JET TAGGING WITH LORENTZ-EQUIVARIANT GEOMETRIC ALGEBRA TRANSFORMERS

Víctor Bresó Pla

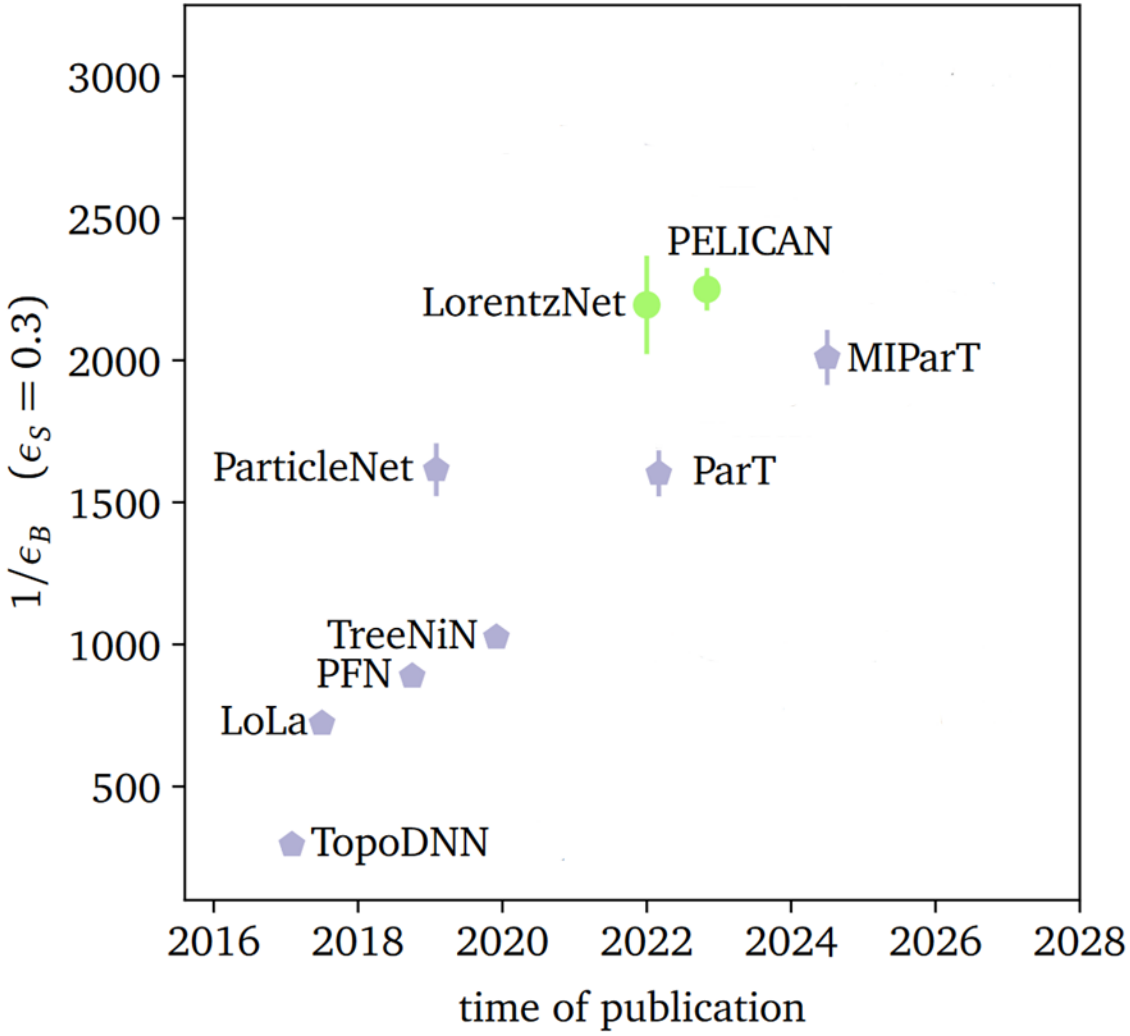
In collaboration with Jonas Spinner, Johann Brehmer, Pim de Haan, Tilman Plehn, Huilin Qu & Jesse Thaler

arXiv:2405.14806 [physics.data-an]

arXiv:2411.00104 [hep-ph, hep-ex]

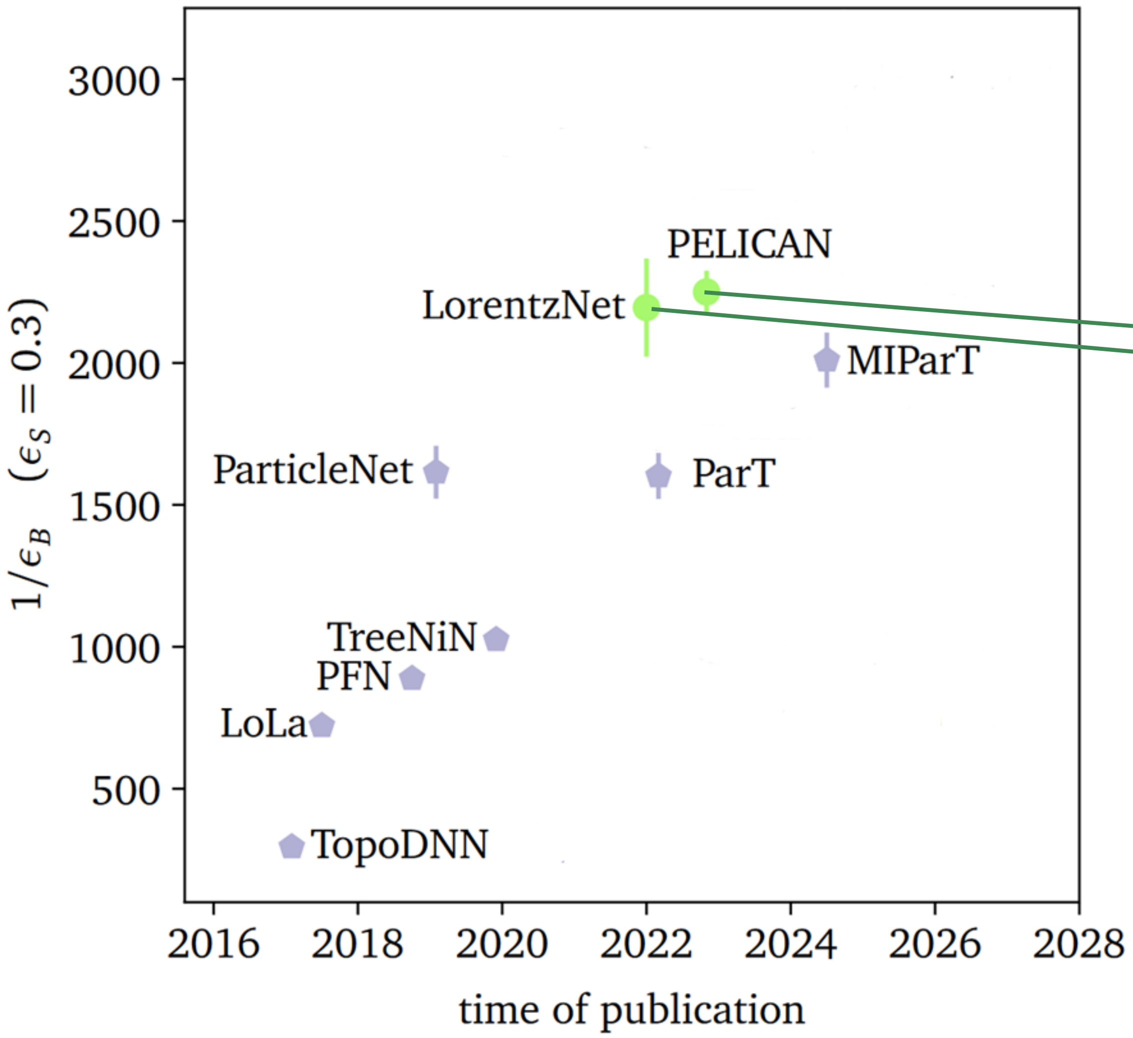
UNIVERSITÄT HEIDELBERG
ZUKUNFT SEIT 1386

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

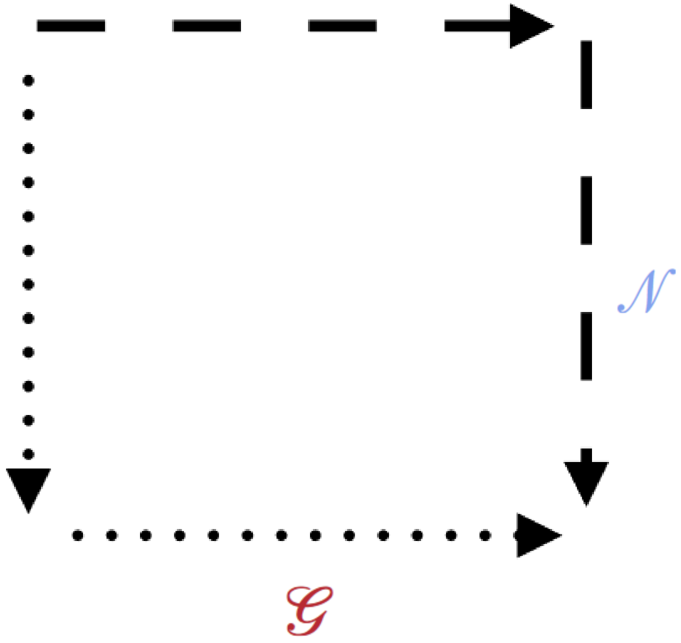
EQUIVARIANT NEURAL NETWORKS

What are equivariant neural networks?

