Machine Learning in Particle Physics

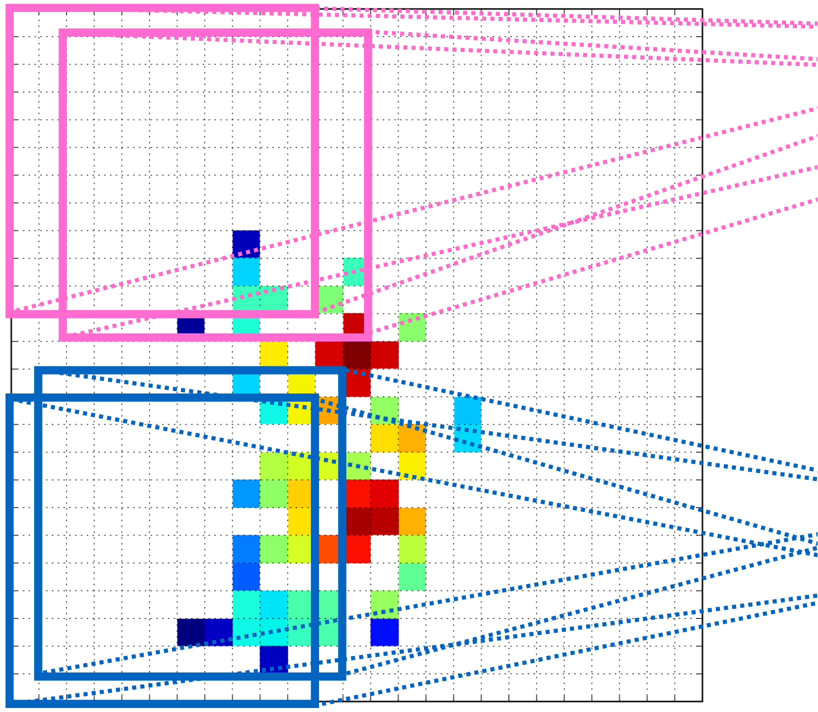
Benjamin Nachman

Lawrence Berkeley National Laboratory

bpnachman.com

bpnachman@lbl.gov

@bpnachman bnachman



UC Berkeley Physics

Physics 290E

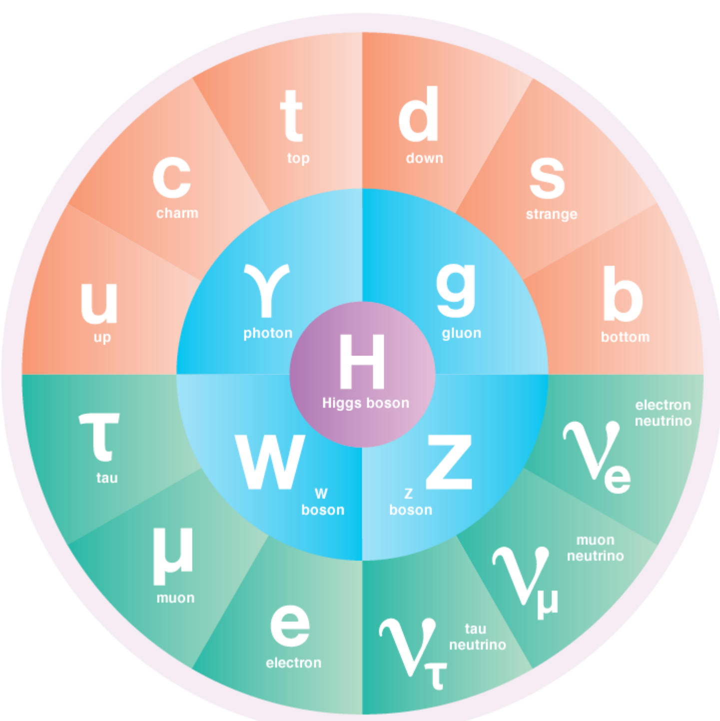
February 23, 2022

Questions in fundamental physics

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

Why is the Higgs boson so light?

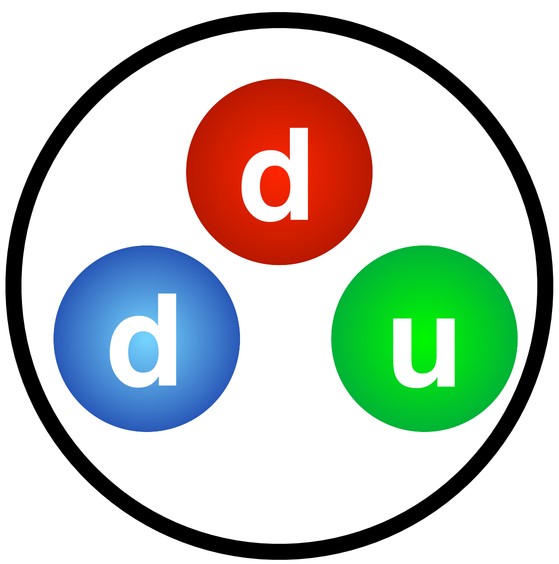
Hierarchy problem



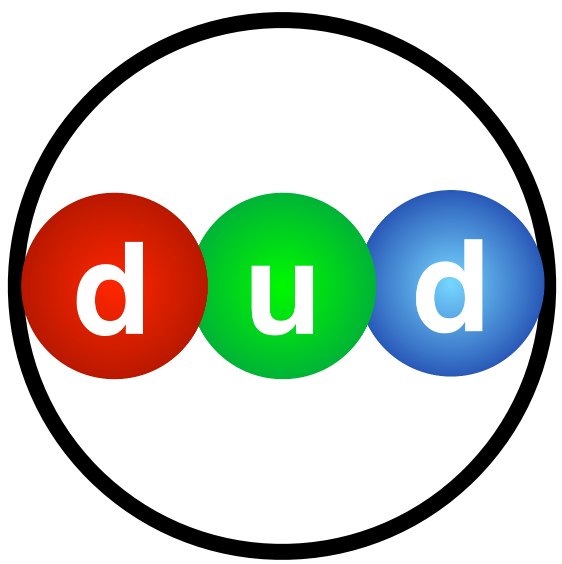
See also: quantum gravity

Why do neutrons have no dipole moment?

Strong CP



>99% of pictures on the internet



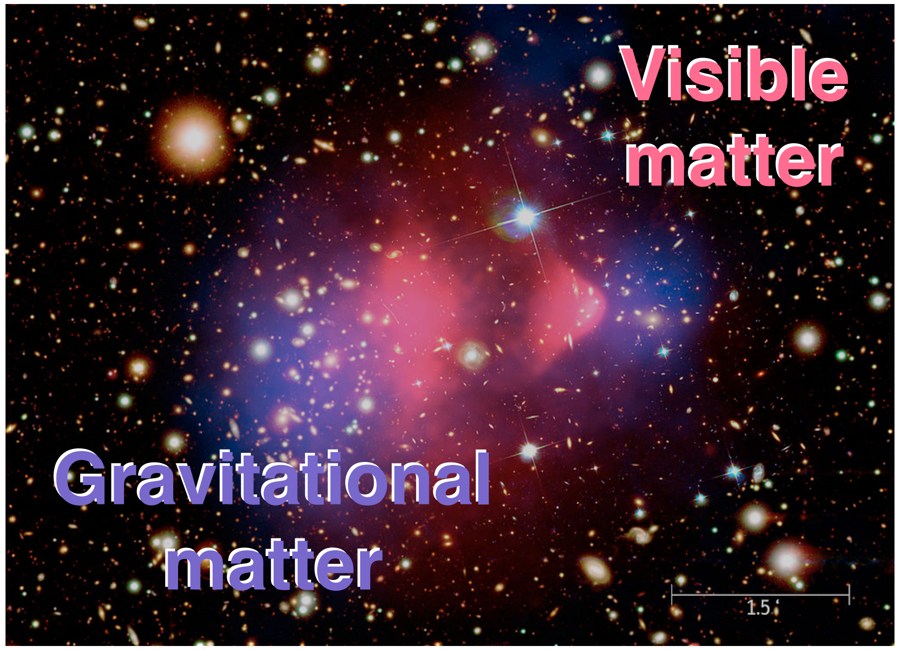
Reality

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

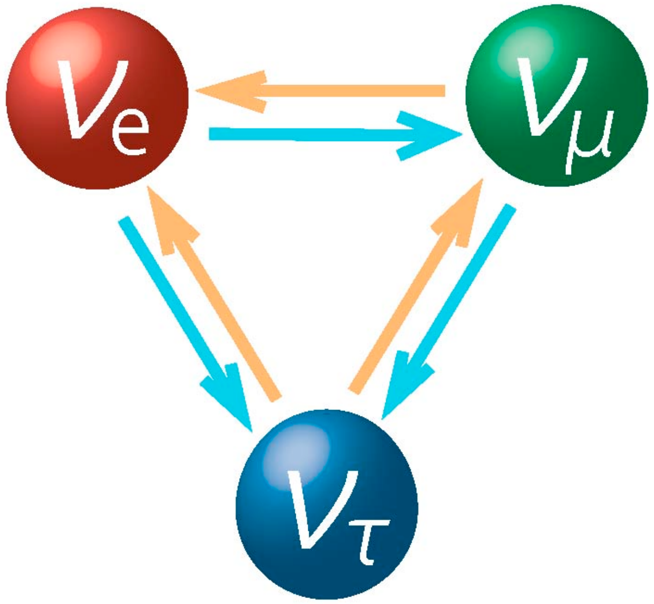


image source: T2K collaboration

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

Questions in fundamental physics

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

Dark matter

Strong CP

Hierarchy problem

Flavor puzzles

Baryogengesis

Dark energy

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

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

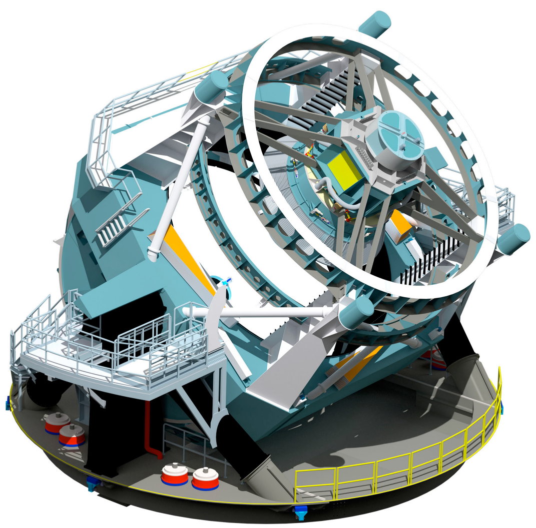
Baryogengesis

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

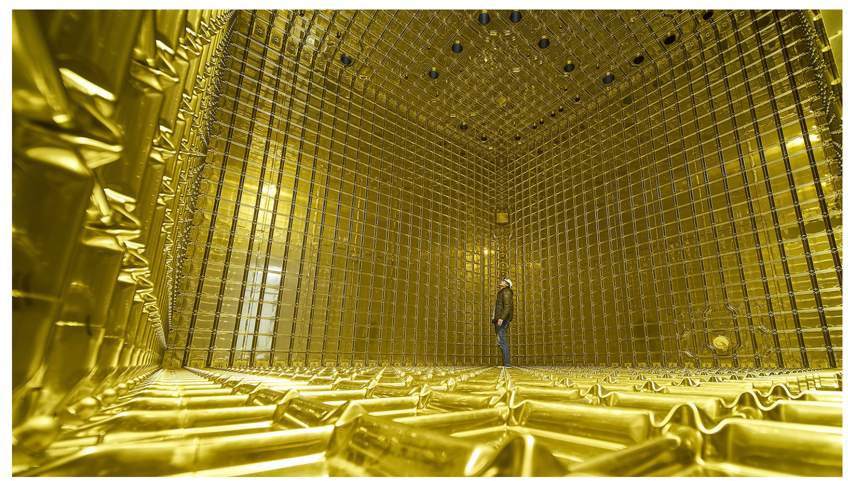
We will need new tools to explore our data in new ways!

New tools: detectors

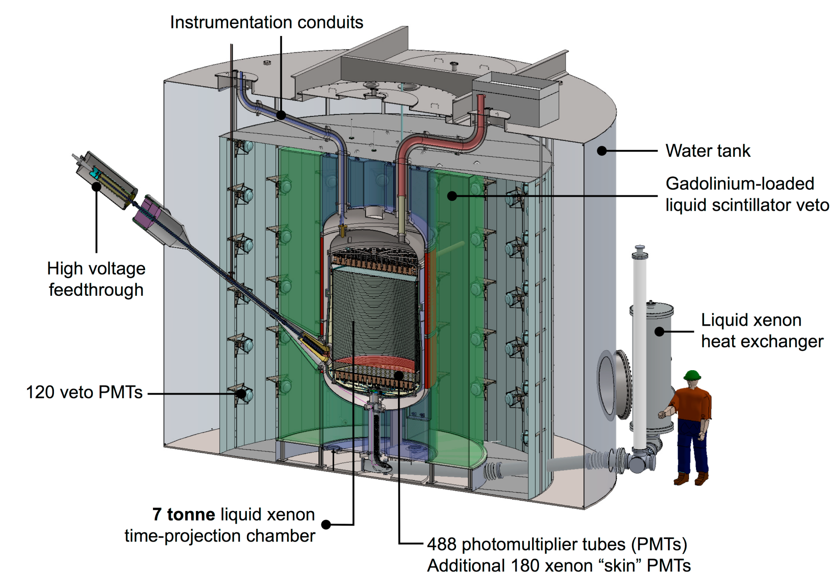
Dark matter/energy with LSST



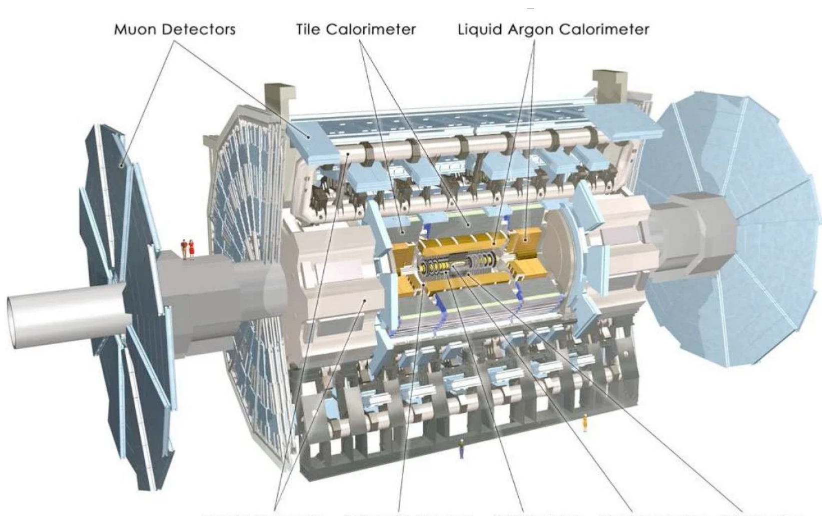
Fermilab neutrino experiments



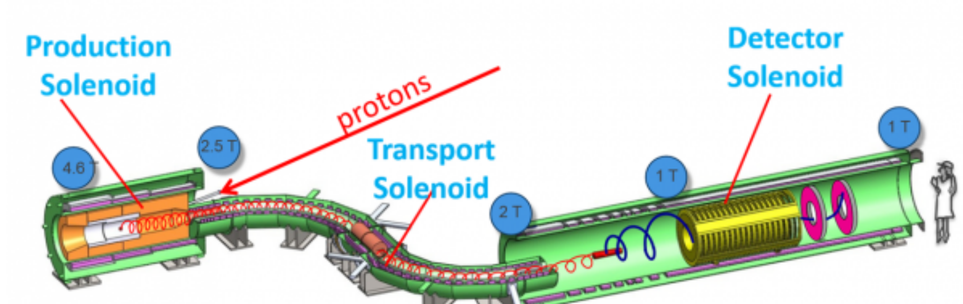
Dark Matter with LZ



Large Hadron Collider



Mu2e experiment

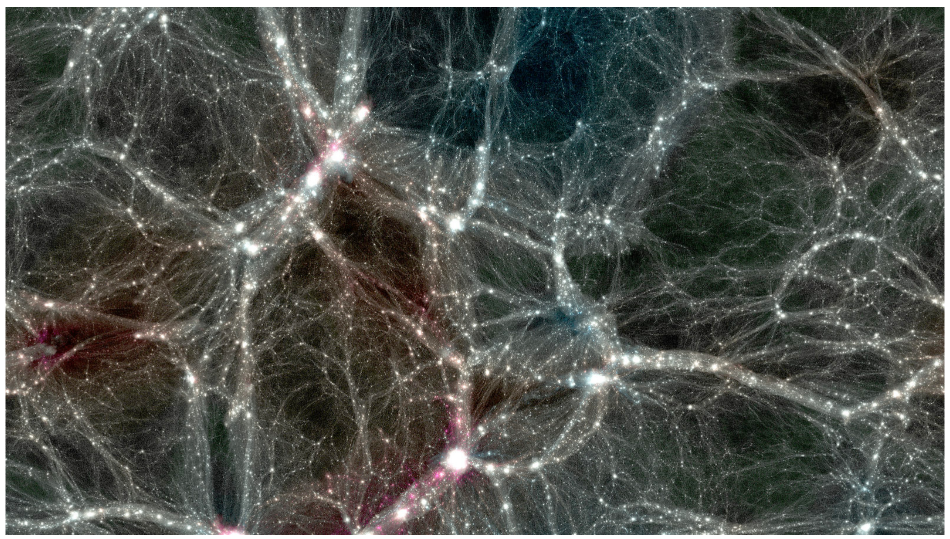


+ others !

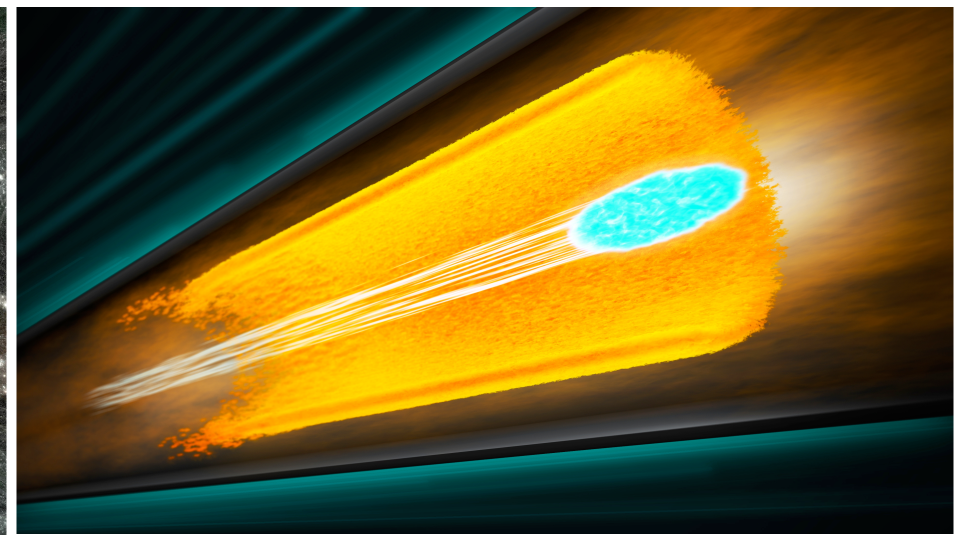
Image sources: LSST Corporation, CERN courier (CERN-201710-248-3), Berkeley Lab, CERN, HPS Collaboration

New tools: methodology

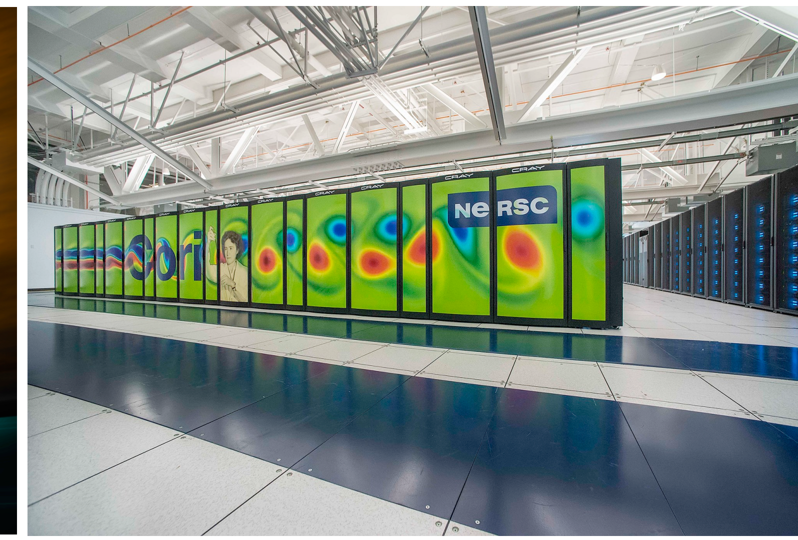
N-body simulations



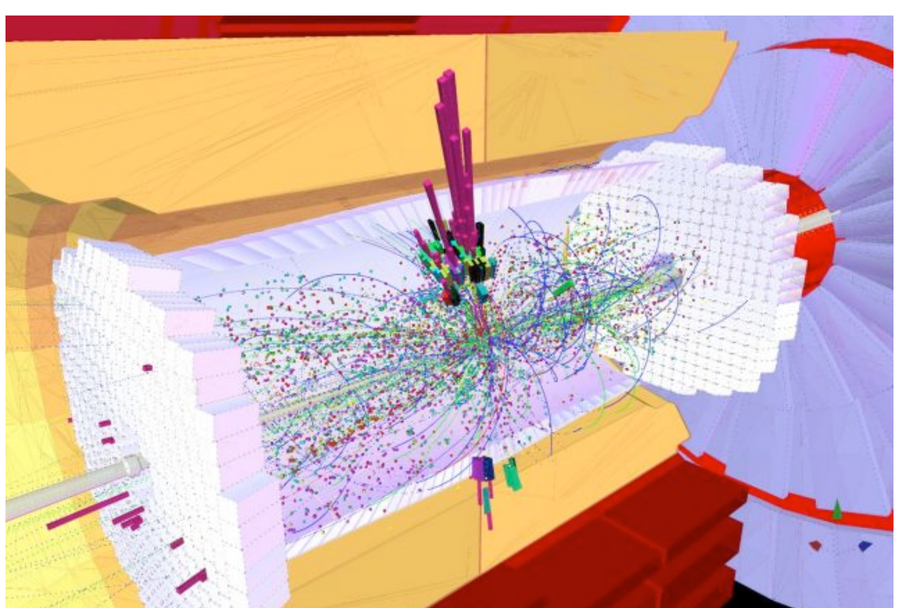
Advanced accelerators



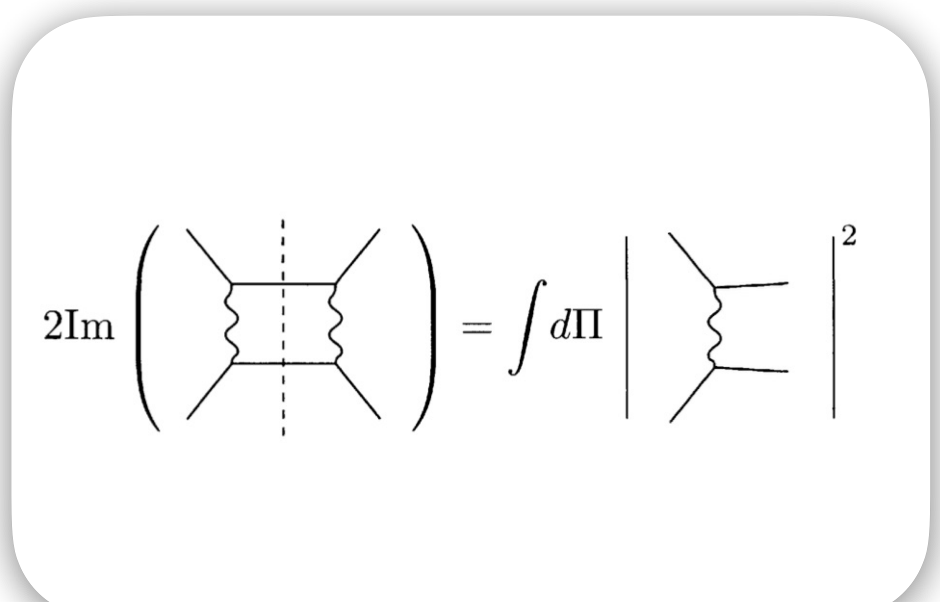
Supercomputers



Material Simulations



Theory Calculations

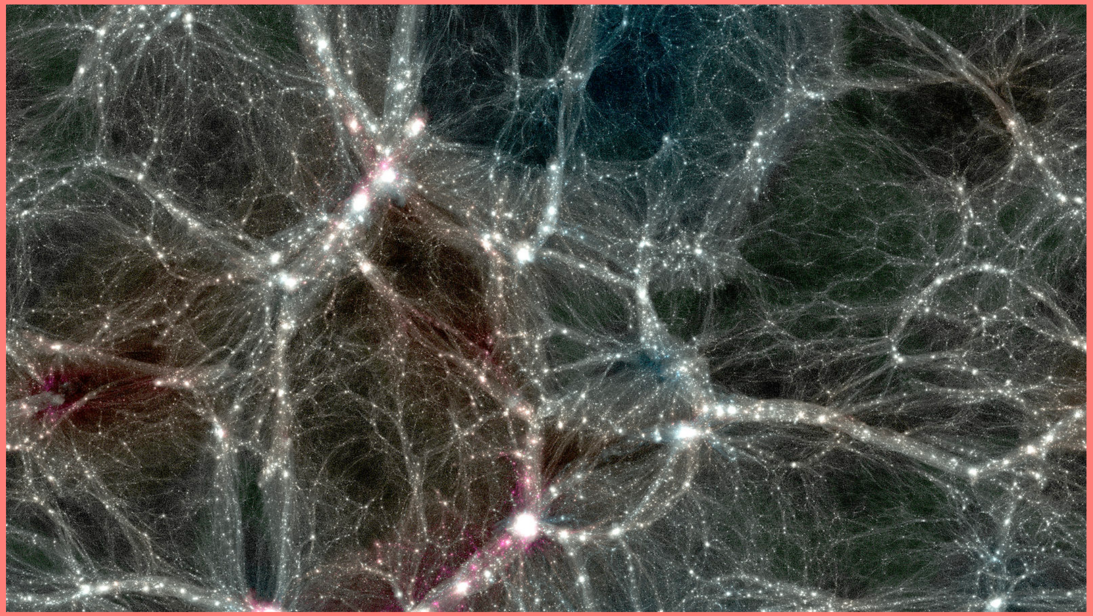


+ others !

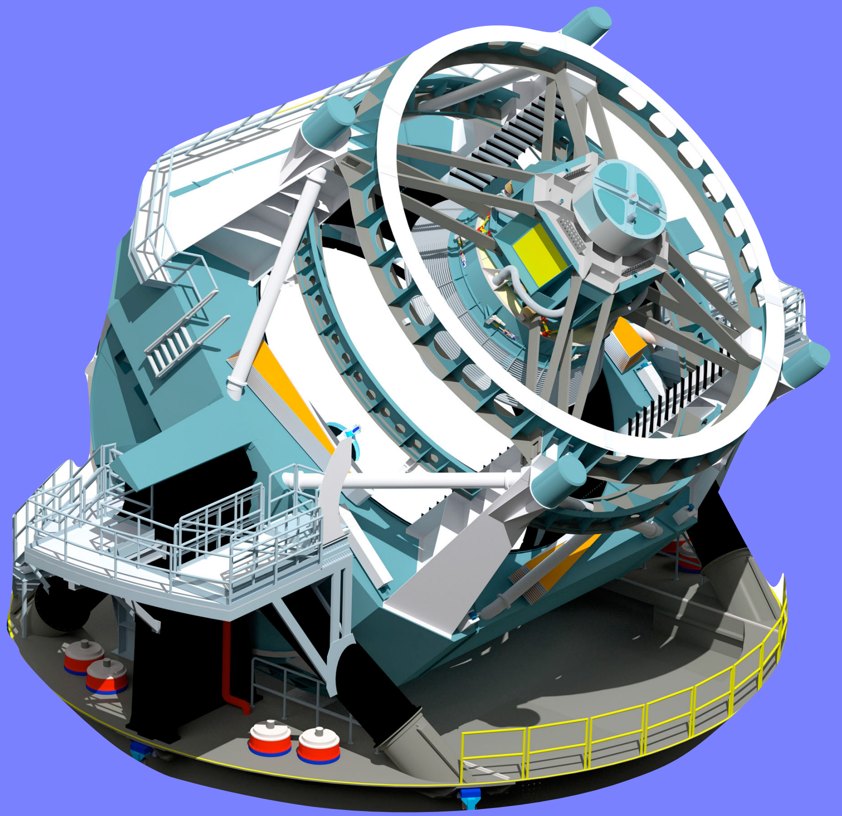
Image sources: Dark Sky Simulations collaboration, SLAC, NERSC, Fermilab Today / Geant4, Peskin and Schroeder

A *hyper* challenge

Key **challenge** and **opportunity**: *hypervariate phase space* & *hyper spectral data*



Methodology



Detectors

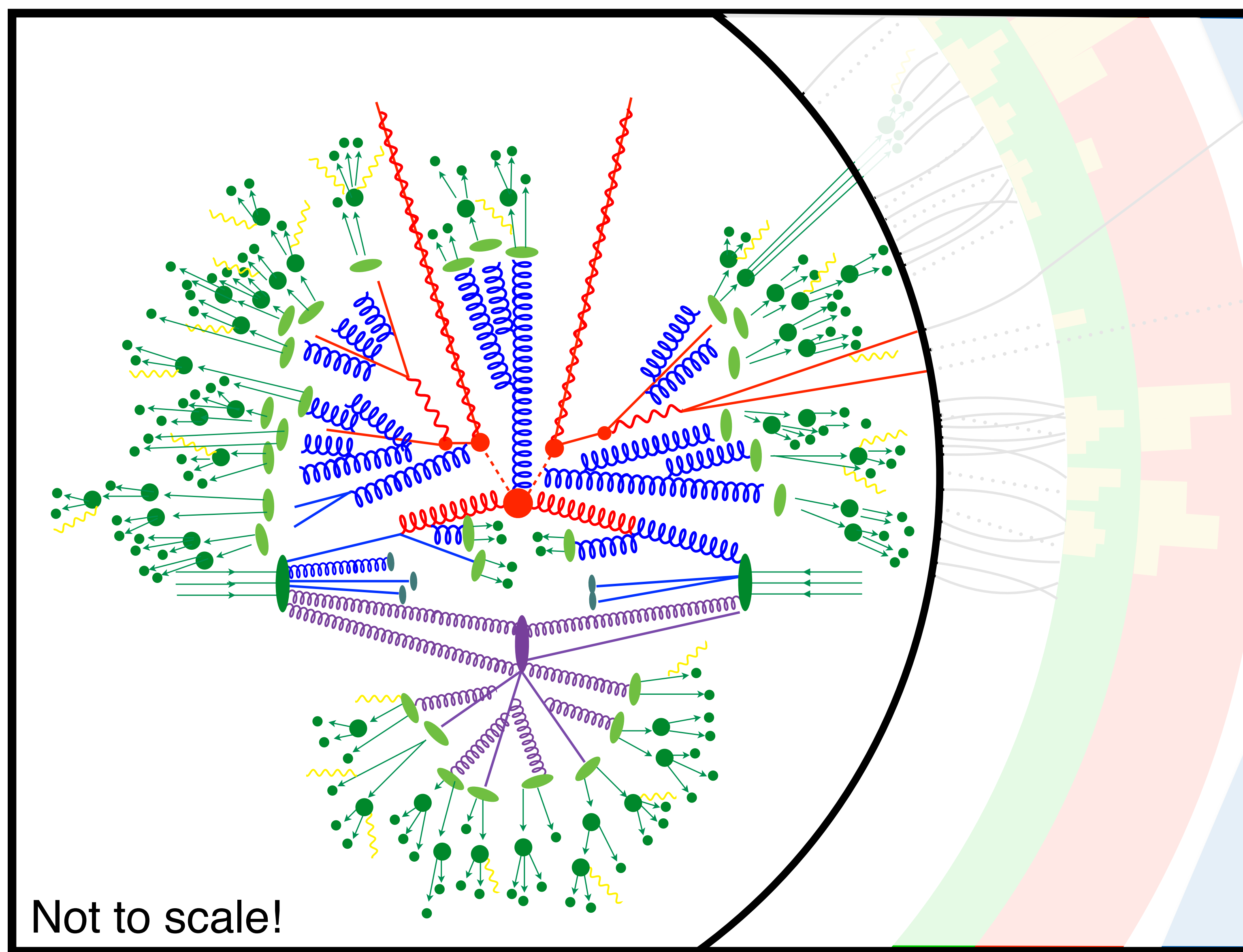
A *hyper* challenge

Key **challenge** and **opportunity**: *hypervariate phase space* & *hyper spectral data*

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



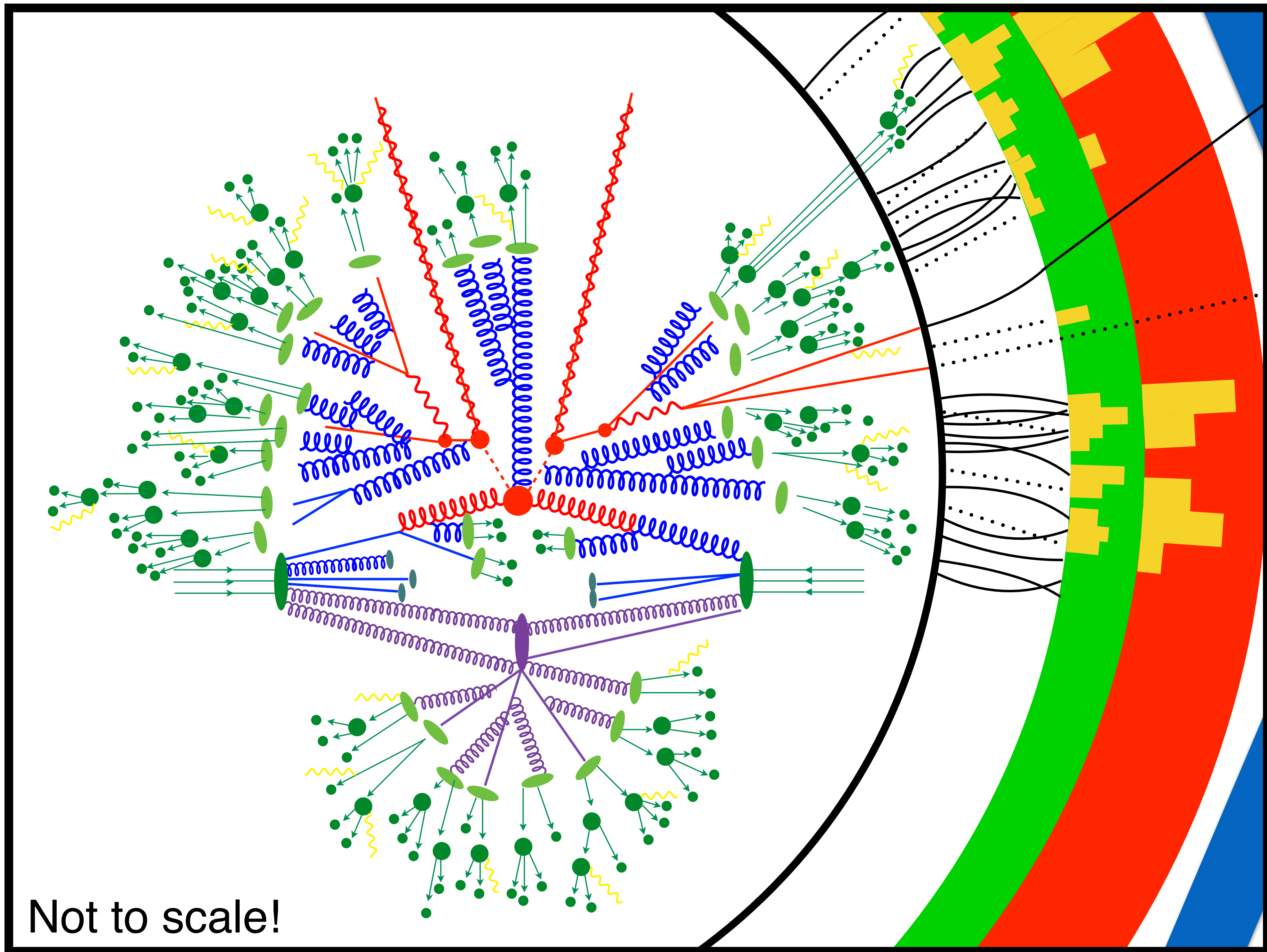
A *hyper* challenge

Key **challenge** and **opportunity**: *hypervariate phase space* & *hyper spectral data*

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

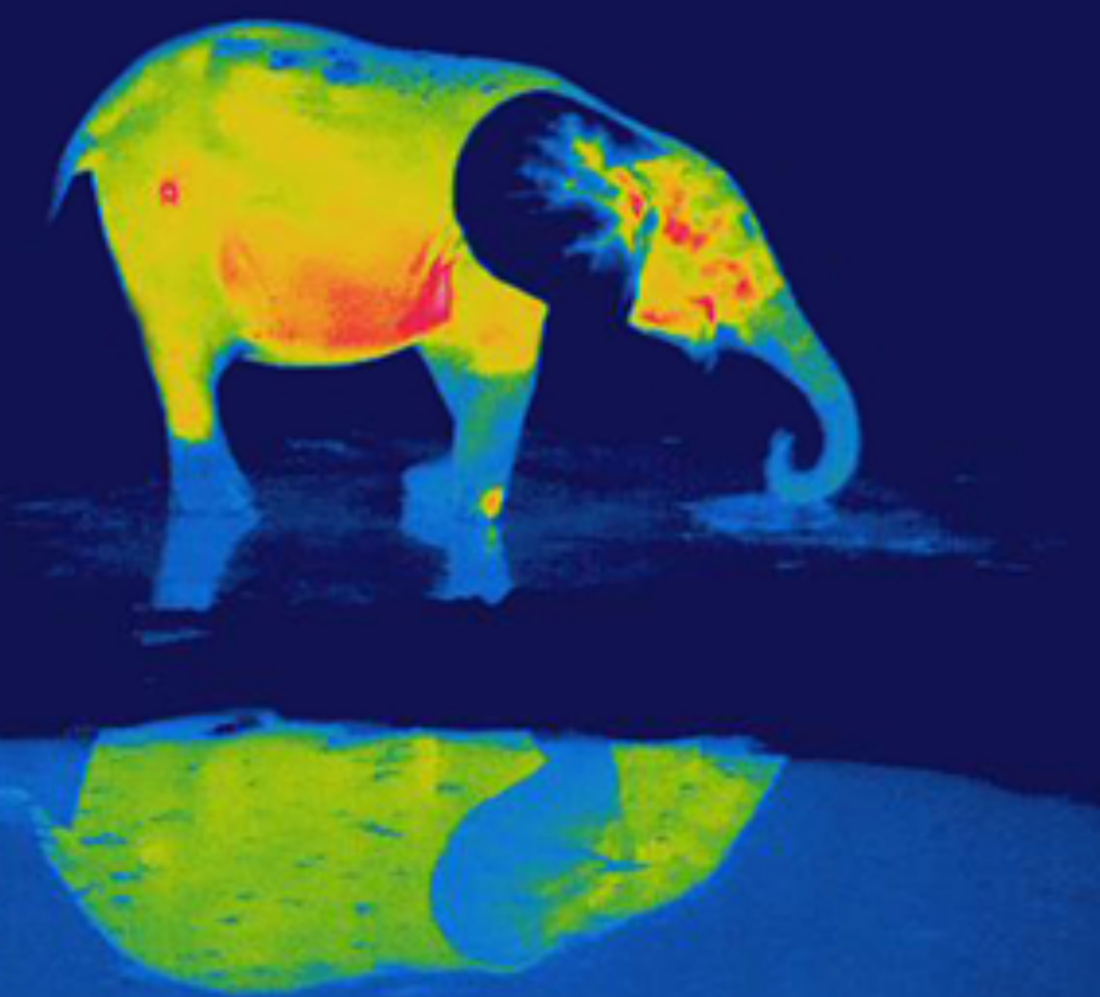


Hypervariate vision with deep learning

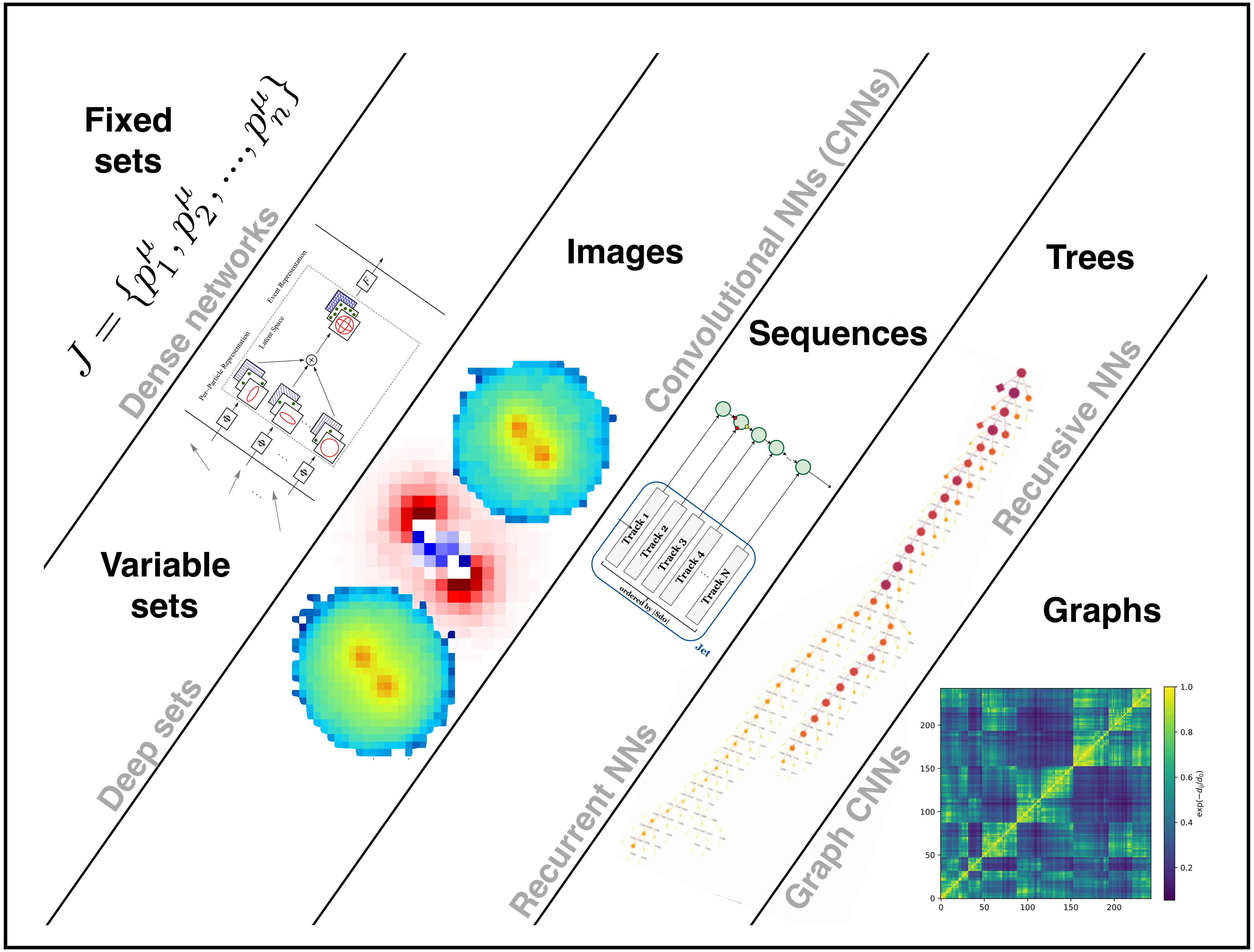
We have been conducting "multivariate" analysis of collision events for many years

Everyone is aware that there must be new physics, but maybe we need hypervariate vision to see it?

However, recent advances have opened up a **new way** of looking at our data. This **hypervariate vision** will lead to a deeper understanding of nature and perhaps surprises along the way…



Representing our data

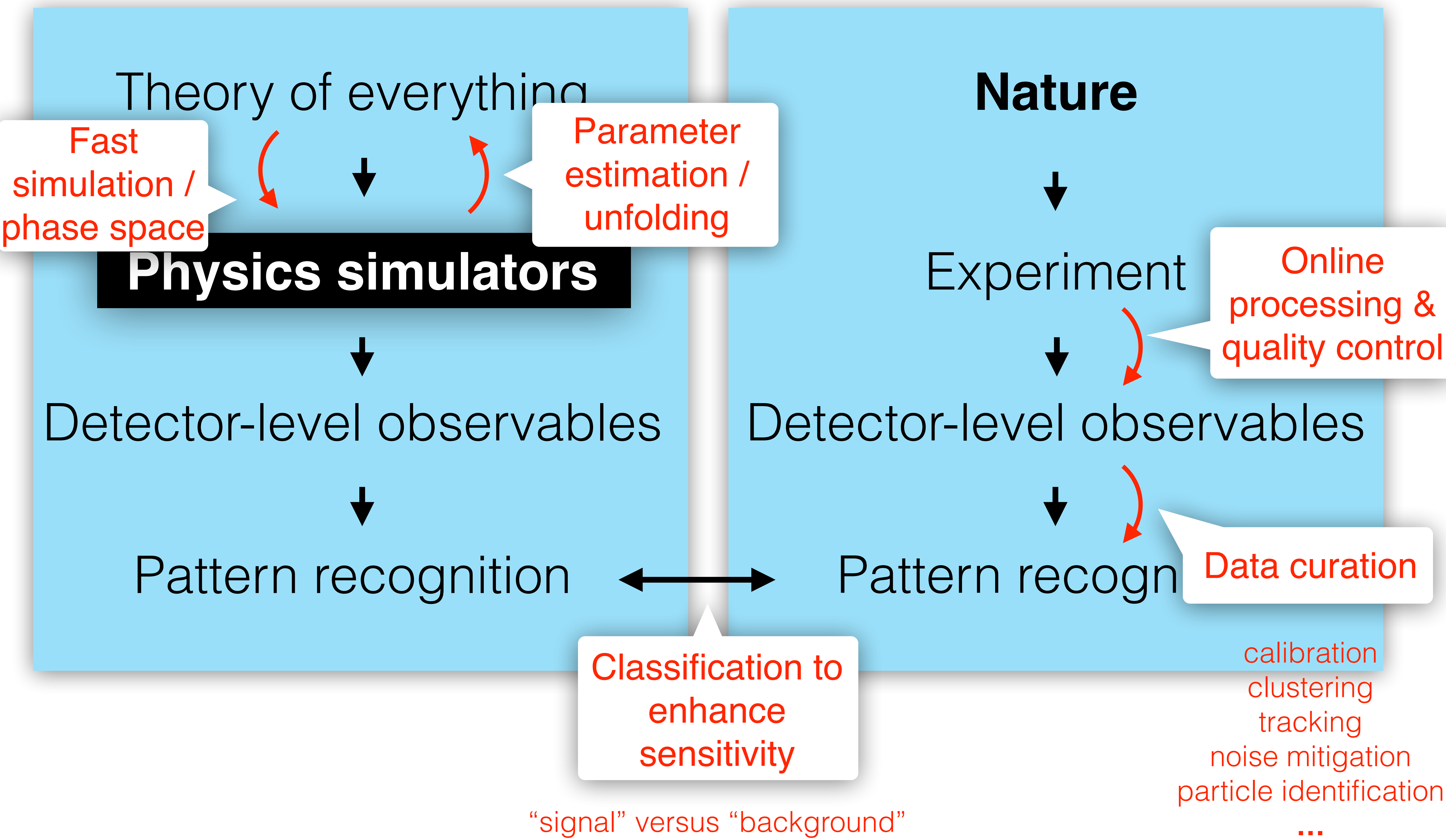


1709.04464, with images from 1810.05165; ATLAS, PUB-2017-003; 1711.02633; Henrion et al. MLPS @ NeurIPS 2017

Data analysis in particle physics

Theory of everything	**Nature**
↓	↓
Physics simulators	Experiment
↓	↓
Detector-level observables	Detector-level observables
↓	↓
Pattern recognition ⟷	Pattern recognition

Data analysis in particle physics + ML



Data analysis in particle physics

Theory of everything

↓

Physics simulators

↓

Detector-level observables

↓

Pattern recognition

↔

Nature

↓

Experiment

↓

Detector-level observables

↓

Pattern recognition

This is where most machine learning is being applied.

