Lorentz-GATr

Lorentz-Equivariant Geometric Algebra Transformers for High-Energy Physics

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

Young Scientists Meeting
of the CRC TRR 257

UNIVERSITÄT
HEIDELBERG
ZUKUNFT
SEIT 1386

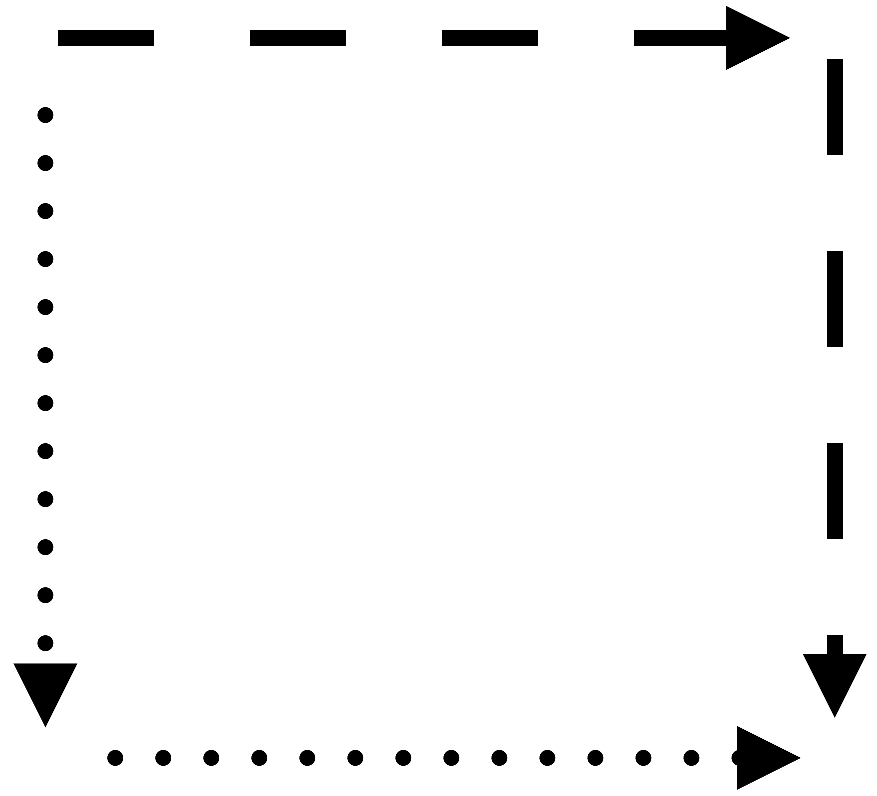
Lorentz symmetry is key in high-energy physics…

$$
\begin{aligned}
\mathcal{L} = & -\frac{1}{4} F_{\mu\nu} F^{\mu\nu} \\
& + i \bar{\psi} \not{D} \psi + h.c \\
& + \psi_i y_{ij} \psi_j \phi + h.c \\
& + |D_\mu \phi|^2 - V(\phi)
\end{aligned}
$$

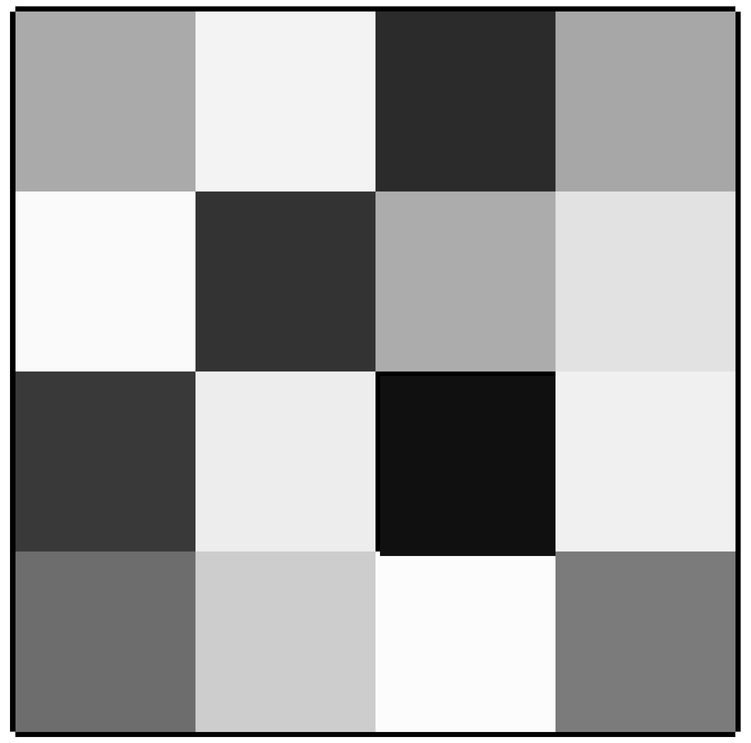
… so let's build it into our neural networks

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

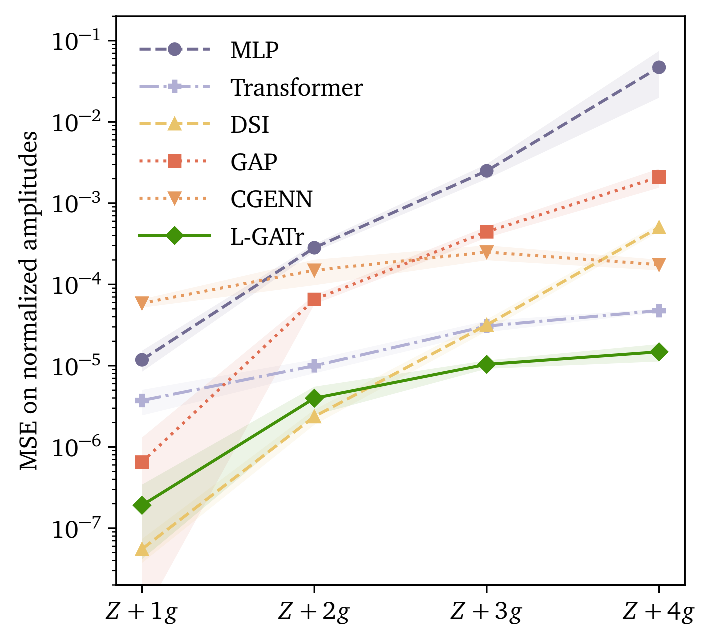
Geometric algebra
representations



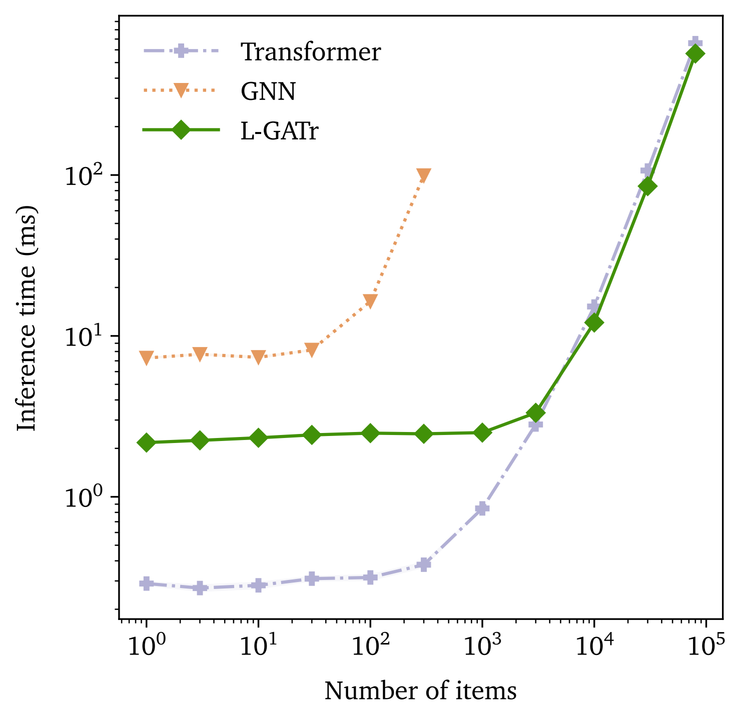
Equivariant
layers



Transformer
architecture

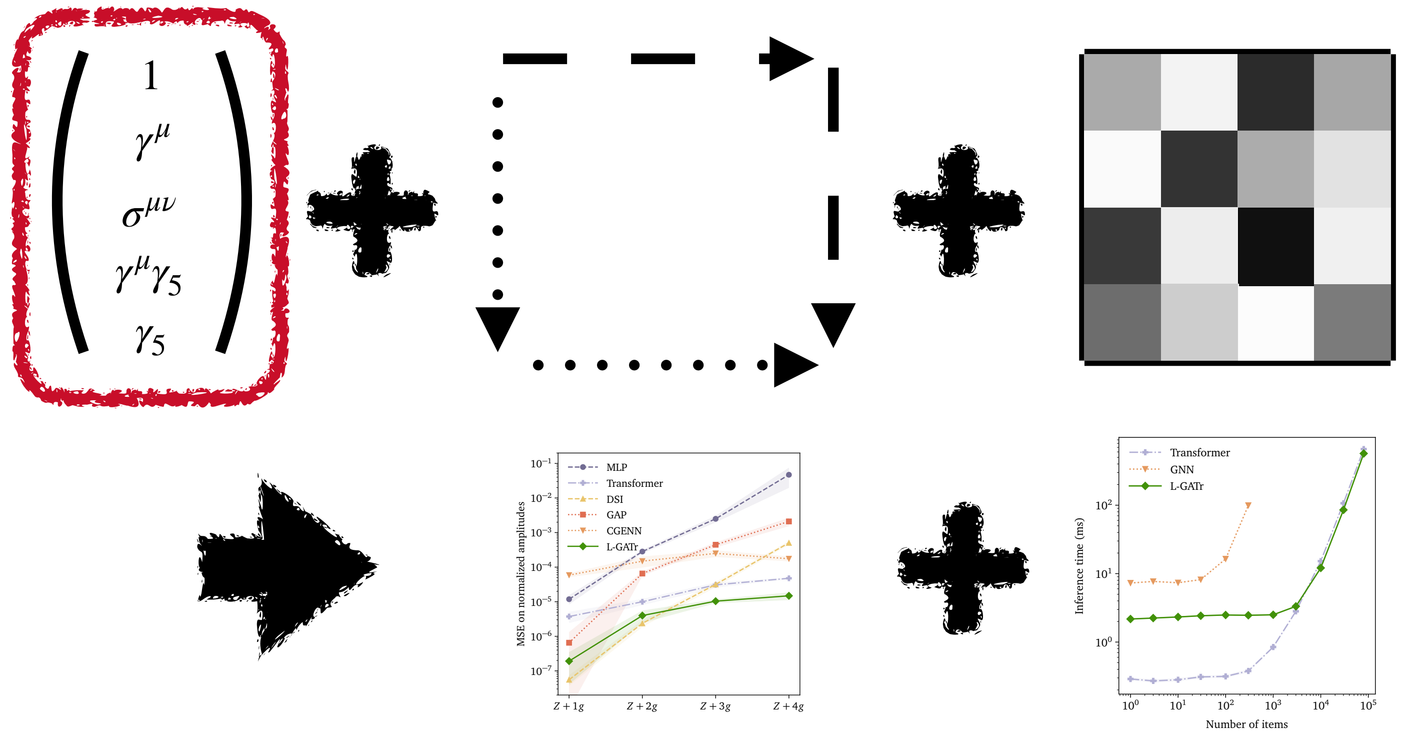


Strong performance
on diverse problems



Scalable
to thousands of tokens

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



Ingredients

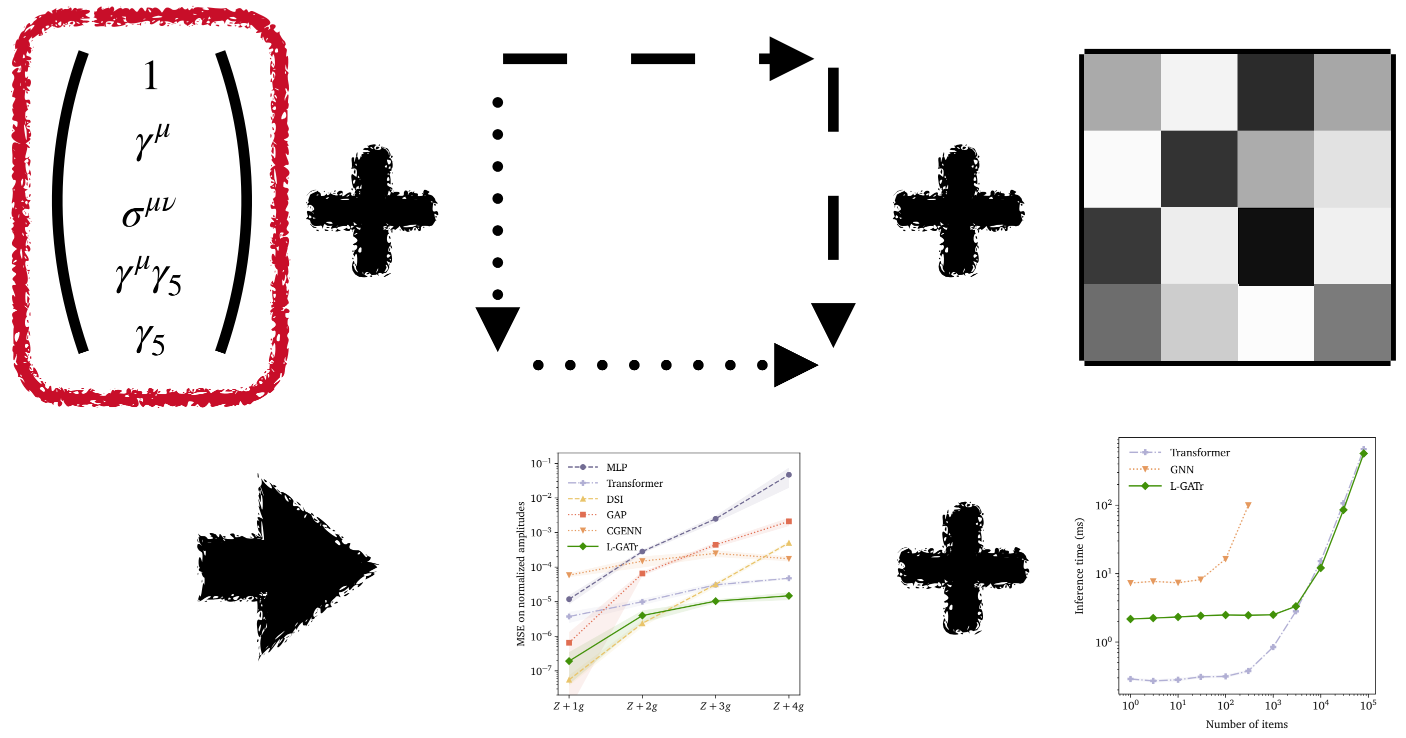
Geometric algebra representations

- Basis elements γ^μ of the geometric algebra defined by $\{\gamma^\mu, \gamma^\nu\} = 2g^{\mu\nu}$
  - Operations: $\alpha x, \quad x + y, \quad x \cdot y$
  - General multivector: $x = x^S 1 + x^V_\mu \gamma^\mu + x^T_{\mu\nu}\sigma^{\mu\nu} + x^A_\mu \gamma^\mu\gamma_5 + x^P\gamma_5$

Geometric algebra = Clifford algebra

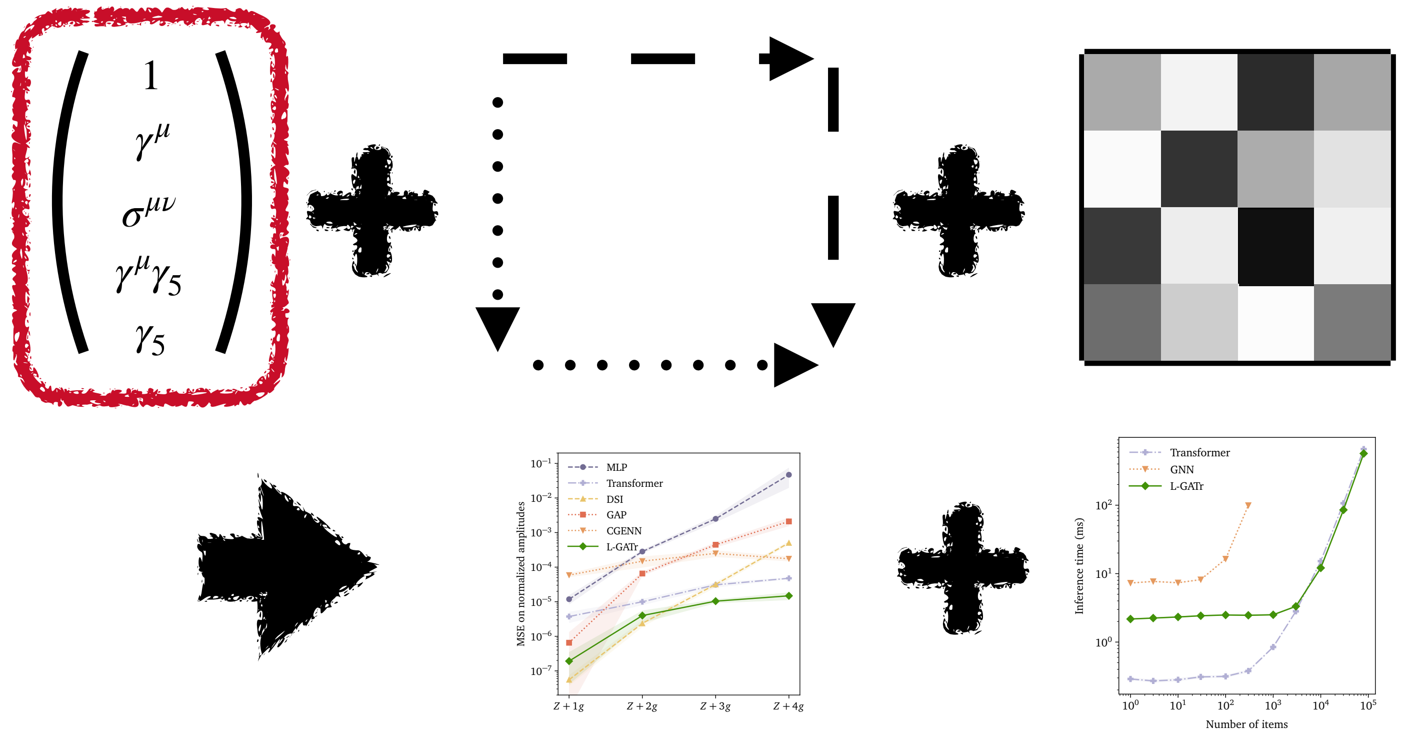
Ingredients

Geometric algebra representations



- Basis elements γ^μ of the geometric 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 $(x^S, x^V_0 \cdots x^V_3, x^T_{01} \cdots x^T_{23}, x^A_0 \cdots x^A_3, x^P) \in \mathbb{R}^{16}$
  - Usually: $x^S = \text{PID}, \quad x^V_\mu = p_\mu, \quad x^{T,A,P} = 0$
- L-GATr has n multivector and m scalar representations for each particle