EQUIVARIANT NEURAL NETWORKS

What are equivariant neural networks?

$$\mathscr{G}(\mathscr{N}(x)) = \mathscr{N}(\mathscr{G}(x))$$

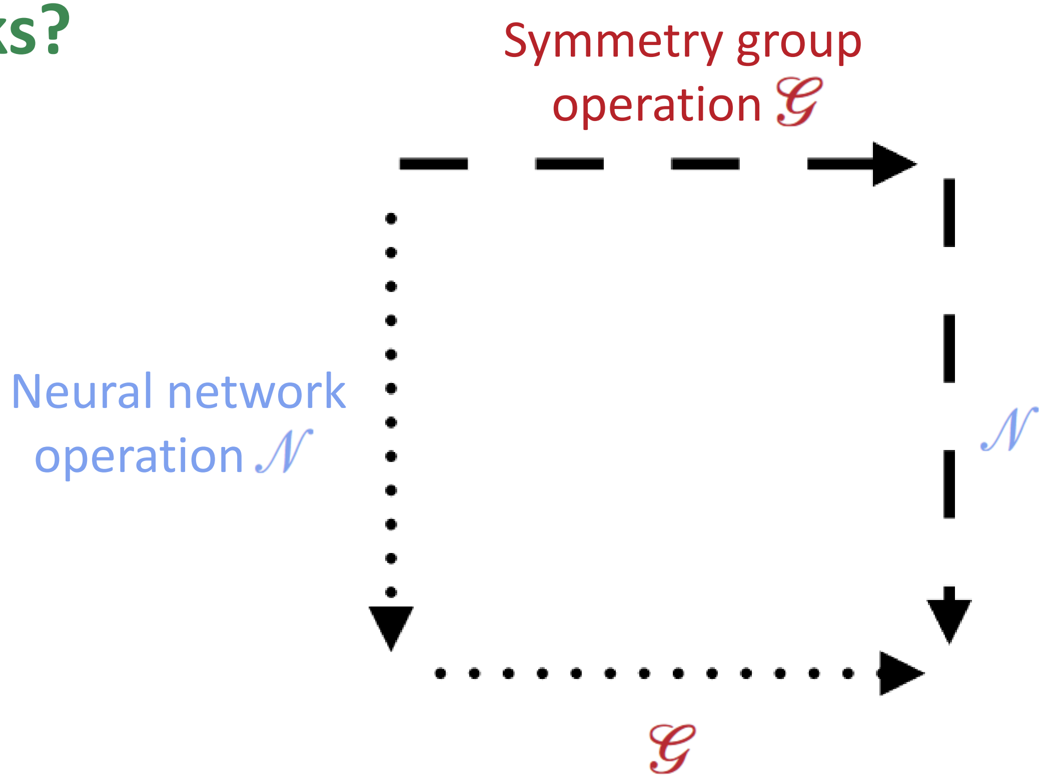


EQUIVARIANT NEURAL NETWORKS

What are equivariant neural networks?

$$\mathscr{G}(\mathscr{N}(x)) = \mathscr{N}(\mathscr{G}(x))$$

Why equivariance?



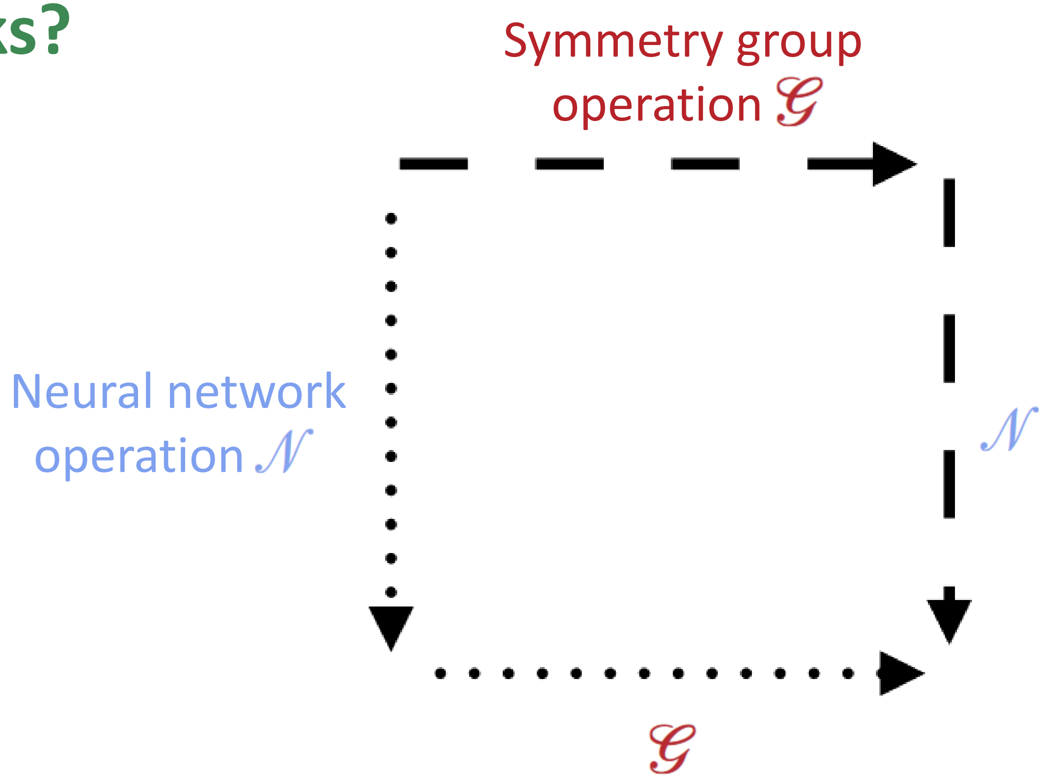
EQUIVARIANT NEURAL NETWORKS

What are equivariant neural networks?

$$\mathscr{G}(\mathscr{N}(x)) = \mathscr{N}(\mathscr{G}(x))$$

Why equivariance?

- Symmetries are important



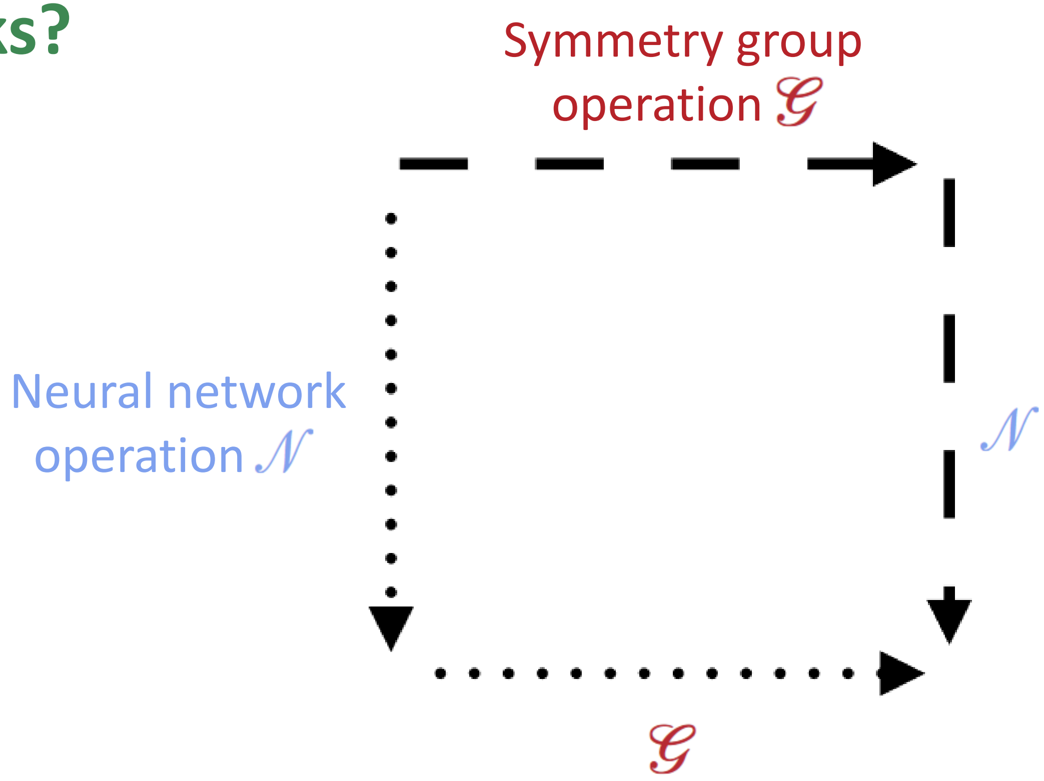
EQUIVARIANT NEURAL NETWORKS

What are equivariant neural networks?

$$\mathscr{G}(\mathscr{N}(x)) = \mathscr{N}(\mathscr{G}(x))$$

Why equivariance?

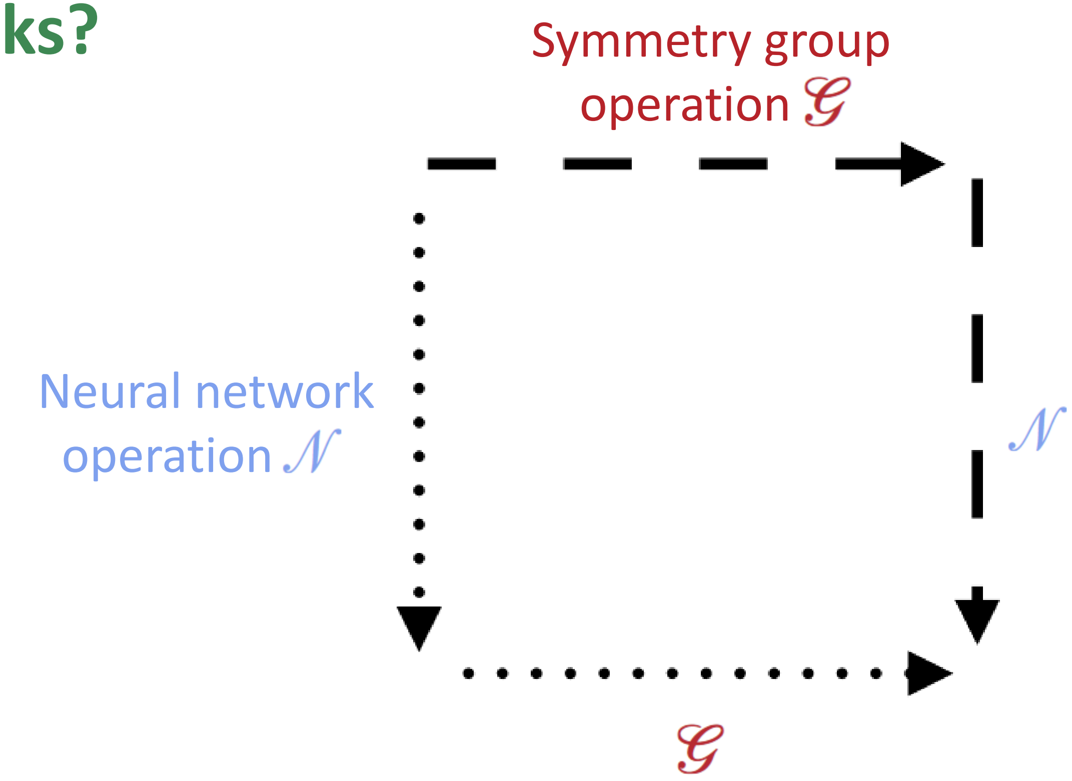
- Symmetries are important
- Symmetries are hard to learn



