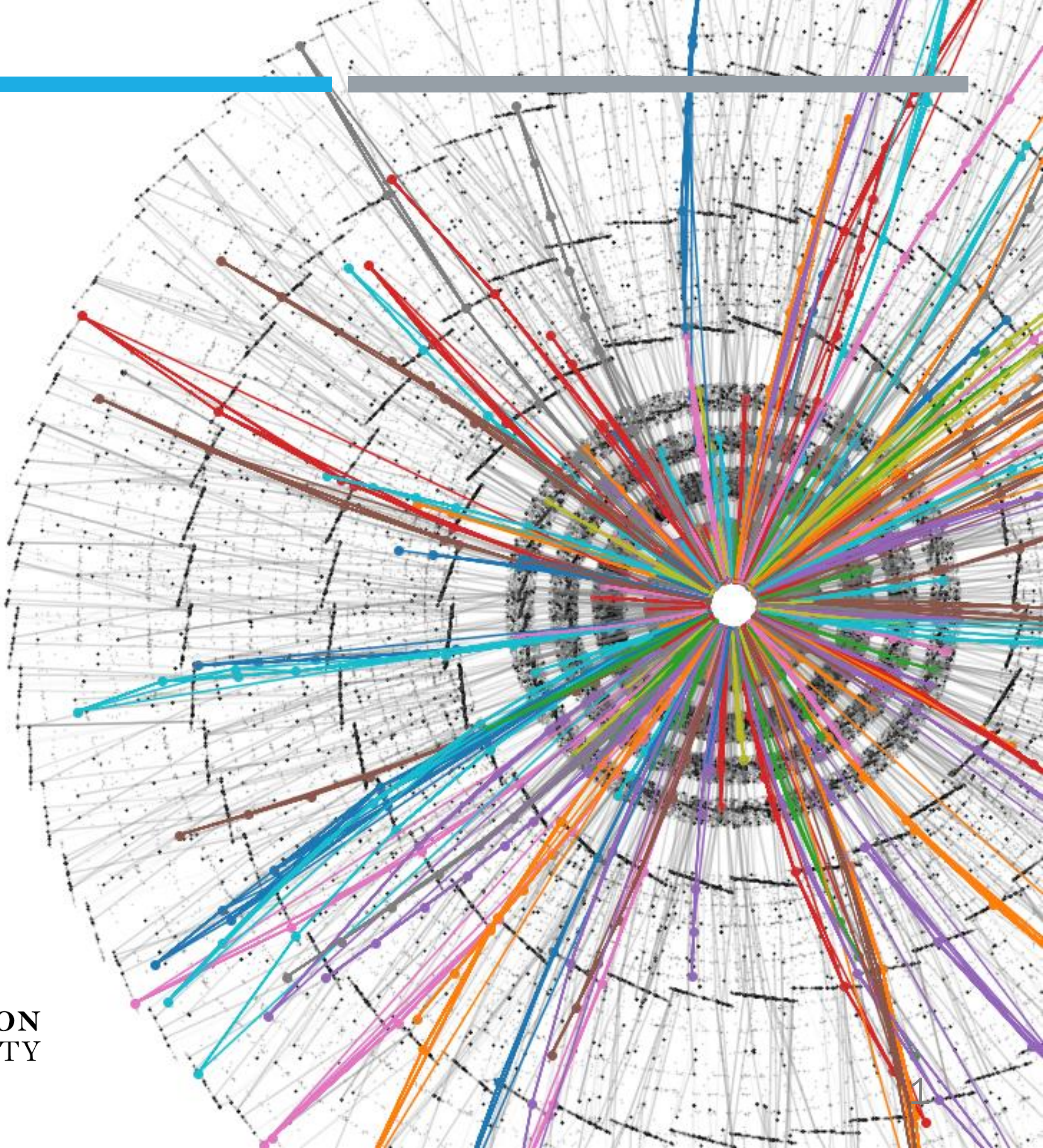
EQUIVARIANT GRAPH NEURAL NETWORKS FOR HIGH ENERGY PHYSICS

CONNECTING THE DOTS MINI-WORKSHOP
3RD JUNE, 2022, PRINCETON

DANIEL MURNANE & SAVANNAH THAIS
ON BEHALF OF THE EXATRKX PROJECT

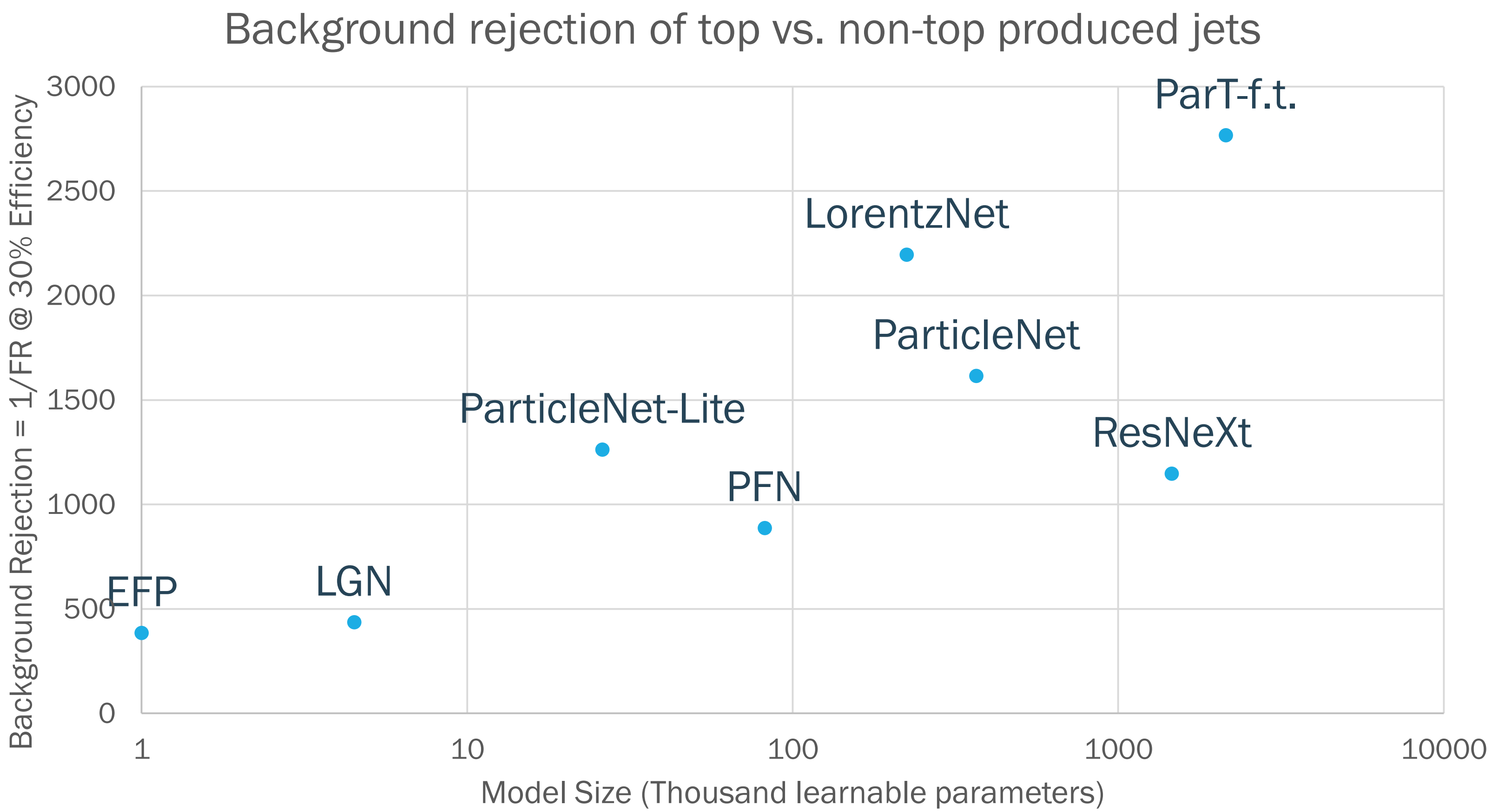
ExaTrkX · BERKELEY LAB · PRINCETON UNIVERSITY



