

|    |   |           |
|----|---|-----------|
| 1  | <b>Appendix Table of Contents</b>             |           |
| 2  | <b>A Training Details and Hyperparameters</b> | <b>2</b>  |
| 3  | A.1 Two-stage training . . . . .              | 2         |
| 4  | A.2 Parallelisms . . . . .                    | 2         |
| 5  | A.3 Model . . . . .                           | 2         |
| 6  | A.4 MPtrj and sAlex Fine-tuning . . . . .     | 3         |
| 7  | A.5 Training Resources . . . . .              | 3         |
| 8  | <b>B Training Data</b>                        | <b>3</b>  |
| 9  | B.1 Materials . . . . .                       | 4         |
| 10 | B.2 Molecules . . . . .                       | 4         |
| 11 | B.3 Catalysis . . . . .                       | 4         |
| 12 | B.4 Molecular Crystals . . . . .              | 5         |
| 13 | B.5 Metal-organic frameworks (MOFs) . . . . . | 5         |
| 14 | B.6 Referencing . . . . .                     | 5         |
| 15 | <b>C Scaling Laws Methods</b>                 | <b>5</b>  |
| 16 | C.1 FLOP counting . . . . .                   | 5         |
| 17 | C.2 Compute optimal fits . . . . .            | 6         |
| 18 | C.3 Fitting loss vs. parameters . . . . .     | 6         |
| 19 | <b>D Inference</b>                            | <b>7</b>  |
| 20 | <b>E Evaluations</b>                          | <b>7</b>  |
| 21 | E.1 UMA-sm-v1 Results . . . . .               | 7         |
| 22 | E.2 Materials . . . . .                       | 8         |
| 23 | E.3 Catalysis . . . . .                       | 9         |
| 24 | E.4 Molecules . . . . .                       | 9         |
| 25 | E.5 Molecular Crystals . . . . .              | 11        |
| 26 | E.6 DAC . . . . .                             | 12        |
| 27 | <b>F Single-task vs Multi-task</b>            | <b>12</b> |
| 28 | <b>G MoLE Expert Analysis</b>                 | <b>12</b> |

## 29 A Training Details and Hyperparameters

### 30 A.1 Two-stage training

31 While conservative models have been found to provide reliable performance in diverse physical  
32 property prediction tasks [16], the backward pass required for force/stress prediction significantly  
33 increases inference costs. Models with direct force prediction are much more efficient, and have been  
34 found to be effective as a pre-training strategy to save compute when subsequently fine-tuned as a  
35 conservative model [16, 7]. The UMA-sm and UMA-md both follow this procedure. In addition,  
36 we introduce low-precision training, max-atom batching, max neighbor switching, and activation  
37 checkpointing to further enhance the scalability and efficiency of our training process. These novel  
38 strategies are discussed in turn below. Detailed hyper-parameters are summarized in Table 1.

39 **Precision.** For pretraining, we used BF16 numerical format, commonly used in training LLMs but  
40 uncommon in MLIP models due to high numerical precision requirements. In our experiments, we  
41 found that BF16 is significantly more stable than AMP-FP16 (automatic mixed precision) especially  
42 in our multi-modal setting where data distributions can vary dramatically, frequently causing gradient  
43 and loss spikes that would destabilize AMP training. However, it suffers an accuracy drop of 20–50%  
44 compared to AMP-FP16 and FP32. We found the degradation can be nearly completely recovered  
45 after a very small number of finetuning steps in FP32 (<1% of data).

46 **Max-atom Batching.** Due to the large differences in system topology and number of atoms/edges  
47 per system, using a fixed number of systems as batch size is infeasible. Instead we chose to use a  
48 max-atom batching scheme where we randomly pack batches that contain as close to an upper bound  
49 (max atoms) as possible without going over to guarantee an upper bound on memory usage.

50 **Max Neighbors.** For training efficiency purposes, we use a significantly smaller number of neigh-  
51 bors per atom during pretraining and found that it has no effect on the final performance, energy  
52 conservation, and smoothness properties of the model after finetuning with effectively “infinite” max  
53 neighbors.

### 54 A.2 Parallelisms

55 Although our models are designed with inference efficiency in mind, training models with a large  
56 number of MoLE experts is memory intensive. In particular, for the finetuning stages, a combination  
57 of infinite neighbors, FP32 precision, and autograd forces puts significant constraints on memory and  
58 training speed. We used a combination three parallelism training techniques summarized as follows:

- 59 • Graph parallelism (GP) [42]: Partitioning graphs across GPUs during message passing  
60 layers is used when scaling up to a large number of atoms at large model sizes. Graph  
61 partitioning is only used within a node with a fixed graph parallel rank size (2 or 4) during  
62 conservative fine-tuning stages.
- 63 • Fully-sharded data parallel (FSDP): We use the Pytorch FSDPv1 implementation on MoLE  
64 expert layers only for models with a total parameter count exceeding 1B during conservative  
65 fine-tuning when memory is scarce. Parameters are sharded within a node and replicated  
66 across nodes.
- 67 • Distributed Data Parallel (DDP): We use the standard PyTorch DDP implementation, with  
68 modifications for per-atom loss averaging and compatibility with graph parallelism.

69 Furthermore, we leveraged Pytorch’s Distributed Checkpointing framework to ensure saving and  
70 loading extremely large checkpoints is efficient and stable across different node configurations.  
71 Exponential moving average (EMA) is used for stable validation performance.

### 72 A.3 Model

73 **One model for all tasks.** UMA is designed for multi-task learning under diverse DFT settings. For  
74 two inputs with exactly the same atomic numbers and positions, the DFT labels will be different  
75 when different DFT settings are used. Such DFT settings include the level of theory and system  
76 total charge/spin. These task specifications are global information of the atomic system, and we  
77 process them through initial embedding layers. In this paper, five levels of theories are involved –

Table 1: Summary of main training-related hyper-parameters for the pre-training and fine-tuning stages. These hyper-parameters are shared among model sizes.

| Hyper-parameter                  | Pre-training       | Fine-tuning        |
|----------------------------------|--------------------|--------------------|
| precision                        | BF16               | FP32               |
| radius cutoff Å                  | 6                  | 6                  |
| max neighbors                    | 30                 | 300                |
| force prediction                 | Direct             | Autograd           |
| stress prediction                | None               | Autograd           |
| Optimizer                        | AdamW              | AdamW              |
| Learning rate scheduling         | Cosine             | Cosine             |
| Maximum learning rate            | $8 \times 10^{-4}$ | $4 \times 10^{-4}$ |
| Warmup epochs                    | 0.01               | 0.01               |
| Warmup factor                    | 0.2                | 0.2                |
| Gradient clipping norm threshold | 100                | 100                |
| Model EMA decay                  | 0.999              | 0.999              |
| Weight decay                     | $1 \times 10^{-3}$ | $1 \times 10^{-3}$ |
| Energy loss coefficient          | 10                 | 20                 |
| OMol energy loss coefficient     | 30                 | -                  |
| Force loss coefficient           | 30                 | 2                  |
| Stress loss coefficient          | -                  | 1                  |

Table 2: Hyper-parameters for UMA models of different sizes.

| Hyper-parameters                        | UMA-sm | UMA-md | UMA-lg |
|---|--------|--------|--------|
| Number of MoLE experts                  | 32     | 32     | Dense  |
| Number of layer blocks                  | 4      | 10     | 16     |
| Maximum degree $L_{\max}$               | 2      | 4      | 6      |
| Maximum order $M_{\max}$                | 2      | 2      | 2      |
| Number of channels $N_{\text{channel}}$ | 128    | 128    | 256    |
| Number of radial basis functions        | 64     | 128    | 256    |
| Global batch size (atoms)               | 88k    | 44k    | 44k    |
| Number of pre-training steps            | 1.68M  | 2.08M  | 2.58M  |
| Number of fine-tuning steps             | 1M     | 545k   | 350k   |

OMat24, OC20, OMol25, OMC25, and ODAC23 all use different DFT levels of theory. OMol25 further contains systems with non-neutral charge/spin.

Furthermore, to use a single model for all tasks, it is crucial to normalize the labels such that targets from different datasets fall into similar numerical ranges. We specifically design a referencing scheme that brings the diverse datasets to a similar level, detailed in Appendix B. The model hyperparameters are shown in Table 2.

#### A.4 MPTrj and sAlex Fine-tuning

For materials evaluations, UMA models are fine-tuned on the MPTrj [15] and sAlex [41, 5] datasets to ensure consistent DFT settings. The fine-tuning procedure is the same as eSEN-30M-OAM as documented in [16].

#### A.5 Training Resources

The resources used to train UMA models are described in Table 3.

## B Training Data

A summary of the five datasets used to train UMA is shown in Table 4. In total, the dataset has 459 million training examples, containing up to 350 atoms. The average number of atoms varies based on the dataset, from 19 for OMat24 to 178 for ODAC23.

