoriev[9]**, Ruslan Lukin**[9]**, Adel Yarullin**[9]**, Max Faleev**[9]

Runner-up team participants
[9] *Innopolis University, Russia*

Abstract

In this report, we describe the Open Catalyst Challenge held at NeurIPS 2021, focusing on using machine learning (ML) to accelerate the search for low-cost catalysts that can drive reactions converting renewable energy to storable forms. Specifically, the challenge required participants to develop ML approaches for relaxed energy prediction, *i.e.* given atomic positions for an adsorbate-catalyst system, the goal was to predict the energy of the system's relaxed or lowest energy state. To perform well on this task, ML approaches need to approximate the quantum mechanical computations in Density Functional Theory (DFT). By modeling these accurately, the catalyst's impact on the overall rate of a chemical reaction may be estimated; a key factor in filtering potential electrocatalyst materials. The challenge encouraged community-wide progress on this task and the winning approach improved direct relaxed energy prediction by ∼15% relative over the previous state-of-the-art.

Keywords: catalysis, graph neural networks, renewable energy, density functional theory
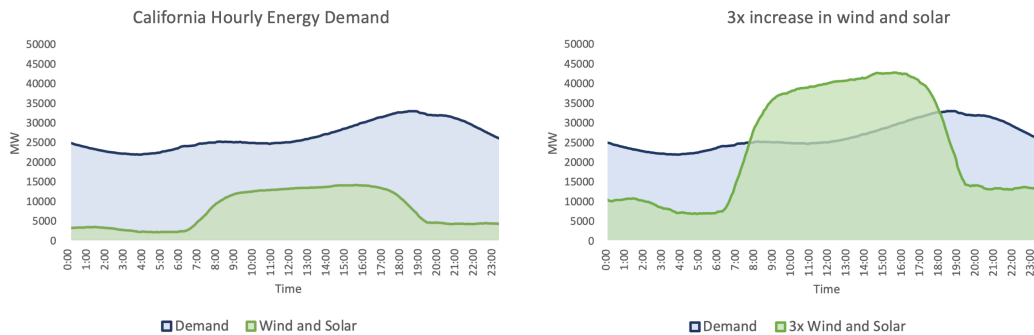
---

* corresponding author: abhshkdz@fb.com

Figure 1. Hourly electricity demand for California on a typical summer day (August 6th, 2020) as reported by California ISO (left). The demand (blue line) peaks around 19:00, the same time that output from solar and wind (green line) declines. If we naively scale up solar and wind by 3x (right), excess power is generated during the afternoon hours, but evening peak demand remains unmet. For renewable energy to meet the grid's needs, energy storage is needed to transfer excess power from times of peak generation (noon hours, summer) to times of peak demand (evenings, winter).

1. Introduction

Scalable and cost-effective solutions to renewable energy storage are essential for addressing the world's rising energy needs while mitigating climate change. As we increase our reliance on renewable energy sources such as wind and solar, which produce intermittent power, storage is needed to transfer power from times of peak generation to peak demand (Fig. 1). This may require the storage of power for hours, days, or months. One solution that offers the potential of scaling to nation-sized grids is the conversion of renewable energy to chemical fuels, such as hydrogen. To be widely adopted, this process requires cost-effective solutions to running chemical reactions (Zitnick et al., 2020).

An open challenge is finding low-cost catalysts to drive these reactions at high rates (Chanussot* et al., 2021). Through the use of quantum mechanical simulations (Density Functional Theory, DFT (Sholl and Steckel, 2009)), new catalyst structures can be tested and evaluated. Unfortunately, the high computational cost of these simulations limits the number of structures that may be tested. The use of machine learning may provide a method to efficiently approximate these calculations; reducing the time required from ∼24 hours to a second. This improvement in throughput would transform the search for new catalysts from the present day practice of evaluating 1,000s of handpicked candidates to brute force search over millions or even billions of candidates.