EQUIVARIANT NEURAL NETWORKS

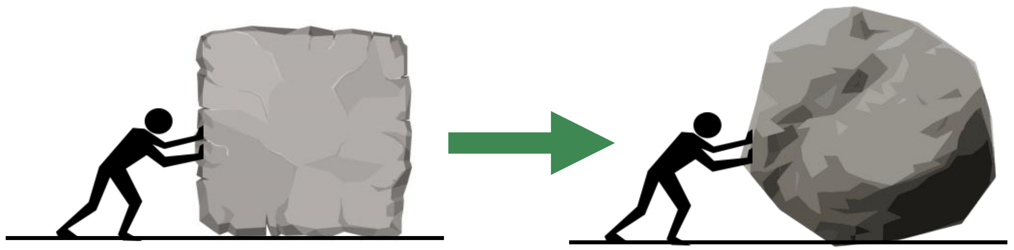
What are equivariant neural networks?

$$\mathscr{G}(\mathscr{N}(x)) = \mathscr{N}(\mathscr{G}(x))$$



Why equivariance?

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



EQUIVARIANT NEURAL NETWORKS

What are equivai

$\mathcal{G}(\mathcal{N}(x)) = \mathcal{N}(\mathcal{G}($

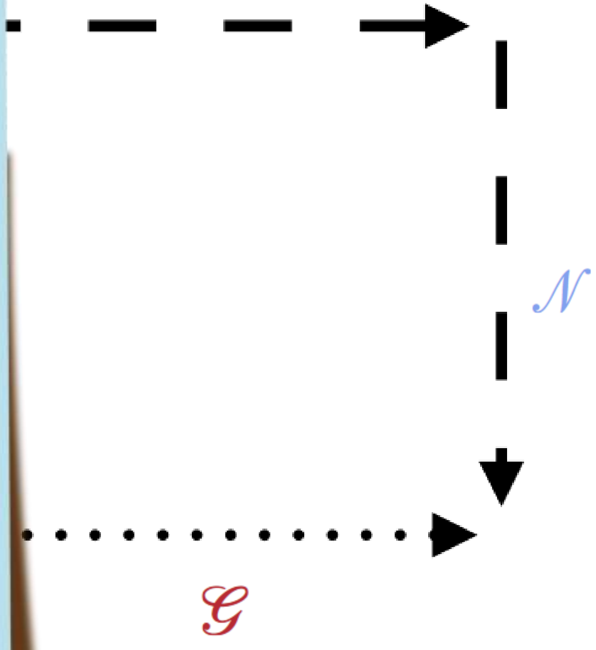
Why equivarianc

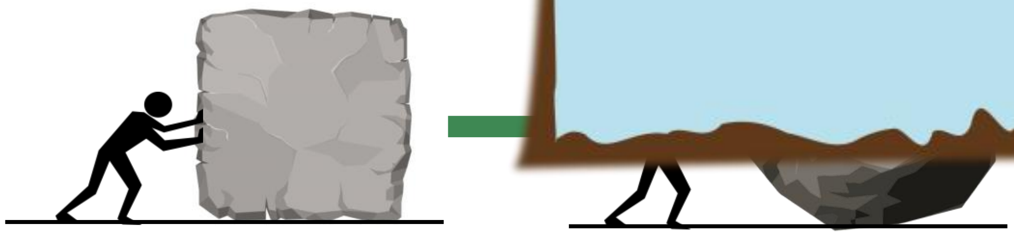
- Symmetries ar
- Symmetries ar
- More efficient

What is our recipe?

- Geometric Algebra
- Transformer

Symmetry group operation $\mathcal{G}$





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

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

Same properties as the gamma matrices!!

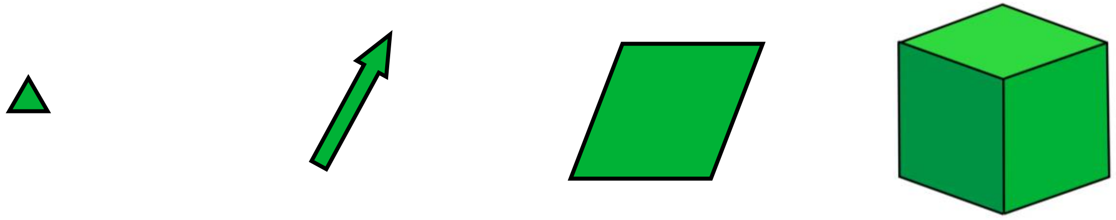
SPACETIME ALGEBRA

- **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

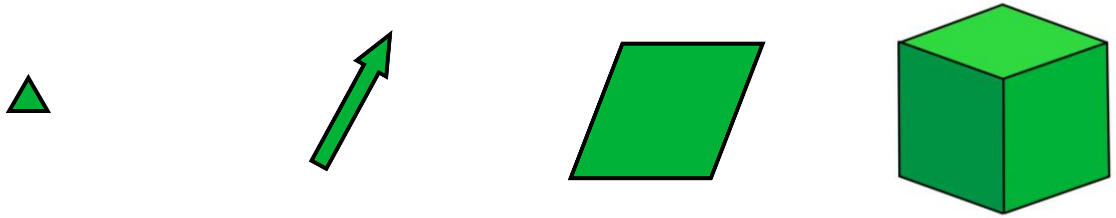
SPACETIME ALGEBRA

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

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

Same properties as the gamma matrices!!

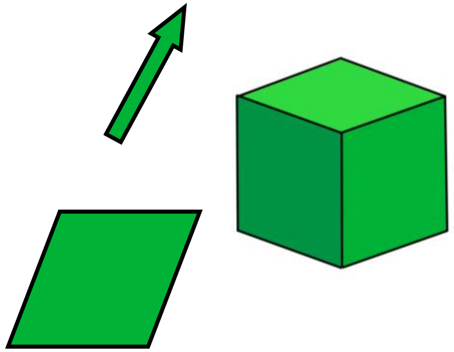
Scalar | Vector | Bivector | Axial vector | Pseudoscalar



