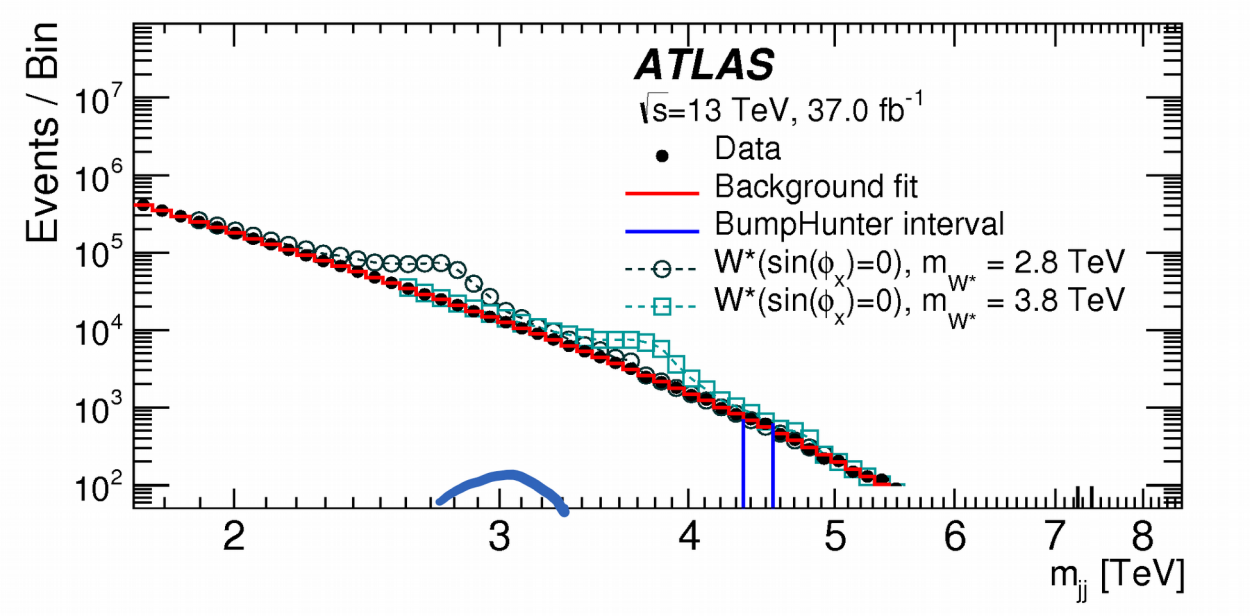
CWoLa Hunting:

Extending the Bump Hunt with Machine Learning

Based on:

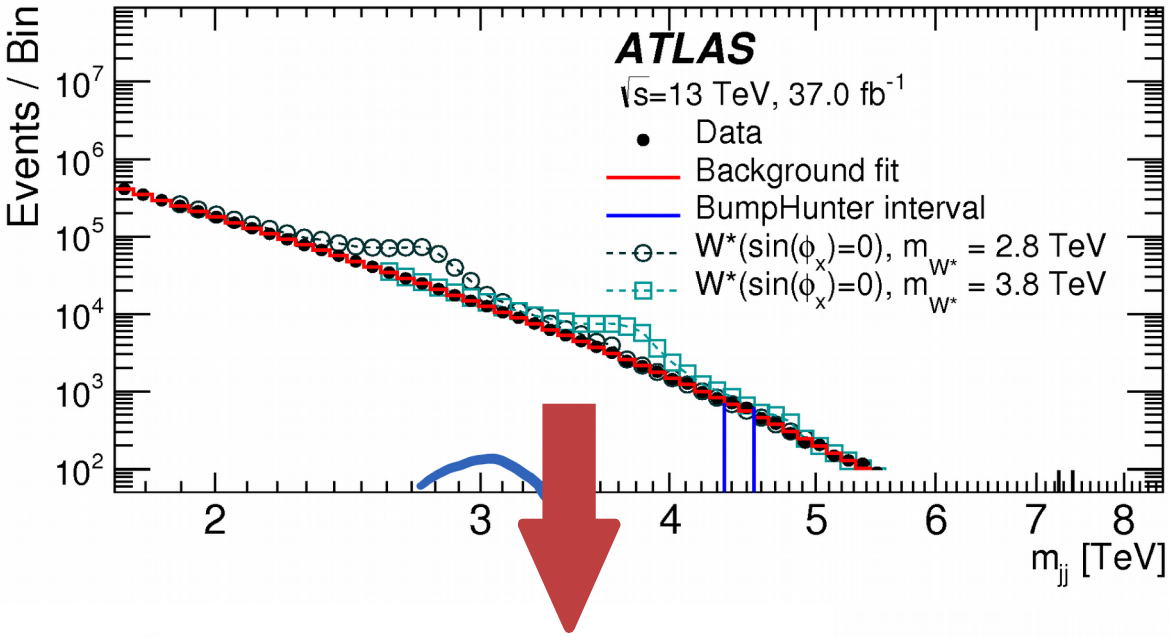
[1805.02664] Jack Collins, Kiel Howe, Ben Nachman

Dijet Resonances



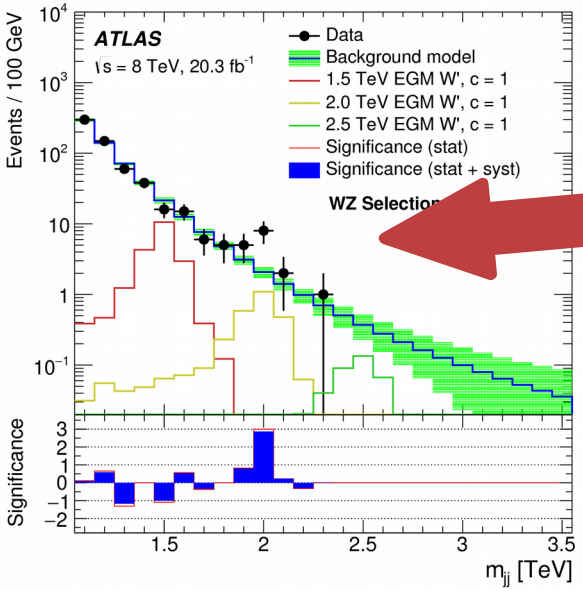
Edited from [1703.01927]

Dijet Resonances



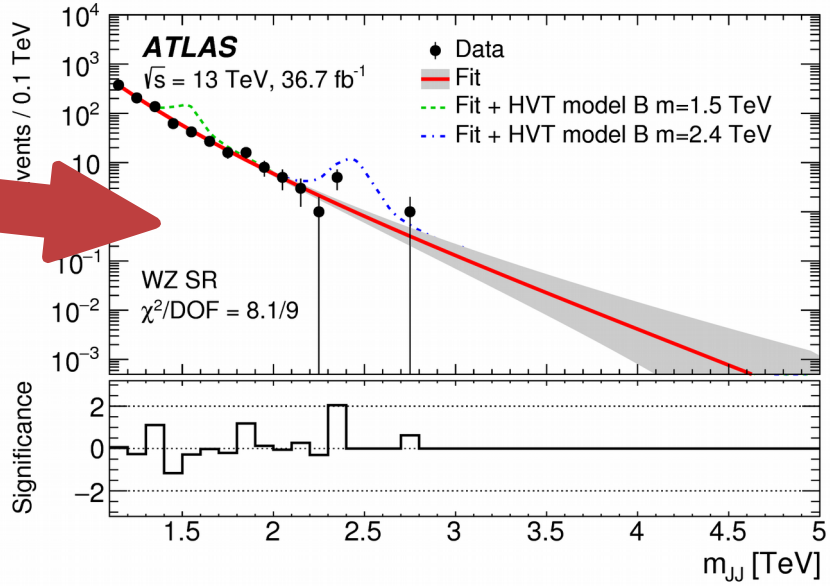
Edited from [1703.01927]

[1506.00962]

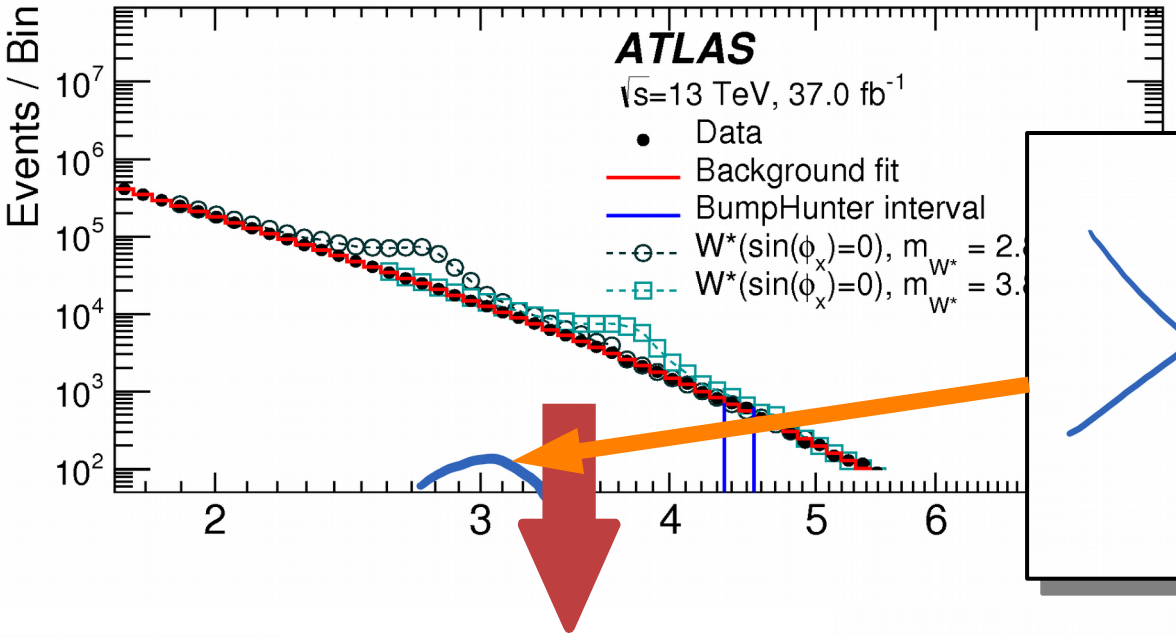


1) Theorist comes up with specific model with some specific prediction (e.g. W' → WZ).
2) Choose dedicated substructure variables.
3) Simulate signal to optimize cuts
4) Calibrate in some data sample
5) Apply cuts to events and look for a bump in the new distribution

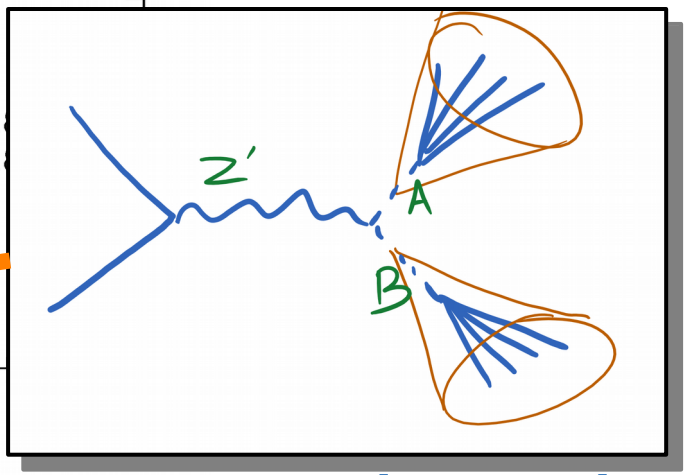
[1708.04445]



Dijet Resonances

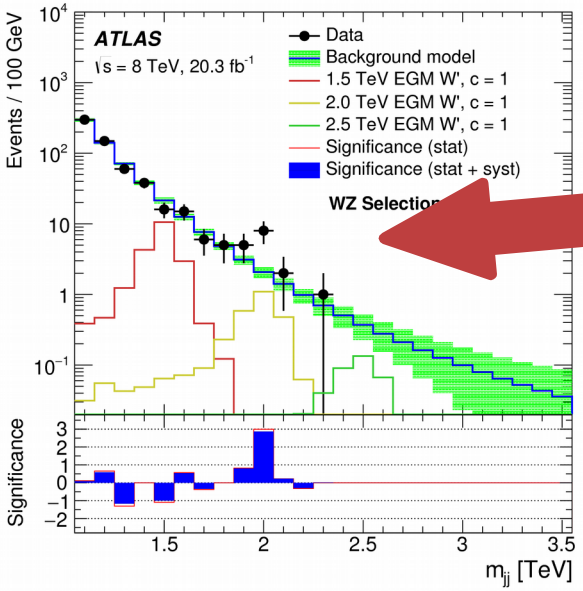


Edited from [1703.01927]

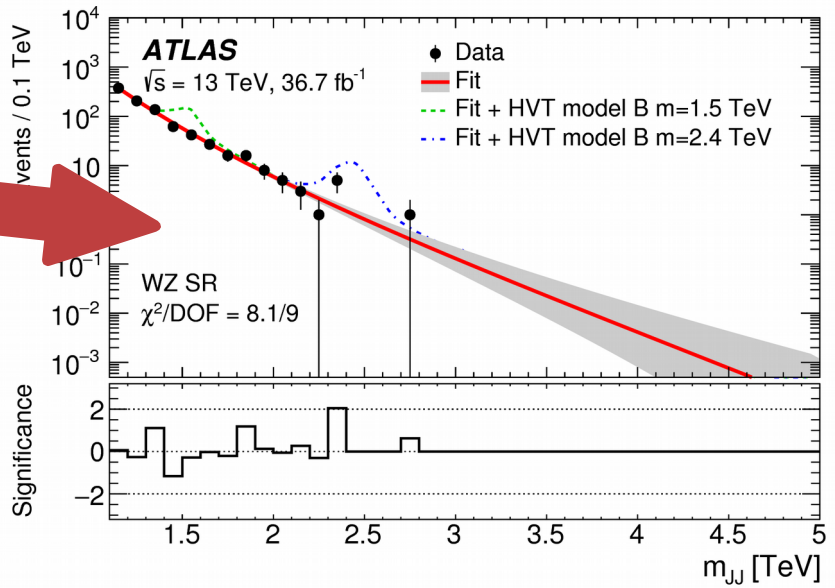


[1708.04445]

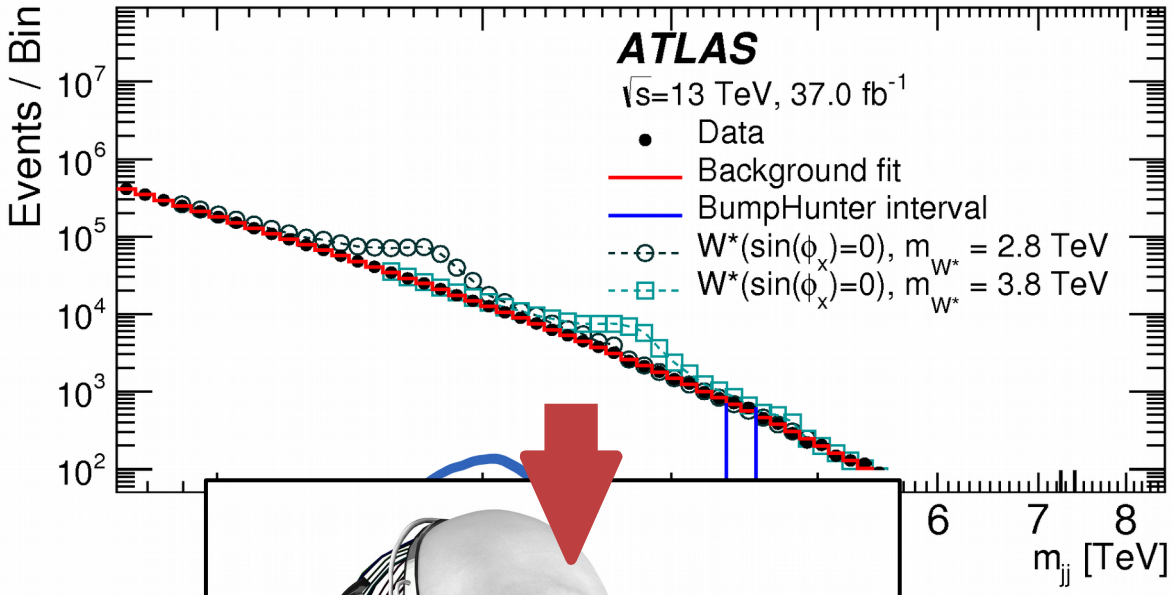
[1506.00962]



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

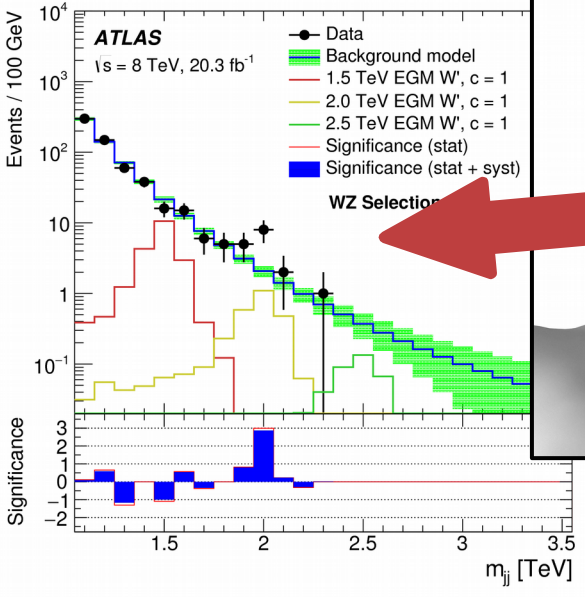


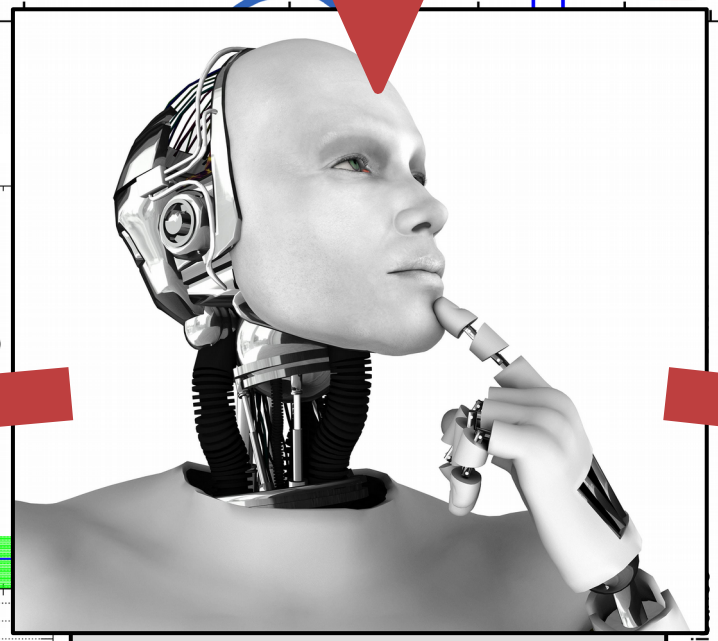
Dijet Resonances



Edited from [1703.01927]

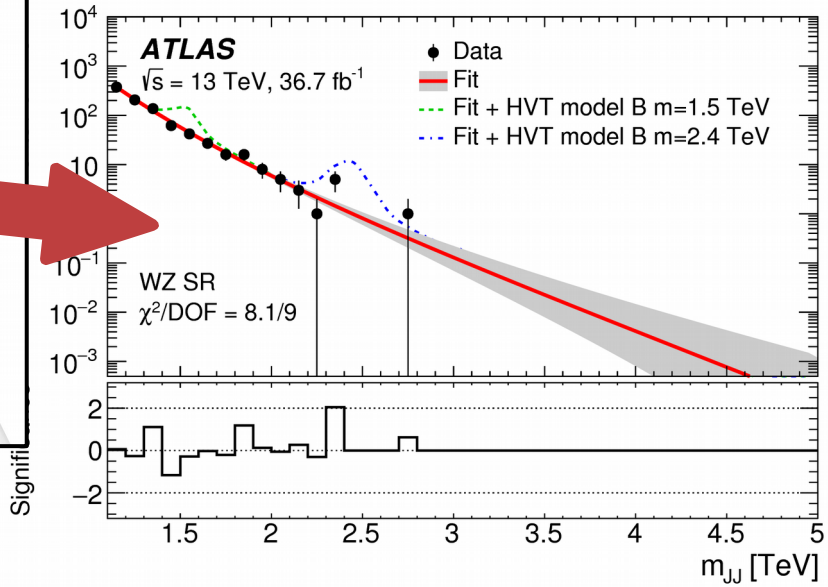
[1506.00962]