Table 3: Training Times for UMA models.

| Model and Stage          | GPUs in Parallel | Training Days | GPU-Type   |
|--------------------------|------------------|---------------|------------|
| UMA sm direct pretrain   | 128              | 5             | H200 140GB |
| UMA sm conserve finetune | 256              | 5             | H200 140GB |
| UMA md direct pretrain   | 128              | 14            | H200 140GB |
| UMA md conserve finetune | 256              | 14            | H200 140GB |
| UMA lg direct pretrain   | 128              | 25            | H100 80GB  |
| UMA lg stress finetune   | 128              | 4             | H100 80GB  |
| UMA lg fp32 finetune     | 128              | 2             | H100 80GB  |

Table 4: Overview of the five datasets used to train UMA. For each dataset various statistics are provided alongside the sampling ratio used for training.

| Dataset      | Domain             | Training Size | Labels | # Elements | Avg Size | Force RMS | Sampling ratio |
|--------------|--------------------|---------------|--------|------------|----------|-----------|----------------|
| OMat24       | Materials          | 100,824,585   | E,F,S  | 89         | 19       | 2.83      | 4              |
| OMol25       | Molecules          | 75,889,983    | E,F    | 83         | 52       | 0.985     | 4              |
| OC20++       | Catalysis          | 229,054,043   | E,F    | 56         | 77       | 0.624     | 1              |
| OMC25        | Molecular Crystals | 24,870,226    | E,F,S  | 12         | 130      | 0.103     | 2              |
| ODAC23       | MOFs               | 28,517,826    | E,F    | 70         | 178      | 0.046     | 1              |
| <b>Total</b> |                    | 459,156,663   |        |            |          |           |                |

## 94 B.1 Materials

95 The field of inorganic bulk materials is moving at an incredibly fast pace. Here we train on the  
 96 Open Materials (OMat24) dataset (100M) [5], one of the largest and most diverse datasets in the  
 97 community. All DFT calculations from this domain were run with VASP [27, 26, 28, 29] and used the  
 98 PBE [37] functional. Due to the differences in pseudopotential version and different pseudopotentials  
 99 for certain elements in OMat24 and Materials Project [5] calculation settings used for the data in most  
 100 third party benchmarks, finetuning was also performed on MPtrj [15] and subsampled Alexandria  
 101 (sAlex) [41] to ensure consistent evaluation on the materials benchmarks.

## 102 B.2 Molecules

103 The community has seen dozens of molecular datasets spanning different scales for a variety of  
 104 applications []. However, the varying levels of DFT theory and quality makes it challenging to  
 105 unify under a single model. The release of the Open Molecules 2025 (OMol25) dataset [3] helps  
 106 address this by providing the largest single dataset (100M+) spanning 80+ elements covering metal-  
 107 complexes, biomolecules, electrolytes, and several existing datasets under a single, high-quality level  
 108 of theory. All DFT calculations were performed using Orca [34] ( $\omega$ B97M-V/def2-TZVPD). At the  
 109 time of training, only 75M samples from OMol25 were available for use, and we refer to this as  
 110 OMol-preview. Splits were constructed to ensure that this snapshot of the dataset is consistent with  
 111 the full dataset release. All ablations and results were trained with this OMol-preview, unless stated  
 112 otherwise. Released models, however, will be retrained with the full OMol25 dataset to ensure the  
 113 best models are accessible by the community.

## 114 B.3 Catalysis

115 The Open Catalyst (OC20) dataset [11] provides the largest adsorbate+surface dataset in the com-  
 116 munity. OC20 enumerates 1M+ unique surface + adsorbate combinations, spanning 55 elements,  
 117 and runs local geometry optimizations. Here, we train on the OC20 All (133M), MD (38M), and  
 118 Rattled (17M) datasets. Unlike prior work, we also leverage OC20’s clean surface data (14M) since  
 119 models here are trained on total energies. One limitation of OC20 is that it only contains single  
 120 adsorbates that interact with a surface. To address this, we introduce the OC20-Multi-Adsorbate  
 121 (mAds) dataset (22M) to better capture coverage effects and adsorbate-adsorbate interactions. All  
 122 DFT calculations were performed using VASP [27, 26, 28, 29] with the RPBE exchange-correlation  
 123 functional functional [19].

## 124 B.4 Molecular Crystals

125 The most recent 7th Crystal Structure Prediction (CSP) Blind Test organized by Cambridge Crystallo-  
126 graphic Data Center (CCDC) demonstrated the effectiveness of tailored machine learning interatomic  
127 potentials (MLIPs) in predicting, filtering, and ranking molecular crystal structures [22, 21]. However,  
128 despite the widespread applications of molecular crystals, there has been limited focus on developing  
129 universal MLIPs for molecular crystals, mostly because of the scarcity of publicly available datasets.  
130 Currently, publicly available datasets of molecular crystals are scarce, with at most 60,000 materials  
131 represented [1, 9]. To address this data gap, we use the Open Molecular Crystals (OMC25) dataset,  
132 which comprises 25 million molecular crystal structures. The dataset includes multiple relaxation  
133 trajectories of various molecular crystal packings generated with Genarris [45] starting from organic  
134 molecules from the OE62 dataset [44]. The dataset includes 12 elements (all elements from OE62,  
135 excluding Li, As, Se, and Te) and the maximum number of atoms is capped at 300. The dataset is  
136 computed using VASP [27, 26, 28, 29] with the PBE exchange-correlation functional [37] and D3  
137 dispersion correction [18]. The OMC25 dataset, along with in-depth details on data and polymorph  
138 structure generation, will be released in an upcoming publication [2].

## 139 B.5 Metal-organic frameworks (MOFs)

140 The Open Direct Air Capture (ODAC23) dataset represents the largest MOF dataset (28M) for DAC  
141 applications [43]. Derived from the CoREMOF [12, 13] dataset, ODAC studies the adsorption and  
142 co-adsorption of CO<sub>2</sub> and H<sub>2</sub>O in MOFs. This work uses an enhanced version of the ODAC23 dataset  
143 where the DFT calculated energies are upgraded to a higher k-points sampling grid density []. All  
144 DFT calculations were performed using VASP [27, 26, 28, 29] with the PBE exchange-correlation  
145 functional [37] and D3 dispersion correction [18].

## 146 B.6 Referencing

147 Although each dataset used in this work comes with its own very specific set of DFT settings, we  
148 wanted a mechanism to try and ensure the utility of our models’ predicted energies for even slightly  
149 different DFT settings researchers may be interested in. We do this through a "heat of formation"  
150 (HOF) reference that is applied to the energies:

$$E_{ref} = E_{DFT} - \sum_i^N [E_{i,DFT} - \Delta H_{f,i}]$$

151 Where  $E_{DFT}$  corresponds to the total DFT energy of the system,  $i$  is the atom number,  $N$  is the  
152 total number of atoms in the system,  $E_{i,DFT}$  is the DFT energy of an isolated atom  $i$  in a box,  
153 and  $\Delta H_{f,i}$  is the heat of formation of atomic number  $i$  as taken directly from Mendelev [33].  
154  $E_{i,DFT}$  was calculated using the DFT settings for each of the unique datasets in this work. In this  
155 referencing scheme, a researcher who may be interested in evaluating different DFT settings needs  
156 only to compute  $E_{i,DFT}$  for their level of theory to then apply to our models’ predicted energies.  
157 The fundamental underlying physics is, of course, limited to the DFT settings used in this work;  
158 this referencing merely provides a way for researchers to make energy magnitudes comparable to  
159 different DFT settings, a common problem we have seen in the community.

160 Additionally, we apply a linear reference to the above energies to help with the convergence and  
161 training stability of our models. We follow the same protocol as described in the OC22 Appendix  
162 [46].

## 163 C Scaling Laws Methods

164 The scaling law experiments were only performed on pretraining; due to compute constraints, we did  
165 not study the effect on finetuning with energy conservation. We used an 8-expert MoLE version of  
166 the model with the UMA-md settings ( $l_{max} = 4$ ,  $m_{max} = 2$ , and  $n_{neighbors} = 30$ ) for consistency.

### 167 C.1 FLOP counting

168 We approximate our training FLOPs by

$$C(N, D) \approx \kappa ND, \tag{1}$$

where  $N$  is the total number of parameters in the network, and  $D$  is the number of inputs (tokens for LLMs and atoms or edges for MLIPs) computed on. This relationship holds for any network that is dominated by linear layers where  $\kappa$  indicates how many times a parameter is re-used on an input. For a single forward pass of a linear network,  $\kappa = 2$ . A full training cycle for such a network with a single backward pass brings  $\kappa = 6$  as we need to compute the gradient with respect to both the parameters and inputs. Thus, for most LLMs  $\kappa = 6$  FLOPs/parameter/token [24] for a single forward and backwards pass where  $D$  is in units of tokens.

For our edge-based SO2 equivariant networks [36],  $\kappa$  is a function of  $l_{max}$  and  $m_{max}$  spherical harmonic orders and the number of edges per atom  $n_{neighbors}$ . For the scaling law experiments, we used  $l_{max} = 4$ ,  $m_{max} = 2$ , and  $n_{neighbors} = 30$  which corresponds to the settings of UMA-md model, resulting in  $\kappa \approx 270$  FLOPs/parameter/atom (or 9 FLOPs/parameter/edge) per training step, which can be computed or experimentally determined. As long as the number of input edges and parameters are sufficiently large (which holds for UMA models), this flop approximation holds as contributions from all other operators are marginal. We also verified this assumption with direct FLOP counting in our model code.

Overall, parameter reuse is significantly higher in Equivariant-GNNs compared to LLMs, and hence the flop count is 2-3 orders of magnitude higher for a similar parameter-sized LLM network.

## C.2 Compute optimal fits

The compute optimal model and dataset sizes can then be fitted to power laws [24, 20]:

$$\begin{aligned}\log(N^*(C)) &= \alpha \log(C) + A \\ \log(D^*(C)) &= \beta \log(C) + B\end{aligned}\tag{2}$$

Where  $C$  is the compute in FLOPs described in Appendix C.1.  $N^*(C)$  represents the optimal model size (in parameters) as a function of compute, and  $D^*(C)$  represents the optimal dataset size (in units of atoms).  $N^*(C)$  is determined by finding the minima of fitted parabolas for each Isoflop curve. The 10% and 90% percentile bootstrap errors are shown in Figure 3(e).  $\alpha$  and  $\beta$  are the scaling coefficients w.r.t. model size and dataset size respectively.  $A$  and  $B$  are offset constants of the fit. Fit coefficients and bootstrap errors are shown in Table 5.

