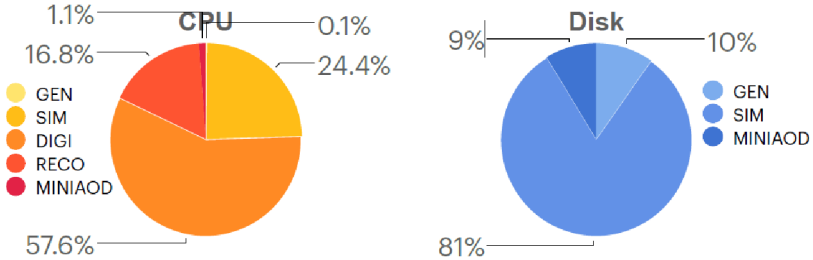
Particle-based Fast Jet Simulation at the LHC with Variational Autoencoders

Mary Touranakou, Nadezda Chernyavskaya, Javier Duarte, Dimitrios Gunopulos, Raghav Kansal, Breno Orzari, Maurizio Pierini, Thiago Tomei, Jean-Roch Vlimant

5th Inter-experiment Machine Learning Workshop - 10/05/2022

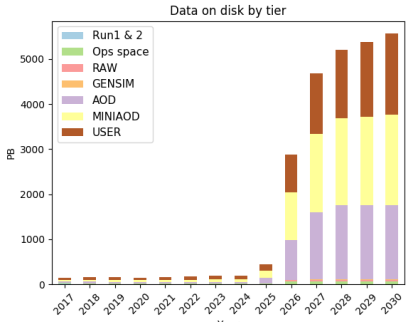
Motivations

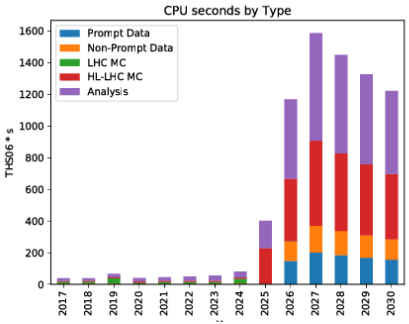
- HEP data analysis heavily relies on the production and the storage of large datasets of simulated events
- Current HEP solutions are very accurate, but also very slow
  - Costly in time resources
  - Costly in storage (as data needs to be saved till researchers need it)
  - Cannot be used as on-demand services



Motivations

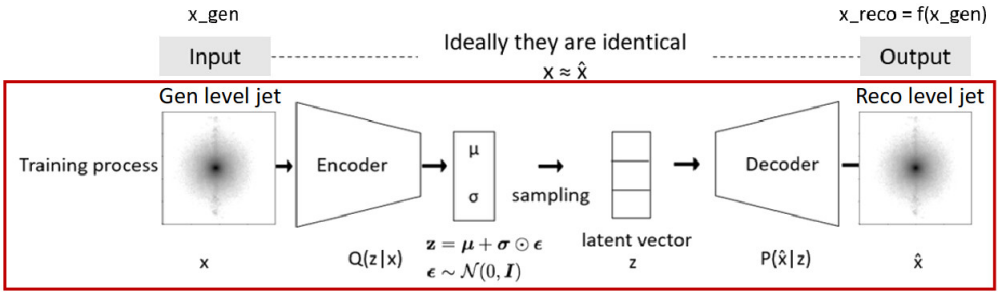
- Foreseen upgrade for the High-Luminosity phase of the LHC will increase the number of collisions and the need of more accurate, synthetic data for analysis
- Current solutions do not scale with fixed budget → Deep Generative models proposed as a tool for speeding up HEP workflows





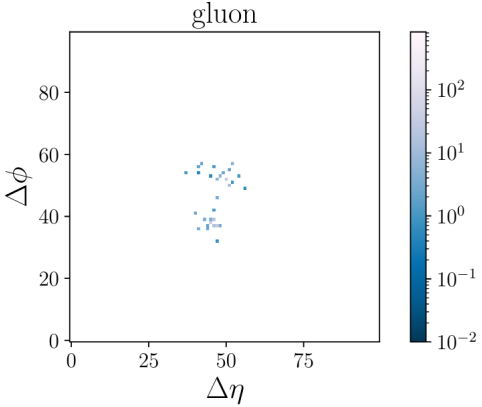
Particle-based Fast Jet Simulation with Variational Autoencoders

- Starting from a simulation of the jet before detector effects, we train a Variational Autoencoder (VAE) to return the corresponding list of constituents after detection
- Run end-to-end VAE (from GEN to RECO) to learn a parametric description of the detector response



