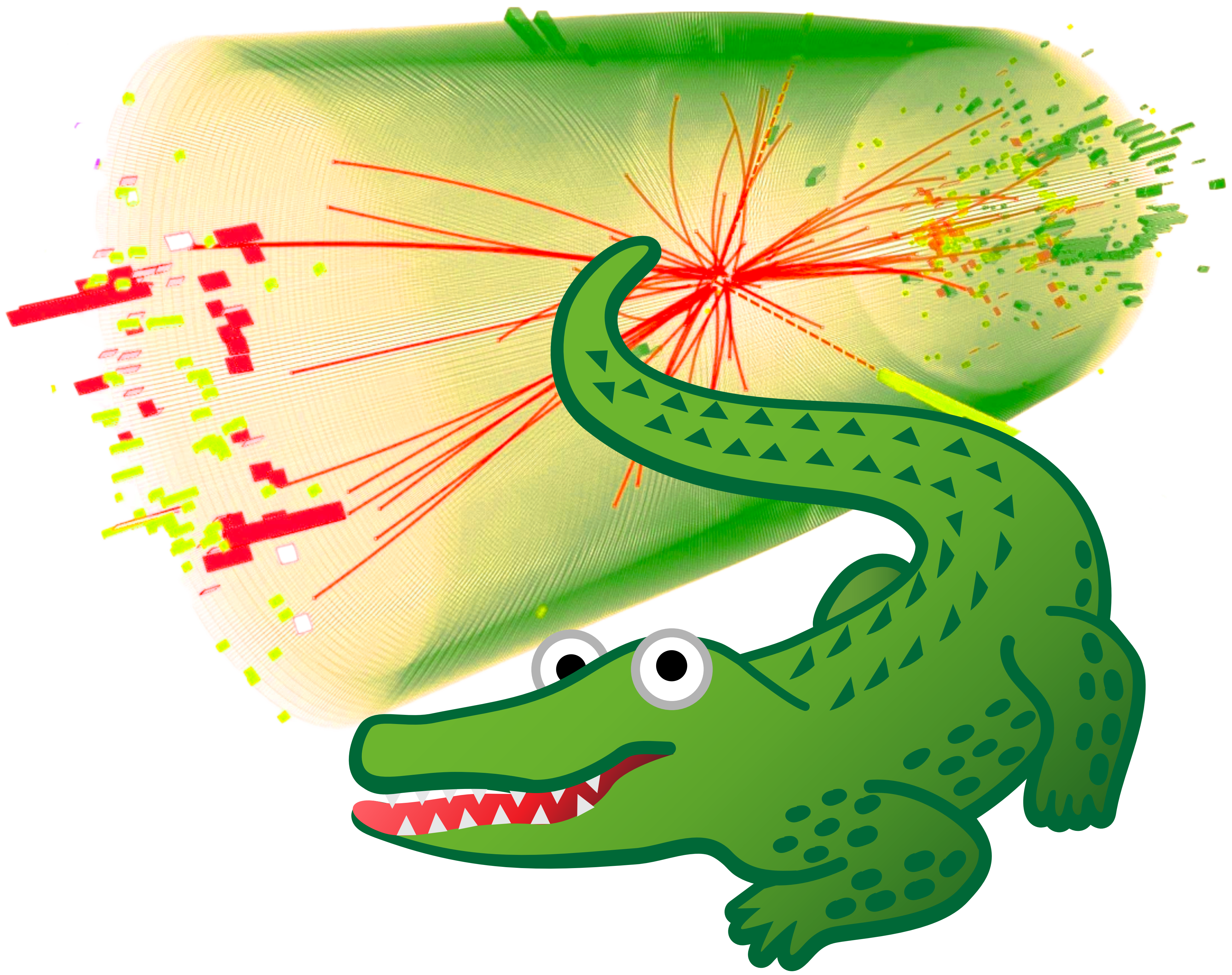
A Lorentz Equivariant Transformer for All of the LHC

Víctor Bresó Pla

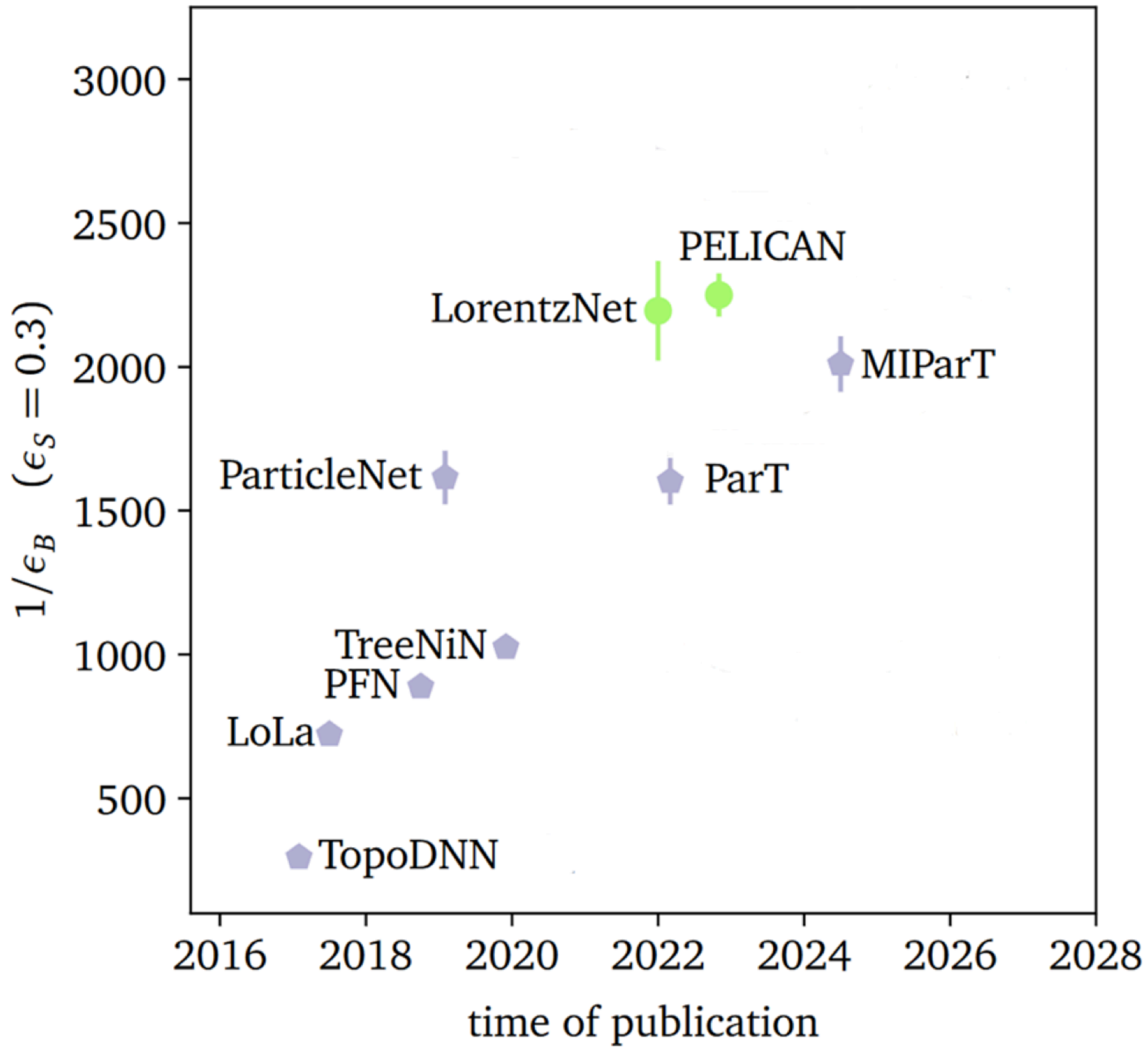
In collaboration with Jonas Spinner, Johann Brehmer, Pim de Haan, Tilman Plehn, Huilin Qu & Jesse Thaler

arXiv:2405.14806 [physics.data-an]
arXiv:2411.00104 [hep-ph, hep-ex]

Public lgatr package repository

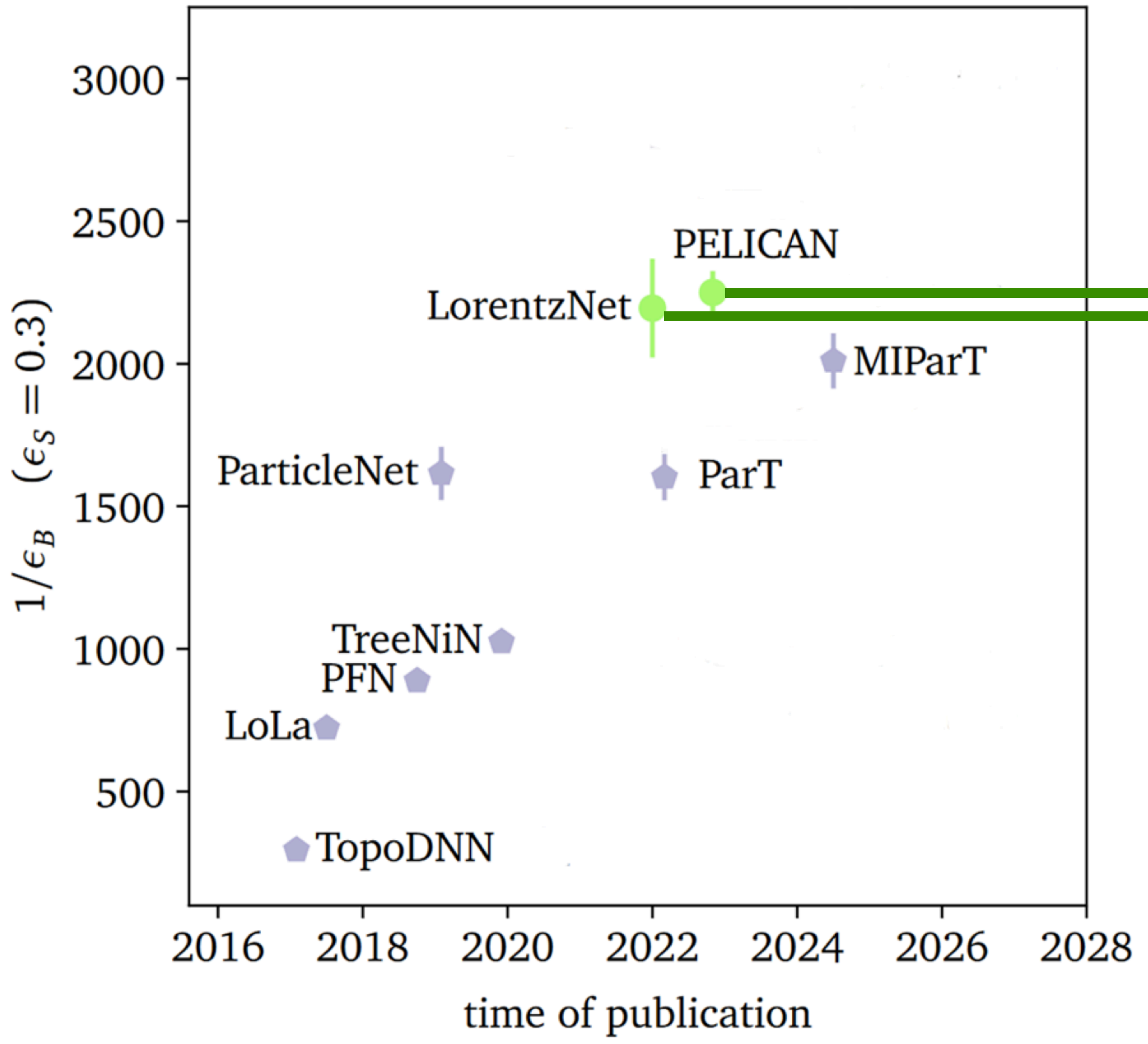


History of Top Tagging



A. Bogatskiy et al., 2211.00454
S. Gong et al., 2201.08187
D. Ruhe et al., 2305.11141

History of Top Tagging



Equivariant Neural Networks!!

A. Bogatskiy et al., 2211.00454
S. Gong et al., 2201.08187
D. Ruhe et al., 2305.11141

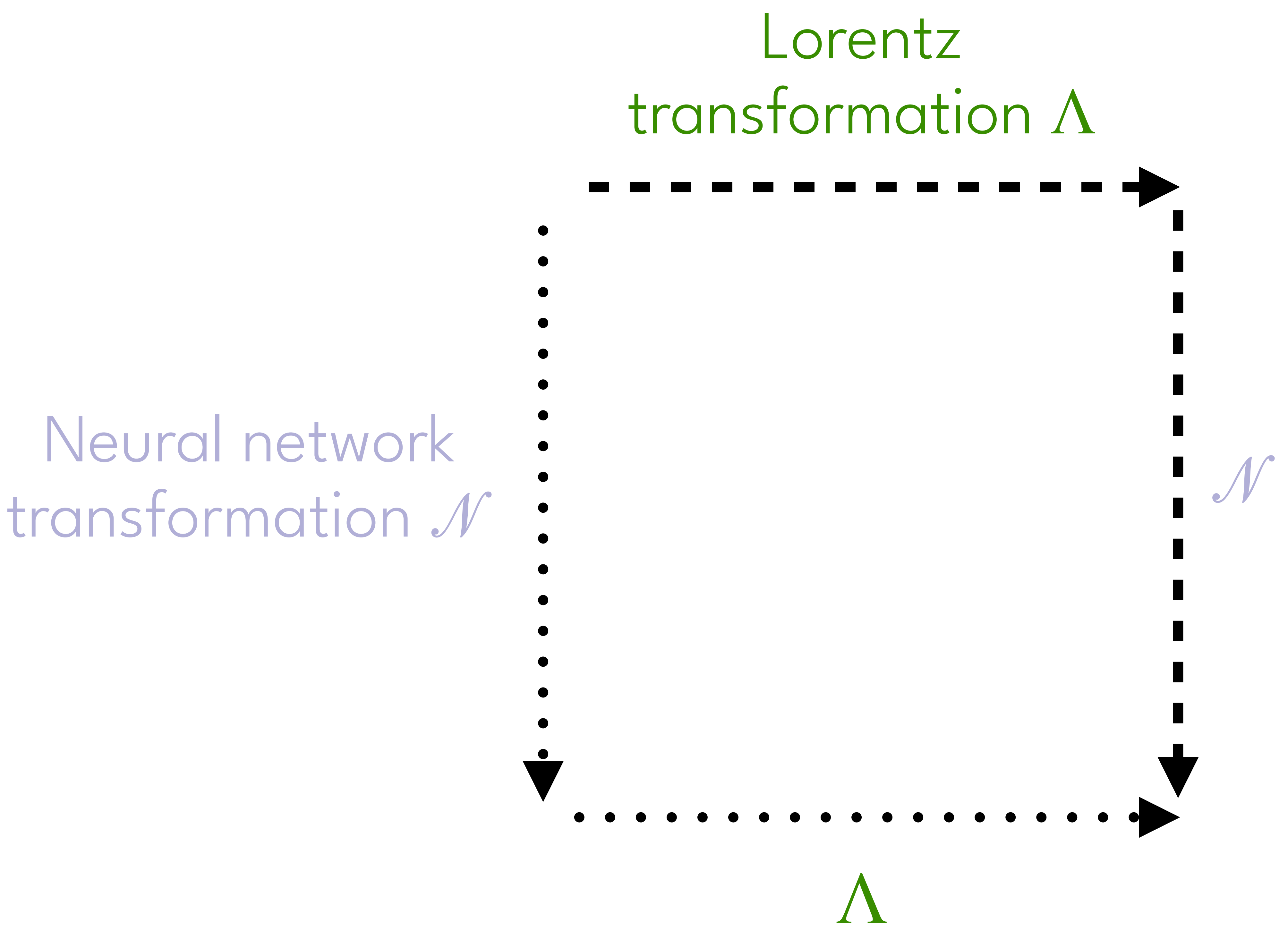
Equivariant Neural Networks

What are equivariant neural networks?

Equivariant Neural Networks

What are equivariant neural networks?

$$\mathcal{N}\left(\Lambda\left(x\right)\right) = \Lambda\left(\mathcal{N}\left(x\right)\right)$$

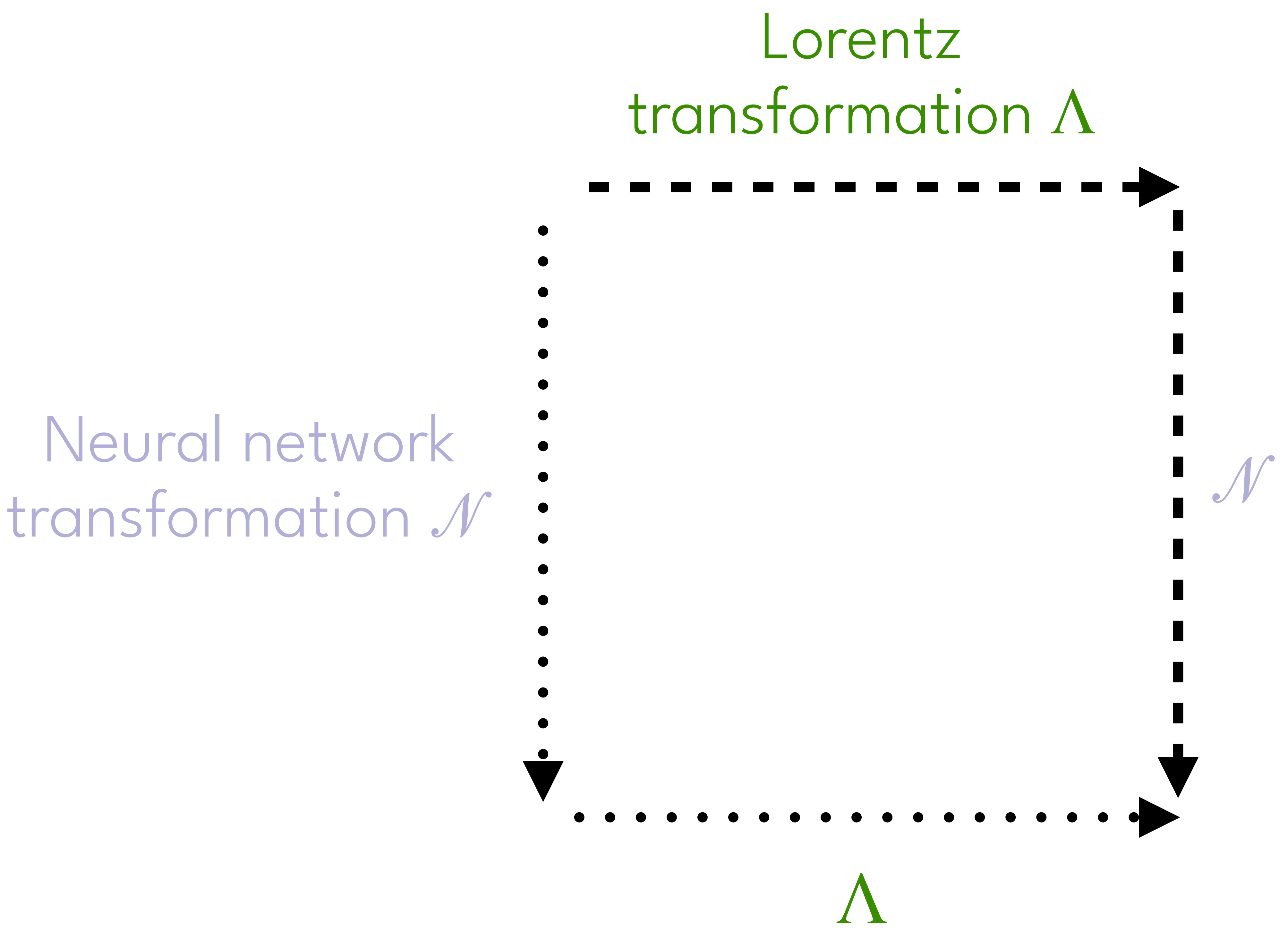


Equivariant Neural Networks

What are equivariant neural networks?

$$\mathcal{N}\left(\Lambda\left(x\right)\right) = \Lambda\left(\mathcal{N}\left(x\right)\right)$$

Why equivariance?



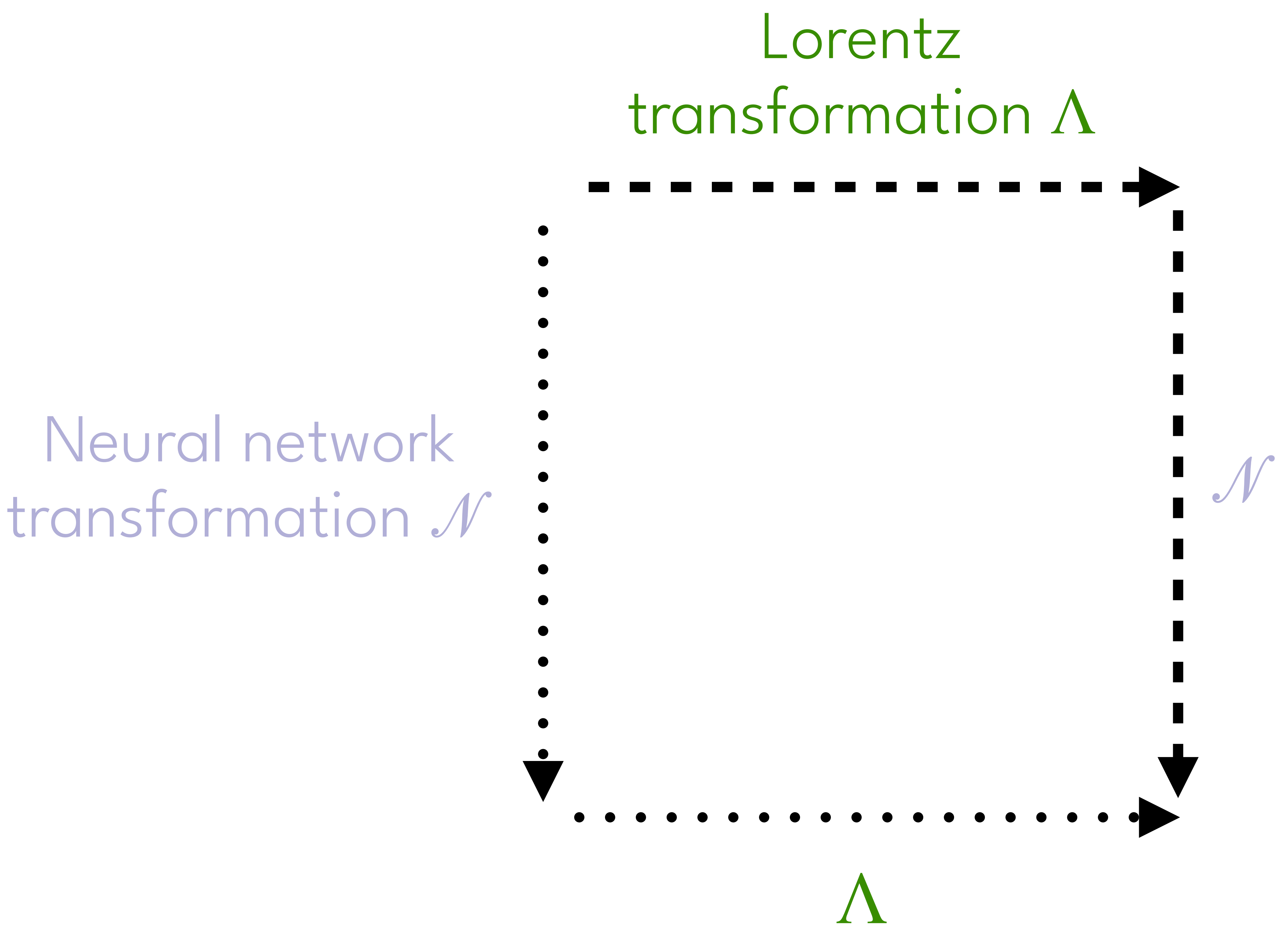
Equivariant Neural Networks

What are equivariant neural networks?

$$\mathcal{N}\left(\Lambda\left(x\right)\right) = \Lambda\left(\mathcal{N}\left(x\right)\right)$$

Why equivariance?

