Deep generative models for generating Drell-Yan events in the ATLAS collaboration at the LHC

Xiangyang Ju
On behalf of the ATLAS Collaboration,
CHEP 2023, Norfolk Virginia
9 May 2023

This project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No *800945* — NUMERICS — H2020-MSCA-COFUND-2017

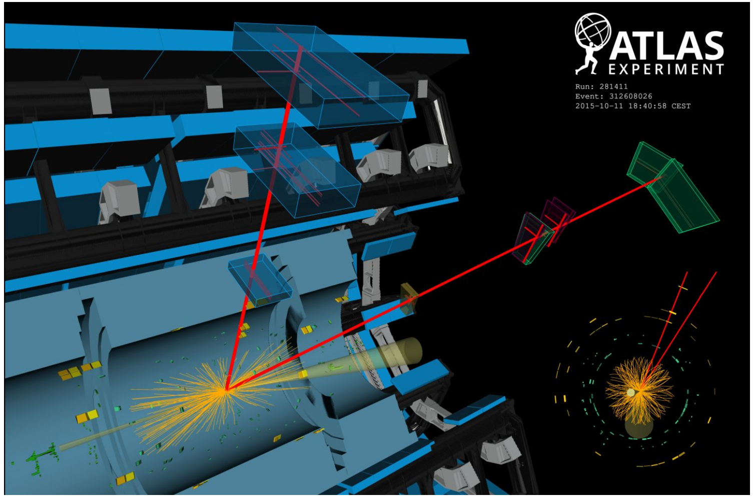
With Peter Fitzhugh and Rosy Nikolaidou

Introduction

Study of Higgs boson decay to two muons (μ)
- Discovery of the Higgs boson, by ATLAS and CMS at the LHC in 2012
- The H—>μμ process essential to measure the couplings of the Higgs boson to the 2nd generation of fermions
  - Full Run-2 data analysis already published

In the context of this process:
- Application of AI techniques to simulate the main background of the H—>μμ process with a large number of statistics.

Phys. Lett. B 812 (2021)



Photograph: ATLAS Collaboration

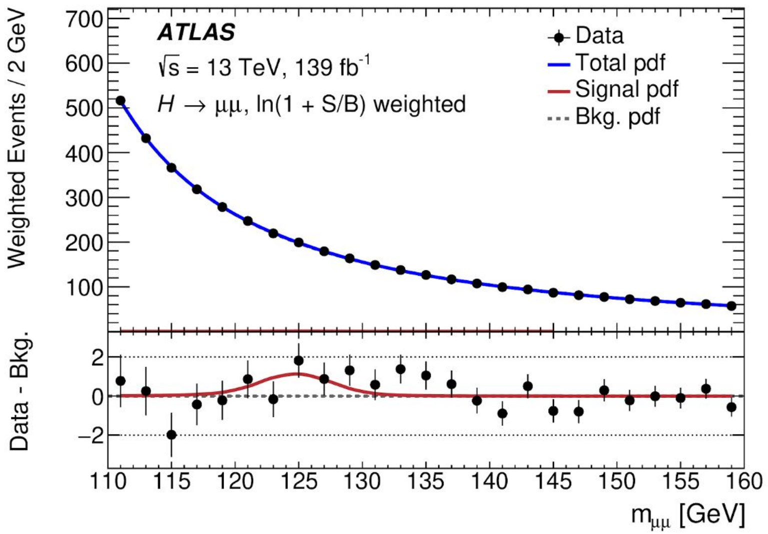
Motivation

H→μμ decay is a rare decay with small S/B ratio

- Main background: the dimuon Drell-Yan process
- Accurate background determination highly important
- $m_{\mu\mu}$ parametrized by analytic functions
  - *background functions based on the **spurious signal**, which measures the residual signal events obtained from signal-plus-background fits to background-only MC templates*
- High statistics needed to derive these MC templates to reduce possible fluctuations