Data analysis in particle physics

Analysis
Observables

Calibration

Pattern
recognition

Noise
mitigation

Readout

Nature

↓

Experiment

↓

Detector-level observables

↓

Pattern recognition

This is where most machine learning is being applied.

Data analysis in particle physics

Analysis Observables

Calibration

Pattern recognition

Noise mitigation

Readout

Nature

↓

Experiment

↓

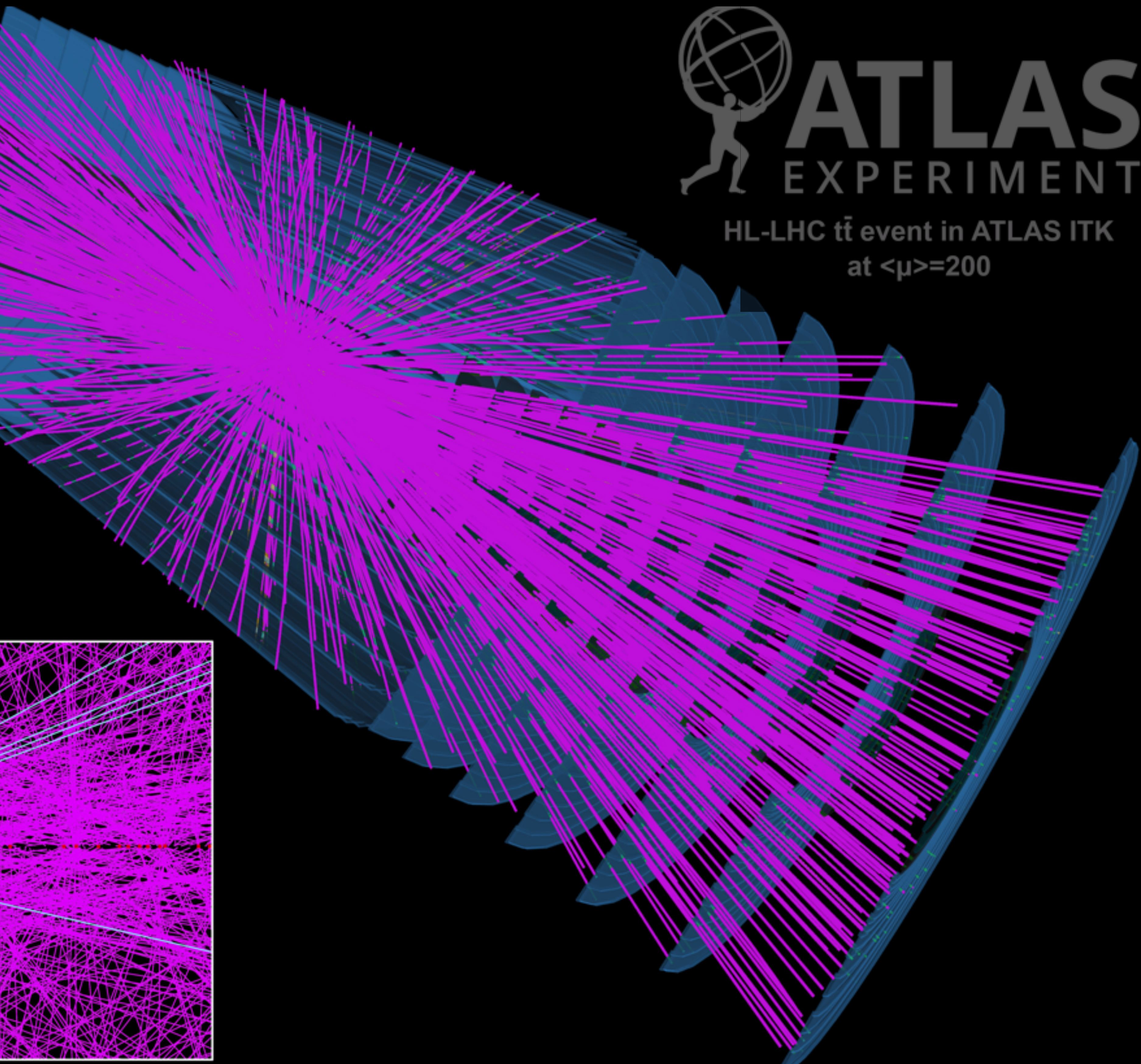
Detector-level observables

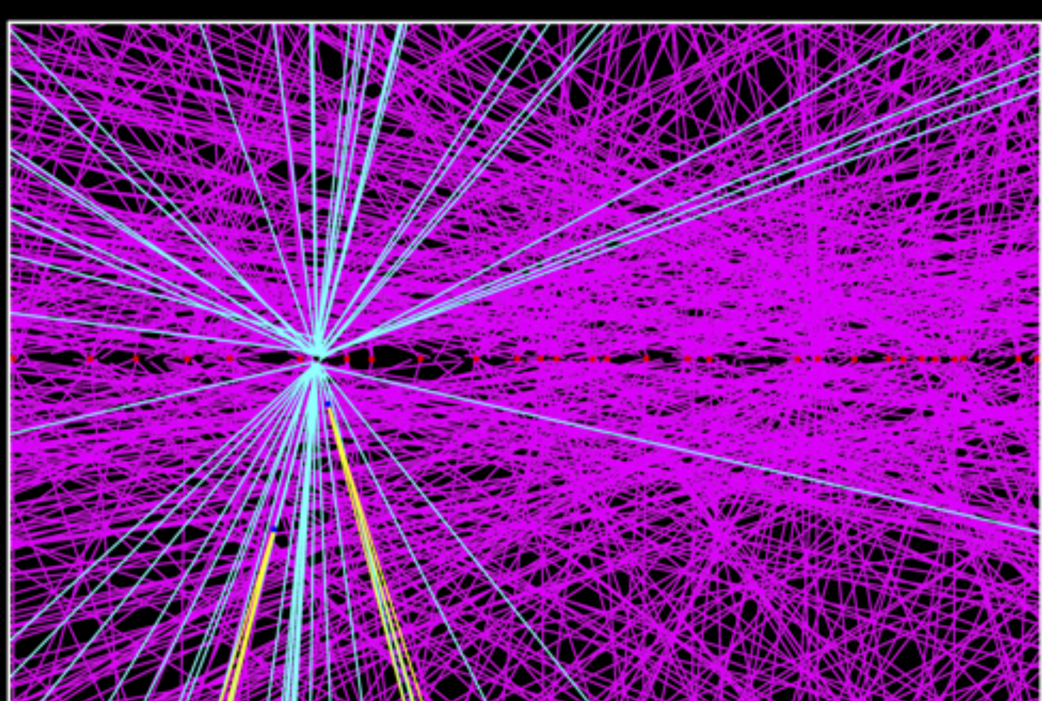
↓

Pattern recognition

This is where most machine learning is being applied.

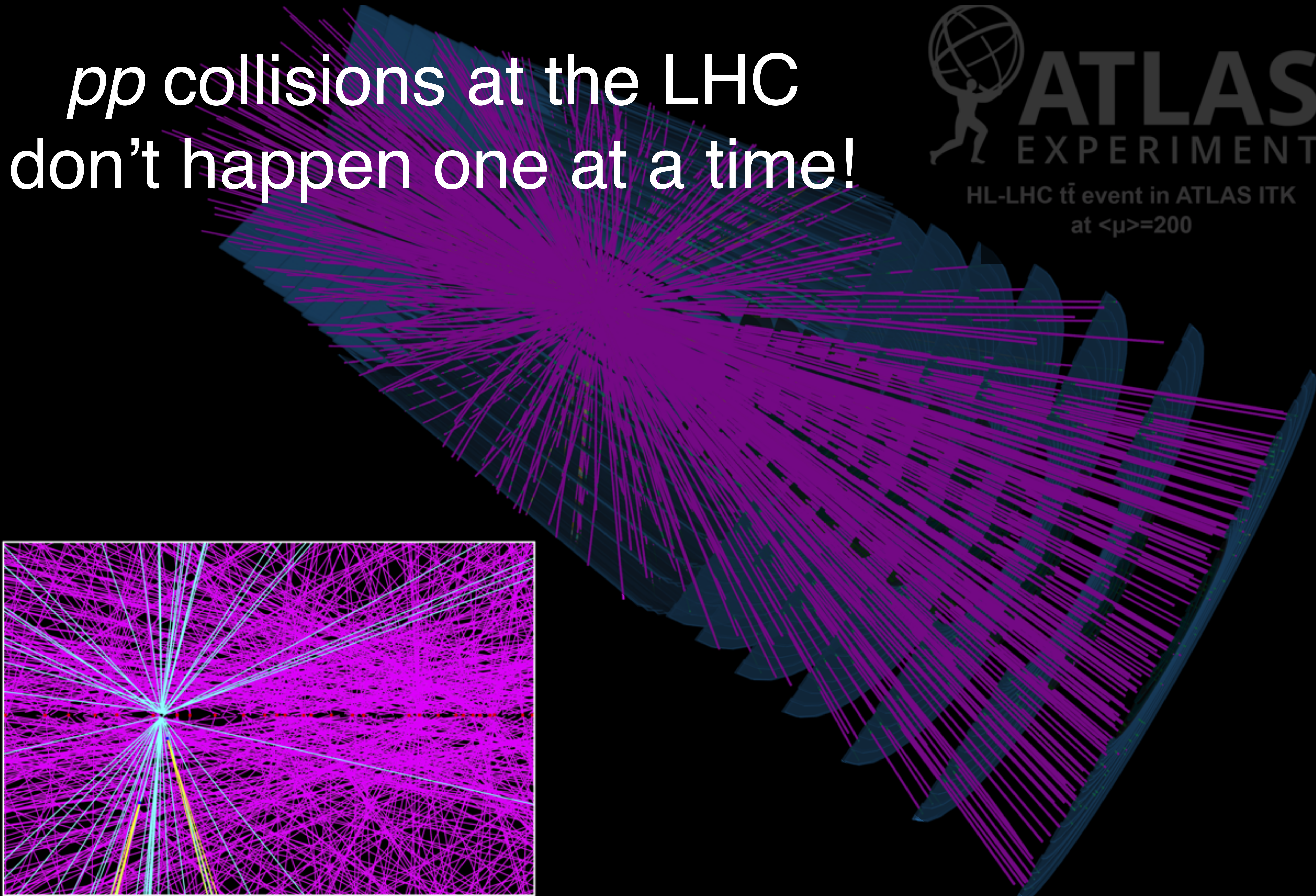
Example: Removing Noise





Example: Removing Noise

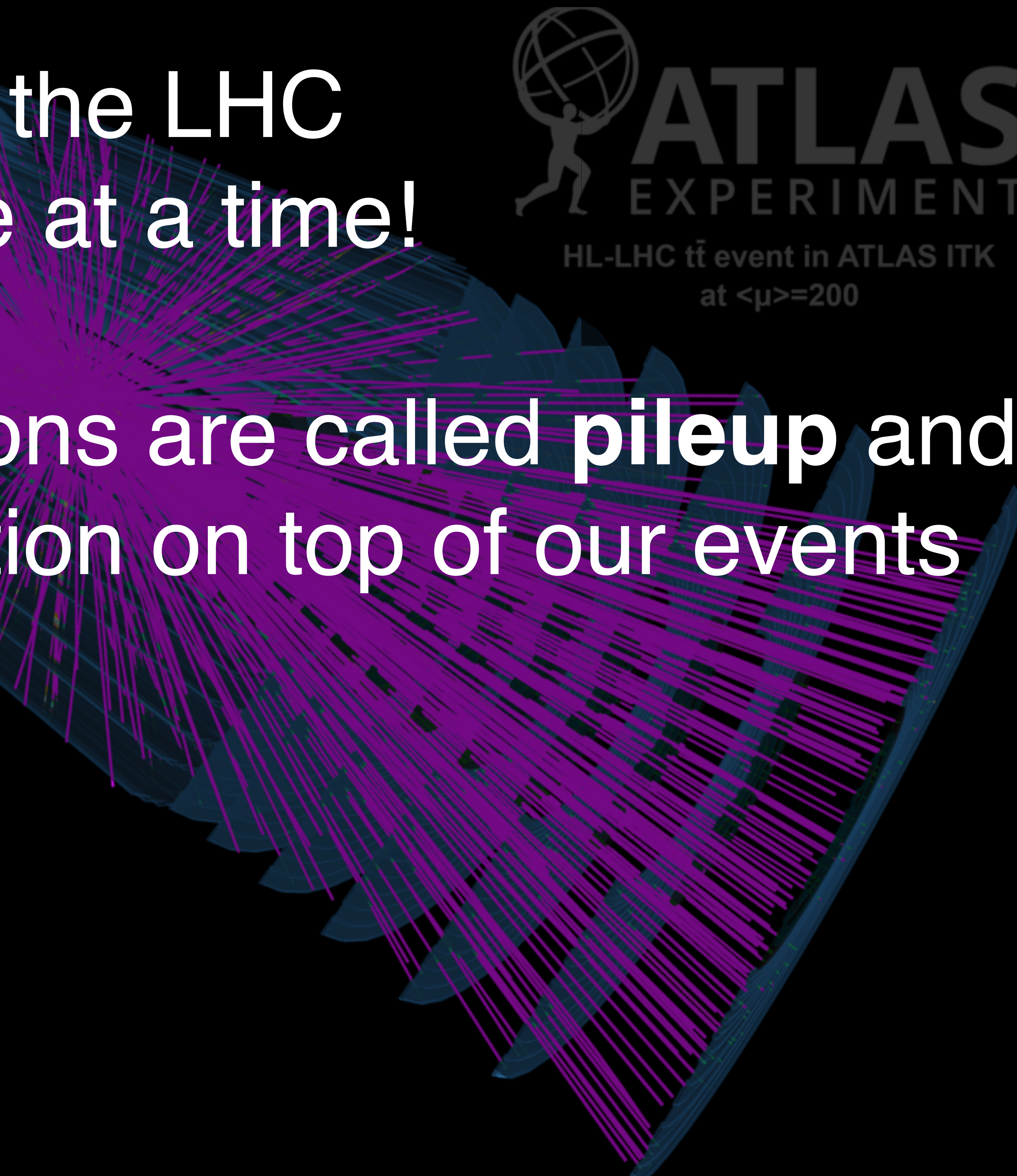
pp collisions at the LHC
don’t happen one at a time!

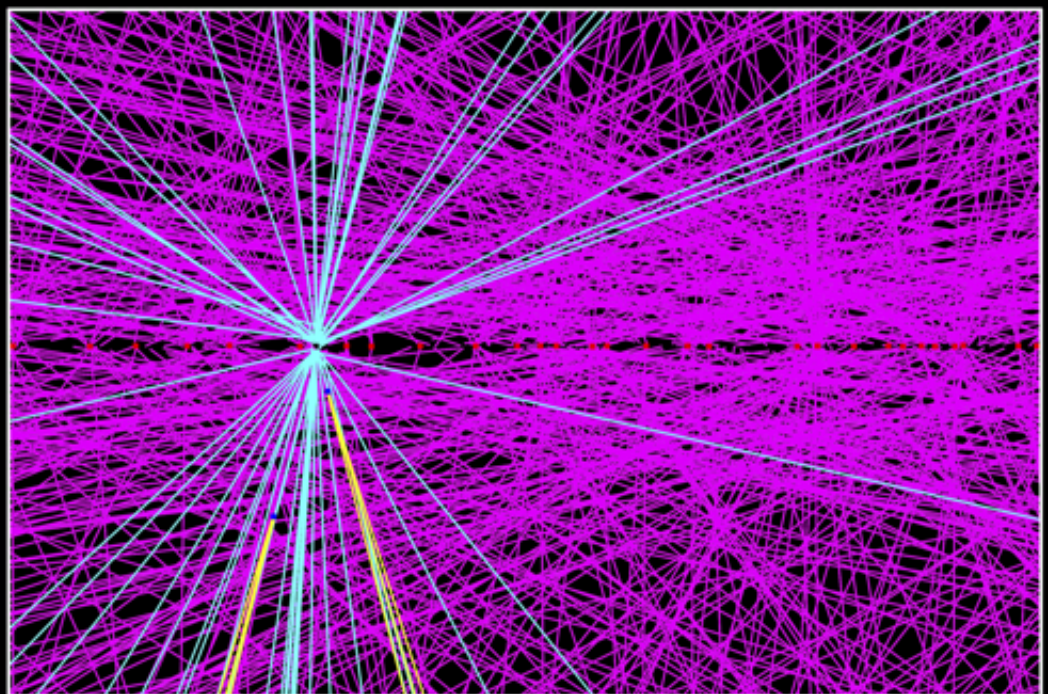


Example: Removing Noise

pp collisions at the LHC
don’t happen one at a time!

the extra collisions are called **pileup** and
add soft radiation on top of our events

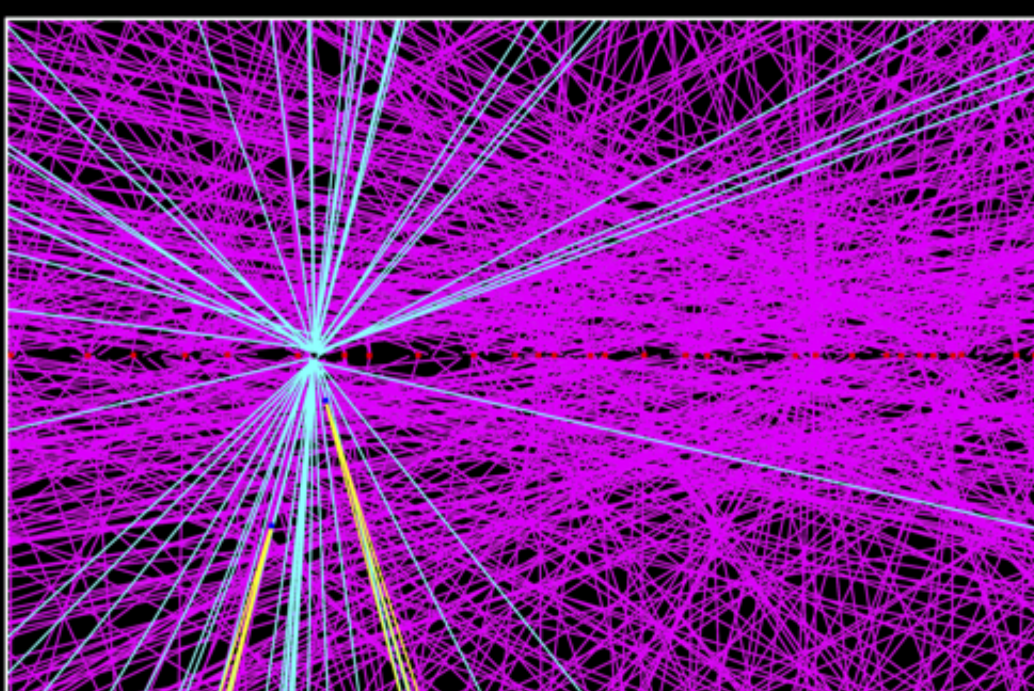




Example: Removing Noise

pp collisions at the LHC don’t happen one at a time!

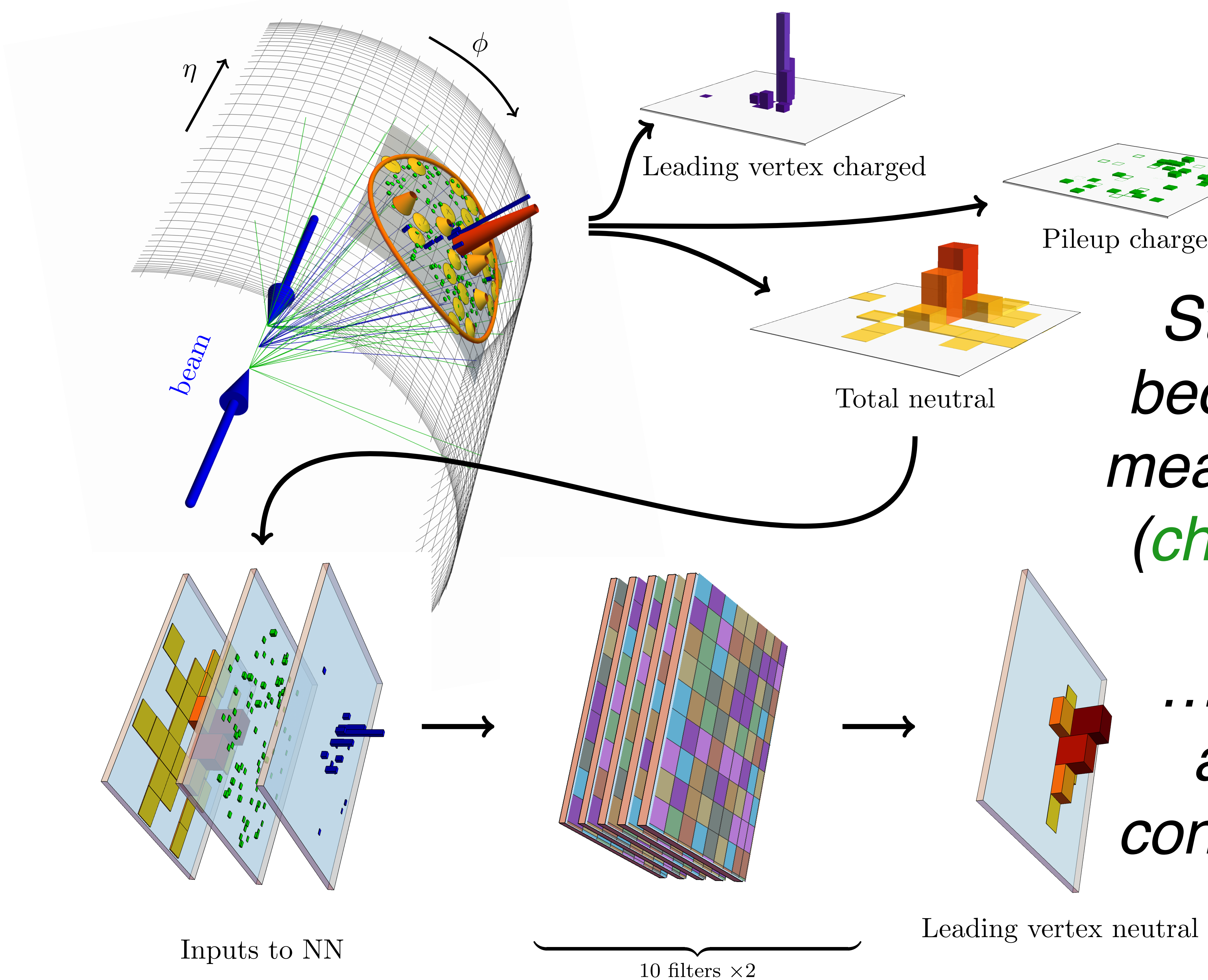
the extra collisions are called **pileup** and add soft radiation on top of our events



this is akin to image de-noising - we can use ML for that!

Example: Removing Noise

Image: Journal of High Energy Physics 12 (2017) 51



Strange noise because we can measure ~2/3 of it (charged pileup)

…also a natural application of convolutional NNs!

Data analysis in particle physics

Analysis
Observables

Calibration

Pattern
recognition

Noise
mitigation

Readout

Nature

↓

Experiment

↓

Detector-level observables

↓

Pattern recognition

This is where most machine learning is being applied.

Data analysis in particle physics

Analysis Observables

Calibration

Pattern recognition

Noise mitigation

Readout

Nature

↓

Experiment

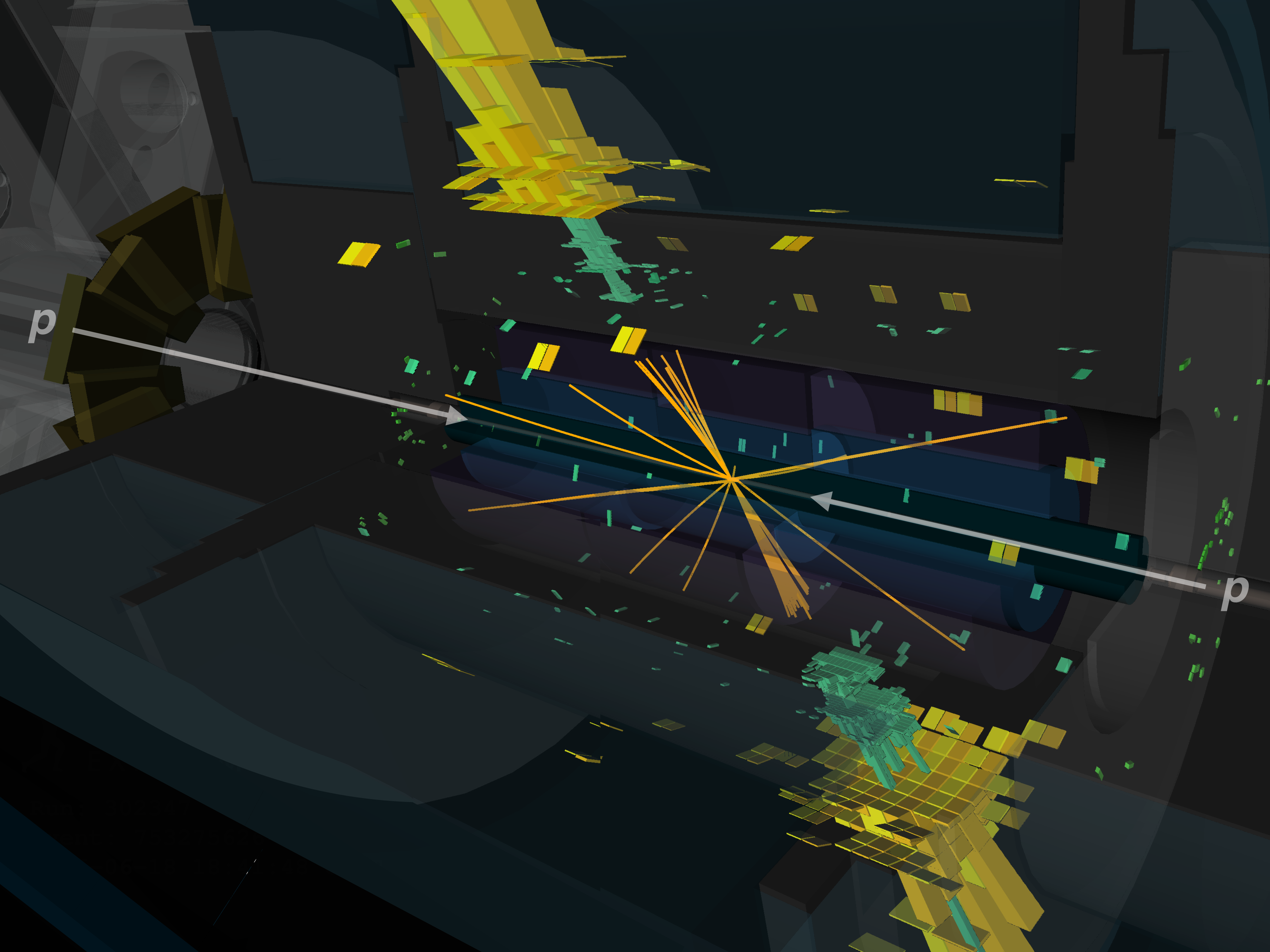
↓

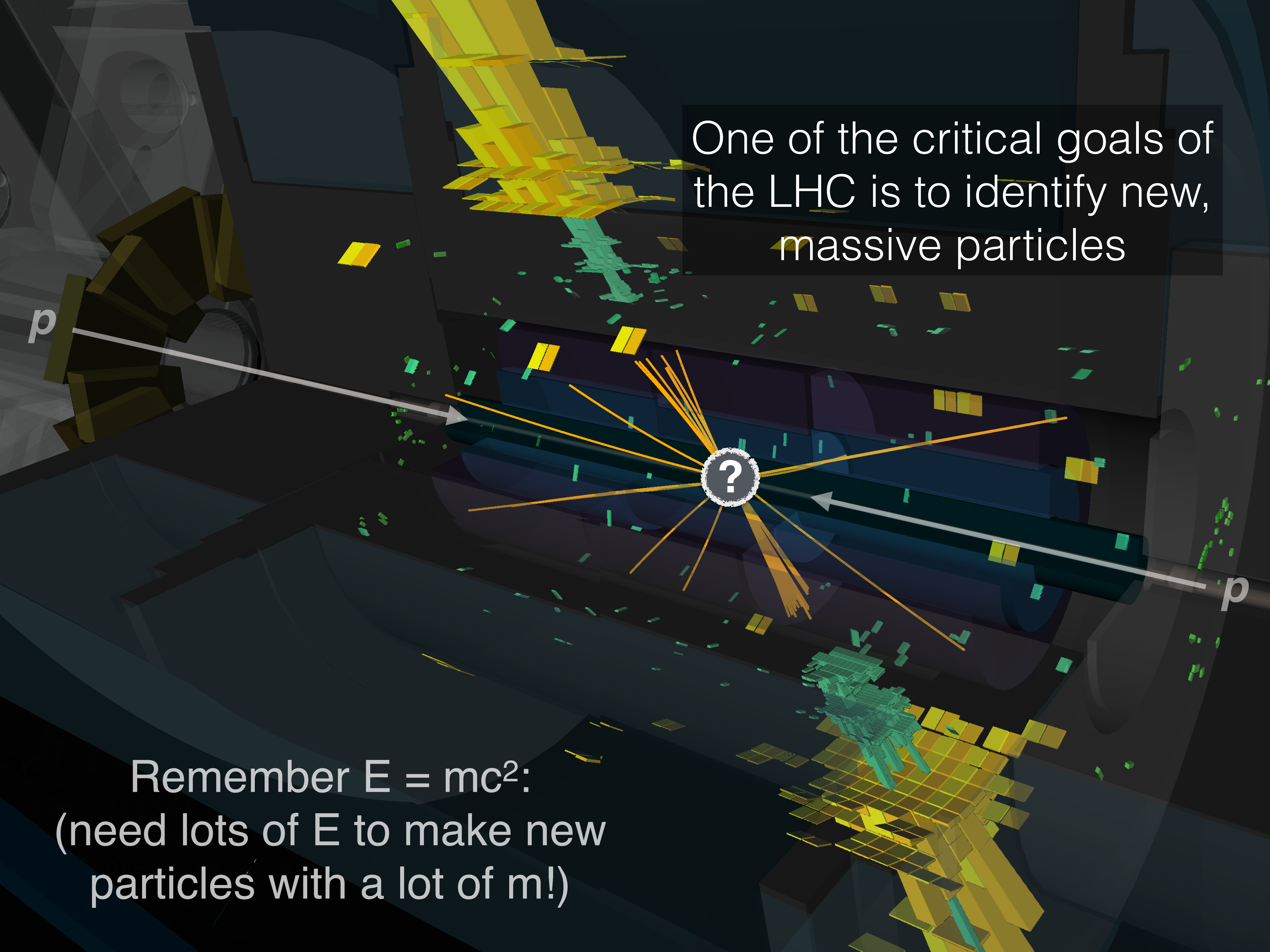
Detector-level observables

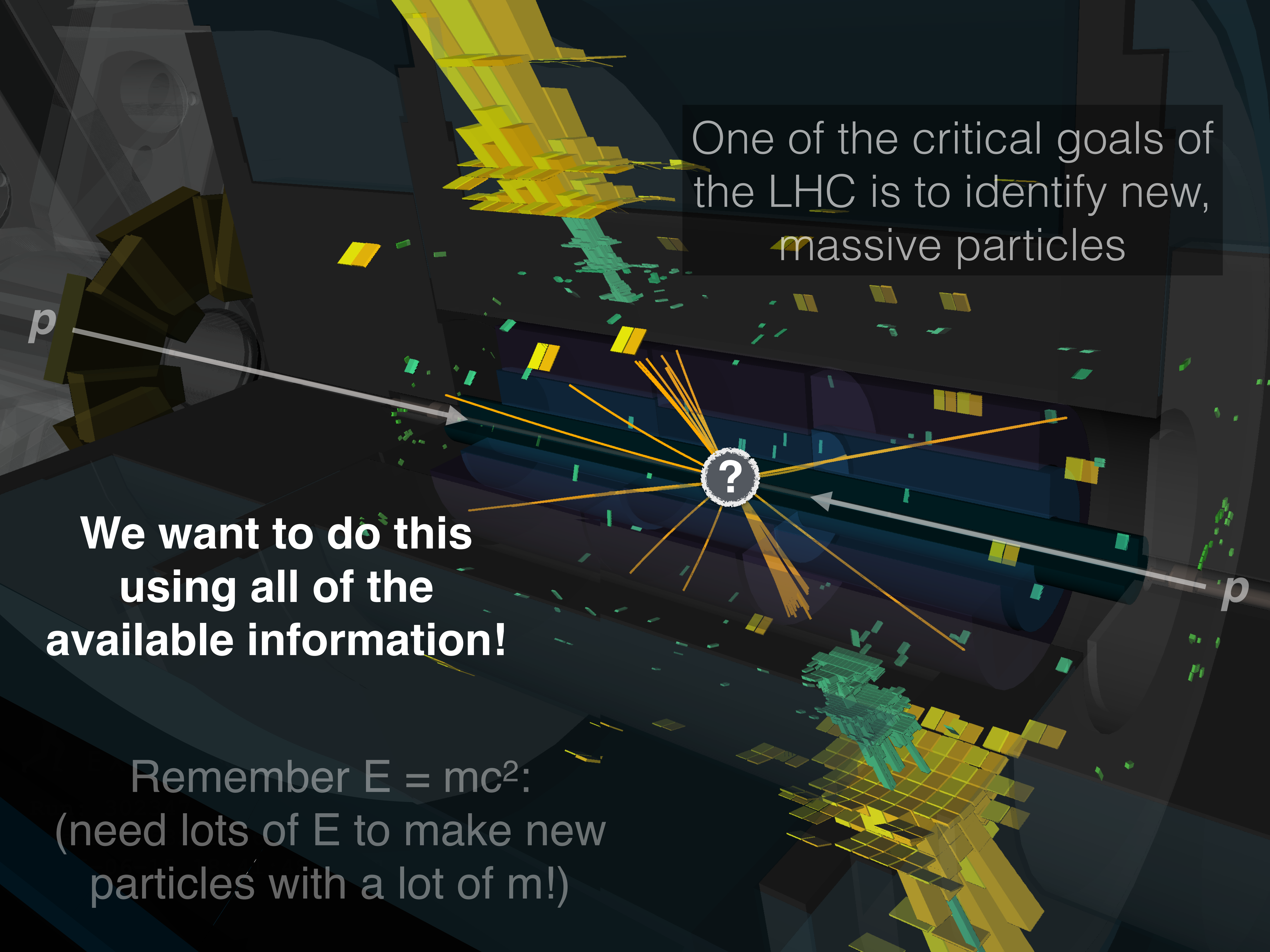
↓

Pattern recognition

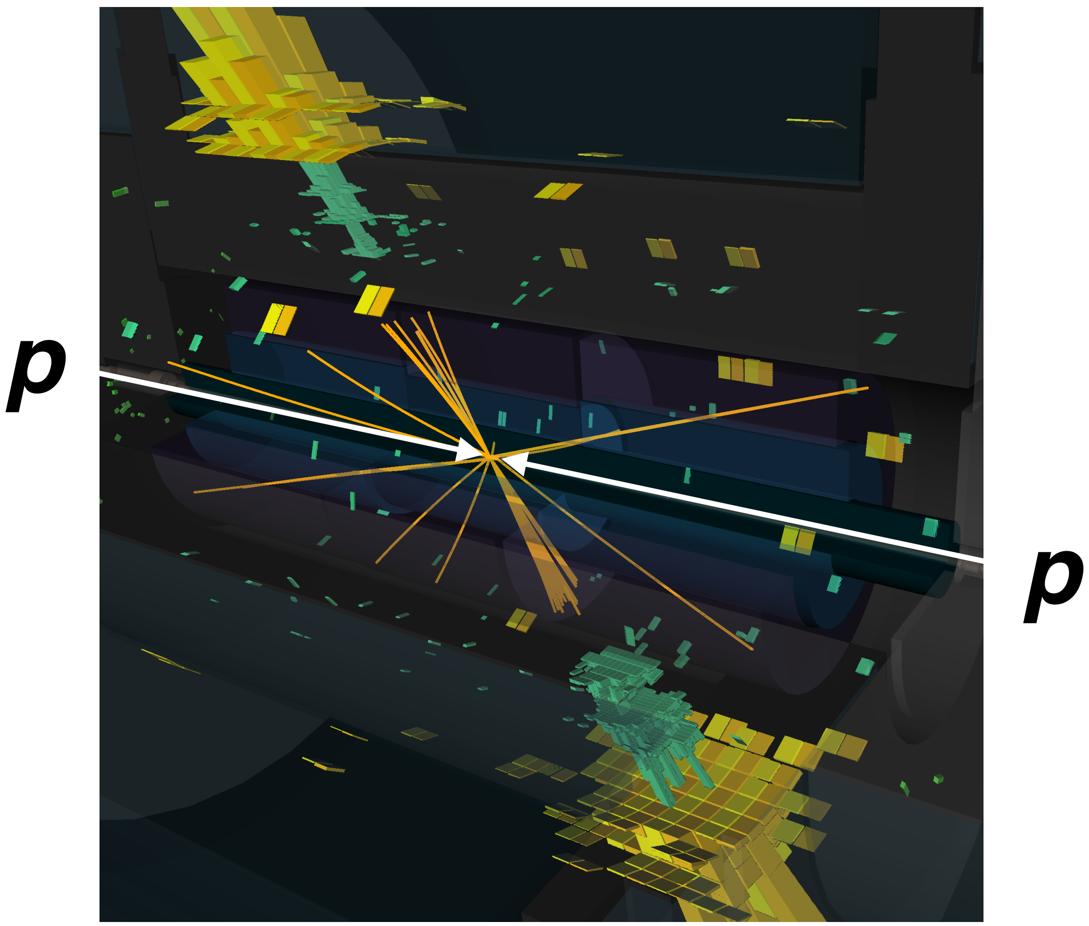
This is where most machine learning is being applied.


p
p


One of the critical goals of the LHC is to identify new, massive particles
p
?
p
Remember E = mc²:
(need lots of E to make new particles with a lot of m!)


One of the critical goals of the LHC is to identify new, massive particles
p
p
?
We want to do this using all of the available information!
Remember E = mc²: (need lots of E to make new particles with a lot of m!)

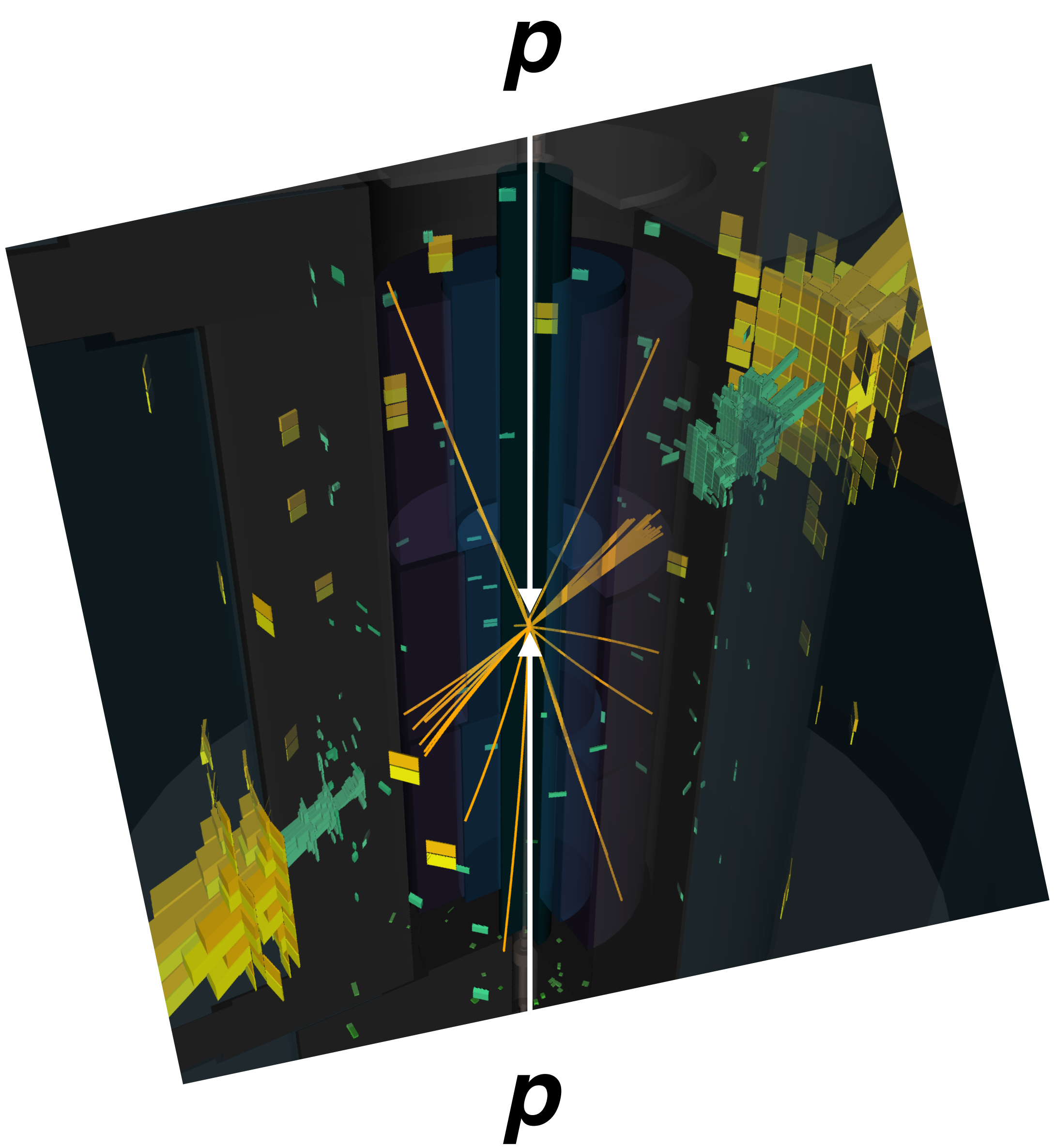
Analyzing collider events with ML



Think of an event as an image + convolution neural network

O(100) x O(100) pixels = O(10^4) dimensions!   (state-of-the-art image processing tool)

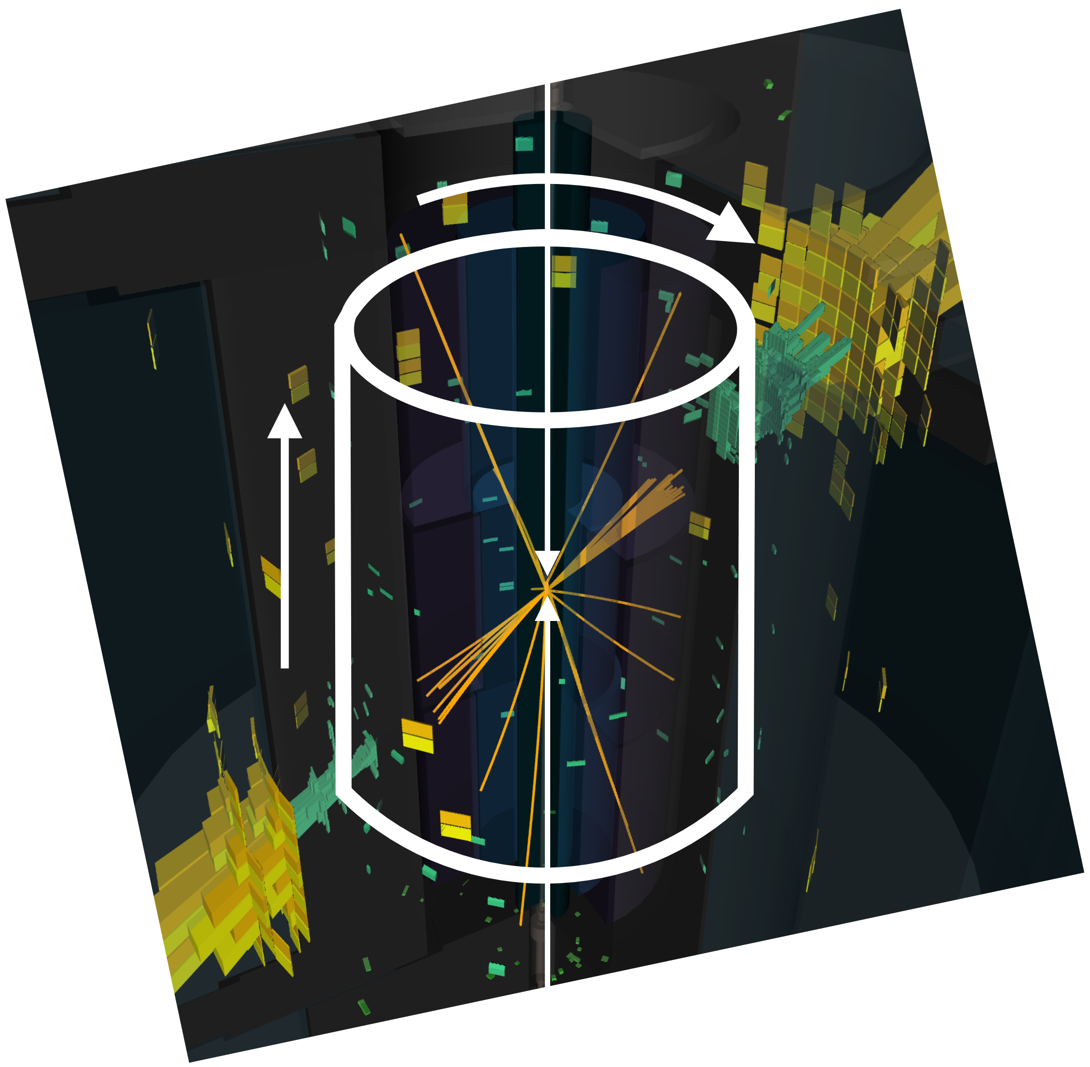
Analyzing collider events with ML



Think of an event as an image + convolution neural network

O(100) x O(100) pixels = O(10^4) dimensions! (state-of-the-art image processing tool)