- Symmetries are important



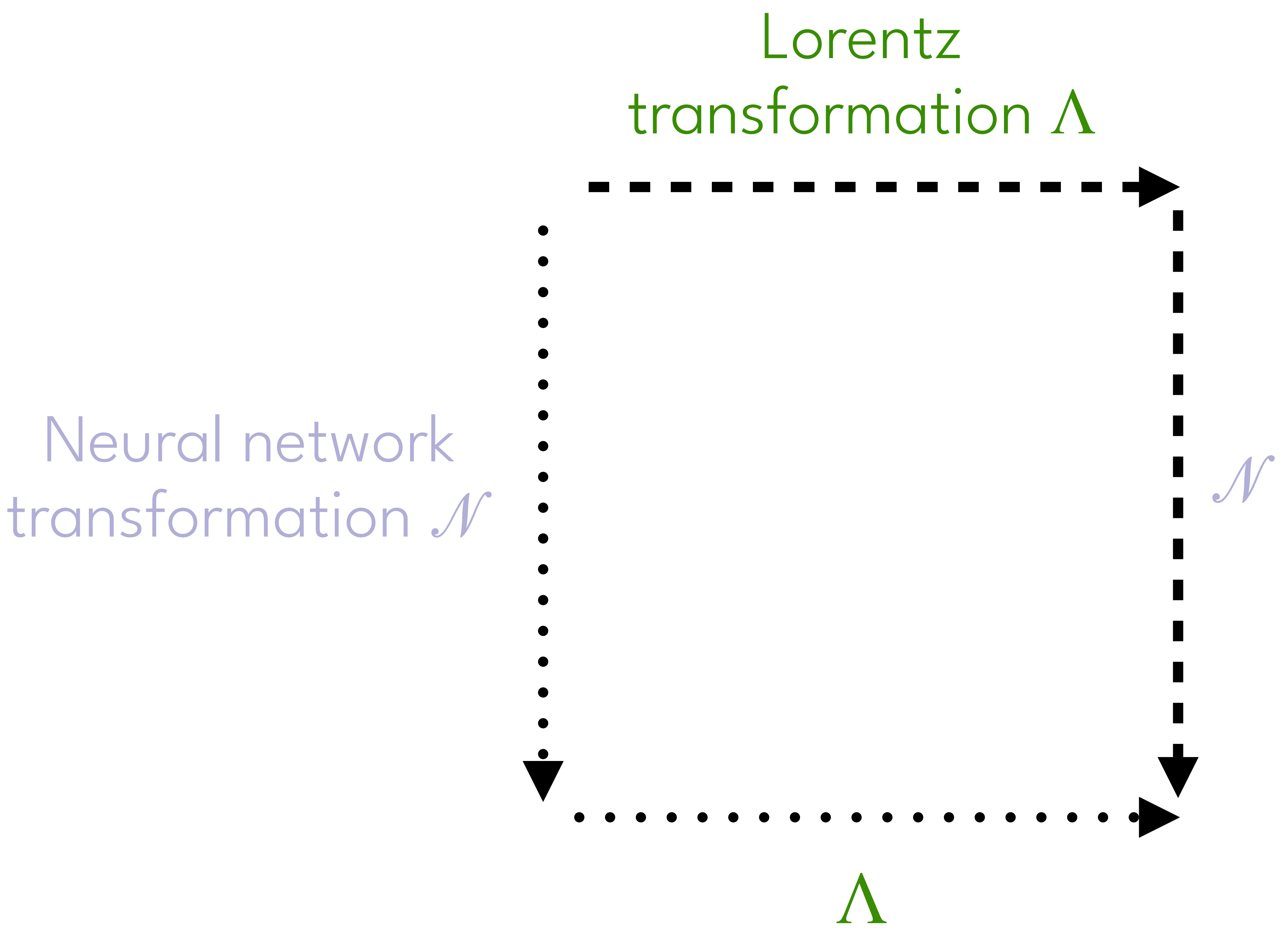
Equivariant Neural Networks

What are equivariant neural networks?

$$\mathcal{N}\left(\Lambda\left(x\right)\right) = \Lambda\left(\mathcal{N}\left(x\right)\right)$$

Why equivariance?

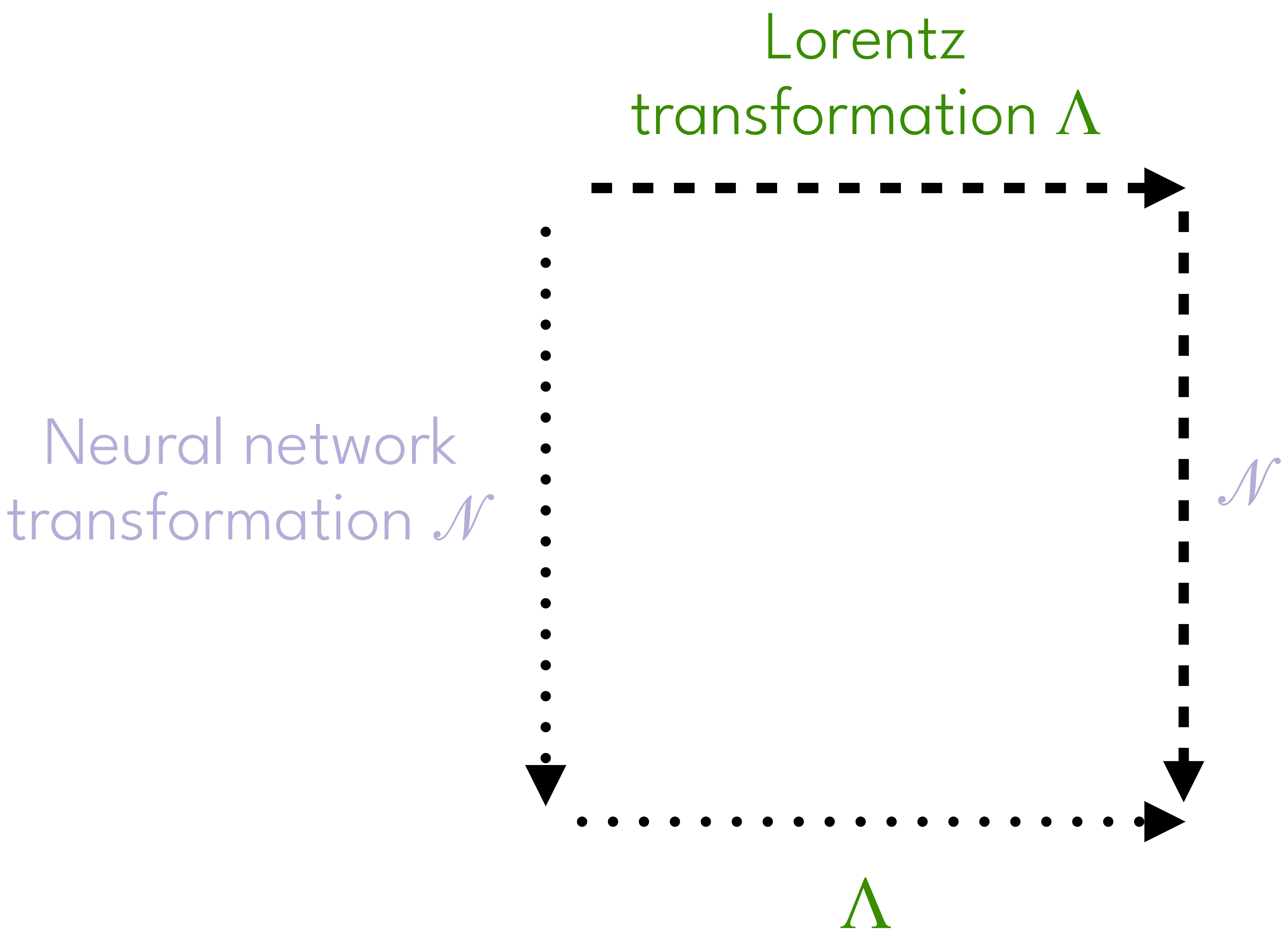
- Symmetries are important
- Symmetries are hard to learn



Equivariant Neural Networks

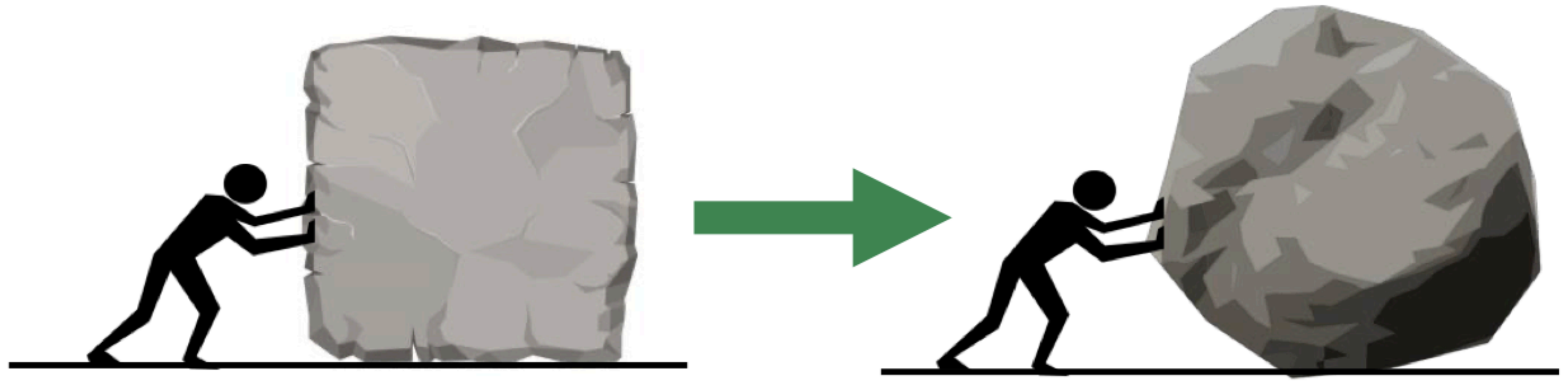
What are equivariant neural networks?

$$\mathcal{N}\left(\Lambda\left(x\right)\right) = \Lambda\left(\mathcal{N}\left(x\right)\right)$$



Why equivariance?

- Symmetries are important
- Symmetries are hard to learn
- More efficient networks



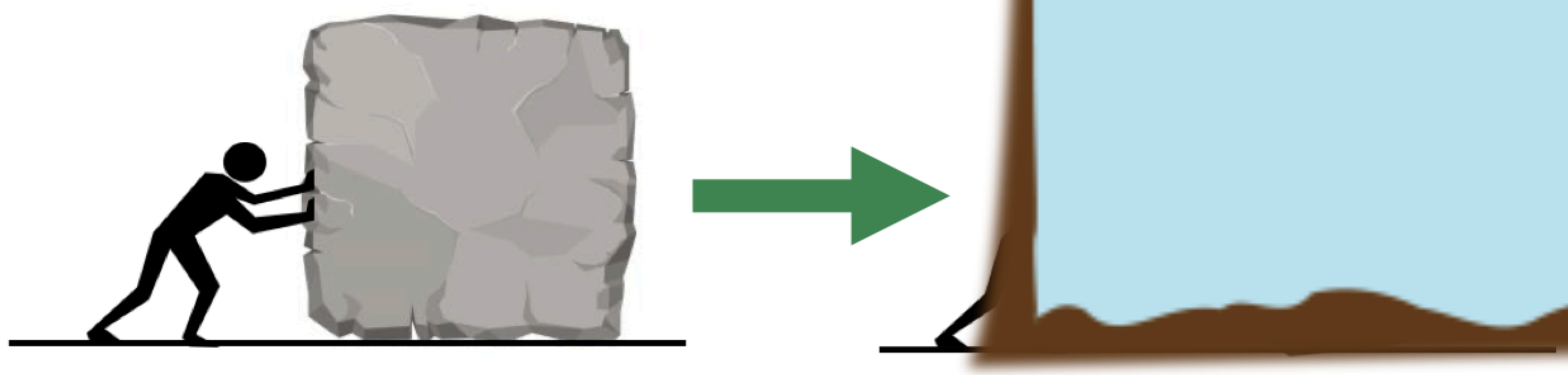
Equivariant Neural Networks

What are equivari

$$\mathcal{N}\left(\Lambda\left(x\right)\right) = \Lambda\left(\mathcal{N}\right.$$

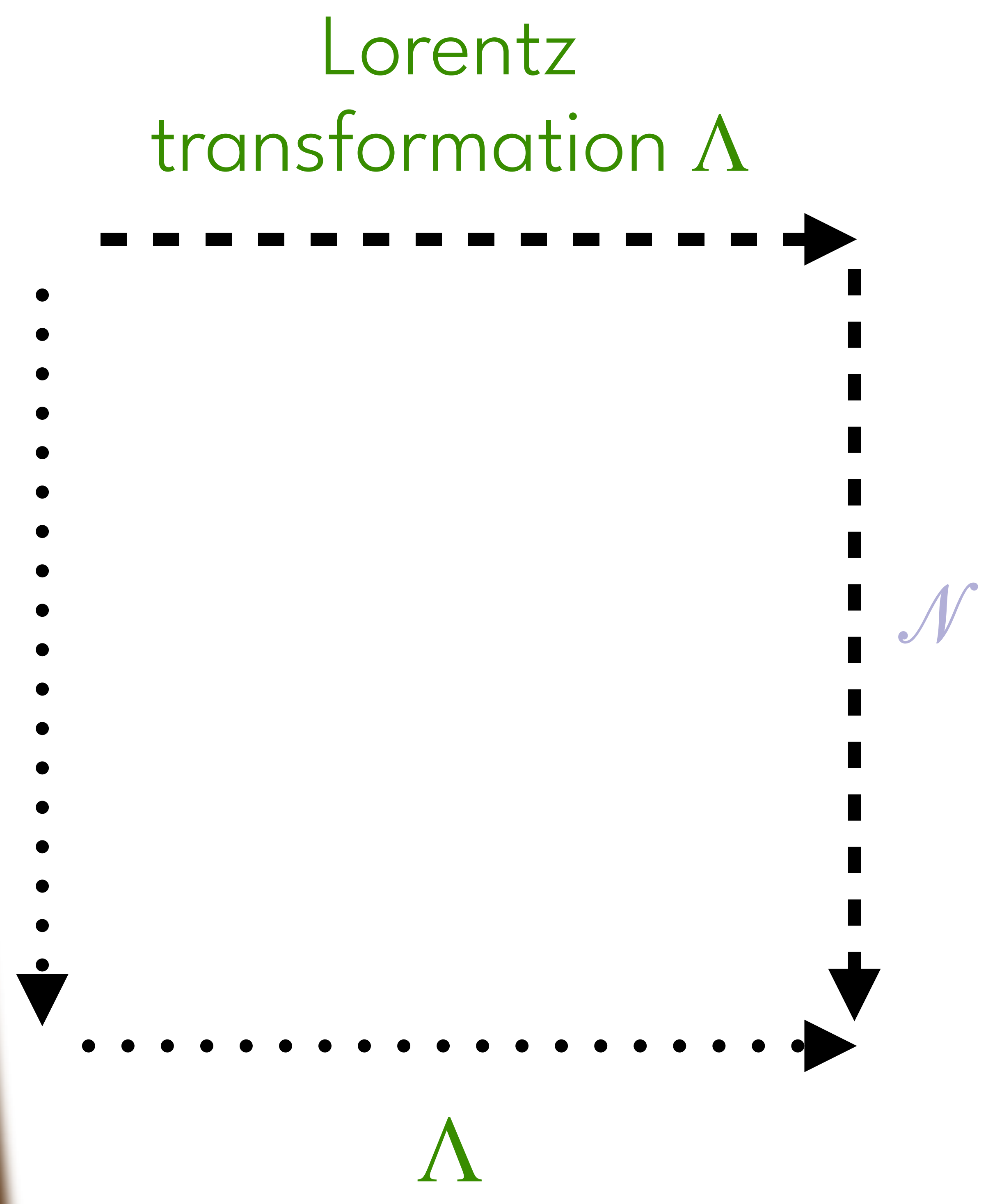
Why equivariance

- Symmetries are i
- Symmetries are h
- More efficient ne



What is our recipe?

- Geometric Algebra
- Transformer



Spacetime Algebra

- What is a **geometric algebra**?

Spacetime Algebra

- What is a **geometric algebra**?

Vector space + Geometric product

Spacetime Algebra

- What is a **geometric algebra**?

Vector space + Geometric product

- We work with the **spacetime algebra**, built from the Minkowski vector space

Spacetime Algebra

- What is a **geometric algebra**?

Vector space + Geometric product

- We work with the **spacetime algebra**, built from the Minkowski vector space

- **How do we build it?** We start with **real** vectors γ^μ with the property

$$\{\gamma^\mu, \gamma^\nu\} = 2g^{\mu\nu}$$

Spacetime Algebra

- What is a **geometric algebra**?

Vector space + Geometric product

- We work with the **spacetime algebra**, built from the Minkowski vector space

- **How do we build it?** We start with **real** vectors γ^μ with the property

$$\{\gamma^\mu, \gamma^\nu\} = 2g^{\mu\nu}$$

→ Same properties as the gamma matrices!

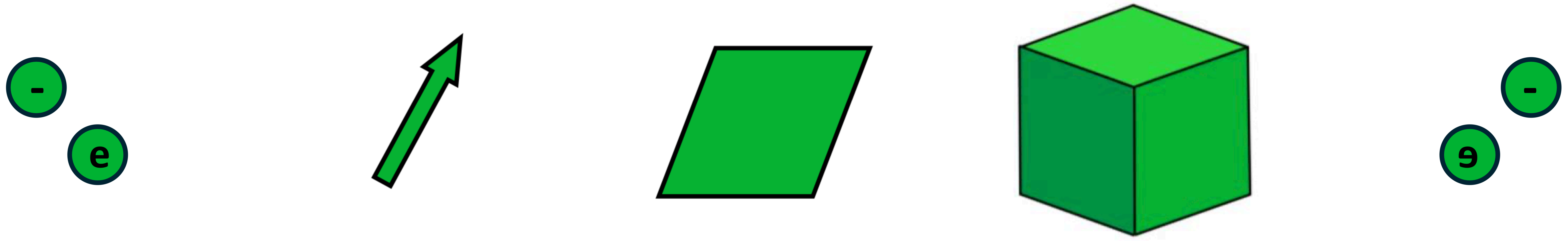
Spacetime Algebra

- **How do we build it?** We start with **real** vectors γ^μ with the property

$$\{\gamma^\mu, \gamma^\nu\} = 2g^{\mu\nu}$$

→ Same properties as the gamma matrices!

Scalar	Vector	Bivector	Axial vector	Pseudoscalar
1	γ^μ	$\sigma^{\mu\nu}$	$\gamma^\mu\gamma^5$	γ^5



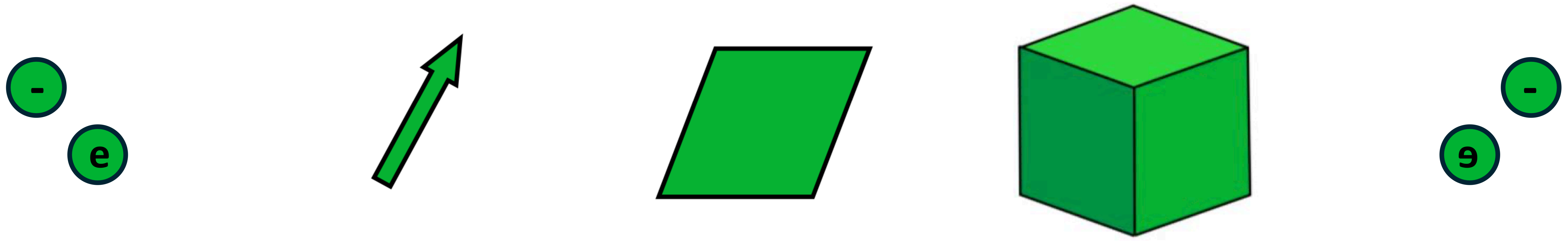
Spacetime Algebra

- **How do we build it?** We start with **real** vectors γ^μ with the property

$$\{\gamma^\mu, \gamma^\nu\} = 2g^{\mu\nu}$$

→ Same properties as the gamma matrices!