Geometric algebra = Clifford algebra



Ingredients

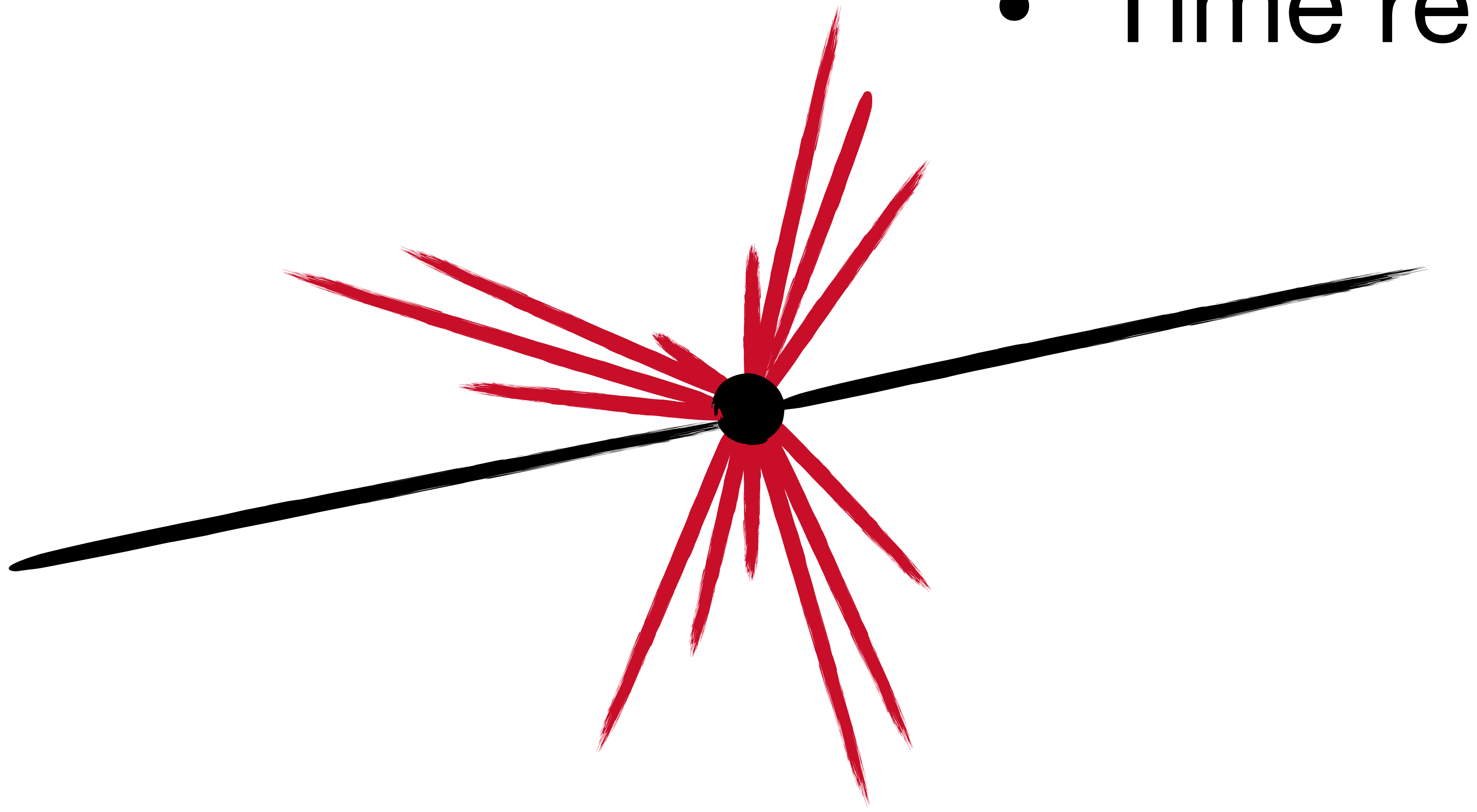
Symmetry breaking with spurious

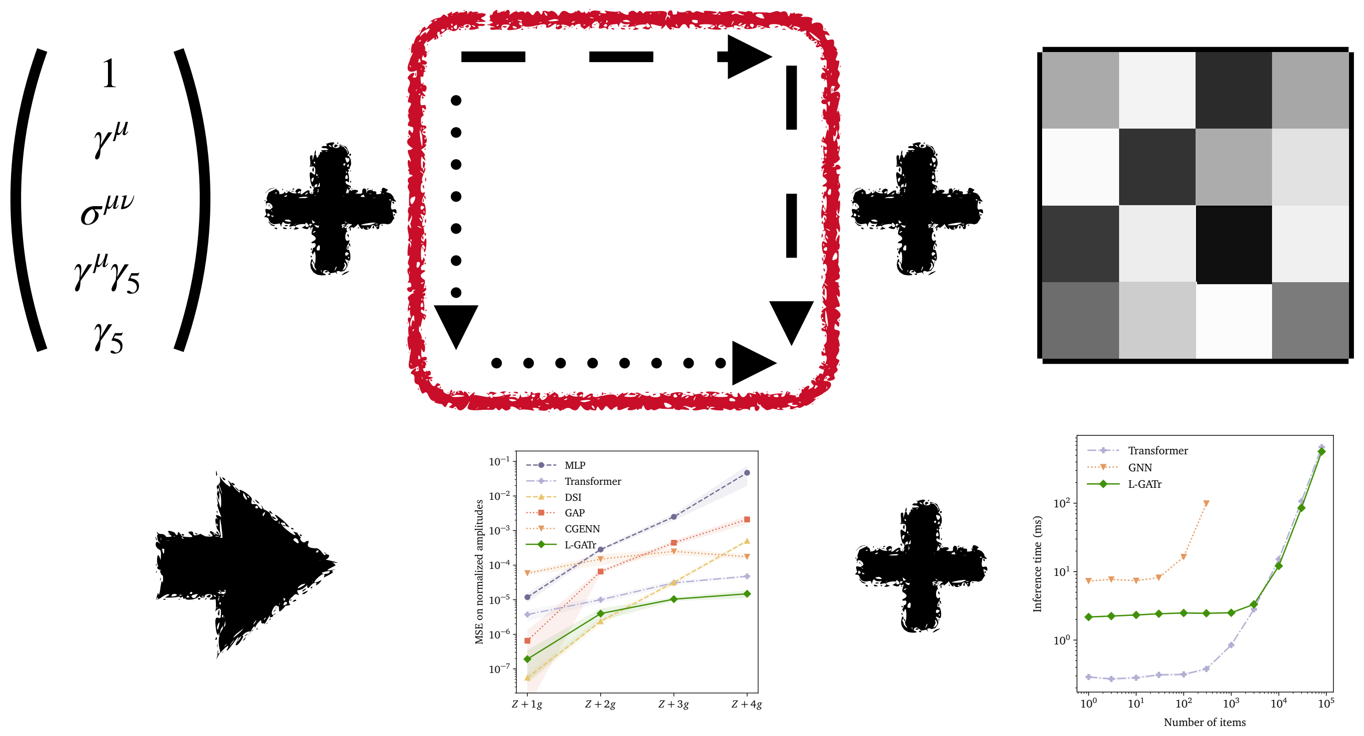
Lorentz symmetry is rarely exact

- Beam direction in collider
- Detector effects
- …?

Add a spurion to the particle list (either as token or channel)

- Beam reference: $p^\mu = (0,0,0,\pm 1)$
- Time reference: $p^\mu = (1,0,0,0)$





Ingredients

Equivariance

symmetry group transformation $\mathscr{G}$

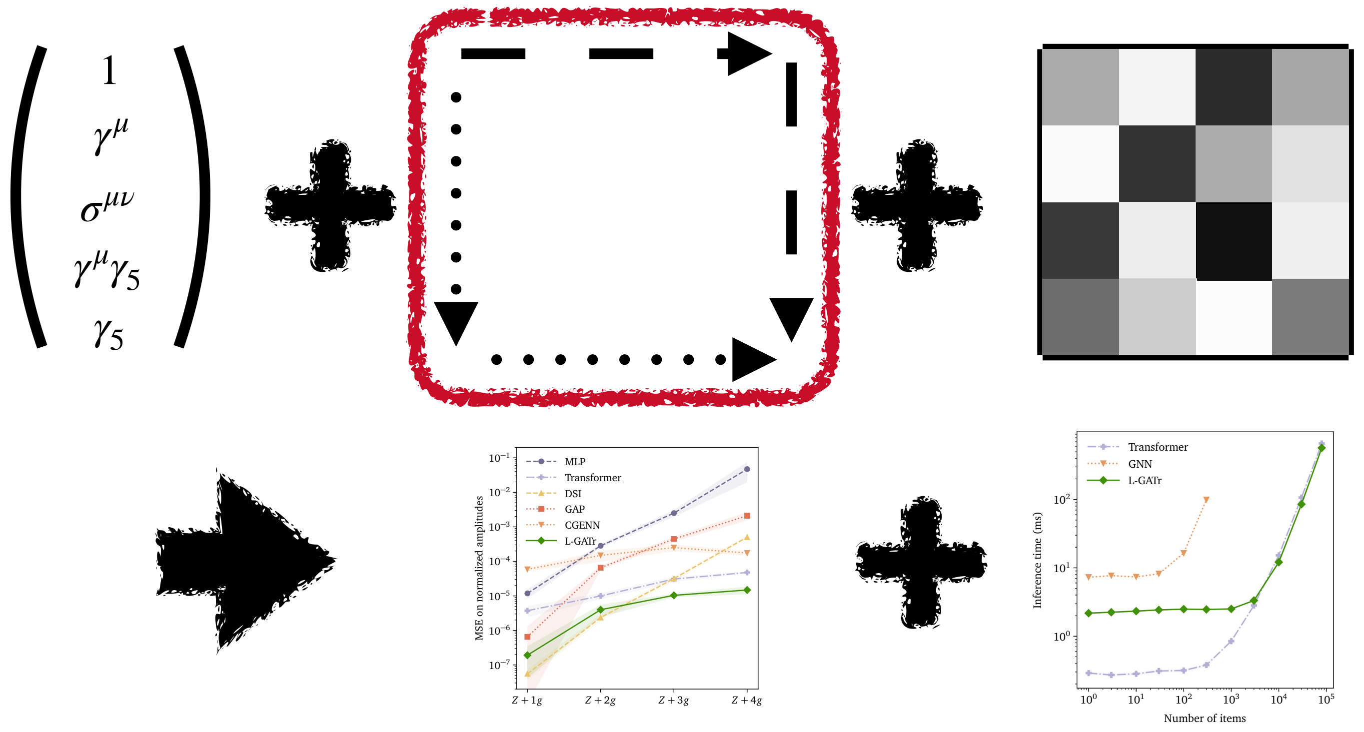
neural network transformation $\mathscr{N}$

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

$\mathscr{N}$

$\mathscr{G}$

Equivariance = Covariance



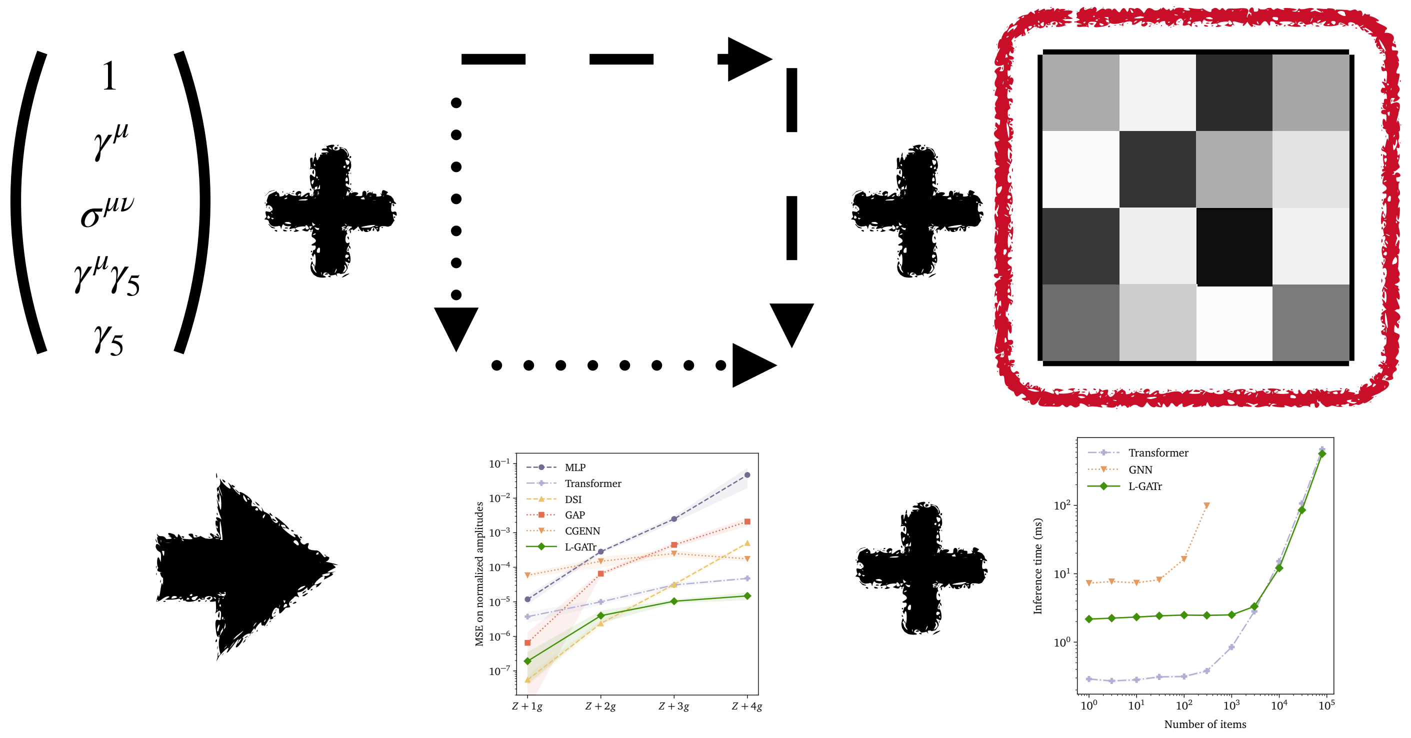
Ingredients

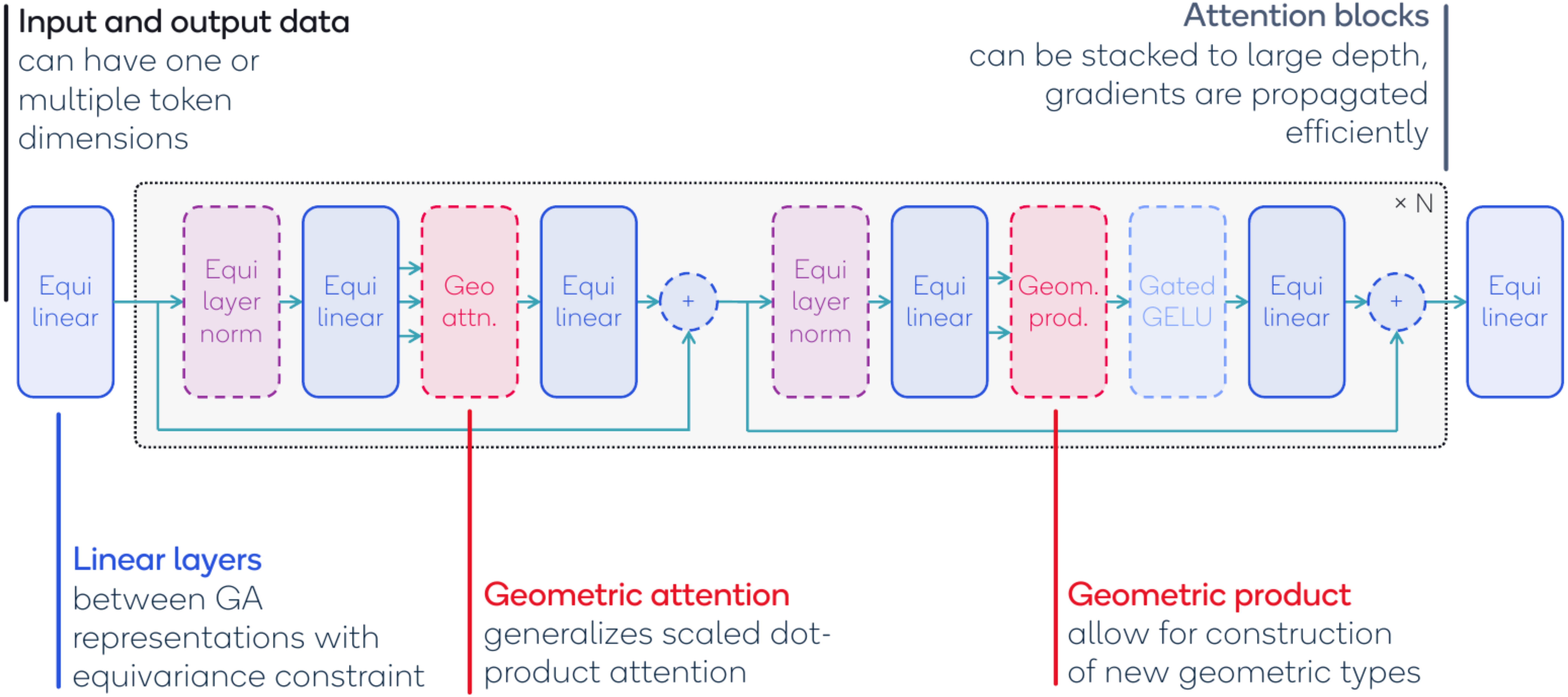
Equivariant layers

	Transformer	L-GATr
Linear(x)	$v\, x + c$	$\sum_{k=0}^{4} v_k \langle x \rangle_k + \sum_{k=0}^{4} w_k \gamma_5 \langle x \rangle_k$
Attention$(q,k,v)_{i\alpha}$	$\sum_{j,\beta} \mathrm{Softmax}_j\left(\frac{q_{i\beta}, k_{j\beta}}{\sqrt{n}}\right) v_{j\alpha}$	$\sum_{j,\beta} \mathrm{Softmax}_j\left(\frac{\langle q_{i\beta}, k_{j\beta} \rangle}{\sqrt{16n}}\right) v_{j\alpha}$
GeometricProduct(x,y)	–	$x \cdot y$
Dropout, LayerNorm, activation function, …	See bonus material	See bonus material

