WEIZMANN INSTITUTE OF SCIENCE

BERKELEY LAB

Parnassus

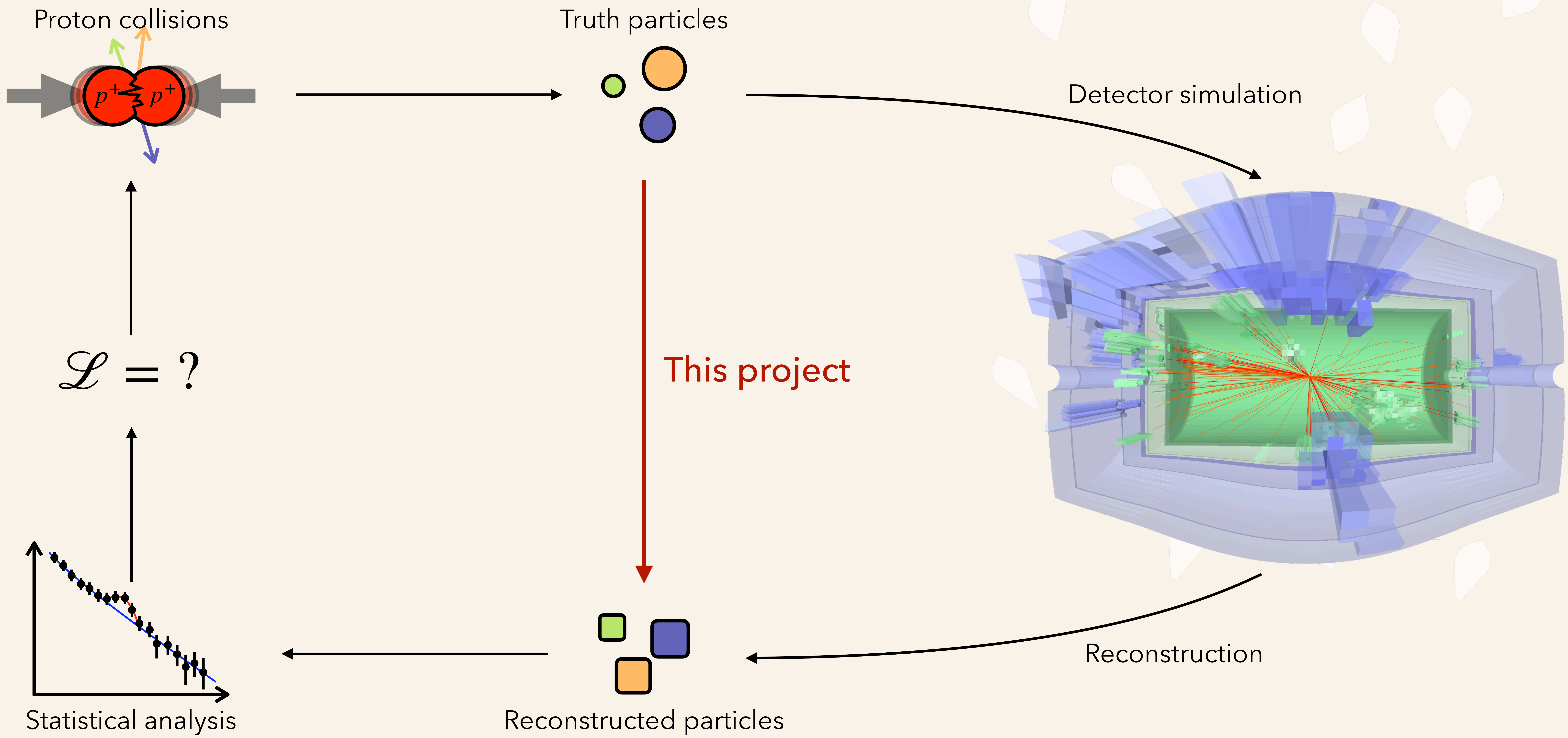
An Automated Approach to Accurate, Precise, and Fast Detector Simulation and Reconstruction

Etienne Dreyer, Eilam Gross, Dmitrii Kobylianskii,
Vinicius Mikuni, Benjamin Nachman, Nathalie Soybelman

ML4Jets 2024

WEIZMANN INSTITUTE OF SCIENCE

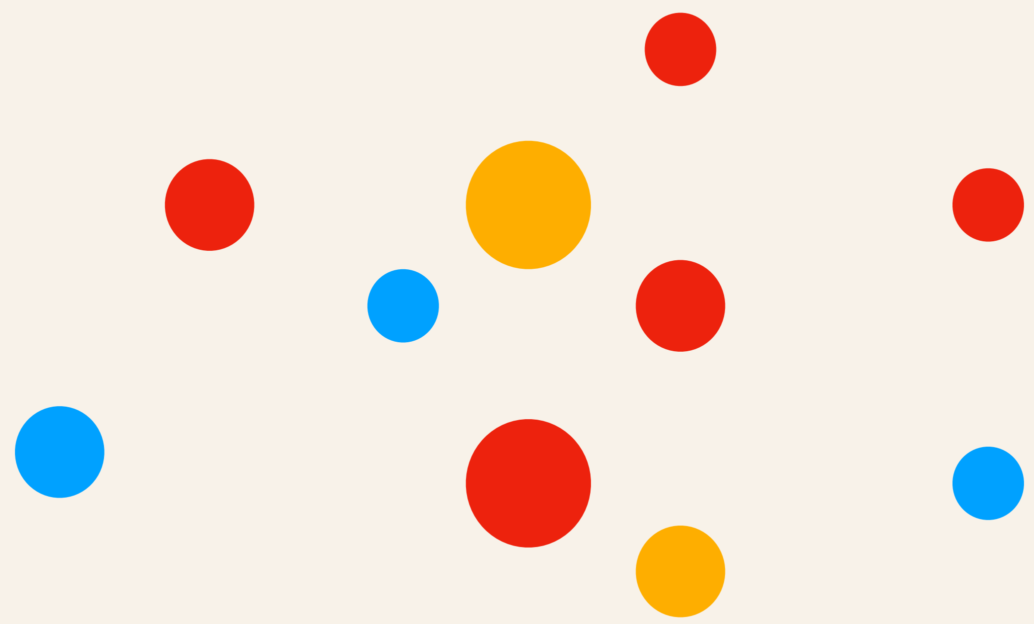
Motivation



WEIZMANN INSTITUTE OF SCIENCE

Problem to solve

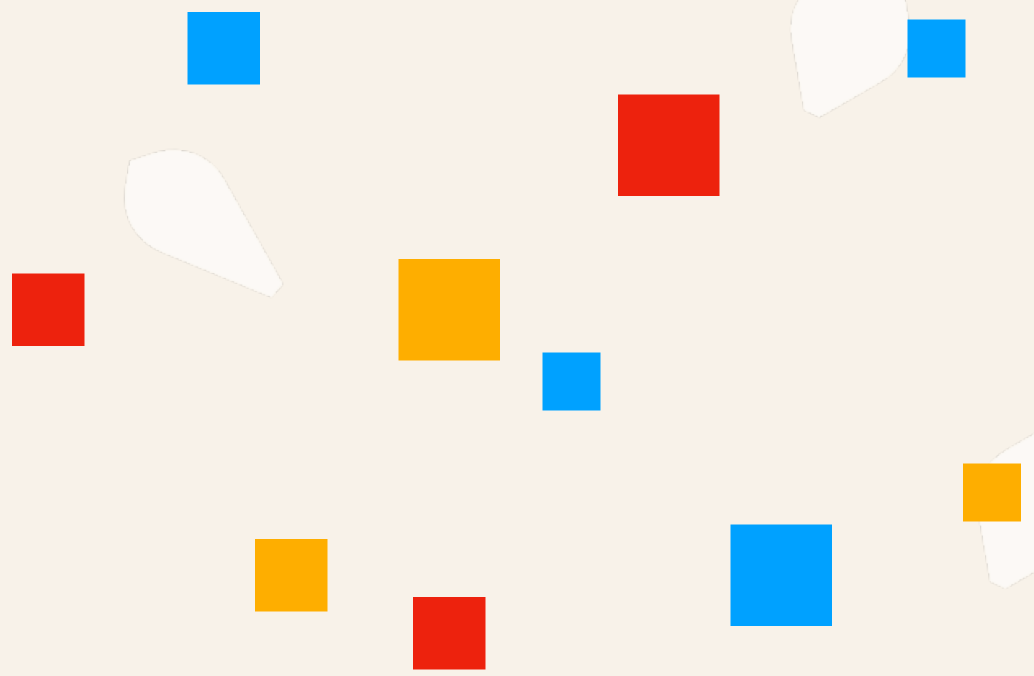
Final state truth particles



Set to Set

?

