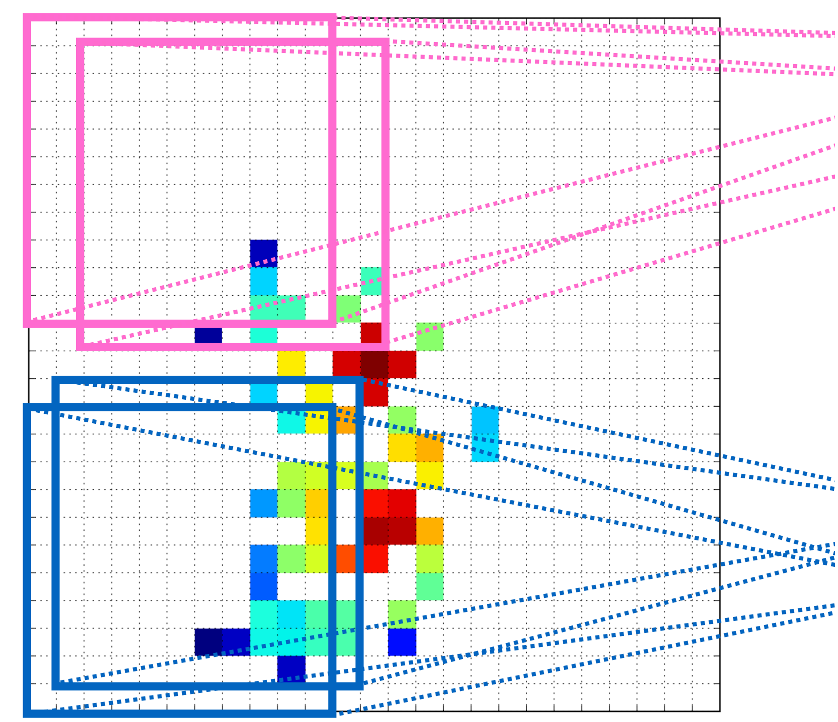
High-dimensional Anomaly Detection with Radiative Return in e^+e^- Collisions

Benjamin Nachman

Lawrence Berkeley National Laboratory

bpnachman.com
bpnachman@lbl.gov

@bpnachman bnachman



BERKELEY LAB

BIDS
BERKELEY INSTITUTE FOR DATA SCIENCE

FCC physics meeting

September 26, 2021

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

Julia Gonski,[a] **Jerry Lai,**[b] **Benjamin Nachman,**[c,d] **and Inês Ochoa**[e]

[a] *Nevis Laboratories, Columbia University, 136 S Broadway, Irvington NY, USA*

[c] *Department of Electrical Engineering and Computer Sciences, University of California, Berkeley, CA 94720, USA*

[c] *Physics Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA*

[d] *Berkeley Institute for Data Science, University of California, Berkeley, CA 94720, USA*

[e] *Laboratory of Instrumentation and Experimental Particle Physics, Lisbon, Portugal*

E-mail: julia.gonski@cern.ch, thejerrylai@berkeley.edu, bpnachman@lbl.gov, ines.ochoa@cern.ch

Abstract: Experiments at a future e^+e^- collider will be able to search for new particles with masses below the nominal centre-of-mass energy by analyzing collisions with initial-state radiation (radiative return). We show that machine learning methods based on semisupervised and weakly supervised learning can achieve model-independent sensitivity to the production of new particles in radiative return events. In addition to a first application of these methods in e^+e^- collisions, our study is the first to combine weak supervision with variable-dimensional information by deploying a deep sets neural network architecture. We have also investigated some of the experimental aspects of anomaly detection in radiative return events and discuss these in the context of future detector design.

(Brief) Motivation

Theoretical and **experimental** questions motivate a deep exploration **of the fundamental structure of nature**

Dark matter

Hierarchy problem

Strong CP

Flavor puzzles

Dark energy

Baryogengesis

We have performed thousands of hypothesis tests & have no significant evidence for physics beyond the Standard Model

Three possibilities

(Brief) Motivation

Theoretical and **experimental** questions motivate a deep exploration **of the fundamental structure of nature**

Dark matter

Hierarchy problem

Strong CP

Dark energy

Flavor puzzles

Baryogengesis

We have performed thousands of hypothesis tests & have no significant evidence for physics beyond the Standard Model

Three possibilities

(1) There is nothing new at LHC energies

(Brief) Motivation

Theoretical and **experimental** questions motivate a deep exploration **of the fundamental structure of nature**

Dark matter

Hierarchy problem

Strong CP

Dark energy

Flavor puzzles

Baryogengesis

We have performed thousands of hypothesis tests & have no significant evidence for physics beyond the Standard Model

Three possibilities

(1) There is nothing new at ~~LHC~~ FCC energies

(Brief) Motivation

Theoretical and **experimental** questions motivate a deep exploration **of the fundamental structure of nature**

Dark matter
Strong CP
Hierarchy problem
Dark energy
Flavor puzzles
Baryogengesis

We have performed thousands of hypothesis tests & have no significant evidence for physics beyond the Standard Model

Three possibilities

(1) There is nothing new at FCC energies

(2) Patience! (new physics is rare)

(Brief) Motivation

Theoretical and **experimental** questions motivate a deep exploration **of the fundamental structure of nature**

Dark matter

Hierarchy problem

Strong CP

Flavor puzzles

Dark energy

Baryogengesis

We have performed thousands of hypothesis tests & have no significant evidence for physics beyond the Standard Model

Three possibilities

(1) There is nothing new at FCC energies

(2) Patience! (new physics is rare)

(3) We are not looking in the right place

(Brief) Motivation

Theoretical and **experimental** questions motivate a deep exploration **of the fundamental structure of nature**

Dark matter

Hierarchy problem

Strong CP

Flavor puzzles

Dark energy

Baryogengesis

We have performed thousands of hypothesis tests & have no significant evidence for physics beyond the Standard Model

Three possibilities

This is what motivated this work!

(3) We are not looking in the right place

(Brief) Motivation

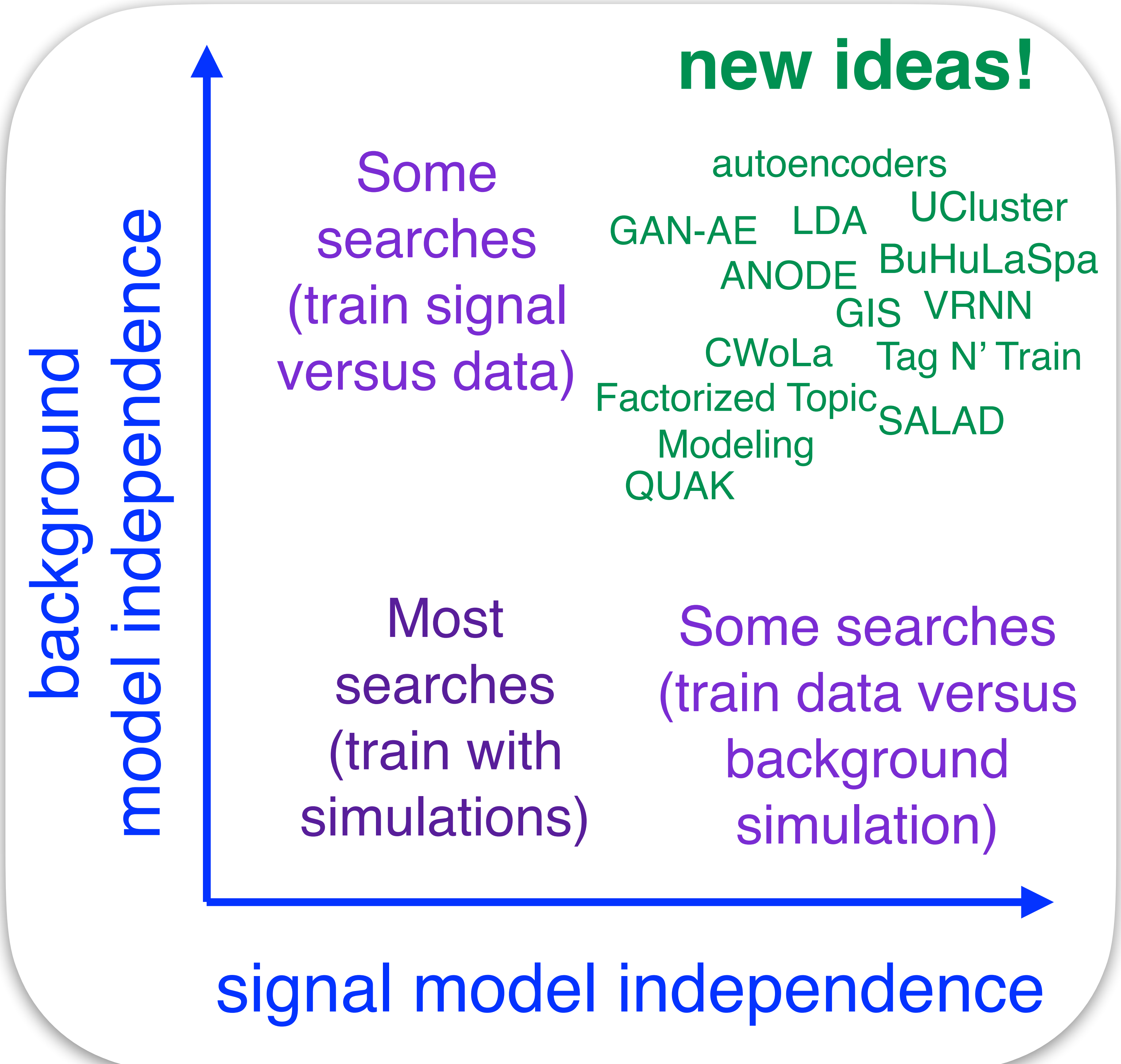
		e	μ	τ	q/g	b	t	γ	Z/W	H	BSM $\to$ SM$_1$ × SM$_1$: q/g	γ/π^0's	b	$\cdots$	BSM $\to$ SM$_1$ × SM$_2$: tZ/H	bH	$\cdots$	BSM $\to$ complex: $\tau qq'$	eqq'	$\mu qq'$	$\cdots$
	e	[37, 38]	[39, 40]	[39]	∅	∅	∅	[41]	[42]	∅	∅	∅	∅		∅	∅	∅	∅	[43, 44]	∅	
	μ		[37, 38]	[39]	∅	∅	∅	[41]	[42]	∅	∅	∅	∅		∅	∅	∅	∅	∅	[43, 44]	
	τ			[45, 46]	∅	[47]	∅	∅	∅	∅	∅	∅	∅		∅	∅	∅	[48, 49]	∅	∅	
	q/g				[29, 30, 50, 51]	[52]	∅	[53, 54]	[55]	∅	∅	∅	∅		∅	∅	∅	∅	∅	∅	
	b					[29, 52, 56]	[57]	[54]	[58]	[59]	∅	∅	∅		[60]	∅	∅	∅	∅	∅	
	t						[61]	∅	[62]	[63]	∅	∅	∅		[64]	[60]	∅	∅	∅	∅	
	γ							[65, 66]	[67–69]	[68, 70]	∅	∅	∅		∅	∅	∅	∅	∅	∅	
	Z/W								[71]	[71]	∅	∅	∅		∅	∅	∅	∅	∅	∅	
	H									[72, 73]	[74]	∅	∅		∅	∅	∅	∅	∅	∅	
BSM $\to$ SM$_1$ × SM$_1$	q/g										∅	∅	∅		∅	∅	∅	∅	∅	∅	
	γ/π^0's											[75]	∅		∅	∅	∅	∅	∅	∅	
	b												[76, 77]		∅	∅	∅	∅	∅	∅	
	⋮																				
	⋮																				
BSM $\to$ SM$_1$ × SM$_2$	tZ/H																				
	bH																				
	⋮																				
	⋮																				
BSM $\to$ complex	$\tau qq'$																				
	eqq'																				
	$\mu qq'$																				
	⋮																				
	⋮																				

There are also many uncovered scenarios - we cannot possibly do a search for every possible topology!

