

High-Dimensional Anomaly Detection with Radiative Return in e^+e^- Collisions

Julia Gonski

28 October 2021

ILC Workshop on Potential Experiments (ICLX)

[arXiv:2108.13451](https://arxiv.org/abs/2108.13451)



Berkeley
UNIVERSITY OF CALIFORNIA

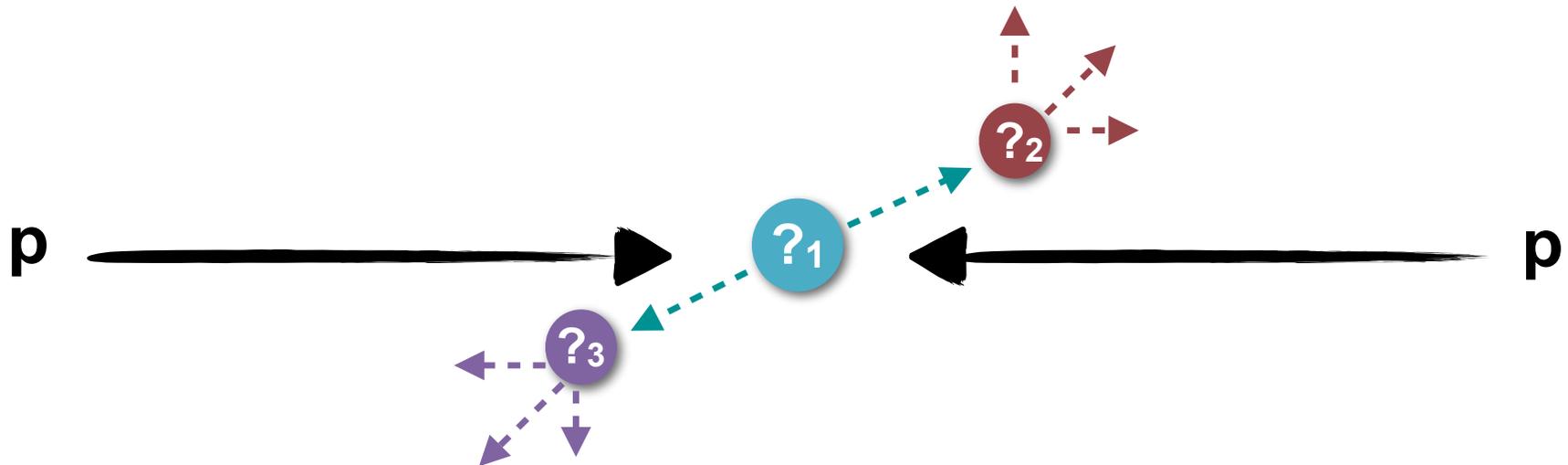


Outline

- **Motivation**
 - Anomaly detection in HEP
 - e^+e^- & radiative return
- **Strategy**
 - Dataset details/samples
 - Training setup
- **Results**
 - Data vs. simulation (semi-supervised)
 - Weakly supervised via the CWoLa method
- **Experimental outlook & conclusions**

Anomaly Detection in HEP

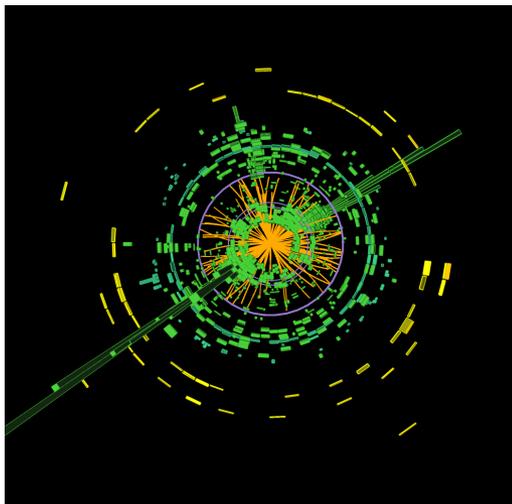
- **Anomaly detection (AD)** = identify features of the data that are inconsistent with a background-only model
 - “Semi-supervised” training: some events have labels (eg. signal contaminated)
 - “Weakly supervised” training: noisy labels (eg. some labels are wrong)
- At the **Large Hadron Collider**: no recent new physics + many exclusion results → develop strong model independent search program
 - [LHC Olympics 2020](#): cross-experiment/theory “competition” of AD methods
 - [Dark Machines](#) anomaly score challenge



AD Beyond the LHC

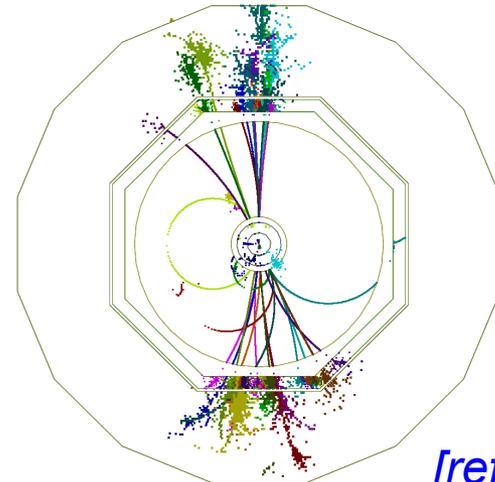
- **Snowmass 2022:** e^+e^- colliders, specifically ILC, are strong candidates for the next international accelerator
- How to exploit anomaly detection in an entirely different type of particle collision?
 - Many crucial differences in hadron vs. e^+e^- events: initial state knowledge, background processes, pileup, detector info

$pp \rightarrow$ dijet, $\sqrt{s}=13$ TeV LHC



[\[ref\]](#)

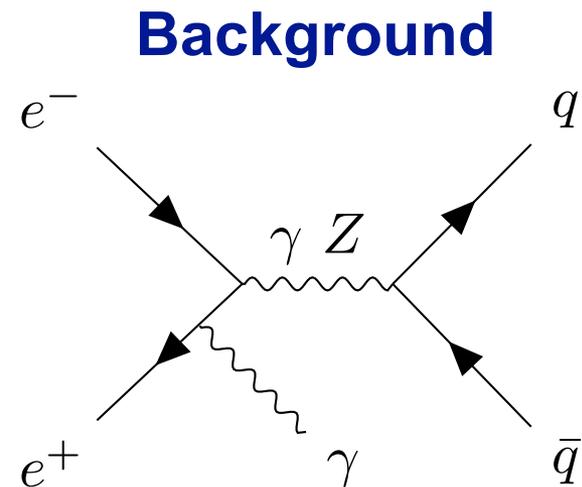
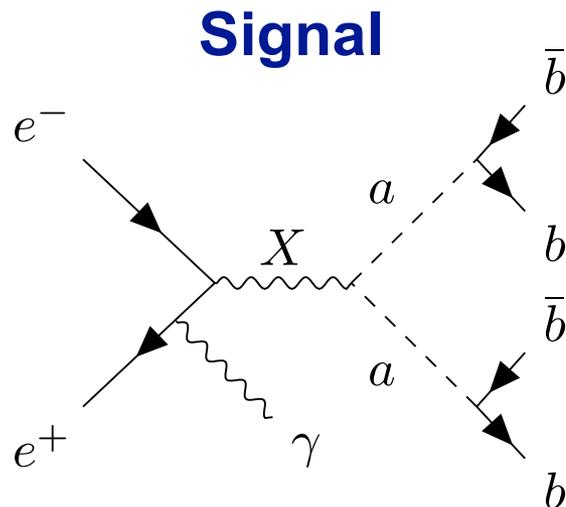
$e^+e^- \rightarrow WW \rightarrow qq\bar{q}q\bar{q}$, $\sqrt{s}=1$ TeV ILC



[\[ref\]](#)

