

Dijet resonance search with weak supervision using $\sqrt{s}=13$ TeV pp collisions in the **ATLAS** detector CWoLa Hunting in ATLAS

Aviv Cukierman
May 28, 2020



Analysis Status

- [Public webpage](#)
- [arXiv submission](#)

Dijet resonance search with weak supervision using $\sqrt{s} = 13$ TeV pp collisions in the ATLAS detector

6 May 2020

Contact: [ATLAS Higgs and Diboson Searches conveners](#)

Content	Preview
e-print arXiv:2005.02983	pdf from arXiv
Figures	-
Abstract	
This Letter describes a search for resonant new physics using a machine-learning anomaly detection procedure that does not rely on a signal model hypothesis. Weakly supervised learning is used to train classifiers directly on data to enhance potential signals. The targeted topology is dijet events and the features used for machine learning are the masses of the two jets. The resulting analysis is essentially a three-dimensional search $A \rightarrow BC$, for $m_A \sim \mathcal{O}(\text{TeV})$, $m_B, m_C \sim \mathcal{O}(100 \text{ GeV})$ and B, C are reconstructed as large-radius jets, without paying a penalty associated with a large trials factor in the scan of the masses of the two jets. The full Run 2 $\sqrt{s} = 13$ TeV pp collision data set of 139 fb^{-1} recorded by the ATLAS detector at the Large Hadron Collider is used for the search. There is no significant evidence of a localized excess in the dijet invariant mass spectrum between 1.8 and 8.2 TeV. Cross-section limits for narrow-width A, B, and C particles vary with m_A, m_B, and m_C. For example, when $m_A = 3 \text{ TeV}$ and $m_B \gtrsim 200 \text{ GeV}$, a production cross section between 1 and 5 fb is excluded at 95% confidence level, depending on m_C. For certain masses, these limits are up to 10 times more sensitive than those obtained by the inclusive dijet search.	
Figures	

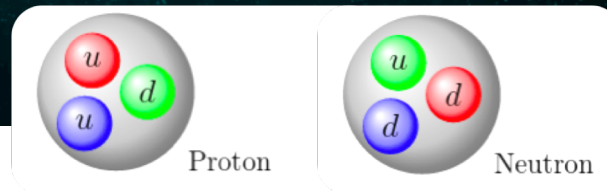
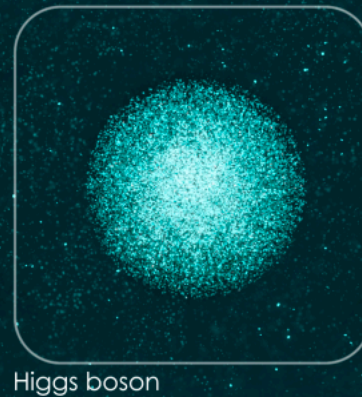
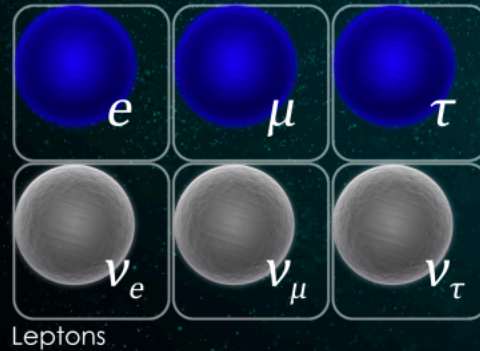
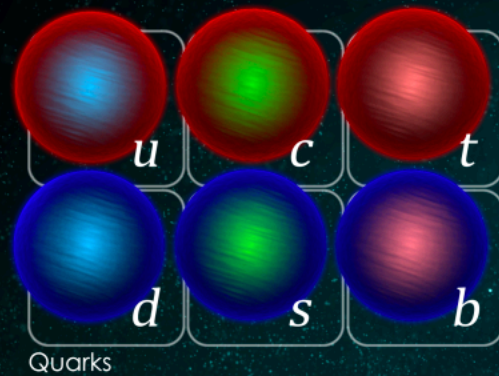


- Introduction:
 - SM, LHC, and ATLAS
 - Dijet resonance searches
 - Classification without Labels (CWoLa)
 - CWoLa Hunting
- CWoLa hunting in ATLAS
- Results
- Conclusion

Introduction

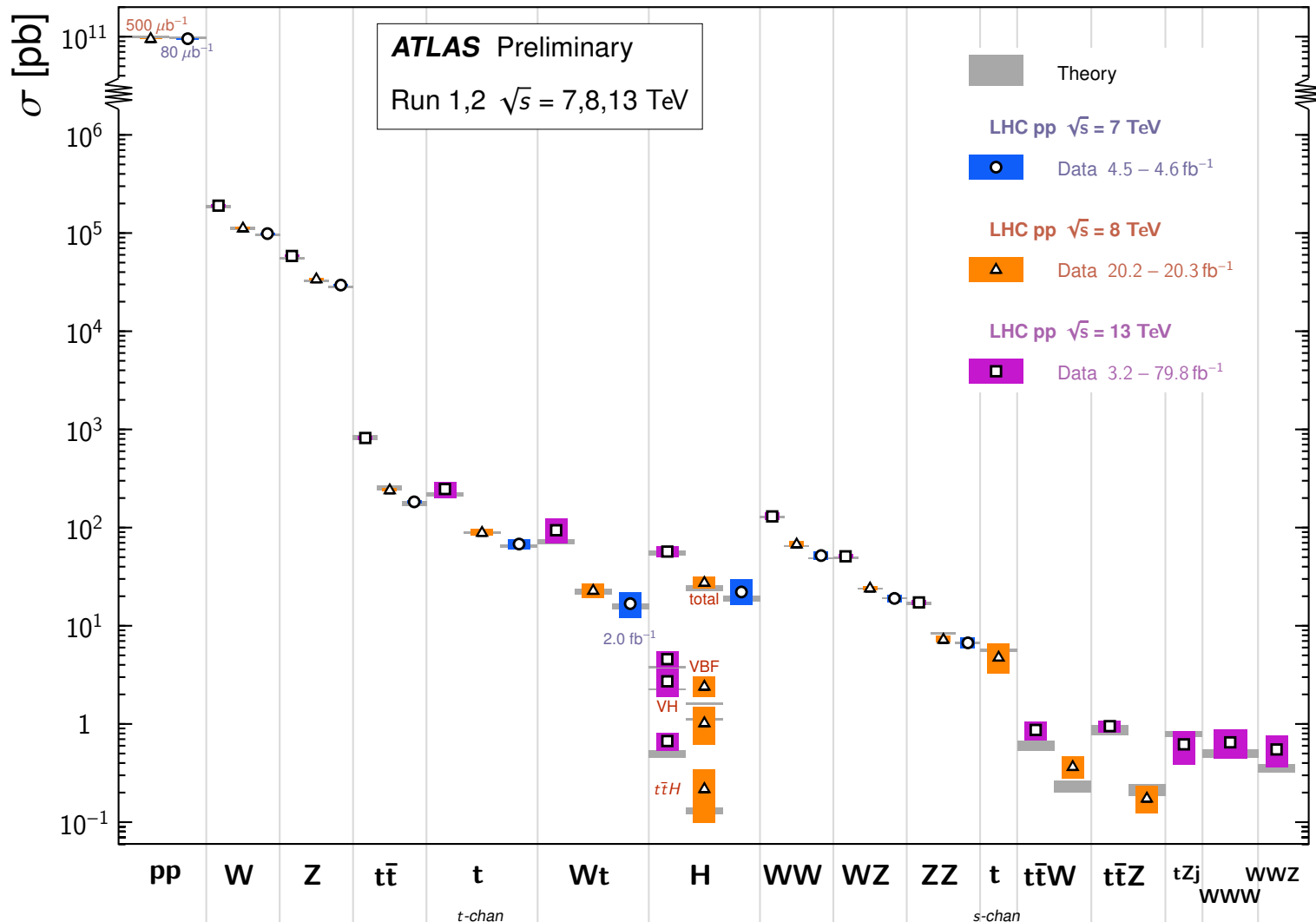
Standard Model and Jets

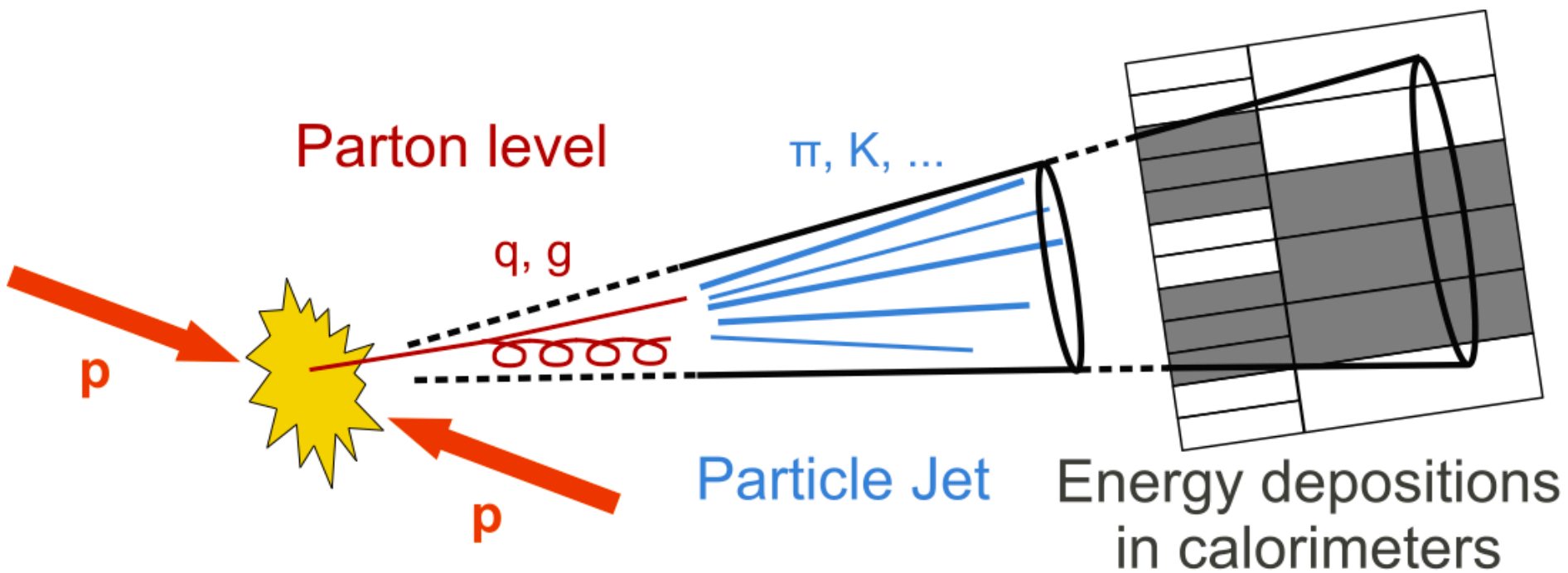
Standard Model of Particle Physics



Standard Model of Particle Physics

Standard Model Total Production Cross Section Measurements Status: November 2019





Beyond the Standard Model

- Problems

- Gravity

- Relevant at much higher energy scales (*Planck scale*)

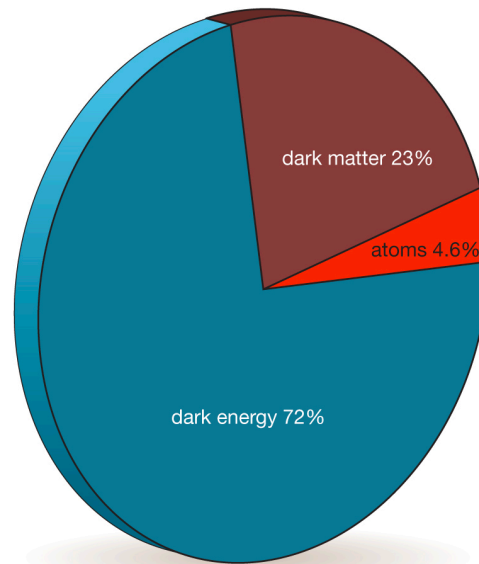
- Hierarchy Problems

$$m_t / m_u \sim 10^5$$

$$m_e / m_\nu \sim 10^6$$

$$m_P / m_h \sim 10^{19}$$

- Dark Matter



- Strong CP Problem; Muon g-2; Matter-antimatter asymmetry

Beyond the Standard Model

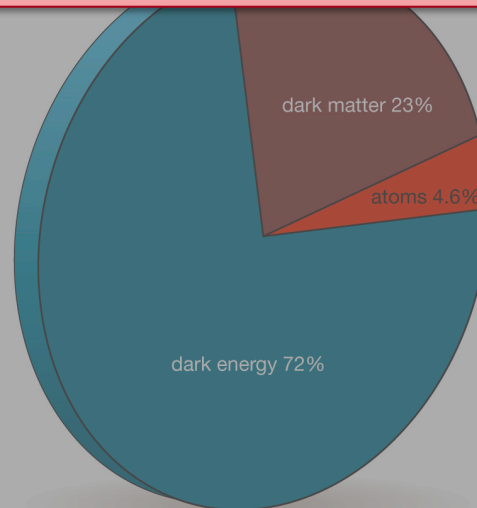
- Problems
 - Gravity
 - Relevant at much higher energy scales (*Planck scale*)
 - Hierarchy Problems

$$m_t / m_u \sim 10^5$$

$$m_e / m_\nu \sim 10^6$$

$$m_P / m_h \sim 10^{19}$$

Many Beyond-the-Standard-Model theories propose the existence of new particles that can be produced and/or detected at the LHC



- Others: Strong CP Problem; Muon $g-2$; Matter-antimatter asymmetry

Beyond the Standard Model

- Problems
 - Gravity
 - Relevant at much higher energy scales (*Planck scale*)
 - Hierarchy Problems

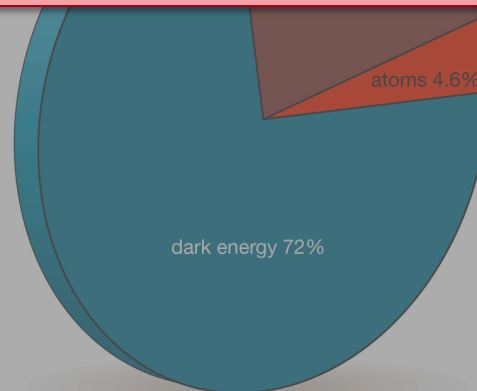
$$m_t / m_u \sim 10^5$$

$$m_e / m_\nu \sim 10^6$$

$$m_P / m_h \sim 10^{19}$$

Many Beyond-the-Standard-Model theories propose the existence of new particles that can be produced and/or detected at the LHC

Many of these new particles decay into **jets**



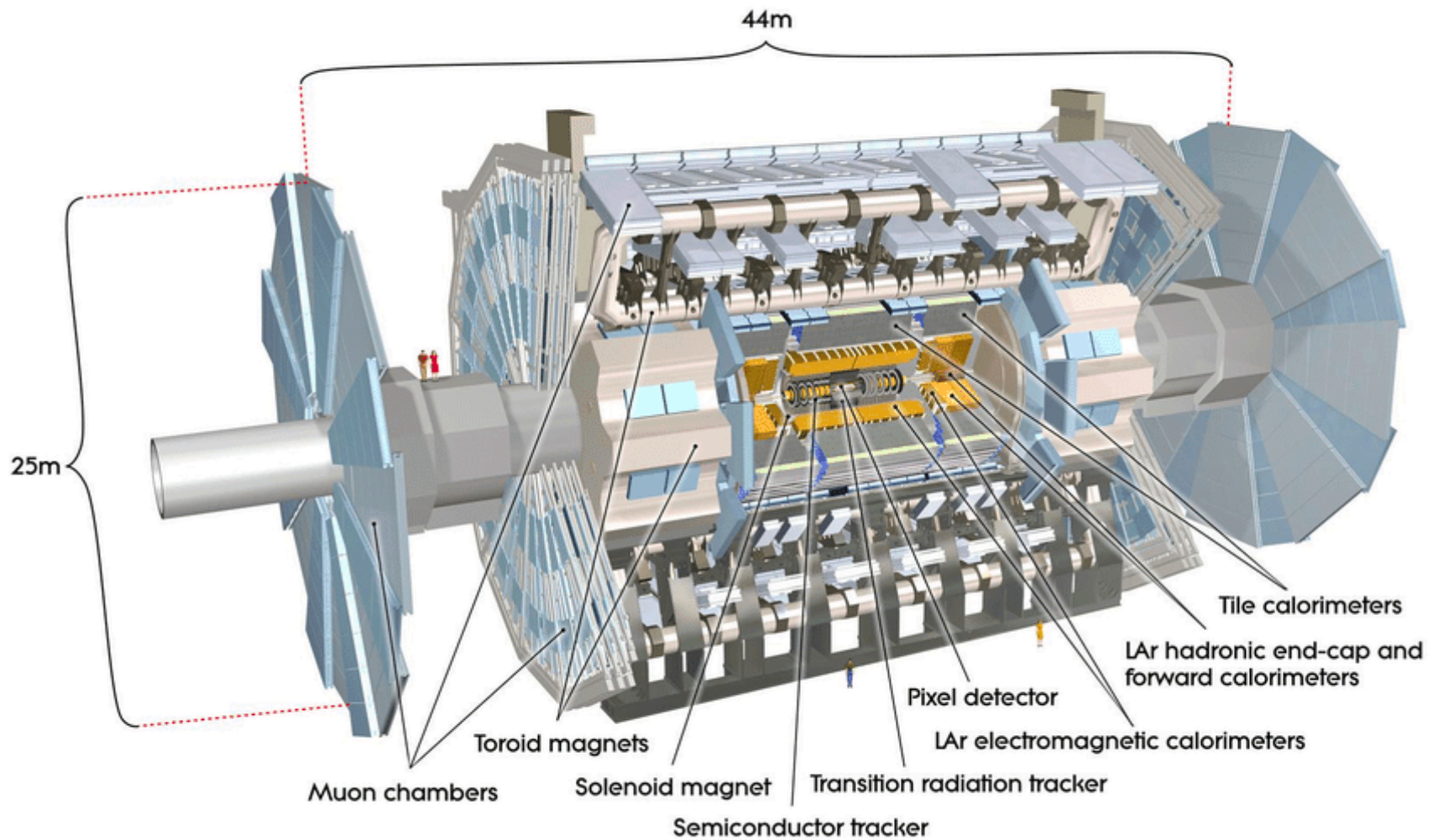
- Others: Strong CP Problem; Muon g-2; Matter-antimatter asymmetry

Introduction

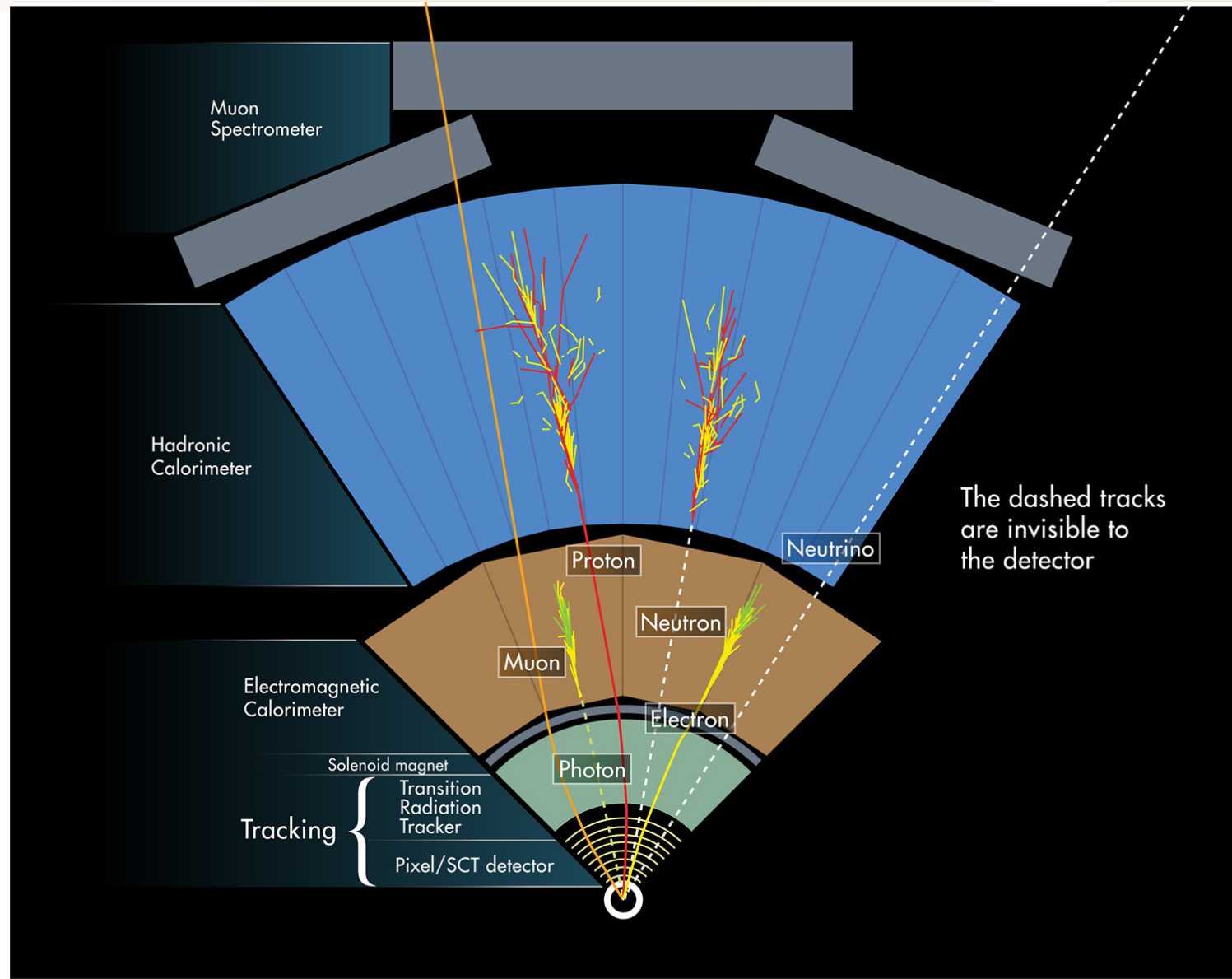
LHC and ATLAS

Large Hadron Collider (LHC)

