(Variational) Auto-Encoders for New Physics Mining at the LHC

Olmo Cerri
Model-independent tagger for unexpected events

Flag non-SM events, despite their nature or particular features

1. Auto-Encoders & Anomaly Detection
2. Online trigger
3. Offline analysis
Physics anomaly detection

- Data mining concept
- Based on Variational Auto-Encoders [1]

1. Define what is “standard” through a set of example events
   - The Standard Model

2. Fit a function which gives the probability of belonging to the standard set
   - No assumption on the anomaly

3. Use this function to tag new events
   - Anomaly: low probability of belonging to the standard set
   - SM rare region or BSM

[1]: https://arxiv.org/abs/1312.6114
Auto-encoders in one slide

- **Map an input onto itself** passing through a latent representation
- **Unsupervised algorithm**, used for data compression, generation, clustering, etc.
- **Anomaly**: any event whose output is “far” from the input
AE anomaly detection: Training

Training:
Fit the VAE params to minimize the input-output distance
AE anomaly detection: Inference

Evaluate:
One-side hypothesis test on the input-output distance

Observed value

OK
AE anomaly detection: Inference

Evaluate:
One-side hypothesis test on the input-output distance

Observed value

d(in, out)
Online anomaly detection: a model independent trigger

O. Cerri\textsuperscript{a}, T. Q. Nguyen\textsuperscript{a}, M. Pierini\textsuperscript{b}, M. Spiropulu\textsuperscript{a}, J. R. Vlimant\textsuperscript{a}

\textsuperscript{a} CalTech, \textsuperscript{b} CERN

Anomalies in lepton stream

- Stream of data with at least one interesting lepton (e or μ)
  - \( p_T > 23 \text{ GeV} \) & ISO < 0.45
  - Too high rate to be tolerated at LHC

- SM contribution
  - W (59%), QCD (34%), Z (6%), tt (1%)

- Events represented by 21 high level features (HLF)
  - Broad general choice, not BSM tailored
Procedure

1. **Train** one (or more) VAE(s):
   a. Train on MC or **on data** (robust against signal injection - backup or ask)

2. Run the **VAE(s) in the High Level Trigger**
   a. Evaluate each event
   b. Acceptance threshold tuned to $O(10)$ SM events/day

3. Collect events in a **dedicated dataset**
   a. Visual inspection
   b. Develop targeted analysis

Remind: VAE does not see the BSM (if any) until it’s evaluated on new events
BSM benchmark models

Light BSM which are usually very hard to trigger with standard strategies

- $A \rightarrow 4\ell$: neutral scalar, $M = 50$ GeV
- $LQ \rightarrow b\tau$: leptoquark, $M = 80$ GeV
- $h^0 \rightarrow \tau\tau$: neutral scalar, $M = 60$ GeV
- $h^\pm \rightarrow \tau\nu$: charged scalar, $M = 60$ GeV

**BENCHMARKING ONLY, NOT USED FOR TRAINING**
Results

- (Unsupervised) VAE
  - Single one, trained only on SM
  - No info on BSM \(\rightarrow\) less powerful
  - Applies to any BSM \(\rightarrow\) more general

\[ \varepsilon_{SM} = 5.4 \cdot 10^{-6} \Leftrightarrow 30 \text{ evts/day} \]
Results

--- Model dep.  VAE

- (Unsupervised) VAE
  - Single one, trained only on SM
  - No info on BSM $\rightarrow$ less powerful
  - Applies to any BSM $\rightarrow$ more general

- (Supervised) Model-dependent classifiers
  - 4 algorithms, each trained on a specific BSM scenario
  - Set optimal performances

$\varepsilon_{SM} = 5.4 \times 10^{-6} \Leftrightarrow 30$ evts/day
Results

--- Model dep. ——— VAE
--- Model dep. on a different model

- (Unsupervised) VAE
  - Single one, trained only on SM
  - No info on BSM $\rightarrow$ less powerful
  - Applies to any BSM $\rightarrow$ more general