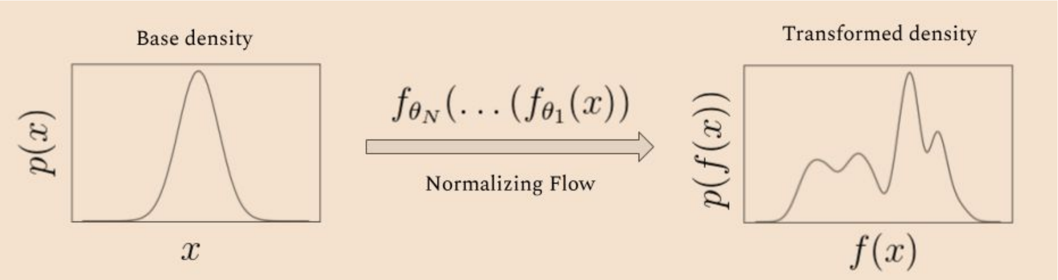
- Main production method of simulated samples time consuming and computationally intensive

Phys. Lett. B 812 (2021)



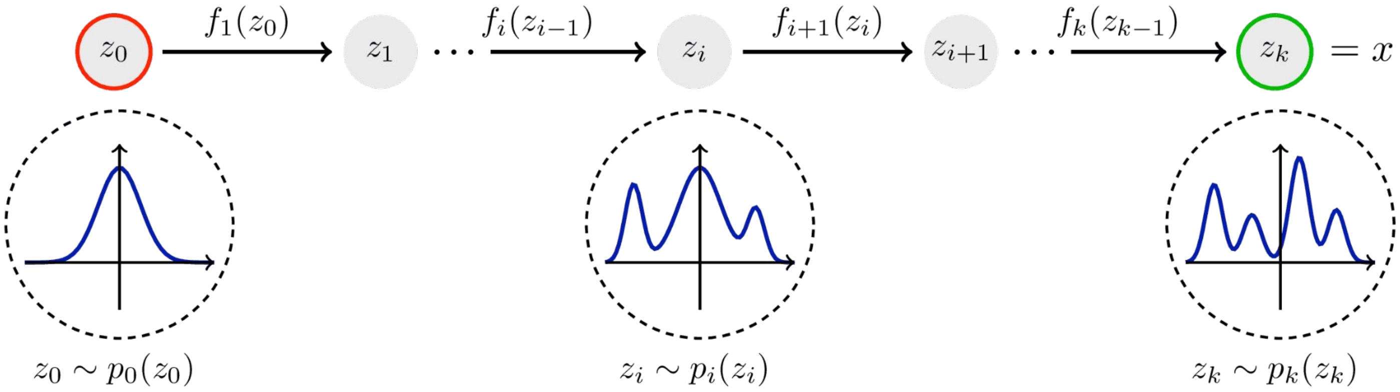
What we do

- Use of generative models trained on the existing ATLAS fully simulated and fully reconstructed events containing 7 variables in order to generate billions of events using GPUs for the spurious signal study, and to test the statistical independence of these events

- 

- Alternative procedure for the generation of events with high statistical power that could be used in the future by default in many analyses at the LHC

Normalising Flow (NF) Model

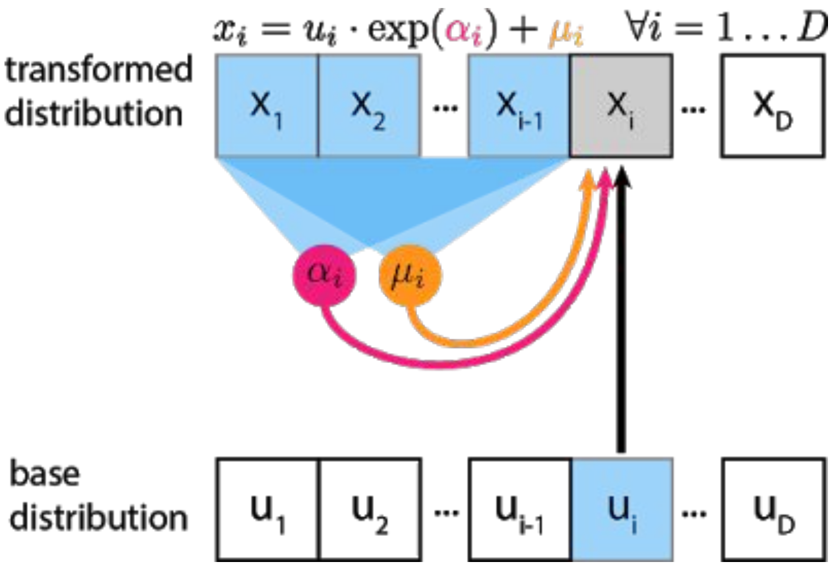


- A NF is a model that transforms a simple base density function to a more complex target density distribution using a bijective, differential function known as a bijection.

