CWoLa Hunting:

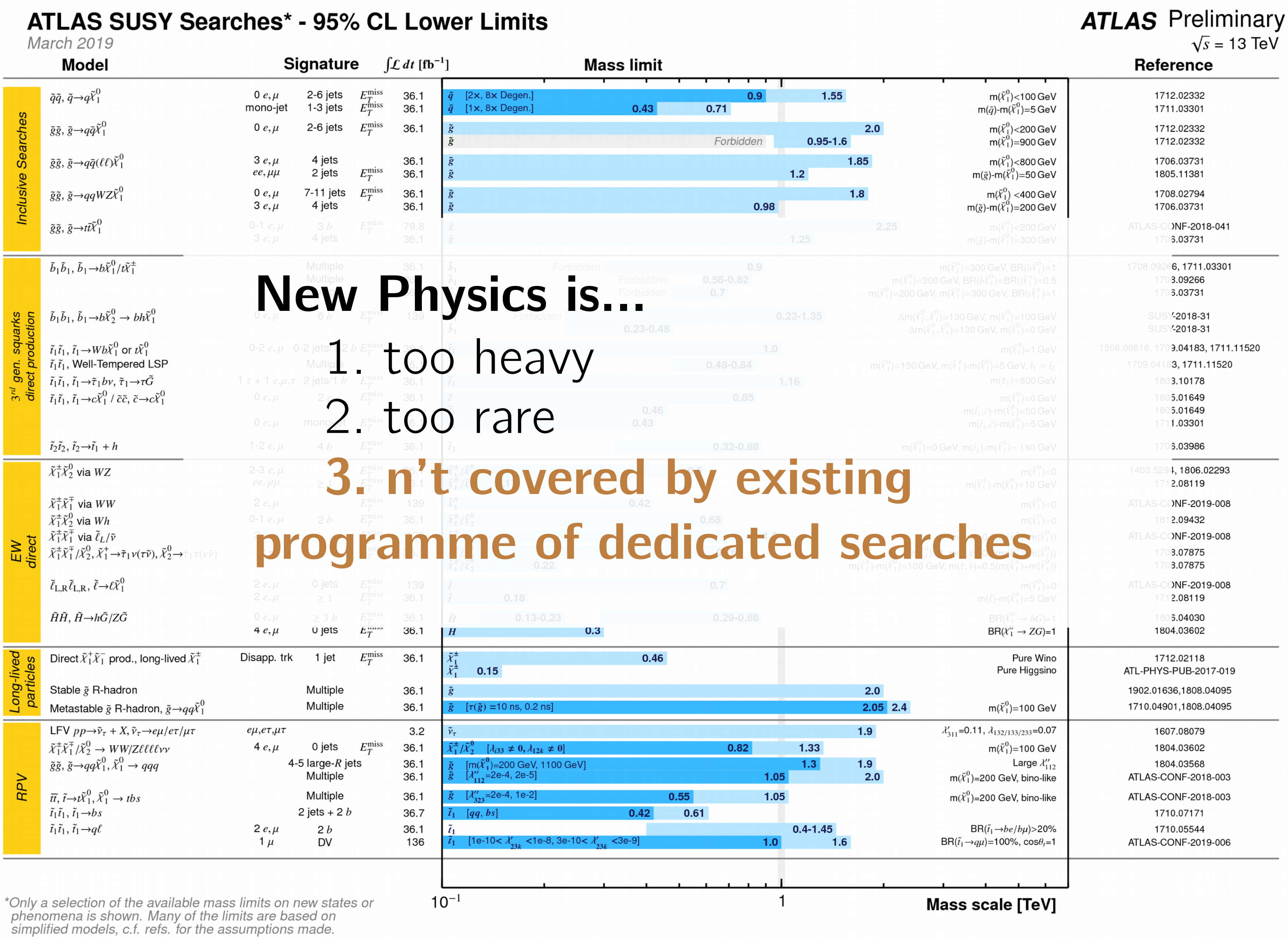
Extending the Bump Hunt with Machine Learning

Based on:

Phys. Rev. Lett. 121, 241803 (2018)

[1805.02664] **Jack Collins, Kiel Howe, Ben Nachman**

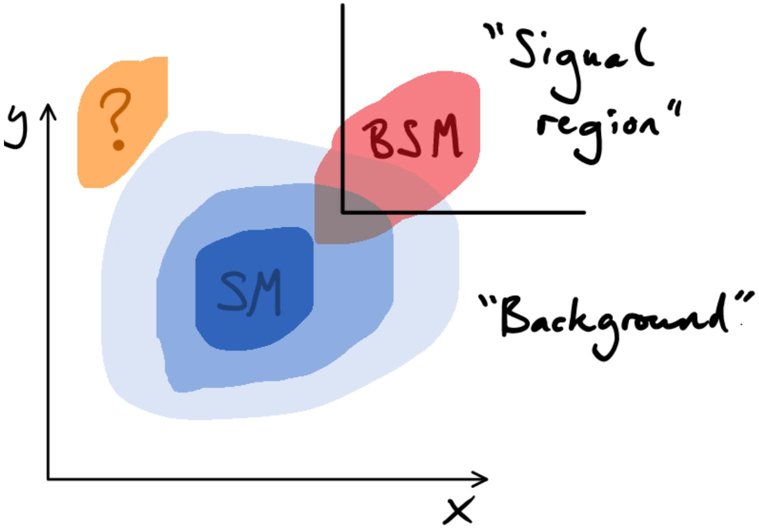
Is there new physics at the LHC?



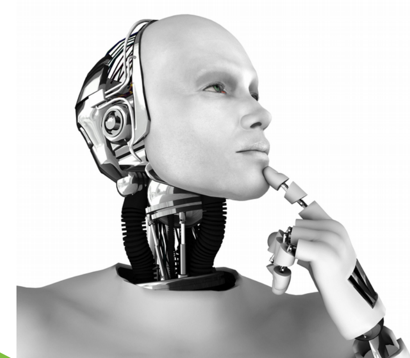
New Physics is…

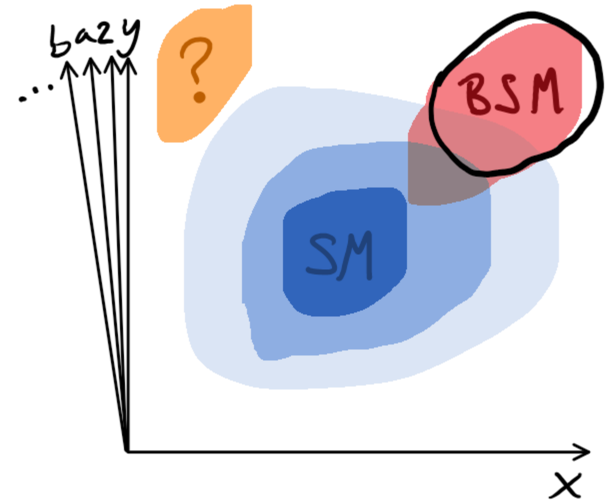
1. too heavy
2. too rare
3. n't covered by existing programme of dedicated searches

LHC Searches

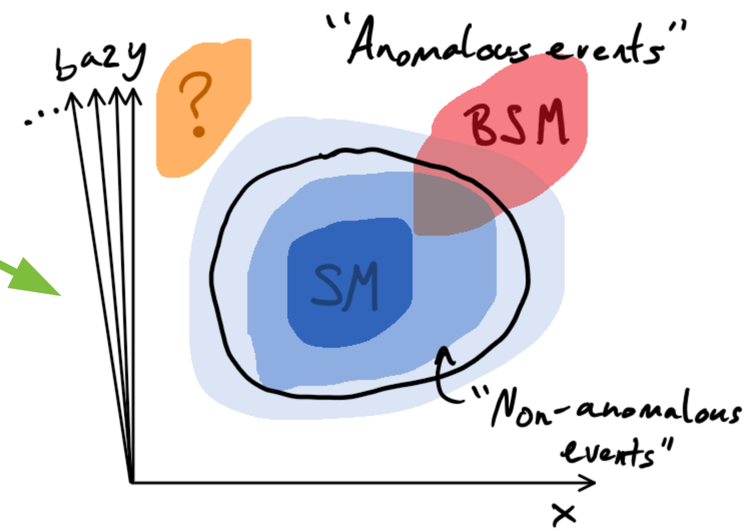


ML-optimised dedicated search

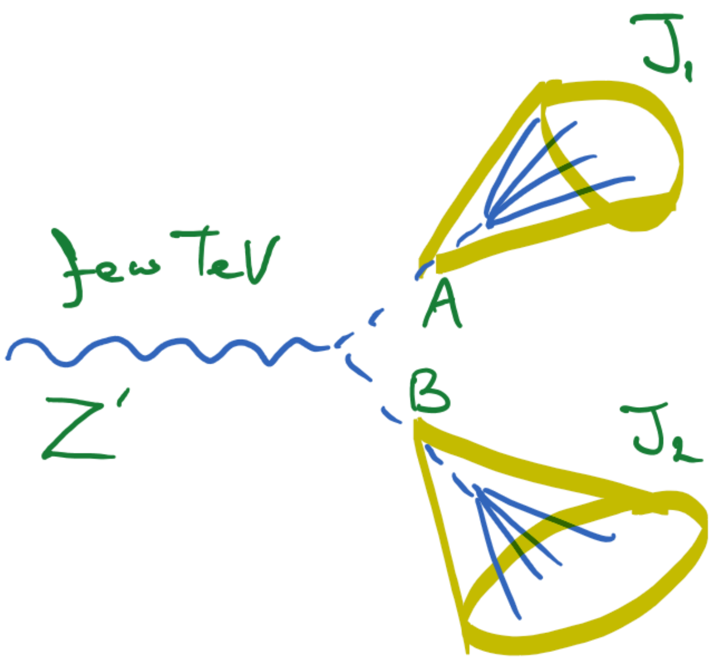




ML Anomaly Detection



An Example Target



$$A \to qq$$

$$A \to WW \to 4q$$

$$A \to t\bar{t} \to bbqqqq$$

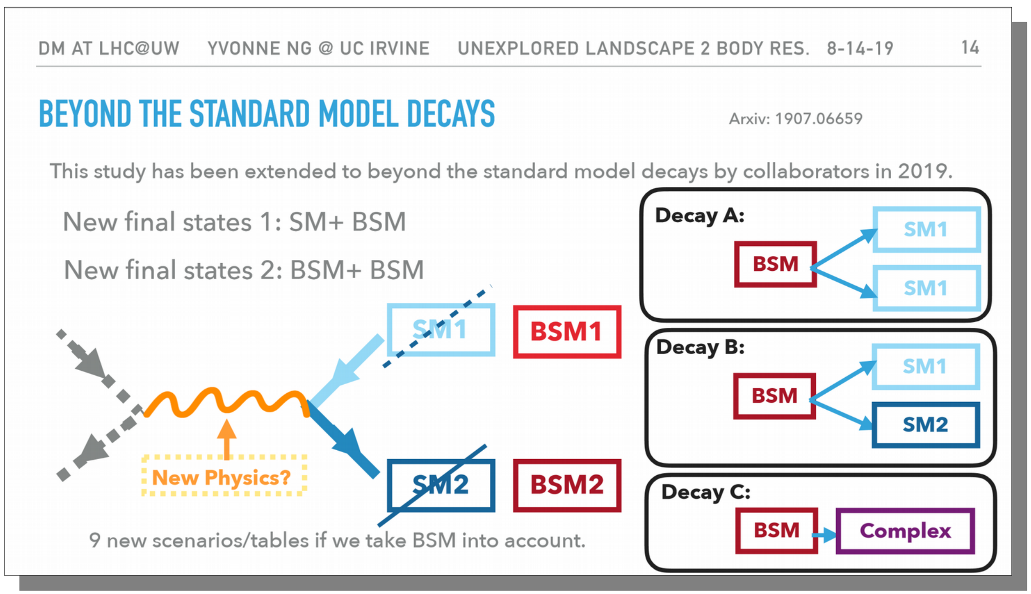
$$bbqq\ell\nu$$

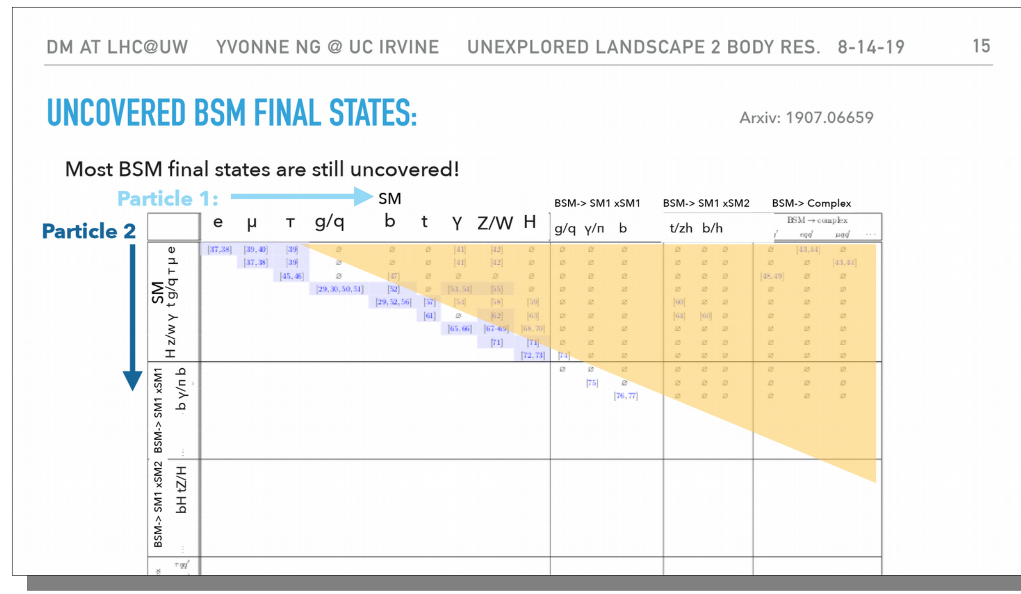
$$A \to 2\phi_1 \to 4\phi_2 \to \ldots \to Nq$$

