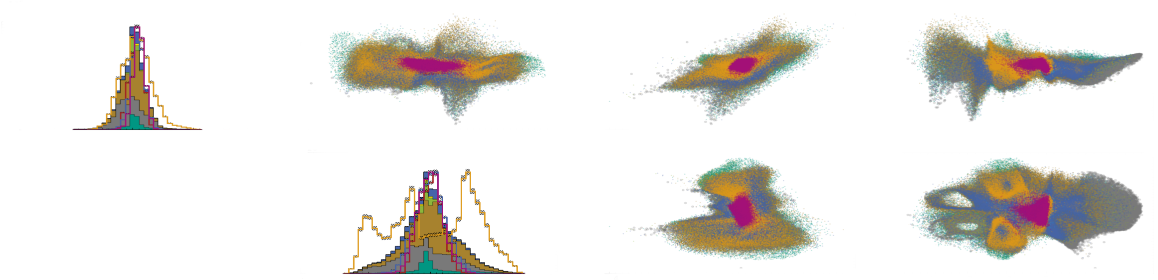
Karlsruhe Institute of Technology

Inelastic Dark Matter with Dark Higgs

Jonas Eppelt | November 14, 2022

ETP



Introduction

Anomaly Detection in general

- Something that is unlike any other: out-of-distribution
- In HEP typically over-densities

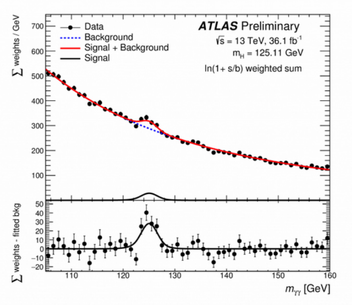


Figure: ATLAS collaboration, https://atlas.cern/updates/briefing/new-atlas-

Anomaly Detection in HEP

- physics case: Jet Identification at CMS or ATLAS
- prevalent methods: Classification without labels, Autoencoders

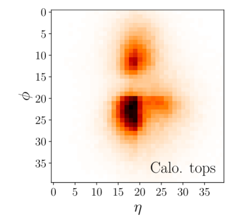
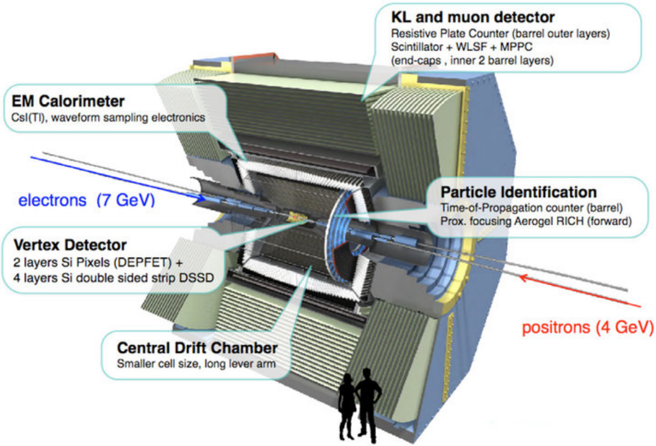
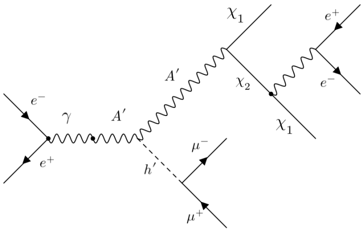


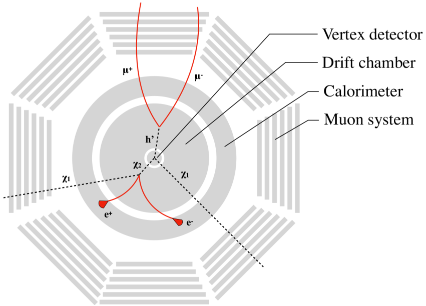
Figure: B. Dillon, T. Plehn, C. Sauer, P. Sorrenson, "Better Latent Spaces for Better Autoencoders", SciPost 11 no.3 (Sep, 2021)

Belle II - a e^+e^- collider experiment



Inelastic Dark Matter with Dark Higgs





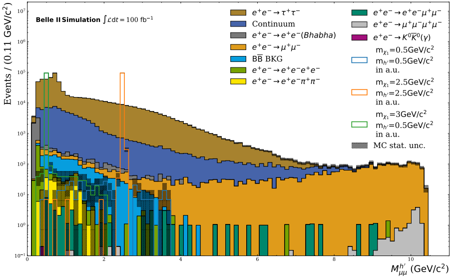
7 free parameters

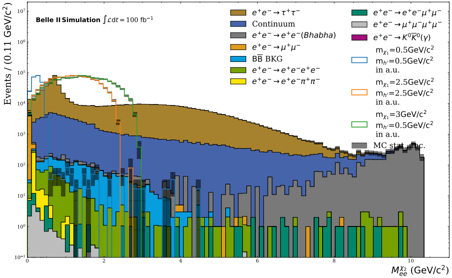
- Coupling constants for Dark Higgs h' to Dark Matter χ_1/χ_2 and Dark Photon A' to Dark Matter
- Mixing angles for the A' and h'
- Masses for the χ_1, h' and A'
- described in *Long-lived Dark Higgs and inelastic Dark Matter at Belle II* (DOI:2012.08595)

Motivating Anomaly Detection

Searches for Inelastic Dark Matter

- Non prompt decays have low background ($\longrightarrow$ search by Patrick Ecker)
- Prompt decays hard to identify
- Curse of dimensionality: number of points $\propto n^d$ $\longrightarrow$ Find a model-independent way to select signal





Signal Simulation

model parameter	values
m_{χ_1}	[0.25, 0.5, 1.0, 1.5, 2.0, 2.5, 3.0]
$m_{h'}$	[0.25, 0.5, 1.0, 1.5, 2.0, 2.5, 3.0]
θ	10^{-2}
ϵ	10^{-2}
$m_{A'}$	$4 \cdot m_{\chi_1}$
	2.746×10^{-1}
	1.12

example signals

- light dm: $m_{h'} = 0.5\,\mathrm{G}$, $m_{\chi_1} = 0.5\,\mathrm{G}$
- heavy dm: $m_{h'} = 2.5\,\mathrm{G}$, $m_{\chi_1} = 2.5\,\mathrm{G}$
- high mass splitting: $m_{h'} = 0.5\,\mathrm{G}$, $m_{\chi_1} = 3\,\mathrm{G}$

Reconstruction

Goals

- Reduce ´unphysical´ background sources (beam background)
- Reduce the amount of events to a ‘handleable‘ level

final state particles

- leptons (e/μ)
  - Trackinghits > 20
  - binaryPID (μ vs e)
    $e :> 0.1$
    $\mu :< 0.9$
  - θ in CDCAcceptance
  - nCDCHits > 20
  - dr < 0.5 cm
  - $|\text{dz}| < 2$ cm

reconstruct particles

$\mu^+\mu^- \rightarrow h' \,/\, e^+e^- \rightarrow \chi_2$

- dr $<= 0.2$ cm
- $\chi_{\text{Prob}} > 0$ & min one $\chi_{\text{Prob}} >= 0.01$