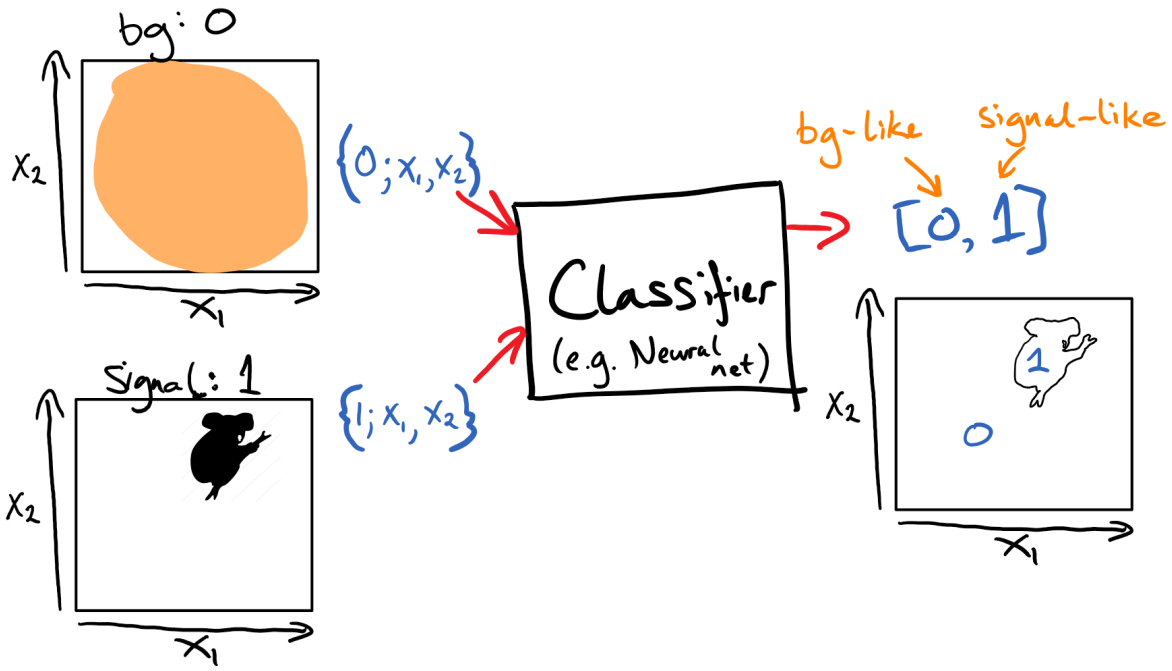


5) Apply cuts to events and look for a bump in the new distribution

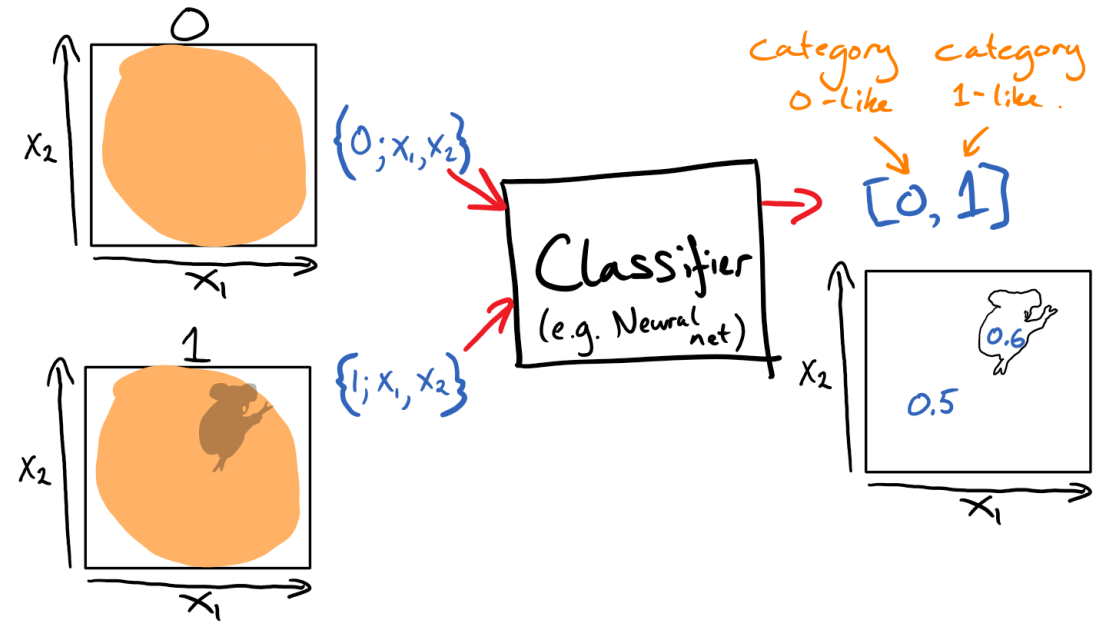
[1708.04445]



Fully Supervised learning vs Weak Supervision



$$\frac{r}{1-r} \sim \frac{p(\text{data}\,|\,1)}{p(\text{data}\,|\,0)}$$

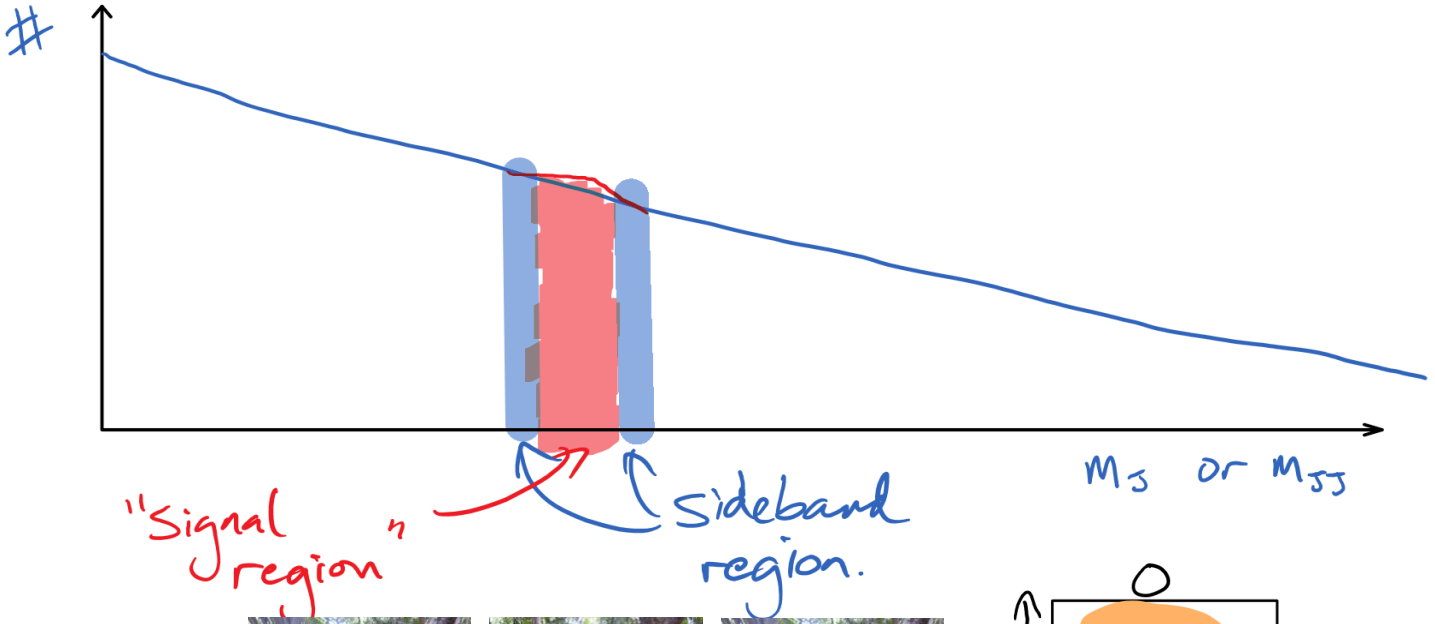


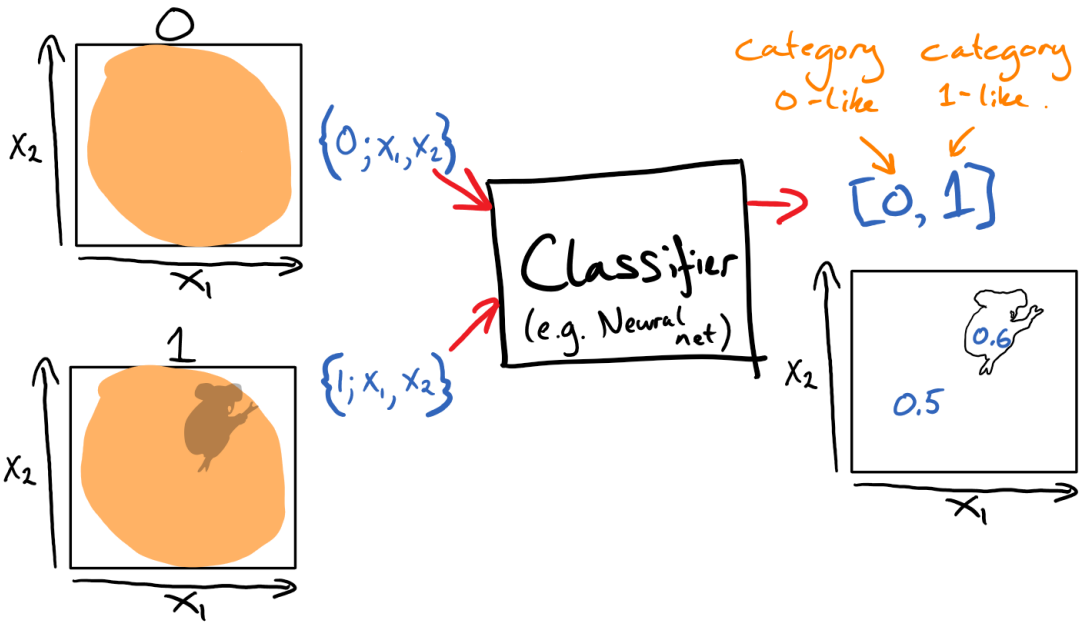
CWoLa (and other weak supervision):

[1708.02949] E. M. Metodiev, B. Nachman, J. Thaler
[1702.00414] L. M. Dery, B. Nachman, F. Rubbo, A Schwartzman
[1801.10158] P. T. Komiske, E. M. Metodiev, B. Nachman, M. D. Schwartz
[1706.09451] T. Cohen, M. Freytsis, B. Ostdiek

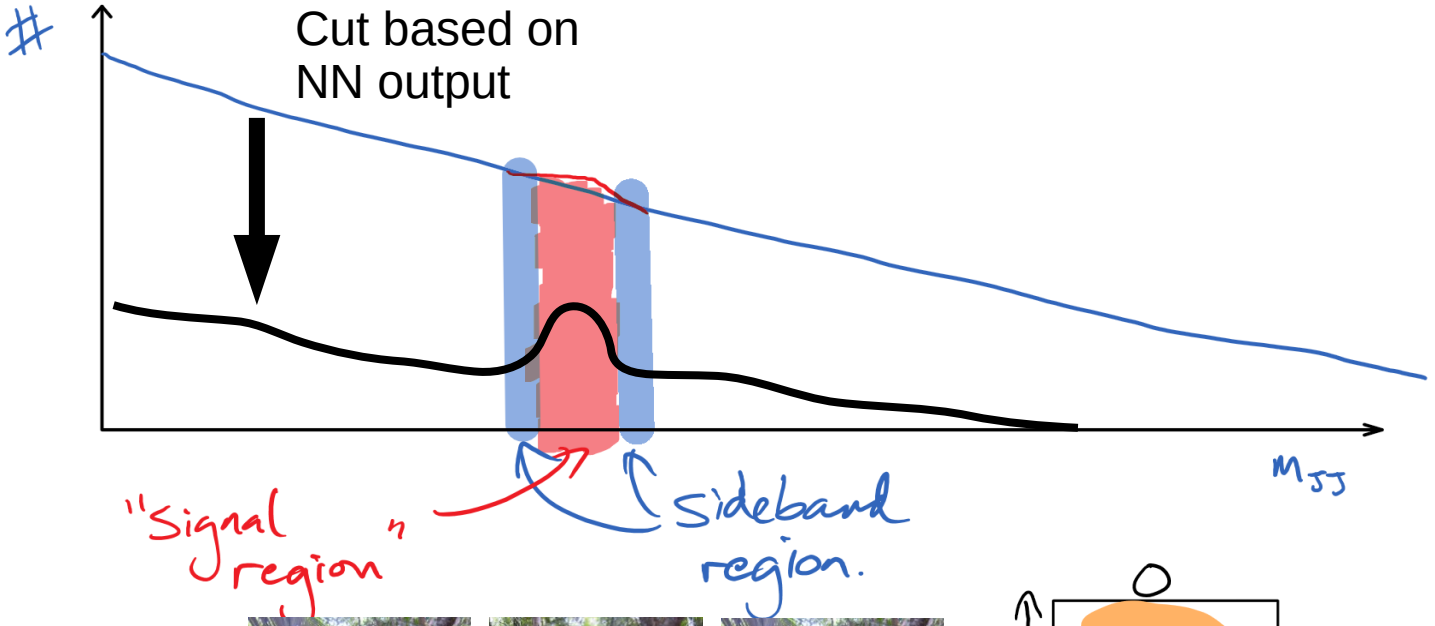
$$\frac{r}{1-r} \sim \frac{p(\text{data}\,|\,S+B)}{p(\text{data}\,|\,B)} = \frac{p(\text{data}\,|\,S)}{p(\text{data}\,|\,B)} + 1$$