Ingredients

Transformer architecture

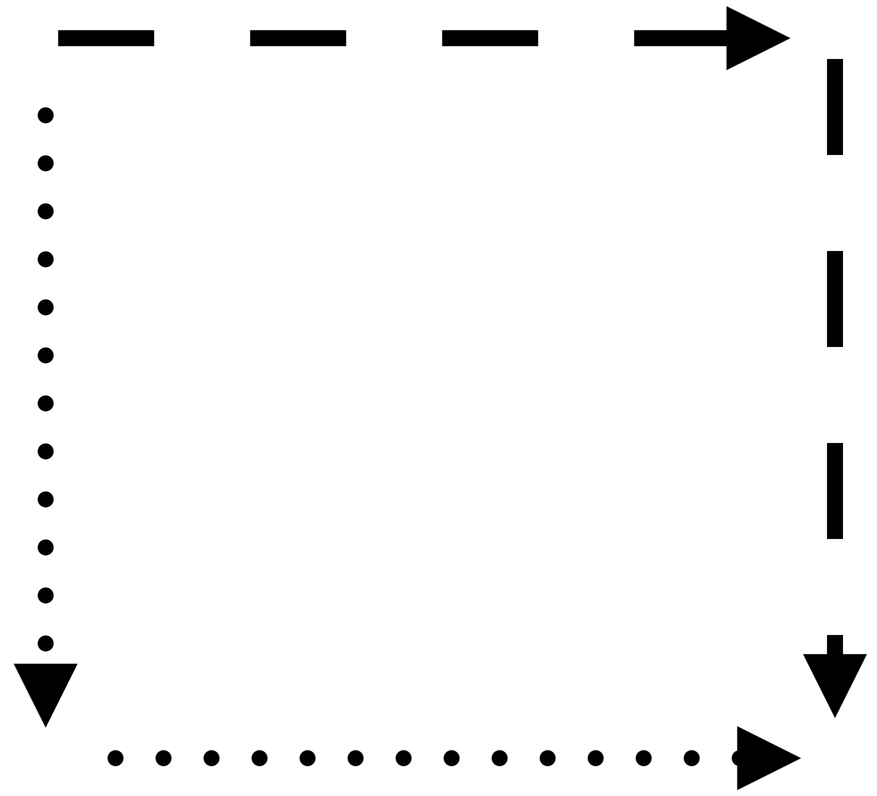




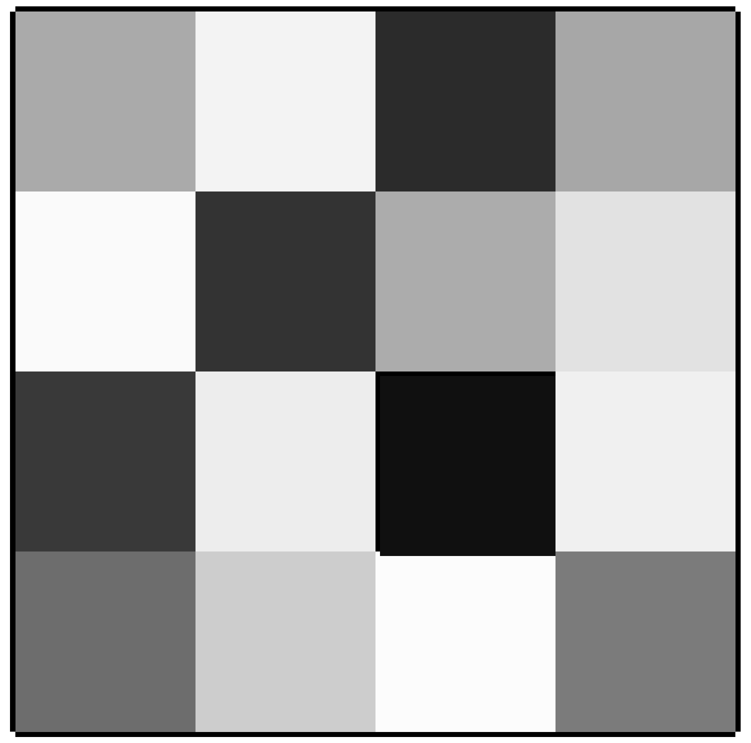
Credits to Johann Brehmer

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

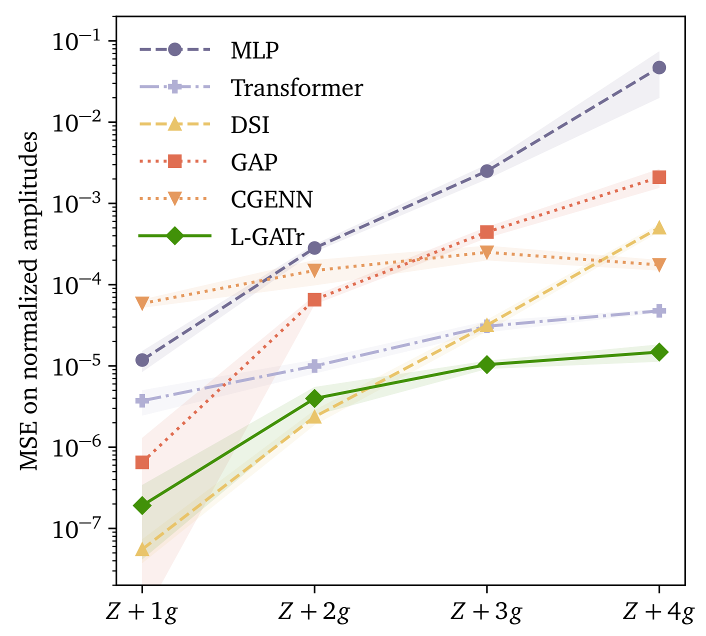
Geometric algebra representations



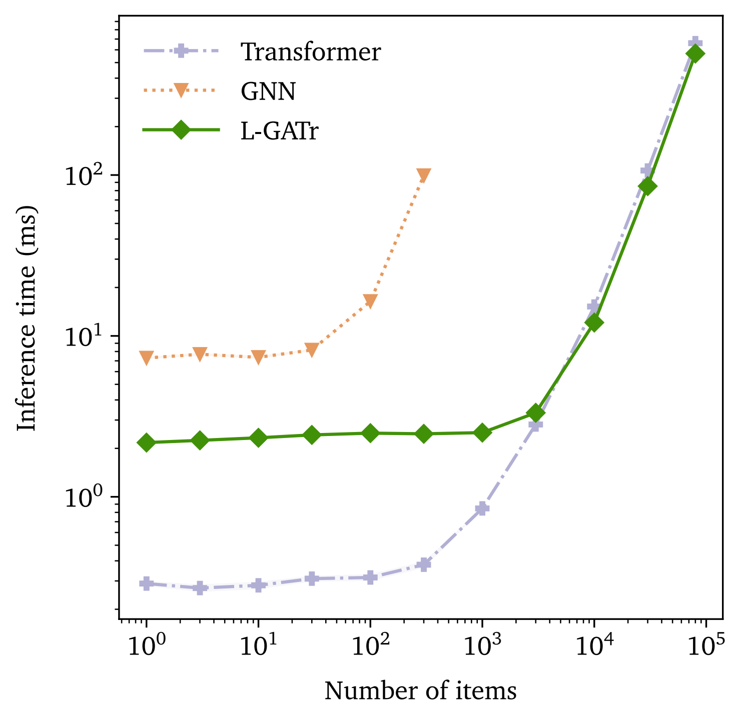
Equivariant layers



Transformer architecture

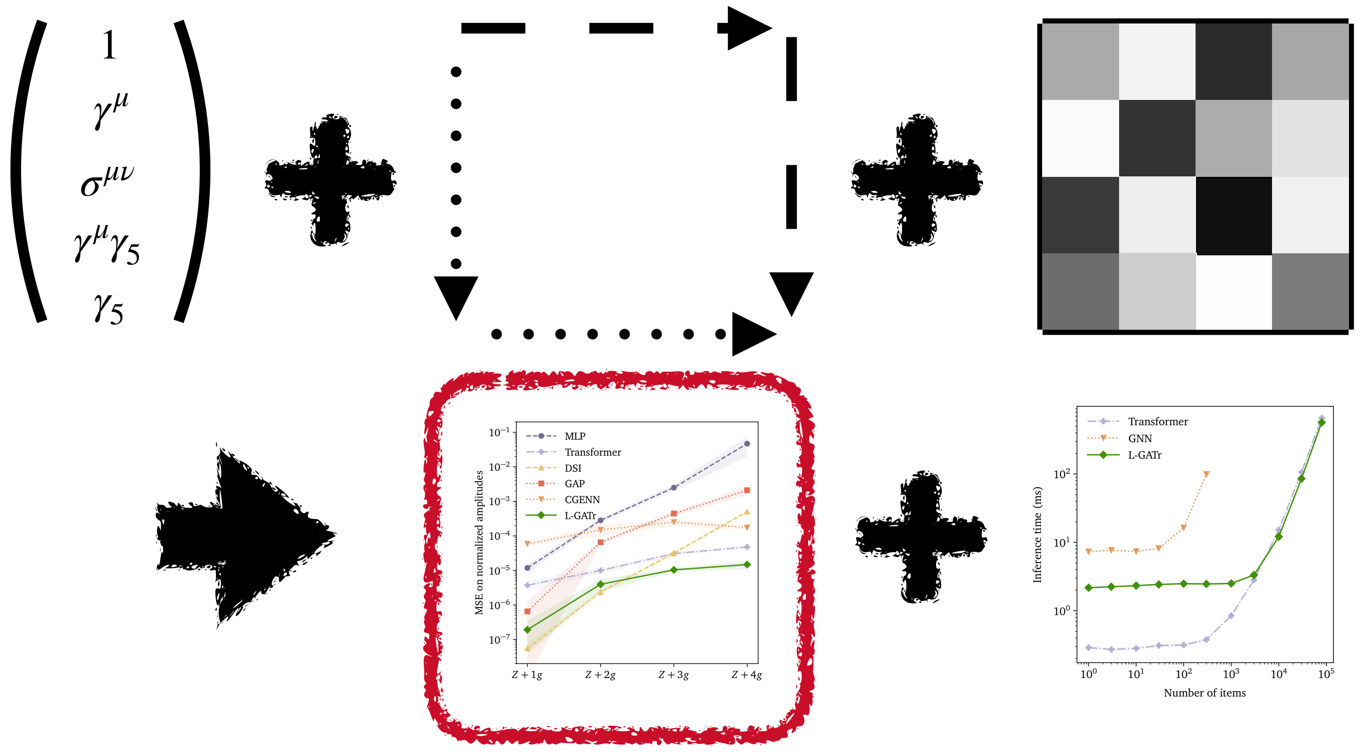


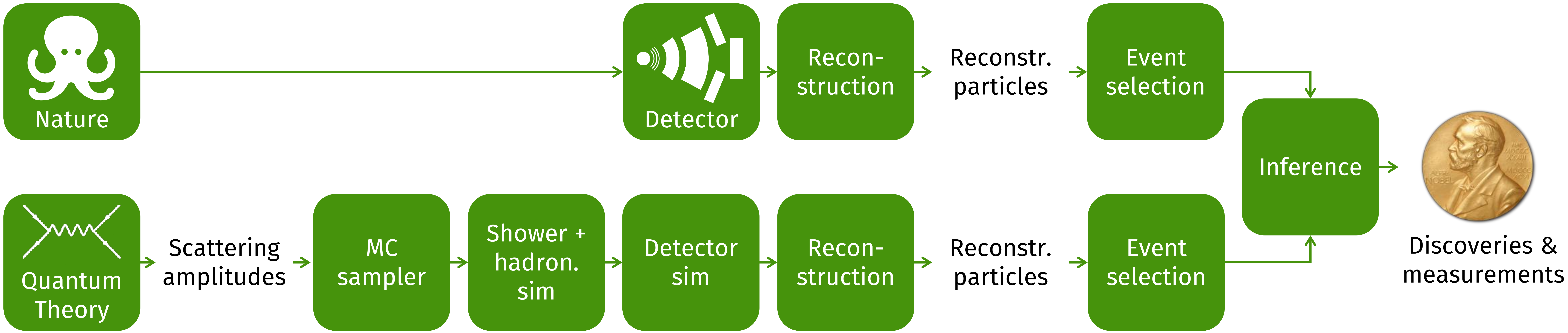
Strong performance on diverse problems



Scalable to thousands of tokens

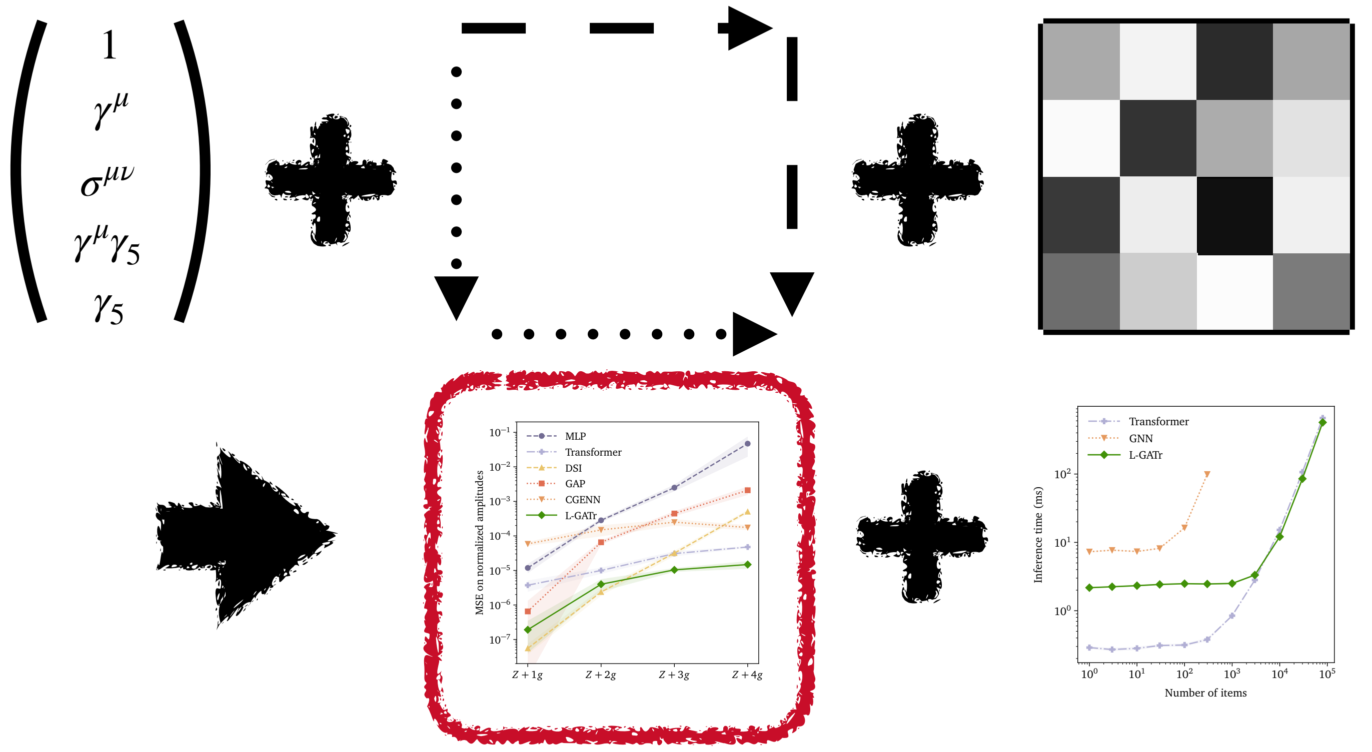
