

# Searching for New Physics with Deep Autoencoders

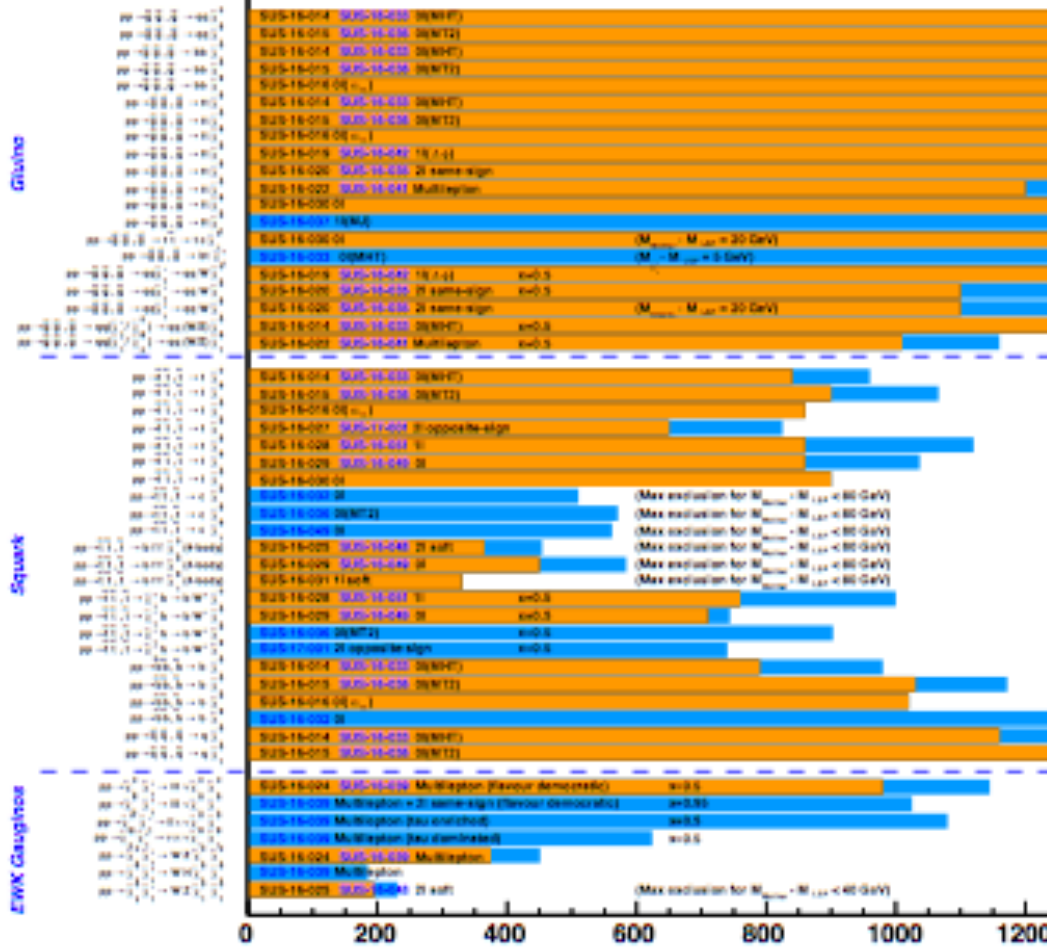
David Shih  
NHETC, Rutgers University

IML Working Group Seminar  
Oct 30, 2018

Based on Farina, Nakai & DS 1808.08992

Selected CMS SUSY Results\* - SMS Interpretation

ICHEP '16 - Moriond '17



\*Observed limits at 95% C.L. - theory uncertainties not included  
Only a selection of available mass limits. Probe "up to" the quoted mass limit

ATLAS SUSY Searches\* - 95% CL Lower Limits  
May 2017

Model	$e, \mu, \tau, \gamma$	Jets	$E_T^{miss}$	$\int \mathcal{L} dt [fb^{-1}]$	Mass limit	$\sqrt{s} = 7, 8 \text{ TeV}$	$\sqrt{s} = 13 \text{ TeV}$	Reference
Inclusive Searches	MSUGRA/CMSSM	$0-3 e, \mu / 1-2 \tau$	2-10 jets/3 b	Yes	20.3	$\tilde{g}, \tilde{q}$	1.65 TeV	$m(\tilde{g})=m(\tilde{q})$ 1507.05525
	$\tilde{g}\tilde{g}, \tilde{q}\tilde{q} \rightarrow q\bar{q}g$	0	2-6 jets	Yes	36.1	$\tilde{g}$	1.57 TeV	$m(\tilde{g}) < 200 \text{ GeV}, m(1^{st} \text{ gen. } \tilde{q}) = m(2^{nd} \text{ gen. } \tilde{q})$ ATLAS-CONF-2017-022
	$\tilde{g}\tilde{g}, \tilde{q}\tilde{q} \rightarrow q\bar{q}g$ (compressed)	mono-jet	1-3 jets	Yes	3.2	$\tilde{g}$	608 GeV	$m(\tilde{g}) = m(\tilde{q}) < 5 \text{ GeV}$ 1604.07773
	$\tilde{g}\tilde{g}, \tilde{q}\tilde{q} \rightarrow q\bar{q}g$	0	2-6 jets	Yes	36.1	$\tilde{g}$	2.02 TeV	$m(\tilde{g}) < 200 \text{ GeV}$ ATLAS-CONF-2017-022
	$\tilde{g}\tilde{g}, \tilde{q}\tilde{q} \rightarrow q\bar{q}g, W^{\pm}Z$	0	2-6 jets	Yes	36.1	$\tilde{g}$	2.01 TeV	$m(\tilde{g}) < 200 \text{ GeV}, m(\tilde{q}) = 0.5(m(\tilde{g}) + m(\tilde{q}))$ ATLAS-CONF-2017-022
	$\tilde{g}\tilde{g}, \tilde{q}\tilde{q} \rightarrow q\bar{q}g, W^{\pm}Z, \tau\tau$	3 e, $\mu$	4 jets	-	36.1	$\tilde{g}$	1.825 TeV	$m(\tilde{g}) < 400 \text{ GeV}$ ATLAS-CONF-2017-033
	$\tilde{g}\tilde{g}, \tilde{q}\tilde{q} \rightarrow q\bar{q}g, W^{\pm}Z, \tau\tau$	0	7-11 jets	Yes	36.1	$\tilde{g}$	1.8 TeV	$m(\tilde{g}) < 400 \text{ GeV}$ 1607.05979
	GMSB ( $\tilde{g}$ NLSP)	$1-2 \tau + 0-1 \ell$	0-2 jets	Yes	3.2	$\tilde{g}$	2.0 TeV	$cr(NLSP) < 0.1 \text{ mm}$ 1606.09150
	GGM (bino NLSP)	$2 \gamma$	-	Yes	3.2	$\tilde{g}$	1.65 TeV	$m(\tilde{g}) < 950 \text{ GeV}, cr(NLSP) < 0.1 \text{ mm}, \mu < 0$ 1507.05493
	GGM (higgsino-bino NLSP)	$\gamma$	1 b	Yes	20.3	$\tilde{g}$	1.37 TeV	$m(\tilde{g}) < 680 \text{ GeV}, cr(NLSP) < 0.1 \text{ mm}, \mu < 0$ ATLAS-CONF-2016-066
3 <sup>rd</sup> gen. squarks & med.	$\tilde{g}\tilde{g}, \tilde{q}\tilde{q} \rightarrow b\bar{b}g$	0	3 b	Yes	36.1	$\tilde{g}$	1.92 TeV	$m(\tilde{g}) < 600 \text{ GeV}$ ATLAS-CONF-2017-021
	$\tilde{g}\tilde{g}, \tilde{q}\tilde{q} \rightarrow t\bar{t}g$	0-1 e, $\mu$	3 b	Yes	36.1	$\tilde{g}$	1.97 TeV	$m(\tilde{g}) < 200 \text{ GeV}$ ATLAS-CONF-2017-021
	$\tilde{g}\tilde{g}, \tilde{q}\tilde{q} \rightarrow b\bar{b}g$	0-1 e, $\mu$	3 b	Yes	20.1	$\tilde{g}$	1.37 TeV	$m(\tilde{g}) < 300 \text{ GeV}$ 1407.0600
3 <sup>rd</sup> gen. squarks direct production	$\tilde{b}_1\tilde{b}_1, \tilde{b}_1 \rightarrow b\bar{b}g$	0	2 b	Yes	36.1	$\tilde{b}_1$	950 GeV	$m(\tilde{g}) < 420 \text{ GeV}$ ATLAS-CONF-2017-038
	$\tilde{b}_1\tilde{b}_1, \tilde{b}_1 \rightarrow t\bar{t}g$	2 e, $\mu$ (SS)	1 b	Yes	36.1	$\tilde{b}_1$	275-700 GeV	$m(\tilde{g}) < 200 \text{ GeV}, m(\tilde{q}) = m(\tilde{t}) + 100 \text{ GeV}$ ATLAS-CONF-2017-030
	$\tilde{t}_1\tilde{t}_1, \tilde{t}_1 \rightarrow b\bar{b}g$	0-2 e, $\mu$	1-2 b	Yes	4.7/13.3	$\tilde{t}_1$	117-170 GeV	$m(\tilde{g}) = 2m(\tilde{t}_1), m(\tilde{q}) = 55 \text{ GeV}$ 1209.2102, ATLAS-CONF-2016-077
	$\tilde{t}_1\tilde{t}_1, \tilde{t}_1 \rightarrow Wb\bar{c}g$ or $\tilde{t}_1^0$	0-2 e, $\mu$	0-2 jets/1-2 b	Yes	20.3/36.1	$\tilde{t}_1$	90-198 GeV	$m(\tilde{g}) = 1 \text{ GeV}$ 1506.08616, ATLAS-CONF-2017-020
EW direct	$\tilde{t}_1\tilde{t}_1, \tilde{t}_1 \rightarrow c\bar{c}g$	0	mono-jet	Yes	3.2	$\tilde{t}_1$	90-323 GeV	$m(\tilde{g}) = m(\tilde{t}_1) = 5 \text{ GeV}$ 1604.07773
	$\tilde{t}_1\tilde{t}_1$ (natural GMSB)	2 e, $\mu$ (Z)	1 b	Yes	20.3	$\tilde{t}_1$	150-600 GeV	$m(\tilde{g}) < 150 \text{ GeV}$ 1403.5222
	$\tilde{t}_2\tilde{t}_2, \tilde{t}_2 \rightarrow t\bar{t}g$	3 e, $\mu$ (Z)	1 b	Yes	36.1	$\tilde{t}_2$	290-790 GeV	$m(\tilde{g}) = 0 \text{ GeV}$ ATLAS-CONF-2017-019
	$\tilde{t}_2\tilde{t}_2, \tilde{t}_2 \rightarrow t\bar{t}g$	1-2 e, $\mu$	4 b	Yes	36.1	$\tilde{t}_2$	320-680 GeV	$m(\tilde{g}) = 0 \text{ GeV}$ ATLAS-CONF-2017-019
	$\tilde{t}_1\tilde{t}_2, \tilde{t}_1 \rightarrow c\bar{c}g$	2 e, $\mu$	0	Yes	36.1	$\tilde{t}_1$	90-440 GeV	$m(\tilde{g}) = 0$ ATLAS-CONF-2017-039
	$\tilde{t}_1\tilde{t}_2, \tilde{t}_1 \rightarrow t\bar{t}g$	2 e, $\mu$	0	Yes	36.1	$\tilde{t}_1$	710 GeV	$m(\tilde{g}) = 0, m(\tilde{t}_1, \nu) = 0.5(m(\tilde{t}_1) + m(\tilde{t}_2))$ ATLAS-CONF-2017-039
	$\tilde{t}_1\tilde{t}_2, \tilde{t}_1 \rightarrow t\bar{t}g$	2 e, $\mu$	0	Yes	36.1	$\tilde{t}_1$	760 GeV	$m(\tilde{g}) = 0, m(\tilde{t}_1, \nu) = 0.5(m(\tilde{t}_1) + m(\tilde{t}_2))$ ATLAS-CONF-2017-035
	$\tilde{t}_1\tilde{t}_2, \tilde{t}_1 \rightarrow t\bar{t}g$	3 e, $\mu$	0	Yes	36.1	$\tilde{t}_1$	1.16 TeV	$m(\tilde{g}) = m(\tilde{t}_2), m(\tilde{q}) = 0, m(\tilde{t}_1, \nu) = 0.5(m(\tilde{t}_1) + m(\tilde{t}_2))$ ATLAS-CONF-2017-039
	$\tilde{t}_1\tilde{t}_2, \tilde{t}_1 \rightarrow t\bar{t}g$	2-3 e, $\mu$	0-2 jets	Yes	36.1	$\tilde{t}_1$	580 GeV	$m(\tilde{g}) = m(\tilde{t}_2), m(\tilde{q}) = 0, \tilde{t}$ decoupled ATLAS-CONF-2017-039
	$\tilde{t}_1\tilde{t}_2, \tilde{t}_1 \rightarrow t\bar{t}g$	e, $\mu, \gamma$	0-2 b	Yes	20.3	$\tilde{t}_1$	270 GeV	$m(\tilde{g}) = m(\tilde{t}_2), m(\tilde{q}) = 0, \tilde{t}$ decoupled 1501.07110
Long-lived particles	$\tilde{t}_1\tilde{t}_2, \tilde{t}_1 \rightarrow t\bar{t}g$	4 e, $\mu$	0	Yes	20.3	$\tilde{t}_1$	635 GeV	$m(\tilde{g}) = m(\tilde{t}_2), m(\tilde{q}) = 0, m(\tilde{t}_1, \nu) = 0.5(m(\tilde{t}_1) + m(\tilde{t}_2))$ 1405.5086
	GGM (wino NLSP) weak prod., $\tilde{X}_1^0 \rightarrow \gamma G$	1 e, $\mu + \gamma$	-	Yes	20.3	$\tilde{X}_1^0$	115-370 GeV	$m(\tilde{g}) = m(\tilde{X}_1^0), m(\tilde{q}) = 0, m(\tilde{t}_1, \nu) = 0.5(m(\tilde{t}_1) + m(\tilde{t}_2))$ 1507.05493
	GGM (bino NLSP) weak prod., $\tilde{X}_1^0 \rightarrow \gamma G$	2 $\gamma$	-	Yes	20.3	$\tilde{X}_1^0$	590 GeV	$cr < 1 \text{ mm}$ 1507.05493
	Direct $\tilde{X}_1^0 \tilde{X}_1^0$ prod., long-lived $\tilde{X}_1^0$	Disapp. trk	1 jet	Yes	36.1	$\tilde{X}_1^0$	430 GeV	$m(\tilde{g}) = m(\tilde{X}_1^0) = 160 \text{ MeV}, \tau(\tilde{X}_1^0) = 0.2 \text{ ns}$ ATLAS-CONF-2017-017
	Direct $\tilde{X}_1^0 \tilde{X}_1^0$ prod., long-lived $\tilde{X}_1^0$	dE/dx trk	-	Yes	18.4	$\tilde{X}_1^0$	495 GeV	$m(\tilde{g}) = m(\tilde{X}_1^0) = 160 \text{ MeV}, \tau(\tilde{X}_1^0) < 15 \text{ ns}$ 1506.05332
	Stable, stopped $\tilde{g}$ R-hadron	0	1-5 jets	Yes	27.9	$\tilde{g}$	850 GeV	$m(\tilde{g}) = 100 \text{ GeV}, 10 \mu\text{s} < \tau(\tilde{g}) < 1000 \text{ s}$ 1310.6584
	Stable $\tilde{g}$ R-hadron	trk	-	-	-	$\tilde{g}$	1.58 TeV	$m(\tilde{g}) = 100 \text{ GeV}, \tau > 10 \text{ ns}$ 1606.05129
	Metastable $\tilde{g}$ R-hadron	dE/dx trk	-	-	-	$\tilde{g}$	1.57 TeV	$m(\tilde{g}) = 100 \text{ GeV}, \tau > 10 \text{ ns}$ 1604.04520
	GMSB, stable $\tilde{t}, \tilde{X}_1^0 \rightarrow \tau(\tilde{t}, \tilde{\mu}) + \tau(e, \mu)$	1-2 $\mu$	-	-	-	$\tilde{X}_1^0$	537 GeV	$10 < \tau < 50$ 1411.6795
	GMSB, $\tilde{X}_1^0 \rightarrow \gamma G$ , long-lived $\tilde{X}_1^0$	2 $\gamma$	-	Yes	20.3	$\tilde{X}_1^0$	440 GeV	$1 < \tau(\tilde{X}_1^0) < 3 \text{ ns}$ , SPS8 model 1409.5542
RPV	$\tilde{g}\tilde{g}, \tilde{X}_1^0 \rightarrow ee\gamma/\mu\mu\gamma$	displ. ee/ $\mu\mu$	-	-	20.3	$\tilde{X}_1^0$	1.0 TeV	$7 < \tau(\tilde{X}_1^0) < 740 \text{ mm}, m(\tilde{g}) = 1.3 \text{ TeV}$ 1504.05162
	GGM $\tilde{g}\tilde{g}, \tilde{X}_1^0 \rightarrow ZG$	displ. vtx + jets	-	-	20.3	$\tilde{X}_1^0$	1.0 TeV	$6 < \tau(\tilde{X}_1^0) < 480 \text{ mm}, m(\tilde{g}) = 1.1 \text{ TeV}$ 1504.05162
	LFV $pp \rightarrow \tilde{\nu}_i + X, \tilde{\nu}_i \rightarrow e\mu/\tau\mu/\mu\tau$	$e\mu, \tau\mu, \mu\tau$	-	-	-	$\tilde{\nu}_i$	1.9 TeV	$A'_{11} = 0.11, A'_{22} = 0.233, \tau > 0.07$ 1607.08079
	Bilinear RPV CMSSM	2 e, $\mu$ (SS)	0-3 b	Yes	20.3	$\tilde{g}, \tilde{q}$	1.45 TeV	$m(\tilde{g}) = m(\tilde{q}), cr_{LSP} < 1 \text{ mm}$ 1404.2500
	$\tilde{X}_1^0 \tilde{X}_1^0, \tilde{X}_1^0 \rightarrow WZ, \tilde{X}_1^0 \rightarrow ee\nu, \mu\nu, \mu\nu\nu$	4 e, $\mu$	-	Yes	13.3	$\tilde{X}_1^0$	1.14 TeV	$m(\tilde{g}) > 400 \text{ GeV}, A_{1,2,3} \neq 0 (k = 1, 2)$ ATLAS-CONF-2016-075
	$\tilde{X}_1^0 \tilde{X}_1^0, \tilde{X}_1^0 \rightarrow WZ, \tilde{X}_1^0 \rightarrow \tau\nu, \tau\nu, \tau\nu, \tau\nu$	3 e, $\mu + \tau$	-	Yes	20.3	$\tilde{X}_1^0$	450 GeV	$m(\tilde{g}) > 0.2 \times m(\tilde{X}_1^0), A_{1,2,3} \neq 0$ 1405.5086
	$\tilde{g}\tilde{g}, \tilde{q}\tilde{q} \rightarrow q\bar{q}g$	0	4-5 large-R jets	-	14.8	$\tilde{g}$	1.08 TeV	$BR(\tilde{g}) = BR(\tilde{q}) = BR(\tilde{t}) = 0\%$ ATLAS-CONF-2016-057
	$\tilde{g}\tilde{g}, \tilde{q}\tilde{q} \rightarrow q\bar{q}g$	0	4-5 large-R jets	-	14.8	$\tilde{g}$	1.55 TeV	$m(\tilde{g}) = 800 \text{ GeV}$ ATLAS-CONF-2016-057
	$\tilde{g}\tilde{g}, \tilde{q}\tilde{q} \rightarrow q\bar{q}g, \tilde{X}_1^0 \rightarrow q\bar{q}q$	1 e, $\mu$	8-10 jets/0-4 b	-	36.1	$\tilde{g}$	2.1 TeV	$m(\tilde{g}) = 1 \text{ TeV}, A_{1,2,3} \neq 0$ ATLAS-CONF-2017-013
	$\tilde{g}\tilde{g}, \tilde{q}\tilde{q} \rightarrow q\bar{q}g, \tilde{X}_1^0 \rightarrow q\bar{q}q$	1 e, $\mu$	8-10 jets/0-4 b	-	36.1	$\tilde{g}$	1.65 TeV	$m(\tilde{g}) = 1 \text{ TeV}, A_{1,2,3} \neq 0$ ATLAS-CONF-2017-013
Other	$\tilde{t}_1\tilde{t}_1, \tilde{t}_1 \rightarrow b\bar{b}g$	0	2 jets + 2 b	-	15.4	$\tilde{t}_1$	410 GeV	$BR(\tilde{t}_1 \rightarrow b\bar{c}) = 20\%$ ATLAS-CONF-2016-084
	$\tilde{t}_1\tilde{t}_1, \tilde{t}_1 \rightarrow b\bar{b}g$	2 e, $\mu$	2 b	-	36.1	$\tilde{t}_1$	450-510 GeV	$BR(\tilde{t}_1 \rightarrow b\bar{c}) = 20\%$ ATLAS-CONF-2017-036
	Scalar charm, $Z \rightarrow c\bar{c}$	0	2 c	Yes	20.3	$\tilde{Z}$	510 GeV	$m(\tilde{g}) < 200 \text{ GeV}$ 1501.01325

\*Only a selection of the available mass limits on new states or phenomena is shown. Many of the limits are based on simplified models, c.f. refs. for the assumptions made.