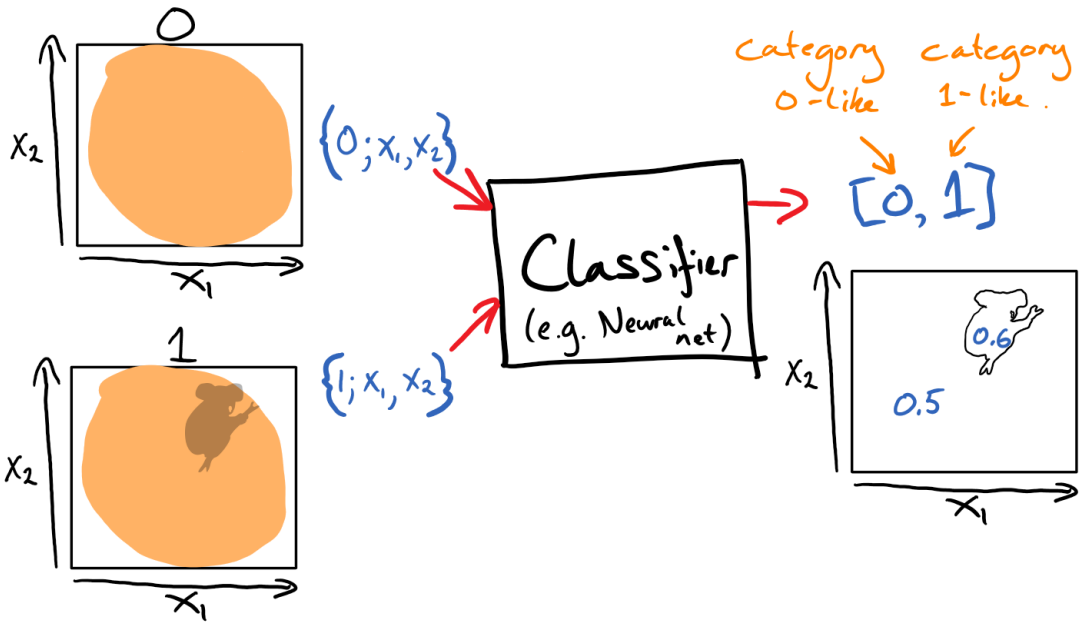
CwoLa Hunting: Basic Picture





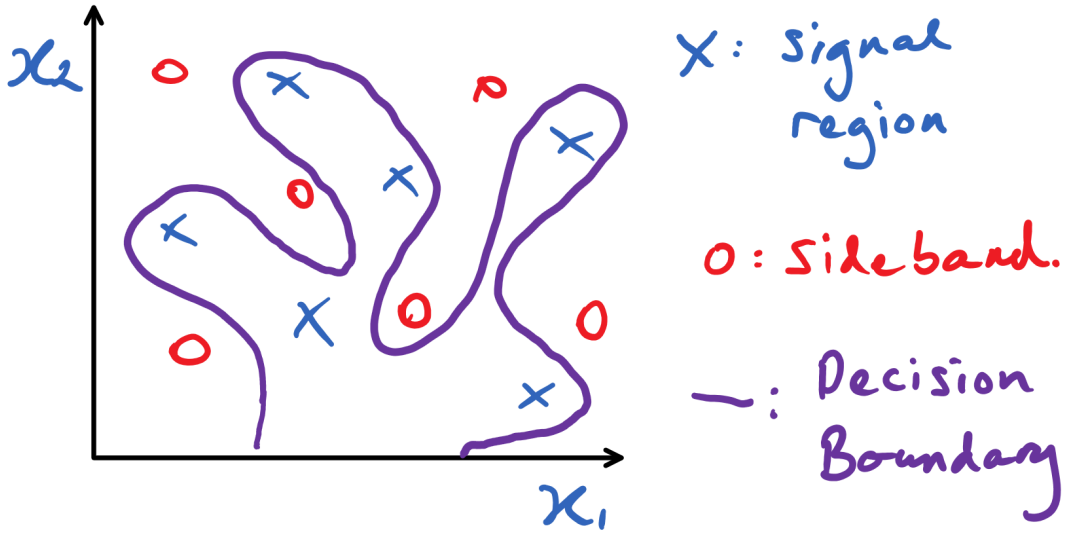
CwoLa Hunting: Basic Picture



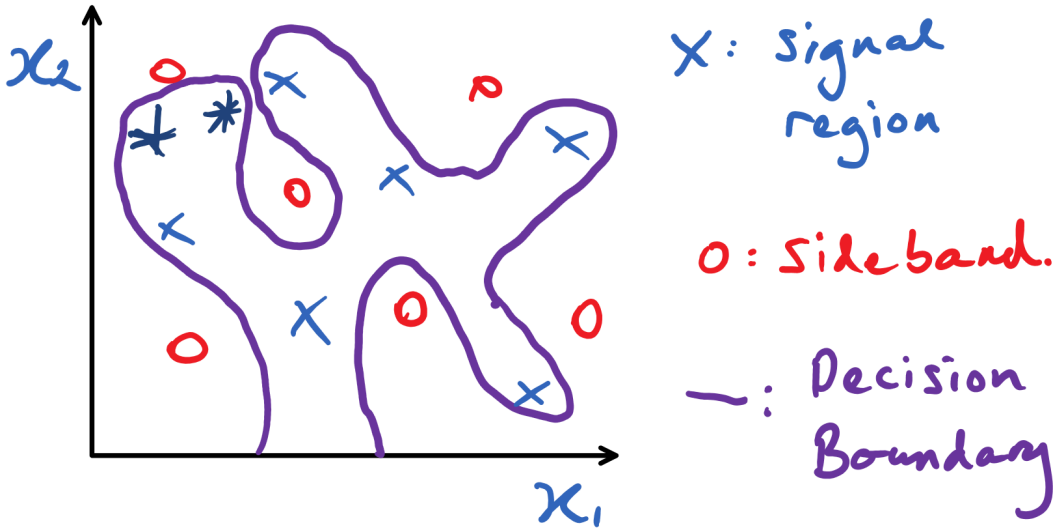


The problem(s) of overfitting

1) Create fake bumps out of background fluctuations

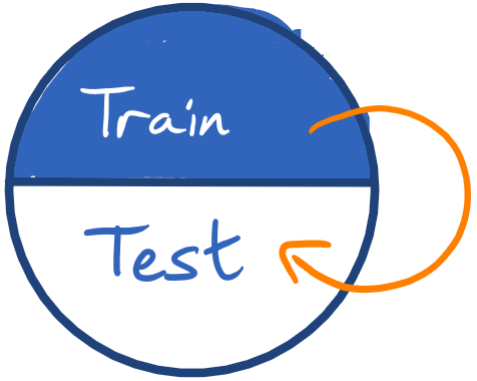


2) Leads to sub-optimal discriminant for true signal



Avoiding fake bumps: Simplest Solution

Train/Test Split

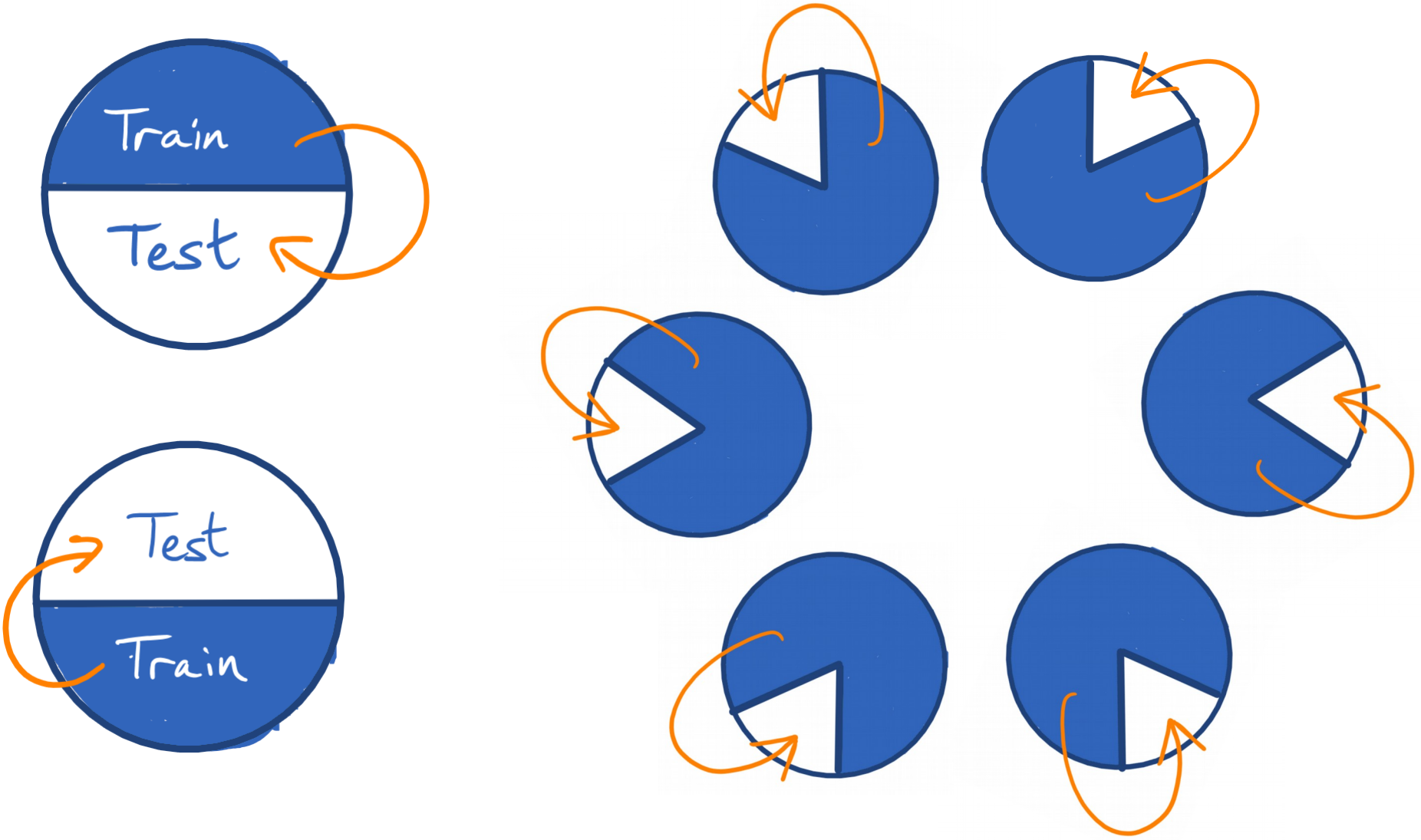


'Wastes' half of the data

Train & Test on same dataset

Create fake bumps by overfitting to statistical fluctuations

Avoiding fake bumps: Cross Validation



Nested Cross-Validation

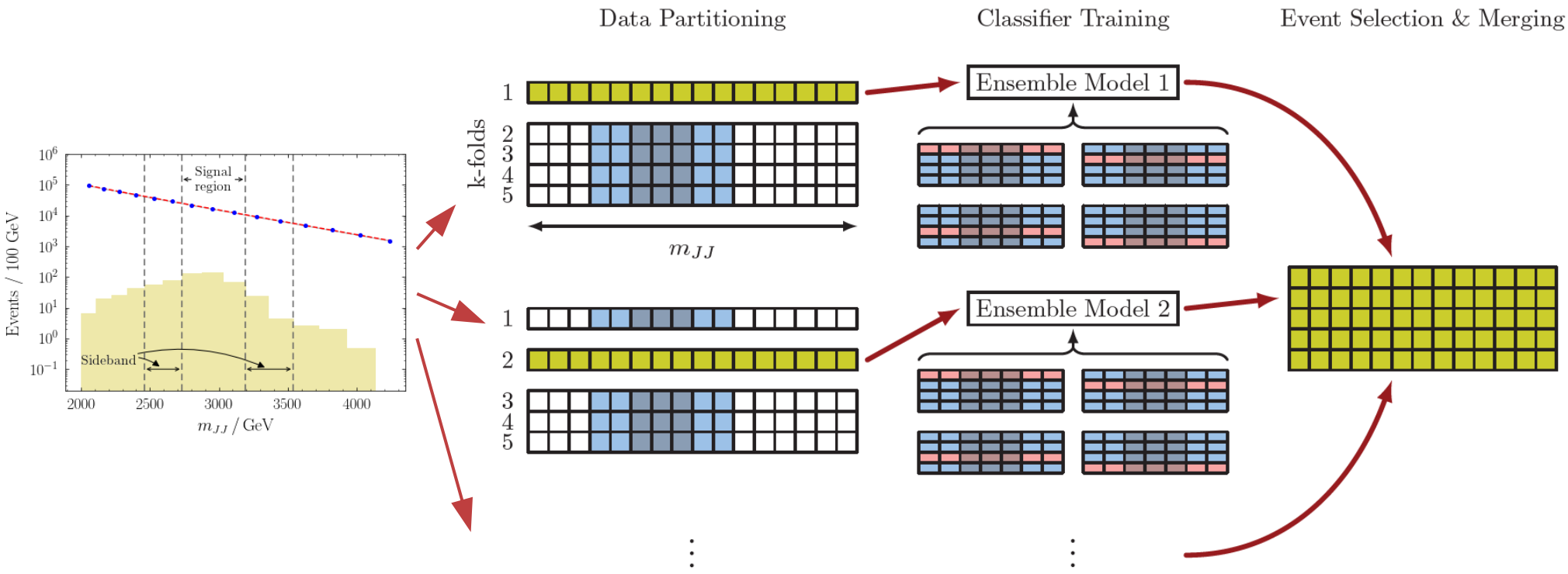
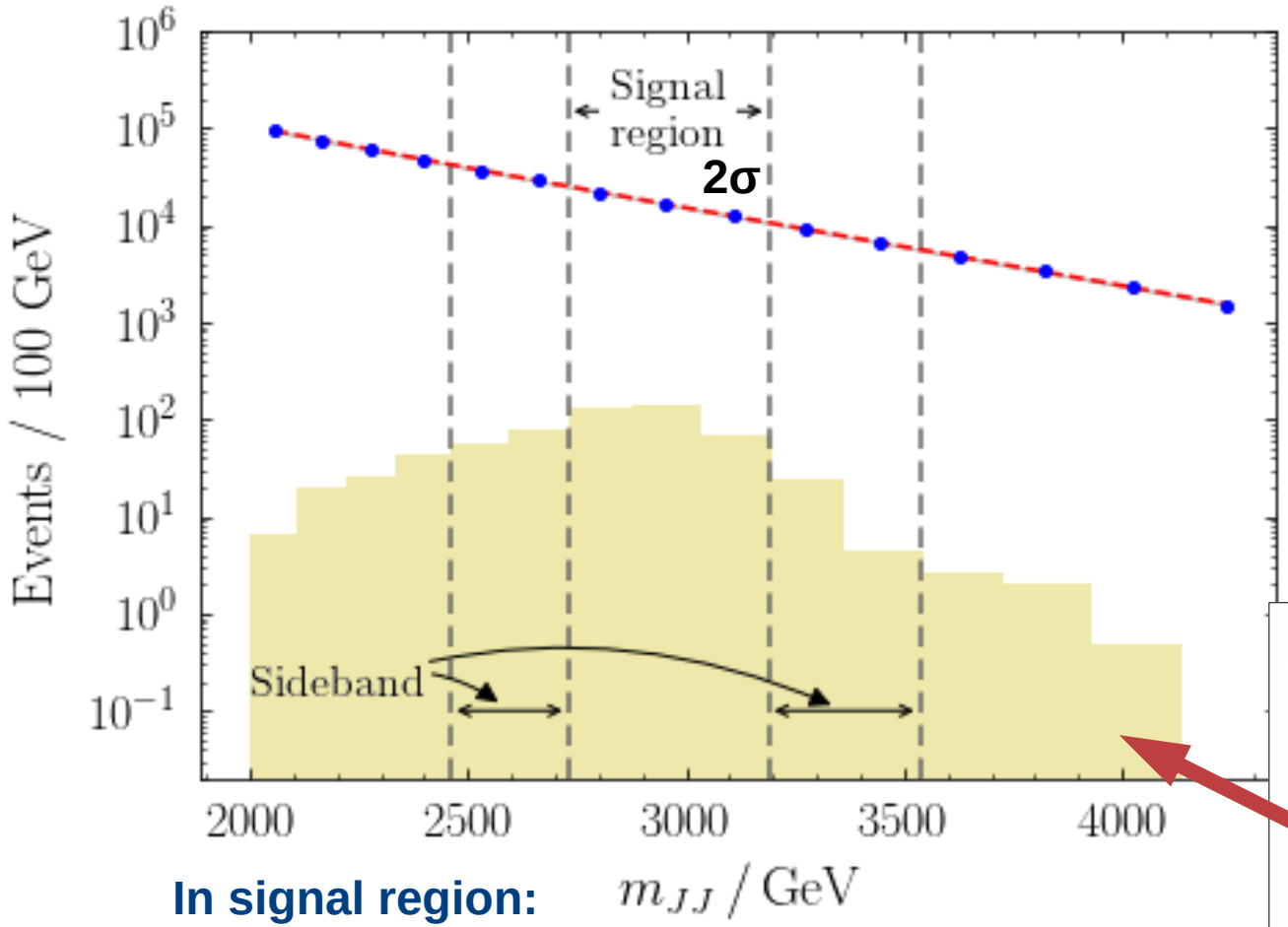


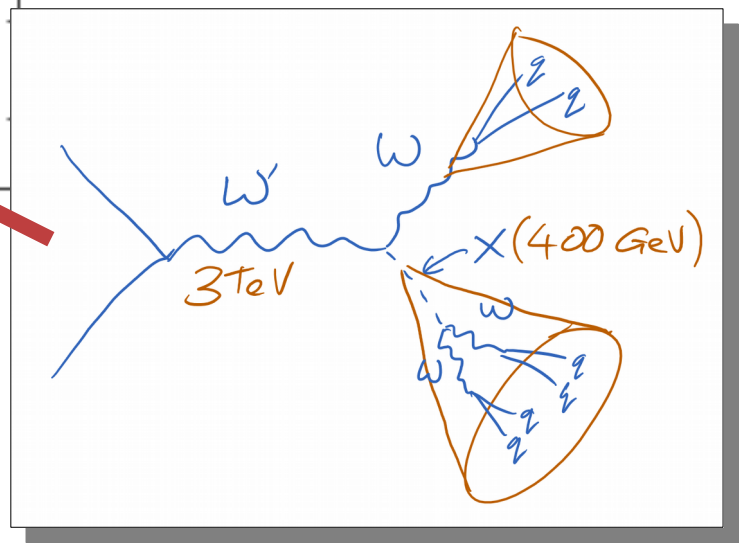
Figure 7. Illustration of the nested cross-validation procedure. **Left:** the dataset is randomly partitioned bin-by-bin into five groups. **Center:** for each group, an ensemble classifier is trained on the remaining groups. For each of the four possible combinations of these four groups into three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

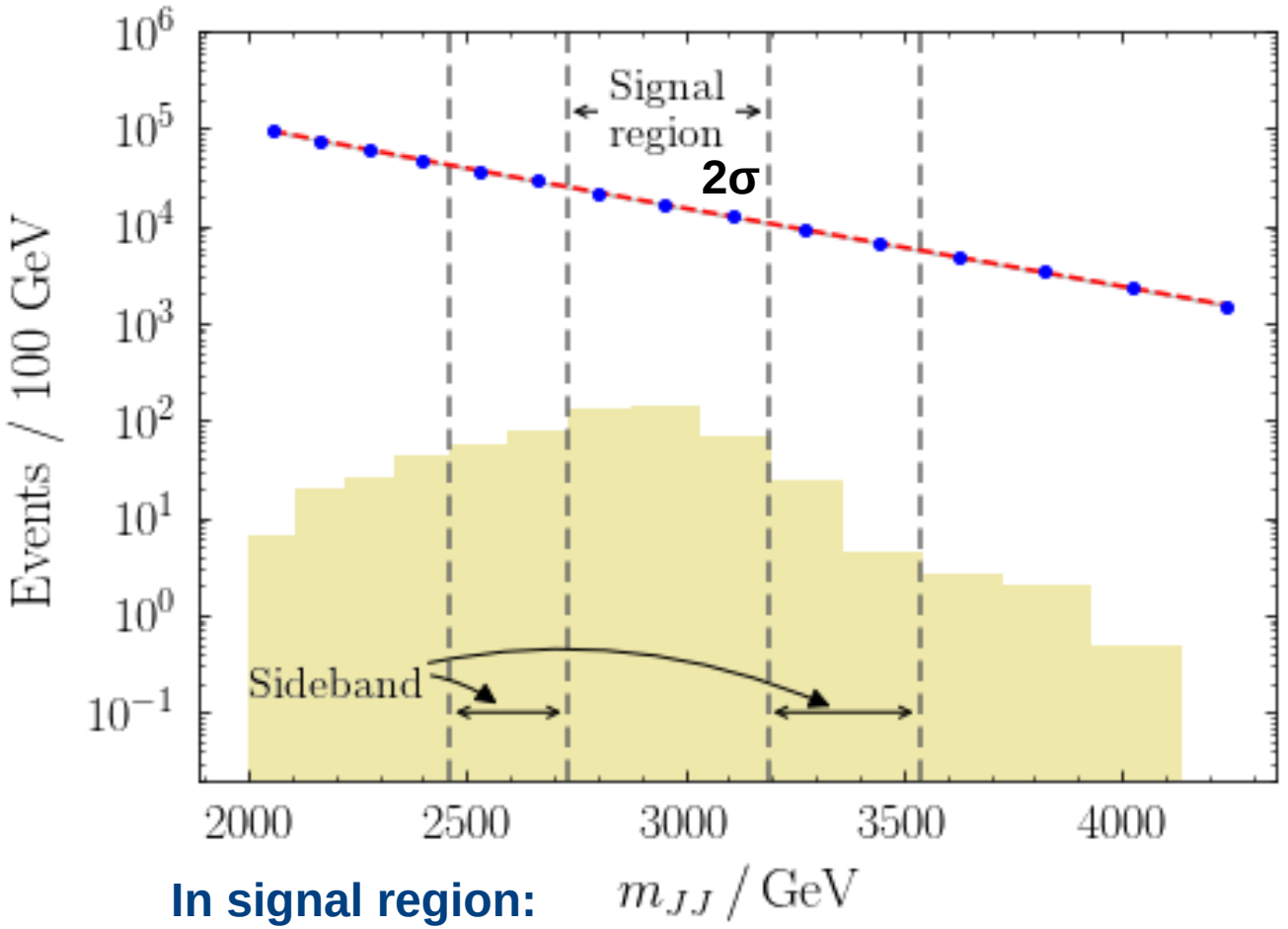


$$\frac{dN}{dm_{JJ}} = p_0 \frac{(1 - m_{JJ}/\sqrt{s})^{p_1}}{(m_{JJ}/\sqrt{s})^{p_2}}$$