Experiments
LHC simulation chain

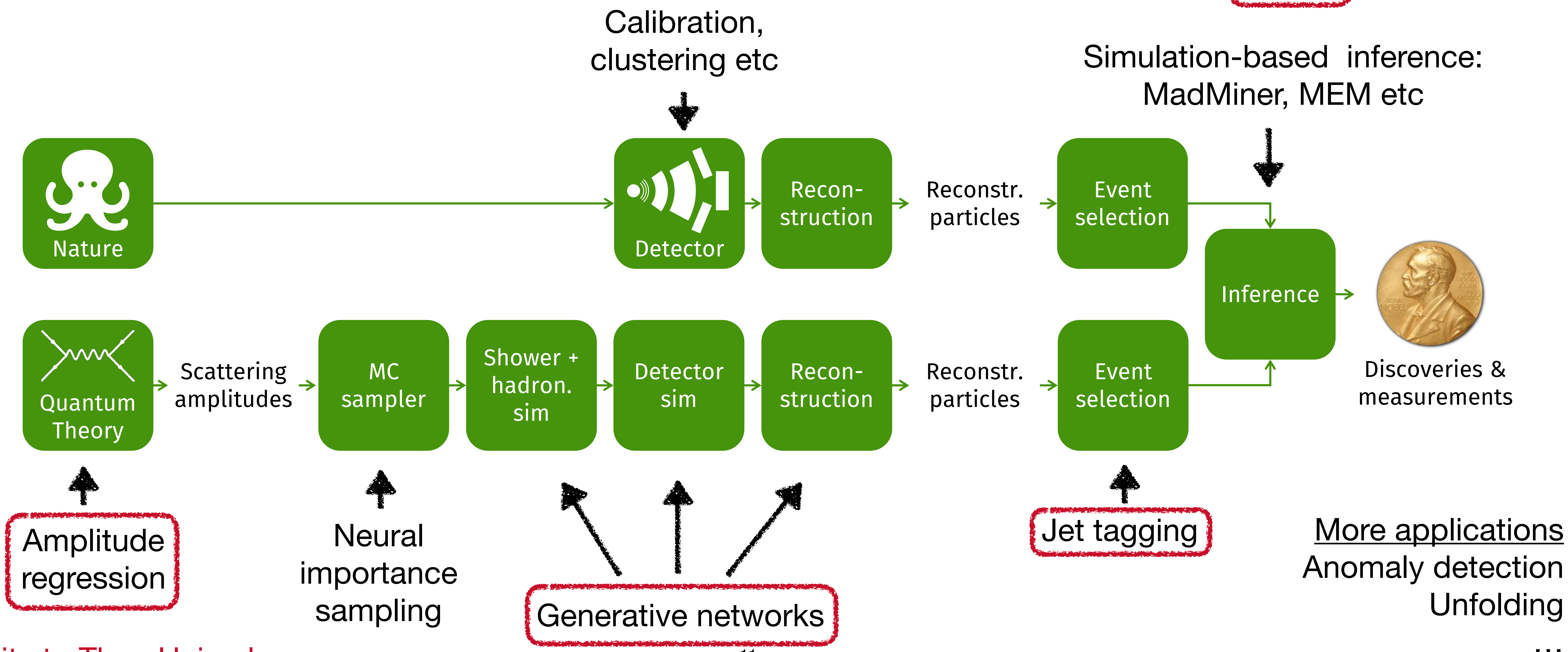




Experiments

LHC simulation chain meets ML

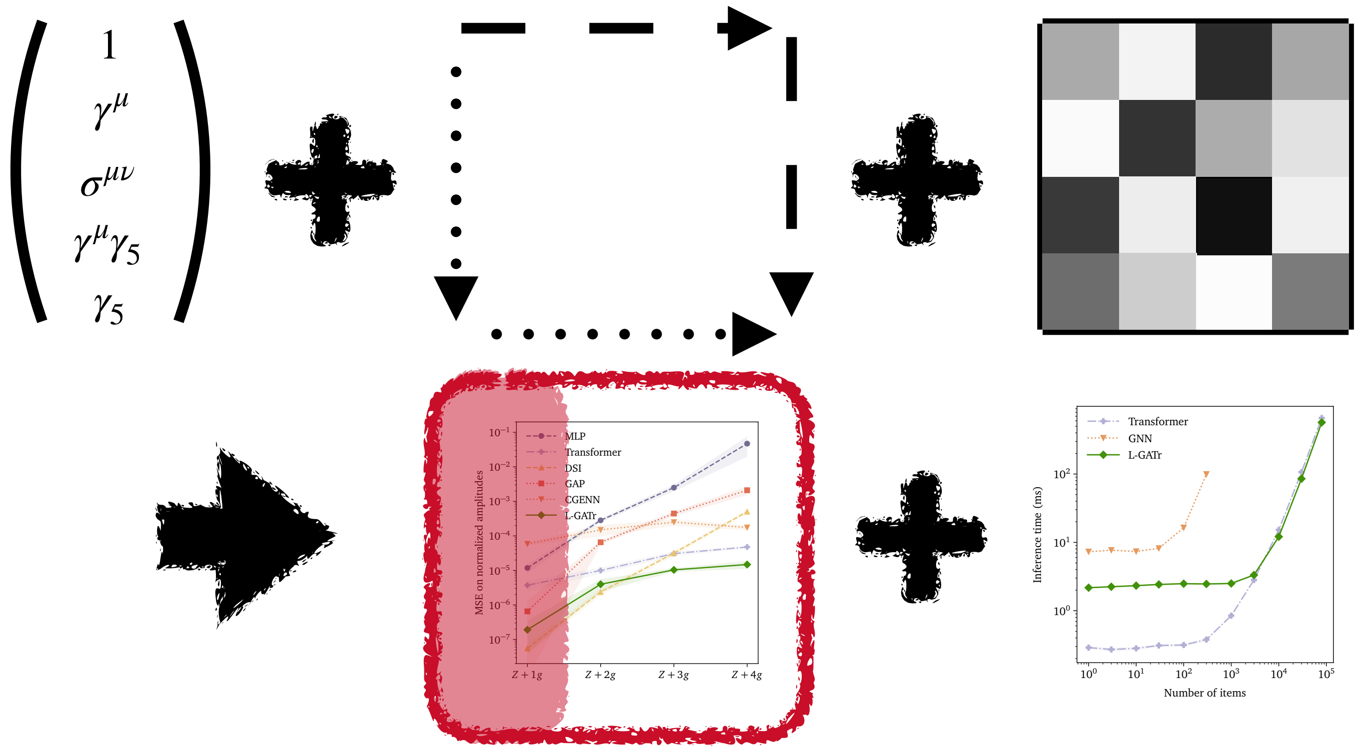


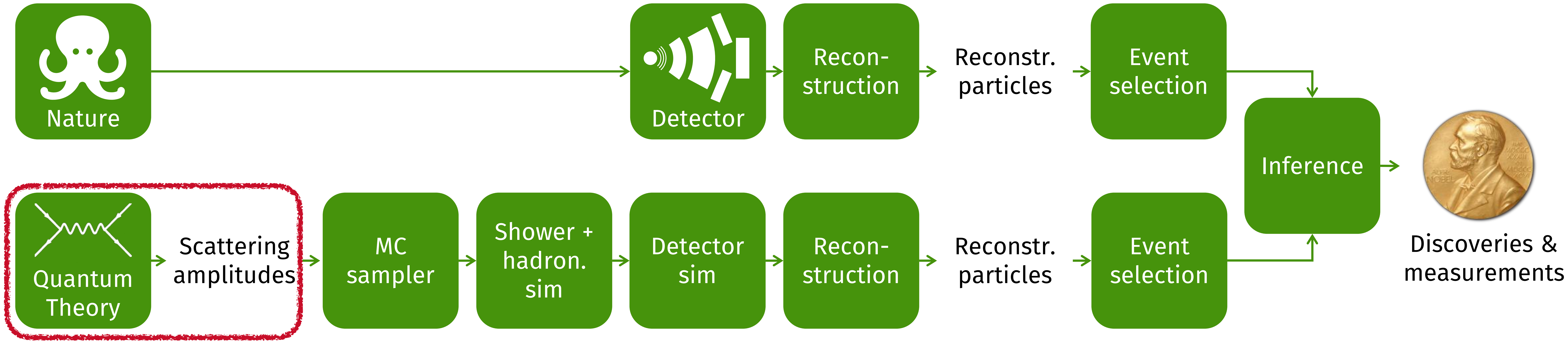


Credits to Theo Heimel

Experiments

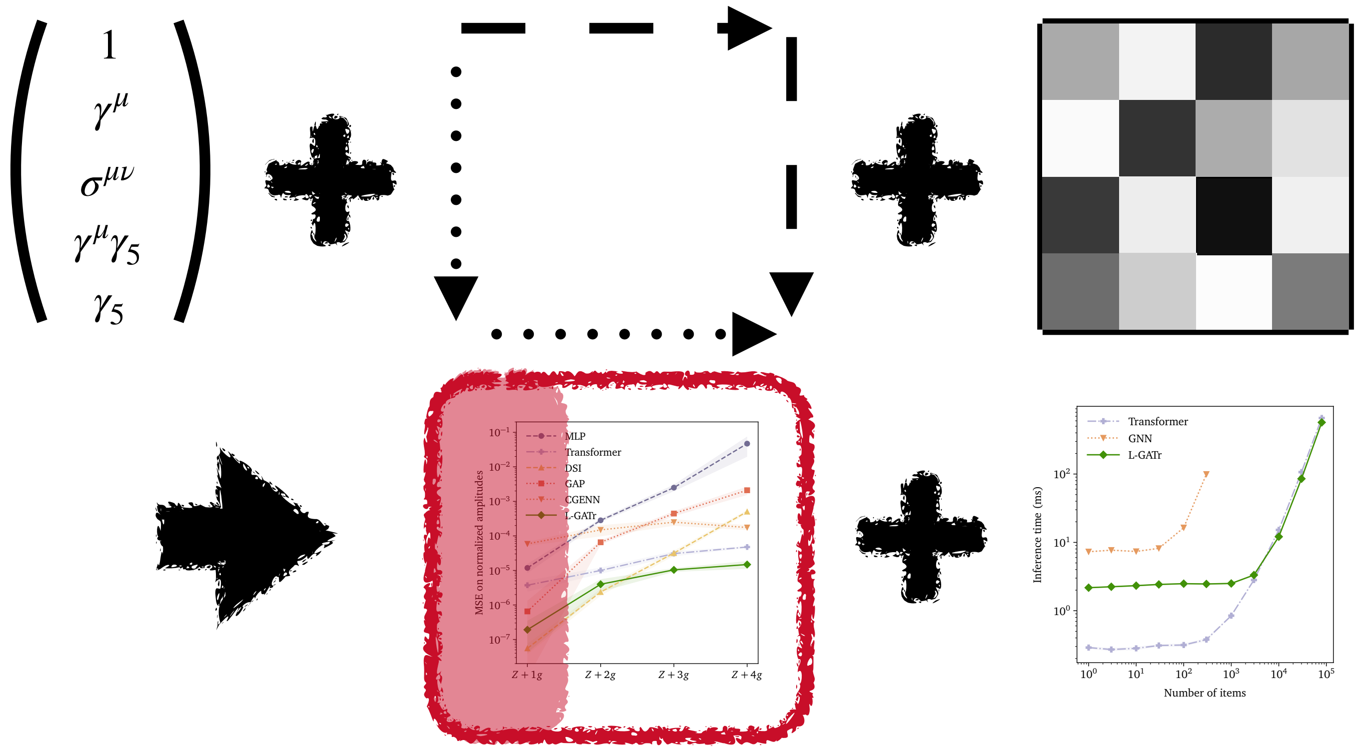
Amplitude regression

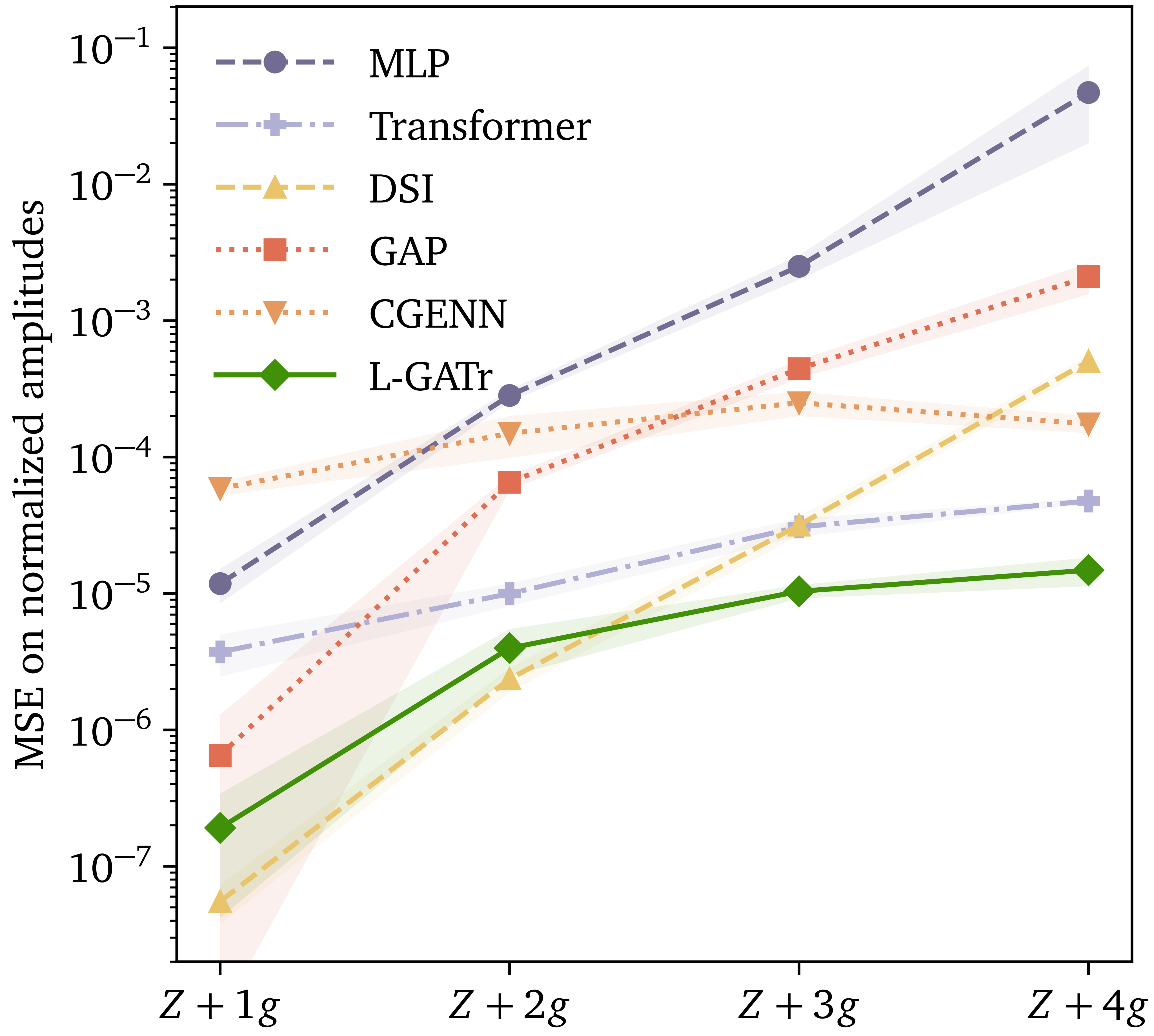




Experiments

Amplitude regression

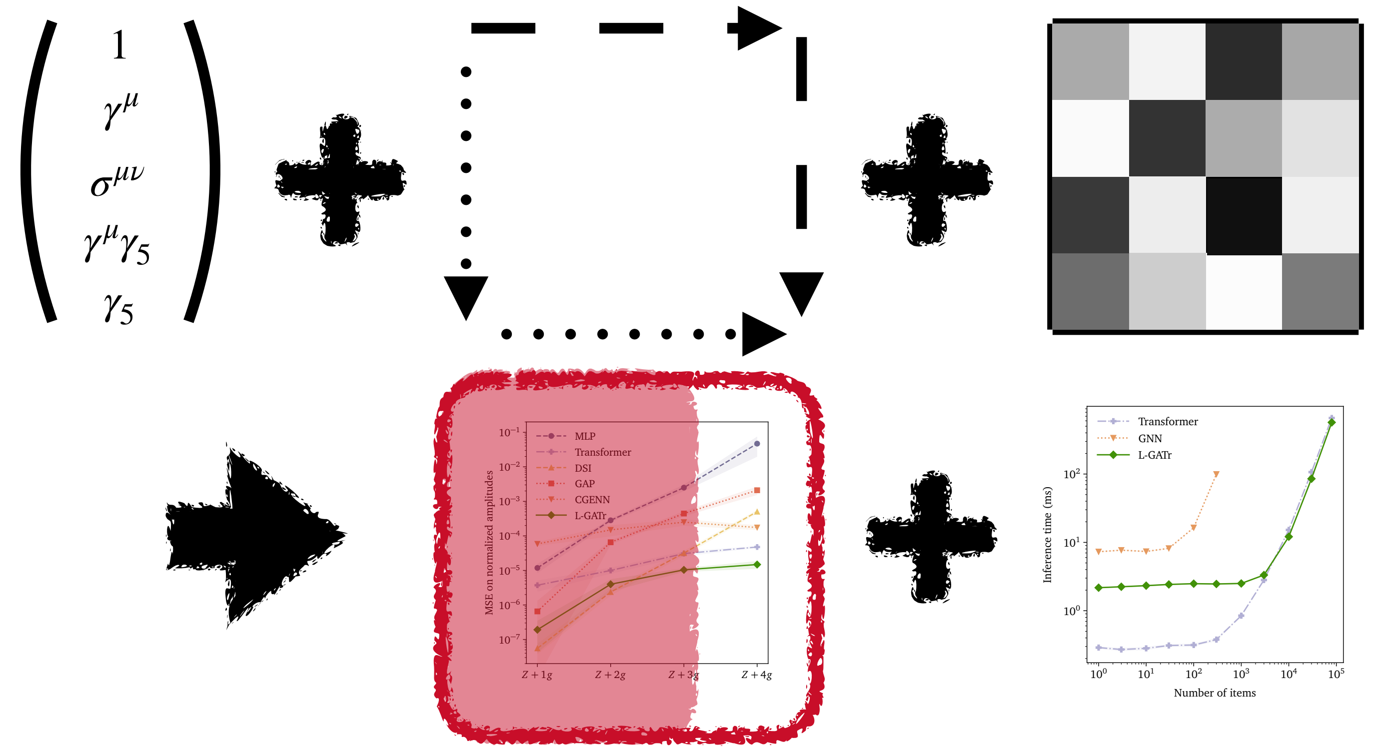


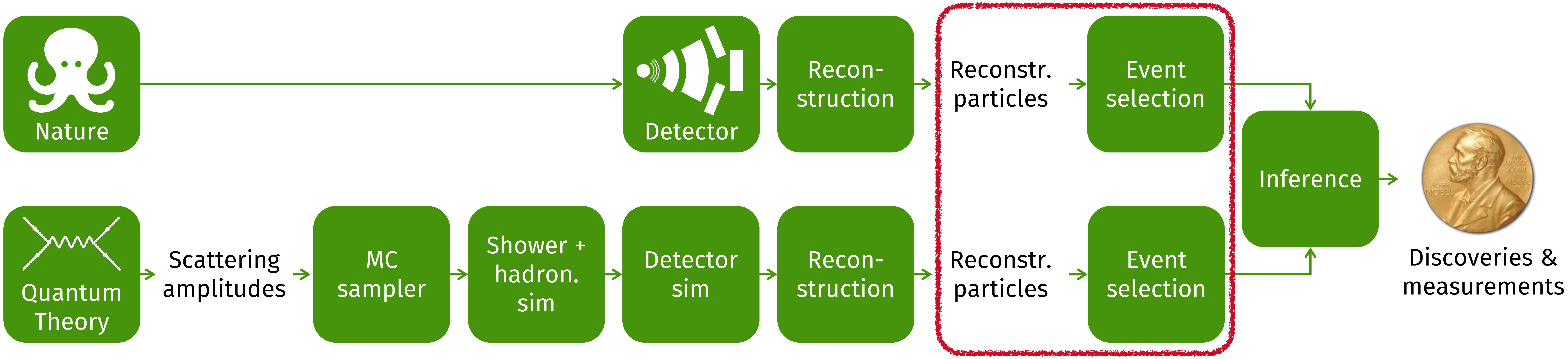


L-GATr scales best to high multiplicity, where amplitude surrogates are most useful