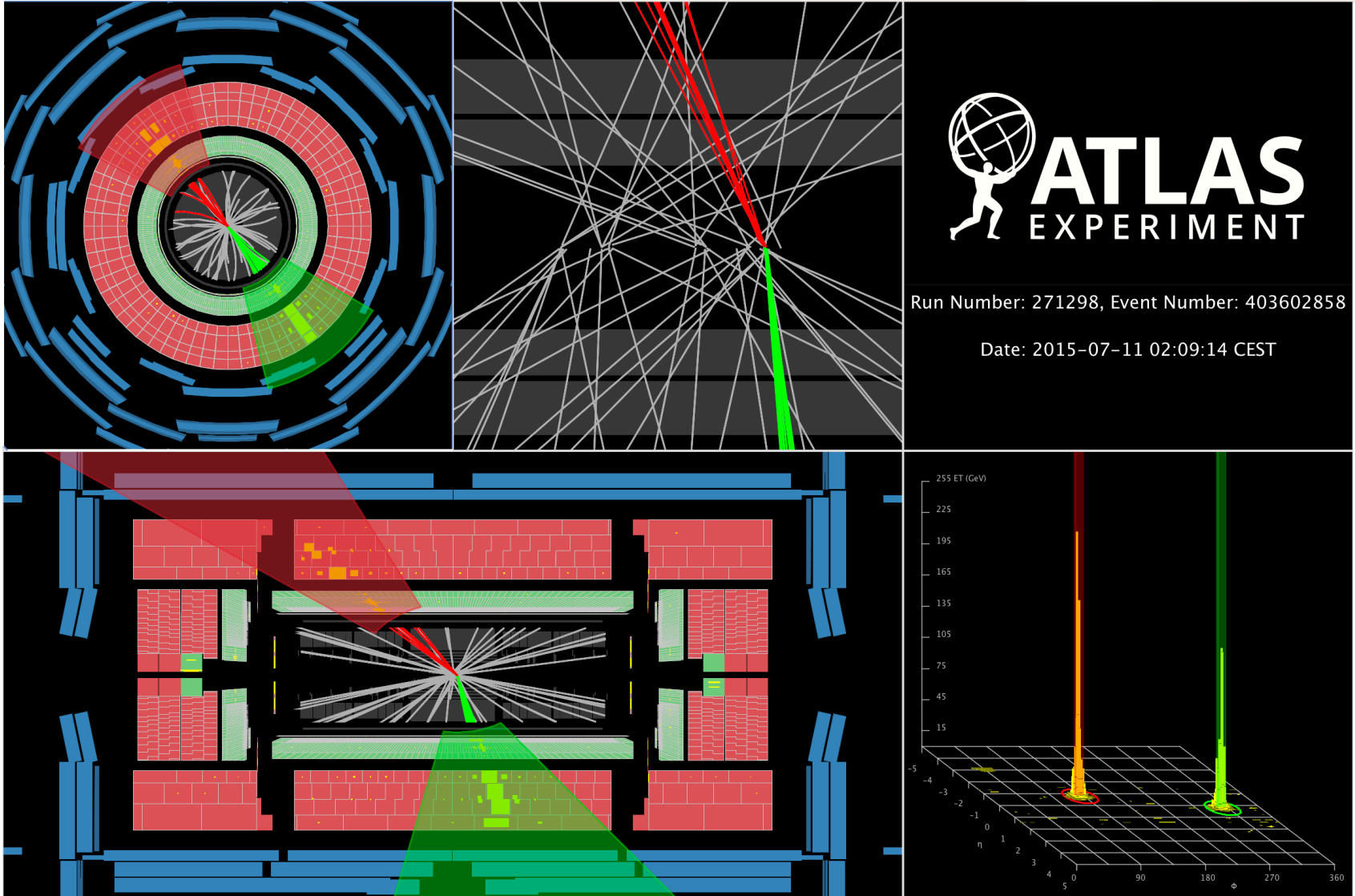




ATLAS Detector



Jets in ATLAS

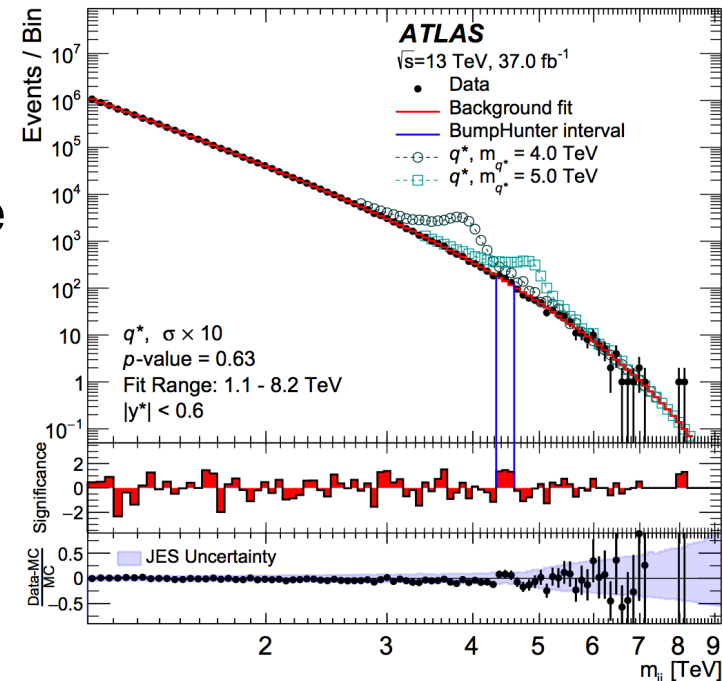


Introduction

Dijet Resonance Searches
Classification without Labels (CWoLa)
and
CWoLa Hunting

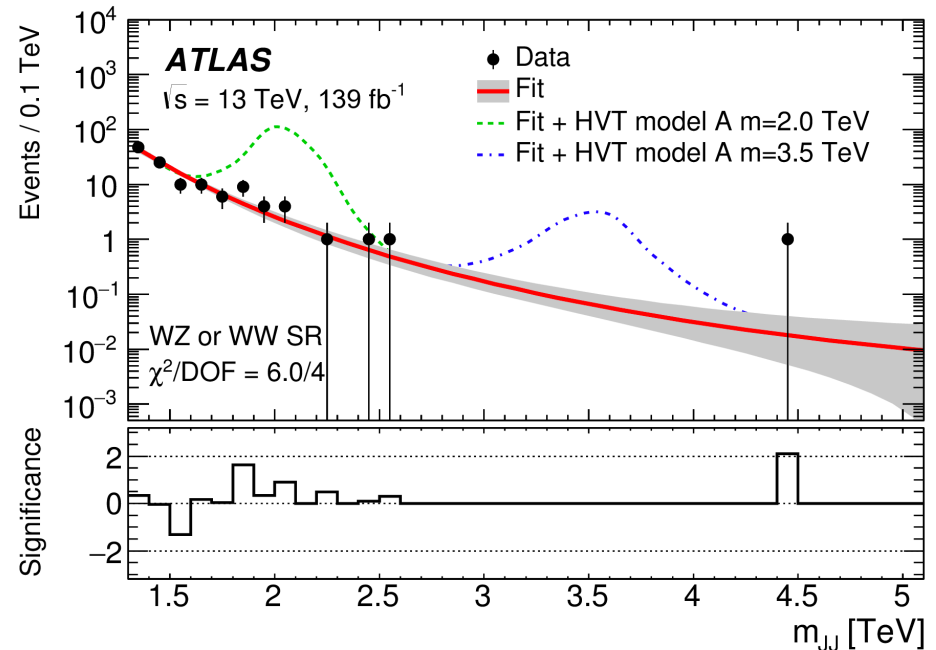
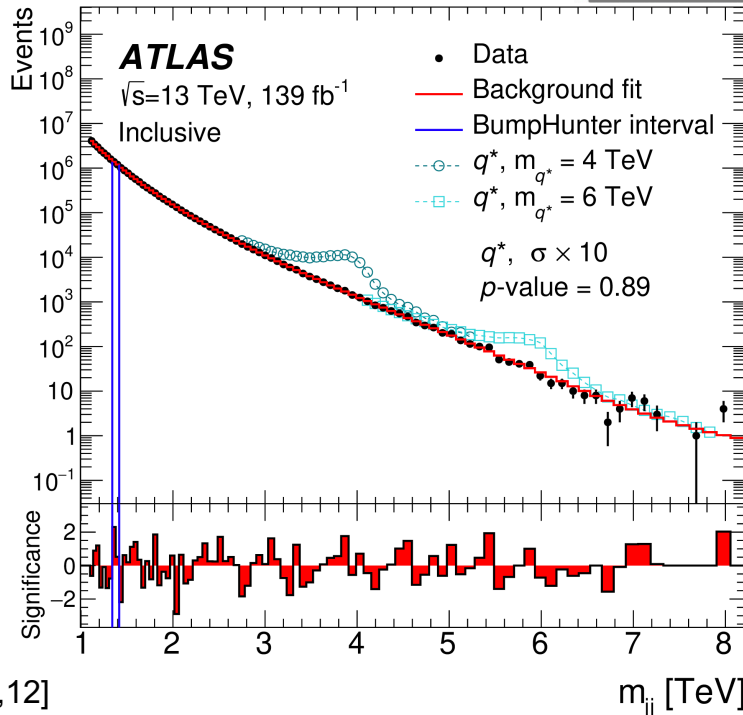
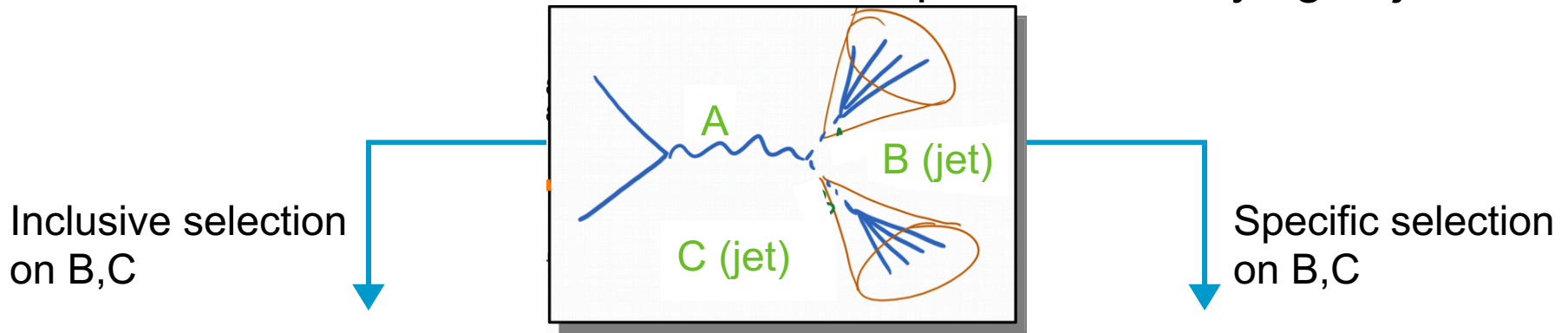
Motivation

- Dijet searches have a long history in ATLAS and in HEP in general
 - Search for “bump” in dijet invariant mass spectrum
- Dedicated searches for $A \rightarrow \{e/\gamma, g, q, b, W, Z, H\} \times \{e/\gamma, g, q, b, W, Z, H\}$
- Need generic search for $A \rightarrow BC$
 - One or both of B,C could be BSM
- CWoLa hunting ([arXiv:1805.02664](https://arxiv.org/abs/1805.02664)):
New bump finding technique sensitive to these exotic signatures
- Builds on dijet search to be more sensitive to a broad class of models



Dijet Resonance Searches

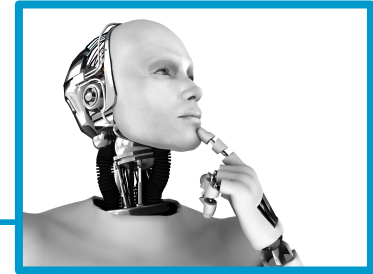
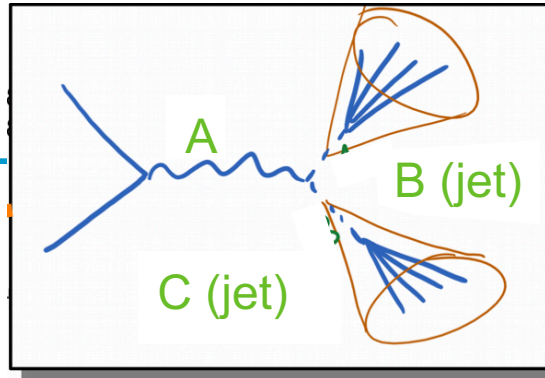
- Generic searches for new massive particles decaying to jets



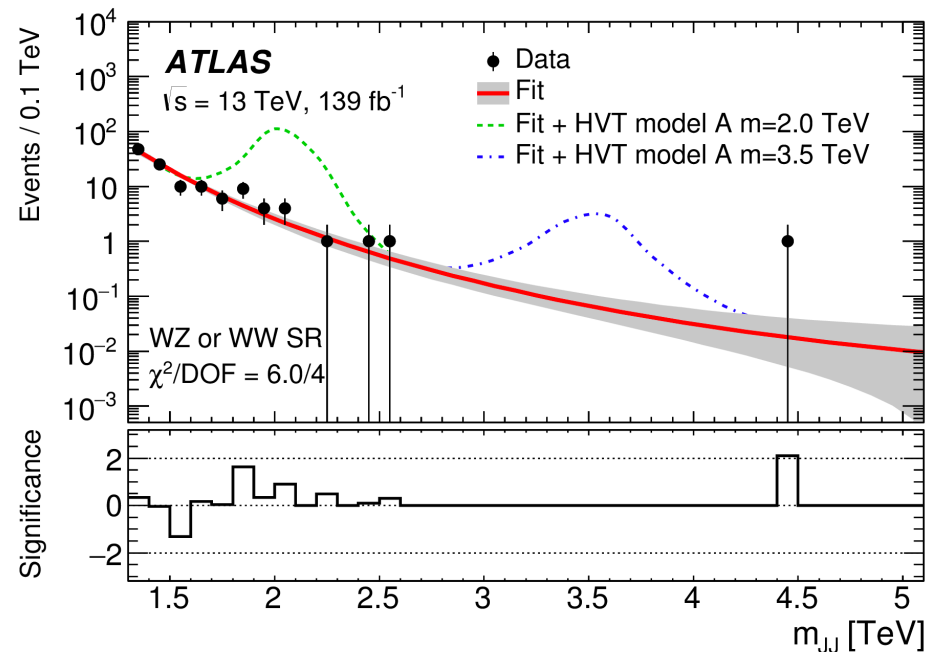
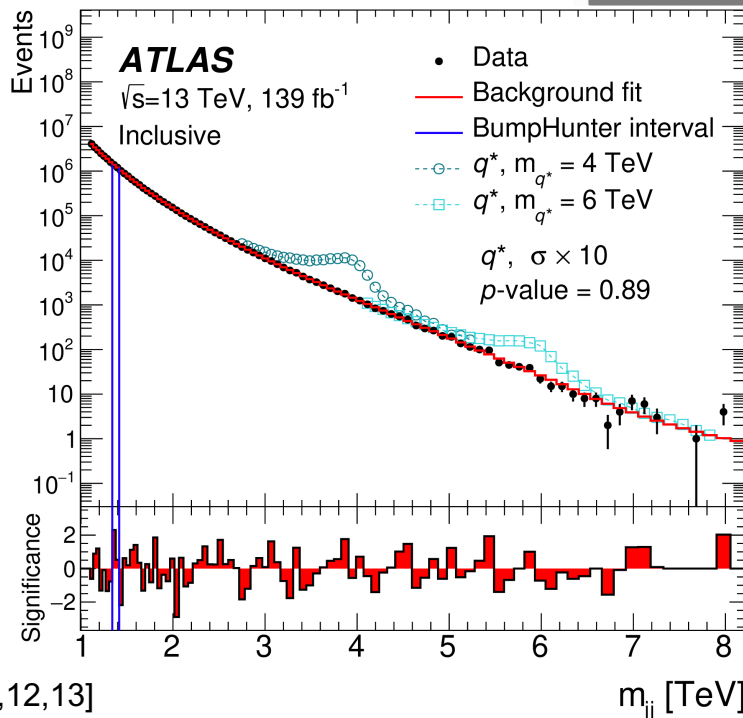
Resonance Searches

- Generic searches for new massive particles

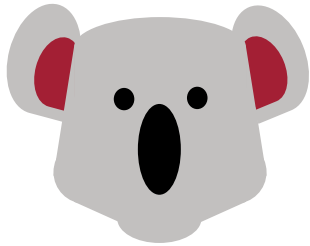
Inclusive selection
on B,C



Learned selection
on B,C

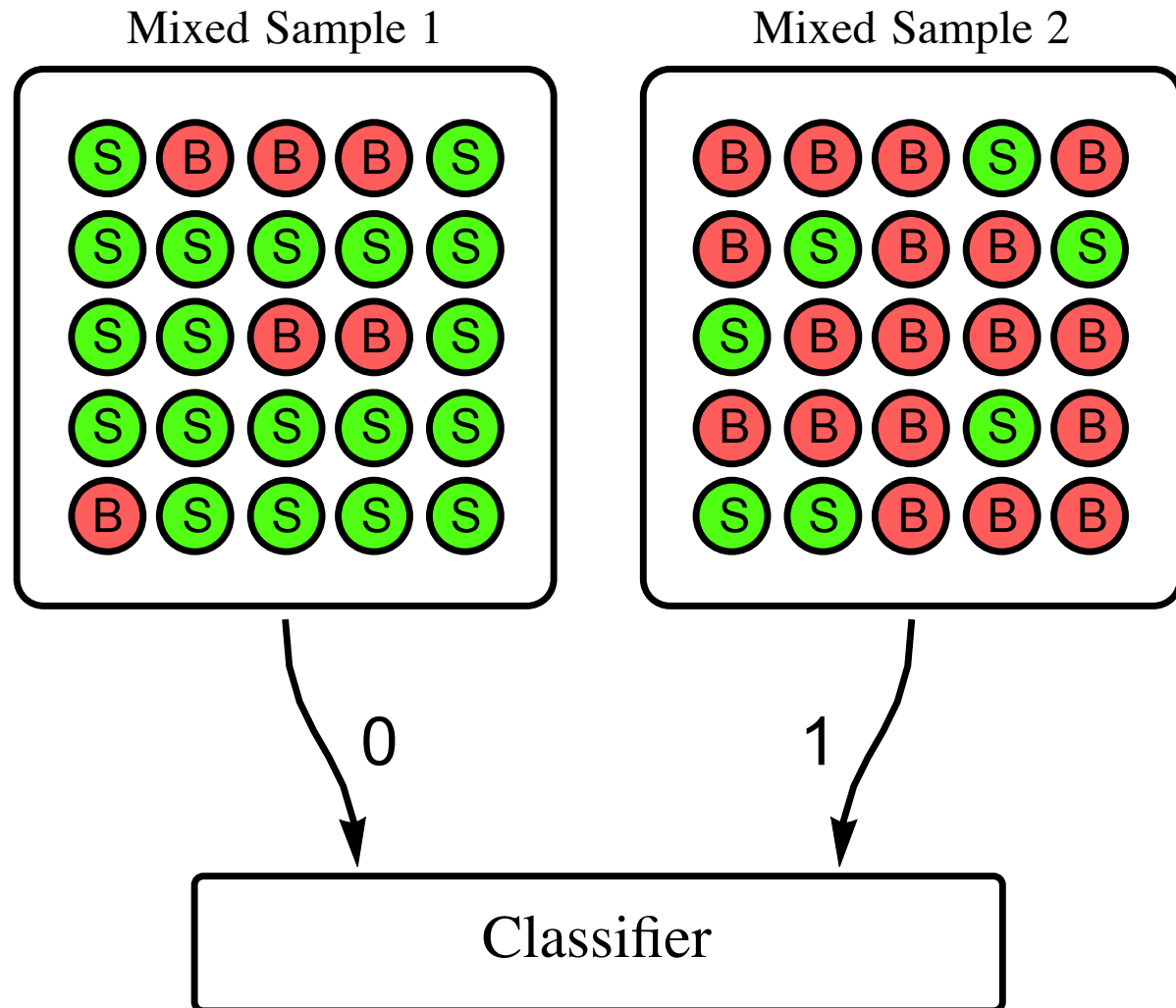


Classification Without Labels (CWoLa)

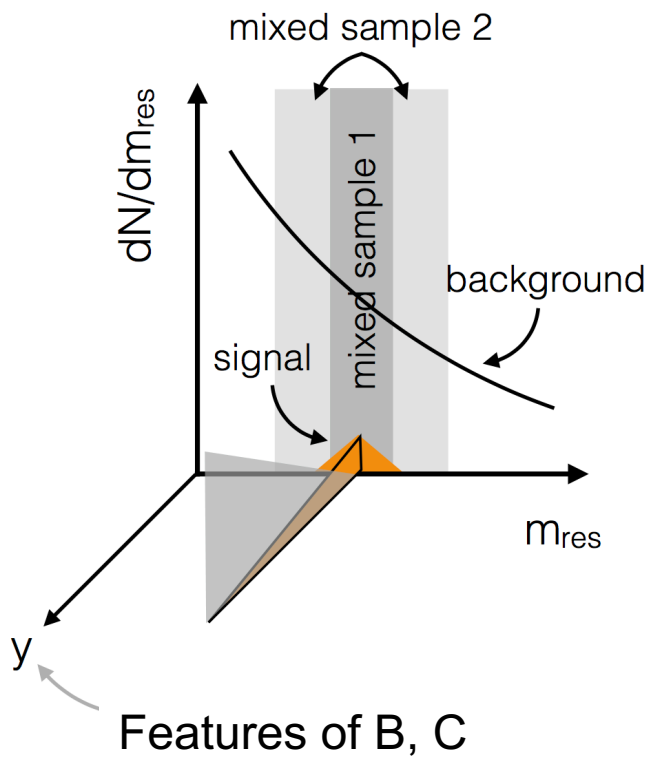
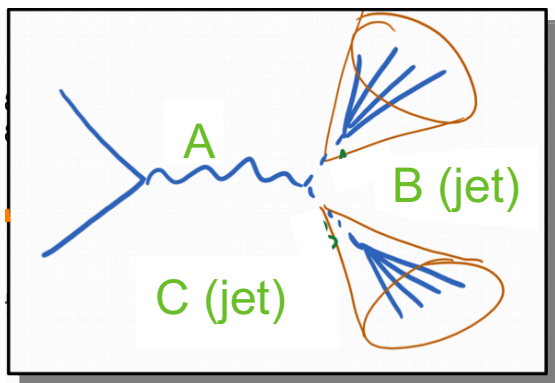


CWoLa

*Classification
Without Labels*



CWoLa Hunting



Sensitivity of CWoLa Hunting

$A \rightarrow BC$	e	μ	τ	q/g	b	t	γ	Z/W	H
e	[37, 38]	[39, 40]	[39]	$\emptyset$	$\emptyset$	$\emptyset$	[41]	[42]	$\emptyset$
μ		[37, 38]	[39]	$\emptyset$	$\emptyset$	$\emptyset$	[41]	[42]	$\emptyset$
τ			[45, 46]	$\emptyset$	[47]	$\emptyset$	$\emptyset$	$\emptyset$	$\emptyset$
q/g				[29, 30, 50, 51]	[52]	$\emptyset$	[53, 54]	[55]	$\emptyset$
b					[29, 52, 56]	[57]	[54]	[58]	[59]
t						[61]	$\emptyset$	[62]	[63]
γ							[65, 66]	[67–69]	[68, 70]
Z/W								[71]	[71]
H									[72, 73]

CWoLa sensitive to all!
(except maybe μ)

For citations
see [16]

$A \rightarrow BC$	$B = \text{SM}$								
	e	μ	τ	q/g	b	t	γ	Z/W	H
$C = \text{SM}$	e	Z'	$\bar{R}$	$\bar{R}$	LQ	LQ	LQ	L^*	L^*
	μ		Z'	$\bar{R}$	LQ	LQ	LQ	L^*	L^*
	τ			Z'	LQ	LQ	LQ	L^*	L^*
	q/g				Z'	W'	T'	Q^*	Q'
	b					Z'	W'	Q^*	B'
	t						Z'	Q^*	T'
	γ							H	Z_{KK}
	Z/W	Group I							$H^\pm/A$
	H								H

Sensitivity of CWoLa Hunting

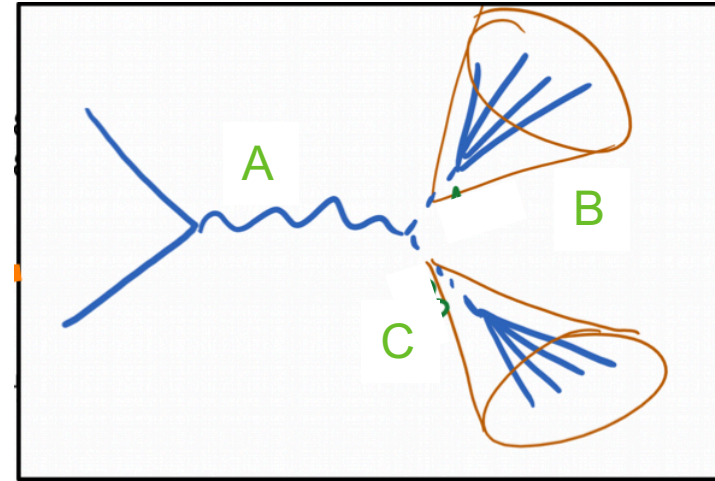
	e	μ	τ	q/g	b	t	γ	Z/W	H	BSM $\rightarrow$ SM ₁ $\times$ SM ₁				BSM $\rightarrow$ SM ₁ $\times$ SM ₂			BSM $\rightarrow$ complex			
										q/g	$\gamma/\pi^{0,s}$	b	...	tZ/H	bH	...	$\tau qq'$	eqq'	$\mu qq'$	...
e	[37, 38]	[39, 40]	[39]	$\emptyset$	$\emptyset$	$\emptyset$	[41]	[42]	$\emptyset$	$\emptyset$	$\emptyset$	$\emptyset$		$\emptyset$	$\emptyset$	$\emptyset$	$\emptyset$	[43, 44]	$\emptyset$	
μ		[37, 38]	[39]	$\emptyset$	$\emptyset$	$\emptyset$	[41]	[42]	$\emptyset$	$\emptyset$	$\emptyset$	$\emptyset$		$\emptyset$	$\emptyset$	$\emptyset$	$\emptyset$	$\emptyset$	[43, 44]	
τ			[45, 46]	$\emptyset$	[47]	$\emptyset$	$\emptyset$	$\emptyset$	$\emptyset$	$\emptyset$	$\emptyset$	$\emptyset$		$\emptyset$	$\emptyset$	$\emptyset$	[48, 49]	$\emptyset$	$\emptyset$	
q/g				[29, 30, 50, 51]	[52]	$\emptyset$	[53, 54]	[55]	$\emptyset$	$\emptyset$	$\emptyset$	$\emptyset$		$\emptyset$	$\emptyset$	$\emptyset$	$\emptyset$	$\emptyset$	$\emptyset$	
b					[29, 52, 56]	[57]	[54]	[58]	[59]	$\emptyset$	$\emptyset$	$\emptyset$		[60]	$\emptyset$	$\emptyset$	$\emptyset$	$\emptyset$	$\emptyset$	
t						[61]	$\emptyset$	[62]	[63]	$\emptyset$	$\emptyset$	$\emptyset$		[64]	[60]	$\emptyset$	$\emptyset$	$\emptyset$	$\emptyset$	
γ							[65, 66]	[67–69]	[68, 70]	$\emptyset$	$\emptyset$	$\emptyset$		$\emptyset$	$\emptyset$	$\emptyset$	$\emptyset$	$\emptyset$	$\emptyset$	
Z/W								[71]	[71]	$\emptyset$	$\emptyset$	$\emptyset$		$\emptyset$	$\emptyset$	$\emptyset$	$\emptyset$	$\emptyset$	$\emptyset$	
H								[72, 73]		[74]	$\emptyset$	$\emptyset$		$\emptyset$	$\emptyset$	$\emptyset$	$\emptyset$	$\emptyset$	$\emptyset$	
BSM $\rightarrow$ SM ₁ $\times$ SM ₁	q/g									$\emptyset$	$\emptyset$	$\emptyset$		$\emptyset$	$\emptyset$	$\emptyset$	$\emptyset$	$\emptyset$	$\emptyset$	
	$\gamma/\pi^{0,s}$										[75]	$\emptyset$		$\emptyset$	$\emptyset$	$\emptyset$	$\emptyset$	$\emptyset$	$\emptyset$	
	b											[76, 77]		$\emptyset$	$\emptyset$	$\emptyset$	$\emptyset$	$\emptyset$	$\emptyset$	
	...													$\emptyset$	$\emptyset$	$\emptyset$	$\emptyset$	$\emptyset$	$\emptyset$	
BSM $\rightarrow$ SM ₁ $\times$ SM ₂	tZ/H																			
	bH																			
	...																			
	...																			
BSM $\rightarrow$ complex	$\tau qq'$																			
	eqq'																			
	$\mu qq'$																			
	...																			
...																				

CWoLa sensitive to all!