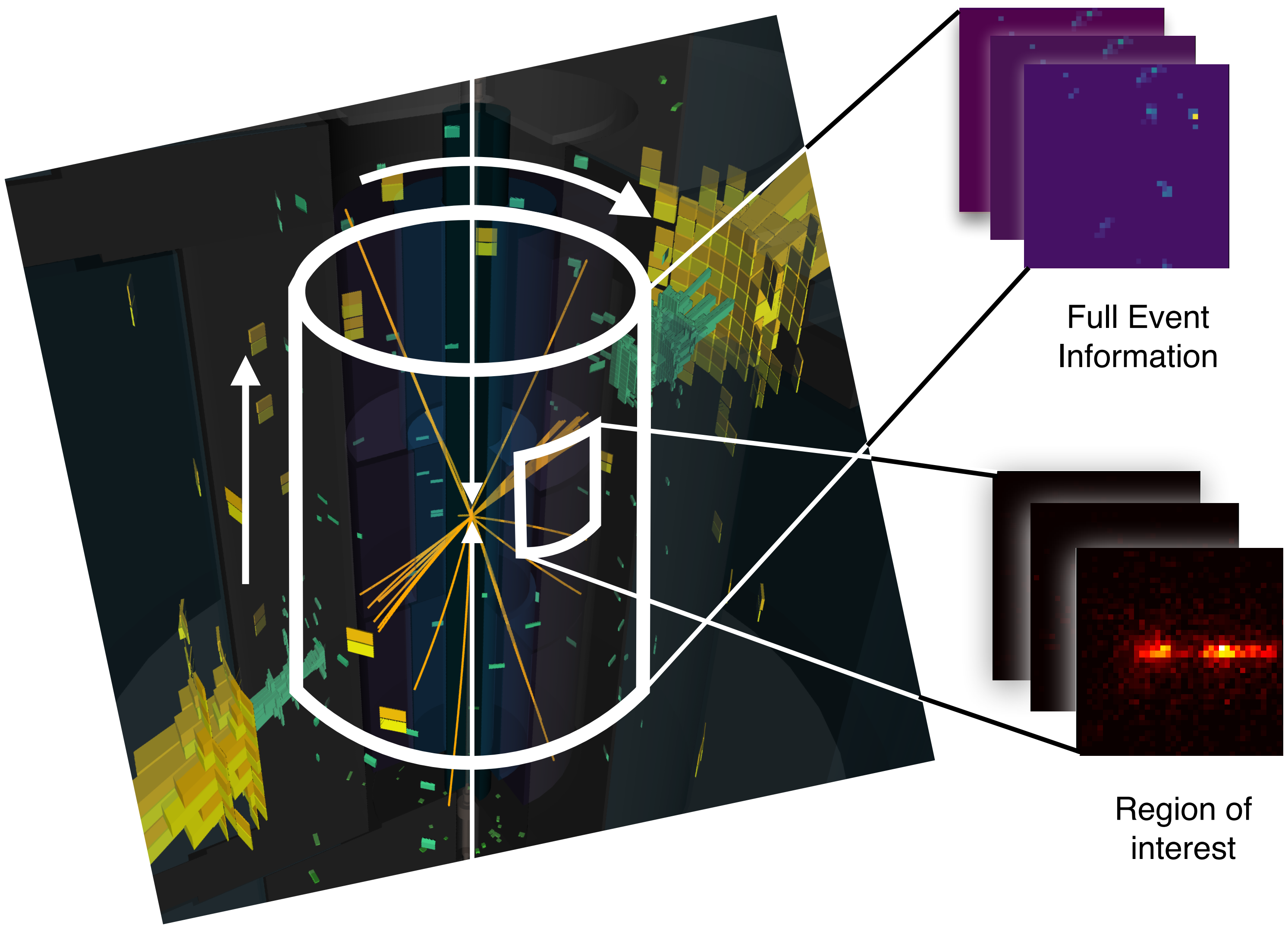
Analyzing collider events with ML



Think of an event as an image + convolution neural network

O(100) x O(100) pixels = $O(10^4)$ dimensions! (state-of-the-art image processing tool)

Analyzing collider events with ML

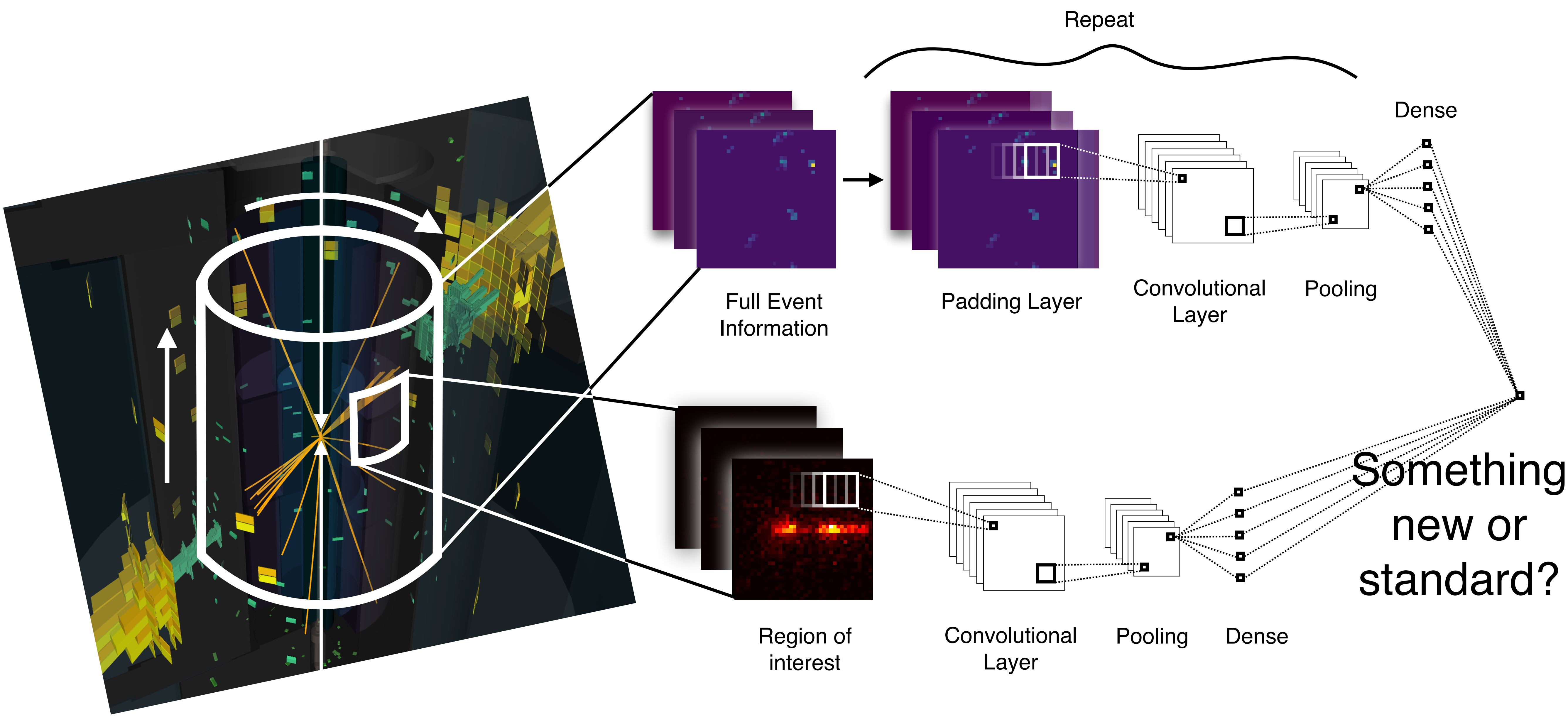


Think of an event as an image + convolution neural network

O(100) x O(100) pixels = O(10^4) dimensions! (state-of-the-art image processing tool)

Analyzing collider events with ML

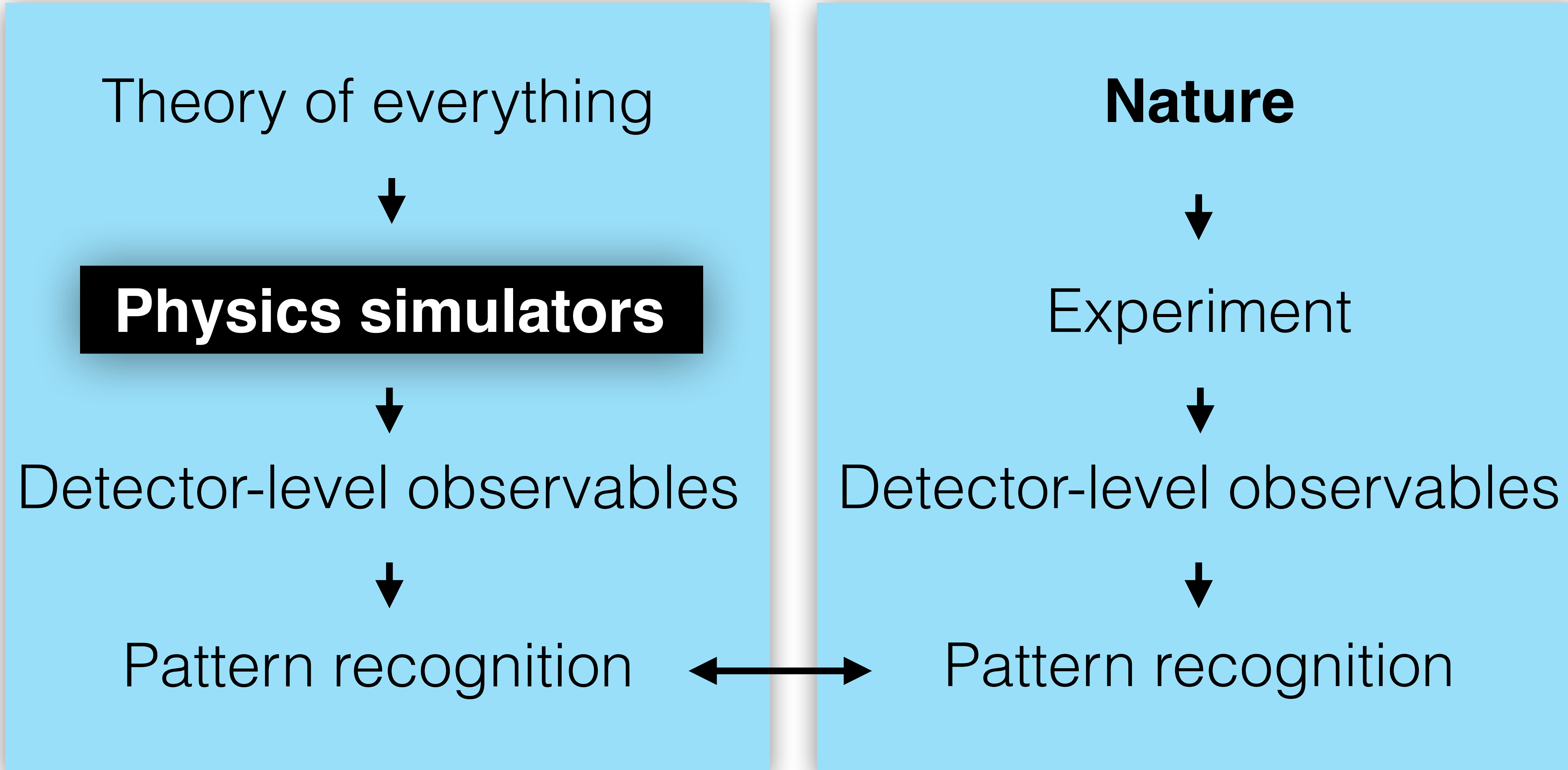
Image: Journal of High Energy Physics 10 (2018) 101



Think of an event as an image + convolution neural network

O(100) x O(100) pixels = O(10^4) dimensions!   (state-of-the-art image processing tool)

Data analysis in particle physics



Data analysis in particle physics

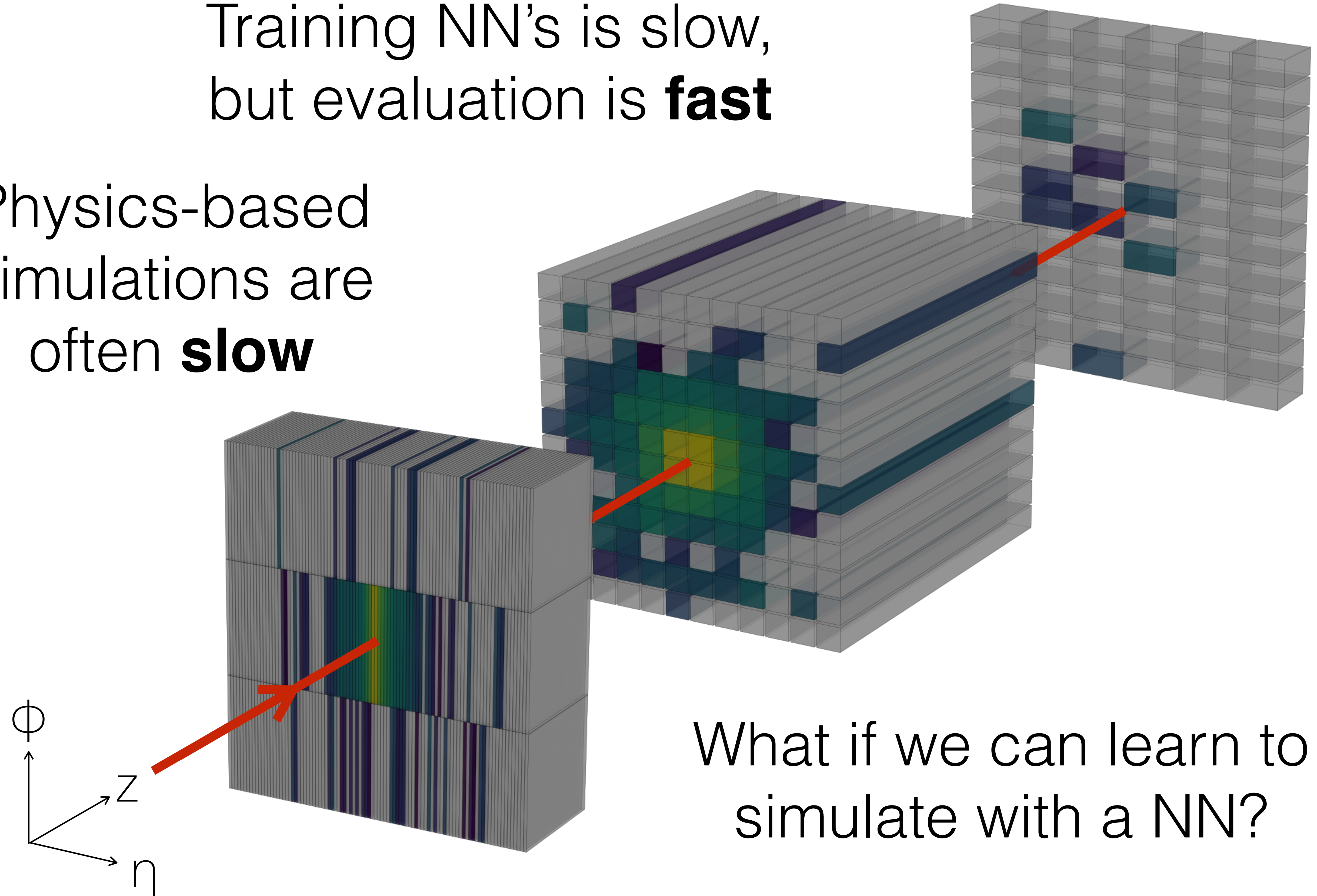
Theory of everything	**Nature**
↓	↓
Physics simulators	Experiment
↓	↓
Detector-level observables	Detector-level observables
↓	↓
Pattern recognition ⟷	Pattern recognition

A growing toolkit called “generative models” are being developed to accelerate or augment simulations.

Generative models

Training NN's is slow, but evaluation is **fast**

Physics-based simulations are often **slow**



What if we can learn to simulate with a NN?

Deep Generative Models

Can we combine our physics-simulator with deep learning?

A **generator** is nothing other than a function that maps random numbers to structure.

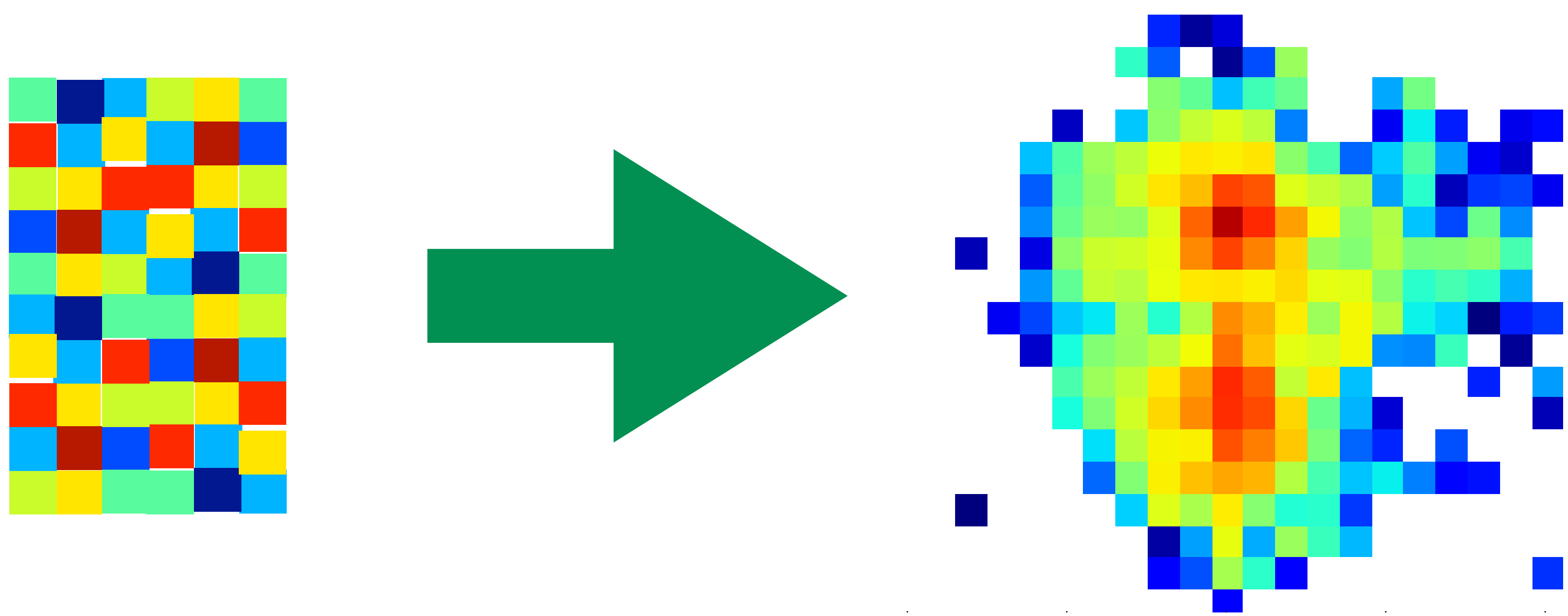
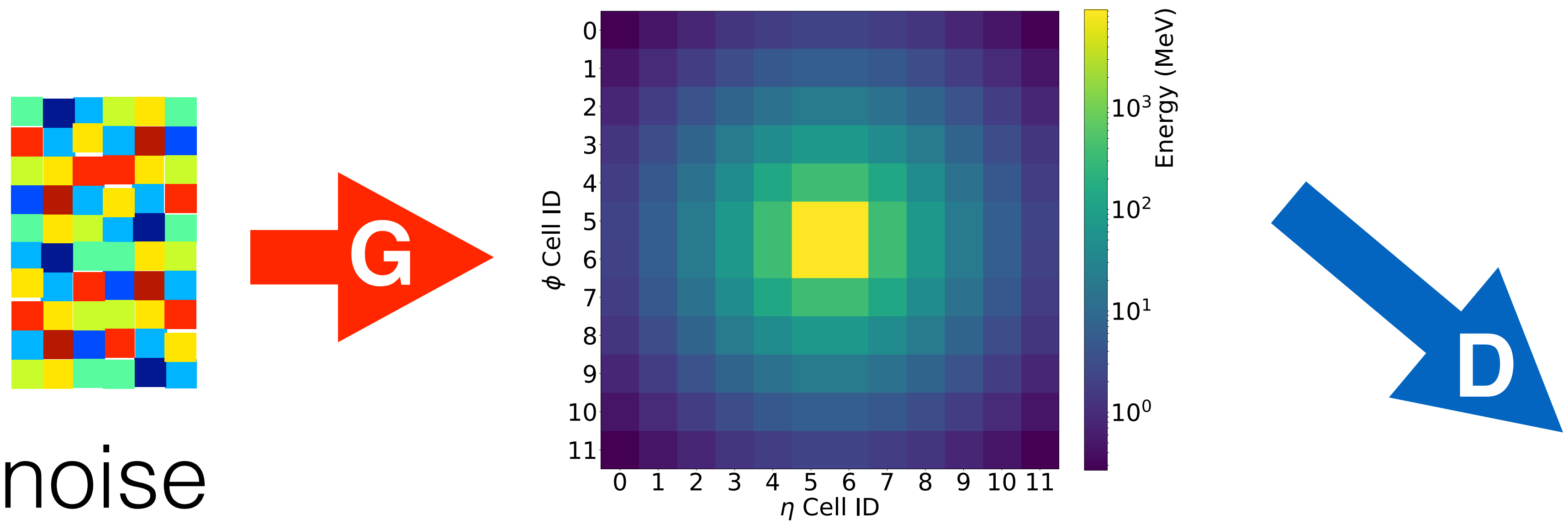


Image: Physical Review Letters 120 (2018) 042003

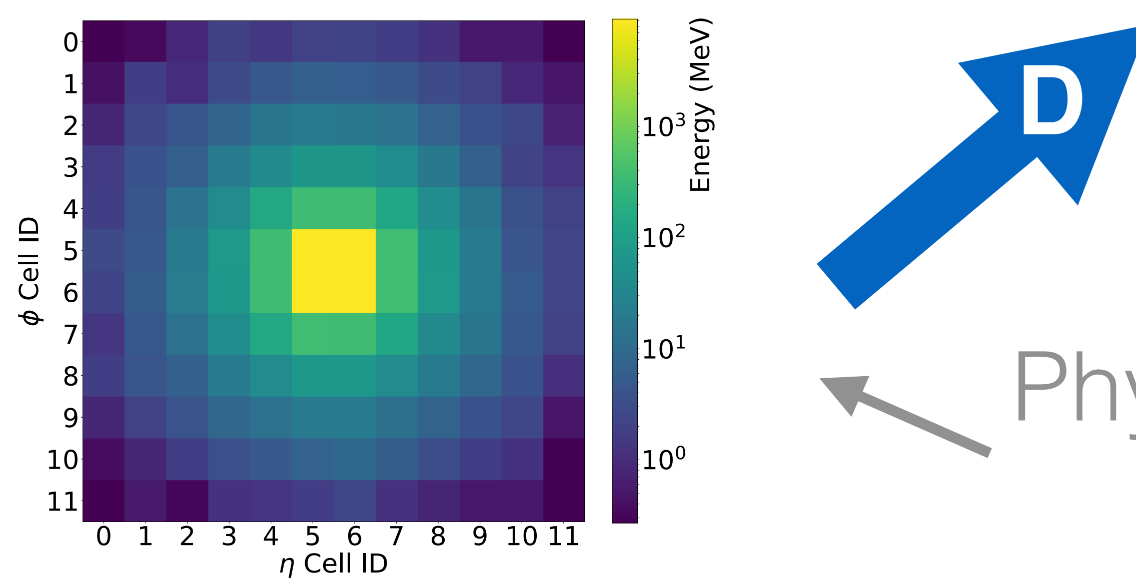
A deep learning solution: GANs

Generative Adversarial Networks (GAN):
A two-network game where one ***maps noise to images*** *and one* ***classifies images as fake or real****.*



noise

{real,fake}

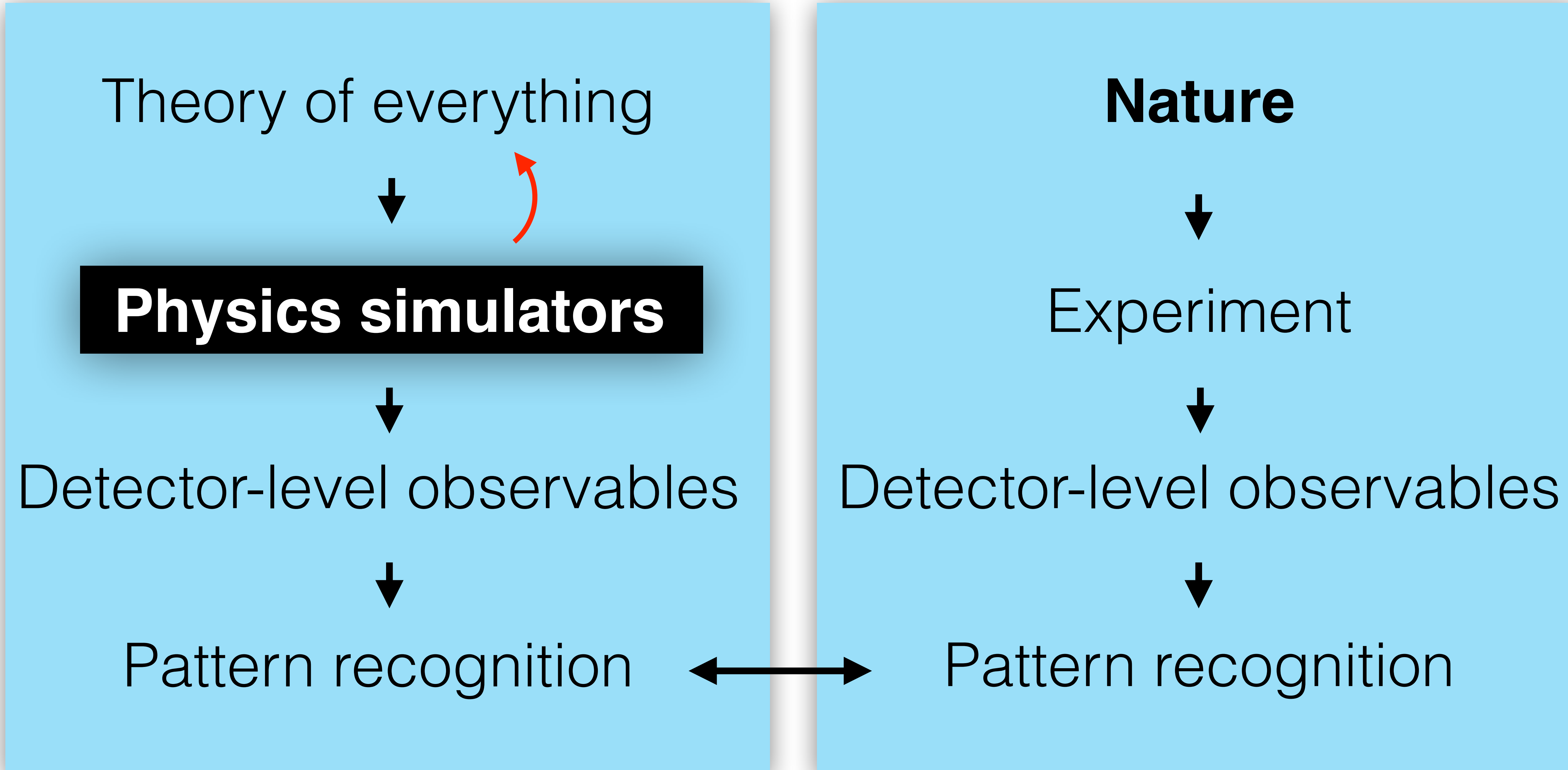


When **D** is maximally confused, **G** will be a good generator

Physics-based simulator

[I. Goodfellow et al., NIPS 2014]

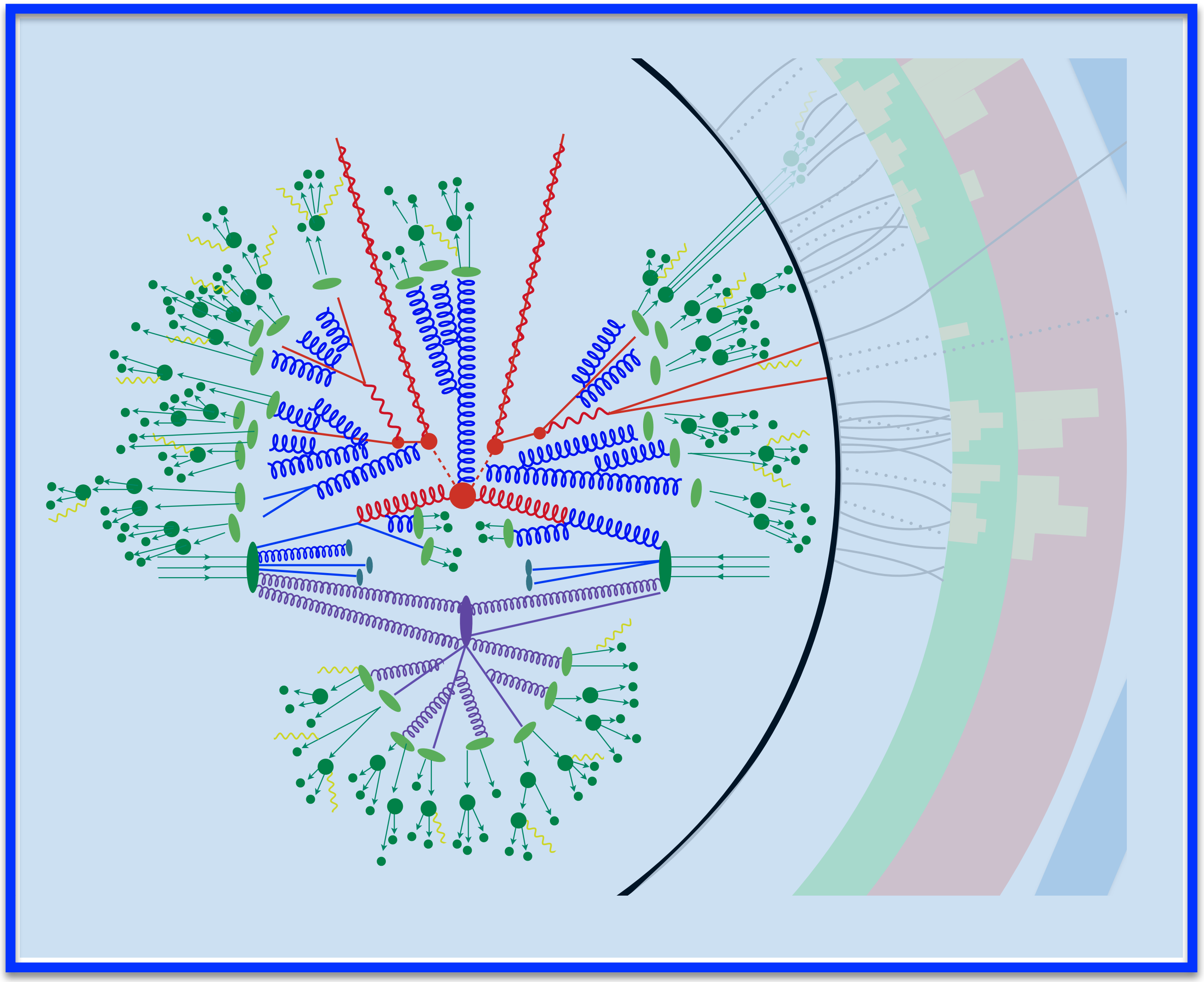
Data analysis in particle physics



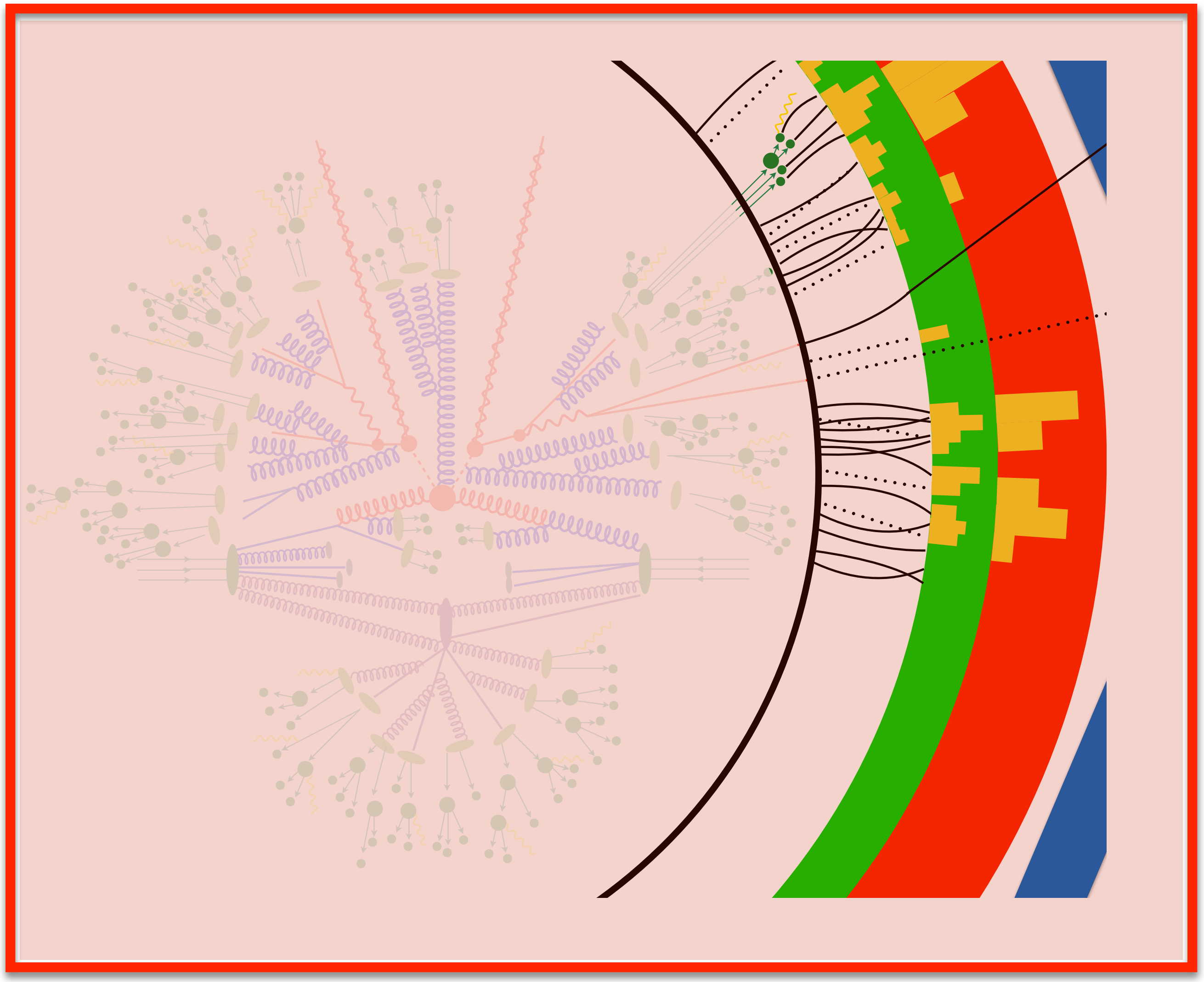
Simulators are a unique and powerful aspect of particle physics, but, they do not allow us to go “backwards” !!

The Inference Challenge

Want this



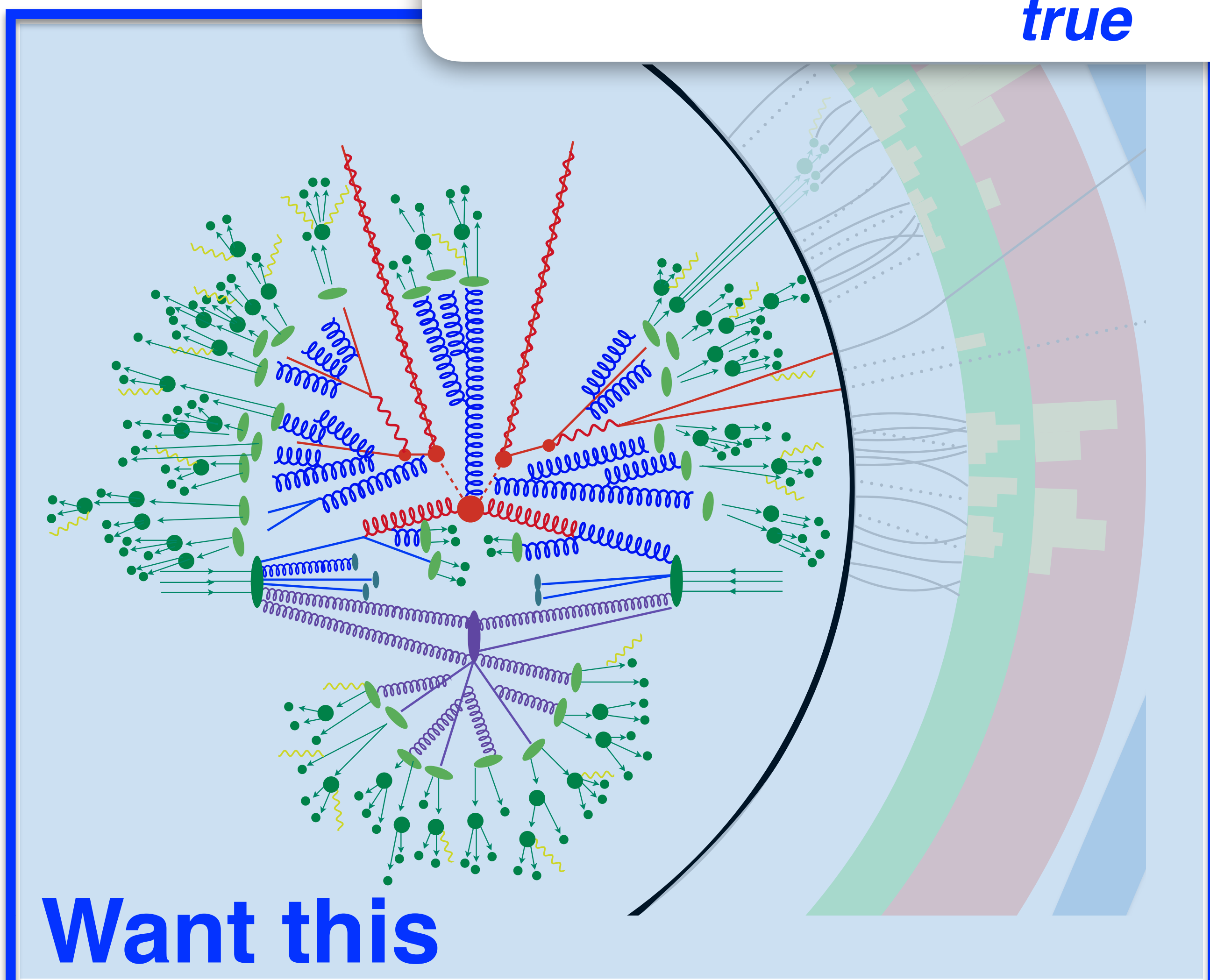
Measure this

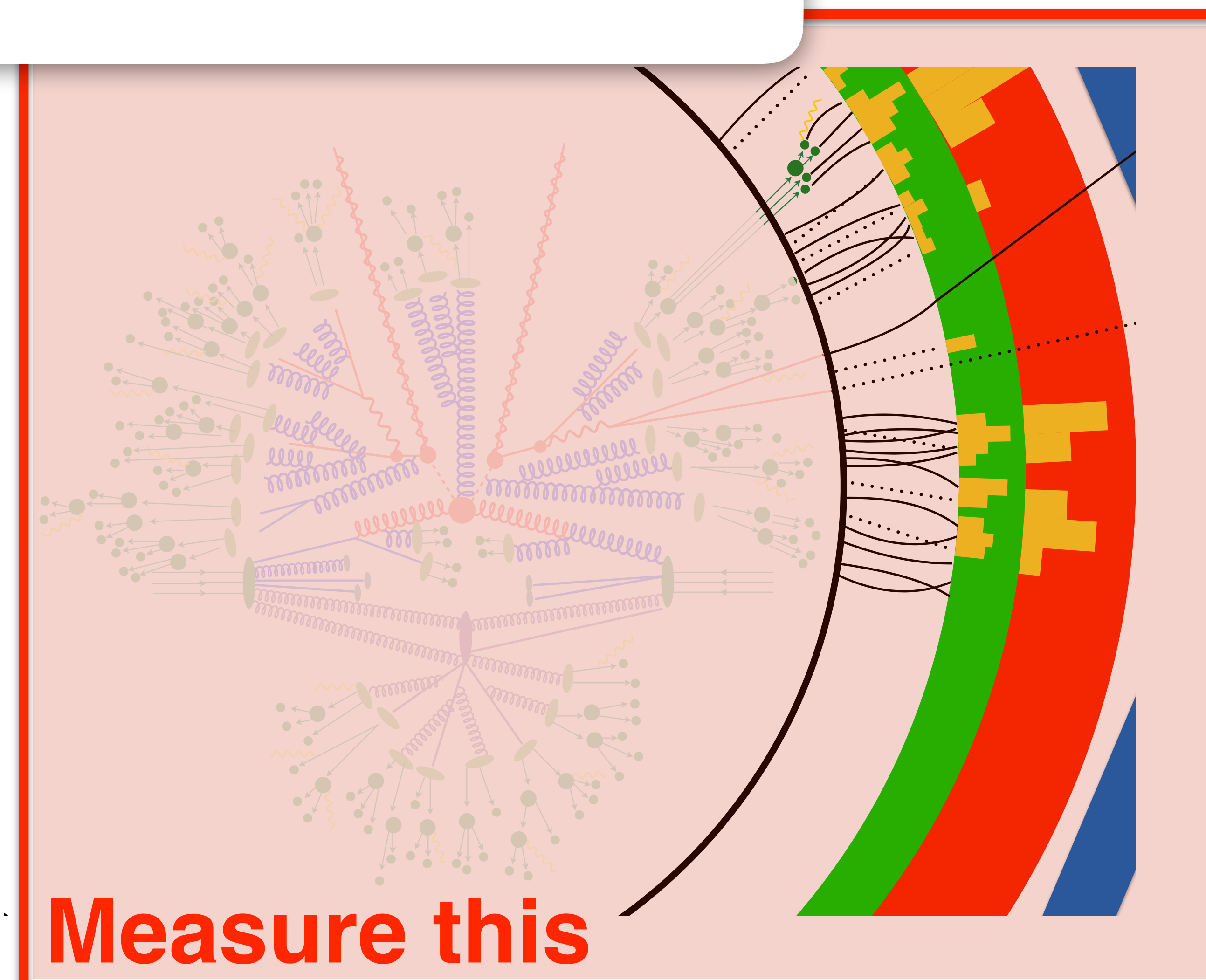


The Inference Challenge

If you know *p(**meas.** / **true**)*, could do maximum likelihood, i.e.

$$\text{unfolded} = \underset{\text{true}}{\operatorname{argmax}}\; p(\text{measured} \mid \text{true})$$





The Inference Challenge

If you know *p(**meas.** | **true**)*, could do maximum likelihood, i.e.

$$\textit{unfolded} = \underset{\textit{true}}{\operatorname{argmax}}\; p(\textit{measured} \mid \textit{true})$$

Challenge: **measured** is hyperspectral and **true** is hypervariate … *p(**meas.** | **true**) is **intractable** !*

The Inference Challenge

If you know *p(**meas.** / **true**)*, could do maximum likelihood, i.e.

$$\textit{unfolded} = \underset{\textbf{true}}{\operatorname{argmax}}\; p(\textbf{measured} \,/\, \textbf{true})$$

Challenge: **measured** is hyperspectral and **true** is hypervariate … *p(**meas.** | **true**) is **intractable** !*

However: we have **simulators** that we can use to sample from *p(**meas.** | **true**)*

→ **Simulation-based (likelihood-free) inference !**

…an area of machine learning were particle physics is making a key contribution!

I’ll briefly show you one solution to give you a sense of the power of likelihood-free inference.

Reweighting

I'll briefly show you one solution to give you a sense of the power of likelihood-free inference.

The solution will be built on ***reweighting***

dataset 1: sampled from ***p(x)***
dataset 2: sampled from ***q(x)***

Create weights ***w(x) = q(x)/p(x)*** so that when dataset 1 is weighted by ***w***, it is statistically identical to dataset 2.

Reweighting

I'll briefly show you one solution to give you a sense of the power of likelihood-free inference.

The solution will be built on ***reweighting***

dataset 1: sampled from ***p(x)***
dataset 2: sampled from ***q(x)***

Create weights ***w(x) = q(x)/p(x)*** so that when dataset 1 is weighted by ***w***, it is statistically identical to dataset 2.

What if we don't (and can't easily) know ***q*** and ***p***?

Classification for reweighting

Fact: Neutral networks learn to approximate the likelihood ratio $= q(x)/p(x)$

Solution: train a neural network to distinguish the two datasets!

