

# Open Catalyst 2020 (OC20) Dataset and Community Challenges

Lowik Chanussoot,<sup>¶</sup> Abhishek Das,<sup>¶</sup> Siddharth Goyal,<sup>¶</sup> Thibaut Lavril,<sup>¶</sup> Muhammed Shuaibi,<sup>¶</sup> 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\*



Cite This: *ACS Catal.* 2021, 11, 6059–6072



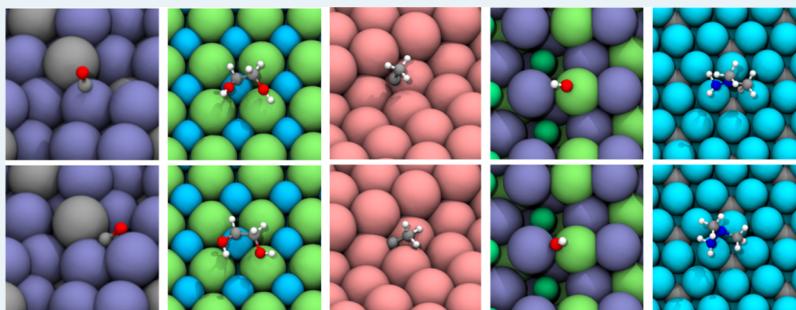
Read Online

ACCESS |

Metrics & More

Article Recommendations

Supporting Information



**ABSTRACT:** Catalyst discovery and optimization is key to solving many societal and energy challenges including solar fuel synthesis, long-term energy storage, and renewable fertilizer production. Despite considerable effort by the catalysis community to apply machine learning models to the computational catalyst discovery process, it remains an open challenge to build models that can generalize across both elemental compositions of surfaces and adsorbate identity/configurations, perhaps because datasets have been smaller in catalysis than in related fields. To address this, we developed the OC20 dataset, consisting of 1,281,040 density functional theory (DFT) relaxations (~264,890,000 single-point evaluations) across a wide swath of materials, surfaces, and adsorbates (nitrogen, carbon, and oxygen chemistries). We supplemented this dataset with randomly perturbed structures, short timescale molecular dynamics, and electronic structure analyses. The dataset comprises three central tasks indicative of day-to-day catalyst modeling and comes with predefined train/validation/test splits to facilitate direct comparisons with future model development efforts. We applied three state-of-the-art graph neural network models (CGCNN, SchNet, and DimeNet++) to each of these tasks as baseline demonstrations for the community to build on. In almost every task, no upper limit on model size was identified, suggesting that even larger models are likely to improve on initial results. The dataset and baseline models are both provided as open resources as well as a public leader board to encourage community contributions to solve these important tasks.

**KEYWORDS:** *catalysis, renewable energy, datasets, machine learning, graph convolutions, force field*

## 1. INTRODUCTION

Advancements to renewable energy processes are needed urgently to address climate change and energy scarcity around the world.<sup>1,2</sup> These include the generation of electricity through fuel cells, fuel generation from renewable resources, and the production of ammonia for fertilization. Catalysis plays a key role in each of these by enabling new reactions and improving process efficiencies.<sup>3–5</sup> Unfortunately, discovering or optimizing catalysts remains a time-intensive process. The space of possible catalyst materials that can be synthesized or engineered is vast, and modeling their full complexity under reaction conditions remains elusive. Simulation tools such as density functional theory (DFT)<sup>6</sup> have greatly expanded our field's ability to develop reaction mechanisms for specific materials, rationalize experimental measurements, and suggest more active or selective structures for experimental testing.

Despite steady growth in computing resources from Moore's law, the computational complexity of DFT remains a limiting factor in the large-scale exploration of new catalysts.<sup>7,8</sup> Given its societal importance, finding computationally efficient methods for molecular simulations is of utmost necessity. One potentially promising approach is the use of efficient machine learning (ML) models trained with data produced from computationally expensive models, such as DFT.

Received: October 19, 2020

Revised: March 16, 2021

Published: May 4, 2021



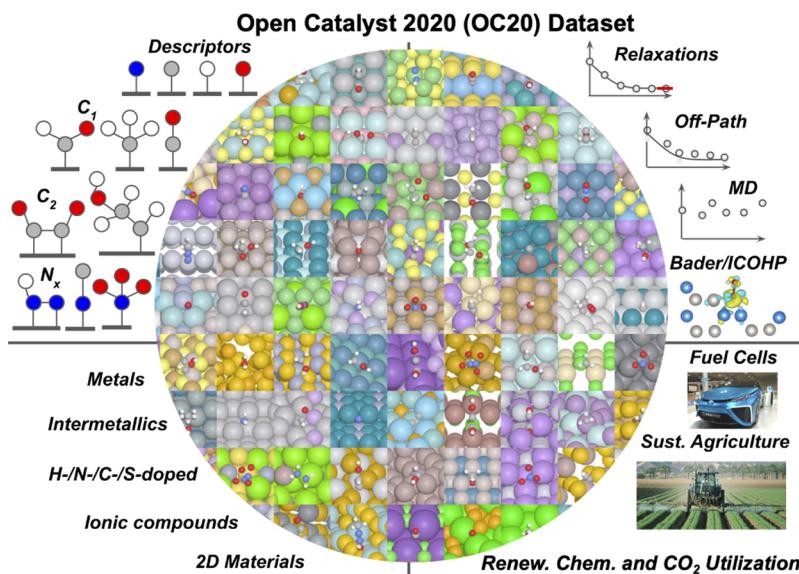


Figure 1. Adsorbates, materials, calculations, and impact areas of the OC20 dataset. Images are a random sample of the dataset.

Indeed, the application of artificial intelligence and machine learning (AI/ML) to molecular simulations has increased in popularity recently due to its ability to efficiently model complex functions in data-rich domains. There have been a number of demonstrations from domain scientists for specific challenges such as reaction network elucidation,<sup>9–11</sup> thermochemistry prediction,<sup>12–20</sup> structure optimization,<sup>21–25</sup> accelerating individual calculations,<sup>26–29</sup> and integration with characterization<sup>30</sup> (see recent reviews for a more thorough discussion<sup>8,31–43</sup>). Most of these tasks are variations on the same fundamental problem: modeling heterogeneous catalysis. The dataset developed seeks to target a specific subclass of this problem, periodic slab models. Such modeling involves predicting the energy and forces of various configurations of adsorbate molecules at inorganic interfaces.

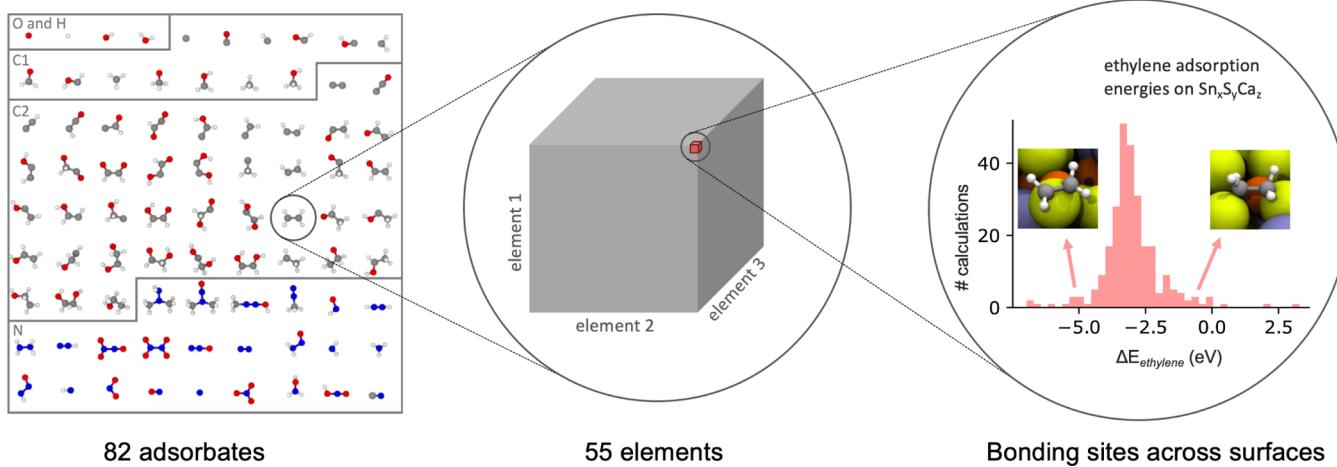
Unfortunately, modeling of heterogeneous catalysts entails all the known difficulties of modeling both organic and inorganic chemistry. In organic chemistry, modeling involves an overwhelming space of molecules and reactions and many similar, low-energy conformers. Inorganic chemistry involves a large diversity in elements, coordination environments, lattice structures, and long-range interactions. The result is a complex space of compositions and chemistries for which computationally efficient modeling methods are needed for thorough exploration.

A critical factor in building ML models is the data used for training. Despite the importance of heterogeneous catalysis, datasets for it remain smaller than those in other related fields<sup>44,45</sup> due to additional complexity and a higher computational cost. Much of the progress in applying AI/ML in heterogeneous catalysis has been driven by increasingly large and diverse datasets of electronic structure calculations. In the past few years, there has been a push toward larger datasets in catalysis, going from O(100)<sup>46–50</sup> to O(1000)<sup>51–53</sup> and then O(100,000)<sup>15,54,55</sup> relaxations. Most focus on relaxed adsorption energies of simple adsorbates with smaller datasets of transition-state calculations. State-of-the-art ML methods are still improving as data are added to these datasets, so there is no indication that we have saturated the performance of these models. Further, models trained on these datasets have shown limited ability to generalize, which suggests that the models are

not yet learning fundamental physical representations. As has been shown in other ML tasks,<sup>56–58</sup> we expect that significantly larger datasets will lead to improved accuracy and better generalization.

In this paper, we present the Open Catalyst 2020 (OC20) dataset (Figure 1), which comprises over 1.2 million DFT relaxations of molecular adsorptions onto surfaces (*ca.* 250 million single-point calculations) across a substantially larger structure and chemistry space than previously realized. We envision OC20 to serve as a crucial stepping stone in the development of ML models for practical catalysis applications.

While a dataset of this magnitude will lead to significant improvements in ML models, this is still an extremely sparse sampling of all possibilities. We consider 82 different adsorbates (small adsorbates, C<sub>1</sub>/C<sub>2</sub> compounds, and N/O-containing intermediates) that are relevant for renewable energy and environmental applications. Relaxations are performed on randomly sampled low-Miller-index facets of stable materials from the Materials Project,<sup>59</sup> resulting in surfaces from 55 different elements and mixtures thereof. For each of the calculations, we include relaxation trajectories, Bader charges, and LOBSTER<sup>60,61</sup>-calculated orbital information. To aid in training more robust models, we additionally compute short, high-temperature ab initio molecular dynamics (MD) trajectories on a randomly sampled subset of the relaxed states. We also randomly perturb the atomic positions in a subset of the structures along the relaxation pathways and perform single-point DFT calculations for these perturbed/rattled structures. We recognize that OC20 addresses a simplified version of heterogeneous catalysis—single adsorbates on idealized structures. Although useful as a first step to informing reaction pathways, the reality involves a number of additional complexities that impact catalyst performance, including reaction conditions, solvation effects, kinetics, and so forth. While we believe OC20's approximations to be a reliable step forward, it is important to understand the limits of models developed from this dataset. Future work that incorporates more of the complexities mentioned will undoubtedly benefit from the developments related to OC20. The dataset is publicly available at <http://opencatalystproject.org>. We also plan to upload the dataset



**Figure 2.** Adsorbates used to generate the Open Catalyst Dataset contain oxygen, hydrogen, C<sub>1</sub>, C<sub>2</sub>, and nitrogen molecules useful for renewable energy applications. Adsorbates that contain both carbon and nitrogen were counted both as C<sub>x</sub> adsorbates and as nitrogen-containing adsorbates. For each adsorbate, up to 55<sup>3</sup> different catalyst compositions were considered, with up to dozens of adsorption energy calculations per adsorbate–composition pairing.

to other open systems (e.g., NOMAD or Zenodo) for long-term availability.