Reconstructed objects



Amount of objects

+

Their properties

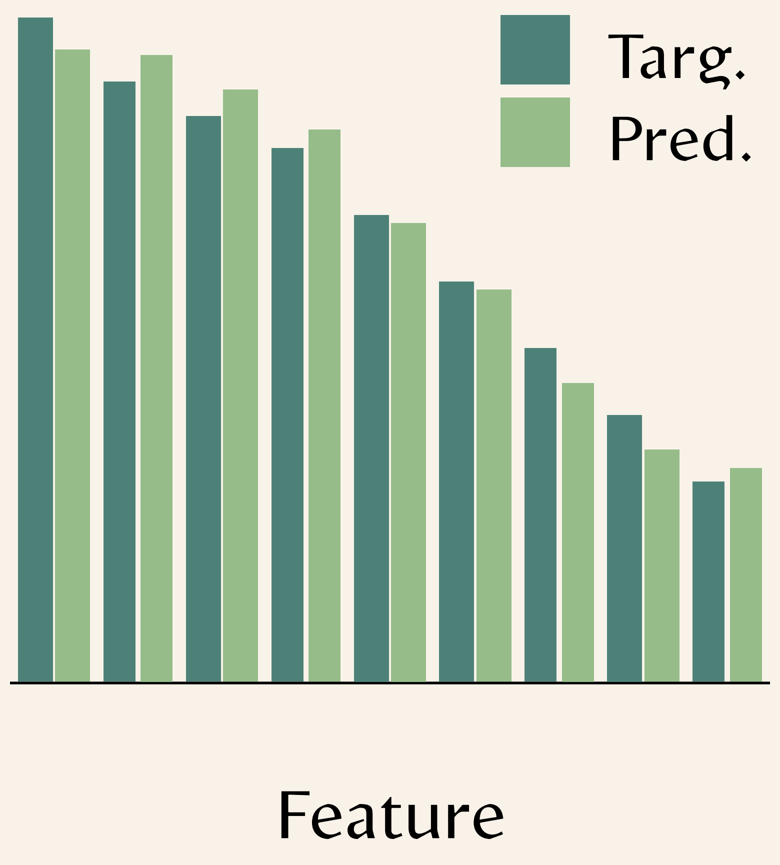
WEIZMANN INSTITUTE OF SCIENCE

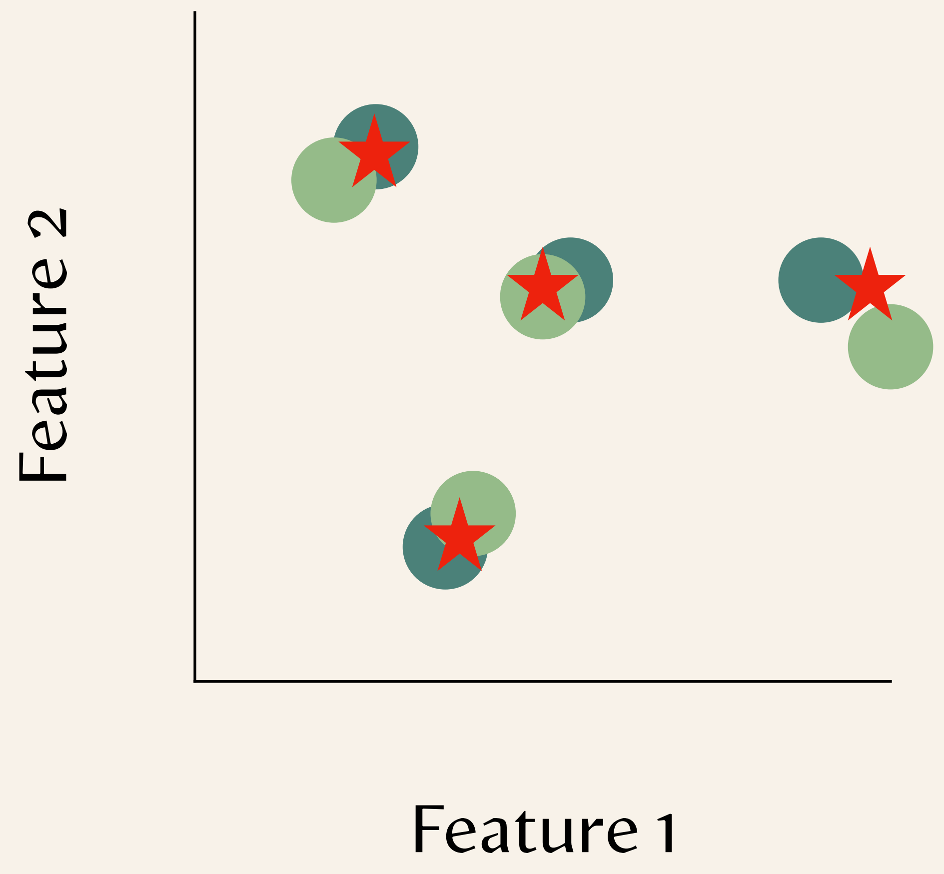
Goals

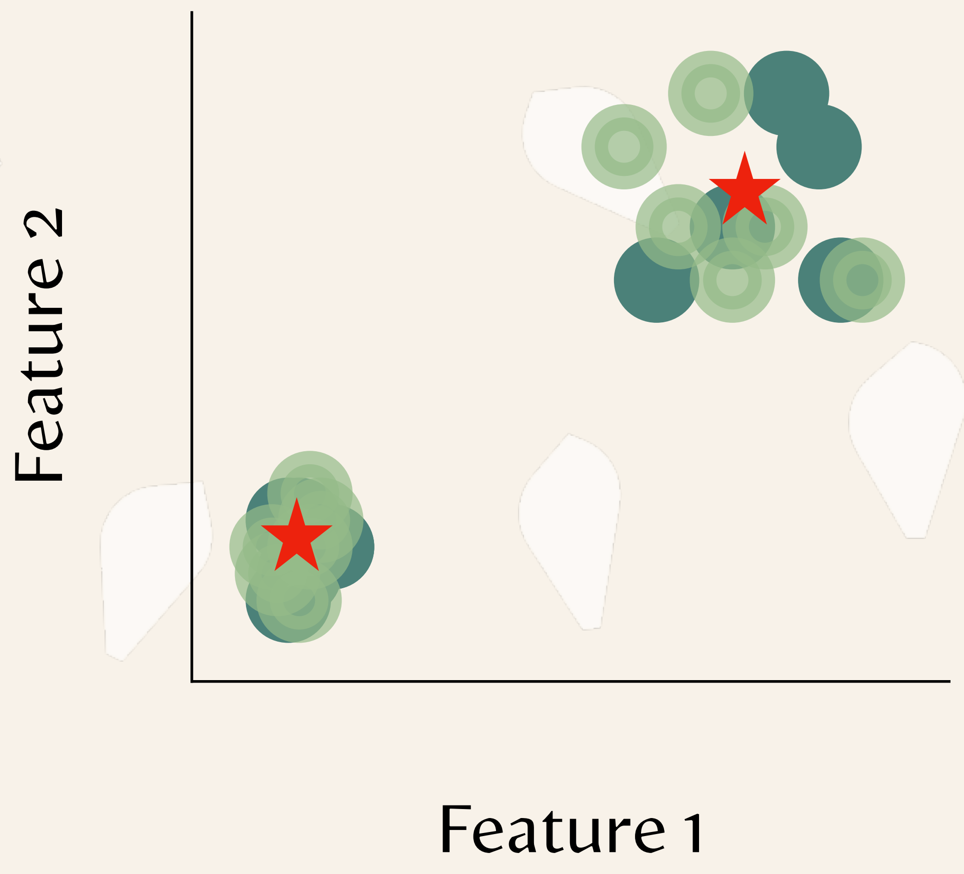
Marginal distributions

Reconstruct constituents

Resolution

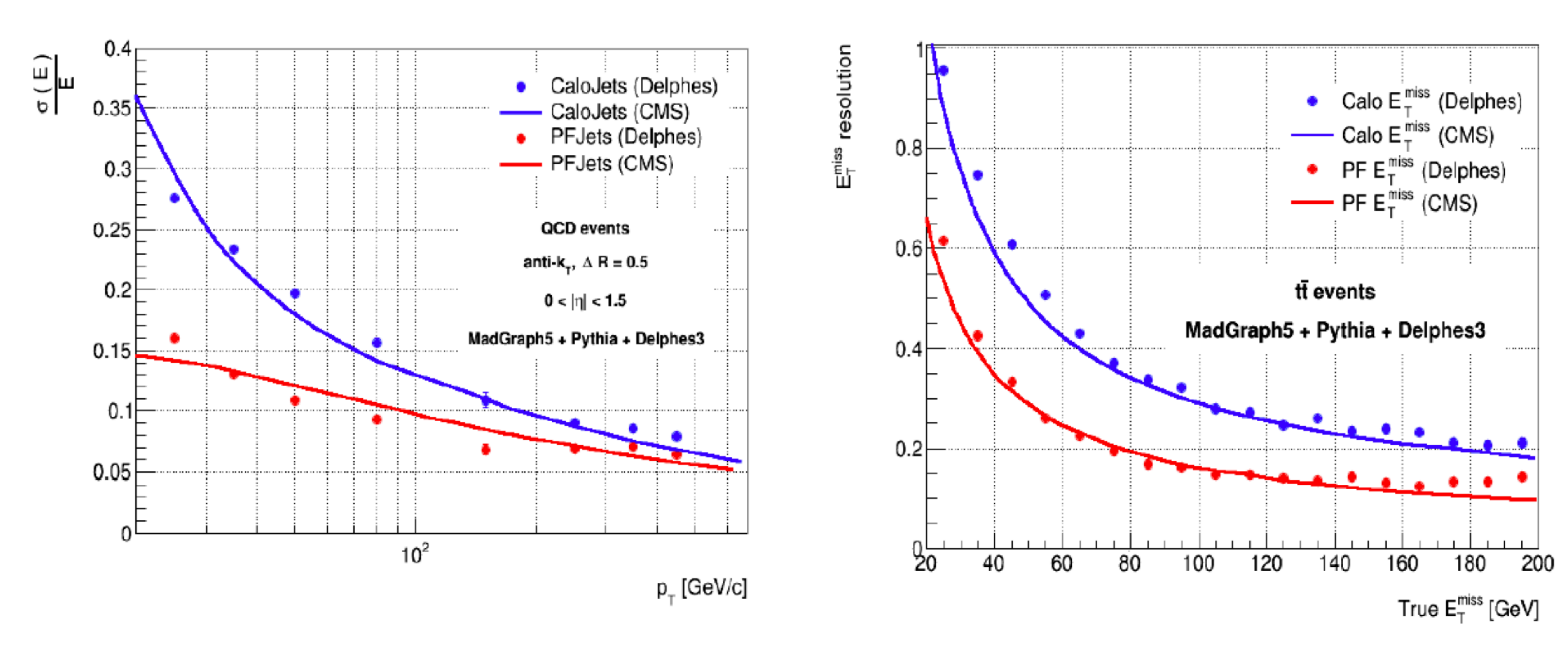






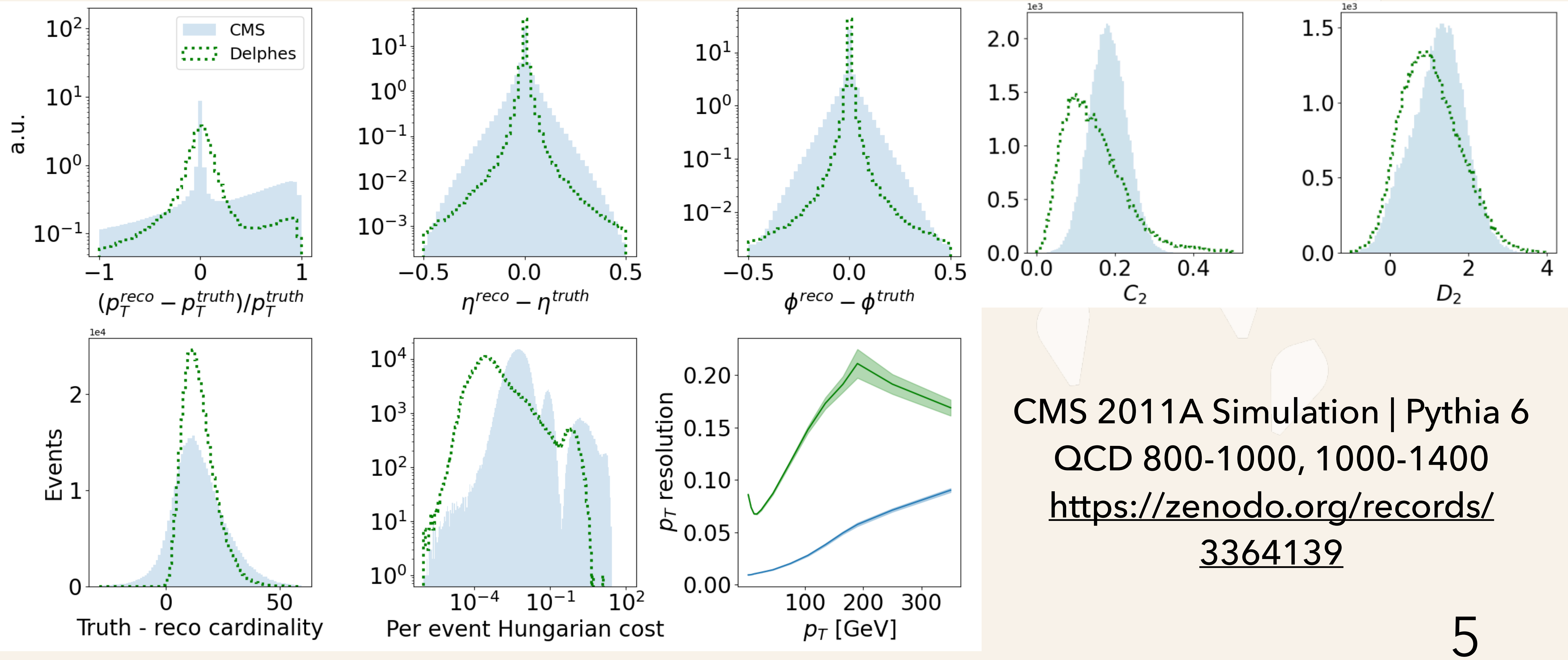
Existing approach: Delphes 3