For citations
see [\[16\]](#)

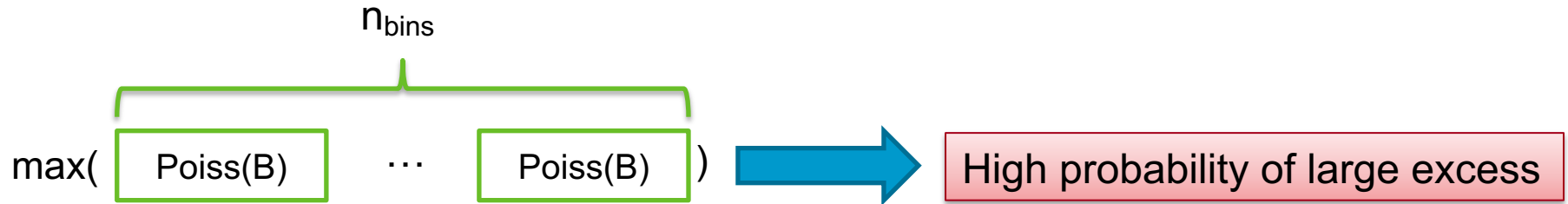
Sensitivity of CWoLa Hunting

- Mostly uncovered: $A \rightarrow BC(JJ)$ (B and/or C BSM)
- BSM Masses, SM properties
 - Impact parameter, number of prongs, etc.
 - Exotic signatures and combinations
- **CWoLa sensitive to all!** (depending on features used)
- This analysis:
 - Using reduced feature set ($Y_i = \{m_{JJ}\}$), provide limits on $A \rightarrow BC$ generic narrow resonance
 - Smart way of doing 3D bump hunt
 - Currently no dedicated 3D bump hunt in ATLAS $\rightarrow$ new result



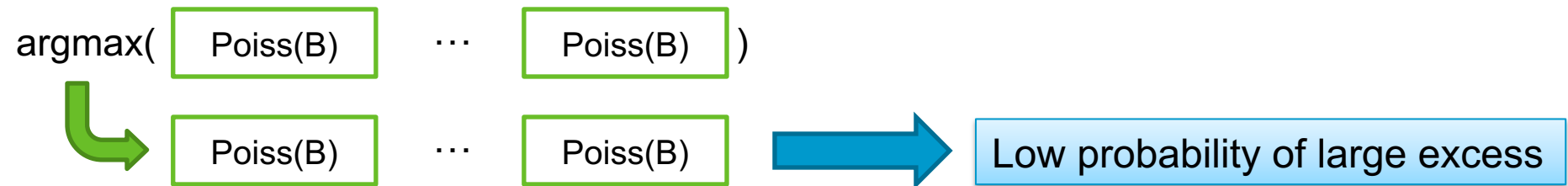
Trials Factors

- Trials factor for discovery potential with large numbers of bins

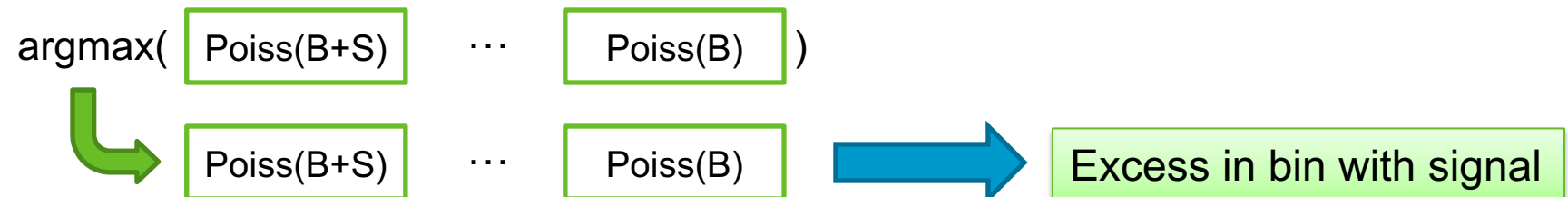


- In 3D m_A, m_B, m_C space, $n_{\text{bins}} \gg 1$
- CWoLa hunting (for fixed m_A):

Background-only



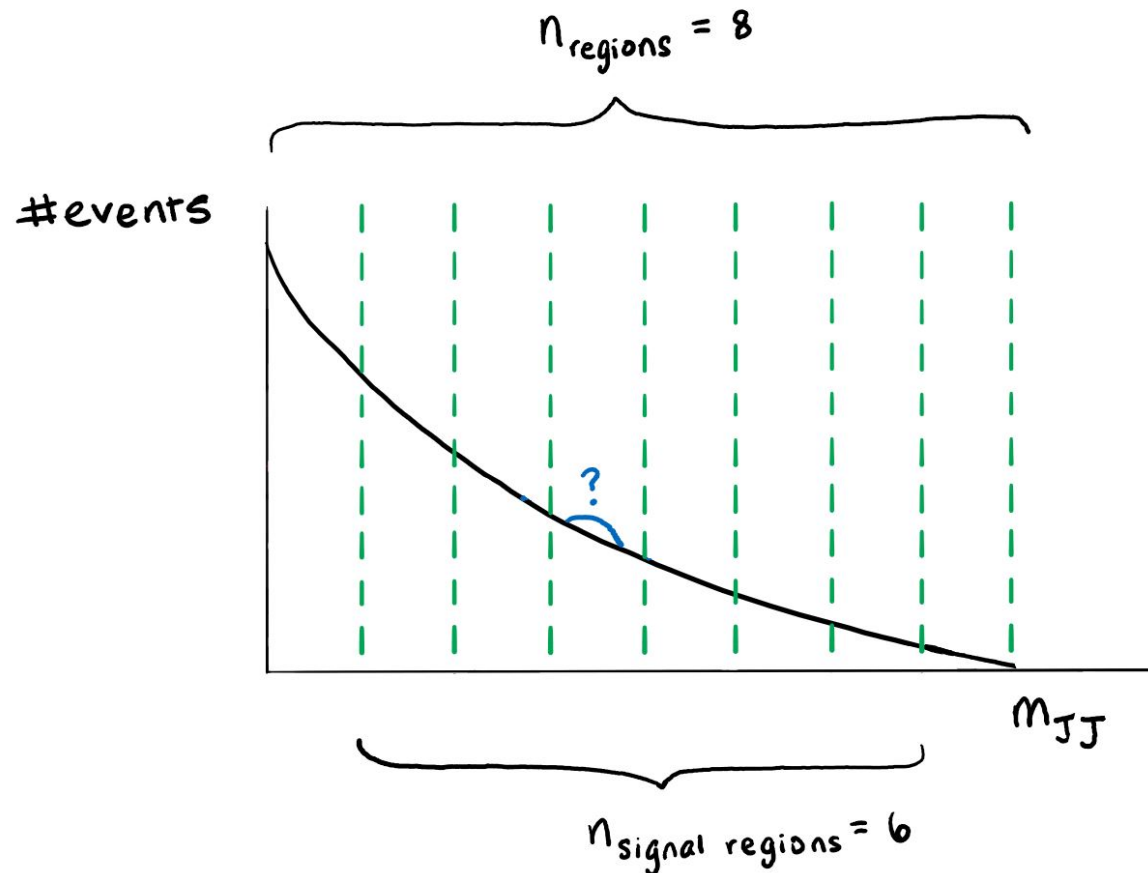
True signal present



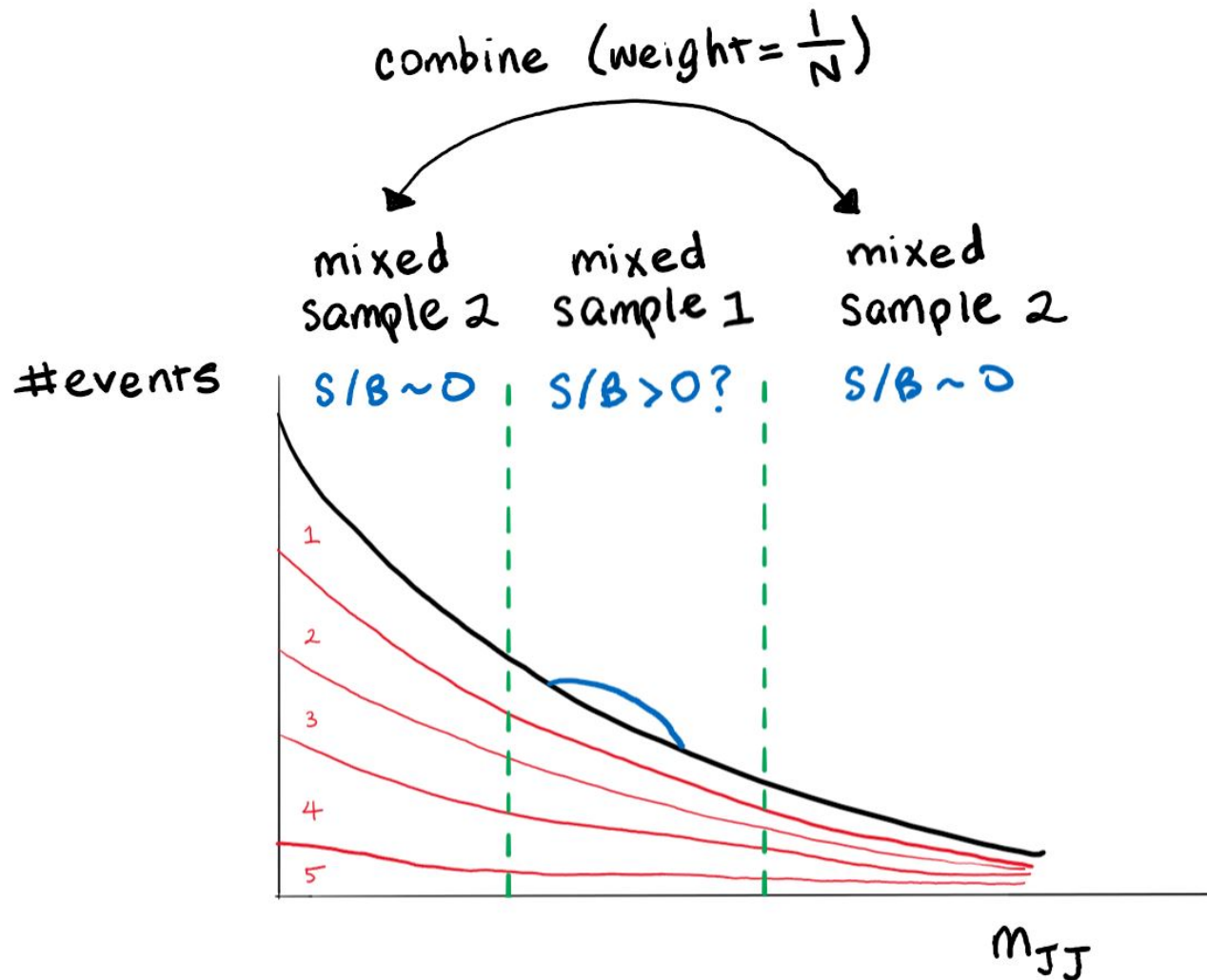
CWoLa Hunting in ATLAS

Setup - m_{JJ}

- Bump hunt in dijet invariant mass spectrum
- Signal regions: Bins set by dijet mass resolution (20%)
- Bins used for training signal vs sideband

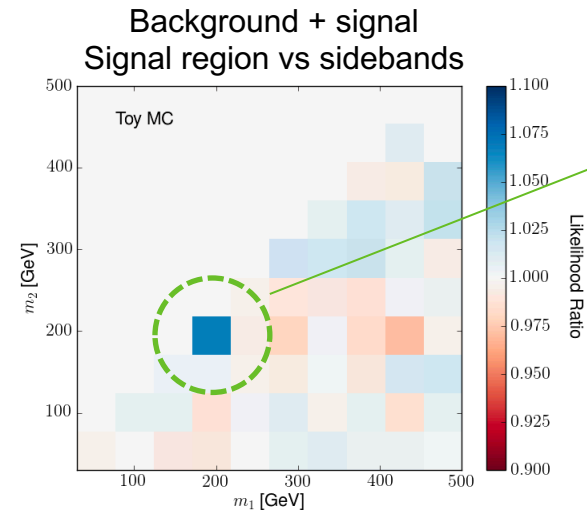
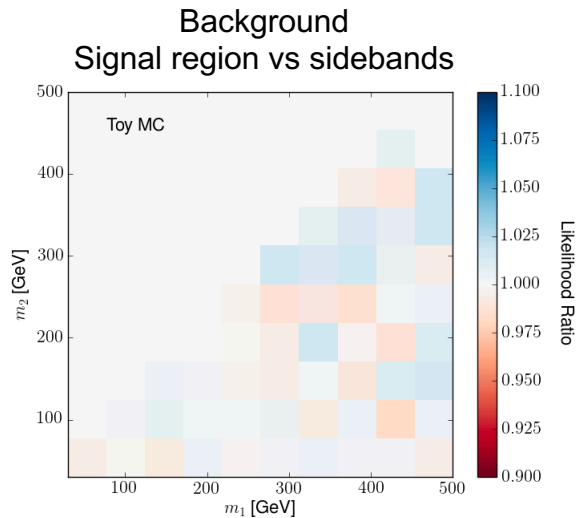
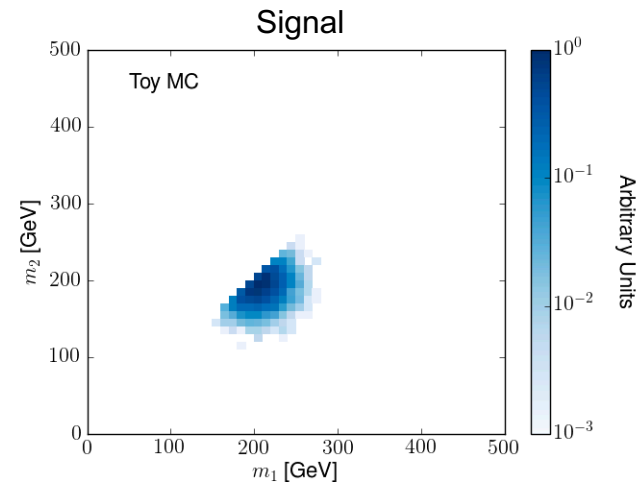
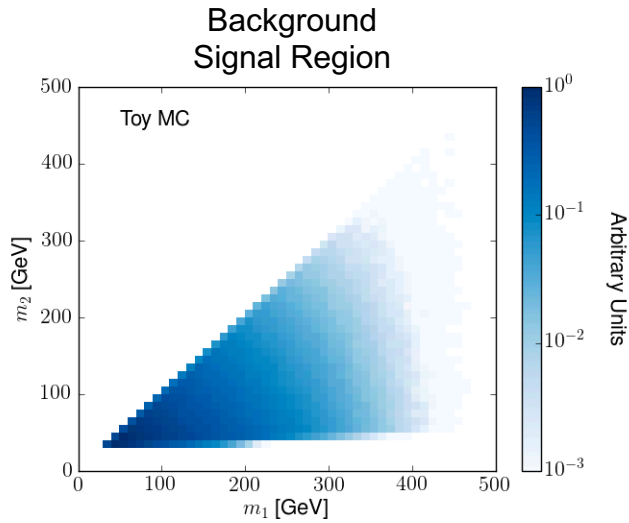


Learning Setup



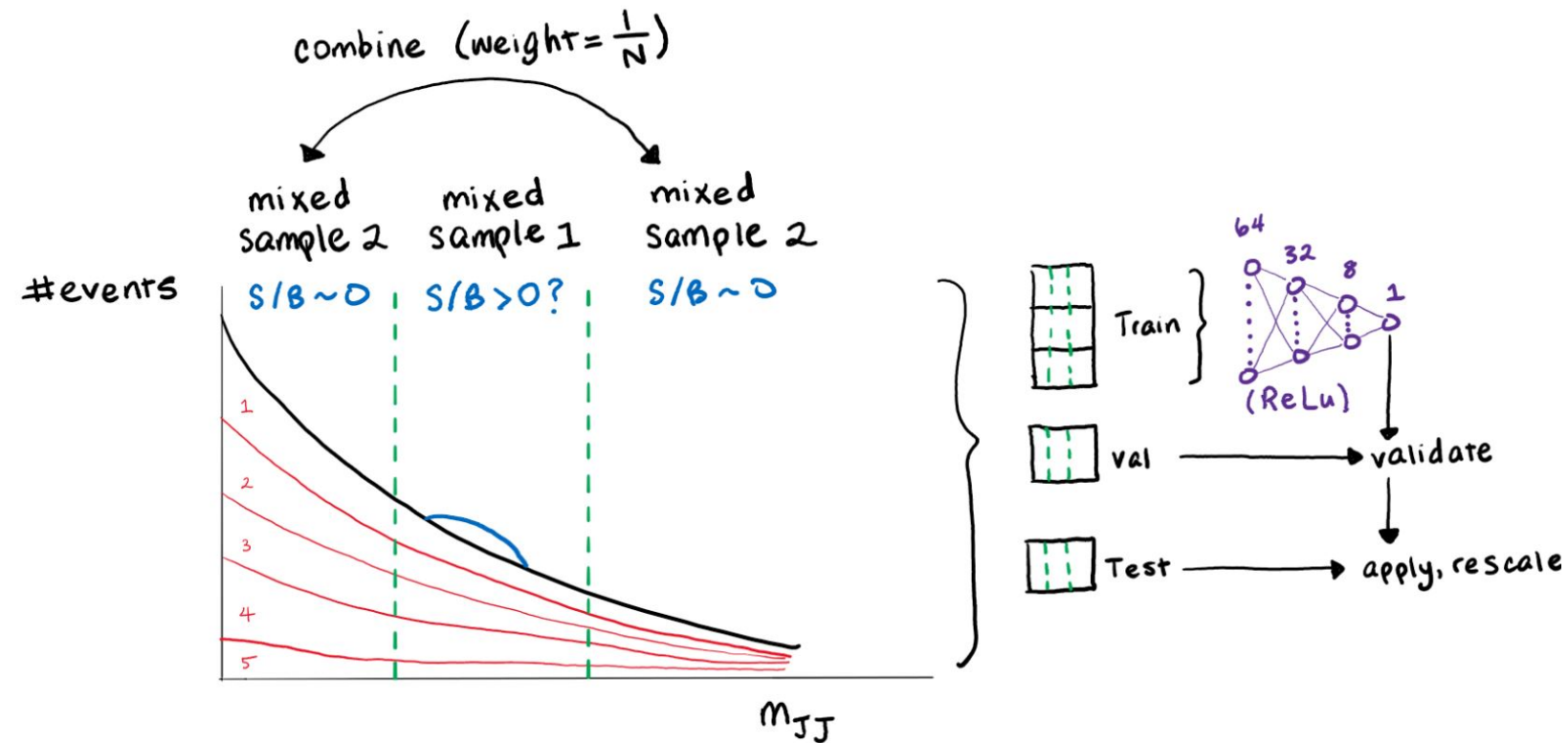
Setup - Features

- Features are $X = (m_1, m_2)$ (decorrelated)

