UNIVERSITY OF OREGON

BERKELEY LAB

Non-resonant Anomaly Detection with background extrapolation

Kehang Bai, Radha Mastandrea, Ben Nachman

Physics Seminar, University of Ljubljana, 12/21/2023

Based on arXiv:2311.12924

The need for signal-agnostic searches in hadronic final states

ATLAS hadronic resonance search summary.

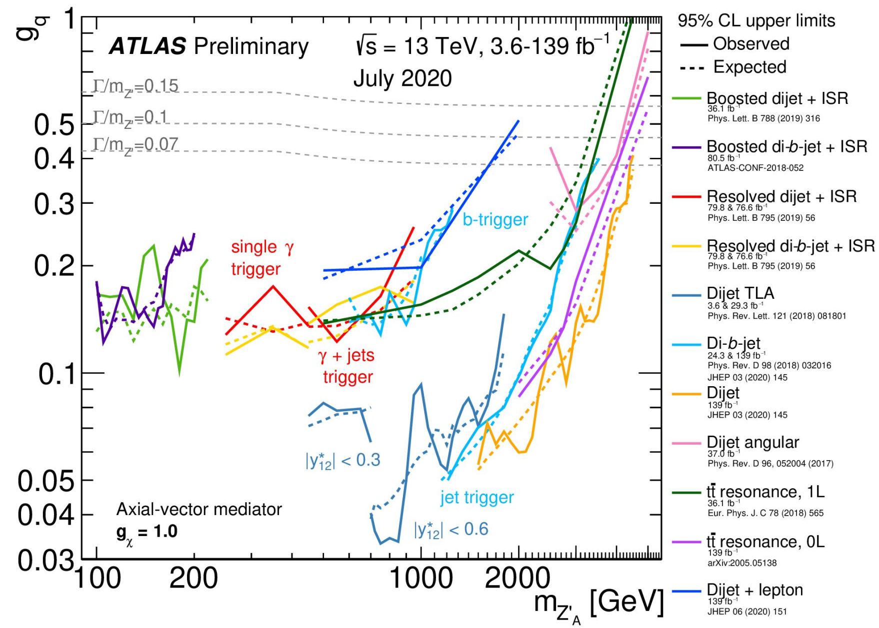
- Only an interpretation in a specific Z' model.
- There are many more models, each with a large parameter space.

It is impossible to scan all models and phase space.

Model-independent searches are complementary,

➔ **but also necessary!**

ML-based Anomaly Detection (AD) provides us with more sophisticated tools for model-independent searches.



[ATLAS exotics results]

Requirements for complete anomaly detection

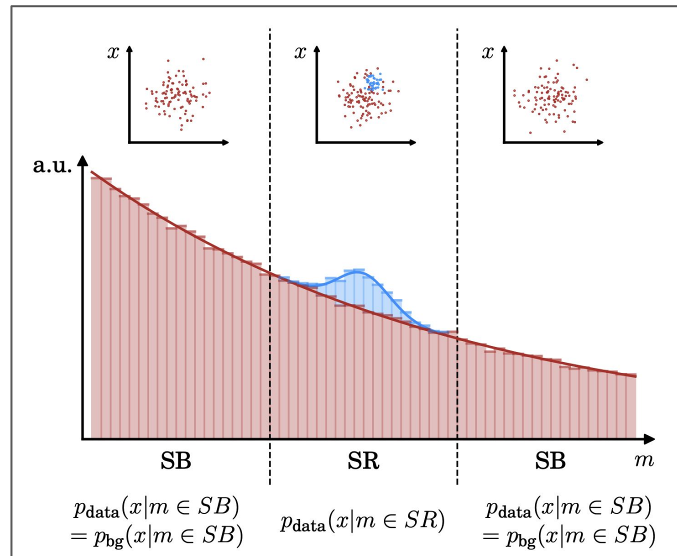
AD in particle physics looks for "outliers" or "overdensities" in the SM background without model assumptions.

Successful background estimation + Sensitive to signal

Overdensity AD example:

Classifying Anomalies THrough Outer Density Estimation (CATHODE) [2109.00546]

Weakly-supervised.



Current landscape of overdensity anomaly detection

Many proposals and all data results have targeted resonant signals by interpolating background from a sideband.

Background interpolation from sideband + **Resonant signal**

Existing methods for background interpolation:

- MC-to-data reweighting (SALAD [2001.05001])
- Density estimation using generative models (CATHODE [2109.00546])
- MC-to-data feature morphing (FETA [2212.11285])
- Other variations…

Available technologies

➔ **A binary classifier for reweighting**

Neural networks ($f(x)$) can learn to approximate the likelihood ratio.

$$w(x) \equiv \frac{f(x)}{1-f(x)} = \frac{p(x|\text{data})}{p(x|\text{simulation})} \begin{array}{l} \longrightarrow \text{Class 1} \\ \longrightarrow \text{Class 0} \end{array}$$

Available technologies

➔ **A binary classifier for reweighting**

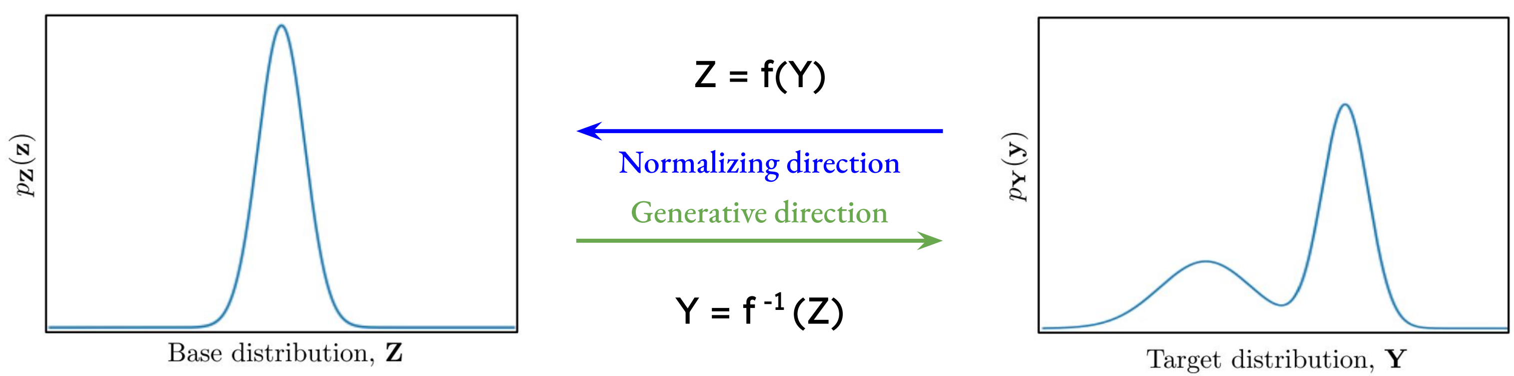
Neural networks (*f(x)*) can learn to approximate the likelihood ratio.

$$w(x) \equiv \frac{f(x)}{1-f(x)} = \frac{p(x|\text{data})}{p(x|\text{simulation})}$$

