Spherical Channels for Modeling Atomic Interactions, Zitnick et al.

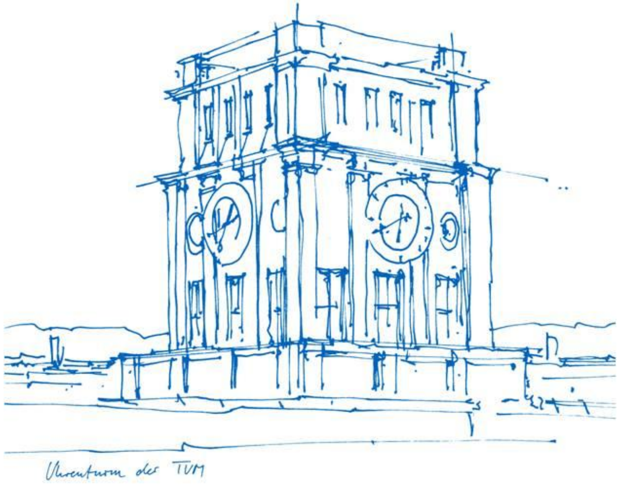
Jean-Pasqual Sindermann

Technische Universität München

Selected Topics in Deep Learning –

Equivariance & Dynamics

Munich, 17. June 2025



Structure

- Introduction
- Spherical Harmonics
- Message Passing in Spherical Channel Networks (SCN)
- Energy and Force Prediction
- Equivariance Considerations
- Paper results
- Additional discussion points

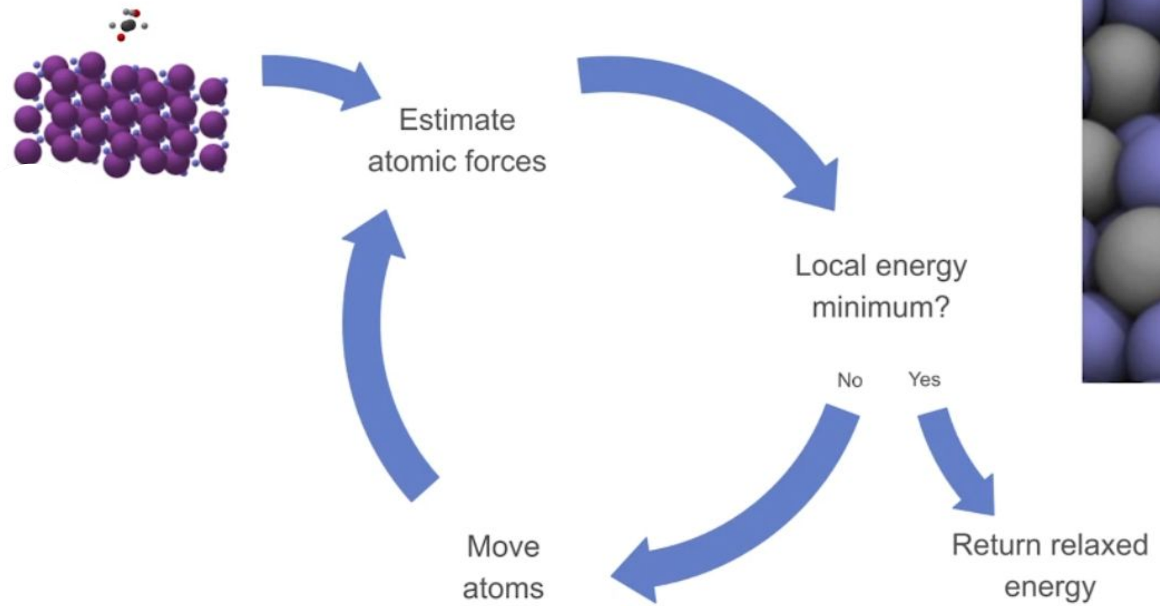
Introduction

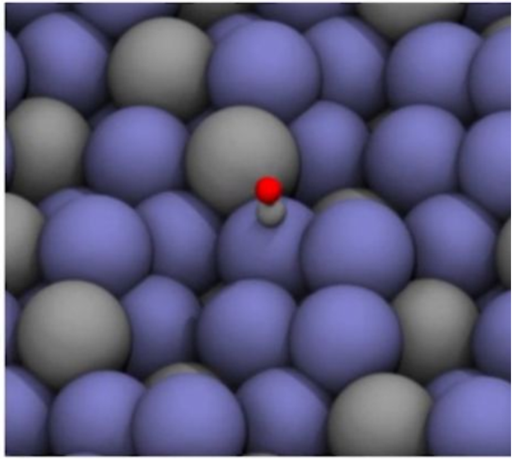
Lets develop a new catalyst/chemical/material,
Ideally to tackle the climate change!

