

## Appendix

### A Additional background

#### A.1 Group theory

#### A.2 Equivariance

#### A.3 eSCN Convolution

### B Details of architecture

### C Details of experiments on OC20

#### C.1 Training details

#### C.2 Details of Running Relaxations

#### C.3 Details of AdsorbML

## A Additional Background

We first provide relevant mathematical background on group theory and equivariance. We note that most of the content is adapted from Equiformer [17] and that these works [69, 70] have more in-depth and pedagogical discussions. Then, we provide mathematical details of eSCN convolutions.

### A.1 Group Theory

**Definition of Groups.** A group is an algebraic structure that consists of a set  $G$  and a binary operator  $\circ : G \times G \rightarrow G$ . Typically denoted as  $G$ , groups satisfy the following four axioms:

1. Closure:  $g \circ h \in G$  for all  $g, h \in G$ .
2. Identity: There exists an identity element  $e \in G$  such that  $g \circ e = e \circ g = g$  for all  $g \in G$ .
3. Inverse: For each  $g \in G$ , there exists an inverse element  $g^{-1} \in G$  such that  $g \circ g^{-1} = g^{-1} \circ g = e$ .
4. Associativity:  $g \circ h \circ i = (g \circ h) \circ i = g \circ (h \circ i)$  for all  $g, h, i \in G$ .

In this work, we consider 3D Euclidean symmetry, and relevant groups are:

1. The Euclidean group in three dimensions  $E(3)$ : 3D rotation, translation and inversion.
2. The special Euclidean group in three dimensions  $SE(3)$ : 3D rotation and translation.
3. The orthogonal group in three dimensions  $O(3)$ : 3D rotation and inversion.
4. The special orthogonal group in three dimensions  $SO(3)$ : 3D rotation.

Since eSCN [18] and this work only consider equivariance to 3D rotation and invariance to 3D translation but not inversion, we mainly discuss  $SE(3)$ -equivariance in the main text and in appendix and note that more details of  $E(3)$ -equivariance can be found in the work of Equiformer [17].

**Group Representations.** Given a vector space  $X$ , the way a group  $G$  acts on  $X$  is given by the group representation  $D_X$ .  $D_X$  is parameterized by  $g \in G$ , with  $D_X(g) : X \rightarrow X$ . Group representations  $D_X$  are invertible matrices, and group transformations, or group actions, take the form of matrix multiplications. This definition of group representations satisfies the requirements of groups, including associativity,  $D(g)D(h) = D(g \circ h)$  for all  $g, h \in G$ . We say that the two group representations  $D(g)$  and  $D'(g)$  are equivalent if there exists a change-of-basis  $N \times N$  matrix  $P$  such that  $P^{-1}D(g)P = D'(g)$  for all  $g \in G$ .  $D(g)$  is reducible if  $D'(g)$  is block diagonal for all  $g \in G$ , meaning that  $D'(g)$  acts on multiple independent subspaces of the vector space. Otherwise, the representation  $D(g)$  is said to be irreducible. Irreducible representations, or irreps, are a class of representations that are convenient for composing different group representations. Specifically, for the case of  $SO(3)$ , Wigner-D matrices are irreducible representations, and we can express any group representation of  $SO(3)$  as a direct sum (concatentation) of Wigner-D matrices [55, 69, 70]:

$$D(g) = P^{-1} \left( \bigoplus_i D^{(L_i)}(g) \right) P = P^{-1} \begin{pmatrix} D^{(L_0)}(g) & & \\ & D^{(L_1)}(g) & \\ & & \dots \end{pmatrix} P \quad (2)$$

where  $D^{(L_i)}(g)$  are Wigner-D matrices of degree  $L_i$ .

## 396 A.2 Equivariance

397 A function  $f$  mapping between vector spaces  $X$  and  $Y$  is equivariant to a group of transformations  $G$  if  
 398 for any input  $x \in X$ , output  $y \in Y$  and group element  $g \in G$ , we have  $f(D_X(g)x) = D_Y(g)f(x) =$   
 399  $D_Y(g)y$ , where  $D_X(g)$  and  $D_Y(g)$  are transformation matrices or group representations parametrized  
 400 by  $g$  in  $X$  and  $Y$ . Additionally,  $f$  is invariant when  $D_Y(g)$  is an identity matrix for any  $g \in G$ .

401 As neural networks comprise many composable operations, equivariant neural networks comprise  
 402 many equivariant operations to maintain the equivariance of input, intermediate, and output features.  
 403 Incorporating equivariance as a strong prior knowledge can help improve data efficiency and gen-  
 404 eralization of neural networks [5, 9, 71]. In this work, we achieve equivariance to 3D rotation by  
 405 operating on vector spaces of  $SO(3)$  irreps, incorporate invariance to 3D translation by acting on  
 406 relative positions, but do not consider inversion.

## 407 A.3 eSCN Convolution

408 Message passing is used to update equivariant irreps features and is typically implemented as  $SO(3)$   
 409 convolutions. A traditional  $SO(3)$  convolution interacts input irrep features  $x_{m_i}^{(L_i)}$  and spherical har-  
 410 monic projections of relative positions  $Y_{m_f}^{(L_f)}(\vec{r}_{ts})$  with an  $SO(3)$  tensor product with Clebsch-Gordan  
 411 coefficients  $C_{(L_i, m_i), (L_f, m_f)}^{(L_o, m_o)}$ . Since tensor products are compute-intensive, eSCN convolutions [18]  
 412 are proposed to reduce the complexity of tensor products when they are used in  $SO(3)$  convolutions.  
 413 Rotating the input irreps features  $x_{m_i}^{(L_i)}$  based on the relative position vectors  $\vec{r}_{ts}$  simplifies the tensor  
 414 products and enables reducing  $SO(3)$  convolutions to  $SO(2)$  linear operations. Below we provide  
 415 the mathematical details of  $SO(3)$  convolutions built from tensor products and how rotation can  
 416 reduce their computational complexity.