The LHC has searched for new physics in many places.

So far, there has been no evidence of anything.

Many well-motivated models (SUSY, composite higgs, dark matter, ...) have not turned up as expected.

People are losing interest in “well-motivated models”...

Fraction of hep-ph papers on **SUSY** @ LHC



We need new ideas!

Can we search for new physics in the data without knowing what we're looking for?

Can we find the unexpected?

Can we find a needle in a haystack, without knowing what needles are?

Sounds hopelessly difficult...

**Maybe deep learning can help!**

# Deep learning at LHC

Recently there has been a lot of interest in applications of deep learning to the LHC.

- classification (eg quark/gluon tagging, boosted resonance tagging)
- pile-up removal
- event generation
- triggering
- anomaly detection
- .....

# Deep learning at LHC

Recently there has been a lot of interest in applications of deep learning to the LHC.

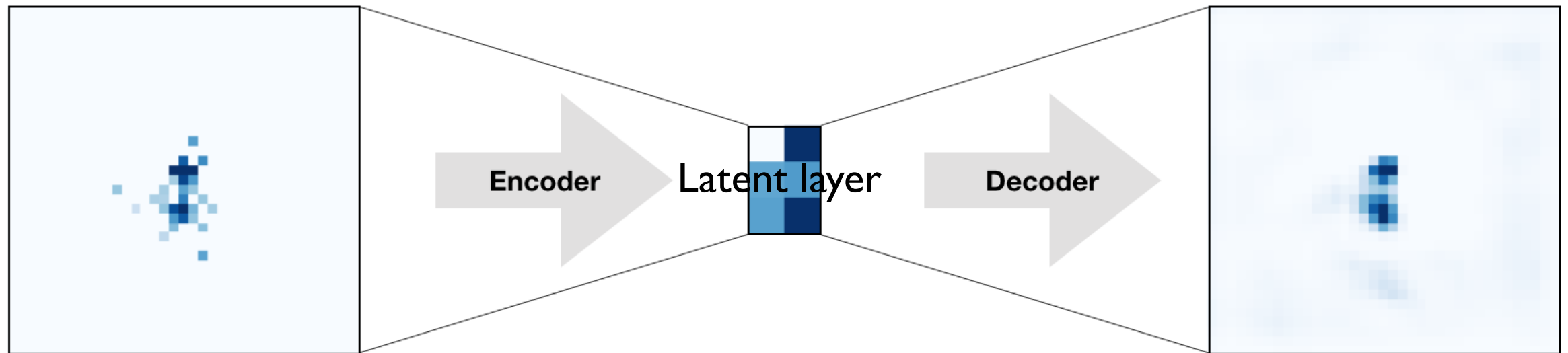
- classification (eg quark/gluon tagging, boosted resonance tagging)
- pile-up removal
- event generation
- triggering
- **anomaly detection**
- .....

**This talk**

A promising idea:

# Deep autoencoders

Heimel et al I808.08979; Farina, Nakai & DS I808.08992



An autoencoder maps an input into a “latent representation” and then attempts to reconstruct the original input.

The encoding is lossy (“information bottleneck”), so the decoding cannot be perfect.

Some previous approaches:

Aguilar-Saavedra et al, “A generic anti-QCD jet tagger” I709.01087

Collins et al, “CWoLa Hunting” I805.02664

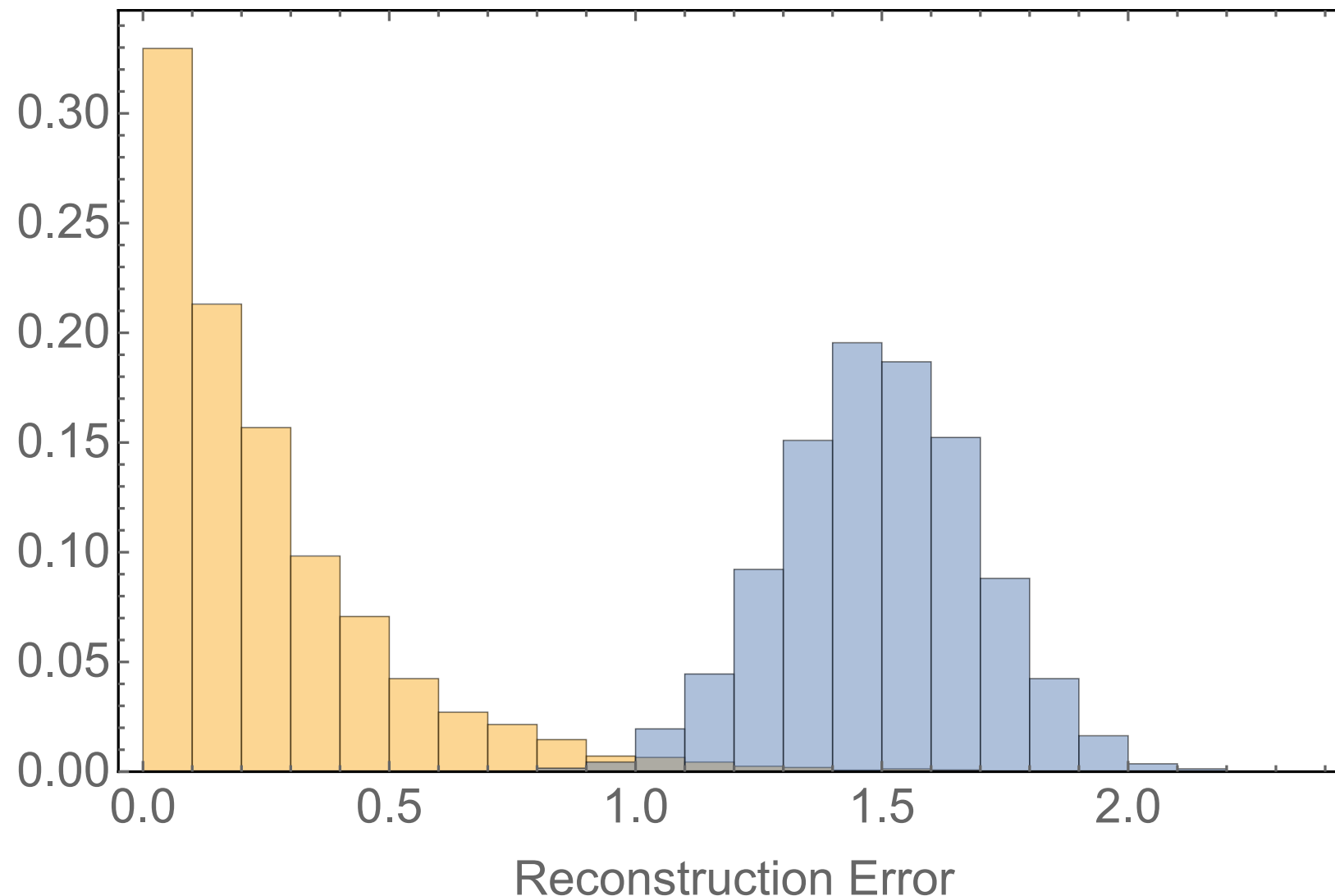
Hajer et al “Novelty Detection Meets Collider Physics” I807.10261

# Deep autoencoders

Heimel et al I808.08979; Farina, Nakai & DS I808.08992

Quantify AE performance using reconstruction error:

$$L = \frac{1}{N} \sum_{i=1}^N (x_i^{in} - x_i^{out})^2$$



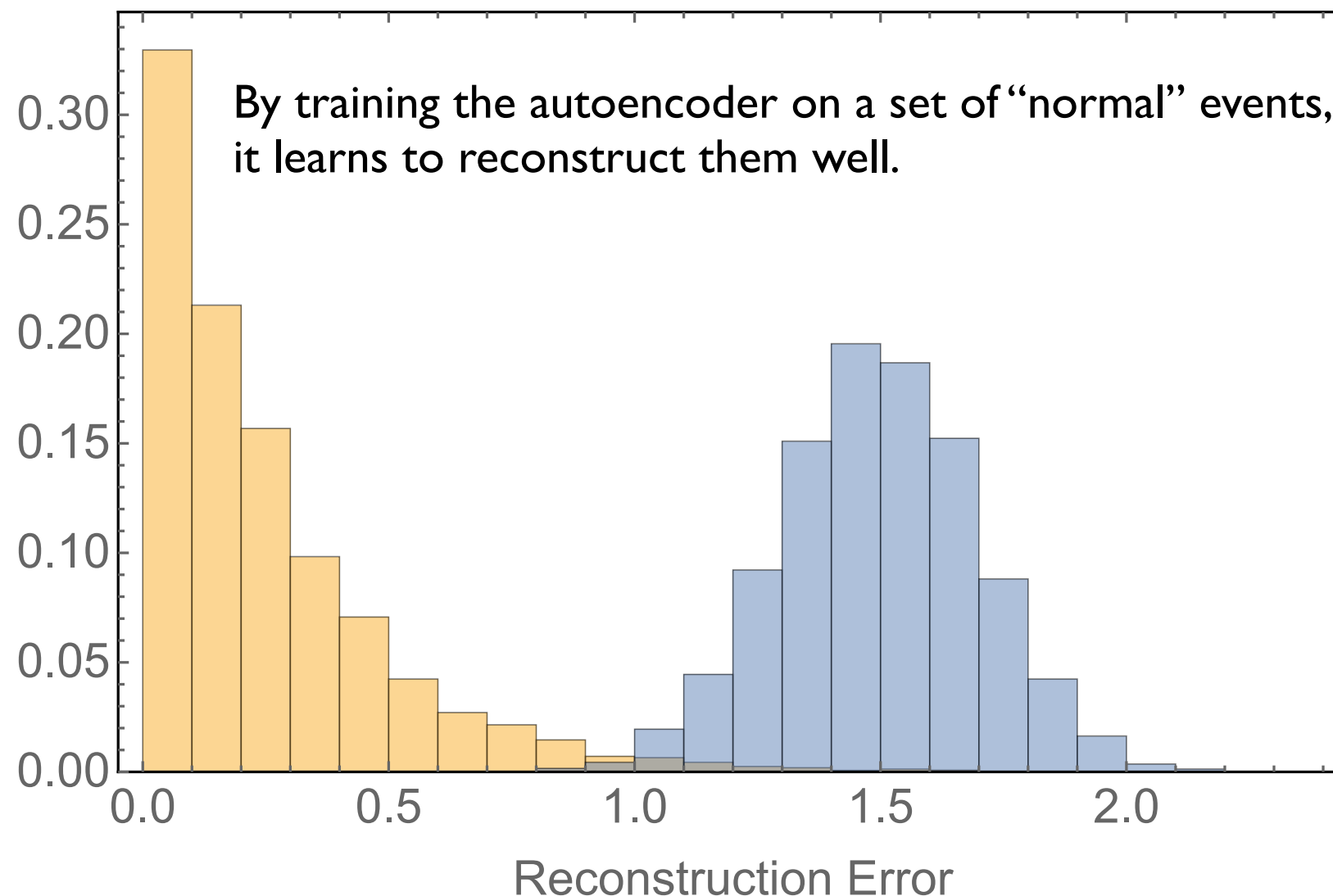


# Deep autoencoders

Heimel et al I808.08979; Farina, Nakai & DS I808.08992

Quantify AE performance using reconstruction error:

$$L = \frac{1}{N} \sum_{i=1}^N (x_i^{in} - x_i^{out})^2$$

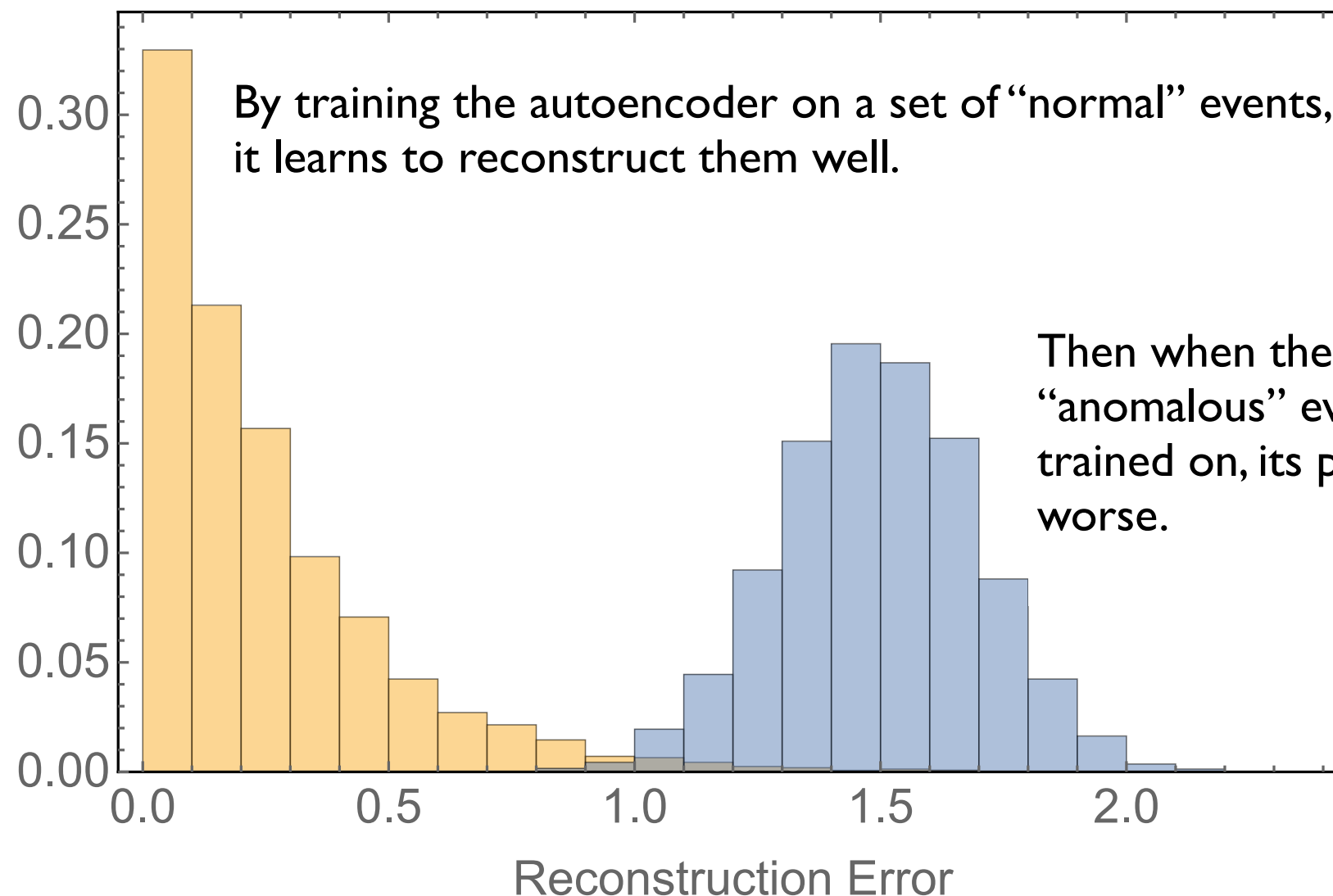


# Deep autoencoders

Heimel et al I808.08979; Farina, Nakai & DS I808.08992

Quantify AE performance using reconstruction error:

$$L = \frac{1}{N} \sum_{i=1}^N (x_i^{in} - x_i^{out})^2$$

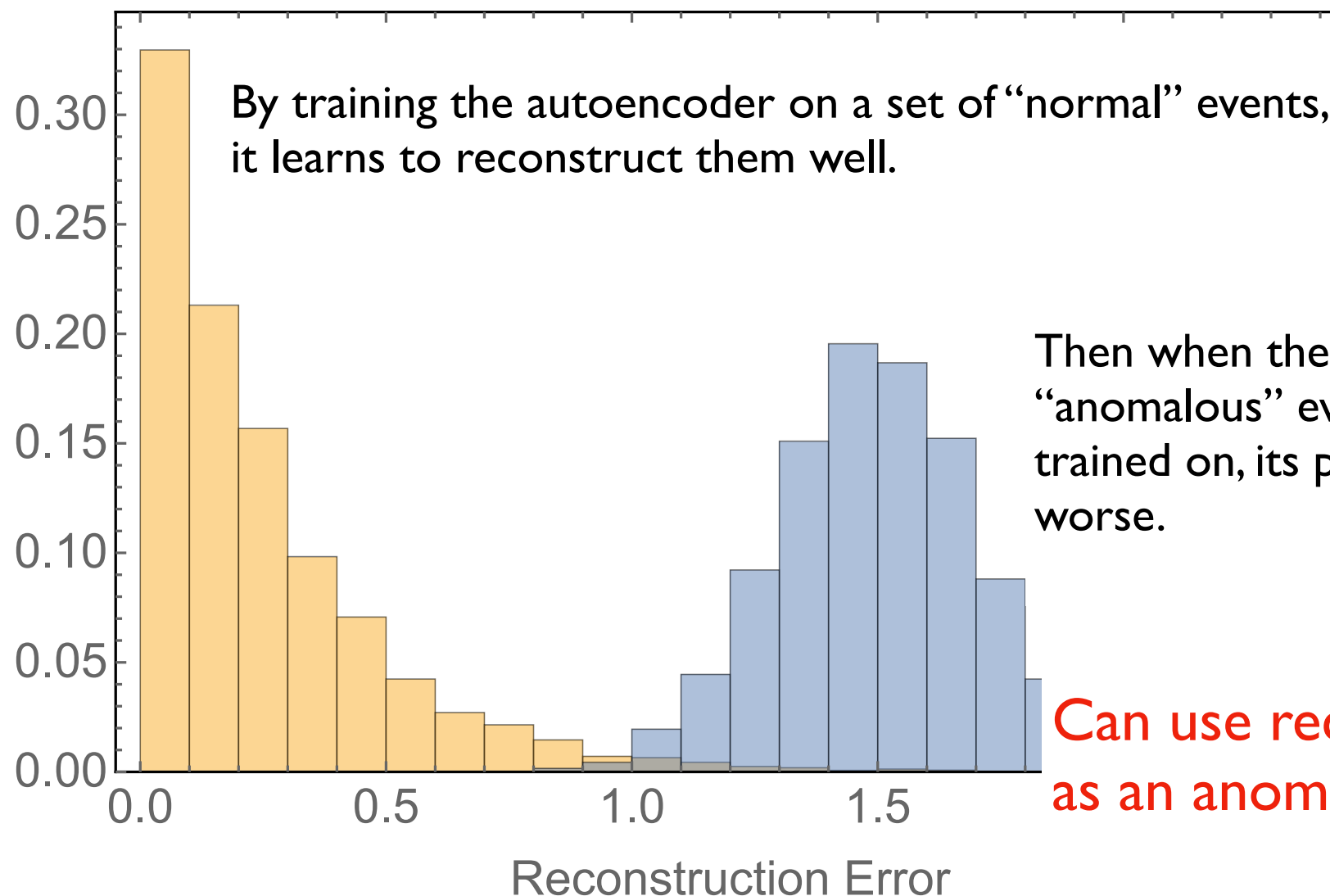


# Deep autoencoders

Heimel et al I808.08979; Farina, Nakai & DS I808.08992

Quantify AE performance using reconstruction error:

$$L = \frac{1}{N} \sum_{i=1}^N (x_i^{in} - x_i^{out})^2$$



By training the autoencoder on a set of “normal” events, it learns to reconstruct them well.

Then when the autoencoder encounters “anomalous” events that it was not trained on, its performance should be worse.

Can use reconstruction error as an anomaly threshold!

