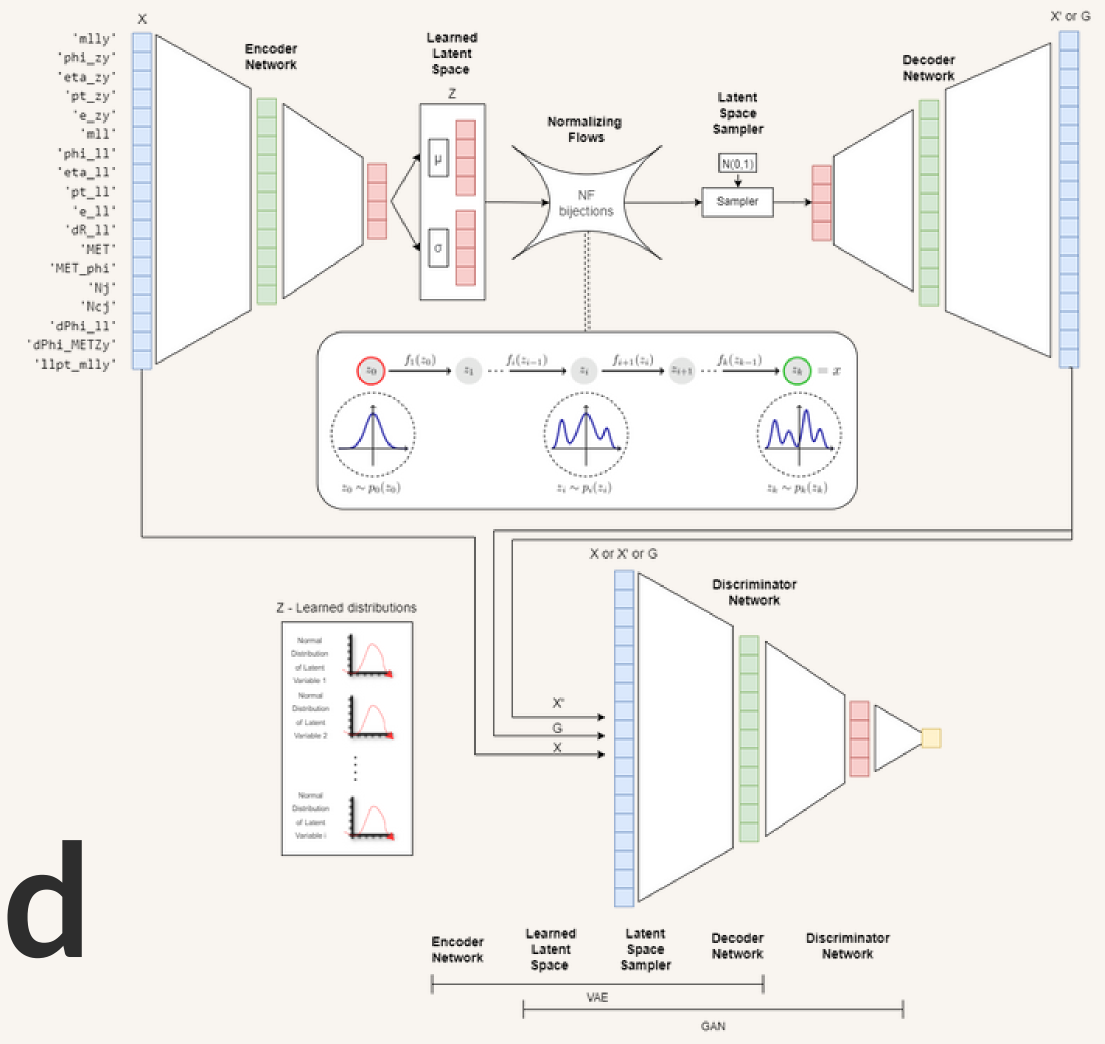
Kruger 2022, Discovery Physics at the LHC



The Assessment of Different VAE Derivatives for Data Generation and Event Classification in Zγ Final State Background Data

Finn Stevenson

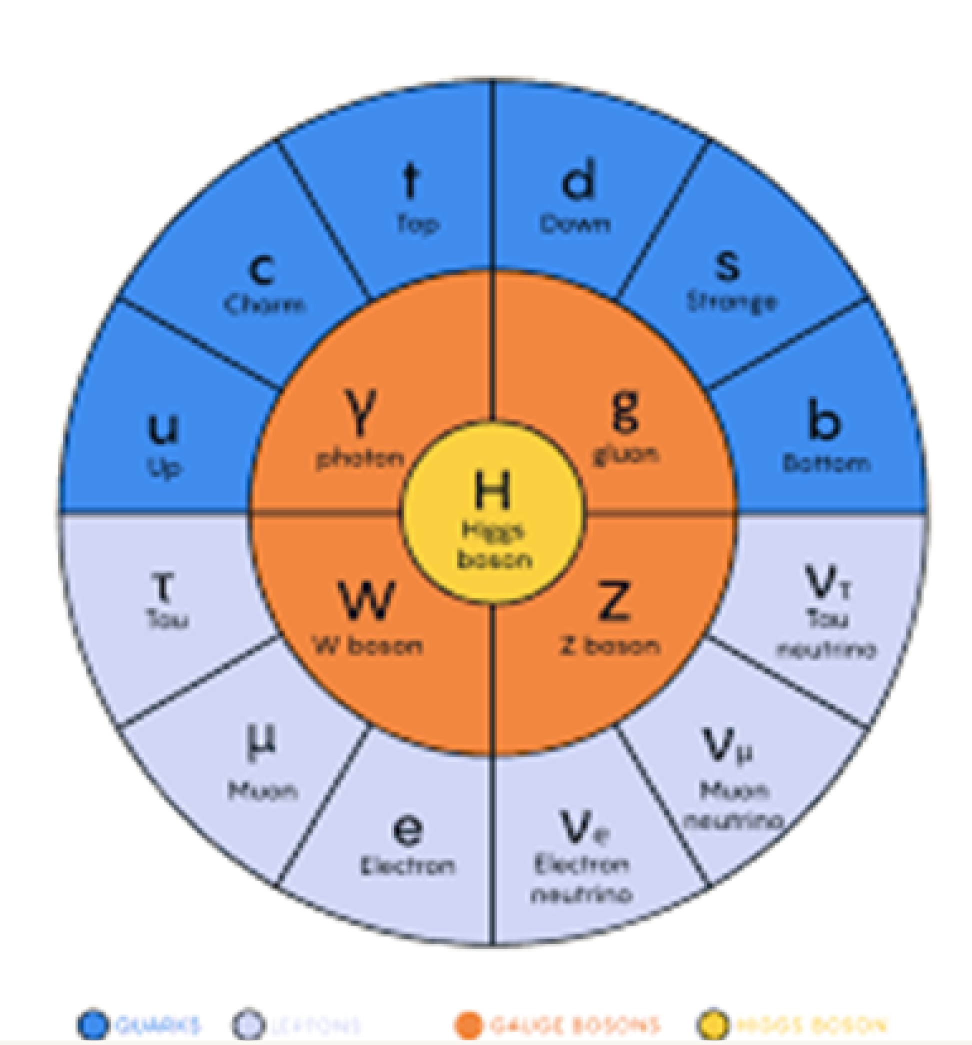
2022 December 7th

Institute for Collider Particle Physics
University of the Witwatersrand

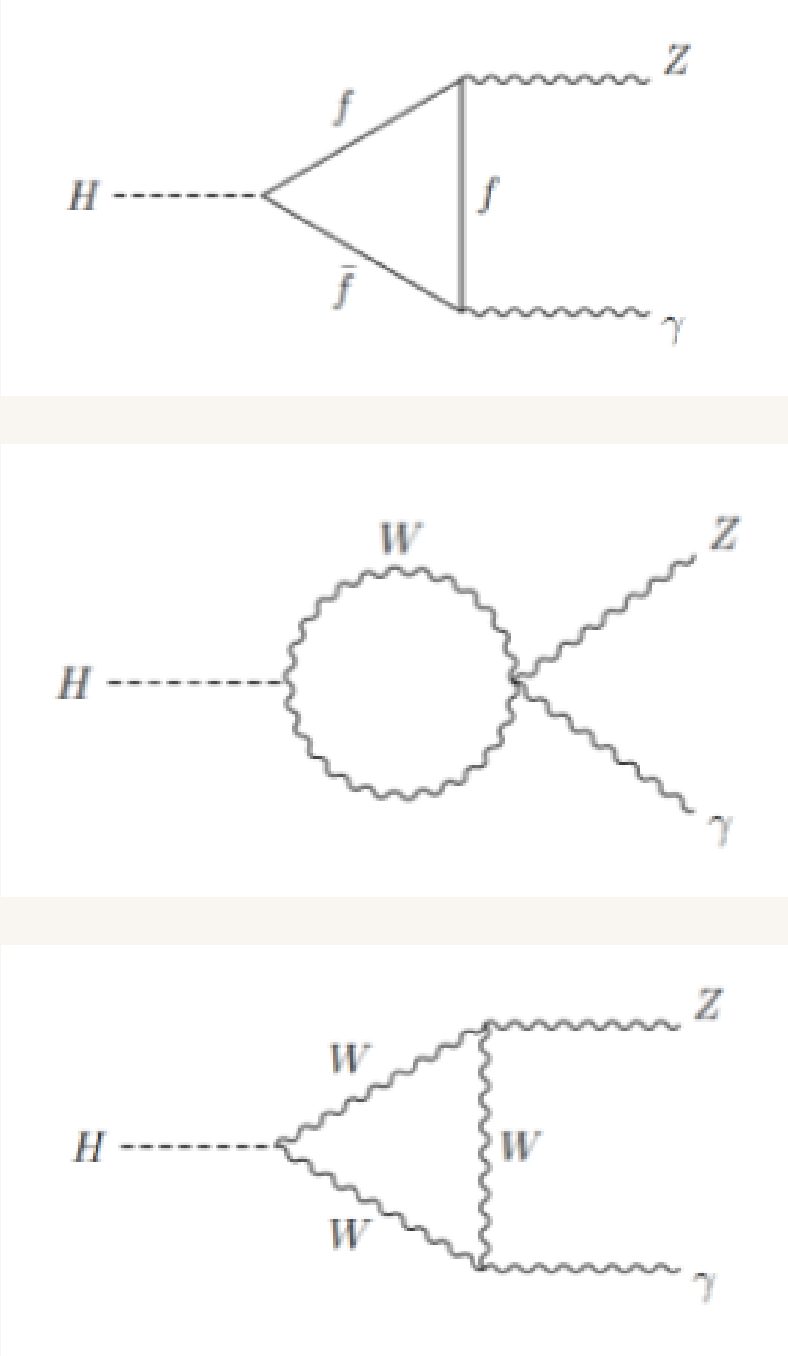
Background & Motivation

- The **Standard Model (SM) explains how the basic building blocks of matter interact,** governed by four fundamental forces.
- SM was **completed by the discovery of the Higgs boson** in 2012 by the ATLAS and CMS collaborations.
- However, the SM is **not able to explain a number of phenomena** seen in the data.
- These discrepancies to the SM **motivate the search for new bosons.**
- **Many theories have been developed to extend the SM** and propose a theoretical explanation for phenomena beyond the Standard Model (BSM).
- Some theories include the prediction of **novel massive bosons**.
- **2HDM+S,** where S is a singlet scalar
- Certain models predicted the **decay of the new massive boson to Zγ final state**.
- In this research, we are searching for **Zγ resonances (pp →H →Zγ)**

$$pp \rightarrow H \rightarrow Z\gamma \rightarrow (\ell^{+}\ell^{-})\gamma$$



Fast Simulation Monte Carlo Zγ Data generated:
Using Madgraph5 with NNPDF3.0 parton distribution function. Parton level generation is done using Pythia and detector level simulation is done using Delphes(v3)



The leading order Feynman diagrams of massive boson H to Zγ final state

Background & Motivation

MC simulation at the LHC -> **CPU intensive**

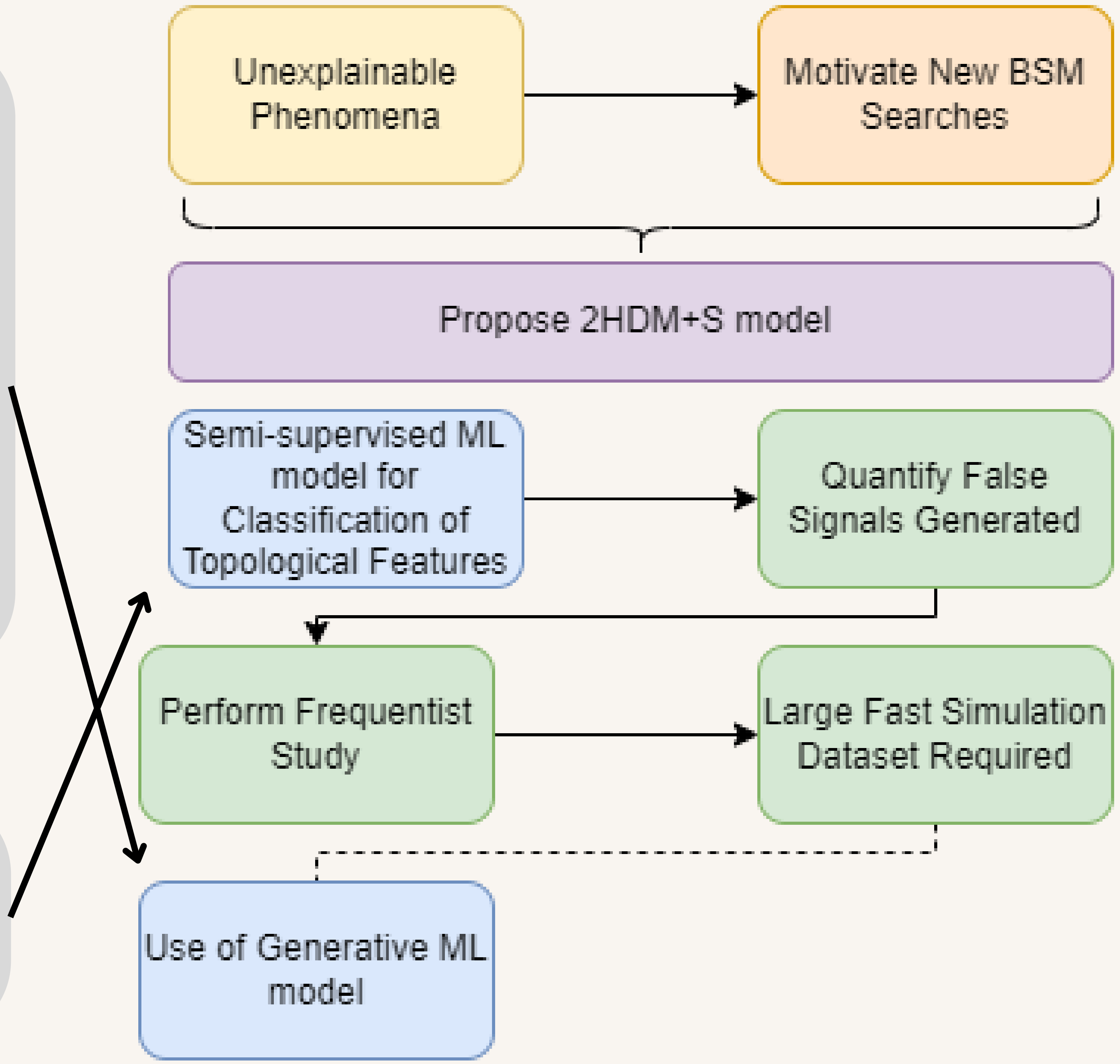
Deep learning models for fast simulation -> **Alleviate CPU pressure**