J. Kim, K. Kong, BN, and D. Whiteson, 1907.06659

(3) We are not looking in the right place

New Methods



Adapted from BN and D. Shih, 2001.04990

There are many new ideas that make use of modern machine learning

The goal is to learn **directly from data**, injecting as little bias as possible

*N.B. this is just for signal sensitivity - there is **also model dependence** for determining the background* https://lhco2020.github.io/homepage/ 2101.08320

A method testing ground: the LHCO

The LHC Olympics 2020

A Community Challenge for Anomaly Detection in High Energy Physics

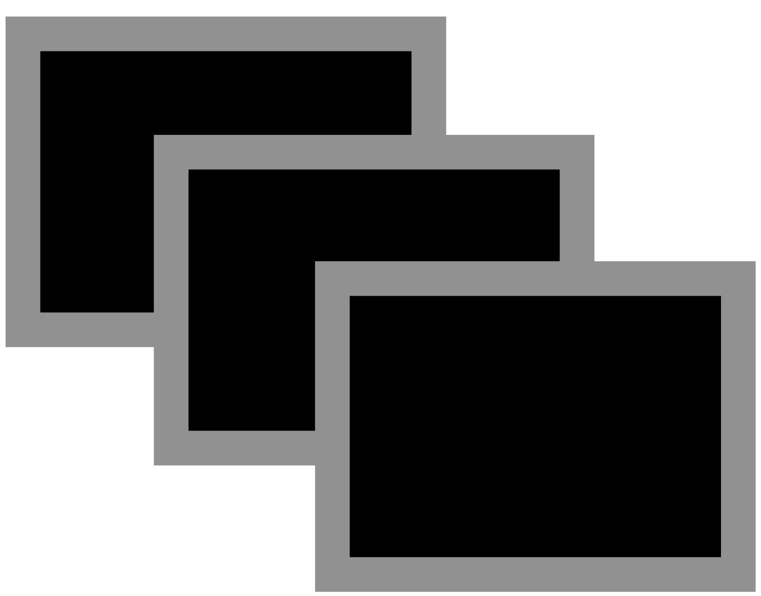
Gregor Kasieczka (ed),[1] Benjamin Nachman (ed),[2,3] David Shih (ed),[4] Oz Amram,[5] Anders Andreassen,[6] Kees Benkendorfer,[2,7] Blaz Bortolato,[8] Gustaaf Brooijmans,[9] Florencia Canelli,[10] Jack H. Collins,[11] Biwei Dai,[12] Felipe F. De Freitas,[13] Barry M. Dillon,[8,14] Ioan-Mihail Dinu,[5] Zhongtian Dong,[15] Julien Donini,[16] Javier Duarte,[17] D. A. Faroughy[10] Julia Gonski,[9] Philip Harris,[18] Alan Kahn,[9] Jernej F. Kamenik,[8,19] Charanjit K. Khosa,[20,30] Patrick Komiske,[21] Luc Le Pottier,[2,22] Pablo Martín-Ramiro,[2,23] Andrej Matevc,[8,19] Eric Metodiev,[21] Vinicius Mikuni,[10] Inês Ochoa,[24] Sang Eon Park,[18] Maurizio Pierini,[25] Dylan Rankin,[18] Veronica Sanz,[20,26] Nilai Sarda,[27] Uroš Seljak,[2,3,12] Aleks Smolkovic,[8] George Stein,[2,12] Cristina Mantilla Suarez,[5] Manuel Szewc,[28] Jesse Thaler,[21] Steven Tsan,[17] Silviu-Marian Udrescu,[18] Louis Vaslin,[16] Jean-Roch Vlimant,[29] Daniel Williams,[9] Mikaeel Yunus[18]

The Challenge

We provided a list of particle for each event (700 particles with the 3-vector of each particle)

1 dataset for R&D with labeled signal and background

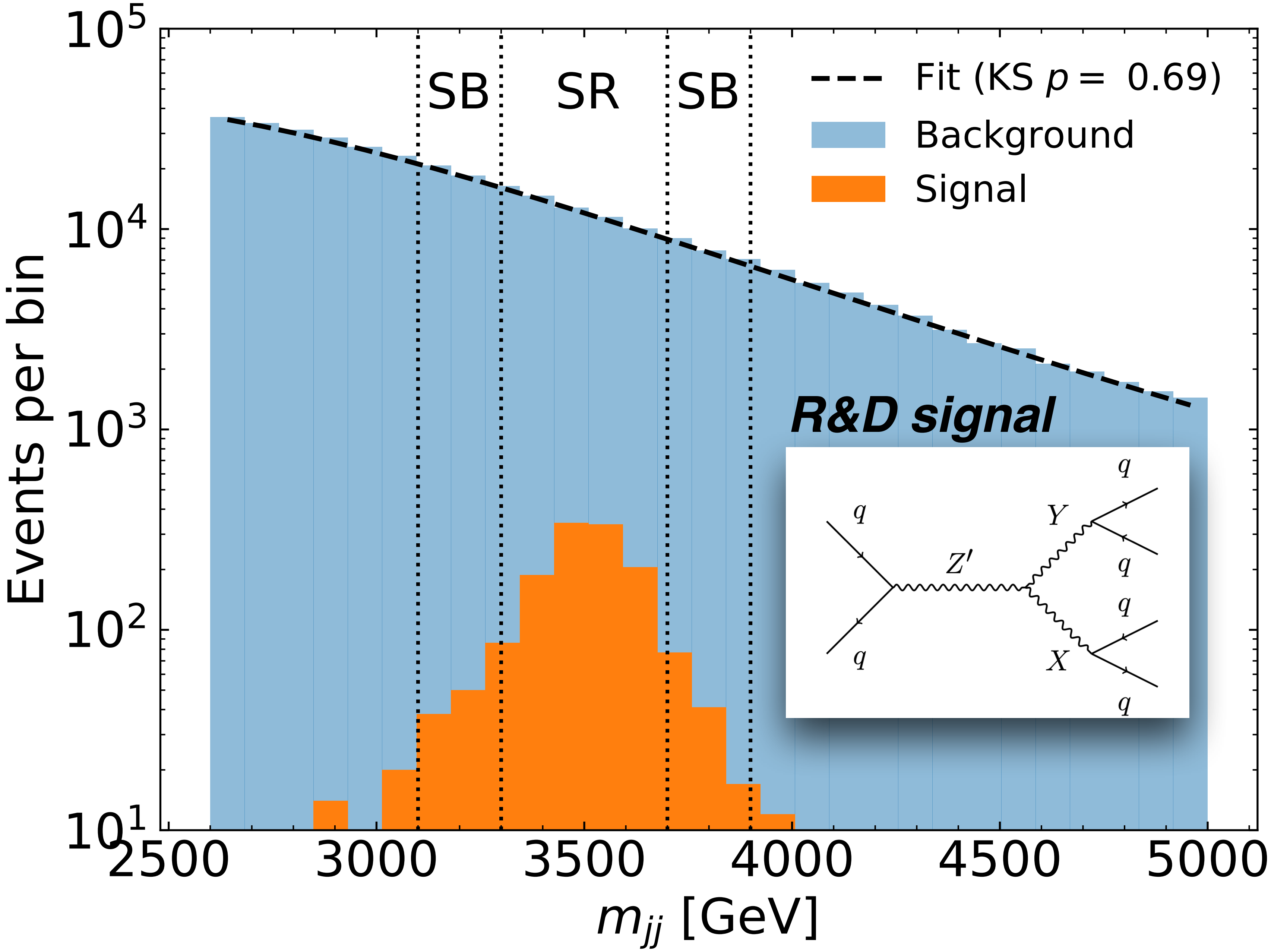
3 black boxes with unlabeled data

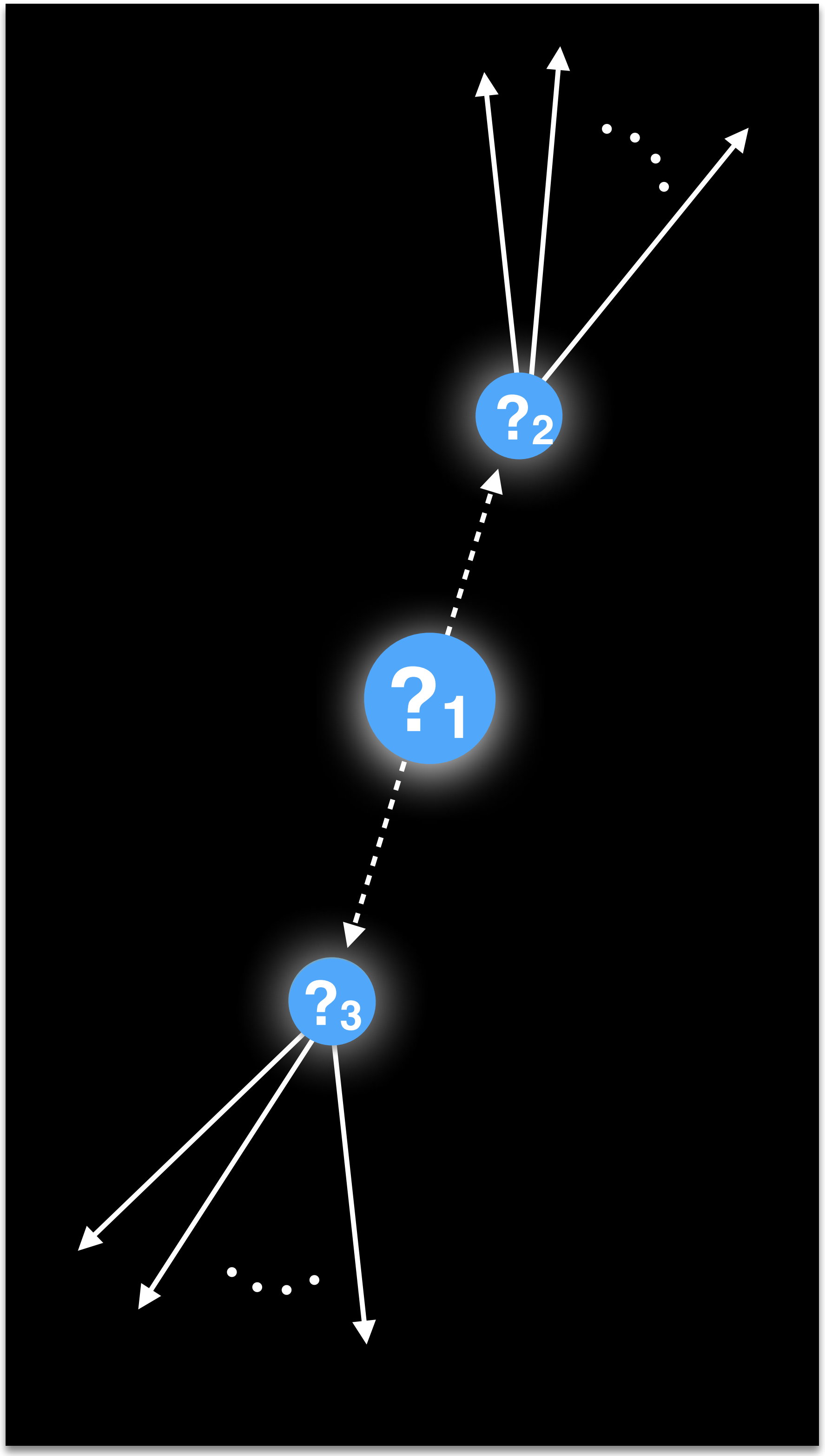


The particle-level + detector-level simulation for background in the black boxes was modified for each dataset (think Pythia/Herwig, etc.)

Actually, all of the parameters are now public on Zenodo

The dataset





Dijet final state (allow for data-driven background + complex final state).

Solutions

I don't have time to cover all of them - please see the paper for details! I'll just highlight some general ideas.

Section	Short Name	Method Type	Results Type
3.1	VRNN	Unsupervised	(i) (BB2,3) and (ii) (BB1)
3.2	ANODE	Unsupervised	(iii)
3.3	BuHuLaSpa	Unsupervised	(i) (BB2,3) and (ii) (BB1)
3.4	GAN-AE	Unsupervised	(i) (BB2-3) and (ii) (BB1)
3.5	GIS	Unsupervised	(i) (BB1)
3.6	LDA	Unsupervised	(i) (BB1-3)
3.7	PGA	Unsupervised	(ii) (BB1-2)
3.8	Reg. Likelihoods	Unsupervised	(iii)
3.9	UCluster	Unsupervised	(i) (BB2-3)
4.1	CWoLa	Weakly Supervised	(ii) (BB1-2)
4.2	CWoLa AE Compare	Weakly/Unsupervised	(iii)
4.3	Tag N' Train	Weakly Supervised	(i) (BB1-3)
4.4	SALAD	Weakly Supervised	(iii)
4.5	SA-CWoLa	Weakly Supervised	(iii)
5.1	Deep Ensemble	Semisupervised	(i) (BB1)
5.2	Factorized Topics	Semisupervised	(iii)
5.3	QUAK	Semisupervised	(i) (BB2,3) and (ii) (BB1)
5.4	LSTM	Semisupervised	(i) (BB1-3)

BB = black box; (i) = blinded, (ii) = unblinded

Solutions

Supervision refers to the type of label information provided to the ML during training.