- Public parametrized simulation
- Commonly used for research
- Very fast
- Shows good agreement of jet kinematics and resolution
- Not used by ATLAS/CMS
- Not very suitable for substructure and individual particle properties



Delphes 3

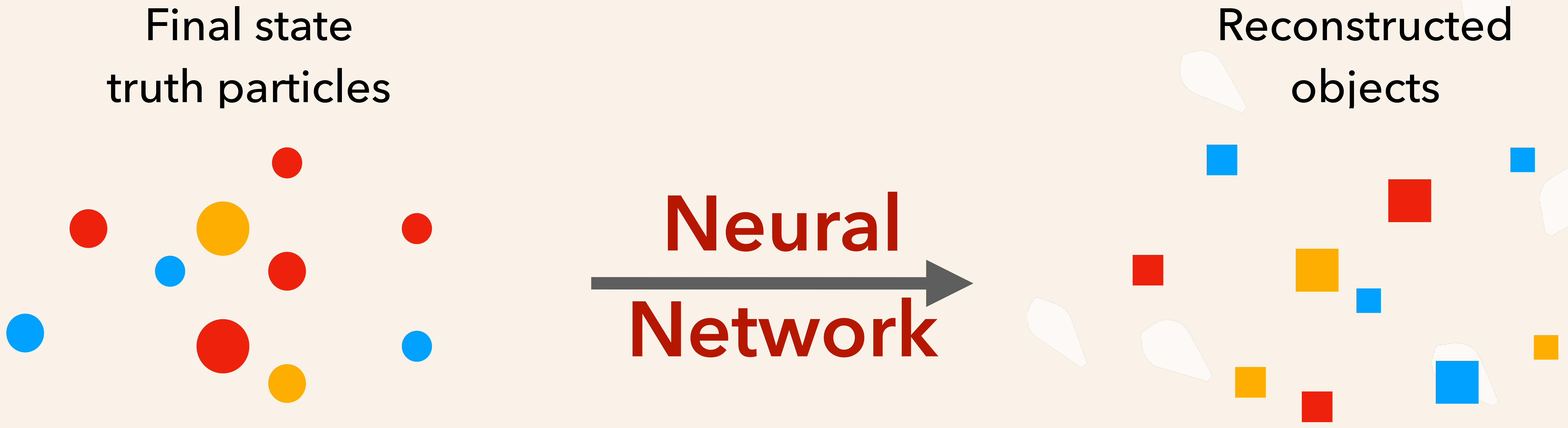
http://arxiv.org/abs/1307.6346



CMS 2011A Simulation | Pythia 6
QCD 800-1000, 1000-1400
https://zenodo.org/records/3364139