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

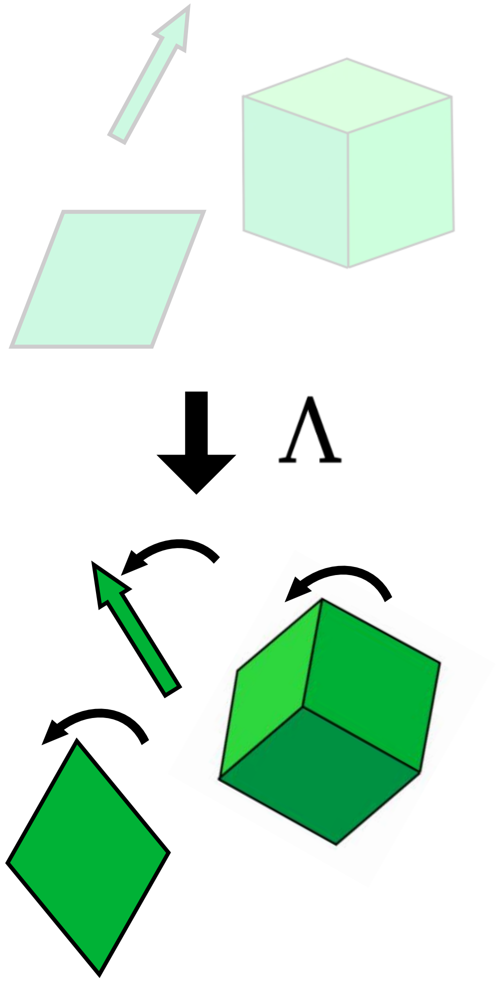
SPACETIME ALGEBRA

- Lorentz transformations



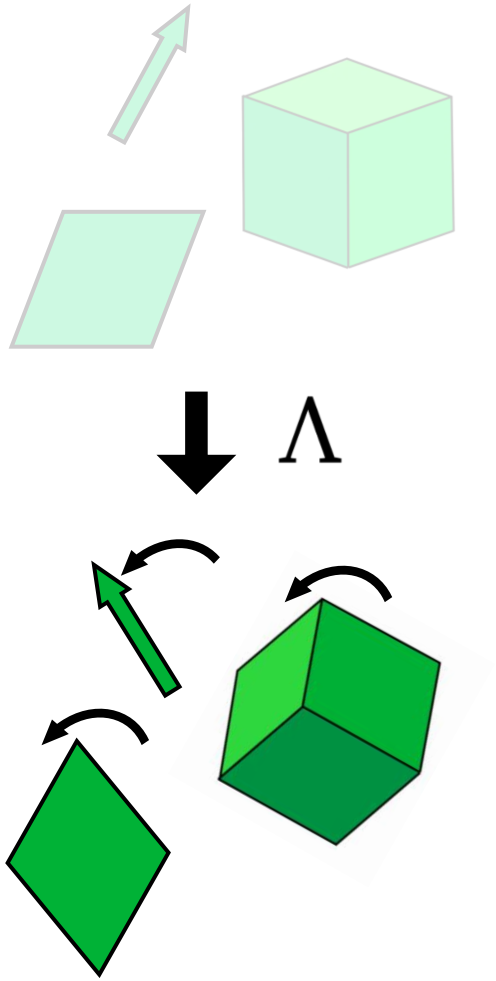
SPACETIME ALGEBRA

- Lorentz transformations



SPACETIME ALGEBRA

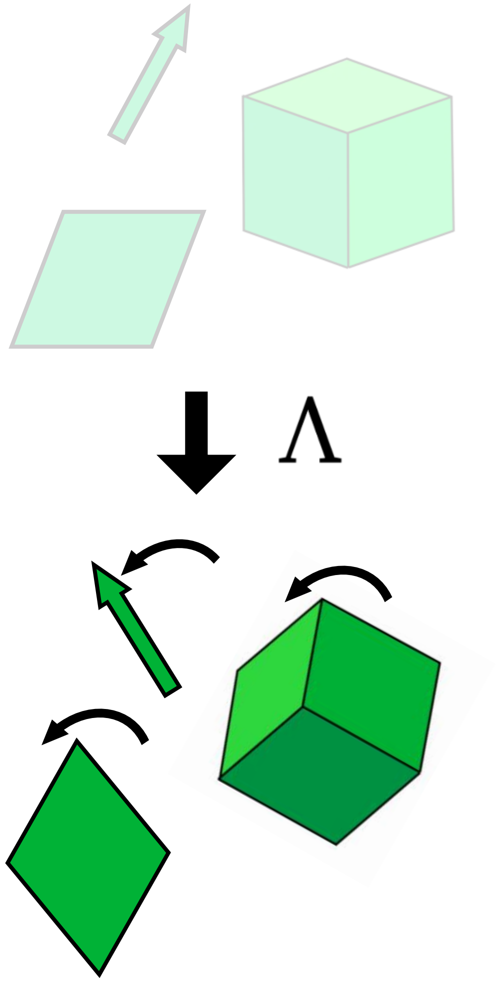
- Lorentz transformations



- Each grade transforms separately

SPACETIME ALGEBRA

- Lorentz transformations



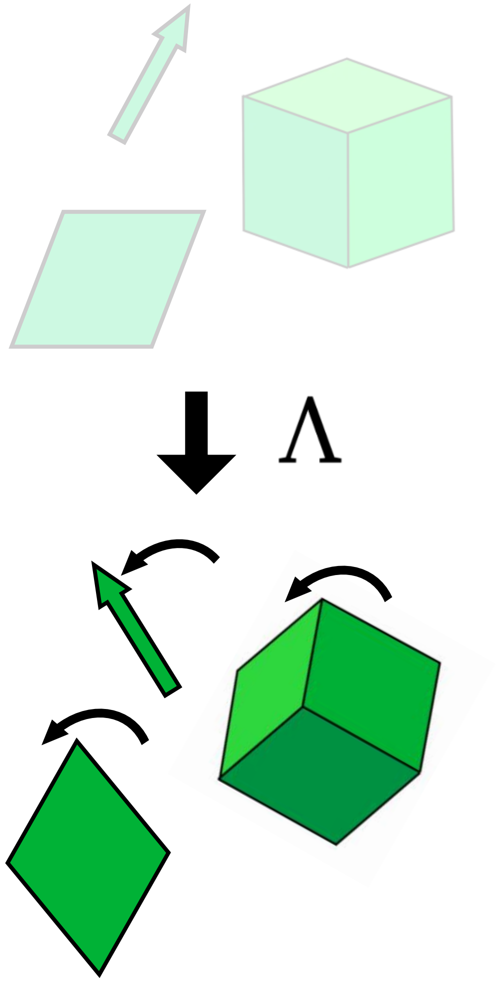
- Each grade transforms separately

⇓

- Equivariant maps can never mix grade components

SPACETIME ALGEBRA

- Lorentz transformations



- Each grade transforms separately

⇩

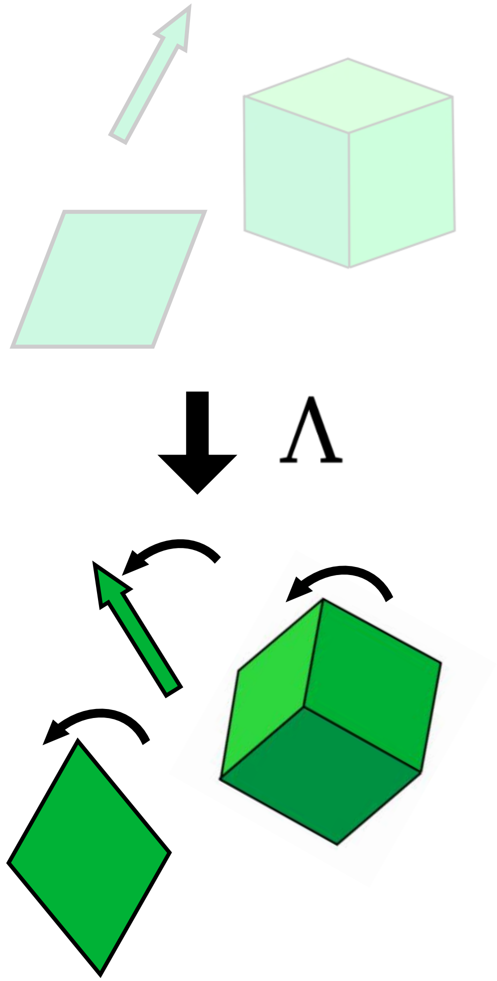
- Equivariant maps can never mix grade components

⇩

- Easy to build equivariant versions of ordinary networks

SPACETIME ALGEBRA

- Lorentz transformations



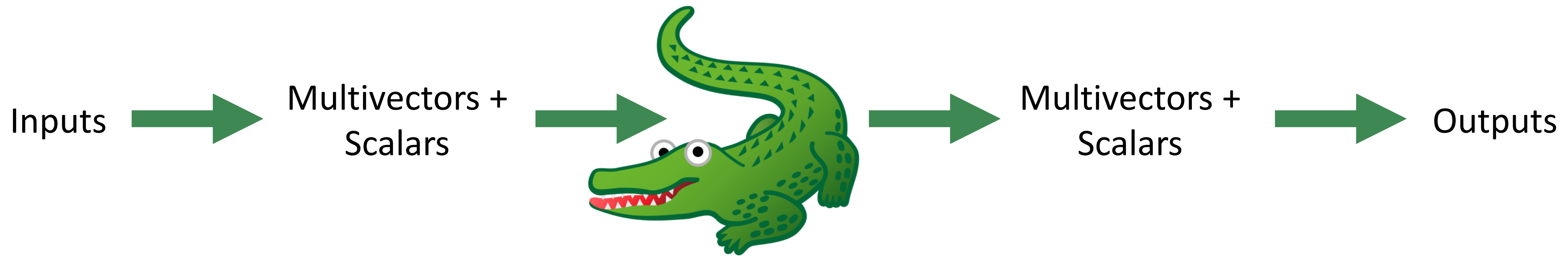
- Each grade transforms separately

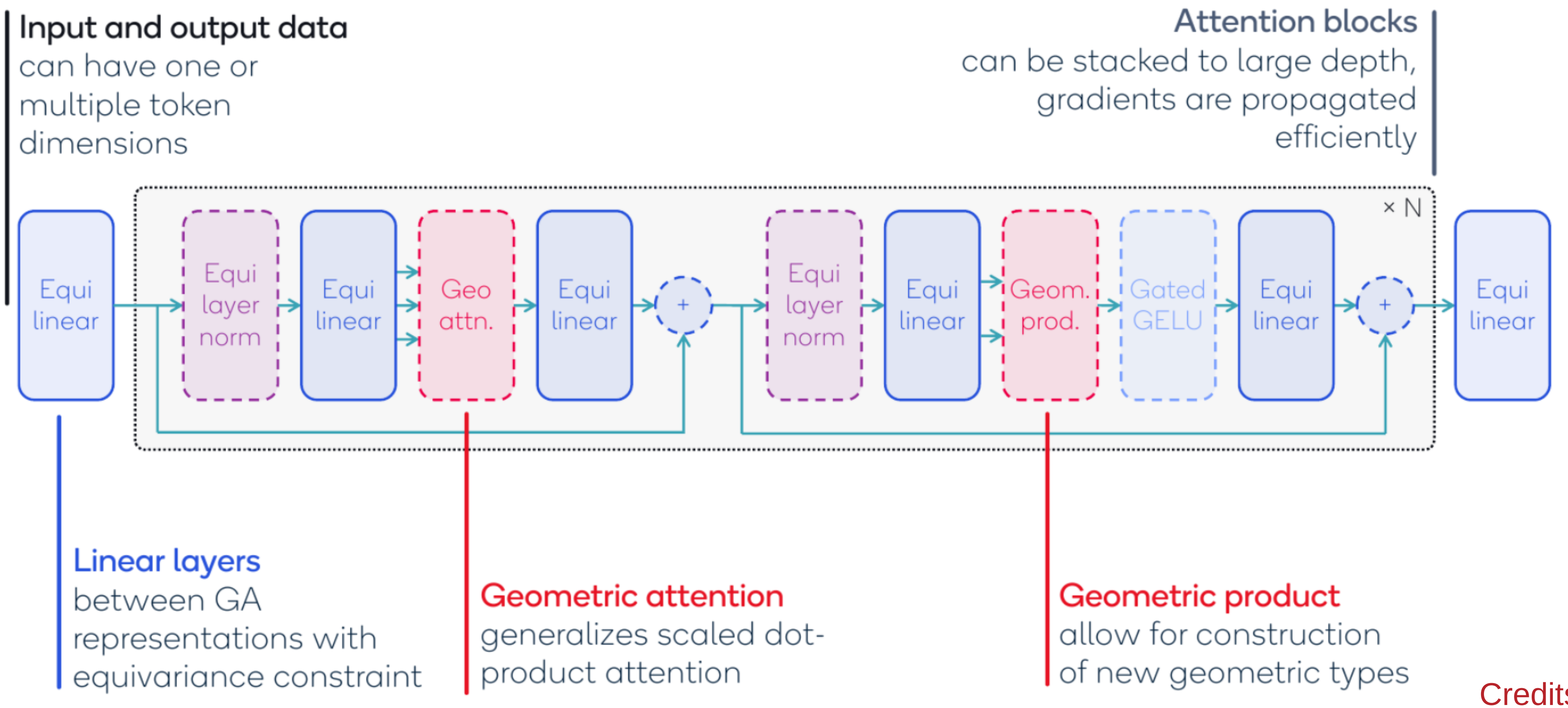
- Equivariant maps can never mix grade components