How do we do this?

Introduction

Atomic Relaxation





Introduction

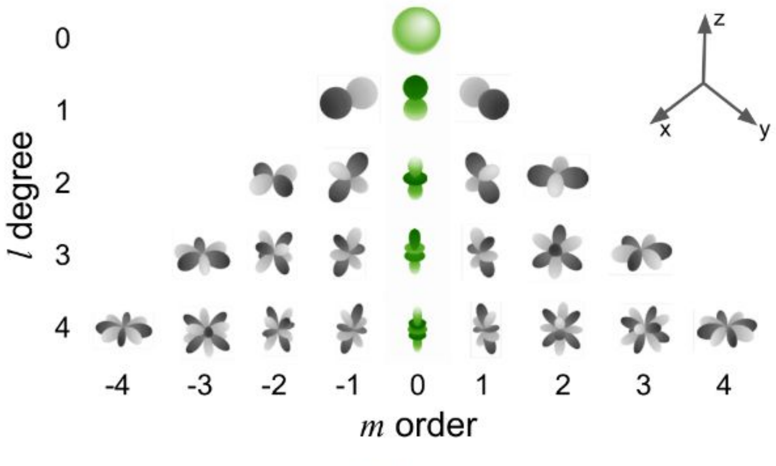
- Prediction of Atomic Energy and Forces is key
- Mathematical perfect description: Density Functional Theory
  - Complicated to calculate/approximate
  - Takes ages
- Can we speed things up?

Introduction

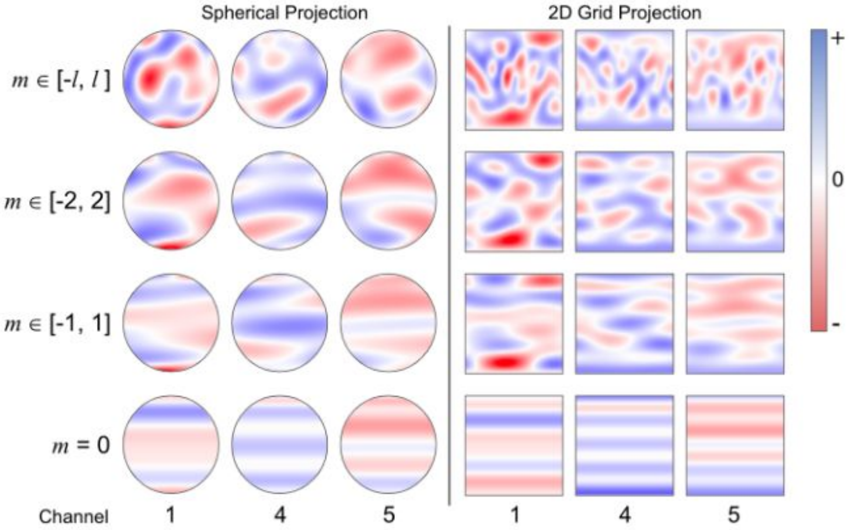
- Prediction of Atomic Energy and Forces is key
- Mathematical perfect description: Density Functional Theory
  - Complicated to calculate/approximate
  - Takes ages
- Can we speed things up?
  - Solution to everything: Machine Learning (and Neural Networks)

- Energy: Invariant to Rotation (and other Transformations)
- Forces: Equivariant to Rotation (and other Transformations)
  - Per Atom Forces

Spherical Harmonics



(a)



(b)

Spherical Harmonics

$$Y_{0,0}(\theta,\phi) = \sqrt{\frac{1}{4\pi}}$$

$$Y_{1,-1}(\theta,\phi) = \sqrt{\frac{3}{4\pi}} \sin\phi \sin\theta$$

$$Y_{1,0}(\theta,\phi) = \sqrt{\frac{3}{4\pi}} \cos\theta$$

$$Y_{1,1}(\theta,\phi) = \sqrt{\frac{3}{4\pi}} \cos\phi \sin\theta$$

$$Y_{2,-2}(\theta,\phi) = \sqrt{\frac{15}{16\pi}} \sin(2\phi) \sin^2\theta$$

$$Y_{2,-1}(\theta,\phi) = \sqrt{\frac{15}{4\pi}} \sin\phi \sin\theta \cos\theta$$

