CWoLa Hunting:

Extending the Bump Hunt with Machine Learning

Based on:

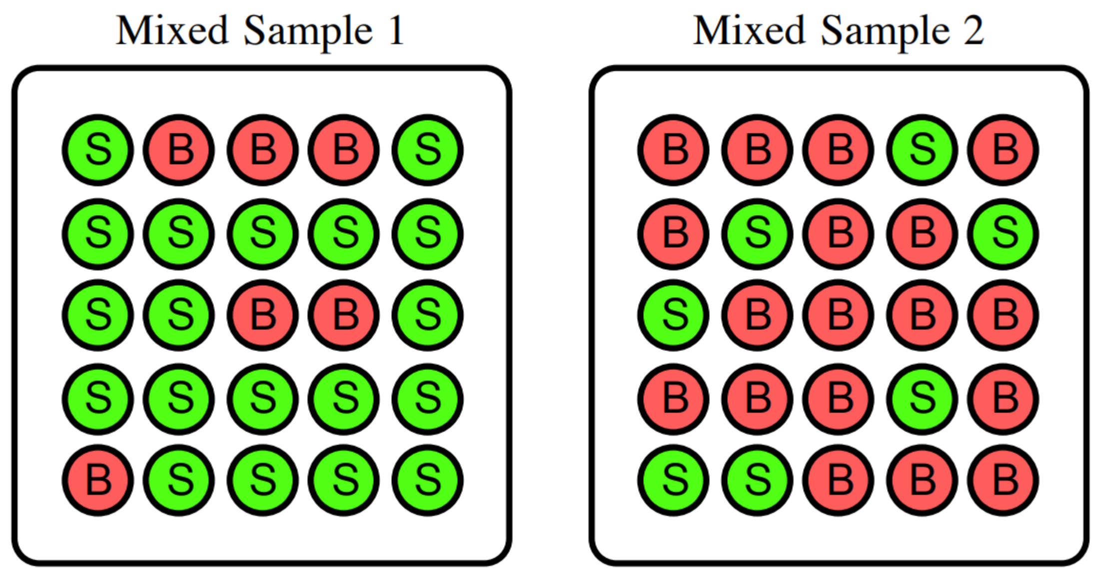
[1805.02664] Jack Collins, Kiel Howe, Ben Nachman

UNIVERSITY OF MARYLAND

JOHNS HOPKINS UNIVERSITY

CWoLa

CWoLa

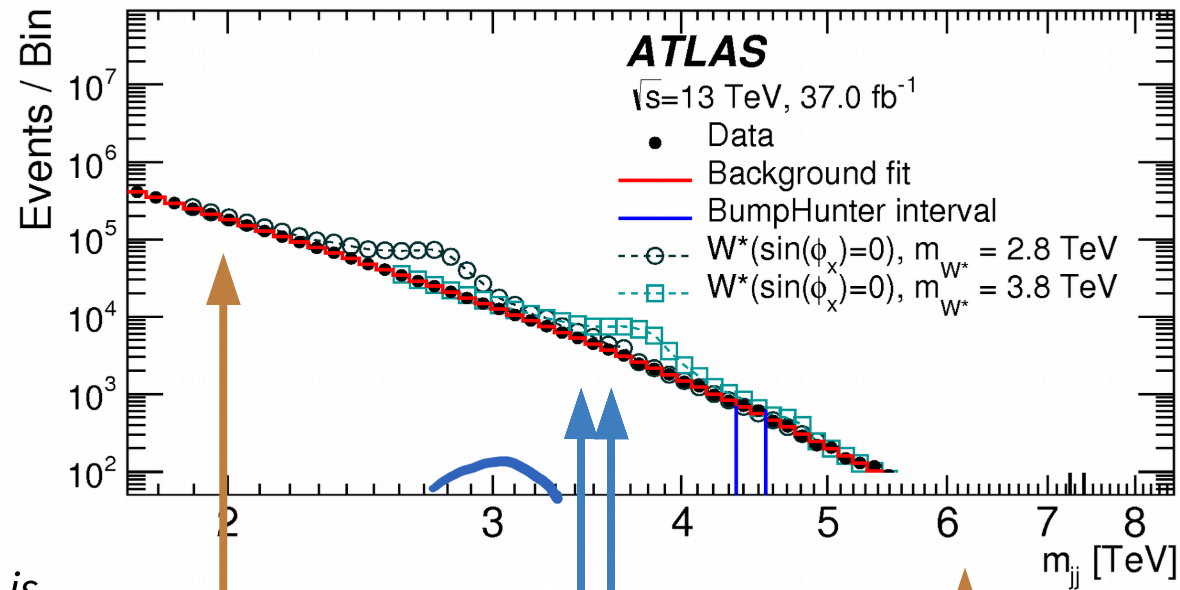


Classifier trained to optimally discriminate mixed sample 1 from mixed sample 2 *is also optimal* for discriminating S from B, so long as:

– Samples 1 and 2 contain different fractions of S and B
– S in sample 1 is drawn from the same distribution as S in sample 2
– B in sample 1 is drawn from the same distribution as B in sample 2
– Training statistics are sufficiently large

How to use this for a search where S is new physics and B is SM background?

Dijet Resonances



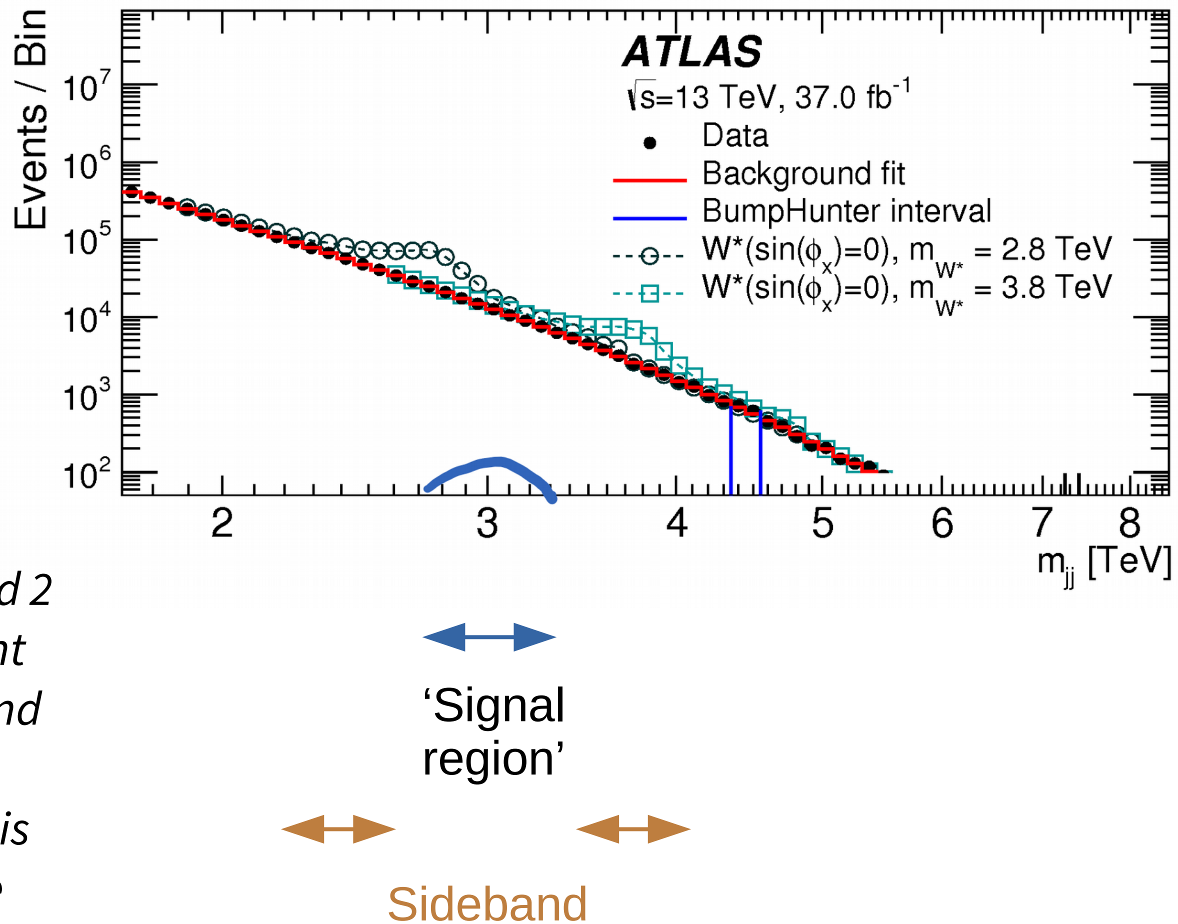
Edited from [1703.01927]

– B in sample 1 is drawn from the same distribution as B in sample 2

QCD events with **very similar** event characteristics (kinematics, **substructure**)

QCD events with **quite different** event characteristics (kinematics, **substructure**)

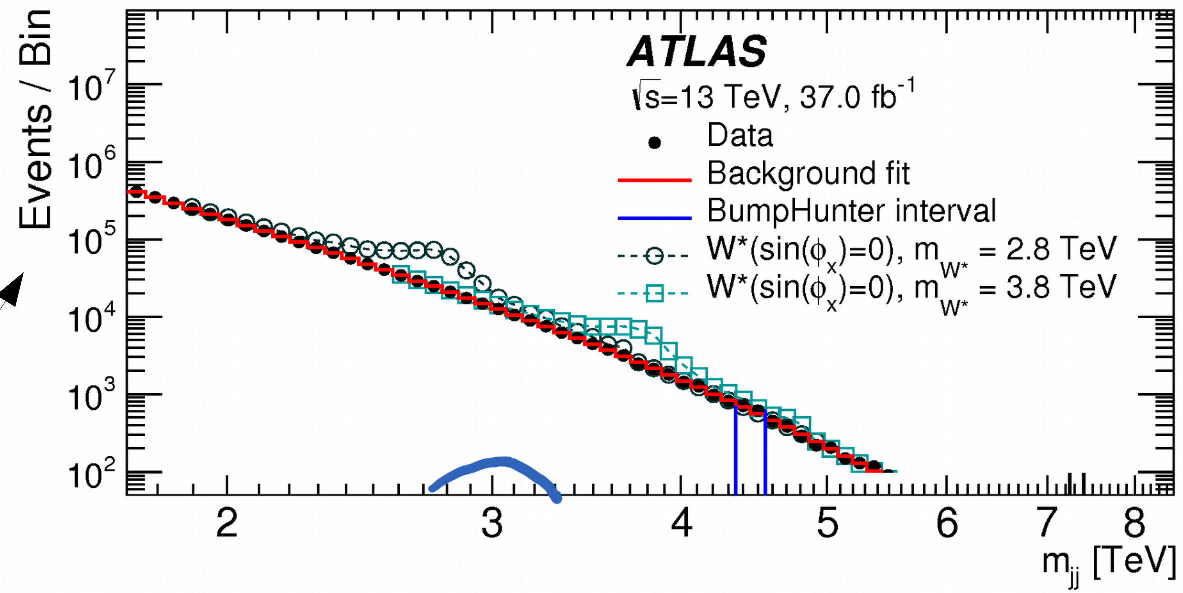
Dijet Resonances



Edited from [1703.01927]

– Samples 1 and 2 contain different fractions of S and B
– S in sample 1 is drawn from the same distribution as S in sample 2

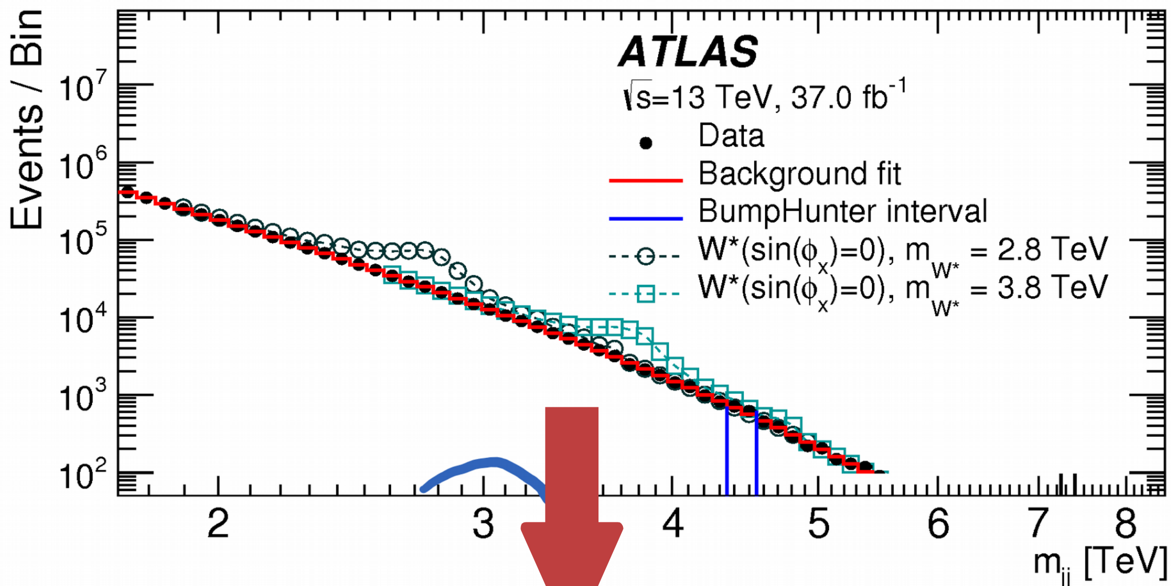
Dijet Resonances



Edited from [1703.01927]

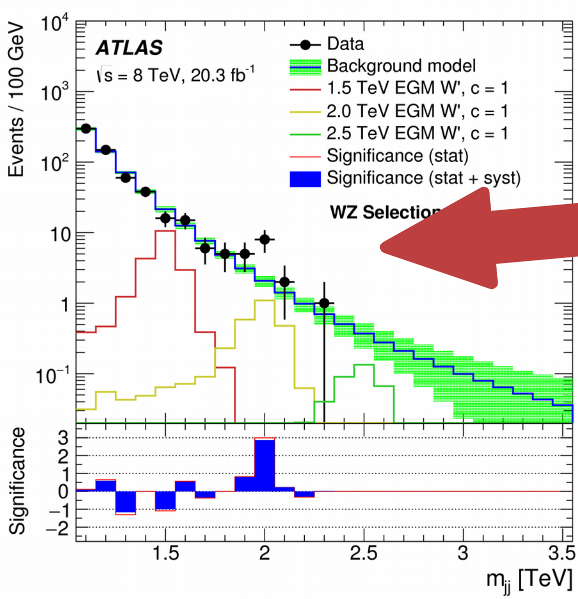
– *Training statistics are sufficiently large*

Dijet Resonances



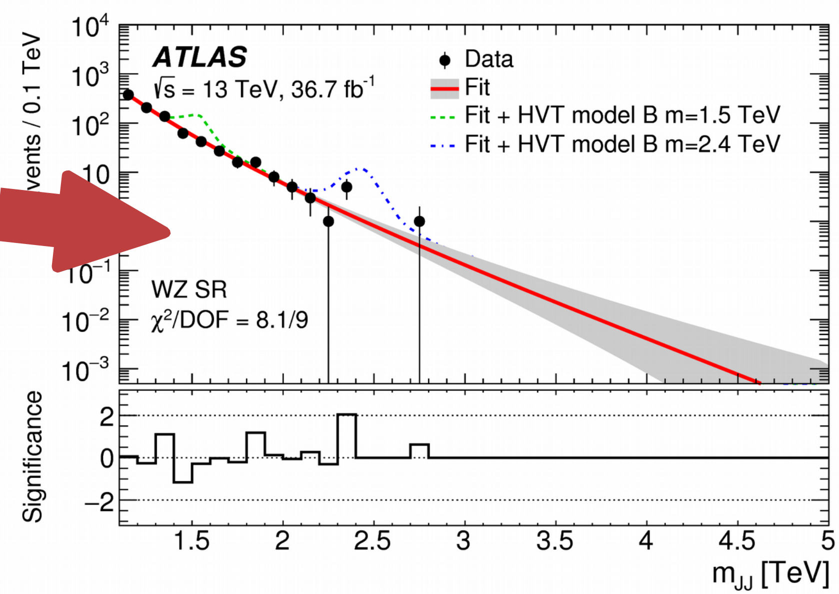
Edited from [1703.01927]

[1506.00962]

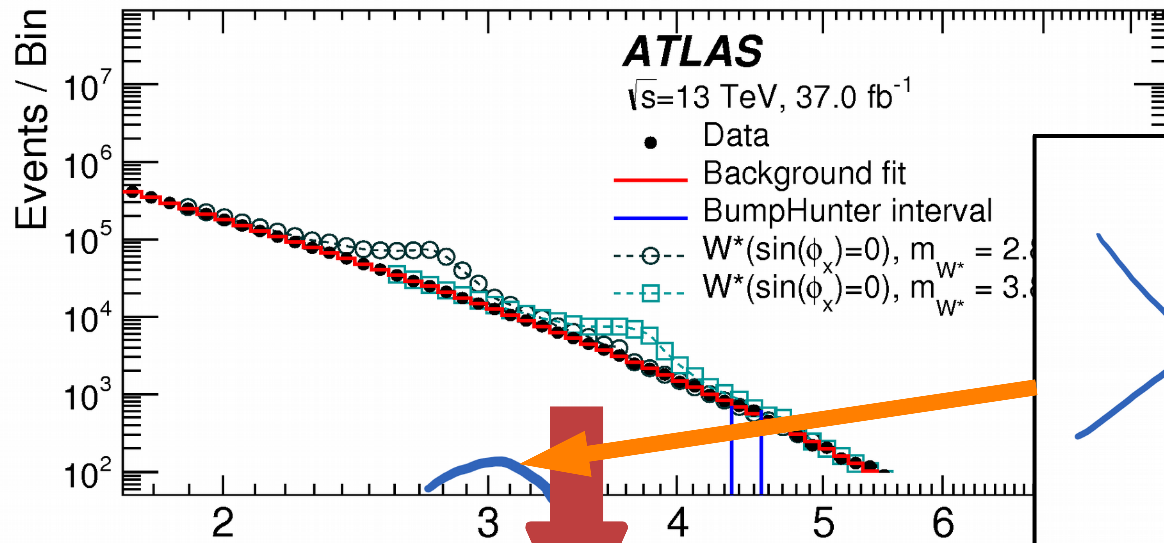


1) Theorist comes up with specific model with some specific prediction (e.g. W’ → WZ).
2) Choose dedicated substructure variables.
3) Simulate signal to optimize cuts
4) Calibrate in some data sample
5) Apply cuts to events and look for a bump in the new distribution

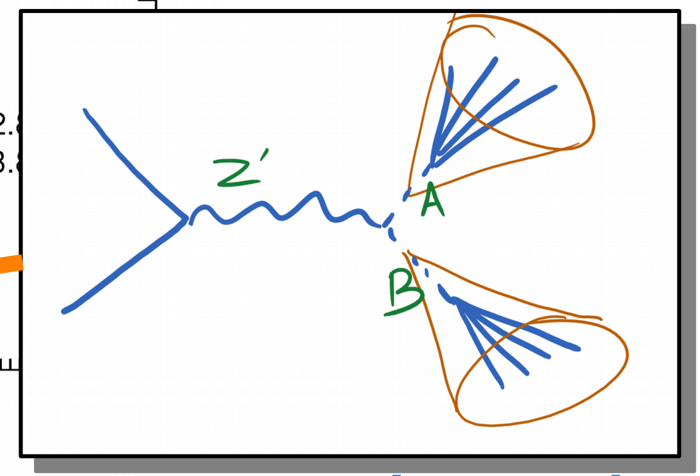
[1708.04445]



