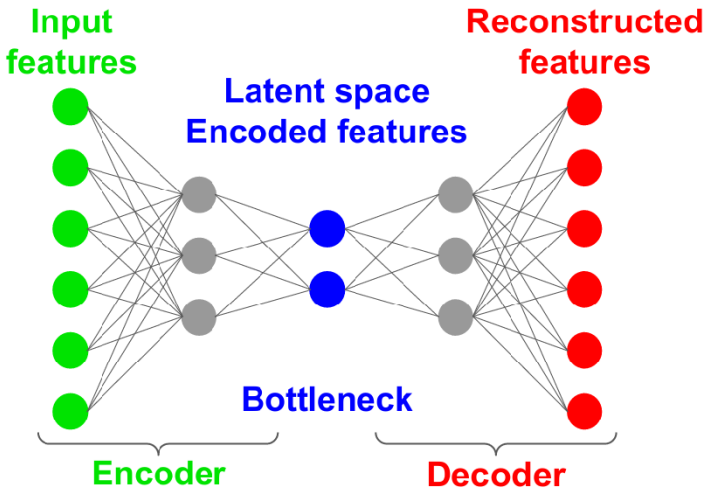
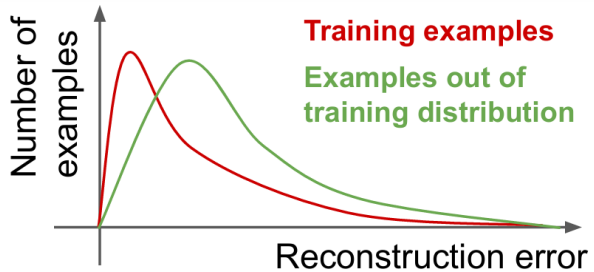
# WASSERSTEIN NORMALIZED AUTOENCODERS FOR DETECTING ANOMALOUS JETS IN CMS

Roberto Seidita, on behalf of the CMS collaboration

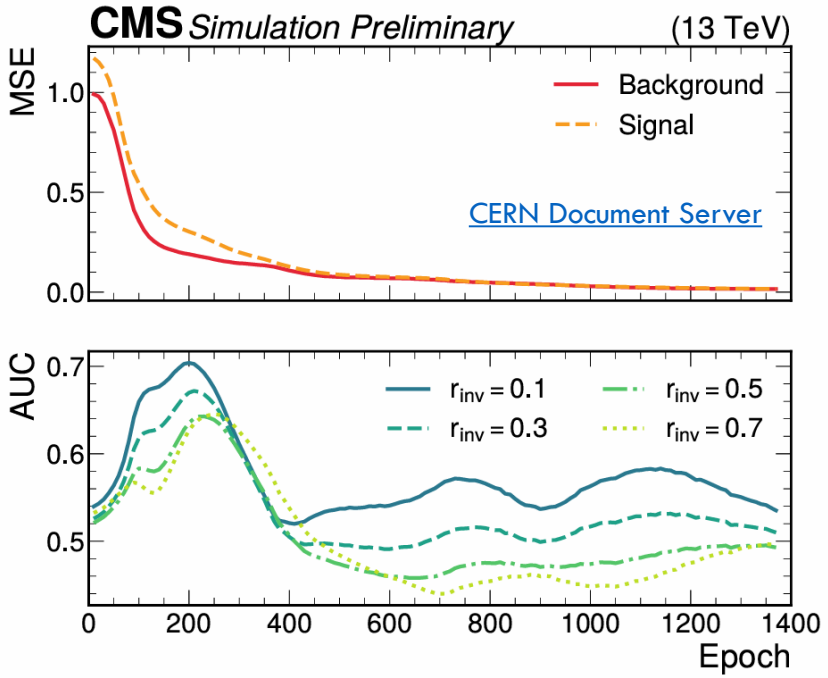
Variations on the autoencoder (AE) idea are among the most widely used tools for anomaly detection in HEP

Simple idea: learn compressed representation of the data, anomalies are spotted by failing to reconstruct them




In reality, this does not always work



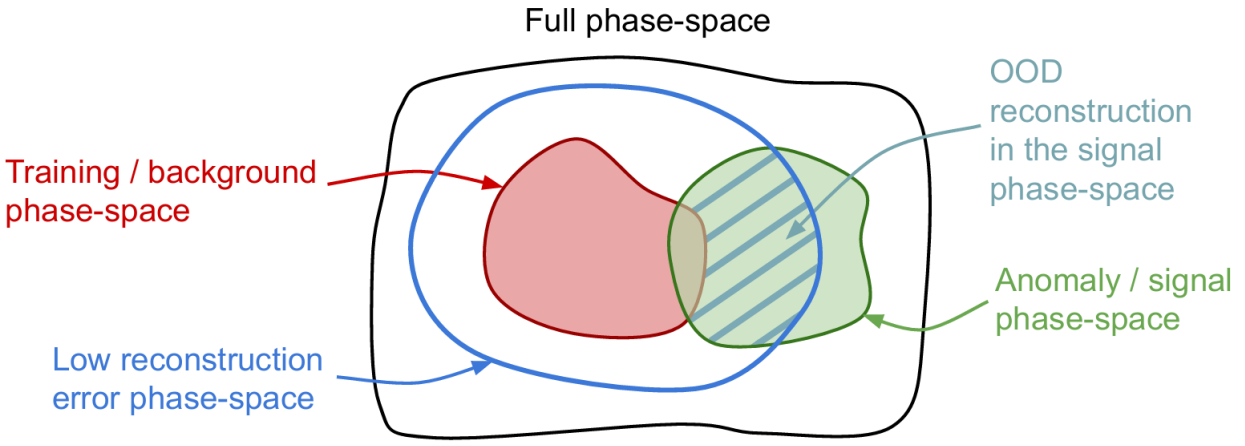

***Signal:*** *strongly-interacting dark sector events producing hadronic jets with fraction* $r_{inv}$ *of invisible particles*