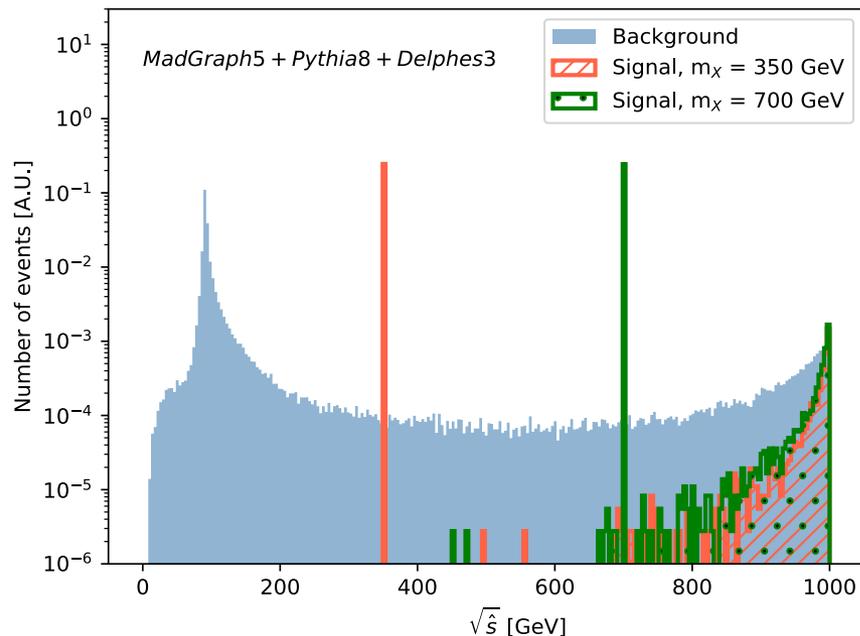
e^+e^- Dataset

- **Signal:** 700 GeV scalar $X \rightarrow 2$ 100 GeV scalars $aa \rightarrow b\bar{b}b\bar{b}$
- **Background:** Drell-Yan hadronic decays
- Reconstruct dijet final state with $R=1.0$ jets built from particle flow objects
- Generate e^+e^- collisions at $\sqrt{s} = 1$ TeV (Madgraph)
 - Pythia showering/hadronization + Delphes detector simulation (using [general ILC card](#))
 - Emulate training scenario with “[full dataset](#)” of ~ 6.5 ab^{-1}



Radiative Return in e^+e^-

- Require events to have at least 1 photon with $E > 10$ GeV from initial state radiation (ISR)
 - ISR photon can have any energy
 - Initial CoM energy in e^+e^- is exactly known
- Can use to “scan” new particle masses, à la dijet invariant mass bump hunts at the LHC



350 GeV X
~ 650 GeV γ

700 GeV X
~ 300 GeV γ

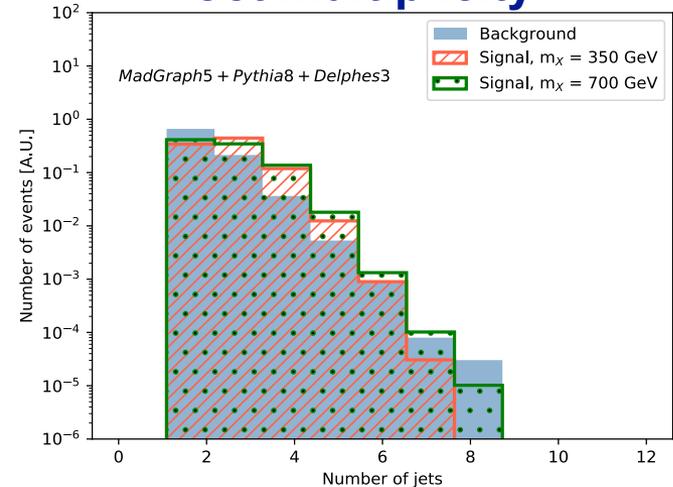
Training Setup

- Using Particle Flow Networks implemented in [EnergyFlow framework](#): model an event as an unordered, variable-length set of jets
 - Up to 15 jets per event
 - 10 features per jet: 4 vector (p_T , η , ϕ , m), b-tagging bit, 5 N-subjettiness variables τ

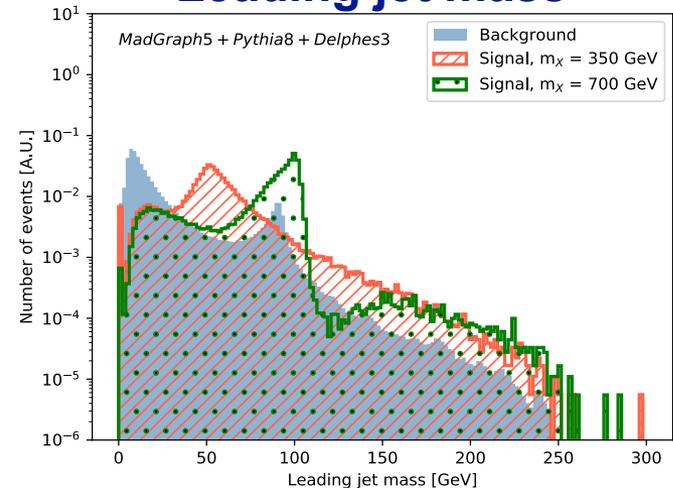
➔ 150 input features per event

- **Normalization**: normalize jets to average $p_T/\eta/\phi$ in event — critical to not induce \sqrt{s} sculpting
- **Ensembling**: train 50 models per setup with random signal injections, quantile scale outputs, and average