This turns the problem of **density estimation** (**hard**) into a problem of **classification** (**easy**)

Classification for reweighting

Particularly useful for particle physics, where collisions may produce a variable # of particles which are interchangeable

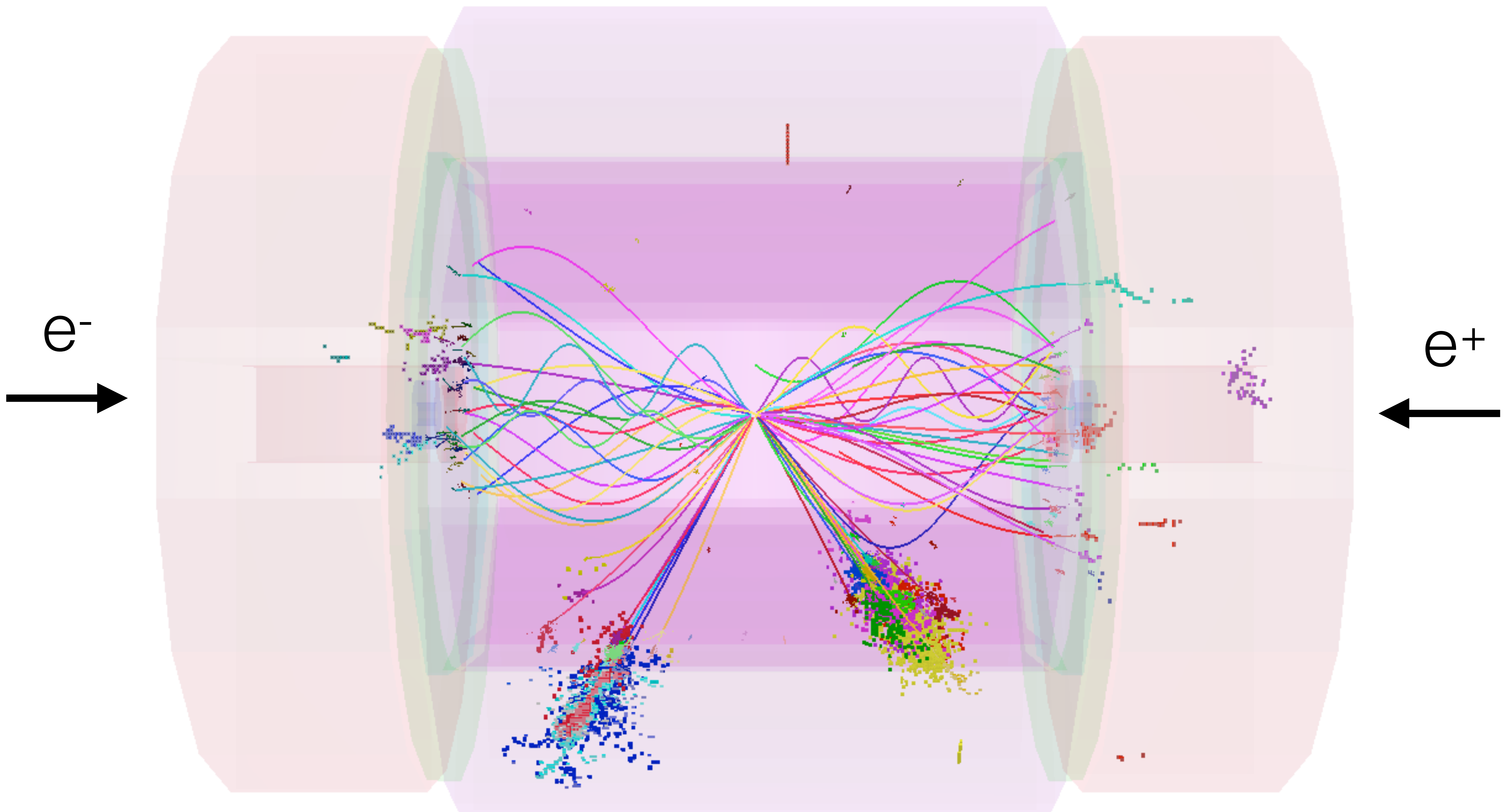
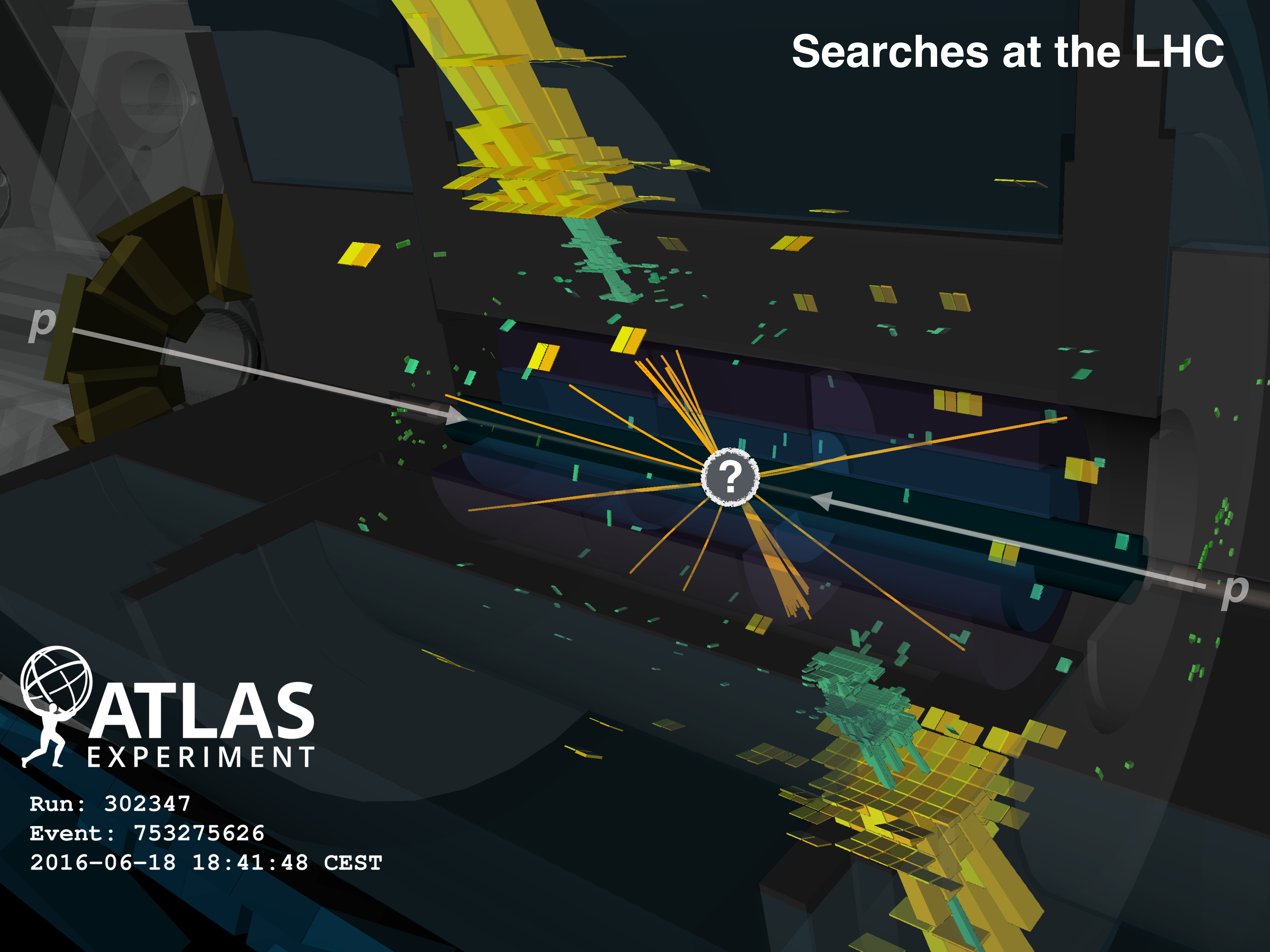


Image: Linear Collider Detector Project

Data analysis in fundamental physics

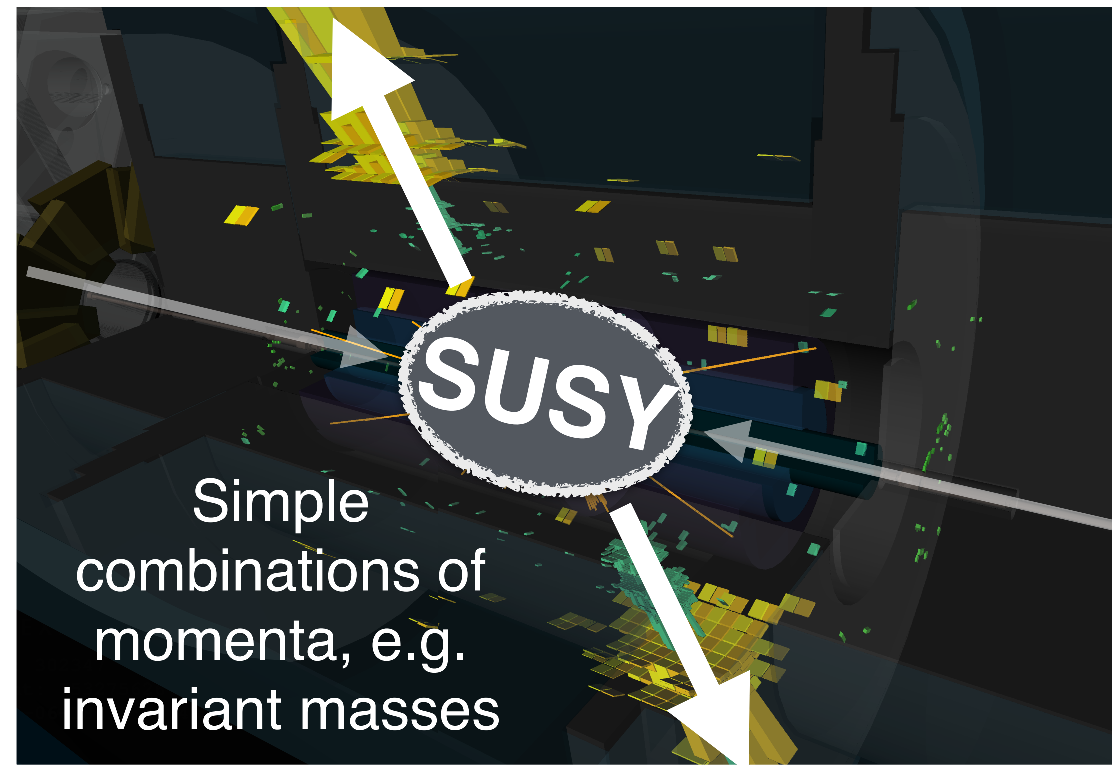
Theory of everything	**Nature**
↓	↓
Physics simulators	Experiment
↓	↓
Detector-level observables	Detector-level observables
↓	↓
Pattern recognition ⟷	Pattern recognition

Anomaly detection


Searches at the LHC
p
?
p
ATLAS
EXPERIMENT
Run: 302347
Event: 753275626
2016-06-18 18:41:48 CEST

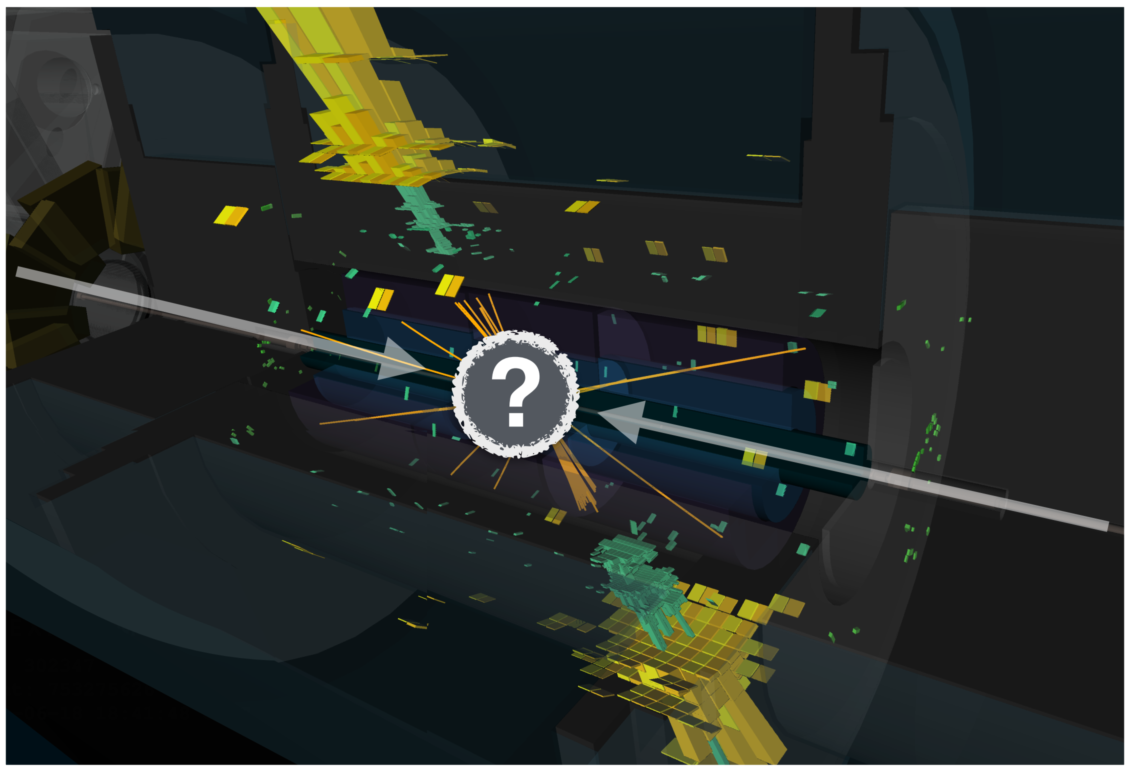
Current Search Paradigm

SUSY = Supersymmetry



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

Current Search Paradigm

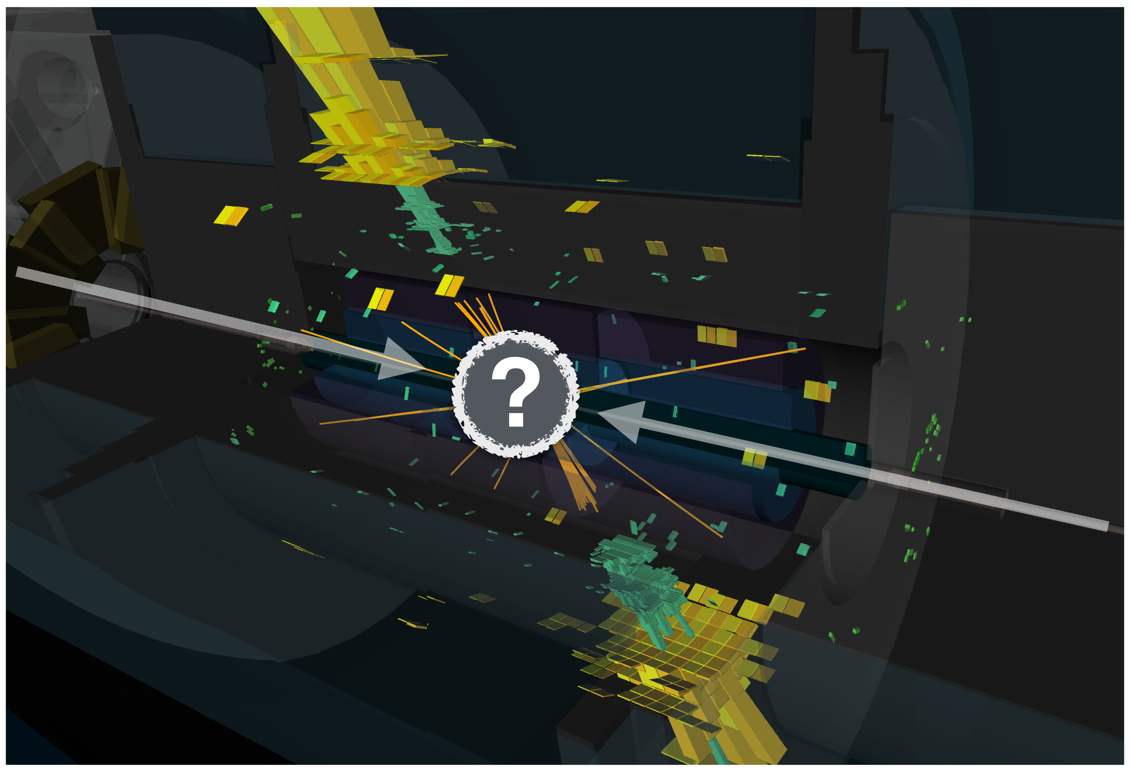


SUSY = Supersymmetry

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

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

Current Search Paradigm



What if we are not looking in the right place for the new phenomena?!

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

What is the problem?

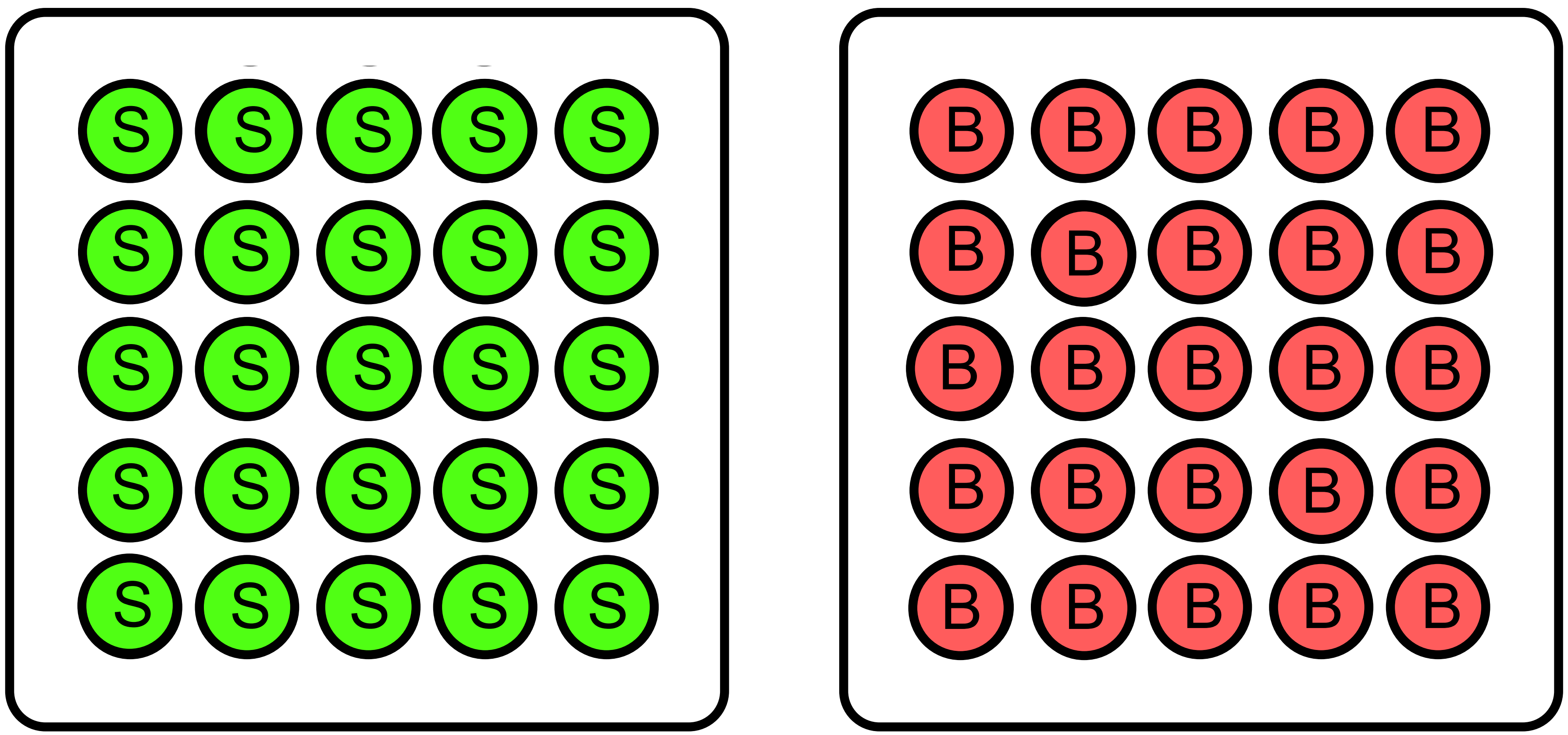
Why can't I just pay some physicists to label events and then train a neural network using those labels?

Image credit: pixabay.com

Answer: this is not cats-versus-dogs … thanks to quantum mechanics it is **not possible to know** what happened.

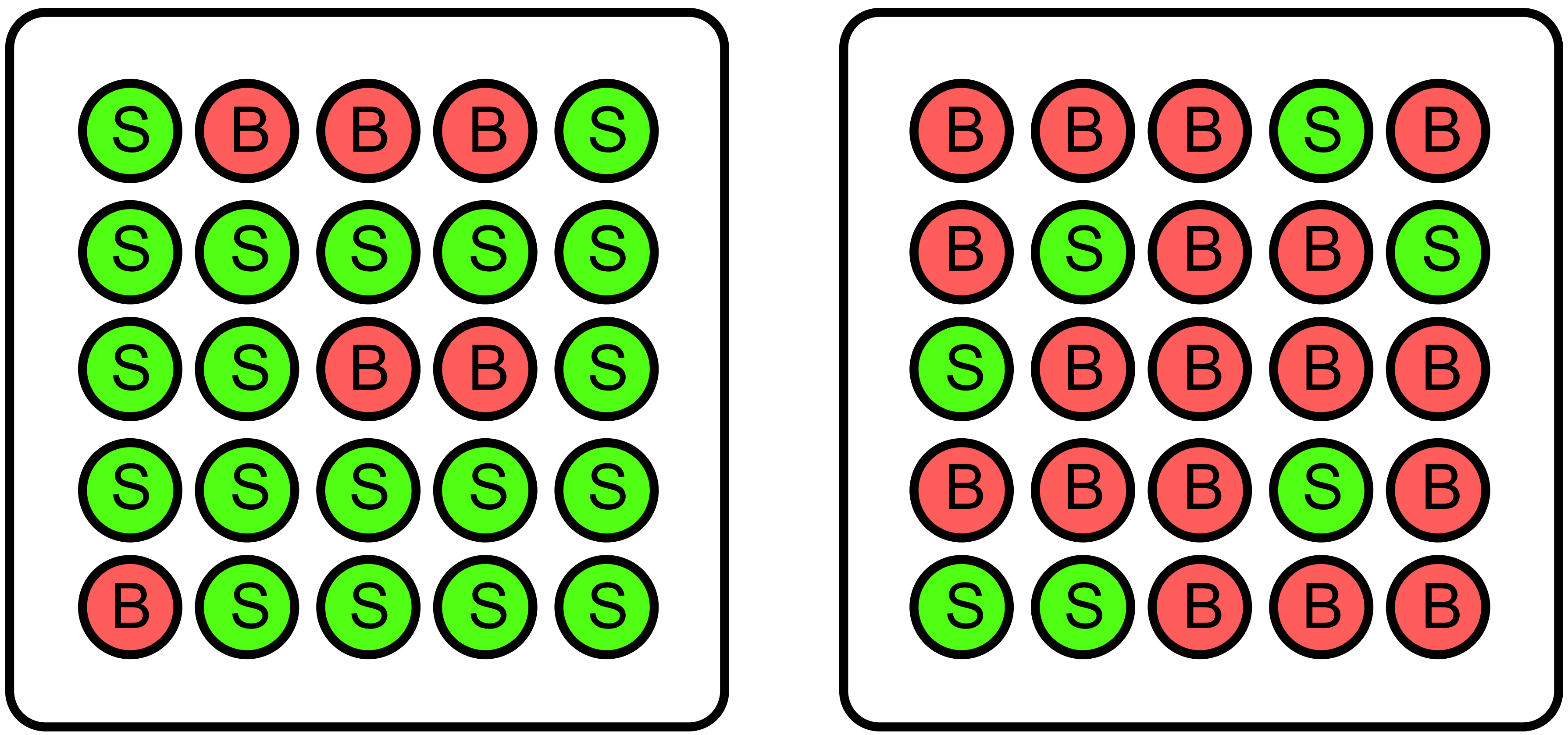
Usual case: train with labels

Usually, we train using simulations where we know which events are "**signal**" and which are "**background**".



No labels, no problem!

Can we still do machine learning when reality is like this?



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

No labels, no problem!

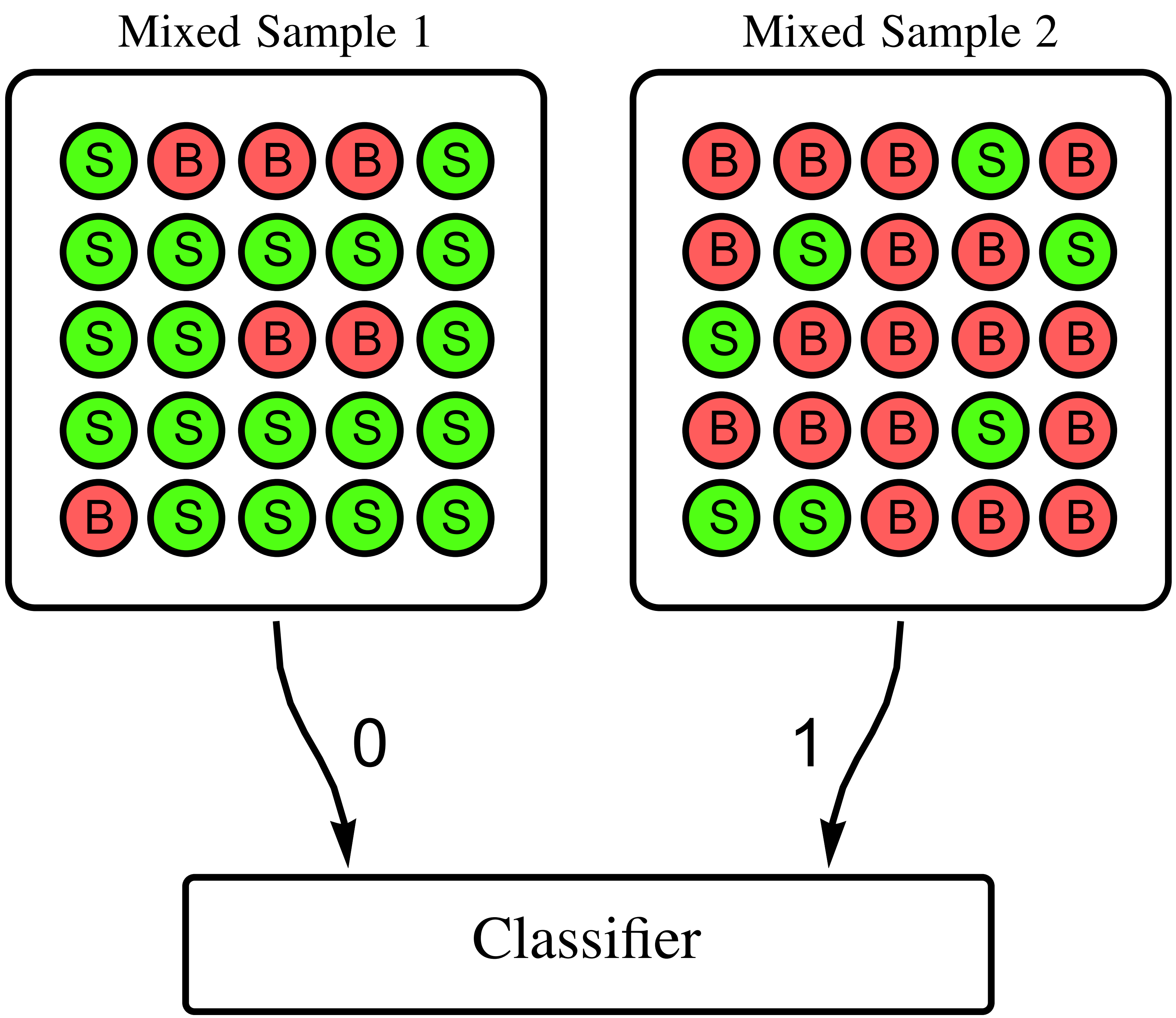
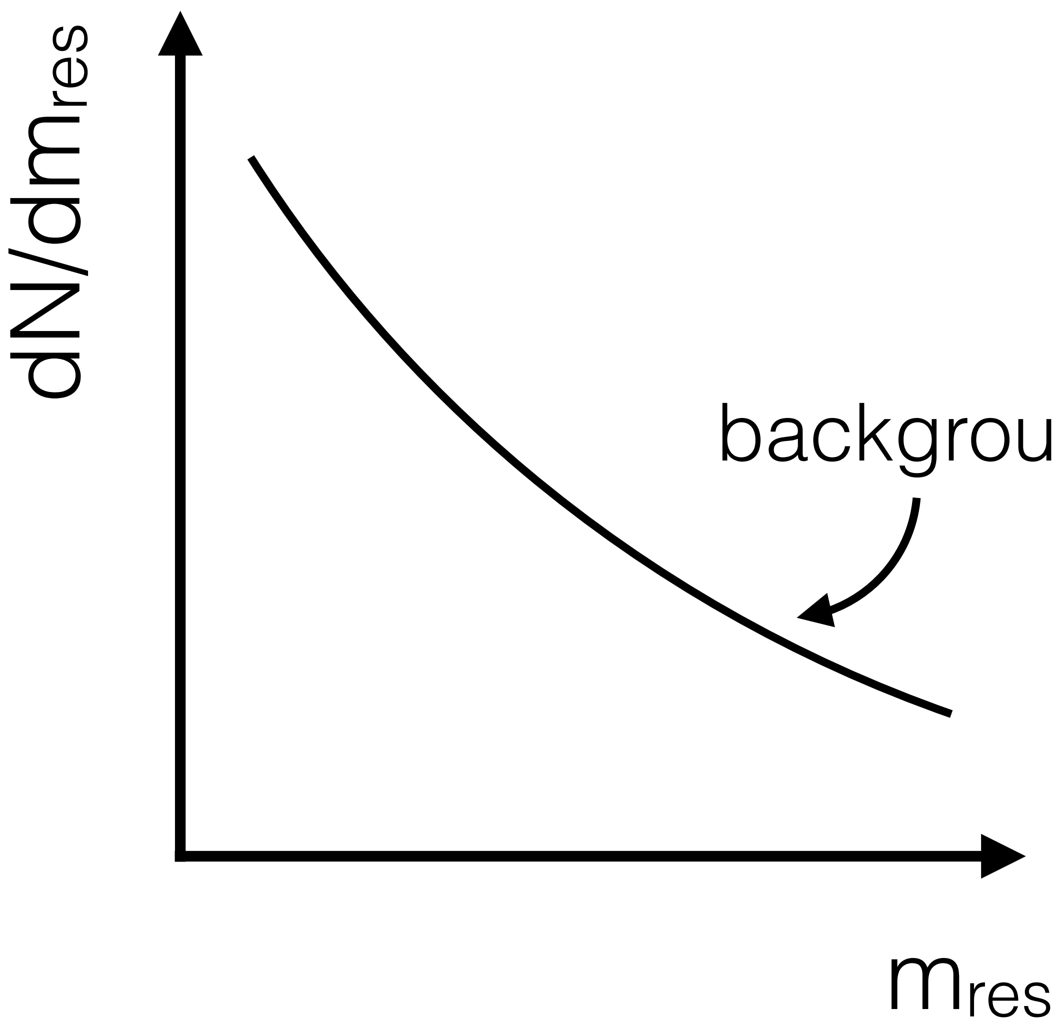


Image: Journal of High Energy Physics 10 (2017) 51

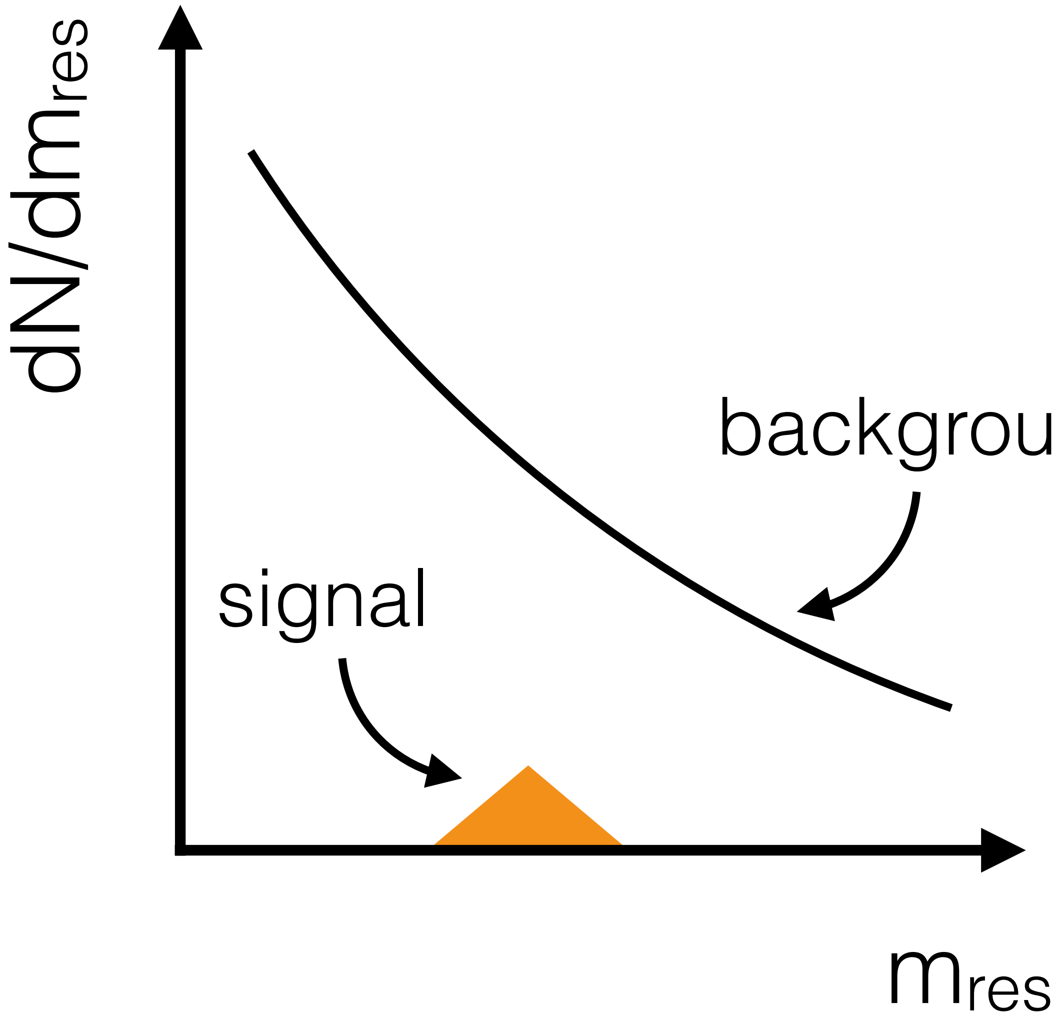
One simple, but powerful idea: “weak supervision”

No labels, no problem!



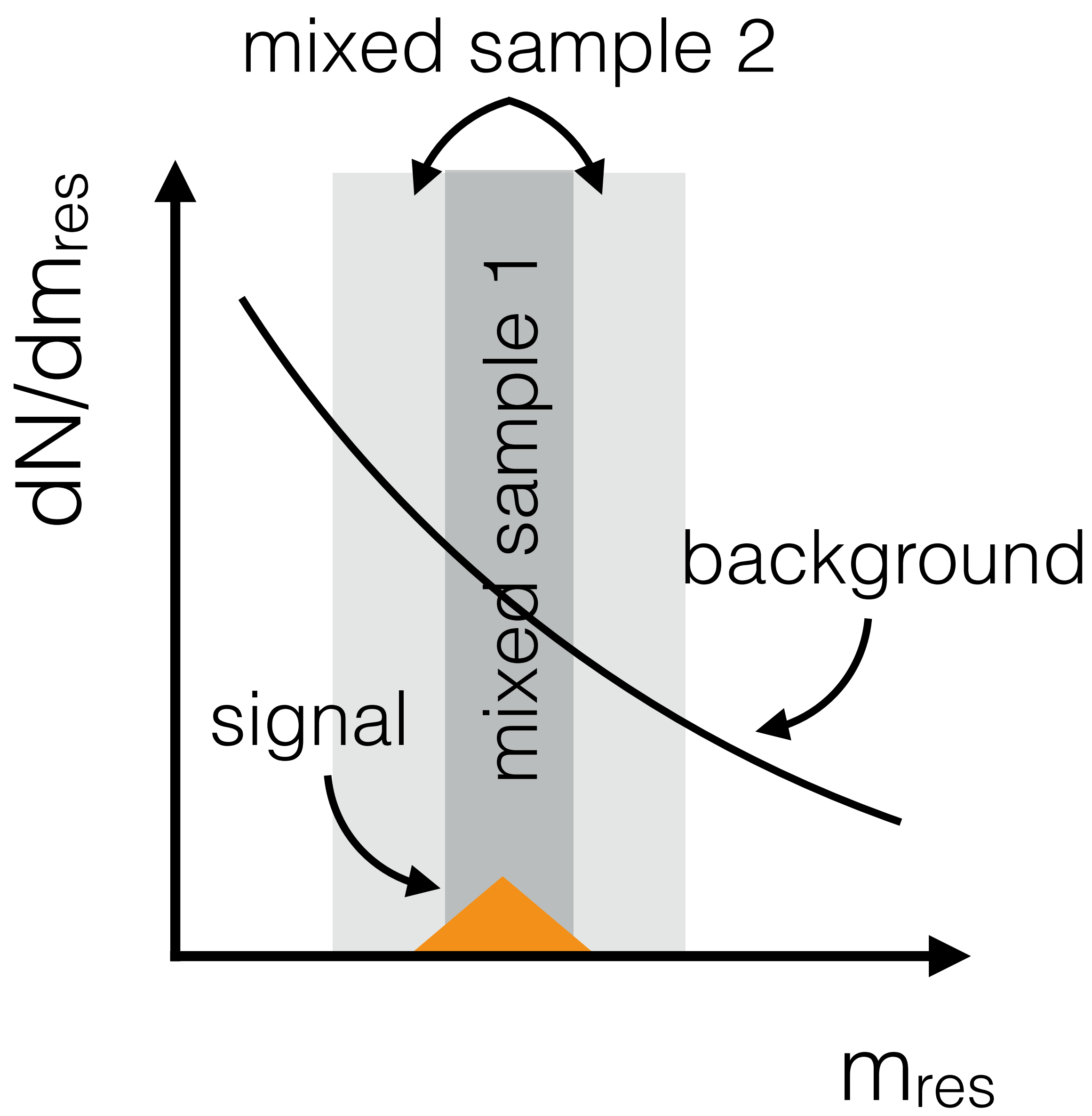
Assumption: there is a feature that we know about where the background is smooth and the signal (if it exists) is localized.

No labels, no problem!



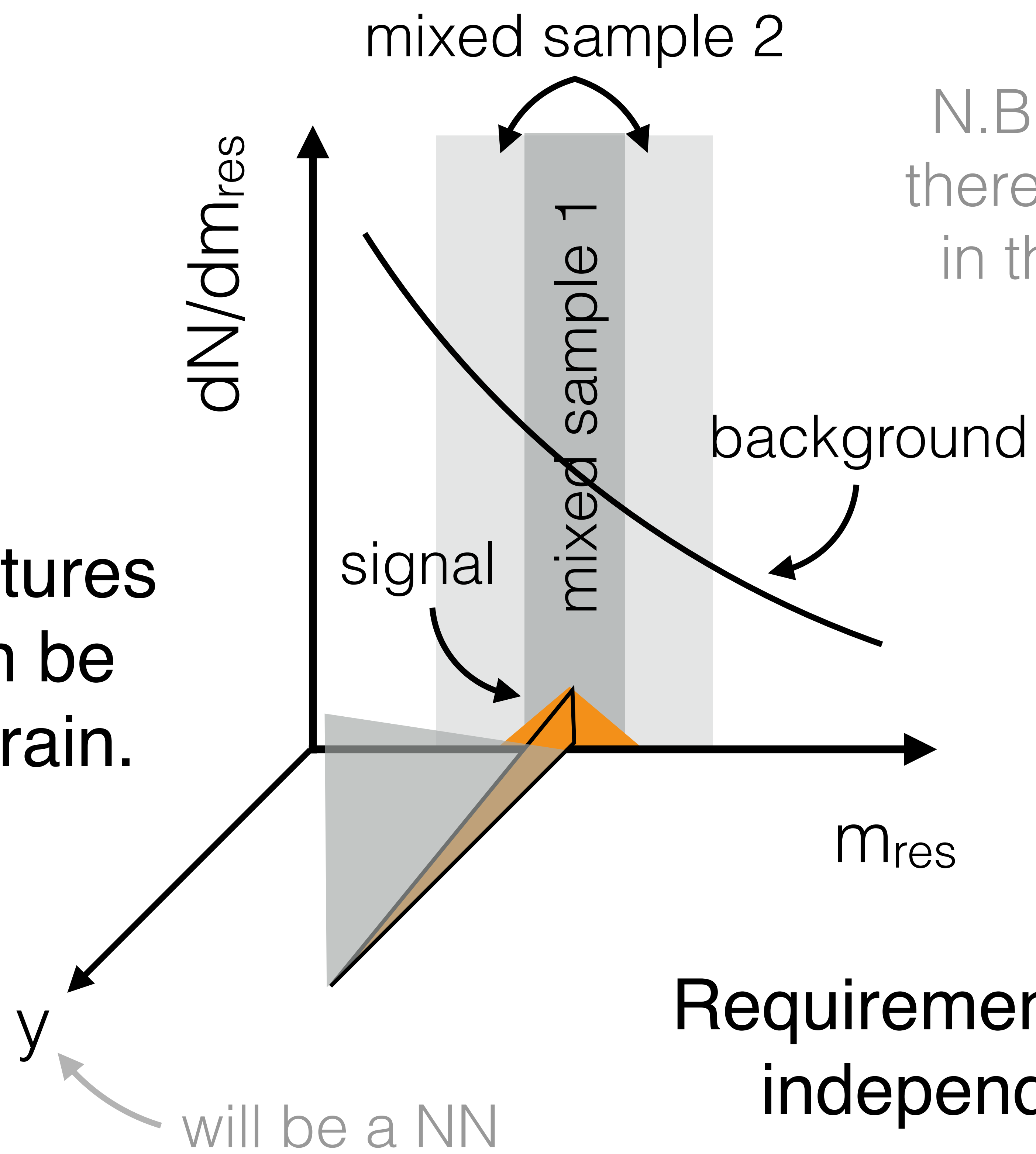
Assumption: there is a feature that we know about where the background is smooth and the signal (if it exists) is localized.

No labels, no problem!



We don’t know where the signal is, but for a given hypothesis, we can make signal windows and sidebands.

No labels, no problem!

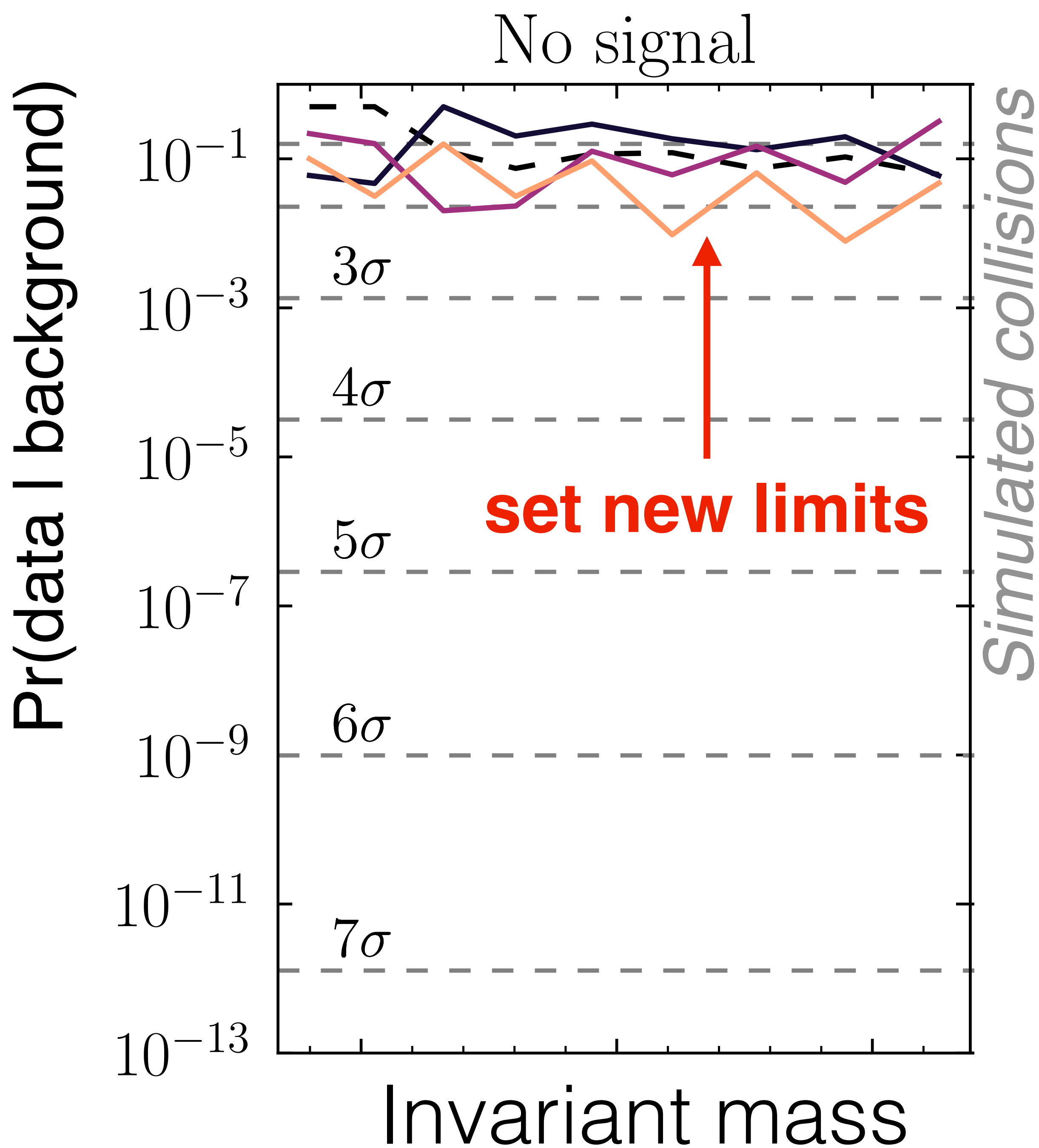


N.B. it is okay that there is some signal in the sidebands!

Other features can then be used to train.

Requirement: y is (nearly) independent of m_{res}.

Potential of weak supervision



Necessary: when there is no anomaly, the procedure does not find an anomaly.