Scalar | Vector | Bivector | Axial vector | Pseudoscalar

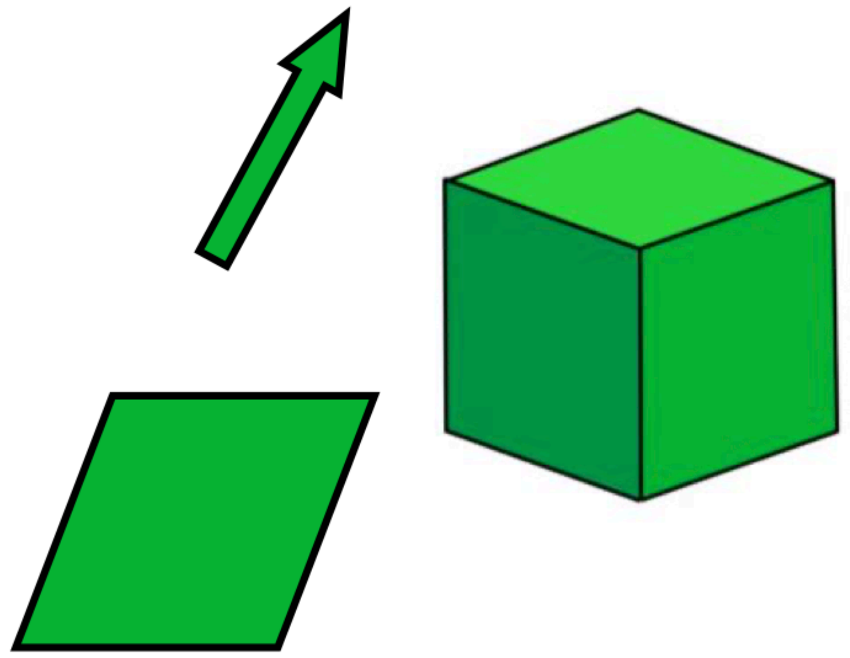


$$\underbrace{1 + \gamma^\mu + \sigma^{\mu\nu} + \gamma^\mu\gamma^5 + \gamma^5}_{\textbf{Multivectors}}$$

Multivectors

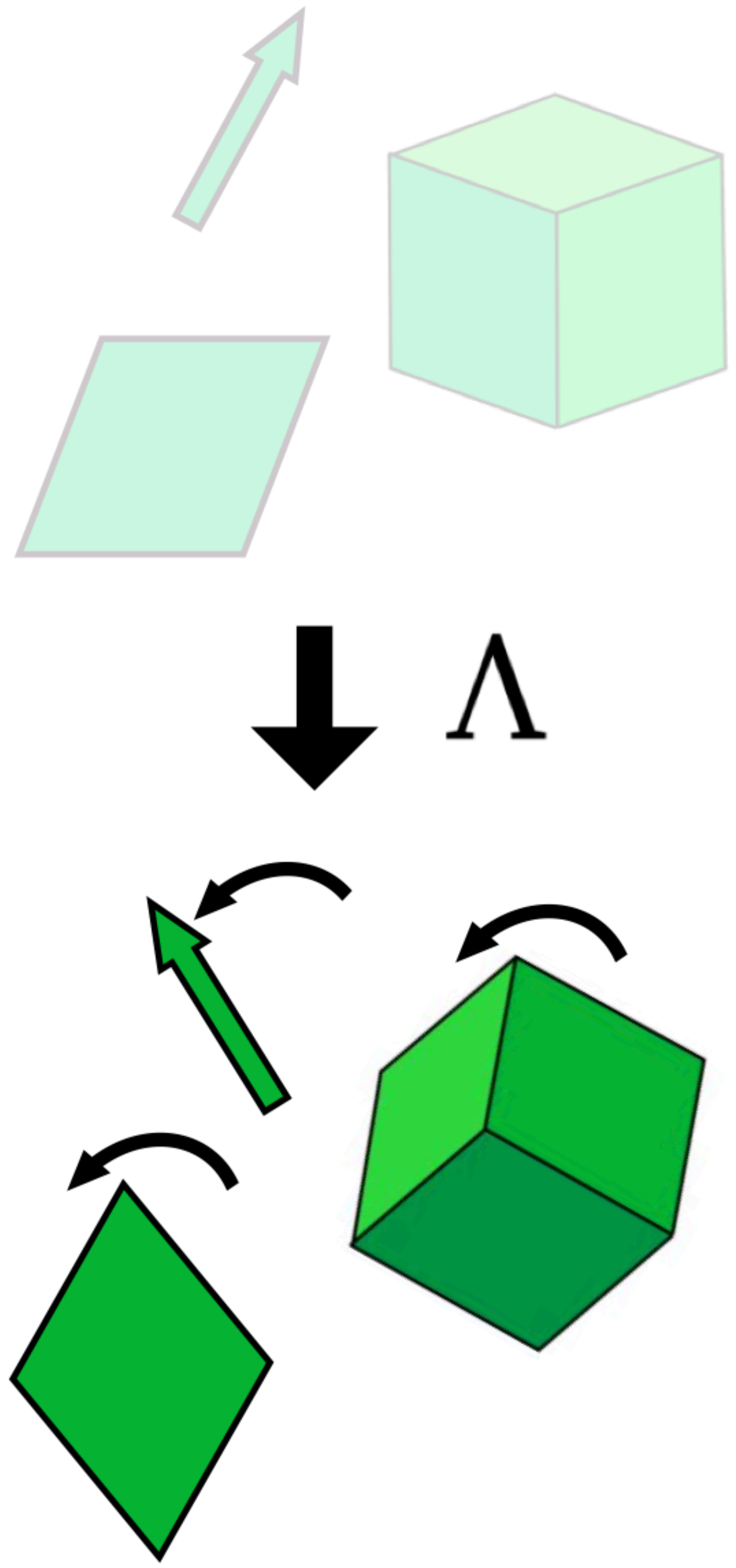
Spacetime Algebra

- Lorentz transformations:



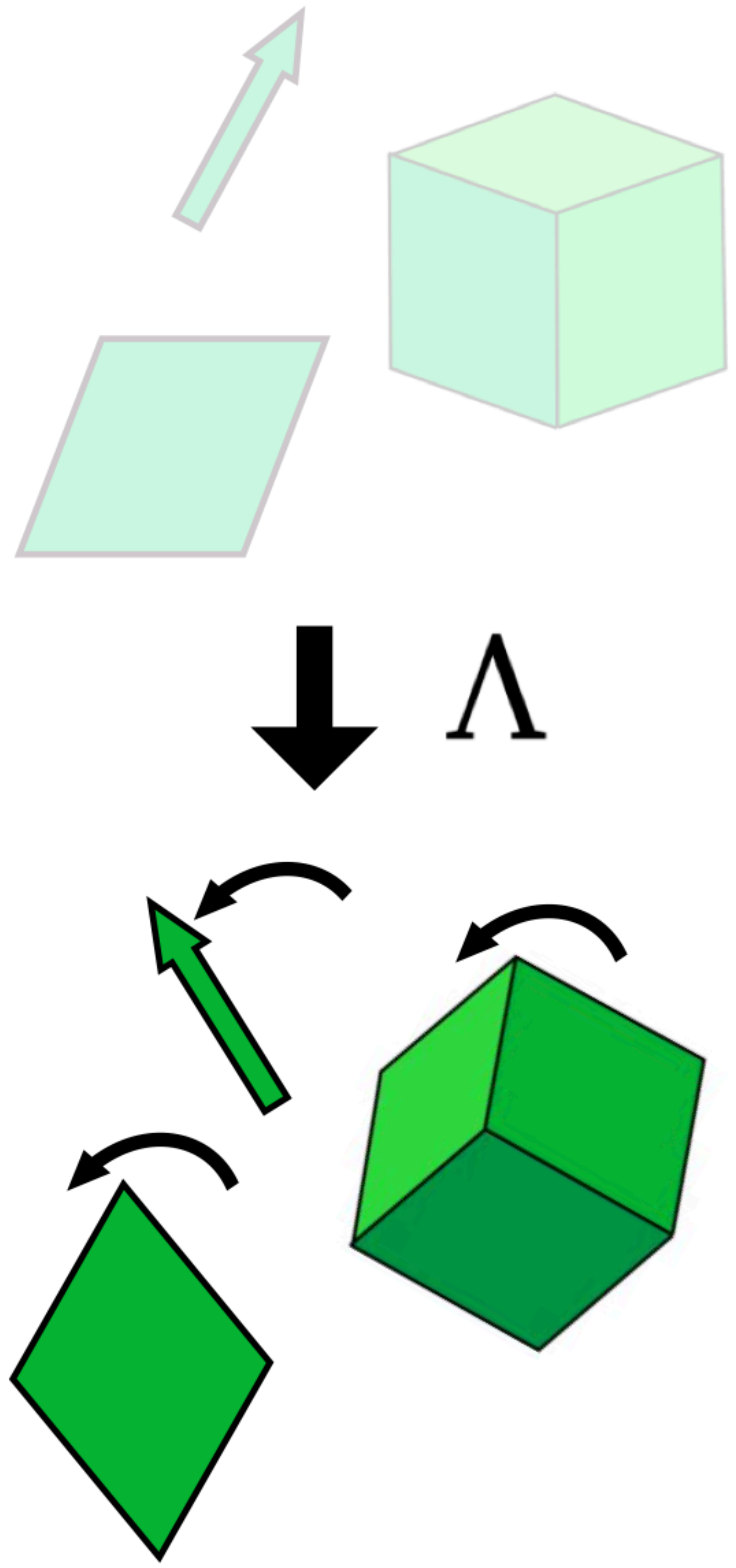
Spacetime Algebra

- Lorentz transformations:



Spacetime Algebra

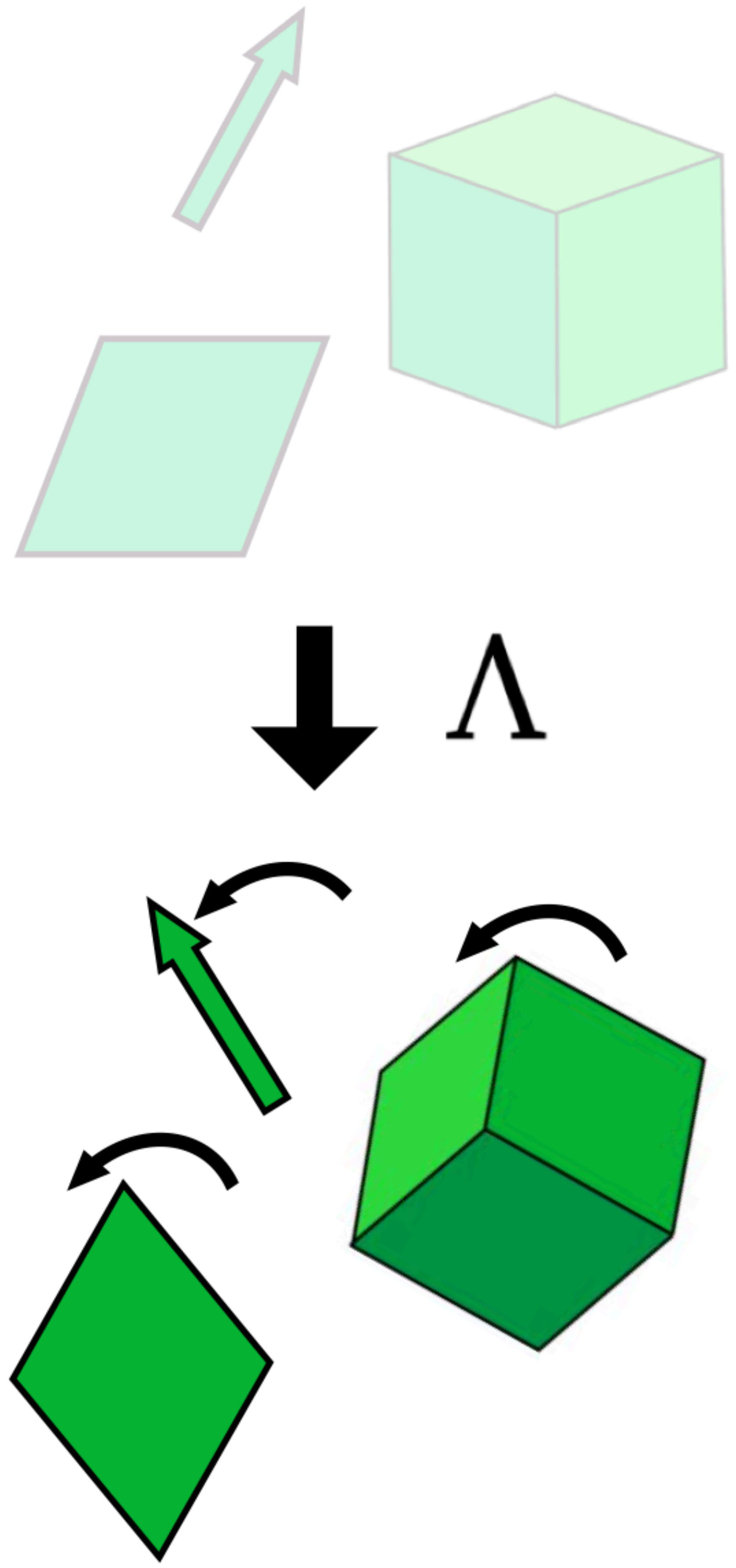
- Lorentz transformations:



Each grade
transforms separately

Spacetime Algebra

- Lorentz transformations:

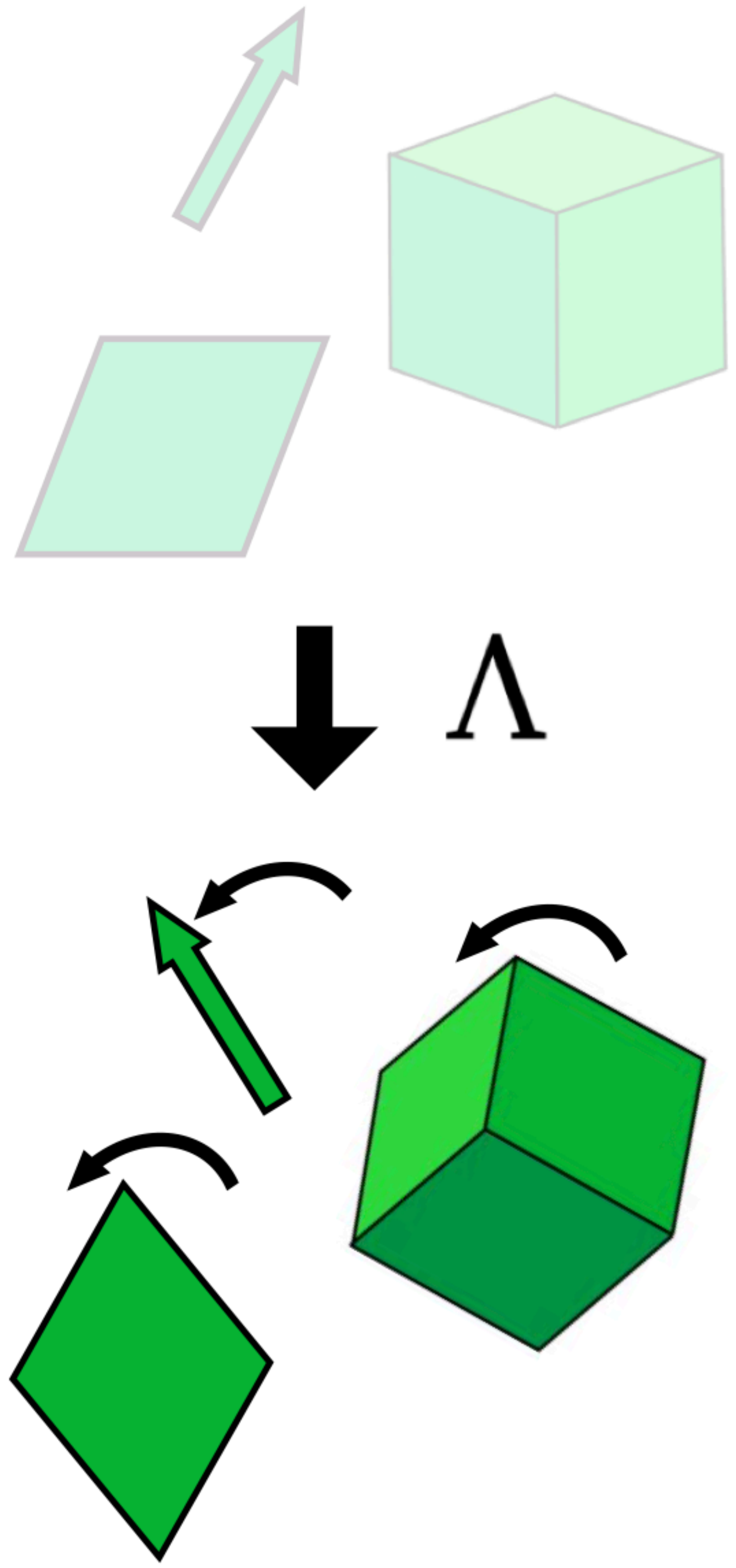


Each grade
transforms separately

Equivariant maps never
mix grade components

Spacetime Algebra

- Lorentz transformations:



Each grade transforms separately

↓

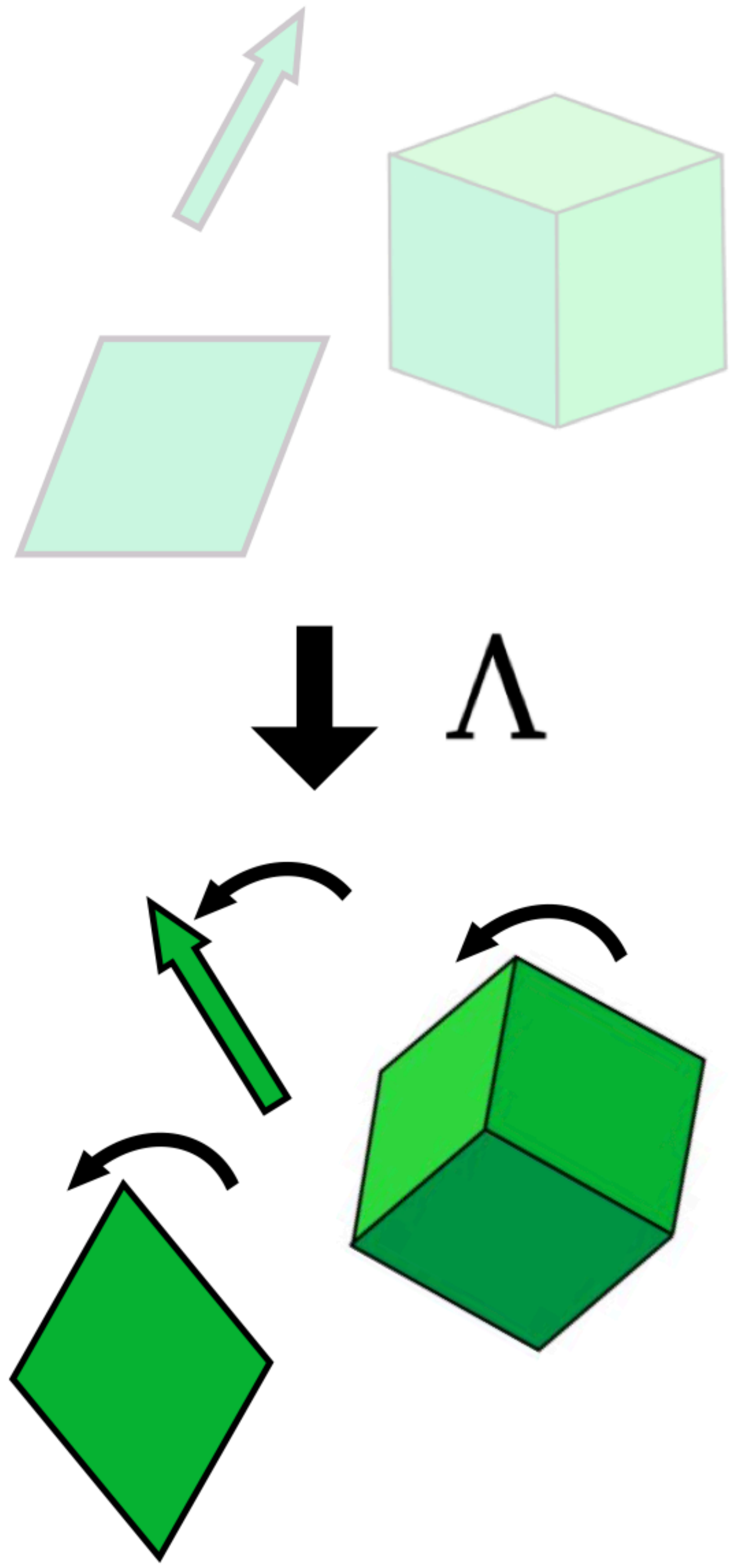
Equivariant maps never mix grade components

↓

Easy to build equivariant versions of ordinary networks

Spacetime Algebra

- Lorentz transformations:



Each grade transforms separately

↓

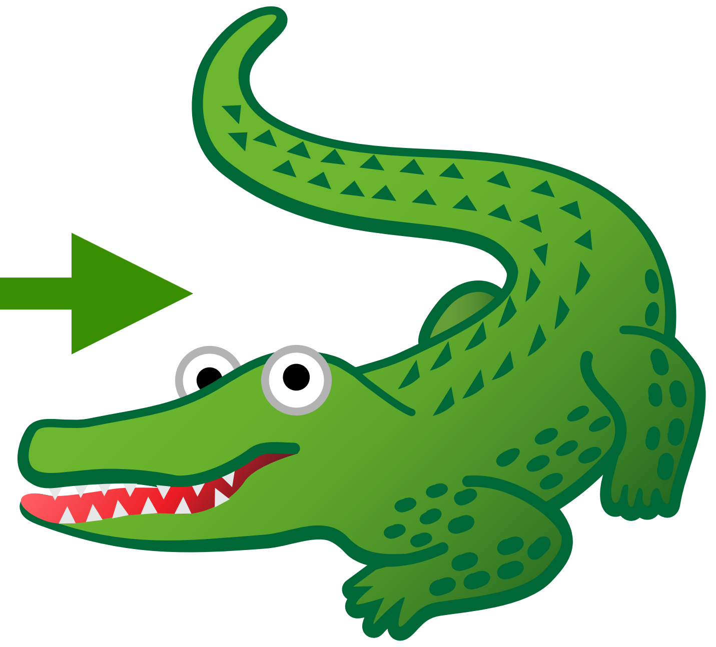
Equivariant maps never mix grade components

↓

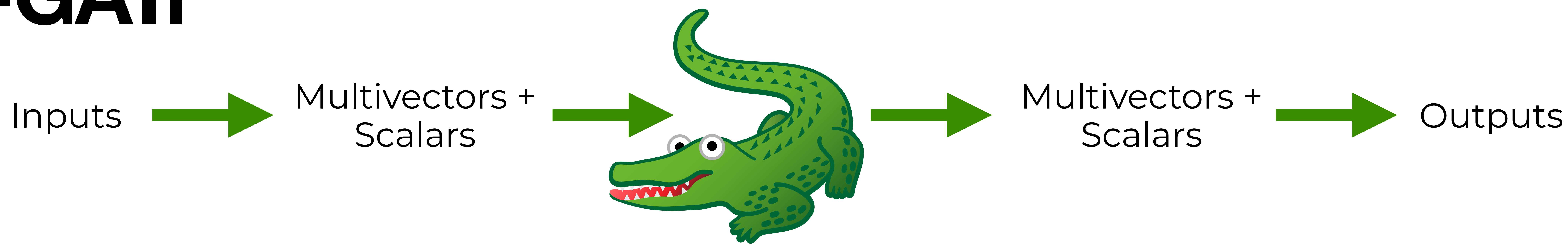
Easy to build equivariant versions of ordinary networks

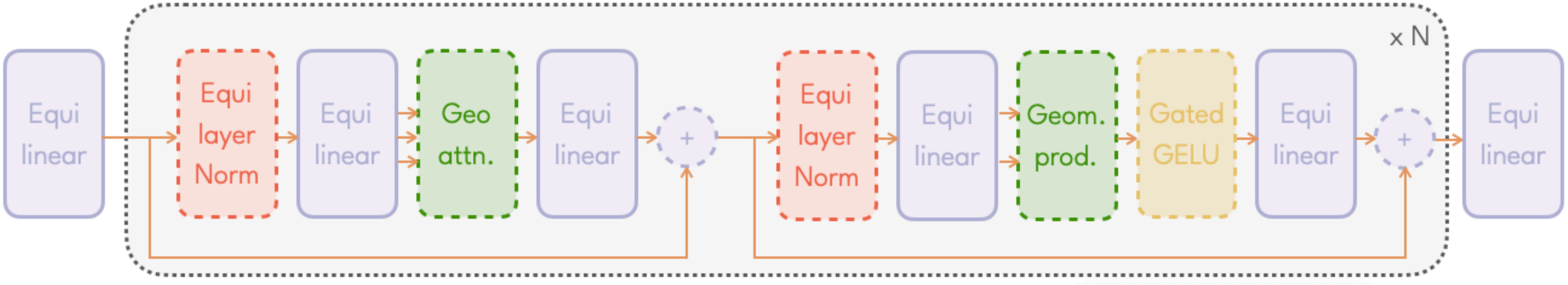
Lorentz Equivariance + Geometric Algebra + Transformer = **L-GATr**