417 Tensor products interact type- $L_i$  vector  $x^{(L_i)}$  and type- $L_f$  vector  $f^{(L_f)}$  to produce type- $L_o$   
 418 vector  $y^{(L_o)}$  with Clebsch-Gordan coefficients  $C_{(L_i, m_i), (L_f, m_f)}^{(L_o, m_o)}$ . Clebsch-Gordan coefficients  
 419  $C_{(L_i, m_i), (L_f, m_f)}^{(L_o, m_o)}$  are non-zero only when  $|L_i - L_o| \leq L_f \leq |L_i + L_o|$ . Each non-trivial com-  
 420 bination of  $L_i \otimes L_f \rightarrow L_o$  is called a path, and each path is independently equivariant and can be  
 421 assigned a learnable weight  $w_{L_i, L_f, L_o}$ .

422 We consider the message  $m_{ts}$  sent from source node  $s$  to target node  $t$  in an  $SO(3)$  convolution. The  
 423  $L_o$ -th degree of  $m_{ts}$  can be expressed as:

$$m_{ts}^{(L_o)} = \sum_{L_i, L_f} w_{L_i, L_f, L_o} \left( x_s^{(L_i)} \otimes_{L_i, L_f}^{L_o} Y^{(L_f)}(\hat{r}_{ts}) \right) \quad (3)$$

424 where  $x_s$  is the irreps feature at source node  $s$ ,  $x_s^{(L_i)}$  denotes the  $L_i$ -th degree of  $x_s$ , and  $\hat{r}_{ts} = \frac{\vec{r}_{ts}}{|\vec{r}_{ts}|}$ .  
 425 The spherical harmonic projection of relative positions  $Y^{(L_f)}(\hat{r}_{ts})$  becomes sparse if we rotate  $\hat{r}_{ts}$   
 426 with a rotation matrix  $R_{ts}$  to align with the direction of  $L = 0$  and  $m = 0$ , which corresponds to the  
 427 z axis traditionally but the y axis in the conventions of e3nn [55]. Concretely, given  $R_{ts}\hat{r}_{ts}$  aligned  
 428 with the y axis,  $Y_{m_f}^{(L_f)}(R_{ts}\hat{r}_{ts}) \neq 0$  only for  $m_f = 0$ . Without loss of equivariance, we re-scale  
 429  $Y_0^{(L_f)}(R_{ts}\vec{r}_{ts})$  to be one. Therefore, by rotating  $x_s^{(L_i)}$  and  $Y^{(L_f)}$  based on  $\hat{r}_{ts}$ , we can simplify Eq. 3

430 as follows:

$$\begin{aligned}
m_{ts}^{(L_o)} &= \left(D^{(L_o)}(R_{ts})\right)^{-1} \sum_{L_i, L_f} w_{L_i, L_f, L_o} \left(D^{(L_i)}(R_{ts}) x_s^{(L_i)} \otimes_{L_i, L_f}^{L_o} Y^{(L_f)}(R_{ts} \hat{r}_{ts})\right) \\
&= \left(D^{(L_o)}\right)^{-1} \sum_{L_i, L_f} w_{L_i, L_f, L_o} \bigoplus_{m_o} \left( \sum_{m_i, m_f} \left(D^{(L_i)} x_s^{(L_i)}\right)_{m_i} C_{(L_i, m_i), (L_f, m_f)}^{(L_o, m_o)} \left(Y^{(L_f)}(R_{ts} \hat{r}_{ts})\right)_{m_f} \right) \\
&= \left(D^{(L_o)}\right)^{-1} \sum_{L_i, L_f} w_{L_i, L_f, L_o} \bigoplus_{m_o} \left( \sum_{m_i} \left(D^{(L_i)} x_s^{(L_i)}\right)_{m_i} C_{(L_i, m_i), (L_f, 0)}^{(L_o, m_o)} \right) \\
&= \left(D^{(L_o)}\right)^{-1} \sum_{L_i, L_f} w_{L_i, L_f, L_o} \bigoplus_{m_o} \left( \sum_{m_i} \left(\tilde{x}_s^{(L_i)}\right)_{m_i} C_{(L_i, m_i), (L_f, 0)}^{(L_o, m_o)} \right)
\end{aligned} \tag{4}$$

431 where  $D^{(L_i)}(R_{ts}) = D^{(L_i)}$  and  $D^{(L_o)}(R_{ts}) = D^{(L_o)}$  denote Wigner-D matrices of degrees  $L_i$  and  
432  $L_o$  based on rotation matrix  $R_{ts}$ , respectively,  $\bigoplus$  denotes concatenation, and  $D^{(L_i)} x_s^{(L_i)} = \tilde{x}_s^{(L_i)}$ .  
433 Additionally, given  $m_f = 0$ , Clebsch-Gordan coefficients  $C_{(L_i, m_i), (L_f, 0)}^{(L_o, m_o)}$  are sparse and are non-zero  
434 only when  $m_i = \pm m_o$ , which further simplifies Eq. 4:

$$m_{ts}^{(L_o)} = \left(D^{(L_o)}\right)^{-1} \sum_{L_i, L_f} w_{L_i, L_f, L_o} \bigoplus_{m_o} \left( \left(\tilde{x}_s^{(L_i)}\right)_{m_o} C_{(L_i, m_o), (L_f, 0)}^{(L_o, m_o)} + \left(\tilde{x}_s^{(L_i)}\right)_{-m_o} C_{(L_i, -m_o), (L_f, 0)}^{(L_o, m_o)} \right) \tag{5}$$

435 By re-ordering the summations and concatenation in Eq. 5, we have:

$$\left(D^{(L_o)}\right)^{-1} \sum_{L_i} \bigoplus_{m_o} \left( \left(\tilde{x}_s^{(L_i)}\right)_{m_o} \sum_{L_f} \left(w_{L_i, L_f, L_o} C_{(L_i, m_o), (L_f, 0)}^{(L_o, m_o)}\right) + \left(\tilde{x}_s^{(L_i)}\right)_{-m_o} \sum_{L_f} \left(w_{L_i, L_f, L_o} C_{(L_i, -m_o), (L_f, 0)}^{(L_o, m_o)}\right) \right) \tag{6}$$

