Hadronic Simulation with conditional Masked Autoregressive Flow

Tuan Pham and Xiangyang Ju

26^{th} International Conference on Computing in High Energy Physics and Nuclear Physics
May 11, 2023

WISCONSIN
UNIVERSITY OF WISCONSIN–MADISON
BERKELEY LAB

Hadronic interaction simulation in GEANT4

- In particle colliders, hadrons ($\pi^{\pm}$, $K^{\pm}$, p, n, ...) are copiously produced and interact with the detector material, creating a shower of secondary particles.
- A full detector simulation requires modelling hadronic interaction

$$p_1 + \text{stationary target} \rightarrow q_1 + q_2 + \dots$$

  where p_1 is the hadron projectile 4-vector momentum and q_i are secondary product momenta.
- Full simulations are increasingly expensive, motivating the exploration of deep generative architectures such as GAN [1, 3] and normalising flow [5] for to enhance simulation speed.
- We explore here the ability of a normalizing flow architecture to simulate the final state of the interaction between hadrons and nuclei, using data generated with the GEANT4 toolkit.

Normalizing flow

- A normalizing flow seeks a bijective transformation $f : U \to X$, in which $U \sim \pi_U$ is the base variable and $X \sim p_X$ the target.
- The base distribution π_U is known and simple, whereas the target p_X unknown and complex.
- The bijector f is parametrized by a Multilayer Perceptron (MLP) and multiple bijectors are chained together to create a deeper and more expressive flow.
- The weights w of the MLPs are learned from GEANT4 data. On inference, we sample u from π_U and compute $x = f(u)$

Autoregressive density estimators I

- Let the target random variable be $X \in \mathbb{R}^d$, $X \sim p_X$. Decompose p_X as a product of $p_{X_i|X_{1:i-1}}$ and parameterise each conditional as a Gaussian [4] (autoregressive property).

$$p_X(x) = \prod_{i=1}^{d} p_{X_i|X_{1:i-1}}(x_i|x_{1:i-1}) = \boxed{\prod_{i=1}^{d} \mathcal{N}\left(x_i|\mu_i, \sigma_i^2\right)}, \tag{1}$$

in which $\mu_i = \mu_i(X_{1:i-1})$ and $\sigma_i = \sigma_i(X_{1:i-1})$. Now let

$$U_i = \frac{X_i - \mu_i}{\sigma_i} \Rightarrow X_i = U_i \sigma_i + \mu_i = f_i(U_i, X_{1:i-1}) \tag{2}$$

- Since $X_i \sim \mathcal{N}(\mu_i, \sigma_i^2)$, $\boxed{U_i \sim \mathcal{N}}$. Perform a change of variable,

$$p_{X_i|X_{1:i-1}}(x_i|x_{1:i-1}) = \mathcal{N}\left(f_i^{-1}(x_{1:i})\right) \left| \frac{\partial f_i^{-1}}{\partial X_i} \right| = \frac{\mathcal{N}\left(f_i^{-1}(x_{1:i})\right)}{\sigma_i} \tag{3}$$

Autoregressive density estimators II

- f is learned from data by minimizing the KL-divergence of $p_X(x)$ and the sample distribution $\pi_X(x)$, tantamount to minimizing the negative log likelihood of the data

$$-\frac{1}{N}\sum_{i=1}^{N}\log(p_X(x_i)), \quad x_i \sim \pi_X \tag{4}$$

- f is implemented using the Masked Autoencoder for Density Estimation (MADE) architecture [2], which ensures $\mu_i = \mu_i(x_{<i})$ and $\sigma_i = \sigma_i(x_{<i})$. We feed into each MADE block the COM-frame total energy E_{com} as the condition and concatenate multiple blocks to create a deep conditional Masked Autoregressive Flow (MAF).

MADE architecture

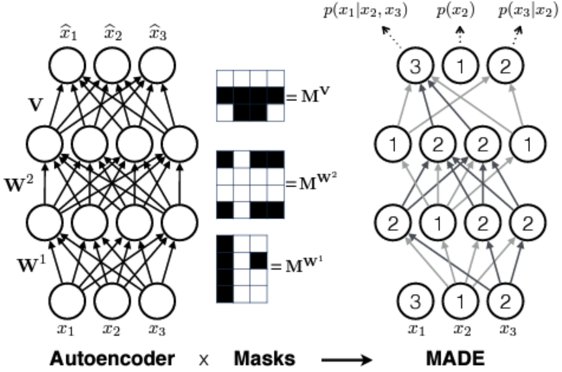

Figure: Masked Autoencoder for Density Estimation (MADE) architecture [2] as building block of the MAF. The connections are systematically dropped out to guarantee the autoregressive property.