- A NF often contains a chain of bijections

- Using the change of variables technique in mathematics, a normalizing flow can estimate the target density distribution with the input vector

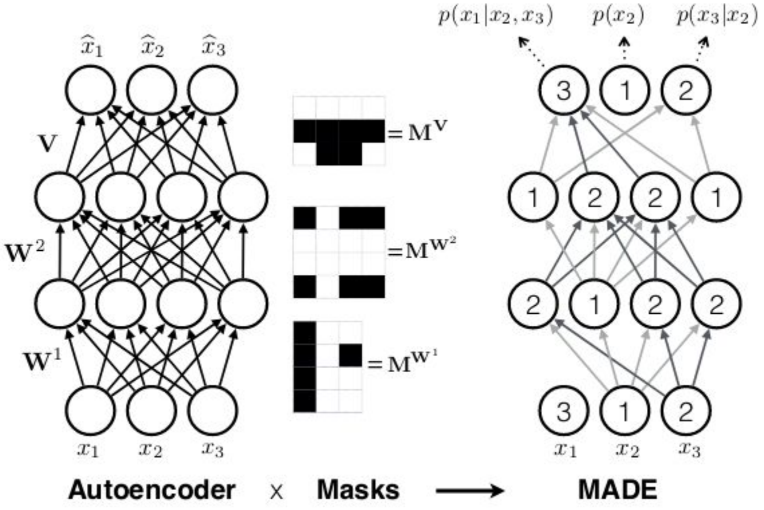
Masked Autoregressive Flow



- Autoregressive Flow sequentially transforms each dimension based on the previously transformed dimensions

So it depends on the ordering of the input vectors and slow for sampling, but fast for training

Masked Autoregressive Flow



Another key component is the Masked Autoencoder

- It masks the weights of the Neural Network so that the output dimension i depending on the previous dimensions

- By construction, it fulfill the autoregressive requirements

Methodology

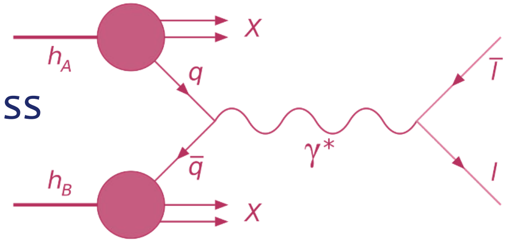
- Use of the ATLAS fully simulated dimuon background events to train the NF model

- Split these events into a training and a testing portion.

- The training portion is used throughout a number of training epochs

- Metric used to determine performance: **Wasserstein distance**
  - Wasserstein Distance is defined as the minimum ‘cost’ required to change one probability distribution into another
  - Aim was to minimize the Wasserstein distance between the Fully Simulated and Network Generated events

Training Dataset Used



- Used fully simulated samples of the Drell-Yan Process

- Separate the samples by the number of jets and train for each case separately
  - Chosen as these are the same categories used for the Hμμ analysis and for the Higgs production modes

For this study, only the 0 jet case is shown

- The training dataset contains the Pt , η, φ for each of the two muons and the invariant mass distribution containing the final state radiation (FSR)
- Extra variables were calculated after the training to verify results
  - Examples shown: Dimuon η , Cosθ*, Dimuon φ

