Lorentz-GATr

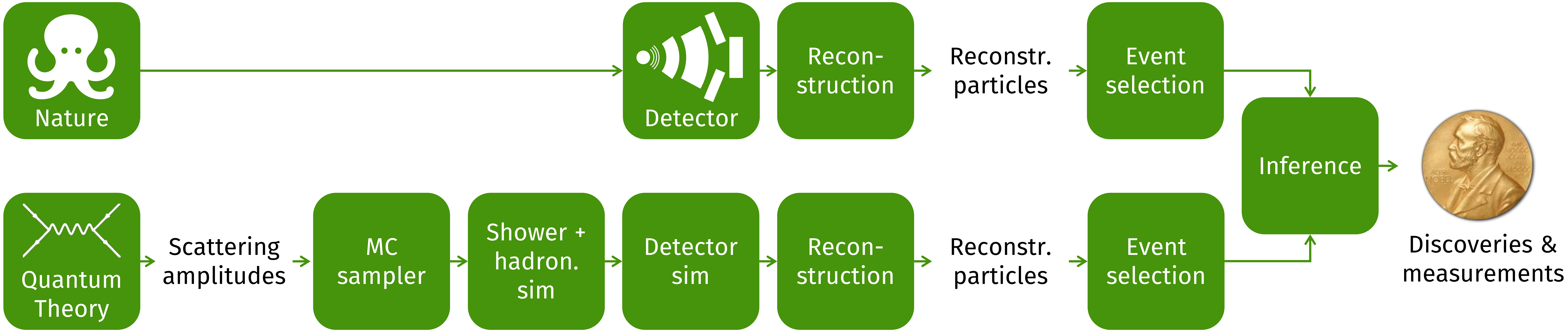
Event Generation with Lorentz-Equivariant Geometric Algebra Transformers

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

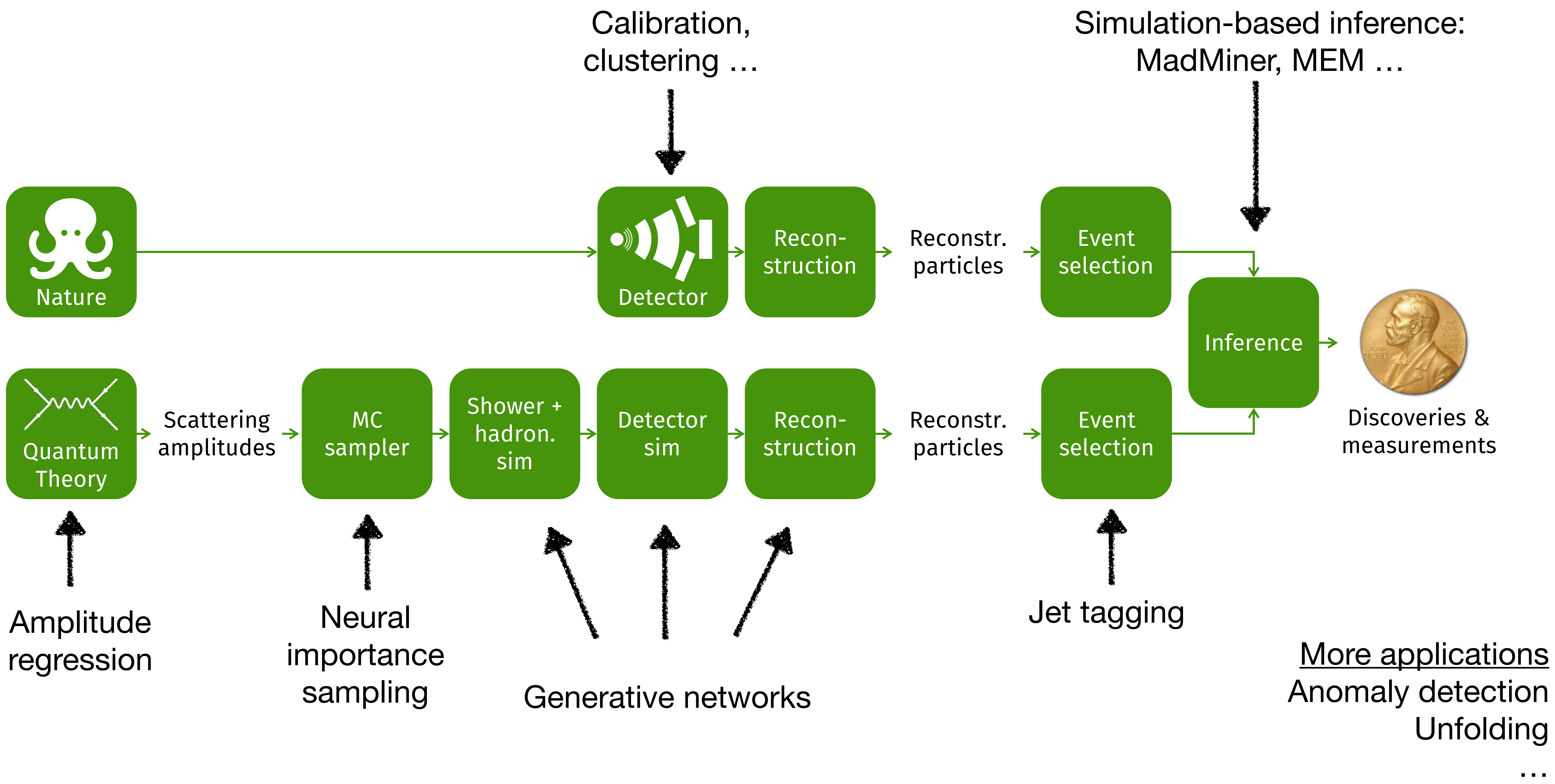
arXiv:2405.14806, arXiv:2411.00446

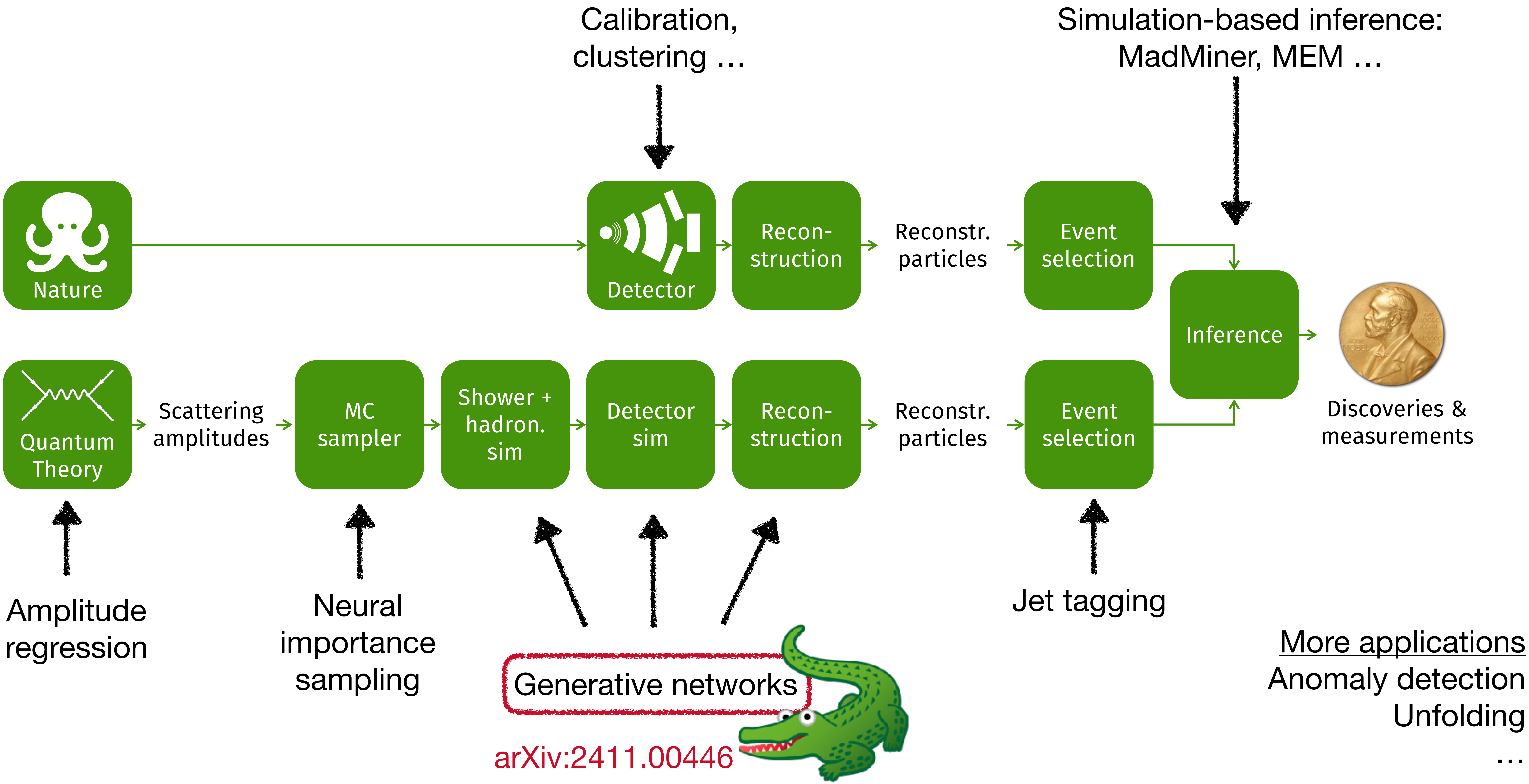
ML4Jets 2024
Paris

UNIVERSITÄT
HEIDELBERG
ZUKUNFT
SEIT 1386



Nature
Detector
Recon-
struction
Reconstr.
particles
Event
selection
Inference
Discoveries &
measurements
Quantum
Theory
Scattering
amplitudes
MC
sampler
Shower +
hadron.
sim
Detector
sim
Recon-
struction
Reconstr.
particles
Event
selection


Calibration,
clustering …
Simulation-based inference:
MadMiner, MEM …
Nature
Detector
Recon-
struction
Reconstr.
particles
Event
selection
Inference
Discoveries &
measurements
Quantum
Theory
Scattering
amplitudes
MC
sampler
Shower +
hadron.
sim
Detector
sim
Recon-
struction
Reconstr.
particles
Event
selection
Amplitude
regression
Neural
importance
sampling
Generative networks
Jet tagging
More applications
Anomaly detection
Unfolding
…



Calibration, clustering …
Simulation-based inference: MadMiner, MEM …
Nature
Detector
Recon-struction
Reconstr. particles
Event selection
Inference
Quantum Theory
Scattering amplitudes
MC sampler
Shower + hadron. sim
Detector sim
Recon-struction
Reconstr. particles
Event selection
Discoveries & measurements
Amplitude regression
Neural importance sampling
Generative networks
arXiv:2411.00446
Jet tagging
More applications
Anomaly detection
Unfolding
…


Calibration, clustering …
Giovanni's talk
arXiv:2410.07315
Simulation-based inference: MadMiner, MEM …
Nature
Detector
Recon-struction
Reconstr. particles
Event selection
Inference
Discoveries & measurements
Quantum Theory
Scattering amplitudes
MC sampler
Shower + hadron. sim
Detector sim
Recon-struction
Reconstr. particles
Event selection
Amplitude regression
arXiv:2411.00446
Neural importance sampling
Generative networks
arXiv:2411.00446
Jet tagging
Víctor's talk
arXiv:2411.00446
More applications
Anomaly detection
Unfolding
…

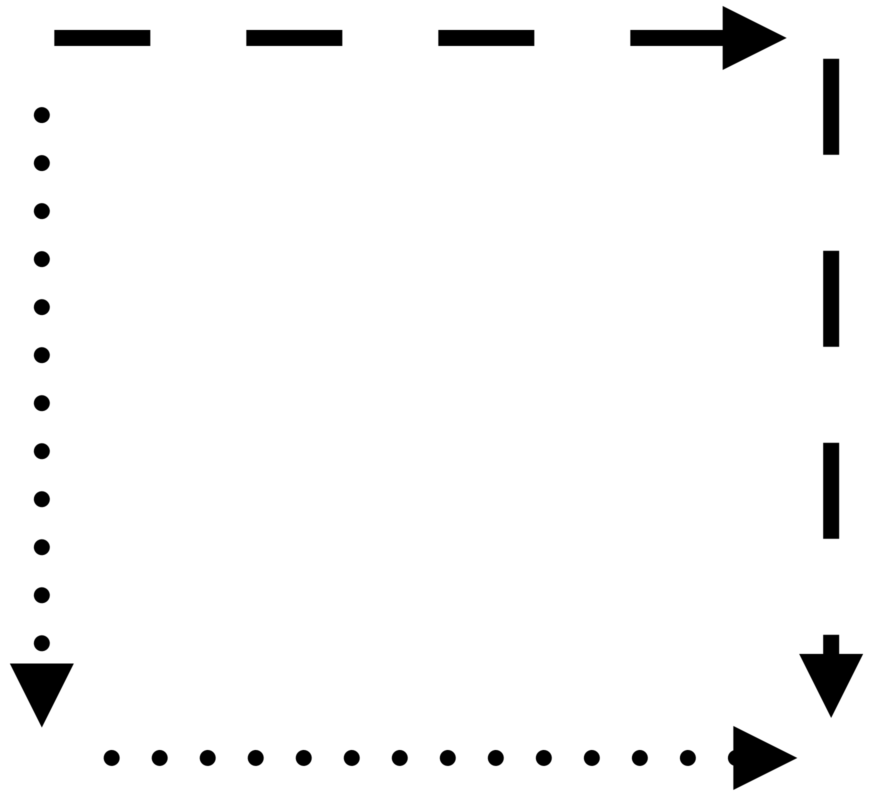
Lorentz-Equivariant
Geometric **A**lgebra
Transformer