$$Y_{2,0}(\theta,\phi) = \sqrt{\frac{5}{16\pi}} (3\cos^2\theta - 1)$$

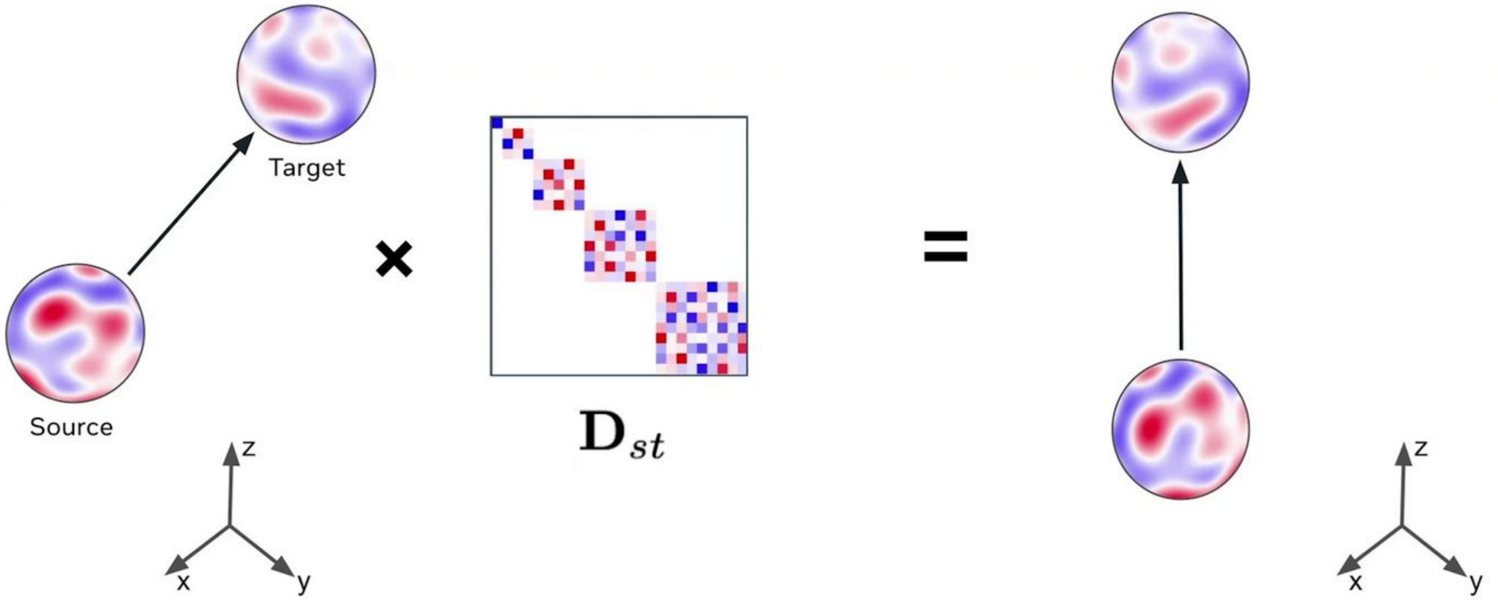
$$Y_{2,1}(\theta,\phi) = \sqrt{\frac{15}{4\pi}} \cos\phi \sin\theta \cos\theta$$

