Lorentz-GATr

Lorentz-Equivariant Geometric Algebra Transformers for High-Energy Physics

Jonas Spinner*, Victor Breso*,
Pim de Haan, Tilman Plehn,
Jesse Thaler, Johann Brehmer

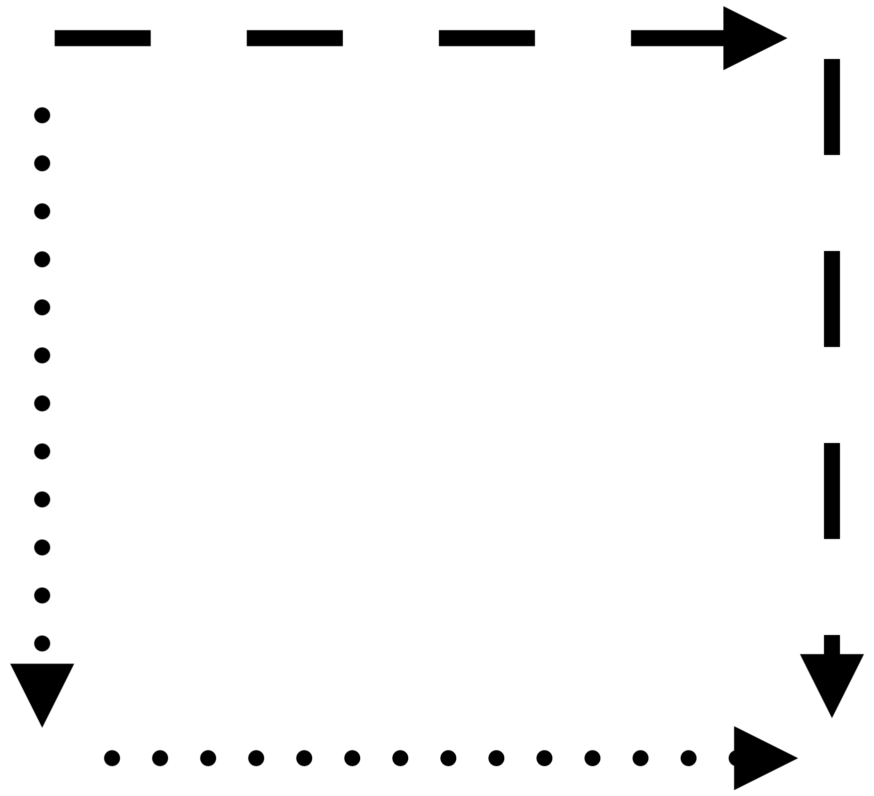
PHYSTAT 2024
Statistics meets machine learning

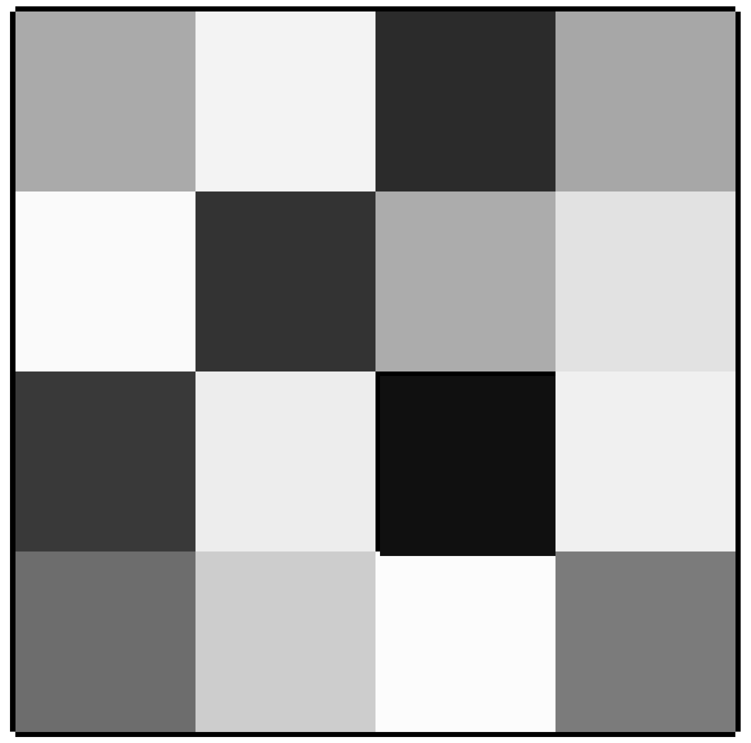
UNIVERSITÄT HEIDELBERG
ZUKUNFT SEIT 1386

Symmetries are key in high-energy physics…

… but its hard to build them into networks in a scalable way.

$$\begin{pmatrix} 1 \\ \gamma^{\mu} \\ \sigma^{\mu\nu} \\ \gamma^{\mu}\gamma_5 \\ \gamma_5 \end{pmatrix}$$

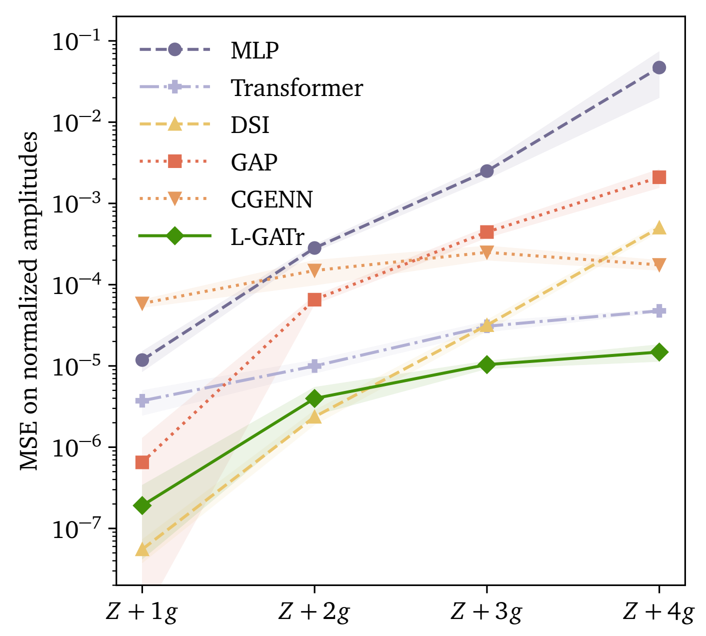




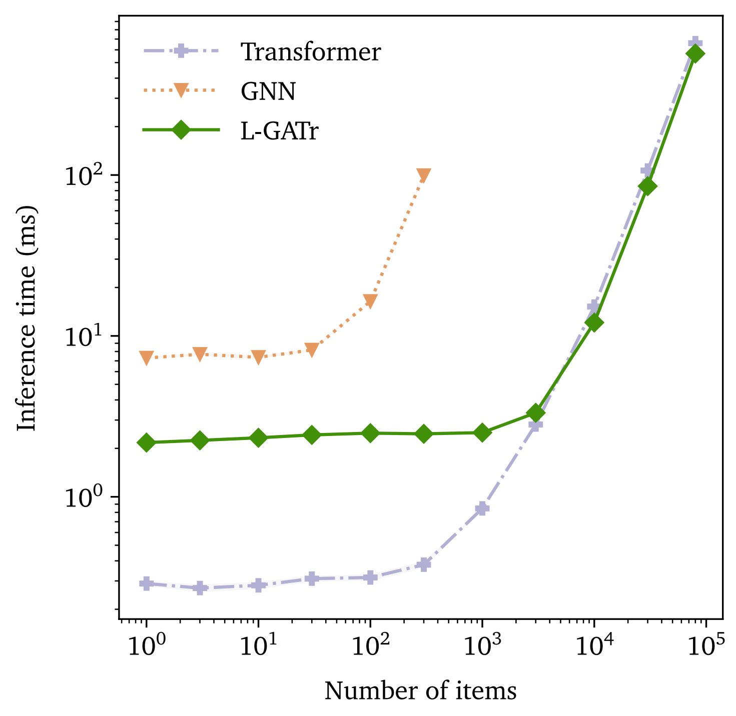
Geometric algebra representations

Equivariant layers

Transformer architecture



Strong performance on diverse problems

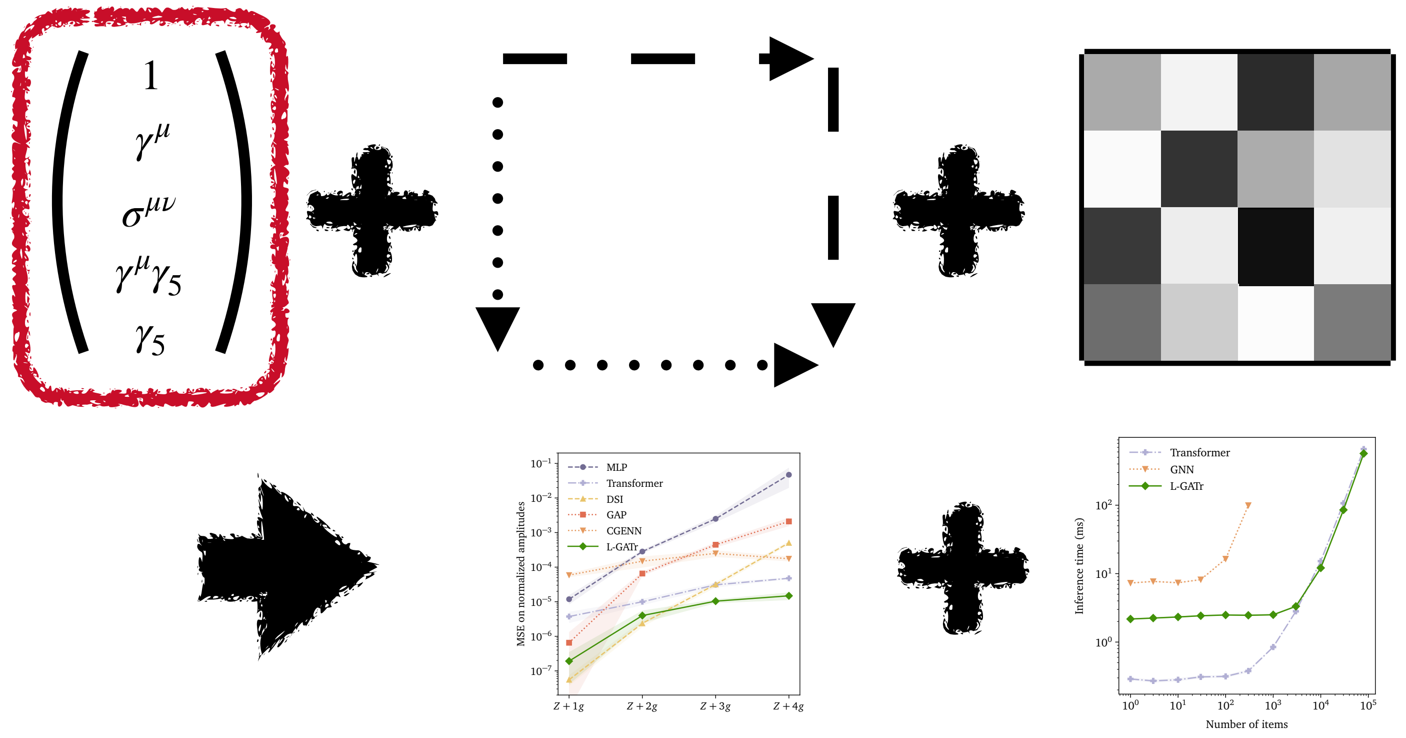


Scalable to thousands of tokens

GATr was originally developed for E(3) arXiv:2305.18415

Ingredients

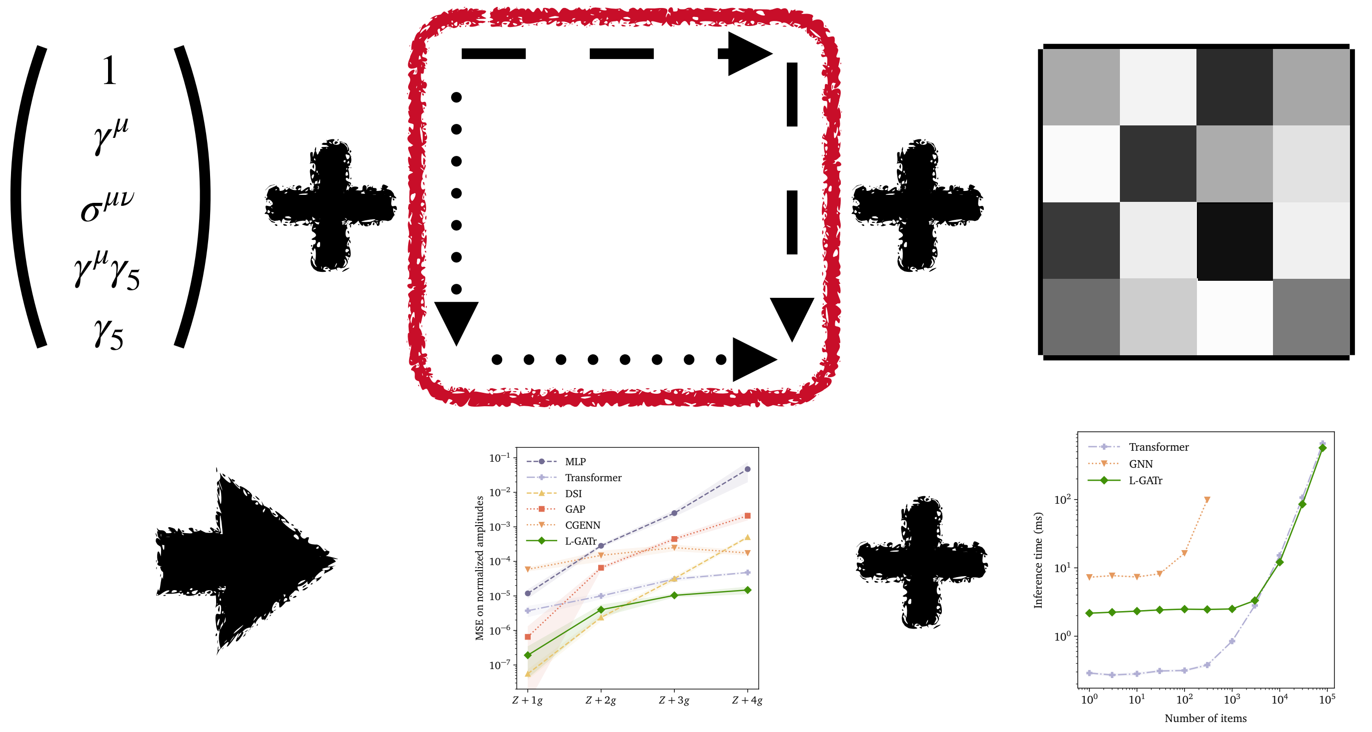
Geometric algebra representations

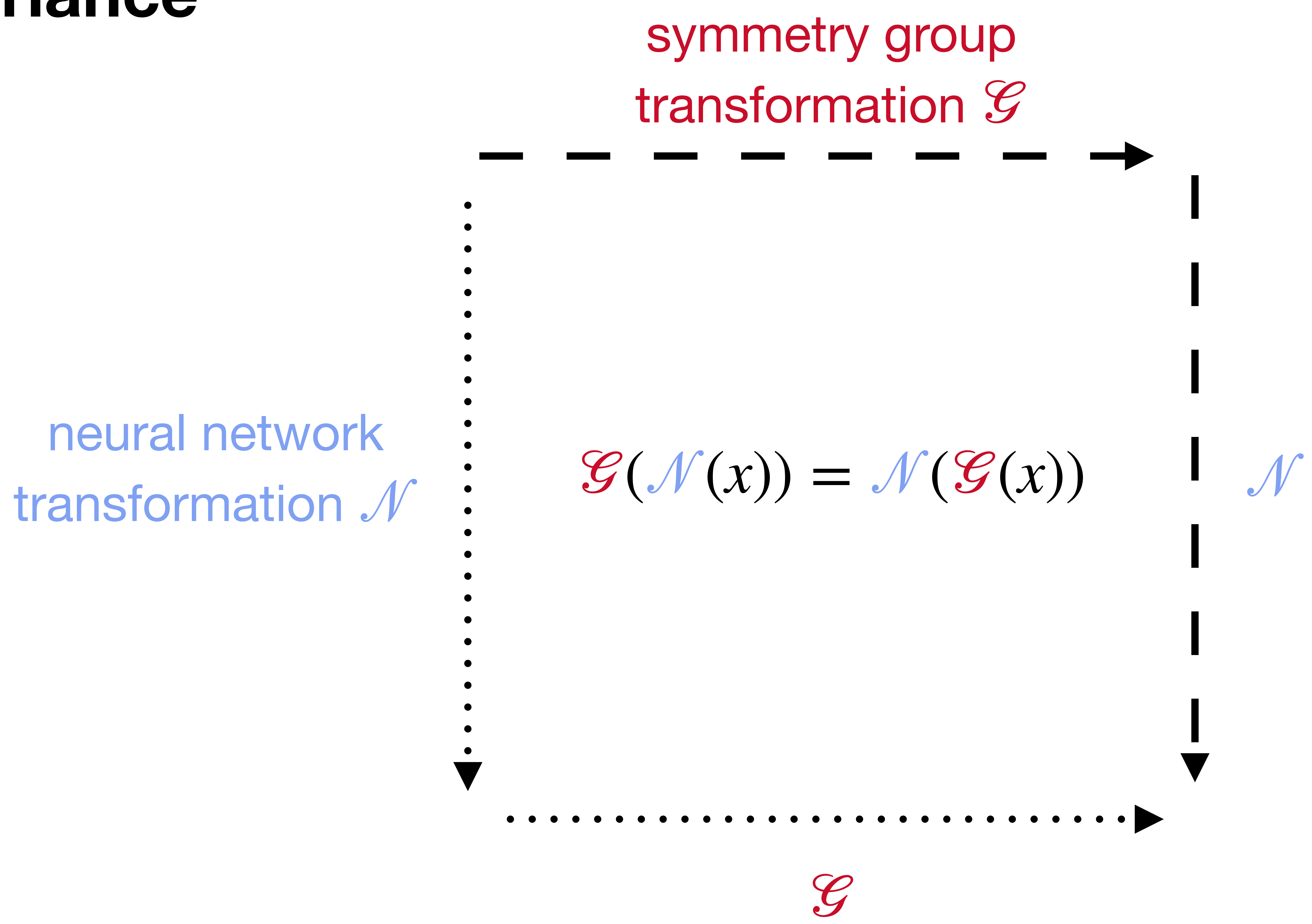


- Basis elements γ^μ of the geometric/Clifford algebra defined by $\{\gamma^\mu, \gamma^\nu\} = 2g^{\mu\nu}$
  - Operations: $\alpha x, \quad x + y, \quad x \cdot y$
  - General multivector: $x = x^S 1 + x^V_\mu \gamma^\mu + x^T_{\mu\nu}\sigma^{\mu\nu} + x^A_\mu \gamma^\mu \gamma_5 + x^P \gamma_5$
- We embed multivectors as $(x^S, x^V_0 \cdots x^V_3, x^T_{01} \cdots x^T_{23}, x^A_0 \cdots x^A_3, x^P) \in \mathbb{R}^{16}$
- Each GATr token contains n multivector and m scalar representations