- (Supervised) Model-dependent classifiers
  - 4 algorithms, each trained on a specific BSM scenario
  - Set target performances

- Model dep. clf applied to a different BSM model

$\varepsilon_{SM} = 5.4 \cdot 10^{-6} \Leftrightarrow 30$ evts/day
Results

--- Model dep. VAE

••• Model dep. on a different model

$\varepsilon_{SM} = 5.4 \cdot 10^{-6} \Leftrightarrow 30 \text{ evts/day}$

<table>
<thead>
<tr>
<th>Process</th>
<th>VAE selection</th>
<th>Sample composition</th>
<th>Event/month</th>
</tr>
</thead>
<tbody>
<tr>
<td>$W$</td>
<td>$3.6 \pm 0.7 \cdot 10^{-6}$</td>
<td>32%</td>
<td>$379 \pm 74$</td>
</tr>
<tr>
<td>QCD</td>
<td>$6.0 \pm 2.3 \cdot 10^{-6}$</td>
<td>29%</td>
<td>$357 \pm 143$</td>
</tr>
<tr>
<td>$Z$</td>
<td>$21 \pm 3.5 \cdot 10^{-6}$</td>
<td>21%</td>
<td>$256 \pm 43$</td>
</tr>
<tr>
<td>$t\bar{t}$</td>
<td>$400 \pm 9 \cdot 10^{-6}$</td>
<td>18%</td>
<td>$212 \pm 5$</td>
</tr>
<tr>
<td>Tot</td>
<td></td>
<td></td>
<td>$1204 \pm 167$</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>$A \rightarrow 4\ell$</td>
<td>$2.8 \cdot 10^{-3}$</td>
<td>7.1</td>
<td>27</td>
</tr>
<tr>
<td>$LQ \rightarrow b\tau$</td>
<td>$6.7 \cdot 10^{-4}$</td>
<td>30</td>
<td>110</td>
</tr>
<tr>
<td>$h^0 \rightarrow \tau\tau$</td>
<td>$3.6 \cdot 10^{-4}$</td>
<td>55</td>
<td>210</td>
</tr>
<tr>
<td>$h^\pm \rightarrow \tau\nu$</td>
<td>$1.2 \cdot 10^{-3}$</td>
<td>17</td>
<td>65</td>
</tr>
</tbody>
</table>

Efficiency drop $\lesssim 10$ w.t.r. to model-dependent classifier (i.e. optimal limit)
Offline anomaly detection: a model independent analysis

O. Cerri\textsuperscript{a}, J. M. G. Duarte\textsuperscript{b}, J. Ngadiuba\textsuperscript{c}, T. Q. Nguyen\textsuperscript{a}, M. Pierini\textsuperscript{c}, M. Spiropulu\textsuperscript{a}, J. R. Vlimant\textsuperscript{a}, K. A. Wozniak\textsuperscript{c,d}

\textsuperscript{a} CalTech, \textsuperscript{b} Fermilab, \textsuperscript{c} CERN, \textsuperscript{d} U. of Vienna

arXiv:19XX.XXXXX
Extend an existing supervised search

Di-Jet bump hunt

1. Train a VAE on SM jets from a control region [1,2]

2. Choose a VAE selection with constant efficiency for the control region as a function of a variable of interest (M_{jj}) [3]

New search strategy

- Selection force the rejected and accepted events to have the same mass spectrum
  - In control region by construction
  - In signal region, if no BSM is present

- BSM-agnostic search:
  - Hypothesis test comparing rejected and accepted shapes
  - Improve state-of-art [1]

Results

- Di-jet events
  - Injected 1.5 TeV graviton
  - Control region $|\Delta \eta_{jj}| > 1.4$

- Image-based AE

- Defined threshold to retain 1% of SM events as anomalous
  - Model independent: comparing standard to anomaly $M_{jj}$ spectrum

- Model Dependent bump hunt

Discovery significance

Only a factor 2 drop wtr to optimal performances
Conclusions

● **VAE as model-independent BSM trigger**
  ○ Can be trained on data
  ○ Select otherwise lost anomalous events for further studies
  ○ Allows benchmark models to be probed down to 10-100 pb cross section

● **VAE based model-independent analysis**
  ○ New method proposed to turn each LHC search with jets into a model-independent test
  ○ Provides data driven estimation of the background shape
  ○ Define anomalies and provide a test to look for non-SM processes

● **Alternative strategies** might open new directions
BACKUP
Train on data

If BSM is rare enough, having it in the training sample will not spoil performances.