Data Representation of Jets

- Jets can be characterized as sparse sets of particles that are intrinsically unordered
- Jets can be represented as a list of constituents characterized by their momenta:
  - In cartesian coordinates $p_x,\ p_y,\ p_z$
  - In hadronic coordinates $p_T,\ \eta,\ \phi$
- Although, sometimes, an ordering might be given to the data (e.g. ordering particles by decreasing p_T), it is also important to preserve the permutation invariance in it (depending on application-specific requirements)

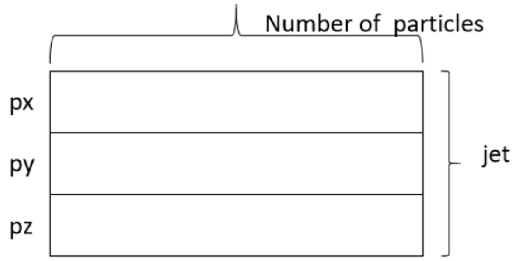


Graphical representation of a gluon jet

arXiv:1908.05318

Data Representation of Jets

- We represent jets as a list of N highest p_T constituents, characterized by their momenta p_x, p_y, p_z. If less than N particles are present in the jet, zero-padding is applied.
  - Avoids sparsity of data dealing only with useful information
  - Fits well on the convolutiuonal layers of our network
- We apply feature-dependent standardization such that each feature (p_x, p_y, p_z) has zero mean and unit variance.
- Our dataset consists of W boson jets with up to 50 particles.



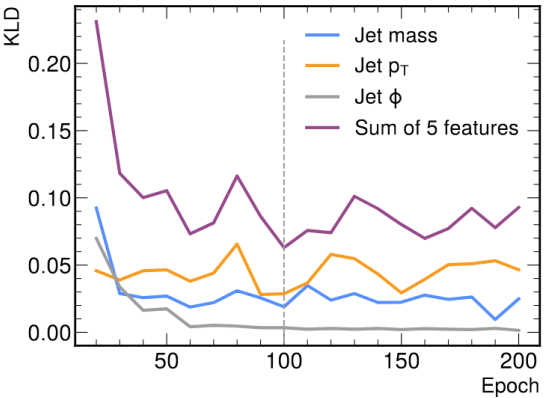
Reconstruction Loss Function

- Use of a permutation-invariant, pairwise Nearest Neighbour Distance (NND) known as the Chamfer distance (arXiv:1906.02795) for the reconstruction loss
- Customization of the reconstruction loss by adding two extra terms, the jet mass and the jet p_T to enforce the model to learn jet kinematics
- The jet mass and the jet p_T (target and reconstructed) are computed from the sum of the momenta of the particles on the jet

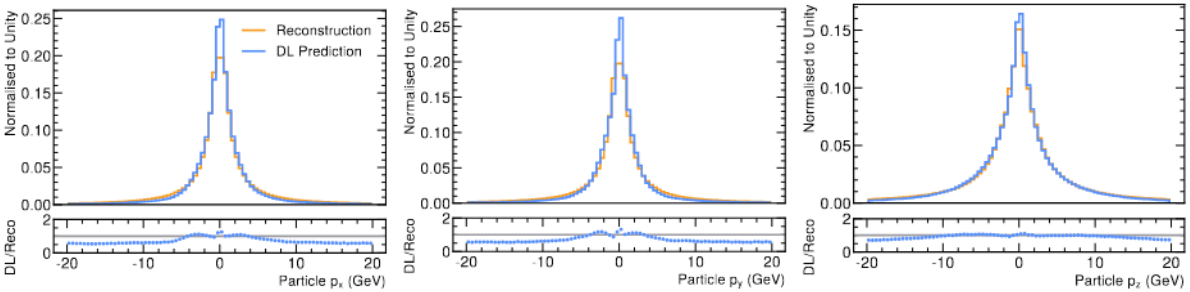
$$
\begin{aligned}
L_{reco} &= \frac{1}{N}\sum_{i=0}^{N}\left[\sum_{j}\min_{k}(p_j^i-\hat{p}_k^i)^2+\sum_{k}\min_{j}(p_j^i-\hat{p}_k^i)^2\right. \\
&+ \left.\alpha_m(m_{\text{jet}}^i-\hat{m}_{\text{jet}}^i)^2+\alpha_{p_T}(p_{\text{T}}^{\text{jet},i}-\hat{p_{\text{T}}}^{\text{jet},i})^2\right]
\end{aligned}
$$