Image: Physical Review Letters 121 (2018) 241803

Potential of weak supervision

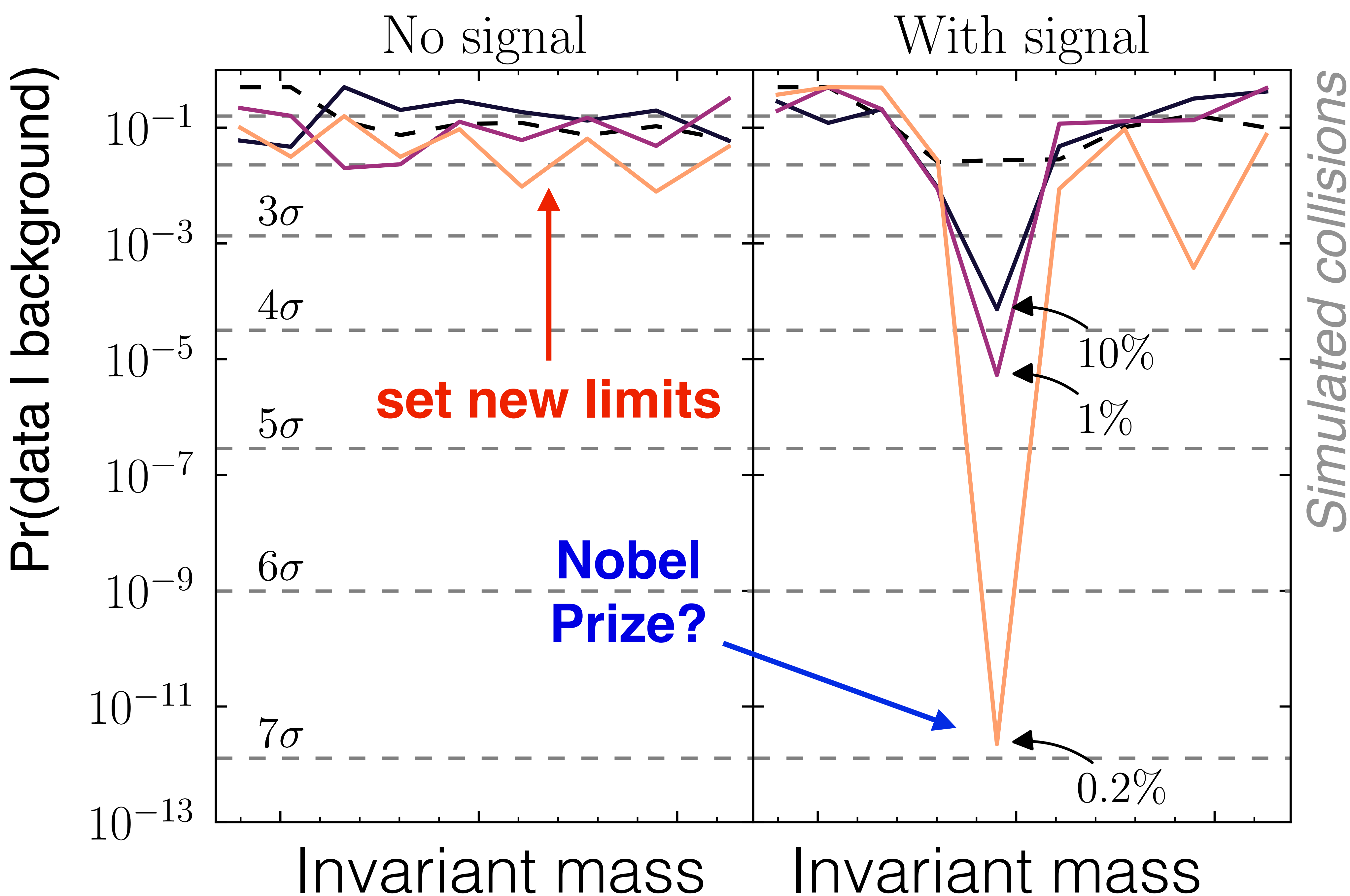
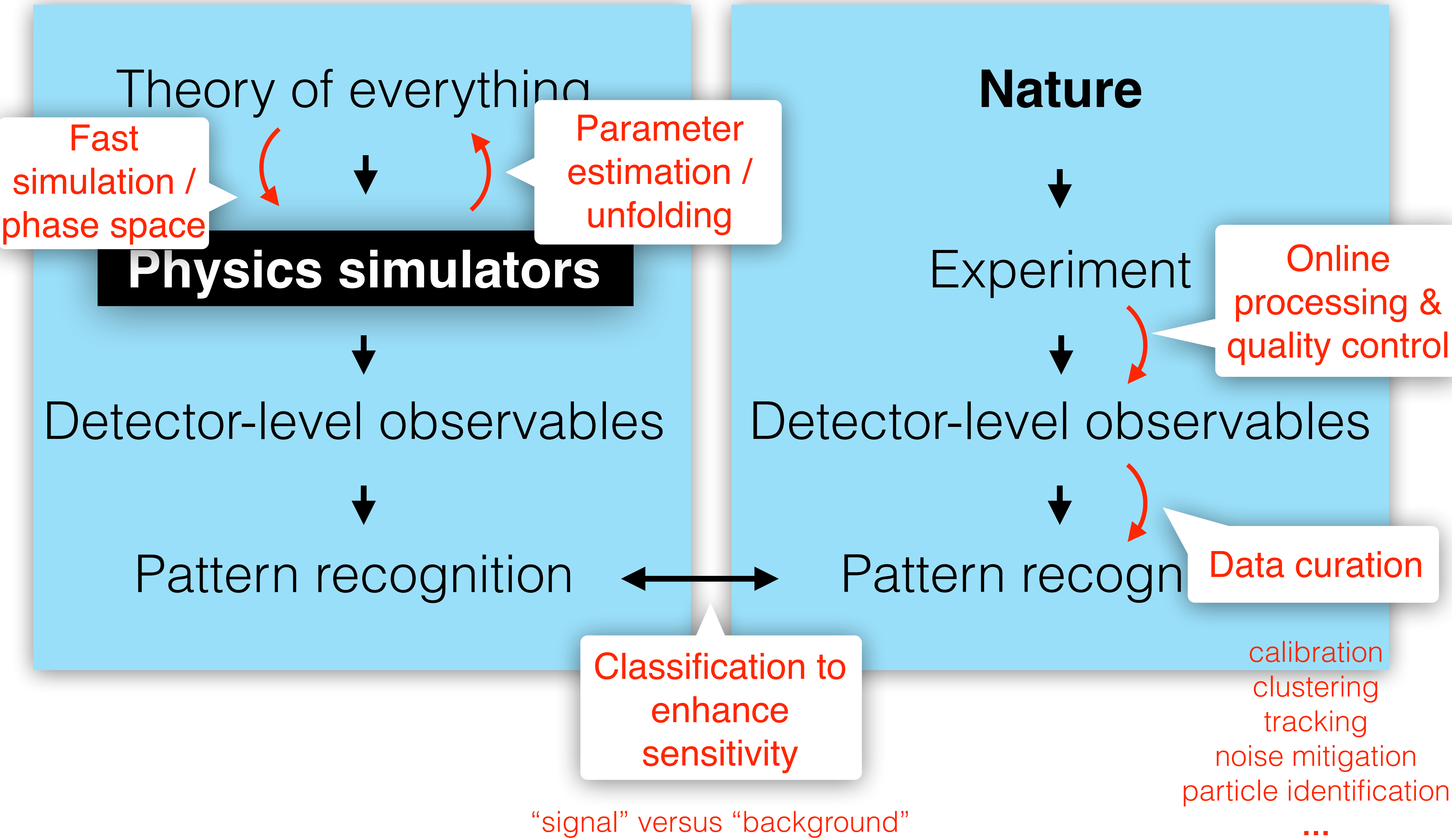


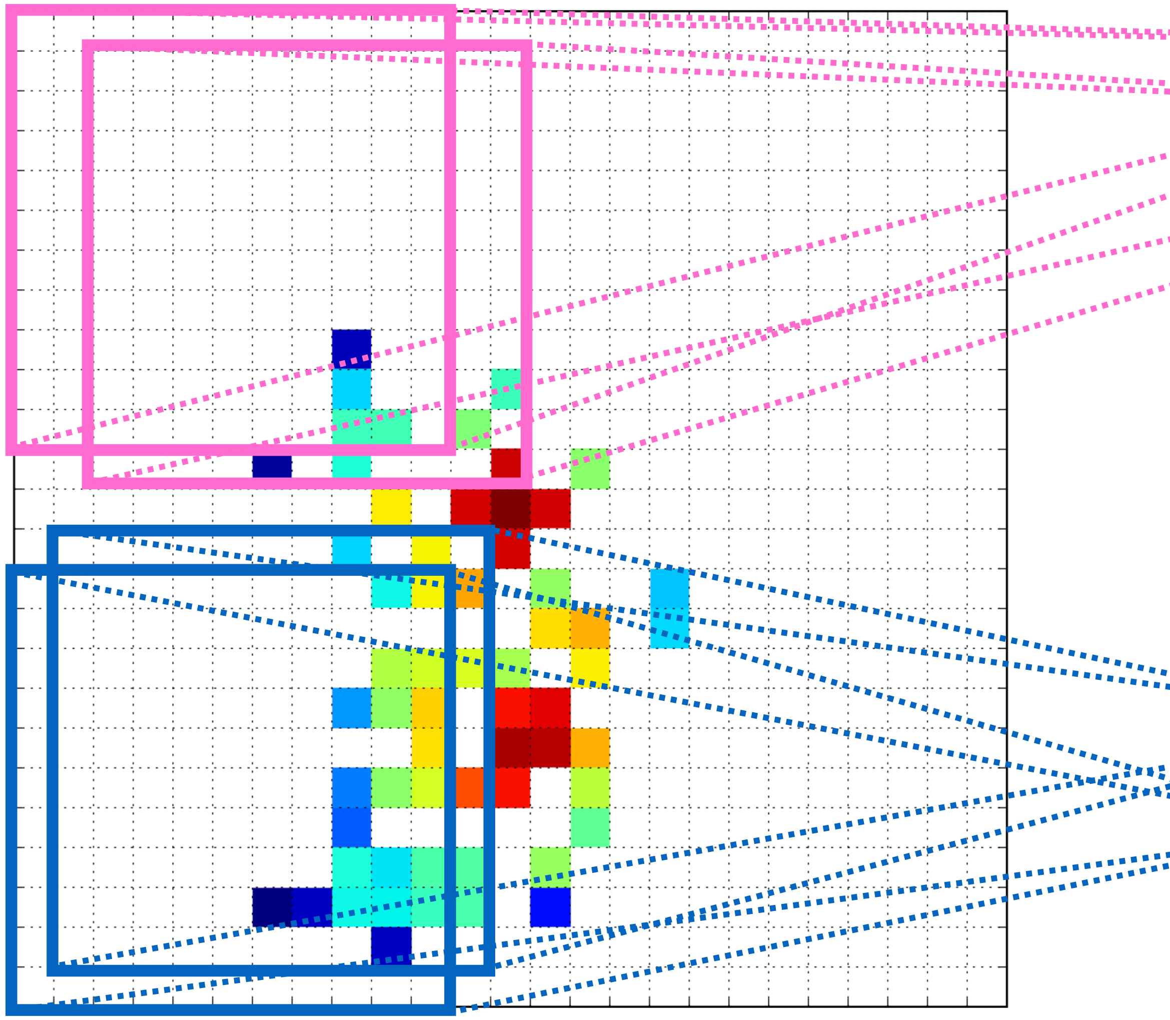
Image: Physical Review Letters 121 (2018) 241803

Overview: Particle physics and ML



Conclusions and outlook

Deep learning has a great potential to **enhance**, **accelerate**, and **empower** discoveries in particle physics.



This set of tools is growing in importance and no matter what you do, they will help you do it better.

Consider taking a statistics / statical learning / machine learning / applied statistics course(s)!

How else to get involved?

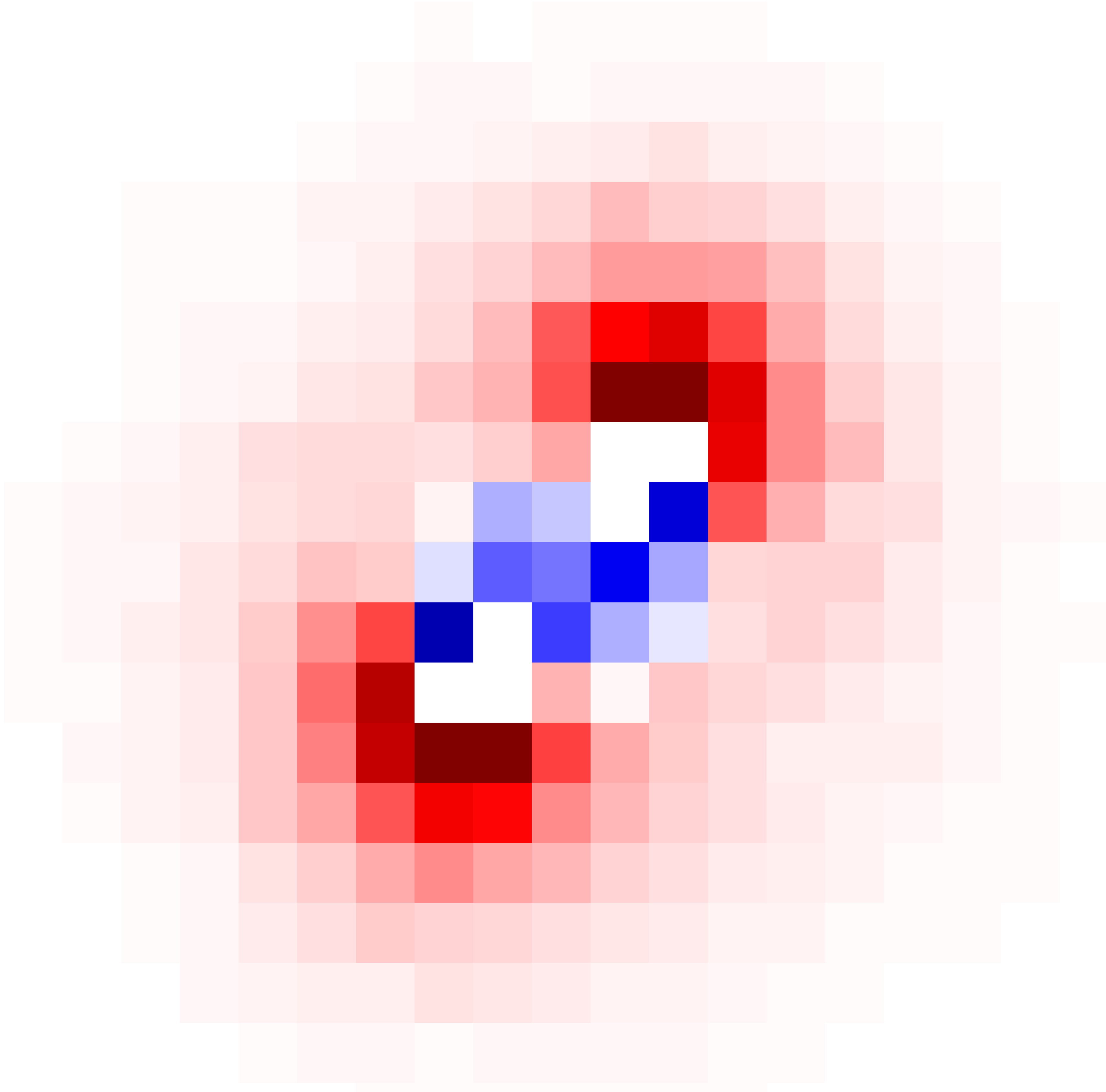
BERKELEY INSTITUTE FOR DATA SCIENCE

ML and Science Forum,
biweekly on Mondays at 11 AM

BERKELEY LAB

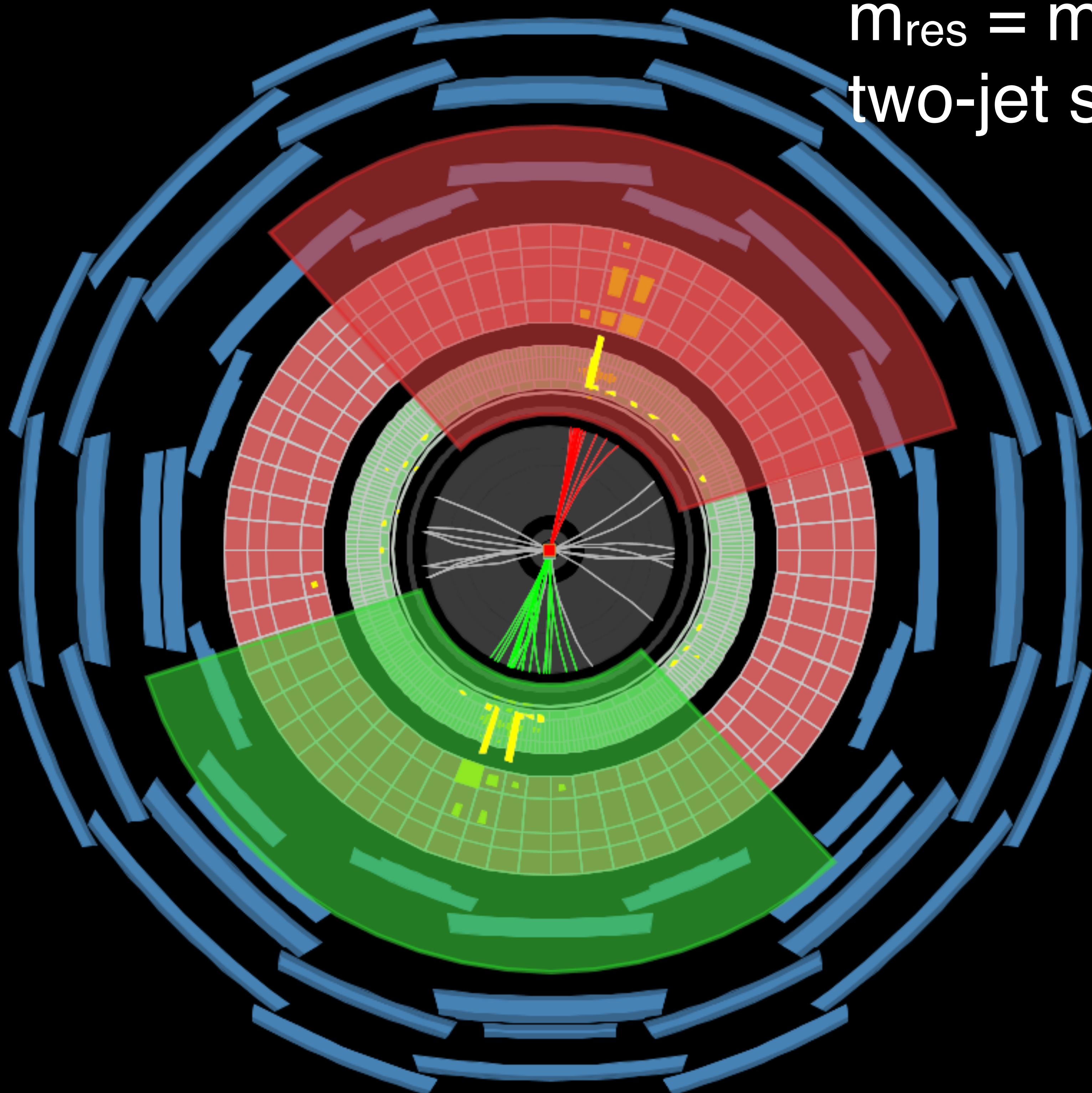
Physics Division ML meetings,
weekly on Thursdays at 1 PM

(open to all)

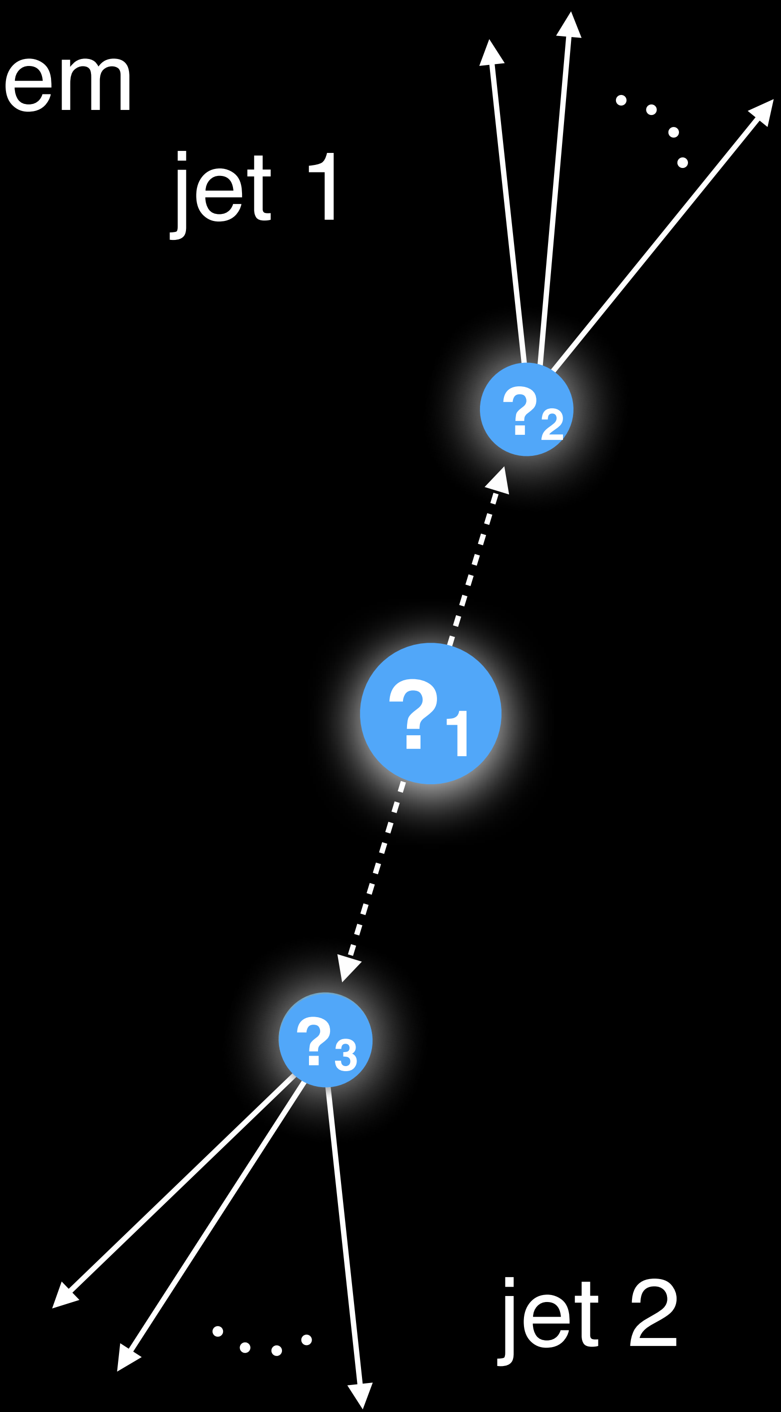


Fin.