The **luminosity** of the detectors at the LHC is **increasing continuously.**

Variational Autoencoders (**VAEs) are assessed as a deep learning-based event production mechanism.**

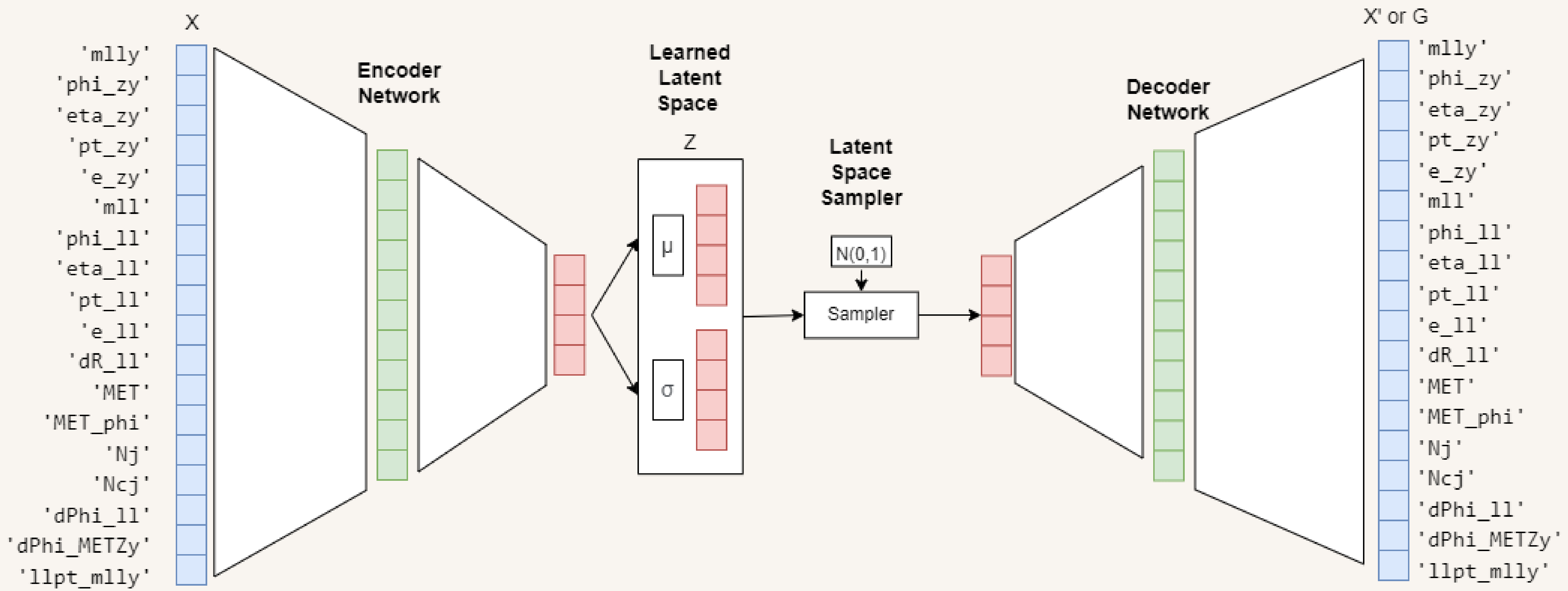
Signal classification -> crucial in the search for new bosons.

A **VAE** can be a **Signal Classification** and a **Data Generation** model in one.



Hypothesis

A VAE well-trained for data generation purposes can be used in the search for new bosons at the LHC simultaneously for data generation and signal classification.



VAE Use in HEP

VAEs have previously been utilised for many applications in HEP.

ATLAS EXPERIMENT

ATLAS PUB Note

ATL-SOFT-PUB-2018-001

10th July 2018

CERN

Deep generative models for fast shower simulation in ATLAS

The ATLAS Collaboration

The need for large scale and high fidelity simulated samples for the extensive physics program of the ATLAS experiment at the Large Hadron Collider motivates the development of new simulation techniques. Building on the recent success of deep learning algorithms, Variational Auto-Encoders and Generative Adversarial Networks are investigated for modeling the response of the ATLAS electromagnetic calorimeter for photons in a central calorimeter region over a range of energies. The properties of synthesized showers are compared to showers from a full detector simulation using Geant4. This feasibility study demonstrates the potential of using such algorithms for fast calorimeter simulation for the ATLAS experiment in the future and opens the possibility to complement current simulation techniques. To employ generative models for physics analyses, it is required to incorporate additional particle types and regions of the calorimeter and enhance the quality of the synthesized showers.

2018 Deep generative models for fast shower simulation in ATLAS Tech. rep. CERN Geneva all figures including auxiliary figures are available at https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PUBNOTES/ATL-SOFT-PUB-2018-001 URL http://cds.cern.ch/record/2630433

Variational Autoencoders for Jet Simulation

Kosei Dohi

21kdohi@tenafly.k12.nj.us

Tenafly High School

Tenafly, NJ 07670 USA

Abstract

We introduce a novel variational autoencoder (VAE) architecture that can generate realistic and diverse high energy physics events. The model we propose utilizes several techniques from VAE literature in order to simulate high fidelity jet images. In addition to demonstrating the model's ability to produce high fidelity jet images through various assessments, we also demonstrate its ability to control the events it generates from the latent space. This can be potentially useful for other tasks such as jet tagging, where we can test how well jet taggers can classify signal from background for events generated by the VAE. We test this idea by seeing the signal efficiency vs background rejection for different types of jet images produced by our model. We compare our VAE with generative adversarial networks (GAN) in several ways, most notably in speed. The architecture we propose is ultimately a fast, stable, and easy-to-train deep generative model that demonstrates the potential of VAEs in simulating high energy physics events.

Dohi K 2020 Variational autoencoders for jet simulation URL https://arxiv.org/abs/2009.04842

Bump Hunting in Latent Space

Blaž Bortolato* and Aleks Smolkovič†
Jožef Stefan Institute, Jamova 39, 1000 Ljubljana, Slovenia

Barry M. Dillon‡
Institut für Theoretische Physik, Universität Heidelberg, Germany

Jernej F. Kamenik§
Jožef Stefan Institute, Jamova 39, 1000 Ljubljana, Slovenia and
Faculty of Mathematics and Physics, University of Ljubljana, Jadranska 19, 1000 Ljubljana, Slovenia
(Dated: June 8, 2022)

Unsupervised anomaly-detection could be crucial in future analyses searching for rare phenomena in large datasets, as for example collected at the LHC. To this end, we introduce a physics inspired variational autoencoder (VAE) architecture which performs competitively and robustly on the LHC Olympics Machine Learning Challenge datasets. We demonstrate how embedding some physical observables directly into the VAE latent space, while at the same time keeping the anomaly-detection manifestly agnostic to them, can help to identify and characterise features in measured spectra as caused by the presence of anomalies in a dataset.

B. Bortolato, A. Smolkovič, B.M. Dillon and J.F. Kamenik, Bump hunting in latent space, Physical Review D 105 (2022) 115009.

Methodology

Dataset Exploration & Kinematic Variable Selection

Kinematic variables selected from the Zγ (pp →H →Zγ) MC fast simulation dataset were used to test the hypothesis.

Model Development (Generative)

A Variational Autoencoder was selected and developed as the base model. VAEs are a tried and tested generative and classification models.

Model Derivative Developement

The base model has derivatives that could be better suited to the generation of the specific Zγ (pp →H →Zγ) dataset. These need to be compared to the VAE base.

Optimisation

The base model and derivatives require extensive hyper-parameter optimization to achieve results that are favourable fo the search.

Generative Capability Evaluation

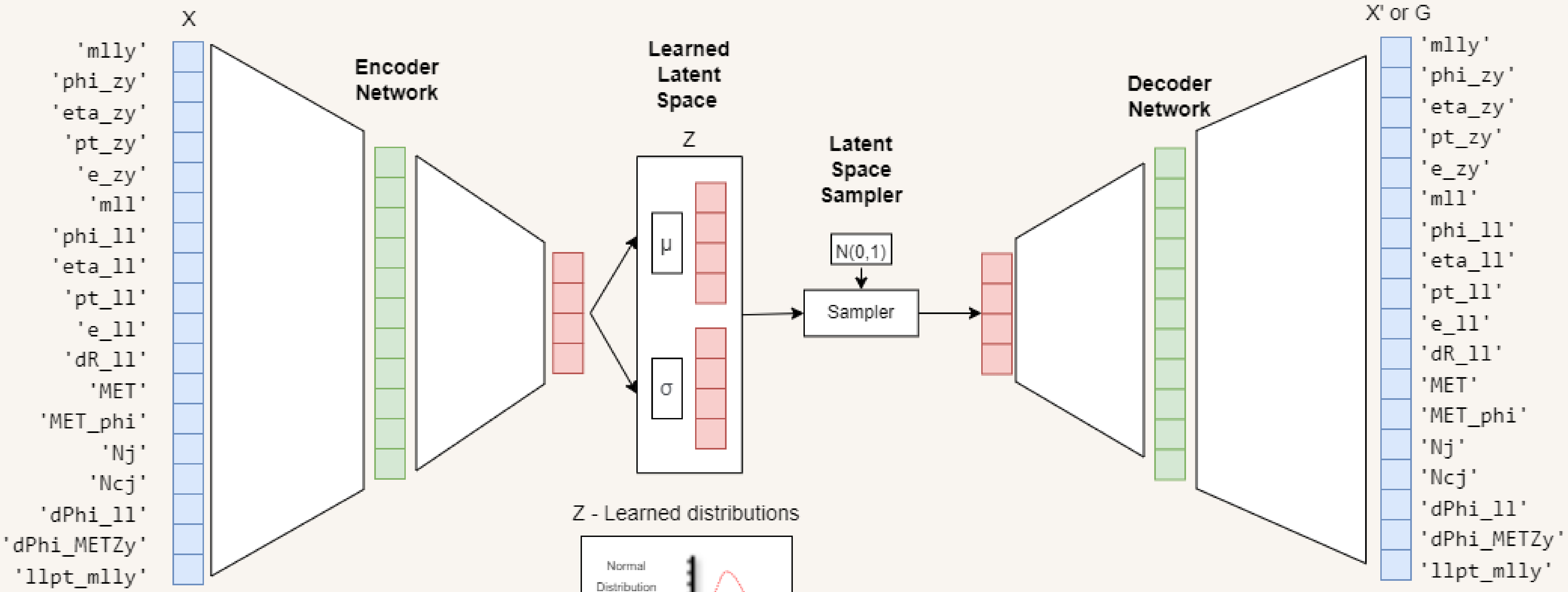
A number of evaluation metrics have been selected to evaluate the model's generative capability.

Classification Capability Evaluation

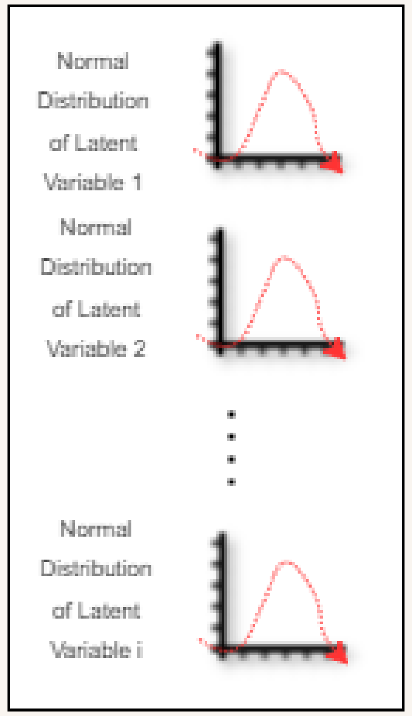
A number of evaluation metrics have been selected to evaluate the models classification capabilities.

VAE

During training, the VAE learns to **reconstruct the training events** that are fed forward through the network. Weights are learned using backpropagation and optimization of the loss function.



Z - Learned distributions



The VAE loss function consists of 2 terms:

- The **reconstruction loss**
- The **KL divergence**

The **reconstruction loss is responsible for minimizing the difference between an input event and a reconstructed event**, during training.
The **KL divergence loss term is a regularization term.**

The Variational-beta parameter is used to weight the importance of the KL-divergence loss term against that of the reconstruction loss term.

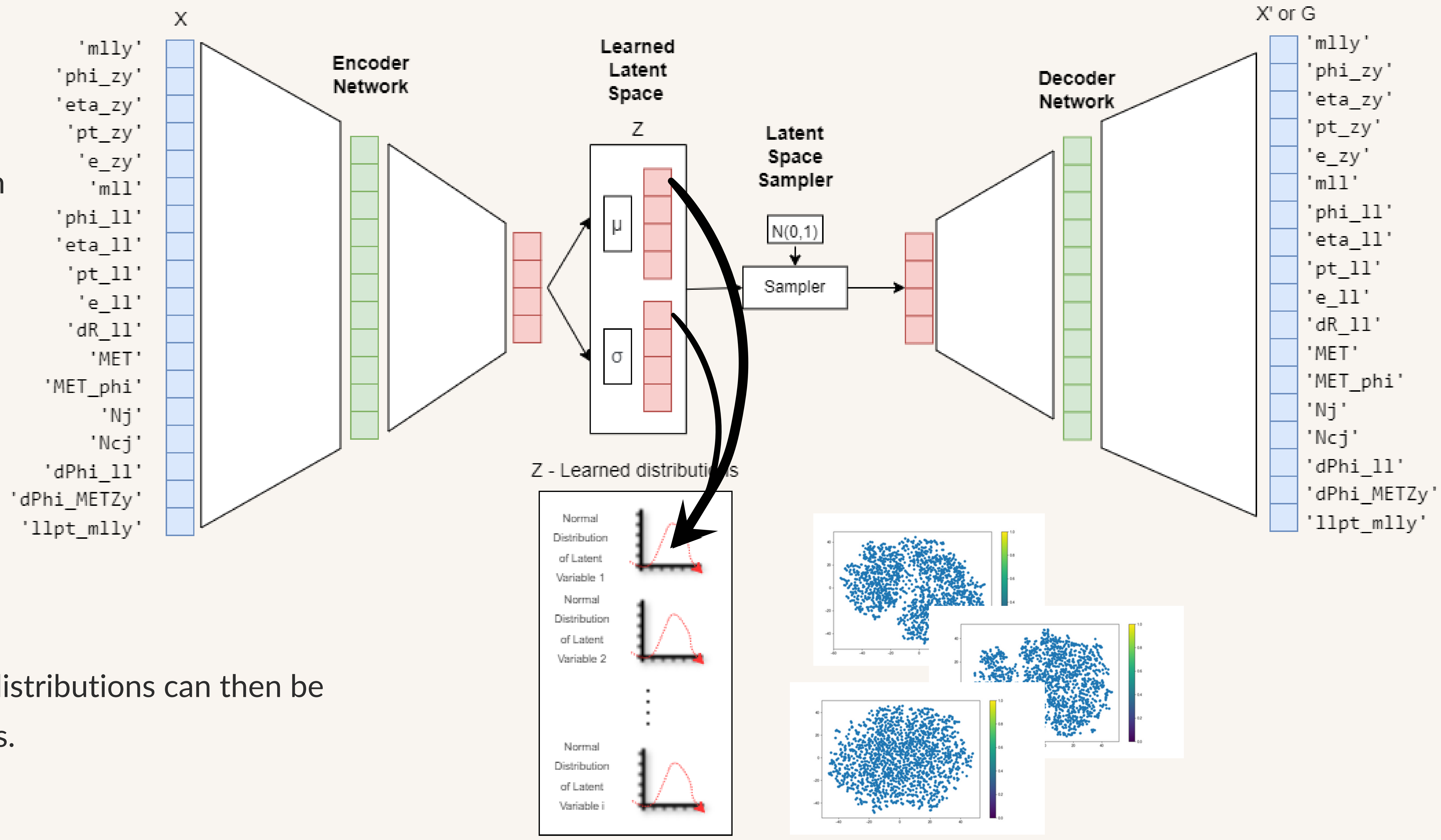
$$VAE_{loss} = L_R + \beta_{var} * L_{KL}$$

$$L_R = MSE(X, X')$$

$$L_{KL} = \sum p(x) \log \frac{p(x)}{q(x)}$$

VAE

Latent space variables (mean and variance) for each latent variable normal distribution are learned through variational inference (as a result of the addition of the KL divergence to the loss function).