Simulation data - low projectile energy

- We simulated the data from $\pi^- + H \rightarrow q_1 + q_2$ reaction using *FTFP_BERT_ATL* physics list from GEANT4, which comprises of 2 models operating at two regimes of projectile energy.
- The Bertini cascade model operates at $0\,GeV < k_\pi < 12\,GeV$

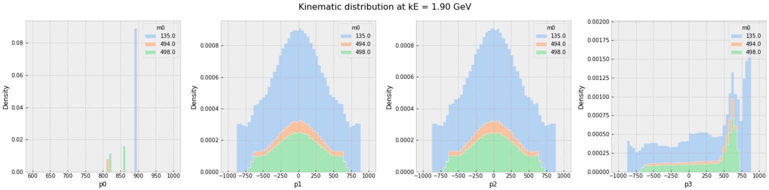

Figure: GEANT4 simulated of final state particle in a $2 \rightarrow 2$ hadronic interaction at $k_{\pi^-} = 1.9$ GeV. Notice the irregularity the p_z distribution.

Simulation data - high projectile energy

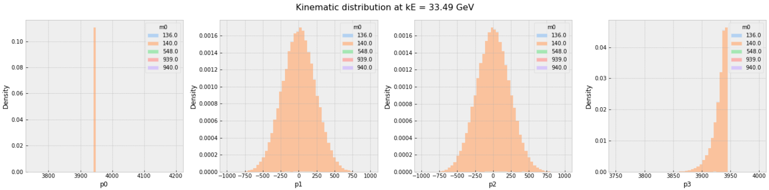

Figure: GEANT4 simulated of final state particle in a $2 \to 2$ hadronic interaction at $k_{\pi^-} = 33.5$ GeV.

- Fritiof parton model operates at $9\,GeV < k_\pi < 100\,TeV$. Data generated by Fritiof are considerably less complex than by Bertini $\Rightarrow$ Train density estimator separately for 2 regimes.
- Train each model to generate the first particle's momentum (E, p_x, p_y, p_z), conditioned on the COM-frame total energy.
- Test model ability to interpolate to projectile energy it never sees during training.

Model training

- Model contains 30 MADE blocks, each consisting of a 2-layer MLP, each layer having 128 nodes.
- Training objective: minimize the negative log likelihood of the input data.
- Trained over 2000 epochs, with learning rate decaying from 10^{-4} to 10^{-6}.
- Quantify the closeness of truth distribution (f_1) and generated distribution (f_2) by Wasserstein (earth-mover) distance

$$W(f_1, f_2) = \int_0^1 |F_1^{-1}(q) - F_2^{-1}(q)| dq \tag{5}$$

where F_i is the cumulative distribution function i.

Qualitative comparison of MAF and truth data - low energy I

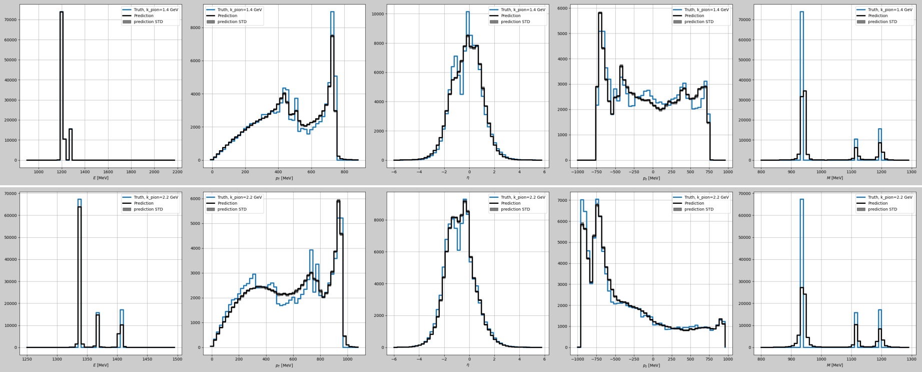

Figure: Data generated by MAF and GEANT4 at $k_{\pi^-} = 1.4$ GeV (up) and $k_{\pi^-} = 2.2$ GeV (down).