Ingredients

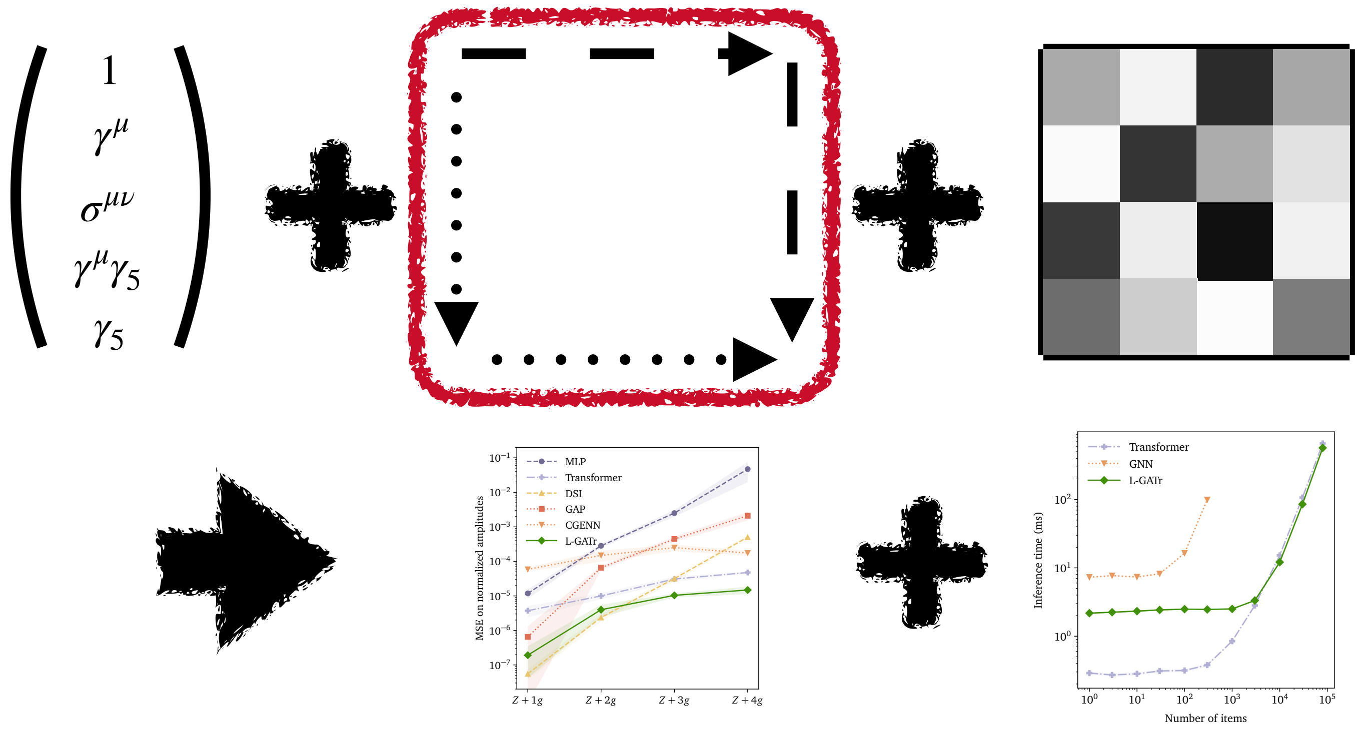
Equivariance





Ingredients

Equivariant layers



EquiLinear

$$\phi(x) = \sum_{k=0}^{4} v_k \langle x \rangle_k + \sum_{k=0}^{4} w_k \gamma_5 \langle x \rangle_k$$

Geometric product

$$\psi(x, y) = x \cdot y$$

Geometric attention

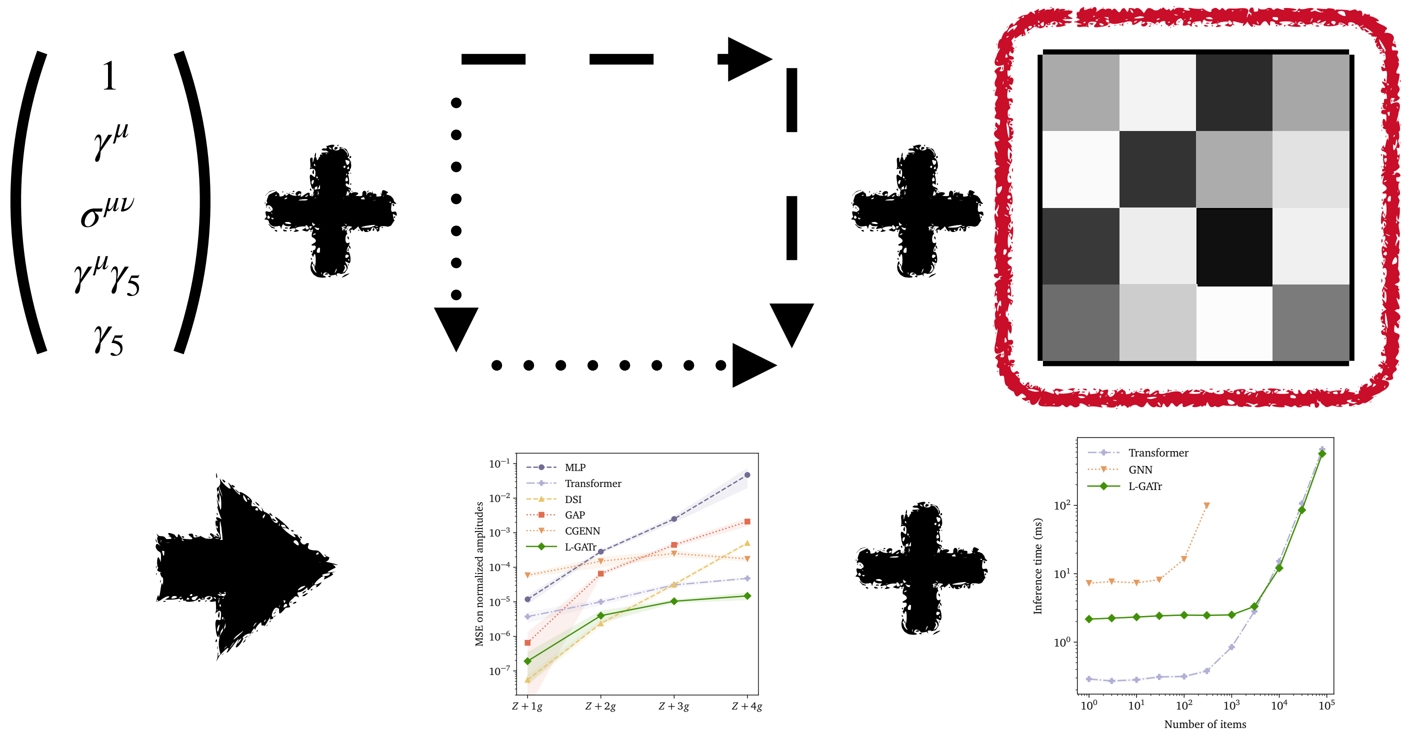
$$\text{Attention}(q, k, v)_{i\alpha} = \text{Softmax}_j \left(\frac{\langle q_{i\beta}, k_{j\beta} \rangle}{\sqrt{16n}} \right) v_{j\alpha}$$

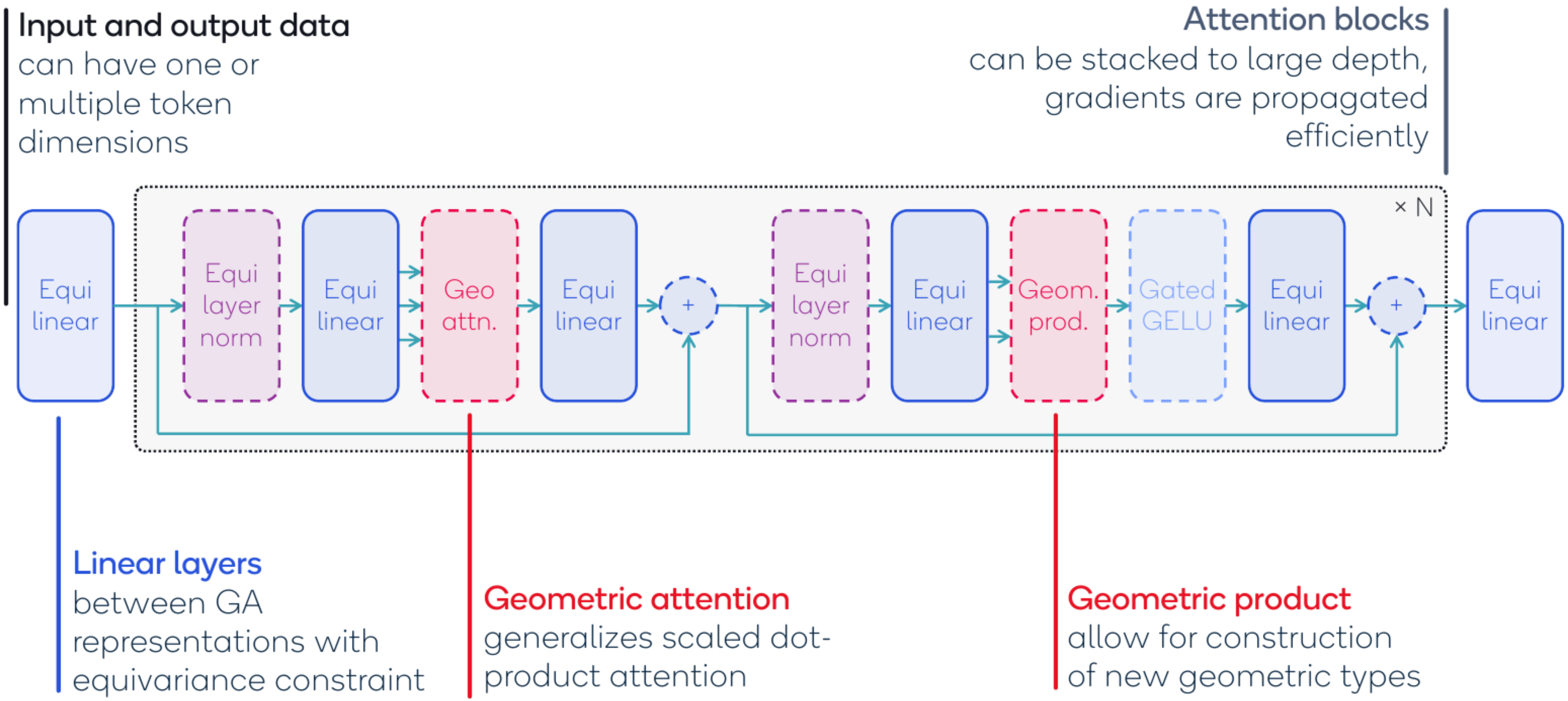
LayerNorm, dropout, activation function…

See bonus material

Ingredients

Transformer architecture

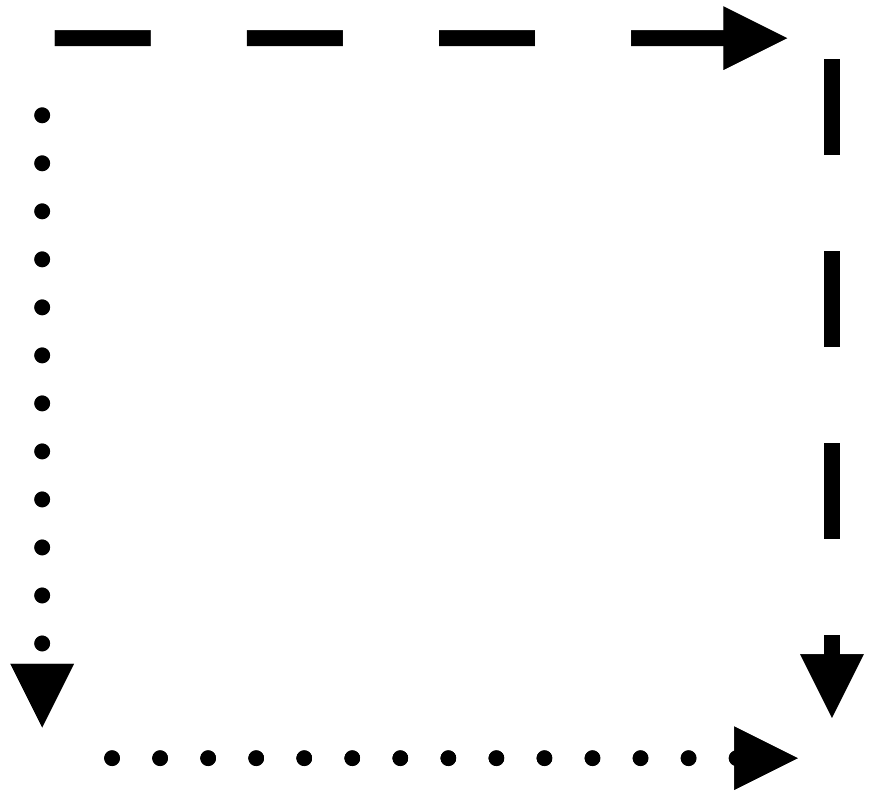




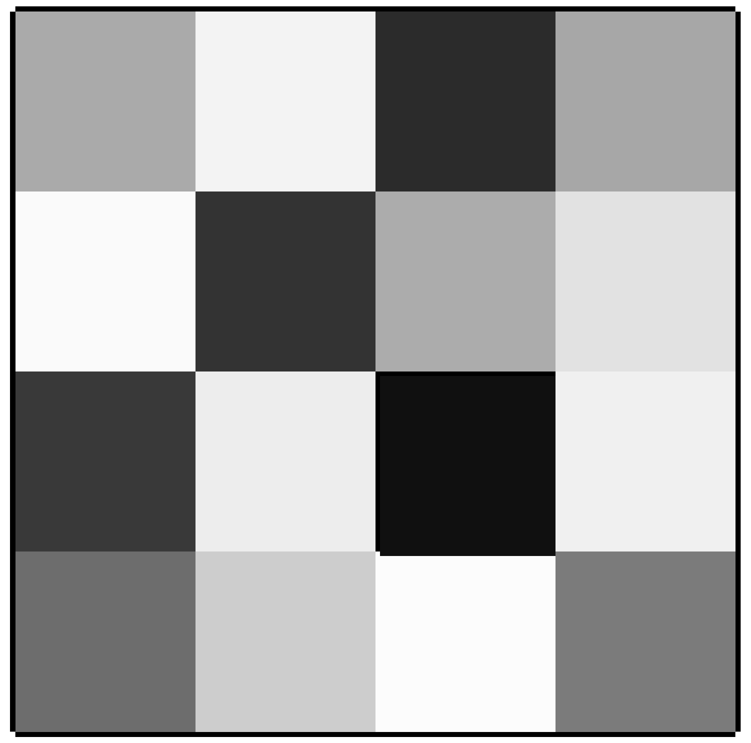
Credits to Johann Brehmer

$$\begin{pmatrix} 1 \\ \gamma^\mu \\ \sigma^{\mu\nu} \\ \gamma^\mu\gamma_5 \\ \gamma_5 \end{pmatrix}$$

