Tools for Discovery at High Energy Colliders

Julia Gonski

22 September 2021

Rising Stars in Experimental Particle Physics Symposium
University of Chicago

Nevis Laboratories
Columbia University

ATLAS Experiment

Outline

- Introduction to high energy collider physics

1. The ATLAS Experiment & the HL-LHC upgrade
    - LAr calorimeter readout electronics

2. Analysis innovation for new physics searches
    - Long-lived particles
    - Boosted topologies
    - Anomaly detection

3. Future accelerator experiments
    - Forays into e+e-

- Conclusions & ideas

Outline

- Introduction to high energy collider physics

1. The ATLAS Experiment & the HL-LHC upgrade
    - LAr calorimeter readout electronics

2. Analysis innovation for new physics searches
    - Long-lived particles
    - Boosted topologies
    - Anomaly detection

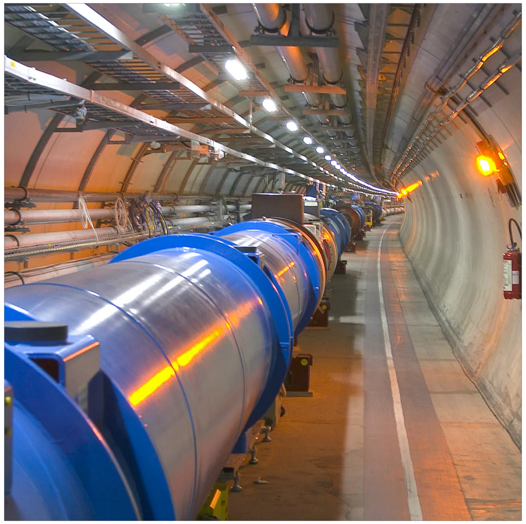
3. Future accelerator experiments
    - Forays into e+e-

- Conclusions & ideas

Collider Experiment Strategy

1. **Large Hadron Collider**: 27km proton synchrotron at CERN responsible for discovery and precise measurement of the Higgs boson

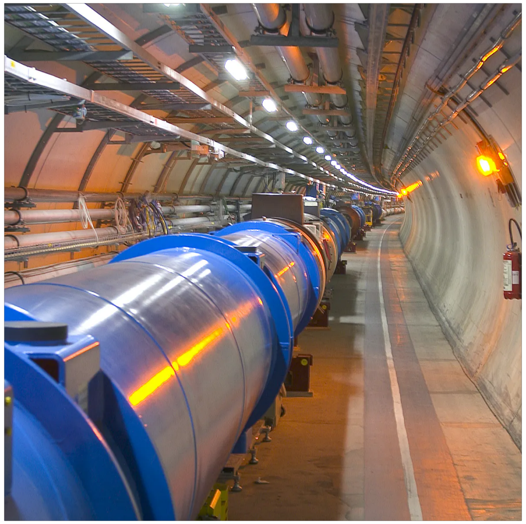
High Energy Accelerator



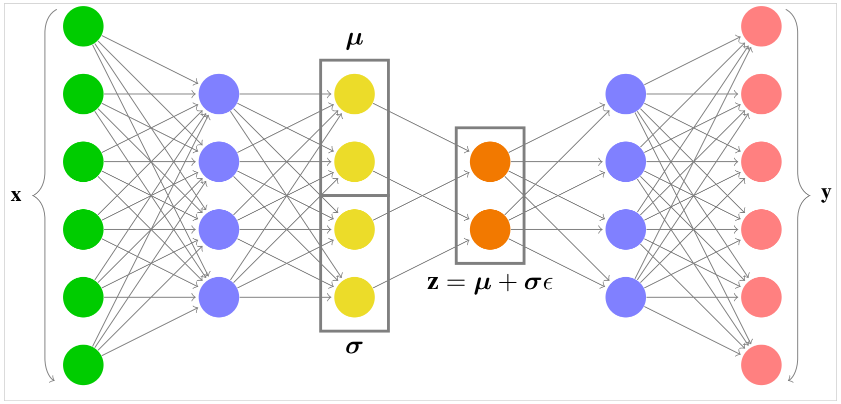
Collider Experiment Strategy

1. **Large Hadron Collider**: 27km proton synchrotron at CERN responsible for discovery and precise measurement of the Higgs boson
2. Lots of high energy data ($\sqrt{s}$ = 13 TeV, 139 fb^{-1}) + many physics goals → improved **analysis strategies** to better explore available datasets

High Energy Accelerator



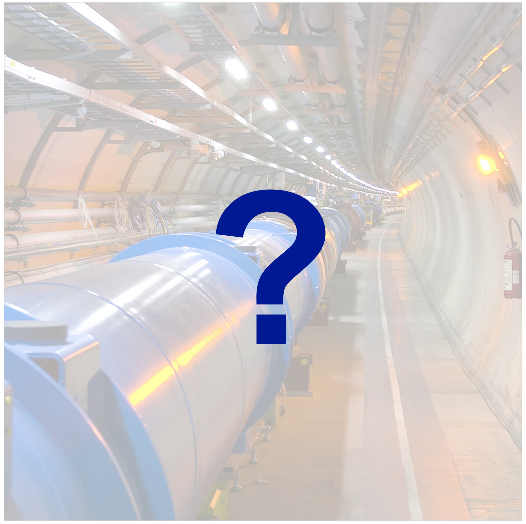
Data Analysis



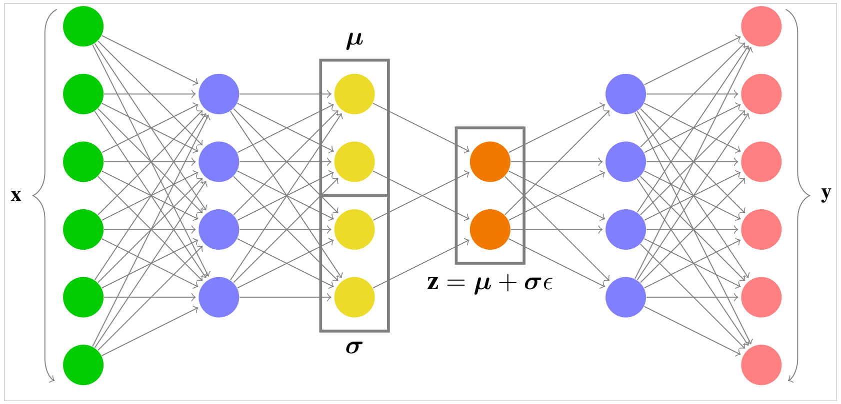
Collider Experiment Strategy

1. **Large Hadron Collider**: 27km proton synchrotron at CERN responsible for discovery and precise measurement of the Higgs boson
2. Lots of high energy data ($\sqrt{s}$ = 13 TeV, 139 fb^{-1}) + many physics goals → improved **analysis strategies** to better explore available datasets
3. Sophisticated analysis informs picture of new physics → motivation for **next generation accelerator**

High Energy Accelerator



Data Analysis



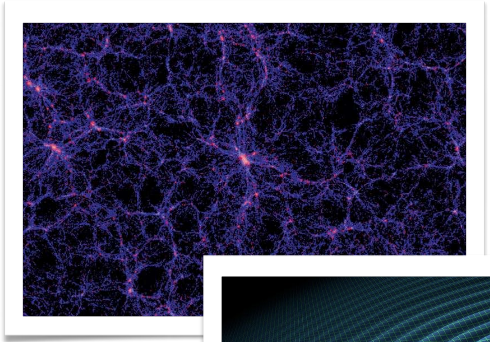
Physics Motivations

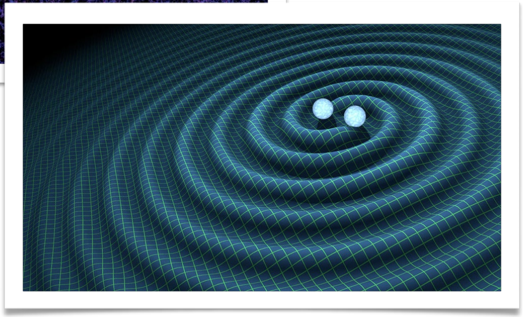
What we need? → **Where to look?**

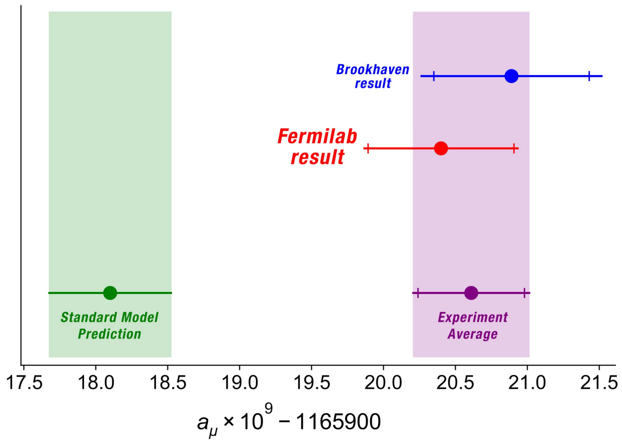
Physics Motivations

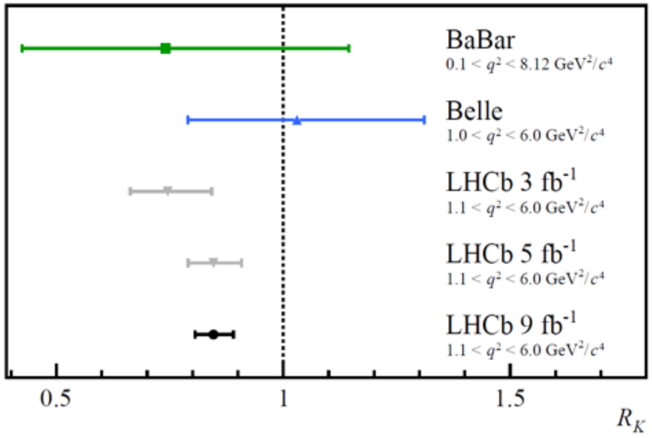
What we need?

