Anomaly detection with machine learning at the LHC

Benjamin Nachman

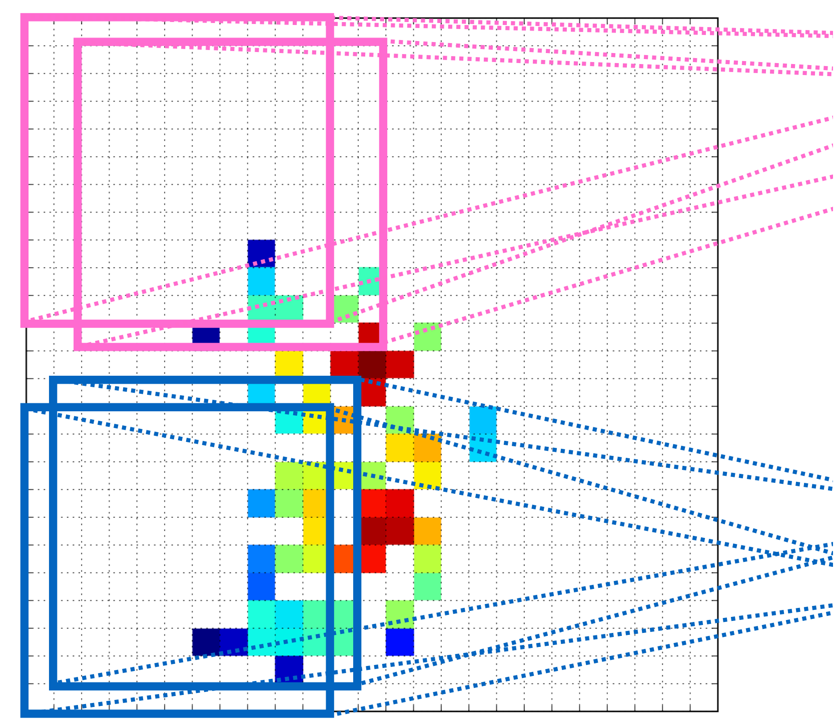
Lawrence Berkeley National Laboratory

bpnachman.com

bpnachman@lbl.gov

@bpnachman

bnachman



BERKELEY LAB

BIDS

BERKELEY INSTITUTE FOR DATA SCIENCE

Liverpool HEP Seminar

Nov. 11, 2020

Questions in fundamental physics

2

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

Why is the Higgs boson so light?

Hierarchy problem

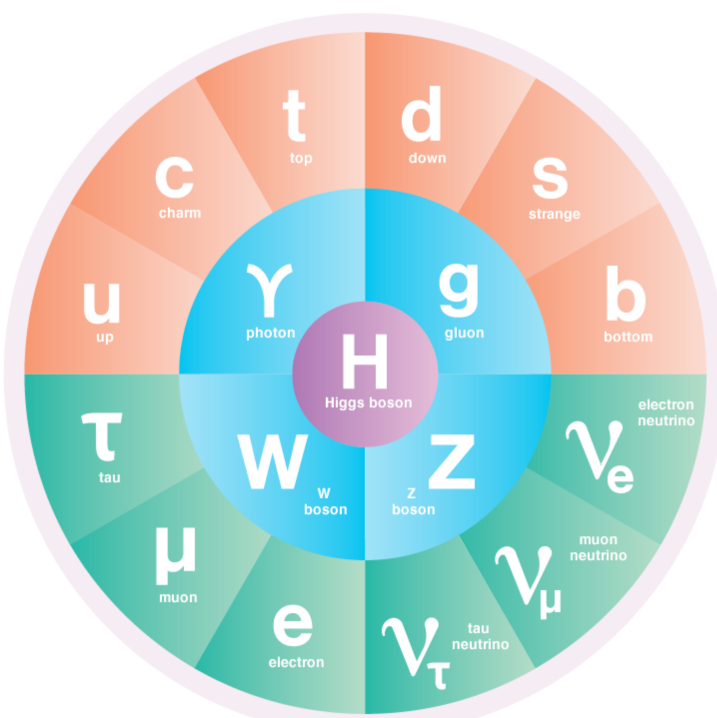
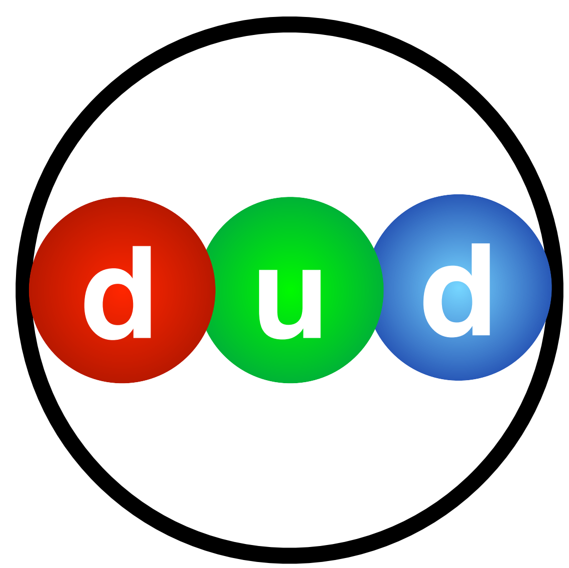


image source: symmetry magazine

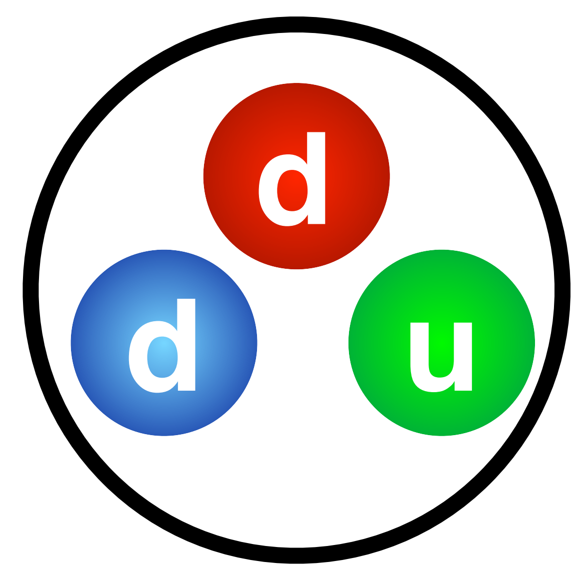
See also: quantum gravity

Why do neutrons have no dipole moment?

Strong CP



Reality



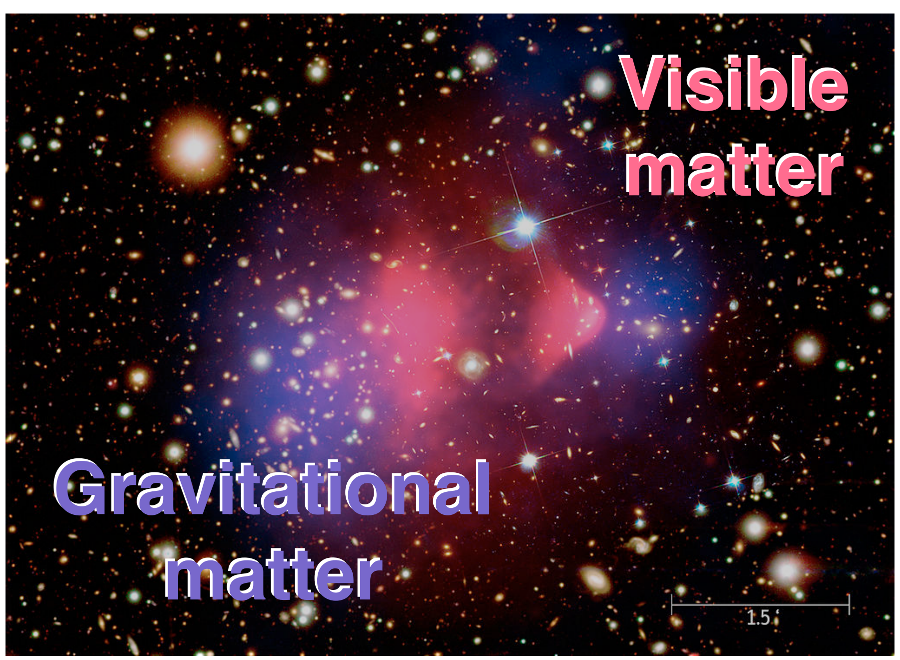
>99% of pictures on the internet

Questions in fundamental physics

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

What is the extra gravitational matter?

Dark Matter

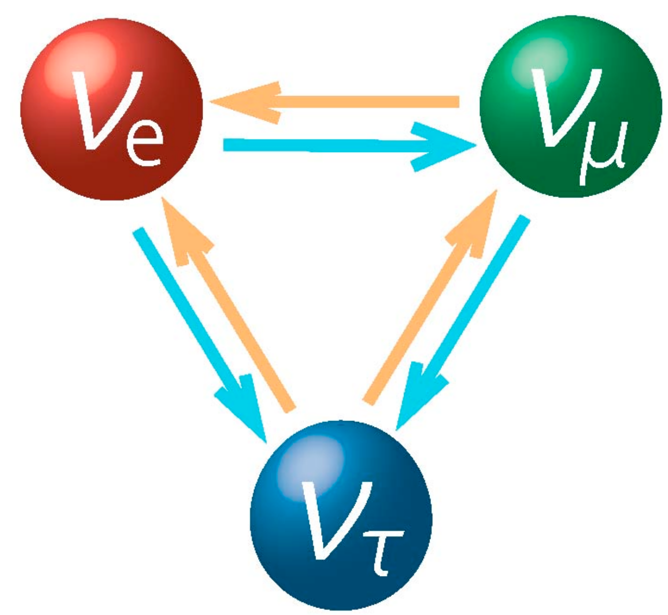


NASA/CXC/M. Weiss - Chandra X-Ray Observatory: 1E 0657-56

See also: dark energy

Why do neutrinos have a mass?

Flavor puzzles



See also: Where did all the anti-particles go? (Baryogengesis)

image source: T2K collaboration

Questions in fundamental physics

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

Dark matter

Strong CP

Hierarchy problem

Dark energy

Flavor puzzles

Baryogengesis

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

Three possibilities

Questions in fundamental physics

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

Dark matter
Strong CP
Hierarchy problem
Dark energy
Flavor puzzles
Baryogengesis

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

Three possibilities

(1) Nothing new at accessible energies

Questions in fundamental physics

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

Dark matter

Hierarchy problem

Strong CP

Flavor puzzles

Dark energy

Baryogengesis

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

Three possibilities

(1) Nothing new at accessible energies

(2) Patience! (new physics is rare)

Questions in fundamental physics

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

Dark matter

Strong CP

Hierarchy problem

Dark energy

Flavor puzzles

Baryogengesis

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

Three possibilities

(1) Nothing new at accessible energies

(2) Patience! (new physics is rare)

(3) We are not looking in the right place

Questions in fundamental physics

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

Dark matter

Hierarchy problem

Strong CP

Flavor puzzles

Dark energy

Baryogengesis

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

Three possibilities

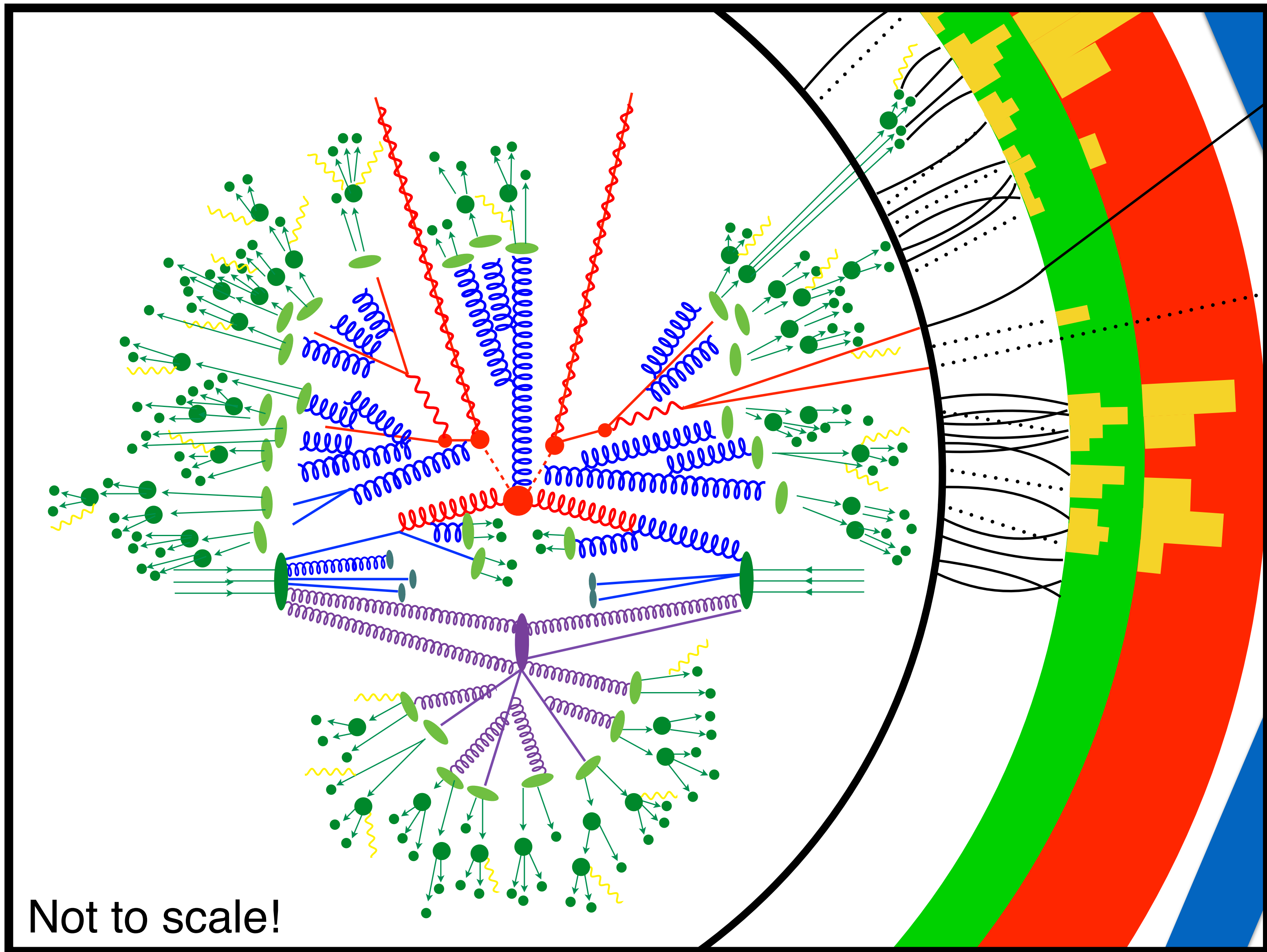
This is what keeps me up at night!

(3) We are not looking in the right place

Large Hadron Collider

Many of the deep questions in fundamental physics can be probed at the LHC.

Image inspired by JHEP 02 (2009) 007



Large Hadron Collider

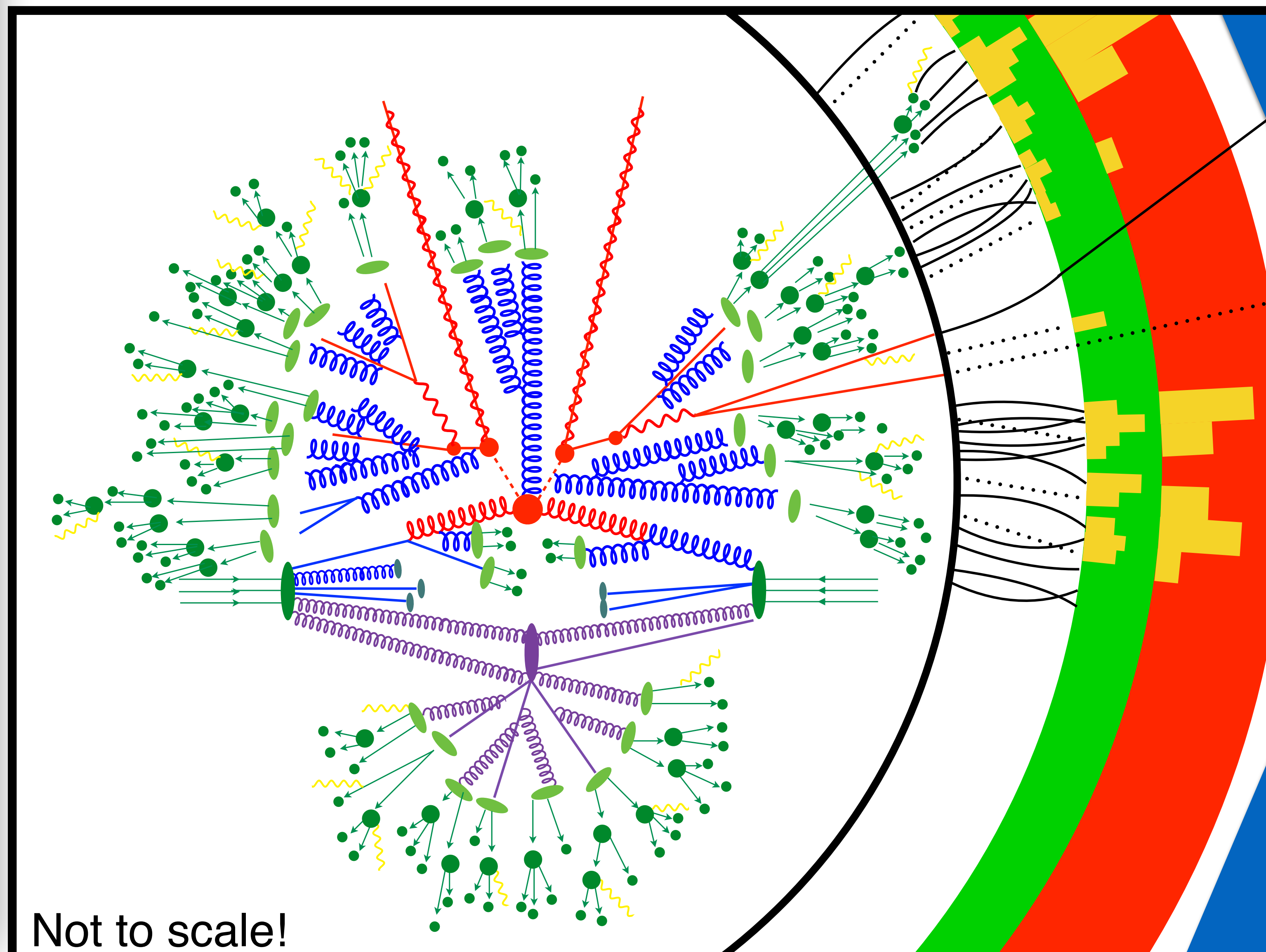
Many of the deep questions in fundamental physics can be probed at the LHC.

Key challenges (and opportunity!)

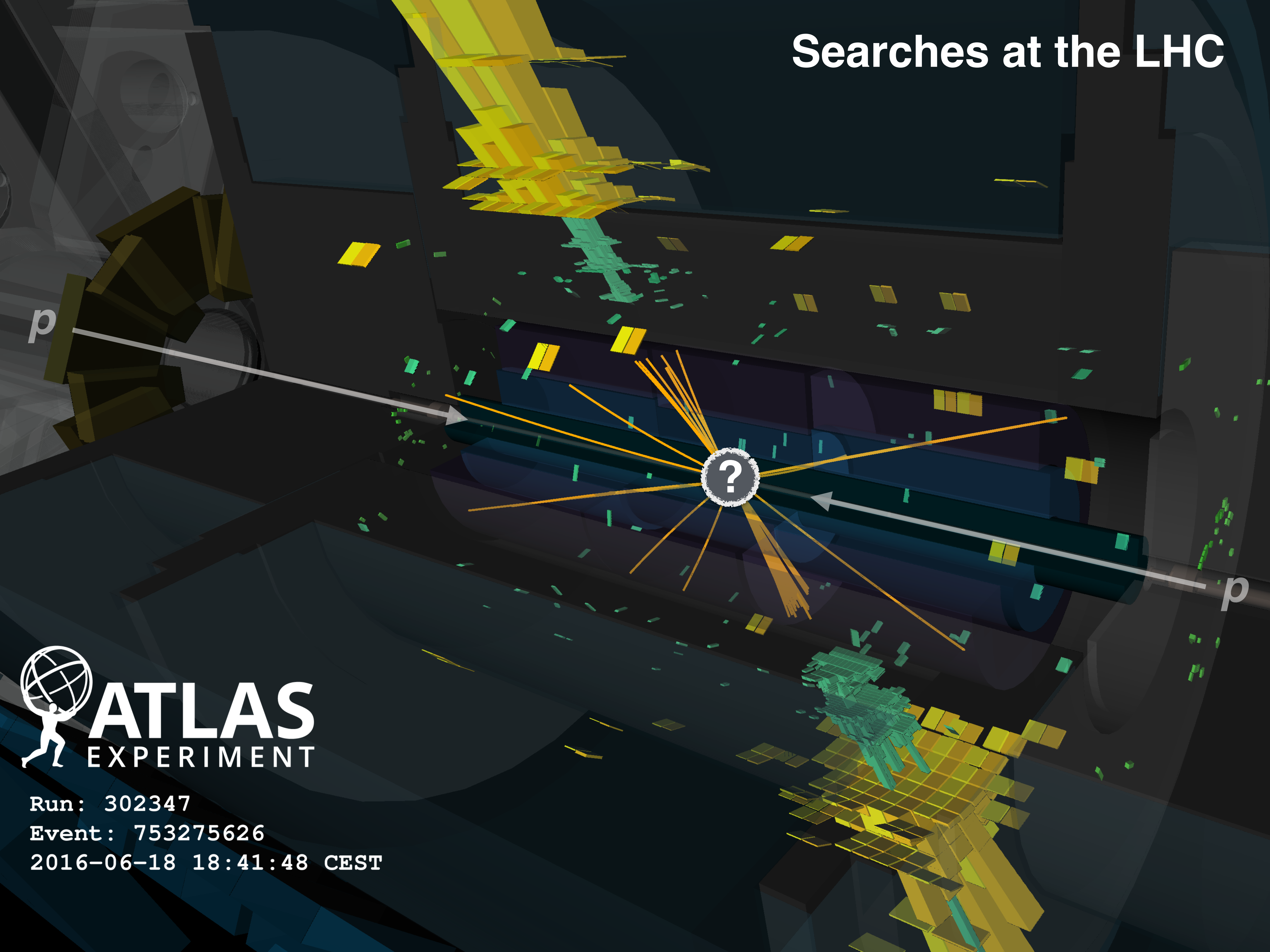
Typical collision events at the LHC produce **O(1000+)** particles

We detect these particles with **O(100 M)** readout channels

Image inspired by JHEP 02 (2009) 007

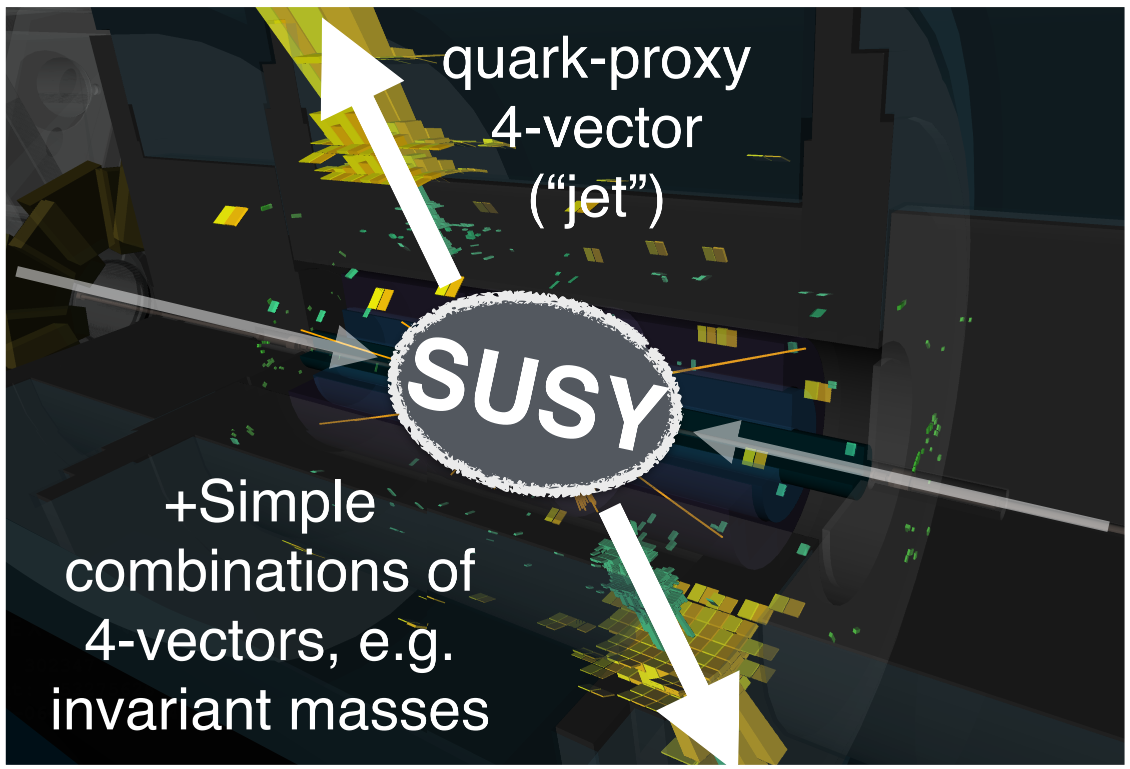


Searches at the LHC



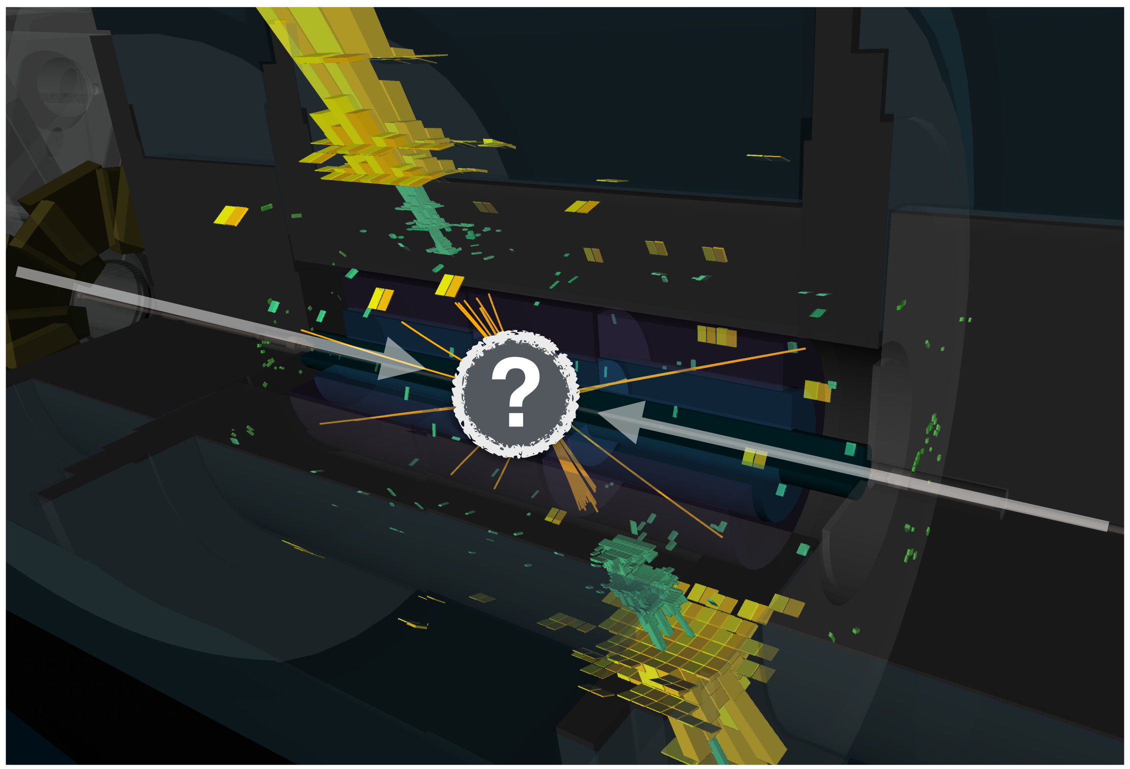
Current Search Paradigm

SUSY = Supersymmetry



(well-motivated) theory-biased
& low-dimensional observables

Current paradigm for searches



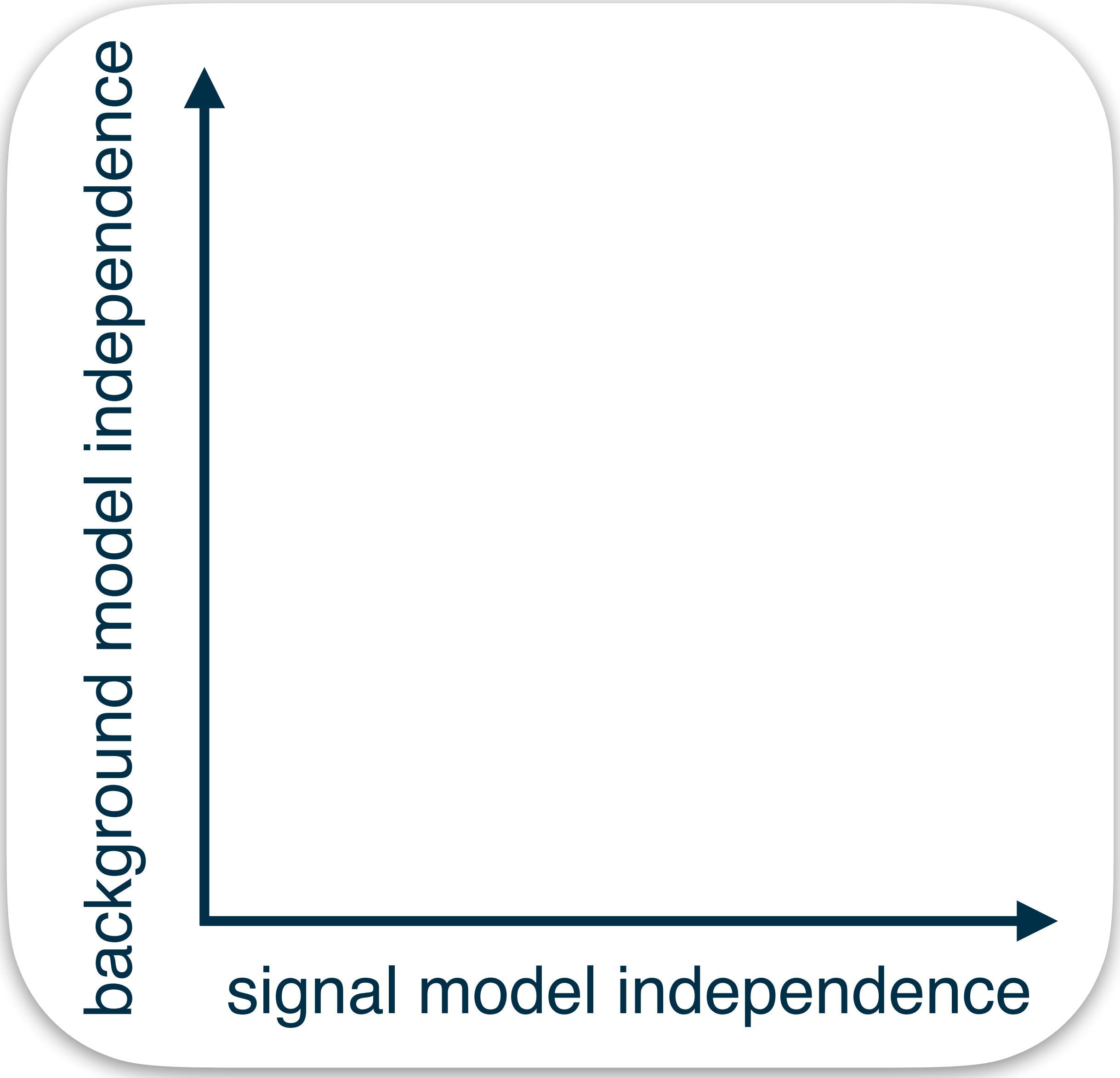
SUSY = Supersymmetry

Can we relax model assumptions and explore high-dimensional feature spaces?

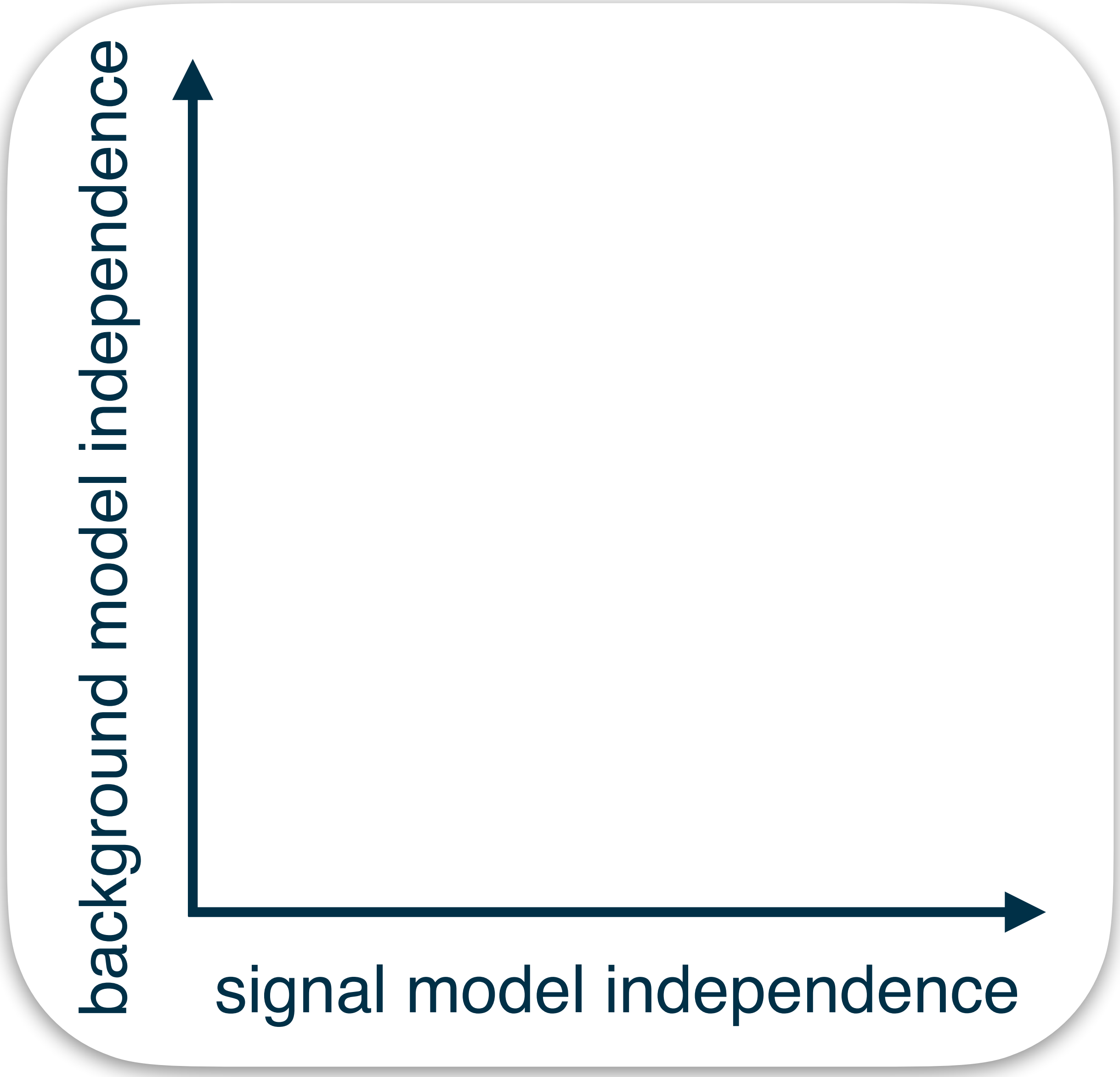
(well-motivated) theory-biased & low-dimensional observables

(clearly, we should still do model-dependent searches as well!)

