1st COMETA General Meeting, Izmir 2024

An Anomaly Detection strategy for New Physics searches at the LHC

Giacomo Boldrini[3], Simone Gennai[1], Raffaele Gerosa[1], Pietro Govoni[1,2], Andrea Massironi[1], **Giulia Lavizzari**[1,2]

[1] INFN sezione di Milano Bicocca
[2] Università degli Studi di Milano Bicocca
[3] Ecole Polytechnique, LLR

Istituto Nazionale di Fisica Nucleare
Sezione di Milano-Bicocca

UNIVERSITA' DEGLI STUDI DI MILANO BICOCCA

Aim of the work

Probing the use of **unsupervised learning methods for anomaly detection** to identify signals of new physics at LHC → Variational AutoEncoders

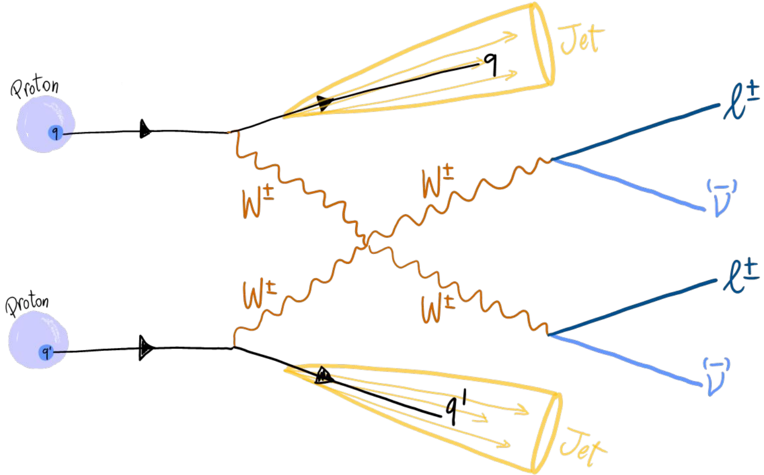
- anomalous contributions modeled through **SMEFT** (**dimension-6** effects)
- physics use case: **same-sign WW scattering** in a fully leptonic final state

Why VBS for BSM searches:

Deeply connected to Higgs mechanism (probe of the SM sensitive to modifications of the EWK sector)

Tree level sensitivity to:

- triple and quartic gauge couplings
- Higgs-gauge couplings away from mass shell



SM Effective Field Theory (SMEFT)

The **SM** is seen as a **low-energy approximation** of a more complete theory:

$$\mathcal{L}_{EFT} = \mathcal{L}_{SM} + \sum_{i,d>4} \frac{c_i}{\Lambda^{d-4}} \mathcal{O}^{(d_i)}$$

Λ – new physics scale

$\mathcal{O}^{(d_i)}$ – EFT operator of dimension d_i

c_i – Wilson coefficient

- BSM effects are parametrized as additional terms to the SM lagrangian, which contain **higher order operators**
- Their intensity is gauged by **Wilson coefficients**

→ The first non-zero term (after SM) is **dimension 6**

- **parton-level**, @LO generations of SSWW events (SMEFTsim)
- integrated luminosity of 350/fb
- **backgrounds neglected**
- Dim 6 operators, chosen from the Warsaw basis
- one operator at a time

Why unsupervised learning

EFT is a complex, multidimensional problem:

- o (2500) parameters to constrain
- each operator affects differently each variable
  - hard to define a single observable to detect all operators

→ We want to **build a strategy that maximizes the observation of anything that is not Standard Model** (in principle we should see all the operators):