Unsupervised = no labels
Weakly-supervised = noisy labels
Semi-supervised = partial labels
Supervised = full label information

Section	Short Name	Method Type	Results Type
3.1	VRNN	Unsupervised	(i) (BB2,3) and (ii) (BB1)
3.2	ANODE	Unsupervised	(iii)
3.3	BuHuLaSpa	Unsupervised	(i) (BB2,3) and (ii) (BB1)
3.4	GAN-AE	Unsupervised	(i) (BB2-3) and (ii) (BB1)
3.5	GIS	Unsupervised	(i) (BB1)
3.6	LDA	Unsupervised	(i) (BB1-3)
3.7	PGA	Unsupervised	(ii) (BB1-2)
3.8	Reg. Likelihoods	Unsupervised	(iii)
3.9	UCluster	Unsupervised	(i) (BB2-3)
4.1	CWoLa	Weakly Supervised	(ii) (BB1-2)
4.2	CWoLa AE Compare	Weakly/Unsupervised	(iii)
4.3	Tag N' Train	Weakly Supervised	(i) (BB1-3)
4.4	SALAD	Weakly Supervised	(iii)
4.5	SA-CWoLa	Weakly Supervised	(iii)
5.1	Deep Ensemble	Semisupervised	(i) (BB1)
5.2	Factorized Topics	Semisupervised	(iii)
5.3	QUAK	Semisupervised	(i) (BB2,3) and (ii) (BB1)
5.4	LSTM	Semisupervised	(i) (BB1-3)

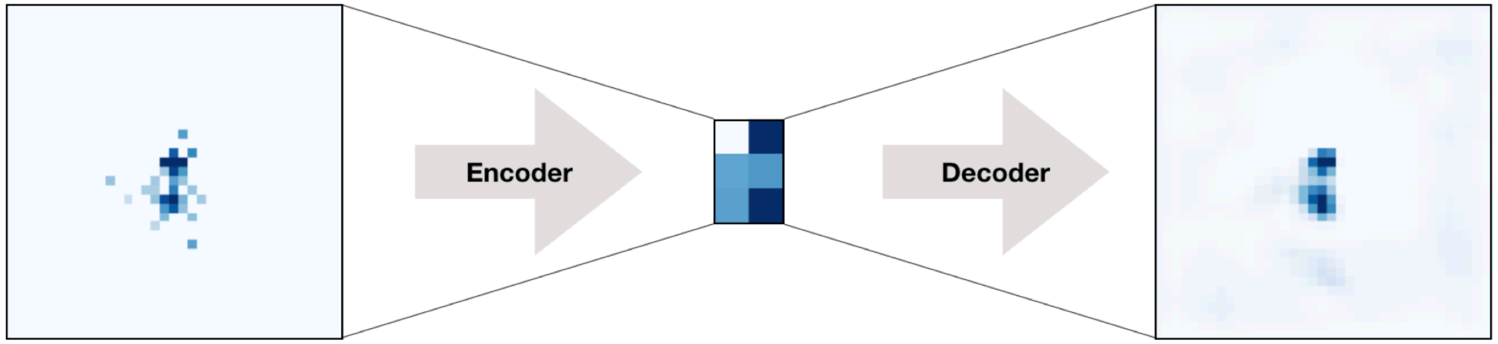
These categories are not exact and the boundaries are not rigid!

*N.B. Not everyone agrees on the boundary between semi-supervised and weakly supervised.

Solutions: Unsupervised

Unsupervised = no labels

Typically, the goal of these methods is to look for events with low *p(background)*



Section	Short Name	Method Type	Results Type
3.1	VRNN	Unsupervised	(i) (BB2,3) and (ii) (BB1)
3.2	ANODE	Unsupervised	(iii)
3.3	BuHuLaSpa	Unsupervised	(i) (BB2,3) and (ii) (BB1)
3.4	GAN-AE	Unsupervised	(i) (BB2-3) and (ii) (BB1)
3.5	GIS	Unsupervised	(i) (BB1)
3.6	LDA	Unsupervised	(i) (BB1-3)
3.7	PGA	Unsupervised	(ii) (BB1-2)
3.8	Reg. Likelihoods	Unsupervised	(iii)
3.9	UCluster	Unsupervised	(i) (BB2-3)
4.1	CWoLa	Weakly Supervised	(ii) (BB1-2)
4.2	CWoLa AE Compare	Weakly/Unsupervised	(iii)
4.3	Tag N' Train	Weakly Supervised	(i) (BB1-3)
4.4	SALAD	Weakly Supervised	(iii)
4.5	SA-CWoLa	Weakly Supervised	(iii)
5.1	Deep Ensemble	Semisupervised	(i) (BB1)
5.2	Factorized Topics	Semisupervised	(iii)
5.3	QUAK	Semisupervised	(i) (BB2,3) and (ii) (BB1)
5.4	LSTM	Semisupervised	(i) (BB1-3)

One strategy (autoencoders) is to try to compress events and then uncompress them. When x = uncompres(compress(x)), then x probably has low p(x).

M. Farina, Y. Nakai, D. Shih, 1808.08992; T. Heimel, G. Kasieczka, T. Plehn, J. Thompson, 1808.08979; + many more

Solutions: Weakly-supervised

Weakly-supervised = noisy labels

Typically, the goal of these methods is to look for events with high *p(possibly signal-enriched)/p(possibly signal-depleted)*

Section	Short Name	Method Type	Results Type
3.1	VRNN	Unsupervised	(i) (BB2,3) and (ii) (BB1)
3.2	ANODE	Unsupervised	(iii)
3.3	BuHuLaSpa	Unsupervised	(i) (BB2,3) and (ii) (BB1)
3.4	GAN-AE	Unsupervised	(i) (BB2-3) and (ii) (BB1)
3.5	GIS	Unsupervised	(i) (BB1)
3.6	LDA	Unsupervised	(i) (BB1-3)
3.7	PGA	Unsupervised	(ii) (BB1-2)
3.8	Reg. Likelihoods	Unsupervised	(iii)
3.9	UCluster	Unsupervised	(i) (BB2-3)
4.1	CWoLa	Weakly Supervised	(ii) (BB1-2)
4.2	CWoLa AE Compare	Weakly/Unsupervised	(iii)
4.3	Tag N' Train	Weakly Supervised	(i) (BB1-3)
4.4	SALAD	Weakly Supervised	(iii)
4.5	SA-CWoLa	Weakly Supervised	(iii)
5.1	Deep Ensemble	Semisupervised	(i) (BB1)
5.2	Factorized Topics	Semisupervised	(iii)
5.3	QUAK	Semisupervised	(i) (BB2,3) and (ii) (BB1)
5.4	LSTM	Semisupervised	(i) (BB1-3)

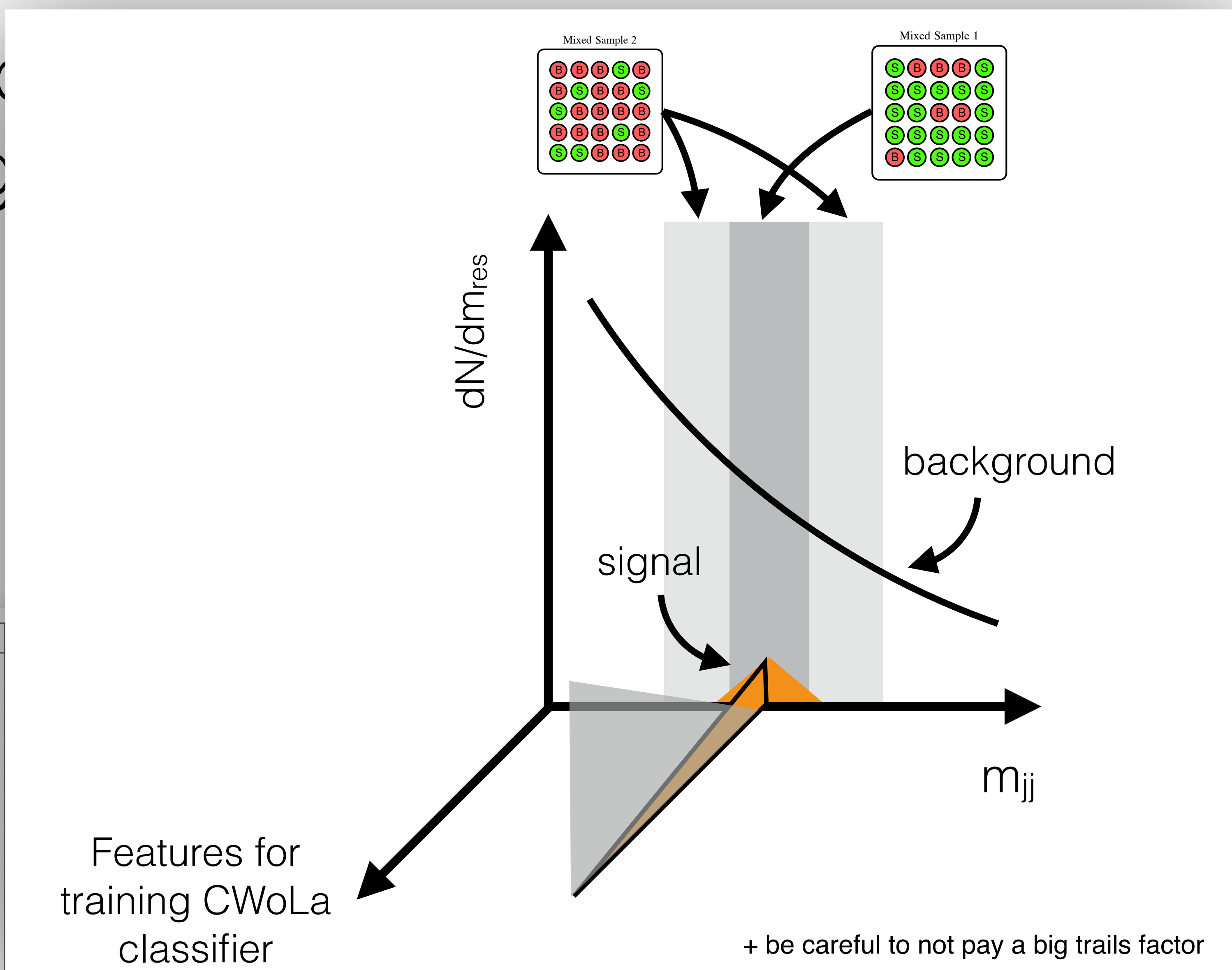
e.g. Classification Without Labels (CWoLa), events in a signal region are labeled "signal" and events in a sideband are labeled "background". These labels are "noisy" but a classifier trained with them can detect the presence of a signal.

E. Metodiev, BN, J. Thaler, 1708.02949; J. Collins, K. Howe, BN, 1805.02664

Solutions: Weakly-supervised

Weakly-supervised = noisy labels

Typ… to look for events with hig… *…ssibly signal-depleted)*



…fication Without Labels …ents in a signal region are …” and events in a sideband …ackground”. These labels …ut a classifier trained with …ct the presence of a signal.

Section			
3.1			
3.2			
3.3			
3.4			
3.5			
3.6			
3.7			
3.8			
3.9			
4.1			
4.2			
4.3	Tag N' Train	Weakly Supervised	(i) (BB1-3)
4.4	SALAD	Weakly Supervised	(iii)
4.5	SA-CWoLa	Weakly Supervised	(iii)
5.1	Deep Ensemble	Semisupervised	(i) (BB1)
5.2	Factorized Topics	Semisupervised	(iii)
5.3	QUAK	Semisupervised	(i) (BB2,3) and (ii) (BB1)
5.4	LSTM	Semisupervised	(i) (BB1-3)