- Easy to build equivariant versions of ordinary networks

- Equivariance + Transformer = **L-GATr**

L-GATr

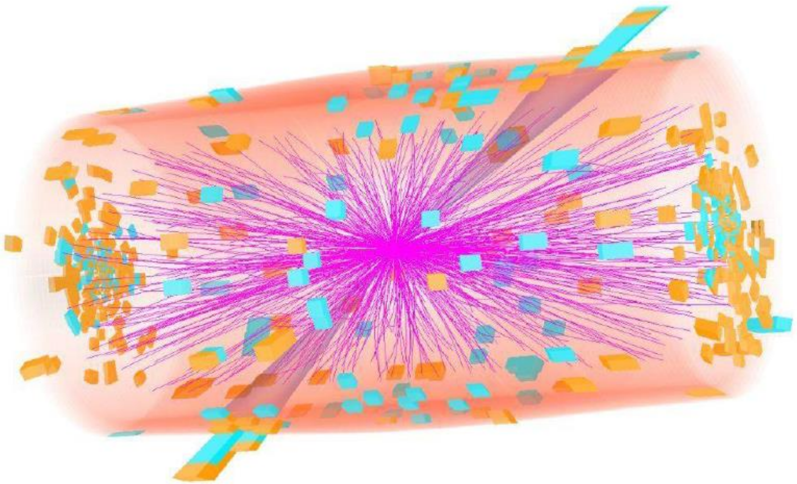




Credits to Johann Brehmer

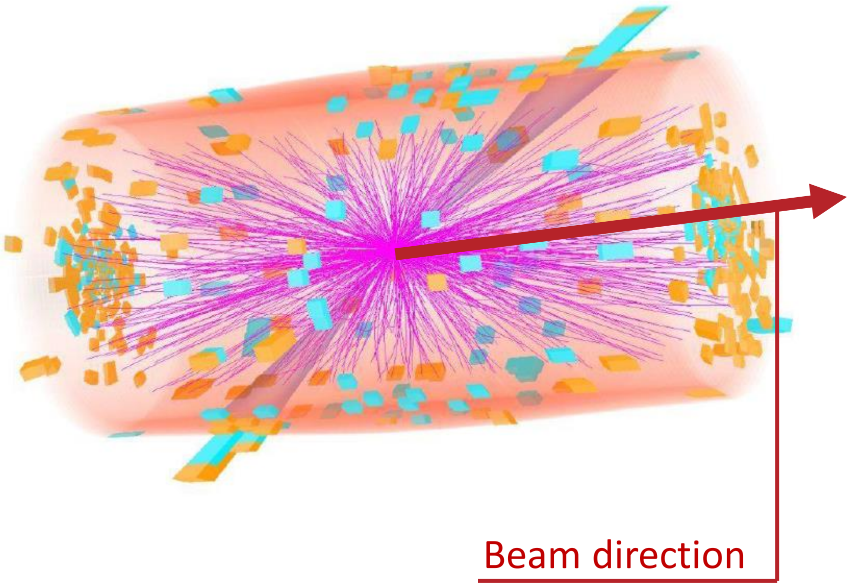
L-GATr

- Key feature: **Symmetry breaking**



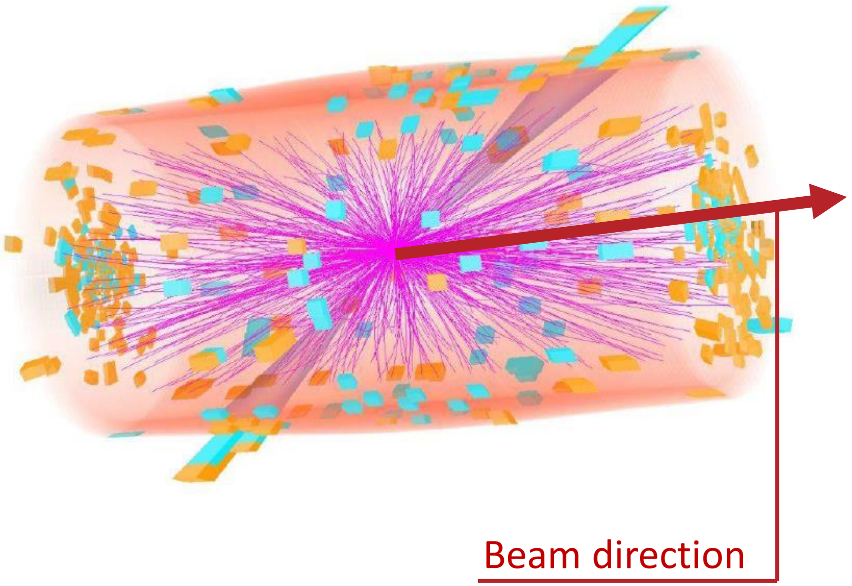
L-GATr

- Key feature: **Symmetry breaking**



L-GATr

- Key feature: **Symmetry breaking**

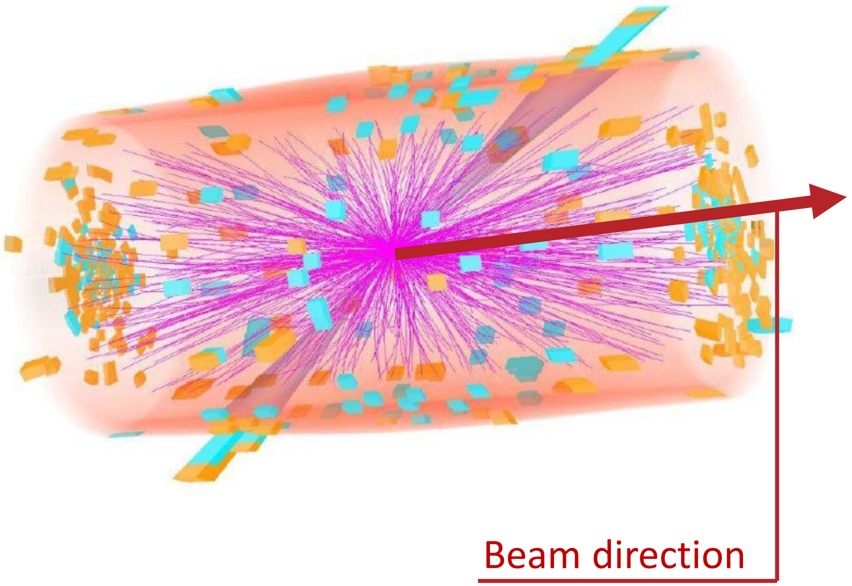


-Beam reference: $(1, 0, 0, \pm 1)$

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

L-GATr

- Key feature: **Symmetry breaking**



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

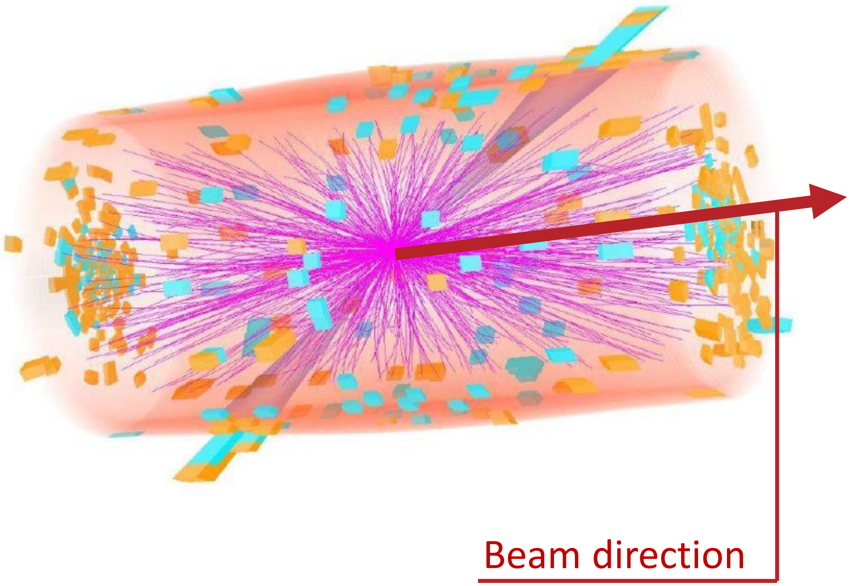
$SO(1,3) \rightarrow$ boosts + rotations around the beam