$$A \to \text{Dark shower}$$

An Example Target

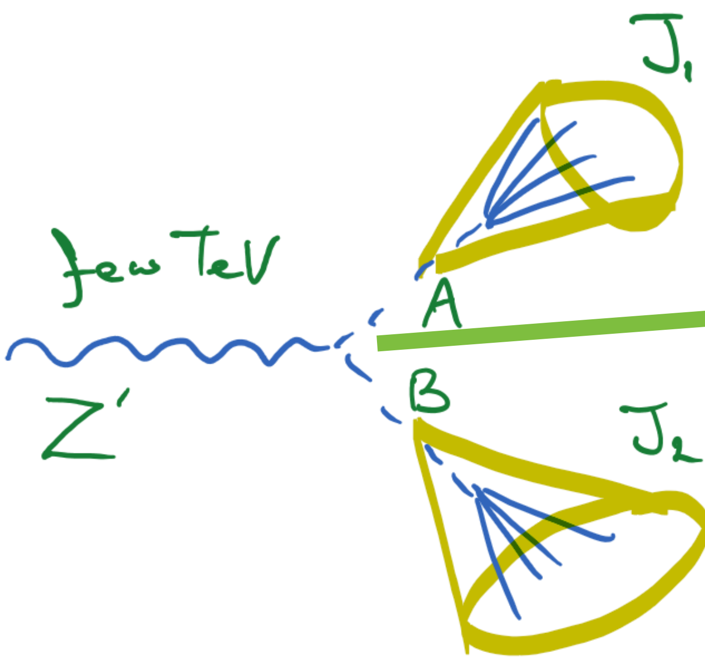
Yesterday from Yvonne

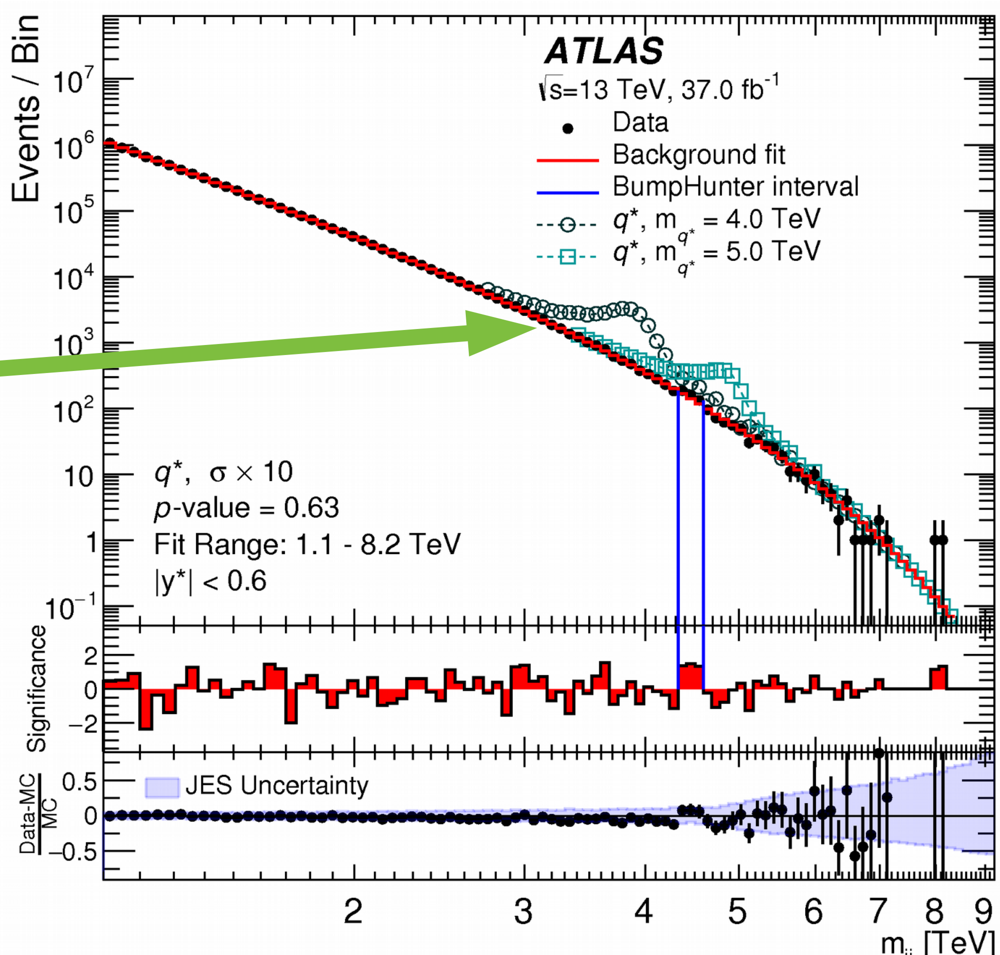




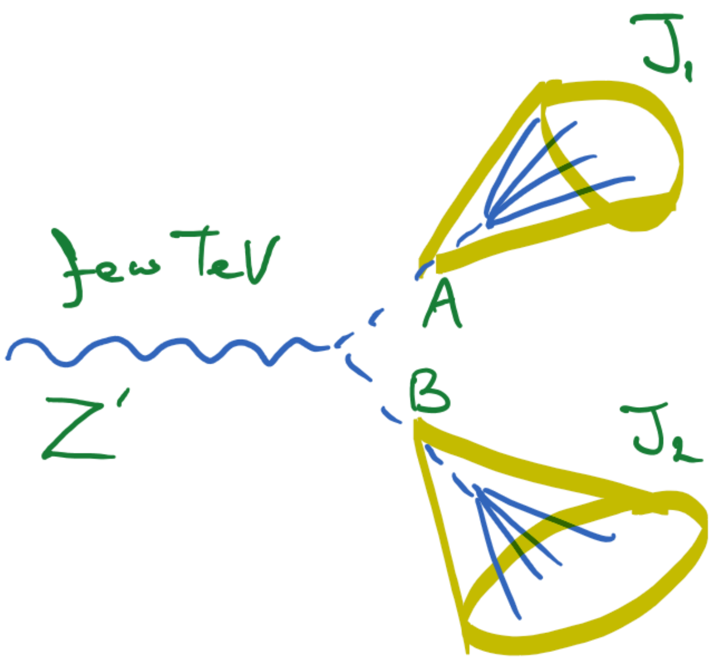
An Example Target

1. It is very simple (bump hunt)

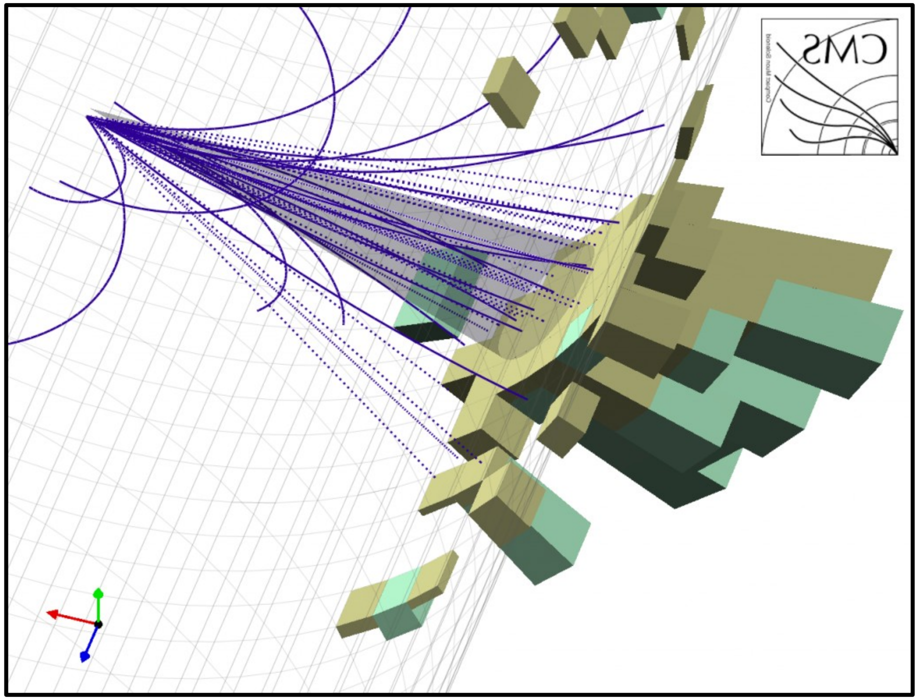




An Example Target



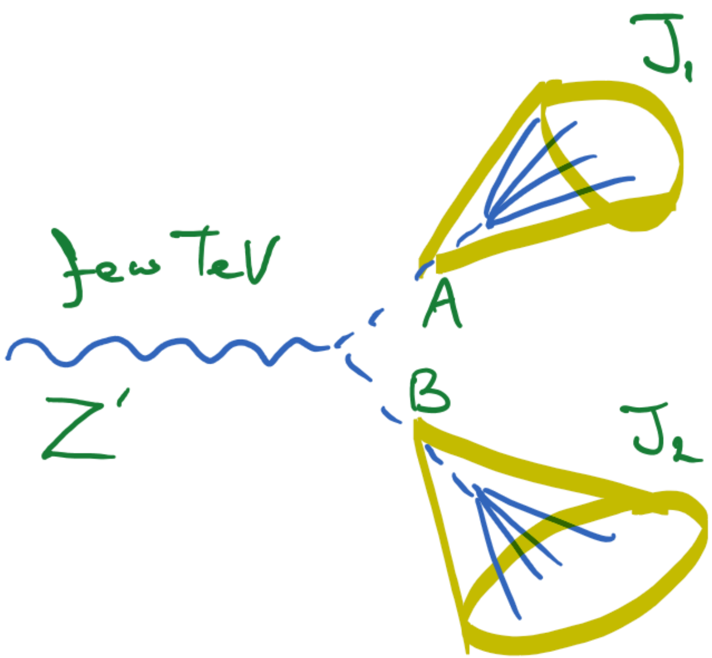
1. It is very simple (bump hunt)
2. It is very complex (jet substructure)



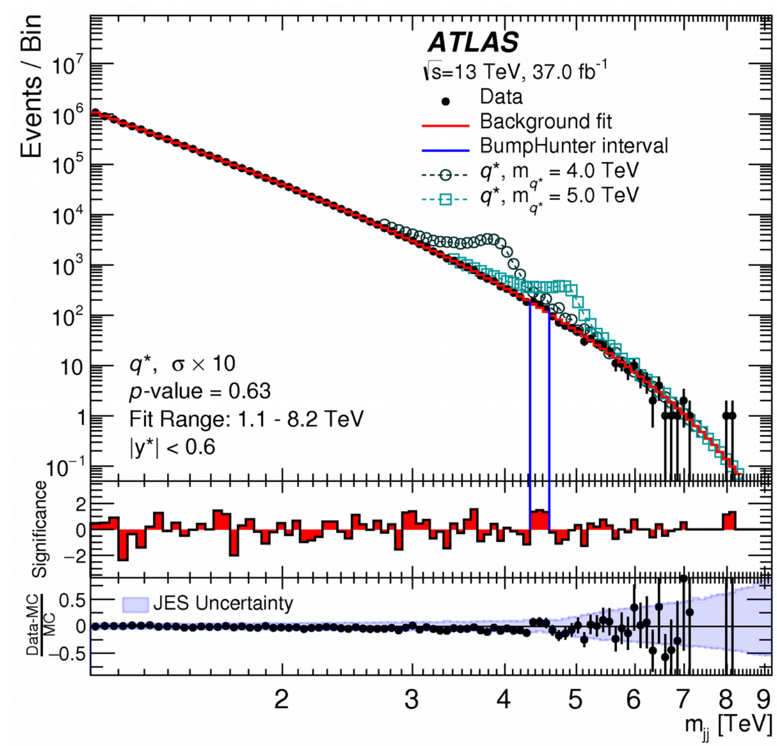
O(100)-dimensional phase space:

– Particle 3-momenta – Vertices
– Particle ID (non-isolated leptons, photons)

An Example Target

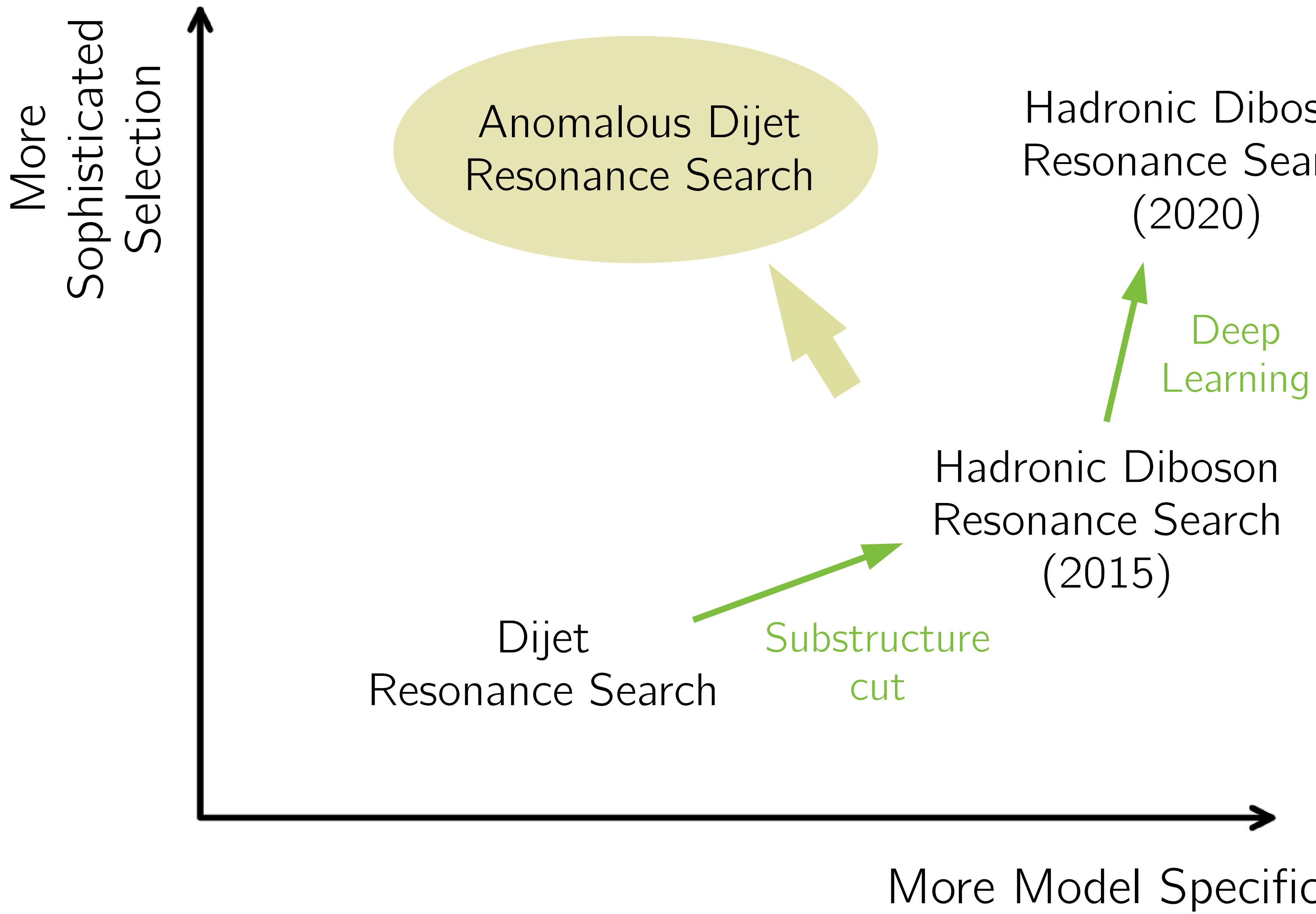


1. It is very simple (bump hunt)
2. It is very complex (jet substructure)
3. It is easily missed



With 150/fb, exclusion on 3 TeV dijet resonance is 5000 events

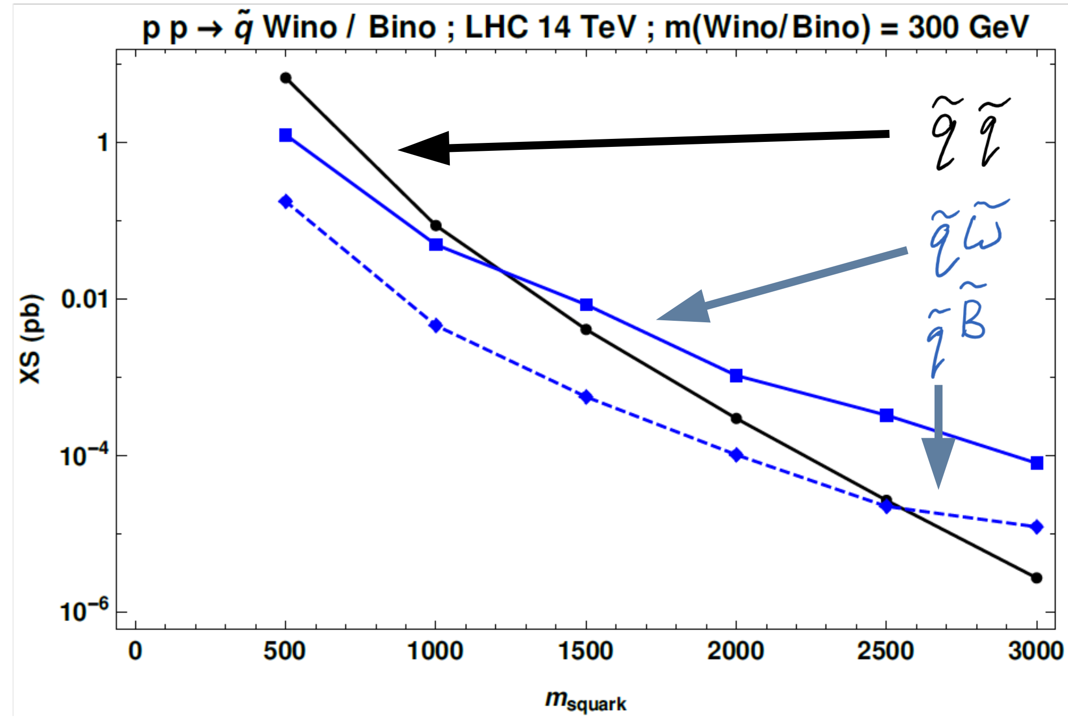
An Example Target

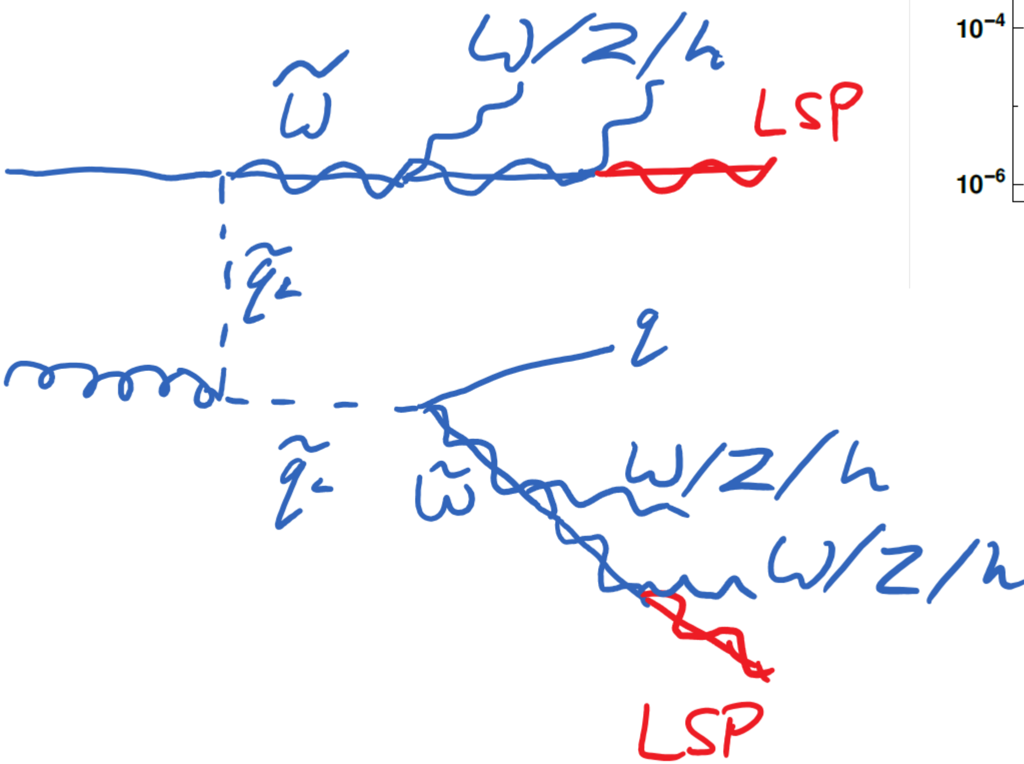


Dark Matter Detour

An Example Target (Dark Matter version)

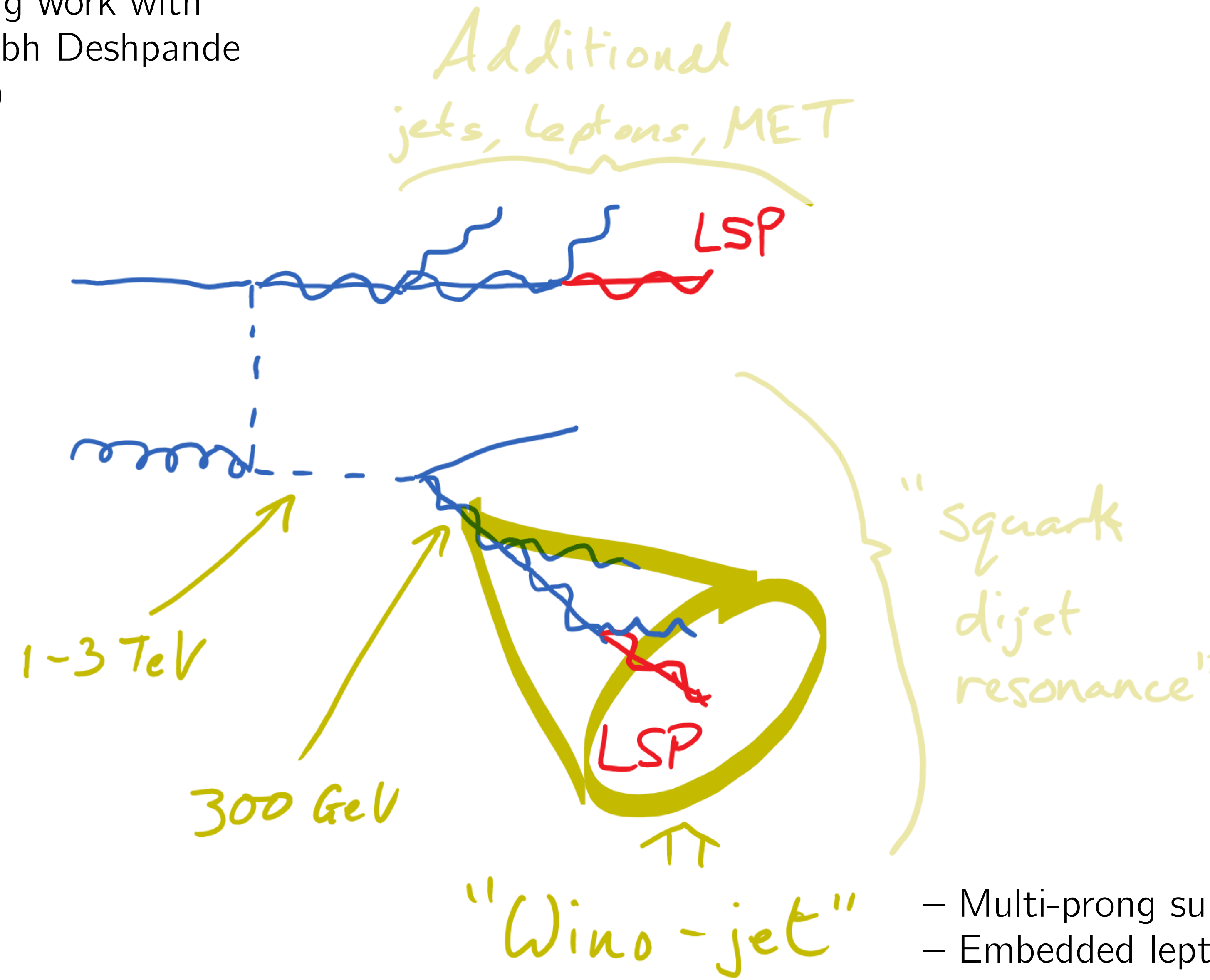
Ongoing work with Kaustubh Deshpande (UMD)





An Example Target (Dark Matter version)