In addition to generating and sharing the dataset, we propose three related domain challenges as an open competition: (1) predict the energy and force for a given state, (2) predict a nearby relaxed state given an initial starting state, and (3) predict the relaxed adsorption energy given an initial state. The dataset is split into train/validation/test splits indicative of common situations in catalysis: predicting these properties for a previously unseen adsorbate, for a previously unseen crystal structure or composition or both. To boot-strap research and the competition, we also provide an open software repository (<https://github.com/Open-Catalyst-Project/ocp>) containing a set of baseline models, data loaders, and training scripts for each of these tasks. While we focus on a subset of tasks, we believe that models capable of solving these tasks on the OC20 dataset will also be able to address a large number of related catalysis problems.

## 2. TASKS

Our goal is to improve the efficiency with which inorganic and organic interfaces can be simulated for use in catalysis. Since the primary computational bottlenecks are the DFT calculations used to compute a structure's forces and energy, we focus on the general challenge of efficient DFT approximation. We focus on structure relaxation—a fundamental calculation in catalysis used in determining a structure's activity and selectivity. We define three related tasks in that success in one task may aid other tasks. These are not the only possibilities for this dataset, and future tasks may be added with additional data generation and input from the community.

In all our tasks, the structure contains a surface and an adsorbate. The surface is defined by a unit cell that is periodic in all directions with a vacuum layer of at least 20 Å applied in the z-direction. Initial structures are heuristically determined. Ground truth data are computed for all tasks using DFT. Dataset details and evaluation metrics are provided in following sections.

**2.1. Structure to Energy and Forces (S2EF).** Structure to energy and forces is to take the positions of the atoms as input and predict the energy and per-atom forces as calculated

by DFT. For the purposes of this article, energy refers to adsorption energy unless otherwise noted. The adsorption energy is defined as the energy of the combined surface and adsorbate system (relaxed or not) minus the energy of the relaxed slab and the relaxed gas-phase adsorbate molecule. The force is defined as the negative gradient of the energy with respect to the atomic positions.

This is our most general task and has the broadest applicability across catalysis and related fields. It is essentially identical to existing challenges in developing ML potentials.<sup>62</sup> However, the inclusion of both inorganic and organic materials and the dataset size make this challenge unique.

### 2.2. Initial Structure to Relaxed Structure (IS2RS).

Initial structure to relaxed structure takes as input an initial structure and predicts the atomic positions in their final, relaxed state. Traditional relaxations are performed through an iterative process that estimates the atomic forces using DFT, which are in turn used to update atom positions until convergence. This very computationally expensive process typically requires hundreds of DFT calculations to converge.

If the IS2RS task is approached using ML approximations of DFT to estimate atomic forces (S2EF task), evaluation on the IS2RS task may help determine whether models built for S2EF are sufficiently accurate for practical applications. Alternatively, it may be possible to predict the relaxed structure directly, without estimating a structure's energy or forces (Figure 3B) as many of the changes during relaxation (say due to particular initial guess strategies) are systematic. These direct IS2RS approaches may lead to even further improvements in computational efficiency.

### 2.3. Initial Structure to Relaxed Energy (IS2RE).

This task is to take the initial structure as input and predict the structure's energy in the relaxed state. This is the most common task in catalysis as the relaxed energies are often correlated with catalyst activity and selectivity, and the energies are important parameters for detailed microkinetic models. Similar to IS2RS, this task may be approached by estimating the relaxed structure and energy by iteratively applying S2EF or by directly regressing the energy from the initial structure without estimating the intermediate or relaxed structures.

### 3. OC20 DATASET

The OC20 dataset is constructed to provide both training and evaluation data for our three previously defined tasks involving DFT approximation and structure relaxation. Modern ML models, especially those employing deep learning, require sufficiently large datasets to learn accurate models. For training, we provide 640,081 relaxations across a wide variety of surfaces and adsorbates. The intermediate structures and their corresponding energy and forces are provided for each relaxation resulting in over 133 million training structures. To potentially aid in training and to provide additional information for the catalysis community, we performed DFT calculations on rattled and *ab initio* MD data. We also computed Bader charges and LOBSTER analyses (over 1.8 million examples each) as these computed properties may be useful for models by explaining why the energies are what they are.

**3.1. Dataset Generation.** The dataset is constructed in four stages: (1) adsorbate selection, (2) surface selection, (3) initial structure generation, and (4) structure relaxation. We describe each of these four stages in turn, followed by a description of the additional data provided with the main dataset. All source codes to generate the configurations are provided in the Open Catalyst Dataset repository (<https://github.com/Open-Catalyst-Project/Open-Catalyst-Dataset>).

**3.1.1. Adsorbate Selection.** Adsorbates are sampled randomly from a set of 82 molecules that are chosen for their utility to renewable energy applications. As shown in Figure 2, this includes adsorbates that contain only oxygen or hydrogen, C<sub>1</sub> molecules, C<sub>2</sub> molecules, and nitrogen-containing molecules. We enumerated the oxygen and hydrogen molecules for their ubiquitous presence in water-solvated electrochemical reactions. C<sub>1</sub> and C<sub>2</sub> molecules are important for solar fuel synthesis, while nitrogen-containing molecules have applicability in solar fuel and solar chemical synthesis. Note that some of the C<sub>2</sub> molecules have two binding sites; we refer to these as bidentate adsorbates. The list of all 82 adsorbates is provided in the Supporting Information.

**3.1.2. Surface Selection.** Surfaces are sampled in three stages. First, the number of elements is selected with a 5% chance of choosing a unary material, a 65% chance for a binary material, and a 30% chance for a ternary material. Greater emphasis is given to binary and ternary materials because these sets contain a wider variety of understudied materials. Next, a stable bulk material is randomly selected from the 11,451 materials in the Materials Project<sup>59</sup> with the number of elements chosen in the first step. Finally, all symmetrically distinct surfaces from the material with Miller indices less than or equal to 2 are enumerated, including possibilities for different absolute positions of the surface plane. From this list of surfaces, one is randomly selected. The surface atoms were replicated to a depth of at least 7 Å and a width of at least 8 Å.

Pymatgen<sup>63</sup> was used to search over all bulk materials in the Materials Project with nonpositive formation energies and energies above lower hulls of at most 0.1 eV/atom. The enumeration of symmetrically distinct surfaces was also performed using Pymatgen.<sup>63</sup> Elements for the bulk materials were chosen from a set of 55 elements comprising reactive nonmetals, alkali metals, alkaline earth metals, metalloids, transition metals, and post-transition metals.

Note that DFT was used to reoptimize the bulk structures prior to surface enumeration to ensure that differences

between the DFT settings used in the Materials Project and OC20 did not induce unintended stress or strain effects. Any bulks that we could not successfully relax were omitted from this dataset.

**3.1.3. Initial Structure Generation.** The initial structures are generated by placing the selected adsorbates on the selected surfaces using CatKit<sup>64</sup> and the atomic simulation environment.<sup>65</sup> Surface atoms are identified by their positions above the center of mass, their z-distance within 2 Å of the uppermost atom, and their undercoordination relative to the bulk atoms. Atomic coordination environments were calculated using Pymatgen's Voronoi tessellation algorithm.<sup>63</sup> Next, we manually tagged the adsorbates' binding atoms for both mono- and bidentate adsorbates. Finally, we gave the surface structure, the adsorbate, the identified surface atoms, and identified adsorbate binding sites to CatKit.<sup>64</sup> CatKit used this information to enumerate a list of symmetrically distinct adsorption sites along with suggested per-site orientations for the adsorbates. From this list, an adsorption configuration is randomly selected. The sites selected are not necessarily the most stable adsorption sites on each surface. Since one of our goals is to calculate adsorption energies, we generate two sets of inputs for each system: (1) the adsorbate placed over the catalyst atoms and (2) just the catalyst atoms without the adsorbate. This resulted in a total of 1,919,165 and 616,124 unique inputs for (1) and (2), respectively, which were later filtered and segregated into suitable train, validation, and test validation splits as described later in this section.

**3.1.4. Structure Relaxation.** All structure relaxations were performed using the Vienna ab initio simulation package (VASP)<sup>66–70</sup> until all per-atom forces are less than 0.03 eV/Å. Calculations were allowed up to 144 h (12 cores) for the relaxation. Systems that timed out before reaching the specified force threshold were set aside for the S2EF task. All intermediate structures, energies, and forces are stored for future training and evaluation. During the relaxations, only adsorbate and surface atoms (as defined during the generation above) were allowed to move; subsurface atoms were maintained at fixed positions. This was done to avoid unrealistic structure deformations and to simulate a semi-infinite condition with the bulk material far below the catalyst surface. Given the intended scale of OC20, the careful consideration of DFT settings was a nontrivial challenge. Relaxations generally followed previous high-throughput catalysis efforts with reasonable trade-offs between accuracy for surface chemistry and computational cost<sup>16</sup> (VASP,<sup>66–70</sup> RPBE,<sup>71</sup> no spin polarization, etc.). The choices made for DFT were a result of several important considerations: ensuring that calculations were representative, concerns associated with inconsistent cutoffs/settings, and representative of typical numerical/convergence issues the computational chemistry field faces. The assumptions made were necessary to achieve the dataset's scale. Detecting small numerical or convergence errors is a nontrivial problem that could be improved with this dataset. Most importantly, we anticipate that models and methods that solve the S2EF, IS2RE, or IS2RS tasks for this dataset are very likely to solve future challenges for future surface science datasets with different DFT modeling choices.

System DFT energies were referenced to represent adsorption energies. Adsorption energies were calculated according to the equation below, where  $E_{\text{sys}}$  is the DFT energy of the combined surface (i.e., slab) and adsorbate—this energy can be from both relaxed and intermediate structures.

The reference energies for each system,  $E_{\text{slab}}$  and  $E_{\text{gas}}$ , are the DFT energies of the relaxed surface and adsorbate molecule, respectively. The value of  $E_{\text{gas}}$  for each adsorbate was computed as a linear combination of N<sub>2</sub>, H<sub>2</sub>O, CO, and H<sub>2</sub> resulting in the atomic energies found in the supplementary

$$E_{\text{ad}} = E_{\text{sys}} - E_{\text{slab}} - E_{\text{gas}}$$

The resulting trajectories were further analyzed for the per-atom force criterion, numerical issues, or catastrophic reconstructions as described below in the **Train, Validation, and Test Splits** section.

**3.1.5. MD and Rattled Calculations.** The intermediate structures from the relaxations may result in a dataset biased toward structures with lower energies. To learn robust models, training samples with higher forces and greater configurational diversity may be needed. We adopted two strategies for generating additional training data: (1) partial MD in VASP<sup>66–70</sup> and (2) normally distributed random position perturbation methods colloquially known in molecular simulations as “rattling.”

MD calculations simulate the atomic interactions when heat is added to the system. Partial MD calculations were carried out on previously relaxed structures with random initial velocities generated from a Maxwell–Boltzmann distribution at a temperature of 900 K. We integrated the MD trajectories over 80 or 320 fs with integration steps of 2 fs in the NVE ensemble. Timescales were selected to allow systems to explore local configurations while minding computational costs.

To diversify the distribution of single-point structures in the dataset, we “rattled” some of the structures by adding random displacements to the atomic positions with ASE.<sup>65</sup> For each relaxation, 20% of the images in the trajectories were selected for rattling. The atomic displacements were sampled from a heuristically generated normal distribution with  $\mu = 0$  and  $s = 0.05$ . Single-point DFT calculations were then performed on the rattled structures.

