What's Anomalous in LHC Jets?

Thorsten Buss[1], Barry M. Dillon[1], Thorben Finke[2], Michael Krämer[2], Alessandro Morandini[2], Alexander Mück[2], Ivan Oleksiyuk[2], Tilman Plehn[1]

[1]Institute for Theoretical Physics
University of Heidelberg
[2]Institute for Theoretical Particle Physics and Cosmology (TTK)
RWTH Aachen University

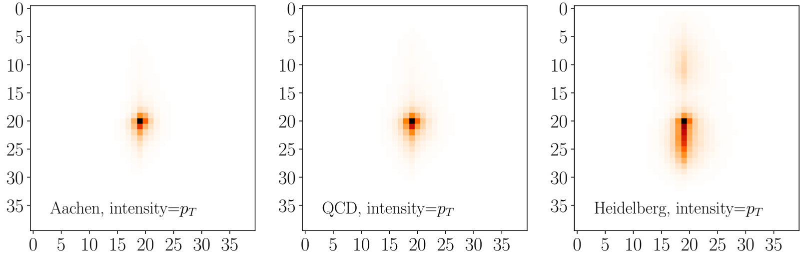
March 30, 2022

What is anomalous?

- traditional searches for new physics based on signal hypotheses
  $\rightarrow$ hard to generalize $\rightarrow$ potentially misses discovery
- possible way out: machine learning based anomaly detection
- any possible LHC jet will occur with a finite probability
  $\rightarrow$ the concept of outliers is not defined
- define jets lying in low-density phase space regions as anomalous
  $\rightarrow$ need for background density estimation

Dark jets

Aachen	**Heidelberg**
$pp \to Z' \to \bar{q}_d q_d$	$pp \to \bar{q}_d q_d$
$m_{q_d} = 500$ MeV	$m_{q_d} = 50$ GeV
$r_{inv} = 0.75$	$r_{inv} \simeq 0.0$

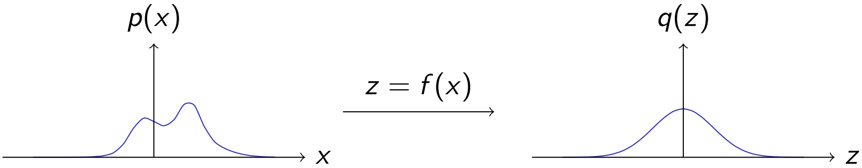


[1] tools: Madgraph5, Pythia8, Delphes3, FastJet

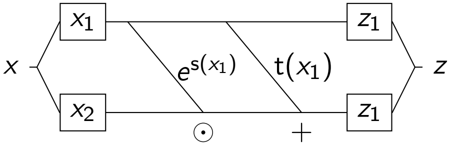
Normalizing flows

- diffeomorphism between physics space and latent space
- transform physics space distribution into a simple prior distribution
- change of variables formula allows for physics space density estimate
- training: minimize negative log-likelihood

$$p(x) = q(f(x))\,|J(x)| \qquad \mathcal{L} = -\log q(f(x)) - \log|J(x)|$$



Invertible Neuronal Network (INN)



$$\begin{pmatrix} z_1 \\ z_2 \end{pmatrix} = \begin{pmatrix} x_1 \\ x_2 \odot e^{s(x_1)} + t(x_1) \end{pmatrix} \Leftrightarrow \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} z_1 \\ (z_2 - t(z_1)) \odot e^{-s(z_1)} \end{pmatrix}$$

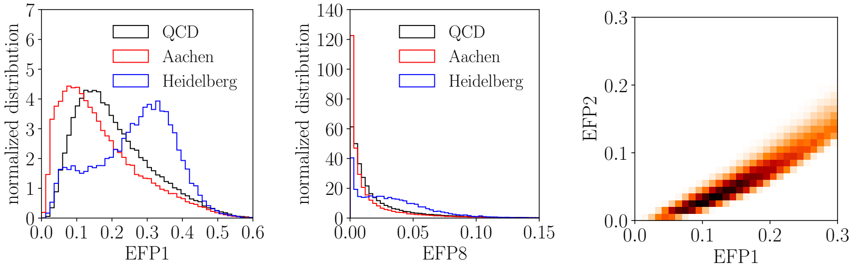
- class of neuronal networks
- learn bijective mappings
- forward and backward direction fast
- tractable jacobian

- ideal for normalizing flows
- 24 affine coupling blocks
- each followed by an orthogonal transformation