The Open Catalyst Challenge 2021 was designed to spawn interest and research in the consequential and complex problem of discovery of catalyst materials, an important effort in meeting the world's energy needs in the decades ahead. An important quantity in screening catalysts is the adsorption energy for the molecules, referred to as 'adsorbates', involved in the reaction of interest. The adsorption energy may be found by simulating the interaction of the adsorbate molecule on the surface of the catalyst to find their resting or relaxed energy, *i.e.* how tightly the adsorbate binds to the catalyst's surface. The rate of the chemical reaction, a value of high practical importance, is then commonly approximated using simple functions of the adsorption energy. Specifically, given an initial set of atomic positions for an adsorbate-catalyst structure, the task in this competition was for ML models to predict the energy of the structure at the local minimum or relaxed state.
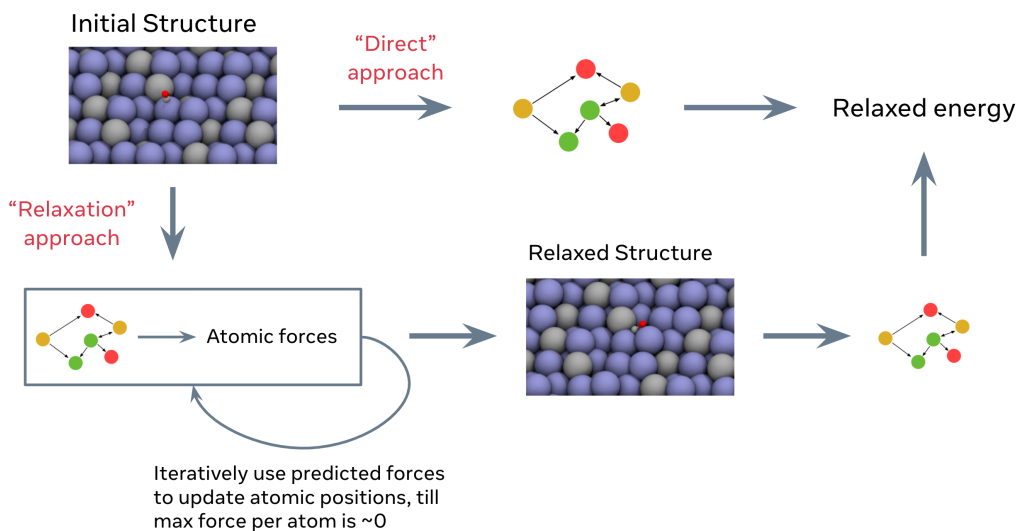
Figure 2. Two pathways to IS2RE – 1) iteratively estimating atomic forces and updating atomic positions until a relaxed state is reached and predicting the energy of that state, 2) directly predicting the energy of the relaxed state from the initial state, without estimating intermediate states.

As part of the broader Open Catalyst Project (opencatalystproject.org) effort, we publicly released the world's largest quantum mechanical simulation dataset – OC20 Chanussot* et al. (2021) – in the Fall of 2020 along with a suite of baselines and evaluation metrics that formed the basis for this challenge. The creation of the dataset required over 200 million hours of compute. This dataset enabled the exploration of techniques that will generalize across different catalyst materials and adsorbates. If successful, models trained on the dataset could enable the computational testing of millions of catalyst materials for a wide variety of chemical reactions. However, techniques that achieve the accuracies required for practical impact are still beyond reach and remain an open area for research.

2. Competition Details

2.1. Task – Initial Structure to Relaxed Energy (IS2RE)

The challenge consisted of one primary task – Initial Structure to Relaxed Energy (IS2RE) – as originally proposed in our OC20 dataset paper (Chanussot* et al., 2021). Here the input consists of the atomic positions for an initial structure *i.e.* an adsorbate-catalyst system, and the goal is to predict the energy of the structure's relaxed state.

Relaxed energies are a critical indicator in determining the reaction rate resulting from the use of a catalyst. By placing an adsorbate in multiple locations above a catalyst's surface and relaxing the structure, the binding site between the adsorbate and catalyst with the lowest relaxed energy can be determined. This lowest energy binding site is likely to be the one realized in practice under experimental conditions. The energy of the lowest energy binding site is also highly correlated with the reaction rate or selectivity. If successful, these techniques could be used to screen millions or even billions of potential catalyst materials for the reactions involved in renewable energy storage and solar fuel generation.
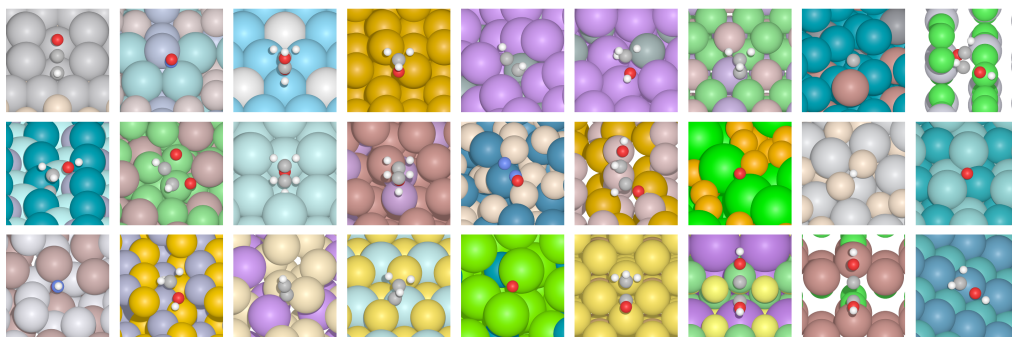
Figure 3. Illustrations of some systems from the OC20 dataset (Chanussot* et al., 2021). Each system consists of an adsorbate (the small molecule on the surface) and a catalyst (the large grid-like structure sitting below the adsorbate), and is repeated in the direction of the horizontal axes infinitely.