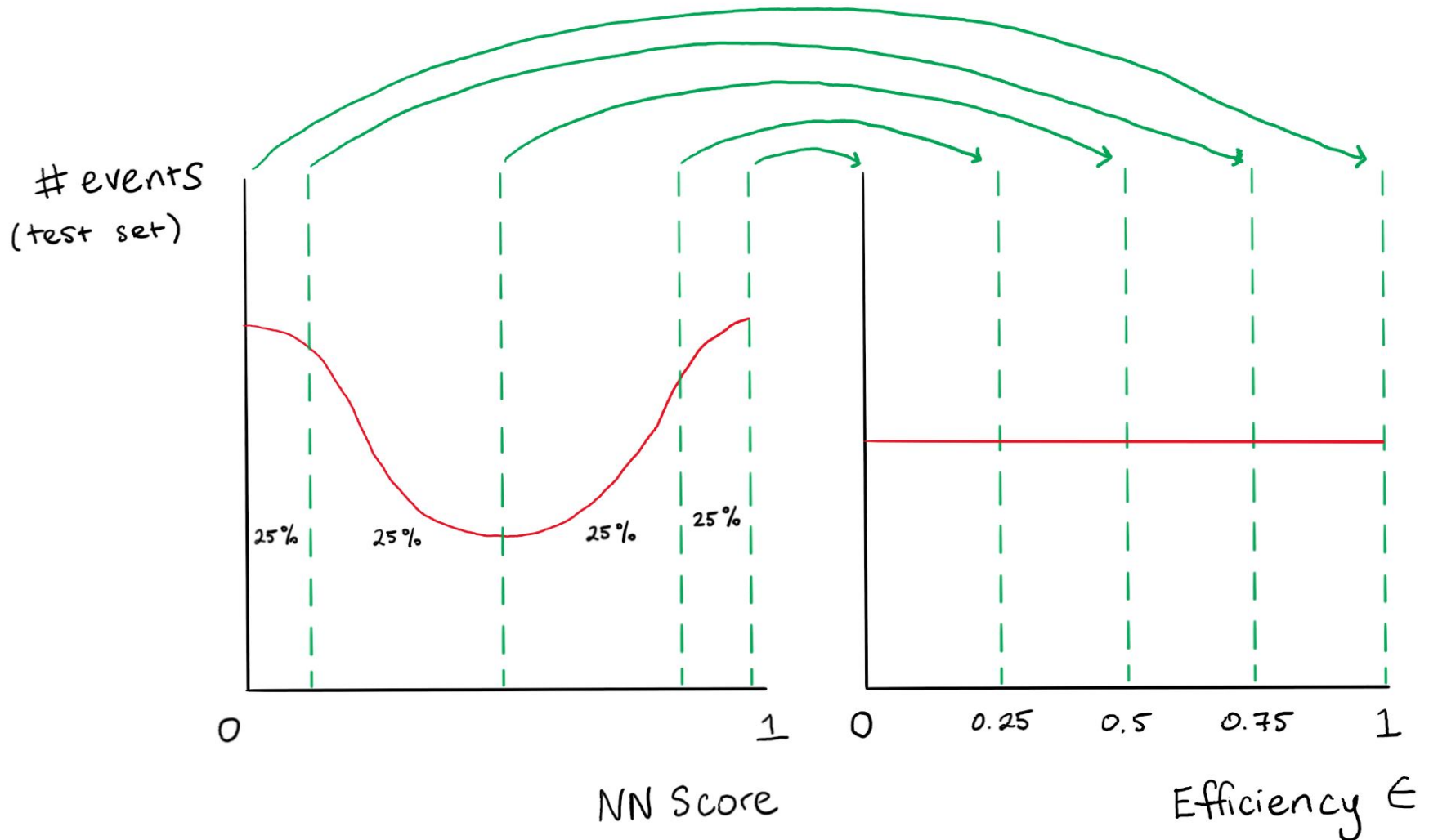


Excess! NN
should learn
to target

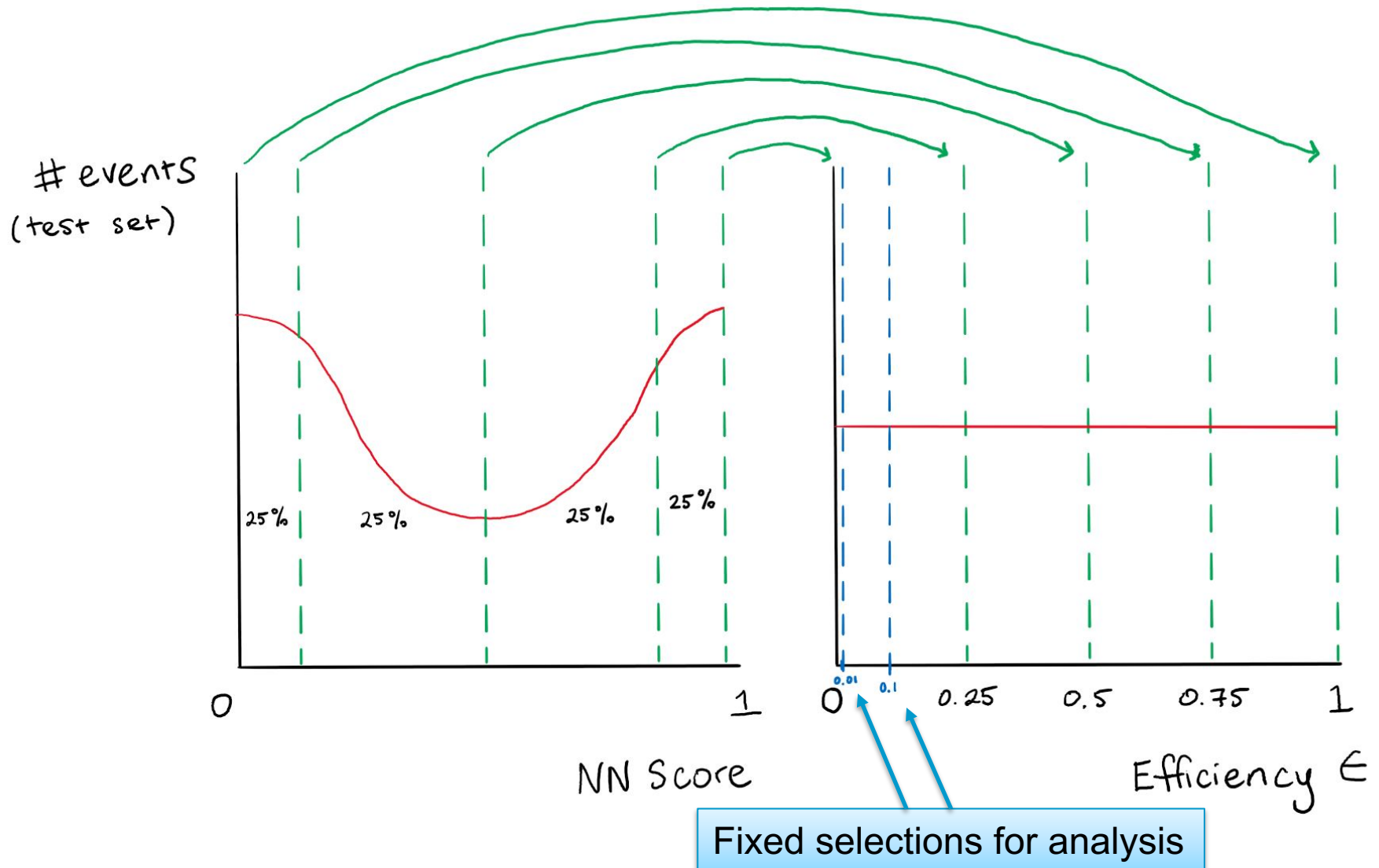
Learning Setup



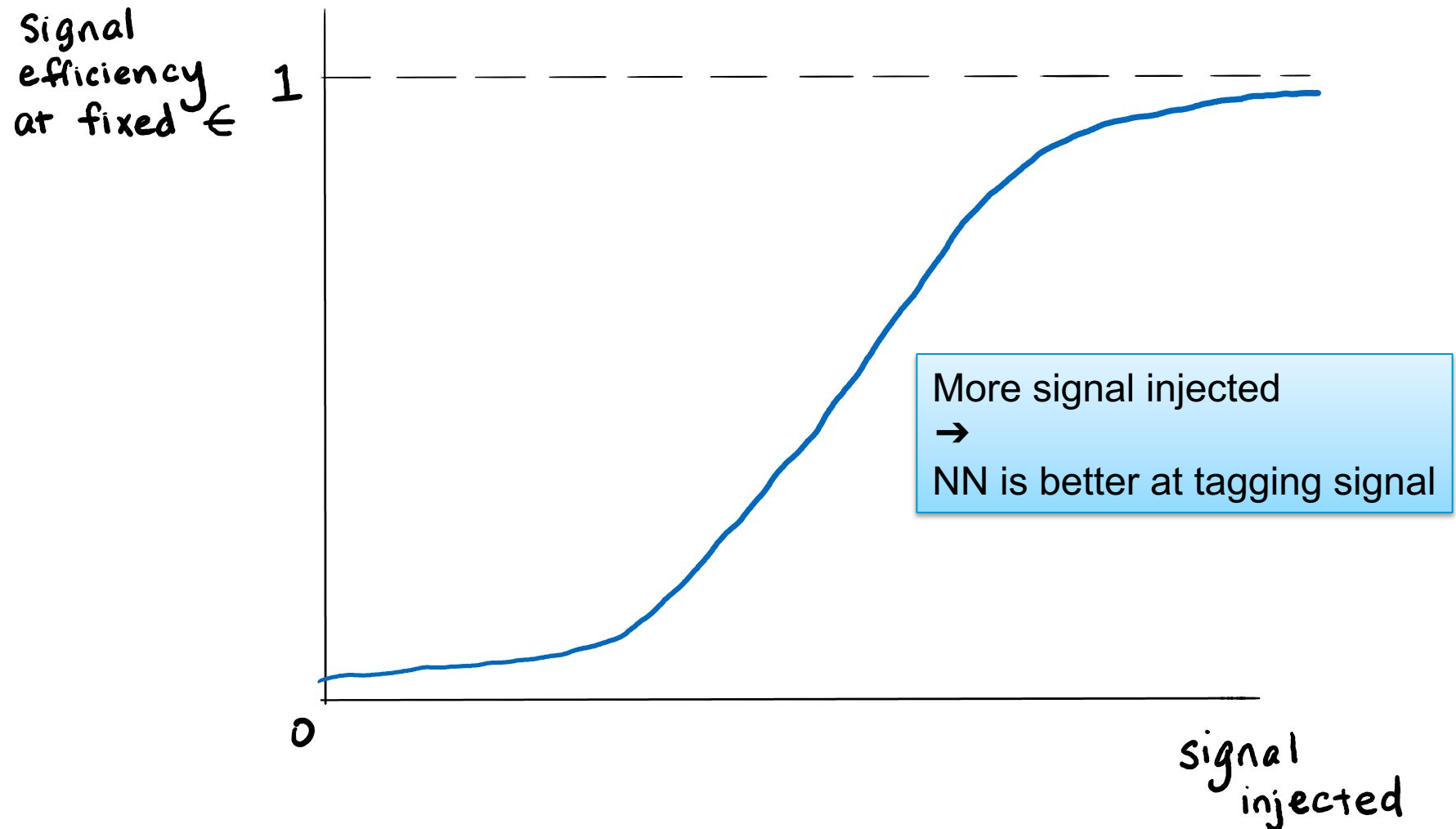
Learning Setup – NN Rescaling



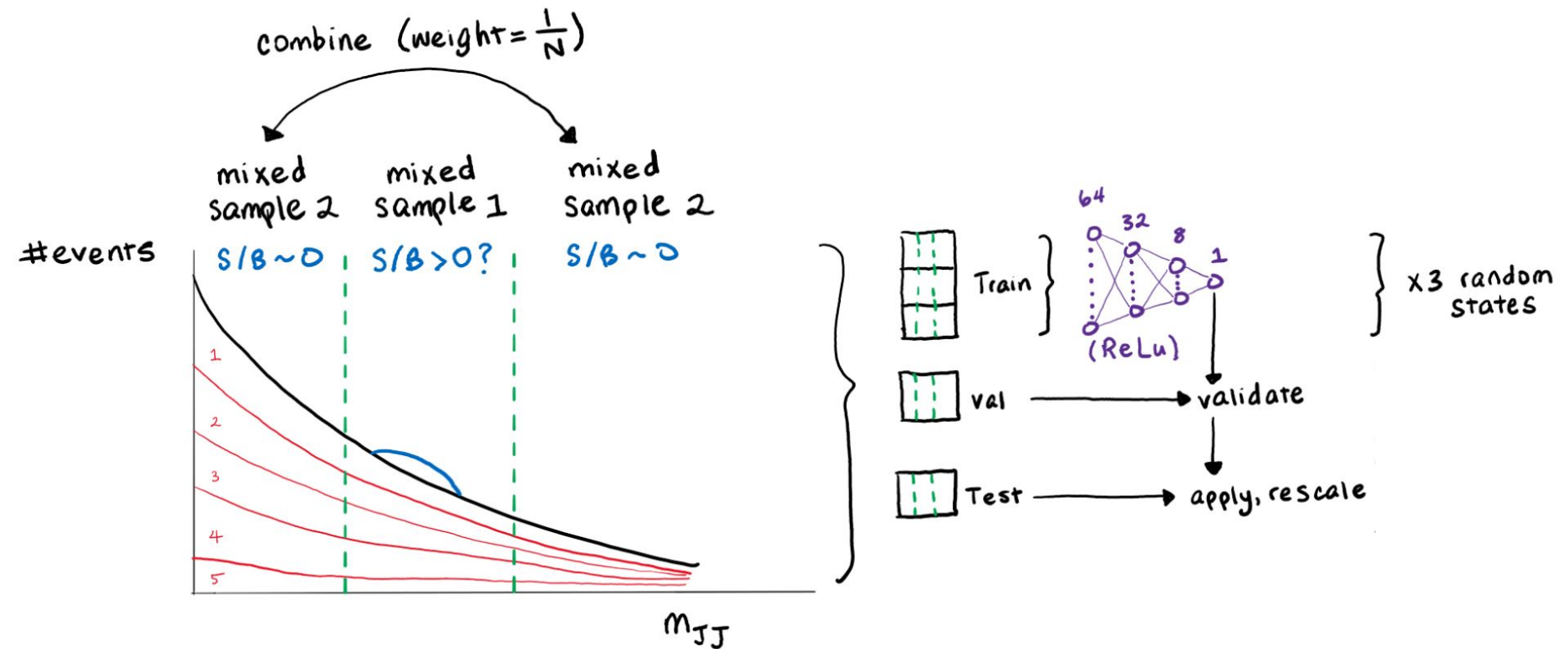
Learning Setup – NN Rescaling



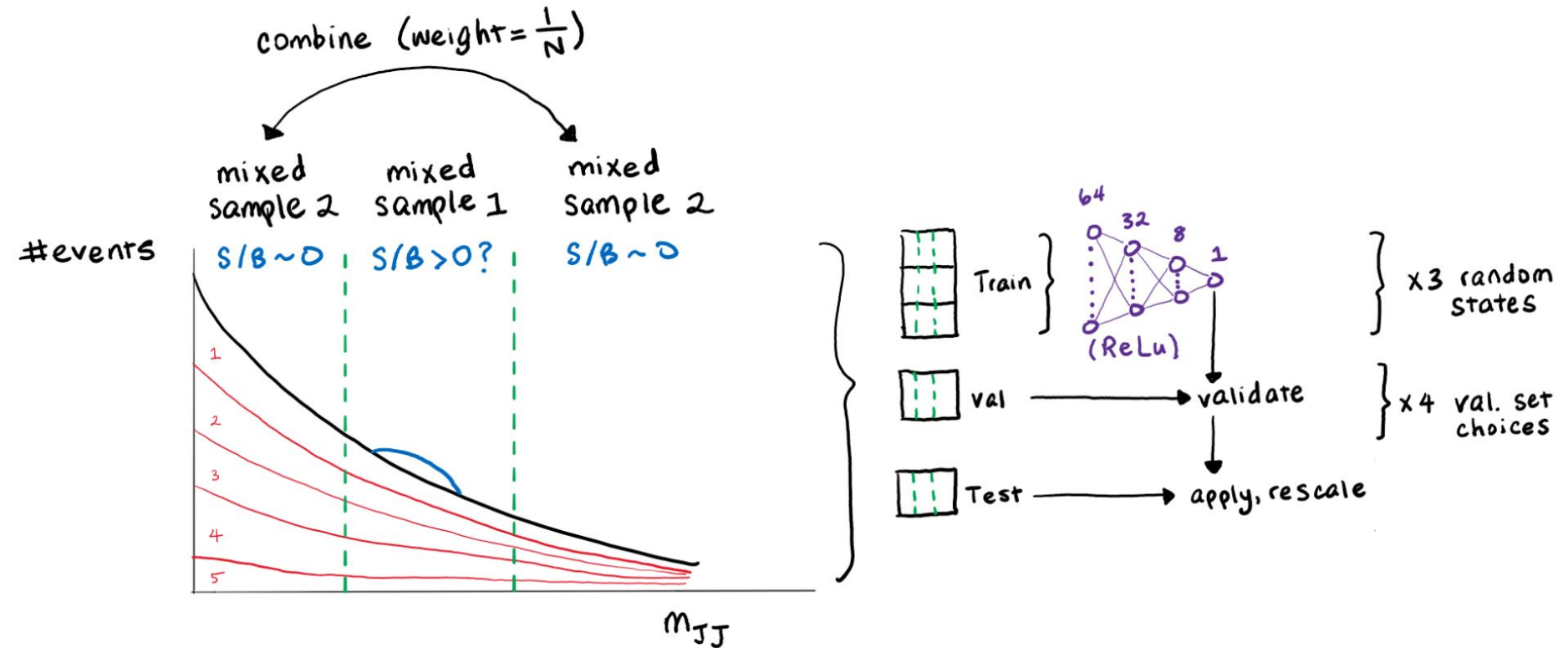
NN Output



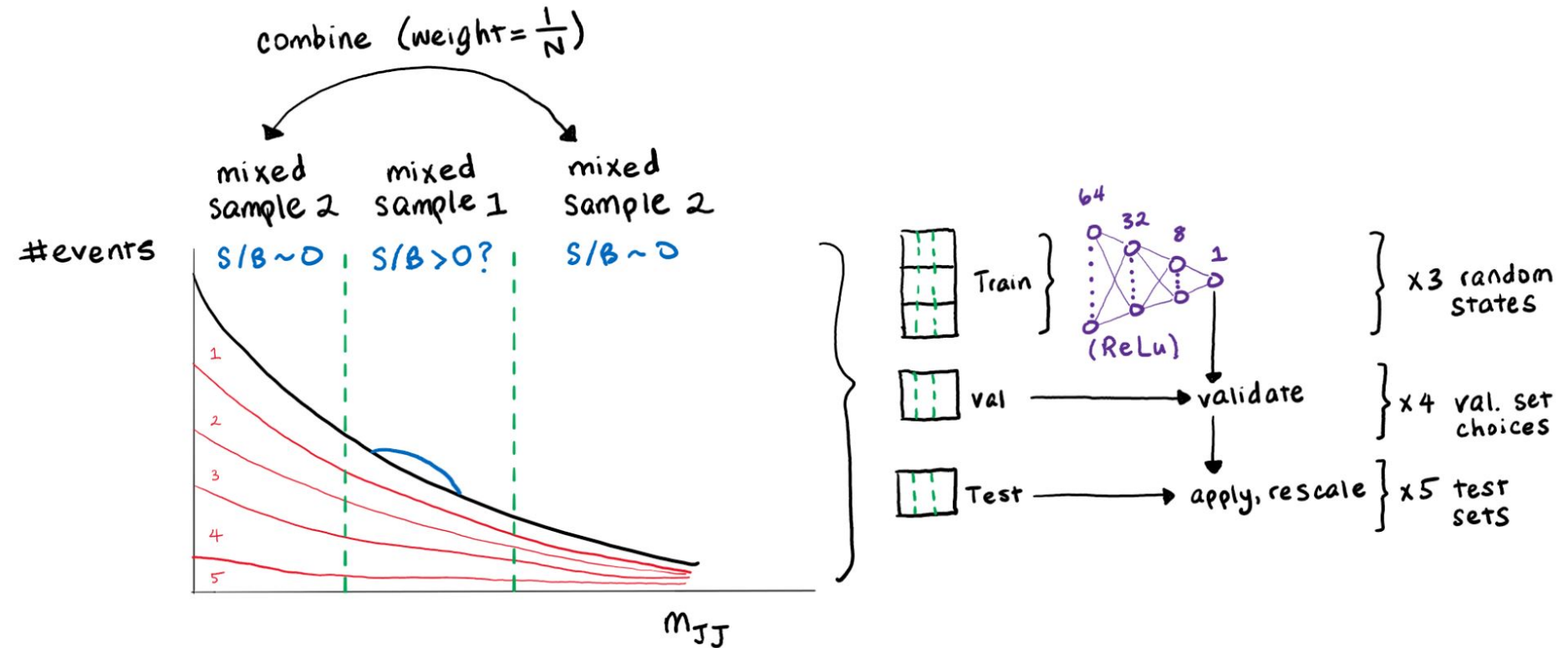
Learning Setup



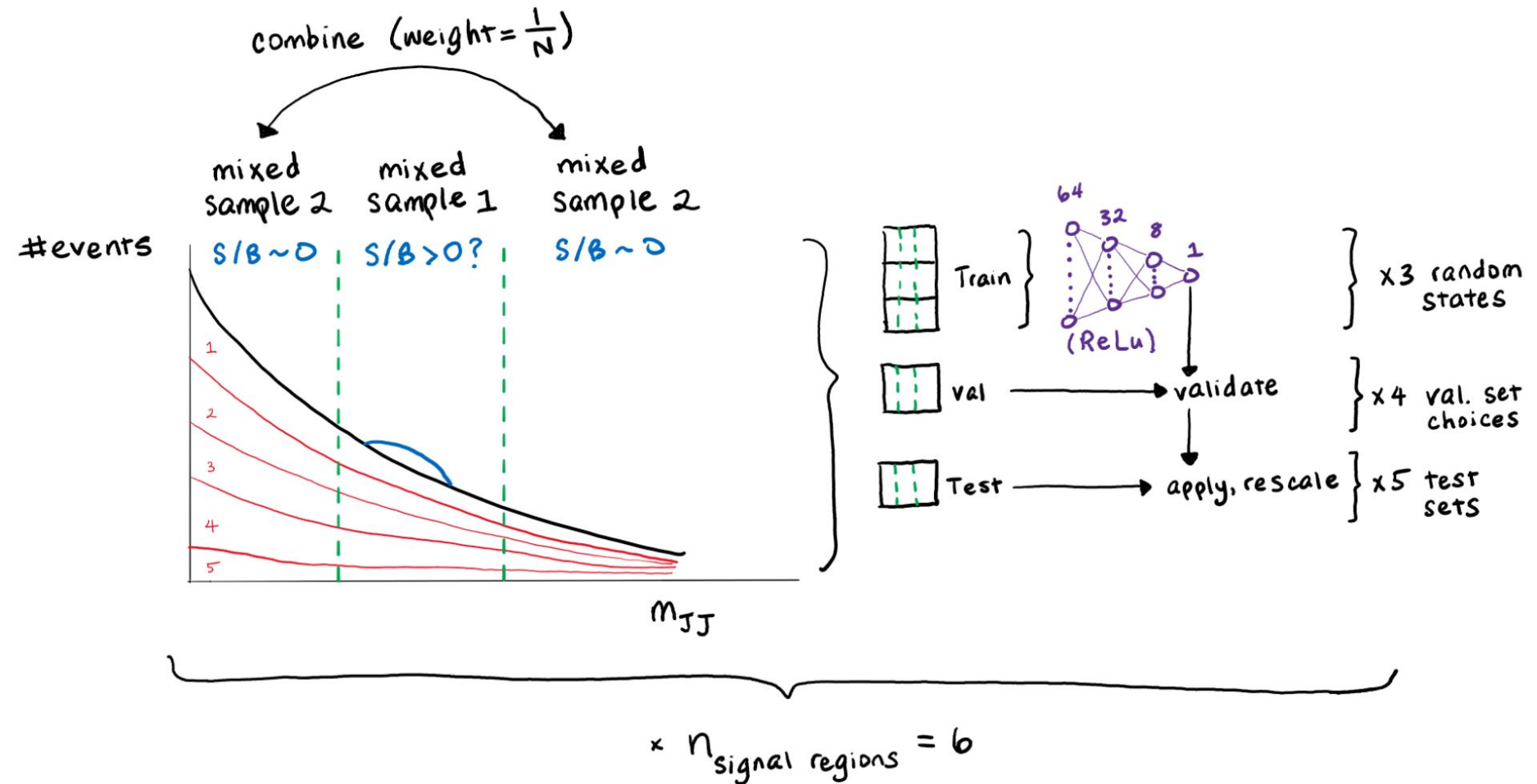
Learning Setup



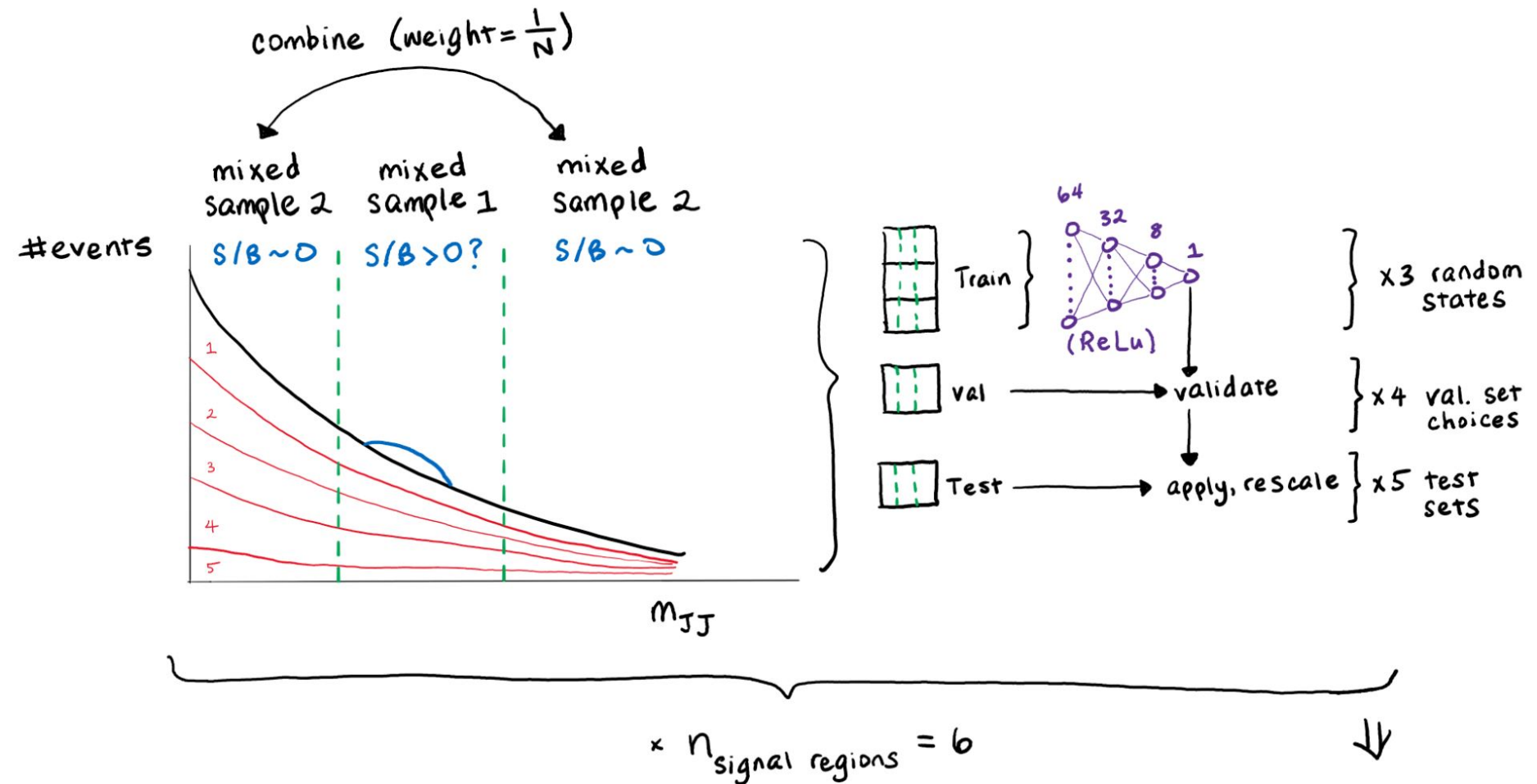
Learning Setup



Learning Setup



Learning Setup

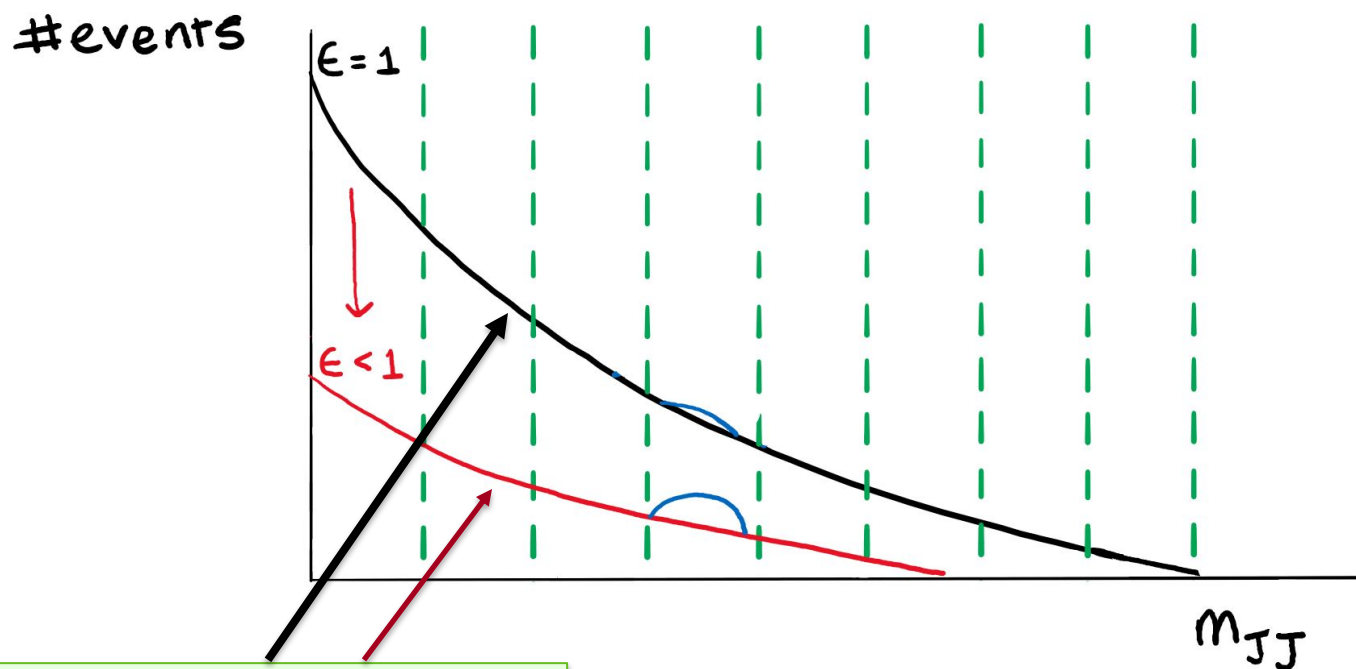


With injected signal: $\times (n_{\mu} \approx 10) \times (5 \text{ random samplings}) \Rightarrow \sim 18000 \text{ NNs!}$

Statistical Interpretation – Model-independent Results

SLAC

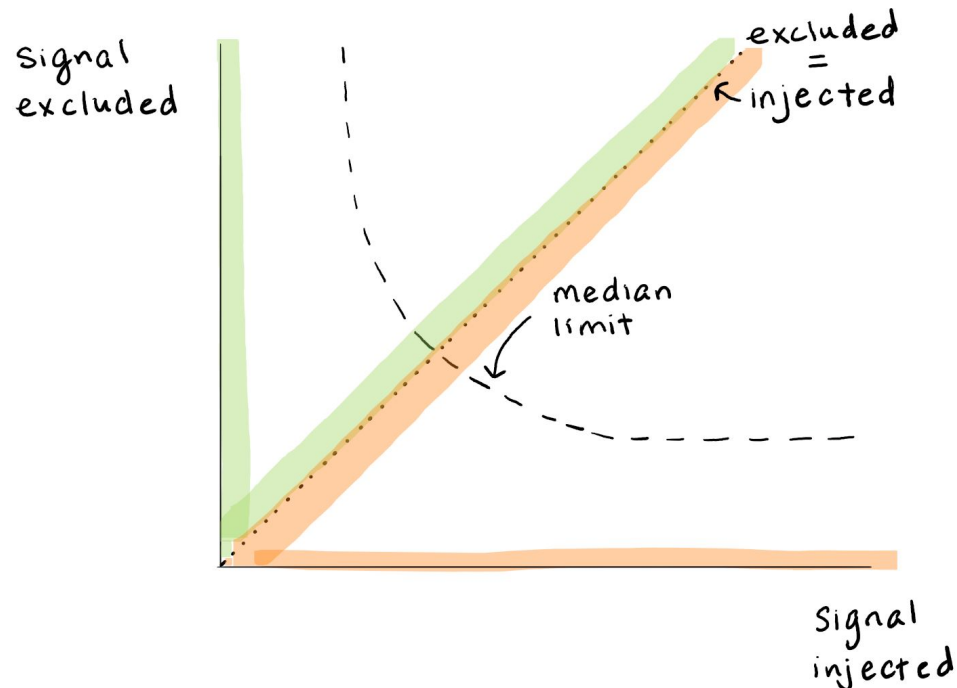
- p-value in signal regions for each m_{JJ}



Smooth parameterized fit function

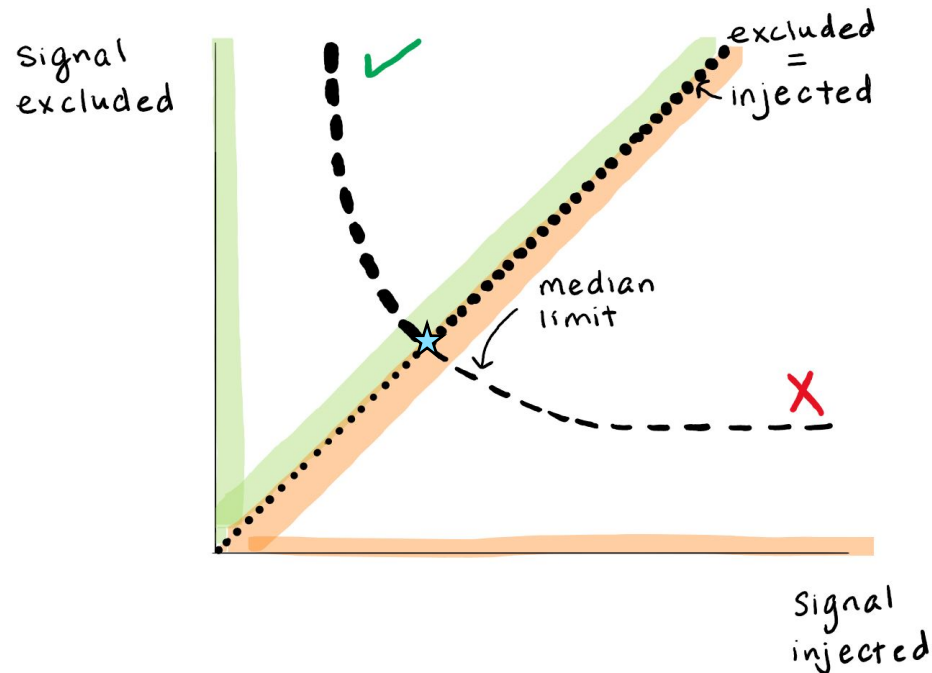
Statistical Interpretation – Model-dependent Results

- With injected signal strength μ :
 - Apply NN selection ϵ (depends on μ)
 - Let μ be POI and set 95% CL limits in the usual way
 - Limit depends on injected signal strength!