→ Class 1

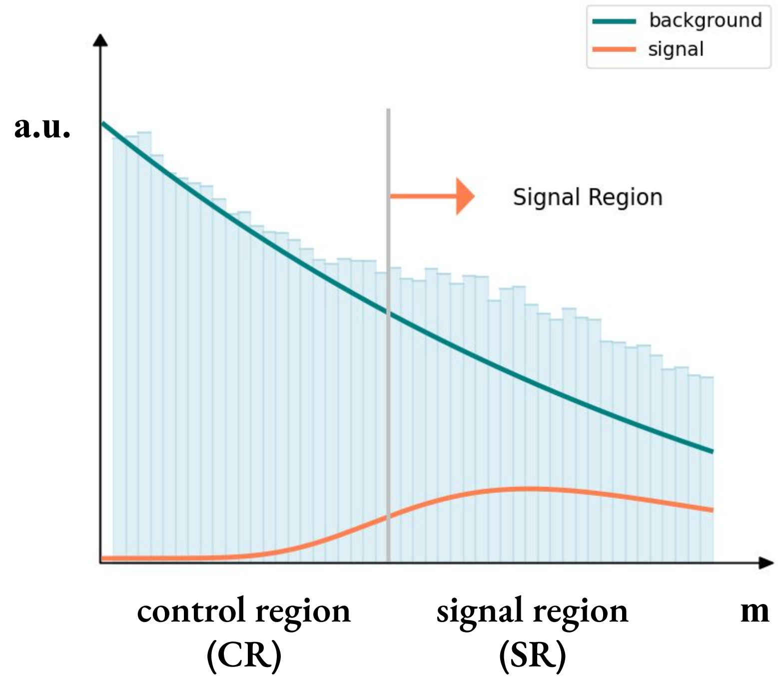
→ Class 0

➔ **Normalizing flow for generating and morphing**



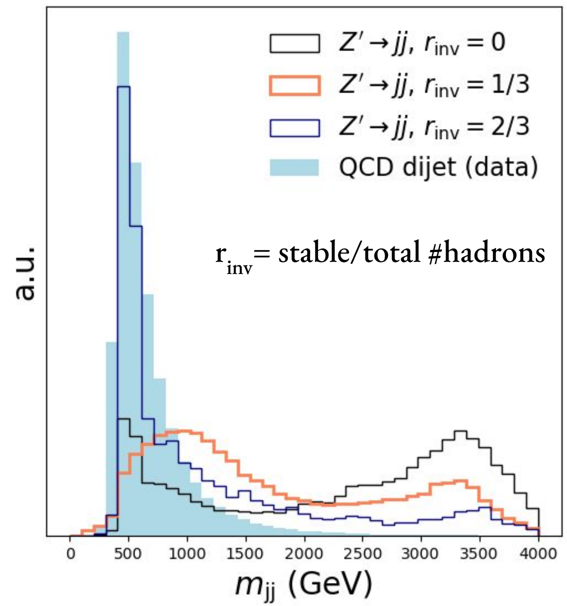
This proposal

Background extrapolation from one side ➕ non-resonant signal



Non–resonant signals

Case 1: large missing energy in the final states of a resonant production. Example here: semi-visible jets from Z'.



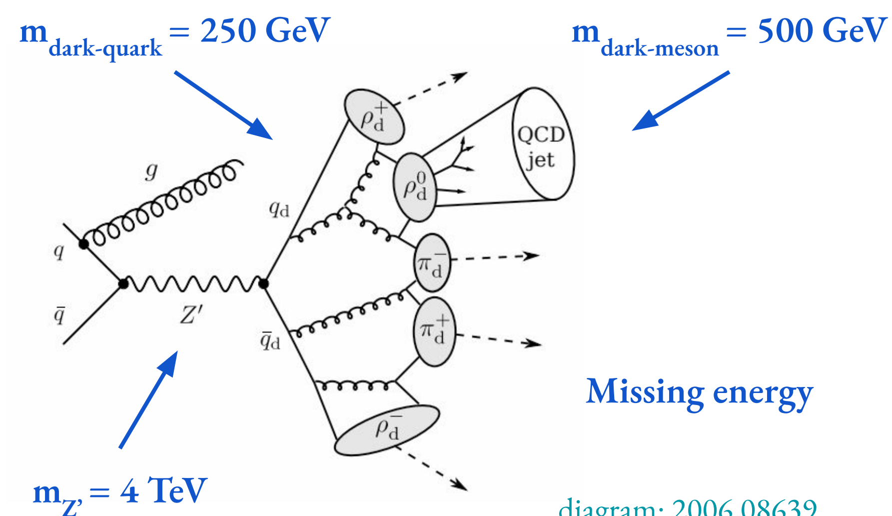


diagram: 2006.08639

Case 2: off-shell effects from heavy particles. Example: modifications to SMEFT coefficients, but not explored here.

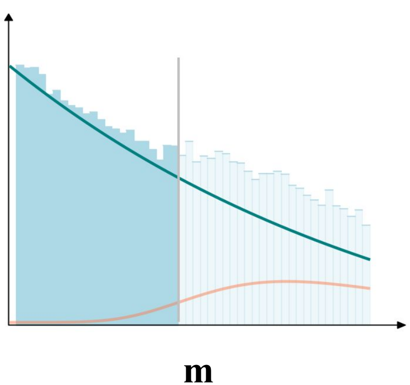
The idea

➔ **Why don't we just extrapolate with existing methods?**

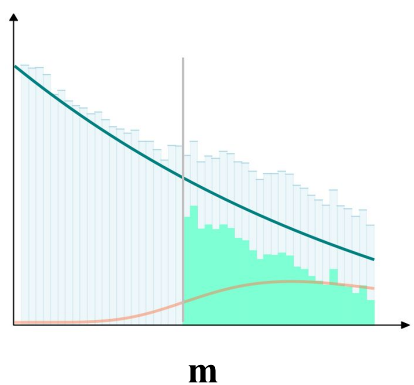
In principle we could directly use MC-to-data reweighting.

However, density estimation with generative models are **less robust without a sideband.**

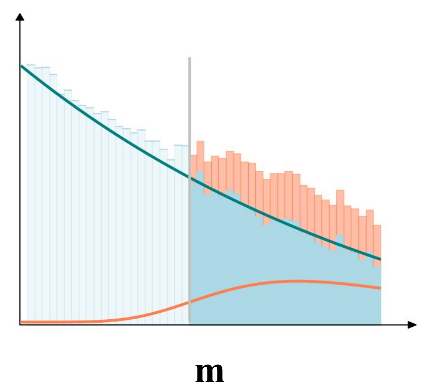
❶ Learn P(x|m) in CR.



❷ Predict P(x) in SR.



❸ Classification in SR



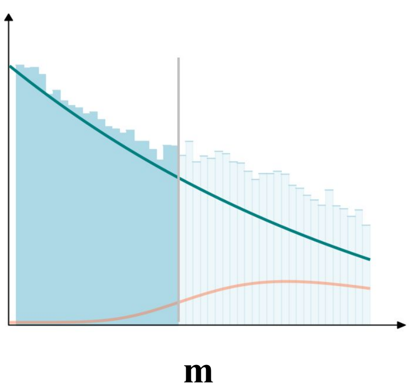
The idea

➔ Combine generative models and ML reweighting!

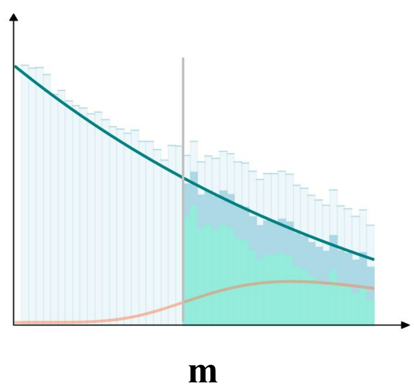
Reweight conditional variables from MC to data in SR.

$$
\begin{aligned}
P_{\text{data}}(x, m) &= P_{\text{data}}(x|m)P_{\text{data}}(m) \\
&= P_{\text{data}}(x|m)P_{\text{MC}}(m)\boxed{\frac{P_{\text{data}}(m)}{P_{\text{MC}}(m)}} \\
&= P_{\text{data}}(x|m)P_{\text{MC}}(m)\boxed{w(m)}
\end{aligned}
$$

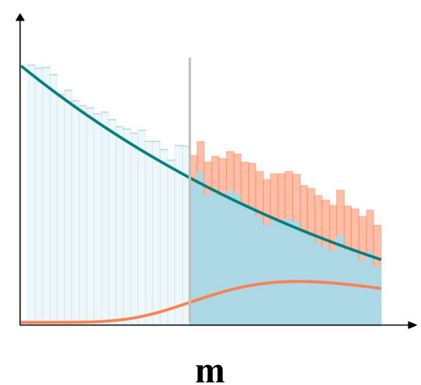
❶ Learn P(x|m) in CR.



❷ Predict P(x) in SR with reweighting.