$$Y_{2,2}(\theta,\phi) = \sqrt{\frac{15}{16\pi}} \cos(2\phi) \sin^2\theta$$

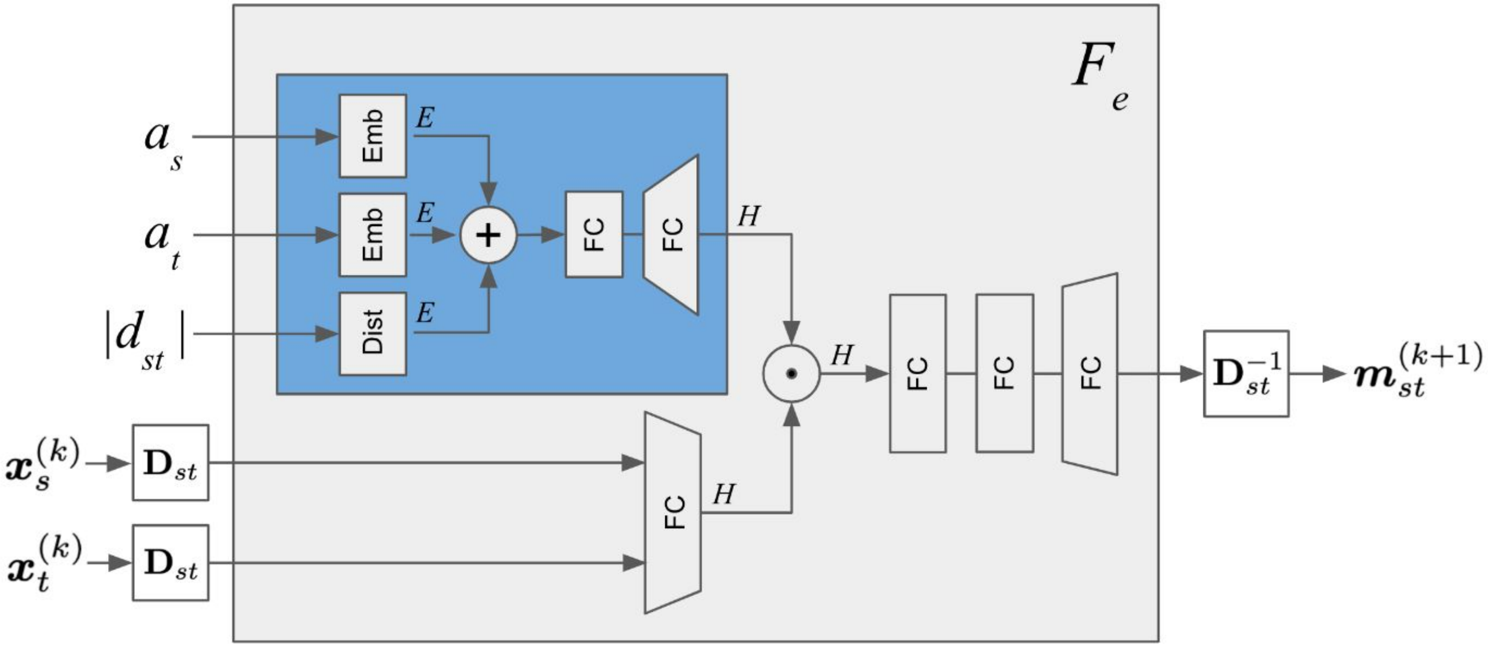
Spherical Channel Networks

- Graph Neural Networks
  - Atoms are Nodes
  - Edges are Neighbours
- Embeddings are Coefficients of Spherical Harmonics
- Each Atom encodes a function on a sphere
  - with multiple channels like CNNs
- Messages calculated per Edge
  - rotate edge to align with Z-Axis
  - compute message value
  - rotate back