Model Detection & Evaluation

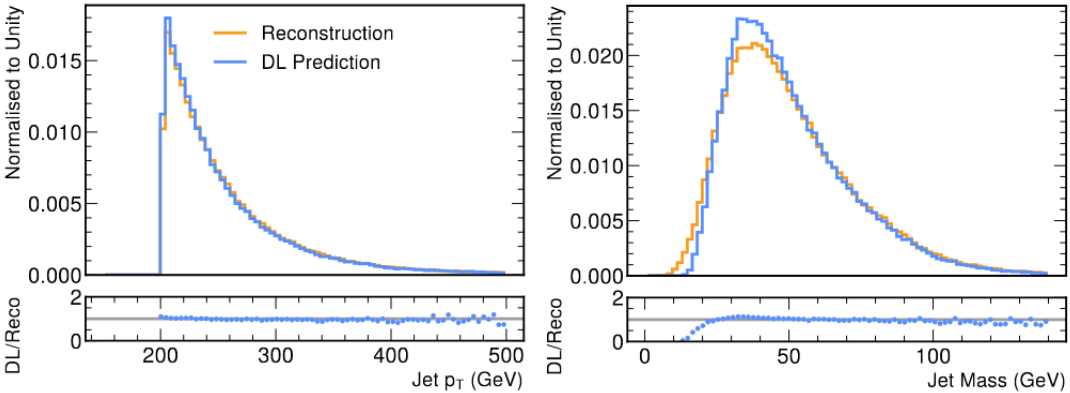
- KL divergence (KLD) between the output (DL) and the target (Reco) jet features histograms used as a quantitative metric to select the best model for analysis
- Define as best the model that satisfies joint conditions for KLD metrics:
  - minimizes KLD sum of all jet features
  - keeps the KLD sum of jet mass and jet pT below an empirically-set threshold
- KLD as a figure of merit to find a compromise between the multi-objective reco loss



Results: Constituents Properties



Results: Jets Properties

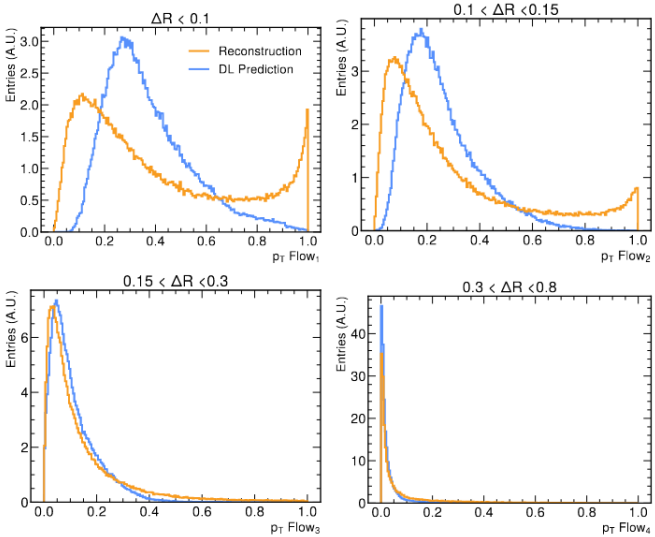


Results: Jets Substructure

- Jet substructure computed using momentum flows

$$Flow_n = \sum_p \frac{p_{\mathrm{t}}^p}{p_{\mathrm{T}}^{\mathrm{jet}}}$$

- The sum runs over all particles with $\Delta R = \sqrt{\Delta\phi^2 + \Delta\eta^2}$ from jet axis, within $(n-1)/4 \times R$ and $n/4 \times R$



Conclusions and Perspectives

- CNN VAE
  - Working baseline model with a permutation-invariant loss
  - Achieved fast-simulation precision on jets (constituents) properties
    - Retained fast simulation time of 10^{-3} s per event
    - Could already be used to generate events for most LHC physics analysis
  - Main challenge: fixed-length tensors and zero-padding
  - Possible solutions: learnable masking feature to deal with zero-padding, or Graph NNs.
- Graph VAE
  - Moving into a graph implementation of the VAE
  - Permutation-invariant architecture + permutation-invariant loss $\rightarrow$ Replace CNN VAE with a Graph VAE + Chamfer loss
  - Varied length tensors instead of fixed length $\rightarrow$ Avoid zero-padding

Backup

VAE Loss Function

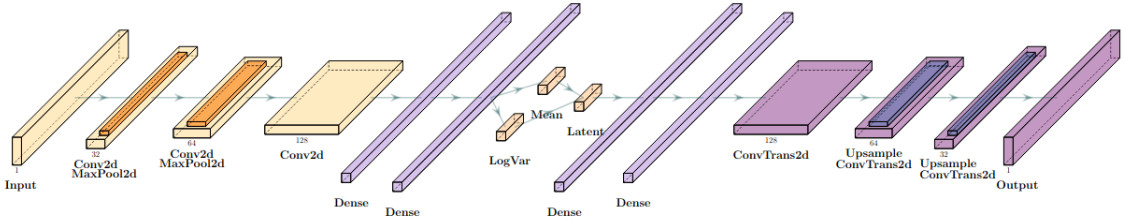
- The loss function of a VAE consists of two terms:
  - The reconstruction loss (e.g. traditionally a generic loss function such as the MSE) – the error between the original samples (inputs/targets) and the produced outputs (reconstructed inputs).
  - The Kullback-Leibler (KL) divergence used as a loss function between the encoder's distribution $q_\phi(z|x)$ and the $p_\theta(z)$ that optimizes the probability distribution parameters (μ and σ) to closely resemble those of the target distribution.