GOALS

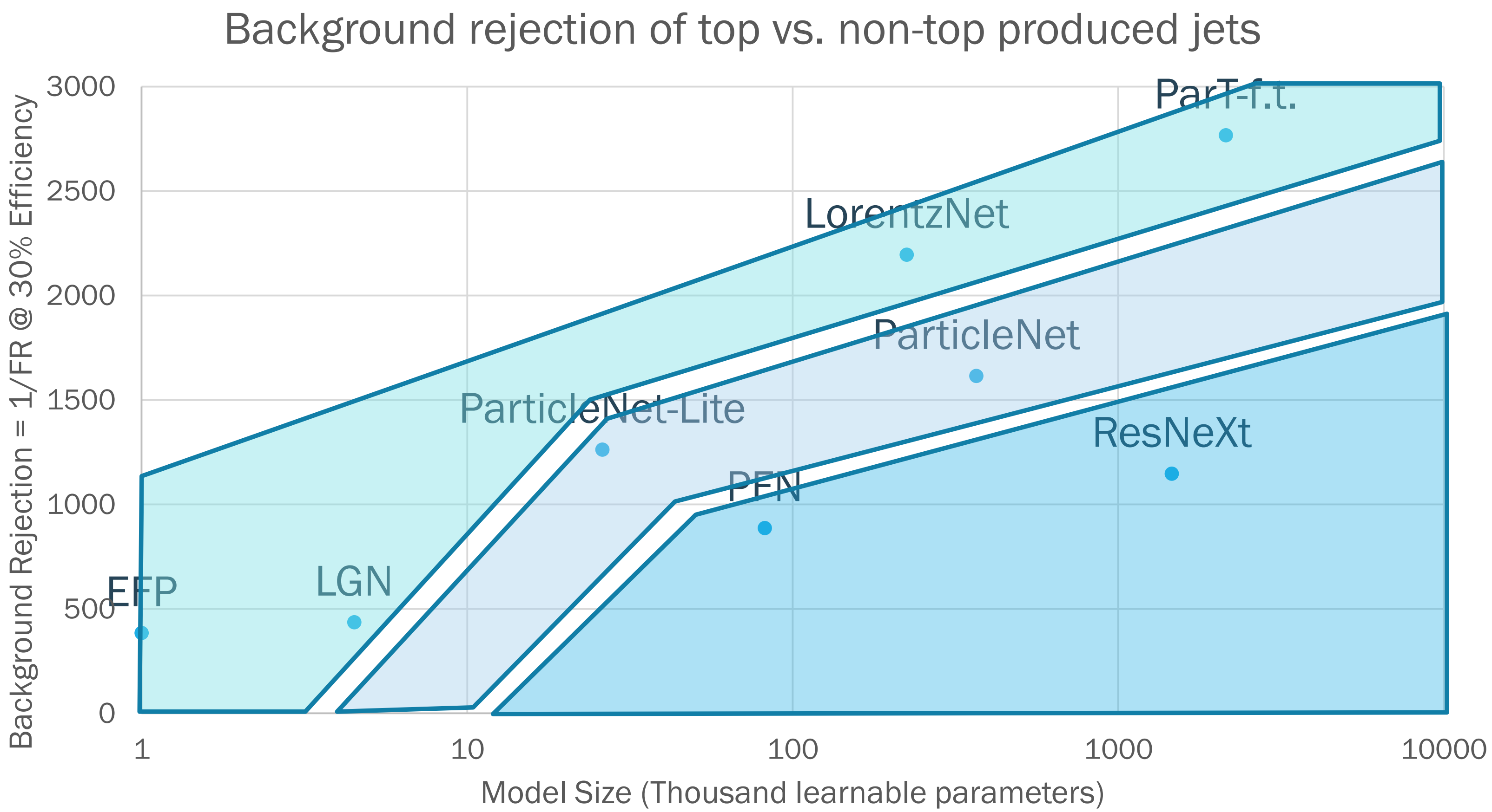
- Show that Equivariant GNNs are one of several highly active fields in physics-informed ML for HEP
- Show a concrete example of how to implement equivariance
- **Avoid group theory**
- Review the landscape of EquiGNNs
- Suggest some benefits of EquiGNNs
- Wonder aloud about how to incorporate them into track reconstruction
- Solicit feedback and discussion

WHY EQUI-GNN: NAIVELY IMPROVING MODEL PERFORMANCE



Would love to add LundNet-5 and JEDI-net to this plot, but don't have apples-to-apples rejection rate

WHY EQUI-GNN: NAIVELY IMPROVING MODEL PERFORMANCE



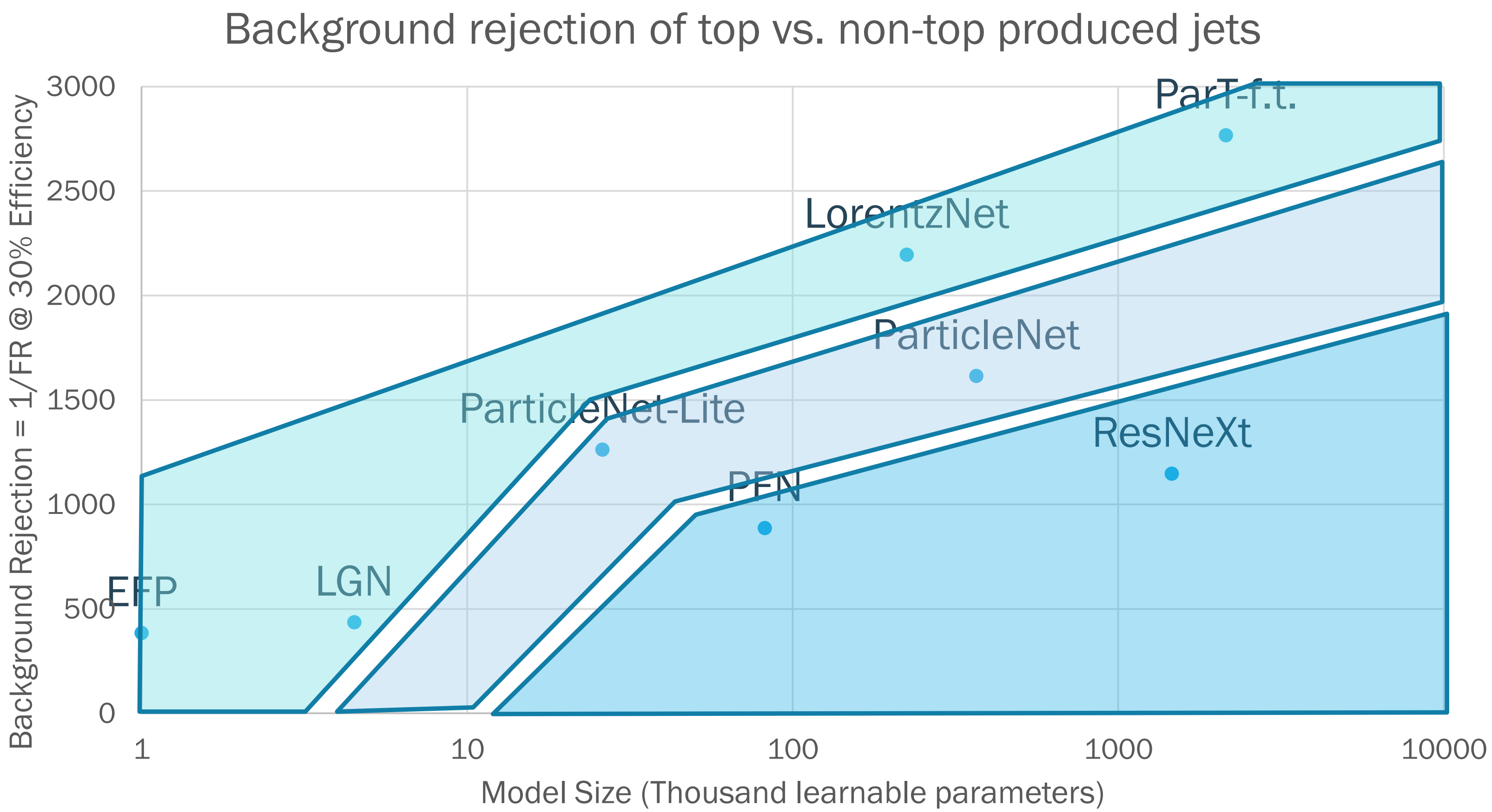
PHYSICS-MOTIVATED ML (SYMMETRY, DATA)

RELATIONAL ML (GRAPHS)

NON-RELATIONAL ML (SETS, IMAGES)

Would love to add LundNet-5 and JEDI-net to this plot, but don't have apples-to-apples rejection rate

WHY EQUI-GNN: NAIVELY IMPROVING MODEL PERFORMANCE



- Given a particular ML structure (a.k.a relational bias), diminishing returns on simply increasing model size
- Graph-structured appears to be as general as one can get structurally
- GNN-based models seem to perform best at large size
- Physics-based models seem to perform best at small size
- Motivates us to constrain graph-structured ML with physics knowledge

Would love to add LundNet-5 and JEDI-net to this plot, but don't have apples-to-apples rejection rate

JET TAGGING: KINDS OF PHYSICS KNOWLEDGE

- A variety of knowledge about the physics case can be included in the algorithm
- Quantum field theory: Feynman diagram structure (EFP)
- QCD: Decay processes in the Lund plane (LundNet)
- Permutation invariance of the jet constituents (PFN, ParticleNet)
- QCD + permutation invariance: Lund features with GNN (ParT: ParticleTransformer)
- 2D translation invariance in the calorimeter (ResNeXt)
- Special relativity: Frame-invariance under Lorentz transformations (LorentzNet, VecNet, Covariant ParT, …)

Good summary of theory-based tagging in Kasieczka, et al.