Jet multiplicity

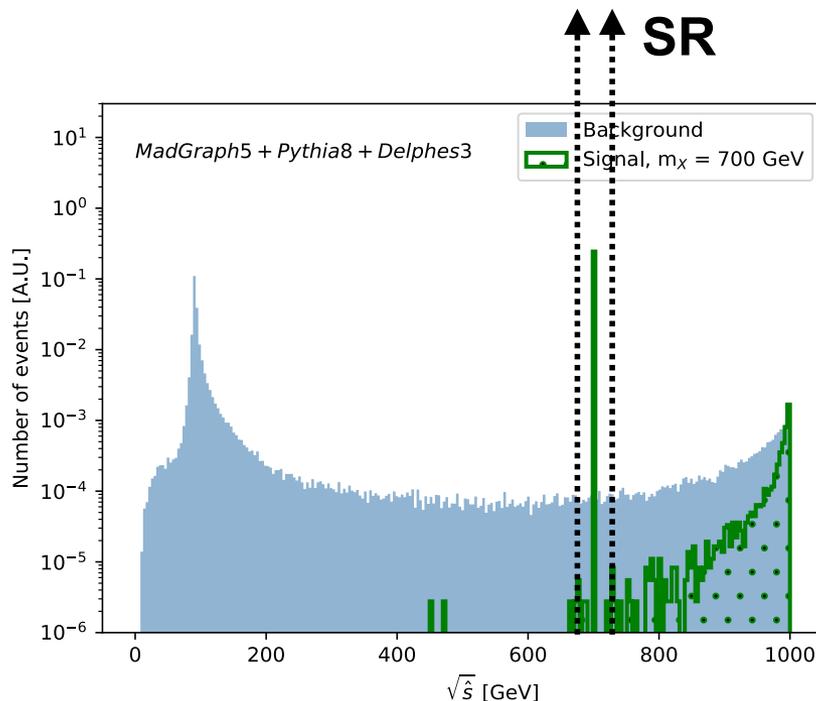


Leading jet mass



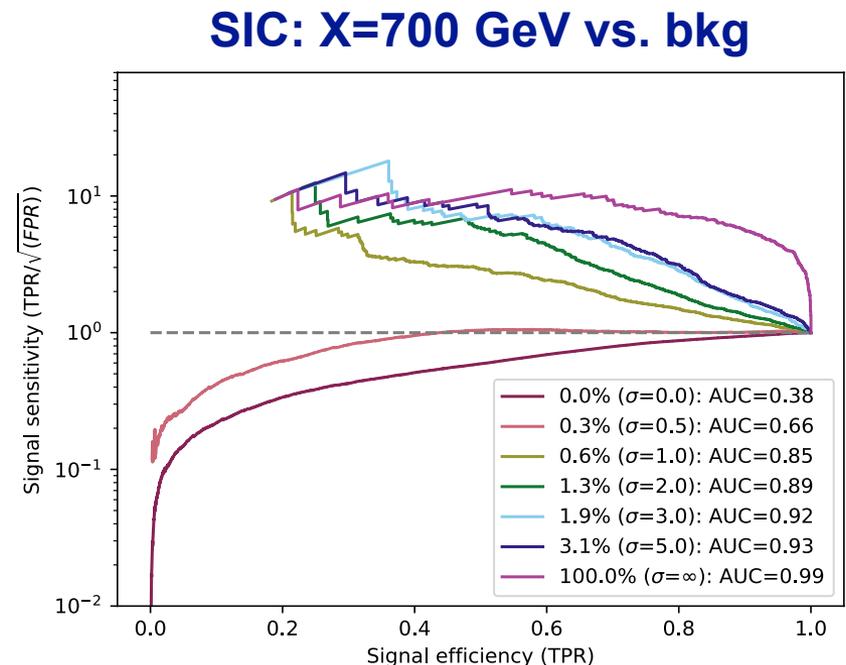
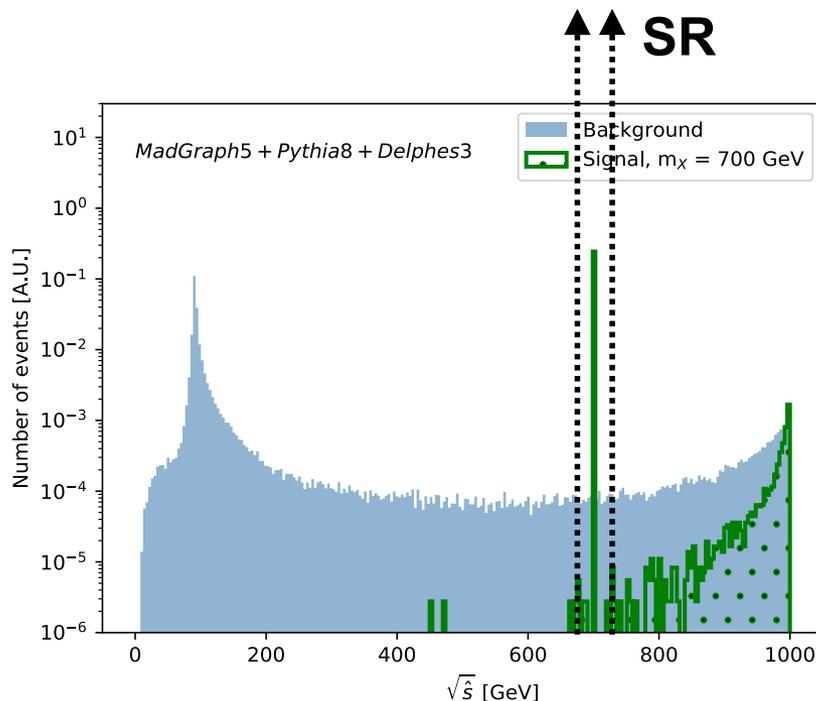
Semi-supervised (Data vs. Simulation)

- Select signal and background in ± 25 GeV windows in \sqrt{s} around the resonance mass: SR = [675, 725)
- Train with a variety of signal contaminations: $\sigma=0.0, 0.5, 1.0, 2.0, 3.0, 5.0,$ and ∞ (eg. all S vs. all B)



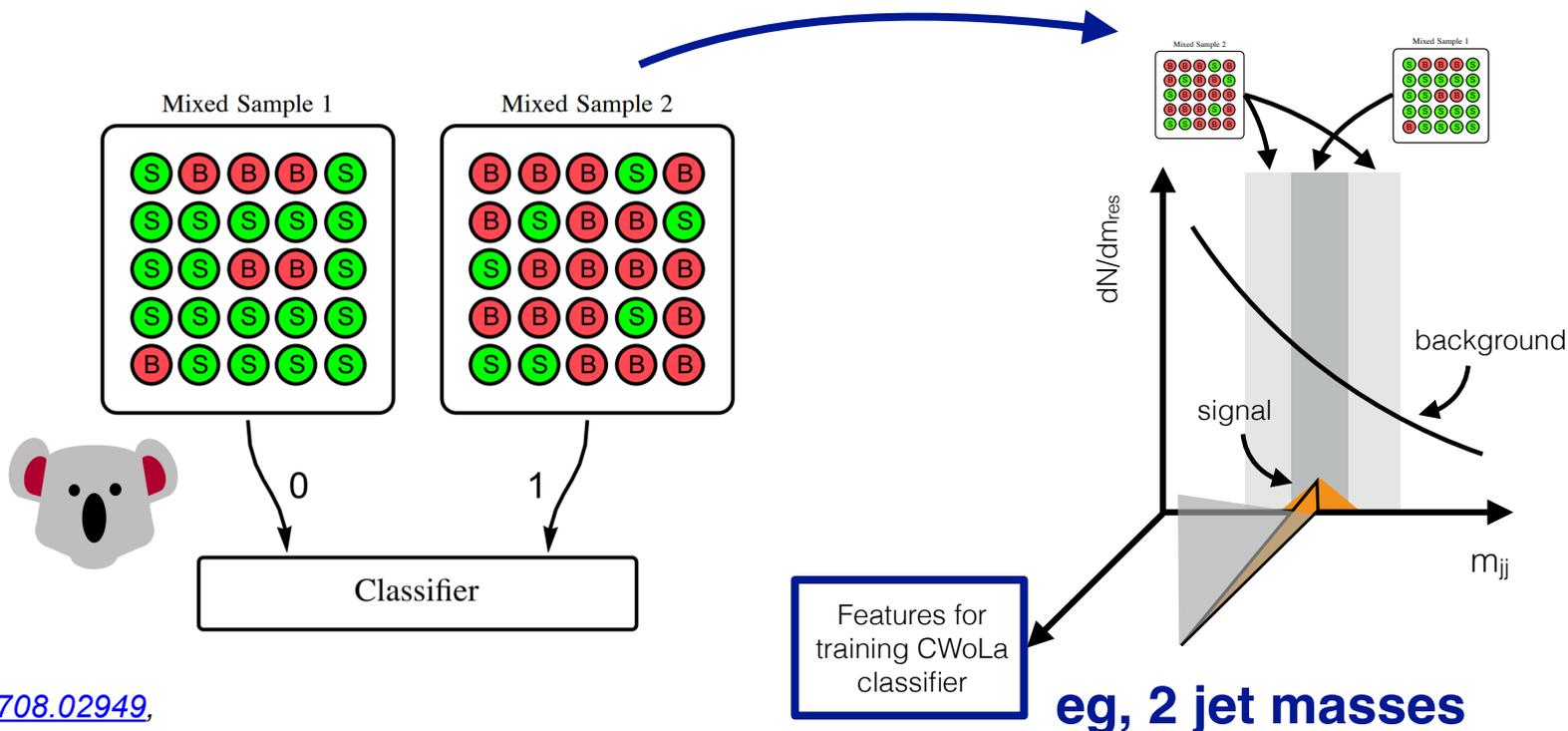
Semi-supervised (Data vs. Simulation)

- Select signal and background in ± 25 GeV windows in \sqrt{s} around the resonance mass: SR = [675, 725)
- Train with a variety of signal contaminations: $\sigma=0.0, 0.5, 1.0, 2.0, 3.0, 5.0,$ and ∞ (eg. all S vs. all B)
- ➔ **Significance Improvement Characteristic (SIC):** can enhance a 0.6% signal contamination from 1.0σ to $\sim 10.0\sigma$



Data-Driven/Weakly Supervised (CWoLa)