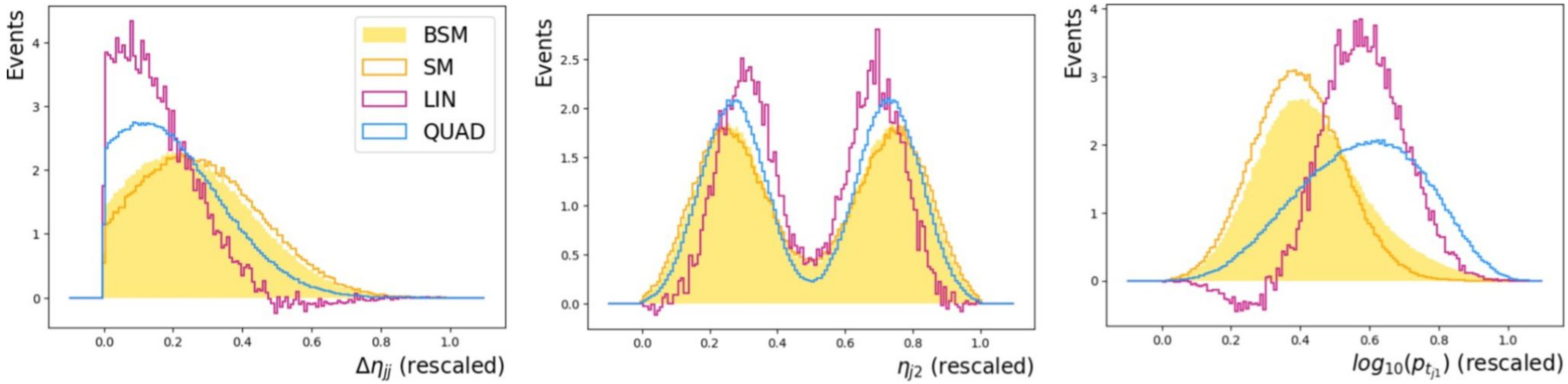
- Variational AutoEncoders
  - Unsupervised learning is an increasingly popular choice 2101.08320
- idea: **train a model on know physics, and later use it to detect outliers** (anomaly detection task)

Input distributions

The EFT operators modify the distributions of the variables, that now comprise:

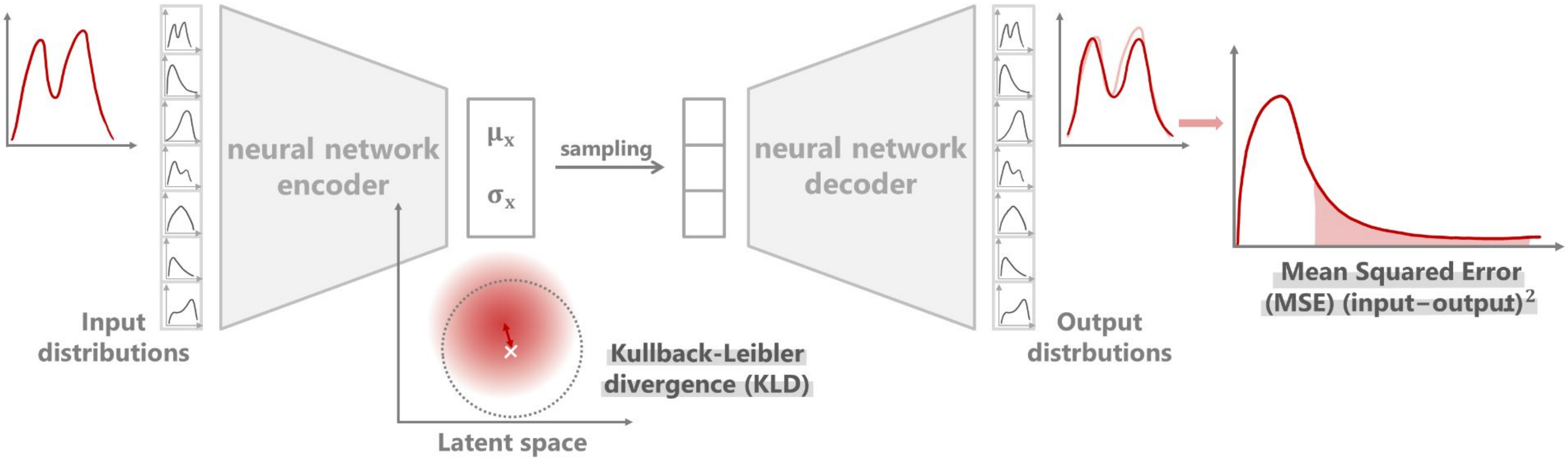
- A pure **SM contribution**
- Additional terms with **linear** and **quadratic** dependence on the EFT operator

$$|A_{EFT}|^2 = \underbrace{|A_{SM}|^2} + \underbrace{2Re(A_{SM}A^*_{op})} + \underbrace{|A_{op}|^2}$$