L-GATr

Inputs → Multivectors + Scalars →  → Multivectors + Scalars → Outputs

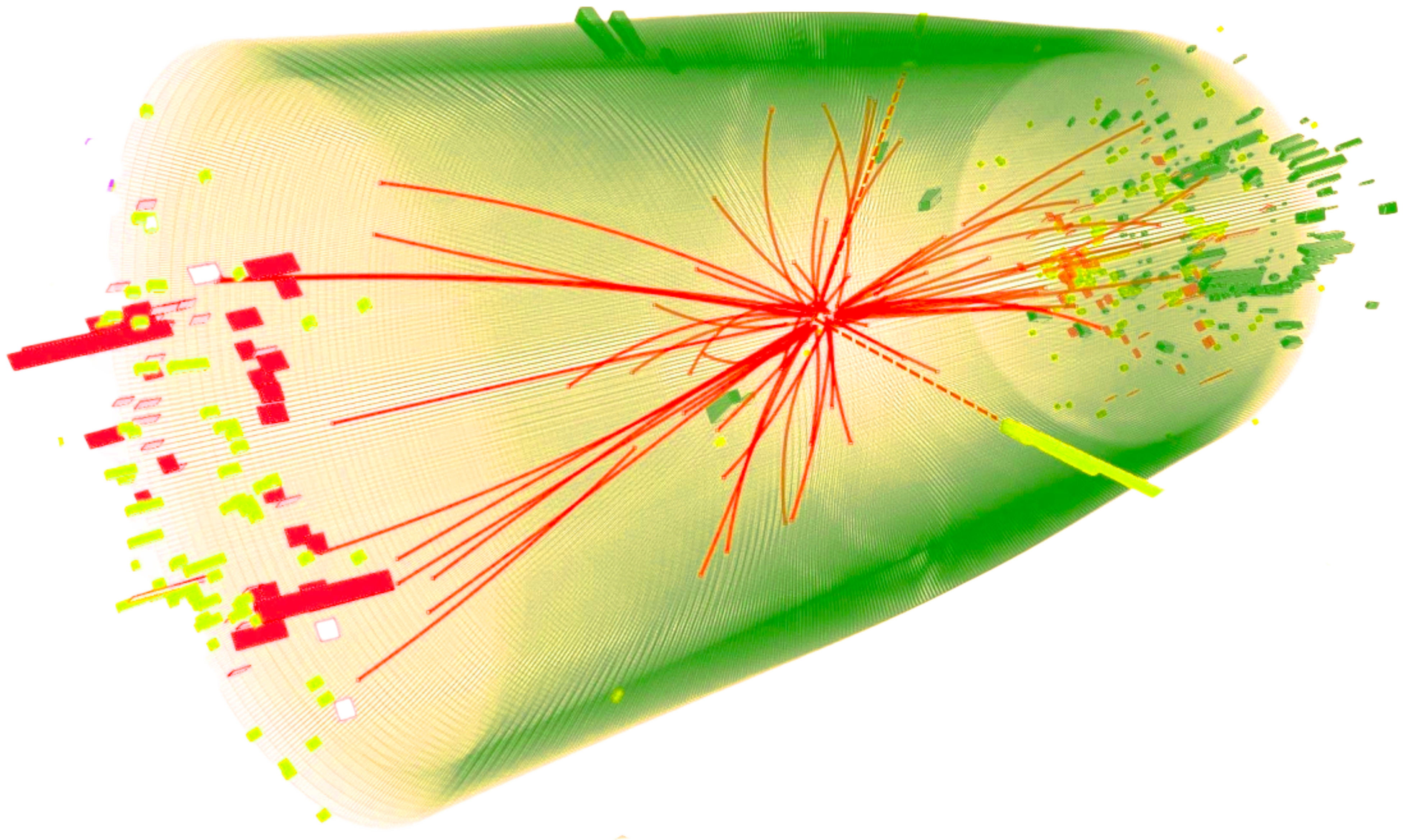
L-GATr


Inputs
Multivectors + Scalars
Multivectors + Scalars
Outputs


Equi linear
Equi layer Norm
Equi linear
Geo attn.
Equi linear
+
Equi layer Norm
Equi linear
Geom. prod.
Gated GELU
Equi linear
+
x N
Equi linear

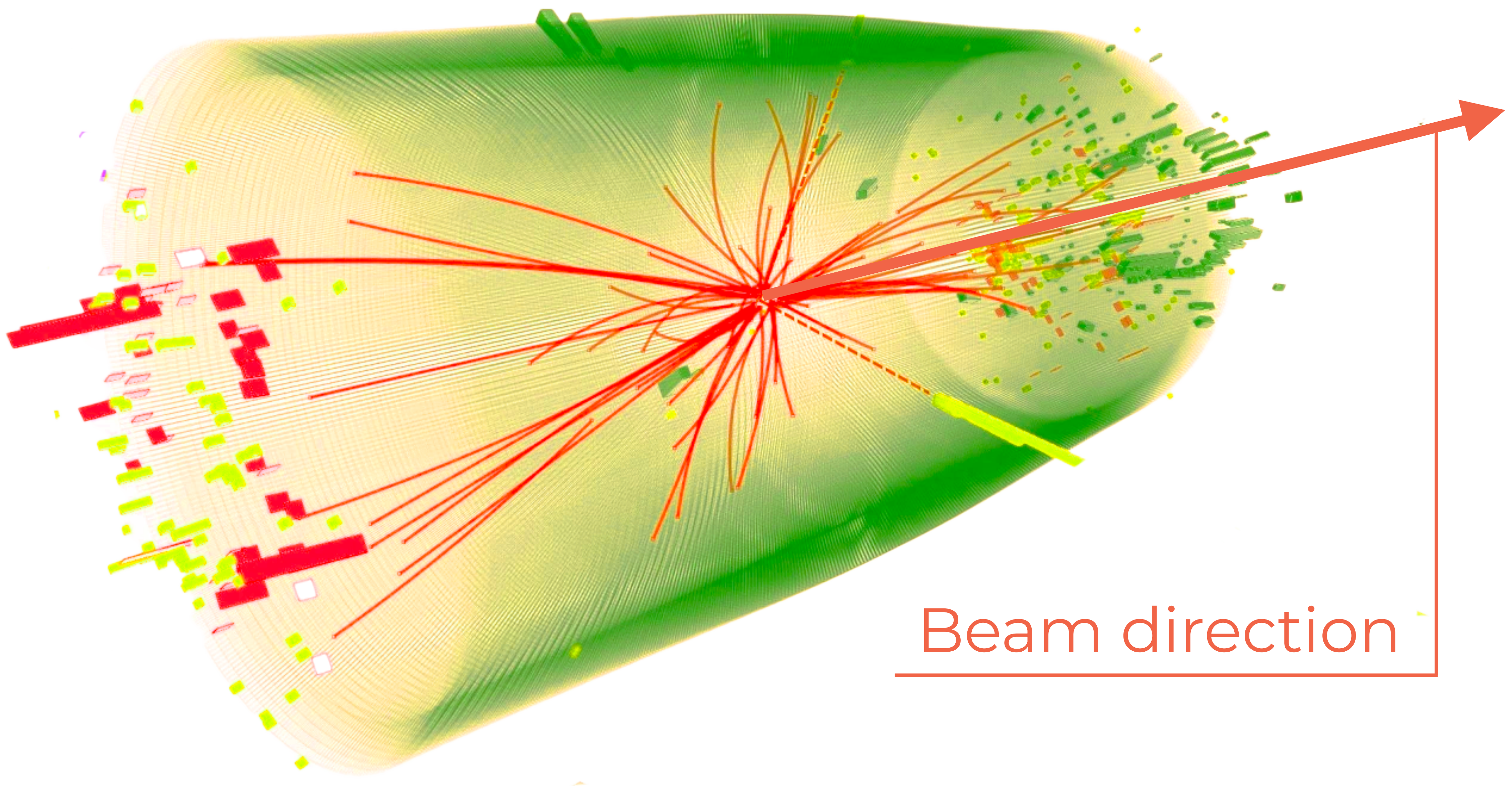
L-GATr

Key feature: Symmetry breaking



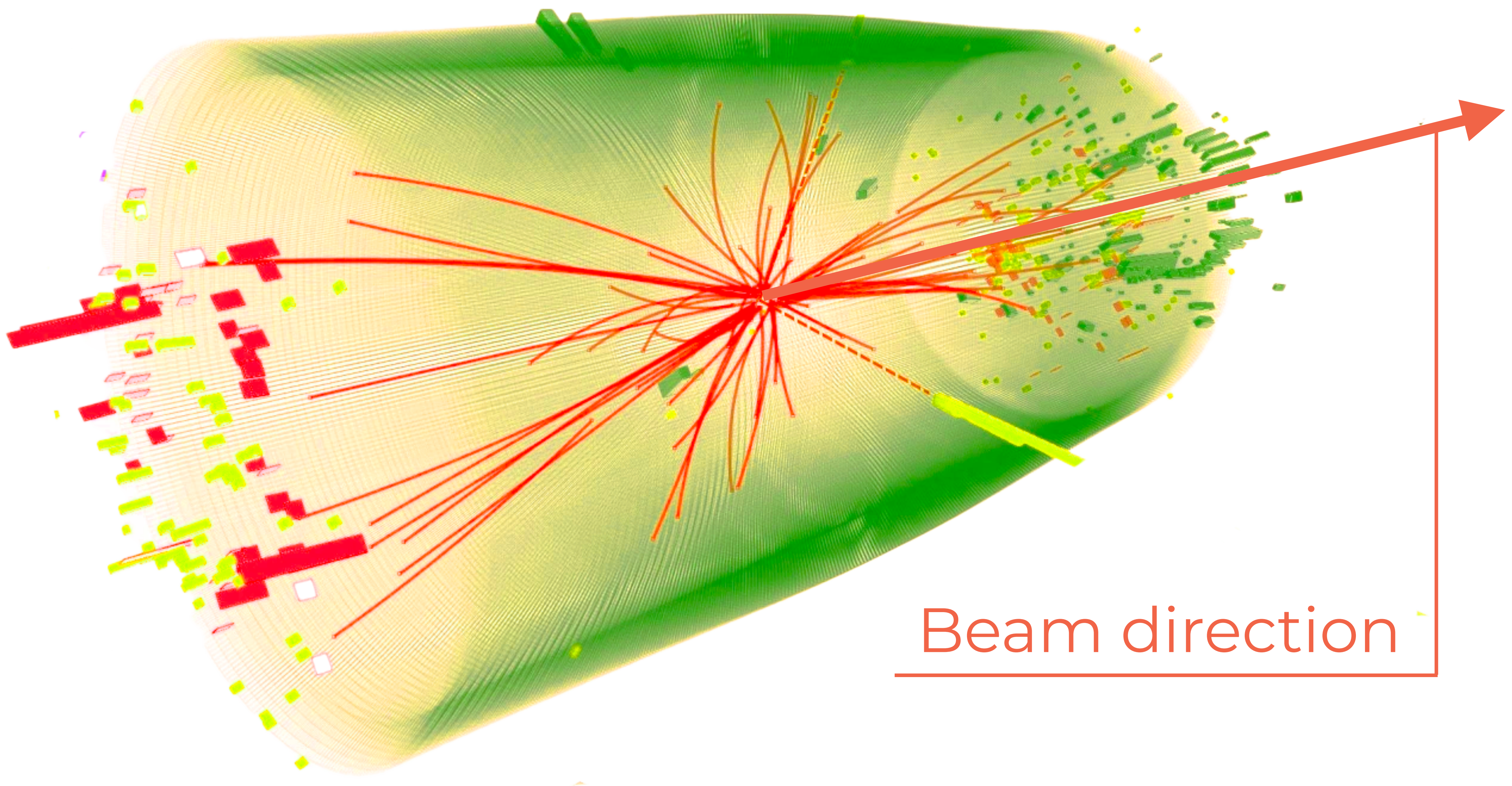
L-GATr

Key feature: Symmetry breaking



L-GATr

Key feature: Symmetry breaking



- Beam reference: $(1, 0, 0, \pm 1)$

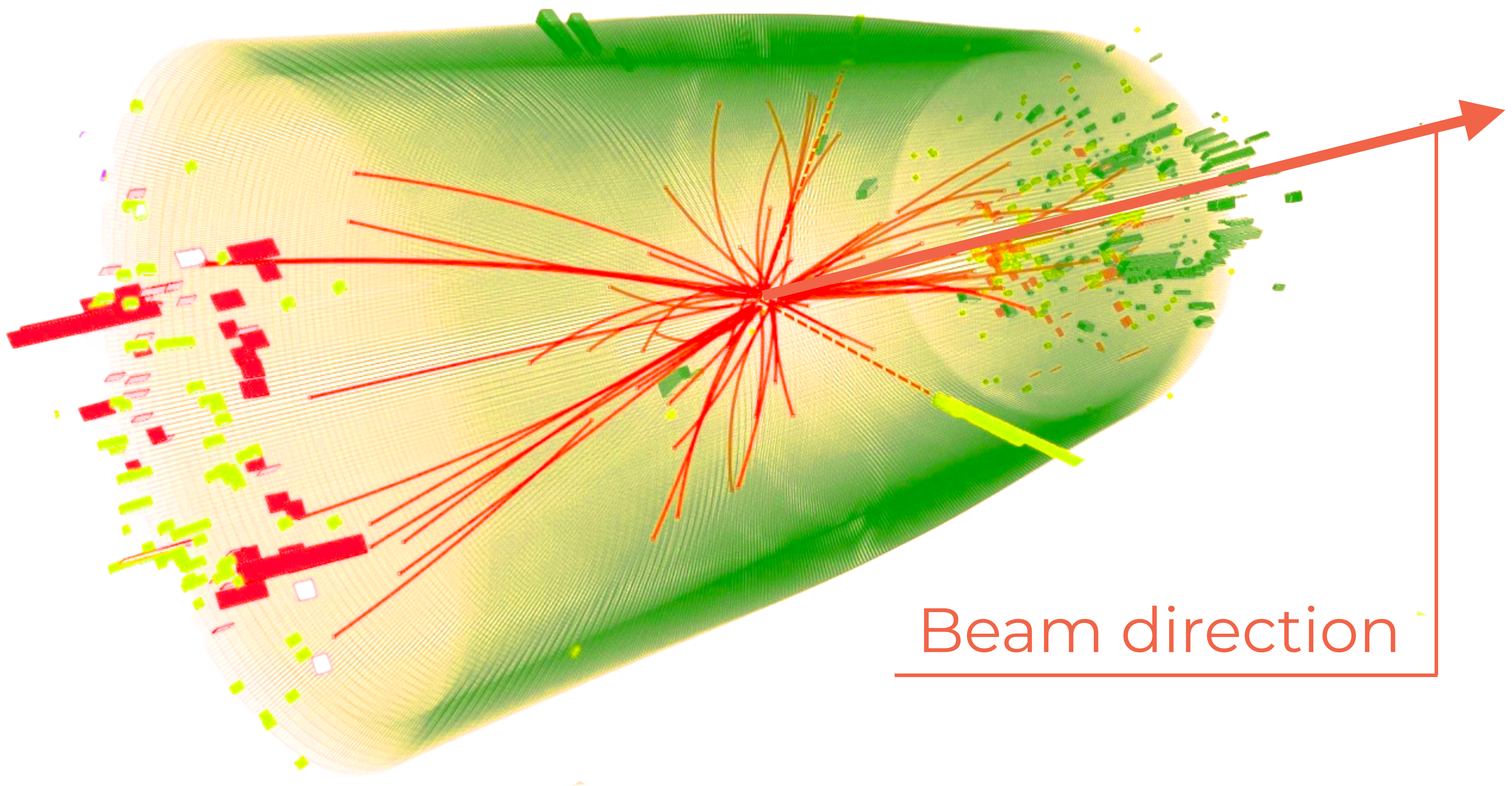
$SO(1,3) \rightarrow$ Boosts + rotations around the beam

L-GATr

Key feature: Symmetry breaking

- Beam reference: $(1, 0, 0, \pm 1)$

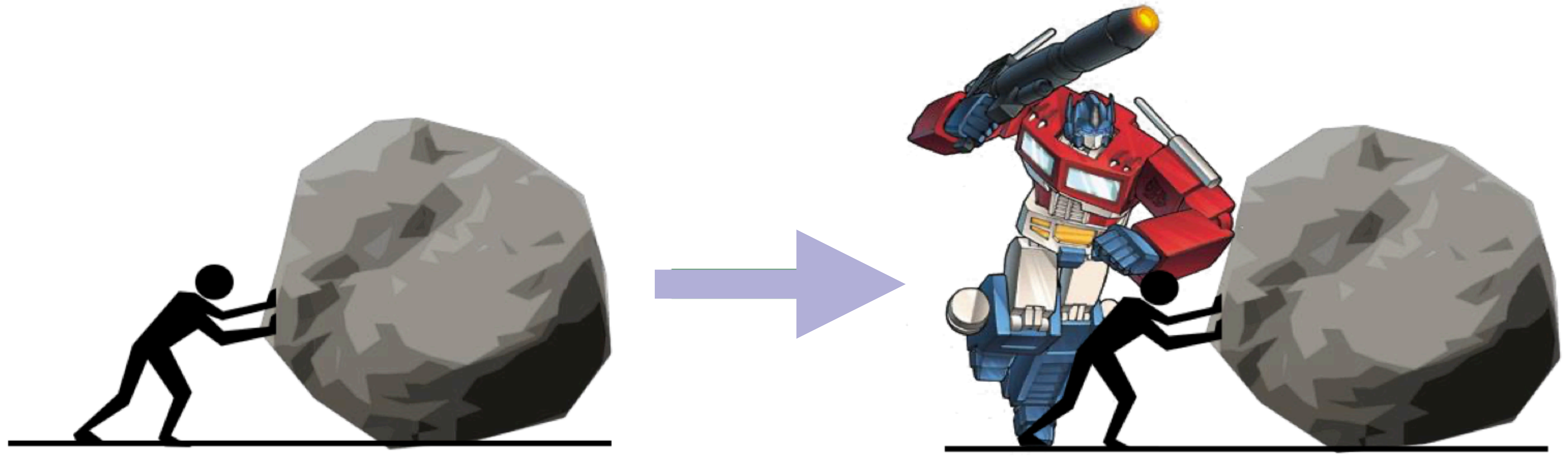
$SO(1,3) \rightarrow$ Boosts + rotations around the beam

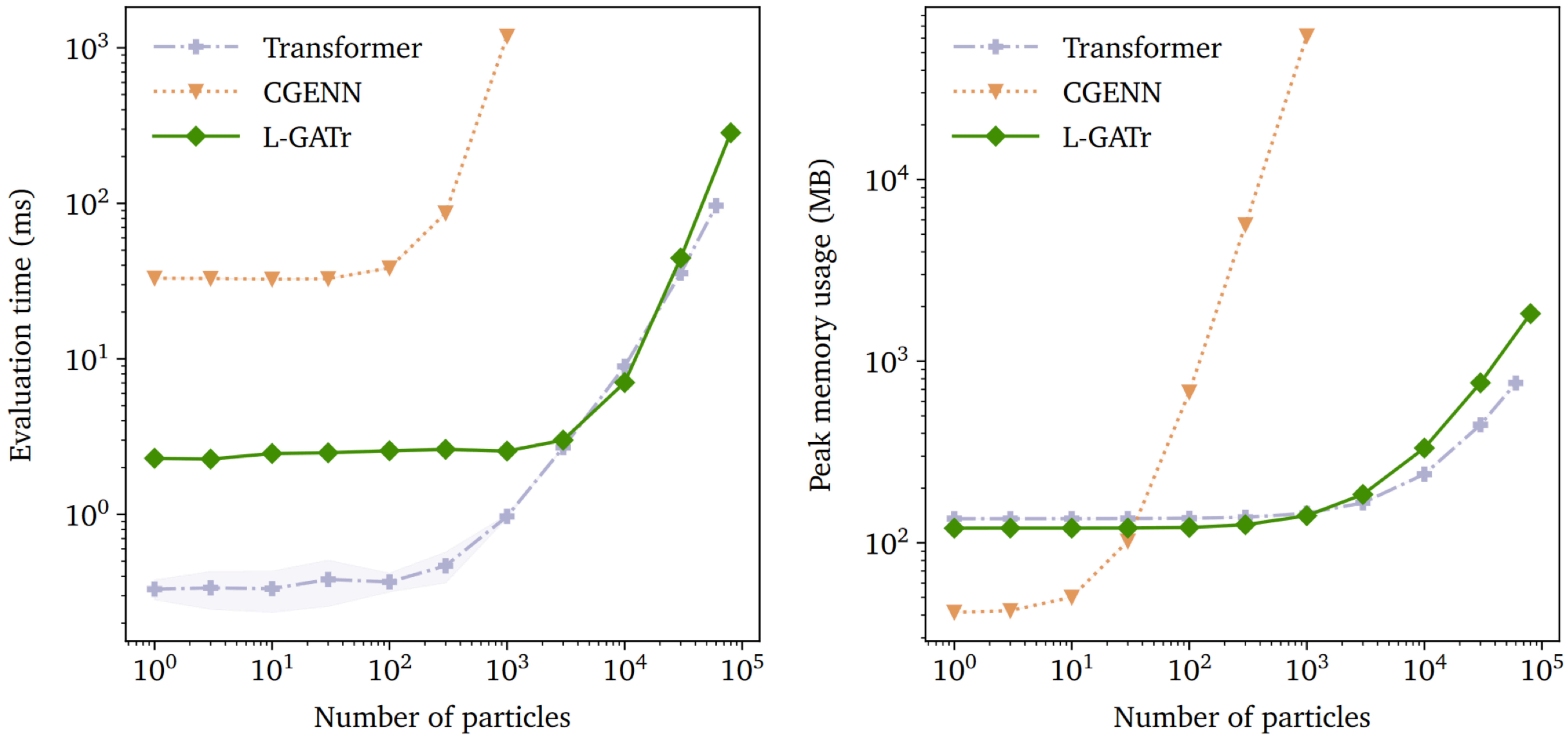


Beam	Time	Embedding	$1/\epsilon_B$ $(\epsilon_S = 0.3)$
–	✗	Particle	1422
Spacelike	✗	Particle	1905
All planes	✓	Particle	2009
–	✓	Token	1923
xy plane	✓	Channel	2060
Spacelike	✓	Particle	2152
Lightlike	✓	Particle	2114
xy plane	✓	Particle	2240

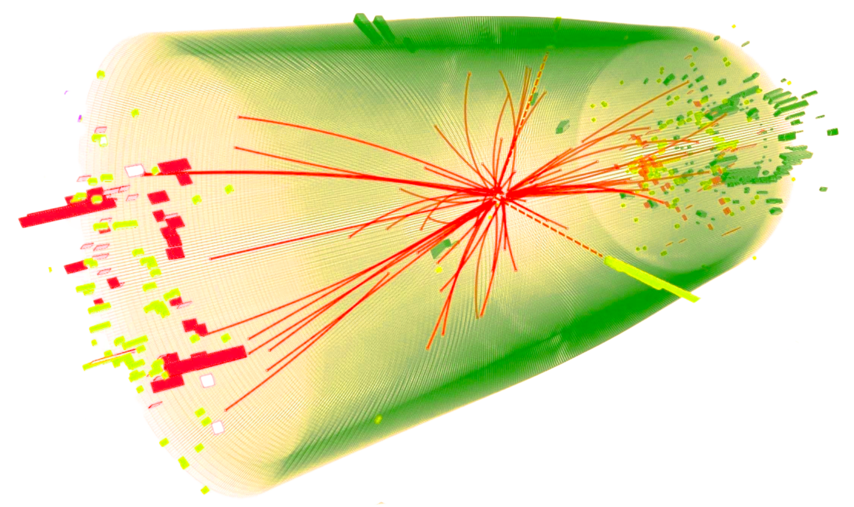
L-GATr

Key feature: Transformer scaling





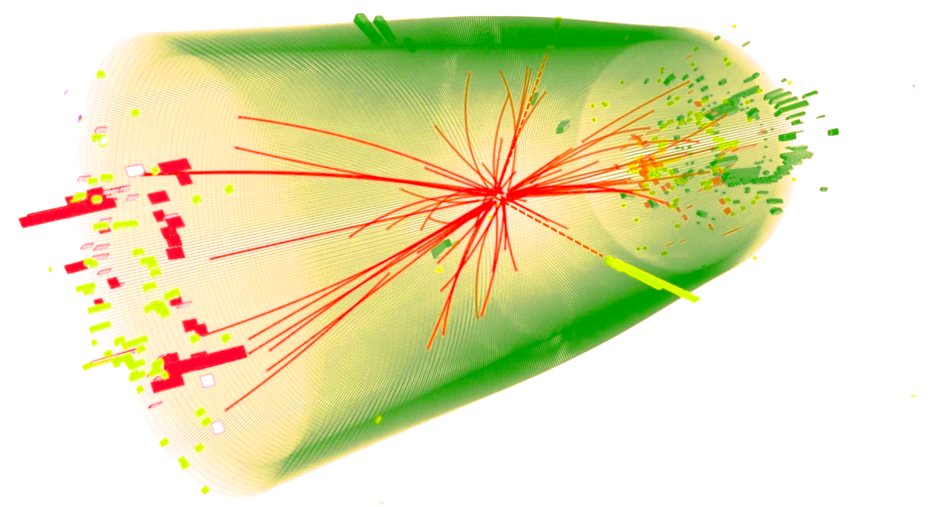
Experiment: Jet Tagging



Top Tagging

Network	Accuracy	AUC	$1/\epsilon_B$ ($\epsilon_S = 0.5$)	$1/\epsilon_B$ ($\epsilon_S = 0.3$)
TopoDNN [52]	0.916	0.972	–	295 ± 5
LoLa [9]	0.929	0.980	–	722 ± 17
N-subjettiness [53]	0.929	0.981	–	867 ± 15
PFN [54]	0.932	0.9819	247 ± 3	888 ± 17
TreeNiN [55]	0.933	0.982	–	1025 ± 11
ParticleNet [56]	0.940	0.9858	397 ± 7	1615 ± 93
ParT [57]	0.940	0.9858	413 ± 16	1602 ± 81
MIParT [58]	0.942	0.9868	505 ± 8	2010 ± 97
LorentzNet* [10]	0.942	0.9868	498 ± 18	2195 ± 173
CGENN* [12]	0.942	0.9869	500	2172
PELICAN* [40]	0.9426 ± 0.0002	0.9870 ± 0.0001	–	2250 ± 75
L-GATr* [33]	0.9423 ± 0.0002	0.9870 ± 0.0001	540 ± 20	2240 ± 70