E. Metodiev, BN, J. Thaler, 1708.02949; J. Collins, K. Howe, BN, 1805.02664

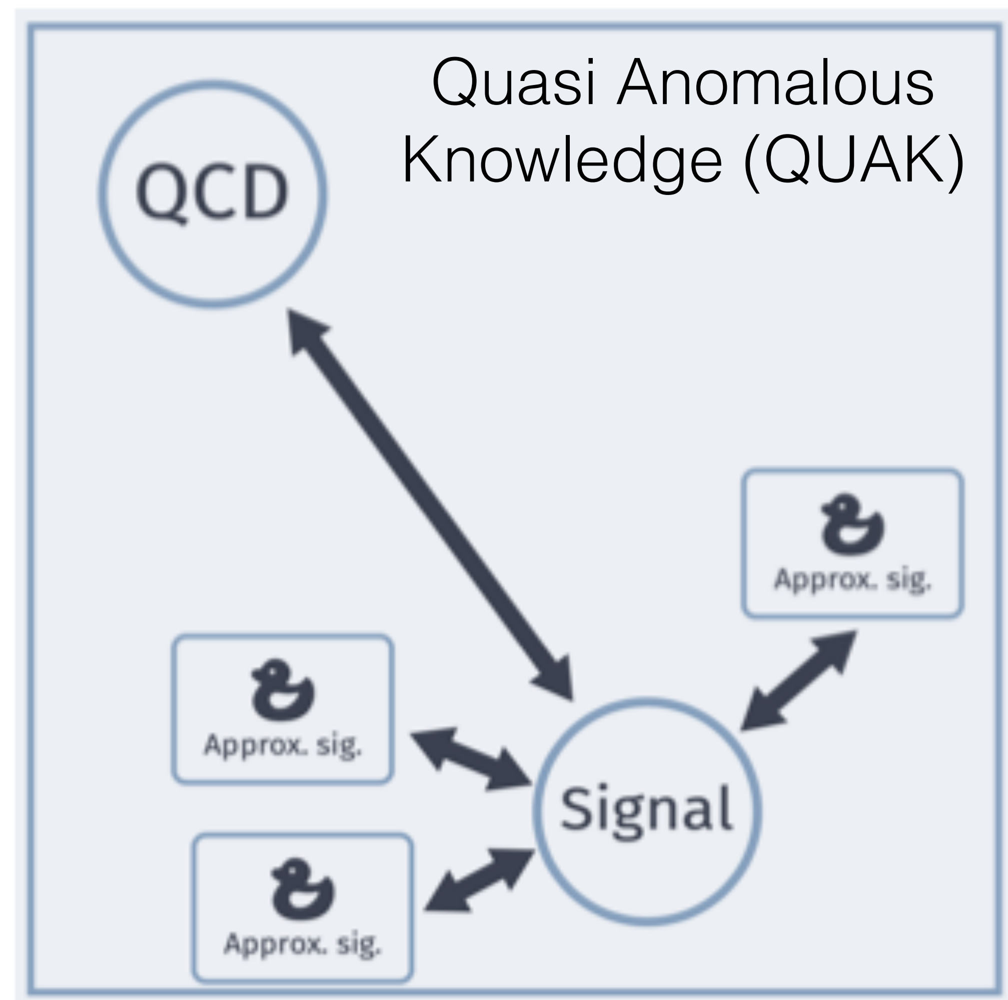
Solutions: Semi-supervised

Semi-supervised = partial labels

Typically, these methods use some signal simulations to build signal sensitivity

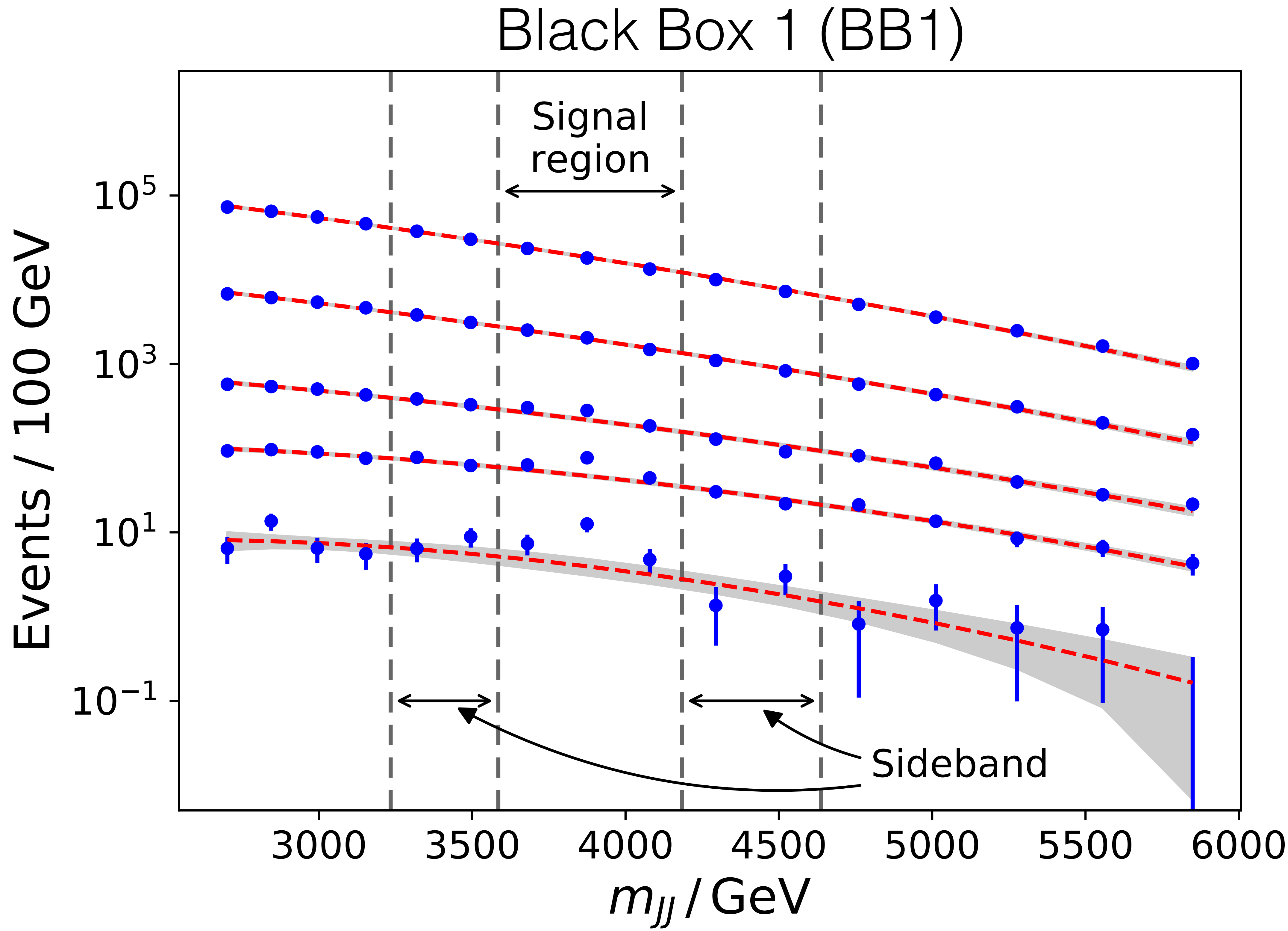
(We did not give bonus points for the best acronyms !)

Section	Short Name	Method Type	Results Type
3.1	VRNN	Unsupervised	(i) (BB2,3) and (ii) (BB1)
3.2	ANODE	Unsupervised	(iii)
3.3	BuHuLaSpa	Unsupervised	(i) (BB2,3) and (ii) (BB1)
3.4	GAN-AE	Unsupervised	(i) (BB2-3) and (ii) (BB1)
3.5	GIS	Unsupervised	(i) (BB1)
3.6	LDA	Unsupervised	(i) (BB1-3)
3.7	PGA	Unsupervised	(ii) (BB1-2)
3.8	Reg. Likelihoods	Unsupervised	(iii)
3.9	UCluster	Unsupervised	(i) (BB2-3)
4.1	CWoLa	Weakly Supervised	(ii) (BB1-2)
4.2	CWoLa AE Compare	Weakly/Unsupervised	(iii)
4.3	Tag N' Train	Weakly Supervised	(i) (BB1-3)
4.4	SALAD	Weakly Supervised	(iii)
4.5	SA-CWoLa	Weakly Supervised	(iii)
5.1	Deep Ensemble	Semisupervised	(i) (BB1)
5.2	Factorized Topics	Semisupervised	(iii)
5.3	QUAK	Semisupervised	(i) (BB2,3) and (ii) (BB1)
5.4	LSTM	Semisupervised	(i) (BB1-3)



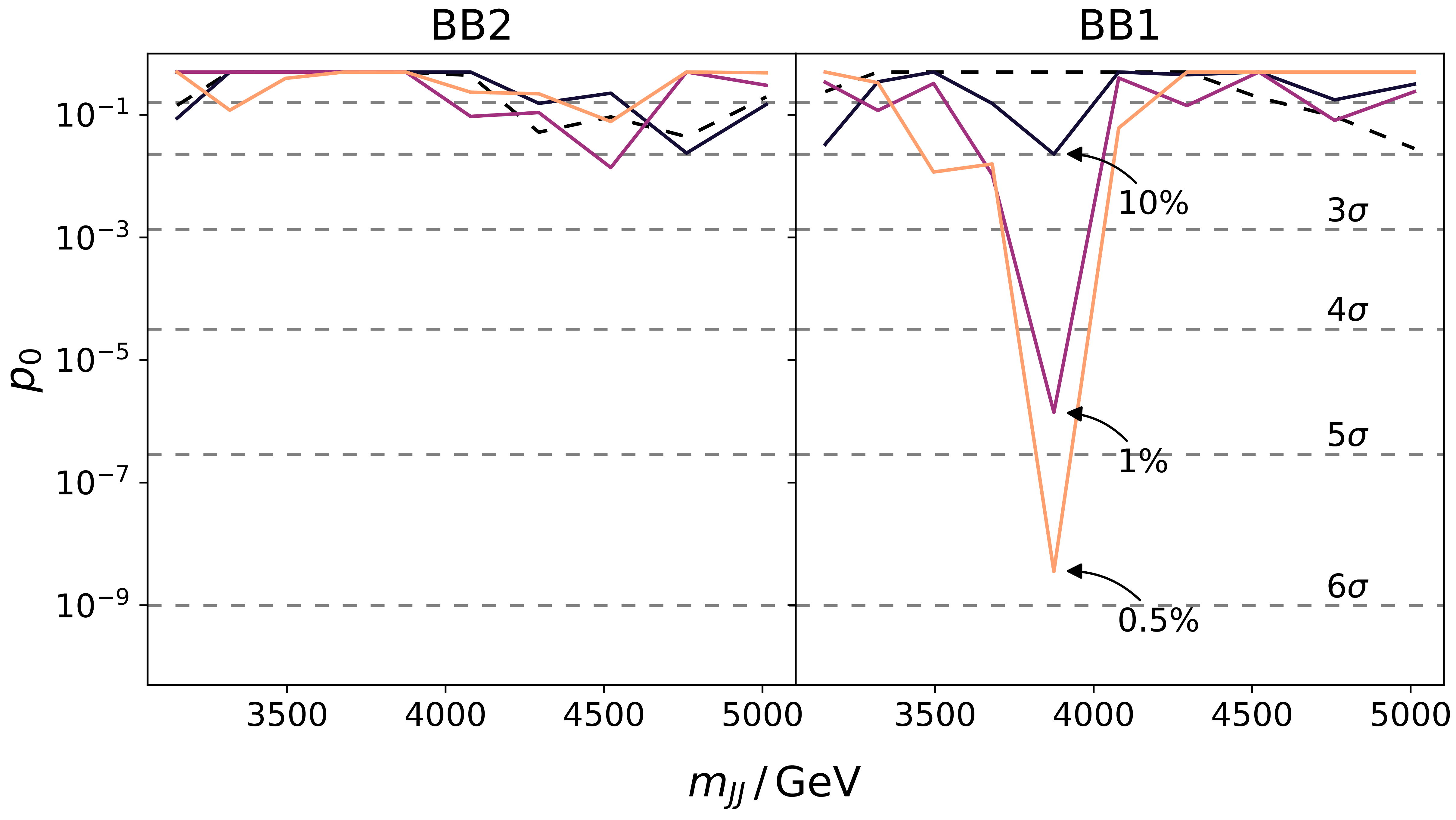
S. Park, D. Rankin, S.-M. Udrescu, M. Yunus, P. Harris, 2011.03550