[2] L. Dinh, J. Sohl-Dickstein, and S. Bengio. *Density estimation using Real NVP.*

Data representation

- jet images are very sparsely populated
  $\rightarrow$ makes density estimation hard
- energy flow polynomials[3] (EFP): set of jet substructure observables
- using the eight prime EFPs up to order d = 3
- preprocessing: PCA

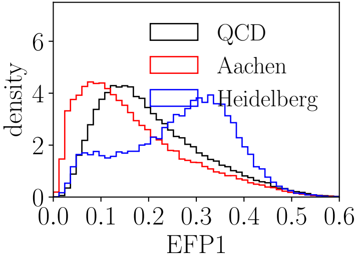


[3] P. T. Komiske, E. M. Metodiev, and J. Thaler. *Energy flow polynomials: A complete linear basis for jet substructure.*

Anomaly scores

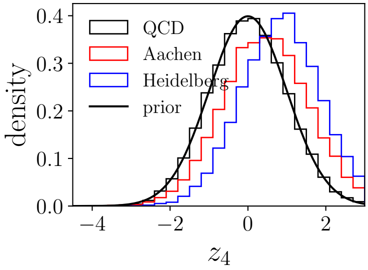
physics space density

- depends on data representation
- independent of network architecture



latent space density

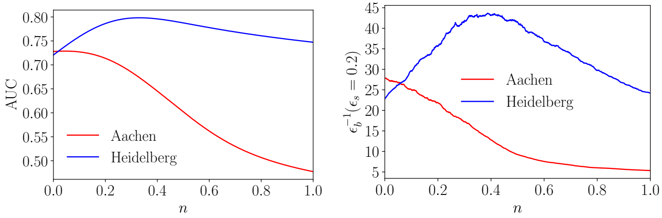
- less dependent on data representation
- depends possibly on network architecture



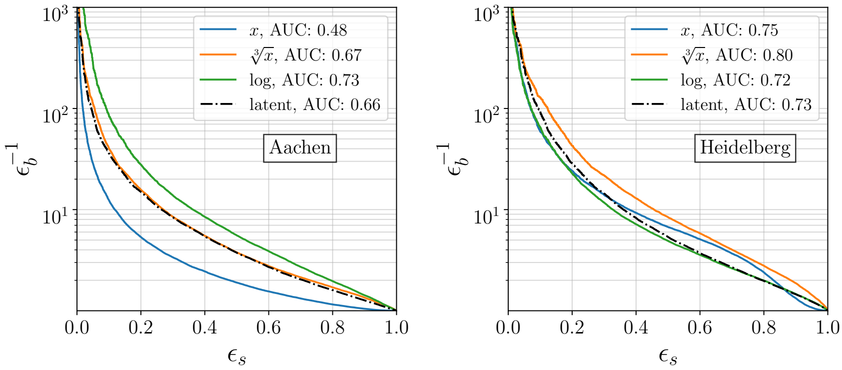
Reweighting

- using density based tagging on the EFP space leads to a bias
  $\rightarrow$ more complex jets are detected as more anomalous
- bias can be compensated for by using a reweighting of the input
- by using the change of variables formula we can study the effect of reweighting without having to retrain the network

$$g(x) = x^n \qquad \mathcal{L} \rightarrow \mathcal{L} + (n-1)\sum \log x + \sum \log n$$



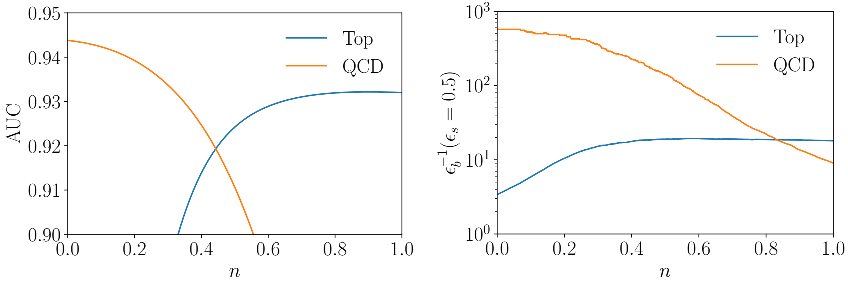
Performance



Summary

- defining jets lying in low-density phase space regions as anomalous
  $\rightarrow$ need for background density estimation
- INNs in a normalizing flows setup allow for density estimation
- reweighting of the input changes the definition of anomalies
  $\rightarrow$ can compensate biases
- even hard to find anomalies like our dark jets can be found with this method

Reweighting Top



Performance Top