Example: two-jet search

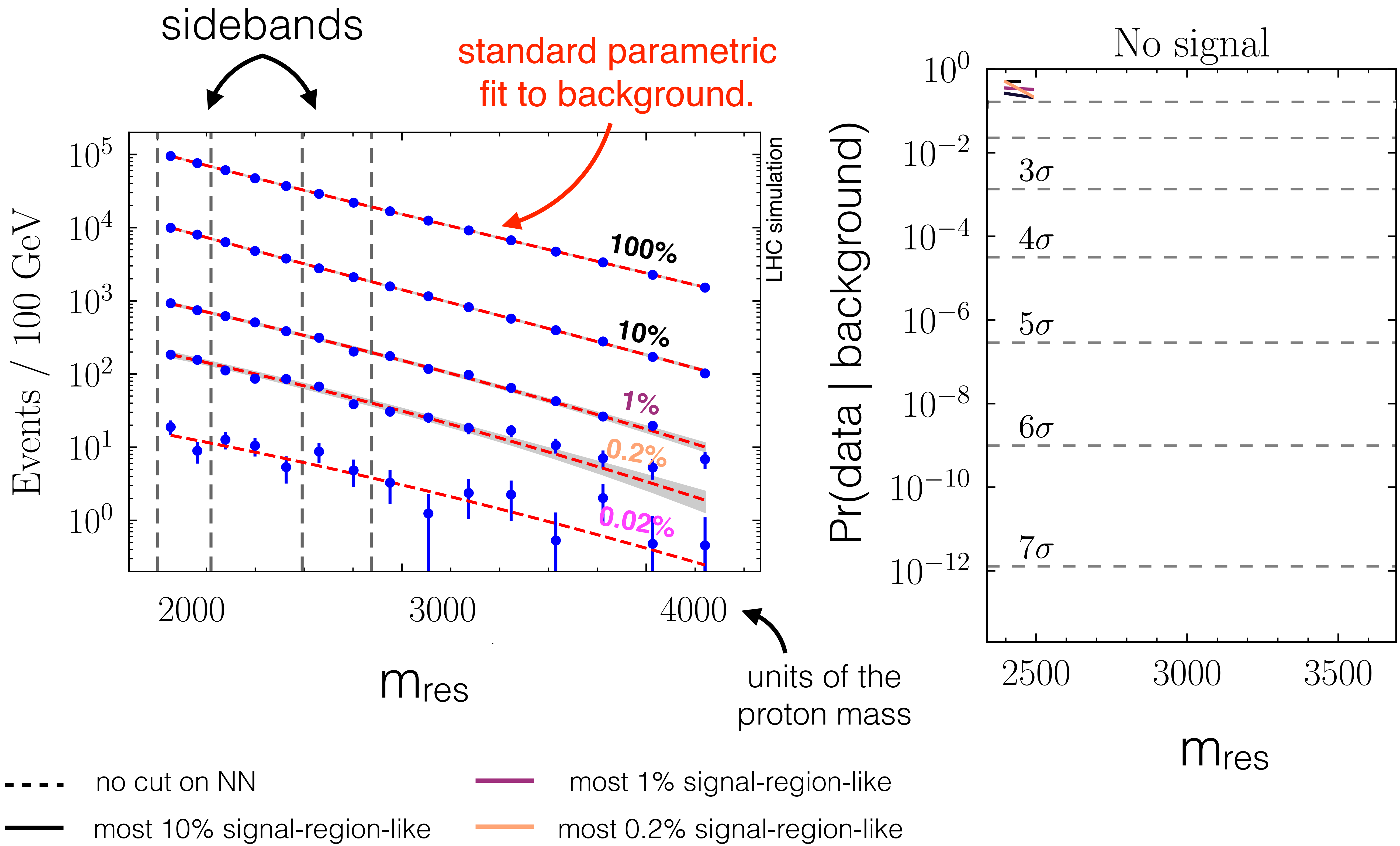


m_{res} = mass of two-jet system

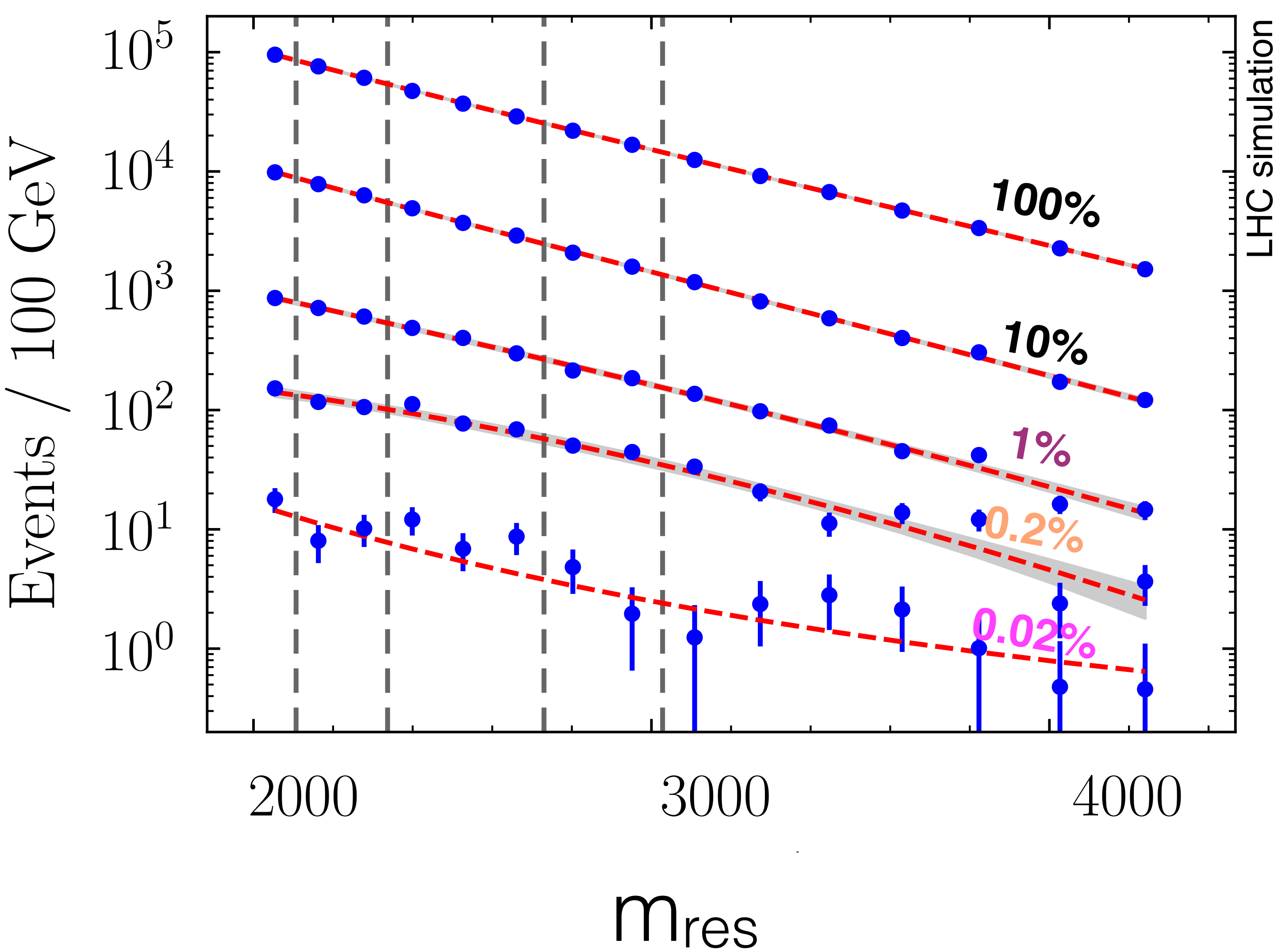


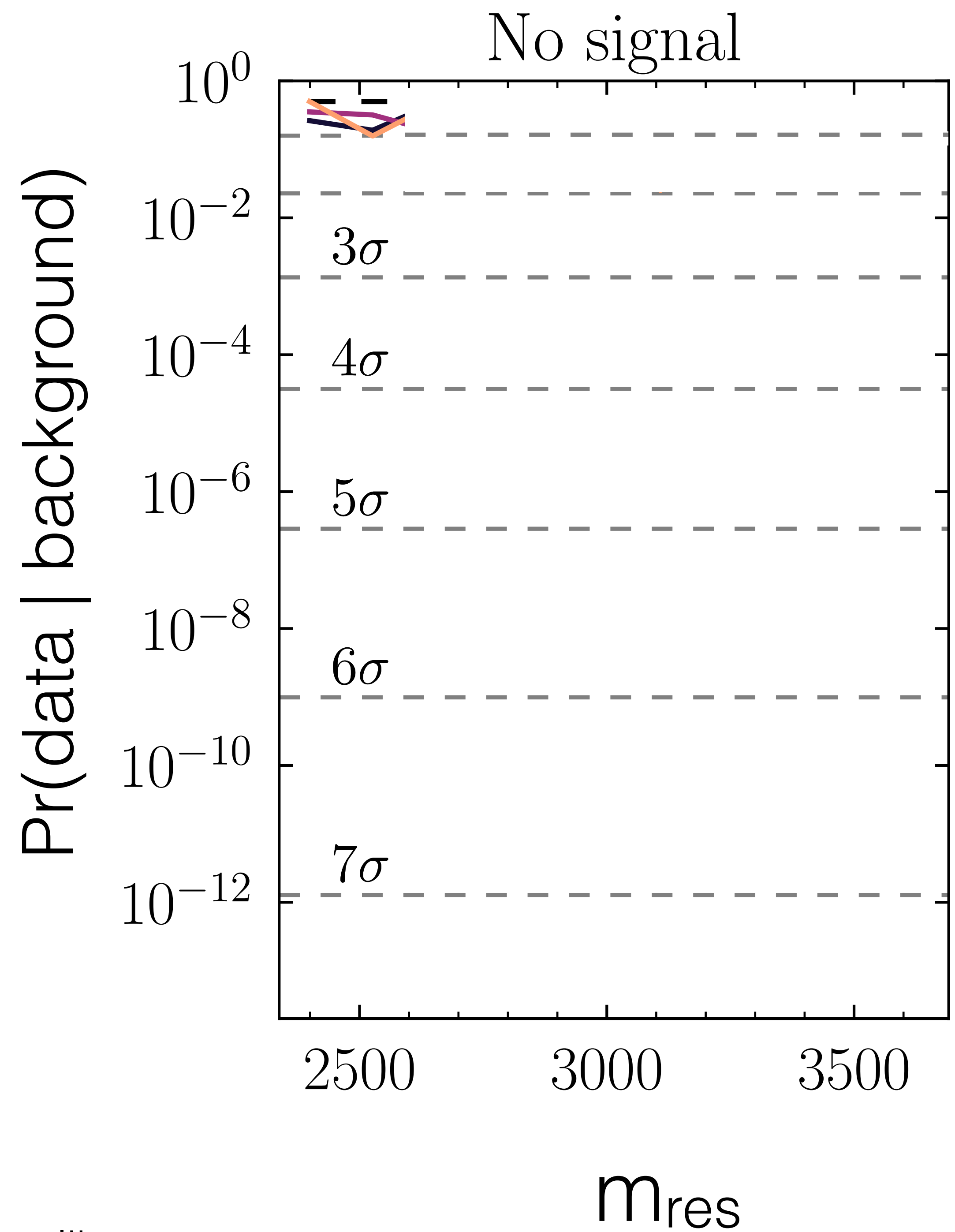
y = many features of the two jets

Example: two-jet search



Example: two-jet search





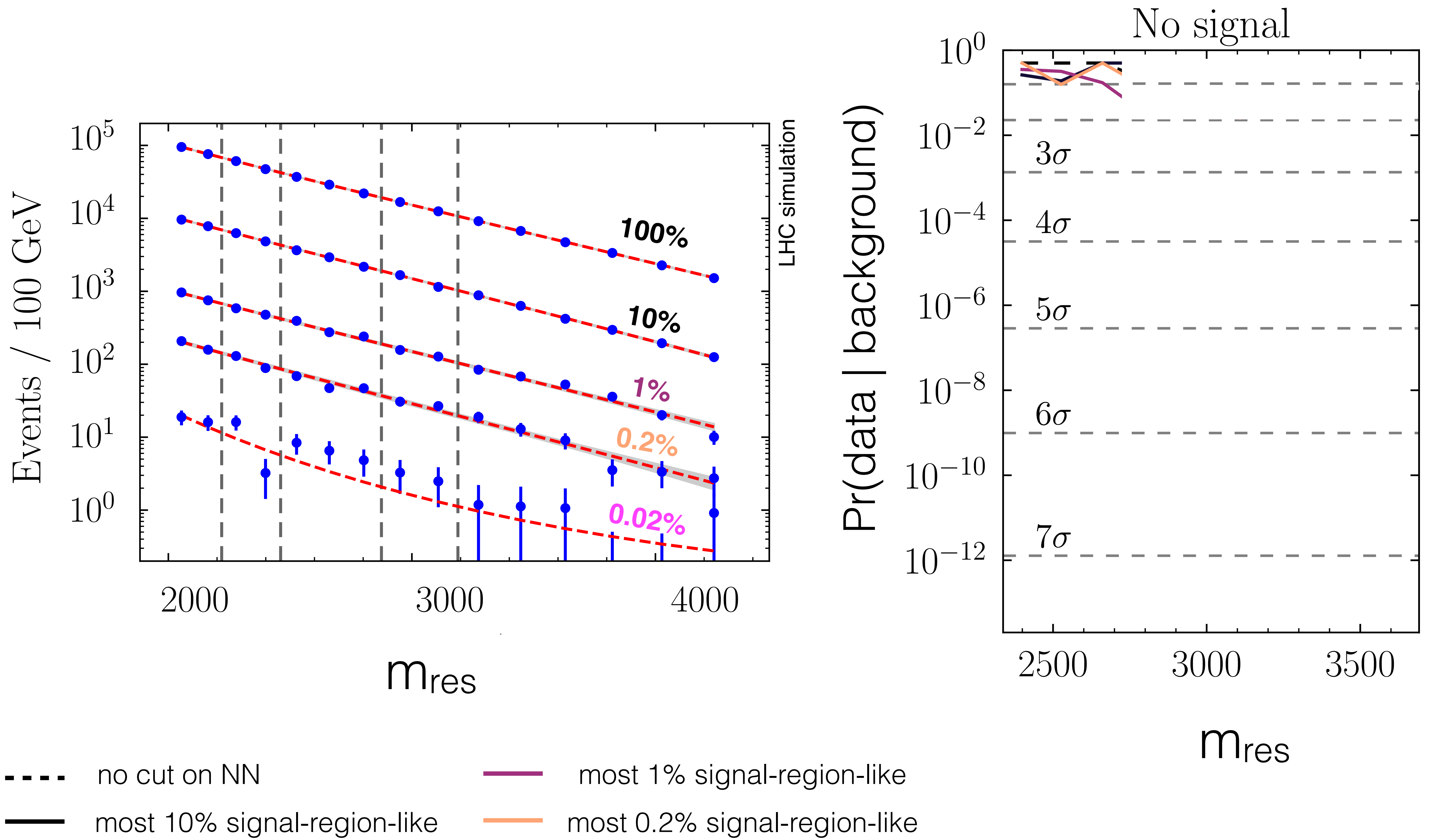
no cut on NN

most 10% signal-region-like

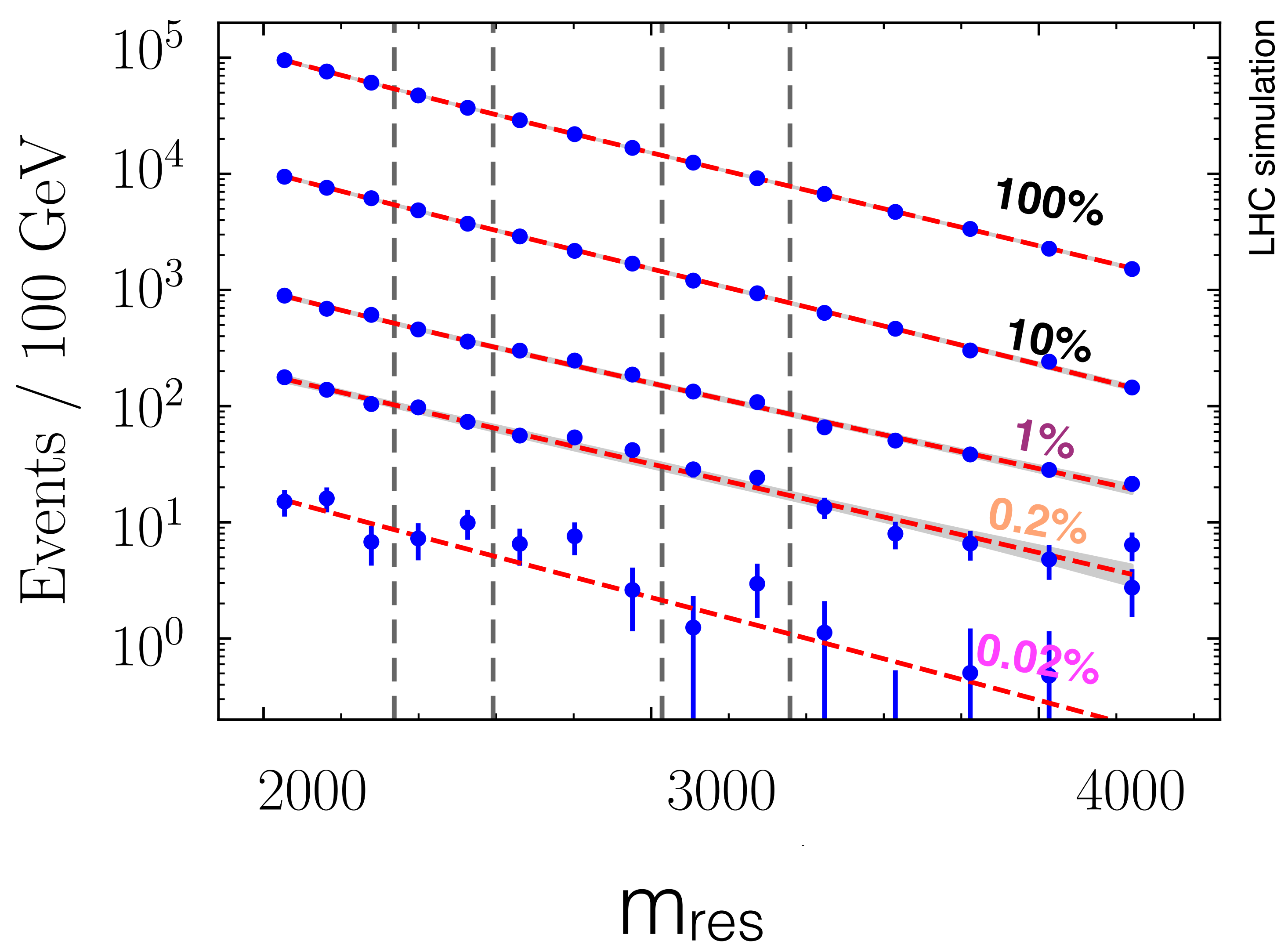
most 1% signal-region-like

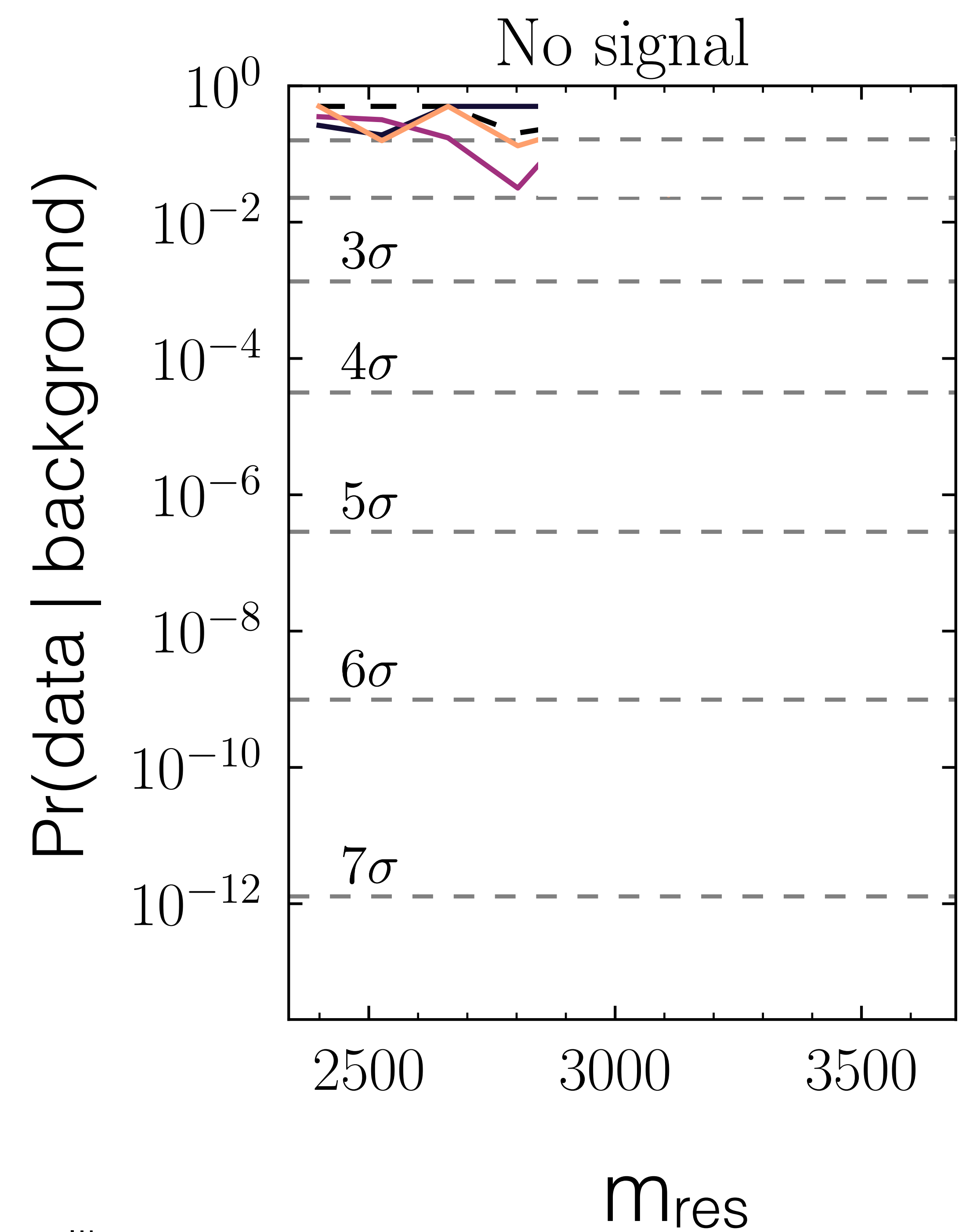
most 0.2% signal-region-like

Example: two-jet search



Example: two-jet search





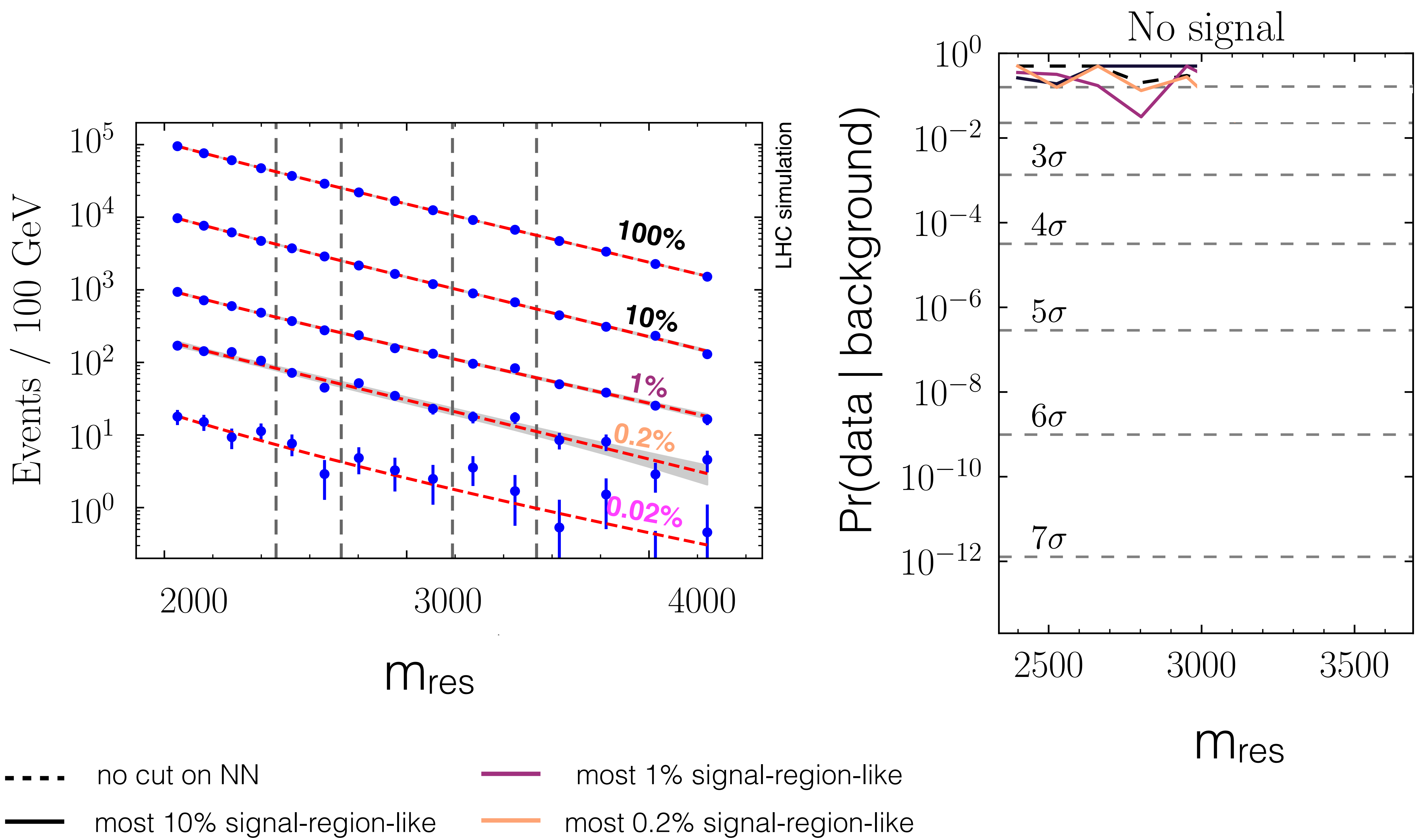
no cut on NN

most 10% signal-region-like

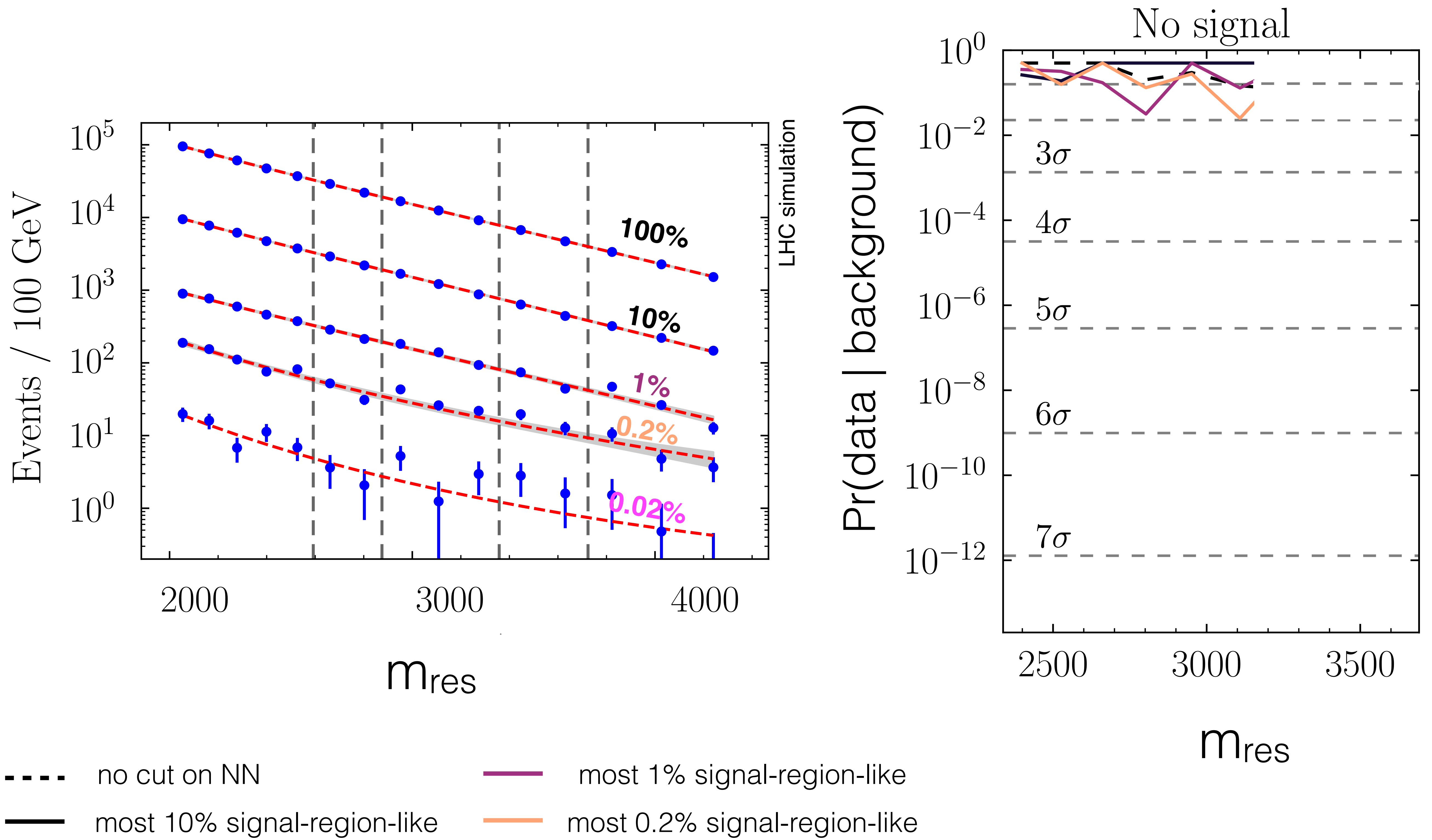
most 1% signal-region-like

most 0.2% signal-region-like

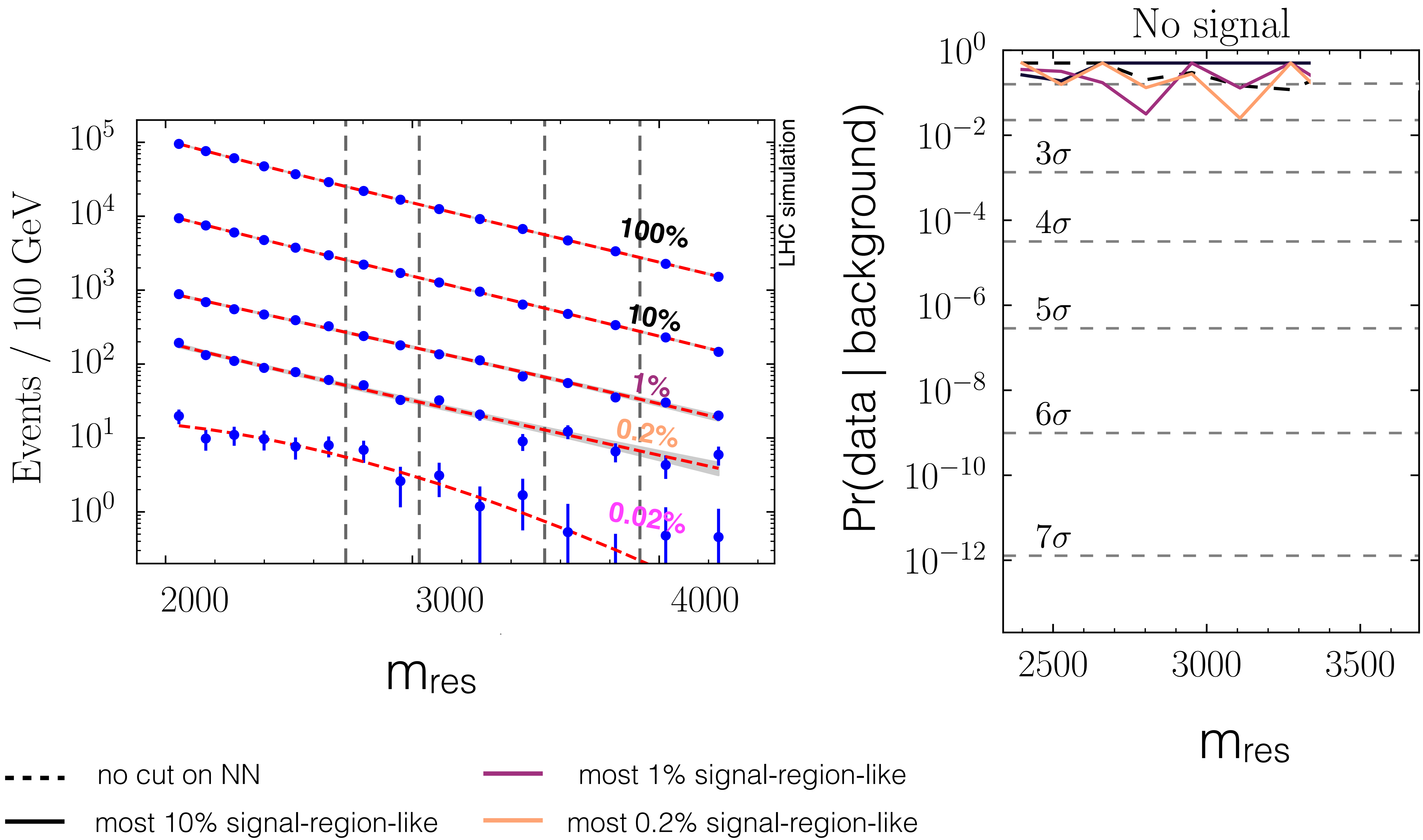
Example: two-jet search



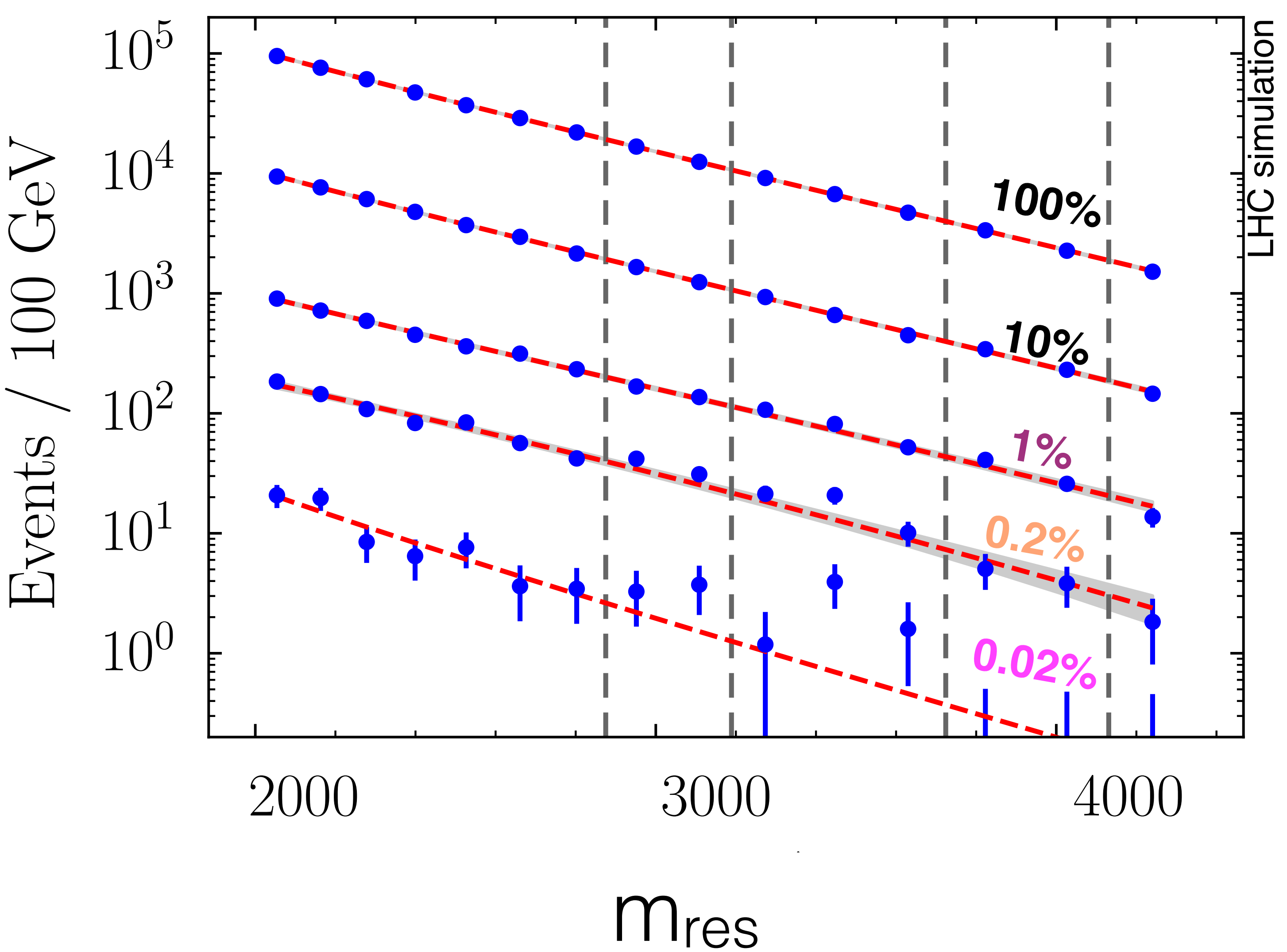
Example: two-jet search

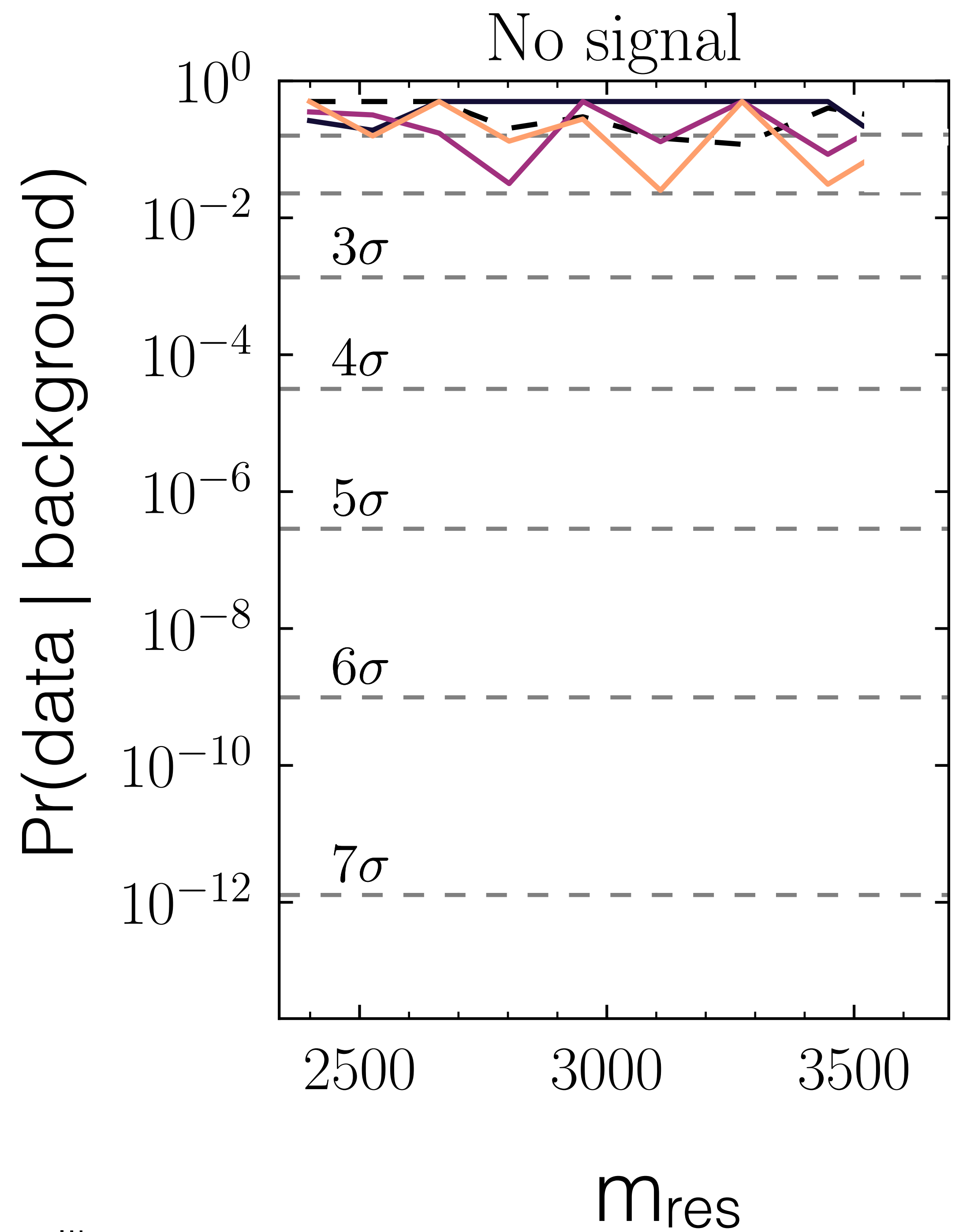


Example: two-jet search

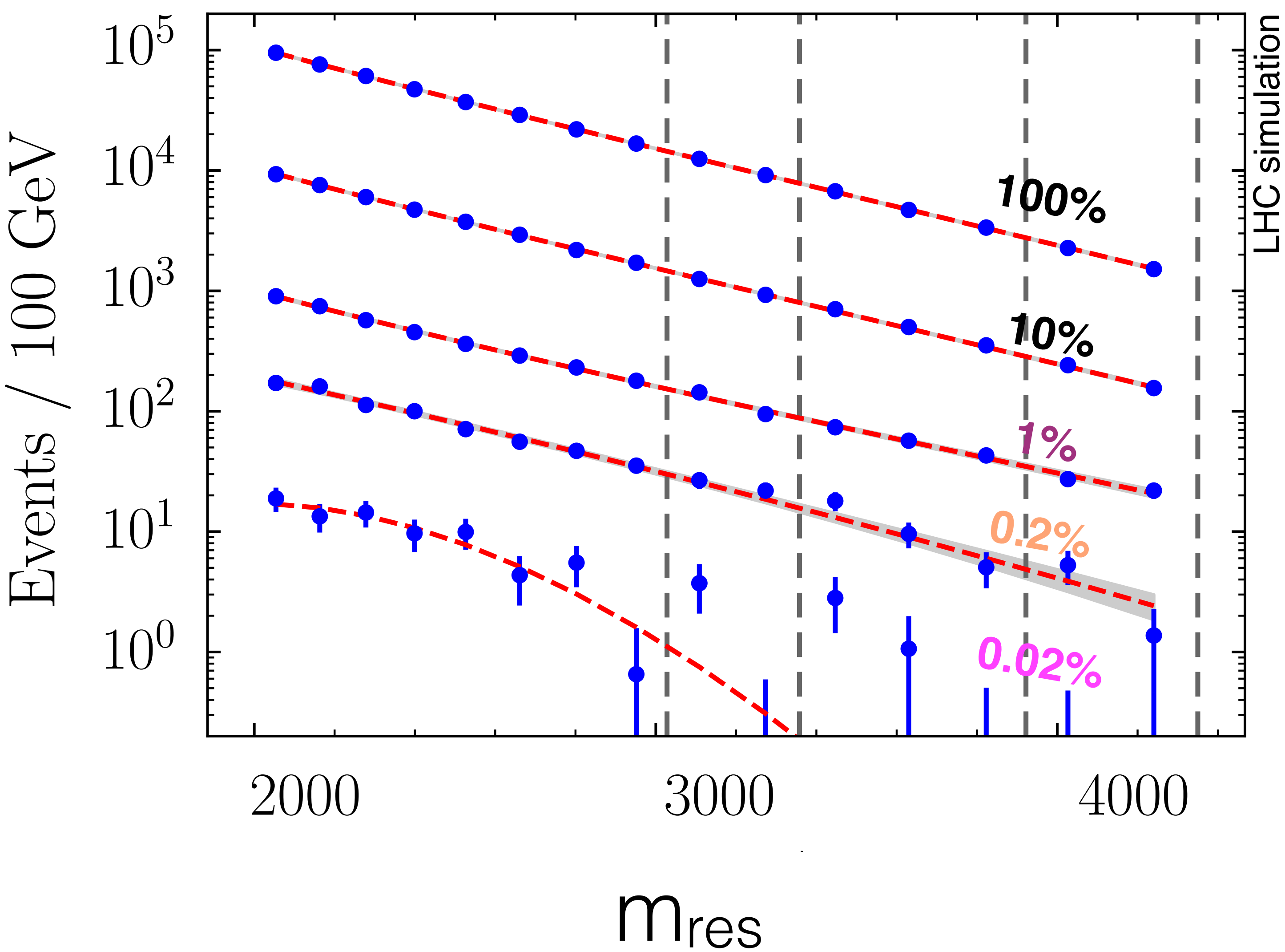


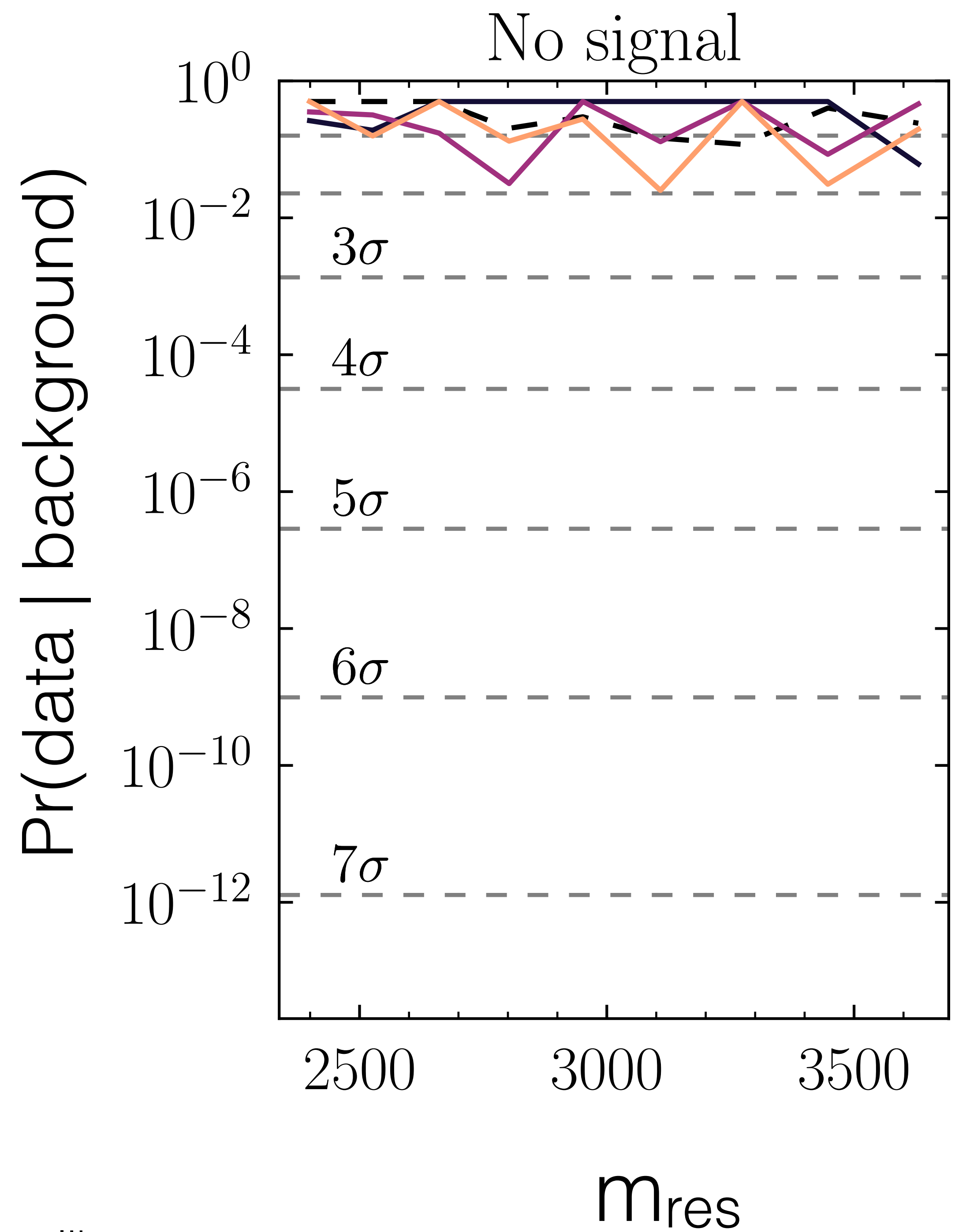
Example: two-jet search





Example: two-jet search





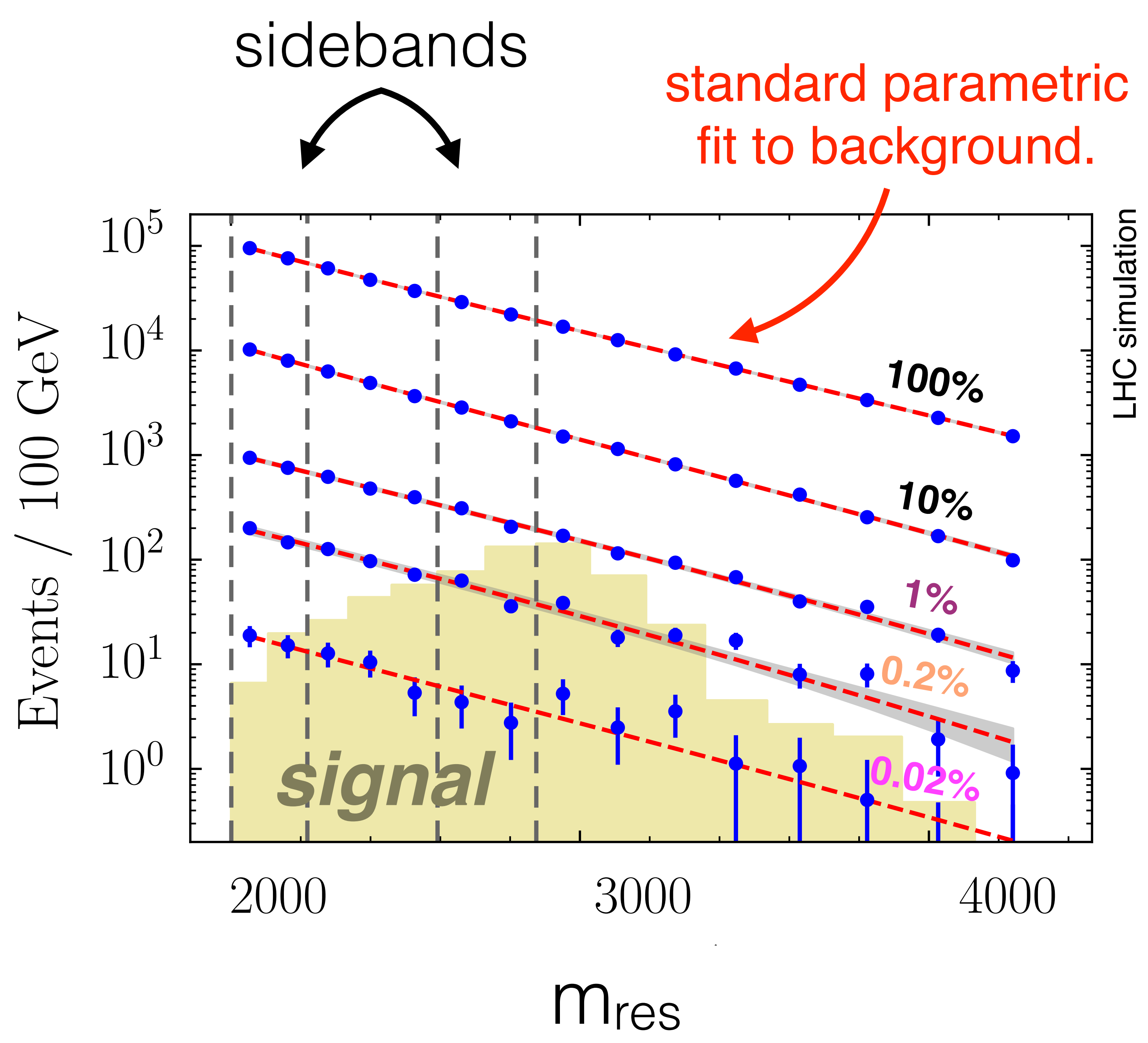
no cut on NN

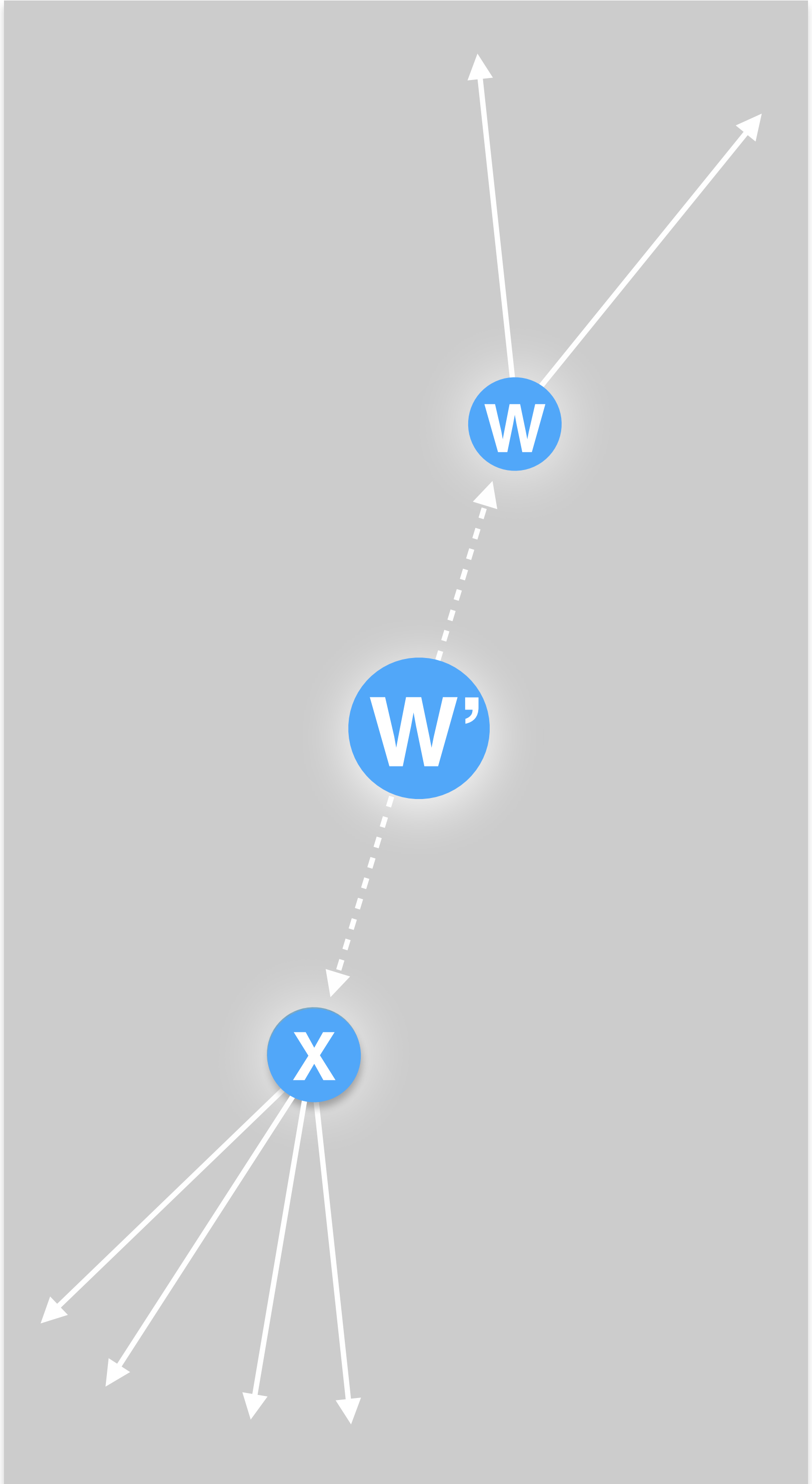
most 10% signal-region-like

most 1% signal-region-like

most 0.2% signal-region-like

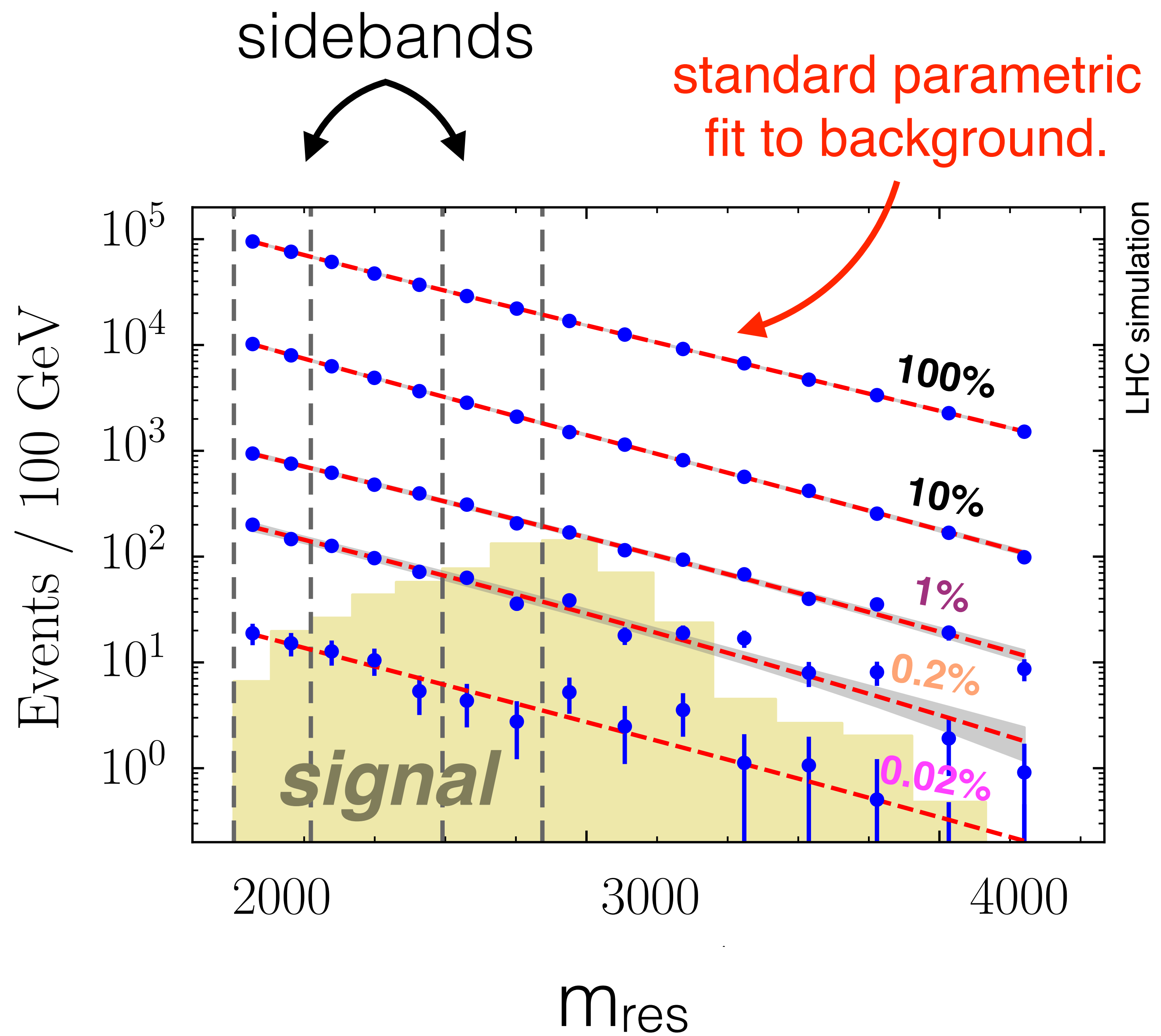
…and when there is a signal?

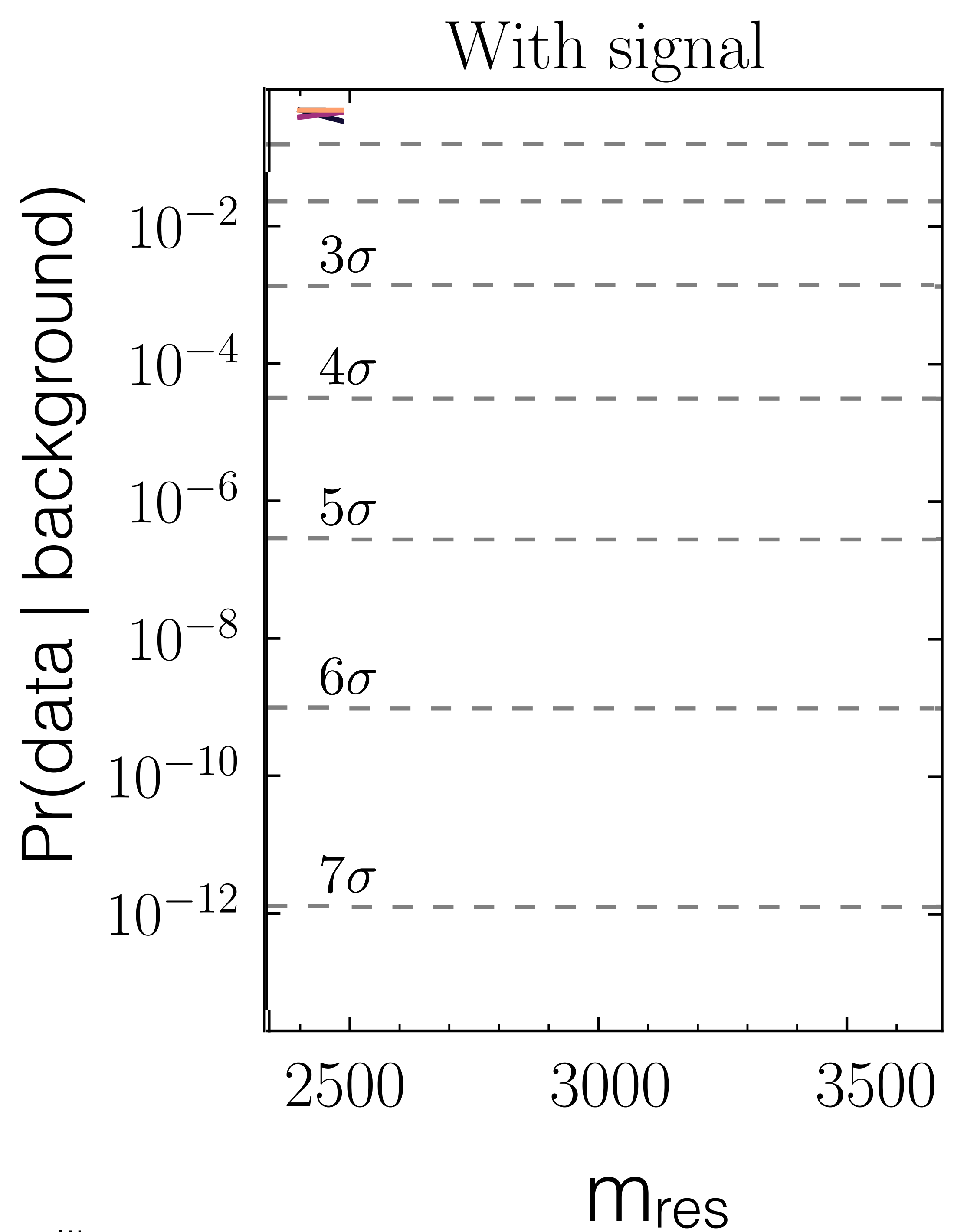




- - - - no cut on NN
— most 10% signal-region-like
— most 1% signal-region-like
— most 0.2% signal-region-like

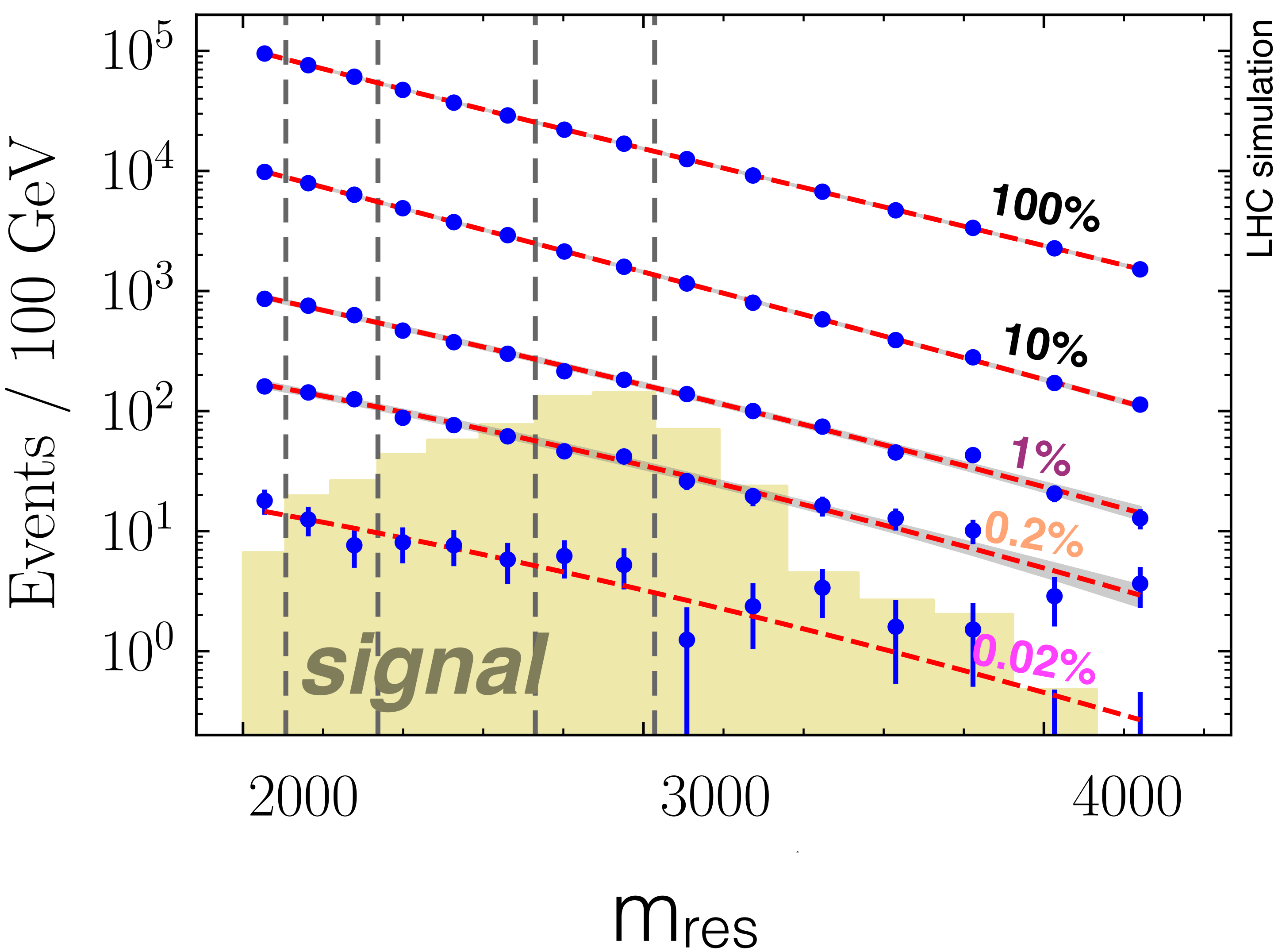
…and when there is a signal?

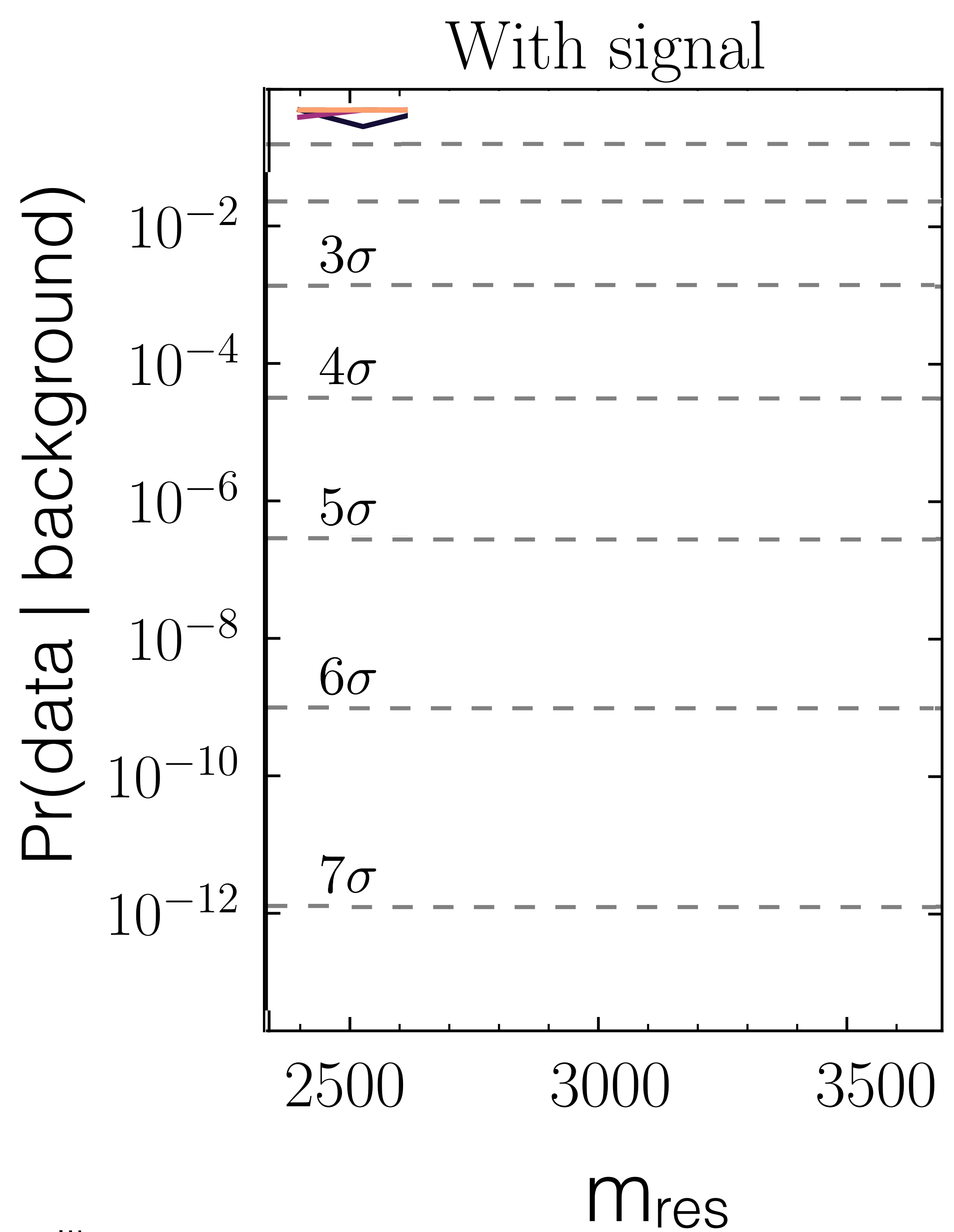




no cut on NN
most 10% signal-region-like
most 1% signal-region-like
most 0.2% signal-region-like

…and when there is a signal?