## C.3 Fitting loss vs. parameters

To understand the minimum achievable loss for dense vs MoLE models we can fit the more general parameterized ansatz of  $L(N, D)$  proposed by [24].

$$\tilde{L}(N, D) = \hat{E} + \frac{\hat{A}}{N^{\hat{\alpha}}} + \frac{\hat{B}}{D^{\hat{\beta}}}\tag{3}$$

This maps the power law coefficients from Equation 3 to those in Equation 2 by using  $\alpha = \frac{\hat{\beta}}{\hat{\alpha} + \hat{\beta}}$  and  $\beta = \frac{\hat{\alpha}}{\hat{\alpha} + \hat{\beta}}$ . Here we can either minimize the  $\tilde{L}(N, D)$  directly by fitting the 5 parameters  $\hat{E}, \hat{A}, \hat{B}, \hat{\alpha}, \hat{\beta}$  with a iterative minimization procedure such as LBFGS [20, 10] or by examining the loss as a power relationship of  $N^*$  using

$$\log \tilde{L}(N^*) = \hat{\alpha} \log(N^*) + \gamma\tag{4}$$

where  $\gamma = \log([1 + \frac{\hat{\alpha}}{\hat{\beta}}]\hat{A})$  with  $\hat{E} \approx 0$ . We found both methods yielded similar results but the minimization of  $\tilde{L}$  was more sensitive to the choice of hyperparameters.

Table 5: Power Law Coefficients determined from fitting Equations 2 and 4. Error bounds are determined by bootstrap sampling 1000 times and taking the 10th and 90th percentile values, quoted in brackets.

| Parameter      | Dense                | MoLE                |
|----------------|----------------------|---------------------|
| $\alpha$       | 0.61 (0.57,0.65)     | 0.56 (0.49,0.59)    |
| $\beta$        | 0.39 (0.35, 0.43)    | 0.44 (0.39, 0.43)   |
| $A$            | -4.5 (-3.8, -5.3)    | -3.8 (-2.56, -4.65) |
| $B$            | 3.6 (2.9, 4.4)       | 2.9 (1.6, 3.7)      |
| $\hat{\alpha}$ | -0.29 (-0.27, -0.31) | -0.25 (-0.2, -0.3)  |
| $\gamma$       | 2.16 (2.02, 2.34)    | 1.82 (1.61, 2.12)   |

## D Inference

For inference benchmarking we use a periodic fcc carbon system with lattice constant  $a = 3.8\text{\AA}$ . This results in a fixed density of approximately 50 edges per atom within  $6\text{\AA}$ . For UMA models, we use a combination of torch.compile, cuda graphs and pre-merged MoLE experts for inference speed. For large number of atoms ( $> 1000$ ), we use edge-based activation checkpointing to trade off memory for some speed, allowing us to fit 100k+ atoms for the UMA-sm into memory. We checked all our optimizations chosen does not degrade simulation accuracy, equivariance or energy conservation properties. While our benchmarks do not include graph generation, our internal CUDA based graph generation algorithm is very fast and decreases throughput by no more than 10% even for the largest systems tested. In the case of non-MoLE merging, we found the inference speed was comparable but the parameters require more GPU memory to store.

For fair comparisons against other models, we used pytorch2.6.0, cuda12.4, python3.12 and TF-32 precision universally on a H100 80GB GPU. We use standard torch.compile settings whenever possible (only MACE-MPA-0 failed to compile). Different models have different radius cutoffs and max neighbors settings. We made sure that all models was receiving roughly 50 neighbors per atom for the same number of atoms.

Table 6: Single-GPU simulation speeds for energy-conservative models in *steps per second*: comparing UMA models to the top two models (eSEN and OrbV3) on the Matbench Discovery leaderboard [40] and the MACE models. We exclude UMA-lg being non-conservative. Benchmarks are run, excluding graph generation, on a single Nvidia H100 80GB GPU using FP32 (TF32-high precision) and torch compile when possible. Test systems are a standard periodic atomic system that have  $\approx 50$  neighbors per atom with a  $6\text{\AA}$  cutoff. OOM indicates the model ran out of memory.

| Atoms   | UMA-sm<br>(6.6M) | UMA-md<br>(50M) | eSEN-30M-<br>OAM<br>(30M) | Orb-v3<br>conservative-<br>inf-omat<br>(25M) | MACE-<br>MPA-0<br>(9M) | MACE-<br>OFF23-L<br>(4.7M) |
|---------|------------------|-----------------|---------------------------|--|------------------------|----------------------------|
| 100     | 50               | 21              | 8                         | 77   | 38                     | 89                         |
| 1,000   | 16               | 3               | 1.7                       | 30   | 24                     | 20                         |
| 10,000  | 1.6              | 0.2             | OOM                       | 3.7  | 2.9                    | OOM                        |
| 50,000  | 0.2              | OOM             | OOM                       | OOM  | OOM                    | OOM                        |
| 100,000 | 0.1              | OOM             | OOM                       | OOM  | OOM                    | OOM                        |

## E Evaluations

### E.1 UMA-sm-v1 Results

In this section, we provide results on the UMA-sm-v1 model trained with the entire OMol25 dataset, instead of the preview OMol25 dataset used by UMA-sm in the main body of this paper. The other datasets used for training were not changed. Tables 7 and 8 correspond to Tables 2 and 3 in the main text.

Table 7: Test MAE results on held out test splits for materials [40], catalysis [11], molecules [3], molecular crystals [2] and ODAC [43]. All energies are in meV, forces are in meV/Å and stresses are in meV/Å<sup>3</sup>. Results for UMA are compared against the SOTA literature results when available and other strong baselines trained only on the domain-specific dataset. Target accuracies for practical utility are provided as an approximate guide for reference.

| Model               | Materials          |        |        |                    |        |        | Catalysis         |        |                         |        | Molecules               |        |                       |        | Molecular crystals      |        |        | ODAC                  |        |
|---------------------|--------------------|--------|--------|--------------------|--------|--------|-------------------|--------|-------------------------|--------|-------------------------|--------|-----------------------|--------|-------------------------|--------|--------|-----------------------|--------|
|                     | WBM<br>Energy/Atom | Forces | Stress | HEA<br>Energy/Atom | Forces | Stress | ID<br>Ads. Energy | Forces | OOD-Both<br>Ads. Energy | Forces | OOD-Comp<br>Energy/Atom | Forces | PDB-TM<br>Energy/Atom | Forces | OMC-Test<br>Energy/Atom | Forces | Stress | OOD-LT<br>Ads. Energy | Forces |
| <b>UMA</b>          |                    |        |        |                    |        |        |                   |        |                         |        |                         |        |                       |        |                         |        |        |                       |        |
| UMA-sm              | 20.0               | 60.8   | 4.4    | 22.0               | 72.8   | 3.1    | 52.1              | 24.3   | 70.2                    | 30.9   | 3.64                    | 10.80  | 0.88                  | 16.12  | 0.91                    | 4.77   | 0.97   | 292.4                 | 16.0   |
| UMA-sm-v1           | 19.4               | 62.2   | 4.5    | 21.9               | 73.5   | 3.1    | 53.1              | 24.5   | 70.4                    | 31.2   | 0.96                    | 8.25   | 0.93                  | 15.56  | 0.93                    | 5.15   | 1.01   | 287.1                 | 13.6   |
| <b>Literature</b>   |                    |        |        |                    |        |        |                   |        |                         |        |                         |        |                       |        |                         |        |        |                       |        |
| eSEN-OMat [16]      | 16.2               | 49.6   | 4.1    | 20.0               | 59.5   | 3.2    | -                 | -      | -                       | -      | -                       | -      | -                     | -      | -                       | -      | -      | -                     | -      |
| eqV2-OMat [5]       | 14.9               | 46.3   | 3.6    | 20.3               | 47.0   | 2.7    | -                 | -      | -                       | -      | -                       | -      | -                     | -      | -                       | -      | -      | -                     | -      |
| eqV2-OC20 [31]      | -                  | -      | -      | -                  | -      | -      | 149.1             | 11.6   | 306.5                   | 15.7   | -                       | -      | -                     | -      | -                       | -      | -      | -                     | -      |
| GemNet-OC20 [17]    | -                  | -      | -      | -                  | -      | -      | 163.5             | 16.3   | 343.3                   | 23.1   | -                       | -      | -                     | -      | -                       | -      | -      | -                     | -      |
| eSEN-sm-cons. [3]   | -                  | -      | -      | -                  | -      | -      | -                 | -      | -                       | -      | 1.35                    | 7.39   | 0.83                  | 12.72  | -                       | -      | -      | -                     | -      |
| eqv2-ODAC [43]      | -                  | -      | -      | -                  | -      | -      | -                 | -      | -                       | -      | -                       | -      | -                     | -      | -                       | -      | -      | 316.0                 | 7.2    |
| <b>ST Baselines</b> |                    |        |        |                    |        |        |                   |        |                         |        |                         |        |                       |        |                         |        |        |                       |        |
| UMA-sm-OMC          | -                  | -      | -      | -                  | -      | -      | -                 | -      | -                       | -      | -                       | -      | -                     | -      | 1.05                    | 5.39   | 0.94   | -                     | -      |
| <b>Target</b>       |                    |        |        |                    |        |        |                   |        |                         |        |                         |        |                       |        |                         |        |        |                       |        |
| Practical Utility   | 10-20              | -      | -      | 10-20              | -      | -      | 100               | -      | 100                     | -      | 1-3                     | -      | 1-3                   | -      | 1-3                     | -      | -      | 100                   | -      |