WEIZMANN INSTITUTE OF SCIENCE

ML-based approach



ML-based approach: Our journey

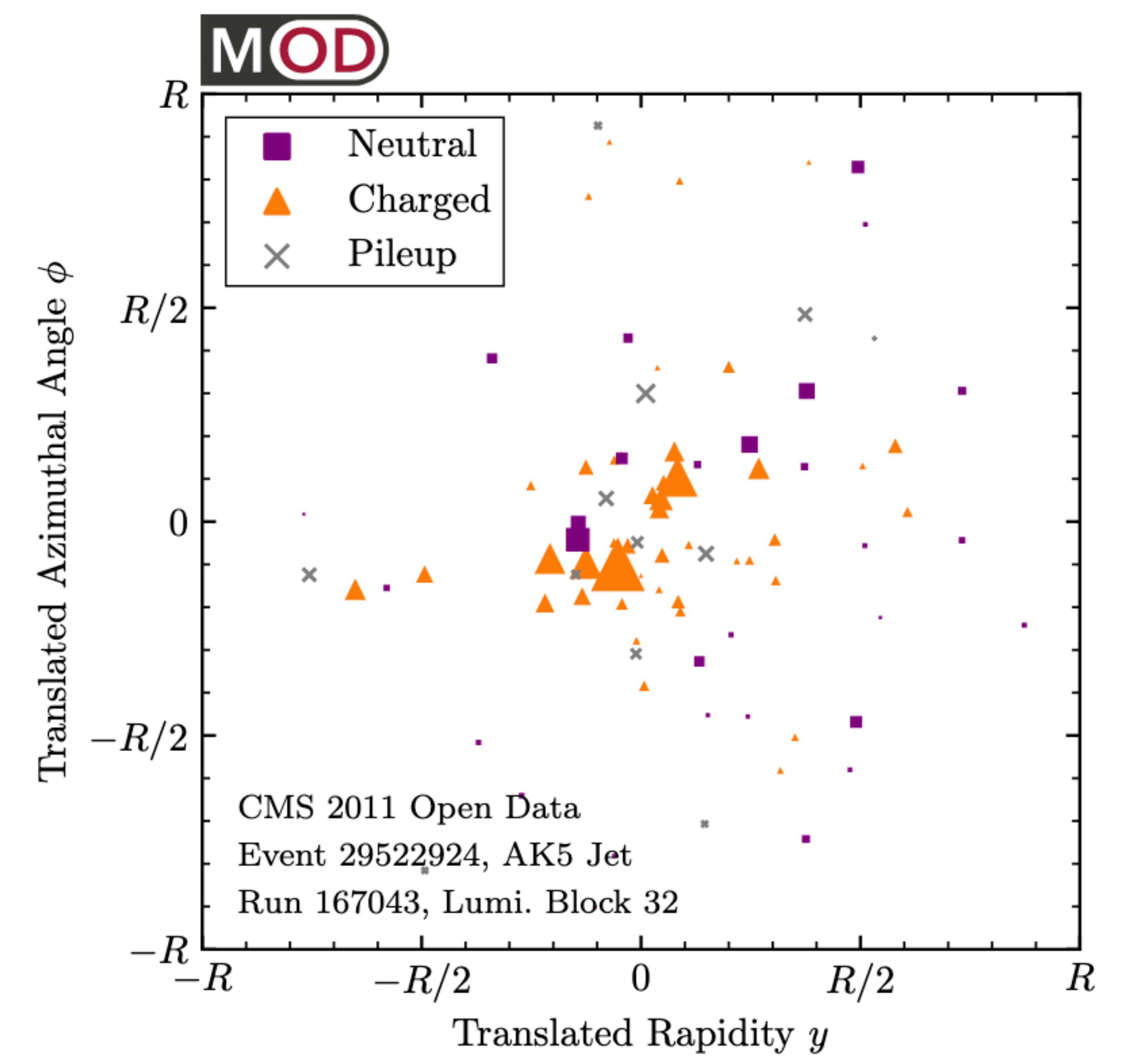
Toy model: emulated tracks	Full GEANT4 reconstruction and simulation	CMS 2011A Simulation QCD, Single jets	CMS 2011A Simulation QCD, TTbar, H4lep **Full event**
Only charged particles	Charged + Neutral	Charged + Neutral	Charged + Neutral
p_T, η, ϕ	p_T, η, ϕ, q	p_T, η, ϕ	$p_T, \eta, \phi, \vec{v}$, PID
GNN with Slot-Attention	Diffusion and Flow matching models	Flow matching	Flow matching
arXiv:2211.06406 Published in MLST	arXiv:2405.10106 Accepted in PRD	arXiv:2406.01620 Accepted in PRL	

This talk

Single-Jet (arXiv:2406.01620)

- CMS 2011A Simulation dataset
- Full CMS simulation
- QCD dijets events, jets clustered with anti-kt 0.5 $p_T > 375, |\eta| < 1.9$
- 200 particles max

p_T^{min} - p_T^{max} [GeV]	Type	Training	Testing
470 - 600	Out-of-distribution		✓
600 - 800	Out-of-distribution		✓
800 - 1000	In-distribution	✓	✓
1000 - 1400	In-distribution	✓	✓
1400 - 1800	Out-of-distribution		✓
1800 - ∞	Out-of-distribution		✓



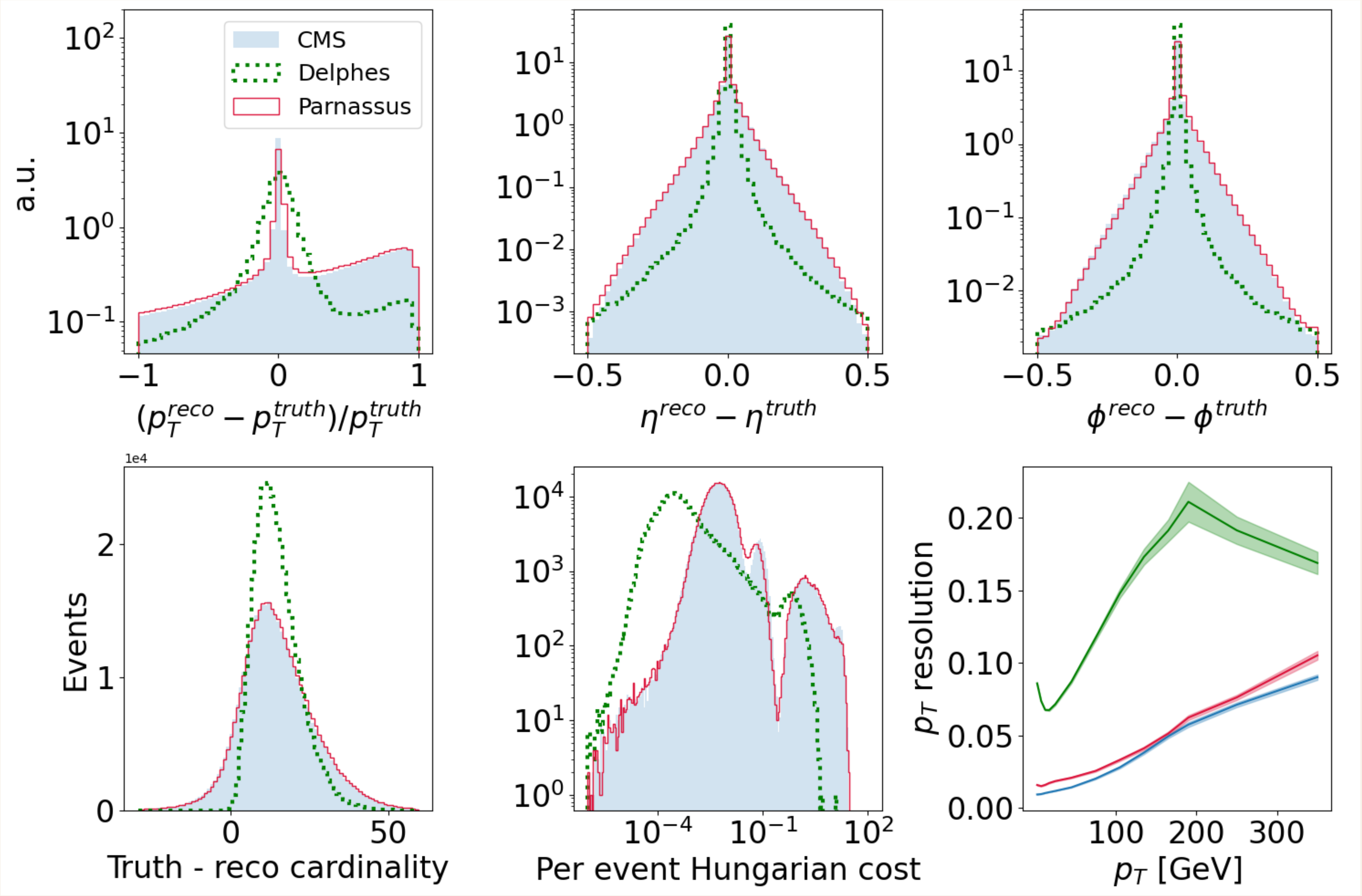
Example QCD jet (http://arxiv.org/abs/1908.08542)

Delphes dataset for comparison was simulated by us with addition of CMS PileUp Minimum bias events.