Rest of event (ROE)

- θ in CDCAcceptance
- nCDCHits > 20
- dr < 0.5 cm
- $|\text{dz}| < 2$ cm
- E > 0.05 GeV

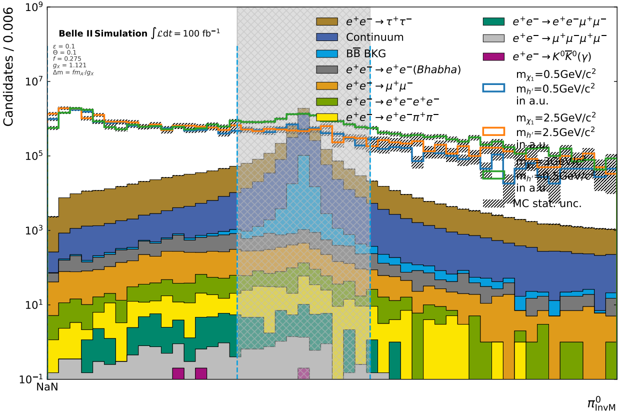
π_o veto

γ:

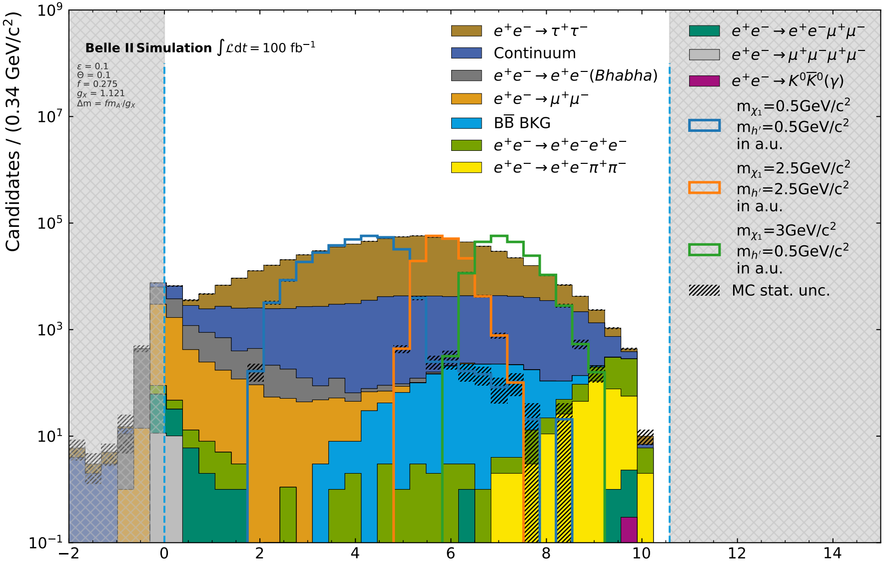
- n cl.Hits > 1.5
- $0.2967 < \theta_{\text{cl}} < 2.6180$
- E< 0.25 GeV ,bzw.
  E> 0.04 GeV
- $|\text{clusterTiming}| < 200$

$\gamma \rightarrow \pi^0$:

- $0\,\text{GeV} < \text{InvM} < 0.3\,\text{GeV}$
- candidate closest to π^0_{InvM}



Missing energy selection

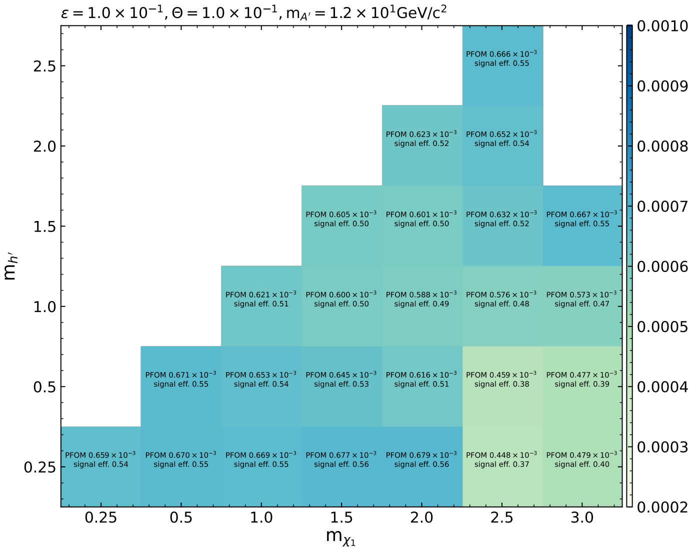


Sensitivities after Reconstruction

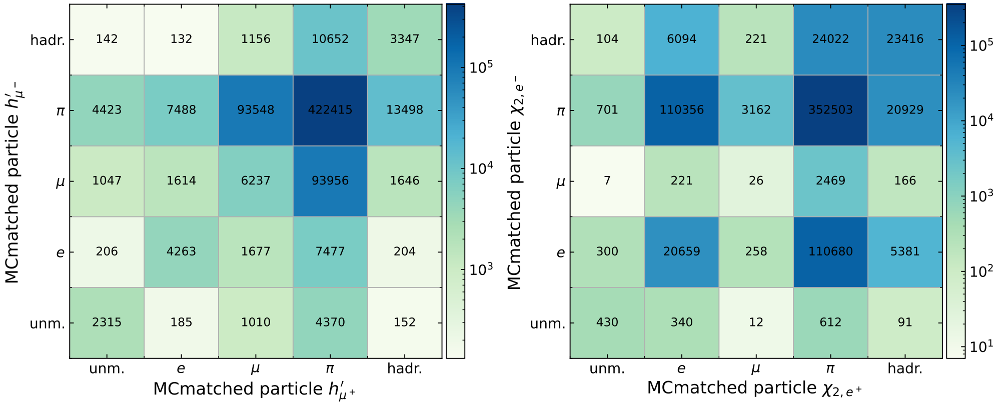
benchmark sensitivity: Punzi Figure of Merit (PFOM)

$$\text{PFOM} = \frac{eff_{\text{signal}}}{\frac{1}{2} + \sqrt{N_{\text{background, after}}}},$$

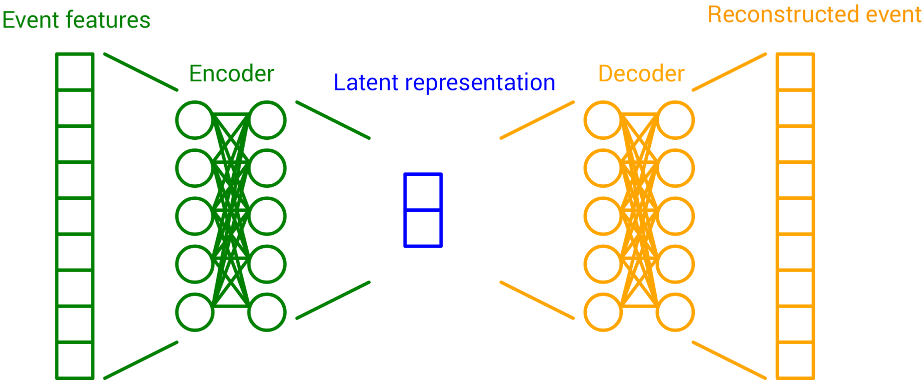
eff_{signal} is calculated w.r.t to generated samples (25000).