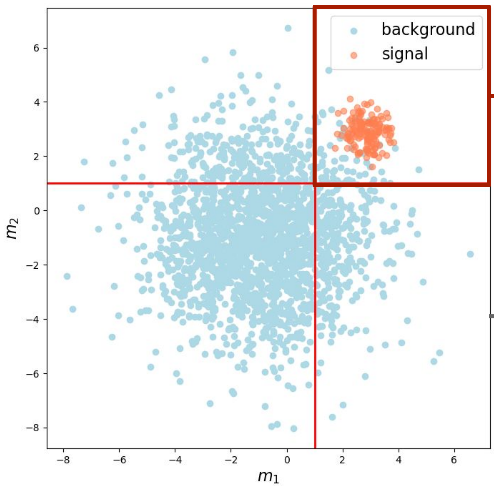
❸ Classification in SR



First, let's take a look at a toy example

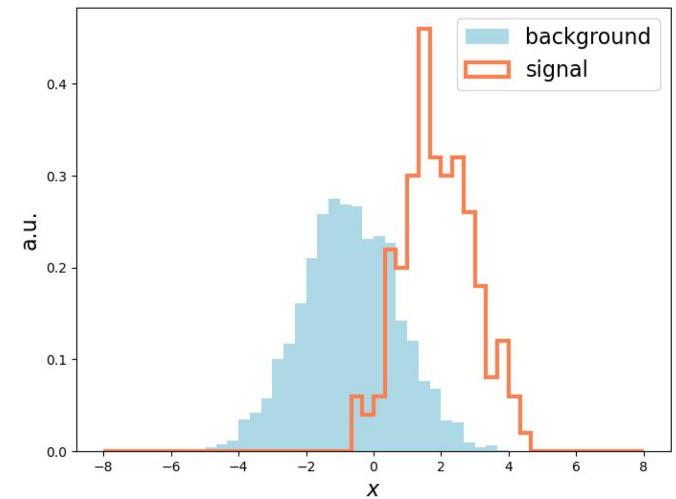
Context variables m1 and m2, used to define SR & CR.

A feature variable x.



Signal Region (SR)

Control Region (CR)



Extrapolate the background using Reweight, Generate, and Morph methods.

Anomaly detection methods

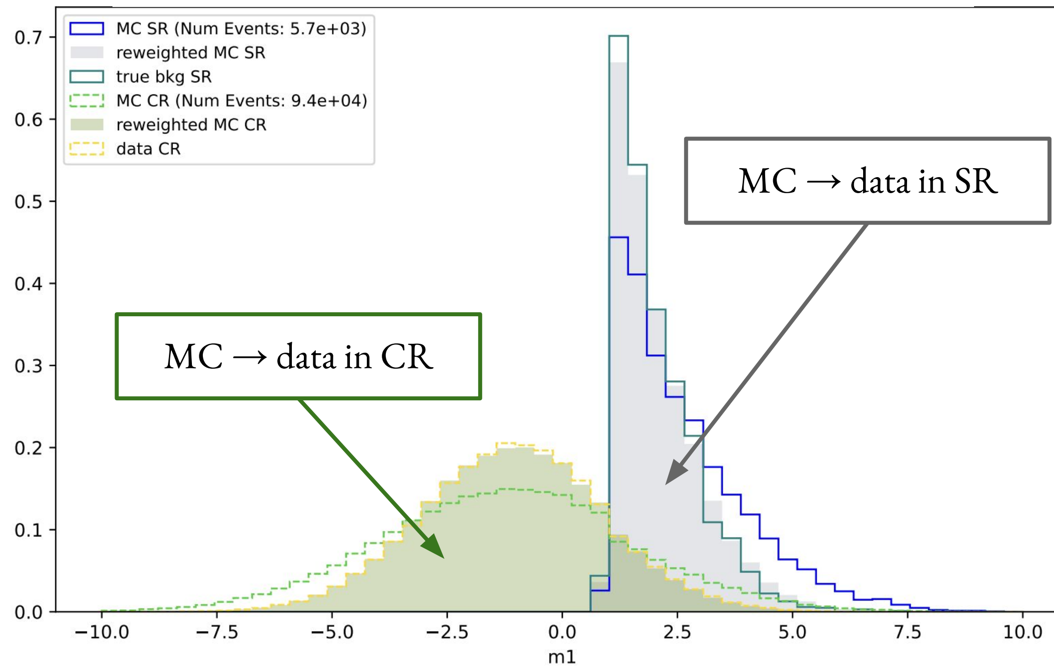
- **Reweight**

Assumption:

$$w(m) = \frac{P_{data}(m \in CR)}{P_{MC}(m \in CR)}$$

$$= \frac{P_{data}(m \in SR)}{P_{MC}(m \in SR)}$$

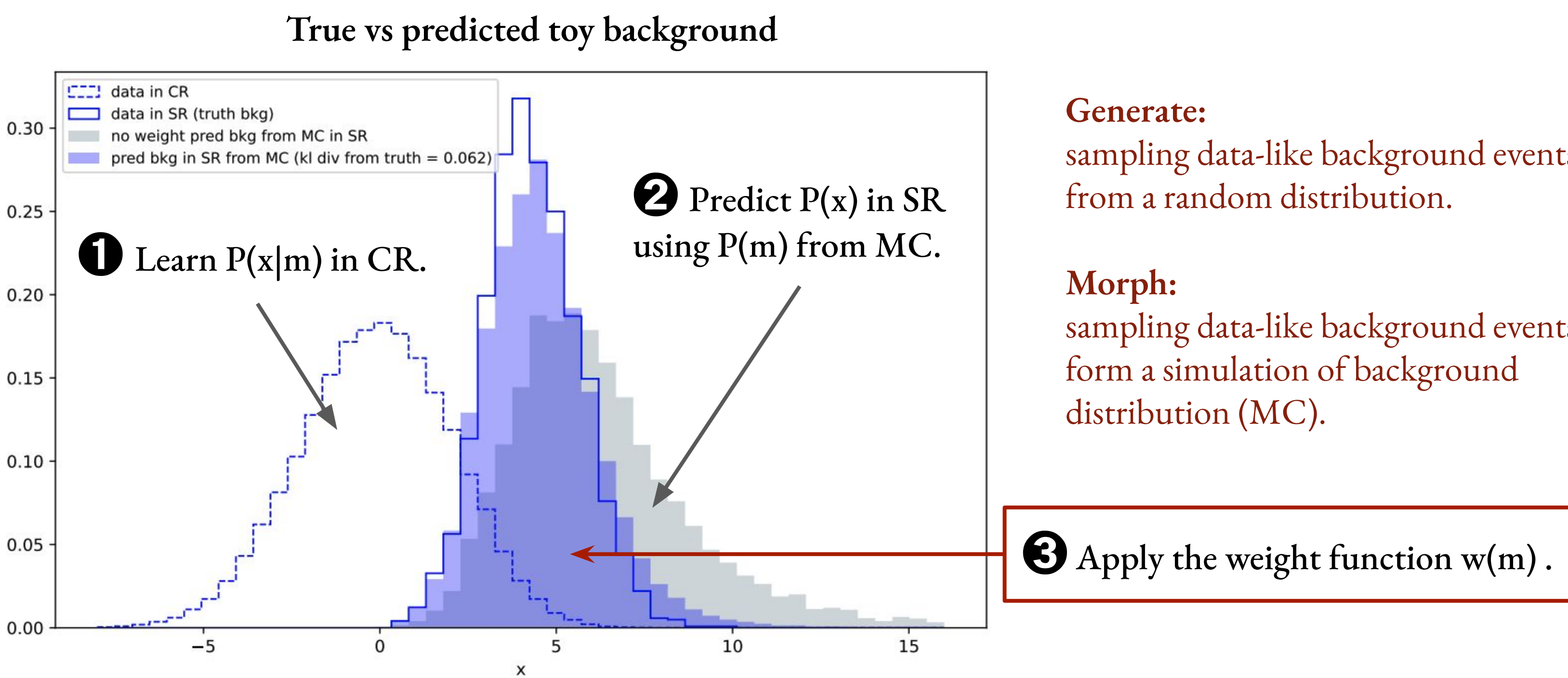
Reweighted MC vs data for the toy background



* This plot is used for illustration purpose only, not the final result.