- Train on a dataset with signal injected:

<table>
<thead>
<tr>
<th>Injected evts</th>
<th>Training set fraction</th>
<th>VAE selected evts/month</th>
<th>Anomaly fraction</th>
</tr>
</thead>
<tbody>
<tr>
<td>700</td>
<td>$2 \cdot 10^{-4}$</td>
<td>134</td>
<td>12%</td>
</tr>
<tr>
<td>7k</td>
<td>$2 \cdot 10^{-3}$</td>
<td>957</td>
<td>48%</td>
</tr>
<tr>
<td>70k</td>
<td>$2 \cdot 10^{-2}$</td>
<td>6</td>
<td>0.6%</td>
</tr>
</tbody>
</table>

- SM size: 3.5M evts $\approx$ 100 pb$^{-1}$ $\approx$ few hours

No performance drop up to $10^{-3}$ signal contamination in training set (huge, S/B = 1):
⇒ Can be trained on data without impacting BSM efficiency
Standard events sample

Train → NN (VAE) → Function $f(x) \approx P_{SM}(x)$ → $f(x) < \varepsilon$?

Save it → Discard
Working hypothesis:

- Each event has a set of features: $x \in \mathbb{R}^n$
- Relevant information can be summarized in: $z \in \mathbb{R}^m$ ($n>m$)
  - Lost information for is somehow stored in the encoding/decoding function

Goal:

- Creating a function that, ON THE STD DATASET, allow to consistently compress and decompress the event information
  - the VAE should underperform on a different dataset because the lost information is different from the one of the training
- Consistency can be directly checked by comparing input and output
Training loss function technicalities

\[ \text{Loss}_{\text{Tot}} = \text{Loss}_{\text{reco}} + \lambda D_{\text{KL}} \]

Reconstruction likelihood:

- “True” loss (NLL)
- Force the autoencoded distribution to describe the \( x \)
- The goodness of the VAE depends on the ability of \( f_j \) to describe \( p(x \mid z) \)

\[
\text{Loss}_{\text{reco}} = -\frac{1}{k} \sum_i \ln \left( P(x \mid \alpha_1, \alpha_2, \alpha_3) \right)
= -\frac{1}{k} \sum_{i,j} \ln \left( f_j(x_{i,j} \mid \alpha_{1,j}^i, \alpha_{2,j}^i, \alpha_{3,j}^i) \right)
\]

Regularization term:

- Force the \( z \) distribution to a Normal
- To avoid strange latent variable

\[
D_{\text{KL}} = \frac{1}{k} \sum_i D_{\text{KL}} \left( N(\mu_z^i, \sigma_z^i) \parallel N(\mu_P, \sigma_P) \right)
\]
The Variational Auto-Encoder

Encoder:

- For each value of $x$, tell what is the pdf of $z$
- Practically:
  - A functional form $f_e[z; \alpha_e(x)]$ is fixed

Decoder:

- For each value of $z$, tell what is the pdf of $x$
- Practically:
  - A functional form $f_d[x; \alpha_d(z)]$ is fixed

The encoder function $g_e : x \rightarrow \alpha_e$ gives the value of the $z$ distribution parameters.

$x$ and $z$ are swapped w.r.t. to Encoder.