Table 8: Evaluation results on Matbench-Discovery [40], MDR phonon [32], elastic tensor [14, 23], and AdsorbML benchmarks [30]. Results are also provided for a diverse set of molecule [3] and molecular crystal [21, 2] benchmarks. NVE MD [16] tests whether energy conservation is observed when running the model for molecular dynamics. SOTA results from literature are reported where available. Note the UMA-sm-OMol and UMA-sm-OMC models are only trained on the preview OMol25 and OMC25 datasets respectively. Additionally, for the materials evaluations, UMA models were fine-tuned on MPtrj [15] and sAlex [41, 5] to be consistent with the benchmark’s DFT settings.

| Model                 | Materials           |       |               |                             |   |                         |   |                        |                         | Catalysis                     | Molecules                            |                     |               |               |                         | Molecular Crystals                             |              |          |
|-----------------------|---------------------|-------|---------------|-----------------------------|---|-------------------------|---|------------------------|-------------------------|-------------------------------|--------------------------------------|---------------------|---------------|---------------|-------------------------|--|--------------|----------|
|                       | Matbench [40]<br>F1 | RMSD  | MAE [eV/atom] | $\kappa_{\text{syme}}$ [38] | Phonons [32]<br>$\omega_{\text{max}}$ [K] | Free Energy<br>[kJ/mol] | Elasticity [14, 23]<br>$G_{\text{vib}}$ [GPa] | $K_{\text{vib}}$ [GPa] | NVE MD [16]<br>Conserve | AdsorbML [30]<br>Success Rate | OMol25 [3]<br>Ligand-strain<br>[meV] | PDB-pocket<br>[meV] | Dist-SR [meV] | Dist-LR [meV] | NVE MD [16]<br>Conserve | CSP Targets [21]<br>Lattice Energy<br>[kJ/mol] | Kendall Rank | RMSD [Å] |
| <b>UMA</b>            |                     |       |               |                             |   |                         |   |                        |                         |                               |                                      |                     |               |               |                         |  |              |          |
| UMA-sm                | 0.916               | 0.064 | 0.020         | 0.203                       | 17.59                                     | 5.0                     | 8.54  | 4.96                   | ✓                       | 68.35%                        | 4.39                                 | 150.3               | 67.6          | 432.1         | ✓                       | 2.70   | 0.82         | 0.12     |
| UMA-sm-v1             | 0.914               | 0.064 | 0.020         | 0.231                       | 18.65                                     | 5.51                    | 13.72   | 5.19                   | ✓                       | 64.85%                        | 5.19                                 | 138.3               | 23.8          | 87.1          | ✓                       | 2.57   | 0.81         | 0.13     |
| <b>Literature</b>     |                     |       |               |                             |   |                         |   |                        |                         |                               |                                      |                     |               |               |                         |  |              |          |
| eSEN-30M-OAM [16]     | 0.925               | 0.061 | 0.018         | 0.170                       | 15.00                                     | 4.00                    | 9.13  | 5.73                   | ✓                       | -                             | -                                    | -                   | -             | -             | -                       | -  | -            | -        |
| ORB v3 [39]           | 0.905               | 0.075 | 0.024         | 0.210                       | -   | -                       | -   | -                      | -                       | -                             | -                                    | -                   | -             | -             | -                       | -  | -            | -        |
| SevenNet-MF-ompa [25] | 0.901               | 0.064 | 0.021         | 0.317                       | -   | -                       | -   | -                      | -                       | -                             | -                                    | -                   | -             | -             | -                       | -  | -            | -        |
| GRACE-2L-OAM [8]      | 0.880               | 0.067 | 0.023         | 0.294                       | -   | -                       | -   | -                      | -                       | -                             | -                                    | -                   | -             | -             | -                       | -  | -            | -        |
| MACE-MPA-0 [6]        | 0.852               | 0.073 | 0.028         | 0.412                       | -   | -                       | -   | -                      | -                       | -                             | -                                    | -                   | -             | -             | -                       | -  | -            | -        |
| eqv2-OC20 [30]        | -                   | -     | -             | -                           | -   | -                       | -   | -                      | -                       | 60.80%                        | -                                    | -                   | -             | -             | -                       | -  | -            | -        |
| GemNet-OC20 [30]      | -                   | -     | -             | -                           | -   | -                       | -   | -                      | -                       | 54.88%                        | -                                    | -                   | -             | -             | -                       | -  | -            | -        |
| eSEN-sm-cons. [3]     | -                   | -     | -             | -                           | -   | -                       | -   | -                      | -                       | -                             | 4.52                                 | 147.3               | 28.6          | 268.6         | ✓                       | -  | -            | -        |
| <b>ST Baselines</b>   |                     |       |               |                             |   |                         |   |                        |                         |                               |                                      |                     |               |               |                         |  |              |          |
| UMA-sm-OMC            | -                   | -     | -             | -                           | -   | -                       | -   | -                      | -                       | -                             | -                                    | -                   | -             | -             | -                       | 6.18   | 0.74         | 0.18     |

## 225 E.2 Materials

226 Full results on materials’ benchmarks are in Tables 9 and 10. Table 9 shows both validation and  
 227 test results following OMat24 [5] along with the new high entropy alloy (HEA) test set introduced  
 228 in this paper. The HEA dataset contains relaxation trajectories for over 5000 alloys with atomic  
 229 configuration disorder. Input structures were generated by sampling metallic element combinations  
 230 of up to 6 different unique elements and using the special quasirandom structure method [47, 4]  
 231 to decorate face-centered cubic, body-centered cubic and hexagonal close packed structures. DFT  
 232 relaxations were carried out following the settings used in the OMat24 dataset [5].

233 In Table 10 we show full results for Matbench discovery [40], MDR phonon, elastic tensors, high  
 234 entropy alloy IS2RE and NVE MD conservation benchmarks. The Matbench-Discovery benchmark



Table 9: Materials validation and test evaluations from OMat24 [5] and HEA. All energies are in meV, forces are in meV/Å and stresses are in meV/Å<sup>3</sup>.

| Model              | Val [5]     |        |        |              | Test        |        |        |                     |        |        |                 |        |        |             |        |        |
|--------------------|-------------|--------|--------|--------------|-------------|--------|--------|---------------------|--------|--------|-----------------|--------|--------|-------------|--------|--------|
|                    |             |        |        |              | WBM [5]     |        |        | OOD Composition [5] |        |        | OOD Element [5] |        |        | HEA         |        |        |
|                    | Energy/Atom | Forces | Stress | Force Cosine | Energy/Atom | Forces | Stress | Energy/Atom         | Forces | Stress | Energy/Atom     | Forces | Stress | Energy/Atom | Forces | Stress |
|                    |             |        |        |              |             |        |        |                     |        |        |                 |        |        |             |        |        |
| UMA                |             |        |        |              |             |        |        |                     |        |        |                 |        |        |             |        |        |
| UMA-sm             | 11.3        | 57.1   | 2.9    | 0.98         | 20.0        | 60.8   | 4.4    | 11.5                | 57.0   | 3.0    | 9.9             | 56.9   | 2.6    | 22.0        | 72.8   | 3.1    |
| UMA-md             | 10.0        | 47.3   | 2.7    | 0.99         | 18.1        | 51.4   | 4.3    | 10.2                | 47.6   | 2.9    | 8.5             | 47.1   | 2.4    | 19.0        | 62.2   | 3.2    |
| UMA-lg             | 9.7         | 43.5   | 2.5    | 0.99         | 17.6        | 45.5   | 3.8    | 9.8                 | 43.6   | 2.6    | 8.1             | 43.6   | 2.3    | 24.8        | 48.3   | 2.8    |
| Literature         |             |        |        |              |             |        |        |                     |        |        |                 |        |        |             |        |        |
| eSEN-30M-OMat [16] | 10.7        | 47.3   | 2.6    | 0.99         | 16.2        | 49.6   | 4.1    | 10.7                | 47.3   | 2.8    | 9.0             | 47.2   | 2.3    | 20.0        | 59.5   | 3.2    |
| eqV2-86M-OMat [31] | 10.0        | 44.9   | 2.4    | 0.99         | 14.9        | 46.3   | 3.6    | 10.0                | 44.5   | 2.5    | 8.8             | 44.7   | 2.1    | 20.3        | 47.0   | 2.7    |

Table 10: **Materials evals results.** We note that UMA models are fine-tuned on the MPTrj [15] and sAlex [41, 5] datasets to be consistent with the DFT settings of the benchmarks.

| Model             | Matbench [40] |      |           |          | Kappa 103 [38] |                |          |                       | MDR Phonons [32]       |                                 |                                 |                        | Elasticity [14, 23]   |                           | Binary Elasticity            |                              | NVE MD [16]  |
|-------------------|---------------|------|-----------|----------|----------------|----------------|----------|-----------------------|------------------------|---------------------------------|---------------------------------|------------------------|-----------------------|---------------------------|------------------------------|------------------------------|--------------|
|                   | F1            | DAF  | Precision | Accuracy | MAE [eV/atom]  | RMSE [eV/atom] | $\rho^2$ | $\kappa_{\text{ste}}$ | $\kappa_{\text{srne}}$ | $\omega_{\text{max}}$ , MAE [K] | $\omega_{\text{avg}}$ , MAE [K] | Entropy, MAE [J/K/mol] | $C_v$ , MAE [J/K/mol] | Free Energy, MAE [kJ/mol] | $G_{\text{vib}}$ , MAE [GPa] | $K_{\text{vib}}$ , MAE [GPa] | Conservative |
| UMA               |               |      |           |          |                |                |          |                       |                        |                                 |                                 |                        |                       |                           |                              |                              |              |
| UMA-sm            | 0.92          | 6.00 | 0.92      | 0.97     | 0.02           | 0.07           | 0.87     | 0.09                  | 0.20                   | 17.59                           | 7.41                            | 13.59                  | 3.49                  | 5.00                      | 8.54                         | 4.96                         | ✓            |
| UMA-md            | 0.93          | 6.08 | 0.93      | 0.98     | 0.02           | 0.07           | 0.87     | 0.09                  | 0.20                   | 13.91                           | 5.11                            | 9.63                   | 2.66                  | 3.39                      | 8.40                         | 4.76                         | ✓            |
| UMA-lg            | 0.93          | 6.09 | 0.93      | 0.98     | 0.02           | 0.07           | 0.86     | 0.45                  | 0.67                   | 78.50                           | 27.68                           | 43.04                  | 15.85                 | 18.20                     | 20.56                        | 14.48                        | ✗            |
| Literature        |               |      |           |          |                |                |          |                       |                        |                                 |                                 |                        |                       |                           |                              |                              |              |
| eSEN-30M-OAM [16] | 0.93          | 6.07 | 0.93      | 0.98     | 0.02           | 0.07           | 0.87     | -                     | 0.17                   | 15.00                           | 10.21                           | 10.00                  | 3.00                  | 4.00                      | 9.13                         | 5.73                         | ✓            |
| eqV2-86M-OAM [5]  | 0.92          | 6.05 | 0.92      | 0.98     | 0.02           | 0.07           | 0.85     | 1.82                  | 1.94                   | 840.33                          | 377.96                          | 426.79                 | 102.72                | 251.14                    | 19.60                        | 26.25                        | ✗            |