Statistical Interpretation – Model-dependent Results

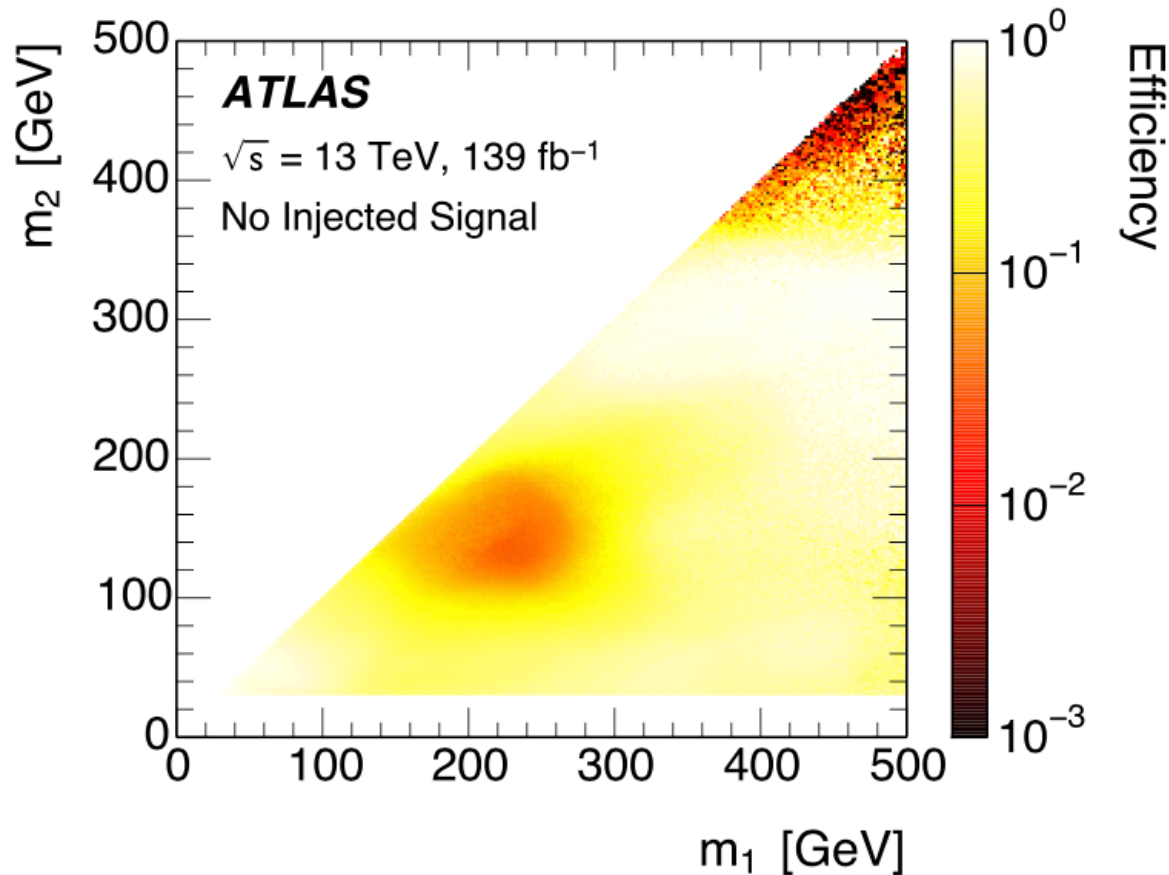
- If excluded $\geq$ injected, then trust limit
 - Network should be better if truly more signal than injected
- If excluded $<$ injected, then don't trust limit
 - Network could be worse if truly less signal than injected
- -> Best limit is $\min_{\mu}(\max(\text{CL}(\mu), \mu))$



Results

NN Output

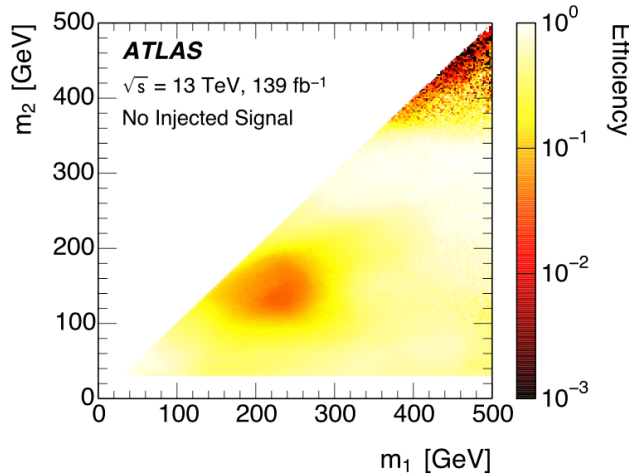
- NN output with no signal injected
- N.B.: Efficiency=0 (high NN score) somewhere by definition



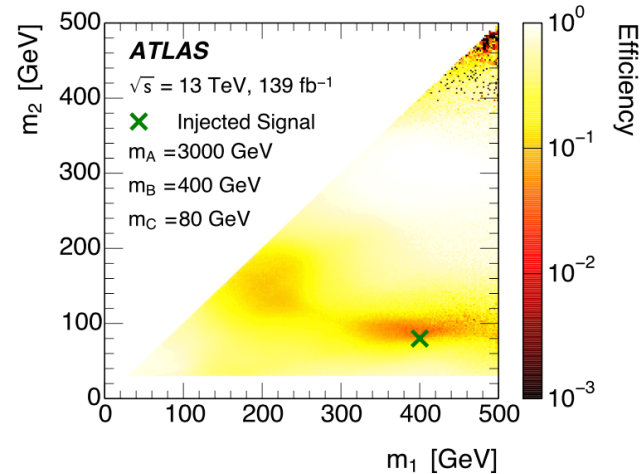
$m_{JJ} = [2.74, 3.28]$
TeV

NN Output

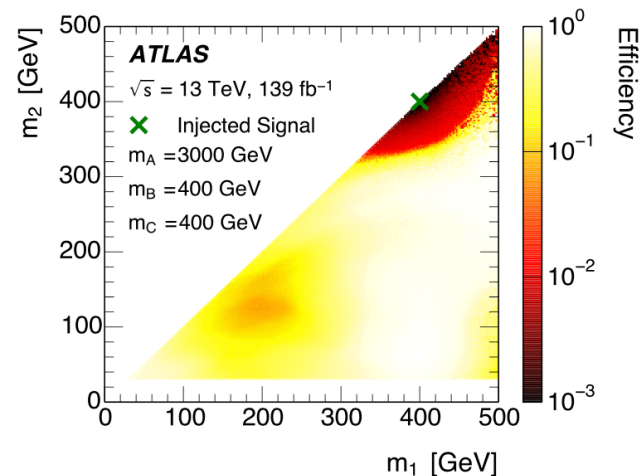
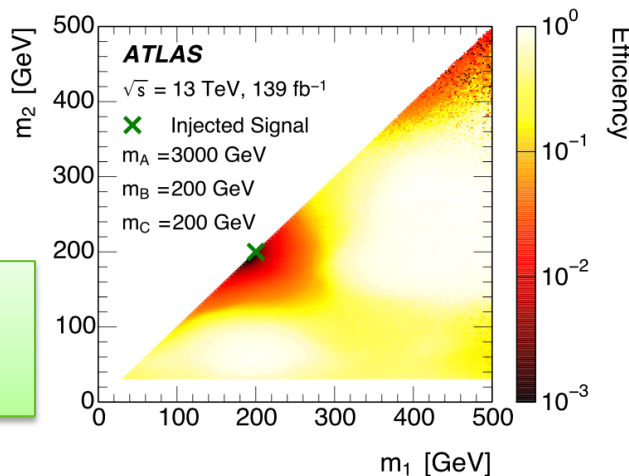
- NN output with no signal injected and with signal injected
- N.B.: Efficiency=0 (high NN score) somewhere by definition



(a)



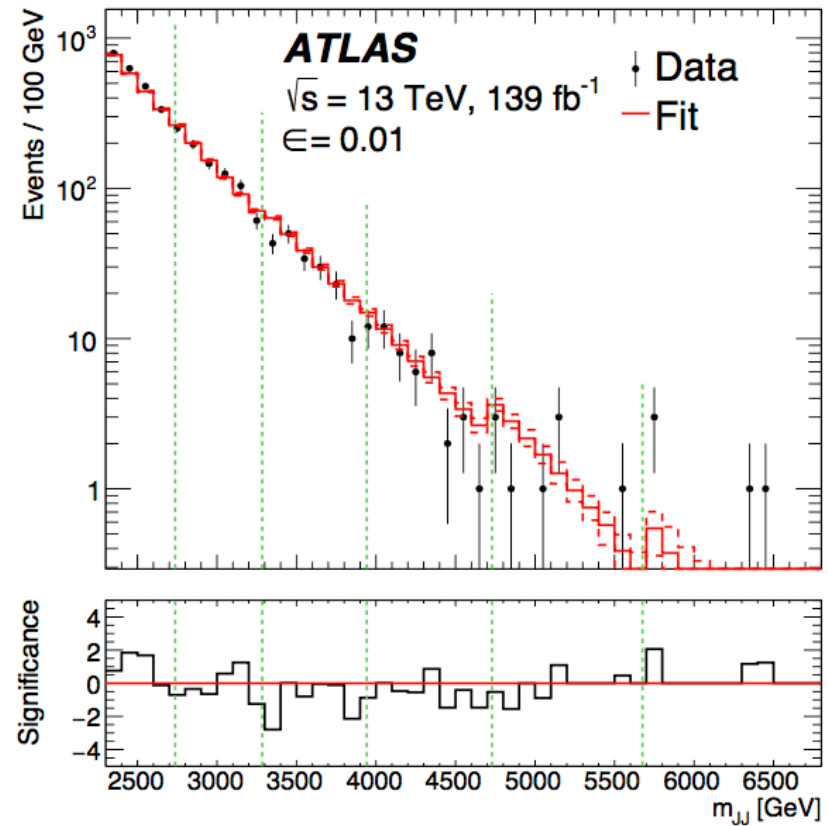
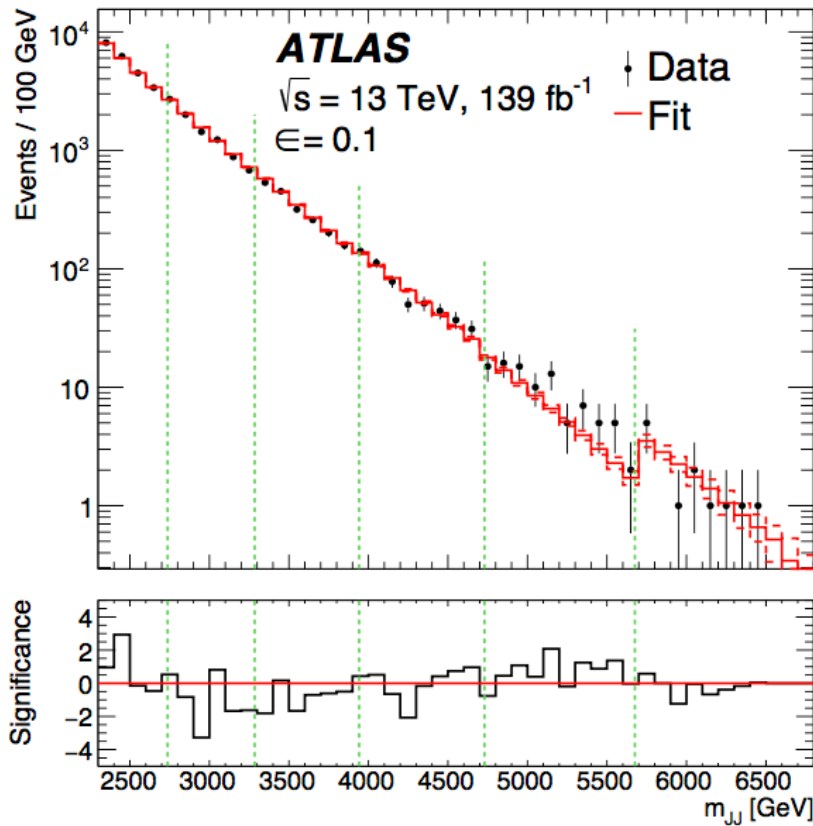
(b)



For all:
 $m_{JJ} = [2.74, 3.28]$
 TeV

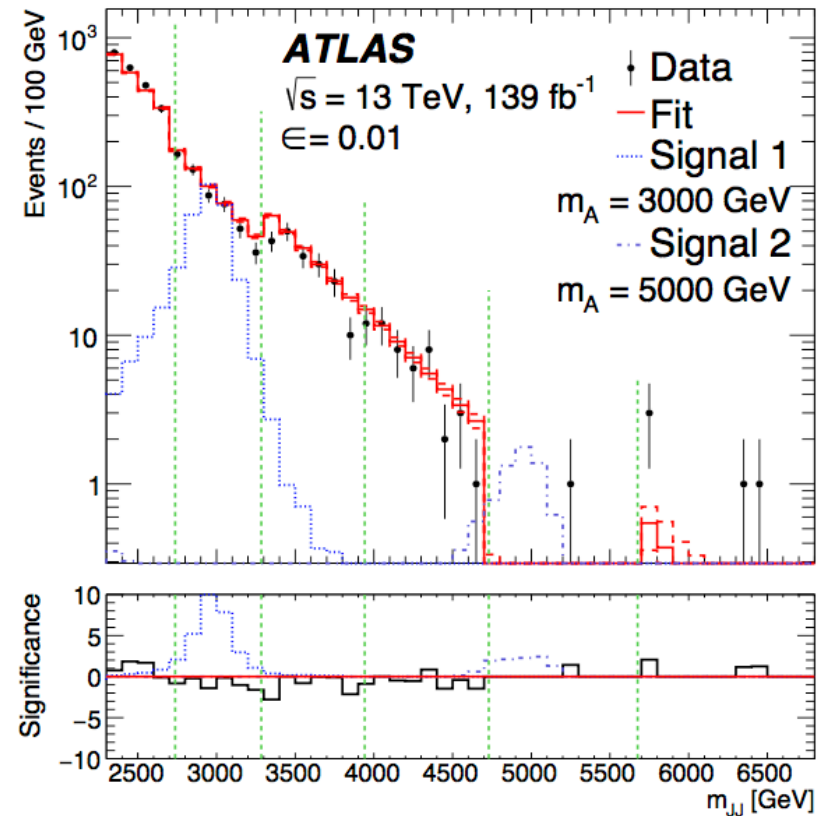
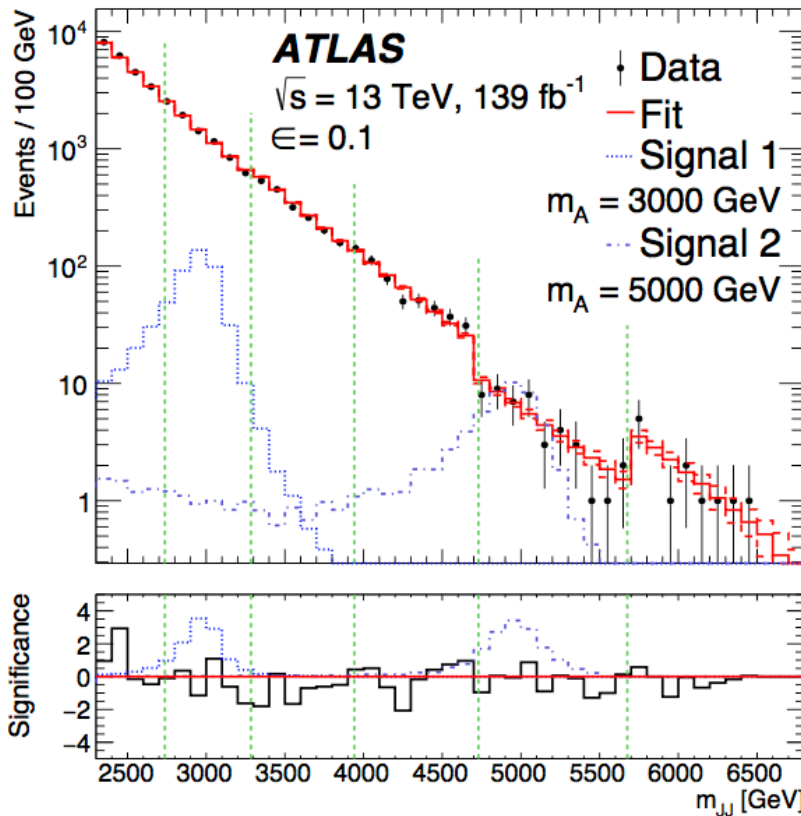
Fits in Signal Regions

- Significances of data in signal regions with respect to background-only fit
- Signal regions stitched together -> can be discontinuous



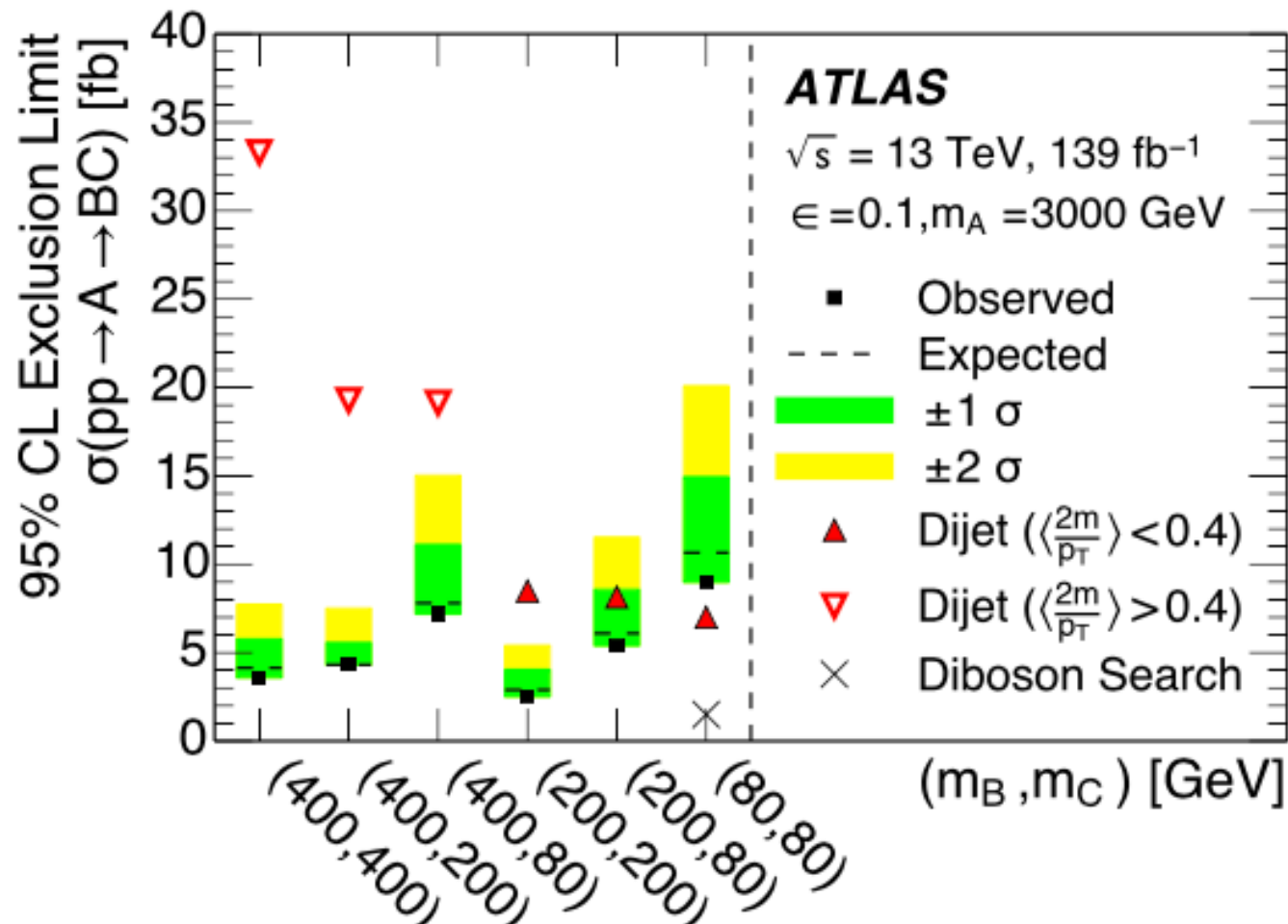
Fits in Signal Regions

- Significances of data in signal regions with respect to background-only fit
- With injected signal to demonstrate power of method



Limits

- Limits (after injecting signal)

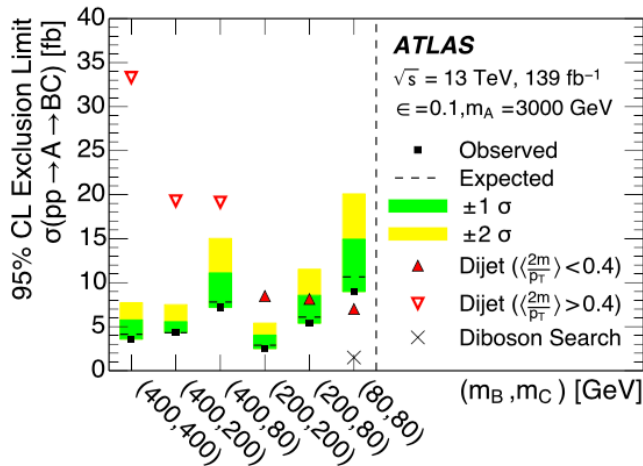


Limits for some signals >2x better!

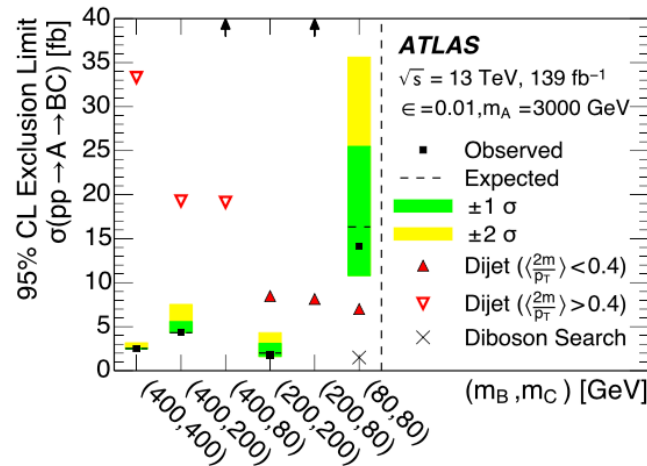
Dijet search: small-R jets

Diboson search: more sensitive to one signal (uses ntrack, D2)

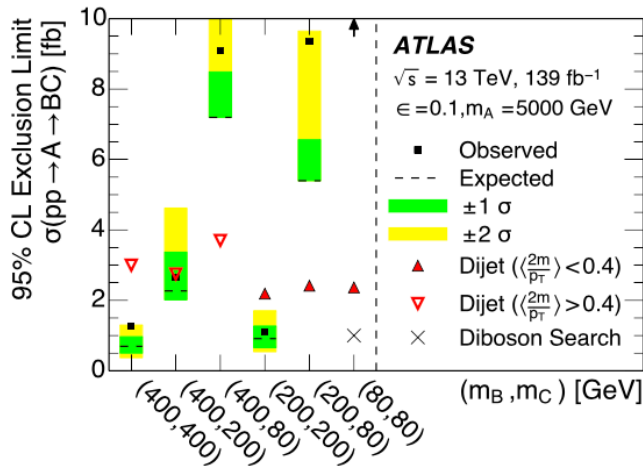
- Limits (after injecting signal)



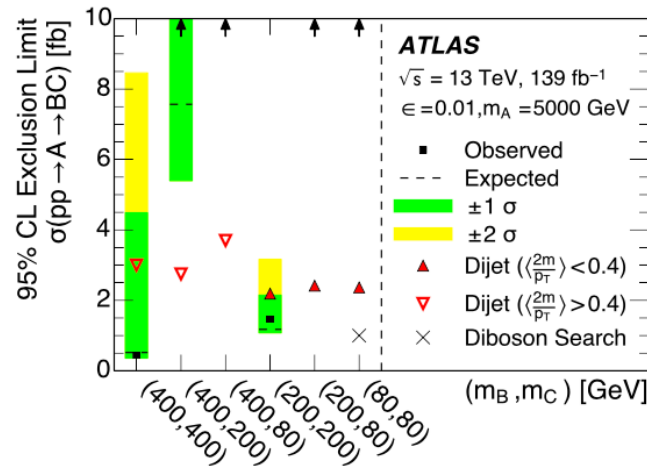
(a)



(b)



(c)

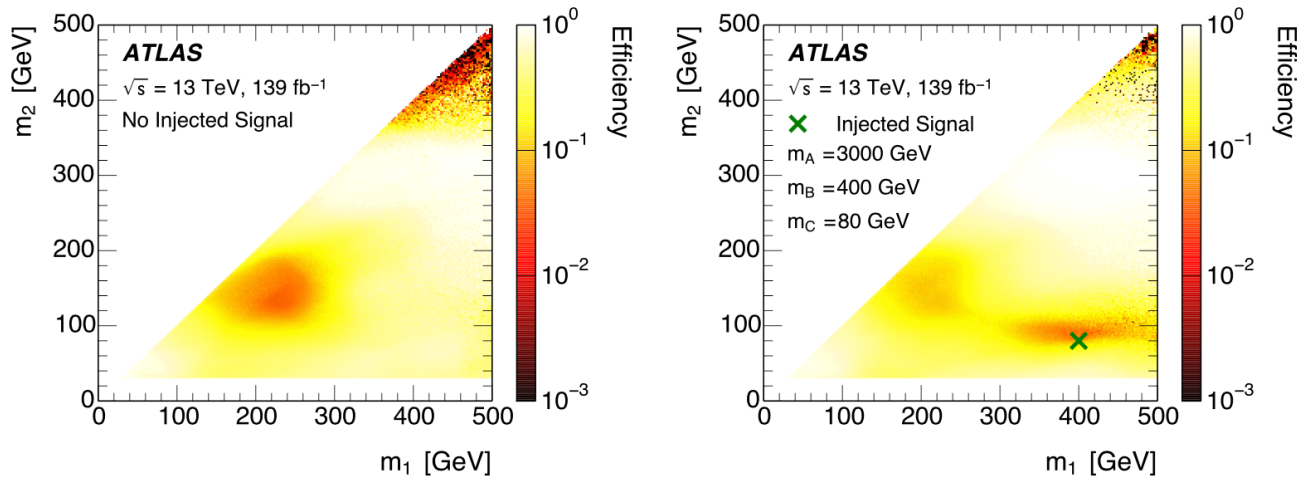


(d)

Limits for
some signals
>2x better!