436 Instead of using learnable parameters for  $w_{L_i, L_f, L_o}$ , eSCN proposes to parametrize  $\tilde{w}_{m_o}^{(L_i, L_o)}$  and  
437  $\tilde{w}_{-m_o}^{(L_i, L_o)}$  as below:

$$\begin{aligned}
\tilde{w}_{m_o}^{(L_i, L_o)} &= \sum_{L_f} w_{L_i, L_f, L_o} C_{(L_i, m_o), (L_f, 0)}^{(L_o, m_o)} = \sum_{L_f} w_{L_i, L_f, L_o} C_{(L_i, -m_o), (L_f, 0)}^{(L_o, -m_o)} \quad \text{for } m_o \geq 0 \\
\tilde{w}_{-m_o}^{(L_i, L_o)} &= \sum_{L_f} w_{L_i, L_f, L_o} C_{(L_i, m_o), (L_f, 0)}^{(L_o, -m_o)} = - \sum_{L_f} w_{L_i, L_f, L_o} C_{(L_i, -m_o), (L_f, 0)}^{(L_o, m_o)} \quad \text{for } m_o > 0
\end{aligned} \tag{7}$$

438 The parametrization of  $\tilde{w}_{m_o}^{(L_i, L_o)}$  and  $\tilde{w}_{-m_o}^{(L_i, L_o)}$  enables removing the summation over  $L_f$  and further  
439 simplifies the computation. By combining Eq. 6 and Eq. 7, we have:

$$\begin{aligned}
m_{ts}^{(L_o)} &= \left(D^{(L_o)}\right)^{-1} \sum_{L_i} \bigoplus_{m_o} \left(y_{ts}^{(L_i, L_o)}\right)_{m_o} \\
\left(y_{ts}^{(L_i, L_o)}\right)_{m_o} &= \tilde{w}_{m_o}^{(L_i, L_o)} \left(\tilde{x}_s^{(L_i)}\right)_{m_o} - \tilde{w}_{-m_o}^{(L_i, L_o)} \left(\tilde{x}_s^{(L_i)}\right)_{-m_o} \quad \text{for } m_o > 0 \\
\left(y_{ts}^{(L_i, L_o)}\right)_{-m_o} &= \tilde{w}_{-m_o}^{(L_i, L_o)} \left(\tilde{x}_s^{(L_i)}\right)_{m_o} + \tilde{w}_{m_o}^{(L_i, L_o)} \left(\tilde{x}_s^{(L_i)}\right)_{-m_o} \quad \text{for } m_o > 0 \\
\left(y_{ts}^{(L_i, L_o)}\right)_{m_o} &= \tilde{w}_{m_o}^{(L_i, L_o)} \left(\tilde{x}_s^{(L_i)}\right)_{m_o} \quad \text{for } m_o = 0
\end{aligned} \tag{8}$$

440 The formulation of  $y_{ts}^{(L_i, L_o)}$  coincides with performing  $SO(2)$  linear operations [18, 72]. Addition-  
441 ally, eSCN convolutions can further simplify the computation by considering only a subset of  $m_o$   
442 components in Eq. 8, i.e.,  $|m_o| \leq M_{max}$ .

443 In summary, efficient  $SO(3)$  convolutions can be achieved by first rotating irreps features  $x_s^{(L_i)}$  based  
444 on relative position vectors  $\vec{r}_{ts}$  and then performing  $SO(2)$  linear operations on rotated features.  
445 The key idea is that rotation simplifies the computation as in Eq. 4, 5, 7, and 8. Please refer to their

work [18] for more details. We note that eSCN convolutions consider only simplifying the case of taking tensor products between input irreps features and spherical harmonic projections of relative position vectors. eSCN convolutions do not simplify general cases such as taking tensor products between input irreps features and themselves [38] since the relative position vectors used to rotate irreps features are not clearly defined.

## B Details of Architecture

In this section, we define architectural hyper-parameters like maximum degrees and numbers of channels in certain layers in EquiformerV2, which are used to specify the detailed architectures in Sec. C.1. Besides, we note that eSCN [18] and this work mainly consider  $SE(3)$ -equivariance.

We denote embedding dimensions as  $d_{embed}$ , which defines the dimensions of most irreps features. Specifically, the output irreps features of all modules except the output head in Figure 1a have dimension  $d_{embed}$ . For separable  $S^2$  activation as illustrated in Figure 2c, we denote the resolution of point samples on a sphere as  $R$ , which can depend on maximum degree  $L_{max}$ , and denote the unconstrained functions after projecting to point samples as  $F$ .

For equivariant graph attention in Figure 1b, the input irreps features  $x_i$  and  $x_j$  have dimension  $d_{embed}$ . The dimension of the irreps feature  $f_{ij}^{(L)}$  is denoted as  $d_{attn\_hidden}$ . Equivariant graph attention can have  $h$  parallel attention functions. For each attention function, we denote the dimension of the scalar feature  $f_{ij}^{(0)}$  as  $d_{attn\_alpha}$  and denote the dimension of the value vector, which is in the form of irreps features, as  $d_{attn\_value}$ . For the separable  $S^2$  activation used in equivariant graph attention, the resolution of point samples is  $R$ , and we use a single SiLU activation for  $F$ . We share the layer normalization in attention re-normalization across all  $h$  attention functions but have different  $h$  linear layers after that. The last linear layer projects the dimension back to  $d_{embed}$ . The two intermediate  $SO(2)$  linear layers operate with maximum degree  $L_{max}$  and maximum order  $M_{max}$ .

For feed forward networks (FFNs) in Figure 1d, we denote the dimension of the output irreps features of the first linear layer as  $d_{ffn}$ . For the separable  $S^2$  activation used in FFNs, the resolution of point samples is  $R$ , and  $F$  consists of a two-layer MLP, with each linear layer followed by SiLU, and a final linear layer. The linear layers have the same number of channels as  $d_{ffn}$ .

For radial functions, we denote the dimension of hidden scalar features as  $d_{edge}$ . For experiments on OC20, same as eSCN [18], we use Gaussian radial basis to represent relative distances and additionally embed the atomic numbers at source nodes and target nodes with two scalar features of dimension  $d_{edge}$ . The radial basis and the two embeddings of atomic numbers are fed to the radial function to generate distance embeddings.