evaluates a model’s ability to predict ground-state thermodynamic stability by optimizing geometry and predicting energy. The thermal conductivity prediction task demands accurate modeling of harmonic and anharmonic phonons, which are crucial for precise predictions of thermal transport. The MDR Phonon benchmark assesses a model’s performance in predicting phonon and vibrational thermodynamic properties. The MP elastic constant benchmark tests a model’s accuracy in predicting bulk and shear moduli, requiring precise calculations of stress tensors and their derivatives with respect to cell deformations.

### E.3 Catalysis

For catalysis, we show full validation and test results for OC20 [11] in Table 11. The structures in the dataset contain molecules, called adsorbates, interacting with surfaces. The goal is to estimate the adsorption energy, which is the change in energy as the adsorbates come into contact with the surface, and the forces on the atoms. Force MAEs are comparable across UMA-md and UMA-lg to other state-of-the-art models. Adsorption energies for UMA are calculated by subtracting two total energy calculations as described in the main text, which improves results over prior models.

### E.4 Molecules

We report results on molecules following the Open Molecules 2025 [3]. These include energy and force estimates for validation and test splits, as well as, numerous other tasks. The validation and test results are shown in Tables 12 and 13, respectively. Note that the UMA-sm-OMol model is only trained on the preview OMol25 dataset, which is  $\approx 70\%$  of the full OMol25 dataset for a fair comparison with the UMA models that were trained on the same subset. In general, the UMA-sm and UMA-sm-OMol models provide comparable results.

OMol25 [3] provides numerous other tasks for evaluation. These include evaluations which only require the estimation of a single point DFT calculation and evaluations that require optimizations. The comparison for single point DFT calculations is shown in Table 14. Similar to the test results, the

Table 11: Catalysis validation and test results on OC20 [11] metrics. All energies are in meV and forces are in meV/Å.

| Model             | Val (Total Energy) |        |              |          |        |              | Test (Ads. Energy) |       |          |       |        |       |
|-------------------|--------------------|--------|--------------|----------|--------|--------------|--------------------|-------|----------|-------|--------|-------|
|                   | ID                 |        |              | OOD-Both |        |              | ID                 |       | OOD-Both |       |        |       |
|                   | Energy             | Forces | Force Cosine | Energy   | Forces | Force Cosine | Energy             | Force | Energy   | Force | Energy | Force |
| <b>UMA</b>        |                    |        |              |          |        |              |                    |       |          |       |        |       |
| UMA-sm            | 63.6               | 24.1   | 0.63         | 107.0    | 29.2   | 0.65         | 52.1               | 24.3  | 70.2     | 30.9  |        |       |
| UMA-md            | 43.1               | 15.8   | 0.73         | 70.0     | 19.2   | 0.75         | 33.4               | 16.0  | 46.5     | 21.0  |        |       |
| UMA-lg            | 32.6               | 12.0   | 0.77         | 49.8     | 14.5   | 0.79         | 32.4               | 12.2  | 43.5     | 15.9  |        |       |
| <b>Literature</b> |                    |        |              |          |        |              |                    |       |          |       |        |       |
| eqV2-OC20 [31]    | -                  | -      | -            | -        | -      | -            | 149.1              | 11.63 | 306.5    | 15.74 |        |       |
| GemNet-OC20 [17]  | -                  | -      | -            | -        | -      | -            | 163.5              | 16.33 | 343.3    | 23.11 |        |       |

Table 12: Open Molecule 2025 [3] validation evaluations across biomolecules, electrolytes, metal complexes, neutral organics and OOD-comp. All energies are in meV and forces are in meV/Å. All models are trained with preview OMol25.

| Model           | Biomolecules |       | Electrolytes |       | Metal Complexes |       | Neutral Organics |       | OOD-Comp    |       |
|-----------------|--------------|-------|--------------|-------|-----------------|-------|------------------|-------|-------------|-------|
|                 | Energy/Atom  | Force | Energy/Atom  | Force | Energy/Atom     | Force | Energy/Atom      | Force | Energy/Atom | Force |
| <b>UMA</b>      |              |       |              |       |                 |       |                  |       |             |       |
| UMA-sm          | 0.53         | 5.69  | 2.69         | 11.65 | 4.63            | 37.85 | 1.00             | 13.15 | 3.62        | 12.02 |
| UMA-md          | 0.44         | 3.95  | 2.28         | 10.21 | 3.60            | 28.81 | 0.68             | 7.00  | 3.21        | 9.90  |
| UMA-lg          | 0.33         | 2.90  | 1.13         | 4.52  | 3.35            | 24.85 | 0.65             | 5.02  | 2.39        | 5.83  |
| <b>Baseline</b> |              |       |              |       |                 |       |                  |       |             |       |
| UMA-sm-OMol     | 0.54         | 6.06  | 2.52         | 12.63 | 4.27            | 37.30 | 0.84             | 13.00 | 3.69        | 12.78 |

Table 13: Open Molecule 2025 [3] test evaluations across biomolecules, electrolytes, metal complexes, neutral organics and OOD-comp, metal ligand, PDB-TM, reactivity, COD and anions. All energies are in meV and forces are in meV/Å. All models are trained with preview OMol25.

| Model           | Biomolecules |       | Electrolytes |       | Metal Complexes |       | Neutral Organics |       | OOD-Comp    |       | Metal Ligand |       | PDB-TM      |       | Reactivity  |       | COD         |       | Anions      |       |
|-----------------|--------------|-------|--------------|-------|-----------------|-------|------------------|-------|-------------|-------|--------------|-------|-------------|-------|-------------|-------|-------------|-------|-------------|-------|
|                 | Energy/Atom  | Force | Energy/Atom  | Force | Energy/Atom     | Force | Energy/Atom      | Force | Energy/Atom | Force | Energy/Atom  | Force | Energy/Atom | Force | Energy/Atom | Force | Energy/Atom | Force | Energy/Atom | Force |
| <b>UMA</b>      |              |       |              |       |                 |       |                  |       |             |       |              |       |             |       |             |       |             |       |             |       |
| UMA-sm          | 0.51         | 5.70  | 3.80         | 13.95 | 3.07            | 33.56 | 1.49             | 20.33 | 3.64        | 10.80 | 1.23         | 17.71 | 0.88        | 16.12 | 4.82        | 61.80 | 2.92        | 29.82 | 0.66        | 10.85 |
| UMA-md          | 0.42         | 3.89  | 3.29         | 13.19 | 2.40            | 24.85 | 0.98             | 10.83 | 3.26        | 9.09  | 0.99         | 12.04 | 0.69        | 10.37 | 3.88        | 47.75 | 2.19        | 20.11 | 0.50        | 8.03  |
| UMA-lg          | 0.34         | 2.92  | 1.41         | 5.42  | 2.47            | 21.67 | 1.03             | 6.91  | 2.33        | 5.19  | 1.05         | 10.00 | 0.81        | 8.76  | 3.93        | 41.97 | 2.57        | 15.44 | 0.62        | 5.90  |
| <b>Baseline</b> |              |       |              |       |                 |       |                  |       |             |       |              |       |             |       |             |       |             |       |             |       |
| UMA-sm-OMol     | 0.52         | 6.06  | 3.52         | 15.23 | 2.86            | 33.30 | 1.35             | 20.06 | 3.67        | 11.56 | 1.18         | 17.49 | 0.79        | 14.11 | 4.89        | 61.16 | 2.93        | 24.12 | 0.47        | 10.38 |

Table 14: Open Molecule 2025 [3] single point evaluations for protein-ligand, IE/EA, spin gap and distance scaling. All energies are in meV and forces are in meV/Å. All models are trained with preview OMol25.