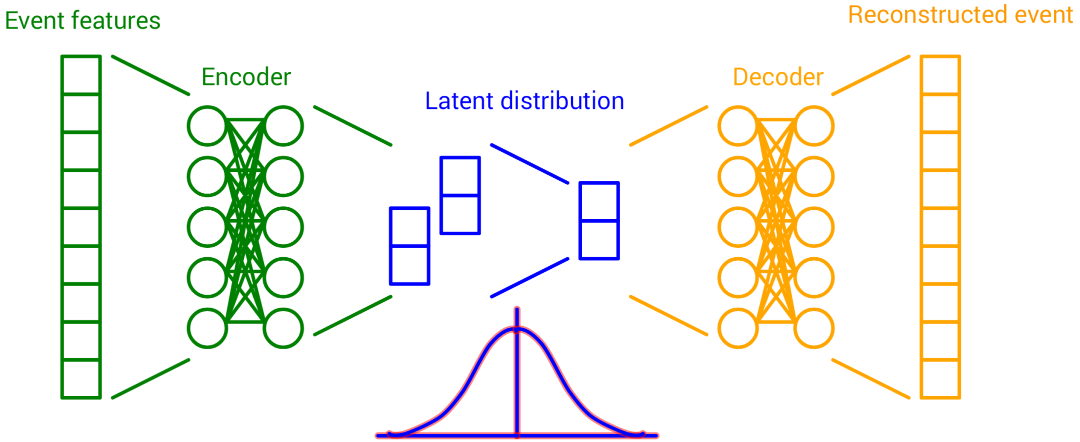
Truthmatching



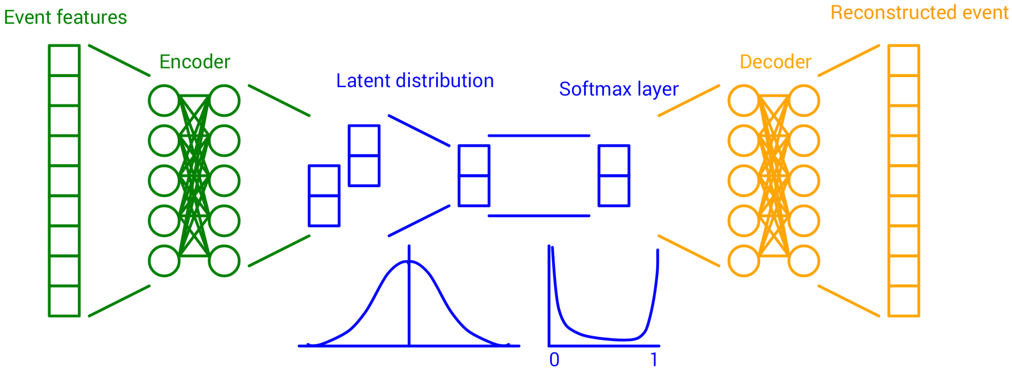
Autoencoder



Variational Autoencoder



Dirichlet Variational Autoencoder



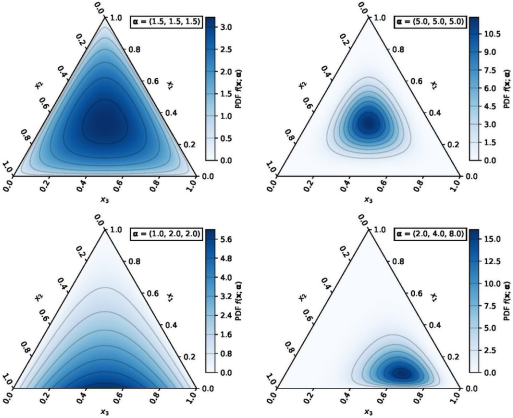
About Dirichlet

basics

- probability for categorical distributions, e.g.:
  - 30% 'anomalous' $\rightarrow$ 70% 'normal'
  - 10% 'Bhabha', 80% '$\tau\tau$, 10% 'anomalous'
- with k categories k parameters α_k

approximation

- $D_\alpha(r) \approx \text{softmax}[N(z, \mu, \sigma)]$ with
  - $\mu_i = log\alpha_i - \frac{1}{R}\sum_i^d \log \alpha_i$
  - $\sigma_i = \frac{1}{\alpha_i}(1 - \frac{2}{R}) + \frac{1}{R^2}\sum_i^d \frac{1}{\alpha_i}$
- add a softmax layer after resampling
- distribution characterized by d parameter α_i



Prescaling

Karlsruhe Institute of Technology

Input features

- final state leptons 4-vectors (E, px, py, pz)
- order of (e^-, e^+, μ^-, μ^+)
- missing Energy as 4-vector
- prescale by standardize:

$$x' = \frac{x - \hat{x}}{\sigma(x)} \tag{1}$$

Hyperparameters

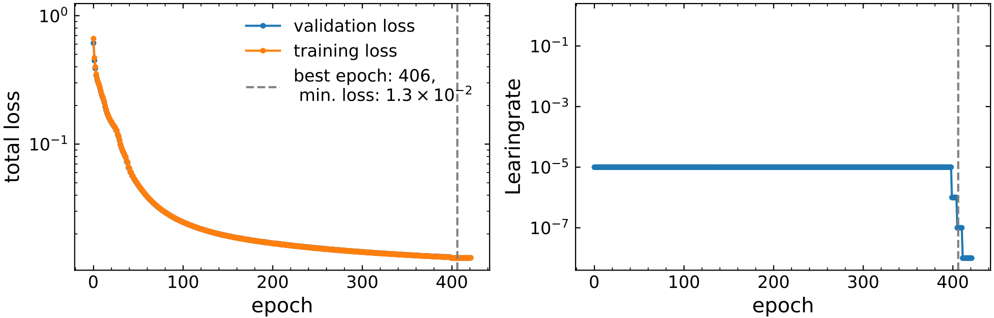
Architecture

- both parts 3 hidden layers @ 100 neurons
- 1 to 10 latent dimensions
- ReLU activation