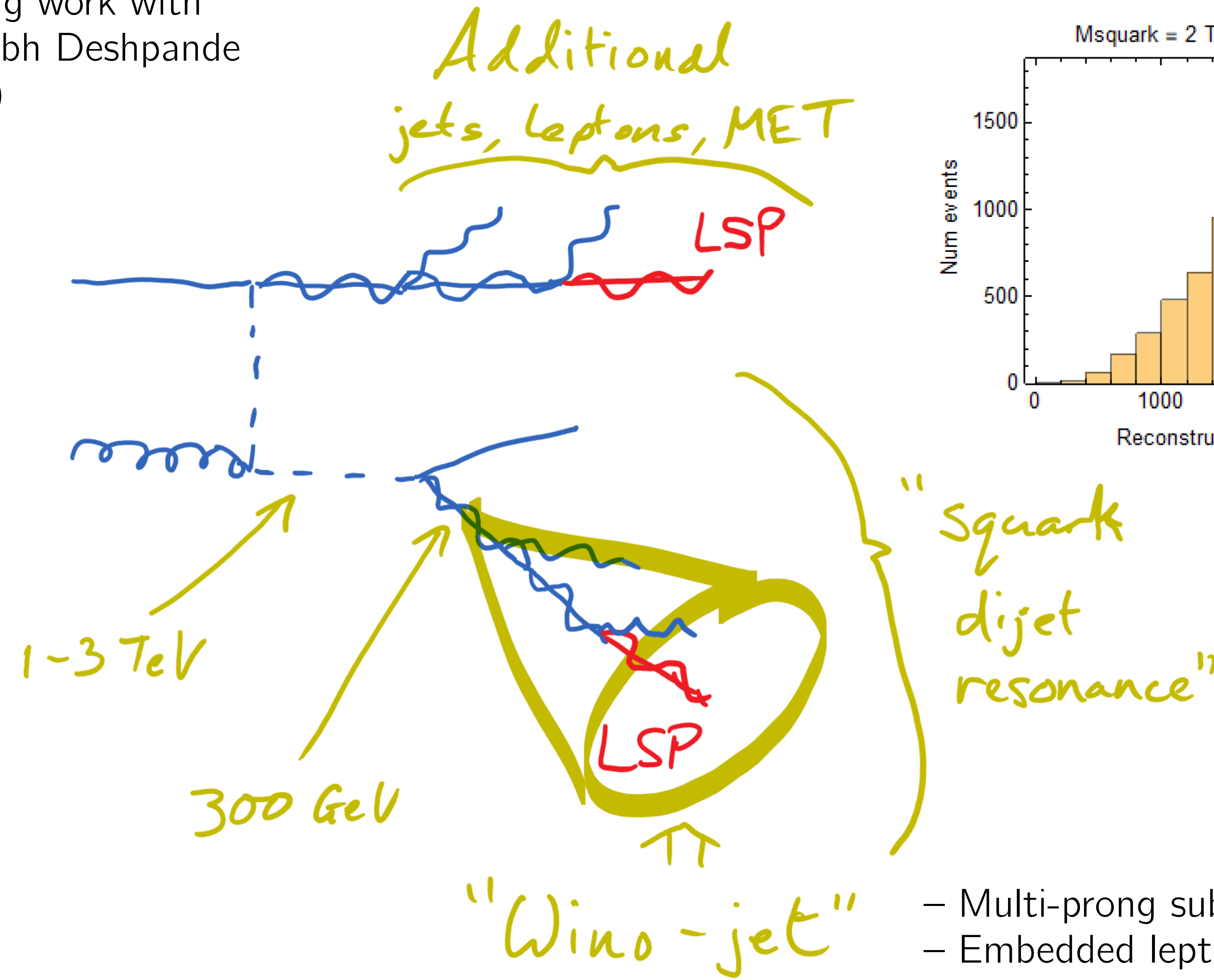
Ongoing work with Kaustubh Deshpande (UMD)

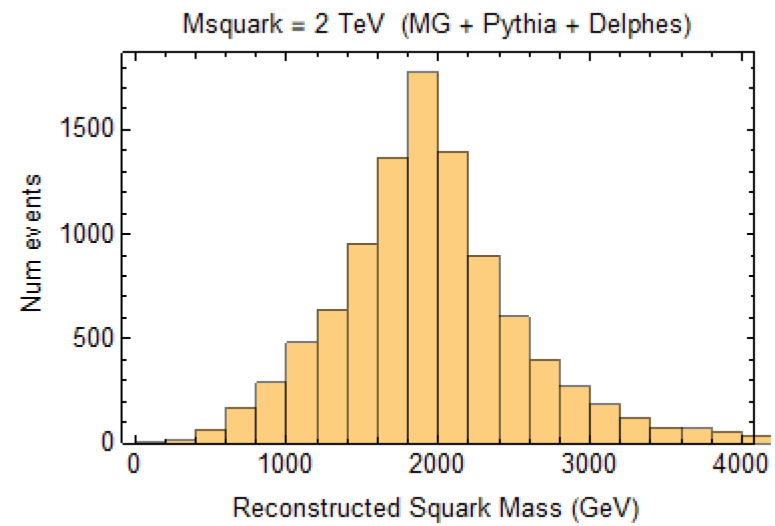


– Multi-prong substructure
– Embedded leptons
– MET collinear with jet

An Example Target (Dark Matter version)

Ongoing work with Kaustubh Deshpande (UMD)





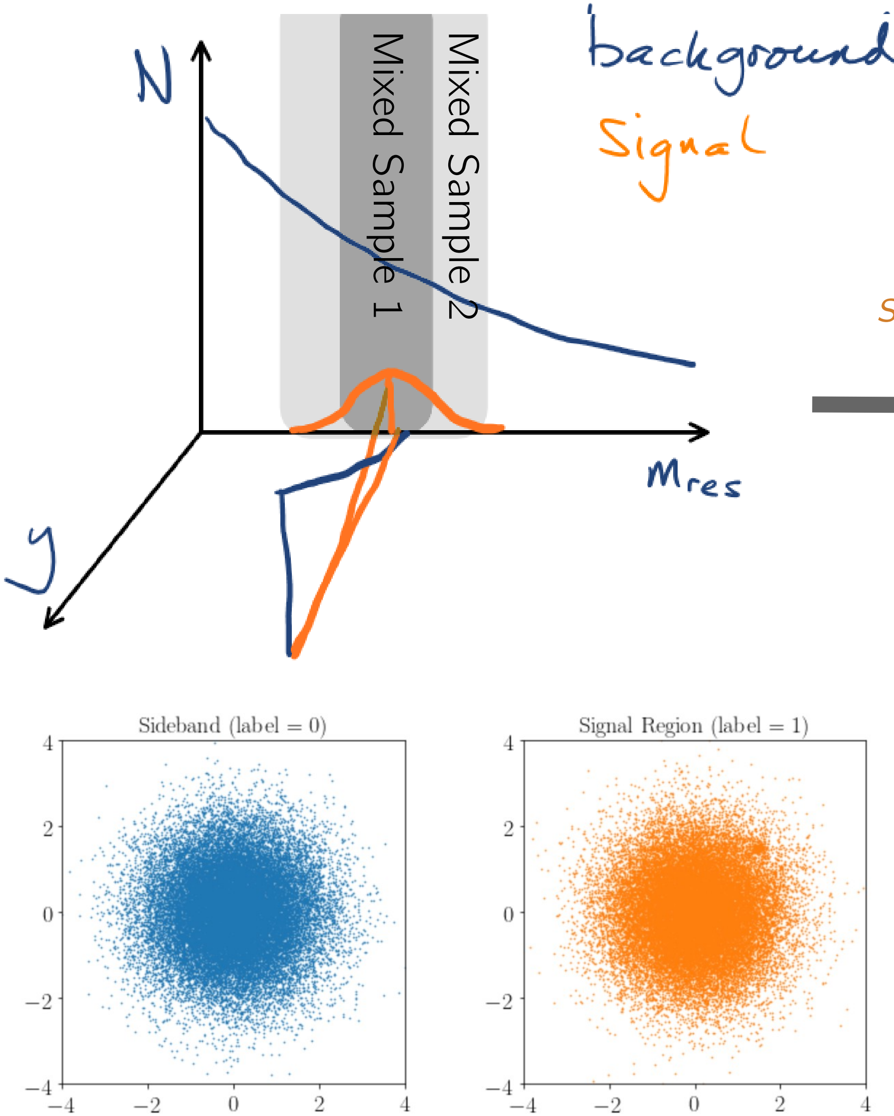
– Multi-prong substructure
– Embedded leptons
– MET collinear with jet

CWoLa Hunting

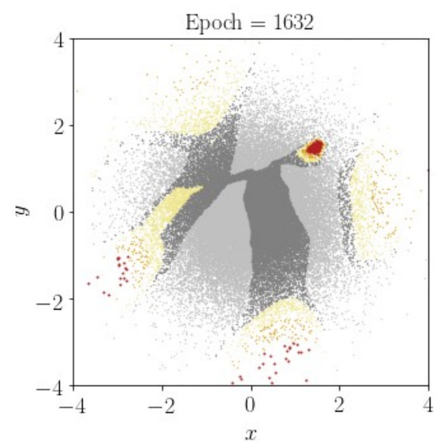
CWoLa Hunting

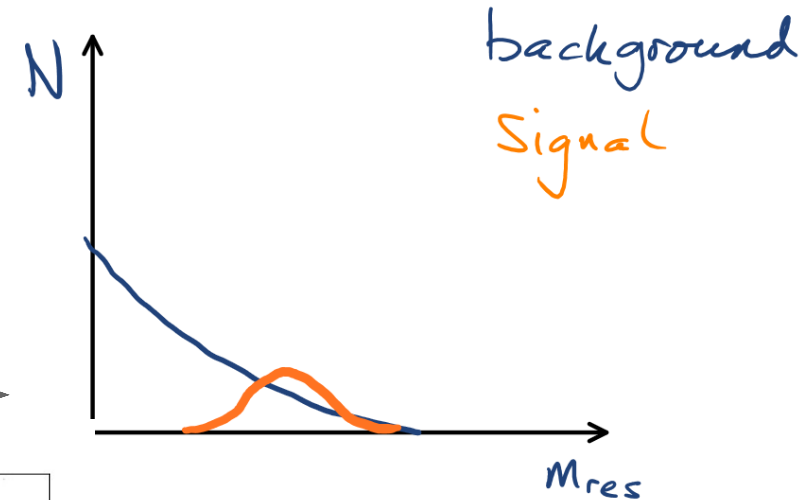
Weak Supervision References:

[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



Selection for *signal-region*-like events

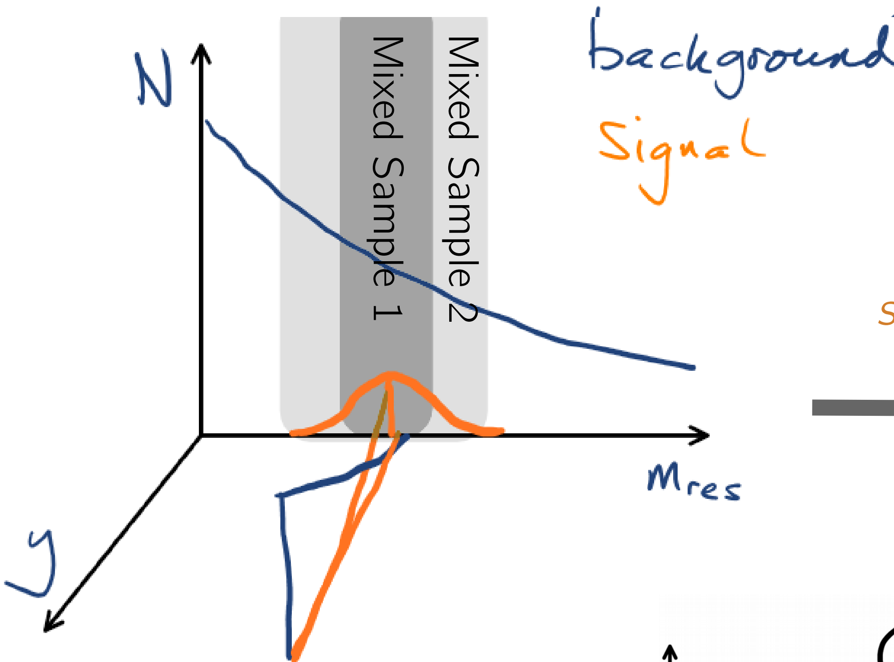




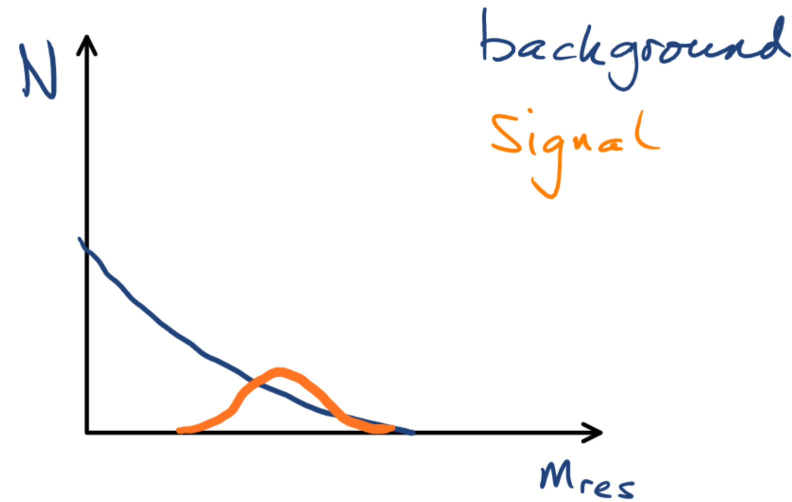
CWoLa Hunting

Weak Supervision References:

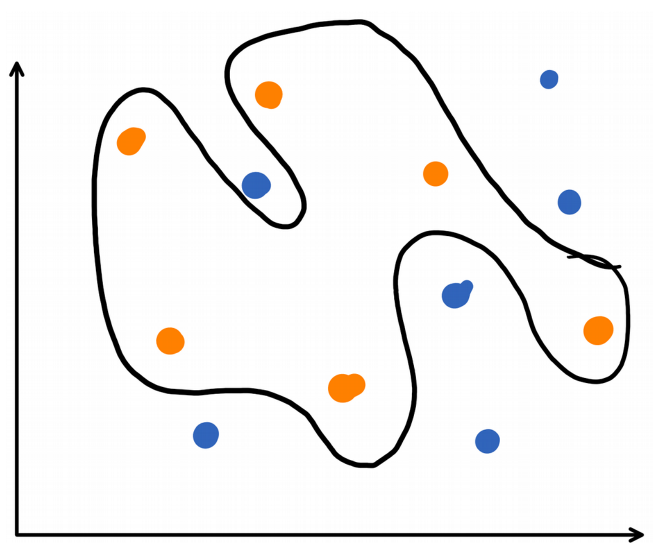
[1708.02949] E. M. Metodiev, B. Nachman, J. Thaler
[1702.00414] L. M. Dery, B. Nachman, F. Rubbo, A Schwartzman
[1801.10158] P. T. Komiske, E. M. Metodiev, B. Nachman, M. D. Schwartz
[1706.09451] T. Cohen, M. Freytsis, B. Ostdiek



Selection for *signal-region*-like events

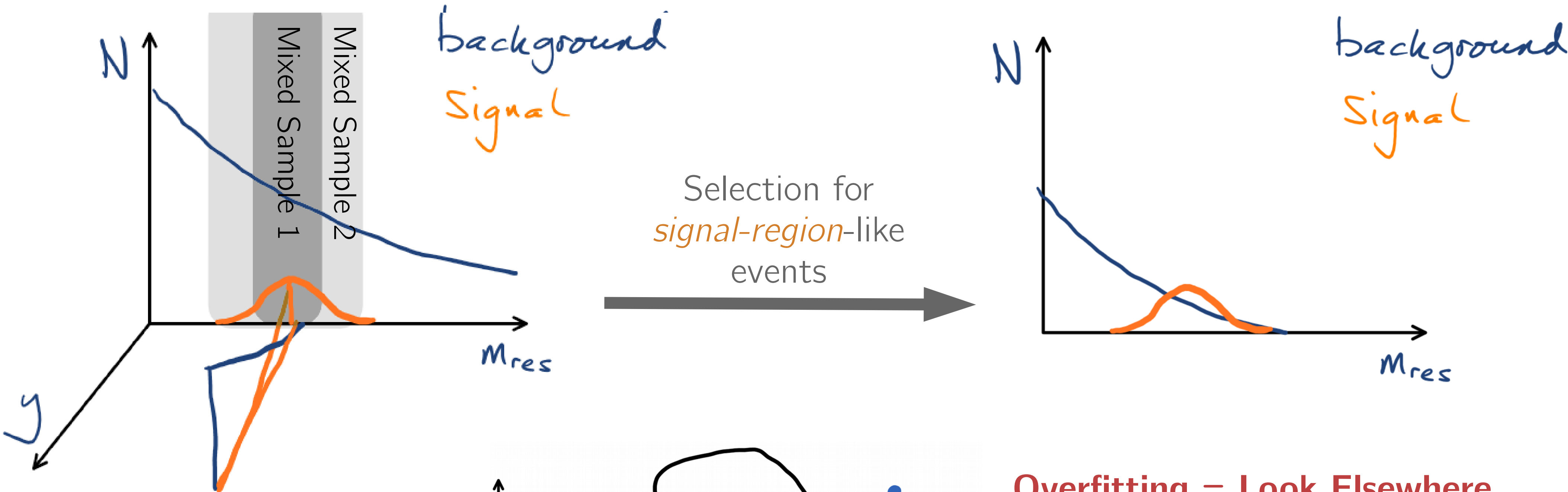


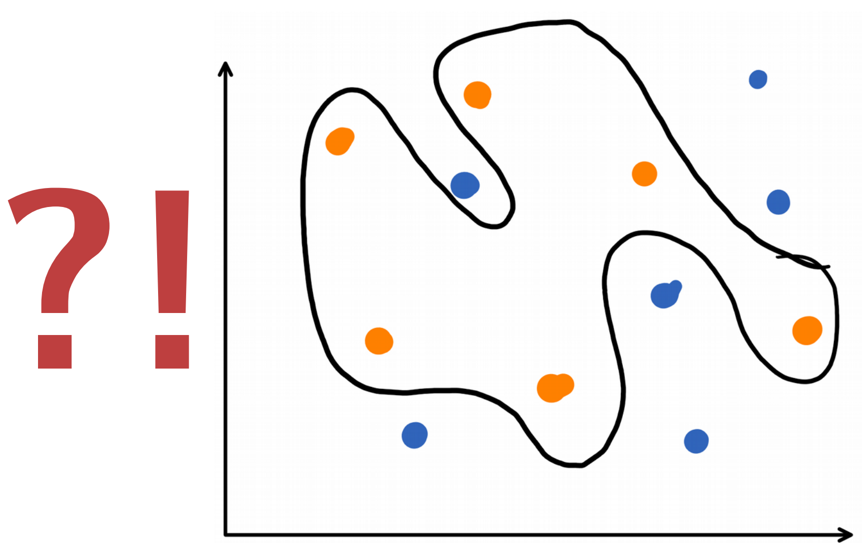
?!