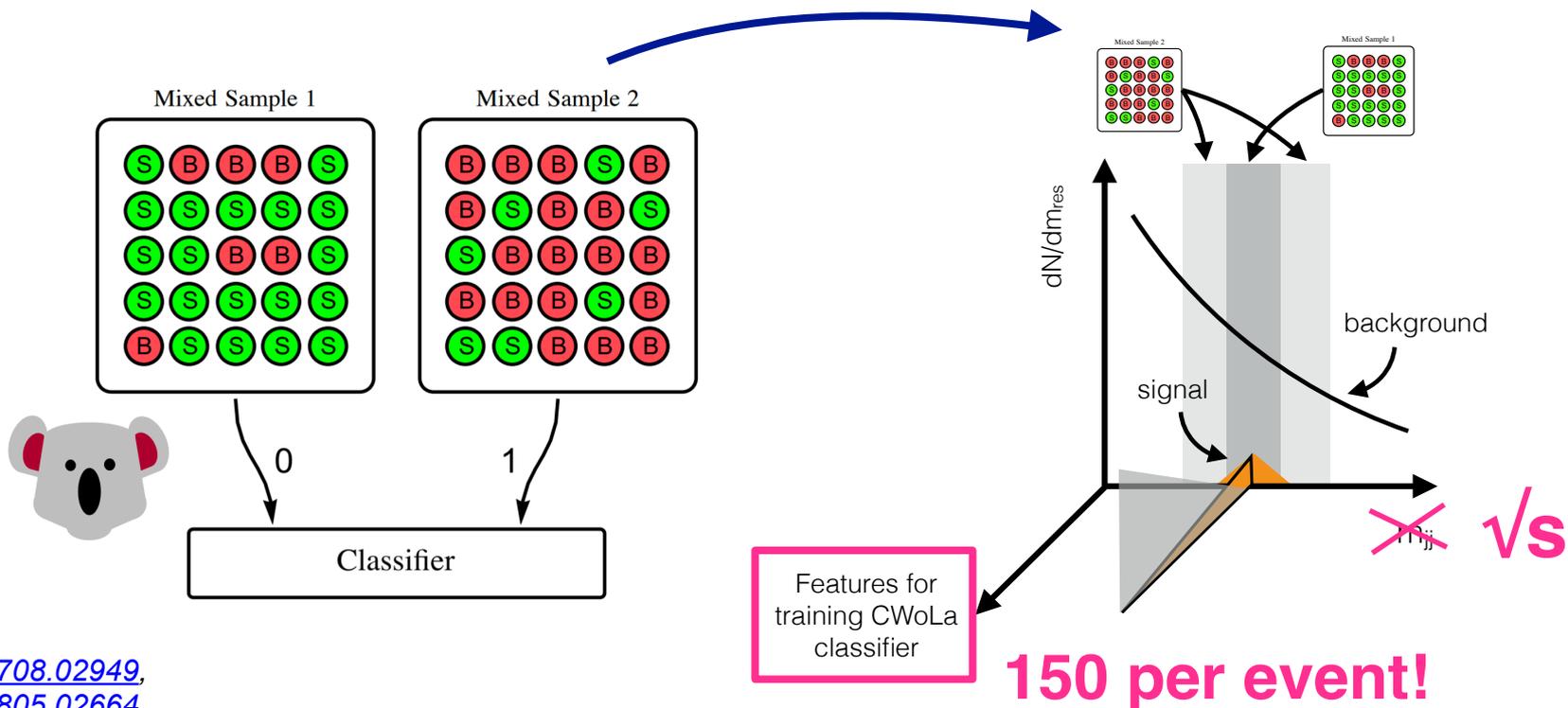
- Applied to first weakly supervised learning for [ATLAS](#) search in dijet final states
- NN trained in signal region vs. sideband is sensitive to signal vs. background characteristics
 - SR and SB defined in windows of m_{jj} , each region has different fraction of signal



Data-Driven/Weakly Supervised (CWoLa)

- Applied to first weakly supervised learning for [ATLAS](#) search in dijet final states
- NN trained in signal region vs. sideband is sensitive to signal vs. background characteristics
 - SR and SB defined in windows of m_{jj} , each region has different fraction of signal

In our case...

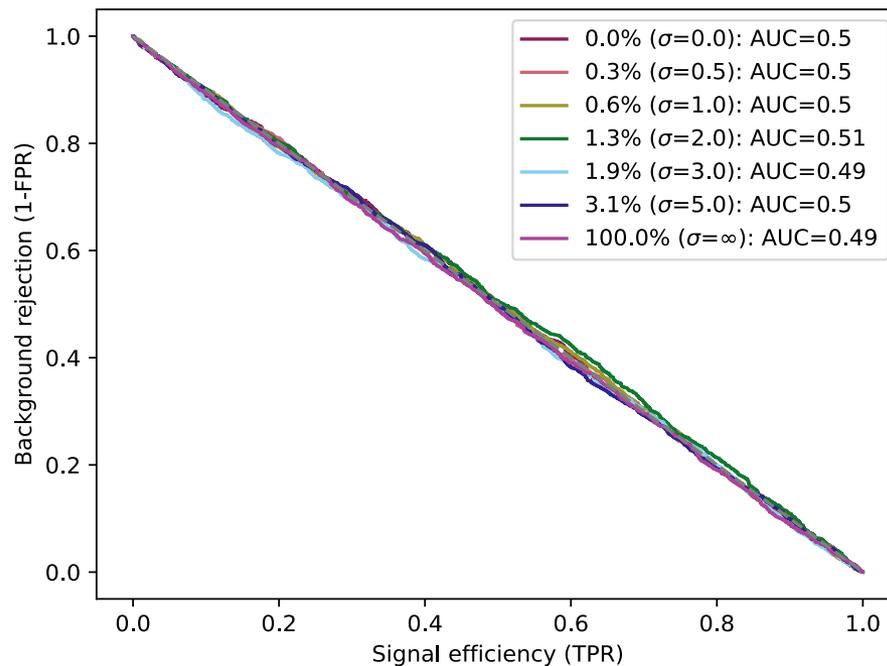


[1708.02949](#),
[1805.02664](#)

Monitoring \sqrt{s} Correlation

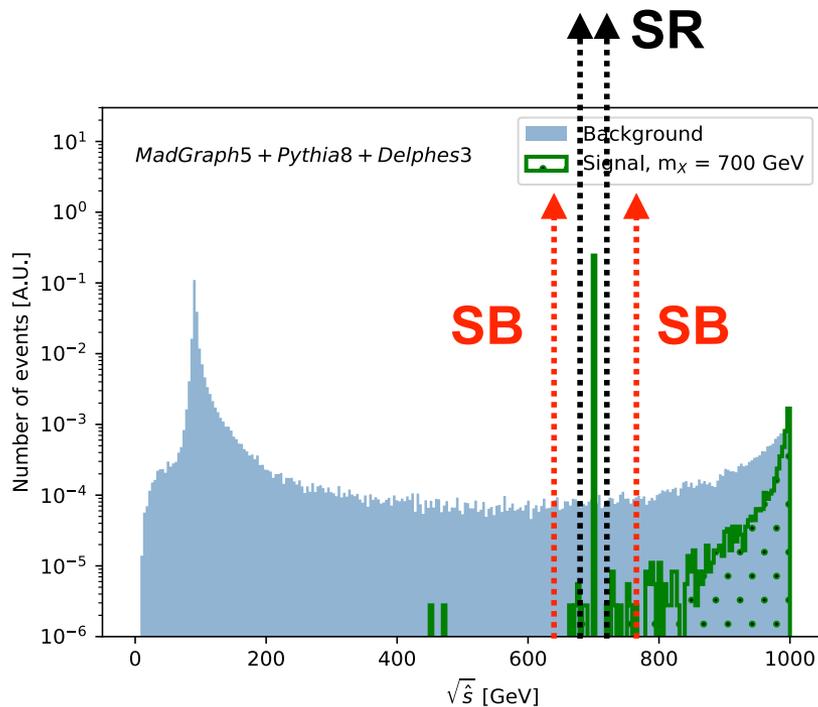
- Shift to CWoLa training means signal and background are in different \sqrt{s} bins
→ Need to validate extrapolation across \sqrt{s} & ensure little/no correlations
- CWoLa-trained classifier tested on background in SB vs. background in SR has learned nothing to discriminate on \sqrt{s} alone