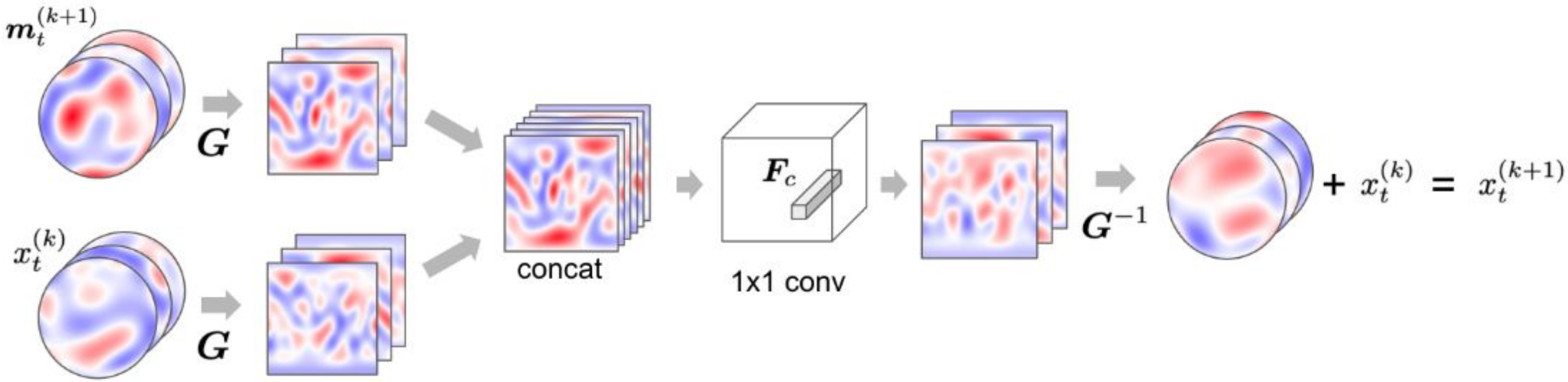
Message Passing



Message Passing

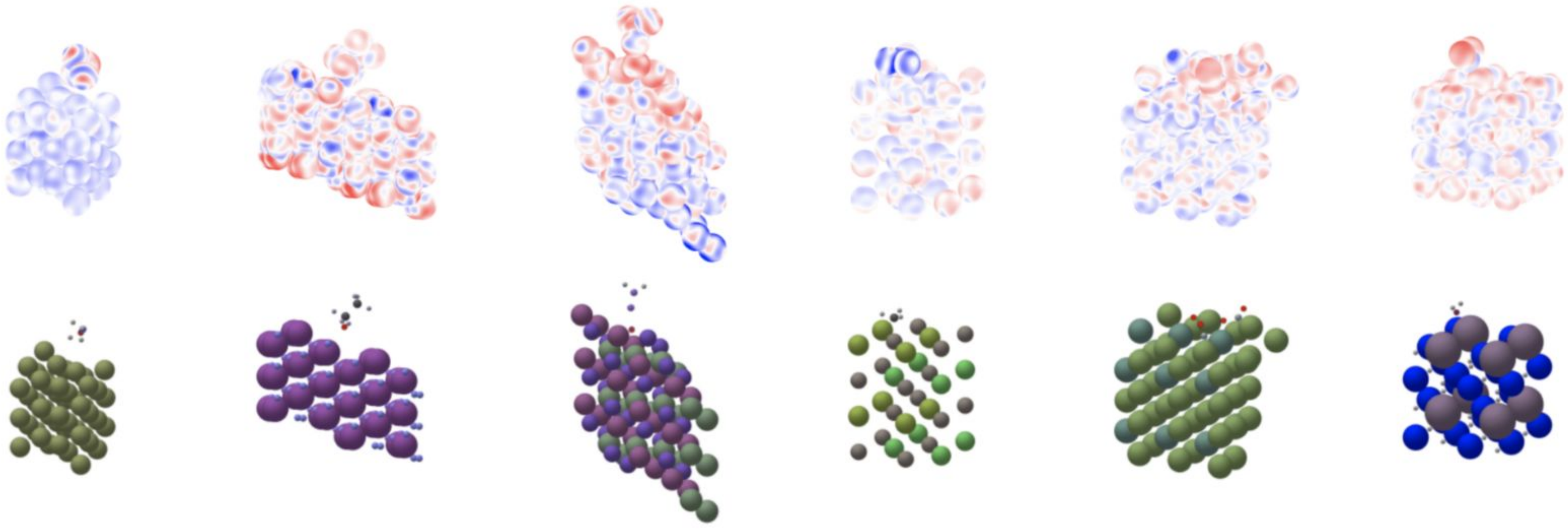


Message Passing



- $m_t^{(k+1)} = \sum_s m_{st}^{(k+1)}$
- 3 layer 1x1 CNN
- SiLU activation per layer

Energy and Force Prediction



Energy and Force Prediction

$$E = \sum_i \int \boldsymbol{F}_{energy} \left(s_i^{(K)}(\hat{\boldsymbol{r}}) \right) d\hat{\boldsymbol{r}},$$

- F_{energy} is three layer FC with SiLU activations

$$\boldsymbol{f}_i = \int \hat{\boldsymbol{r}} \boldsymbol{F}_{force} \left(s_i^{(K)}(\hat{\boldsymbol{r}}) \right) d\hat{\boldsymbol{r}},$$

- F_{force} is three layer FC

Equivariance

Equivariance

How long does this takes you to read?