Traditionally, relaxed energies are found by first performing structure relaxations through an iterative local optimization process that calculates the energy gradient, *i.e.* atomic forces, using the energy from DFT. Atomic forces are in turn used to update atom positions until convergence. This very computationally expensive process typically requires hours or days of DFT calculations to converge and forms the basis of most computational catalysis efforts. One approach to the IS2RE task is to use ML to approximate DFT relaxations, *i.e.* iteratively estimate atomic forces and update atomic positions until a relaxed state is reached and finally predict the energy of that state. Evaluation of the IS2RE task on models built for approximating DFT relaxations will help determine whether this approach is sufficiently accurate and fast for practical applications. These models have the additional benefit of predicting the relaxed structure and accelerating future DFT calculations. Alternatively, it may be possible to predict the relaxed energy directly, without estimating intermediate relaxation states, as many of the changes during a relaxation (say due to particular initial guess strategies) are systematic. These direct IS2RE approaches may lead to even greater improvements in computational efficiency. These two pathways are illustrated in Fig. 2. As such, we placed no restrictions on the possible ML approaches to solve this task and used to participate in this challenge. We encouraged submissions that are significantly more computationally efficient than DFT. For example, a standard relaxation using DFT takes ~24 hours, while ML approaches have the potential to bring this down to < 10 seconds per relaxation or < 1 second per direct prediction, at least a 1000x improvement! To ensure consistent and fair evaluation, we used a public evaluation server hosted on EvalAI. Data loaders, metrics, and baseline models were made available at github.com/open-catalyst-project/ocp.

2.2. Data & Splits

The competition was conducted on the Open Catalyst Dataset (OC20) (Chanussot* et al., 2021), Fig. 3. OC20 was used to provide training and validation data for the competition. OC20 is a freely available and public dataset containing approximately 1.2M DFT relaxations, the world's largest catalyst dataset. Due to its significant scale, the dataset required over 200M hours of compute to generate. Computations were performed on servers Meta has committed to be 100% supported by renewable energy since 2020. Each relaxation contains a sequence of structures as the atoms move from an initial structure to a relaxed structure. Each structure contains the atoms corresponding to the adsorbate and catalyst.

Task	Train	Val In Domain	Val OOD Adsorbate	Val OOD Catalyst	Val OOD Both
S2EF	133,934,018	999,866	999,838	999,809	999,944
IS2RE	460,328	24,943	24,961	24,963	24,987