- The model captures the irregular shape of the p_T and p_z distributions.
- Despite being trained on (E, p_x, p_y, p_z), it reproduces the mass spectrum $\Rightarrow$ learned physically relevant information.

Qualitative comparison of MAF and truth data - low energy II

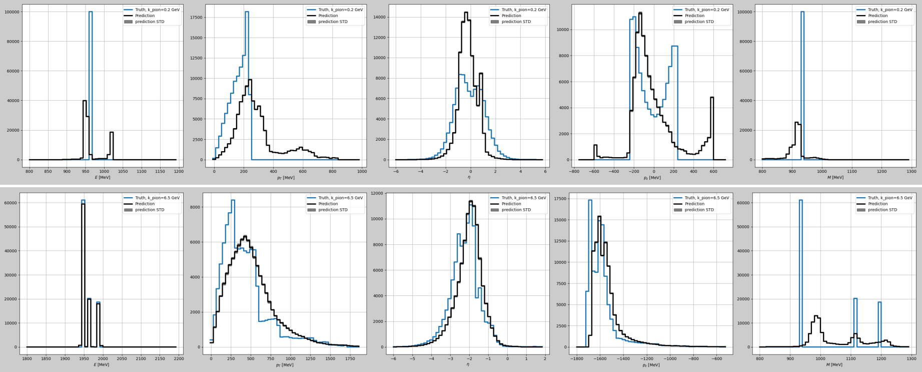

Figure: Data generated by the MAF and GEANT4 at $k_{\pi^-} = 0.2$ GeV (up) and $k_{\pi^-} = 6.6$ GeV (down).

The performance degrades near both upper and lower boundaries of the training conditional input, possibly due to uneven sampling of training conditional input (see next slide).

Quantitative result

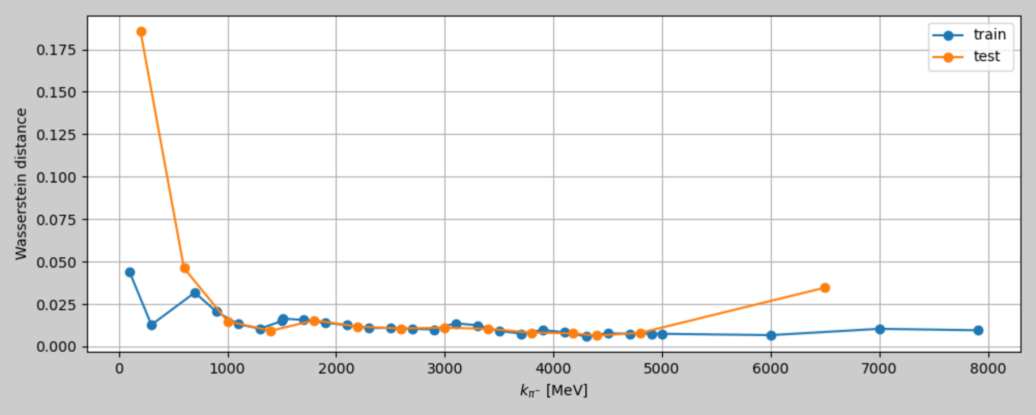

Figure: Comparison of train and test performance of the MAF on at low-projectile regime.

For $k_{\pi^-} \in [1, 5]$ GeV, test performance agrees with training performance. Where the grid of conditional inputs is more sparse, we have sub-optimal performance $\Rightarrow$ Use more training data at these regimes.