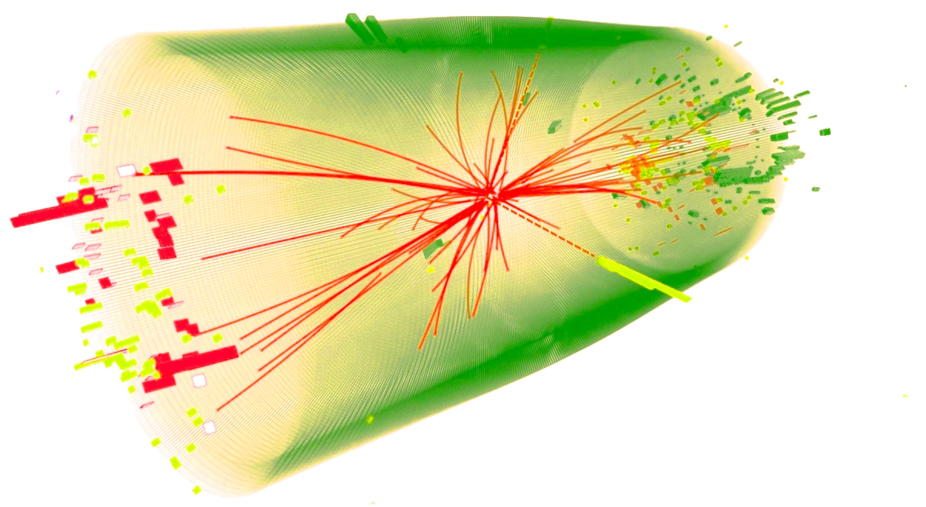
Experiment: Jet Tagging



JetClass Tagging

- Large and comprehensive jet dataset
- 100M events
- 10 classes

Experiment: Jet Tagging

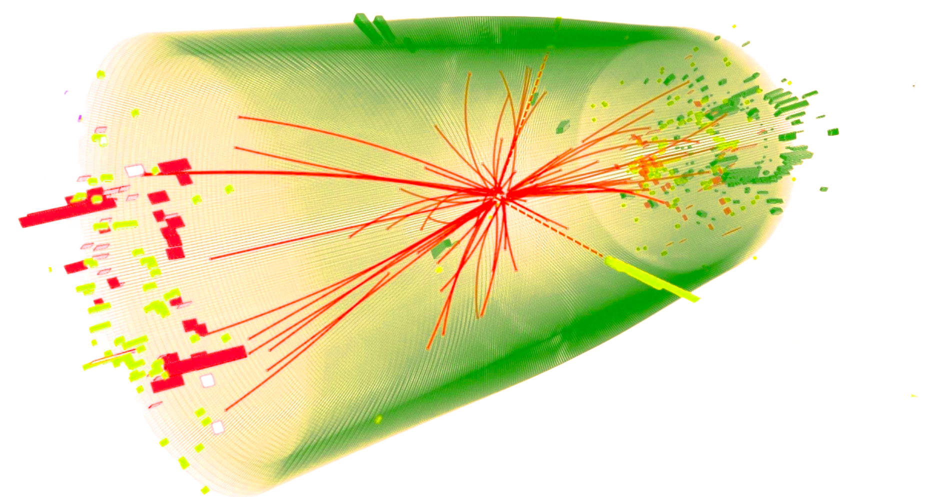


JetClass Tagging

- Large and comprehensive jet dataset
- 100M events
- 10 classes

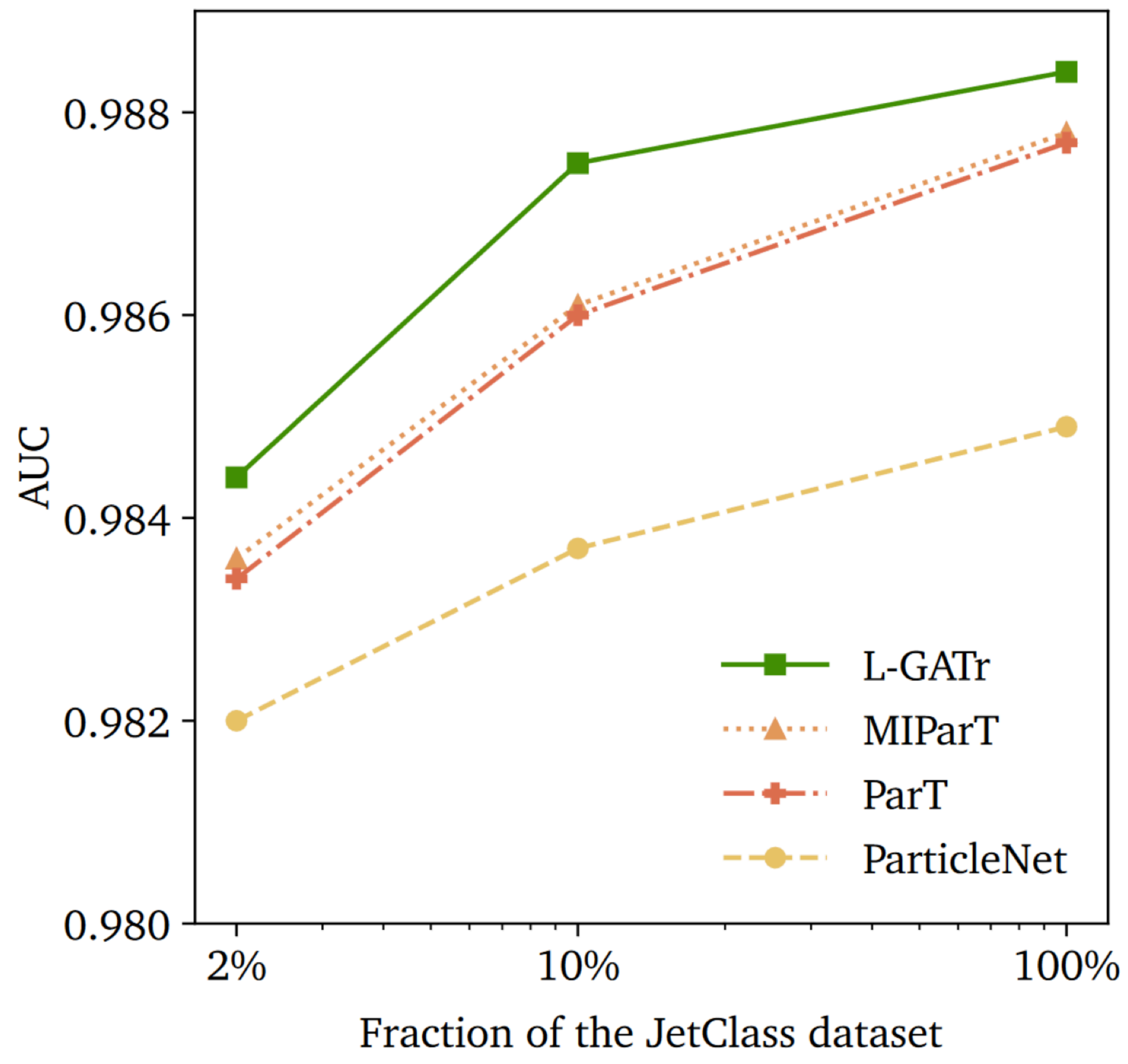
	All classes Accuracy	AUC	$H \to b\bar{b}$ $\text{Rej}_{50\%}$	$H \to c\bar{c}$ $\text{Rej}_{50\%}$	$H \to gg$ $\text{Rej}_{50\%}$	$H \to 4q$ $\text{Rej}_{50\%}$	$H \to l\nu q\bar{q}'$ $\text{Rej}_{99\%}$	$t \to bq\bar{q}'$ $\text{Rej}_{50\%}$	$t \to bl\nu$ $\text{Rej}_{99.5\%}$	$W \to q\bar{q}'$ $\text{Rej}_{50\%}$	$Z \to q\bar{q}$ $\text{Rej}_{50\%}$
ParticleNet [56]	0.844	0.9849	7634	2475	104	954	3339	10526	11173	347	283
ParT [57]	0.861	0.9877	10638	4149	123	1864	5479	32787	15873	543	402
MIParT [58]	0.861	0.9878	10753	4202	123	1927	5450	31250	16807	542	402
L-GATr	0.865	0.9884	12195	4819	128	2304	5764	37736	19231	580	427

Experiment: Jet Tagging



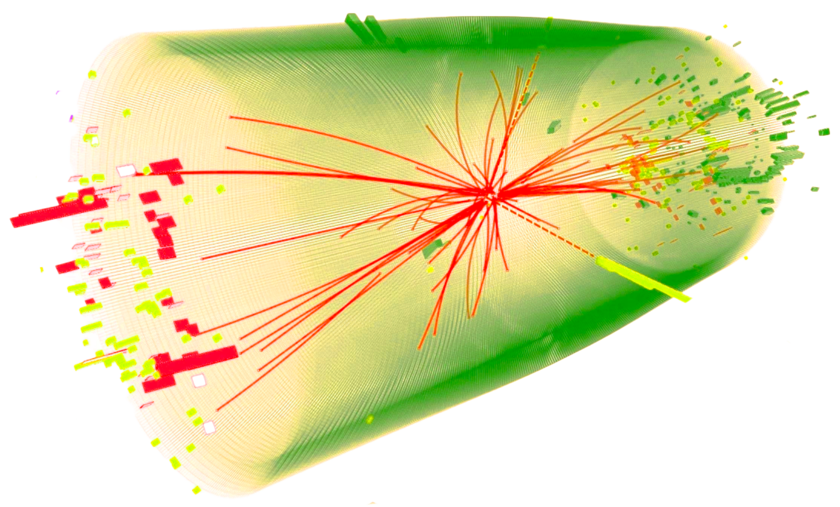
JetClass Tagging

- Large and comprehensive jet dataset
- 100M events
- 10 classes

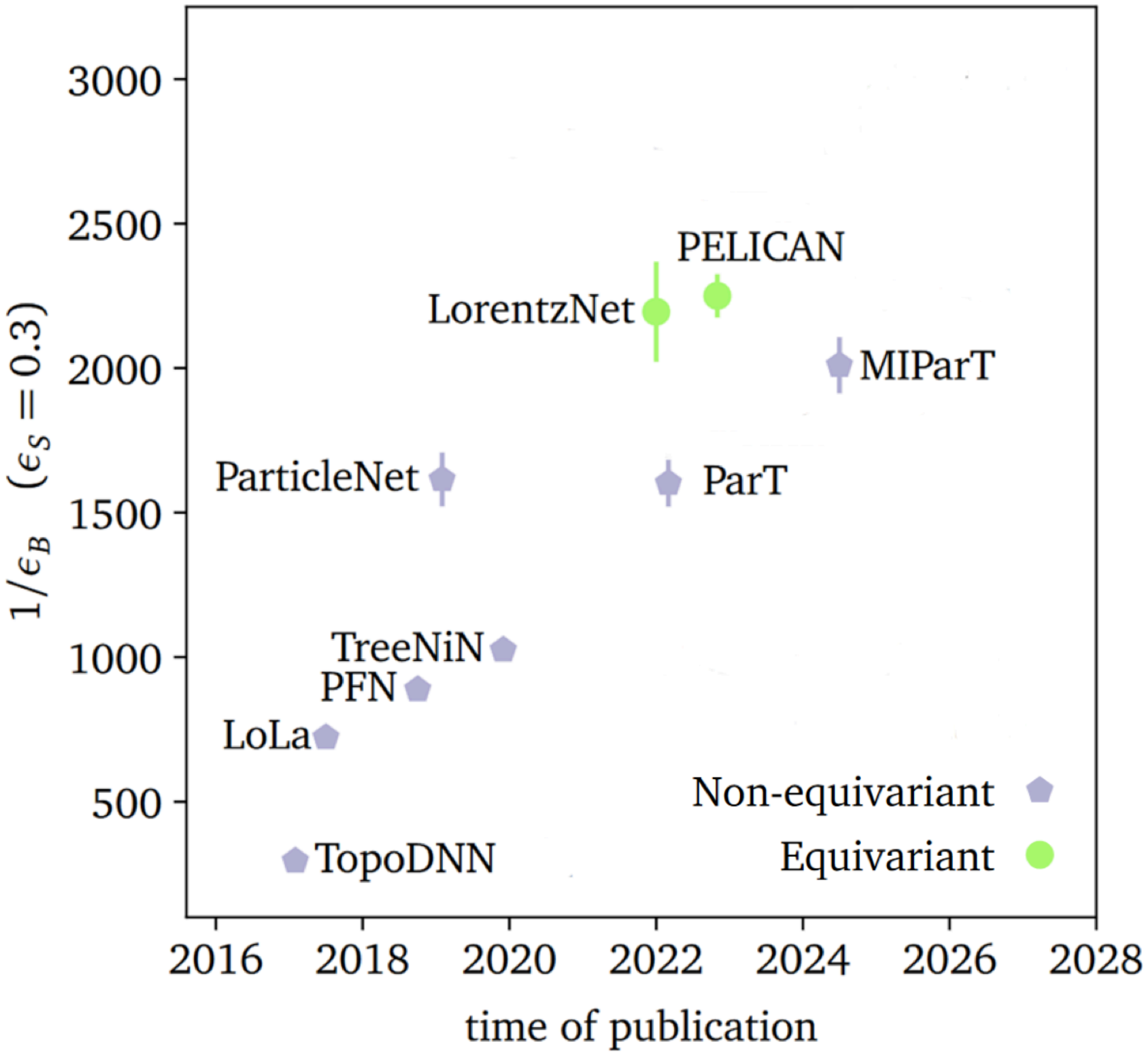


	All classes Accuracy	All classes AUC	$H \to b\bar{b}$ $\mathrm{Rej}_{50\%}$	$H \to c\bar{c}$ $\mathrm{Rej}_{50\%}$	$H \to gg$ $\mathrm{Rej}_{50\%}$	$H \to 4q$ $\mathrm{Rej}_{50\%}$	$H \to l\nu q\bar{q}'$ $\mathrm{Rej}_{99\%}$	$t \to bq\bar{q}'$ $\mathrm{Rej}_{50\%}$	$t \to bl\nu$ $\mathrm{Rej}_{99.5\%}$	$W \to q\bar{q}'$ $\mathrm{Rej}_{50\%}$	$Z \to q\bar{q}$ $\mathrm{Rej}_{50\%}$
ParticleNet [56]	0.844	0.9849	7634	2475	104	954	3339	10526	11173	347	283
ParT [57]	0.861	0.9877	10638	4149	123	1864	5479	32787	15873	543	402
MIParT [58]	0.861	0.9878	10753	4202	123	1927	5450	31250	16807	542	402
L-GATr	0.865	0.9884	12195	4819	128	2304	5764	37736	19231	580	427