Equivariance

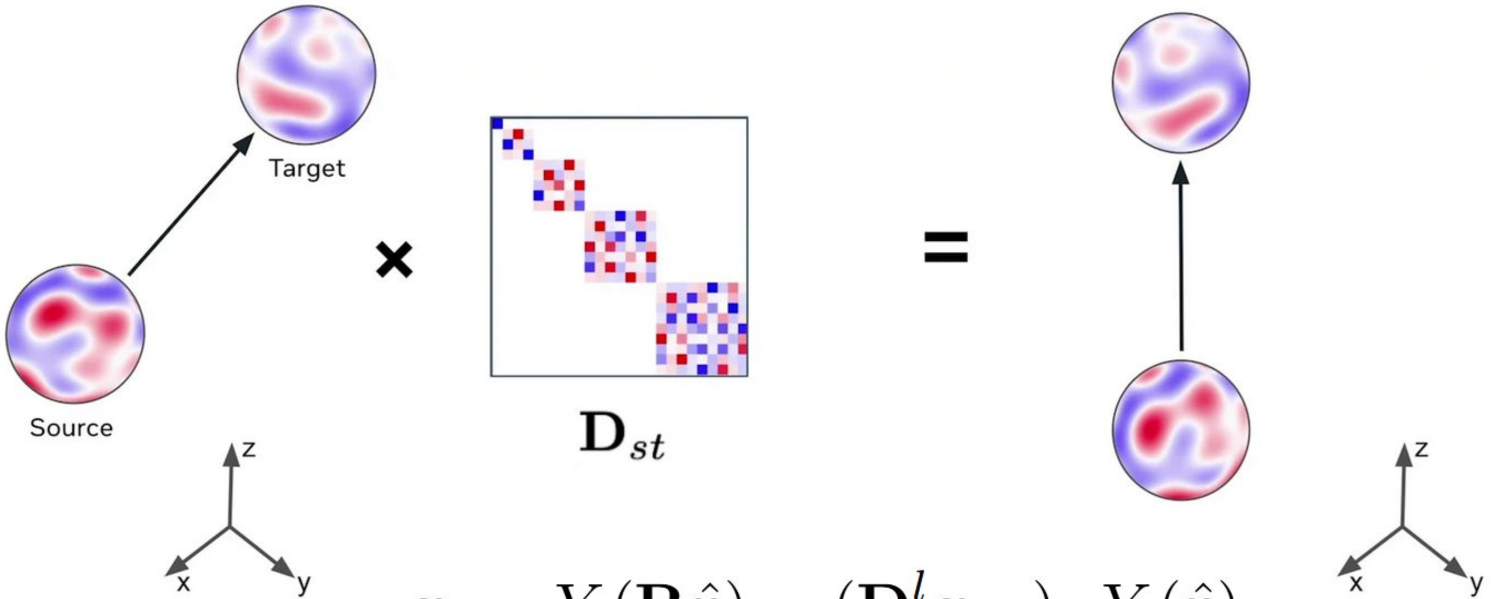
Are SCNs equivariant?

Why or why not?

Equivariance

- Embeddings are coefficients of Spherical Harmonics
    - Rotation can be achieved with Wigner D-matrices