=

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

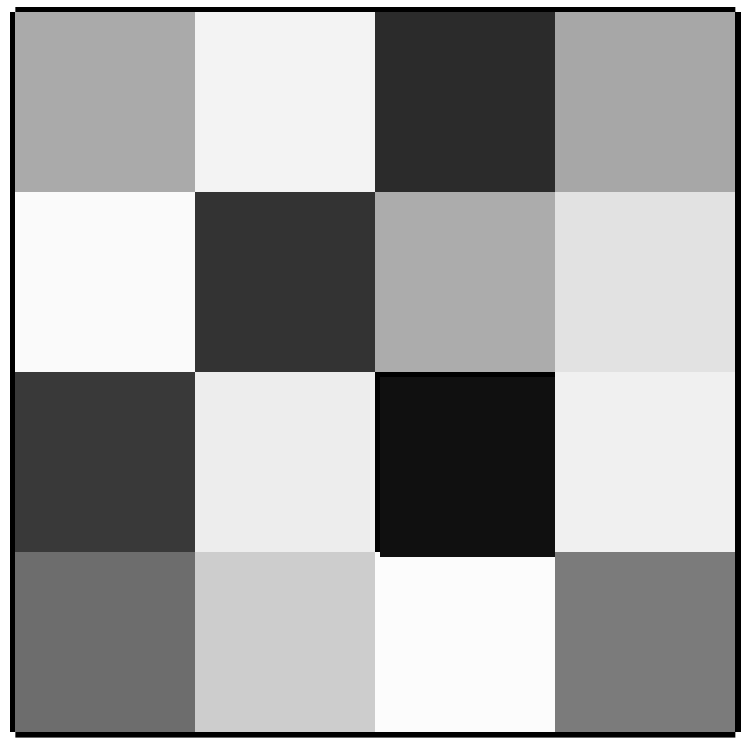
Geometric algebra
representations

+



Lorentz-Equivariant
layers

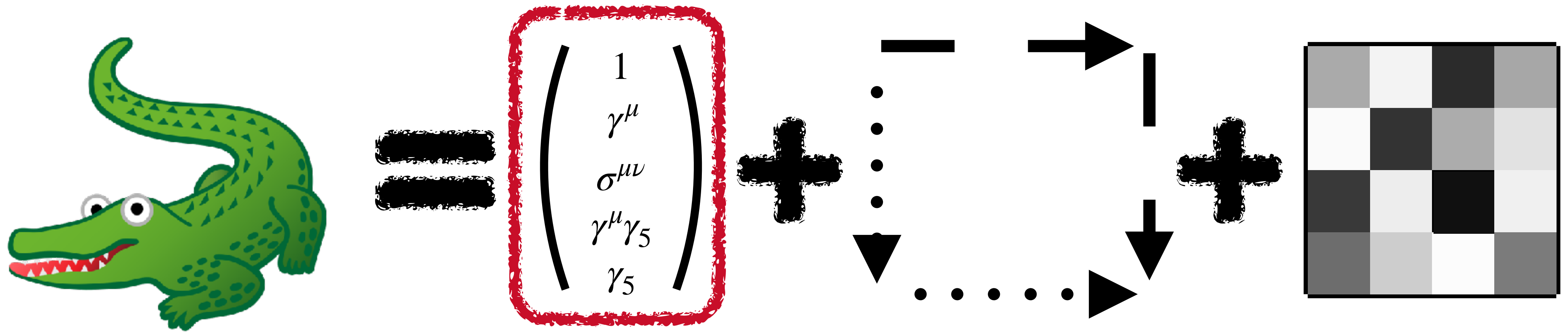
+



Transformer
architecture

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

L-GATr

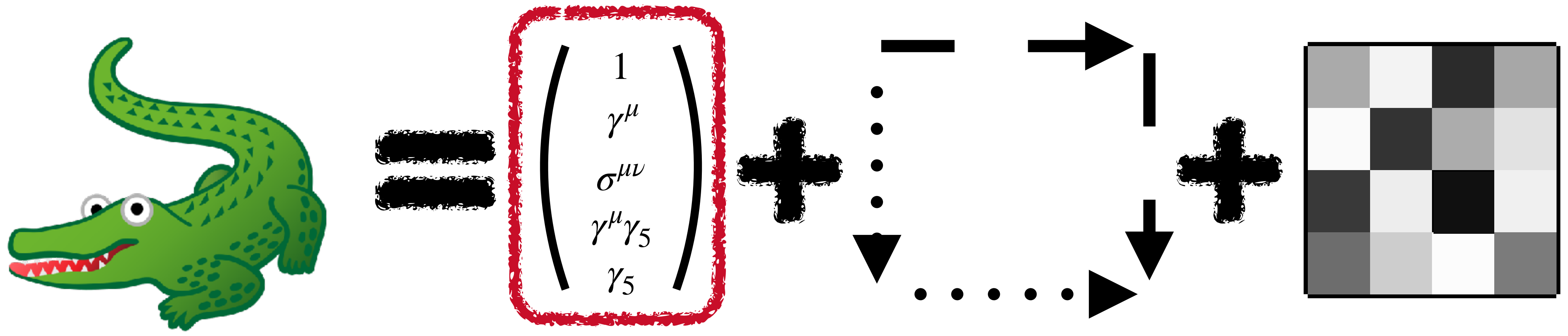


Geometric algebra = Clifford algebra

Geometric algebra = Vector space + geometric product $xy = \frac{\{x, y\}}{2} + \frac{[x, y]}{2}$

- Symmetric part $\{x, y\}$: scalar/inner product
- Antisymmetric part $[x, y]$: outer product (yields higher-order objects)

L-GATr



Geometric algebra = Clifford algebra

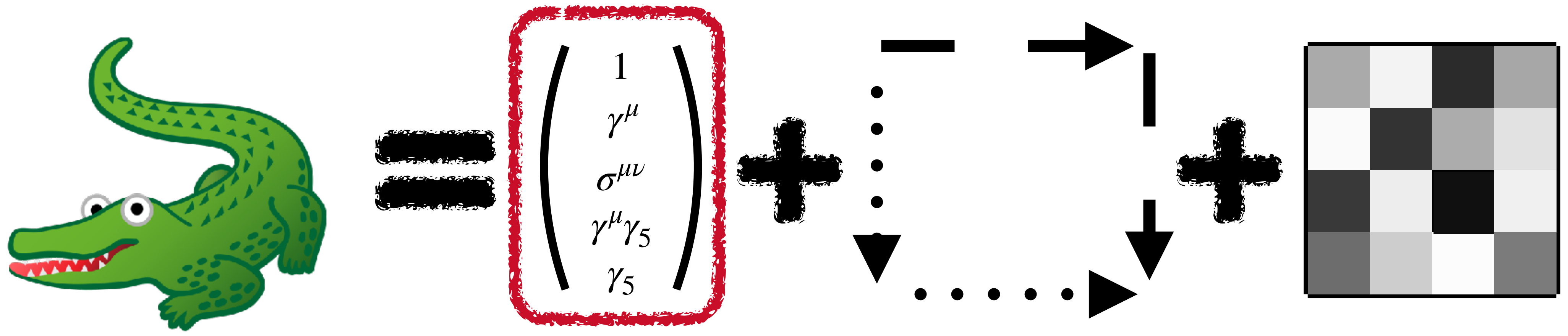
Geometric algebra = Vector space + geometric product $xy = \frac{\{x, y\}}{2} + \frac{[x, y]}{2}$

- Symmetric part $\{x, y\}$: scalar/inner product
- Antisymmetric part $[x, y]$: outer product (yields higher-order objects)

Spacetime geometric algebra: Geometric algebra over vector space $\mathbb{R}^4$ with Minkowski metric $g = \mathrm{diag}(1, -1, -1, -1)$

- Basis elements γ^μ are orthonormal: $\{\gamma^\mu, \gamma^\nu\} = 2g^{\mu\nu}$
- Dirac algebra is the same up to $\mathbb{R} \to \mathbb{C}$

L-GATr



Building multivectors

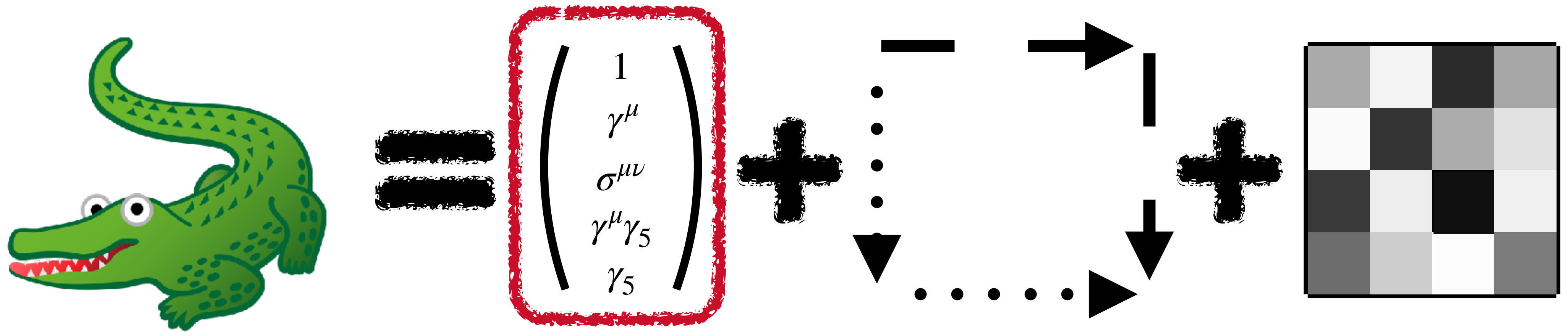
- Scalar and vectors $1,\ \gamma^\mu$ (1+4 objects)
- Product of two vectors: $\gamma^\mu\gamma^\nu = \frac{\{\gamma^\mu,\gamma^\nu\}}{2} + \frac{[\gamma^\mu,\gamma^\nu]}{2} = g^{\mu\nu} + \sigma^{\mu\nu}$ (6 new objects)
- Axial vector: $\epsilon_{\mu\nu\rho\sigma}\gamma^\nu\gamma^\rho\gamma^\sigma$ (4 new objects)
- Pseudoscalar: $\gamma^5 = \gamma^0\gamma^1\gamma^2\gamma^3 = \frac{1}{4!}\epsilon_{\mu\nu\rho\sigma}\gamma^\mu\gamma^\nu\gamma^\rho\gamma^\sigma$ (1 new object)

Multivector: $x = x^S 1 + x^V_\mu \gamma^\mu + x^B_{\mu\nu}\sigma^{\mu\nu} + x^A_\mu \gamma^\mu\gamma^5 + x^P\gamma^5$ with $(x^S, x^V_\mu, x^B_{\mu\nu}, x^A_\mu, x^P) \in \mathbb{R}^{16}$

Particle: $x^V = (E, p_x, p_y, p_z), \quad x^S = \mathrm{PID}$

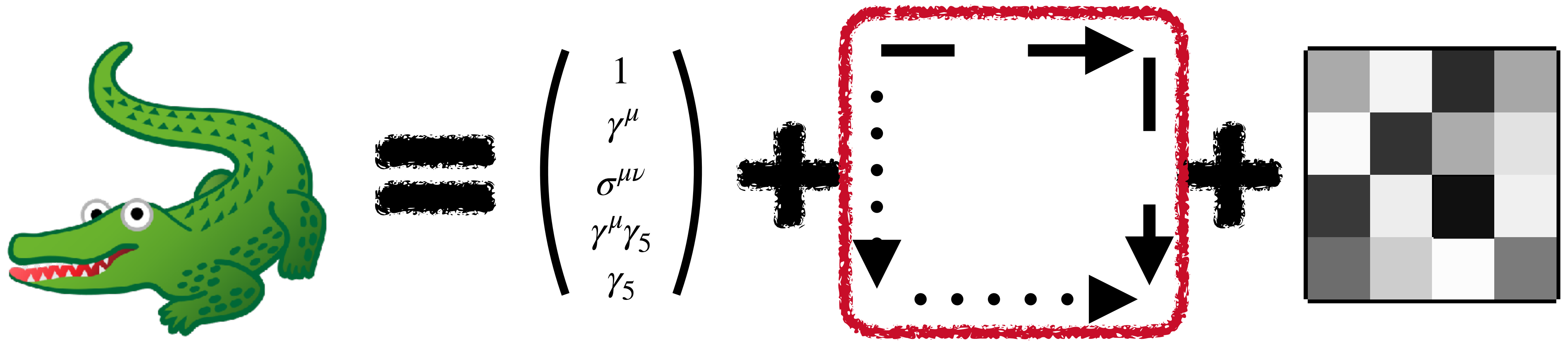
L-GATr

Geometric algebra representations



$$x^S \in \mathbb{R} \quad \Rightarrow \quad \begin{pmatrix} x^S \\ x^V_\mu \\ x^B_{\mu\nu} \\ x^A_\mu \\ x^P \end{pmatrix} \in \mathbb{R}^{16}$$