Similar to the relaxations, only the top surface atom layers were allowed to move in both the MD and rattled calculations with the rest of the atom positions held fixed. All calculations were performed at the same theoretical level and energy/force convergence criteria as in the relaxation calculations. Approximately 950,000 MD (*ca.* 64 million single-point energies/forces) and 30 million rattled calculations were carried out.

**3.1.6. Bader Charges and LOBSTER Analyses.** We performed electronic structure calculations for general use by the catalysis research field. These calculations (i.e., Bader charges<sup>60,72,73</sup> and LOBSTER<sup>74,75</sup> analyses) were carried out on relaxed structures and also on randomly selected snapshots from both MD and rattled trajectories. Bader charge analyses provide charge density maxima at each atomic center and the Bader volume for each atom through the zero-flux partitioning method.<sup>61</sup> LOBSTER enables chemical-bonding analysis based on periodic DFT outputs.<sup>74</sup> LOBSTER calculates atom-projected densities of states or projected crystal orbital Hamilton population curves, among others. The literature has demonstrated that such electronic structure information can provide valuable insights into the theoretical and ML communities.<sup>76–78</sup>

**3.1.7. Dataset Profile.** Approximately 872,000 adsorption energies were calculated successfully. Of these, 3.7% were calculations on unary catalysts, 61.4% were on binaries, and 34.9% were on ternaries. Among these calculations, 28.9% of

them had reactive nonmetal elements in the catalyst, 8.1% of them had alkali metals, 10.2% had alkaline earth metals, 26.4% had metalloids, 81.3% had transition metals, and 37.2% had post-transition metals. Considering adsorbates, 6.6% of the calculations had adsorbates containing only oxygen or hydrogen, 25.2% of the calculations had C<sub>1</sub> adsorbates, 44.4% had C<sub>2</sub> adsorbates, and 29.0% had nitrogen-containing adsorbates.

Despite this dataset’s large size compared to previous catalytic datasets, it still very small in comparison to the number of potential calculations. Of the  $\binom{55}{3} + \binom{55}{2} + \binom{55}{1} = 27,775$  possible compositions, only 5,243 (18.9%) of them were successfully sampled here. Of the compositions sampled, there were an average of 249 successful adsorption calculations for each. Additionally, if we compare the number of sites we sampled here to rough estimates of the number of sites we could have sampled given our constraints on adsorbates, surfaces, and bulks, then we find that we performed *ca.* 0.07% of the possible calculations. This severe sparsity in the data compared to the large scale emphasizes the need for surrogate models.

**3.2. Train, Validation, and Test Splits.** We split our dataset into training, validation, and testing sets. The training set is used to learn model parameters, the validation set is used to tune model hyperparameters and to perform ablation studies, and the test set is used to report model performance.

A careful choice of validation and test splits can help evaluate a model’s performance on both interpolative and extrapolative tasks. Interpolative evaluation tests the ability to model variations of the training data and is performed by sampling examples from the same distribution as the training dataset. Extrapolative evaluation tests a model’s performance on unseen tasks, for example, new materials or adsorbates. In the context of catalytic development, we strive to extrapolate beyond data we have already seen so that we can discover new materials and search spaces.<sup>79,80</sup>

We explore extrapolation along two dimensions, new adsorbates and new catalyst compositions. Adsorbate extrapolation is performed by holding out 14 adsorbates from the training dataset sampled from all types (O, H, C<sub>1</sub>, C<sub>2</sub>, and N) of adsorbates. Similarly, for catalyst compositions, a subset of element combinations for catalysts is held out from the training dataset. These were sampled from the 1485 binary and 26,235 ternary material combinations of the 55 elements used in the dataset. No surfaces with unary materials are in the extrapolative subsplits for validation and testing. A full list of the adsorbate materials in train and validation splits is in the **Supporting Information**.

We used four subsplits for each of the validation and test sets by considering all combinations of potential extrapolations (**Table 1**). These include the in-domain (sampled from the training distribution), the out-of-domain adsorbate (OOD adsorbate), the OOD catalyst, and OOD both (both unseen adsorbate and unseen catalyst compositions). As shown in **Table 1**, each subsplit in validation and testing contains *ca.* 25,000 relaxations. For the S2EF task, we randomly select a 1 million structure subset from the relaxations in each subsplit. Note that the extrapolative subsplits of our validation set are completely exclusive to the extrapolative subsplits in the test set, for example, the adsorbates in the validation adsorbate subsplit are unique from the adsorbates in the test adsorbate

**Table 1. Size of Test/Validation Splits (Number of Structures for S2EF and Initial Structures for IS2RS and IS2RE)<sup>a</sup>**

task	train	in domain	OOD adsorbate	OOD catalyst	OOD both
S2EF	133,934,018	999,866	999,838	999,809	999,944
IS2RS	460,328	24,943	24,961	24,963	24,987
IS2RE	460,328	24,943	24,961	24,963	24,987

<sup>a</sup>The structures for S2EF are sampled from 640,081 relaxations for the train split and from 30k to 70k relaxations for each validation and the test split. Subsplits of validation and test are of the same size but are exclusive of each other. Subsplits include sampling from the same distribution as training (in the domain), unseen adsorbates (OOD adsorbate), unseen element compositions for catalysts (OOD catalyst), and unseen adsorbates and catalysts (OOD both). Test sizes are similar.

subsplit. This helps ensure that overfitting to the test set does not occur during hyperparameter tuning on the validation set.

#### 4. BASELINE GNN MODELS

We evaluate our tasks using a set of baseline models that are representative of the current state of the art. The set of models we evaluate is by no means comprehensive, but they demonstrate what is feasible with current models. Code and pretrained models for our baseline ML approaches implemented in PyTorch Geometric<sup>81,82</sup> are publicly available at the Open Catalyst Project (<http://opencatalystproject.org>).

Our baseline ML approaches are all based on graph neural networks (GNNs)<sup>83</sup> that operate over a graph structure containing nodes and edges. In our domain, the nodes represent atoms and edges represent the relationship between neighboring atoms. At each node, an atom embedding is iteratively updated based on messages passed along the edges. During this message-passing phase, GNNs employ neural networks to learn the atomic representations,<sup>84,85</sup> and unlike traditional descriptor-based models, they do not require hand crafting. Node embeddings are initialized based on the atom's properties, such as the atomic number, group number, electronegativity, atomic volume, and so forth.<sup>86</sup> Outputs for the GNN may be computed from individual node (atom) embeddings for node-specific information (per-atom forces) or over the pooled node embeddings for system outputs (structure energy).

We benchmark three recent GNN methods: Crystal Graph Convolutional Neural Network (CGCNN),<sup>86</sup> SchNet,<sup>87</sup> and DimeNet++.<sup>88,89</sup> CGCNN is one of the first approaches to use GNNs on periodic crystal systems and uses a diverse set of features as input to the node embeddings. The original model encoded edge information using the discretized distances between atoms. SchNet proposed using continuous edge filters, which allows for the computation of per-atom forces through partial derivatives of the structure's energy with respect to the atom positions. To allow CGCNN to compute per-atom forces in the same manner, we updated the distance encoding to use Gaussian basis functions but without the envelope distance function used in SchNet in our experiments. Finally, to encode not only distance information but also angular information between triplets of atoms, DimeNet++ introduced the use of directional message passing. DimeNet++, an extension to DimeNet, replaces the Bilinear layer with a Hadamard product and additional multilayer perceptrons,

providing reported speed improvements of 8× and a 10% accuracy boost on QM9.<sup>90</sup>

For all approaches, graph edges were determined by a nearest neighbor search limited by a cutoff radius of 6 Å, retaining up to the 50 nearest neighbors. When computing distances, periodic boundary conditions were taken into consideration. Atoms were tagged as three types, slab (fixed), surface (free), and adsorbate (free), to allow loss functions to emphasize free atoms over fixed atoms. The number of hidden channels is 128, 1024, and 192 for CGCNN, SchNet, and DimeNet++, respectively, unless stated otherwise, resulting in 3.6 million (CGCNN), 7.4 million (SchNet), and 1.8 million (DimeNet++) parameters. Model sizes were chosen so that runtimes were roughly equivalent. Note that the size of the models was increased from their original implementations to account for OC20's larger size. Model hyperparameters and additional modifications can be found in the supplementary.

Since both the computed energies and forces are evaluated, the baseline loss function<sup>27,89</sup> uses the following form

$$\mathcal{L} = \lambda_E \sum_i |E_i - E_i^{\text{DFT}}| + \lambda_F \sum_{i,j} \frac{1}{N_i} |F_{i,j} - F_{i,j}^{\text{DFT}}|$$

where  $\lambda_E$  and  $\lambda_F$  are empirical parameters,  $E_i$  is the energy of image  $i$ ,  $F_{i,j}$  is the force of the  $j$ th free atom in image  $i$ , and  $N_i$  is the number of free atoms in image  $i$ . For the IS2RE task, in which only the energy is evaluated, only the first term of the loss function is used ( $\lambda_F = 0$ ).

All the models are ML-based as there are currently no physical models that operate over such a large composition space with reasonable accuracy and elemental parameterizations. In particular, the recently developed GFNO-xTB method<sup>91</sup> is parameterized for all the elements in this dataset and is fast enough (approx 1000× faster than DFT) to compete on these benchmarks, and preliminary results are reported in the Supporting Information. However, since the method was not fit for inorganic surfaces and the xTB code<sup>92</sup> is still under active development for periodic boundary conditions, the results were excluded from the summaries here. We hope that the release of our dataset will inspire future efforts on parameterizing tight-binding DFT codes or reactive force field methods for these materials.

#### 5. EXPERIMENTS

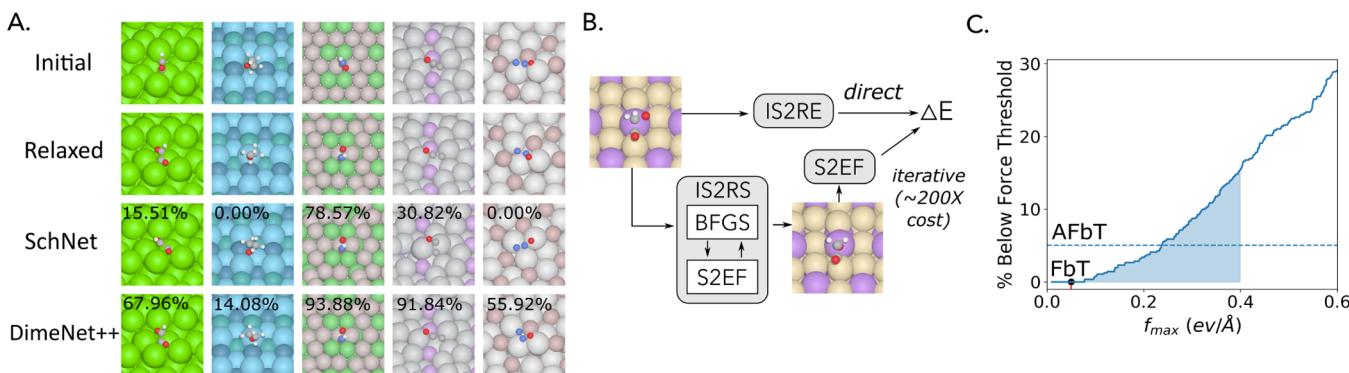
We begin by describing the metrics used to evaluate our three tasks, followed by the results of our baseline models.

**5.1. Evaluation Metrics.** For each task, we define evaluation metrics to track the progress in the field as well as to measure the practical utility of the approaches. All ground truth values are computed using DFT. Our evaluation metrics are as follows:

**5.1.1. S2EF.** The S2EF task has three metrics: the mean absolute error (MAE) for energy, MAE for forces on free atoms, and a combined metric. Our combined metric, energy and forces within threshold (EFwT), is designed to measure the practical usefulness of a model for replacing DFT by evaluating whether both the computed energy and forces are close to the ground truth.