Model dependence



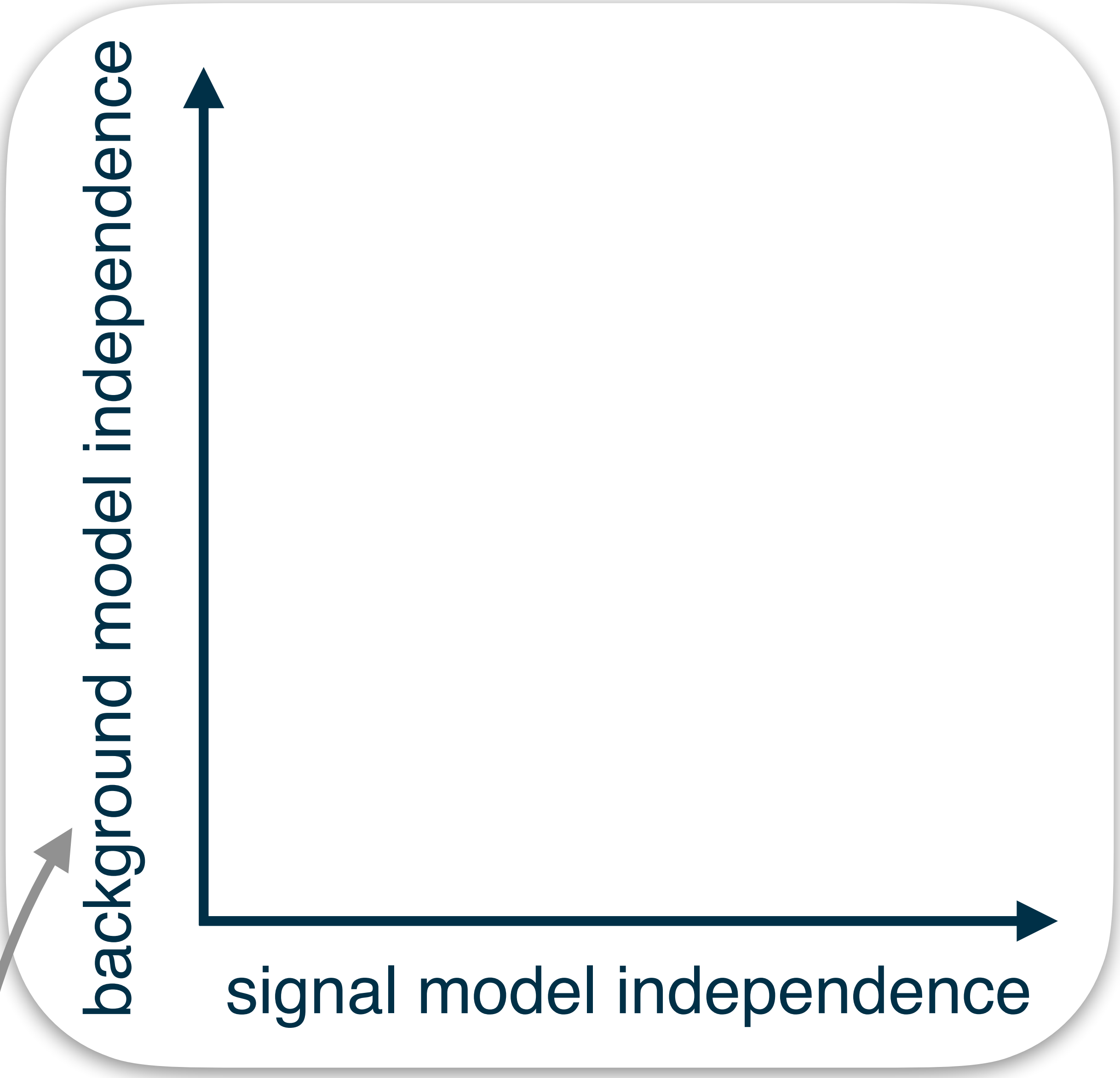
Signal sensitivity



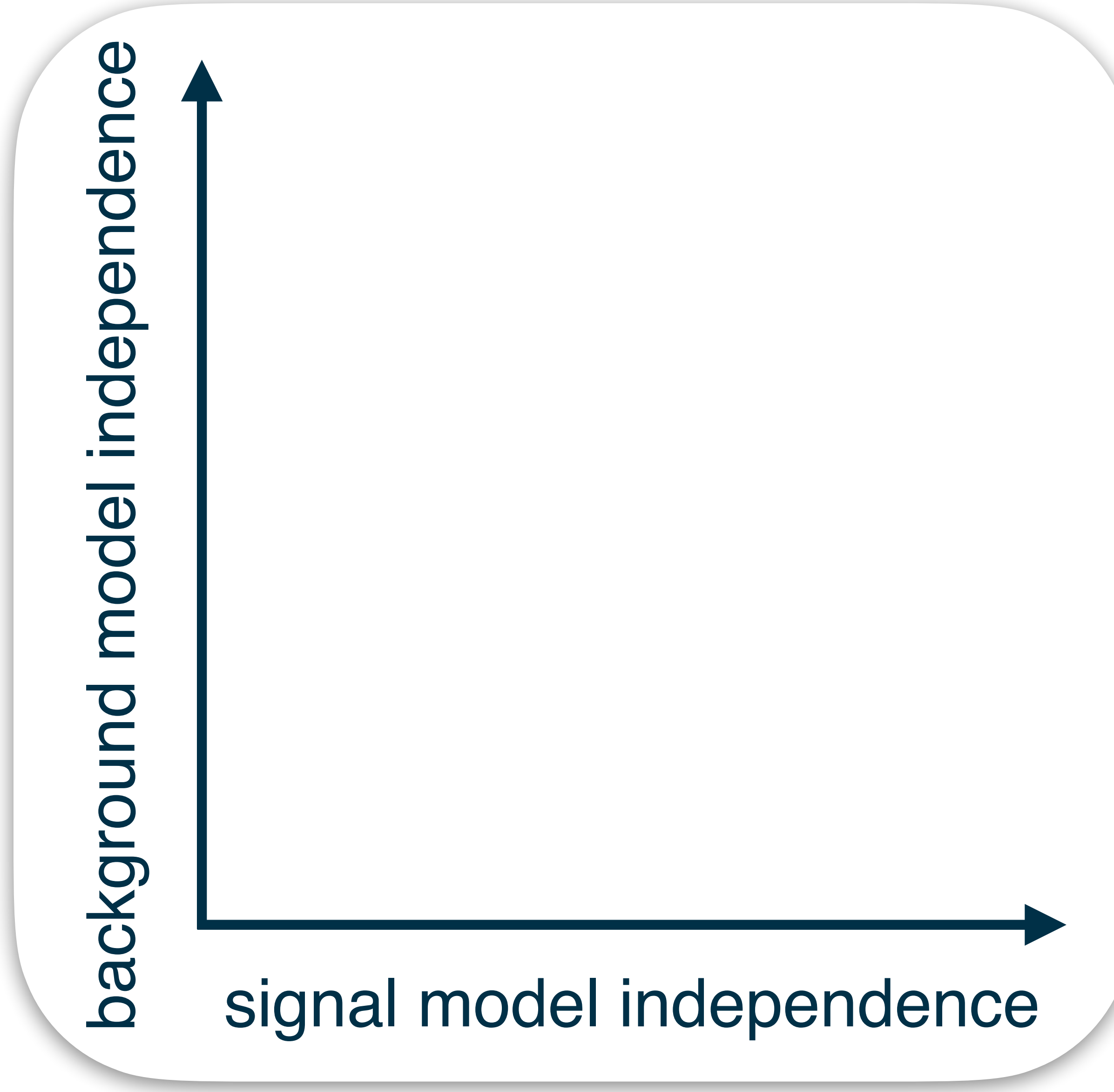
Background specificity

Suppose you want to search for a new signal process

Model dependence



Signal sensitivity

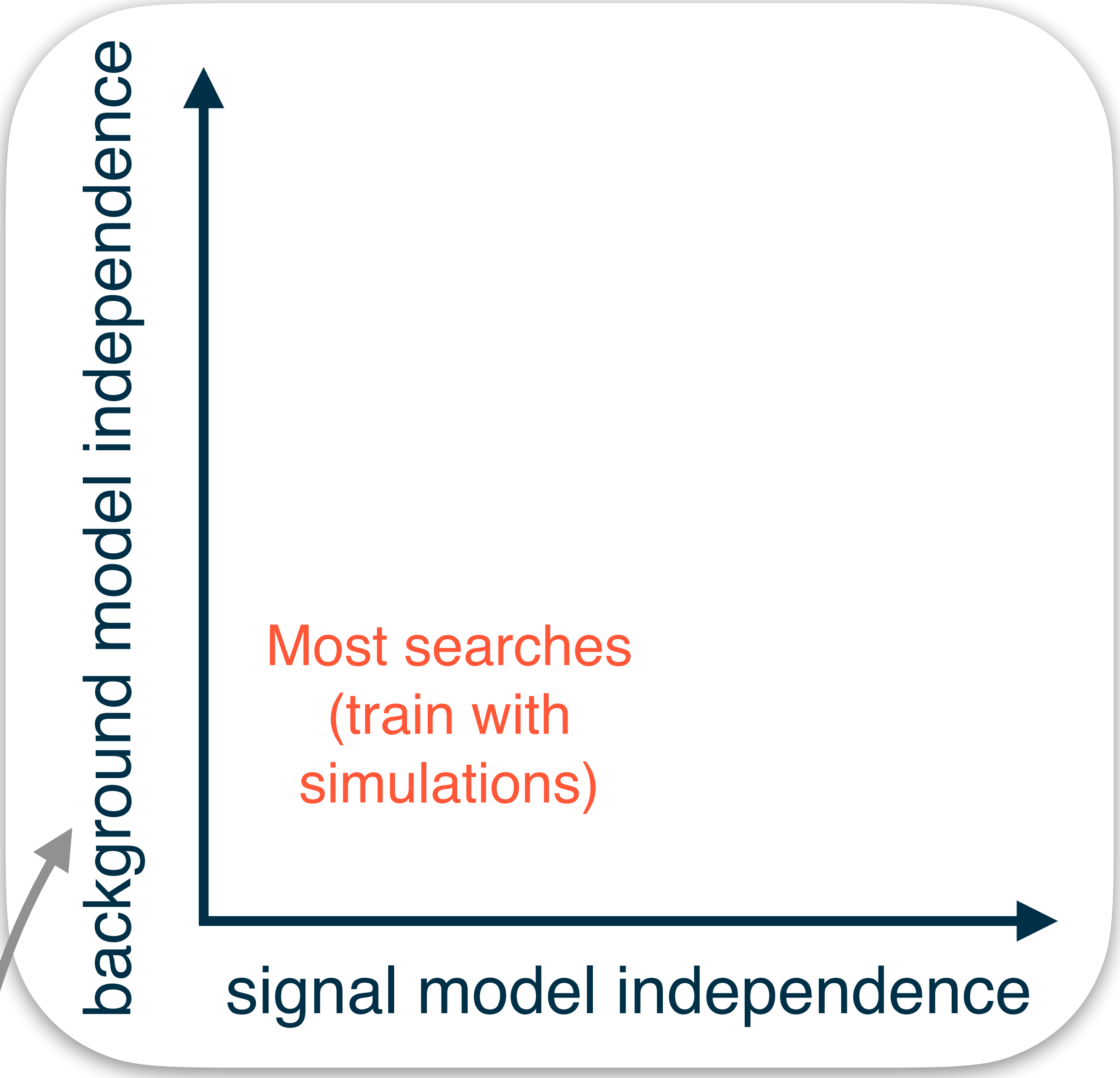


Background specificity

Standard Model

B. Nachman, D. Shih, Phys. Rev. D 101, 075042 (2020)

Model dependence

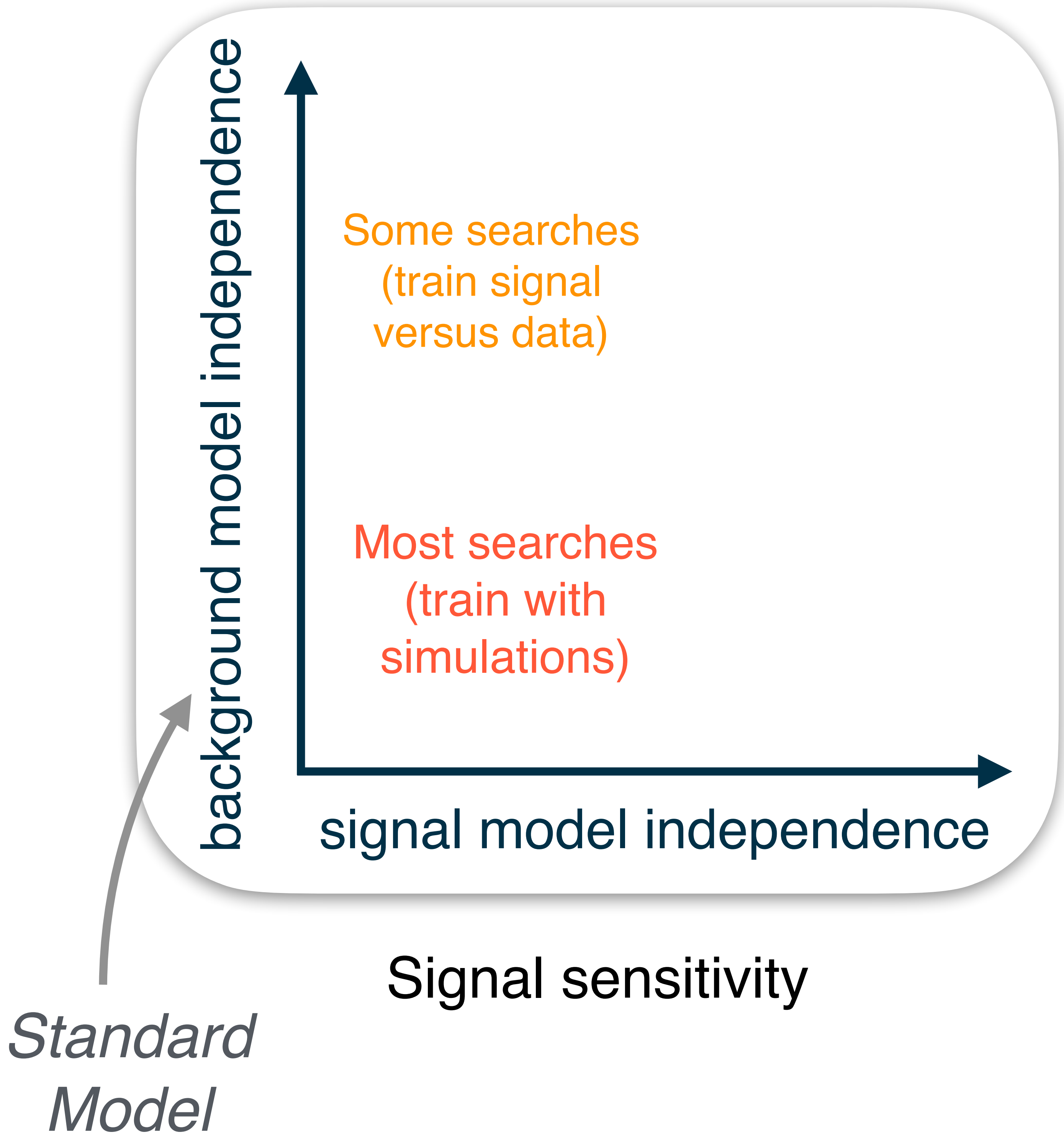


Signal sensitivity

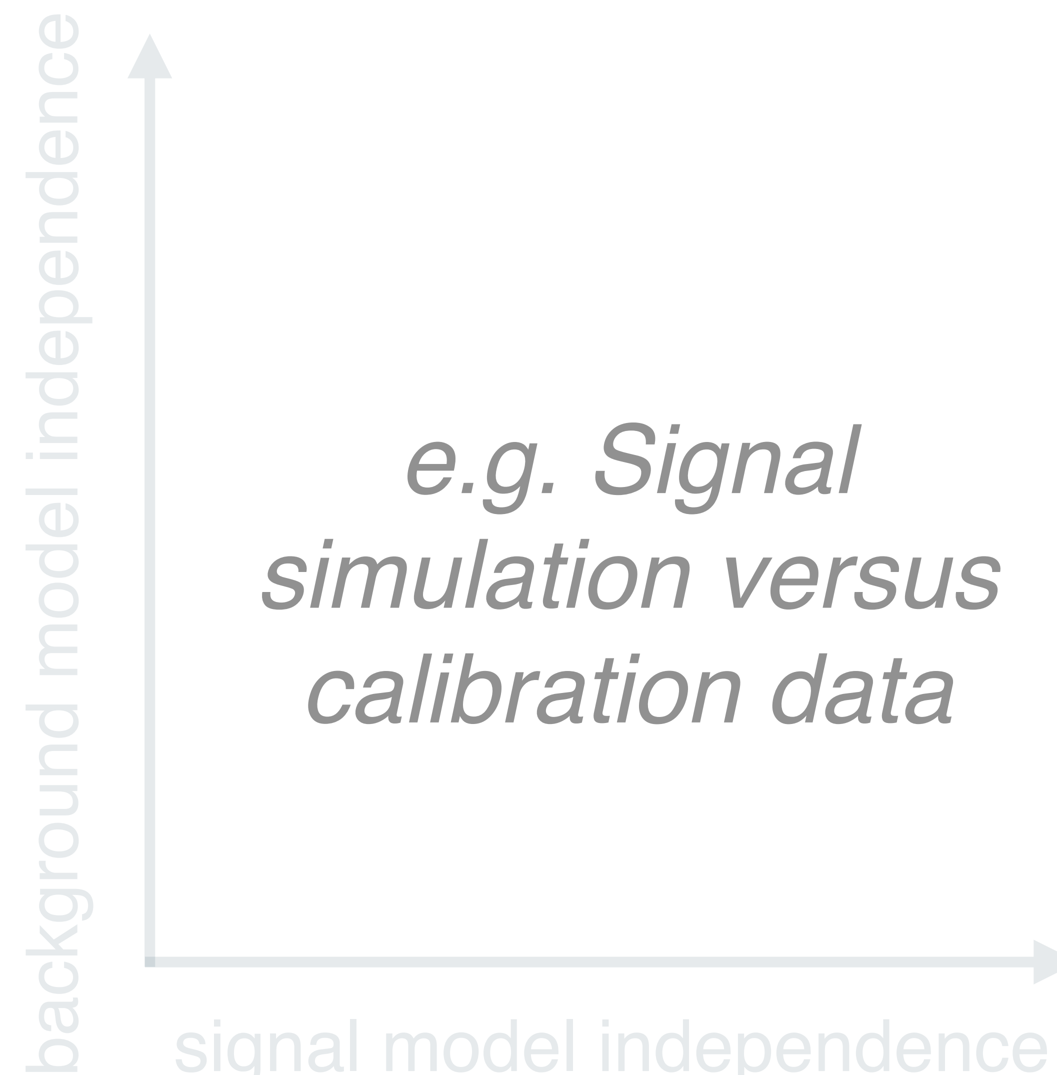
Standard Model

> 99% of searches at the LHC are of this type

Model dependence

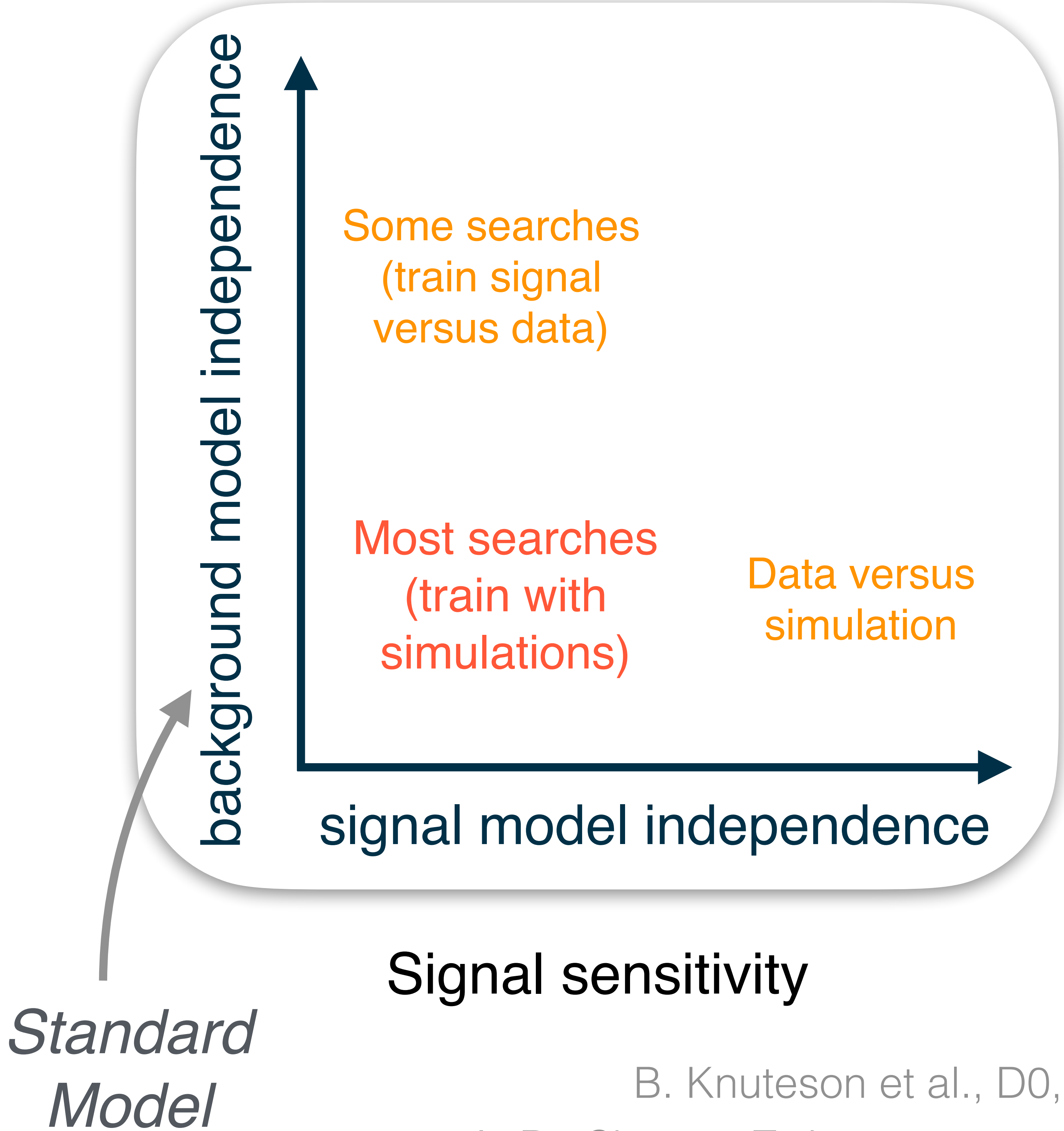


Signal sensitivity



Background specificity

Model dependence

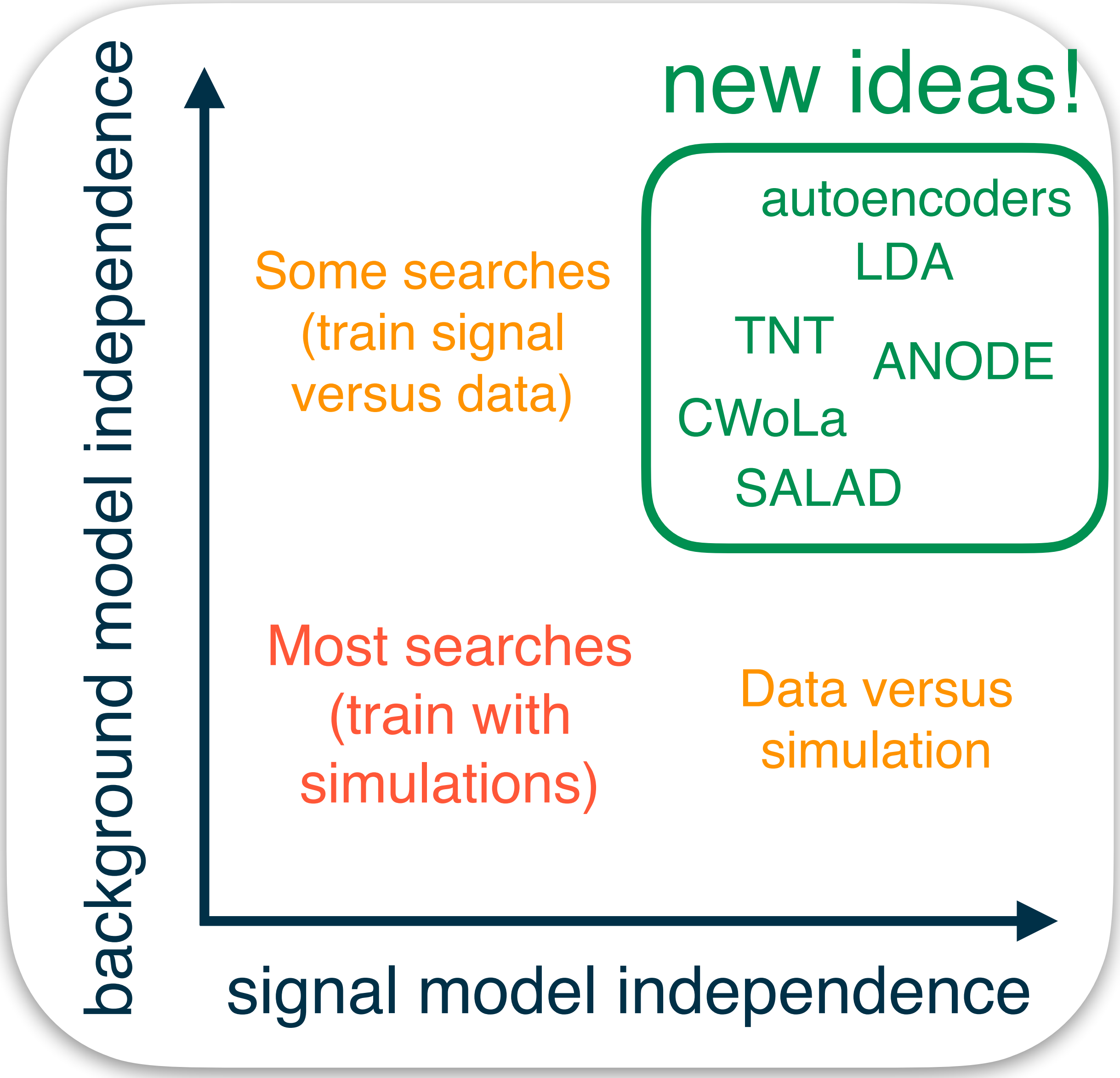


Signal sensitivity

signal model *independent*
background model *dependent*

B. Knuteson et al., D0, H1, CDF, CMS (“MUSiC”), ATLAS (“General Search”)
A. De Simone, T. Jacques, 1807.06038, A. Casa, Giovanna, 1809.02977, and others
R. T. D’Agnolo and A. Wulzer, PRD 99 (2019) 015014, R. T. D’Agnolo et al. 1912.12155

Model dependence



Signal sensitivity

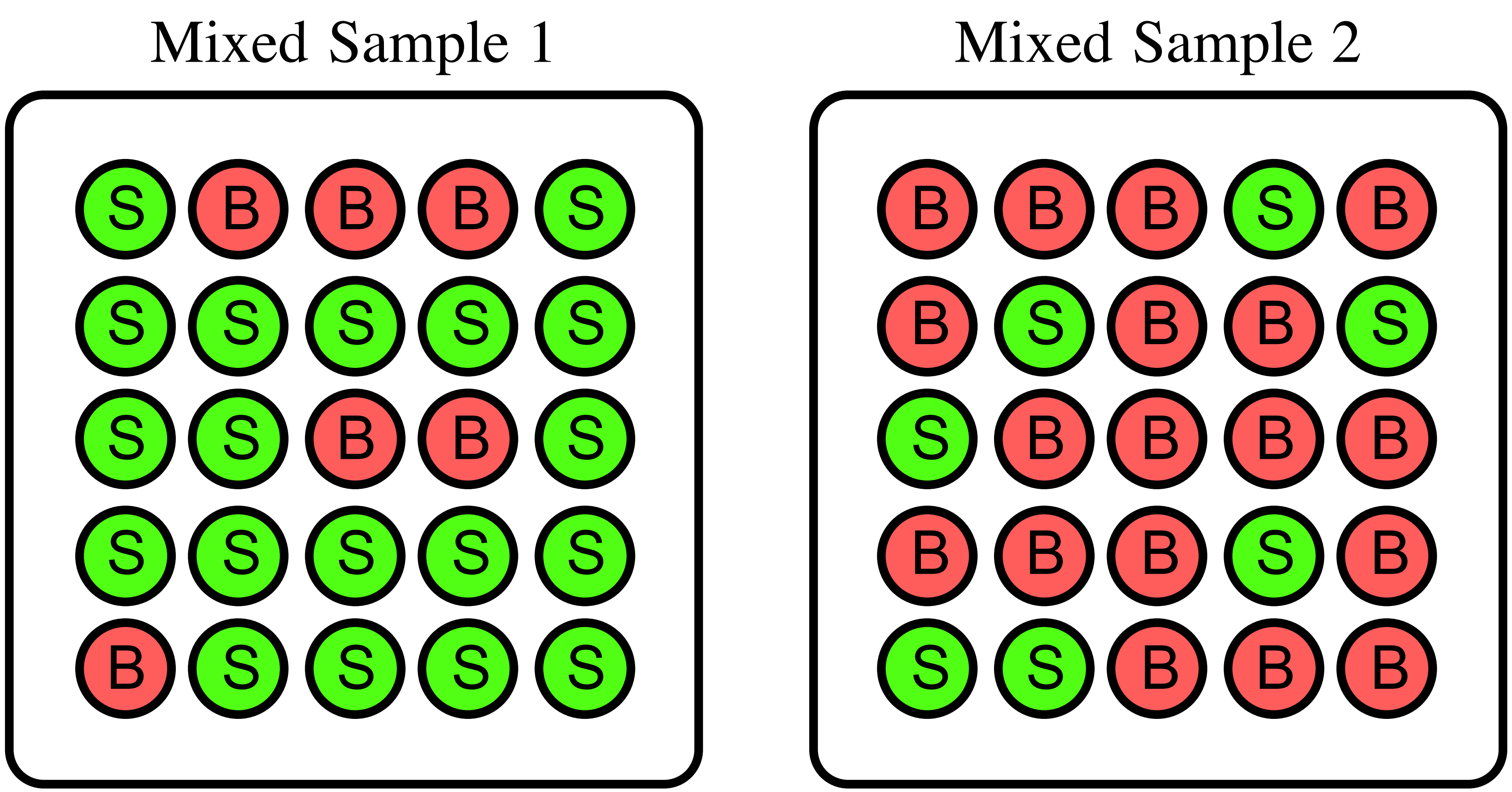
Can we develop new methods that also assume as little as possible about the signal and learn from data (no simulation)?

M. Farina, Y, Nakai, D. Shih, 1808.08992,
T. Heimel et al. SciPost Phys. 6 (2019) 030, and others
B. Dillon et al., PRD 100 (2019) 056002
B. Nachman, D. Shih, 2001.04990

O. Amram, C. Suarez, 2002.12376
J. Collins, K. Howe, B. Nachman, PRL 121 (2018) 241803
A. Andreassen, B. Nachman, D. Shih, 2001.05001

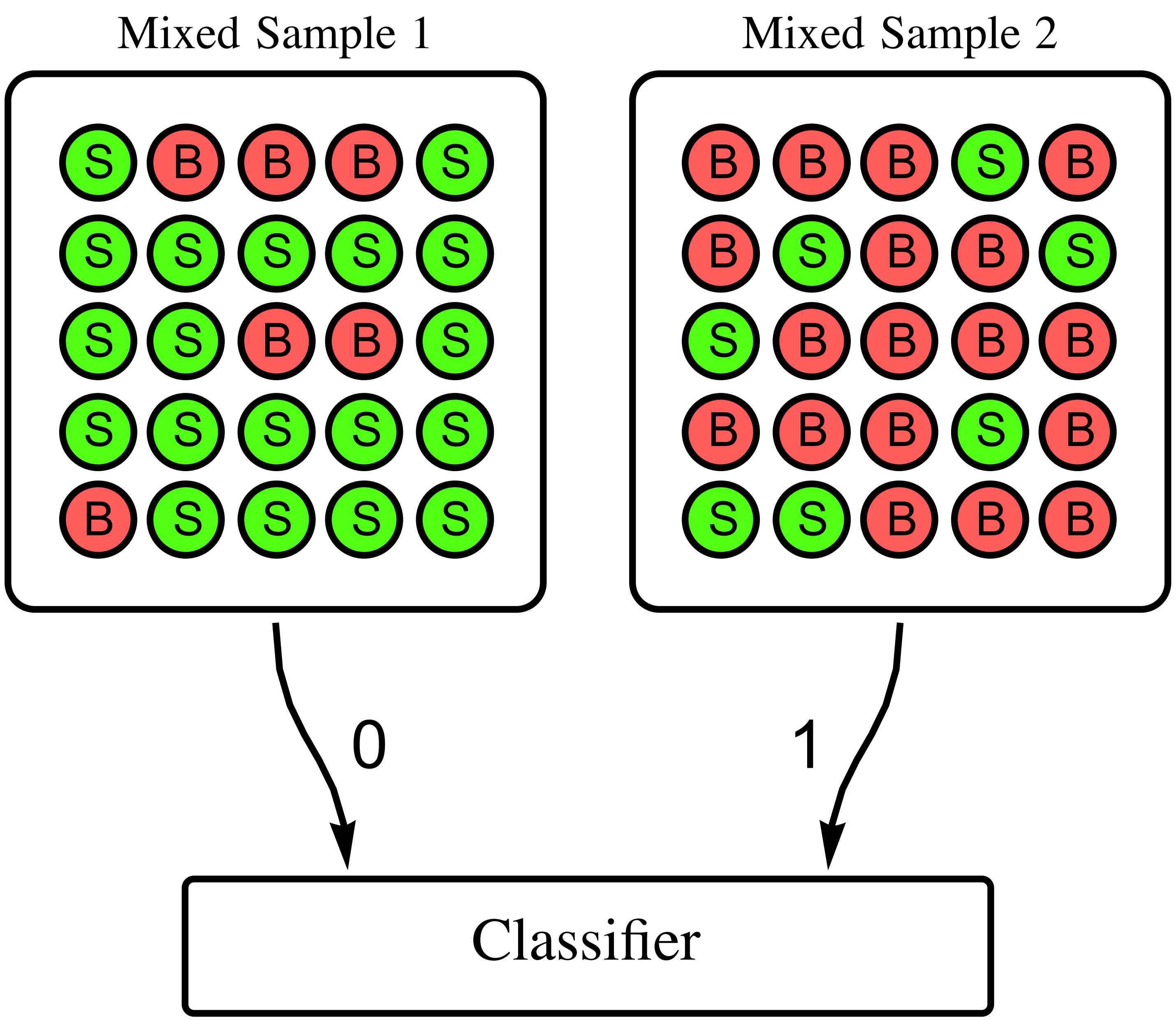
One idea: weak supervision

Can we learn signal events from two datasets that are mixtures of signal and background with no labels?



(we don't get to observe the color of the circles)

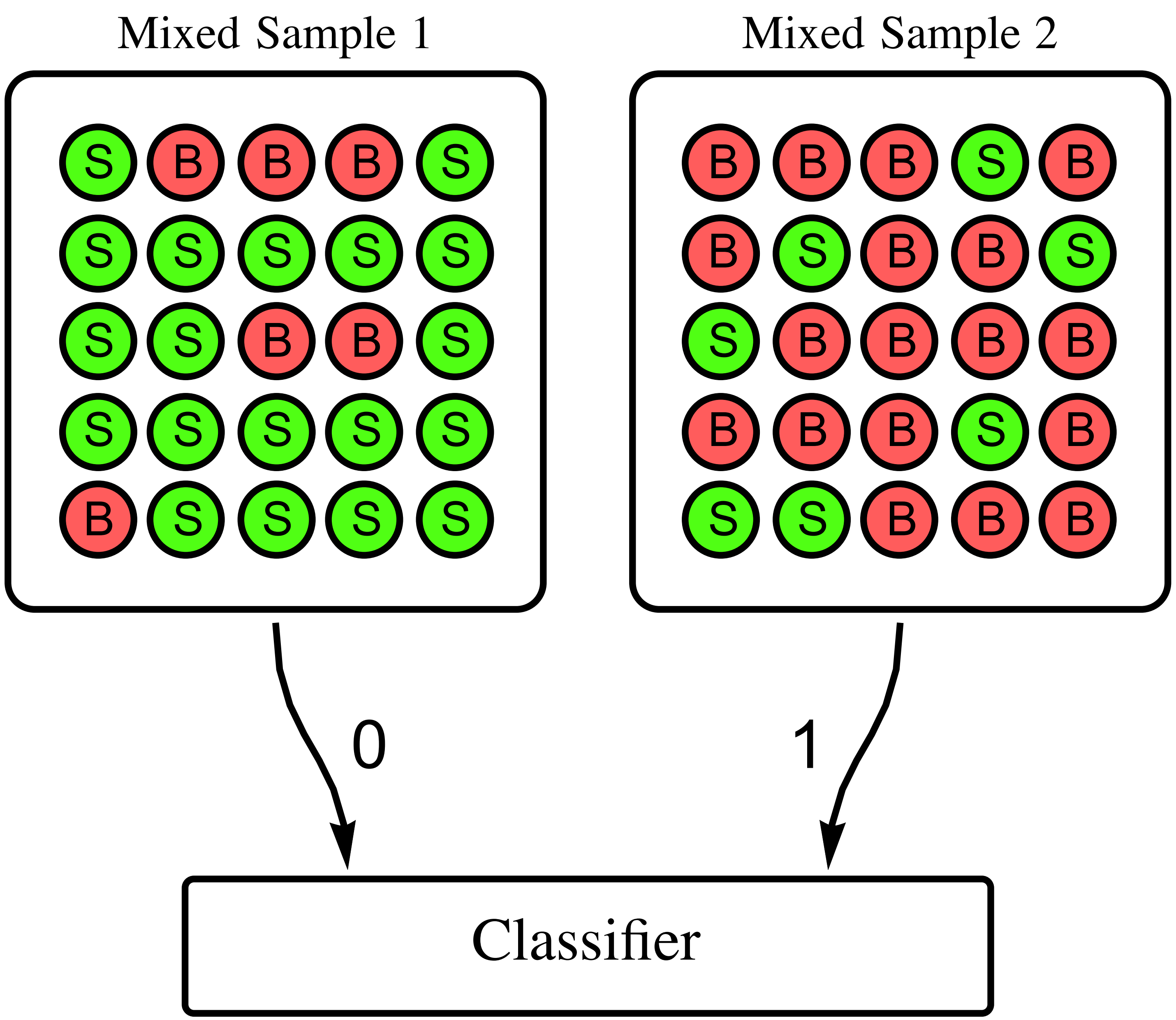
Classification Without Labels



Yes !