Table 1. Sizes of OC20 training and validation splits for the Structure to Energy & Forces (S2EF) and Initial Structure to Relaxed Energy (IS2RE) tasks. The structures for S2EF are sampled from $640,081$ relaxations for the training split, and from 30k-70k relaxations for each validation split.
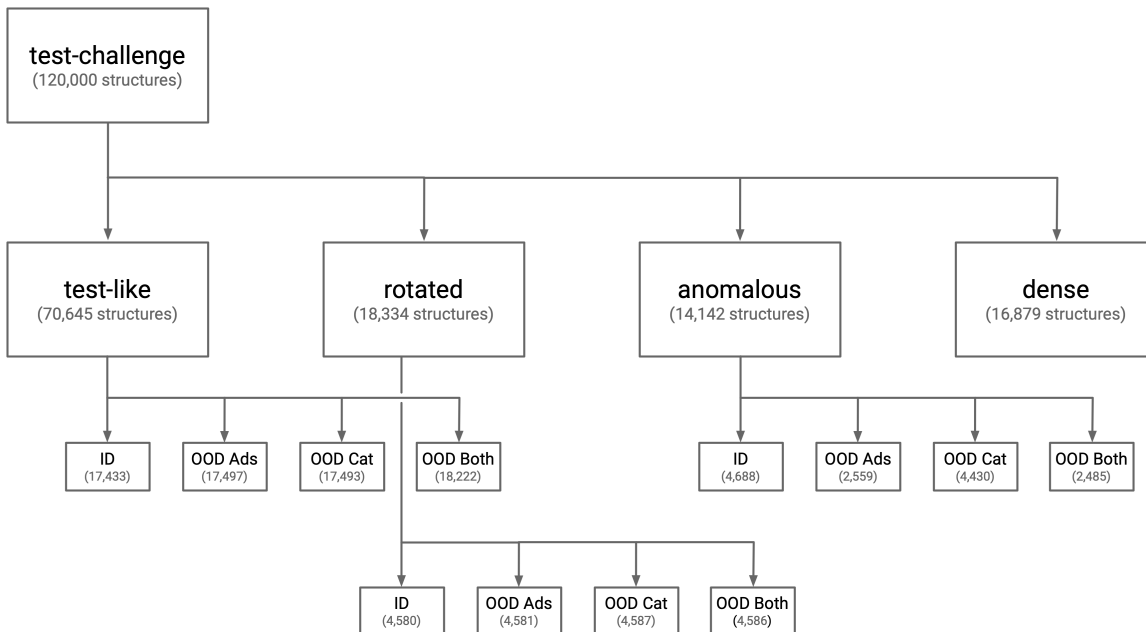

Figure 4. Overview and structure counts of all subsplits in test-challenge.

The initial structures are heuristically determined and the relaxed structures correspond to a state in which the atoms are at a local energy minima. The IS2RE training dataset, that consists of these heuristic initial structures paired with relaxed structure energies and per-atom positions as annotations, has ∼460k points. Alternatively, the Structure to Energy & Forces (S2EF) dataset, that consists of every relaxation step as a training sample, has ∼134M points. See Table 1. In the S2EF splits, for each structure, DFT-computed system energy, per-atom forces and per-atom positions are available as annotations. The OC20 validation and test splits have several subsplits to help evaluate a model's performance on interpolative and extrapolative tasks. A model's interpolative ability is evaluated on samples from the same distribution as the training dataset (In Domain). Extrapolation is evaluated on two dimensions – new adsorbates and new catalyst compositions. Subsplits are created by considering all combinations of potential extrapolations – Out-of-Domain Adsorbate (OOD Adsorbate), OOD Catalyst, and OOD Both (both unseen adsorbate and unseen catalyst).

A new **test-challenge** split, consisting of 120k structures, was released specifically for the competition to ensure there is no overfitting on the test data. Relaxed energy and force
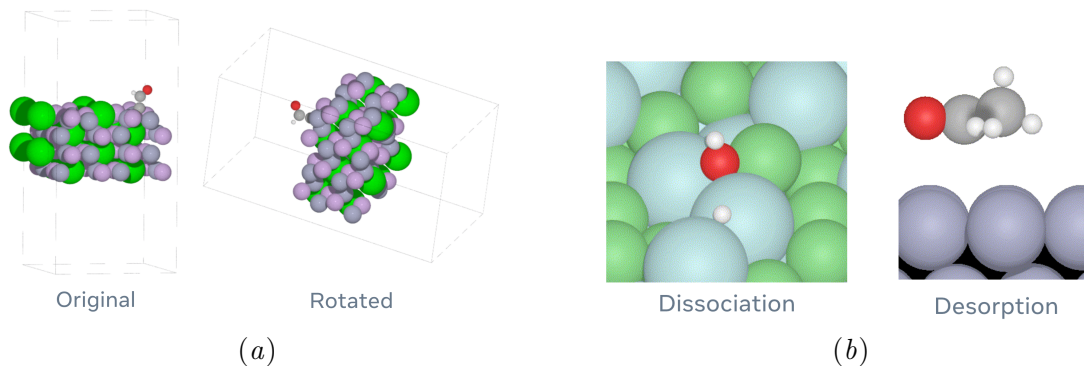
Figure 5. Example structures from the 'rotated' and 'anomalous' subsplits of test-challenge.

annotations were removed from the released test-challenge data. Only initial structures were released. Fig. 4 provides an overview of all subsplits in test-challenge, detailed below:

- Test-like: Similarly generated as the OC20 test split and used to pick winners.