Dijet Resonances

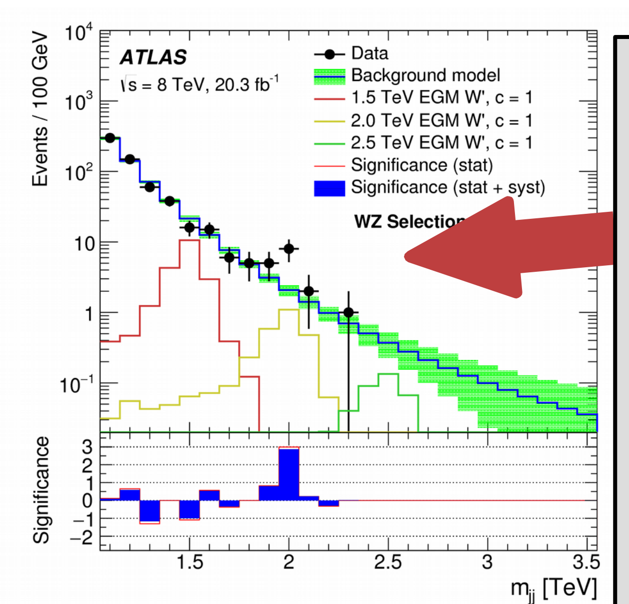


Edited from [1703.01927]



[1708.04445]

[1506.00962]



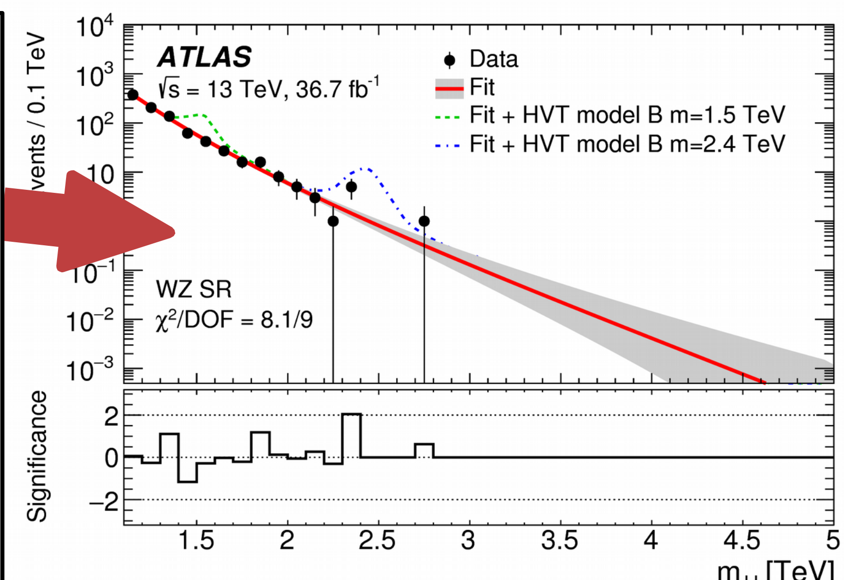
1) Theorist comes up with specific model with some specific prediction (e.g. W' → WZ).

2) Choose dedicated substructure variables.

3) Simulate signal to optimize cuts

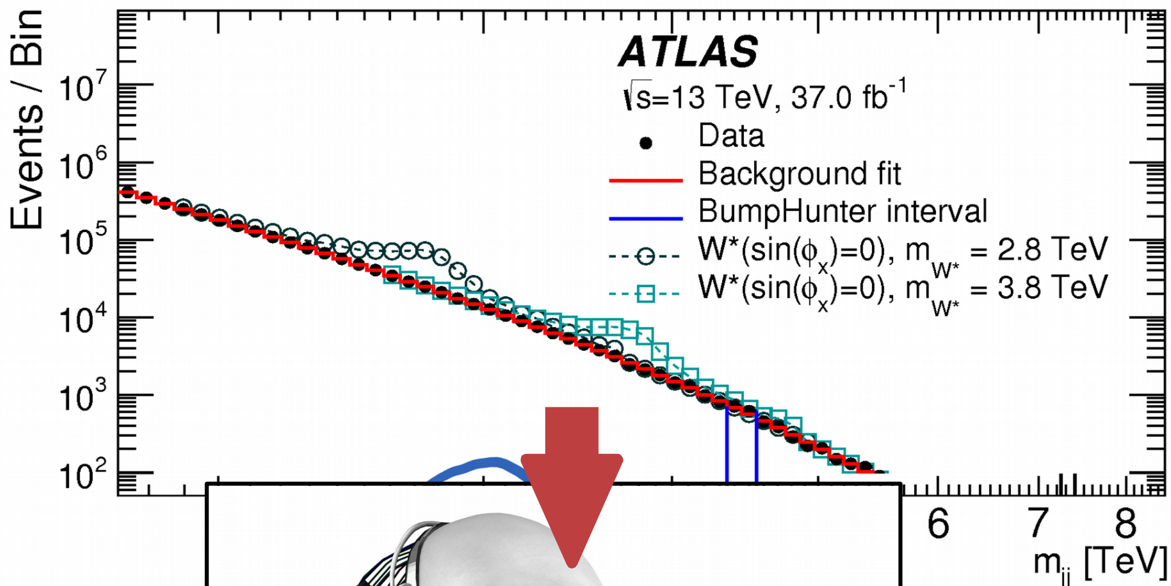
4) Calibrate in some data sample

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

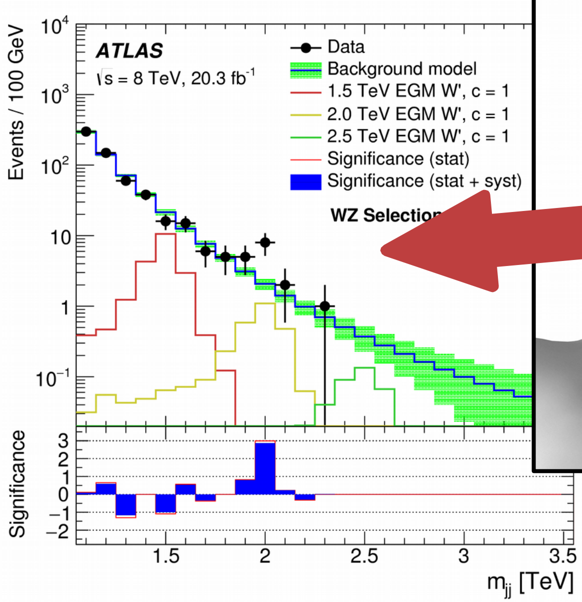


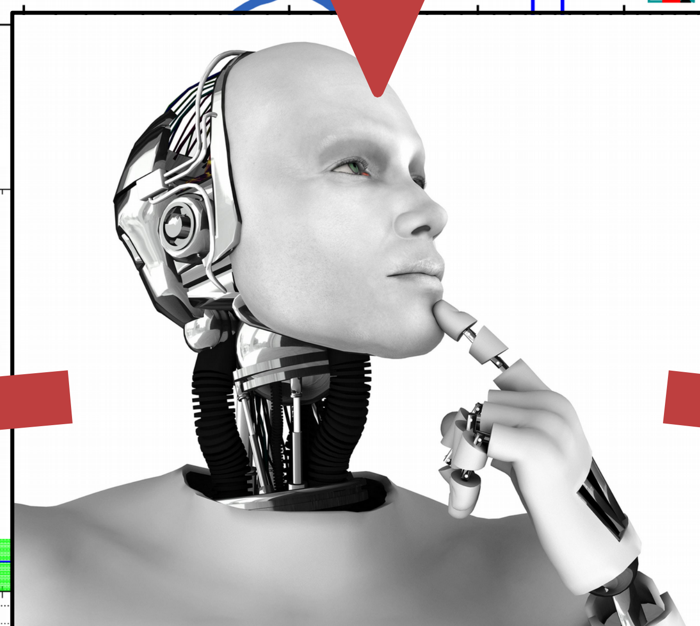
Dijet Resonances

Edited from [1703.01927]

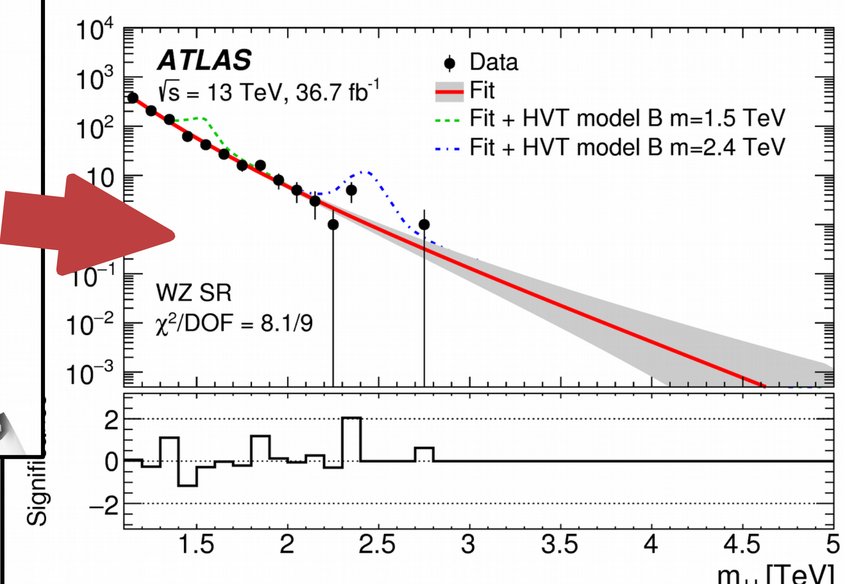


[1506.00962]



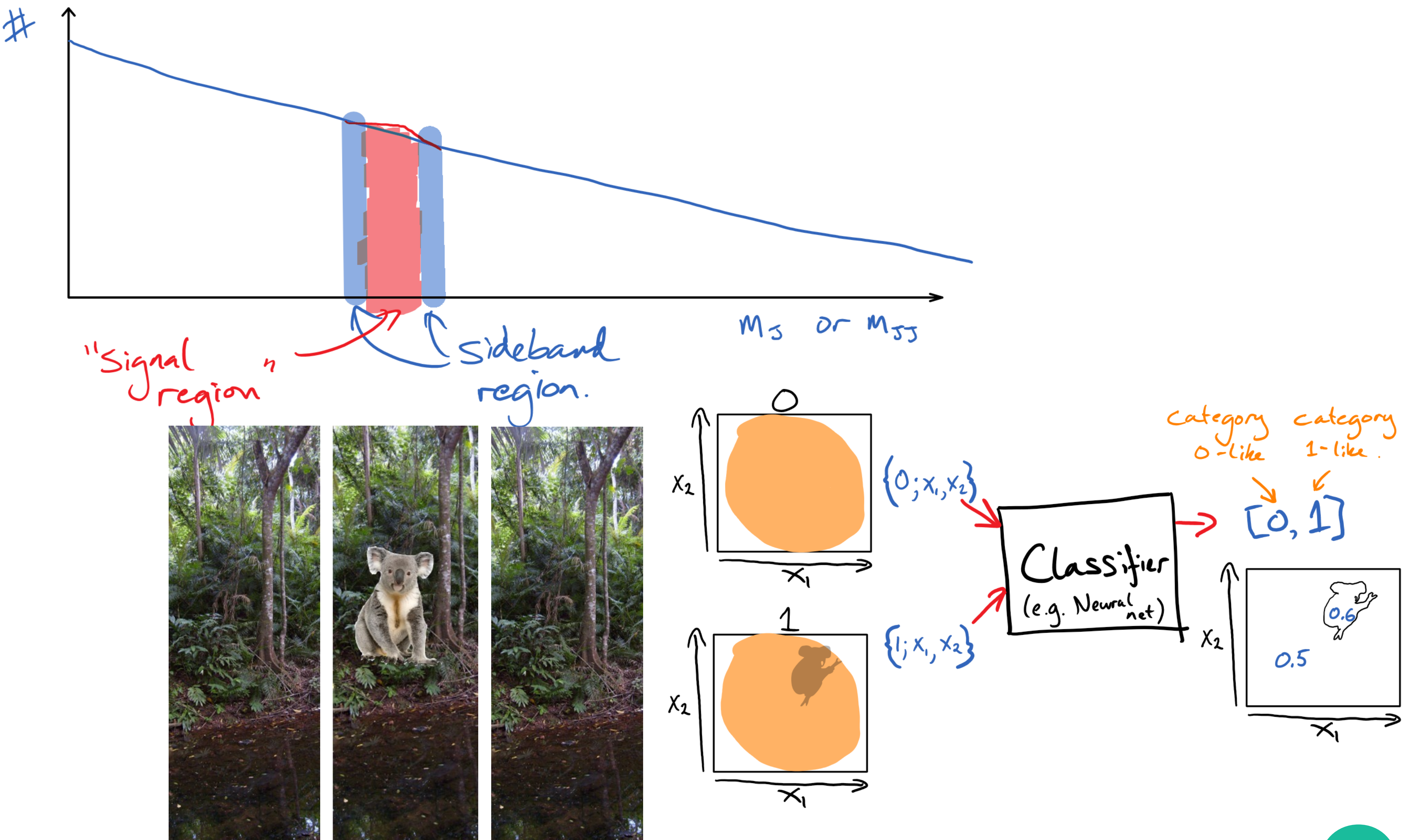


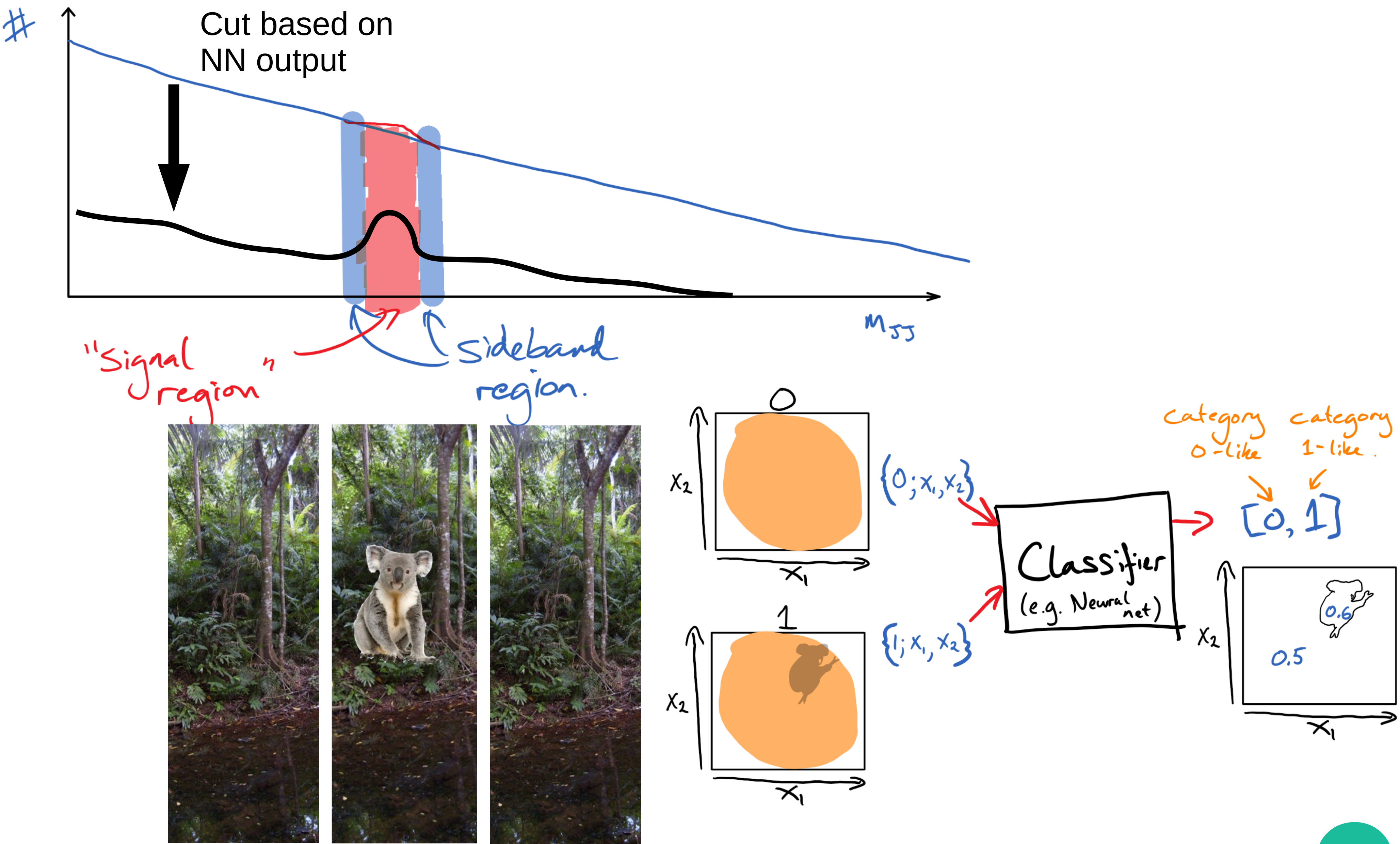
[1708.04445]



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

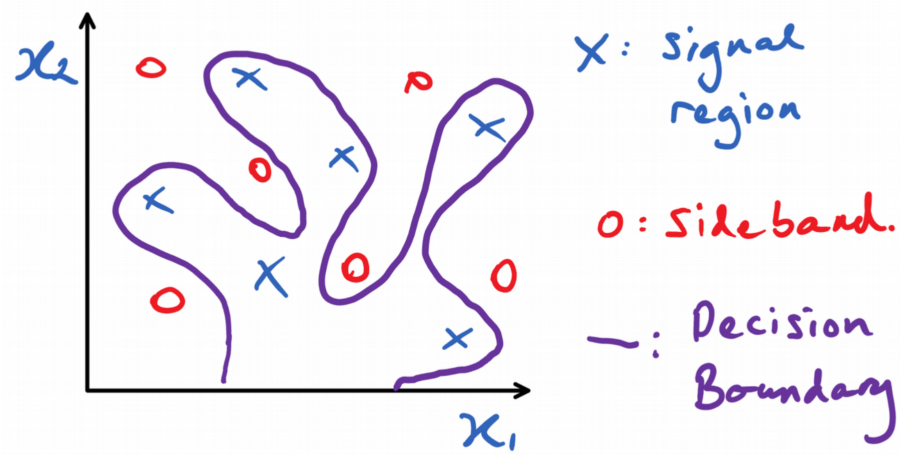
CwoLa Hunting: Basic Picture





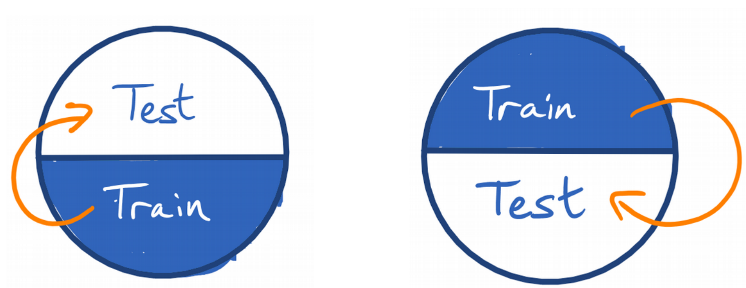
#
Cut based on
NN output
m_{JJ}
"Signal region"
sideband region.
0
x_2
x_1
1
$\{0; x_1, x_2\}$
$\{1; x_1, x_2\}$
Classifier
(e.g. Neural net)
category 0-like
category 1-like.
[0, 1]
0.6
0.5