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

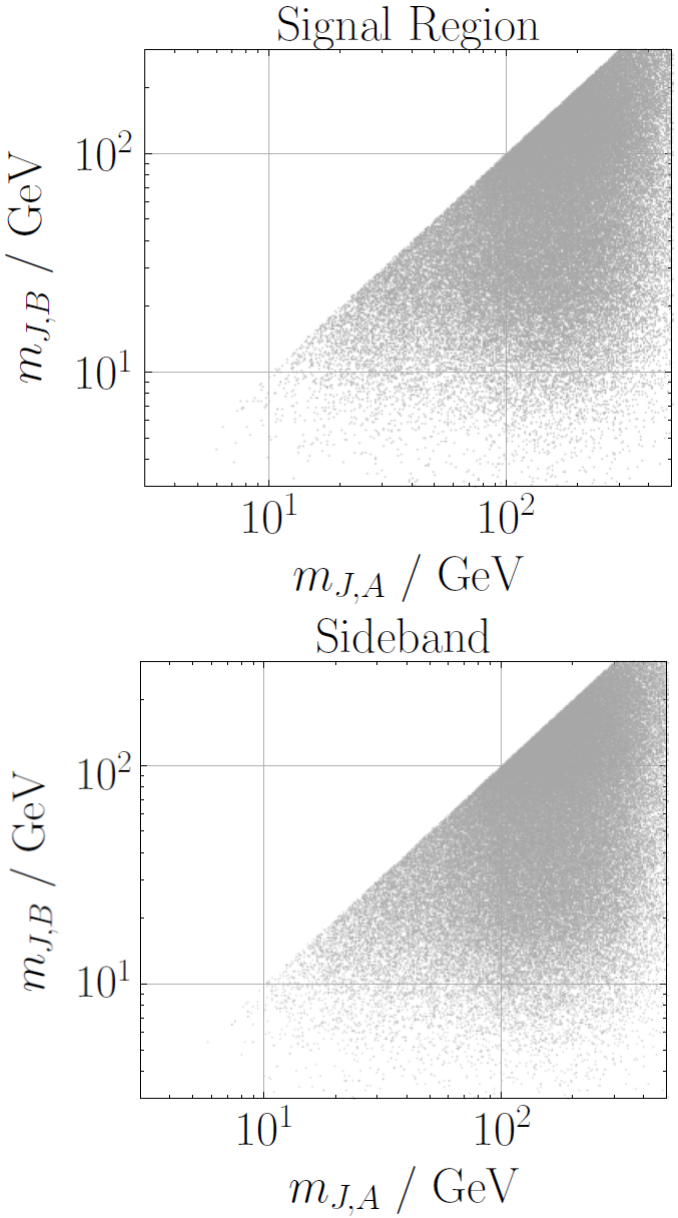


Application to Bump Hunt

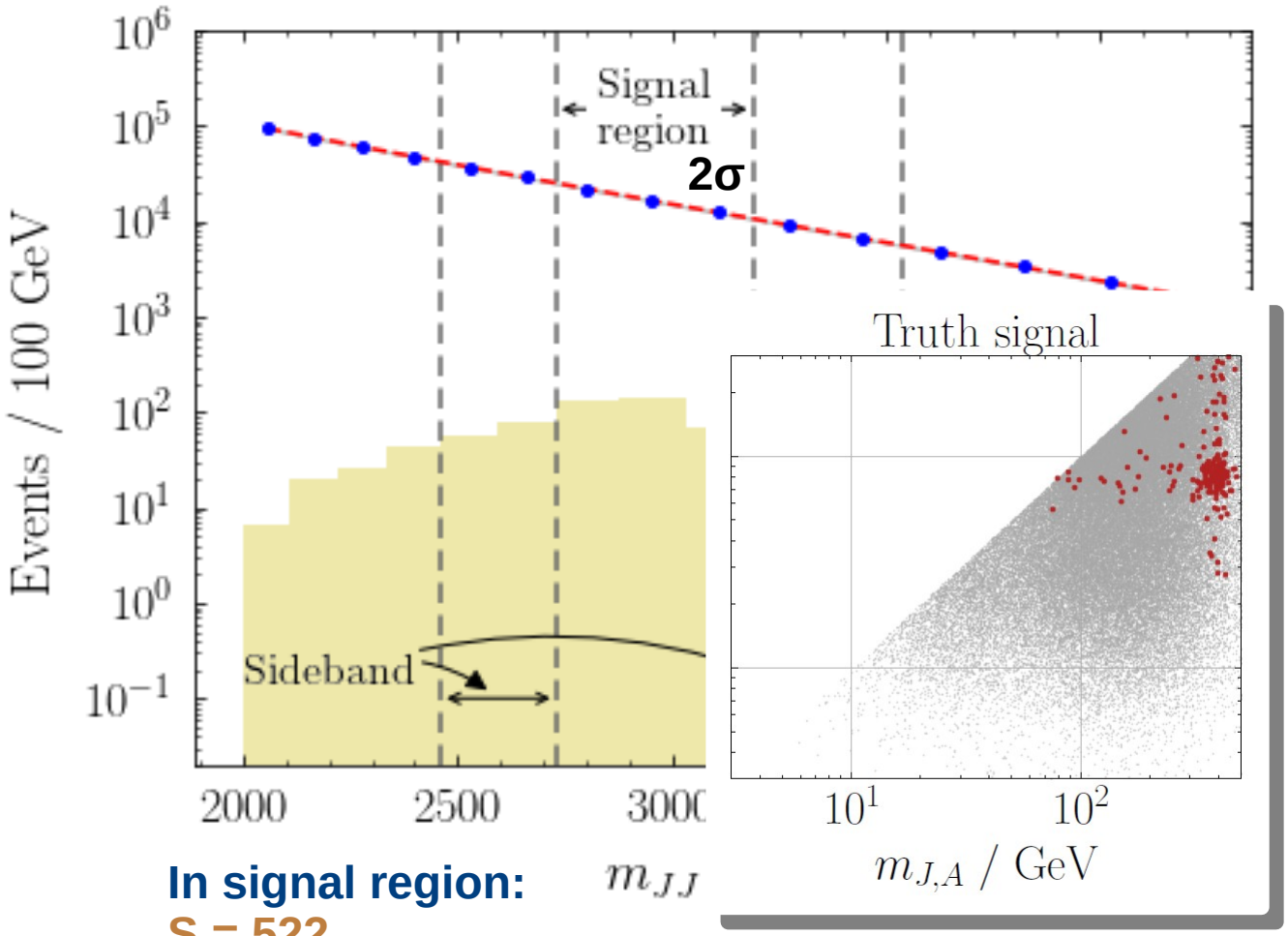


In signal region:
S = 522,
S/B = 0.64%

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

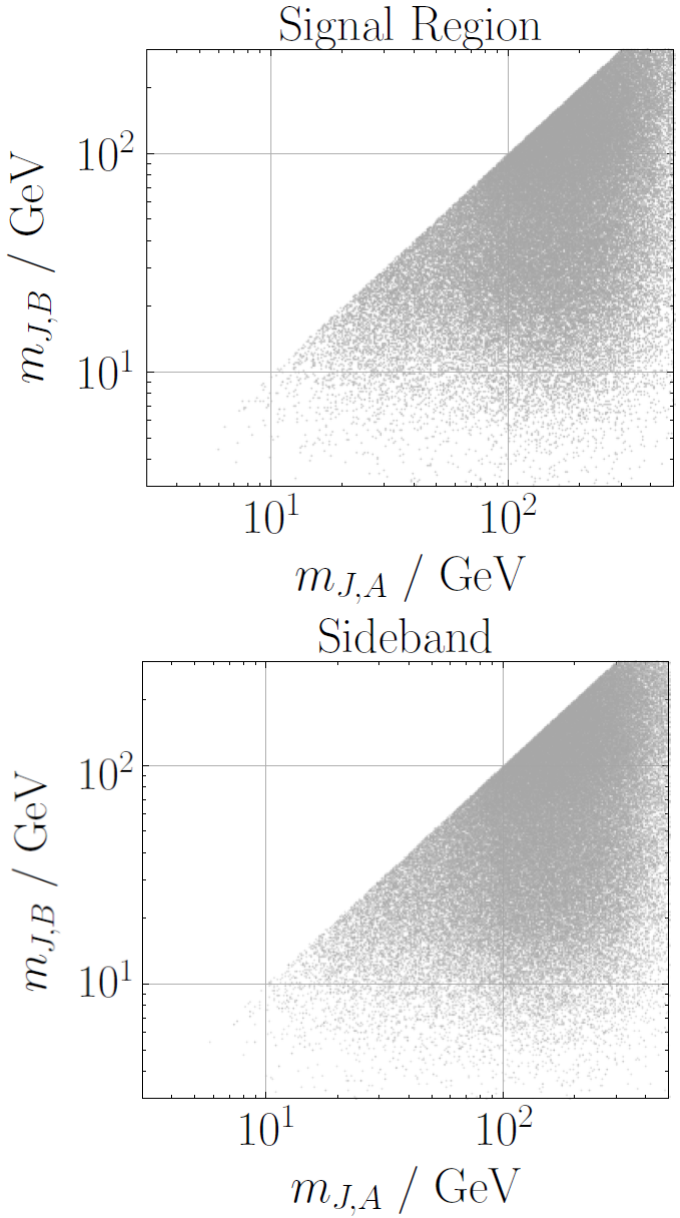


Application to Bump Hunt

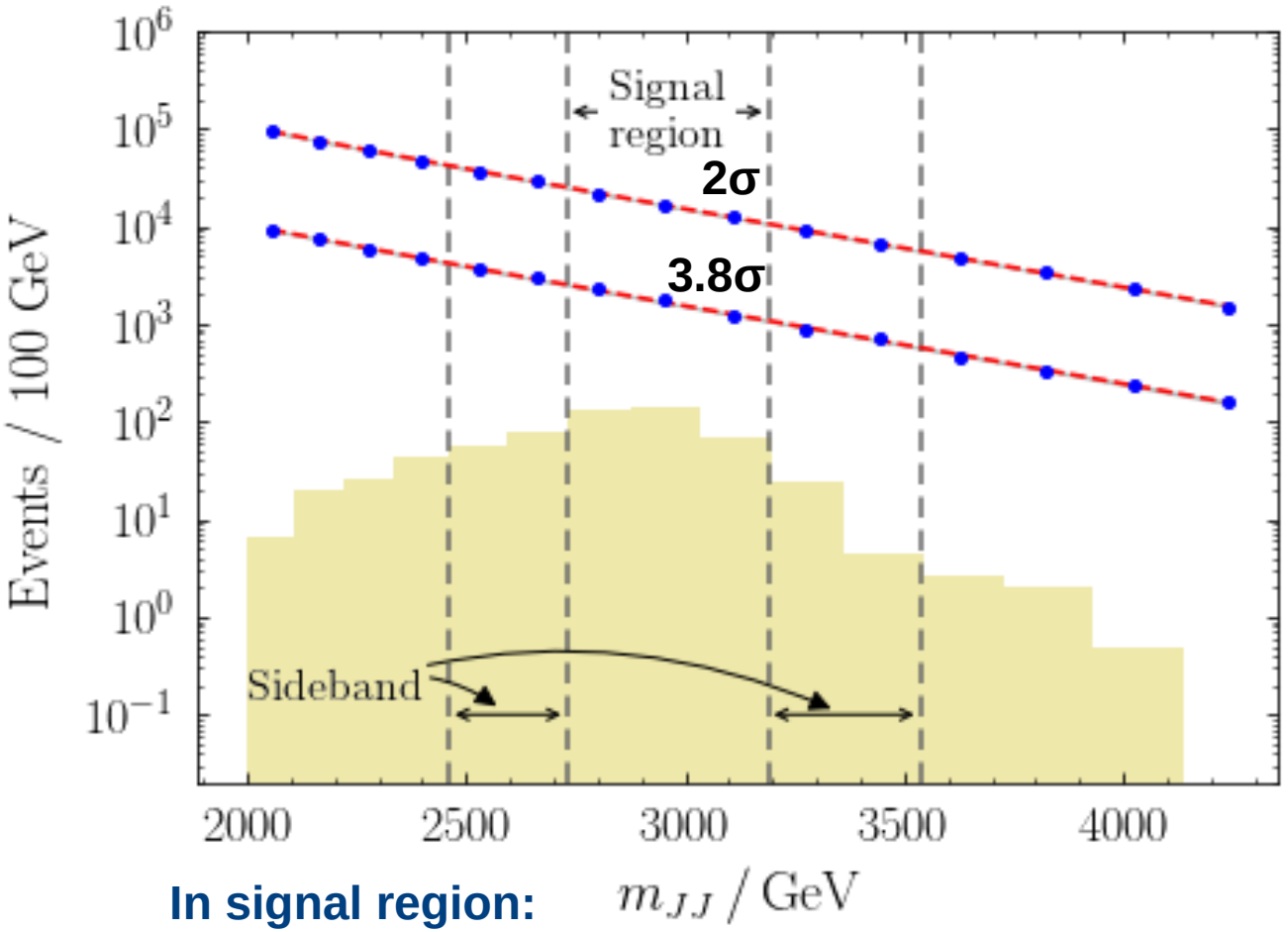


In signal region:
S = 522,
S/B = 0.64%

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

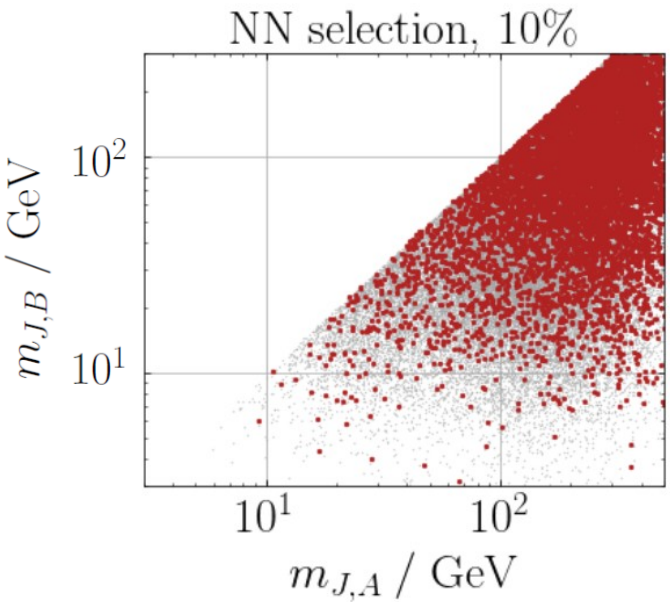


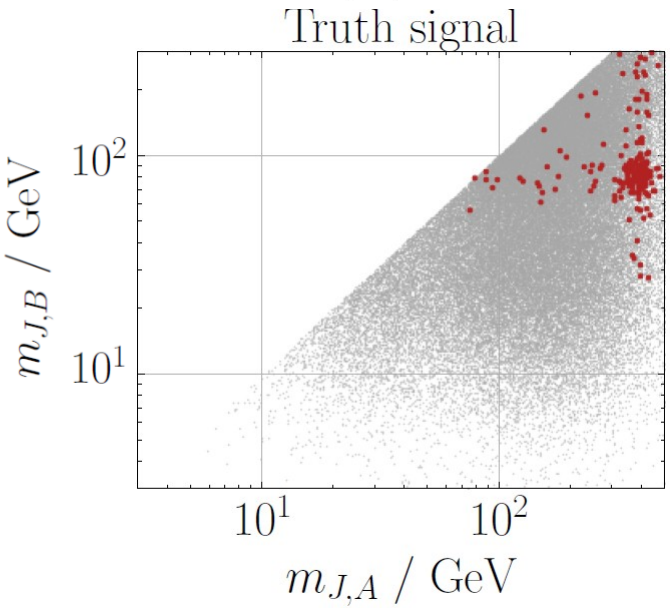
Application to Bump Hunt



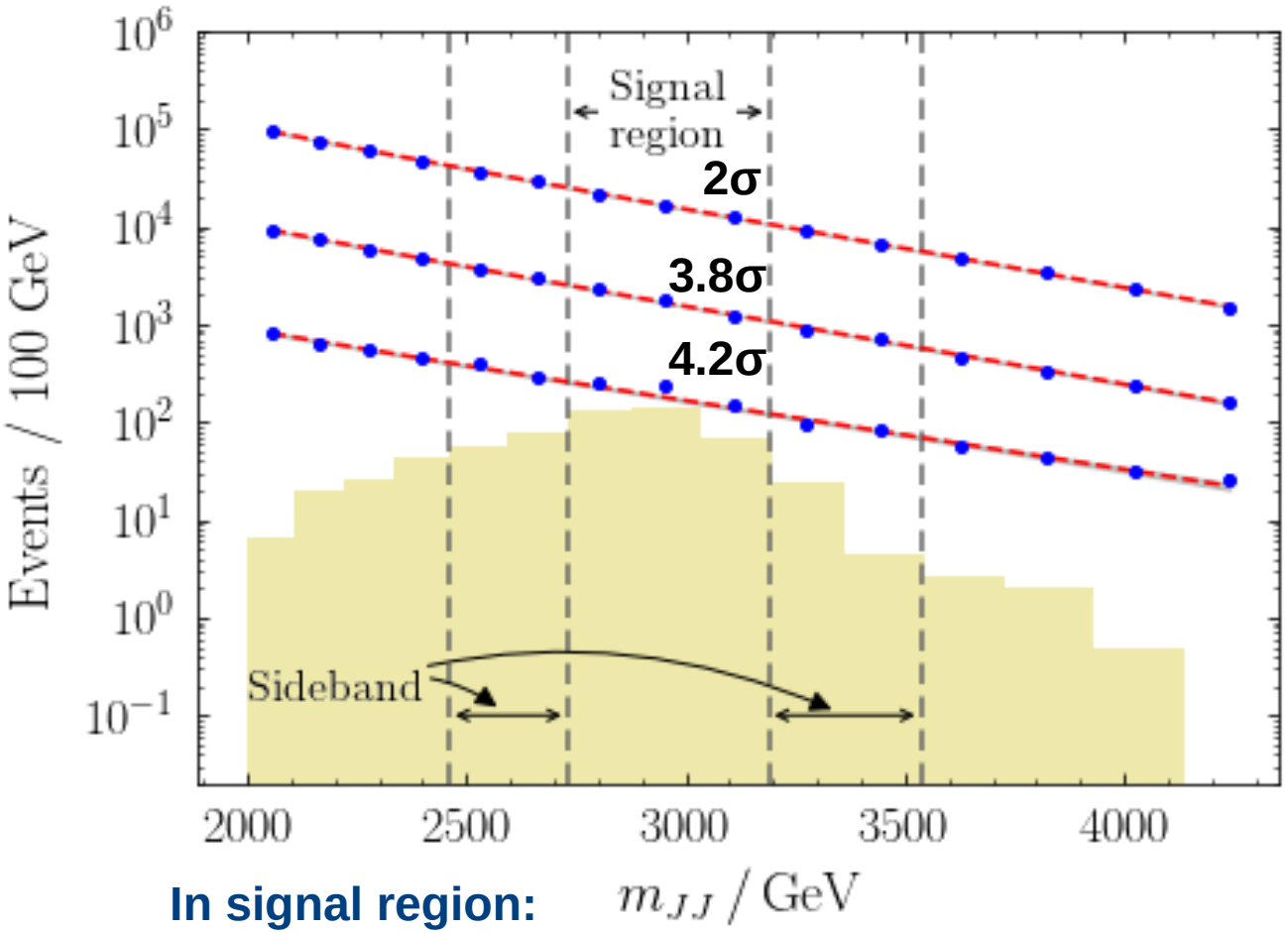
In signal region:
S = 522,
S/B = 0.64%

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



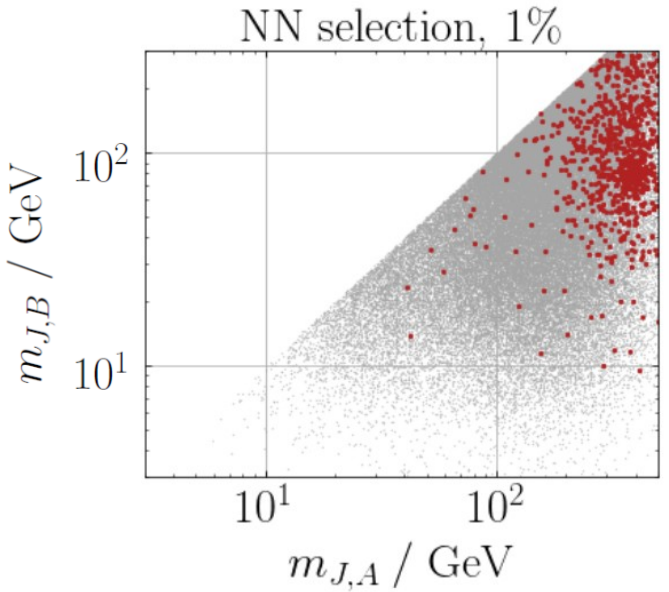