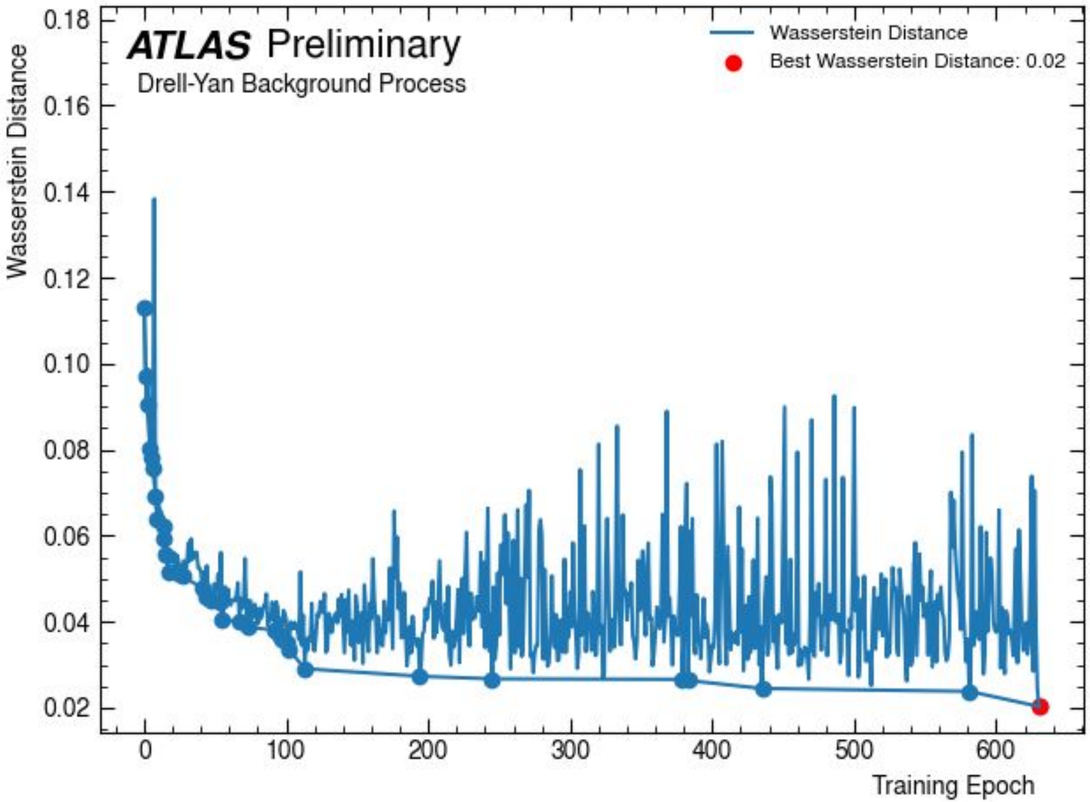
Training the Normalising Flow Model

- Parameters of the model:
  - Number of Events: 1,000,000
  - Percentage of events used for testing: 10%
  - Multilayer Perceptron Shape= 128*2
  - Number of Bijector Layers = 20
  - Learning Rate = 1e-3 to 1e-4, polynomial decay with a power of 0.5
  - Batch Size = 512
  - Number of Epochs = 1000

- 2 sets of data presented: ‘Fully Simulated’ and ‘Network Generated’:
  - ‘Fully Simulated’: Fully reconstructed events from the SHERPA simulated samples of the Drell-Yan
  - ‘Network Generated’: Events generated from the trained normalising flow model