QFT Symmetries

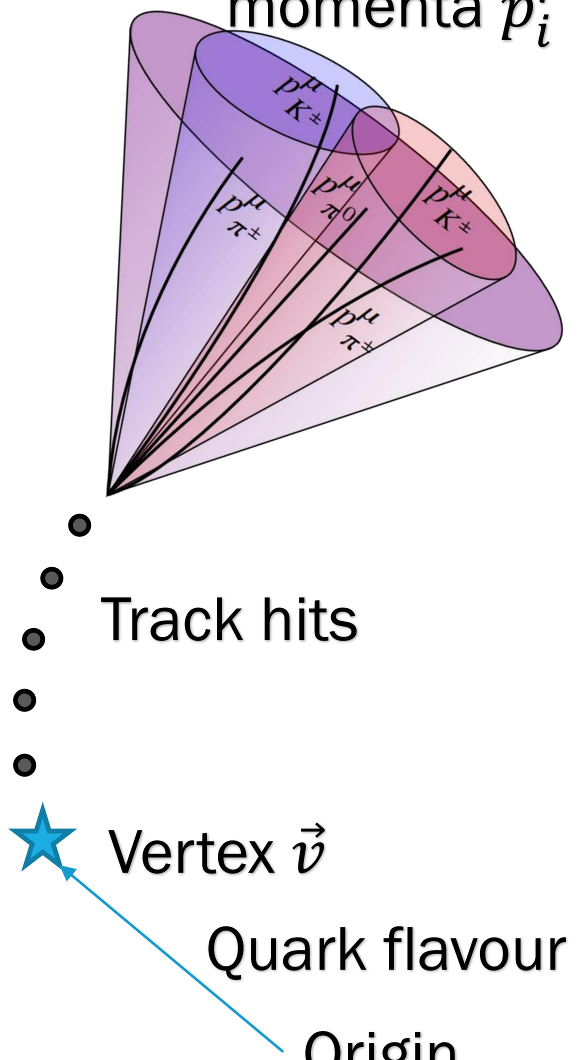
Spacetime Symmetries

Physics-informed Features

Data Augmentation

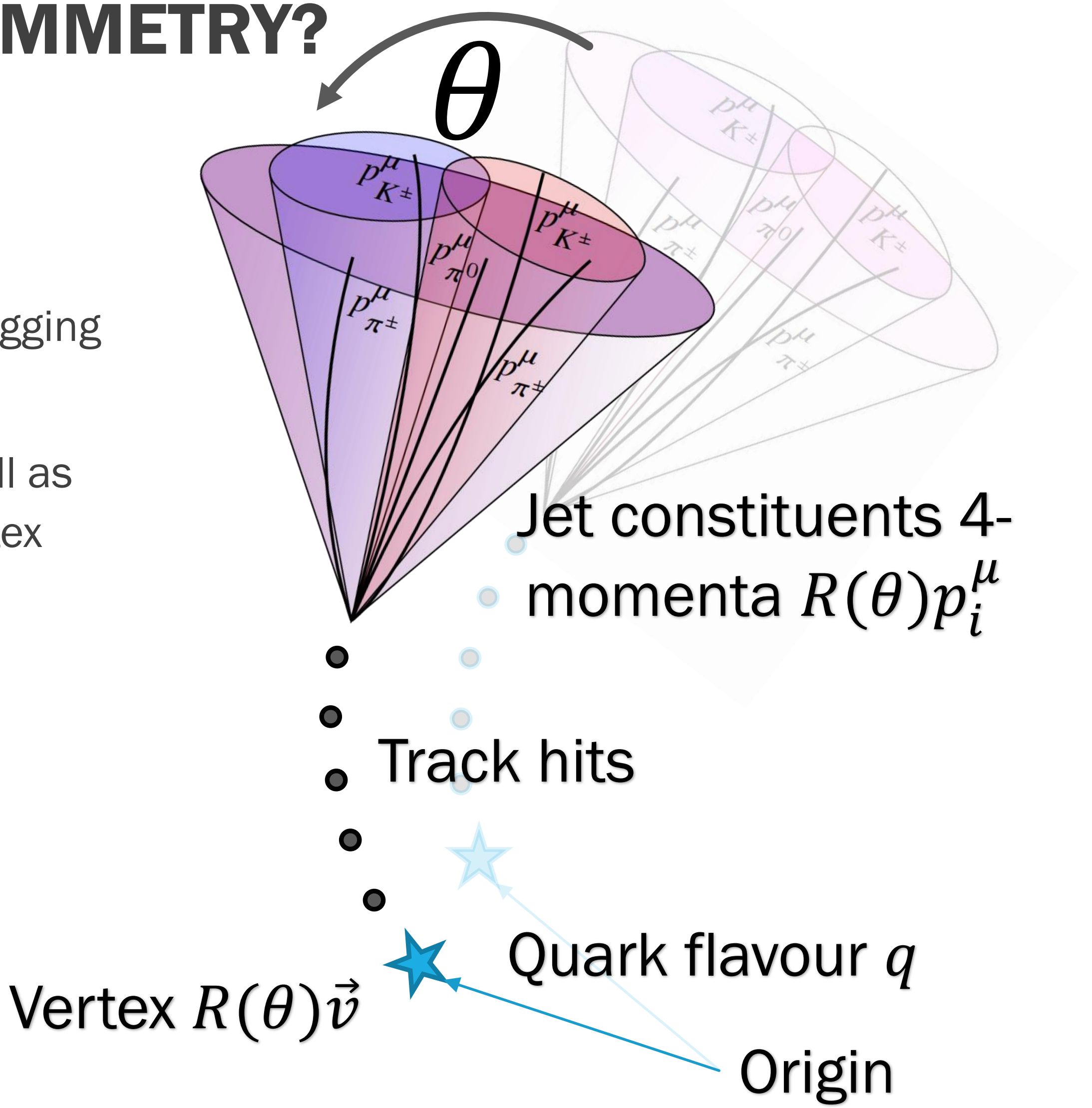
WHAT DOES IT MEAN TO INCLUDE A SYMMETRY?

- Consider a task described very well on Wednesday: Jet flavor tagging
- Nilotpal Kakati outlined a GNN for predicting the source of jet production (b quark, c quark, tau jet or a lighter particle), as well as auxiliary predictions: track production vertex and track-pair vertex compatibility [ATL-PHYS-PUB-2022-027]
- Consider rotating the jet by angle ϕ, using rotation matrix $R(\theta)$



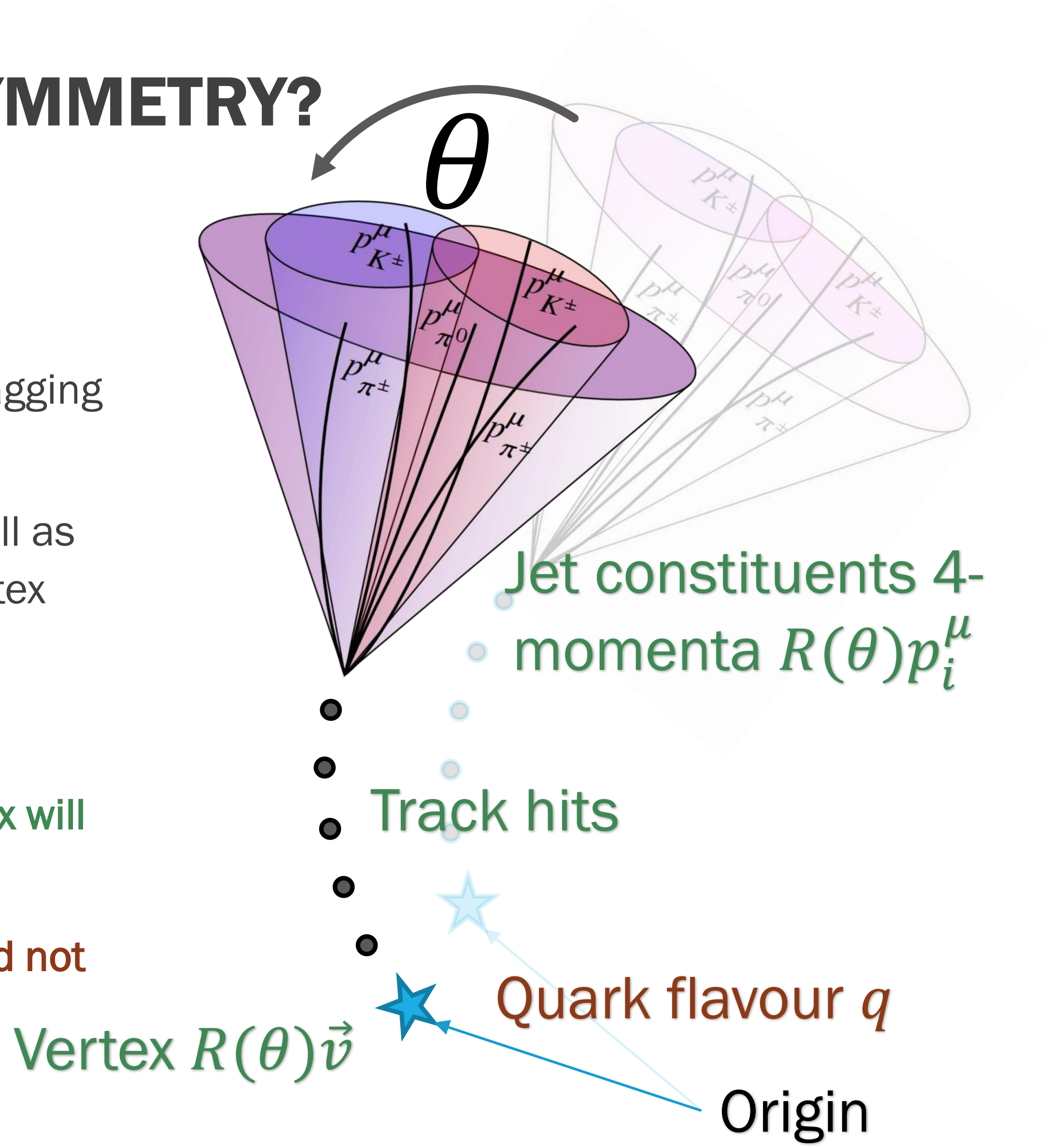
WHAT DOES IT MEAN TO INCLUDE A SYMMETRY?