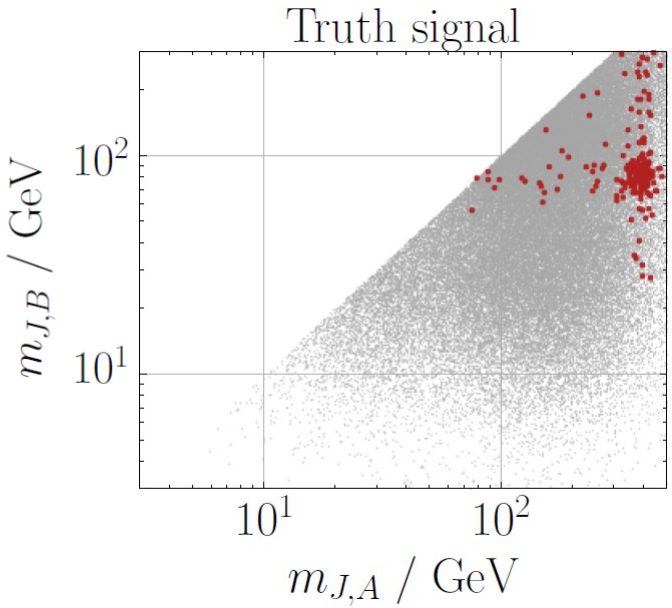
Application to Bump Hunt



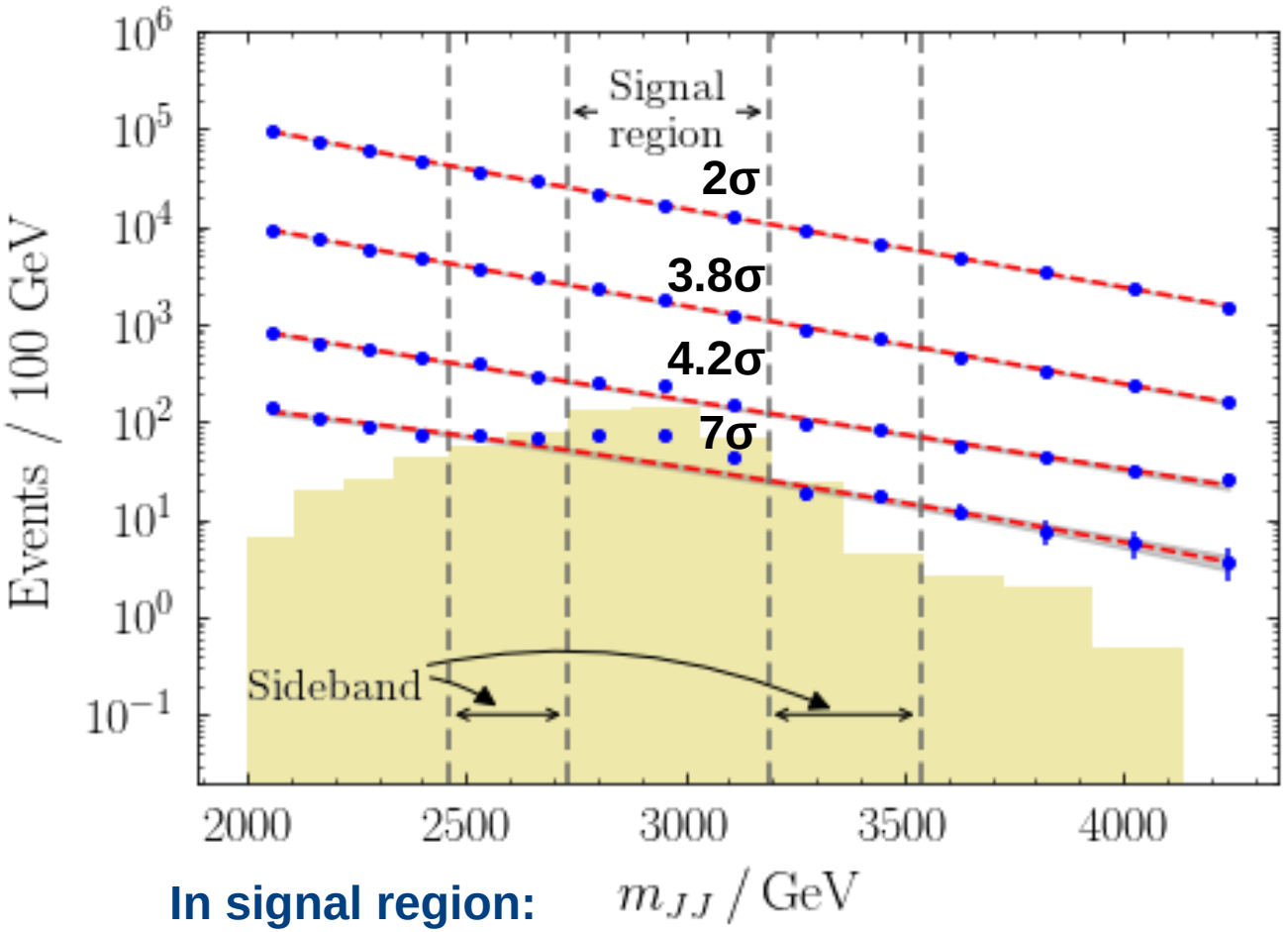
In signal region:
S = 522,
S/B = 0.64%

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



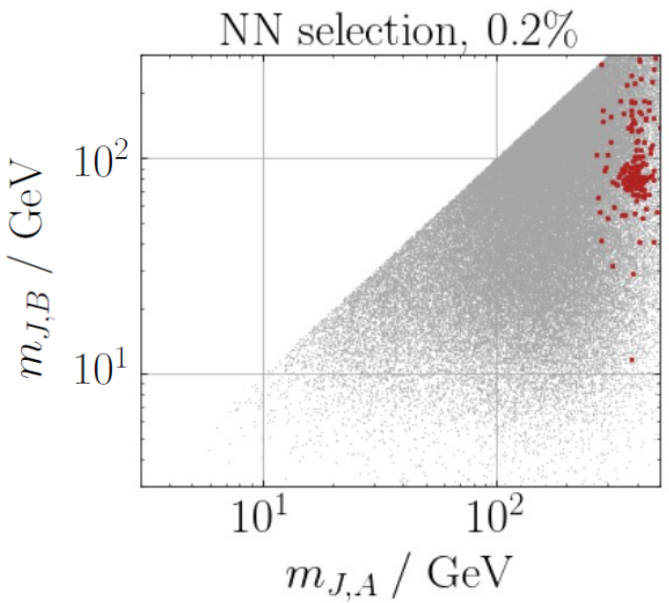


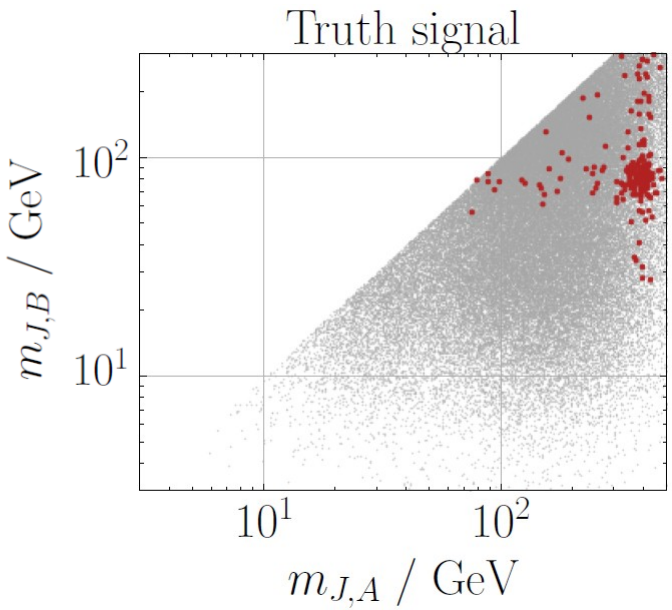
Application to Bump Hunt



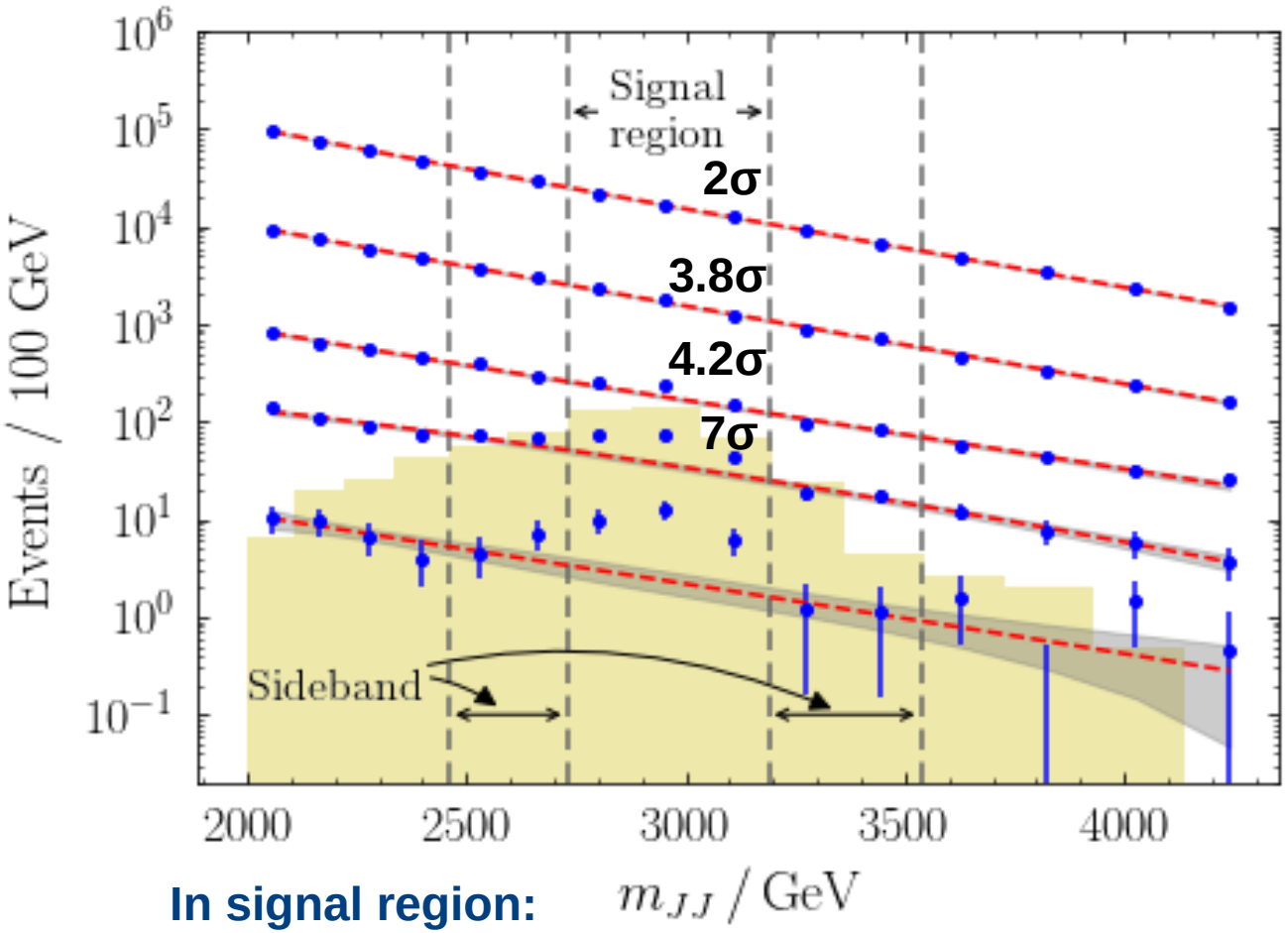
In signal region:
S = 522,
S/B = 0.64%

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



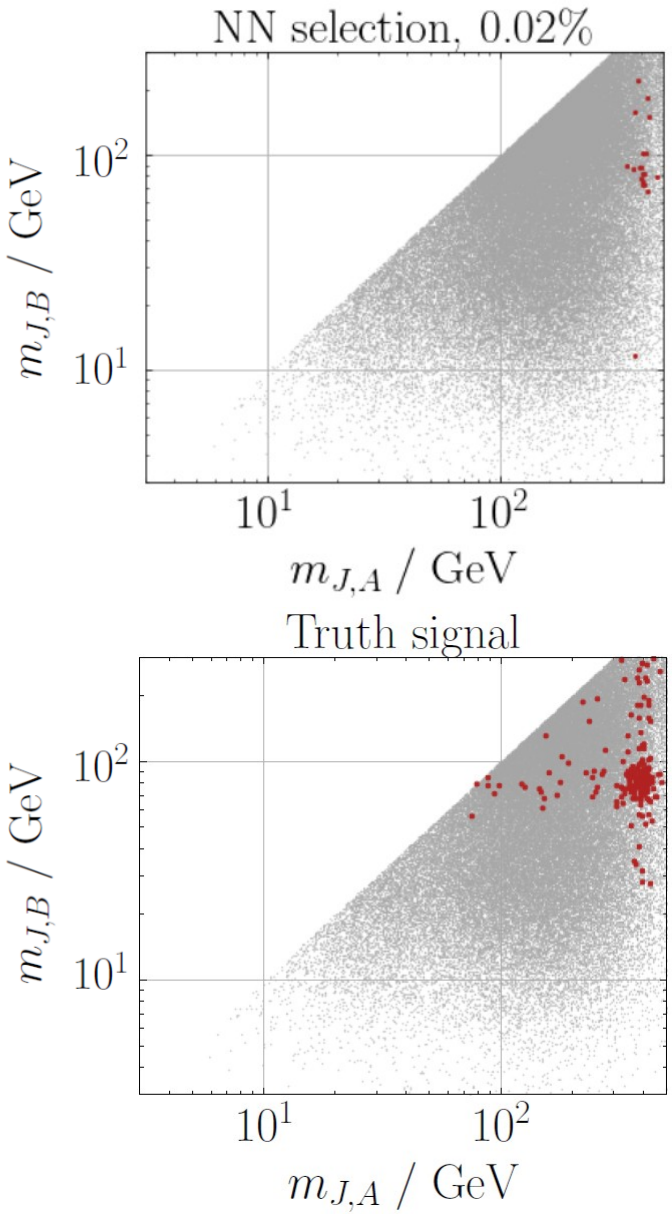


Application to Bump Hunt



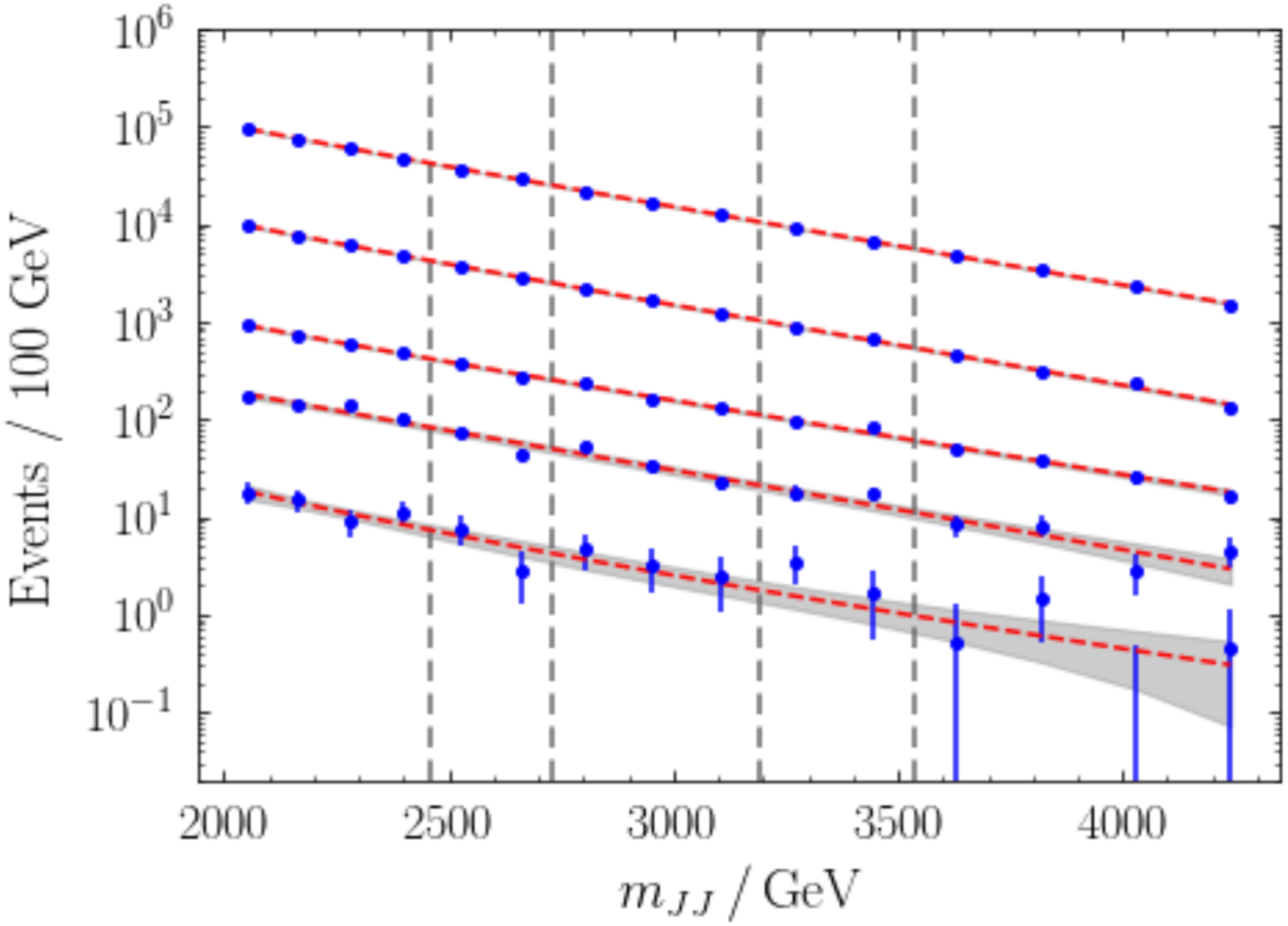
In signal region:
S = 522,
S/B = 0.64%

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

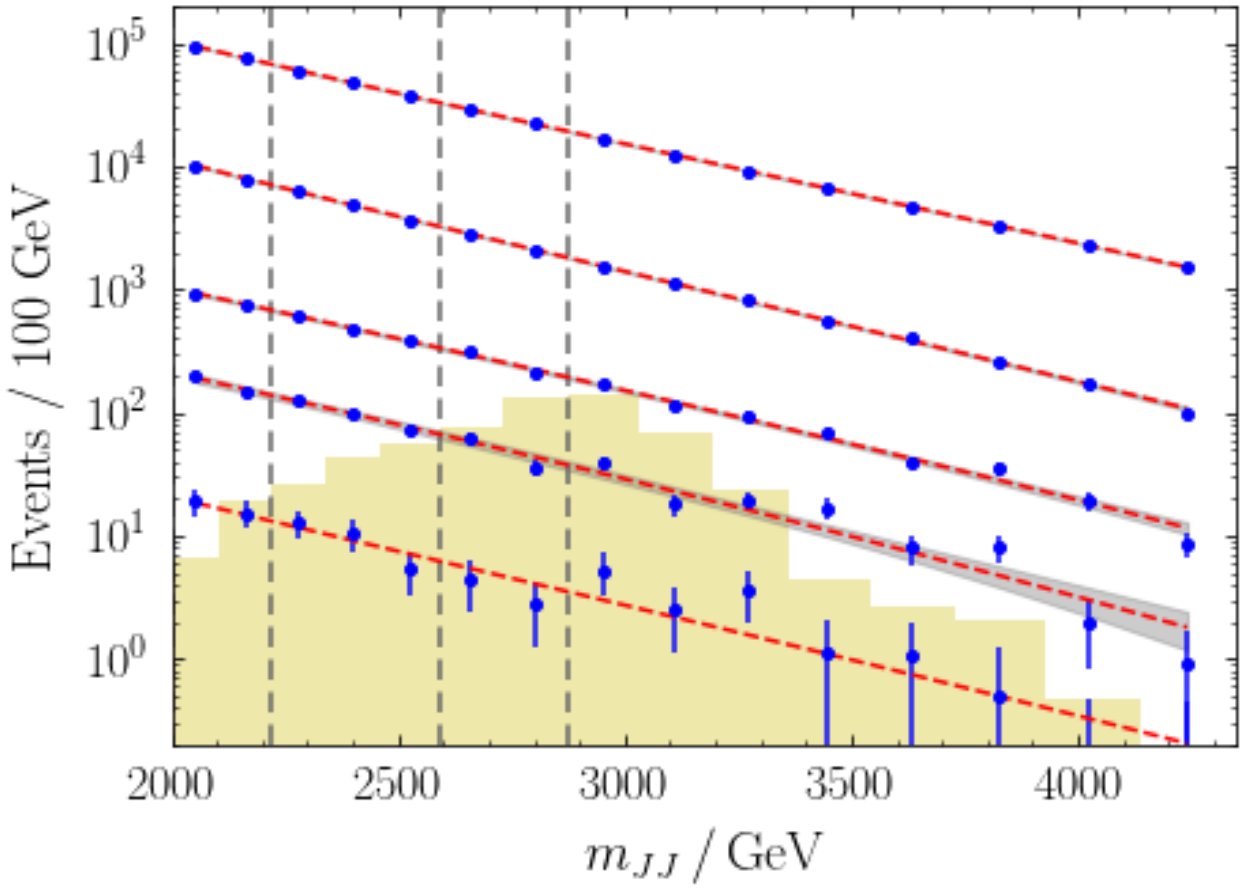


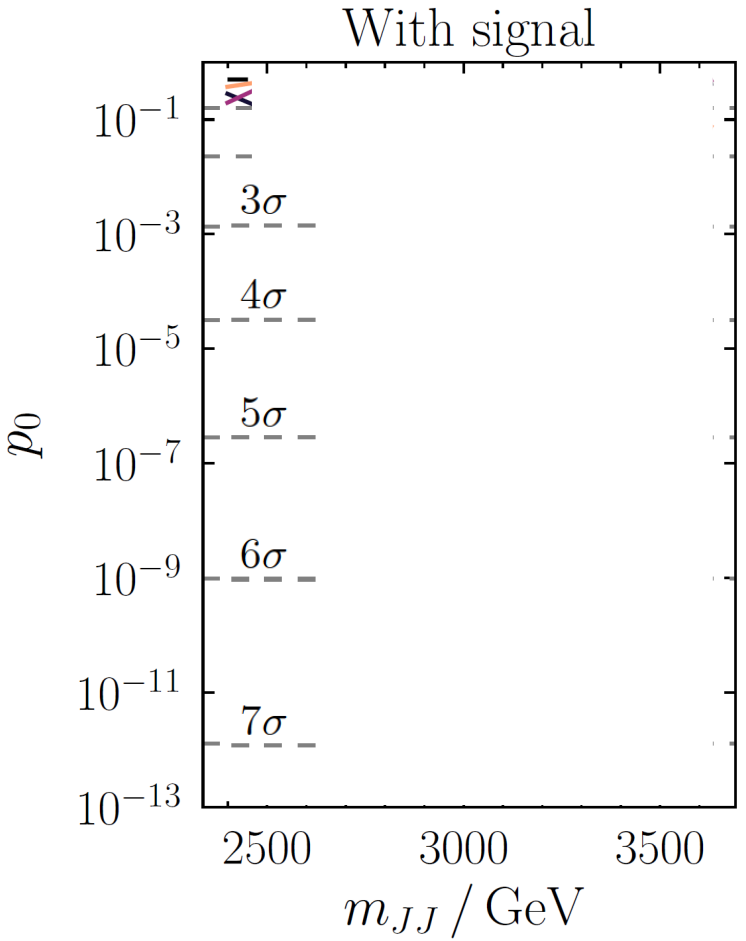
No Signal → No Bump!