CWoLa Hunting

Weak Supervision References:
[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



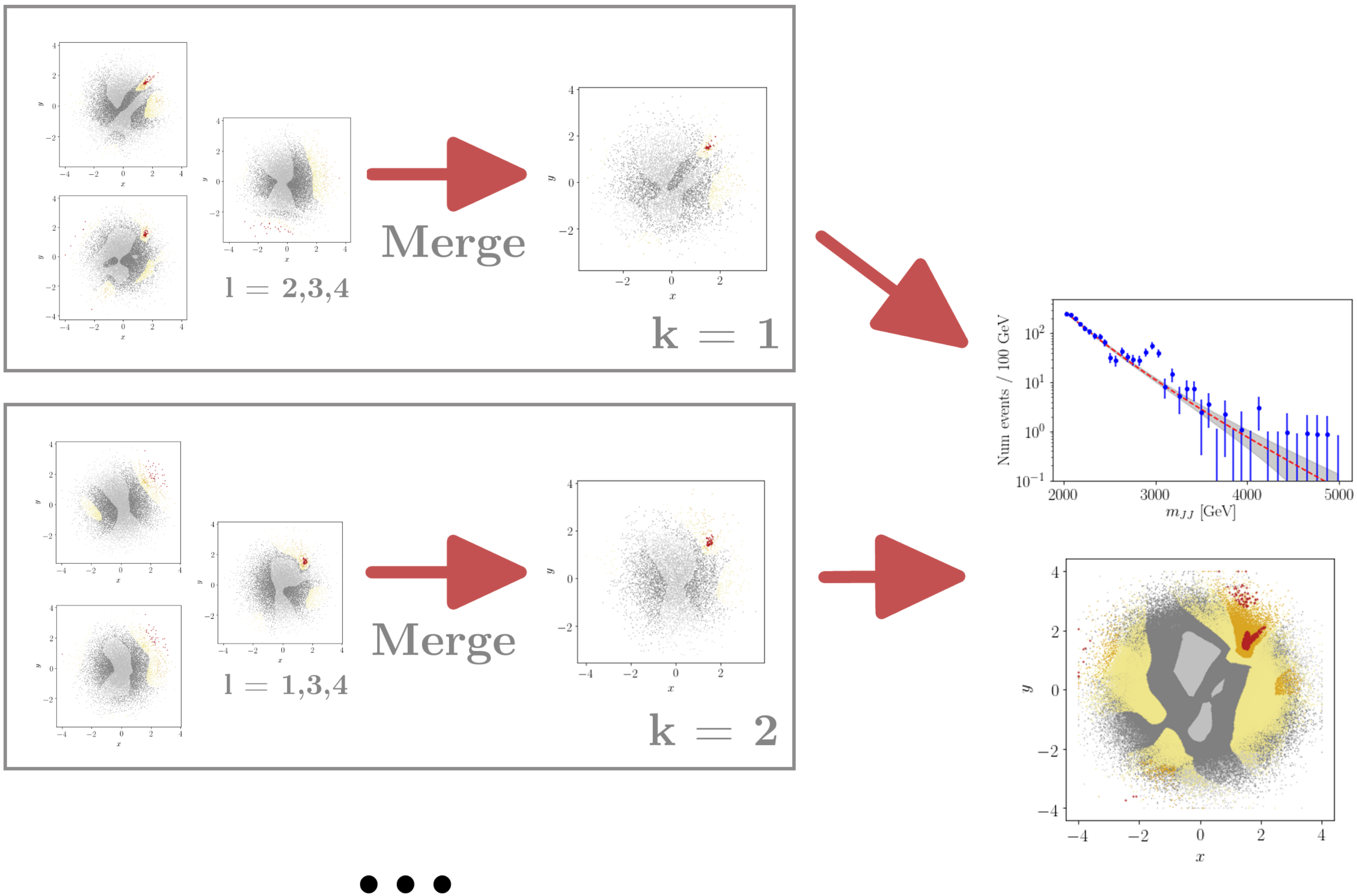


Overfitting = Look Elsewhere Effect

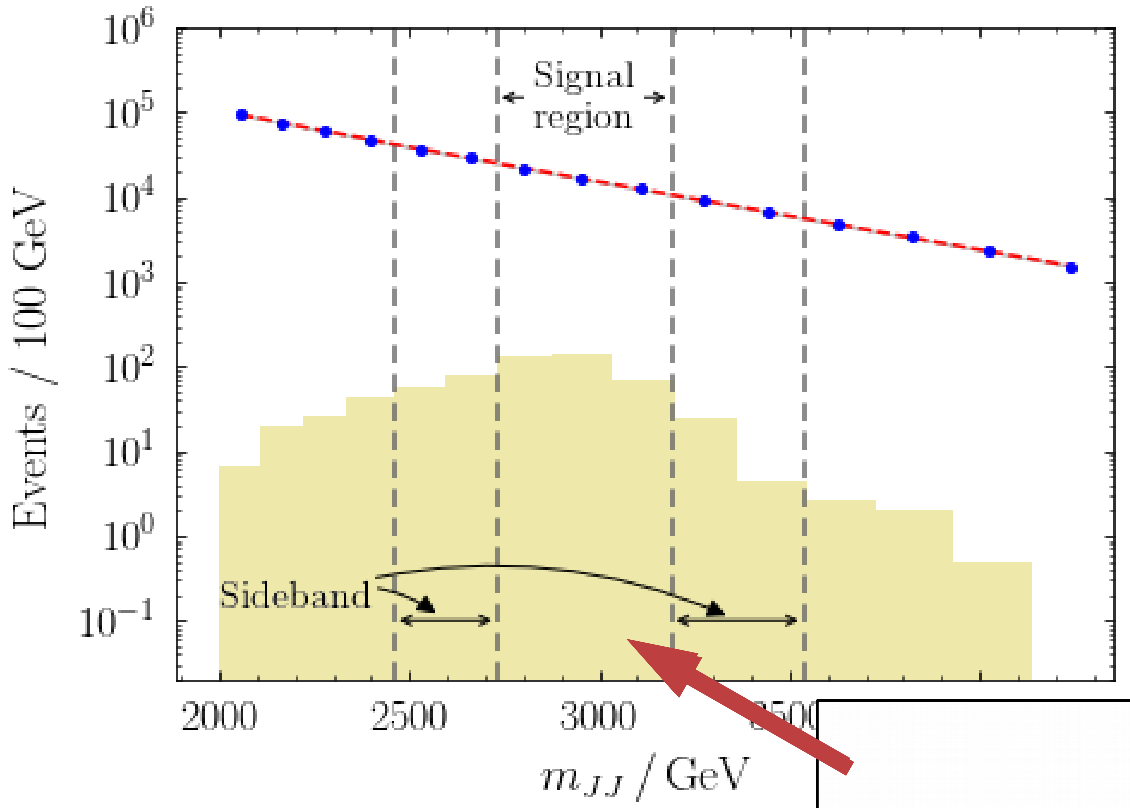
Overfitting leads to bumps even in absence of signal

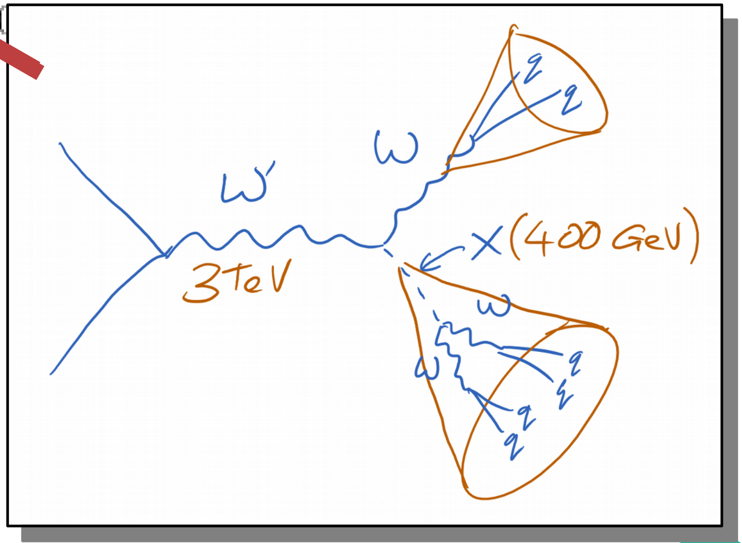
Solution: Train and test on statistically independent samples.

Cross Validation



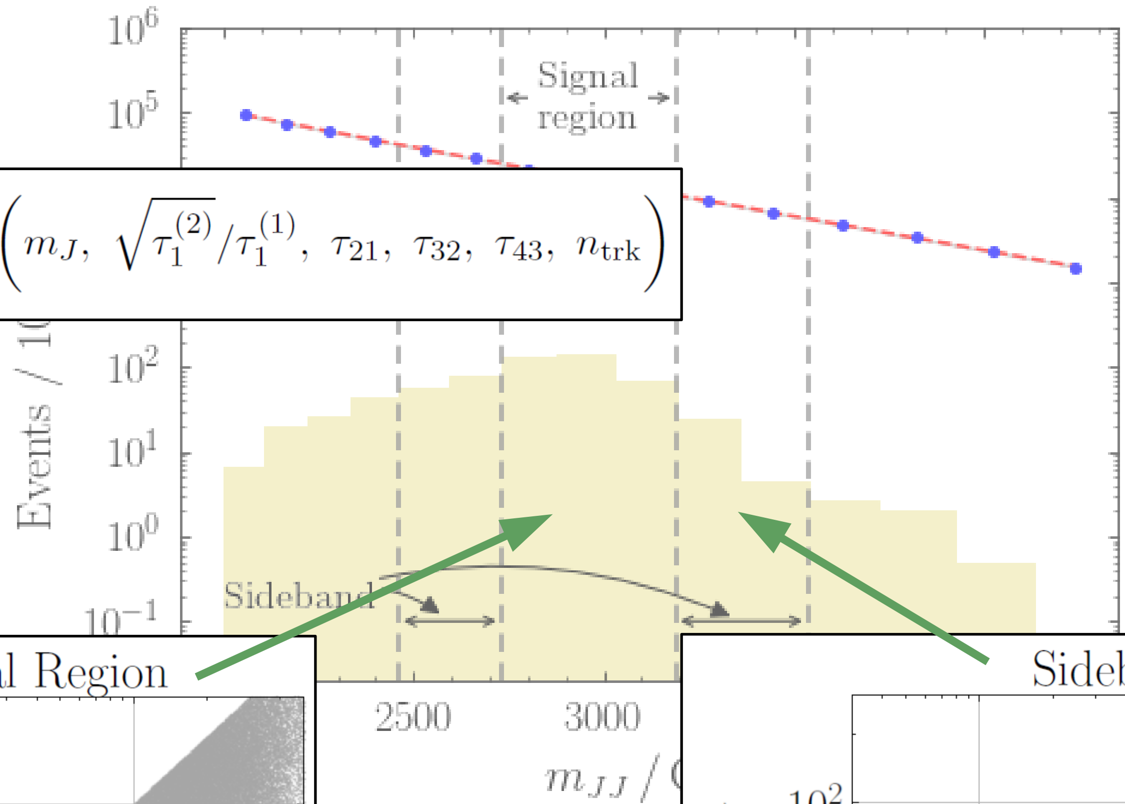
Case Study



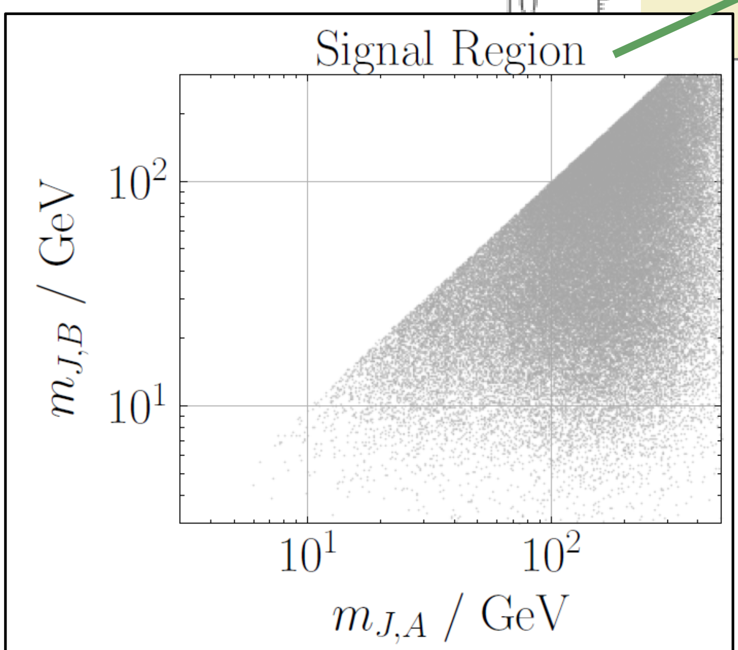


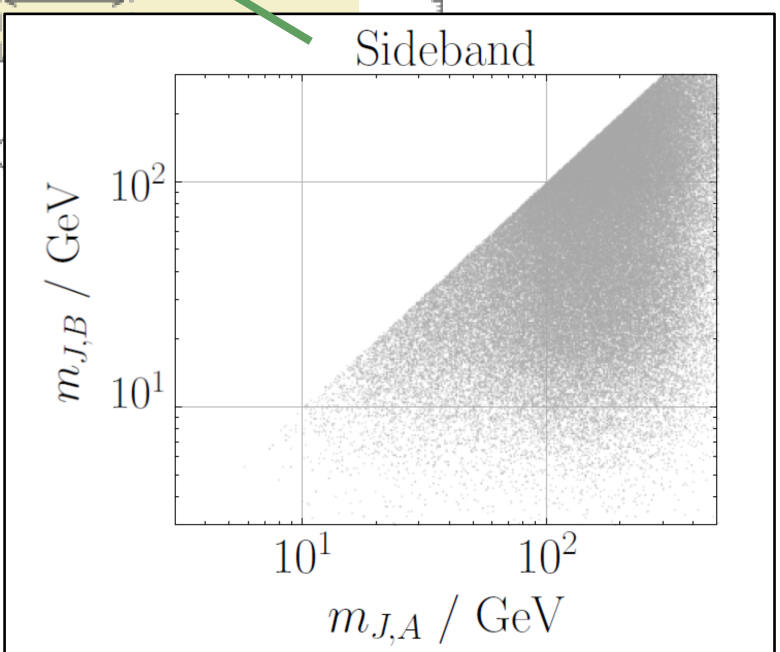
In signal region:
S = 522,
S/B = 0.64%

Case Study



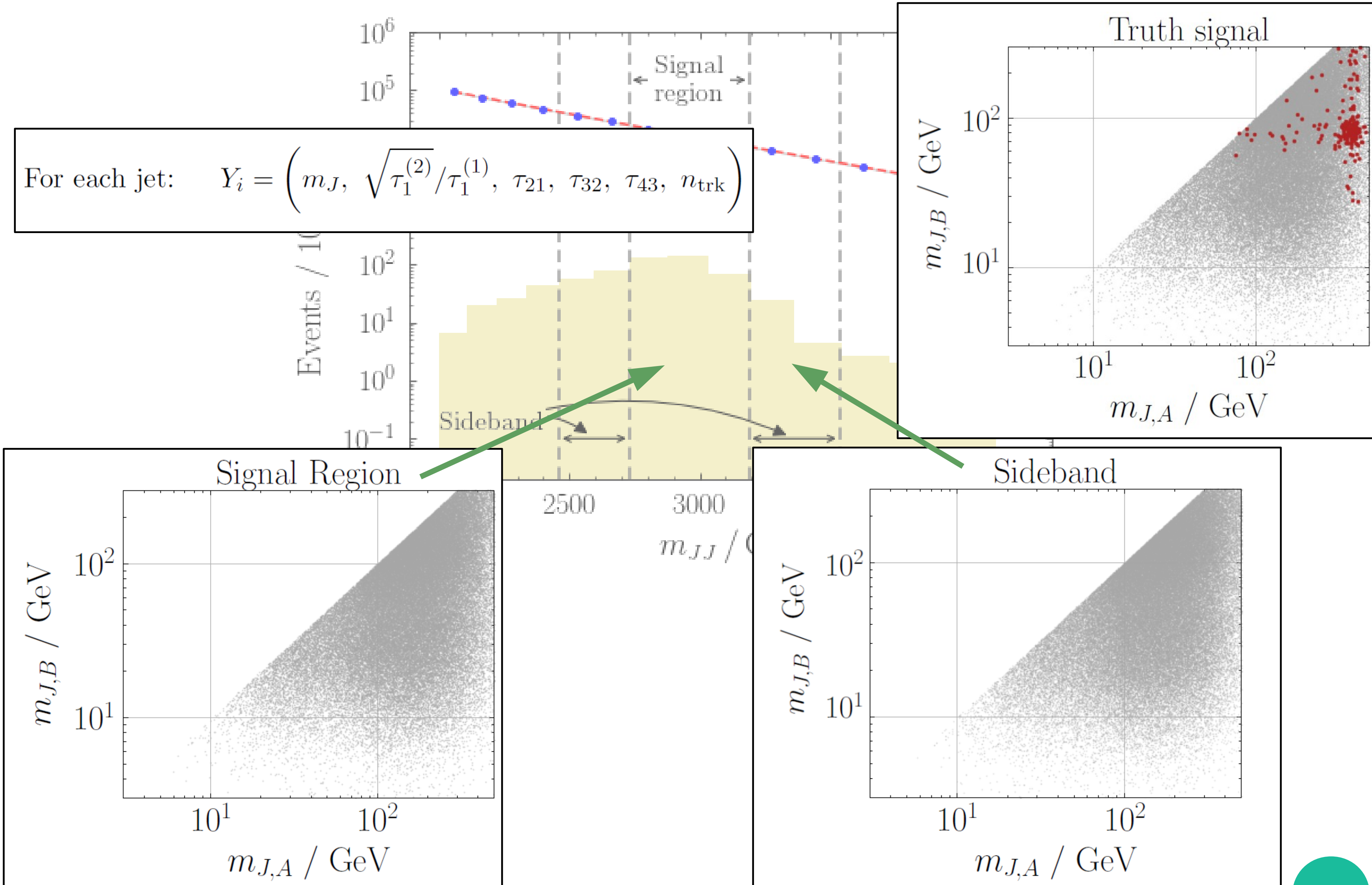
For each jet: $Y_i = \left(m_J,\ \sqrt{\tau_1^{(2)}/\tau_1^{(1)}},\ \tau_{21},\ \tau_{32},\ \tau_{43},\ n_{\mathrm{trk}} \right)$