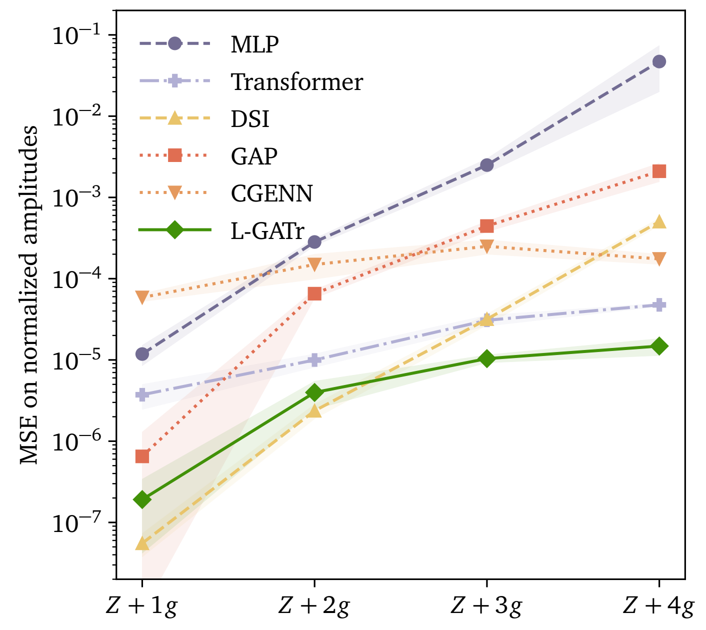
Geometric algebra
representations



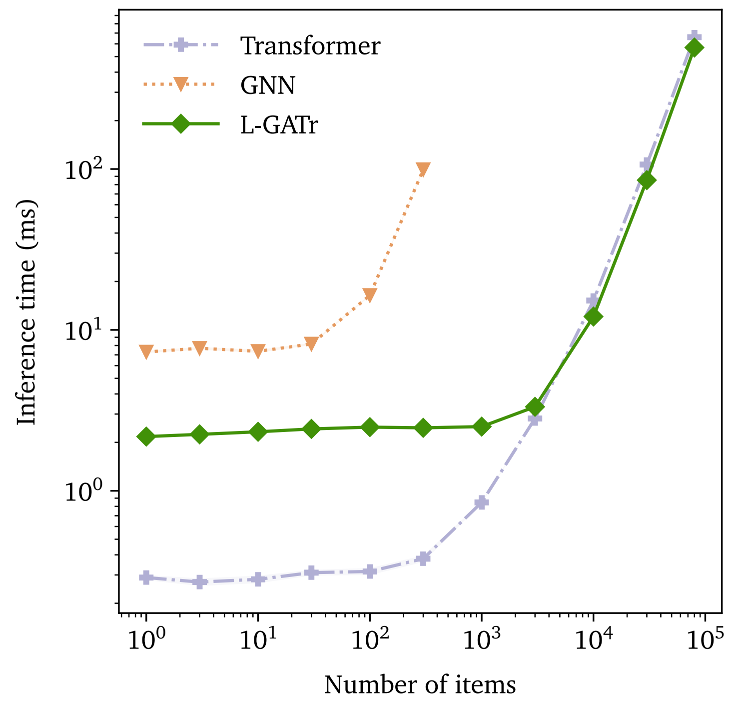
Equivariant
layers



Transformer
architecture



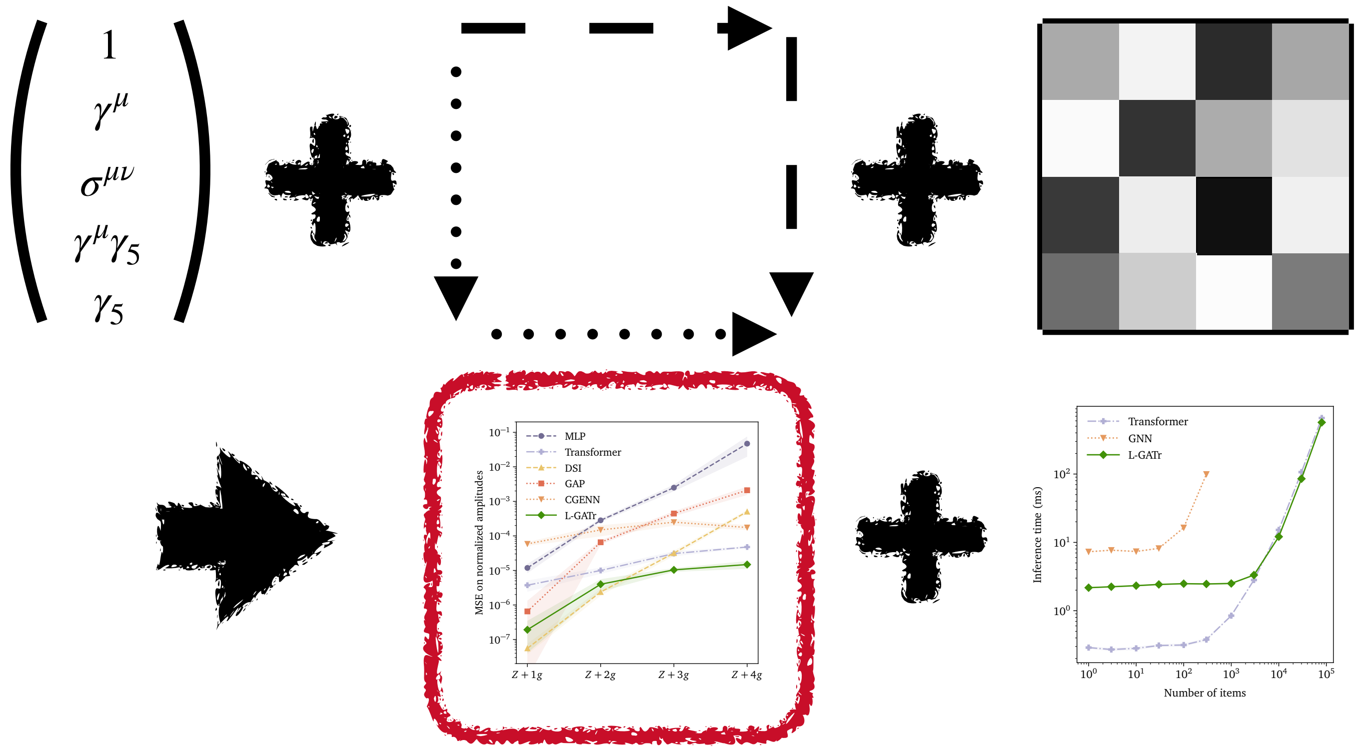
Strong performance
on diverse problems

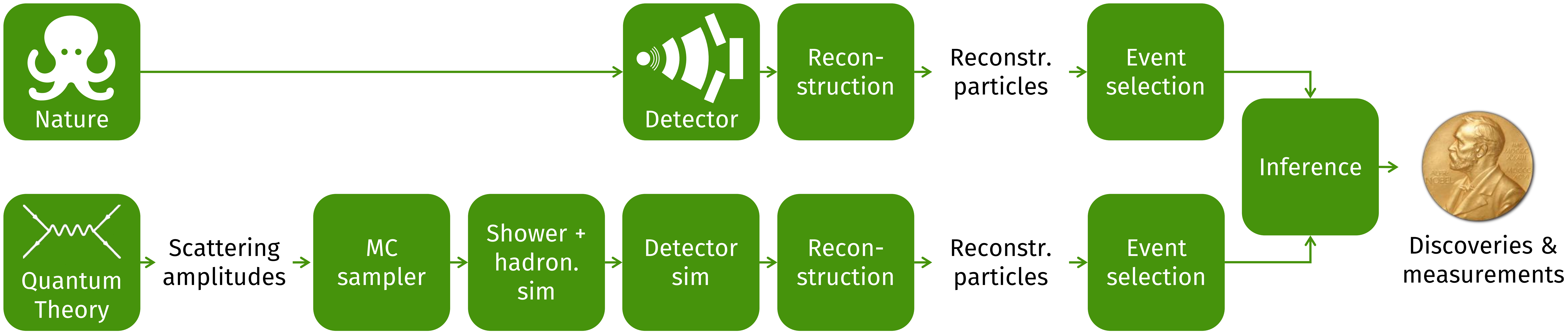


Scalable
to thousands of tokens

Experiments

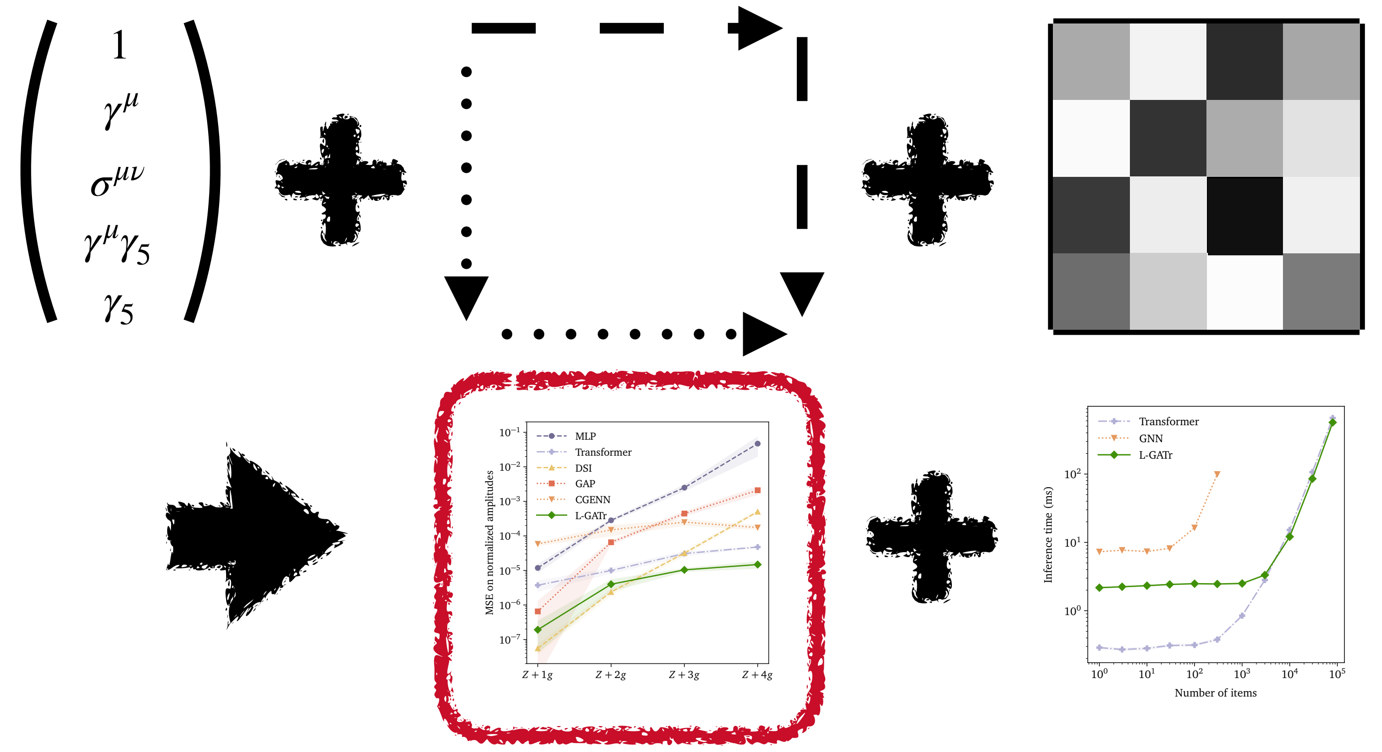
LHC simulation chain

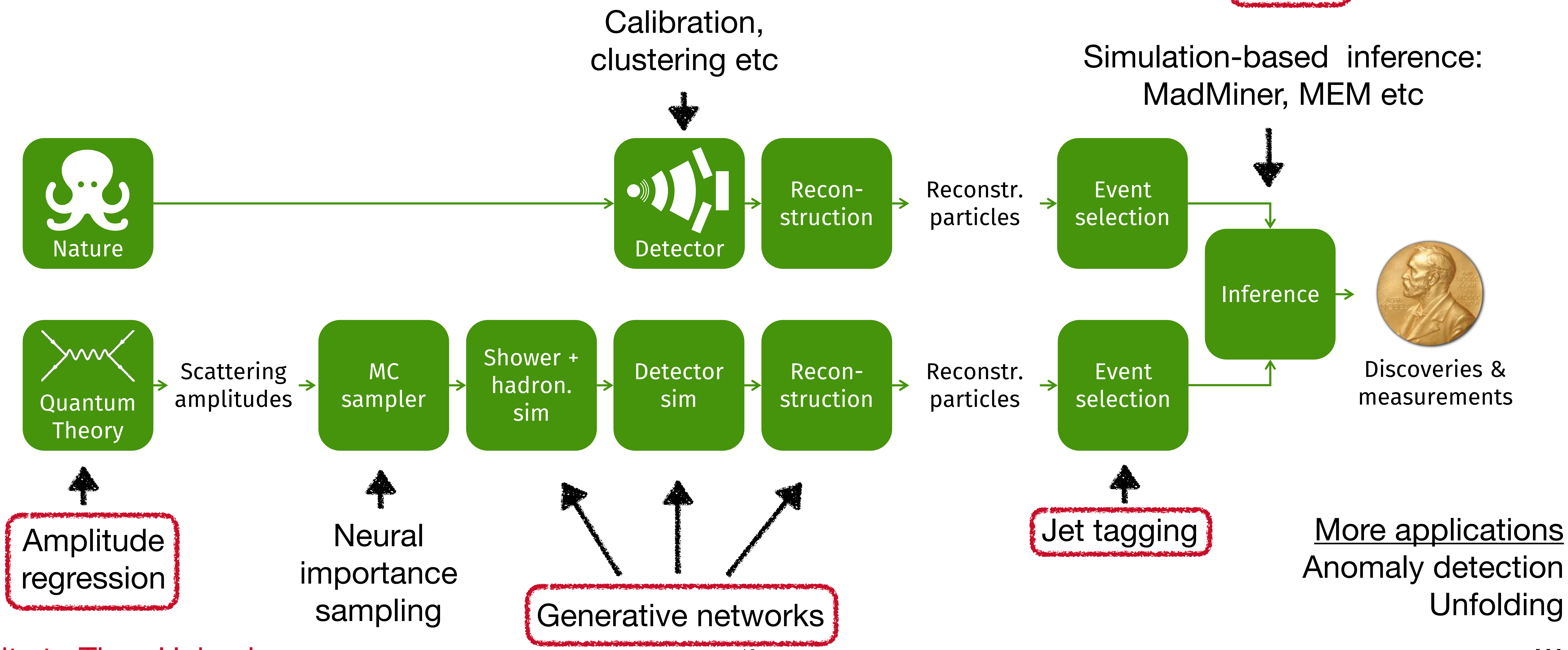




Experiments