[Komiske, Metodiev, BPN, Schwartz, PRD 98 (2018) 011502]
[Cohen, Freytsis, Ostdiek, JHEP 02 (2018) 034]
[Metodiev, BPN, Thaler, JHEP 10 (2017) 51]
[Dery, BPN, Rubbo, Schwartzman, JHEP 05 (2017) 145]

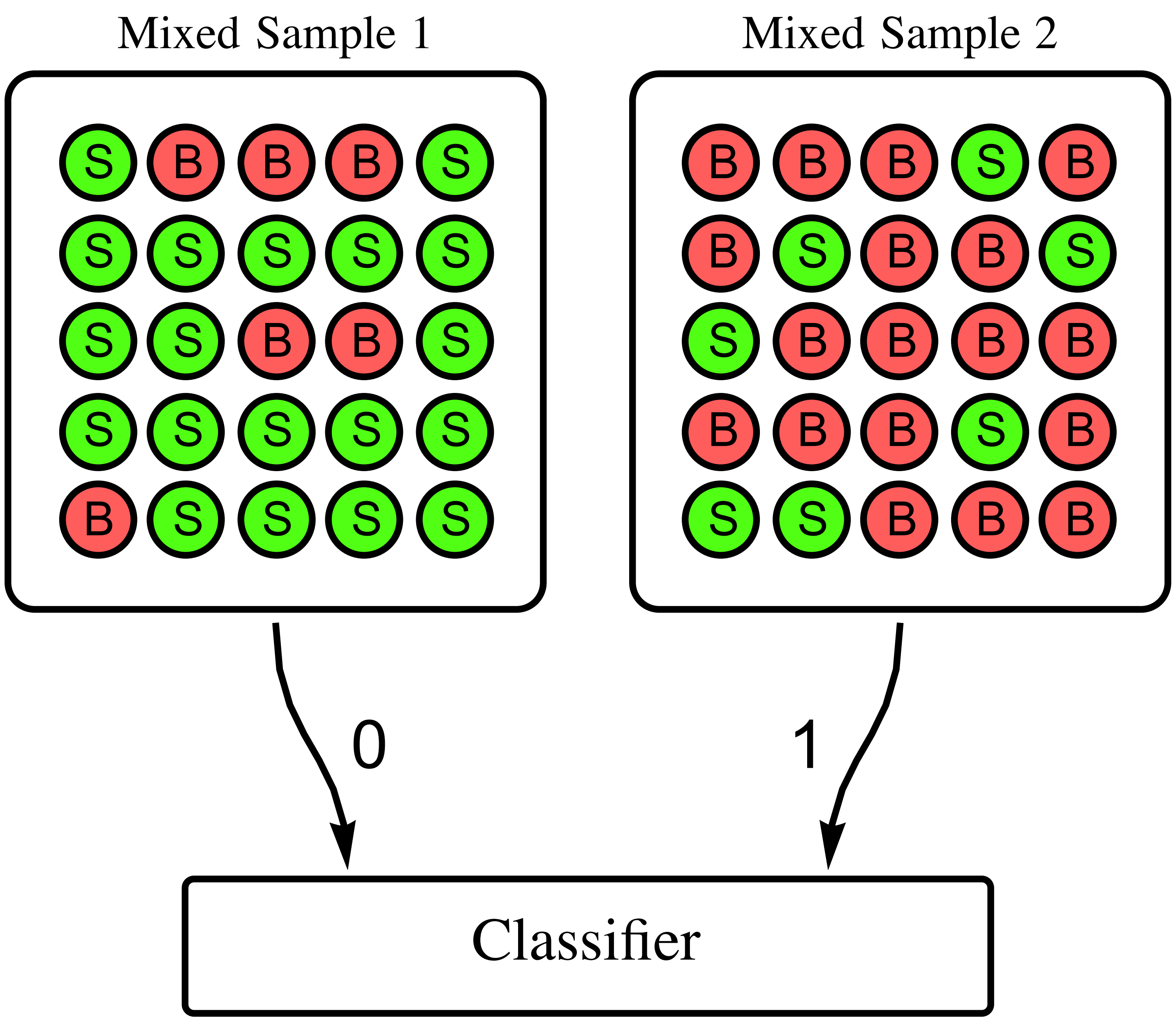
Classification Without Labels



Yes !

[Komiske, Metodiev, BPN, Schwartz, PRD 98 (2018) 011502]
[Cohen, Freytsis, Ostdiek, JHEP 02 (2018) 034]
[Metodiev, BPN, Thaler, JHEP 10 (2017) 51]
[Dery, BPN, Rubbo, Schwartzman, JHEP 05 (2017) 145]

Classification Without Labels



Yes !

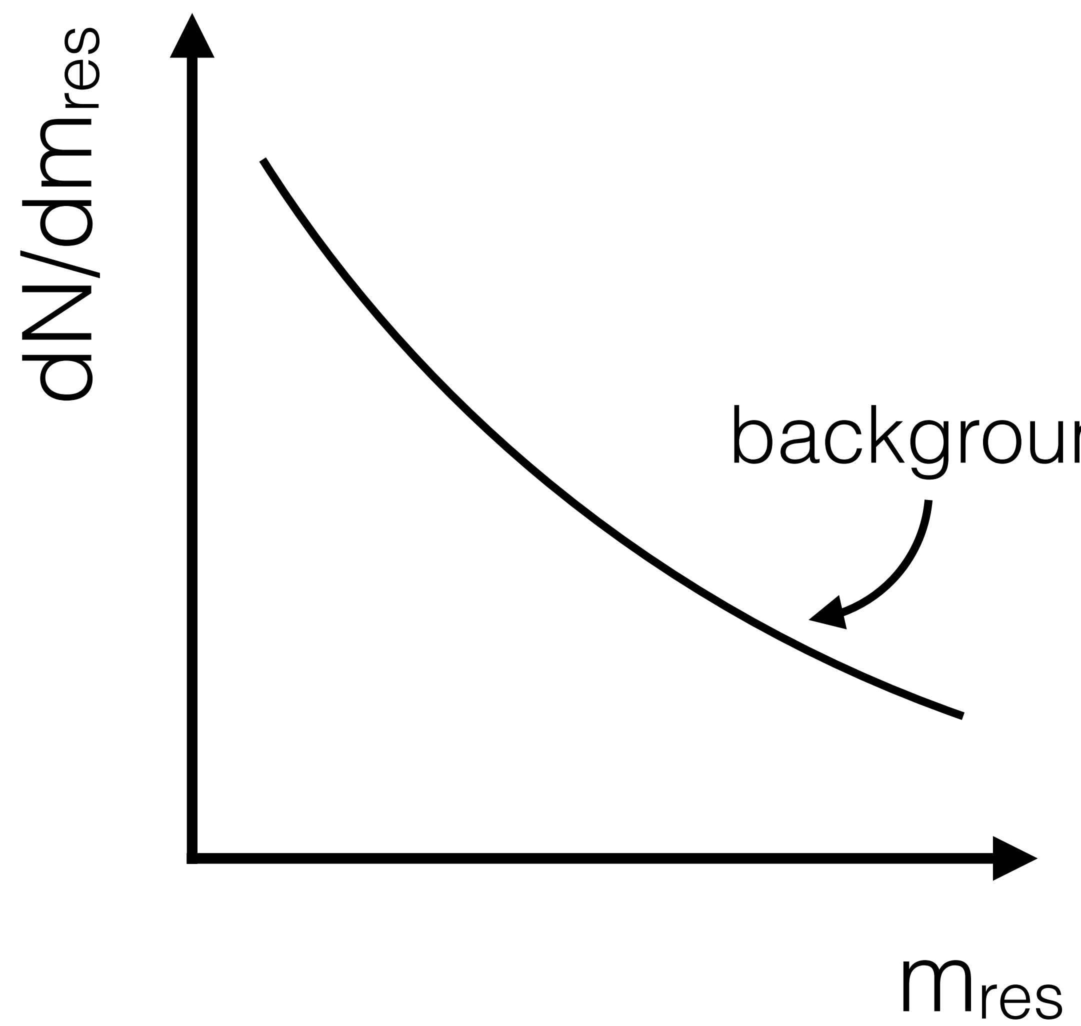
One can show that this procedure asymptotically converges to the optimal classifier (with labels).

[Komiske, Metodiev, BPN, Schwartz, PRD 98 (2018) 011502]
[Cohen, Freytsis, Ostdiek, JHEP 02 (2018) 034]
[Metodiev, BPN, Thaler, JHEP 10 (2017) 51]
[Dery, BPN, Rubbo, Schwartzman, JHEP 05 (2017) 145]

CWoLa for anomaly detection

J. Collins, K. Howe, BPN
PRL 121 (2018) 241803

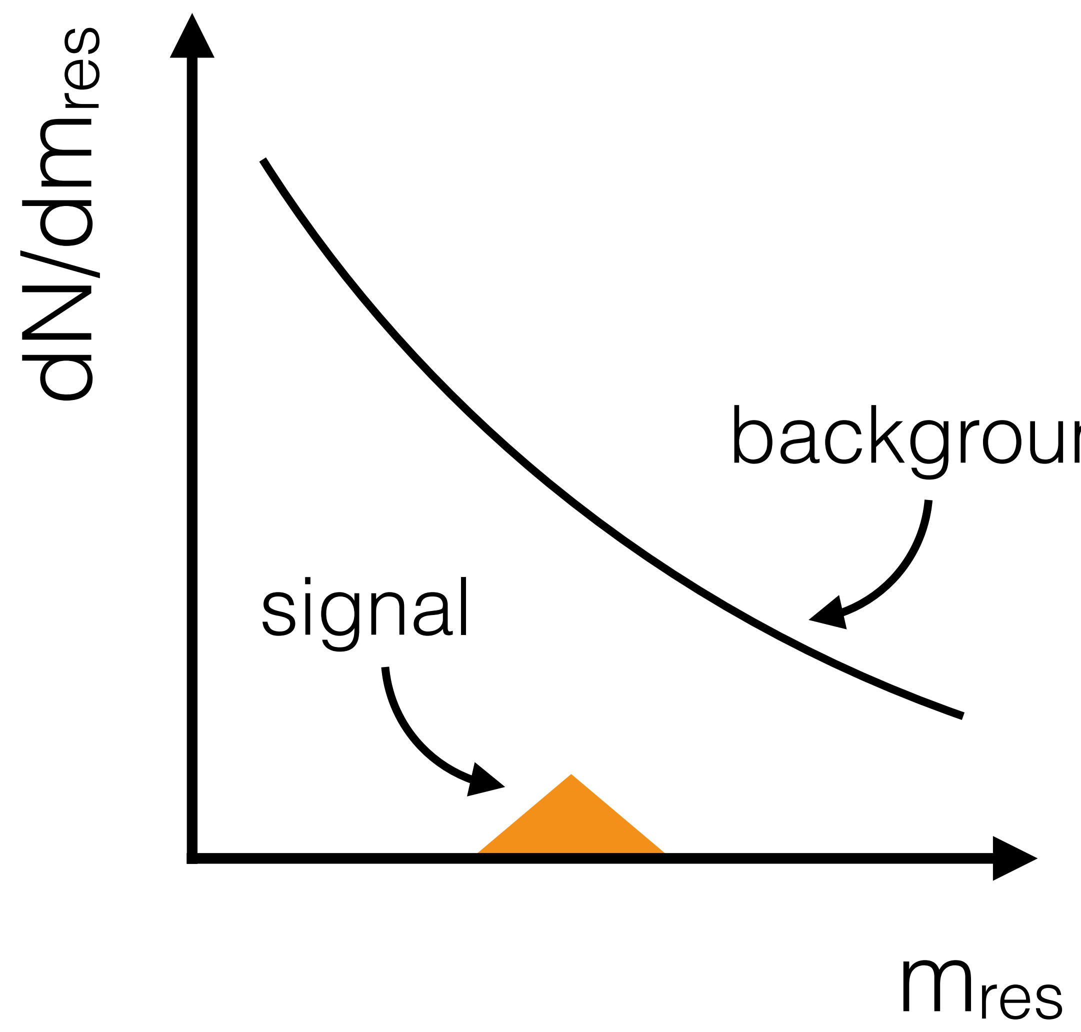
J. Collins, K. Howe, BPN
PRD 99 (2019) 014038



CWoLa for anomaly detection

J. Collins, K. Howe, BPN
PRL 121 (2018) 241803

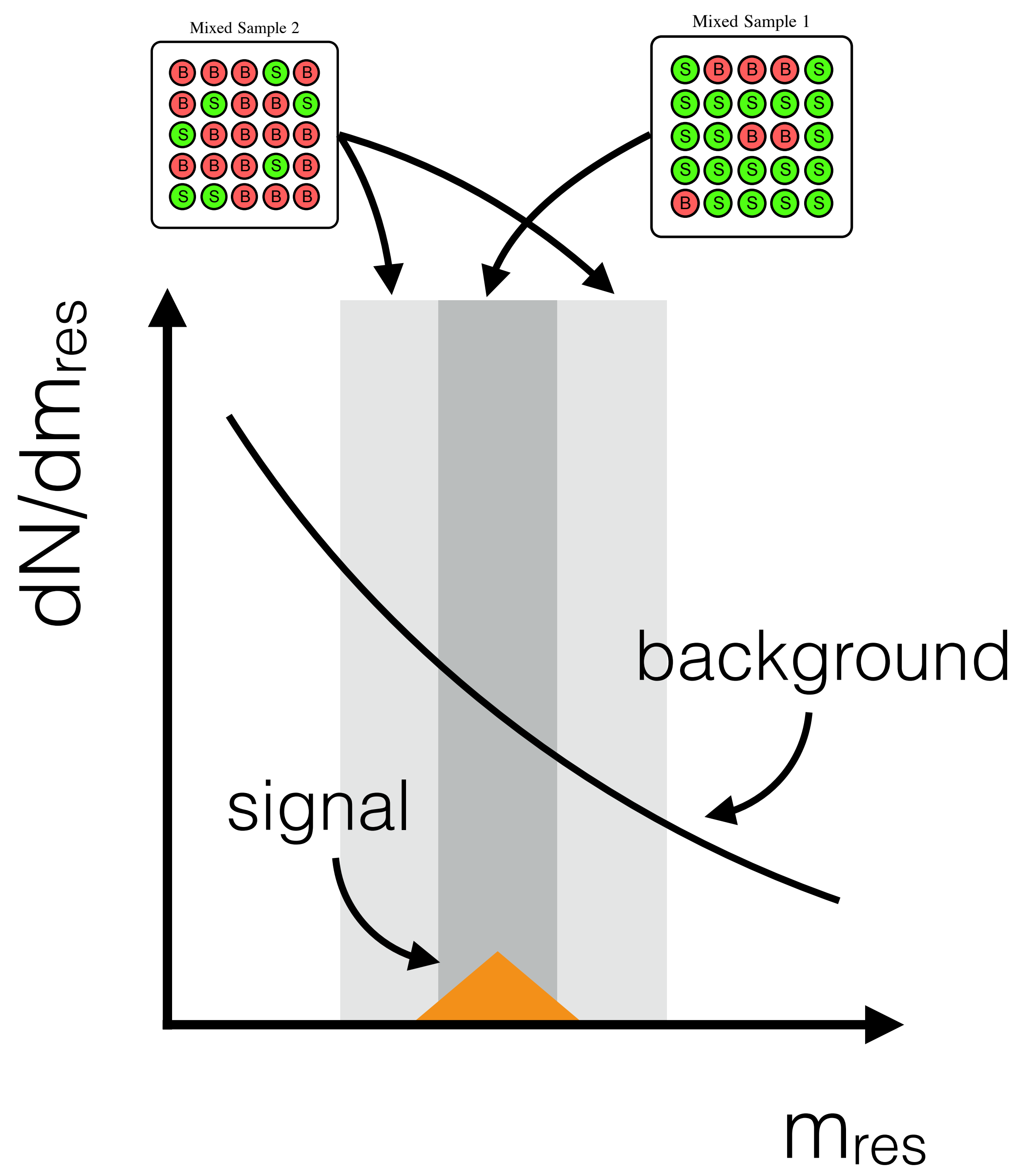
J. Collins, K. Howe, BPN
PRD 99 (2019) 014038



CWoLa for anomaly detection

J. Collins, K. Howe, BPN
PRL 121 (2018) 241803

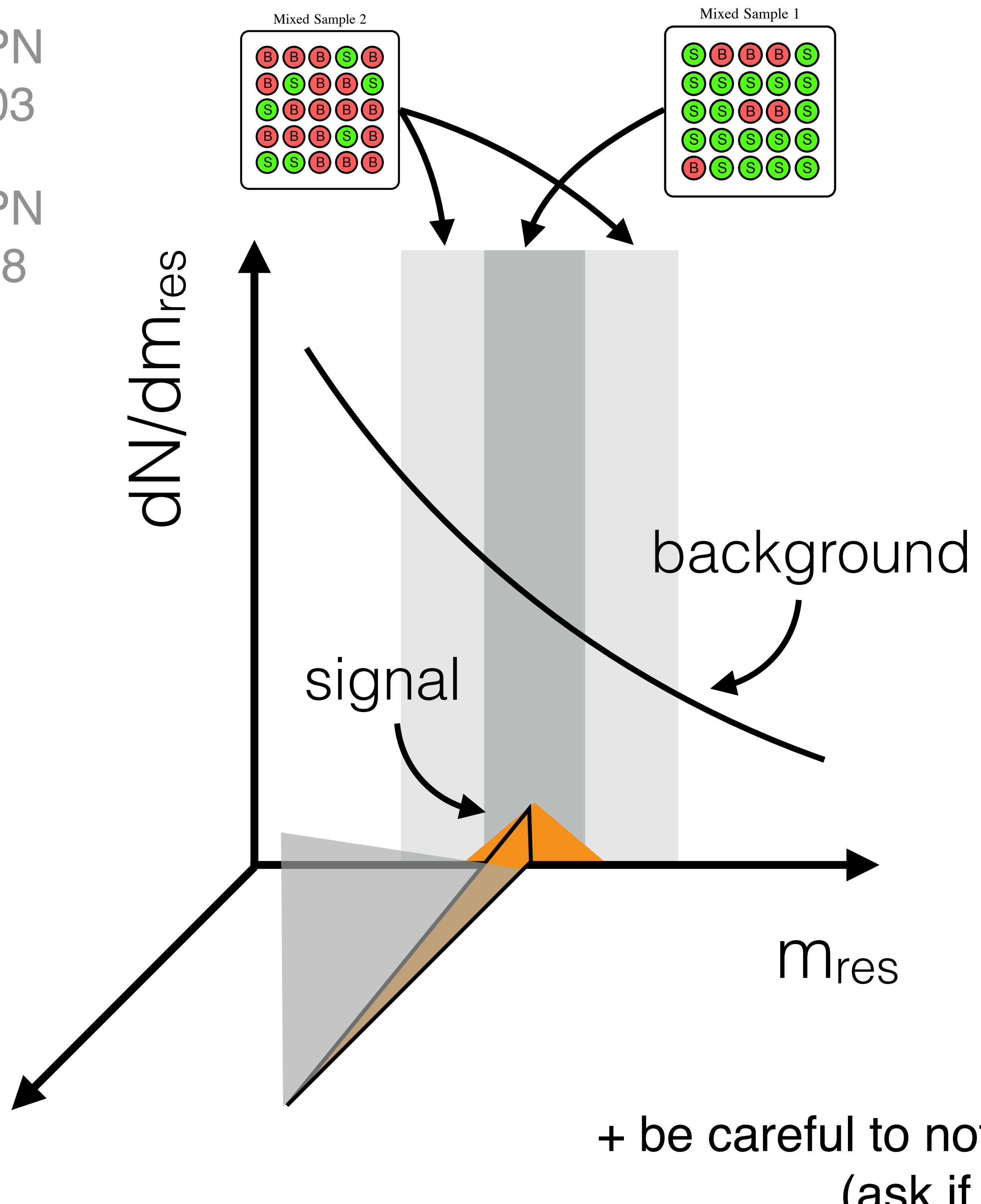
J. Collins, K. Howe, BPN
PRD 99 (2019) 014038



CWoLa for anomaly detection

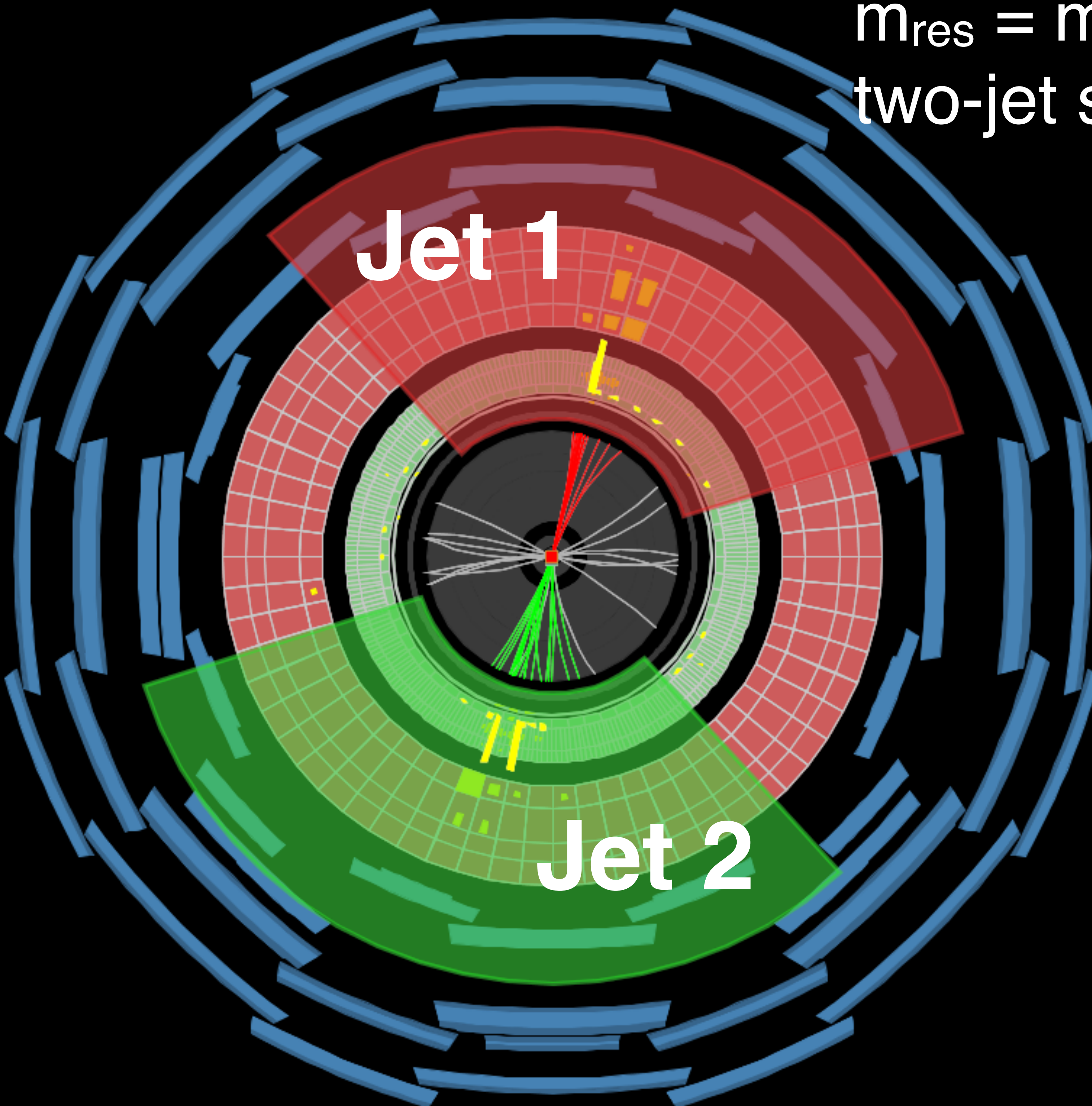
J. Collins, K. Howe, BPN
PRL 121 (2018) 241803

J. Collins, K. Howe, BPN
PRD 99 (2019) 014038



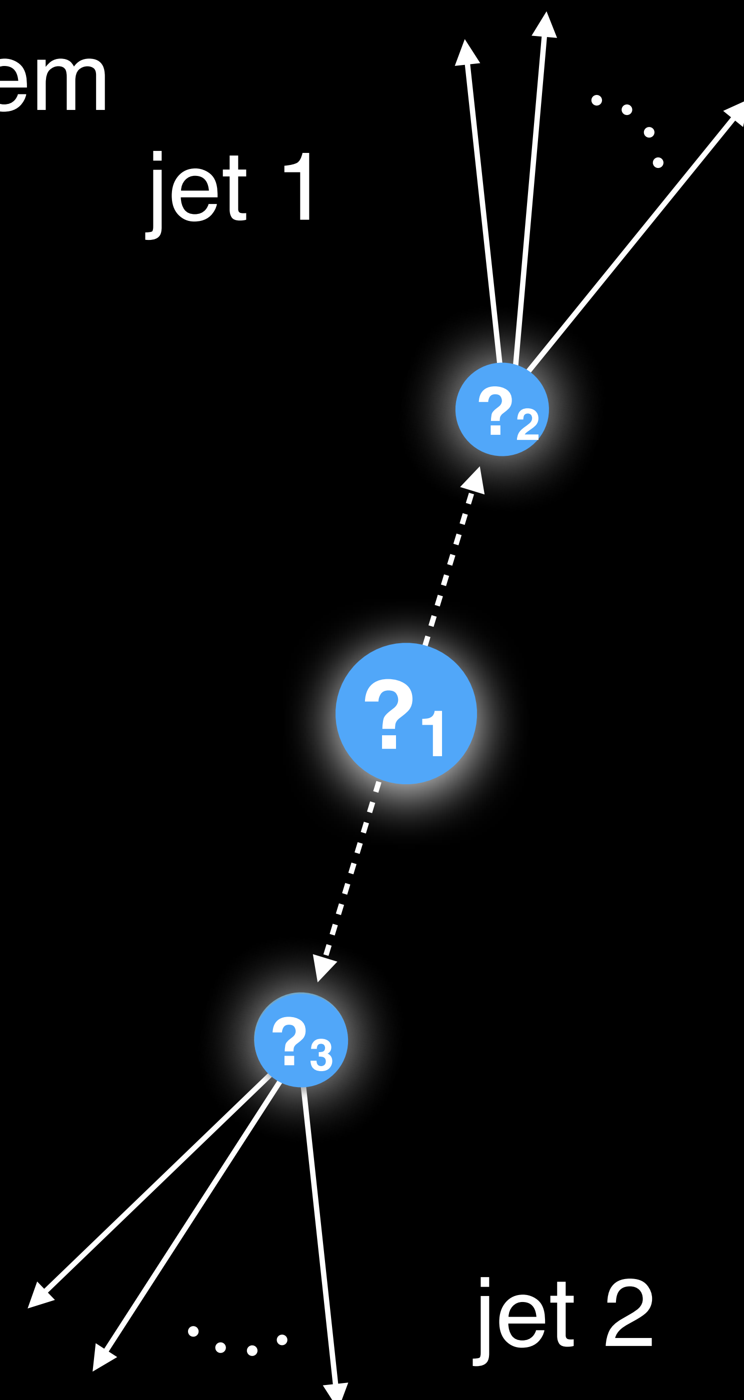
+ be careful to not pay a big trials factor
(ask if interested)