Experiments

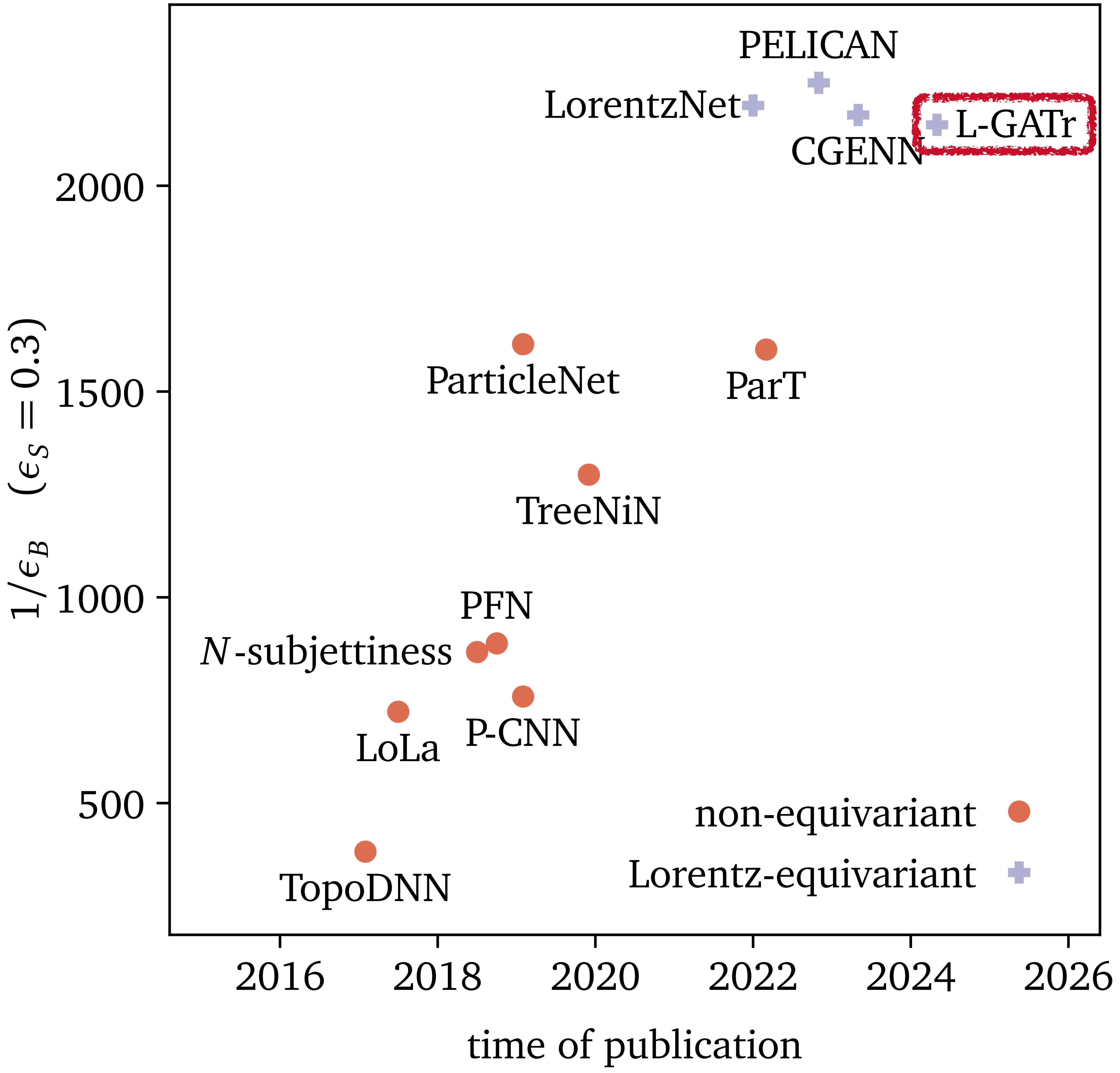
Top tagging

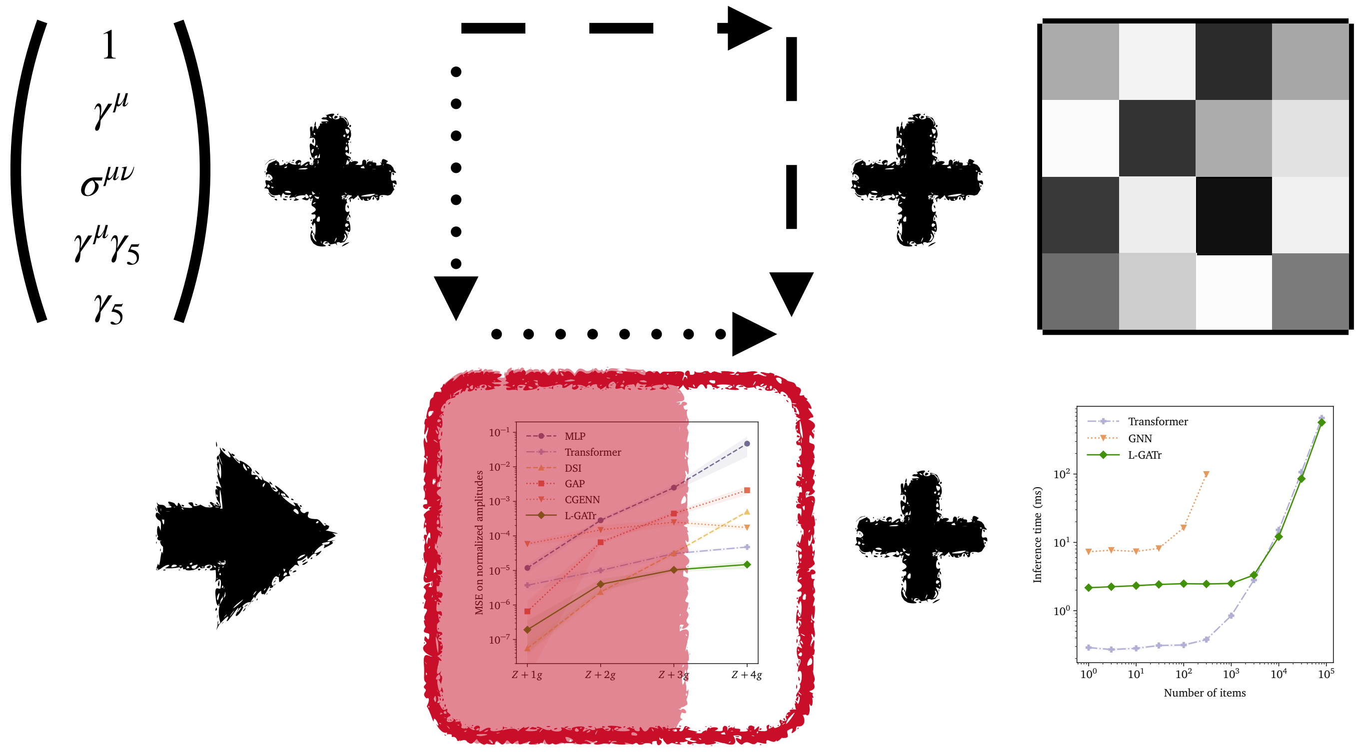




Experiments

Top tagging

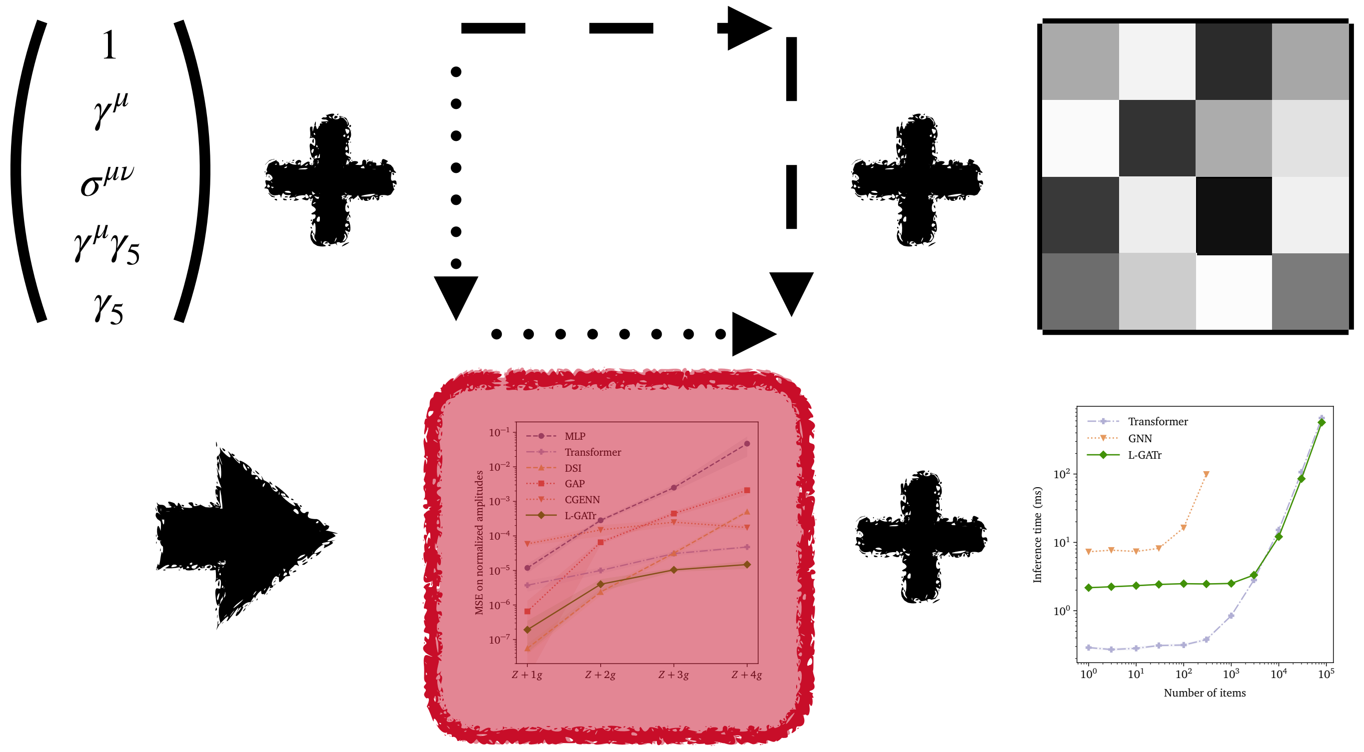


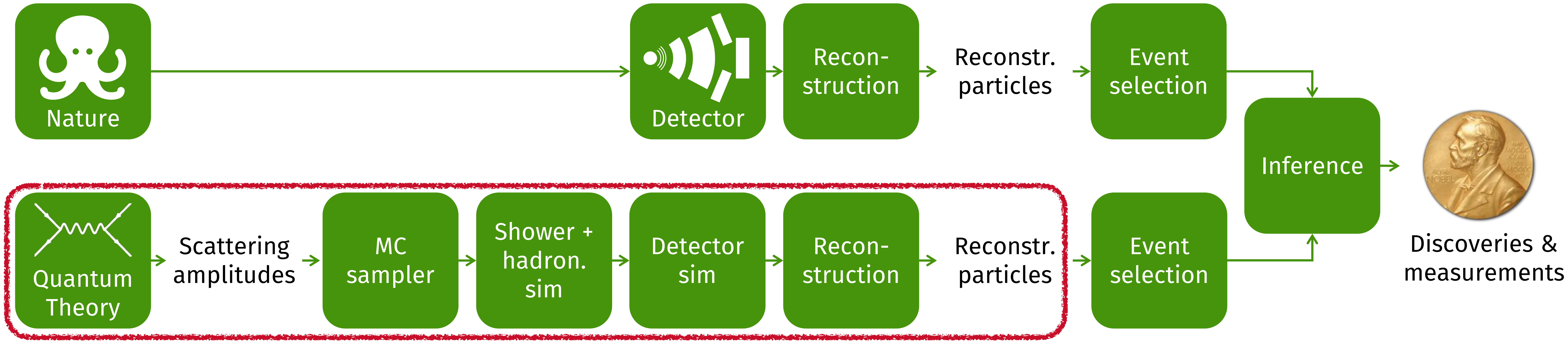


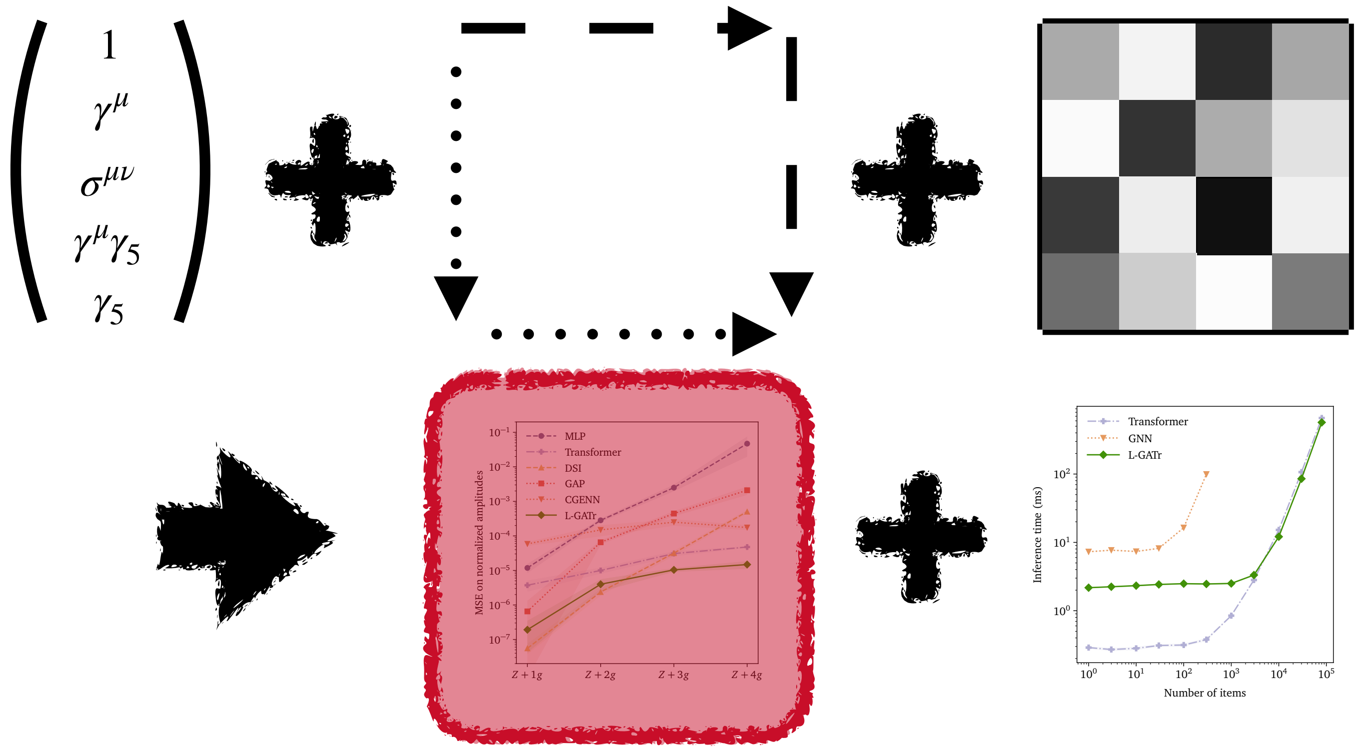
L-GATr is on par with the best equivariant (*) baselines

