DEPARTMENT OF INFORMATICS & TELECOMMUNICATIONS
HELLENIC REPUBLIC
National and Kapodistrian University of Athens
EST. 1837

Particle-based Simulation of Jets at the LHC with Variational Autoencoders

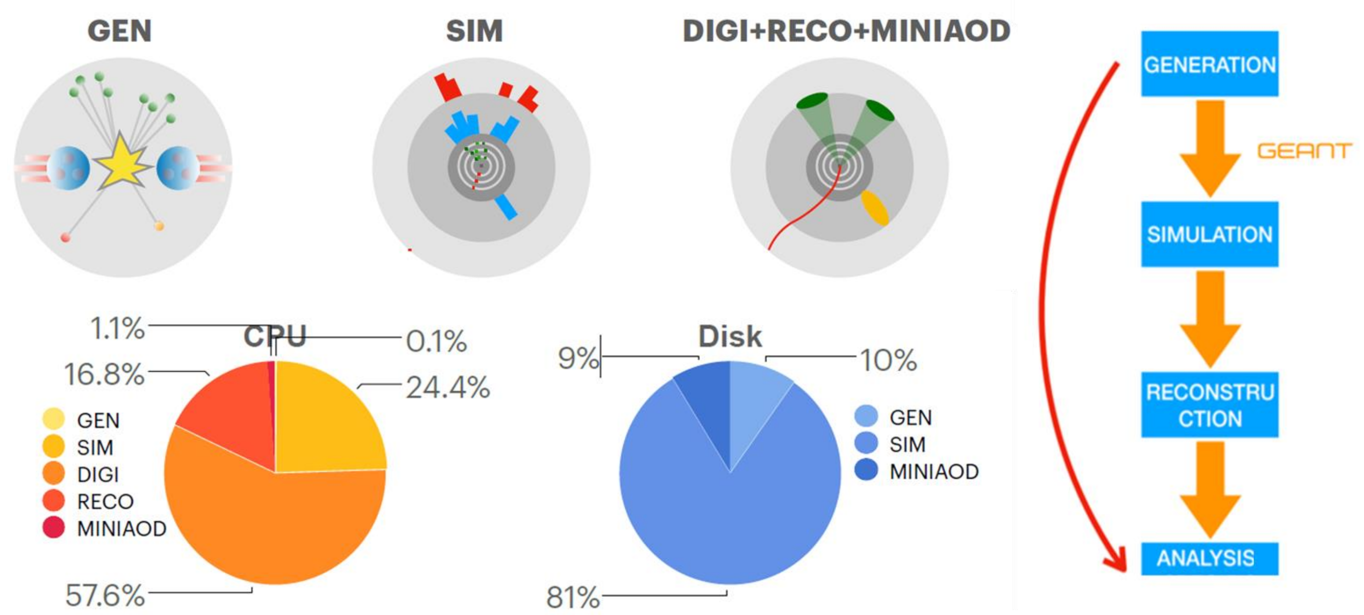
Mary Touranakou [1, 2], Breno Orzari [3], Maurizio Pierini [1], Thiago Tomei [3], Dimitrios Gunopulos [2], Raghav Kansal [4], Javier Duarte [4], Jean-Roch Vlimant [5]

[1]CERN, [2]NKUA, [3]SPRACE-UNESP, [4]UCSD, [5]Caltech – mary.touranakou@cern.ch , breno.orzari@cern.ch

INTRODUCTION

- Simulated events are of high importance for HEP experiments
- Existing workflows are costly in time and storage resources

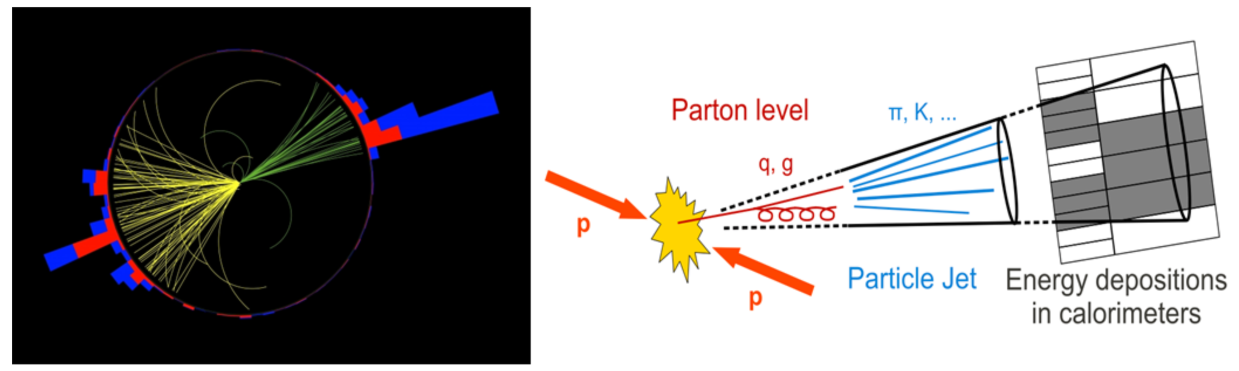
→ Deep Generative models proposed for speedup [1] (ms/event instead of s/event)