- Beyond the Standard Model physics: explanations for **dark matter**, **gravity**…
- Understand recent anomalies/excesses: **Muon g-2**, **LHCb** lepton non-universality… ?

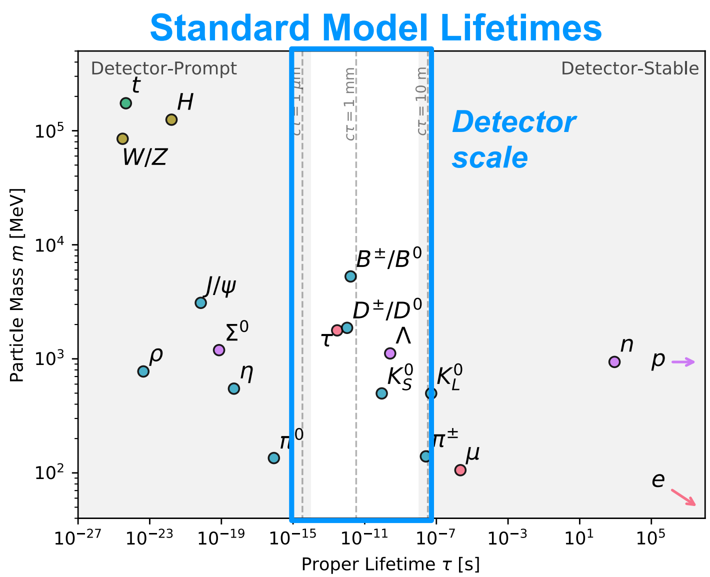


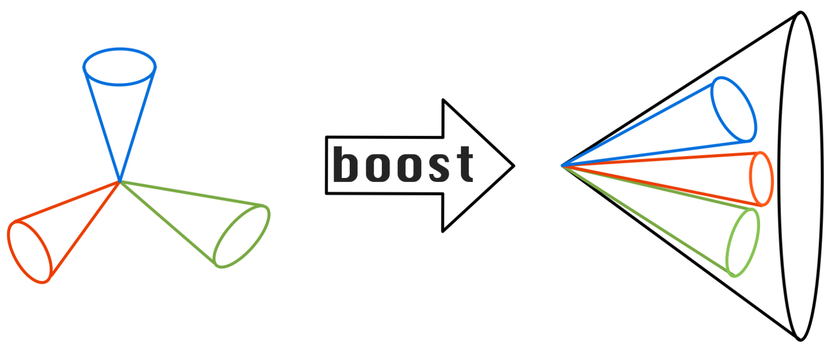






Physics Motivations





Where to look?

- ❖ Cool uncovered/ challenging signatures
  - **Long lived particles**: common in SM + relatively unconstrained at LHC
  - **Boosted topologies**: collimation of particles present in high mass parent decays
- ❖ Anything that is different/unexpected: **anomaly detection**

Outline

- Introduction to high energy collider physics

1. The ATLAS Experiment & the HL-LHC upgrade
    - LAr calorimeter readout electronics

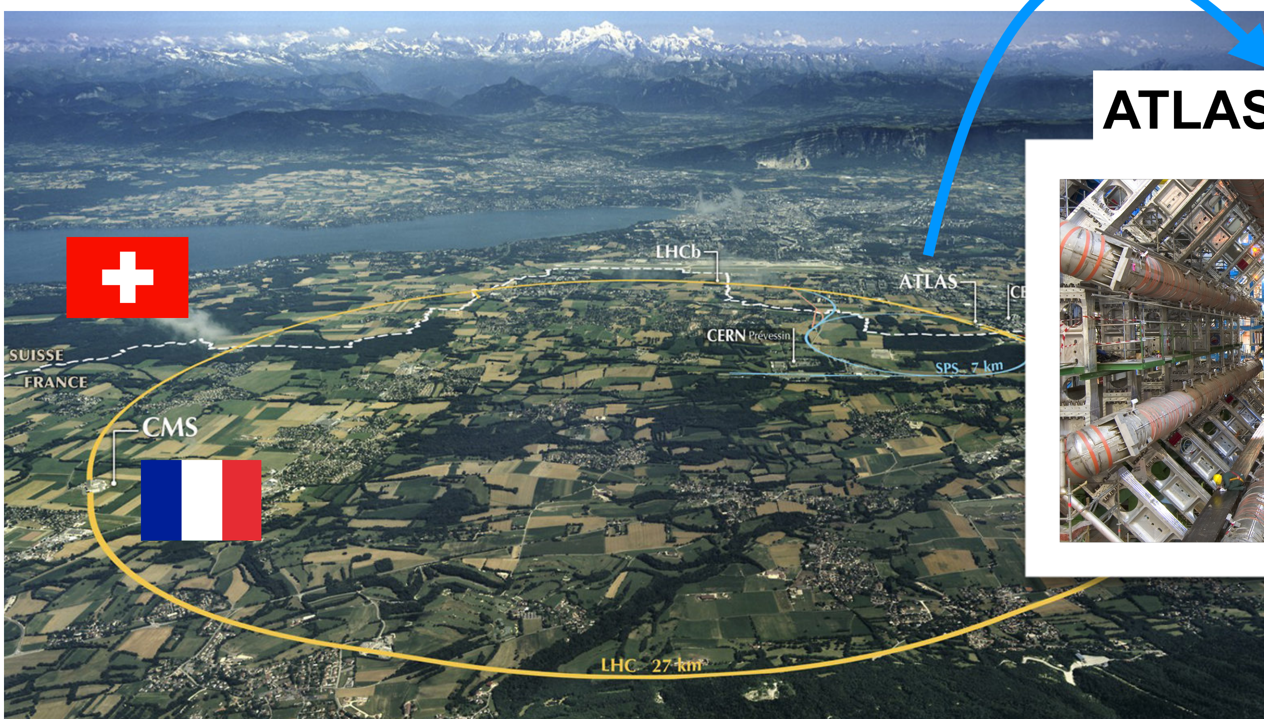
2. Analysis innovation for new physics searches
    - Long-lived particles
    - Boosted topologies
    - Anomaly detection

3. Future accelerator experiments
    - Forays into e+e-

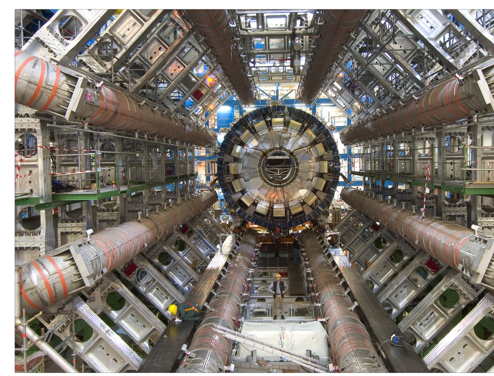
- Conclusions & ideas

Status of the LHC

- **High Luminosity LHC (HL-LHC) in ~2027**: up to 200 simultaneous pp collisions (>20x larger datasets) to give better handle on very rare new physics processes
  - Many detector subsystems getting upgraded or completely new readout to ensure fast and rad-hard electronics