CWoLa Hunting on the LHCO



J. Collins and BN, 2101.08320

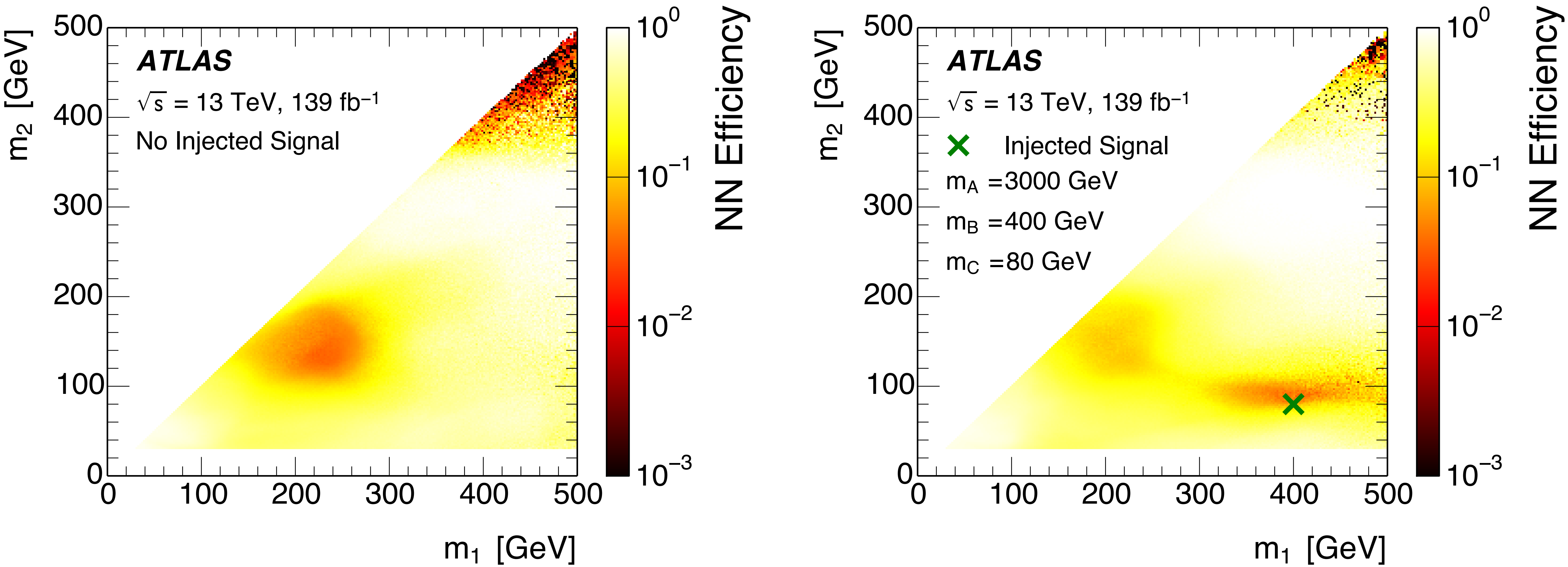
CWoLa Hunting on the LHCO



J. Collins and BN, 2101.08320

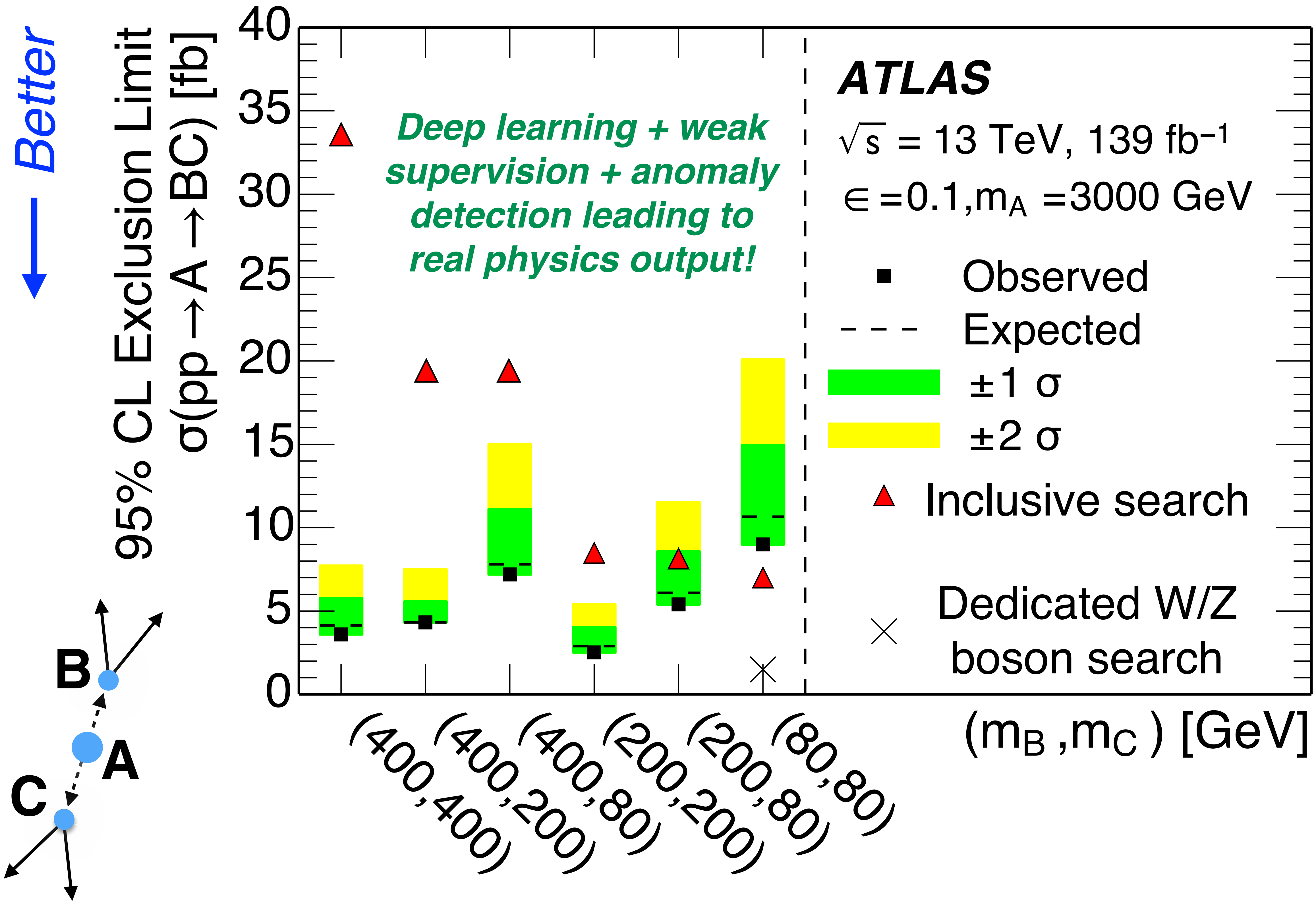
CWoLa Hunting with ATLAS Data

ATLAS Collaboration
PRL 125 (2020) 131801, 2005.02983



First round, keep it simple: feature space is 2D (jet masses)

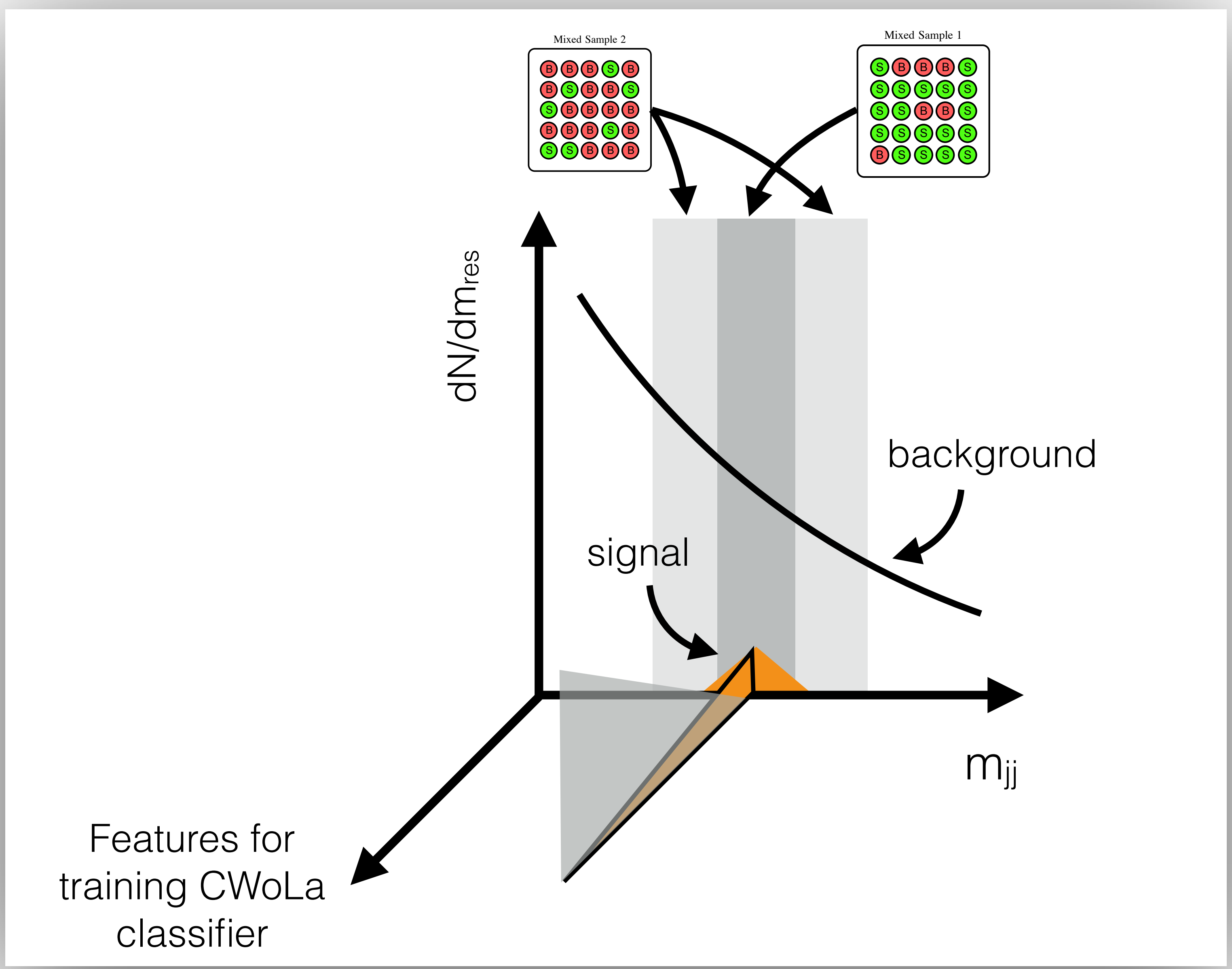
CWoLa Hunting with ATLAS Data



ATLAS Collaboration
PRL 125 (2020) 131801, 2005.02983

CWoLa hunting at e^+e^-?

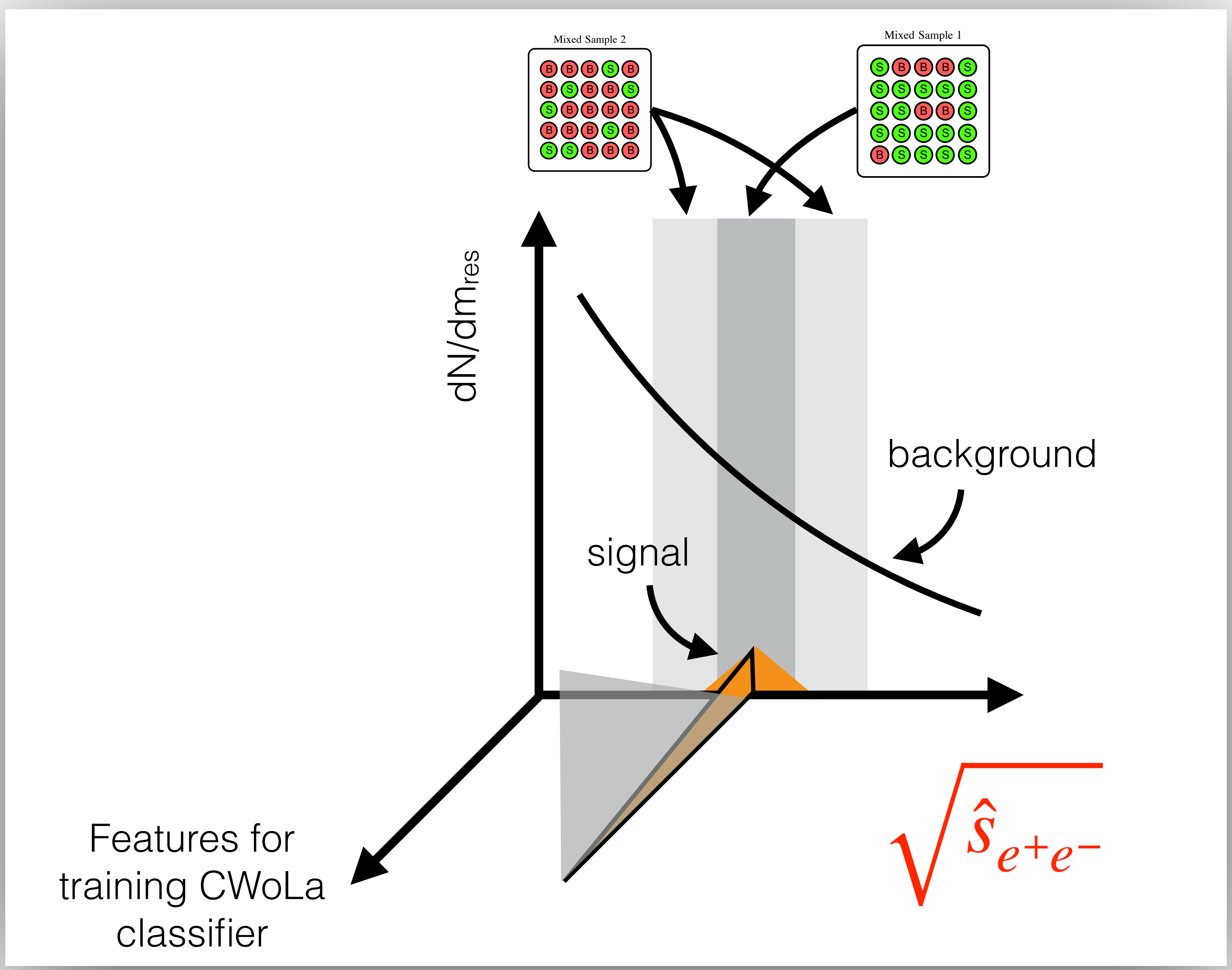
To apply CWoLa, need a resonant feature



CWoLa hunting at e^+e^-?

To apply CWoLa, need a resonant feature

…we can scan an invariant mass in e^+e^- with radiative return!