- Consider a task described very well on Wednesday: Jet flavor tagging
- Nilotpal Kakati outlined a GNN for predicting the source of jet production (b quark, c quark, tau jet or a lighter particle), as well as auxiliary predictions: track production vertex and track-pair vertex compatibility [ATL-PHYS-PUB-2022-027]
- Consider rotating the jet by angle ϕ, using rotation matrix $R(\theta)$



WHAT DOES IT MEAN TO INCLUDE A SYMMETRY?

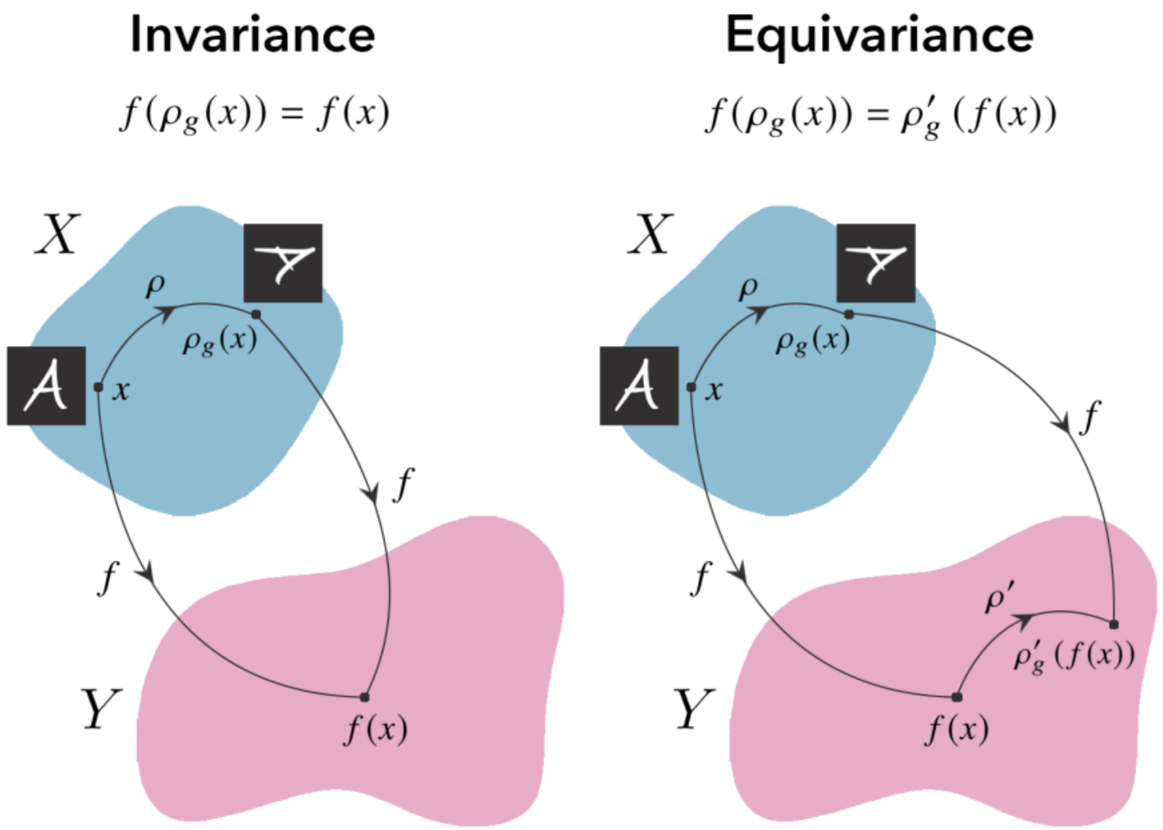
- Consider a task described very well on Wednesday: Jet flavor tagging
- Nilotpal Kakati outlined a GNN for predicting the source of jet production (b quark, c quark, tau jet or a lighter particle), as well as auxiliary predictions: track production vertex and track-pair vertex compatibility [ATL-PHYS-PUB-2022-027]
- Consider rotating the jet by angle ϕ, using rotation matrix $R(\theta)$
- Some predictions (and input features) like the production vertex will rotate with the transformation: "equivariant"
- Some predictions (and input features) like the jet flavour should not be affected: "invariant"



INVARIANCE VS. EQUIVARIANCE

- For some neural network f and some input feature x
- For a group element $g \in G$ transformation ρ_g
- Invariant network leaves output unaffected

  $f\left(\rho_g(x)\right) = f(x)$
- Equivariant (under G) network gives an output that is also transformed by $g \in G$
- May be same representation ρ_g or another representation ρ'_g

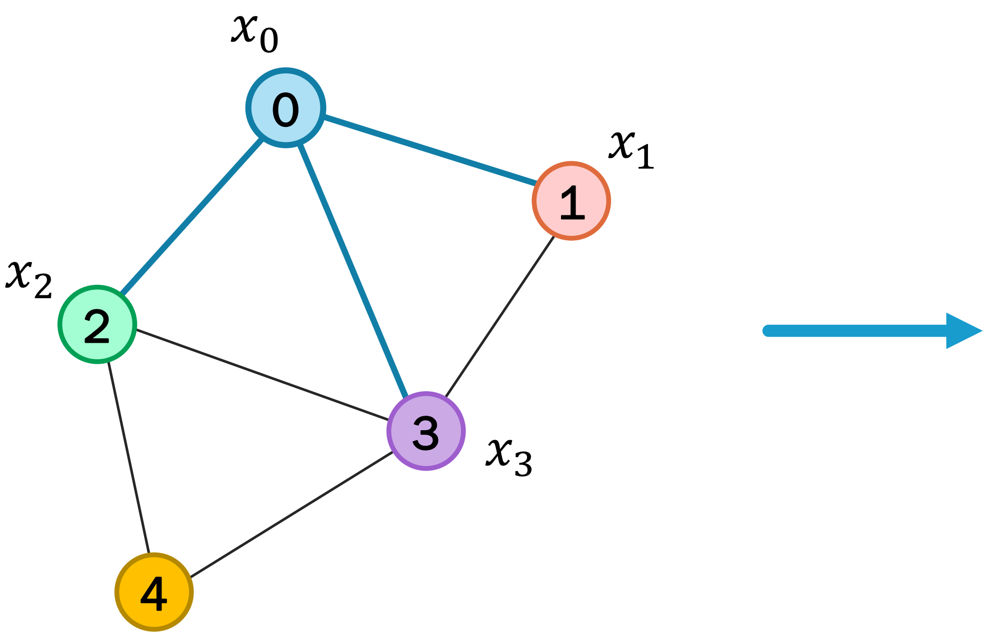


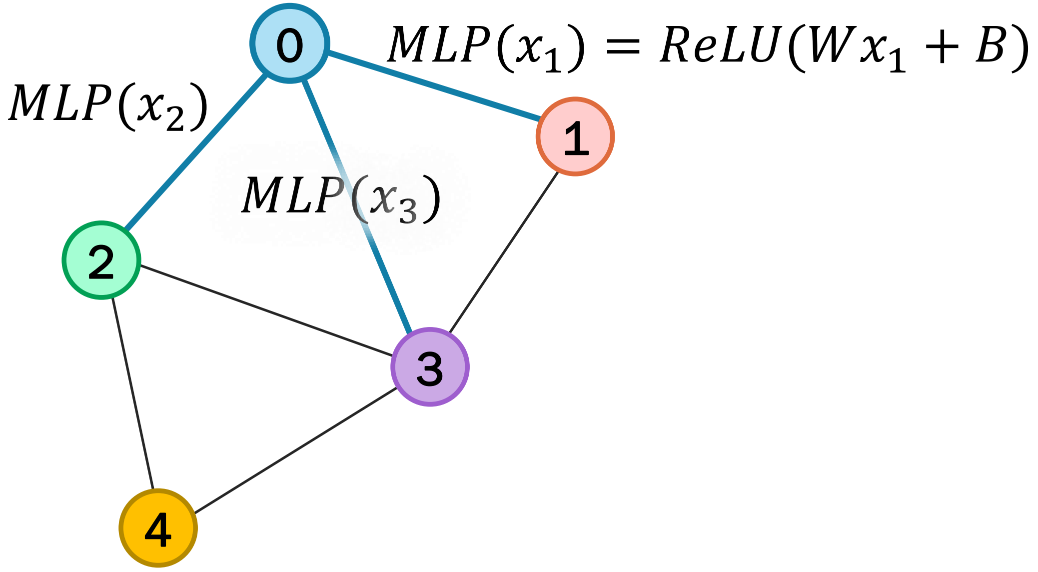
Lovely plot from Mariel Pettee: *Symmetry Group Equivariant Architectures for Physics – Snowmass White Paper*

WHAT DOES IT MEAN TO INCLUDE A SYMMETRY?