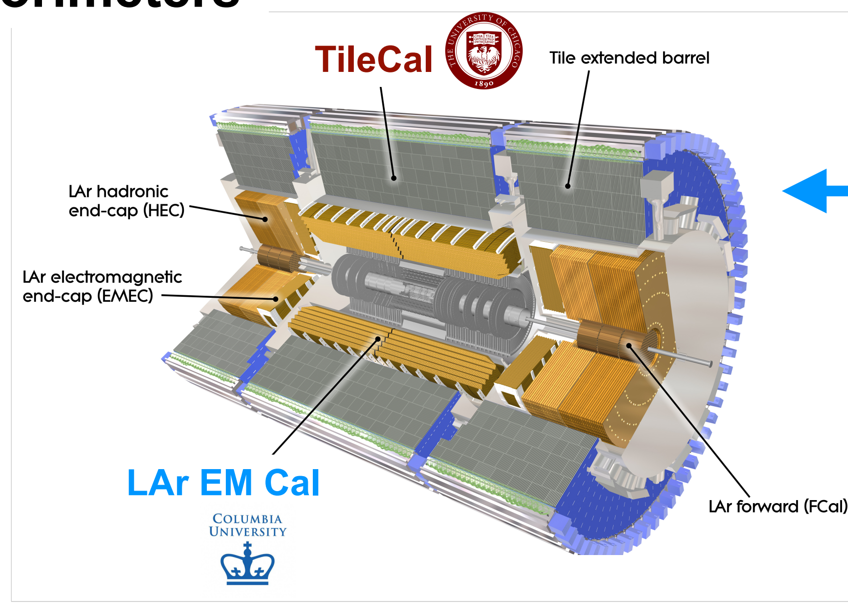
ATLAS Detector



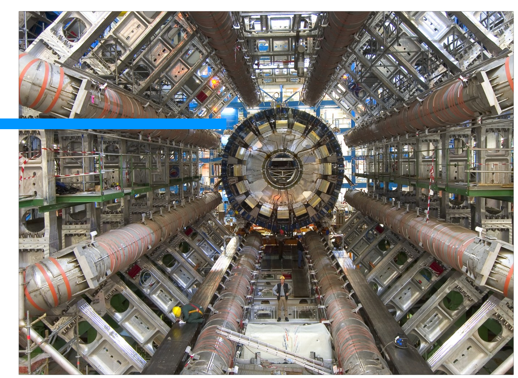
Status of the LHC

- **High Luminosity LHC (HL-LHC) in ~2027**: up to 200 simultaneous pp collisions (>20x larger datasets) to give better handle on very rare new physics processes
  - Many detector subsystems getting upgraded or completely new readout to ensure fast and rad-hard electronics
- **ATLAS calorimeters:** detect energy/timing information of photons, electrons, jets, with readout electronics systems that sample calo cells at LHC bunch crossing frequency of 40 MHz & send digitized pulse off detector

ATLAS Calorimeters



ATLAS Detector



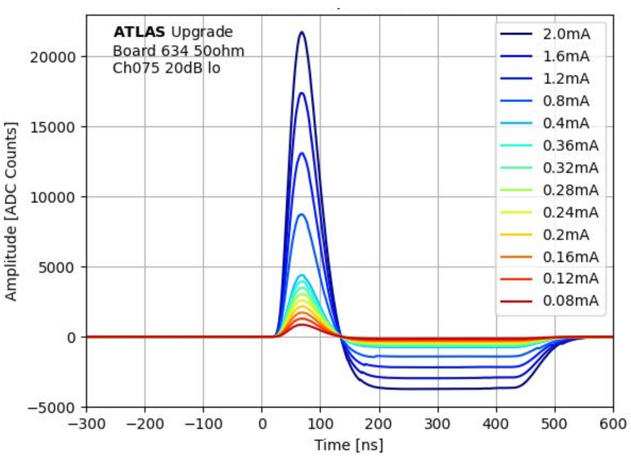
LAr @ HL-LHC: FEB2 Pre-Prototype

- Columbia is responsible for the **ADC** in the LAr frontend readout chain (custom 40 MSPS 14-bit in 65 nm CMOS) and the integration of all custom chips (**Front-End Board 2**)

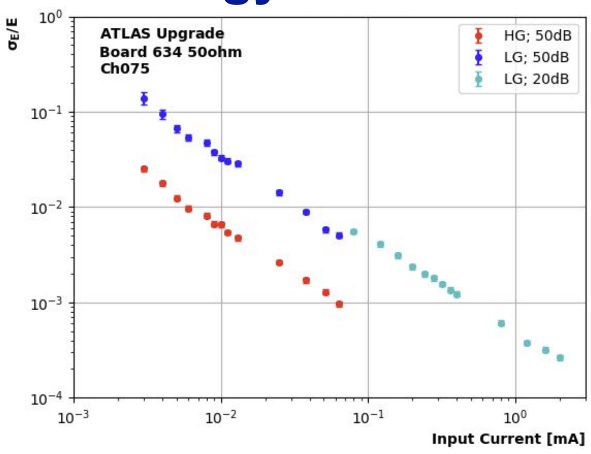
➡ First performance measurements from 32-channel Slice Testboard prototype well within specs!
  - For large pulses, **energy resolution < 0.1% (cf. spec 0.25%), timing resolution ~50 ps**

- **Next steps**:
  - Recently taped out new version of Columbia-UTAustin (COLUTA) ADC
  - Full 128-channel FEB2 prototype in ~2022 + system tests

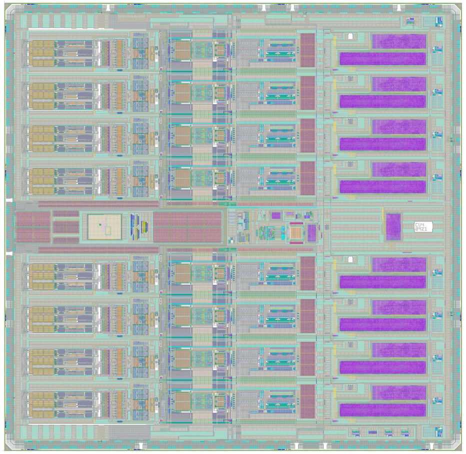
LAr Pulses



Energy Resolution



COLUTAv4



Outline

- Introduction to high energy collider physics

1. The ATLAS Experiment & the HL-LHC upgrade
    - LAr calorimeter readout electronics

2. Analysis innovation for new physics searches
    - Long-lived particles
    - Boosted topologies
    - Anomaly detection

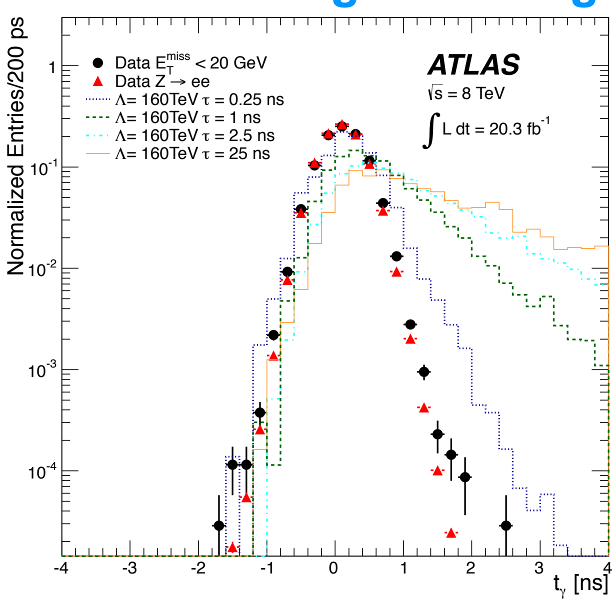
3. Future accelerator experiments
    - Forays into e+e-

- Conclusions & ideas

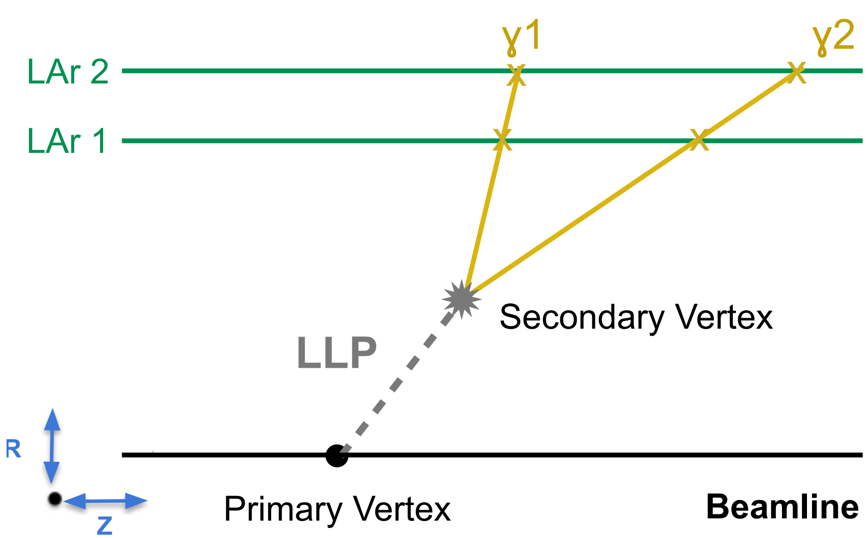
Delayed Photon Searches

- Signature = long-lived particles decaying to **displaced/out-of-time photons:** exploit ~100s ps timing resolution from LAr calorimeter
  1. Unblinded search for Higgs decaying to BSM particles with final state photons: *public soon!*
  2. Finalizing R&D on novel trackless calo-vertexing method searching for displaced diphoton vertices

LAr Calo Signal Timing



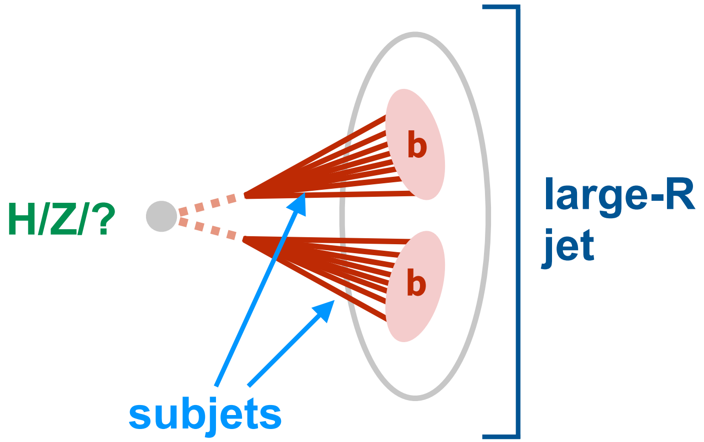
LAr "Calo-Vertexing"