Example: two-jet search



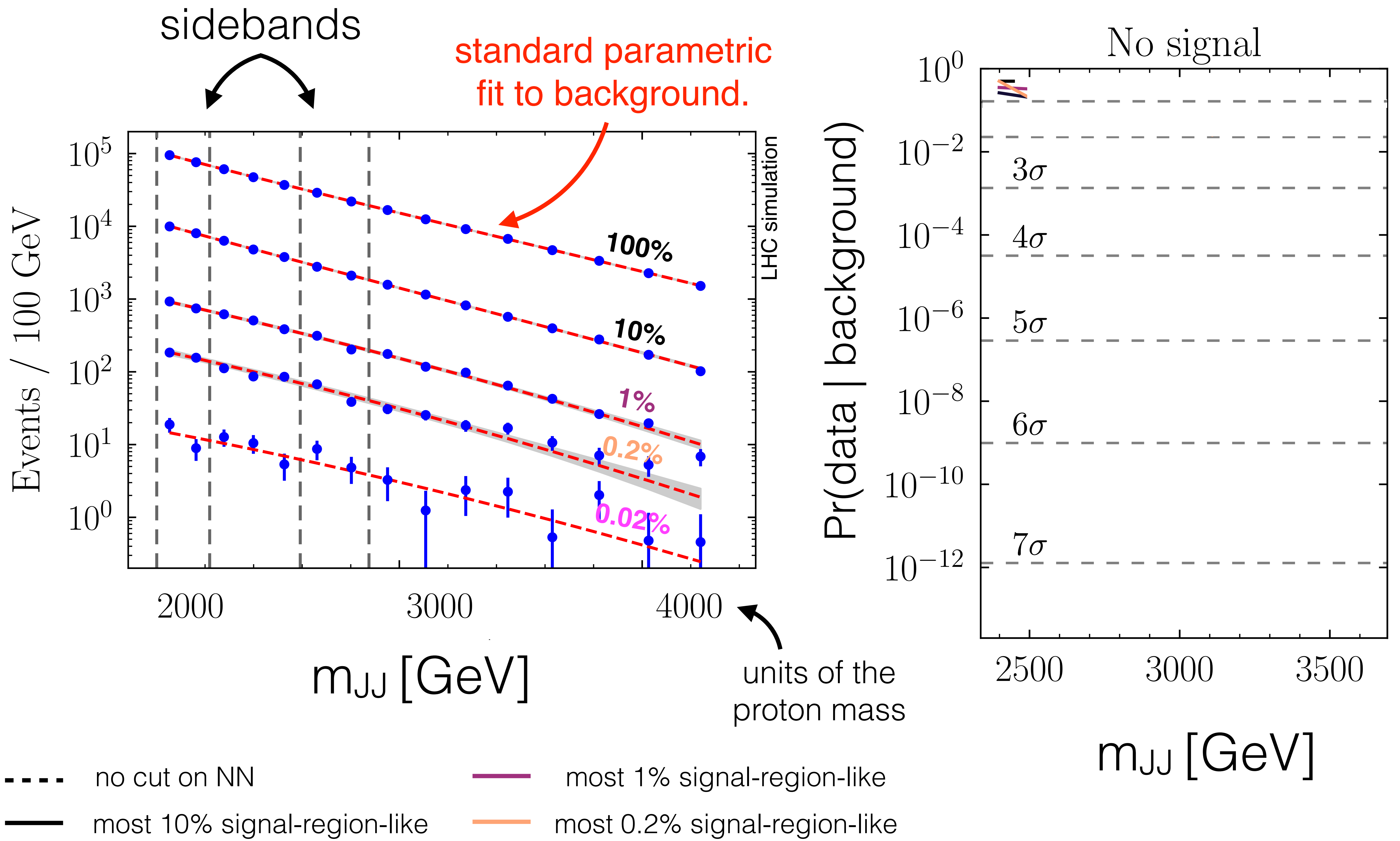
collisions in/out of page

m_{res} = mass of two-jet system

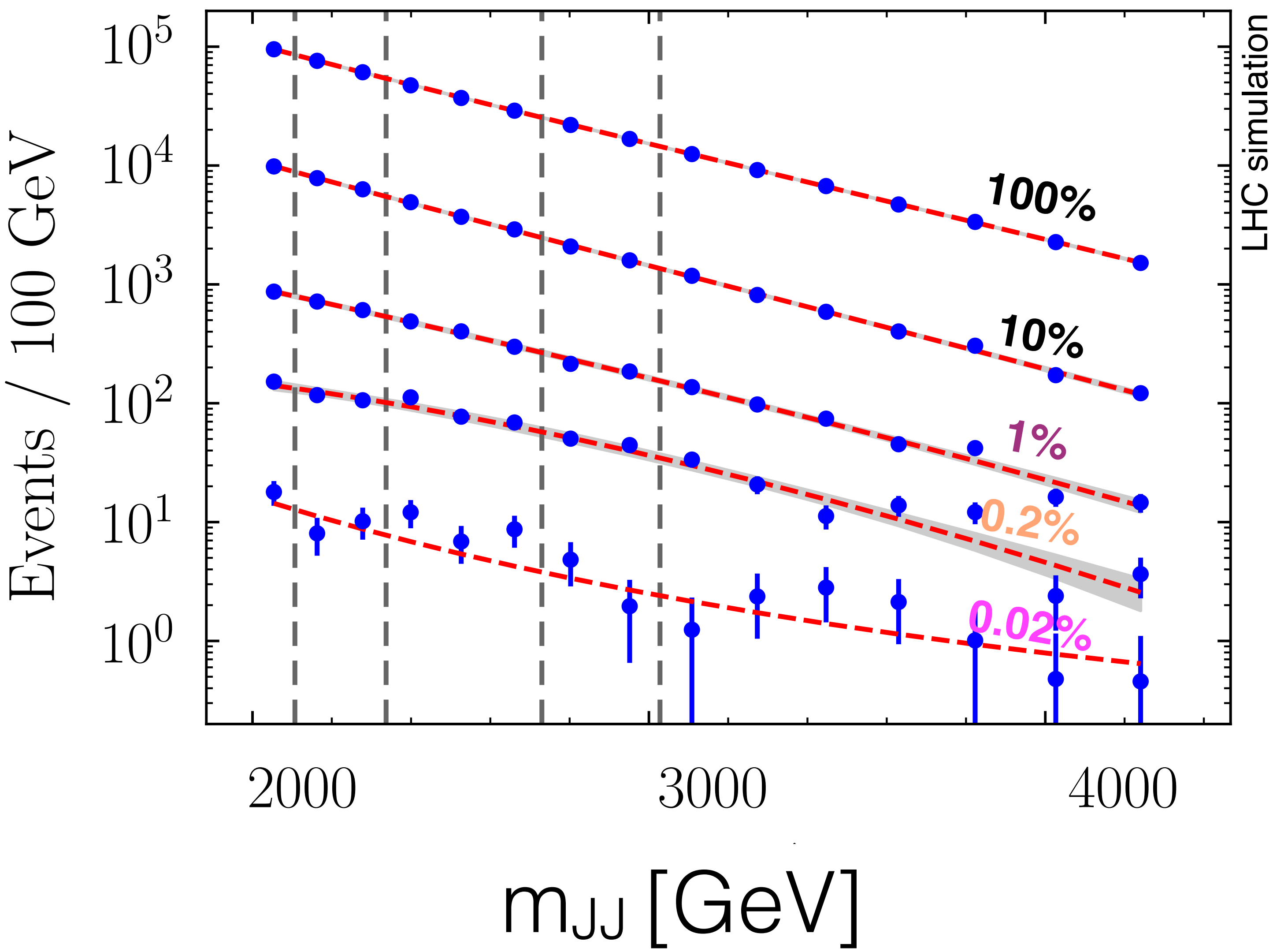


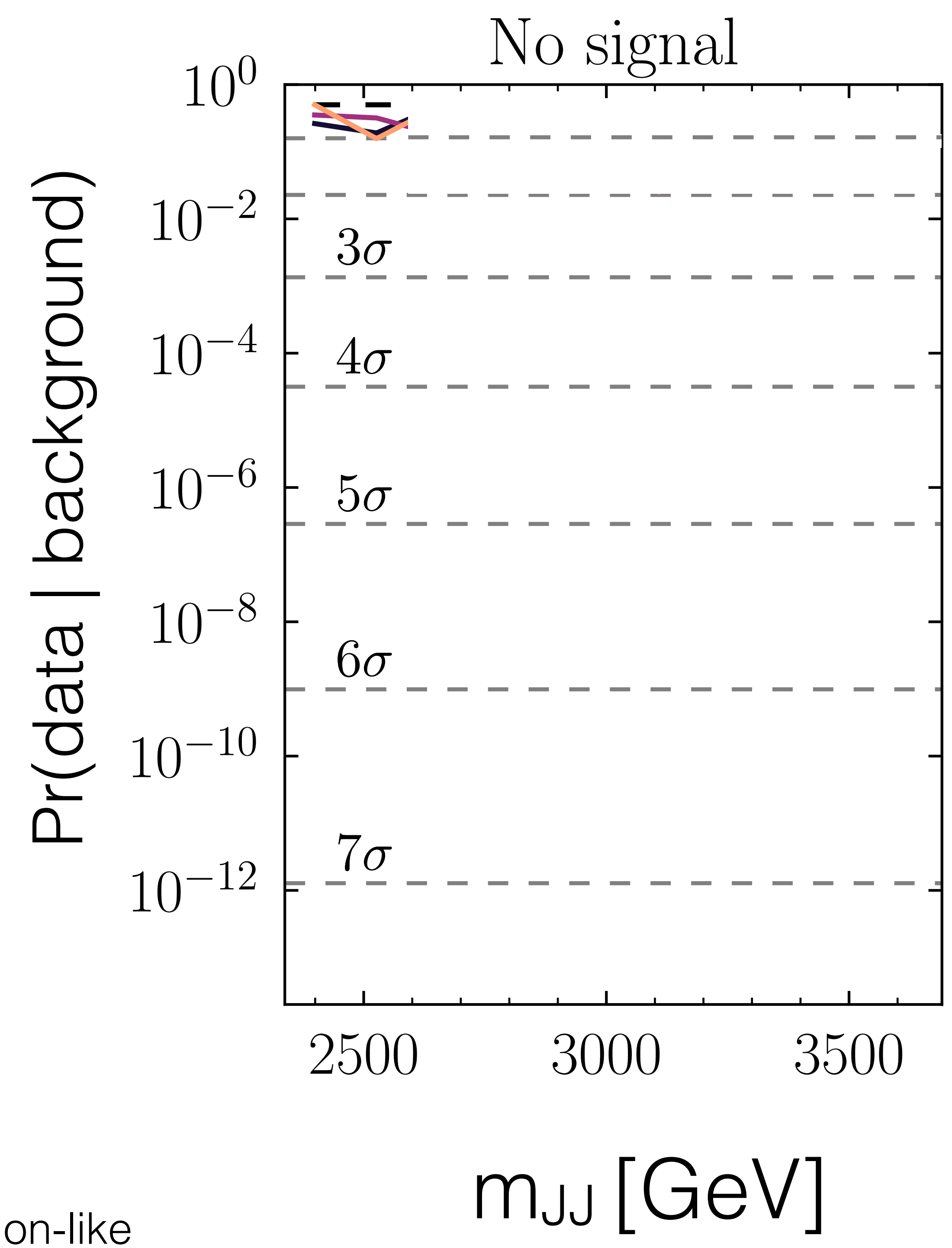
y = many features of the two jets

Example: two-jet search



Example: two-jet search





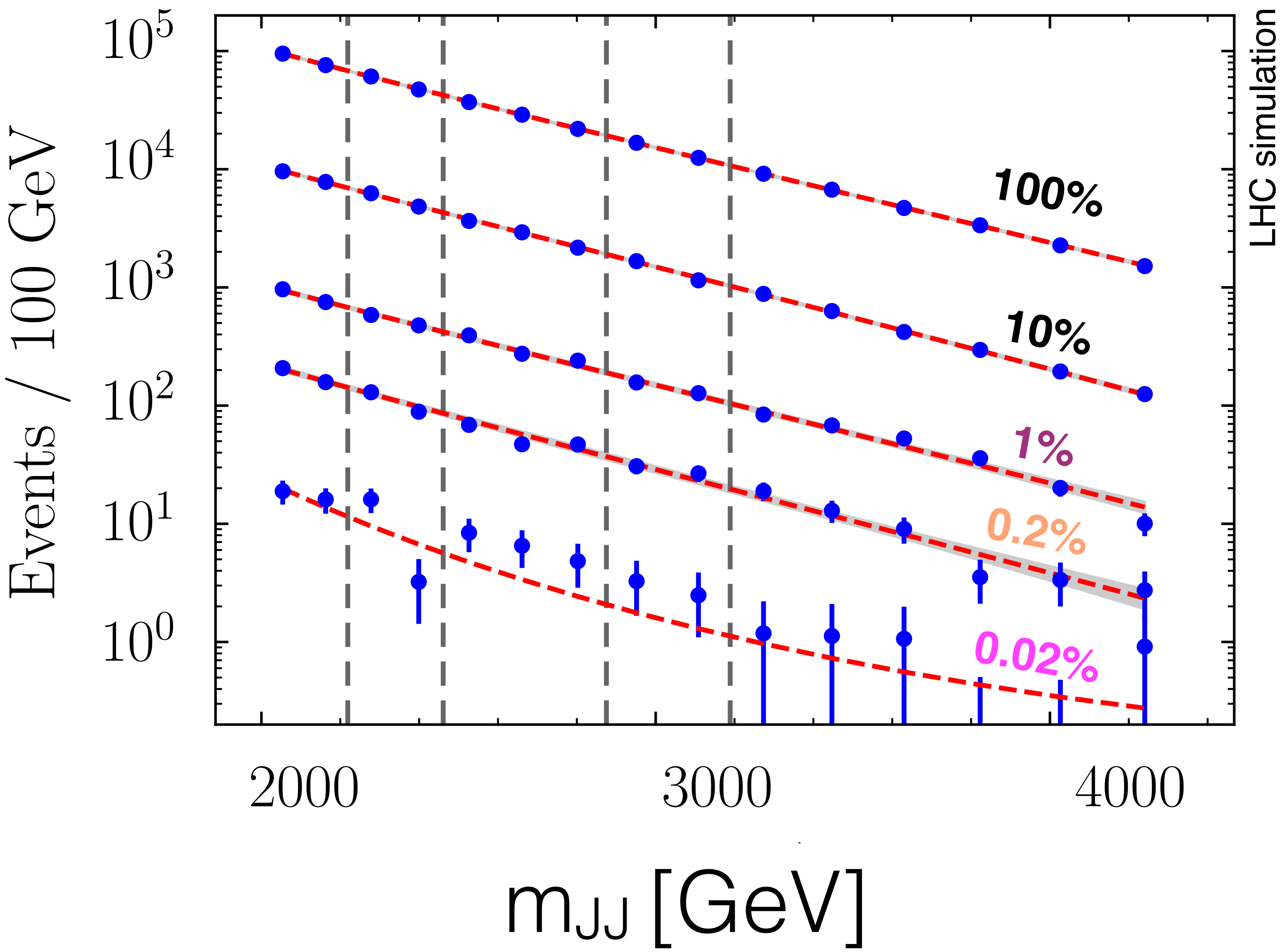
no cut on NN

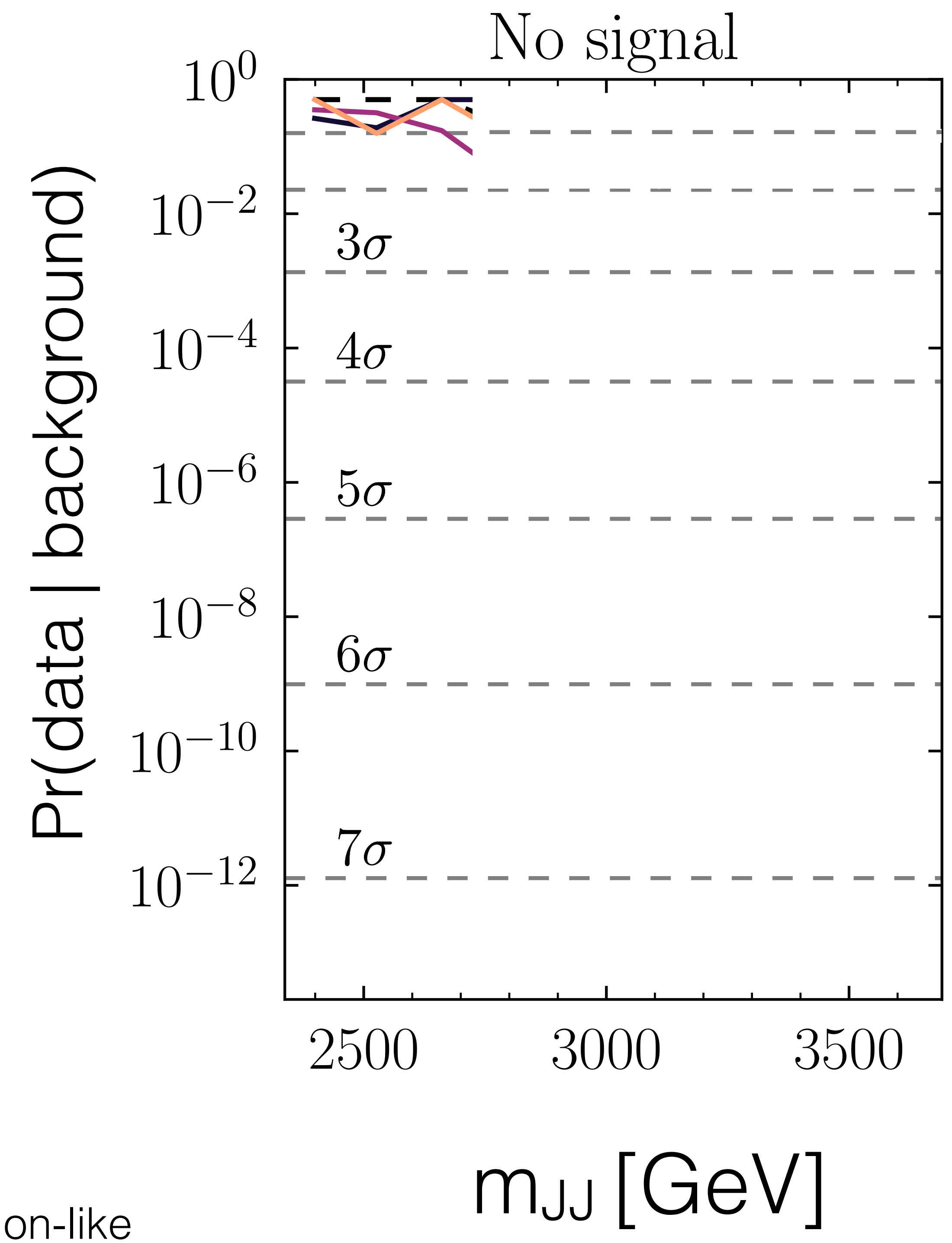
most 10% signal-region-like

most 1% signal-region-like

most 0.2% signal-region-like

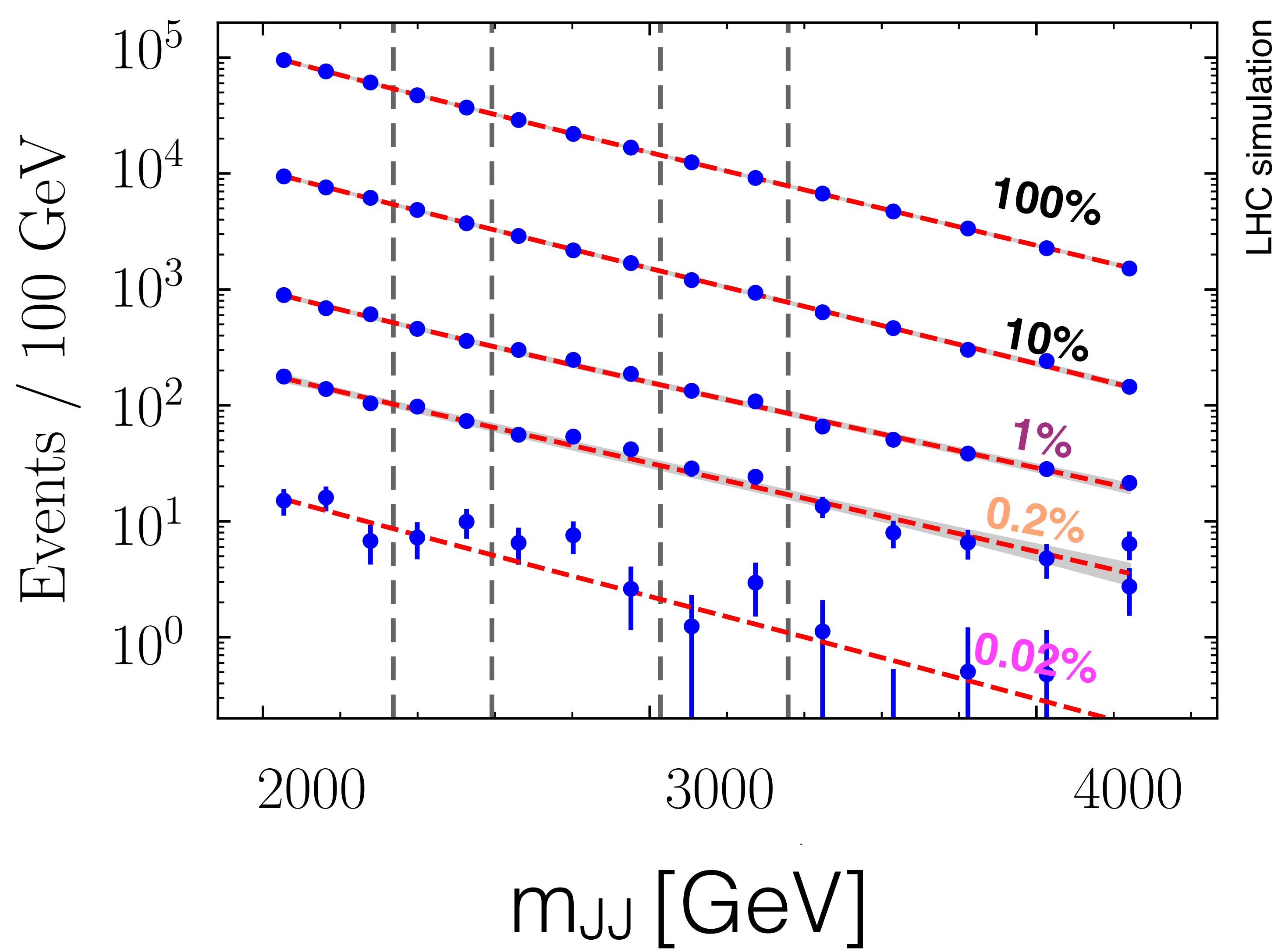
Example: two-jet search

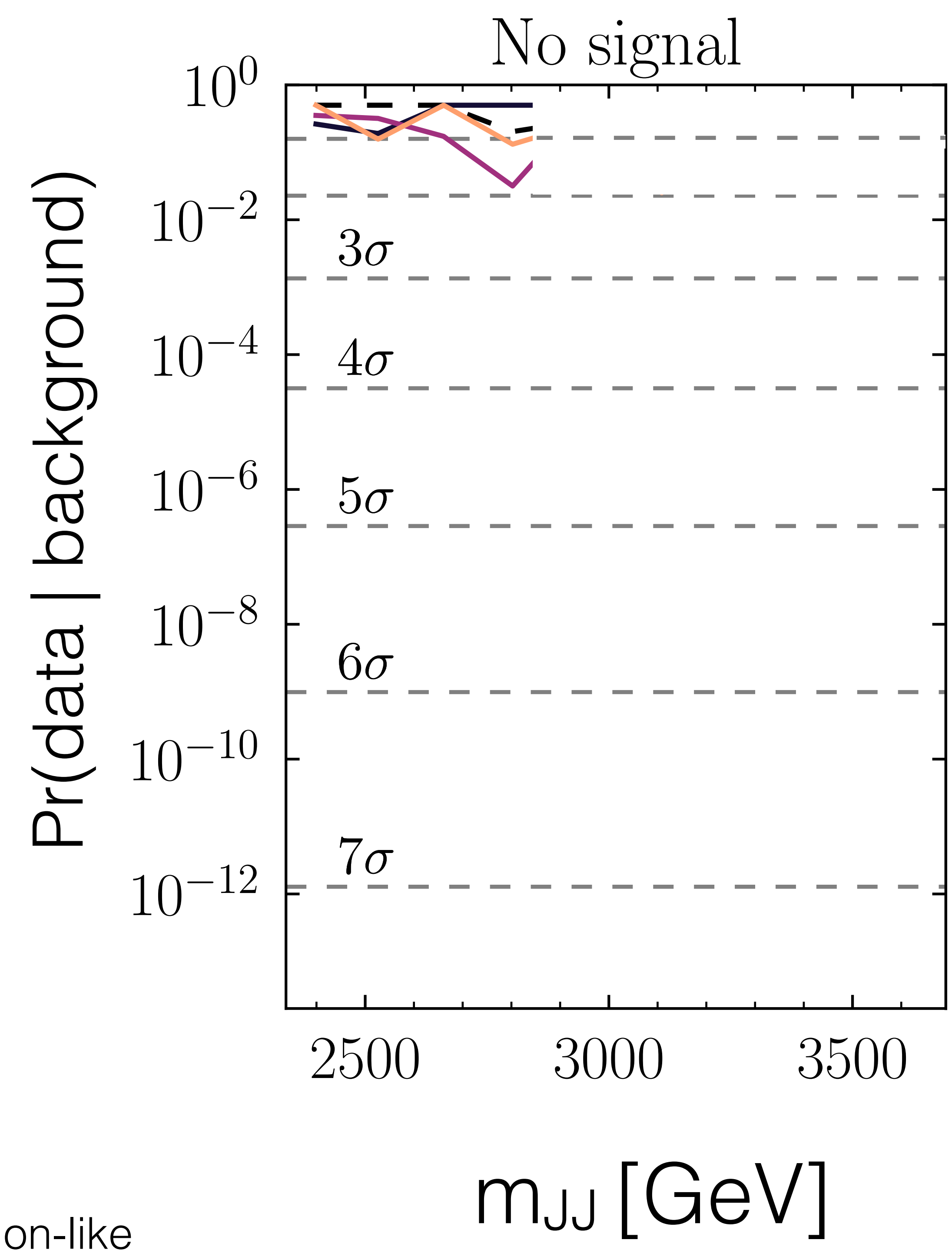




no cut on NN
most 10% signal-region-like
most 1% signal-region-like
most 0.2% signal-region-like

Example: two-jet search





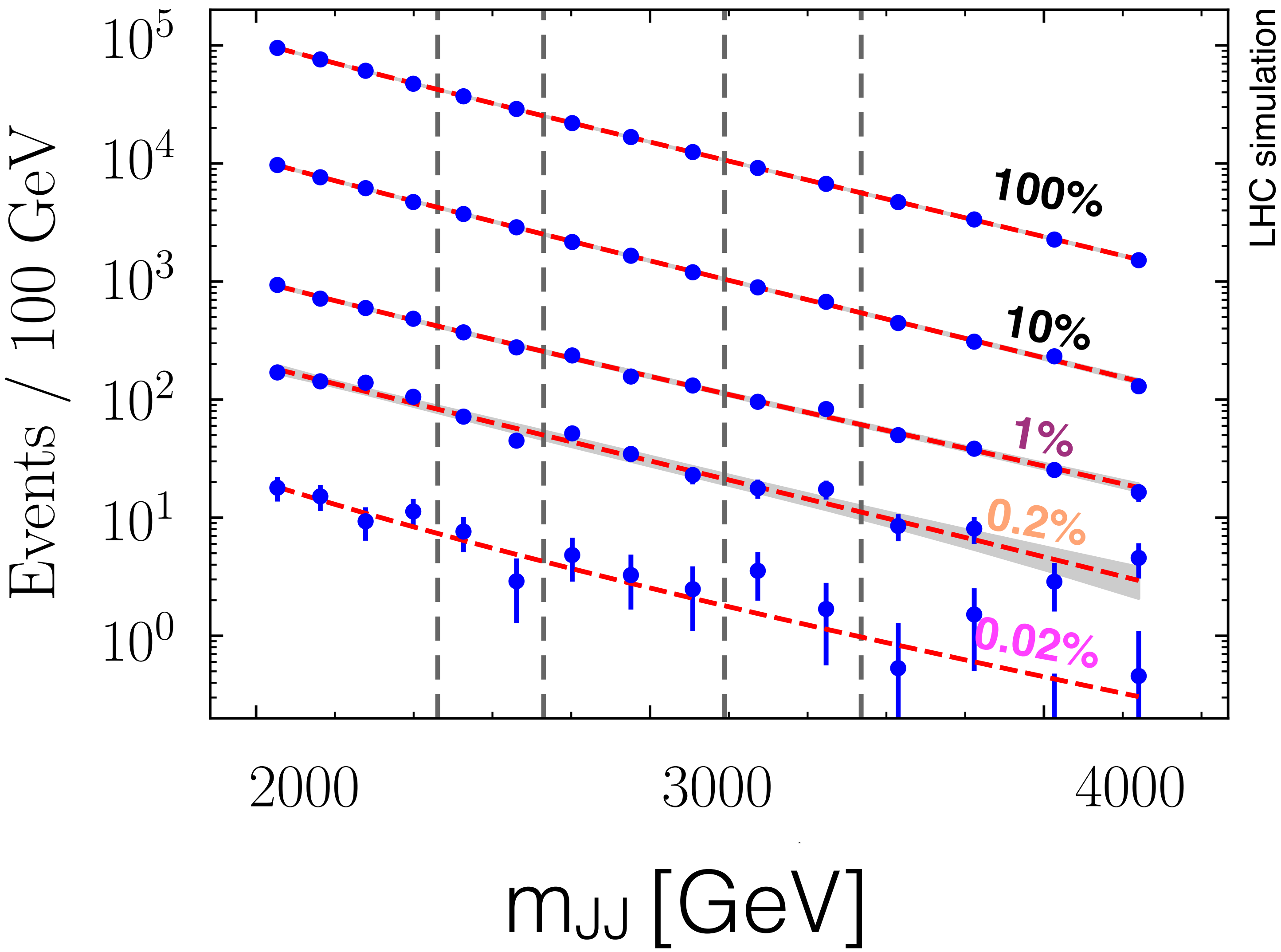
no cut on NN

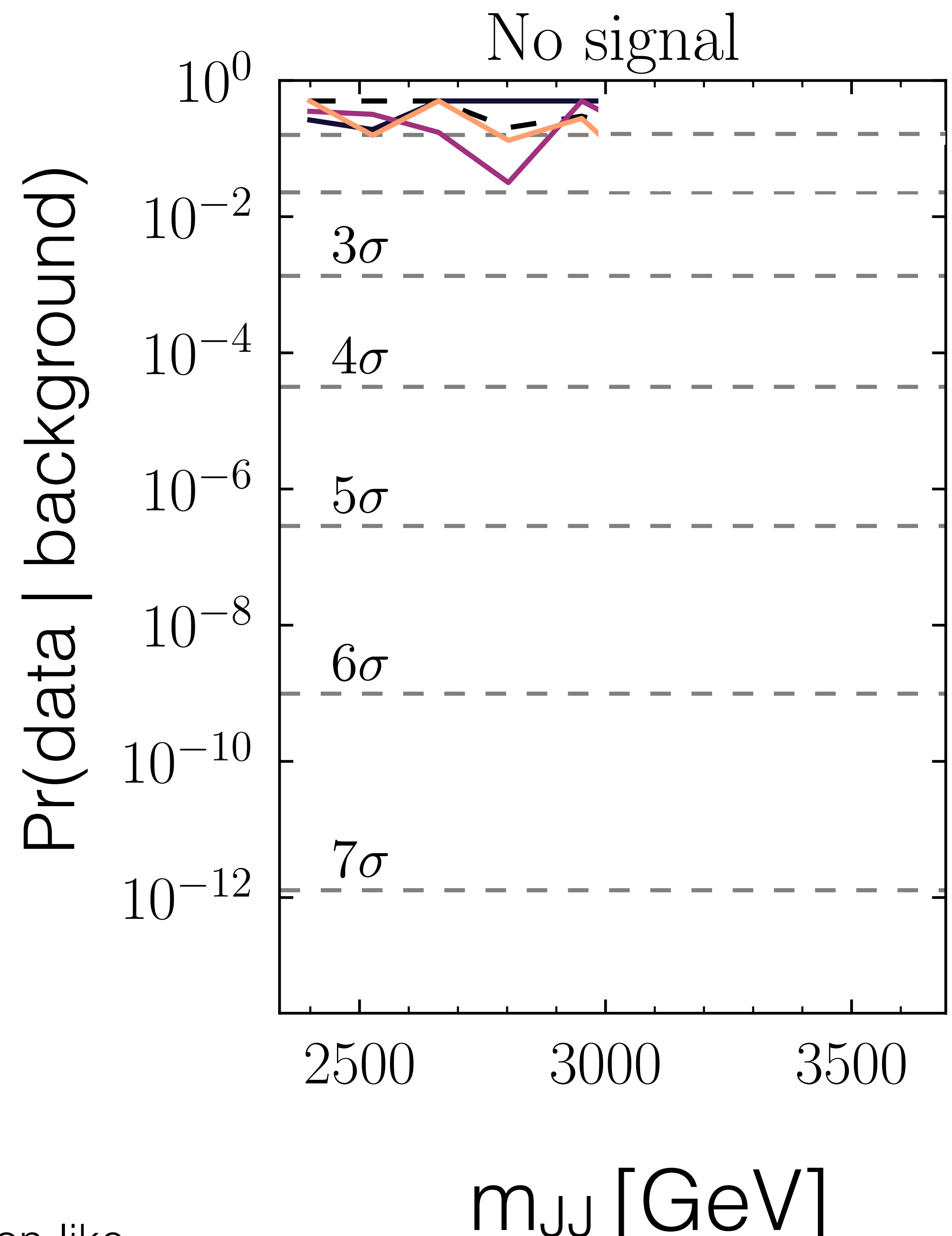
most 10% signal-region-like

most 1% signal-region-like

most 0.2% signal-region-like

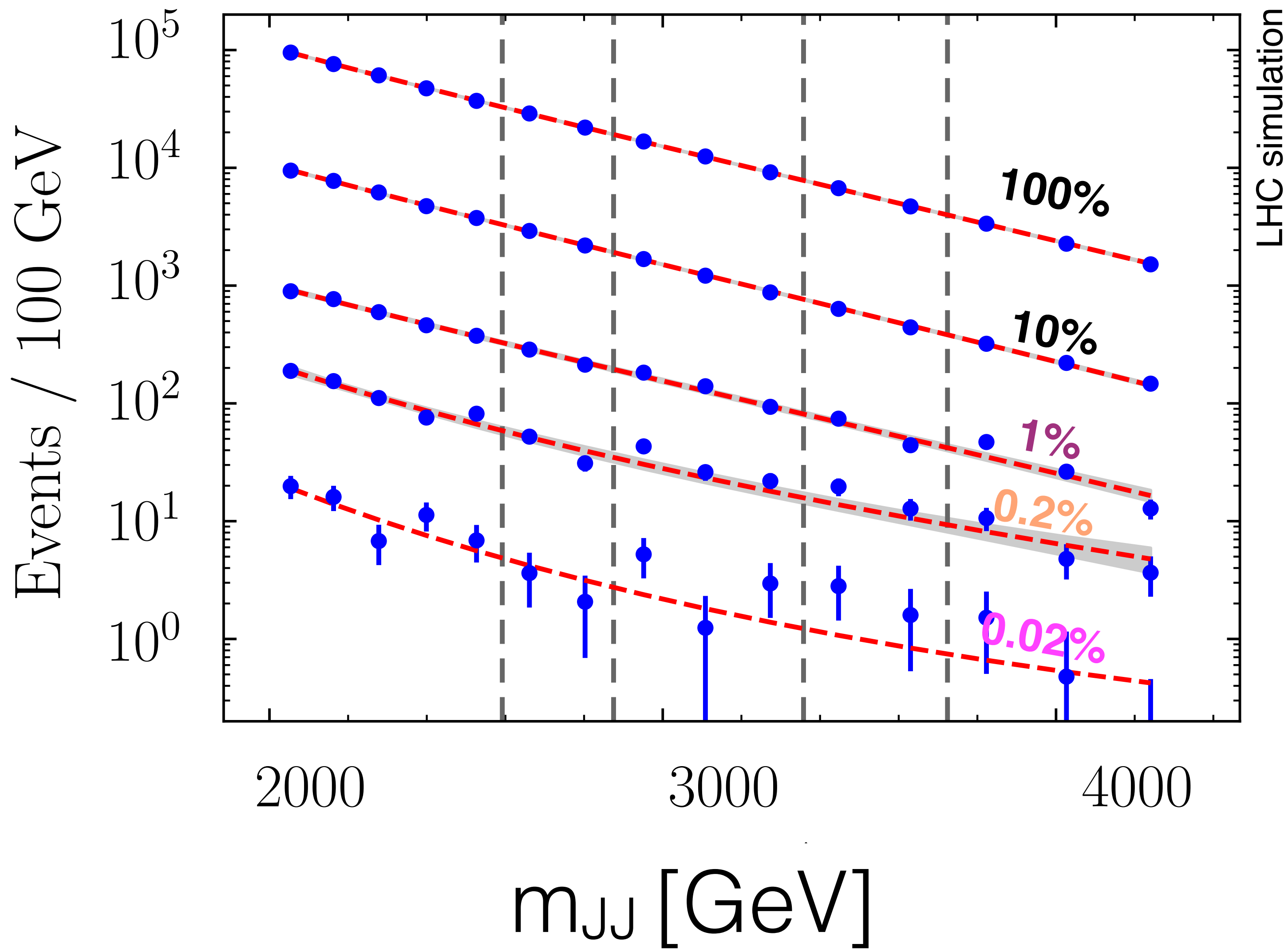
Example: two-jet search

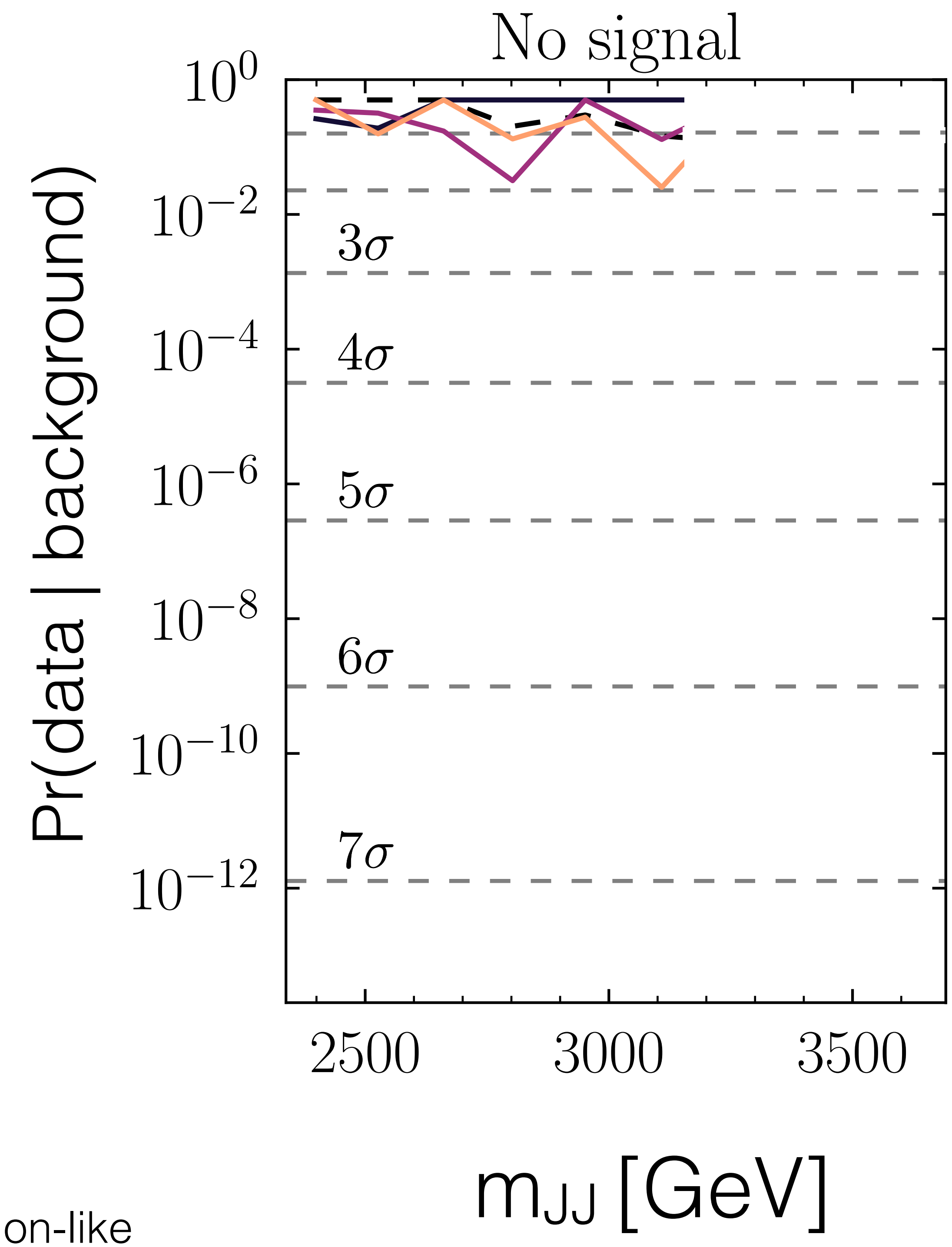




no cut on NN
most 10% signal-region-like
most 1% signal-region-like
most 0.2% signal-region-like

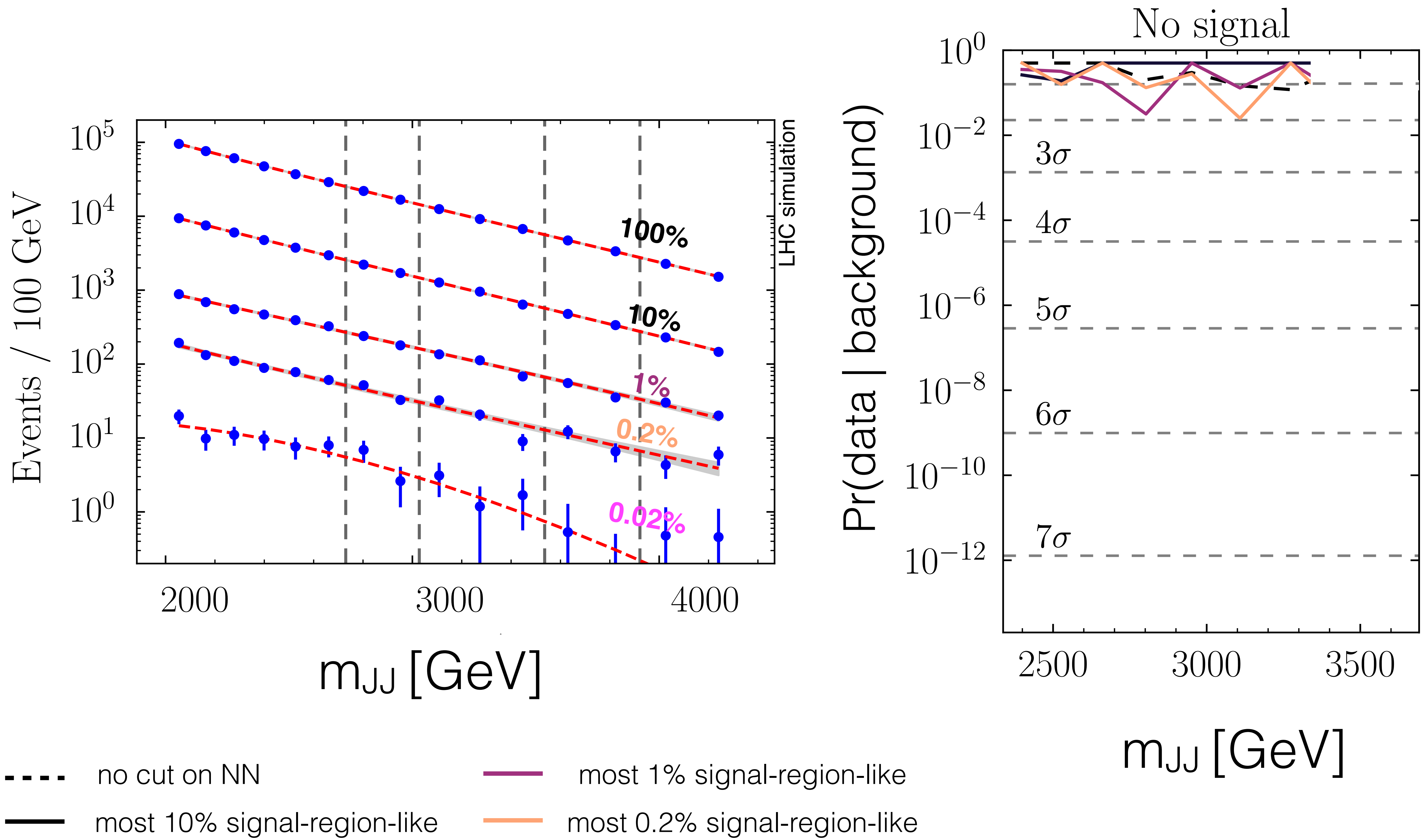
Example: two-jet search



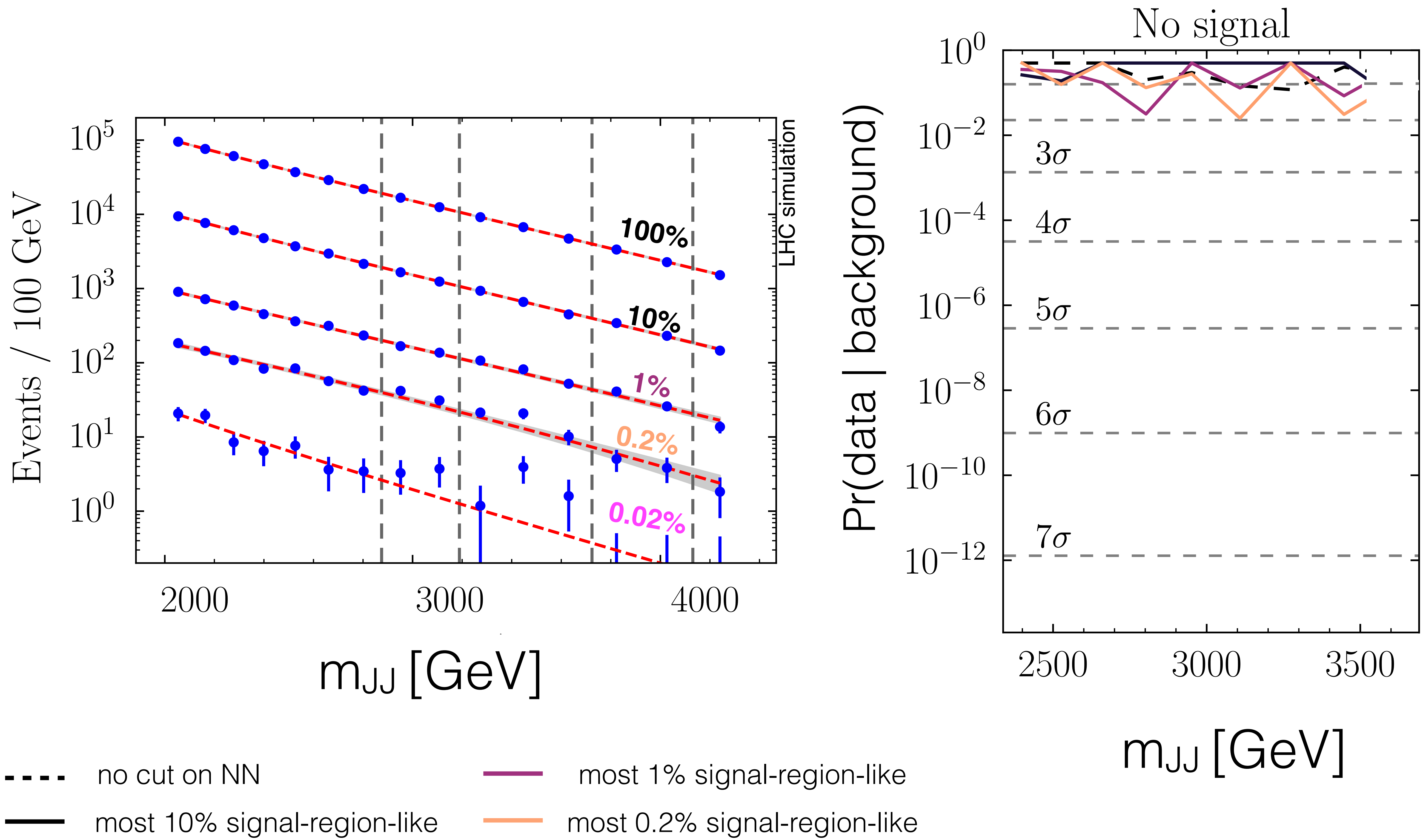


no cut on NN
most 10% signal-region-like
most 1% signal-region-like
most 0.2% signal-region-like