- Consider a point cloud, with behavior that you expect to be invariant under E3 symmetry – 3 dimensional Euclidean (rotational and translational) transformations
- Observe how a transformation R propagates in some arbitrary GNN convolution:

$$x_0' = \sum_i MLP(x_i)$$

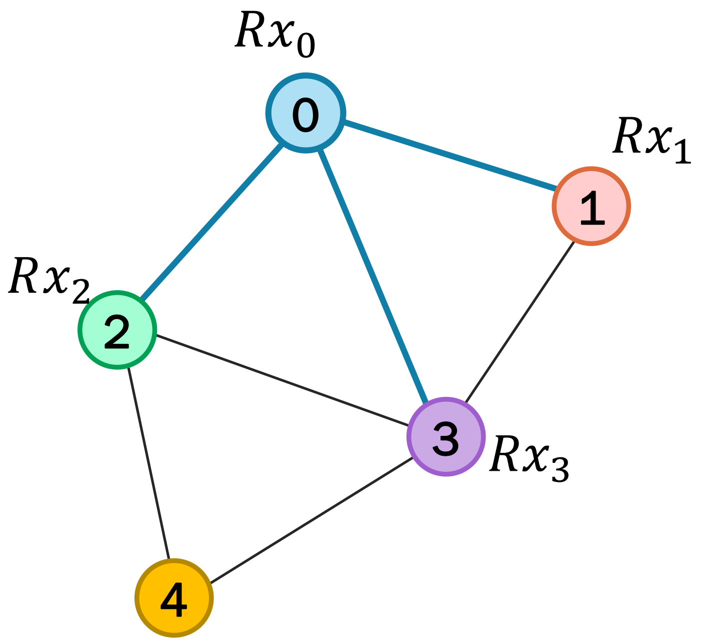


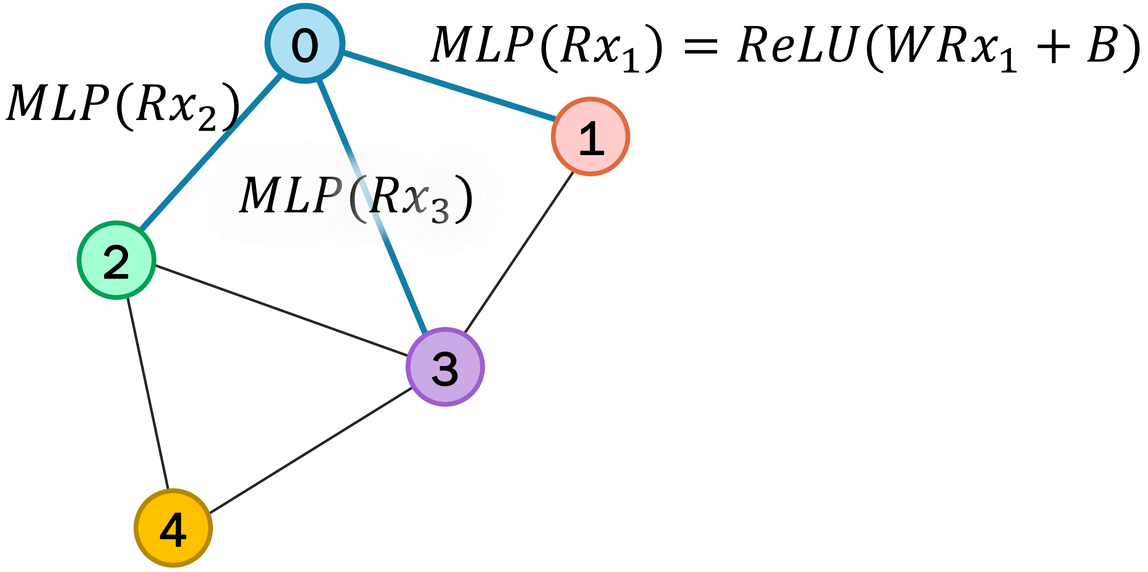


WHAT DOES IT MEAN TO INCLUDE A SYMMETRY?

- Consider a point cloud, with behavior that you expect to be invariant under E3 symmetry – 3 dimensional Euclidean (rotational and translational) transformations
- Observe how a transformation R propagates in some arbitrary GNN convolution:

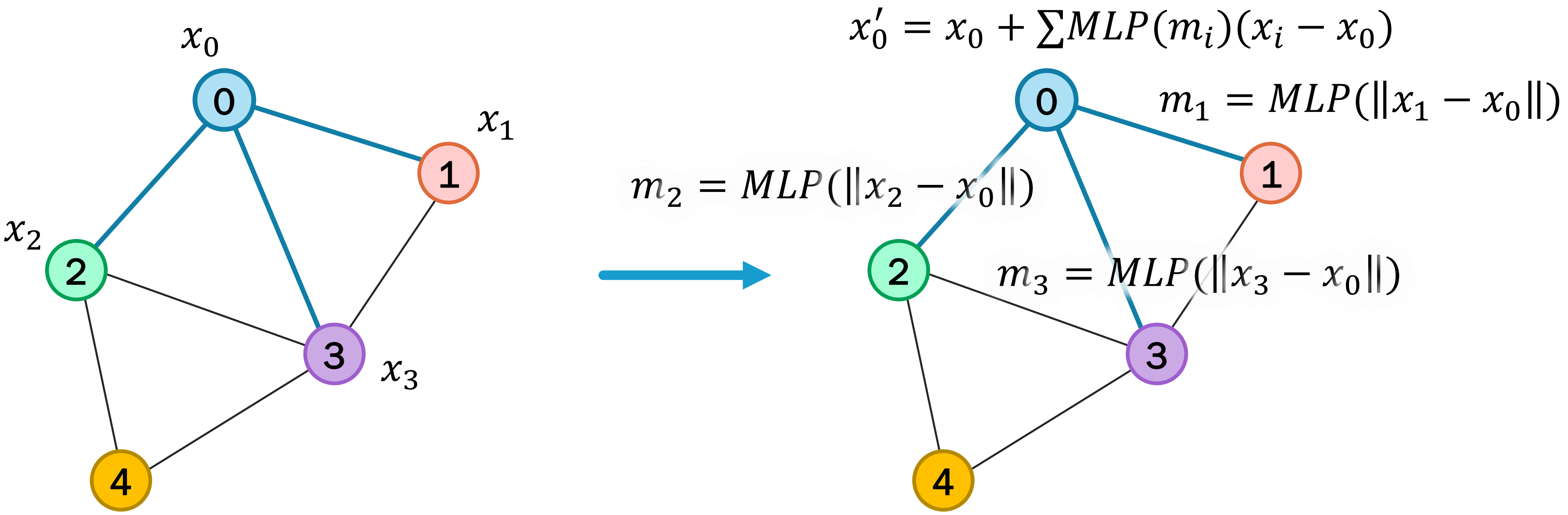
$$x_0' = \sum_i MLP(Rx_i) = ??$$





WHAT DOES IT MEAN TO INCLUDE A SYMMETRY?