ROC: Bkg in SB vs. Bkg in SR



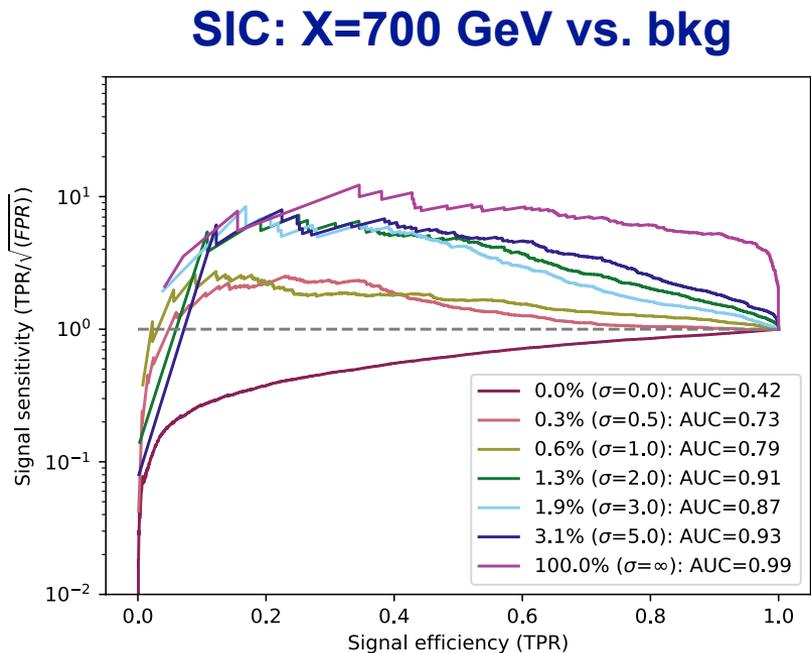
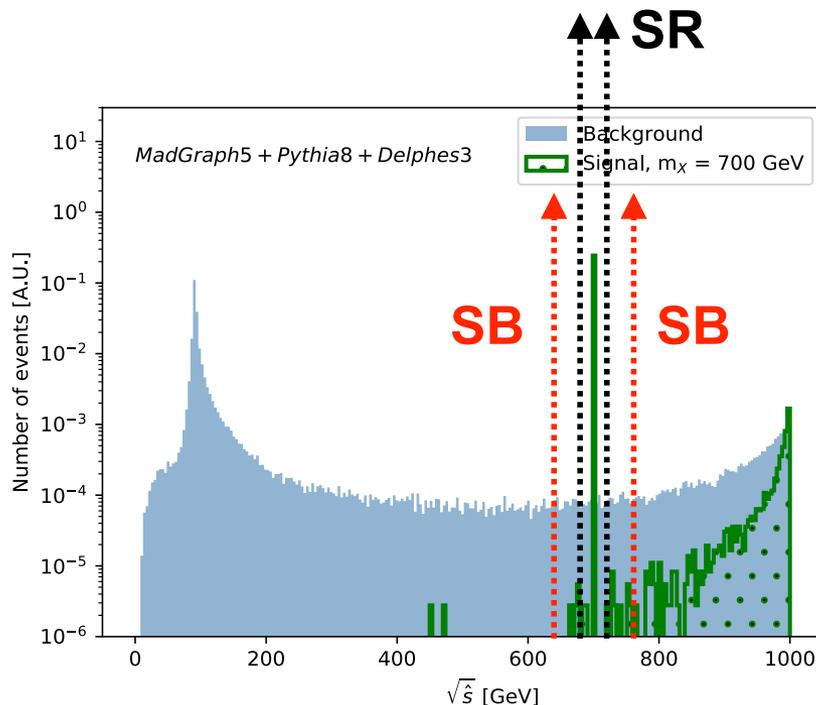
Weakly Supervised Results

- SR = [675, 725); sideband in ± 50 GeV windows around SR= [625,675) U [725,775)



Weakly Supervised Results

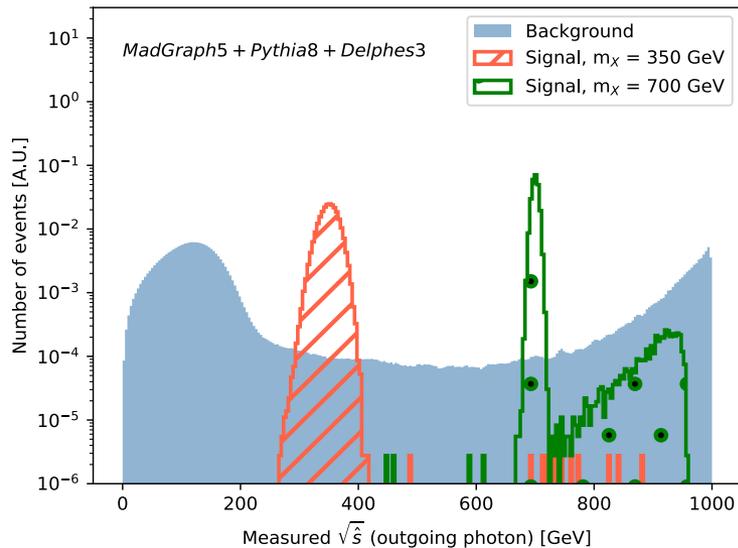
- SR = [675, 725); sideband in ± 50 GeV windows around SR= [625,675) U [725,775)
- ➔ Little degradation in performance from removing signal hypothesis: 1.0σ excess enhanced to $\sim 3.0\sigma$
 - Other contaminations have even smaller discrepancies



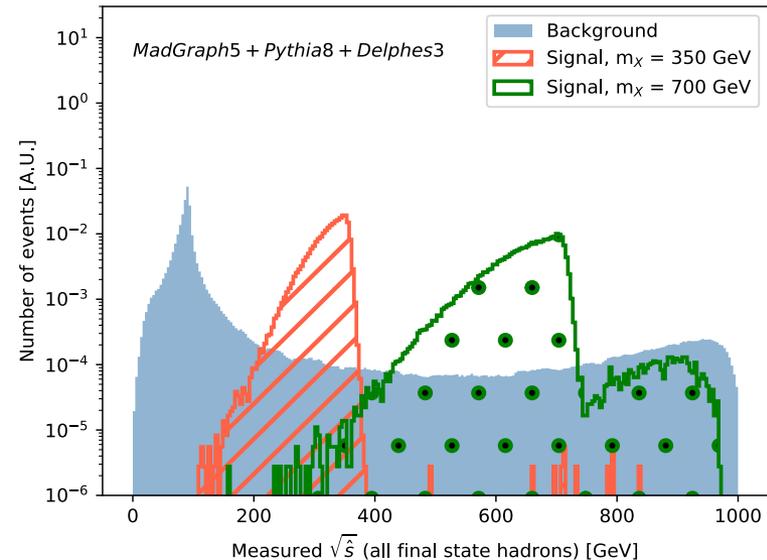
ILC Outlook

- Efficacy of method is not limited to radiative return
 - Highly analogous to exotic Higgs decays with Z associated production (eg. [1612.09284](#))
 - Further AD application to key BSM opportunities in e^+e^- , eg. SMEFT deviations...
- Detector features such as mass resolution and forward acceptance have strong impact on radiative return AD analyses
 - We are investigating different \sqrt{s} reconstruction measures to understand dependency and inform e^+e^- detector design