Result at high projectile energy

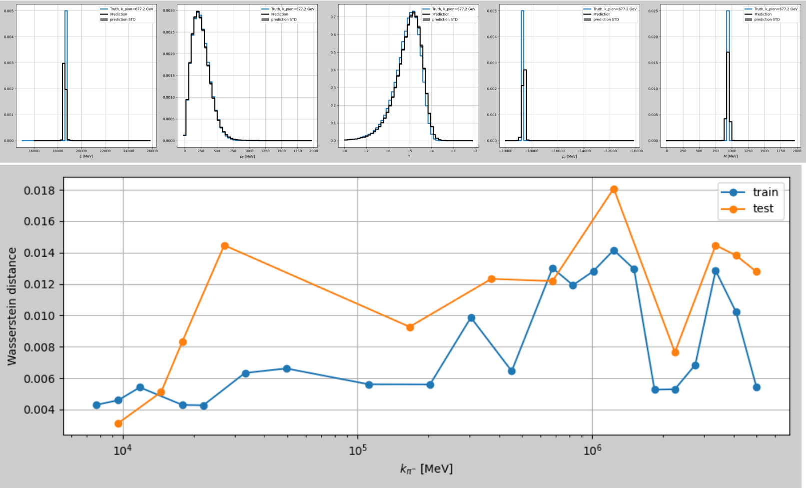

Figure: Qualitative comparison of MAF- and GEANT4 data at $k_{\pi^-} = 67.7$ GeV (upper) and quantitative performance over the range of test conditions (lower). At high projectile energy (Fritiof regime), the model reaches better agreement to GEANT4 than at lower energy (Bertini regime).

Preliminary $2 \to 3$ simulation

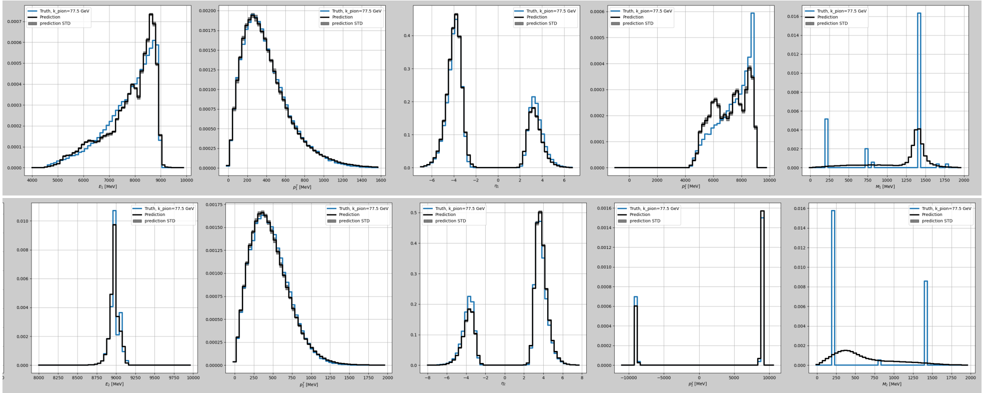

Figure: Qualitative comparison between MAF- and GEANT4-generated data for $2 \to 3$ interaction at $k_{\pi^-} = 77.5$ GeV.

Reasonable agreement between generated and truth data. However, the mass spectrum is more spread out. Will investigate ways to mitigate this short-coming.

Summary and future directions

- Result shows that the MAF architecture can learn the kinematic distribution of hadronic interaction and interpolate to unseen projectile energies.
- The model can captures non-smooth features of the kinematic distribution, but shows poor performance with low training statistics.
- Moving forward, we will obtain training data on a more dense grid of conditional inputs, experiment with different ways to scale the conditional input and representation of the data.
- Explore other initial conditions (other projectiles and target) and final states (3-particle, 4-particle final states, etc.).

References I

[1] Luke de Oliveira, Michela Paganini, and Benjamin Nachman. Learning particle physics by example: Location-aware generative adversarial networks for physics synthesis. *Computing and Software for Big Science*, 1(1), sep 2017. doi: 10.1007/s41781-017-0004-6. URL https://doi.org/10.1007%2Fs41781-017-0004-6.

[2] Mathieu Germain, Karol Gregor, Iain Murray, and Hugo Larochelle. Made: Masked autoencoder for distribution estimation, 2015.

[3] Michela Paganini, Luke de Oliveira, and Benjamin Nachman. Accelerating science with generative adversarial networks: An application to 3d particle showers in multilayer calorimeters. *Physical Review Letters*, 120(4), jan 2018. doi: 10.1103/physrevlett.120.042003. URL https://doi.org/10.1103%2Fphysrevlett.120.042003.

[4] George Papamakarios, Theo Pavlakou, and Iain Murray. Masked autoregressive flow for density estimation, 2018.

[5] Danilo Jimenez Rezende and Shakir Mohamed. Variational inference with normalizing flows, 2016.