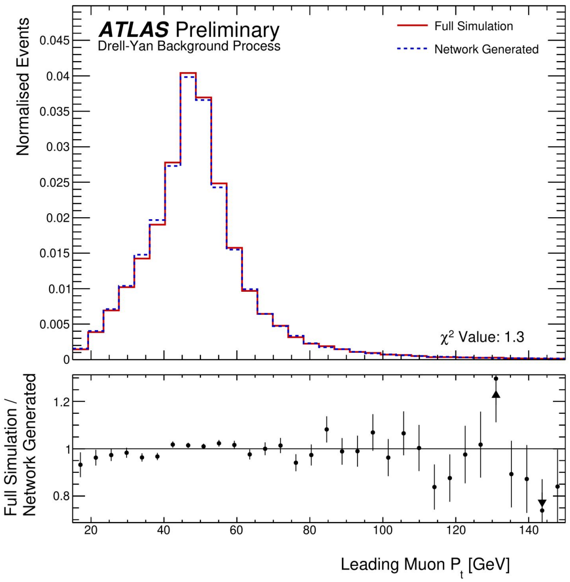
Results

Best Wasserstein Distance: 0.02 at epoch 630

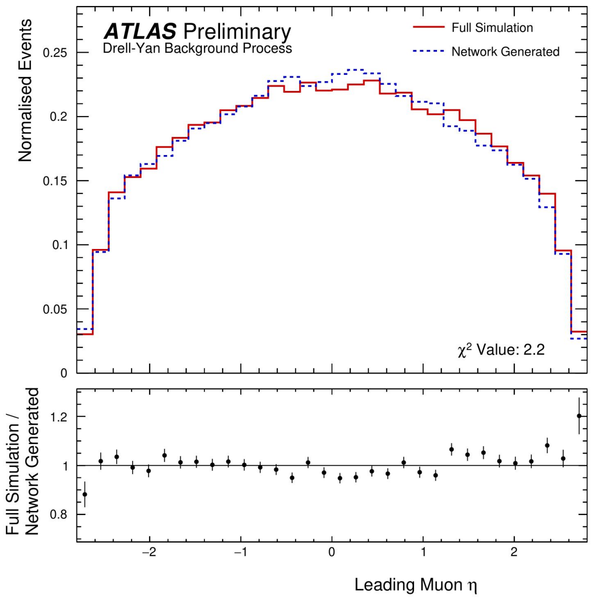


Leading Muon kinematics

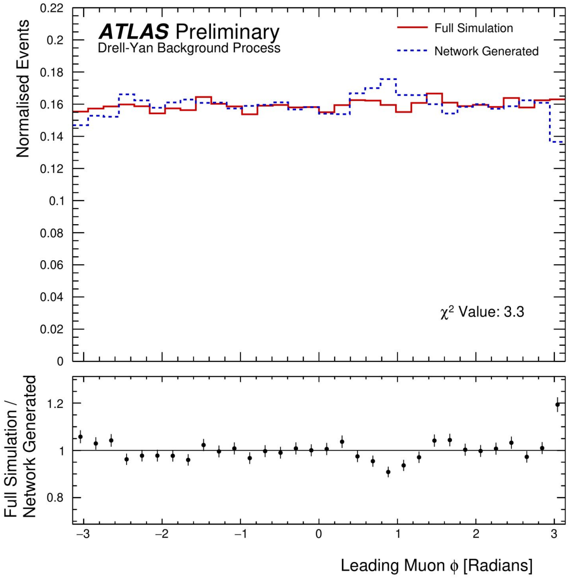
Leading Muon p_T



Leading Muon η



Leading Muon ϕ

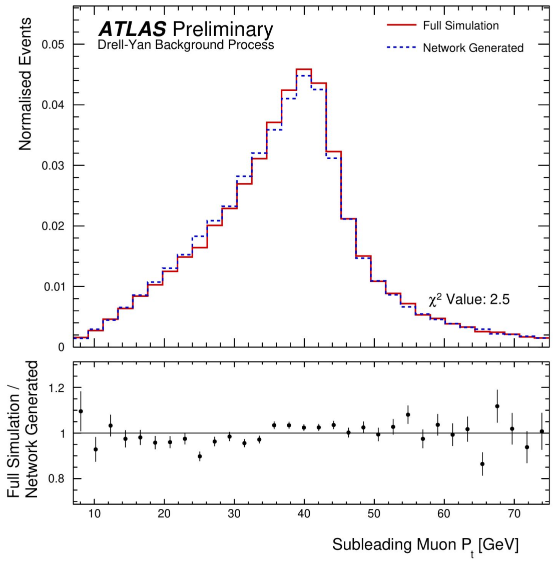


Full Simulation