- Consider a point cloud, with behavior that you expect to be invariant under E3 symmetry – 3 dimensional Euclidean (rotational and translational) transformations
- Observe how a transformation R propagates in some arbitrary GNN convolution
- To preserve E3 symmetry, we must choose a specific kind of message passing function:

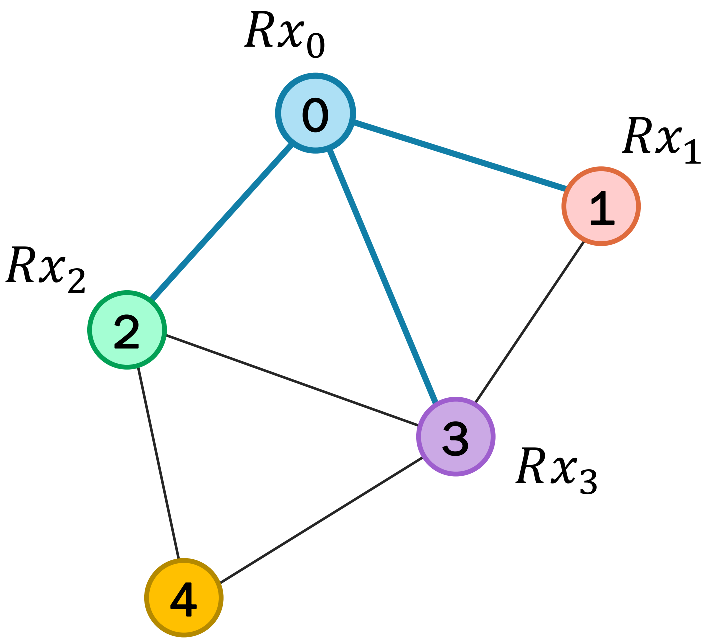


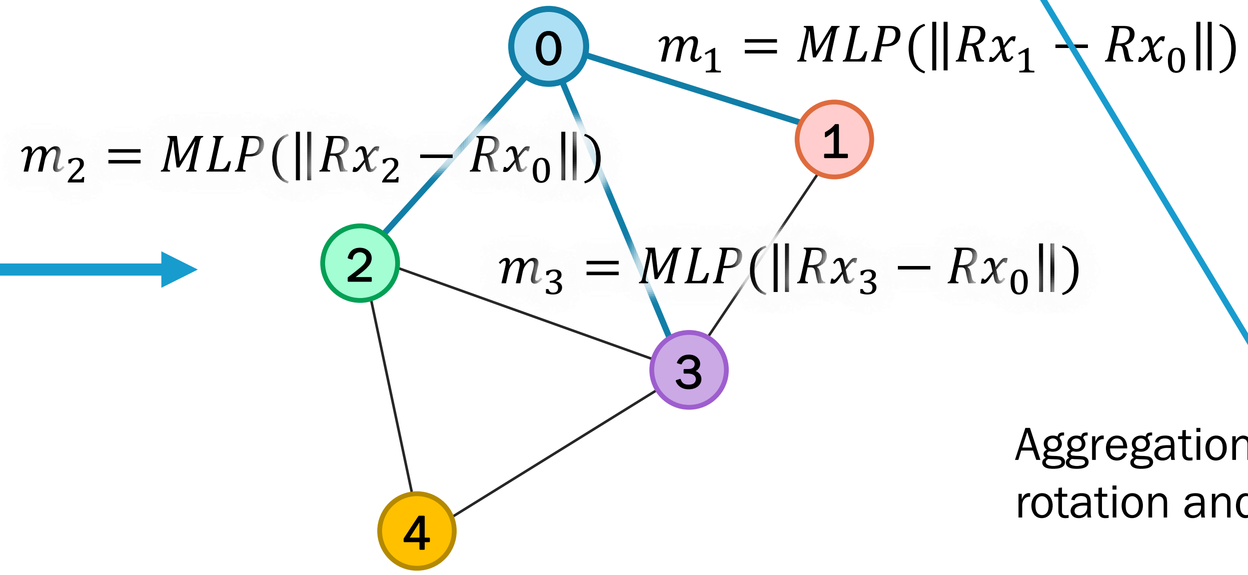
WHAT DOES IT MEAN TO INCLUDE A SYMMETRY?

$$
\begin{aligned}
\|Rx_3 - Rx_0\|^2 &= (Rx_3 - Rx_0)^T(Rx_3 - Rx_0) \\
&= (x_3 - x_0)^T R^T R(x_3 - x_0) \\
&= \|x_3 - x_0\|^2
\end{aligned}
$$

Message passing invariant to rotation and translation

$$Rx_0 + \sum MLP(m_i)(Rx_i - Rx_0) = Rx_0'$$

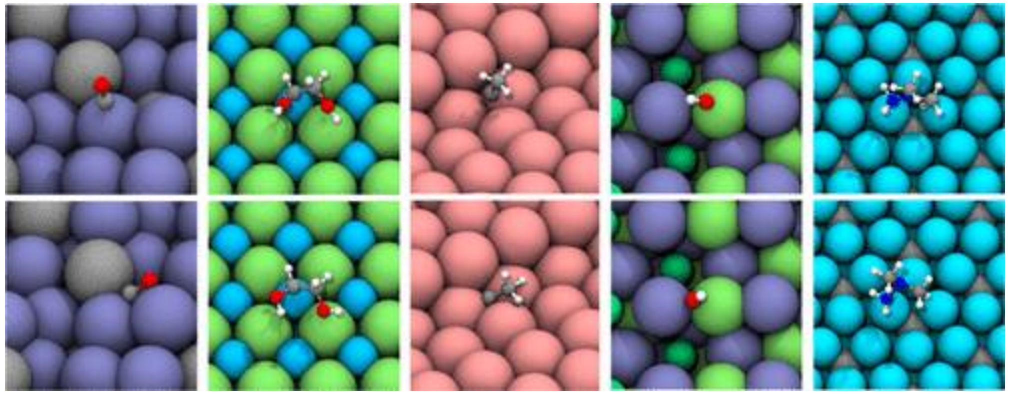




Aggregation equivariant to rotation and translation

E(n) Equivariant Graph Neural Networks, Satorras, et al

LANDSCAPE OF EQUIVARIANT GNNS



- Most investigation on equivariant GNNs (and point-based convolutions) done in context of molecular property prediction, e.g. Open Catalyst 2020 Dataset
- The kind of symmetry-preserving GNN is described by the message passing and update functions
- Nice comparison of popular EquiGNNs in *E(n) Equivariant Graph Neural Networks, Satorras, et al*:

	GNN	Radial Field	TFN	Schnet	EGNN
Edge	$\mathbf{m}_{ij} = \phi_e(\mathbf{h}_i^l, \mathbf{h}_j^l, a_{ij})$	$\mathbf{m}_{ij} = \phi_{\text{rf}}(\lVert\mathbf{r}_{ij}^l\rVert)\mathbf{r}_{ij}^l$	$\mathbf{m}_{ij} = \sum_k \mathbf{W}^{lk}\mathbf{r}_{ji}^l\mathbf{h}_i^{lk}$	$\mathbf{m}_{ij} = \phi_{\text{cf}}(\lVert\mathbf{r}_{ij}^l\rVert)\phi_{\text{s}}(\mathbf{h}_j^l)$	$\mathbf{m}_{ij} = \phi_e(\mathbf{h}_i^l, \mathbf{h}_j^l, \lVert\mathbf{r}_{ij}^l\rVert^2, a_{ij})$ $\hat{\mathbf{m}}_{ij} = \mathbf{r}_{ij}^l\phi_x(\mathbf{m}_{ij})$
Agg.	$\mathbf{m}_i = \sum_{j\in\mathcal{N}(i)}\mathbf{m}_{ij}$	$\mathbf{m}_i = \sum_{j\neq i}\mathbf{m}_{ij}$	$\mathbf{m}_i = \sum_{j\neq i}\mathbf{m}_{ij}$	$\mathbf{m}_i = \sum_{j\neq i}\mathbf{m}_{ij}$	$\mathbf{m}_i = \sum_{j\neq i}\mathbf{m}_{ij}$ $\hat{\mathbf{m}}_i = C\sum_{j\neq i}\hat{\mathbf{m}}_{ij}$
Node	$\mathbf{h}_i^{l+1} = \phi_h(\mathbf{h}_i^l, \mathbf{m}_i)$	$\mathbf{x}_i^{l+1} = \mathbf{x}_i^l + \mathbf{m}_i$	$\mathbf{h}_i^{l+1} = w^{ll}\mathbf{h}_i^l + \mathbf{m}_i$	$\mathbf{h}_i^{l+1} = \phi_h(\mathbf{h}_i^l, \mathbf{m}_i)$	$\mathbf{h}_i^{l+1} = \phi_h\left(\mathbf{h}_i^l, \mathbf{m}_i\right)$ $\mathbf{x}_i^{l+1} = \mathbf{x}_i^l + \hat{\mathbf{m}}_i$
	Non-equivariant	$\mathrm{E}(n)$-Equivariant	$\mathrm{SE}(3)$-Equivariant	$\mathrm{E}(n)$-Invariant	$\mathrm{E}(n)$-Equivariant

See also Steerable EGNNs for a further generalisation to EGNNs (isotropic to anisotropic)