Avoiding Overfitting



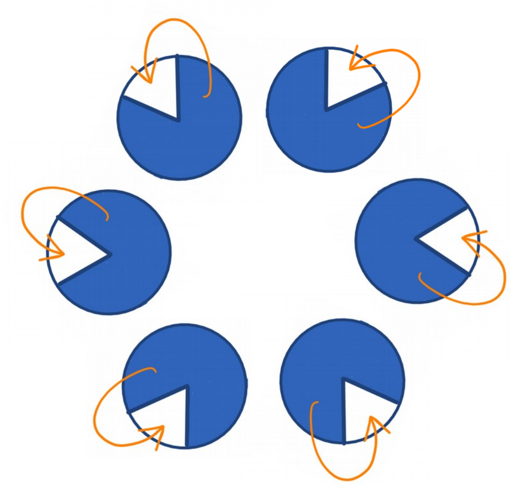
Train/Test Split

(and 'waste' half of your dataset)



Cross Validation

(and introduce new complications in statistical analysis)



Nested Cross-Validation

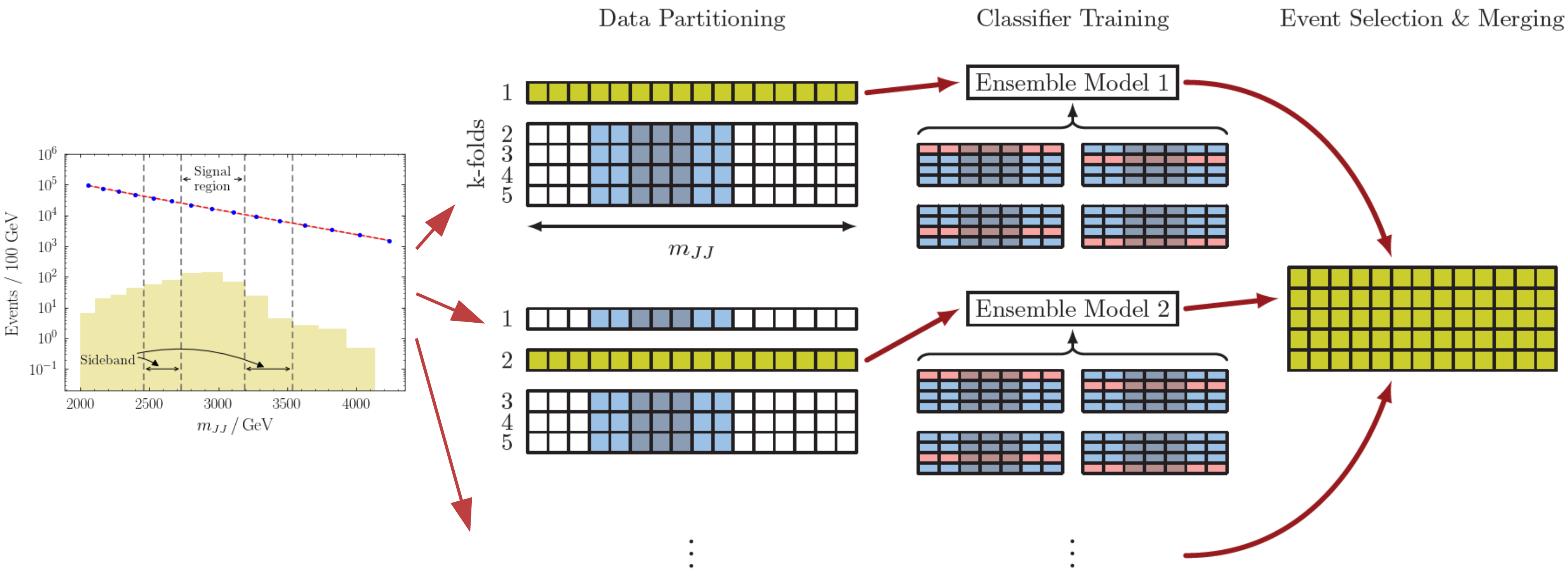
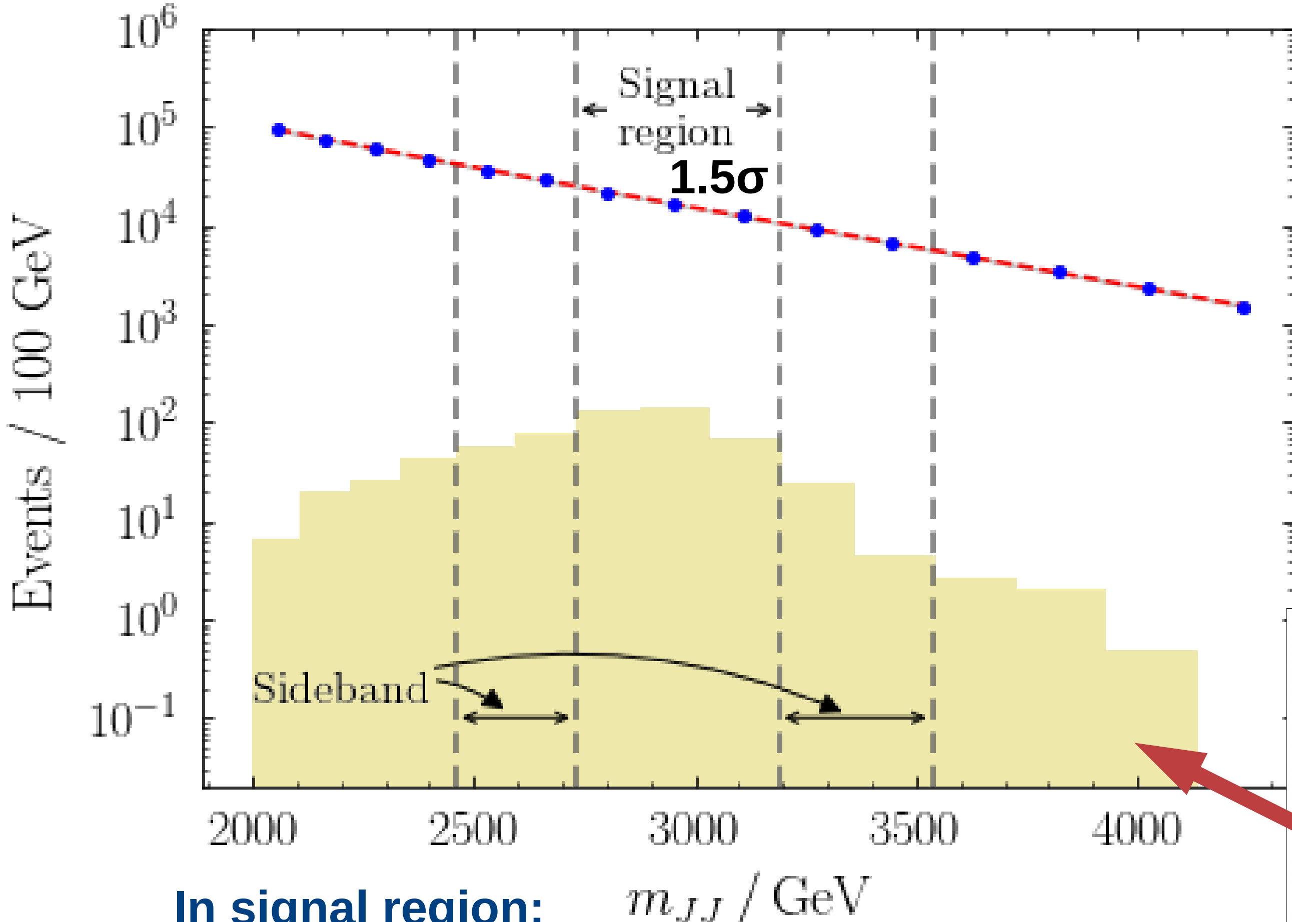


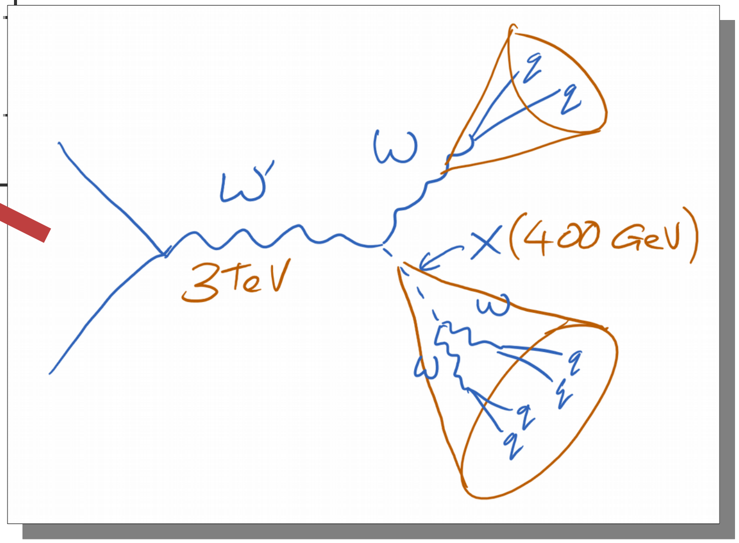
Figure 7. Illustration of the nested cross-validation procedure. **Left:** the dataset is randomly partitioned bin-by-bin into five groups. **Center:** for each group, an ensemble classifier is trained on the remaining groups. For each of the four possible combinations of these four groups into three training groups and one validation group, a set of invidual 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.

Application to Bump Hunt

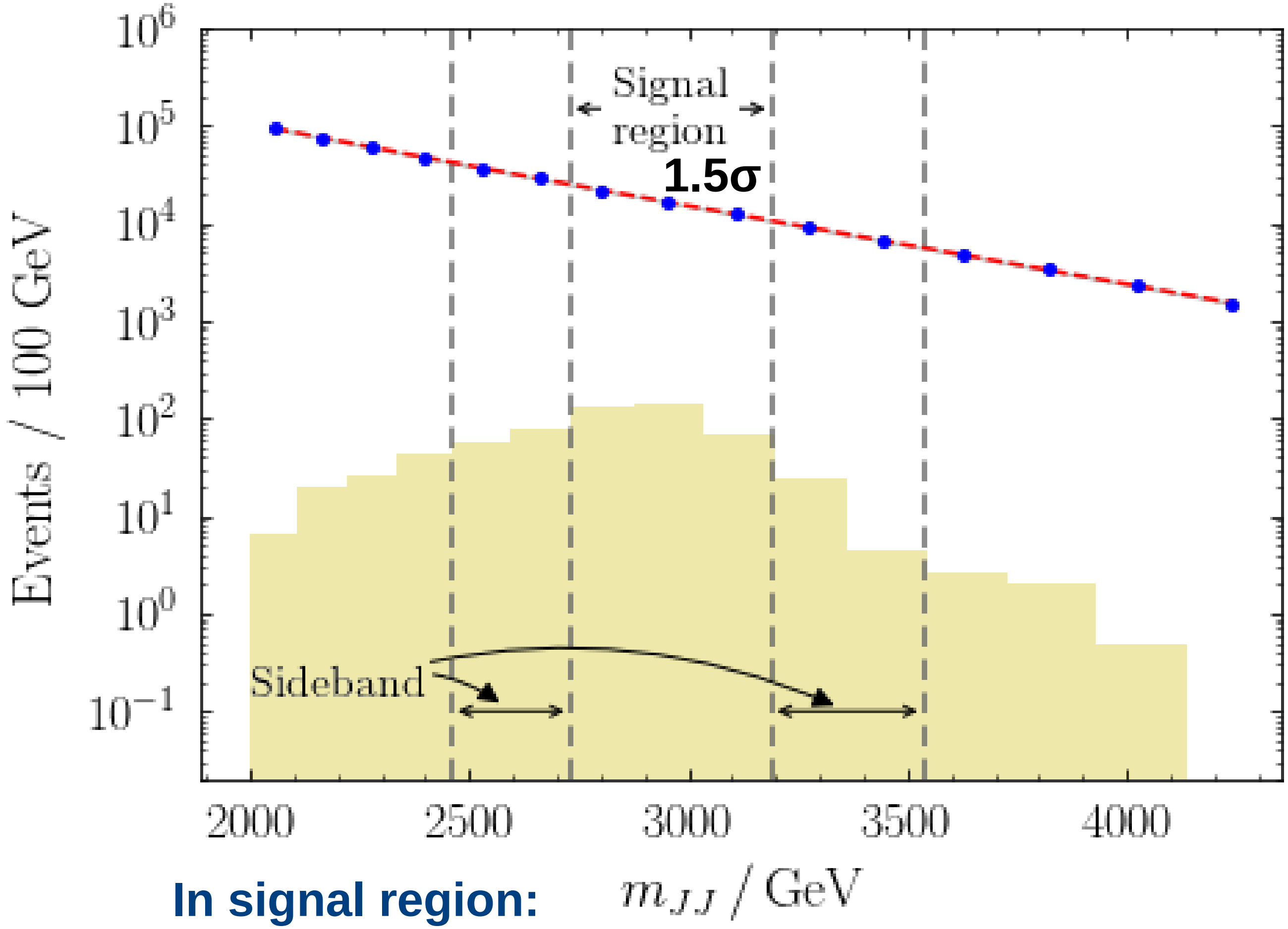


$$\frac{dN}{dm_{JJ}} = p_0 \frac{(1 - m_{JJ}/\sqrt{s})^{p_1}}{(m_{JJ}/\sqrt{s})^{p_2}}$$