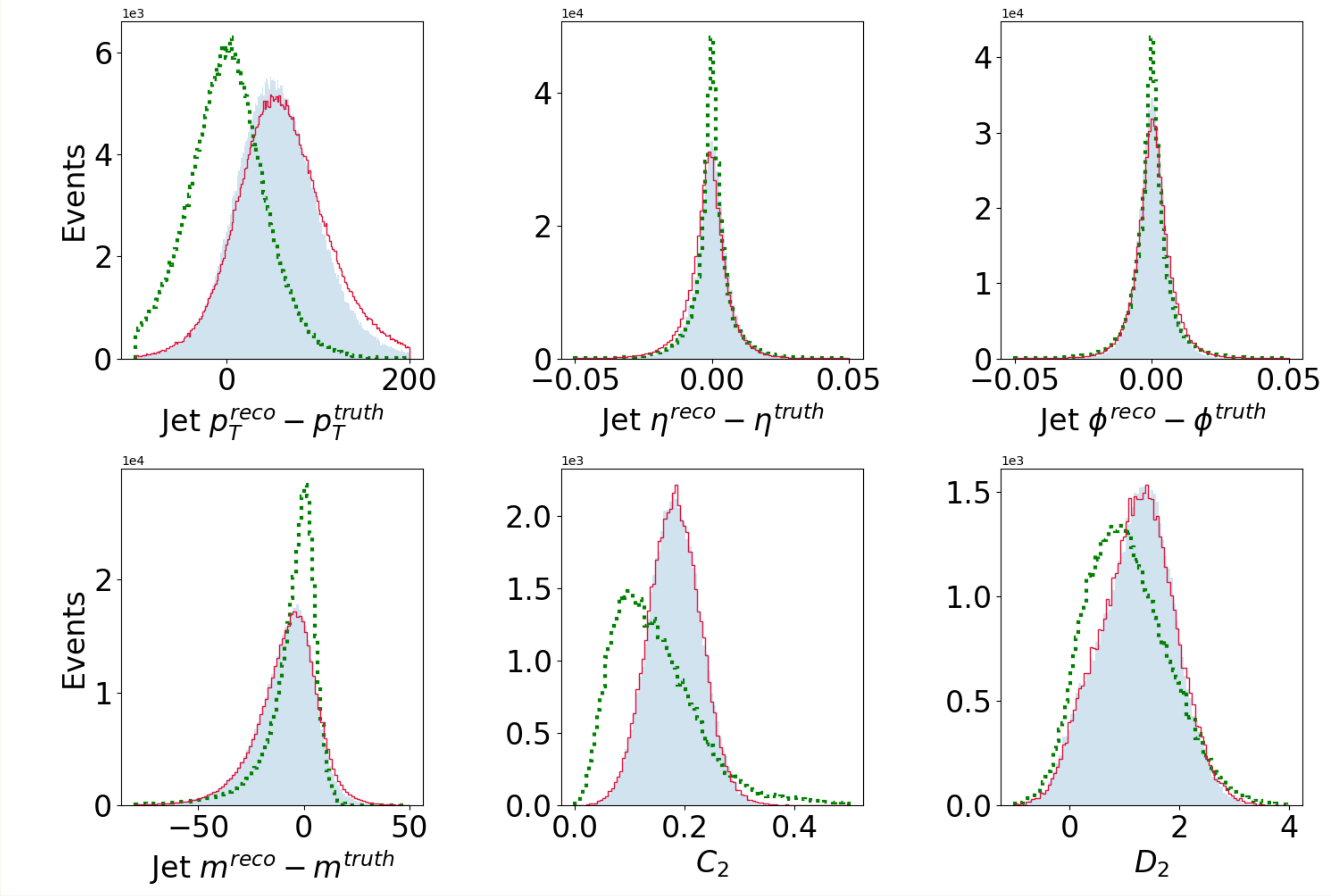
Results: Single Jets

Very good agreement with CMS Pflow

Outperform Delphes everywhere



Particle features

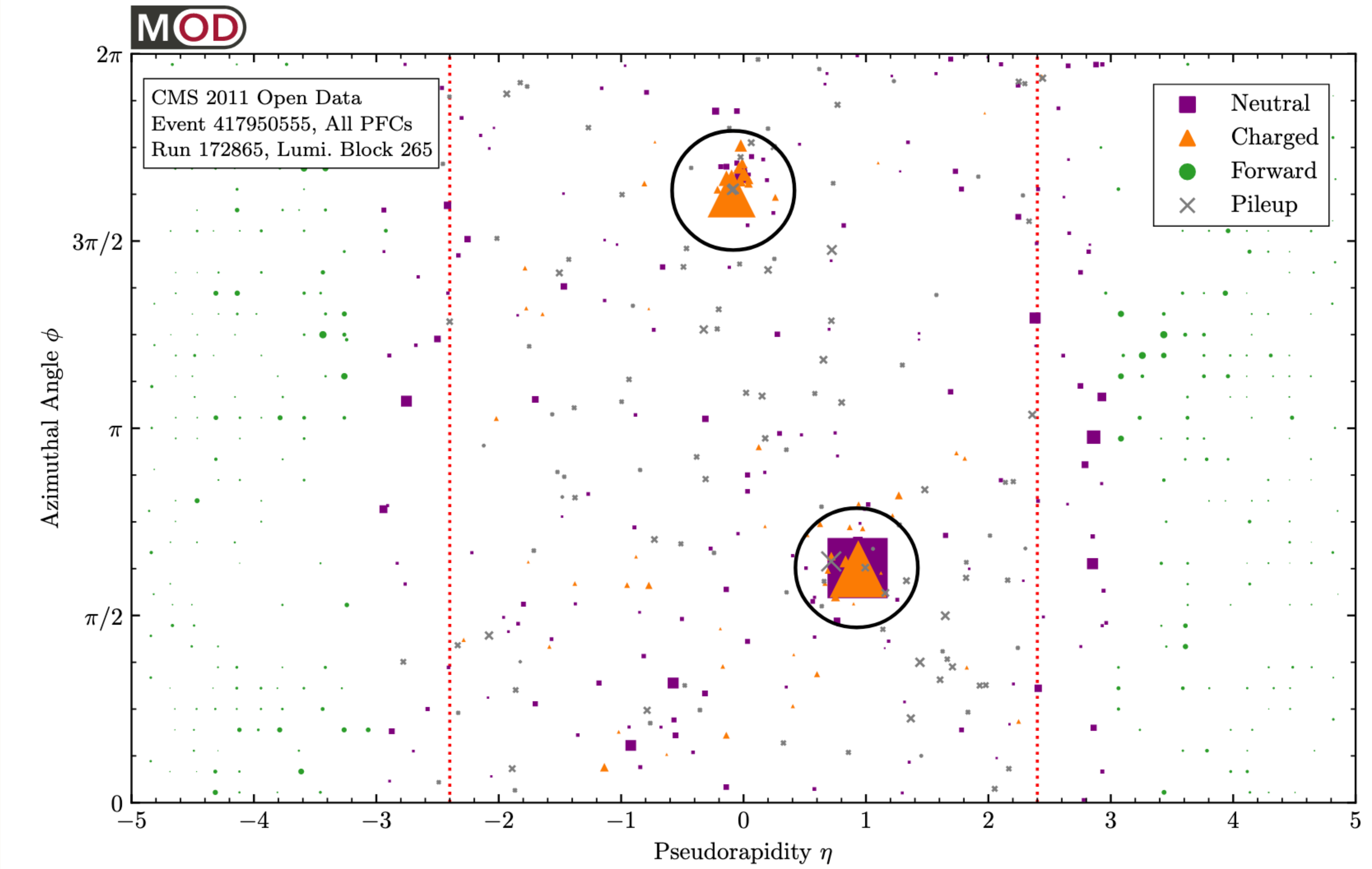


Jet features

Full event

- CMS Open Data, Simulation Datasets 2011
- **Full event**
- $p_T > 1$ GeV, $|\eta| < 2.7$ cut on PFOs and truth particles
- 3M events for training

Dataset	Training	Testing
QCD 470-600 GeV	✓	✓
TTbar	✓	✓
Higgs → 4 leptons		✓
QCD 1000-1400 GeV		✓



Example QCD event (http://arxiv.org/abs/1908.08542)

Image is for illustrative purposes, datasets were extracted by us from CMS Open Data.