The maximum degree of irreps features is denoted as  $L_{max}$ . All irreps features have degrees from 0 to  $L_{max}$  and have  $C$  channels for each degree. We denote the dimension as  $(L_{max}, C)$ . For example, irreps feature  $x_{irreps}$  of dimension (6, 128) has maximum degree 6 and 128 channels for each degree. The dimension of scalar feature  $x_{scalar}$  can be expressed as  $(0, C_{scalar})$ .

Following Equiformer [17], we apply dropout [40] to attention weights and stochastic depth [41] to outputs of equivariant graph attention and feed forward networks. However, we do not apply dropout or stochastic depth to the output head.

## C Details of Experiments on OC20

### C.1 Training Details

**Hyper-Parameters.** We summarize the hyper-parameters for the base model setting on OC20 S2EF-2M dataset and the main results on OC20 S2EF-All and S2EF-All+MD datasets in Table 4. For the ablation studies on OC20 S2EF-2M dataset, when trained for 20 or 30 epochs as in Table 1b, we increase the learning rate from  $2 \times 10^{-4}$  to  $4 \times 10^{-4}$ . When using  $L_{max} = 8$  as in Table 1c, we increase the resolution of point samples  $R$  from 18 to 20. We vary  $L_{max}$  and the widths for speed-accuracy trade-offs in Figure 4. Specifically, we first decrease  $L_{max}$  from 6 to 4. Then, we multiply  $h$  and the number of channels of  $(d_{embed}, d_{attn\_hidden}, d_{ffn})$  by 0.75 and 0.5. We train all models for 30 epochs. The same strategy to scale down eSCN models is adopted for fair comparisons.

Hyper-parameters	S2EF-2M	S2EF-All/S2EF-All+MD
Optimizer	AdamW	AdamW
Learning rate scheduling	Cosine learning rate with linear warmup	Cosine learning rate with linear warmup
Warmup epochs	0.1	0.01
Maximum learning rate	$2 \times 10^{-4}$	$4 \times 10^{-4}$
Batch size	64	256 for S2EF-All, 512 for S2EF-All+MD
Number of epochs	12	1
Weight decay	$1 \times 10^{-3}$	$1 \times 10^{-3}$
Dropout rate	0.1	0.1
Stochastic depth	0.05	0.1
Energy coefficient $\lambda_E$	2	2 for S2EF-All, 2, 4 for S2EF-All+MD
Force coefficient $\lambda_F$	100	100
Gradient clipping norm threshold	100	100
Model EMA decay	0.999	0.999
Cutoff radius ( $\text{\AA}$ )	12	12
Maximum number of neighbors	20	20
Number of radial basis	600	600
Dimension of hidden scalar features in radial functions $d_{edge}$	(0, 128)	(0, 128)
Maximum degree $L_{max}$	6	6
Maximum order $M_{max}$	2	3
Number of Transformer blocks	12	20
Embedding dimension $d_{embed}$	(6, 128)	(6, 128)
$f_{ij}^{(L)}$ dimension $d_{attn\_hidden}$	(6, 64)	(6, 64)
Number of attention heads $h$	8	8
$f_{ij}^{(0)}$ dimension $d_{attn\_alpha}$	(0, 64)	(0, 64)
Value dimension $d_{attn\_value}$	(6, 16)	(6, 16)
Hidden dimension in feed forward networks $d_{ffn}$	(6, 128)	(6, 128)
Resolution of point samples $R$	18	18

Table 4: Hyper-parameters for the base model setting on OC20 S2EF-2M dataset and the main results on OC20 S2EF-All and S2EF-All+MD datasets.

Training set	Attention re-normalization	Activation	Normalization	$L_{max}$	$M_{max}$	Training time (GPU-hours)	Inference speed (Samples / GPU sec.)	Number of parameters
S2EF-2M	✗	Gate	LN	6	2	965	19.06	91.06M
	✓	Gate	LN	6	2	998	19.07	91.06M
	✓	$S^2$	LN	6	2	1476	12.80	81.46M
	✓	Sep. $S^2$	LN	6	2	1505	12.51	83.16M
	✓	Sep. $S^2$	SLN	6	2	1412	13.22	83.16M
	✓	Sep. $S^2$	SLN	4	2	965	19.86	44.83M
	✓	Sep. $S^2$	SLN	8	2	2709	7.86	134.28M
	✓	Sep. $S^2$	SLN	6	3	1623	11.92	95.11M
	✓	Sep. $S^2$	SLN	6	4	2706	7.98	102.14M
	✓	Sep. $S^2$	SLN	6	6	3052	7.13	106.63M
S2EF-All	✓	Sep. $S^2$	SLN	6	3	20499	6.08	153.60M
S2EF-All+MD ( $\lambda_E = 2$ )	✓	Sep. $S^2$	SLN	6	3	32834	6.08	153.60M
S2EF-All+MD ( $\lambda_E = 4$ )	✓	Sep. $S^2$	SLN	6	3	37692	6.08	153.60M

Table 5: Training time, inference speed and numbers of parameters of different models trained on OC20 S2EF-2M, S2EF-All and S2EF-All+MD datasets. All numbers are measured on V100 GPUs with 32GB.

**Training Time, Inference Speed and Numbers of Parameters.** Table 5 summarizes the training time, inference speed and numbers of parameters of models in Tables 1a (rows 1, 2, 3, 4, 5), 1c, 1d and 2. We use 16 V100 GPUs with 32GB to train each individual model on S2EF-2M dataset, 64 GPUs for S2EF-All and 128 GPUs for S2EF-All+MD.

## C.2 Details of Running Relaxations

A structural relaxation is a local optimization where atom positions are iteratively updated based on forces to minimize the energy of the structure. We perform ML relaxations using the LBFGS optimizer (quasi-Newton) implemented in the Open Catalyst Github repository [30]. The structural relaxations for OC20 IS2RE and IS2RS tasks are allowed to run for 200 steps or until the maximum predicted force per atom  $F_{max} \leq 0.02 \text{ eV/\AA}$ , and the relaxations for AdsorbML are allowed to run for 300 steps or until  $F_{max} \leq 0.02 \text{ eV/\AA}$ . These settings are chosen to be consistent with prior works. We run relaxations on V100 GPUs with 32GB. The computational cost of running relaxations for OC20 IS2RE and IS2RS tasks is 1011 GPU-hours, and that of running ML relaxations of AdsorbML is 1075 GPU-hours.