scalars — multivectors

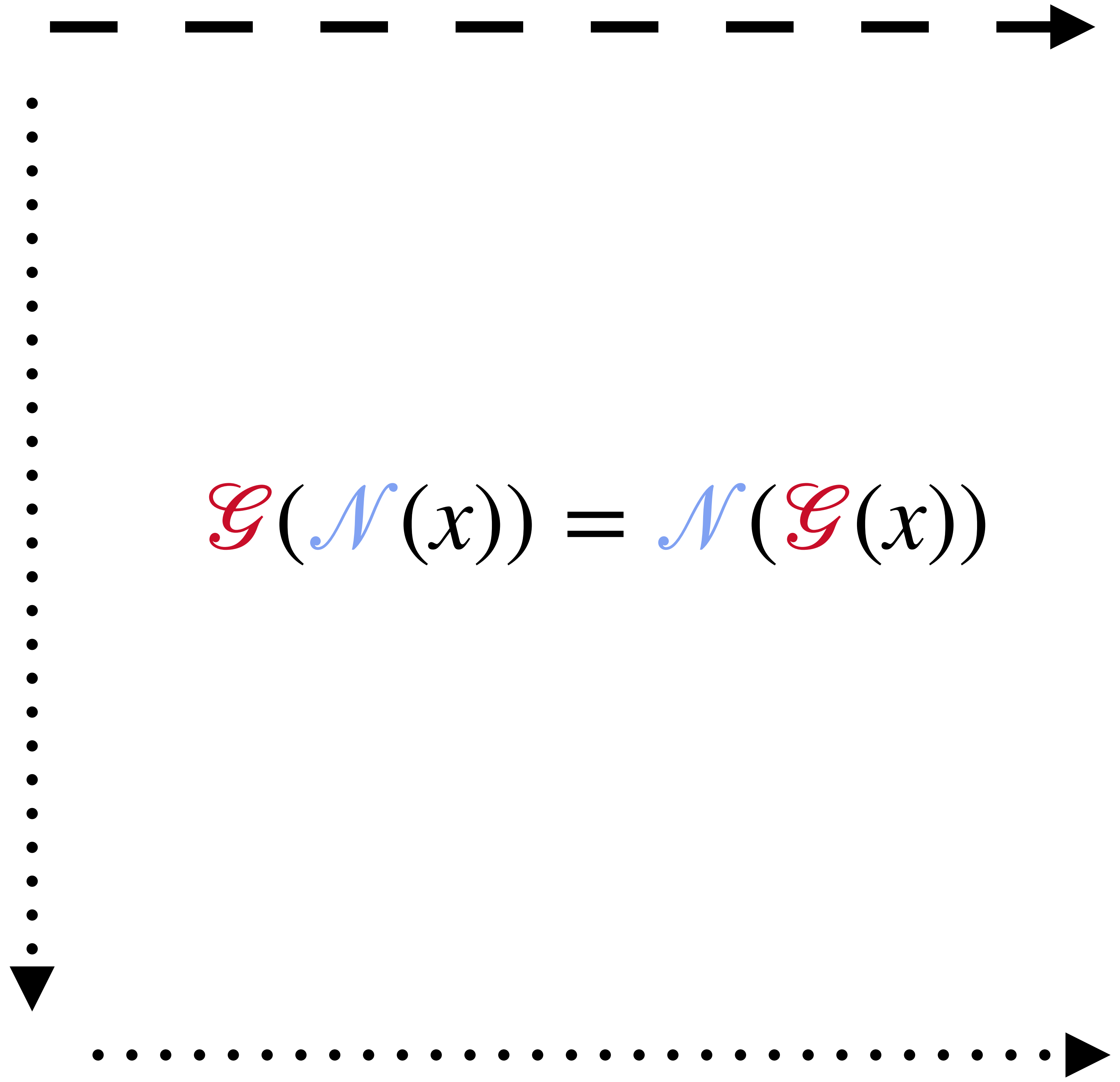


L-GATr

Lorentz Equivariance

Lorentz transformation $\mathscr{G}$

neural network transformation $\mathscr{N}$



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

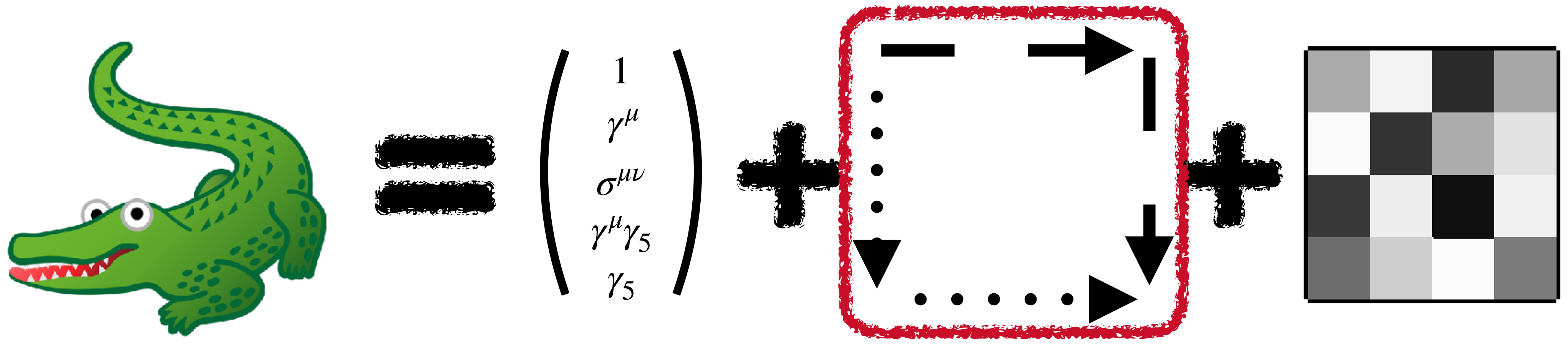
$\mathscr{N}$

$\mathscr{G}$

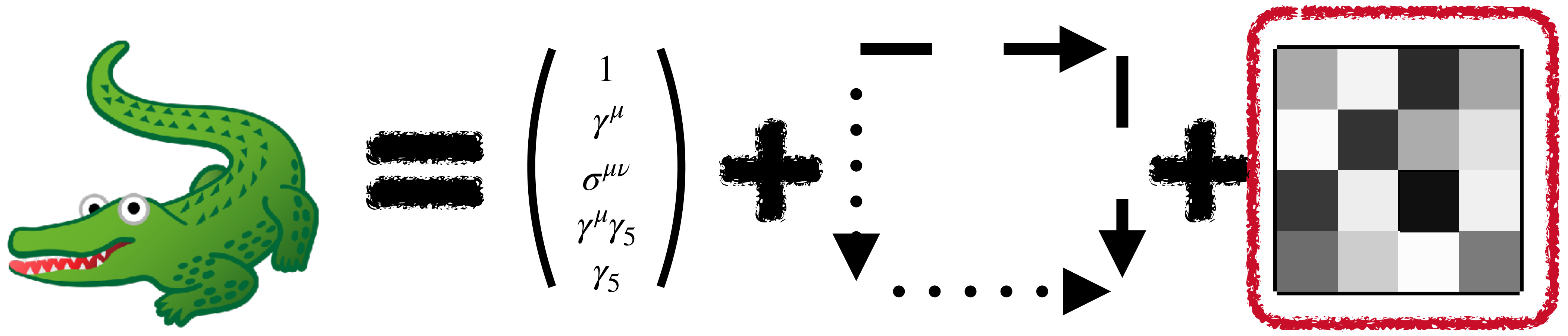
Equivariance = Covariance

L-GATr

GATr-ing all transformer layers

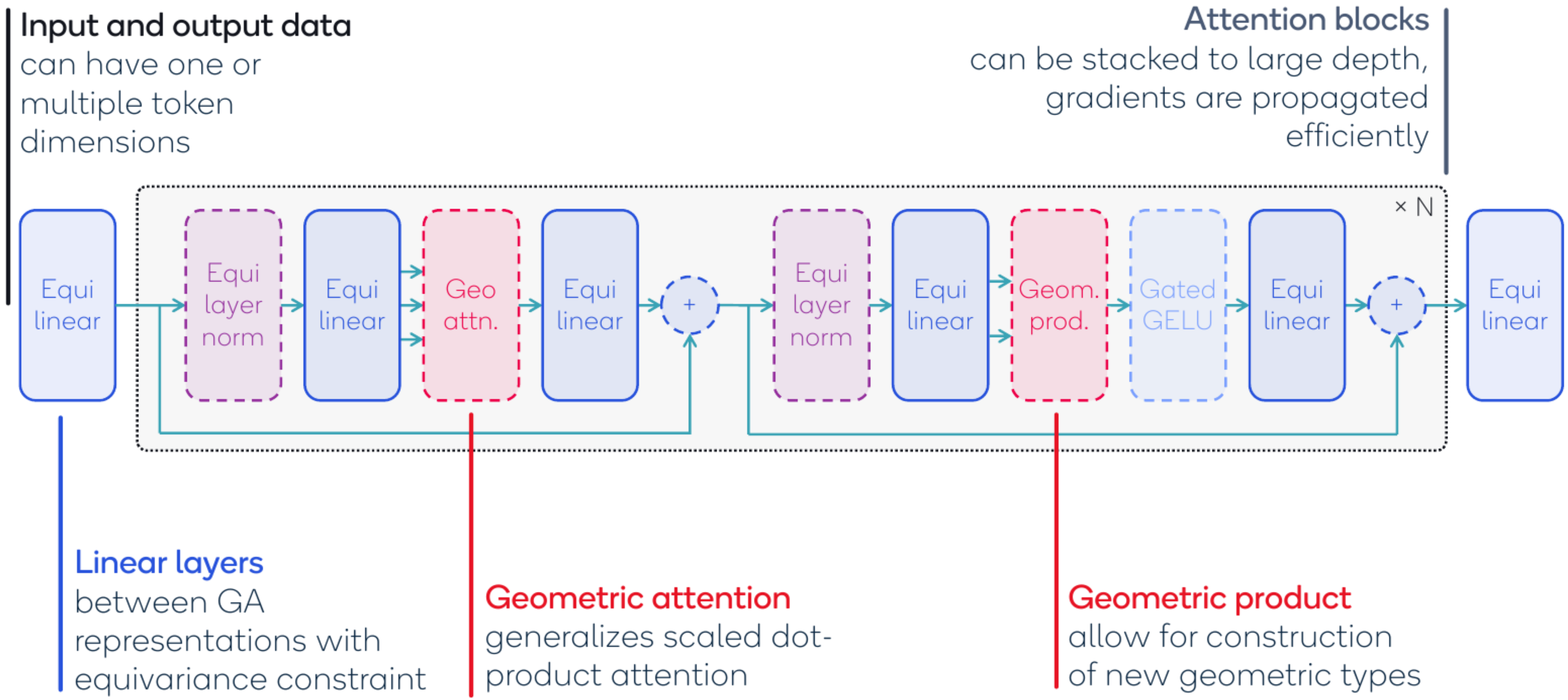


Layer type	Transformer	L-GATr
$\mathrm{Linear}(x)$	$vx + w$	$\sum_{k=0}^{4} v_k \langle x \rangle_k \left(+ \sum_{k=0}^{4} w_k \gamma^5 \langle x \rangle_k \right)$
$\mathrm{Attention}(q,k,v)_{ic}$	$\sum_{j=1}^{n_t} \mathrm{Softmax}_j \left(\sum_{c'=1}^{n_c} \frac{q_{ic'} k_{jc'}}{\sqrt{n_c}} \right) v_{jc}$	$\sum_{j=1}^{n_t} \mathrm{Softmax}_j \left(\sum_{c'=1}^{n_c} \frac{\langle q_{ic'}, k_{jc'} \rangle}{\sqrt{16 n_c}} \right) v_{jc}$
$\mathrm{LayerNorm}(x)$	$x \left[\frac{1}{n_c} \sum_{c=1}^{n_c} x_c^2 + \epsilon \right]^{-1/2}$	$x \left[\frac{1}{n_c} \sum_{c=1}^{n_c} \sum_{k=0}^{4} \left| \left\langle \langle x_c \rangle_k, \langle x_c \rangle_k \right\rangle \right| + \epsilon \right]^{-1/2}$
$\mathrm{Activation}(x)$	$\mathrm{GELU}(x)$	$\mathrm{GELU}(\langle x \rangle_0) x$
$\mathrm{GP}(x,y)$	—	xy