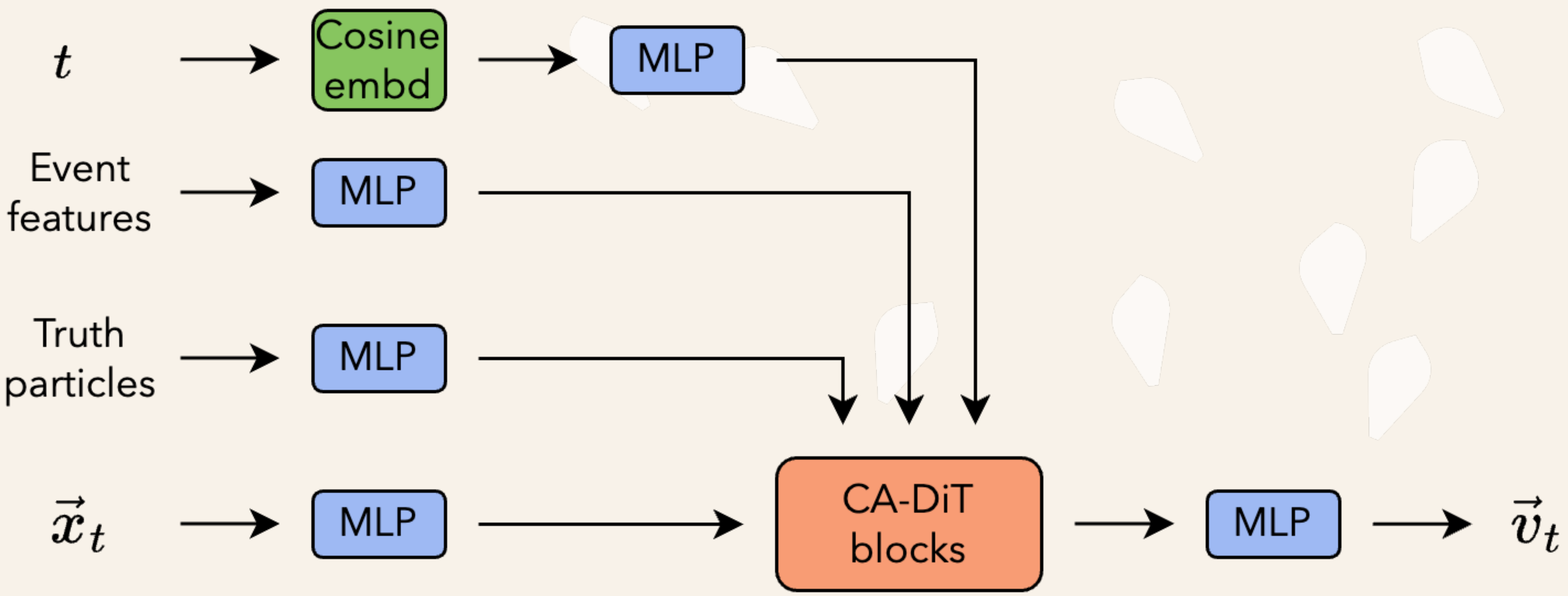
Full event: model description

- Conditional Flow Matching model
- Separate ResNet CFM network for (cardinality, $E_x^{miss}, E_y^{miss}, H_T$) prediction
- Cross-Attention Diffusion Transformer architecture for particle properties
- Maximum 400 particles
- $p_T, \eta, \phi, \vec{v}$, PID prediction

$$p_t(x|z) = \mathcal{N}\left(x \mid tx_1, (t\sigma - t + 1)^2\right)$$

$$u_t(x|z) = \frac{x_1 - (1-\sigma)x}{1 - (1-\sigma)t}$$

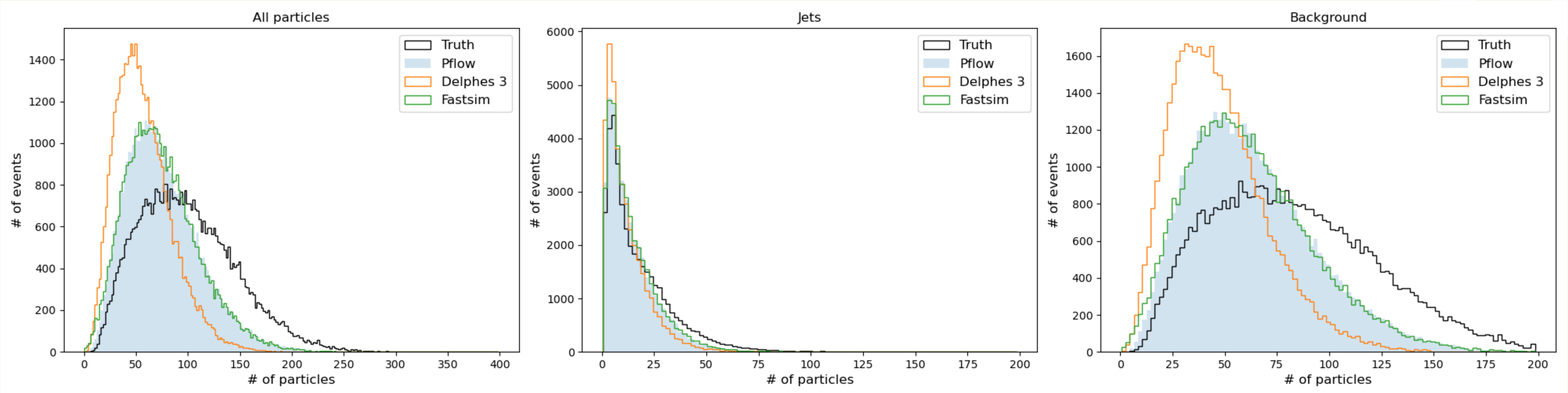
$$\mathcal{L}(\theta) = \mathbb{E}_{t,q(x_1),p_t(x|x_1)} \| v_\theta(x,t) - u_t(x|x_1) \|^2$$