509 **C.3 Details of AdsorbML**

510 We run the AdsorbML algorithm on the OC20-Dense dataset in accordance with the procedure laid  
511 out in the paper [10], which is summarized here:

- 512 1. Run ML relaxations on all initial structures in the OC20-Dense dataset. There are around  
513 1000 different adsorbate-surface combinations with about 90 adsorbate placements per  
514 combination, and therefore we have roughly 90k structures in total.
- 515 2. Remove invalid ML relaxed structures based on physical constraints and rank the other ML  
516 relaxed structures in order of lowest to highest ML predicted energy.
- 517 3. Take the top  $k$  ML relaxed structures with the lowest ML predicted energies for each  
518 adsorbate-surface combination and run DFT single-point calculations. The single-point  
519 calculations are performed on the ML relaxed structures to improve the energy predictions  
520 without running a full DFT relaxation and are run with VASP using the same setting as the  
521 original AdsorbML experiments. As shown in Table 3, we vary  $k$  from 1 to 5.
- 522 4. Compute success and speedup metrics based on our lowest DFT single-point energy per  
523 adsorbate-surface combination and the DFT labels provided in the OC20-Dense dataset.

524 We visualize some examples of relaxed structures from eSCN [18], EquiformerV2 and DFT in  
525 Figure 5.



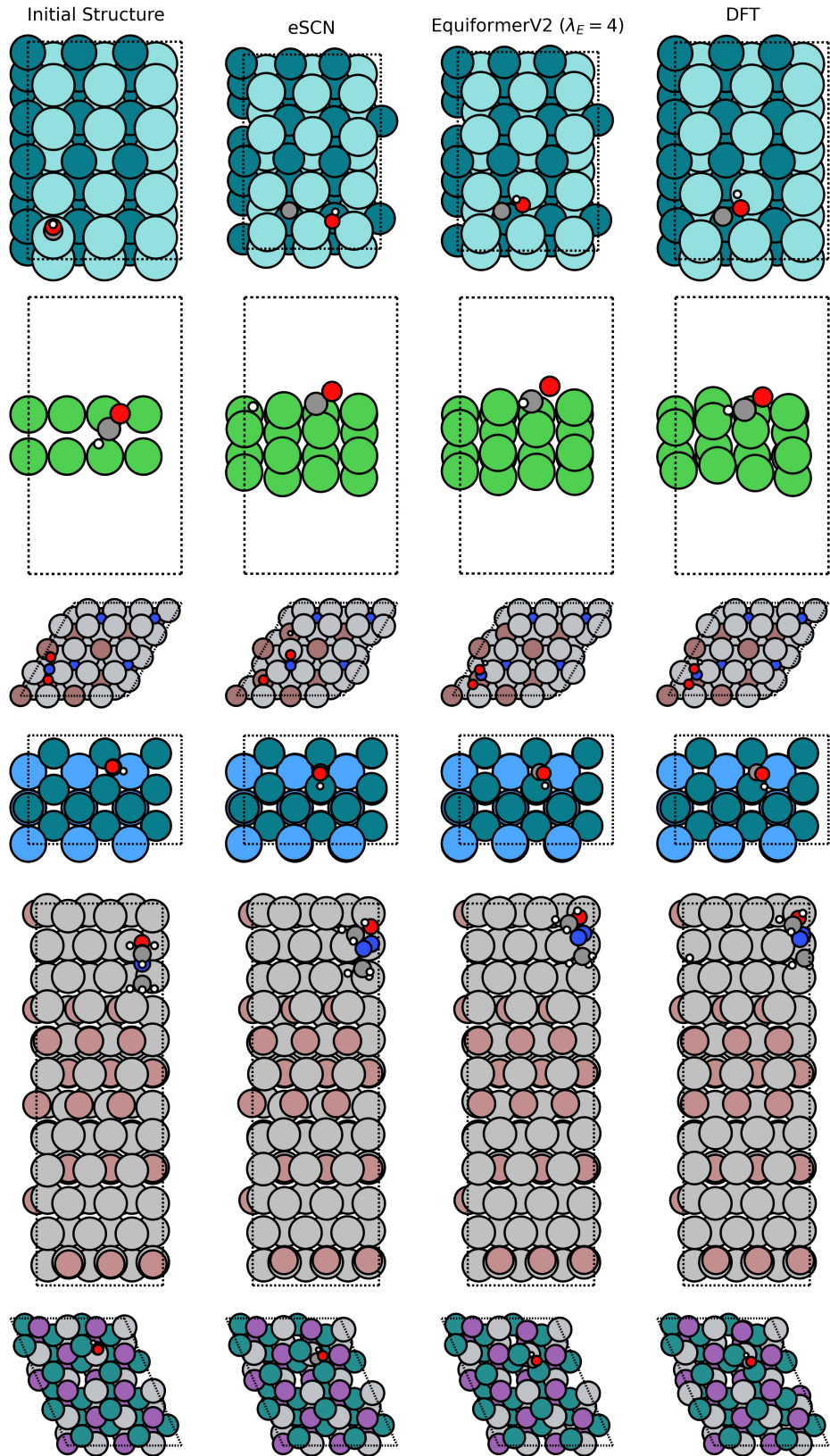


Figure 5: Qualitative examples of the initial configuration of an adsorbate on a catalyst surface (column 1), and corresponding relaxed configurations obtained from eSCN [18] (column 2), EquiformerV2 (column 3), and DFT (column 4). All examples are selected from the OC20-Dense dataset [10]. We show top-down views of each structure, with dashed lines showing the boundary of the unit cell repeating in the  $x$  and  $y$  directions.