Equivariance

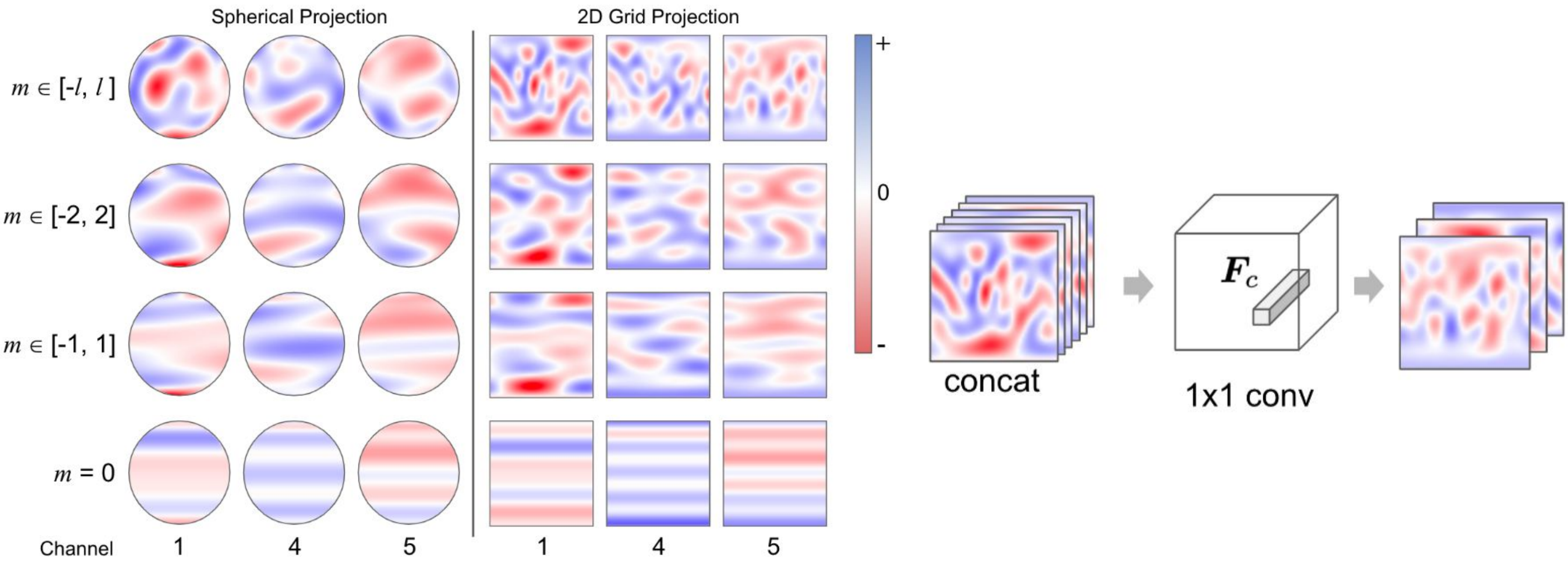


$$\boldsymbol{x}_{ilc} \cdot Y_l(\mathbf{R}\hat{\boldsymbol{r}}) = (\mathbf{D}^l \boldsymbol{x}_{ilc}) \cdot Y_l(\hat{\boldsymbol{r}})$$

Equivariance

- Embeddings are coefficients of Spherical Harmonics
    - Rotation can be achieved with Wigner D-matrices
- Message Passing
    - Edges are rotated to align with Z-Axis
        - Roll is unknown and randomly sampled in practice
        - Only harmonics of order $m=0$ are strictly equivariant
        - Possible: Average over different angles similar to G-CNNs

Equivariance



Equivariance

- Embeddings are coefficients of Spherical Harmonics
  - Rotation can be achieved with Wigner D-matrices
- Message Passing
  - Edges are rotated to align with Z-Axis
  - Projection onto 2D-Grid and non-linearities
    - non-linearities may add frequencies higher than degree of L
    - grid-resolution / sampling rate has to be high enough

Equivariance

- Embeddings are coefficients of Spherical Harmonics
  - Rotation can be achieved with Wigner D-matrices
- Message Passing
  - Edges are rotated to align with Z-Axis
  - Projection onto 2D-Grid and non-linearities
- Energy and Force prediction
  - Fully connected layers with activation
  - Integrated (numerically) over the sphere
    - neglectable effect

Equivariance

Model	Activation	Energy MAE [meV]	MAD [meV]	Force MAE Force [meV/Å]	MAD [meV/Å]
SCN 3-Layer	SiLU	302	0.95	24.6	0.05

Rotational Error of Energy/Force Prediction

Model	Activation	% MAD
SCN 1-Layer	Linear	0.0%
SCN 1-Layer	SiLU	4.6%
SCN 2-Layer	SiLU	6.7%
SCN 1-Layer	ReLU	4.2%
SCN 2-Layer	ReLU	6.8%

Rotational Error of Activations

Equivariance

		Energy				
			No AMP		AMP	
Model		**MAE [meV] ↓**	**MAD [meV] ↓**	**% Error ↓**	**MAD [meV] ↓**	**% Error ↓**
SCN	$m = 0$	307	7.9	2.6%	12.4	4.0%
SCN	$m \in [-1, 1]$	302	39.1	12.9%	42.8	14.2%
SCN	$m \in [-1, 1]$, No 1x1 conv	313	30.3	9.7%	37.3	11.9%
SCN	$m \in [-1, 1]$, 4-tap	294	6.9	2.3%	19.4	6.6%

		Force				
			No AMP		AMP	
Model		**MAE [meV/Å] ↓**	**MAD [meV/Å] ↓**	**% Error ↓**	**MAD [meV/Å] ↓**	**% Error ↓**
SCN	$m = 0$	26.5	0.4	1.5%	0.9	3.6%
SCN	$m \in [-1, 1]$	24.6	3.1	12.6%	4.4	17.9%
SCN	$m \in [-1, 1]$, No 1x1 conv	26.2	0.5	2.0%	0.6	2.2%
SCN	$m \in [-1, 1]$, 4-tap	23.1	0.4	1.7%	2.6	11.3%