In signal region:
S = 522,
S/B = 0.64%

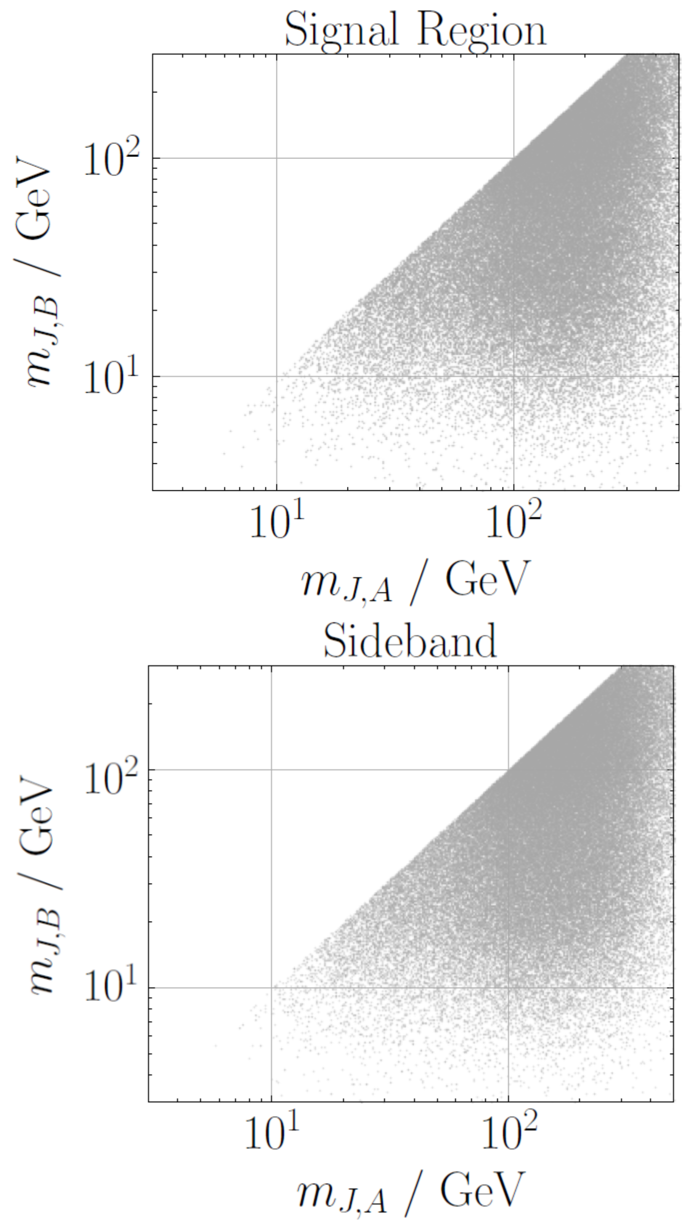


Application to Bump Hunt

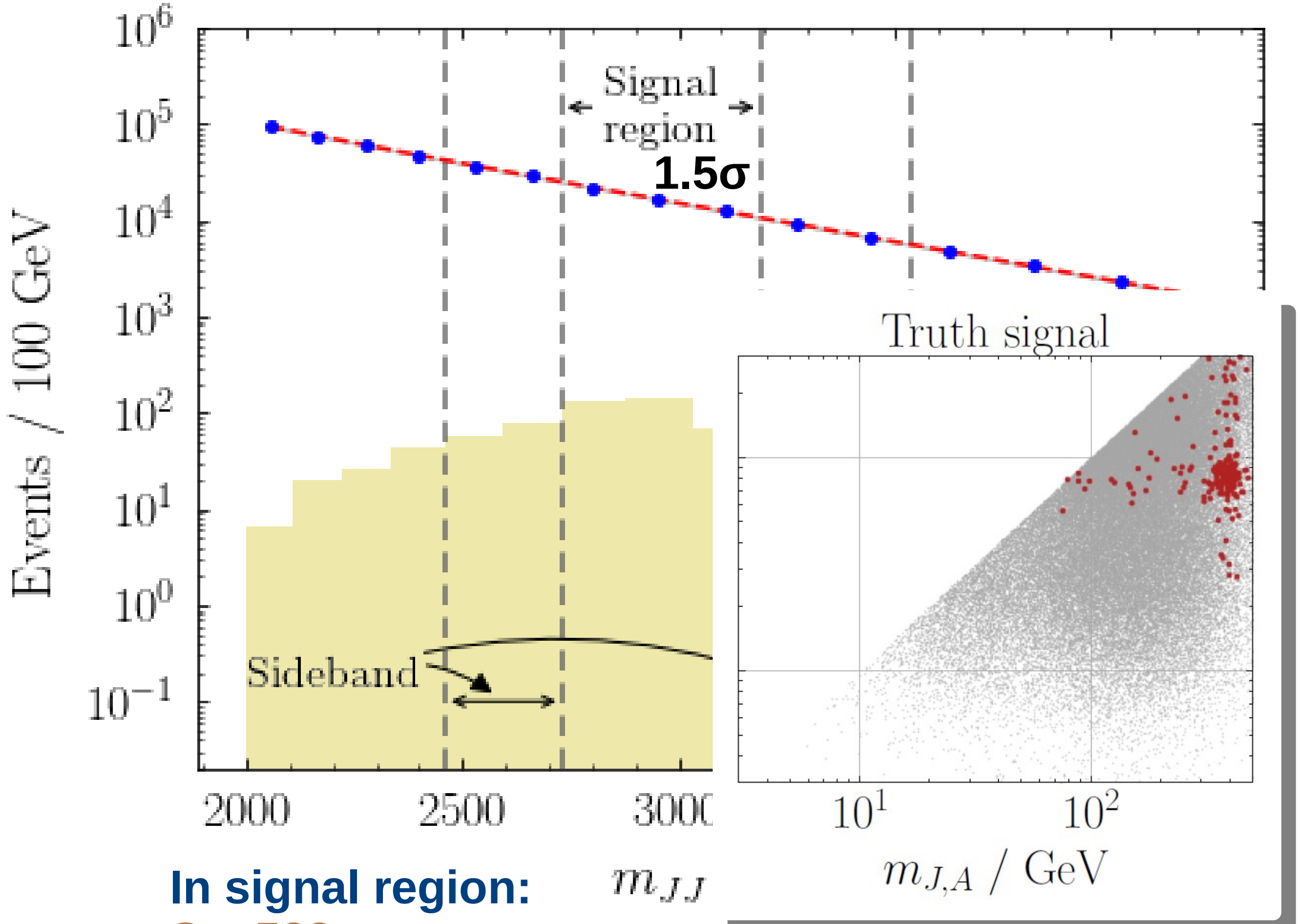


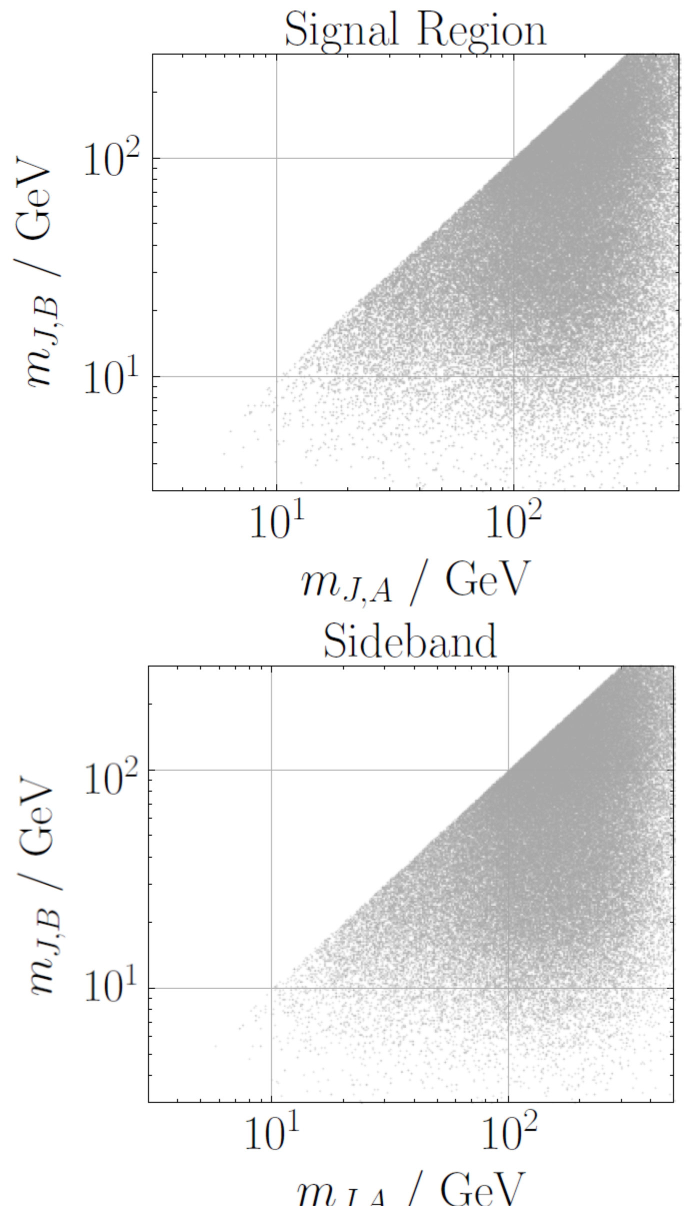
In signal region:
S = 522,
S/B = 0.64%

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



Application to Bump Hunt

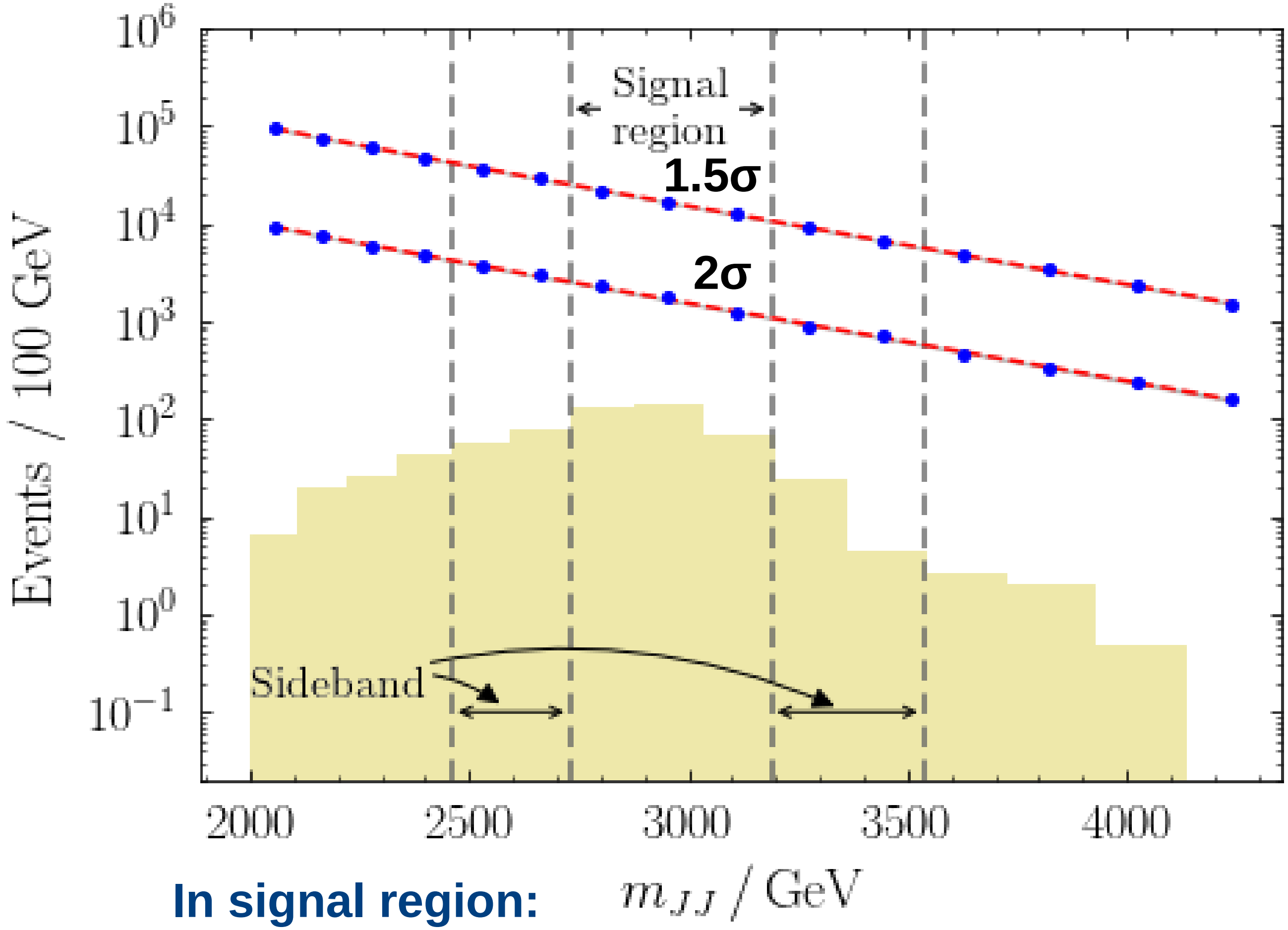


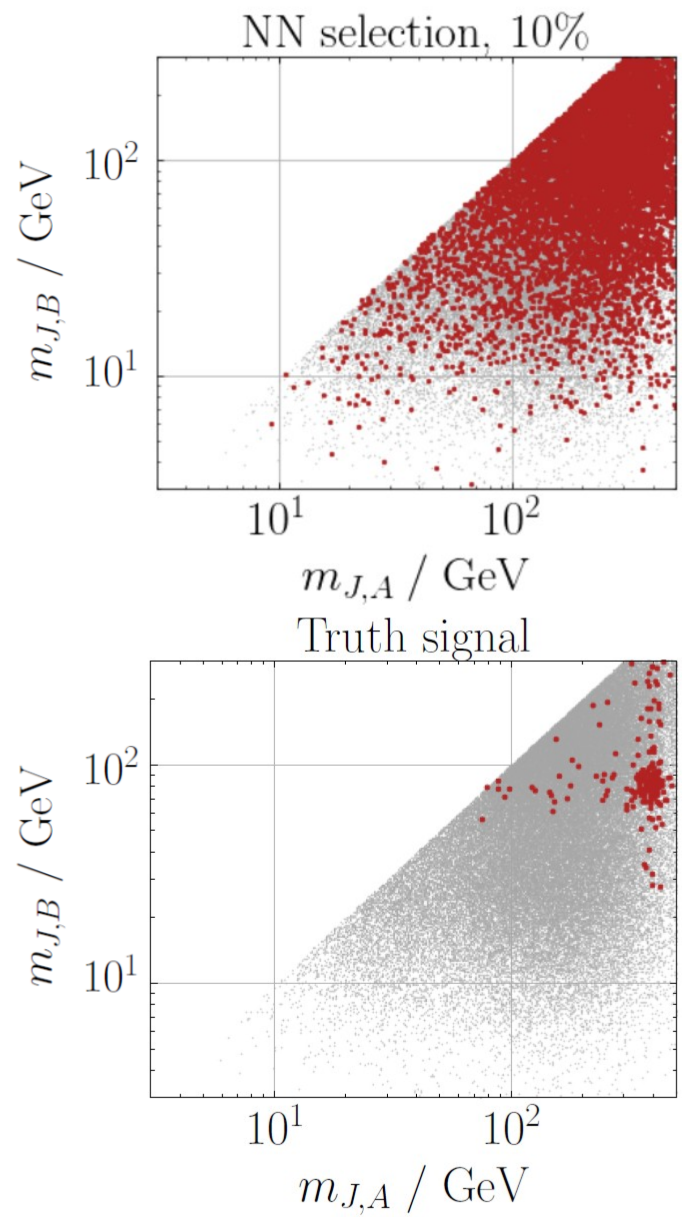


In signal region:
S = 522,
S/B = 0.64%

For each jet: $$Y_i = \left(m_J,\ \sqrt{\tau_1^{(2)}}/\tau_1^{(1)},\ \tau_{21},\ \tau_{32},\ \tau_{43},\ n_{\mathrm{trk}} \right)$$

Application to Bump Hunt

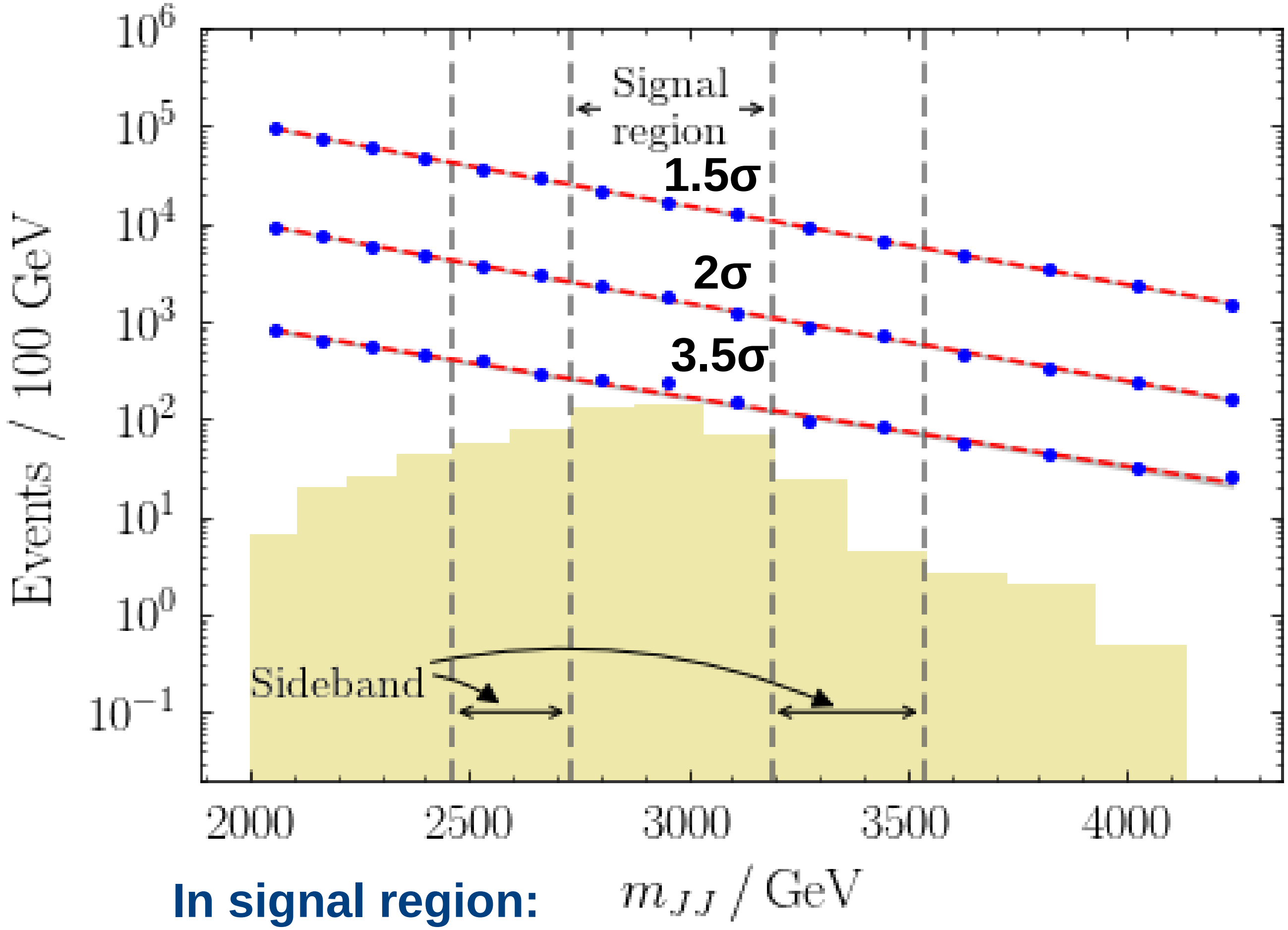


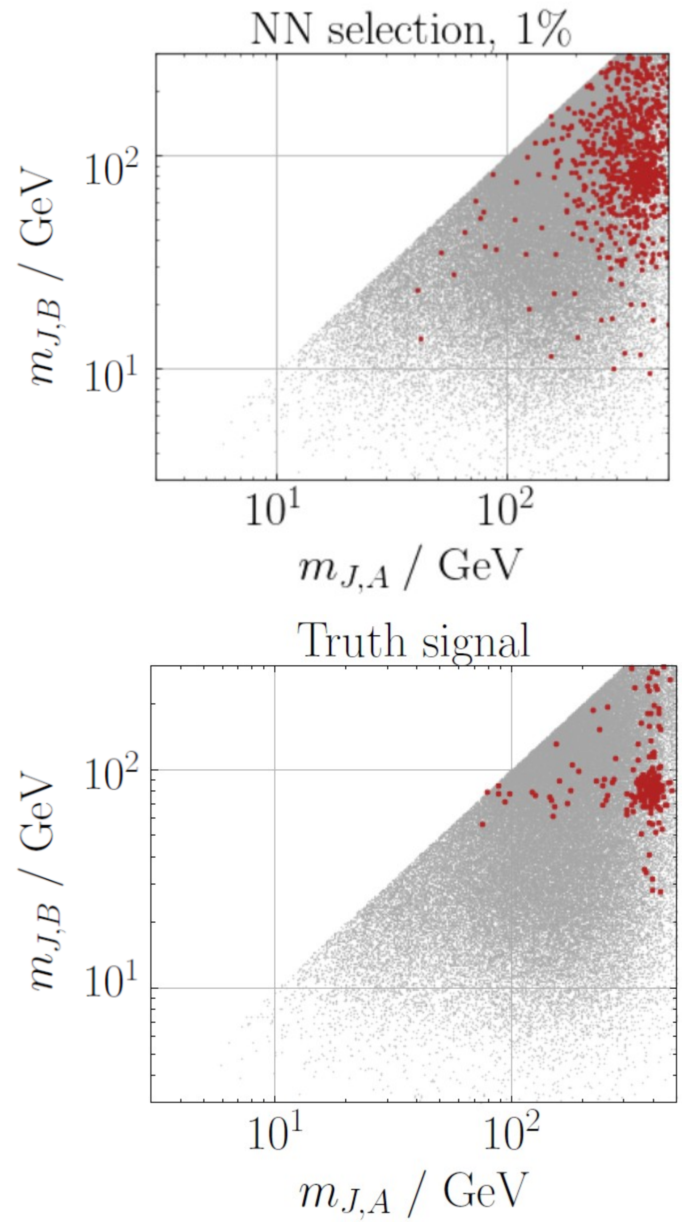


In signal region:
S = 522,
S/B = 0.64%

For each jet: $Y_i = \left(m_J,\ \sqrt{\tau_1^{(2)}}/\tau_1^{(1)},\ \tau_{21},\ \tau_{32},\ \tau_{43},\ n_{\mathrm{trk}} \right)$