| Model       | Protein-ligand    |                   | IE/EA                  |                        | Spin gap                       |                        |                        | Distance scaling               |                             |                             |                             |                             |
|-------------|-------------------|-------------------|------------------------|------------------------|--------------------------------|------------------------|------------------------|--------------------------------|-----------------------------|-----------------------------|-----------------------------|-----------------------------|
|             | Idx Energy<br>MAE | Idx Forces<br>MAE | $\Delta$ Energy<br>MAE | $\Delta$ Forces<br>MAE | $\Delta$ Forces<br>cosine sim. | $\Delta$ Energy<br>MAE | $\Delta$ Forces<br>MAE | $\Delta$ Forces<br>cosine sim. | $\Delta$ Energy (SR)<br>MAE | $\Delta$ Forces (SR)<br>MAE | $\Delta$ Energy (LR)<br>MAE | $\Delta$ Forces (LR)<br>MAE |
| UMA         |                   |                   |                        |                        |                                |                        |                        |                                |                             |                             |                             |                             |
| UMA-sm      | 150.25            | 5.09              | 336.16                 | 80.60                  | 0.77                           | 665.75                 | 112.09                 | 0.69                           | 67.60                       | 4.11                        | 432.14                      | 5.54                        |
| UMA-md      | 89.69             | 4.06              | 236.76                 | 66.28                  | 0.81                           | 547.73                 | 98.15                  | 0.74                           | 41.60                       | 3.86                        | 588.74                      | 8.73                        |
| UMA-lg      | 71.68             | 2.27              | 244.23                 | 57.18                  | 0.82                           | 568.36                 | 93.09                  | 0.73                           | 16.55                       | 2.03                        | 246.10                      | 2.34                        |
| Baselines   |                   |                   |                        |                        |                                |                        |                        |                                |                             |                             |                             |                             |
| UMA-sm-OMol | 154.48            | 5.59              | 310.19                 | 77.48                  | 0.77                           | 634.02                 | 110.28                 | 0.70                           | 73.02                       | 4.16                        | 608.91                      | 7.69                        |

Table 15: Open Molecule 2025 [3] optimization evaluations including ligand-strain, conformer prediction, and protonation states. Results are reported across a variety of energy and structure based metrics for each task. All energies are in meV. All models are trained with preview OMol25.

| Model            | Ligand strain               |                   |                       | Conformers          |                        |                     |                           | Protonation           |                        |                     |                           |
|------------------|-----------------------------|-------------------|-----------------------|---------------------|------------------------|---------------------|---------------------------|-----------------------|------------------------|---------------------|---------------------------|
|                  | Strain energy<br>MAE [meV]↓ | RMSD<br>min. [Å]↓ | RMSD<br>ensemble [Å]↓ | RMSD<br>boltz. [Å]↓ | Δ Energy<br>MAE [meV]↓ | RMSD<br>reopt. [Å]↓ | Δ Energy<br>reopt. [meV]↓ | RMSD<br>ensemble [Å]↓ | Δ Energy<br>MAE [meV]↓ | RMSD<br>reopt. [Å]↓ | Δ Energy<br>reopt. [meV]↓ |
| <b>UMA</b>       |                             |                   |                       |                     |                        |                     |                           |                       |                        |                     |                           |
| UMA-sm           | 4.39                        | 0.28              | 0.06                  | 0.05                | 5.76                   | 0.02                | 3.06                      | 0.13                  | 40.42                  | 0.04                | 18.35                     |
| UMA-md           | 2.45                        | 0.19              | 0.04                  | 0.03                | 3.19                   | 0.01                | 1.47                      | 0.08                  | 24.08                  | 0.02                | 10.99                     |
| UMA-lg           | 3.37                        | 0.25              | 0.05                  | 0.05                | 4.97                   | 0.01                | 2.27                      | 0.11                  | 30.31                  | 0.02                | 12.54                     |
| <b>Baselines</b> |                             |                   |                       |                     |                        |                     |                           |                       |                        |                     |                           |
| UMA-sm-OMol      | 5.15                        | 0.31              | 0.06                  | 0.05                | 5.25                   | 0.03                | 3.24                      | 0.13                  | 32.82                  | 0.04                | 18.65                     |

UMA-sm and UMA-sm-OMol models perform similarly with the UMA-md and UMA-lg performing significantly better. Table 15 shows the results for tasks that require optimizations, which require repeated calls to the model to update atoms positions until a local energy minima is found. The small models report similar performance, and the UMA-md demonstrates the highest accuracies. It is likely that UMA-md outperforms UMA-lg due to UMA-md being energy conserving and better behaved during optimization tasks.

## E.5 Molecular Crystals

Table 16: Open Molecular Crystals 2025 [2] validation and test table. All energies are in meV, forces are in meV/Å and stresses are in meV/Å<sup>3</sup>.

| Model            | Val         |        |        |              | Test        |        |        |              |
|------------------|-------------|--------|--------|--------------|-------------|--------|--------|--------------|
|                  | Energy/Atom | Forces | Stress | Force Cosine | Energy/Atom | Forces | Stress | Force Cosine |
| <b>UMA</b>       |             |        |        |              |             |        |        |              |
| UMA-sm           | 0.9         | 4.9    | 1.0    | 0.92         | 0.9         | 4.8    | 1.0    | 0.93         |
| UMA-md           | 0.8         | 3.1    | 1.0    | 0.95         | 0.8         | 3.0    | 1.0    | 0.95         |
| UMA-lg           | 0.6         | 2.3    | 0.1    | 0.96         | 0.6         | 2.3    | 0.1    | 0.96         |
| <b>Baselines</b> |             |        |        |              |             |        |        |              |
| UMA-sm-OMC       | 1.06        | 5.58   | 0.96   | 0.92         | 1.05        | 5.39   | 0.94   | 0.92         |

Table 17: Open Molecular Crystals 2025 [2] evaluation for polymorphs from the Structure Ranking Phase of CCDC 7th CSP Blind Test [21]. All lattice energies are per molecule basis in kJ/mol.

| Model            | CCDC 7th CSP Blind Test Polymorphs |               |                |                             |          |                        |             |            |
|------------------|------------------------------------|---------------|----------------|-----------------------------|----------|------------------------|-------------|------------|
|                  | Lattice Energy<br>MAE [kJ/mol]     | RMSE [kJ/mol] | r <sup>2</sup> | Rank correlation<br>Kendall | Spearman | Structures<br>RMSE [Å] | RMSE sd [Å] | Match rate |
| <b>UMA</b>       |                                    |               |                |                             |          |                        |             |            |
| UMA-sm           | 2.69                               | 3.67          | 0.73           | 0.82                        | 0.93     | 0.12                   | 0.07        | 0.99       |
| UMA-md           | 2.66                               | 3.71          | 0.60           | 0.81                        | 0.91     | 0.13                   | 0.07        | 0.99       |
| UMA-lg           | 2.49                               | 3.70          | 0.81           | 0.84                        | 0.95     | 0.12                   | 0.07        | 1.00       |
| <b>Baselines</b> |                                    |               |                |                             |          |                        |             |            |
| UMA-sm-OMC       | 6.18                               | 7.38          | 0.07           | 0.74                        | 0.87     | 0.18                   | 0.08        | 0.91       |

Open Molecular Crystals (OMC25) [2] evaluates whether a model can predict the packing of molecule into crystal structures. This task requires the accurate estimation of inter-molecular forces. Results on validation and test splits are shown in Table 16. It is notable that all sizes of UMA models outperform the UMA-sm-OMC model trained on only OMC25. This indicates that the other datasets, such as OMol25, provide useful complementary information for the task. One important and real-world task for molecular crystals is to predict the lowest energy packing, called a polymorph, for a molecule. Results for this task for a subset of molecular crystal polymorphs from the most recent

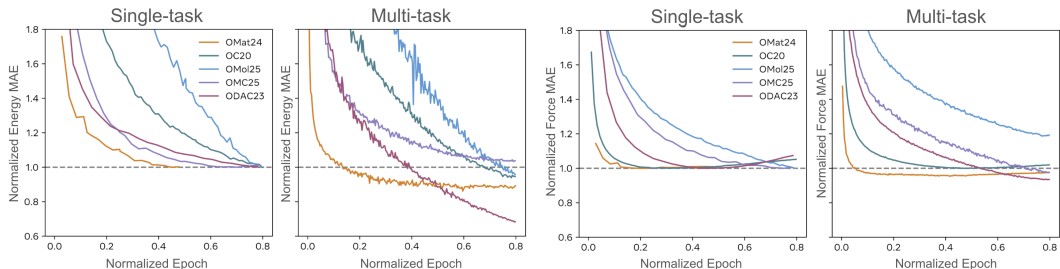


Figure 1: Pre-training curves of UMA-Ig for both single-task and multi-task models. Errors are normalized based on single-task performance. Note single-task models can overfit (forces on right), and the multi-task model generally converges to lower errors.

7th Crystal Structure Prediction (CSP) Blind Test [21] are shown in Table 17. The pymatgen’s [35] StructureMatcher class with default settings is used to match DFT and UMA-relaxed polymorphs, and root mean square deviation (RMSD) is computed for matches. Similar to the test metrics, the UMA models outperform the UMA-sm-OMC trained on only OMC25.

## E.6 DAC

Table 18: OpenDAC [43] val and test table. All energies are in meV and forces are in meV/Å.

| Model             | Val (Total Energy) |        |              | Test (Ads. Energy) |        |              |        |        |              |
|-------------------|--------------------|--------|--------------|--------------------|--------|--------------|--------|--------|--------------|
|                   |                    |        |              | ID                 |        |              | OOD-LT |        |              |
|                   | Energy             | Forces | Force CoVine | Energy             | Forces | Force CoVine | Energy | Forces | Force CoVine |
| <b>UMA</b>        |                    |        |              |                    |        |              |        |        |              |
| UMA-sm            | 60.4               | 5.9    | 0.82         | 169.5              | 16.7   | 0.63         | 292.4  | 16.0   | 0.57         |
| UMA-md            | 59.3               | 3.8    | 0.91         | 167.3              | 14.8   | 0.62         | 290.2  | 10.7   | 0.76         |
| UMA-Ig            | 38.7               | 3.3    | 0.91         | 177.1              | 7.8    | 0.82         | 291.1  | 6.5    | 0.91         |
| <b>Literature</b> |                    |        |              |                    |        |              |        |        |              |
| eqV2-ODAC [43]    | -                  | -      | -            | 145.0              | 8.2    | 0.69         | 316.0  | 7.2    | 0.72         |

The results on OpenDAC [43] are shown in Table 18. The OpenDAC dataset contains Metal Organic Frameworks (MOFs) with CO<sub>2</sub> and water molecules. The goal is to estimate the change in energy in the presence with and without the CO<sub>2</sub> and water molecules. These adsorption energies are computed in the same manner as for catalysts. The use of total energies leads to significantly better adsorption energy estimates, similar to catalysis. The forces of UMA-md and UMA-Ig are similar to the SoTA eqV2-ODAC [43] model.