Variational AutoEncoders

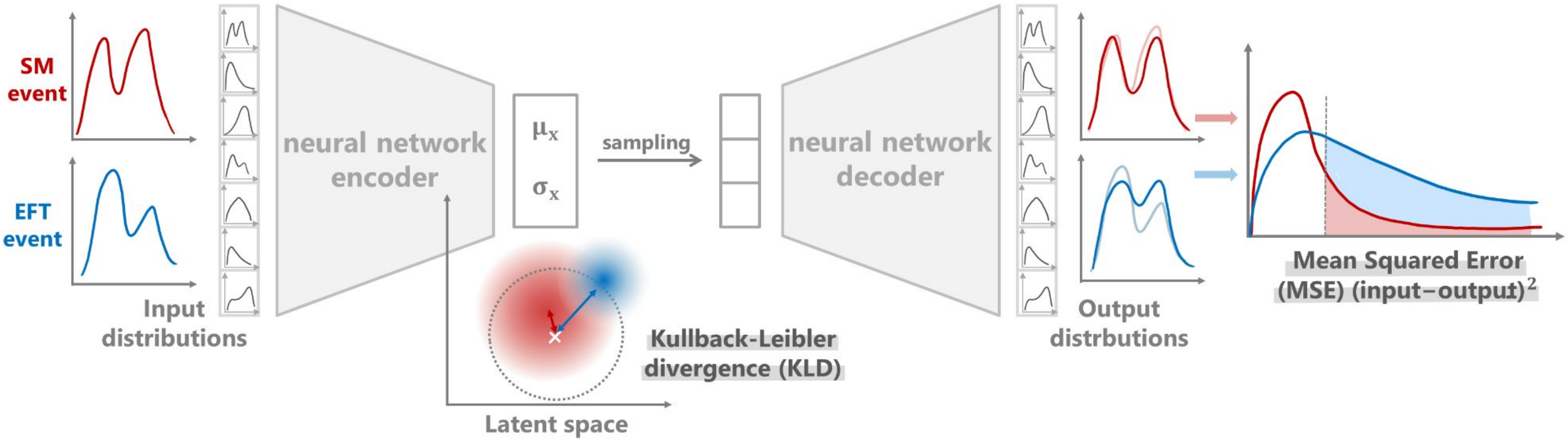
- The latent space is forced to be regular, namely described by a multidimensional **gaussian distribution**
  - via minimization of a **regularization loss (KLD)** + **reconstruction loss (MSE)**
- A point is sampled from the latent space and decoded



Variational AutoEncoders for Anomaly Detection

Generative model: it **learns to decode samples drawn from the same probability distribution of the original dataset**