Application to Bump Hunt

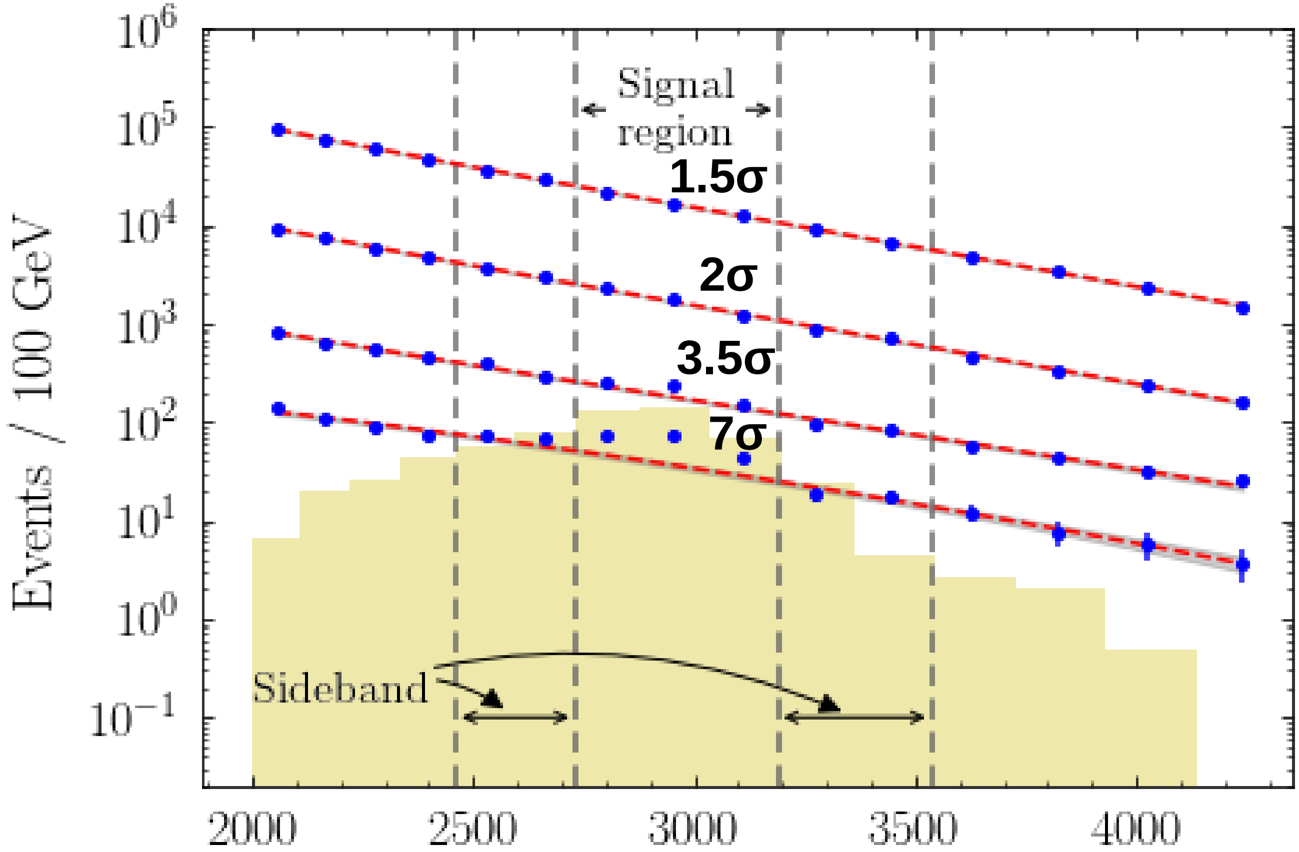


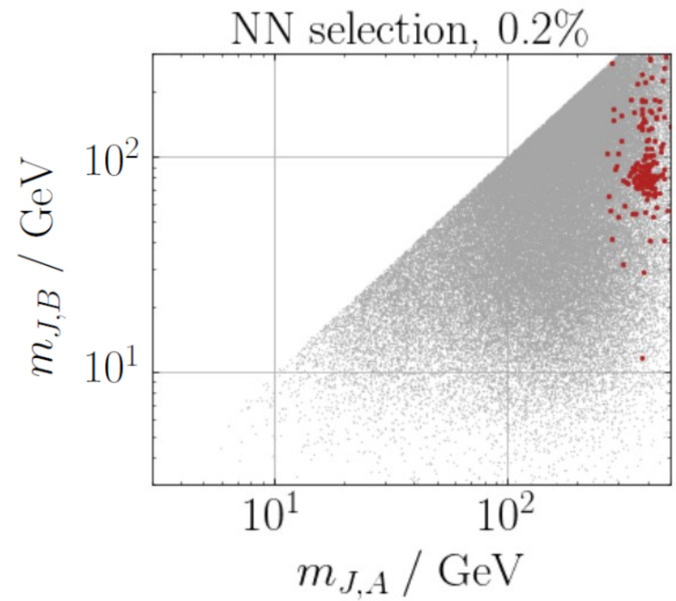


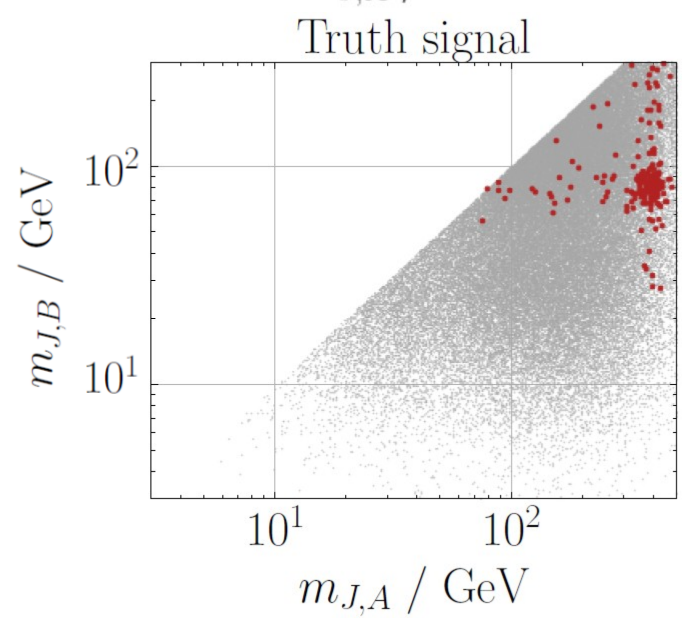
In signal region:
S = 522,
S/B = 0.64%

For each jet: $$Y_i = \left(m_J,\ \sqrt{\tau_1^{(2)}}/\tau_1^{(1)},\ \tau_{21},\ \tau_{32},\ \tau_{43},\ n_{\mathrm{trk}}\right)$$

Application to Bump Hunt



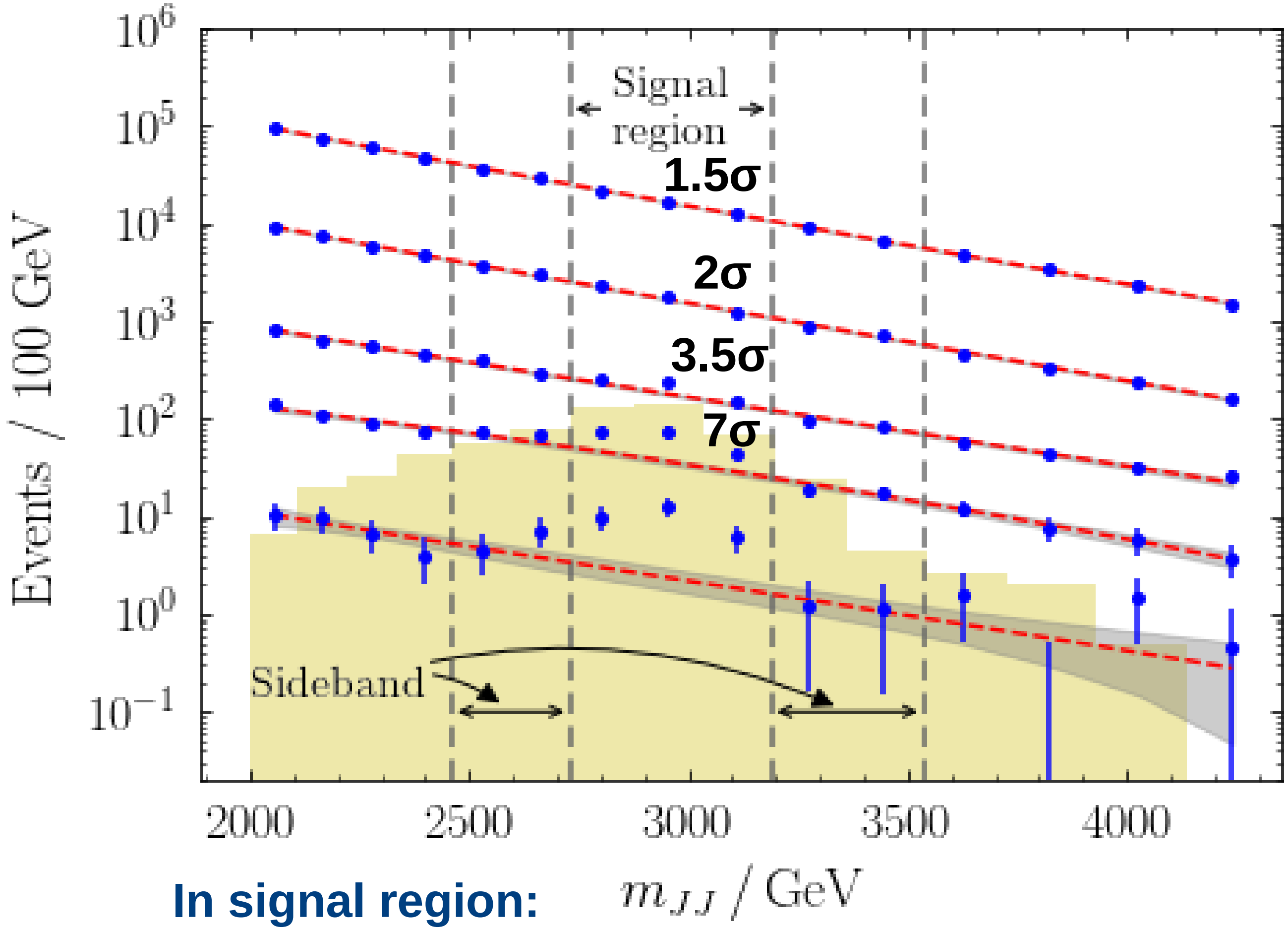


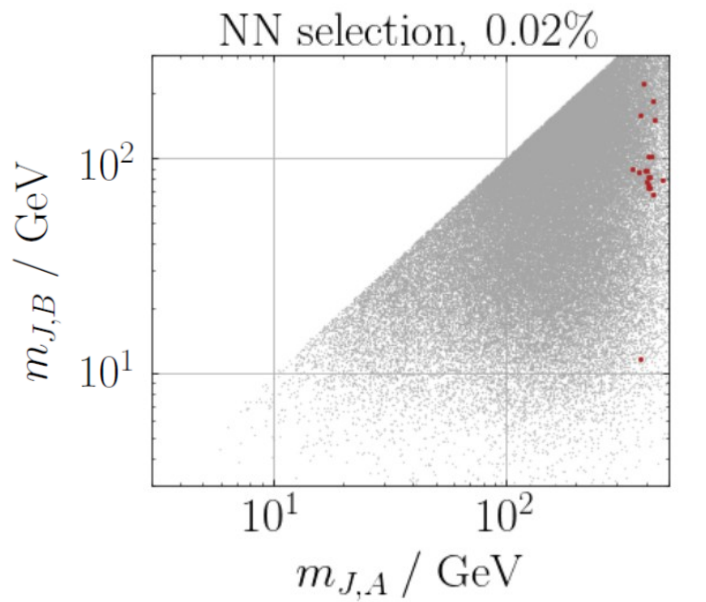


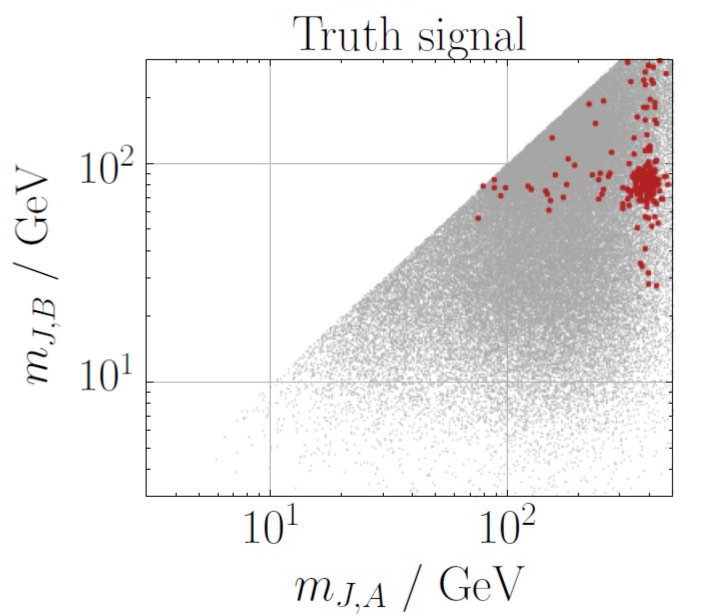
In signal region:
S = 522,
S/B = 0.64%

For each jet: $$Y_i = \left(m_J,\ \sqrt{\tau_1^{(2)}}/\tau_1^{(1)},\ \tau_{21},\ \tau_{32},\ \tau_{43},\ n_{\mathrm{trk}} \right)$$

Application to Bump Hunt







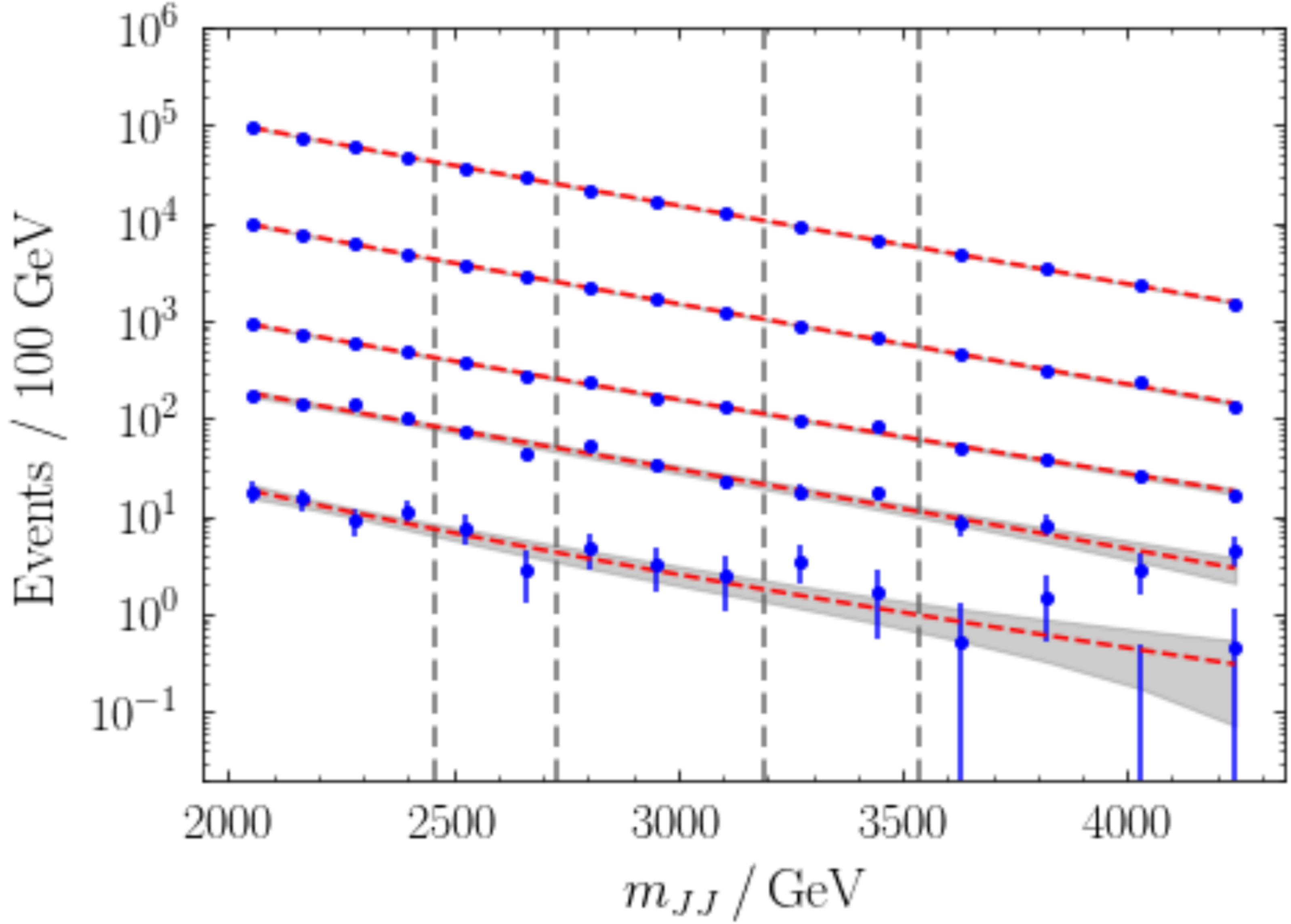
In signal region:

S = 522,

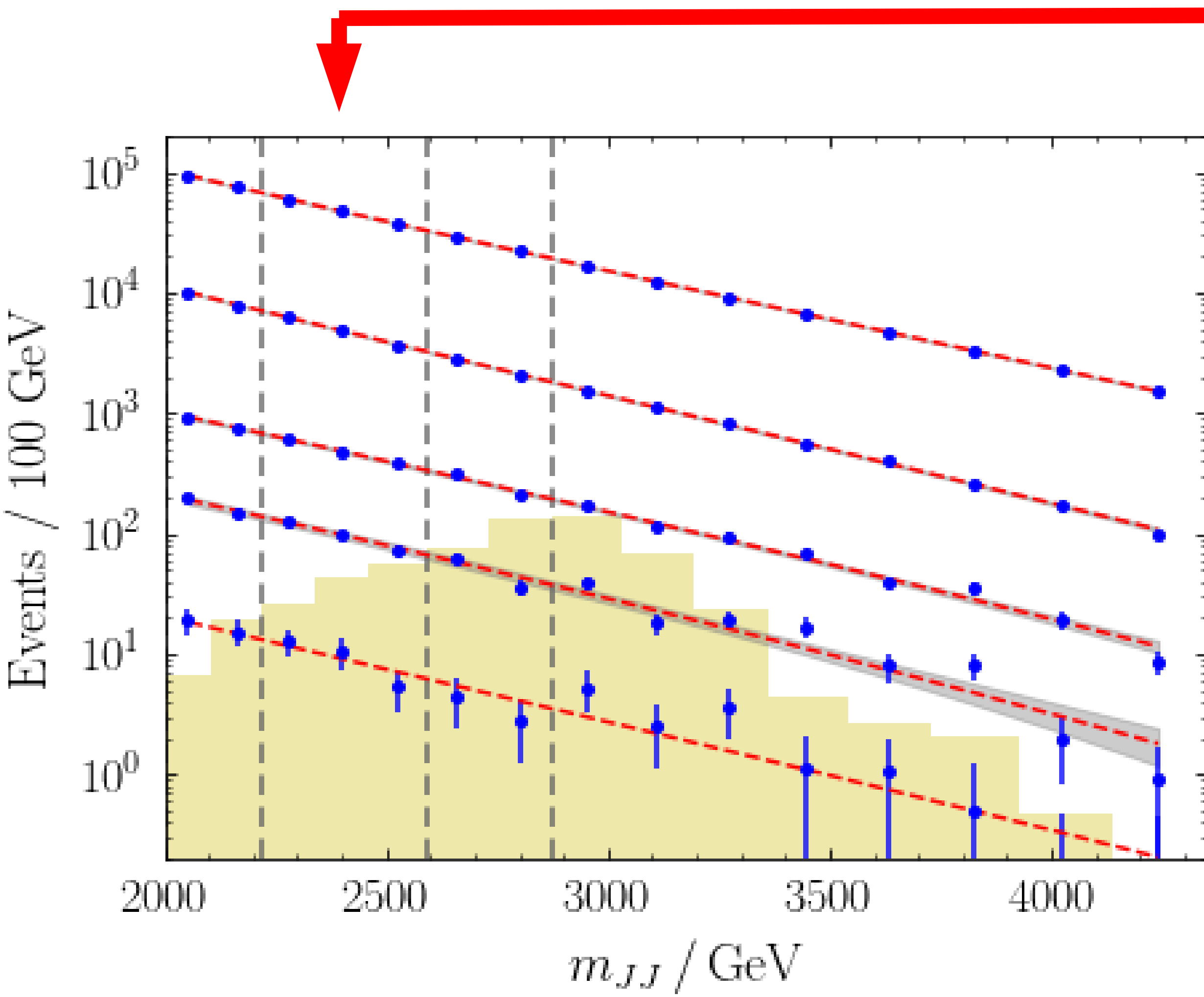
S/B = 0.64%

For each jet: $$Y_i = \left(m_J,\ \sqrt{\tau_1^{(2)}}/\tau_1^{(1)},\ \tau_{21},\ \tau_{32},\ \tau_{43},\ n_{\mathrm{trk}}\right)$$