Training

- Adam with learning rate 10^{-5}
- early stopping after 15 epochs without new best epoch (max 500)
- Cross-validation with 20 % split
- Stochastic Weighted Average
- Batch size 128
- LR scheduling on plateau (factor 0.1 10 epochs, relative change of 10^{-4})

Training AE(l=10)



Trainings results

Karlsruhe Institute of Technology

AE

lat. dim	h	mse
1	1.6	0.7
2	17	0.3
3	22	0.22
4	26	0.14
5	23	0.1
6	27	0.08
7	28	0.06
8	25	0.04
9	21	0.028
10	27	0.013

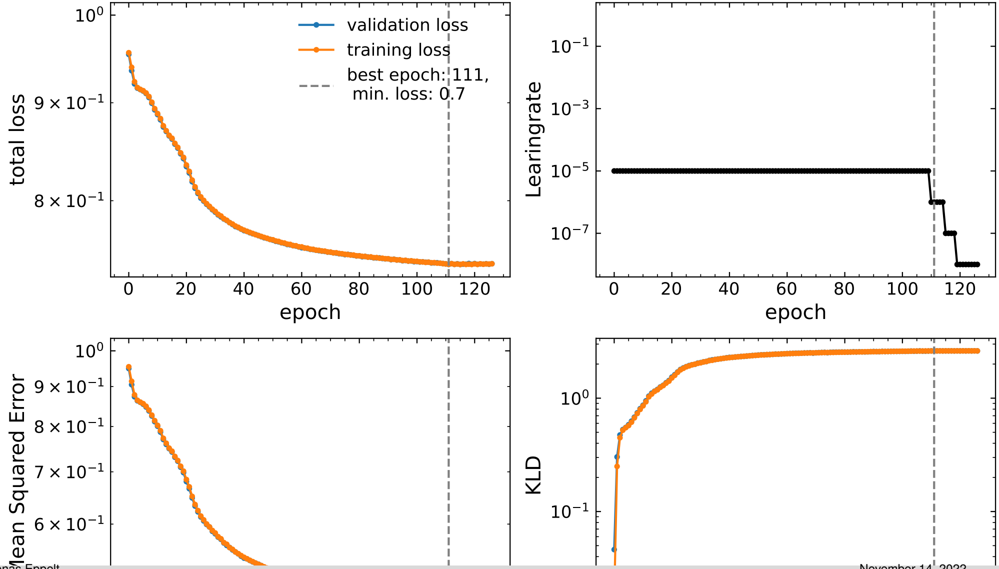
VAE

lat. dim	h	mse
1	2.1	0.9
2	8	0.8
3	9	0.7
4	7	0.7
5	9	0.7
6	7	0.7
7	9	0.7
8	8	0.7
9	8	0.8
10	8	0.7

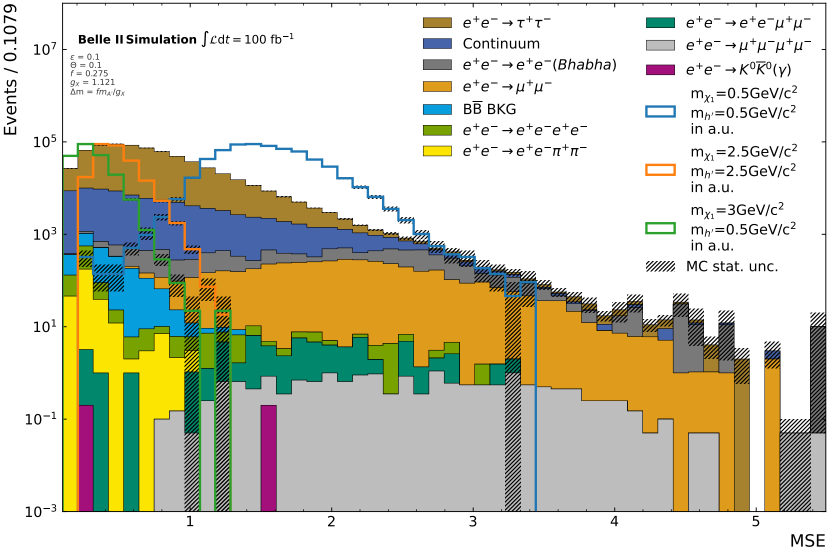
DVAE

lat. dim	h	mse
2	8	0.7
3	18	0.4
4	21	0.29
5	20	0.28
6	30	0.2
7	30	0.15
8	30	0.12
9	28	0.09
10	40	0.08

Training VAE(l=5)



MSE of AE(l=1)



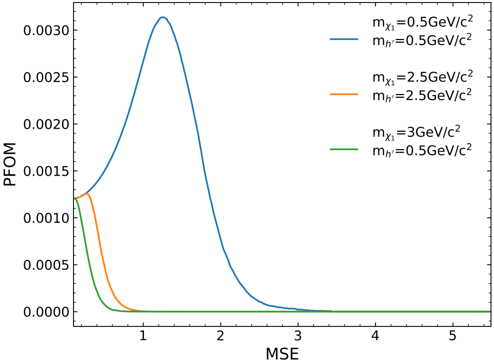
Belle II Simulation $\int \mathcal{L}dt = 100\ \text{fb}^{-1}$
ε = 0.1
Θ = 0.1
f = 0.275
g_X = 1.121
$\Delta m = f m_{A'}/g_X$
$e^+e^- \to \tau^+\tau^-$
Continuum
$e^+e^- \to e^+e^-(Bhabha)$
$e^+e^- \to \mu^+\mu^-$
$B\overline{B}$ BKG
$e^+e^- \to e^+e^-e^+e^-$
$e^+e^- \to e^+e^-\pi^+\pi^-$
$e^+e^- \to e^+e^-\mu^+\mu^-$
$e^+e^- \to \mu^+\mu^-\mu^+\mu^-$
$e^+e^- \to K^0\overline{K}^0(\gamma)$
m_{χ_1}=0.5GeV/c^2 $m_{h'}$=0.5GeV/c^2 in a.u.
m_{χ_1}=2.5GeV/c^2 $m_{h'}$=2.5GeV/c^2 in a.u.
m_{χ_1}=3GeV/c^2 $m_{h'}$=0.5GeV/c^2 in a.u.
MC stat. unc.
Events / 0.1079
MSE

Punzi curves

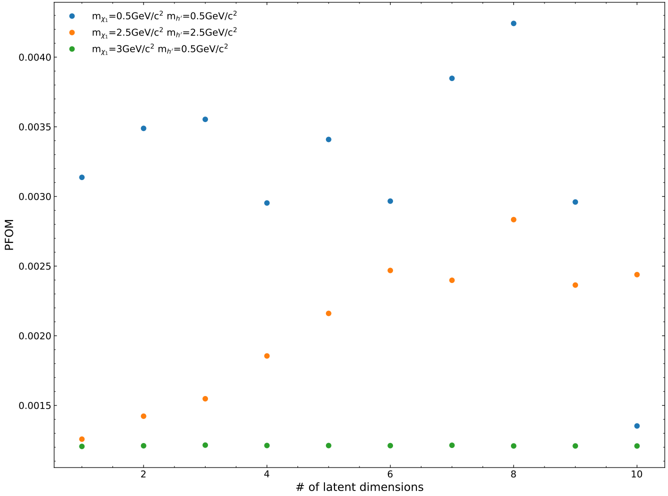
benchmark sensitivity: Punzi Figure of Merit (PFOM)

$$\text{PFOM} = \frac{eff_{\text{signal}}}{\frac{1}{2} + \sqrt{N_{\text{background, after}}}},$$