**5.1.2. Energy MAE.** Energy MAE is MAE between the computed energy and the ground truth energy.

**5.1.3. Force MAE.** Force MAE is MAE between the computed per-atom forces and the ground truth forces. Errors are only computed for the free catalyst and adsorbate atoms.



**Figure 3.** Demonstration of baseline SchNet and DimeNet++ models for solving the IS2RE, S2EF, and IS2RS tasks and the inter-relationships. (A) Snapshots of five representative initial adsorbate configurations before DFT relaxations, the same adsorbates after DFT relaxation, and the relaxed structures as relaxed by SchNet and DimeNet++ after fitting the S2EF task. ADwT metrics are overlaid on the model snapshots. (B) Three ways to predict the relaxed energy: directly through IS2RE, indirectly through IS2RS, and confirmation of the relaxed structure with a single DFT single point. (C) SchNet force-only performance as characterized by the percentage of structures within the desired max force threshold of  $0.05 \text{ eV}/\text{\AA}$  (FbT) and an average percentage of force below threshold (AFbT) of  $0.4 \text{ eV}/\text{\AA}$  (shaded area).

**5.1.4. Force Cosine.** Force cosine is the mean cosine of the angle between the computed per-atom forces and the ground-truth forces. Similar to MAE, these are only computed for free atoms.

**5.1.5. EFwT.** EFwT is the percentage of structures in which the computed energy is within  $\epsilon = 0.02 \text{ eV}$  of the ground truth energy and the maximum error in per-atom forces is below  $\alpha = 0.03 \text{ eV}/\text{\AA}$ . Both these criteria must be met for the structure to be labeled as “correct”.

**5.1.6. IS2RS.** Several methods exist for determining the accuracy of relaxed structures predicted by ML models. The simplest is to measure the distance between the predicted 3D positions of the atoms and those of the ground truth. However, small changes in position can lead to significant changes in the per-atom forces and a structure’s energy. For this reason, a better measure of a proposed relaxed structure is the magnitude of its per-atom forces as measured by a single-point DFT calculation. If the proposed relaxed structure represents a true local energy minimum, the forces should be close to 0.

**5.1.7. ADwT.** ADwT is the average DwT (distance within threshold) across thresholds ranging from  $\beta = 0.01$  to  $\beta = 0.5 \text{ \AA}$  in increments of  $0.001 \text{ \AA}$ . DwT is computed as the percentage of structures with an atom position MAE below the threshold. MAE is only computed for the free catalyst and adsorbate atom positions while taking into account periodic boundary conditions. We use ADwT as opposed to the MAE on 3D atom positions since ADwT is robust to outliers and better indicates the percentage of relaxations that are likely to be successful.

**5.1.8. FbT.** FbT is the percentage of relaxed structures with maximum DFT-calculated per-atom force magnitudes below a threshold of  $\alpha = 0.05 \text{ eV}/\text{\AA}$ . Force magnitudes of only free catalyst and adsorbate atoms are used. A value of  $\alpha = 0.05 \text{ eV}/\text{\AA}$  represents a practical threshold by which DFT relaxations are commonly assumed to have converged. To ensure that the ML relaxations find a relaxed structure that is not significantly different from the ground truth relaxed structures, for example, the adsorbate moves to a different binding site, an additional filtering step is applied. We filter on the atom position MAE (free catalyst and adsorbate atoms) with a threshold of  $\beta = 0.5 \text{ \AA}$ . Thus, to be considered correct, a relaxed structure must meet both the FbT and DwT criteria.

**5.1.9. AFbT.** AFbT is the average FbT (forces below threshold) over a range of thresholds ranging from  $\alpha = 0.01$  to  $\alpha = 0.4 \text{ eV}/\text{\AA}$  in increments of  $0.001 \text{ eV}/\text{\AA}$ , Figure 3C. This metric measures progress over a wider range of thresholds, which may be important for early algorithm development that may need thresholds more lenient than  $\alpha = 0.05 \text{ eV}/\text{\AA}$  to see improvement. Similar to FbT, the relaxed structures must also meet the same DwT criterion with  $\beta = 0.5 \text{ \AA}$ .

Note that FbT and AFbT require the computation of single-point DFT calculations, which are computationally expensive. For this reason, a random subset of 500 relaxed structures is chosen from the validation and test set splits (2000 total for each) for evaluating these metrics. If a DFT calculation fails to converge within 60 electronic steps or a wall time of 2 h, the system is assumed to be incorrect with forces beyond the thresholds for both FbT and AFbT.

**5.1.10. IS2RE.** Similar to the S2EF task, we propose two metrics for IS2RE. The first measures the MAE on the computed and ground truth energy. The second measures the energies within a threshold (EwT) of the ground truth, which once again measures the percentage of estimated energies that are likely to be practically useful.

**5.1.11. Energy MAE.** Energy MAE is MAE between the computed relaxed energy and the ground truth relaxed energy.

**5.1.12. EwT.** EwT is the percentage of computed relaxed energies within  $\epsilon = 0.02 \text{ eV}$  of the ground truth relaxed energy.

While our evaluation metrics focus on accuracy, it is important to note that methods should also be significantly faster than conventional DFT. As a rough benchmark, we desire energy and force estimates at approximately 10 ms which would significantly improve the applicability of DFT. Significantly faster than this (closer in speed to classical force fields) would open up even more interesting applications. We ask that users self-report timing results, but we are not going to make this a core part of the challenge as computation time can likely be further optimized for the best models and with hardware acceleration.

**5.2. Leaderboard.** To ensure consistent and fair evaluation, a public leaderboard is available on the Open Catalyst Project webpage (<http://opencatalystproject.org>). The results on any of the tasks’ test datasets may be uploaded for evaluation. Ground truth test data are not publicly released to reduce potential overfitting. Evaluation on the test set may

only be done through the leaderboard. Ablation studies and hyperparameter tuning may be done and reported on using the validation datasets.

## 6. RESULTS

To provide baselines for the OC20 dataset, we report the results using three state-of-the-art approaches: CGCNN,<sup>86</sup> SchNet,<sup>87</sup> and DimeNet++.<sup>88,89</sup> Details of the models' implementations can be found in the Supporting Information.

**6.1. S2EF.** The results on CGCNN,<sup>86</sup> SchNet,<sup>87</sup> and DimeNet++<sup>88,89</sup> are evaluated. All approaches predict structure energies in their forward pass and per-atom forces by the negative gradient of the predicted energy with respect to atomic positions.<sup>93</sup> Across most metrics, DimeNet++ performs the best, with SchNet marginally outperforming DimeNet++ and CGCNN on EFwT. SchNet outperforms CGCNN across all metrics. Since trade-offs exist in the prediction of energy and forces, we trained three variants of SchNet and DimeNet++ with  $\{\lambda_E, \lambda_F\} = \{1, 30\}, \{0, 100\}, \{100, 1\}$  for SchNet/DimeNet++, SchNet/DimeNet++ force-only, and SchNet/DimeNet++ energy-only, respectively. As expected, the energy-only model performs the best on energy MAE, while the force-only model performs the best on force MAE. Both DimeNet++ and SchNet provide a balance between the two and the best results on EFwT. All approaches perform badly on the EFwT metric, indicating that the results are still far from being practically useful. Table 2 and Figure 4 show the results across subsplits. As expected, the in domain (ID) achieves the best results and the OOD both performs the worst. However, the results are not dramatically different between the in domain, OOD adsorbate, and OOD catalyst, which shows some generalization to new adsorbates and catalysts. Increases in training data sizes result in significant improvements, Figure 6A. The rate and amount of improvement vary based on the model. Finally, wider and deeper models are shown to improve accuracies in Figure 4. Both an increased depth (medium to large) and width (small to medium) show improvements.

**6.2. IS2RS.** For IS2RS, we use our S2EF baselines to drive ML relaxations from the given initial structures to estimate the relaxed structures using L-BGFS;<sup>94</sup> examples are shown in Figure 3A. Table 3 shows that DimeNet++ outperforms SchNet in the ADwT and AFbT metrics. However, the FbT metrics indicate that both methods do not produce relaxed structures with forces below thresholds used in practice. Since only the computed forces are used for the IS2RS task and not the energies, it is not surprising that the DimeNet++ force-only model performs the best. It was trained using only force losses and performs significantly better on AFbT and ADwT but still is near zero when measured by FbT. A plot of FbT across thresholds from 0.01 to 0.6 for SchNet is shown in Figure 3C. Both methods show better generalization to new adsorbates versus new catalyst material compositions. Similar to S2EF, improved results are found with more training data, especially for DimeNet++ and SchNet, Figure 6B. Experiments using the additional rattled and MD data are shown in Figure 5. Interestingly, the force cosine metric appears to better correlate with AFbT scores than force MAE. A discussion on these results may be found in the supplementary.

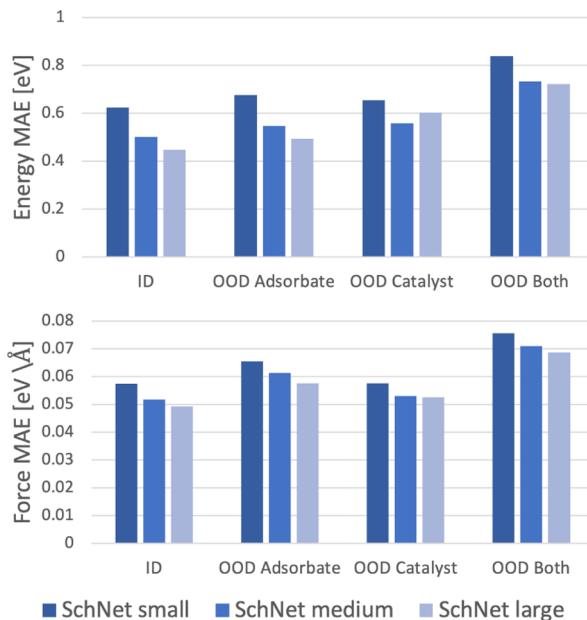
**6.3. IS2RE.** For IS2RE, we explore two pathways for computing the relaxed energy from the initial state, Figure 3B. The first directly computes the relaxed energy given the initial state. The same model architectures are used as the S2EF task, but with new weights learned. The second approach uses

**Table 2. Predicting Energy and Forces from a Structure (S2EF) as Evaluated by MAE of the Energies, Force MAE, and the Percentage of EFwT<sup>a</sup>**