## F Single-task vs Multi-task

For large models, multi-task training offers benefits even without MoLE. In Figure 1, we plot the direct-force pre-training curves of UMA-Ig and single-task models with the same model architecture and size. All metrics are normalized to those achieved with the models trained on single tasks to easily compare their relative performance with multi-task models. We observe that single task models frequently overfit to forces (OMat overfits upon further training), while the multi-task UMA model does not. Furthermore, UMA achieves lower losses in most cases. The one exception is OMol forces, for which errors are already small ( $< 10$  meV/Å) for both models.

## G MoLE Expert Analysis

Using a limited validation set consisting of 10,000 OMat24, 5,000 OC20, 20,000 OMol25, 10,000 OMC25, and 5,000 ODAC23 samples, we calculate the mean expert coefficient for each element-expert pair across all systems where the pair appears. We visualize these results using 32 periodic tables, each representing one of the 32 experts in UMA-sm (Figure 2).

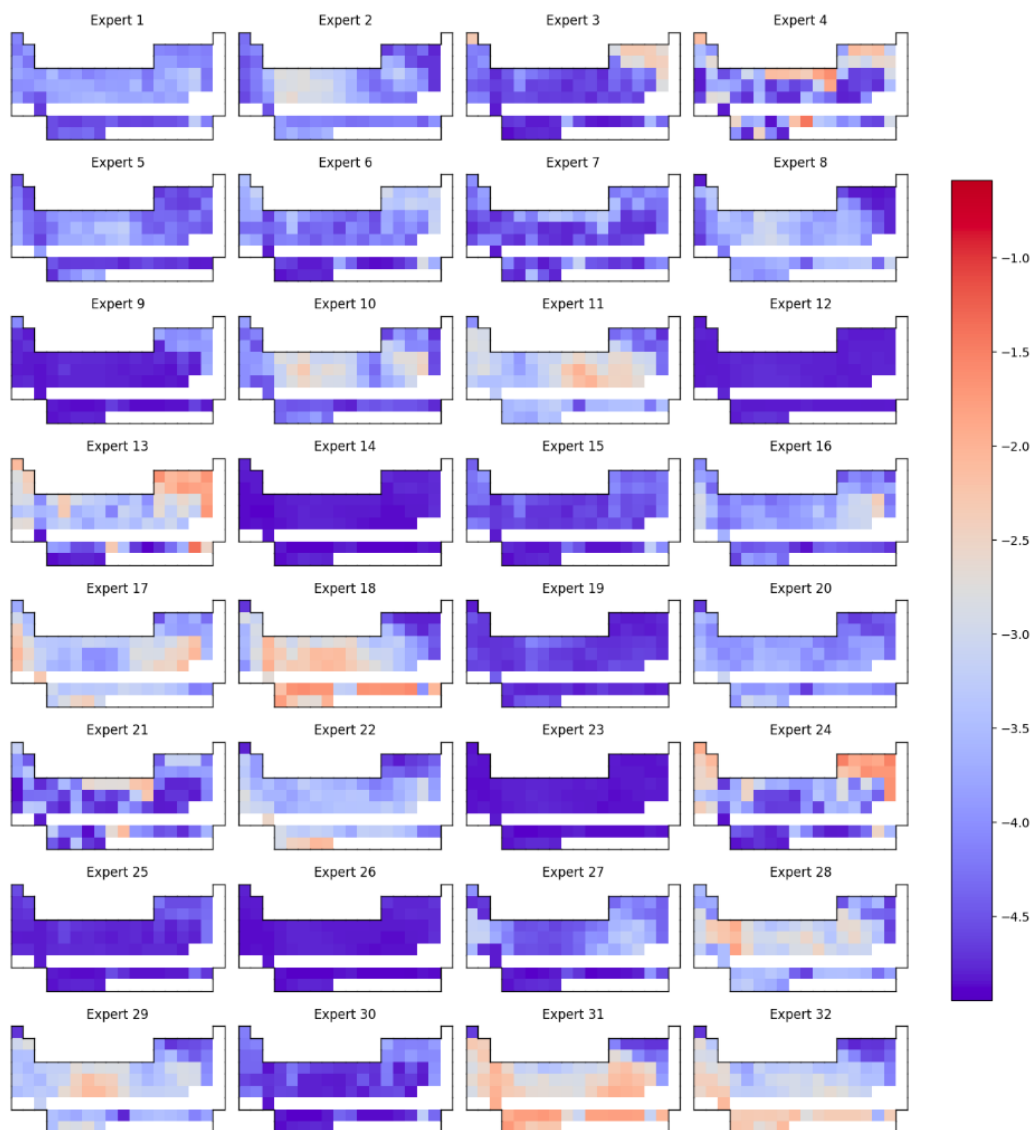


Figure 2: Log mean expert coefficient across element-expert pairs.

## References

- [1] Qianxiang Ai, Vinayak Bhat, Sean M. Ryno, Karol Jarolimek, Parker Sornberger, Andrew Smith, Michael M. Haley, John E. Anthony, and Chad Risko. OCELOT: An infrastructure for data-driven research to discover and design crystalline organic semiconductors. *The Journal of Chemical Physics*, 154(17):174705, May 2021. ISSN 0021-9606. doi: 10.1063/5.0048714.
- [2] Anonymous authors. The Open Molecular Crystals 2025 (OMC25) dataset. Forthcoming.
- [3] Anonymous authors. The Open Molecules 2025 (OMol25) Dataset, Evaluations, and Models. Provided in supplementary material.
- [4] Luis Barroso-Luque, Julia H. Yang, Fengyu Xie, Tina Chen, Ronald L. Kam, Zinab Jadidi, Peichen Zhong, and Gerbrand Ceder. SMOL: A Python package for cluster expansions and beyond. *Journal of Open Source Software*, 7(77):4504, 2022. ISSN 2475-9066. doi: 10.21105/joss.04504.

- [5] Luis Barroso-Luque, Muhammed Shuaibi, Xiang Fu, Brandon M Wood, Misko Dzamba, Meng Gao, Ammar Rizvi, C Lawrence Zitnick, and Zachary W Ulissi. Open materials 2024 (omat24) inorganic materials dataset and models. *arXiv preprint arXiv:2410.12771*, 2024.
- [6] Ilyes Batatia, Philipp Benner, Yuan Chiang, Alin M Elena, Dávid P Kovács, Janosh Riebesell, Xavier R Advincula, Mark Asta, Matthew Avaylon, William J Baldwin, et al. A foundation model for atomistic materials chemistry. *arXiv preprint arXiv:2401.00096*, 2023.
- [7] Filippo Bigi, Marcel Langer, and Michele Ceriotti. The dark side of the forces: assessing non-conservative force models for atomistic machine learning. *arXiv preprint arXiv:2412.11569*, 2024.
- [8] Anton Bochkarev, Yury Lysogorskiy, and Ralf Drautz. Graph atomic cluster expansion for semilocal interactions beyond equivariant message passing. *Physical Review X*, 14(2):021036, 2024.
- [9] Stanislav S. Borysov, R. Matthias Geilhufe, and Alexander V. Balatsky. Organic materials database: An open-access online database for data mining. *PLOS ONE*, 12(2):e0171501, February 2017. ISSN 1932-6203. doi: 10.1371/journal.pone.0171501.
- [10] Johann Brehmer, Sönke Behrends, Pim de Haan, and Taco Cohen. Does equivariance matter at scale? *arXiv preprint arXiv:2410.23179*, 2024.
- [11] Lowik Chanussot, Abhishek Das, Siddharth Goyal, Thibaut Lavril, Muhammed Shuaibi, Morgane Riviere, Kevin Tran, Javier Heras-Domingo, Caleb Ho, Weihua Hu, et al. Open catalyst 2020 (oc20) dataset and community challenges. *Acs Catalysis*, 11(10):6059–6072, 2021.
- [12] Yongchul G. Chung, Jeffrey Camp, Maciej Haranczyk, Benjamin J. Sikora, Wojciech Bury, Vaiva Krungleviciute, Taner Yildirim, Omar K. Farha, David S. Sholl, and Randall Q. Snurr. Computation-ready, experimental metal–organic frameworks: A tool to enable high-throughput screening of nanoporous crystals. *Chem. Mater.*, 26(21):6185–6192, 2014. doi: 10.1021/cm502594j.
- [13] Yongchul G. Chung, Emmanuel Haldoupis, Benjamin J. Bucior, Maciej Haranczyk, Seulchan Lee, Hongda Zhang, Konstantinos D. Vogiatzis, Marija Milisavljevic, Sanliang Ling, Jeffrey S. Camp, Ben Slater, J. Ilja Siepmann, David S. Sholl, and Randall Q. Snurr. Advances, updates, and analytics for the computation-ready, experimental metal–organic framework database: Core mof 2019. *J. Chem. Eng. Data*, 64(12):5985–5998, 2019. doi: 10.1021/acs.jced.9b00835.
- [14] Maarten de Jong, Wei Chen, Thomas Angsten, Anubhav Jain, Randy Notestine, Anthony Gamst, Marcel Sluiter, Chaitanya Krishna Ande, Sybrand van der Zwaag, Jose J. Plata, Cormac Toher, Stefano Curtarolo, Gerbrand Ceder, Kristin A. Persson, and Mark Asta. Charting the complete elastic properties of inorganic crystalline compounds. *Scientific Data*, 2(1):150009, March 2015. ISSN 2052-4463. doi: 10.1038/sdata.2015.9.
- [15] Bowen Deng, Peichen Zhong, KyuJung Jun, Janosh Riebesell, Kevin Han, Christopher J Bartel, and Gerbrand Ceder. Chgnet as a pretrained universal neural network potential for charge-informed atomistic modelling. *Nature Machine Intelligence*, 5(9):1031–1041, 2023.
- [16] Xiang Fu, Brandon M Wood, Luis Barroso-Luque, Daniel S Levine, Meng Gao, Misko Dzamba, and C Lawrence Zitnick. Learning smooth and expressive interatomic potentials for physical property prediction. *arXiv preprint arXiv:2502.12147*, 2025.
- [17] Johannes Gasteiger, Muhammed Shuaibi, Anuroop Sriram, Stephan Günnemann, Zachary Ulissi, C Lawrence Zitnick, and Abhishek Das. Gemnet-oc: developing graph neural networks for large and diverse molecular simulation datasets. *arXiv preprint arXiv:2204.02782*, 2022.
- [18] Stefan Grimme, Jens Antony, Stephan Ehrlich, and Helge Krieg. A consistent and accurate ab initio parametrization of density functional dispersion correction (DFT-D) for the 94 elements H-Pu. *The Journal of Chemical Physics*, 132(15):154104, April 2010. ISSN 0021-9606. doi: 10.1063/1.3382344.