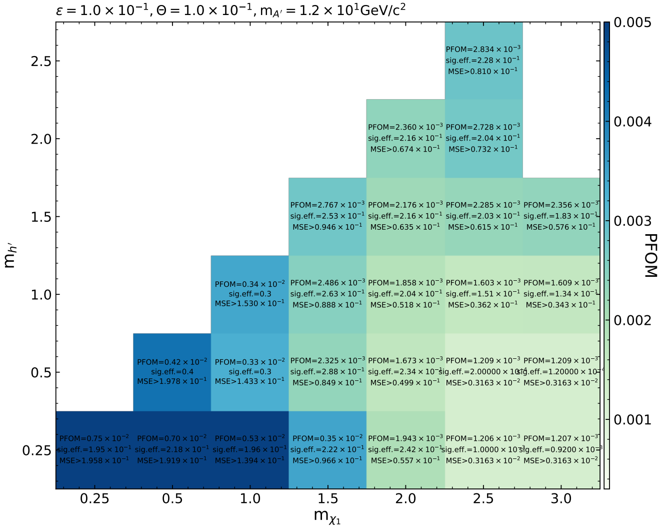
eff_{signal} relative to selected samples



Punzi comparison AE

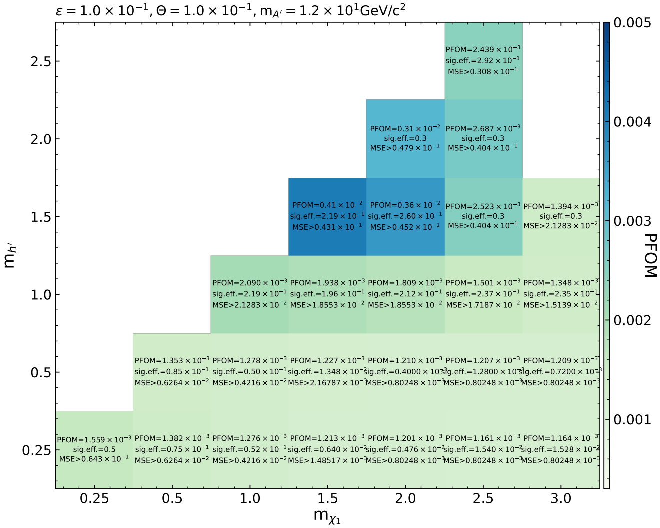


AE(l=8)



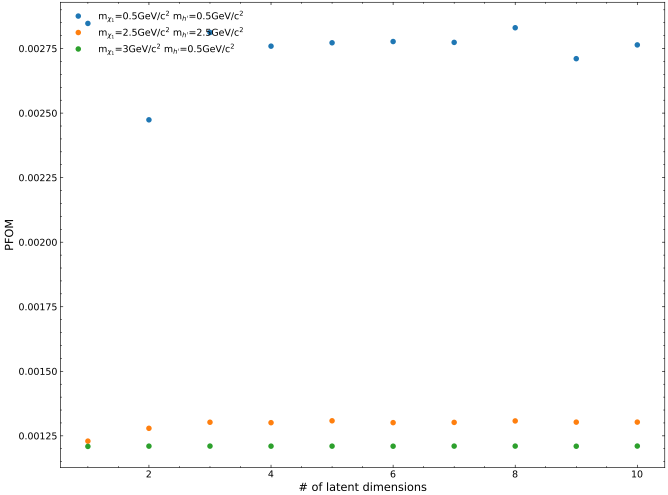
$\varepsilon = 1.0 \times 10^{-1}, \Theta = 1.0 \times 10^{-1}, m_{A'} = 1.2 \times 10^{1} \mathrm{GeV/c^2}$
$m_{h'}$
m_{χ_1}
PFOM
0.25
0.5
1.0
1.5
2.0
2.5
3.0
0.001
0.002
0.003
0.004
0.005
PFOM=2.834 × 10⁻³ sig.eff.=2.28 × 10⁻¹ MSE>0.810 × 10⁻¹
PFOM=2.360 × 10⁻³ sig.eff.=2.16 × 10⁻¹ MSE>0.674 × 10⁻¹
PFOM=2.728 × 10⁻³ sig.eff.=2.04 × 10⁻¹ MSE>0.732 × 10⁻¹
PFOM=2.767 × 10⁻³ sig.eff.=2.53 × 10⁻¹ MSE>0.946 × 10⁻¹
PFOM=2.176 × 10⁻³ sig.eff.=2.16 × 10⁻¹ MSE>0.635 × 10⁻¹
PFOM=2.285 × 10⁻³ sig.eff.=2.03 × 10⁻¹ MSE>0.615 × 10⁻¹
PFOM=2.356 × 10⁻³ sig.eff.=1.83 × 10⁻¹ MSE>0.576 × 10⁻¹
PFOM=0.34 × 10⁻² sig.eff.=0.3 MSE>1.530 × 10⁻¹
PFOM=2.486 × 10⁻³ sig.eff.=2.63 × 10⁻¹ MSE>0.888 × 10⁻¹
PFOM=1.858 × 10⁻³ sig.eff.=2.04 × 10⁻¹ MSE>0.518 × 10⁻¹
PFOM=1.603 × 10⁻³ sig.eff.=1.51 × 10⁻¹ MSE>0.362 × 10⁻¹
PFOM=1.609 × 10⁻³ sig.eff.=1.34 × 10⁻¹ MSE>0.343 × 10⁻¹
PFOM=0.42 × 10⁻² sig.eff.=0.4 MSE>1.978 × 10⁻¹
PFOM=0.33 × 10⁻² sig.eff.=0.3 MSE>1.433 × 10⁻¹
PFOM=2.325 × 10⁻³ sig.eff.=2.88 × 10⁻¹ MSE>0.849 × 10⁻¹
PFOM=1.673 × 10⁻³ sig.eff.=2.34 × 10⁻¹ MSE>0.499 × 10⁻¹
PFOM=1.209 × 10⁻³ sig.eff.=2.00000 × 10⁻⁴ MSE>0.3163 × 10⁻²
PFOM=1.209 × 10⁻³ sig.eff.=1.20000 × 10⁻⁴ MSE>0.3163 × 10⁻²
PFOM=0.75 × 10⁻² sig.eff.=1.95 × 10⁻¹ MSE>1.958 × 10⁻¹
PFOM=0.70 × 10⁻² sig.eff.=2.18 × 10⁻¹ MSE>1.919 × 10⁻¹
PFOM=0.53 × 10⁻² sig.eff.=1.96 × 10⁻¹ MSE>1.394 × 10⁻¹
PFOM=0.35 × 10⁻² sig.eff.=2.22 × 10⁻¹ MSE>0.966 × 10⁻¹
PFOM=1.943 × 10⁻³ sig.eff.=2.42 × 10⁻¹ MSE>0.557 × 10⁻¹
PFOM=1.206 × 10⁻³ sig.eff.=1.0000 × 10⁻³ MSE>0.3163 × 10⁻²
PFOM=1.207 × 10⁻³ sig.eff.=0.9200 × 10⁻³ MSE>0.3163 × 10⁻²