Model	S2EF test			
	ID	OOD Ads	OOD Cat	OOD Both
Energy MAE [eV] ↓				
median baseline	2.0430	2.4203	1.9916	2.5770
CGCNN <sup>86</sup>	0.5272	0.6322	0.5372	0.7675
SchNet <sup>87</sup>	0.4426	0.4907	0.5288	0.7161
SchNet <sup>87</sup> —force-only	34.0316	33.769	35.2982	38.4652
SchNet <sup>87</sup> —energy-only	0.3948	0.4460	0.5510	0.7031
DimeNet++ <sup>88,89</sup>	0.4858	0.4702	0.5331	0.6482
DimeNet++ <sup>88,89</sup> —force-only	28.2134	28.9428	28.9069	34.9049
DimeNet++ <sup>88,89</sup> —energy-only	0.3586	0.4022	0.5060	0.6540
DimeNet++ <sup>88,89</sup> -large—force-only	29.3382	30.0365	30.0461	36.7537
Force MAE [eV/Å] ↓				
median baseline	0.0809	0.0801	0.0787	0.0978
CGCNN <sup>86</sup>	0.0684	0.0728	0.0671	0.0851
SchNet <sup>87</sup>	0.0493	0.0527	0.0508	0.0652
SchNet <sup>87</sup> —force-only	0.0443	0.0469	0.0459	0.0590
SchNet <sup>87</sup> —energy-only	0.4828	0.4984	0.4854	0.5447
DimeNet++ <sup>88,89</sup>	0.0443	0.0458	0.0444	0.0558
DimeNet++ <sup>88,89</sup> —force-only	0.0331	0.0341	0.0340	0.0417
DimeNet++ <sup>88,89</sup> —energy-only	0.3404	0.3395	0.3397	0.3639
DimeNet++ <sup>88,89</sup> -large—force-only	0.0281	0.0289	0.0312	0.0371
Force Cosine ↑				
median baseline	0.0164	0.0161	0.0151	0.0149
CGCNN <sup>86</sup>	0.1544	0.1390	0.1483	0.1469
SchNet <sup>87</sup>	0.3180	0.2960	0.2943	0.3001
SchNet <sup>87</sup> —force-only	0.3590	0.3382	0.3279	0.3401
SchNet <sup>87</sup> —energy-only	0.0870	0.0821	0.0828	0.0849
DimeNet++ <sup>88,89</sup>	0.3623	0.3470	0.3462	0.3685
DimeNet++ <sup>88,89</sup> —force-only	0.4867	0.4717	0.4601	0.4955
DimeNet++ <sup>88,89</sup> —energy-only	0.1065	0.0959	0.1046	0.1015
DimeNet++ <sup>88,89</sup> -large—force-only	0.5634	0.5502	0.5109	0.5518
EFwT ↑				
median baseline	0.01%	0.00%	0.01%	0.00%
CGCNN <sup>86</sup>	0.01%	0.00%	0.01%	0.00%
SchNet <sup>87</sup>	0.11%	0.06%	0.07%	0.01%
SchNet <sup>87</sup> —force-only	0.00%	0.00%	0.00%	0.00%
SchNet <sup>87</sup> —energy-only	0.00%	0.00%	0.00%	0.00%
DimeNet++ <sup>88,89</sup>	0.10%	0.03%	0.05%	0.01%
DimeNet++ <sup>88,89</sup> —force-only	0.00%	0.00%	0.00%	0.00%
DimeNet++ <sup>88,89</sup> —energy-only	0.00%	0.00%	0.00%	0.00%
DimeNet++ <sup>88,89</sup> -large—force-only	0.00%	0.00%	0.00%	0.00%

<sup>a</sup>Results reported for model training on the entire training dataset.

models trained on the S2EF task to perform ML relaxations from which the resulting energy is returned. Note that the ML relaxation approach is about 200 times more expensive to compute since energies need to be computed at each relaxation step.



**Figure 4.** Predicting structure to energy and forces (S2EF) as evaluated by MAE of the energies and forces. The small, medium, and large SchNet models have the following sizes: small: 256 hidden, 4 message-passing layers, 1,316,097 params; medium: 1024 hidden, 3 message-passing layers, 5,704,193 params; large: 1024 hidden, 4 message-passing layers, 7,396,353 params. Results reported for models trained on the entire training dataset.

As shown in Table 4, the direct approaches outperformed those using relaxation across all metrics. The percentage of predicted energies within a tight threshold (EwT) ranged from 2 to 4%, indicating that accuracies are still below practical usefulness. Generalization to new catalyst compositions performed better than new adsorbates. As shown in Figure 6C, larger dataset sizes could significantly improve performance. For the relaxation-based approaches, DimeNet++ performed the best. The use of DimeNet++ force-only to perform the relaxation, followed by DimeNet++ energy-only to compute the relaxed energy, slightly outperformed the use of a single model (optimized for EFwT) to compute both.

## 7. OUTLOOK AND FUTURE DIRECTIONS

The baseline models in this work give significant insights into the complexity of day-to-day challenges in catalysis and what it will take to achieve generalizable models. Motivated by previous efforts,<sup>97</sup> we analyzed model performance for increasing dataset sizes to illustrate the differences between catalysis and related efforts—for example, materials sciences or small-molecule property prediction. Figure 7 (left) and Figure 7(middle) show the performance of GNN models similar to the baseline models in this work on datasets for small molecules (QM9) and materials (formation energies from the materials project). The scaling of model accuracy with respect to dataset size is related to the effective dimensionality of the task and the effective representation in the model. Comparing DimeNet++ performance across all three tasks shows that the aggressive scaling for small molecules is reduced for inorganic materials and further reduced for surfaces. Focusing on the results from this study in Figure 7(right) shows that the scaling is similar for the same baseline models trained on the OC20 dataset and a related literature dataset of CO adsorption energies (see the Supporting Information). Importantly, this

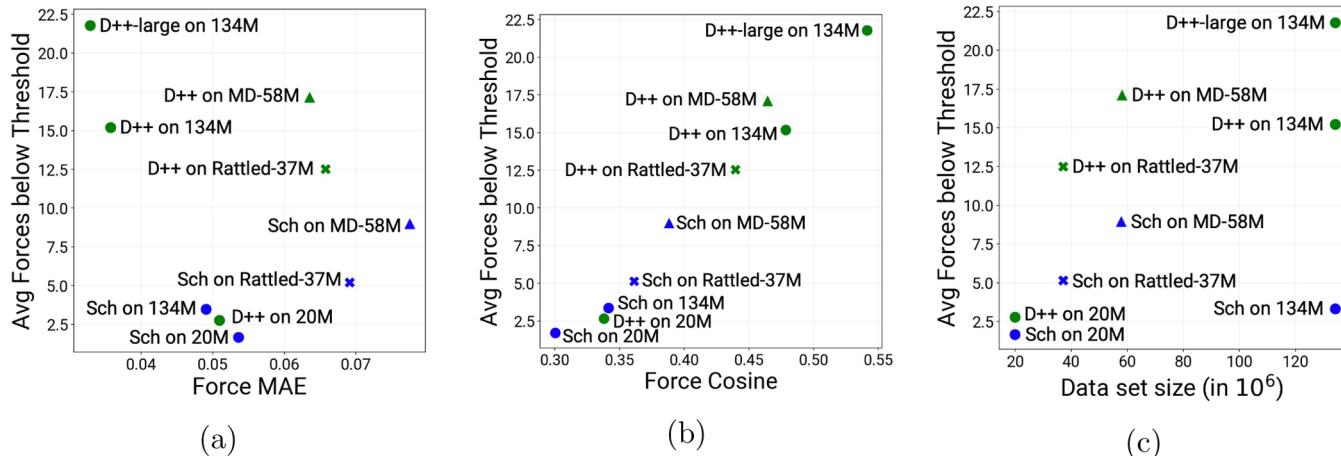
**Table 3. Predicting Relaxed Structure from Initial Structure (IS2RS) as Evaluated by Average Distance within Threshold (ADwT), Forces below Threshold (FbT), and Average Forces below Threshold (AFbT)<sup>a</sup>**

model	IS2RS test			
	ID (%)	OOD Ads (%)	OOD Cat (%)	OOD both (%)
ADwT ↑				
IS baseline	21.39	19.09	21.45	26.25
SchNet <sup>87</sup>	15.69	13.21	14.61	15.29
SchNet <sup>87</sup> —force-only	32.49	28.59	30.99	35.08
DimeNet++ <sup>88,89</sup>	30.65	26.66	30.06	32.27
DimeNet++ <sup>88,89</sup> —force-only	48.76	45.19	48.59	53.14
DimeNet++ <sup>88,89</sup> -large—force-only	52.45	48.47	50.99	54.82
FbT ↑				
IS baseline	0.00	0.00	0.00	0.00
SchNet <sup>87</sup>	0.00	0.00	0.00	0.00
SchNet <sup>87</sup> —force-only	0.00	0.00	0.00	0.00
DimeNet++ <sup>88,89</sup>	0.00	0.20	0.00	0.00
DimeNet++ <sup>88,89</sup> —force-only	0.60	0.20	0.00	0.20
DimeNet++ <sup>88,89</sup> -large—force-only	1.00	0.40	0.00	0.20
AFbT ↑				
IS baseline	0.05	0.34	0.18	0.00
SchNet <sup>87</sup>	5.28	2.82	2.62	2.73
SchNet <sup>87</sup> —force-only	3.69	3.01	2.78	2.32
DimeNet++ <sup>88,89</sup>	17.52	14.67	14.32	14.43
DimeNet++ <sup>88,89</sup> -large—force-only	25.65	20.73	20.24	20.67

<sup>a</sup>All values in percentages; higher is better. Results reported for structure to force models trained on the all training datasets. The initial structure was used as a naive baseline (IS baseline). FbT and AFbT metrics are only computed when ADwT metrics are greater than 20.26%.

suggests that achieving the desired accuracy using the current baseline models would require a dataset nearly 10 orders of magnitude larger than the current dataset. This implies that this problem will not be solved through brute-force methods alone and that significantly improved ML representations are also necessary. This is an exciting opportunity for the broader community.

For the computer science and ML communities, we expect that this dataset will provide unique challenges and spur innovation in atomistic simulations. Many state-of-the-art methods for organic and inorganic materials are based on graph convolutional networks,<sup>86</sup> which have seen rapid progress. With the above perspective, we expect that additional creative solutions will be necessary to fully solve these tasks. While they have not been demonstrated for inorganic materials, physics-informed tensor representations for small molecules may be helpful.<sup>98–101</sup> Element embeddings and representations will be important to scale across materials. Incorporation of lower-level physics-based potentials is welcomed and encouraged. This includes the use of related datasets (organic molecules or inorganic materials) for pretraining or learning priors. Incorporating other electronic features in the training set, such as charge distribution to

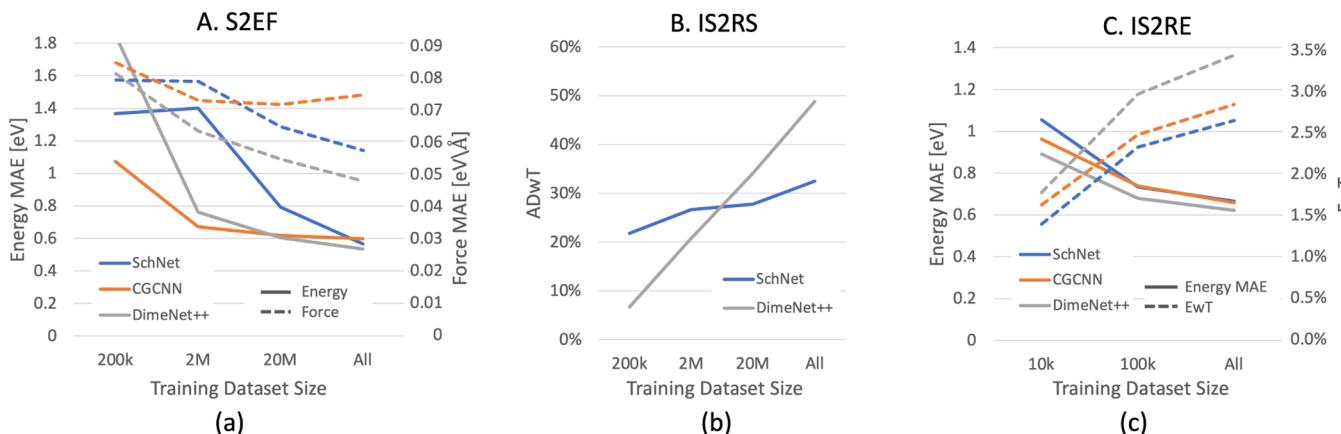


**Figure 5.** Results of force-only SchNet (denoted by “Sch”) and DimeNet++ (“D++”) S2EF models trained on S2EF-20M, S2EF-100M, S2EF-20M + Rattled (“Rattled-37M”), and S2EF-20M + MD (“MD-58M”) dataset splits used to drive relaxations from given initial structures (IS2RS). We plot IS2RS AFbT performance against S2EF force cosine, S2EF force MAE, and the number of training samples for the different variants. Sa,Sb: IS2RS AFbT appears to correlate better with S2EF force cosine than S2EF force MAE, especially when analyzing models trained on Rattled-37M or MD-58M data. Sc: further, both DimeNet++ and SchNet achieve higher AFbT when trained on MD-58M than S2EF-134M. Additional MD data appear to offer a stronger learning signal than additional S2EF data.