DATASET

Jets, high-level objects of particular interest: **sparse**, **highly granular**, **irregularly distributed** in space, can be represented as sparse sets of constituents.

We define each jet as a list of particles where each particle is characterized by its cartesian momenta p_x, p_y, p_z and we employ different datasets per use case.



Jet at CMS

p_{x_1}	p_{x_2}	p_{x_3}	...	p_{x_N}
p_{y_1}	p_{y_2}	p_{y_3}	...	p_{y_N}
p_{z_1}	p_{z_2}	p_{z_3}	...	p_{z_N}

Jet represented as a list of constituents

Datasets [4] [5]

PARTICLE-BASED SIMULATION OF JETS WITH VAEs – TWO ALTERNATIVE PARADIGMS

I. **VAE for Collision Simulation (Parton → Reconstructed Jet Constituents)**: Use a trained VAE decoder as a generator

II. **VAE for Detector Parametrization (Generated Jet Constituents → Reconstructed Jet Constituents)**: Use end-to-end VAE to model the detector response

- Both with a permutation-invariant reconstruction loss to increase accuracy and impose physics constraints to the model

Total loss [2]:

$$L^{VAE} = \beta D_{KL} + (1-\beta)\left[L^{rec} + \left(m^{jet} - \hat{m}^{jet}\right)^2 + \left(p_T^{jet} - \hat{p}_T^{jet}\right)^2\right]$$

with differing values of beta per use case. Jet transverse momentum and jet mass regression terms added to the loss to help learning the collective kinematic properties of the particle cloud.

Permutation-invariant Chamfer distance [3] of the particles' features representing the input and output jets as sets of particles :

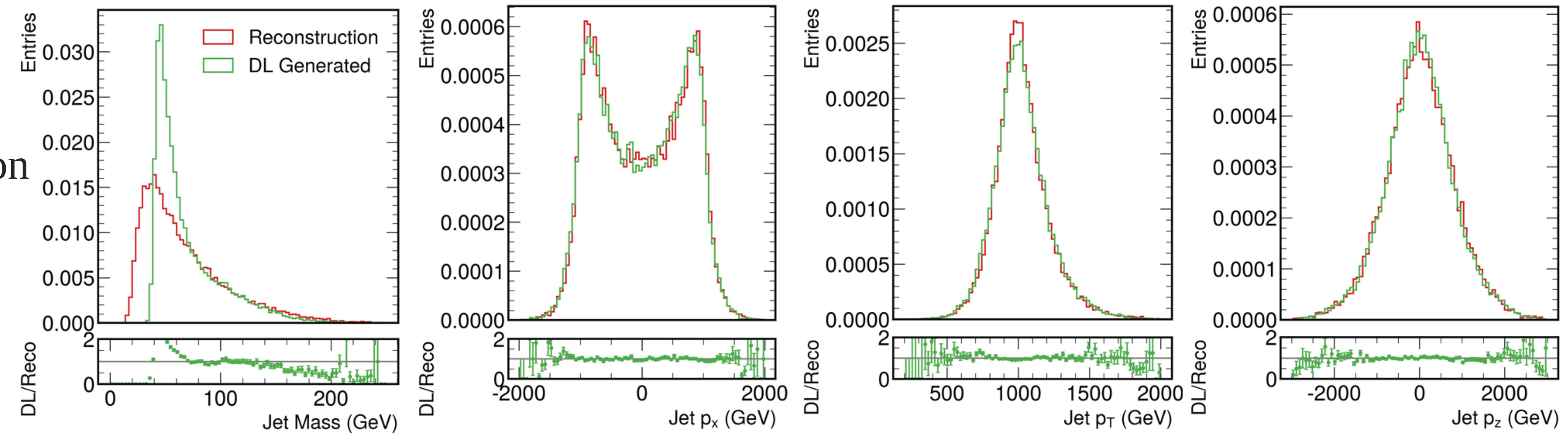
$$L^{rec} = \sum_i \min_j (p_i - \hat{p}_j)^2 + \sum_j \min_i (p_i - \hat{p}_j)^2$$