Experiment: Jet Tagging

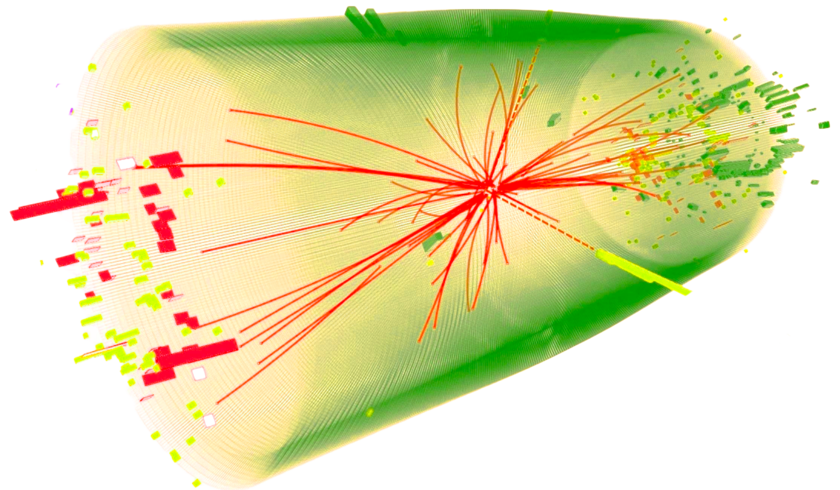


Impact of **pre-training**

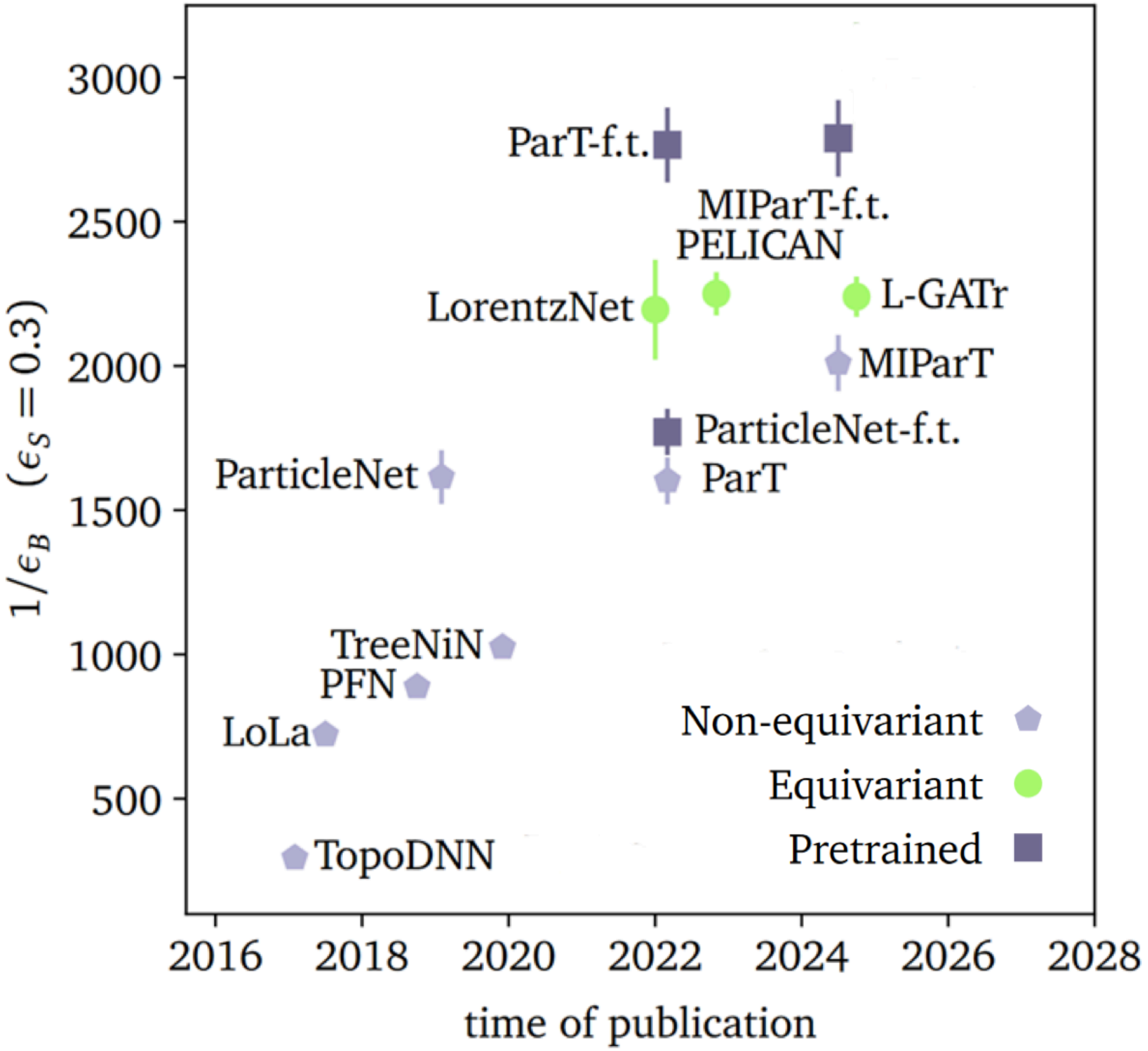


H.Qu et al., 2202.03772

Experiment: Jet Tagging

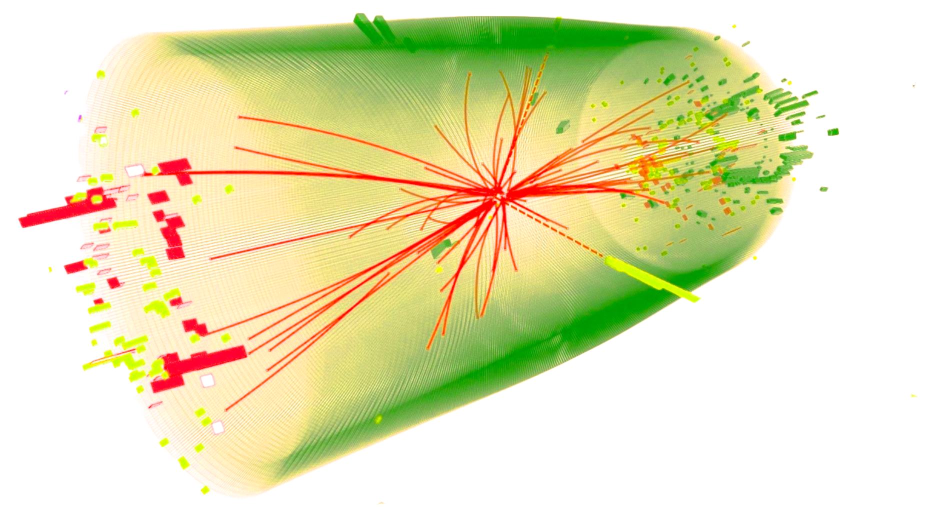


Impact of **pre-training**

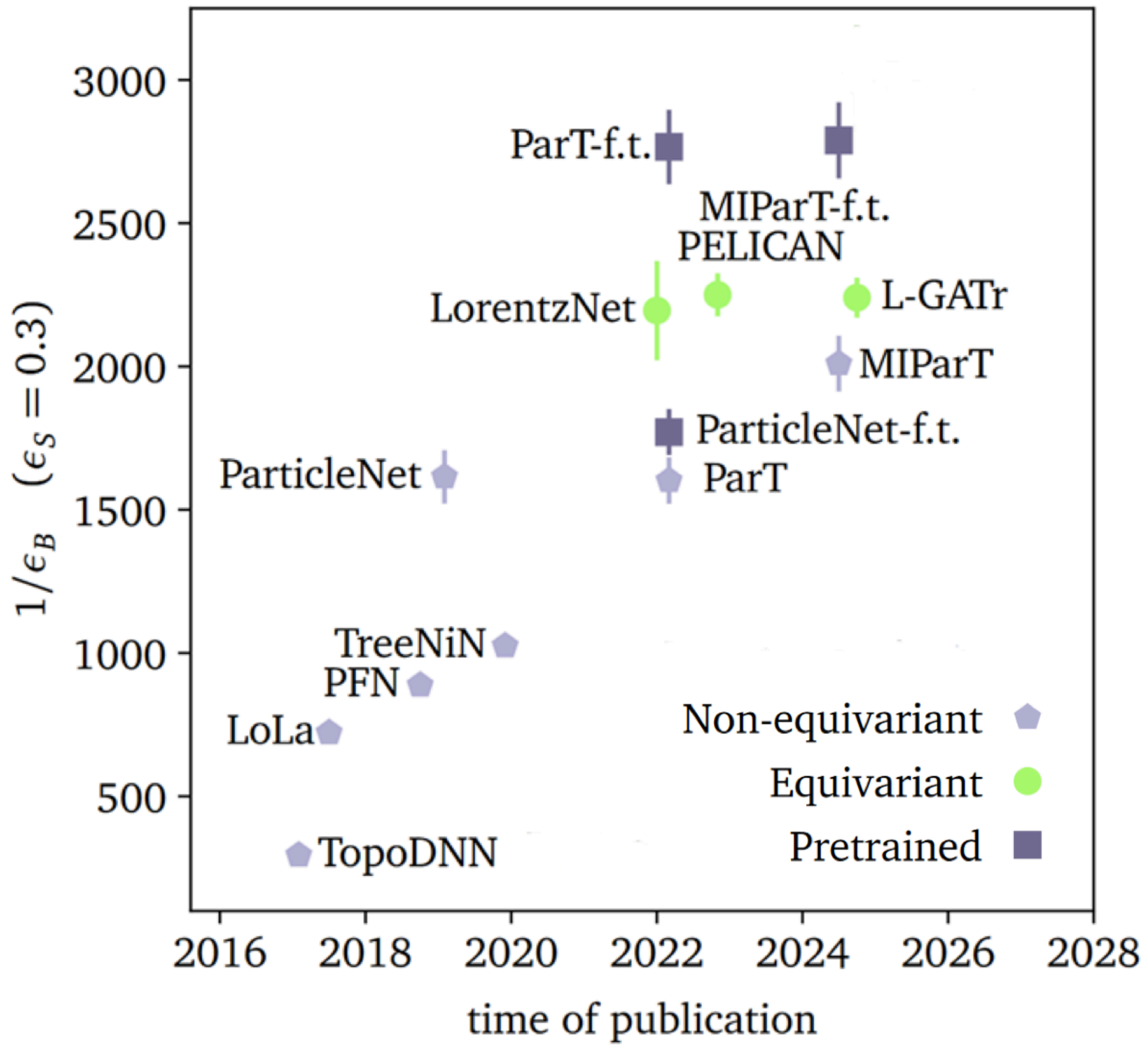


H.Qu et al., 2202.03772

Experiment: Jet Tagging



Impact of **pre-training**



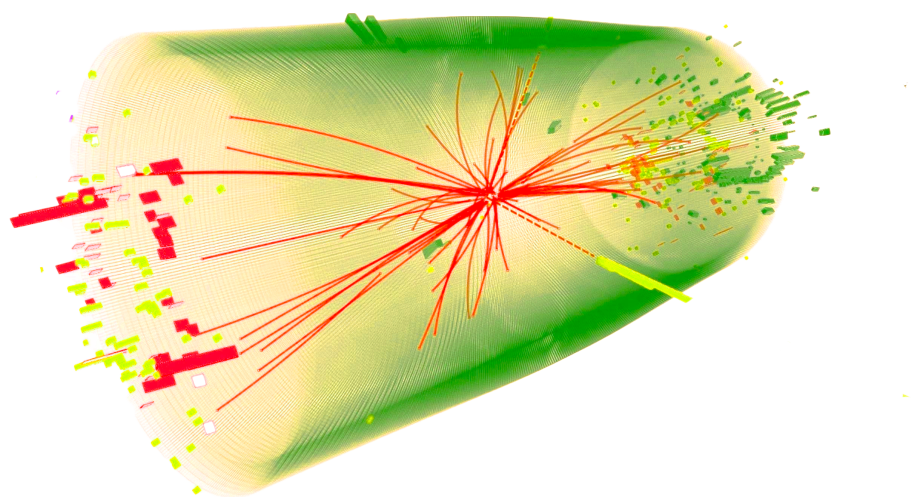
Strategy

----

1. Pre-train on JetClass
2. Restart the output layer
3. Fine-tune on top tagging

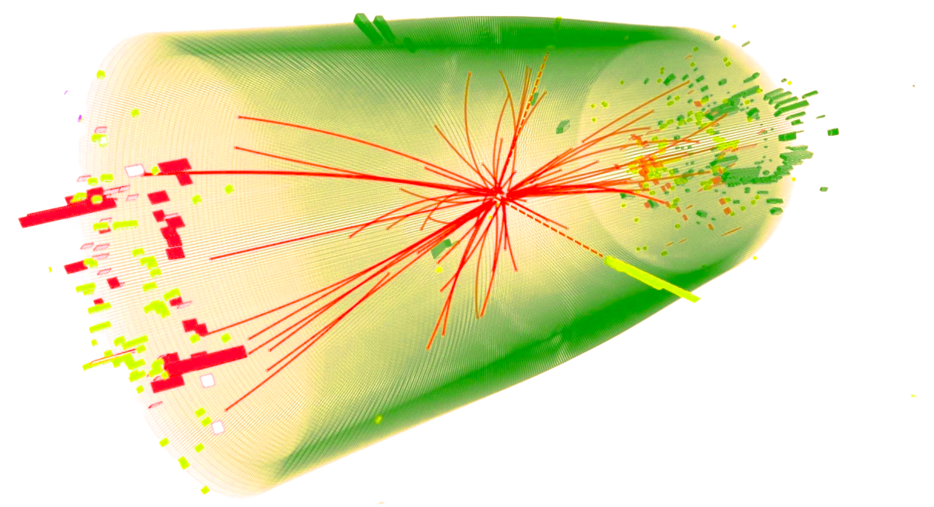
H.Qu et al., 2202.03772

Experiment: Jet Tagging



What if we combine pre-training and equivariance?

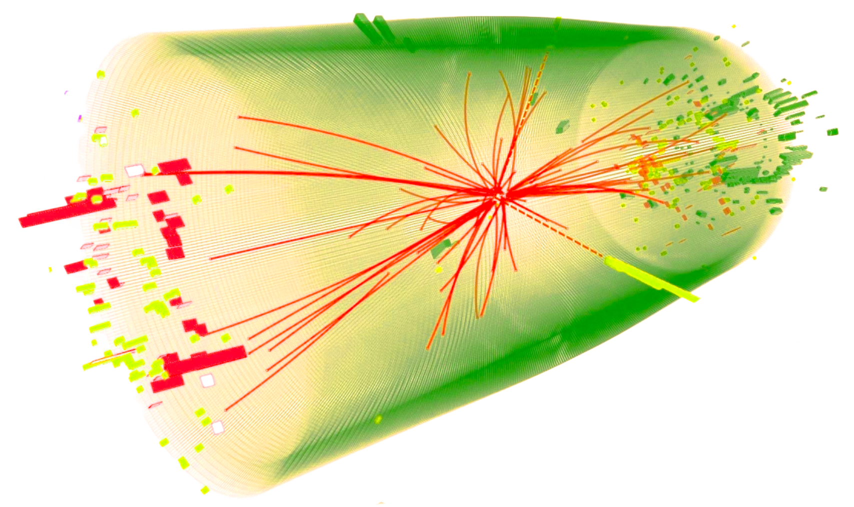
Experiment: Jet Tagging



What if we combine pre-training and equivariance?

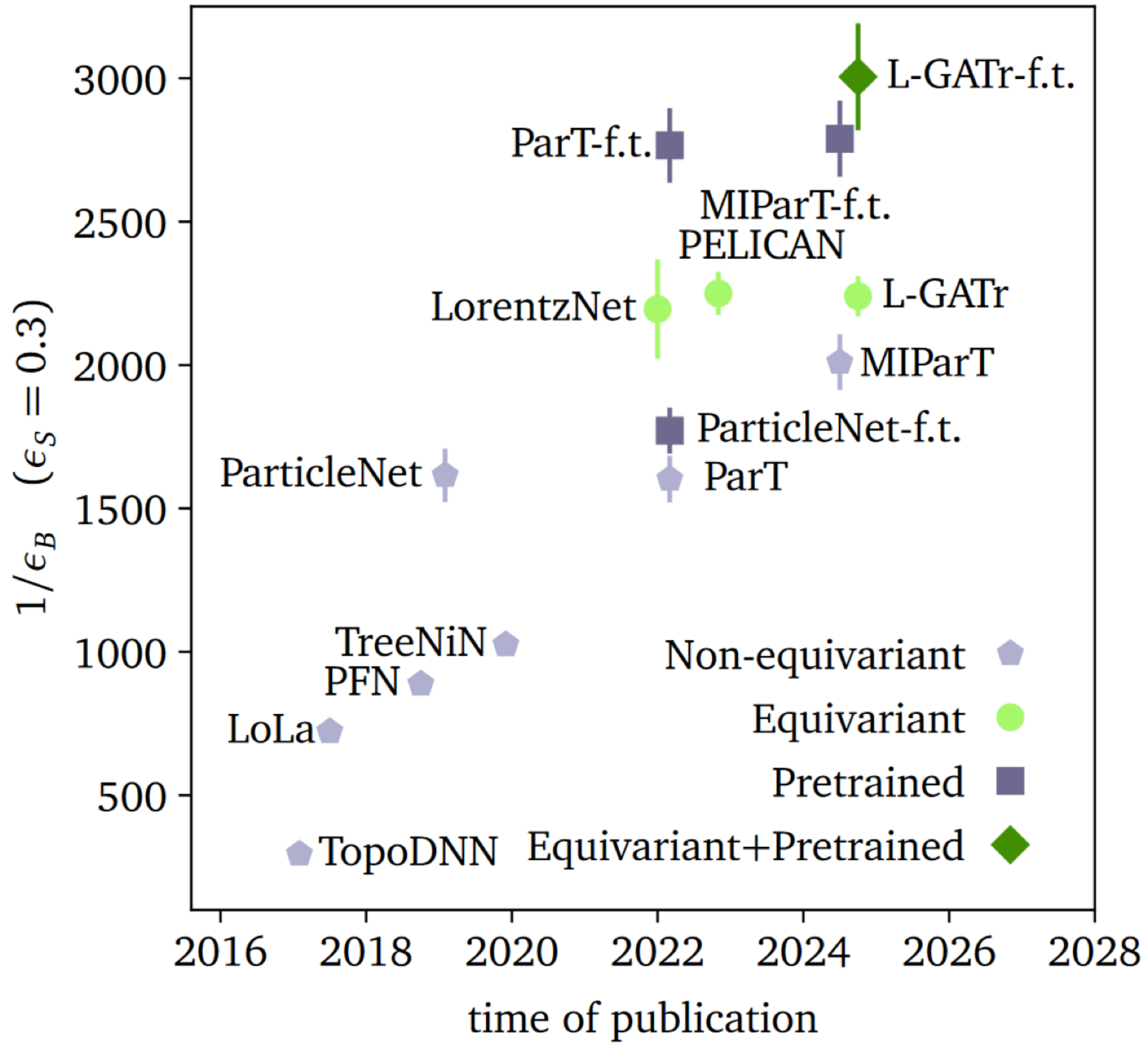
Network	Accuracy	AUC	$1/\epsilon_B$ ($\epsilon_S = 0.5$)	$1/\epsilon_B$ ($\epsilon_S = 0.3$)
LorentzNet	0.942	0.9868	498 ± 18	2195 ± 173
CGENN	0.942	0.9869	500	2172
PELICAN	0.9426 ± 0.0002	0.9870 ± 0.0001	–	2250 ± 75
L-GATr	0.9423 ± 0.0002	0.9870 ± 0.0001	540 ± 20	2240 ± 70
ParticleNet-f.t.	0.942	0.9866	487 ± 9	1771 ± 80
ParT-f.t.	0.944	0.9877	691 ± 15	2766 ± 130
MIParT-f.t.	0.944	0.9878	640 ± 10	2789 ± 133
L-GATr-f.t. (new)	0.9442 ± 0.0002	0.98792 ± 0.00004	661 ± 24	3005 ± 186

Experiment: Jet Tagging

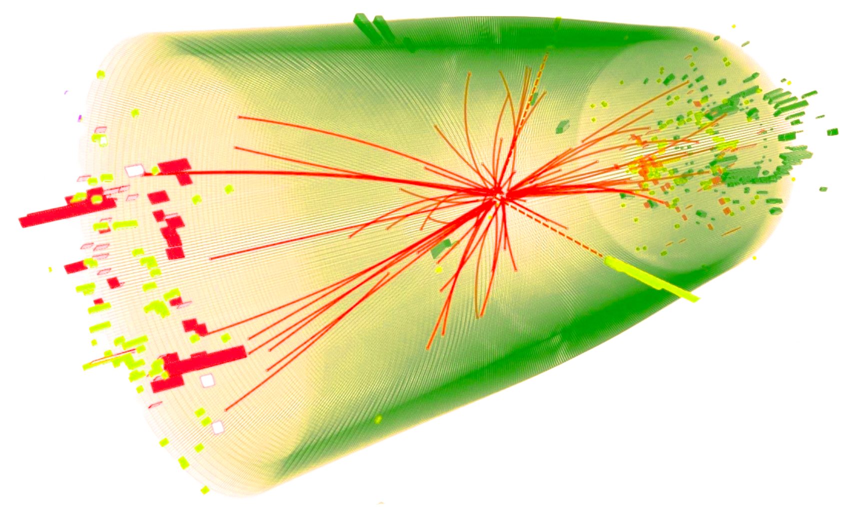


What if we combine pre-training and equivariance?

Network	Accuracy	AUC	$1/\epsilon_B$ ($\epsilon_S = 0.5$)	$1/\epsilon_B$ ($\epsilon_S = 0.3$)
LorentzNet	0.942	0.9868	498 ± 18	2195 ± 173
CGENN	0.942	0.9869	500	2172
PELICAN	0.9426 ± 0.0002	0.9870 ± 0.0001	–	2250 ± 75
L-GATr	0.9423 ± 0.0002	0.9870 ± 0.0001	540 ± 20	2240 ± 70
ParticleNet-f.t.	0.942	0.9866	487 ± 9	1771 ± 80
ParT-f.t.	0.944	0.9877	691 ± 15	2766 ± 130
MIParT-f.t.	0.944	0.9878	640 ± 10	2789 ± 133
L-GATr-f.t. (new)	0.9442 ± 0.0002	0.98792 ± 0.00004	661 ± 24	3005 ± 186

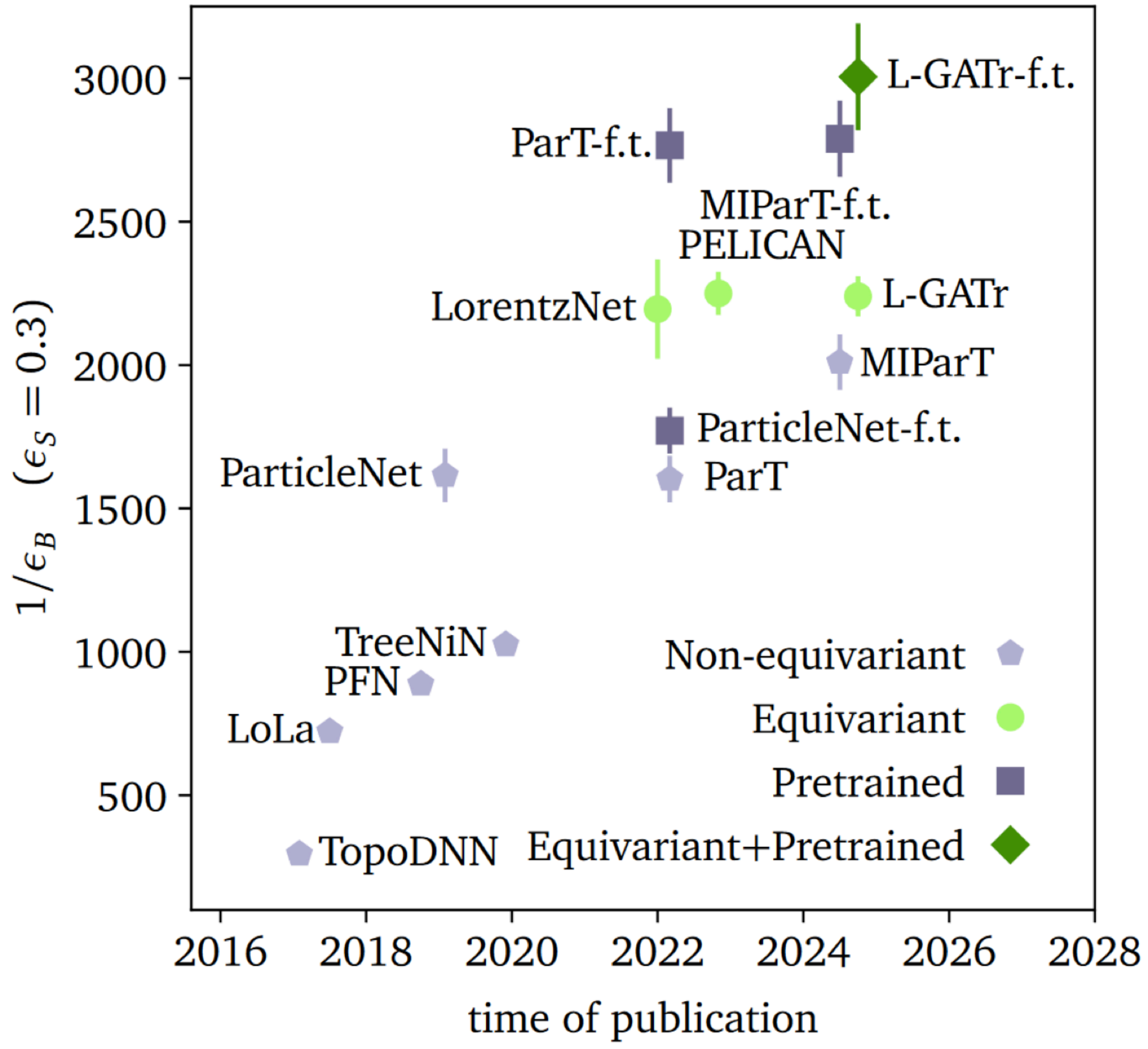


Experiment: Jet Tagging



What if we combine pre-training and equivariance?

Network	Accuracy	AUC	$1/\epsilon_B$ ($\epsilon_S = 0.5$)	$1/\epsilon_B$ ($\epsilon_S = 0.3$)
LorentzNet	0.942	0.9868	498 ± 18	2195 ± 173
CGENN	0.942	0.9869	500	2172
PELICAN	0.9426 ± 0.0002	0.9870 ± 0.0001	–	2250 ± 75
L-GATr	0.9423 ± 0.0002	0.9870 ± 0.0001	540 ± 20	2240 ± 70
ParticleNet-f.t.	0.942	0.9866	487 ± 9	1771 ± 80
ParT-f.t.	0.944	0.9877	691 ± 15	2766 ± 130
MIParT-f.t.	0.944	0.9878	640 ± 10	2789 ± 133
L-GATr-f.t. (new)	0.9442 ± 0.0002	0.98792 ± 0.00004	661 ± 24	3005 ± 186

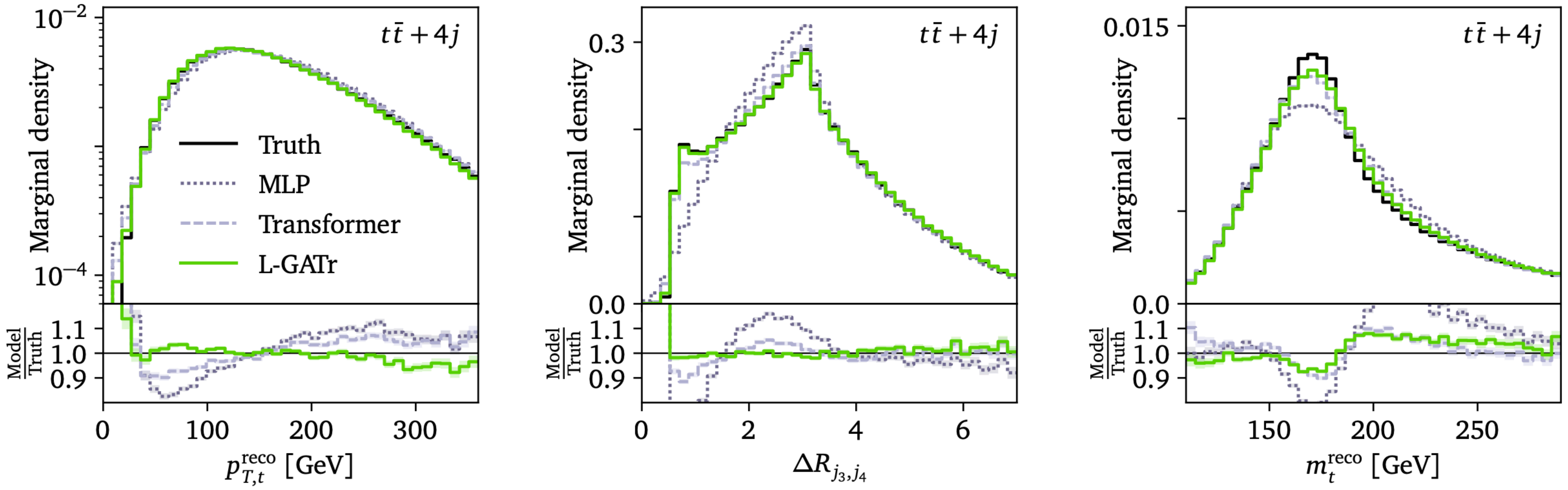


<u>L-GATr is useful for more tasks!!!</u>

Further L-GATr Applications

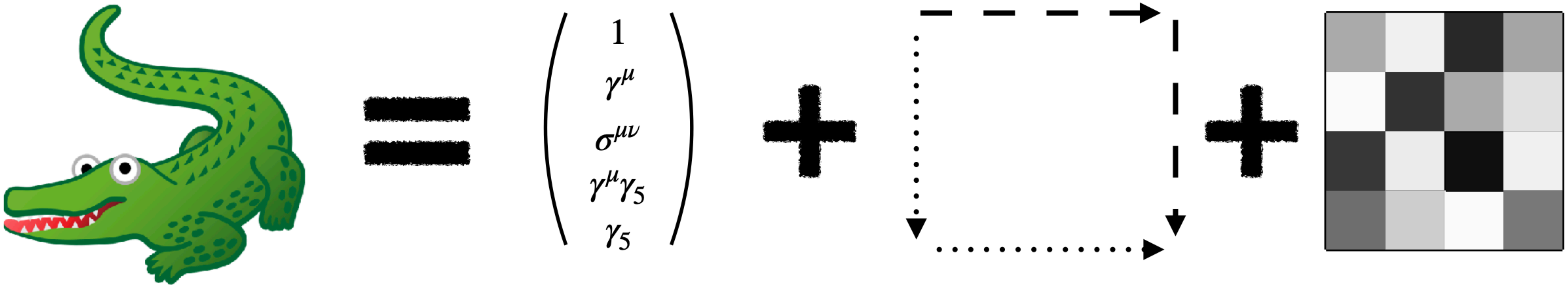
More information on the backup slides!