Some signals
not found by
NN

- Tagging based on B, C jet mass



- Add other jet features
 - Jet substructure
- Other anomaly-detection algorithms

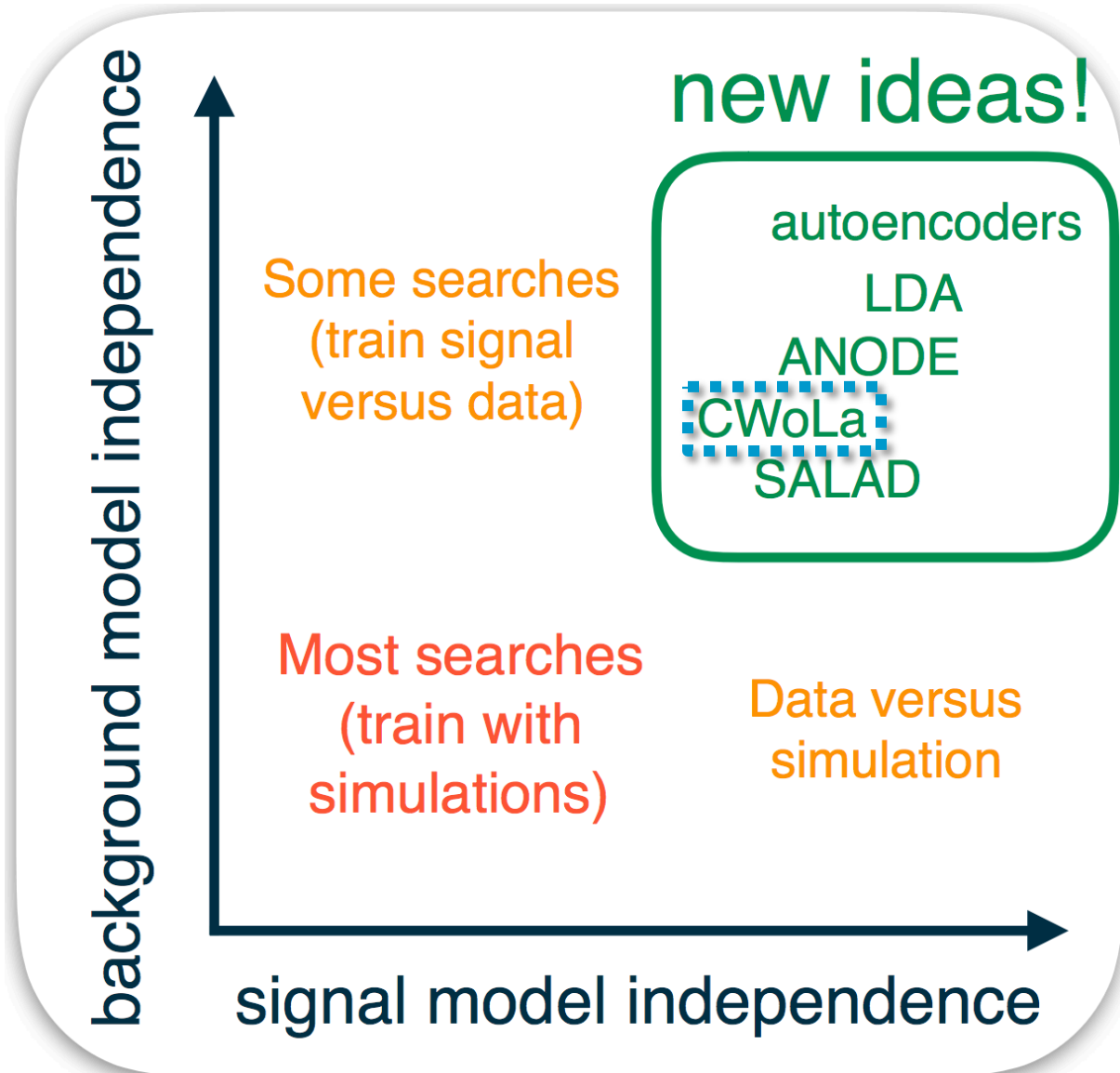
Conclusion

- Generic $A \rightarrow B(\text{jet})C(\text{jet})$ dijet resonance search using CWoLa hunting
- 3D bump hunt in m_A, m_B, m_C
 - Avoids large trials factor
- Learn to tag directly from data
- Fit to m_{JJ} spectrum after tagging
 - With no signal injected, no evidence for excess
 - With signal injected
 - NN learns to tag many signals
 - Exclusion limit improved significantly

- [1] ATLAS Collaboration, G. Aad et. al., Dijet resonance search with weak supervision using $\sqrt{s} = 13$ TeV pp collisions in the ATLAS detector, Submitted to Phys. Rev. Lett. [2005.02983]
- [2] <https://home.cern/science/physics/standard-model>
- [3] ATLAS Collaboration, Standard Model Summary Plots Summer 2019, Tech. Rep. ATL-PHYS-PUB-2019-024, CERN, Geneva, Jul, 2019.
- [4] <http://cms.web.cern.ch/news/jets-cms-and-determination-their-energy-scale>
- [5] <https://www.britannica.com/science/dark-matter>
- [6] R. E. Allen, The Higgs Bridge, Phys. Scripta 89 (2013) 018001 [1311.2647].
- [7] <https://cds.cern.ch/record/1095924>
- [8] <http://iopscience.iop.org/book/978-1-6817-4078-2/chapter/bk978-1-6817-4078-2ch4>
- [9] <https://twiki.cern.ch/twiki/bin/view/AtlasPublic/EventDisplayRun2Collisions>
- [10] ATLAS Collaboration, G. Aad et. al., Search for new resonances in mass distributions of jet pairs using 139 fb⁻¹ of pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector, JHEP 03 (2020) 145 [1910.08447].
- [11] ATLAS Collaboration, G. Aad et. al., Search for diboson resonances in hadronic final states in 139 fb⁻¹ of pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector, JHEP 09 (2019) 091 [1906.08589].
- [12] <https://indico.cern.ch/event/649482/contributions/2993318/>
- [13] <https://blogs.exeter.ac.uk/exeterblog/blog/2016/10/18/how-to-win-friends-and-influence-people-using-robots/>
- [14] E. M. Metodiev, B. Nachman and J. Thaler, Classification without labels: Learning from mixed samples in high energy physics, JHEP 10 (2017) 174 [1708.02949].
- [15] <https://indico.cern.ch/event/558411/contributions/3420889/>
- [16] J. H. Kim, K. Kong, B. Nachman and D. Whiteson, The motivation and status of two-body resonance decays after the LHC Run 2 and beyond, JHEP 04 (2020) 030 [1907.06659]
- [17] B. Nachman and D. Shih, Anomaly Detection with Density Estimation, Phys. Rev. D 101 (2020) 075042 [2001.04990]
- [18] J. H. Collins, K. Howe and B. Nachman, Extending the Bump Hunt with Machine Learning, Phys. Rev. D 99 (2019) 014038 [1902.02634]
- [19] <http://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PAPERS/PERF-2012-02/>
- [20] <https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PLOTS/JETM-2017-002/>

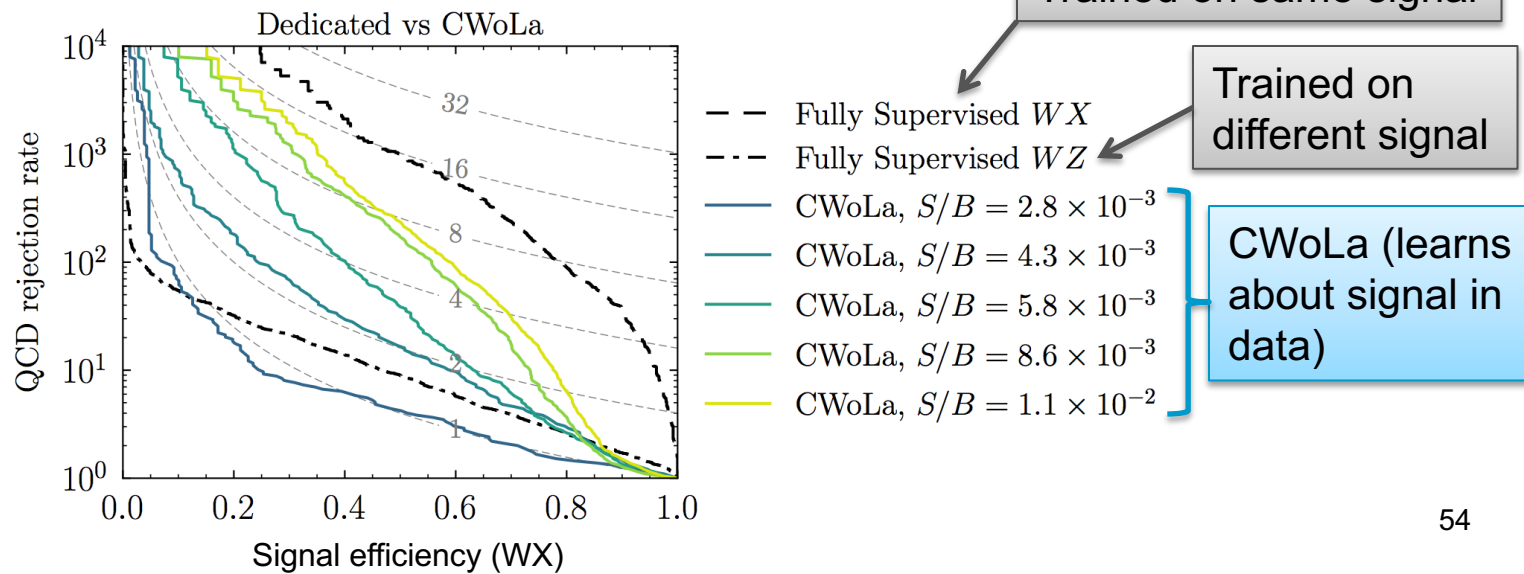
Backup

Landscape of Model-Independent Searches



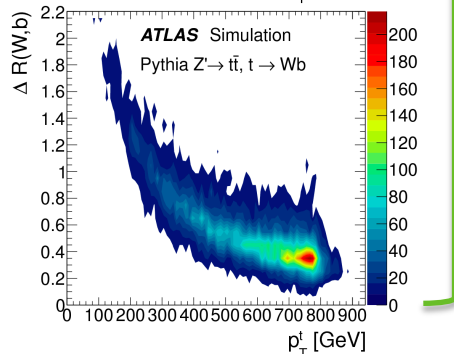
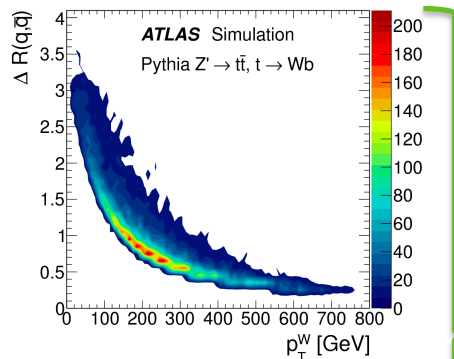
Natural Questions for CWoLa Hunting

- What is CWoLa hunting sensitive to? How to set limits?
- How to make sure cuts don't sculpt background?
 - Correlation of features with m_{JJ}
 - Train/test split
- Why not a direct search?
 - Direct searches – more sensitive, large trials factor
 - CWoLa hunting – sensitive to each signal, avoid large trials factor
 - Train/test split

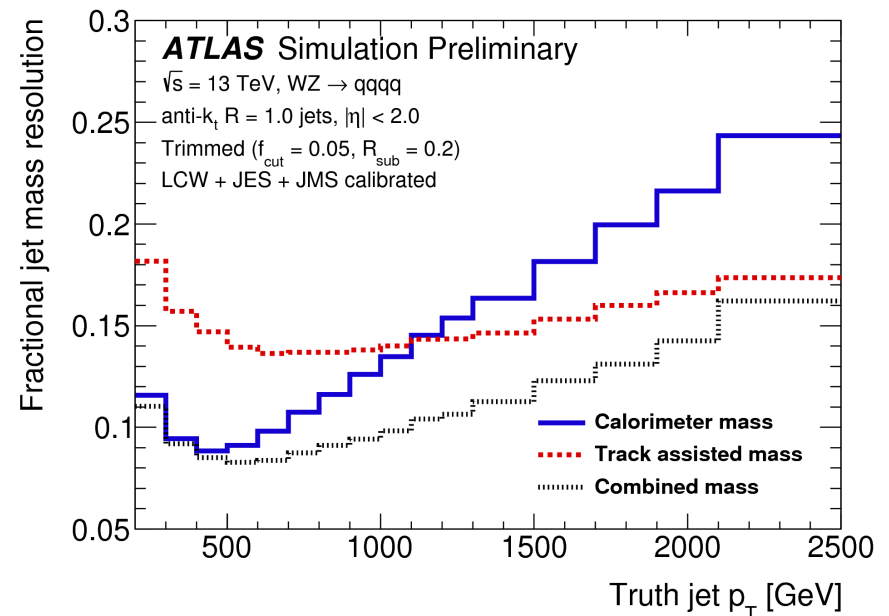


Details

- Full Run 2 (2015-2018) (139 fb^{-1})
- Lowest unrescaled large-R jet trigger (offline $p_T > 500 \text{ GeV}$)
 - Trimmed large R jets, combined jet mass - m_1, m_2
- Selection:
 - Two jets, $p_T > 200 \text{ GeV}$; leading jet $p_T > 500 \text{ GeV}$
 - $m_2 > 30 \text{ GeV}$; $m_1 < 500 \text{ GeV}$ (limit learning differences in tails)

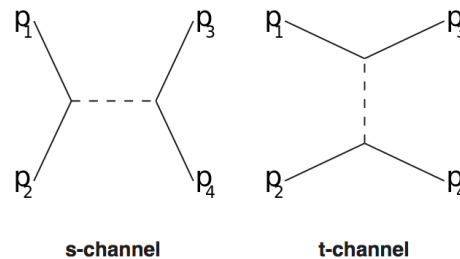


$$\Delta R \sim 2m/p_T$$

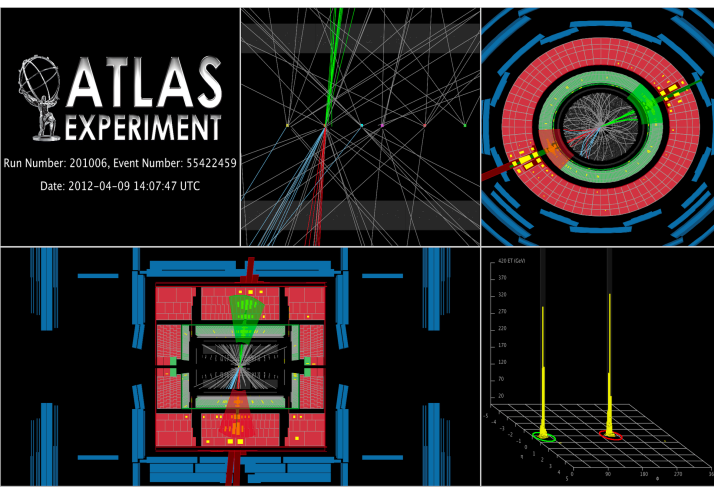


Rapidity Cut

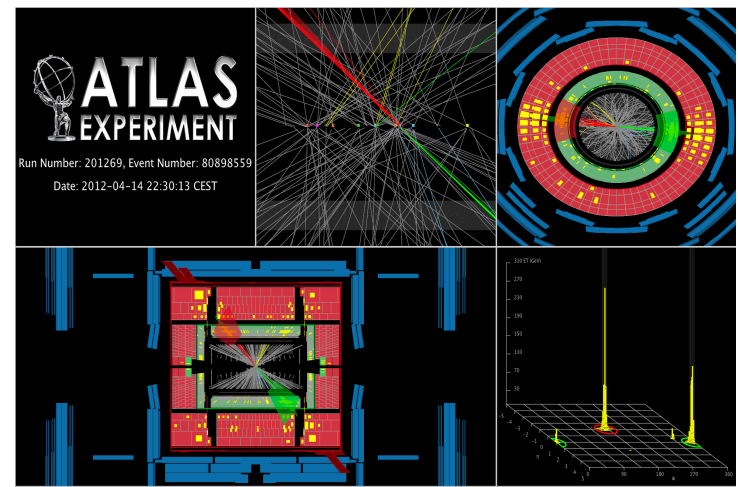
- Cut on difference in rapidity ($|y_1 - y_2| < 1.2$)
 - Reduce QCD t-channel processes
 - Same as VVJJ, inclusive dijet
 - Inverted rapidity cut regime serves as data validation region
- Signals: $W' \rightarrow WZ$, variable W, Z masses



$m_{JJ} = 3.62 \text{ TeV}; |y_1 - y_2| = 0.24$

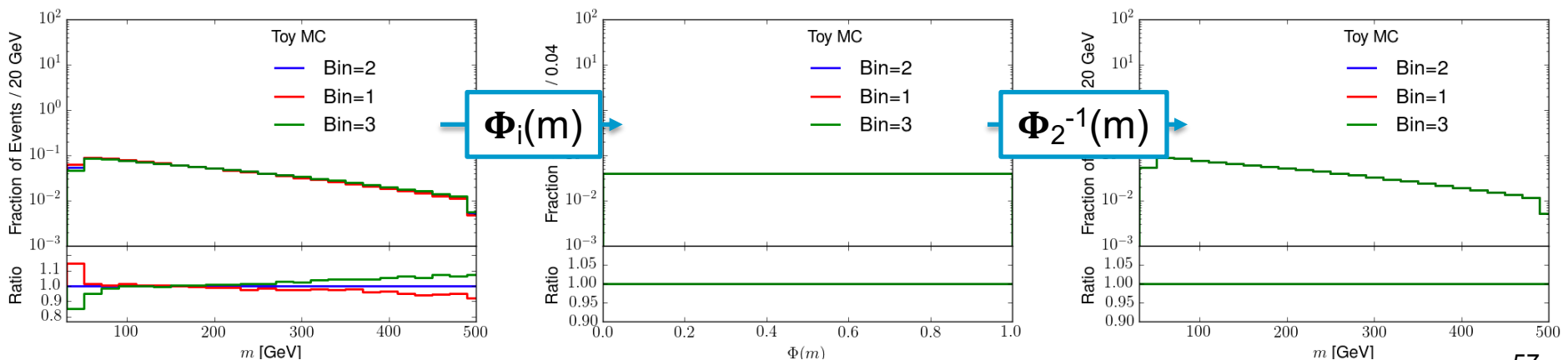


$m_{JJ} = 4.23 \text{ TeV}; |y_1 - y_2| = 2.08$



Setup - Features

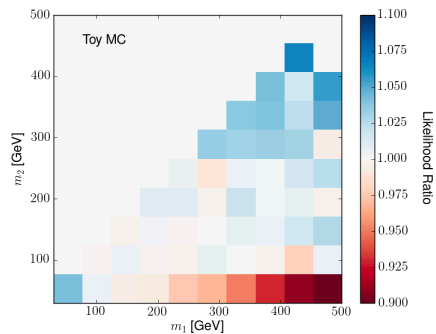
- Train network to distinguish between signal and sideband regions
- Network can learn true differences in features between signal and sideband regions
 - Since learning to distinguish signal region from sidebands, leads to bump!
- Decorrelate 1D $m_J = \{m_1, m_2\}$ distribution by percentile scaling
 - Use empirical distribution function
 - $\Phi_i(x) = (\# \text{ of samples in bin } i \leq x) / (\# \text{ of samples in bin } i)$
 - Uniform by definition



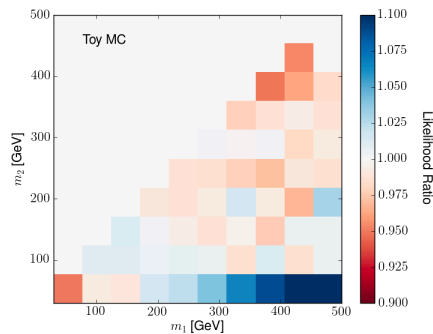
Training vs Sidebands

- Decorrelate 1D m_j distribution
- Remaining correlations in 2D m_1 vs m_2 distribution
 - Combine sidebands (weight = $1/N$) to further cancel out differences

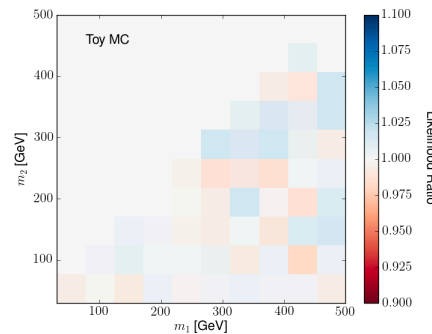
Vs Lower Sideband



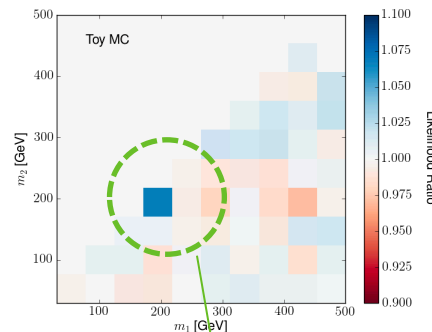
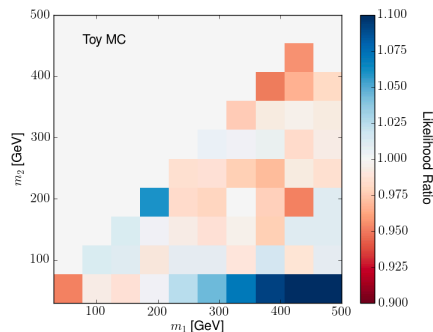
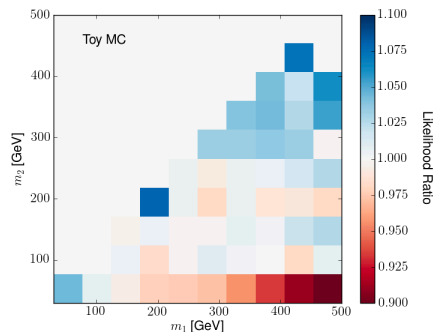
Vs Upper Sideband



Vs Comb. Sidebands



With Signal ($m_B = m_C = 200$ GeV):



Excess! NN should learn to target

- Parametrized background fit
 - Bins: 100 GeV between 1.8 - 8.2 TeV
 - 3 fit functions:

$$\frac{dn}{dx} = p_1(1-x)^{p_2-\xi p_3} x^{-p_3}$$

1

$$\frac{dn}{dx} = p_1(1-x)^{p_2-\xi_1 p_3} x^{-p_3+(p_4-\xi_2 p_3-\xi_3 p_2) \log(x)}$$

2

$$\frac{dn}{dx} = p_1 x^{p_2-\xi_1 p_3} e^{-p_3 x+(p_4-\xi_2 p_3-\xi_3 p_2) x^2}$$

3

- For given signal region, sidebands = signal regions ± 0.5

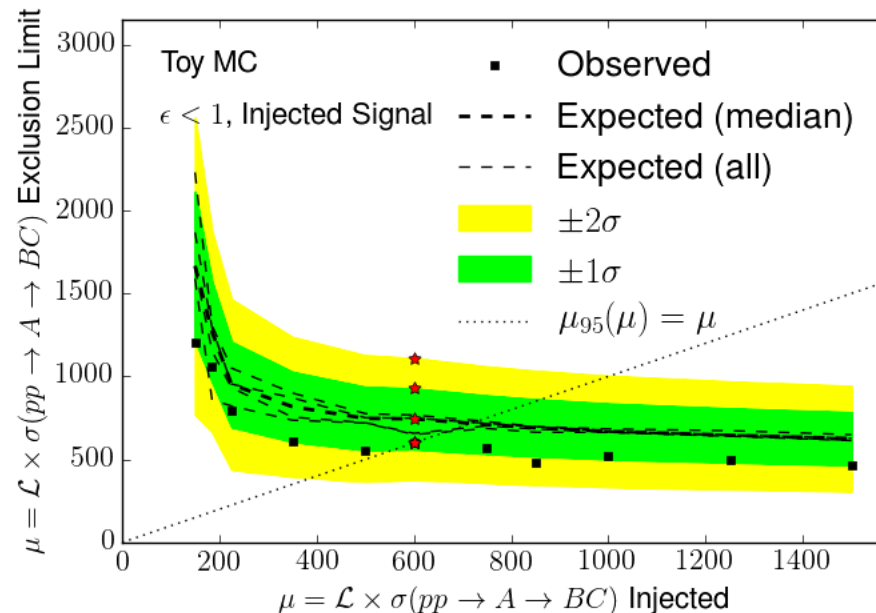
Table 5: m_{jj} bin definitions and the mask regions for the background fit.