-Time reference: $(1, 0, 0, 0)$

$SO(1,3) \rightarrow SO(3)$

L-GATr

- Key feature: **Symmetry breaking**



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

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

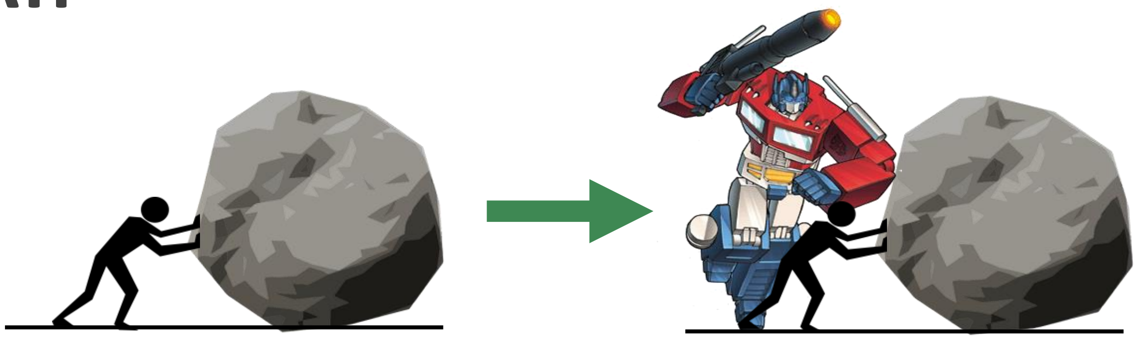
-Time reference: $(1,0,0,0)$

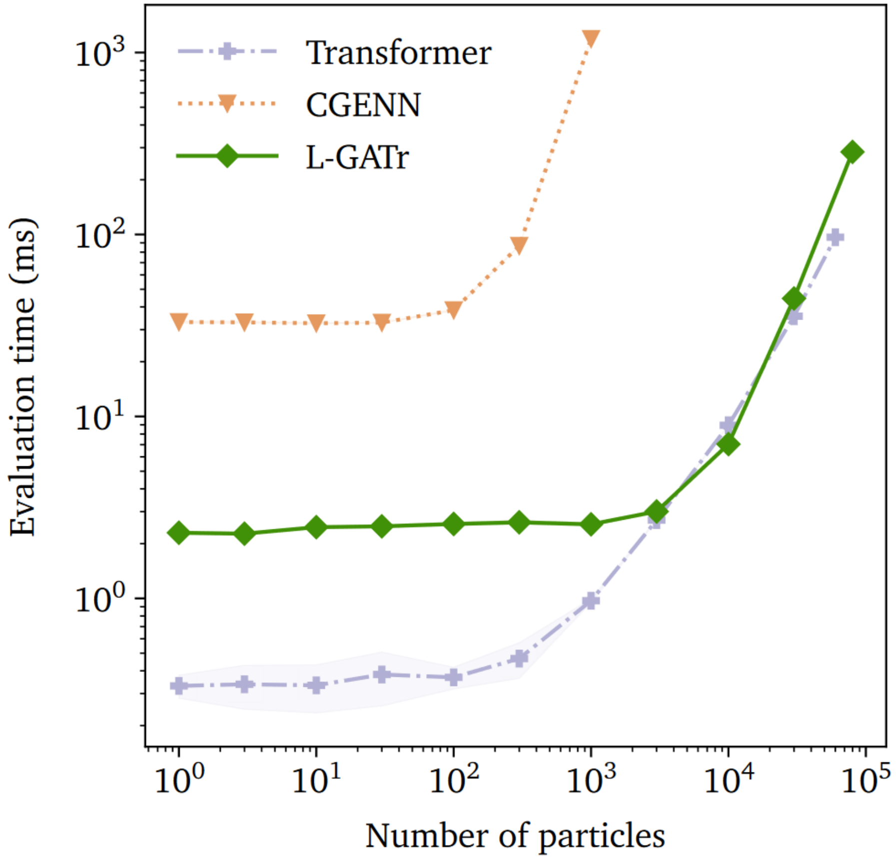
$SO(1,3) \rightarrow SO(3)$

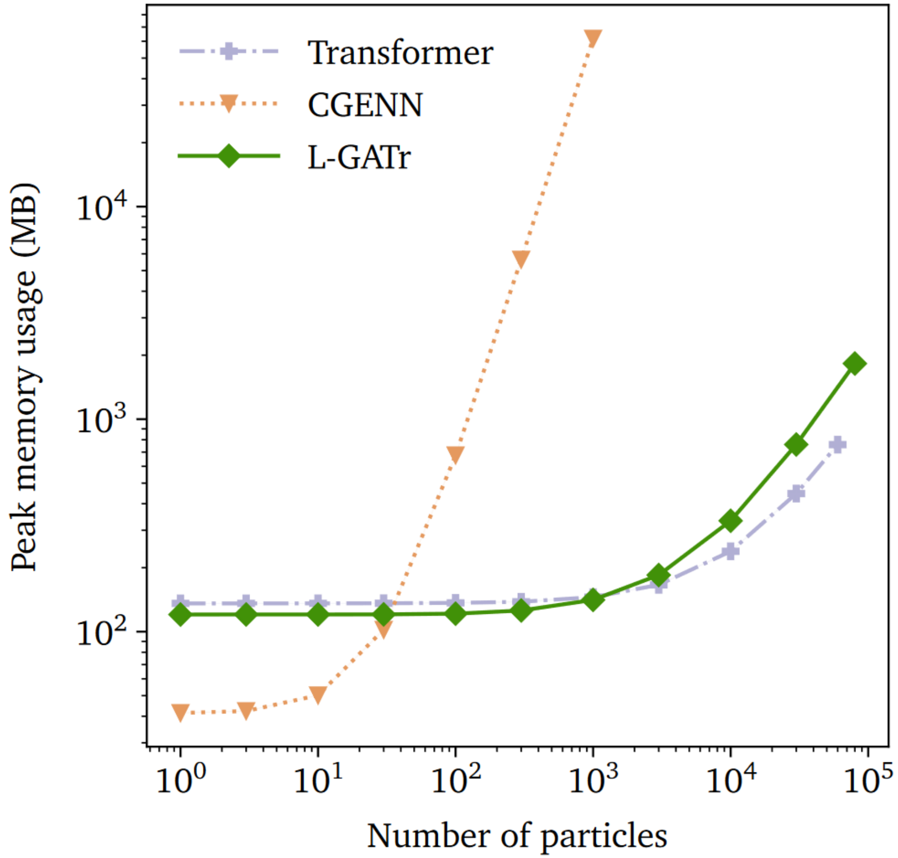
Beam	Time	Embedding	$1/\epsilon_B$ ($\epsilon_S = 0.3$)
–	✗	Particle	1422
Spacelike	✗	Particle	1905
All planes	✓	Particle	2009
–	✓	Token	1923
xy plane	✓	Channel	2060
Spacelike	✓	Particle	2152
Lightlike	✓	Particle	2114
xy plane	✓	Particle	2240

L-GATr

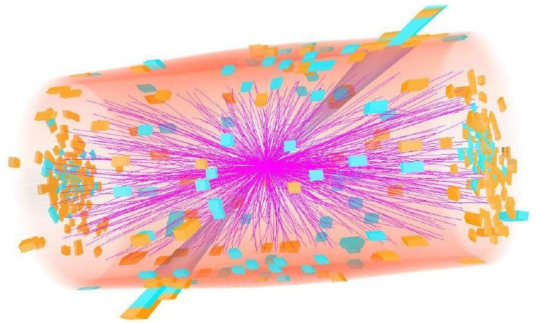
- Key feature: Transformer scaling







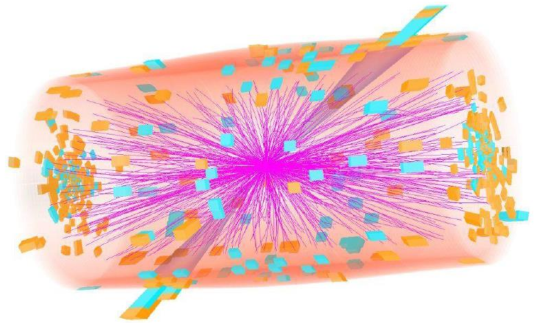
L-GATr TAGGING



- Experiment: **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

L-GATr TAGGING



- Experiment: **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

L-GATr TAGGING

- Experiment: **JetClass tagging**

  - Large and comprehensive jet dataset

  - 100M events

  - 10 classes

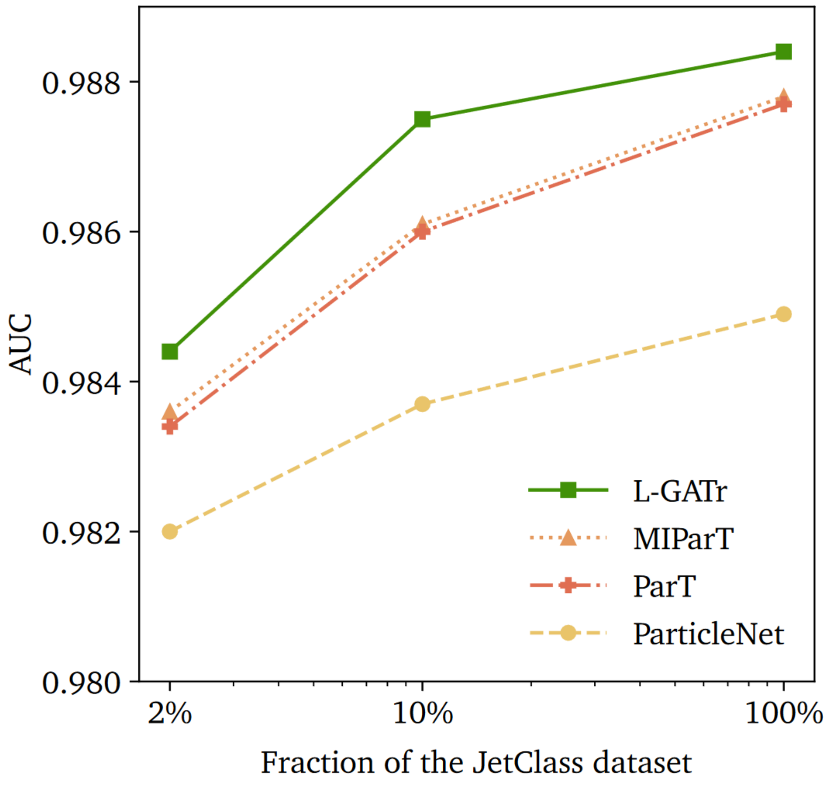
L-GATr TAGGING

- Experiment: **JetClass tagging**
  - Large and comprehensive jet dataset
  - 100M events
  - 10 classes

	All classes Accuracy	AUC	$H \to b\bar{b}$ Rej$_{50\%}$	$H \to c\bar{c}$ Rej$_{50\%}$	$H \to gg$ Rej$_{50\%}$	$H \to 4q$ Rej$_{50\%}$	$H \to l\nu q\bar{q}'$ Rej$_{99\%}$	$t \to bq\bar{q}'$ Rej$_{50\%}$	$t \to bl\nu$ Rej$_{99.5\%}$	$W \to q\bar{q}'$ Rej$_{50\%}$	$Z \to q\bar{q}$ 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