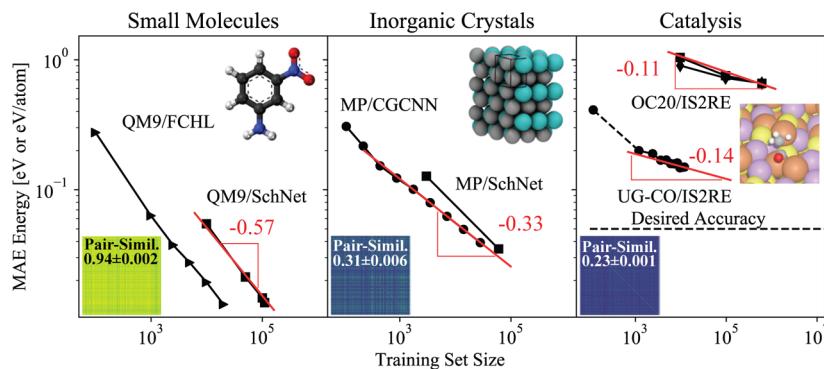
**Table 4. Predicting Relaxed State Energy from Initial Structure (IS2RE) as Evaluated by MAE of the Energies and the Percentage of Energies within a Threshold of the Ground Truth Energy<sup>a</sup>**

Model	approach	IS2RE test			EwT ↑				
		energy MAE [eV] ↓			ID (%)	OOD Ads (%)	OOD Cat (%)	OOD both (%)	
		ID	OOD Ads	OOD Cat					
median baseline		1.7499	1.8793	1.7090	0.71	0.72	0.89	0.74	
CGCNN <sup>86</sup>	direct	0.6149	0.9155	0.6219	0.8511	3.40	1.93	3.10	2.00
SchNet <sup>87</sup>	direct	0.6387	0.7342	0.6616	0.7037	2.96	2.33	2.94	2.21
DimeNet++ <sup>88,89</sup>	direct	0.5620	0.7252	0.5756	0.6613	4.25	2.07	4.10	2.41
SchNet <sup>87</sup>	relaxation	1.8630	1.9351	2.1367	2.0835	0.40	0.80	0.39	0.39
SchNet <sup>87</sup> —force-only + energy-only	relaxation	1.6643	1.6948	1.9577	1.8270	0.59	0.40	0.40	0.59
DimeNet++ <sup>88,89</sup>	relaxation	1.4721	1.4191	1.7509	1.6109	0.83	0.96	0.52	0.70
DimeNet++ <sup>88,89</sup> —force-only + energy-only	relaxation	1.3378	1.2836	1.6180	1.5775	0.87	1.01	0.59	0.76

<sup>a</sup>Results reported for models trained on all training datasets.



**Figure 6.** (A) Predicting energy and forces from a structure (S2EF) as evaluated by MAE of the energies and forces. (B) Predicting relaxed structure from initial structure (IS2RS) as evaluated by average distance within threshold (ADwT) using force-only models. (C) Predicting relaxed state energy from initial structure (IS2RE) as evaluated by MAE of the energies and the percentage of EwT ( $\epsilon = 0.02 \text{ eV}$ ) of the ground truth energy. Results reported for S2EF and IS2RS trained on 200k, 2M, 20M, and all dataset sizes. Results reported for IS2RE trained on 10k, 100k, and all dataset sizes. S2EF and IS2RE values averaged across validation subsplits. IS2RS values evaluated on the test in-domain (ID) subsplit.



**Figure 7.** Model performance versus dataset size across three related atomicistic domains. The insets show pairwise similarity for selected structures from the respective dataset using GraphDot (see the Supporting Information for details) (0/dark-blue/not-similar to 1/yellow/identical).<sup>95,96</sup> (left) Results<sup>62</sup> for FCHL/SchNet models trained on the QM9 small molecule dataset (slope  $-0.57$ ). (middle) Models<sup>86,87</sup> trained on Materials Project formation energies (slope  $-0.33$ , more difficult). (right) Results for catalysis including a literature dataset for CO adsorbates<sup>16</sup> and this work (slope  $-0.11$  to  $-0.14$ , most difficult). Note that reaching the desired accuracy will require several orders of magnitude more data with current models.

correctly localize effects, is also an opportunity to effectively reduce the dimensionality of the problem.

Note that the size of this dataset is larger by 2 orders of magnitude than previous catalyst DFT dataset efforts.<sup>16,102</sup> Along with the potential for more accurate ML models, it provides practical challenges to training atomistic ML models at scale, similar to software engineering challenges in image recognition and NLP.<sup>103,104</sup> The largest baseline models with *ca.* 10 million parameters were trained on upward of 32 GPUs at a time, so we encourage the catalysis community to take advantage of these GPU-enabled resources. This is well timed with the wave of large GPU-enabled supercomputers that are well-suited to these challenges, such as Perlmutter (DOE NERSC) or Summit (DOE OLCF), among many others.

The baseline models in this work represent the state of the art for deep learning methods to predict thermochemistry for small molecules on inorganic surfaces. Solving this challenge with future model development efforts would enable a new generation of computational chemistry methods. In particular, on-the-fly thermochemistry for reaction intermediates would enable reaction mechanism prediction across materials or composition space. Accelerated methods would also enable the more routine use of more accurate computational methods (e.g., hybrid, exact-exchange, or RPA calculations) by focusing these efforts on the most promising and prerelaxed structures. A solution to the S2EF task would enable transition-state calculations, kinetic approximations, vibrational frequency calculations, and the more routine use of long timescale MD for studying these systems. Sensitivity analyses will be necessary to understand the level of accuracy needed for models to be practically relevant for varying applications. Given the sparsity and breadth of OC20, the availability of relevant experimental data will also be a crucial challenge in the next stage of validating model results with experiments. The potential applicability of the OC20 dataset is not just catalysis but also has implications for areas where organic and inorganic materials interact, such as water quality remediation, geochemistry, advanced manufacturing, and durable energy materials.

## ASSOCIATED CONTENT

### Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/acscatal.0c04525>.

Details on the precise DFT calculation methods, adsorption energy reference energies, adsorbates and their assuming binding configurations, details on graph construction, description of the graph similarity metrics, few sample GFN0-XTB relaxations, precise train/test/validation splits, details on the modified CGCNN/SchNet/DimeNet++ implementations, results on the Rattled/MD experiments, hyperparameters for baseline models, list of adsorbates in OC20, and full results on the validation splits ([PDF](#))

## AUTHOR INFORMATION

### Corresponding Authors

C. Lawrence Zitnick – Facebook AI Research (FAIR), Menlo Park, California 94025, United States; Email: [zitnick@fb.com](mailto:zitnick@fb.com)

Zachary Ulissi – Department of Chemical Engineering and Scott Institute for Energy Innovation, Carnegie Mellon University, Pittsburgh, Pennsylvania 15213, United States; [orcid.org/0000-0002-9401-4918](https://orcid.org/0000-0002-9401-4918); Email: [zulissi@andrew.cmu.edu](mailto:zulissi@andrew.cmu.edu)

### Authors

Lowik Chanussot – Facebook AI Research (FAIR), Menlo Park, California 94025, United States

Abhishek Das – Facebook AI Research (FAIR), Menlo Park, California 94025, United States

Siddharth Goyal – Facebook AI Research (FAIR), Menlo Park, California 94025, United States

Thibaut Lavril – Facebook AI Research (FAIR), Menlo Park, California 94025, United States

Muhammed Shuaibi – Department of Chemical Engineering, Carnegie Mellon University, Pittsburgh, Pennsylvania 15213, United States

Morgane Riviere – Facebook AI Research (FAIR), Menlo Park, California 94025, United States

Kevin Tran – Department of Chemical Engineering, Carnegie Mellon University, Pittsburgh, Pennsylvania 15213, United States

Javier Heras-Domingo – Department of Chemical Engineering, Carnegie Mellon University, Pittsburgh, Pennsylvania 15213, United States; [orcid.org/0000-0002-4322-3146](https://orcid.org/0000-0002-4322-3146)

- Caleb Ho – Facebook AI Research (FAIR), Menlo Park, California 94025, United States
- Weihua Hu – Computer Science Department, Stanford University, Stanford, California 94305, United States
- Aini Palizhati – Department of Chemical Engineering, Carnegie Mellon University, Pittsburgh, Pennsylvania 15213, United States
- Anuroop Sriram – Facebook AI Research (FAIR), Menlo Park, California 94025, United States
- Brandon Wood – National Energy Research Scientific Computing Center (NERSC), Berkeley, California 94720, United States
- Junwoong Yoon – Department of Chemical Engineering, Carnegie Mellon University, Pittsburgh, Pennsylvania 15213, United States;  [orcid.org/0000-0003-0206-5978](https://orcid.org/0000-0003-0206-5978)
- Devi Parikh – Facebook AI Research (FAIR), Menlo Park, California 94025, United States; School of Interactive Computing, Georgia Tech, Atlanta, Georgia 30332, United States

Complete contact information is available at:  
<https://pubs.acs.org/10.1021/acscatal.0c04525>

## Author Contributions

<sup>¶</sup>L.C., A.D., S.G., T.L., and M.S. Co-first Author.

## Notes

The authors declare no competing financial interest.  
The full open dataset is provided at <http://opencatalystproject.org> in accessible extxyz format, and the baseline models are provided as an open source repository at <https://github.com/Open-Catalyst-Project/ocp>.

## ACKNOWLEDGMENTS

This research used resources of the National Energy Research Scientific Computing Center (NERSC), a U.S. Department of Energy Office of Science User Facility operated under Contract no. DE-AC02-05CH11231. B.W. acknowledges support from the NERSC Early Science Application Program. The authors acknowledge very helpful discussions with John Kitchin (CMU), Dionisios Vlachos (UD), and Philippe Sautet (UCLA) on the dataset construction and usage.