Bin	Definition	Mask
5	$2.74 \leq m_{jj} < 3.28$ TeV	$2.5 \leq m_{jj} < 3.6$ TeV
6	$3.28 \leq m_{jj} < 3.94$ TeV	$3.0 \leq m_{jj} < 4.3$ TeV
7	$3.94 \leq m_{jj} < 4.73$ TeV	$3.6 \leq m_{jj} < 5.2$ TeV
8	$4.73 \leq m_{jj} < 5.68$ TeV	$4.3 \leq m_{jj} < 6.2$ TeV
9	$5.68 \leq m_{jj} < 6.81$ TeV	$5.2 \leq m_{jj} < 7.5$ TeV

- Fit to all, finish when fit has χ^2 $p > 0.05$ in sidebands:
 - Cycle through fit functions
 - Reduce sideband size by 400 GeV; repeat

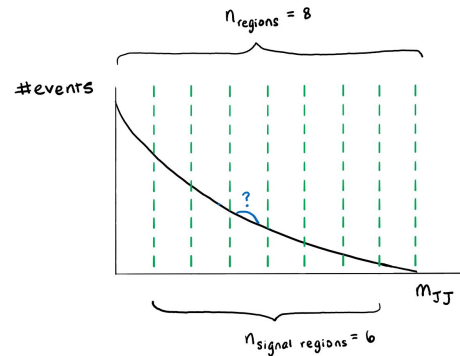
μ Scan Limits

- Coarse scan over injected μ , and train a different NN with each value
 - 5 different random samplings of signal
- For the given NN treat μ as a POI in the fit and find a limit
- Expected/observed limits for that μ are $\text{median}_{\text{samplings}}(\max(\text{CL}(\mu), \mu))$
- Median expected limit is min over all μ
 - Expected limit bands given as bands for median NN $\oplus$ variation in median expected limit

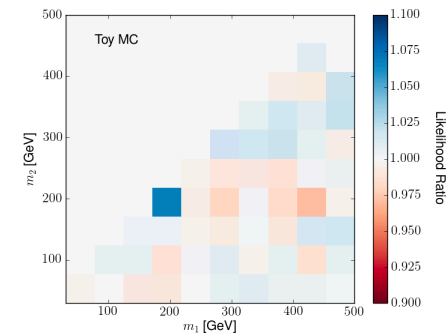
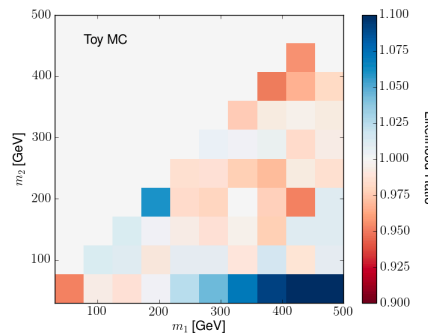
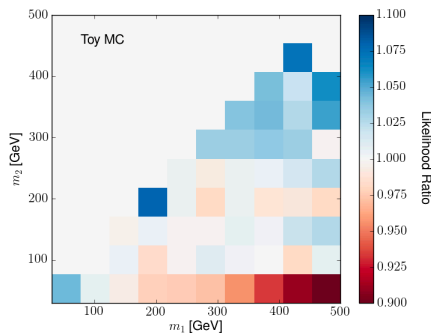


Challenges and Remarks

- Overlapping m_{JJ} regions

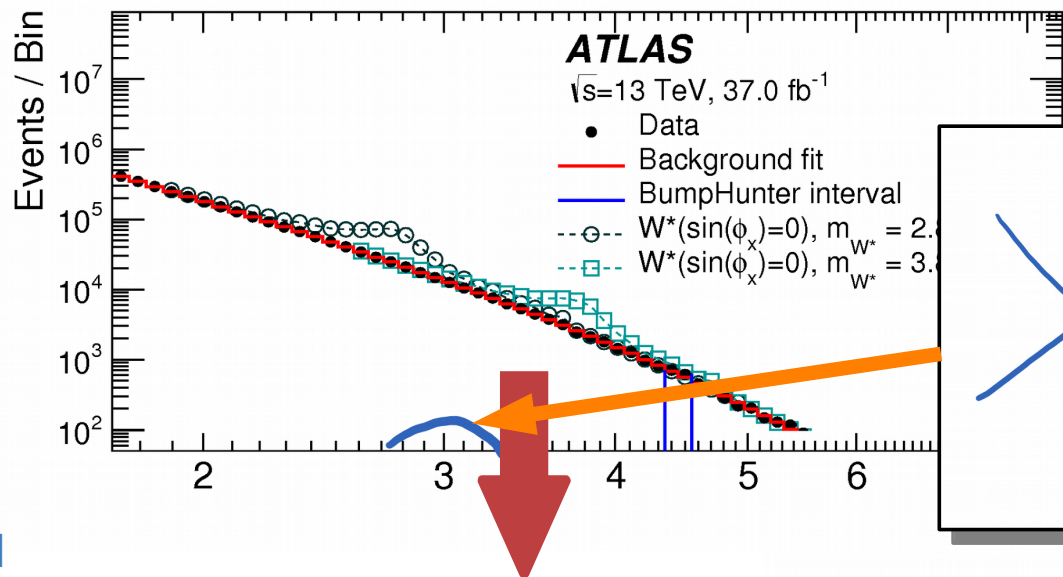


- Feature and m_{res} decorrelation

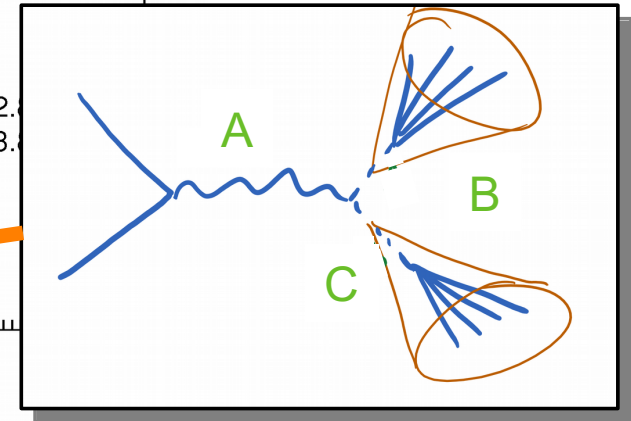


- Parametric fitting with large bin counts
- MC validation
- NN training time
- Limit setting

Dijet Resonances

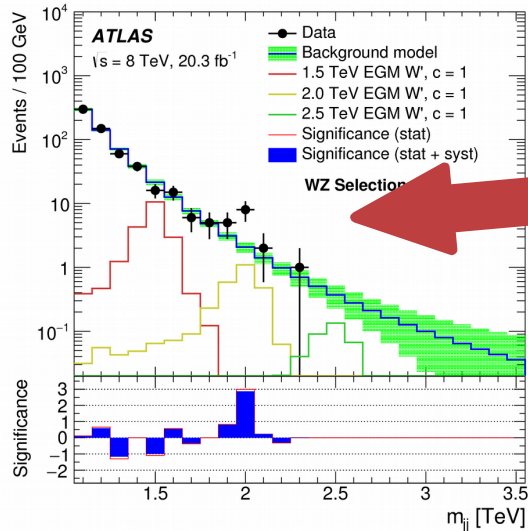


Edited from [1703.01927]

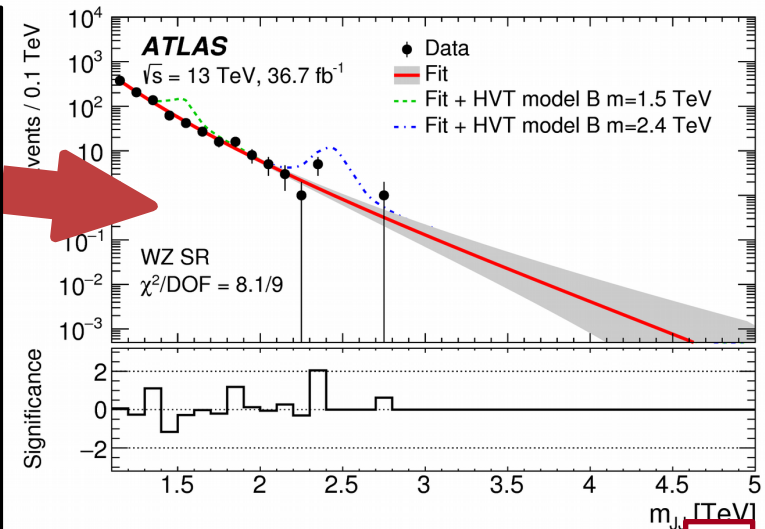


[1708.04445]

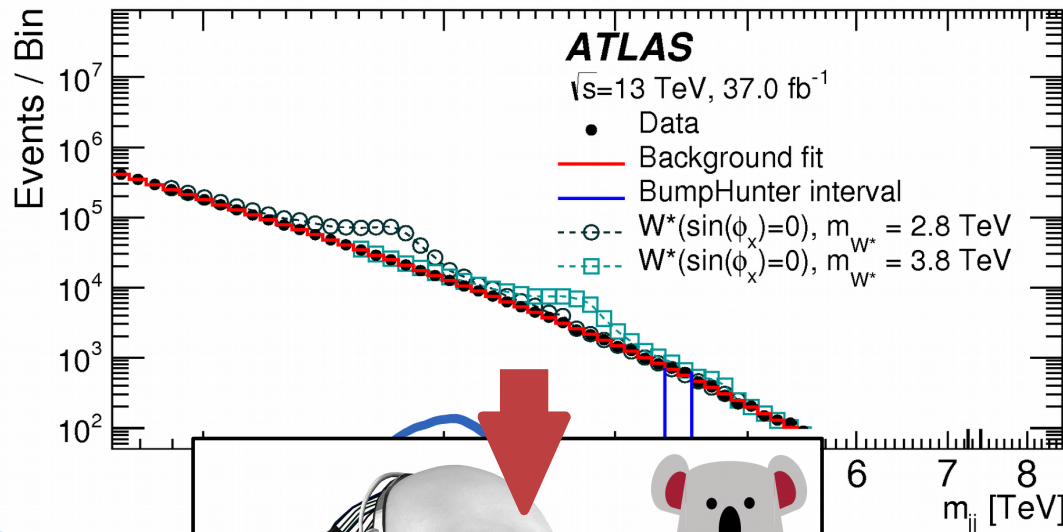
[1506.00962]



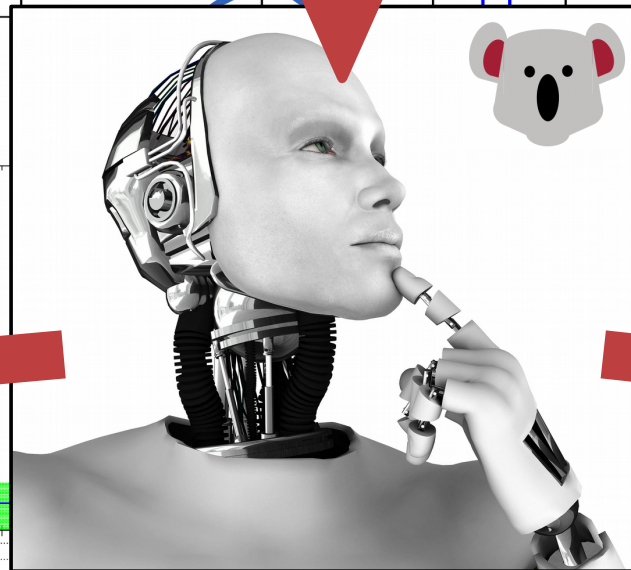
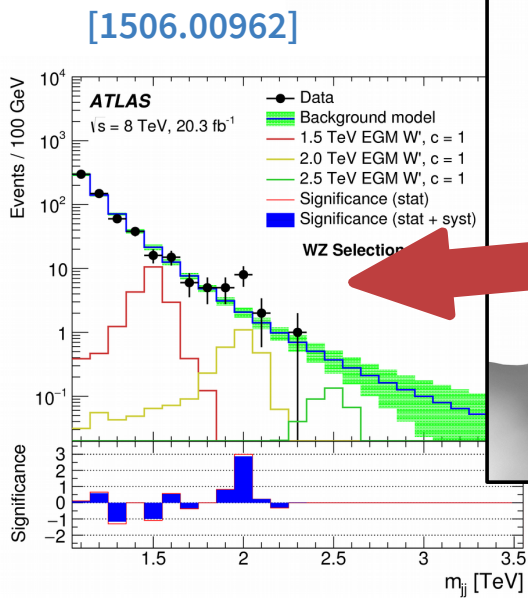
- 1) Theorist comes up with specific model with some specific prediction (e.g. $W' \rightarrow WZ$).
- 2) Choose dedicated substructure variables.
- 3) Simulate signal to optimize cuts
- 4) Calibrate in some data sample
- 5) Apply cuts to events and look for a bump in the new distribution



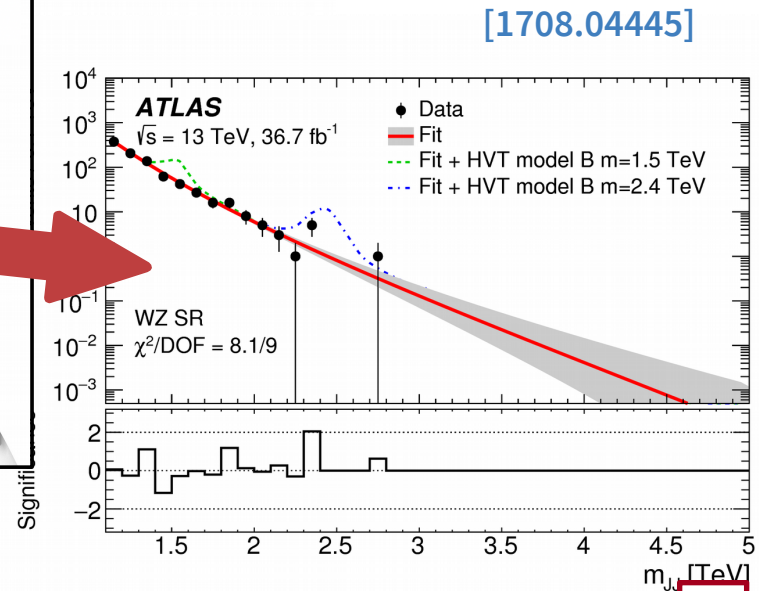
Dijet Resonances



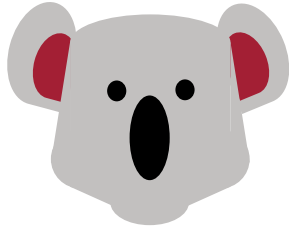
Edited from [1703.01927]