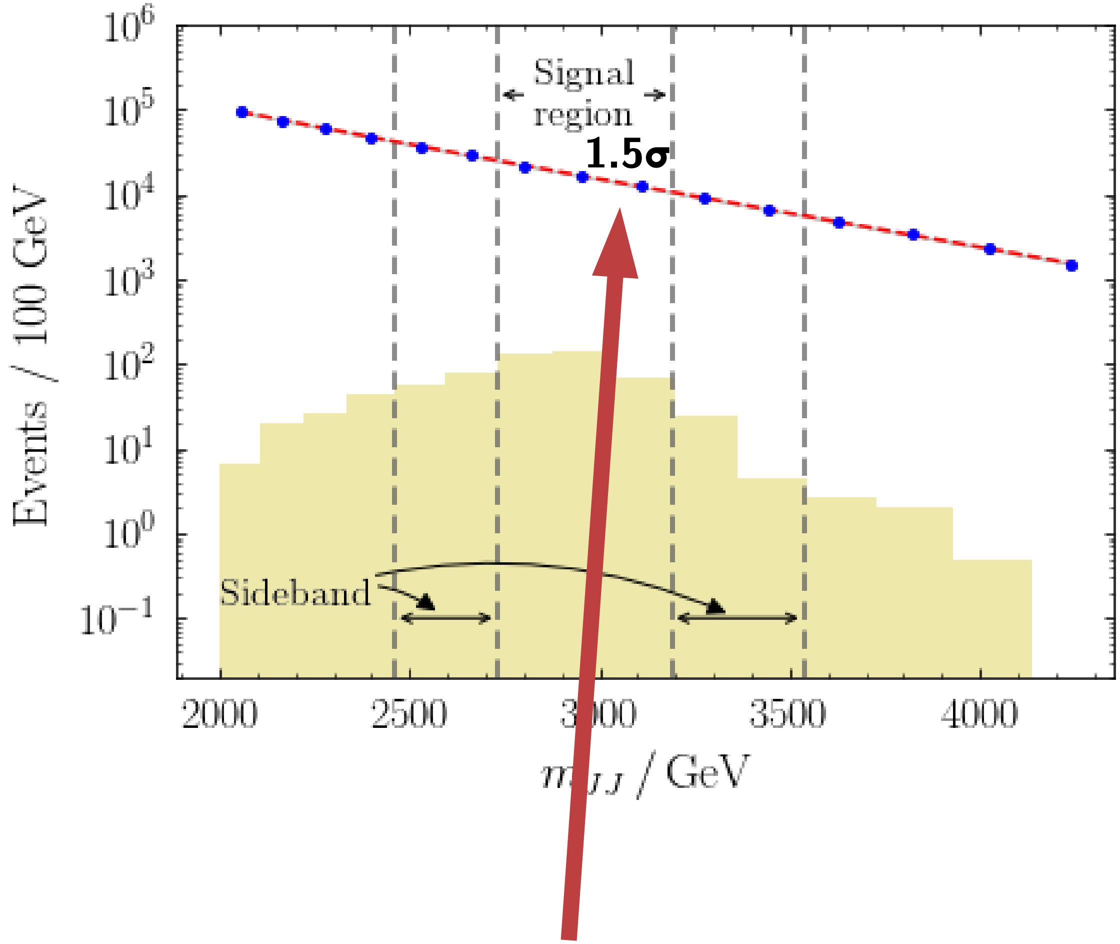


Case Study

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

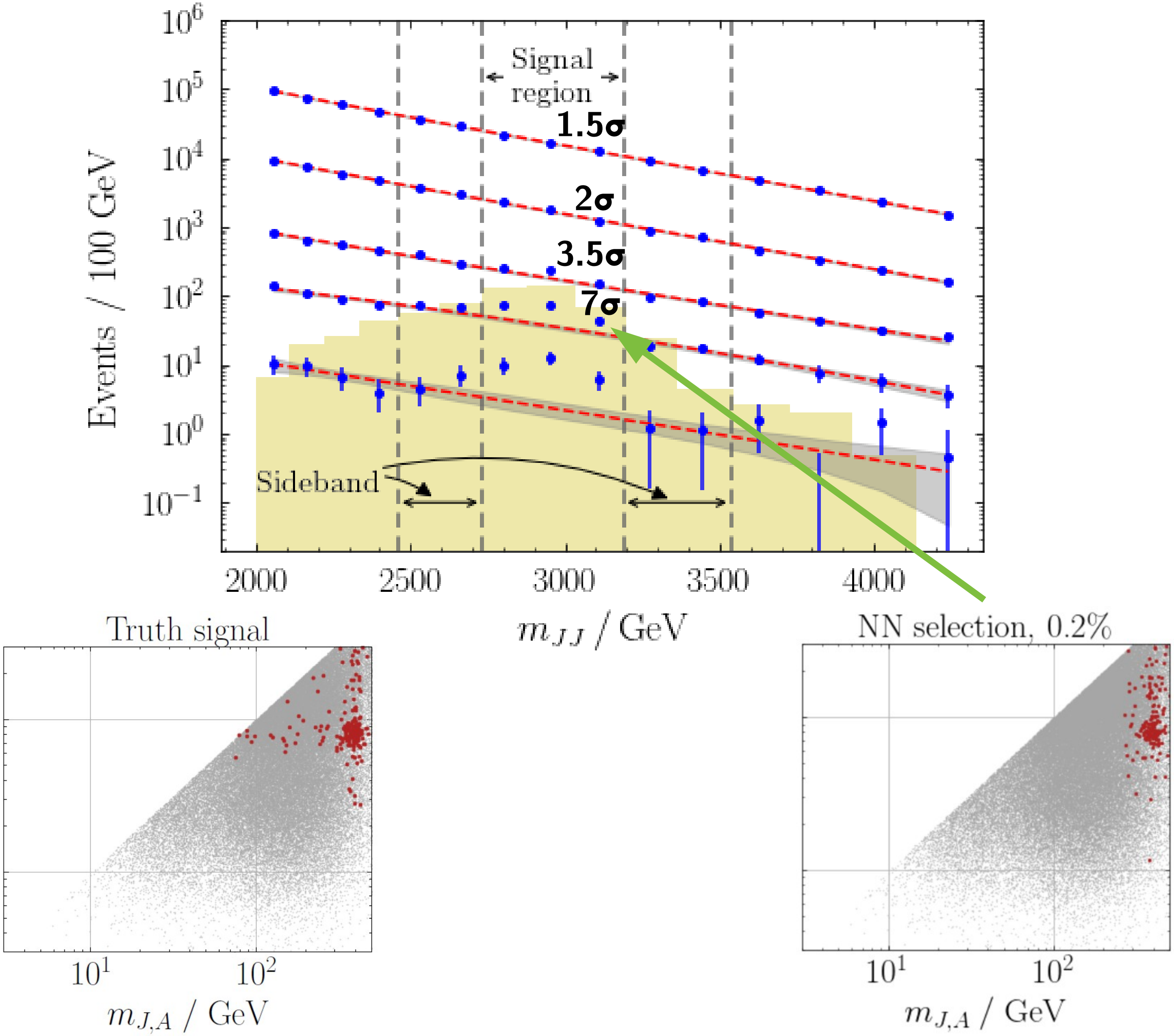


Case Study



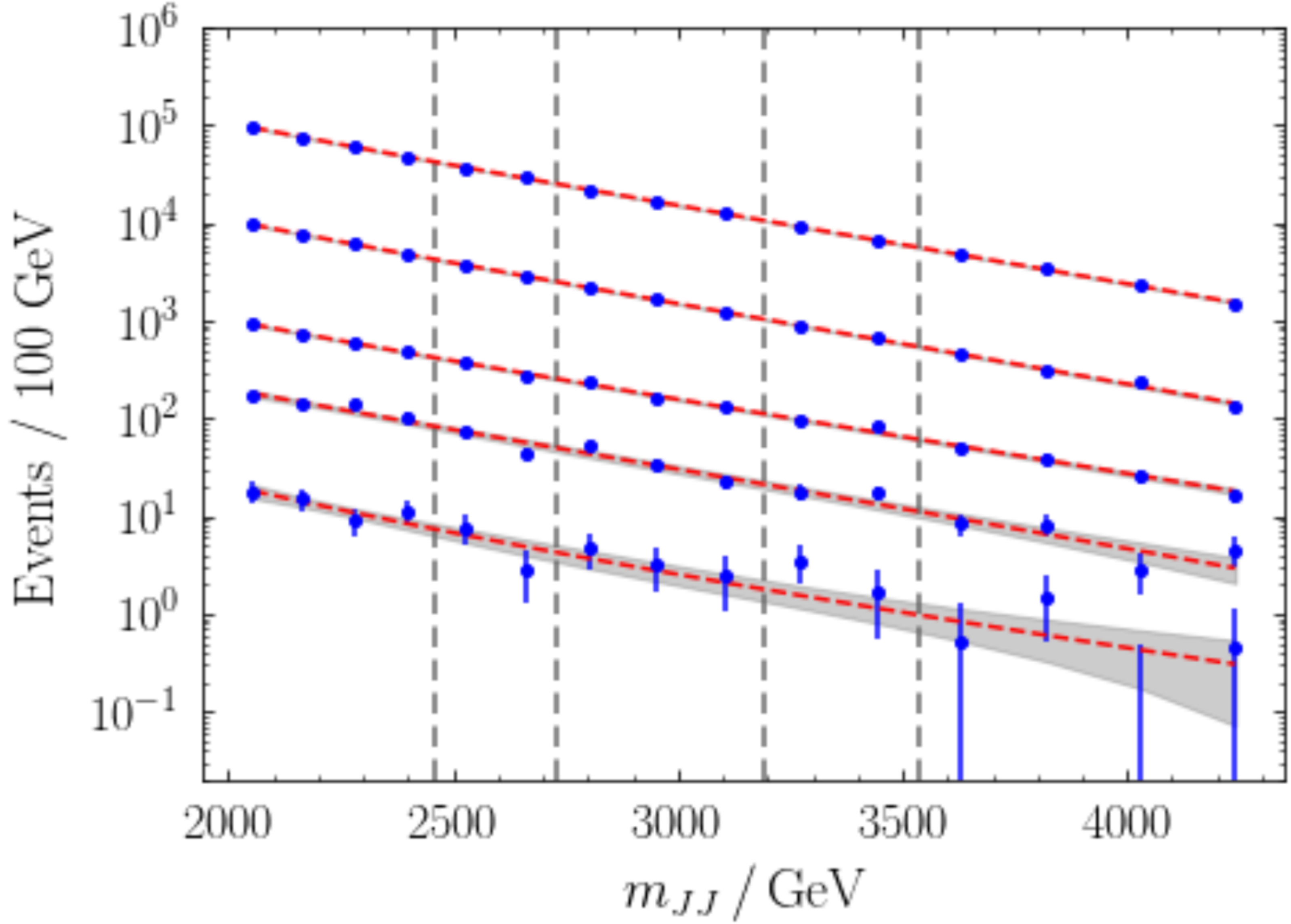
$$\frac{dN}{dm_{JJ}} = p_0 \frac{(1 - m_{JJ}/\sqrt{s})^{p_1}}{(m_{JJ}/\sqrt{s})^{p_2}}$$

Case Study

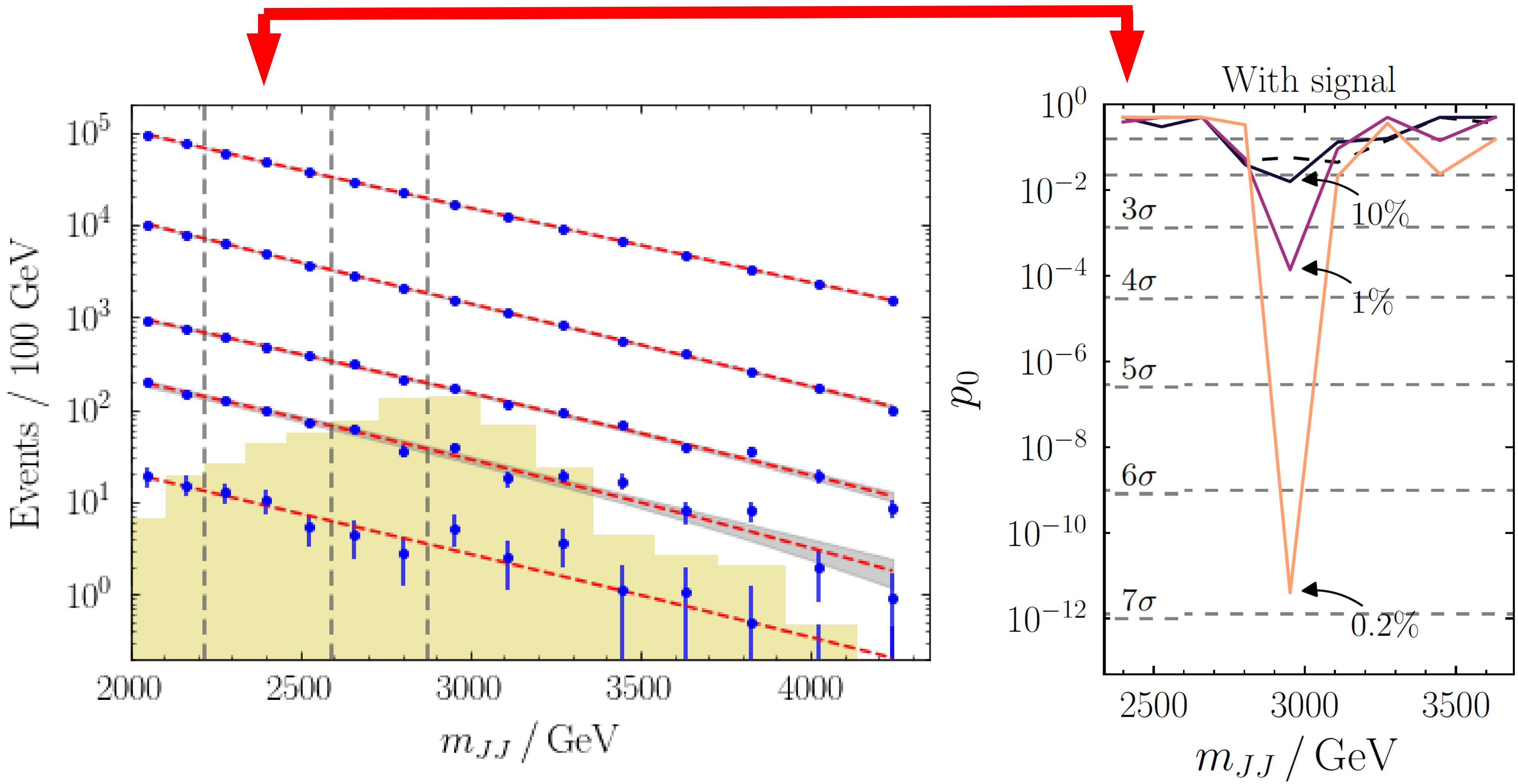


Signal region
Sideband
1.5σ
2σ
3.5σ
7σ
Events / 100 GeV
m_{JJ} / GeV
Truth signal
NN selection, 0.2%
$m_{J,A}$ / GeV
$m_{J,A}$ / GeV

No Signal → No Bump!

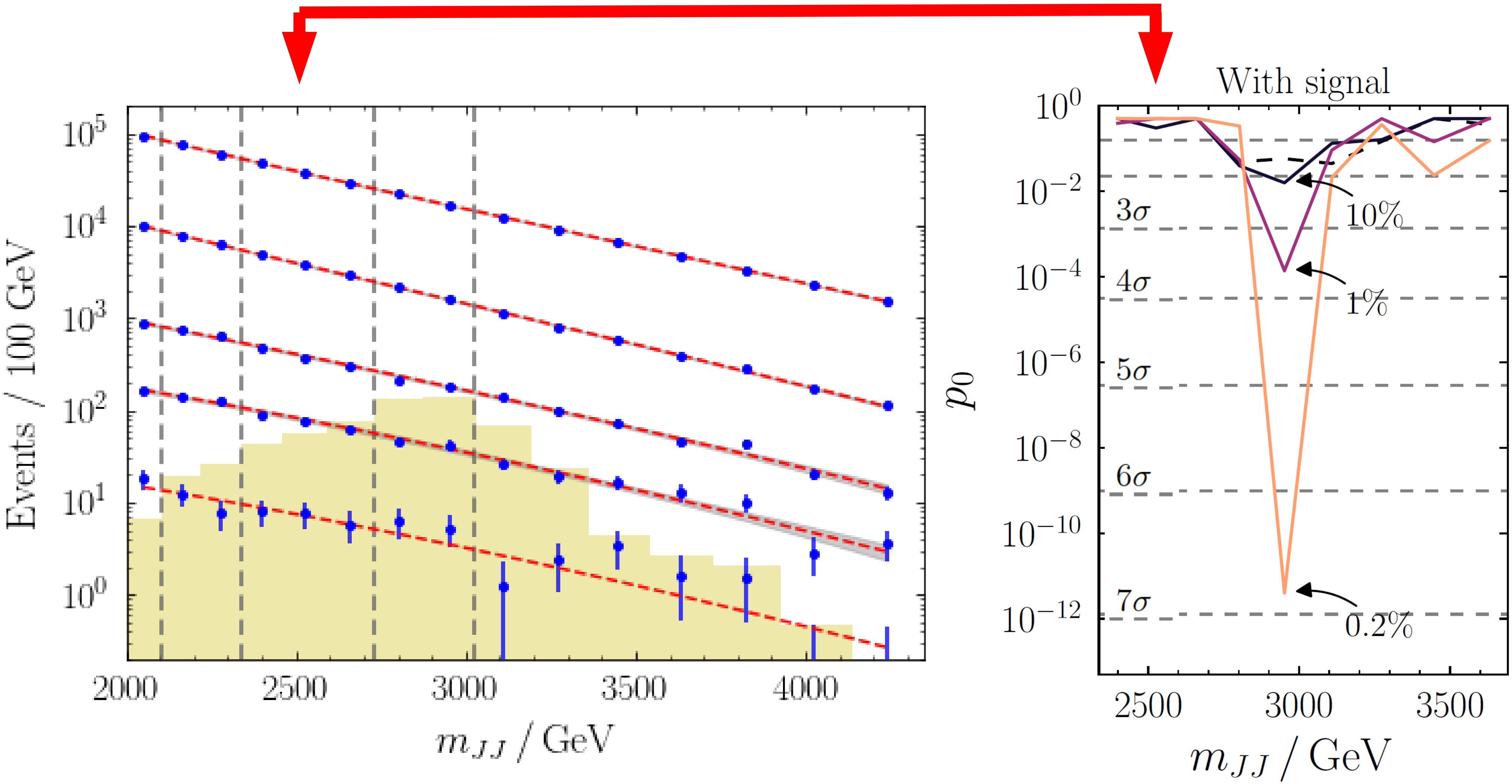


Mass Scan



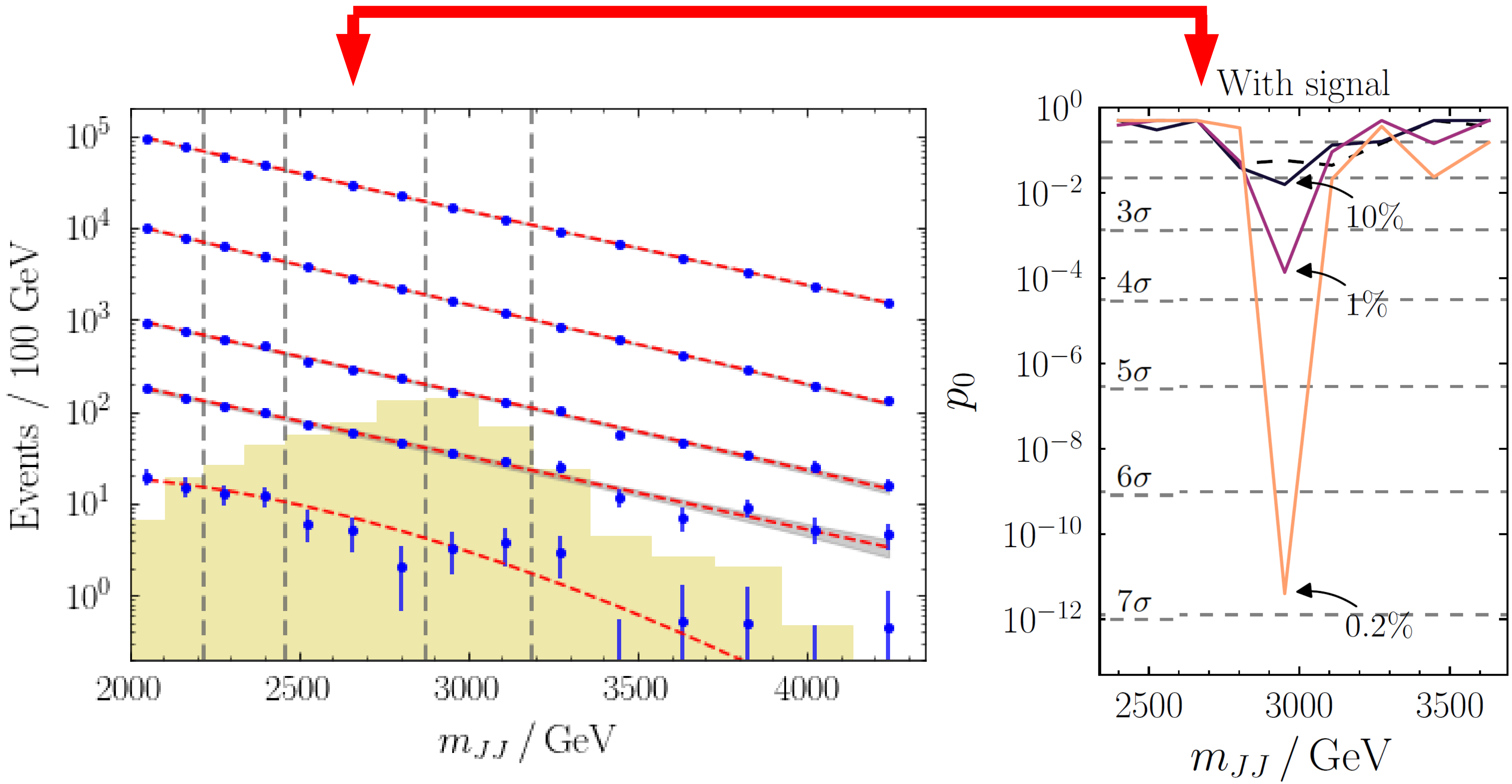
With signal
Events / 100 GeV
m_{JJ} / GeV
p_0
10^5
10^4
10^3
10^2
10^1
10^0
2000
2500
3000
3500
4000
10^0
10^{-2}
10^{-4}
10^{-6}
10^{-8}
10^{-10}
10^{-12}
3σ
4σ
5σ
6σ
7σ
10%
1%
0.2%