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

Julia Gonski,[a] Jerry Lai,[b] Benjamin Nachman,[c,d] and Inês Ochoa[e]

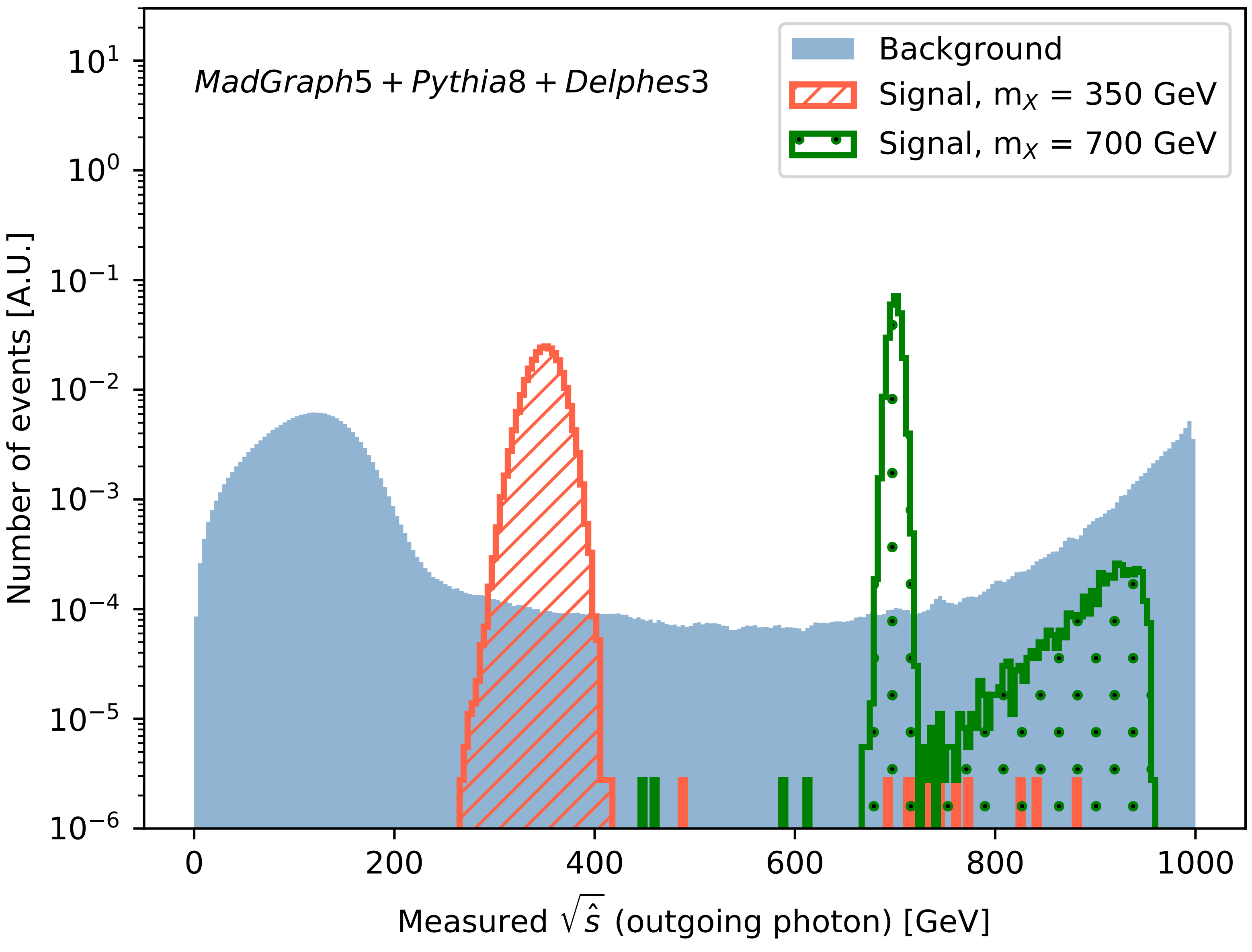
[a] *Nevis Laboratories, Columbia University, 136 S Broadway, Irvington NY, USA*
[c] *Department of Electrical Engineering and Computer Sciences, University of California, Berkeley, CA 94720, USA*
[c] *Physics Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA*
[d] *Berkeley Institute for Data Science, University of California, Berkeley, CA 94720, USA*
[e] *Laboratory of Instrumentation and Experimental Particle Physics, Lisbon, Portugal*

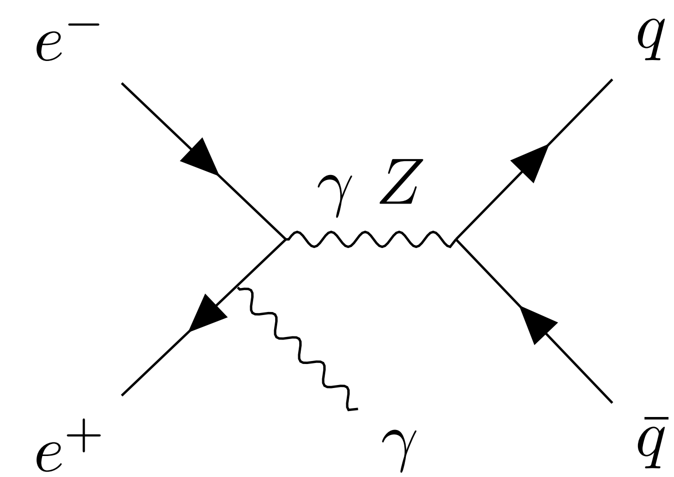
E-mail: julia.gonski@cern.ch, thejerrylai@berkeley.edu, bpnachman@lbl.gov, ines.ochoa@cern.ch

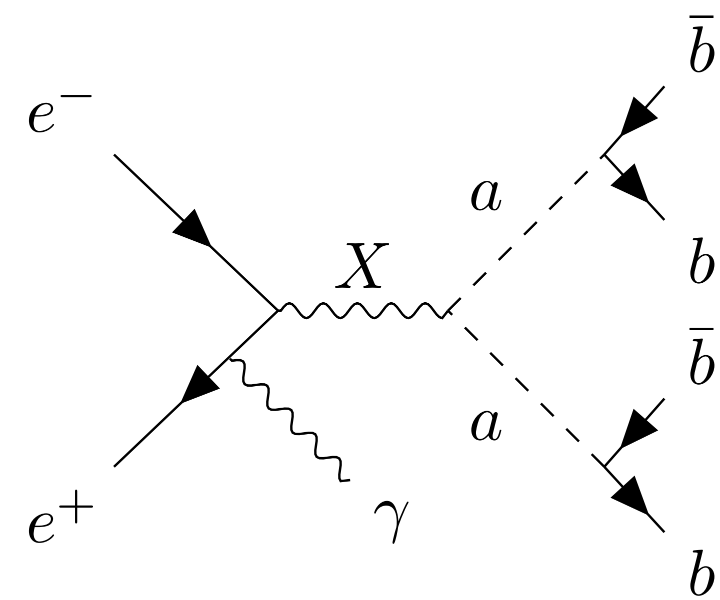
Abstract: Experiments at a future e^+e^- collider will be able to search for new particles with masses below the nominal centre-of-mass energy by analyzing collisions with initial-state radiation (radiative return). We show that machine learning methods based on semisupervised and weakly supervised learning can achieve model-independent sensitivity to the production of new particles in radiative return events. In addition to a first application of these methods in e^+e^- collisions, our study is the first to combine weak supervision with variable-dimensional information by deploying a deep sets neural network architecture. We have also investigated some of the experimental aspects of anomaly detection in radiative return events and discuss these in the context of future detector design.

CWoLa hunting at e+e-?

1 TeV* e+e- radiative return, reconstruct COM energy







*There is nothing special about 1 TeV - we choose it for illustration purposes only

Setup

MadGraph + Pythia + ILD Delphes $|\eta| < 4$

Deep Sets Classifier as EnergyFlow/Particle Flow Networks*

4-vectors for all jets + 5 *n*-subjettiness# variables
+ 4 bit b-tagging discriminant

Scale all energies by the H_T

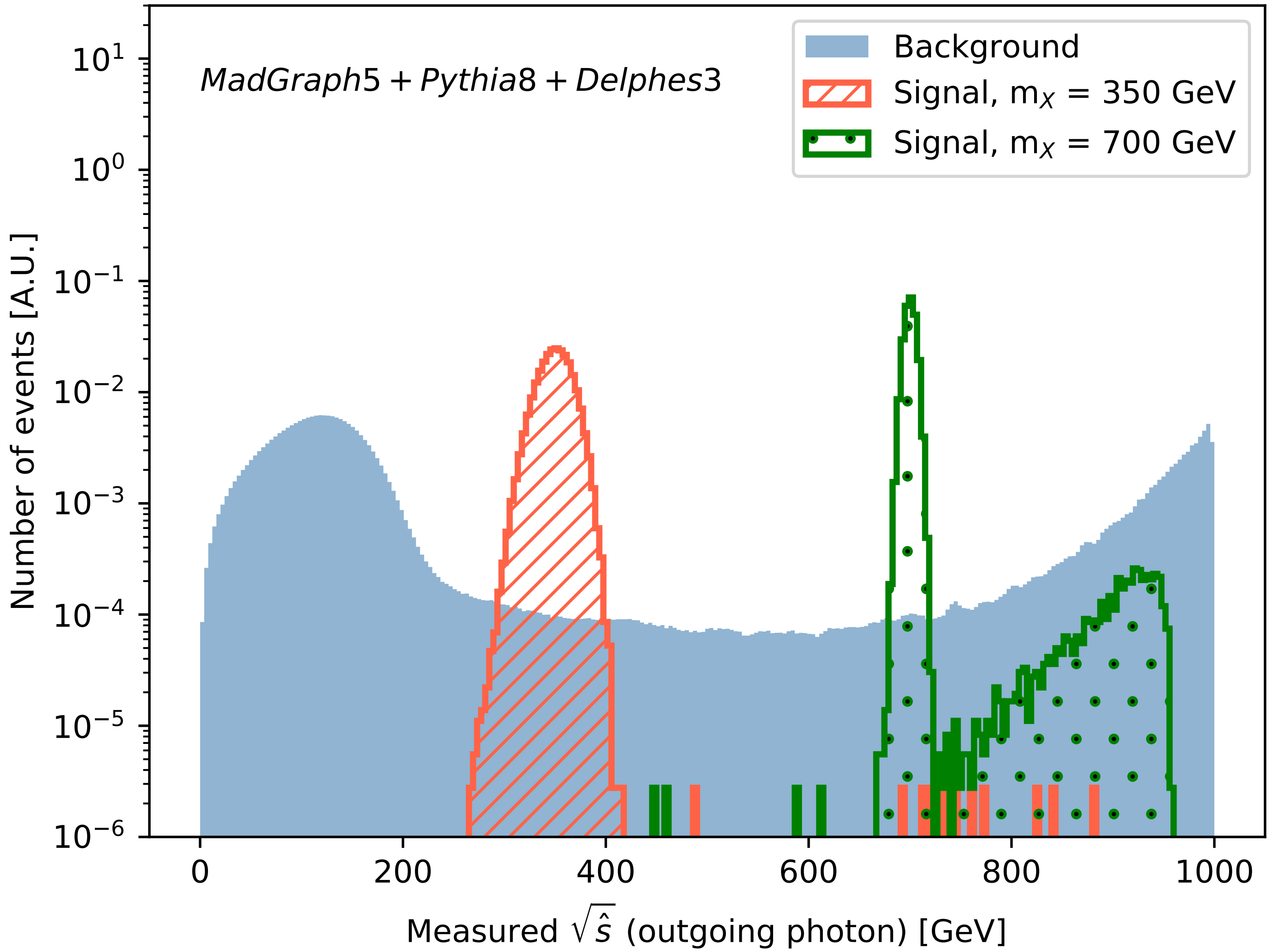
N.B. high-and variable-dimensional!

#J. Thaler, K. Van Tilburg, 1108.2701

*P. Komiske, E. Metodiev, J. Thaler, 1810.05165, https://energyflow.network

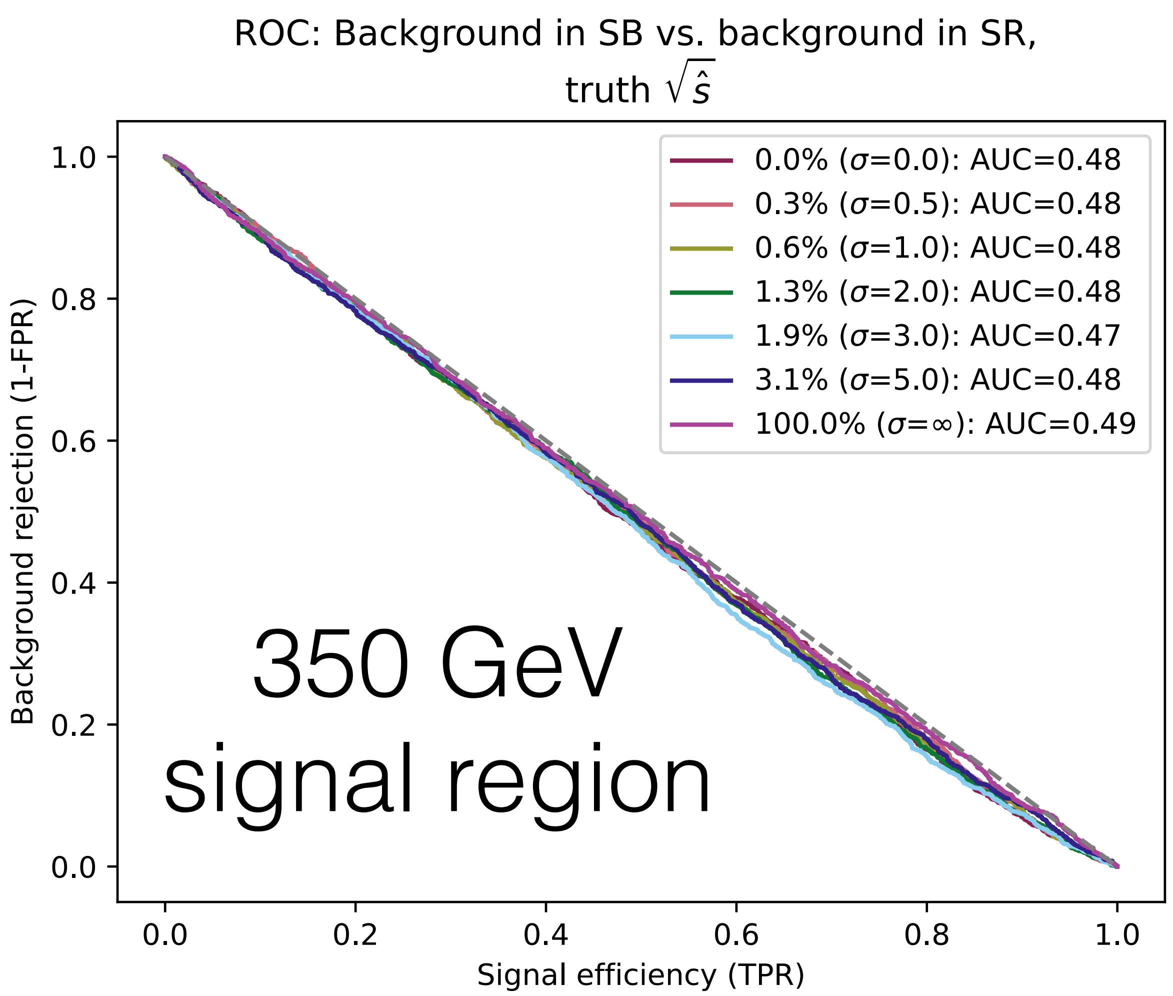
Setup (continued)