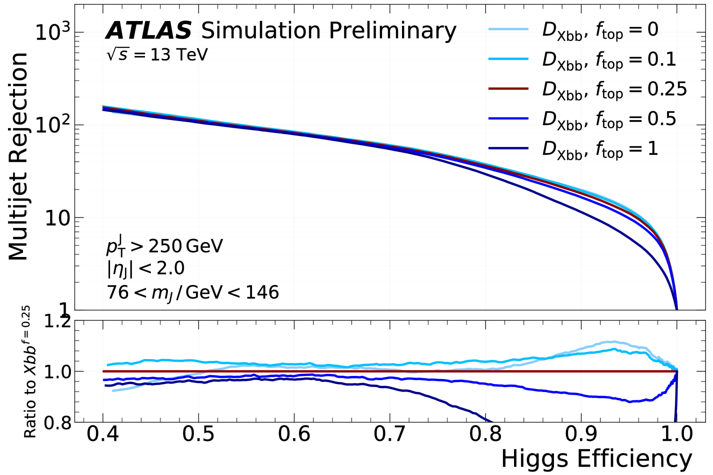
Boosted Higgs Tagging

- 2020: first ATLAS [neural-net based H→bb tagger] for experiment-wide use
  - Highly applicable (bb is most common decay of Higgs)
  - Classifier to distinguish heavy flavor Higgs decays from common backgrounds (multijet, top)
  - Factor x1.5-2 better background rejection w.r.t. previous method

➡ **2021**: calibration of Xbb tagger (scale factors to equate performance in data vs. MC) to be used in upcoming round of ATLAS publications



X→bb vs. Multijet ROC

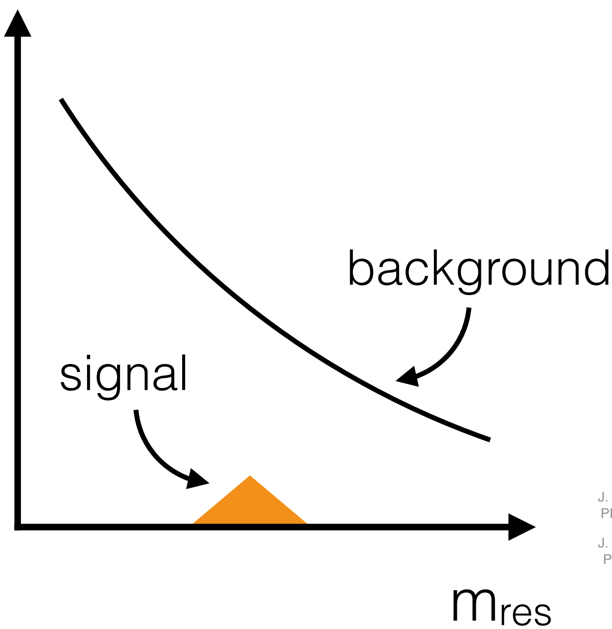


➡ [ATL-PHYS-PUB-2021-035]

Signature-less Searches: Anomaly Detection

- **Anomaly detection (AD)** = identify features of the data that are inconsistent with a background-only model
  - Weakly supervised: train with noisy labels ("signal-contaminated")
  - Unsupervised: train over unlabeled events
- Complementarity of existing model specific efforts (eg. SUSY) with model independent data-driven searches → look under every lamppost!
  - LHC Olympics 2020: cross-experiment/theory "competition" of AD methods

Resonance Search in Dijet Invariant Mass (m_{JJ})



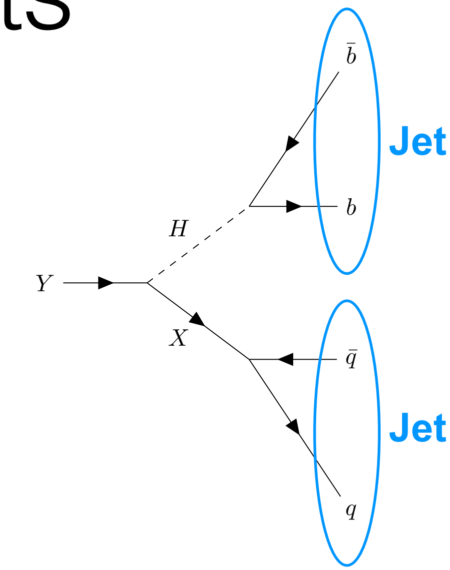
J. Collins, K. Howe, BPN PRL 121 (2018) 241803

J. Collins, K. Howe, BPN PRD 99 (2019) 014038

➡ LHC Olympics [arXiv:2105.09274]

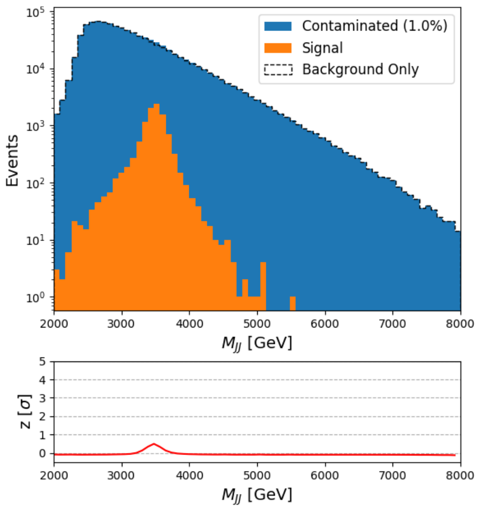
Tagging Anomalous Jets

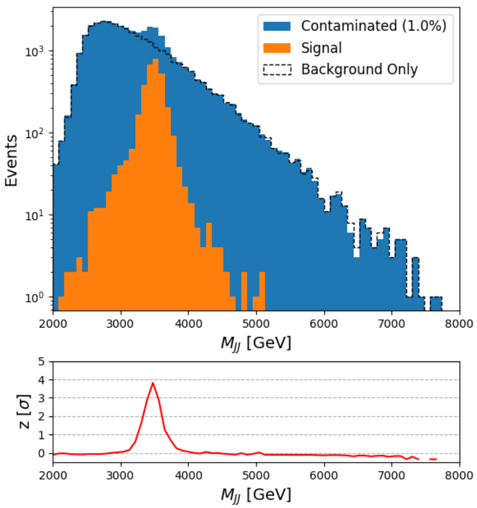
- First adaptation of a **variational recurrent neural network (autoencoder + RNN)** to the tagging of anomalous jets
  - Unsupervised training over jets in data: *no signal model!*
- Application to **Y→XH** search for model-independent tagging of Higgs-associated new bosons
  - ➡Potential to be first unsupervised learning in ATLAS analysis



m_{JJ}, QCD + signal

No selection





Anomaly-selected

➡ A. Kahn, JG, et al [arXiv:2105.09274]

Outline

- Introduction to high energy collider physics

1. The ATLAS Experiment & the HL-LHC upgrade
   - LAr calorimeter readout electronics

2. Analysis innovation for new physics searches
   - Long-lived particles
   - Boosted topologies
   - Anomaly detection

3. Future accelerator experiments
   - Forays into e+e-

- Conclusions & ideas

Beyond the LHC

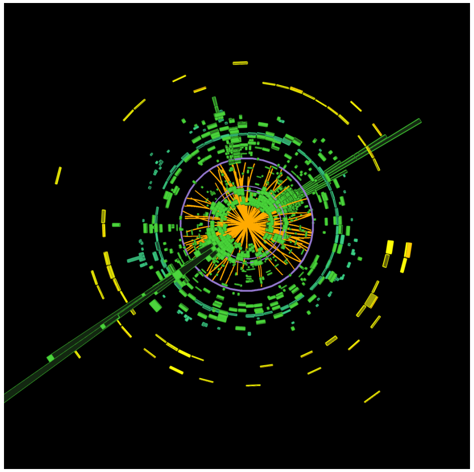
- **Snowmass 2022**: decadal US HEP community planning process to discuss long term physics goals & inform next international experimental plan

- Several e^+e^- colliders (ILC, FCCee, CEPC) are strong candidates for next accelerator

- **How to exploit novel data analysis methods (eg. 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

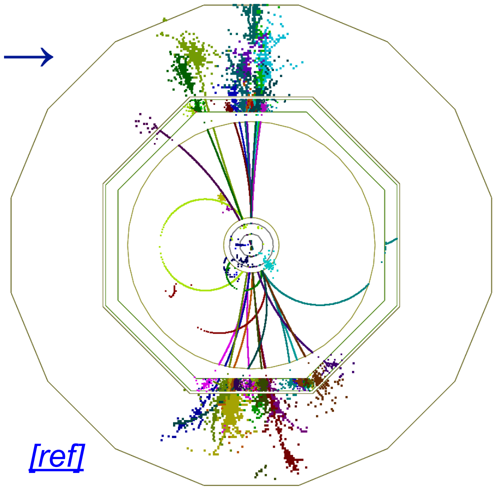
pp → dijet, $\sqrt{s}$=13 TeV LHC

[ref]