- Rotated: Subset of structures in 'test-like' randomly rotated along x, y, z and useful for evaluating invariance to structure rotations. See Fig. 5(a).

- Anomalous: Structures where the adsorbate dissociates or desorps. See Fig. 5(b).

- Dense: Structures with multiple adsorbate placements per adsorbate-catalyst combination, useful for evaluating recall of the lowest energy binding site.

Participants were not provided the breakdown of test-challenge into its subsplits. They were to submit predictions on all 120k structures from test-challenge, and were limited to a maximum of 10 submissions per team to avoid overfitting. Teams did not know the results of their submissions until after the evaluation server was closed at the challenge deadline.

2.3. Evaluation Metrics

All submissions to the Open Catalyst Challenge 2021 were made to the EvalAI server and evaluated on the following metrics:

- Energy MAE: mean absolute error between the predicted relaxed energy and the DFT-computed ground-truth relaxed energy.

- Energy within Threshold (EwT): the percentage of predicted relaxed energies within 0.02 eV of the DFT-computed ground-truth relaxed energy.

Challenge winners were decided based on the Energy MAE metric. Given the large size of the OC20 trajectory data ($\sim$134M training points), we realized that resource availability could become a bottleneck for some participants. Thus, we planned to recognize 2 winners:

1. The best overall performance with no constraints on data used.

2. The best performance using *only* the IS2RE dataset (size 460,328 training points, Table 1)

Participants were prompted while making submissions on EvalAI to specify whether they used only the IS2RE dataset or not. Participants submitting to track (2) were prohibited from using any other datasets and/or pretrained models. Data augmentation was permitted as long as it comes only from the IS2RE dataset. Pretraining in any form that used trajectory data (*e.g.* from S2EF) was disallowed for track (2). Participants submitting to track (1) were free to use any dataset, and were allowed to participate in both tracks if they wished.

3. Results & Analysis

		Energy MAE (eV)			
Rank	Team	Test-like	Rotated	Anomalous	Dense
1	Microsoft Research Asia	**0.5474**	**0.5467**	**1.0312**	**0.6353**
2	Innopolis AI	0.6180	0.6170	1.1859	0.6839
3	Up and Atom	0.6694	0.6707	1.1402	0.7398
3	DIVE @ TAMU	0.6710	0.6712	1.1810	0.7398
5	RedSeaSeed	0.6830	0.6811	1.1876	0.7435
6	air	0.6973	0.6999	1.3089	0.7594
7	EnergyNet	0.7351	0.8842	1.3399	0.8033
Baseline	GemNet-dT (Gasteiger et al., 2021)	0.6410	0.6446	1.1681	0.7078
Baseline	DimeNet++ (Gasteiger et al., 2020)	0.6593	0.6625	1.1937	0.7200

Table 2. Open Catalyst Challenge results on the test-challenge split. Note that energy MAE on the 'test-like' subsplit of test-challenge was used to decide competition winners. Microsoft Research Asia won the competition with an energy MAE of 0.5474, followed by Innopolis AI with 0.6180.

The competition received 30 submissions overall, from 7 teams with participants from 6 different countries. All teams submitted to track (2) – the IS2RE-only track. Table 2 reports results across all test-challenge subsplits. Team 'Microsoft Research Asia' won the competition with an energy MAE of 0.5474 eV on test-like. Compared to our best baseline – GemNet-dT (Gasteiger et al., 2021) – this is a substantial 14.6% relative improvement in energy MAE. This approach also outperforms all other models on the other test-challenge subsplits. Team 'Innopolis AI' placed 2nd with an energy MAE of 0.6180 eV, also beating the GemNet-dT baseline by 3.59% relative. Teams 'Up and Atom' and 'DIVE @ TAMU' shared the 3rd position since their energy MAEs on test-like had overlapping standard errors.