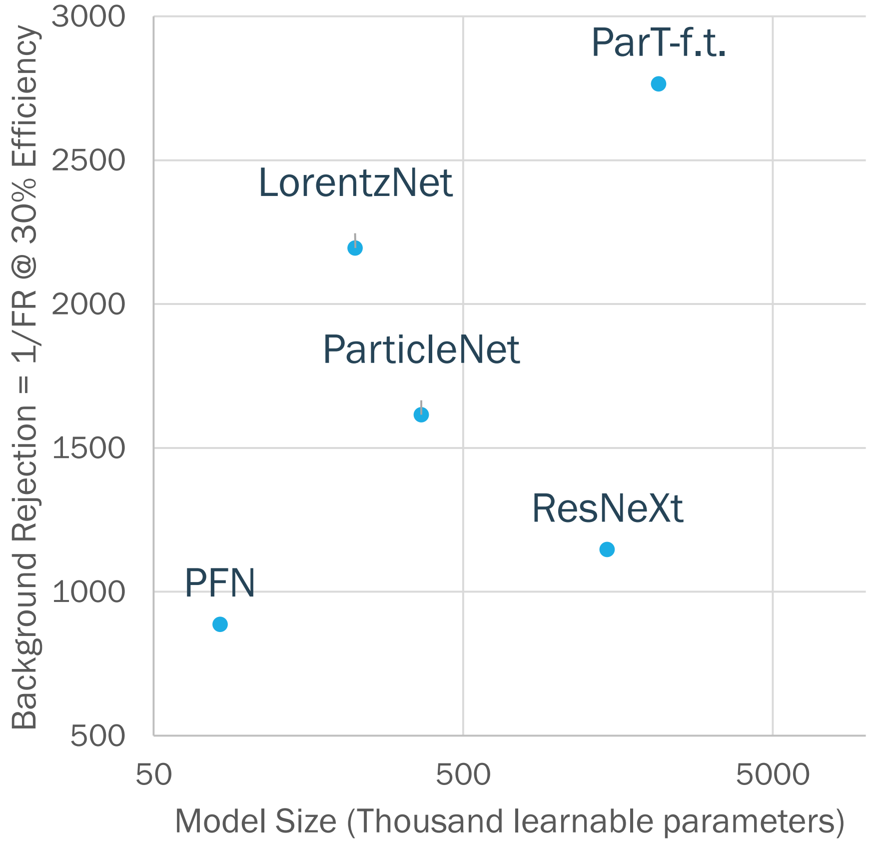
BENEFITS OF EQUIVARIANCE: OVERALL BETTER ACCURACY

- Lorentz equivariance in jet tagging pioneered in Bogatskiy, et al, *Lorentz Group Equivariant Neural Network for Particle Physics*
- Showed recipe for all tensor representations of Lorentz symmetry using Clebsch-Gordon decomposition. However performance not SotA
- LorentzNet used Lorentz extension to EGNN, achieved SotA

Model	Accuracy	AUC	$1/\varepsilon_B$ ($\varepsilon_S = 0.5$)	$1/\varepsilon_B$ ($\varepsilon_S = 0.3$)
ResNeXt	0.936	0.9837	302 ± 5	1147 ± 58
P-CNN	0.930	0.9803	201 ± 4	759 ± 24
PFN	0.932	0.9819	247 ± 3	888 ± 17
ParticleNet	0.940	0.9858	397 ± 7	1615 ± 93
EGNN	0.922	0.9760	148 ± 8	540 ± 49
LGN	0.929	0.9640	124 ± 20	435 ± 95
LorentzNet	**0.942**	**0.9868**	$\mathbf{498 \pm 18}$	$\mathbf{2195 \pm 173}$

Table from *An Efficient Lorentz Equivariant Graph Neural Network for Jet Tagging, Gong et al*

Background rejection of top vs. non-top produced jets



BENEFITS OF EQUIVARIANCE: BETTER POWER-TO-WEIGHT RATIO

- Perhaps ParticleTransformer may be the final choice for at-all-costs accuracy, but EquiGNNs use existing knowledge about the world to get more accuracy-per-parameter
- We suggest in *Semi-Equivariant GNN Architectures for Jet Tagging*, [DM, Thais, Wong], to capture this concept as a figure-of-merit: The "ant factor"
- The better the background rejection, and the smaller the network, the more "ant-like" a tagger is:

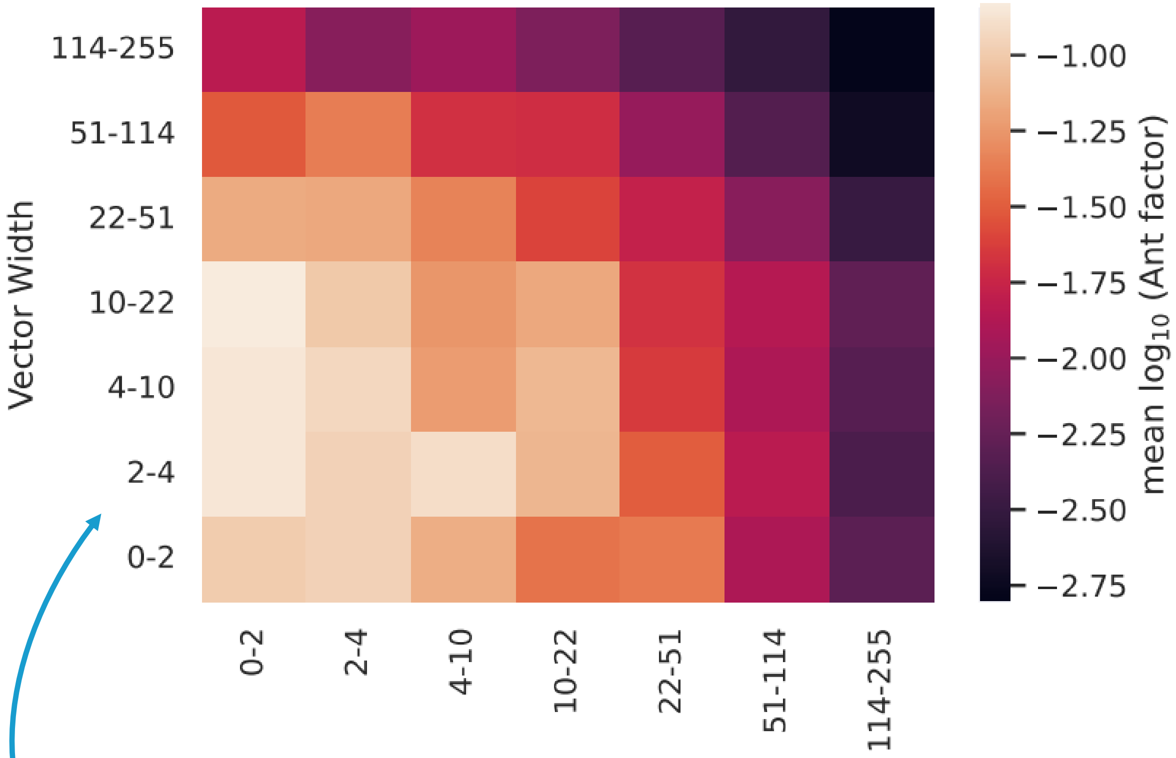
$$ant = \frac{\text{accuracy}}{\text{model size}} = \frac{\epsilon_B^{-1}}{N_{params}}$$

BENEFITS OF EQUIVARIANCE: BETTER POWER-TO-WEIGHT RATIO

- The better the background rejection, and the smaller the network, the more "ant-like" a tagger is:

$$ant = \frac{\text{accuracy}}{\text{model size}} = \frac{\epsilon_B^{-1}}{N_{params}}$$

- Our model *VecNet* uses both equivariant and non-equivariant channels (developed in tandem - but without knowledge of – LorentzNet)
- We use it as a sandbox to interpret the usefulness of equivariant channels
- **Find the best power-to-weight at a *mostly equivariant* configuration**



Lighter colour is more ant-like network

BENEFITS OF EQUIVARIANCE: BETTER SAMPLE EFFICIENCY

- Sample efficiency is the ability to perform well with a minimal dataset
- Constraining a convolution to be equivariant reduces the possible function space
- Smaller model size may also prevent overtraining on small dataset
- Studies of molecular properties and jet flavour tagging show significantly better sample efficiency

Table 1 Energy and Force MAE for molecules on the original MD-17 data set, reported in units of [meV] and [meV/Å], respectively, and a training budget of 1000 reference configurations.

Molecule		SchNet	DimeNet	sGDML	PaiNN	SpookyNet	GemNet-(T/Q)	NewtonNet	UNiTE	NequIP ($l = 3$)
Aspirin	Energy	16.0	8.8	8.2	6.9	6.5	-	7.3	-	**5.7**
	Forces	58.5	21.6	29.5	14.7	11.2	9.4	15.1	**6.8**	8.0
Ethanol	Energy	3.5	2.8	3.0	2.7	2.3	-	2.6	-	**2.2**
	Forces	16.9	10.0	14.3	9.7	4.1	3.7	9.1	4.0	**3.1**
Malonaldehyde	Energy	5.6	4.5	4.3	3.9	3.4	-	4.2	-	**3.3**
	Forces	28.6	16.6	17.8	13.8	7.2	6.7	14.0	6.9	**5.6**
Naphthalene	Energy	6.9	5.3	5.2	5.0	5.0	-	5.1	-	**4.9**
	Forces	25.2	9.3	4.8	3.3	3.9	2.2	3.6	2.8	**1.7**
Salicylic acid	Energy	8.7	5.8	5.2	4.9	4.9	-	5.0	-	**4.6**
	Forces	36.9	16.2	12.1	8.5	7.8	5.4	8.5	4.2	**3.9**
Toluene	Energy	5.2	4.4	4.3	4.1	4.1	-	4.1	-	**4.0**
	Forces	24.7	9.4	6.1	4.1	3.8	2.6	3.8	3.1	**2.0**
Uracil	Energy	6.1	5.0	4.8	**4.5**	4.6	-	4.6	-	**4.5**
	Forces	24.3	13.1	10.4	6.0	5.2	4.2	6.5	4.2	**3.3**

For GemNet, the best result out of the T/Q versions is presented and for PaiNN the best between force-only and joint force and energy training. For UNiTE, we compare to the "direct-learning" results reported in[26].
Best results are marked in bold.