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

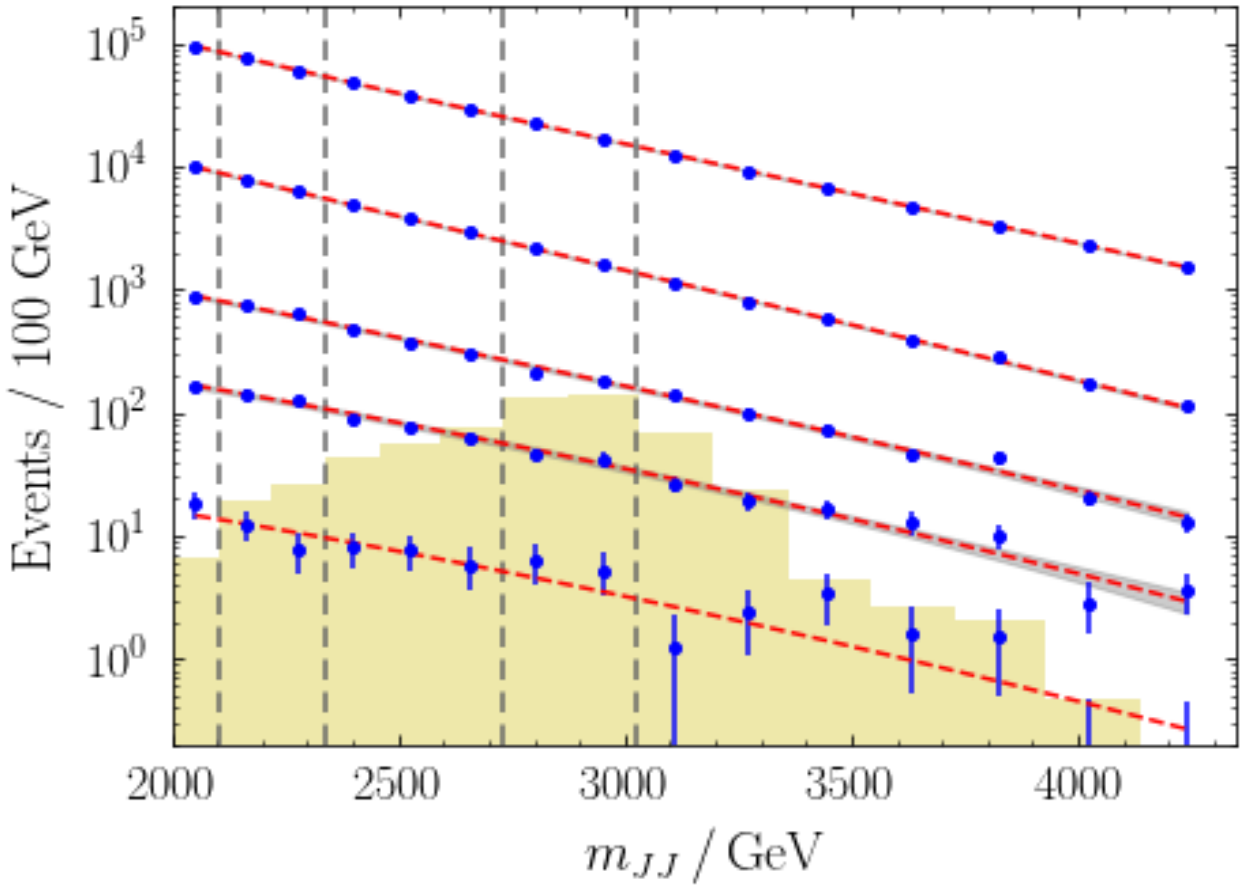


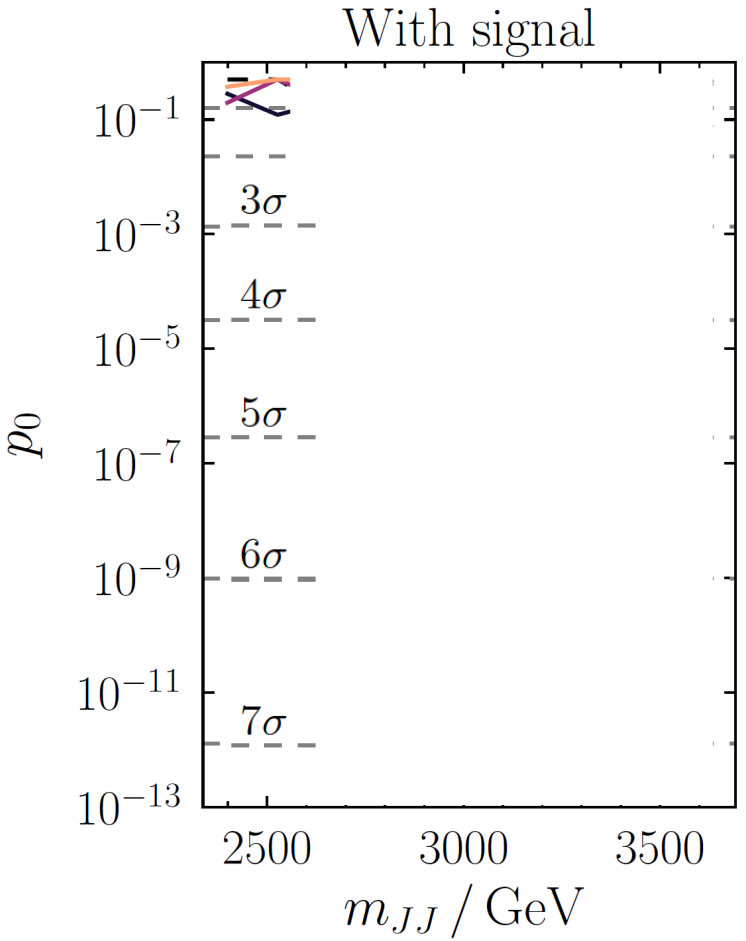
Mass Scan


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

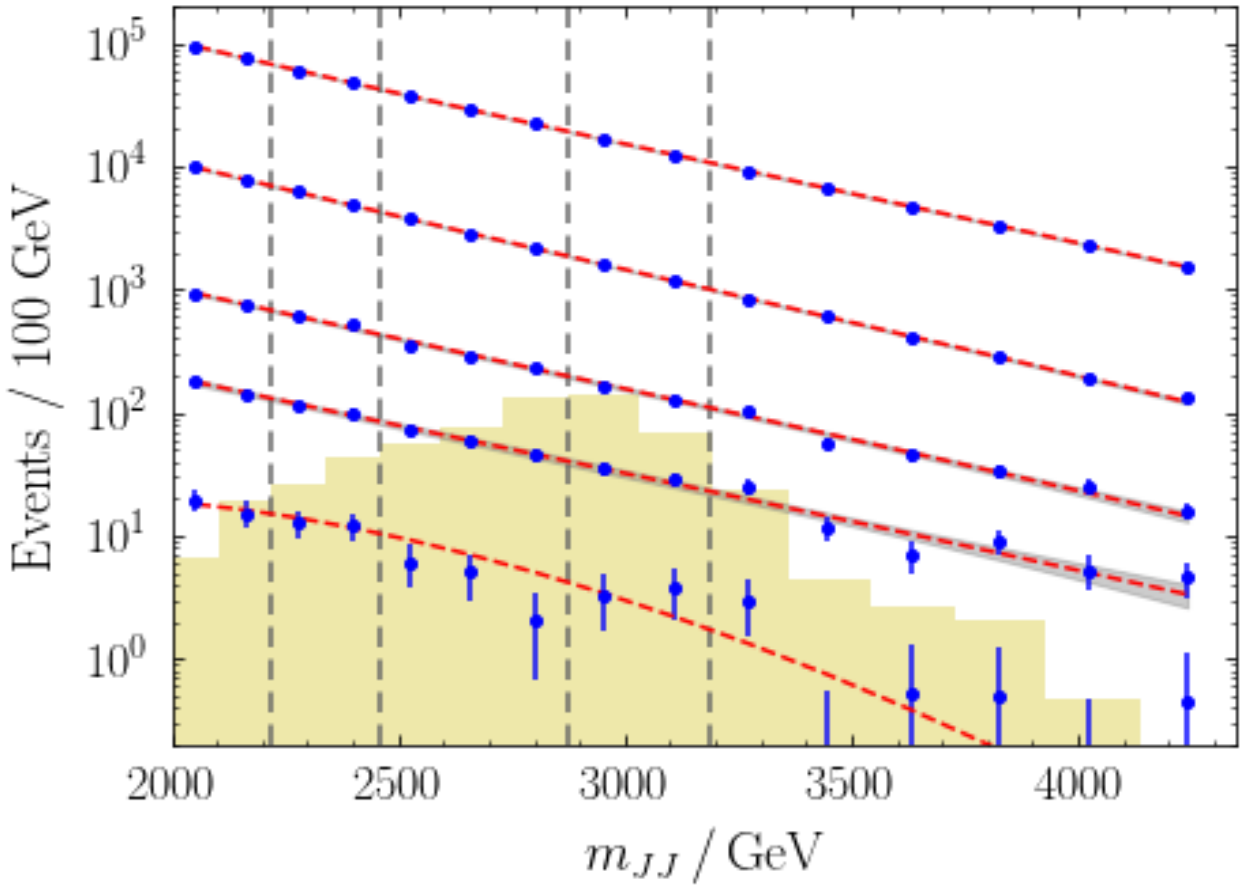

With signal
p_0
10^{-1}
10^{-3}
10^{-5}
10^{-7}
10^{-9}
10^{-11}
10^{-13}
3σ
4σ
5σ
6σ
7σ
2500
3000
3500
m_{JJ} / GeV

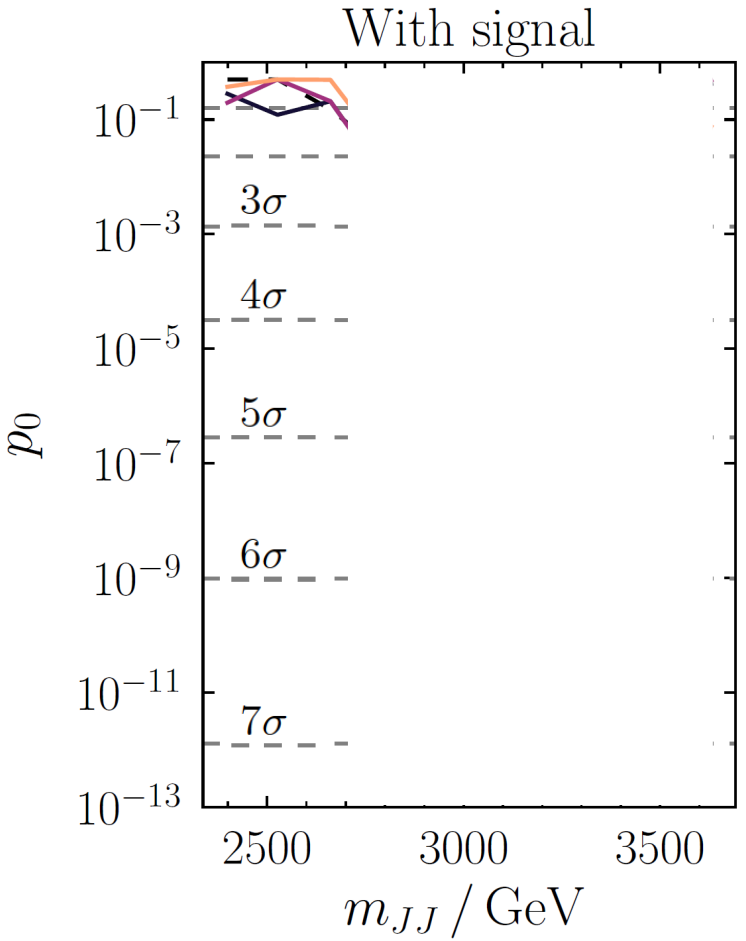
Mass Scan


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

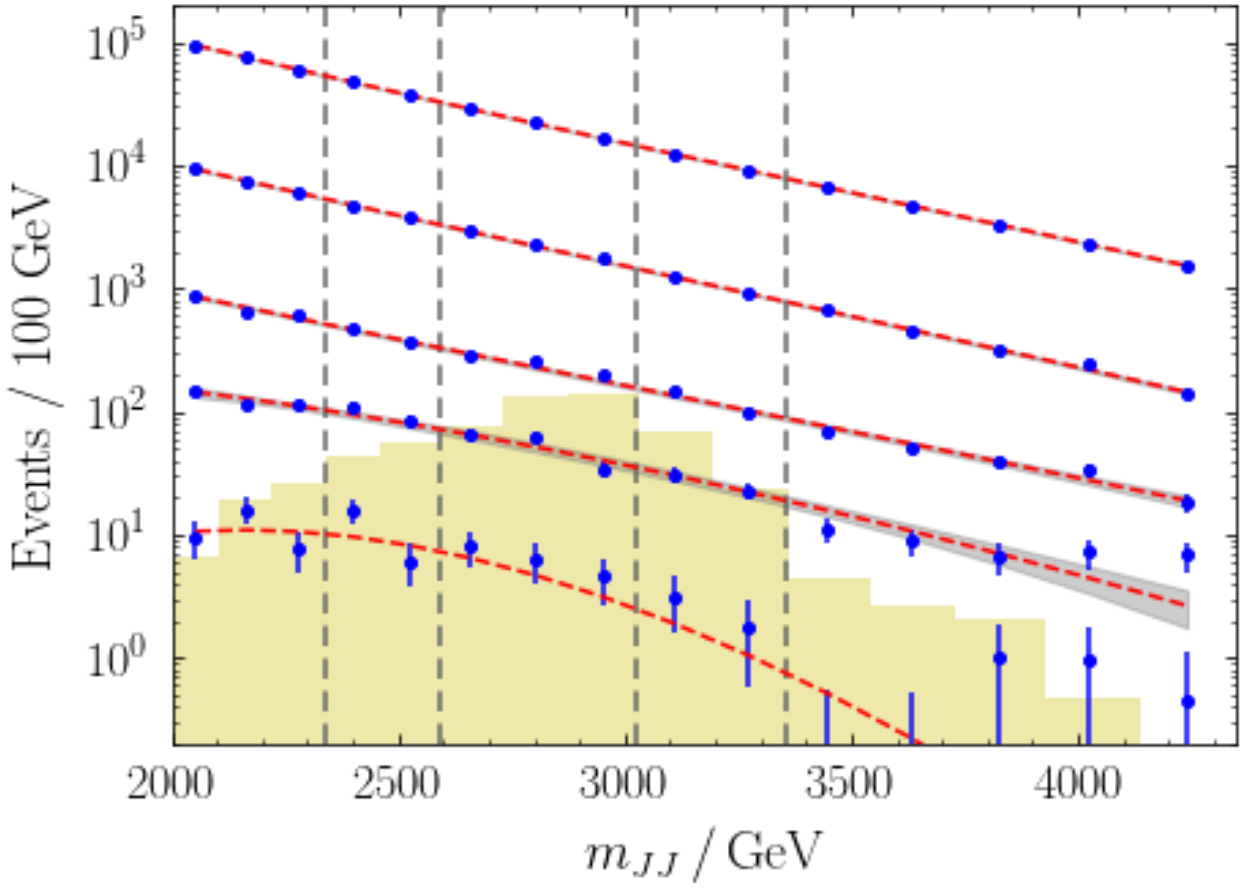

With signal
p_0
10^{-1}
10^{-3}
10^{-5}
10^{-7}
10^{-9}
10^{-11}
10^{-13}
3σ
4σ
5σ
6σ
7σ
2500
3000
3500
m_{JJ} / GeV