Mass Scan



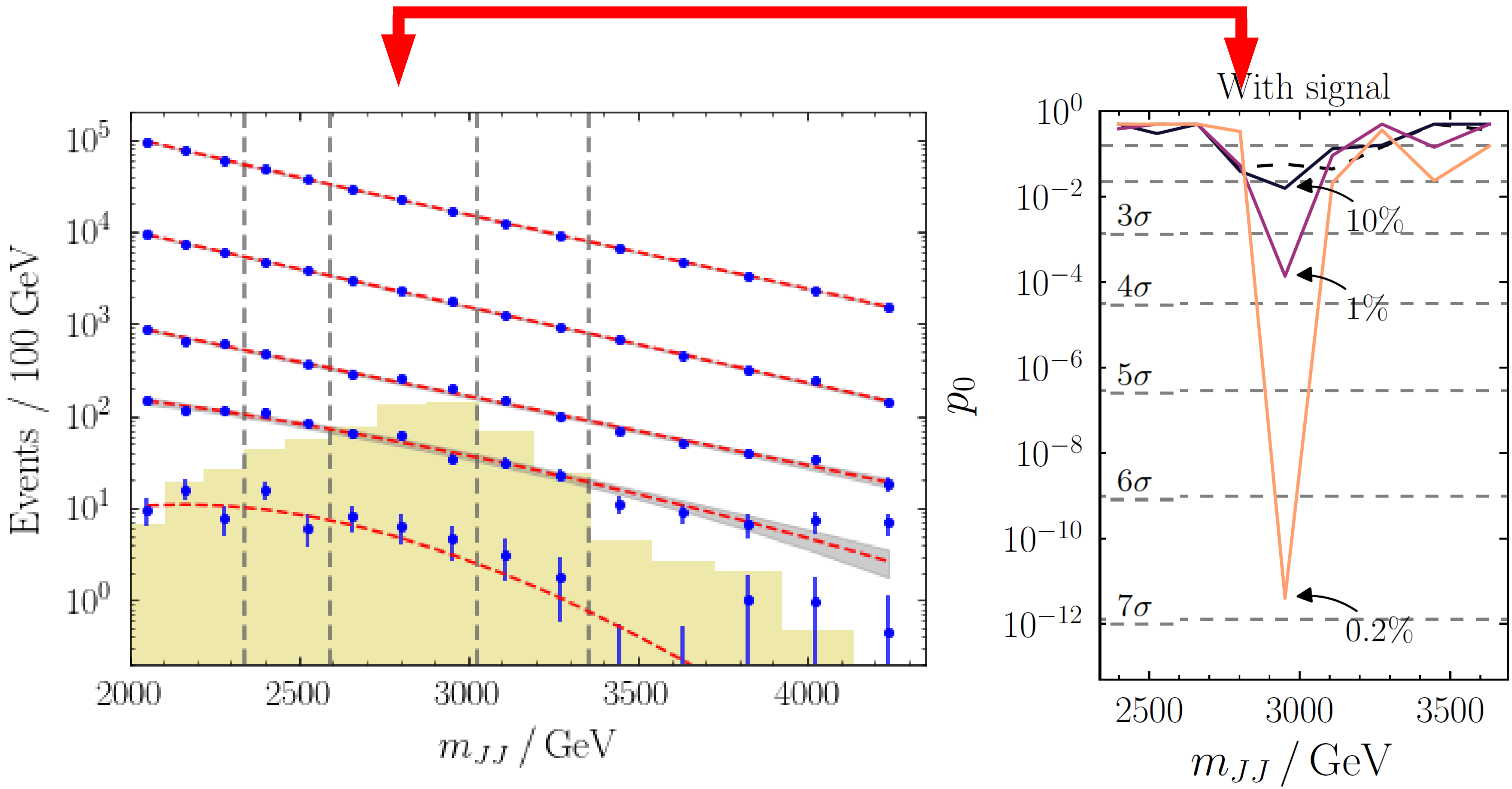
With signal
10^0
10^{-2}
10^{-4}
10^{-6}
10^{-8}
10^{-10}
10^{-12}
3σ
4σ
5σ
6σ
7σ
10%
1%
0.2%
p_0
Events / 100 GeV
10^5
10^4
10^3
10^2
10^1
10^0
2000
2500
3000
3500
4000
m_{JJ} / GeV

Mass Scan



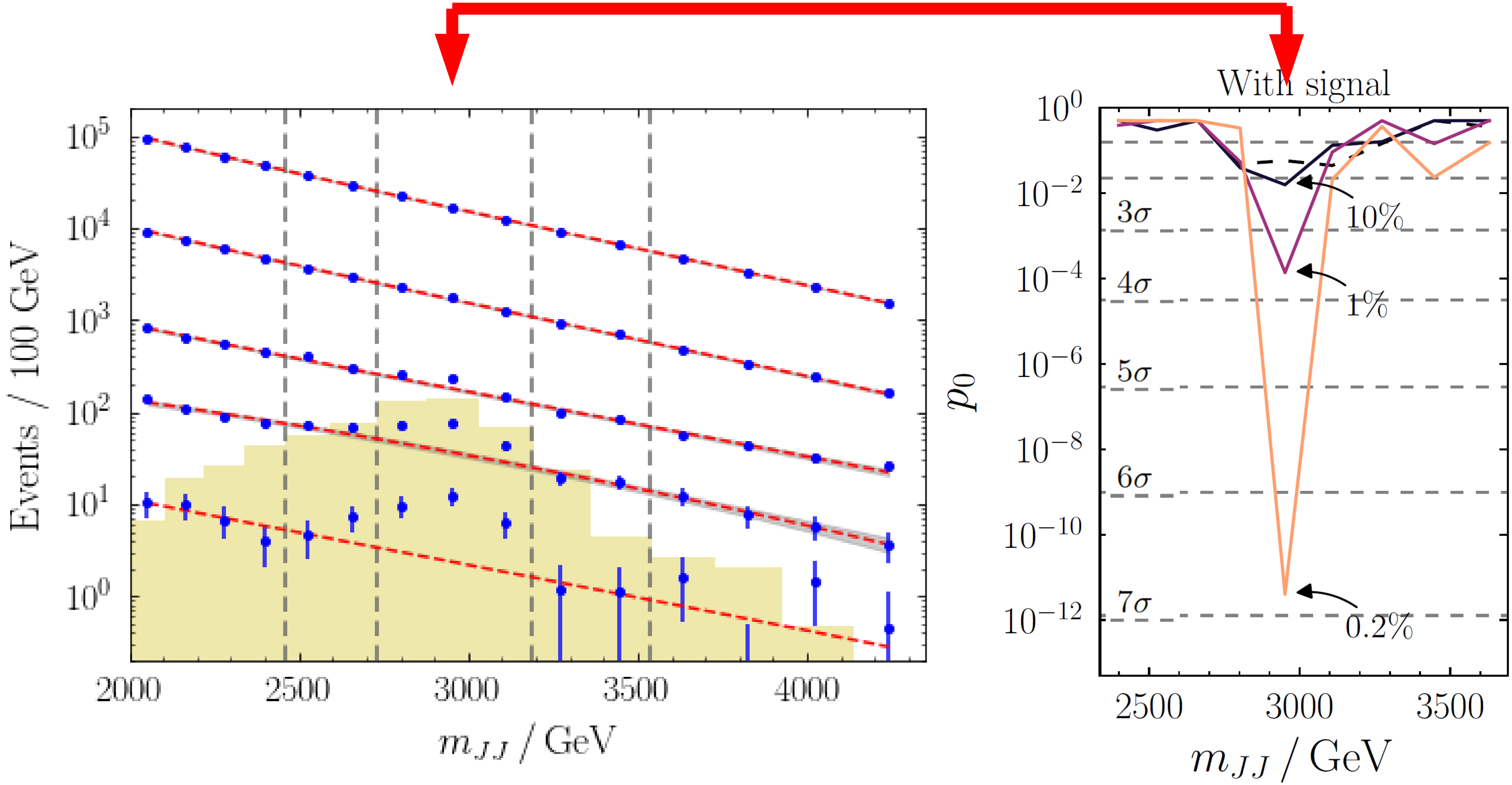
Events / 100 GeV
10^5
10^4
10^3
10^2
10^1
10^0
2000
2500
3000
3500
4000
m_{JJ} / GeV
With signal
p_0
10^0
10^{-2}
10^{-4}
10^{-6}
10^{-8}
10^{-10}
10^{-12}
3σ
4σ
5σ
6σ
7σ
10%
1%
0.2%
2500
3000
3500
m_{JJ} / GeV

Mass Scan



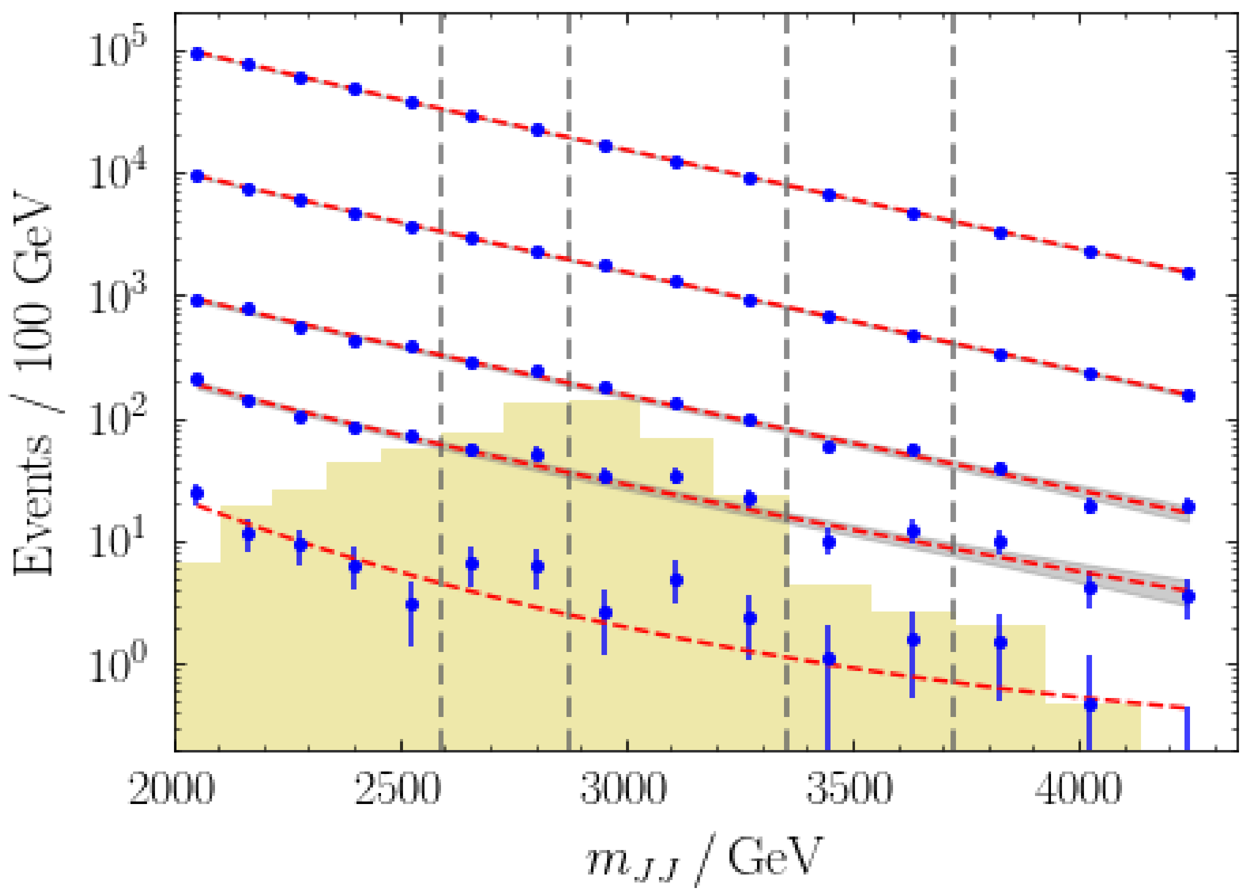
Events / 100 GeV
m_{JJ} / GeV
With signal
p_0
3σ
4σ
5σ
6σ
7σ
10%
1%
0.2%

Mass Scan

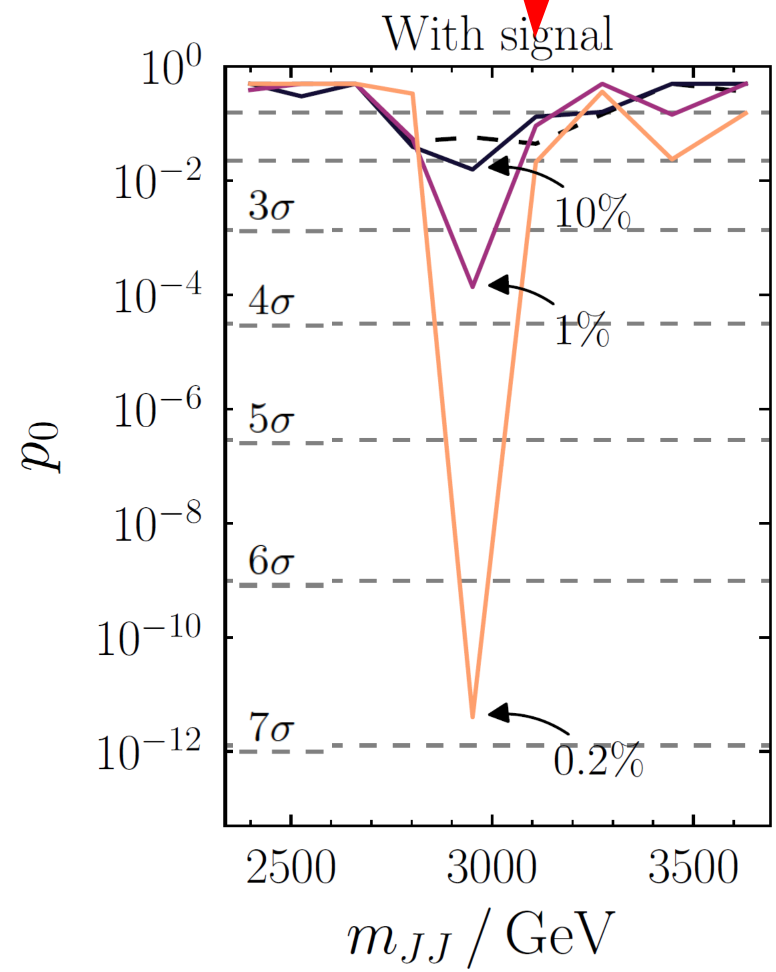


With signal
Events / 100 GeV
m_{JJ} / GeV
p_0
3σ
4σ
5σ
6σ
7σ
10%
1%
0.2%

Mass Scan

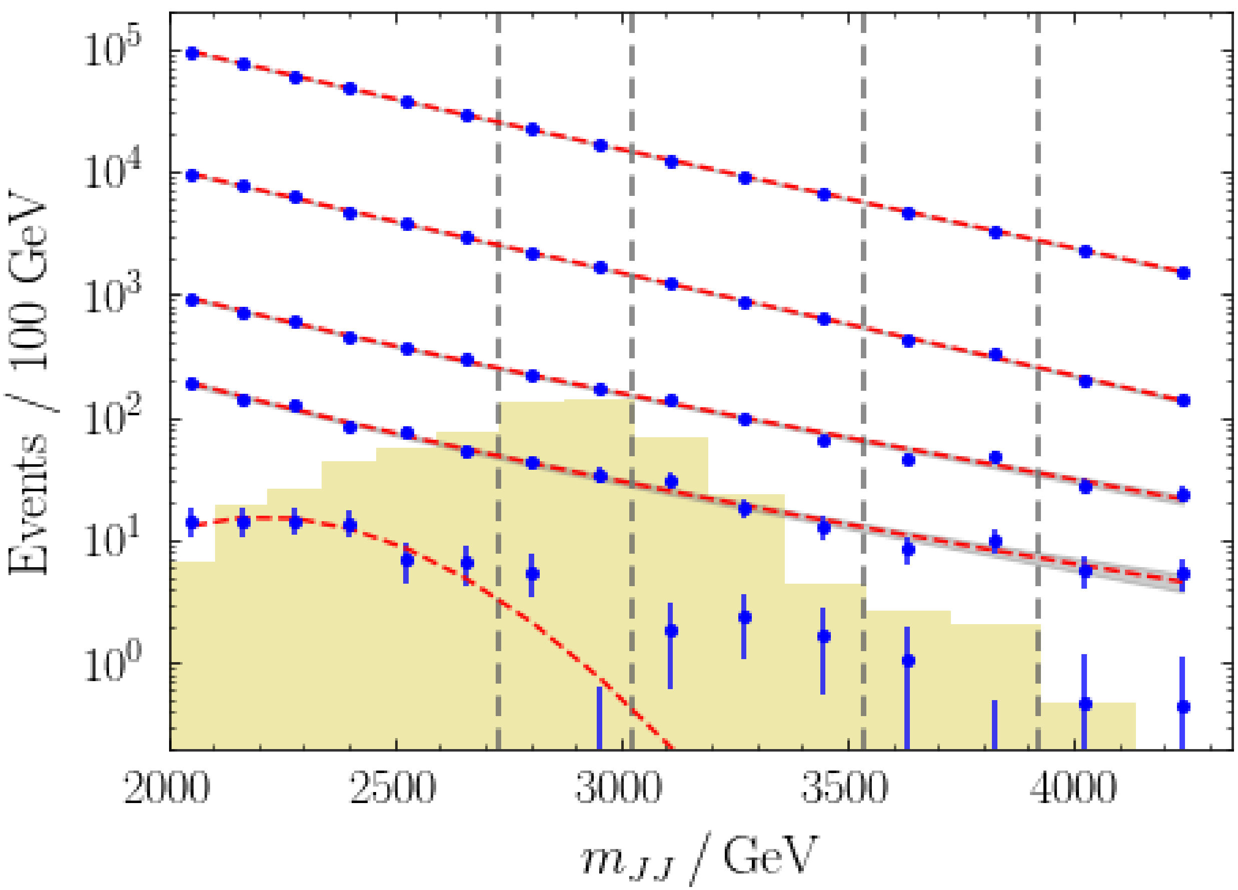


Events / 100 GeV
10^5
10^4
10^3
10^2
10^1
10^0
2000
2500
3000
3500
4000
m_{JJ} / GeV

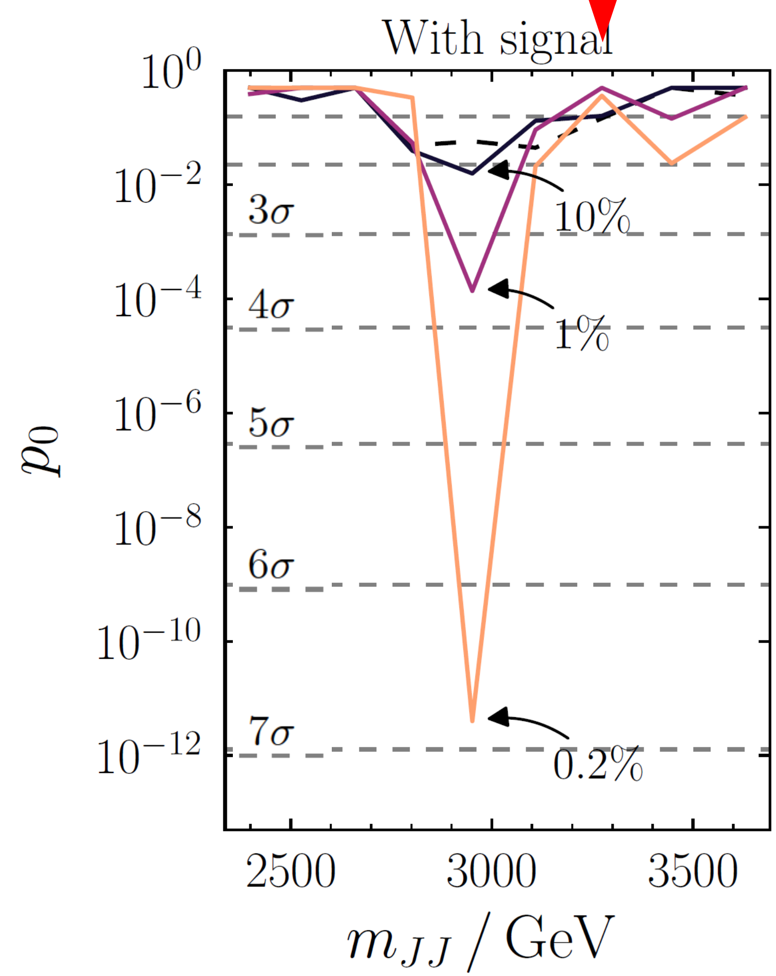


With signal
p_0
10^0
10^{-2}
10^{-4}
10^{-6}
10^{-8}
10^{-10}
10^{-12}
3σ
4σ
5σ
6σ
7σ
10%
1%
0.2%
2500
3000
3500
m_{JJ} / GeV