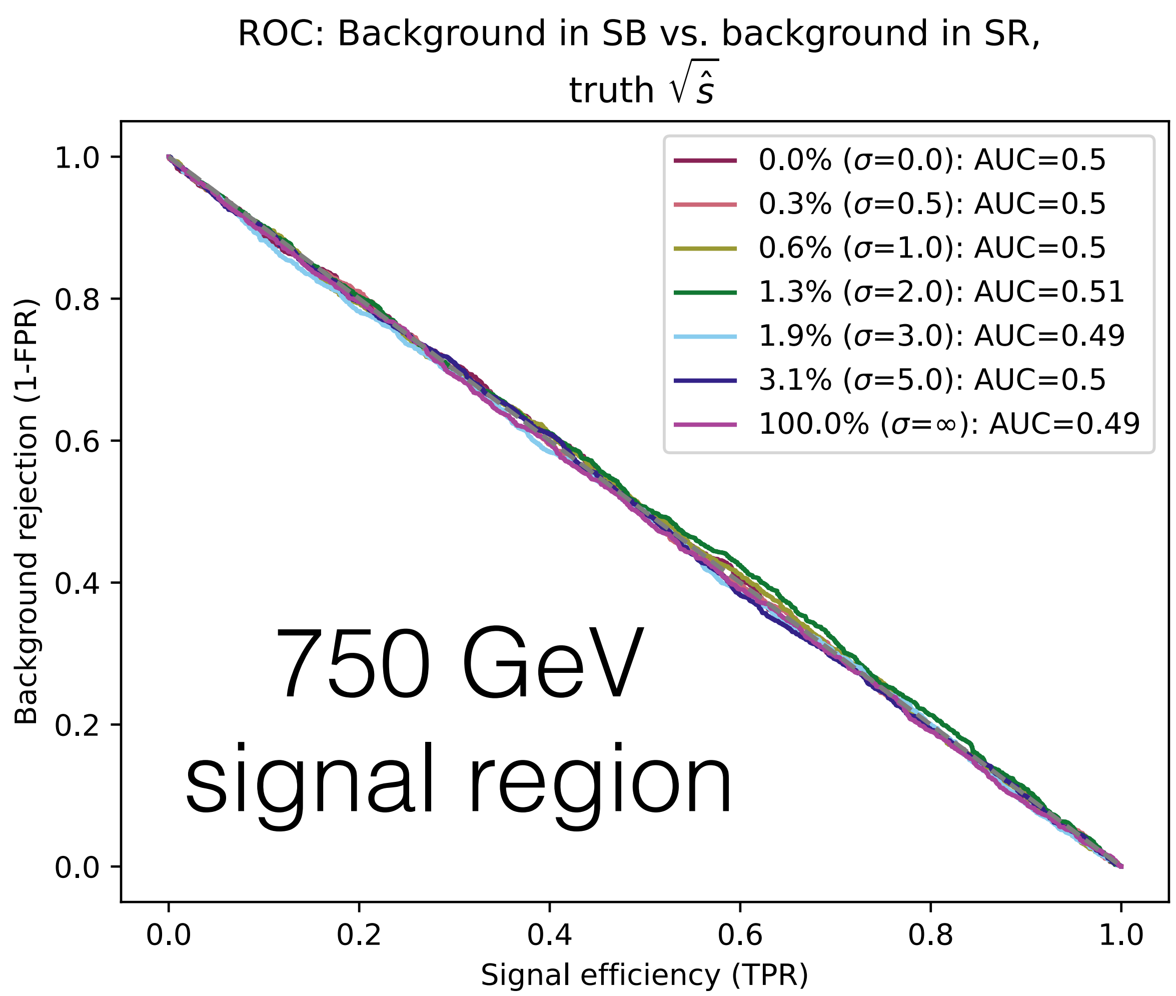
	Signal region [GeV]	Sideband region [GeV]
m_X, m_a= 350 GeV, 40 GeV	[325, 375)	[275, 325) ∪ [375, 425)
m_X, m_a= 700 GeV, 100 GeV	[675, 725)	[625, 675) ∪ [725, 775)



J. Gonski, J. Lai, BN, I. Ochoa, 2108.13451

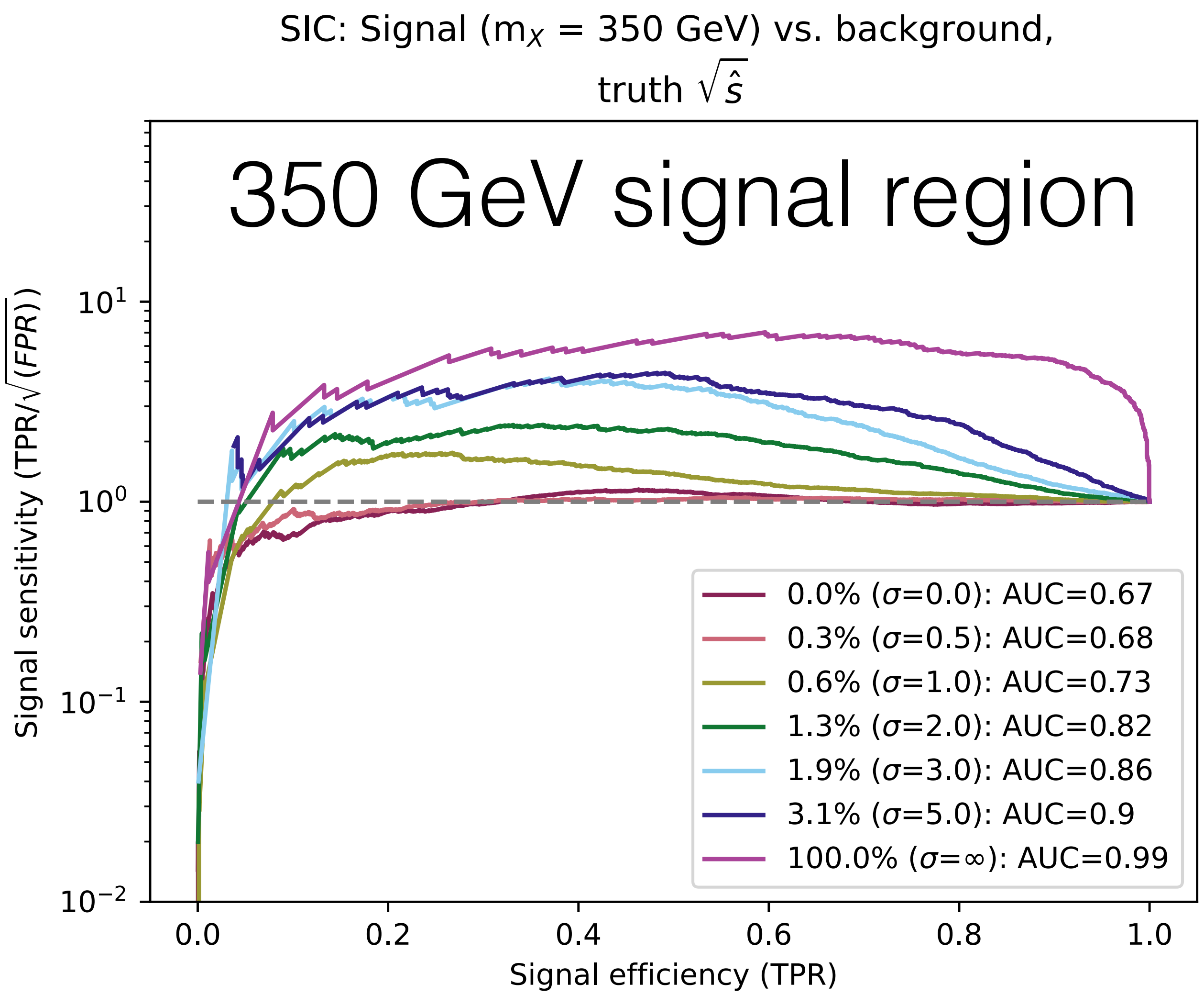
Background-only

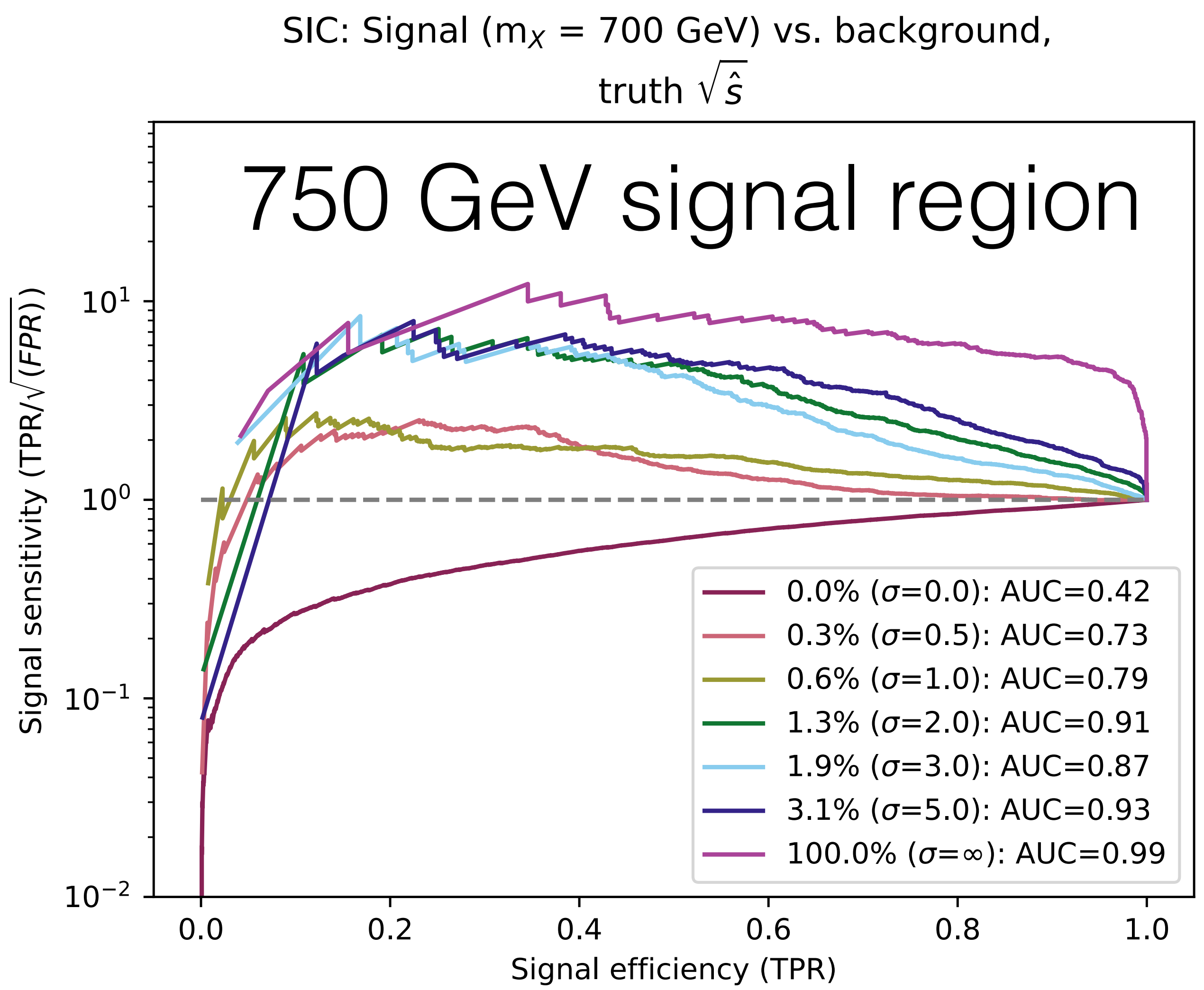




When no signal, does not find anything (H_T scaling critical!)