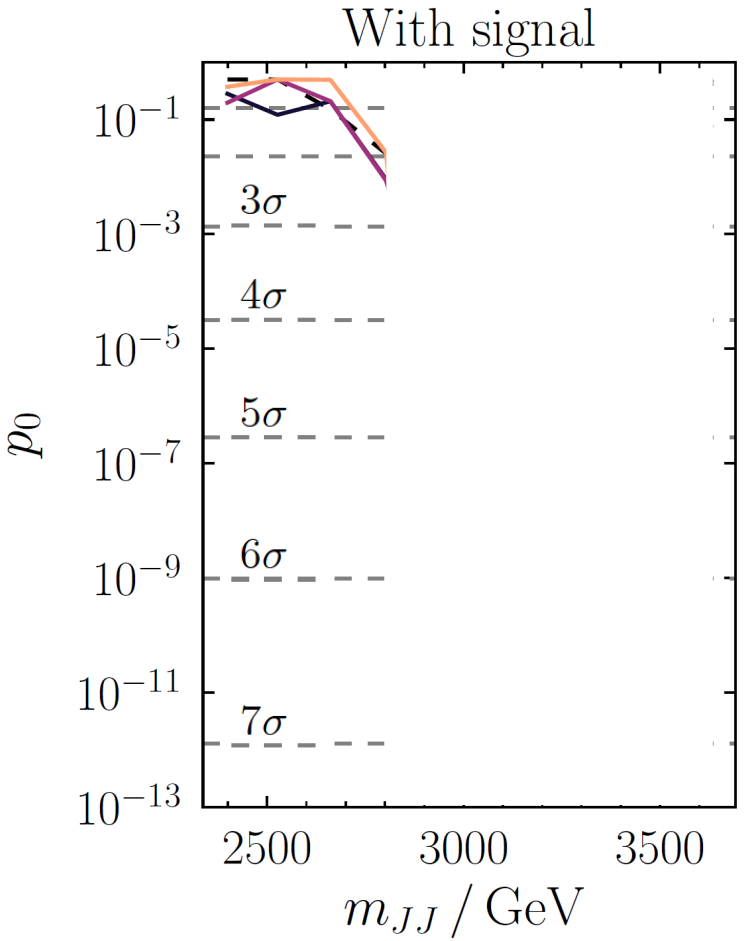
Mass Scan


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

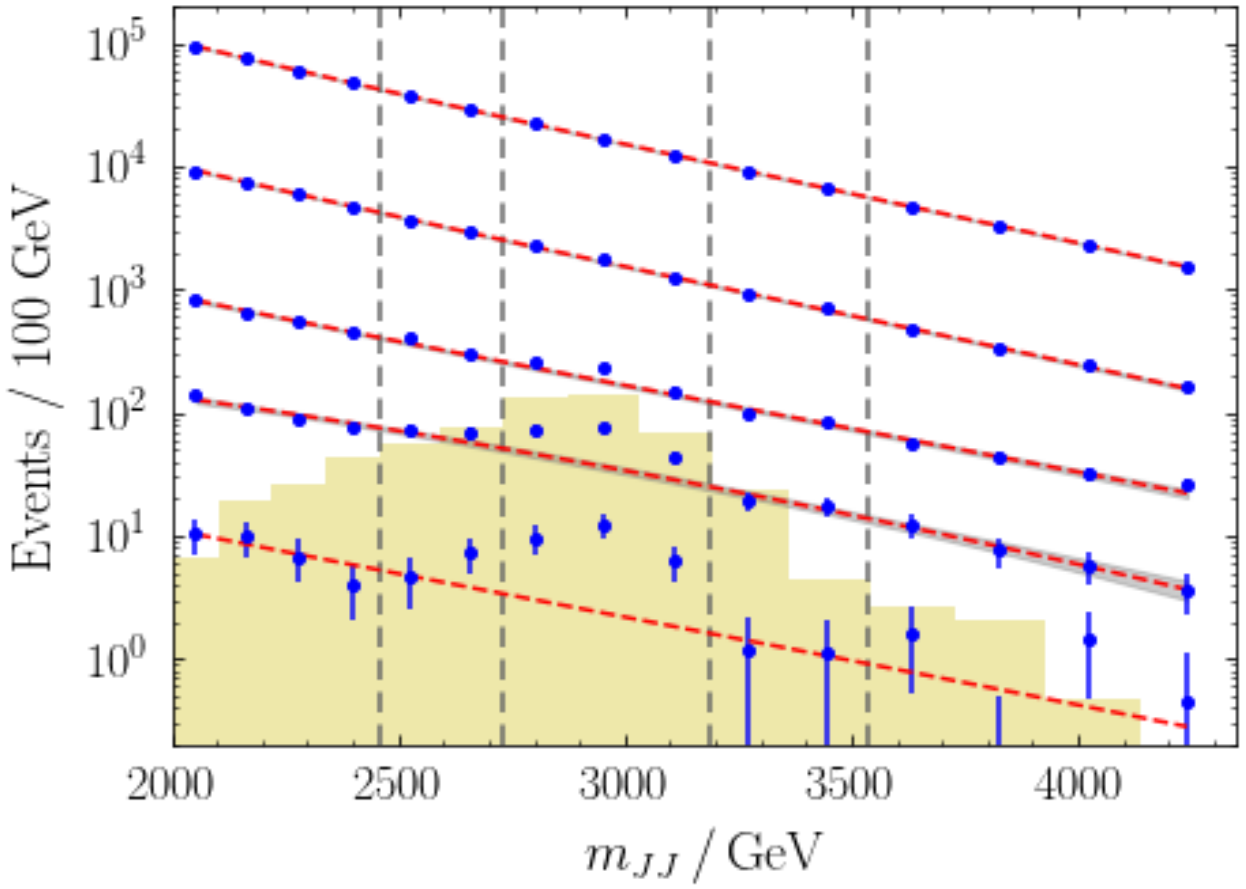

With signal
p_0
10^{-1}
10^{-3}
10^{-5}
10^{-7}
10^{-9}
10^{-11}
10^{-13}
3σ
4σ
5σ
6σ
7σ
2500
3000
3500
m_{JJ} / GeV

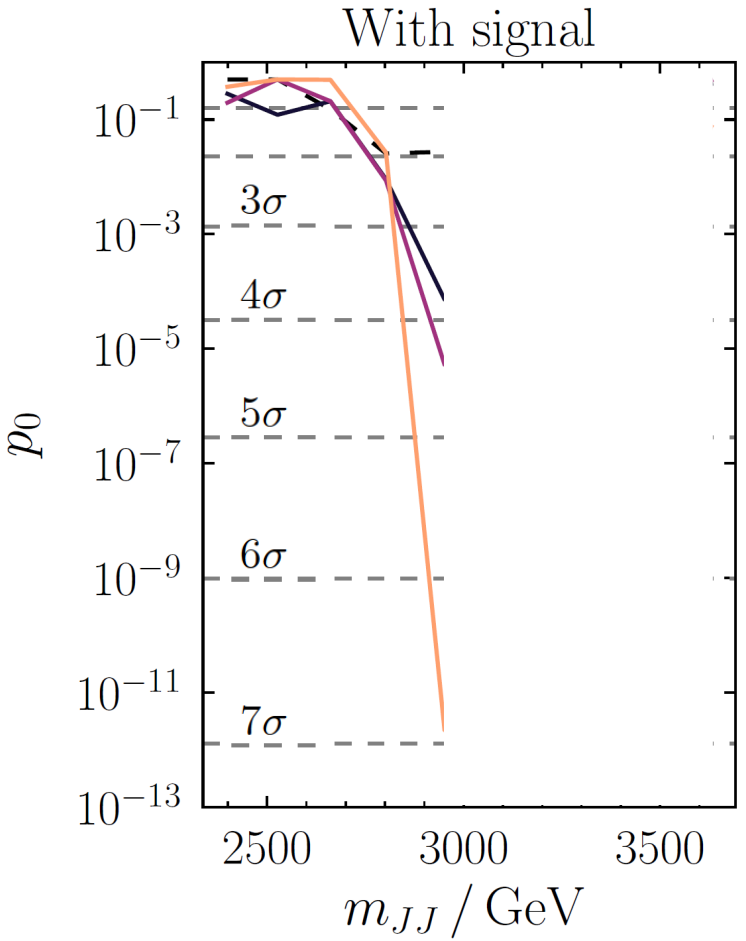
Mass Scan


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

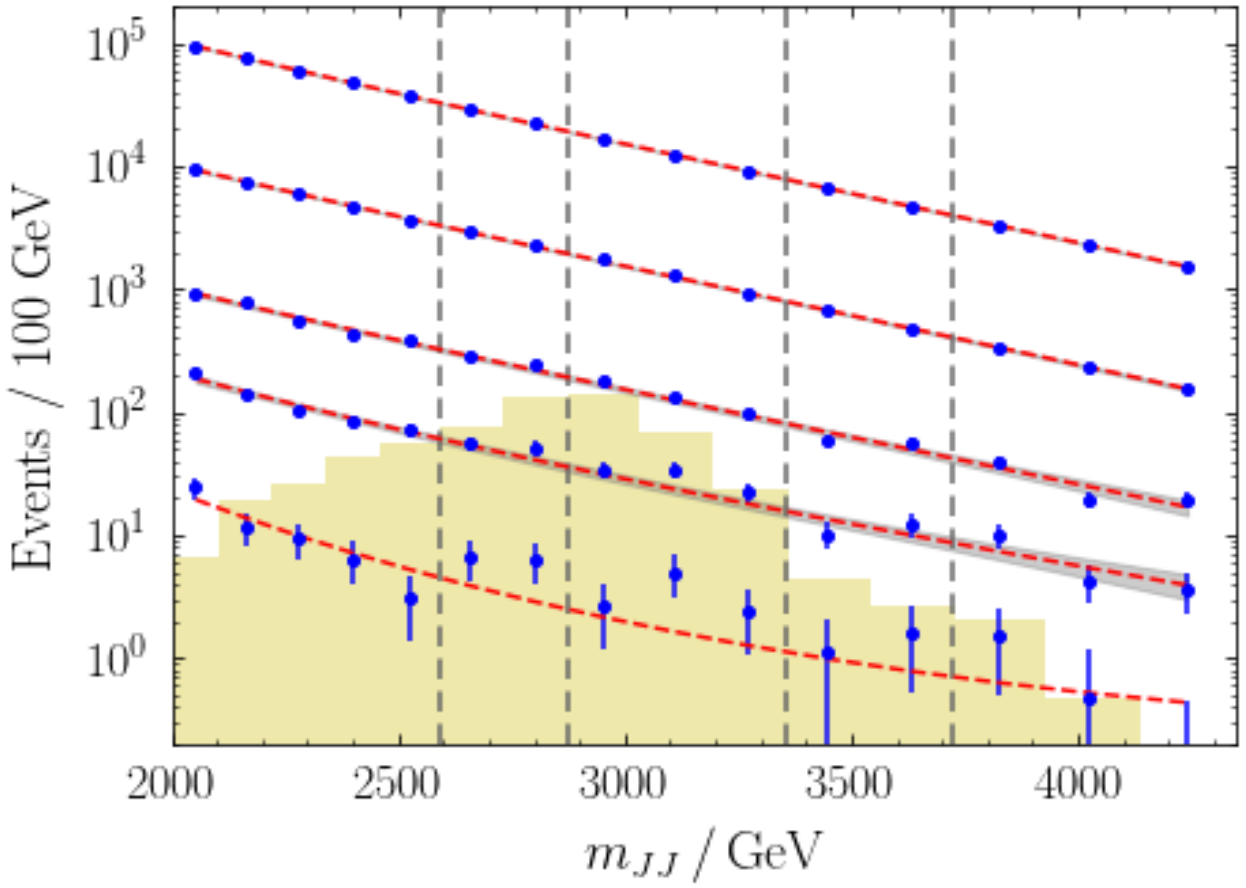

With signal
p_0
10^{-1}
10^{-3}
10^{-5}
10^{-7}
10^{-9}
10^{-11}
10^{-13}
3σ
4σ
5σ
6σ
7σ
2500
3000
3500
m_{JJ} / GeV

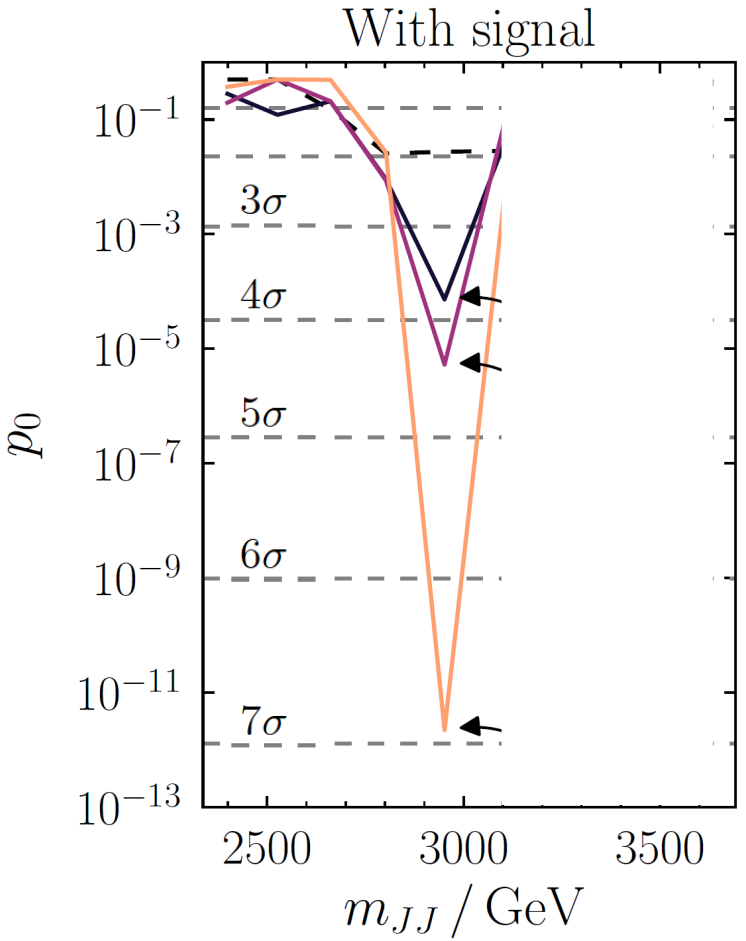
Mass Scan


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

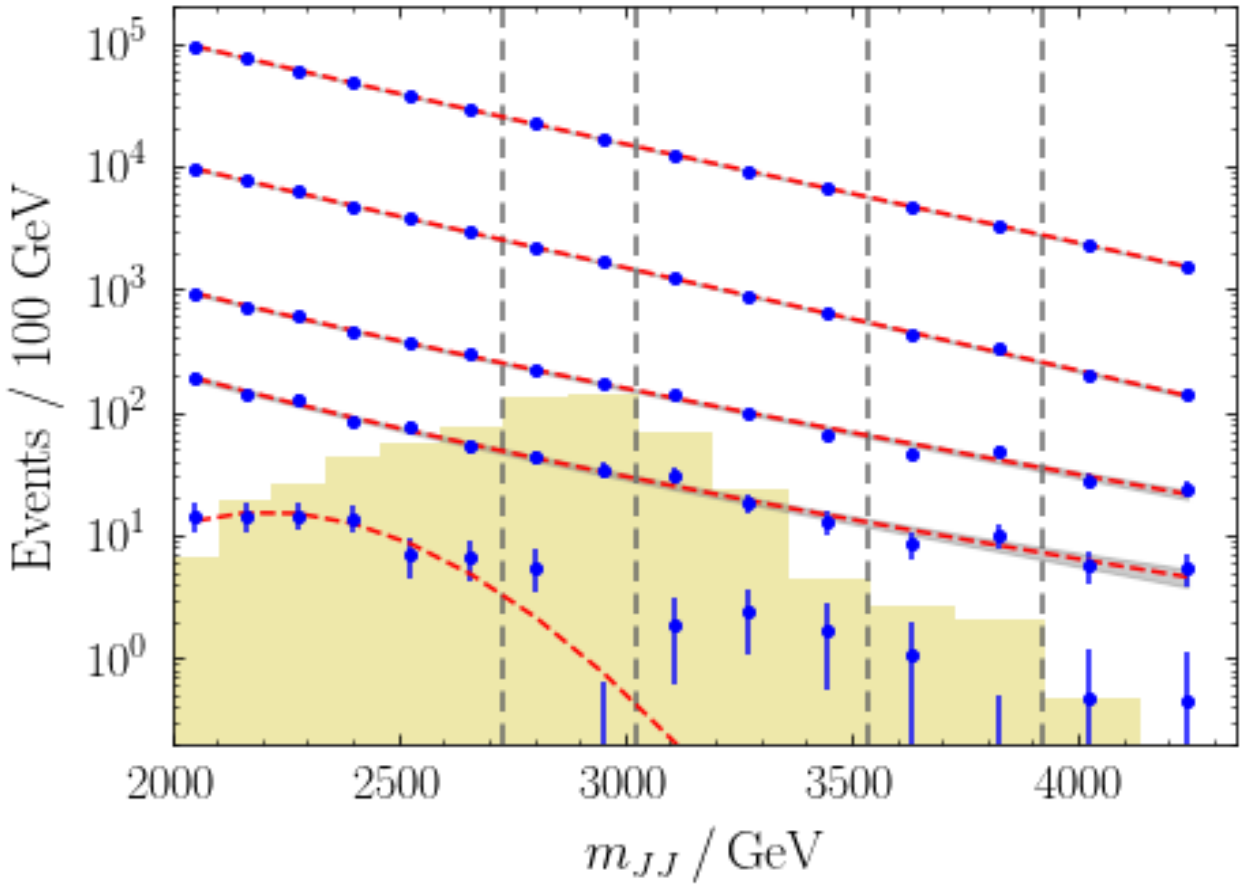

With signal
p_0
10^{-1}
10^{-3}
10^{-5}
10^{-7}
10^{-9}
10^{-11}
10^{-13}
3σ
4σ
5σ
6σ
7σ
2500
3000
3500
m_{JJ} / GeV

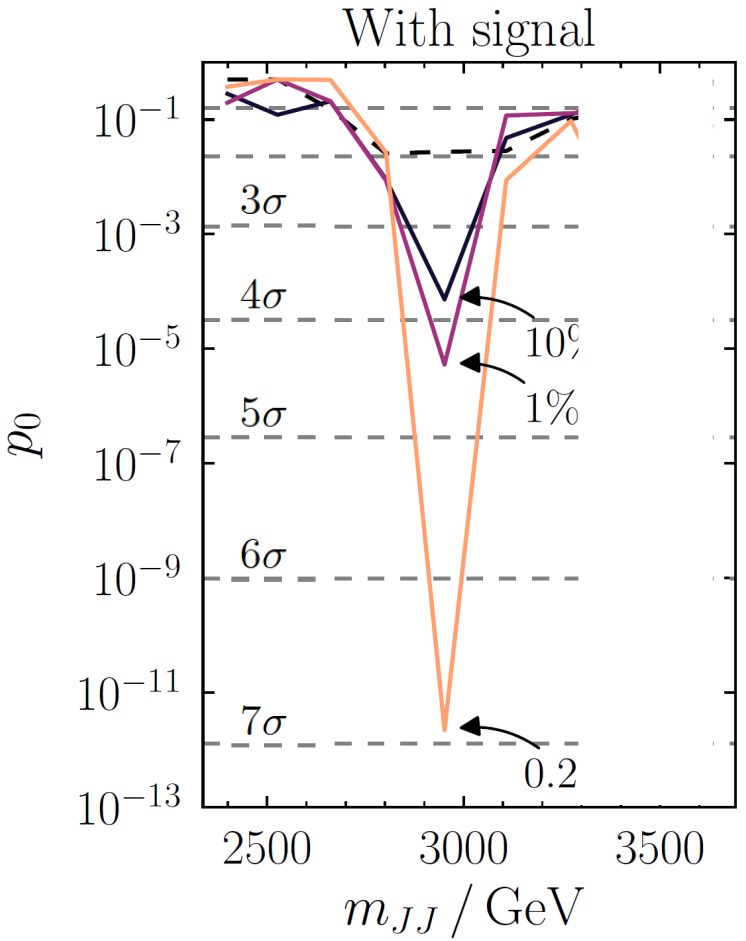
Mass Scan


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

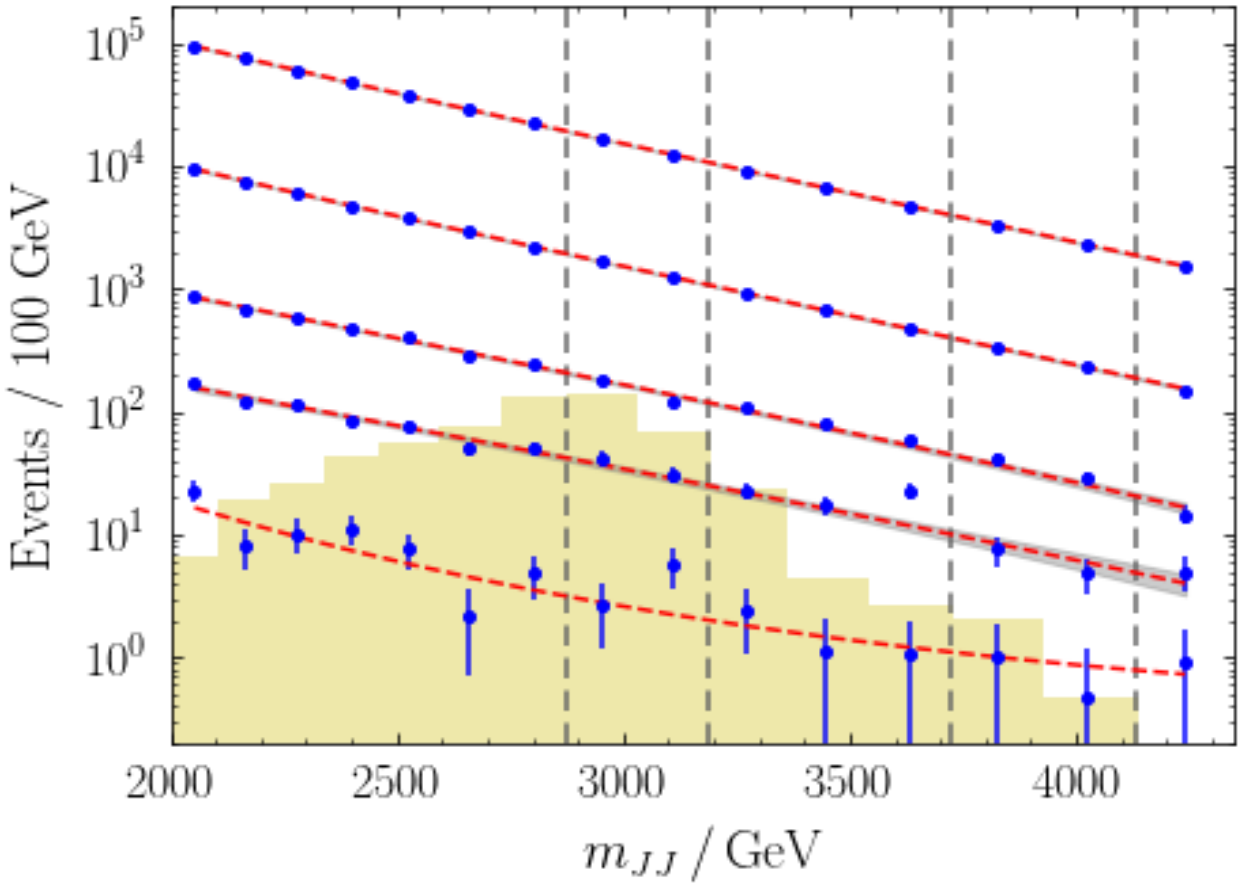

With signal
p_0
10^{-1}
10^{-3}
10^{-5}
10^{-7}
10^{-9}
10^{-11}
10^{-13}
3σ
4σ
5σ
6σ
7σ
2500
3000
3500
m_{JJ} / GeV