Anomaly detection methods

- **Generate/Morph**



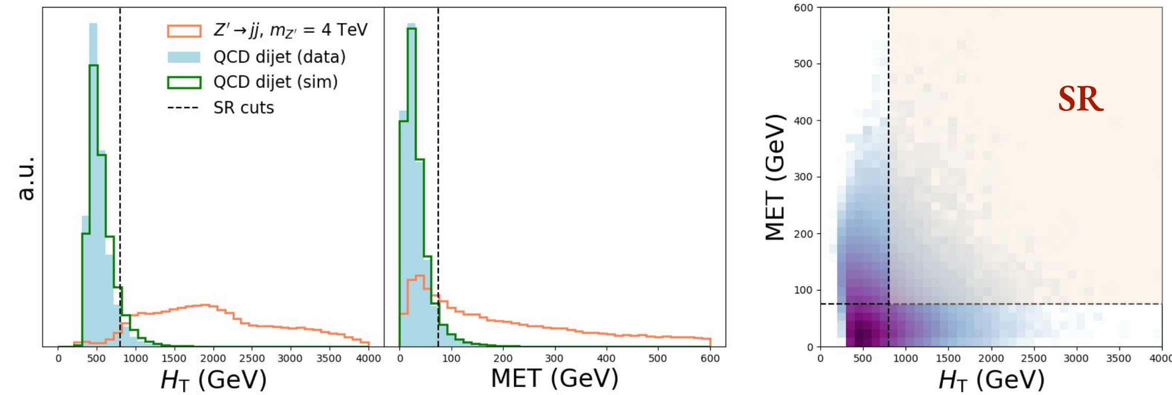
Generate:
sampling data-like background events from a random distribution.

Morph:
sampling data-like background events form a simulation of background distribution (MC).

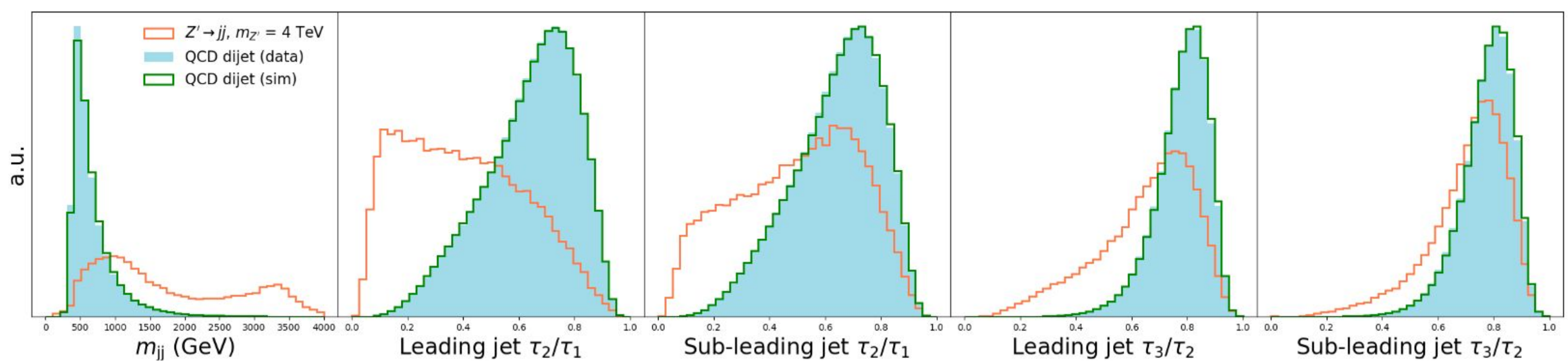
* This plot is used for illustration purpose only, not the final result.

Physics example

Context variables: HT &MET



Feature variables: m_{jj} & N-subjettiness

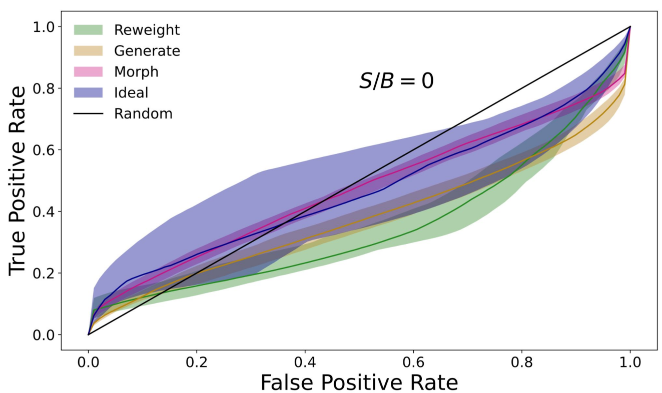


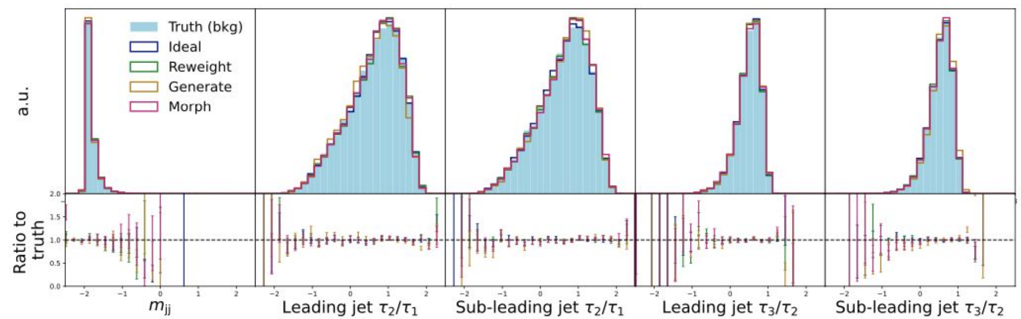
Background estimation results (no signal)

We see closures in CR, and agreements in SR for the 5 feature variables.

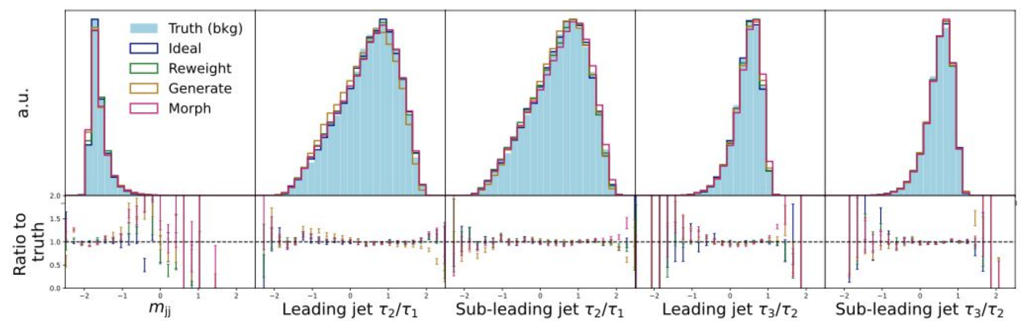
- ➔ Prediction matched truth.
- ➔ Minimum mis-modeling and false-positives.

Receiver Operating Characteristic (ROC)





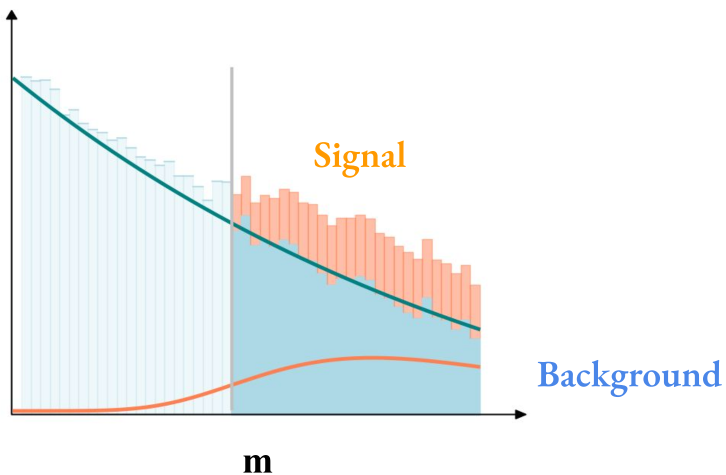
(a) Control region distributions.