LHC simulation chain meets ML

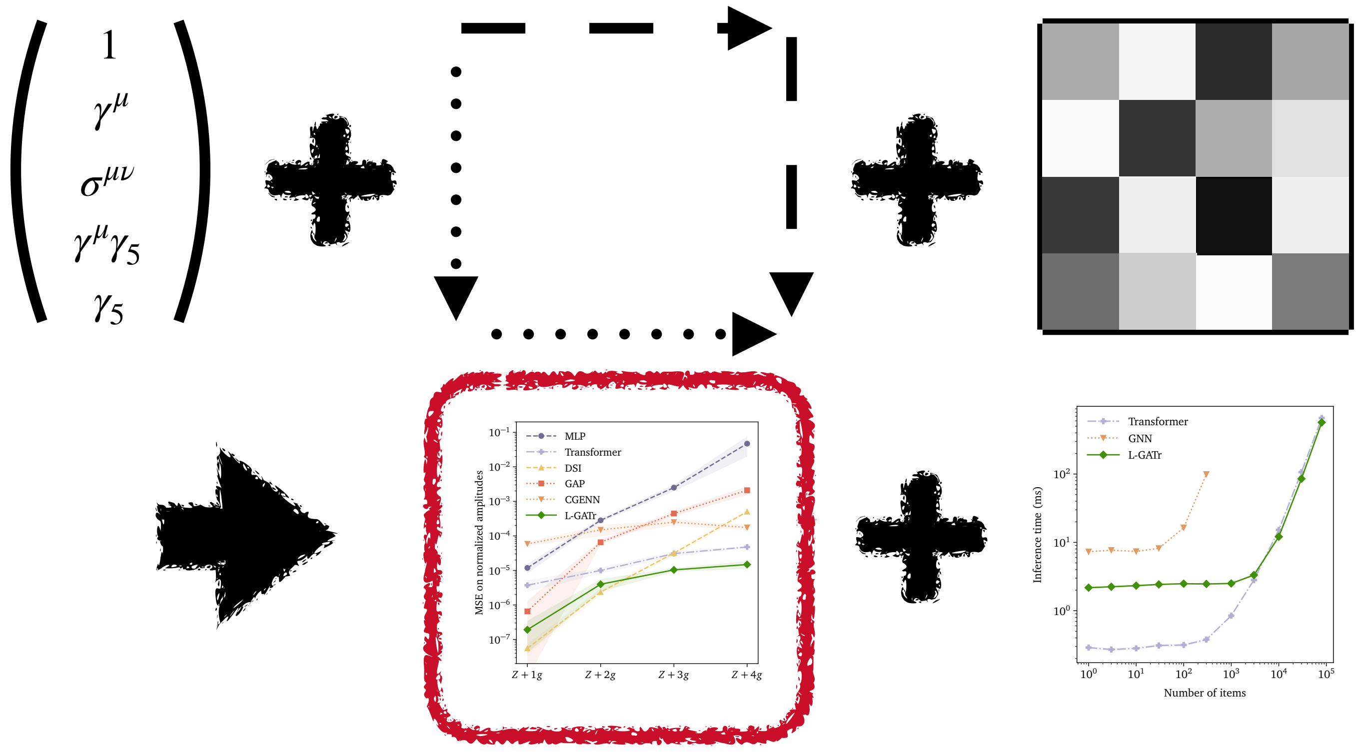


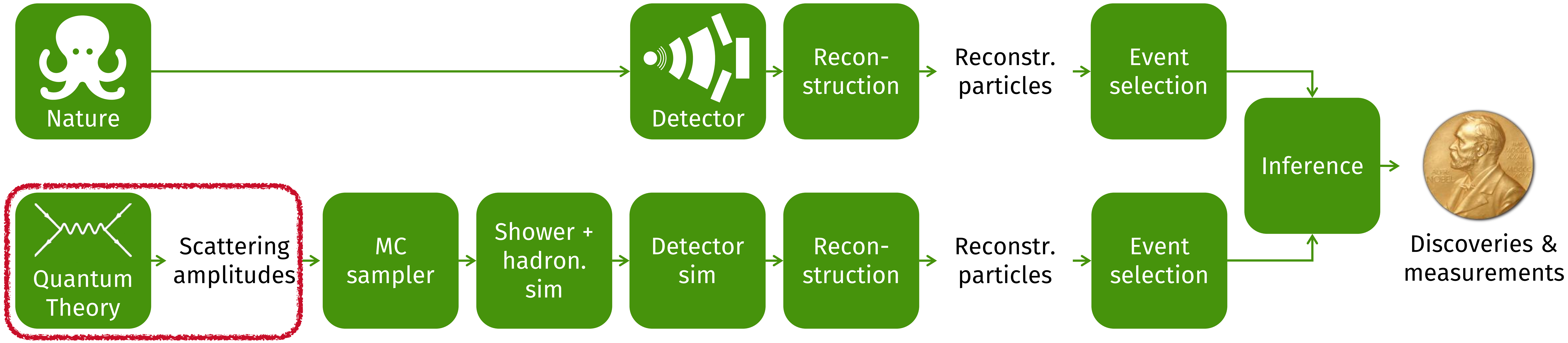


Credits to Theo Heimel

Experiments

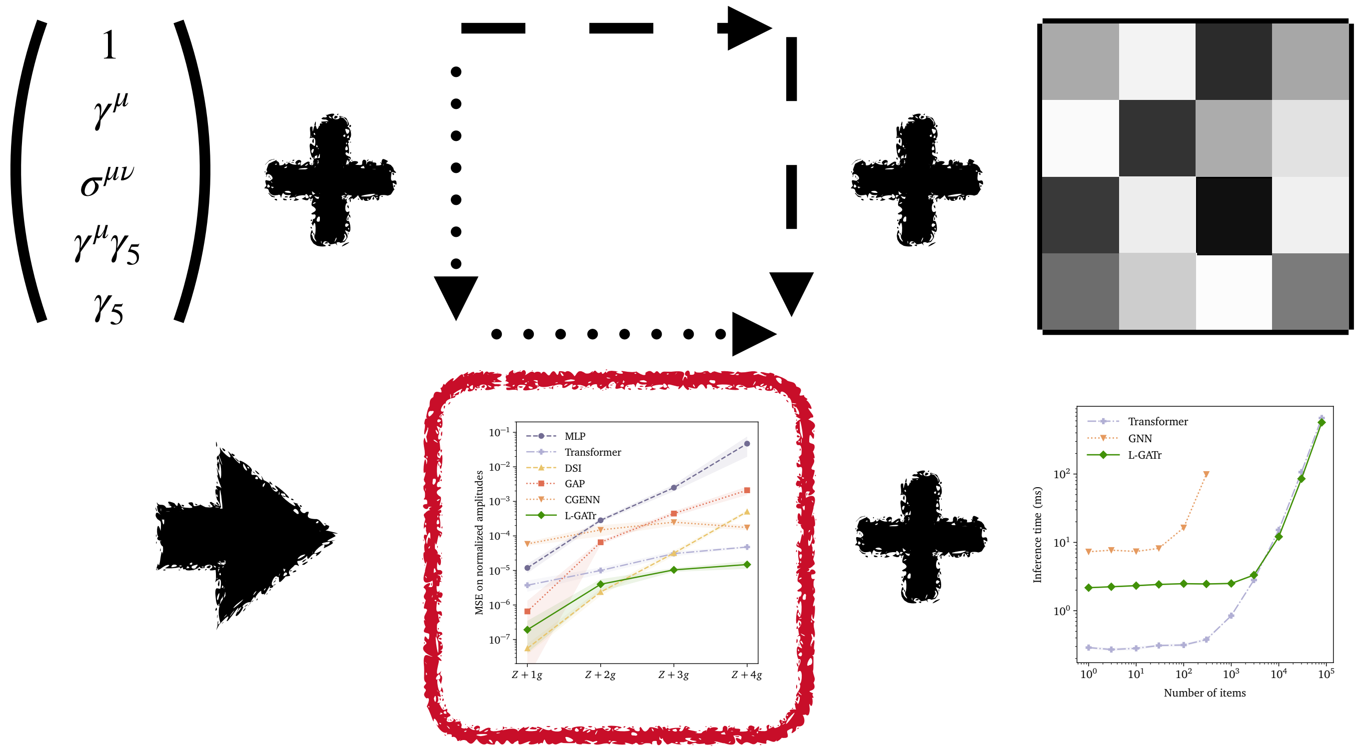
Amplitude regression

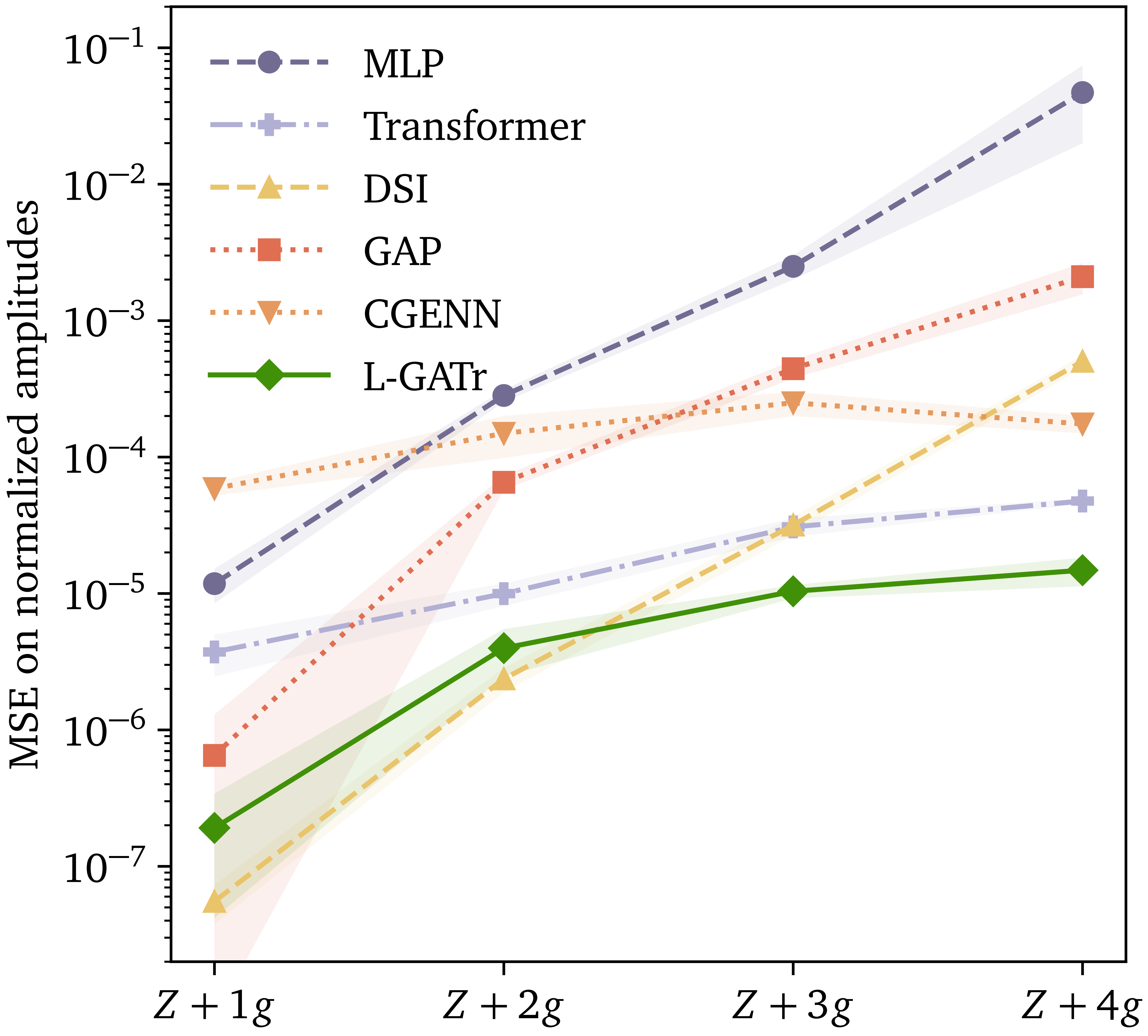




Experiments

Amplitude regression

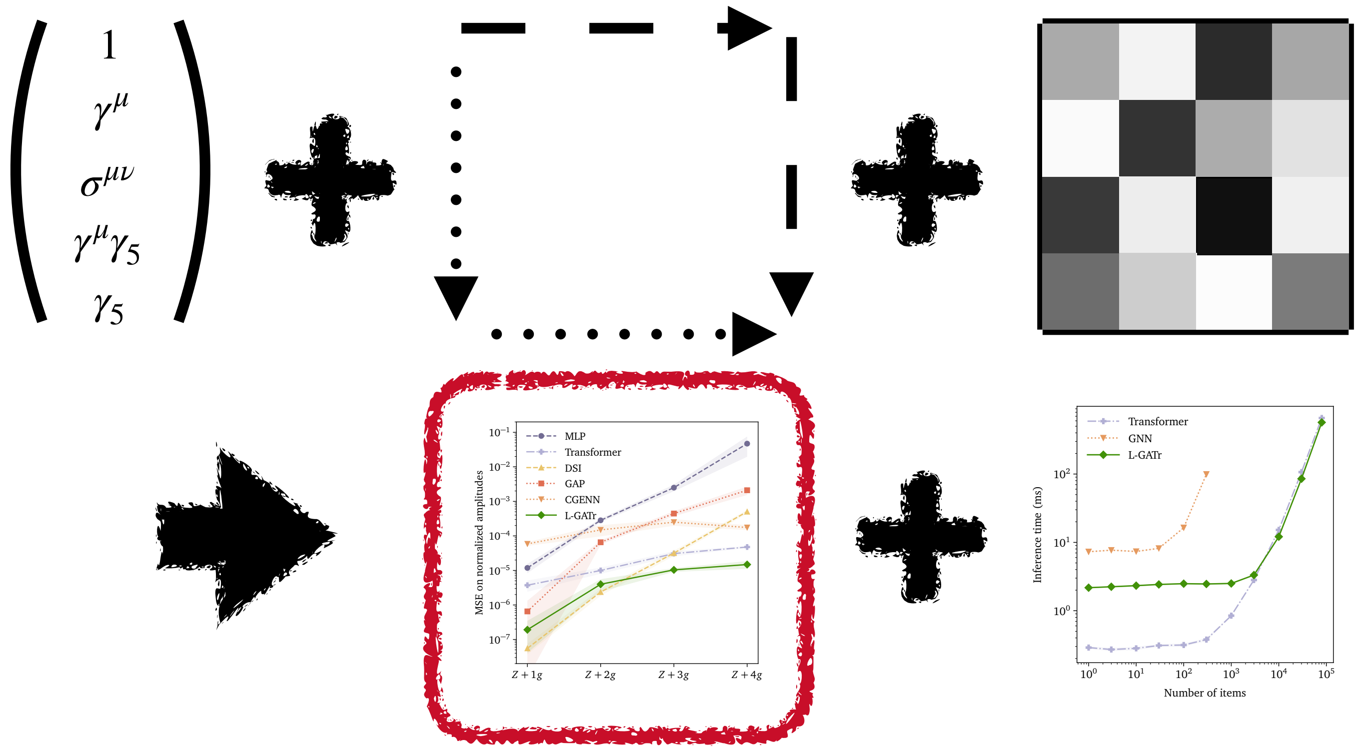


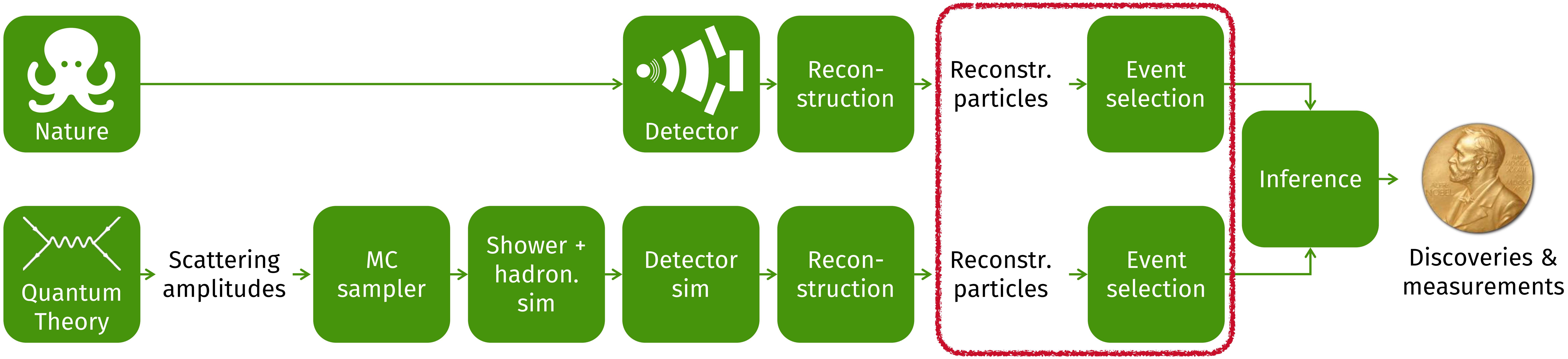


L-GATr scales best to high multiplicity,
where amplitude surrogates are most useful