→ **robust** and **variation-tolerant** Anomaly Detection strategy

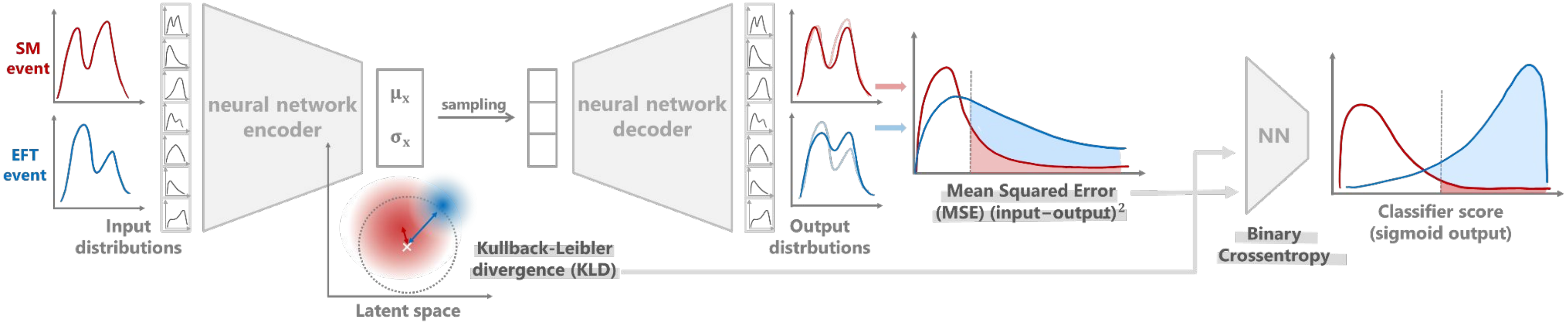


Adding a supervised NN classifier to the VAE

The VAE is only trained to reconstruct a SM sample, while our goal is to isolate EFT events.

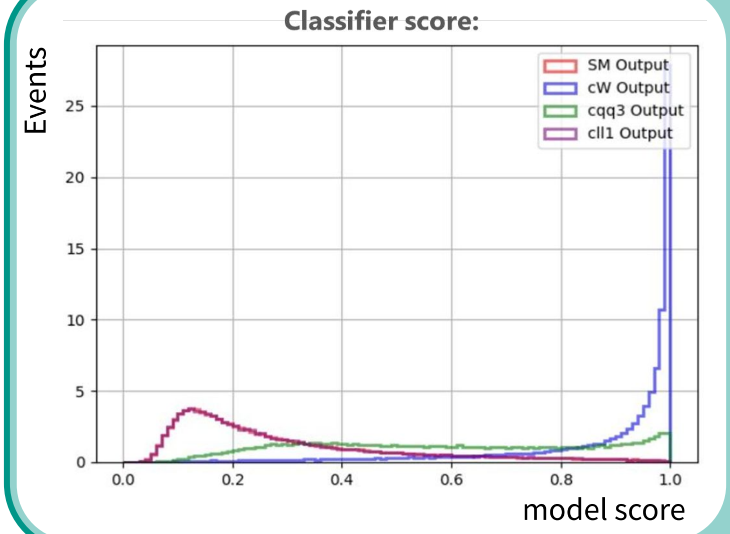
We want to **embed discrimination in the training** → **VAE + NN classifier**

- Trained by minimization of MSE + KLD + **Binary Crossentropy**
- Input data are divided between purely SM and SM + EFT:
  - MSE and KLD coming from **SM** events are added to the model loss
  - MSE and KLD coming from a set of **SM+EFT** events are given to the classifier
  - the binary crossentropy is added to the model loss

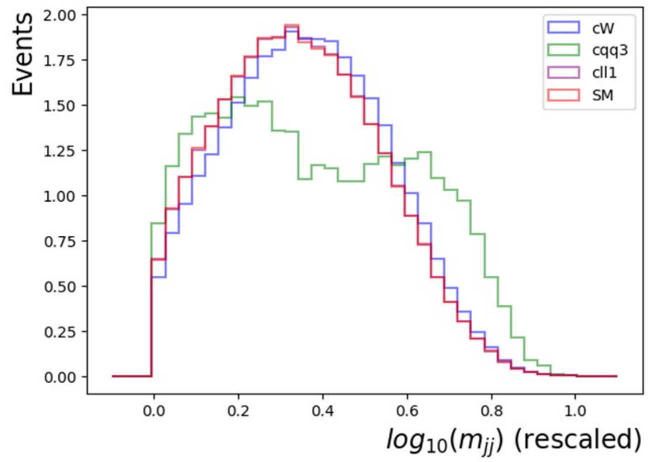


VAE results: shape comparison

The model collects information from various inputs → it provides a variable (**output score**) whose **shape maximizes the separation between EFT and SM**