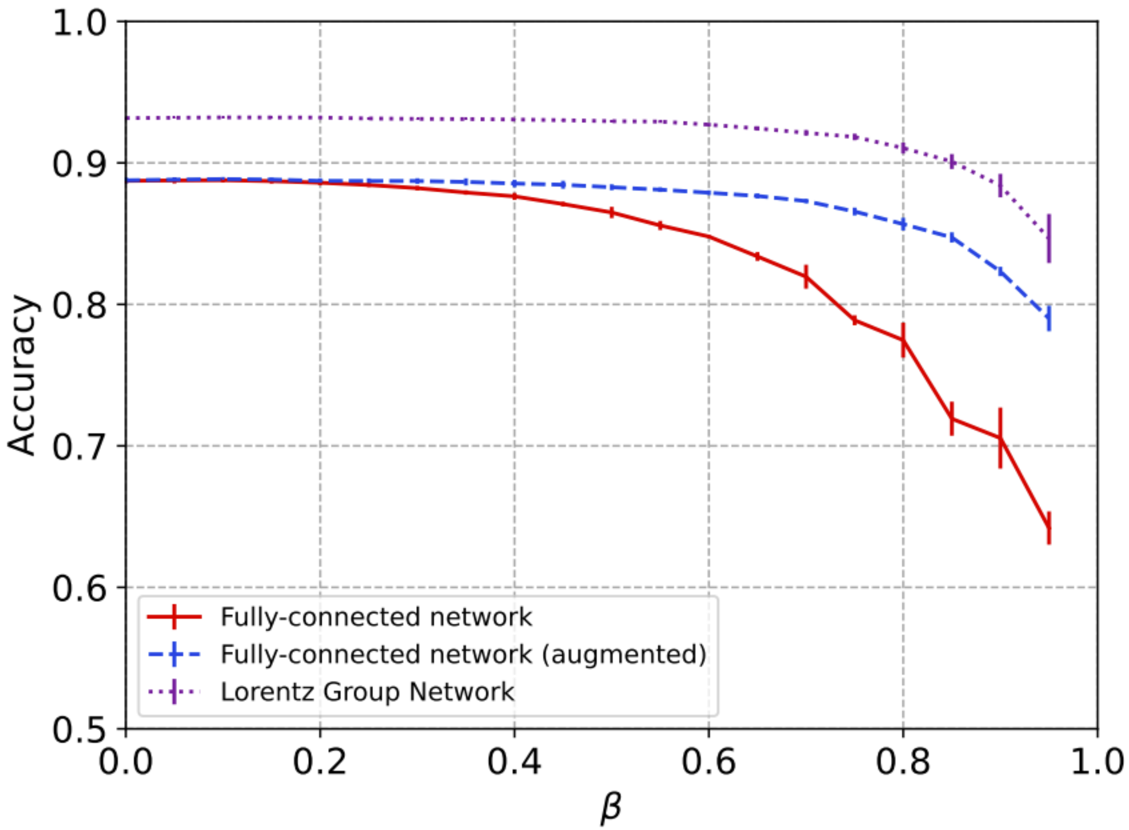
Training Fraction	Model	Accuracy
0.5%	ParticleNet	0.913
	LorentzNet	**0.929**
1%	ParticleNet	0.919
	LorentzNet	**0.932**
5%	ParticleNet	0.931
	LorentzNet	**0.937**

↑ LorentzNet outperforms ParticleNet on a tiny dataset

← NequIP outperforms most other GNNs with small data budget

BENEFITS OF EQUIVARIANCE: BETTER GENERALIZABILITY

- When symmetry is "baked into" the network, generalizability is guaranteed – up to numerical precision
- E.g. Consider the Lorentz Group Network (LGN) vs. a naïve MLP classifier applied to jet tagging
- Red curve shows naïve classifier trained on single boost frame, applied to data across a range of boost values β
- Blue curve shows naïve classifier trained on a range of boost frames
- Purple curve shows LGN trained on a single boost frame
- LGN has similar robustness to data augmentation, without needing to generate and train on general dataset



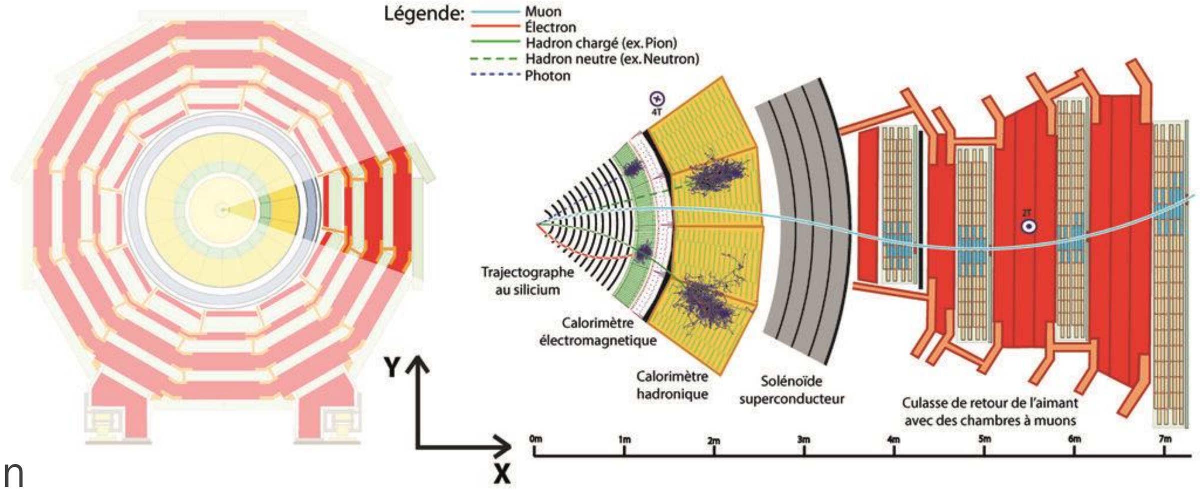
Very nice plot by Jan Offerman. Courtesy of Snowmass Whitepaper

BENEFITS OF EQUIVARIANCE: BETTER INTERPRETABILITY

- At its simplest, the EGNN update rule is just a weighted sum of neighbouring vector values – easy to follow information propagation around graph
- Conversely, if an equivariant network *cannot perform as well as a non-equivariant network*, then we have a tool to probe why the non-equivariant network performs well
- Recent work *Automatic Symmetry Discovery with Lie Algebra Convolutional Network* uses learnable algebra convolutions, allowing any Lie group to be **learned**
- This may be the best of both worlds: A network that can be as (sample / parameter / …) efficient as an unconstrained model, while also delivering an explanation of its own learning (aka "self-interpretable" ML)

PROSPECTS FOR TRACKING: OPEN FOR DISCUSSION!

- Many experiments exhibit *some* spatial symmetry:
  - Parity
  - Rotation around an axis
  - Translational
- Some experiments exhibit Lorentz symmetry:
  - Reconstructed tracks with 4-momenta
  - Hits with timing information
- May be able to incorporate internal symmetries
  - Charge should be preserved through track decay chain
  - Color should be confined through hadronic showers
  - Baryon number, lepton number, …
- Experiment co-design: If we prove the power of equivariance, then *design* an experiment to exploit symmetries



CMS Detector
Courtesy CMS Collaboration

DISCUSSION

1. Do you have a physical problem that you believe obeys some symmetry? Does you instrument/detector/experiment give you the data you need to exploit that symmetry in an ML architecture?
2. Have you tried an equivariant ML model (or more generally applied some physics-informed algorithm)? What was your experience with it?
3. Do we feel that this is a fertile direction of study? Is it possibly a bubble what will be popped by ever-larger unconstrained, generalist GNNs/Transformers? If a generalist GNN can be compressed and pruned, do we need physics-informed architectures?
4. Could EquiGNNs have a place in improving unconstrained models? Either by pre-training a model, or in some student-teacher system?
5. Could EquiGNNs have a place as an interpretability translator? I.e. an unconstrained model teaches an EquiGNN to perform a task, which can then be interpreted by following the symmetry-preserving represenations through the network.
6. Could symmetries themselves be learned from experimental data?
7. Is there anything you've been thinking about or been excited about that is not on this list?

CONCLUSION

- Equivariant GNNs are philosophically attractive and well-motivated
- They also happen to perform very well, e.g. for jet flavour tagging
- They also happen to give good performance for fewer parameters and operations
- As expected, they generalise very well along their enforced symmetry
- They *may* have ramifications for interpretable and explainable AI/ML

THANKS FOR TUNING IN!

Links

Snowmass Equivariance white paper, Snowmass GNN white paper, Lorentz-Equivariant Repo