# Autoencoder architectures

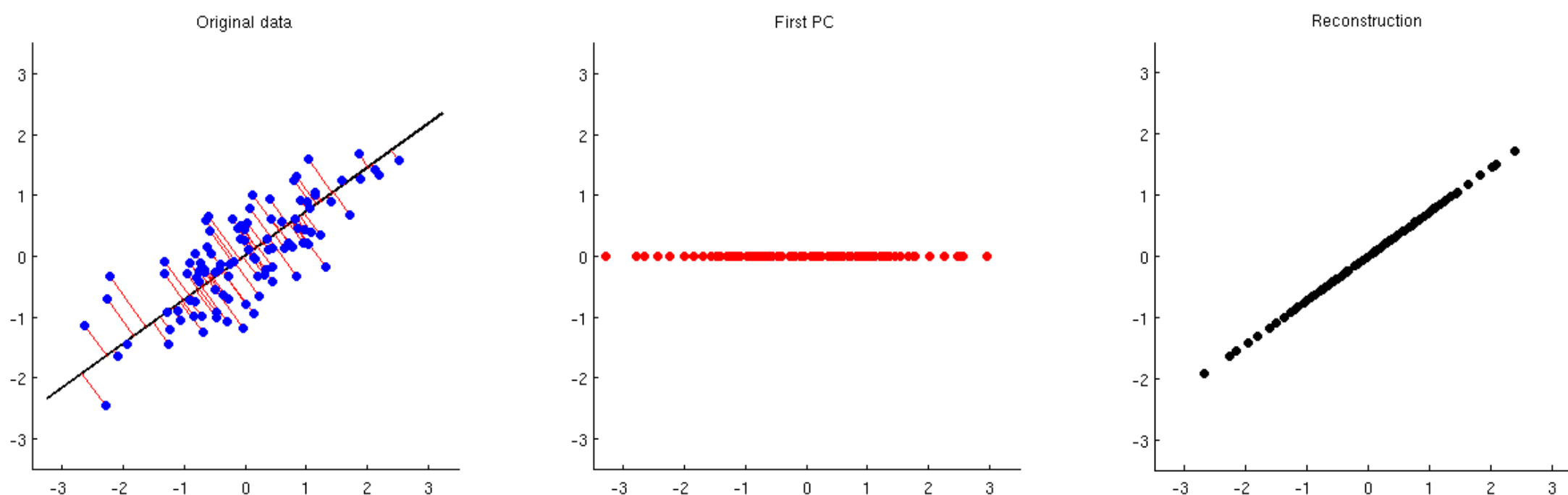
We considered three autoencoder architectures (many more are possible):

- Principal Component Analysis (PCA)
- Dense NN
- Convolutional NN

# Autoencoder architectures

We considered three autoencoder architectures (many more are possible):

- Principal Component Analysis (PCA)



Project onto the first  $d$  PCA eigenvectors

$$z = \mathcal{P}_d x_{in}$$

Inverse transform to reconstruct original input

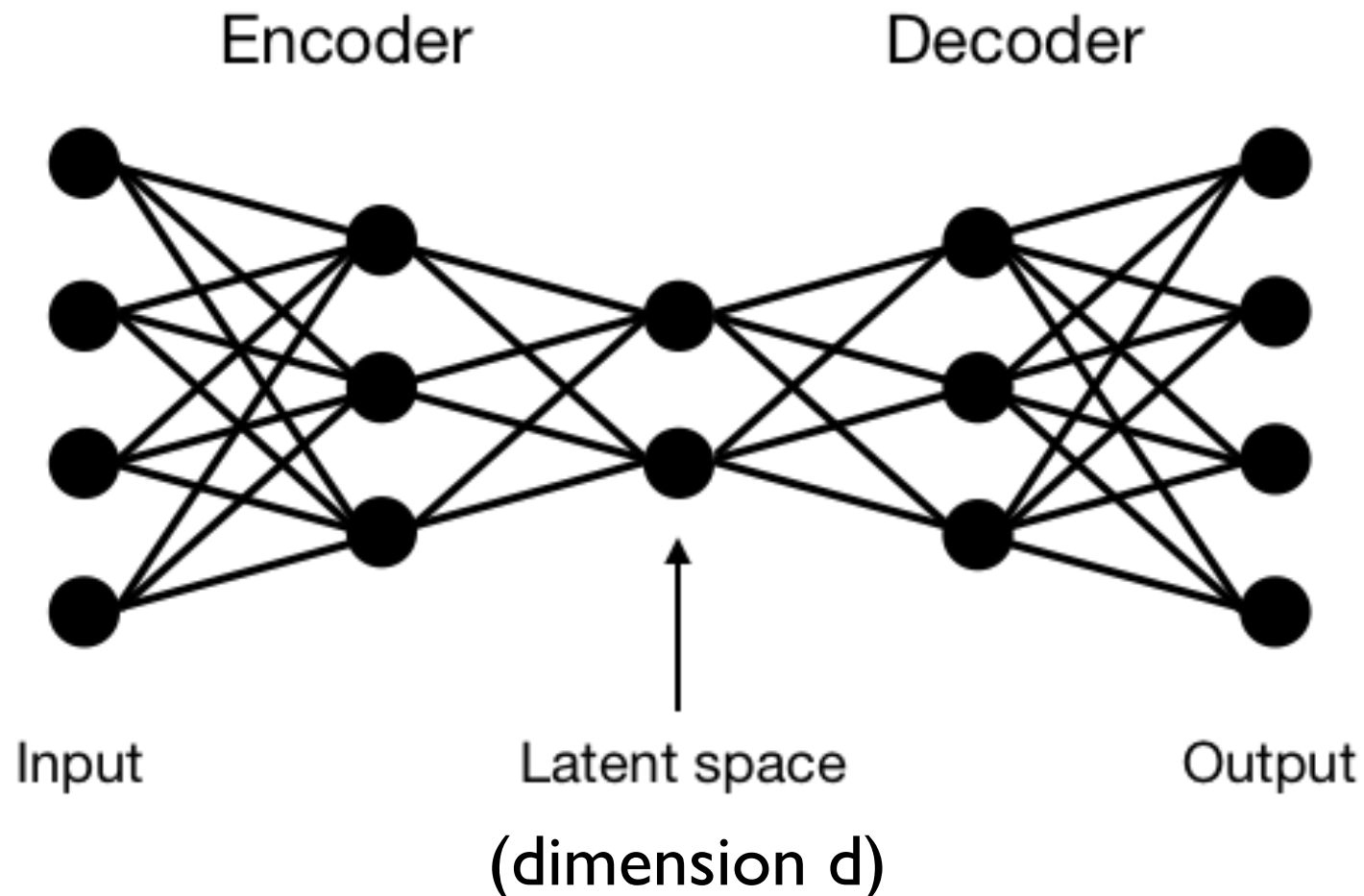
$$x_{out} = \mathcal{P}_d^T z = \mathcal{P}_d^T \mathcal{P}_d x_{in}$$

# Autoencoder architectures

We considered three autoencoder architectures (many more are possible):

- Principal Component Analysis (PCA)
- Dense NN
- Convoluti

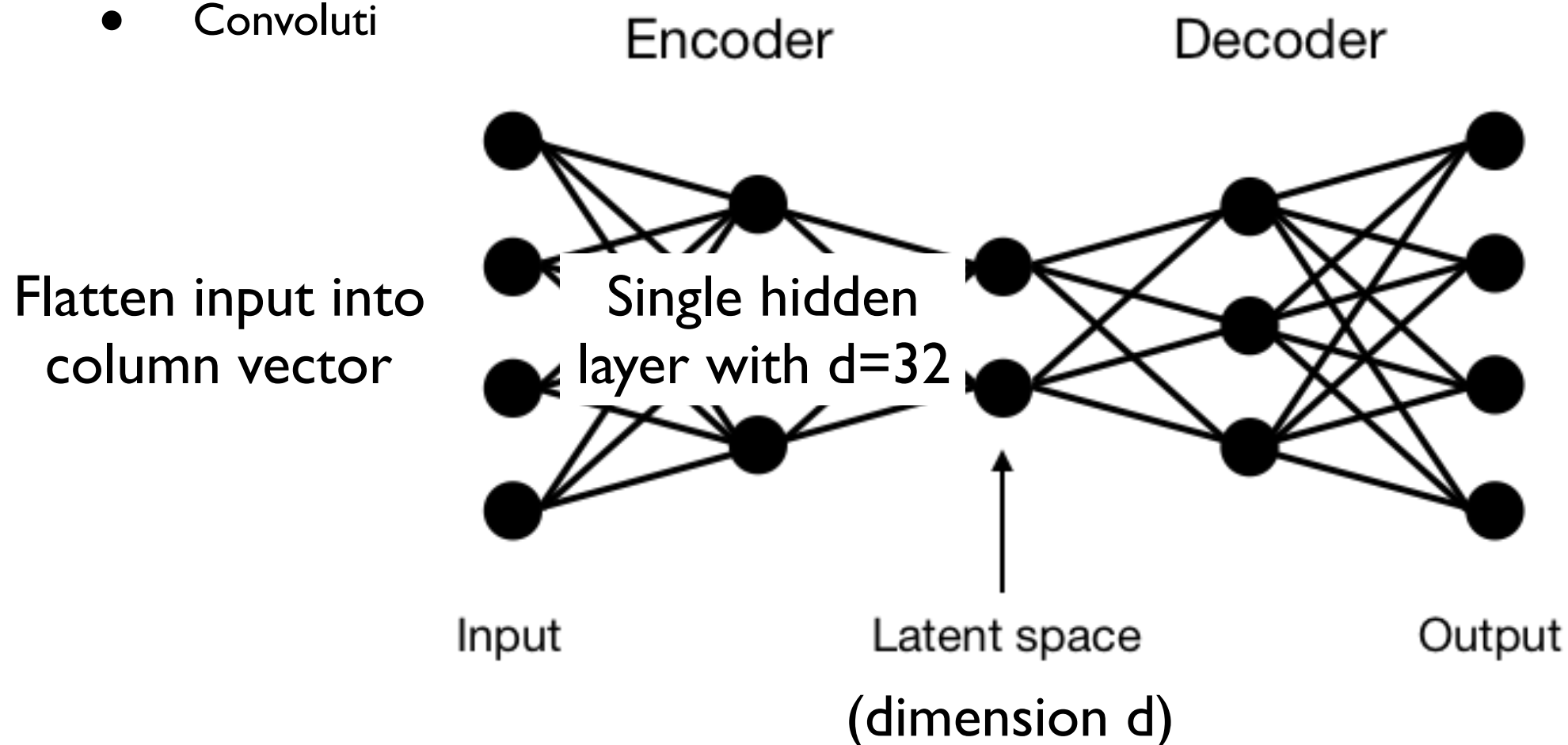
Flatten input into  
column vector



# Autoencoder architectures

We considered three autoencoder architectures (many more are possible):

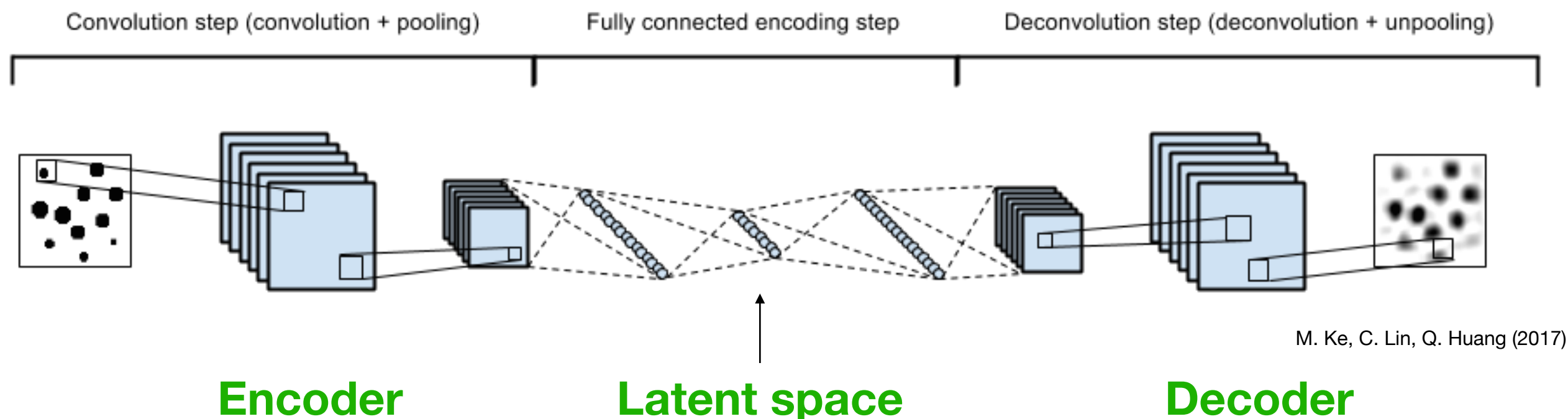
- Principal Component Analysis (PCA)
- Dense NN
- Convoluti



# Autoencoder architectures

We considered three autoencoder architectures (many more are possible):

- Principal Component Analysis (PCA)
- Dense NN
- Convolutional NN



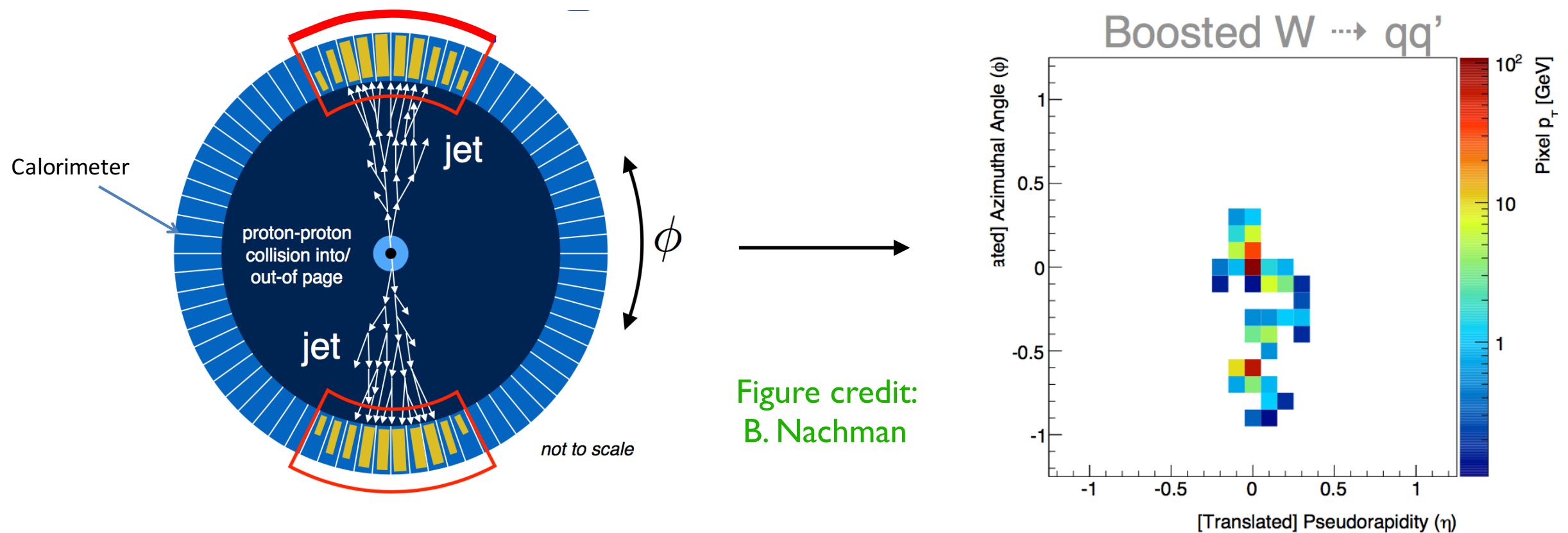
M. Ke, C. Lin, Q. Huang (2017)

I28C3-MP2-I28C3-MP2-I28C3-32N-6N-32N-I2800N-I28C3-US2-I28C3-US2-I C3



# Jet Images

We focused on jet images as inputs to the autoencoders



Can think of a jet as an **image in eta and phi**, with

- Pixelation provided by calorimeter towers
- Pixel intensity =  $p_T$  recorded by each tower

# Sample definitions

We took

- QCD jets as background
- tops and 400 GeV gluinos as signals.

Same jet and image specifications as  
for recent top tagging study  
(Macaluso & DS 1803.00107)

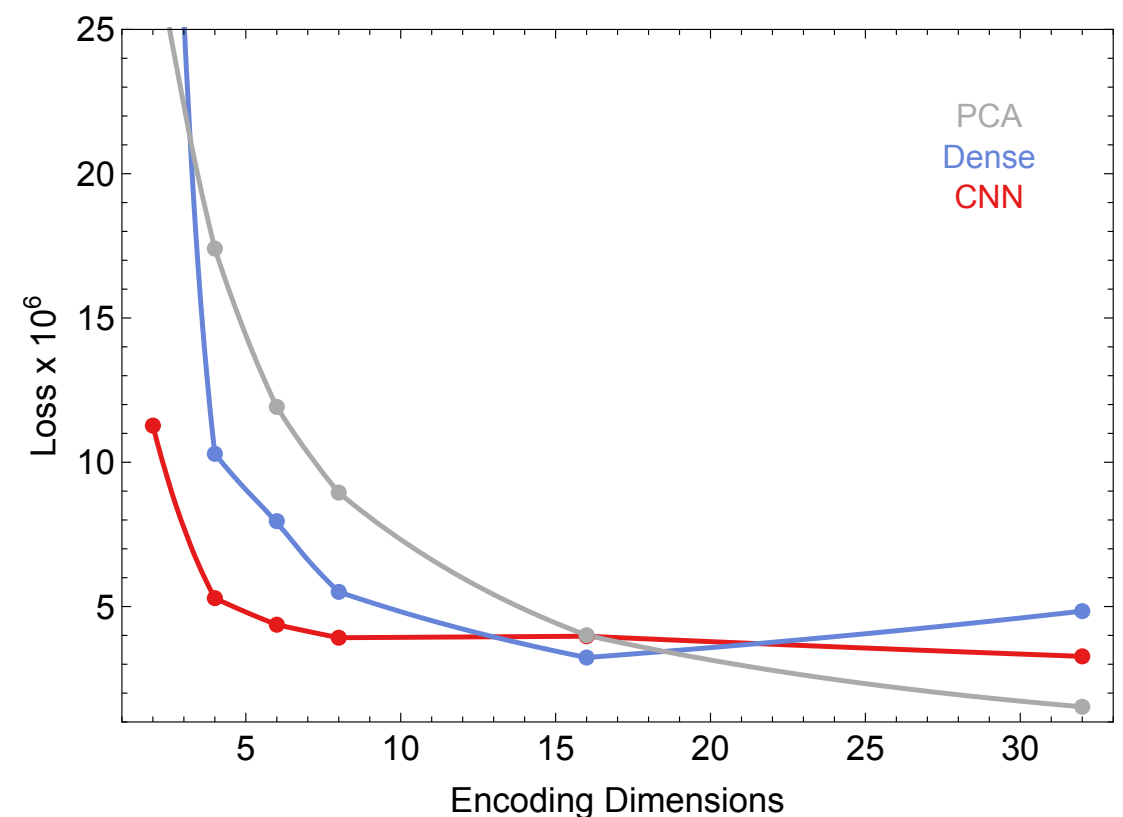
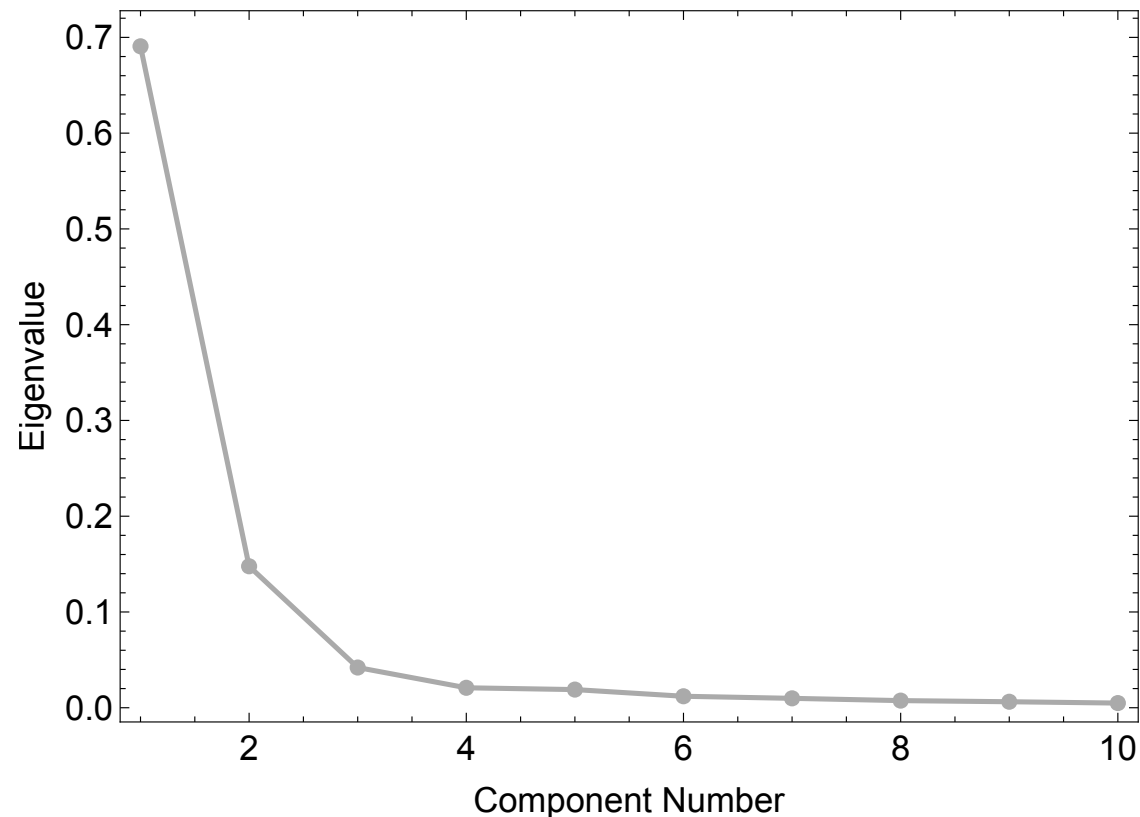
	CMS
Jet sample	13 TeV $p_T \in (800, 900)$ GeV, $ \eta  < 1$ PYTHIA 8 and DELPHES particle-flow match: $\Delta R(t, j) < 0.6$ merge: $\Delta R(t, q) < 0.6$ 1.2M + 1.2M
Image	$37 \times 37$ $\Delta\eta = \Delta\phi = 3.2$
Colors	$(p_T^{neutral}, p_T^{track}, N_{track}, N_{muon})$