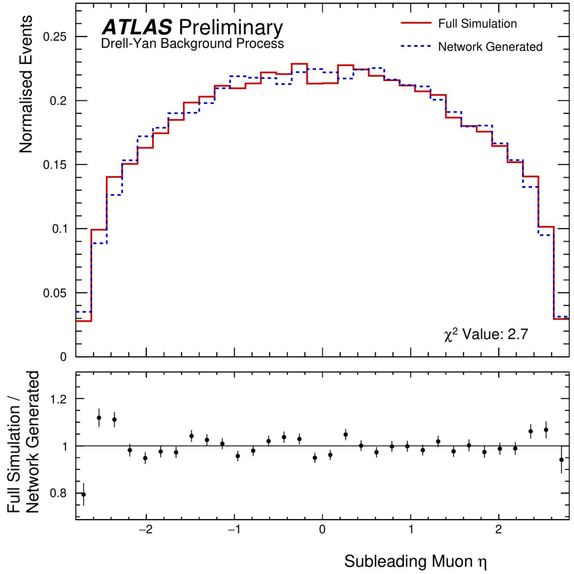
Network Generated

Sub-leading Muon

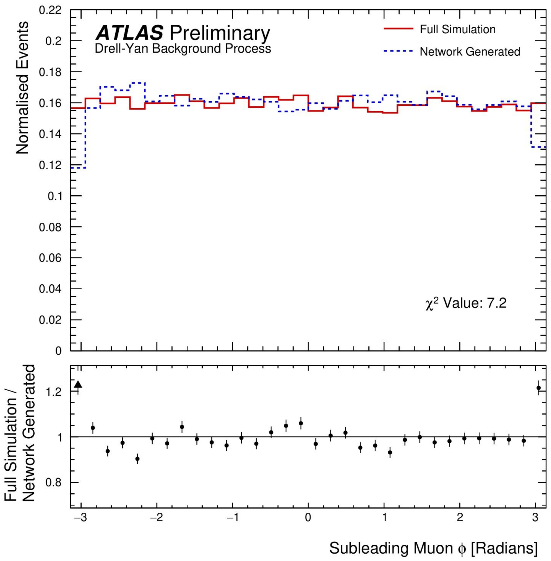
Sub-leading Muon p_T



Subleading Muon η



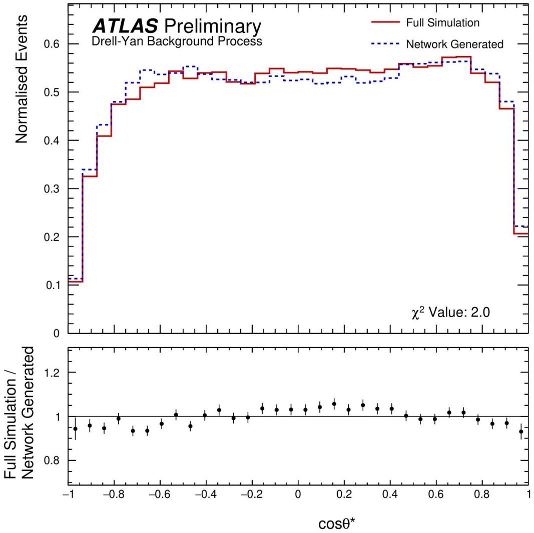
Subleading Muon ϕ



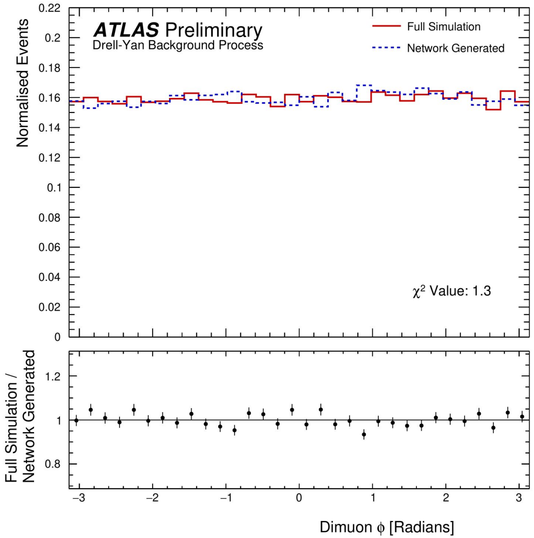
Full Simulation

Network Generated

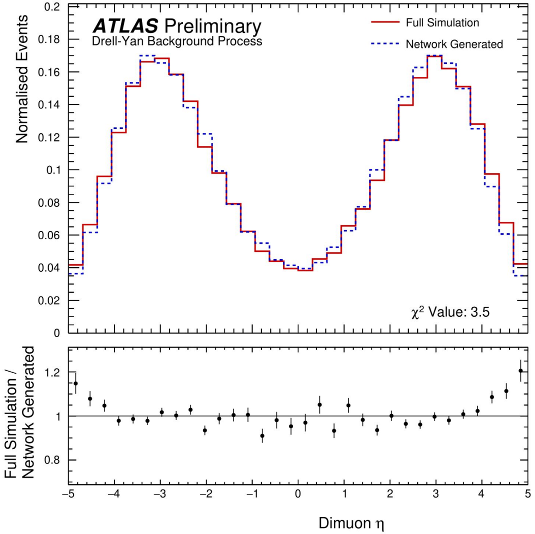