L-GATr TAGGING

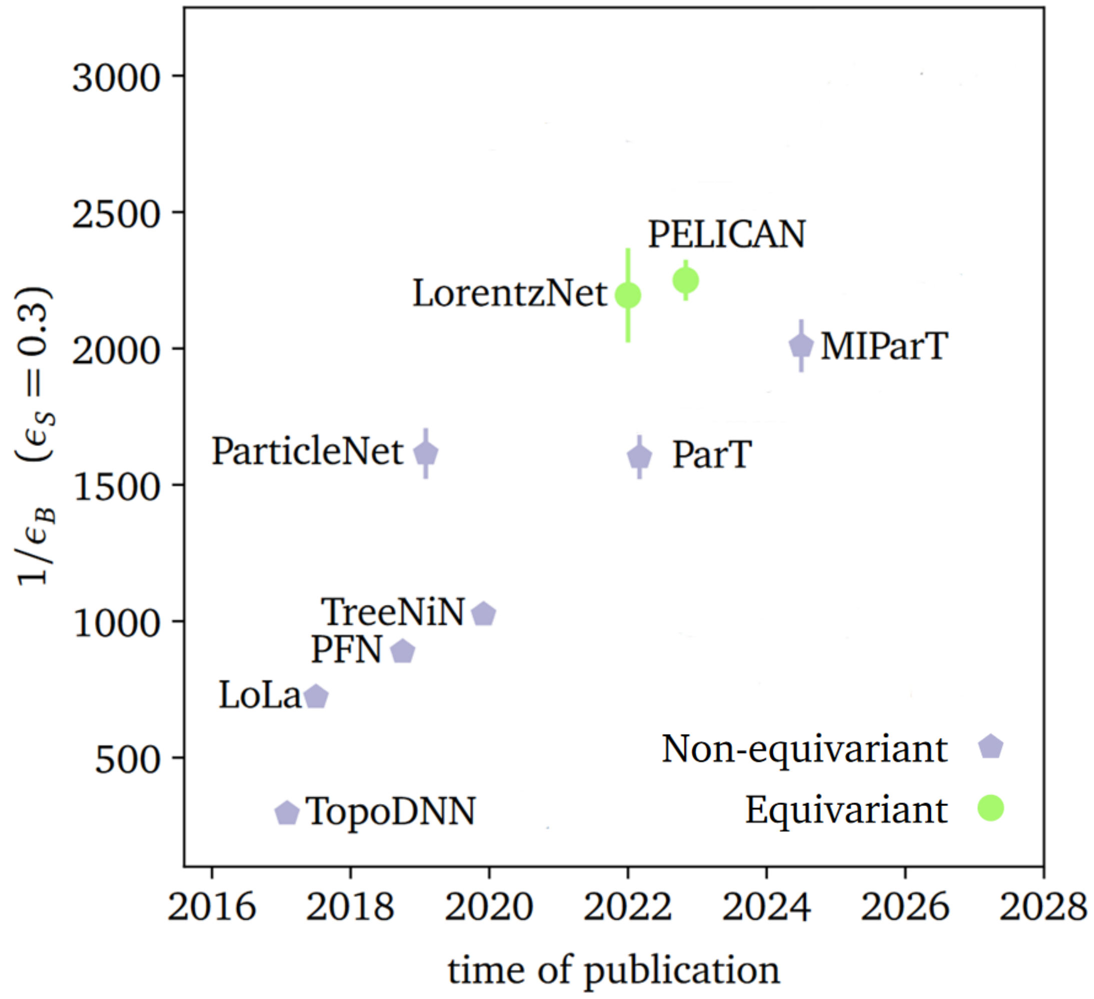
- Experiment: **JetClass tagging**
  - Large and comprehensive jet dataset
  - 100M events
  - 10 classes



	All classes		$H \to b\bar{b}$	$H \to c\bar{c}$	$H \to gg$	$H \to 4q$	$H \to l\nu q\bar{q}'$	$t \to bq\bar{q}'$	$t \to bl\nu$	$W \to q\bar{q}'$	$Z \to q\bar{q}$
	Accuracy	AUC	$\text{Rej}_{50\%}$	$\text{Rej}_{50\%}$	$\text{Rej}_{50\%}$	$\text{Rej}_{50\%}$	$\text{Rej}_{99\%}$	$\text{Rej}_{50\%}$	$\text{Rej}_{99.5\%}$	$\text{Rej}_{50\%}$	$\text{Rej}_{50\%}$
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

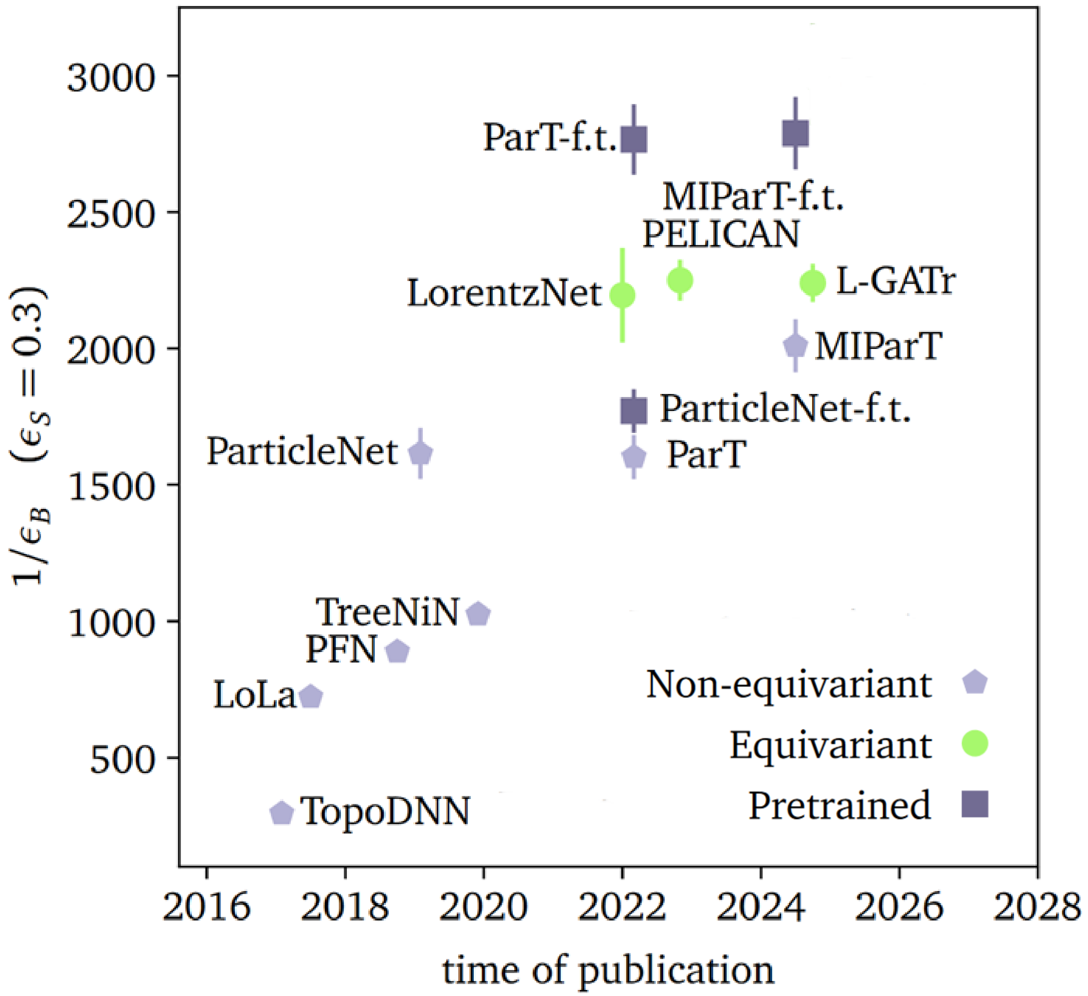
L-GATr TAGGING

- Impact of **pre-training**



L-GATr TAGGING

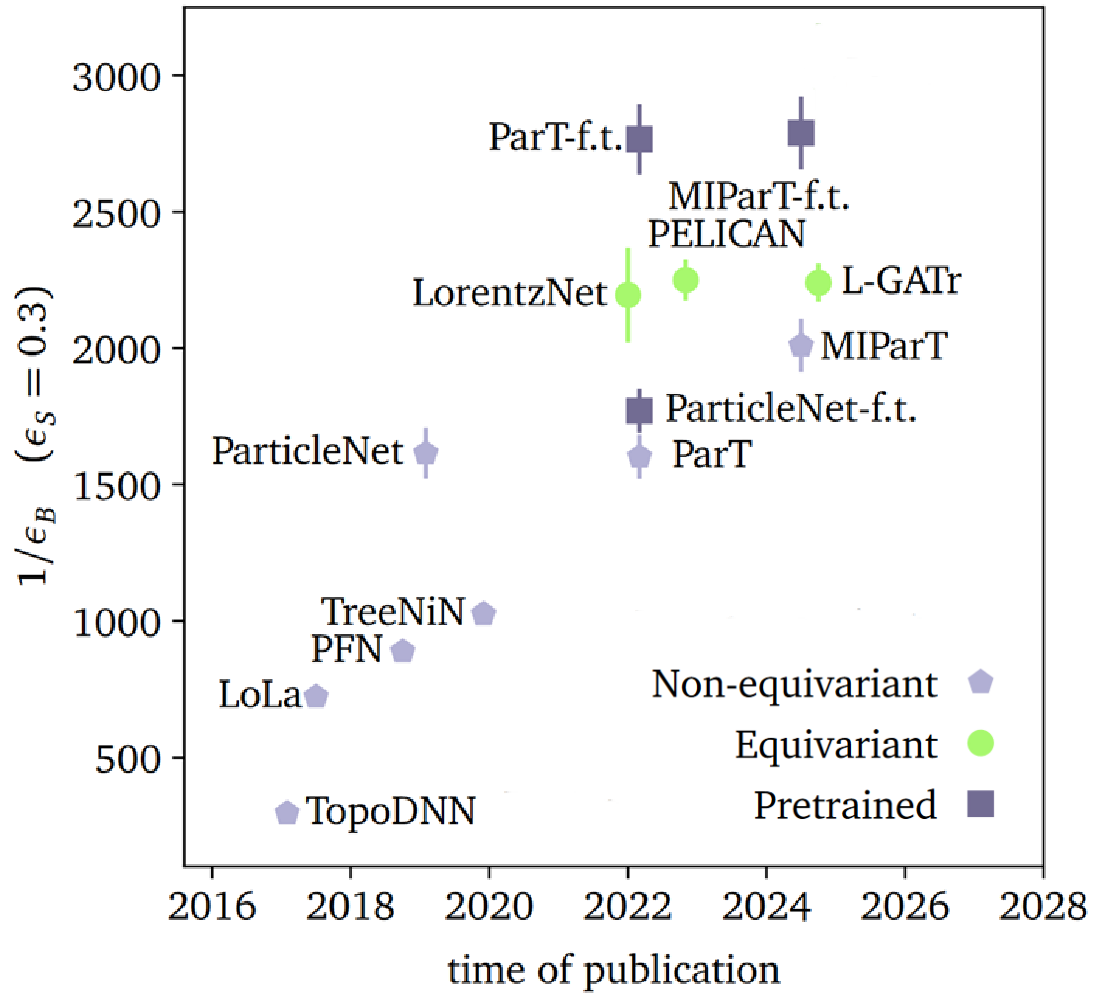
- Impact of **pre-training**



H. Qu et al., 2202,03772

L-GATr 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

L-GATr 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* [10]	0.942	0.9868	498 ± 18	2195 ± 173
CGENN* [14]	0.942	0.9869	500	2172
PELICAN* [42]	0.9426 ± 0.0002	0.9870 ± 0.0001	–	2250 ± 75
L-GATr* [35]	0.9423 ± 0.0002	0.9870 ± 0.0001	540 ± 20	2240 ± 70
ParticleNet-f.t. [60]	0.942	0.9866	487 ± 9	1771 ± 80
ParT-f.t. [60]	0.944	0.9877	691 ± 15	2766 ± 130
MIParT-f.t. [60]	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

