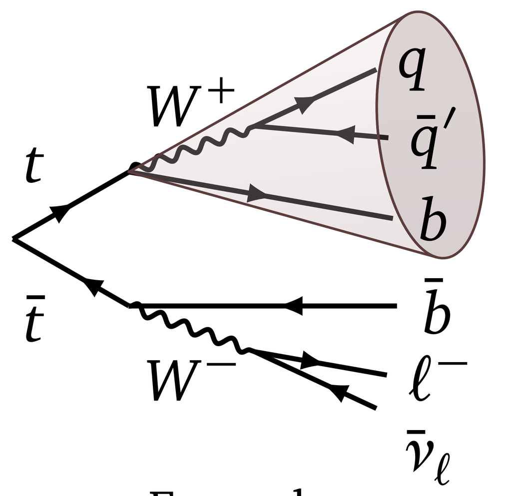
Generative Unfolding of Jets and Their Substructure

Antoine Petitjean[1], Anja Butter[1,2], Kevin Greif[3], Sofia Palacios Schweitzer[1,4], Tilman Plehn[1,5], Jonas Spinner[1,6], Daniel Whiteson[3]

Jet Structure



Example:
Top decays

- Jet structure: set $\{x_i\}_{i=1}^N$ of collimated particles
- Distortions between particle-level jets and jets reconstructed from detector measurements

Particle-level event	Reconstructed event
$x_{\text{part}} = \{x_{\text{part},i}\}_{i=1}^{N_{\text{part}}}$	$x_{\text{reco}} = \{x_{\text{reco},i}\}_{i=1}^{N_{\text{det}}}$

- $x_{\text{reco}} \sim p_{\text{data}}$ available from experiments
- Datasets of pairs $(x_{\text{part}}, x_{\text{reco}}) \sim p_{\text{sim}}$ obtained from simulations

Generative Unfolding

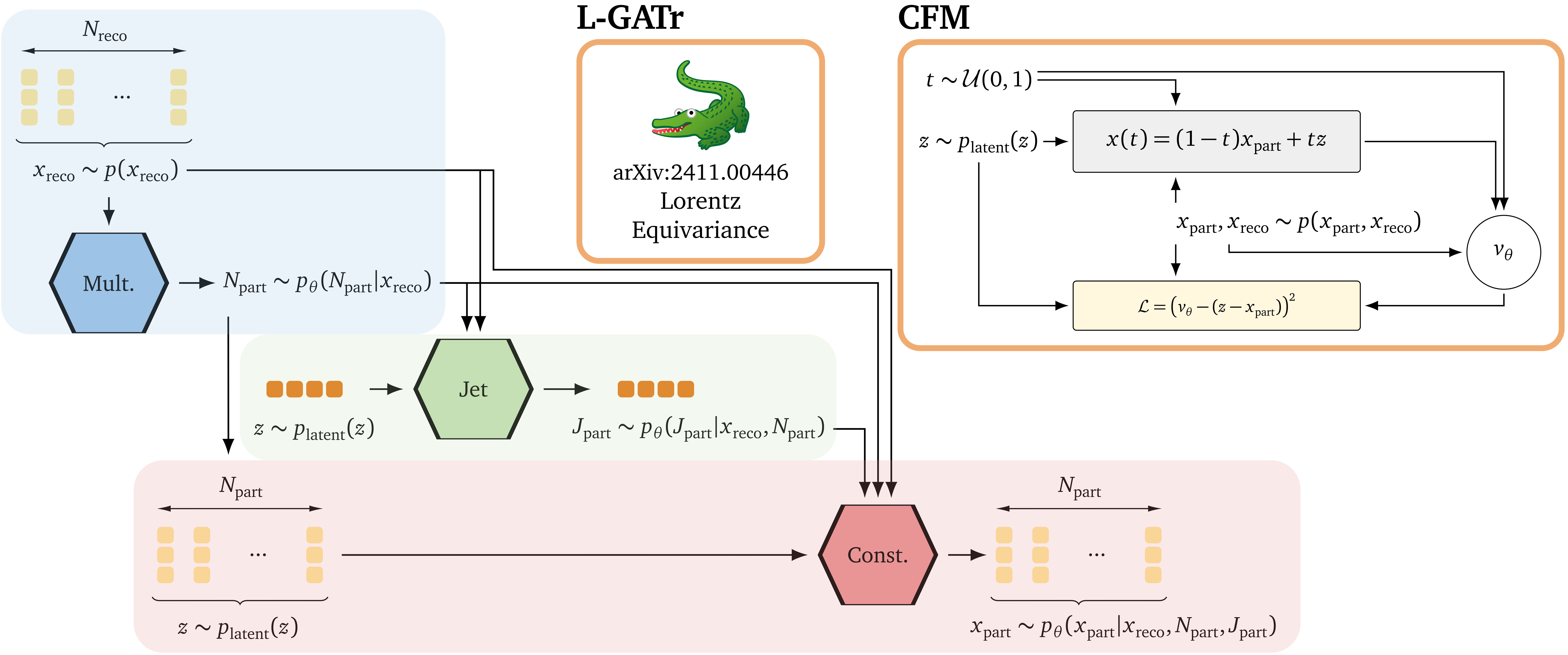
- Unfolding: recover undistorted distribution $p_{\text{unfold}}(x_{\text{part}})$

$$p_{\text{unfold}}(x_{\text{part}}) = \int dx_{\text{reco}}\, p(x_{\text{part}}|x_{\text{reco}})p_{\text{data}}(x_{\text{reco}})$$