(e.g. wrt a **simple kinematic variable**)



VAE results: a proxy metric of the significance

The model is sensitive to several different operators

- We define a proxy metric for the significance σ, which depends on the Wilson coefficients of the operator considered during testing:

$$\sigma(c_{op}) = \frac{|n_{BSM}(c_{op}) - n_{SM}|}{\sqrt{n_{SM}}} = \frac{|n_{LIN}(c_{op}) + n_{QUAD}(c_{op}^2)|}{\sqrt{n_{SM}}}$$

- We consider the model sensitive to an operator if σ reaches the value of 3:

operator	c_W	c_{qq}^{1}	$c_{qq}^{1,1}$	c_{qq}^{3}	$c_{qq}^{3,1}$	c_{Hq}^{1}	c_{HW}
$c_{op} : \sigma(c_{op}) = 3$	0.13	0.17	0.18	0.11	0.11	0.61	0.65

Conclusions and future perspectives

Outcome:

- The VAE allows for detecting many operators → promising!

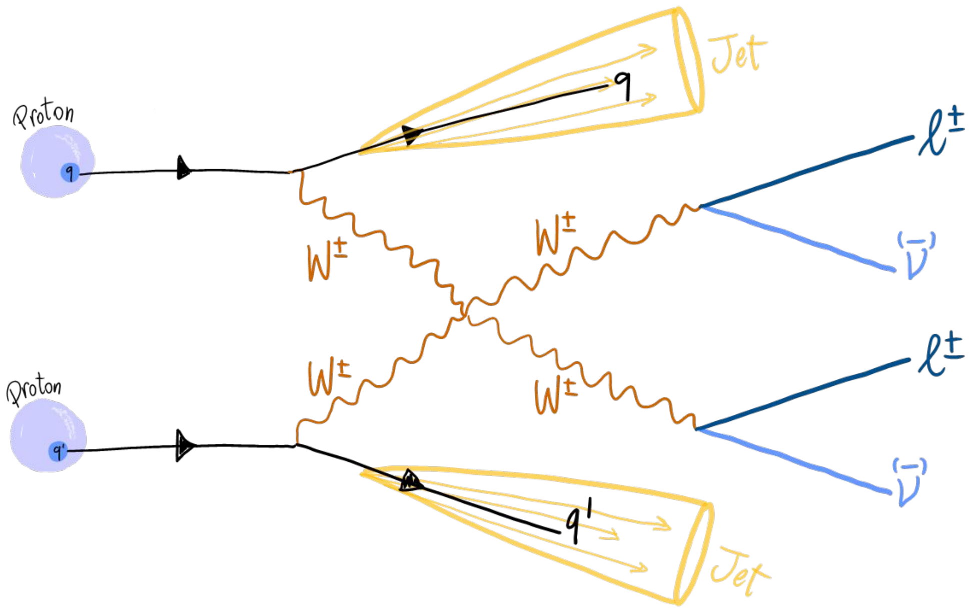
Further steps:

- Introduce estimation of backgrounds (fake leptons)
- Test VAE on fully reconstructed events

BACKUP

fully leptonic SSWW

Scattering of two same-sign W bosons, both decaying leptonically



- **Very clean signature** (very low QCD background):
  - 2 forward jets
  - 2 same sign charged leptons
  - 2 neutrinos
- **Main backgrounds:**
  - fake leptons
  - WZ (QCD and EWK)