AE(l=10)

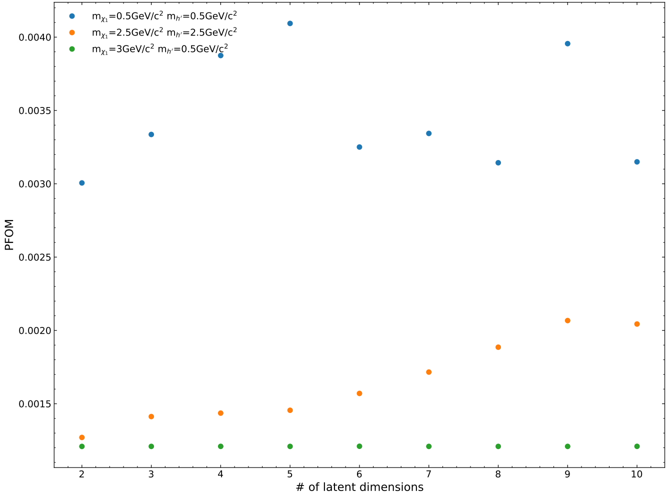


$\varepsilon = 1.0 \times 10^{-1}, \Theta = 1.0 \times 10^{-1}, m_{A'} = 1.2 \times 10^{1} GeV/c^2$
PFOM=2.439×10^{-3} sig.eff.=2.92×10^{-1} MSE>0.308×10^{-1}
PFOM=0.31×10^{-2} sig.eff.=0.3 MSE>0.479×10^{-1}
PFOM=2.687×10^{-3} sig.eff.=0.3 MSE>0.404×10^{-1}
PFOM=0.41×10^{-2} sig.eff.=2.19×10^{-1} MSE>0.431×10^{-1}
PFOM=0.36×10^{-2} sig.eff.=2.60×10^{-1} MSE>0.452×10^{-1}
PFOM=2.523×10^{-3} sig.eff.=0.3 MSE>0.404×10^{-1}
PFOM=1.394×10^{-3} sig.eff.=0.3 MSE>2.1283×10^{-2}
PFOM=2.090×10^{-3} sig.eff.=2.19×10^{-1} MSE>2.1283×10^{-2}
PFOM=1.938×10^{-3} sig.eff.=1.96×10^{-1} MSE>1.8553×10^{-2}
PFOM=1.809×10^{-3} sig.eff.=2.12×10^{-1} MSE>1.8553×10^{-2}
PFOM=1.501×10^{-3} sig.eff.=2.37×10^{-1} MSE>1.7187×10^{-2}
PFOM=1.348×10^{-3} sig.eff.=2.35×10^{-1} MSE>1.5139×10^{-2}
PFOM=1.353×10^{-3} sig.eff.=0.85×10^{-1} MSE>0.6264×10^{-2}
PFOM=1.278×10^{-3} sig.eff.=0.50×10^{-1} MSE>0.4216×10^{-2}
PFOM=1.227×10^{-3} sig.eff.=1.348×10^{-2} MSE>2.16787×10^{-3}
PFOM=1.210×10^{-3} sig.eff.=0.4000×10^{-3} MSE>0.80248×10^{-3}
PFOM=1.207×10^{-3} sig.eff.=1.2800×10^{-3} MSE>0.80248×10^{-3}
PFOM=1.209×10^{-3} sig.eff.=0.7200×10^{-3} MSE>0.80248×10^{-3}
PFOM=1.559×10^{-3} sig.eff.=0.5 MSE>0.643×10^{-1}
PFOM=1.382×10^{-3} sig.eff.=0.75×10^{-1} MSE>0.6264×10^{-2}
PFOM=1.276×10^{-3} sig.eff.=0.52×10^{-1} MSE>0.4216×10^{-2}
PFOM=1.213×10^{-3} sig.eff.=0.640×10^{-2} MSE>1.48517×10^{-3}
PFOM=1.201×10^{-3} sig.eff.=0.476×10^{-2} MSE>0.80248×10^{-3}
PFOM=1.161×10^{-3} sig.eff.=1.540×10^{-2} MSE>0.80248×10^{-3}
PFOM=1.164×10^{-3} sig.eff.=1.528×10^{-2} MSE>0.80248×10^{-3}
$m_{h'}$
2.5
2.0
1.5
1.0
0.5
0.25
0.25 0.5 1.0 1.5 2.0 2.5 3.0
m_{χ_1}
0.005
0.004
0.003
0.002
0.001
PFOM

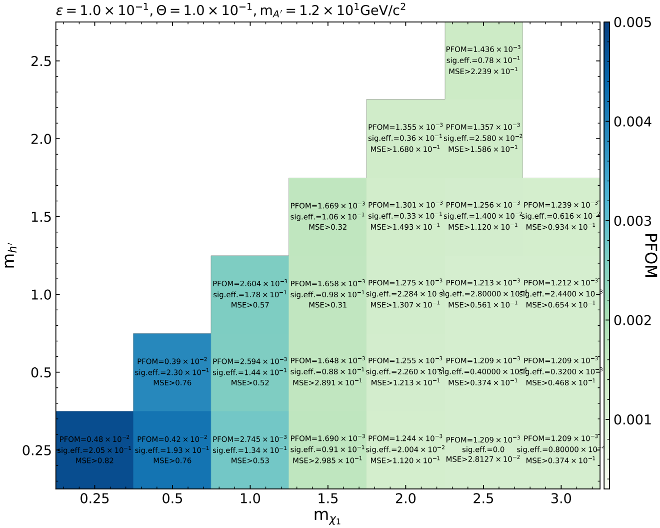
Punzi comparison VAE



Punzi comparison DVAE

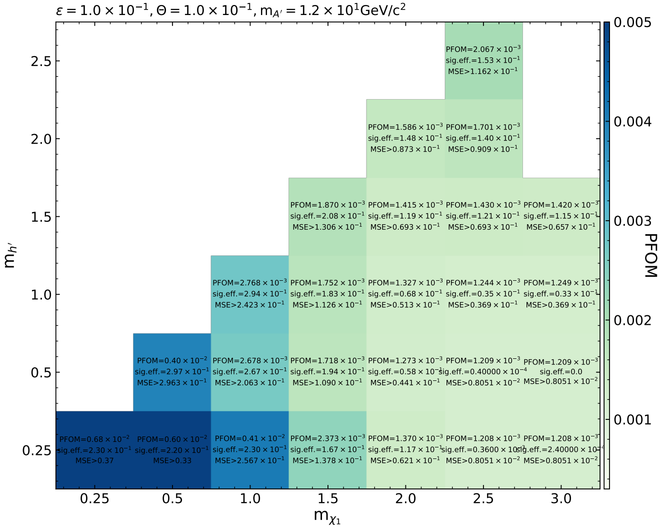


DVAE(l=4)