# Choosing the latent dimension

$d$  too large  $\rightarrow$  autoencoder becomes identity transform

$d$  too small  $\rightarrow$  autoencoder cannot learn all the features

Should choose the latent dimension in an unsupervised manner  
(ie without optimizing on a specific signal)

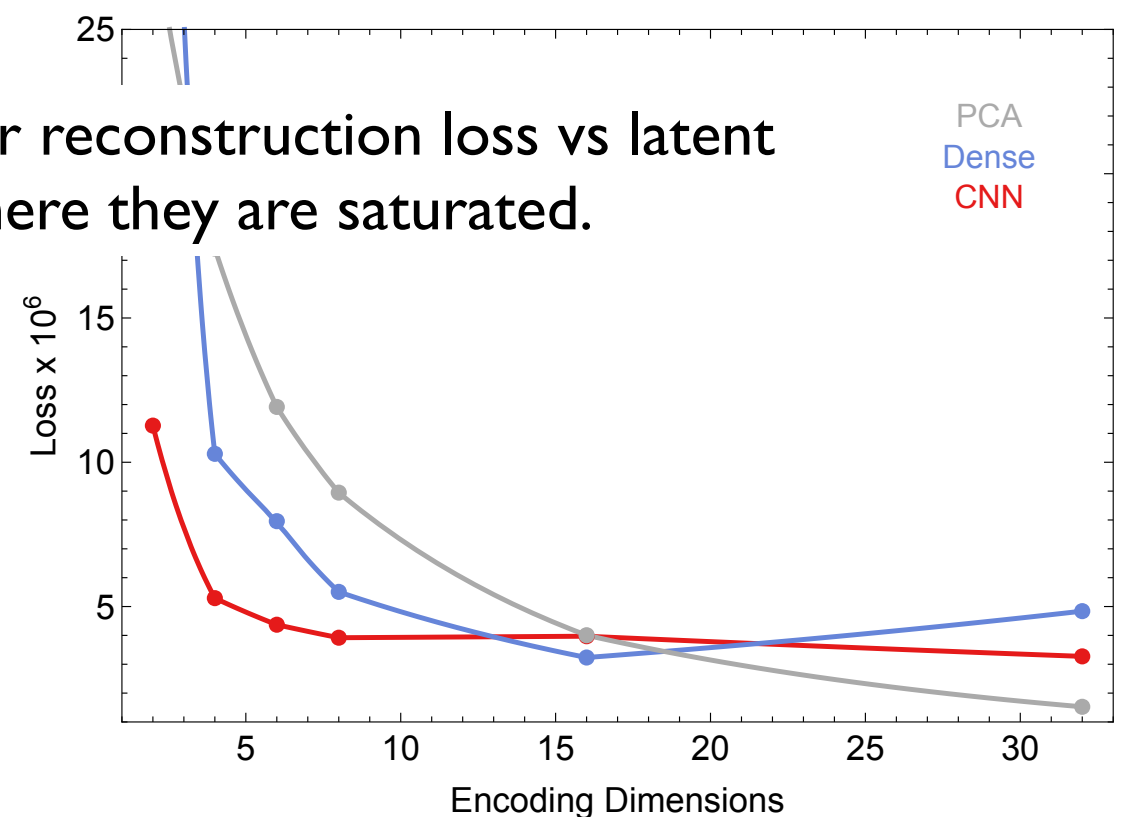
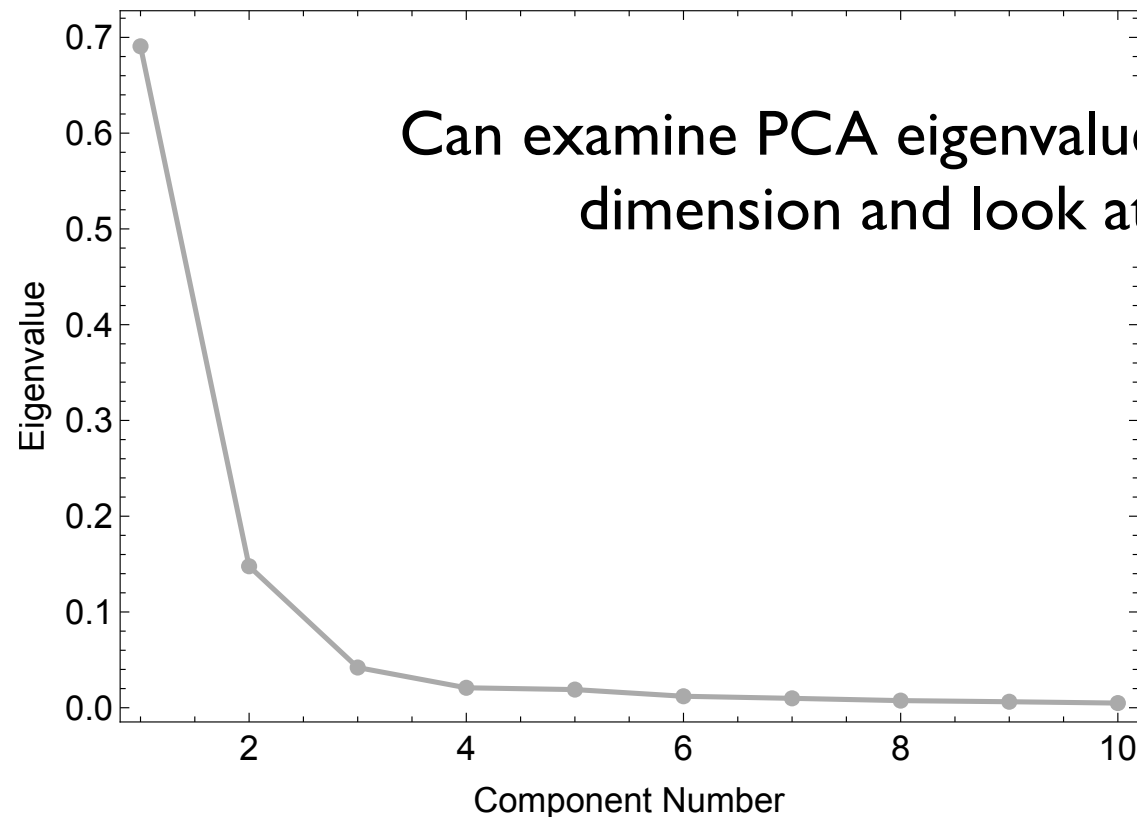


# Choosing the latent dimension

$d$  too large  $\rightarrow$  autoencoder becomes identity transform

$d$  too small  $\rightarrow$  autoencoder cannot learn all the features

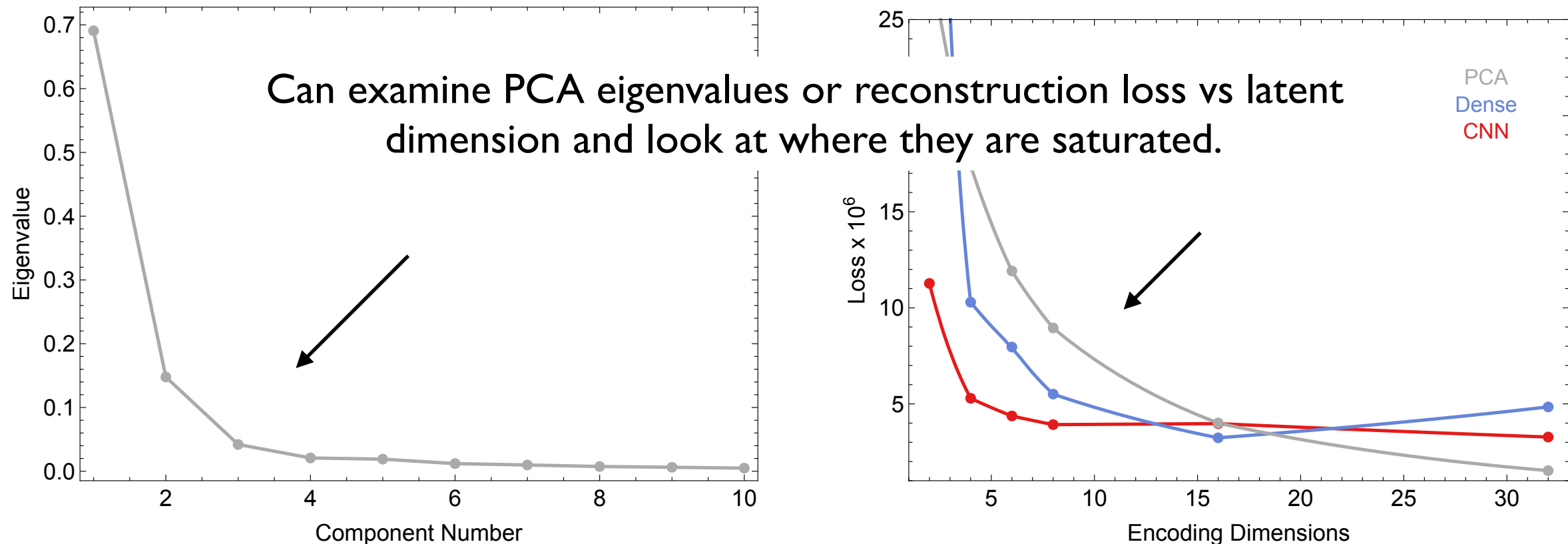
Should choose the latent dimension in an unsupervised manner  
(ie without optimizing on a specific signal)



# Choosing the latent dimension

d too large → autoencoder becomes identity transform  
d too small → autoencoder cannot learn all the features

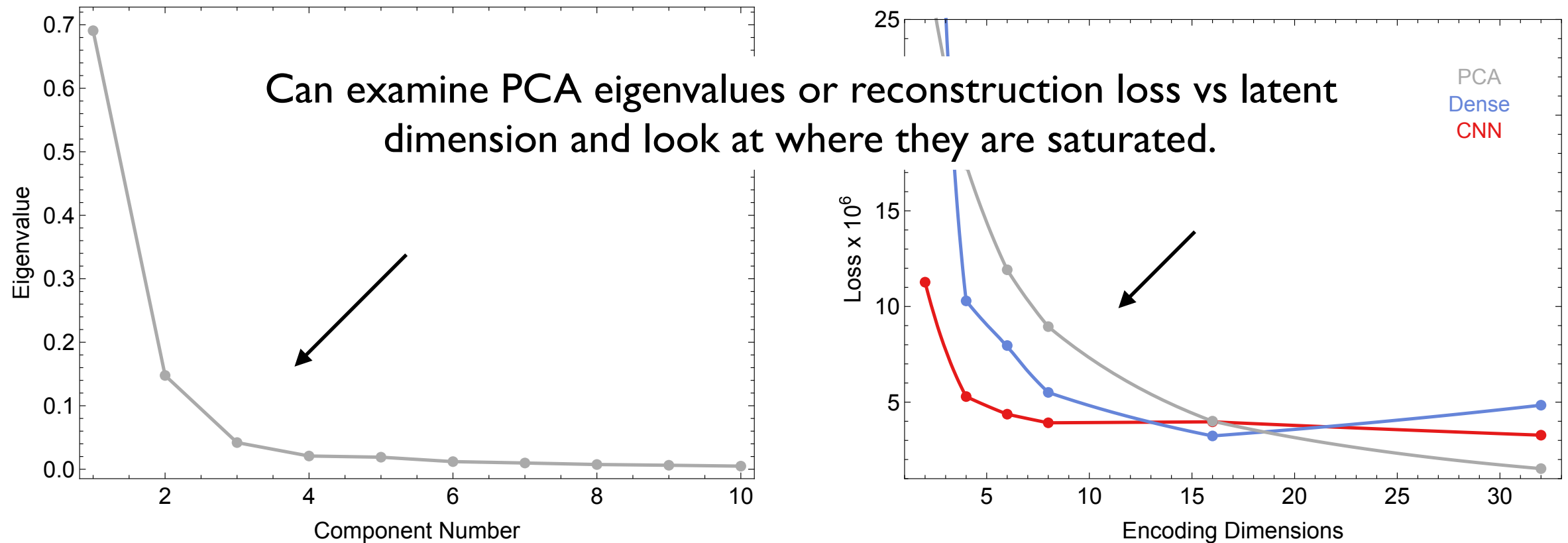
Should choose the latent dimension in an unsupervised manner  
(ie without optimizing on a specific signal)



# Choosing the latent dimension

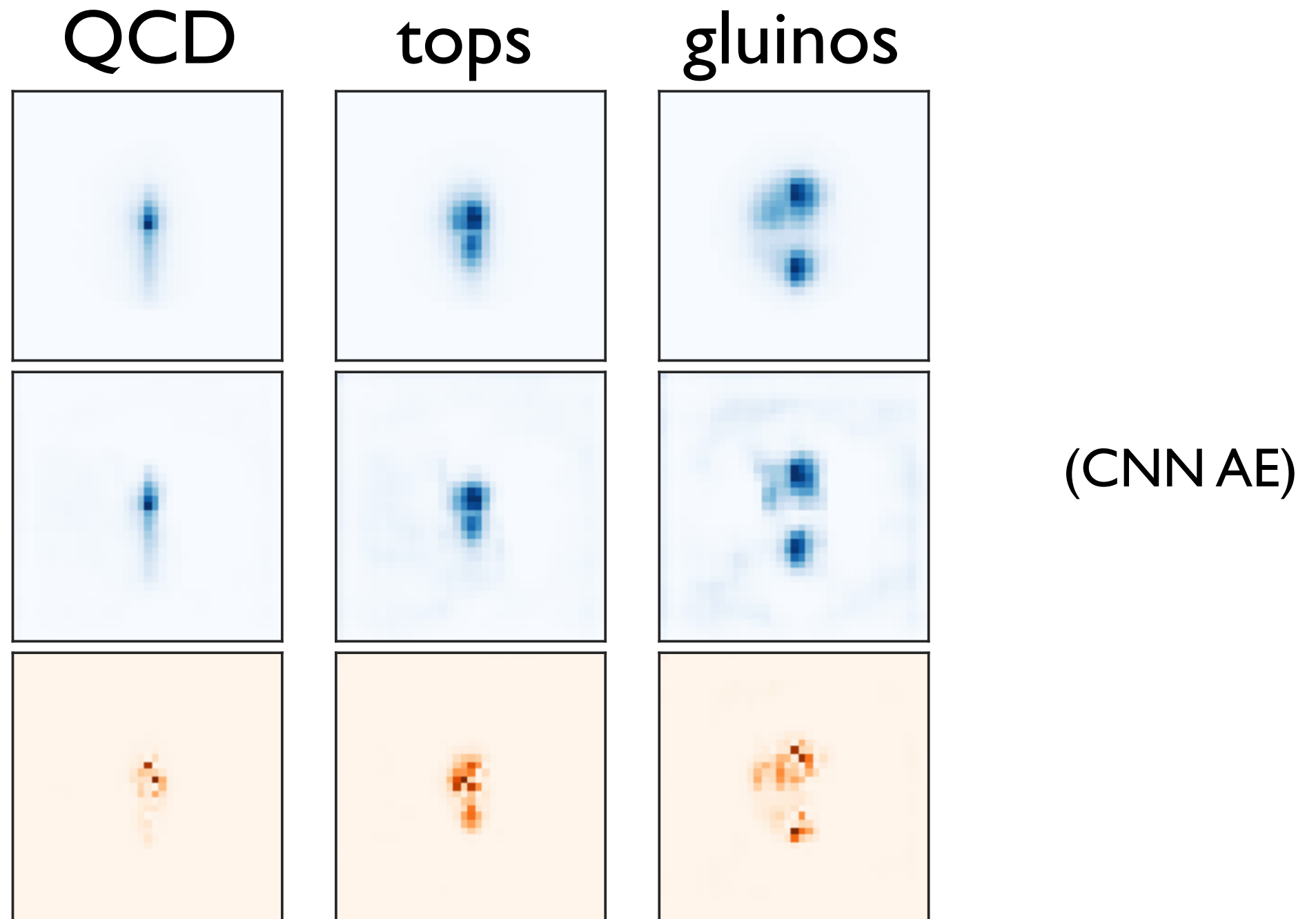
$d$  too large  $\rightarrow$  autoencoder becomes identity transform  
 $d$  too small  $\rightarrow$  autoencoder cannot learn all the features

Should choose the latent dimension in an unsupervised manner  
(ie without optimizing on a specific signal)



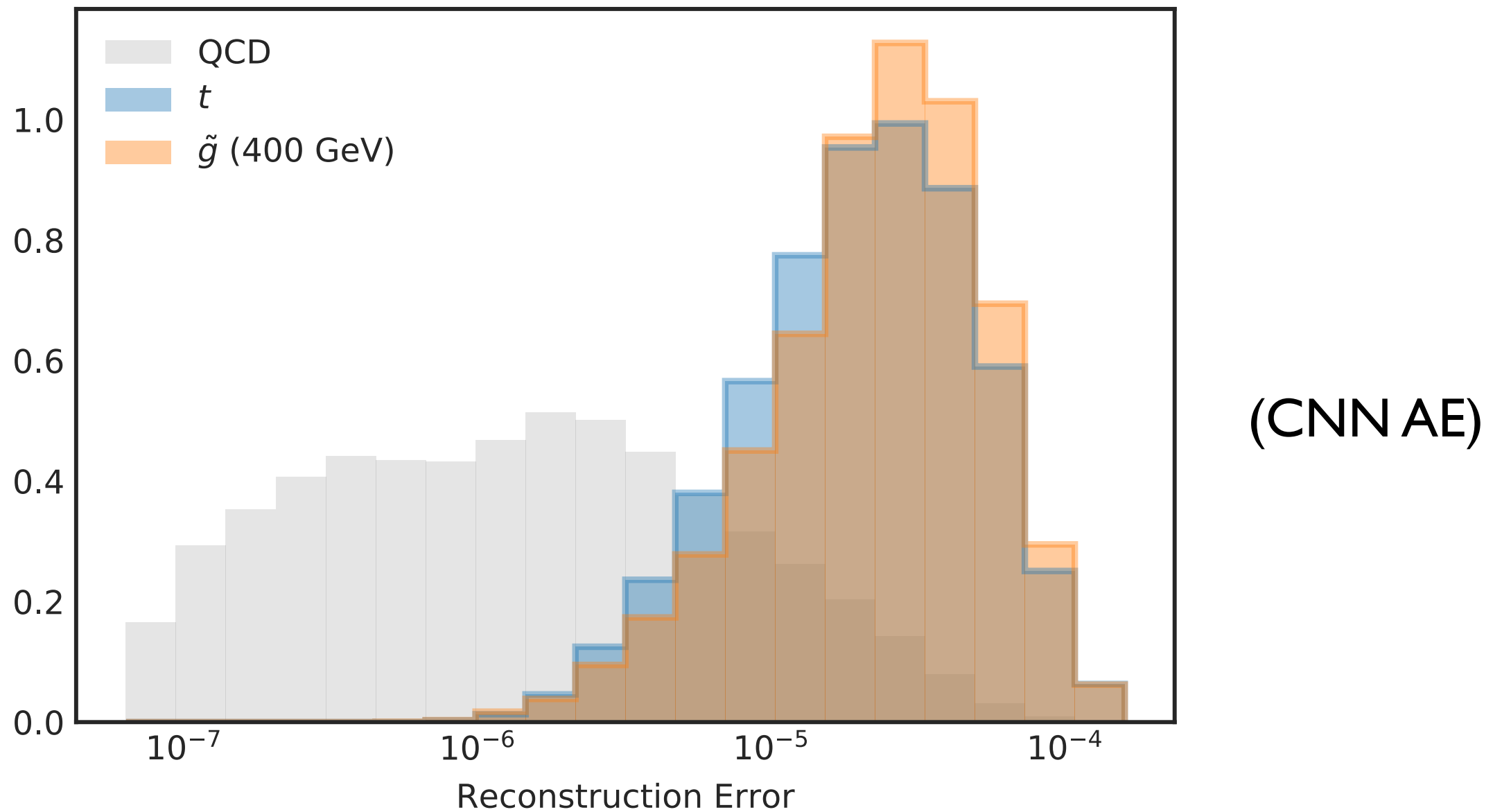
We chose  $d=6$

Performance should be worse on “anomalous” events that autoencoder was not trained on.



The algorithm works when trained on QCD backgrounds!