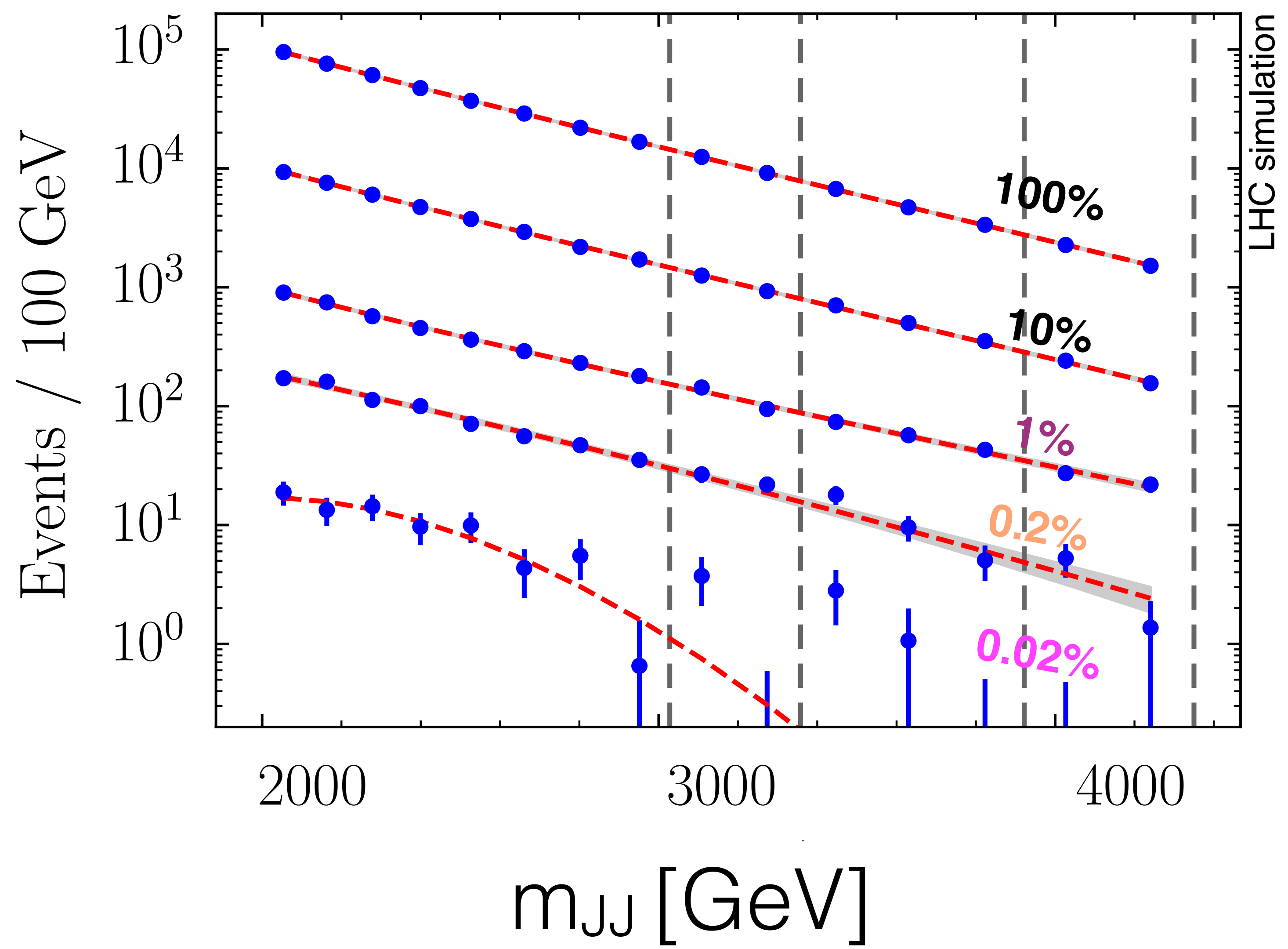
Example: two-jet search

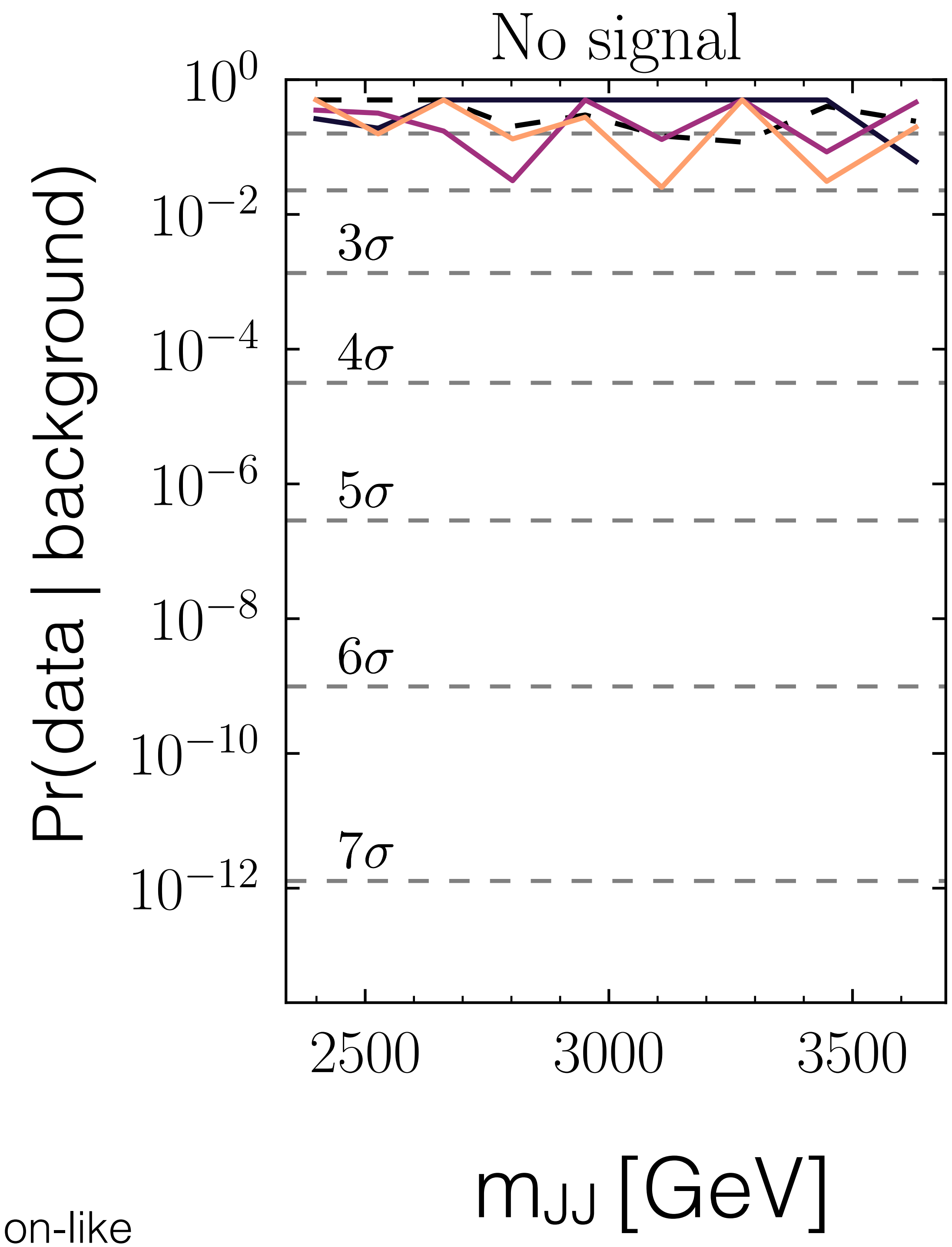


Example: two-jet search



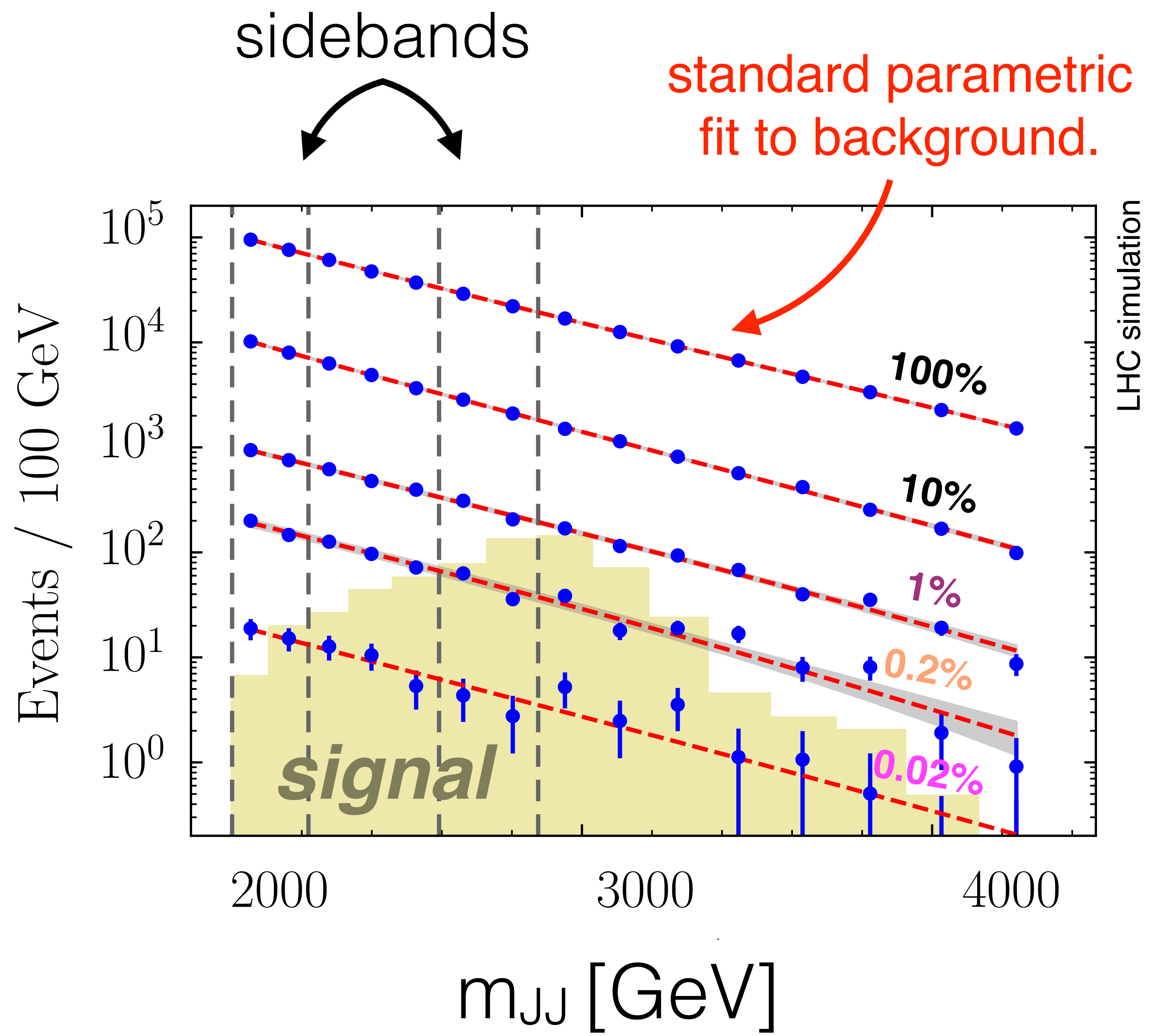
Example: two-jet search

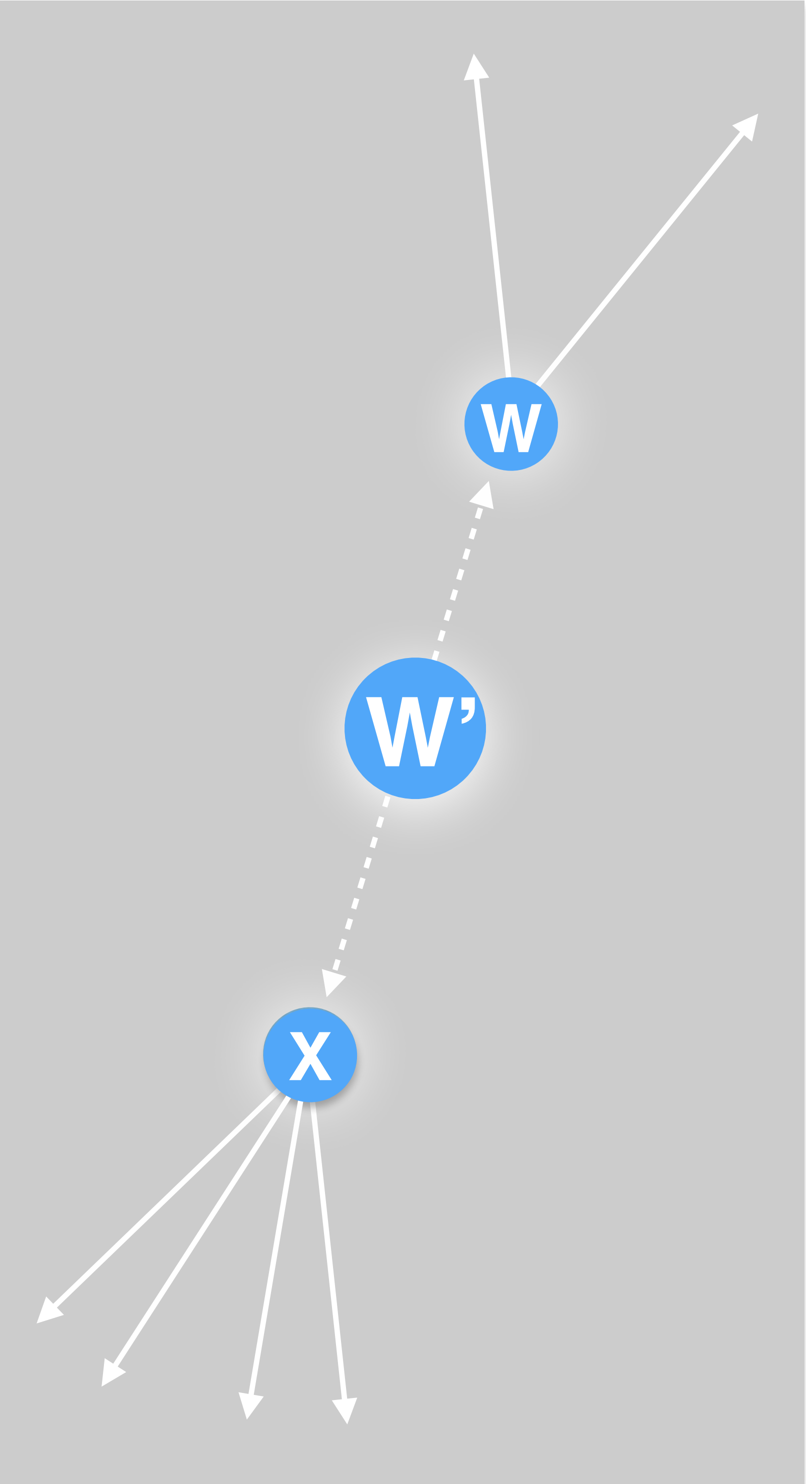




no cut on NN
most 10% signal-region-like
most 1% signal-region-like
most 0.2% signal-region-like

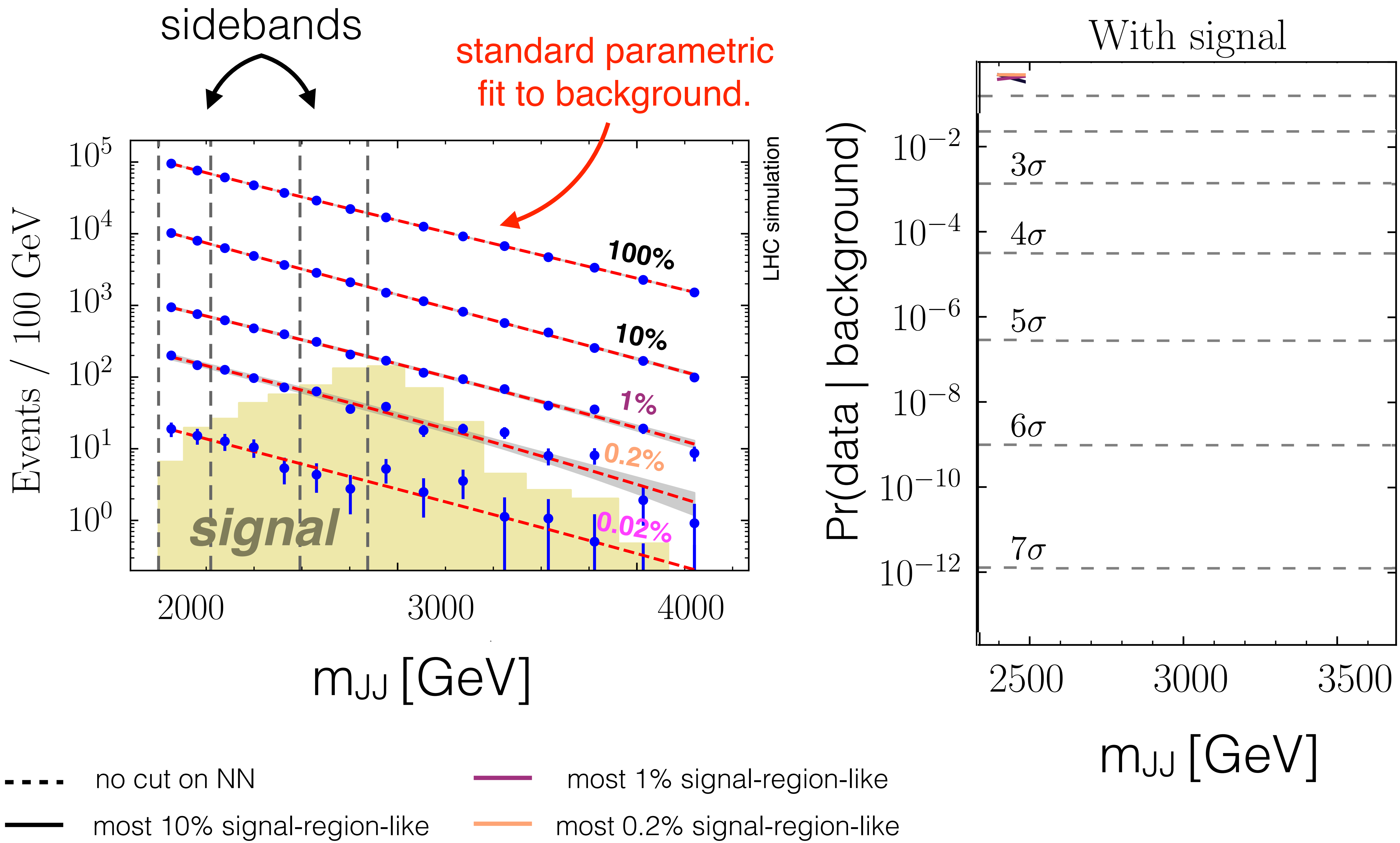
…and when there is a signal?



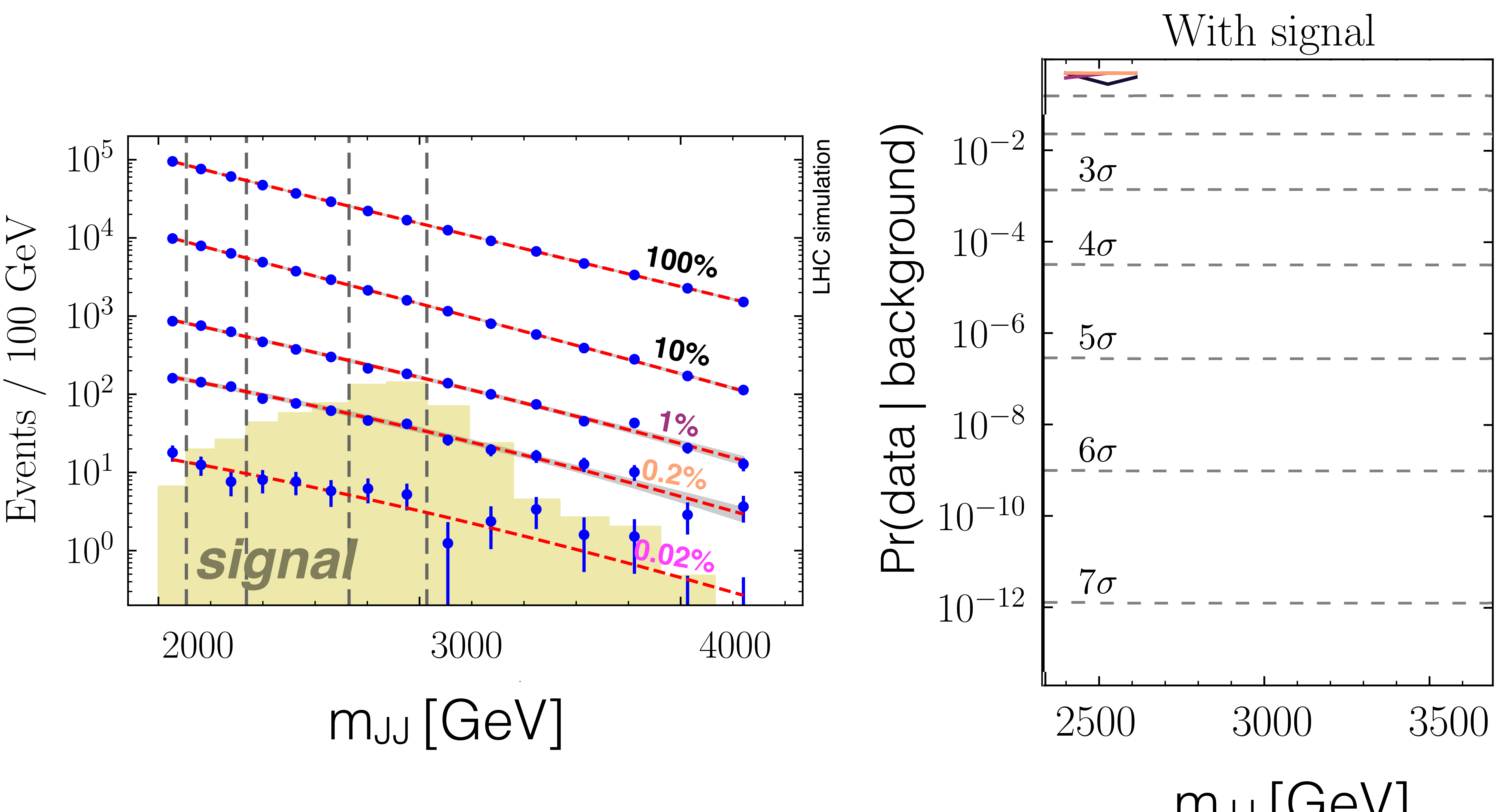


no cut on NN
most 10% signal-region-like
most 1% signal-region-like
most 0.2% signal-region-like

…and when there is a signal?



…and when there is a signal?



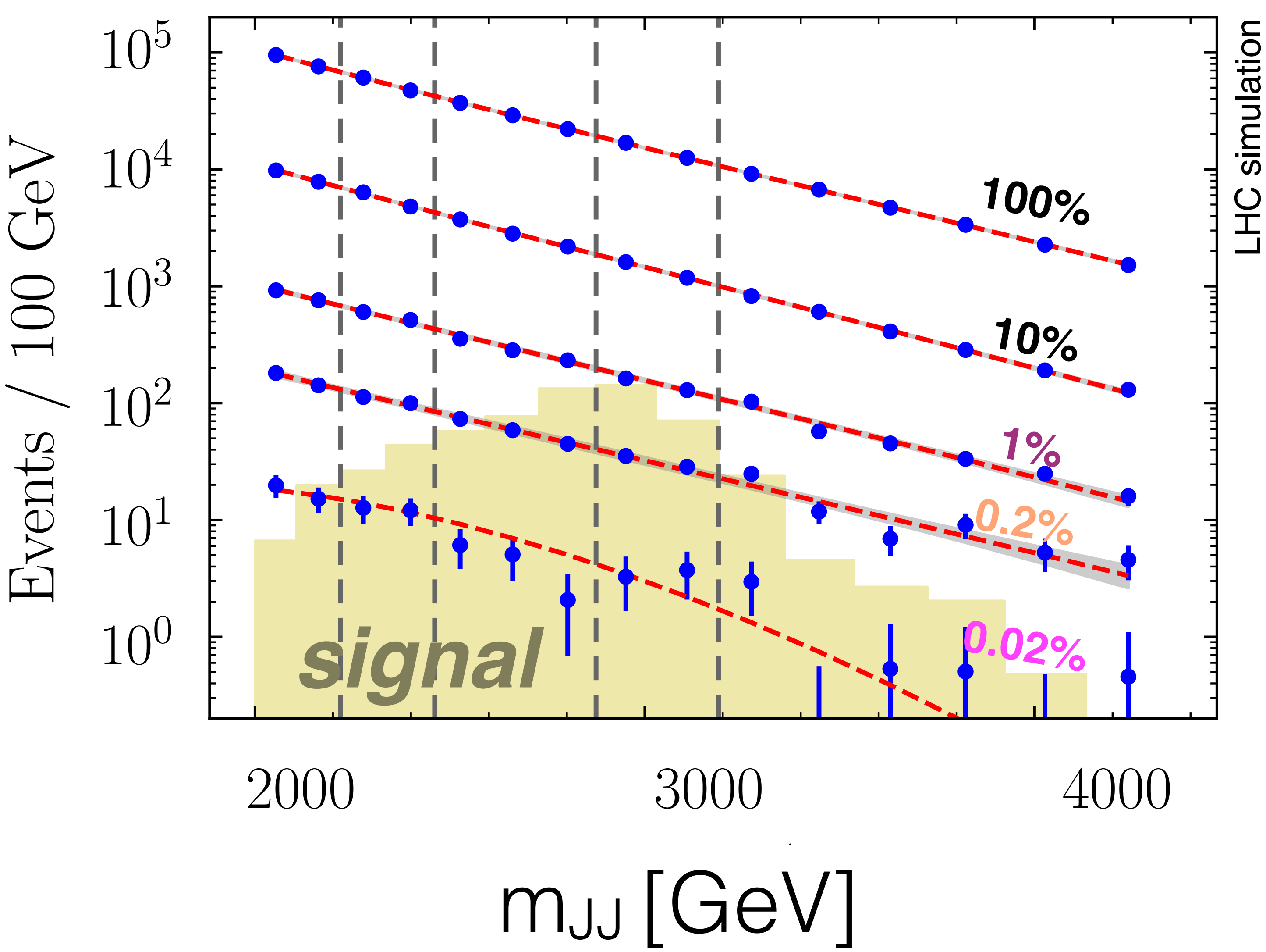
no cut on NN

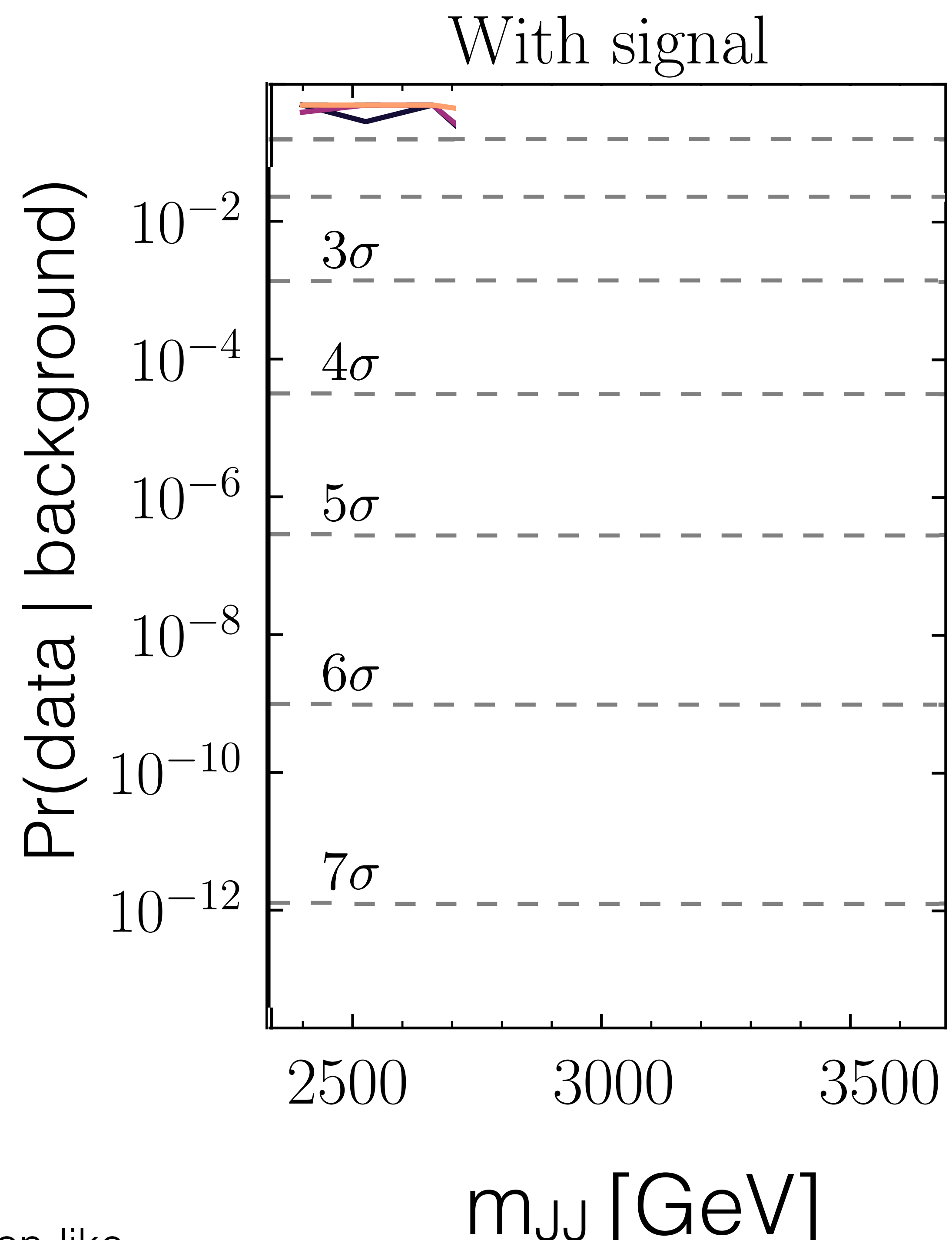
most 10% signal-region-like

most 1% signal-region-like

most 0.2% signal-region-like

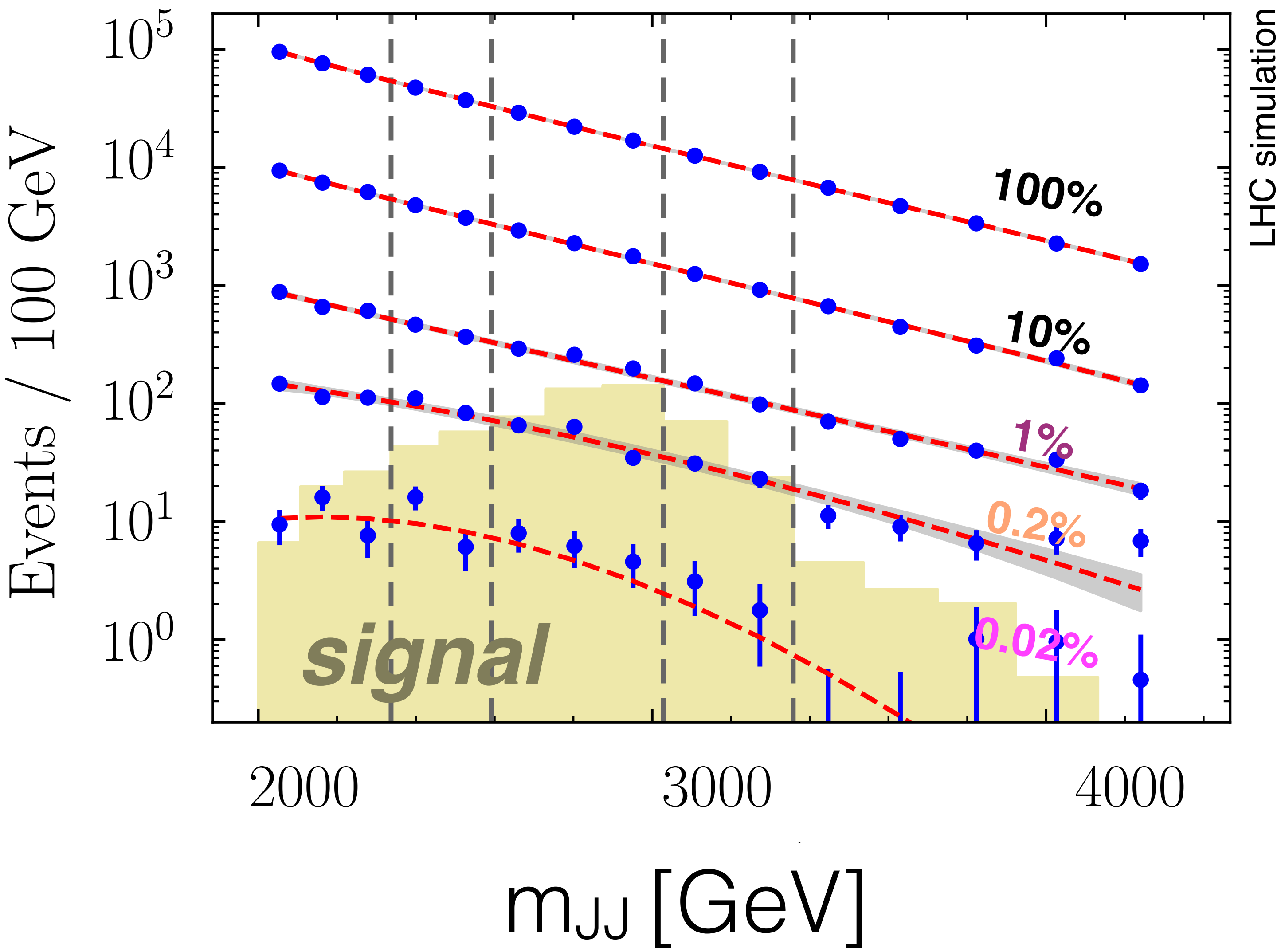
…and when there is a signal?