L-GATr

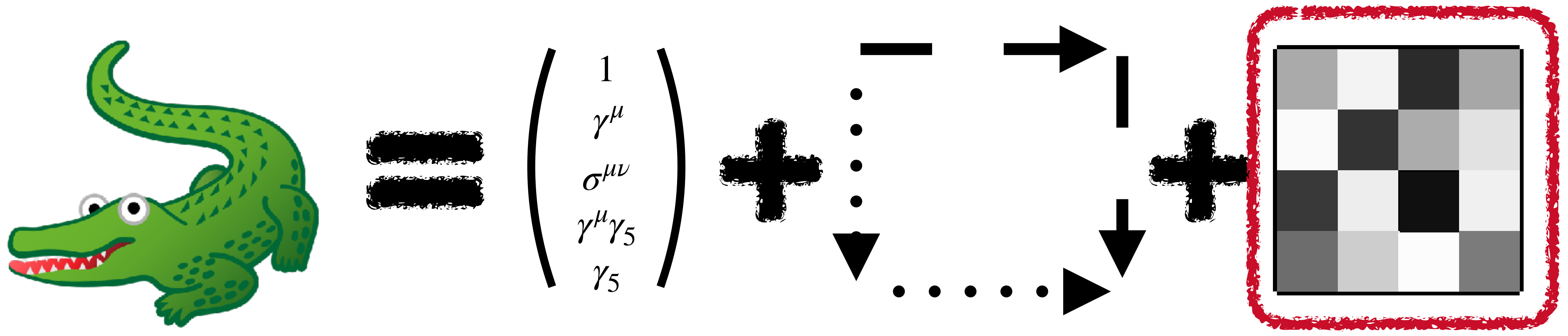
Full architecture

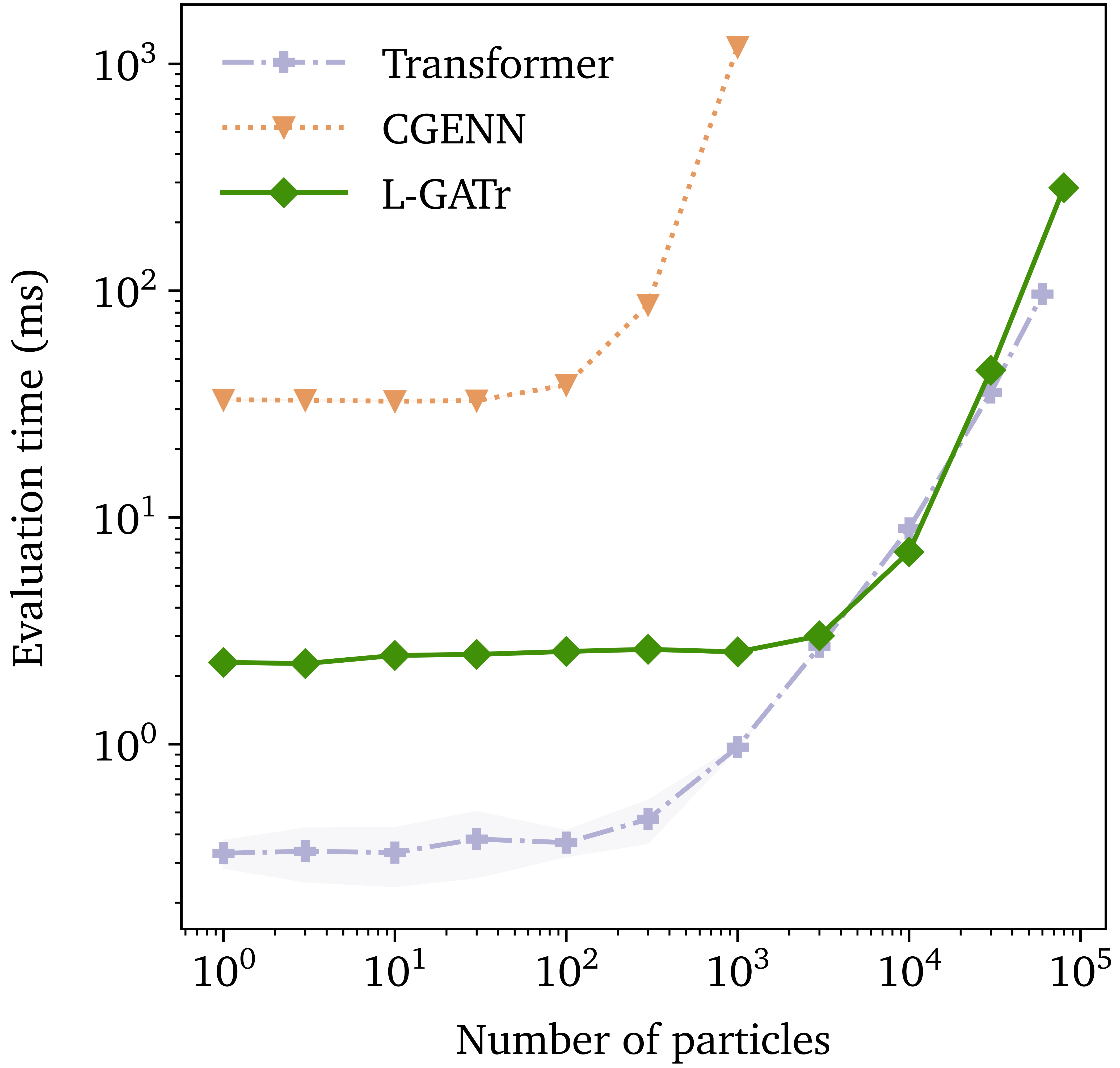


Credits to Johann Brehmer

L-GATr

Processing thousands of particles





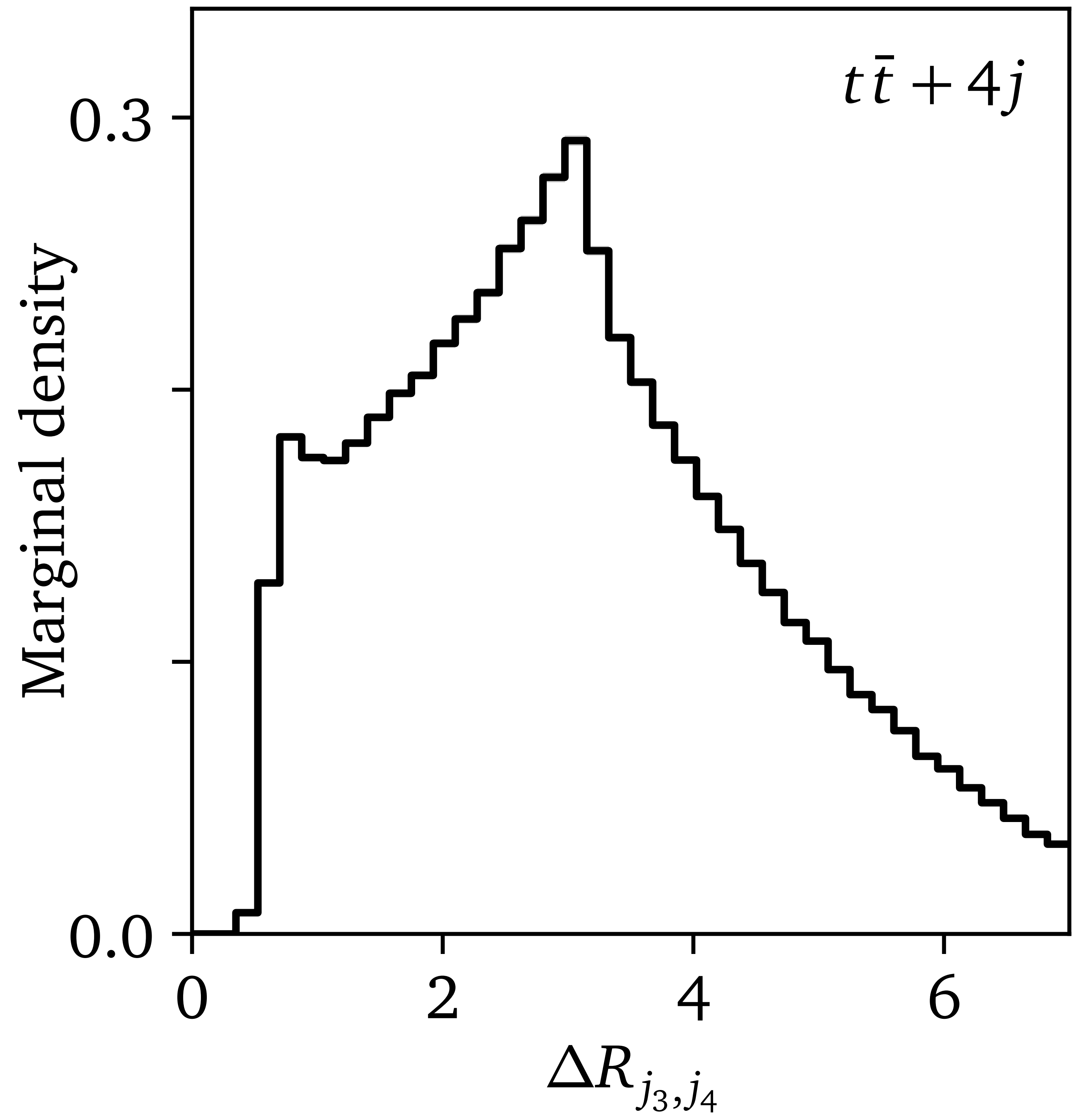
Transformers scale
better than graph networks

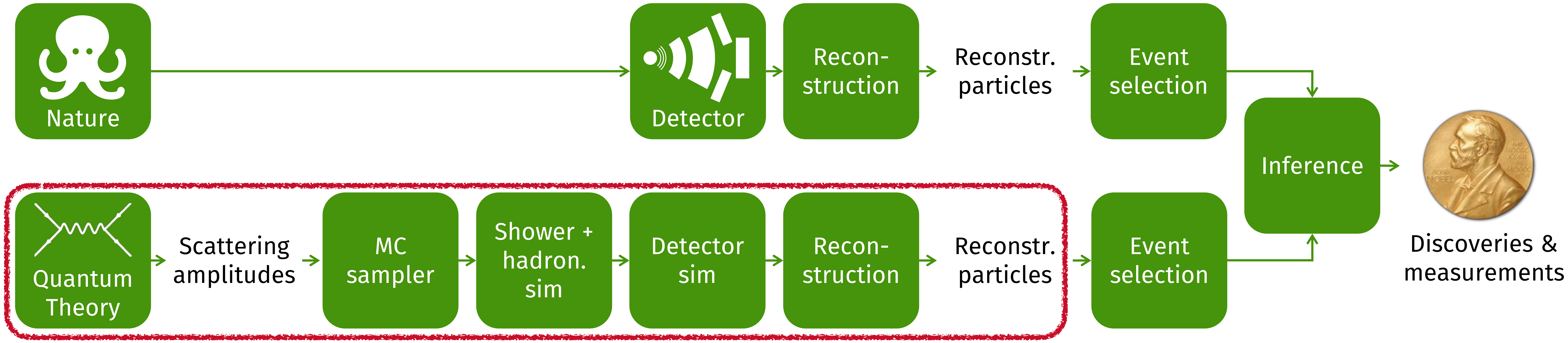
Lorentz-Equivariant CFM

Task

Dataset: $pp \to t_h \bar{t}_h + n\,j, \;\; n = 0 \dots 4$

- MadGraph + Pythia + Delphes + Reconstruction
- Challenging features: $m_t, m_W, \Delta R_{jj} > 0.5$





Lorentz-Equivariant CFM

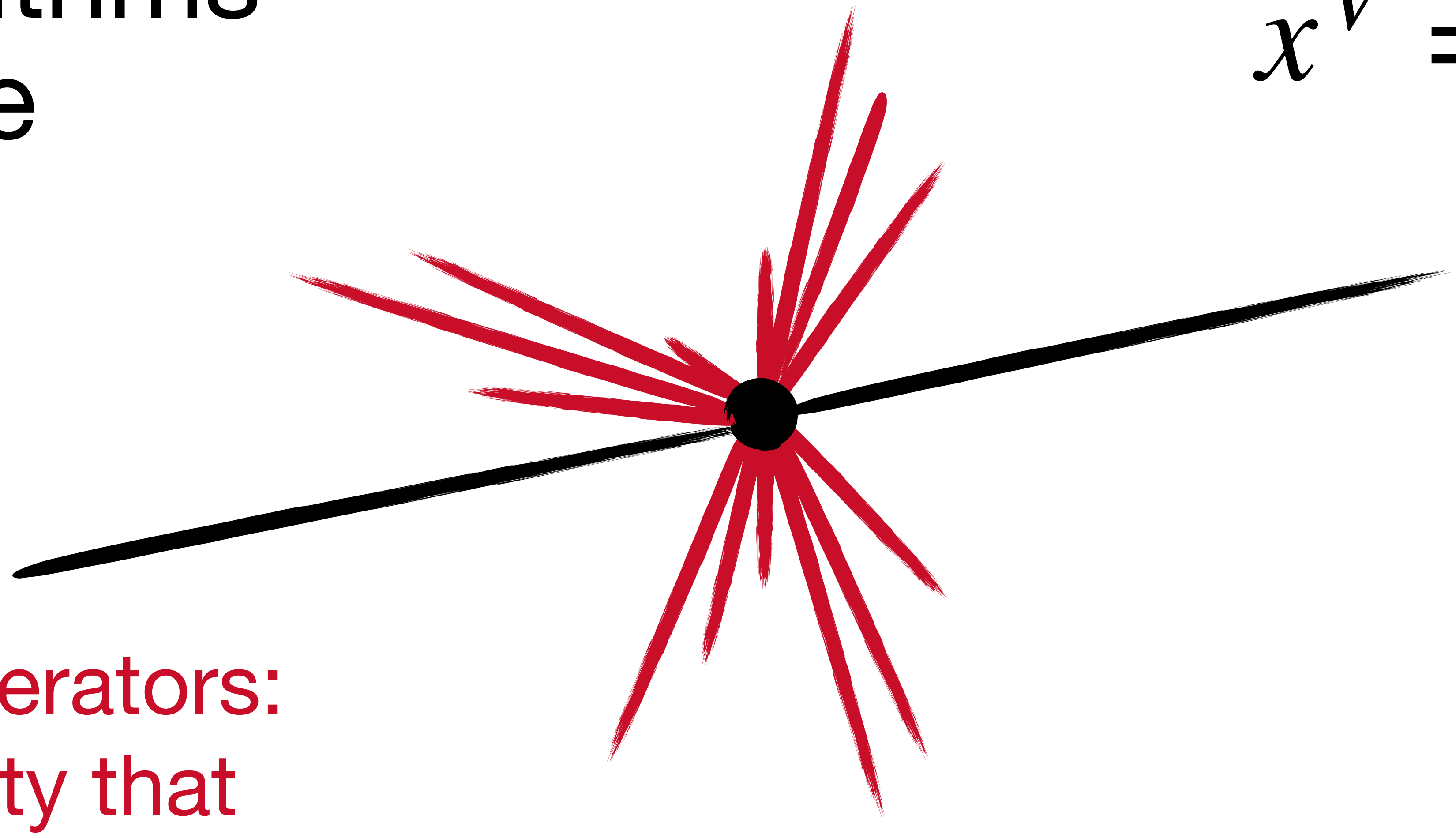
Symmetry breaking with multivectors

The density of reconstructed LHC events is not Lorentz-invariant

- Beam direction breaks invariance under rotations around the x- and y-axis
- Detector and jet algorithms break boost invariance

Add reference multivectors as extra particles to break Lorentz symmetry

- Beam reference multivector: $x^V = (0,0,0,1)$ or $x^B_{12} = 1$
- Time reference multivector: $x^V = (1,0,0,0)$



There are no Lorentz-equivariant generators: It is not possible to construct a density that is invariant under a non-compact group

Lorentz-Equivariant CFM

Conditional Flow Matching

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

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

Continuous normalising flows
arXiv:1806.07366

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

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

Conditional flow matching
arXiv:2210.02747

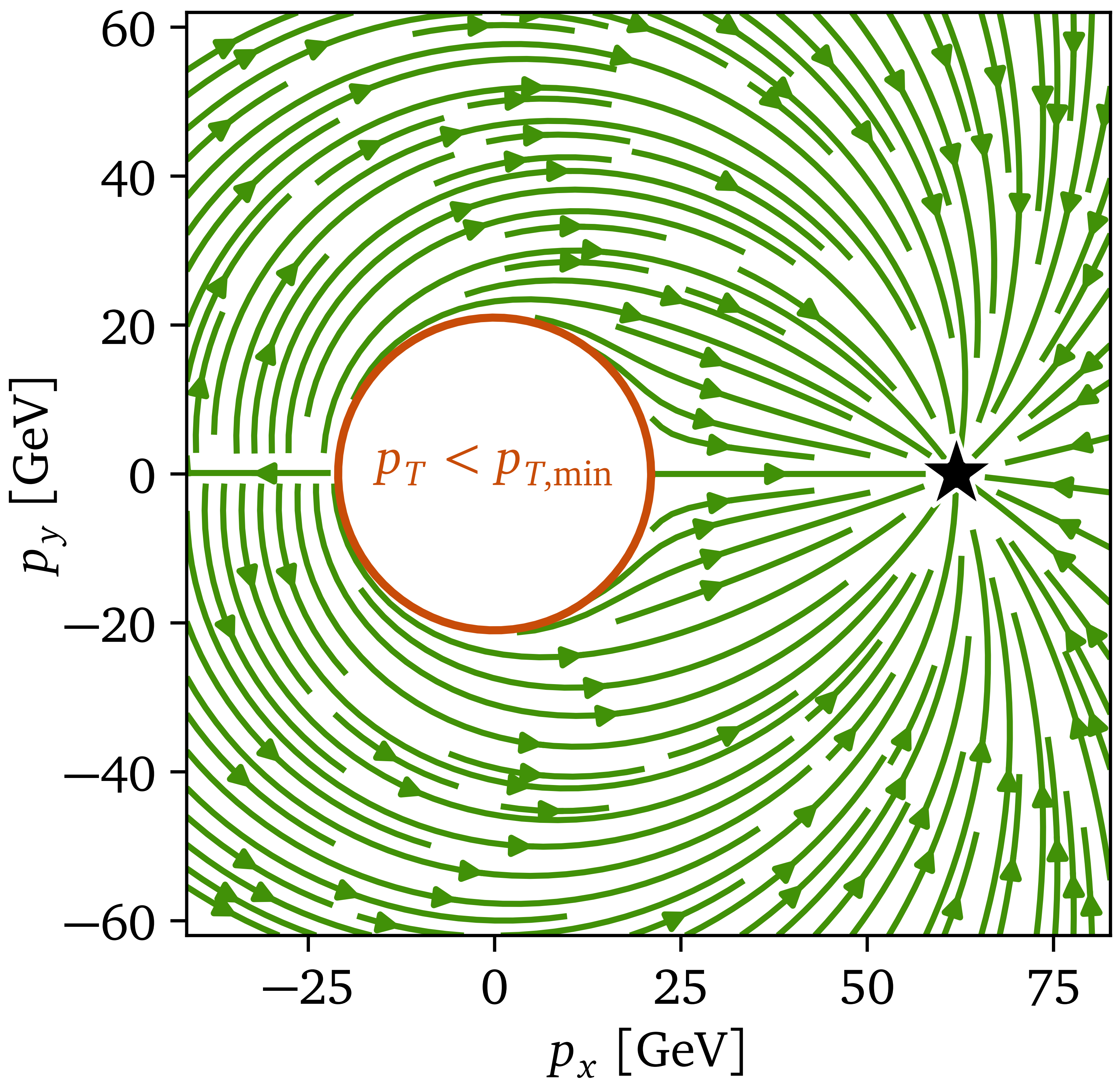
How to pick the target velocity $u_t(x \mid x_1)$?