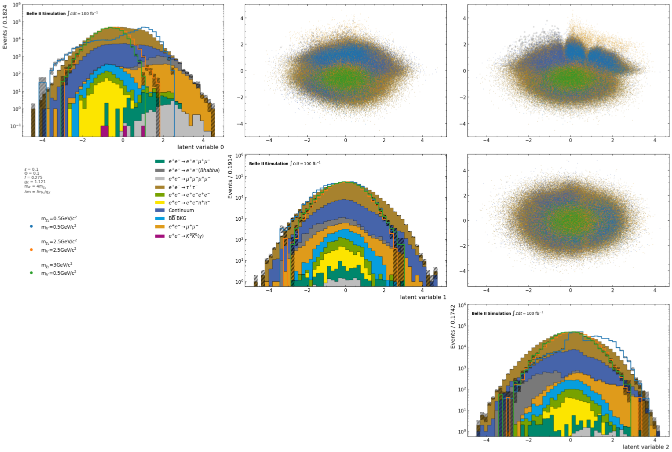
$\varepsilon = 1.0 \times 10^{-1}, \Theta = 1.0 \times 10^{-1}, m_{A'} = 1.2 \times 10^{1} GeV/c^2$
PFOM=1.436×10^{-3}
sig.eff.=0.78×10^{-1}
MSE>2.239×10^{-1}
PFOM=1.355×10^{-3}
sig.eff.=0.36×10^{-1}
MSE>1.680×10^{-1}
PFOM=1.357×10^{-3}
sig.eff.=2.580×10^{-2}
MSE>1.586×10^{-1}
PFOM=1.669×10^{-3}
sig.eff.=1.06×10^{-1}
MSE>0.32
PFOM=1.301×10^{-3}
sig.eff.=0.33×10^{-1}
MSE>1.493×10^{-1}
PFOM=1.256×10^{-3}
sig.eff.=1.400×10^{-2}
MSE>1.120×10^{-1}
PFOM=1.239×10^{-3}
sig.eff.=0.616×10^{-2}
MSE>0.934×10^{-1}
PFOM=2.604×10^{-3}
sig.eff.=1.78×10^{-1}
MSE>0.57
PFOM=1.658×10^{-3}
sig.eff.=0.98×10^{-1}
MSE>0.31
PFOM=1.275×10^{-3}
sig.eff.=2.284×10^{-2}
MSE>1.307×10^{-1}
PFOM=1.213×10^{-3}
sig.eff.=2.80000×10^{-4}
MSE>0.561×10^{-1}
PFOM=1.212×10^{-3}
sig.eff.=2.4400×10^{-3}
MSE>0.654×10^{-1}
PFOM=0.39×10^{-2}
sig.eff.=2.30×10^{-1}
MSE>0.76
PFOM=2.594×10^{-3}
sig.eff.=1.44×10^{-1}
MSE>0.52
PFOM=1.648×10^{-3}
sig.eff.=0.88×10^{-1}
MSE>2.891×10^{-1}
PFOM=1.255×10^{-3}
sig.eff.=2.260×10^{-2}
MSE>1.213×10^{-1}
PFOM=1.209×10^{-3}
sig.eff.=0.40000×10^{-4}
MSE>0.374×10^{-1}
PFOM=1.209×10^{-3}
sig.eff.=0.3200×10^{-3}
MSE>0.468×10^{-1}
PFOM=0.48×10^{-2}
sig.eff.=2.05×10^{-1}
MSE>0.82
PFOM=0.42×10^{-2}
sig.eff.=1.93×10^{-1}
MSE>0.76
PFOM=2.745×10^{-3}
sig.eff.=1.34×10^{-1}
MSE>0.53
PFOM=1.690×10^{-3}
sig.eff.=0.91×10^{-1}
MSE>2.985×10^{-1}
PFOM=1.244×10^{-3}
sig.eff.=2.004×10^{-2}
MSE>1.120×10^{-1}
PFOM=1.209×10^{-3}
sig.eff.=0.0
MSE>2.8127×10^{-2}
PFOM=1.209×10^{-3}
sig.eff.=0.80000×10^{-4}
MSE>0.374×10^{-1}
$m_{h'}$
2.5
2.0
1.5
1.0
0.5
0.25
0.25
0.5
1.0
1.5
2.0
2.5
3.0
m_{χ_1}
0.005
0.004
0.003
0.002
0.001
PFOM

DVAE(l=9)