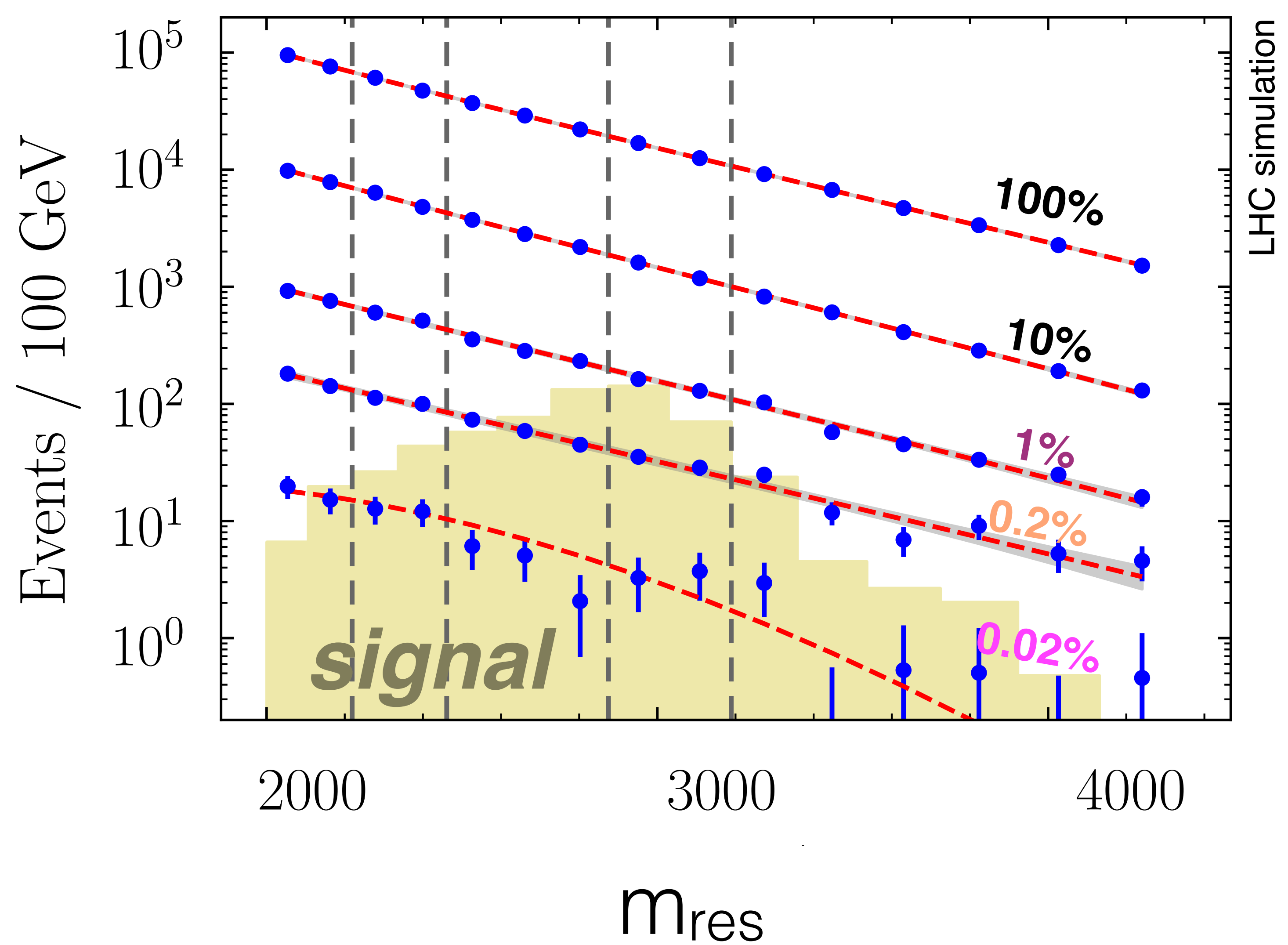
no cut on NN

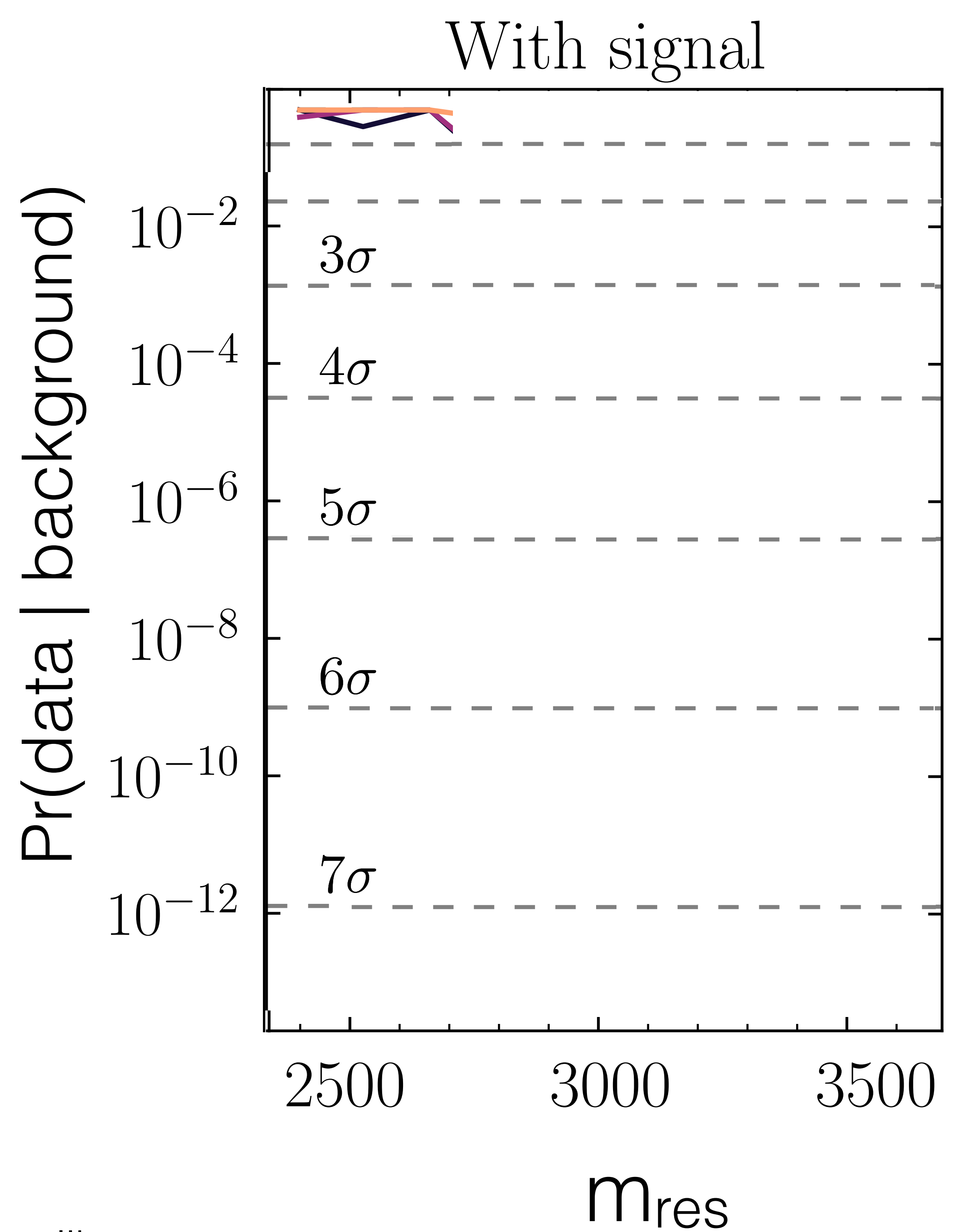
most 10% signal-region-like

most 1% signal-region-like

most 0.2% signal-region-like

…and when there is a signal?





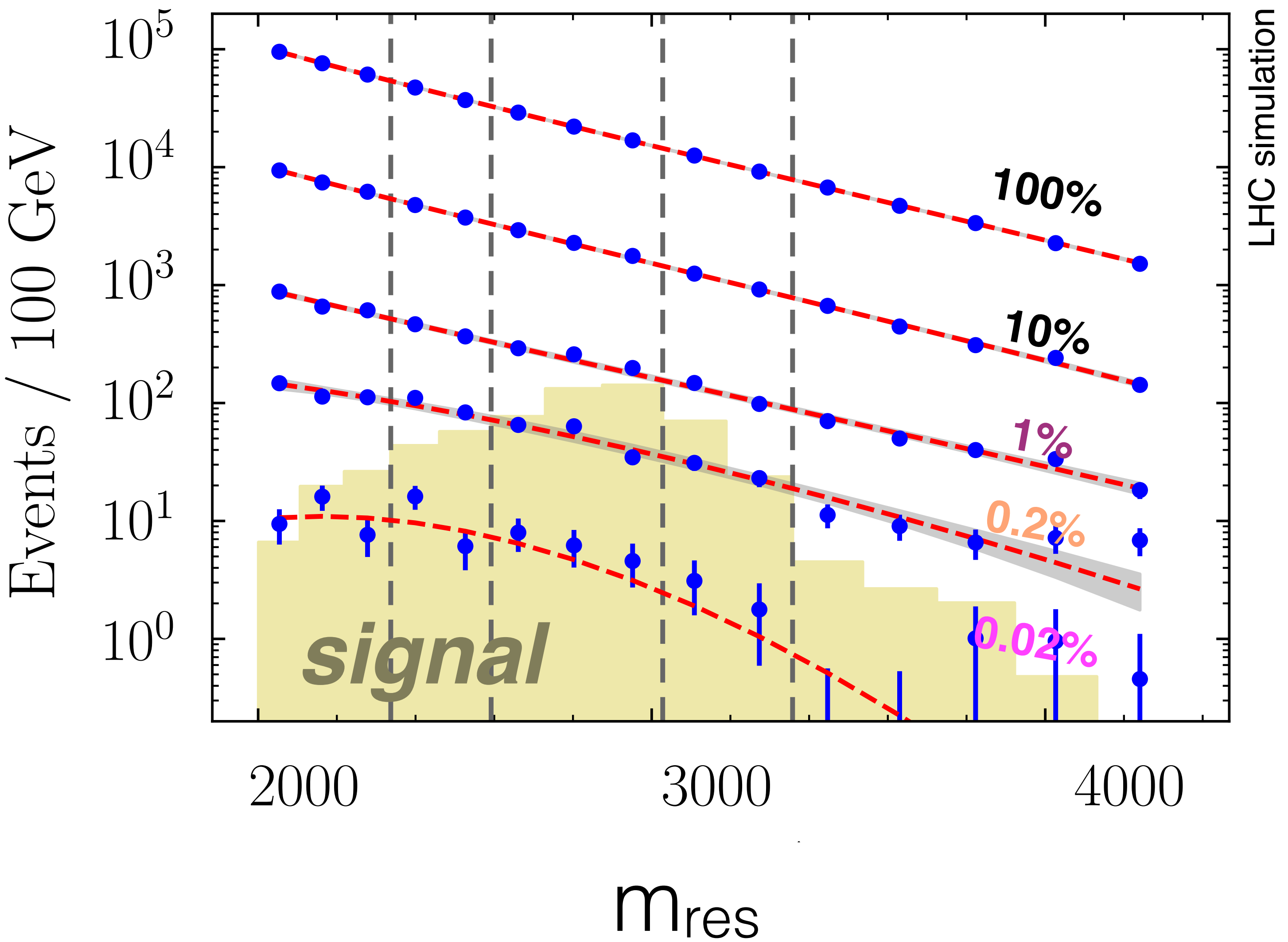
no cut on NN

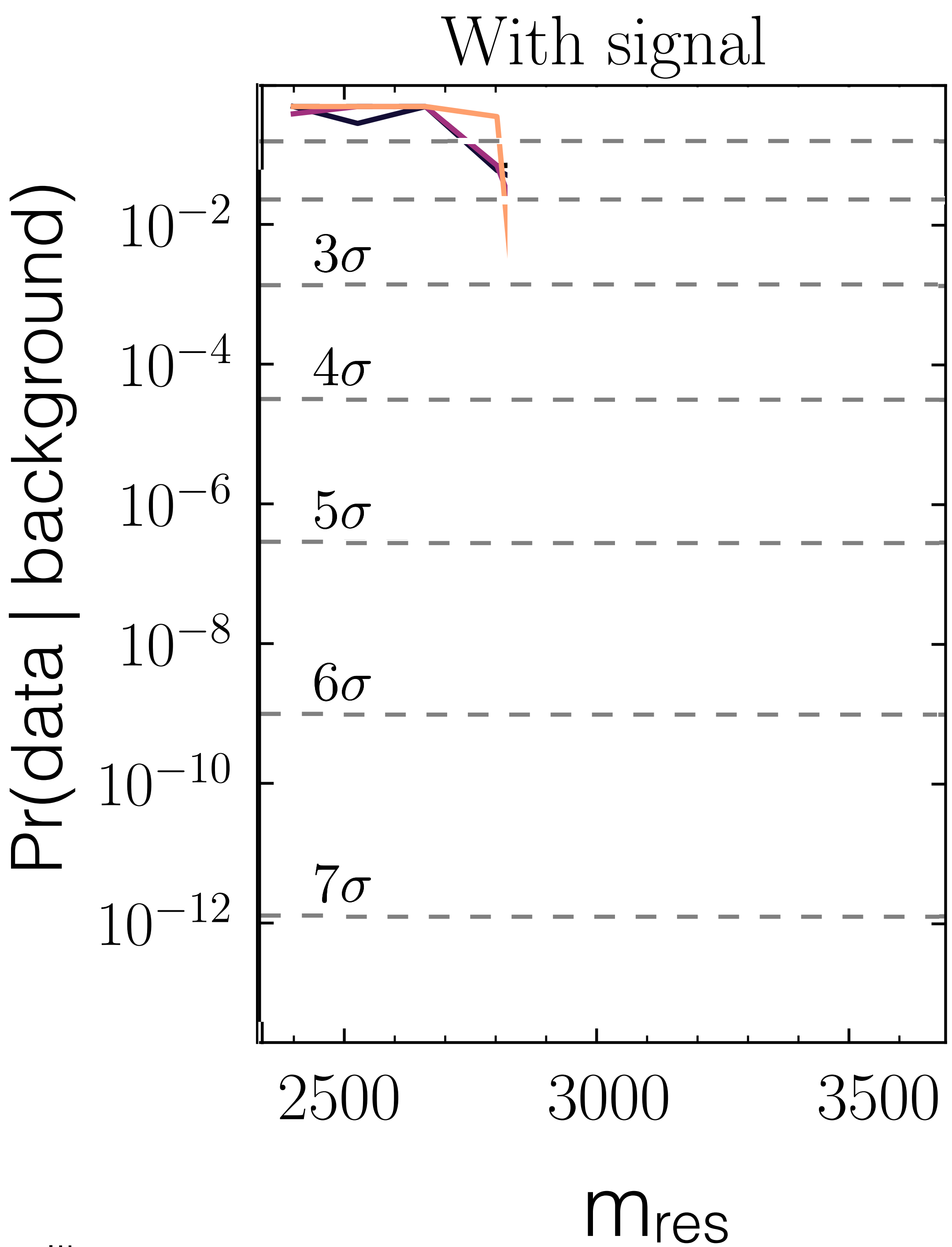
most 10% signal-region-like

most 1% signal-region-like

most 0.2% signal-region-like

…and when there is a signal?





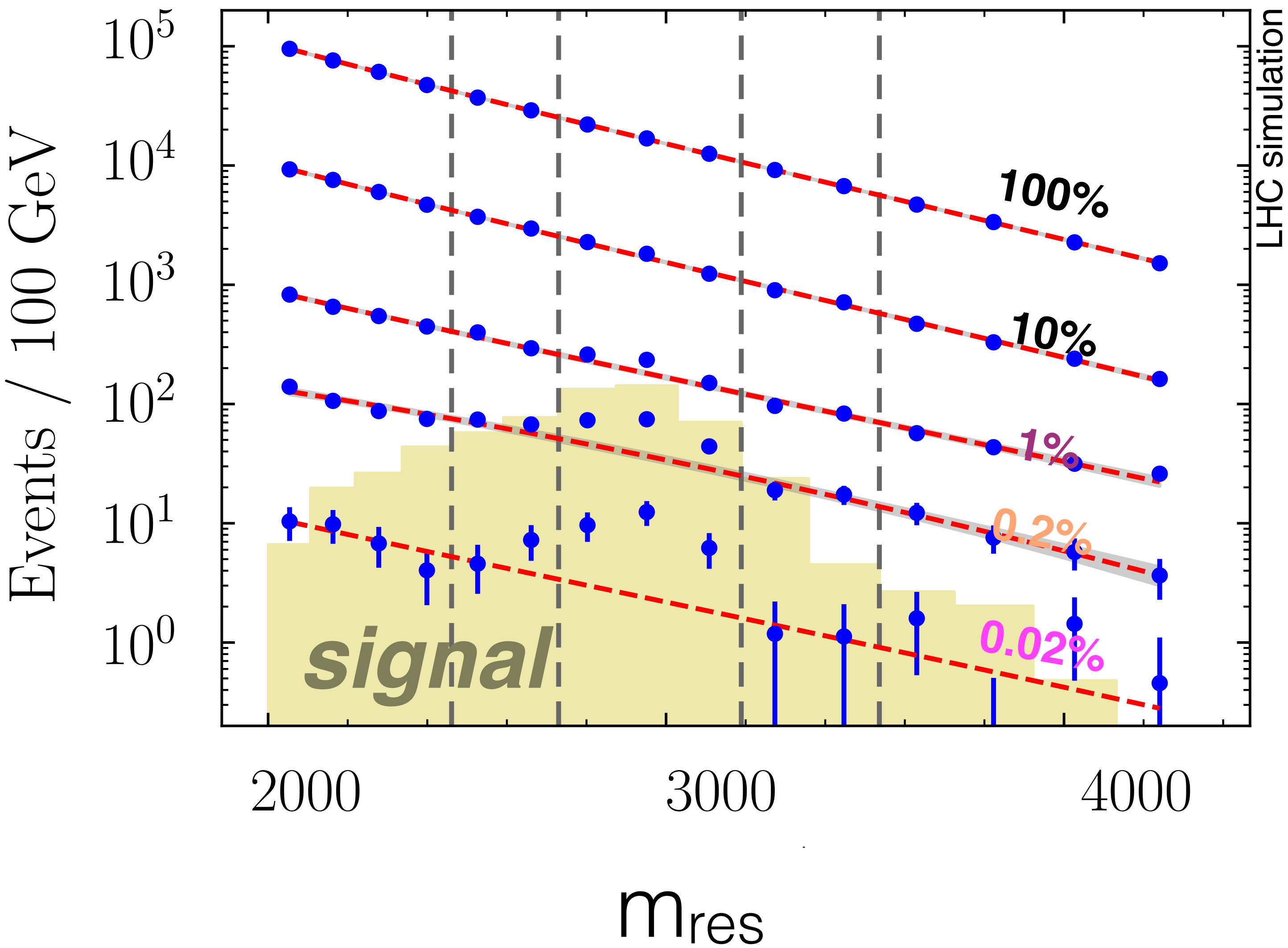
no cut on NN

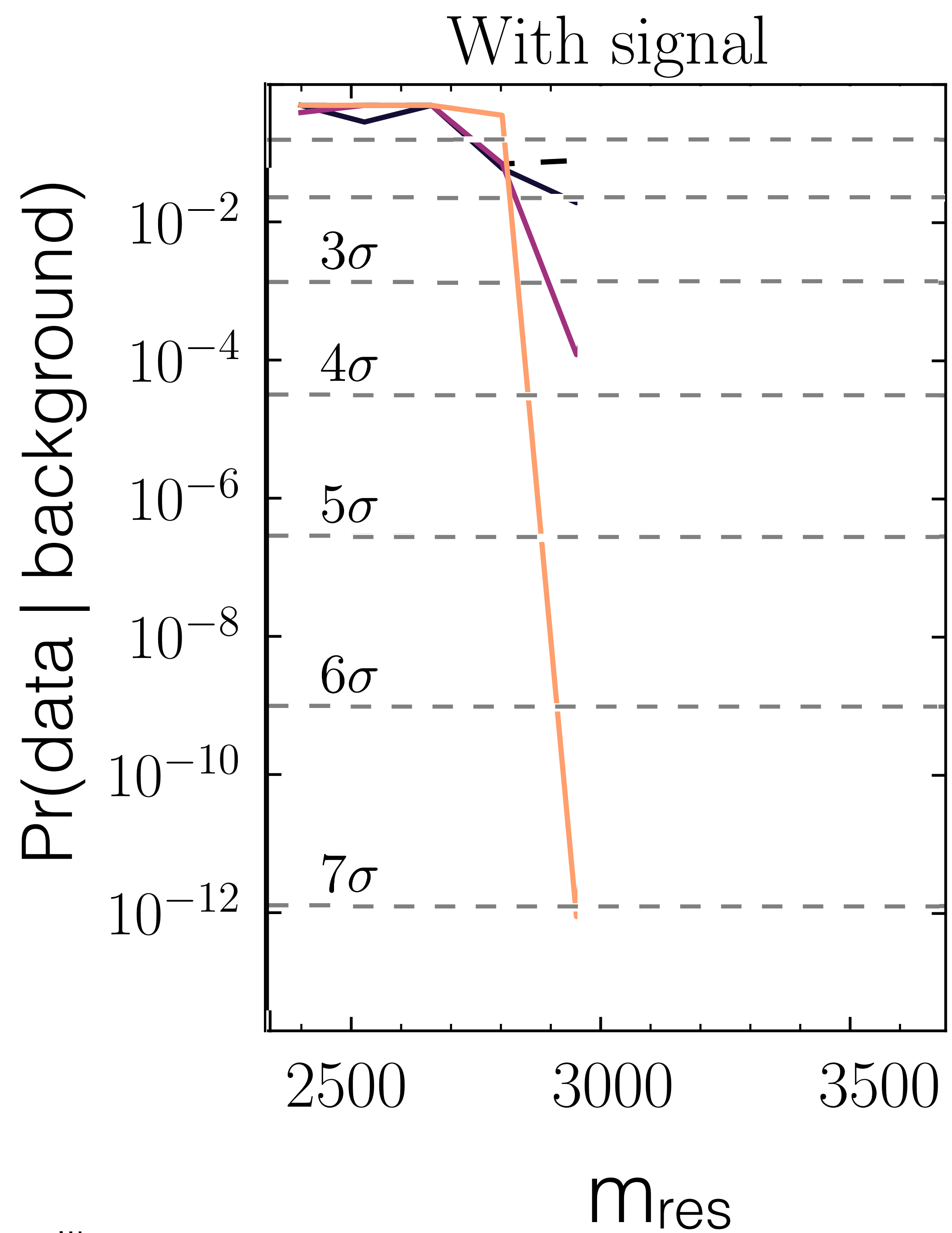
most 10% signal-region-like

most 1% signal-region-like

most 0.2% signal-region-like

...and when there is a signal?





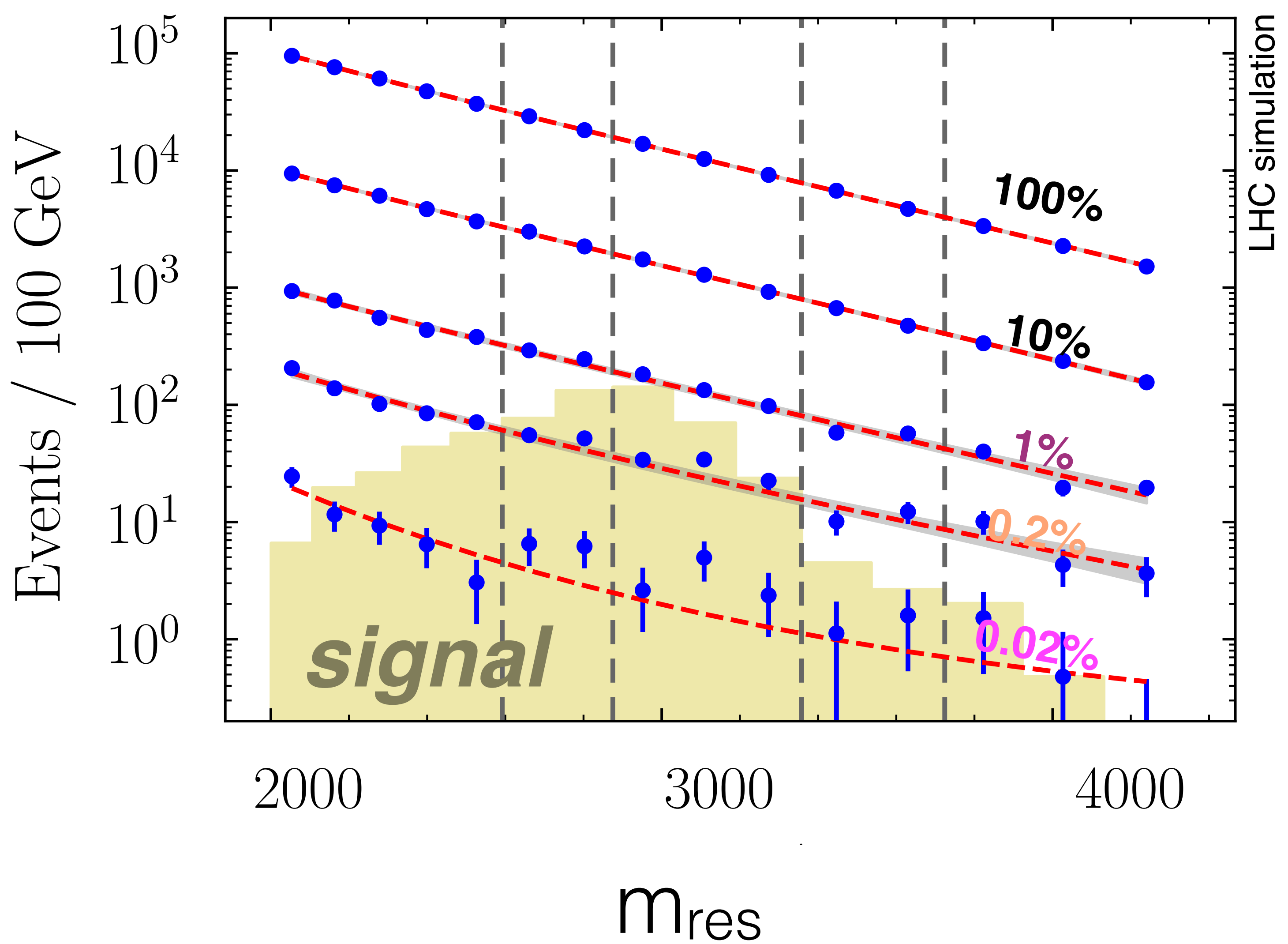
no cut on NN

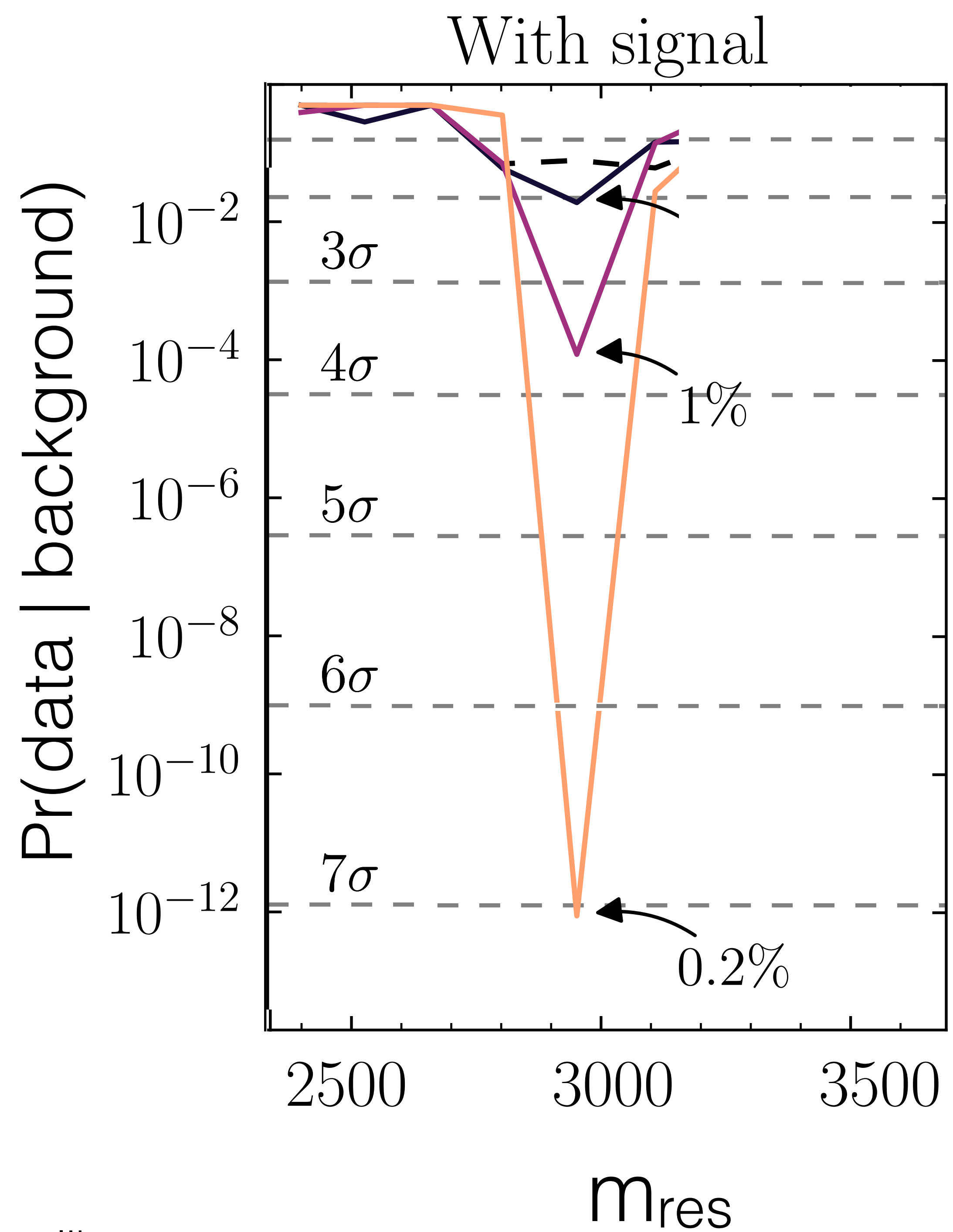
most 10% signal-region-like

most 1% signal-region-like

most 0.2% signal-region-like

…and when there is a signal?





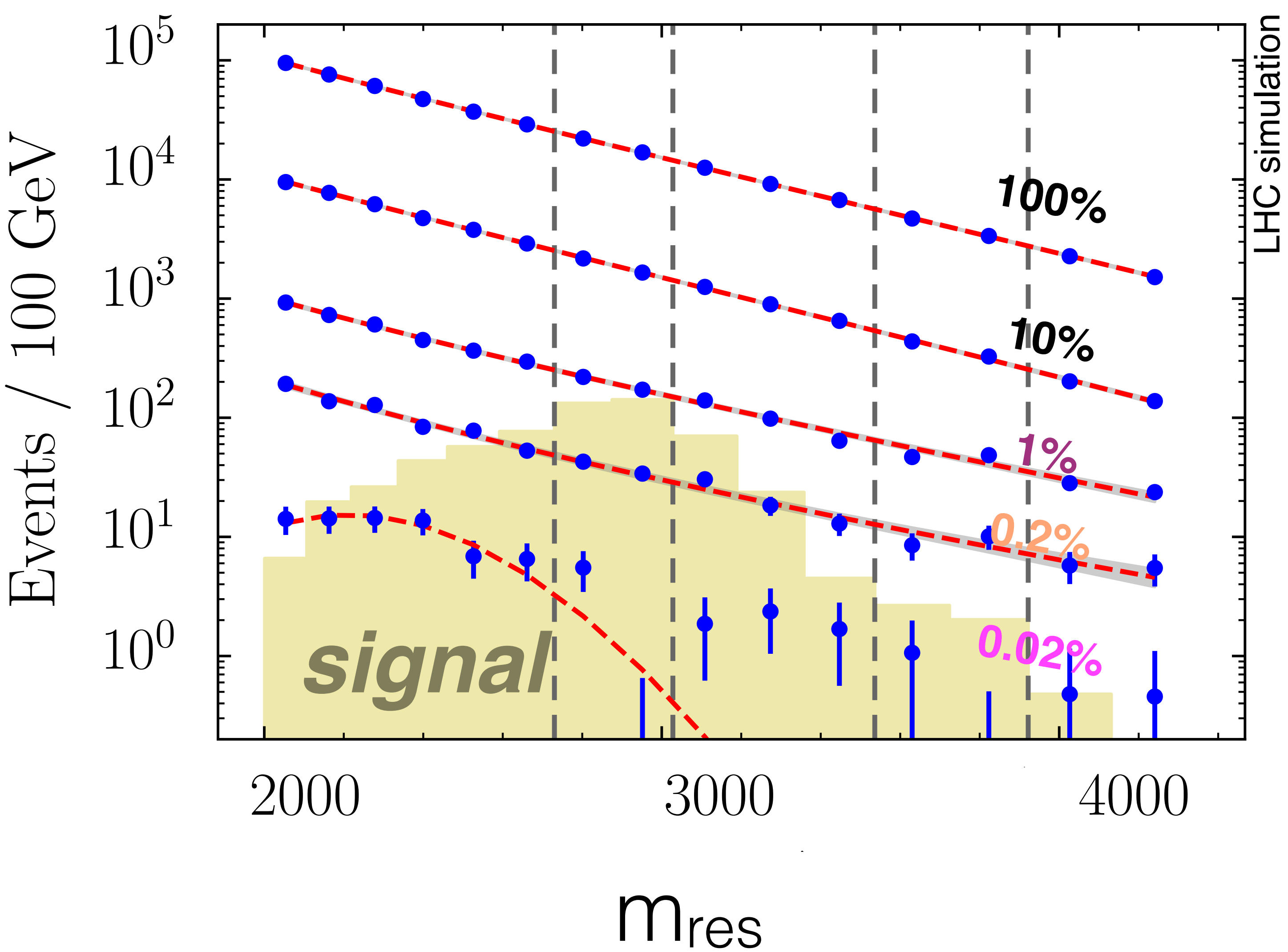
no cut on NN

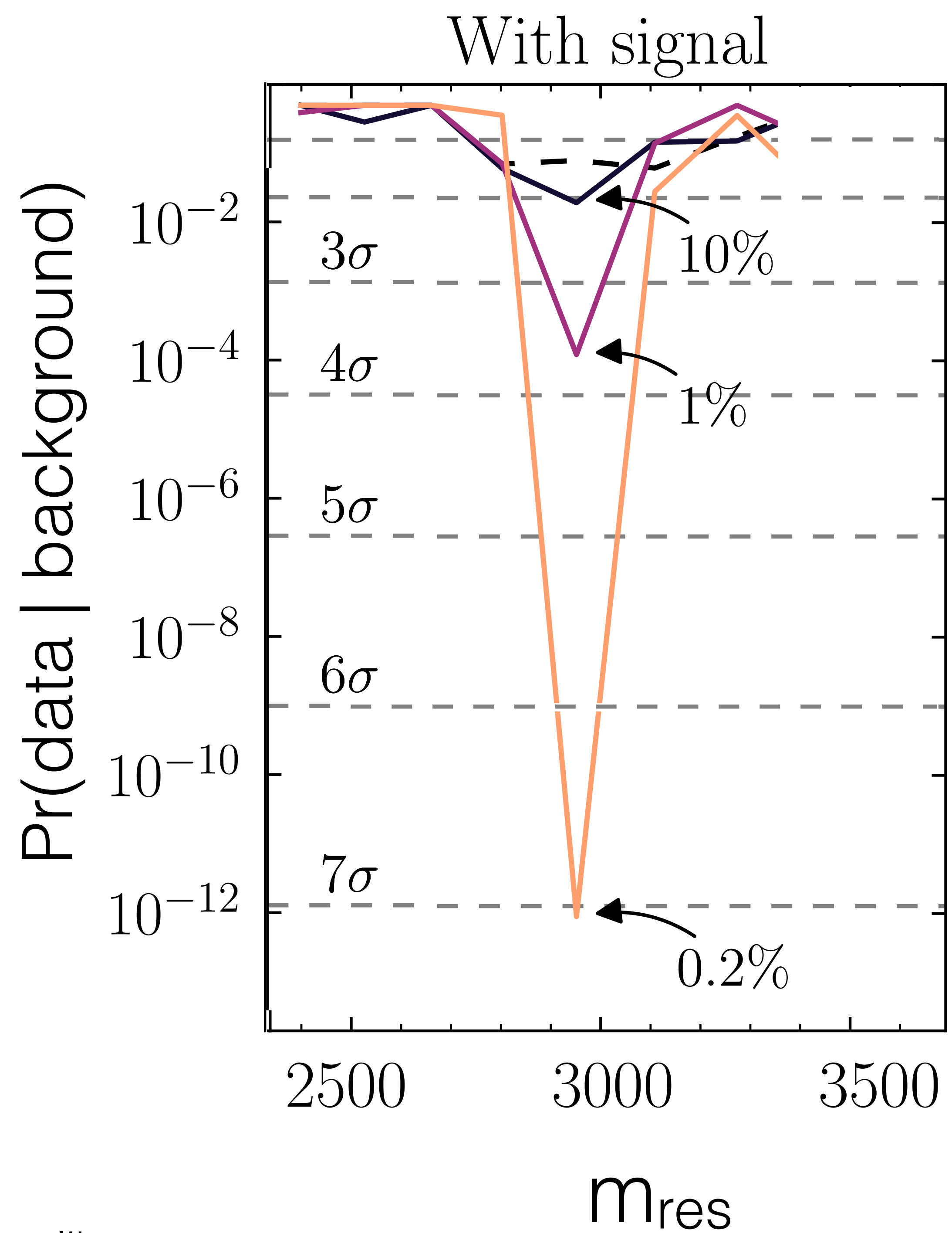
most 10% signal-region-like

most 1% signal-region-like

most 0.2% signal-region-like

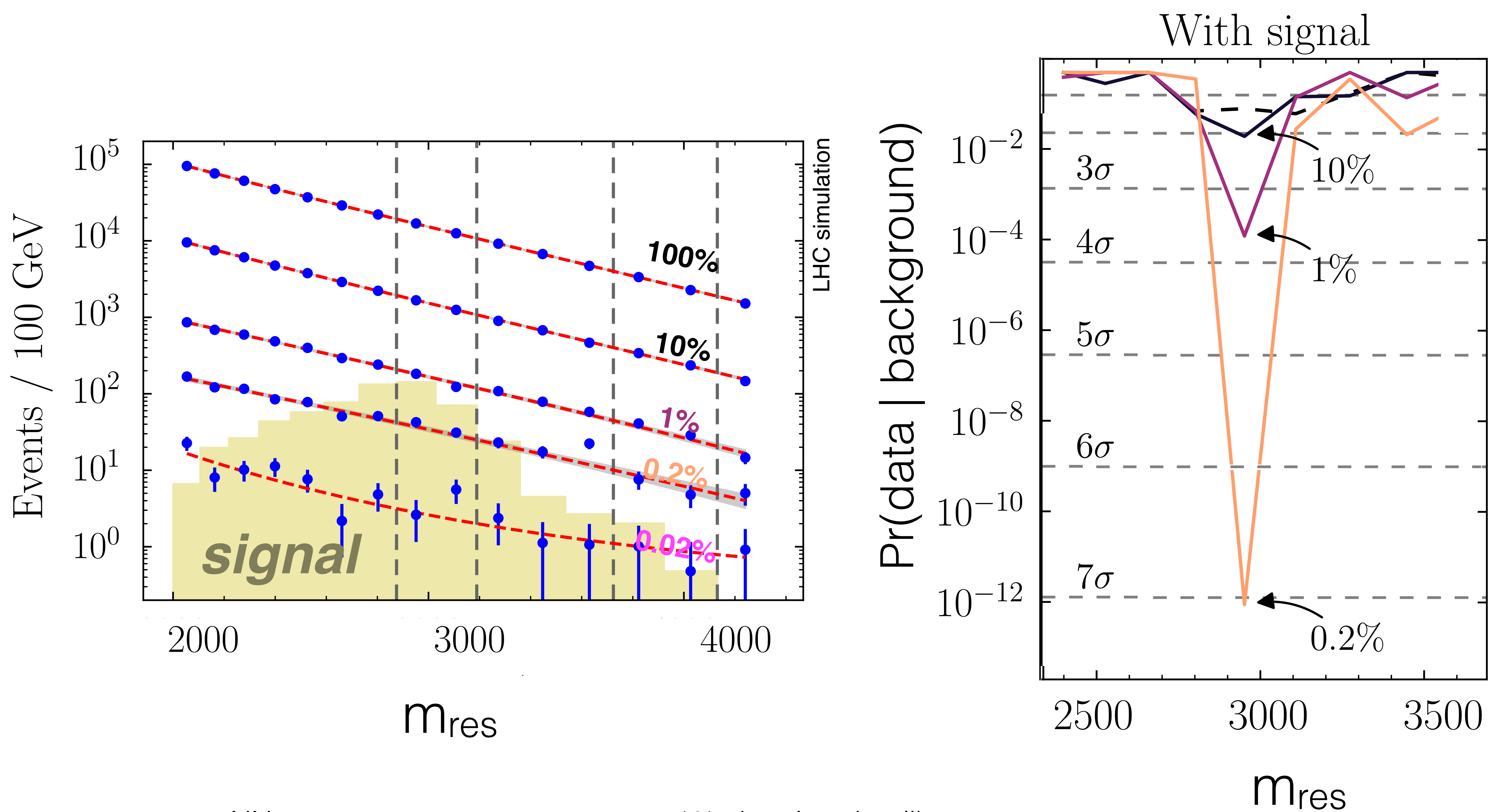
…and when there is a signal?





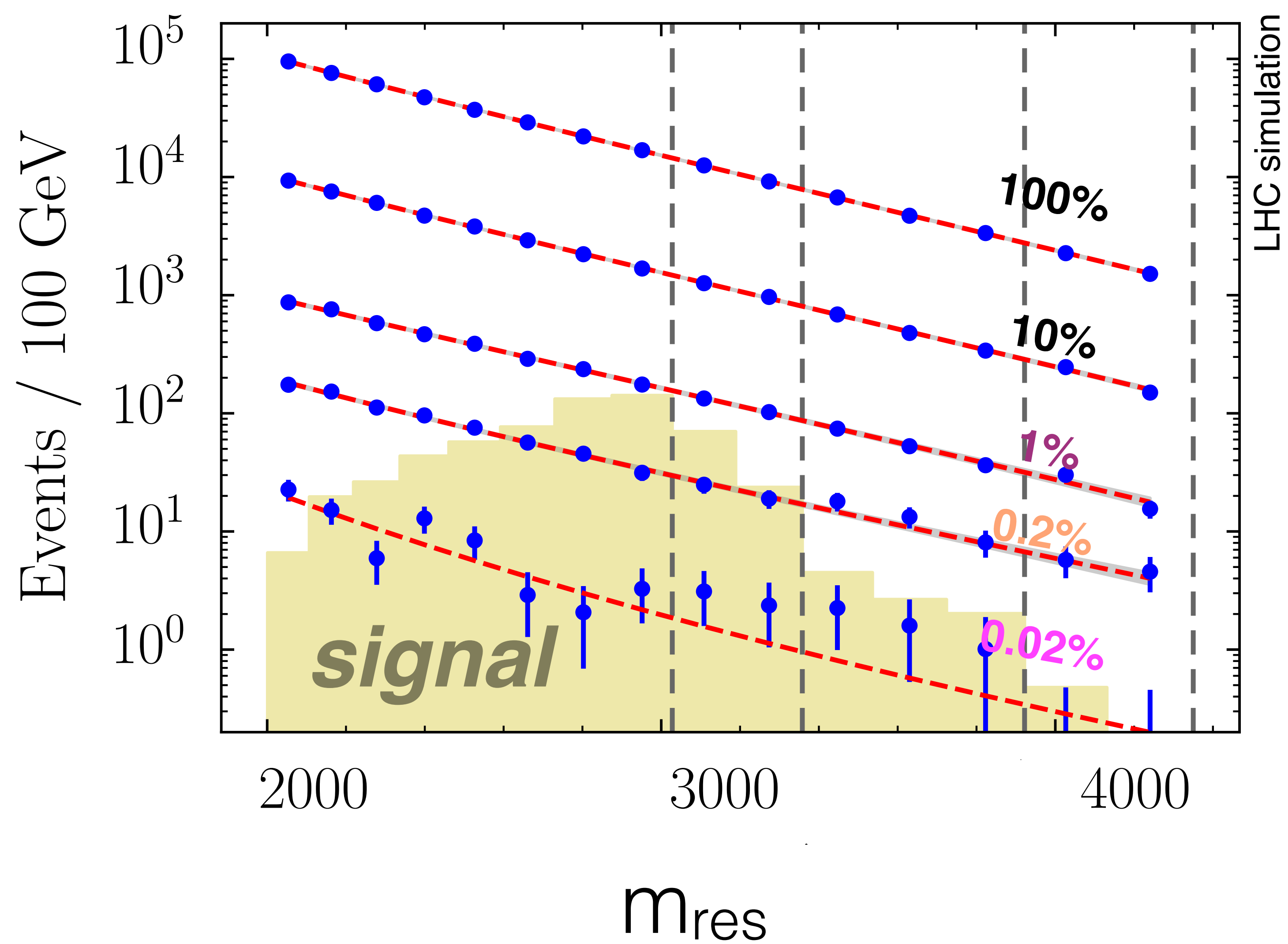
no cut on NN
most 10% signal-region-like
most 1% signal-region-like
most 0.2% signal-region-like

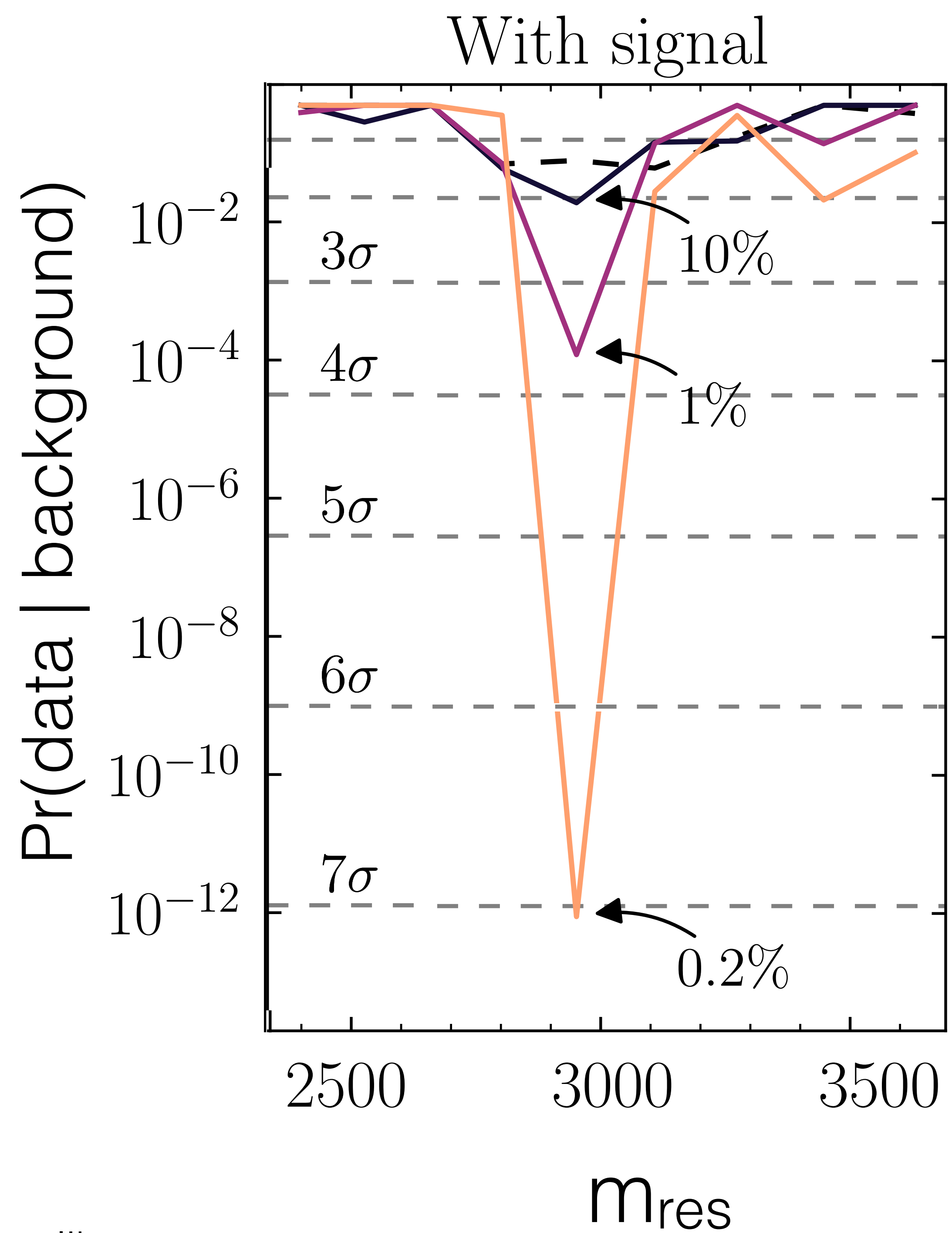
…and when there is a signal?