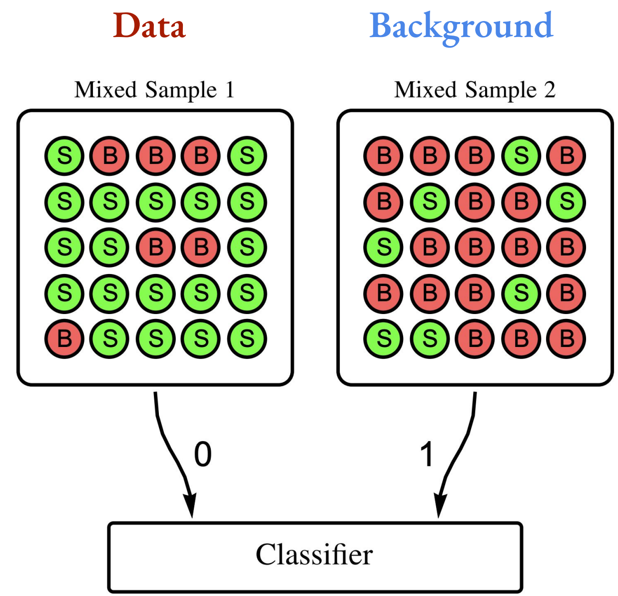


(b) Signal region distributions.

AD performance with injected signal events

❸ Classification in SR

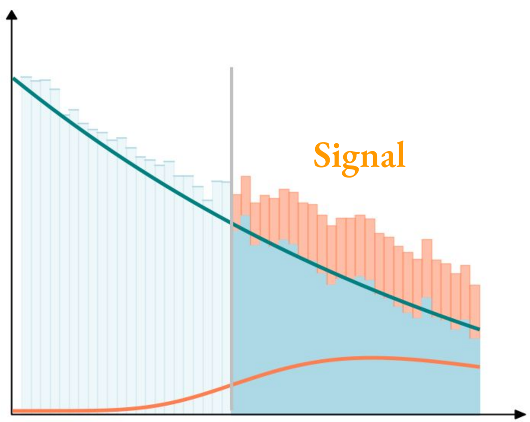




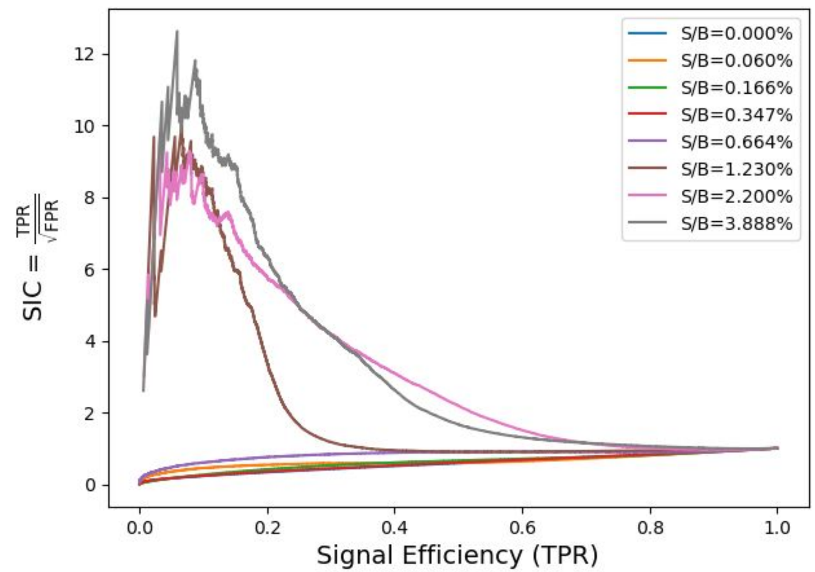
Classification without labels (CWoLa) [1708.02949]

AD performance with injected signal events

❸ Classification in SR



Significance Improvement Characteristic (SIC)

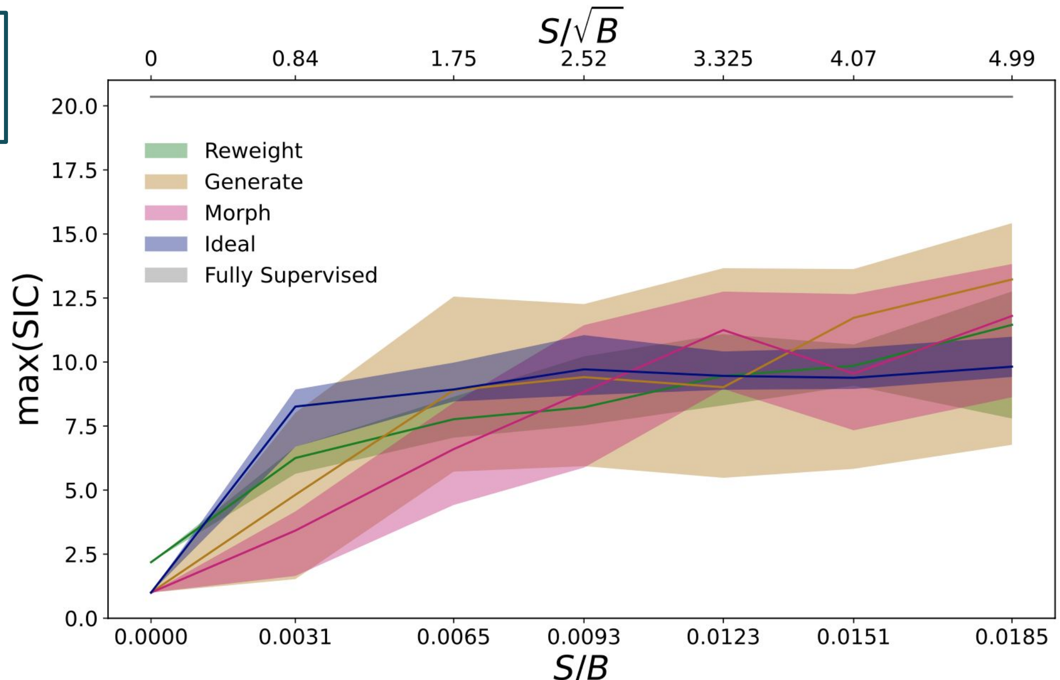


* This plot is used for illustration purpose only, not the final result.