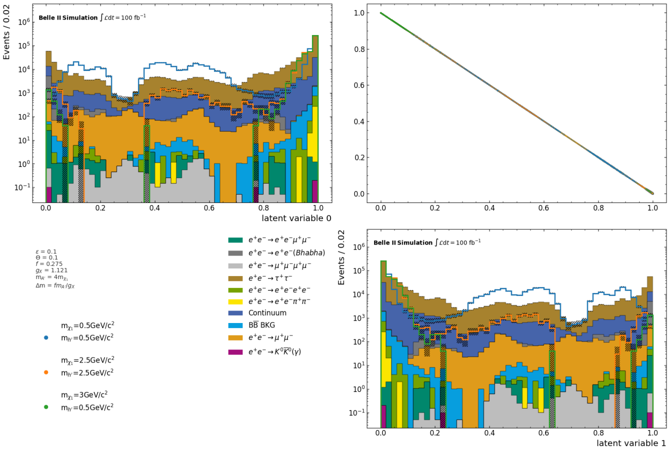


ε = 1.0 × 10^−1, Θ = 1.0 × 10^−1, m_A′ = 1.2 × 10^1 GeV/c^2
PFOM=2.067 × 10^−3 sig.eff.=1.53 × 10^−1 MSE>1.162 × 10^−1
PFOM=1.586 × 10^−3 sig.eff.=1.48 × 10^−1 MSE>0.873 × 10^−1
PFOM=1.701 × 10^−3 sig.eff.=1.40 × 10^−1 MSE>0.909 × 10^−1
PFOM=1.870 × 10^−3 sig.eff.=2.08 × 10^−1 MSE>1.306 × 10^−1
PFOM=1.415 × 10^−3 sig.eff.=1.19 × 10^−1 MSE>0.693 × 10^−1
PFOM=1.430 × 10^−3 sig.eff.=1.21 × 10^−1 MSE>0.693 × 10^−1
PFOM=1.420 × 10^−3 sig.eff.=1.15 × 10^−1 MSE>0.657 × 10^−1
PFOM=2.768 × 10^−3 sig.eff.=2.94 × 10^−1 MSE>2.423 × 10^−1
PFOM=1.752 × 10^−3 sig.eff.=1.83 × 10^−1 MSE>1.126 × 10^−1
PFOM=1.327 × 10^−3 sig.eff.=0.68 × 10^−1 MSE>0.513 × 10^−1
PFOM=1.244 × 10^−3 sig.eff.=0.35 × 10^−1 MSE>0.369 × 10^−1
PFOM=1.249 × 10^−3 sig.eff.=0.33 × 10^−1 MSE>0.369 × 10^−1
PFOM=0.40 × 10^−2 sig.eff.=2.97 × 10^−1 MSE>2.963 × 10^−1
PFOM=2.678 × 10^−3 sig.eff.=2.67 × 10^−1 MSE>2.063 × 10^−1
PFOM=1.718 × 10^−3 sig.eff.=1.94 × 10^−1 MSE>1.090 × 10^−1
PFOM=1.273 × 10^−3 sig.eff.=0.58 × 10^−1 MSE>0.441 × 10^−1
PFOM=1.209 × 10^−3 sig.eff.=0.40000 × 10^−4 MSE>0.8051 × 10^−2
PFOM=1.209 × 10^−3 sig.eff.=0.0 MSE>0.8051 × 10^−2
PFOM=0.68 × 10^−2 sig.eff.=2.30 × 10^−1 MSE>0.37
PFOM=0.60 × 10^−2 sig.eff.=2.20 × 10^−1 MSE>0.33
PFOM=0.41 × 10^−2 sig.eff.=2.30 × 10^−1 MSE>2.567 × 10^−1
PFOM=2.373 × 10^−3 sig.eff.=1.67 × 10^−1 MSE>1.378 × 10^−1
PFOM=1.370 × 10^−3 sig.eff.=1.17 × 10^−1 MSE>0.621 × 10^−1
PFOM=1.208 × 10^−3 sig.eff.=0.3600 × 10^−3 MSE>0.8051 × 10^−2
PFOM=1.208 × 10^−3 sig.eff.=2.40000 × 10^−4 MSE>0.8051 × 10^−2
m_h′
2.5
2.0
1.5
1.0
0.5
0.25
0.25
0.5
1.0
1.5
2.0
2.5
3.0
m_χ1
0.005
0.004
0.003
0.002
0.001
PFOM

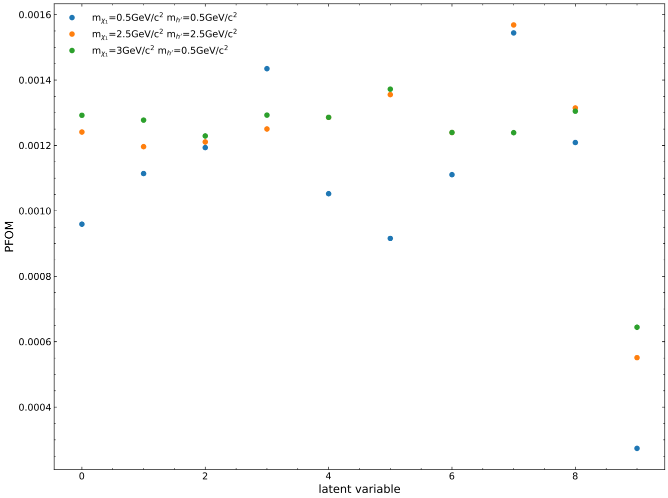
Latent space VAE(l=3)



Latent space DVAE(l=2)



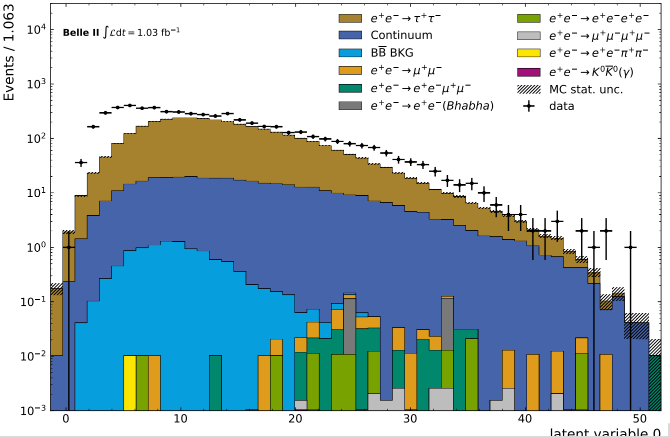
Punzi comparison DVAE(l=10) latent space



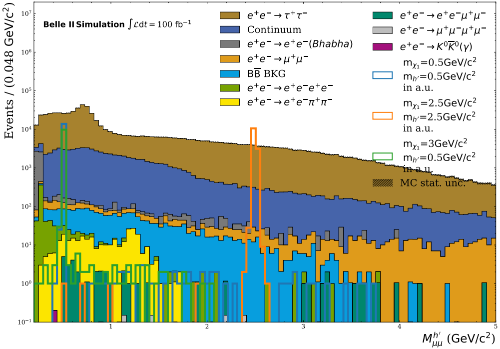
Validation

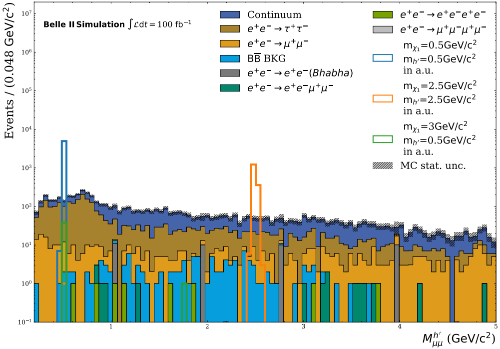
Look in data

- Idea: validate that latent-dimension looks similar for data
- Selecting three track Trigger and HLT



Selecting Anomalies using AE(l=8), MSE > 0.1978





Anomaly Detection for Belle II

Figure: courtesy of L. Reuter

Anomaly Detection for Inelastic Dark Matter

- Differently sized latent spaces are sensitive to different model parameter configurations
- Trade-off between model independence and sensitivities
- only little use of PID information was made

Use and Future of Autoencoders

- AEs outperform VAEs & DVAEs
- Further tuning of Hyperparameters possible
- DVAEs priors for the latent space require more investigation
- architectures could be further improved (Normalized Autoencoder, Graph Neural Networks)