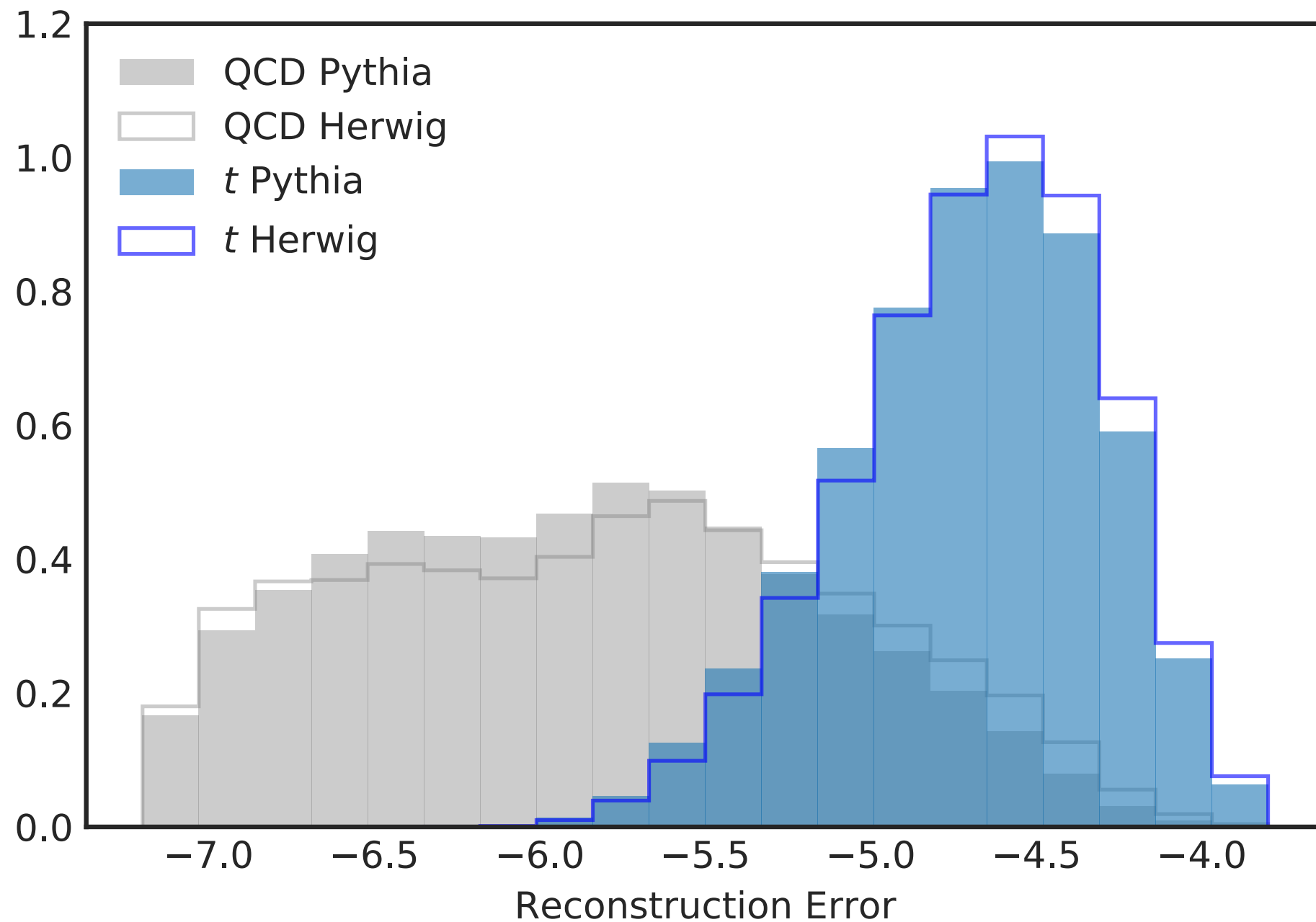
Can use reconstruction error as an anomaly threshold.



The algorithm works when trained on QCD backgrounds!



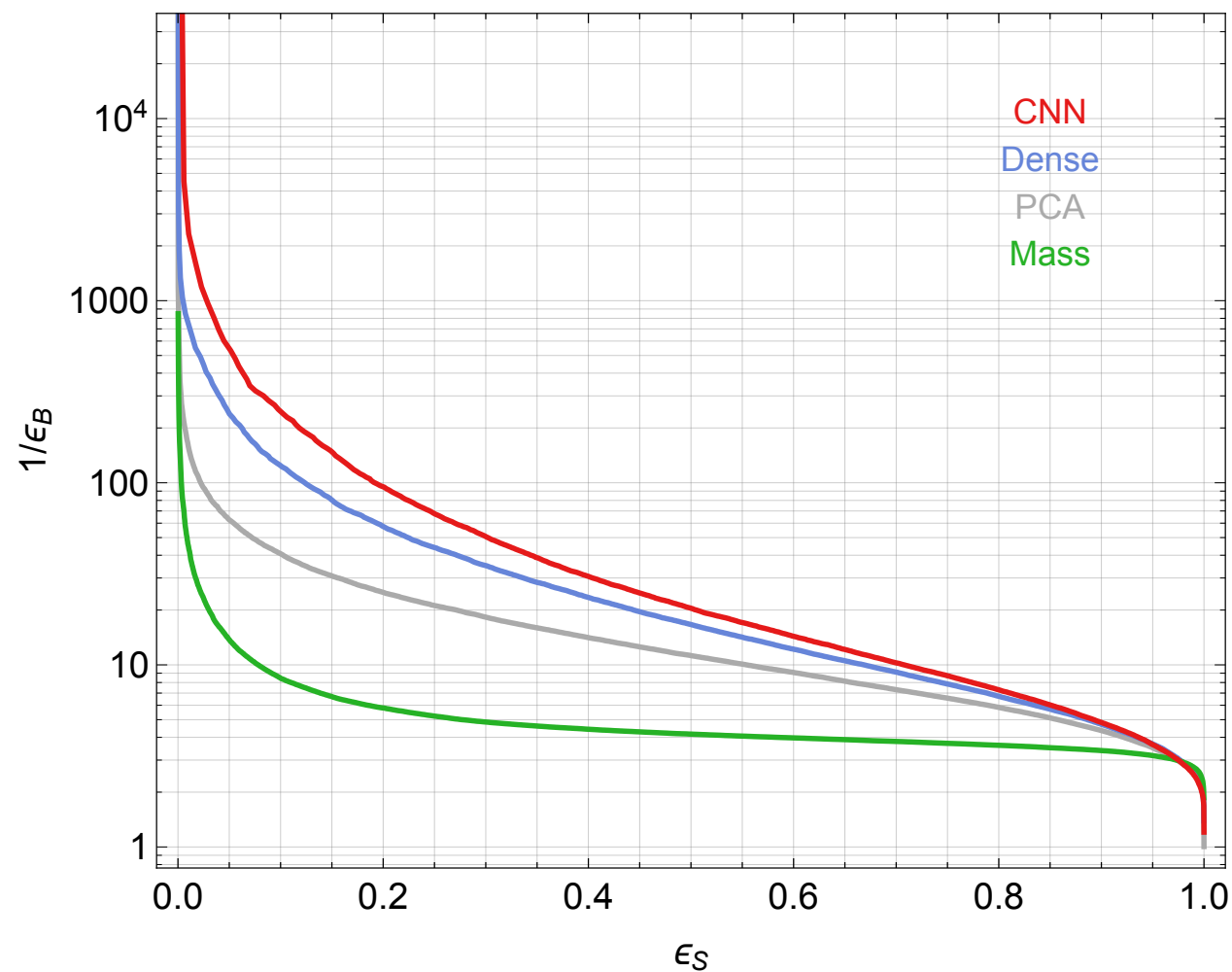
# Robustness with other Monte Carlo



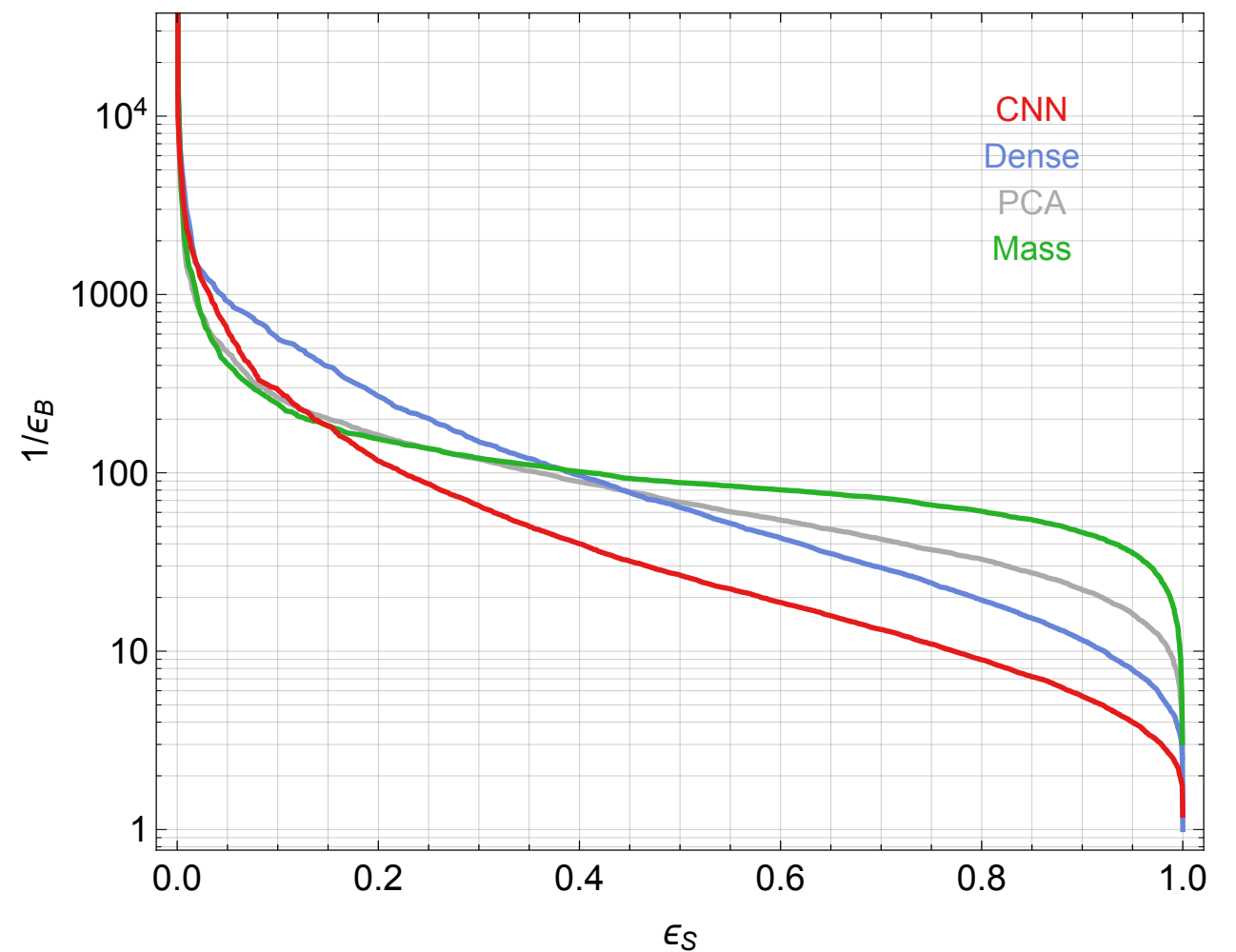
# Comparison vs jet mass

How do our fancy autoencoders compare against a simpler anomaly detection method: jet mass bump hunt?

Tops



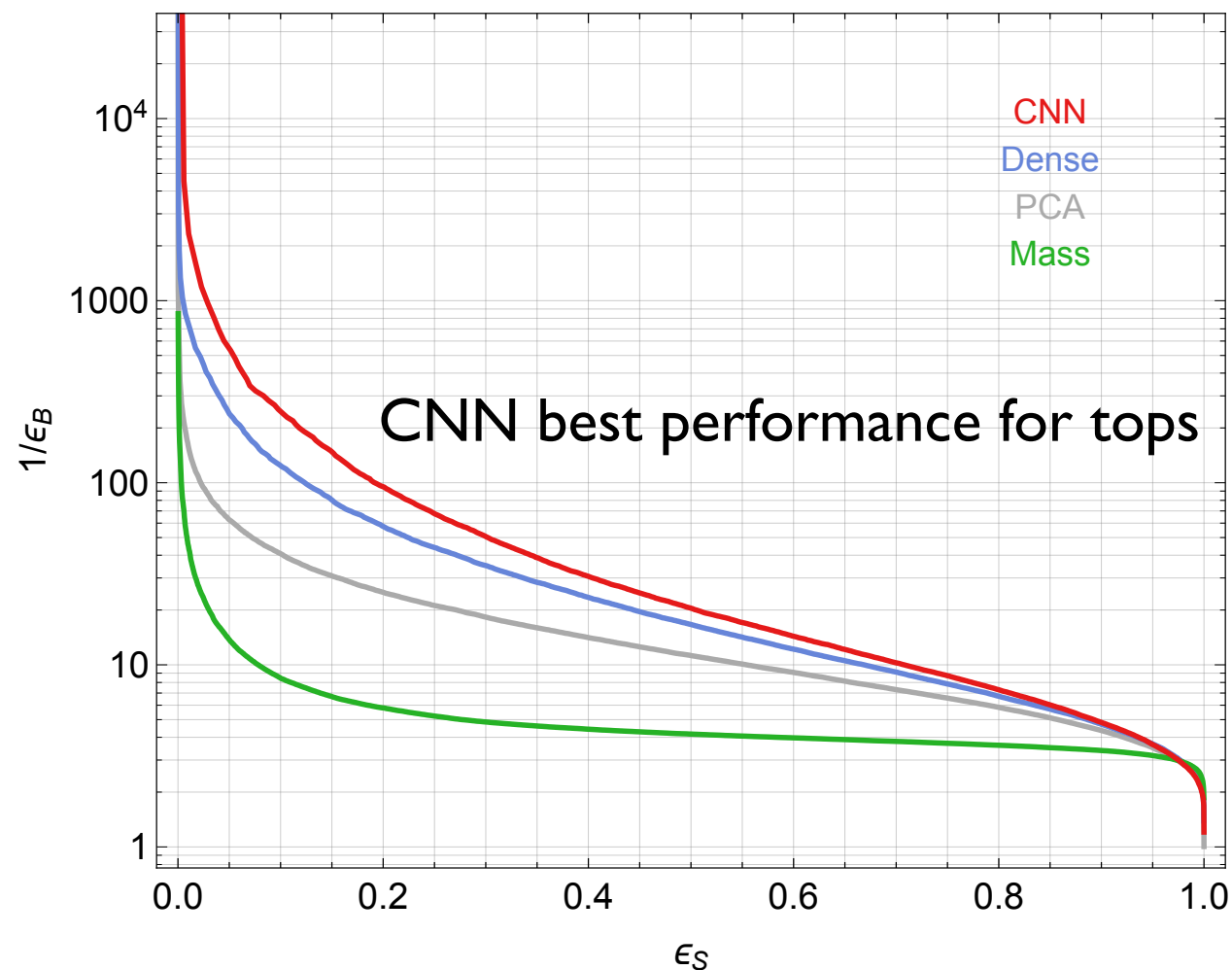
400 GeV gluinos



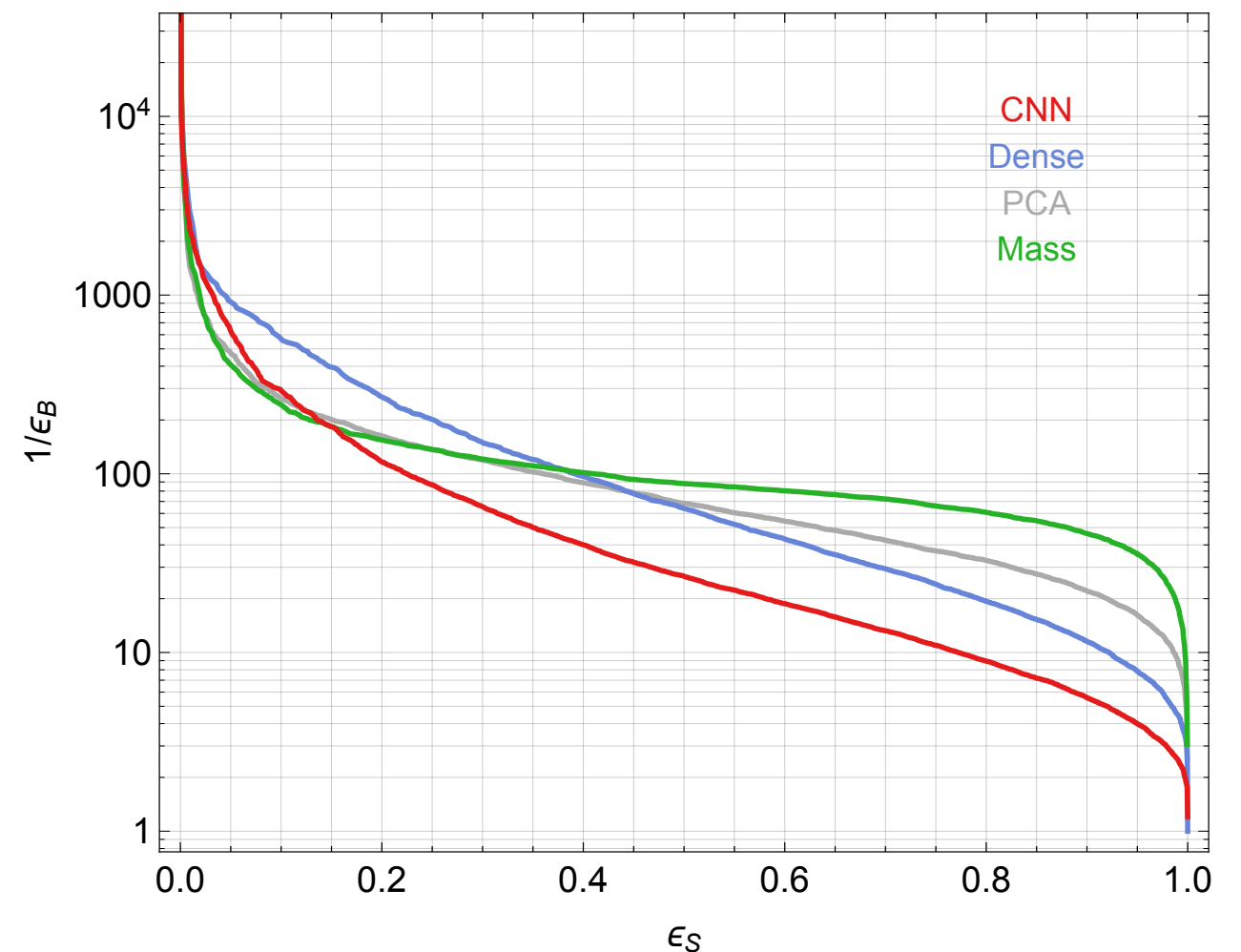
# Comparison vs jet mass

How do our fancy autoencoders compare against a simpler anomaly detection method: jet mass bump hunt?