$$L^{VAE} = \beta D_{KL} + (1 - \beta) L_{reco}$$

where

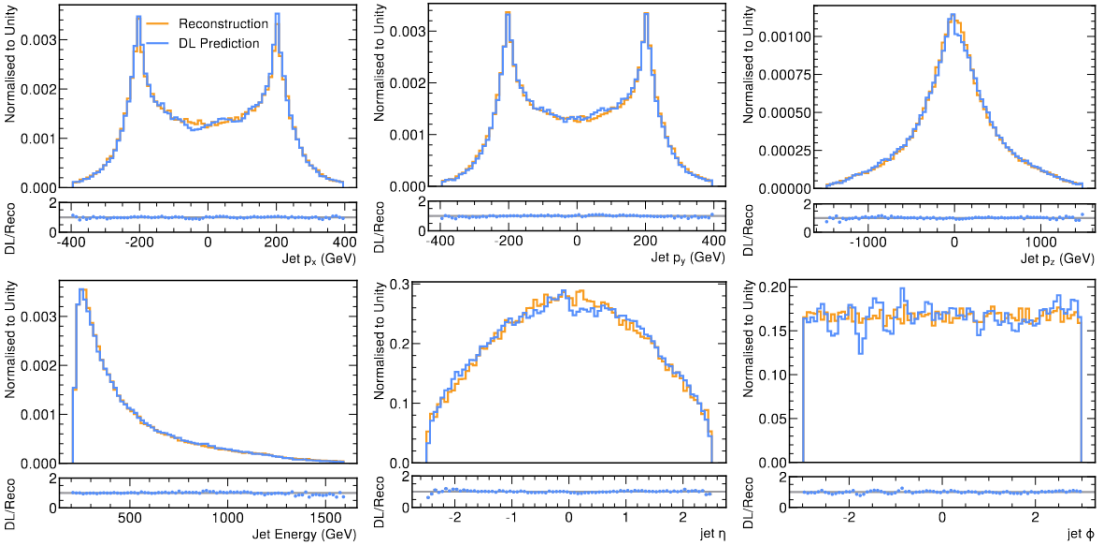
$$D_{KL}(q_\phi(z|x)||p_\theta(z))$$

VAE Architecture

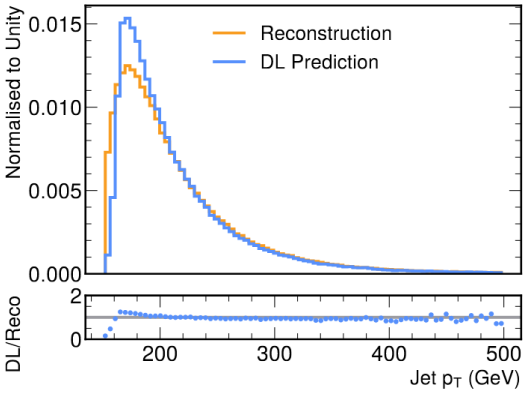


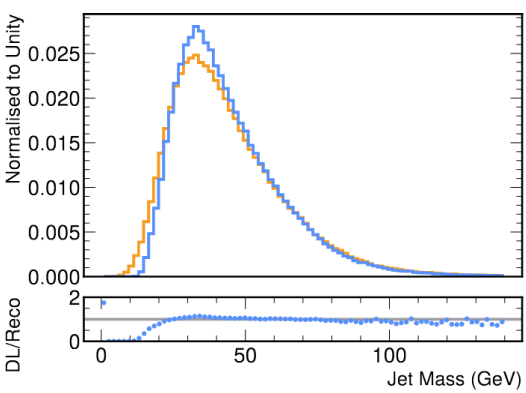
- Leaky ReLU (negative slope = 0.1) is being used as the activation function on all layers except for encoder and decoder output layers where linear activation is used.
- Adam optimization with a learning rate = 0.0001, latent space dimension = 20, and loss function tuned with $\beta = 1/9$, $\alpha_m = 1.0$, $\alpha_{p_T} = 0.1$
- Max pooling (Upsampling) with kernel size 1×2 and stride 2 is used after (before) the first (last) two Conv2D (ConvTrans2D) layers

Results: Jets Properties



Results: Jets Properties (before 200 GeV cut)





Results: Jets Properties (before 200 GeV cut)