Experiments

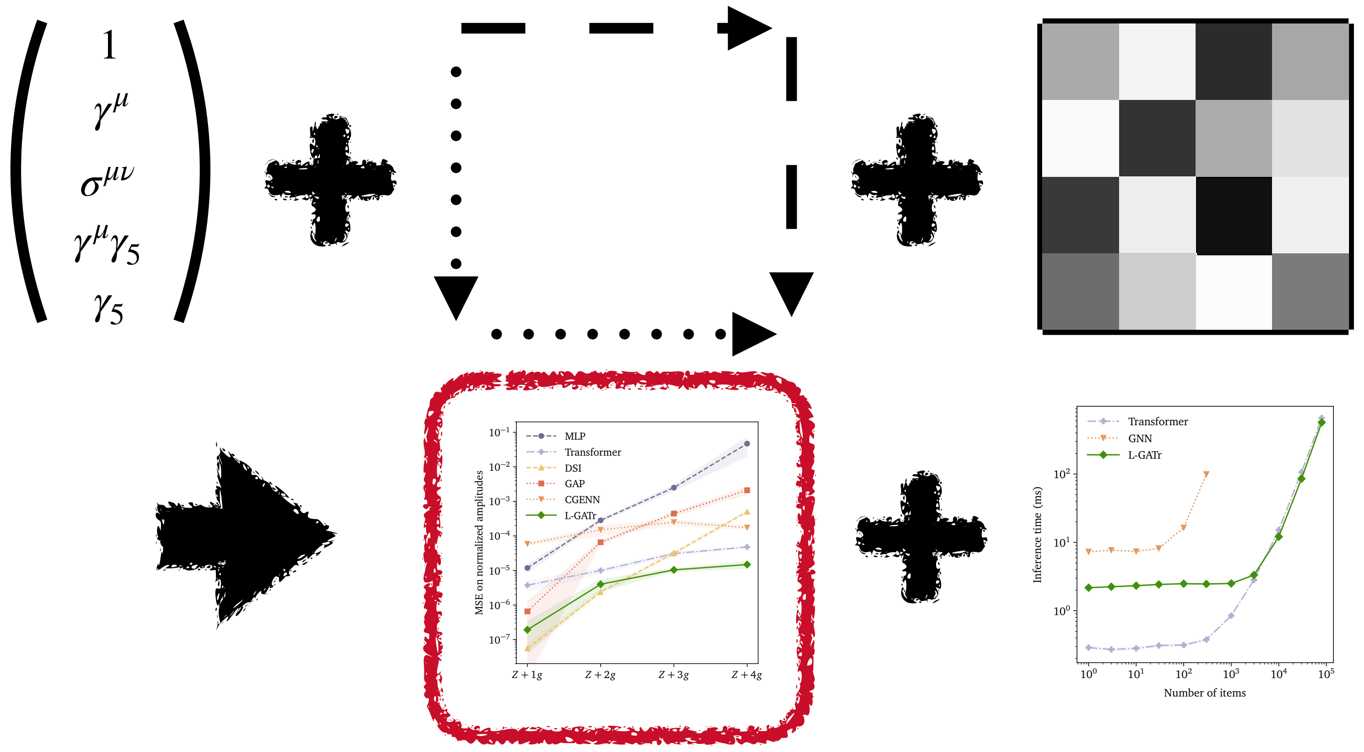
Top tagging





Experiments

Top tagging

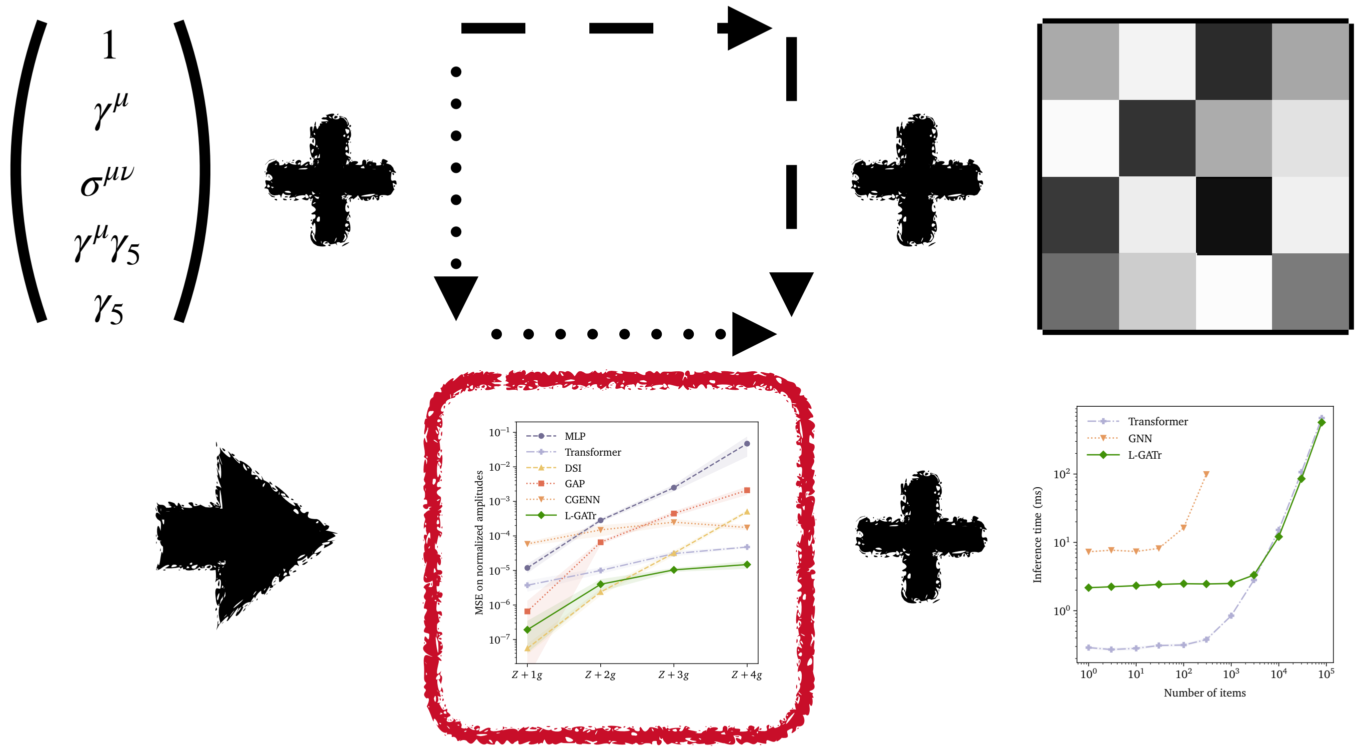


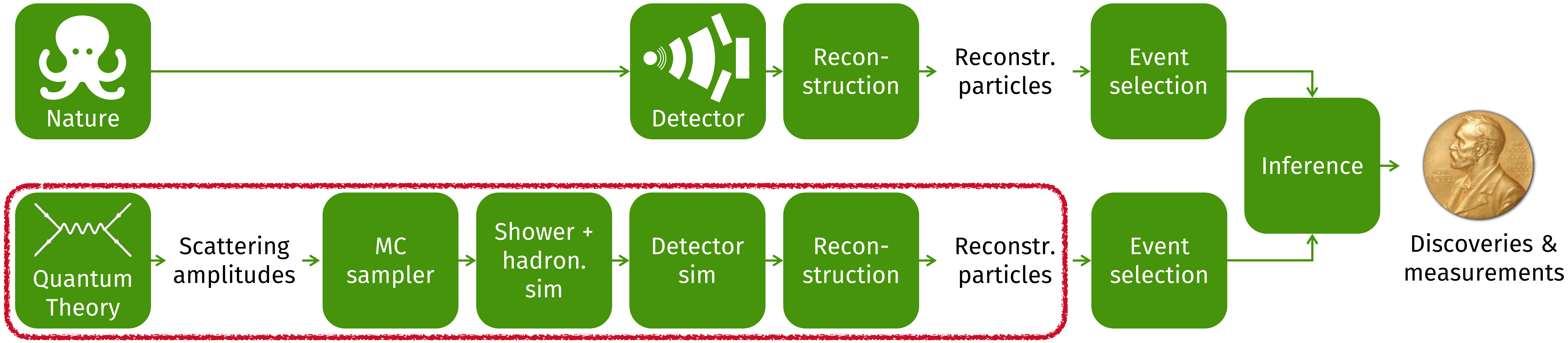
Model	Accuracy	AUC	$1/\epsilon_B$ ($\epsilon_S = 0.5$)	$1/\epsilon_B$ ($\epsilon_S = 0.3$)
TopoDNN [48]	0.916	0.972	–	295 ± 5
LoLa [15]	0.929	0.980	–	722 ± 17
P-CNN [1]	0.930	0.9803	201 ± 4	759 ± 24
N-subjettiness [61]	0.929	0.981	–	867 ± 15
PFN [50]	0.932	0.9819	247 ± 3	888 ± 17
TreeNiN [57]	0.933	0.982	–	1025 ± 11
ParticleNet [63]	0.940	0.9858	397 ± 7	1615 ± 93
ParT [64]	0.940	0.9858	413 ± 16	1602 ± 81
LorentzNet* [41]	0.942	0.9868	498 ± 18	2195 ± 173
CGENN* [67]	0.942	0.9869	500	2172
PELICAN* [9]	**0.9426** ± 0.0002	**0.9870** ± 0.0001	–	**2250** ± 75
L-GATr (ours)*	0.9417 ± 0.0002	0.9868 ± 0.0001	**548** ± 26	2148 ± 106

L-GATr is on par with the best equivariant (*) baselines

Experiments

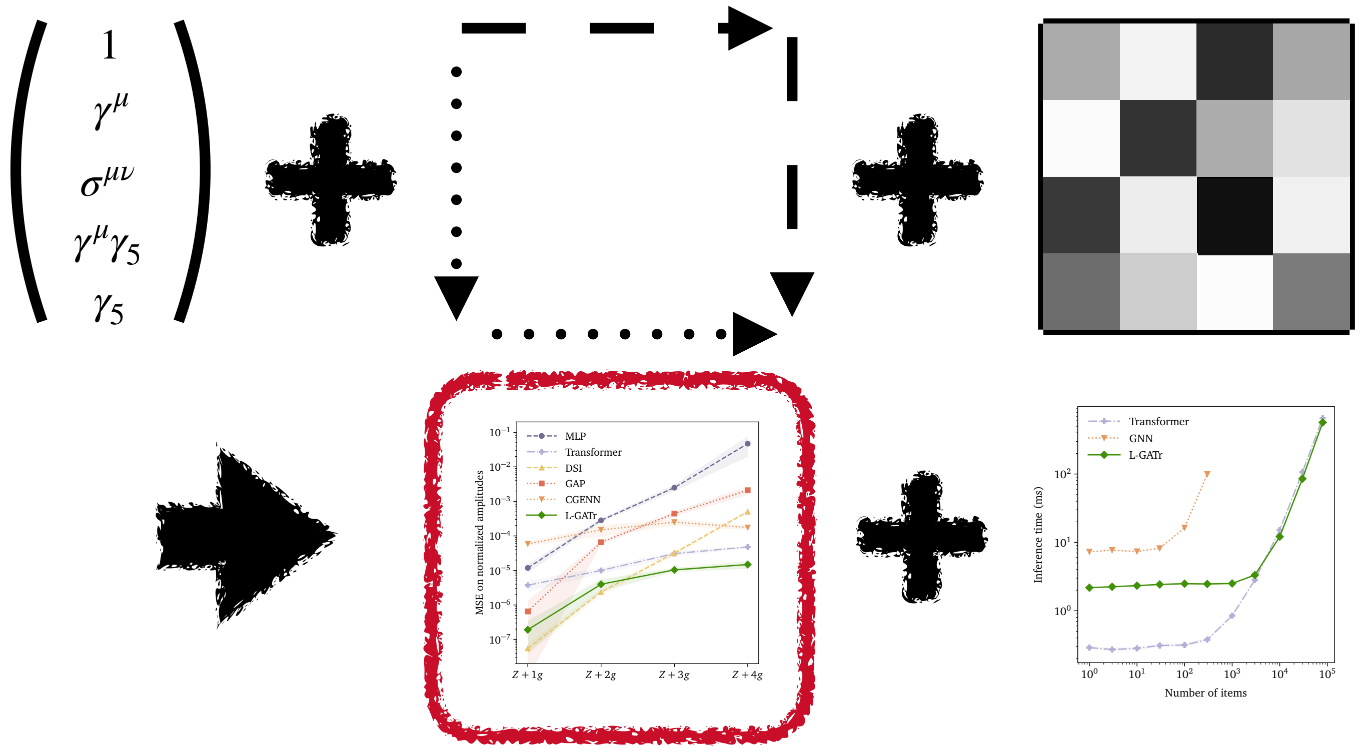
Event generation





Experiments

Event generation



Continuous normalising flows (CNF)
connect a simple base density
to a complex target density
through a neural differential equation

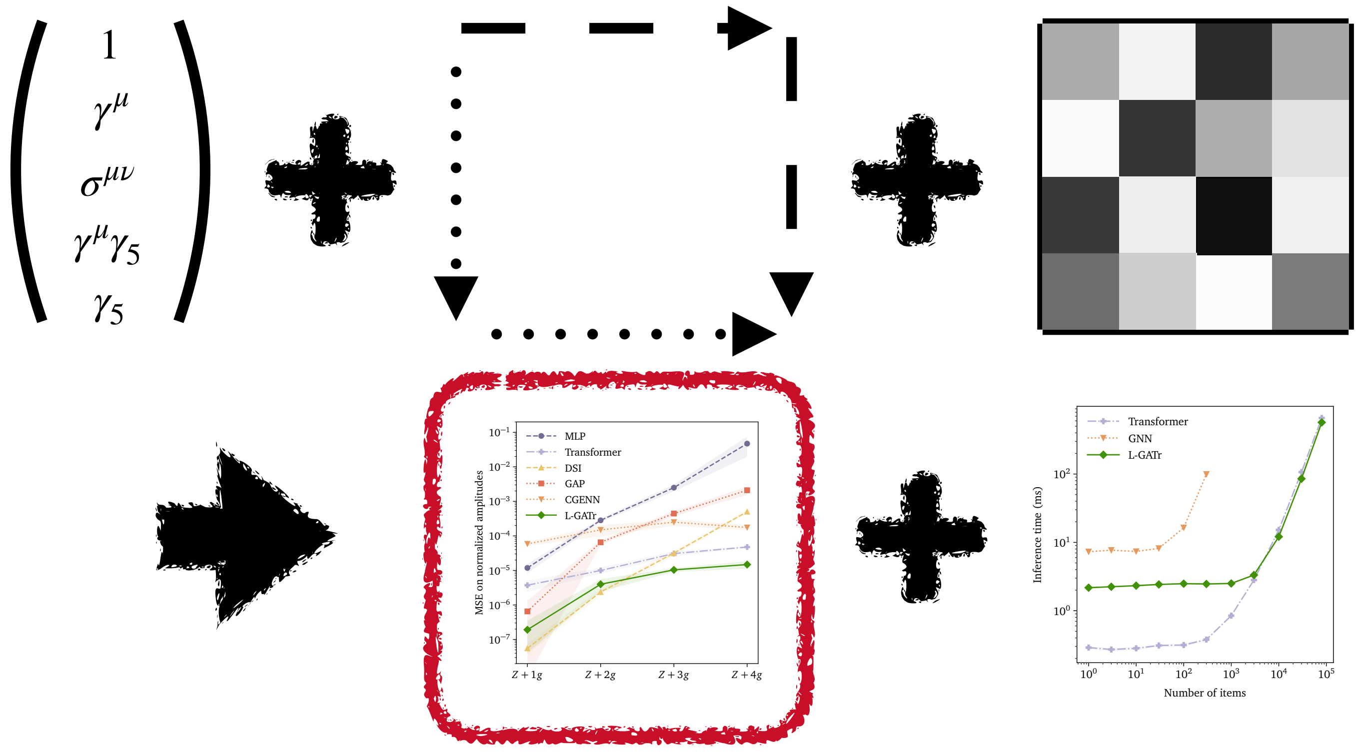
$$\frac{d}{dt}x = v_t(x)$$

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

$$\mathscr{L} = \mathbb{E}_{t,x,x_1} \| v_t(x) - u_t(x\,|\,x_1) \|^2$$

Continuous normalising flows
arXiv:1806.07366

Conditional flow matching
arXiv:2210.02747



Experiments

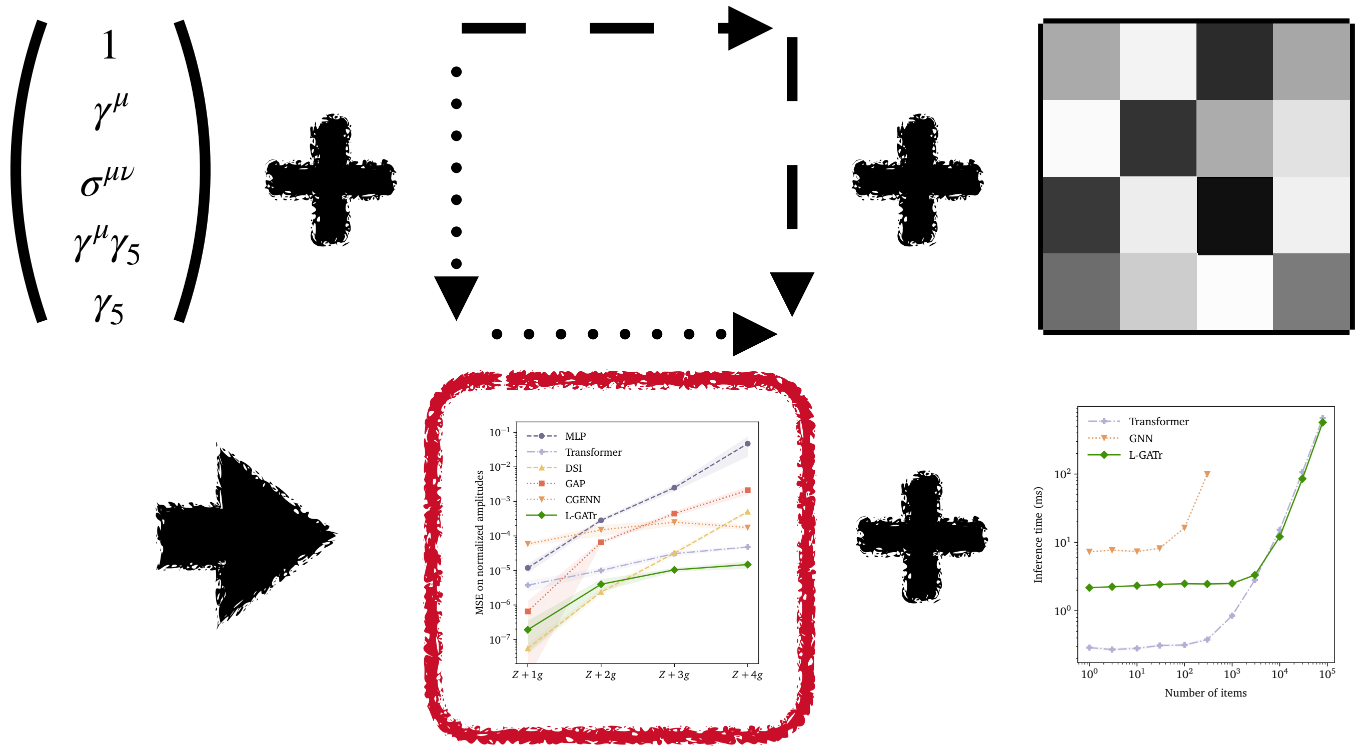
Event generation

In conditional flow matching (CFM), the choice of target velocity can be more important than the architecture

Riemannian Flow Matching
arXiv:2302.03660

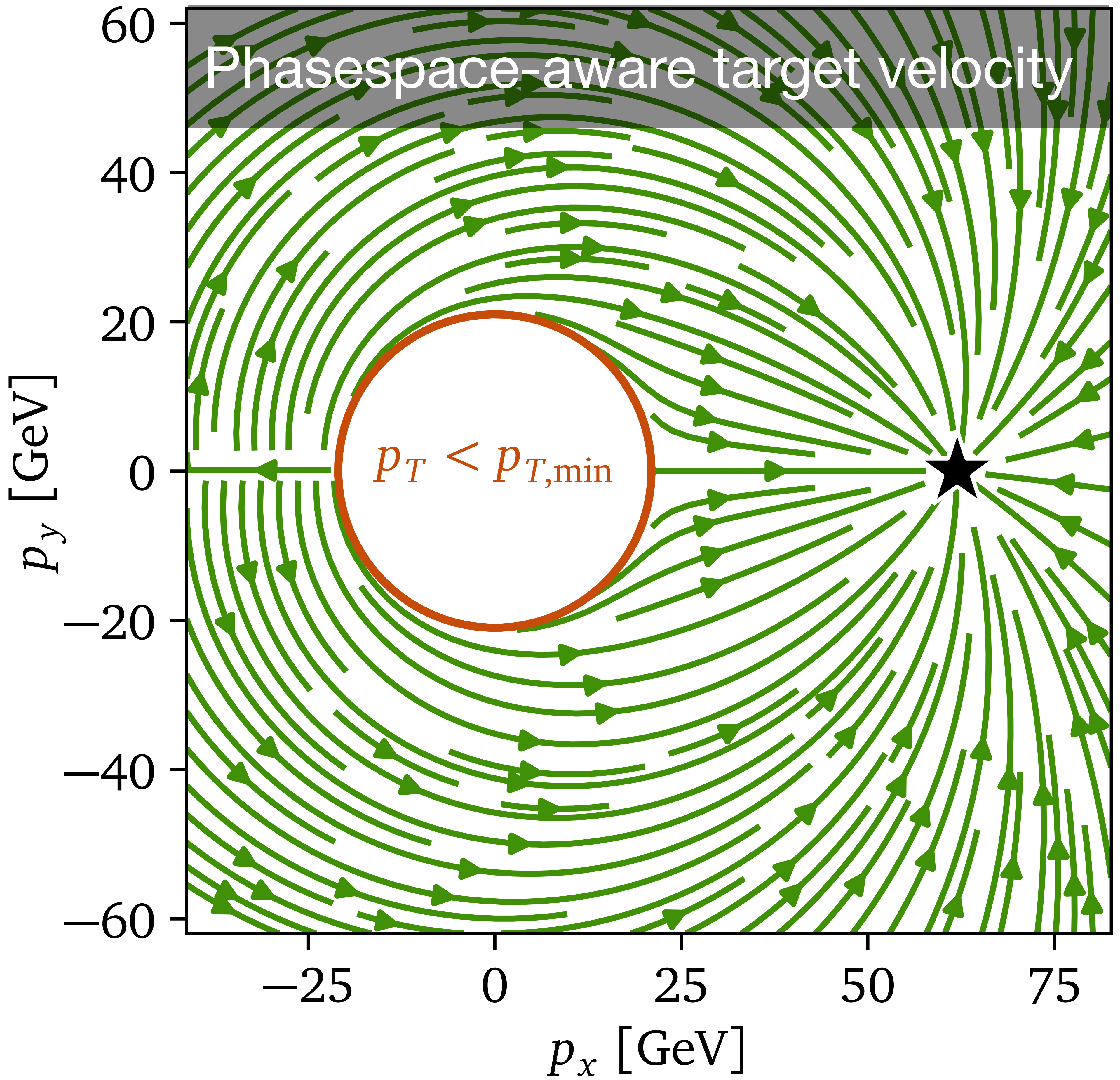
Experiments

Event generation



In conditional flow matching (CFM), the choice of target velocity can be more important than the architecture