Tops



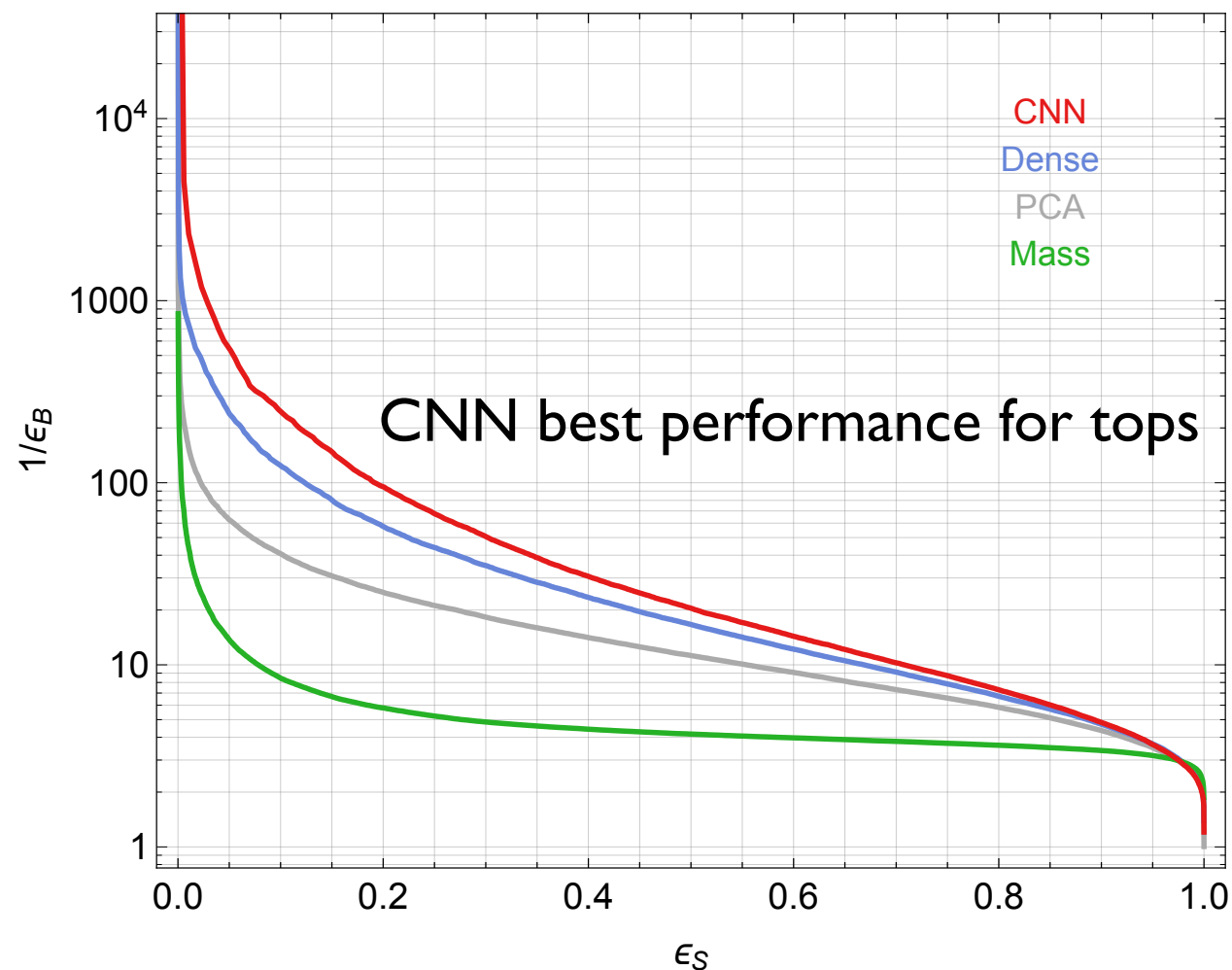
400 GeV gluinos



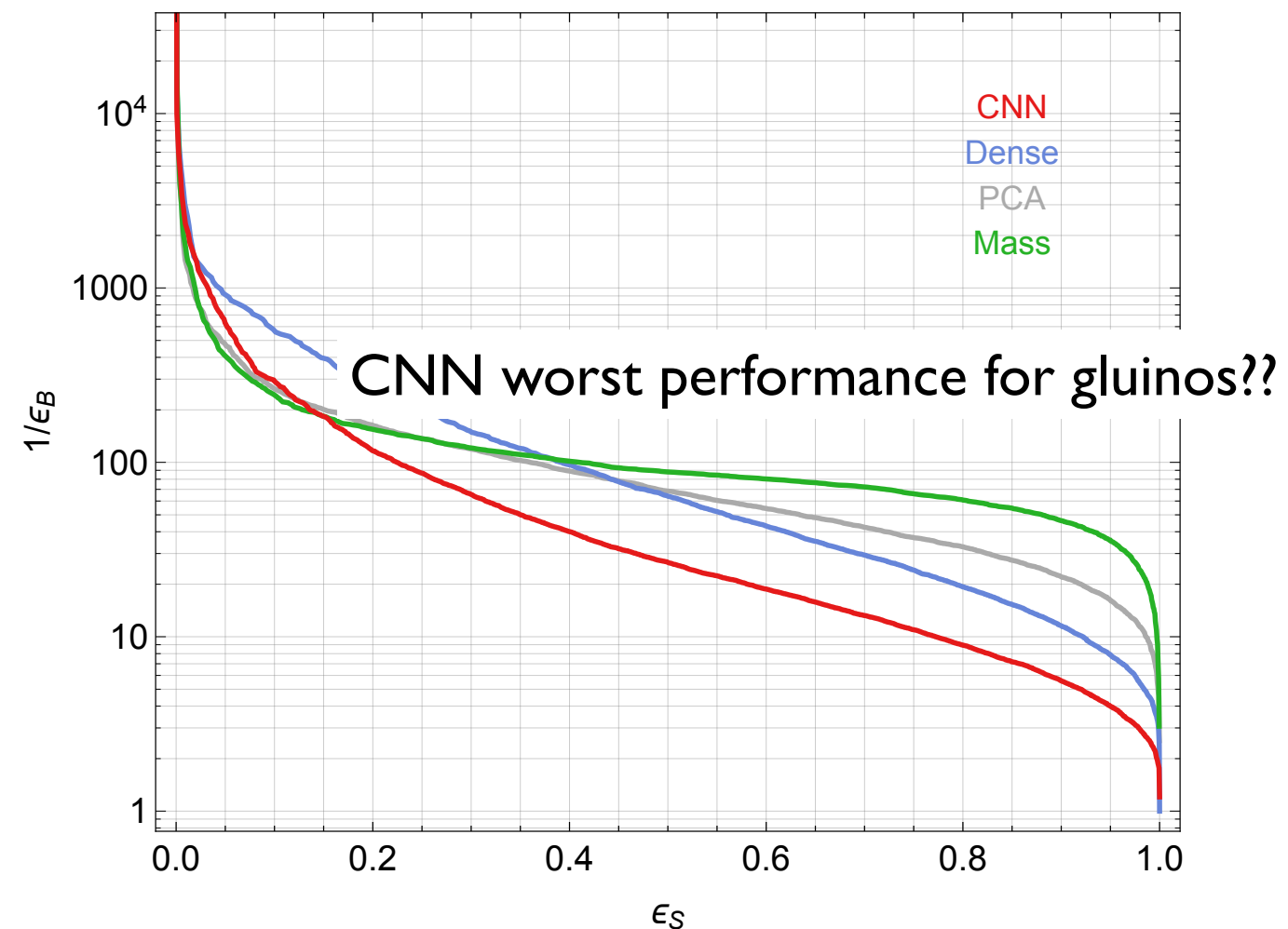
# Comparison vs jet mass

How do our fancy autoencoders compare against a simpler anomaly detection method: jet mass bump hunt?

Tops



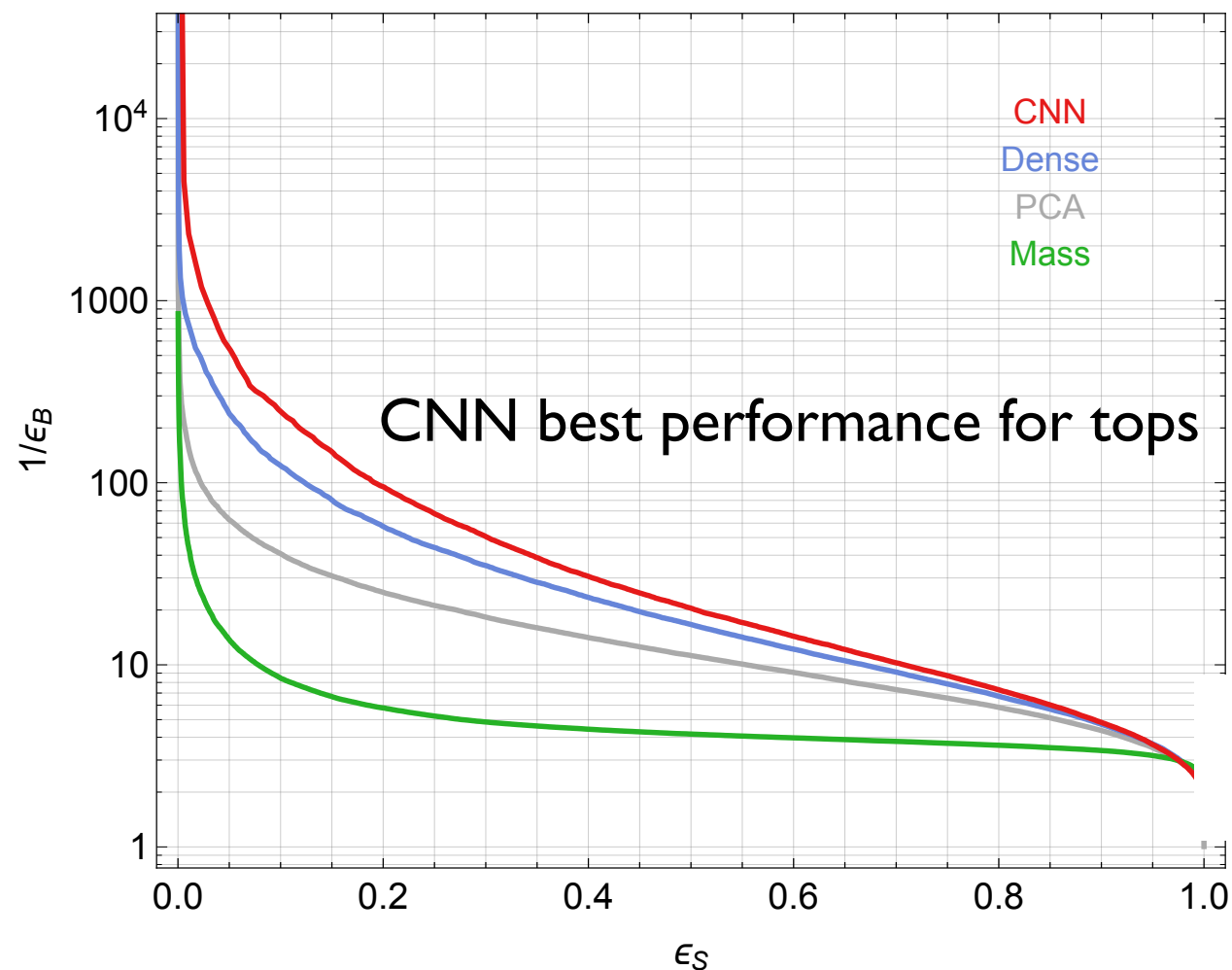
400 GeV gluinos



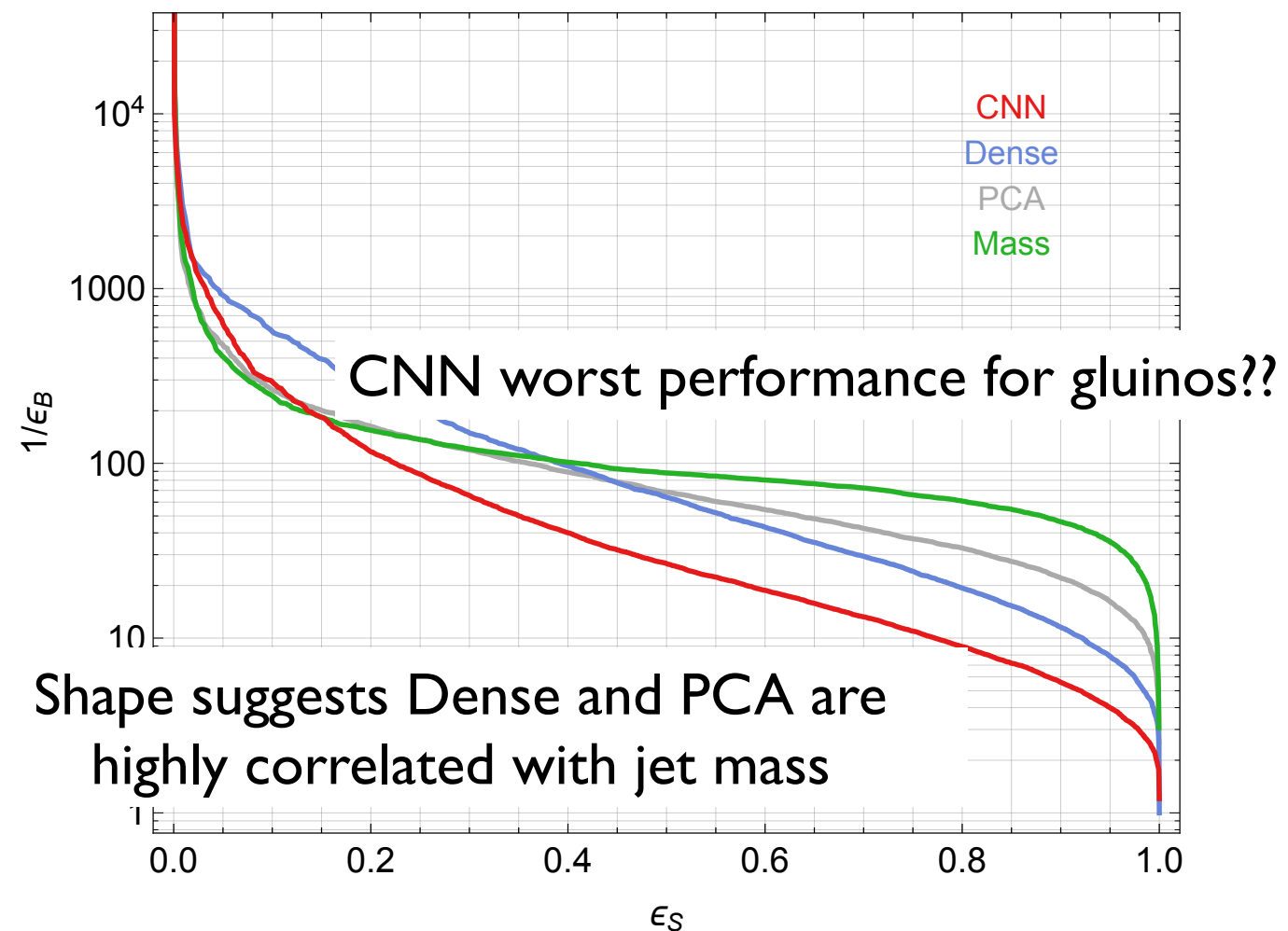
# Comparison vs jet mass

How do our fancy autoencoders compare against a simpler anomaly detection method: jet mass bump hunt?

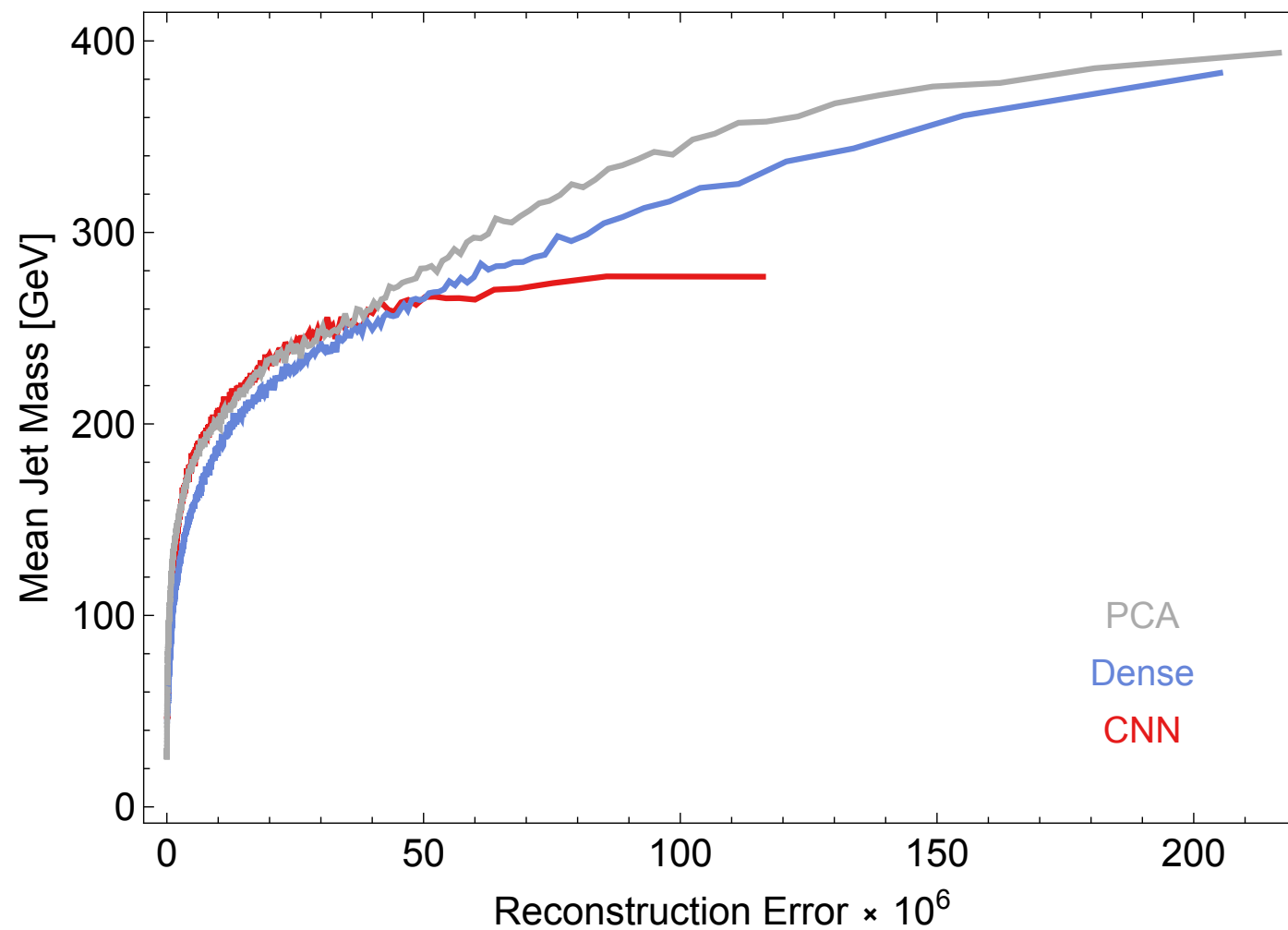
Tops



400 GeV gluinos



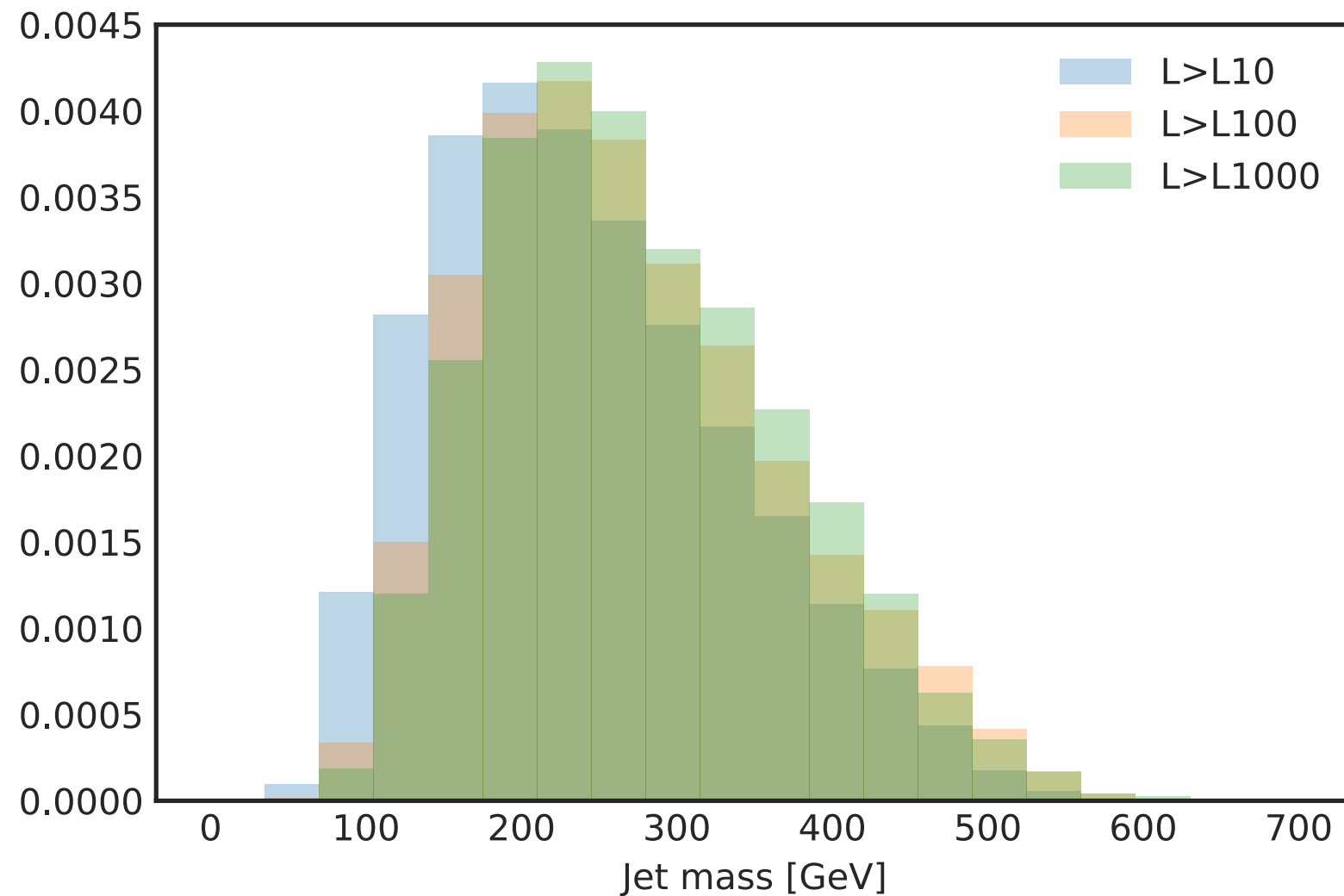
# Correlation with jet mass



Indeed, this is confirmed by looking at mean jet mass in bins of reconstruction error for the QCD background.