- Amplitude regression
- Anomaly detection
- Unfolding
- Simulation-Based Inference
- Reconstructed event generation → **First ever Lorentz-equivariant generative model**



Conclusions

- L-GATr is a versatile architecture for LHC physics
- Equivariance boosts performance in multilple tasks
- L-GAtr has a better scaling than competing baselines

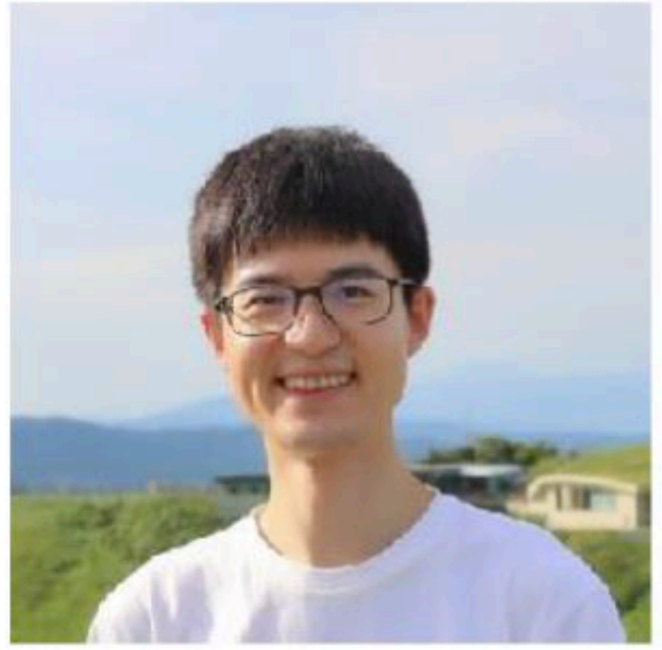


Lorentz-Equivariant Geometric Algebra Transformer = Geometric algebra representations + Lorentz-Equivariant layers + Transformer architecture

Jonas Spinner

Pim de Haan

Tilman Plehn



Huilin Qu

Jesse Thaler

Johann Brehmer

Lorentz-Equivariant Geometric Algebra Transformer for High-Energy Physics

Jonas Spinner*, Victor Breso*, Pim de Haan, Tilman Plehn, Jesse Thaler, Johann Brehmer, NeurIPS 2024, arXiv:2405.14806

CS paper

A Lorentz-Equivariant Transformer for all of the LHC

Johann Brehmer, Víctor Bresó, Pim de Haan, Tilman Plehn, Huilin Qu, Jonas Spinner, Jesse Thaler, arXiv:2411.00446

HEP paper

Try L-GATr for yourself! →

L-GATr code

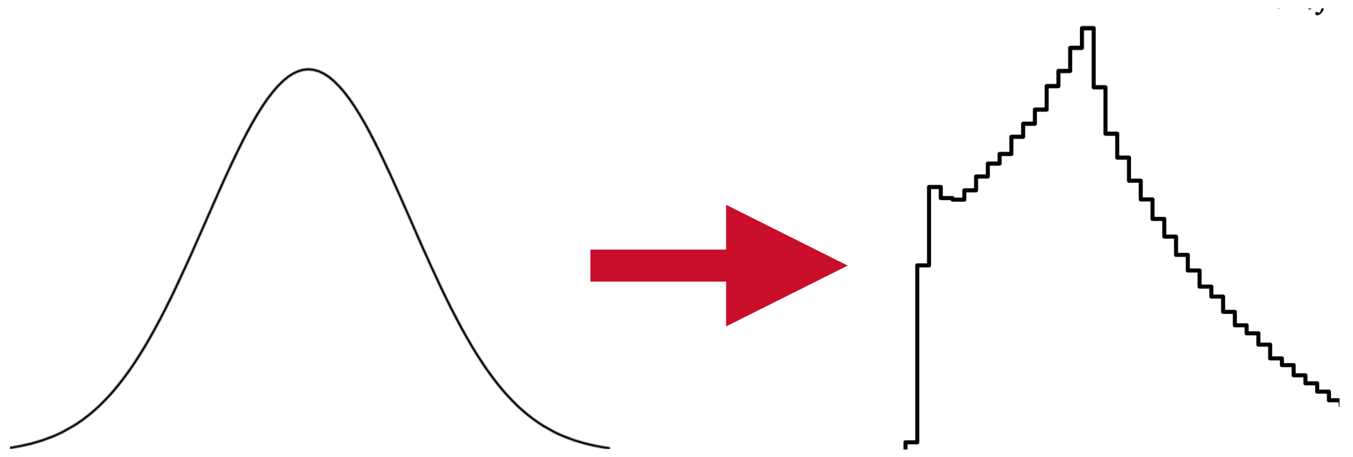
Backup Slides

Experiment: Reconstructed event generation

Lorentz-Equivariant Flow Matching

Continuous normalizing flow (CNF) connects a simple base density to a complex target density through a neural differential equation

$$\frac{d}{dt}\phi(x) = v_t(x)$$



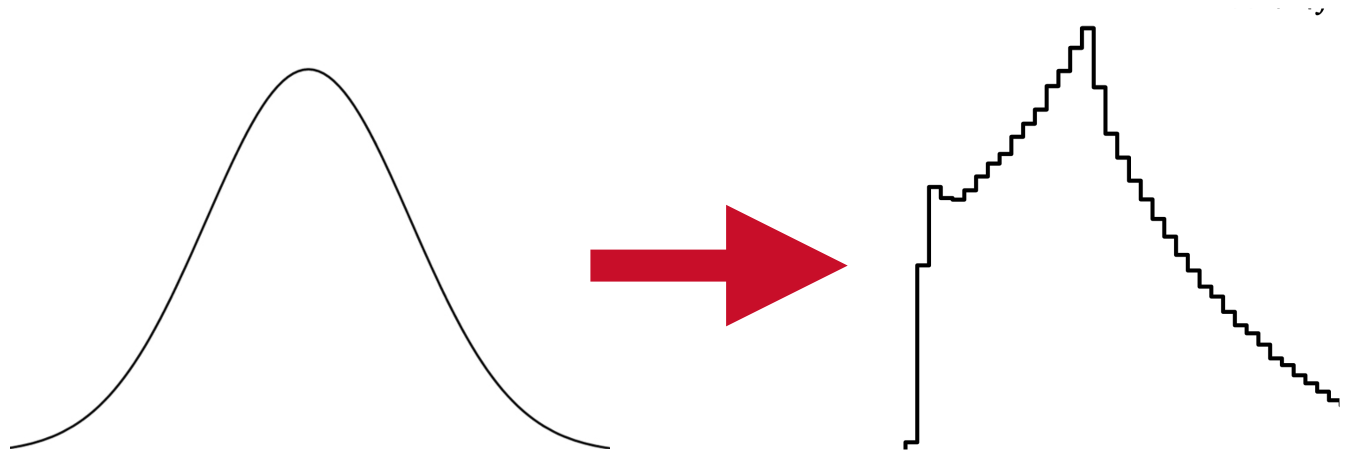
R. Chen et al., 1806.07366
Y. Lipman et al.,2210.02747

Experiment: Reconstructed event generation

Lorentz-Equivariant Flow Matching

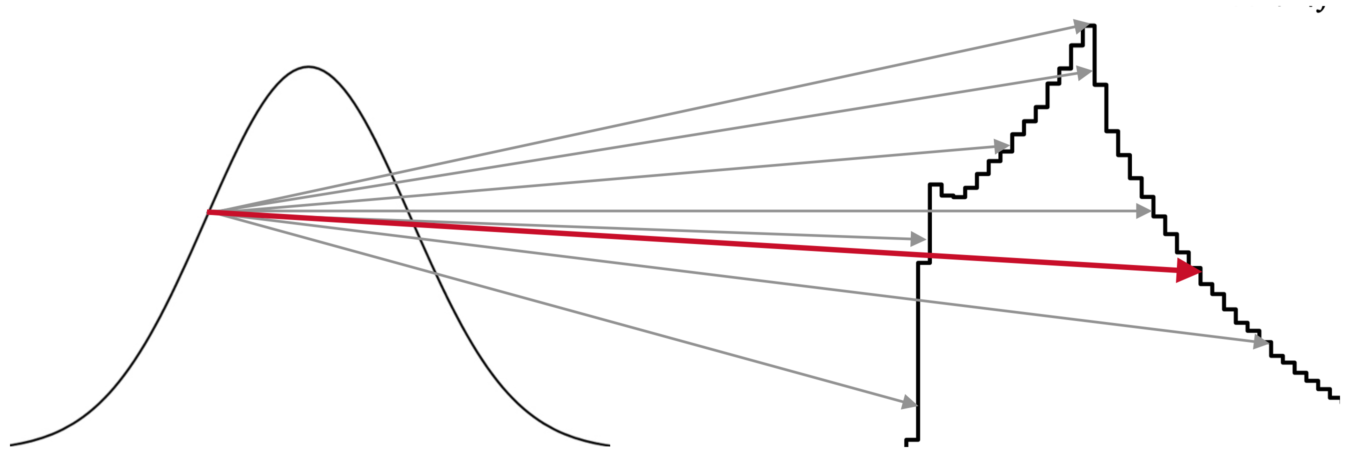
Continuous normalizing flow (CNF) connects a simple base density to a complex target density through a neural differential equation

$$\frac{d}{dt}\phi(x) = v_t(x)$$



Conditional flow matching (CFM) is a simple way to train CNFs by comparing the learned velocity $v_t(x)$ to a conditional target velocity $u_t(x \mid x_1)$

$$\mathscr{L} = \langle (v_t(x) - u_t(x \mid x_1))^2 \rangle$$



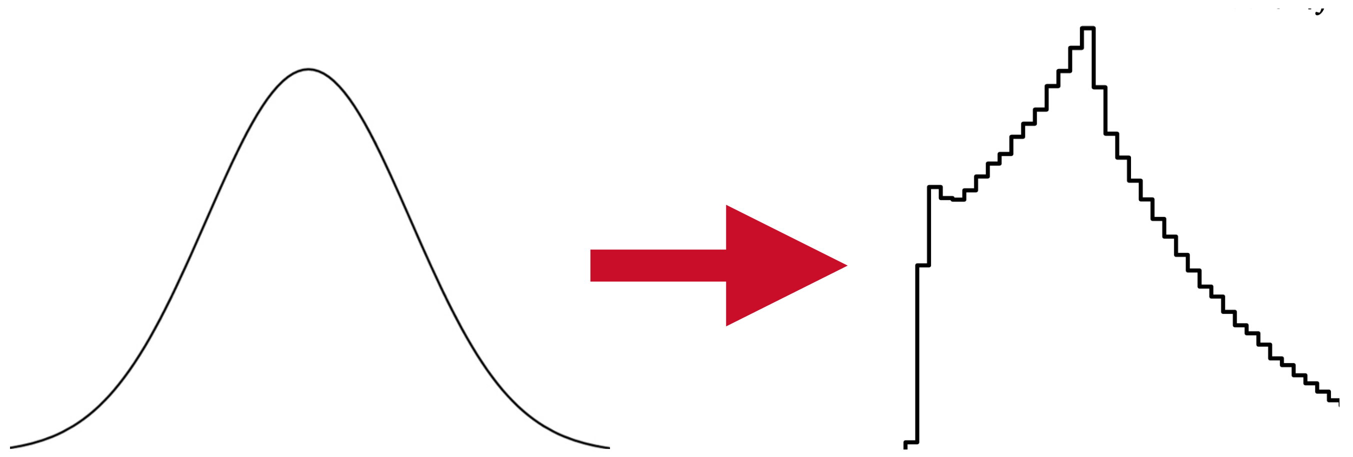
R. Chen et al., 1806.07366
Y. Lipman et al.,2210.02747

Experiment: Reconstructed event generation

Lorentz-Equivariant Flow Matching

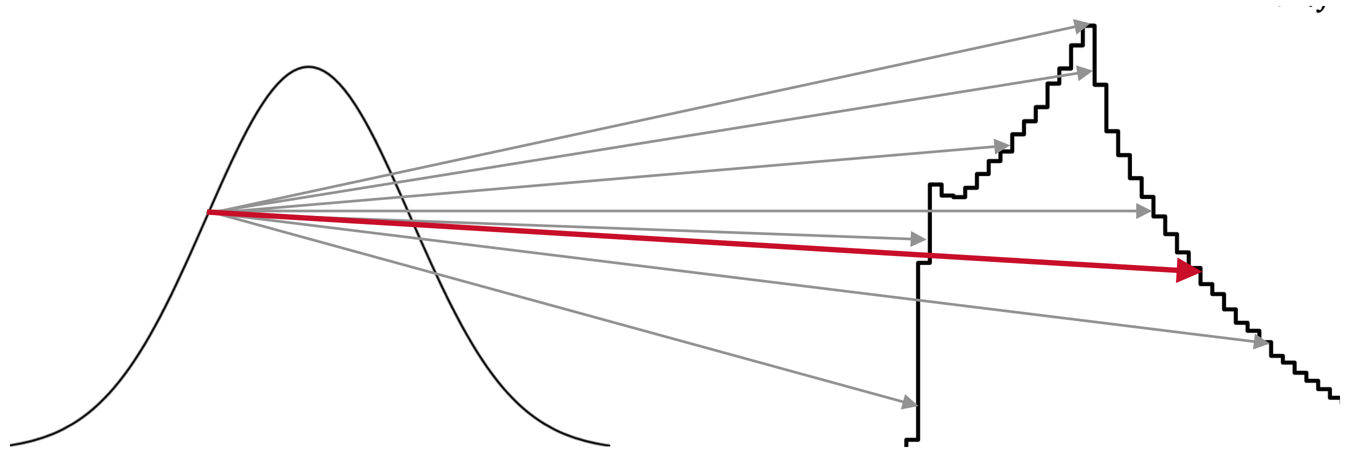
Continuous normalizing flow (CNF)
connects a simple base density to a complex target density through a neural differential equation

$$\frac{d}{dt}\phi(x) = v_t(x)$$



Conditional flow matching (CFM)
is a simple way to train CNFs by comparing the learned velocity $v_t(x)$ to a conditional target velocity $u_t(x\,|\,x_1)$

$$\mathscr{L} = \langle (v_t(x) - u_t(x\,|\,x_1))^2 \rangle$$



How to pick the target velocity $u_t(x\,|\,x_1)$?

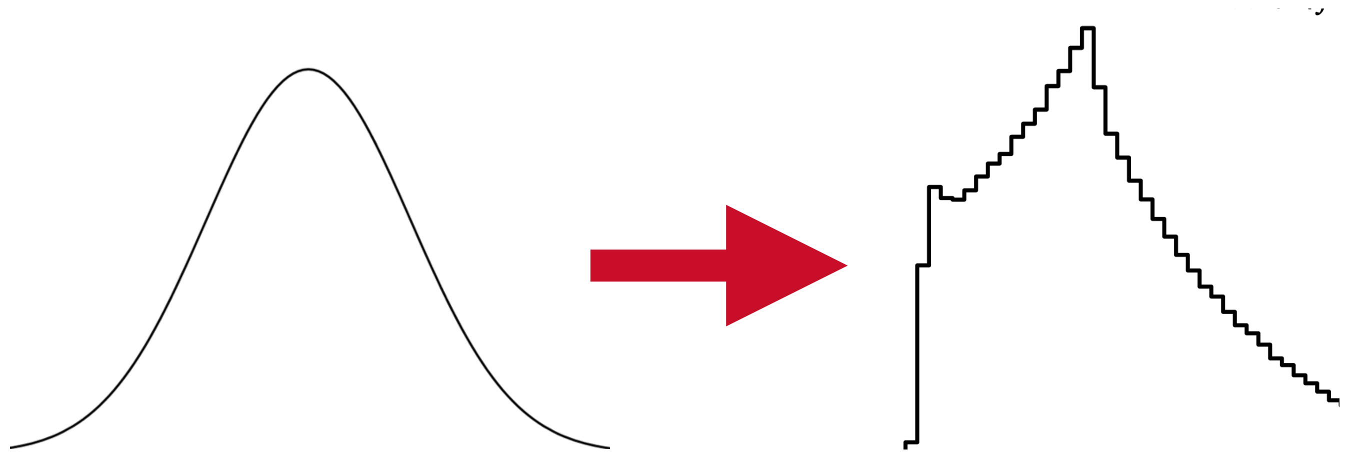
R. Chen et al., 1806.07366
Y. Lipman et al.,2210.02747

Experiment: Reconstructed event generation

Lorentz-Equivariant Flow Matching

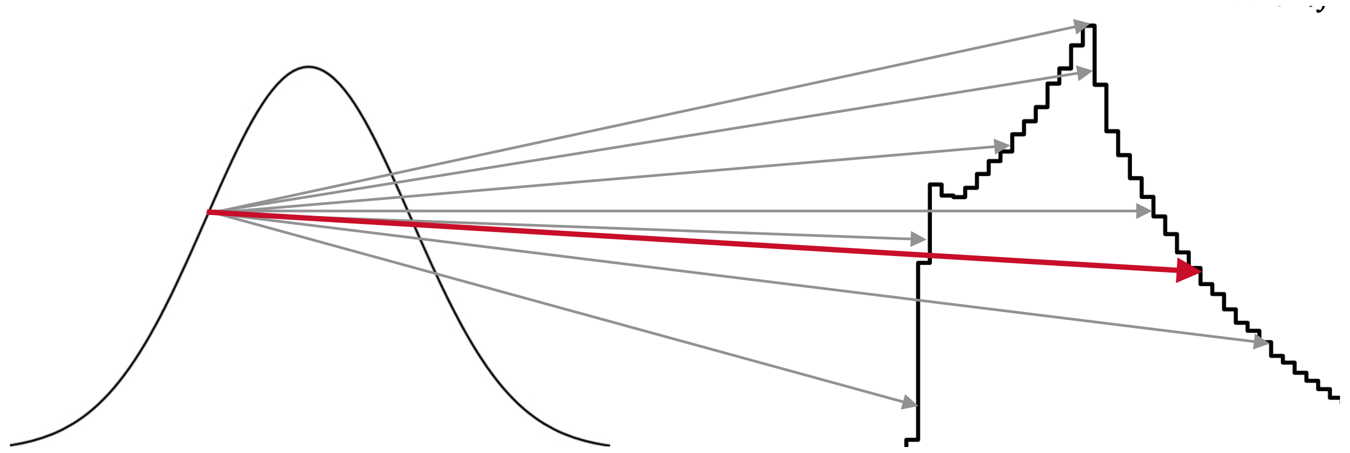
Continuous normalizing flow (CNF) connects a simple base density to a complex target density through a neural differential equation

$$\frac{d}{dt}\phi(x) = v_t(x)$$



Conditional flow matching (CFM) is a simple way to train CNFs by comparing the learned velocity $v_t(x)$ to a conditional target velocity $u_t(x \mid x_1)$

$$\mathscr{L} = \langle (v_t(x) - u_t(x \mid x_1))^2 \rangle$$



How to pick the target velocity $u_t(x \mid x_1)$?

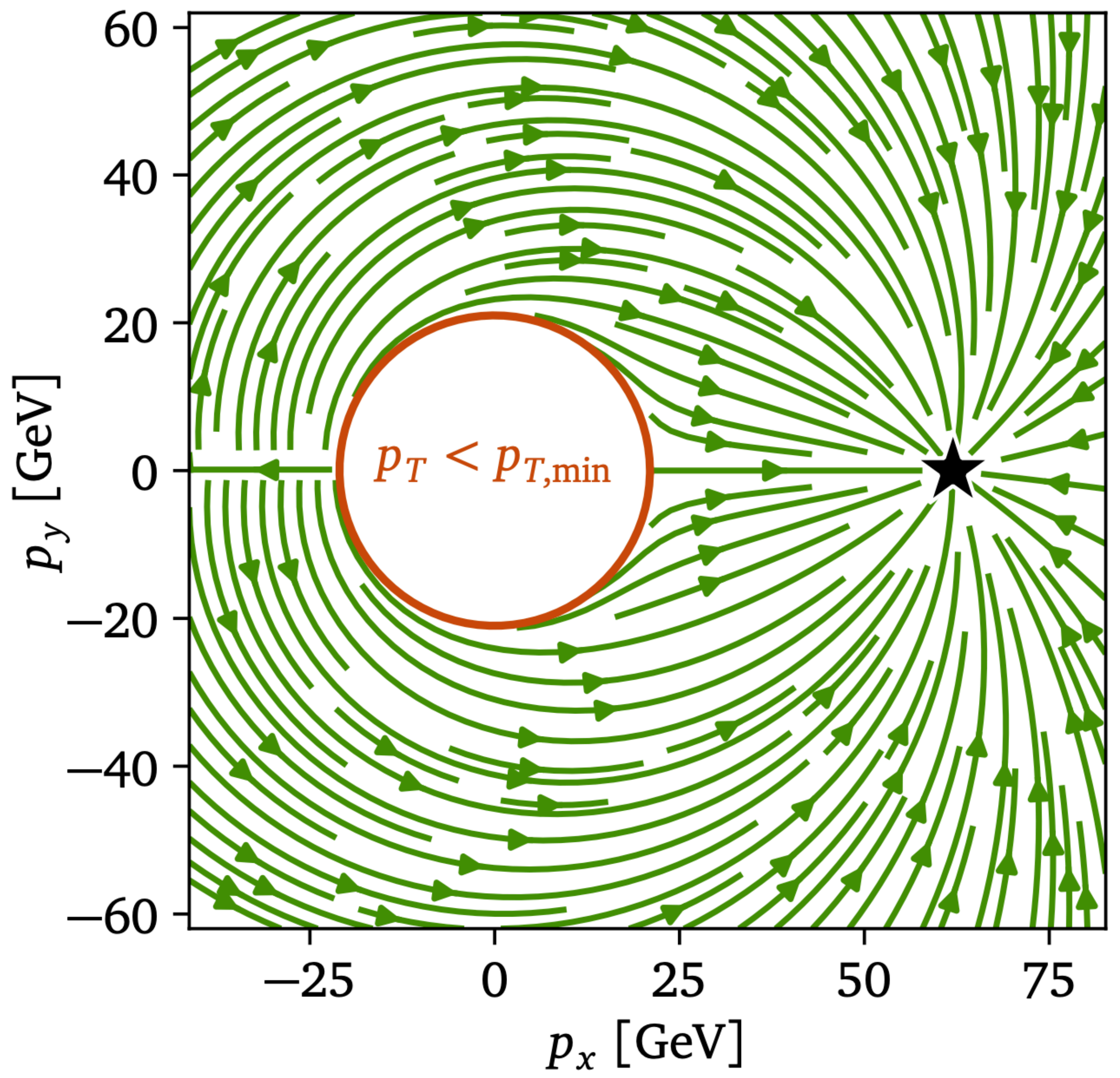
R. Chen et al., 1806.07366
Y. Lipman et al.,2210.02747

Experiment: Reconstructed event generation

Physics-inspired target trajectories

Straight trajectories in 'modified jet momenta' x:

$$p = \begin{pmatrix} E \\ p_x \\ p_y \\ p_z \end{pmatrix} \quad \rightarrow \quad f^{-1}(p) = x = \begin{pmatrix} x_p \\ x_m \\ x_\eta \\ x_\phi \end{pmatrix} \equiv \begin{pmatrix} \log(p_T - p_{T,\text{min}}) \\ \log m^2 \\ \eta \\ \phi \end{pmatrix}$$