Lorentz-Equivariant CFM

Physics-inspired target trajectories

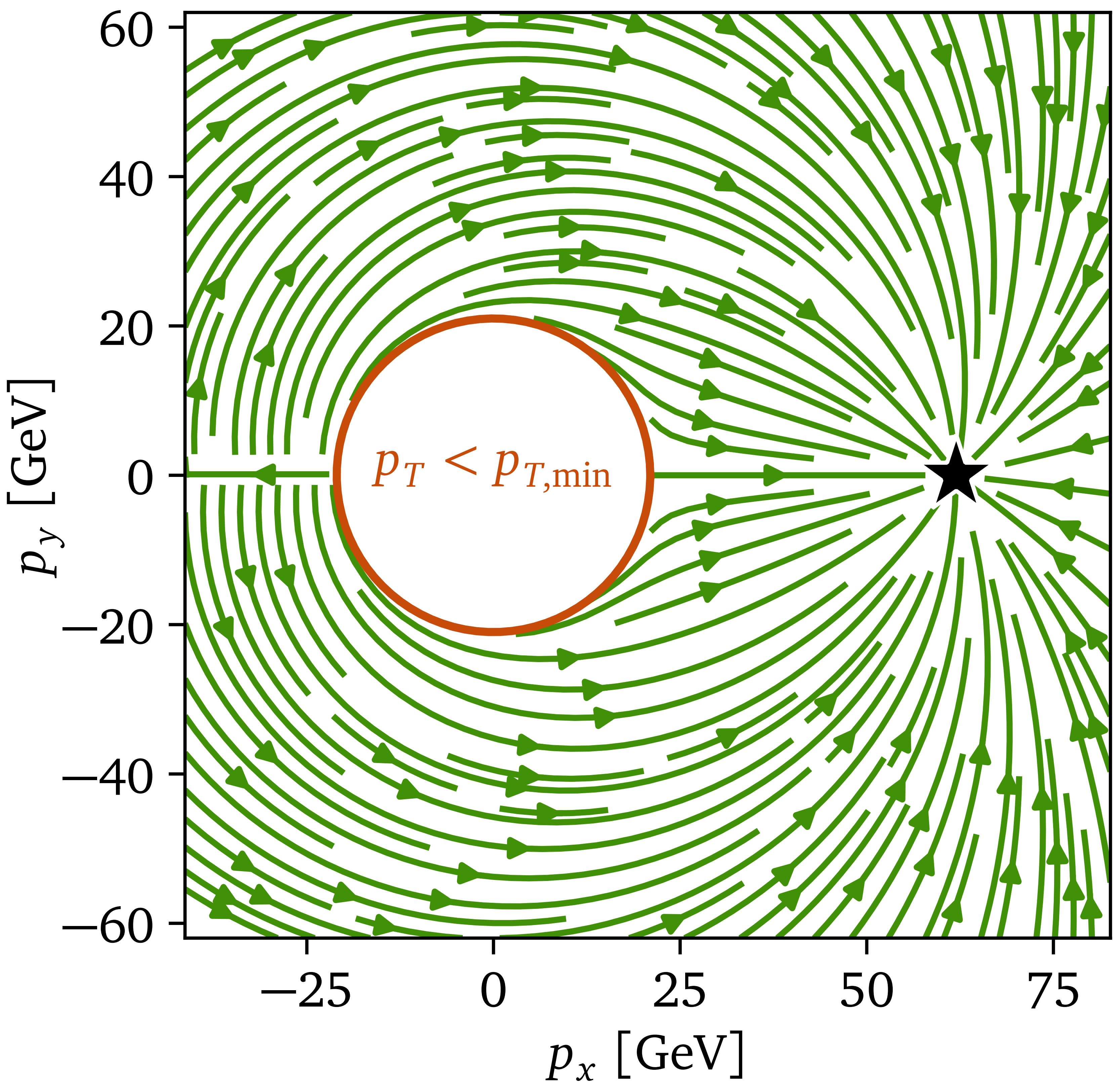
Straight trajectories in 'modified jet momenta' x:

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



Lorentz-Equivariant CFM

Physics-inspired target trajectories



Straight trajectories in ‘modified jet momenta’ x:

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

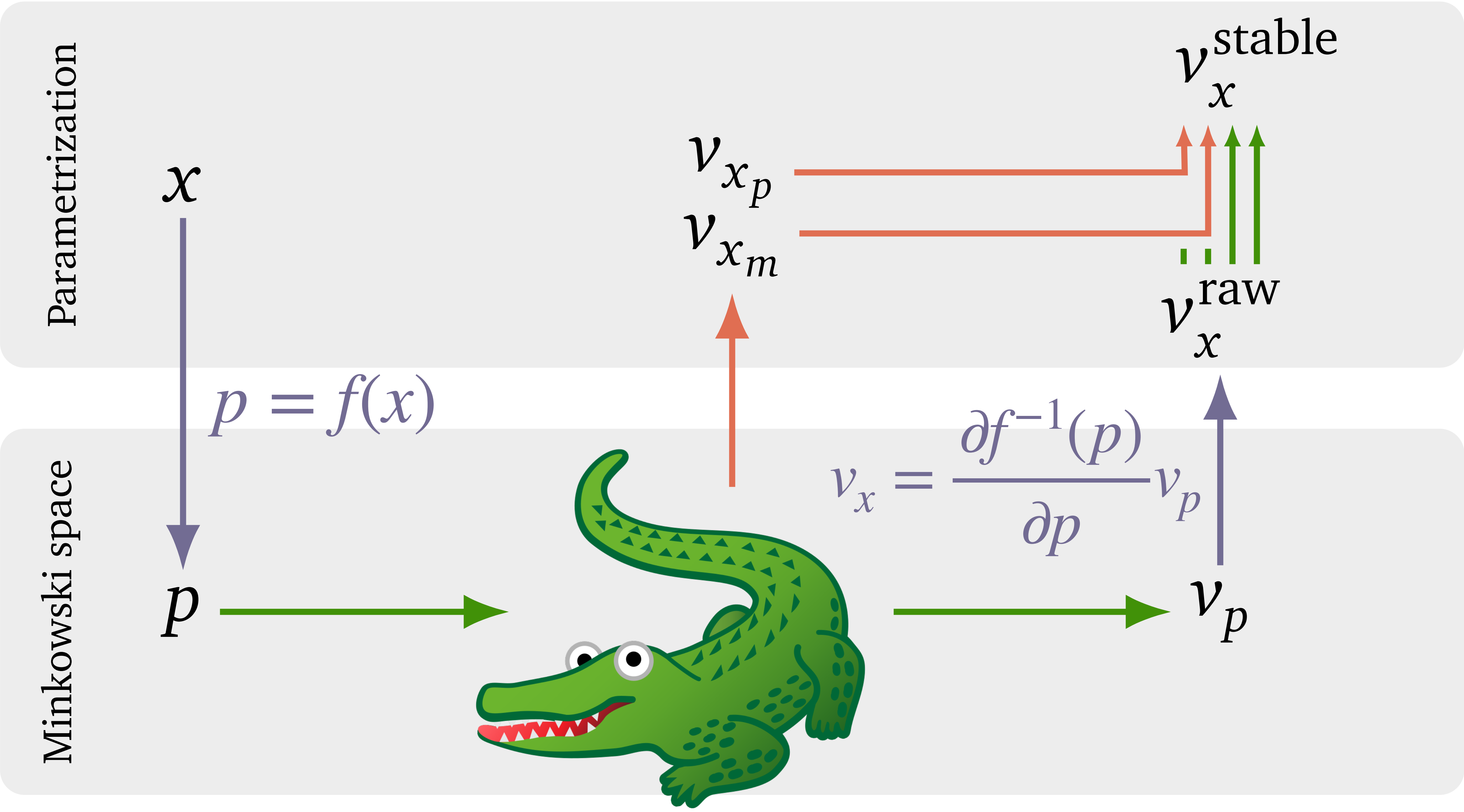
	Data	Architecture	Base distribution	Periodic	Neg. log-likelihood	AUC
	p	L-GATr	rejection sampling	✓	-30.80 ± 0.17	0.945 ± 0.004
	x	MLP	rejection sampling	✓	-32.13 ± 0.05	0.780 ± 0.003
	x	L-GATr	rejection sampling	✗	-32.57 ± 0.05	0.530 ± 0.017
	x	L-GATr	no rejection sampling	✓	-32.58 ± 0.04	0.523 ± 0.014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

Lorentz-Equivariant CFM

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

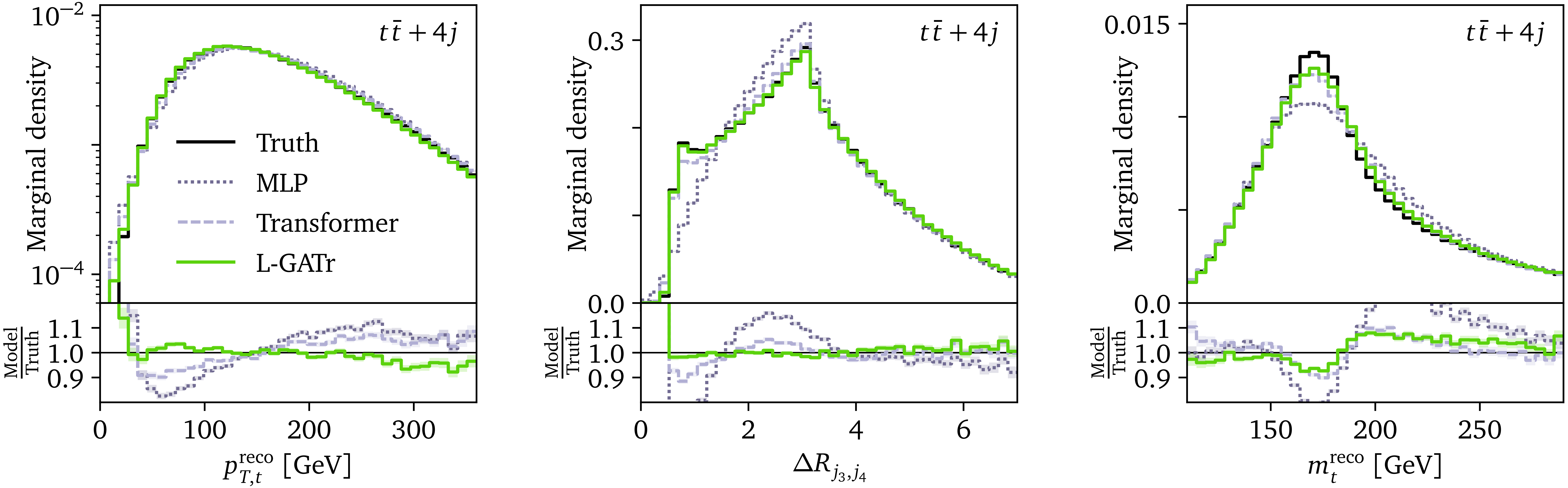
Extend standard CFM workflow with L-GATr:

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



Lorentz-Equivariant CFM

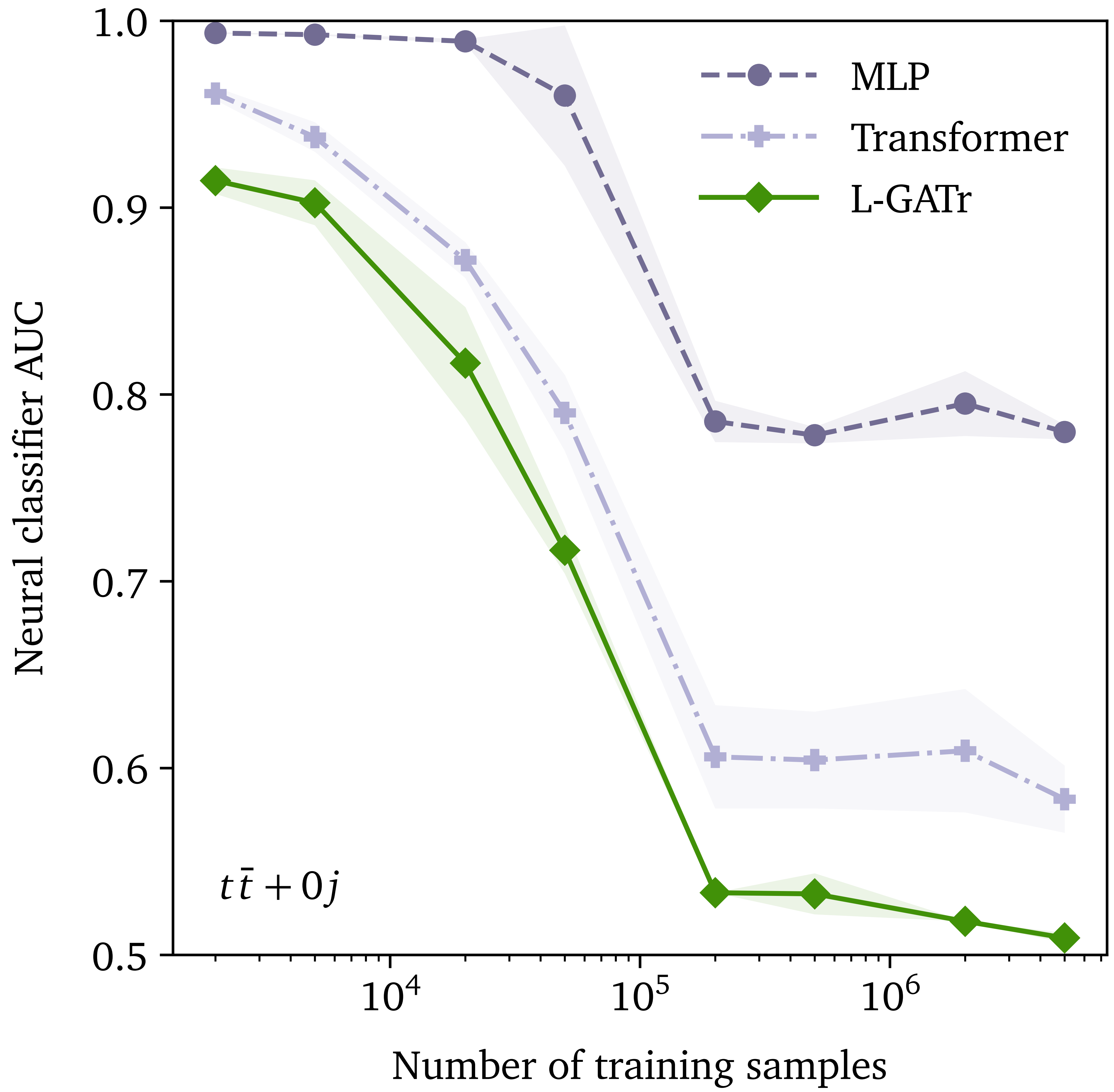
Challenging features



Equivariance helps,
especially for angular correlations

Lorentz-Equivariant CFM

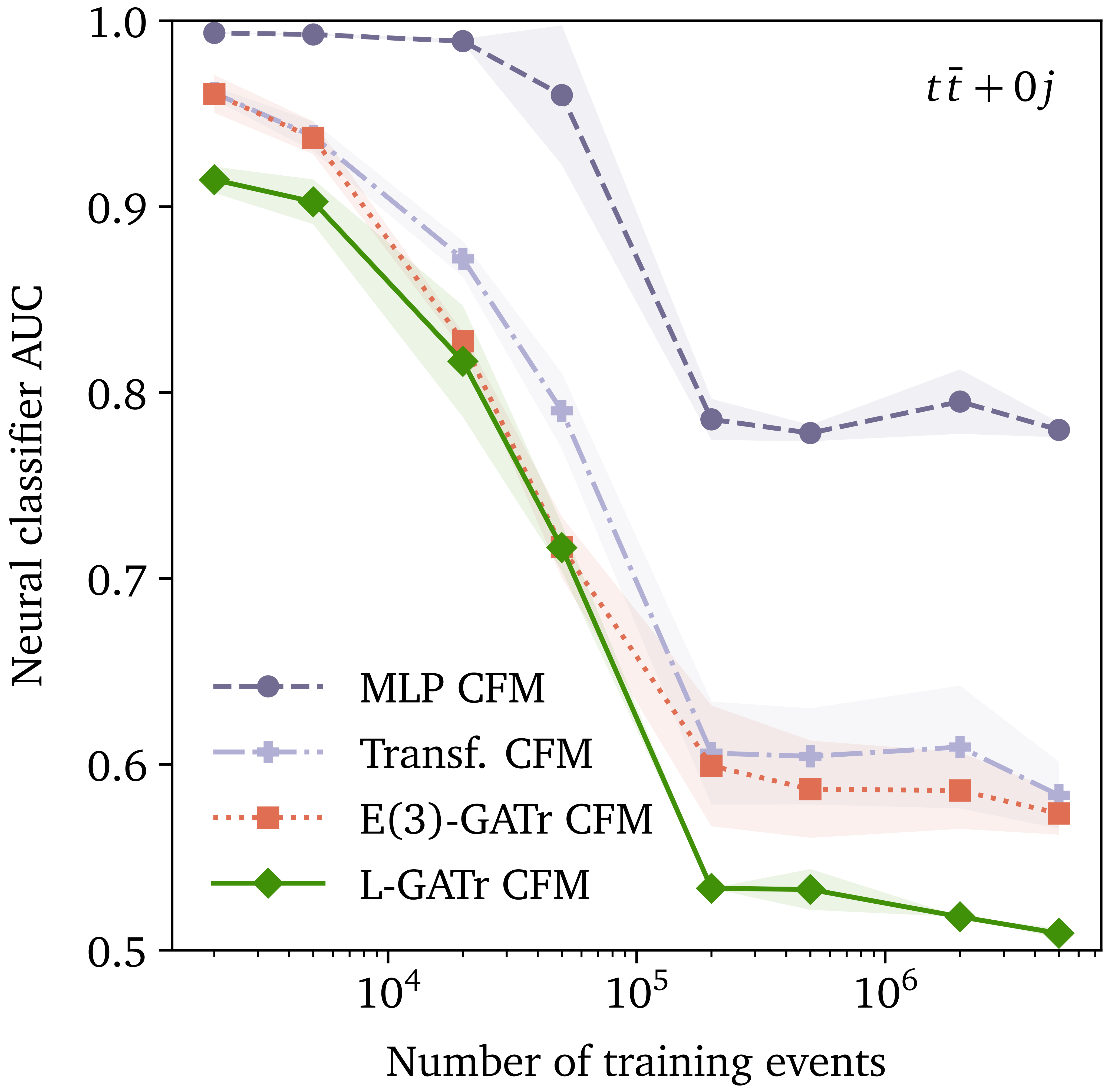
Classifier metric



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

Lorentz-Equivariant CFM

Comparison with E(3)-equivariant GATr



L-GATr and E(3)-GATr are both not boost equivariant:

- E(3)-GATr: No boost equivariance at all
- L-GATr: Boost equivariance broken by reference multivectors

Equivariant networks
with full symmetry breaking
outperform non-equivariant networks

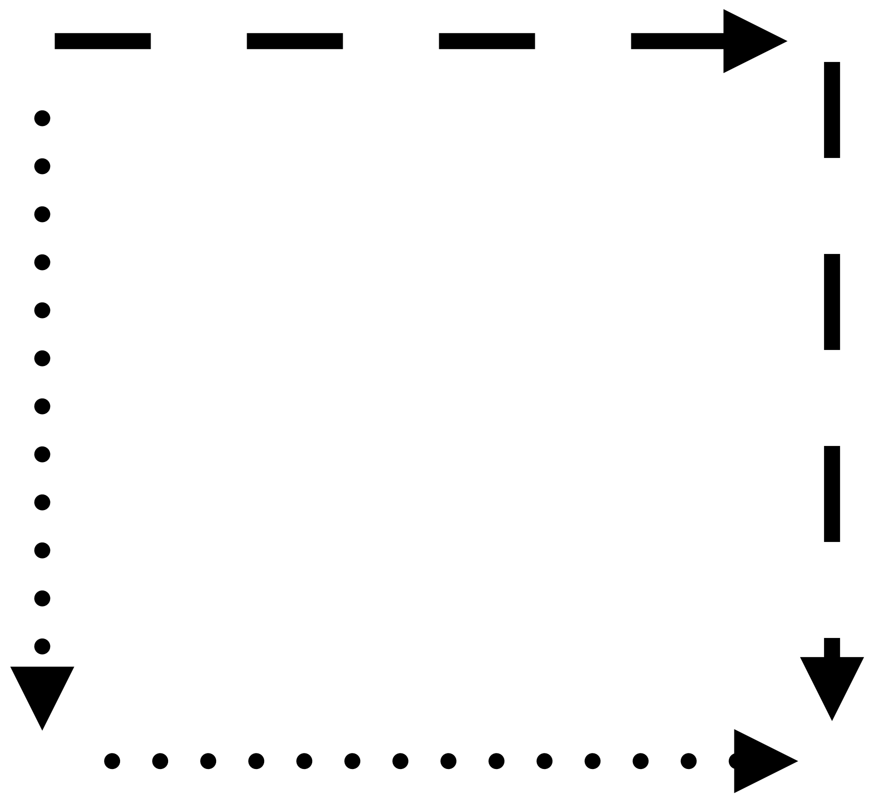
Messages

- L-GATr is a versatile architecture for LHC physics
- Equivariance helps - not only for jet tagging
- Boost equivariance makes the difference, despite being softly broken

Lorentz-Equivariant
Geometric Algebra
Transformer

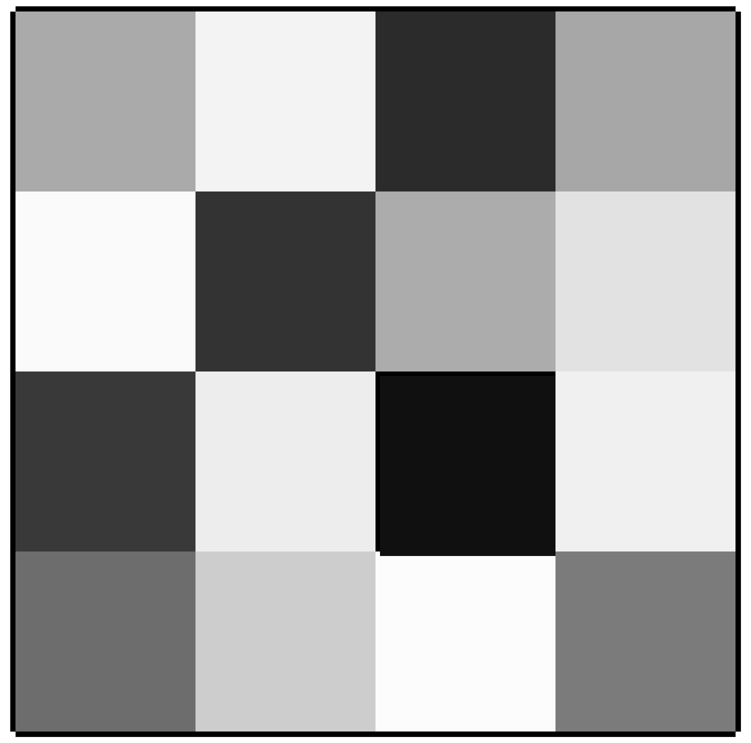
$$= \begin{pmatrix} 1 \\ \gamma^\mu \\ \sigma^{\mu\nu} \\ \gamma^\mu\gamma_5 \\ \gamma_5 \end{pmatrix} +$$