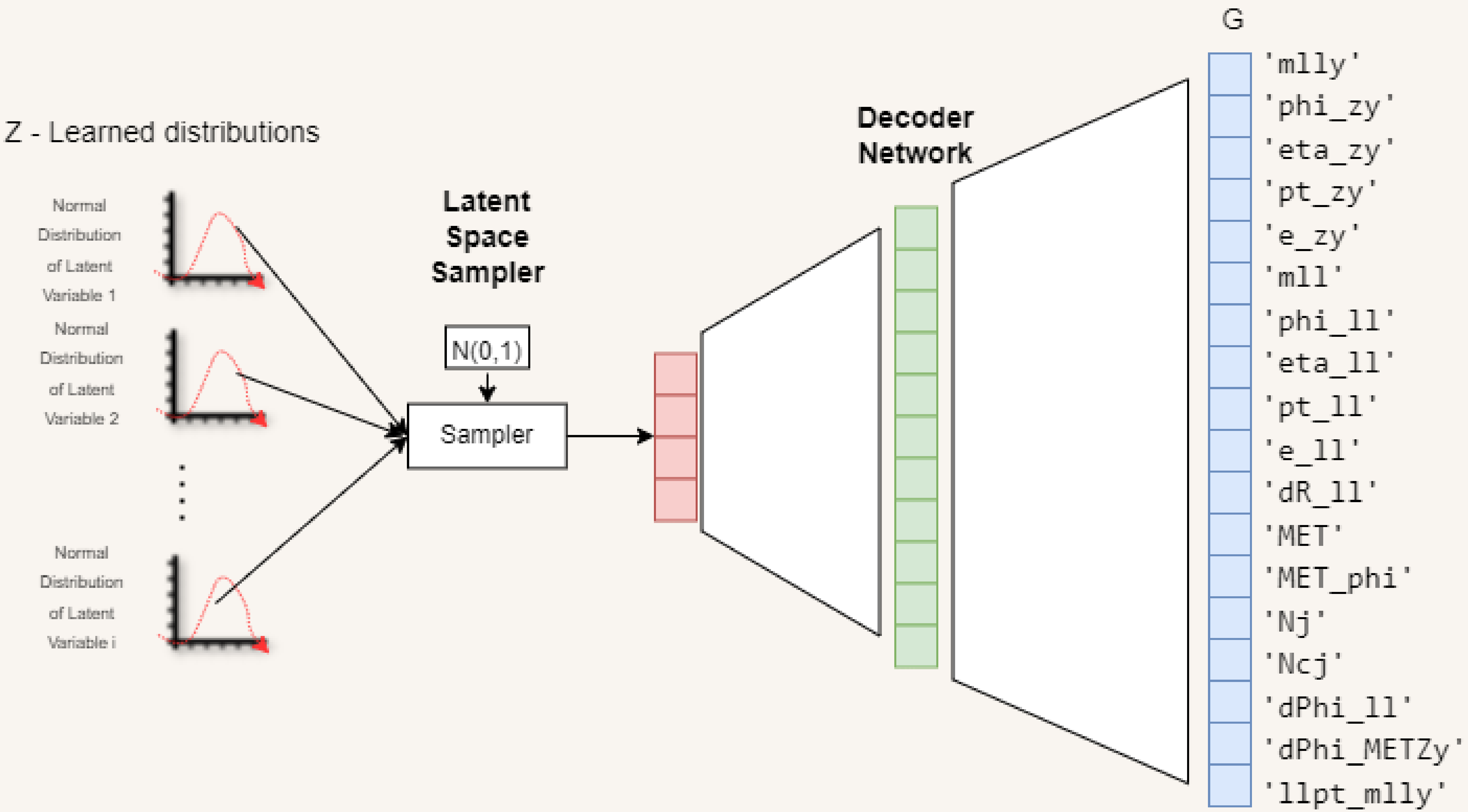
These learned latent space distributions can then be used to generate new events.

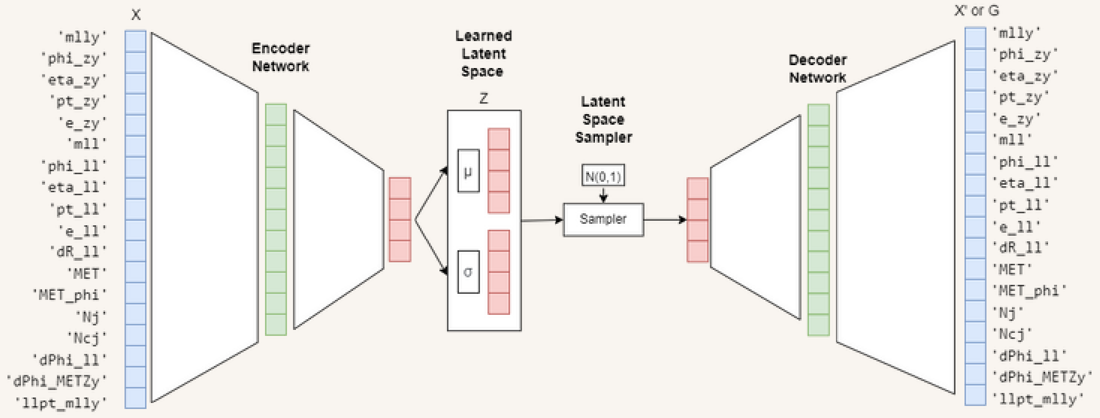
$$VAE_{loss} = L_R + \beta_{var} * L_{KL}$$

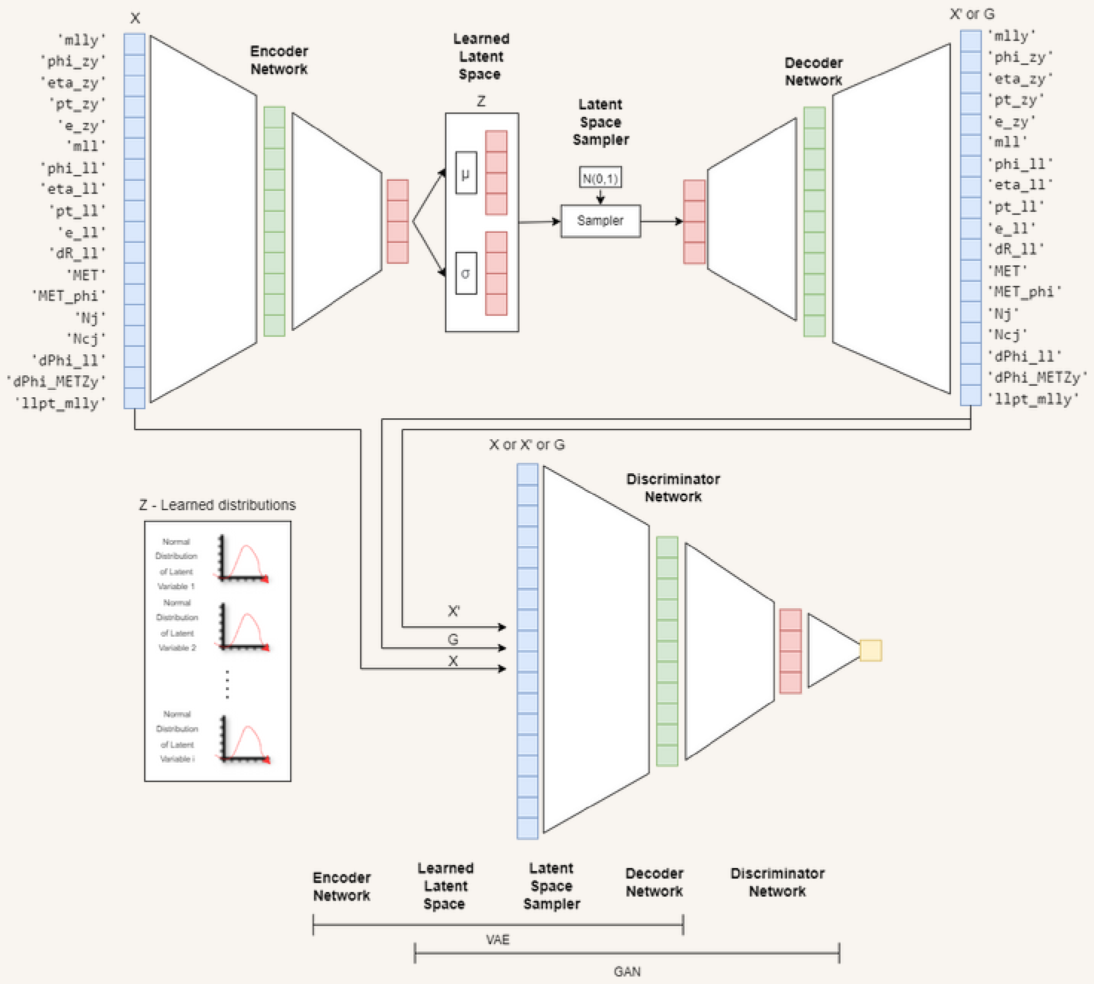
VAE

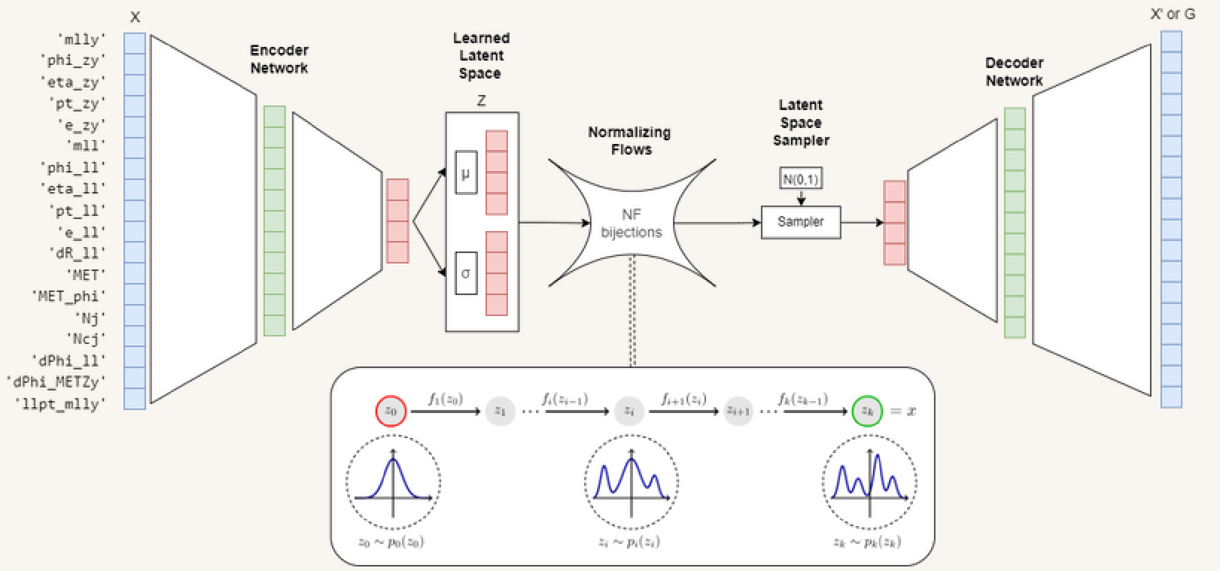
The diagram on the right shows how the latent space learned distributions can be sampled from and the sample can be fed forward through the decoder in order to generate a new event.

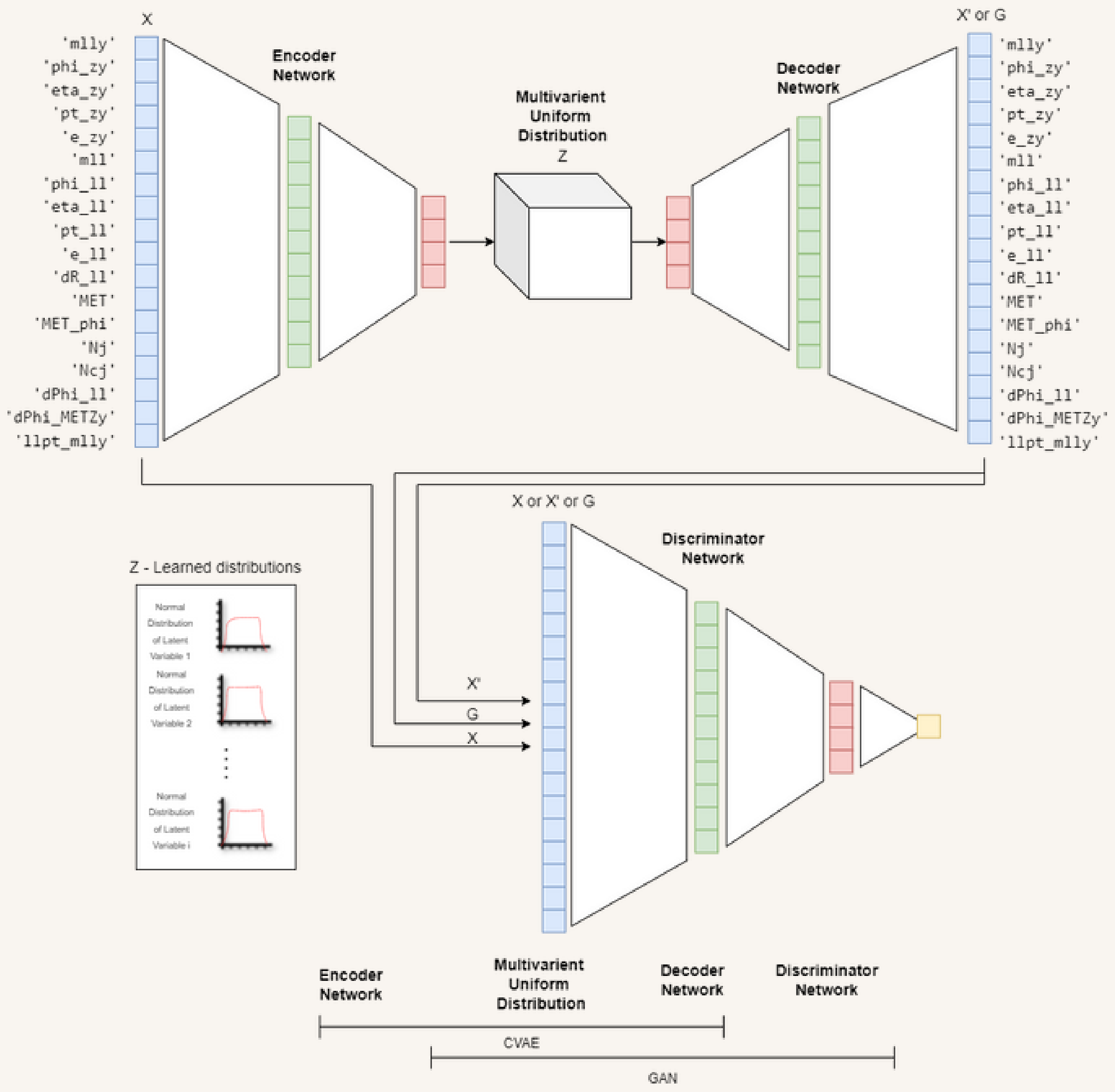
This can be done using random sampling from the latent space to generate many events.