L-GATr TAGGING

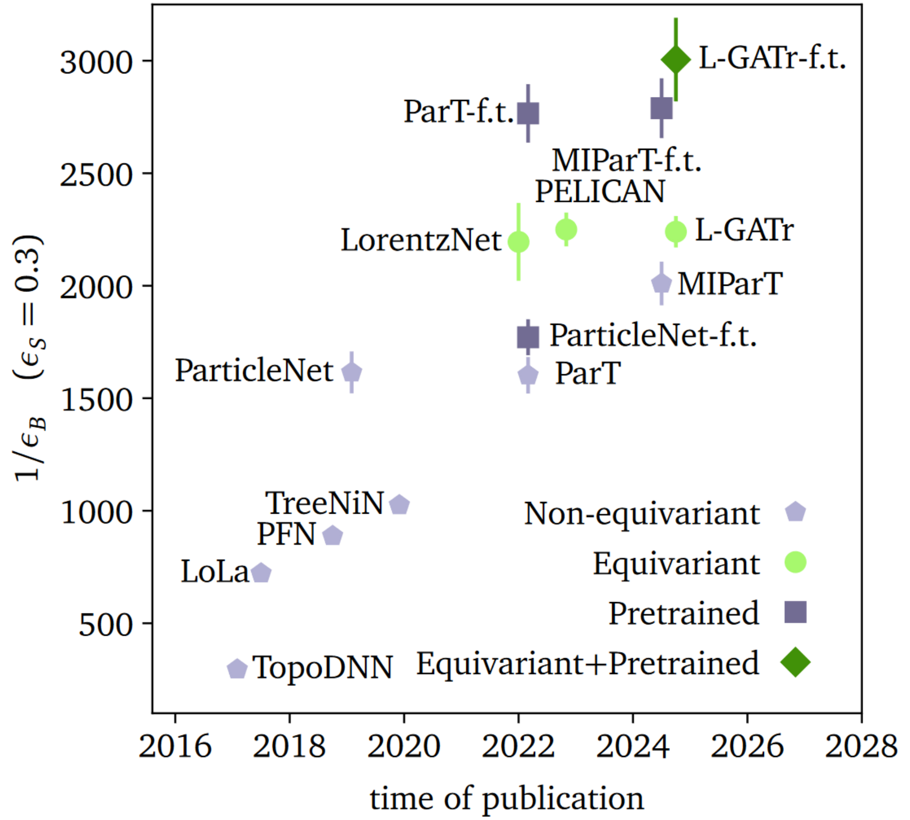
- What if we combine **pre-training** and **equivariance**?

Network	Accuracy	AUC	$1/\epsilon_B$ $(\epsilon_S = 0.5)$	$1/\epsilon_B$ $(\epsilon_S = 0.3)$
LorentzNet* [10]	0.942	0.9868	498 ± 18	2195 ± 173
CGENN* [14]	0.942	0.9869	500	2172
PELICAN* [42]	0.9426 ± 0.0002	0.9870 ± 0.0001	–	2250 ± 75
L-GATr* [35]	0.9423 ± 0.0002	0.9870 ± 0.0001	540 ± 20	2240 ± 70
ParticleNet-f.t. [60]	0.942	0.9866	487 ± 9	1771 ± 80
ParT-f.t. [60]	0.944	0.9877	691 ± 15	2766 ± 130
MIParT-f.t. [60]	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

L-GATr TAGGING

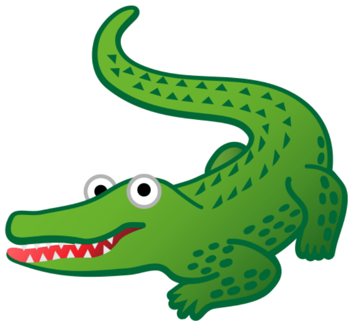
- What if we combine **pre-training** and **equivariance**?

Network	Accuracy	AUC	$1/\epsilon_B$ ($\epsilon_S = 0.5$)	$1/\epsilon_B$ ($\epsilon_S = 0.3$)
LorentzNet* [10]	0.942	0.9868	498 ± 18	2195 ± 173
CGENN* [14]	0.942	0.9869	500	2172
PELICAN* [42]	0.9426 ± 0.0002	0.9870 ± 0.0001	–	2250 ± 75
L-GATr* [35]	0.9423 ± 0.0002	0.9870 ± 0.0001	540 ± 20	2240 ± 70
ParticleNet-f.t. [60]	0.942	0.9866	487 ± 9	1771 ± 80
ParT-f.t. [60]	0.944	0.9877	691 ± 15	2766 ± 130
MIParT-f.t. [60]	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



CONCLUSIONS

- L-GATr sets a new benchmark on multiple tagging tasks
- L-GATr has a better scaling behavior than competing baselines
- L-GATr has a strong performance on multiple problems, namely regression and generation

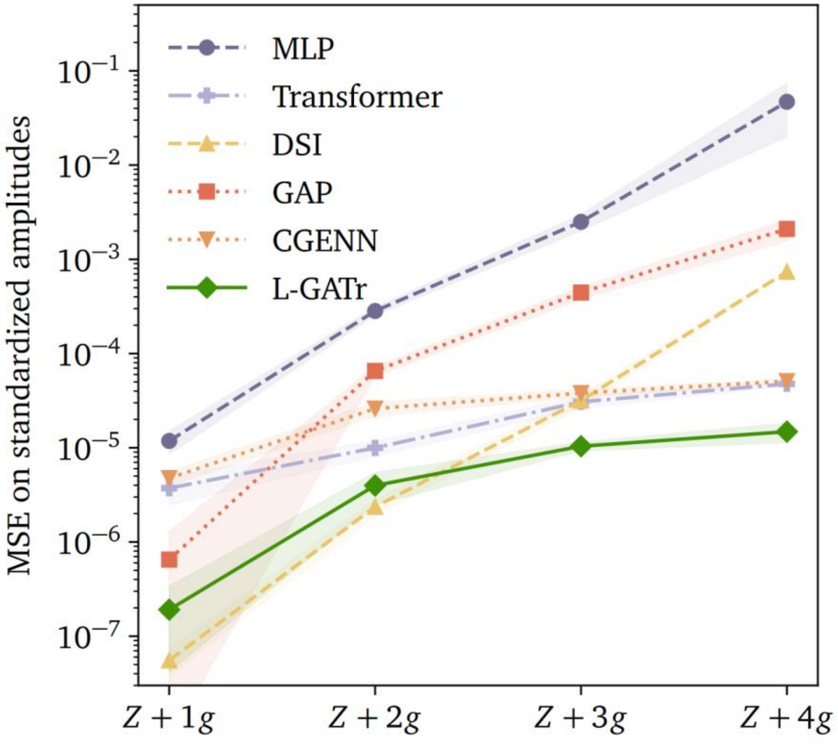


CONCLUSIONS

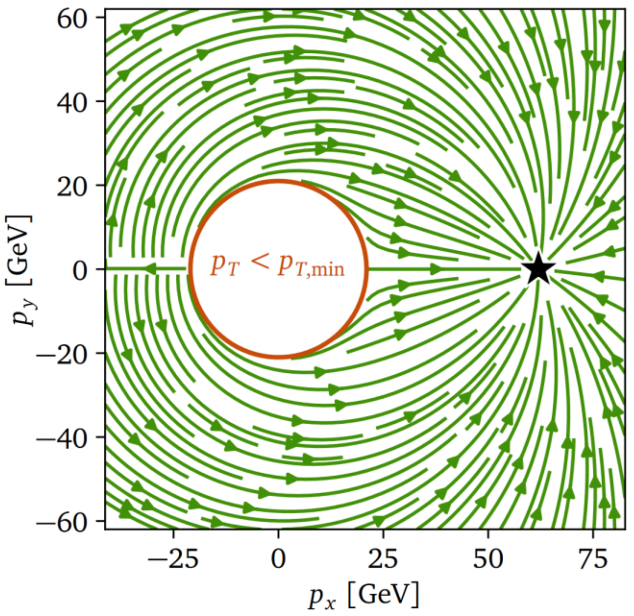
See talks by Jonas Spinner and Giovanni De Crescenzo!

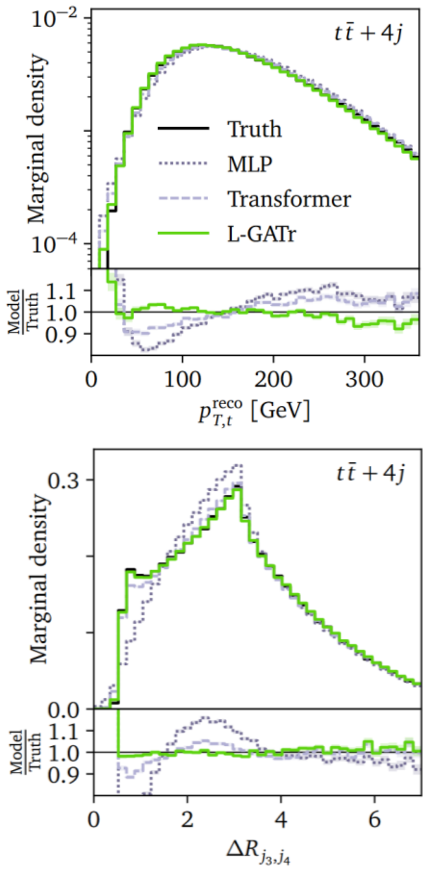
- Sneak peek at our other applications:

Amplitude regression



Event generation

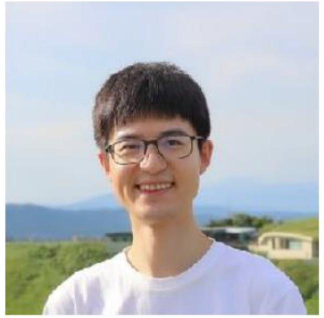





Jonas Spinner


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

What will **you** use L-GATr for?

L-GATr code