Experiments

Event generation







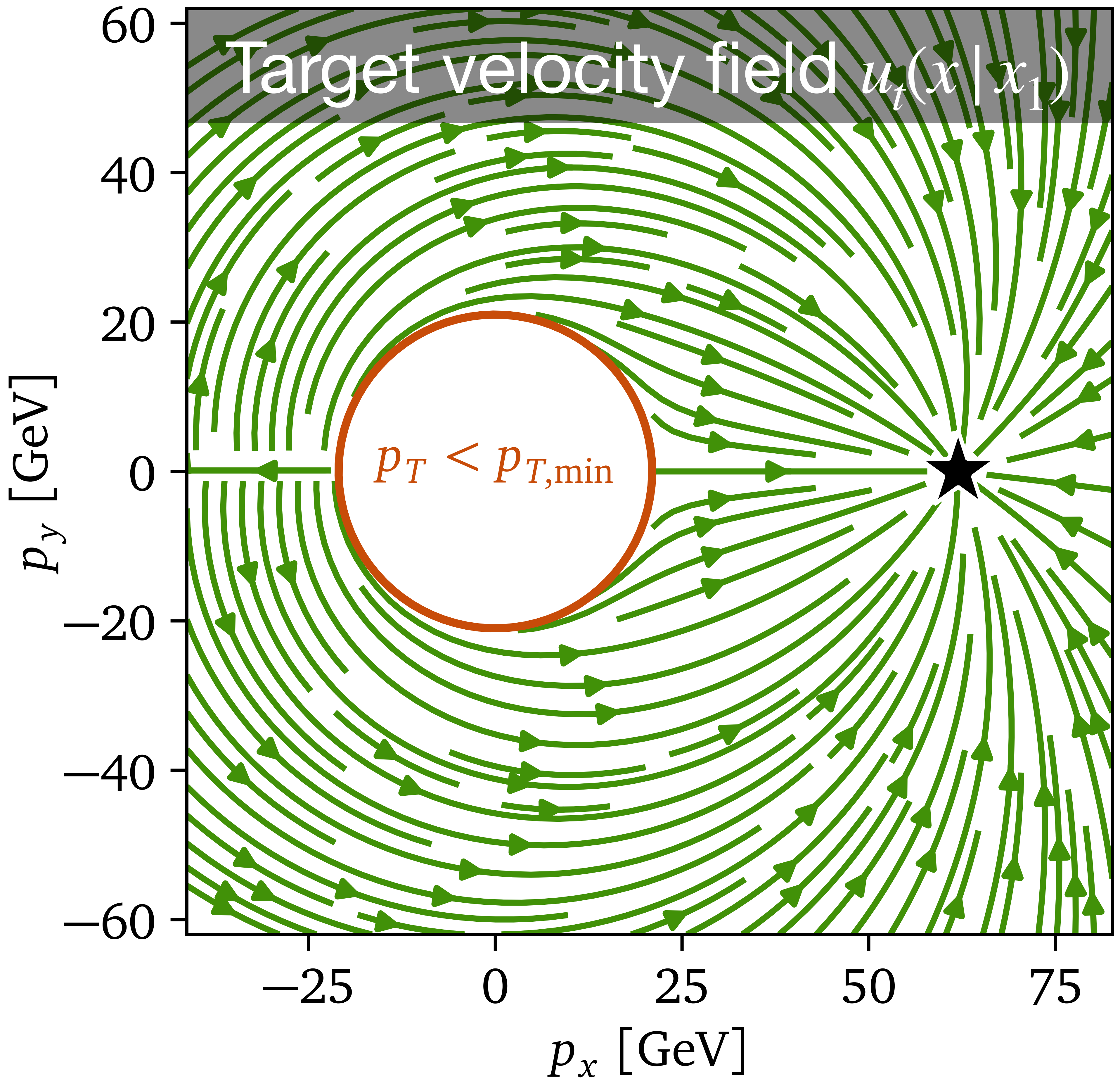
Experiments

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 $u_t(x \mid x_1)$

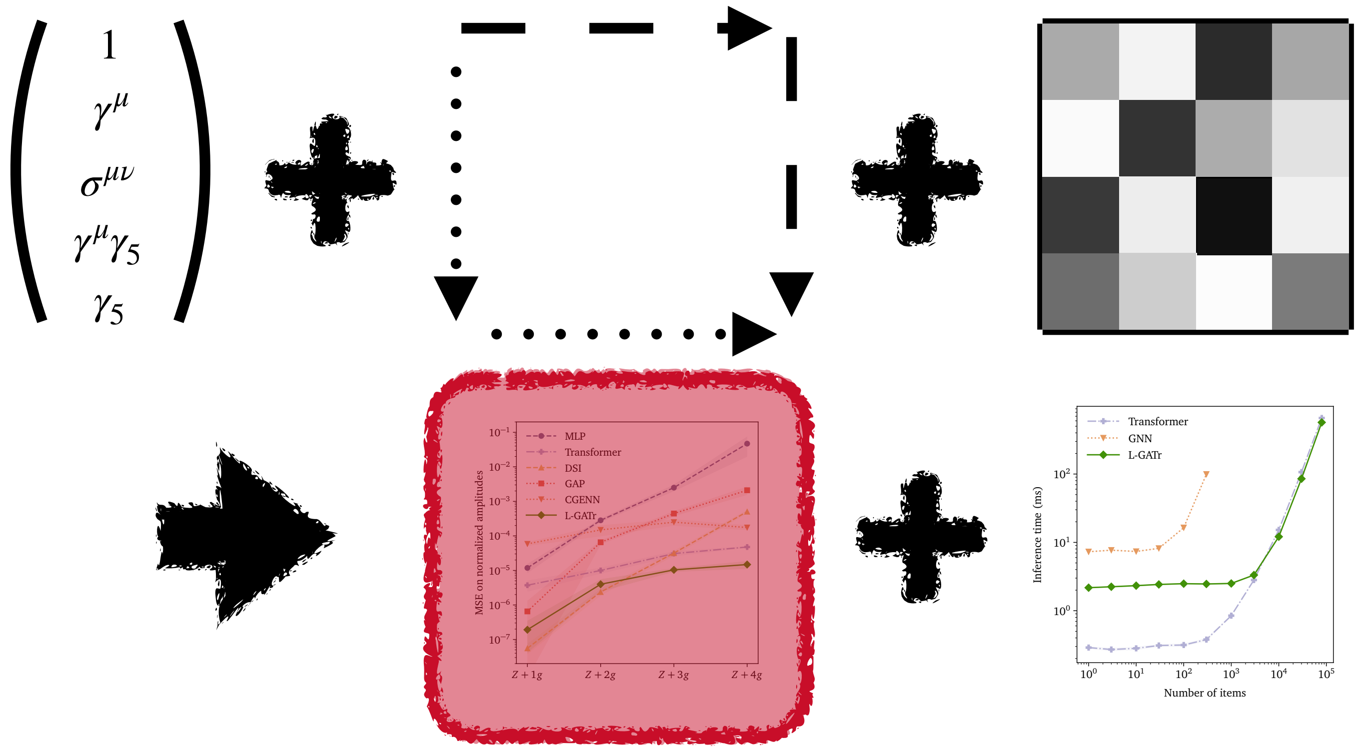


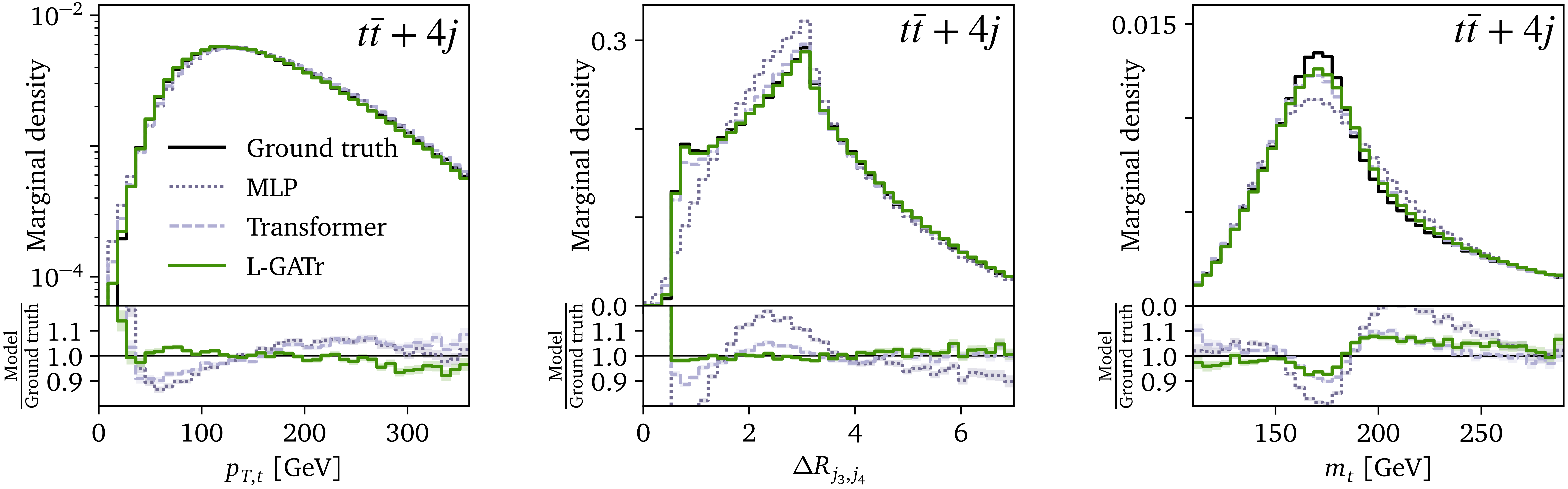
Continuous normalising flows
arXiv:1806.07366

Conditional flow matching
arXiv:2210.02747

Experiments

Event generation

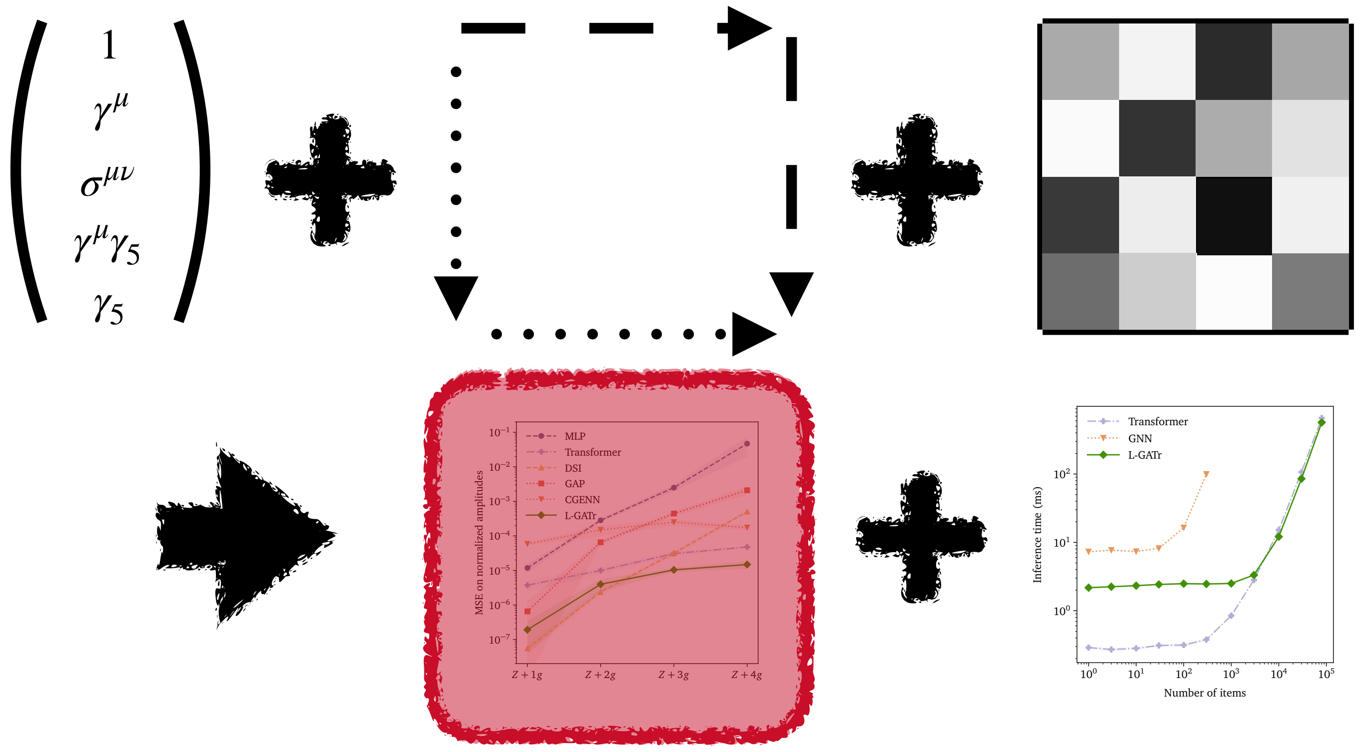


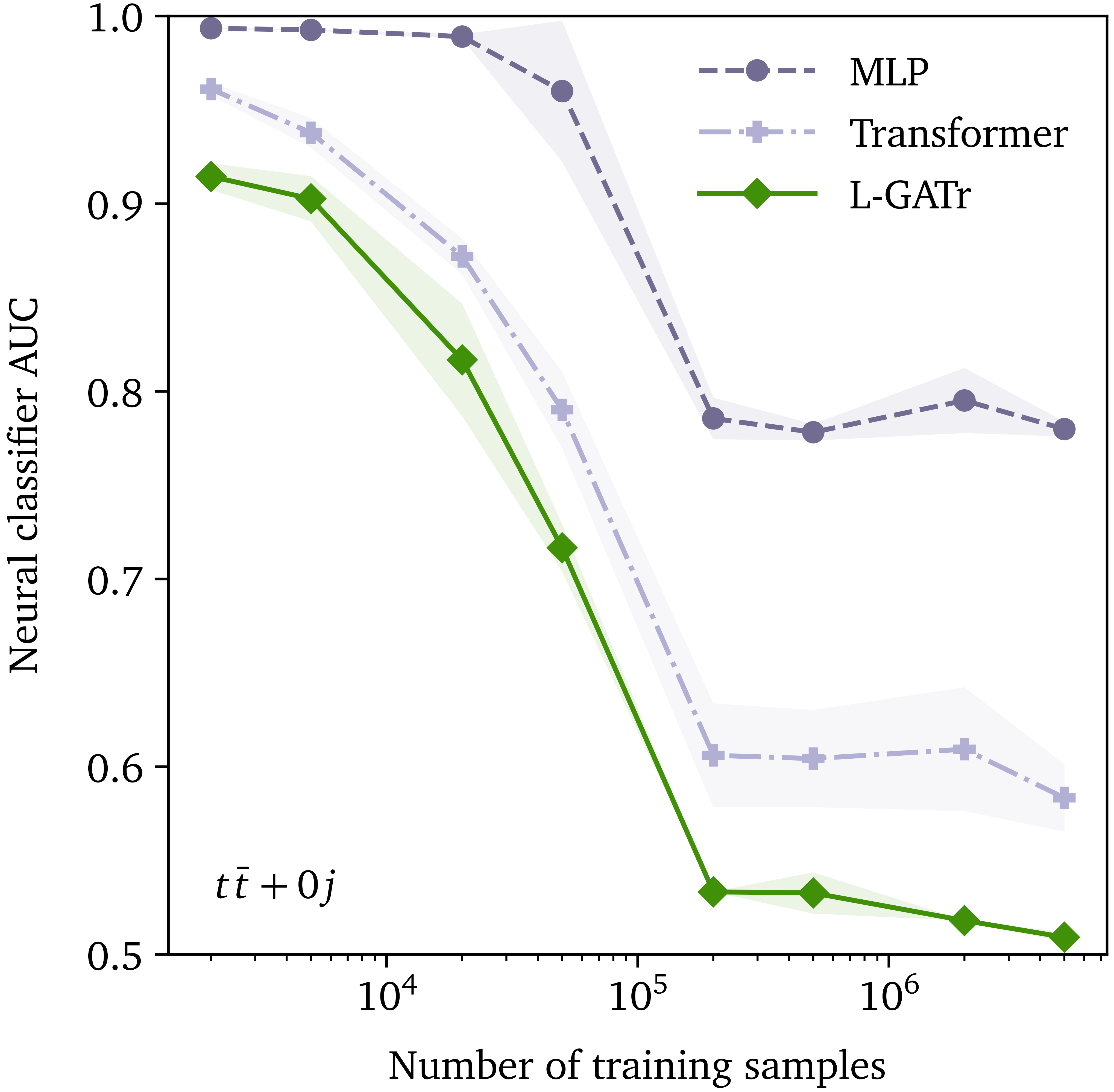


L-GATr helps with tricky kinematic features

Experiments

Event generation

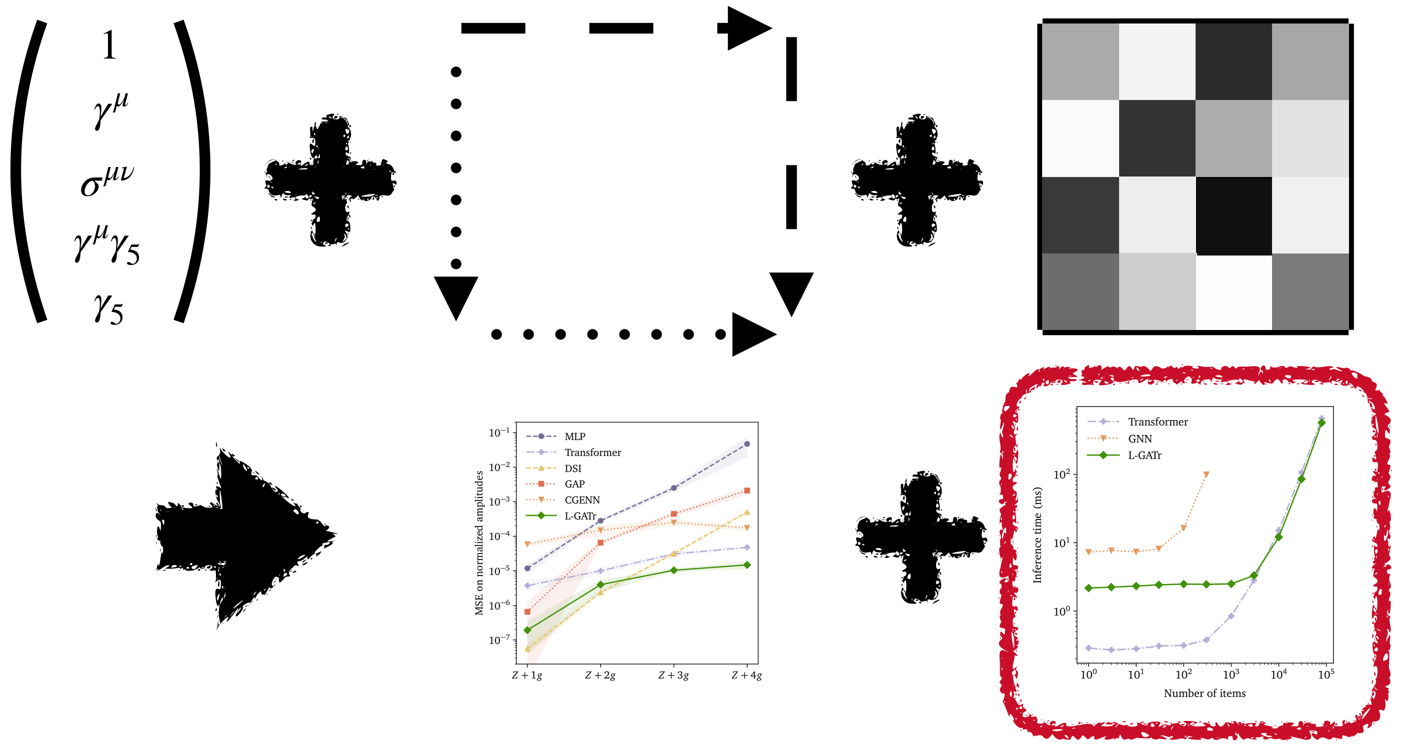


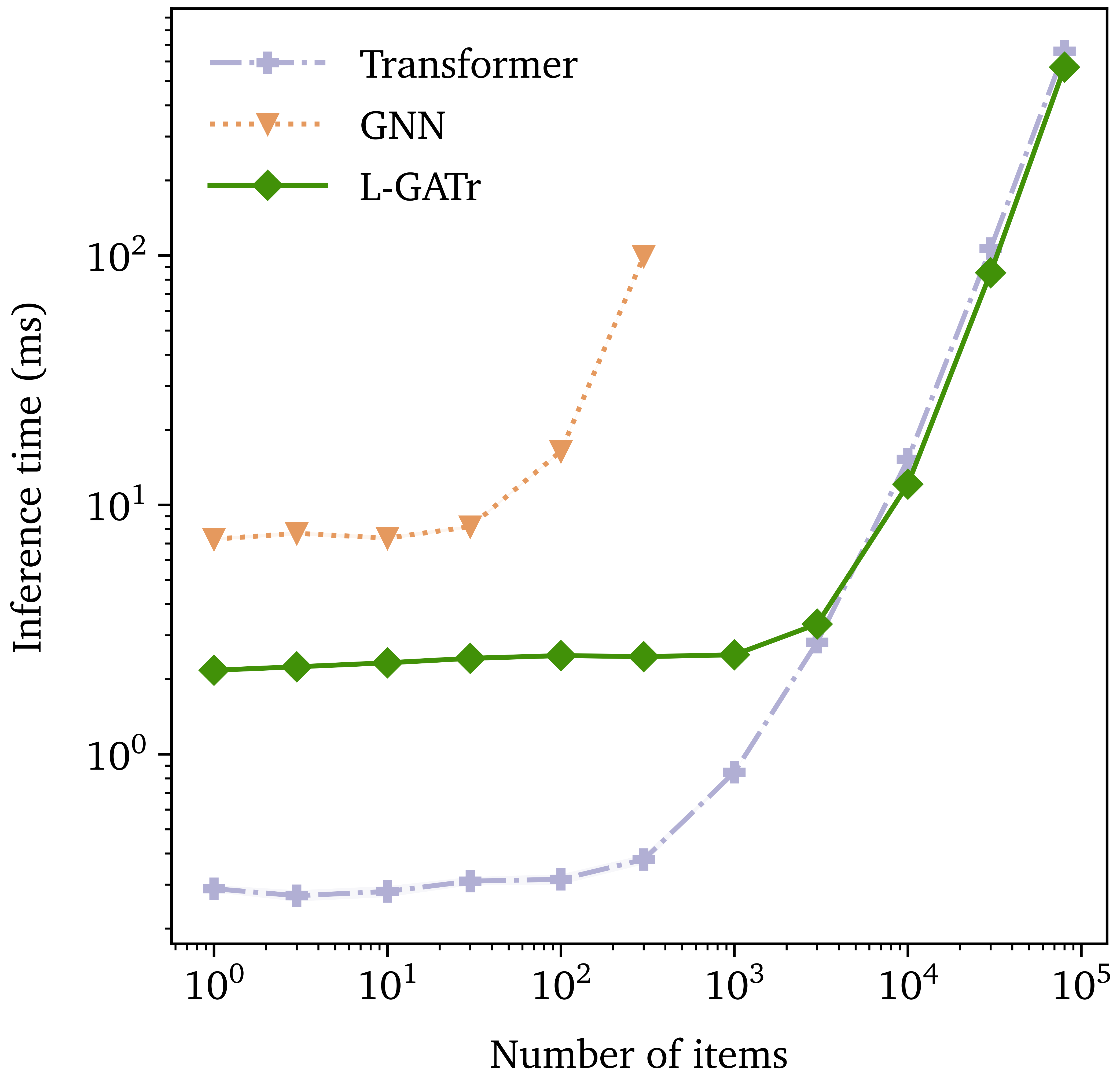


L-GATr generates samples that a classifier can almost not distinguish from the ground truth