5) Apply cuts to events and look for a bump in the new distribution



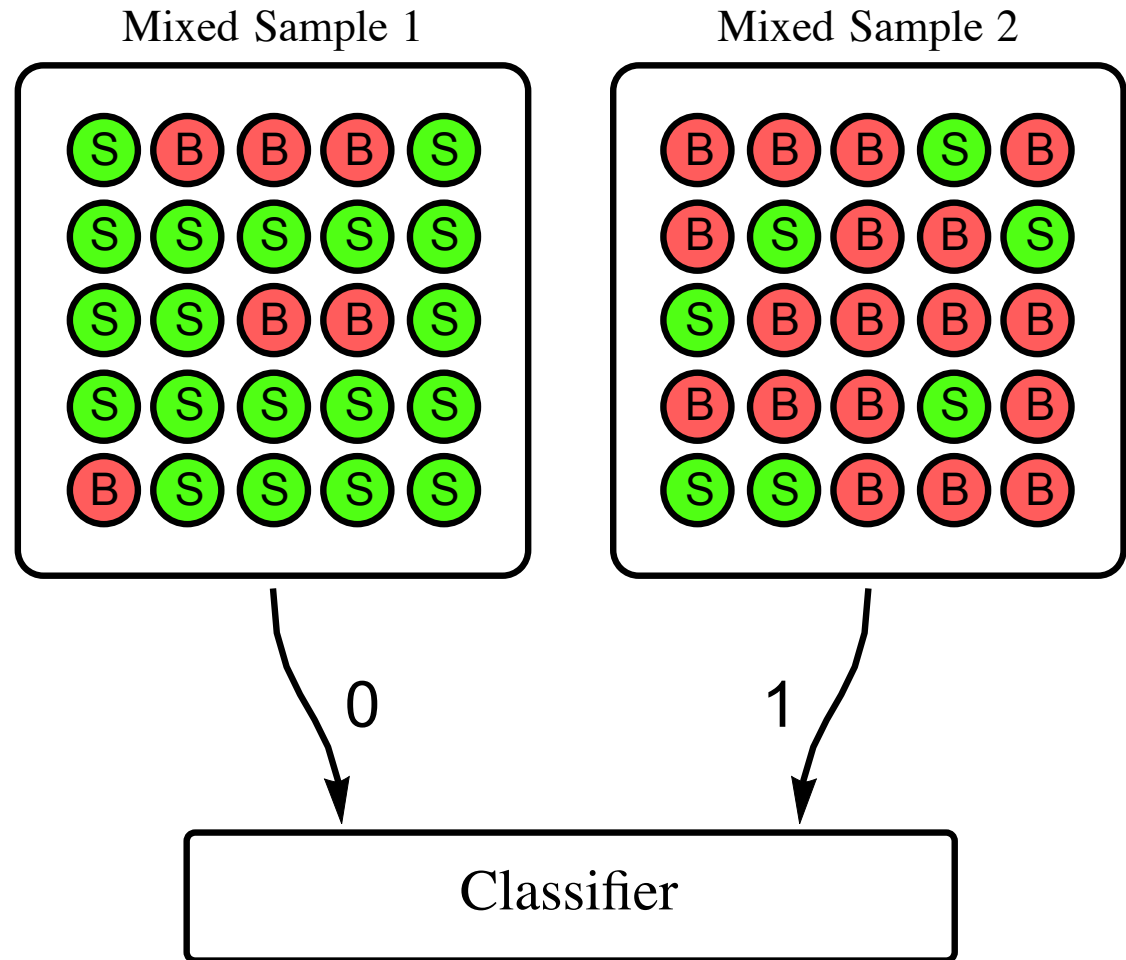
Method 2: Learning without Proportions

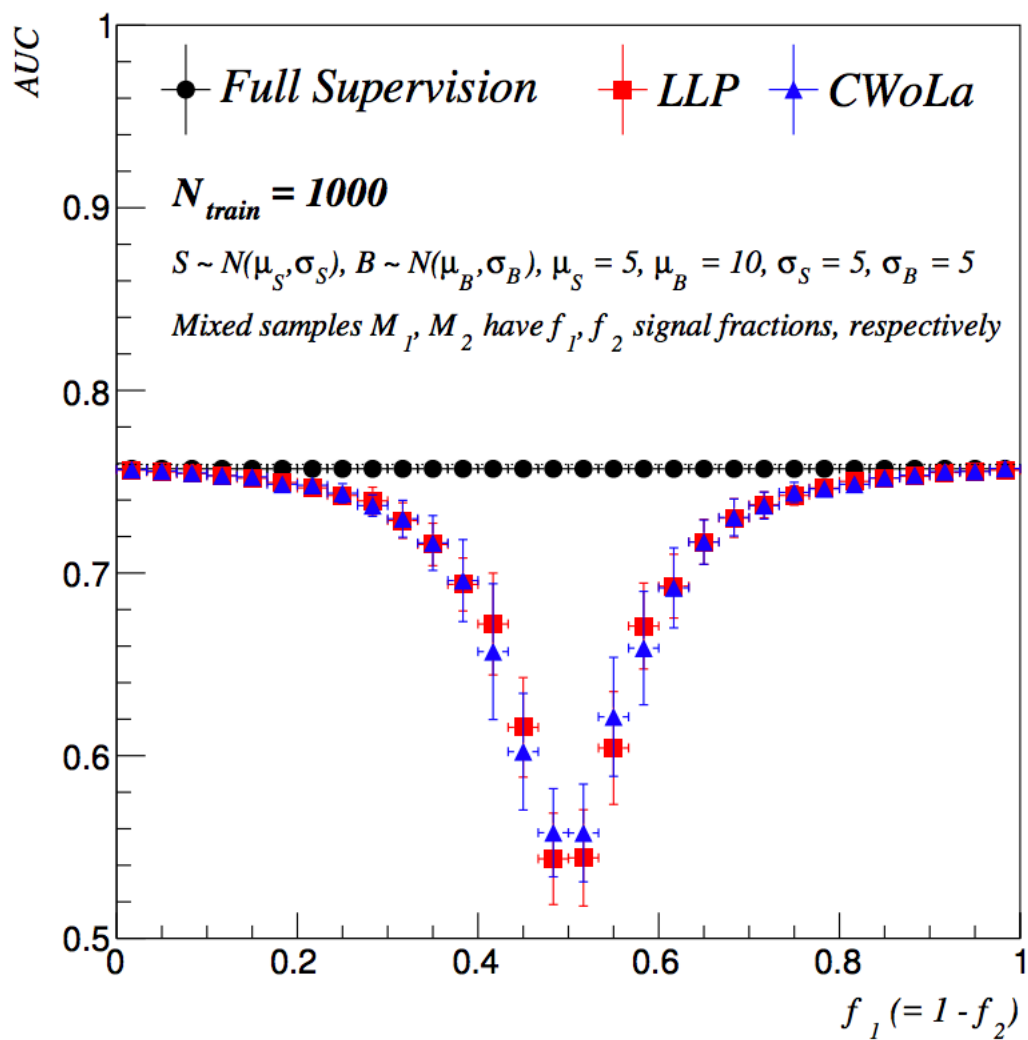


CWoLa

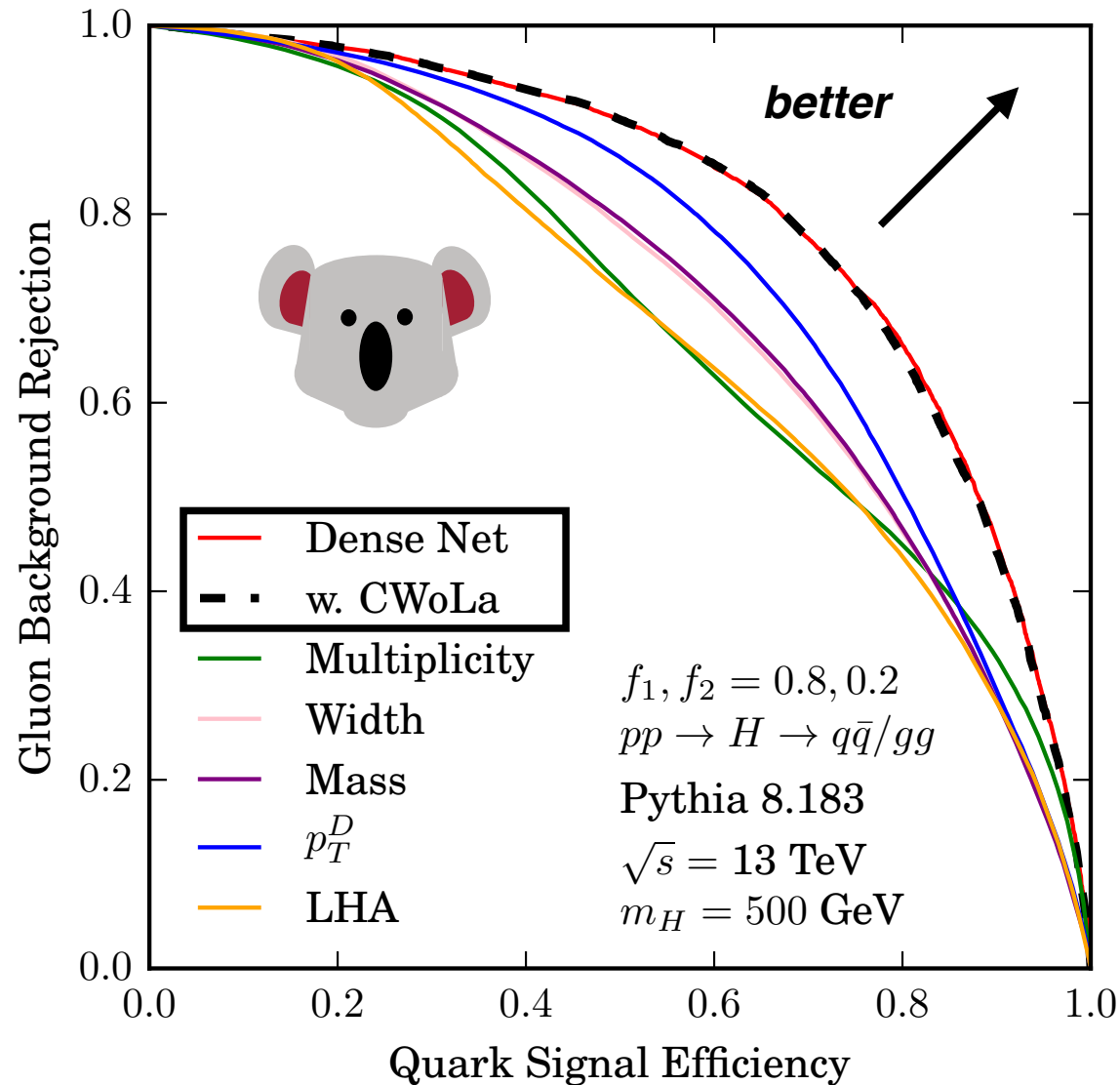
*Classification
Without Labels*

Solution: Train
directly on data using
mixed samples

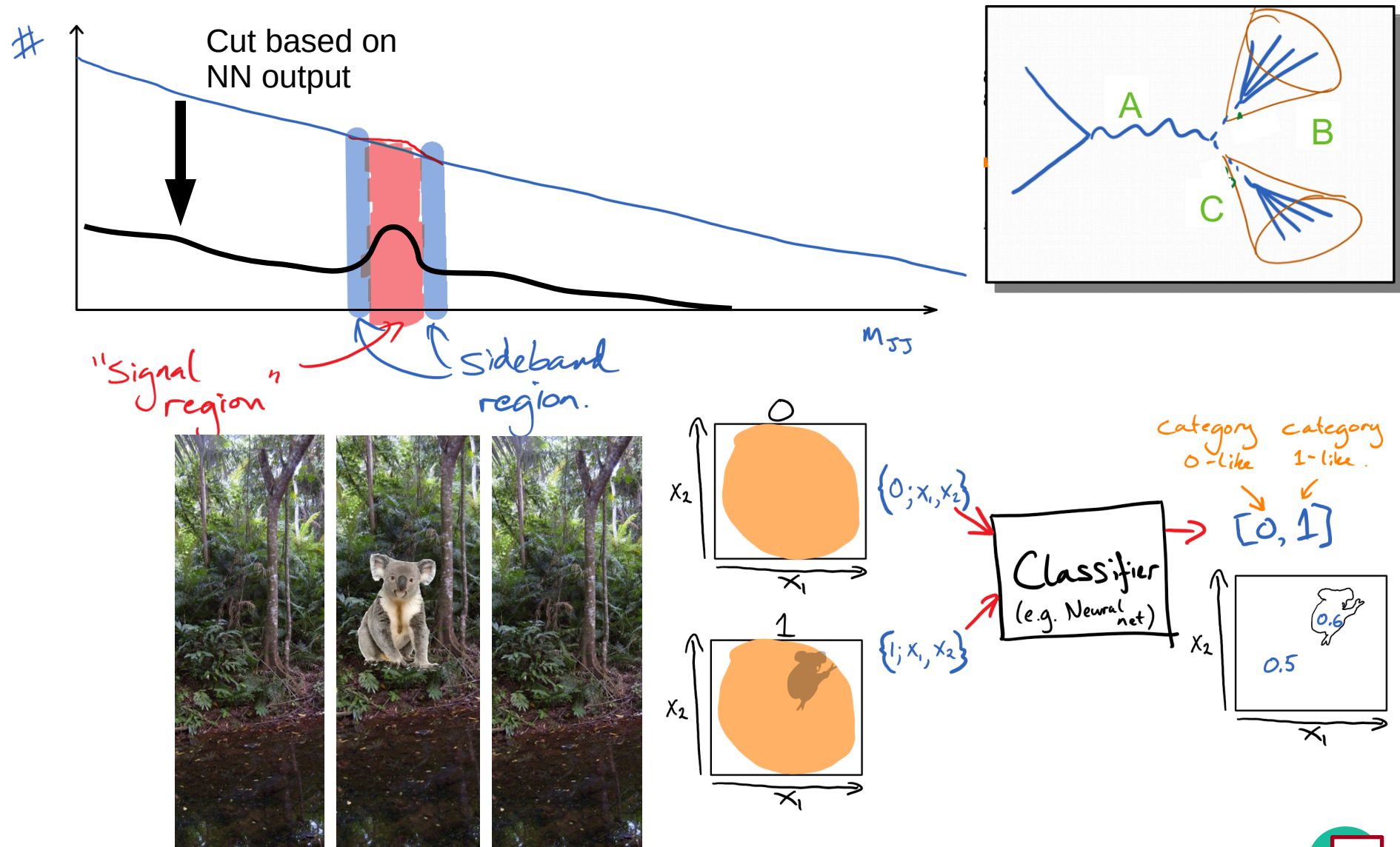




Works in low-dimensions

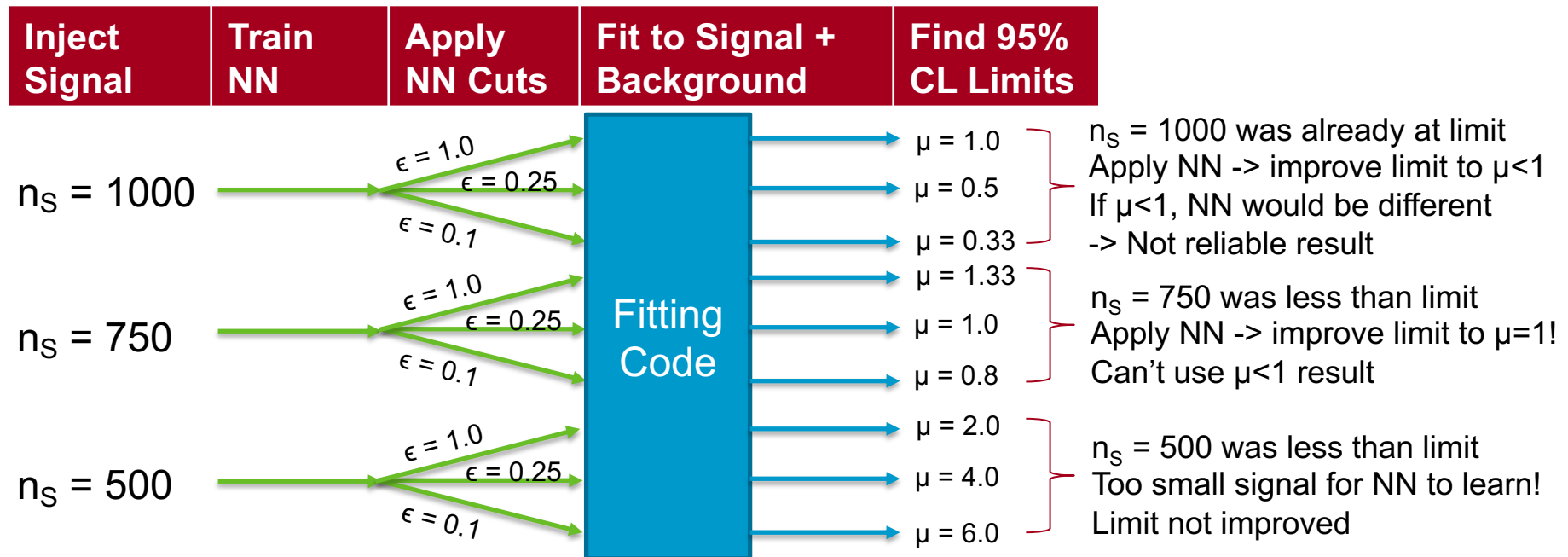


CwoLa Hunting: Basic Picture

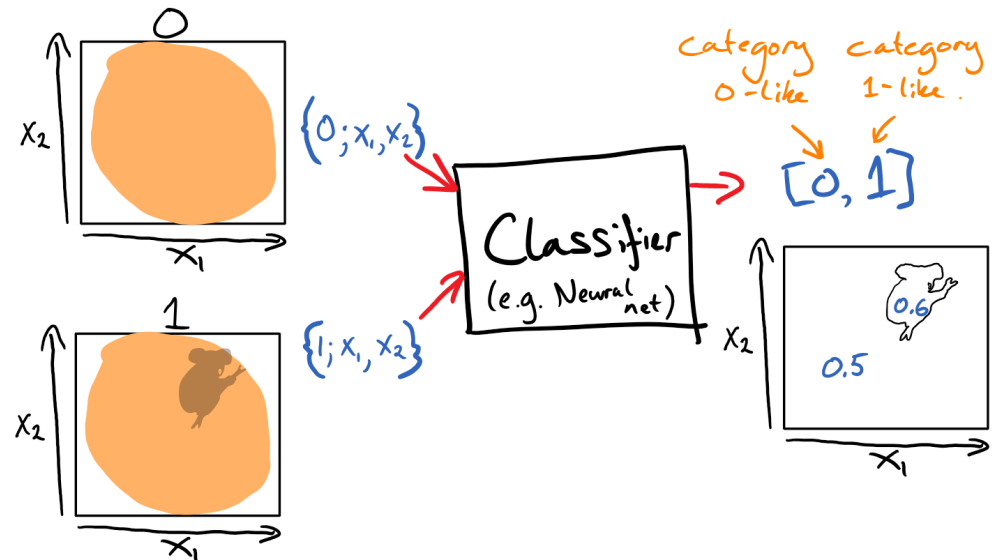
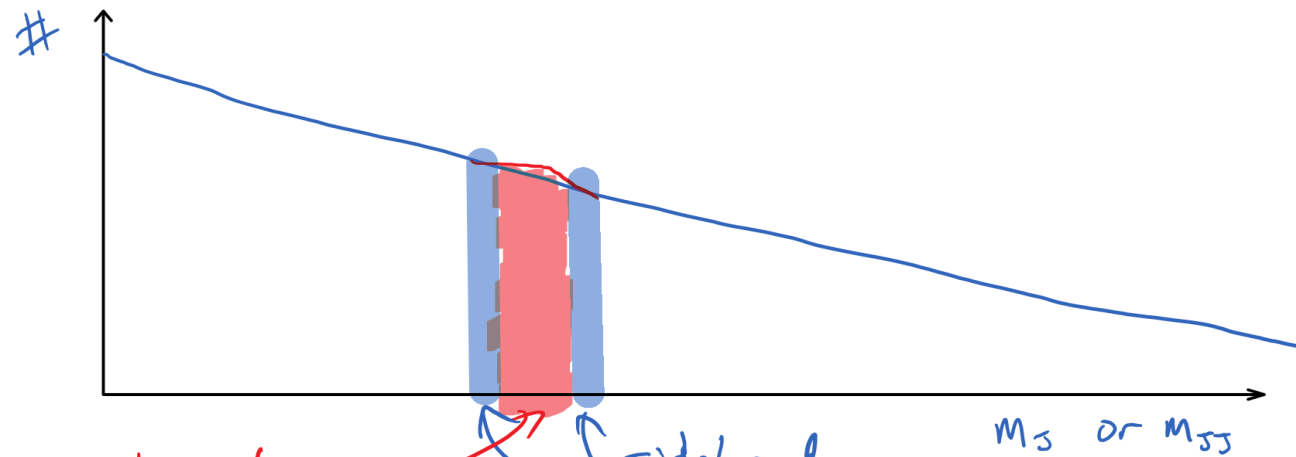


Statistical Analysis Details

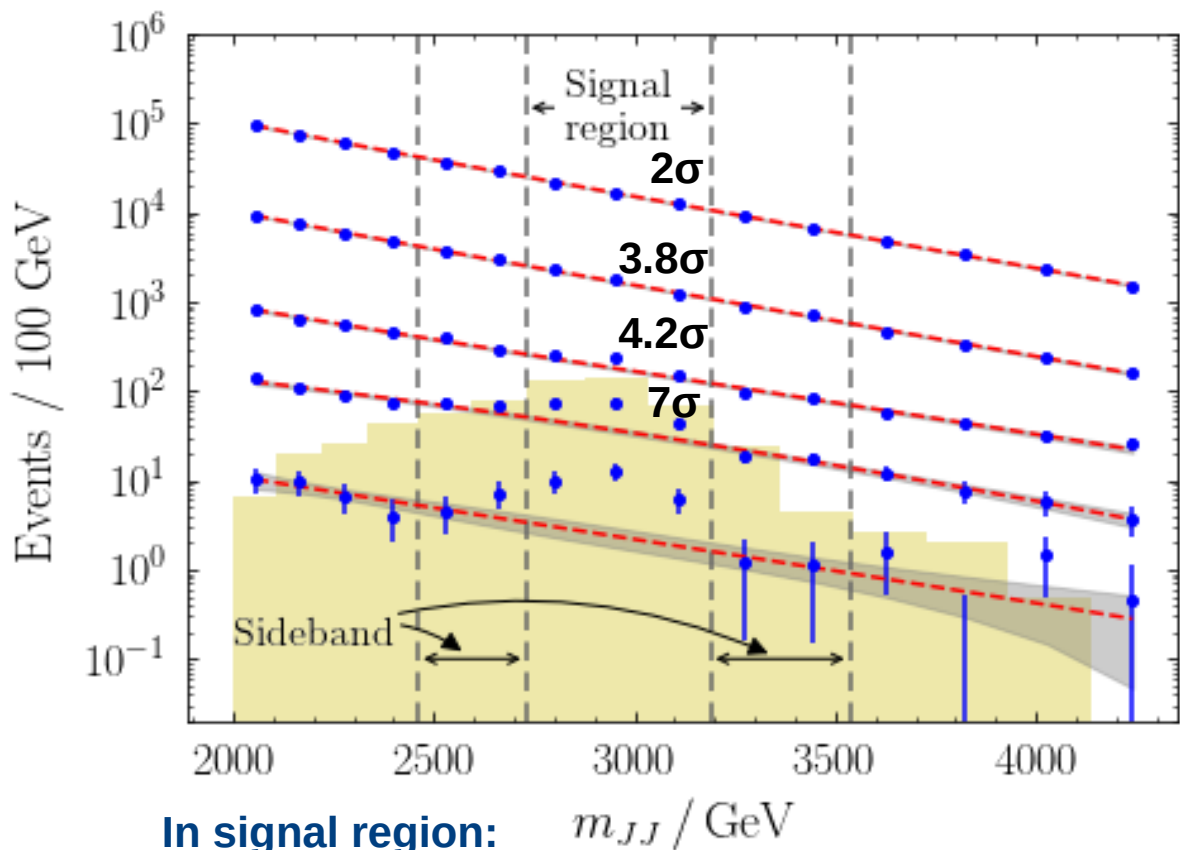
- Exclusion limit illustrative example:
- Injected signal with strength $n_S = L \cdot \sigma \cdot \text{BR}$ (set limits on this quantity)



CwoLa Hunting: Basic Picture



Application to Bump Hunt



In signal region:

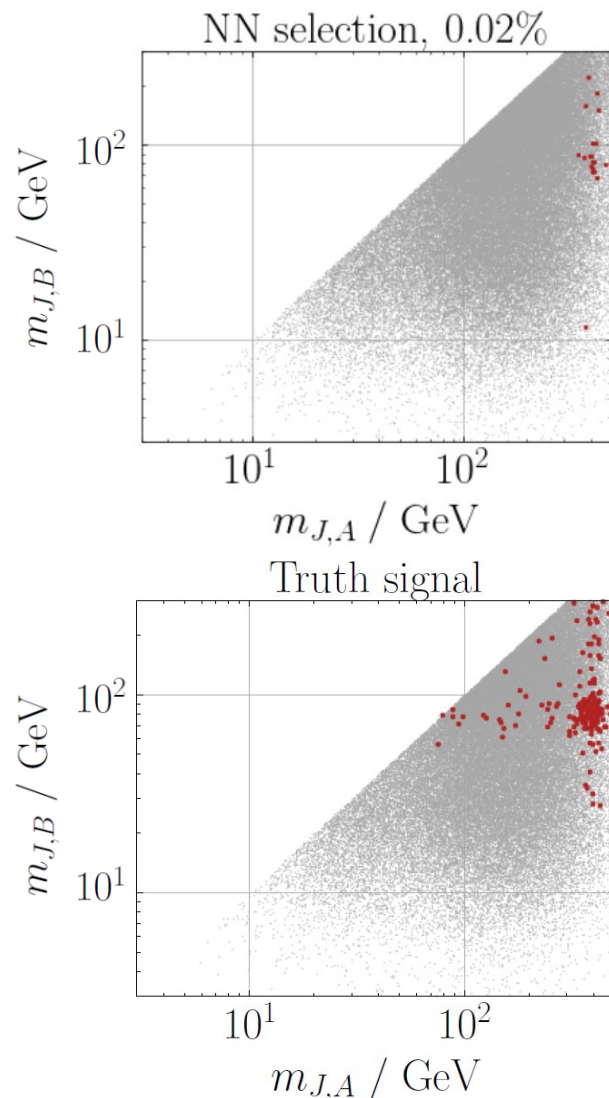
$S = 522$,

$S/B = 0.64\%$

For each jet: $Y_i = \left(m_J, \underbrace{\sqrt{\tau_1^{(2)} / \tau_1^{(1)}}, \tau_{21}, \tau_{32}, \tau_{43}}_{\text{Measure } n\text{-pronginess of jet}}, n_{\text{trk}} \right)$

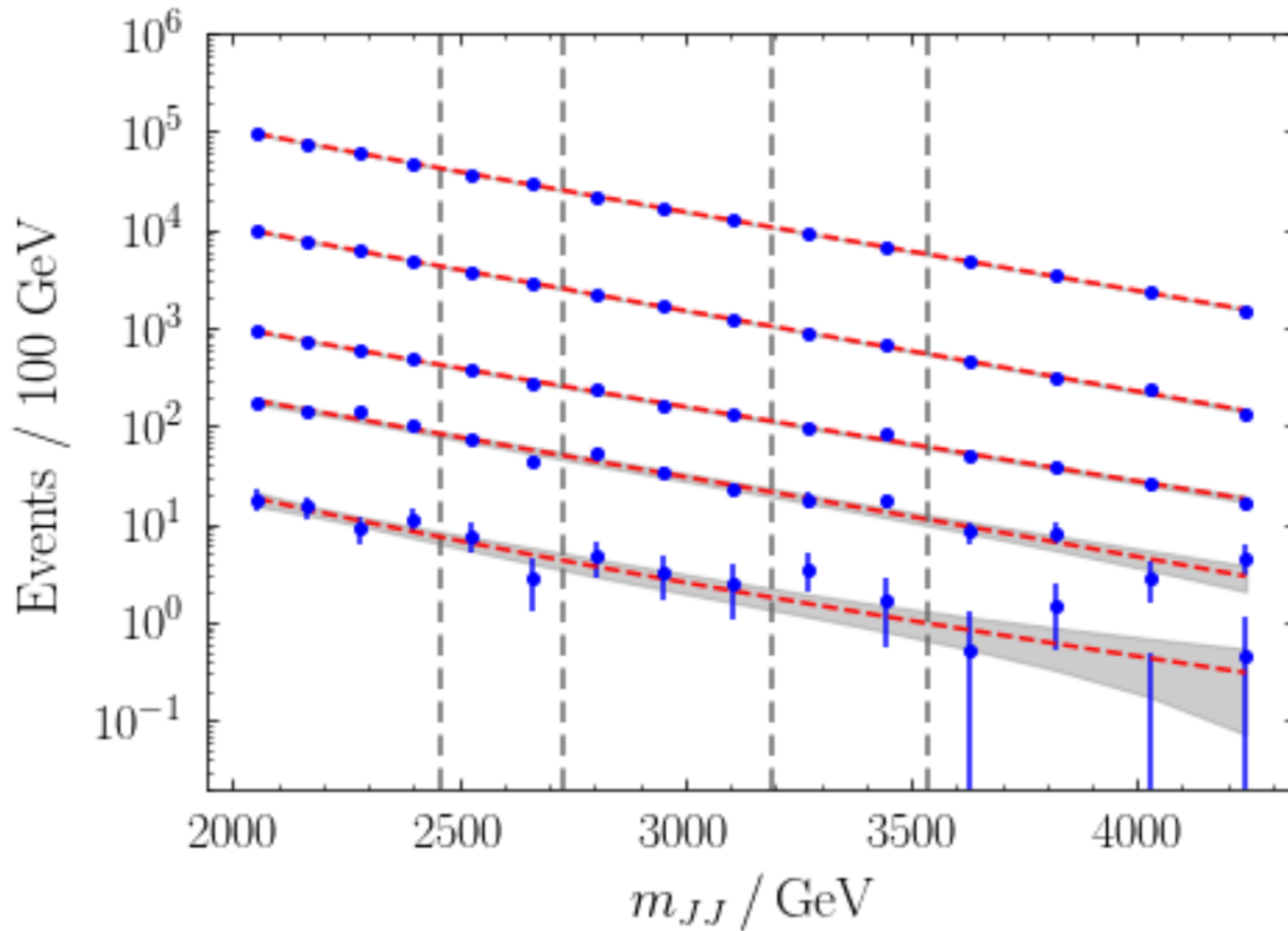
Measure n -pronginess of jet

CWoLa Hunting

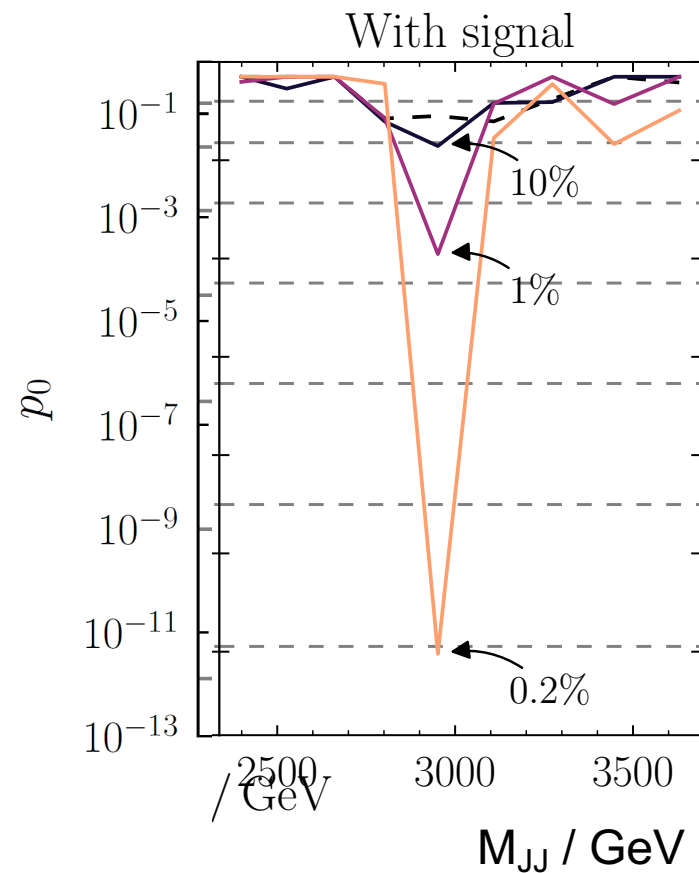
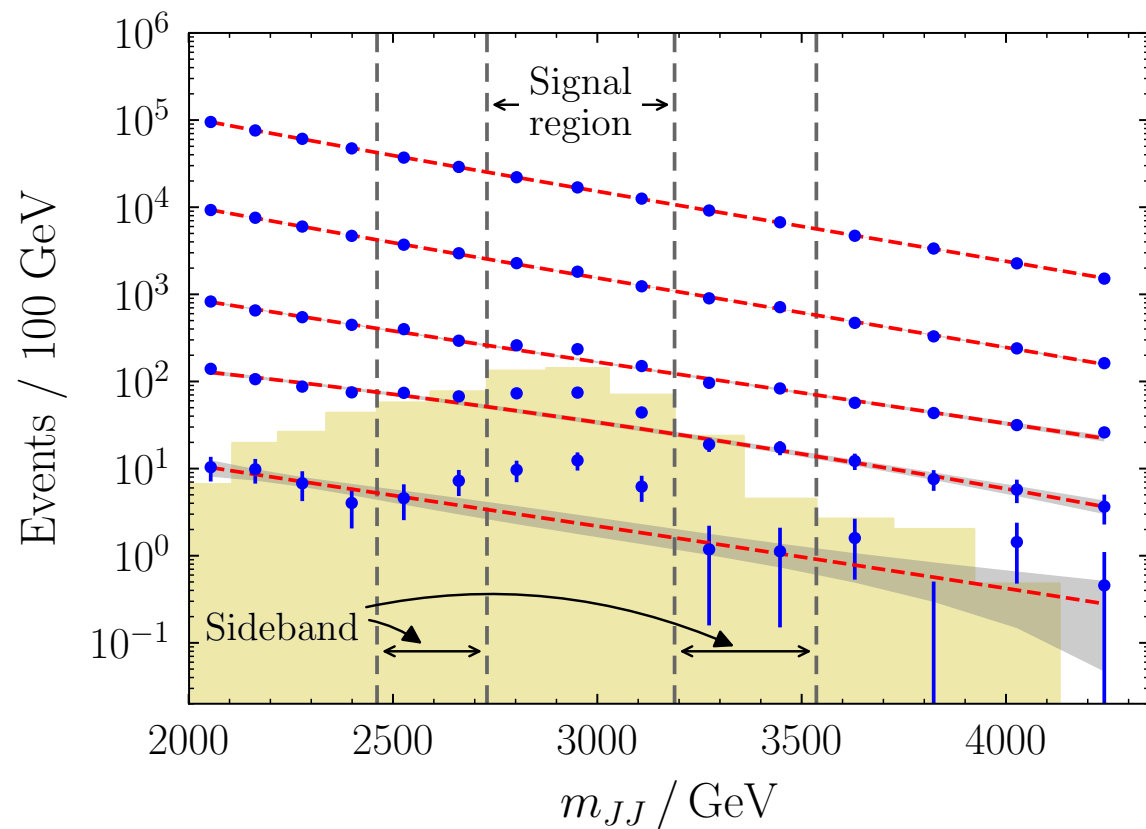


No Signal → No Bump!

(Need to be careful to make this work. Details in backup slides and in paper)



Mass Scan



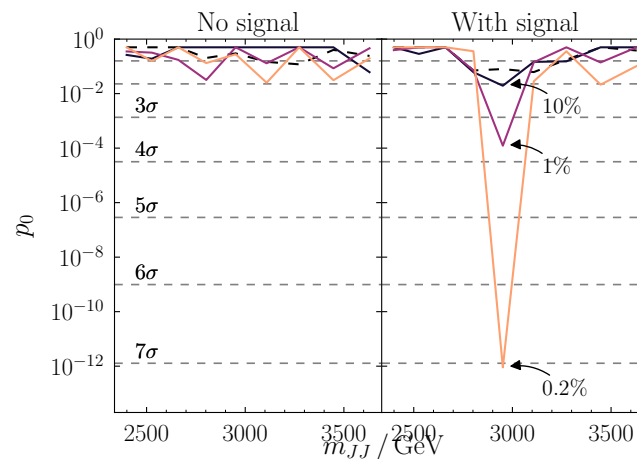
- Proposal for CWoLa Hunting in ATLAS
 - *Generic* – Sensitive to a broad class of familiar and exotic models
 - *Data-driven* – Training and testing done directly on data
 - *Resonance search* – Search requires a resonance in final state
 - *Machine learning* – Method can learn to tag signal (if it exists) from basic features, even if specific signature of signal is unknown in advance

Assumptions of CWoLa Hunting

1. Need some variable X (e.g. m_{JJ}) in which background is smooth and signal is localized
2. Need some other variables $\{Y\}$ (e.g. jet substructure) which may provide discriminating power which may be a-priori unknown
3. Any cuts on $\{Y\}$ should leave background smooth in X

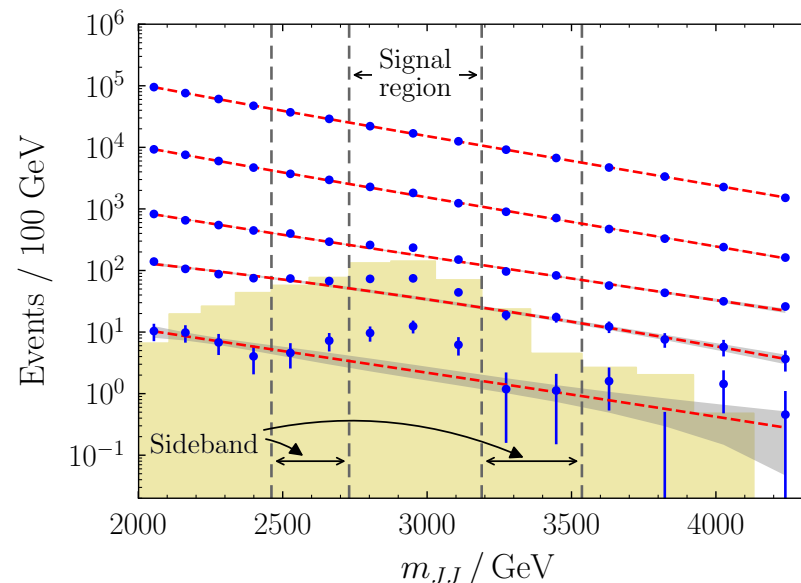
Limits from CWoLa Hunting

- Two sets of results, depending on features used in training
- Using all features, provide *model-independent* information - p-values according to background-only hypothesis
 - No dip: Data is well-described by background-only hypothesis
 - Dip observed: Useful information for experimentalists and theorists
 - Detailed study of features tagging events in the bump
 - E: Can be used as an “alarm system” for a dedicated search in Run 3
 - T: Can publish neural network so theorists can devise models that would be tagged by these features



Limits from CWoLa Hunting

- Two sets of results, depending on features used in training
- Using reduced feature set ($Y_i = \{m_{JJ}\}$), provide *model-dependent* information – limits on $A \rightarrow BC$ generic narrow resonance
 - Inject signal to compare B and S+B
 - Smart way of doing 3D bump hunt
 - Currently no dedicated 3D bump hunt in ATLAS -> new result
 - Can set proper limits because uncertainties available on jet mass



Sculpting Background: Overtraining

J. Collins, BOOST 2018



Nested Cross-Validation

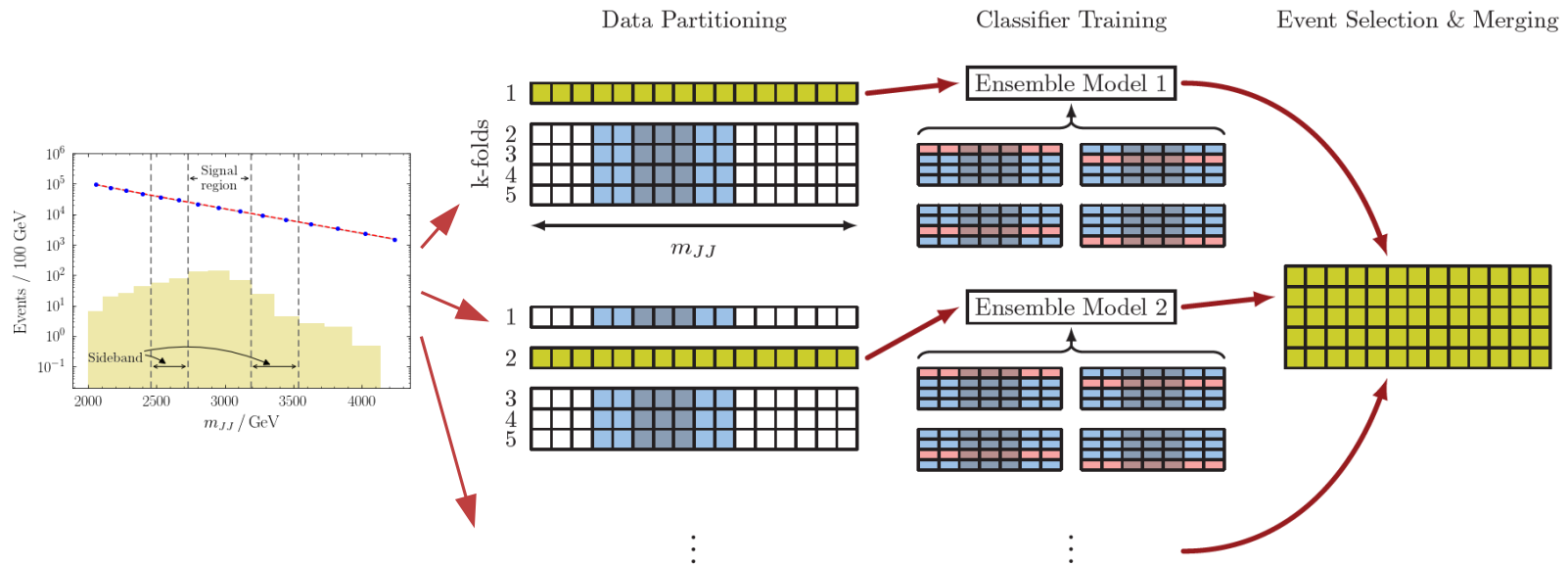


Figure 7. Illustration of the nested cross-validation procedure. **Left:** the dataset is randomly partitioned bin-by-bin into five groups. **Center:** for each group, an ensemble classifier is trained on the remaining groups. For each of the four possible combinations of these four groups into three training groups and one validation group, a set of individual classifiers are trained and the one with best validation performance is selected. The ensemble classifier is formed by the average of the four selected individual classifiers. **Right:** Data are selected from each test group using a threshold cut from their corresponding ensemble classifier. The selected events are then merged into a single m_{JJ} histogram.