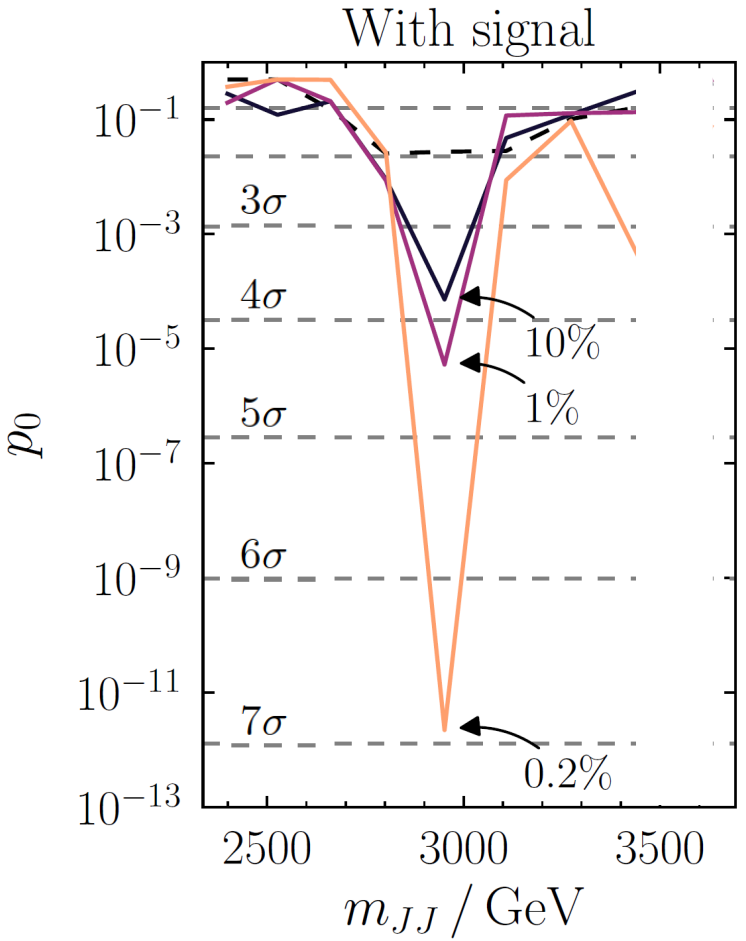
Mass Scan


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

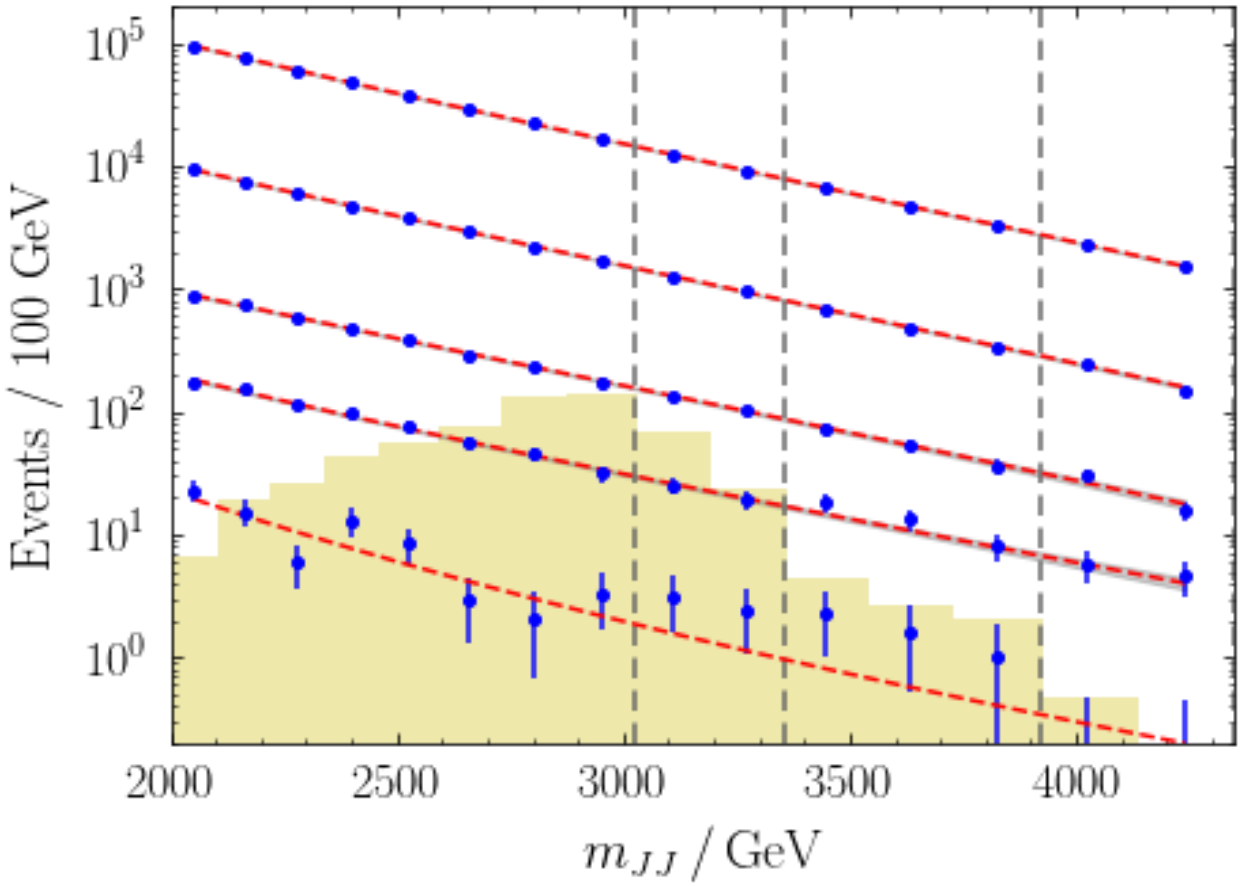

With signal
p_0
10^{-1}
10^{-3}
10^{-5}
10^{-7}
10^{-9}
10^{-11}
10^{-13}
3σ
4σ
5σ
6σ
7σ
10%
1%
0.2
2500
3000
3500
m_{JJ} / GeV

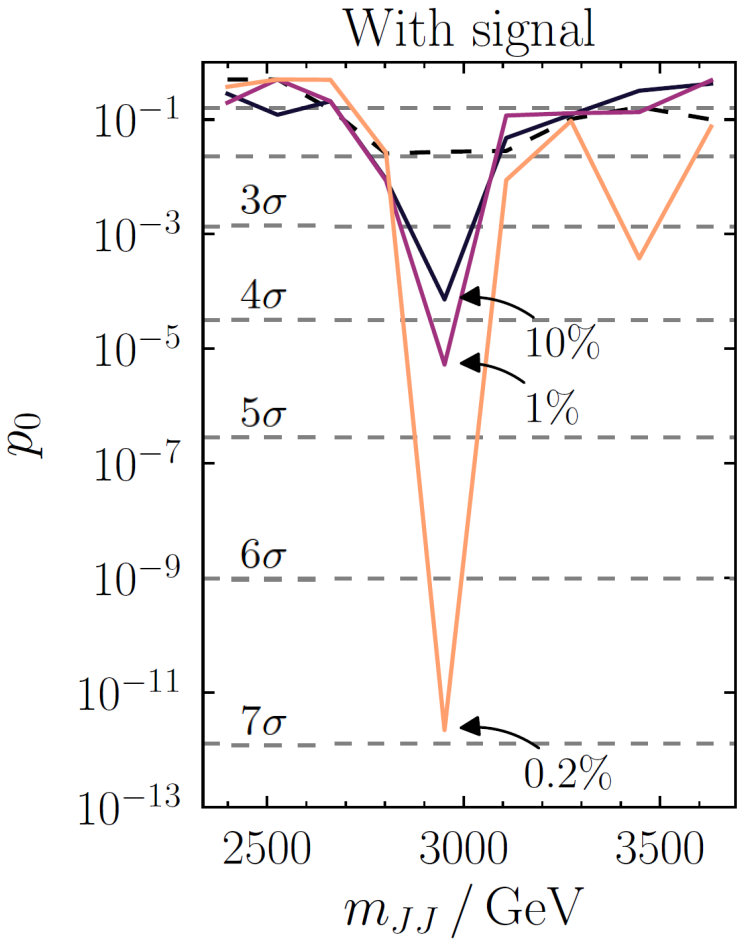
Mass Scan


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

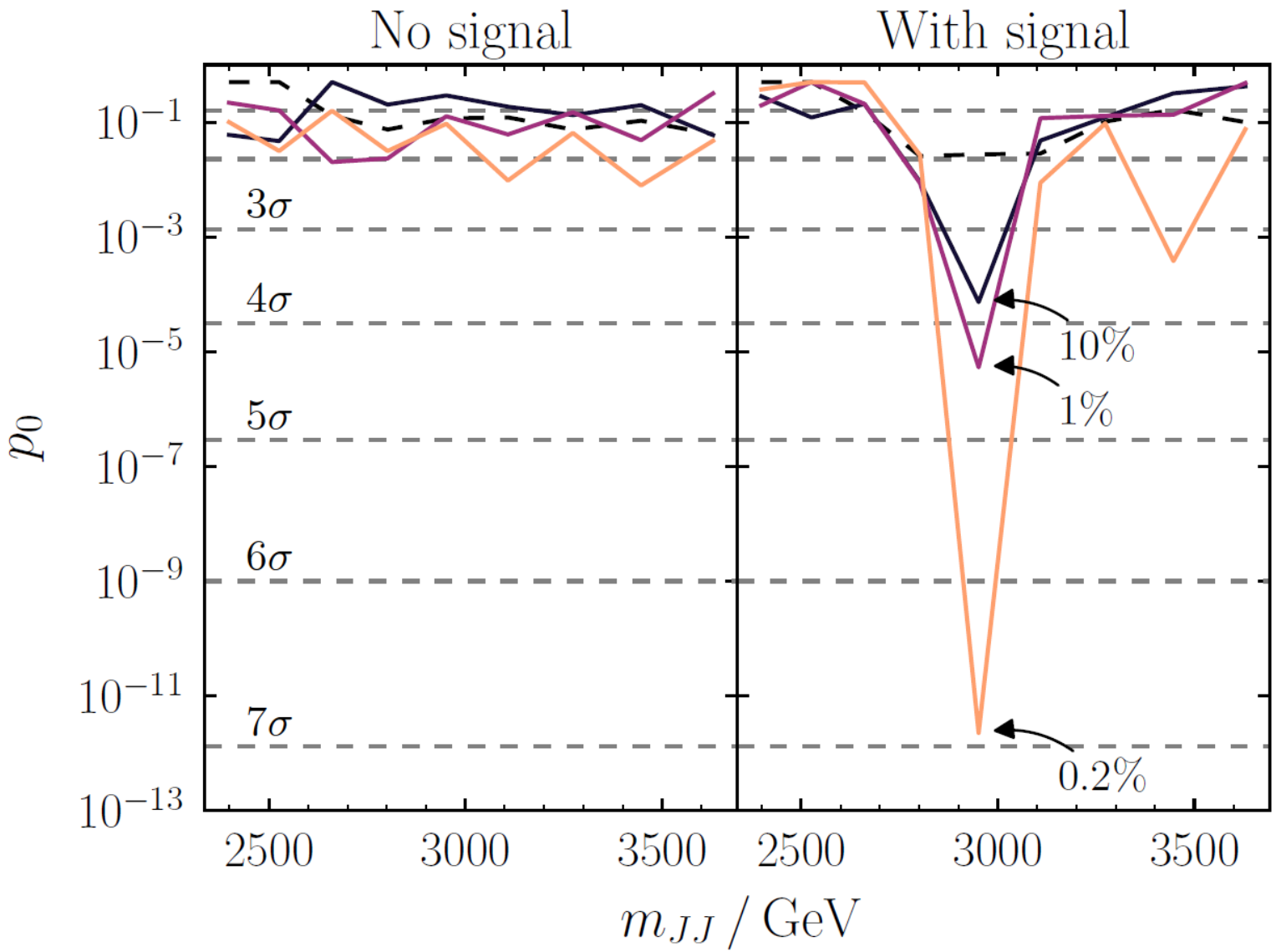

With signal
p_0
10^{-1}
10^{-3}
10^{-5}
10^{-7}
10^{-9}
10^{-11}
10^{-13}
3σ
4σ
5σ
6σ
7σ
10%
1%
0.2%
2500
3000
3500
m_{JJ} / GeV

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^{-1}
10^{-3}
10^{-5}
10^{-7}
10^{-9}
10^{-11}
10^{-13}
3σ
4σ
5σ
6σ
7σ
10%
1%
0.2%
2500
3000
3500
m_{JJ} / GeV

Mass Scan


No signal
With signal
p_0
10^{-1}
10^{-3}
10^{-5}
10^{-7}
10^{-9}
10^{-11}
10^{-13}
3σ
4σ
5σ
6σ
7σ
10%
1%
0.2%
2500
3000
3500
2500
3000
3500
m_{JJ} / GeV

General CWoLa Hunting

1) **Need some variable X (e.g. m_JJ) in which bg 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) **{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).

Performance Comparison

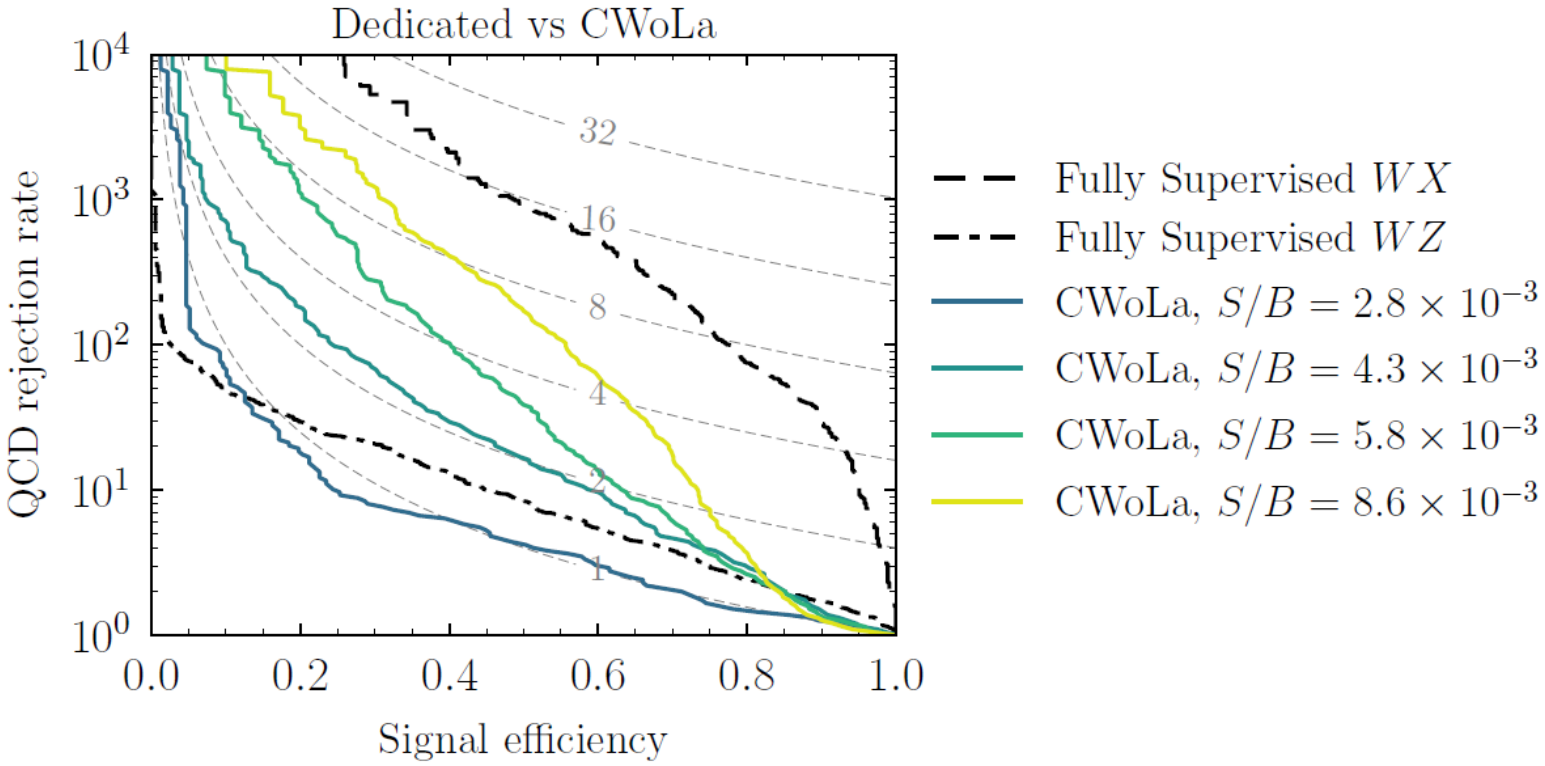


Figure 11. Truth-label ROC curves for taggers trained using CWoLa with varying number of signal events, compared to those for a dedicated tagger trained on pure signal and background samples (solid black) and one trained to discriminate W and Z jets from QCD (dashed black). The CWoLa examples have $B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