Experiments

L-GATr can process thousands of particles

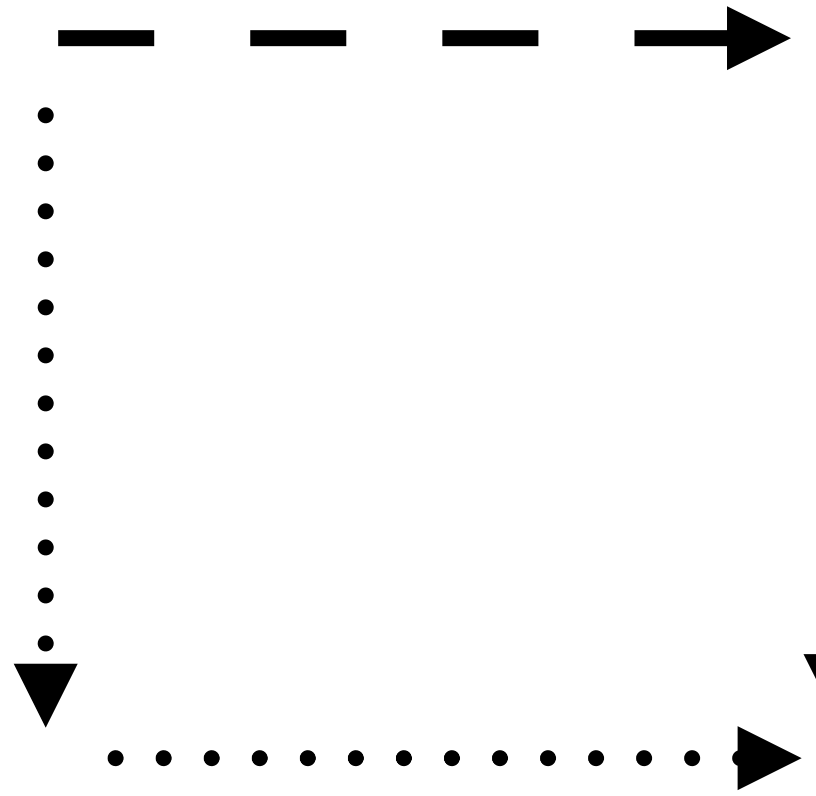




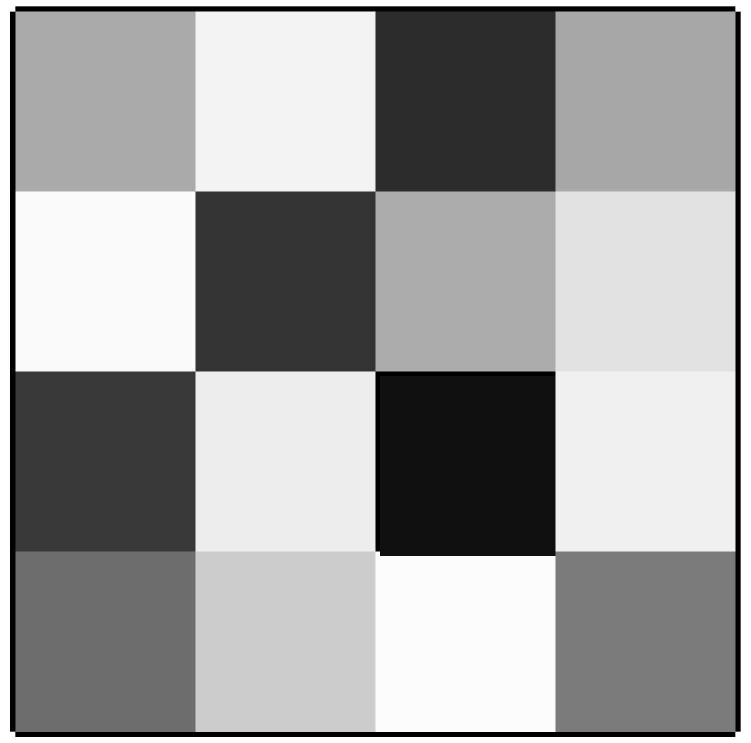
Transformers scale better than graph networks

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

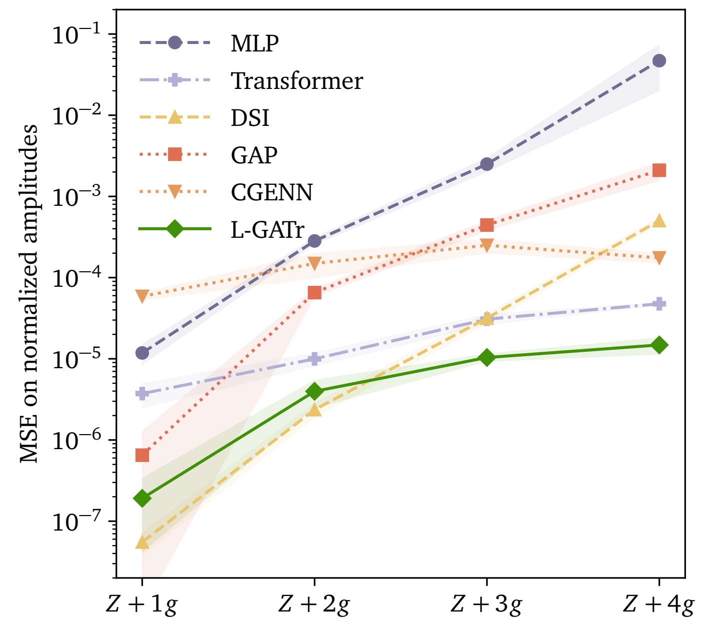
Geometric algebra
representations



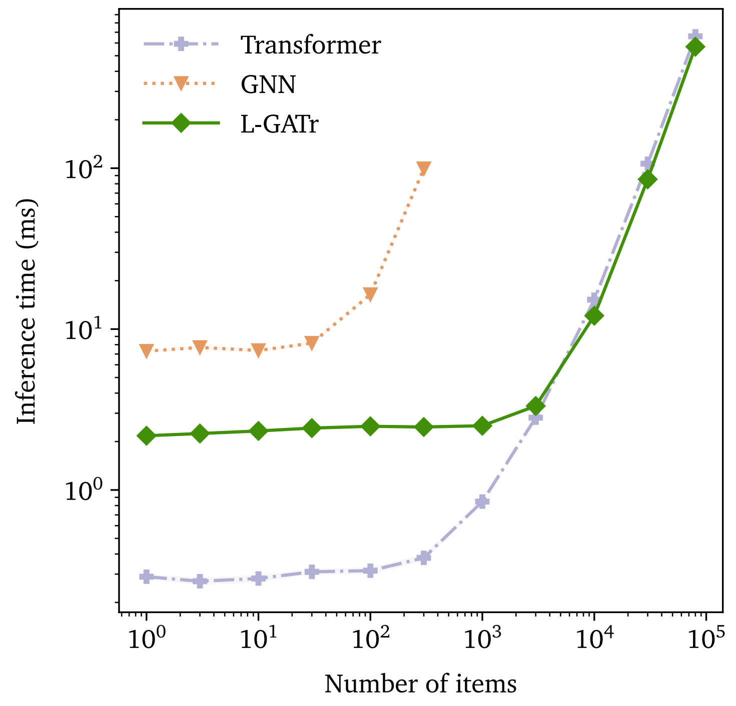
Equivariant
layers



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
NeurIPS 2024, arXiv:2405.14806

L-GATr paper

L-GATr code

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

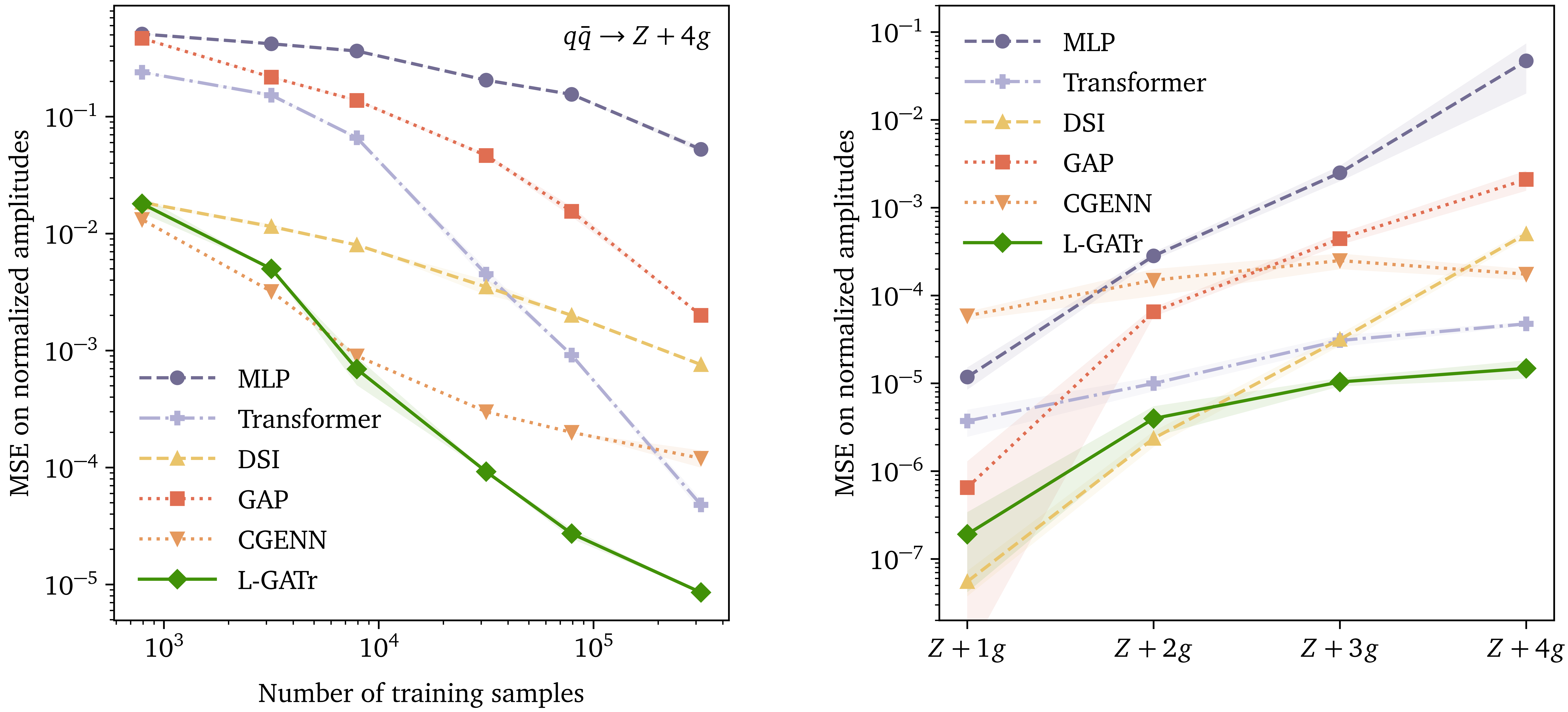
Bonus material

Ingredients

Equivariant layers

	Transformer	L-GATr
Linear(x)	$v\ x + c$	$\sum_{k=0}^{4} v_k \langle x \rangle_k + \sum_{k=0}^{4} w_k \gamma_5 \langle x \rangle_k$
Attention$(q,k,v)_{i\alpha}$	$\sum_{j,\beta} \text{Softmax}_j \left(\frac{q_{i\beta}, k_{j\beta}}{\sqrt{n}} \right) v_{j\alpha}$	$\sum_{j,\beta} \text{Softmax}_j \left(\frac{\langle q_{i\beta}, k_{j\beta} \rangle}{\sqrt{16n}} \right) v_{j\alpha}$
GP(x,y)	–	$x \cdot y$
LayerNorm(x)	$x \Big/ \sqrt{\frac{1}{n} \sum_{c=1}^{n} x_c^2 + \epsilon}$	$x \Big/ \sqrt{\frac{1}{n} \sum_{c=1}^{n} \sum_{k=0}^{4} \left\| \Big\langle \langle x_c \rangle_k, \langle x_c \rangle_k \Big\rangle \right\| + \epsilon}$
Act(x)	$\text{GELU}(x)$	$\text{GELU}(\langle x \rangle_0)\ x$

Amplitude regression



Experiments

Top tagging

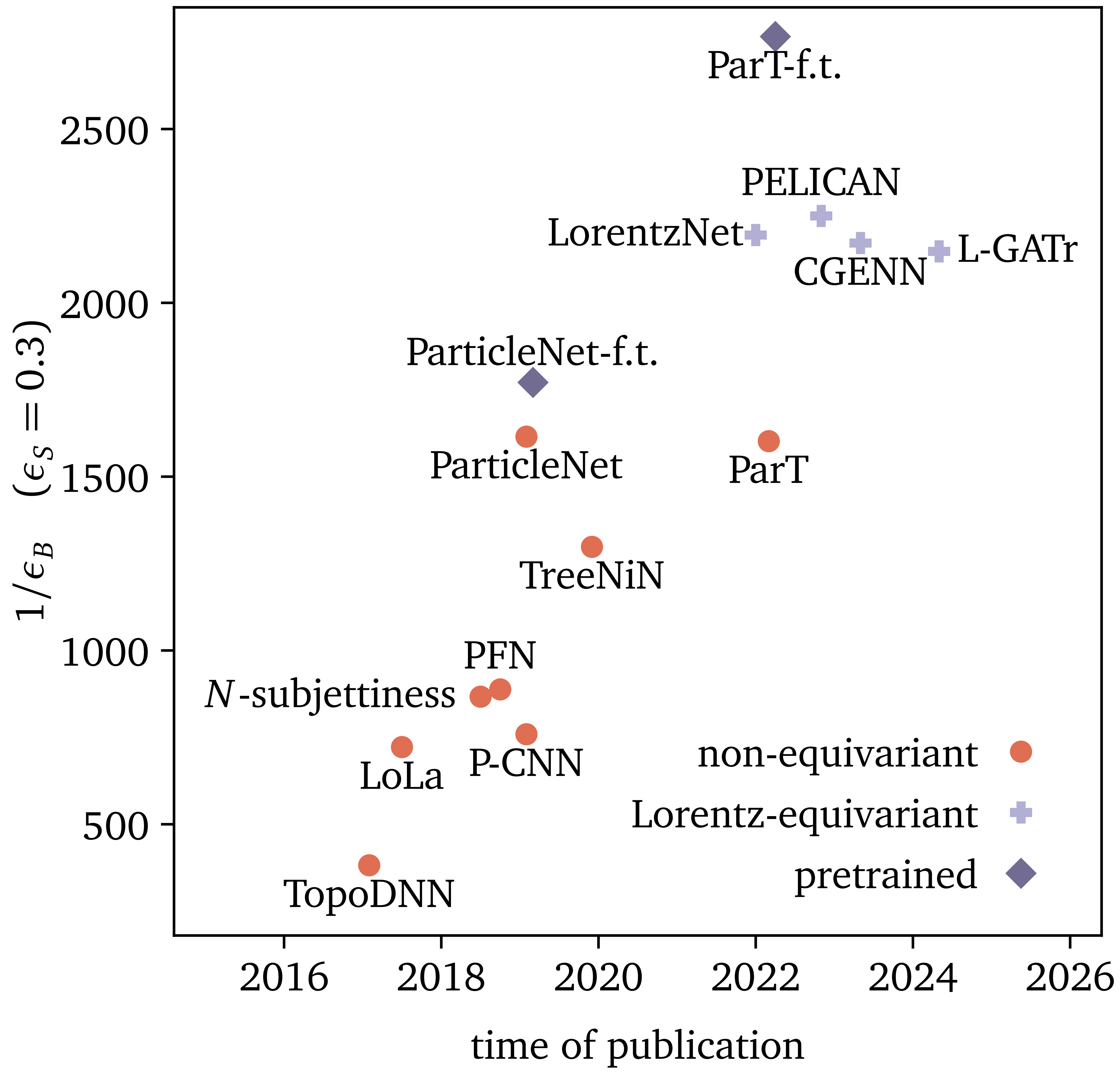
Model	Accuracy	AUC	$1/\epsilon_B$ ($\epsilon_S = 0.5$)	$1/\epsilon_B$ ($\epsilon_S = 0.3$)
TopoDNN [48]	0.916	0.972	–	295 ± 5
LoLa [15]	0.929	0.980	–	722 ± 17
P-CNN [1]	0.930	0.9803	201 ± 4	759 ± 24
N-subjettiness [61]	0.929	0.981	–	867 ± 15
PFN [50]	0.932	0.9819	247 ± 3	888 ± 17
TreeNiN [57]	0.933	0.982	–	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

Experiments

Top tagging

- New paradigm: **Transfer learning**
  Pretrain model on large dataset, then fine-tune on target dataset