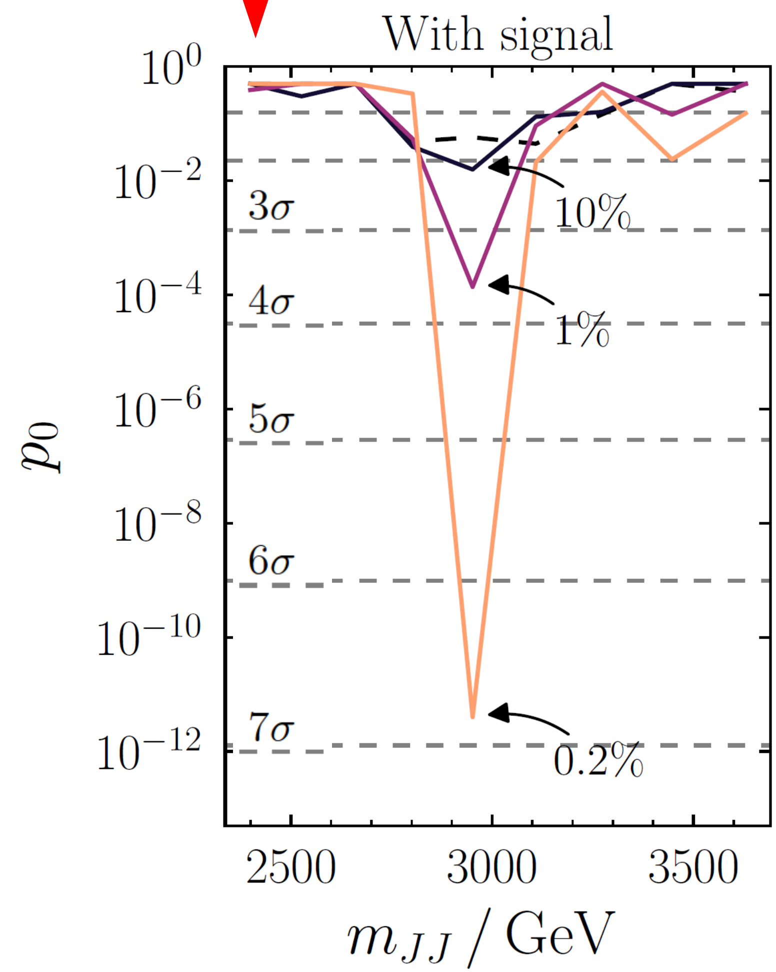
No Signal → No Bump!



Mass Scan

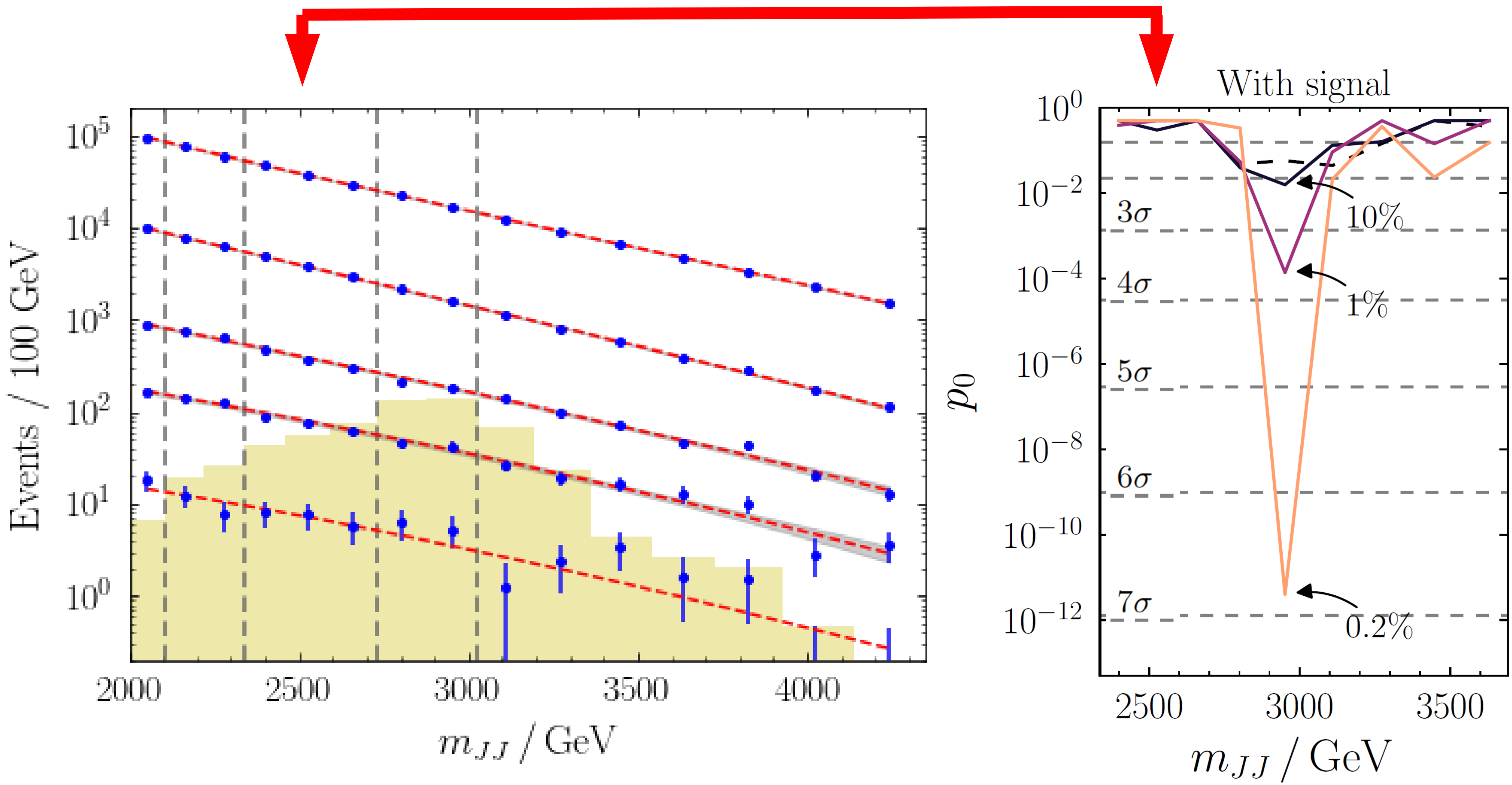


10^5
10^4
10^3
10^2
10^1
10^0
Events / 100 GeV
2000
2500
3000
3500
4000
m_{JJ} / GeV



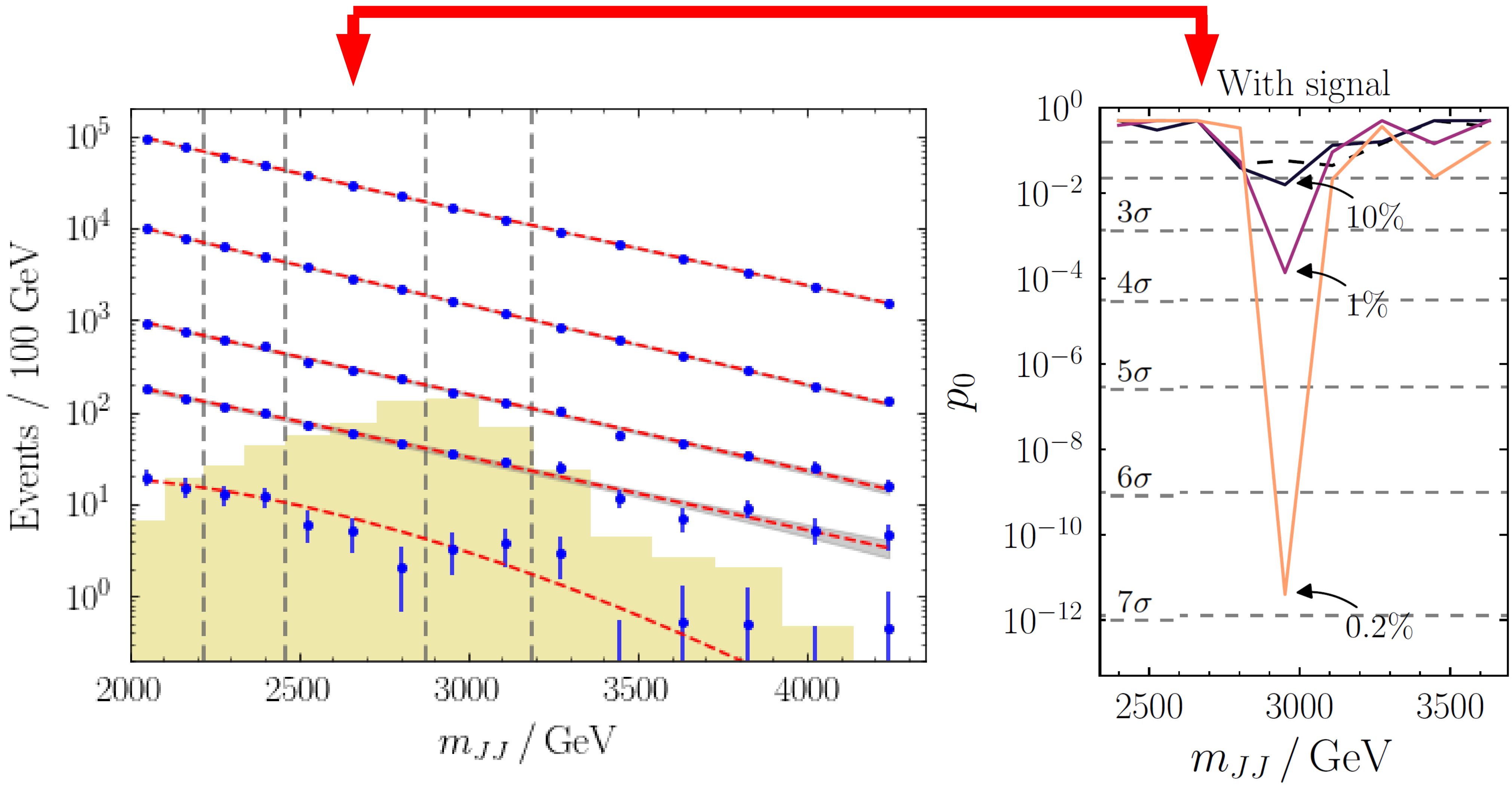
With signal
10^0
10^{-2}
10^{-4}
10^{-6}
10^{-8}
10^{-10}
10^{-12}
p_0
3σ
4σ
5σ
6σ
7σ
10%
1%
0.2%
2500
3000
3500
m_{JJ} / GeV

Mass Scan



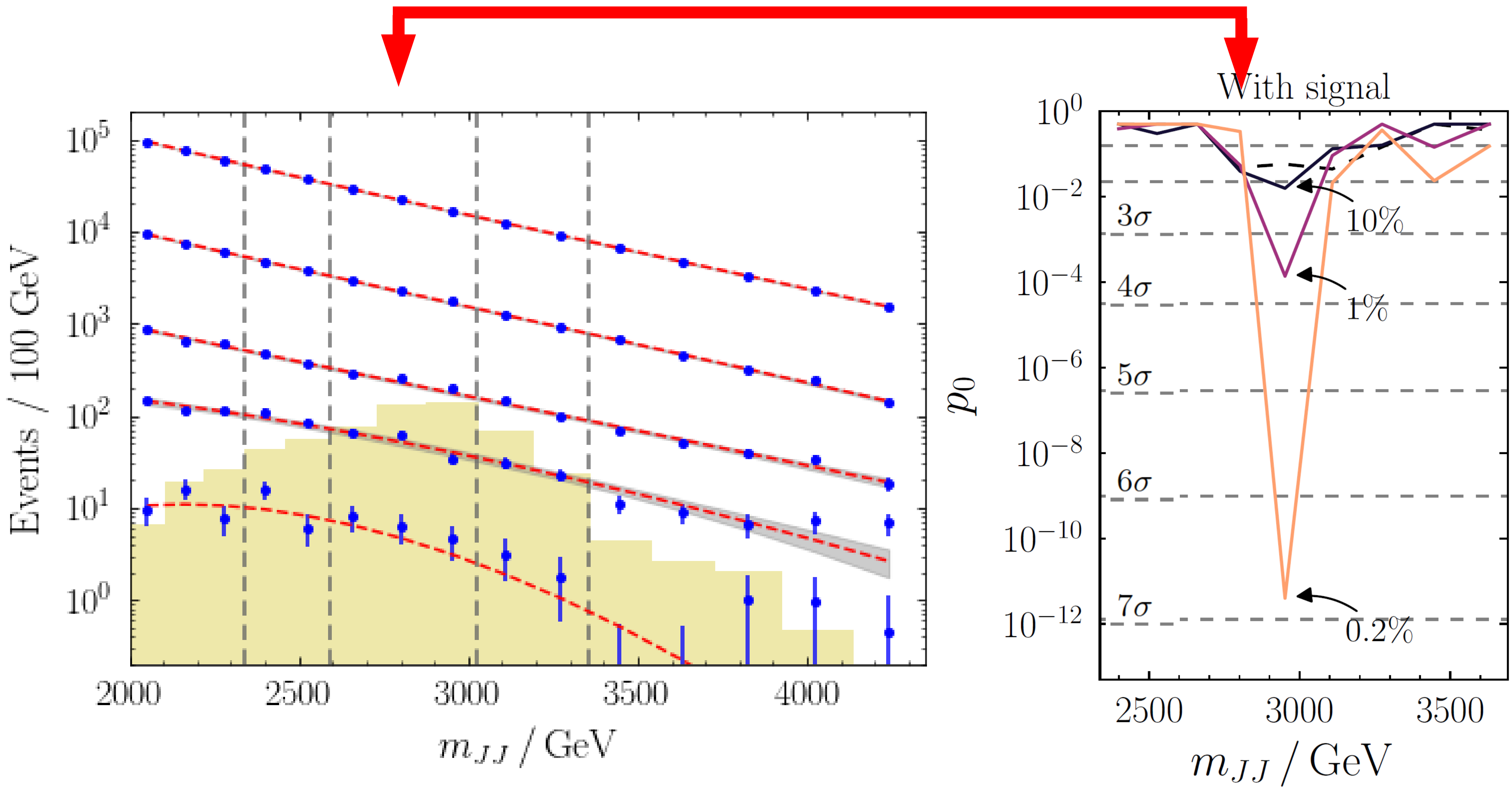
Events / 100 GeV
10^5
10^4
10^3
10^2
10^1
10^0
2000
2500
3000
3500
4000
m_{JJ} / GeV
With signal
p_0
10^0
10^{-2}
10^{-4}
10^{-6}
10^{-8}
10^{-10}
10^{-12}
3σ
4σ
5σ
6σ
7σ
10%
1%
0.2%
2500
3000
3500
m_{JJ} / GeV

Mass Scan



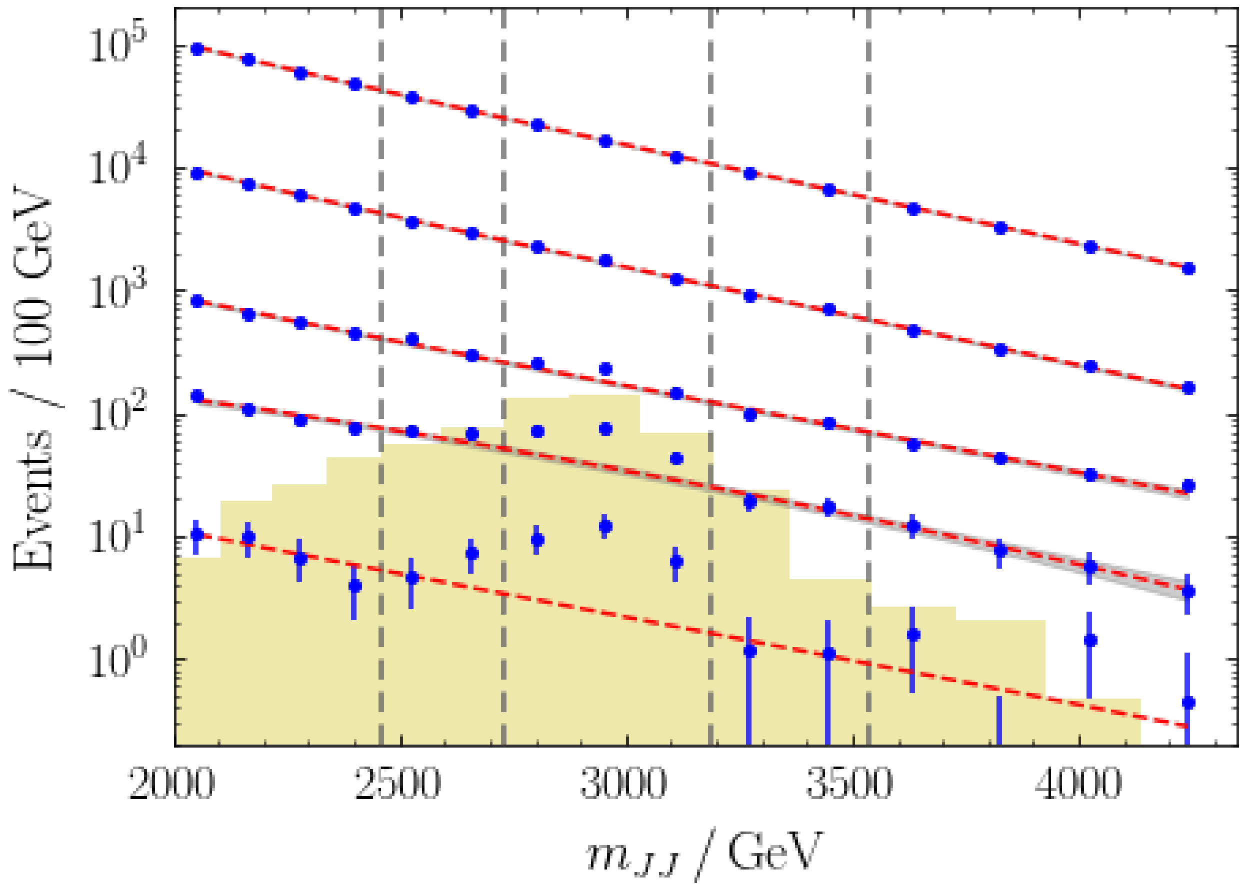
Events / 100 GeV
10^5
10^4
10^3
10^2
10^1
10^0
2000
2500
3000
3500
4000
m_{JJ} / GeV
With signal
p_0
10^0
10^{-2}
10^{-4}
10^{-6}
10^{-8}
10^{-10}
10^{-12}
3σ
4σ
5σ
6σ
7σ
10%
1%
0.2%
2500
3000
3500
m_{JJ} / GeV