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

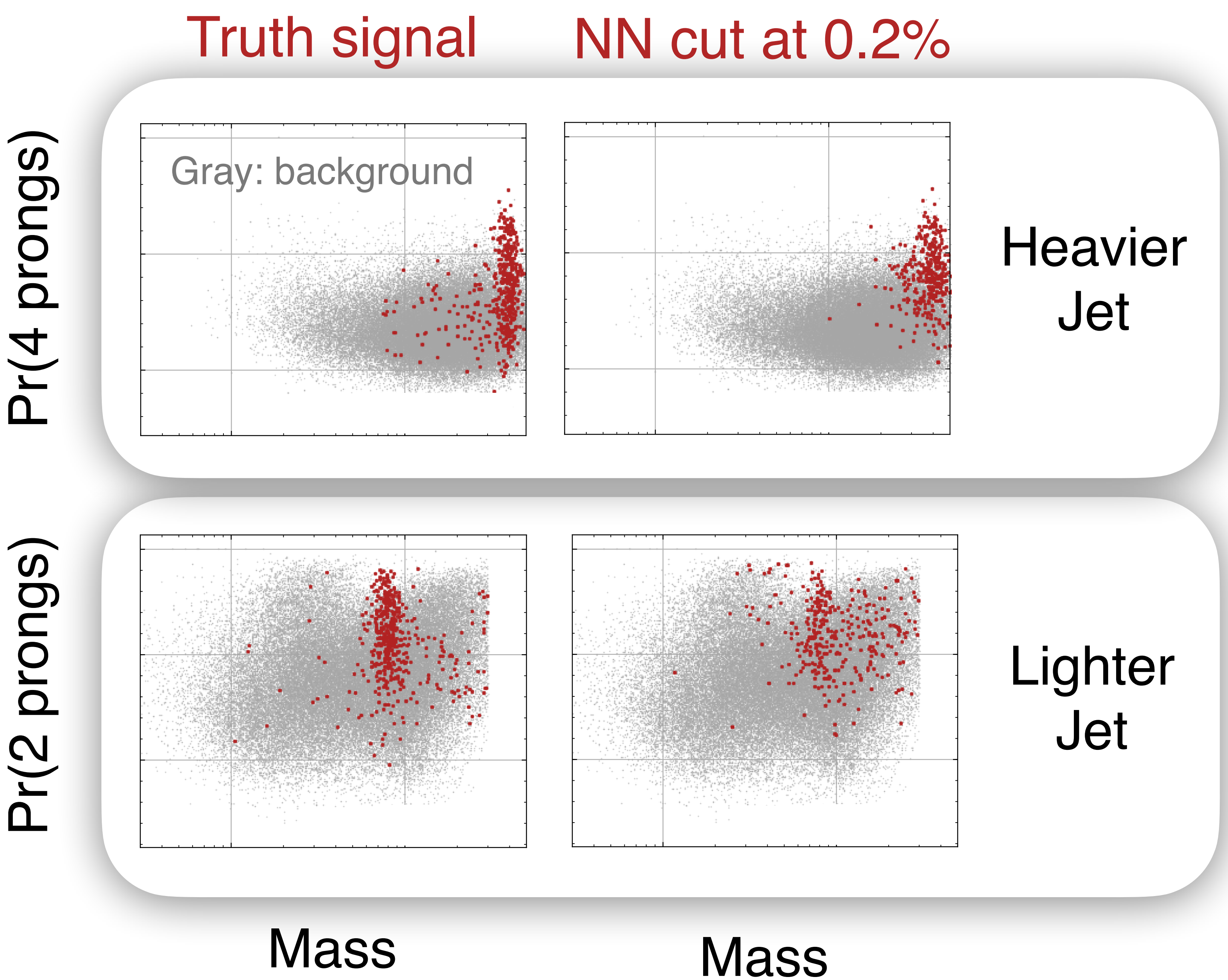
…and when there is a signal?

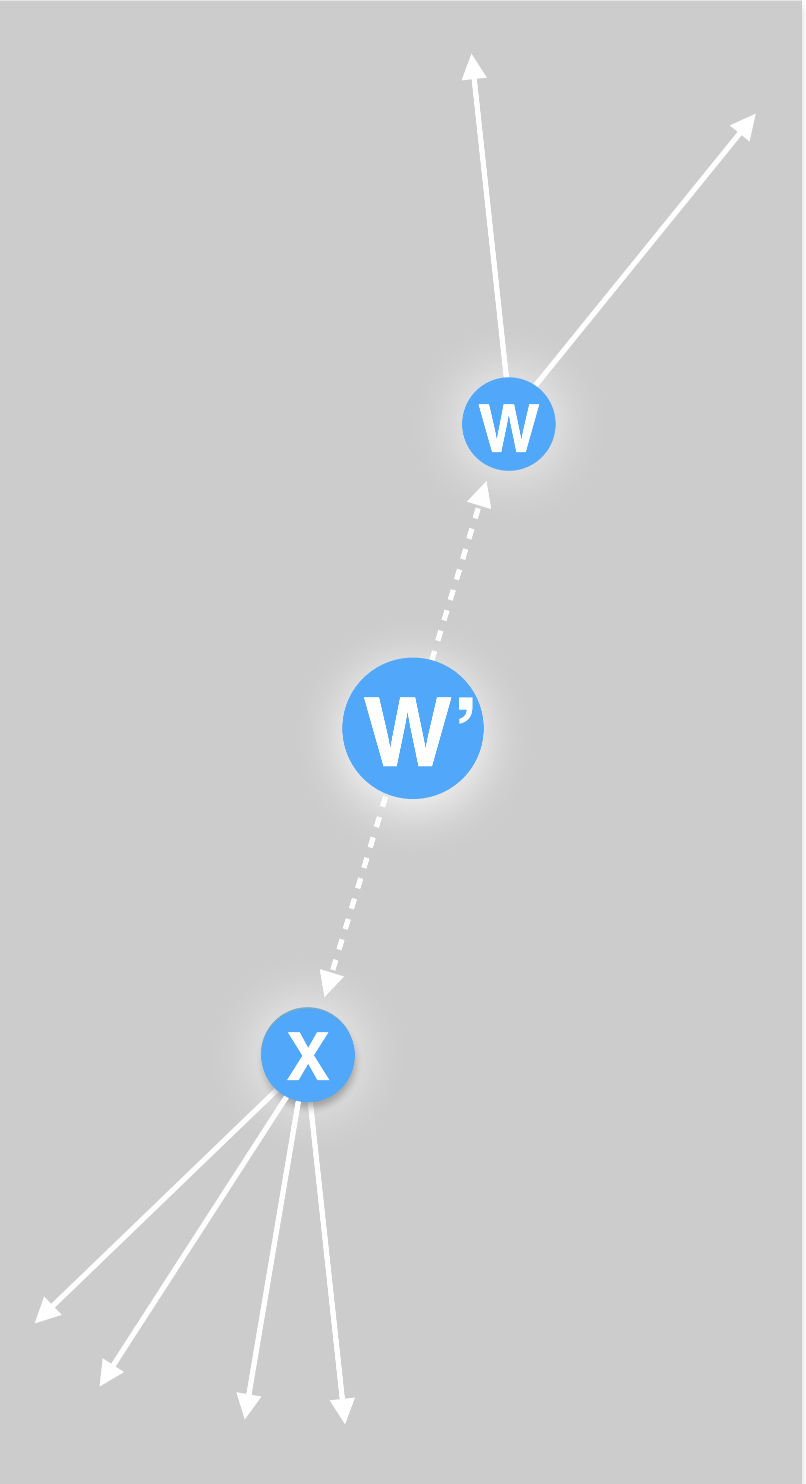




no cut on NN
most 10% signal-region-like
most 1% signal-region-like
most 0.2% signal-region-like

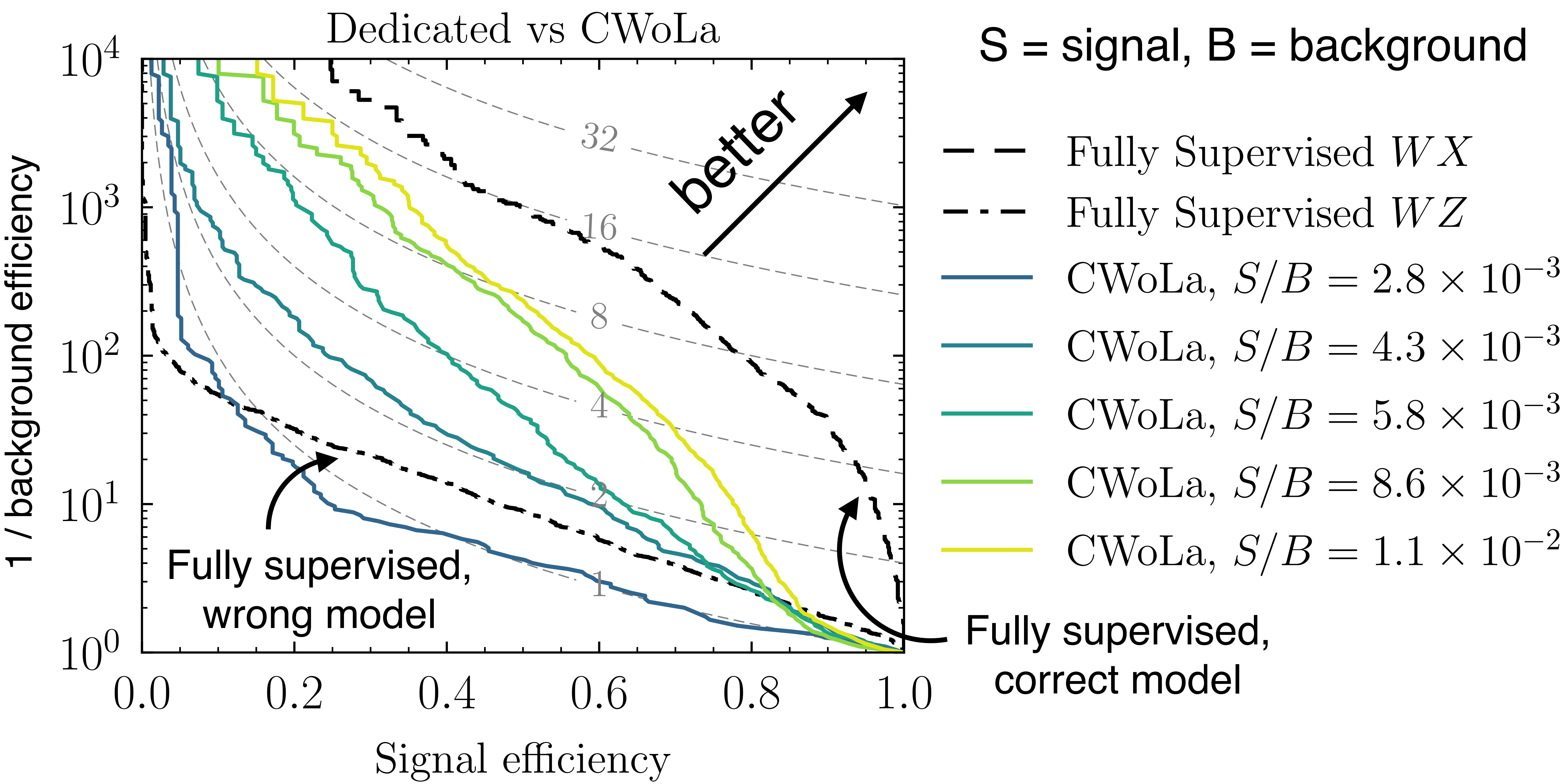
What is the network learning?





Learns to find the signal !

CWoLa hunting vs. Full Supervision

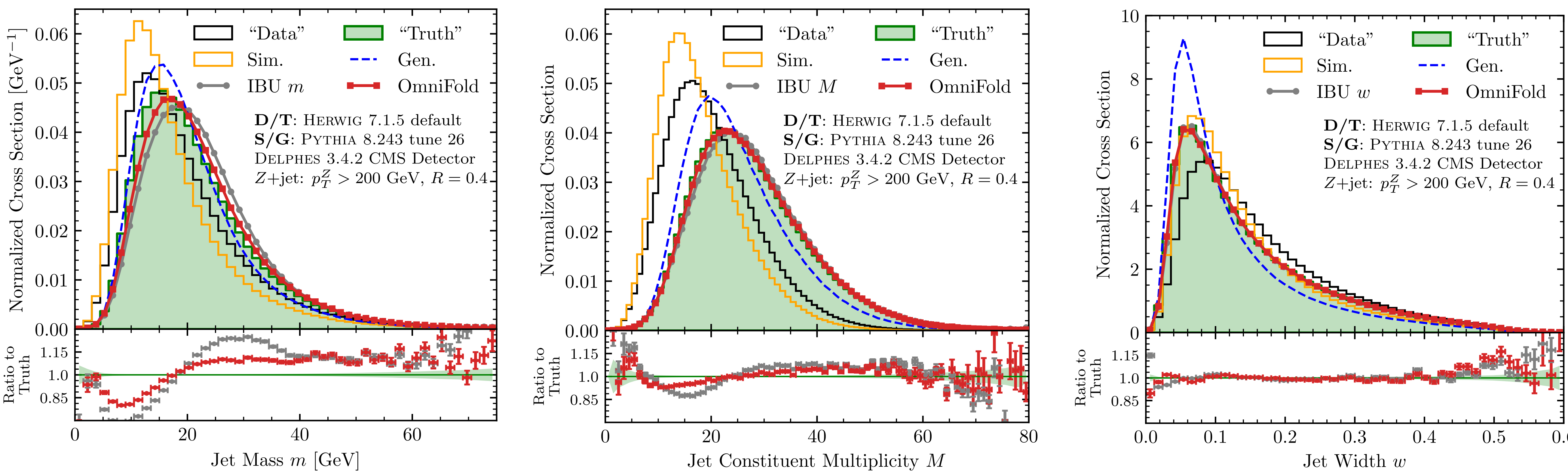


If you know what you are looking for, you should look for it. If you don't know, then CWoLa hunting may be able to catch it!

Example: Unfolding

What if we could unfold all particles simultaneously?
We could then compute observables (and their bins)
AFTER doing the measurement (!)

…stick around for the second part of this
session for more discussions on this point



Andreassen et al., 1911.09107

New search ideas

Supervision refers to the type of label information provided to the ML during training.

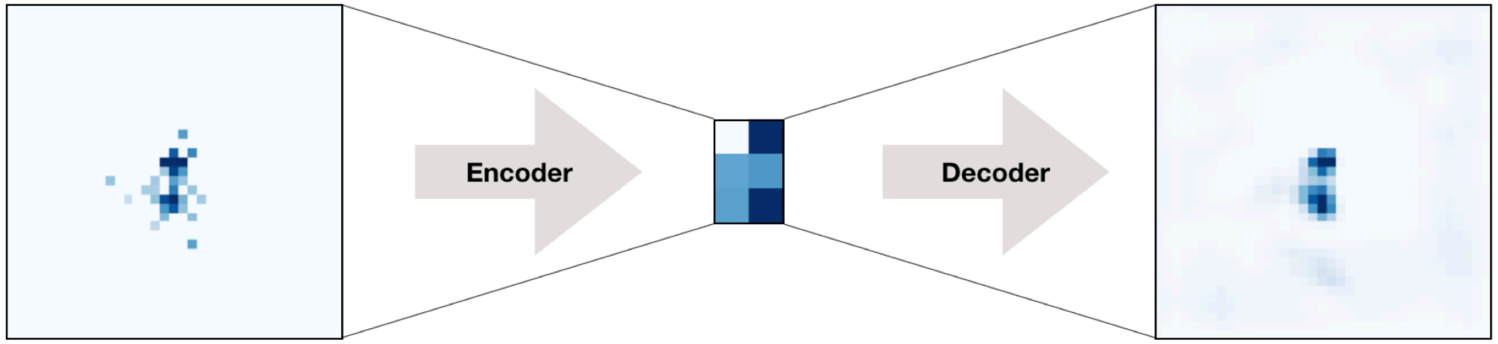
Unsupervised = no labels
Weakly-supervised = noisy labels
Semi-supervised = partial labels
Supervised = full label information

These categories are not exact and the boundaries are not rigid!

Solutions: Unsupervised

Unsupervised = no labels

Typically, the goal of these methods is to look for events with low *p(background)*



One strategy (autoencoders) is to try to compress events and then uncompress them. When x = uncompres(compress(x)), then x probably has low p(x).

Farina, Nakai, Shih, 1808.08992; Heimel, Kasieczka, Plehn, Thompson, 1808.08979; + many more including the recent LHC Olympics (Kasieczka et al., 2101.08320) and Dark Machines (Aarrestad et al., 2105.14027) reports

Solutions: Weakly-supervised

Weakly-supervised = noisy labels

Typically, the goal of these methods is to look for events with high *p(possibly signal-enriched)/p(possibly signal-depleted)*

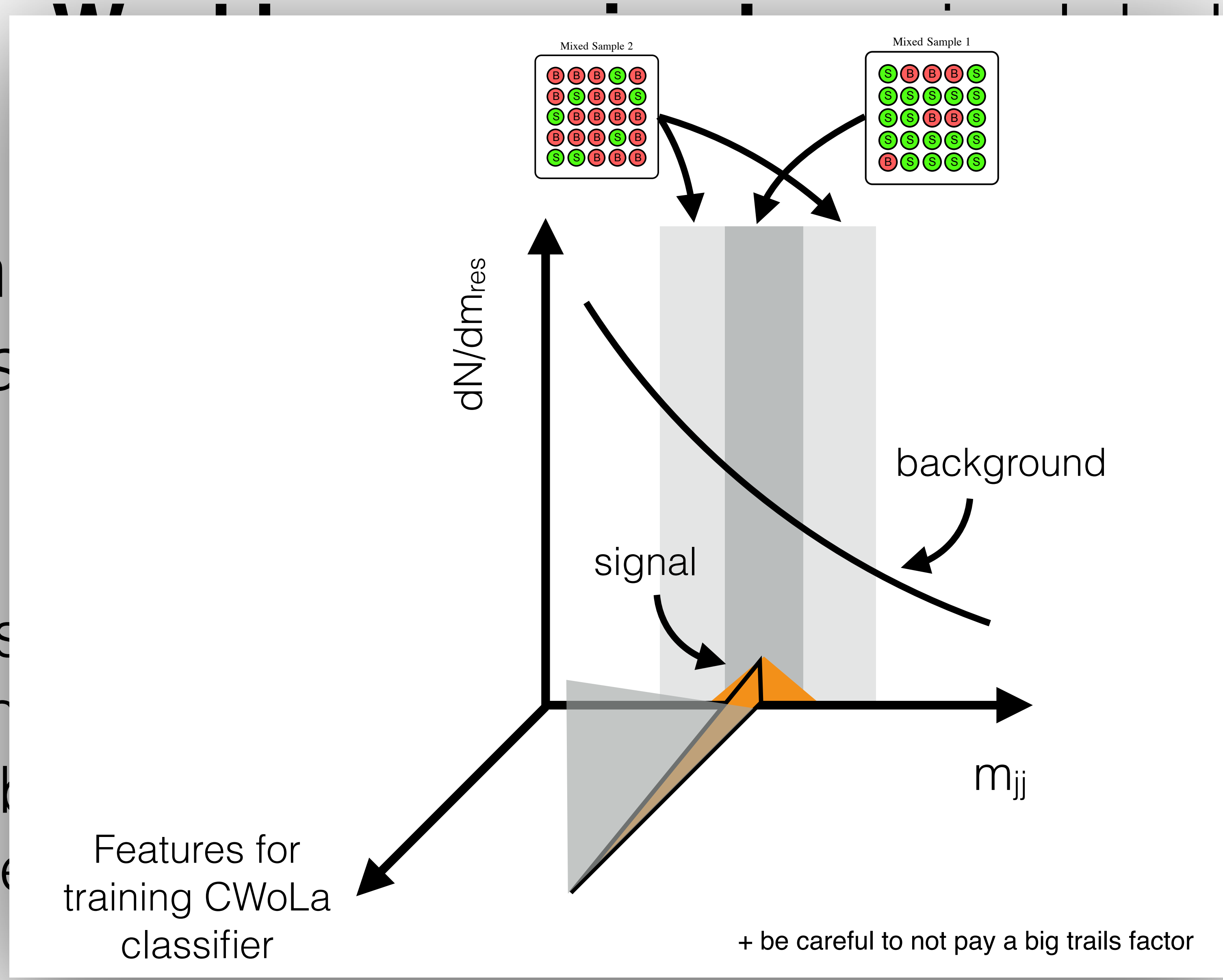
e.g. Classification Without Labels (CWoLa), events in a signal region are labeled "signal" and events in a sideband are labeled "background". These labels are "noisy" but a classifier trained with them can detect the presence of a signal.

Metodiev, Nachman, Thaler, 1708.02949; Collins, Howe, Nachman, 1805.02664
+ many more including the recent LHC Olympics (Kasieczka et al., 2101.08320)

Solutions: Weakly-supervised

Typically, th… …or events with
high *p(poss…* …*nal-depleted)*

e.g. Clas… …in a signal
region… …band are
labeled "… …a classifier
traine… …signal.

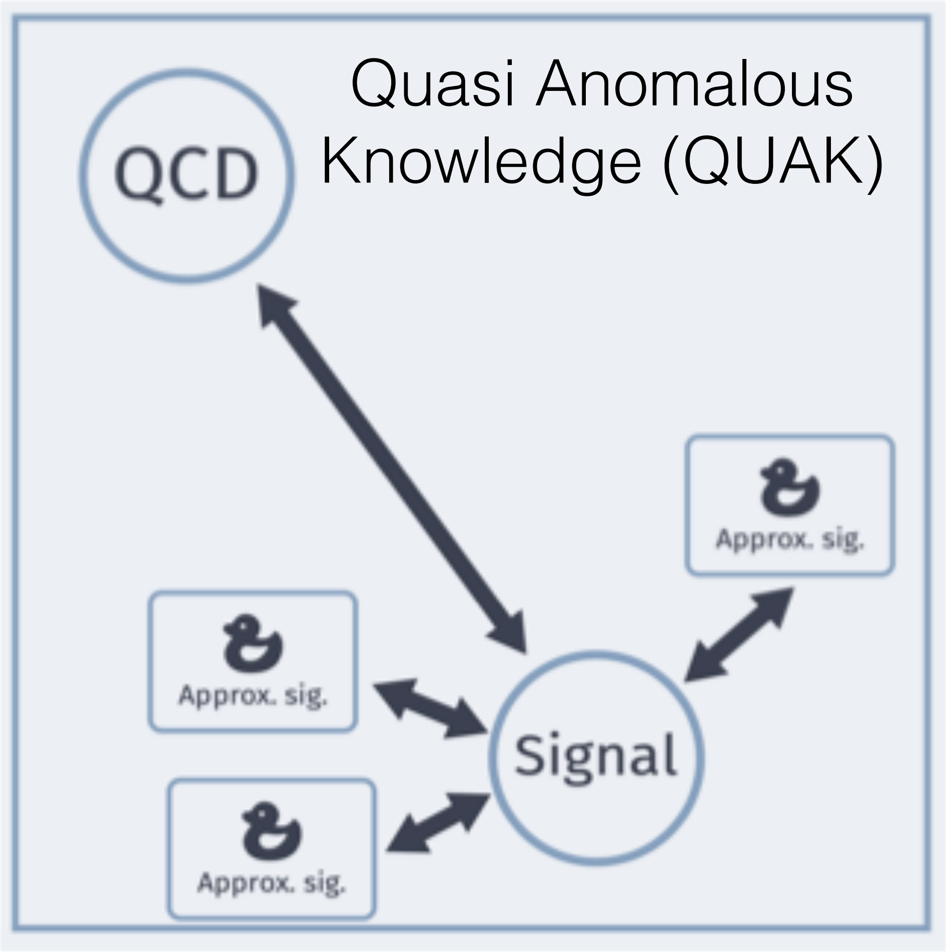


Metodiev, Nachman, Thaler, 1708.02949; Collins, Howe, Nachman, 1805.02664
+ many more including the recent LHC Olympics (Kasieczka et al., 2101.08320)

Solutions: Semi-supervised

Semi-supervised = partial labels

Typically, these methods use some signal simulations to build signal sensitivity



S. Park, D. Rankin, S.-M. Udrescu, M. Yunus, P. Harris, 2011.03550 + many more including the recent LHC Olympics (Kasieczka et al., 2101.08320)

Deep Generative Models

A **generator** is nothing other than a function that maps random numbers to structure.

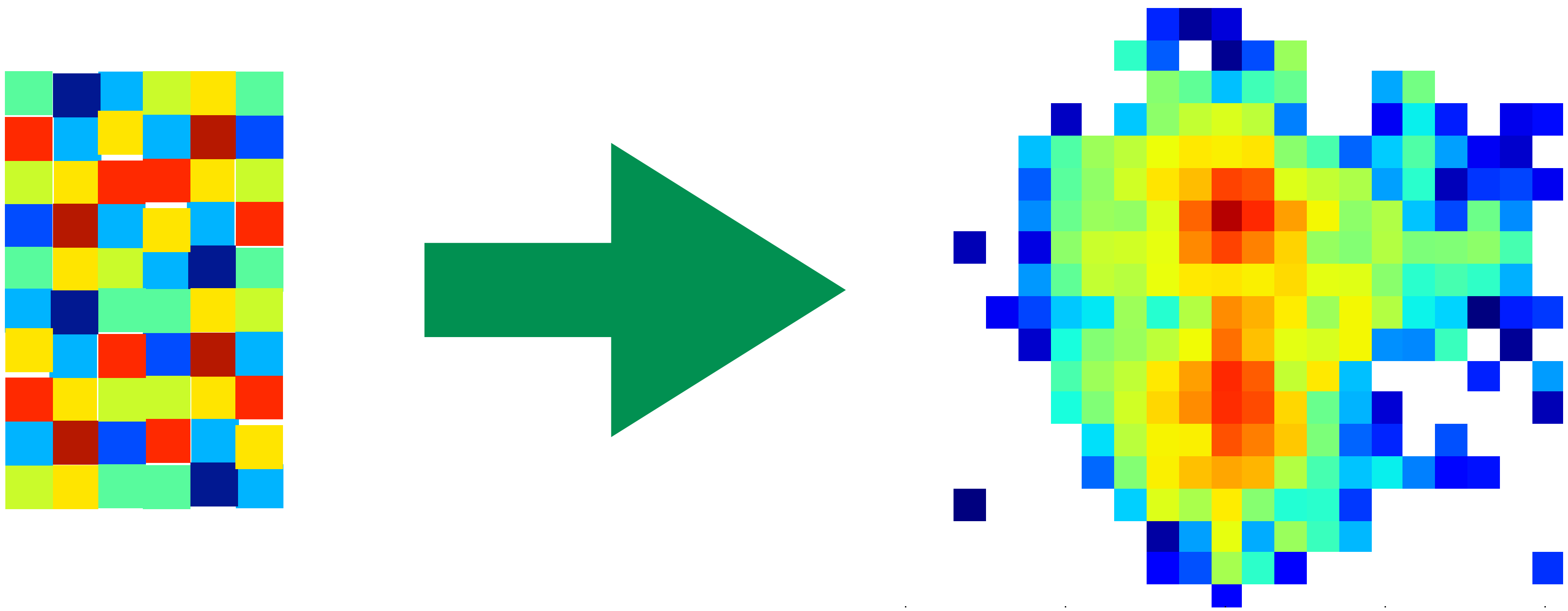


Image: Paganini, Oliveira, Nachman, 1705.02355

Deep Generative Models in HEP

Speeding up slow simulation

Generating Phase space

Estimating SM backgrounds

Measurements and Inference

BSM searches

N.B. being comprehensive with citations would fill up the slide - please see my link to the Living Review at the end for a comprehensive list

Deep Generative Models in HEP

Speeding up slow simulation

Generating Phase space

Estimating SM backgrounds

Measurements and Inference

BSM searches

N.B. being comprehensive with citations would fill up the slide - please see my link to the Living Review at the end for a comprehensive list

N.B. being comprehensive with citations would fill up the slide - please see my link to the Living Review at the end for a comprehensive list

Deep Generative Models in HEP

Speeding up slow simulation

Generating Phase space

Estimating SM backgrounds

Measurements and Inference

BSM searches

N.B. being comprehensive with citations would fill up the slide - please see my link to the Living Review at the end for a comprehensive list

Deep Generative Models in HEP

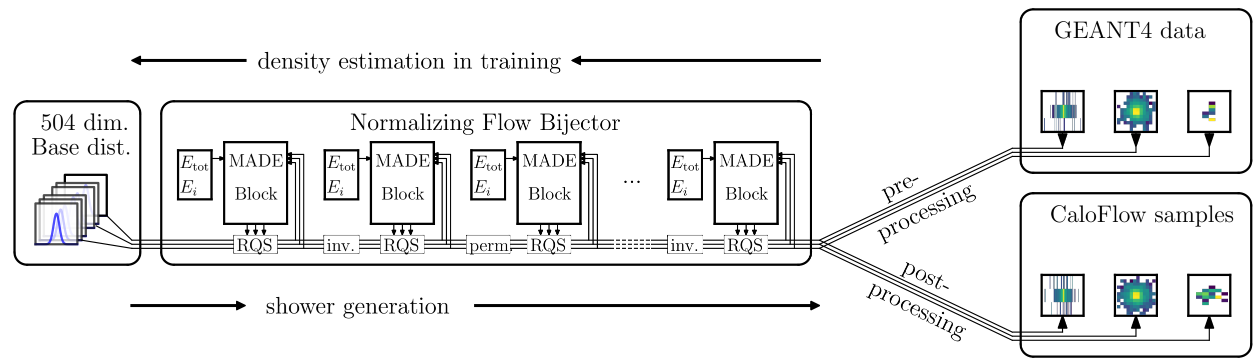
Speeding up slow simulation

Generating Phase space

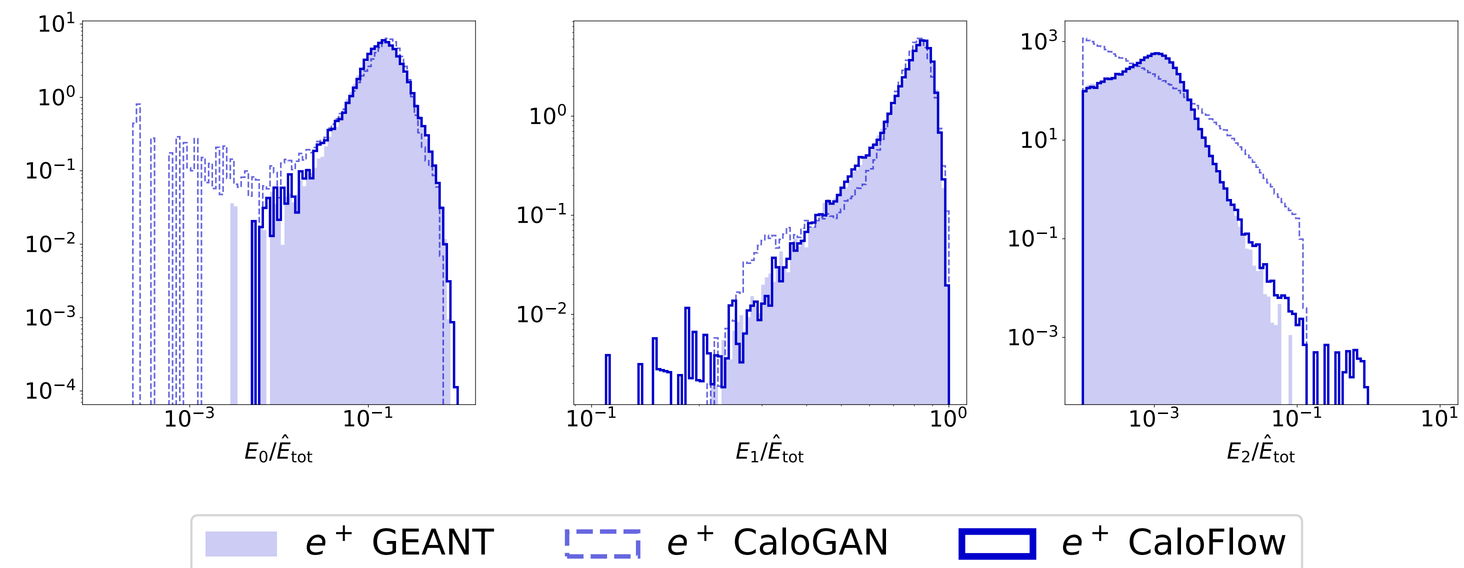
Estimating SM backgrounds

Measurements and Inference

BSM searches



GAN, Flow are NNs



CaloFlow: Krause and Shih, 2106.05285

CaloGAN: Paganini, Oliveira, Nachman, 1705.02355

not quite a fair comparison, but the state-of-the-art accuracy is highly non-trivial and very impressive!

N.B. being comprehensive with citations would fill up the slide - please see my link to the Living Review at the end for a comprehensive list

Deep Generative Models in HEP

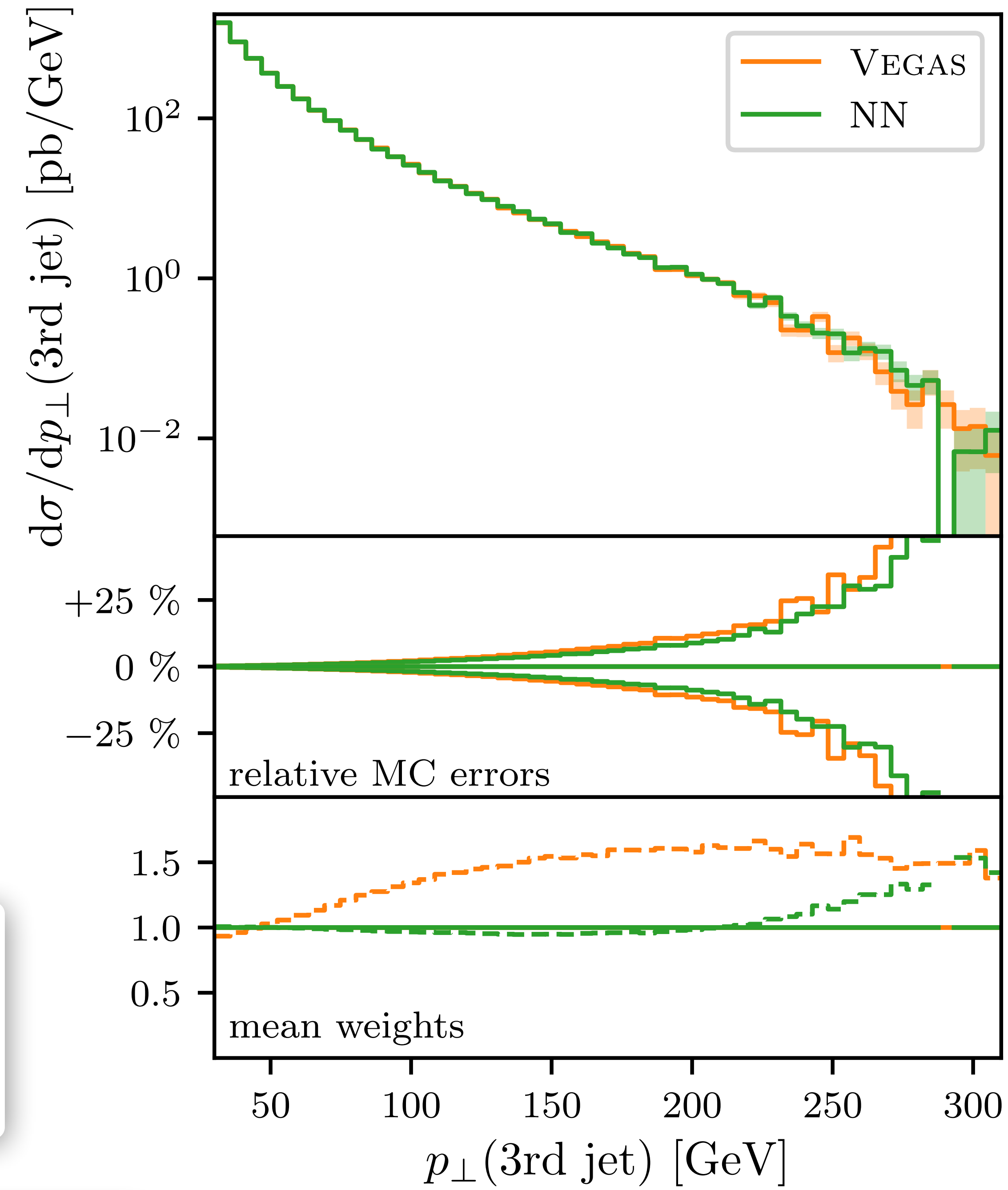
Speeding up slow simulation

Generating Phase space

Estimating SM backgrounds

Measurements and Inference

BSM searches



Bothmann et al., 2001.05478

N.B. being comprehensive with citations would fill up the slide - please see my link to the Living Review at the end for a comprehensive list

Deep Generative Models in HEP

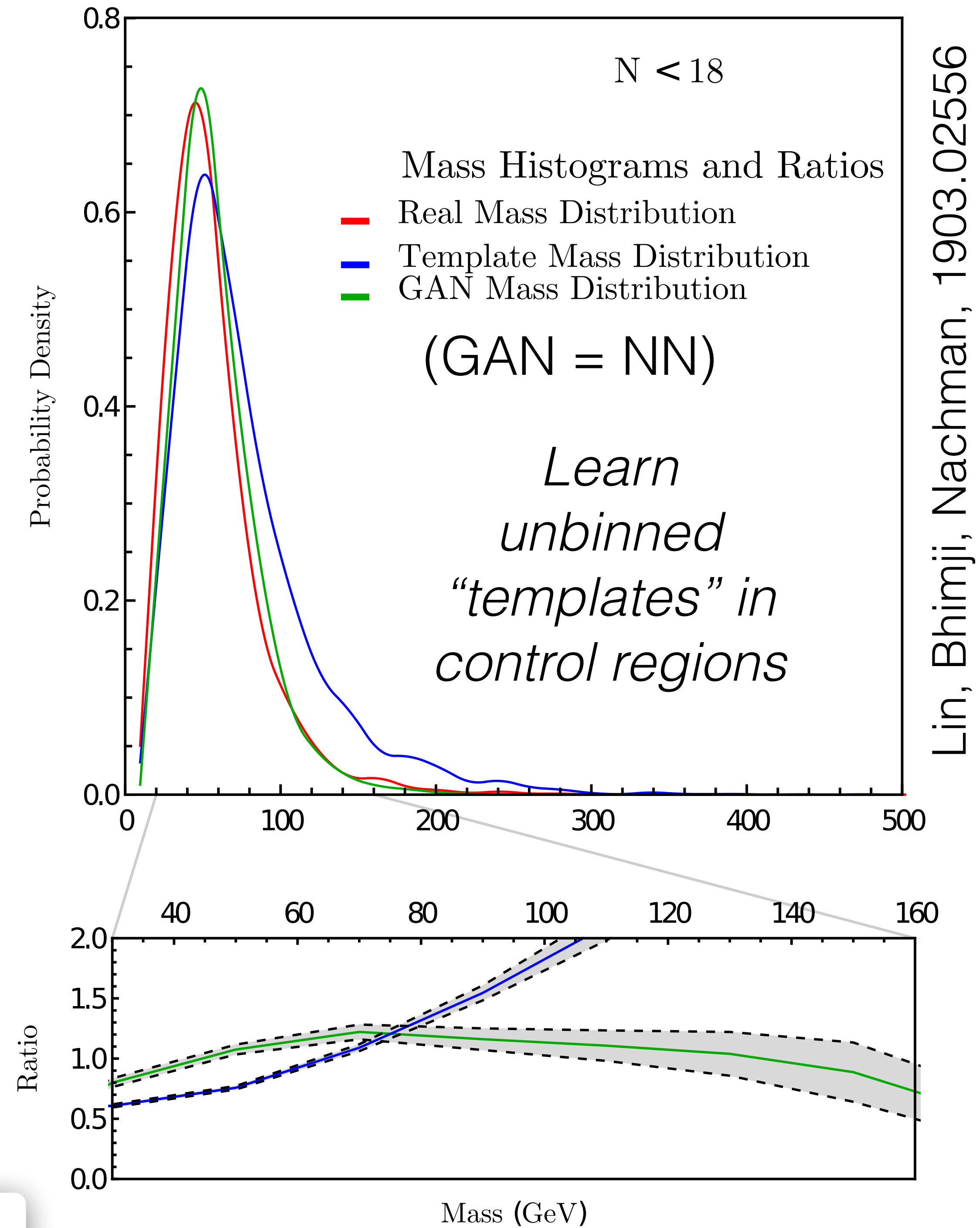
Speeding up slow simulation

Generating Phase space

Estimating SM backgrounds

Measurements and Inference

BSM searches



N.B. being comprehensive with citations would fill up the slide - please see my link to the Living Review at the end for a comprehensive list

Deep Generative Models in HEP

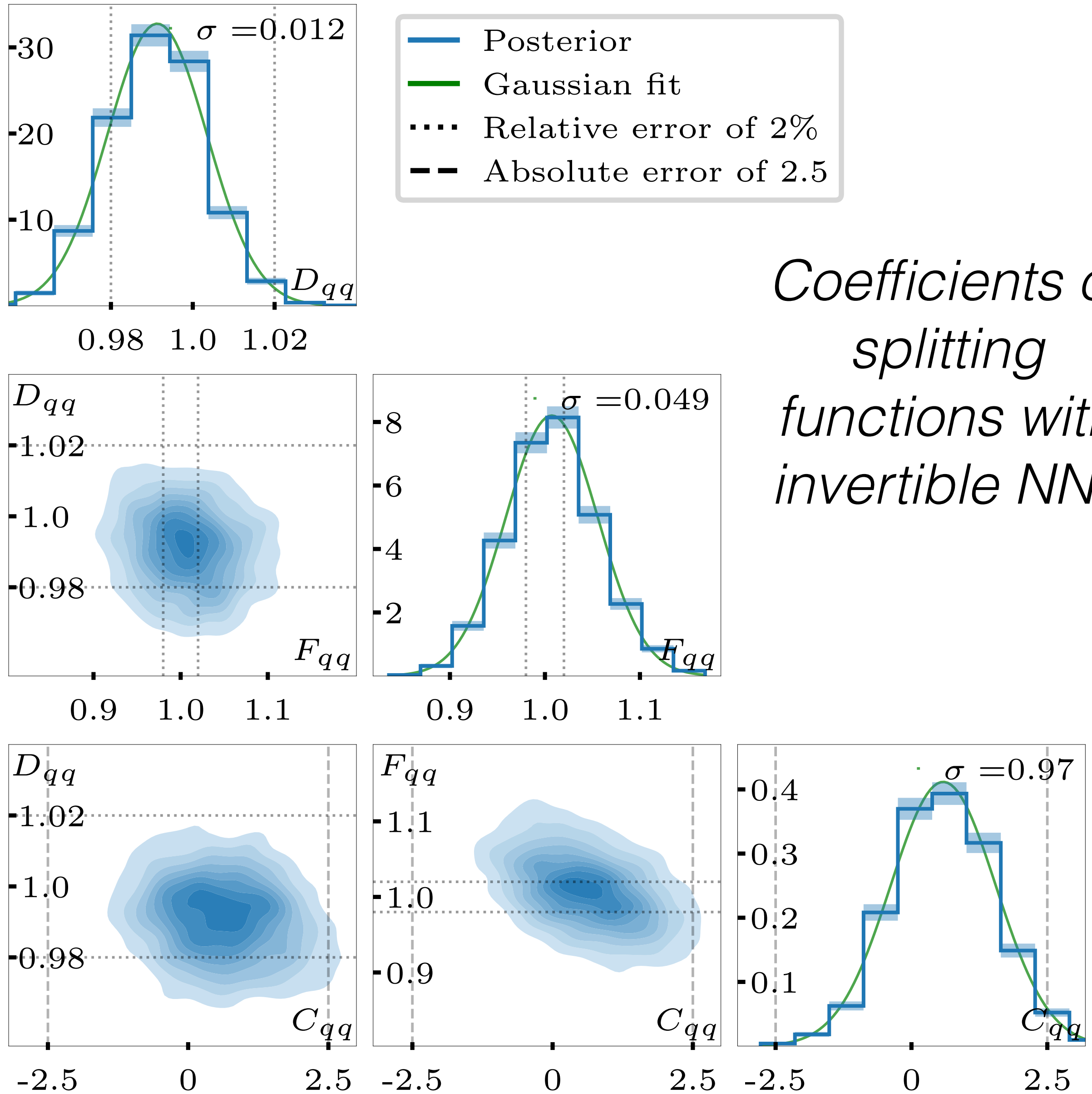
Speeding up slow simulation

Generating Phase space

Estimating SM backgrounds

Measurements and Inference

BSM searches



Coefficients of splitting functions with invertible NNs

Bieringer et al., 2012.09873

N.B. being comprehensive with citations would fill up the slide - please see my link to the Living Review at the end for a comprehensive list

Deep Generative Models in HEP

Speeding up slow simulation

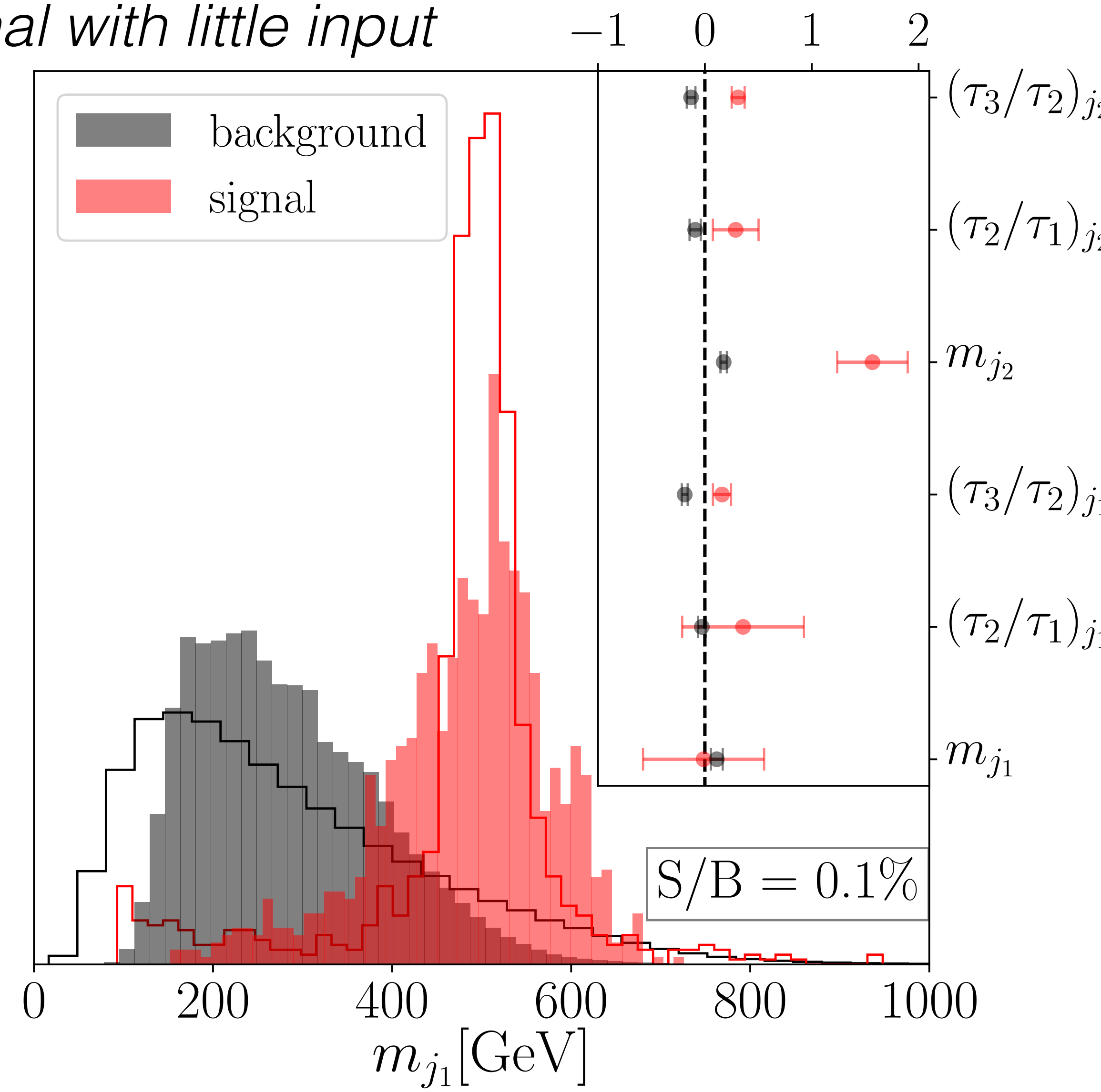
Generating Phase space

Estimating SM backgrounds

Measurements and Inference

BSM searches

Extract features of signal with little input



Bortolato et al., 2103.06595

See also the LHC Olympics 2020

N.B. being comprehensive with citations would fill up the slide - please see my link to the Living Review at the end for a comprehensive list