## REFERENCES

- (1) Newell, R. G.; Raimi, D.; Villanueva, S.; Prest, B. *Global Energy Outlook 2020: Energy Transition or Energy Addition? With Commentary on Implications of the COVID-19 Pandemic; Resources for the Future*, 2020.
- (2) *Annual Energy Outlook 2020*; U.S. Energy Information Administration, 2020.
- (3) Nørskov, J. K.; Studt, F.; Abild-Pedersen, F.; Bligaard, T. *Fundamental Concepts in Heterogeneous Catalysis*; John Wiley & Sons, 2014; pp 1–4.
- (4) She, Z. W.; Kibsgaard, J.; Dickens, C. F.; Chorkendorff, I.; Nørskov, J. K.; Jaramillo, T. F. Combining theory and experiment in electrocatalysis: Insights into materials design. *Science* **2017**, 355, No. eaad4998.
- (5) Nørskov, J. K.; Bligaard, T. *The Catalyst Genome*; Wiley, 2013.
- (6) Sholl, D. S.; Steckel, J. A. *Density Functional Theory*; John Wiley & Sons, Inc.: Hoboken, NJ, USA, 2009; pp 1–31.
- (7) Matera, S.; Schneider, W. F.; Heyden, A.; Savara, A. Progress in Accurate Chemical Kinetic Modeling, Simulations, and Parameter Estimation for Heterogeneous Catalysis. *ACS Catal.* **2019**, 9, 6624–6647.
- (8) Medford, A. J.; Kunz, M. R.; Ewing, S. M.; Borders, T.; Fushimi, R. Extracting Knowledge from Data through Catalysis Informatics. *ACS Catal.* **2018**, 8, 7403–7429.
- (9) Ulissi, Z. W.; Medford, A. J.; Bligaard, T.; Nørskov, J. K. To address surface reaction network complexity using scaling relations machine learning and DFT calculations. *Nat. Commun.* **2017**, 8, 14621.
- (10) Li, B.; Rangarajan, S. Designing compact training sets for data-driven molecular property prediction through optimal exploitation and exploration. *Mol. Syst. Des. Eng.* **2019**, 4, 1048–1057.
- (11) Gu, G. H.; Plechac, P.; Vlachos, D. G. Thermochemistry of gas-phase and surface species via LASSO-assisted subgraph selection. *React. Chem. Eng.* **2018**, 3, 454–466.
- (12) Liu, Y.; Hong, W.; Cao, B. Machine learning for predicting thermodynamic properties of pure fluids and their mixtures. *Energy* **2019**, 188, 116091.
- (13) Jirasek, F.; Alves, R. A. S.; Damay, J.; Vandermeulen, R. A.; Bamler, R.; Bortz, M.; Mandt, S.; Kloft, M.; Hasse, H. Machine Learning in Thermodynamics: Prediction of Activity Coefficients by Matrix Completion. *J. Phys. Chem. Lett.* **2020**, 11, 981–985.
- (14) Kauwe, S. K.; Graser, J.; Vazquez, A.; Sparks, T. D. Machine learning prediction of heat capacity for solid inorganics. *Integrat. Mater. Manufact. Innovat.* **2018**, 7, 43–51.
- (15) Tran, K.; Neiswanger, W.; Yoon, J.; Zhang, Q.; Xing, E.; Ulissi, Z. W. Methods for comparing uncertainty quantifications for material property predictions. *Mach. Learn. Sci. Technol.* **2020**, 1, 025006.
- (16) Tran, K.; Ulissi, Z. W. Active learning across intermetallics to guide discovery of electrocatalysts for CO<sub>2</sub> reduction and H<sub>2</sub> evolution. *Nat. Catal.* **2018**, 1, 696–703.
- (17) Gu, G. H.; Noh, J.; Kim, S.; Back, S.; Ulissi, Z.; Jung, Y. Practical Deep-Learning Representation for Fast Heterogeneous Catalyst Screening. *J. Phys. Chem. Lett.* **2020**, 11, 3185–3191 PMID: 32191473.
- (18) Back, S.; Yoon, J.; Tian, N.; Zhong, W.; Tran, K.; Ulissi, Z. W. Convolutional neural network of atomic surface structures to predict binding energies for high-throughput screening of catalysts. *J. Phys. Chem. Lett.* **2019**, 10, 4401–4408.
- (19) Kim, M.; Yeo, B. C.; Park, Y.; Lee, H. M.; Han, S. S.; Kim, D. Artificial Intelligence to Accelerate the Discovery of N<sub>2</sub> Electro-reduction Catalysts. *Chem. Mater.* **2019**, 32, 709–720.
- (20) Esterhuizen, J. A.; Goldsmith, B. R.; Linic, S. Theory-Guided Machine Learning Finds Geometric Structure-Property Relationships for Chemisorption on Subsurface Alloys. *Chem* **2020**, 6, 3100–3117.
- (21) Sun, G.; Sautet, P. Metastable Structures in Cluster Catalysis from First-Principles: Structural Ensemble in Reaction Conditions and Metastability Triggered Reactivity. *J. Am. Chem. Soc.* **2018**, 140, 2812–2820.
- (22) Li, B.; Huang, C.; Li, X.; Zheng, S.; Hong, J. Non-iterative structural topology optimization using deep learning. *Comput. Aided Des.* **2019**, 115, 172–180.
- (23) Hayashi, K.; Ohsaki, M. Reinforcement Learning and Graph Embedding for Binary Truss Topology Optimization Under Stress and Displacement Constraints. *Front. Built. Environ.* **2020**, 6, 59.
- (24) Schmidt, J.; Marques, M. R.; Botti, S.; Marques, M. A. Recent advances and applications of machine learning in solid-state materials science. *npj Comput. Mater.* **2019**, 5, 1–36.
- (25) Aksöz, Z.; Preisinger, C. An Interactive Structural Optimization of Space Frame Structures Using Machine Learning. *Impact: Design with All Senses*; Springer International Publishing, 2020; pp 18–31.
- (26) Boes, J. R.; Kitchin, J. R. Neural network predictions of oxygen interactions on a dynamic Pd surface. *Mol. Simul.* **2017**, 43, 346–354.
- (27) Khorshidi, A.; Peterson, A. A. Amp: A modular approach to machine learning in atomistic simulations. *Comput. Phys. Commun.* **2016**, 207, 310–324.
- (28) Peterson, A. A. Acceleration of saddle-point searches with machine learning. *J. Chem. Phys.* **2016**, 145, 074106.
- (29) Sun, G.; Sautet, P. Toward fast and reliable potential energy surfaces for metallic Pt clusters by hierarchical delta neural networks. *J. Chem. Theory Comput.* **2019**, 15, 5614–5627.

- (30) Timoshenko, J.; Frenkel, A. I. "Inverting" X-ray Absorption Spectra of Catalysts by Machine Learning in Search for Activity Descriptors. *ACS Catal.* **2019**, *9*, 10192–10211.
- (31) Kitchin, J. R. Machine learning in catalysis. *Nat. Catal.* **2018**, *1*, 230–232.
- (32) Li, Z.; Wang, S.; Xin, H. Toward artificial intelligence in catalysis. *Nat. Catal.* **2018**, *1*, 641–642.
- (33) Goldsmith, B. R.; Esterhuizen, J.; Liu, J. X.; Bartel, C. J.; Sutton, C. Machine learning for heterogeneous catalyst design and discovery. *AIChE J.* **2018**, *64*, 2311–2323.
- (34) Schlexer Lamoureux, P.; Winther, K. T.; Garrido Torres, J. A.; Streibel, V.; Zhao, M.; Bajdich, M.; Abild-Pedersen, F.; Bligaard, T. Machine learning for computational heterogeneous catalysis. *ChemCatChem* **2019**, *11*, 3581–3601.
- (35) Li, H.; Zhang, Z.; Liu, Z. Application of artificial neural networks for catalysis: a review. *Catalysts* **2017**, *7*, 306.
- (36) Tabor, D. P.; Roch, L. M.; Saikin, S. K.; Kreisbeck, C.; Sheberla, D.; Montoya, J. H.; Dwarakanath, S.; Aykol, M.; Ortiz, C.; Tribukait, H.; et al. Accelerating the discovery of materials for clean energy in the era of smart automation. *Nat. Rev. Mater.* **2018**, *3*, 5–20.
- (37) Chen, C.; Zuo, Y.; Ye, W.; Li, X.; Deng, Z.; Ong, S. P. A Critical Review of Machine Learning of Energy Materials. *Adv. Energy Mater.* **2020**, *10*, 1903242.
- (38) Artrith, N. Machine learning for the modeling of interfaces in energy storage and conversion materials. *Journal of Physics: Energy* **2019**, *1*, 032002.
- (39) Gu, G. H.; Noh, J.; Kim, I.; Jung, Y. Machine learning for renewable energy materials. *J. Mater. Chem. A* **2019**, *7*, 17096–17117.
- (40) Toyao, T.; Maeno, Z.; Takakusagi, S.; Kamachi, T.; Takigawa, I.; Shimizu, K.-i. Machine Learning for Catalysis Informatics: Recent Applications and Prospects. *ACS Catal.* **2019**, *10*, 2260–2297.
- (41) Gu, G. H.; Choi, C.; Lee, Y.; Situmorang, A. B.; Noh, J.; Kim, Y. H.; Jung, Y. Progress in Computational and Machine-Learning Methods for Heterogeneous Small-Molecule Activation. *Adv. Mater.* **2020**, *32*, 1907865.
- (42) Schlexer Lamoureux, P.; Winther, K. T.; Garrido Torres, J. A.; Streibel, V.; Zhao, M.; Bajdich, M.; Abild-Pedersen, F.; Bligaard, T. Machine Learning for Computational Heterogeneous Catalysis. *ChemCatChem* **2019**, *11*, 3581.
- (43) Takahashi, K.; Takahashi, L.; Miyazato, I.; Fujima, J.; Tanaka, Y.; Uno, T.; Satoh, H.; Ohno, K.; Nishida, M.; Hirai, K.; Ohyama, J.; Nguyen, T. N.; Nishimura, S.; Taniike, T. The Rise of Catalyst Informatics: Towards Catalyst Genomics. *ChemCatChem* **2019**, *11*, 1146.
- (44) Curtarolo, S.; Setyawan, W.; Wang, S.; Xue, J.; Yang, K.; Taylor, R. H.; Nelson, L. J.; Hart, G. L. W.; Sanvito, S.; Buongiorno-Nardelli, M.; Mingo, N.; Levy, O. AFLOWLIB.ORG: A distributed materials properties repository from high-throughput ab initio calculations. *Comput. Mater. Sci.* **2012**, *58*, 227–235.
- (45) Kirklin, S.; Saal, J. E.; Meredig, B.; Thompson, A.; Doak, J. W.; Aykol, M.; Rühl, S.; Wolverton, C. The Open Quantum Materials Database (OQMD): Assessing the accuracy of DFT formation energies. *npj Comput. Mater.* **2015**, *1*, 15010.
- (46) Calle-Vallejo, F.; Martínez, J. I.; García-Lastra, J. M.; Sautet, P.; Loffreda, D. Fast Prediction of Adsorption Properties for Platinum Nanocatalysts with Generalized Coordination Numbers. *Angew. Chem., Int. Ed.* **2014**, *53*, 8316–8319.
- (47) Abild-Pedersen, F.; Greeley, J.; Studt, F.; Rossmeisl, J.; Munter, T.; Moses, P. G.; Skulason, E.; Bligaard, T.; Nørskov, J. K. Scaling Properties of Adsorption Energies for Hydrogen-containing Molecules on Transition-metal Surfaces. *Phys. Rev. Lett.* **2007**, *99*, 016105.
- (48) Ma, X.; Xin, H. Orbitalwise Coordination Number for Predicting Adsorption Properties of Metal Nanocatalysts. *Phys. Rev. Lett.* **2017**, *118*, 036101.
- (49) Noh, J.; Back, S.; Kim, J.; Jung, Y. Active learning with Non-ab Initio Input Features toward Efficient CO<sub>2</sub> Reduction Catalysts. *Chem. Sci.* **2018**, *9*, 5152–5159.
- (50) Andersen, M.; Levchenko, S. V.; Scheffler, M.; Reuter, K. Beyond Scaling Relations for the Description of Catalytic Materials. *ACS Catal.* **2019**, *9*, 2752–2759.
- (51) Dickens, C. F.; Montoya, J. H.; Kulkarni, A. R.; Bajdich, M.; Nørskov, J. K. An Electronic Structure Descriptor for Oxygen Reactivity at Metal and Metal-oxide Surfaces. *Surf. Sci.* **2019**, *681*, 122–129.
- (52) Li, Z.; Wang, S.; Chin, W. S.; Achenie, L. E.; Xin, H. High-throughput Screening of Bimetallic Catalysts Enabled by Machine Learning. *J. Mater. Chem. A* **2017**, *5*, 24131–24138.
- (53) Batchelor, T. A. A.; Pedersen, J. K.; Winther, S. H.; Castelli, I. E.; Jacobsen, K. W.; Rossmeisl, J. High-entropy Alloys as a Discovery Platform for Electrocatalysis. *Joule* **2019**, *3*, 834–845.
- (54) Mamun, O.; Winther, K. T.; Boes, J. R.; Bligaard, T. High-throughput calculations of catalytic properties of bimetallic alloy surfaces. *Sci. Data* **2019**, *6*, 76.
- (55) Winther, K. T.; Hoffmann, M. J.; Boes, J. R.; Mamun, O.; Bajdich, M.; Bligaard, T. Catalysis-Hub.org, an open electronic structure database for surface reactions. *Sci. Data* **2019**, *6*, 76.
- (56) Deng, J.; Dong, W.; Socher, R.; Li, L.-J.; Li, K.; Fei-Fei, L. Imagenet: A large-scale hierarchical image database. *2009 IEEE Conference on Computer Vision and Pattern Recognition*; IEEE, 2009; pp 248–255.
- (57) Panayotov, V.; Chen, G.; Povey, D.; Khudanpur, S. Librispeech: an asr corpus based on public domain audio books. *2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*; IEEE, 2015; pp 5206–5210.
- (58) Antol, S.; Agrawal, A.; Lu, J.; Mitchell, M.; Batra, D.; Lawrence Zitnick, C.; Parikh, D. Vqa: Visual question answering. *Proceedings of the IEEE International Conference on Computer Vision*; IEEE, 2015; pp 2425–2433.
- (59) Jain, A.; Ong, S. P.; Hautier, G.; Chen, W.; Richards, W. D.; Dacek, S.; Cholia, S.; Gunter, D.; Skinner, D.; Ceder, G.; Persson, K. A. Commentary: The Materials Project: A materials genome approach to accelerating materials innovation. *APL Mater.* **2013**, *1*, 011002.
- (60) Henkelman, G.; Arnaldsson, A.; Jónsson, H. A fast and robust algorithm for Bader decomposition of charge density. *Comput. Mater. Sci.* **2006**, *36*, 354–360.
- (61) Bader, R.; Bader, R. *Atoms In Molecules: A Quantum Theory*; International Series of Monographs on Chemistry; Clarendon Press, 1994; pp 13–52.
- (62) von Lilienfeld, O. A.; Müller, K.-R.; Tkatchenko, A. Exploring chemical compound space with quantum-based machine learning. *Nat. Rev. Chem.* **2020**, *4*, 347.
- (63) Ong, S. P.; Richards, W. D.; Jain, A.; Hautier, G.; Kocher, M.; Cholia, S.; Gunter, D.; Chevrier, V. L.; Persson, K. A.; Ceder, G. Python Materials Genomics (pymatgen): A robust, open-source python library for materials analysis. *Comput. Mater. Sci.* **2013**, *68*, 314–319.
- (64) Boes, J. R.; Mamun, O.; Winther, K.; Bligaard, T. Graph theory approach to high-throughput surface adsorption structure generation. *J. Phys. Chem. A* **2019**, *123*, 2281–2285.
- (65) Larsen, A. H.; et al. The atomic simulation environment—a Python library for working with atoms. *J. Phys.: Condens. Matter* **2017**, *29*, 273002.
- (66) Kresse, G.; Hafner, J. Ab initiomolecular-dynamics simulation of the liquid-metal-amorphous-semiconductor transition in germanium. *Phys. Rev. B* **1994**, *49*, 14251–14269.
- (67) Kresse, G.; Furthmüller, J. Efficiency of ab-initio total energy calculations for metals and semiconductors using a plane-wave basis set. *Comput. Mater. Sci.* **1996**, *6*, 15–50.
- (68) Kresse, G.; Furthmüller, J. Efficient iterative schemes for ab initio total-energy calculations using a plane-wave basis set. *Phys. Rev. B* **1996**, *54*, 11169–11186.
- (69) The calculations in this work have been performed using the ab-initio total-energy and molecular-dynamics package VASP (Vienna ab-initio simulation package) developed at the Institut für Materialphysik of the Universität Wien2,3.