- Transformers transfer better than graph networks



Experiments

Conditional Flow Matching

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 $u_t(x \mid x_1)$

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

Continuous normalising flows
arXiv:1806.07366

Conditional flow matching
arXiv:2210.02747

Experiments

Target velocities for CFM

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

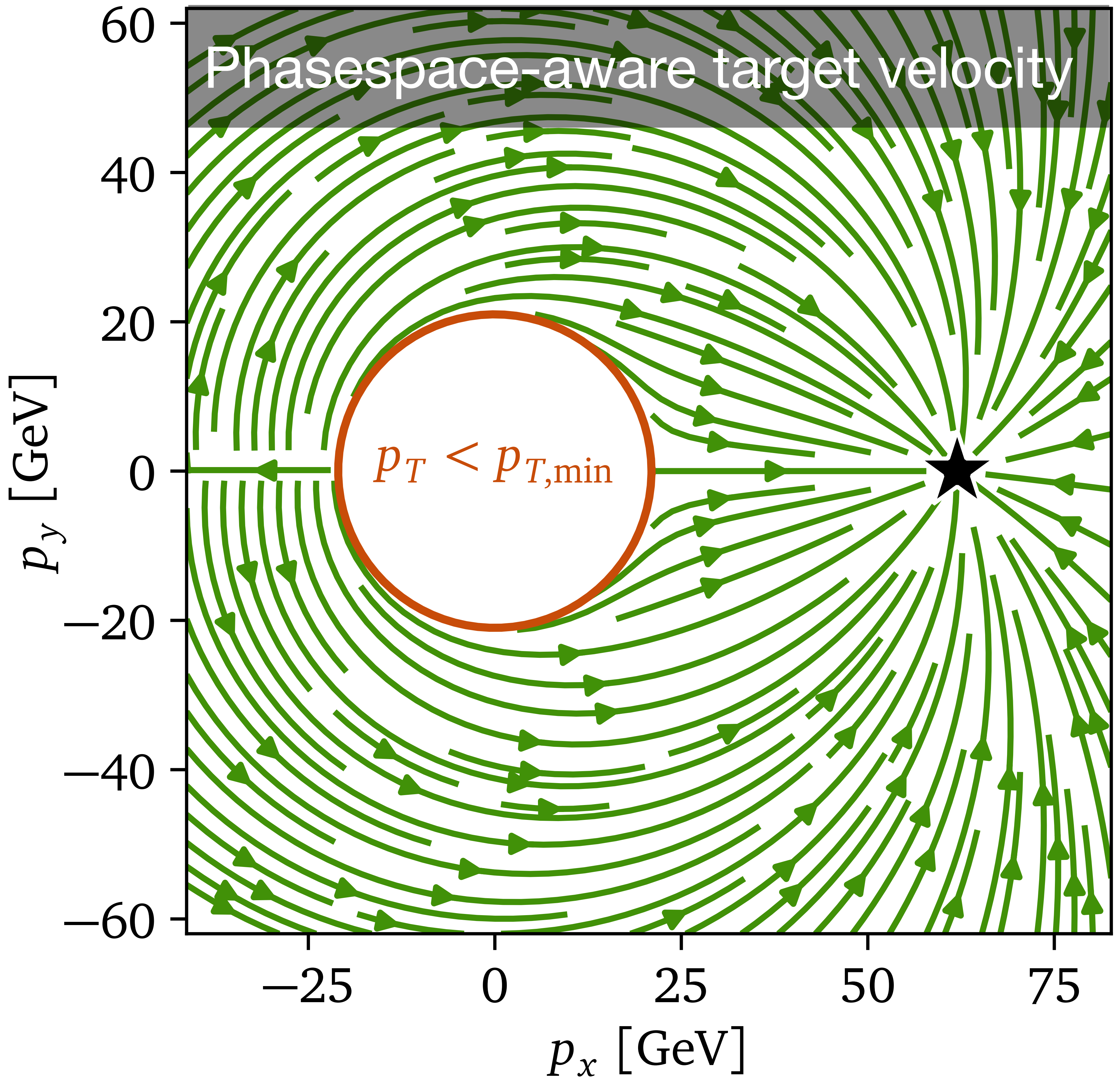
Riemannian Flow Matching
arXiv:2302.03660

Experiments

Target velocities for CFM

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

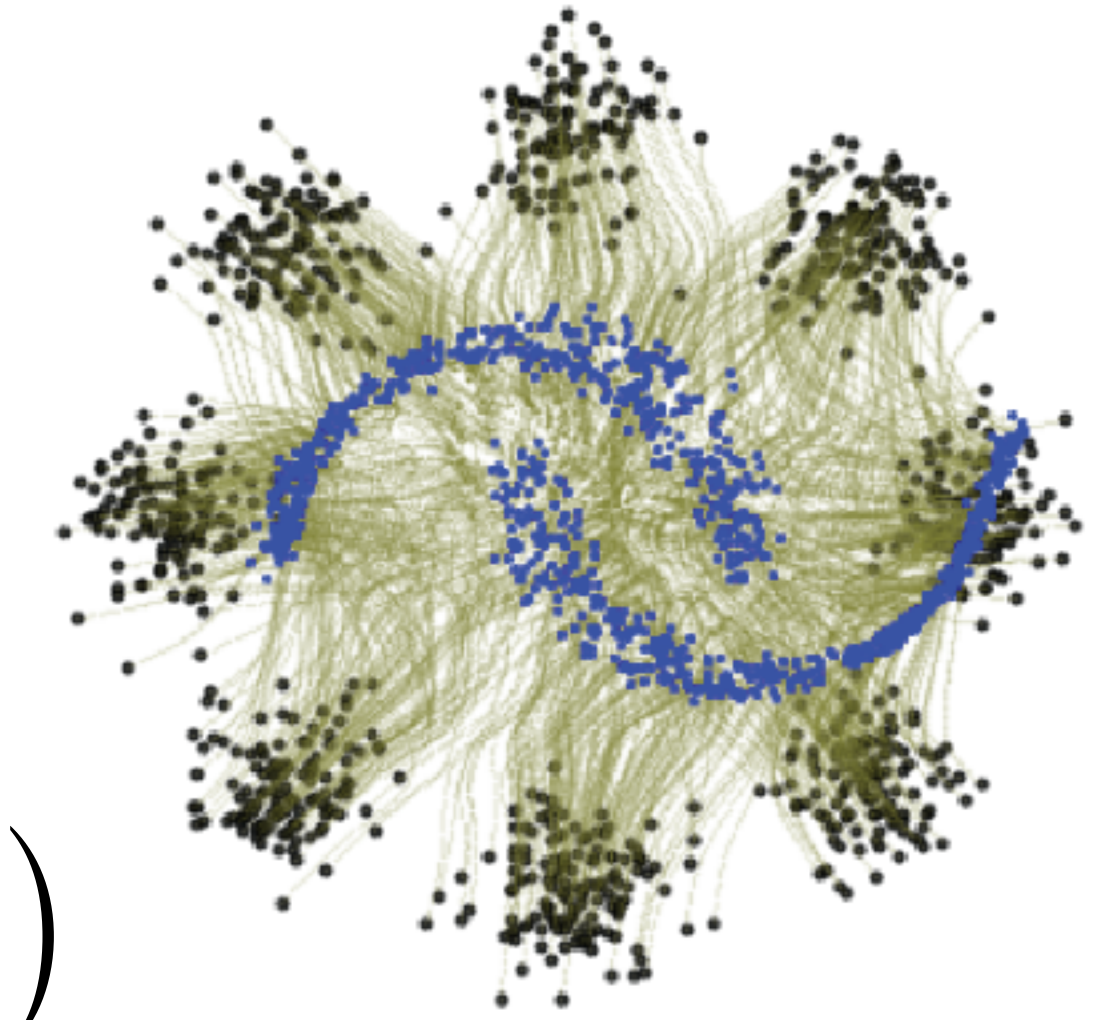
Target velocity	Architecture	AUC
Euclidean	L-GATr	0.99
Phasespace-aware	MLP	0.78
Phasespace-aware	L-GATr	**0.51**



Riemannian Flow Matching
arXiv:2302.03660

Event generation

Target velocities for CFM



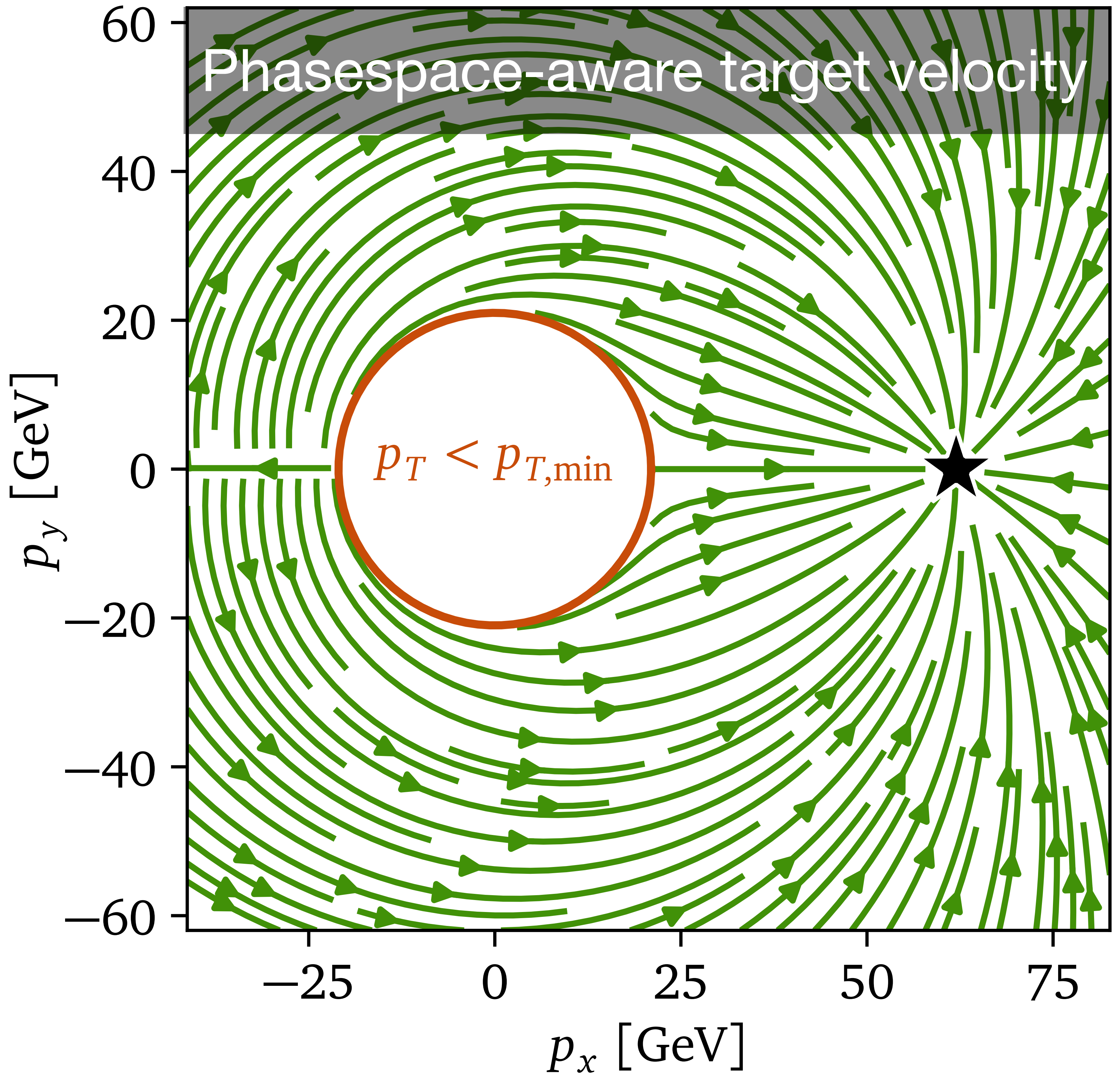
$$p = (E, p_x, p_y, p_z) = f(y) = \left(\sqrt{m^2 + p_T^2 \cosh^2 \eta}, p_T \cos \phi, p_T \sin \phi, p_T \sinh \eta\right)$$

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

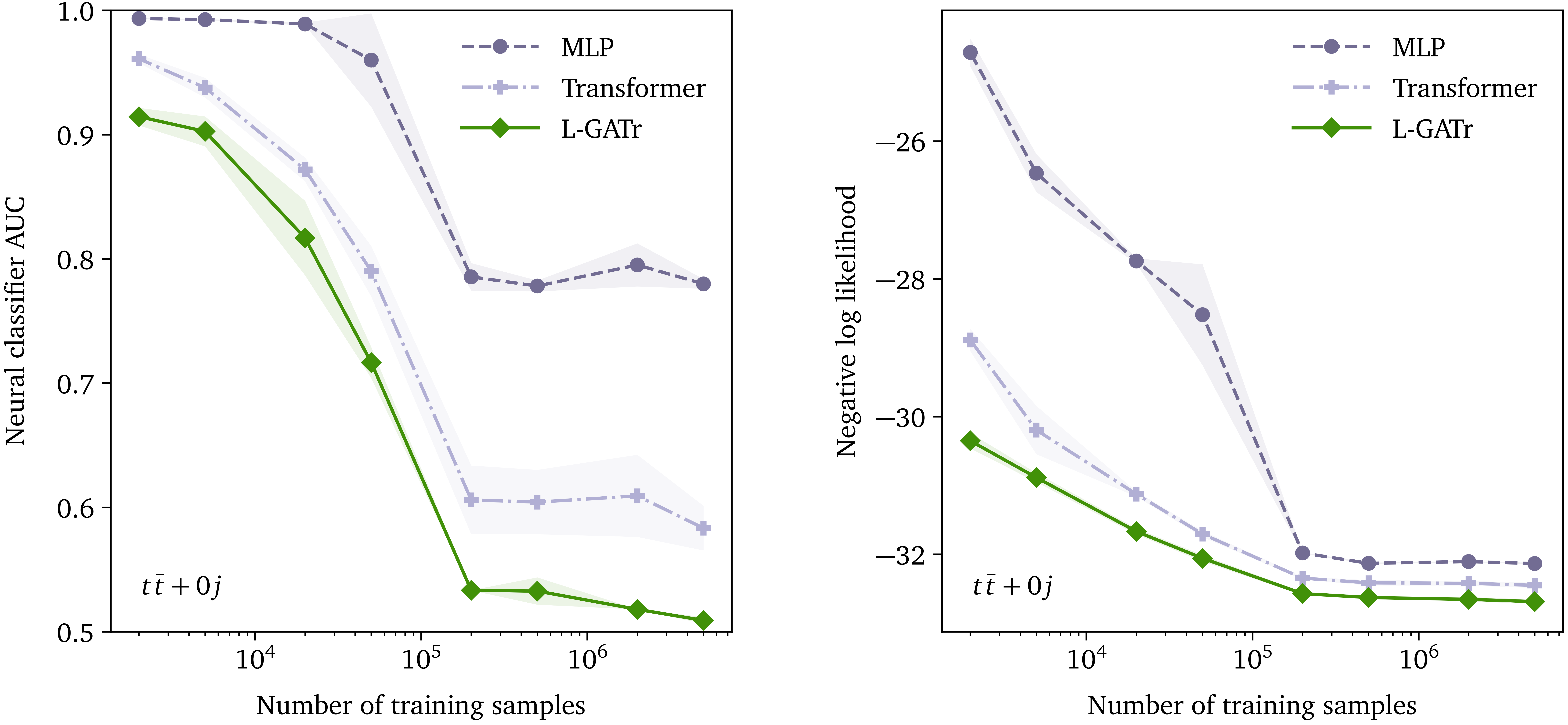
Target velocities can be

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

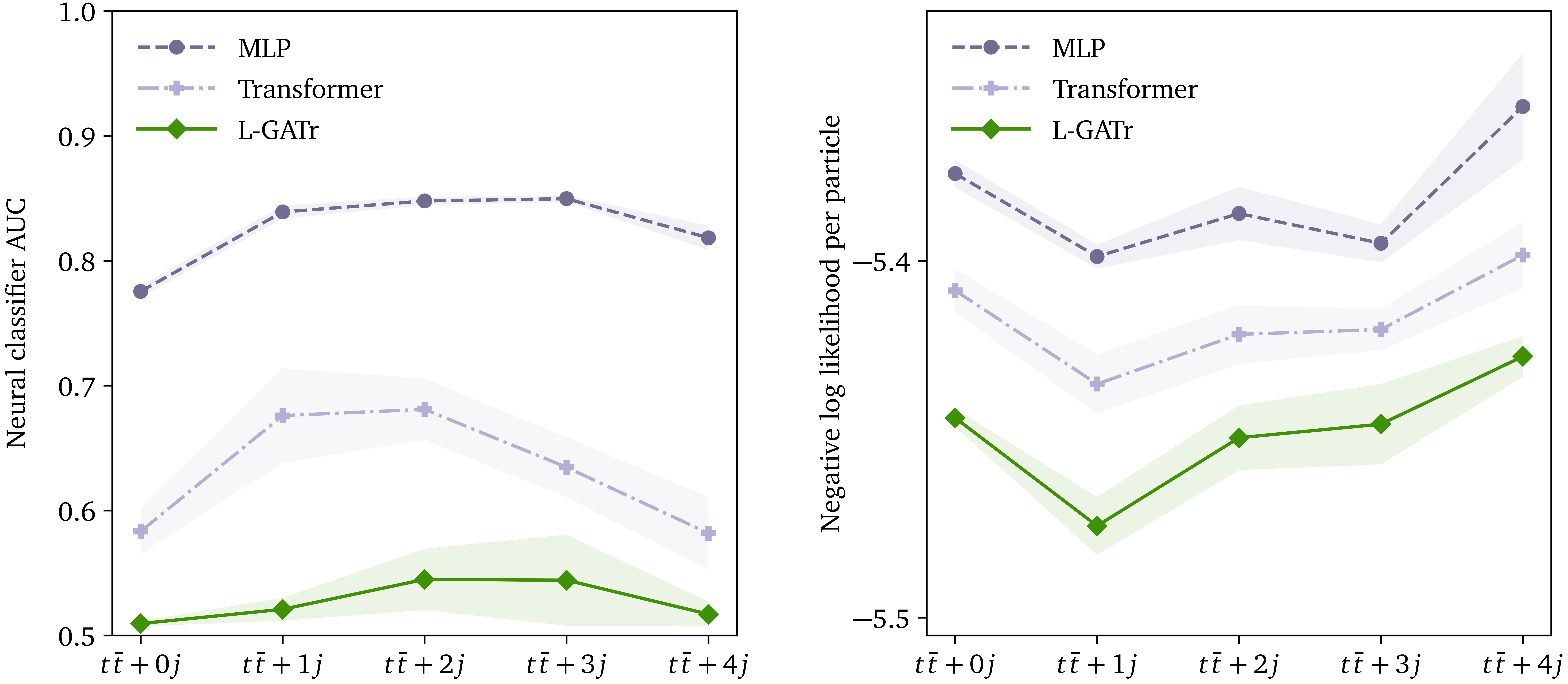
constant in $y = (y_m, y_p, \phi, \eta)$ ('phasespace-aware')



Event generation



Event generation



Symmetry breaking with spurious

Sources of symmetry breaking

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

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

We break the symmetry by adding the spurious as extra token or as extra channel for each token