## References

- [1] J. Gilmer, S. S. Schoenholz, P. F. Riley, O. Vinyals, and G. E. Dahl, “Neural message passing for quantum chemistry,” in *International Conference on Machine Learning (ICML)*, 2017. 1, 2
- [2] L. Zhang, J. Han, H. Wang, R. Car, and W. E, “Deep potential molecular dynamics: A scalable model with the accuracy of quantum mechanics,” *Phys. Rev. Lett.*, vol. 120, p. 143001, Apr 2018. 1
- [3] W. Jia, H. Wang, M. Chen, D. Lu, L. Lin, R. Car, W. E, and L. Zhang, “Pushing the limit of molecular dynamics with ab initio accuracy to 100 million atoms with machine learning,” in *Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis, SC ’20*, IEEE Press, 2020. 1
- [4] J. Gastegger, S. Giri, J. T. Margraf, and S. Günnemann, “Fast and uncertainty-aware directional message passing for non-equilibrium molecules,” in *Machine Learning for Molecules Workshop, NeurIPS*, 2020. 1, 3, 8, 9
- [5] S. Batzner, A. Musaelian, L. Sun, M. Geiger, J. P. Mailoa, M. Kornbluth, N. Molinari, T. E. Smidt, and B. Kozinsky, “E(3)-equivariant graph neural networks for data-efficient and accurate interatomic potentials,” *Nature Communications*, vol. 13, May 2022. 1, 2, 3, 4, 11
- [6] D. Lu, H. Wang, M. Chen, L. Lin, R. Car, W. E, W. Jia, and L. Zhang, “86 pflops deep potential molecular dynamics simulation of 100 million atoms with ab initio accuracy,” *Computer Physics Communications*, vol. 259, p. 107624, 2021. 1
- [7] O. T. Unke, M. Bogojeski, M. Gastegger, M. Geiger, T. Smidt, and K. R. Müller, “SE(3)-equivariant prediction of molecular wavefunctions and electronic densities,” in *Advances in Neural Information Processing Systems (NeurIPS)* (A. Beygelzimer, Y. Dauphin, P. Liang, and J. W. Vaughan, eds.), 2021. 1, 2
- [8] A. Sriram, A. Das, B. M. Wood, and C. L. Zitnick, “Towards training billion parameter graph neural networks for atomic simulations,” in *International Conference on Learning Representations (ICLR)*, 2022. 1, 3, 8
- [9] J. A. Rackers, L. Tecot, M. Geiger, and T. E. Smidt, “A recipe for cracking the quantum scaling limit with machine learned electron densities,” *Machine Learning: Science and Technology*, vol. 4, p. 015027, feb 2023. 1, 11
- [10] J. Lan, A. Palizhati, M. Shuaibi, B. M. Wood, B. Wander, A. Das, M. Uyttendaele, C. L. Zitnick, and Z. W. Ulissi, “AdsorbML: Accelerating adsorption energy calculations with machine learning,” *arXiv preprint arXiv:2211.16486*, 2022. 1, 2, 7, 8, 9, 15, 16
- [11] N. Thomas, T. E. Smidt, S. Kearnes, L. Yang, L. Li, K. Kohlhoff, and P. Riley, “Tensor field networks: Rotation- and translation-equivariant neural networks for 3d point clouds,” *arxiv preprint arXiv:1802.08219*, 2018. 1, 2
- [12] M. Weiler, M. Geiger, M. Welling, W. Boomsma, and T. Cohen, “3D Steerable CNNs: Learning Rotationally Equivariant Features in Volumetric Data,” in *Advances in Neural Information Processing Systems 32*, pp. 10402–10413, 2018. 1, 2, 4, 5
- [13] R. Kondor, Z. Lin, and S. Trivedi, “Clebsch–gordan nets: a fully fourier space spherical convolutional neural network,” in *Advances in Neural Information Processing Systems 32*, pp. 10117–10126, 2018. 1, 2
- [14] F. Fuchs, D. E. Worrall, V. Fischer, and M. Welling, “Se(3)-transformers: 3d roto-translation equivariant attention networks,” in *Advances in Neural Information Processing Systems (NeurIPS)*, 2020. 1, 2
- [15] J. Brandstetter, R. Hesselink, E. van der Pol, E. J. Bekkers, and M. Welling, “Geometric and physical quantities improve e(3) equivariant message passing,” in *International Conference on Learning Representations (ICLR)*, 2022. 1, 2
- [16] A. Musaelian, S. Batzner, A. Johansson, L. Sun, C. J. Owen, M. Kornbluth, and B. Kozinsky, “Learning local equivariant representations for large-scale atomistic dynamics,” *arxiv preprint arxiv:2204.05249*, 2022. 1, 2
- [17] Y.-L. Liao and T. Smidt, “Equiformer: Equivariant graph attention transformer for 3d atomistic graphs,” in *International Conference on Learning Representations (ICLR)*, 2023. 1, 2, 3, 4, 5, 8, 10, 13