Input variables

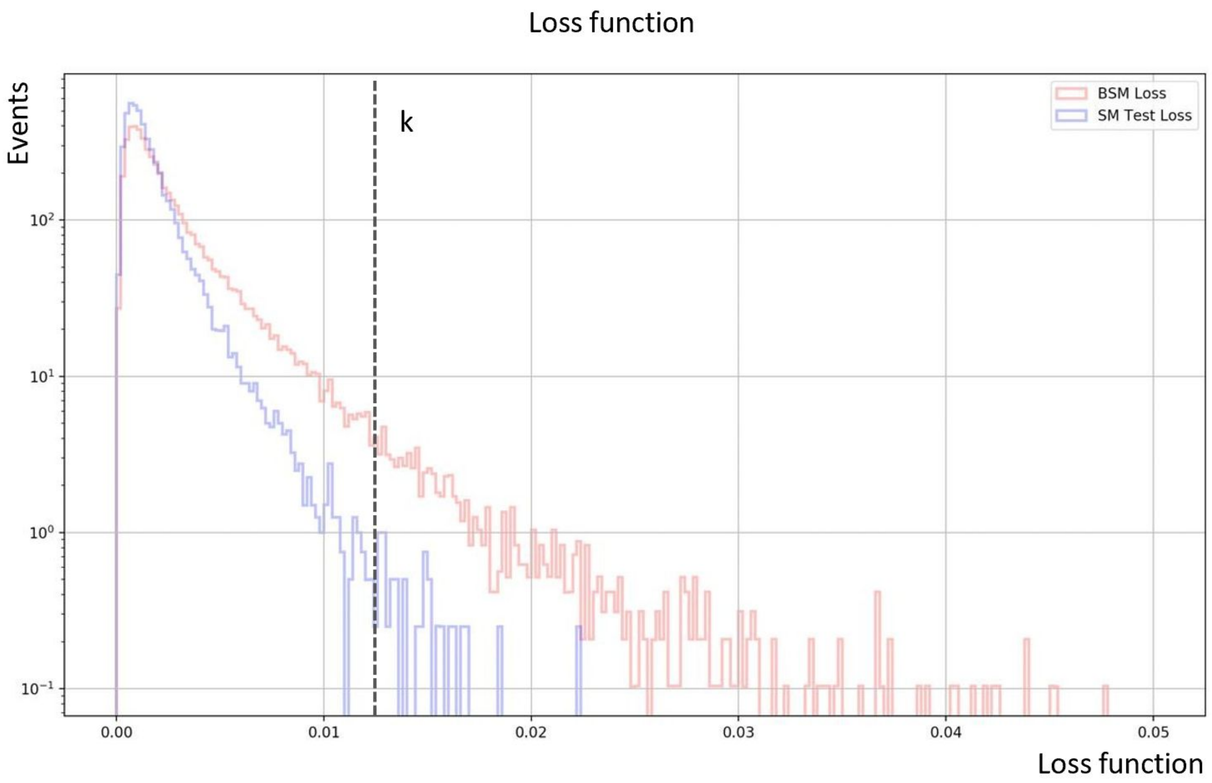
Variables considered:

- m_{ll}, m_{jj}
- $p_{t_{l1}}$, $p_{t_{l2}}$, $p_{t_{j1}}$, $p_{t_{j2}}$
- $p_{t_{ll}}$
- MET
- $\Delta\Phi_{jj}$
- η_{j_1}, η_{j_2}, η_{l_1}, η_{l_2}
- $\Delta\eta_{jj}$

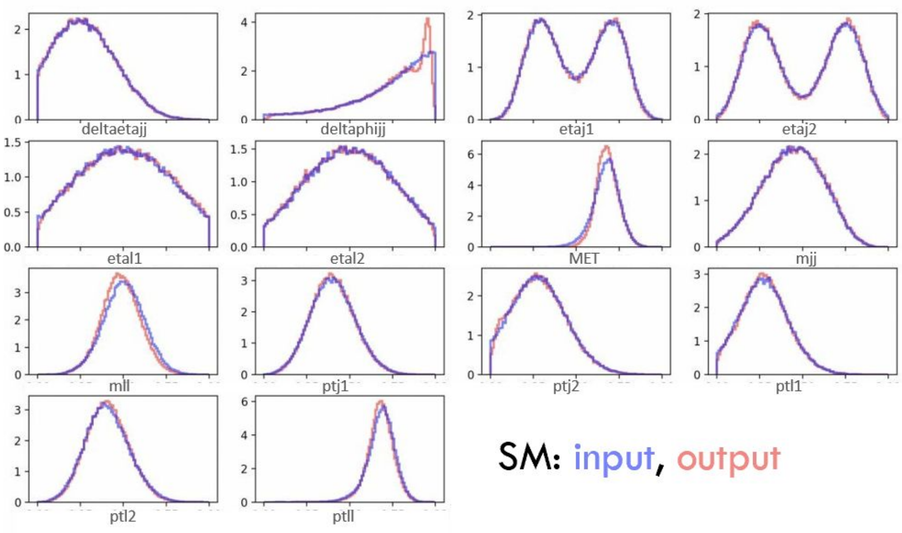
Selections:

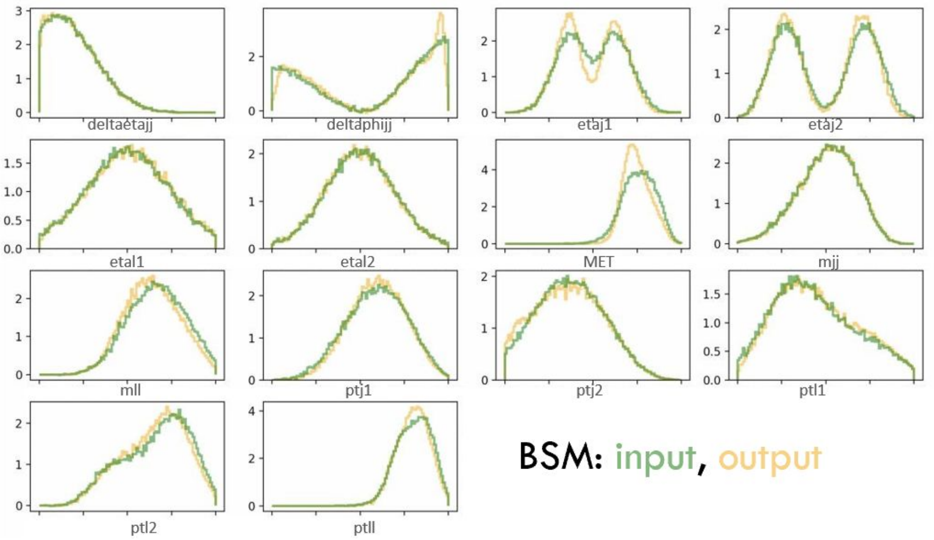
- $p_{t_{j1}}, p_{t_{j2}} > 30GeV$
- $m_{jj} > 200GeV$
- $\Delta\eta_{jj} > 2$

Loss function:



Samples reconstruction: BSM vs SM



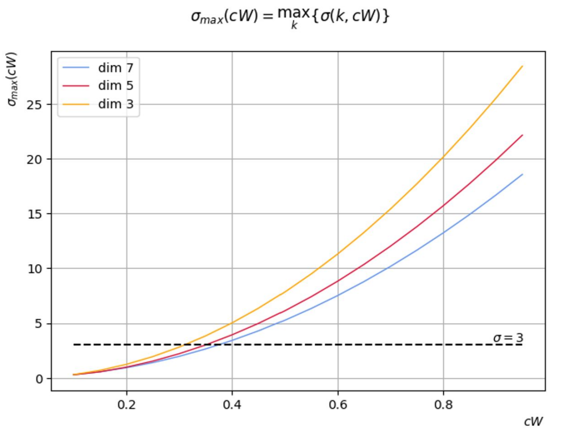


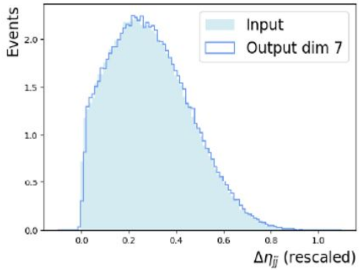
Why model optimization is fundamental

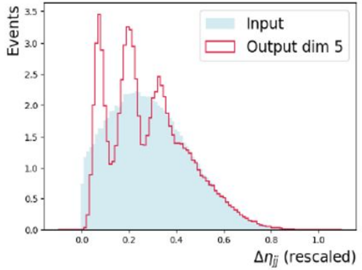
Our aim is to achieve a good discrimination between SM and BSM. However, **the VAE is only trained to produce a good reconstruction of the SM sample:**