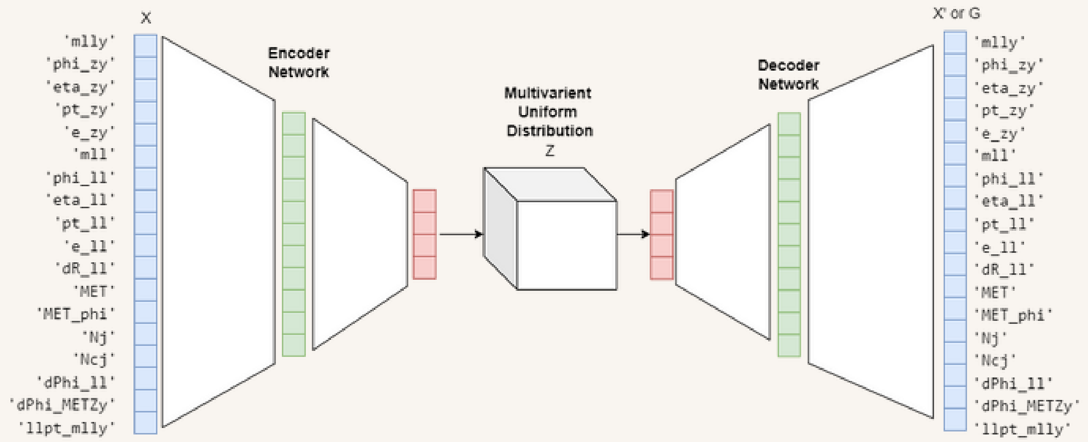


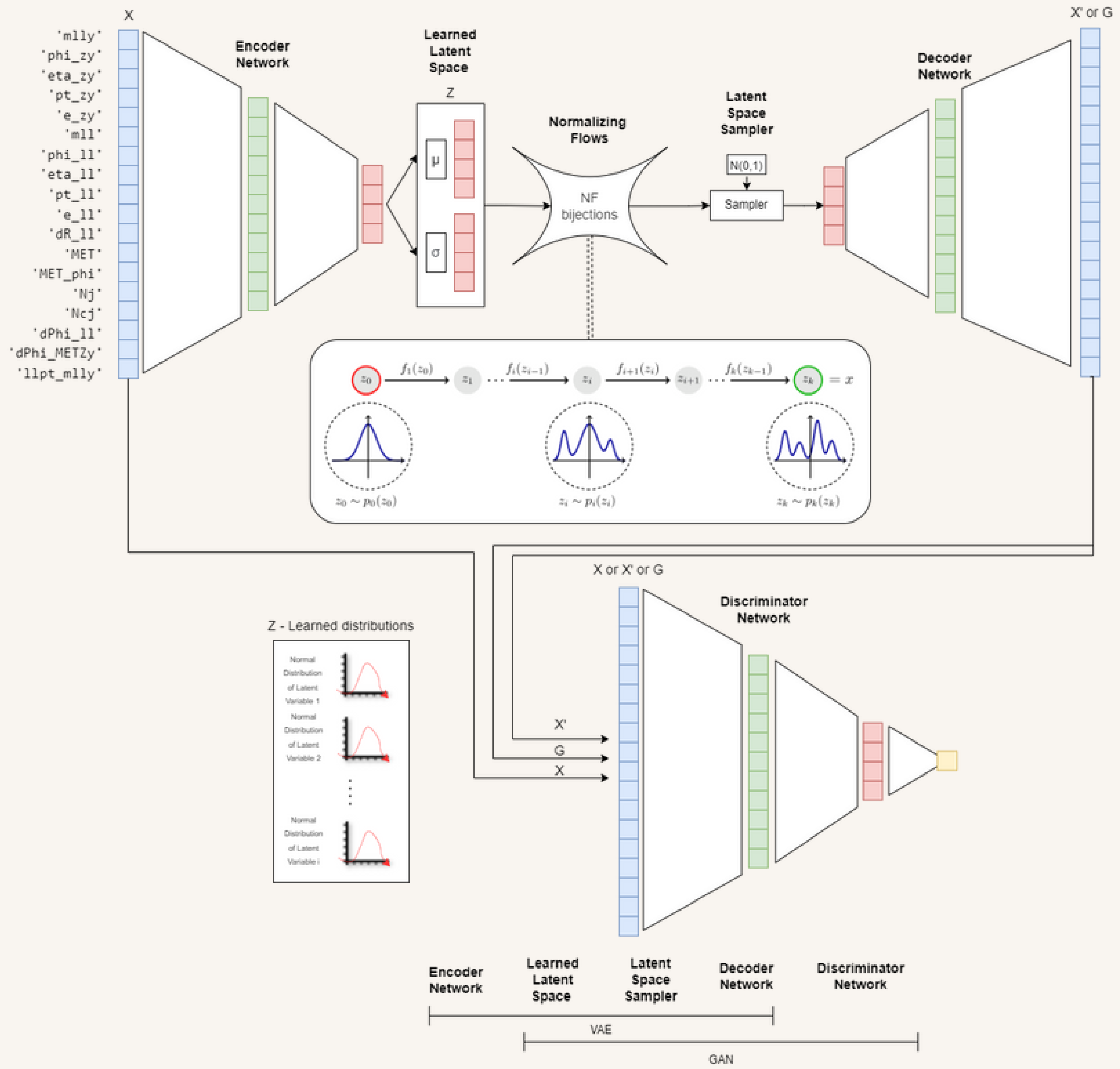






Derivatives

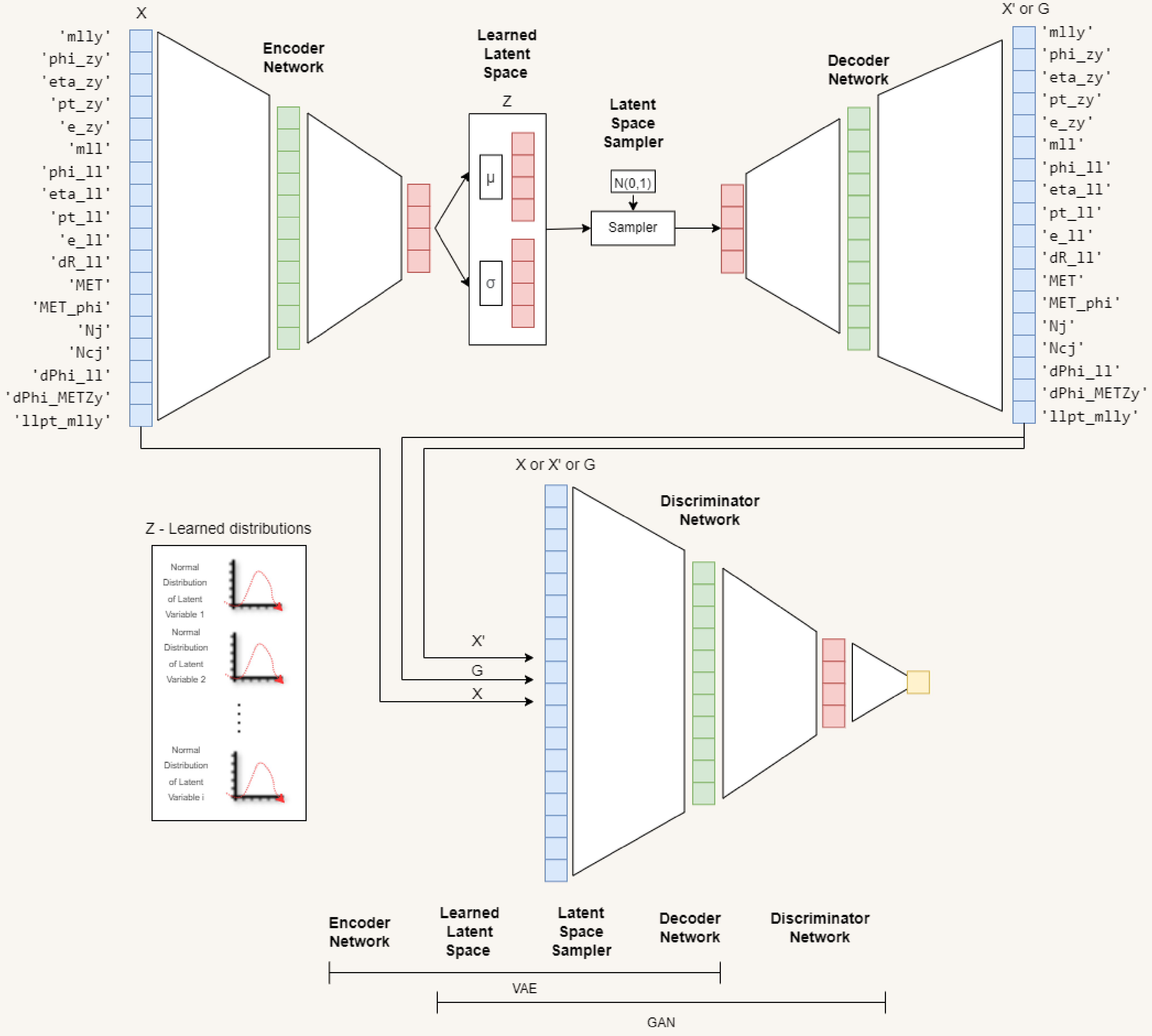




VAE+D

Addition of **Discriminator** to the VAE to make it a VAE+D

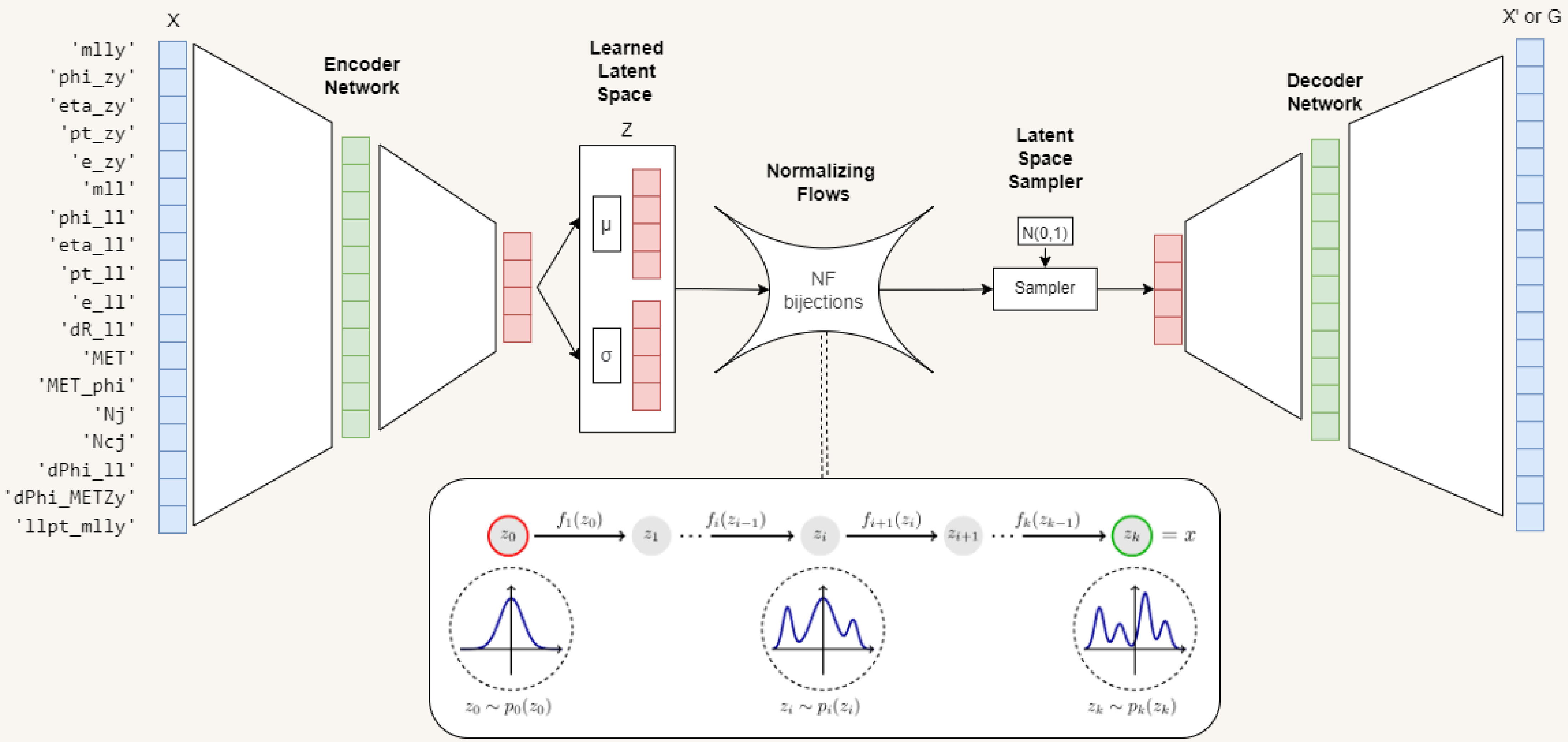
- Different training loop iteration
- 3 separate optimizers
- Notion of adversarial training
- Helps with the training of VAE
- Discriminator can be used for signal classification once trained