AD performance over increasing signal injections

Signal significance

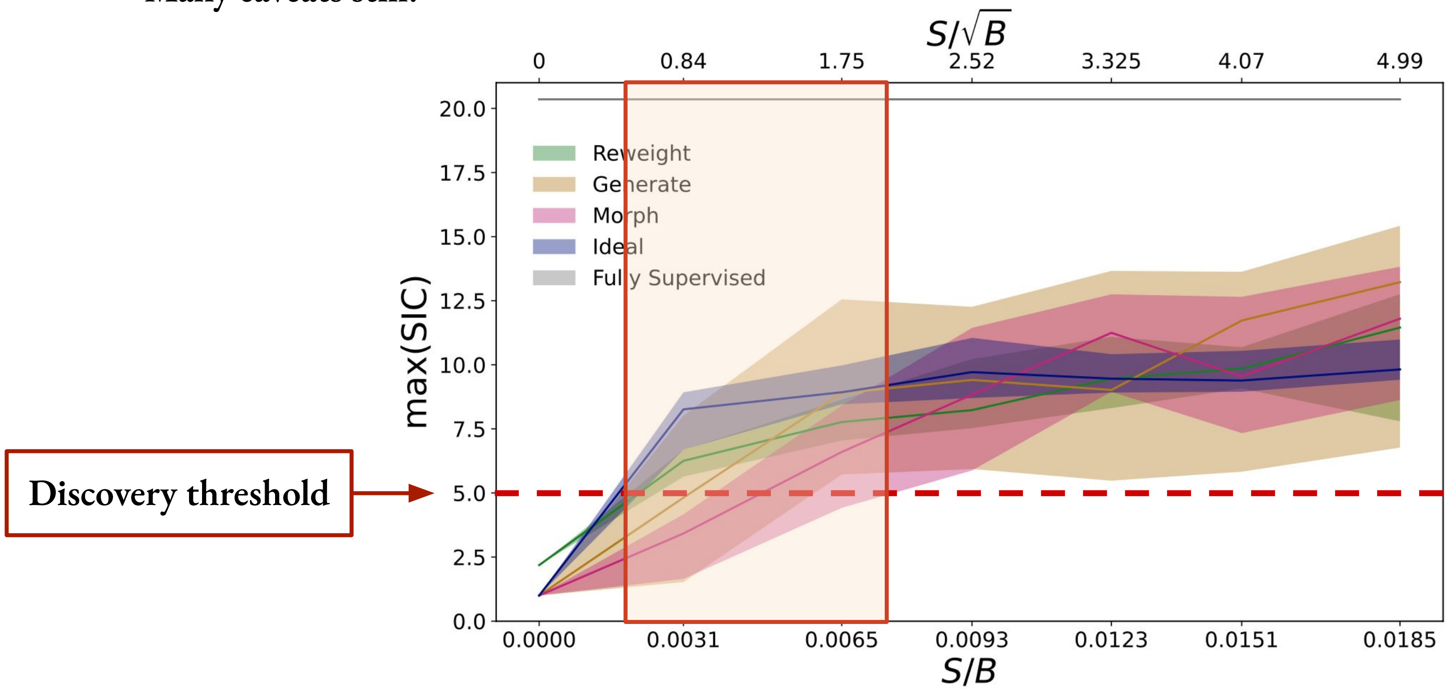
Maximum significance improvement



signal-to-background ratio

AD performance over increasing signal injections

- ➔ A 1-2 sigma signal significance is enhanced to reach the discover threshold of 5 sigma.
- ➔ Many caveats still!



How robust is this performance?

One needs to test these AD methods against many different signal models.

- Tested on lower Z' mass (2 and 3 TeV).
- Found that the performance is worse for lower mass Z' signals.

Did we make the problem easier by choosing a background simulation that looks similar to the "real" background?

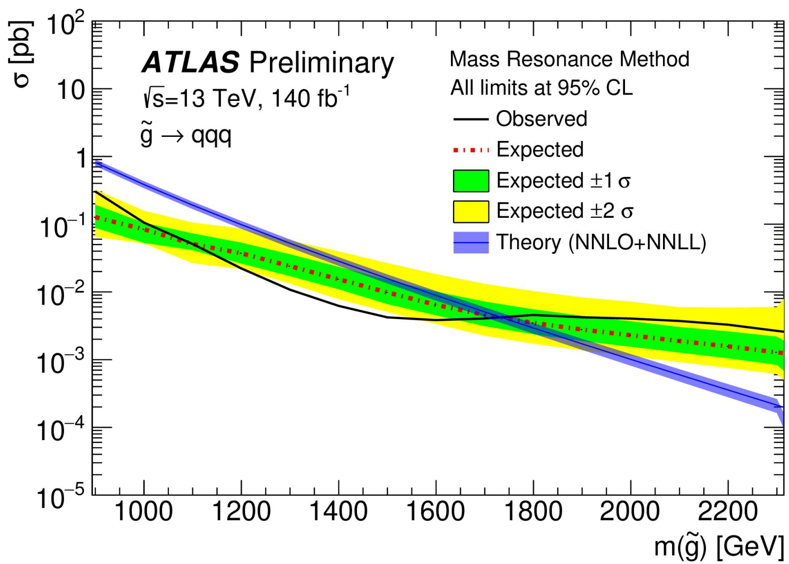
- Repeated the analysis with a hand-tuned simulation daset.
- Found that the performance are compatible with the current simulation.

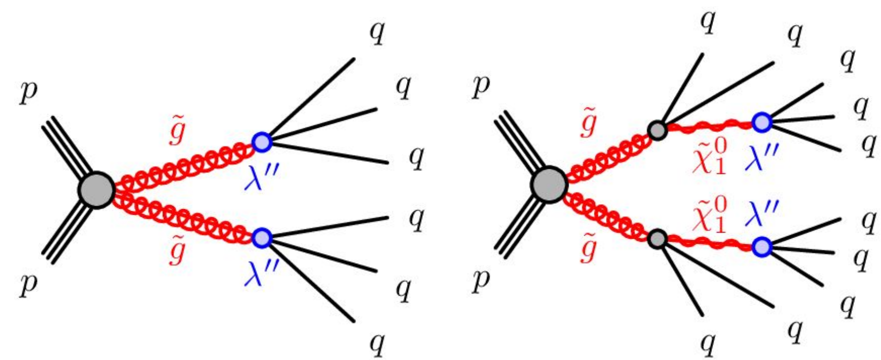
* This discussion will be included in the V2 paper. Stay tuned!

A few words on extensions to RPV SUSY

A full Run2 ATLAS search for R-parity-violating Supersymmetry (RPV SUSY) in hadronic final states is public!

➔ **My first analysis as a graduate student :)**

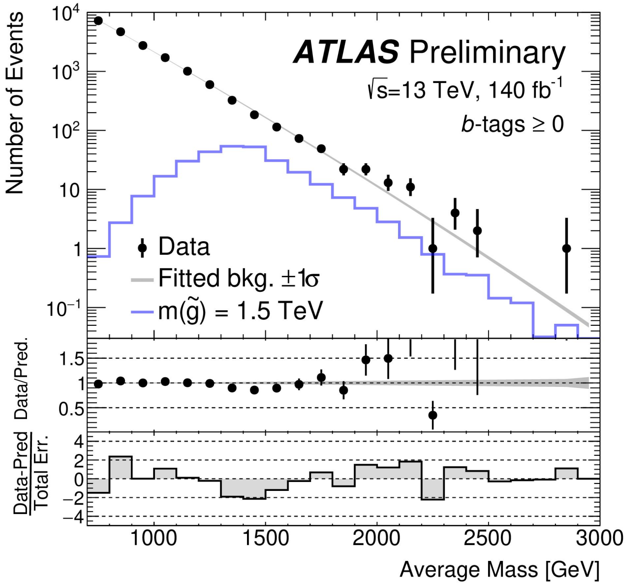


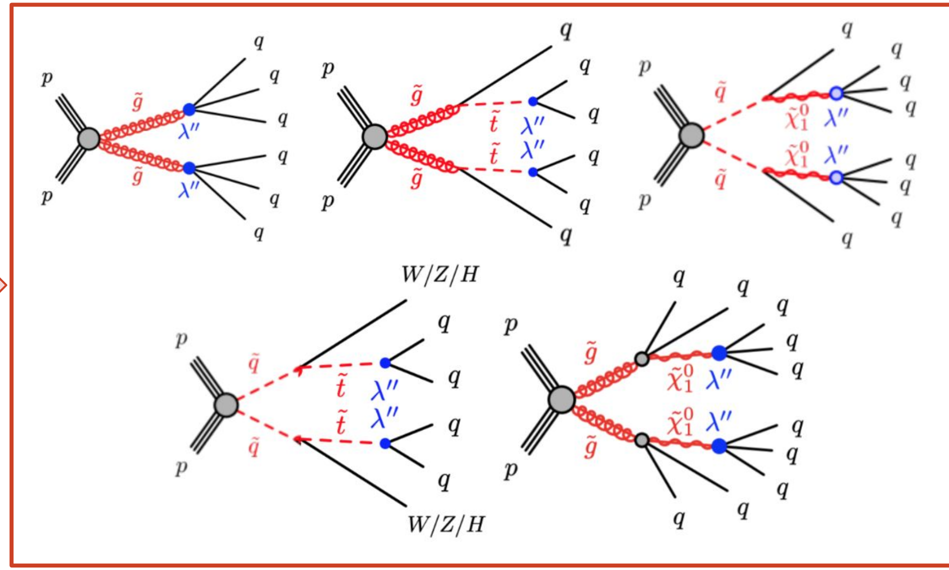


A few words on extensions to RPV SUSY

The analysis reconstructs the gluino mass with a jet-matching NN, additional to a cut & count approach.

The signal is however not so resonant! What about a non-resonant AD for a multi-jet search? ;)





Conclusions

- **This is the first proposal of non-resonant anomaly detection using extrapolation methods!**
- **There are many interesting problems associated with extrapolation that are worth exploring in the future.**
- **We encourage future work to make it possible to look for non-resonant BSM signals!**

Thank you!

Get in touch: kbai@uoregon.edu

Acknowledgement

We thank T. Cohen, S. Diefenbacher, L. Jeanty, S. Knapen, and C. Scherb for the useful discussions!