Mass Scan

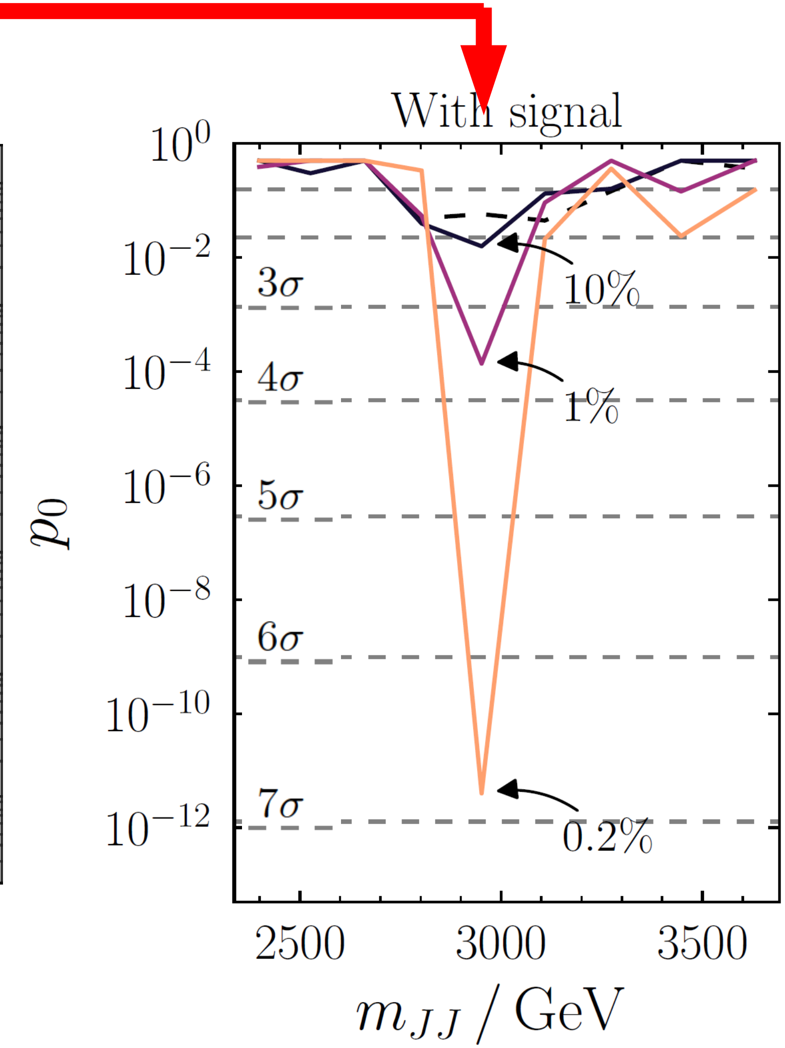


With signal
10^0
10^{-2}
10^{-4}
10^{-6}
10^{-8}
10^{-10}
10^{-12}
p_0
3σ
4σ
5σ
6σ
7σ
10%
1%
0.2%
2500
3000
3500
m_{JJ} / GeV
Events / 100 GeV
10^5
10^4
10^3
10^2
10^1
10^0
2000
4000

Mass Scan

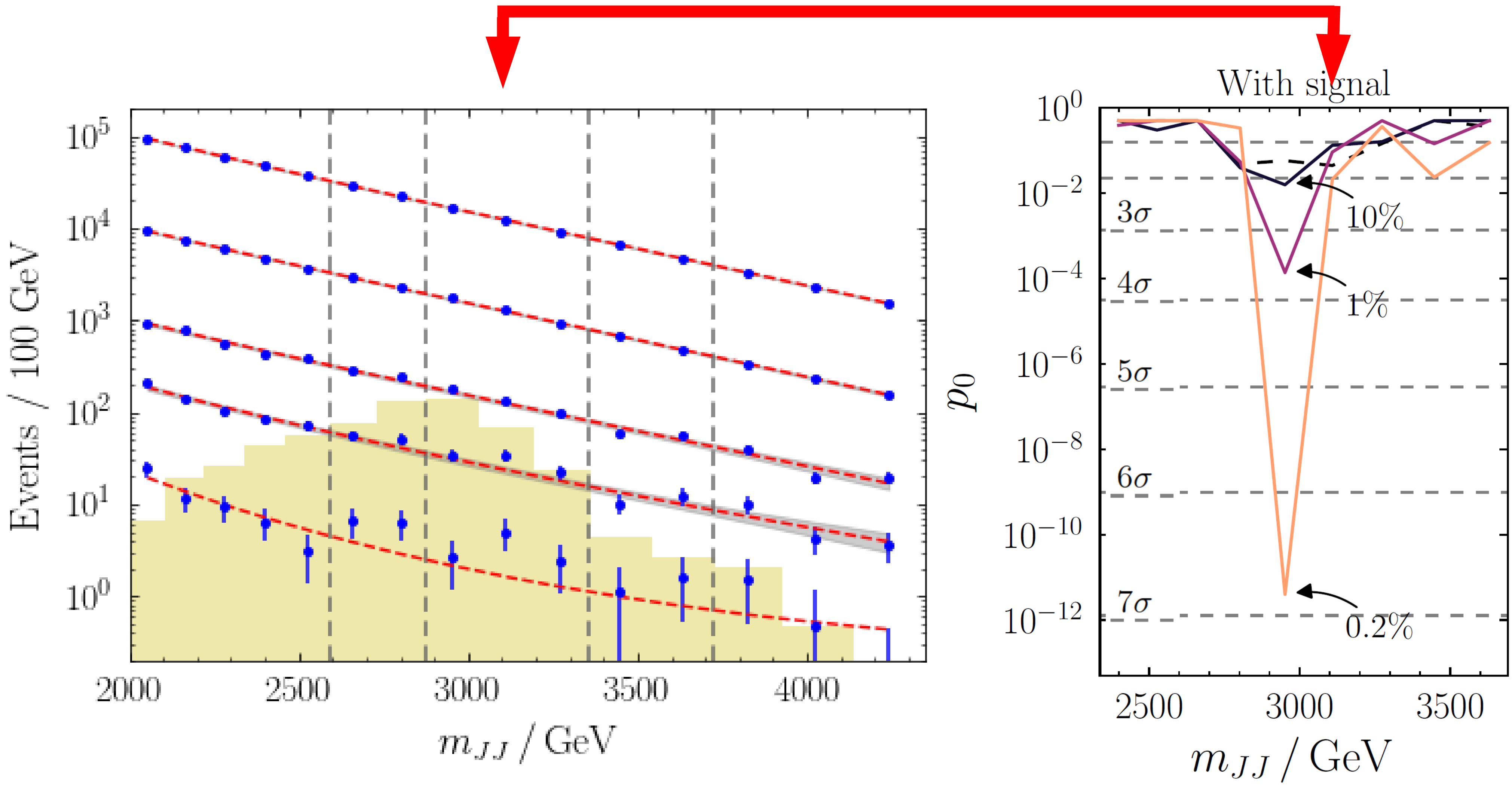


Events / 100 GeV
10^5
10^4
10^3
10^2
10^1
10^0
2000
2500
3000
3500
4000
m_{JJ} / GeV



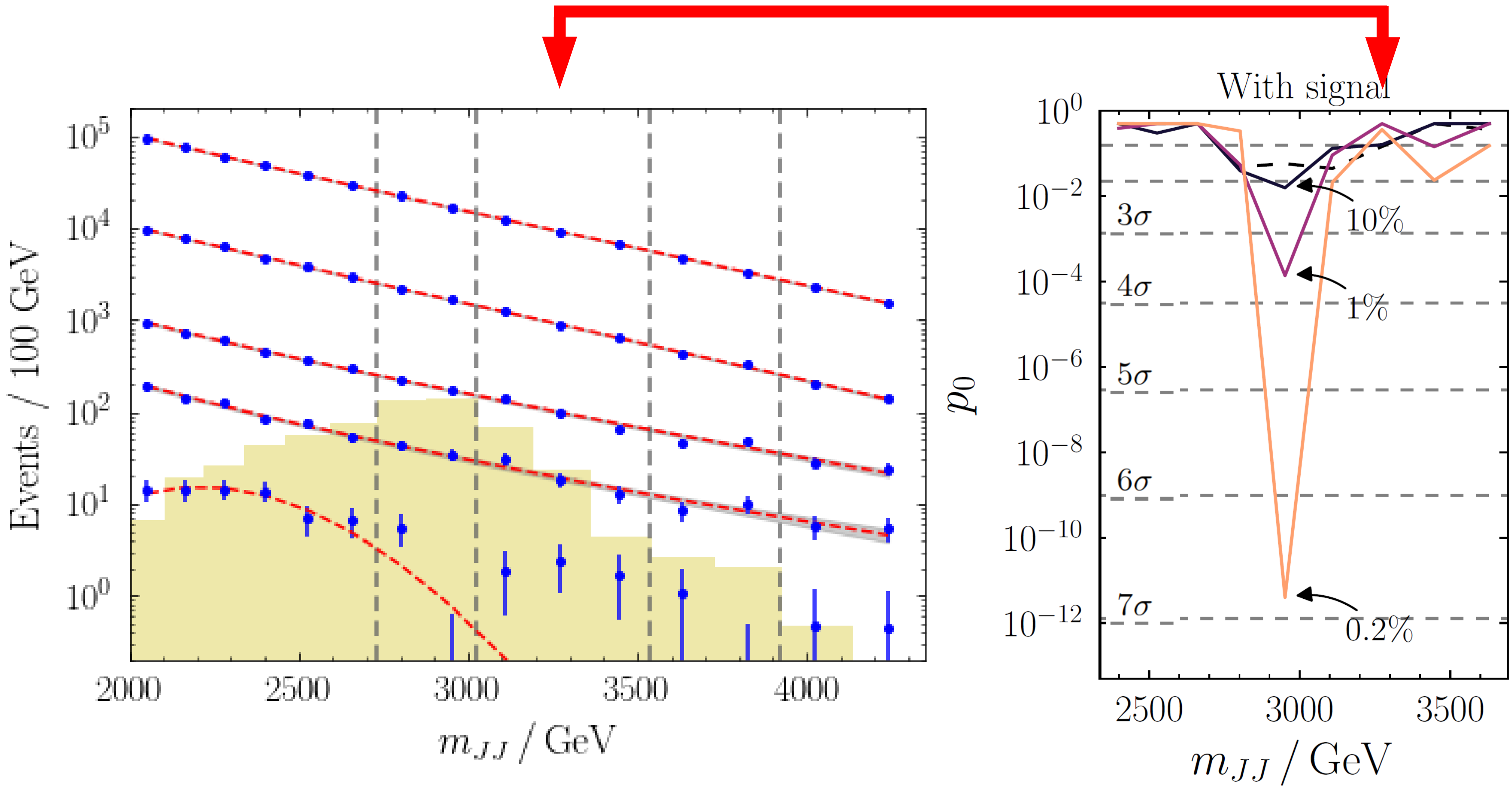
With signal
p_0
10^0
10^{-2}
10^{-4}
10^{-6}
10^{-8}
10^{-10}
10^{-12}
3σ
4σ
5σ
6σ
7σ
10%
1%
0.2%
2500
3000
3500
m_{JJ} / GeV

Mass Scan



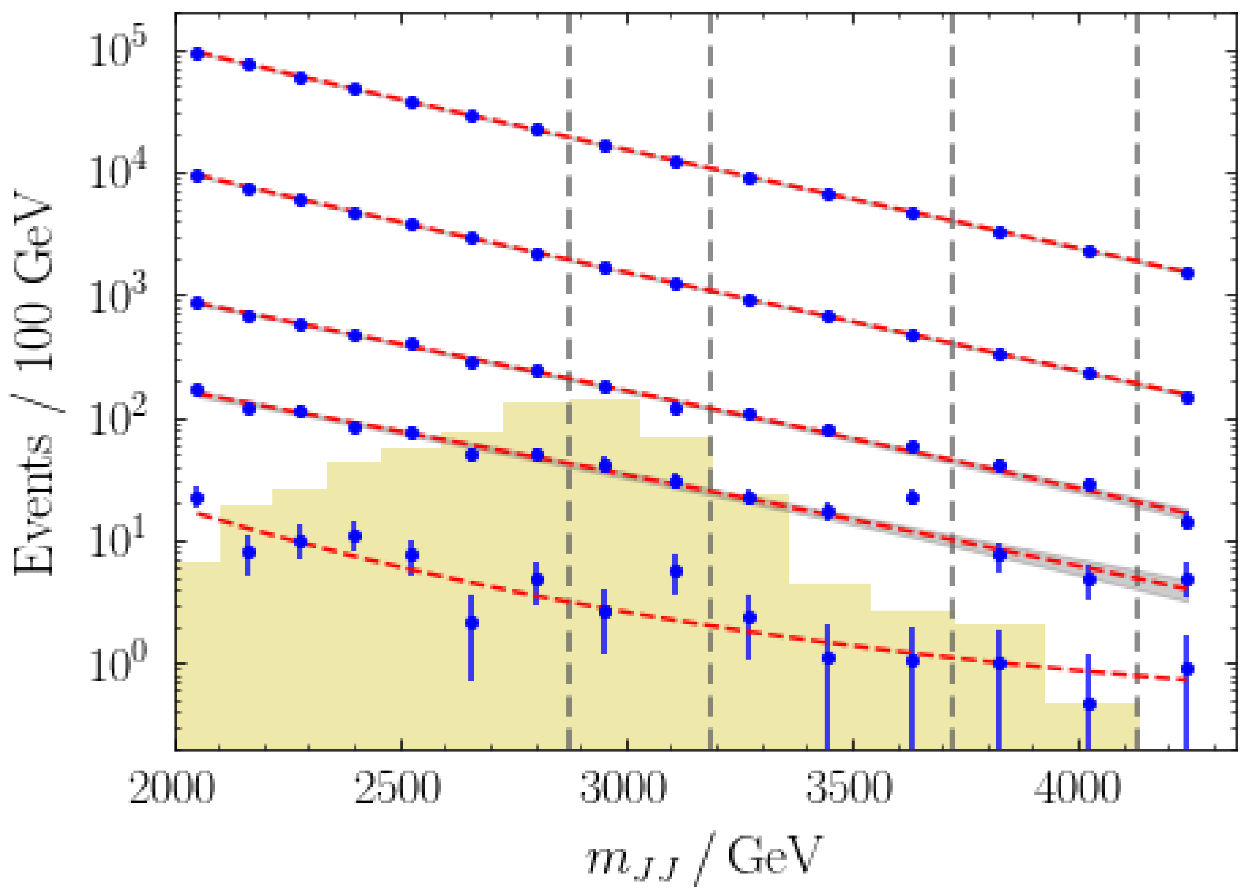
With signal
Events / 100 GeV
m_{JJ} / GeV
p_0
3σ
4σ
5σ
6σ
7σ
10%
1%
0.2%

Mass Scan

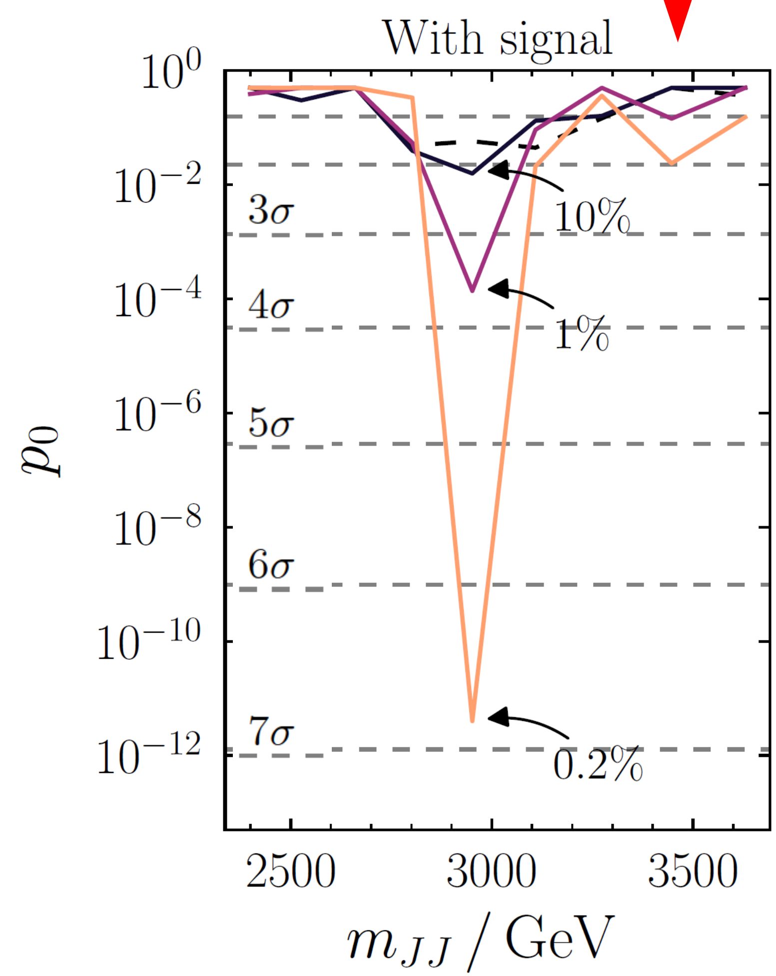


Events / 100 GeV
10^5
10^4
10^3
10^2
10^1
10^0
2000
2500
3000
3500
4000
m_{JJ} / GeV
With signal
p_0
10^0
10^{-2}
10^{-4}
10^{-6}
10^{-8}
10^{-10}
10^{-12}
3σ
4σ
5σ
6σ
7σ
10%
1%
0.2%
2500
3000
3500
m_{JJ} / GeV