3.1. Winning approaches

3.1.1. Microsoft Research Asia

Researchers from Microsoft Research Asia won the first prize based on their Graphormer (Ying et al., 2021) model. Unlike typical graph neural networks, Graphormer is built upon the standard Transformer architecture (Vaswani et al., 2017), which enjoys great expressive power and has good potential for scalability. In addition, compared to conventional message-passing

schemes, Graphormer benefits from a global receptive field by better capturing long-range forces, *e.g.* Coulombic forces that are non-negligible in quantum mechanical computations. The model employed a 48-layer standard Graphormer, where the parameters are shared across each 12-layer block for efficiency. A Gaussian kernel function (Schütt et al., 2017; Unke and Meuwly, 2019) is adopted as the spatial encoding, and the spatial encodings of all nodes are then summed up to obtain the centrality encoding that describes the importance of each atom in the 3D molecular graph. In addition, the output self-attention layer projects the attention scores to x, y, and z axes to keep the overall network equivariant. All implementations and model details are available at `github.com/microsoft/Graphormer`.

3.1.2. INNOPOLIS AI

Researchers from Innopolis University won the second prize based on a graph neural network approach for catalyst design (CatGNNs). Their model consisted of convolutional and message passing layers with physically-informed node and edge attributes for atomistic systems. The authors experimented with three base GNN architectures – EdgeUpdateNet, OFMNet and a variant of DimeNet++ (Gasteiger et al., 2020). The edge attributes consisted of interatomic interaction matrices – Coulomb matrix, Ewald sum matrix, and sine matrix – and the node fingerprints consisted of elements of the orbital field matrix (OFM), a one-hot representation of the electronic state of atoms with surrounding atomic orbitals. This model demonstrated improved accuracy in adsorption energy prediction, especially on out-of-domain examples for both adsorbates and catalysts, making it well-suited for use in catalyst screening. Scripts, models and configurations used are available at `github.com/AI4Materials-lab/catgnns`.

3.2. Rotational invariance

Fig. 6 compares energy MAE on the test-like subsplit against that on the rotated subsplit. Recall that the rotated subsplit consists of a subset of structures from test-like that are randomly rotated along x, y, and z axes. We find that all approaches except 'EnergyNet' demonstrate similar errors across the test-like and rotated structures, *i.e.* their energy predictions are largely robust to structure rotations, a useful property for good generalization.

3.3. Predicting dissociations and desorptions

Recall that the anomalous subsplit of test-challenge consists of adsorbate dissociations – when a chemical bond in an adsorbate breaks during relaxation – and desorptions – when the adsorbate molecule separates from the catalyst surface during relaxation because it is energetically favorable. Fig. 7(*a*) shows the energy MAE on this subsplit. We find that the winning approach performs the best here as well, but the energy MAEs are significantly worse than on the test-like subsplit (1.031 *vs.* 0.547 eV for Microsoft Research Asia).