WEIZMANN INSTITUTE OF SCIENCE

Results: Higgs → 4 leptons

Cardinality

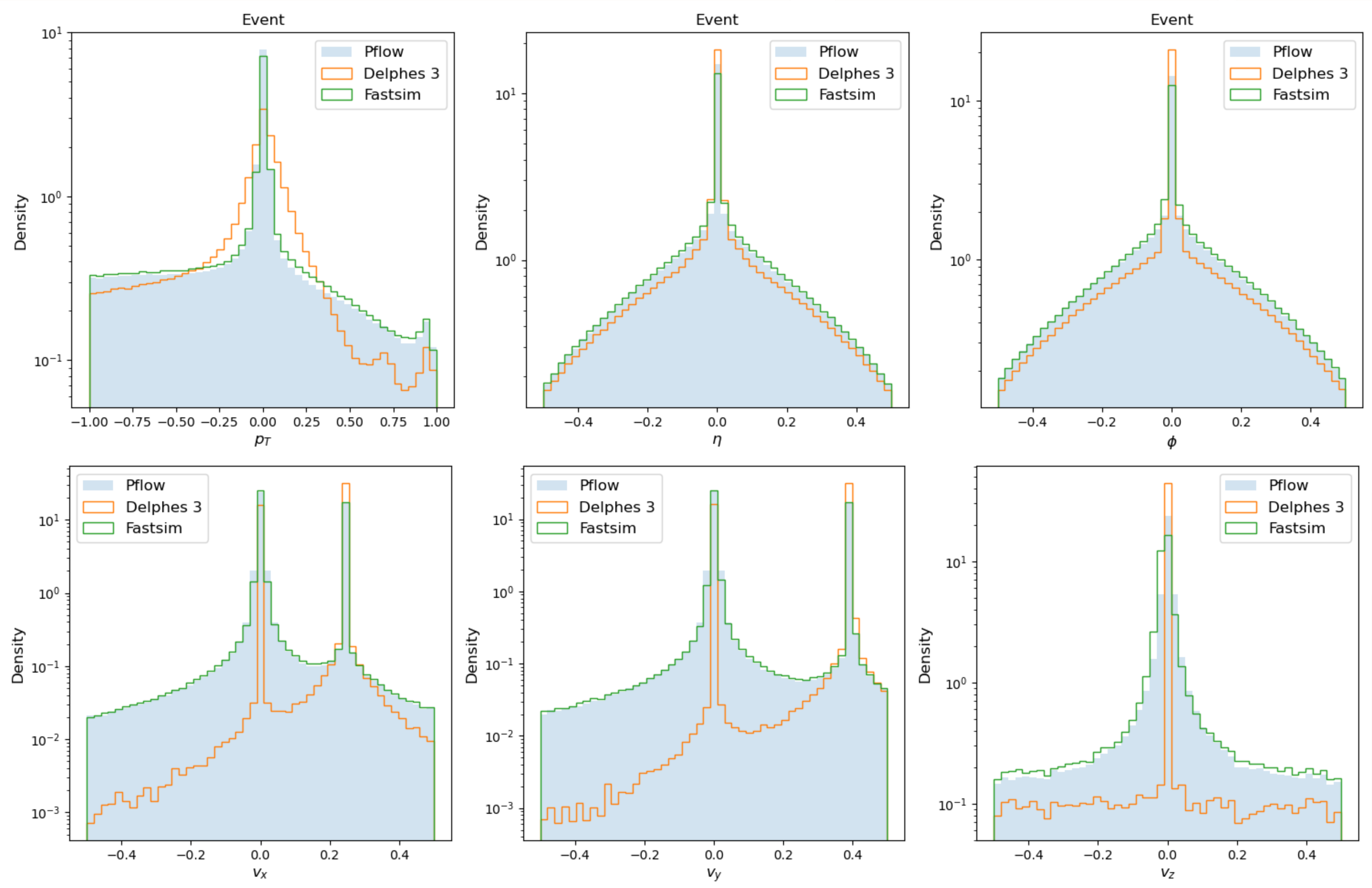


WEIZMANN INSTITUTE OF SCIENCE

Results: Higgs → 4 leptons

Residuals

- Based on Hungarian Matching between PFOs and Truth particles with ΔR metric
- Neutral PFOs are set to zero vertex