Mass Scan

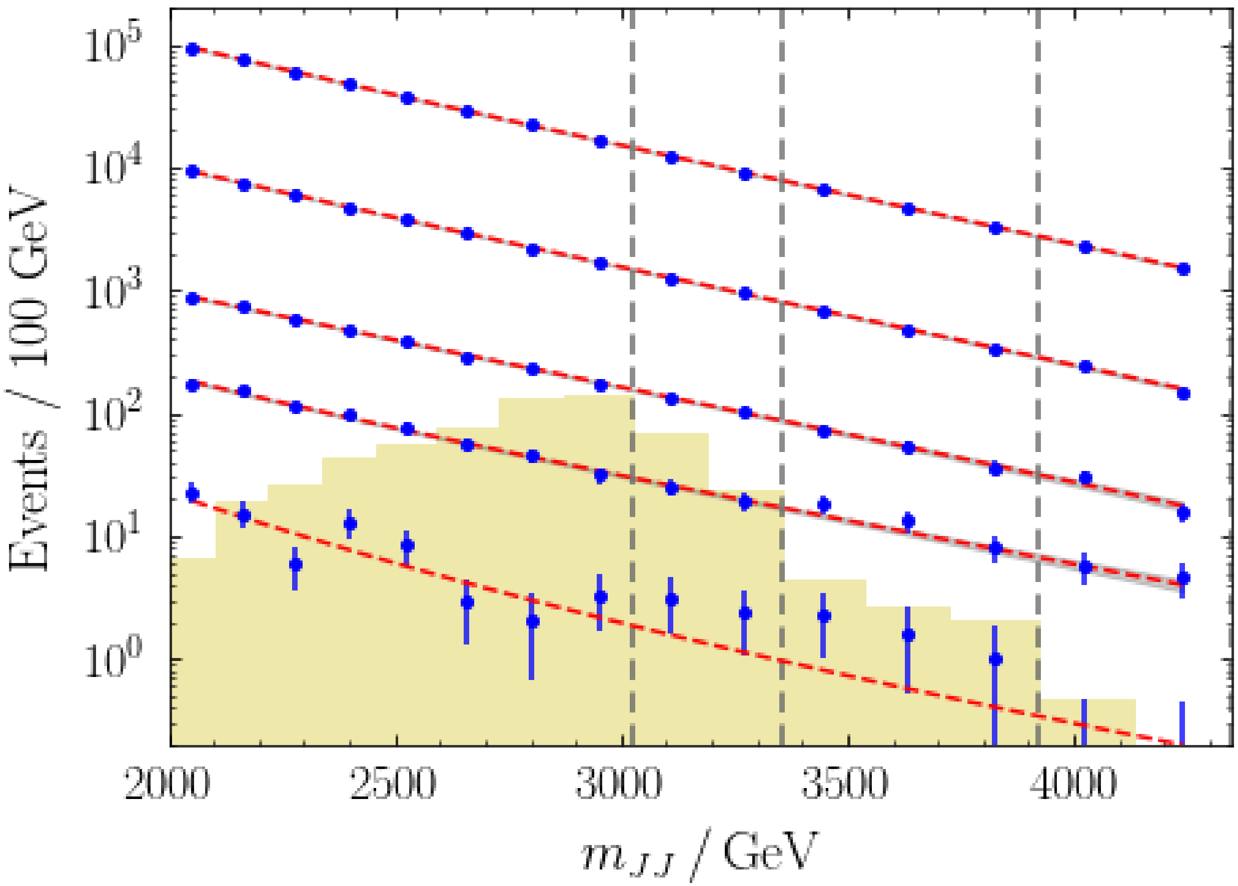


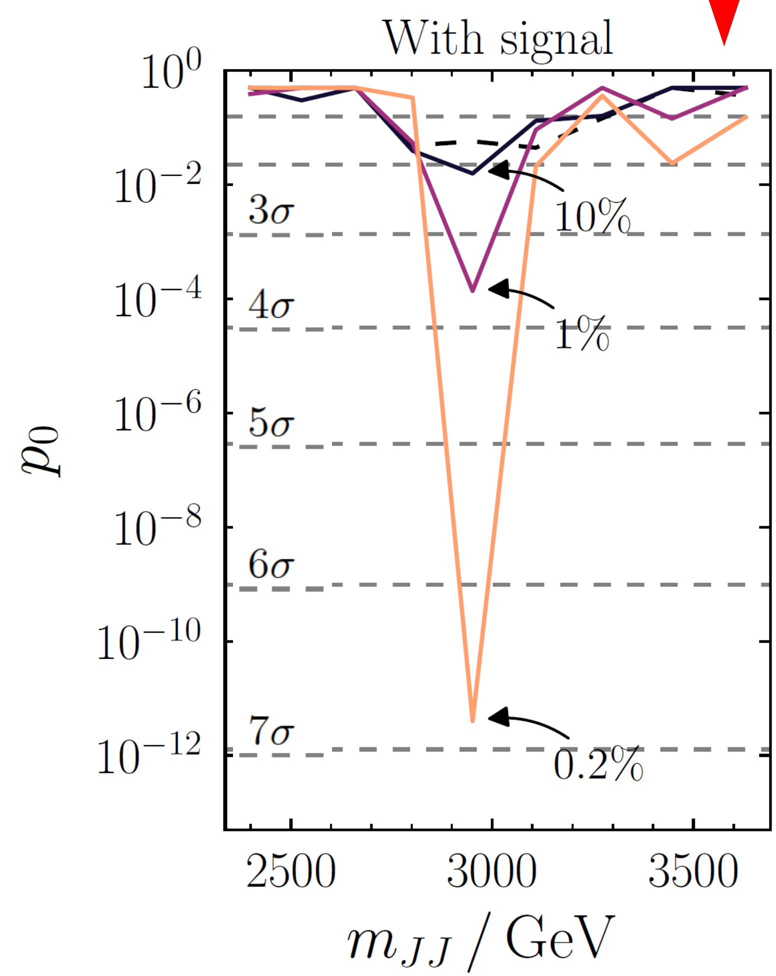
Events / 100 GeV
10^5
10^4
10^3
10^2
10^1
10^0
2000
2500
3000
3500
4000
m_{JJ} / GeV



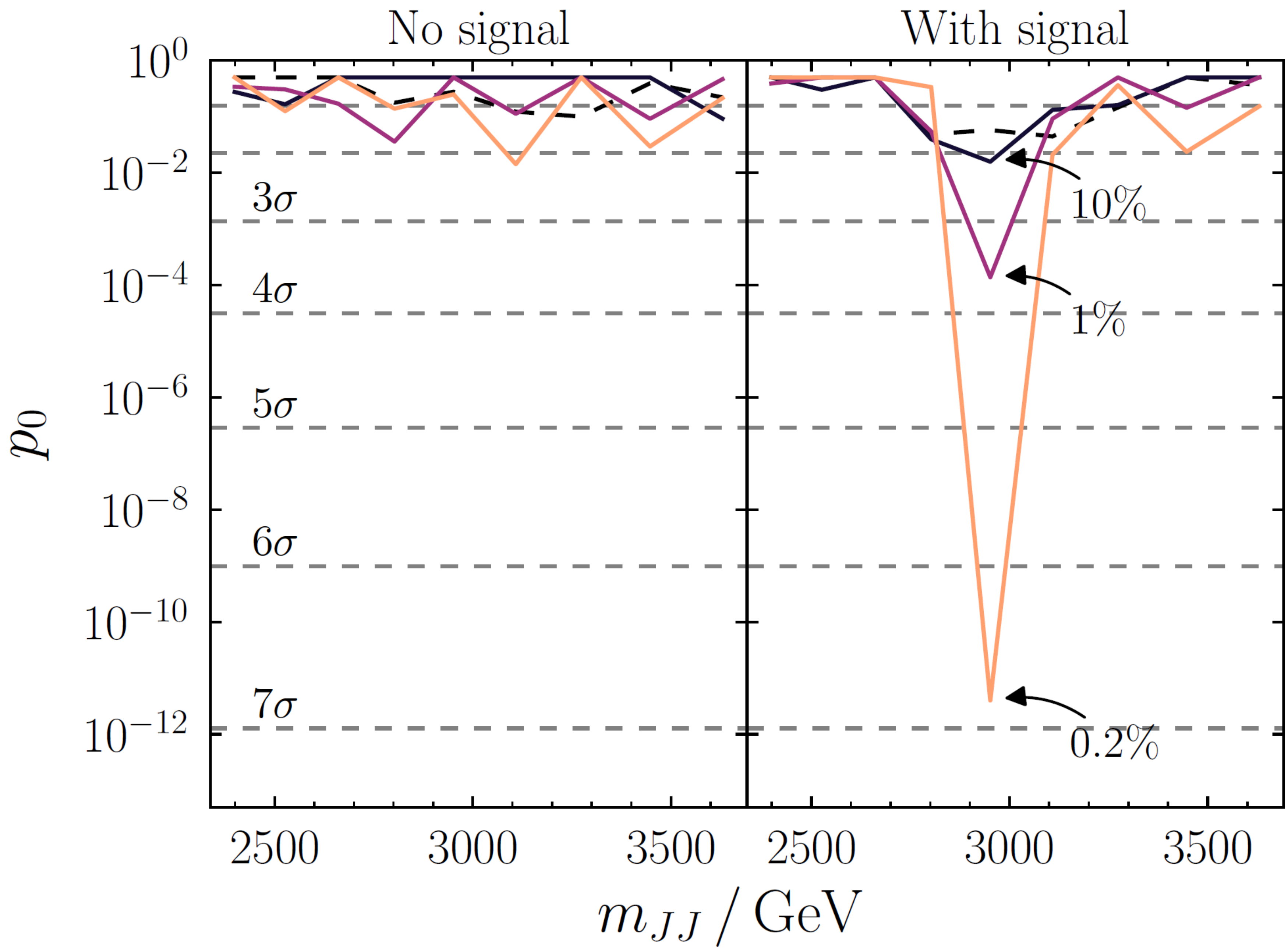
With signal
p_0
10^0
10^{-2}
10^{-4}
10^{-6}
10^{-8}
10^{-10}
10^{-12}
3σ
4σ
5σ
6σ
7σ
10%
1%
0.2%
2500
3000
3500
m_{JJ} / GeV

Mass Scan

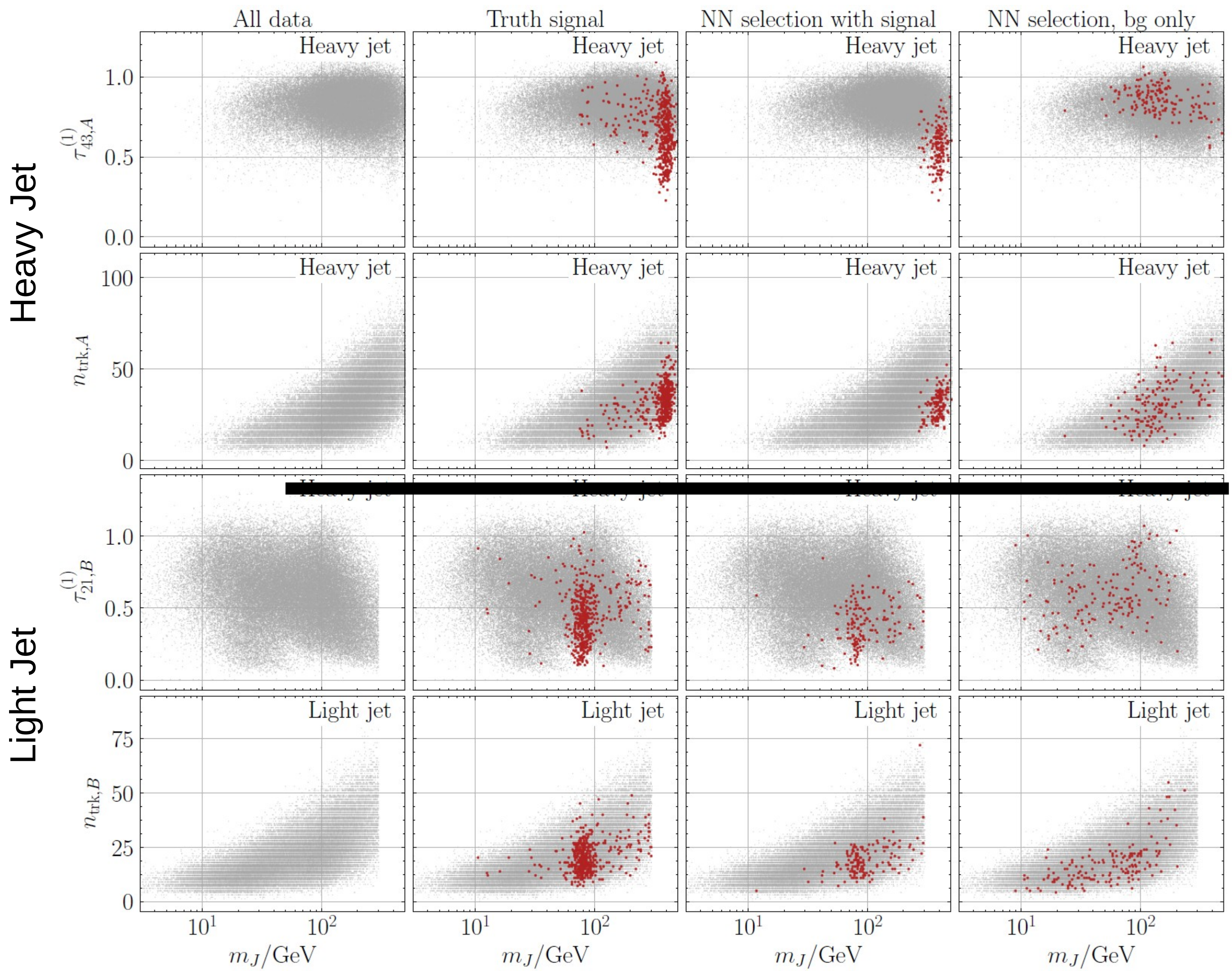

Events / 100 GeV
10^5
10^4
10^3
10^2
10^1
10^0
2000
2500
3000
3500
4000
m_{JJ} / GeV


With signal
p_0
10^0
10^{-2}
10^{-4}
10^{-6}
10^{-8}
10^{-10}
10^{-12}
3σ
4σ
5σ
6σ
7σ
10%
1%
0.2%
2500
3000
3500
m_{JJ} / GeV

Mass Scan


No signal
With signal
10^0
10^{-2}
10^{-4}
10^{-6}
10^{-8}
10^{-10}
10^{-12}
p_0
3σ
4σ
5σ
6σ
7σ
10%
1%
0.2%
2500
3000
3500
2500
3000
3500
m_{JJ} / GeV

Signal Characteristics



Performance Comparison

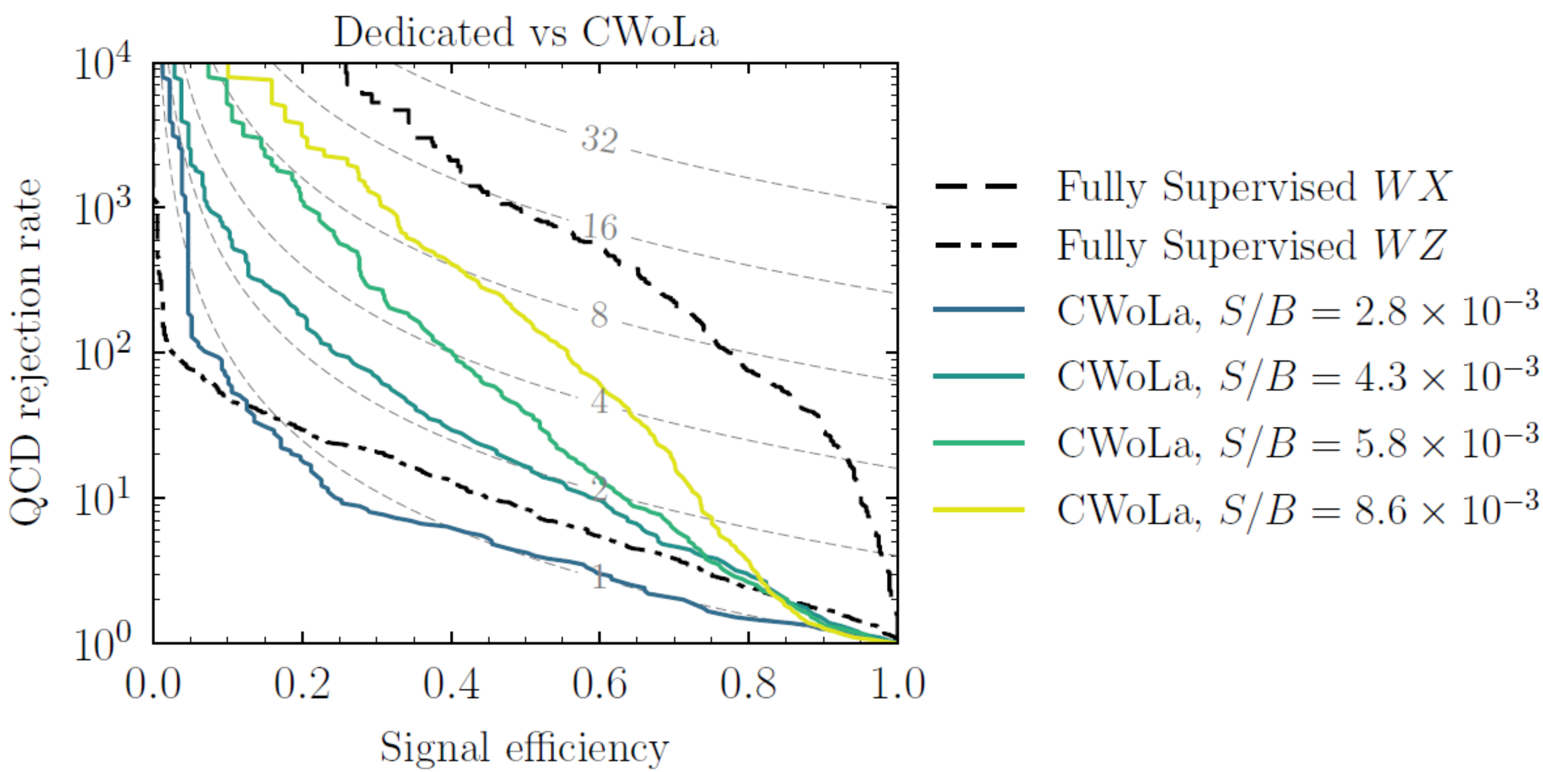


Figure 11. Truth-label ROC curves for taggers trained using CWoLa with varying number of signal events, compared to those for a dedicated tagger trained on pure signal and background samples (solid black) and one trained to discriminate W and Z jets from QCD (dashed black). The CWoLa examples have $B = 81341$ in the signal region and $S = (230, 352, 472, 697)$.

General CWoLa Hunting

1) Need some variable X (e.g. m_JJ) in which bg is smooth and signal is localized or has a sharp feature (could also be used for kinematic edges!)

2) Need some other variables {Y} (e.g. jet substructure) which may provide discriminating power which may be a-priori unknown.

3) {Y} should not be strongly correlated with X over the X-width of the signal.

   Or alternatively, if correlated, there may be a way to decorrelate (e.g. if we can predict or measure the correlation, that can be subtracted away to create new uncorrelated variables).

Map of model-agnostic searches

	Background MC Vs data	Data vs Data
Human Intelligence	'Traditional' CMS and ATLAS model agnostic searches	
Machine Intelligence	A few papers this year (not yet double checked to make 100% sure which ones fit here, but I think: [1806.02350] R. Tito D'Agnolo, A. Wulzer [1807.06038] A. De Simone, T. Jacques [1807.10261] J. Hajer, Y. Li, T. Liu, H. Wang (Please correct me if I made a mistake here!)	**Train on signal region and bg region** [1805.02664] J. H. Collins, K. Howe, B. Nachman **Train only on bg region** [1808.08979] T. Heimel, G. Kasieczka, T. Plehn, J. M. Thompson [1808.08992] M. Farina, Y. Nakai, D. Shih

Background-only training vs signal/sideband:

Background-only

Tagger performance does not depend on signal statistics.

Tagger can never learn the *specific* peculiar features of the signal, and so **cannot improve with greater signal rate**.

Signal / Sideband

Tagger relies on there being sufficient signal statistics for training.

Tagger can learn the *specific* peculiar features of the signal, and so **improves with greater signal rate**, and allows for **signal characterization**.

Stronger in limit of very low signal statistics

??

Stronger in limit of very high signal statistics