CNN is no longer correlated with jet mass for  $m \gtrsim 250$  GeV

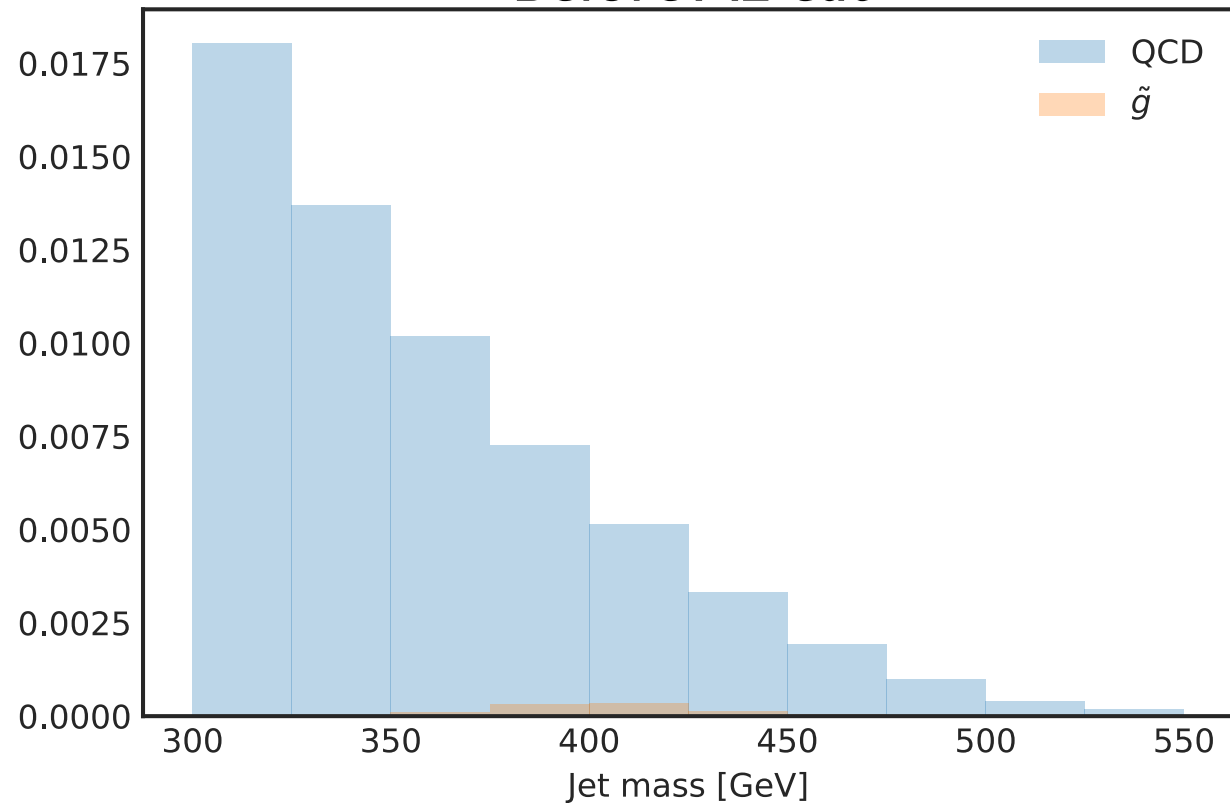
# Correlation with jet mass



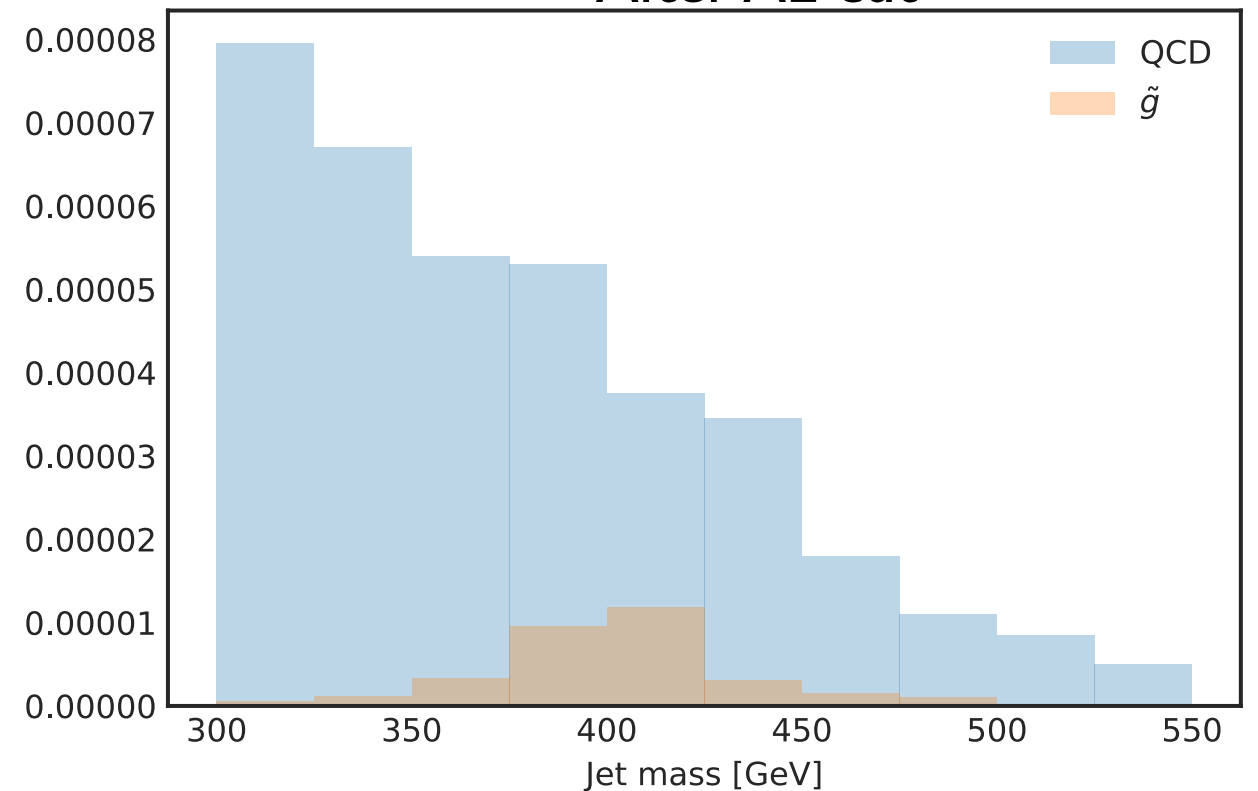
The QCD jet mass distribution is stable against harder cuts on the reconstruction error, for the CNN autoencoder.

# Bump hunt with deep autoencoder

Before AE cut



After AE cut



Can combine the CNN autoencoder with a bump hunt in jet mass.

Use the AE to first clean away “boring” QCD jets.

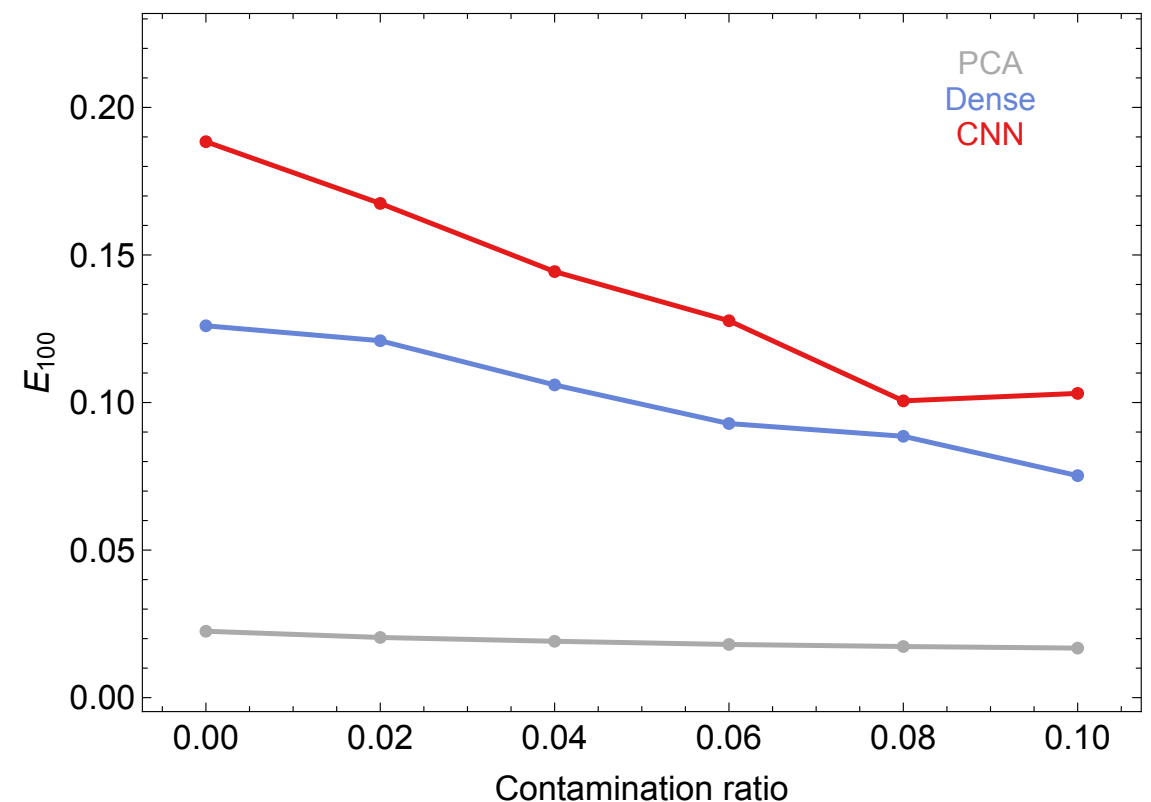
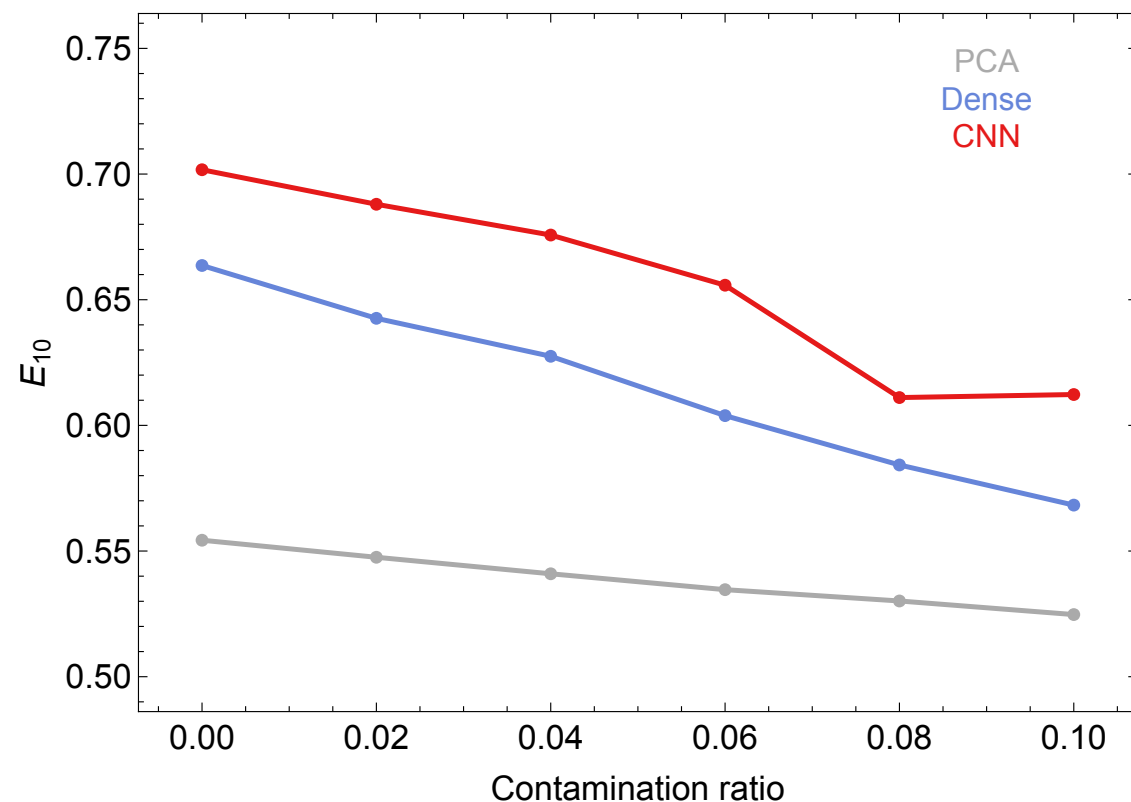
Improve S/B by a lot!



# Fully unsupervised learning

Train on sample of QCD background “contaminated” with a small fraction of signal.

Representative of actual data.

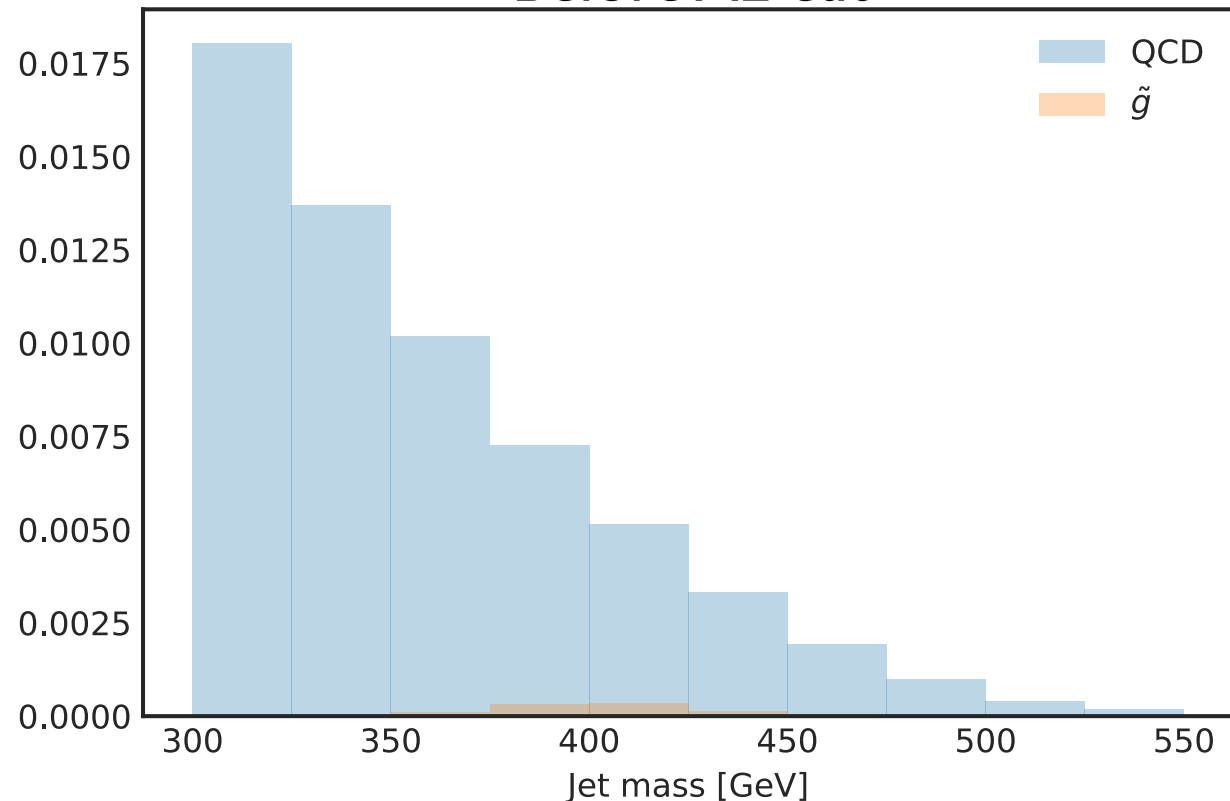


( $E_x$  = signal efficiency at bg rejection =  $x$ )

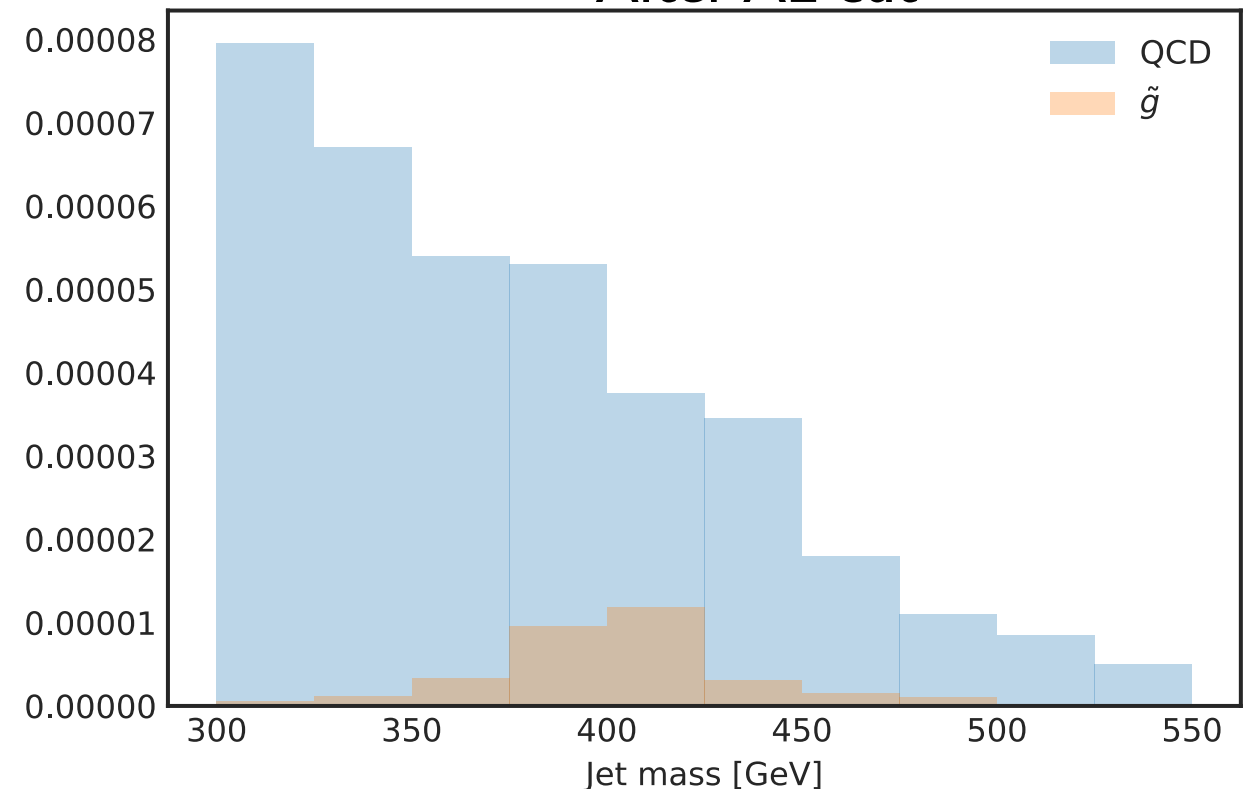
Performance of AE surprisingly robust even up to 10% contamination!

# Bump hunt with deep autoencoder

Before AE cut



After AE cut



Can train directly on data that contains 400 GeV gluinos,  
and still enhance the bump hunt.

Could really discover new physics this way!