Mass Scan

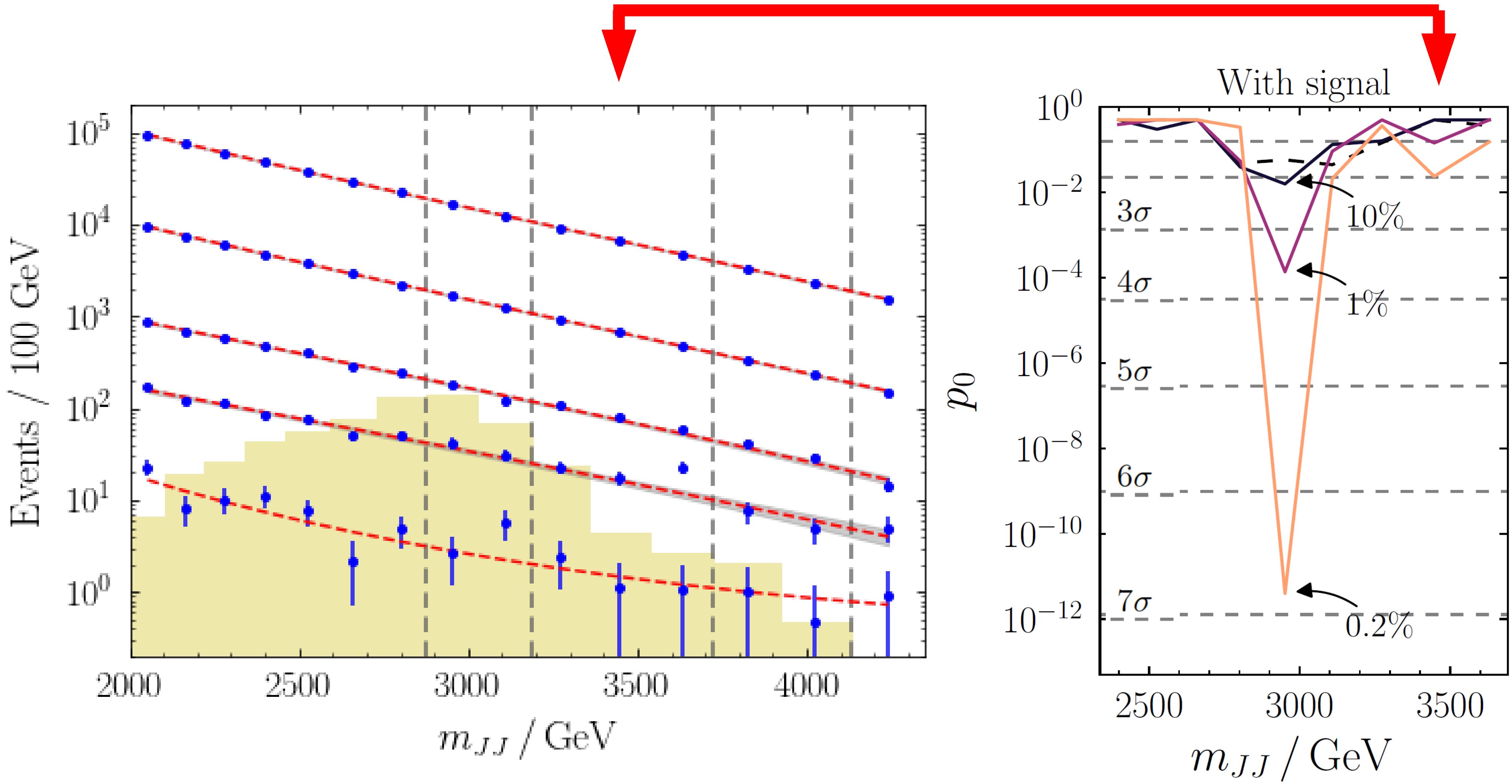


Events / 100 GeV
m_{JJ} / GeV
2000
2500
3000
3500
4000



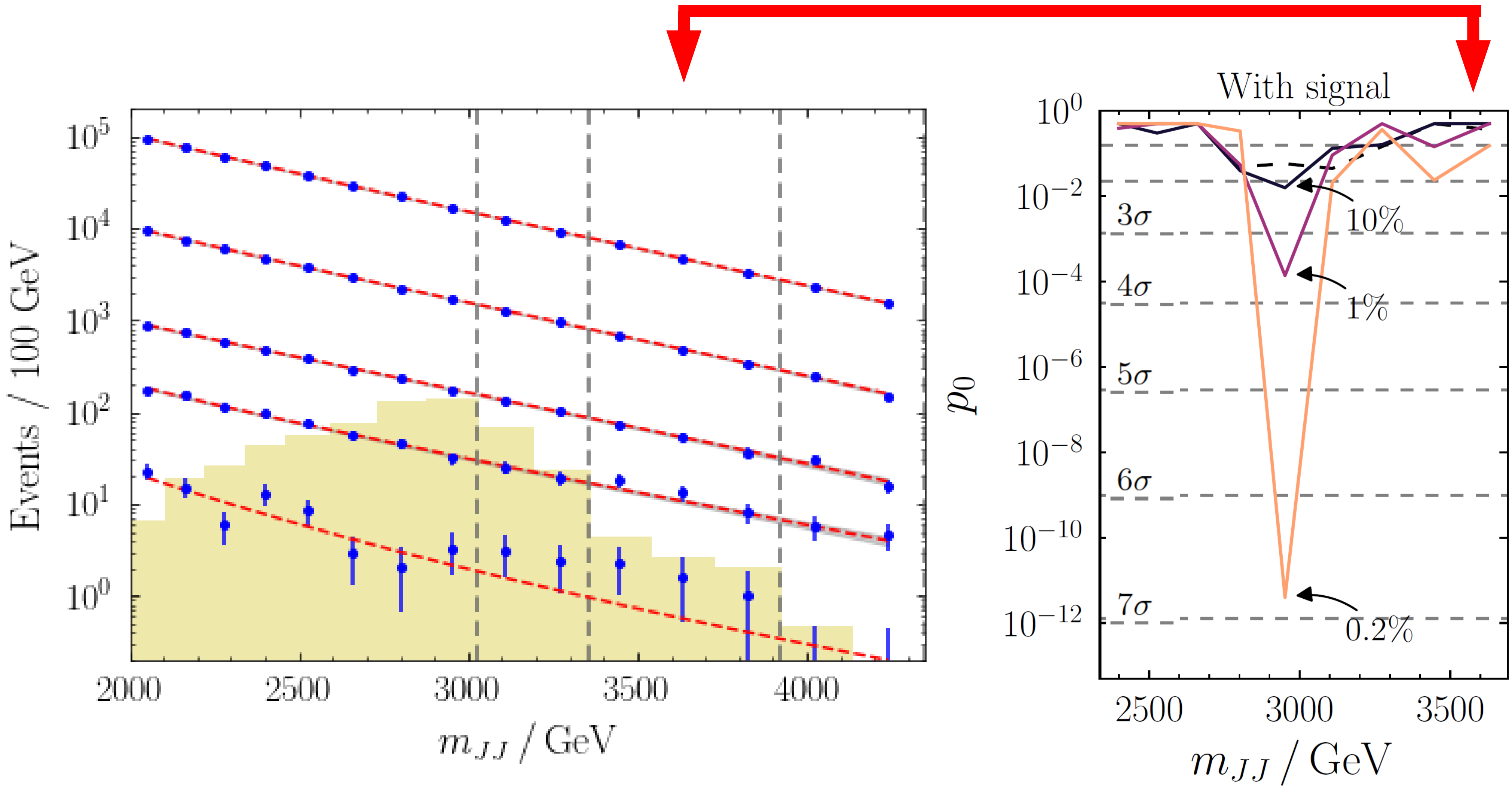
With signal
p_0
m_{JJ} / GeV
3σ
4σ
5σ
6σ
7σ
10%
1%
0.2%
2500
3000
3500

Mass Scan



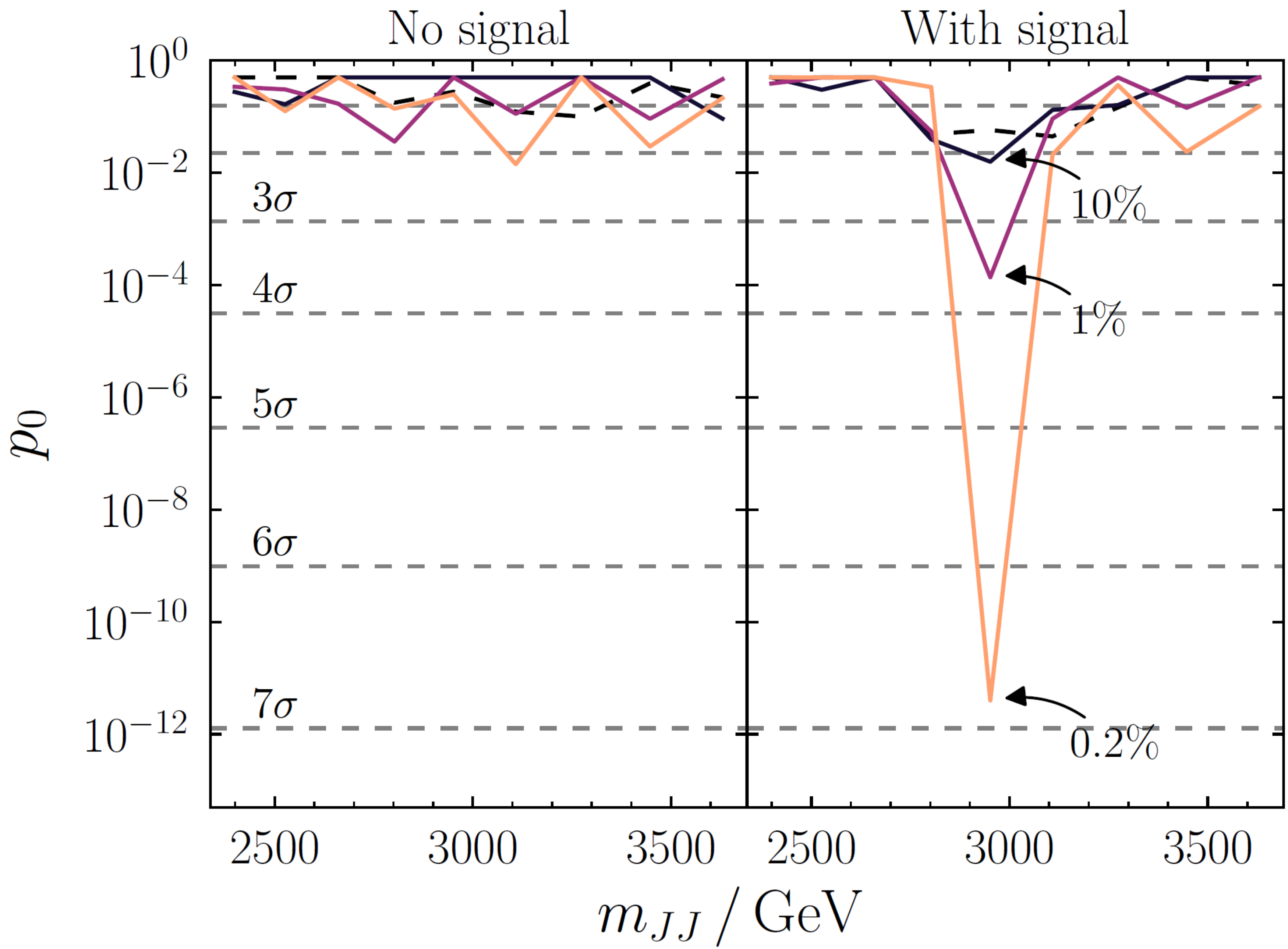
With signal
Events / 100 GeV
m_{JJ} / GeV
p_0
3σ
4σ
5σ
6σ
7σ
10%
1%
0.2%

Mass Scan



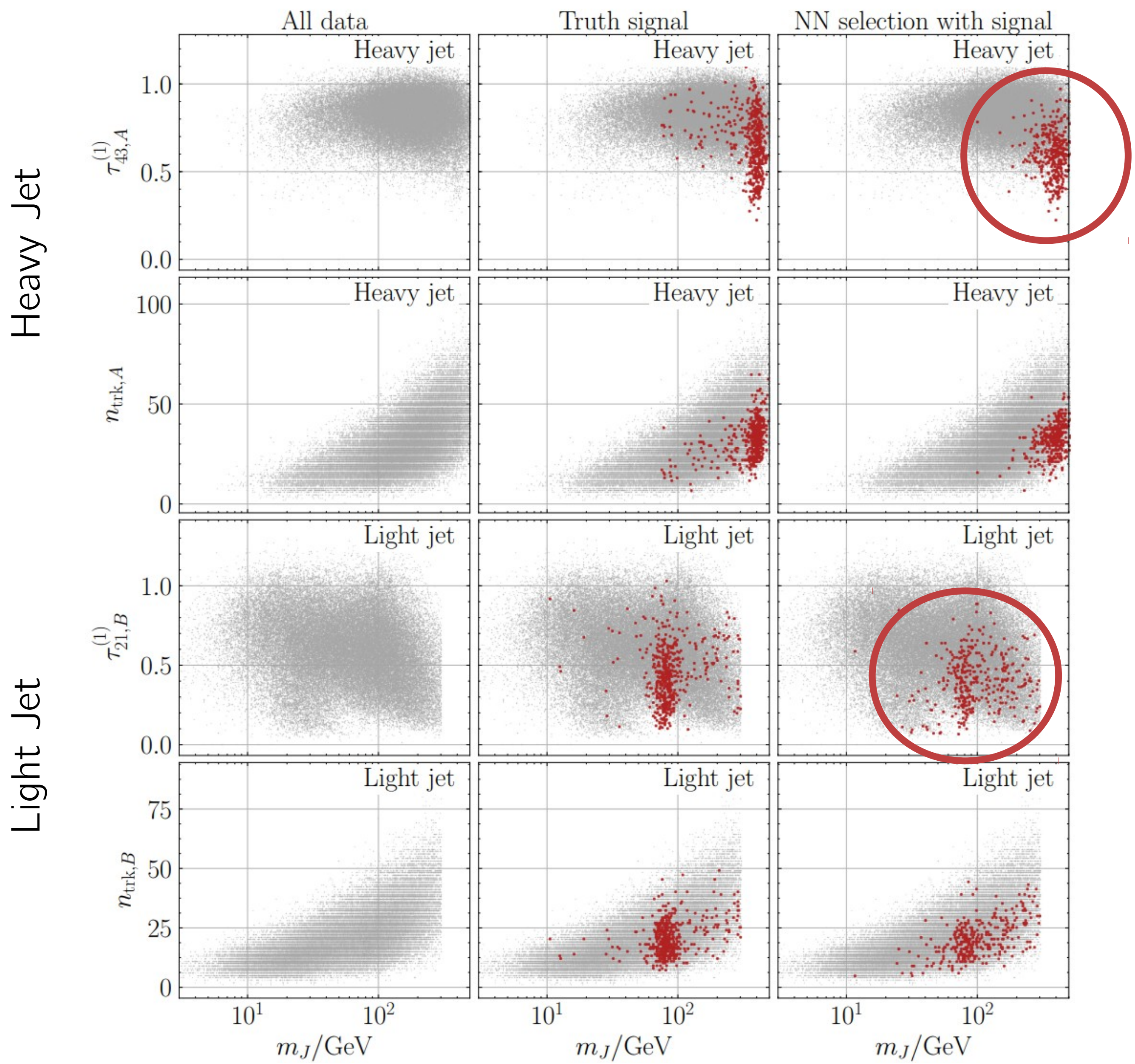
Events / 100 GeV
10^5
10^4
10^3
10^2
10^1
10^0
2000
2500
3000
3500
4000
m_{JJ} / GeV
With signal
10^0
10^{-2}
10^{-4}
10^{-6}
10^{-8}
10^{-10}
10^{-12}
p_0
3σ
4σ
5σ
6σ
7σ
10%
1%
0.2%
2500
3000
3500
m_{JJ} / GeV

Mass Scan



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

Signal Characteristics



400 GeV jet
With 4 prongs

80 GeV jet
With 2 prongs

Summary

1) **Factorize** space of observables into:

   a) One **test observable** (e.g. mJJ) in which bg is smooth and signal has a sharp feature (*doesn't need to be a bump*).

   b) An additional space of **auxiliary observables** (either particle 4-vectors or expert features).

2) May need to decorrelate auxiliary observables form test.

3) Define signal and sideband regions based on test observable

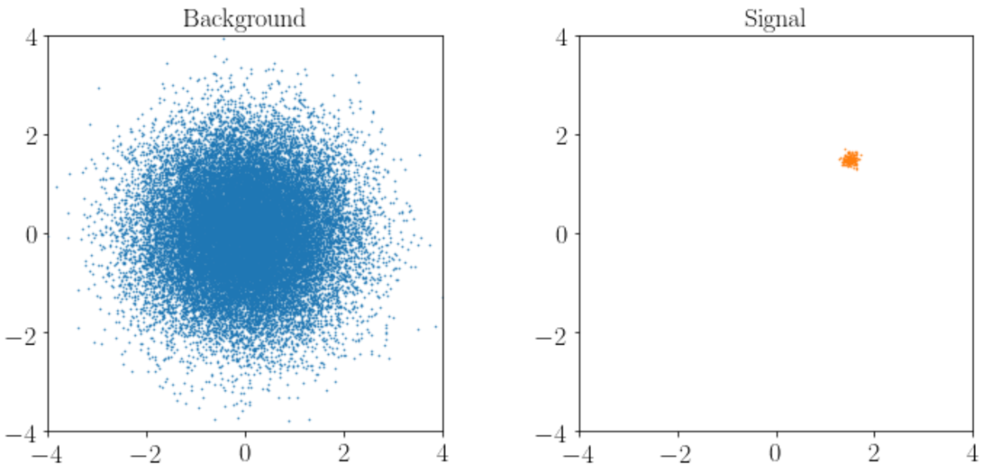
4) **Train NN on auxiliary observables** to discriminate sideband from signal region.

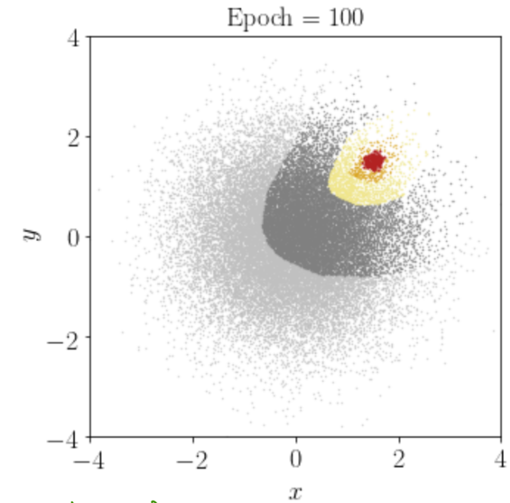
5) Use NN output as a selection cut to select events in a statistically independent sample.