Measured \sqrt{s} , photon captured



Measured \sqrt{s} , photon lost



Conclusions

- Successful application of anomaly detection to a novel e^+e^- collision dataset
 - Strong performance from **high dimensional Particle Flow Network** training inputs in labeled data vs. simulation classification
 - Shift to data-driven training via CWoLa method shows little degradation, and **enhanced sensitivity to a generic new physics signal with no signal prior**
- Interesting avenue for understanding **search potential at ILC!**
- **More studies documented in [arXiv:2108.13451](https://arxiv.org/abs/2108.13451)**
 - Comparison of performance in various measures of \sqrt{s}
 - Sensitivity to lower mass X resonance (350 GeV)
 - Comparison to sensitivity from e^+e^- event-level variables
- **Next steps:**
 - Add all official ILC background processes
 - Benchmark on Z + exotic H decay
 - Understand change in sensitivity/accessible phase space at $\sqrt{s} = 250$ GeV

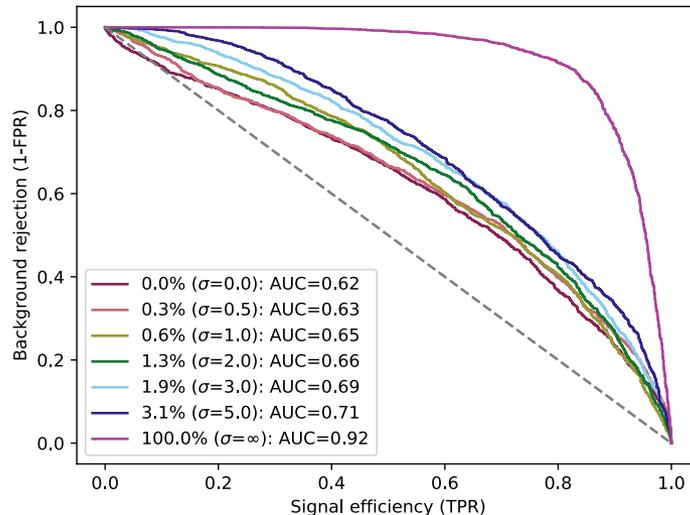
Backup

Event Level Variable Results

- Training features: 15 per event
 - Masses: leading & subleading large-R jet, [total jet mass](#)
 - Transverse momenta: leading & subleading large-R jet, leading photon, j_1 pT / γ_1 pT, reconstructed X, X pT / γ_1 pT
 - Multiplicities: # particle, # jets
 - [Aplanarity](#), [sphericity](#), [transverse sphericity](#)
 - [ln\(y23\)](#), calculated with all jets in the event

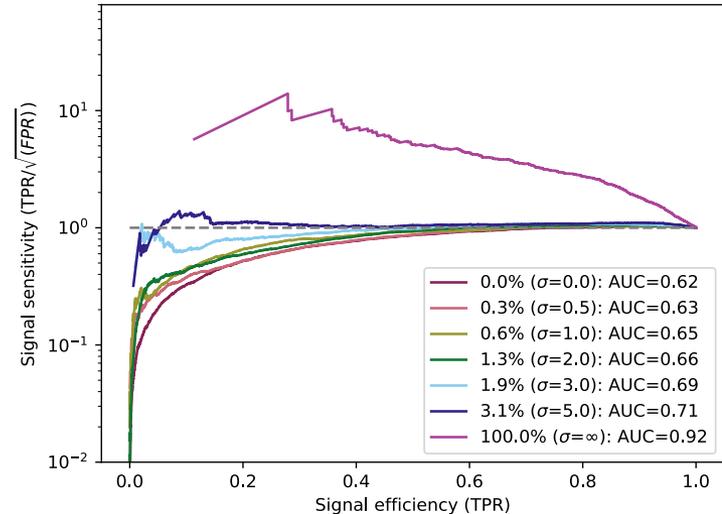
ROC

ROC: Signal ($m_X = 700$ GeV) vs. background,
truth $\sqrt{\hat{S}}$



SIC

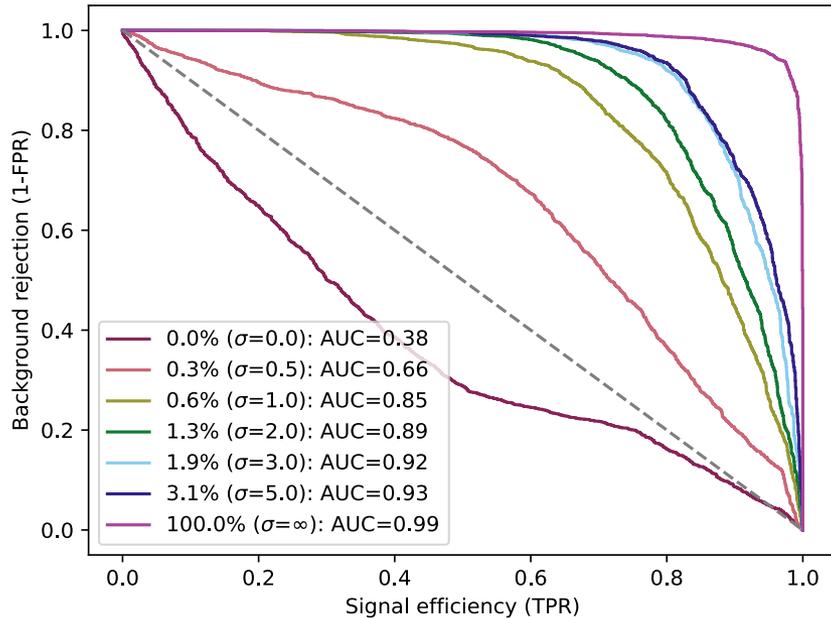
SIC: Signal ($m_X = 700$ GeV) vs. background,
truth $\sqrt{\hat{S}}$