**Thank you for your attention!**

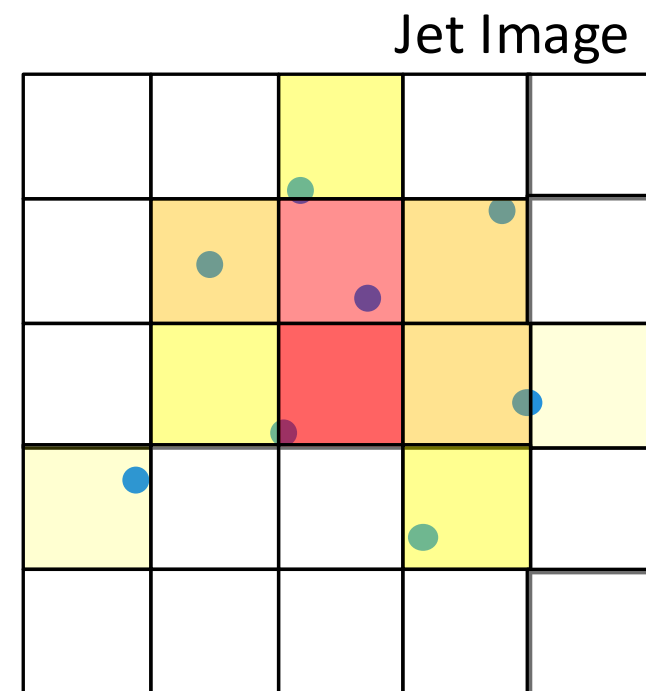
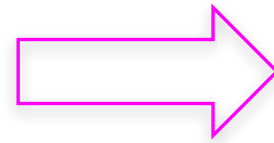
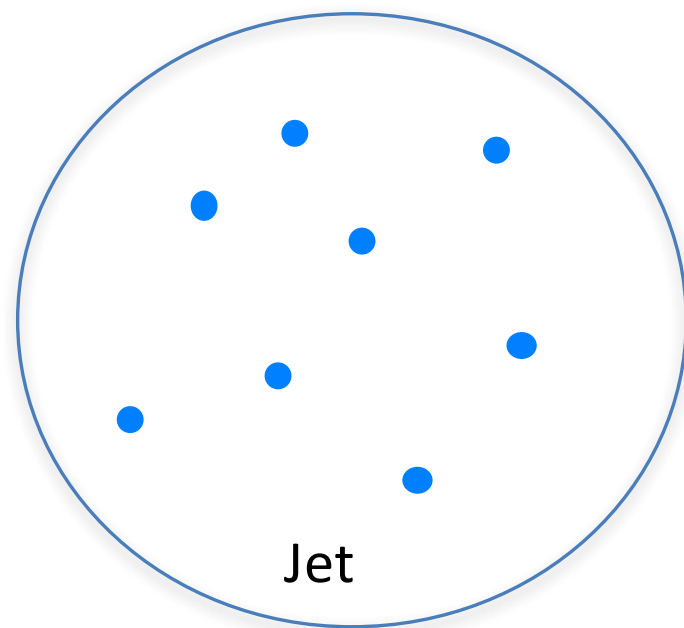
**Backup material**

# CNN Top Tagger Details

	DeepTop minimal	Our final tagger
Training	SGD $\eta = 0.003$ minibatch size=1000 MSE loss	AdaDelta $\eta = 0.3$ with annealing schedule minibatch size=128 cross entropy loss
CNN architecture	8C4-8C4-MP2-8C4-8C4-64N-64N-64N	128C4-64C4-MP2-64C4-64C4-MP2-64N-256N-256N
Preprocessing	pixelate→center → normalize	center→rotate→flip → normalize→pixelate
Sample size	150k+150k	1.2M+1.2M
Color	$p_T^{calo} = p_T^{neutral} + p_T^{track}$	$(p_T^{neutral}, p_T^{track}, N_{track}, N_{muon})$

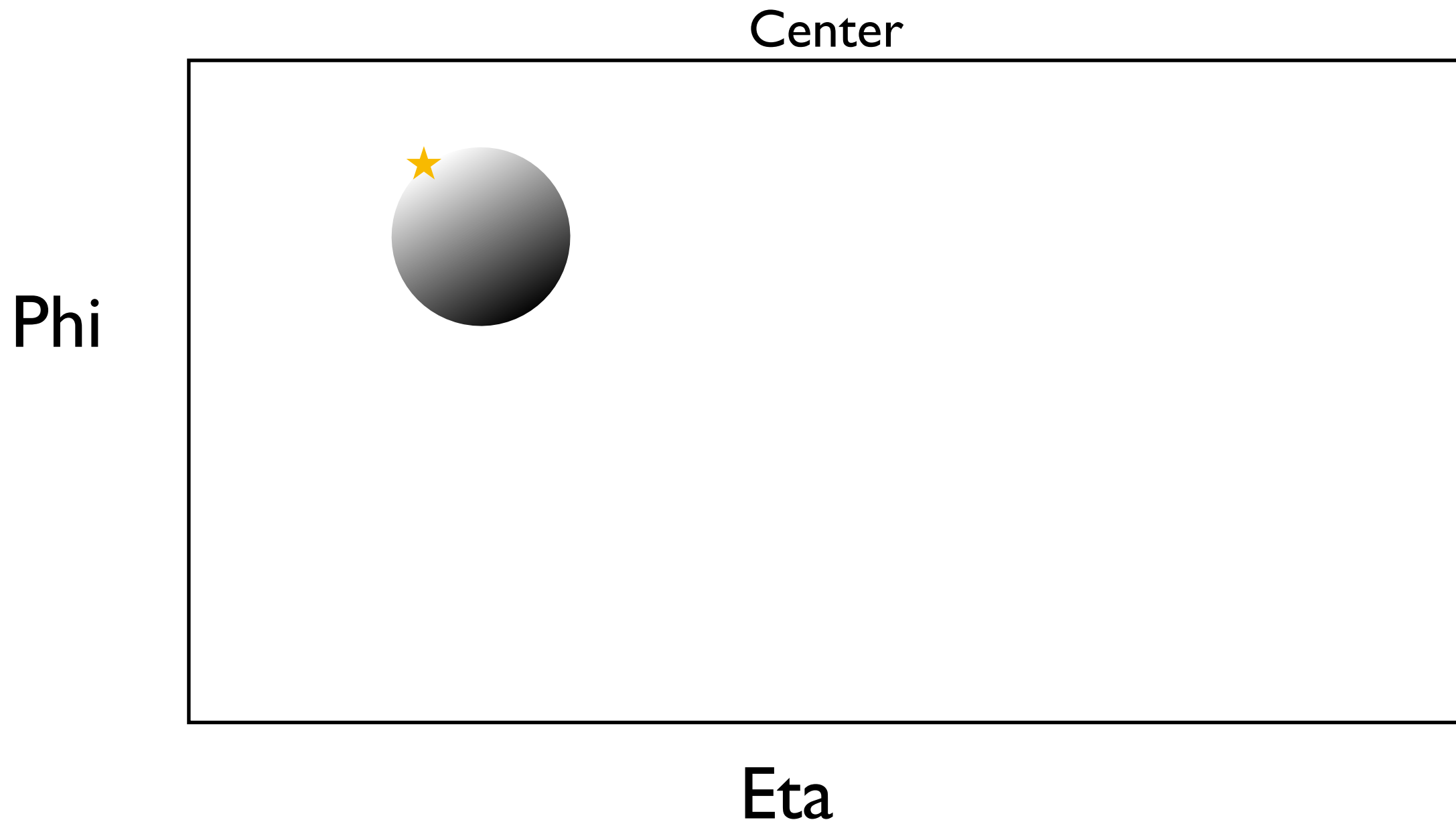
	$t$	$\tilde{g}$
PCA	0.51 / 0.04	0.98 / 0.36
Dense	0.66 / 0.13	0.90 / 0.39
CNN	0.70 / 0.19	0.77 / 0.23

# Jets as images



# Jet Images — Preprocessing

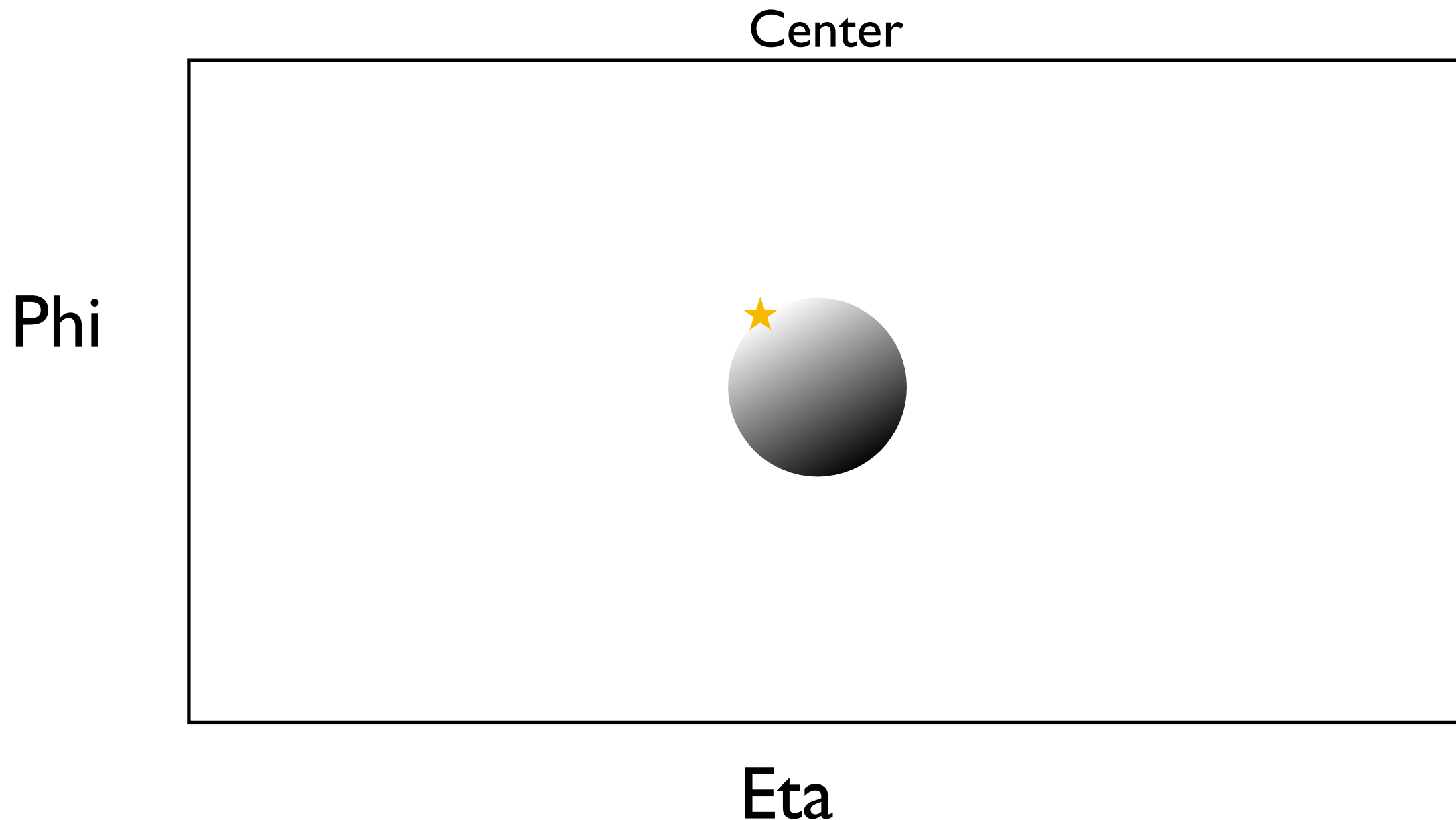
For training the neural network, it is very useful to uniformize the jet images as much as possible, consistently with the physics.





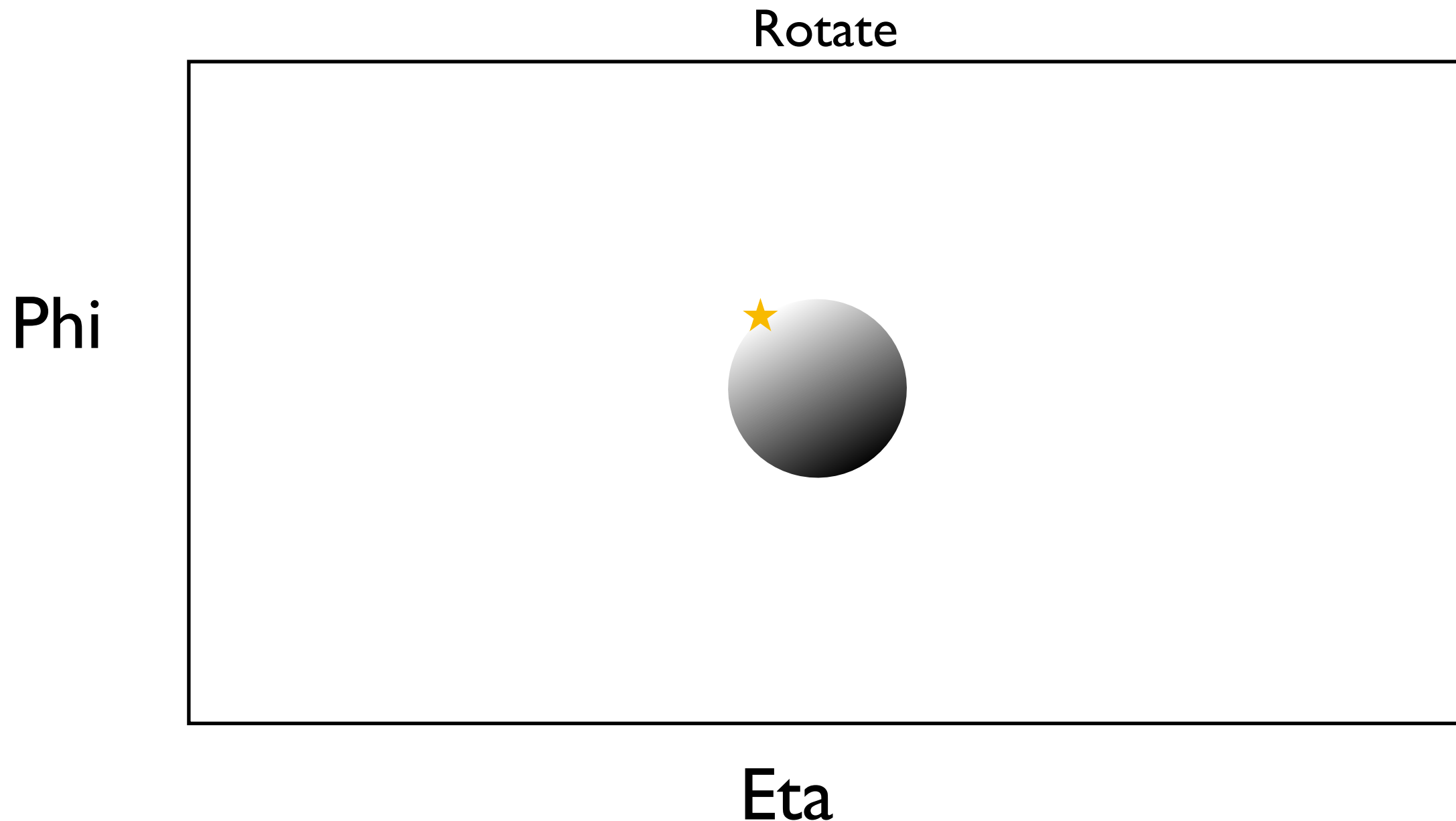
# Jet Images — Preprocessing

For training the neural network, it is very useful to uniformize the jet images as much as possible, consistently with the physics.



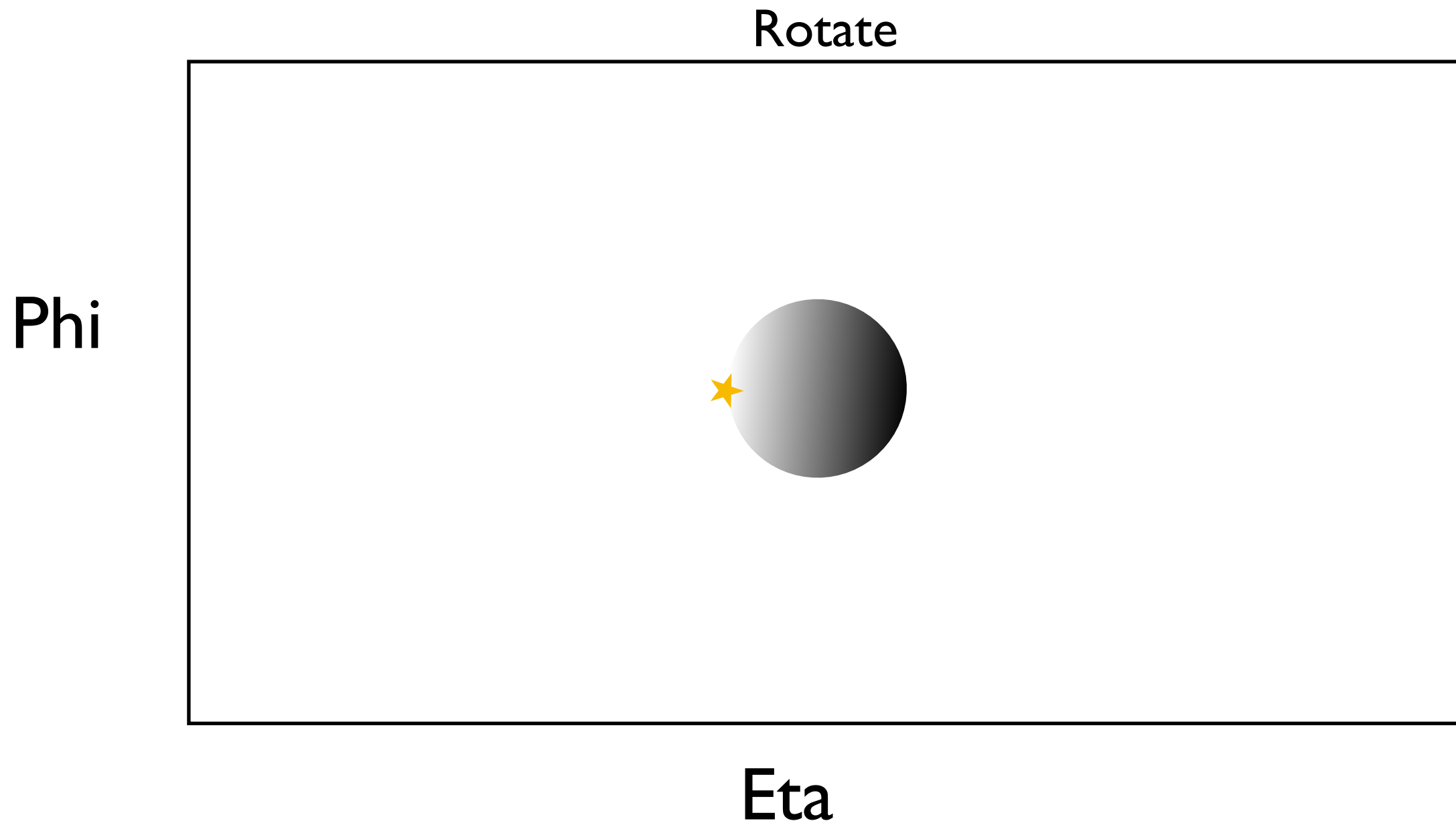
# Jet Images — Preprocessing

For training the neural network, it is very useful to uniformize the jet images as much as possible, consistently with the physics.



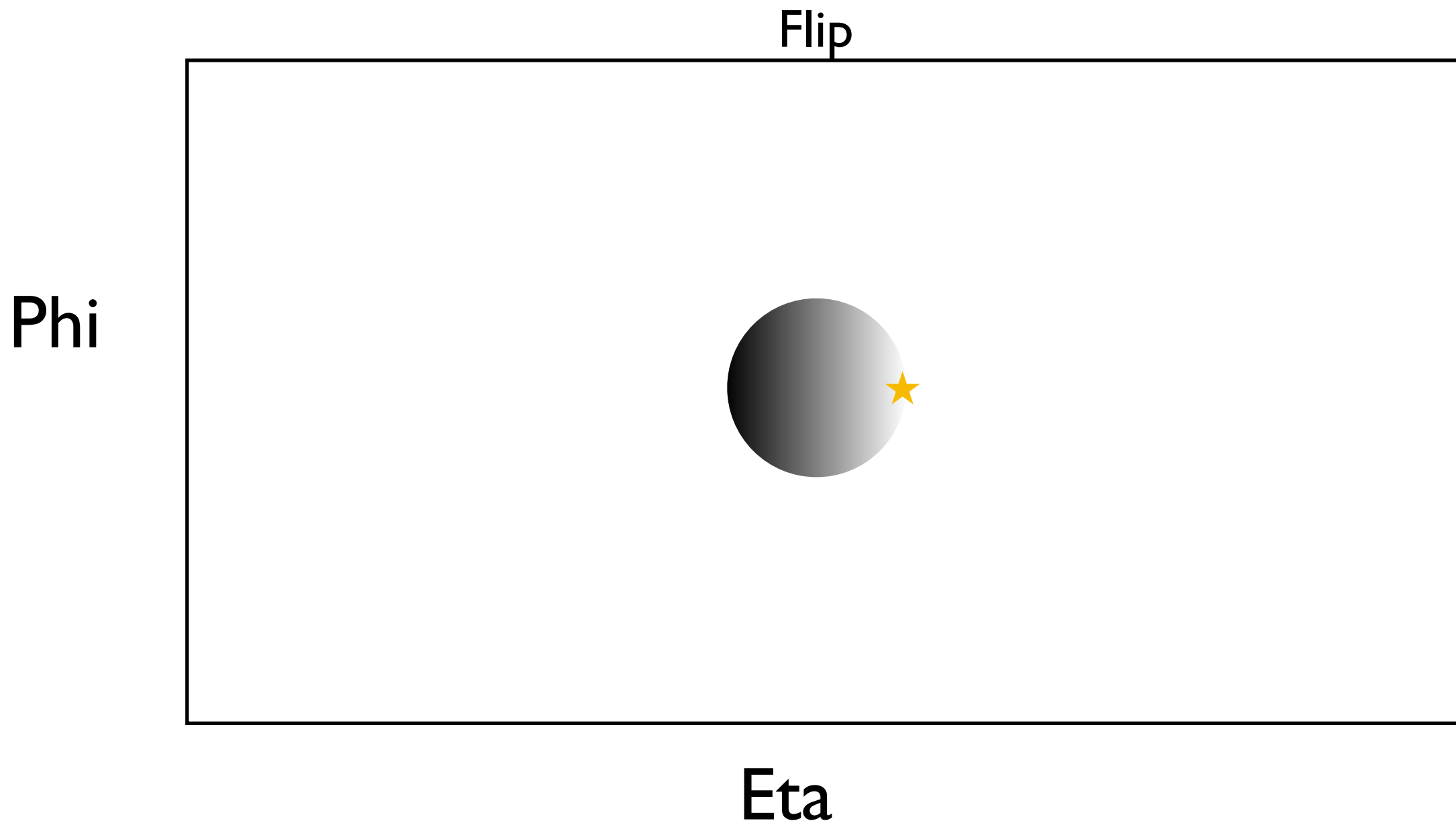
# Jet Images — Preprocessing

For training the neural network, it is very useful to uniformize the jet images as much as possible, consistently with the physics.



# Jet Images — Preprocessing

For training the neural network, it is very useful to uniformize the jet images as much as possible, consistently with the physics.



# Supervised vs Unsupervised ML

Top tagging is a prime example of “supervised machine learning” — training with **labeled datasets**.

Supervised learning is great if you know what you’re looking for.

But we are interested in searching for the unexpected.

If data has a small, **unknown** signal in it, can we train a NN to find it?

We need “unsupervised learning”: training on **unlabeled datasets**.

# Supervised vs Unsupervised ML

## Supervised Learning

## Unsupervised Learning

Training on labeled data

Training on unlabeled data

Need separate training set

Train directly on entire input dataset

Used for prediction

Used for analysis

Classification, regression

Clustering, density estimation,  
dimensionality reduction