6) **Perform a shape-based hypothesis test on the test observable.**

ML with mixed samples

[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





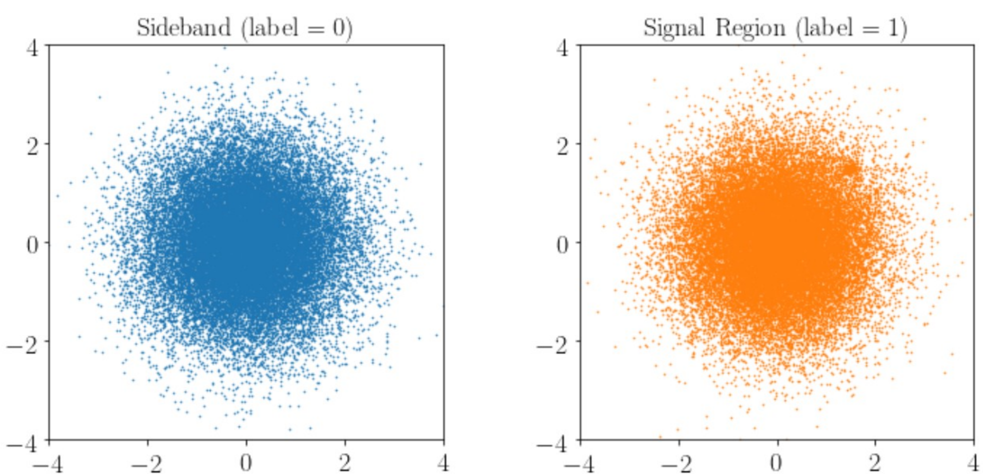
Classification *with* labels (Fully Supervised)

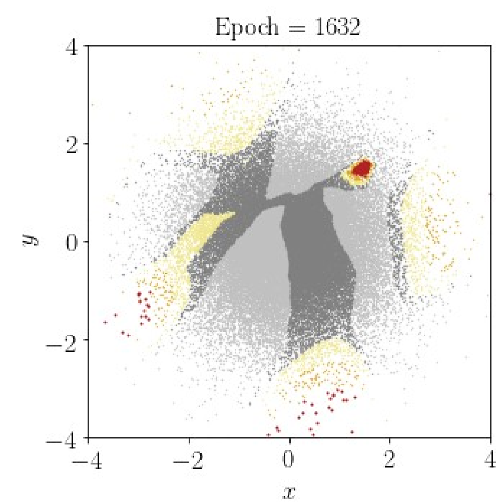
$$\frac{p(\text{data}\mid\text{sig})}{p(\text{data}\mid\text{bg})}$$

Monotonic Rescaling

$$\frac{p(\text{data}\mid 1)}{p(\text{data}\mid 0)}$$

Classification ***'without'*** **la**bels (**CWoLa**)





Machine Learning for Jets

Simulation $\neq$ data

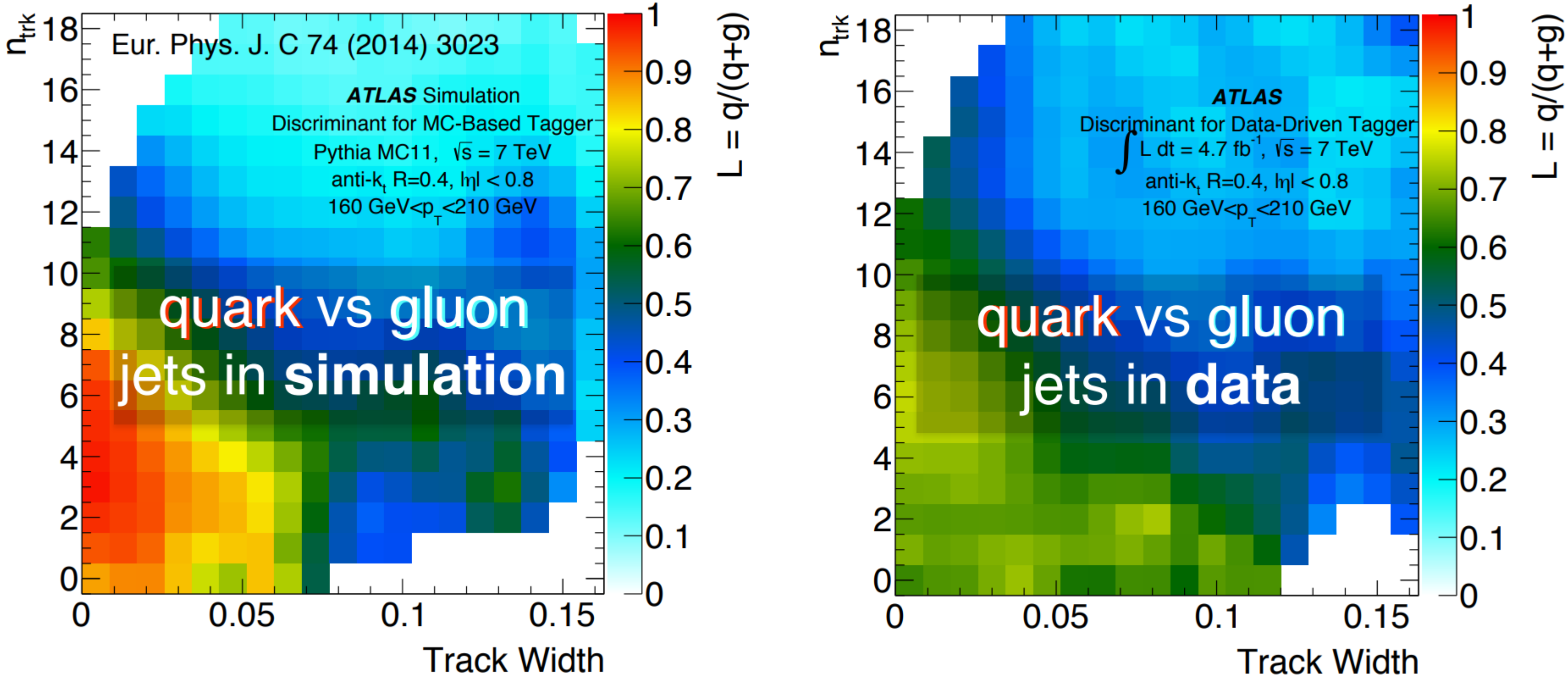
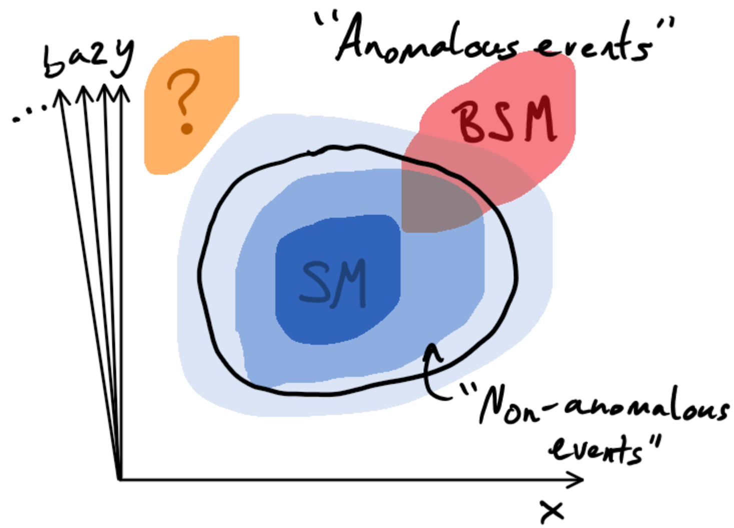


Figure taken from Ben Nachman's talk at BOOST 2018
https://indico.cern.ch/event/649482/contributions/2993322/attachments/1688082/2715256/WeakSupervision_BOOST2018.pdf

Anomaly Detection Landscape

“Anomalous Event Detection”



Autoencoders
(weak supervision)

[1808.08992] Marco Farina, Yuichiro Nakai, David Shih
[1808.08979] Theo Heimel, Gregor Kasieczka, Tilman Plehn, Jennifer M. Thompson
[1811.10276] Olmo Cerri, Thong Nguyen, Maurizio Pierini, Maria Spiropulu, Jean-Roch Vlimant
[1903.02032] Tuhin Roy, Aravind Vijay

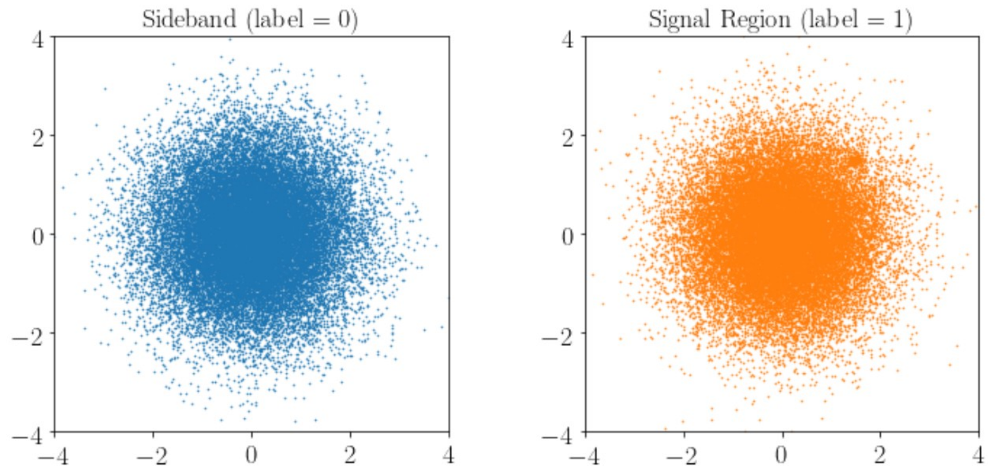
‘Model independent training sample’
(fully supervised)

[1709.01087] Jack H Collins, Rashmish Mishra, Juan Antonio Aguilar-Saavedra

(See also [1707.07084] Amit Chakraborty, Abhishek Iyer, Tuhin Roy for similar, non-ML ideas)

Anomaly Detection Landscape

"Anomalous Overdensity Detection"



Data vs Data
Or Simulation vs data

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

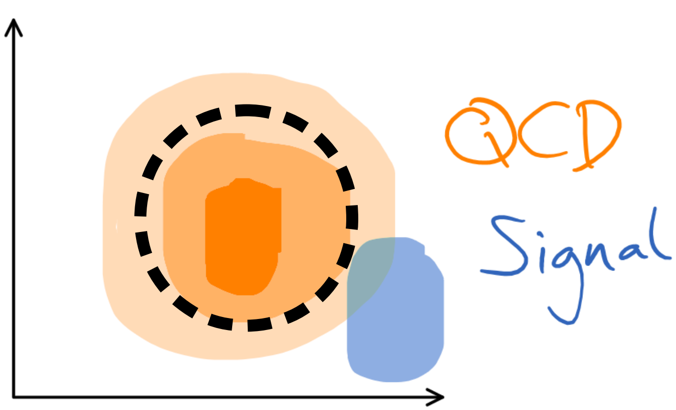
Tagger relies on there being sufficient signal statistics for training.

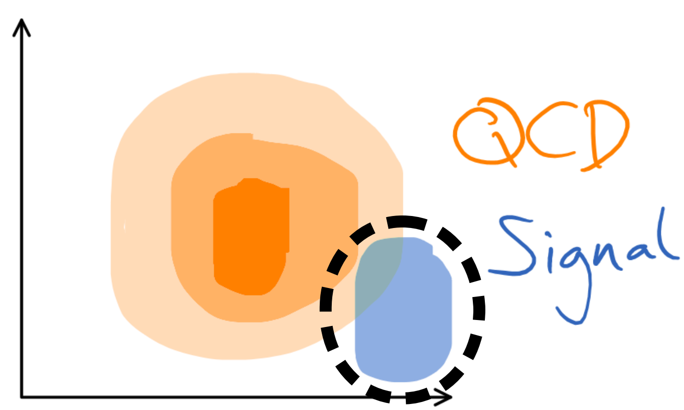
Tagger can learn the *specific* peculiar features of the signal, and so **improves with greater signal rate**, and allows for **signal characterization**.

Stronger in limit of very low signal statistics

??

Stronger in limit of very high signal statistics





Performance Comparison

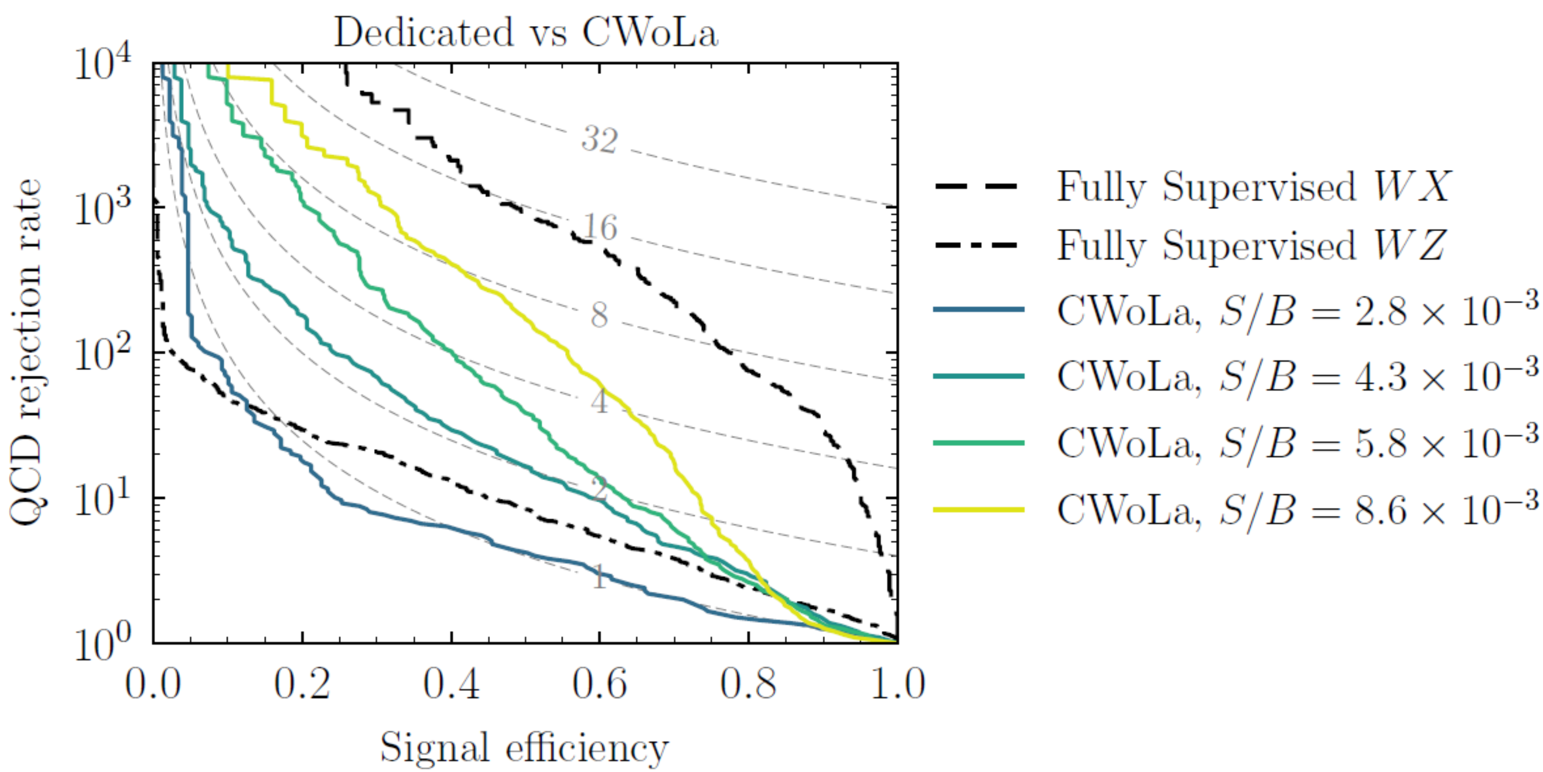


Figure 11. Truth-label ROC curves for taggers trained using CWoLa with varying number of signal events, compared to those for a dedicated tagger trained on pure signal and background samples (solid black) and one trained to discriminate W and Z jets from QCD (dashed black). The CWoLa examples have $B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

Nested Cross-Validation

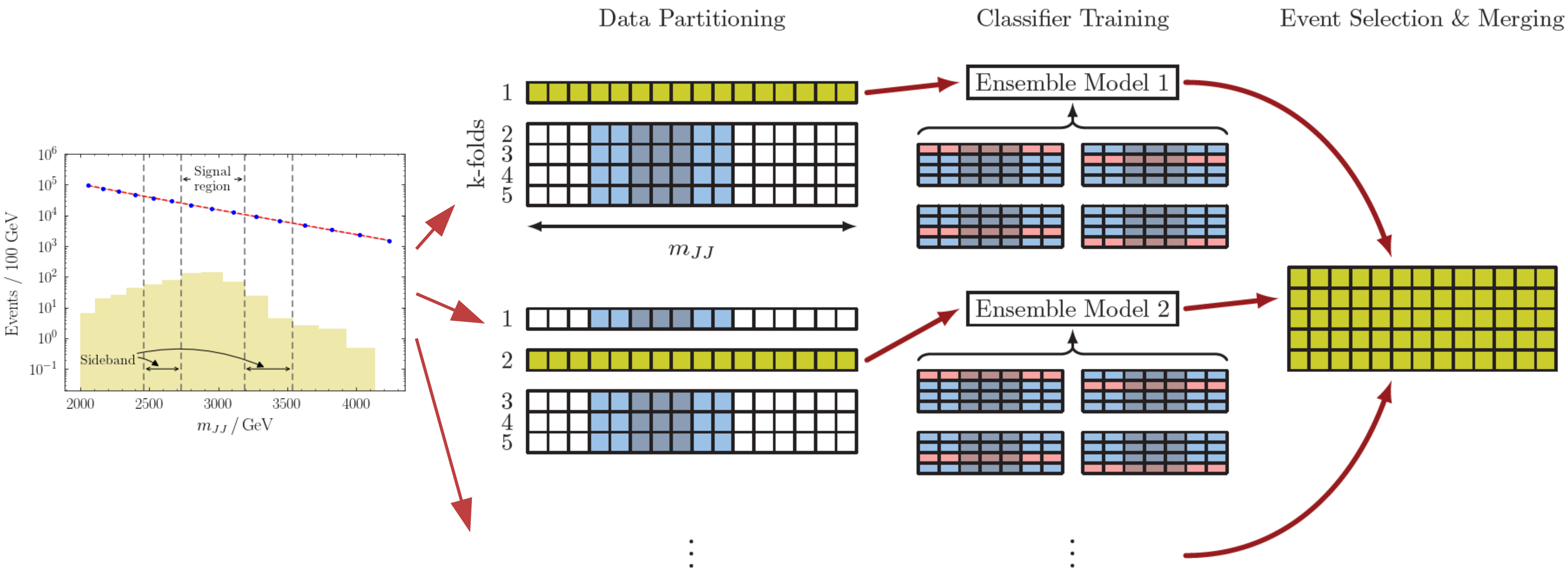


Figure 7. Illustration of the nested cross-validation procedure. **Left:** the dataset is randomly partitioned bin-by-bin into five groups. **Center:** for each group, an ensemble classifier is trained on the remaining groups. For each of the four possible combinations of these four groups into three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

Toy Statistics

$$\mathcal{L}(\mu, \theta) = \text{Poiss}(n|b + \theta + \mu)e^{-\theta^2/(2\sigma^2)}$$

$$\lambda_0 = \frac{\mathcal{L}(\mu = 0, \hat{\hat{\theta}})}{\mathcal{L}(\hat{\mu}, \hat{\theta})}$$