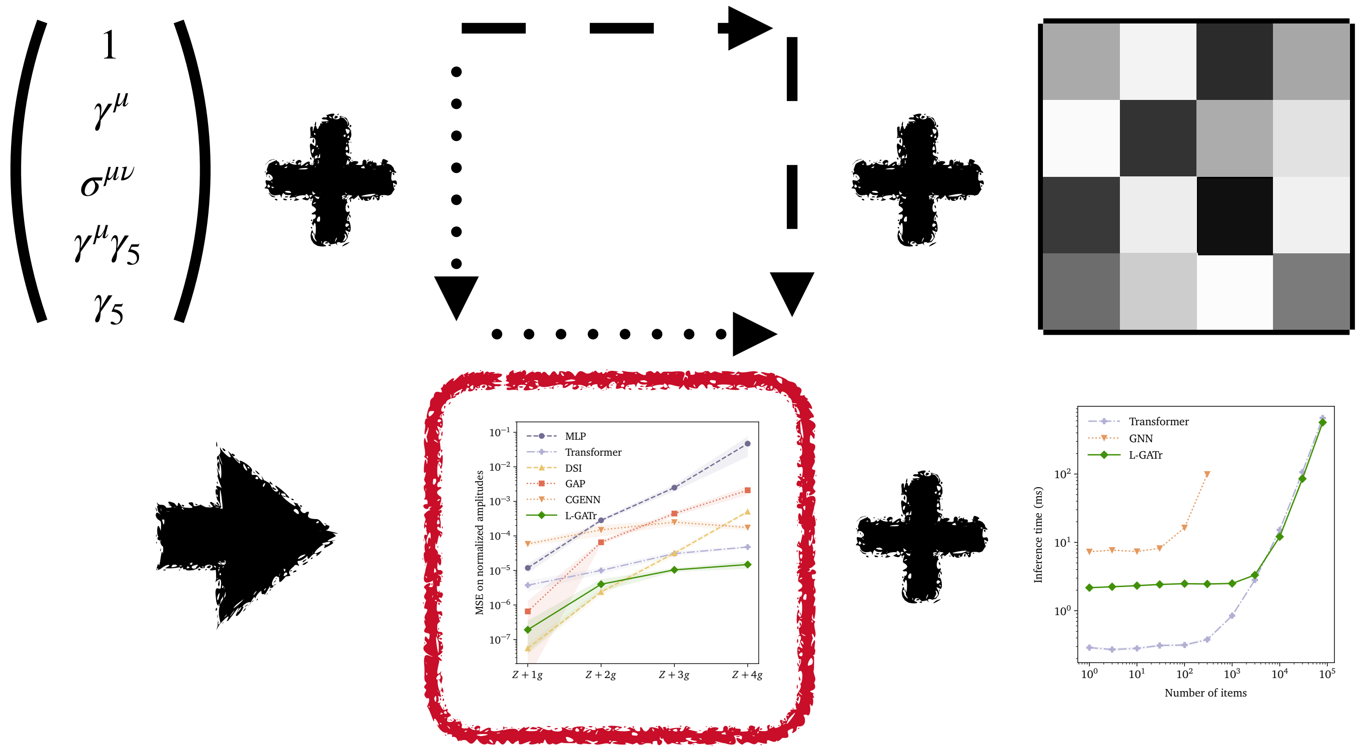
Target velocity	Architecture	AUC
Euclidean	L-GATr	0.99
Phasespace-aware	MLP	0.78
Phasespace-aware	L-GATr	**0.51**

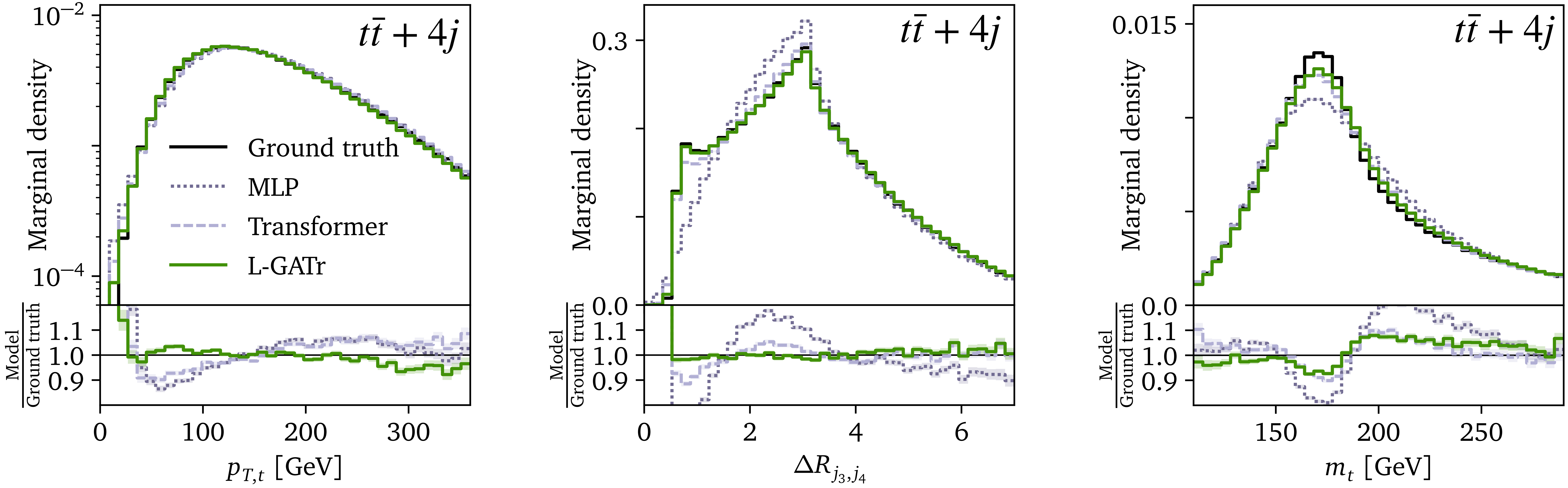


Riemannian Flow Matching
arXiv:2302.03660

Experiments

Event generation

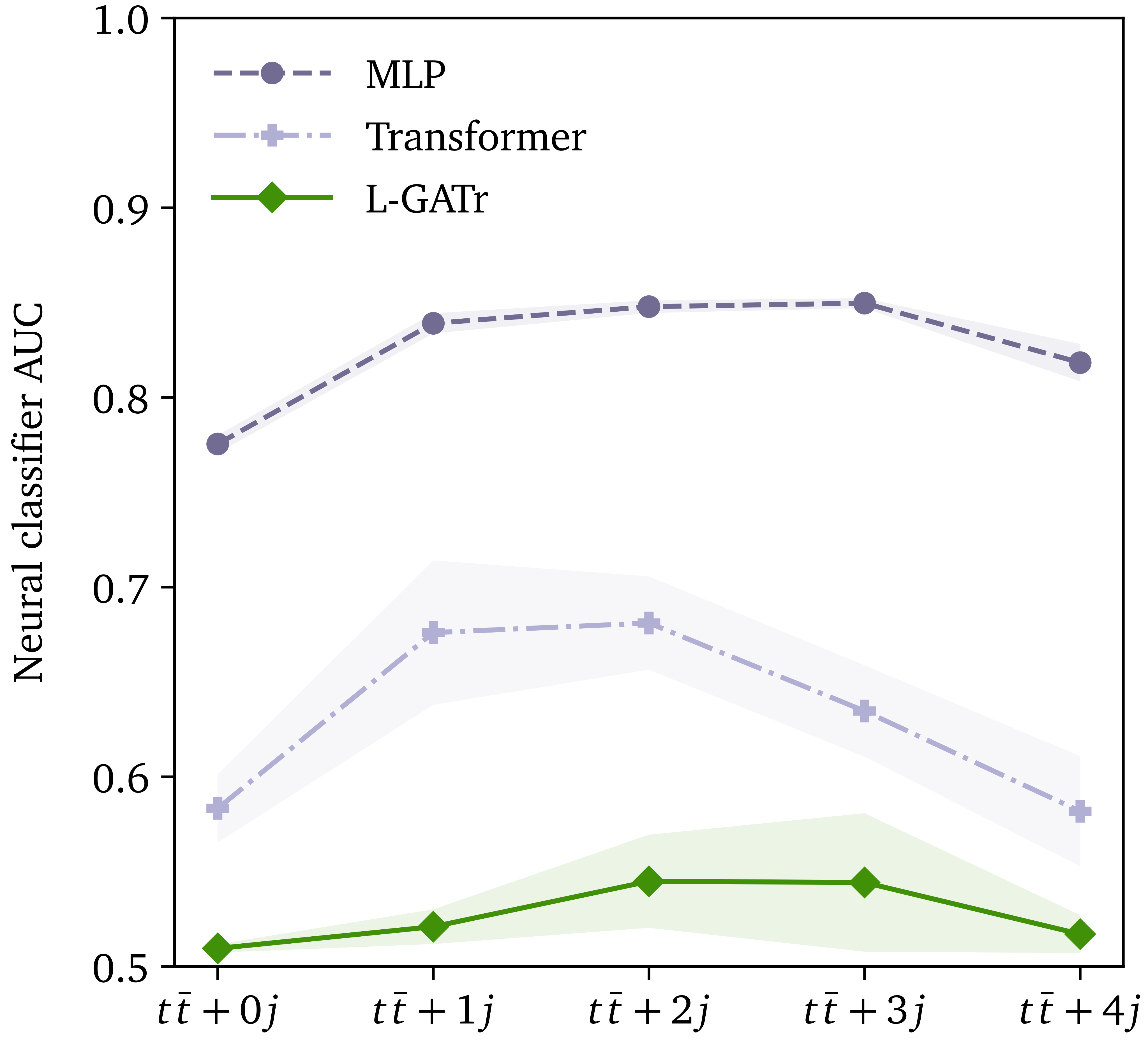




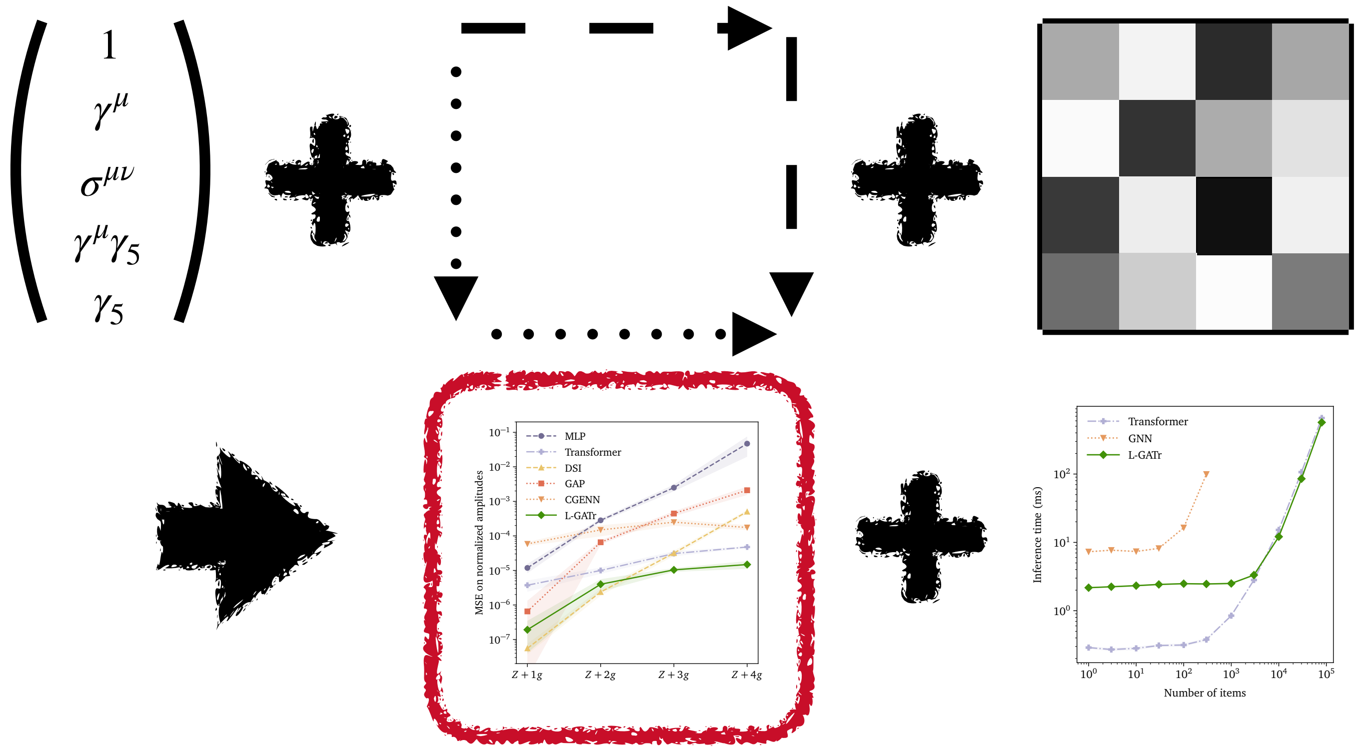
L-GATr helps with tricky kinematic features