VAE+NF

Addition of **Normalising Flows** to the VAE

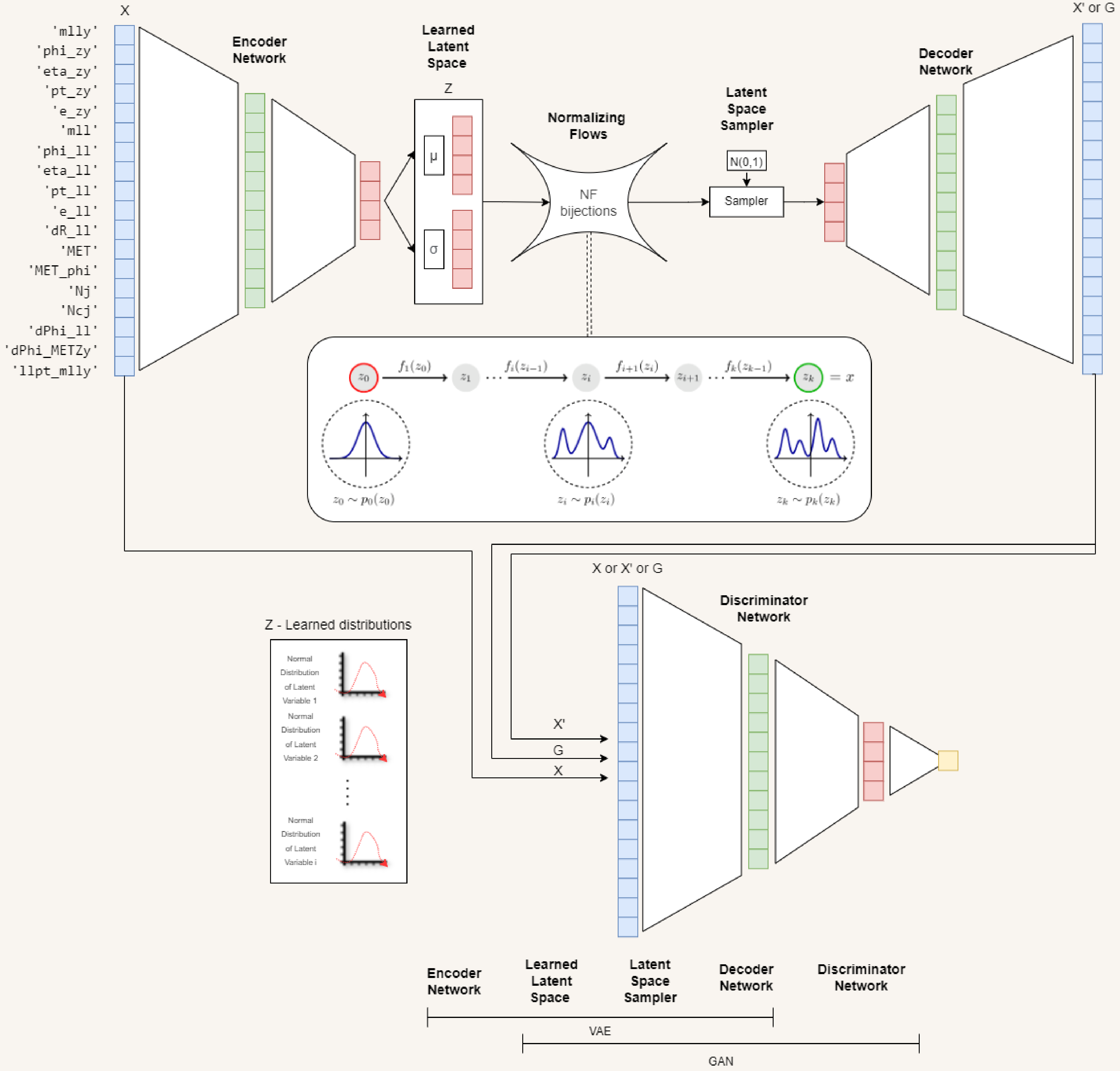
- Change of variable formula used in NF for transformations
- Multiple bijector transformations applied to gaussian latent space
- Allows for more complex information to be retained in latent space



VAE+NF+D

Addition of **Discriminator** and **Normalising Flows** added to the VAE

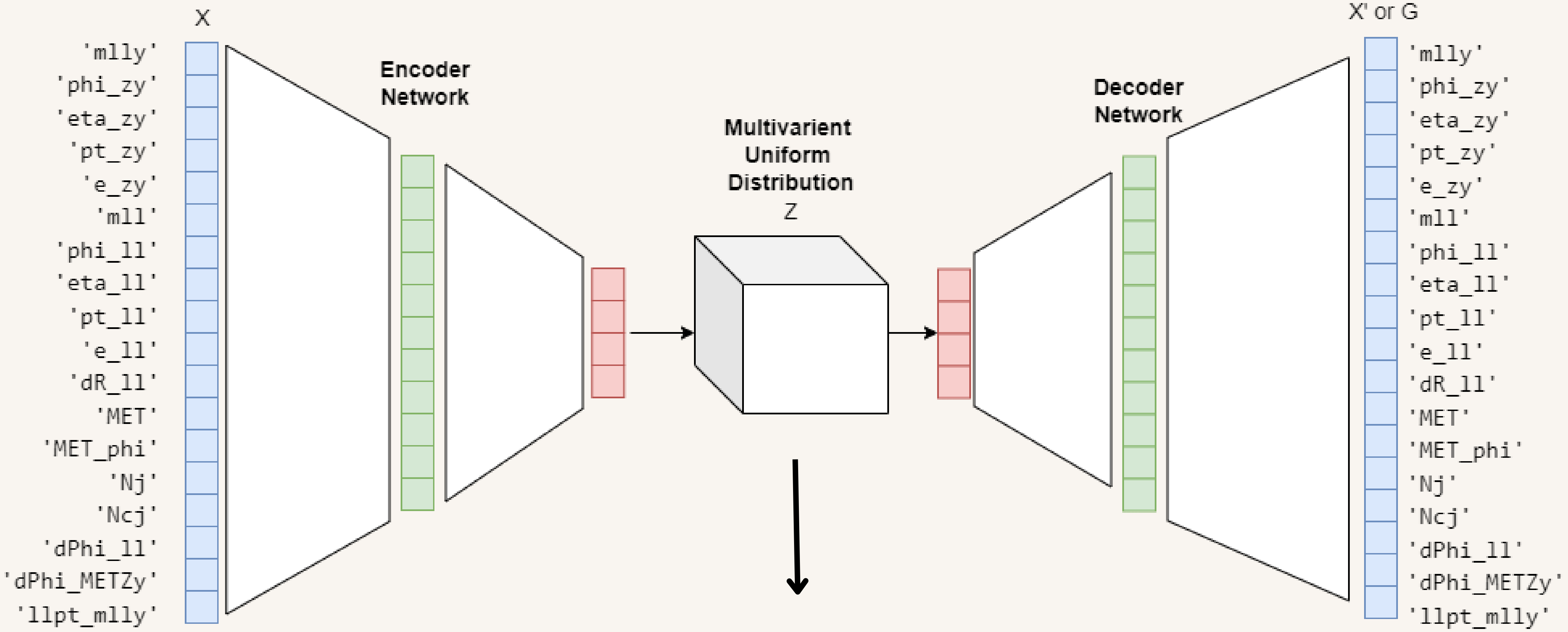
- Helps with the training of VAE
- Discriminator can be used for signal classification once trained
- Multiple bijector transformations applied to gaussian latent space
- Allows for more complex information to be retained in latent space

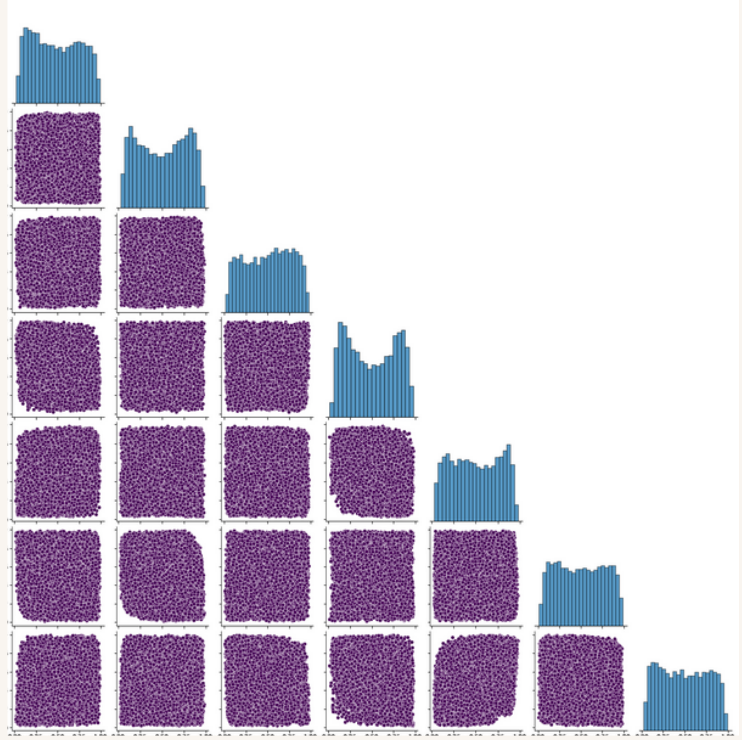


CVAE

Conversion of VAE to **Copula-VAE**

- Latent Space is a multi-variant uniform distribution
- Allows for more valuable information to be retained in the latent space
- Encoder and decoder can be trained separately due to the prior knowledge of the uniform latent space
- Correlations could possibly be better learned due to the ability of the model to learn the CDFs of the input variables



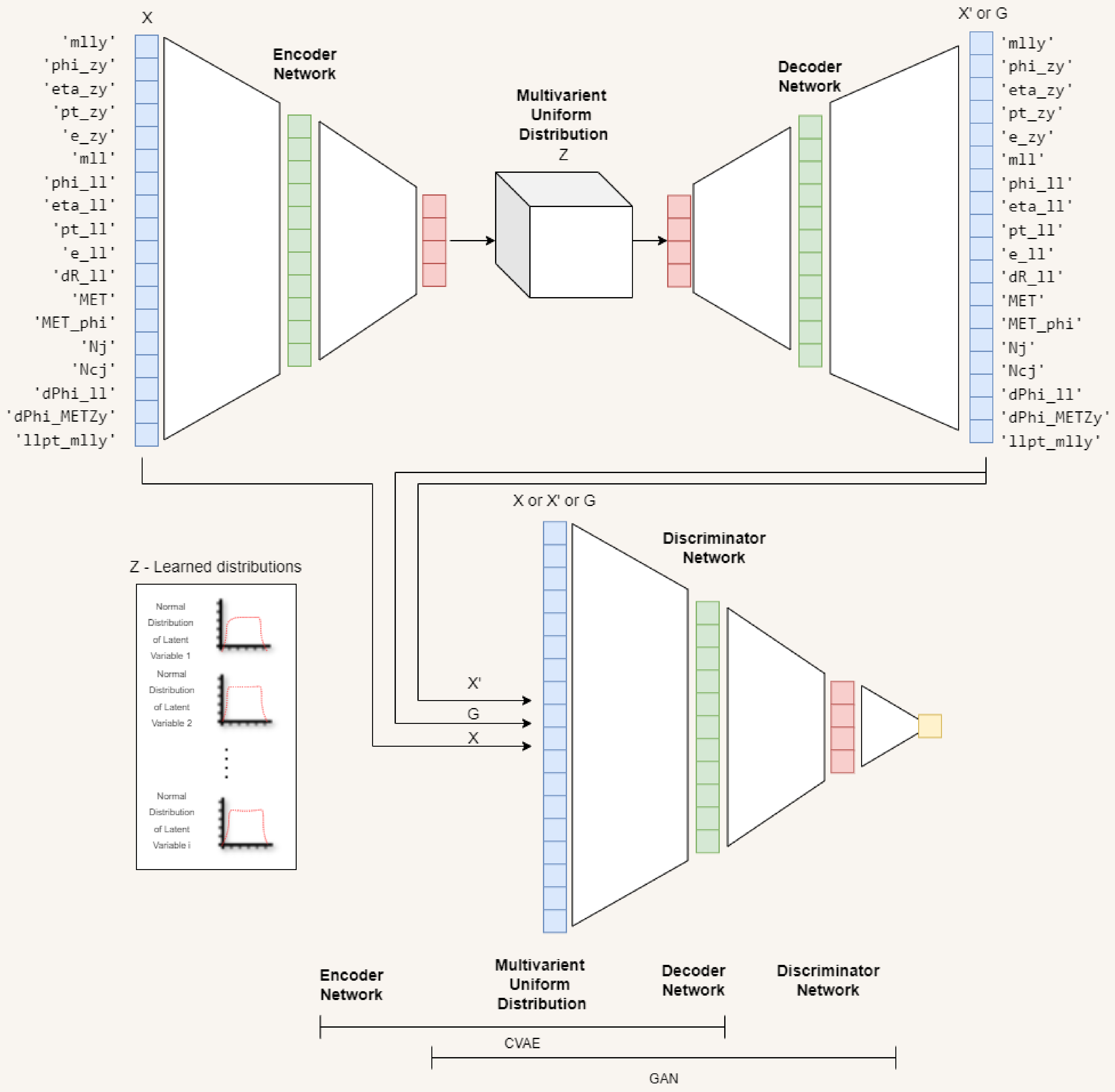


Robbertze, Y. 2022. Neural network libor market model for pricing and hedging interest rate derivatives. . ,Faculty of Science ,Department of Statistical Sciences. http://hdl.handle.net/11427/36545

CVAE +D

Addition of **Discriminator** to the **CVAE**

- Helps with the training of VAE
- Discriminator can be used for signal classification once trained
- Latent Space is a multi-variant uniform distribution
- Allows for more valuable information to be retained in the latent space
- Encoder and decoder can be trained separately due to the prior knowledge of the uniform latent space



VAE Optimisation

Hyper-parameters can be optimized in order to optimize generation accuracy. The following hyper-parameters were selected for hyper-parameter optimization:

VAE:

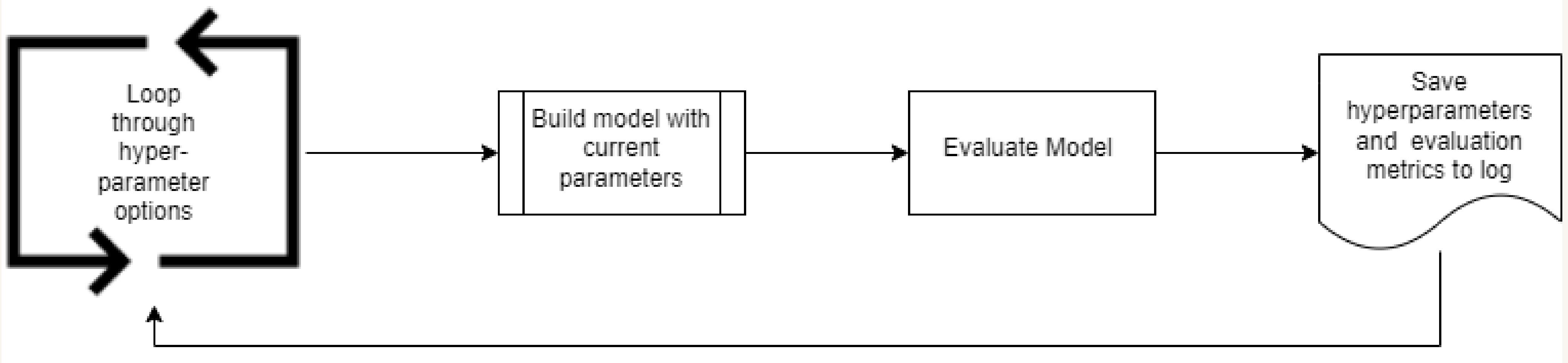
- Training batch_size [real]
- Learning rate [float]
- Activation function [str]
- Latent dimension size [real]
- Number of hidden layers [real]
- Number of nodes in hidden layers [real]
- Variational beta [float]

+D:

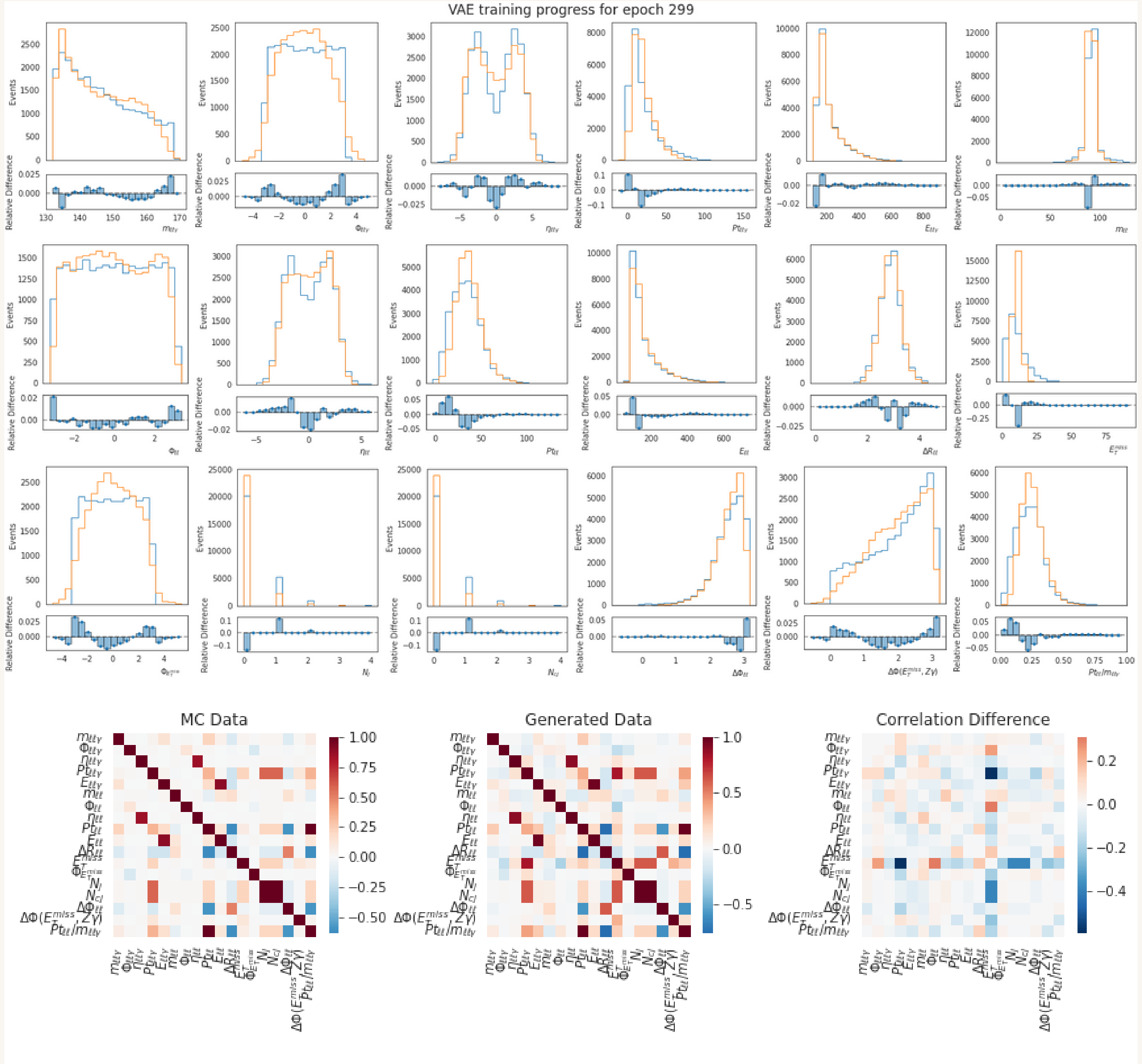
- Alpha [float]
- Gamma [float]

+NF

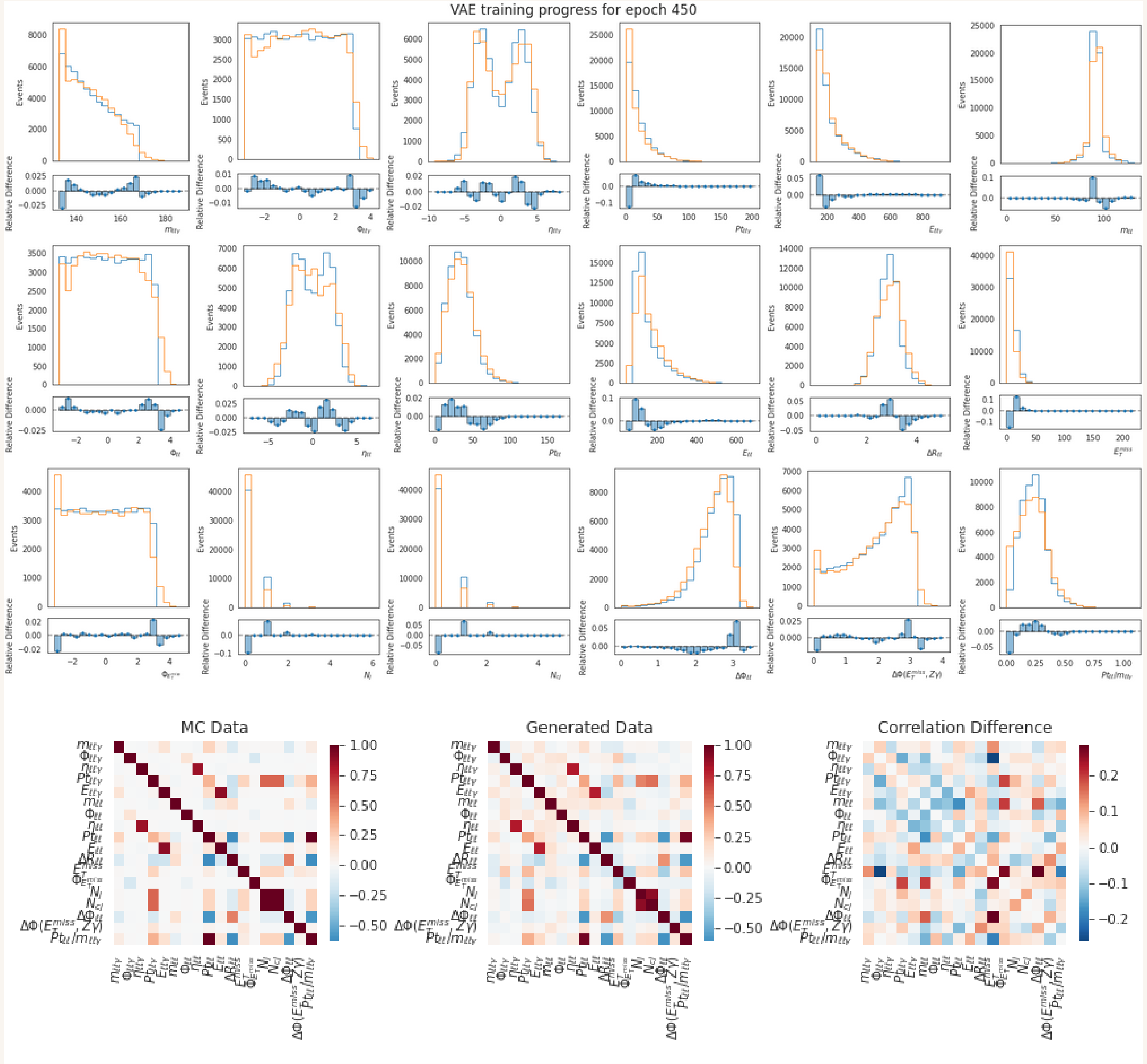
- Flow algorithm [type]
- Flow length [real]