- [18] S. Passaro and C. L. Zitnick, "Reducing SO(3) Convolutions to SO(2) for Efficient Equivariant GNNs," in *International Conference on Machine Learning (ICML)*, 2023. 1, 2, 3, 4, 5, 6, 7, 8, 10, 11, 12, 13, 15, 16
- [19] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, "Attention is all you need," in *Advances in Neural Information Processing Systems (NeurIPS)*, 2017. 1, 2, 3
- [20] S. Khan, M. Naseer, M. Hayat, S. W. Zamir, F. S. Khan, and M. Shah, "Transformers in vision: A survey," *arXiv preprint arxiv:2101.01169*, 2021. 1
- [21] N. Carion, F. Massa, G. Synnaeve, N. Usunier, A. Kirillov, and S. Zagoruyko, "End-to-end object detection with transformers," in *European Conference on Computer Vision (ECCV)*, 2020. 1
- [22] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, J. Uszkoreit, and N. Houlsby, "An image is worth 16x16 words: Transformers for image recognition at scale," in *International Conference on Learning Representations (ICLR)*, 2021. 1, 3
- [23] H. Touvron, M. Cord, M. Douze, F. Massa, A. Sablayrolles, and H. Jégou, "Training data-efficient image transformers & distillation through attention," *arXiv preprint arxiv:2012.12877*, 2020. 1
- [24] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," *arXiv preprint arxiv:1810.04805*, 2019. 1
- [25] T. Brown, B. Mann, N. Ryder, M. Subbiah, J. D. Kaplan, P. Dhariwal, A. Neelakantan, P. Shyam, G. Sastry, A. Askell, S. Agarwal, A. Herbert-Voss, G. Krueger, T. Henighan, R. Child, A. Ramesh, D. Ziegler, J. Wu, C. Winter, C. Hesse, M. Chen, E. Sigler, M. Litwin, S. Gray, B. Chess, J. Clark, C. Berner, S. McCandlish, A. Radford, I. Sutskever, and D. Amodei, "Language models are few-shot learners," in *Advances in Neural Information Processing Systems (NeurIPS)*, 2020. 1
- [26] V. P. Dwivedi and X. Bresson, "A generalization of transformer networks to graphs," *arXiv preprint arxiv:2012.09699*, 2020. 1
- [27] D. Kreuzer, D. Beaini, W. L. Hamilton, V. Létourneau, and P. Tossou, "Rethinking graph transformers with spectral attention," in *Advances in Neural Information Processing Systems (NeurIPS)*, 2021. 1
- [28] C. Ying, T. Cai, S. Luo, S. Zheng, G. Ke, D. He, Y. Shen, and T.-Y. Liu, "Do transformers really perform badly for graph representation?," in *Advances in Neural Information Processing Systems (NeurIPS)*, 2021. 1
- [29] Y. Shi, S. Zheng, G. Ke, Y. Shen, J. You, J. He, S. Luo, C. Liu, D. He, and T.-Y. Liu, "Benchmarking graphormer on large-scale molecular modeling datasets," *arXiv preprint arxiv:2203.04810*, 2022. 1
- [30] L. Chanussot\*, A. Das\*, S. Goyal\*, T. Lavril\*, M. Shuaibi\*, M. Riviere, K. Tran, J. Heras-Domingo, C. Ho, W. Hu, A. Palizhati, A. Sriram, B. Wood, J. Yoon, D. Parikh, C. L. Zitnick, and Z. Ulissi, "Open catalyst 2020 (oc20) dataset and community challenges," *ACS Catalysis*, 2021. 2, 3, 4, 6, 7, 9, 14
- [31] B. K. Miller, M. Geiger, T. E. Smidt, and F. Noé, "Relevance of rotationally equivariant convolutions for predicting molecular properties," *arXiv preprint arxiv:2008.08461*, 2020. 2
- [32] R. J. L. Townshend, B. Townshend, S. Eismann, and R. O. Dror, "Geometric prediction: Moving beyond scalars," *arXiv preprint arxiv:2006.14163*, 2020. 2
- [33] B. Jing, S. Eismann, P. Suriana, R. J. L. Townshend, and R. Dror, "Learning from protein structure with geometric vector perceptrons," in *International Conference on Learning Representations (ICLR)*, 2021. 2
- [34] K. T. Schütt, O. T. Unke, and M. Gastegger, "Equivariant message passing for the prediction of tensorial properties and molecular spectra," in *International Conference on Machine Learning (ICML)*, 2021. 2, 4, 9
- [35] V. G. Satorras, E. Hoogeboom, and M. Welling, "E(n) equivariant graph neural networks," in *International Conference on Machine Learning (ICML)*, 2021. 2

- [36] P. Thölke and G. D. Fabritius, “Equivariant transformers for neural network based molecular potentials,” in *International Conference on Learning Representations (ICLR)*, 2022. 2
- [37] T. Le, F. Noé, and D.-A. Clevert, “Equivariant graph attention networks for molecular property prediction,” *arXiv preprint arXiv:2202.09891*, 2022. 2
- [38] I. Batatia, D. P. Kovacs, G. N. C. Simm, C. Ortner, and G. Csanyi, “MACE: Higher order equivariant message passing neural networks for fast and accurate force fields,” in *Advances in Neural Information Processing Systems (NeurIPS)*, 2022. 2, 13
- [39] A. Sanchez-Gonzalez, J. Godwin, T. Pfaff, R. Ying, J. Leskovec, and P. W. Battaglia, “Learning to simulate complex physics with graph networks,” in *International Conference on Machine Learning (ICML)*, 2020. 2
- [40] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, “Dropout: A simple way to prevent neural networks from overfitting,” *Journal of Machine Learning Research*, vol. 15, no. 56, pp. 1929–1958, 2014. 2, 4, 13
- [41] G. Huang, Y. Sun, Z. Liu, D. Sedra, and K. Q. Weinberger, “Deep networks with stochastic depth,” in *European Conference on Computer Vision (ECCV)*, 2016. 2, 13
- [42] L. Zitnick, A. Das, A. Kolluru, J. Lan, M. Shuaibi, A. Sriram, Z. Ulissi, and B. Wood, “Spherical channels for modeling atomic interactions,” in *Advances in Neural Information Processing Systems (NeurIPS)*, 2022. 2, 3, 5, 7, 8, 9
- [43] K. T. Schütt, P.-J. Kindermans, H. E. Sauceda, S. Chmiela, A. Tkatchenko, and K.-R. Müller, “SchNet: A continuous-filter convolutional neural network for modeling quantum interactions,” in *Advances in Neural Information Processing Systems (NeurIPS)*, 2017. 3, 6, 8, 9
- [44] T. Xie and J. C. Grossman, “Crystal graph convolutional neural networks for an accurate and interpretable prediction of material properties,” *Physical Review Letters*, 2018. 3, 8
- [45] O. T. Unke and M. Meuwly, “PhysNet: A neural network for predicting energies, forces, dipole moments, and partial charges,” *Journal of Chemical Theory and Computation*, vol. 15, pp. 3678–3693, may 2019. 3
- [46] J. Gastegger, J. Groß, and S. Günnemann, “Directional message passing for molecular graphs,” in *International Conference on Learning Representations (ICLR)*, 2020. 3
- [47] Z. Qiao, M. Welborn, A. Anandkumar, F. R. Manby, and T. F. Miller, “OrbNet: Deep learning for quantum chemistry using symmetry-adapted atomic-orbital features,” *The Journal of Chemical Physics*, 2020. 3
- [48] Y. Liu, L. Wang, M. Liu, Y. Lin, X. Zhang, B. Oztekin, and S. Ji, “Spherical message passing for 3d molecular graphs,” in *International Conference on Learning Representations (ICLR)*, 2022. 3
- [49] M. Shuaibi, A. Kolluru, A. Das, A. Grover, A. Sriram, Z. Ulissi, and C. L. Zitnick, “Rotation invariant graph neural networks using spin convolutions,” *arxiv preprint arxiv:2106.09575*, 2021. 3, 8
- [50] J. Klicpera, F. Becker, and S. Günnemann, “Gemnet: Universal directional graph neural networks for molecules,” in *Advances in Neural Information Processing Systems (NeurIPS)*, 2021. 3, 8
- [51] J. Gastegger, M. Shuaibi, A. Sriram, S. Günnemann, Z. Ulissi, C. L. Zitnick, and A. Das, “GemNet-OC: Developing Graph Neural Networks for Large and Diverse Molecular Simulation Datasets,” *Transactions on Machine Learning Research (TMLR)*, 2022. 3, 8, 9
- [52] J. L. Ba, J. R. Kiros, and G. E. Hinton, “Layer normalization,” *arxiv preprint arxiv:1607.06450*, 2016. 4, 5
- [53] P. Veličković, G. Cucurull, A. Casanova, A. Romero, P. Liò, and Y. Bengio, “Graph attention networks,” in *International Conference on Learning Representations (ICLR)*, 2018. 4
- [54] G. Huang, Y. Sun, Z. Liu, D. Sedra, and K. Q. Weinberger, “Deep networks with stochastic depth,” in *European Conference on Computer Vision (ECCV)*, 2016. 4
- [55] M. Geiger, T. Smidt, A. M., B. K. Miller, W. Boomsma, B. Dice, K. Lapchevskyi, M. Weiler, M. Tyszkiewicz, S. Batzner, D. Madiseti, M. Uhrin, J. Frellsen, N. Jung, S. Sanborn, M. Wen, J. Rackers, M. Rød, and M. Bailey, “e3nn/e3nn: 2022-04-13,” Apr. 2022. 4, 10, 11