- Requires conditional distributions $p(x_{\text{part}}|x_{\text{reco}})$ given reconstructed x_{reco}
- Generative unfolding: learn approximate $p_\theta(x_{\text{part}}|x_{\text{reco}}) \simeq p(x_{\text{part}}|x_{\text{reco}})$
  - High-dimensional and unbinned, in contrast to traditional unfolding
  - Can sample from $p_\theta(x_{\text{part}}|x_{\text{reco}})$ for any detector-level event
- Conditional Flow Matching: Learn a push-forward flow

$$z \sim p_{\text{latent}}(z) \mapsto x_{\text{part}} = \left[\int_0^1 dt\, f_\theta(z, t|x_{\text{reco}})\right] \sim p_\theta(x_{\text{part}}|x_{\text{reco}})$$

Three-steps Unfolding



L-GATr

arXiv:2411.00446
Lorentz
Equivariance

CFM

$t \sim \mathcal{U}(0,1)$
$z \sim p_{\text{latent}}(z)$
$x(t) = (1-t)x_{\text{part}} + tz$
$x_{\text{part}}, x_{\text{reco}} \sim p(x_{\text{part}}, x_{\text{reco}})$
v_θ
$\mathcal{L} = (v_\theta - (z - x_{\text{part}}))^2$

$$p(x_{\text{part}}, J_{\text{part}}, N_{\text{part}}|x_{\text{reco}}) = p(N_{\text{part}}|x_{\text{reco}}) \times p(J_{\text{part}}|x_{\text{reco}}, N_{\text{part}}) \times p(x_{\text{part}}|x_{\text{reco}}, N_{\text{part}}, J_{\text{part}})$$

1. Multiplicity

Mixture of 5 Gaussians

$$N_{\text{part}} \sim \sum_{i=1}^{5} w_i \mathcal{N}(\mu_i, \sigma_i)$$

2. Jet Kinematics

Auxiliary jet parametrization

$$J_{\text{part}} = (\log p_{T,J},\ \phi_J,\ \eta_J,\ \log m_J^2)$$

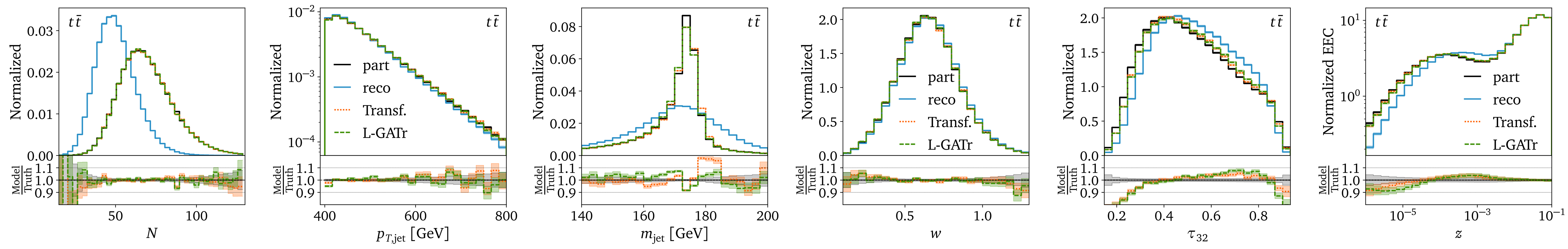
3. Constituents Kinematics

Unfolding of constituents around the auxiliary jet

$$x_{\text{part},i} = (\log p_{T,i} - \log p_{T,J},\ \phi_i - \phi_J,\ \eta_i - \eta_J)$$

- Three-steps sampling process with distinct networks
- Two auxiliary variables: multiplicity N_{part} and jet kinematics J_{part}
- Unfolded jet entirely constructed from the constituents x_{part}
- Transformer-based architecture with encoder and decoder blocks
- Flow learned in specified kinematics coordinates
- End-to-end Lorentz-equivariant version using L-GATr architecture

Results



- Plots of observables for boosted top jets ($p_{T,J} > 400$ GeV)
  - Percent-level accuracy for the constituents kinematics unfolding
  - Precise unfolding of the jet substructure, even for sharp features
  - More plots and additional dataset in the paper

First generative unfolding method to unfold several hundred dimensions at once and achieve high precision on the full jet substructure.

[1]Institut für Theoretische Physik, Universität Heidelberg
[2]LPNHE, Sorbonne Université, Université Paris Cité, CNRS/IN2P3, Paris, France
[3]Department of Physics and Astronomy, University of California, Irvine, USA
[4]NHETC, Department of Physics & Astronomy, Rutgers University, Piscataway, NJ, USA
[5]Interdisciplinary Center for Scientific Computing (IWR), Universität Heidelberg, Germany
[6]Institute for Particle Physics Phenomenology, Department of Physics, Durham University, UK

AI PHY

UNIVERSITÄT HEIDELBERG
ZUKUNFT
SEIT 1386