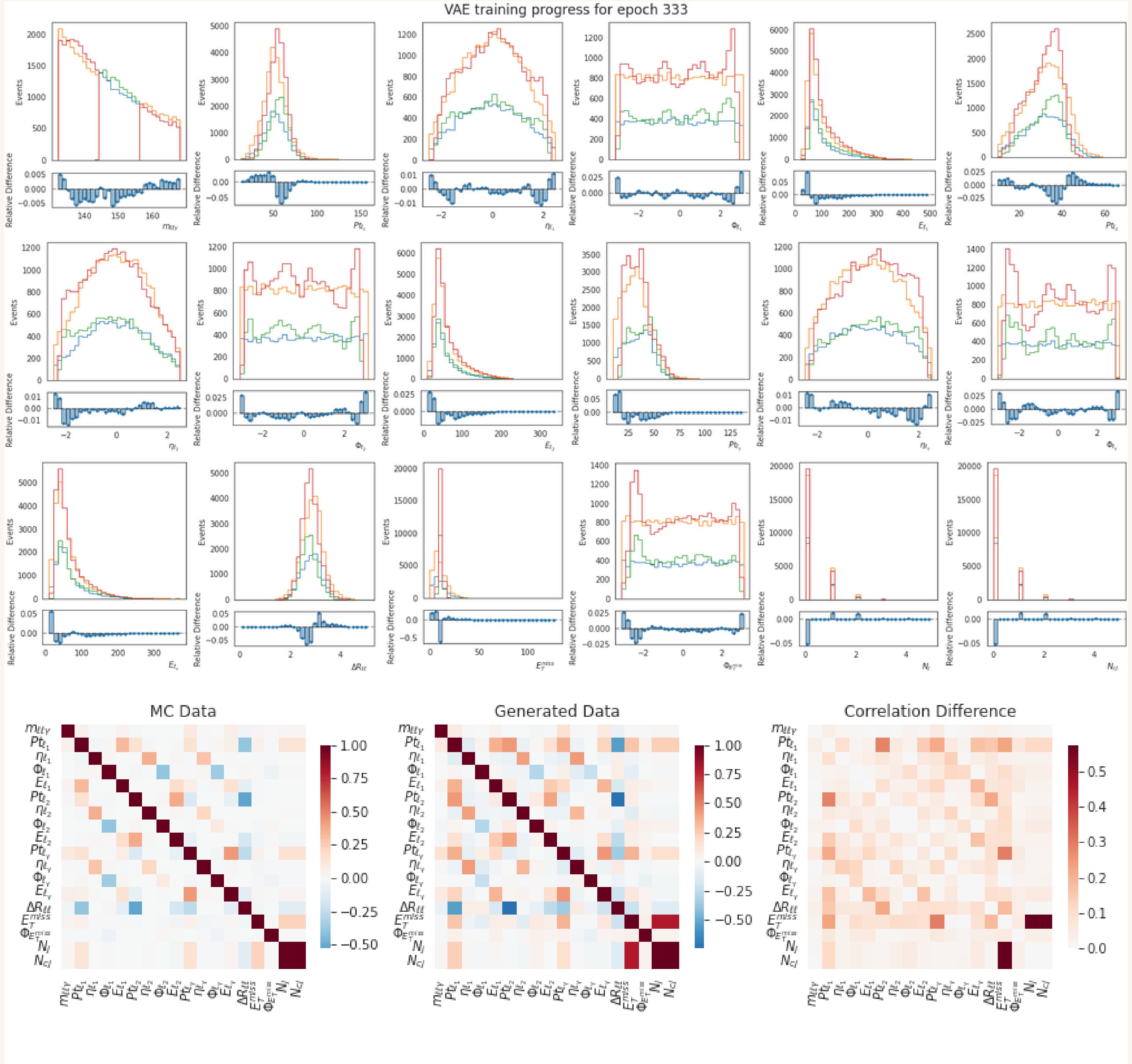
VAE Results



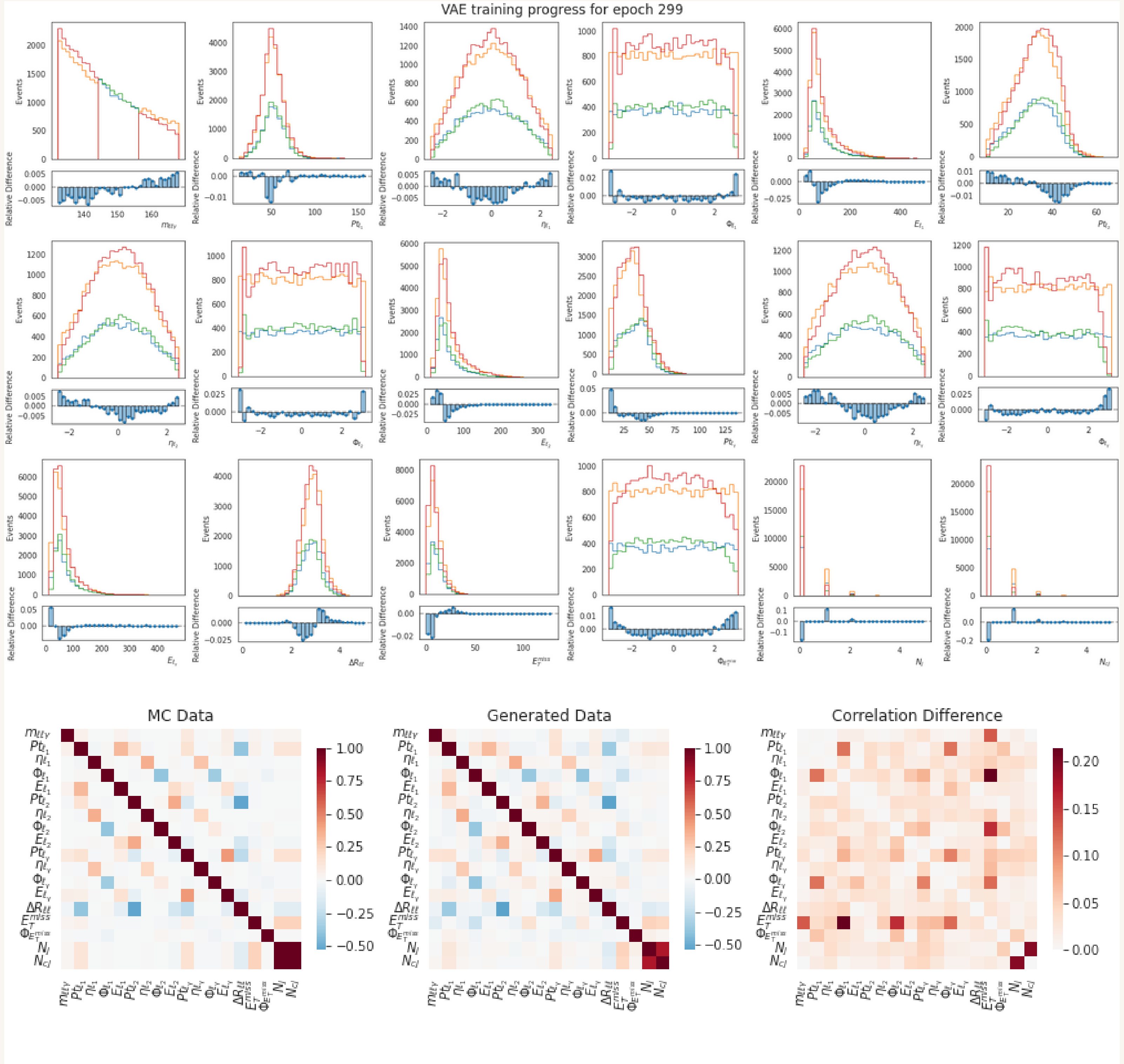
VAE+D Results



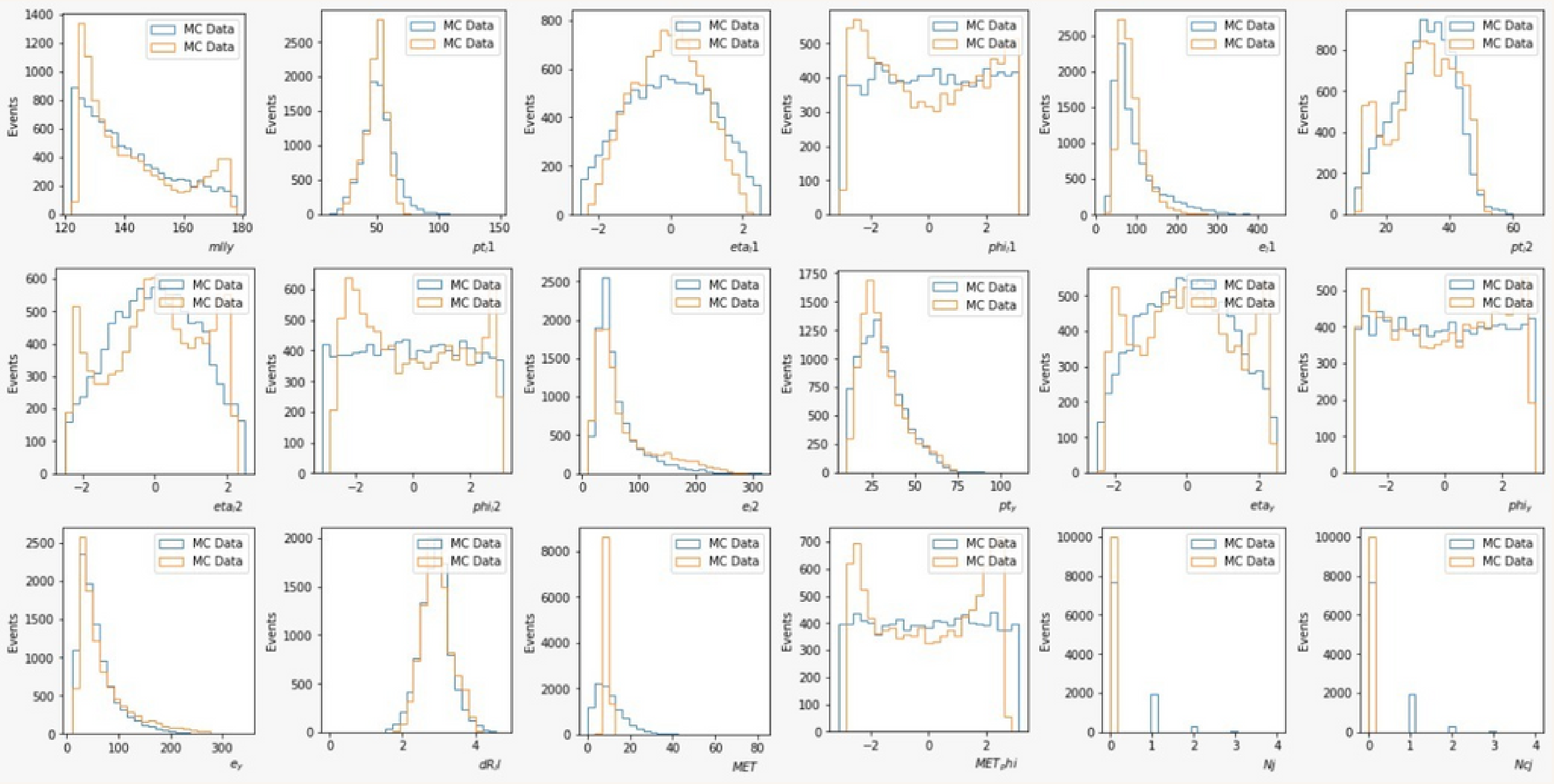
VAE+NF Results

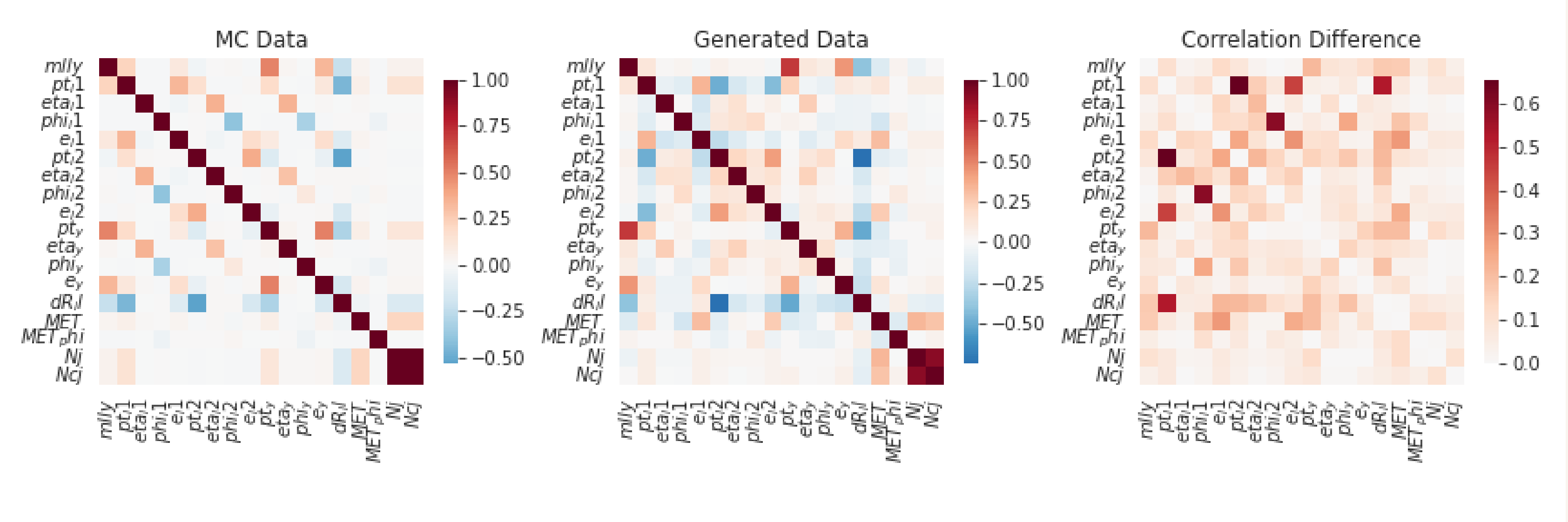


VAE+NF+D Results



CAE Results





Comparison and Disscussion

The results of the Data Generation evaluation are can be simplified as follows:

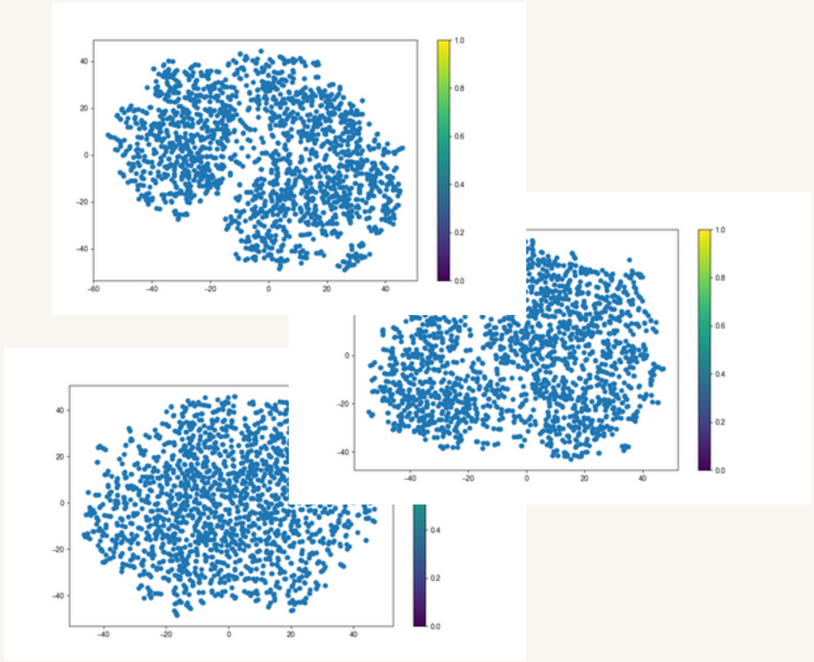
Model	Best for Distribution	Best for Correlation	Best Overall	Comment
VAE				The VAE base model performed well as a base model. However training of the VAE is difficult.
VAE+D	X			Addition of the discriminator network helped to achieve convergence in training. Better results in terms of distributions and correlation were achieved.
VAE+NF				The addition of NF to VAE showed promise in terms of the distribution comparison results. Correlation results worse than VAE+D. Better overall results than VAE+D were not achieved.
VAE+NF+D		X	X	With the addition of both the discriminator and NF, the model was able to generate data that produced much better correlation results and almost as good distribution results.
CVAE				The copula VAE showed promise in terms of the methodology but further optimization is required to achieve better results to compare to the VAE+NF+D.
CVAE+D				Similarly to the CVAE, further optimization is required to properly evaluate the overall generative capability.

VAE + Derivatives for Classification

There are a number of different mechanisms for classification:

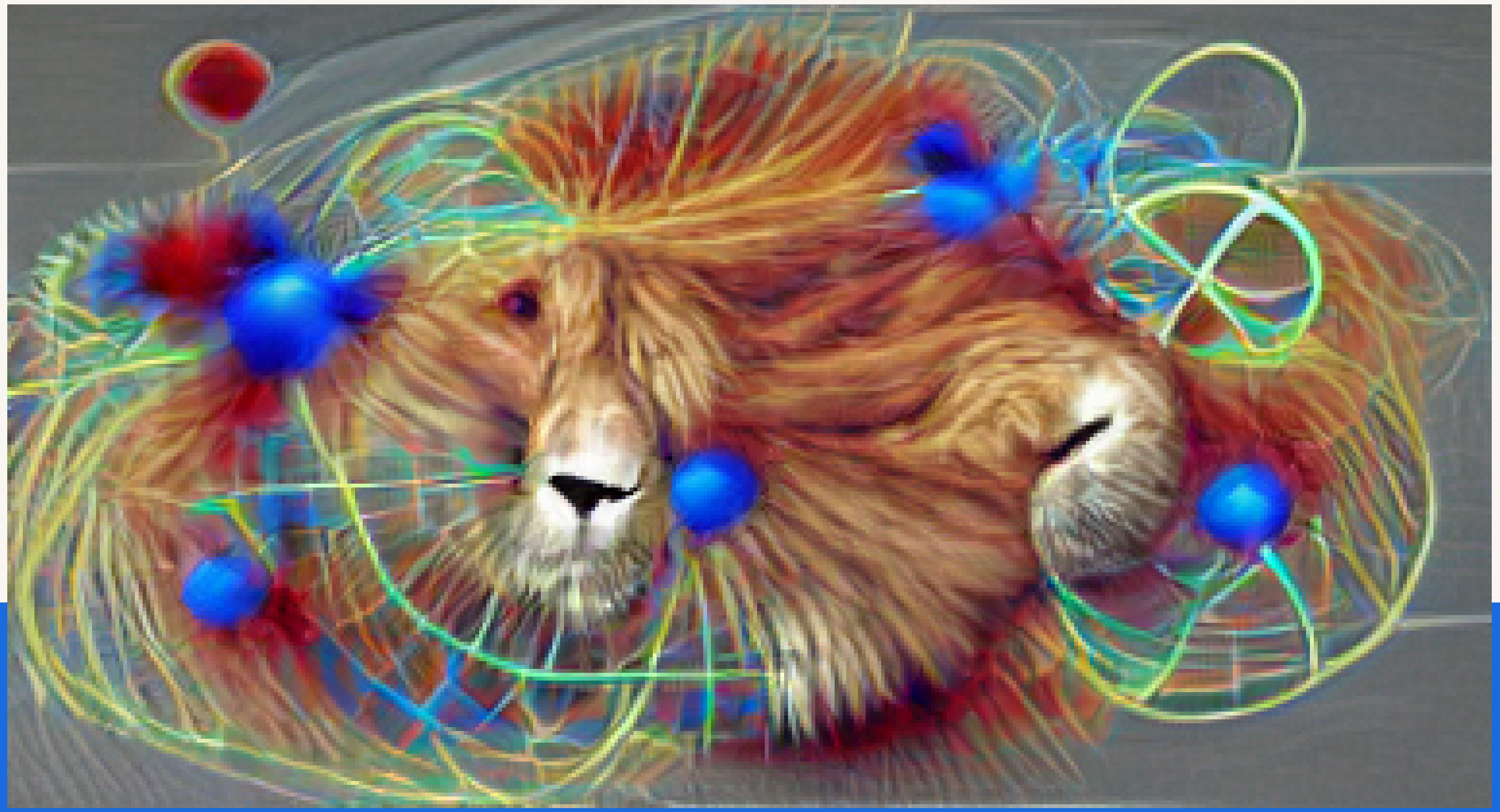
- VAE
  - Use of reconstruction loss
  - Examination of latent space and PCA of latent space
- +D
  - Use of discriminator network
- +NF
  - Latent space properties
- CVAE
  - Outliers in uniform latent space

$$L_R = MSE(X, X')$$



Next Steps

- Complete hyper-parameter optimization for:
  - CVAE
  - CVAE+D
- Complete Final Model Comparison and Evaluation in terms of Data Generation task.
- Use the best model in frequentist study
- Evaluate models on the ability to classify injected signals.