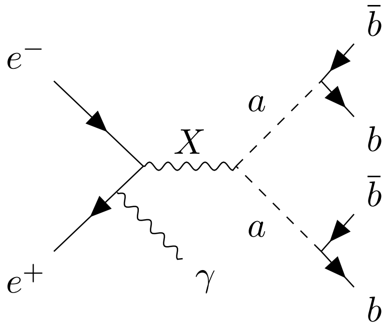
$e^+e^- \to WW \to q\bar{q}q\bar{q}$, $\sqrt{s}$=1 TeV ILC

[ref]

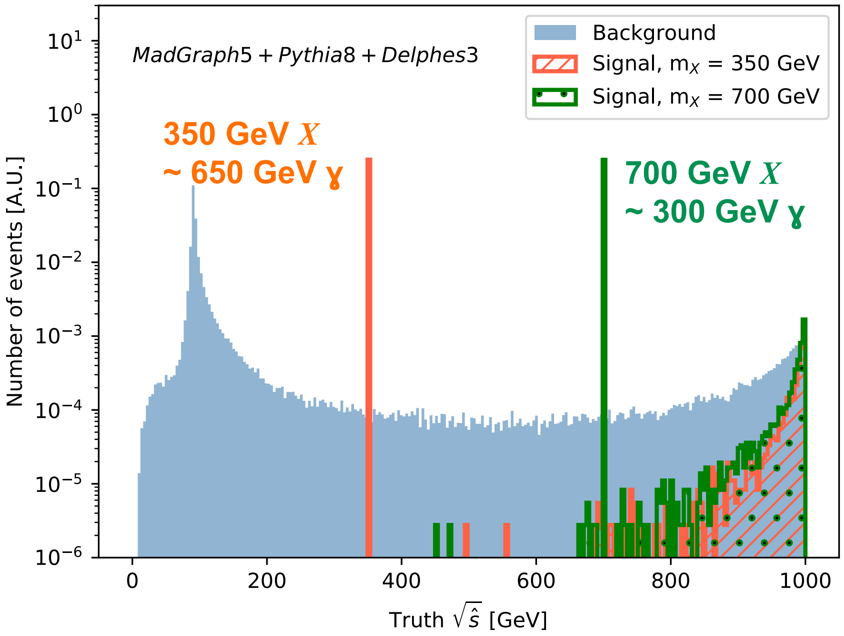


Anomaly Detection in e+e- Collisions

- Radiative return: “scan” new particle masses with ISR photons, à la dijet invariant mass bump hunts



e+e- √s = 1 TeV

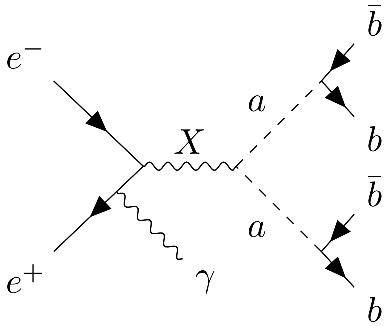


➡ JG, et al [arXiv:2108.13451]

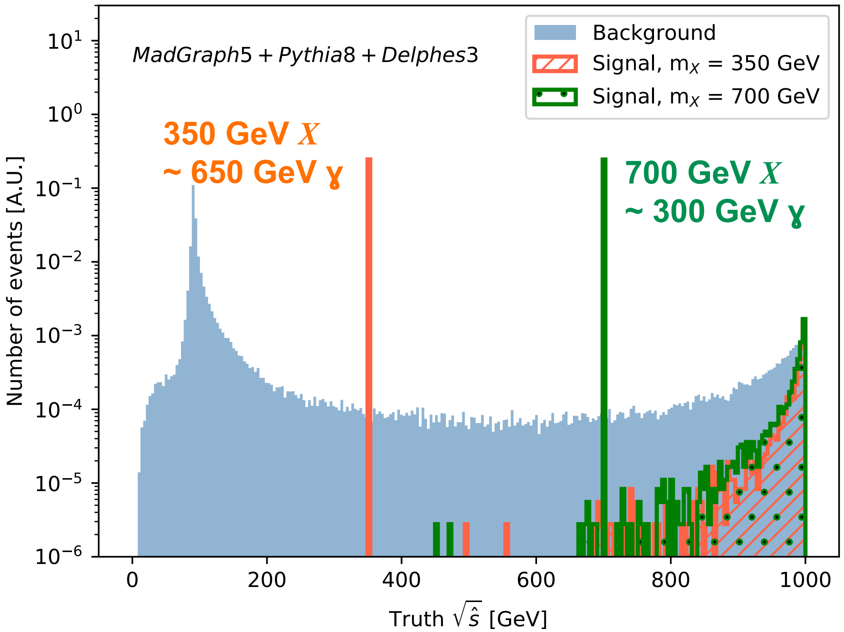
Anomaly Detection in e+e- Collisions

- Radiative return: "scan" new particle masses with ISR photons, à la dijet invariant mass bump hunts
- Weakly supervised learning used to leverage sideband data in S vs. B classification (high-dim PFN inputs)

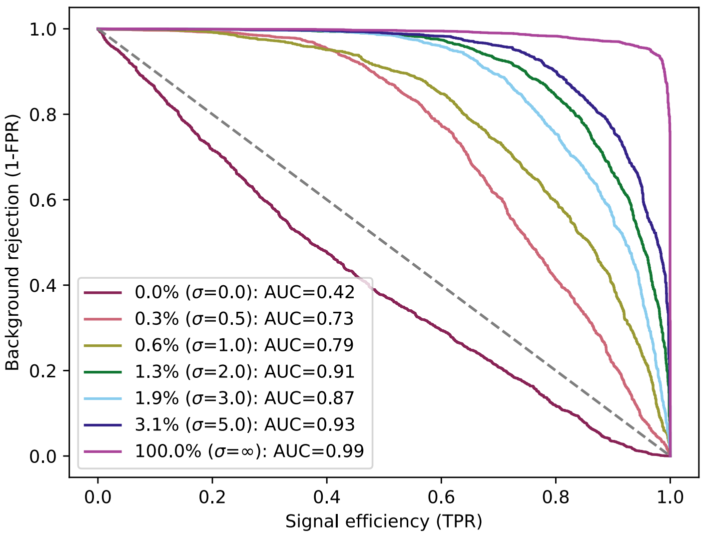
➡ **Gain sensitivity to signal contaminations down to 0.3%!**



e+e- √s = 1 TeV



ROC: X=700 GeV vs. bkg



➡ JG, et al [arXiv:2108.13451]

Conclusions

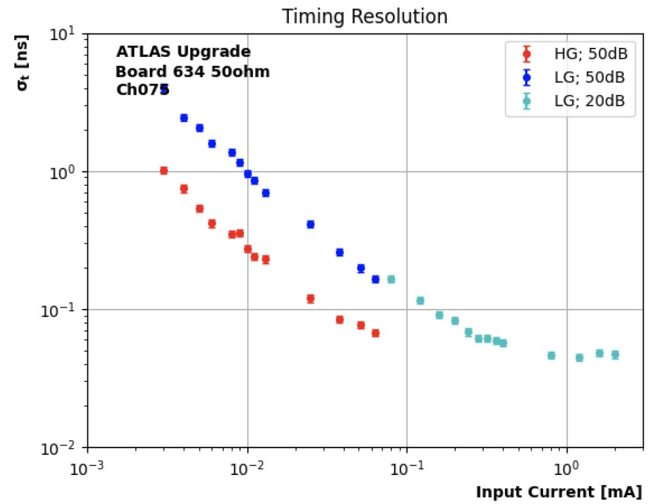
- Energy frontier provides unique reach towards beyond the SM physics prospects
  1. Maximizing utility of LHC through detector upgrades
  2. Better analysis techniques for broad new physics sensitivity
  3. Motivating and brainstorming for next accelerator

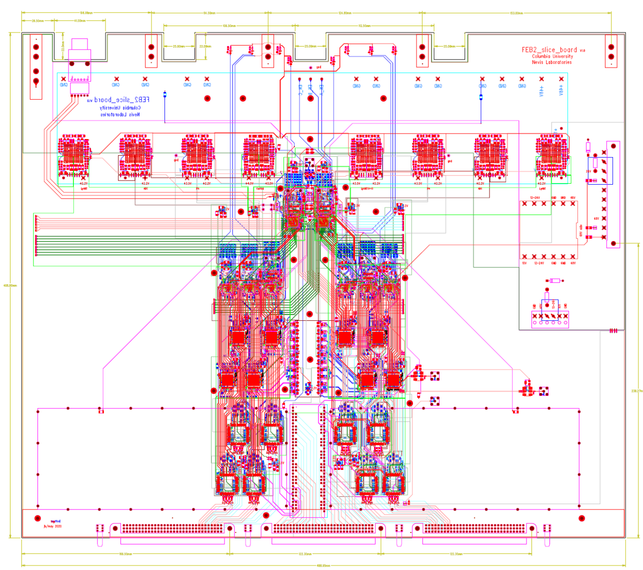
- What I'm excited about:
  - ✓ Dark matter: exploiting ML/AD for challenging dark jet signatures
  - ✓ Readout development of next-generation calorimeters
  - ✓ Snowmass (& beyond) connectivity and community building