Training Result ROC Curves

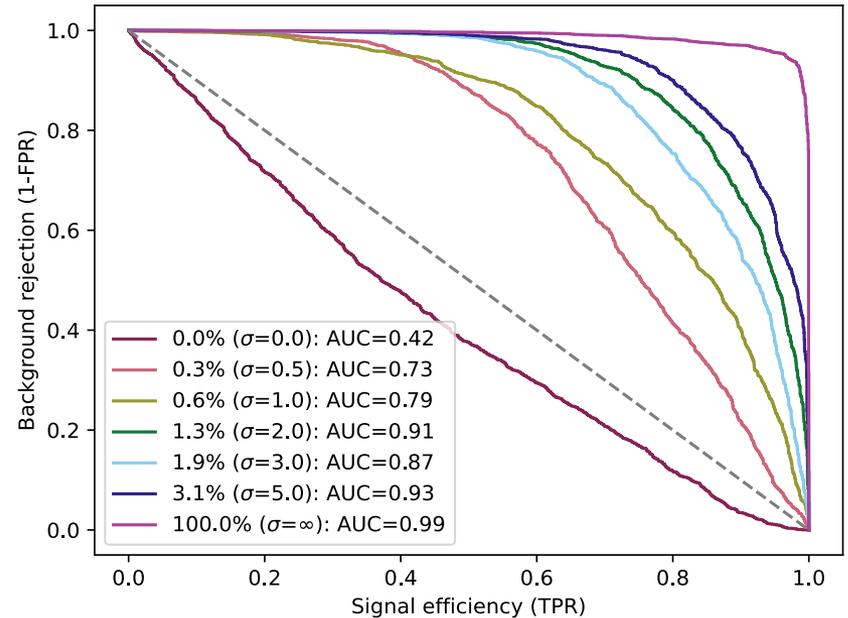
S vs. B Training

ROC: Signal ($m_\chi = 700$ GeV) vs. background,
truth $\sqrt{\hat{s}}$



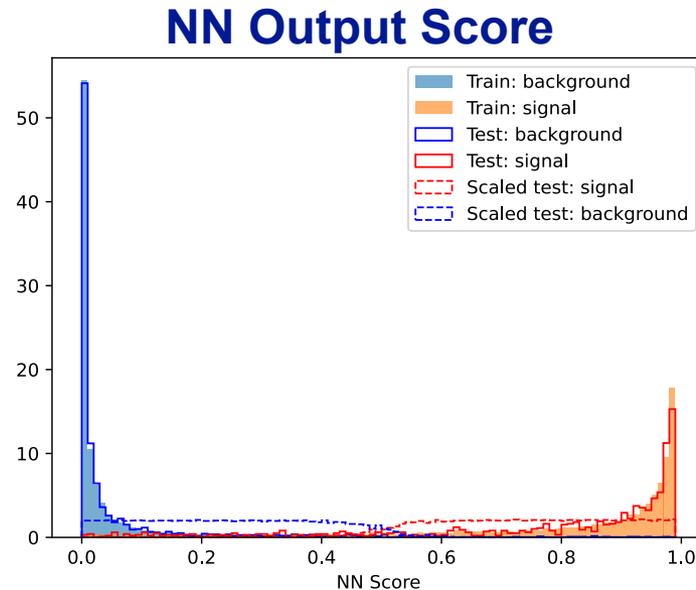
CWoLa Training

ROC: Signal ($m_\chi = 700$ GeV) vs. background,
truth $\sqrt{\hat{s}}$

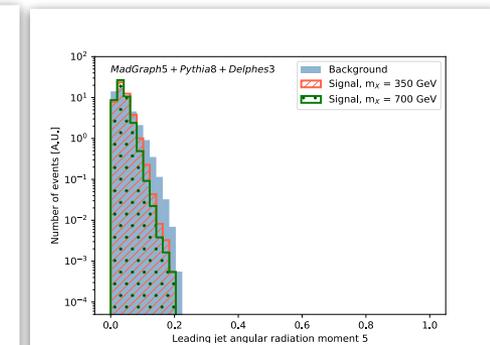
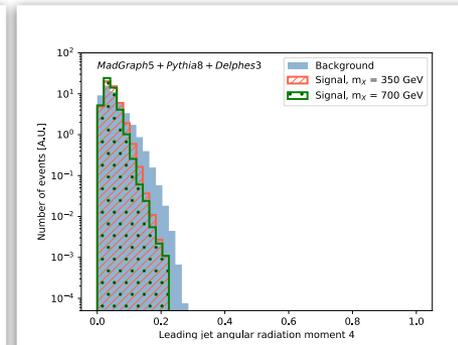
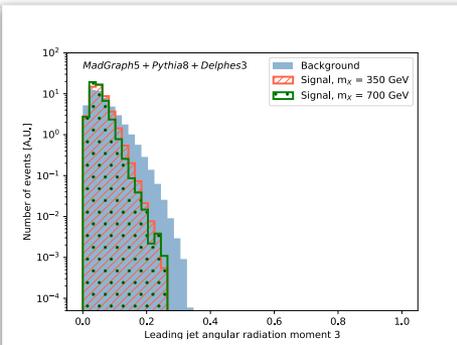
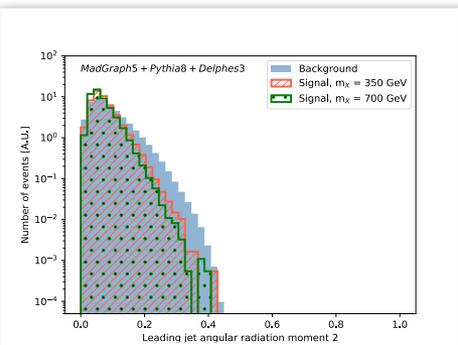
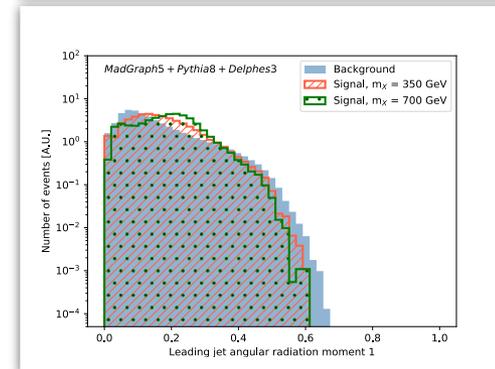
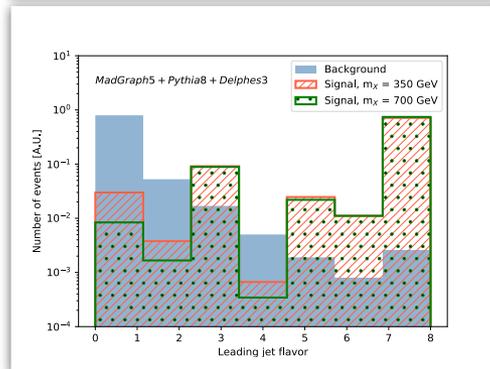
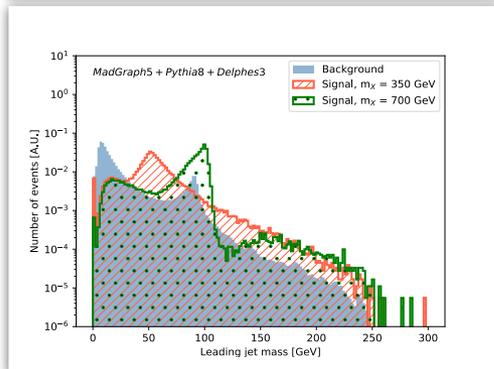
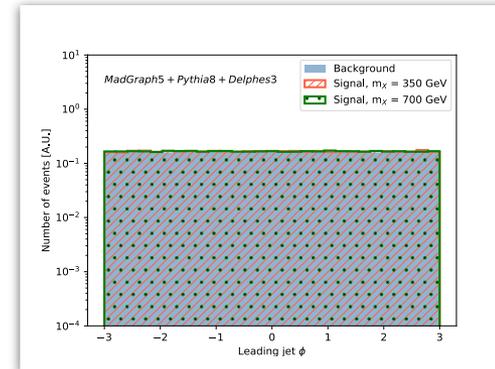
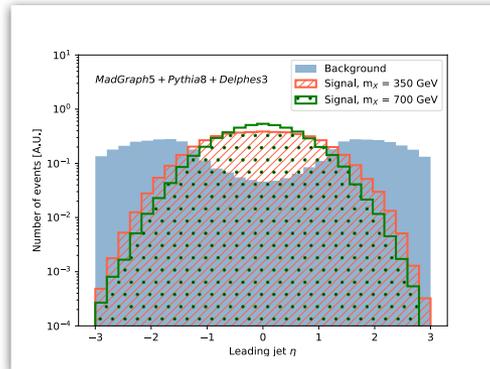
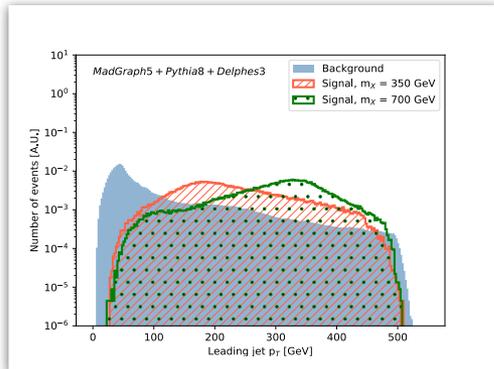