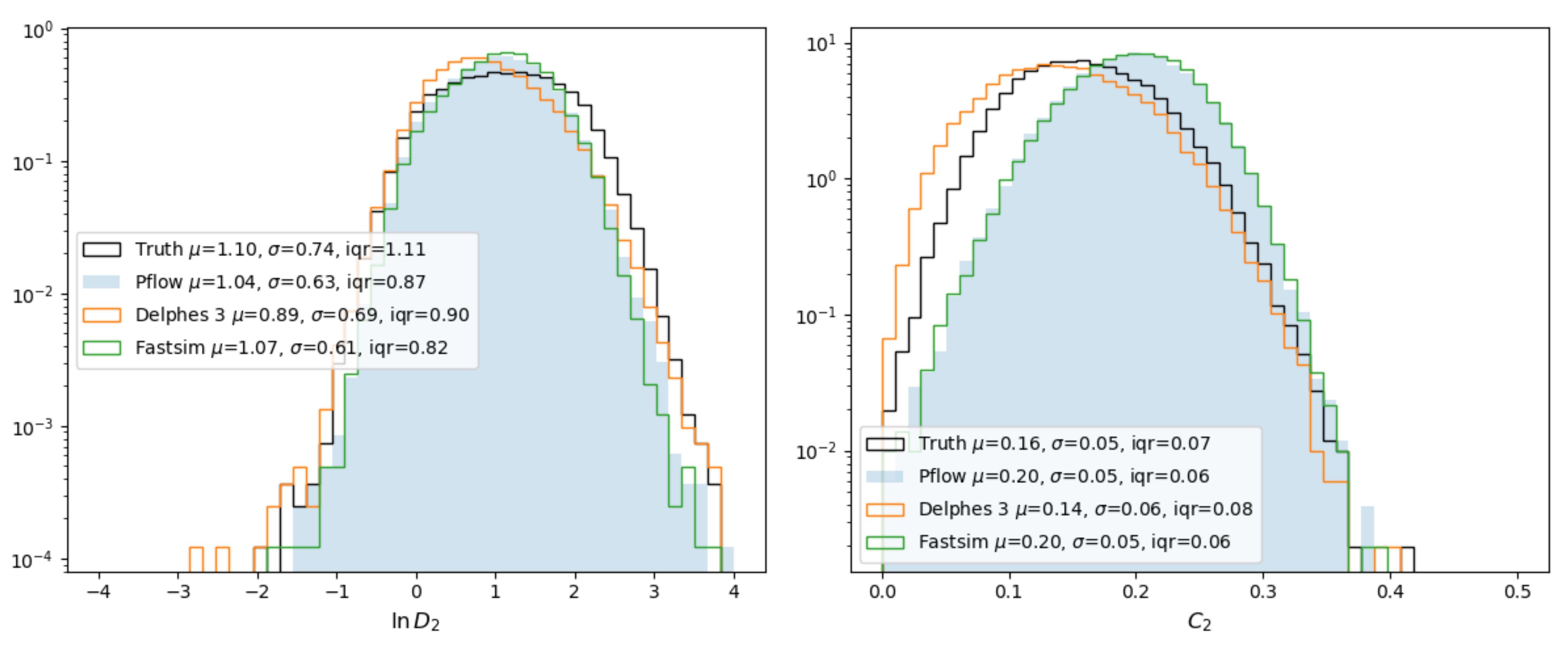


WEIZMANN INSTITUTE OF SCIENCE

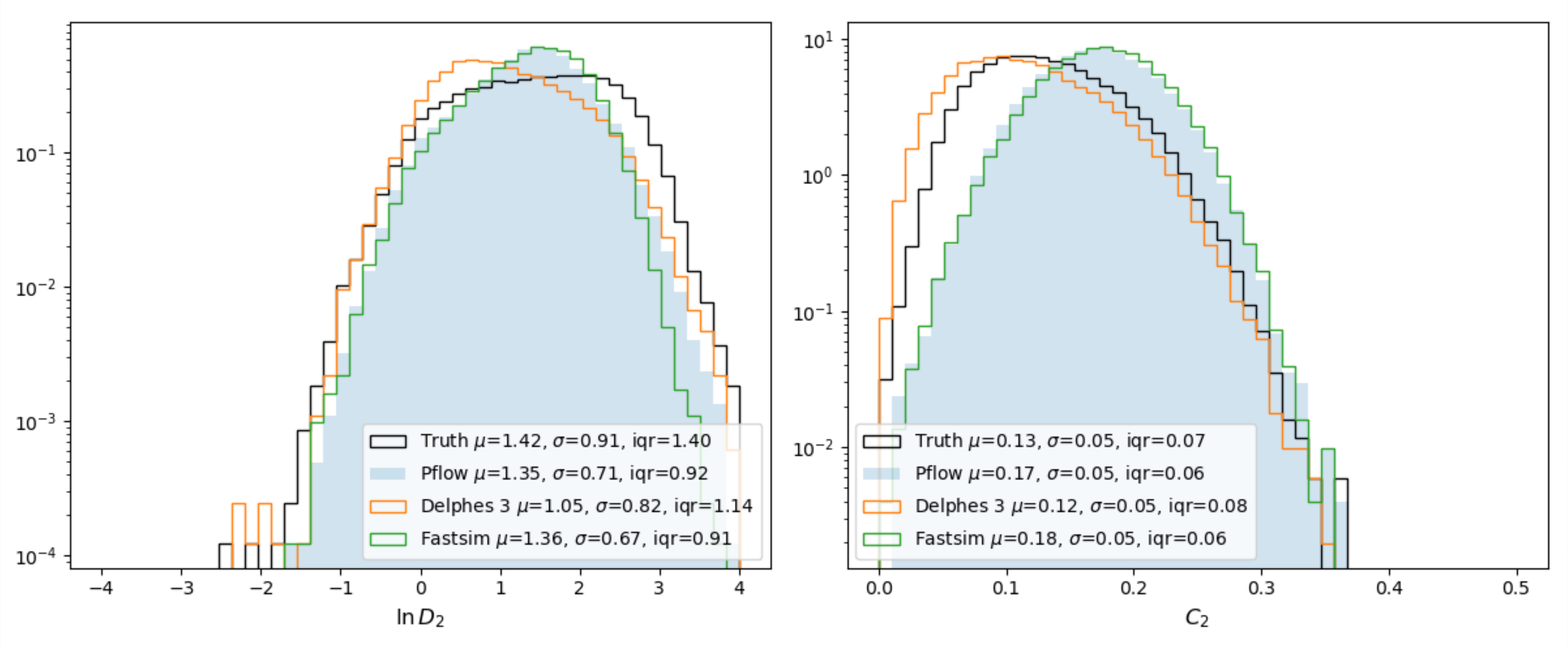
Results: Higgs → 4 leptons, TTbar, QCD

Jet substructure

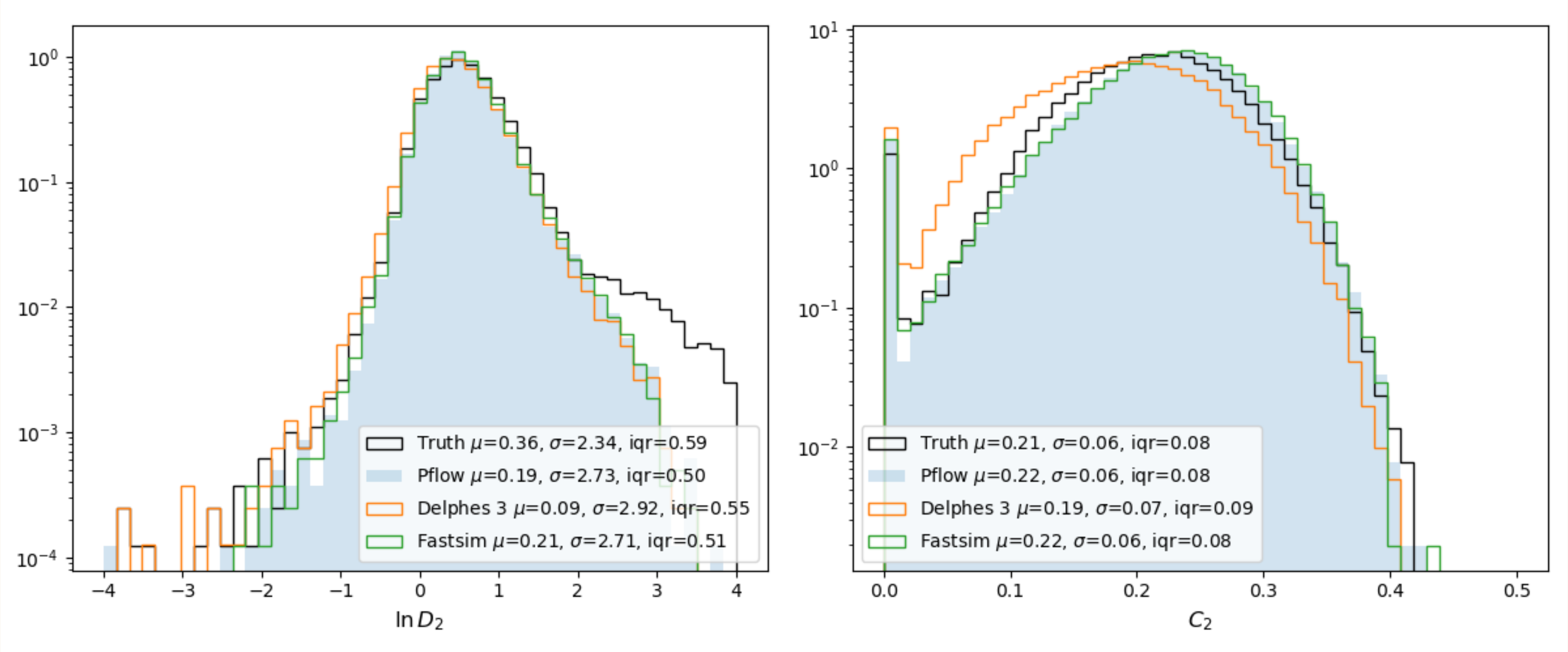
QCD 470-600 GeV



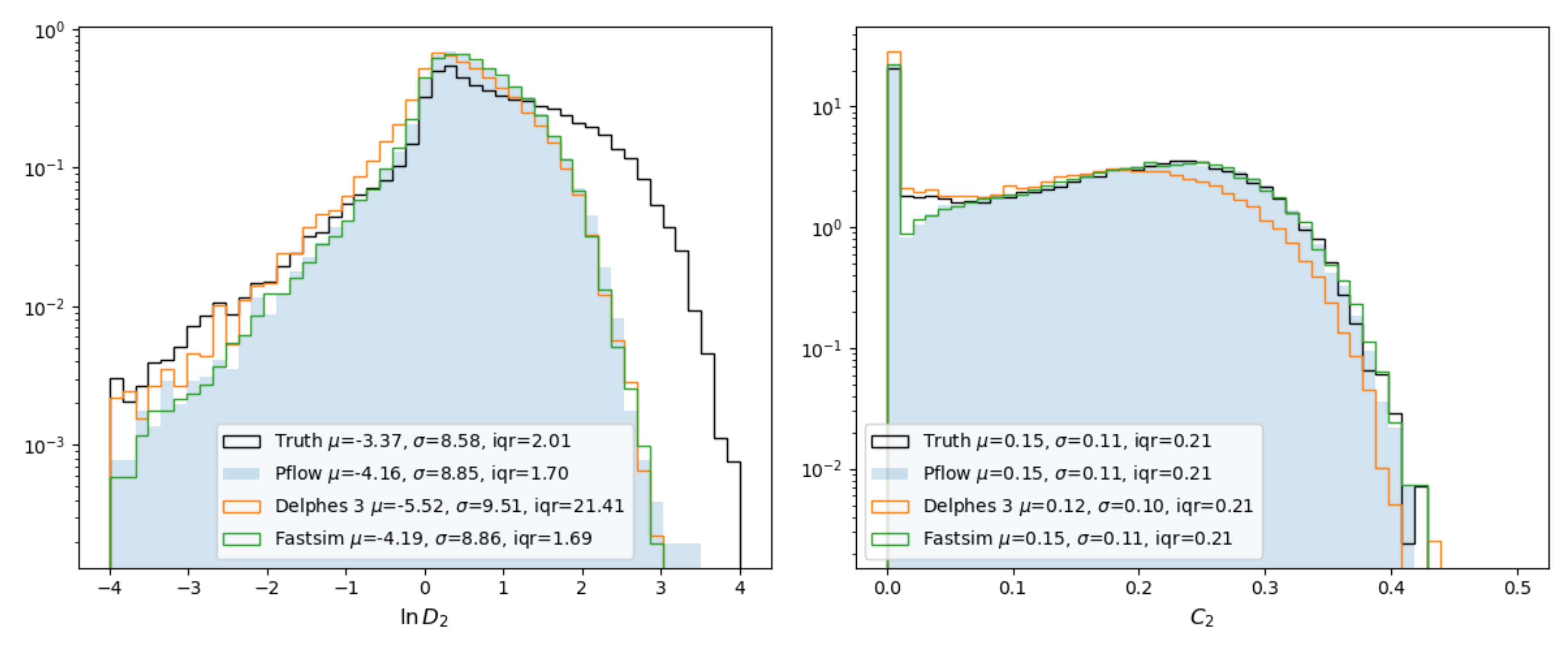
QCD 1000-1400 GeV



TTbar



Higgs → 4 leptons



WEIZMANN INSTITUTE OF SCIENCE

Summary

Conclusions

- CFM is a very powerful tool for PFOs generation
- Model is able to generalize to different processes and phase space regions
- Due to lack of truth pile-up particles model learned it implicitly
- Parnassus outperforms Delphes and is very close to CMS PFOs, especially in substructure and per-particle features

Future work directions

- Implement configurable and user-friendly interface with documentation
- Work with experiments (ATLAS, CMS) to produce and validate specific models
- Facilitate the sharing of such models to the broad physics community

https://github.com/parnassus-hep/cms-flow

WEIZMANN INSTITUTE OF SCIENCE

Backup