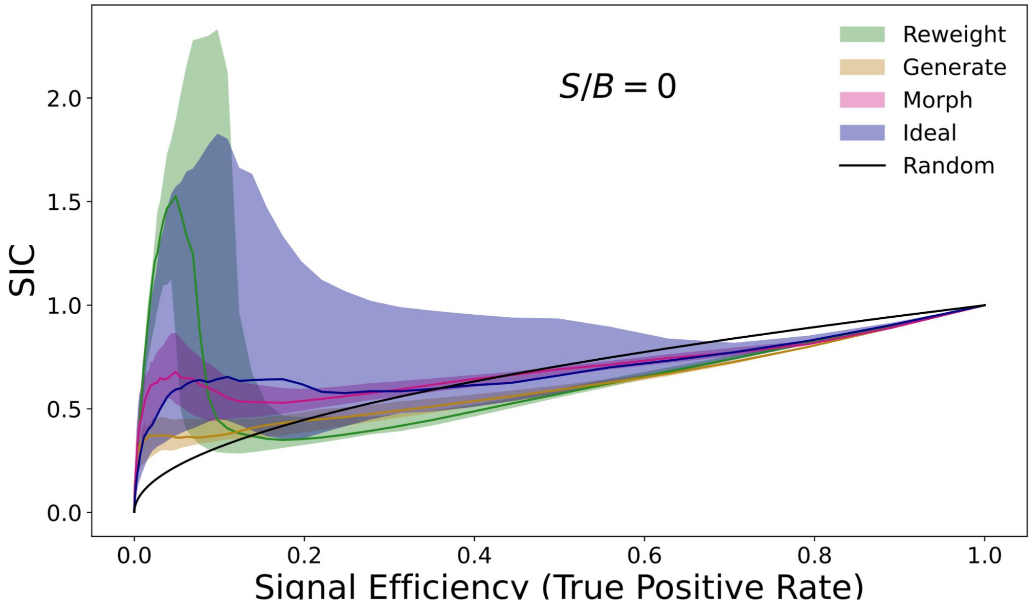
KB is supported by the U.S. Department of Energy, Office of Science, Office of High Energy Physics program under Award Number DE-SC0020244. KB's visit to LBNL for this collaboration is supported by the US ATLAS Center program. BN and RM are supported by the U.S. Department of Energy (DOE), Office of Science under contract DE-AC02-05CH11231 and Grant No. 63038 from the John Templeton Foundation. RM is additionally supported by Grant No. DGE 2146752 from the National Science Foundation Graduate Research Fellowship Program. This research used resources of the National Energy Research Scientific Computing Center, a DOE Office of Science User Facility supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231 using NERSC award HEP-ERCAP0021099.

Back up

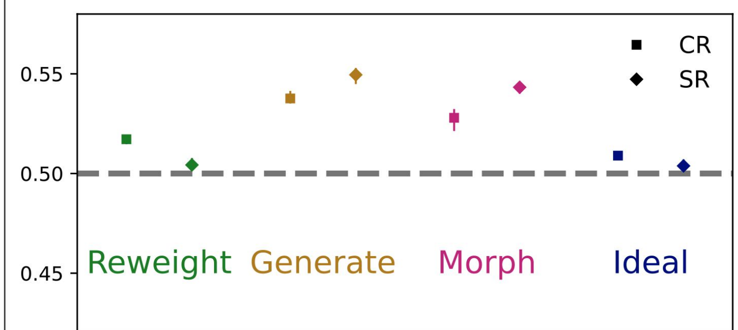
Zero-signal injection

SIC and ROC AUCs at zero-signal injection, showing a non-biased background estimation.

➔ Low false-positives.



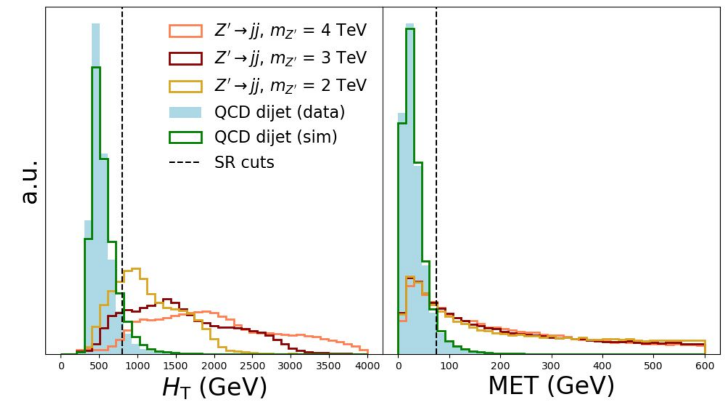
ROC AUCs (area-under-curves) compared with true random

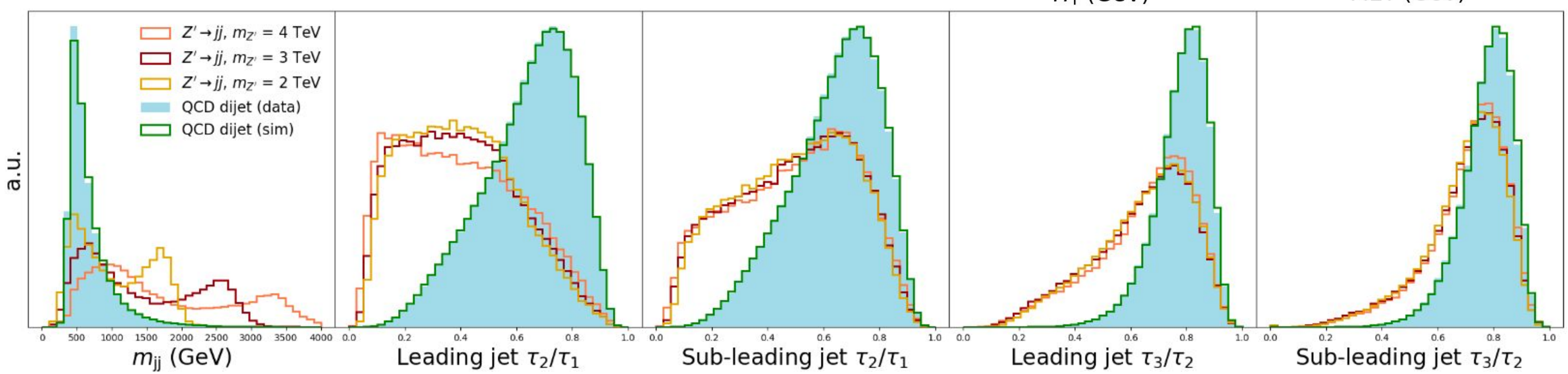


Other signal parameters

Lower Z' masses:

- $mZ' = 2$ TeV, $m_{\pi D} = m_{\rho D} = \Lambda_D = 200$ GeV, and $m_{qD} = 100$ GeV
- $mZ' = 3$ TeV, $m_{\pi D} = m_{\rho D} = \Lambda_D = 300$ GeV, and $m_{qD} = 150$ GeV