Geometric algebra representations



$+$

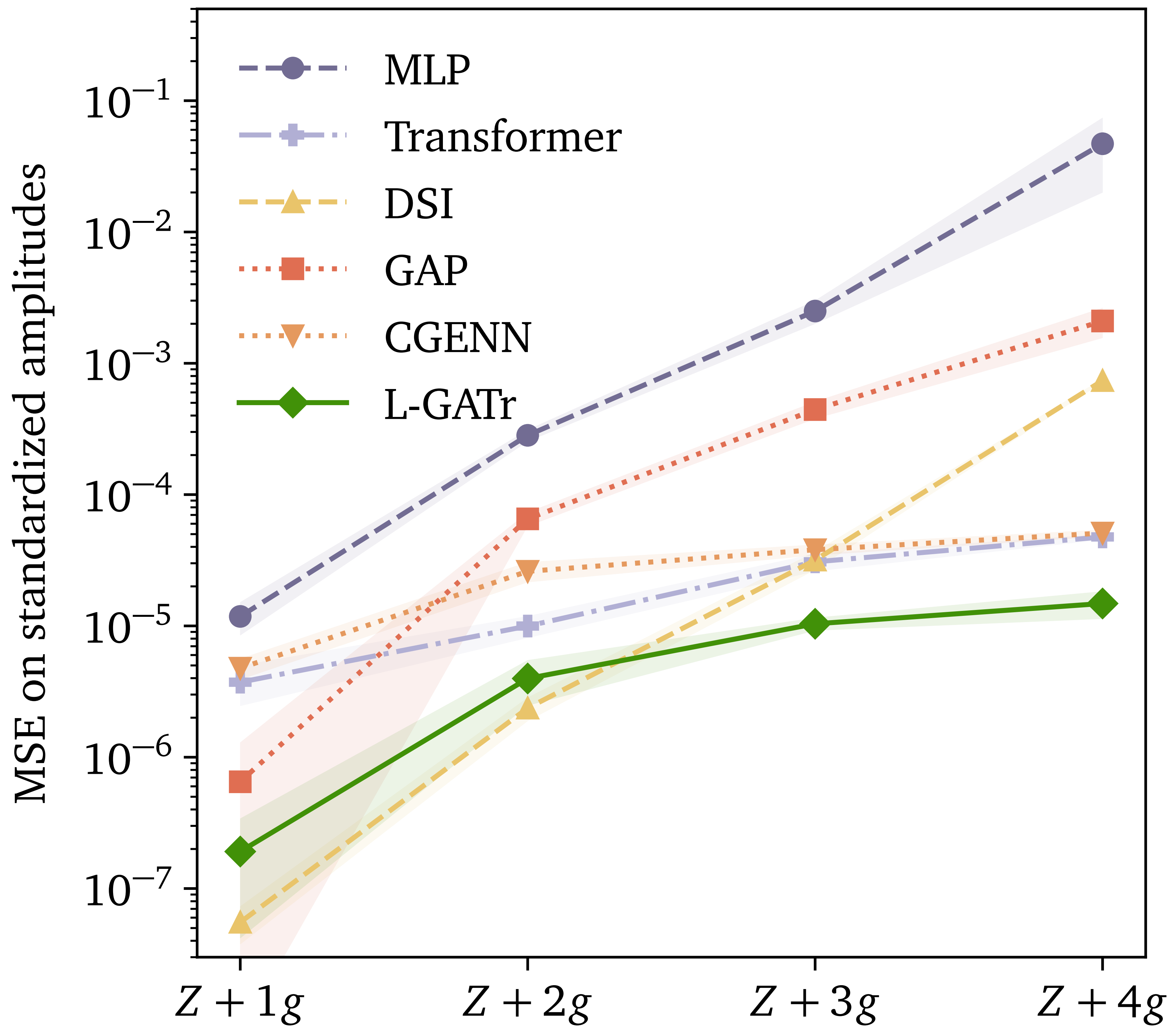
Lorentz-Equivariant layers



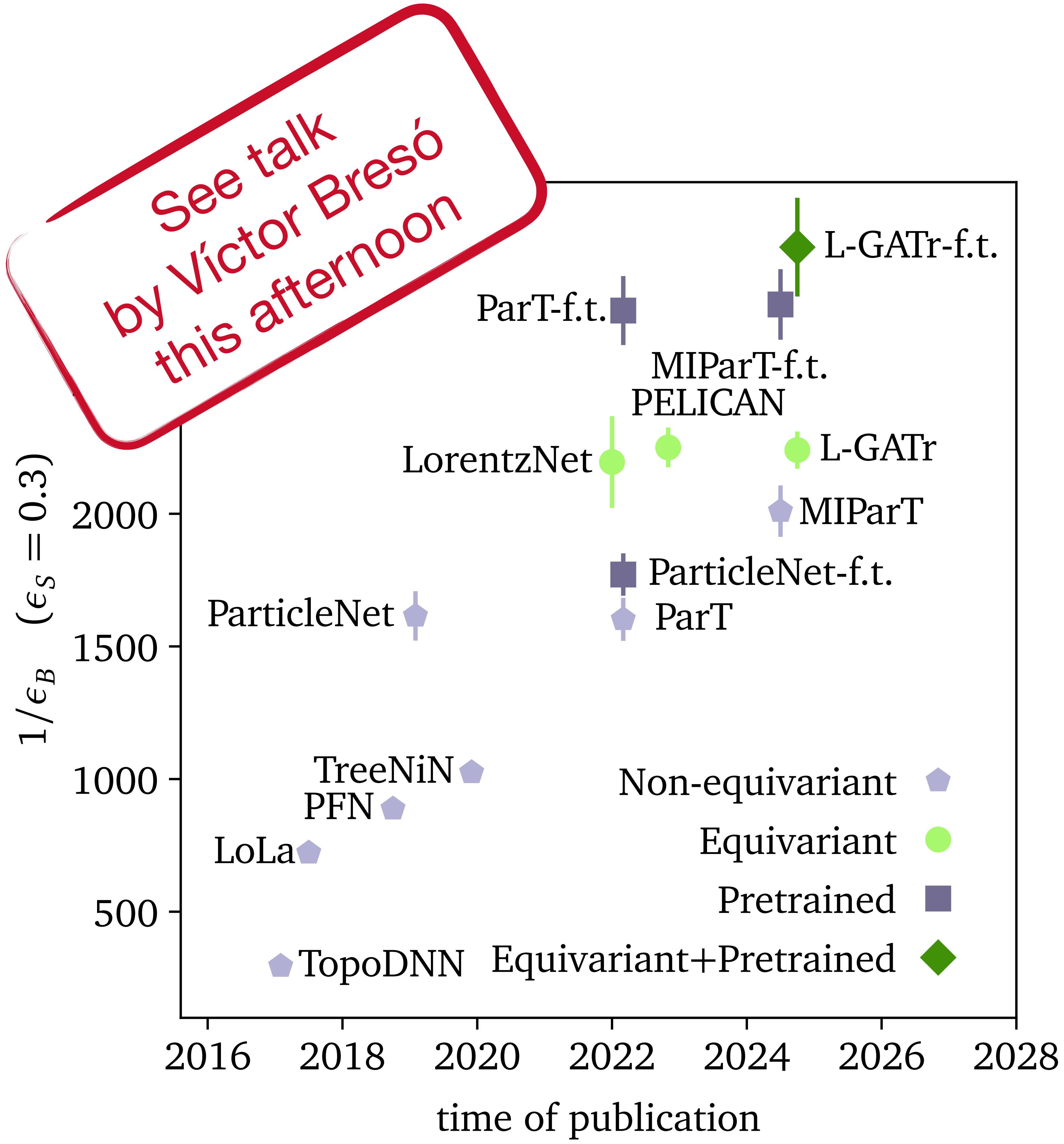
Transformer architecture

arXiv:2411.00446

Spoiler: L-GATr beyond Event Generation



Amplitude regression



Jet tagging

L-GATr is easy to use

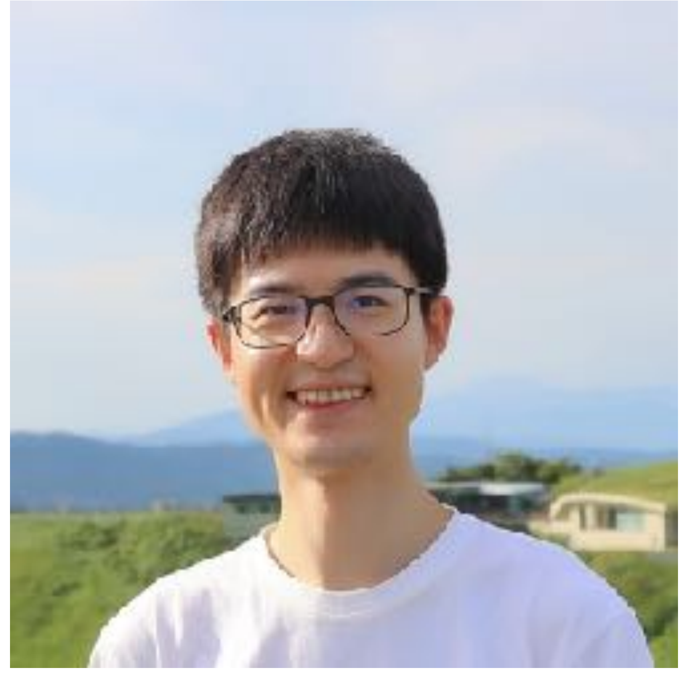
```python
from gatr import GATr, SelfAttentionConfig, MLPConfig
from gatr.interface import embed_vector, extract_scalar, embed_spurions
import torch


class ExampleWrapper(torch.nn.Module):
    """Example wrapper around a L-GATr model.

    Parameters
    ----------
    num_blocks : int
        Number of transformer blocks
    hidden_mv_channels : int
        Number of hidden multivector channels
    hidden_s_channels : int
        Number of hidden scalar channels
    """

    def __init__(self, blocks=6, hidden_mv_channels=16, hidden_s_channels=32):
        super().__init__()
        self.gatr = GATr(
            in_mv_channels=3,
            out_mv_channels=1,
            hidden_mv_channels=hidden_mv_channels,
            in_s_channels=None,
            out_s_channels=None,
            hidden_s_channels=hidden_s_channels,
            num_blocks=num_blocks,
            attention=SelfAttentionConfig(),  # Use default parameters for attention
            mlp=MLPConfig(),  # Use default parameters for MLP
        )
```

```python
def forward(self, fourmomenta):
    """Forward pass.

    Parameters
    ----------
    fourmomenta : torch.Tensor with shape (batchsize, num_points, 4)
        fourmomentum point cloud input data

    Returns
    -------
    outputs : torch.Tensor with shape (batchsize, 1)
        Model prediction: a single scalar for the whole point cloud.
    """
    batchsize, num_points, _ = fourmomenta.shape

    # Embed fourmomentum point cloud inputs in GA
    multivectors = embed_vector(fourmomenta).unsqueeze(-2)  # (batchsize, num_points, 1, 1

    # Append spurions for symmetry breaking (optional)
    spurions = embed_spurions(beam_reference="xyplane", add_time_reference=True, device=fo
    spurions = spurions[None, None, ...].repeat(batchsize, num_points, 1, 1)  # (batchsize
    multivectors = torch.cat((multivectors, spurions), dim=-2)  # (batchsize, num_points,

    # Pass data through GATr
    multivector_outputs, _ = self.gatr(multivectors, scalars=None)  # (batchsize, num_poin

    # Extract scalar outputs
    outputs = extract_scalar(multivector)  # (batchsize, num_points, 1)

    # Mean aggregation to extract a single scalar for the whole point cloud
    score = outputs.mean(dim=1)

    return score
```


Victor Bresó


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

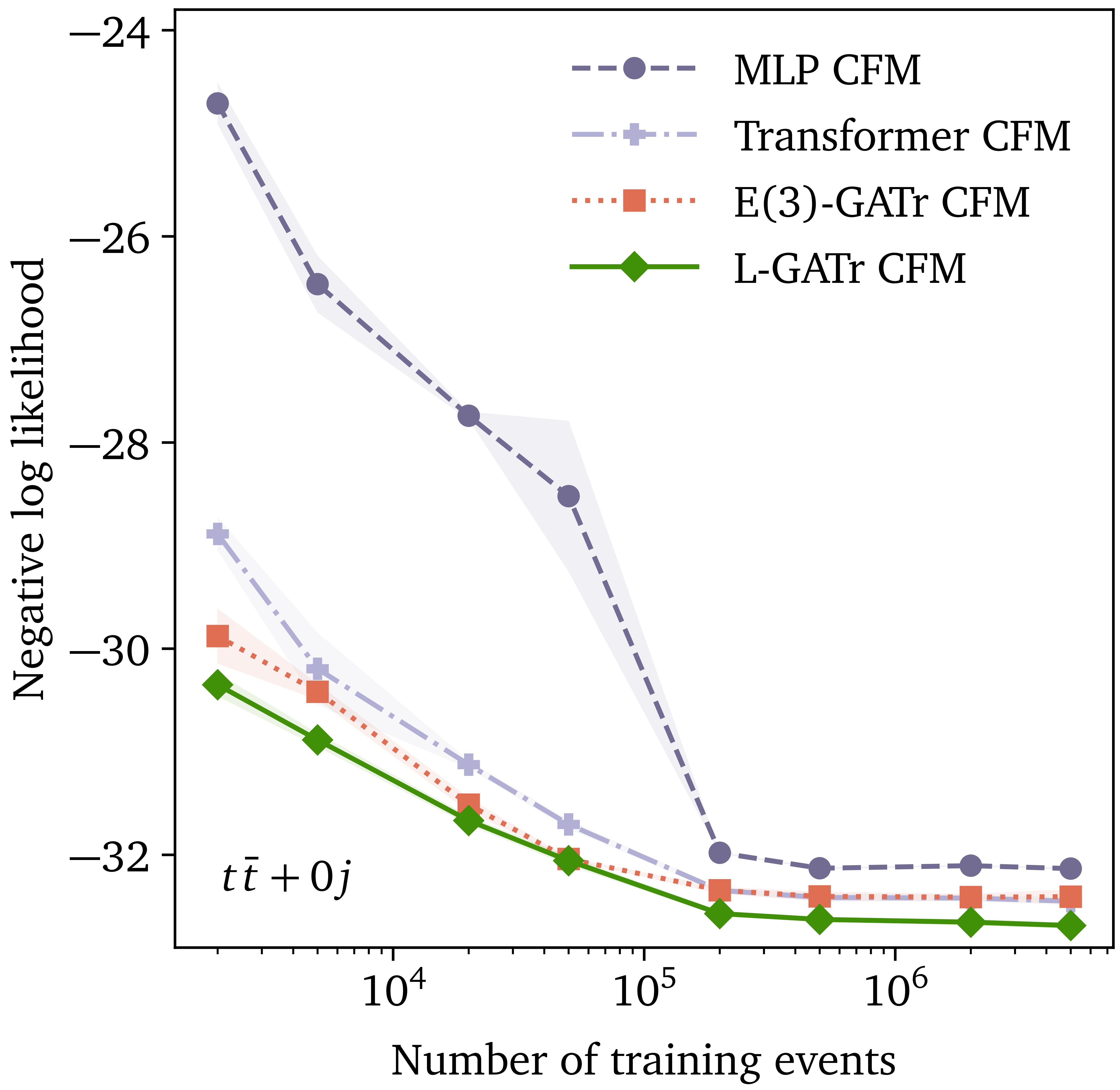
HEP paper

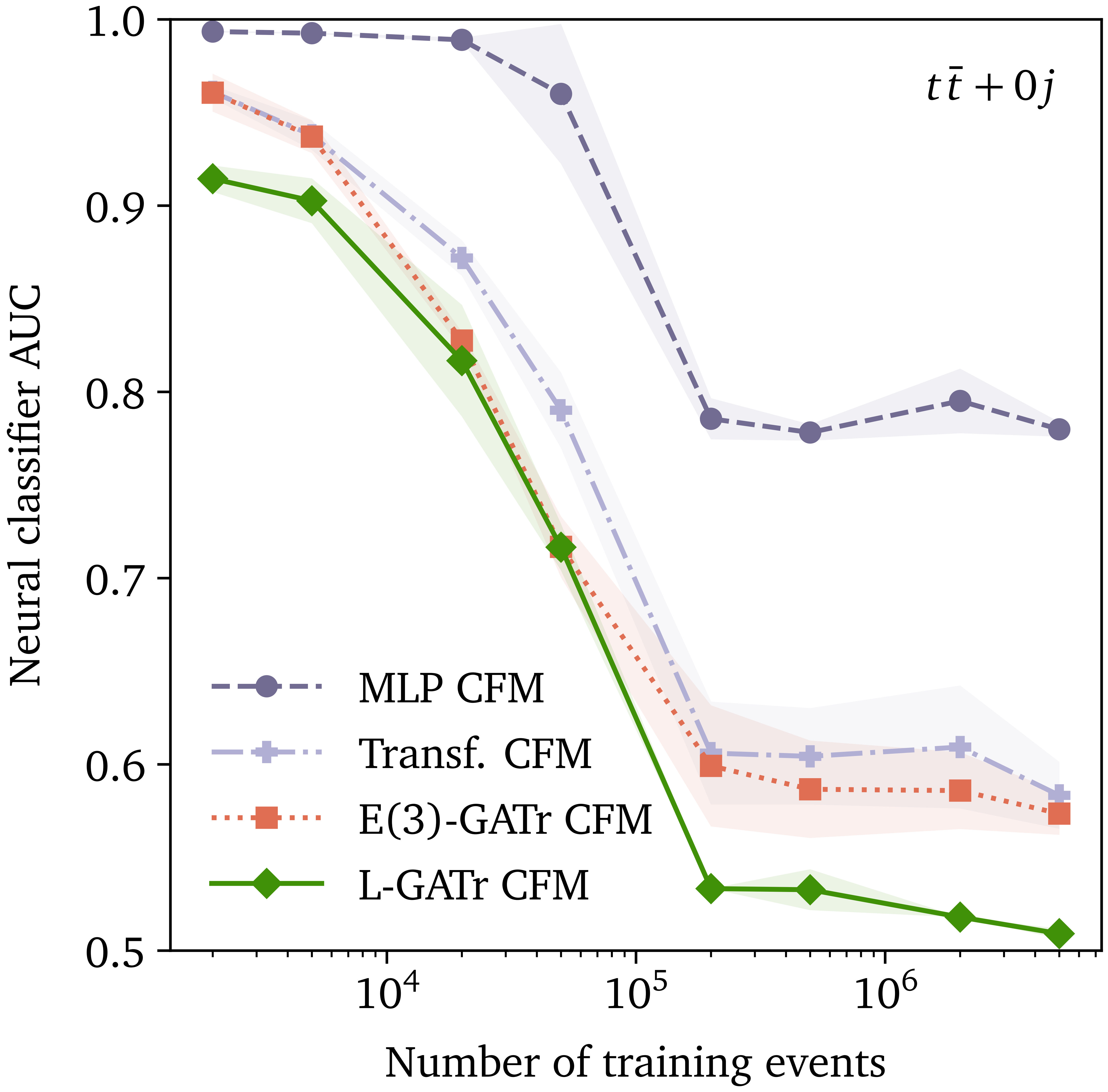
L-GATr code

For what will **you** use L-GATr?