- [56] S. Chmiela, A. Tkatchenko, H. E. Sauceda, I. Poltavsky, K. T. Schütt, and K.-R. Müller, “Machine learning of accurate energy-conserving molecular force fields,” *Science Advances*, vol. 3, no. 5, p. e1603015, 2017. 4, 9
- [57] K. T. Schütt, F. Arbabzadah, S. Chmiela, K. R. Müller, and A. Tkatchenko, “Quantum-chemical insights from deep tensor neural networks,” *Nature Communications*, vol. 8, jan 2017. 4, 9
- [58] S. Chmiela, H. E. Sauceda, K.-R. Müller, and A. Tkatchenko, “Towards exact molecular dynamics simulations with machine-learned force fields,” *Nature Communications*, vol. 9, sep 2018. 4, 9
- [59] T. S. Cohen, M. Geiger, J. Köhler, and M. Welling, “Spherical CNNs,” in *International Conference on Learning Representations (ICLR)*, 2018. 5
- [60] S. Elfving, E. Uchibe, and K. Doya, “Sigmoid-weighted linear units for neural network function approximation in reinforcement learning,” *arXiv preprint arXiv:1702.03118*, 2017. 5
- [61] P. Ramachandran, B. Zoph, and Q. V. Le, “Searching for activation functions,” *arXiv preprint arXiv:1710.05941*, 2017. 5
- [62] B. Hammer, L. B. Hansen, and J. K. Nørskov, “Improved adsorption energetics within density-functional theory using revised perdue-burke-ernzerhof functionals,” *Phys. Rev. B*, 1999. 6
- [63] W. Hu, M. Shuaibi, A. Das, S. Goyal, A. Sriram, J. Leskovec, D. Parikh, and C. L. Zitnick, “Forcenet: A graph neural network for large-scale quantum calculations,” *arxiv preprint arxiv:2103.01436*, 2021. 8
- [64] J. H. Lee, P. Yadollahpour, A. Watkins, N. C. Frey, A. Leaver-Fay, S. Ra, K. Cho, V. Gligorijevic, A. Regev, and R. Bonneau, “Equifold: Protein structure prediction with a novel coarse-grained structure representation,” *bioRxiv*, 2022. 9
- [65] R. Ramakrishnan, P. O. Dral, M. Rupp, and O. A. von Lilienfeld, “Quantum chemistry structures and properties of 134 kilo molecules,” *Scientific Data*, vol. 1, 2014. 9
- [66] L. Ruddigkeit, R. van Deursen, L. C. Blum, and J.-L. Reymond, “Enumeration of 166 billion organic small molecules in the chemical universe database gdb-17,” *Journal of Chemical Information and Modeling*, vol. 52, no. 11, pp. 2864–2875, 2012. PMID: 23088335. 9
- [67] M. Nakata and T. Shimazaki, “Pubchemqc project: A large-scale first-principles electronic structure database for data-driven chemistry,” *Journal of chemical information and modeling*, vol. 57 6, pp. 1300–1308, 2017. 9
- [68] S. Zaidi, M. Schaarschmidt, J. Martens, H. Kim, Y. W. Teh, A. Sanchez-Gonzalez, P. Battaglia, R. Pascanu, and J. Godwin, “Pre-training via denoising for molecular property prediction,” in *International Conference on Learning Representations (ICLR)*, 2023. 9
- [69] A. Zee, *Group Theory in a Nutshell for Physicists*. USA: Princeton University Press, 2016. 10
- [70] M. S. Dresselhaus, G. Dresselhaus, and A. Jorio, *Group theory*. Berlin, Germany: Springer, 2008 ed., Mar. 2007. 10
- [71] N. Frey, R. Soklaski, S. Axelrod, S. Samsi, R. Gomez-Bombarelli, C. Coley, and V. Gadepally, “Neural scaling of deep chemical models,” *ChemRxiv*, 2022. 11
- [72] D. E. Worrall, S. J. Garbin, D. Turmukhambetov, and G. J. Brostow, “Harmonic networks: Deep translation and rotation equivariance,” *arxiv preprint arxiv:1612.04642*, 2016. 12

724