The encoder function $g_d : z \rightarrow \alpha_d$ gives the value of the $x$ distribution parameters.
Input (x) : 21 HLF (HT, MET, njets, …)

2 dense hidden layer (50 neurons)

4-dim latent (z) space
- Trainable mean and sigma

x’ pdf: max 1–3 parameters
- Gaussian
- Binomial
- ...
A whole art exist in choosing the functional form

\[
D_{KL} [N(\mu_z, \sigma_z) \| N(\mu_p, \sigma_p)]
\]

\[
N(\mu_z, \sigma_z)
\]

\[
z(4)
\]

\[
Decoded h1 (50)
\]

\[
Decoded h2 (50)
\]

\[
\alpha_1(21), \alpha_2(17), \alpha_3(10)
\]

\[
P(\mathbf{x} | \alpha_1, \alpha_2, \alpha_3)
\]

\[
Loss = \beta \cdot D_{KL} - \ln(P)
\]

...and architecture details
Training: not a easy beast

- **Optimizer**
  - Adam
  - Callbacks

- **Samples**
  - 3.5 M event for training
  - 3.5 M for validation
  - # evt/# par >> 10

- **The training**
  - Not long, about 1h
  - Spike not unusual
  - Delicate equilibrium of training parameters
1. Start from an existing supervised search with jets (e.g., dijet bumphunt)

2. Train a VAE to “learn” QCD jets on a control region (e.g. $|\Delta\eta_{jj}|>1.4$)
   a. Heimel et al. [1]
      https://arxiv.org/abs/1808.08979
   b. Franco et al. [2]

3. Modulate the cut on the VAE loss as a function of a variable of interest ($M_{jj}$) using DDT
   a. Selected events -> anomaly-enriched sample
   b. Rejected events -> background control region
Latent space distribution
Simulation details:
- Pythia 8
- Delphes
  - CMS phase II default card
- Training on 3.5 M of SM
  - Equivalent of 100 pb$^{-1}$

Machine working conditions:
- 8 months of data taking per year
- $L_{TOT} = 40$ fb$^{-1}$
- $<L_{inst}> = 2.8 \cdot 10^{33}$ cm$^{-2}$s$^{-1}$
- $<PU> = 20$
- $E_{CM} = 13$ TeV
The 21 considered features

- The absolute value of the isolated-lepton transverse momentum $p_T^\ell$.
- The three isolation quantities (ChPFISO, NeuPFISO, GammaPFISO) for the isolated lepton, computed with respect to charged particles, neutral hadrons and photons, respectively.
- The lepton charge.
- A Boolean flag (isELe) set to 1 when the trigger lepton is an electron, 0 otherwise.
- $S_T$, i.e. the scalar sum of the $p_T$ of all the jets, leptons, and photons in the event with $p_T > 30$ GeV and $|\eta| < 2.6$. Jets are clustered from the reconstructed PF candidates, using the FASTJET [24] implementation of the anti-$k_T$ jet algorithm [25], with jet-size parameter $R=0.4$.
- The number of jets entering the $S_T$ sum ($N_J$).
- The invariant mass of the set of jets entering the $S_T$ sum ($M_J$).
- The number of these jets being identified as originating from a $b$ quark ($N_b$).
- The missing transverse momentum, decomposed into its parallel ($p_T^{miss,||}$) and orthogonal ($p_T^{miss,\perp}$) components with respect to the isolated lepton direction. The missing transverse momentum is defined as the negative sum of the PF-candidate $p_T$ vectors:

$$
\vec{p}_T^{miss} = -\sum_q \vec{p}_T^q .
$$

- The transverse mass, $M_T$, of the isolated lepton $\ell$ and the $E_T^{miss}$ system, defined as:

$$
M_T = \sqrt{2p_T^\ell E_T^{miss}(1 - \cos \Delta \phi)} ,
$$

with $\Delta \phi$ the azimuth separation between the $\vec{p}_T^\ell$ and $\vec{p}_T^{miss}$ vectors, and $E_T^{miss}$ the absolute value of $p_T^{miss}$.
- The number of selected muons ($N_\mu$).
- The invariant mass of this set of muons ($M_\mu$).
- The absolute value of the total transverse momentum of these muons ($P_{T,TOT}^{\mu}$).
- The number of selected electrons ($N_e$).
- The invariant mass of this set of electrons ($M_e$).
- The absolute value of the total transverse momentum of these electrons ($P_{T,TOT}^{e}$).
- The number of reconstructed charged hadrons.
- The number of reconstructed neutral hadrons.
Clipped Log-normal + δ function: used to describe $S_T$, $M_J$, $p_T^\mu$, $M_\mu$, $p_T^e$, $M_e$, $p_T^\nu$, ChPFiso, NeuPFiso and GammaPFiso:

$$P(x \mid \alpha_1, \alpha_2, \alpha_3) = \begin{cases} \alpha_3 \delta(x) + \frac{1 - \alpha_3}{x \alpha_2 \sqrt{2\pi}} \exp\left(\frac{(\ln x - \alpha_1)^2}{2 \alpha_2^2}\right) & \text{for } x \geq 10^{-4} \\ 0 & \text{for } x < 10^{-4} \end{cases}.$$ (9)

Gaussian: used for $p_{T,||}^\text{miss}$ and $p_{T,\perp}^\text{miss}$:

$$P(x \mid \alpha_1, \alpha_2) = \frac{1}{\alpha_2 \sqrt{2\pi}} \exp\left(\frac{(x - \alpha_1)^2}{2 \alpha_2^2}\right).$$ (10)

Truncated Gaussian: a Gaussian function truncated for negative values and normalized to unit area for $X > 0$. Used to model $M_T$:

$$P(x \mid \alpha_1, \alpha_2) = \Theta(x) \cdot \frac{1 + 0.5 \cdot (1 + \text{erf}\frac{-\alpha_1}{\alpha_2 \sqrt{2}})}{\alpha_2 \sqrt{2\pi}} \exp\left(\frac{(x - \alpha_1)^2}{2 \alpha_2^2}\right).$$ (11)

Discrete truncated Gaussian: like the truncated Gaussian, but normalized to be evaluated on integers (i.e. $\sum_{n=0}^{\infty} P(n) = 1$). This function is used to describe $N_\mu$, $N_e$, $N_b$ and $N_J$. It is written as:

$$P(n \mid \alpha_1, \alpha_2) = \Theta(x) \left[ \text{erf}\left(\frac{n + 0.5 - \alpha_1}{\alpha_2 \sqrt{2}}\right) - \text{erf}\left(\frac{n - 0.5 - \alpha_1}{\alpha_2 \sqrt{2}}\right) \right] \mathcal{N},$$ (12)

where the normalization factor $\mathcal{N}$ is set to:

$$\mathcal{N} = 1 + \frac{1}{2} \left[ 1 + \text{erf}\left(\frac{-0.5 - \alpha_1}{\alpha_2 \sqrt{2}}\right) \right].$$ (13)

Binomial: used for $\text{IsEle}$ and lepton charge:

$$P(n \mid p) = \delta_{n,m}p + \delta_{n,l}(1-p)$$ (14)

where $m$ and $l$ are the two possible values of the variable (0 or 1 for $\text{IsEle}$ and -1 or 1 for lepton charge) and $p = C_{\text{tanh}}(\alpha_1)$

Poisson: used for charged-particle and neutral-hadron multiplicities:

$$P(n \mid \mu) = \frac{\mu^n e^{-\mu}}{\Gamma(n + 1)}$$ (15)

where $\mu = p-\text{ISRLu}(\alpha_1)$. 
VAE auto-encoding cross-check
Loss distributions

![Graph showing loss distributions for various processes and models.]