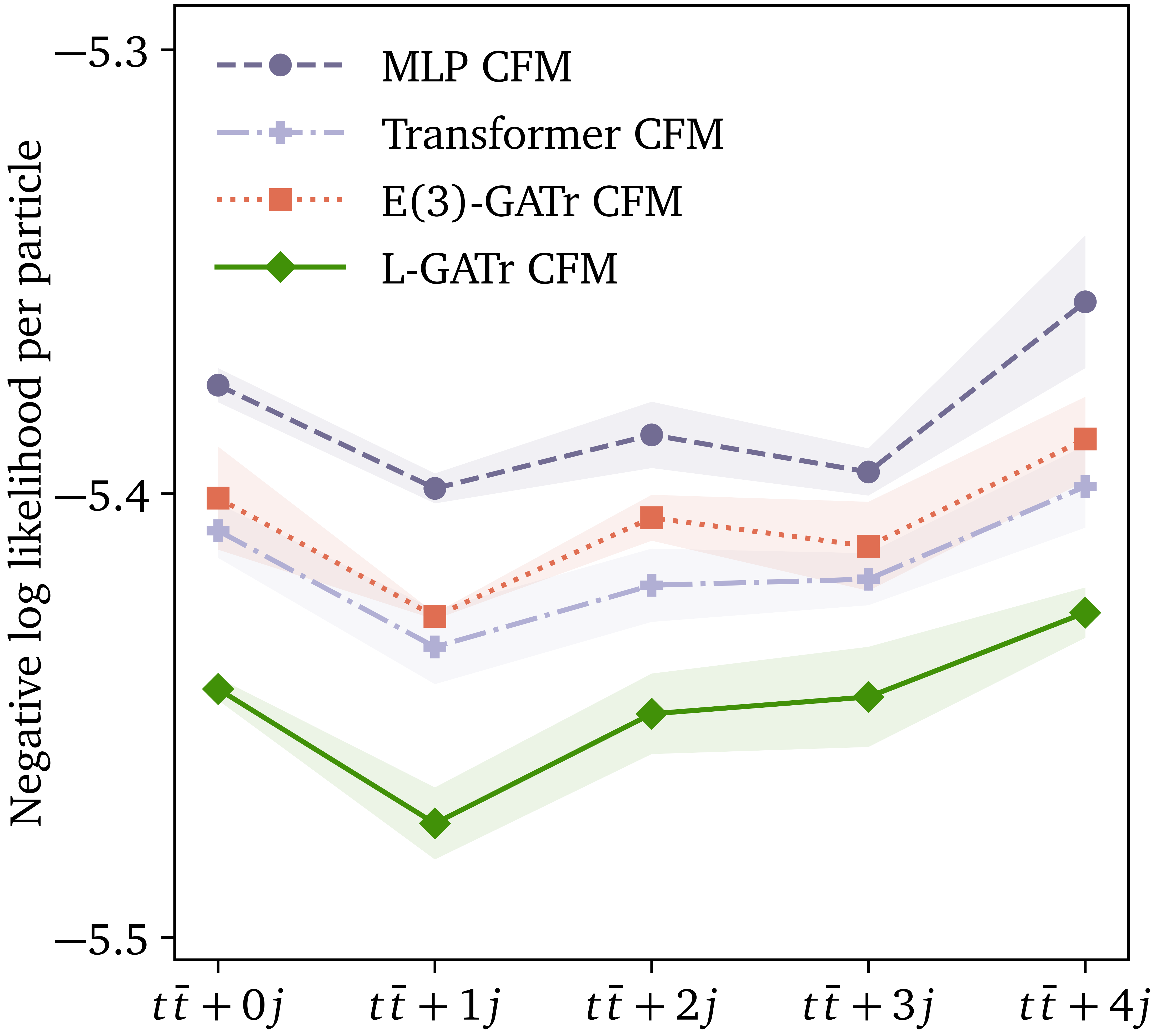
Bonus material

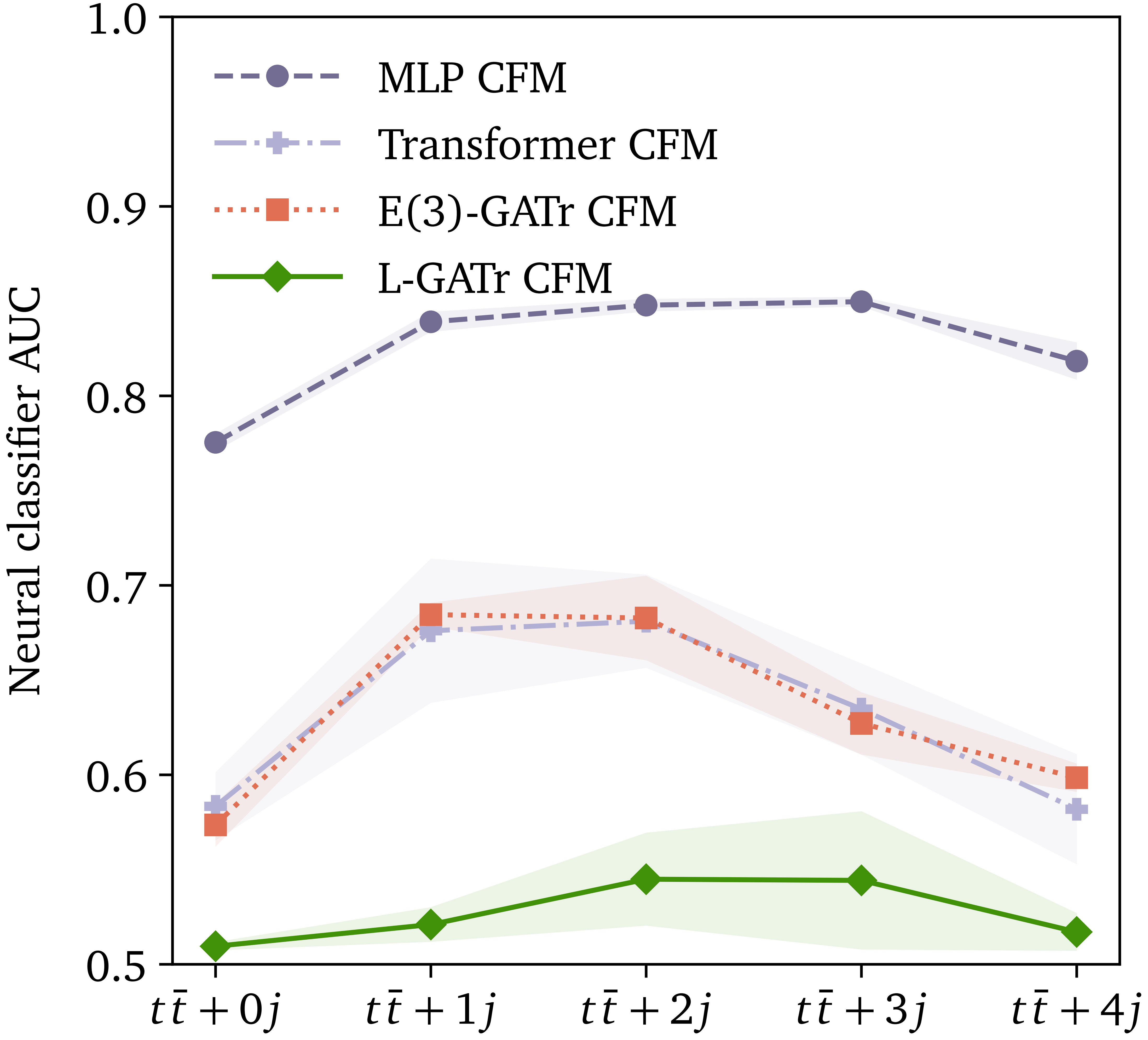
Event generation



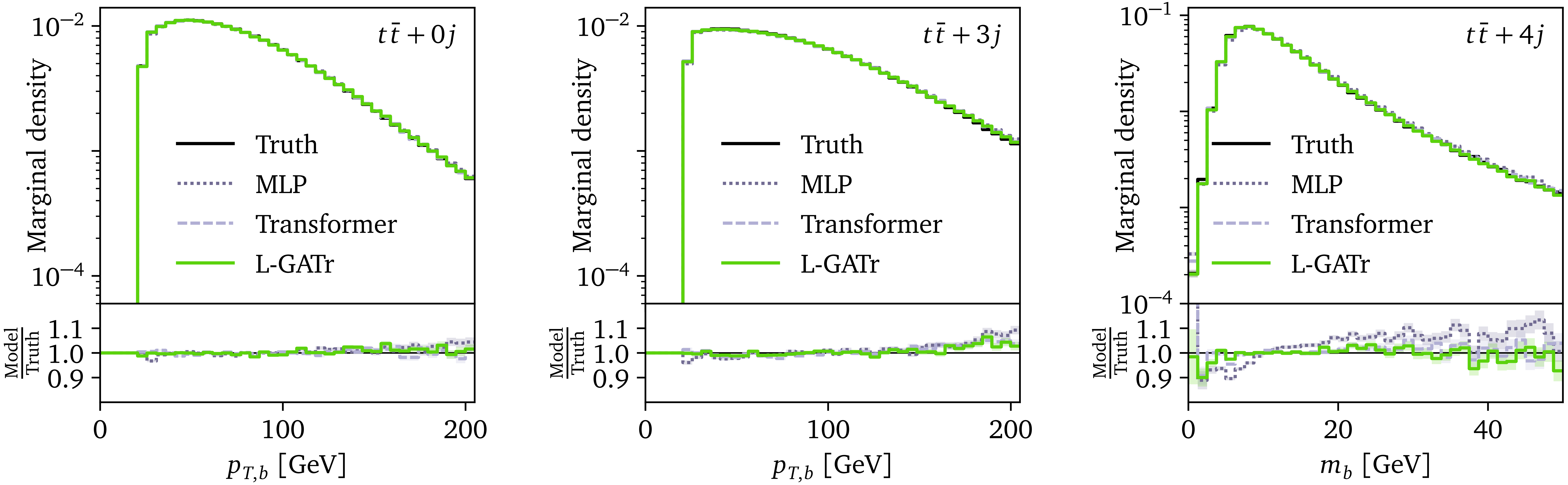


Event generation

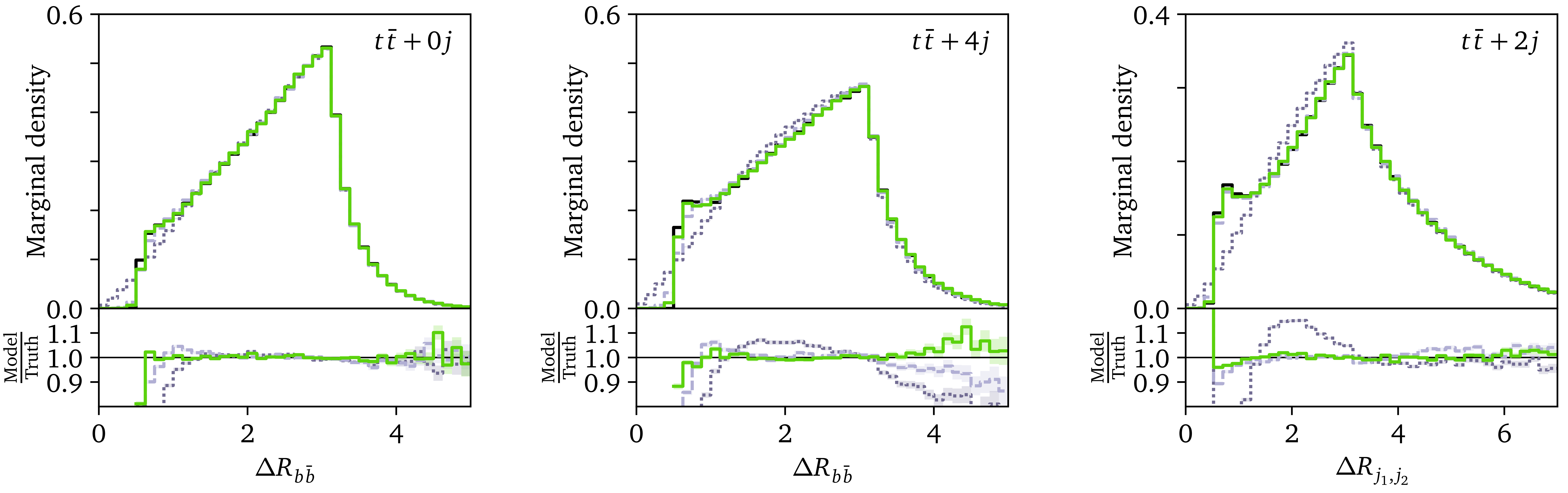




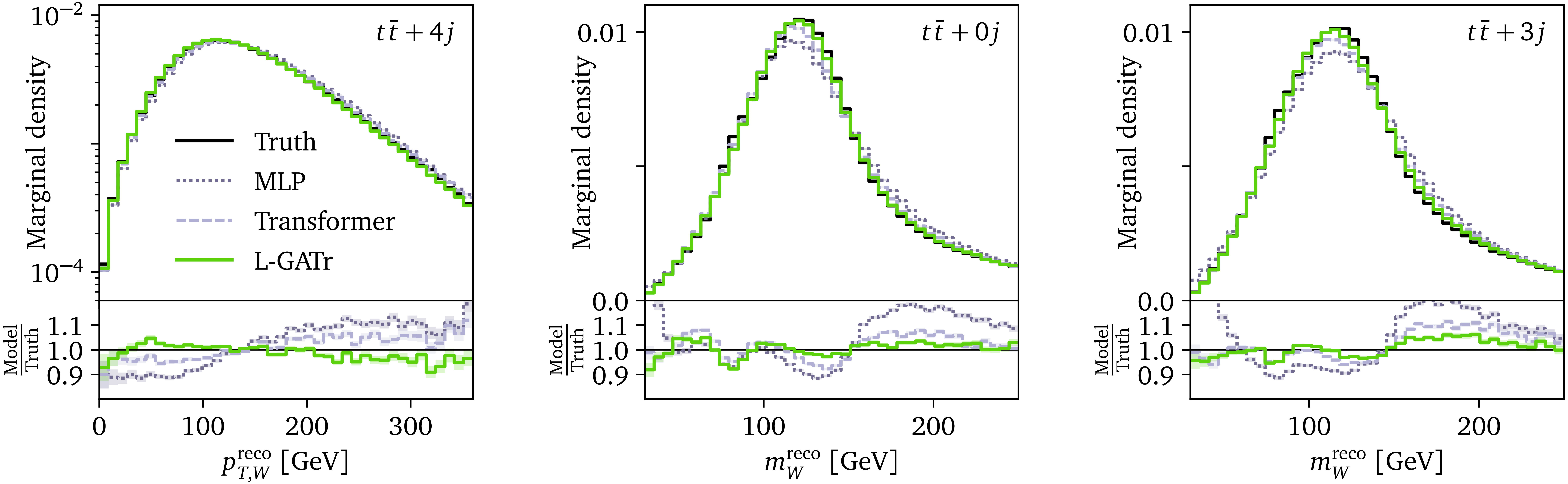
Event generation



Event generation



Event generation



Event generation