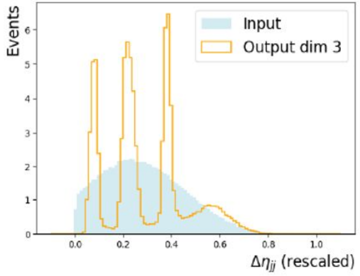
- The model learns the SM distributions
  - The more the model learns, the better the reconstruction of the SM
- The model extrapolates and is able to reconstruct BSM events

Example: **dimension of the latent** space. A bigger latent space allows the model to learn more features
→ better reconstruction of SM
→ better reconstruction of BSM! worse discrimination









Operators:

$Q_{Hl}^{(1)} = (H^\dagger i\overleftrightarrow{D}_\mu H)(\bar{l}_p\gamma^\mu l_p)$	$Q_{Hl}^{(3)} = (H^\dagger i\overleftrightarrow{D}_\mu^i H)(\bar{l}_p\sigma^i\gamma^\mu l_p)$
$Q_{Hq}^{(1)} = (H^\dagger i\overleftrightarrow{D}_\mu H)(\bar{q}_p\gamma^\mu q_p)$	$Q_{Hq}^{(3)} = (H^\dagger i\overleftrightarrow{D}_\mu^i H)(\bar{q}_p\sigma^i\gamma^\mu q_p)$
$Q_{qq}^{(1)} = (\bar{q}_p\gamma_\mu q_p)(\bar{q}_r\gamma^\mu q_r)$	$Q_{qq}^{(1,1)} = (\bar{q}_p\gamma_\mu q_r)(\bar{q}_r\gamma^\mu q_p)$
$Q_{qq}^{(3)} = (\bar{q}_p\gamma_\mu\sigma^i q_p)(\bar{q}_r\gamma^\mu\sigma^i q_r)$	$Q_{qq}^{(3,1)} = (\bar{q}_p\gamma_\mu\sigma^i q_r)(\bar{q}_r\gamma^\mu\sigma^i q_p)$
$Q_{HD} = (H^\dagger D_\mu H)(H^\dagger D^\mu H)$	$Q_{H\Box} = (H^\dagger H)\Box(H^\dagger H)$
$Q_{HWB} = (H^\dagger\sigma^i H)W_{\mu\nu}^i B^{\mu\nu}$	$Q_{HW} = (H^\dagger H)W_{\mu\nu}^i W^{i\mu\nu}$
$Q_W = \varepsilon^{ijk}W_\mu^{i\nu}W_\nu^{j\rho}W_\rho^{k\mu}$	$Q_{ll}^{(1)} = (\bar{l}_p\gamma_\mu l_r)(\bar{l}_r\gamma^\mu l_p)$

- H denotes the SU(2) Higgs doublet
- Wiμν and Bμν denote the gauge fields associate with SU(2) and U(1) symmetries respectively
- l,q denote the left-handed lepton and quark doublets
- u, d, e denote the right-handed quark and charged-lepton fields
- i, j, k denote the SU(2) indexes and sig_i the Pauli matrices

•Enter via modifications of the EW input quantities: $(Q_{Hl}^{(3)}, Q_{ll}^{(1)}, Q_{HD}, Q_{HWB})$

•Induce modifications via:

-Vff couplings $(Q_{Hl}^{(1)}, Q_{Hl}^{(3)}, Q_{Hq}^{(1)}, Q_{Hq}^{(3)})$

-Gauge couplings (Q_W)

-HVV couplings $(Q_{HD}, Q_{HW}, Q_{HWB}, Q_{H\Box})$

-Four-quark contact terms $(Q_{qq}^{(1)}, Q_{qq}^{(3)}, Q_{qq}^{(1,1)}, Q_{qq}^{(3,1)})$