- [19] Bjørk Hammer, Lars Bruno Hansen, and Jens Kehlet Nørskov. Improved adsorption energetics within density-functional theory using revised perdew-burke-ernzerhof functionals. *Physical Review B*, 59(11):7413, 1999.
- [20] Jordan Hoffmann, Sebastian Borgeaud, Arthur Mensch, Elena Buchatskaya, Trevor Cai, Eliza Rutherford, Diego de Las Casas, Lisa Anne Hendricks, Johannes Welbl, Aidan Clark, et al. Training compute-optimal large language models. *arXiv preprint arXiv:2203.15556*, 2022.
- [21] Lily M Hunnisett, Nicholas Francia, Jonas Nyman, Nathan S Abraham, Srinivasulu Aitipamula, Tamador Alkhidir, Mubarak Almehairbi, Andrea Anelli, Dylan M Anstine, John E Anthony, et al. The seventh blind test of crystal structure prediction: Structure ranking methods. *Acta Crystallographica Section B: Structural Science, Crystal Engineering and Materials*, 80(6): 548–574, December 2024. ISSN 2052-5206. doi: 10.1107/S2052520624008679.
- [22] Lily M Hunnisett, Jonas Nyman, Nicholas Francia, Nathan S Abraham, Claire S Adjiman, Srinivasulu Aitipamula, Tamador Alkhidir, Mubarak Almehairbi, Andrea Anelli, Dylan M Anstine, et al. The seventh blind test of crystal structure prediction: Structure generation methods. *Acta Crystallographica Section B: Structural Science, Crystal Engineering and Materials*, 80(6): 517–547, December 2024. ISSN 2052-5206. doi: 10.1107/S2052520624007492.
- [23] Aaron D. Kaplan, Runze Liu, Ji Qi, Tsz Wai Ko, Bowen Deng, Janosh Riebesell, Gerbrand Ceder, Kristin A. Persson, and Shyue Ping Ong. A Foundational Potential Energy Surface Dataset for Materials, March 2025.
- [24] Jared Kaplan, Sam McCandlish, Tom Henighan, Tom B Brown, Benjamin Chess, Rewon Child, Scott Gray, Alec Radford, Jeffrey Wu, and Dario Amodei. Scaling laws for neural language models. *arXiv preprint arXiv:2001.08361*, 2020.
- [25] Jaesun Kim, Jisu Kim, Jaehoon Kim, Jiho Lee, Yutack Park, Youngho Kang, and Seungwu Han. Data-efficient multifidelity training for high-fidelity machine learning interatomic potentials. *Journal of the American Chemical Society*, 147(1):1042–1054, 2024.
- [26] Georg Kresse and Jürgen Furthmüller. Efficient iterative schemes for ab initio total-energy calculations using a plane-wave basis set. *Physical review B*, 54(16):11169, 1996.
- [27] Georg Kresse and Jürgen Hafner. Ab initio molecular dynamics for liquid metals. *Physical review B*, 47(1):558, 1993.
- [28] Georg Kresse and Jürgen Hafner. Norm-conserving and ultrasoft pseudopotentials for first-row and transition elements. *Journal of Physics: Condensed Matter*, 6(40):8245, 1994.
- [29] Georg Kresse and Daniel Joubert. From ultrasoft pseudopotentials to the projector augmented-wave method. *Physical review b*, 59(3):1758, 1999.
- [30] Janice Lan, Aini Palizhati, Muhammed Shuaibi, Brandon M Wood, Brook Wander, Abhishek Das, Matt Uyttendaele, C Lawrence Zitnick, and Zachary W Ulissi. Adsorbml: a leap in efficiency for adsorption energy calculations using generalizable machine learning potentials. *npj Computational Materials*, 9(1):172, 2023.
- [31] Yi-Lun Liao, Brandon Wood, Abhishek Das, and Tess Smidt. Equiformerv2: Improved equivariant transformer for scaling to higher-degree representations. *arXiv preprint arXiv:2306.12059*, 2023.
- [32] Antoine Loew, Dewen Sun, Hai-Chen Wang, Silvana Botti, and Miguel A. L. Marques. Universal Machine Learning Interatomic Potentials are Ready for Phonons, December 2024.
- [33] Łukasz Mentel. mendeleeev - A Python package with properties of chemical elements, ions, isotopes and methods to manipulate and visualize periodic table., March 2021. URL <https://github.com/lmentel/mendeleeev>.
- [34] Frank Neese, Frank Wennmohs, Ute Becker, and Christoph Riplinger. The orca quantum chemistry program package. *The Journal of chemical physics*, 152(22), 2020.

- [35] Shyue Ping Ong, William Davidson Richards, Anubhav Jain, Geoffroy Hautier, Michael Kocher, Shreyas Cholia, Dan Gunter, Vincent L Chevrier, Kristin A Persson, and Gerbrand Ceder. Python materials genomics (pymatgen): A robust, open-source python library for materials analysis. *Computational Materials Science*, 68:314–319, 2013.
- [36] Saro Passaro and C. Lawrence Zitnick. Reducing SO(3) convolutions to SO(2) for efficient equivariant GNNs. In *Proceedings of the 40th International Conference on Machine Learning*, volume 202. PMLR, 2023.
- [37] John P Perdew, Kieron Burke, and Matthias Ernzerhof. Generalized gradient approximation made simple. *Physical review letters*, 77(18):3865, 1996.
- [38] Balázs Póta, Paramvir Ahlawat, Gábor Csányi, and Michele Simoncelli. Thermal Conductivity Predictions with Foundation Atomistic Models, September 2024.
- [39] Benjamin Rhodes, Sander Vandenhaute, Vaidotas Šimkus, James Gin, Jonathan Godwin, Tim Duignan, and Mark Neumann. Orb-v3: atomistic simulation at scale. *arXiv preprint arXiv:2504.06231*, 2025.
- [40] Janosh Riebesell, Rhys EA Goodall, Philipp Benner, Yuan Chiang, Bowen Deng, Alpha A Lee, Anubhav Jain, and Kristin A Persson. Matbench discovery—a framework to evaluate machine learning crystal stability predictions. *arXiv preprint arXiv:2308.14920*, 2023.
- [41] Jonathan Schmidt, Tiago FT Cerqueira, Aldo H Romero, Antoine Loew, Fabian Jäger, Hai-Chen Wang, Silvana Botti, and Miguel AL Marques. Improving machine-learning models in materials science through large datasets. *Materials Today Physics*, 48:101560, 2024.
- [42] Anuroop Sriram, Abhishek Das, Brandon M Wood, and C. Lawrence Zitnick. Towards training billion parameter graph neural networks for atomic simulations. In *International Conference on Learning Representations*, 2022. URL <https://openreview.net/forum?id=OjP2n0YFmKG>.
- [43] Anuroop Sriram, Sihoon Choi, Xiaohan Yu, Logan M Brabson, Abhishek Das, Zachary Ulissi, Matt Uyttendaele, Andrew J Medford, and David S Sholl. The open dac 2023 dataset and challenges for sorbent discovery in direct air capture, 2024.
- [44] Annika Stuke, Christian Kunkel, Dorothea Golze, Milica Todorović, Johannes T. Margraf, Karsten Reuter, Patrick Rinke, and Harald Oberhofer. Atomic structures and orbital energies of 61,489 crystal-forming organic molecules. *Scientific Data*, 7(1):58, February 2020. ISSN 2052-4463. doi: 10.1038/s41597-020-0385-y.
- [45] Rithwik Tom, Timothy Rose, Imanuel Bier, Harriet O’Brien, Álvaro Vázquez-Mayagoitia, and Noa Marom. Genarris 2.0: A random structure generator for molecular crystals. *Computer Physics Communications*, 250:107170, 2020.
- [46] Richard Tran, Janice Lan, Muhammed Shuaibi, Brandon M Wood, Siddharth Goyal, Abhishek Das, Javier Heras-Domingo, Adeesh Kolluru, Ammar Rizvi, Nima Shoghi, et al. The open catalyst 2022 (oc22) dataset and challenges for oxide electrocatalysts. *ACS Catalysis*, 13(5): 3066–3084, 2023.
- [47] Alex Zunger, S.-H. Wei, L. G. Ferreira, and James E. Bernard. Special quasirandom structures. *Physical Review Letters*, 65(3):353–356, July 1990. doi: 10.1103/PhysRevLett.65.353.