Rotational Error for SCN

Results

OC20 2M Validation

Model	L	# layers	H	# batch	Samples / GPU sec.	S2EF Energy MAE [meV] ↓	Force MAE [meV/Å] ↓	Force Cos ↑	EFwT [%] ↑	IS2RE Energy MAE [meV] ↓	EwT [%] ↑
Median											
SchNet [38]						1400	78.3	0.109	0.00	-	-
DimeNet++ [12]						805	65.7	0.217	0.01	-	-
SpinConv [41]						406	36.2	0.479	0.13	-	-
GemNet-dT [14]					25.8	358	29.5	0.557	0.61	438	-
GemNet-OC [15]					18.3	286	25.7	0.598	1.06	407	-
SCN $m = 0$	6	16	1024	96	7.4	300	25.7	0.600	0.96	394	9.6
SCN $m = 0$	8	16	1024	64	4.8	296	25.3	0.608	1.01	389	9.9
SCN	6	16	1024	96	5.9	287	22.8	0.623	1.22	371	10.5
SCN	8	16	1024	96	3.5	**283**	22.7	0.627	1.22	**364**	**11.3**
SCN 4-tap	6	16	1024	64	2.6	**282**	**22.2**	**0.648**	1.37	378	10.7
SCN 4-tap 2-band	6	16	1024	64	2.3	**279**	**21.9**	**0.650**	**1.46**	373	**11.0**

Results on the OC20 2M training dataset

Results

OC20 Test

Model	#Params	Train time	S2EF Energy MAE meV ↓	S2EF Force MAE [meV/Å] ↓	S2EF Force Cos ↑	S2EF EFwT [%] ↑	IS2RS AFbT [%] ↑	IS2RS ADwT [%] ↑	IS2RE Energy MAE meV ↓
Median	–		2258	84.4	0.016	0.01	-	-	-
Train OC20 All									
SchNet [38, 6]	9.1M	194d	540	54.7	0.302	0.00	-	14.4	764
PaiNN [36]	20.1M	67d	341	33.1	0.491	0.46	11.7	48.5	471
DimeNet++-L-F+E [12, 6]	10.7M	1600d	480	31.3	0.544	0.00	21.7	51.7	559
SpinConv (direct-forces) [41]	8.5M	275d	336	29.7	0.539	0.45	16.7	53.6	437
GemNet-dT [14]	32M	492d	292	24.2	0.616	1.20	27.6	58.7	400
GemNet-OC [15]	39M	336d	**233**	20.7	0.666	2.50	35.3	60.3	355
SCN *L*=8 *K*=20	271M	645d	244	**17.7**	**0.687**	**2.59**	**40.3**	**67.1**	**330**
Train OC20 All + MD									
GemNet-OC-L-E [15]	56M	640d	**230**	21.0	0.665	2.80	-	-	-
GemNet-OC-L-F [15]	216M	765d	241	19.0	0.691	**2.97**	40.6	60.4	-
GemNet-OC-L-F+E [15]	-	-	-	-	-	-	-	-	348
SCN *L*=6 *K*=16 4-tap 2-band	168M	414d	**228**	17.8	**0.696**	**2.95**	**43.3**	64.9	328
SCN *L*=8 *K*=20	271M	1280d	237	**17.2**	**0.698**	2.89	**43.6**	**67.5**	**321**

Comparison of SCN to existing GNN models when trained on the All or All+MD datasets.

Results

BENEFITS:

- State-of-the-art performance
- Very large expressive capacity and sample efficient
- One model for Energy and Force prediction

LIMITATIONS:

- Overfitting with small data
- Very slow for enforced energy conservation
- Limited use of $L > 8$
- Equivariance challenges for $|m| > 2$

Questions and Discussion?

Discussion

What are the differences to TFNs?

- both rely on Spherical Harmonics
- TFN constrain architecture (filters, non-linearities, etc.) to enforce equivariance
- TFN uses harmonics as filters for pointwise convolutions
- less/no constraints for SCNs as they apply non-linearities on **rotated embeddings** and projected 2D grid

Discussion

How are Edge Distances embedded?

- 1D Array of evenly spaced Gaussian Radial Basis Functions
  - every 0.02 Å from 0 to 8 Å with σ = 0.04
- plus FC layers

$$[\mathrm{RBF}_1(d_{ij}), \mathrm{RBF}_2(d_{ij}), \ldots, \mathrm{RBF}_K(d_{ij})]$$

$$\mathrm{RBF}_k(d) = \exp\left(-\gamma(d - \mu_k)^2\right)$$

Thank you for your attention!

References

1. Paper used for information and graphics: Spherical Channel Networks for Modeling Atomic Interactions, Zitnick et al. https://arxiv.org/pdf/2006.10503
2. Paper presentation used for additional graphics: Spherical Channel Networks for Modeling Atomic Interactions, Zitnick et al., NeurIPS 2022 slideslive.com/38989937/spherical-channels-for-modeling-atomic-interactions