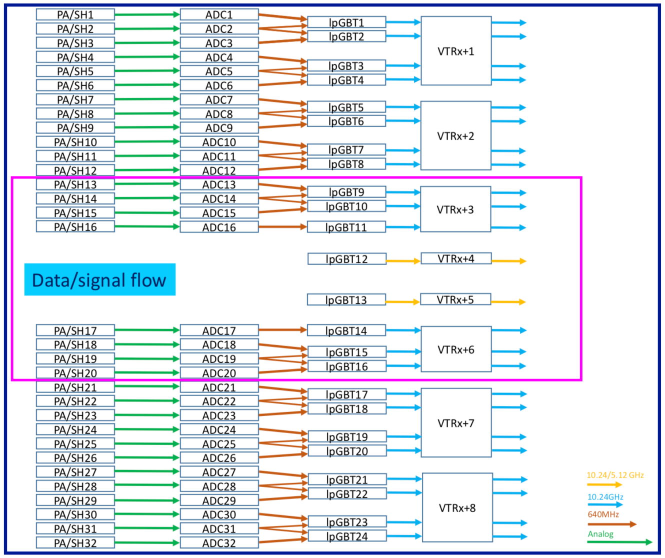
Backup

Slice Testboard

- Done with a sequence of increasingly complex pre-prototypes with new generations of custom ASIC
- ***2019***: Analog Testboard (2 channels)
- ***2020-21***: Slice Testboard (32 channels) with 3rd pre-prototypes of PA/S + ADC + v0 prototype of lpGBT
- ***2022***: Full 128 channel FEB2 prototype

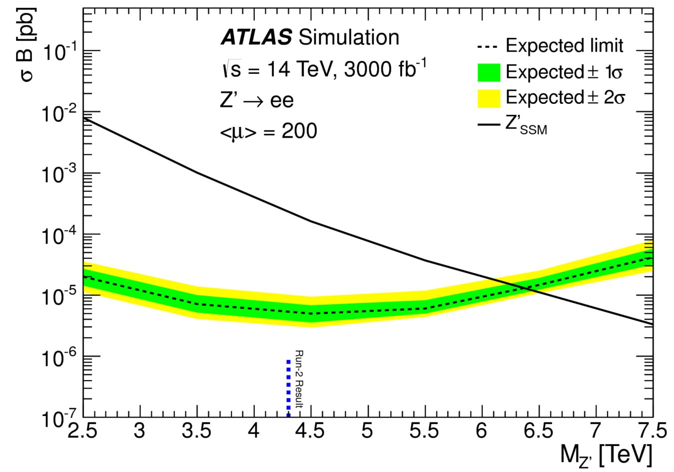




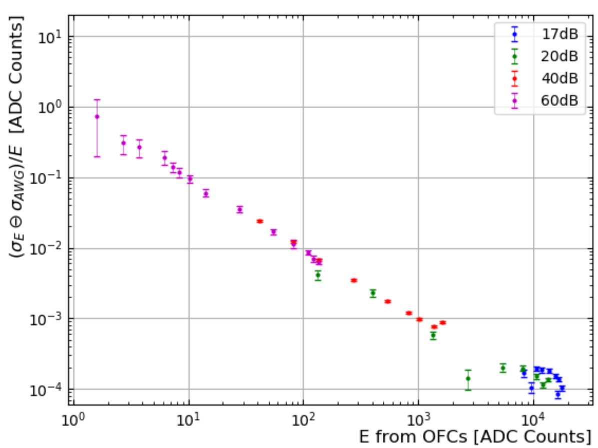


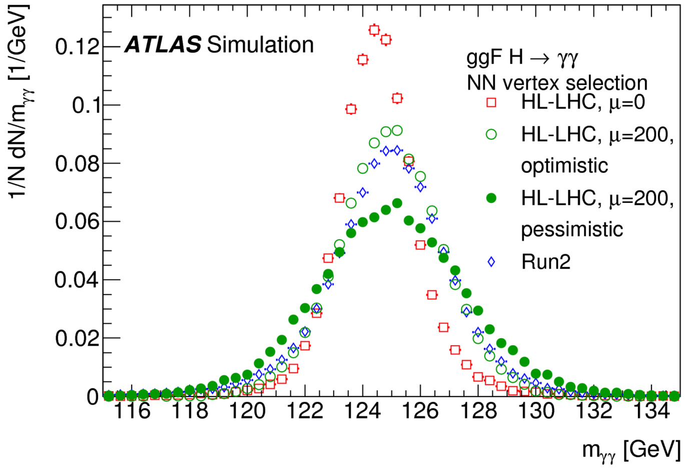
CV3 Physics Requirements

- Dynamic range = 14 bits
  - LAr total dynamic range is 16-bits, set between noise levels and (at high end) by discovery reach for heavy Z' to ee
  - Handled by 2 gain scales, to avoid overlap in Higgs mass range, and to reduce total data flow to LASP
  - Therefore, each gain scale needs 14-bit dynamic range
- Precision = 11-bits (full scale, 8 MHz)
  - Does not significantly degrade LAr energy resolution (eg. maintains Higgs mass resolution, which is critical for precision Higgs studies and HH search; makes limited contribution to constant term for large energies)



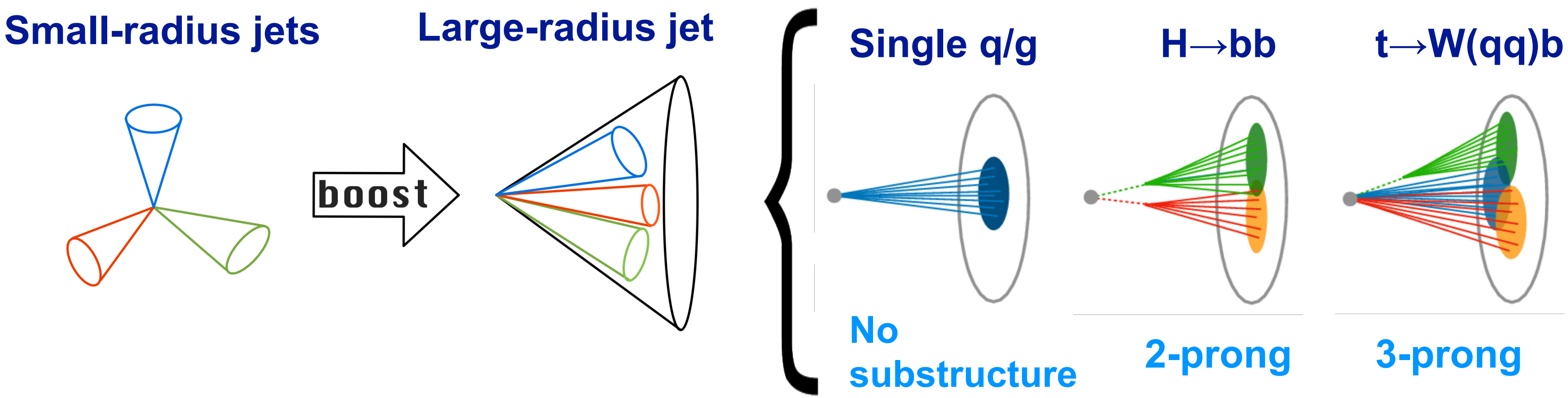
Pulse Energy Resolution





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

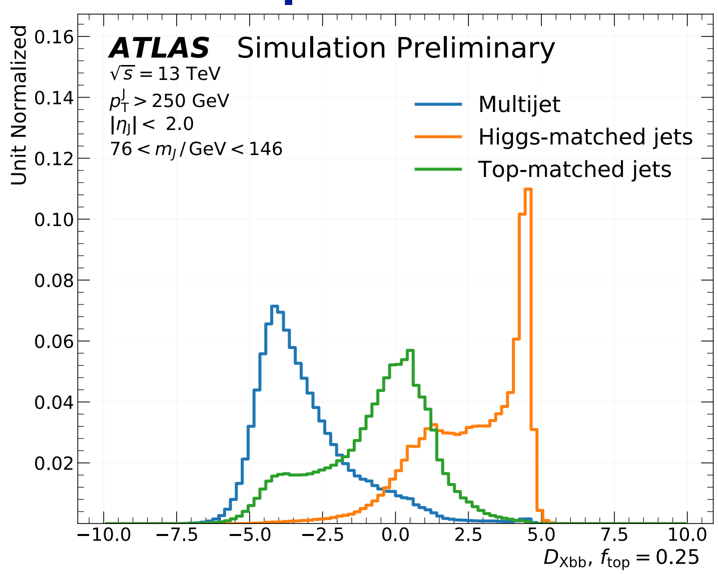


NN-based X→bb Tagging

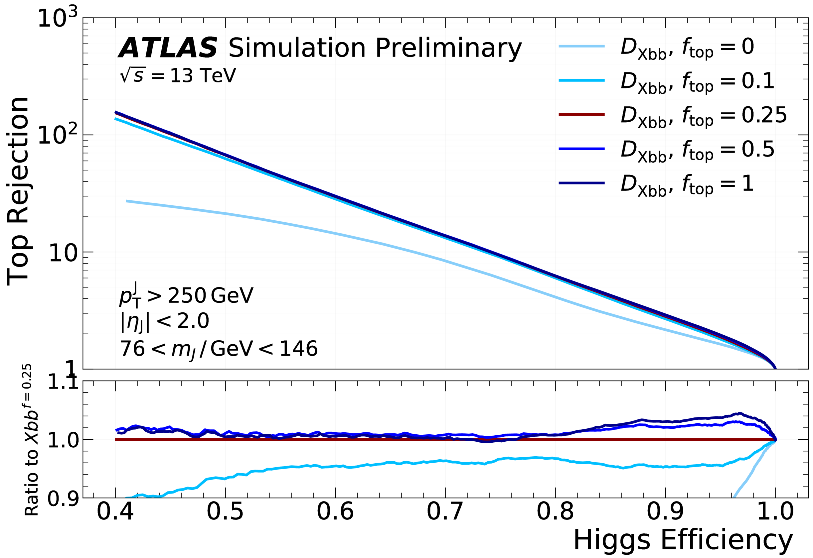
- **Training:** 6 fully connected 250-dim hidden layers, RELU activation, pT reweighted to QCD spectrum
- Inputs = large-R jet pT, η + up to 3 subjet DL1r (small-R b-tagging) scores
- Output: probabilities for 3 classes (multijet, top, Higgs jets)