***Background:*** *jets from top quarks*

# WASSERSTEIN NORMALIZED AUTOENCODERS FOR DETECTING ANOMALOUS JETS IN CMS

Roberto Seidita, on behalf of the CMS collaboration

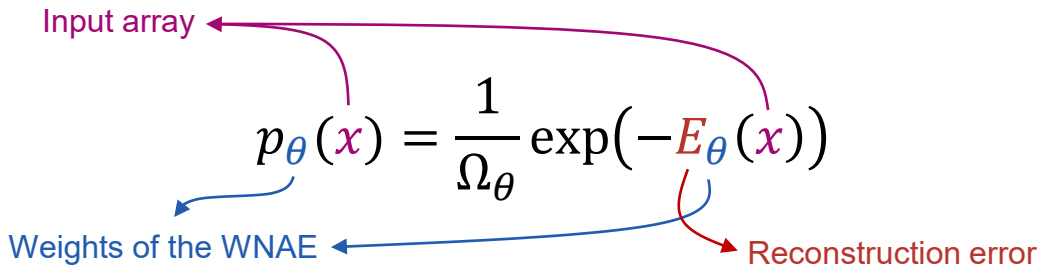
Why does this happen? → Out-of-distribution (OOD) reconstruction




What you want an anomaly detector to do: reconstruct the training data *and only the training data well*

What we ask of AEs: reconstruct the training data well

Can we ask the "correct" question? Yes!

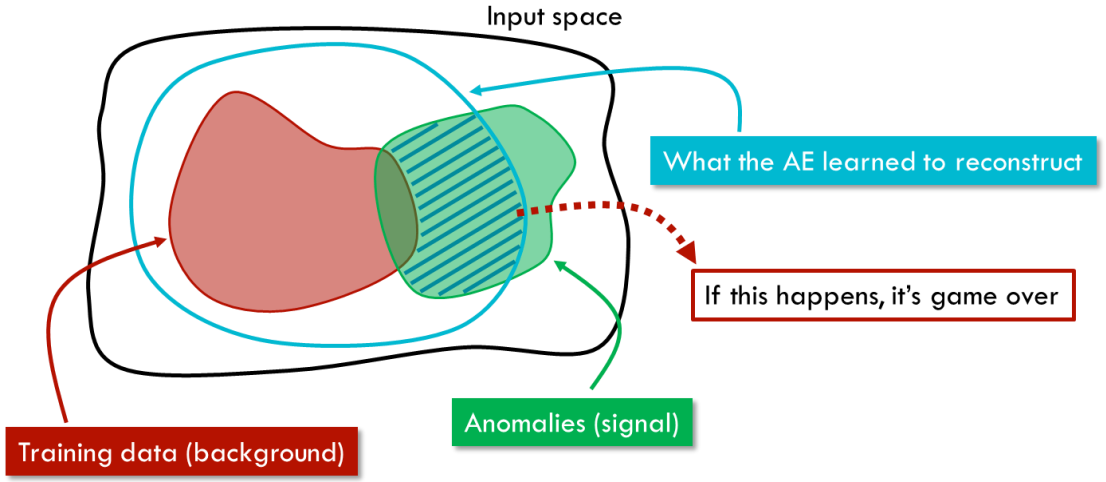


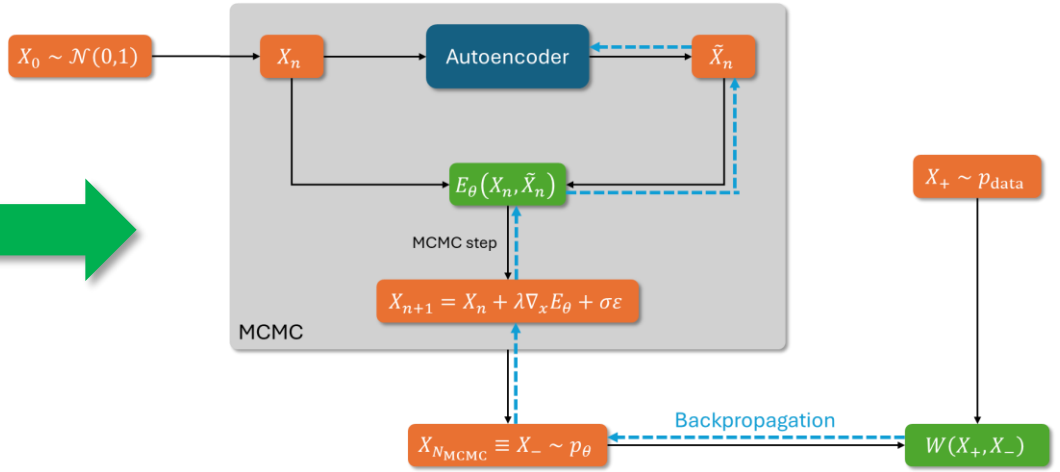

Normalized AE [1] approach:

- Assign a maximum-entropy p.d.f. to the AE
- Use it to sample what the AE is learning to reconstruct
- Enforce that only the training data is learned

# WASSERSTEIN NORMALIZED AUTOENCODERS FOR DETECTING ANOMALOUS JETS IN CMS

Roberto Seidita, on behalf of the CMS collaboration




Wasserstein Normalize AE:

- At training time, sample from induced distribution
- Measure the Wasserstein distance to the training sample
- Minimize it

Many complications, but it works! Come tomorrow for more info and results ☺