RESULTS

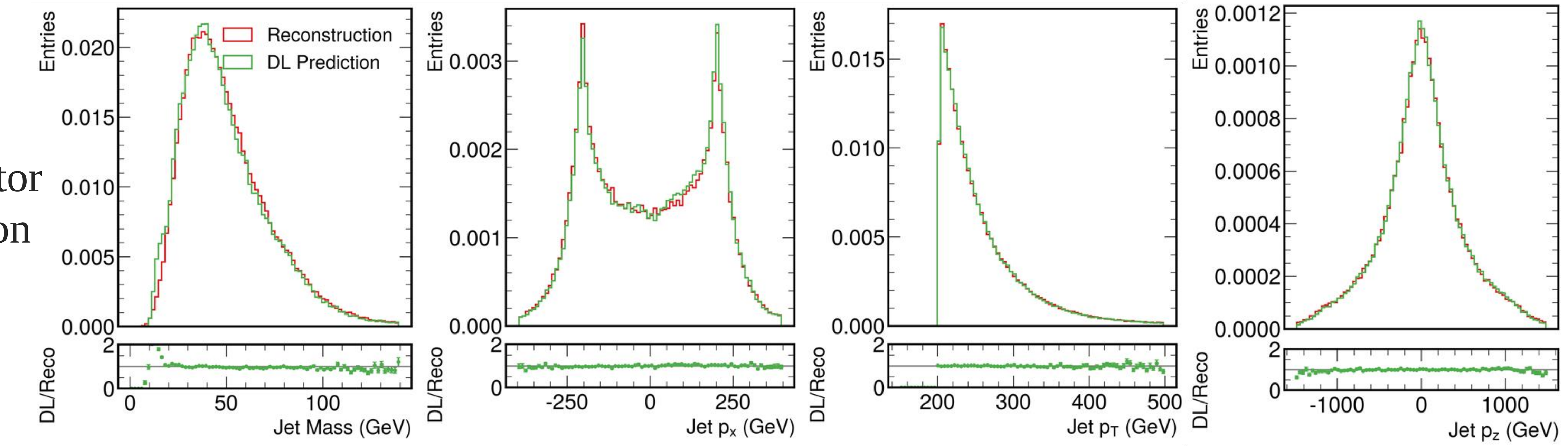
Model evaluation with Earth Mover's Distance (EMD) on jet kinematics distributions.

Comparison of jet features distributions for the target (Reconstruction) and the output (DL Prediction/ DL Generation):

VAE for collision simulation



VAE for detector parametrization



CONCLUSION

Presented two alternative use cases of VAEs for particle-based simulation of jets at the LHC with:

- a permutation-invariant reconstruction loss
- jets represented as lists of particles

Potential to bypass the detector simulation, and the event reconstruction steps of a traditional event processing to speed up the events generation workflow.

Future work will focus on:

- Using normalizing flows [6] to acquire a more appropriate posterior to sample from
- Leveraging the capabilities of VAEs with graph neural networks
- Working with varied length inputs that would better fit the nature of the data

REFERENCES

[1] Chen, C., et al. "Analysis-Specific Fast Simulation at the LHC with Deep Learning." Computing and Software for Big Science 5.1 (2021): 1-17.

[2] Higgins, I., et al. "β-VAE: Learning basic visual concepts with a constrained variational framework." In: 5th International Conference on Learning Representations (2017).

[3] Barrow, H. G. et al. "Parametric correspondence and Chamfer matching: Two new techniques for image matching."

[4] Pierini, M. et al. "HLS4ML LHC Jet dataset (100 particles)."

[5] Sjöstrand, Torbjörn, Stephen Mrenna, and Peter Skands. "PYTHIA 6.4 physics and manual." Journal of High Energy Physics 2006.05 (2006): 026.

[6] Kobyzev, I., Prince, S., and Brubaker, M., "Normalizing flows: An introduction and review of current methods."