Experiments

Event generation

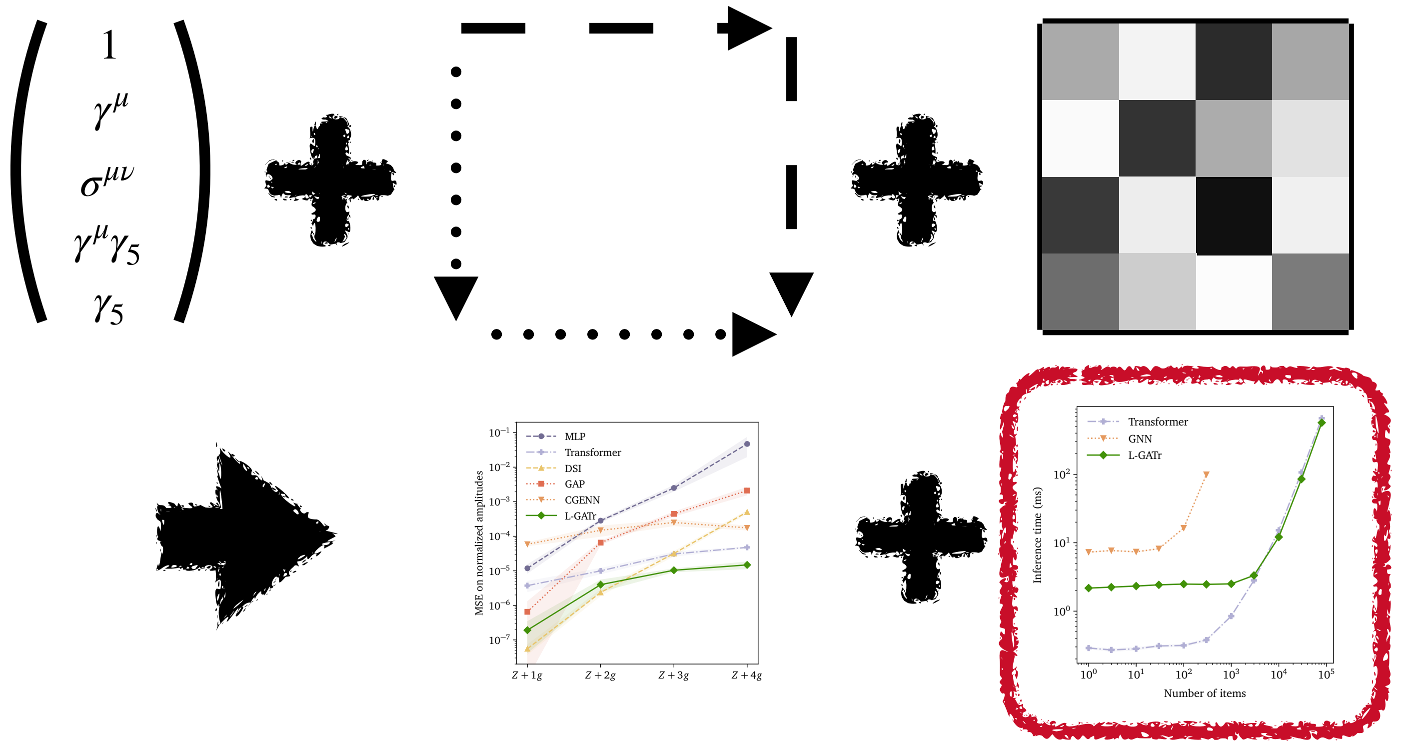


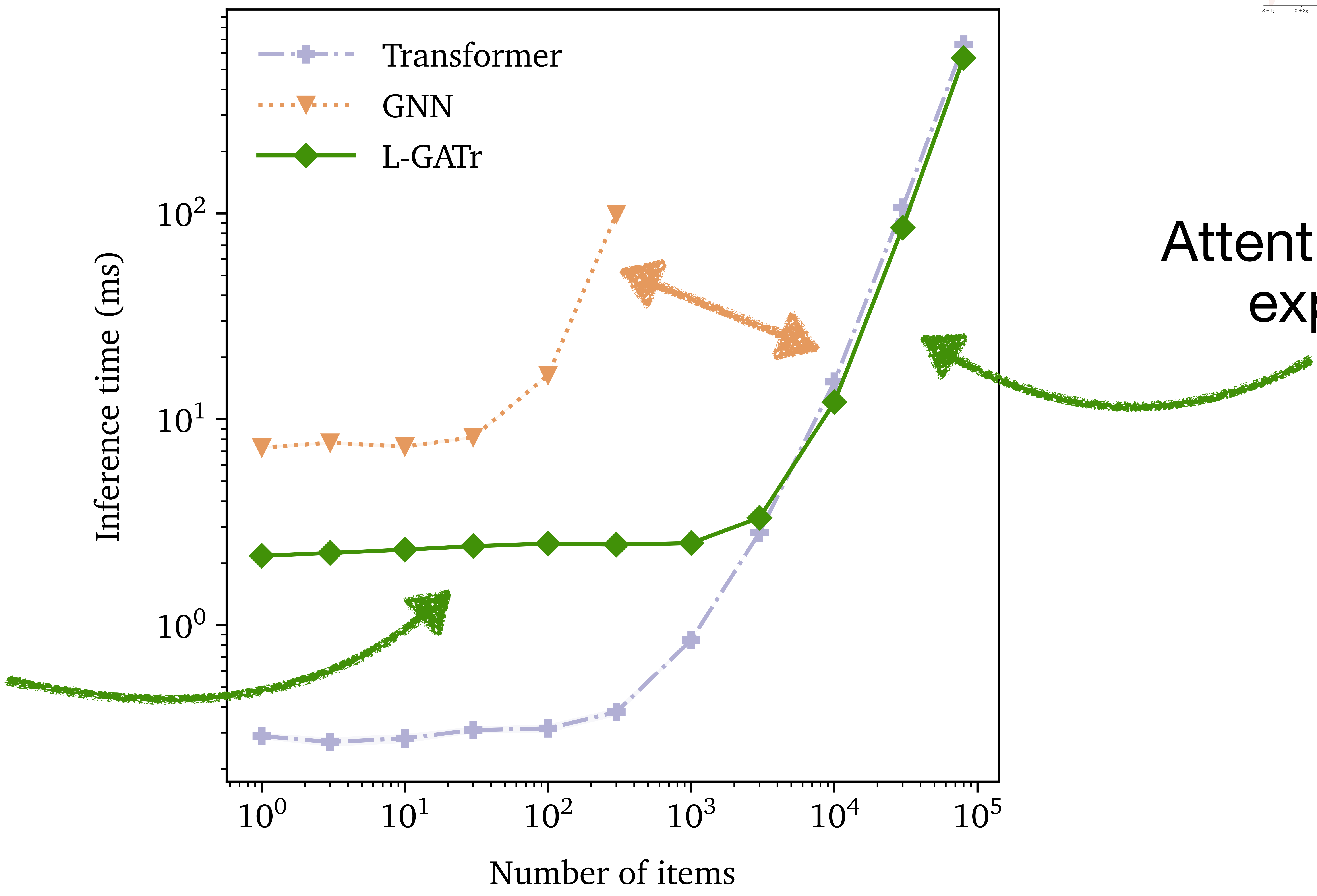
L-GATr generates samples that a classifier can almost not distinguish from the ground truth



Experiments

Scaling L-GATr to thousands of tokens

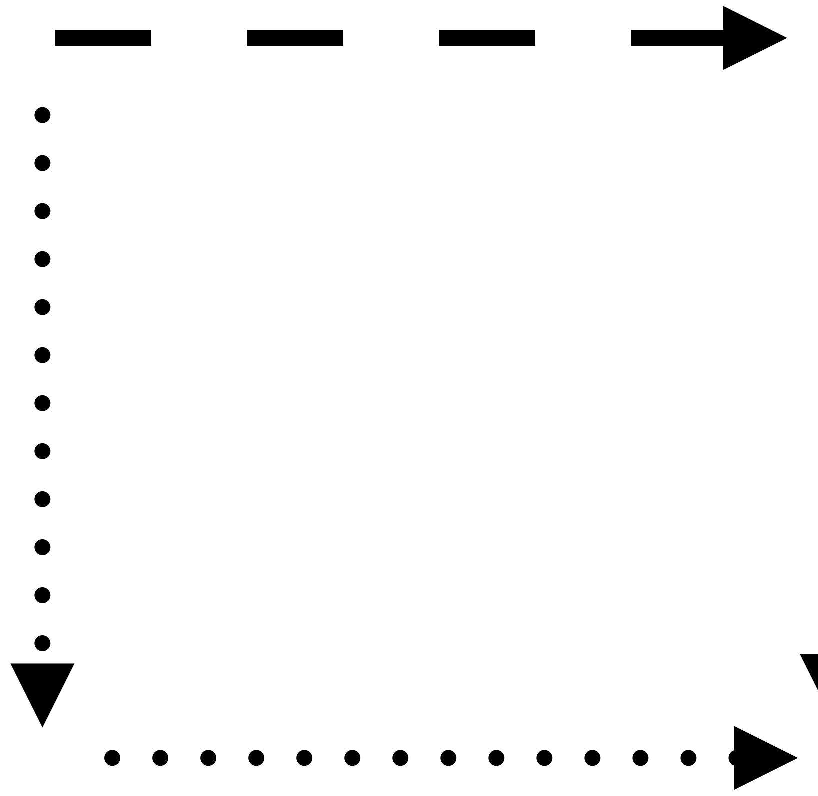


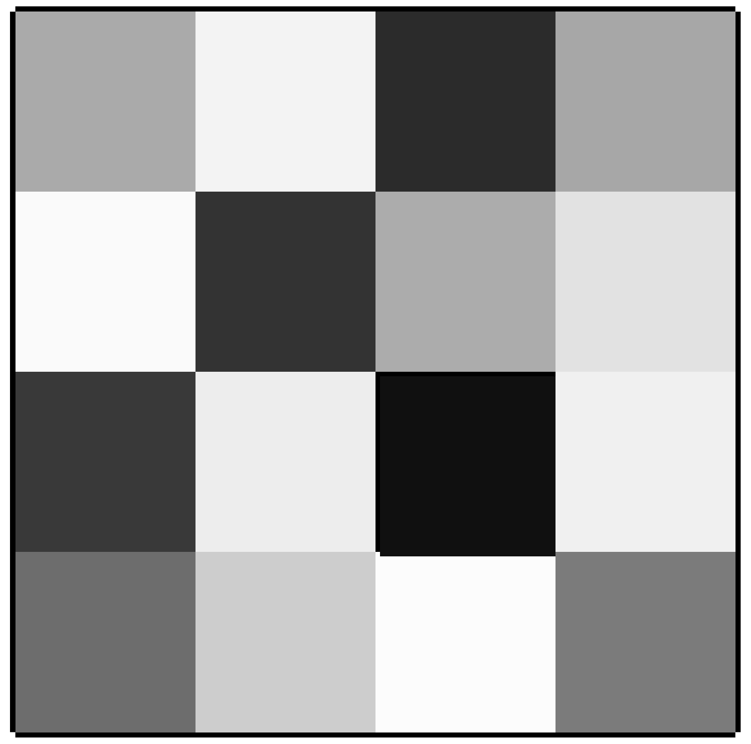


Attention is most expensive

Linear layers are most expensive

$$\begin{pmatrix} 1 \\ \gamma^\mu \\ \sigma^{\mu\nu} \\ \gamma^\mu\gamma_5 \\ \gamma_5 \end{pmatrix}$$

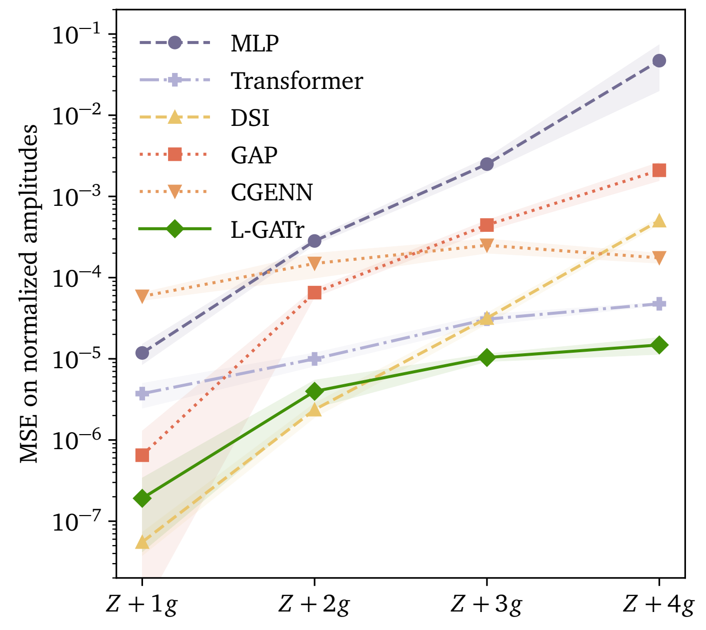


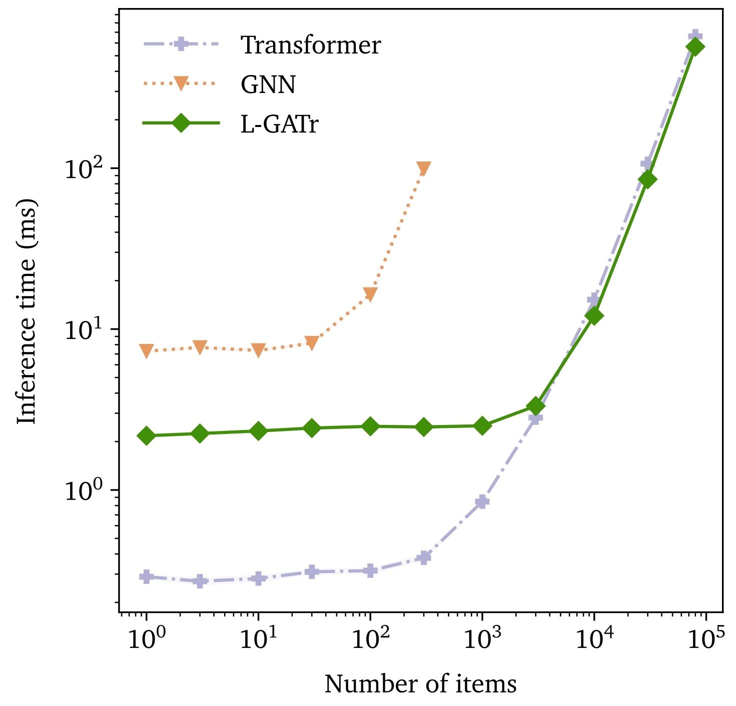


Geometric algebra representations

Equivariant layers

Transformer architecture





Strong performance on diverse problems

Scalable to thousands of tokens

L-GATr combines **equivariance** and **scalability**

Victor Bresó

Pim de Haan

Tilman Plehn

Jesse Thaler

Johann Brehmer

Geometric Algebra Transformer

E(3)-equivariant version
Johann Brehmer*, Pim de Haan*, Sönke Behrends, Taco Cohen
NeurIPS 2023, arXiv:2305.18415

E(3)-GATr paper

E(3)-GATr code

Lorentz-Equivariant Geometric Algebra Transformer for High-Energy Physics

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

L-GATr paper

L-GATr code

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

Bonus material

Ingredients

Equivariant layers

EquiLinear

$$\phi(x) = \sum_{k=0}^{4} v_k \langle x \rangle_k + \sum_{k=0}^{4} w_k \gamma_5 \langle x \rangle_k$$

Geometric product

$$\psi(x, y) = x \cdot y$$

Geometric attention

$$\text{Attention}(q, k, v)_{i\alpha} = \text{Softmax}_j \left(\frac{\langle q_{i\beta}, k_{j\beta} \rangle}{\sqrt{16n}} \right) v_{j\alpha}$$

EquiLayerNorm

$$\text{LN}(x) = x \Big/ \sqrt{\frac{1}{n} \sum_{c=1}^{n} \sum_{k=0}^{4} \left| \Big\langle \langle x_c \rangle_k, \langle x_c \rangle_k \Big\rangle \right| + \epsilon}$$

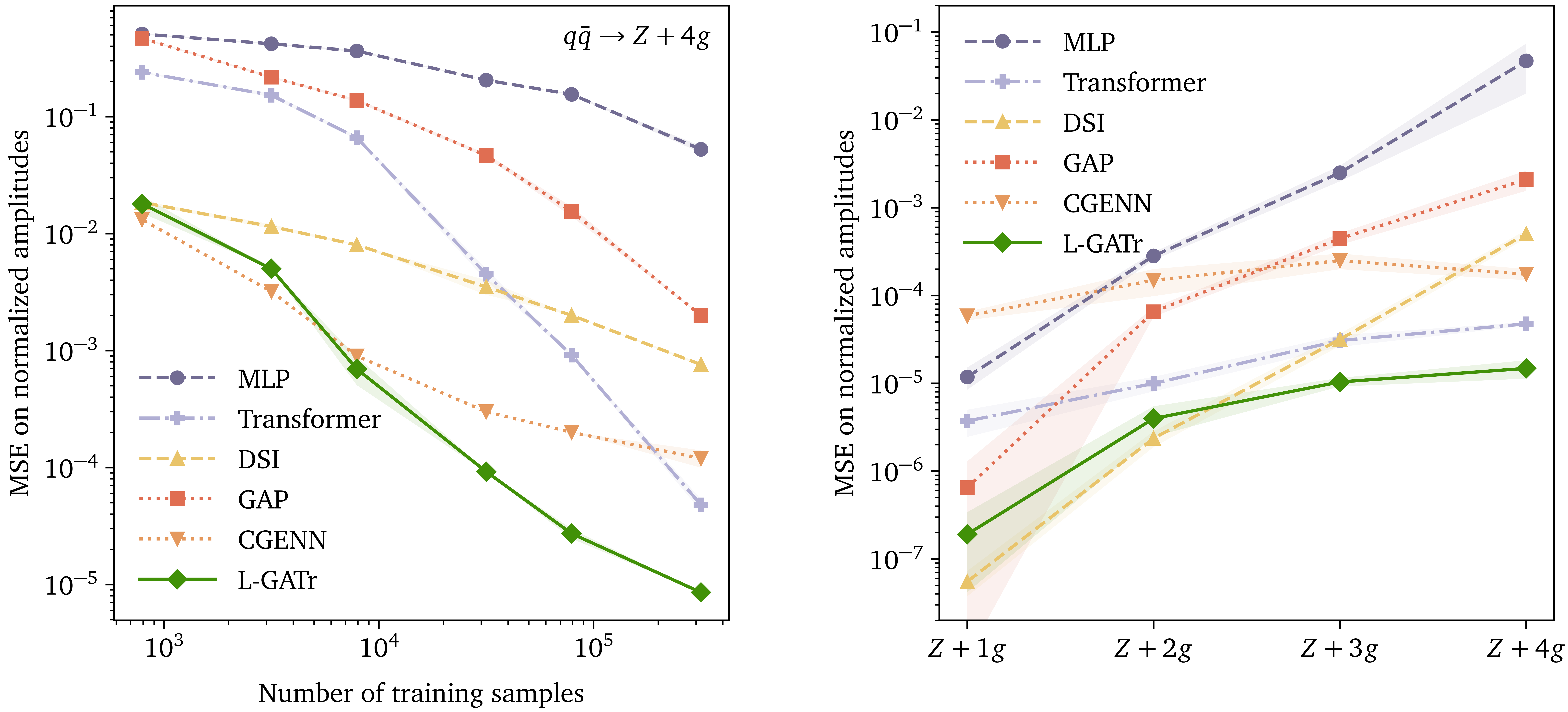
Activation function

$$a(x) = \text{GELU}(\langle x \rangle_0)\, x$$

Dropout

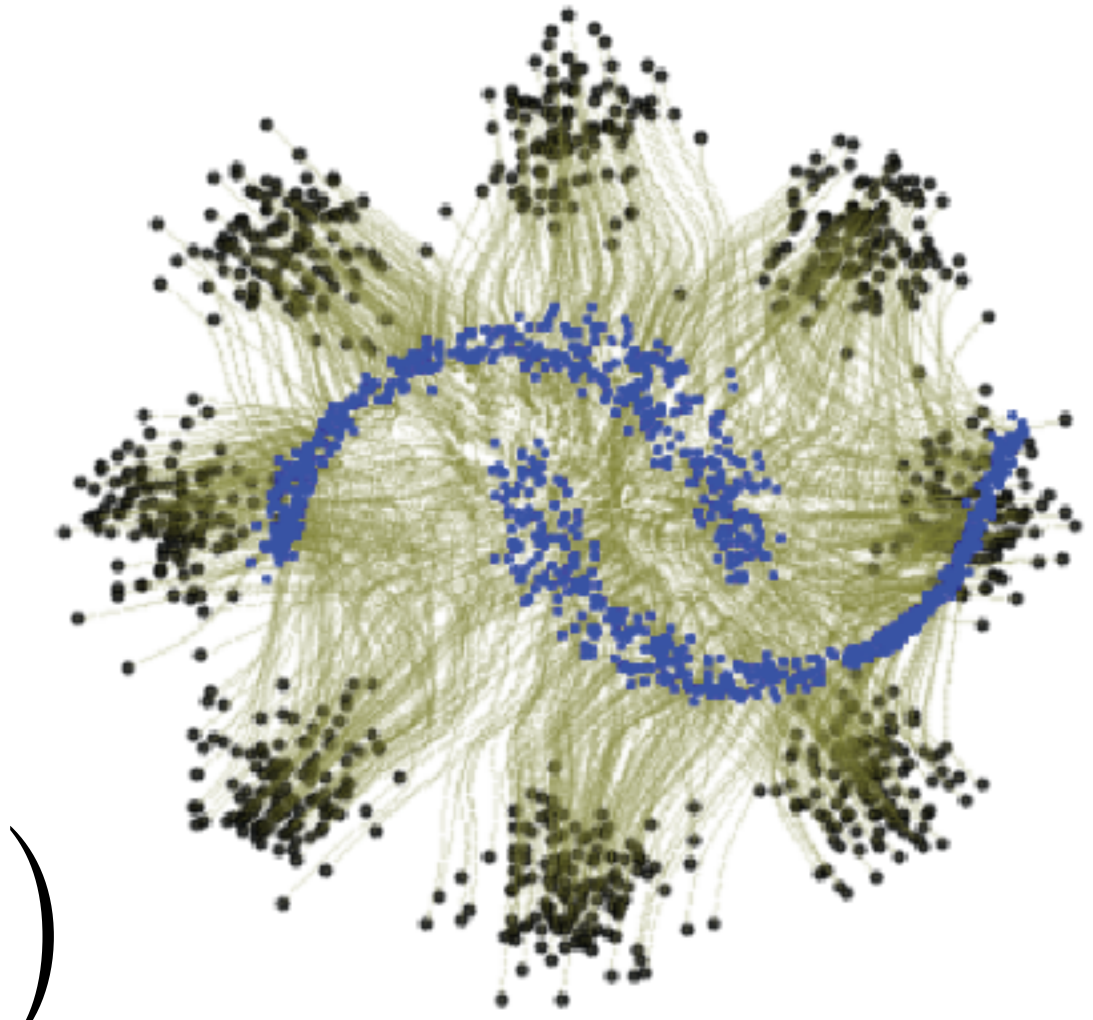
Separate dropout for each multivector blade