$$D_{\mathrm{Xbb}} = \ln \frac{p_{\mathrm{Higgs}}}{f_{\mathrm{top}} \cdot p_{\mathrm{top}} + (1 - f_{\mathrm{top}}) \cdot p_{\mathrm{multijet}}}$$

NN Output Discriminant



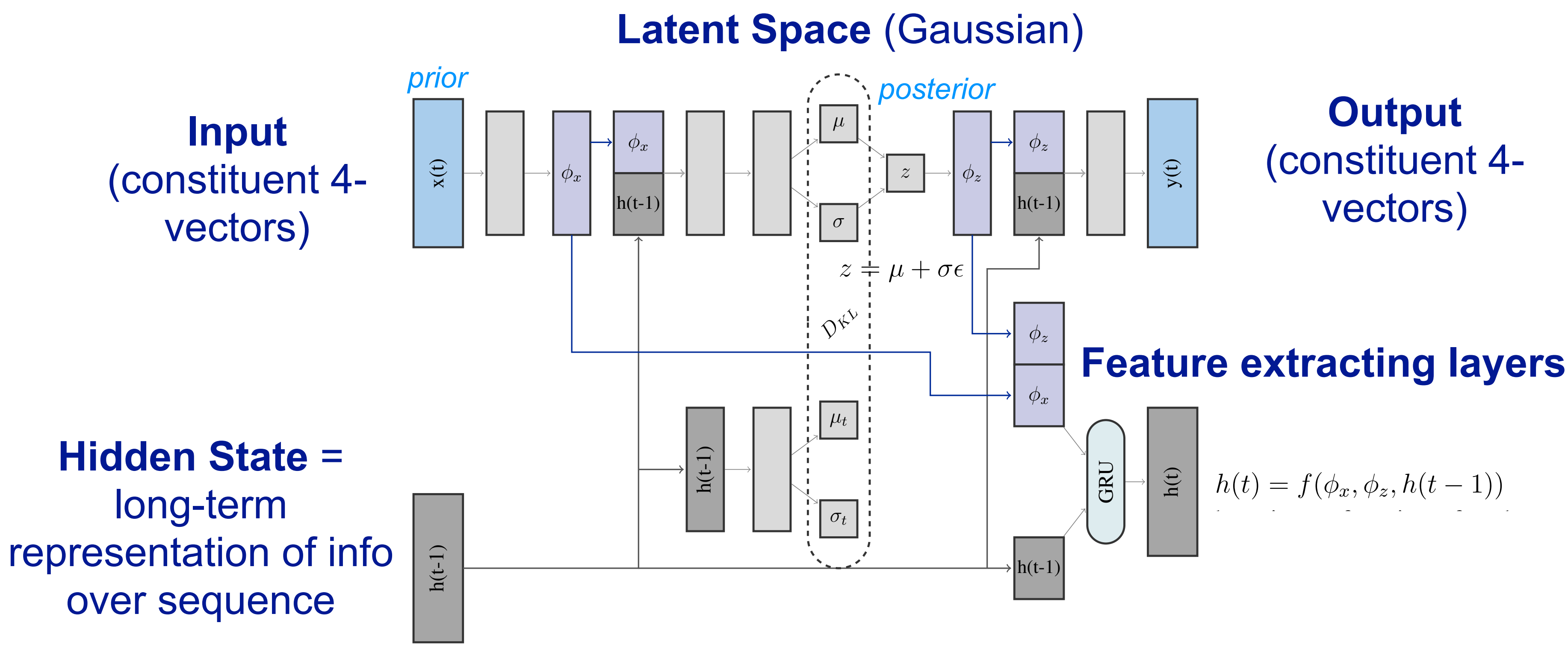
Top ROC



VRNN Architecture

- Train directly on data (avoid data/MC discrepancies in QCD)
- Merge sequence modeling nature of RNN with variational inference capability of VAE

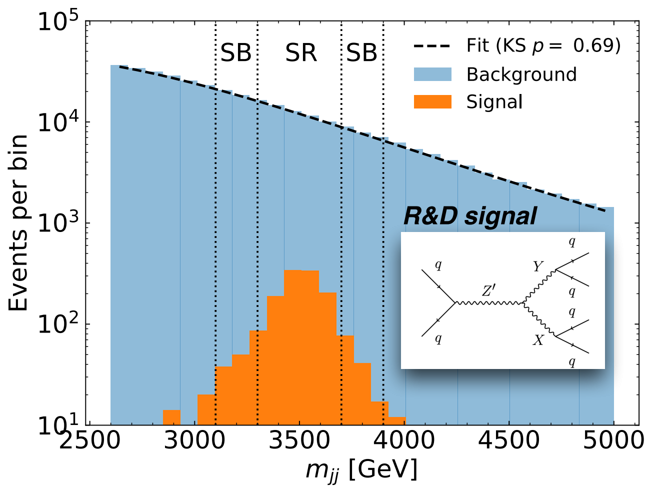
$$\mathcal{L}(t) = |\mathbf{y}(t) - \mathbf{x}(t)|^2 + \lambda D_{KL}(z||z_t)$$



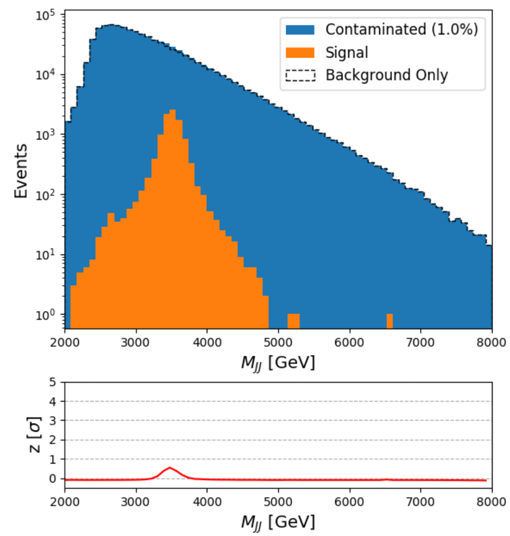
VRNN LHCO Results

- [LHC Olympics dataset](): Pythia generated + Delphes detector simulation (no pileup)
- **Signal:** 3.5 TeV Z' → 500 GeV X + 100 GeV Y
  - Two substructure hypotheses: 2-pronged and 3-pronged X/Y decays
- Reconstruction = two large-radius (R=1.0) jets
  - Trigger: 1 large-R jet with p_T > 1.2 TeV