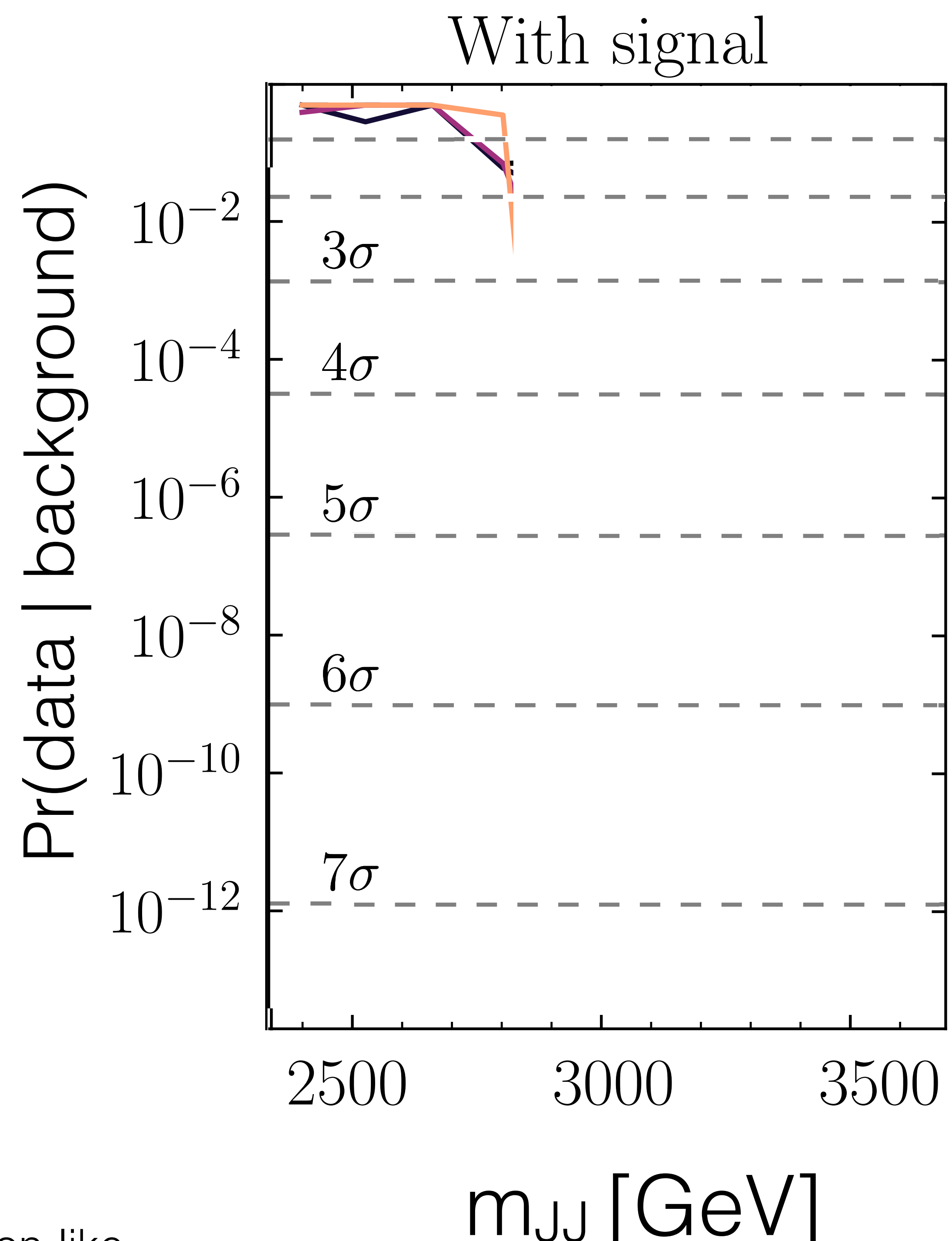




no cut on NN
most 10% signal-region-like
most 1% signal-region-like
most 0.2% signal-region-like

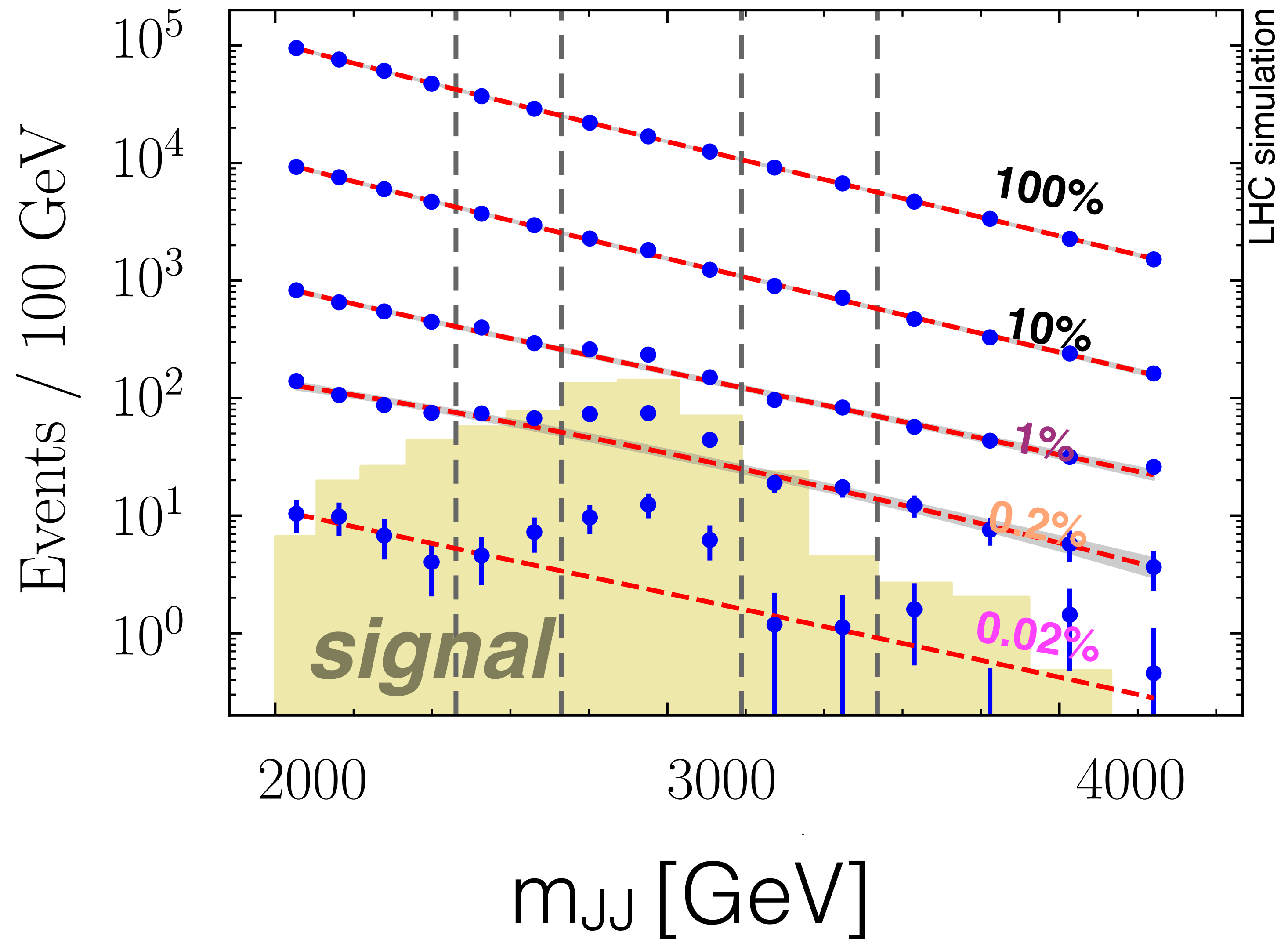
…and when there is a signal?

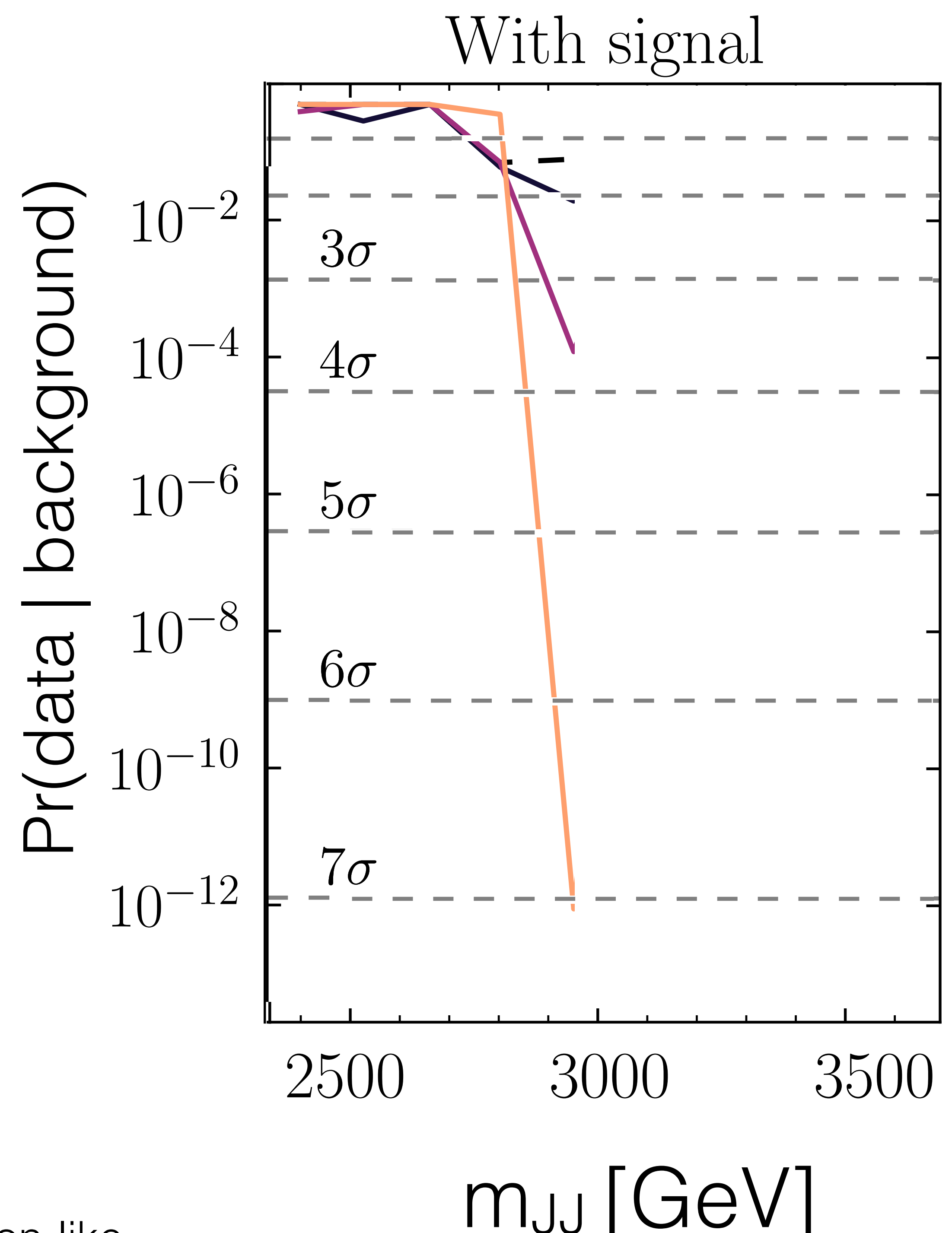




- - - no cut on NN
— most 10% signal-region-like
— most 1% signal-region-like
— most 0.2% signal-region-like

…and when there is a signal?





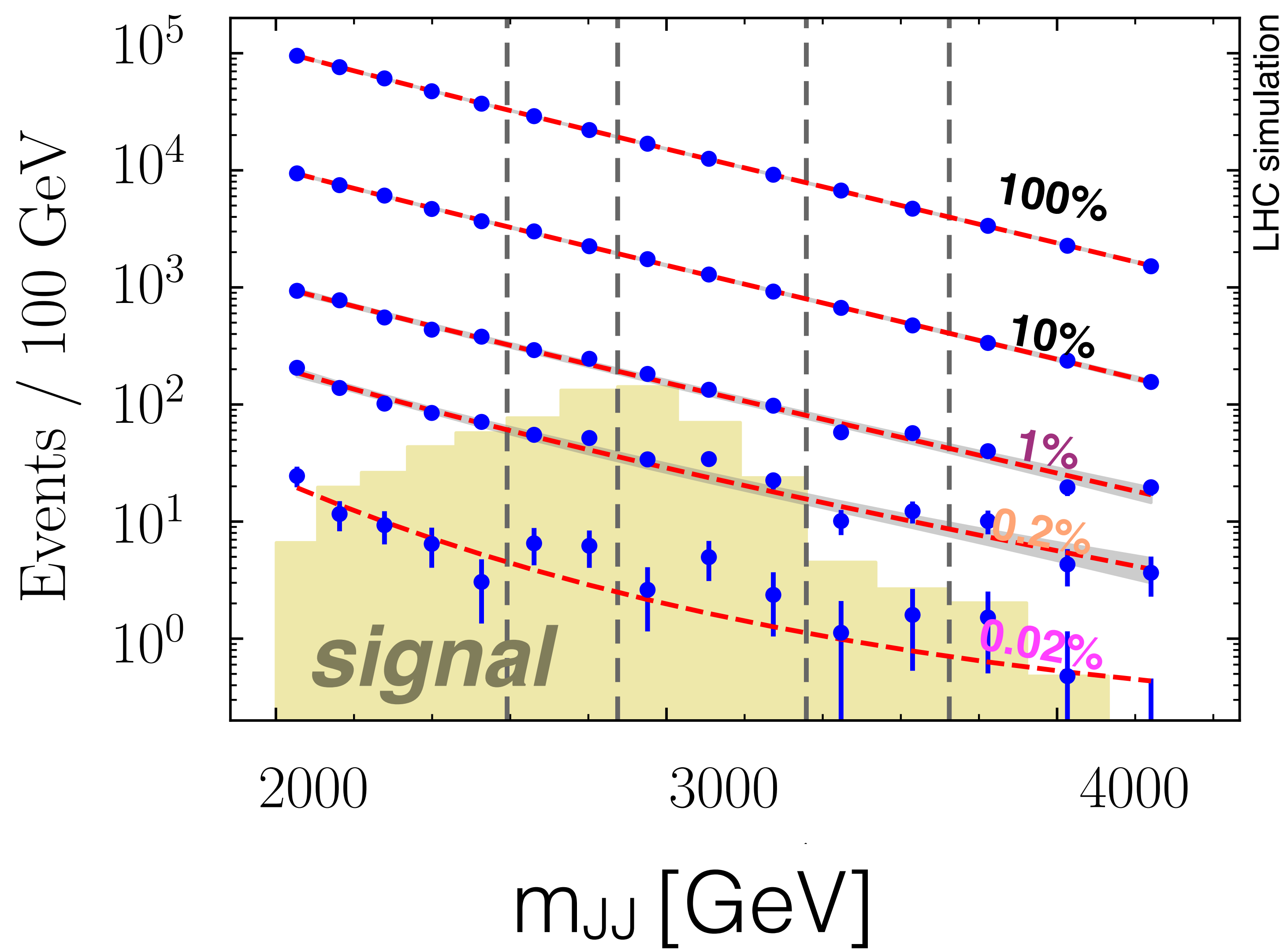
no cut on NN

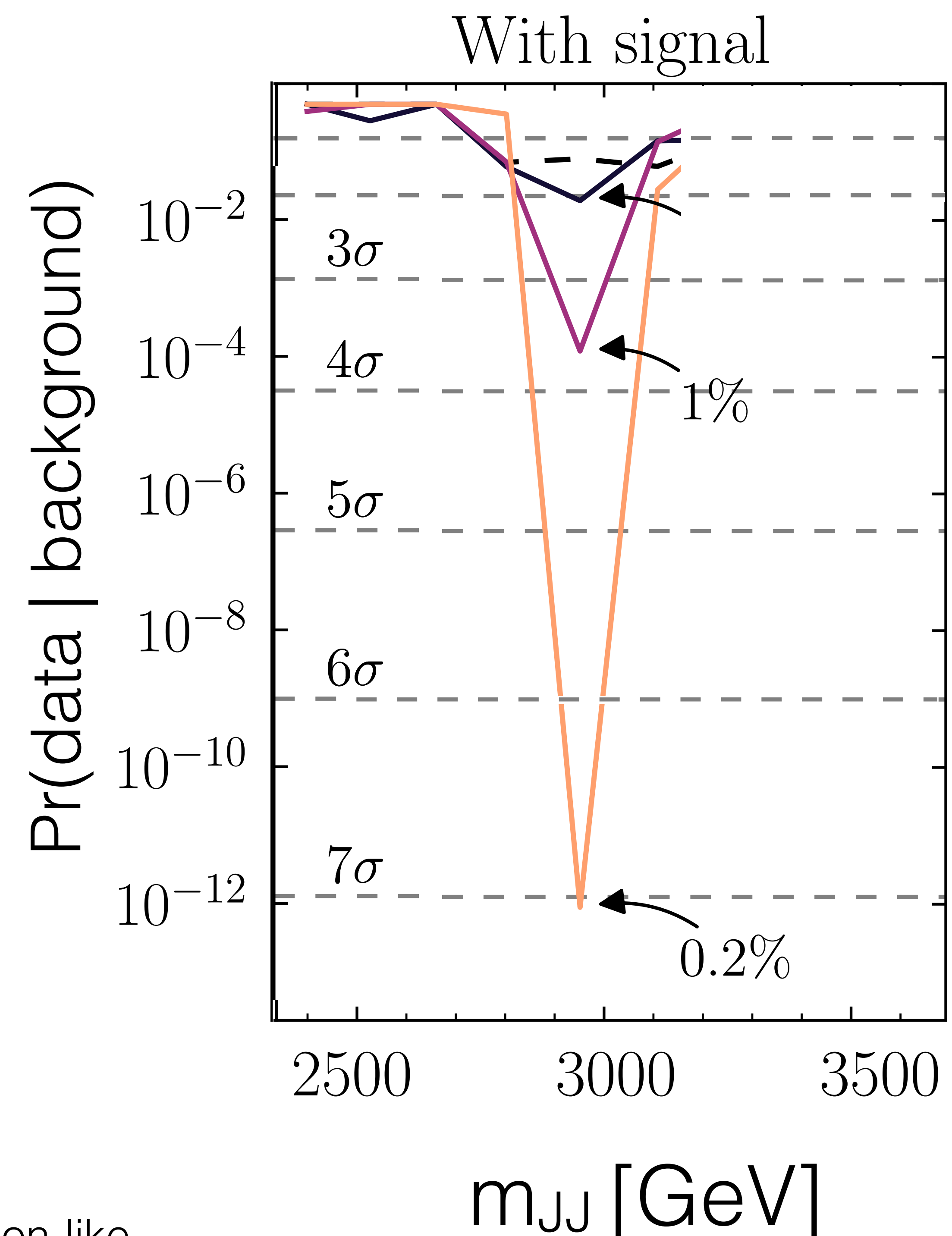
most 10% signal-region-like

most 1% signal-region-like

most 0.2% signal-region-like

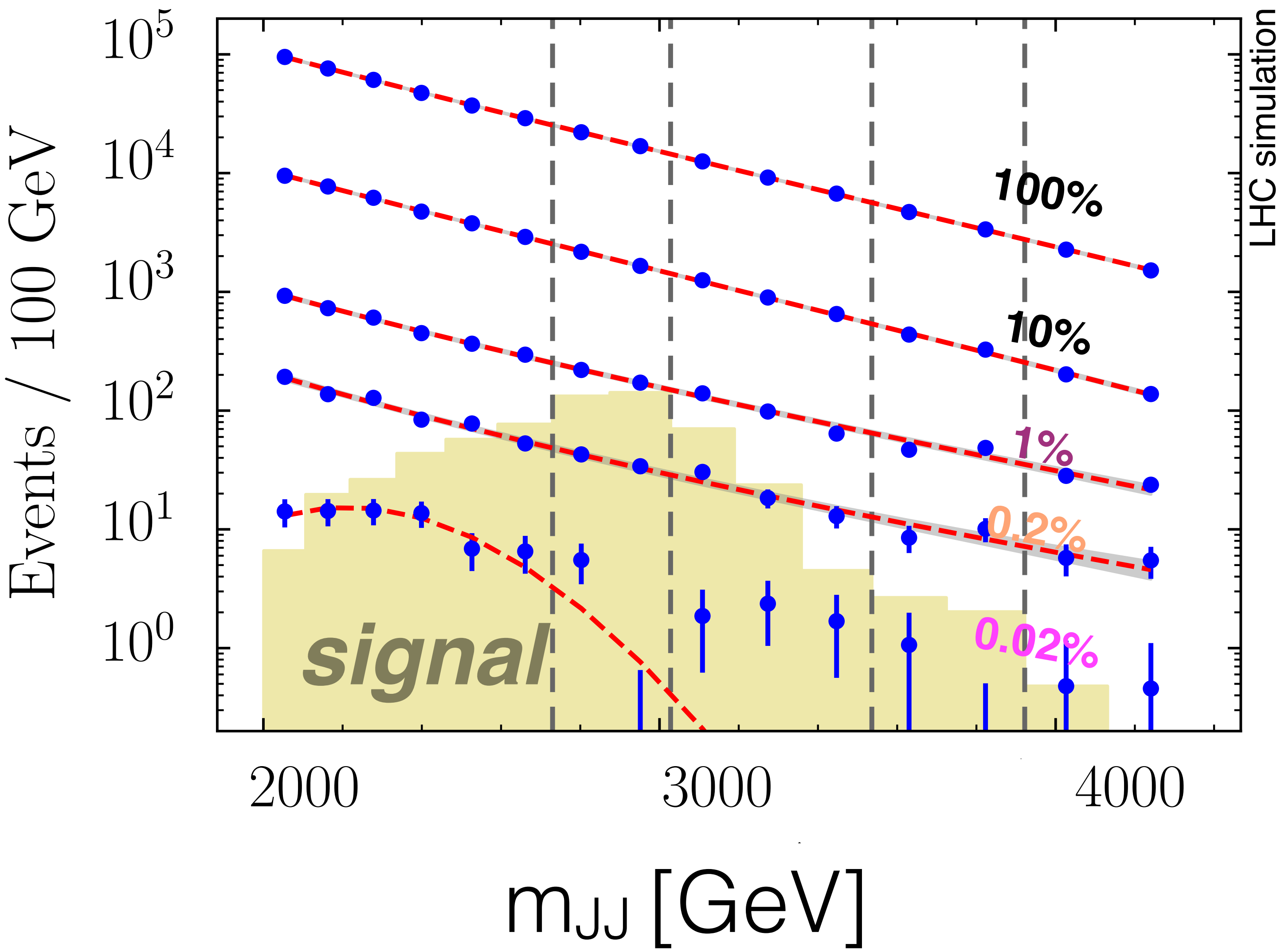
…and when there is a signal?

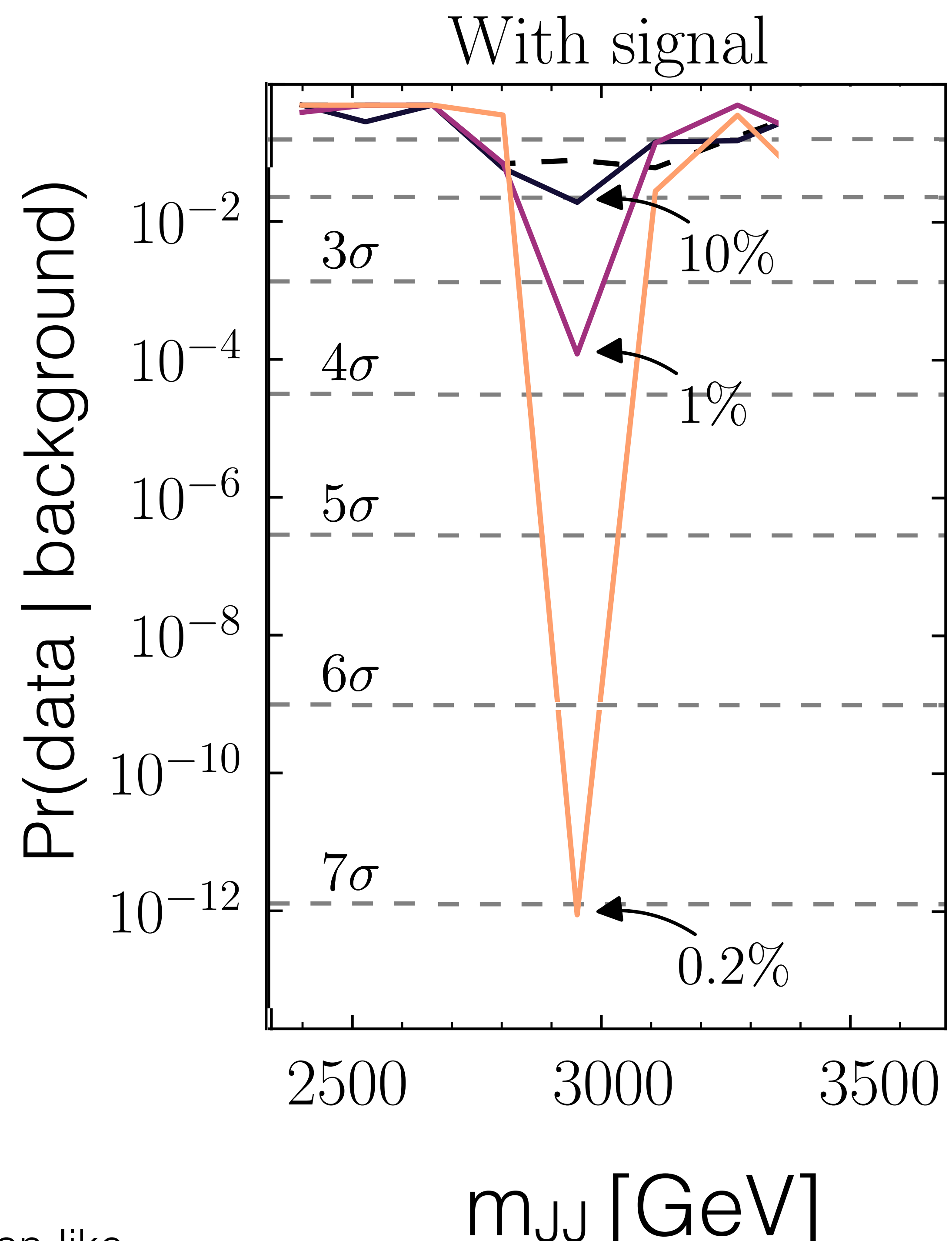




no cut on NN
most 10% signal-region-like
most 1% signal-region-like
most 0.2% signal-region-like

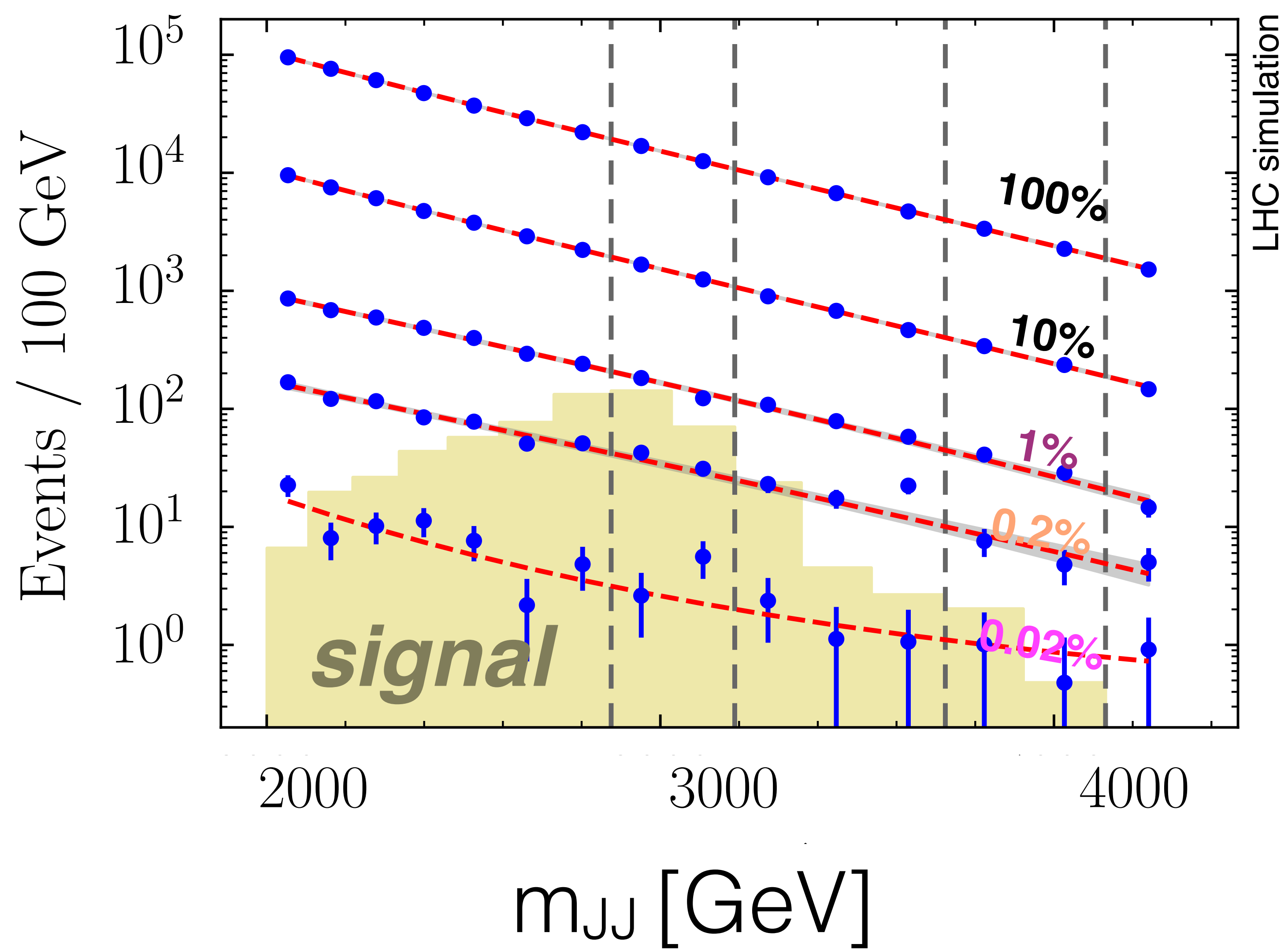
…and when there is a signal?

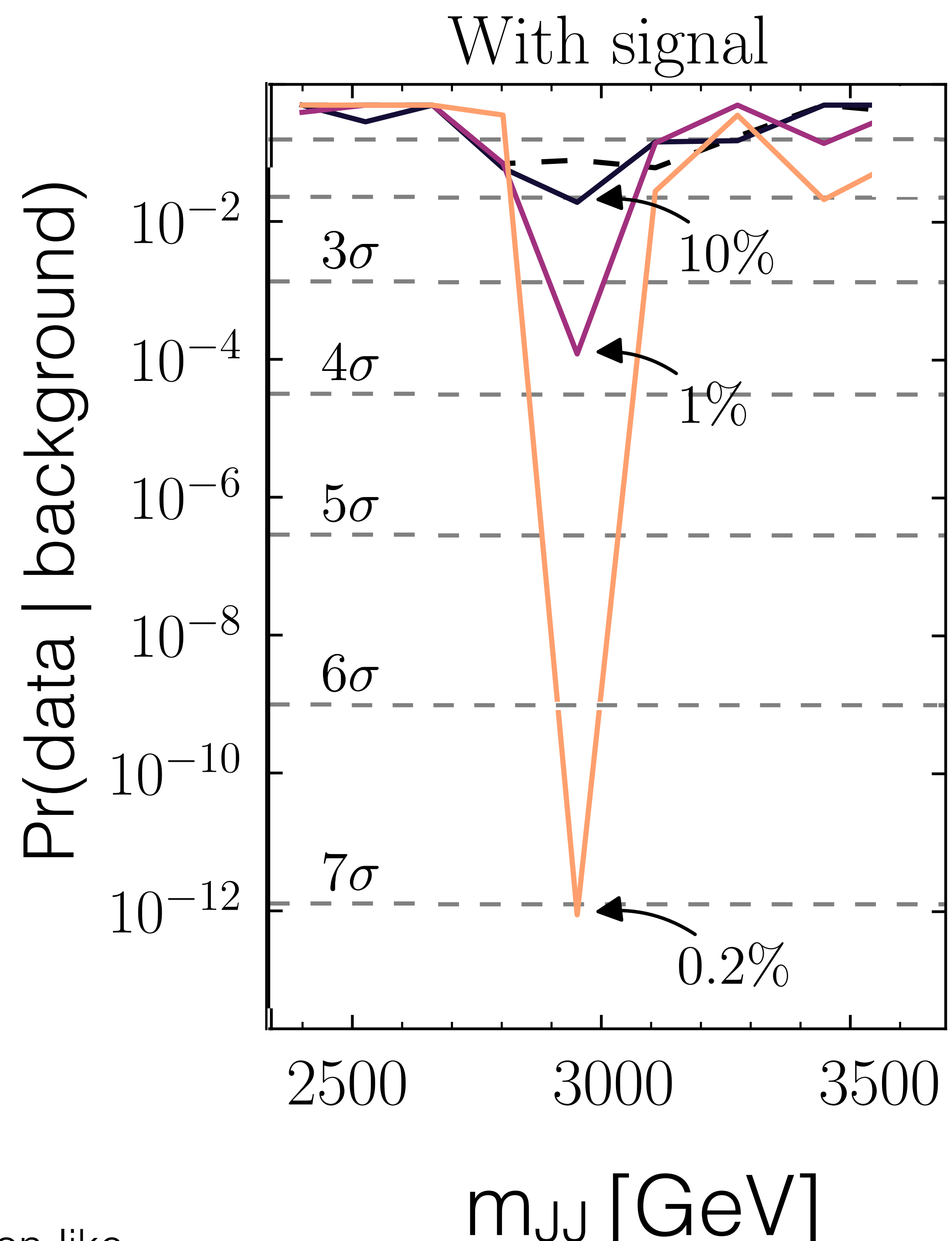




no cut on NN
most 10% signal-region-like
most 1% signal-region-like
most 0.2% signal-region-like

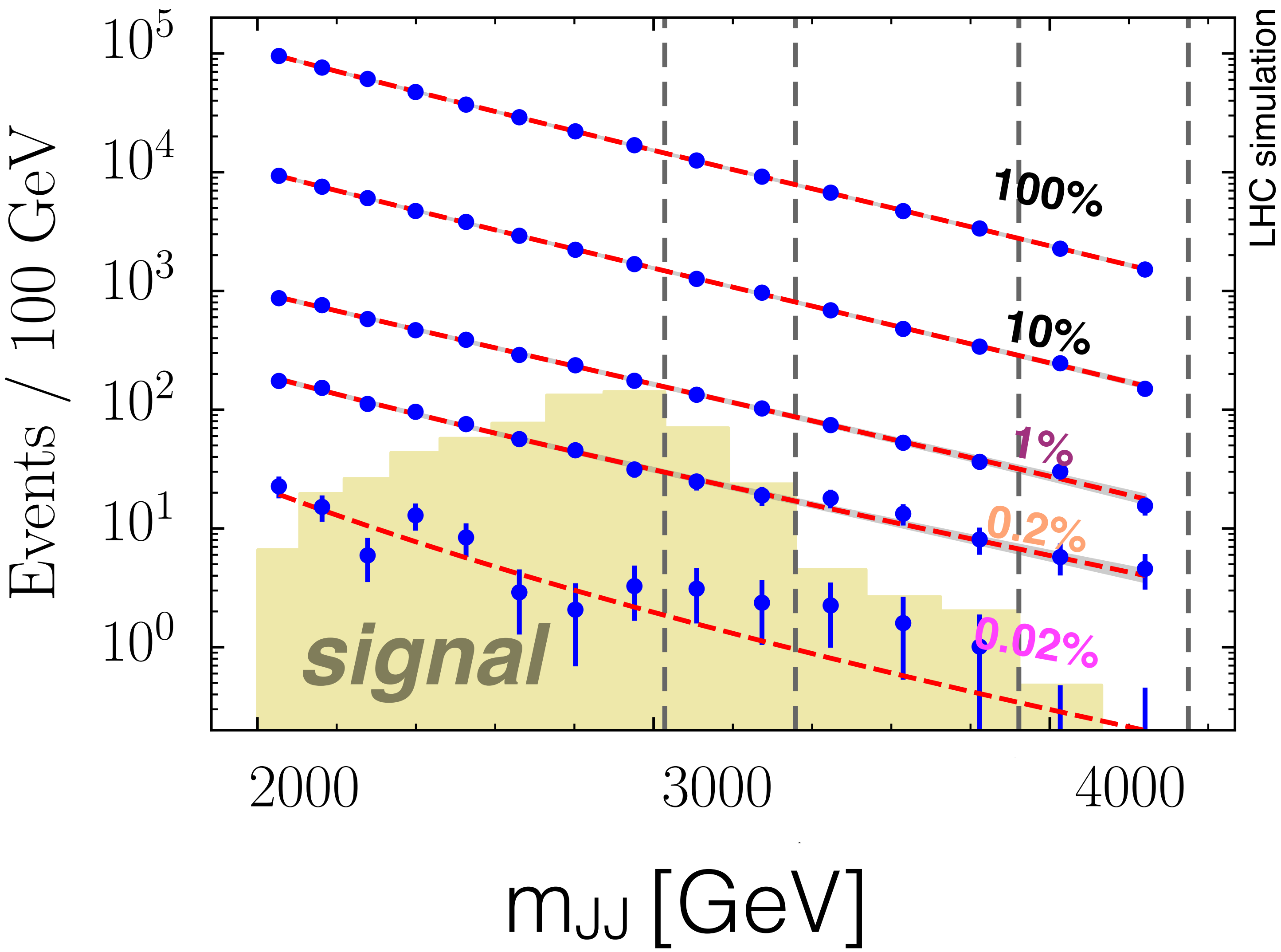
…and when there is a signal?