Calculated Variables



Cosθ*



Dimuon system φ



Dimuon system η

Definition:

$$cos\theta* = \frac{2(P_1^+ P_2^- - P_1^- P_2^+)}{\sqrt{m_{ll}^2 (m_{ll}^2 + p_{T,ll}^2)}} \frac{p_{z,ll}}{|p_{z,ll}|}$$

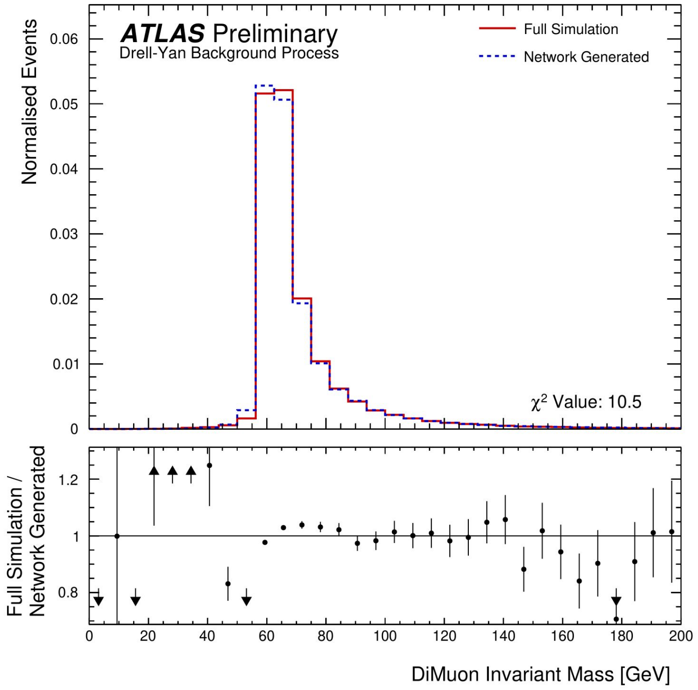
Where: $P_i^{\pm} = \frac{E_i \pm p_{z,i}}{\sqrt{2}}$