Neural Net Setup

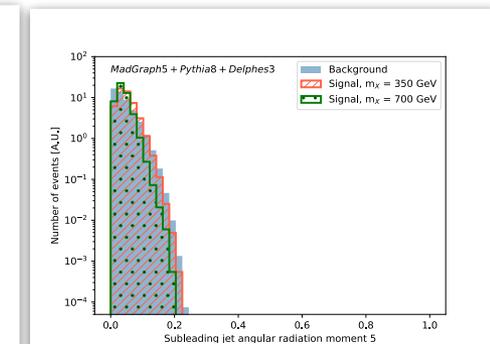
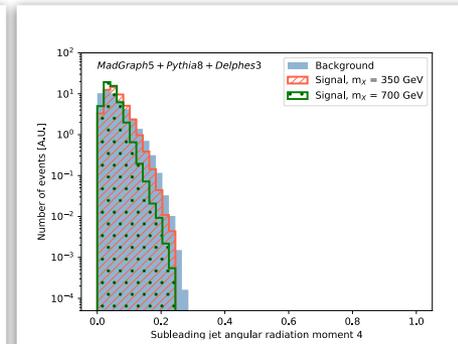
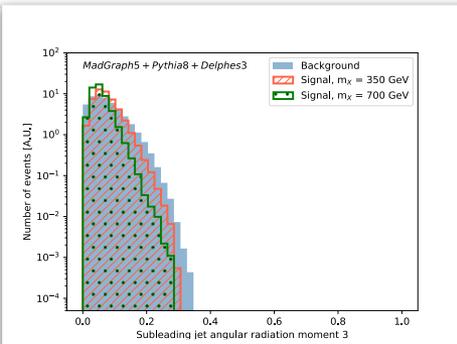
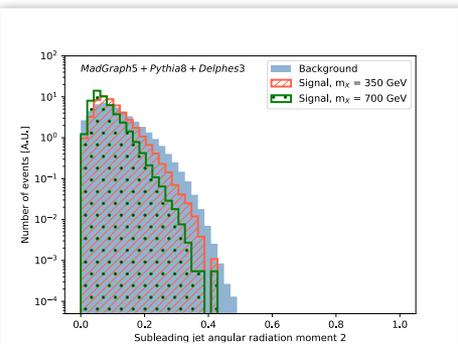
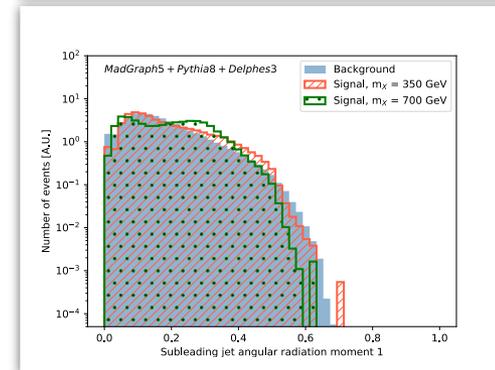
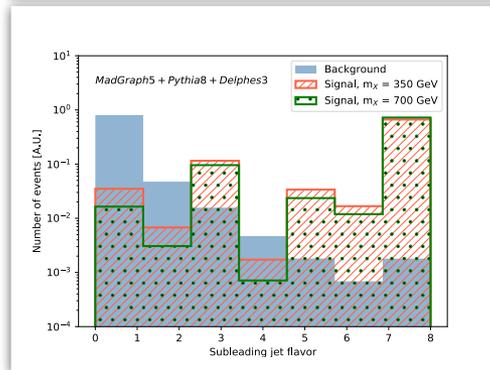
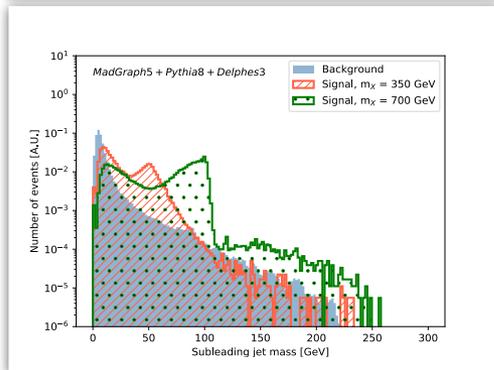
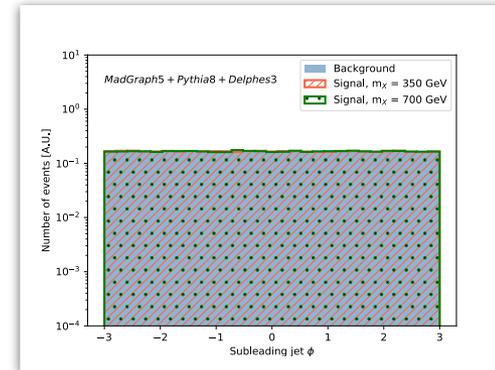
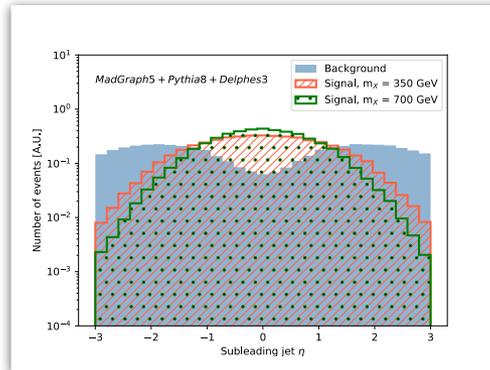
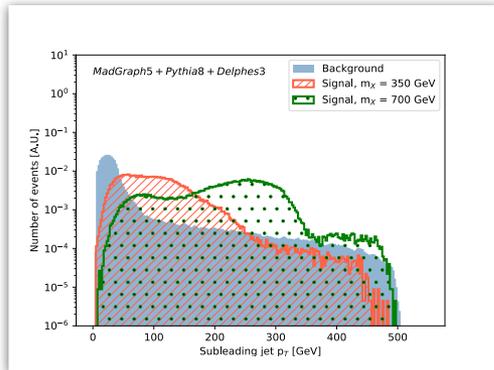
- Train over 50k events:
 - 25k background only
 - 25k with signal contamination of fixed percent
 - Test set = 10% of training (50% signal, 50% background)
- 30 epochs, batch size 100
- Adam optimizer with initial learning rate 0.0001
- Architecture: dense sizes (100, 100), Phi/F sizes (20, 20, 20)



Leading Jet Training Inputs



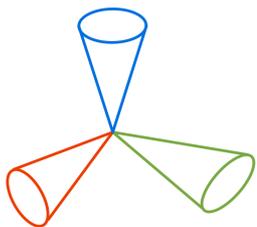
Subleading Jet Training Inputs



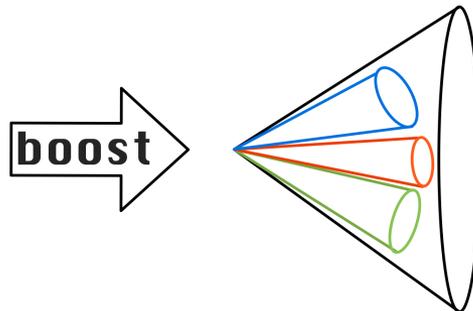
A Word on Jets

- **Jets** = sprays of hadronic particles reconstructed with clustering algorithms into a cone
- Higher mass exclusions for new particles + high energy collisions = high momentum outputs
 - **Constituents**: individual hadrons in jet
 - **Boosting**: collimation of constituents due to high momentum parent
 - **Substructure**: synthesizing correlations between jet constituents to determine particle content in large radius jet

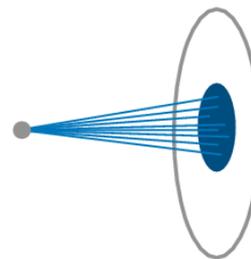
Small-radius jets



Large-radius jet

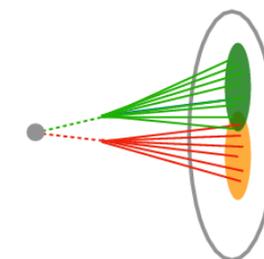


Single q/g



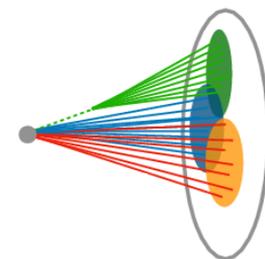
No substructure

$H \rightarrow bb$



2-prong

$t \rightarrow W(qq)b$



3-prong