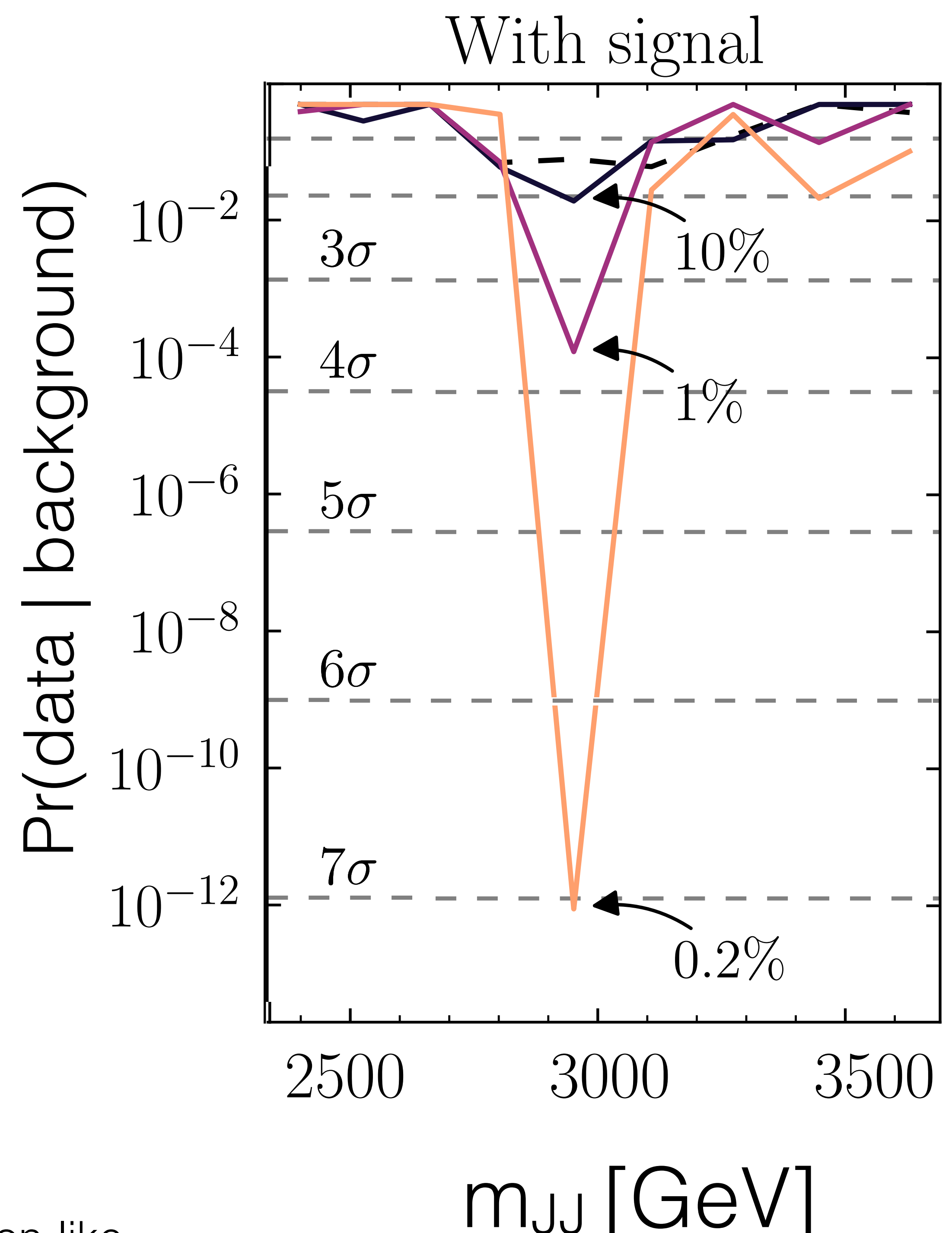




no cut on NN
most 10% signal-region-like
most 1% signal-region-like
most 0.2% signal-region-like

…and when there is a signal?

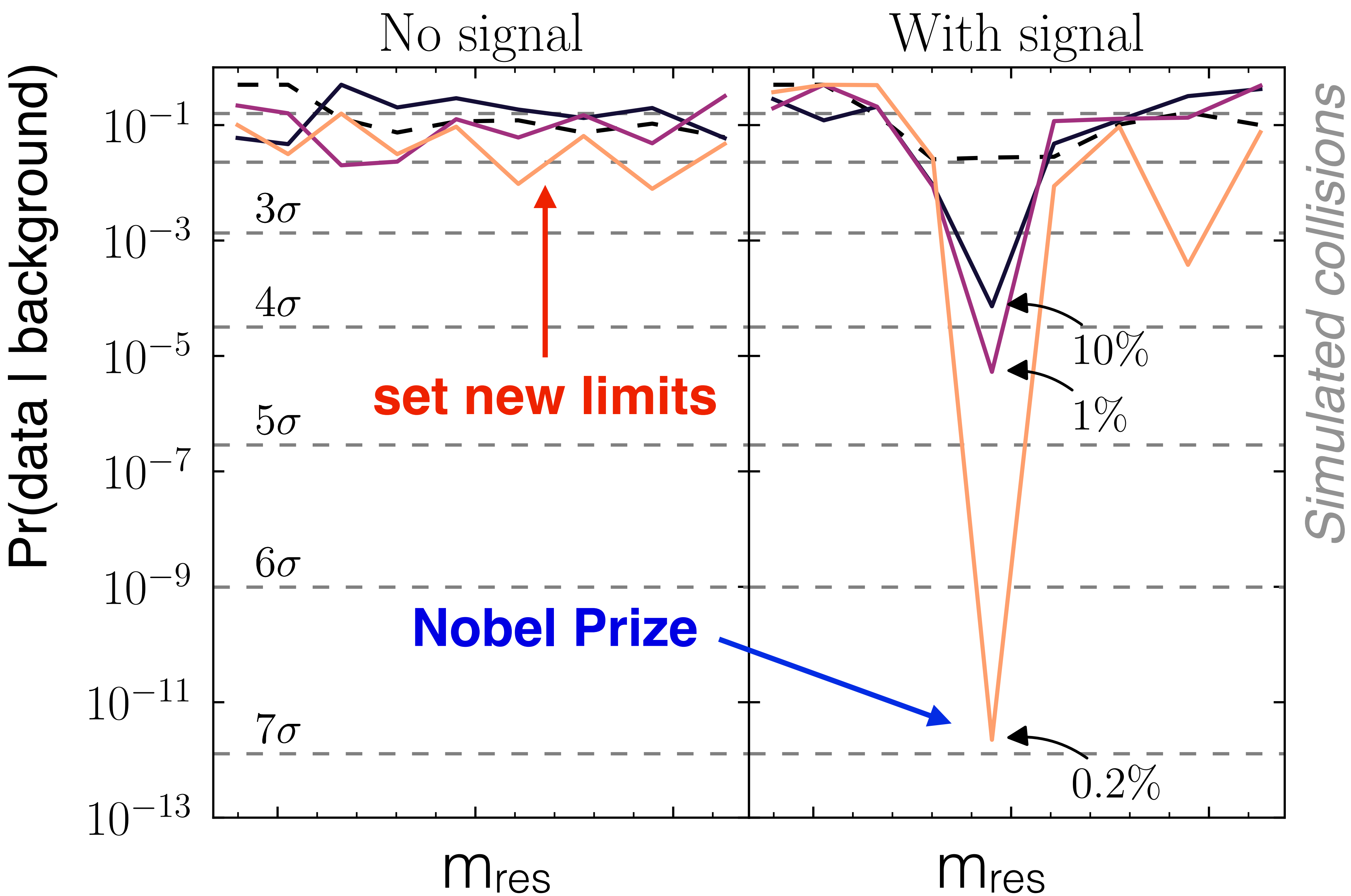




no cut on NN
most 10% signal-region-like
most 1% signal-region-like
most 0.2% signal-region-like

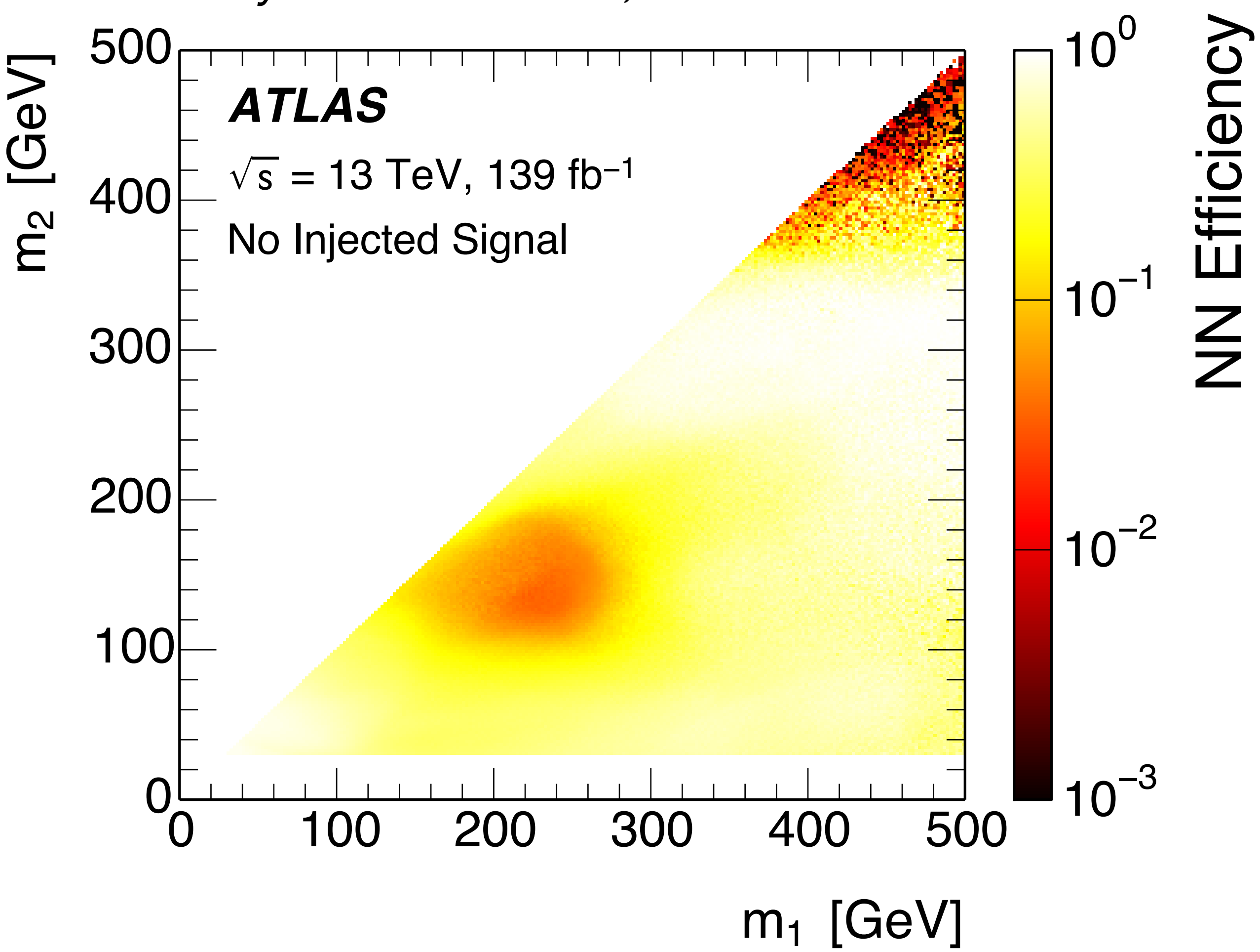
Anomaly detection: Overview

J. Collins, K. Howe, BPN,
Phys. Rev. Lett. 121 (2018)
241803, 1805.02664



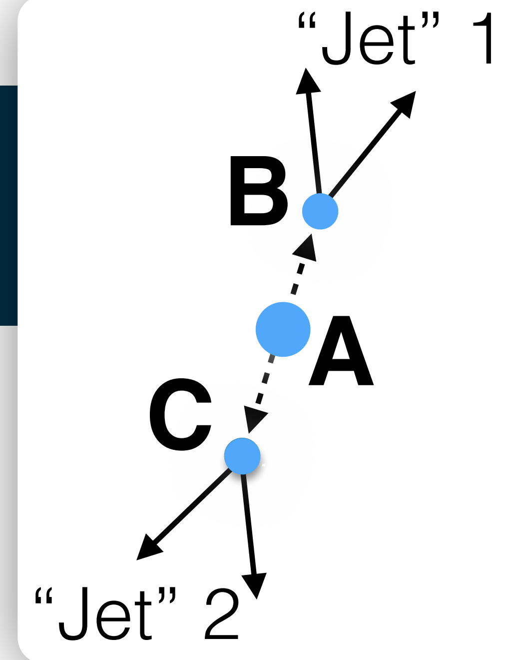
Collision data results New

ATLAS Collaboration, 2005.02983
Phys. Rev. Lett. 125, 131801

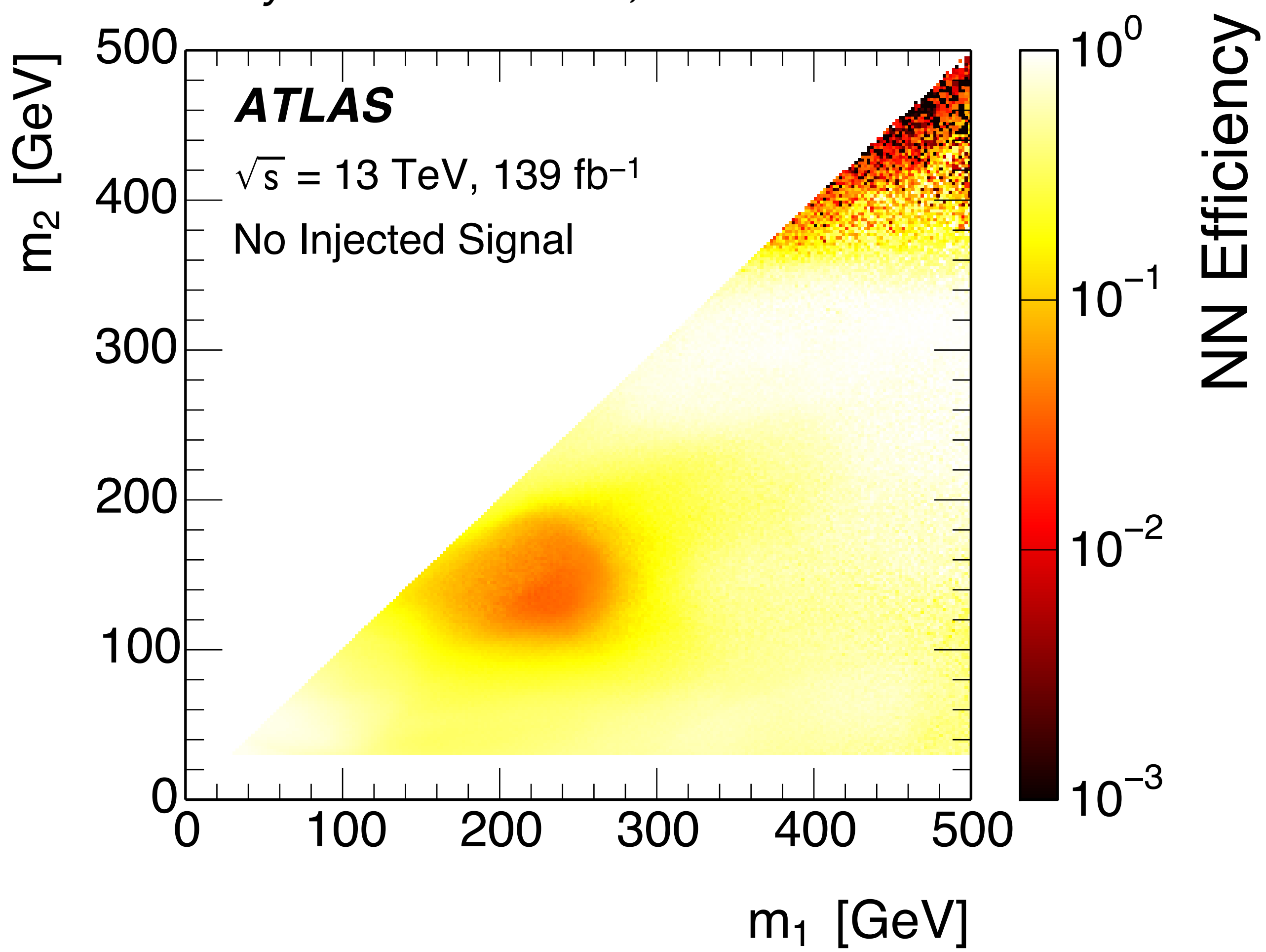


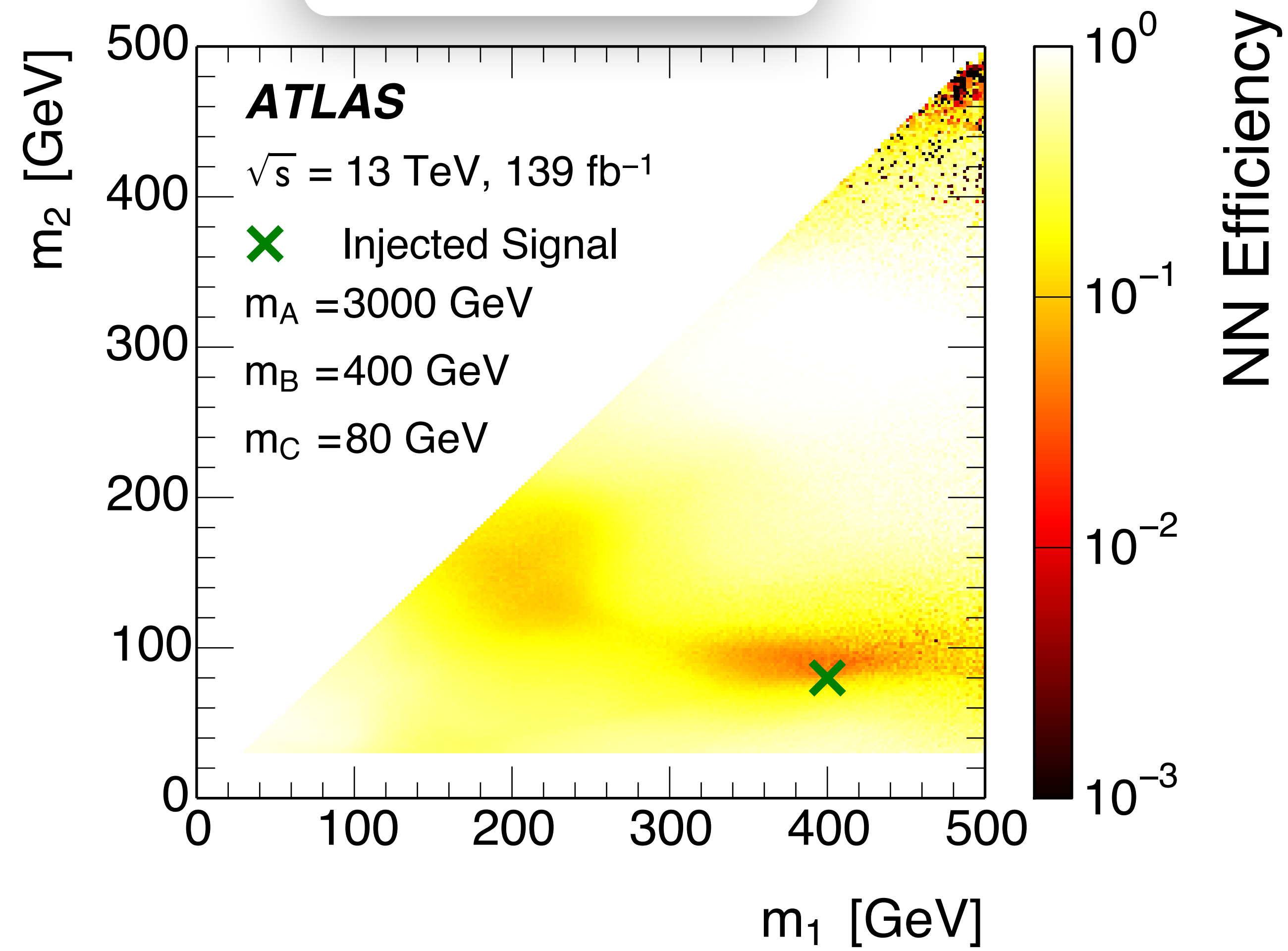
First round, keep it simple: feature space is 2D (jet masses)

Collision data results New



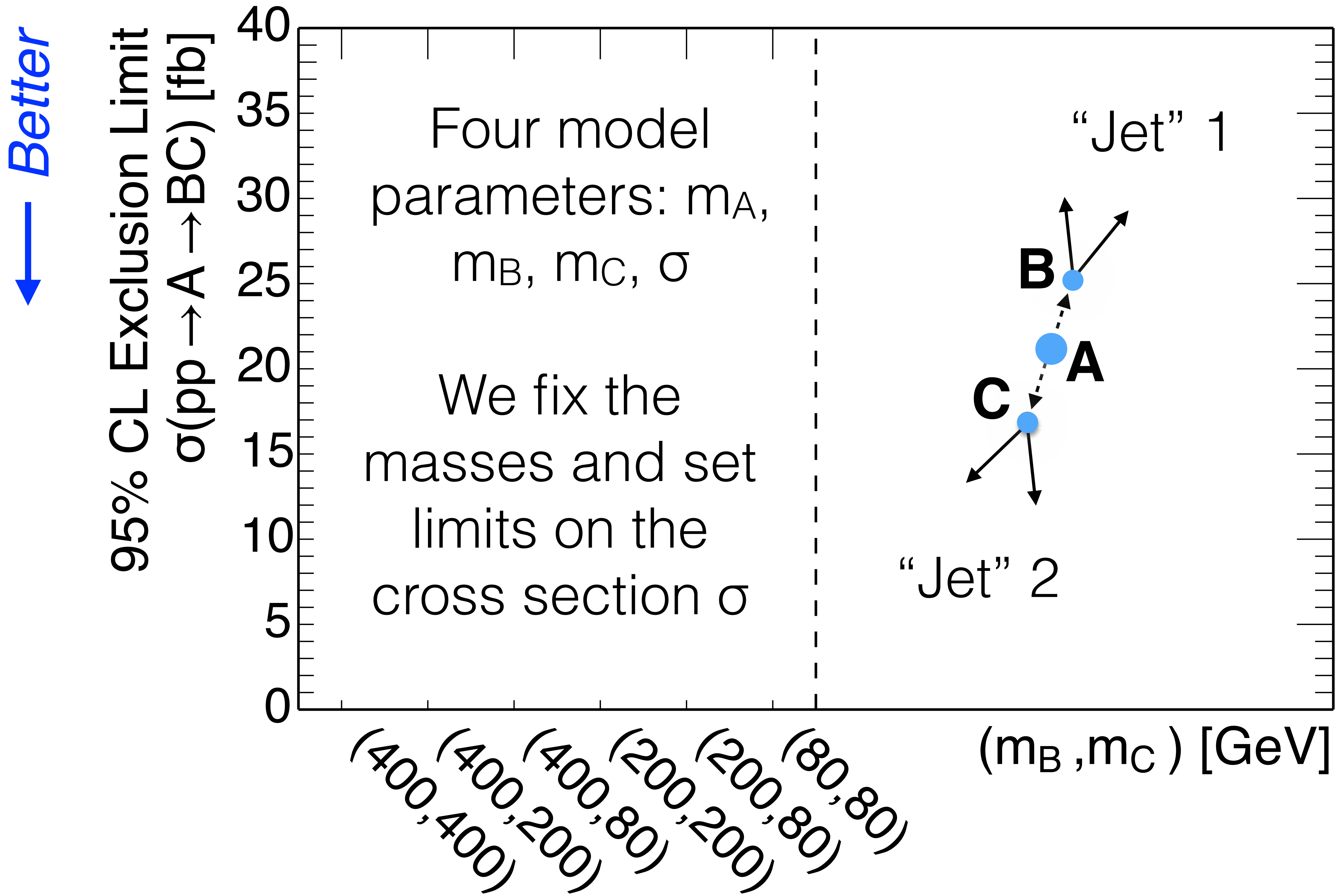
ATLAS Collaboration, 2005.02983
Phys. Rev. Lett. 125, 131801





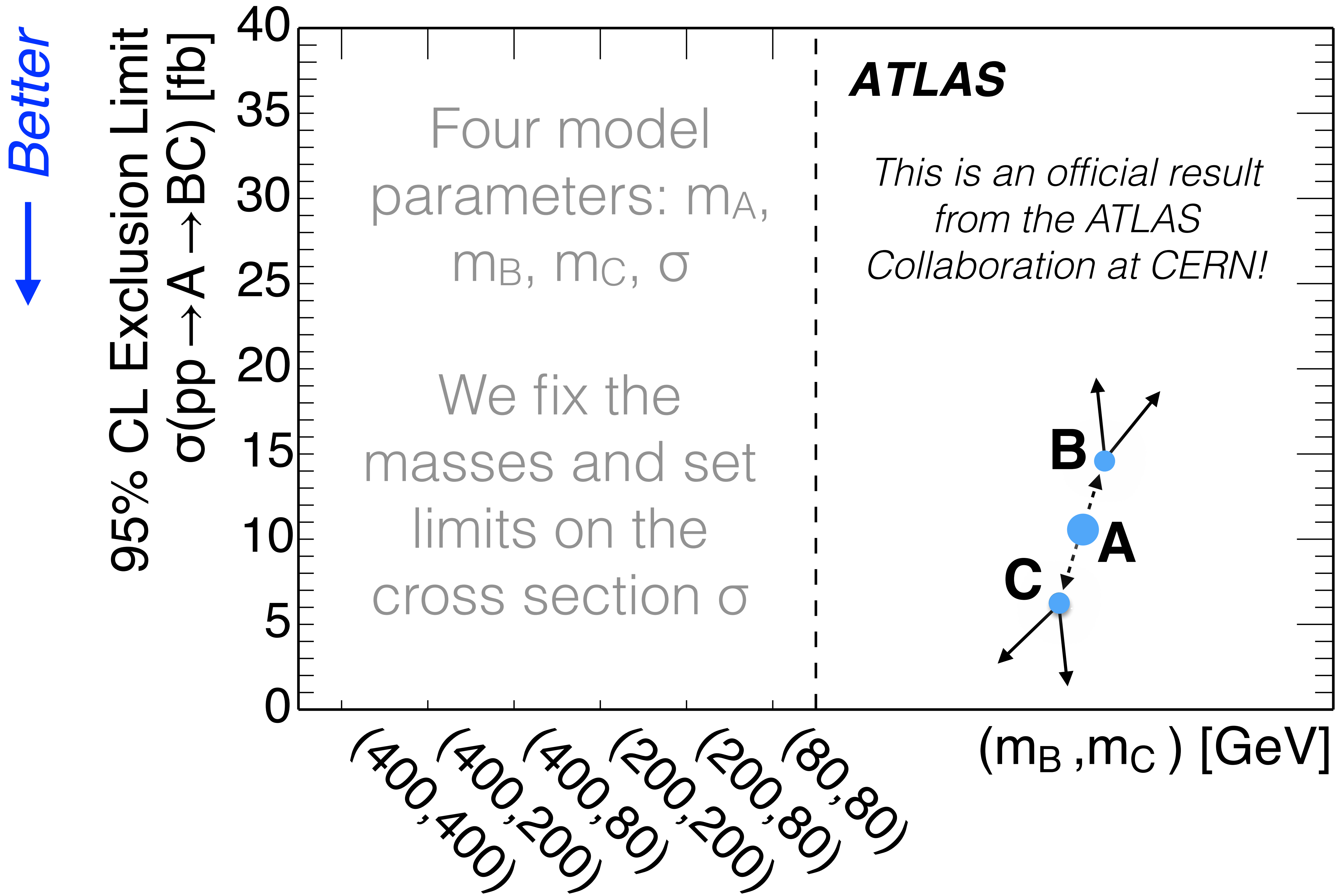
First round, keep it simple: feature space is 2D (jet masses)

Collision data results New



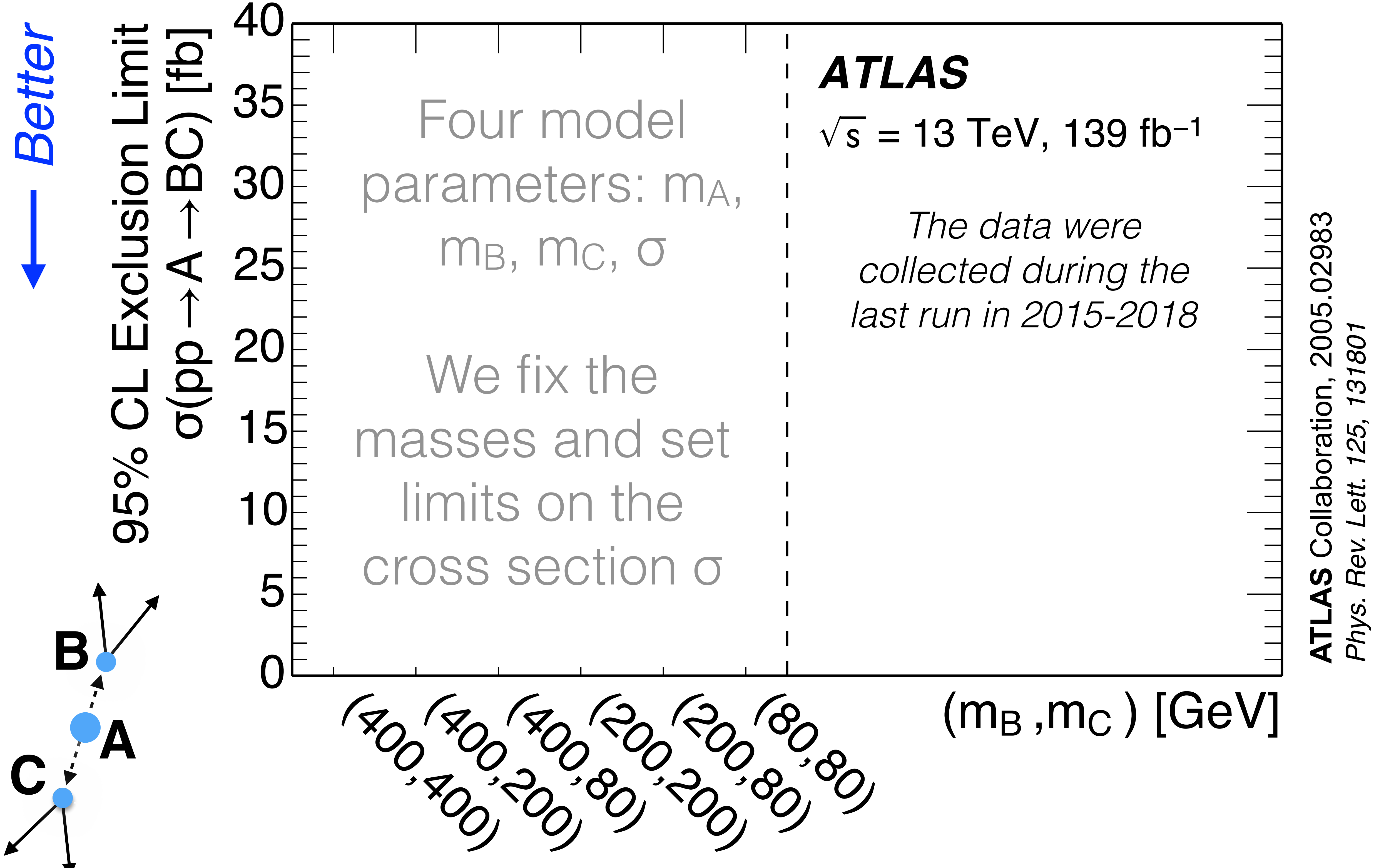
ATLAS Collaboration, 2005.02983
Phys. Rev. Lett. 125, 131801

Collision data results **New**

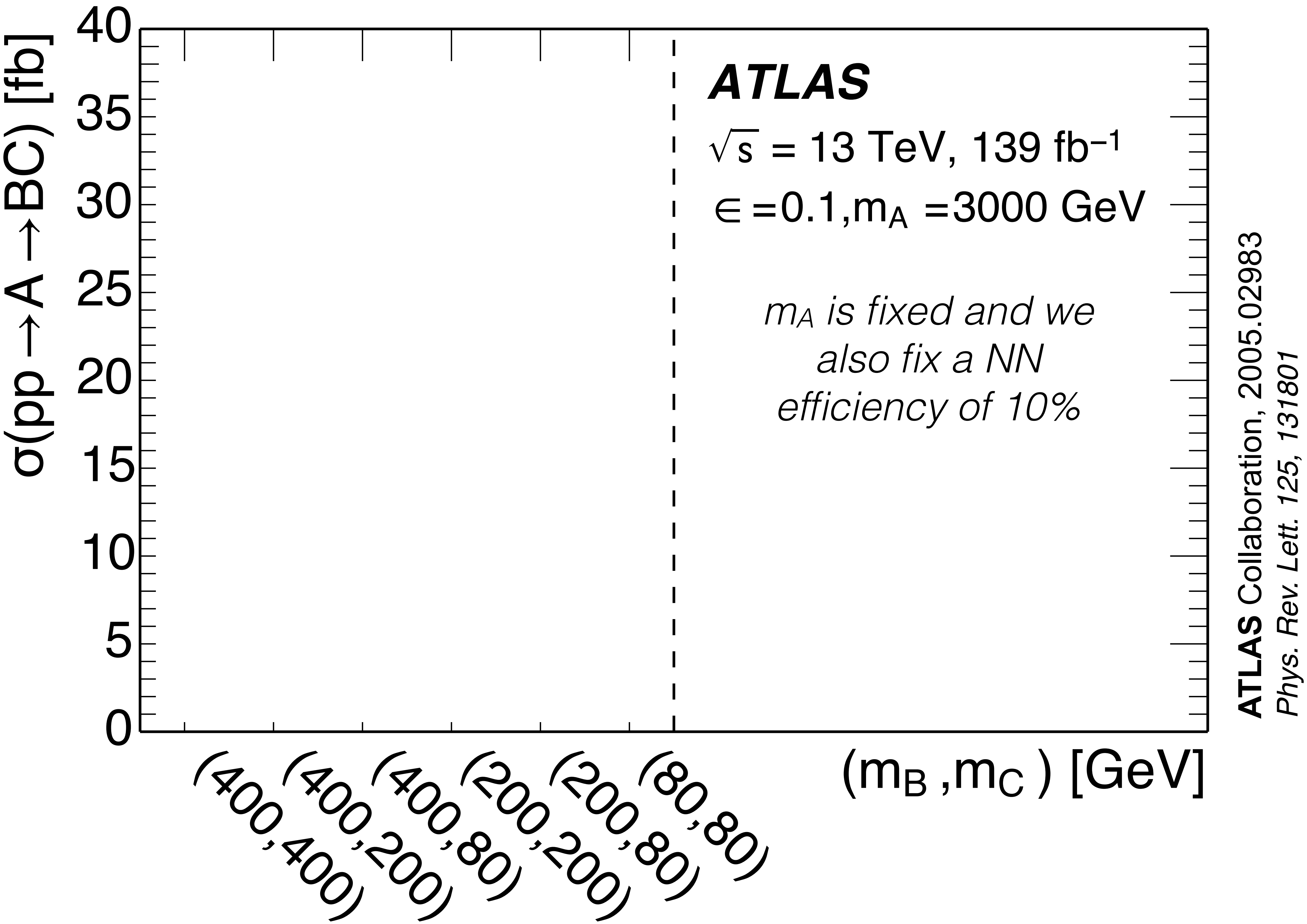


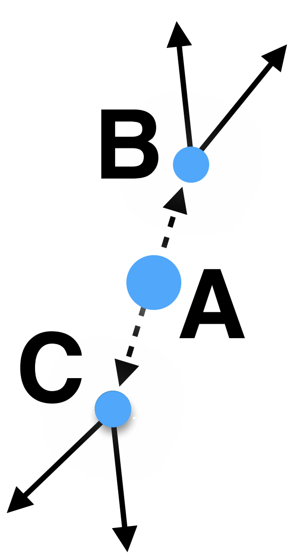
ATLAS Collaboration, 2005.02983
Phys. Rev. Lett. 125, 131801