Amplitude regression



Event generation

Target velocities for CFM



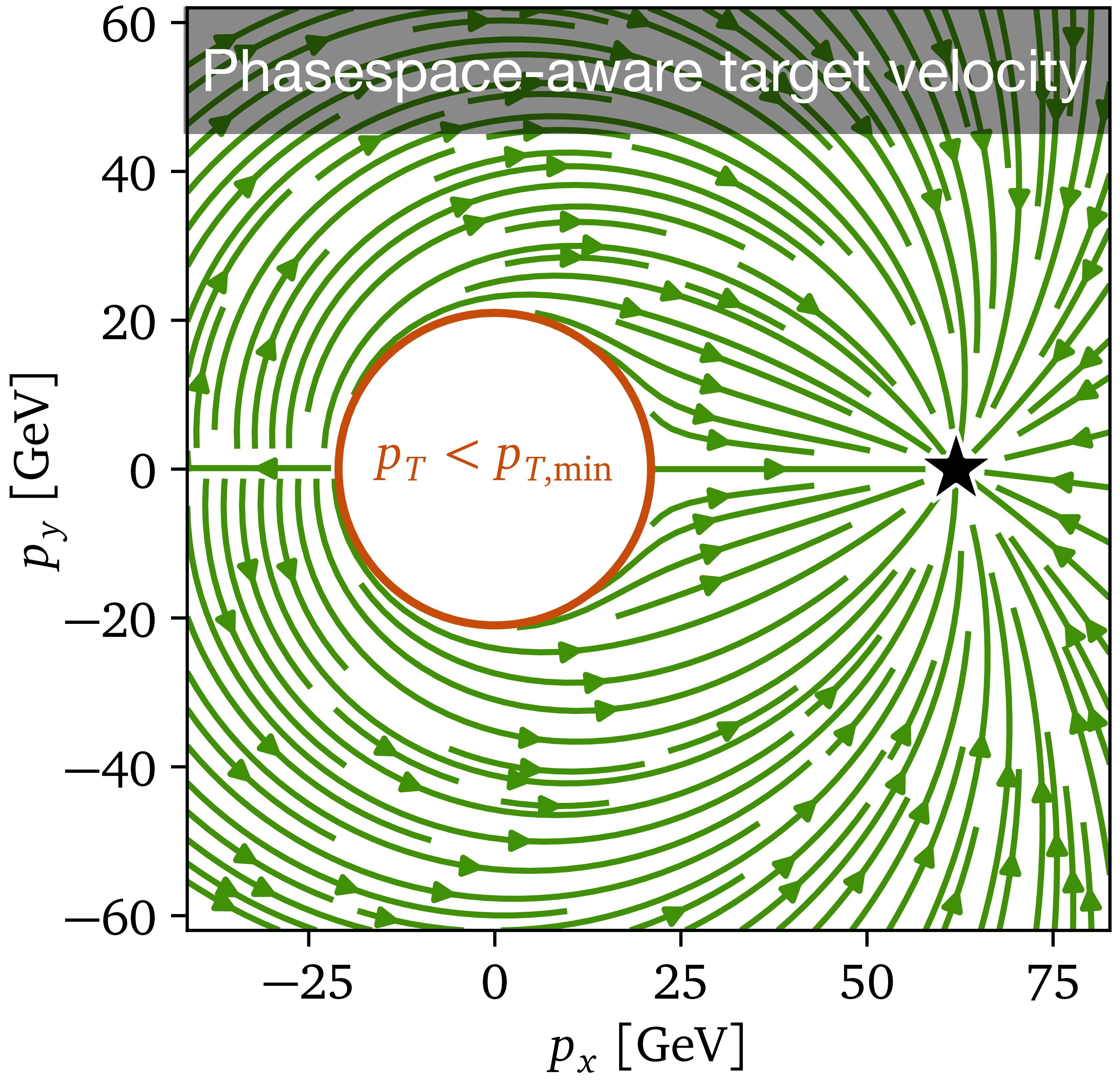
$$p = (E, p_x, p_y, p_z) = f(y) = \left(\sqrt{m^2 + p_T^2 \cosh^2 \eta}, p_T \cos\phi, p_T \sin\phi, p_T \sinh\eta\right)$$

$$y = (y_m, y_p, \phi, \eta), \qquad m^2 = \exp(y_m), \quad p_T = p_{T,\min} + \exp(y_p)$$

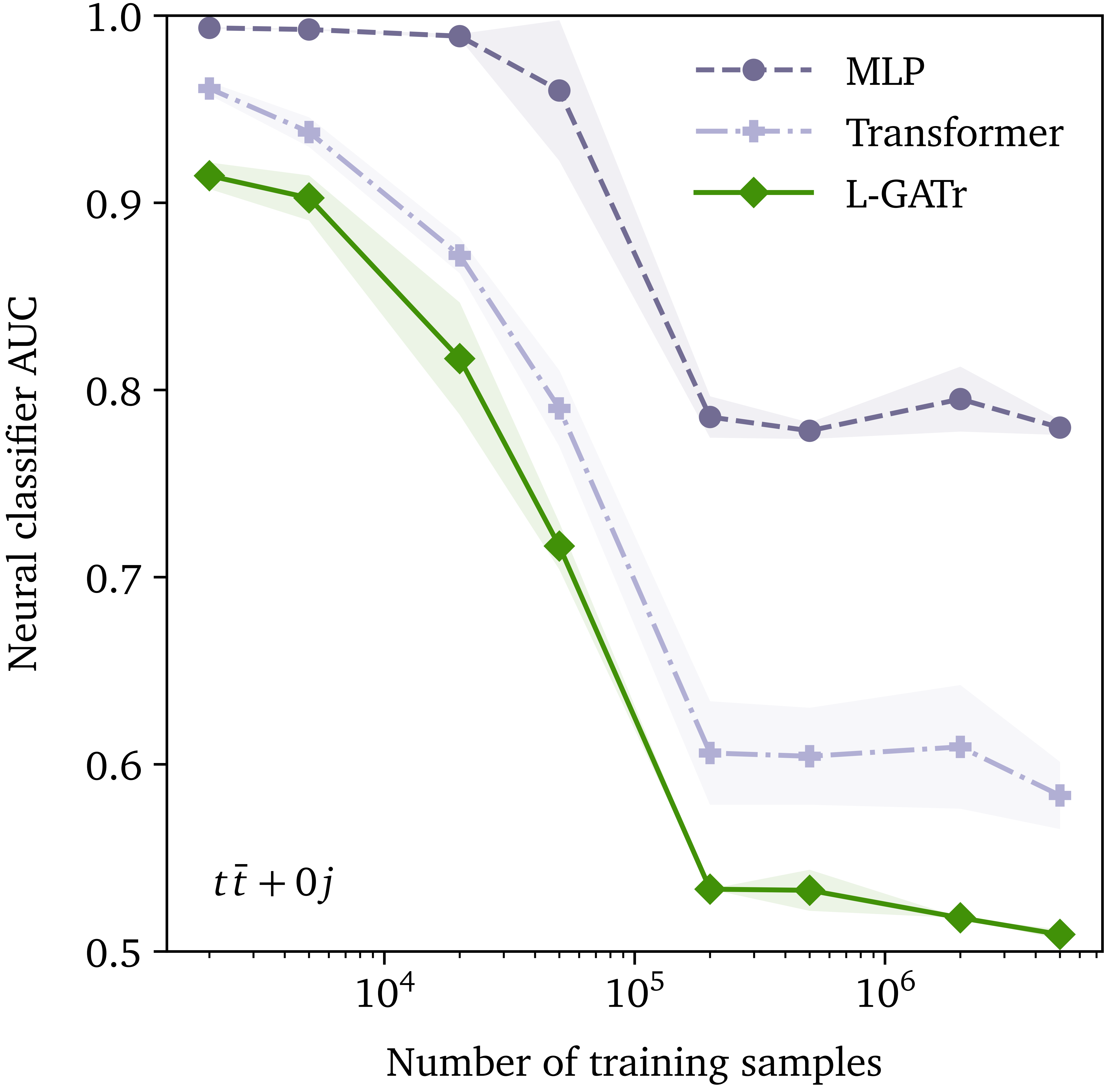
Target velocities can be

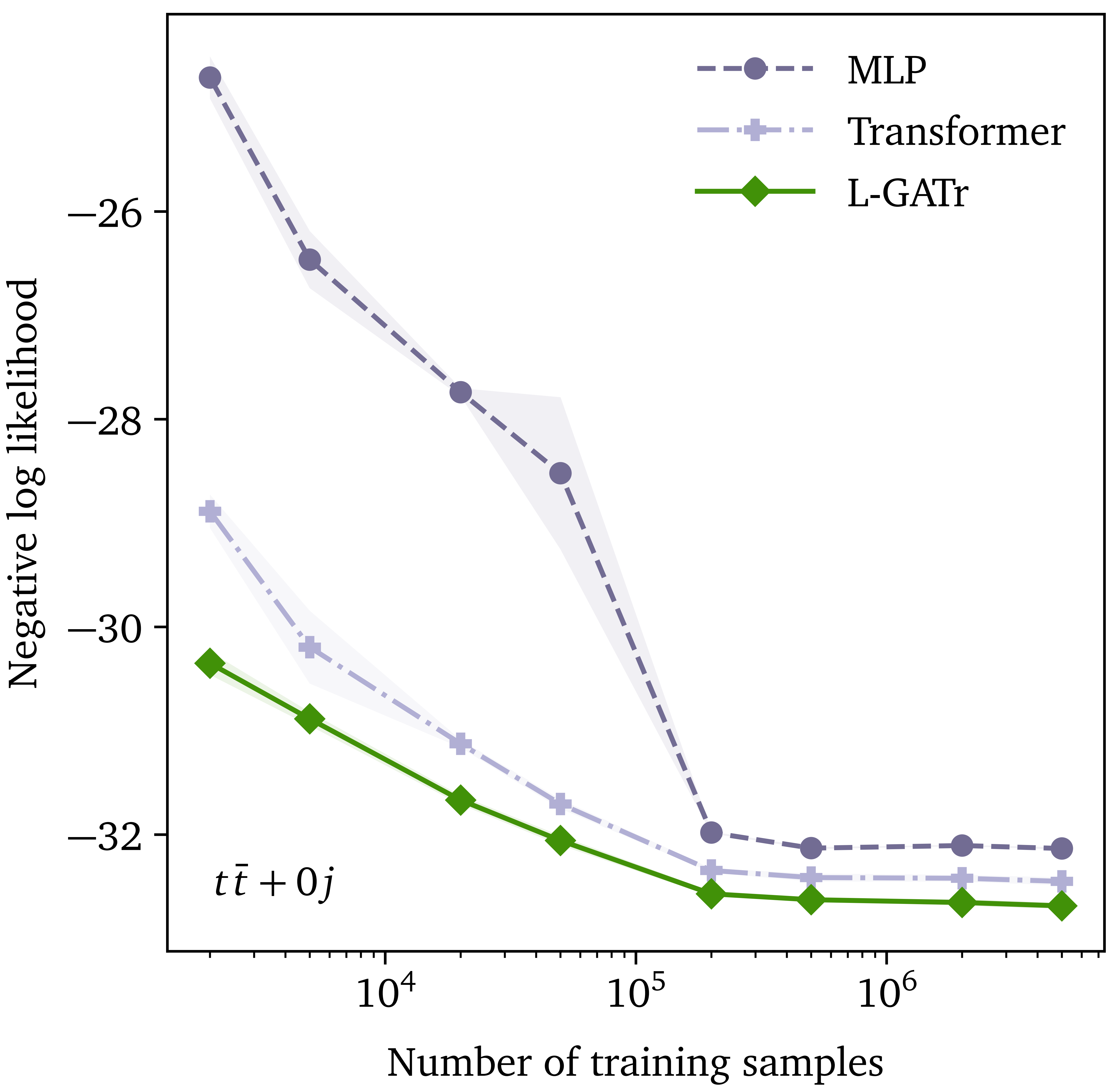
constant in $p = (E, p_x, p_y, p_z)$ ('euclidean')

constant in $y = (y_m, y_p, \phi, \eta)$ ('phasespace-aware')

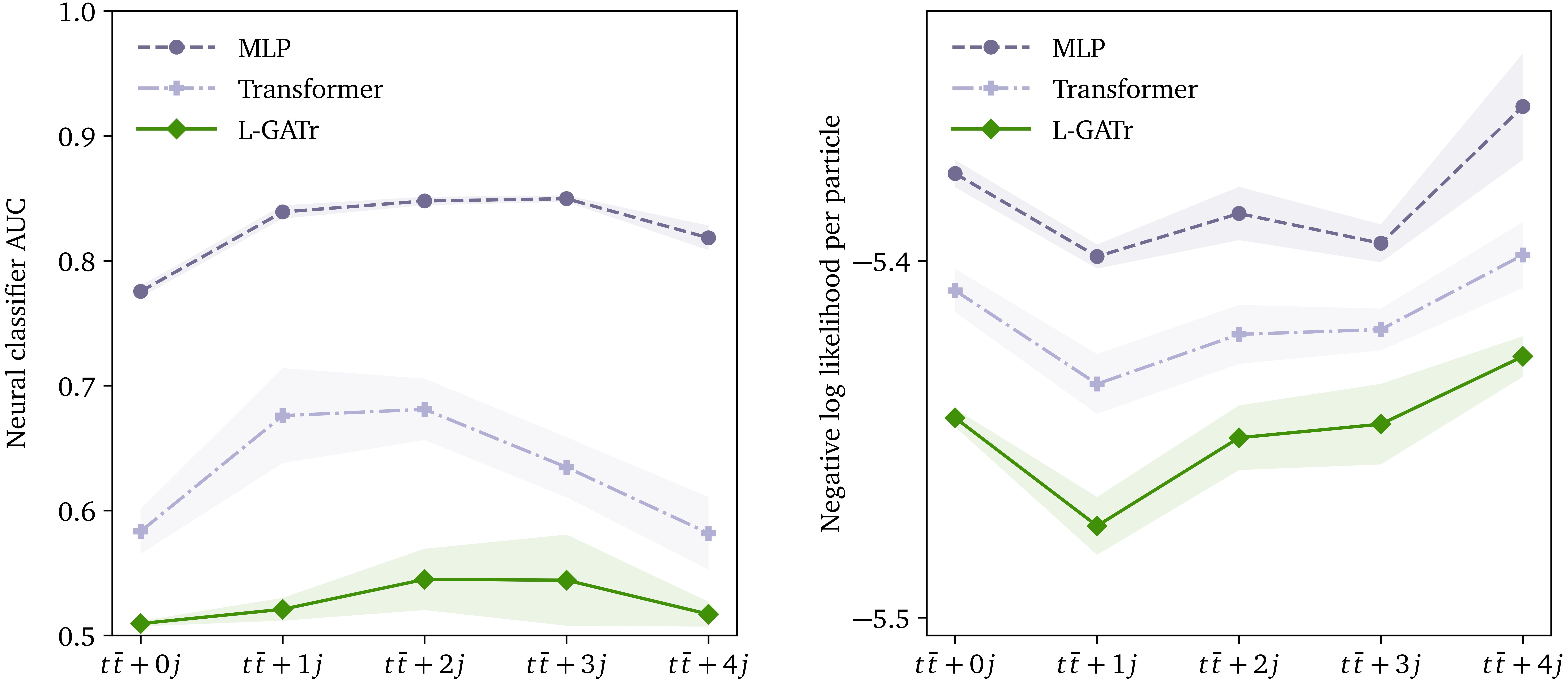


Event generation





Event generation



Tagging + Generation

Symmetry breaking

Sources of symmetry breaking

- Real world: Beam direction, detector geometry…
  Symmetry-breaking object: Beam direction

- Generation: Have to break $SO(1,3) \to SO(3)$ because generative networks can only be defined on compact groups
  Symmetry-breaking object: Time direction

We break the symmetry by adding the symmetry-breaking objects as extra token or as extra channel for each token