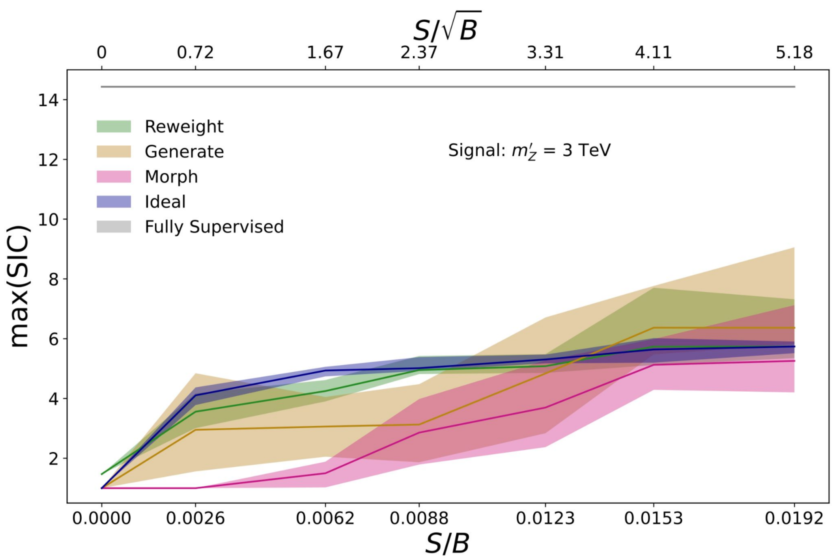


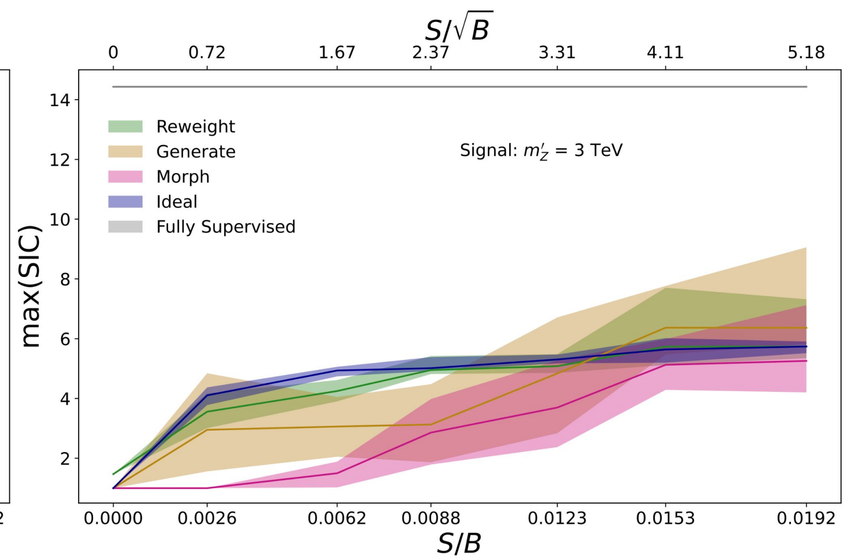


Other signal parameters

Lower Z' masses:

- $mZ' = 2$ TeV, $m_{\pi D} = m_{\rho D} = \Lambda_D = 200$ GeV, and $m_{qD} = 100$ GeV.
- $mZ' = 3$ TeV, $m_{\pi D} = m_{\rho D} = \Lambda_D = 300$ GeV, and $m_{qD} = 150$ GeV.

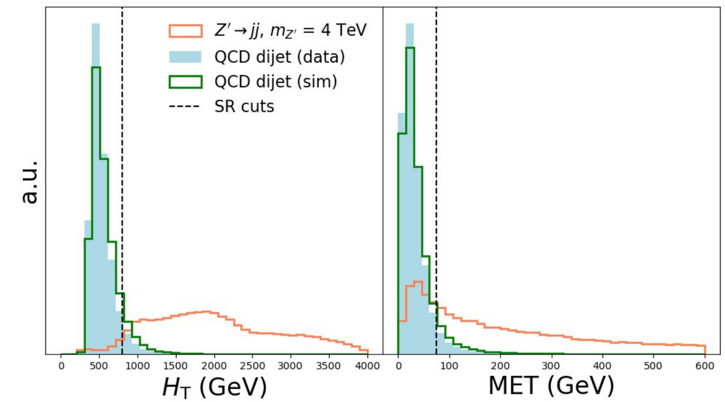


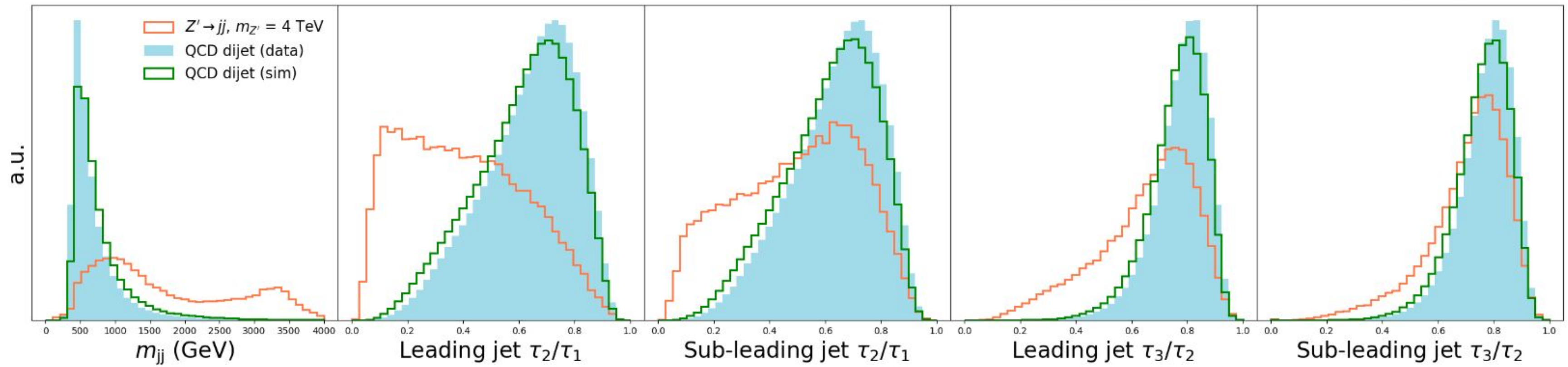


A hand-tuned background simulation

Change the simulation shape to look more different from the data.

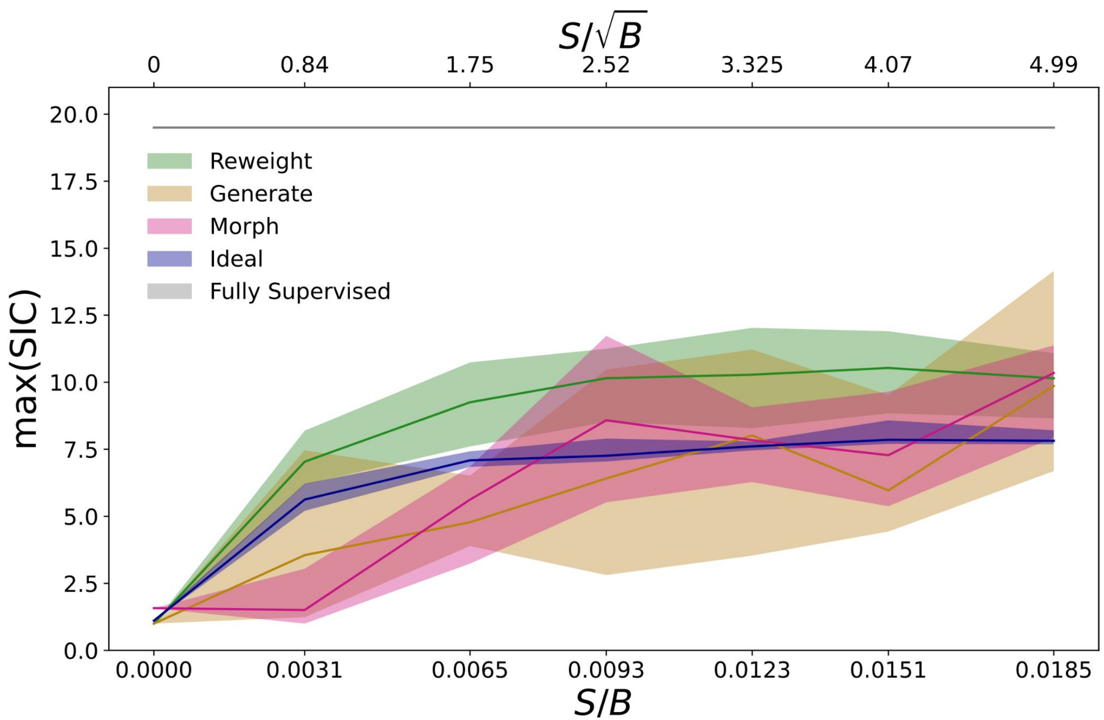
$$H_T: \quad x \to x$$
$$\text{MET}: \quad x \to x(1+\frac{x}{500})$$
$$m_{jj}: \quad x \to x(1+\frac{x}{6000})$$
$$\tau_i: \quad x \to x^{1.1}.$$





A hand-tuned background simulation

Change the simulation shape to look more different from the data.



SIC evaluated at a rejection of 10^3