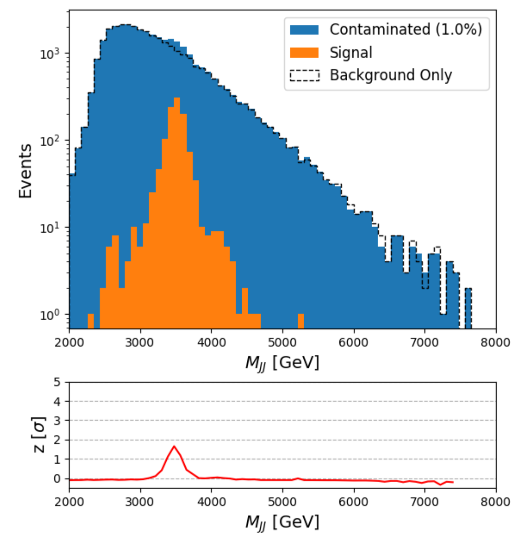
➡arXiv:2101.08320



No cut



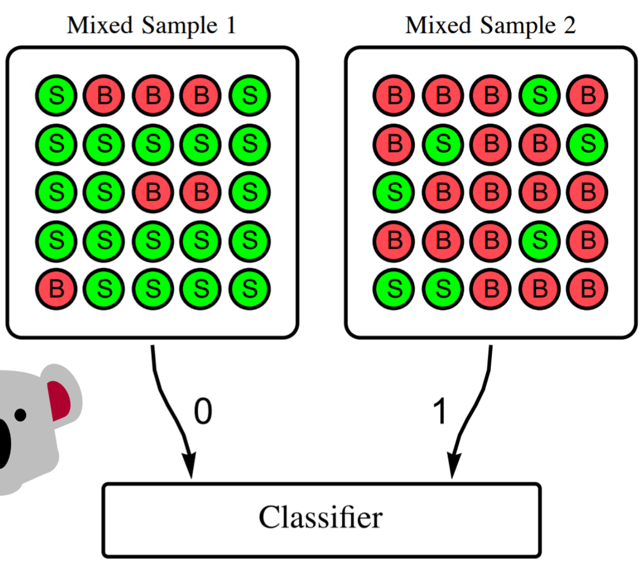
Event Score > 0.65

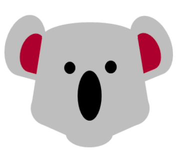


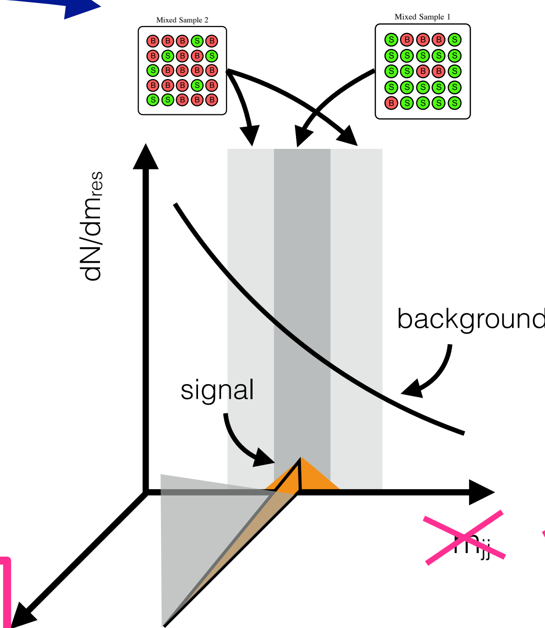
Data-Driven/Weakly Supervised (CWoLa)

- 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…







√s

Features for training CWoLa classifier

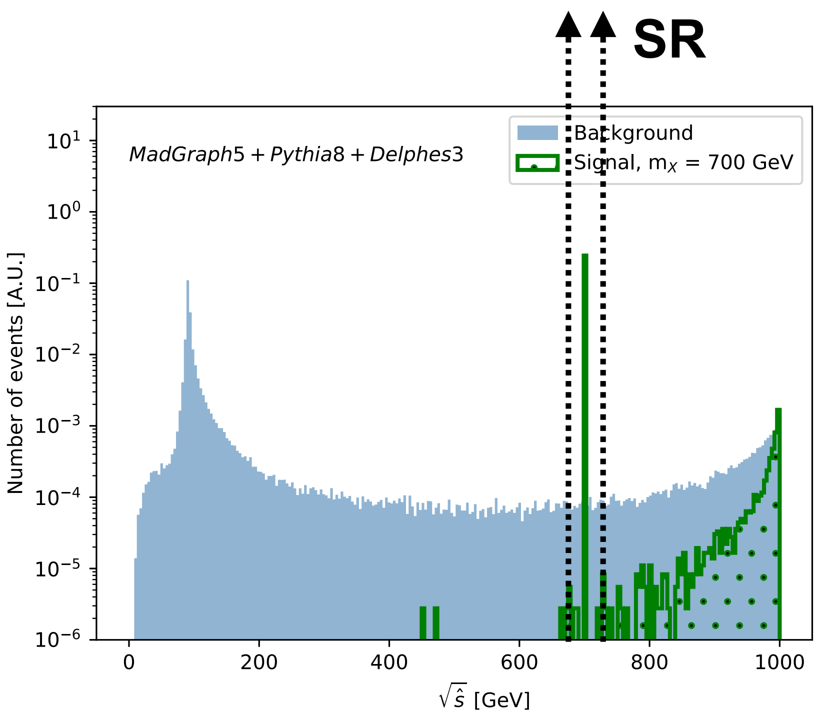
150 per event!

[1708.02949](https://arxiv.org/abs/1708.02949), [1805.02664](https://arxiv.org/abs/1805.02664)

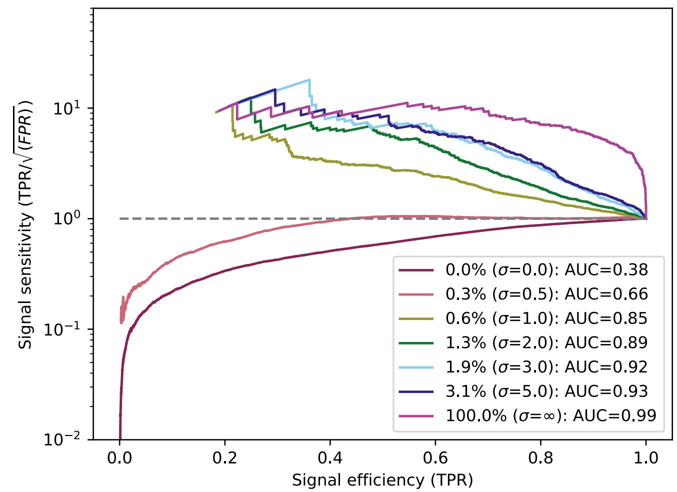
e+e- AD Results

- 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: σ=0.0, 0.5, 1.0, 2.0, 3.0, 5.0, and ∞ (eg. all S vs. all B)

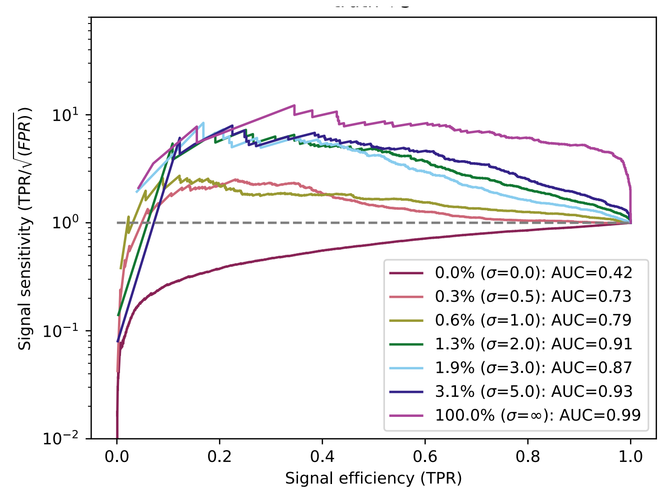
➡ **Significance Improvement Characteristic (SIC): can enhance a 0.6% signal contamination from 1.0σ to ~10.0σ**



Semi-supervised



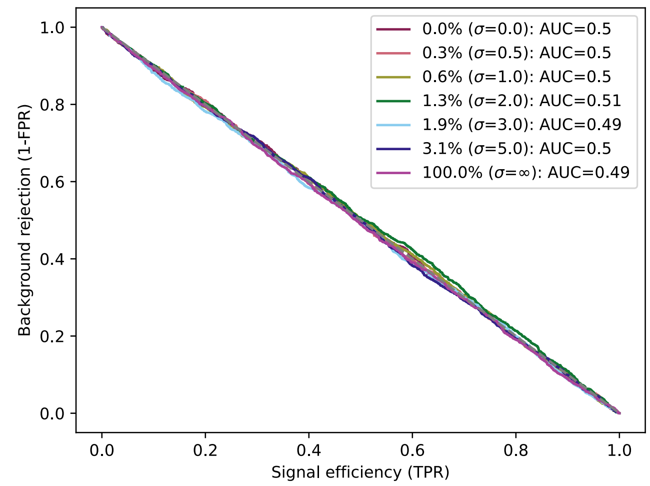
Weakly Supervised



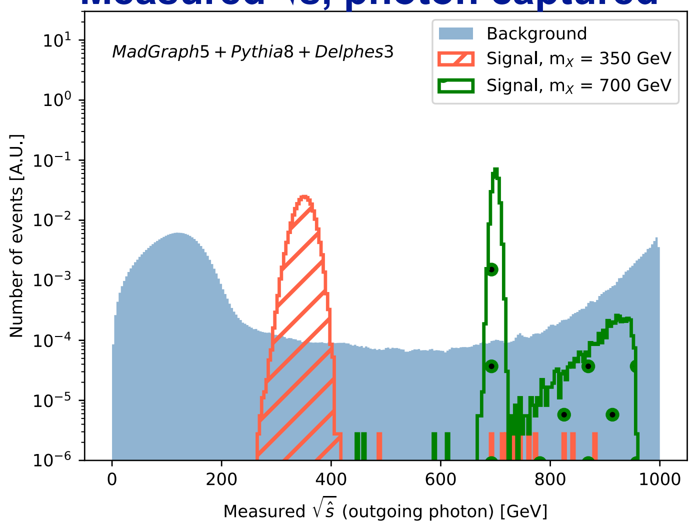
e+e- AD Considerations

- Normalization of inputs: CWoLa-trained classifier tested on background in SB vs. background in SR has minimal √s correlation (right)
- Detector features such as mass resolution and forward acceptance have strong impact on radiative return AD analyses: investigating different √s reconstruction measures to understand dependency and inform e^+e^- detector design

ROC: Bkg in SB vs. Bkg in SR



Measured √s, photon captured



Measured √s, photon lost