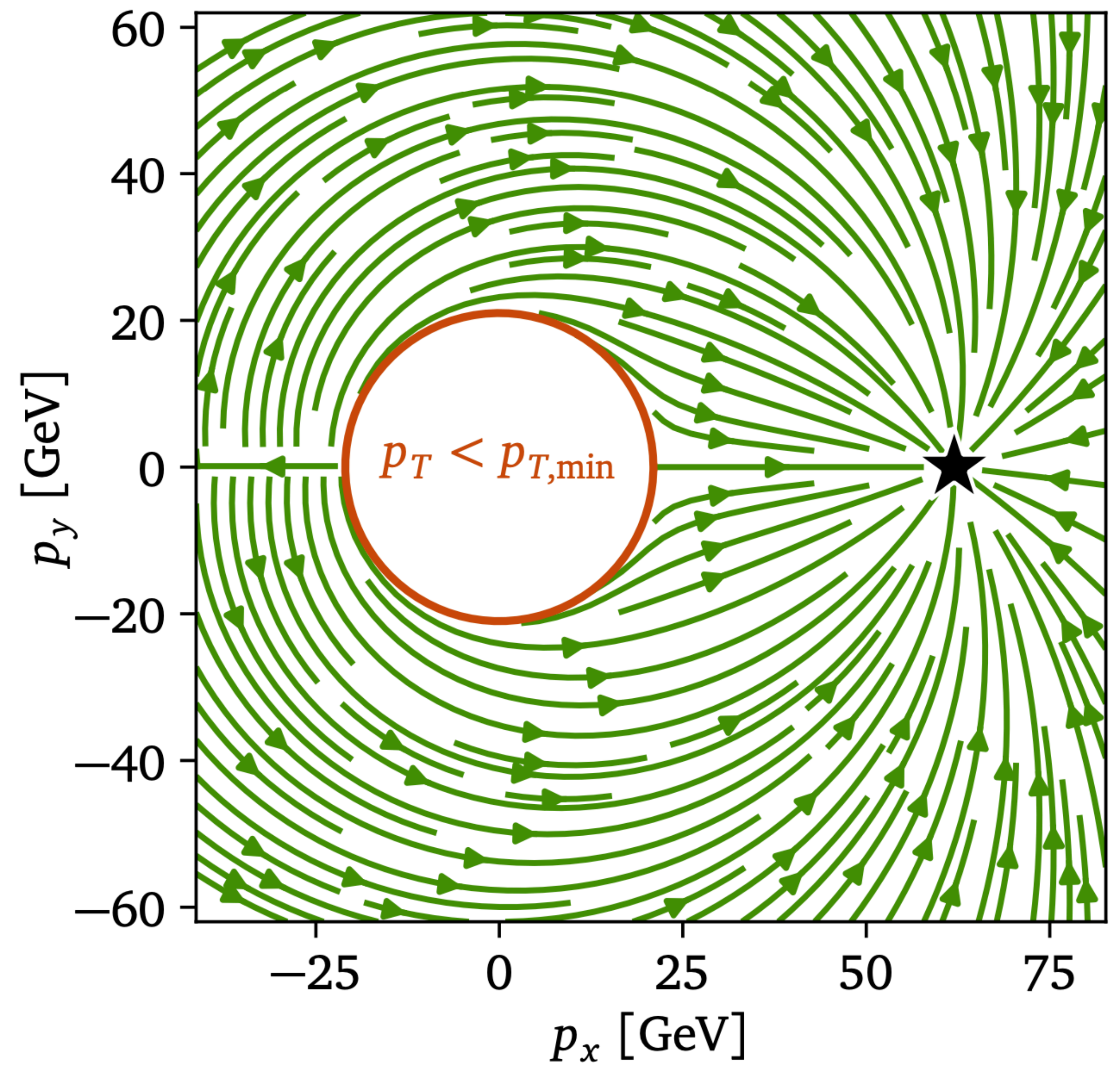


Experiment: Reconstructed event generation

Physics-inspired target trajectories

Straight trajectories in 'modified jet momenta' x:

$$p = \begin{pmatrix} E \\ p_x \\ p_y \\ p_z \end{pmatrix} \quad \rightarrow \quad f^{-1}(p) = x = \begin{pmatrix} x_p \\ x_m \\ x_\eta \\ x_\phi \end{pmatrix} \equiv \begin{pmatrix} \log(p_T - p_{T,\text{min}}) \\ \log m^2 \\ \eta \\ \phi \end{pmatrix}$$



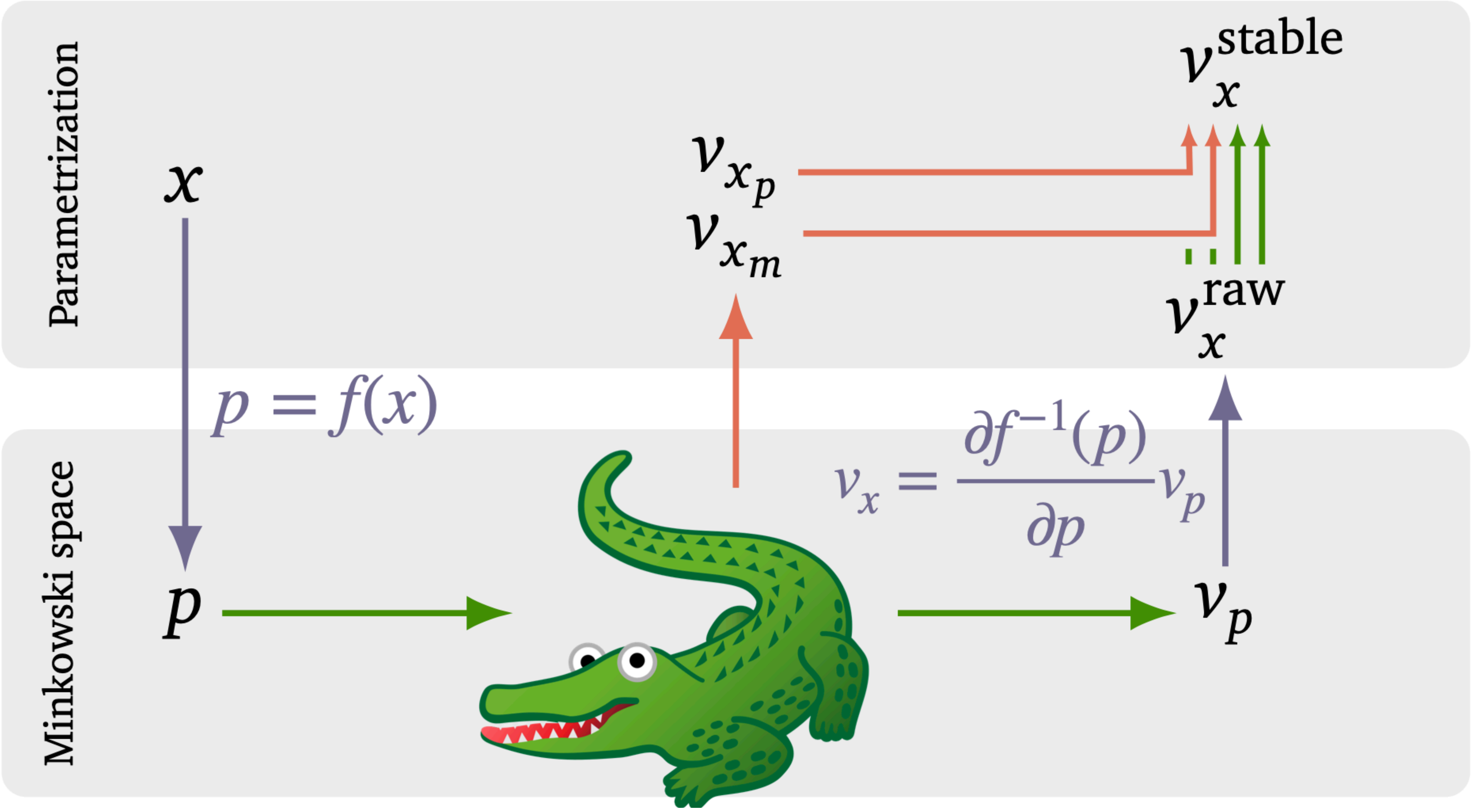
	Data	Architecture	Base distribution	Periodic	Neg. log-likelihood	AUC
	p	L-GATr	rejection sampling	✓	-30.80 ± 0.17	0.945 ± 0.004
	x	MLP	rejection sampling	✓	-32.13 ± 0.05	0.780 ± 0.003
	x	L-GATr	rejection sampling	✗	-32.57 ± 0.05	0.530 ± 0.017
	x	L-GATr	no rejection sampling	✓	-32.58 ± 0.04	0.523 ± 0.014
(default)	x	L-GATr	rejection sampling	✓	-32.65 ± 0.04	0.515 ± 0.009

Experiment: Reconstructed event generation

How to extract the CFM velocity field $v_t(x)$?

Extend standard CFM workflow with L-GATr:

- Transformations $f(x)$ between Minkowski space p and the parametrization x
- Equivariant operations using multivectors
- Symmetry-breaking operations using scalars (required for numerical stability)



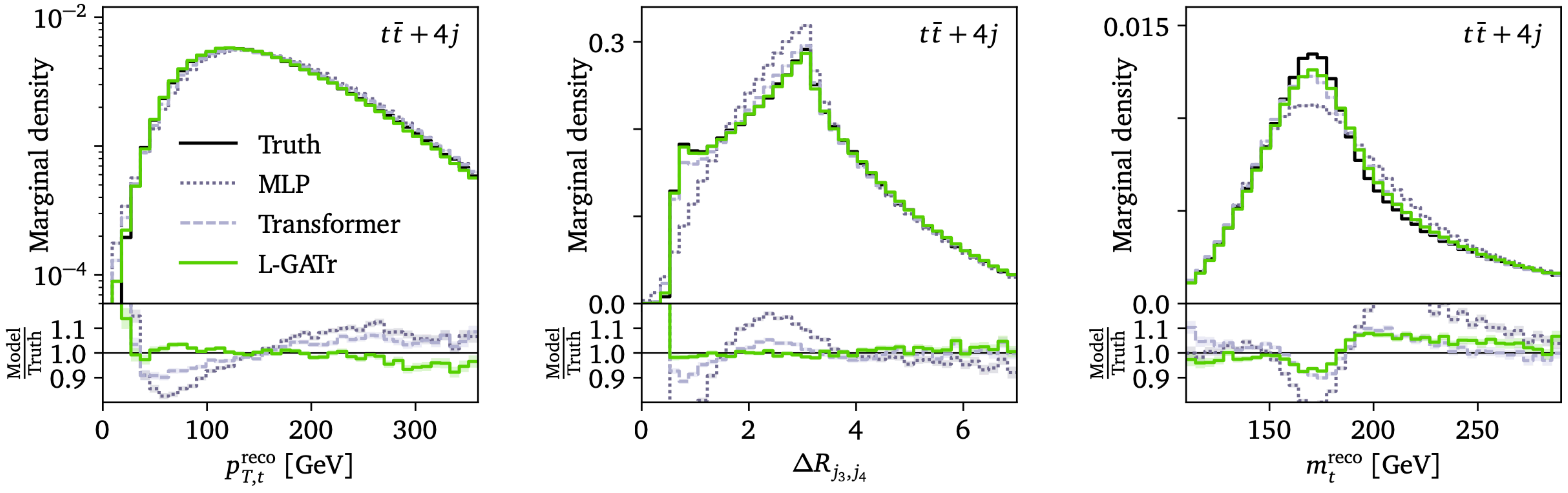
Experiment: Reconstructed event generation

Task: Build a generator that produces reconstructed level distributions

- Dataset: $pp \to t_h \bar{t}_h + nj, \quad n = 0 \ldots 4$

- Simulation chain: MadGraph + Pythia + Delphes + Reconstruction

We develop the first-ever **Lorentz-equivariant generative model** trained with **Riemannian flow matching***

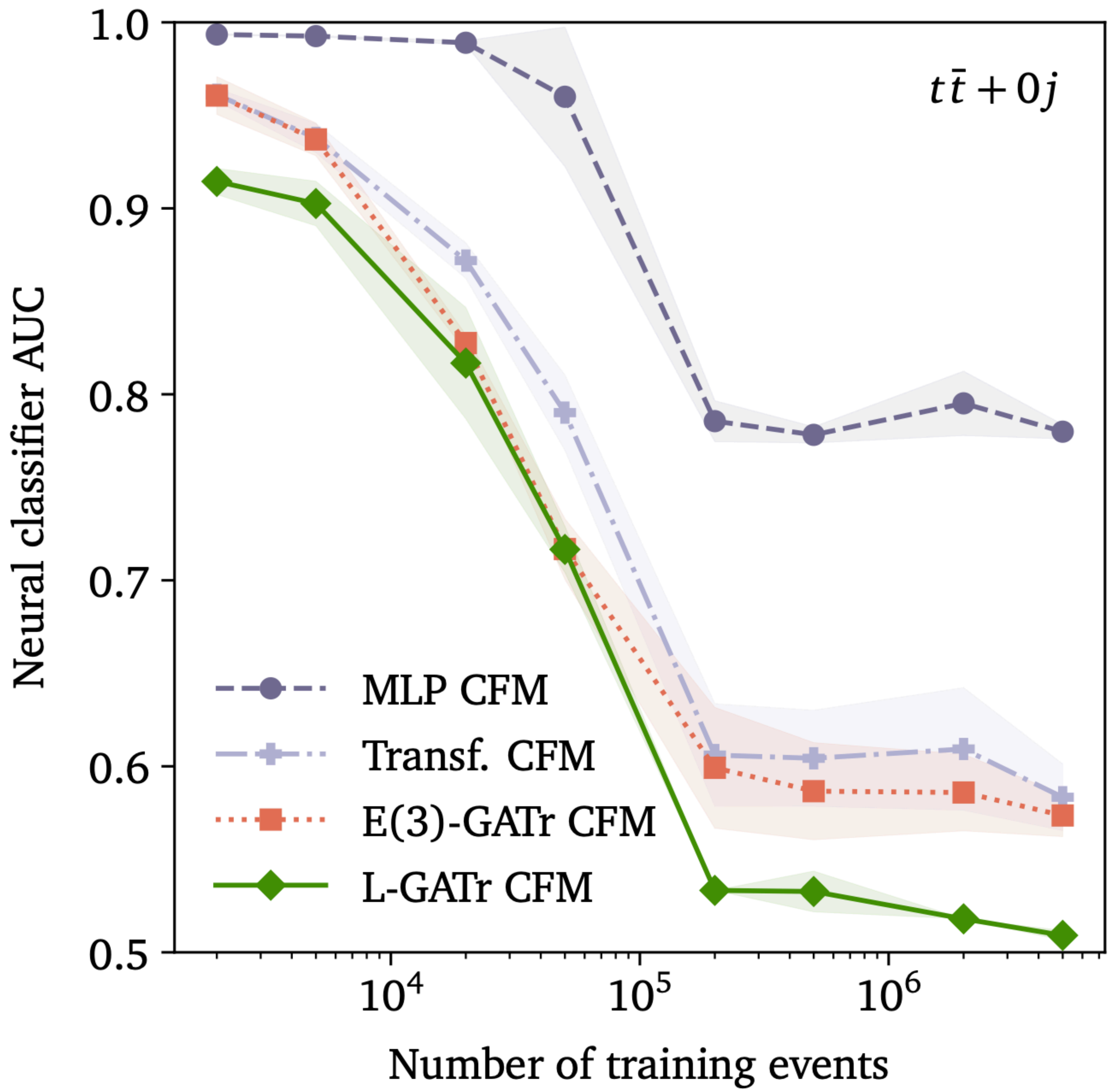


- Equivariance helps with **challenging features**

- L-GATr outperforms all baselines across **multiple process multiplicities**

Experiment: Reconstructed event generation

Result: Classifier metric



- L-GATr generates samples that a classifier can barely distinguish from the ground truth

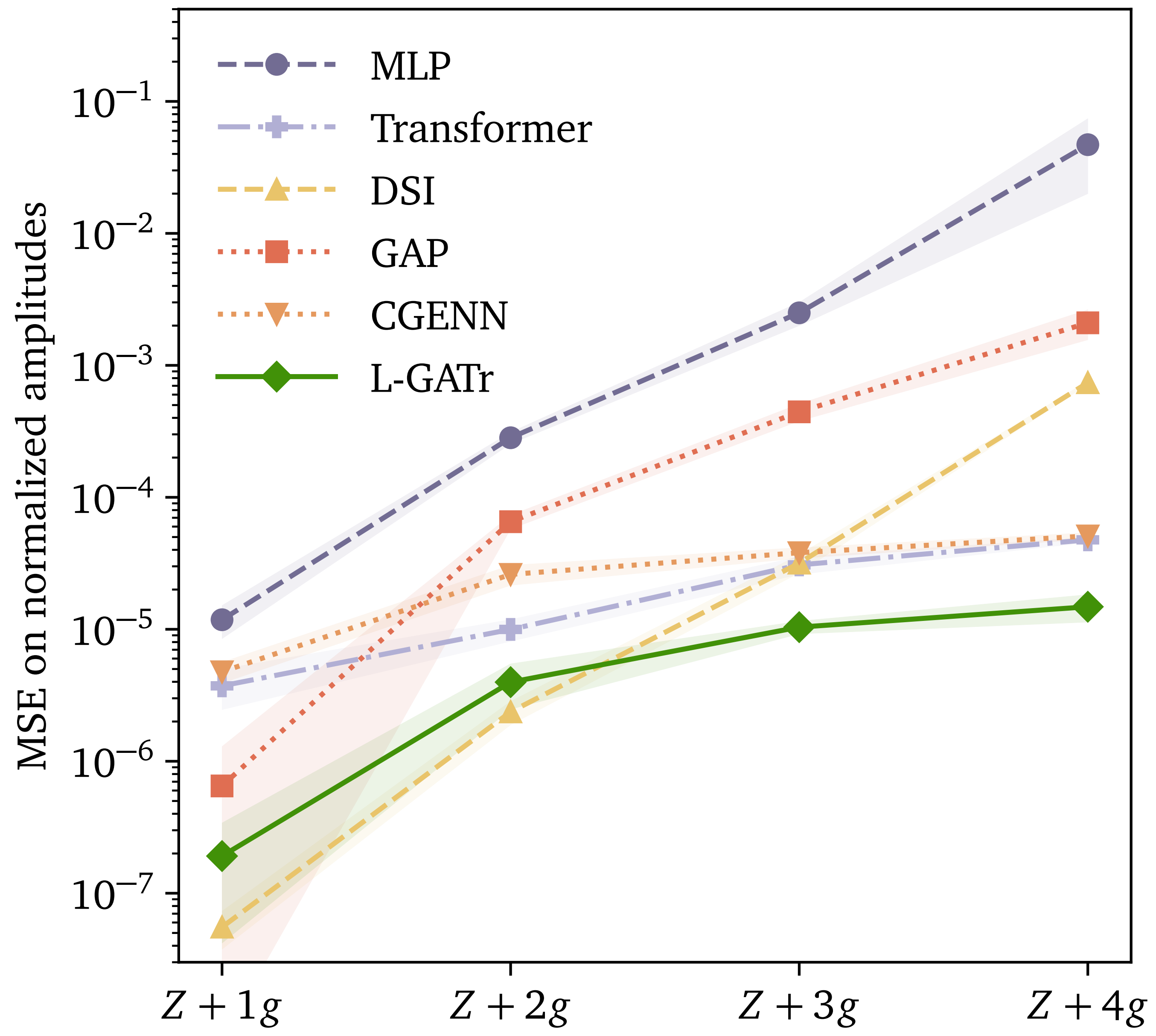
- Equivariant networks with full symmetry-breaking outperform non-equivariant networks

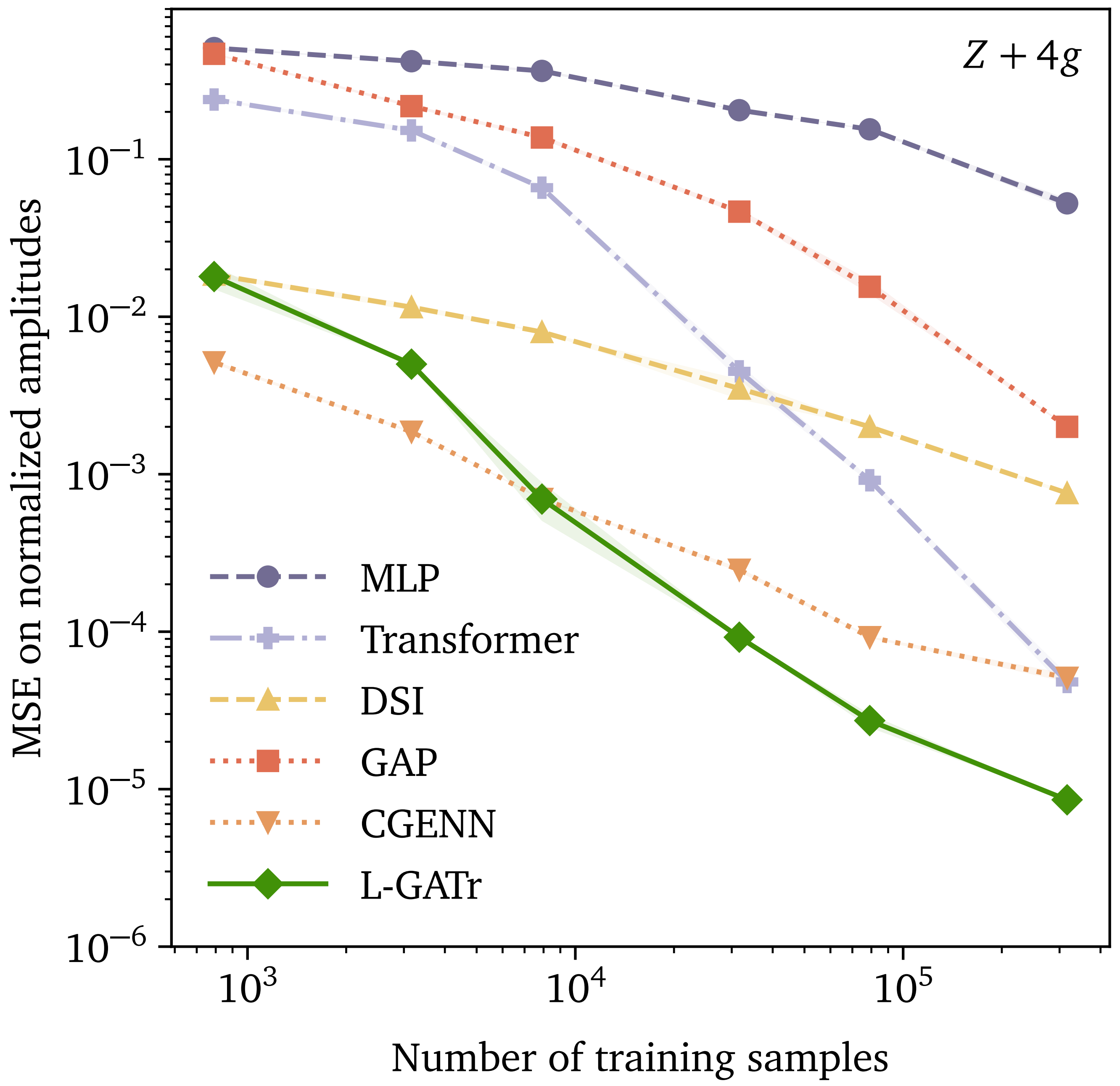
Experiment: QFT amplitude regression

Task: Phase space points $\{p_1, \ldots, p_n\}$ ⟶ Squared amplitude $\mathscr{M}^2$

- Expensive operation at scale for EW processes and NLO calculations

- Neural surrogates are fast, but they don't scale well to high multiplicity





- **Key drivers:** Lorentz and permutation equivariance

- **High data efficiency** (important for interpolation tasks)

VB, G. Heinrich et al., 2412.09534
J. Spinner, L. Favaro et al., 2505.20280