3.4. Recall of lowest energy site

The dense subsplit consists of multiple adsorbate placements per adsorbate-catalyst system and helps determine the lowest energy binding site from a set of relaxations. This is a proxy to determine the effect of the catalyst on the rate of a chemical reaction. Realistically, the adsorption happens at a place on the catalyst surface where the adsorbate is most
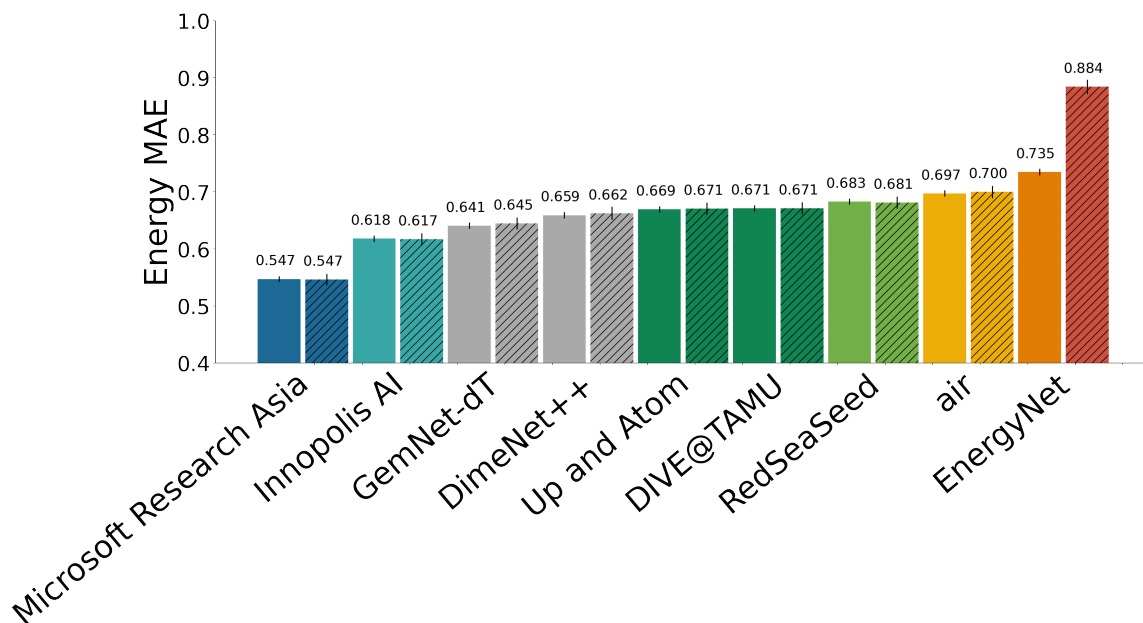
Figure 6. Energy MAE on the test-like (solid) *vs.* rotated (dashed) subsplits for all entries to the competition. Entries with overlapping standard error margins have the same color. Baselines are in grey. All entries except 'EnergyNet' demonstrate similar MAEs on test-like *vs.* rotated.
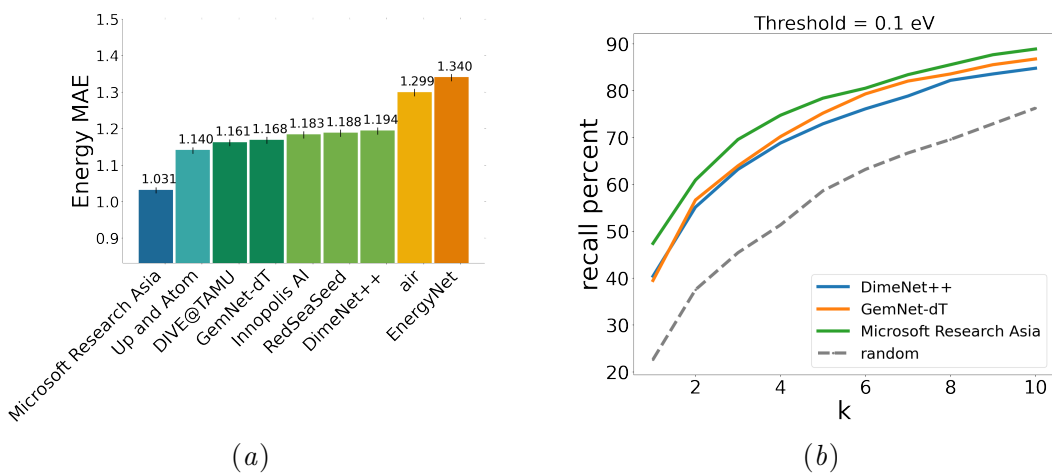

(a)   (b)

Figure 7. (a) Energy MAE on the anomalous subsplit. Entries with overlapping standard error margins have the same color. The winning approach performs the best here as well, but the overall energy MAEs are significantly worse than on the test-like subsplit. (b) Lowest energy site recall rate for the winning approach compared to the DimeNet++ and GemNet-dT baselines. This reports the percentage of times the true lowest energy site for an adsorbate-catalyst pair is included in the k lowest energy sites predicted by models. The winning approach has a recall rate of 89% at $k = 10$.

energetically stable, but we have no knowledge of where that site is in advance, especially on a complicated catalyst surface that consists of multiple elements. The dense subsplit helps solve this by enumerating all the unique sites on a catalyst surface and performing DFT relaxations of an adsorbate on each site to determine the lowest energy. However, DFT is

costly and ML could significantly reduce the computational resources required. We evaluated each entry on this subsplit using a metric called "Lowest energy site recall" (Algorithm 1).

Algorithm 1 Lowest energy site recall

Require: N^a; σ^b; k^c;
  let E_{best} = actual lowest adsorption energy of N sites
  let S_1 ..., S_k = ML-predicted best sites (*i.e.* sites ranked by ML-predicted energies)
  let E_1 ..., E_k = actual DFT energies of the ML-predicted best sites
  if $any(E_1 ..., E_k)$ - $E_{best} \leq \sigma$ **then**
| best site recalled
     end
^{a} Total number of sites enumerated
^{b} Energy threshold (eV)
^{c} Number of sites selected; $0 < k \leq N$

The σ we chose was 0.1 eV, which is a threshold deemed reasonable by the catalysis community (Grabow, 2014; Tian and Rangarajan, 2019). k ranged from 1 to 10. At each k, we calculated the lowest energy site recall percentage for the competition-winning approach as well as the DimeNet++ and GemNet-dT baselines. Fig. 7(b) shows that the winning approach was able to recall 89% of all combinations provided in the dense subsplit at $k = 10$. Since there are on average 26 heuristic adsorbate placements per adsorbate and catalyst surface combination, a recall of 89%@$k = 10$ could approximately reduce 54.7% DFT calculations.