Collision data results New

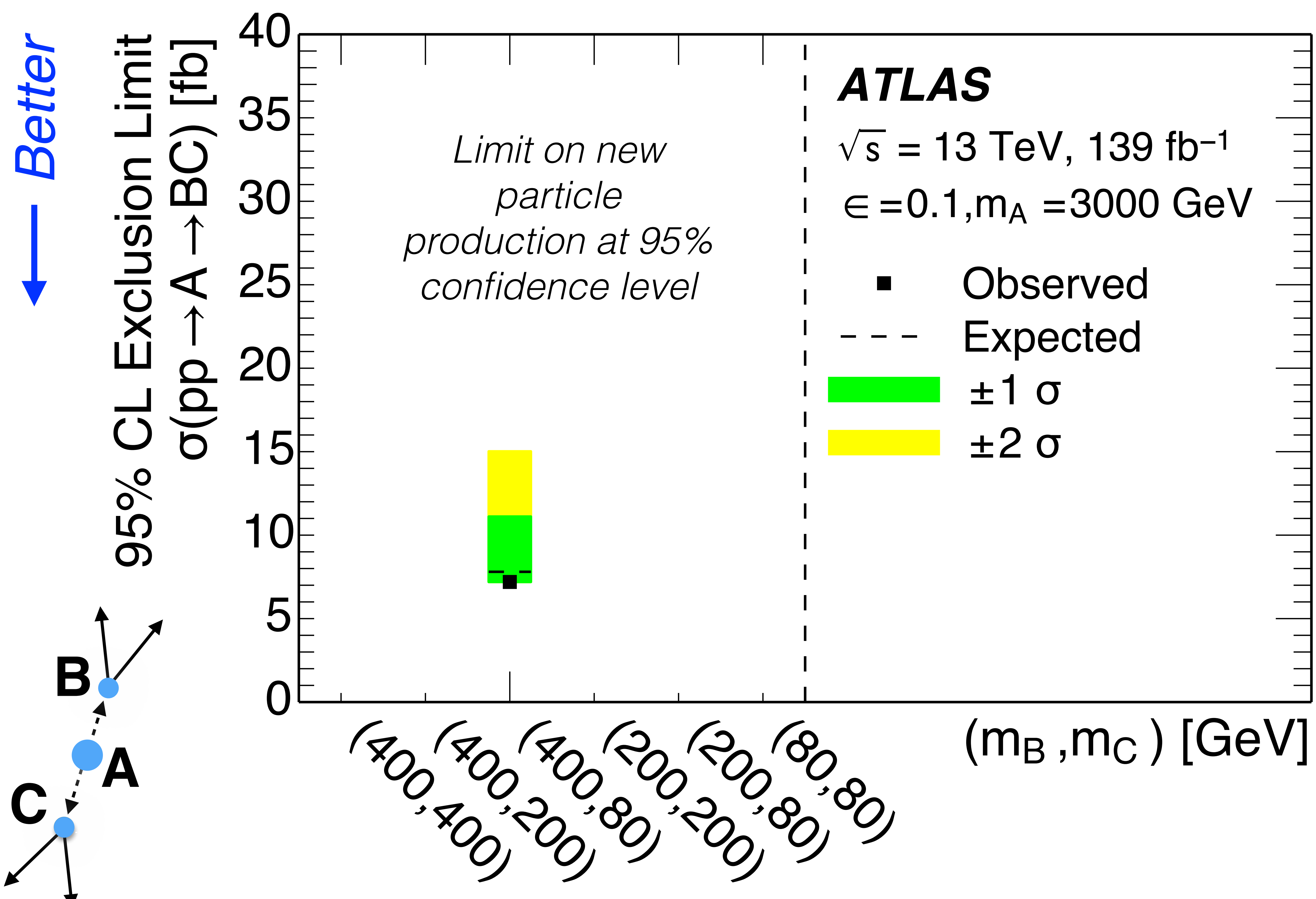


Collision data results New





Collision data results New

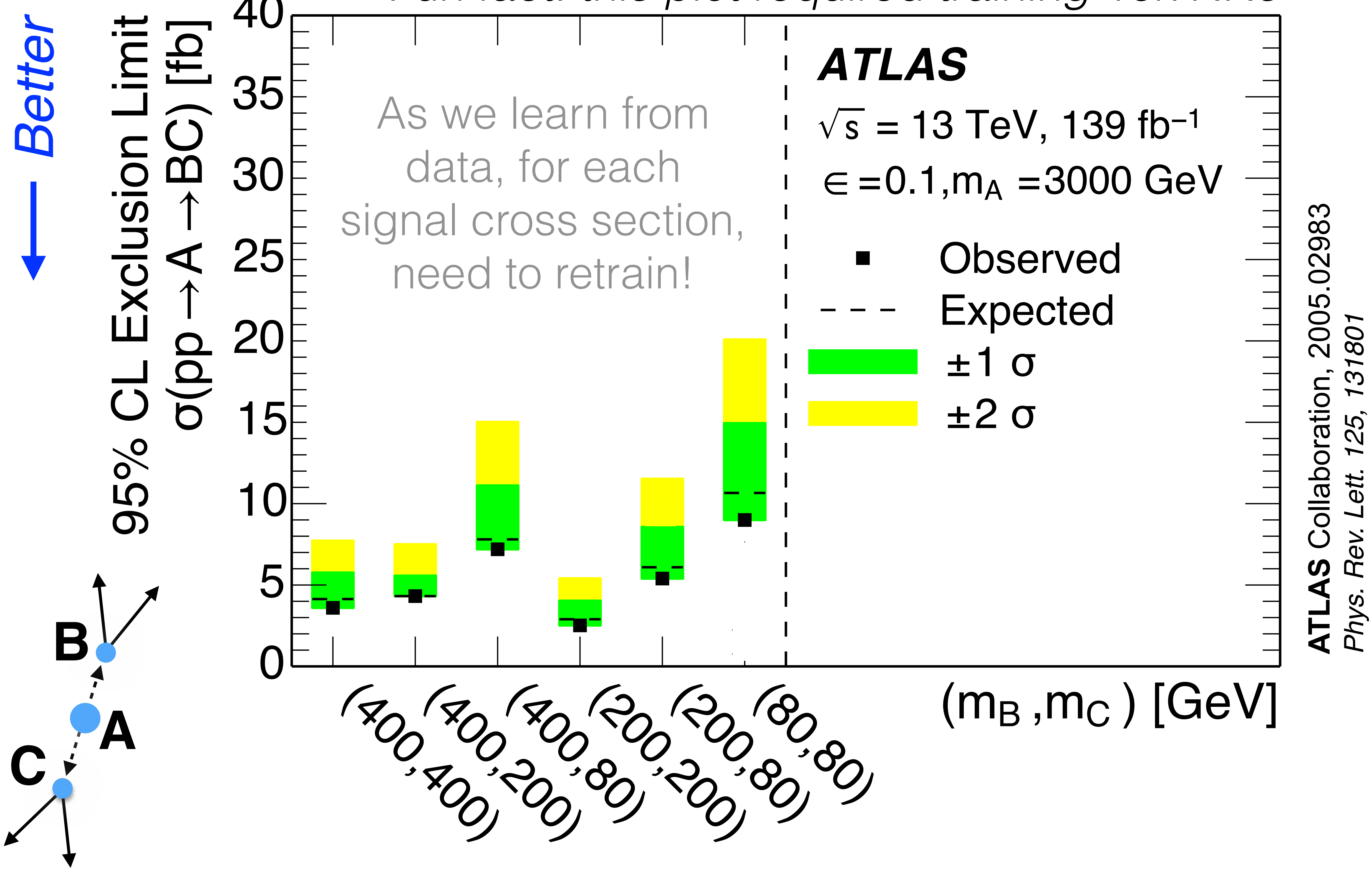


ATLAS Collaboration, 2005.02983

Phys. Rev. Lett. 125, 131801

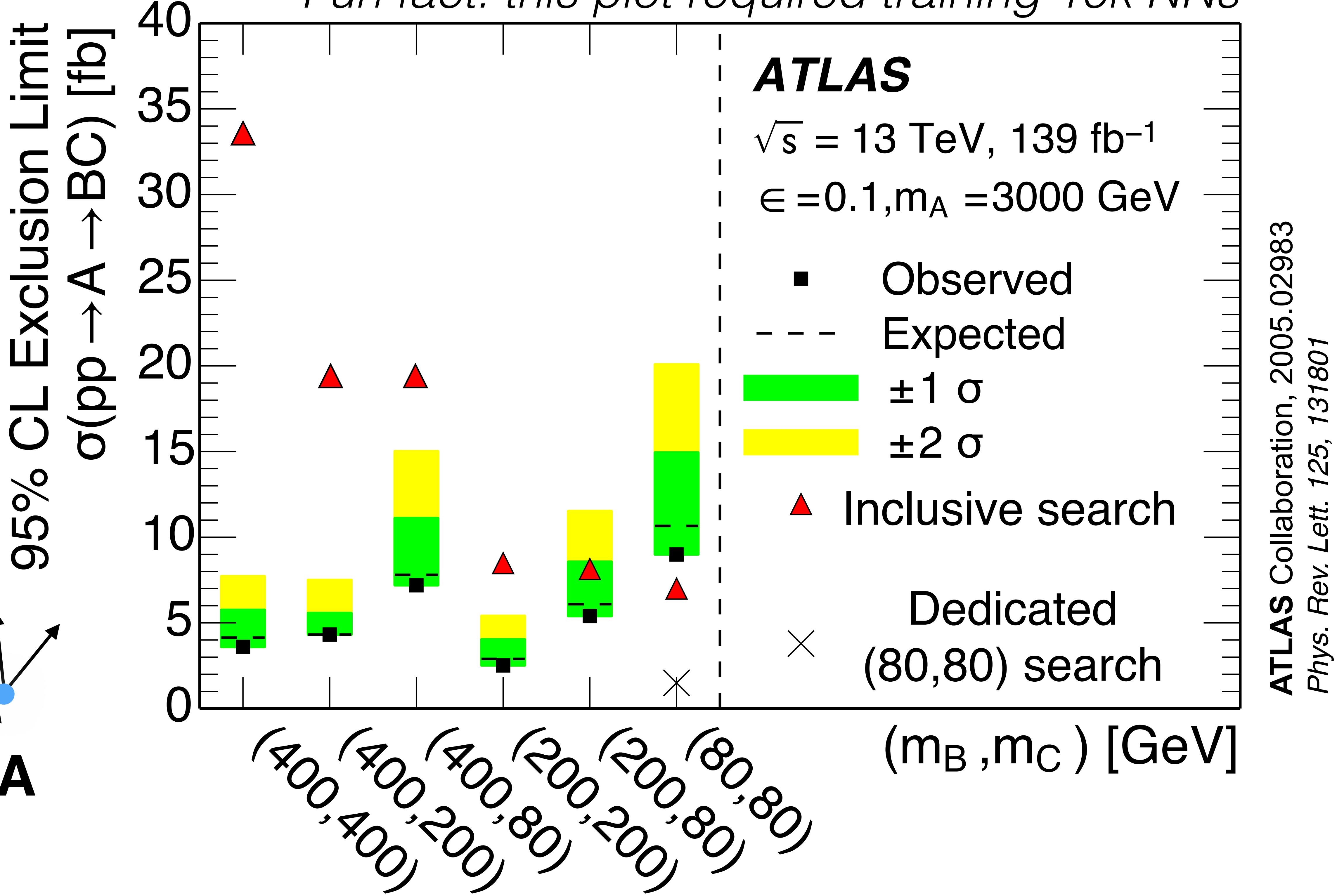
Collision data results New

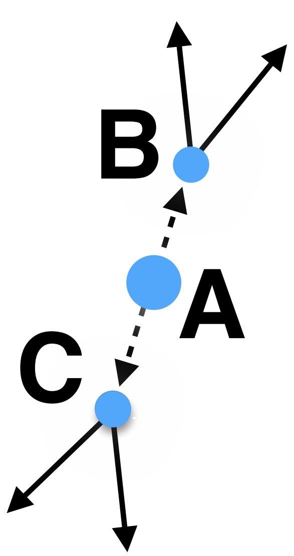
Fun fact: this plot required training 10k NNs



Collision data results New

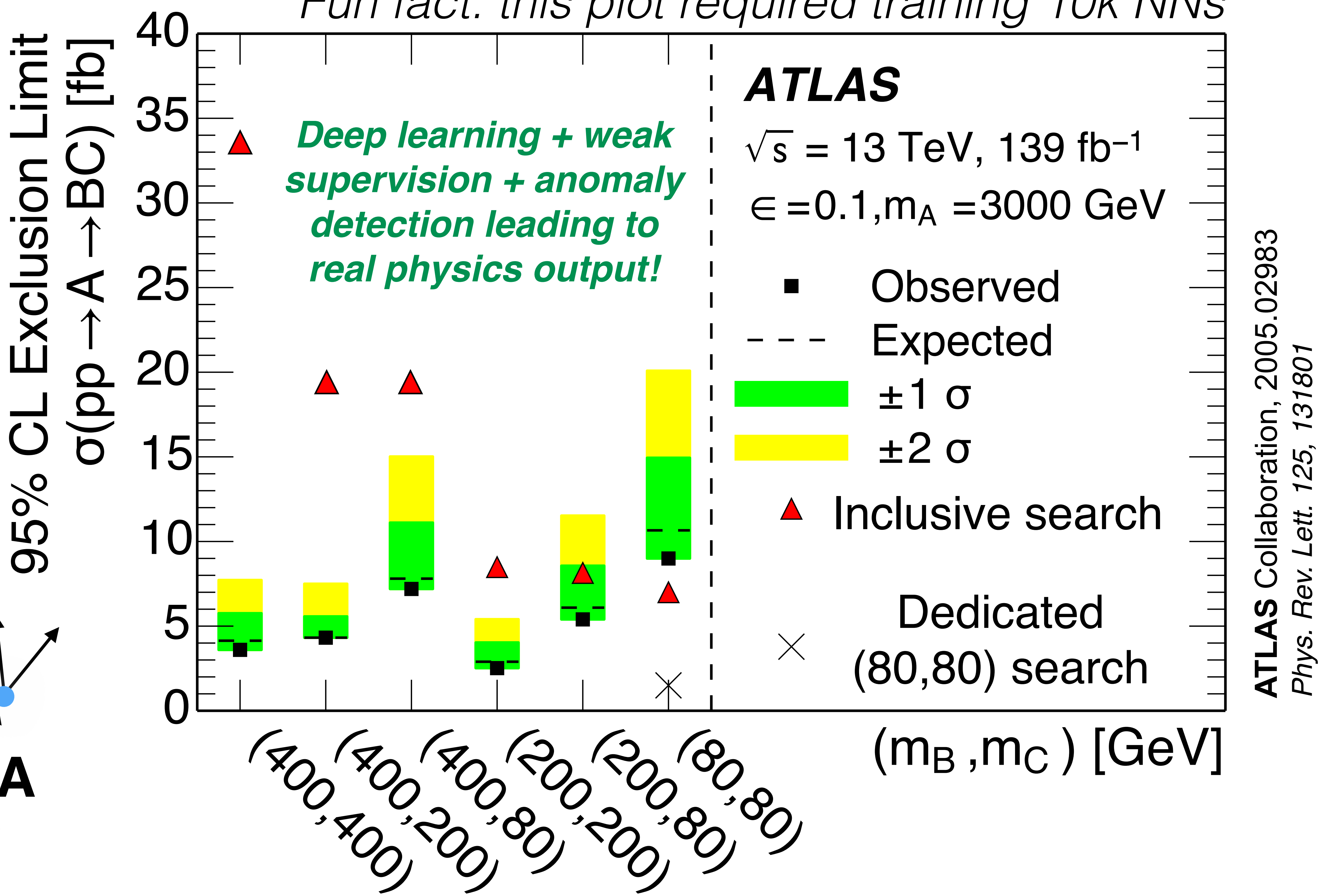
Fun fact: this plot required training 10k NNs

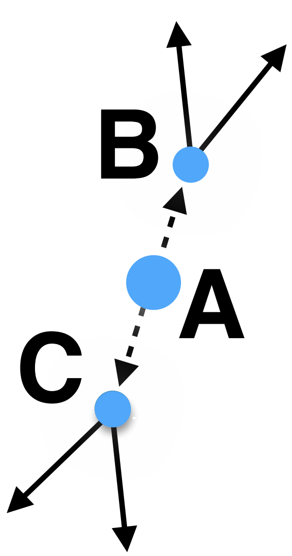




Collision data results New

Fun fact: this plot required training 10k NNs





Computational Challenges

360 NNs

6 signal regions

5-fold cross validation *1 part test, 1 part val, 3 parts train*

Average over 4 possible validation sets

Take the best over 3 different random initializations

(due to very small signal, this is important)

Computational Challenges

10k NNs

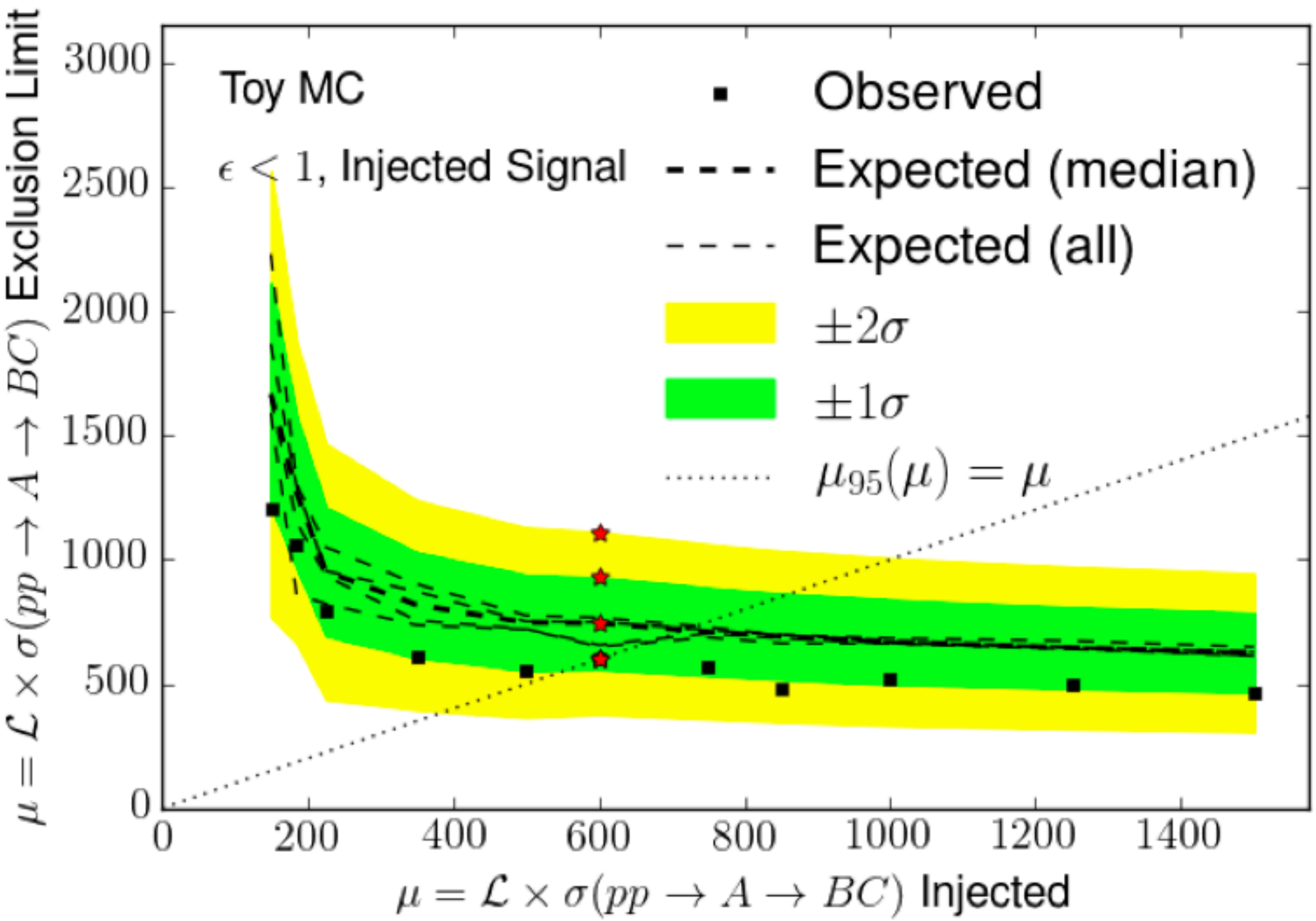
For every signal model (6)

For every signal cross section (5)

[For every systematic uncertainty]

360 NNs

Computational Challenges



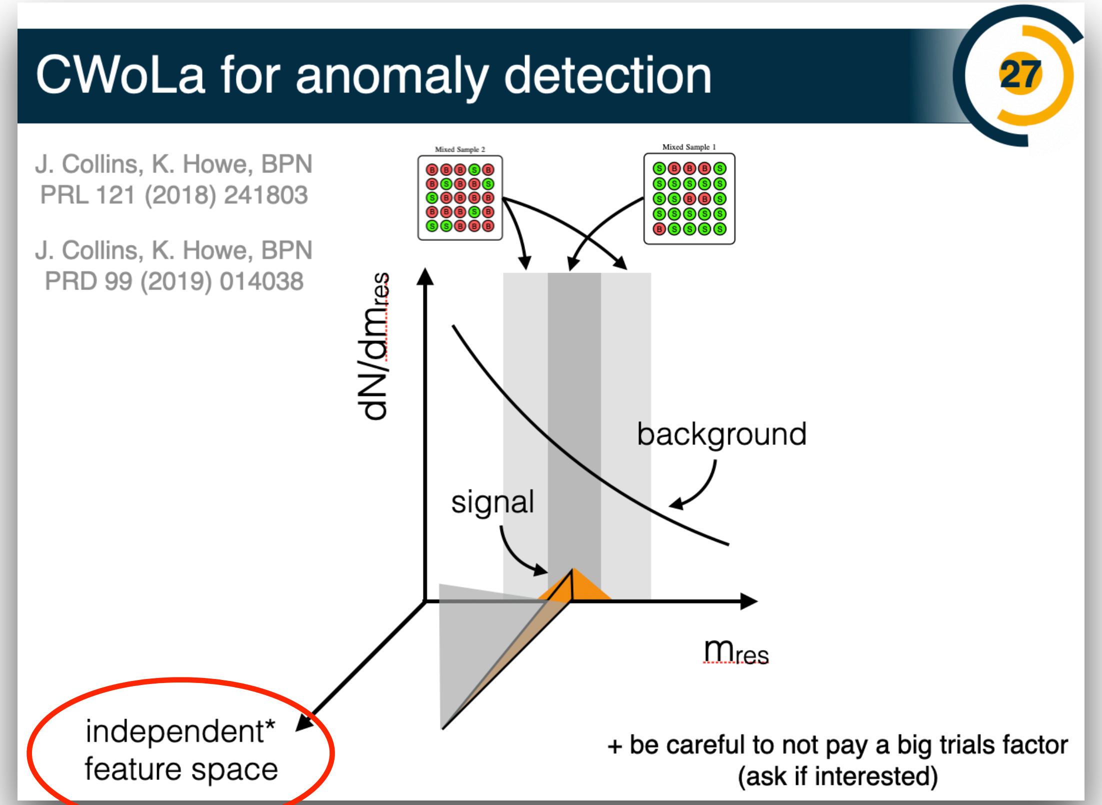
picture from Aviv Cukierman

Computational Challenges

For the first analysis, we used a low-dimensional feature space and so a few days of a generic CPU batch cluster was sufficient.

In the future: we want to increase the number of features and explore many more regions of the parameter space. We will need massive (GPU) compute (!) … fortunately, this is possible with HPCs.

The other challenge of higher-dimensions



For many anomaly detection methods, we need that the classifier does not introduce artificial bumps

There are basically two ways to solve this:

(1) Preprocessing (use features that are ~independent of mass)

(2) Use a training procedure that is robust to correlations

See e.g. K. Benkendorfer, L. Le Pottier, BPN, 2009.02205

Conclusions and outlook

Deep-learning based anomaly detection has a great potential for discovery!

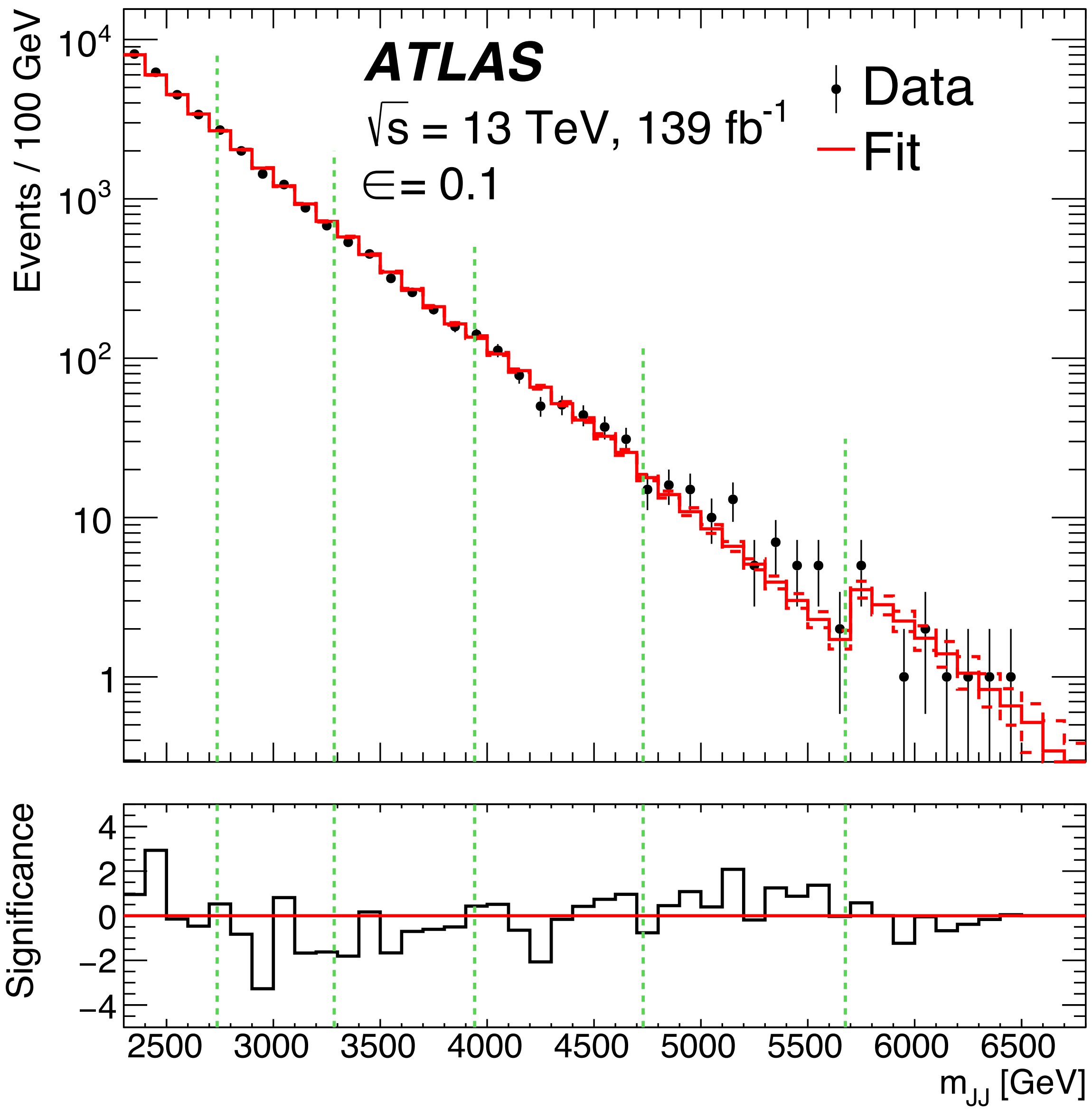
Check out the LHC Olympics: a community challenge for comparing anomaly detection techniques

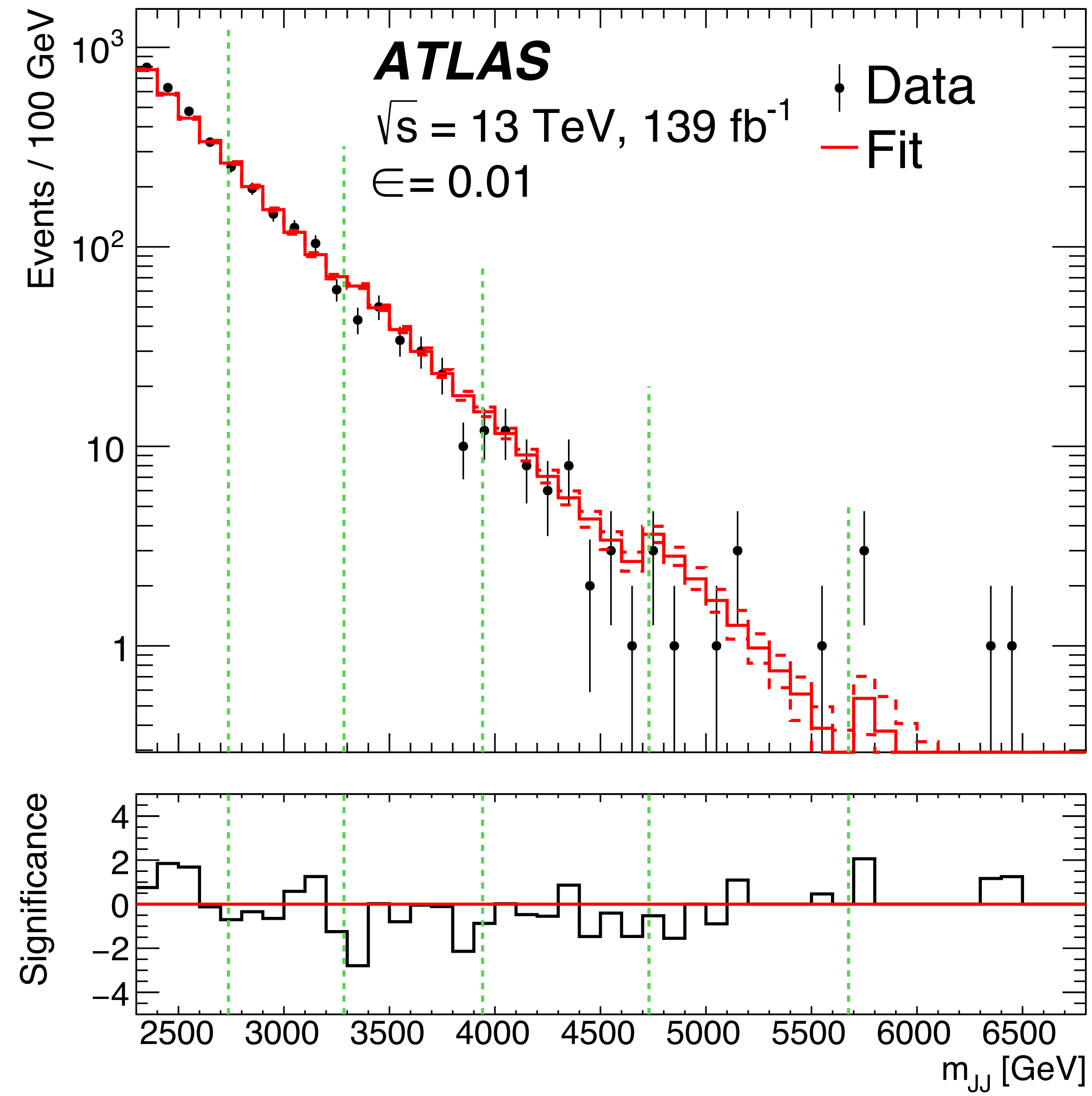
https://lhco2020.github.io/homepage/

As we use more complex less-than-supervised approaches, we will need significant computing power to ensure this program is successful.

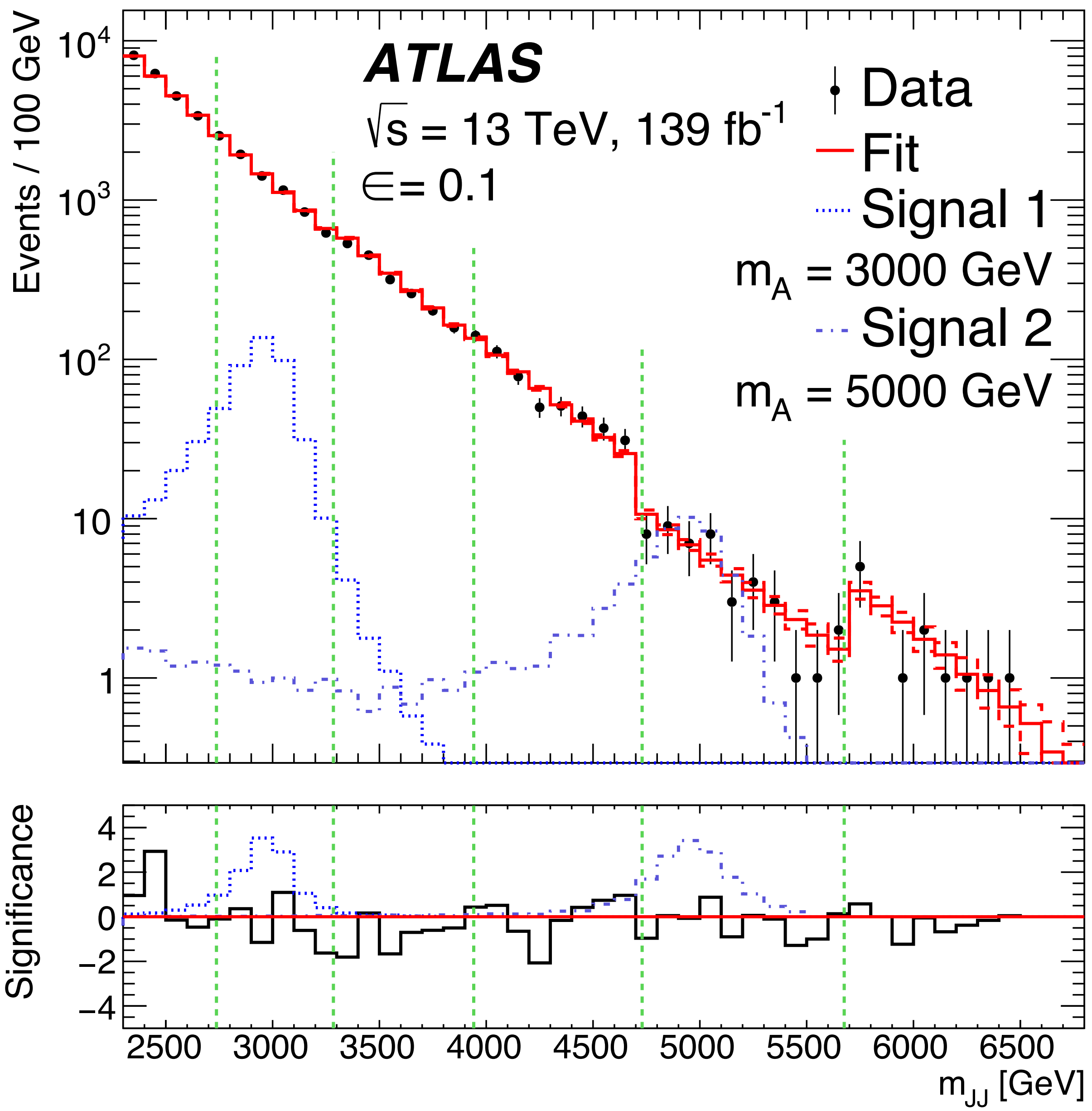
Backup

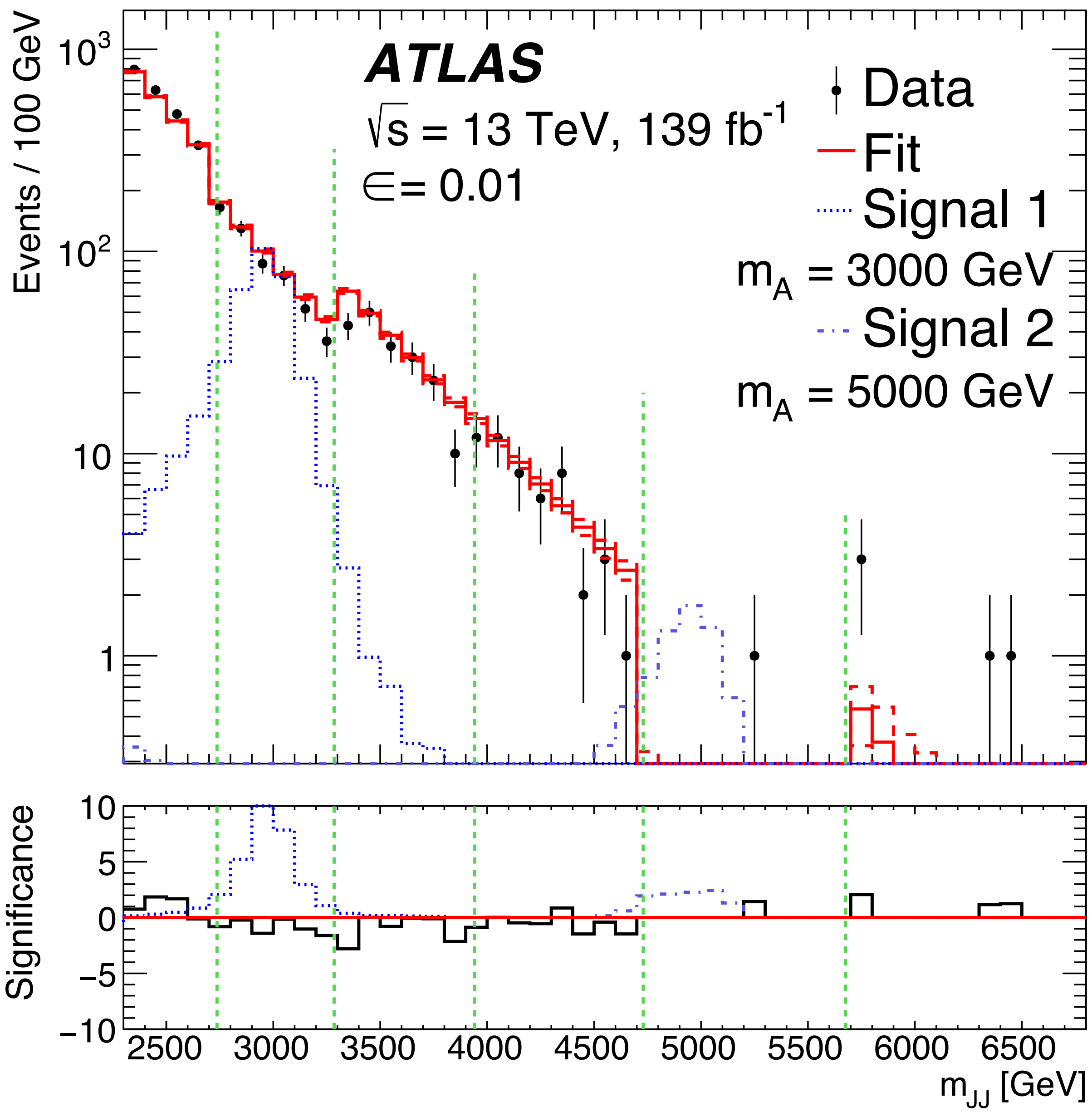
All signal regions





All signal regions





All signal regions