With, Cosθ* in the Collins-Soper rest frame

Full Simulation

Network Generated

Invariant mass of dimuon system

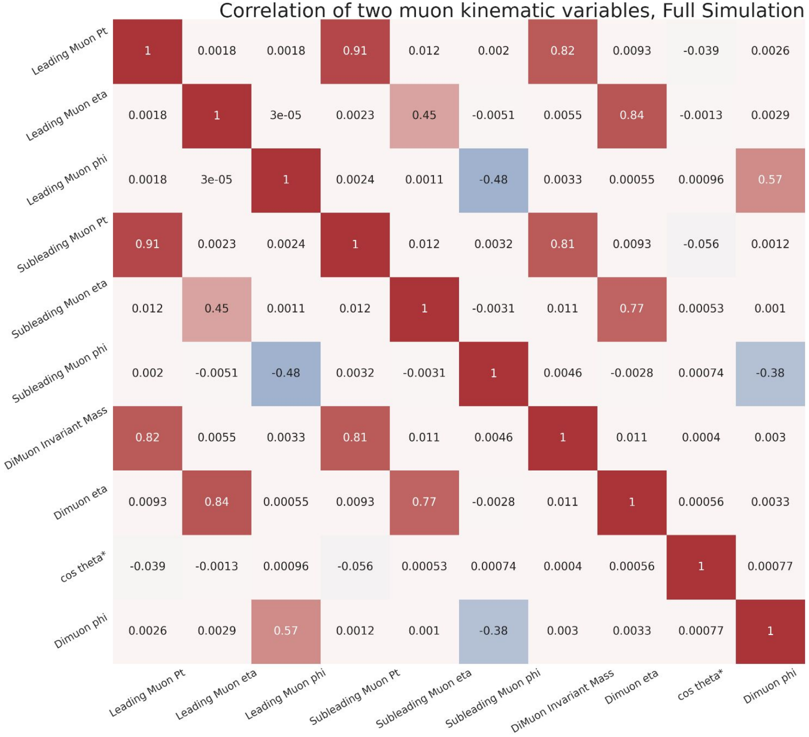


We see very good agreement in the dimuon invariant mass even though this variable is not used in the training.

Correlation Plots

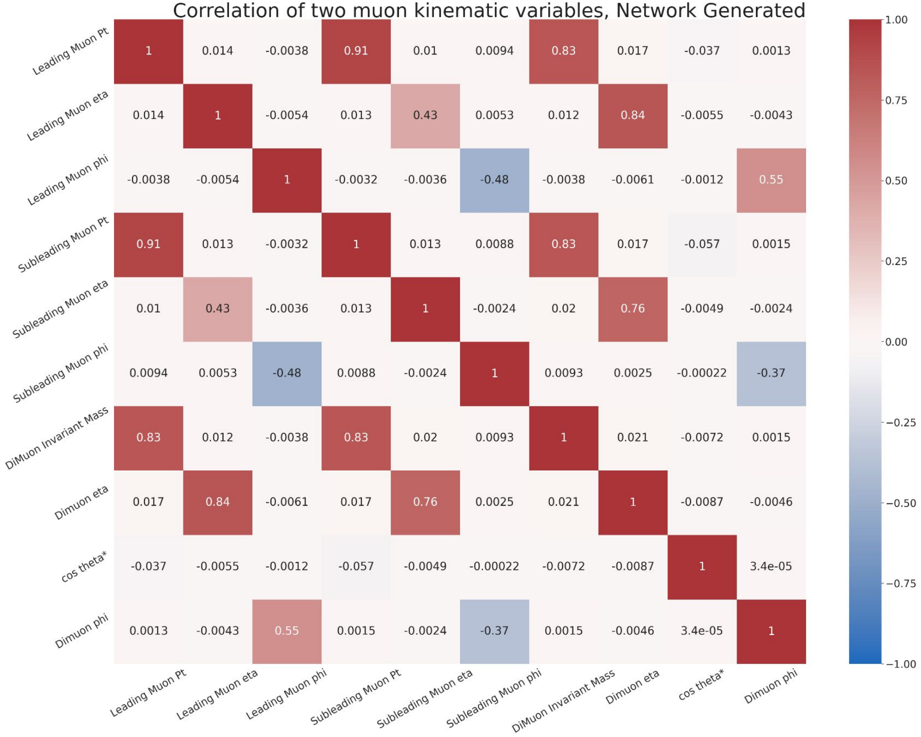
ATLAS **Preliminary**
Drell-Yan Background Process

Full Simulation



ATLAS **Preliminary**
Drell-Yan Background Process

Network Generated



Generated events do a successful job of replicating the correlation between variables

Conclusion

- **We have explored the use of Generative Machine Learning techniques to generate a large number of events for the Drell Yan background process in the $H \rightarrow \mu\mu$ decay in ATLAS for the first time**

- The **Normalising Flow** model used shows promising results in performance

- We demonstrate that the NF can capture the correlations between the two muons in the events and model well those event-level properties

- Work is ongoing to obtain better training results for the 0 jet case and to use the same technique for the cases with jets

- Once trained, the model can generate large quantities of events very quickly