- (70) Kresse, G.; Joubert, D. From ultrasoft pseudopotentials to the projector augmented-wave method. *Phys. Rev. B* **1999**, *59*, 1758.
- (71) Perdew, J. P.; Burke, K.; Ernzerhof, M. Generalized gradient approximation made simple. *Phys. Rev. Lett.* **1996**, *77*, 3865.
- (72) Tang, W.; Sanville, E.; Henkelman, G. A grid-based Bader analysis algorithm without lattice bias. *J. Phys.: Condens. Matter* **2009**, *21*, 084204.
- (73) Sanville, E.; Kenny, S. D.; Smith, R.; Henkelman, G. Improved grid-based algorithm for Bader charge allocation. *J. Comput. Chem.* **2007**, *28*, 899–908.
- (74) Nelson, R.; Ertural, C.; George, J.; Deringer, V. L.; Hautier, G.; Dronskowski, R. LOBSTER : Local orbital projections, atomic charges, and chemical-bonding analysis from projector-augmented-wave-based density-functional theory. *J. Comput. Chem.* **2020**, *41*, 1931–1940.
- (75) Deringer, V. L.; Tchougréeff, A. L.; Dronskowski, R. Crystal orbital Hamilton population (COHP) analysis as projected from plane-wave basis sets. *J. Phys. Chem. A* **2011**, *115*, 5461–5466.
- (76) Gong, S.; Xie, T.; Zhu, T.; Wang, S.; Fadel, E. R.; Li, Y.; Grossman, J. C. Predicting charge density distribution of materials using a local-environment-based graph convolutional network. *Phys. Rev. B* **2019**, *100*, 184103.
- (77) Nagai, R.; Akashi, R.; Sugino, O. Completing density functional theory by machine learning hidden messages from molecules. *npj Comput. Mater.* **2020**, *6*, 43.
- (78) Chandrasekaran, A.; Kamal, D.; Batra, R.; Kim, C.; Chen, L.; Rampasad, R. Solving the electronic structure problem with machine learning. *npj Comput. Mater.* **2019**, *5*, 122.
- (79) Kim, Y.; Kim, E.; Antono, E.; Meredig, B.; Ling, J. Machine-learned metrics for predicting the likelihood of success in materials discovery. *npj Comput. Mater.* **2019**, *6*, 131.
- (80) Meredig, B.; Antono, E.; Church, C.; Hutchinson, M.; Ling, J.; Paradiso, S.; Blaiszik, B.; Foster, I.; Gibbons, B.; Hattrick-Simpers, J.; Mehta, A.; Ward, L. Can machine learning identify the next high-temperature superconductor? Examining extrapolation performance for materials discovery. *Mol. Syst. Des. Eng.* **2018**, *3*, 819–825.
- (81) Fey, M.; Lenssen, J. E. Fast graph representation learning with PyTorch Geometric. **2019**, arXiv preprint arXiv:1903.02428.
- (82) Paszke, A.; Gross, S.; Massa, F.; Lerer, A.; Bradbury, J.; Chanan, G.; Killeen, T.; Lin, Z.; Gimelshein, N.; Antiga, L.; et al. Pytorch: An imperative style, high-performance deep learning library. *Advances in Neural Information Processing Systems* **2019**, pp 8026–8037.
- (83) Hamilton, W. L.; Ying, R.; Leskovec, J. *Representation Learning on Graphs: Methods and Applications*; IEEE Data Engineering Bulletin, 2017.
- (84) Behler, J. Perspective: Machine learning potentials for atomistic simulations. *J. Chem. Phys.* **2016**, *145*, 170901.
- (85) Bartók, A. P.; Payne, M. C.; Kondor, R.; Csányi, G. Gaussian approximation potentials: The accuracy of quantum mechanics, without the electrons. *Phys. Rev. Lett.* **2010**, *104*, 136403.
- (86) Xie, T.; Grossman, J. C. Crystal graph convolutional neural networks for an accurate and interpretable prediction of material properties. *Phys. Rev. Lett.* **2018**, *120*, 145301.
- (87) Schütt, K.; Kindermans, P.-J.; Felix, H. E. S.; Chmiela, S.; Tkatchenko, A.; Müller, K.-R. Schnet: A continuous-filter convolutional neural network for modeling quantum interactions. *Adv. Neural Inf. Process. Syst.* **2017**, 991–1001.
- (88) Klicpera, J.; Giri, S.; Margraf, J. T.; Günnemann, S. Fast and Uncertainty-Aware Directional Message Passing for Non-Equilibrium Molecules. **2020**, arXiv preprint arXiv:2011.14115.
- (89) Klicpera, J.; Groß, J.; Günnemann, S. Directional Message Passing for Molecular Graphs. *International Conference on Learning Representations (ICLR)*, 2020.
- (90) Ramakrishnan, R.; Dral, P. O.; Rupp, M.; von Lilienfeld, O. A. Quantum chemistry structures and properties of 134 kilo molecules. *Sci. Data* **2014**, *1*, 140022.
- (91) Pracht, P.; Caldeweyher, E.; Ehlert, S.; Grimme, S.; A Robust Non-Self-Consistent Tight-Binding Quantum Chemistry Method for large Molecules. **2019**, chemrxiv:8326202.v1.
- (92) Bannwarth, C.; Caldeweyher, E.; Ehlert, S.; Hansen, A.; Pracht, P.; Seibert, J.; Spicher, S.; Grimme, S. Extended tight-binding quantum chemistry methods. *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **2020**, *11*, No. e01493.
- (93) Pukrittayakamee, A.; Malshe, M.; Hagan, M.; Raff, L. M.; Narulkar, R.; Bukkapatnum, S.; Komanduri, R. Simultaneous fitting of a potential-energy surface and its corresponding force fields using feedforward neural networks. *J. Chem. Phys.* **2009**, *130*, 134101.
- (94) Liu, D. C.; Nocedal, J. On the limited memory BFGS method for large scale optimization. *Math. Program.* **1989**, *45*, 503–528.
- (95) Tang, Y.; Selvitopi, O.; Popovici, D. T.; Buluç, A. A High-Throughput Solver for Marginalized Graph Kernels on GPU. *2020 IEEE International Parallel and Distributed Processing Symposium (IPDPS)*; IEEE, 2020; pp 728–738.
- (96) Tang, Y.-H.; de Jong, W. A. Prediction of atomization energy using graph kernel and active learning. *J. Chem. Phys.* **2019**, *150*, 044107.
- (97) Huang, B.; Symonds, N. O.; von Lilienfeld, O. A. The fundamentals of quantum machine learning. **2018**, arXiv preprint arXiv:1807.04259.
- (98) Miller, B. K.; Geiger, M.; Smidt, T. E.; Noé, F. Relevance of Rotationally Equivariant Convolutions for Predicting Molecular Properties. **2020**, arXiv preprint arXiv:2008.08461.
- (99) Bratholm, L. A.; Gerrard, W.; Anderson, B.; Bai, S.; Choi, S.; Dang, L.; Hanchar, P.; Howard, A.; Huard, G.; Kim, S.; et al. A community-powered search of machine learning strategy space to find NMR property prediction models. **2020**, arXiv preprint arXiv:2008.05994.
- (100) Anderson, B.; Hy, T. S.; Kondor, R. Cormorant: Covariant molecular neural networks. *Adv. Neural Inf. Process. Syst.* **2019**, 14537–14546.
- (101) Nigam, J.; Pozdnyakov, S.; Ceriotti, M. Recursive evaluation and iterative contraction of N-body equivariant features. *J. Chem. Phys.* **2020**, *153*, 121101.
- (102) Hummelshøj, J. S.; Abild-Pedersen, F.; Studt, F.; Bligaard, T.; Nørskov, J. K. CatApp: A web application for surface chemistry and heterogeneous catalysis. *Angew. Chem. Int. Ed.* **2012**, *51*, 272–274.
- (103) He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016; pp 770–778.
- (104) Radford, A.; Wu, J.; Child, R.; Luan, D.; Amodei, D.; Sutskever, I. Language models are unsupervised multitask learners. *OpenAI Blog* **2019**, *1*, 9.