- **1k evts/month**
- **SM val. Mix**
- $h^0 \to \tau\tau$
- $A \to 4\ell$
- $h^{\pm} \to \tau\nu$
- LQ

**Axes:**
- **Loss_{reco}**
- **Probability**
- **$D_{KL}$**
- **Probability**
Not a tail-cut algorithm
Other algorithms comparison

--- PCA

--- VAE

--- AE

--- VAE

A→4ℓ

AUC = 0.91 (0.68)

LQ

AUC = 0.85 (0.71)

h⁰→ττ

AUC = 0.75 (0.70)

h⁺→τν

AUC = 0.92 (0.88)

1000 SM evts/month

10⁰

10⁻¹

10⁻²

10⁻³

10⁻⁴

10⁻⁵

10⁻⁶

10⁻⁷

10⁻⁸

10⁻⁹

10⁻¹⁰

10⁻¹¹

10⁻¹²

SM efficiency

BSM efficiency

SM efficiency

10⁰

10⁻¹

10⁻²

10⁻³

10⁻⁴

10⁻⁵

10⁻⁶

10⁻⁷

10⁻⁸

10⁻⁹

10⁻¹⁰

10⁻¹¹

10⁻¹²

10⁻¹³

10⁻¹⁴
Scenario w/o the VAE trigger

Reasonable cuts for single muon full trigger path (i.e. what we can really save on disk):

- $p_T > 27$ GeV
- $\text{ISO} < 0.25$

### VAE trigger improves S/N ratio of 2–3 order of magnitude

<table>
<thead>
<tr>
<th></th>
<th>SM</th>
<th>$A \rightarrow 4\ell$</th>
<th>$h \rightarrow \tau \tau$</th>
<th>$h \rightarrow \tau \nu$</th>
<th>LQ</th>
</tr>
</thead>
<tbody>
<tr>
<td>VAE</td>
<td>5e-6</td>
<td>3e-3</td>
<td>4e-4</td>
<td>1e-3</td>
<td>7e-4</td>
</tr>
<tr>
<td>Single muon trigger</td>
<td>0.6</td>
<td>0.5</td>
<td>0.6</td>
<td>0.7</td>
<td>0.6</td>
</tr>
</tbody>
</table>

The great advantage of VAE is not only the ability to select BSM events but also to produce a high purity sample.
Checking the convergence: sum of pdfs

High input dimension \(\Rightarrow\) Global convergence check

Obtain the distribution of the input as sum of all the predicted pdf

\[
x (\text{input}) \quad \xrightarrow{\text{Prediction}} \quad \text{Predicted pdf for the single event}
\]

Obtain the distribution of the input as sum of all the predicted pdf
Convergence check: SM auto-encoding

- Verifying **encoding-decoding** on validation set
  - Distributions of input vs generated from decoder

- **Good agreement**, with small discrepancy here and there

- Best autoencoder is not necessarily the best anomaly detector
Defining anomaly

- Anomaly defined by a p-value threshold on a given test statistics
- VAE loss function is the natural choice for the test statistics

Loss$_{\text{reco}}$ used as test statistics.
Not a tail-cut algorithm

- Selected events stand on the core of 1D distributions

- Expand the possibility w.r.t. to classical anomaly detection triggers
Results

- Applied to a sample of Dijet events
- Used Image-based Autoencoders
- Defined threshold to retain 1% of the events as anomalous
  - Model Dependent: Bump Hunt on inclusive sample
  - Model Independent: GoF comparing standard to anomaly Mjj distributions
  - (Not Shown) VAE-enhanced Model Dependent: Bump Hunt in the anomaly Mjj distribution (using standard to predict background)
New search strategies

- BSM-agnostic search: goodness-of-fit test comparing the two shapes
- BSM-specific search: classic search with simultaneous fit (e.g., bump hunt)
- Similar to Collins et al (COwLa)
  - but uses data only once
  - generalizes to any analysis with jets
  - no assumption on signal shape
The Variational Auto-Encoder

\[ X = \text{HLF} \]
The Variational Auto-Encoder

\[ X = HLF \]

NN with some free parameter

\[ g_a \]

\[ \mu_z \]

\[ \sigma_z \]
The Variational Auto-Encoder

Latent space ($z$)

model of:

$X = HLF$

$g_a$

$\mu_z$

$\sigma_z$

NN with some free parameter

Latent space ($z$)
The Variational Auto-Encoder

Latent space (z)

model of:

\[ x = HLF \]

NN with some free parameter

get random:

\[ \mathbf{z} \]

\[ \mu_z \]

\[ \sigma_z \]
The Variational Auto-Encoder

X = HLF

g_a

\sigma_z

\mu_z

model of:

Latent space (z)

z_1

z_2

\mu_1

\mu_2

get random:

NN with some free parameter

NN with some free parameter

\mu_x

\sigma_x

NN with some free parameter
The Variational Auto-Encoder

Latent space \( z \)

\[
x = HLF_g \alpha_d \sim \mathcal{N}(\mu_x, \sigma_x)
\]

\[
x = HLF_g \alpha_d \sim \mathcal{N}(\mu_x, \sigma_x)
\]
The Variational Auto-Encoder

Latent space (z)

\[ z \]

\[ \mu_1 \]

\[ \mu_2 \]

\[ \sigma_z \]

\[ \sigma_z \]

\[ HLF \]

\[ g_a \]

\[ \mu_x \]

\[ \sigma_x \]

\[ \alpha_d \]

Probability

\[ X_1 \]

\[ X_{21} \]

Probability

\[ \mu_x^1 \]

\[ \sigma_x^1 \]

\[ \mu_x^n \]

\[ \sigma_x^n \]
The Variational Auto-Encoder

\[ X = \text{HLF} \]

model of:

Latent space (z)

\[ z \]

\[ \mu_1 \]

\[ \sigma_z \]

NN with some free parameter

get random:

\[ z \]

\[ \mu_2 \]

\[ \sigma_z \]

NN with some free parameter

Loss_{reco} = -\ln P[x; \alpha_d(z(x))]
Attentional Particle-VAE

- **Attention**: a function of both list of input particles and the current hidden state of the decoder’s RNN cell.
Particle based VAE performance

Performance (2/2)

- Roughly 10 times worse than the VAE trained on HLFs.
- Optimization in progress, could be improved much further (more data + optimized loss functions).

<table>
<thead>
<tr>
<th>Sample</th>
<th>Efficiency</th>
<th>Rate [Hz]</th>
<th>evts/month</th>
</tr>
</thead>
<tbody>
<tr>
<td>ttbar</td>
<td>2.3E-3 +/- 1.5E-4</td>
<td>5.7E-3</td>
<td>4.8E+3 +/- 3.2E+2</td>
</tr>
<tr>
<td>QCD</td>
<td>1.0E-5 +/- 1.0E-5</td>
<td>2.5E-3</td>
<td>2.1E+3 +/- 2.1E+3</td>
</tr>
<tr>
<td>W1nu</td>
<td>0.0E+1 +/- 0.0E+1</td>
<td>0.0E+1</td>
<td>0.0E+1 +/- 0.0E+1</td>
</tr>
</tbody>
</table>

Expected evts/month: 6883 +/- 5228

<table>
<thead>
<tr>
<th>Sample</th>
<th>Efficiency</th>
<th>xsec (10 evts/month) [fb]</th>
<th>xsec (S/B = 0.3) [fb]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ato4l</td>
<td>3.3e-4 +/- 8.6e-5</td>
<td>7.2E+3</td>
<td>1.5E+6</td>
</tr>
<tr>
<td>leptogluark</td>
<td>5.8e-4 +/- 7.6e-5</td>
<td>4.1E+3</td>
<td>8.5E+5</td>
</tr>
<tr>
<td>HiggsToTauTau</td>
<td>1.1e-3 +/- 1.5e-4</td>
<td>2.2E+3</td>
<td>4.5E+5</td>
</tr>
<tr>
<td>ChHiggsToTauNu</td>
<td>1.4e-3 +/- 1.7e-4</td>
<td>1.7E+3</td>
<td>3.4E+5</td>
</tr>
</tbody>
</table>