4. Conclusion

We described details of the Open Catalyst Challenge held at NeurIPS 2021 – the Initial Structure to Relaxed Energy (IS2RE) task, test-challenge dataset, evaluation metrics, results, and key takeaways. The challenge received 30 submissions overall from 7 teams with participants from 6 different countries. The winning approach achieved an energy MAE of 0.55 eV, an improvement of $\sim$15% relative over the previous state-of-the-art for direct relaxed energy prediction. Given a dense sampling of multiple adsorbate placements per adsorbate-catalyst pair, the winning approach could recall the lowest energy site 89% of times within its $k = 10$ lowest predicted energy sites, roughly translating to a $\sim$55% reduction in DFT calculations. For context, a standard relaxation using DFT takes $\sim$24 hours. These results represent substantial progress from the community in the very first edition of our challenge, but still quite far from 'solving' the task – an energy MAE below $\sim$0.2 eV is considered accurate for practical purposes by the catalysis community – thus motivating further research in developing accurate and fast approaches for approximating DFT.

Acknowledgments

the teams who took part in the competition, as well as to the NeurIPS 2021 competition chairs for all their work in running the program smoothly.

References

Lowik Chanussot*, Abhishek Das*, Siddharth Goyal*, Thibaut Lavril*, Muhammed Shuaibi*, Morgane Riviere, Kevin Tran, Javier Heras-Domingo, Caleb Ho, Weihua Hu, Aini Palizhati, Anuroop Sriram, Brandon Wood, Junwoong Yoon, Devi Parikh, C. Lawrence Zitnick, and Zachary Ulissi. The Open Catalyst 2020 (OC20) Dataset and Community Challenges. *ACS Catalysis*, 2021.

Johannes Gasteiger, Shankari Giri, Johannes T Margraf, and Stephan Günnemann. Fast and uncertainty-aware directional message passing for non-equilibrium molecules. In *NeurIPS Workshop: Machine Learning for Molecules*, 2020.

Johannes Gasteiger, Florian Becker, and Stephan Günnemann. GemNet: Universal Directional Graph Neural Networks for Molecules. In *NeurIPS*, 2021.

Lars C. Grabow. *Computational Catalysis: Chapter 1. Computational Catalyst Screening.* Royal Society of Chemistry, 2014.

Kristof Schütt, Pieter-Jan Kindermans, Huziel Enoc Sauceda Felix, Stefan Chmiela, Alexandre Tkatchenko, and Klaus-Robert Müller. Schnet: A continuous-filter convolutional neural network for modeling quantum interactions. In *NeurIPS*, 2017.

David S. Sholl and Janice A. Steckel. *Density Functional Theory.* John Wiley & Sons, Inc., 2009.

Huijie Tian and Srinivas Rangarajan. On deriving probabilistic models for adsorption energy on transition metals using multi-level ab initio and experimental data. *arXiv preprint arXiv:1901.09253*, 2019.

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*, 2019.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. In *NeurIPS*, 2017.

Deshraj Yadav, Rishabh Jain, Harsh Agrawal, Prithvijit Chattopadhyay, Taranjeet Singh, Akash Jain, Shiv Baran Singh, Stefan Lee, and Dhruv Batra. EvalAI: Towards better evaluation systems for AI agents. *arXiv preprint arXiv:1902.03570*, 2019.

Chengxuan Ying, Tianle Cai, Shengjie Luo, Shuxin Zheng, Guolin Ke, Di He, Yanming Shen, and Tie-Yan Liu. Do transformers really perform badly for graph representation? In *NeurIPS*, 2021.

C Lawrence Zitnick, Lowik Chanussot, Abhishek Das, Siddharth Goyal, Javier Heras-Domingo, Caleb Ho, Weihua Hu, Thibaut Lavril, Aini Palizhati, Morgane Riviere, et al.

An introduction to electrocatalyst design using machine learning for renewable energy storage. *arXiv preprint arXiv:2010.09435*, 2020.