Signal Sensitivity

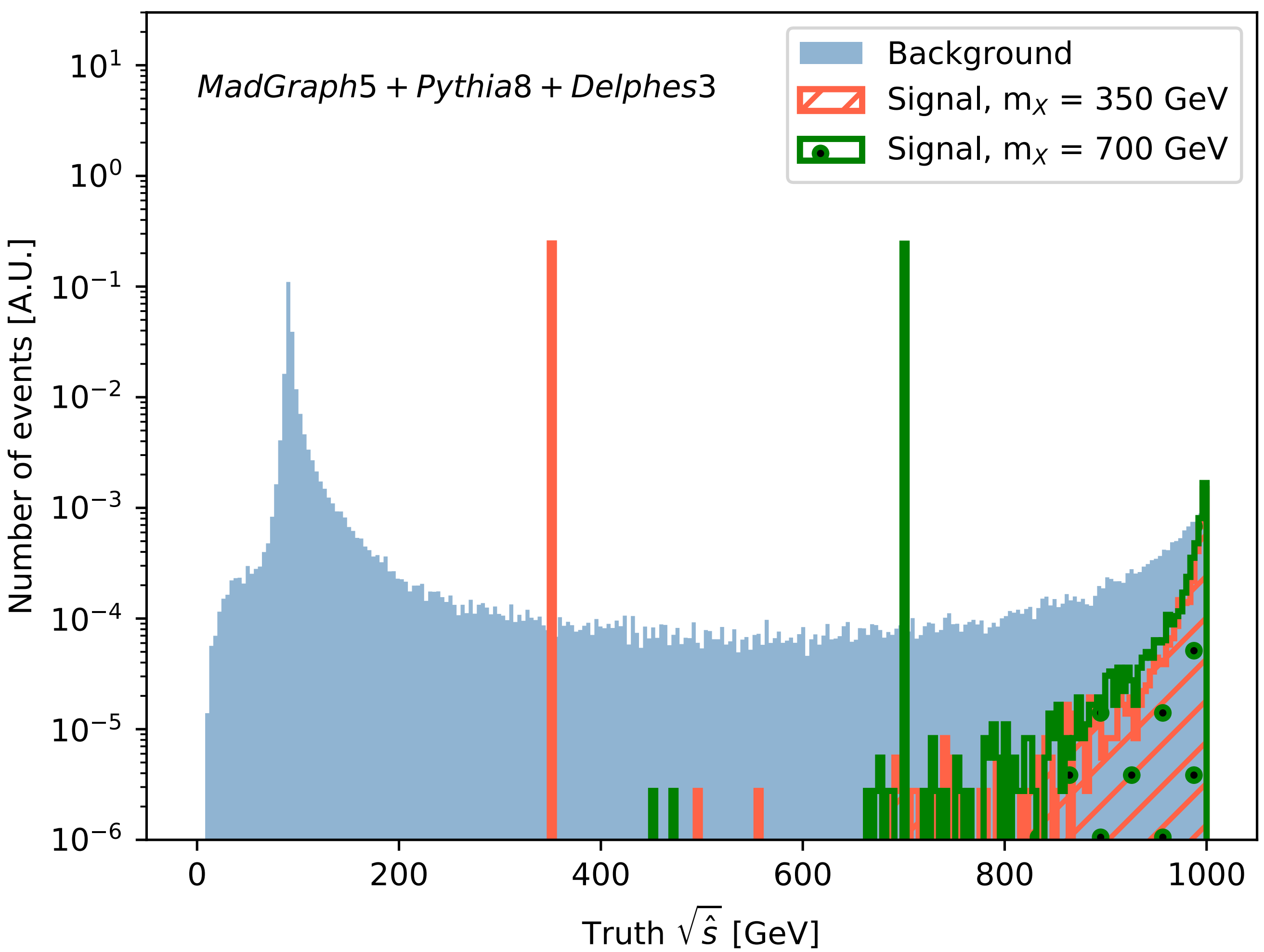


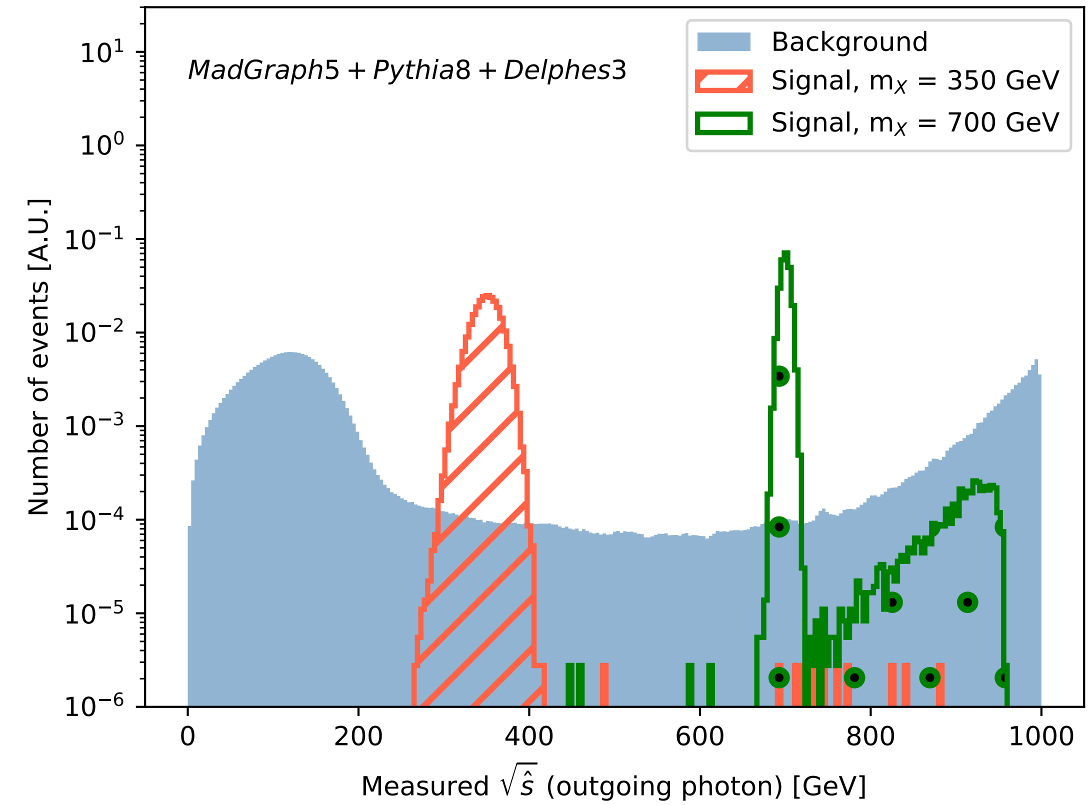


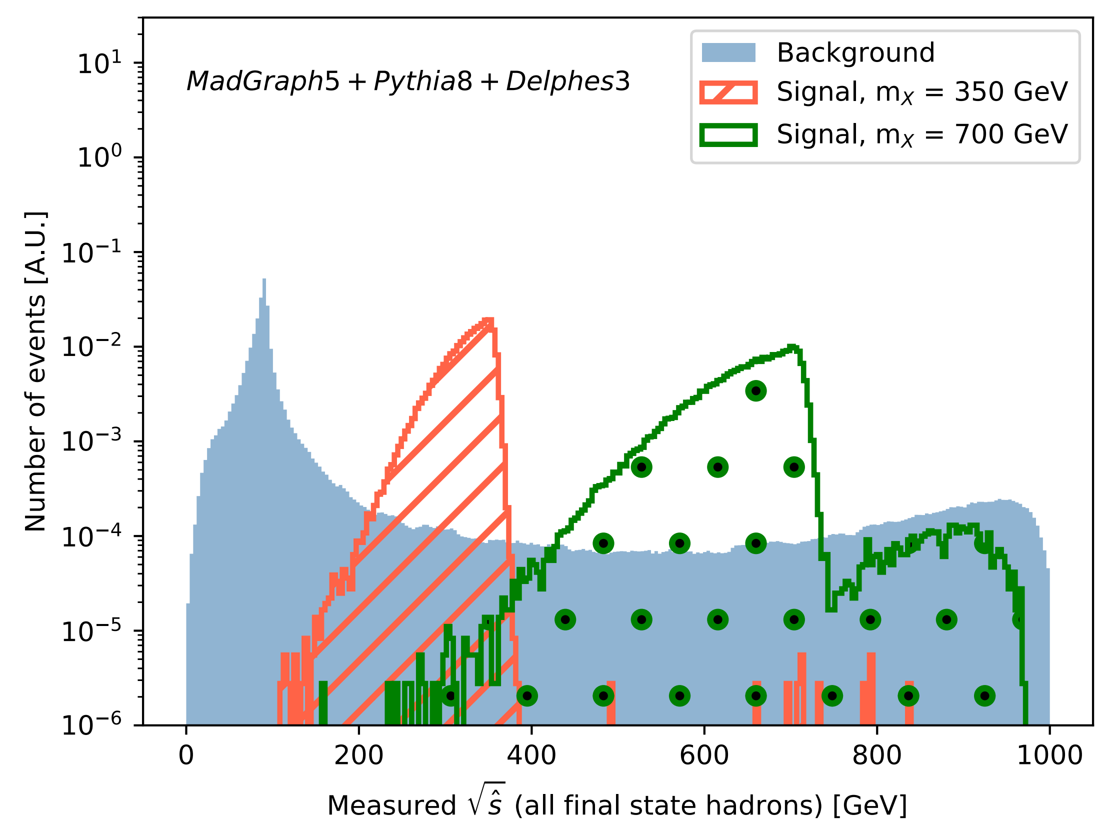
Normalized so > 1 means “better than nothing”

J. Gonski, J. Lai, BN, I. Ochoa, 2108.13451

Detector Considerations

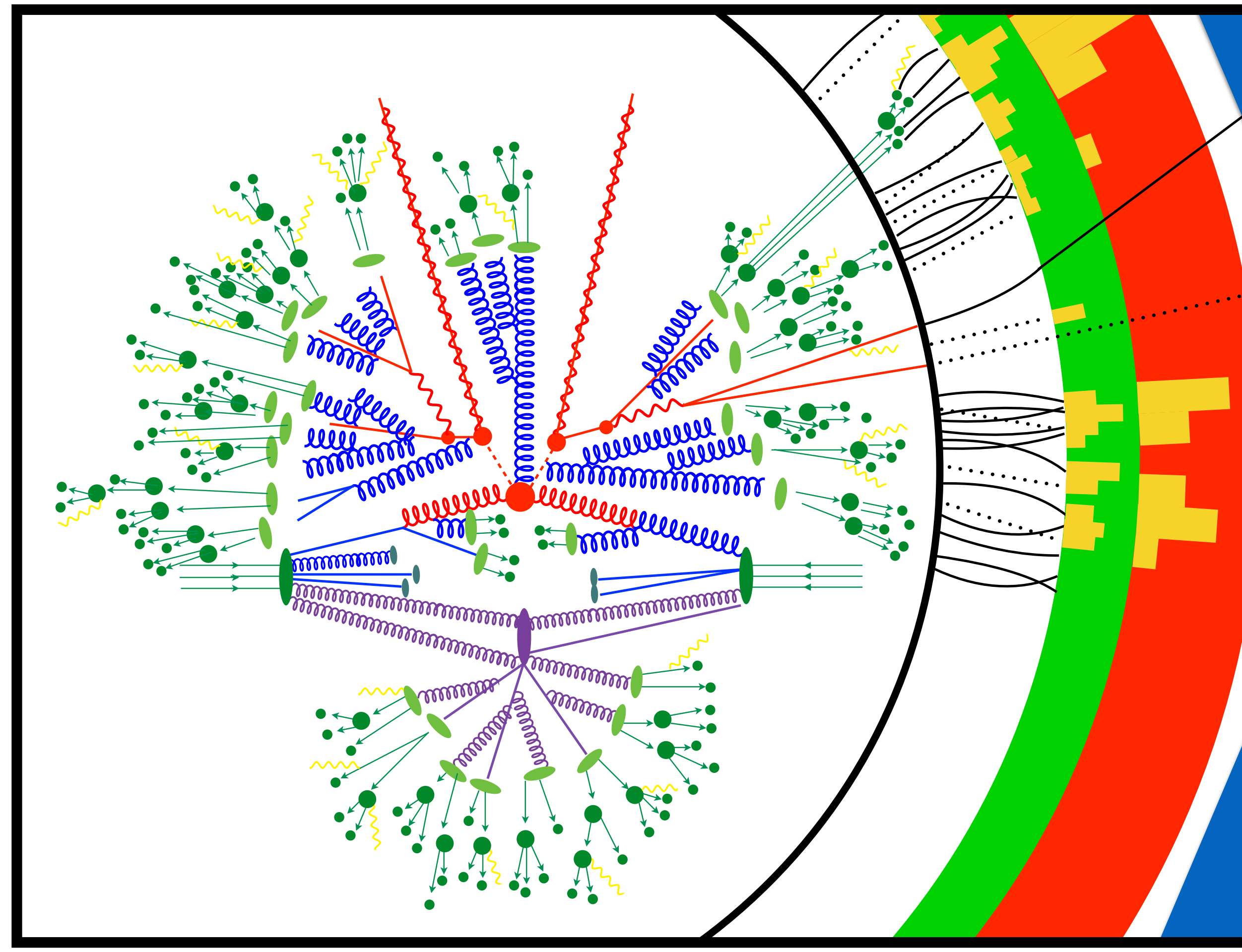






Conclusions and outlook

Deep learning-based anomaly detection is a promising avenue to broaden the energy frontier physics portfolio



I did not cover every proposal
- see the Living Review for more!
Can we extend density estimation techniques like CATHODE to high dimensions?

This methodology can be extended beyond dijets to radiative return in e^+e^-; need to start thinking now about implications for detector, software, and computing!

Backup