Supporting Slides

VAE/VAE+D Training

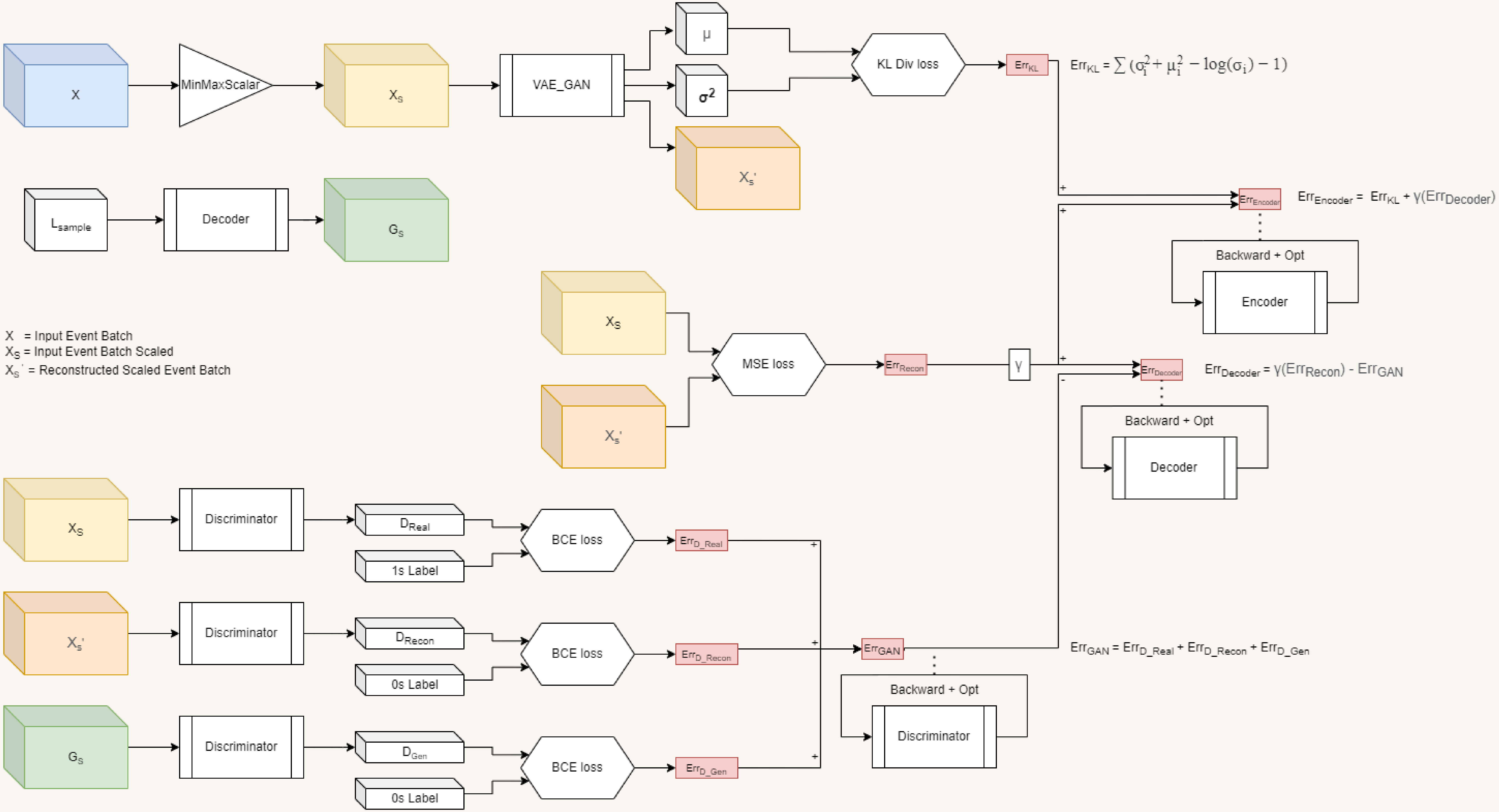
- Training algorithms for VAE and VAE+D
- VAE+D training includes training of discriminator
- VAE+D has three separate optimizers for Encoder, Decoder, and Discriminator

Algorithm 1 VAE Forward Pass

$\mu, \sigma^2, reconstructed \leftarrow VAE \leftarrow eventbatch$
$generated \leftarrow Decoder \leftarrow z_p$
$((reconstructed - eventbatch)^2).mean() \leftarrow Loss_{recon}$
$1 + \sigma - \mu^2 - e^{\sigma} \leftarrow Loss_{KL}$
$Loss_{recon} + \beta_v Loss_{KL} \leftarrow Loss_{VAE}$
VAE backward step

Algorithm 2 VAE+D Forward Pass

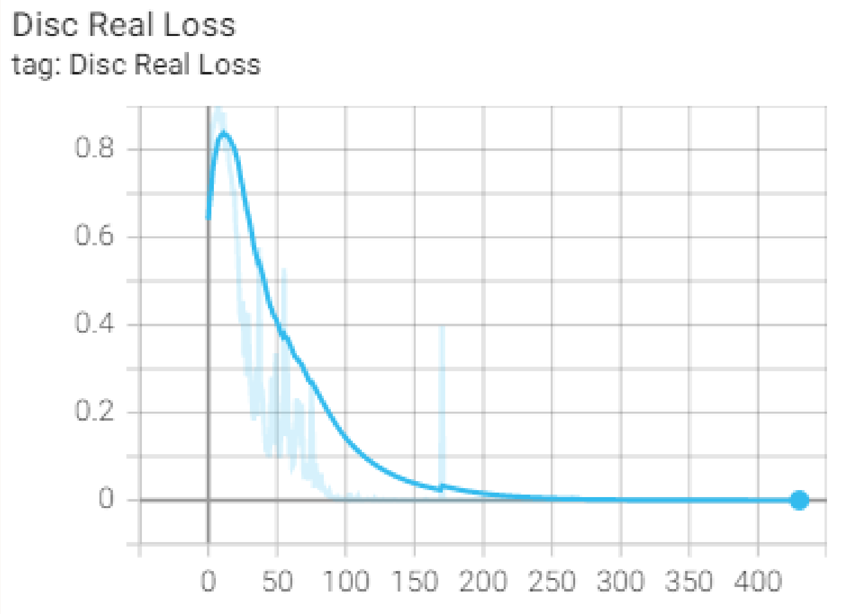
$\mu, \sigma^2, reconstructed \leftarrow VAE \leftarrow eventbatch$
$generated \leftarrow Decoder \leftarrow z_p$
$diss_{real} \leftarrow Diss \leftarrow eventbatch$
$Err_{D-real} \leftarrow criterion_{BCE} \leftarrow diss_{real} vs.1slabel$
$diss_{recon} \leftarrow Diss \leftarrow reconstructed$
$Err_{D-recon} \leftarrow criterion_{BCE} \leftarrow diss_{recon} vs.0slabel$
$diss_{gen} \leftarrow Diss \leftarrow generated$
$Err_{D-gen} \leftarrow criterion_{BCE} \leftarrow diss_{gen} vs.0slabel$
$Err_{D-gen} + Err_{D-recon} + Err_{D-real} \leftarrow Loss_{disc}$
Discriminator backward step
$((diss_{recon} - diss_{real})^2).mean() \leftarrow Loss_{recon}$
$\gamma * Loss_{recon} - Loss_{disc} \leftarrow Err_{Decoder}$
Decoder backward step
$1 + \sigma - \mu^2 - e^{\sigma} \leftarrow Loss_{KL}$
$Loss_{KL} + \gamma * Loss_{recon} \leftarrow Err_{Encoder}$
Encoder backward step

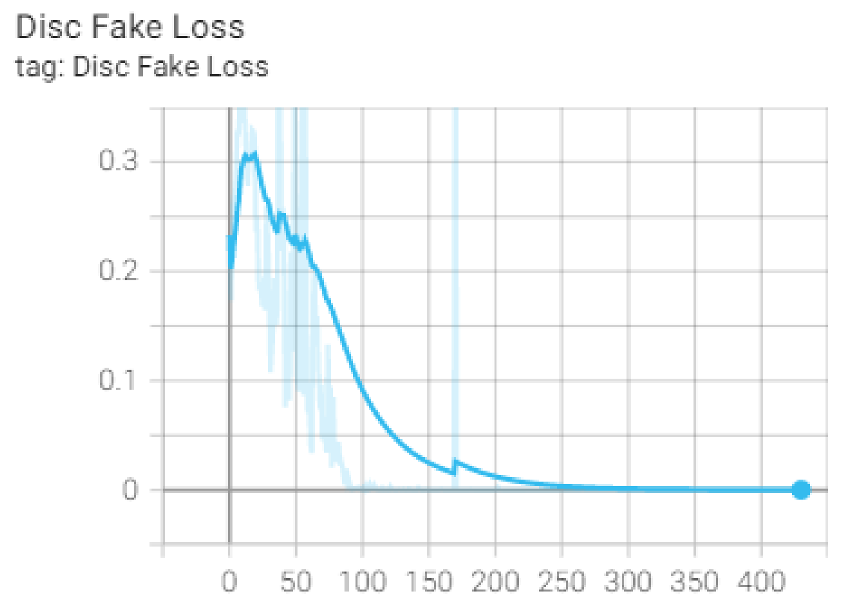


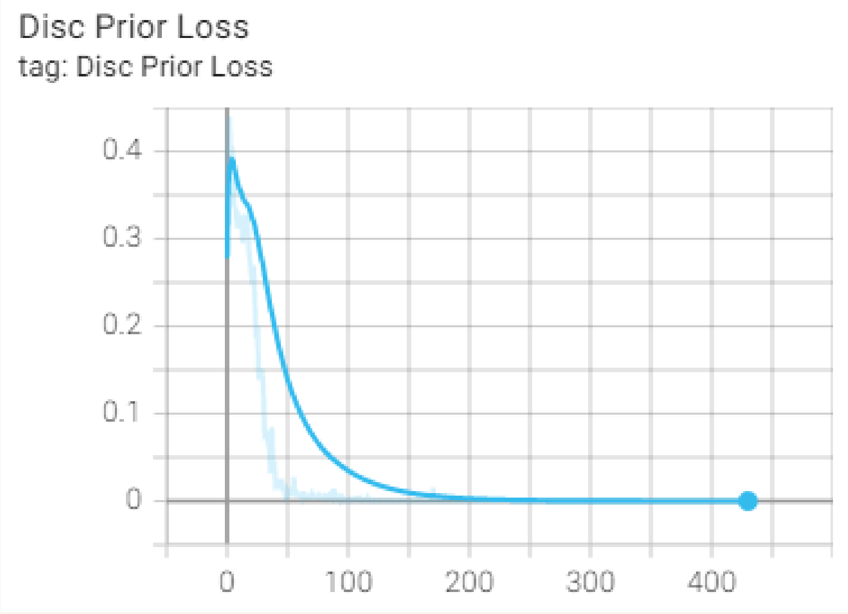
X
MinMaxScalar
X_S
VAE_GAN
μ
σ^2
X_S'
KL Div loss
Err_{KL}
$Err_{KL} = \sum(\sigma_i^2 + \mu_i^2 - \log(\sigma_i) - 1)$
L_{sample}
Decoder
G_S
$Err_{Encoder}$
$Err_{Encoder} = Err_{KL} + \gamma(Err_{Decoder})$
Backward + Opt
Encoder
X = Input Event Batch
X_S = Input Event Batch Scaled
X_S' = Reconstructed Scaled Event Batch
MSE loss
Err_{Recon}
γ
$Err_{Decoder}$
$Err_{Decoder} = \gamma(Err_{Recon}) - Err_{GAN}$
Backward + Opt
Decoder
Discriminator
D_{Real}
1s Label
BCE loss
Err_{D_Real}
D_{Recon}
0s Label
Err_{D_Recon}
Err_{GAN}
$Err_{GAN} = Err_{D\_Real} + Err_{D\_Recon} + Err_{D\_Gen}$
D_{Gen}
Err_{D_Gen}
Backward + Opt
Discriminator

VAE+D Losses and Metrics Best Run

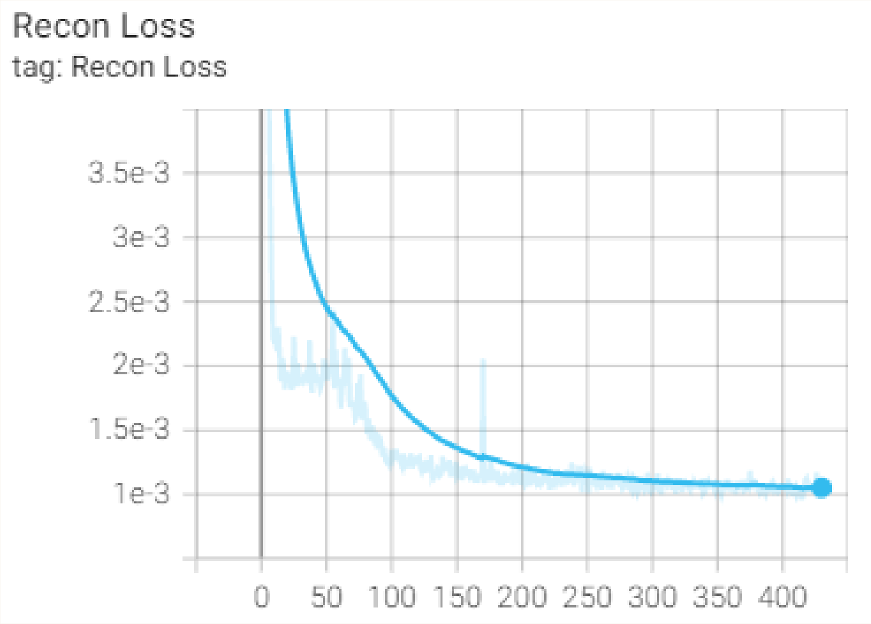
Dsicriminator Losses:

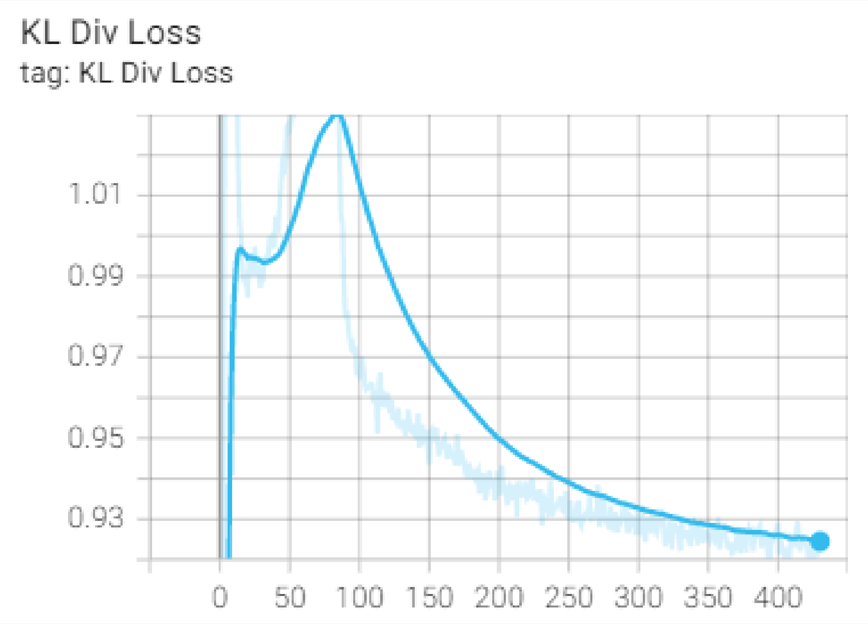




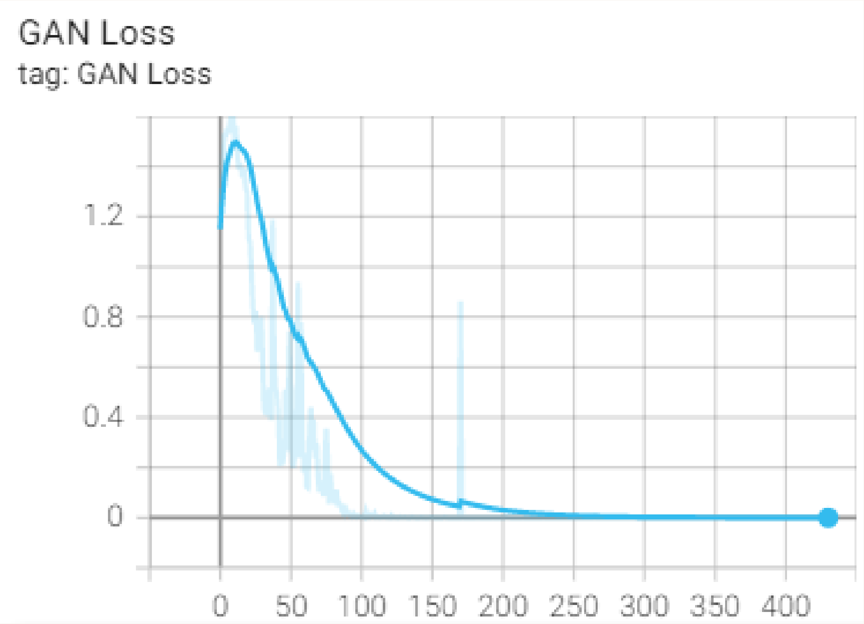


VAE Losses:





GAN Loss:



VAE+D Losses and Metrics Best Run

Eval Metrics: