A new era for fundamental physics with deep learning and quantum computing

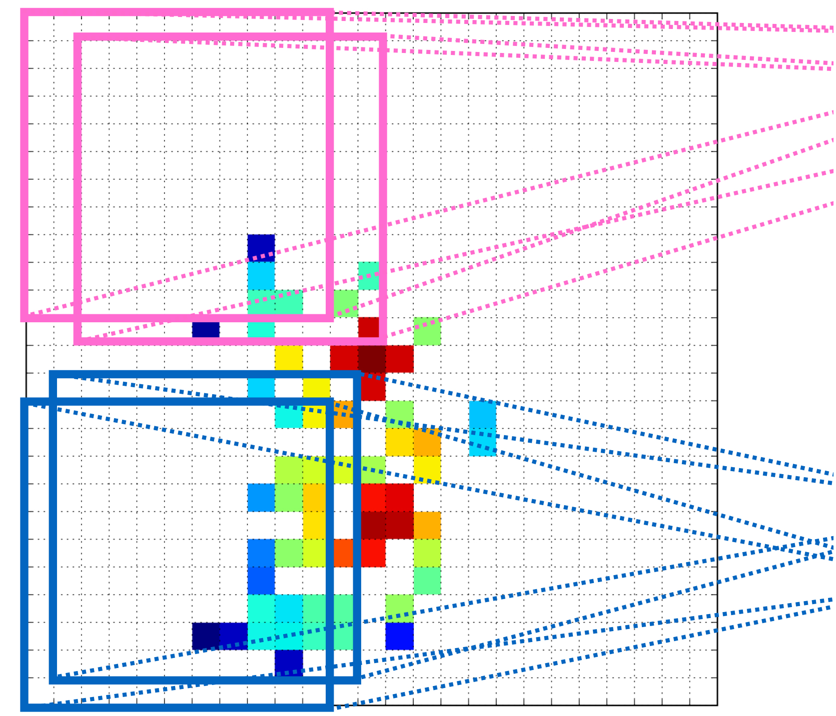
Benjamin Nachman

Lawrence Berkeley National Laboratory

cern.ch/bnachman

bpnachman@lbl.gov

@bpnachman bnachman



SLAC National Lab.

Panofsky Seminar

March 2, 2020

BERKELEY EXPERIMENTAL PARTICLE PHYSICS

NESAP for Learning

QuantISED HEP initiative

Outline for today

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

Outline for today

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

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

Outline for today

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

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

Deep learning & **Quantum computing** for fundamental **physics**

Outline for today

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

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

Likelihood-Free inference

Deep learning & **Quantum computing** for fundamental **physics**

Outline for today

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

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

Likelihood-Free inference

Deep learning & **Quantum computing** for fundamental **physics**

[Deconvolution/Unfolding]

[Generative models]

Outline for today

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

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

Likelihood-Free inference

Deep learning & **Quantum computing** for fundamental **physics**

Label-Free learning

[Deconvolution/Unfolding]

[Generative models]

Outline for today

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

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

Likelihood-Free inference

Deep learning & **Quantum computing** for fundamental **physics**

Label-Free learning

[Deconvolution/Unfolding]

[Generative models]

[Weak supervision]

[Anomaly detection]

Outline for today

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

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

Likelihood-Free inference

Deep learning & **Quantum computing** for fundamental **physics**

Label-Free learning

[Deconvolution/Unfolding]

[Generative models]

[Weak supervision]

[Anomaly detection]

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?

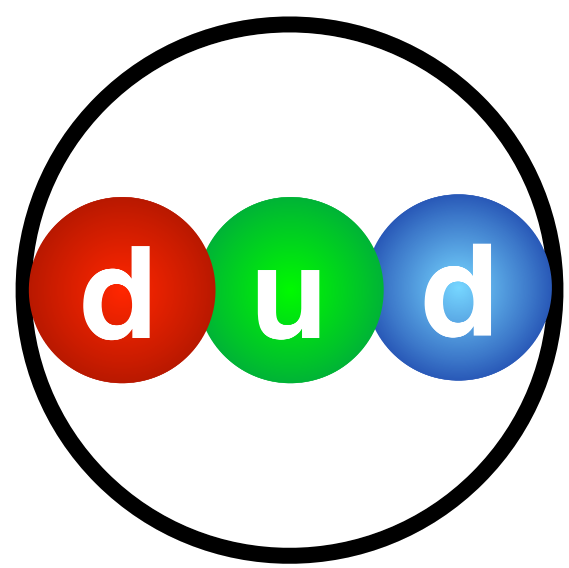
Hierarchy problem

image source: symmetry magazine

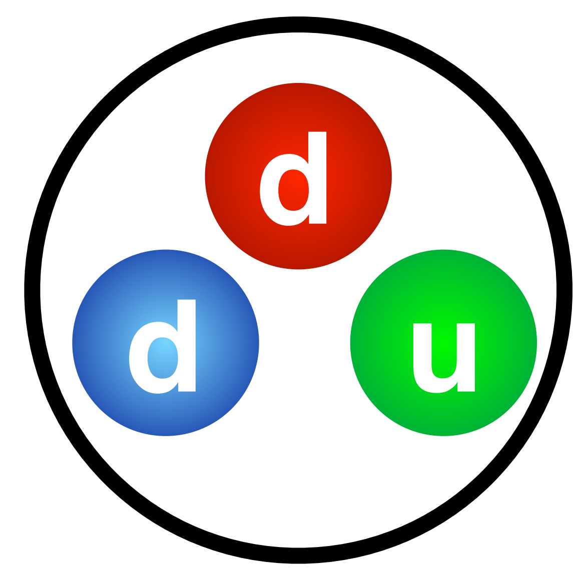
See also: quantum gravity

Why do neutrons have no dipole moment?

Strong CP



Reality



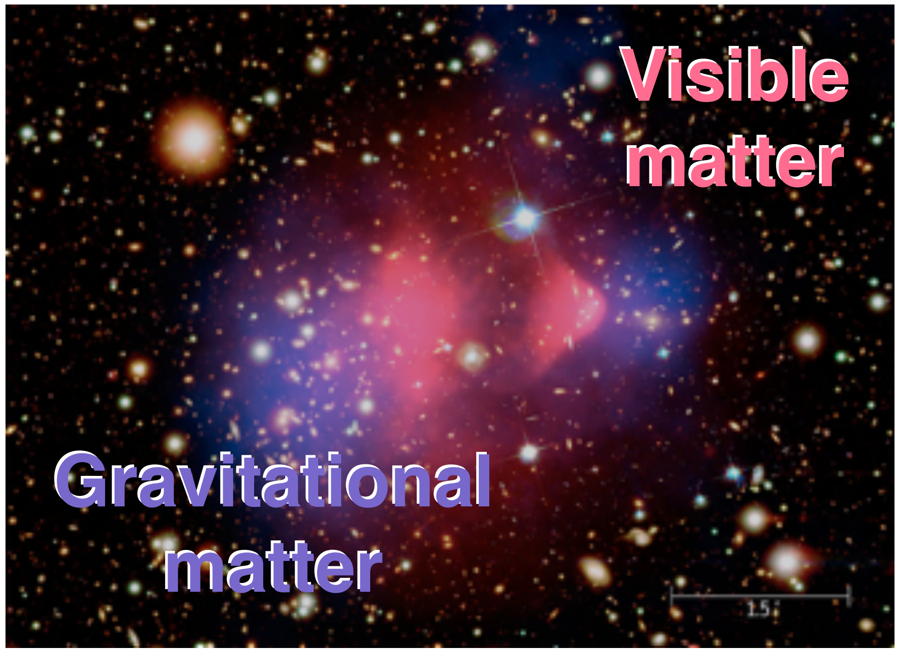
>99% of pictures on the internet

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

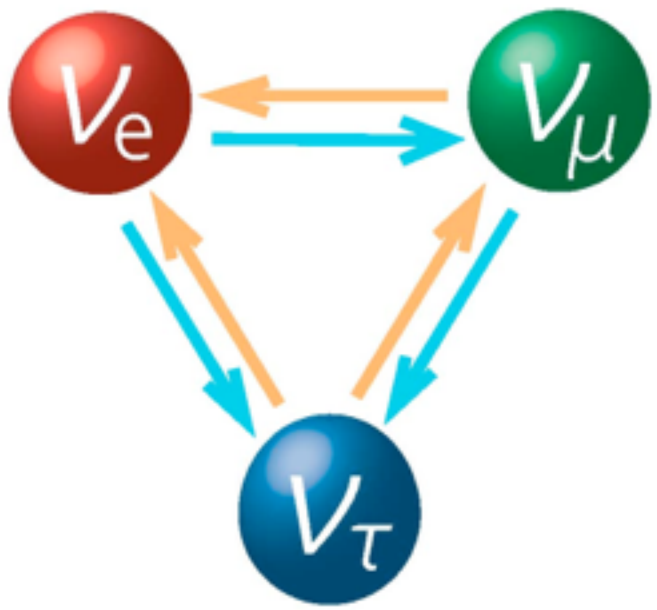


image source: T2K collaboration

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

Hierarchy problem

Strong CP

Flavor puzzles

Dark energy

Baryogengesis

We have performed thousands of hypothesis tests and have never rejected the null (the Standard Model(s))

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 and have never rejected the null (the Standard Model(s))

Three possibilities

(1) There is nothing new

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 and have never rejected the null (the Standard Model(s))

Three possibilities

(1) There is nothing new

(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

Hierarchy problem

Strong CP

Flavor puzzles

Dark energy

Baryogengesis

We have performed thousands of hypothesis tests and have never rejected the null (the Standard Model(s))

Three possibilities

(1) There is nothing new

(2) Patience! (new physics is rare)

(3) Our search methods are not sensitive

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 and have never rejected the null (the Standard Model(s))

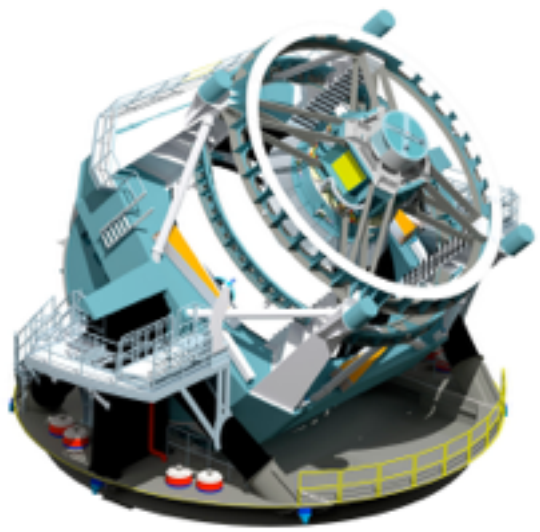
Three possibilities

This is what keeps me up at night !

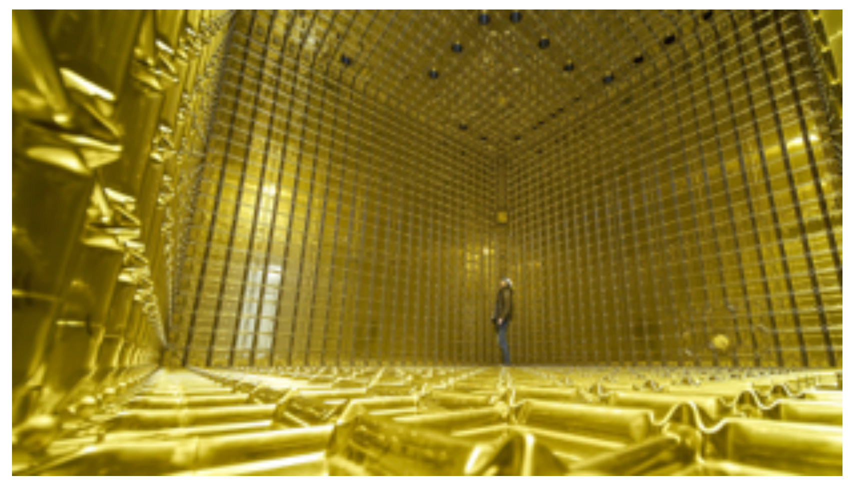
(3) Our search methods are not sensitive

Addressing the questions

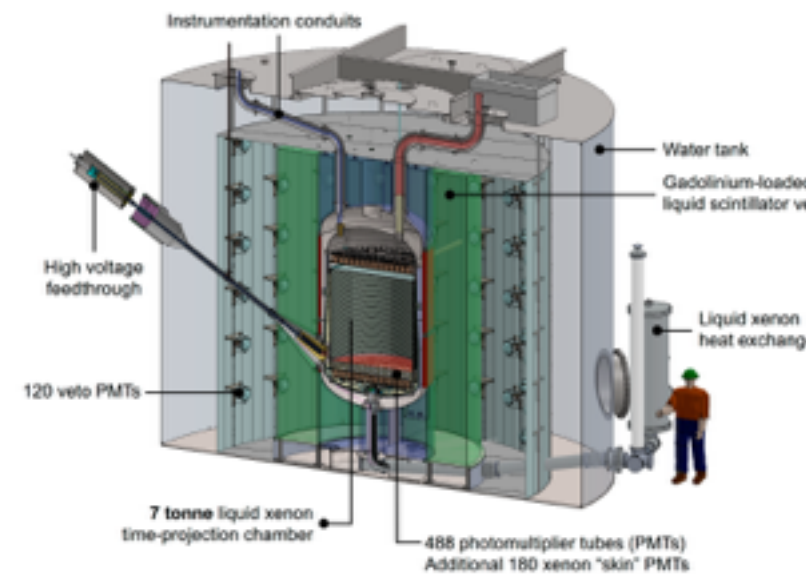
Dark matter/energy with LSST



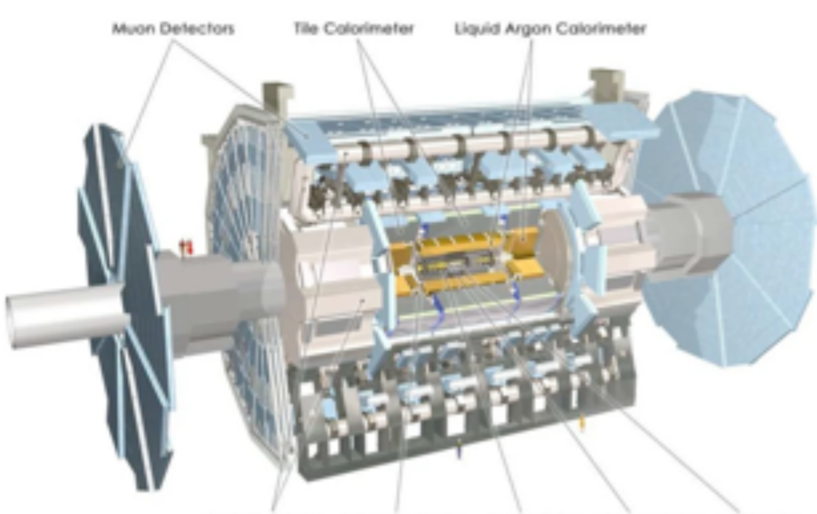
Fermilab neutrino experiments



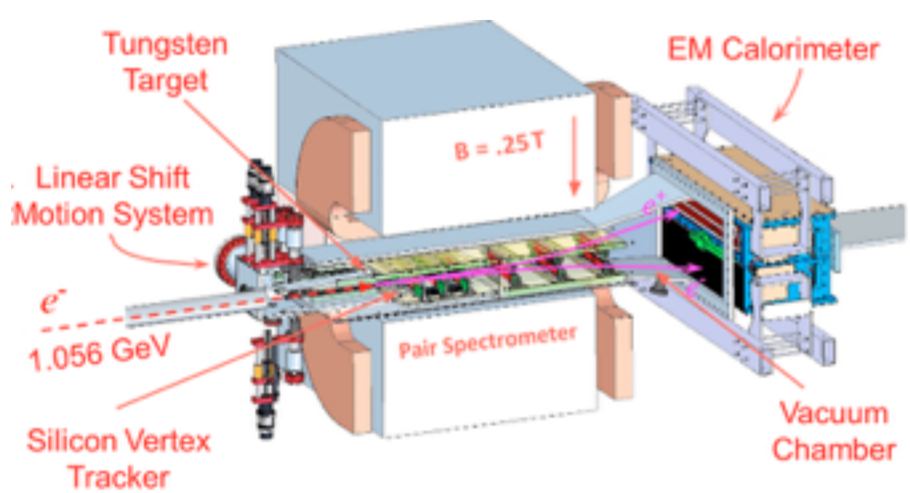
Dark Matter with LZ



Large Hadron Collider



Heavy Photon Search

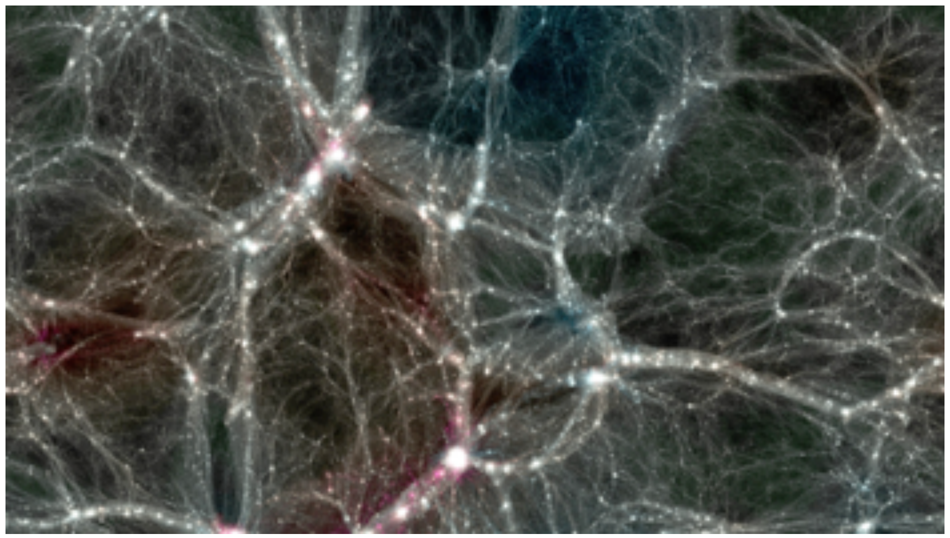


+ others !

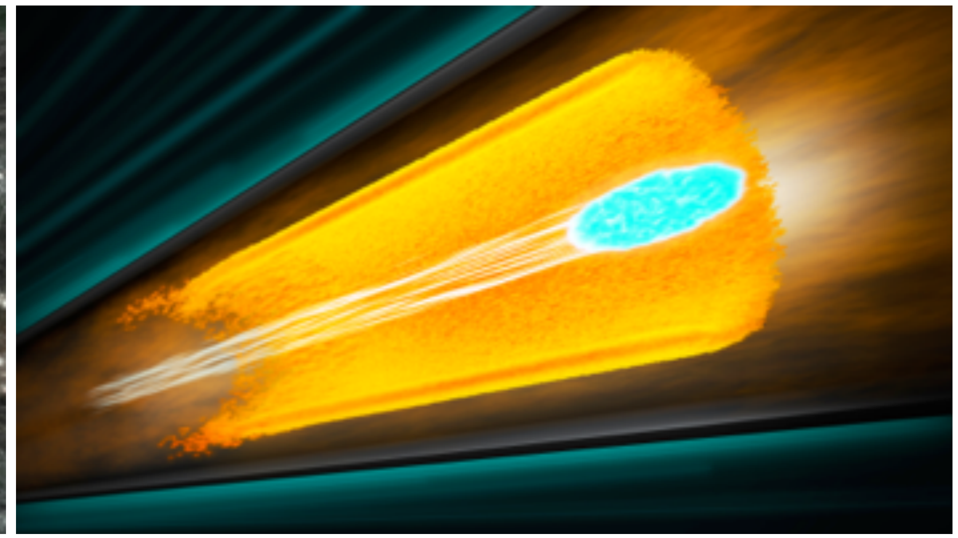
Image sources: LSST Corporation, CERN courier (CERN-201710-248-3), Berkeley Lab, CERN, HPS Collaboration

Addressing the questions

N-body simulations

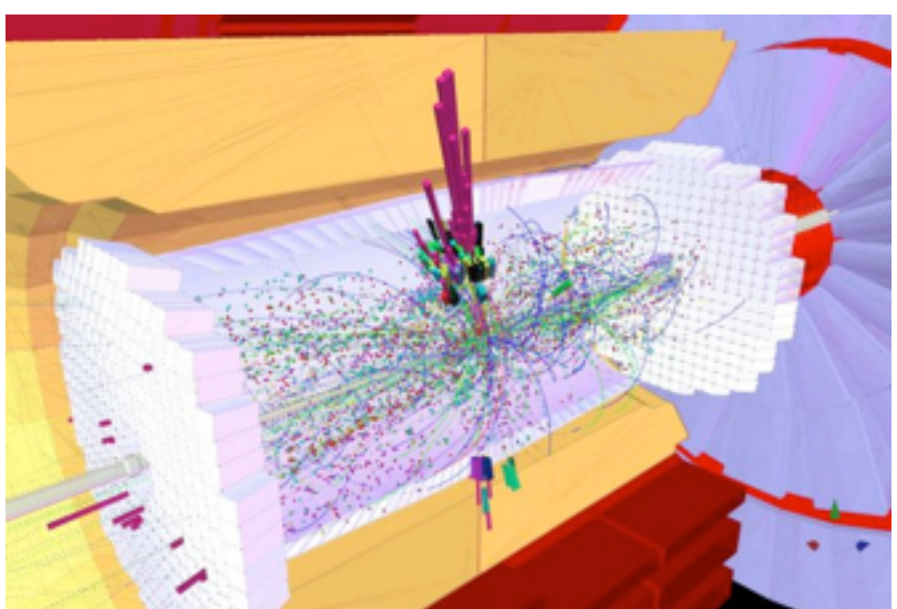


Advanced accelerators

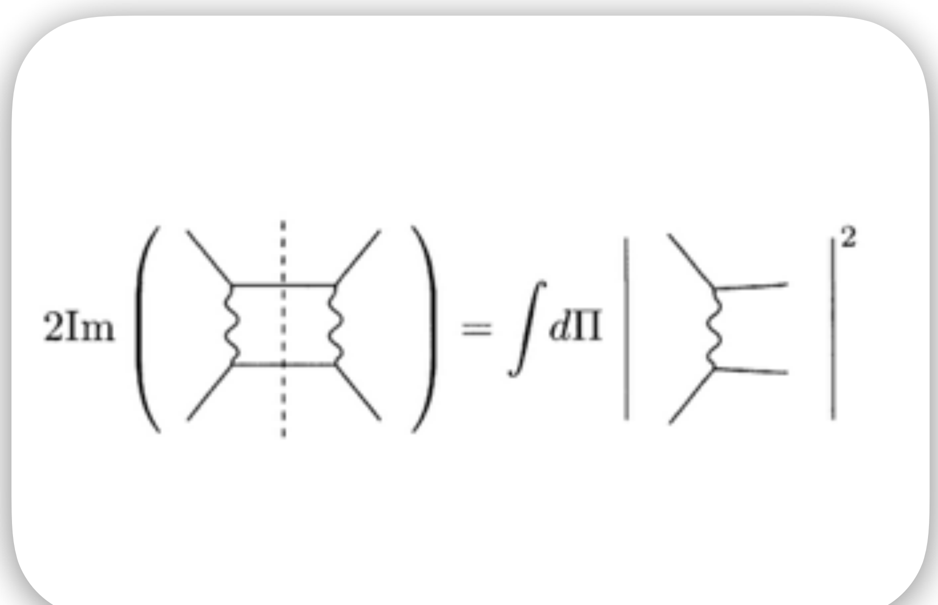


Supercomputers

Material interactions with Geant4



Theory Calculations

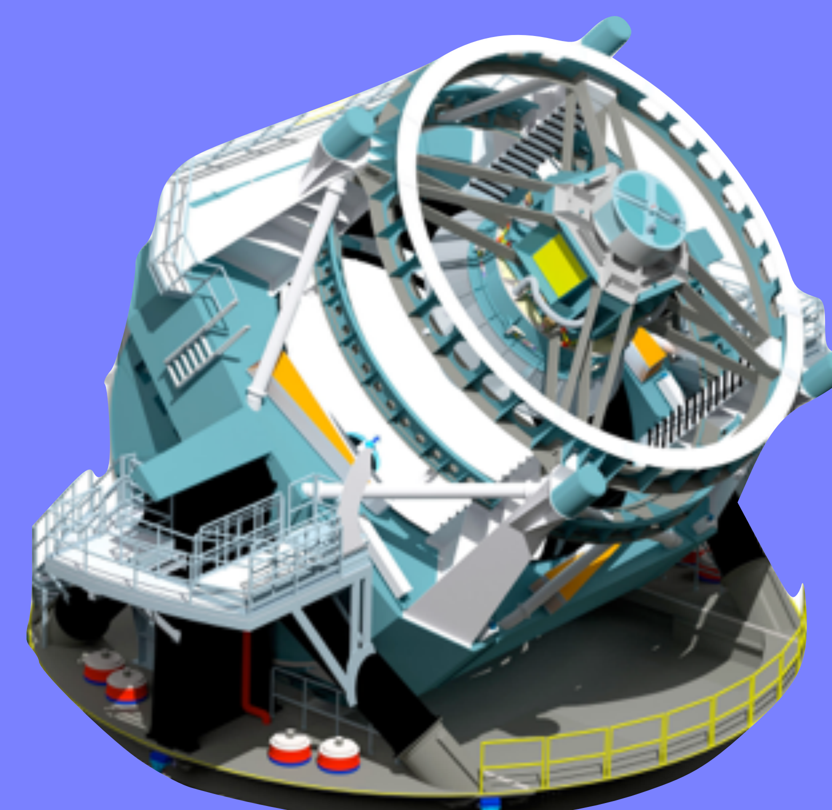


+ others !

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*



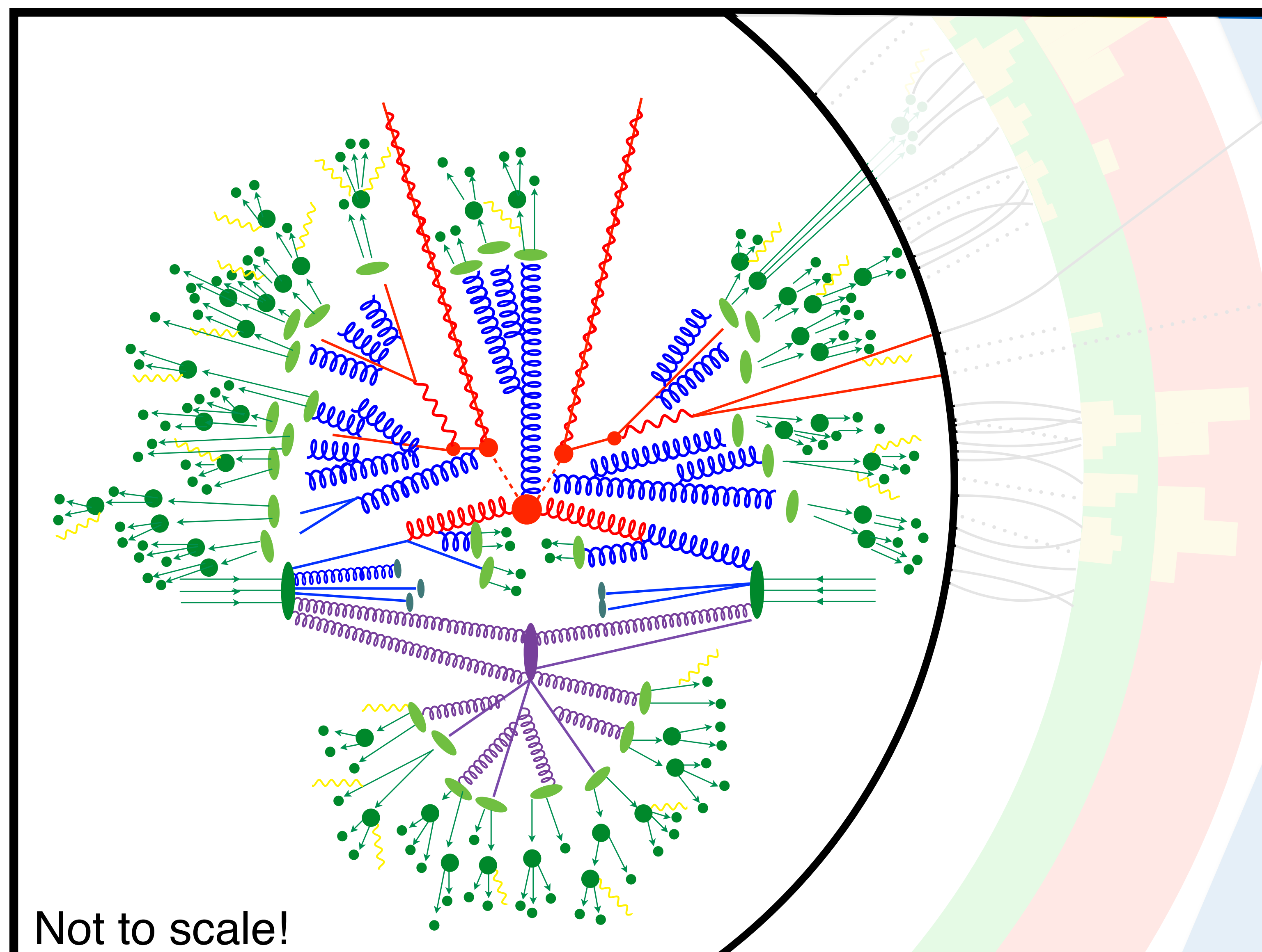
A *hyper* challenge

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



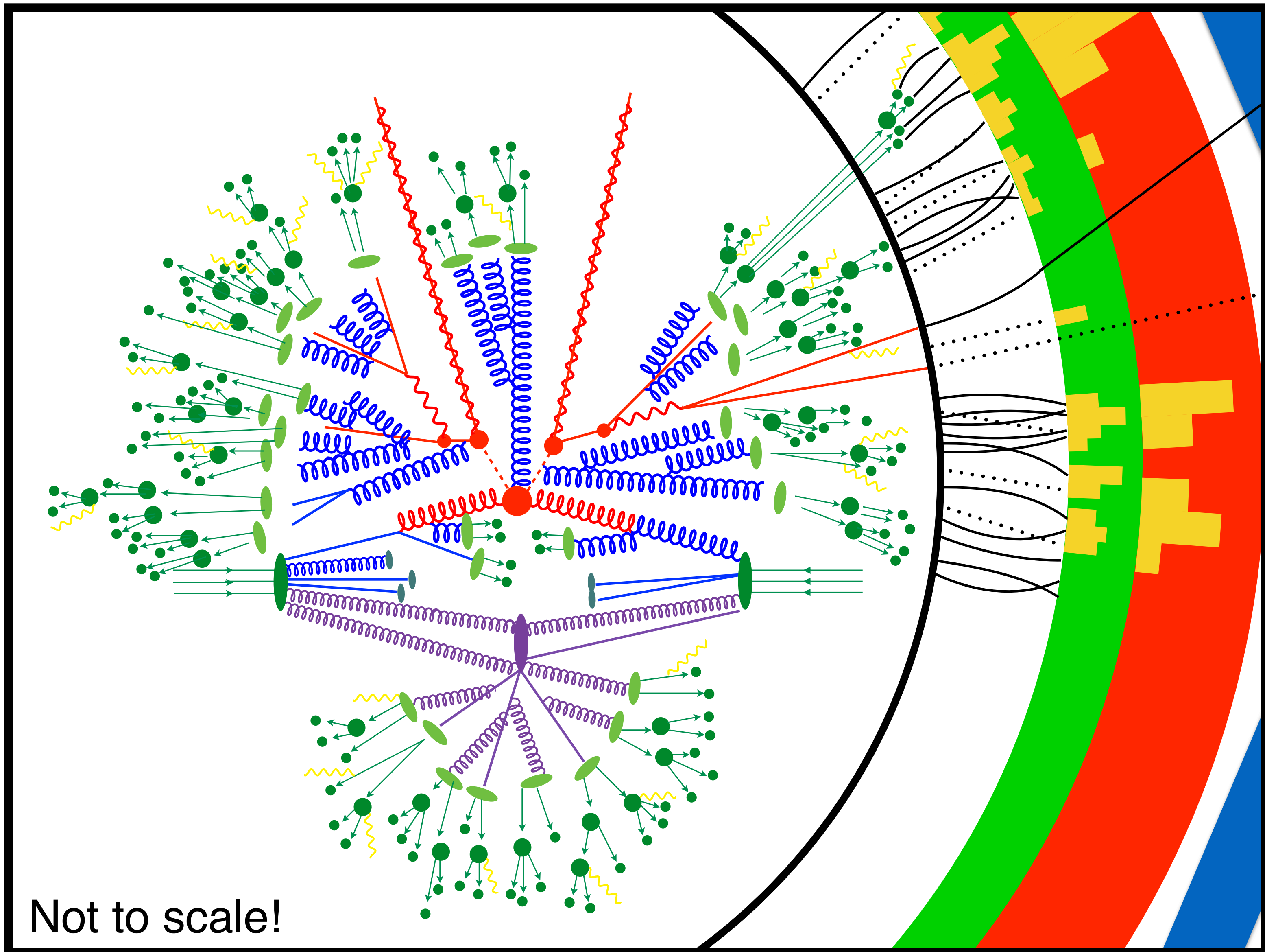
A *hyper* challenge

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

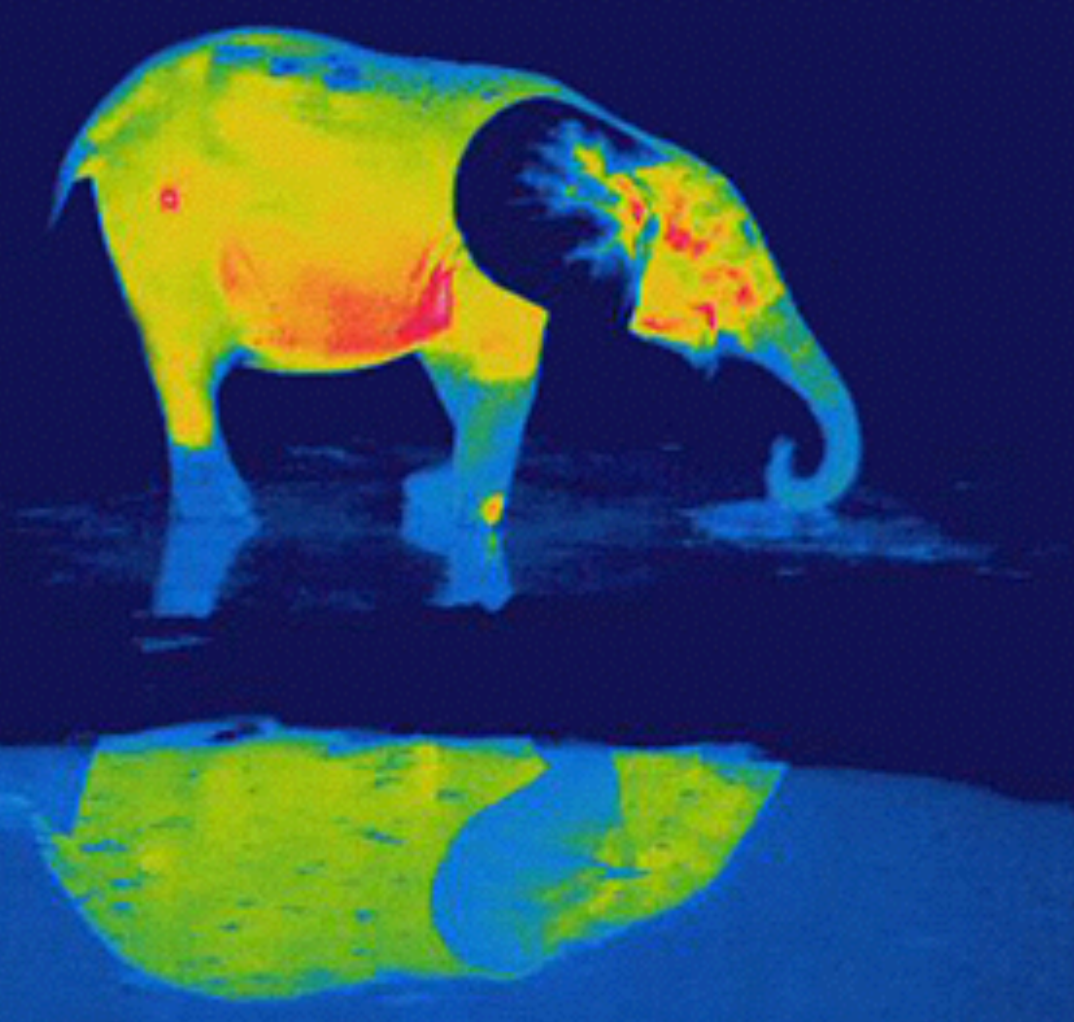


Hypervariate vision with deep learning

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…



©Elephant Listening Project

Data analysis in fundamental physics

Theory of everything

↓

Physics simulators

↓

Detector-level observables

↓

Pattern recognition ⟷ Pattern recognition

Nature

↓

Experiment

↓

Detector-level observables

↓

Pattern recognition

Data analysis in fundamental 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 fundamental 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.

Most of that machine learning is “straightforward”.

Data analysis in fundamental physics

Analysis Observables

[JHEP 10 (2018) 101,
PRD 100 (2019) 095016,
JHEP 07 (2016) 069]

Calibration

[JINST 15 (2020) P01030,
ATL-PHYS-PUB-2020-001,
ATL-PHYS-PUB-2018-013]

Pattern recognition

[NIM A913 (2018) 91,
NIM A951 (2020) 162879]

Noise mitigation

[JHEP 12 (2017) 51,
ATL-PHYS-PUB-2019-028]

Readout

[NIM A902 (2018) 197,
JINST 14 (2019) P09028]

Nature

↓

Experiment

↓

Detector-level observables

↓

Pattern recognition

I won't discuss this area at all, but I am happy to talk later about the potential gains from deep learning on many fronts.

Data analysis in fundamental physics

Theory of everything

↓

Physics simulators

↓

Detector-level observables

↓

Pattern recognition ⟷

Nature

↓

Experiment

↓

Detector-level observables

↓

Pattern recognition

Data analysis in fundamental physics

Theory of everything

↓

Physics simulators

↓

Detector-level observables

↓

Pattern recognition

↔

Nature

↓

Experiment

↓

Detector-level observables

↓

Pattern recognition

Goal of likelihood-free inference is to extract physical information about nature using only the "forward model".

Data analysis in fundamental 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.

Data analysis in fundamental physics

Theory of everything

↓

Physics simulators

↓

Detector-level observables

↓

Pattern recognition

Nature

↓

Experiment

↓

Detector-level observables

↓

Pattern recognition

Anomaly detection

Data analysis in fundamental physics

Theory of everything

↓

Physics simulators

↓

Detector-level observables

↓

Pattern recognition

↔

Nature

↓

Experiment

↓

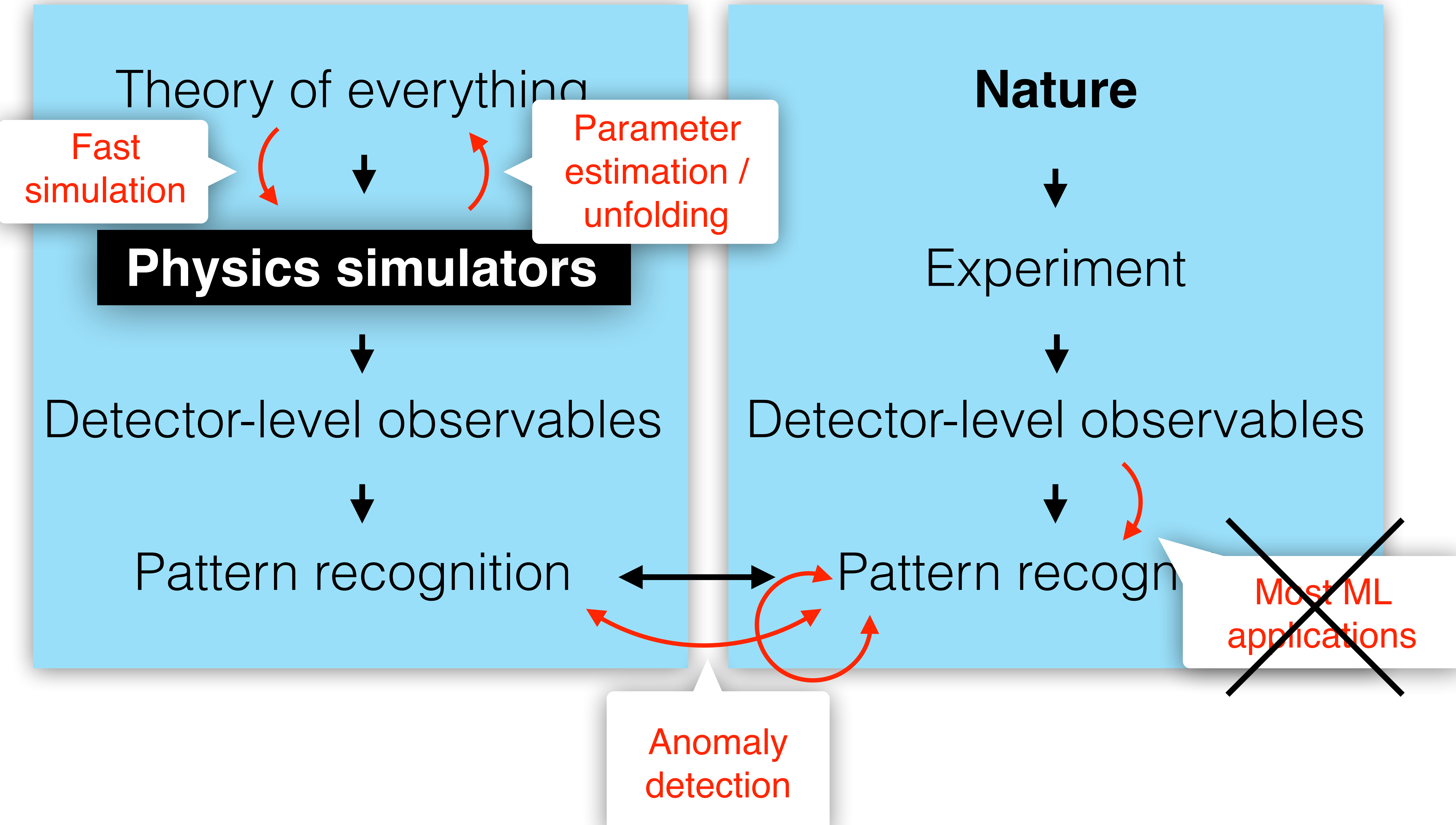
Detector-level observables

↓

Pattern recognition

Anomaly detection

Roadmap for rest of this talk



I will use examples from **collider physics**, but the techniques are **much broader**.

Colliders offer a complex environment for developing methods, but I am actively investigating applications **across fundamental physics domains**.

Outline for today

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

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

Likelihood-Free inference

Deep learning & **Quantum computing** for fundamental **physics**

Label-Free learning

[Deconvolution/Unfolding]

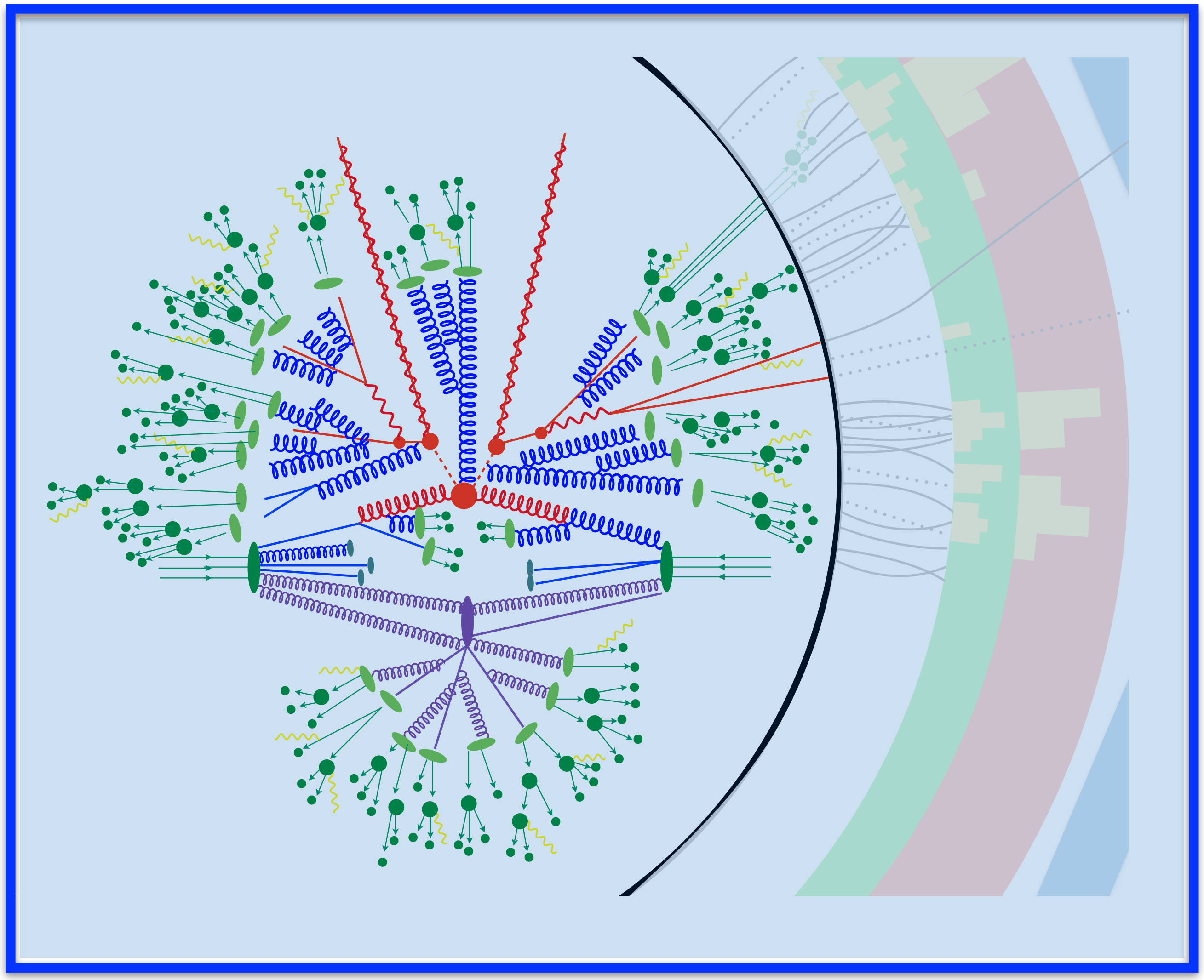
[Generative models]

[Weak supervision]

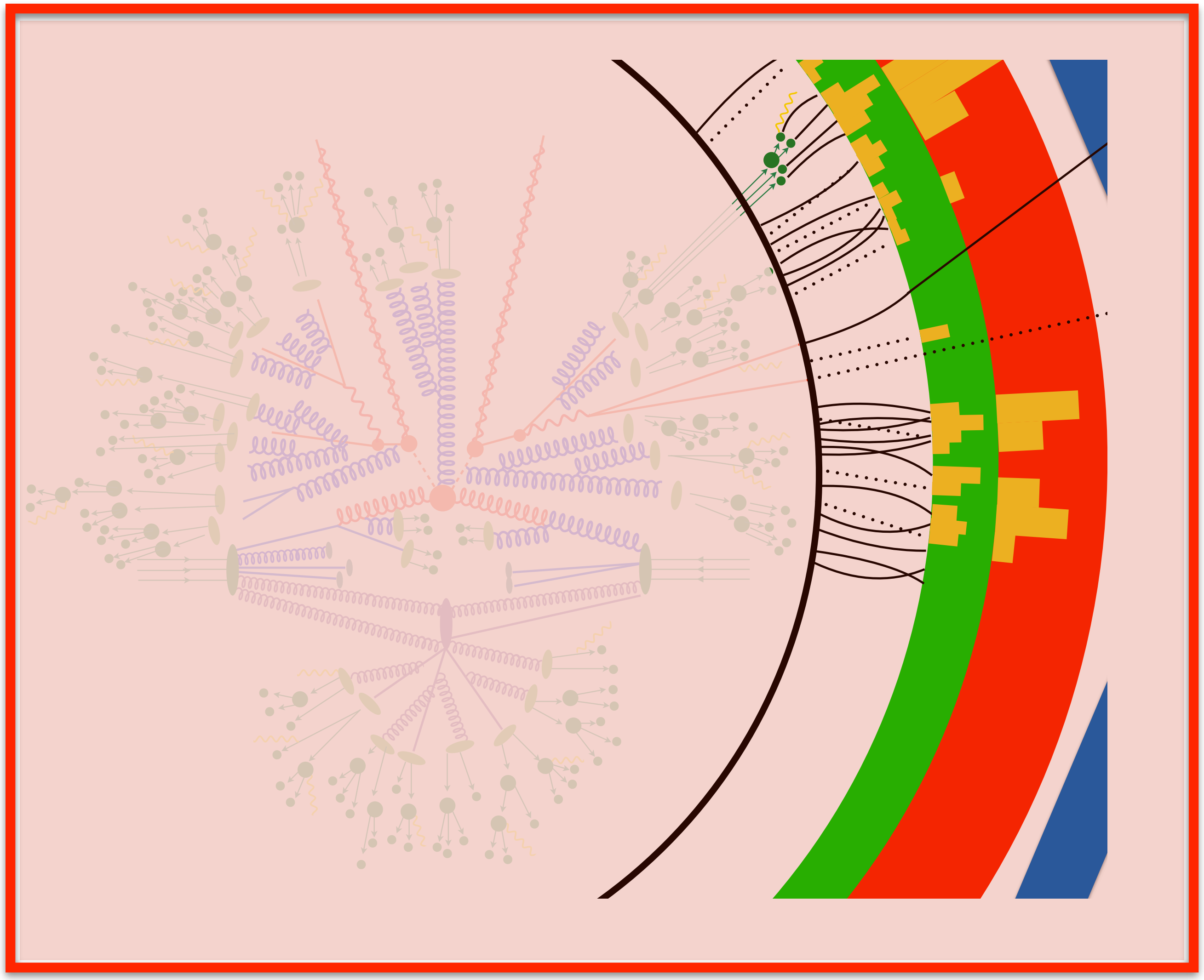
[Anomaly detection]

Deconvolution/Unfolding

Want this



Measure this

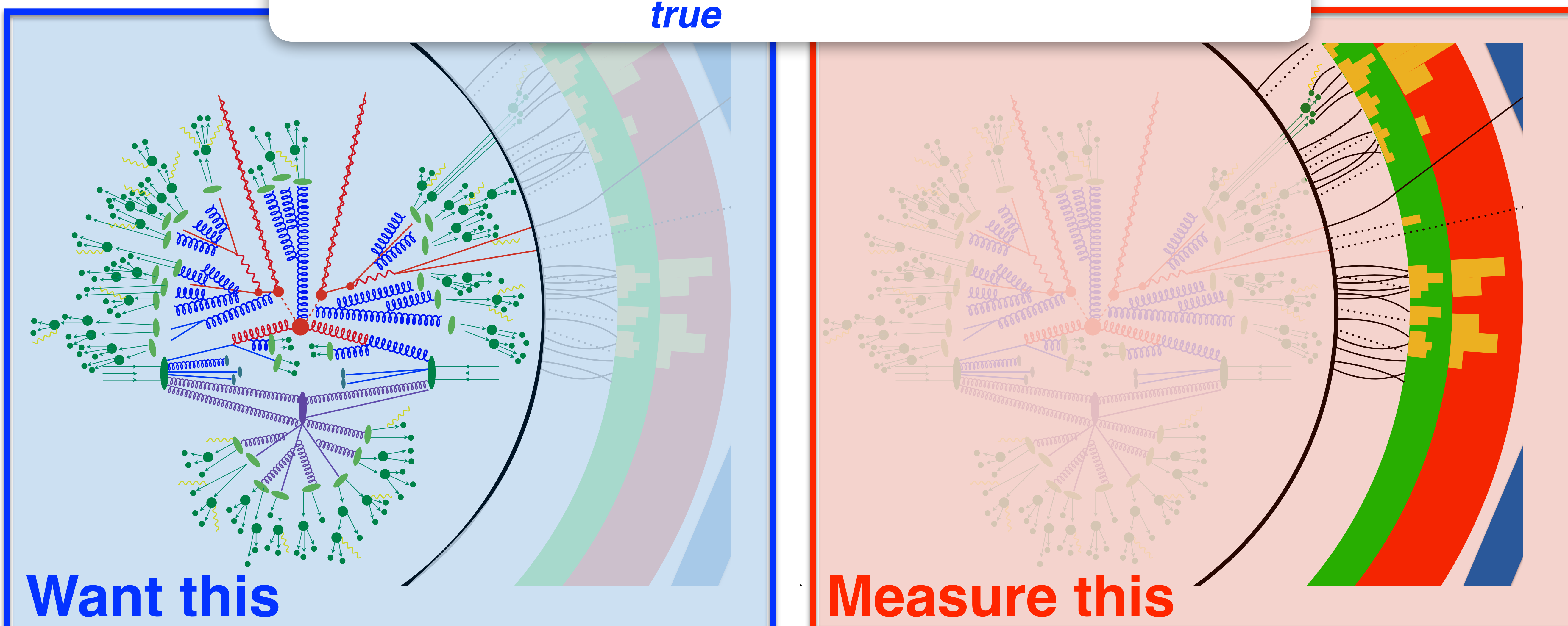


i.e. remove detector distortions

Deconvolution/Unfolding

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

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



p(meas. | true) = "response matrix" or "point spread function"

Deconvolution/Unfolding

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** !*

*p(**meas.** / **true**)* = “response matrix” or “point spread function”

Deconvolution/Unfolding

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

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

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**)*

→ **Likelihood-free inference**

*p(**meas.** / **true**)* = "response matrix" or "point spread function"

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

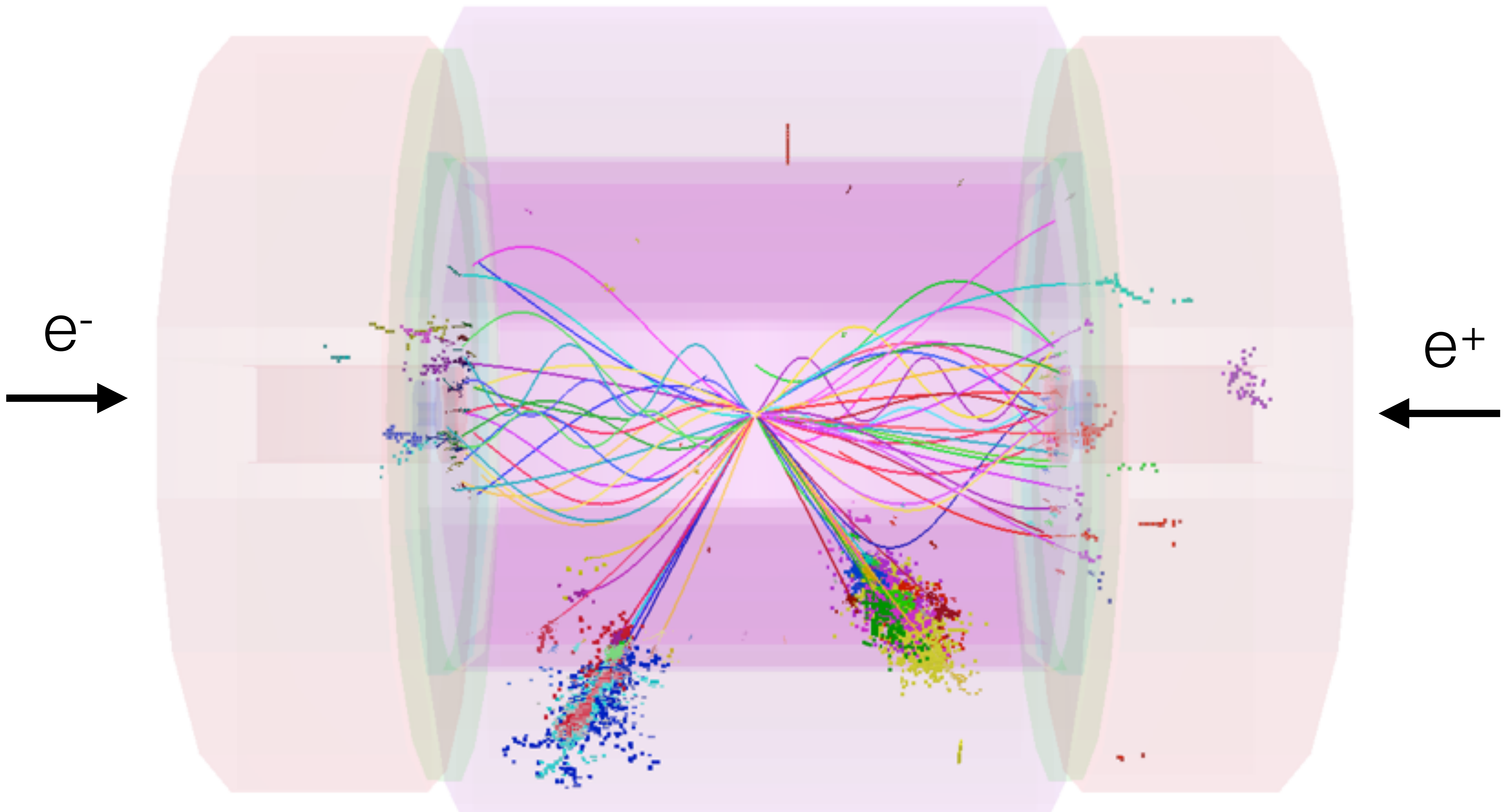
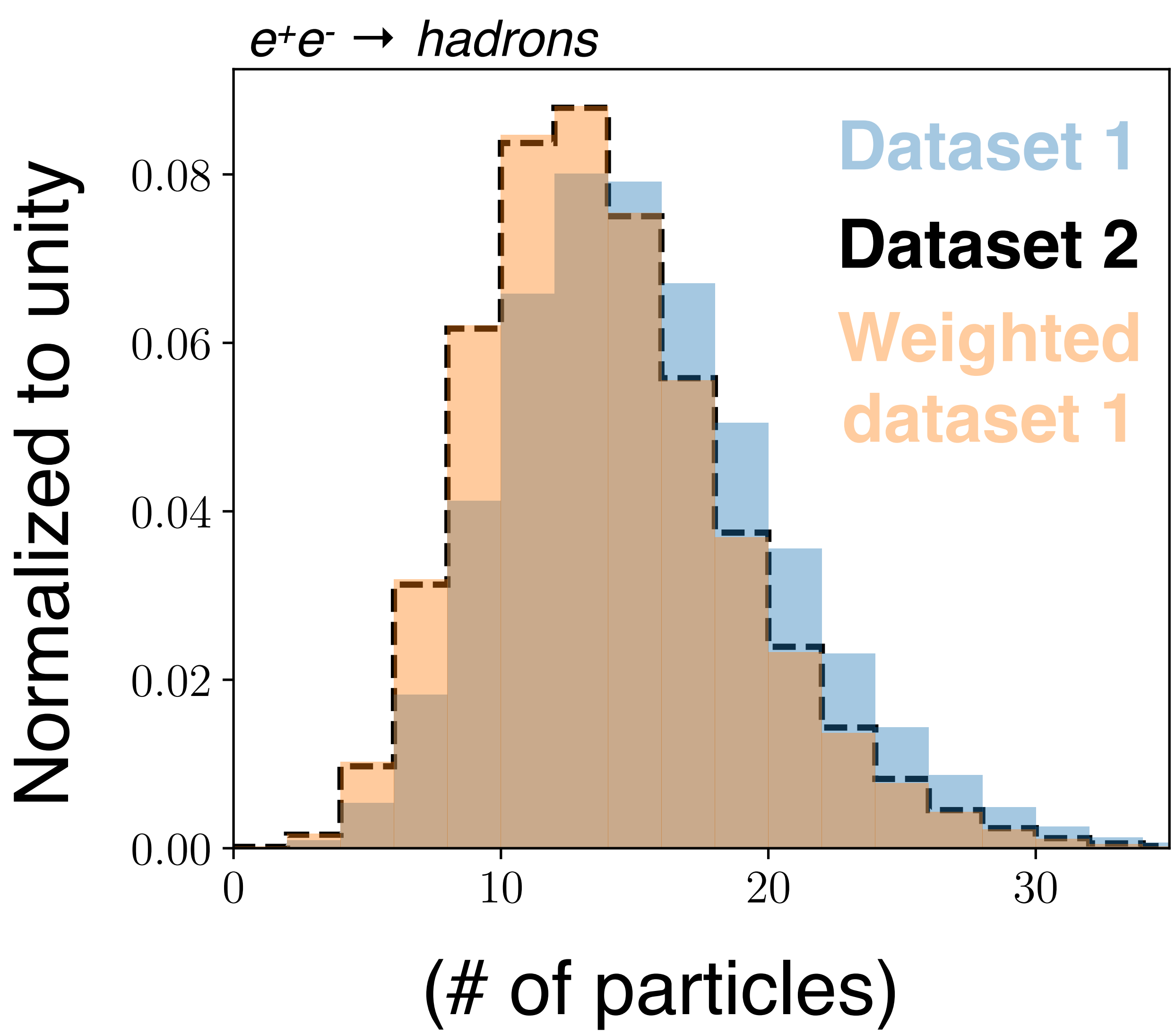
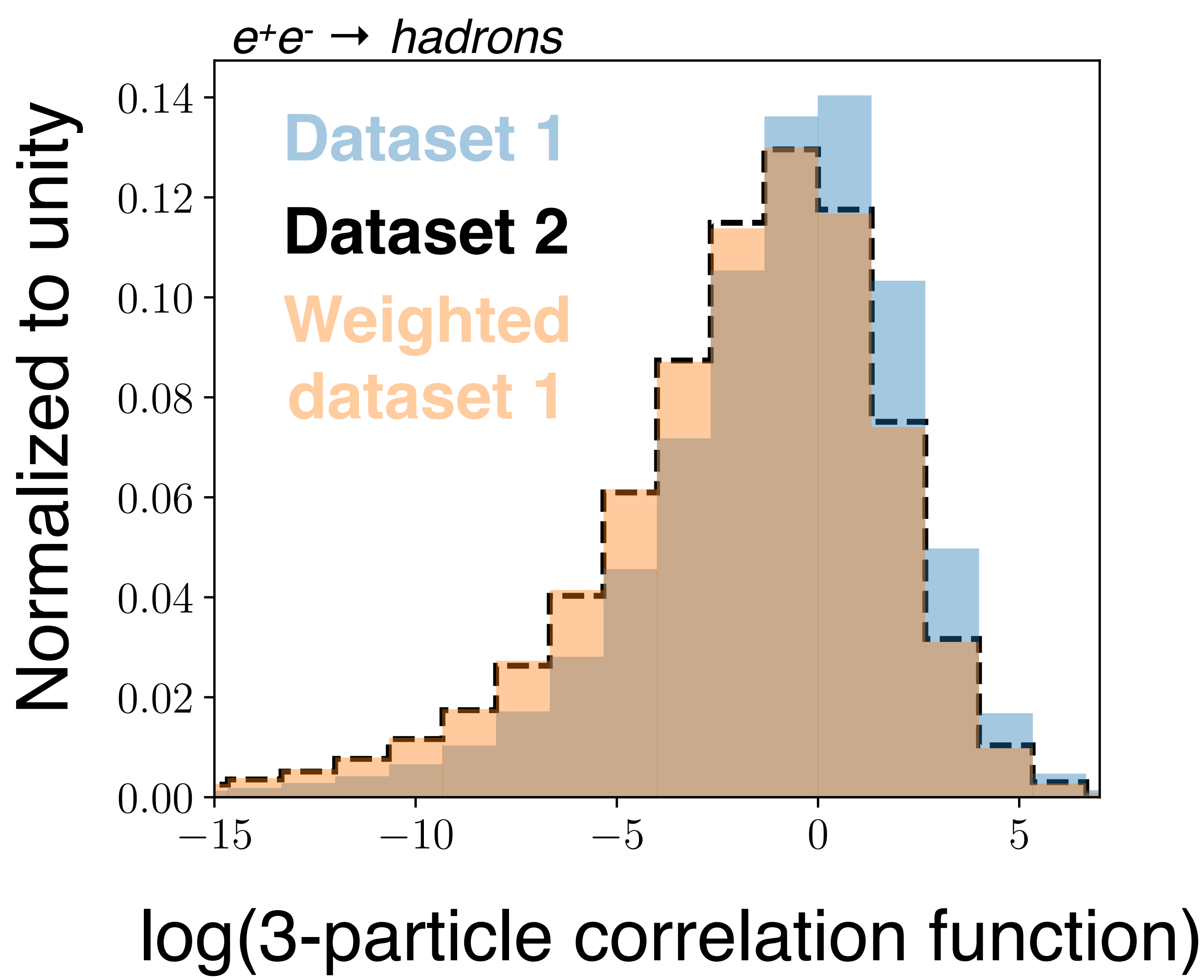


Image: Linear Collider Detector Project

Classification for reweighting

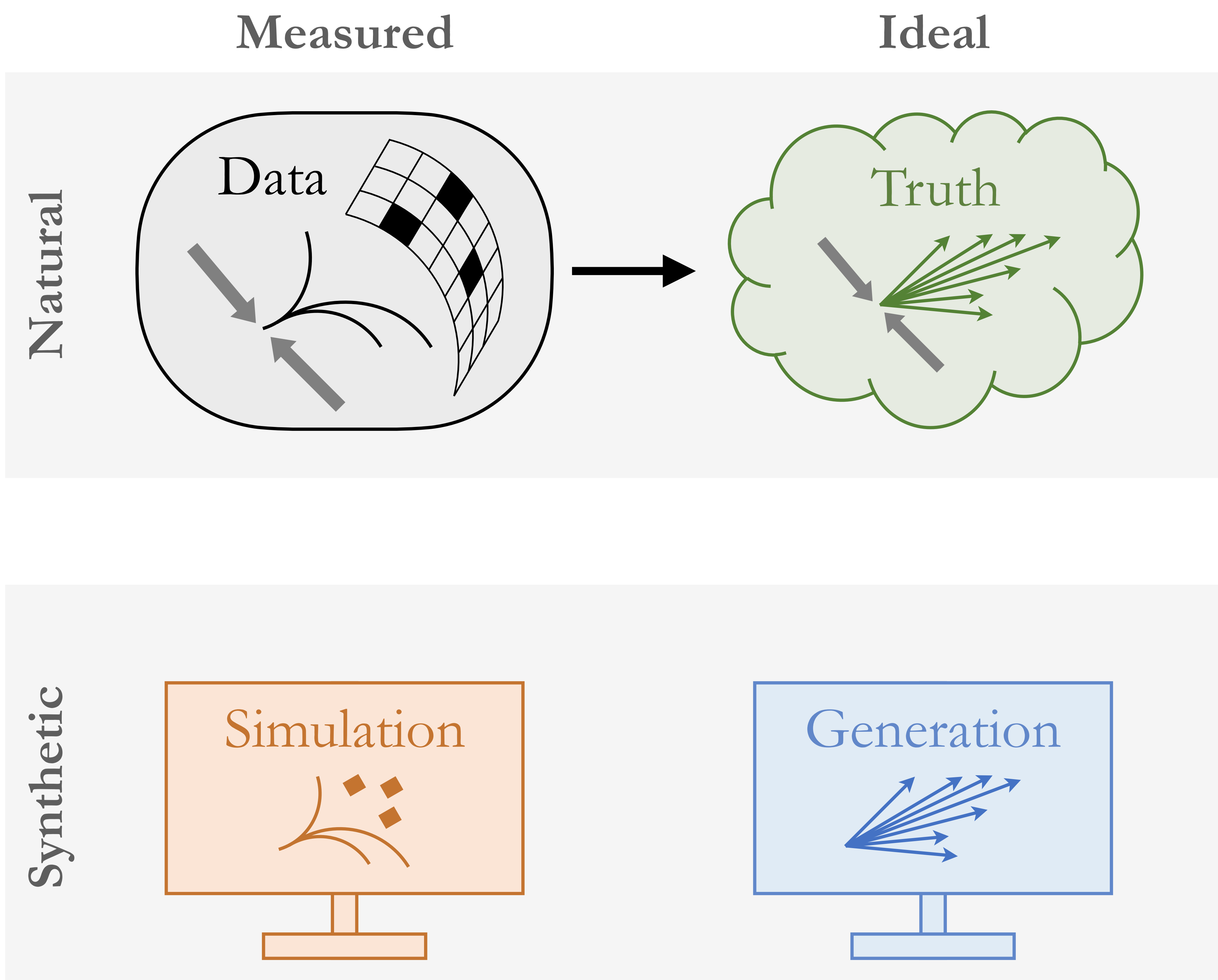
Reweight the **full phase space** and then check for various binned 1D observables.





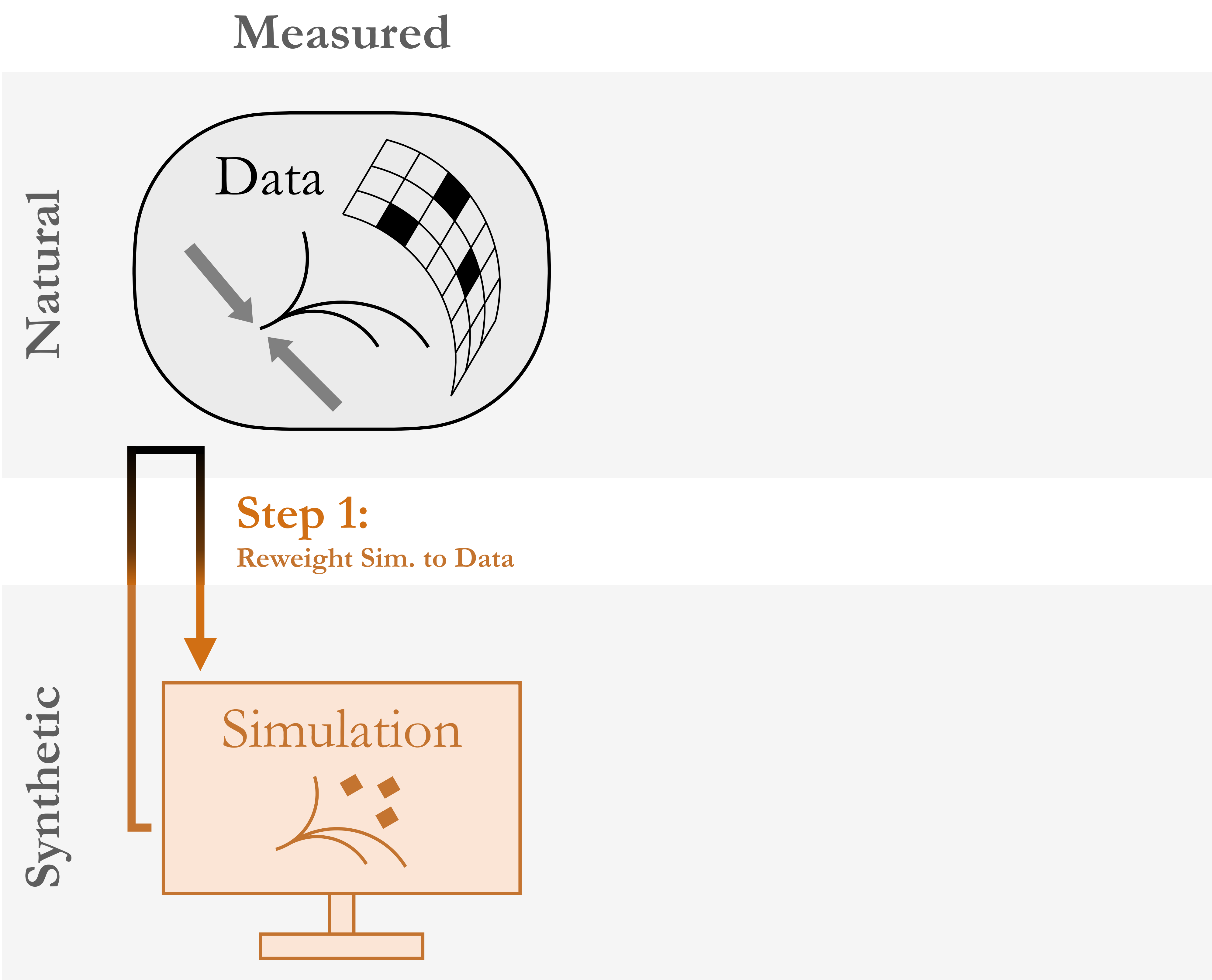
[A. Andreassen, **BPN**, 1907.08209]

Unfold by iterating: OmniFold



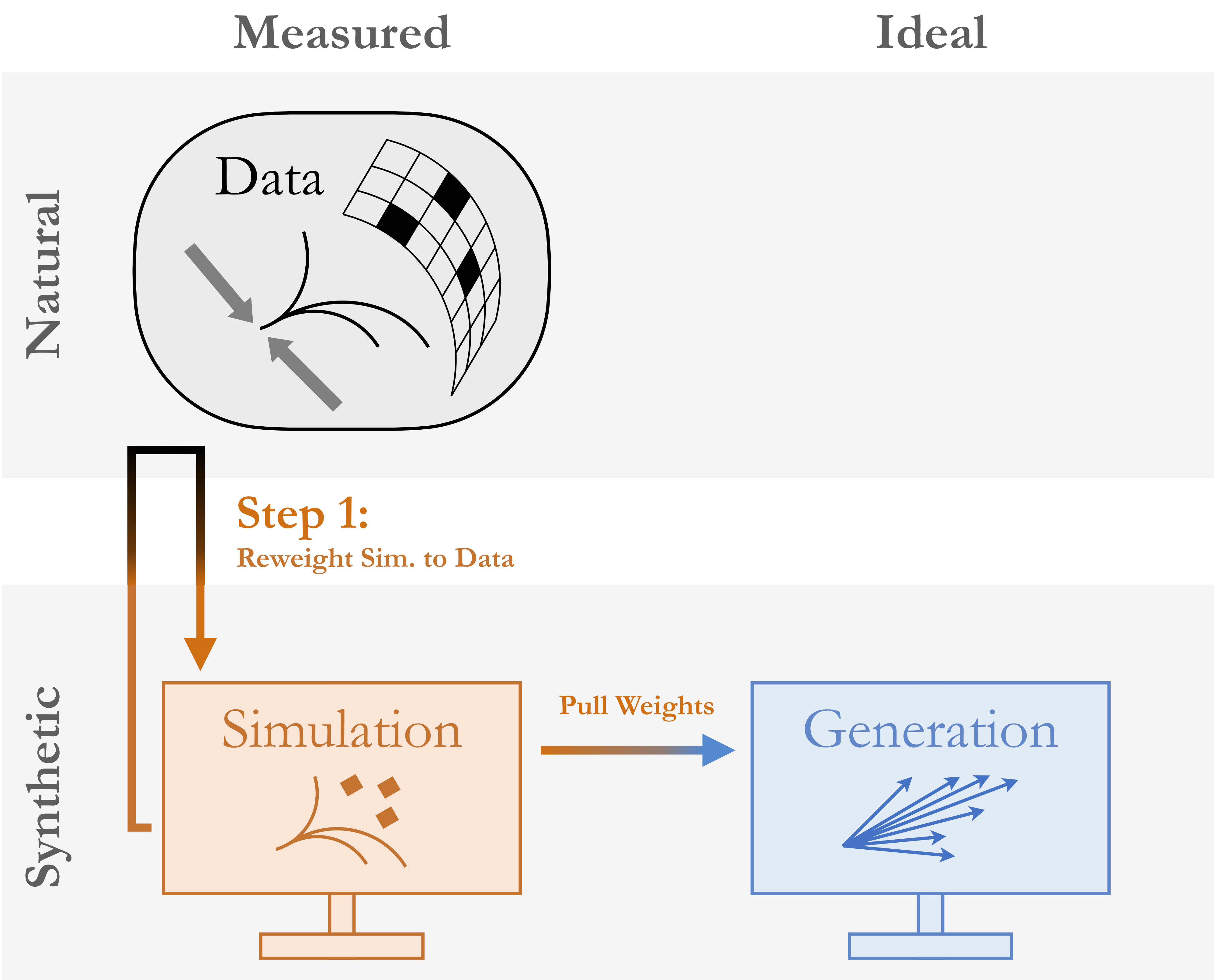
[A. Andreassen, P. Komiske, E. Metodiev, **BPN**, J. Thaler, 1907.08209]

Unfold by iterating: OmniFold



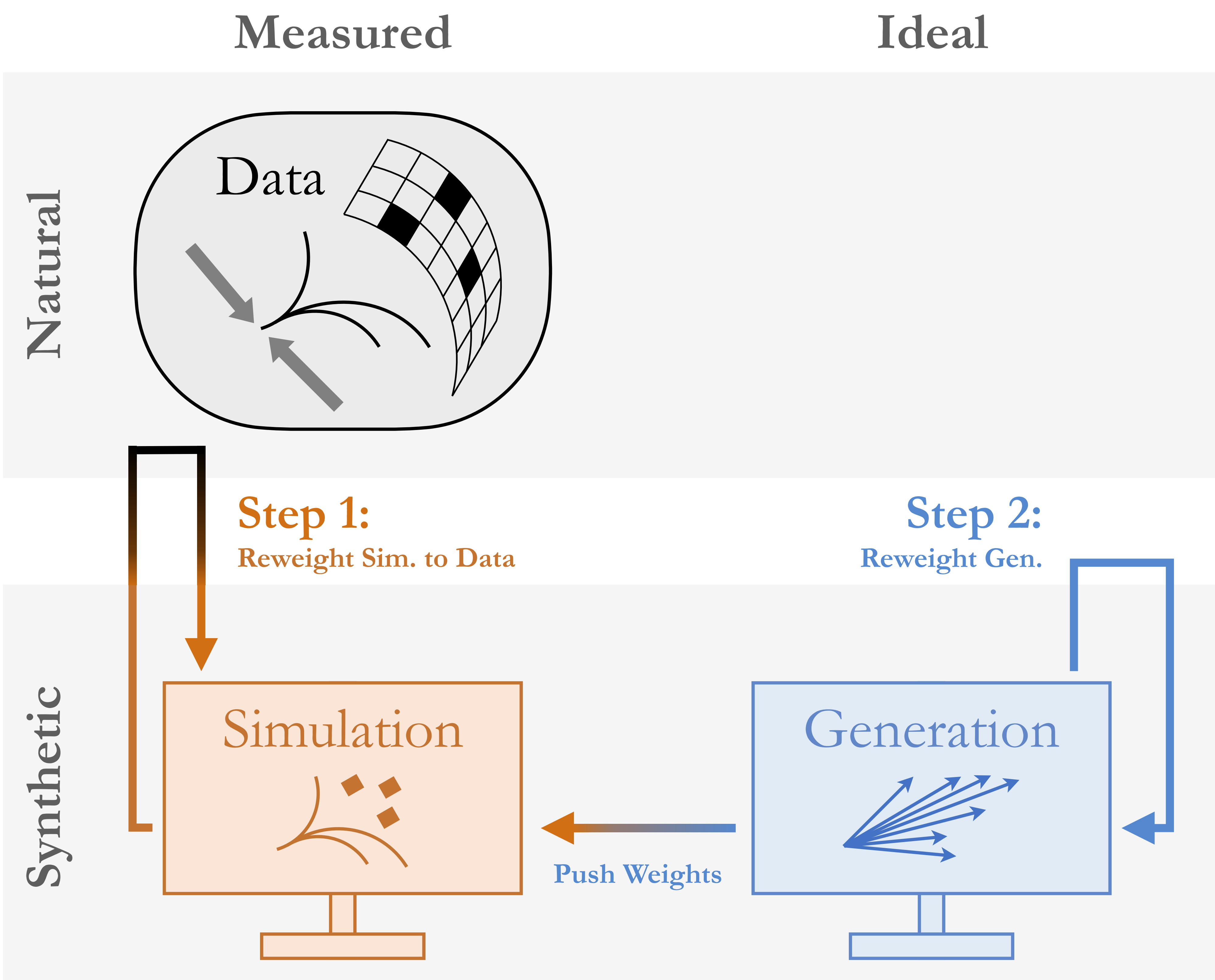
[A. Andreassen, P. Komiske, E. Metodiev, **BPN**, J. Thaler, 1907.08209]

Unfold by iterating: OmniFold



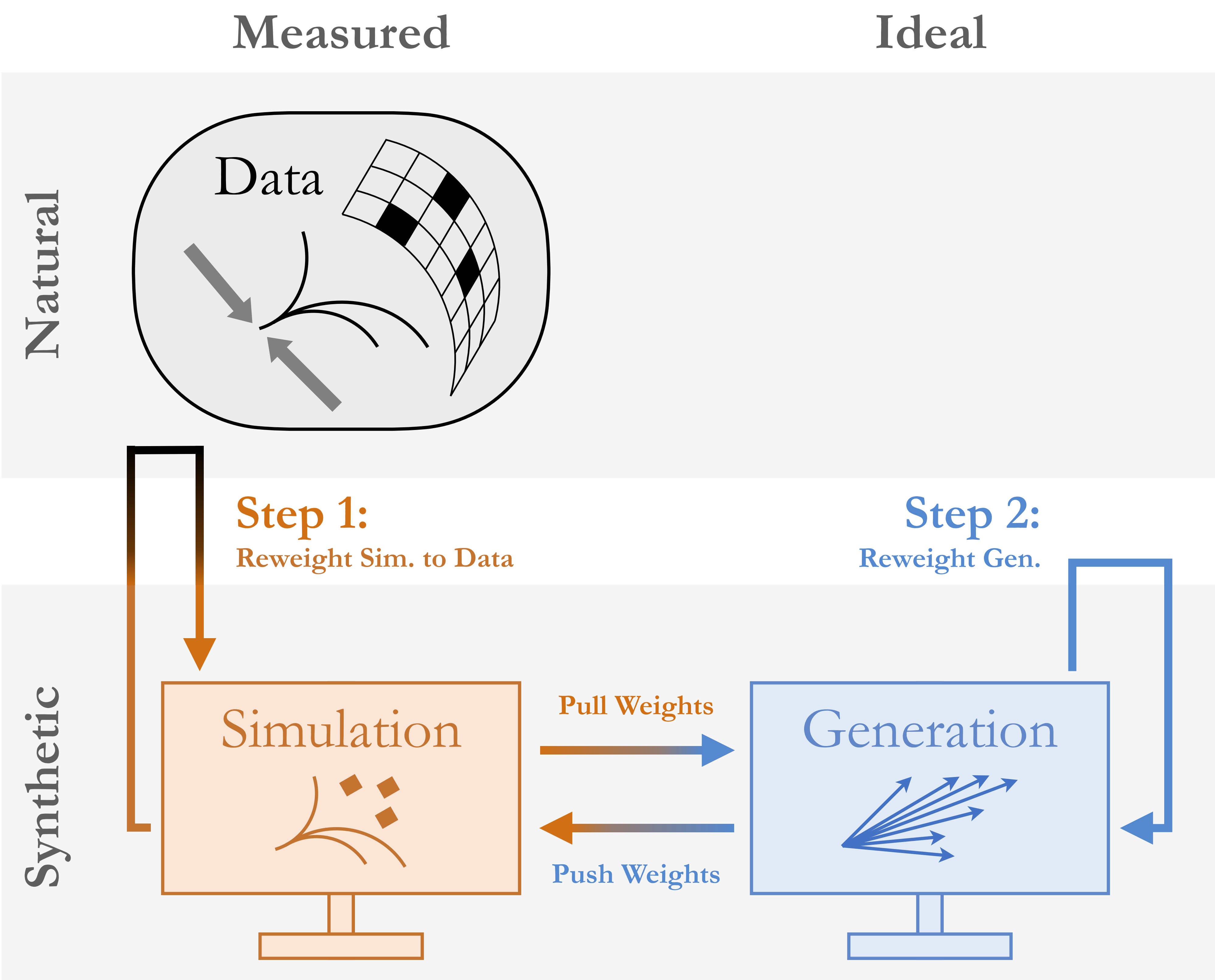
[A. Andreassen, P. Komiske, E. Metodiev, **BPN**, J. Thaler, 1907.08209]

Unfold by iterating: OmniFold



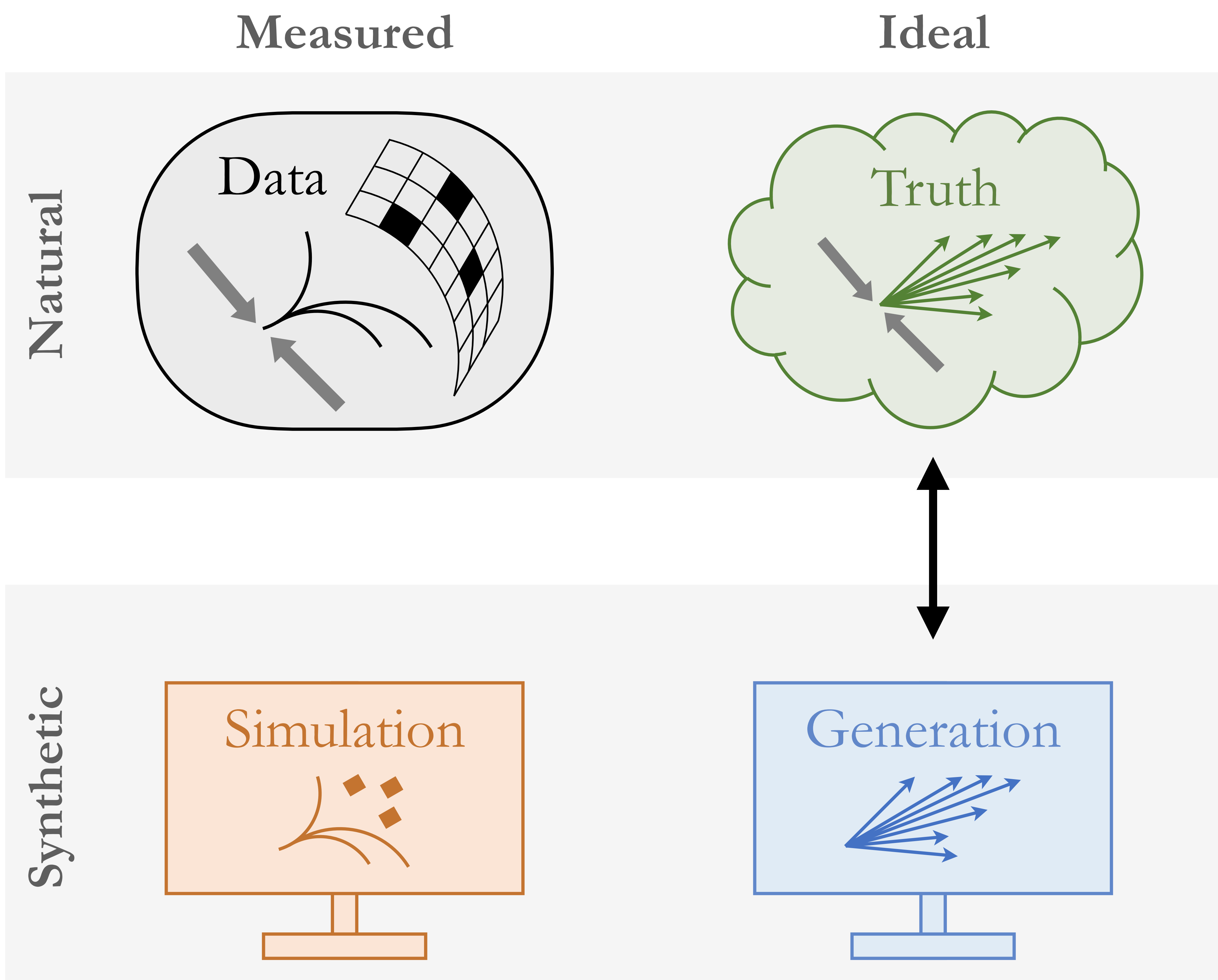
[A. Andreassen, P. Komiske, E. Metodiev, **BPN**, J. Thaler, 1907.08209]

Unfold by iterating: OmniFold



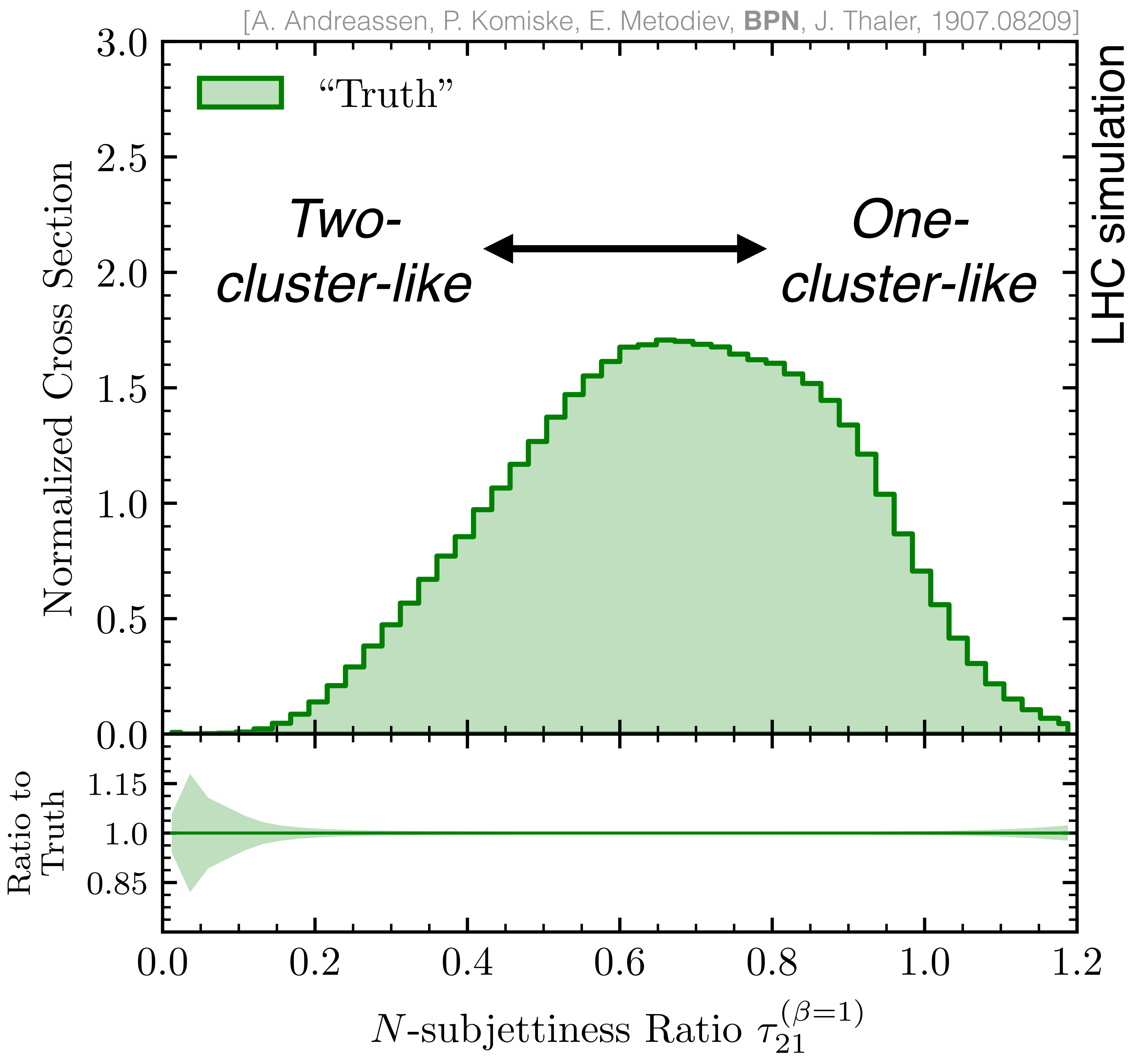
[A. Andreassen, P. Komiske, E. Metodiev, **BPN**, J. Thaler, 1907.08209]

Unfold by iterating: OmniFold



[A. Andreassen, P. Komiske, E. Metodiev, **BPN**, J. Thaler, 1907.08209]

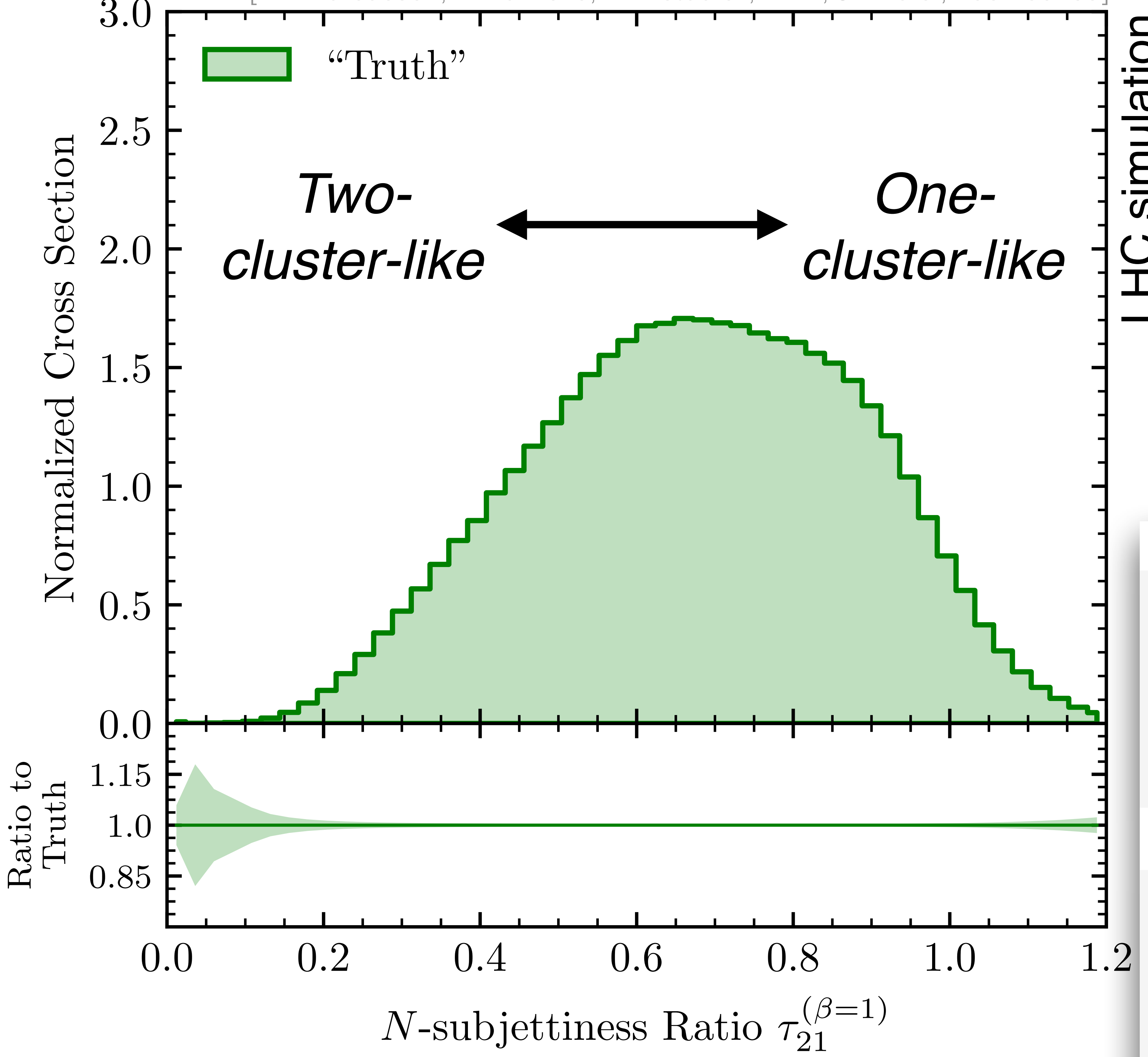
Results



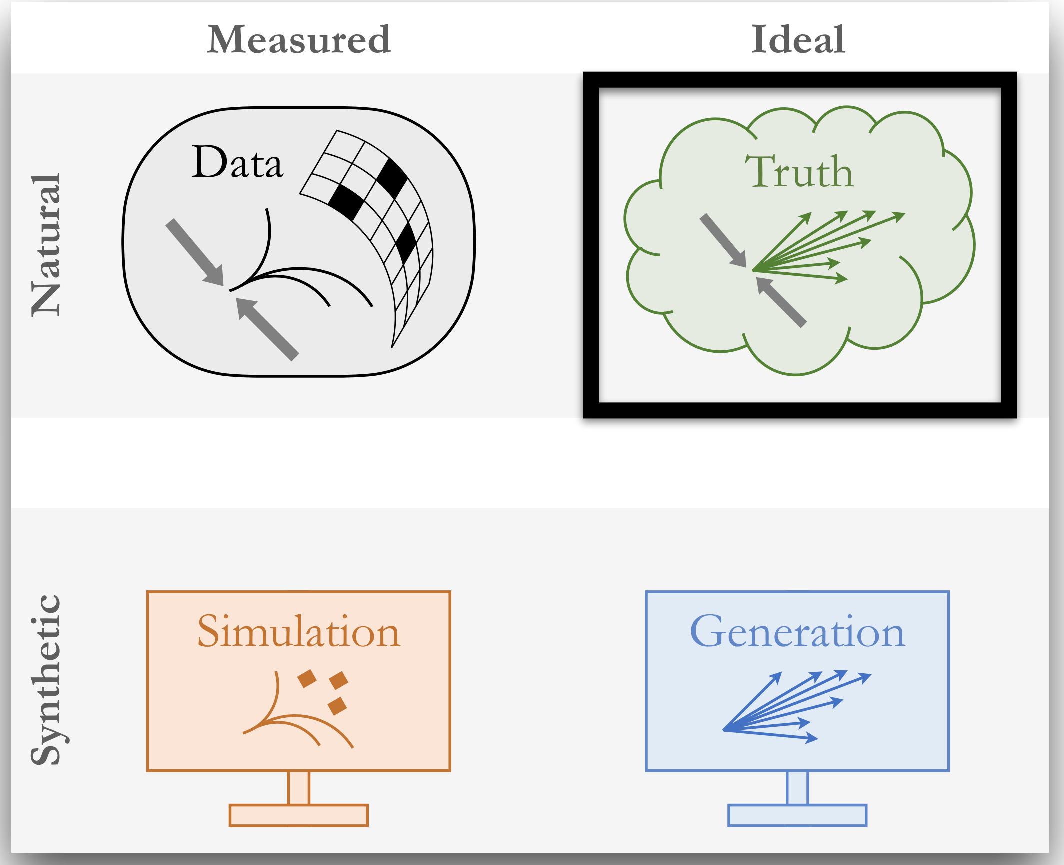
Consider this observable, which characterizes the substructure

Results

[A. Andreassen, P. Komiske, E. Metodiev, **BPN**, J. Thaler, 1907.08209]

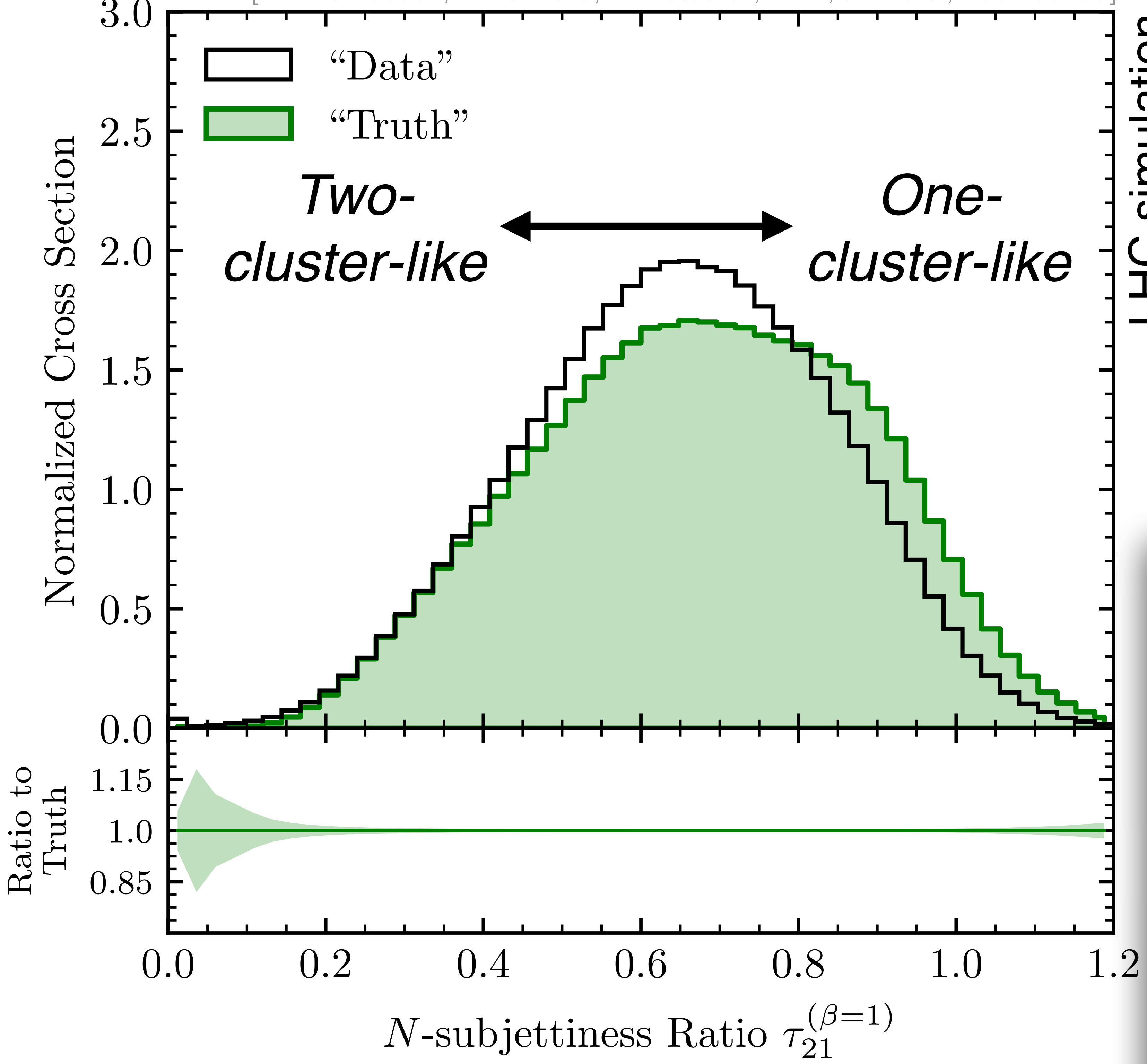


Consider this observable, which characterizes the substructure

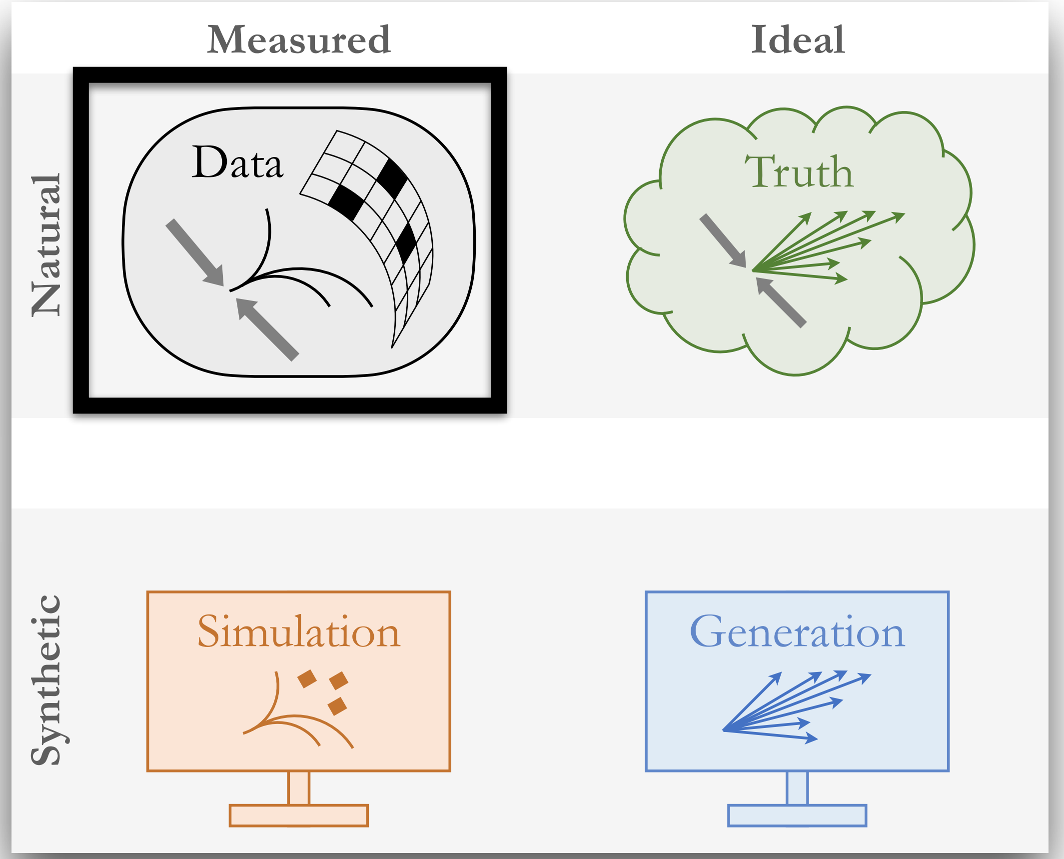


Results

[A. Andreassen, P. Komiske, E. Metodiev, **BPN**, J. Thaler, 1907.08209]

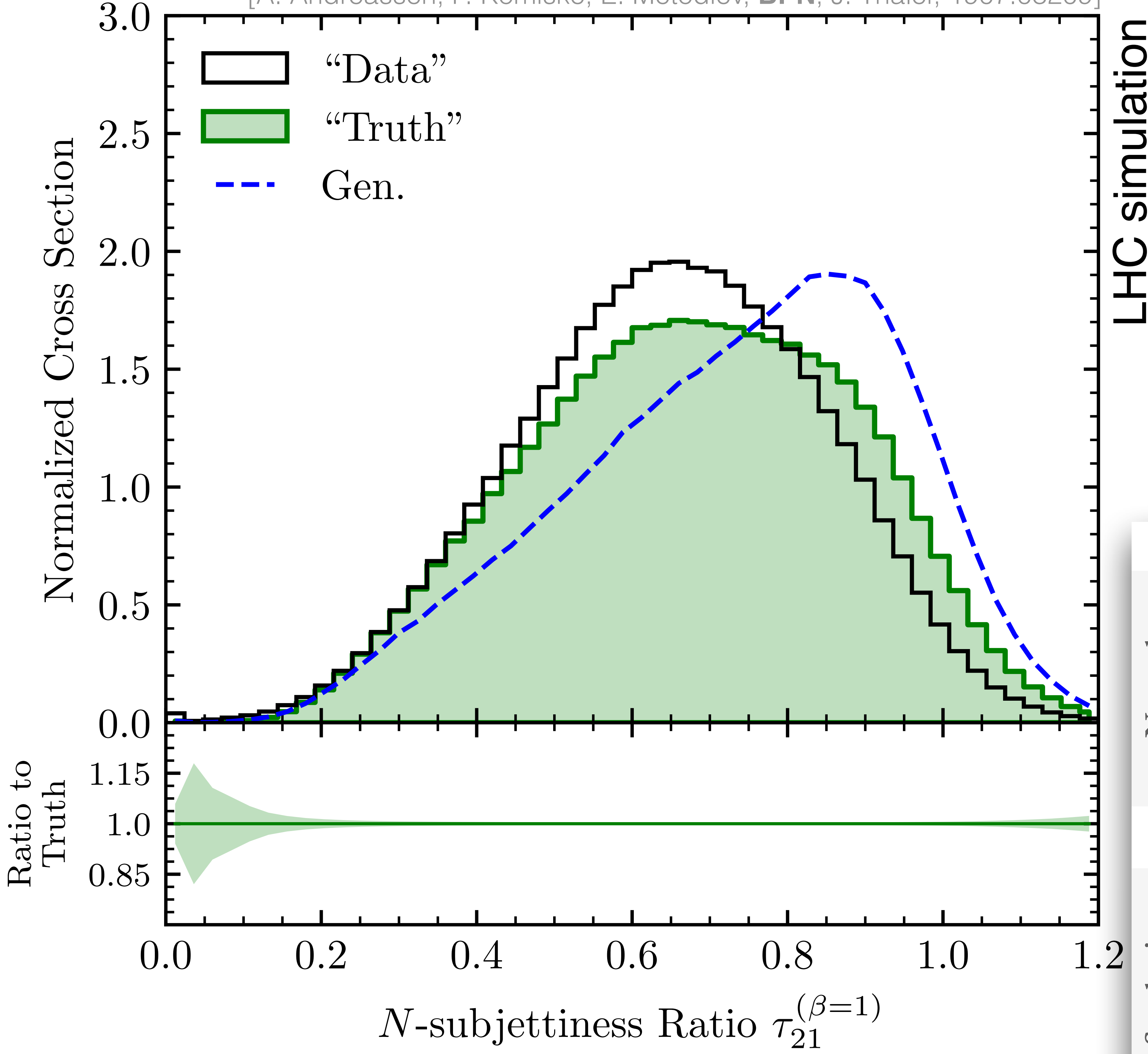


Consider this observable, which characterizes the substructure

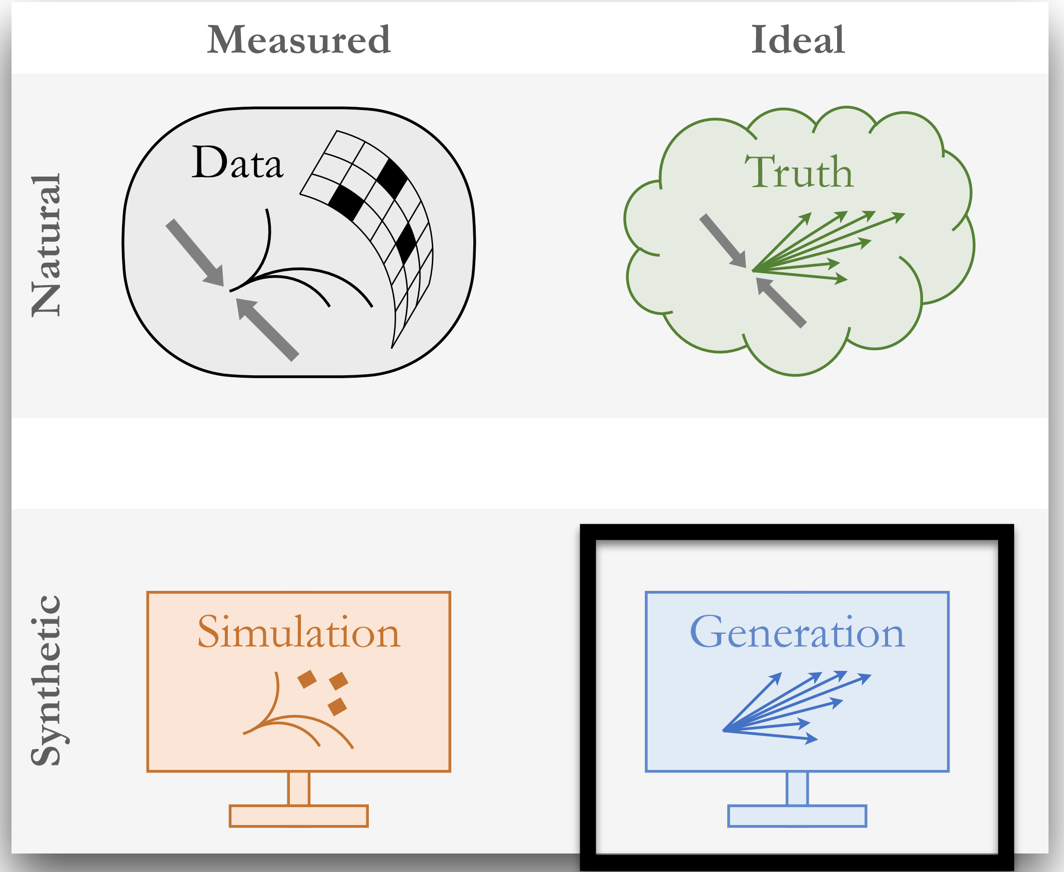


Results

[A. Andreassen, P. Komiske, E. Metodiev, **BPN**, J. Thaler, 1907.08209]

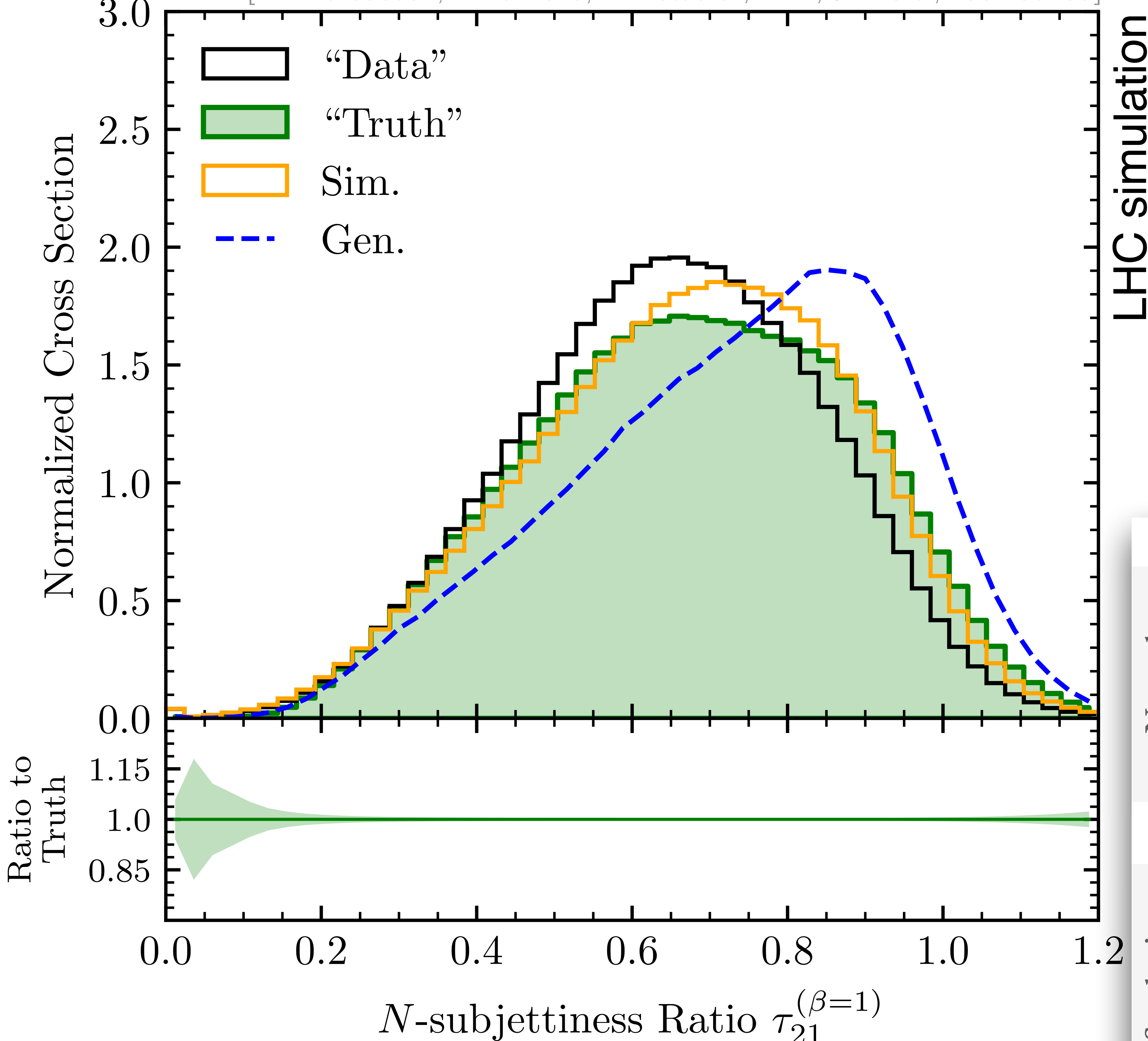


Consider this observable, which characterizes the substructure

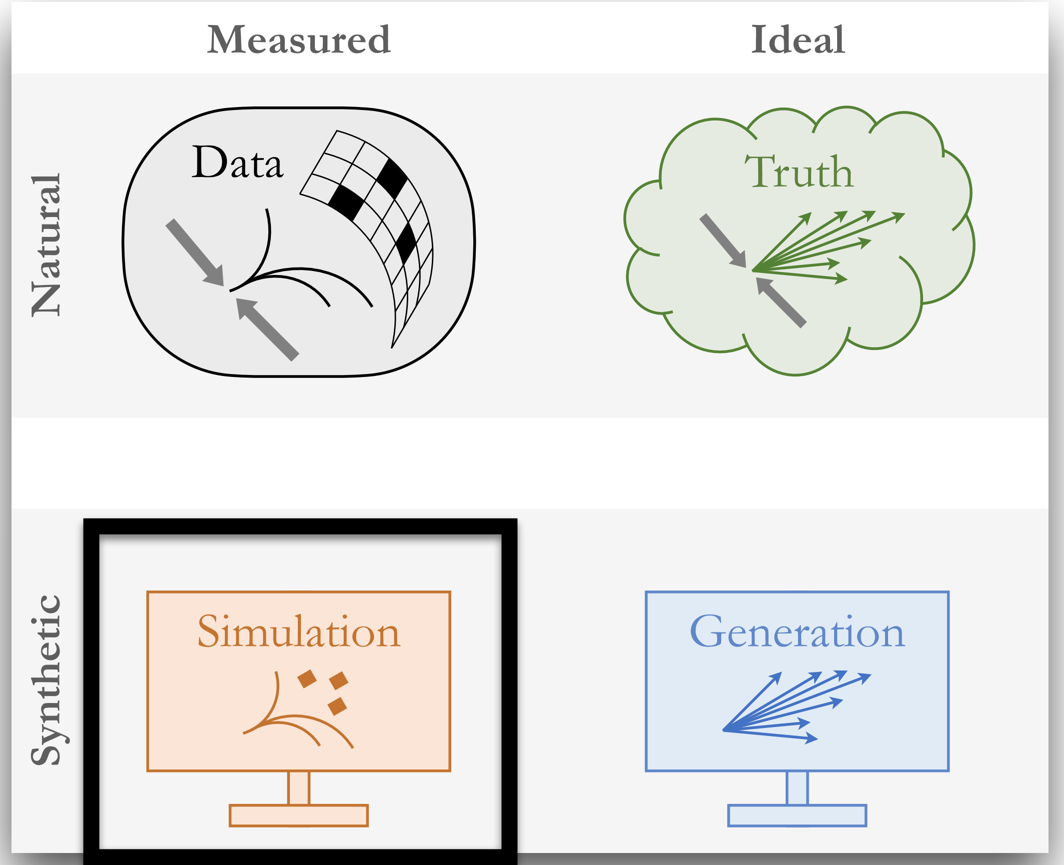


Results

[A. Andreassen, P. Komiske, E. Metodiev, **BPN**, J. Thaler, 1907.08209]

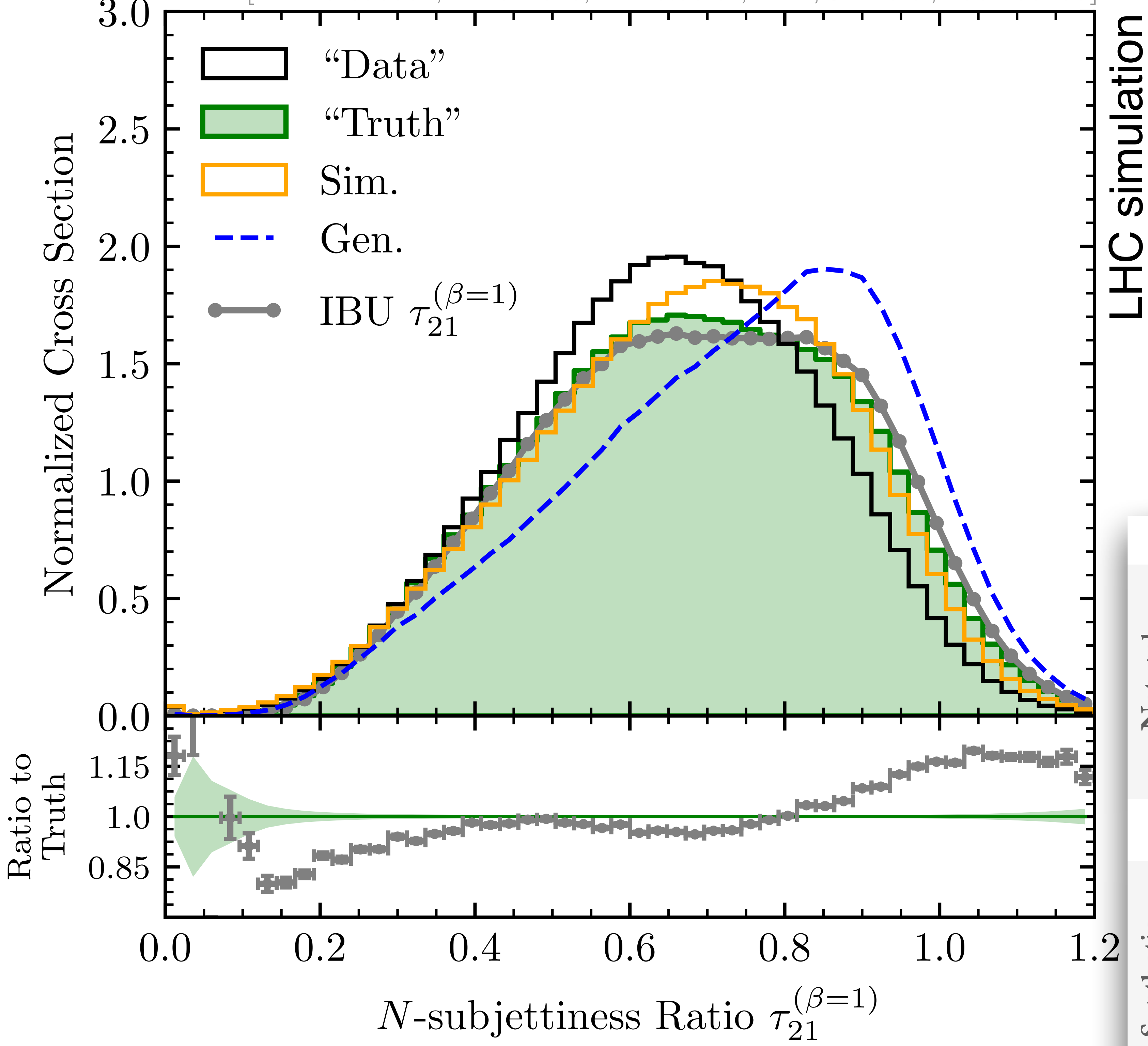


Consider this observable, which characterizes the substructure

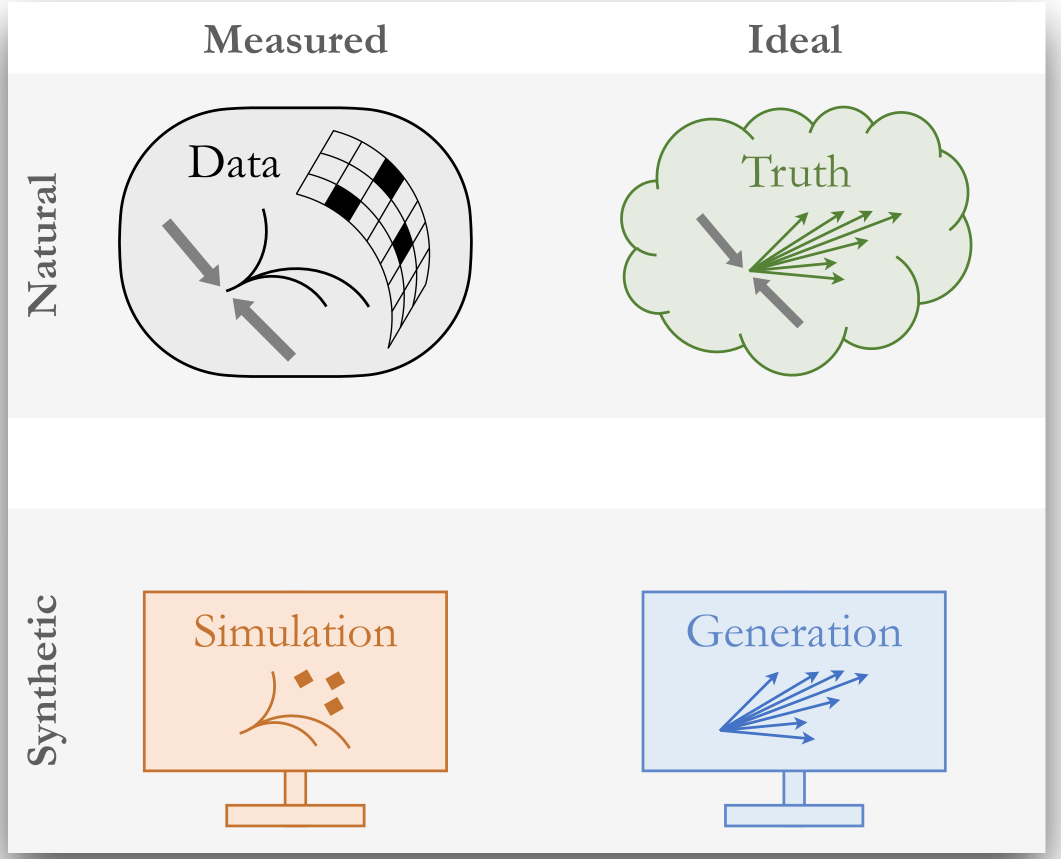


Results

[A. Andreassen, P. Komiske, E. Metodiev, **BPN**, J. Thaler, 1907.08209]

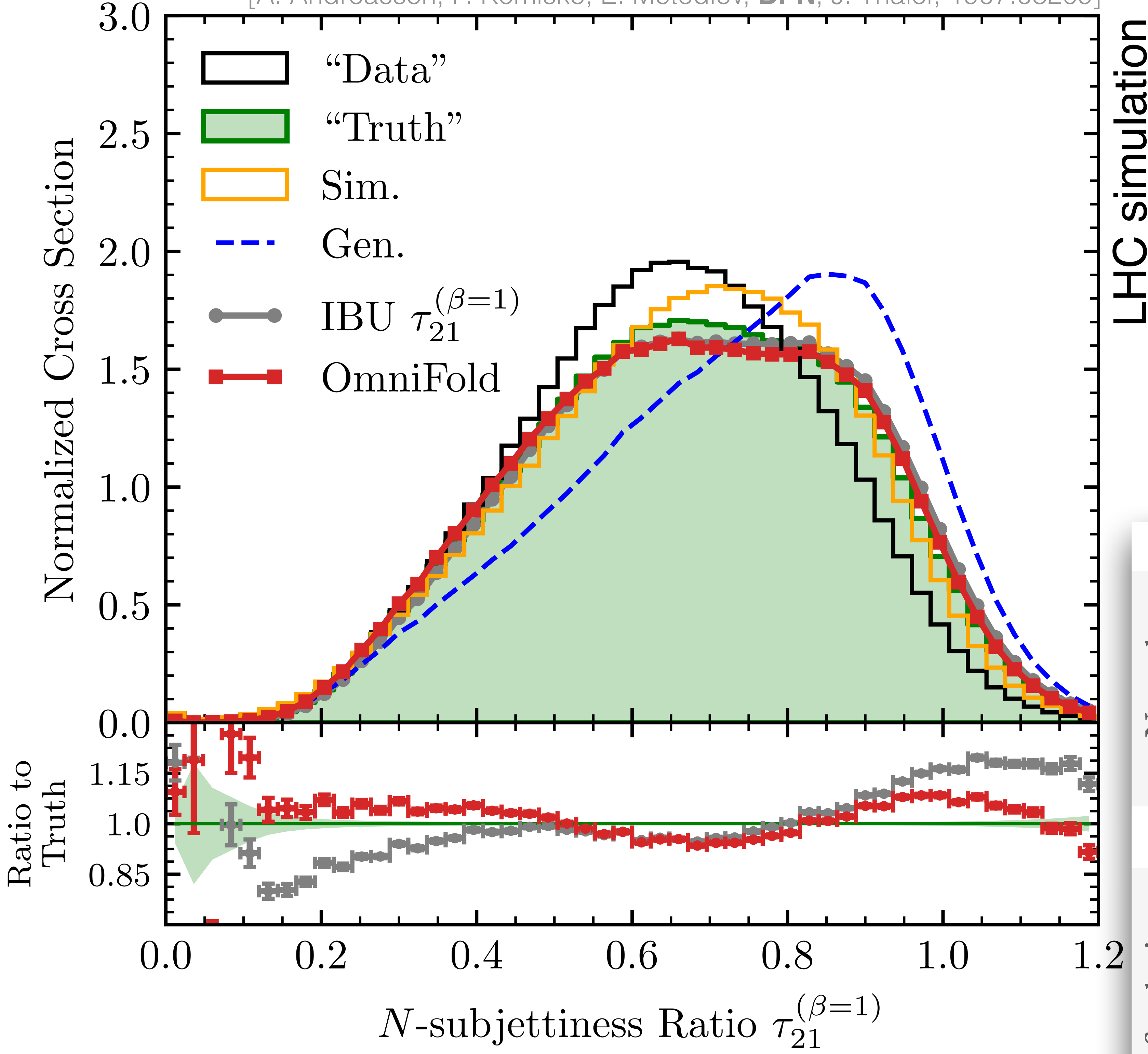


IBU is the current standard. It is a1D binned and iterative approach.

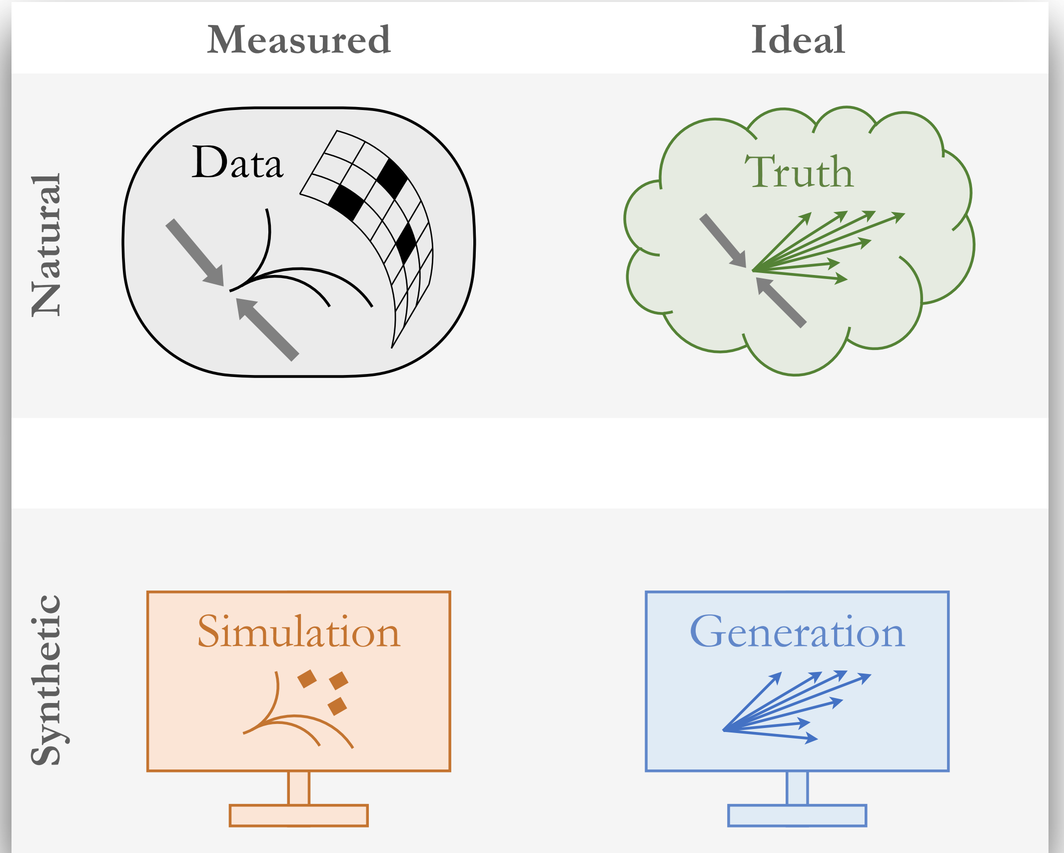


Results

[A. Andreassen, P. Komiske, E. Metodiev, **BPN**, J. Thaler, 1907.08209]

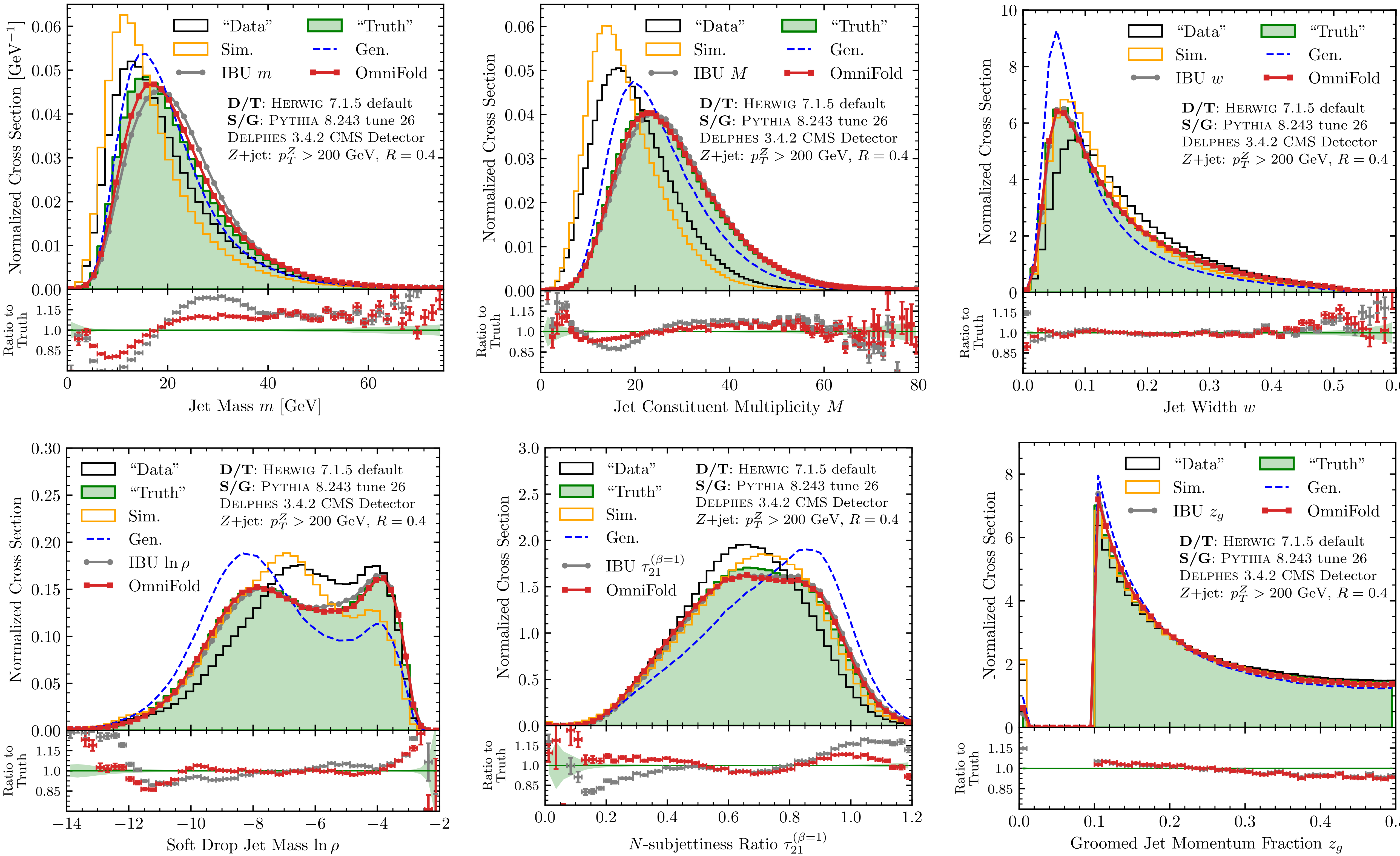


OmniFold outperforms IBU even though it is not tailored to this observable



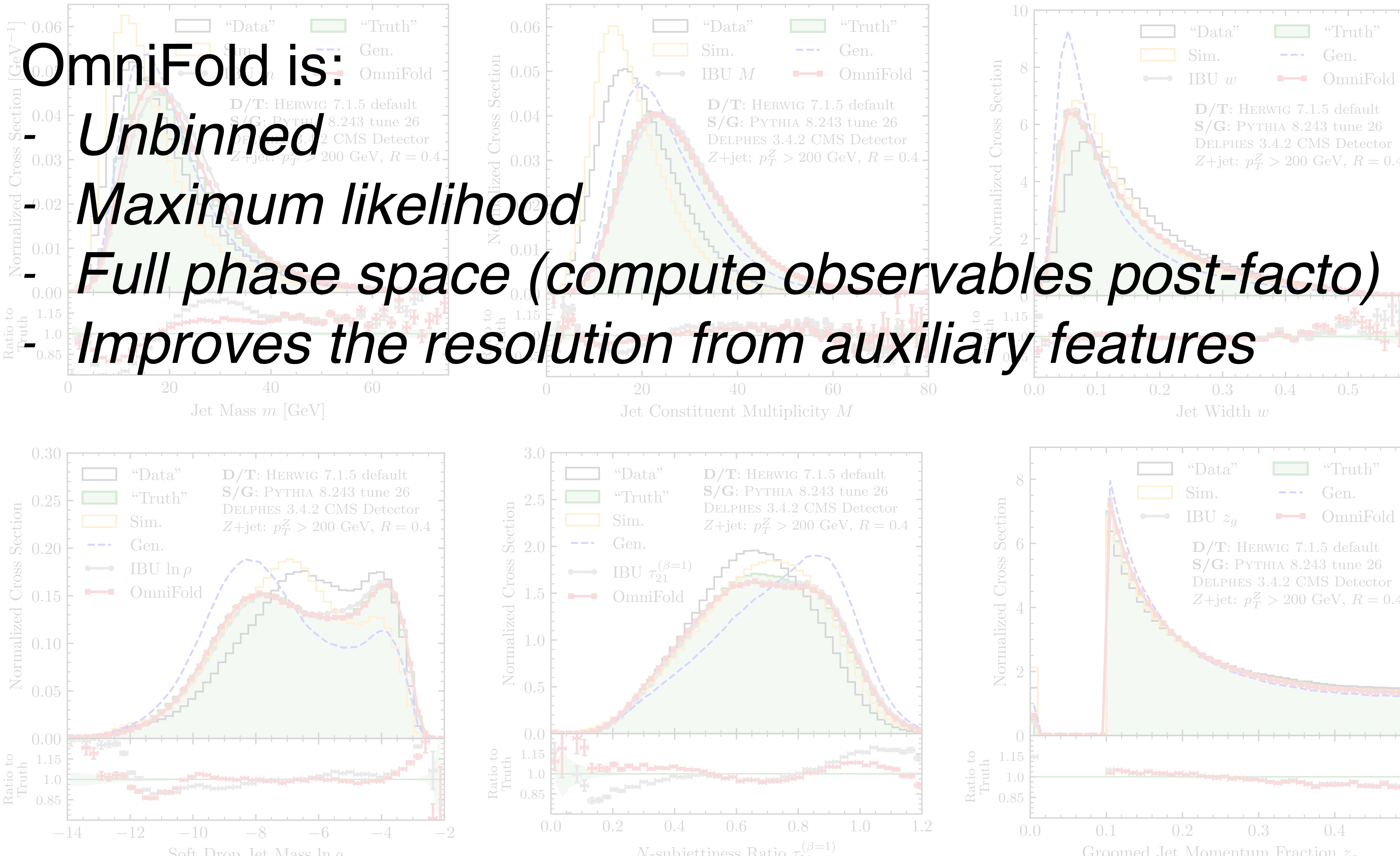
Results

[A. Andreassen, P. Komiske, E. Metodiev, **BPN**, J. Thaler, 1907.08209]



Results

[A. Andreassen, P. Komiske, E. Metodiev, **BPN**, J. Thaler, 1907.08209]



OmniFold is:

- *Unbinned*
- *Maximum likelihood*
- *Full phase space (compute observables post-facto)*
- *Improves the resolution from auxiliary features*

Generative Models

So far, I have assumed that we can readily generate high-fidelity examples from a physics-based simulator.

Generative Models

So far, I have assumed that we can readily generate high-fidelity examples from a physics-based simulator.

What if the simulator is too **expensive**?

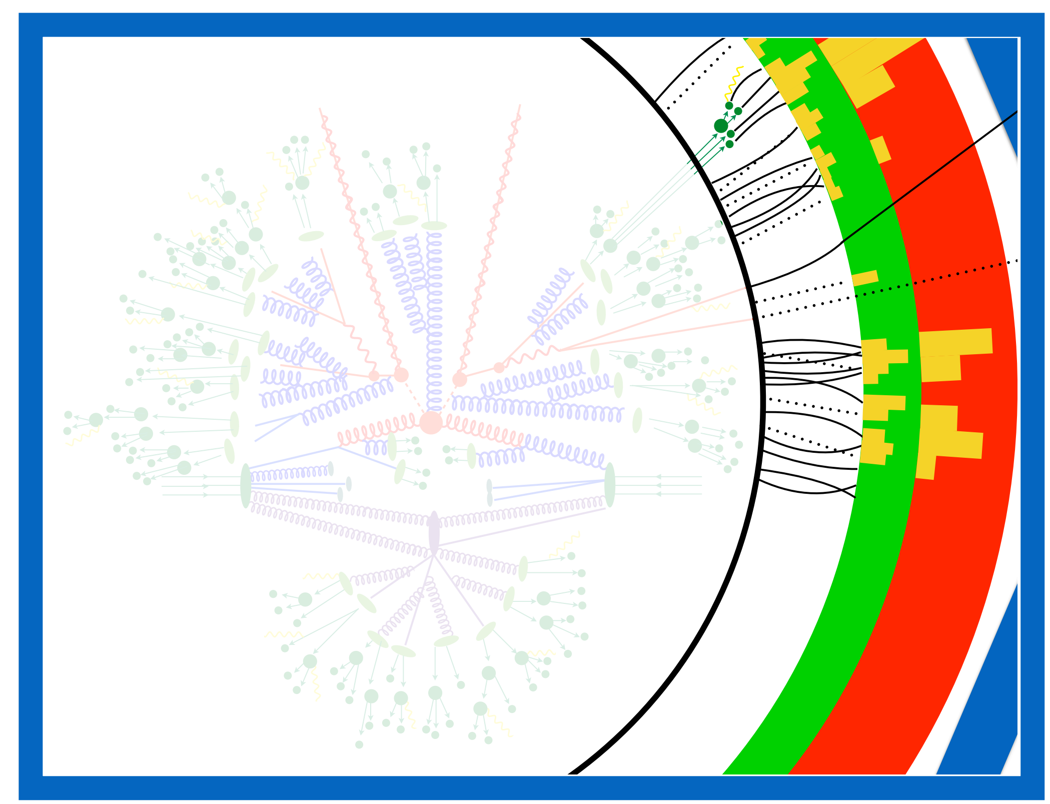
Generative Models

So far, I have assumed that we can readily generate high-fidelity examples from a physics-based simulator.

What if the simulator is too **expensive**?

GEANT4
A SIMULATION TOOLKIT

For example, Geant is an O(1) fraction of all computing in high energy physics

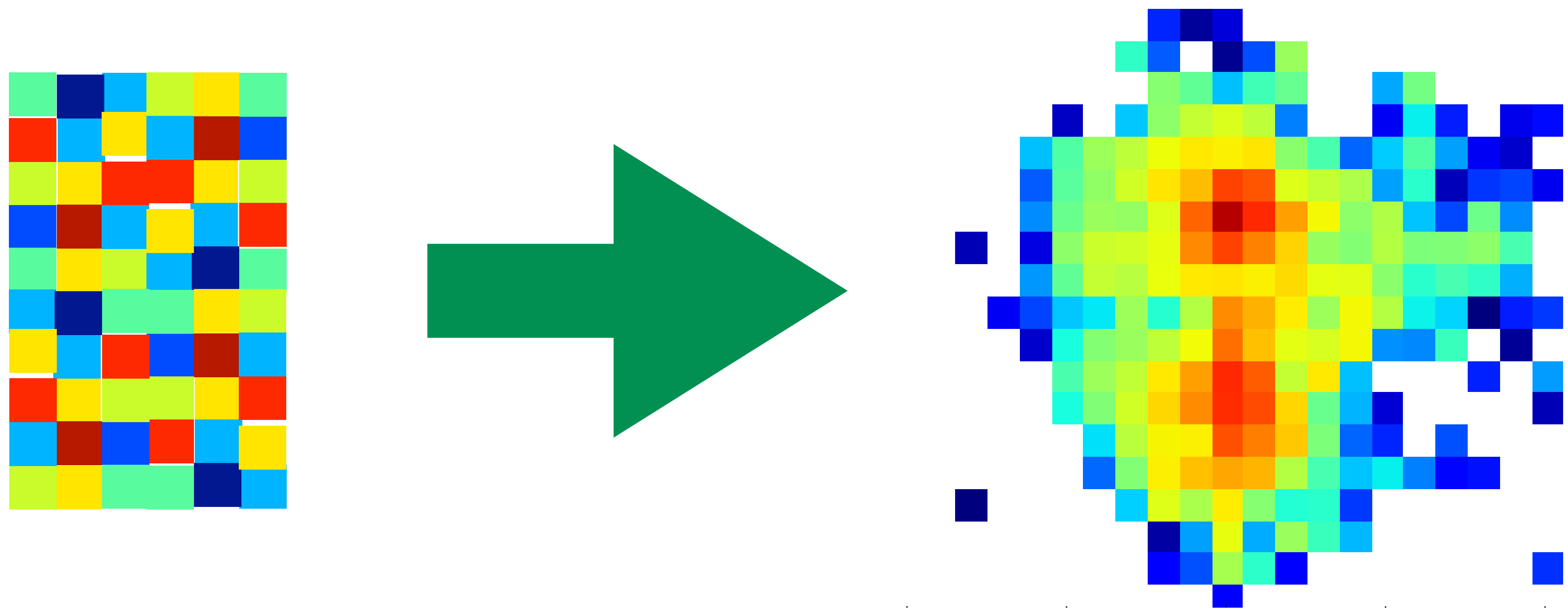


Deep Generative Models

Can we combine our physics-simulator with deep learning?

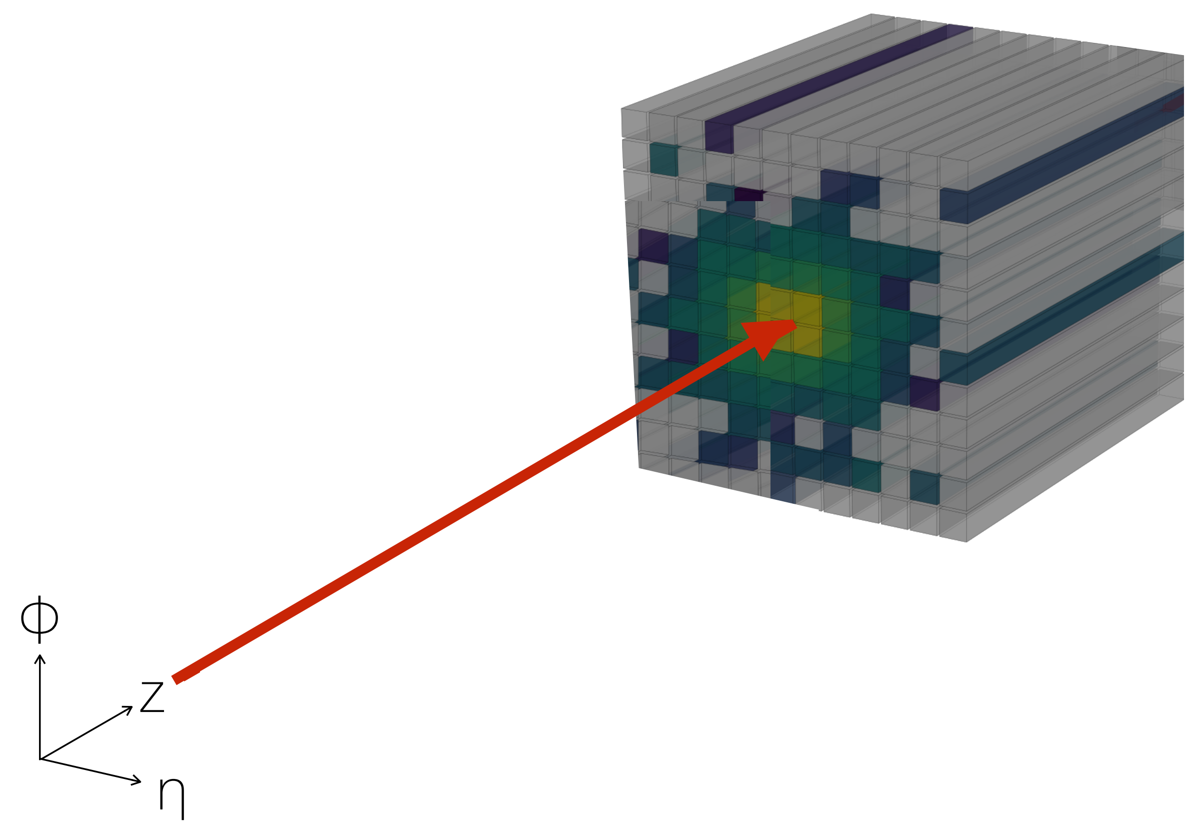
Can we combine our physics-simulator with deep learning?

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



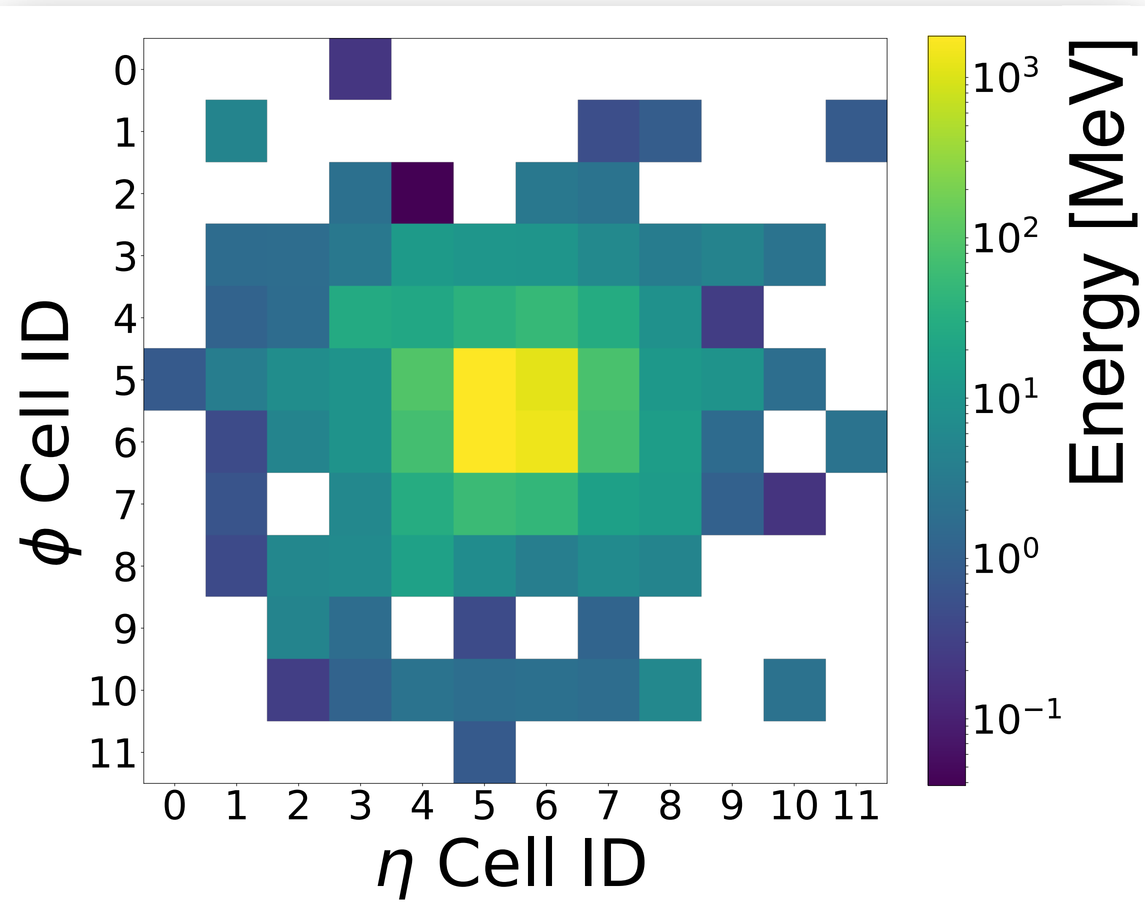
Calorimeter images

The slowest simulation corresponds to the densest material (= calorimeter)



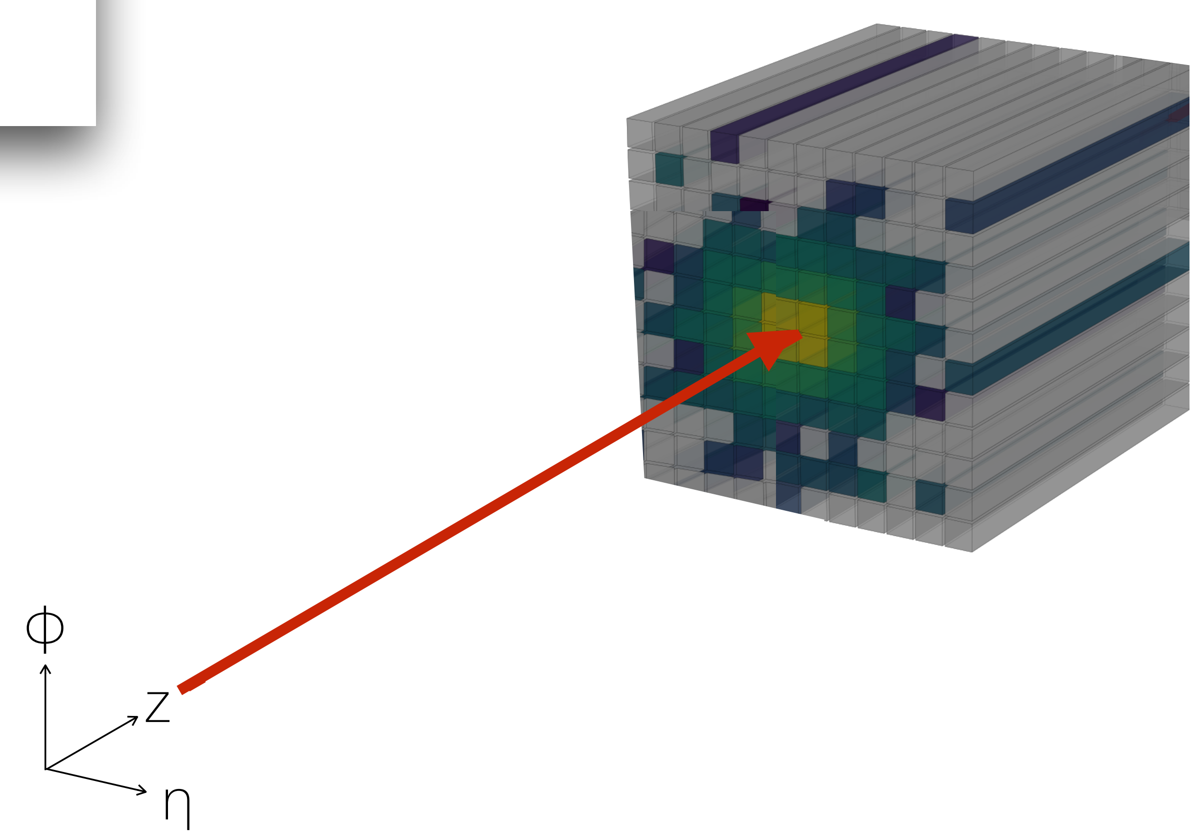
M. Paganini, L. de Oliveira, **BPN**, PRL 120 (2018) 042003

Calorimeter images



The slowest simulation corresponds to the densest material (= calorimeter)

Grayscale images:
Pixel intensity = energy deposited

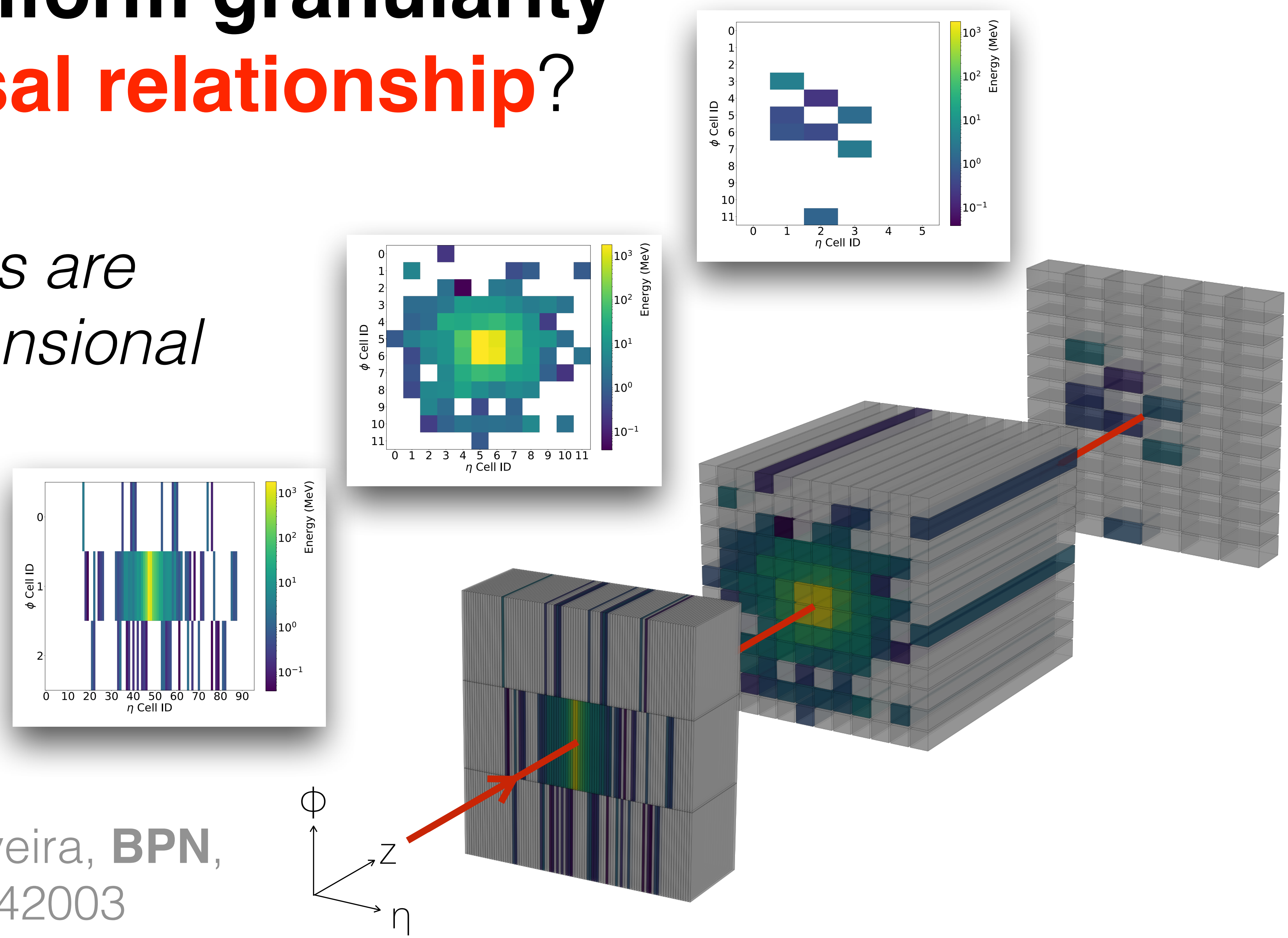


M. Paganini, L. de Oliveira, **BPN**,
PRL 120 (2018) 042003

Calorimeter images

Challenge: **multiple layers** with **non-uniform granularity** and a **causal relationship**?

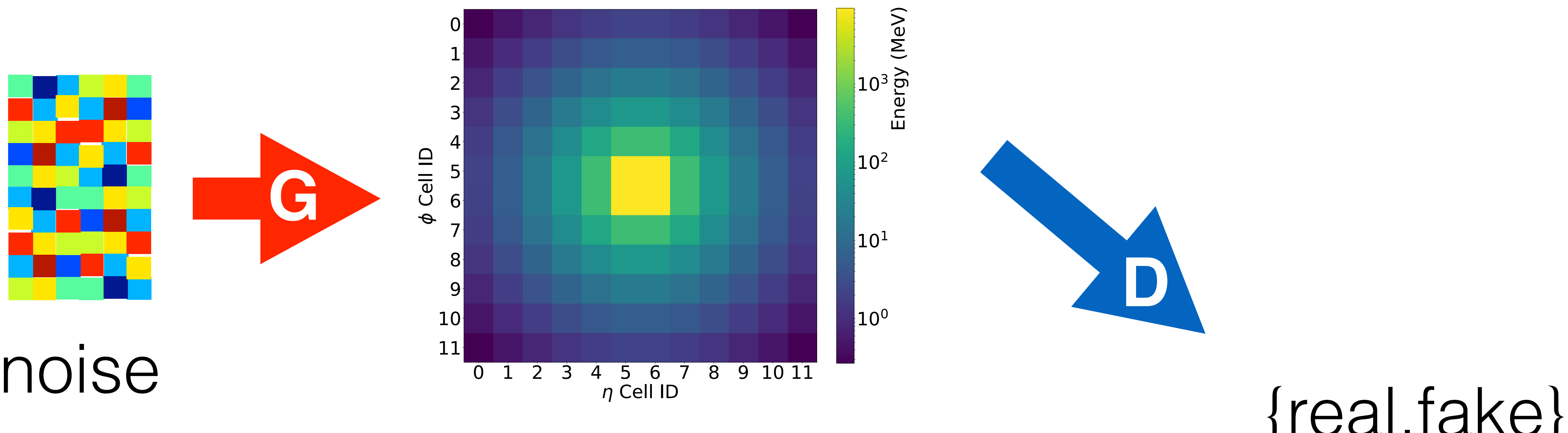
N.B. images are O(1000) dimensional



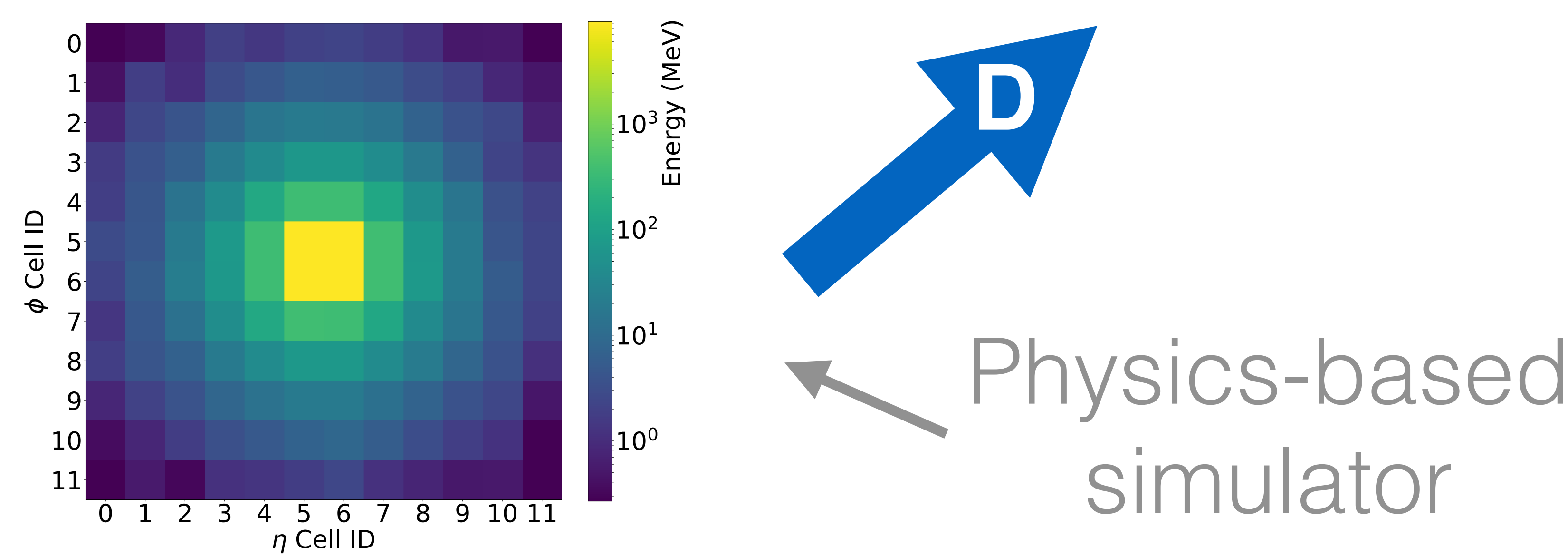
M. Paganini, L. de Oliveira, **BPN**, PRL 120 (2018) 042003

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**.*

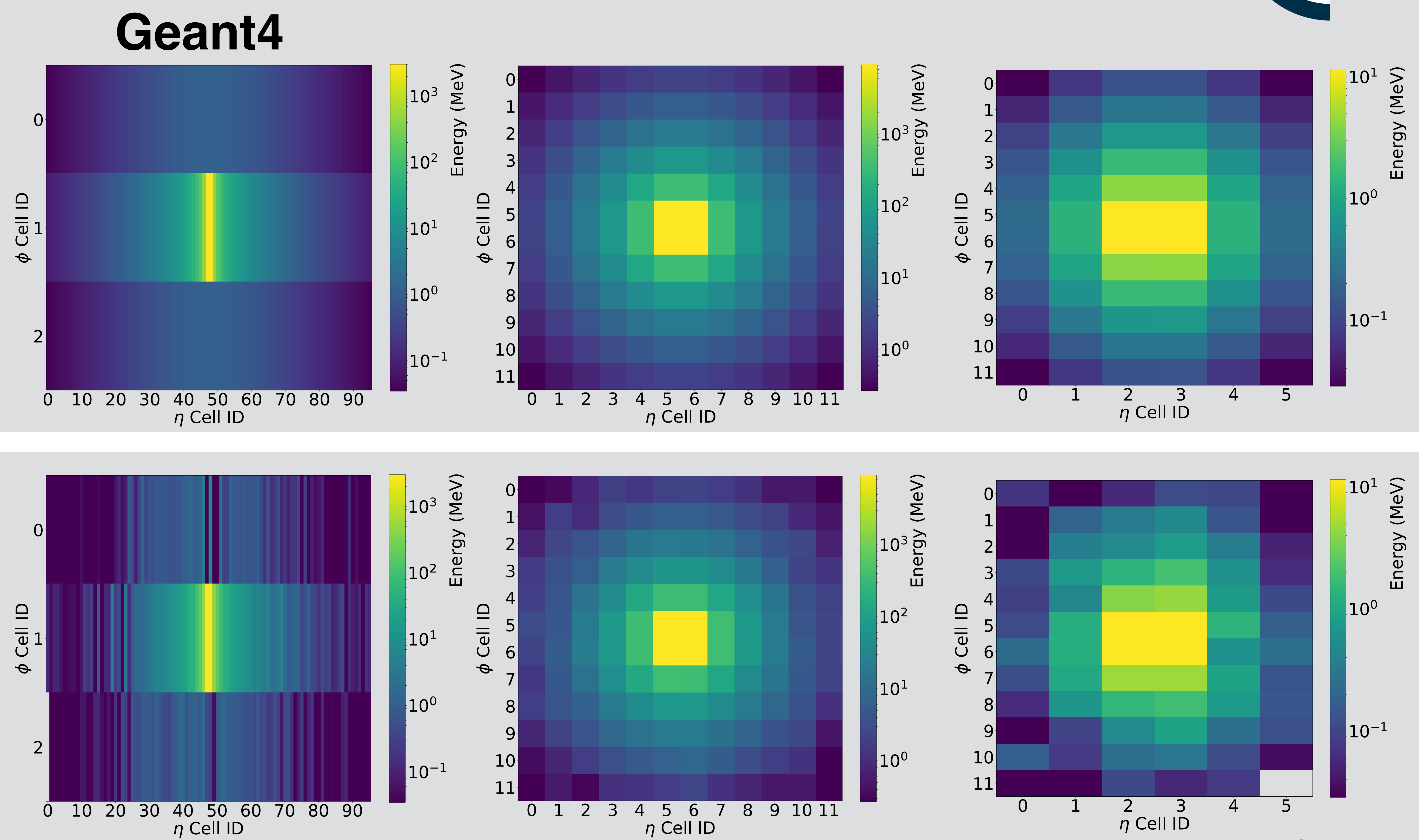


When **D** is maximally confused, **G** will be a good generator



[I. Goodfellow et al., NIPS 2014]

Physics-inspired: CaloGAN



M. Paganini, L. de Oliveira, **BPN**, PRL 120 (2018) 042003

CaloGAN Timing

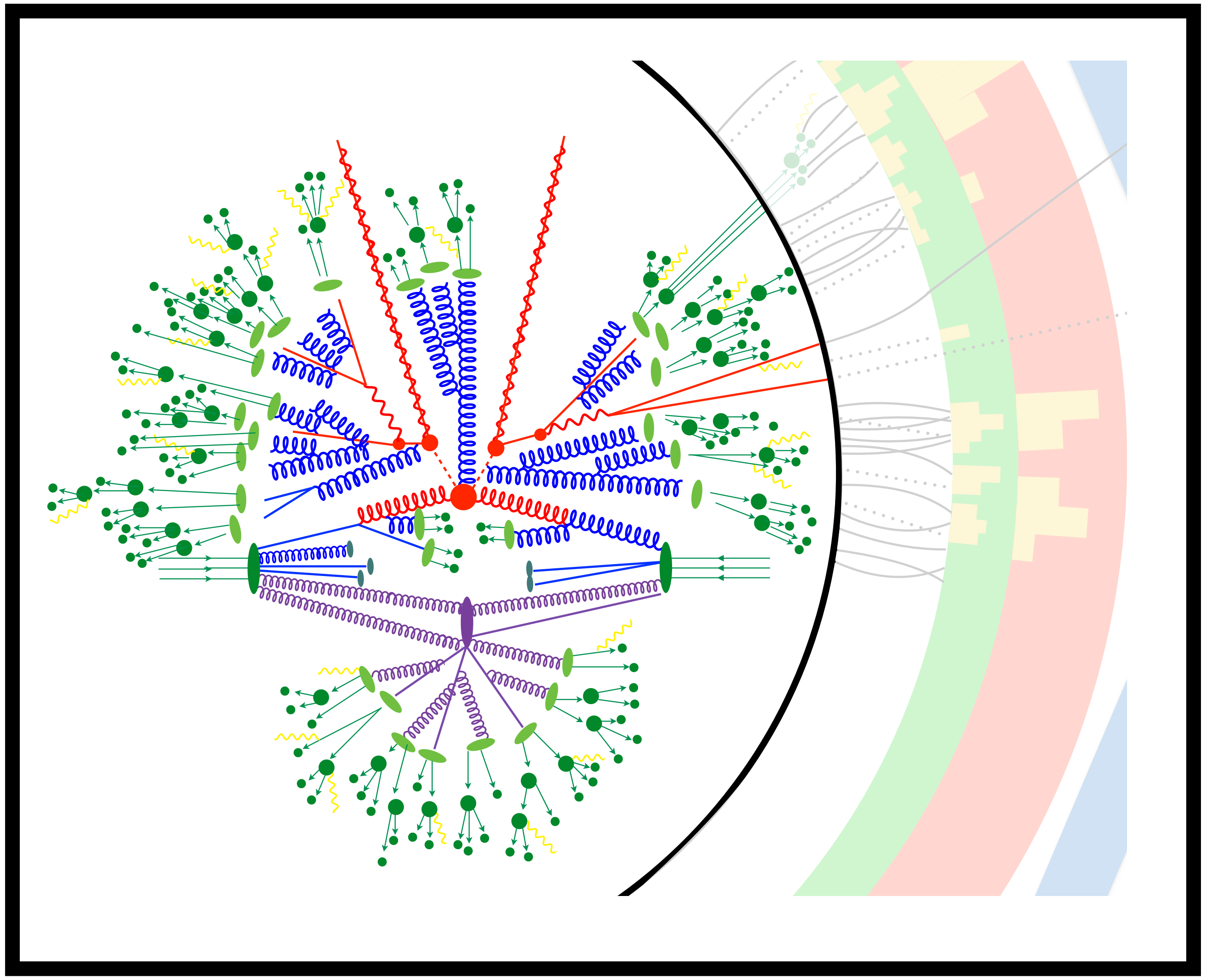
M. Paganini, L. de Oliveira, **BPN**, PRL 120 (2018) 042003

Generation Method	Hardware	Batch Size	milliseconds/shower
GEANT4	CPU	N/A	1772 ←
CALOGAN	CPU	1	13.1
		10	5.11
		128	2.19
		1024	2.03
	GPU	1	14.5
		4	3.68
		128	0.021
		512	0.014
		1024	0.012 ←

(clearly these numbers will change as both technologies improve - this is simply meant to be qualitative and motivating!)

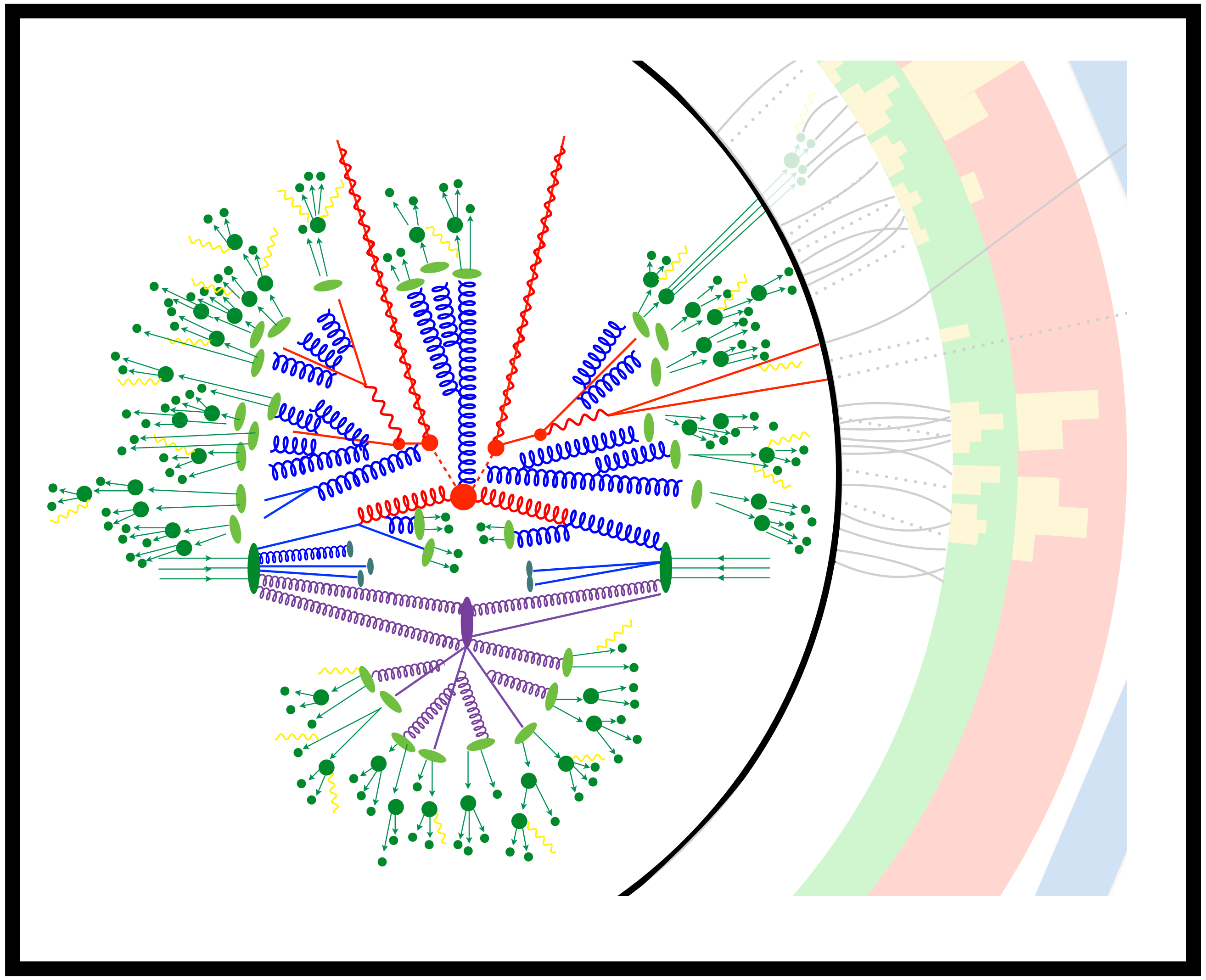
Generative Models

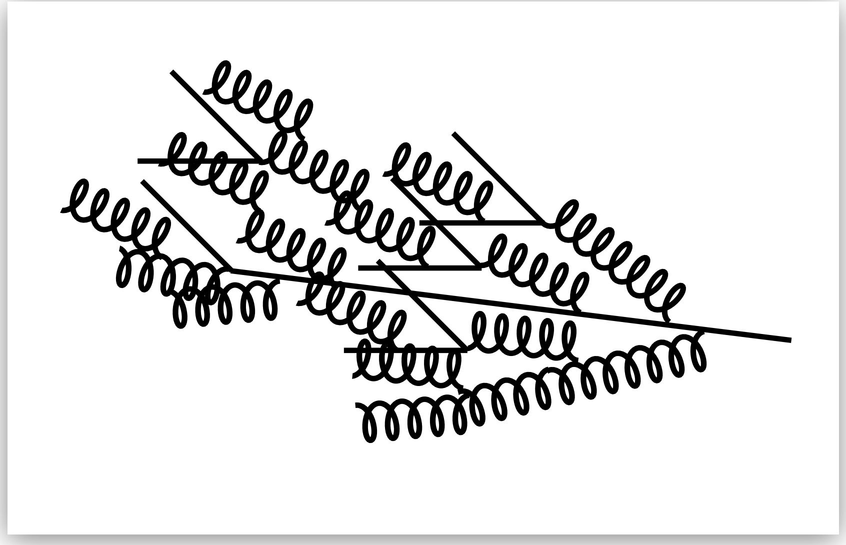
What if the fidelity of the simulation is not good enough?



Generative Models

What if the fidelity of the simulation is not good enough?

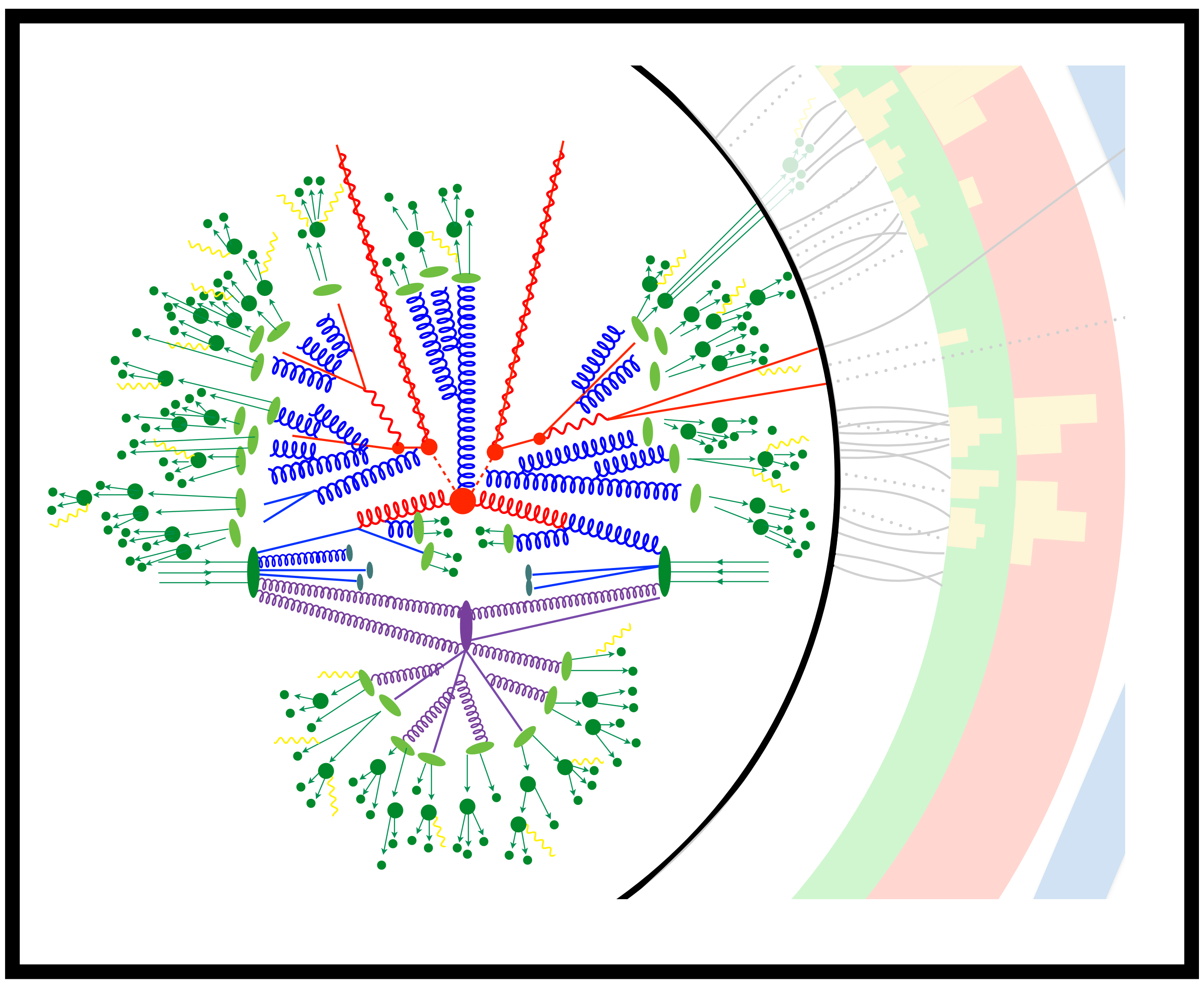


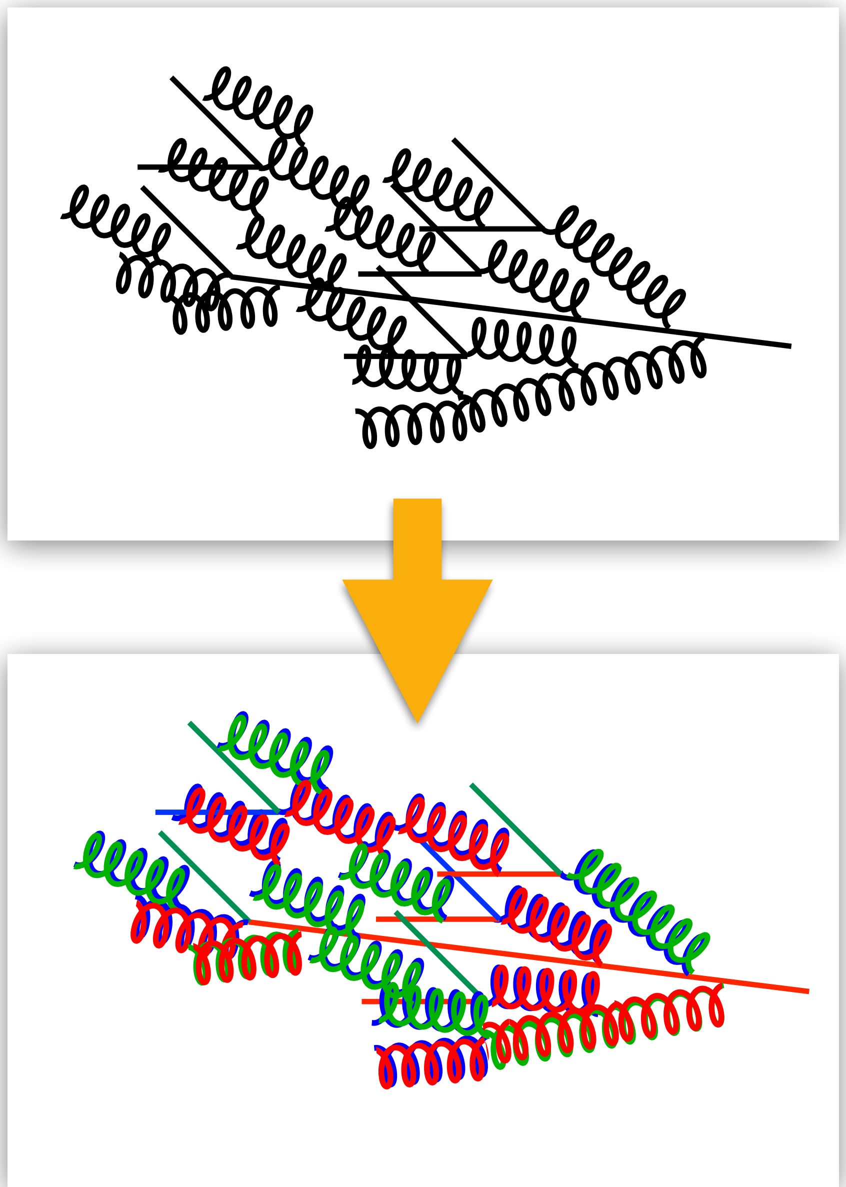


Generative Models

What if the fidelity of the simulation is not good enough?

Quantum simulations with quantum computers?





Analog versus Digital Quantum Circuits

Goal: implement our system's Hamiltonian (e.g. the SM) in a proxy system ("quantum computer") and let it evolve.

The best quantum computer is the one that looks just like the system you are trying to model!

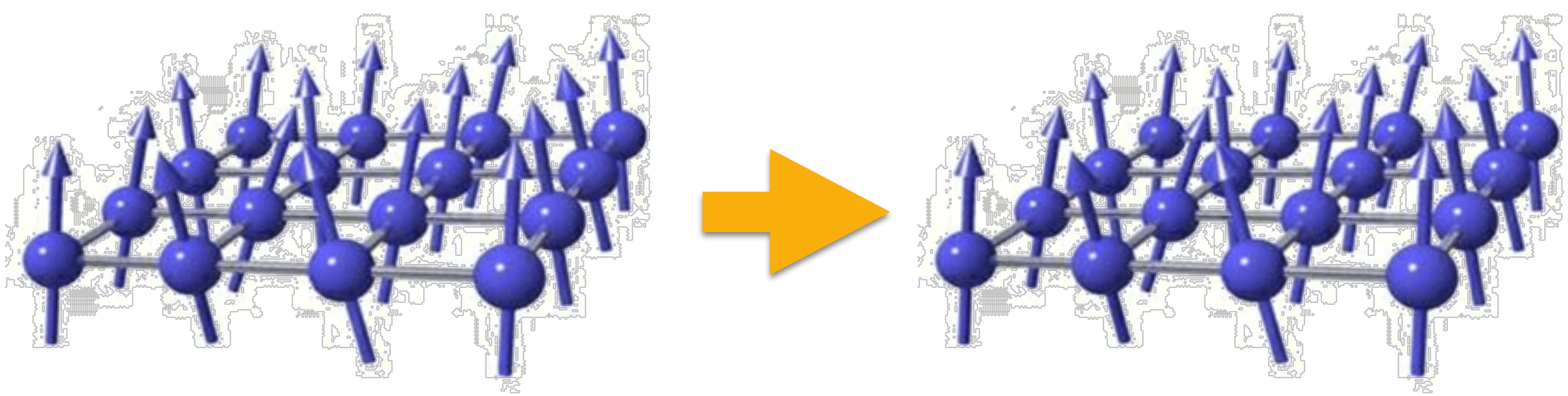


Image credit: https://www3.physik.uni-stuttgart.de/TR21/en/about/research.php

Analog versus Digital Quantum Circuits

Goal: implement our system's Hamiltonian (e.g. the SM) in a proxy system ("quantum computer") and let it evolve.

The best quantum computer is the one that looks just like the system you are trying to model!

Image credit: CERN

Analog versus Digital **Quantum Circuits**

Goal: implement our system's Hamiltonian (e.g. the SM) in a proxy system ("quantum computer") and let it evolve.

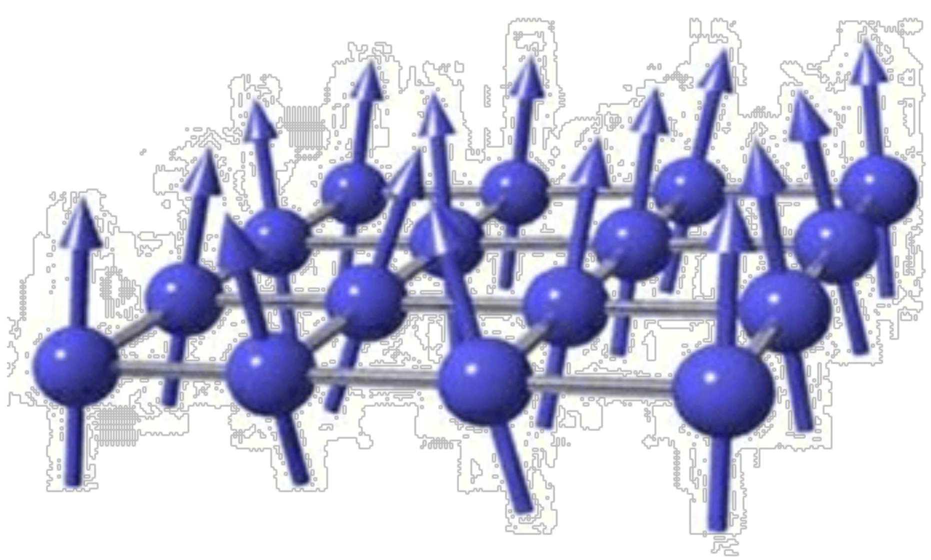
The best quantum computer is the one that looks just like the system you are trying to model!

Image credit: CERN

Analog versus **Digital** Quantum Circuits

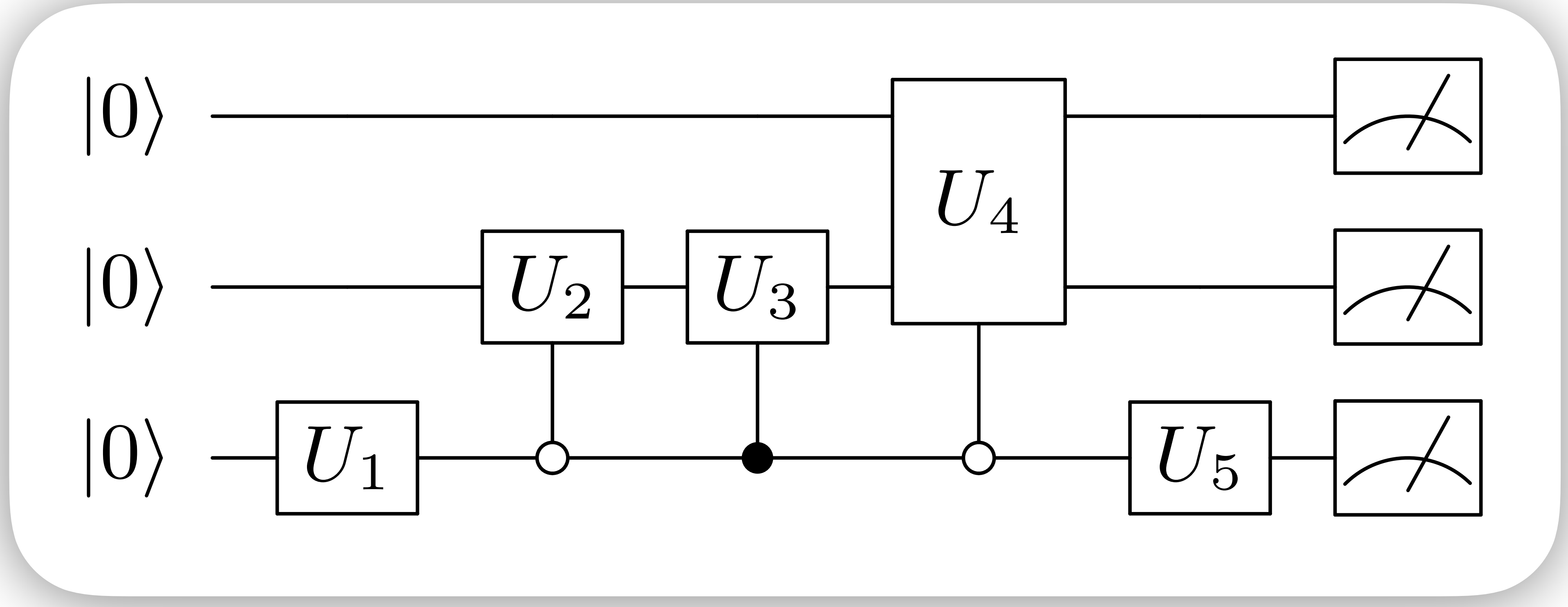
Goal: implement our system's Hamiltonian (e.g. the SM) in a proxy system ("quantum computer") and let it evolve.

In this setup, the possibilities are endless; the key is efficiency.



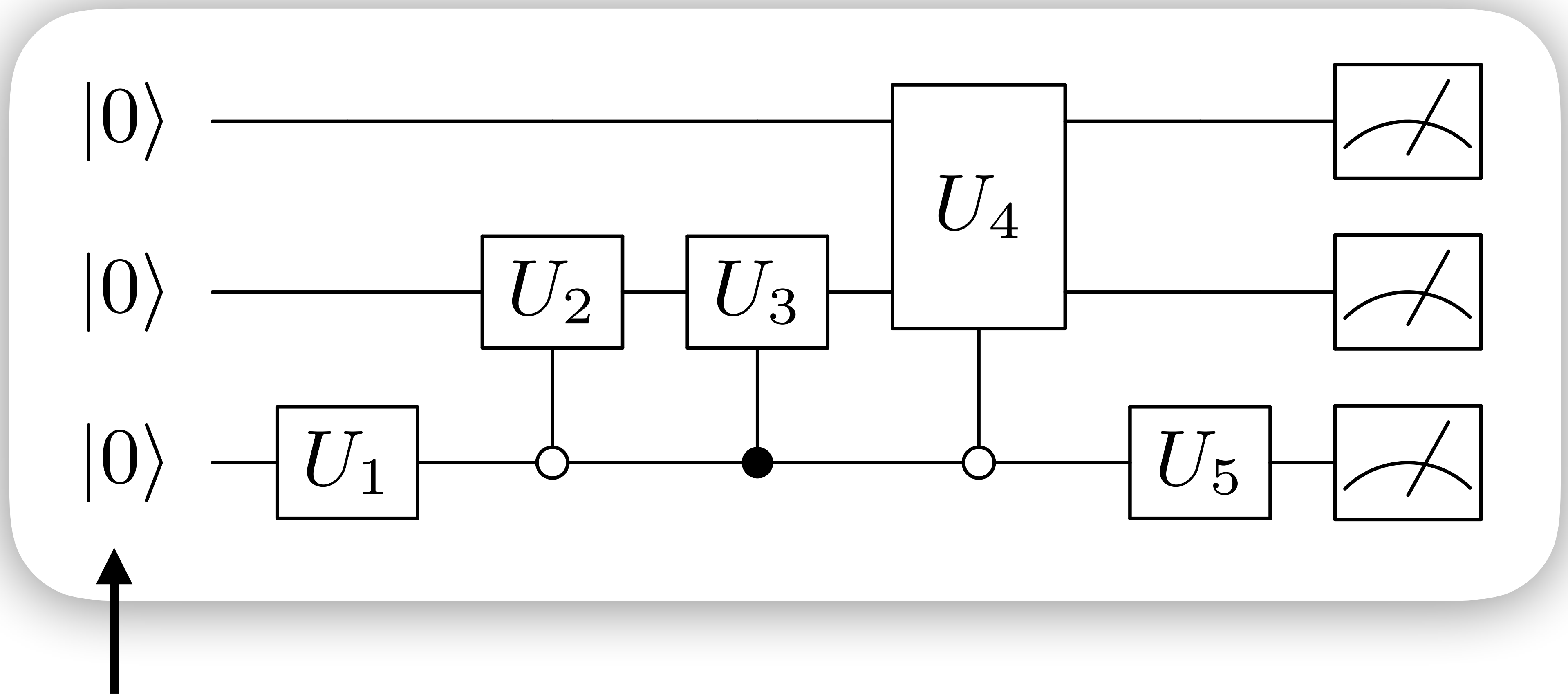
Programming

Just like a classical computer, one can write programs for a universal quantum computer.



Programming

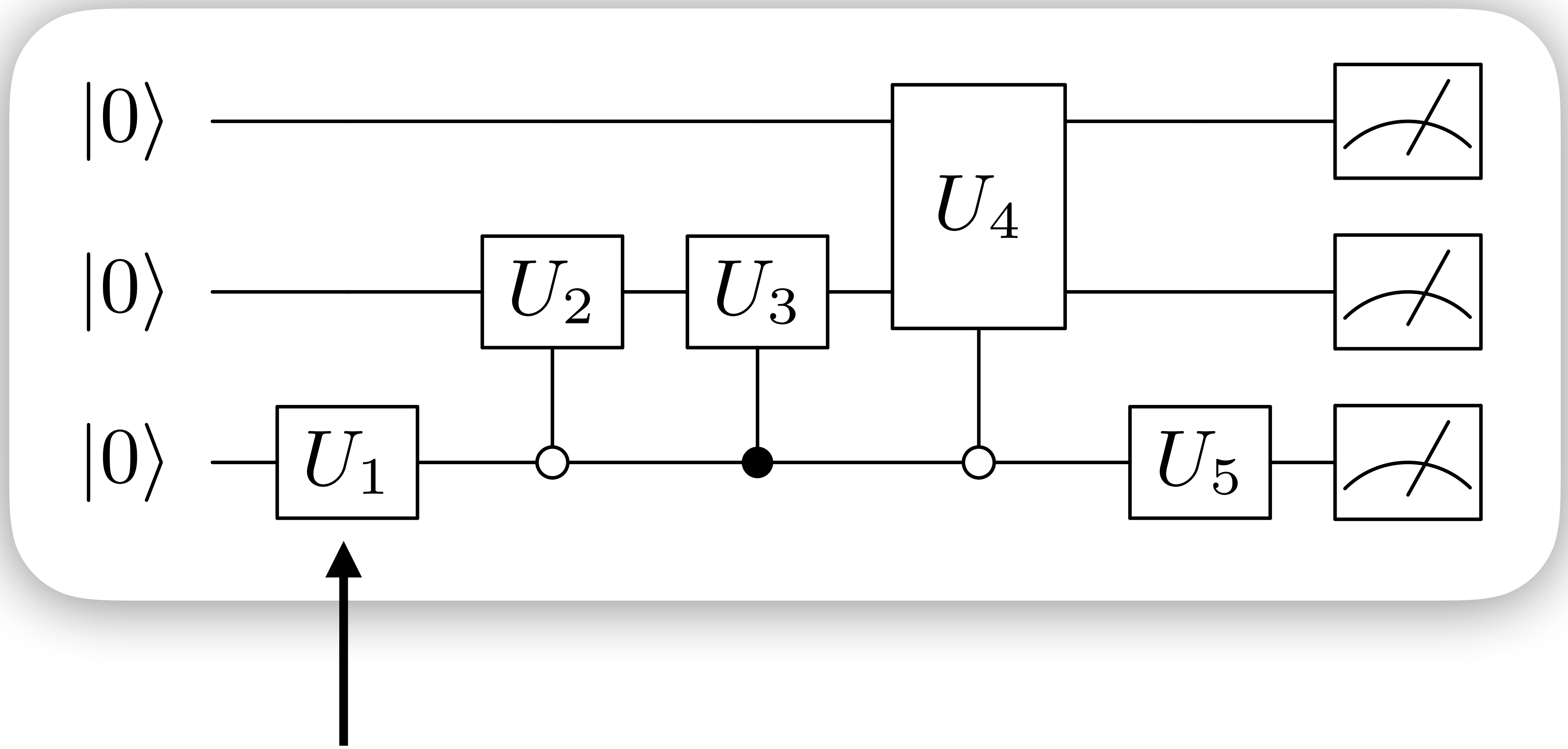
Just like a classical computer, one can write programs for a universal quantum computer.



Initialize in the ground state.

Programming

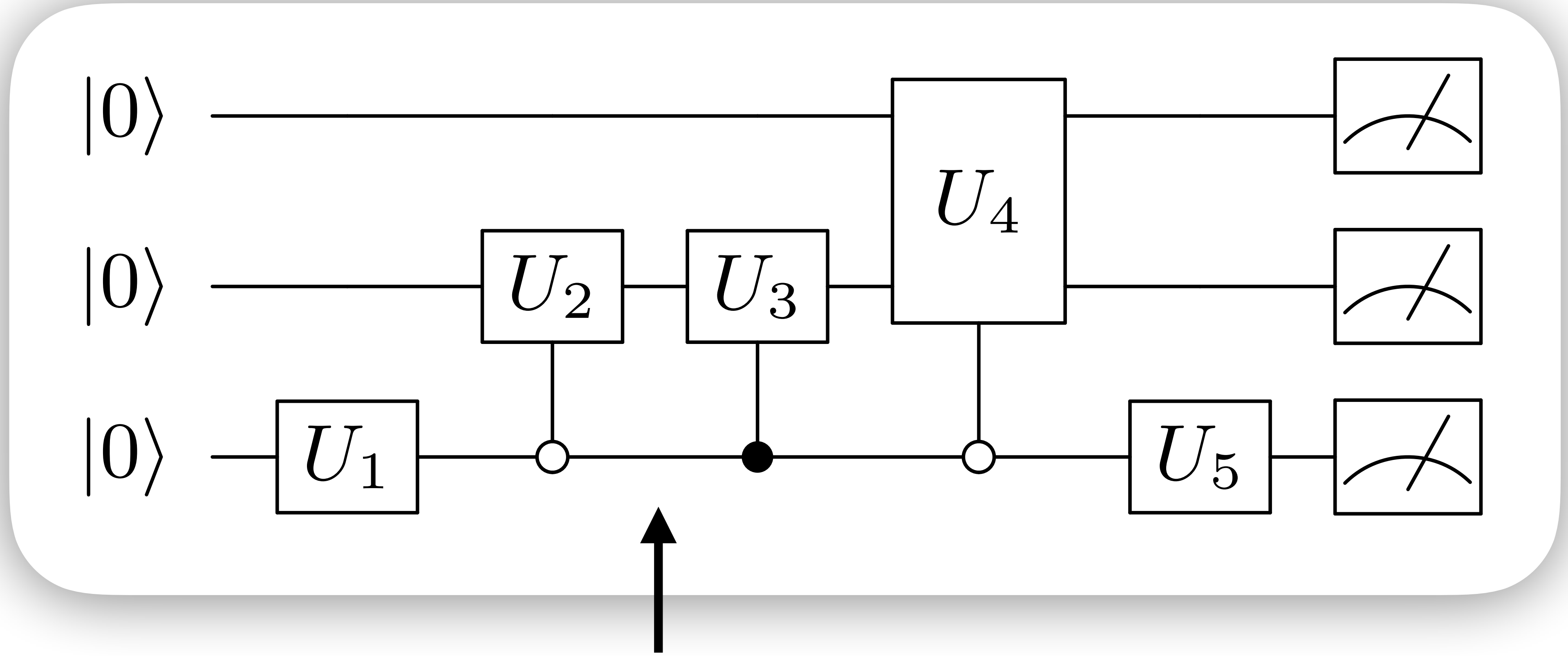
Just like a classical computer, one can write programs for a universal quantum computer.



Apply unitary matrix
U₁ to the third qubit

Programming

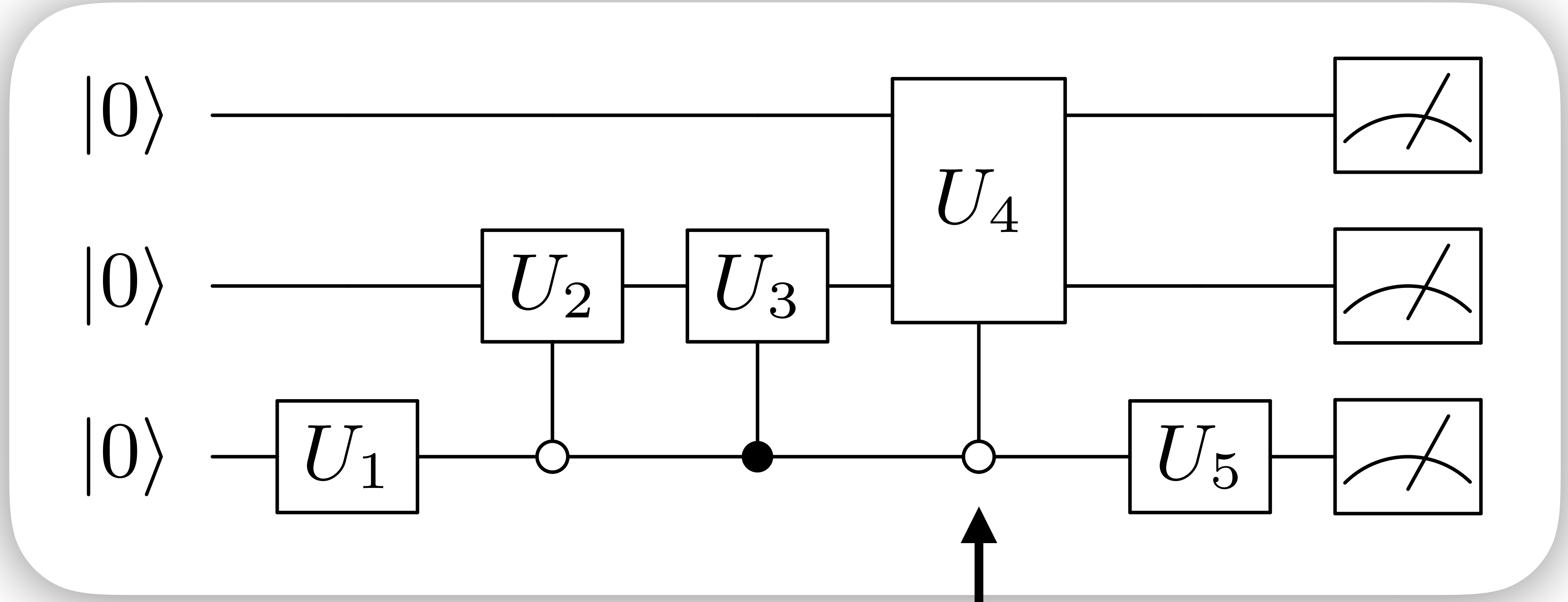
Just like a classical computer, one can write programs for a universal quantum computer.



Apply unitary matrix U_2 to the second qubit when the third is 0, else apply U_3.

Programming

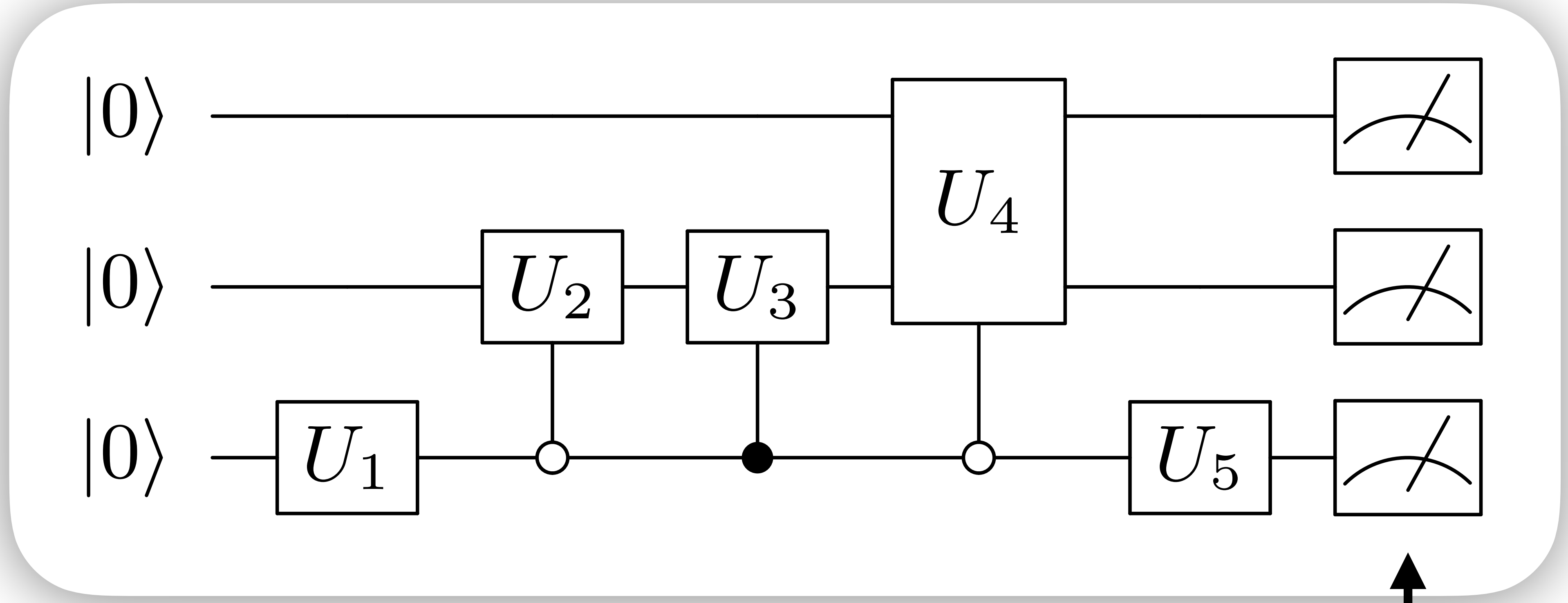
Just like a classical computer, one can write programs for a universal quantum computer.



Apply unitary matrix U_4 to both the first and second quits when the third is 0.

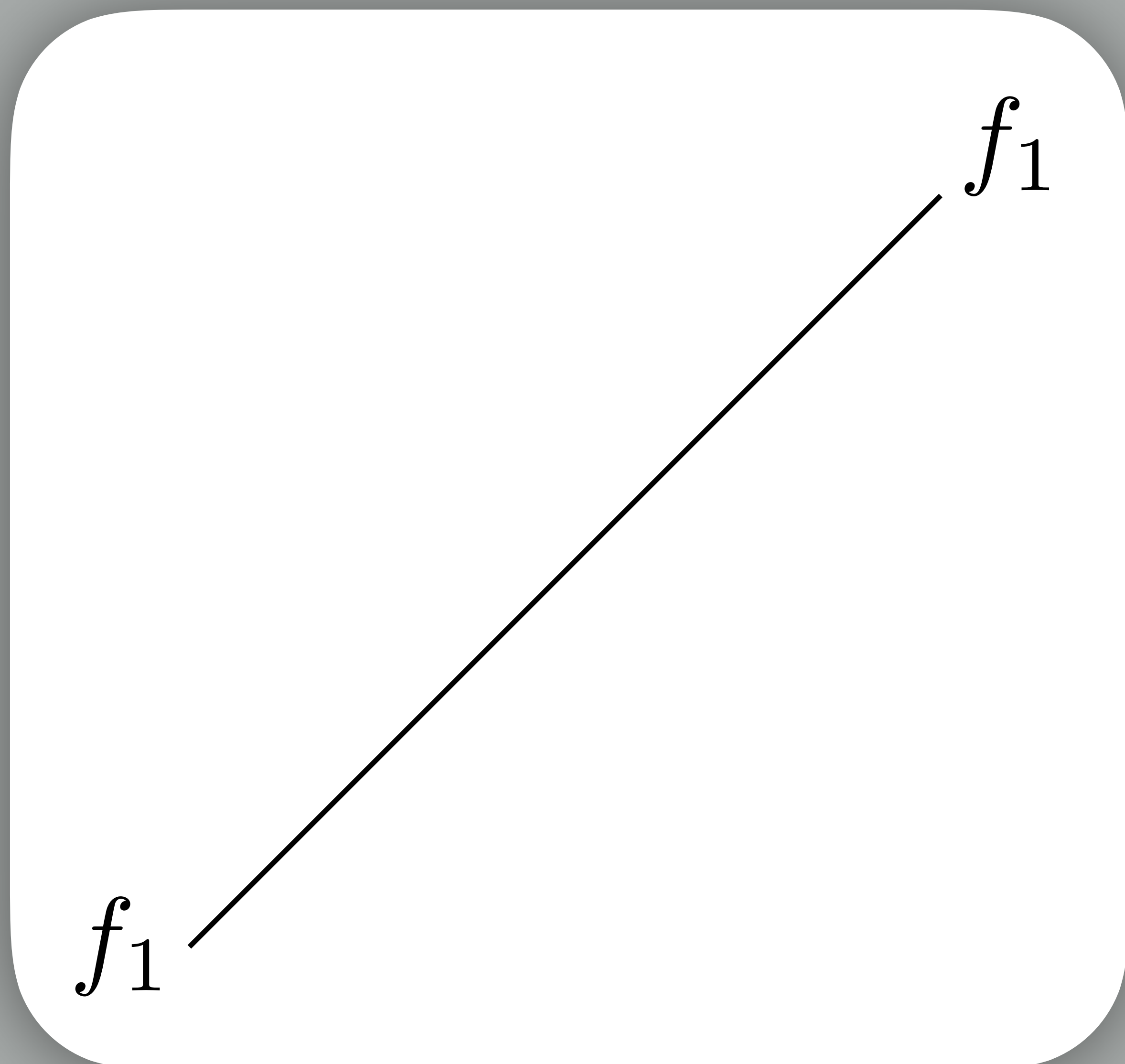
Programming

Just like a classical computer, one can write programs for a universal quantum computer.



Measure all the qubits

A quantum algorithm for radiation

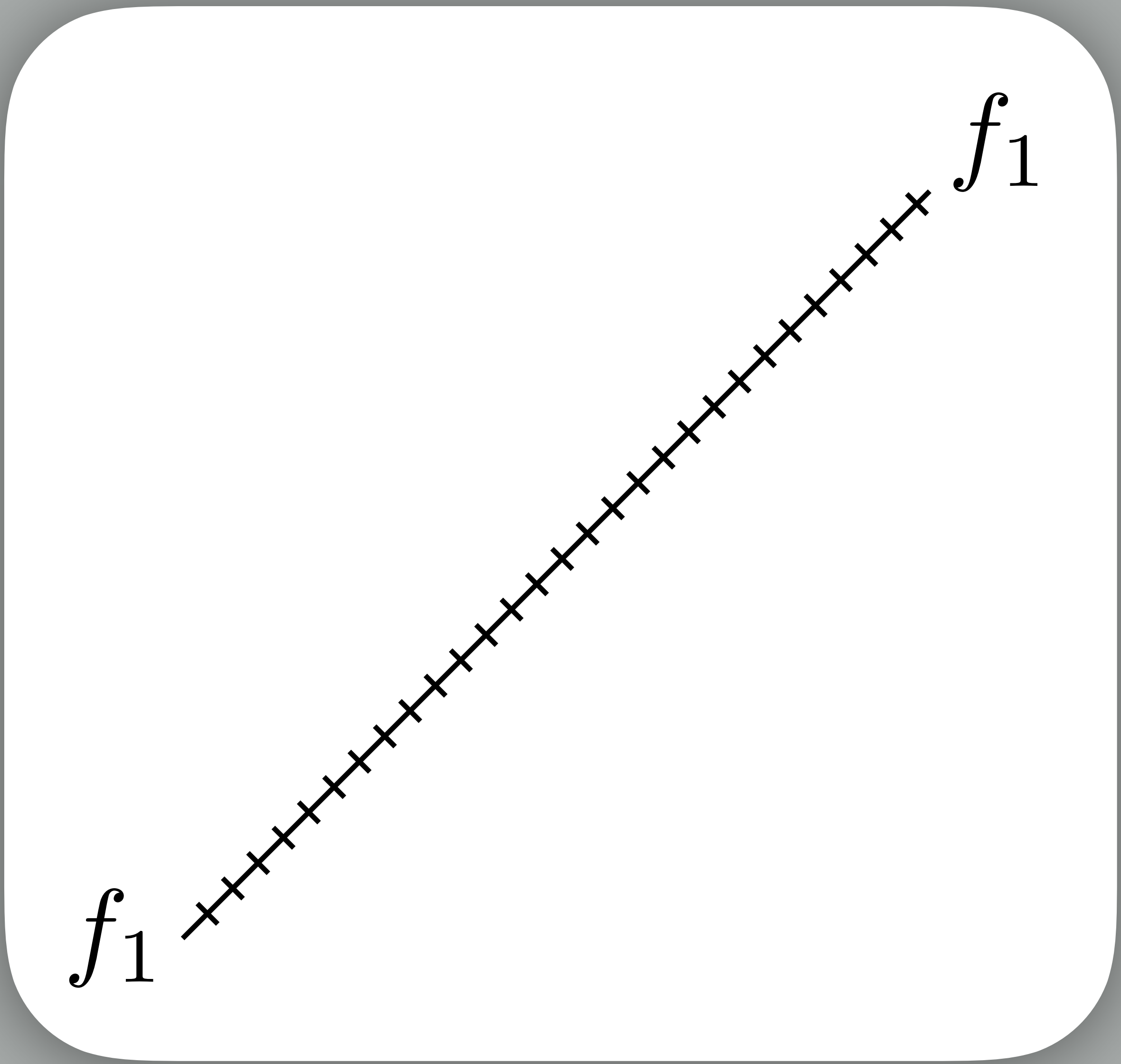


Simplified theory: two fermions f_1 and f_2.

They can radiate scalars and can mix.

BPN, D. Provasoli, W. de Jong, C. Bauer, 1904.03196

A quantum algorithm for radiation

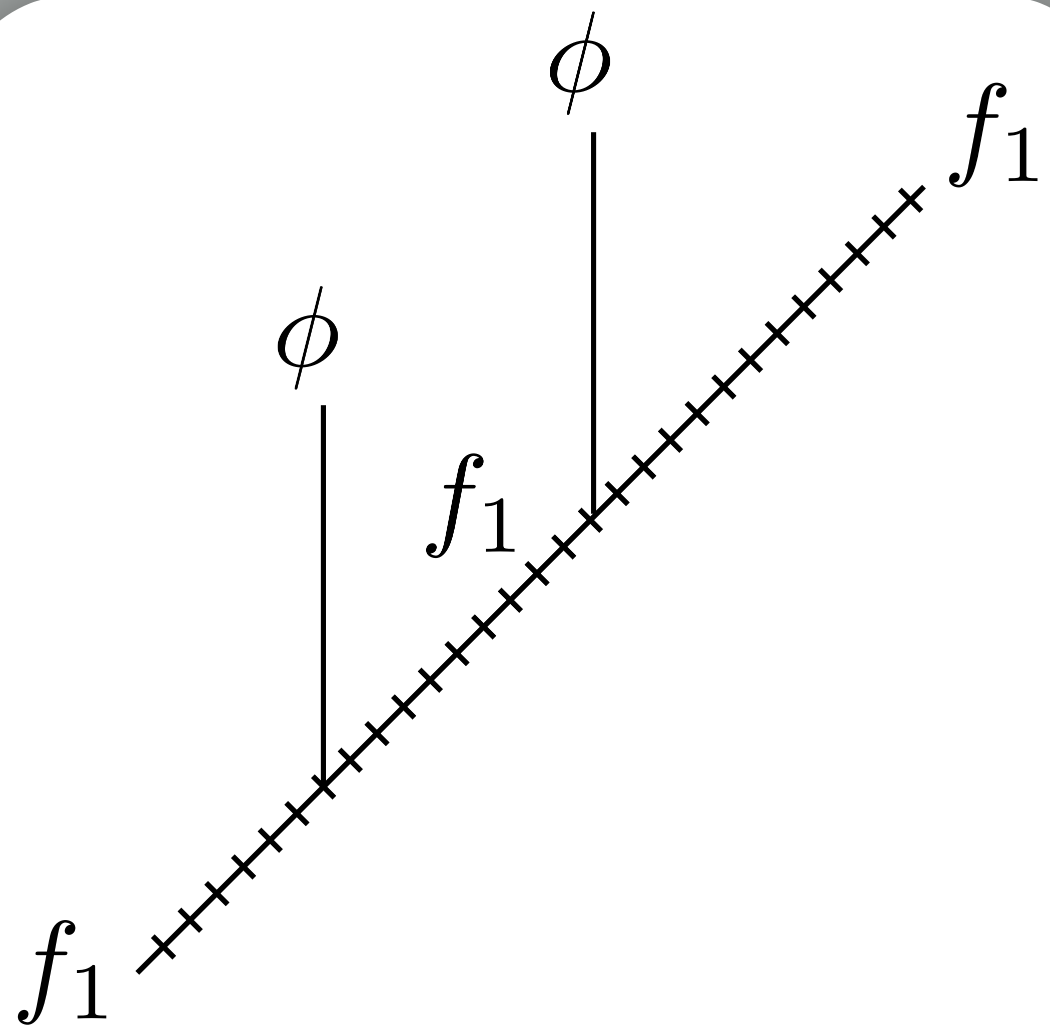


Algorithm: discretize the phase space

Check for emissions at each step.

BPN, D. Provasoli, W. de Jong, C. Bauer, 1904.03196

A quantum algorithm for radiation

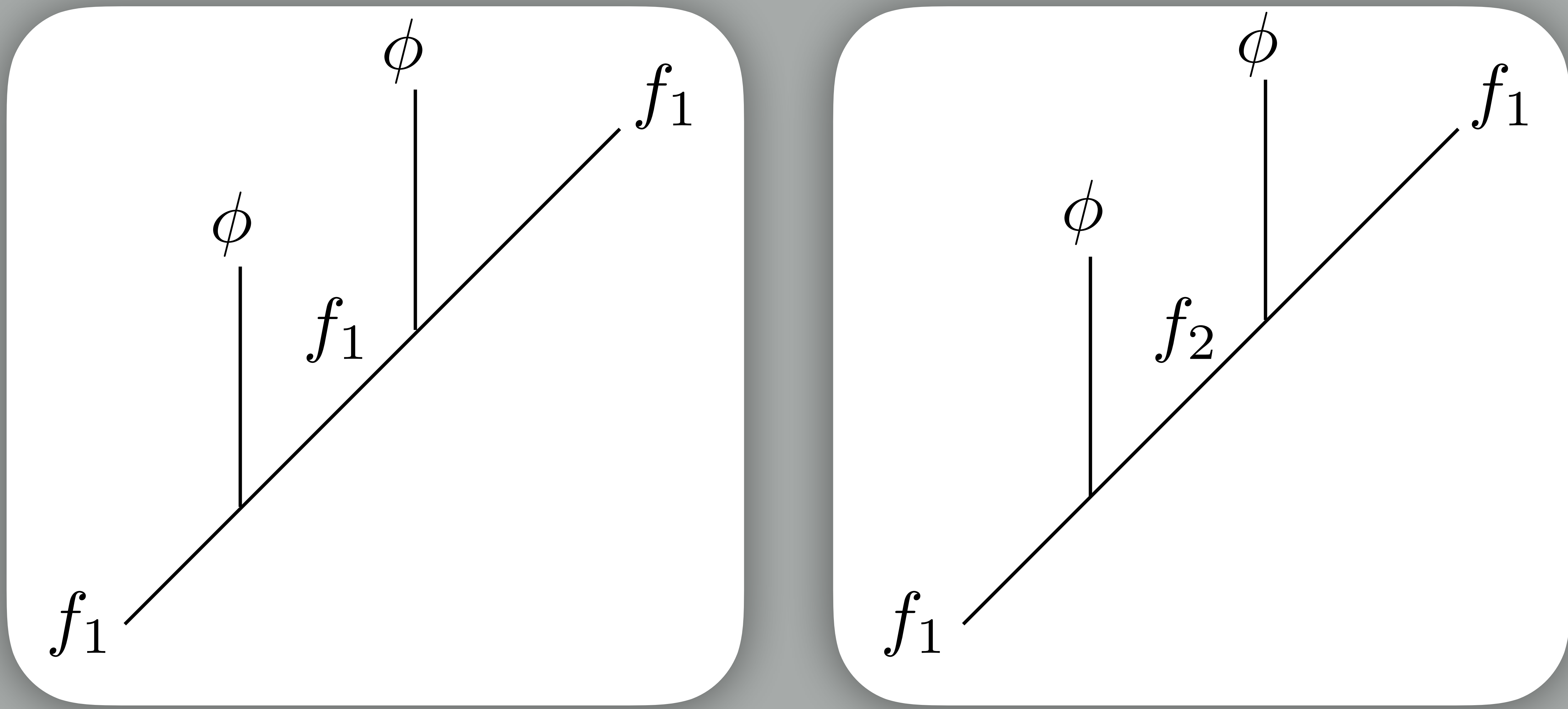


Algorithm: discretize the phase space

Check for emissions at each step.

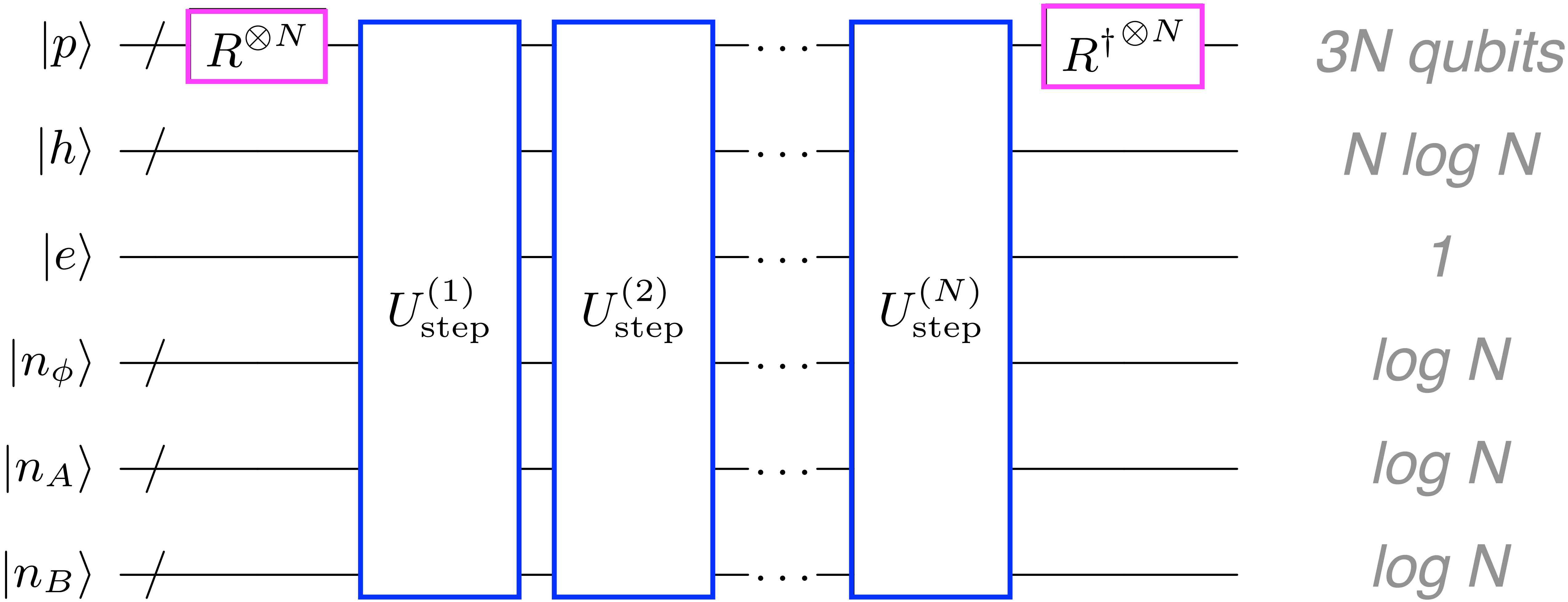
A quantum algorithm for radiation

BPN, D. Provasoli, W. de Jong, C. Bauer, 1904.03196



Multiple routes to the same state - interference!

A quantum algorithm for radiation

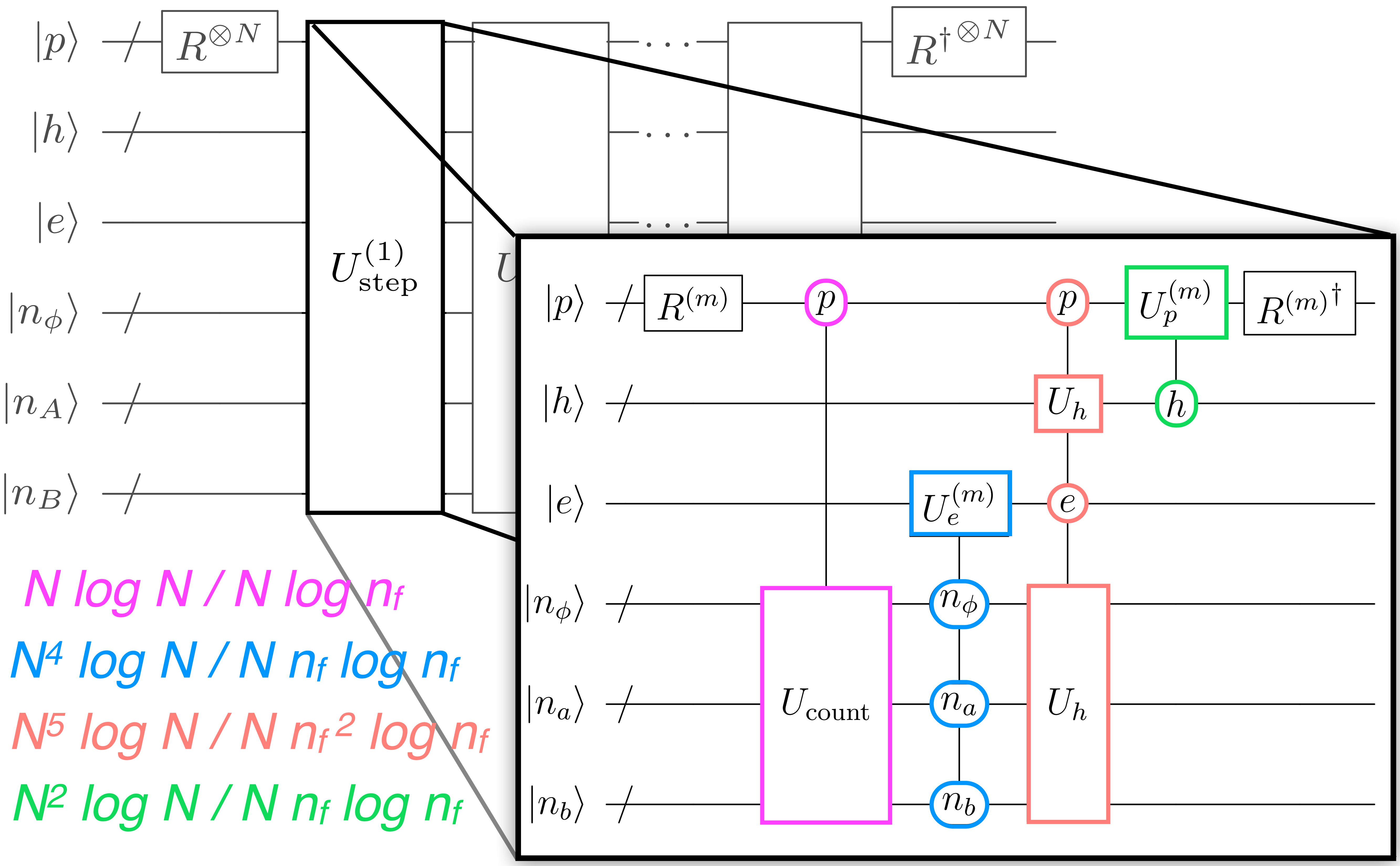


Discretize the shower evolution (e.g. in angle)

& **rotate** to a particle basis without interference.

BPN, D. Provasoli, W. de Jong, C. Bauer, 1904.03196

A quantum algorithm for radiation



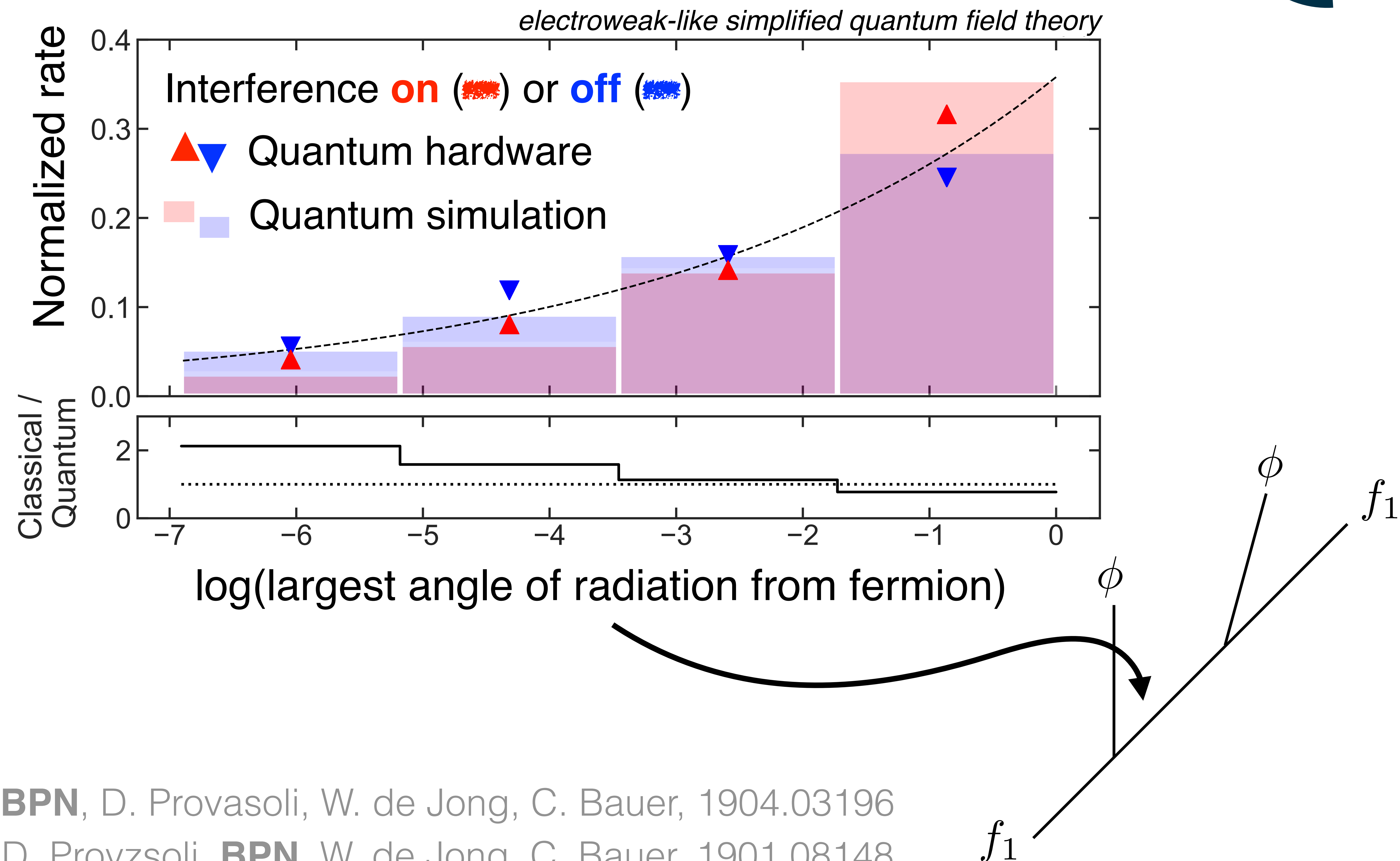
$N \log N$ / $N \log n_f$

$N^4 \log N$ / $N\, n_f \log n_f$

$N^5 \log N$ / $N\, n_f^{\,2} \log n_f$

$N^2 \log N$ / $N\, n_f \log n_f$

A quantum algorithm for radiation



BPN, D. Provasoli, W. de Jong, C. Bauer, 1904.03196
D. Provzsoli, **BPN**, W. de Jong, C. Bauer, 1901.08148

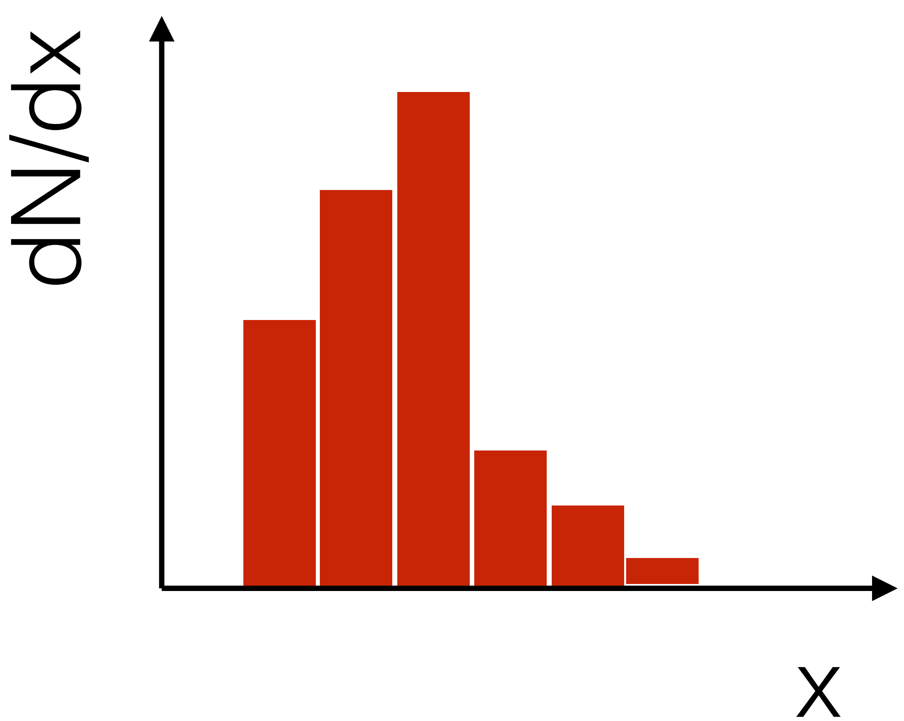
From HEP to QIS

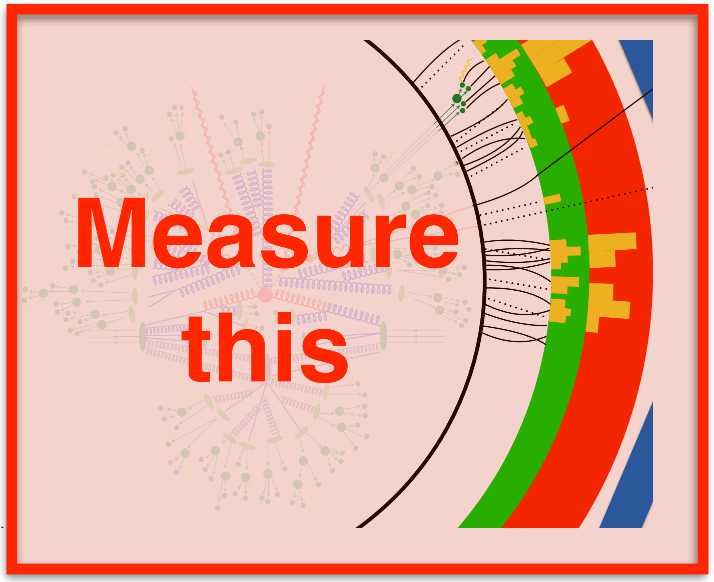
In addition to basic algorithms dev. for QFT, I have discovered new ways to improve quantum computing **in general**

From HEP to QIS

In addition to basic algorithms dev. for QFT, I have discovered new ways to improve quantum computing **in general**

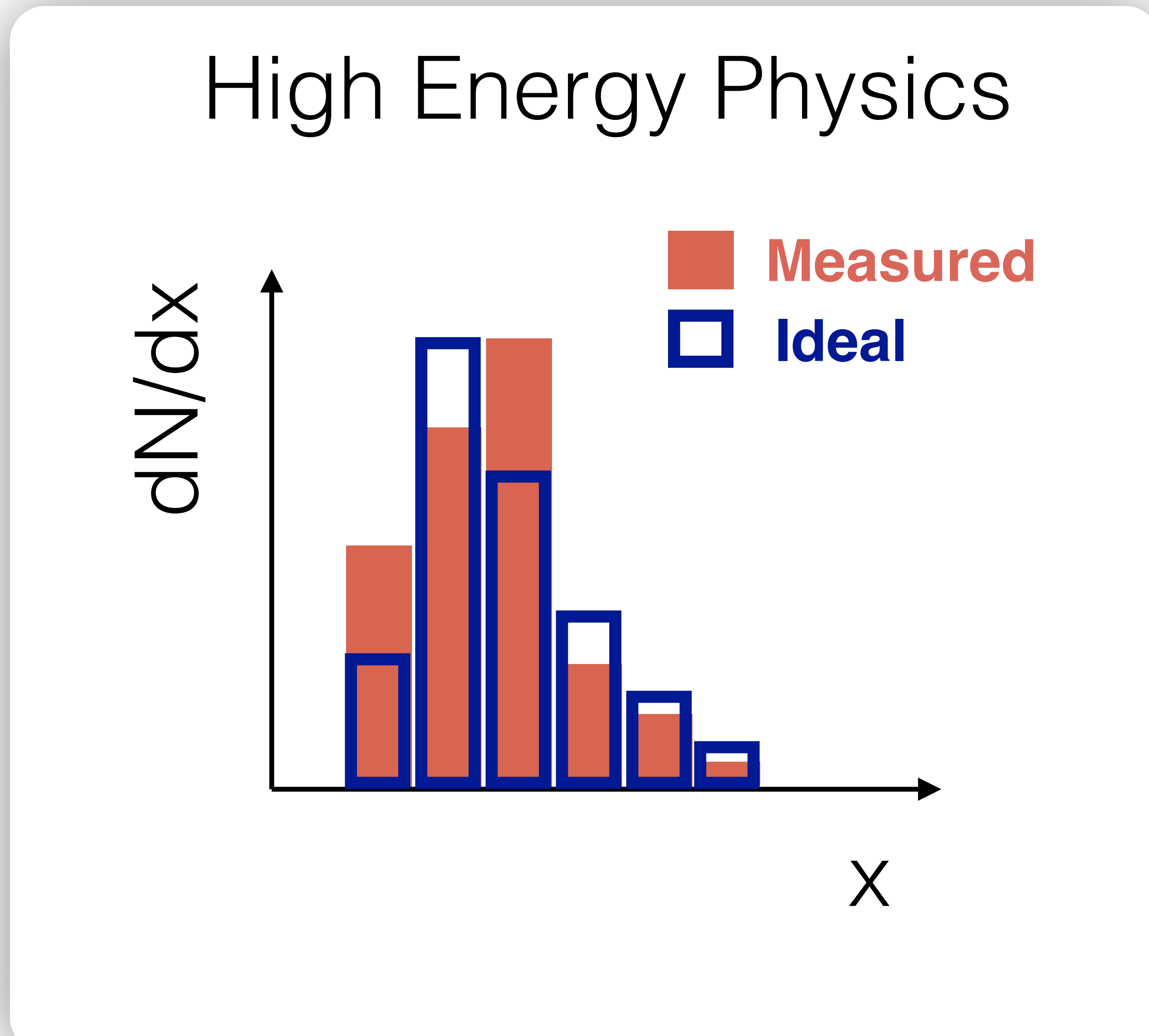
High Energy Physics

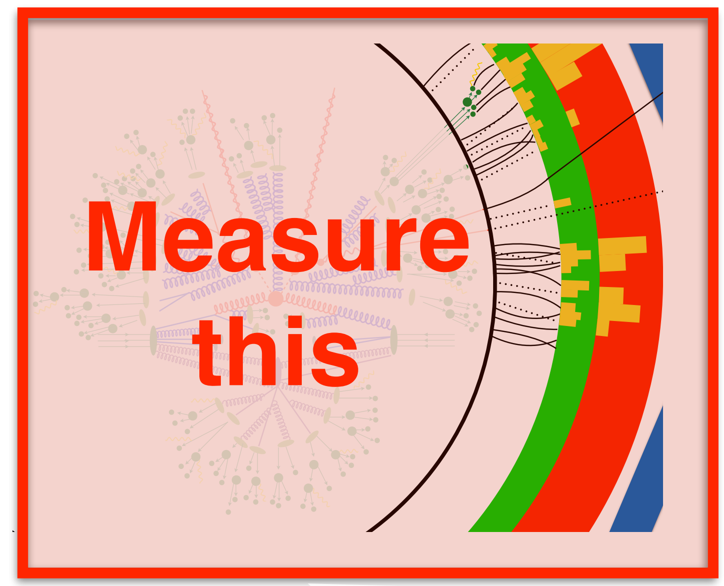


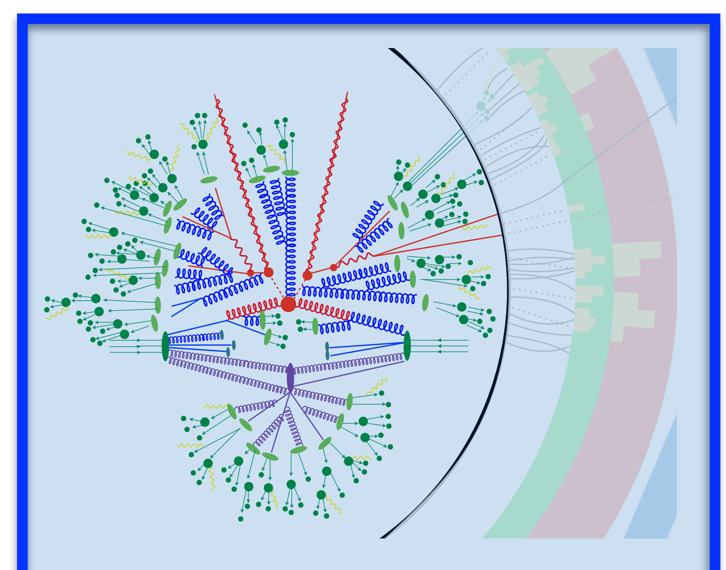


From HEP to QIS

In addition to basic algorithms dev. for QFT, I have discovered new ways to improve quantum computing **in general**



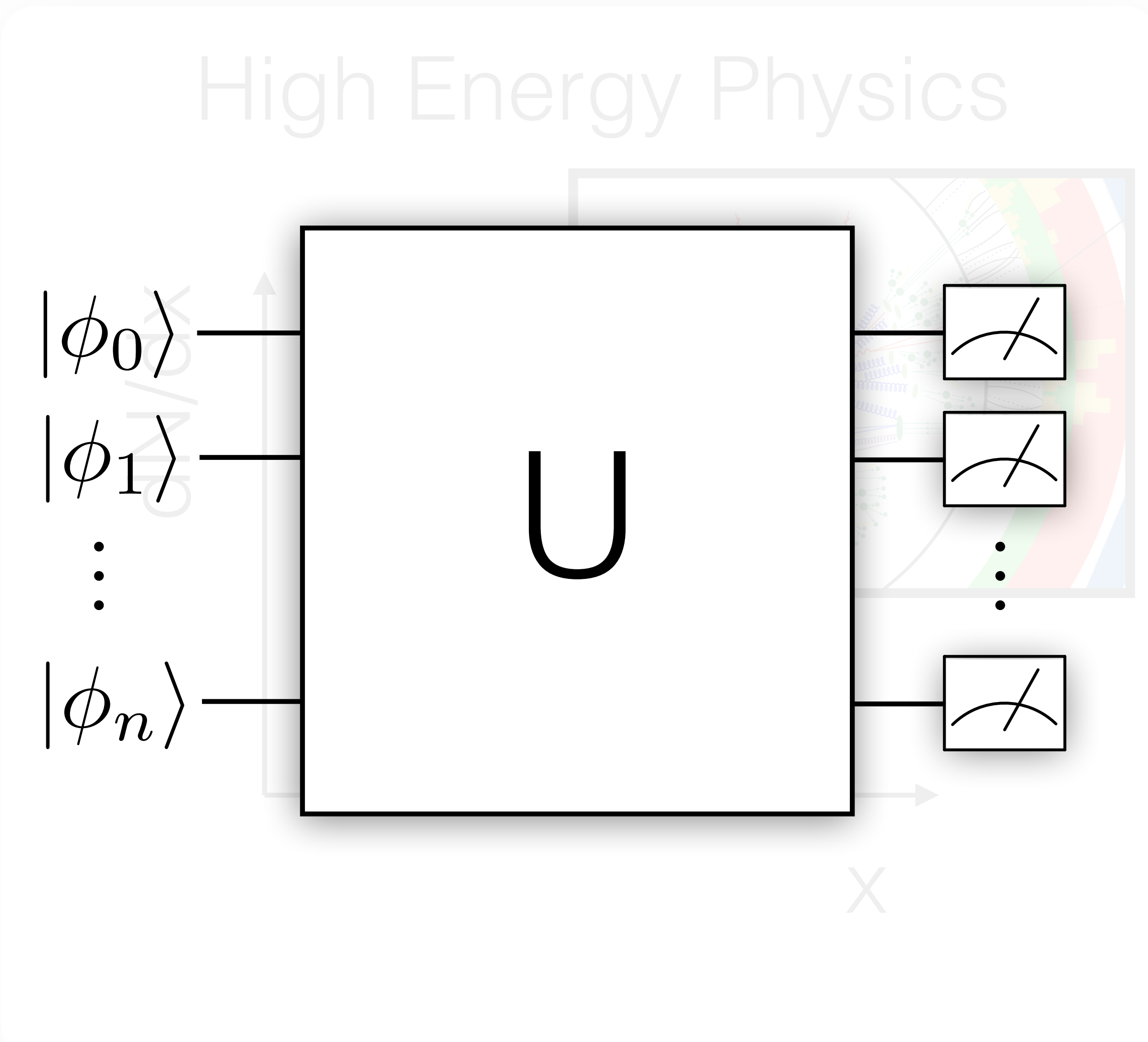


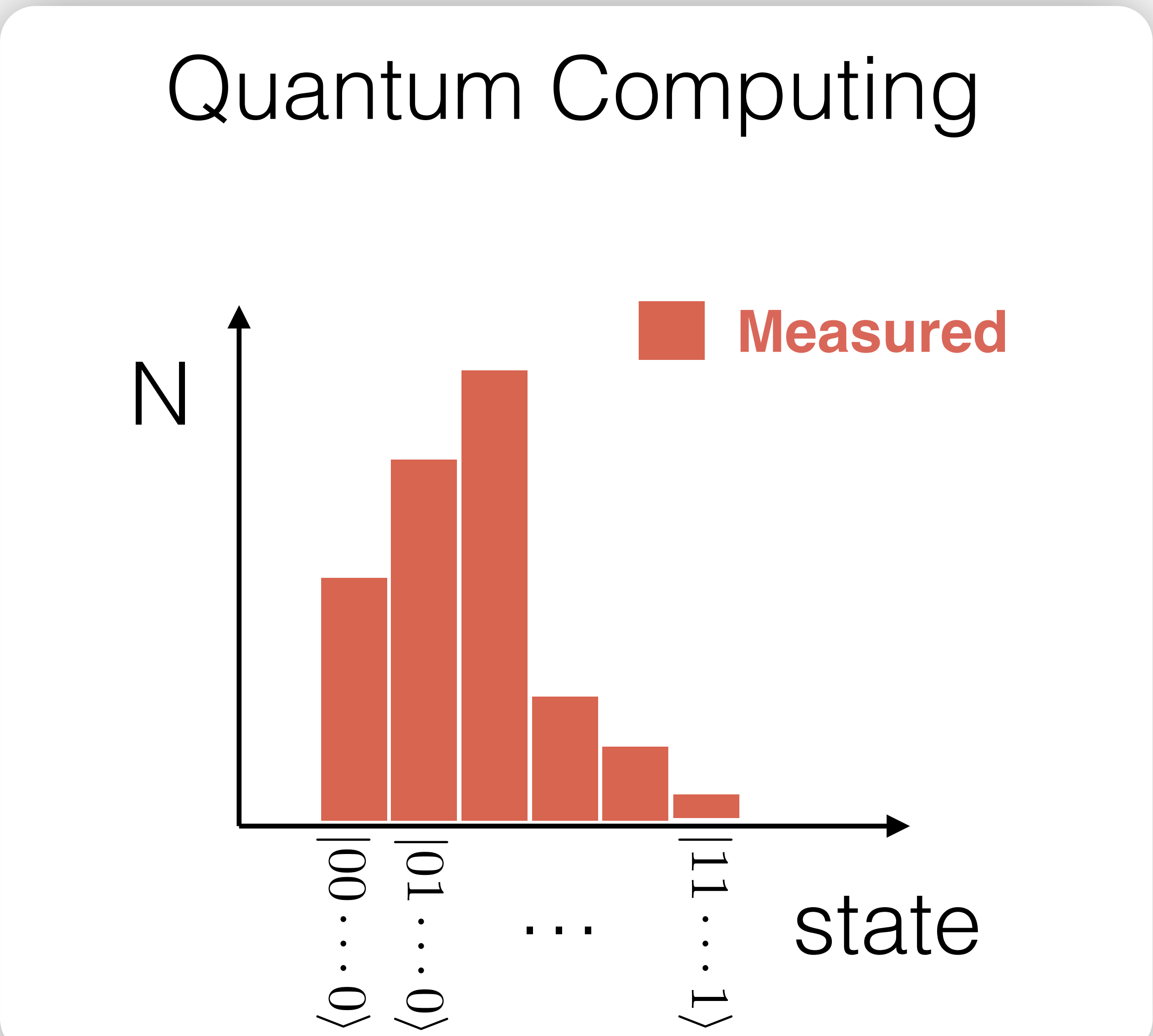


Want this

From HEP to QIS

In addition to basic algorithms dev. for QFT, I have discovered new ways to improve quantum computing **in general**

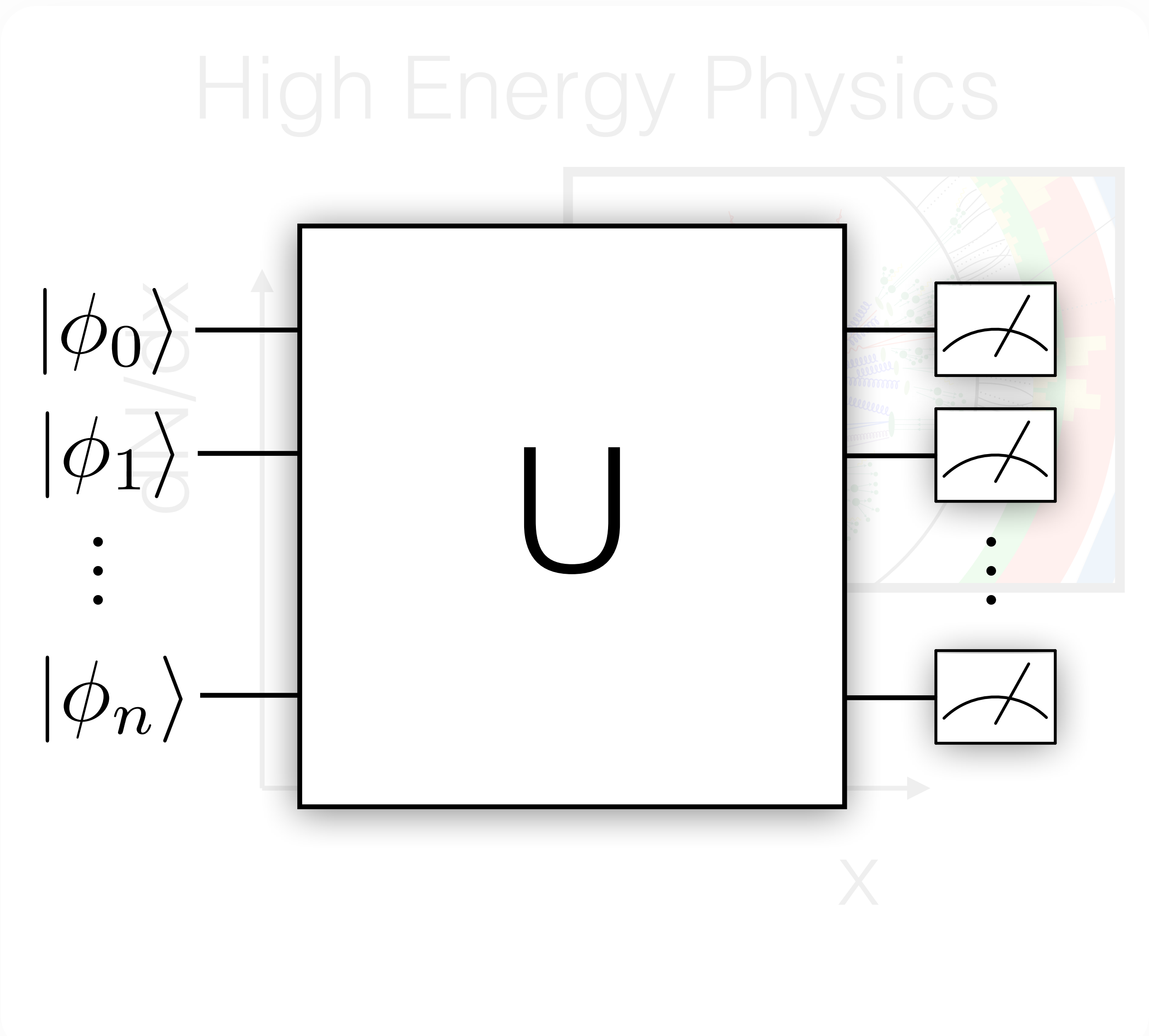


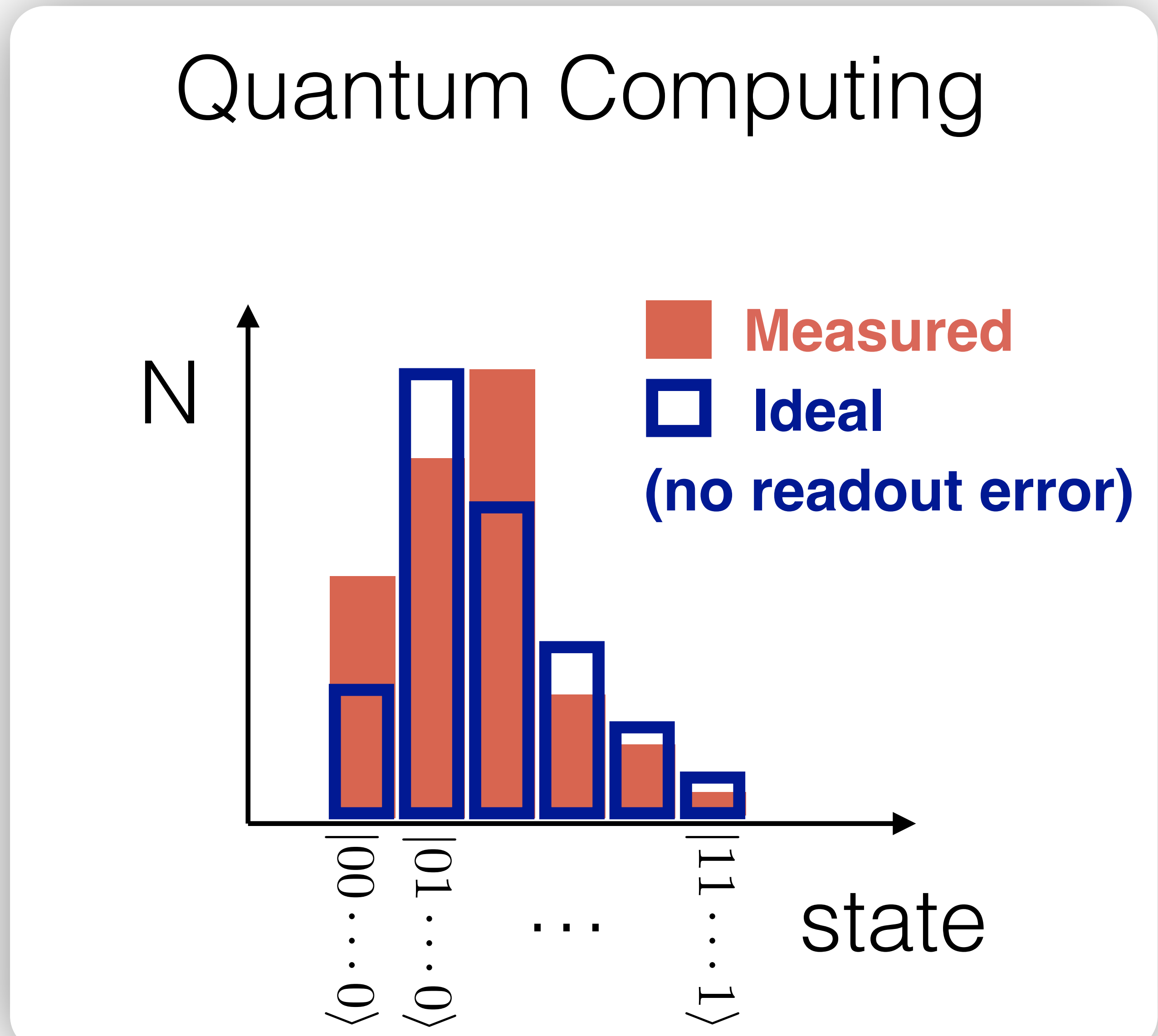


BPN, M. Urbanek, W. de Jong, C. Bauer, 1910.01969, A. He, **BPN**, W. de Jong, C. Bauer, 2003.nextweek

From HEP to QIS

In addition to basic algorithms dev. for QFT, I have discovered new ways to improve quantum computing **in general**



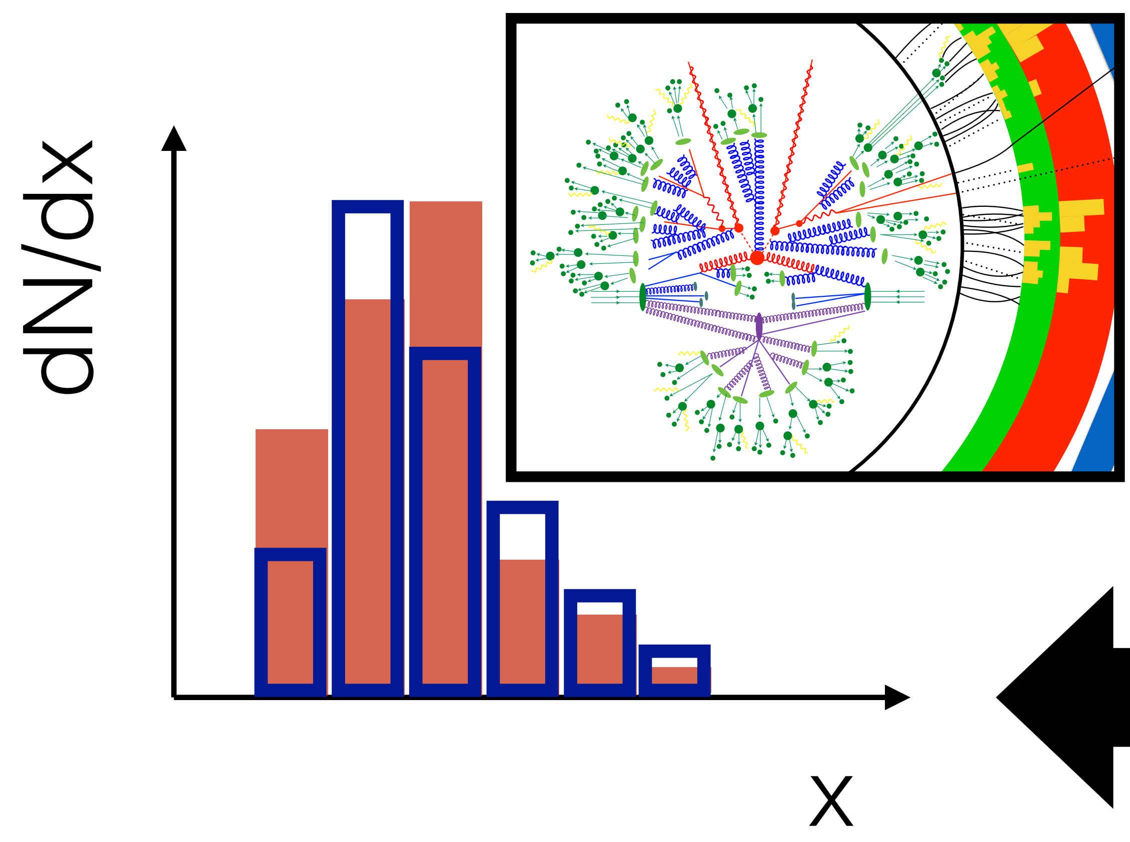


BPN, M. Urbanek, W. de Jong, C. Bauer, 1910.01969, A. He, **BPN**, W. de Jong, C. Bauer, 2003.nextweek

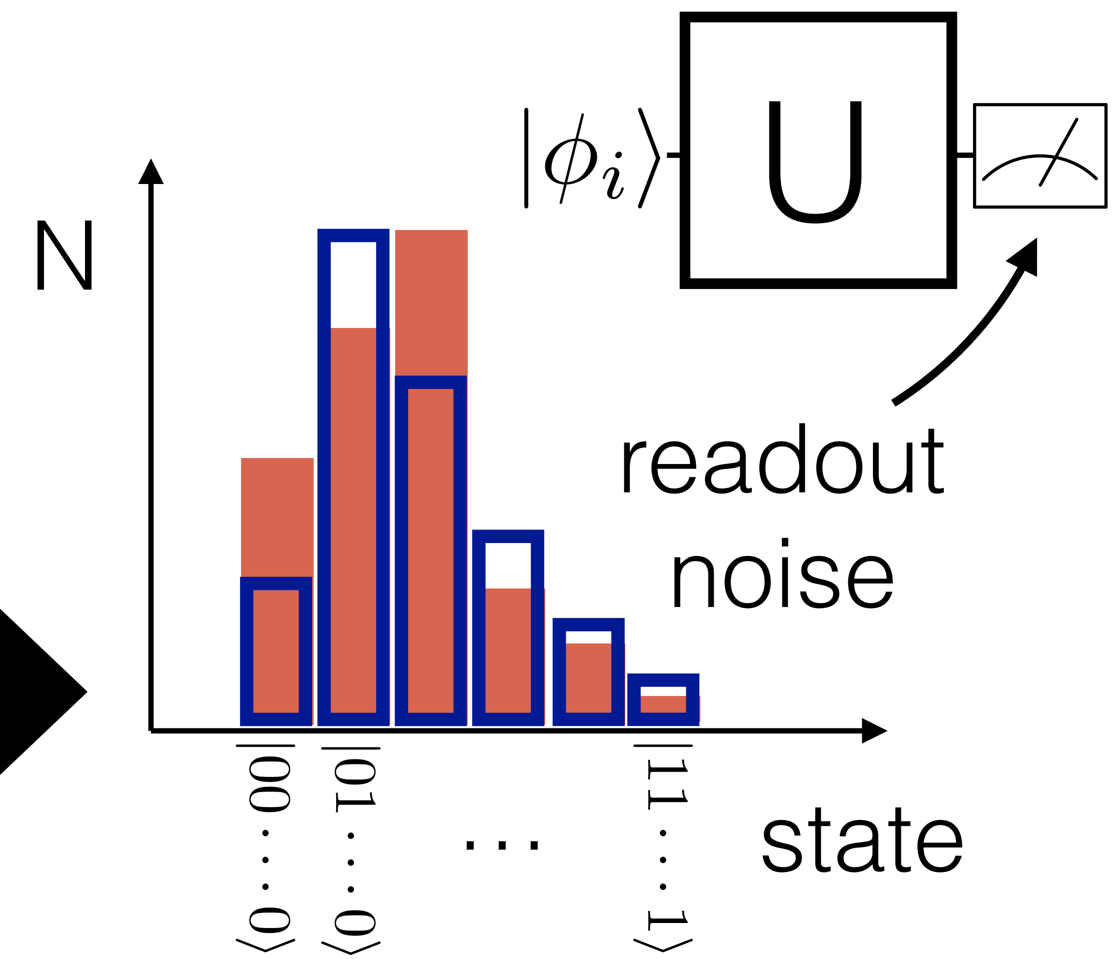
From HEP to QIS

In addition to basic algorithms dev. for QFT, I have discovered new ways to improve quantum computing **in general**

High Energy Physics



Quantum Computing



BPN, M. Urbanek, W. de Jong, C. Bauer, 1910.01969, A. He, **BPN**, W. de Jong, C. Bauer, 2003.nextweek

Outline for today

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

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

Likelihood-Free inference

Deep learning & **Quantum computing** for fundamental **physics**

Label-Free learning

[Deconvolution/Unfolding]

[Generative models]

[Weak supervision]

[Anomaly detection]

Uncertainties

“But what are the uncertainties on your NN”?

- question asked by every reviewer

Uncertainties

“But what are the uncertainties on your NN”?

- question asked by every reviewer

Determining uncertainties associated with a NN analysis can be tricky (ask me more later)

BPN, 1909.03081

Uncertainties

"But what are the uncertainties on your NN"?

- question asked by every reviewer

Determining uncertainties associated with a NN analysis can be tricky (ask me more later)

BPN, 1909.03081

…one point is clear: NN's exploit **high-dimensional correlations**; these are hard to model & it is hard to know how our accuracy.

High-dimensional uncertainties

Option 1: Work hard to understand the true nuisance parameters in the hypervariate parameter space.

*Can borrow from **AI safety** to set bounds.*

BPN and C. Shimmin, MLPS, NeurIPS 2019

High-dimensional uncertainties

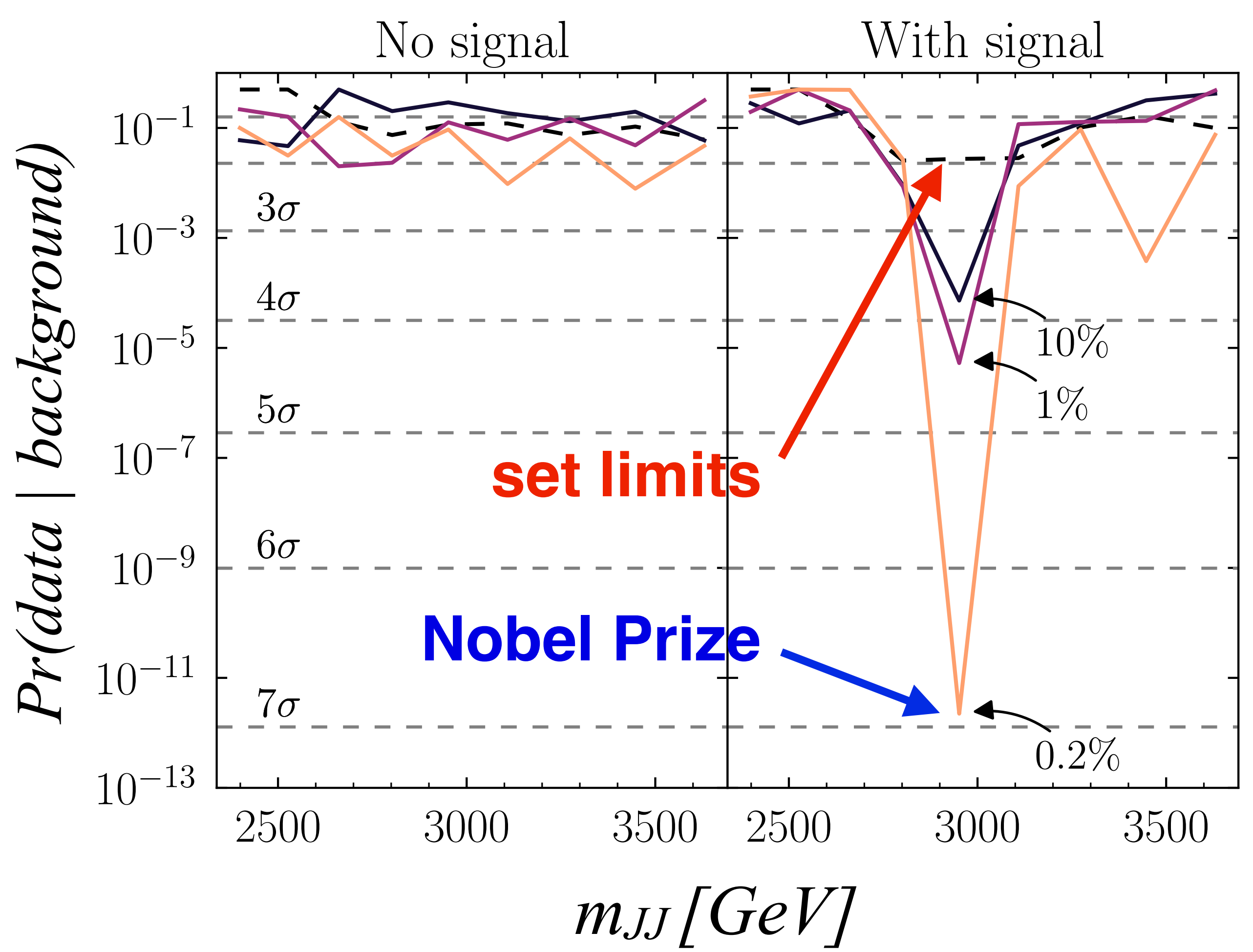
Option 1: Work hard to understand the true nuisance parameters in the hypervariate parameter space.

*Can borrow from **AI safety** to set bounds.*

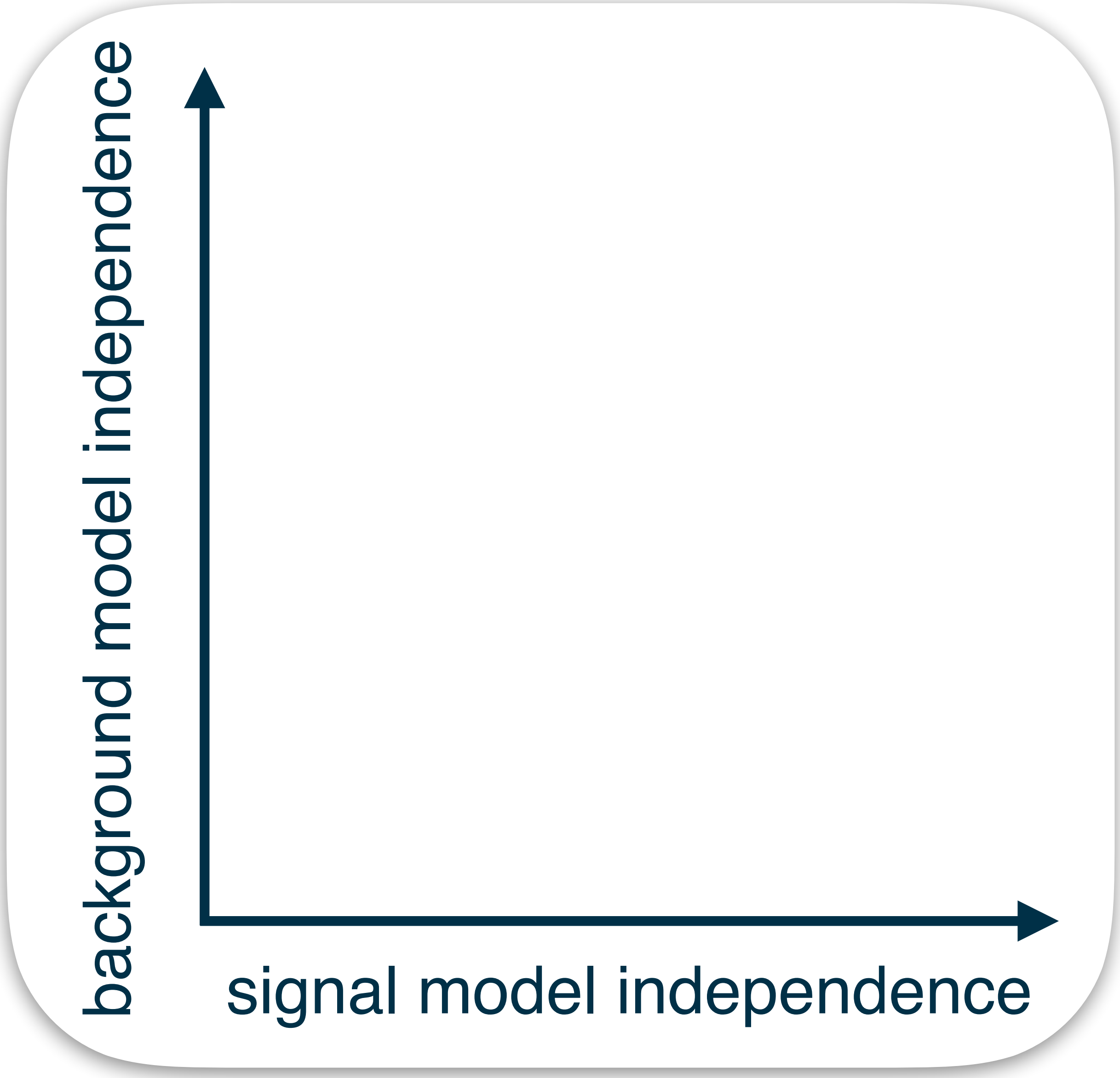
BPN and C. Shimmin, MLPS, NeurIPS 2019

Option 2:
Don't use simulation!
(not always possible!)

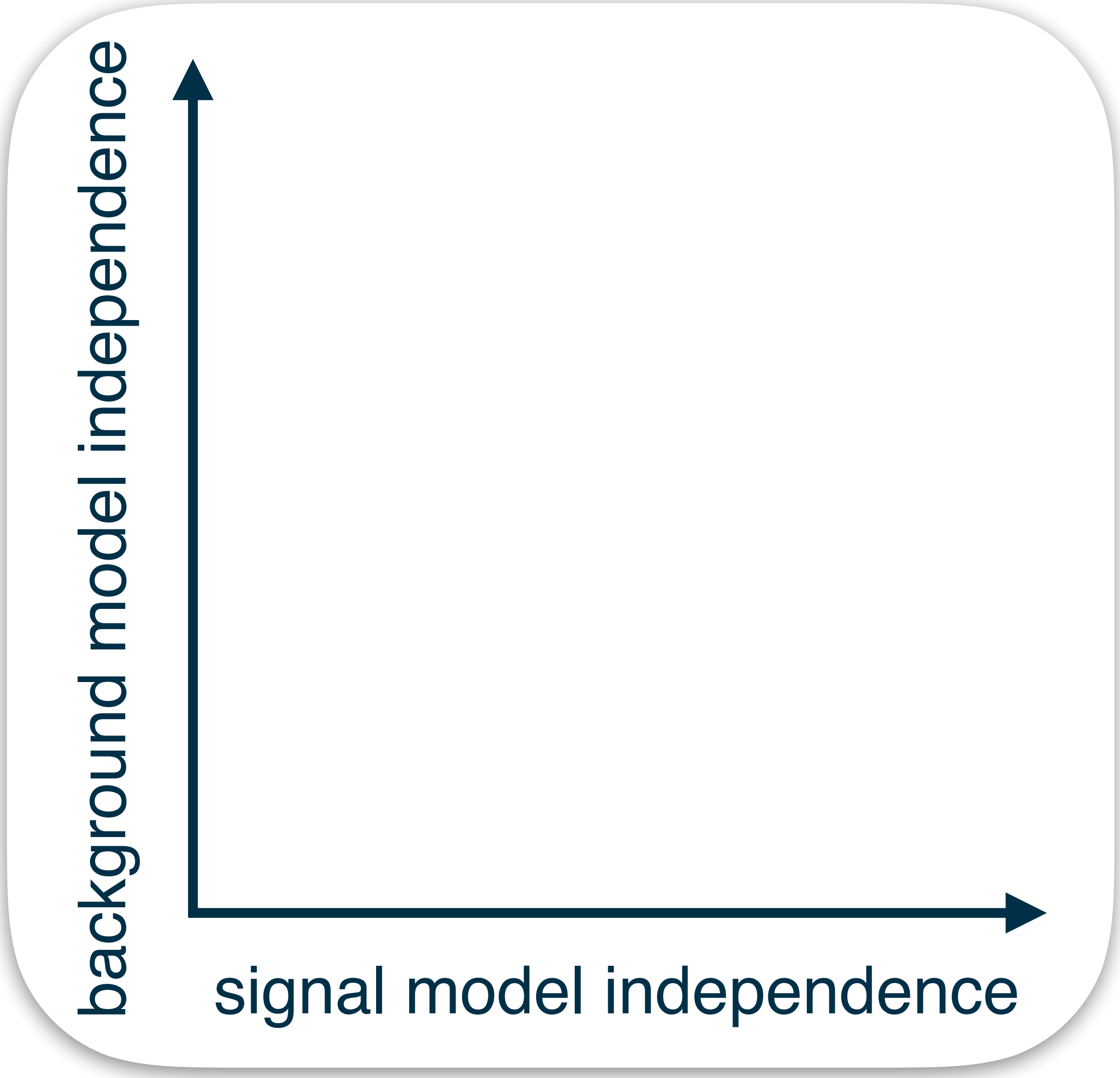
J. Collins, K. Howe, **BPN**,
Phys. Rev. Lett. 121 (2018) 241803



Model dependence



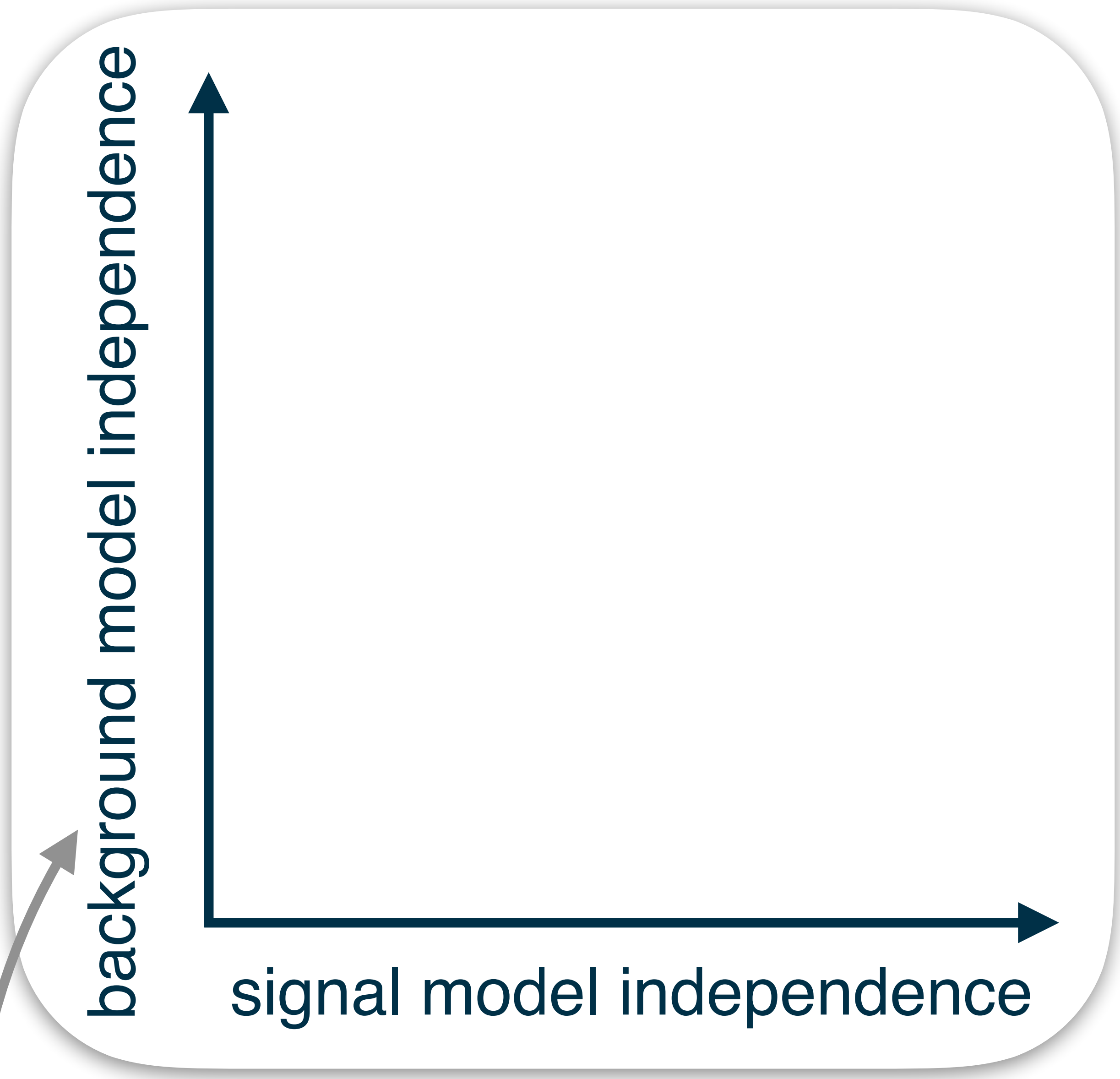
Signal sensitivity



Background specificity

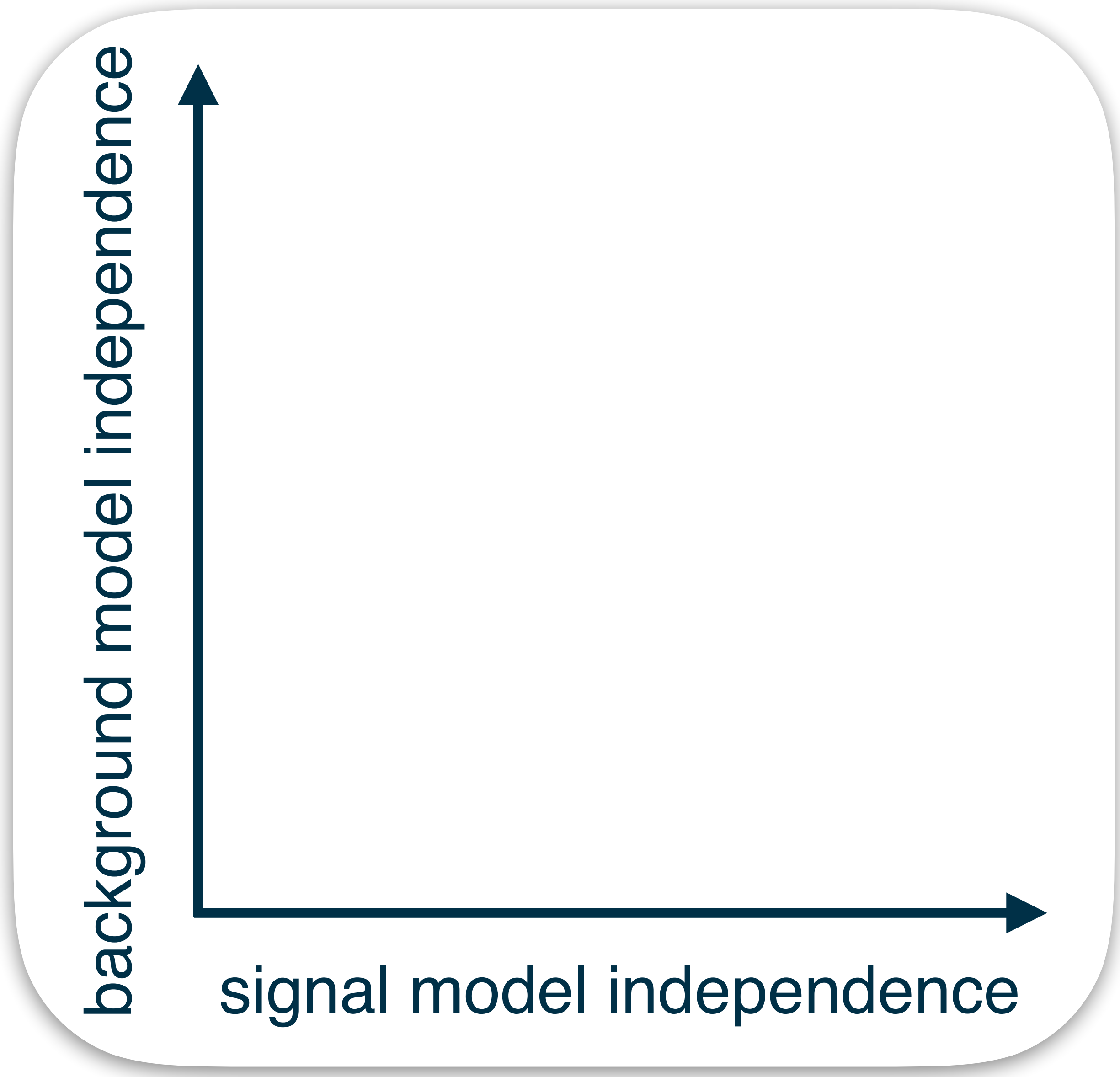
Suppose you want to search for a new signal process

Model dependence



Signal sensitivity

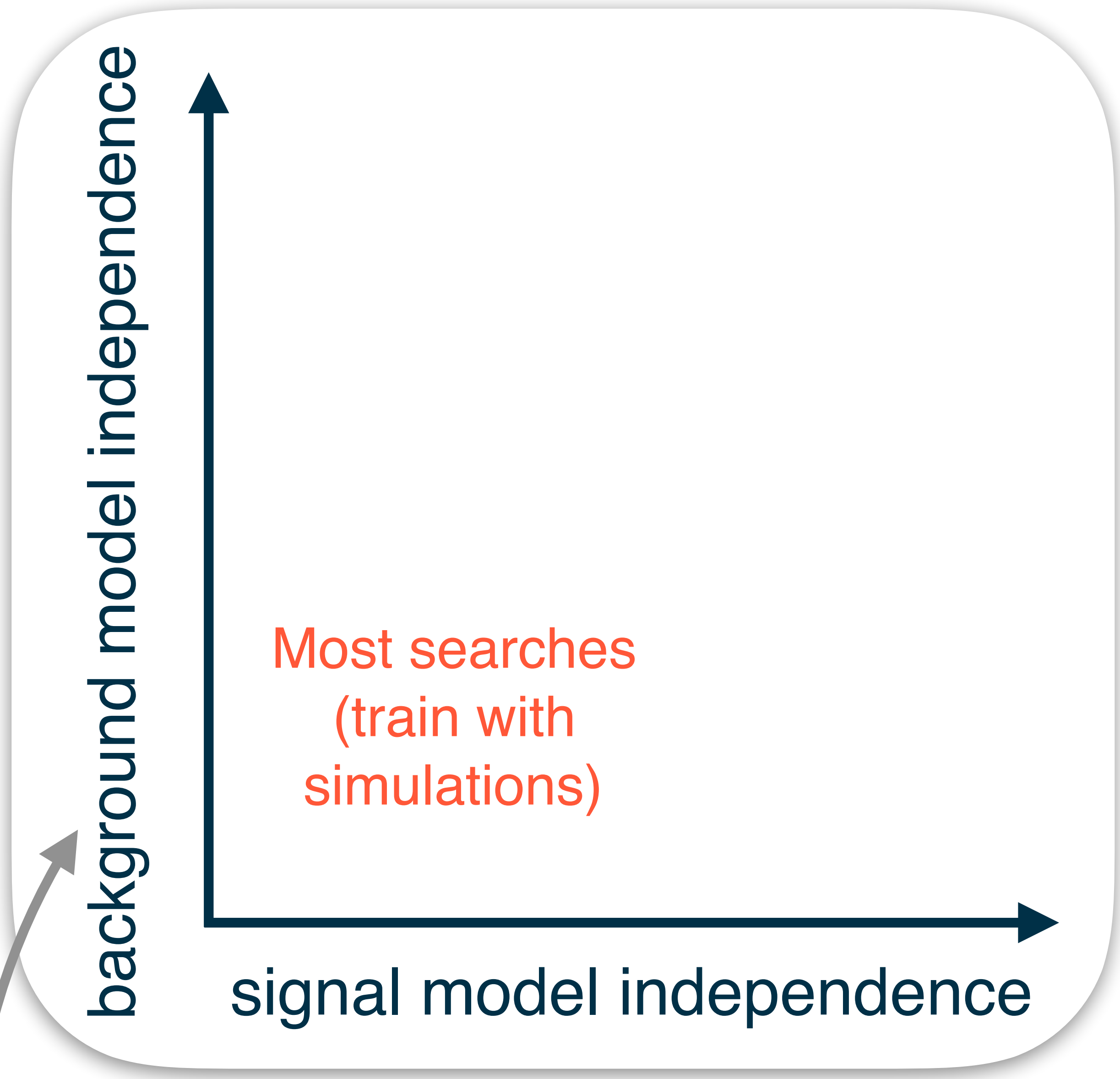
Standard Model



Background specificity

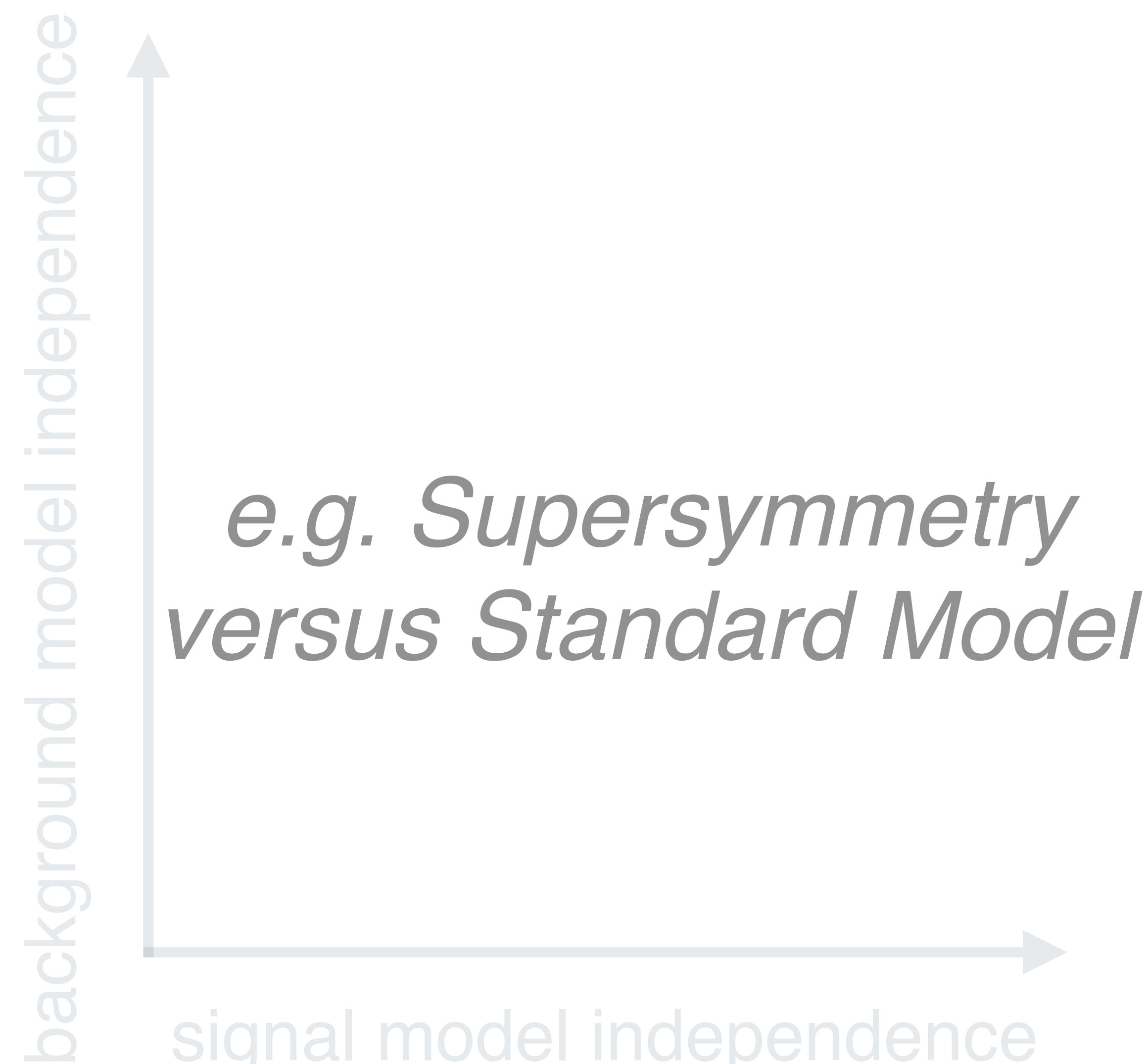
BPN, D. Shih, 2001.04990

Model dependence



Signal sensitivity

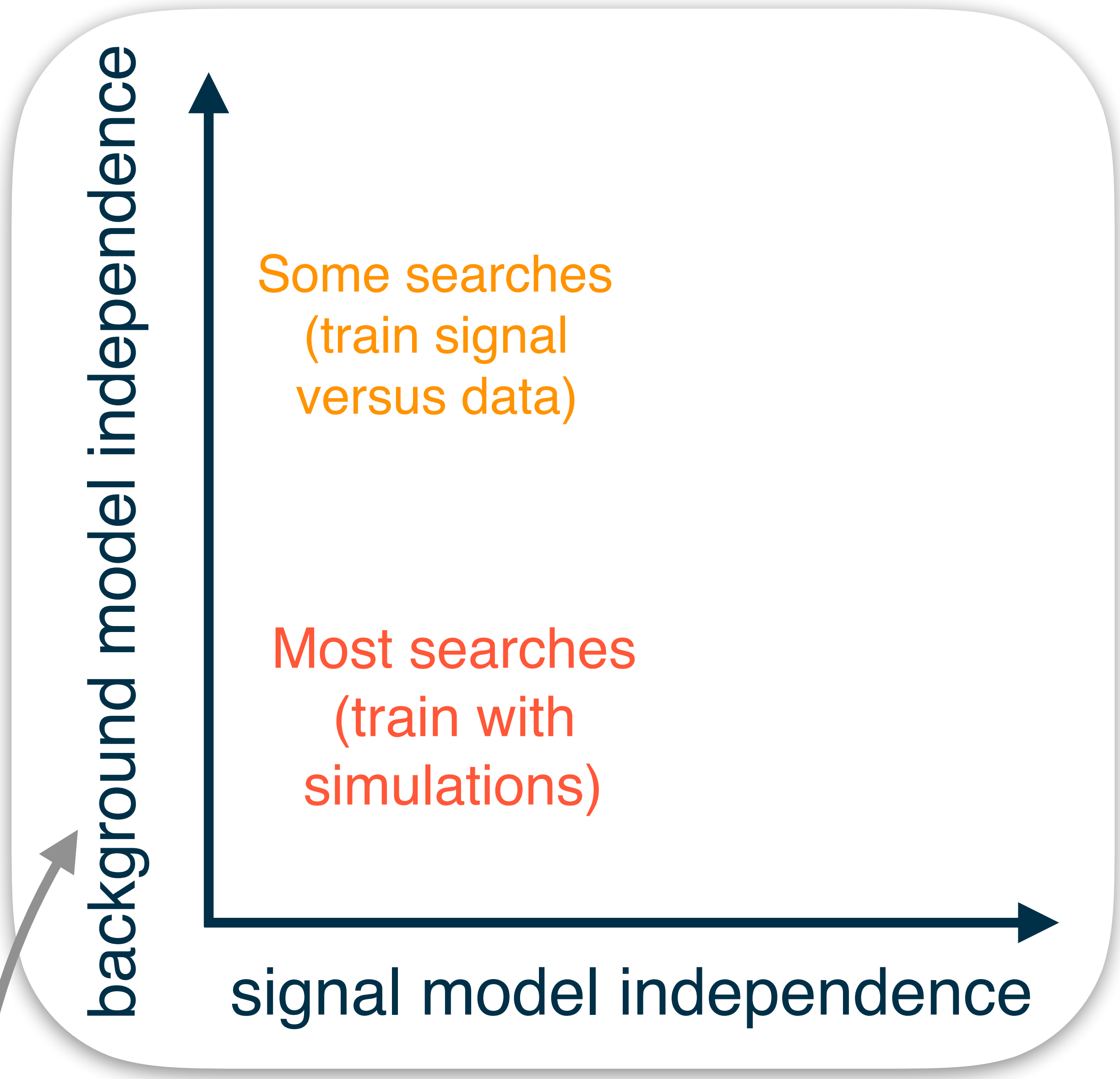
Standard Model



Background specificity

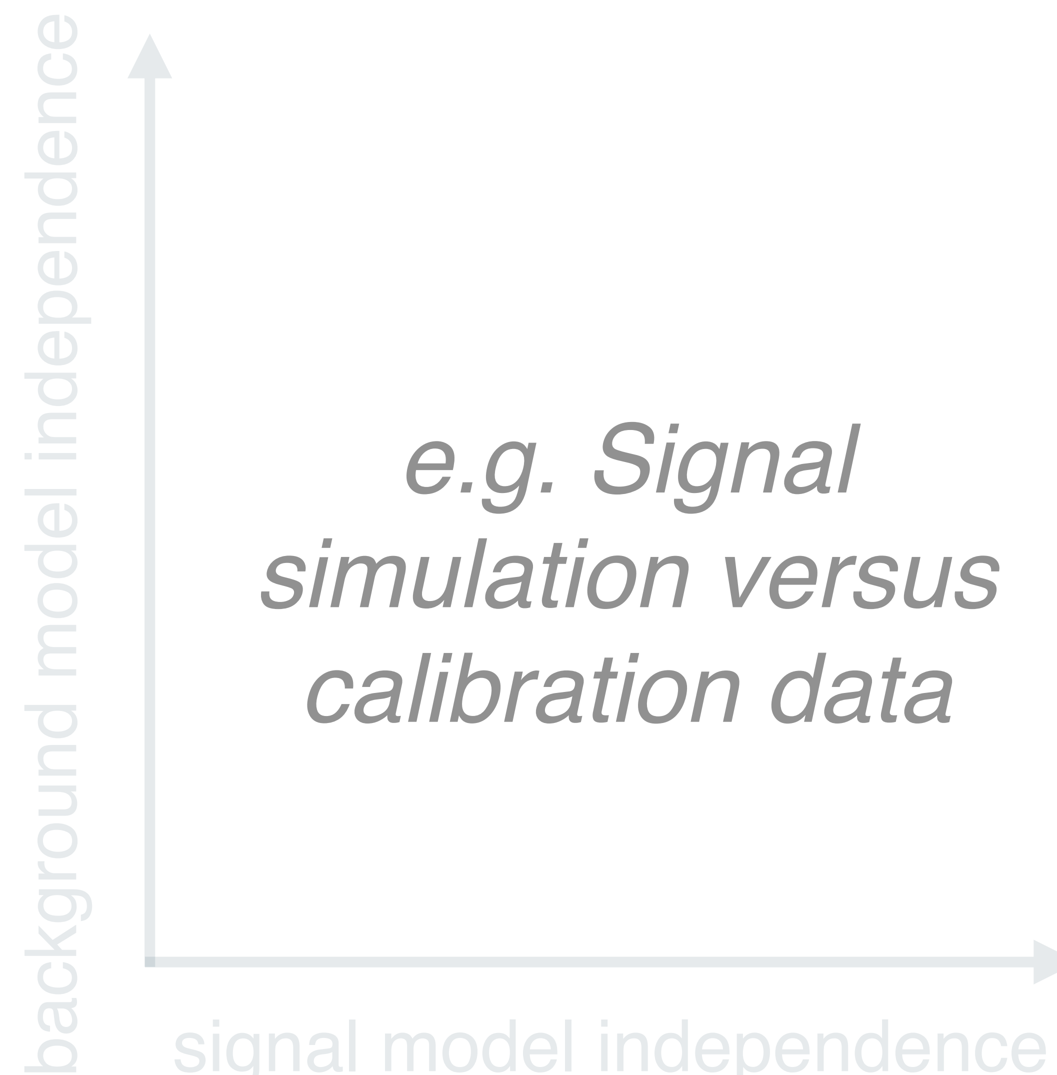
BPN, D. Shih, 2001.04990

Model dependence



Signal sensitivity

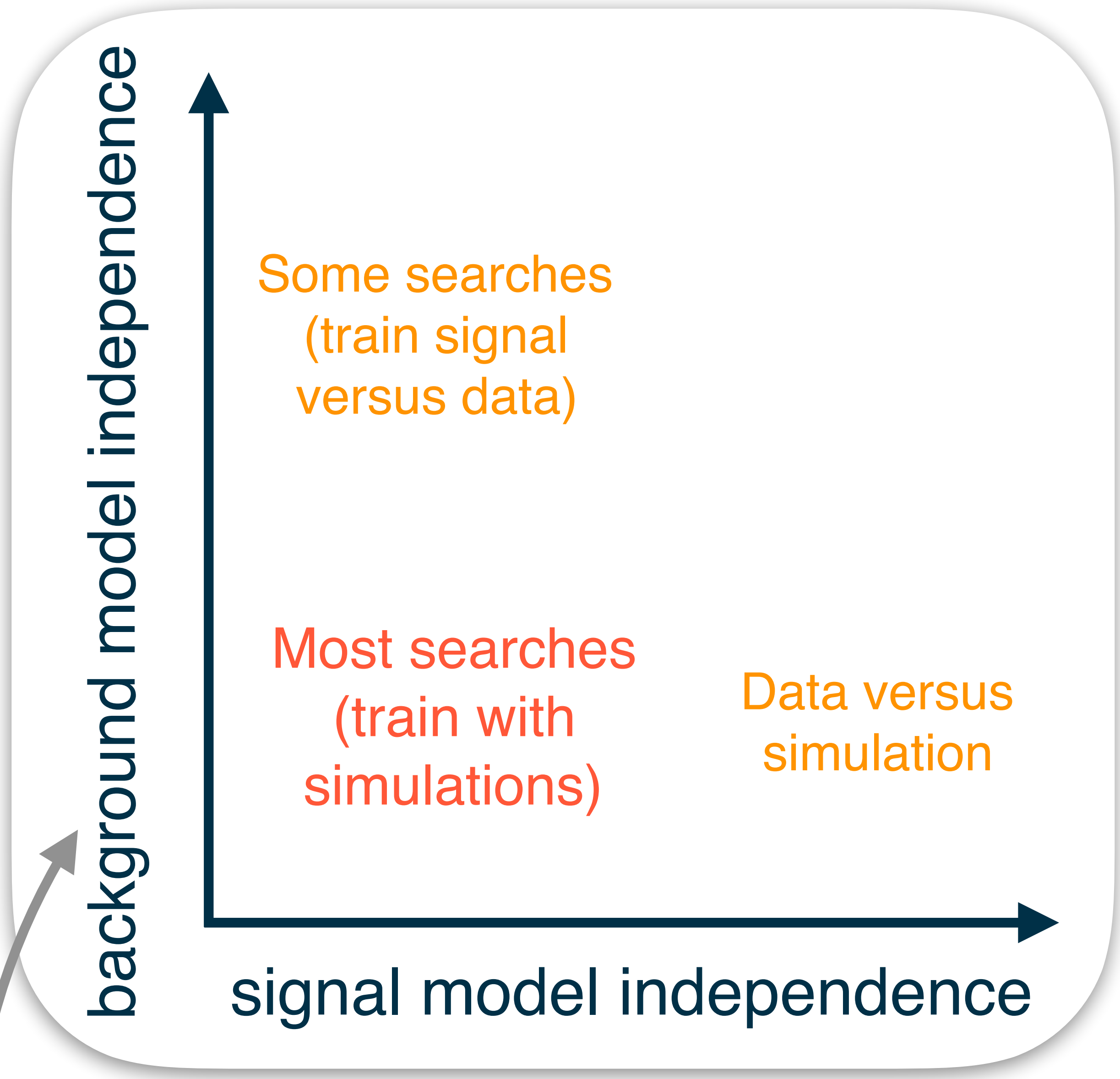
Standard Model



Background specificity

BPN, D. Shih, 2001.04990

Model dependence



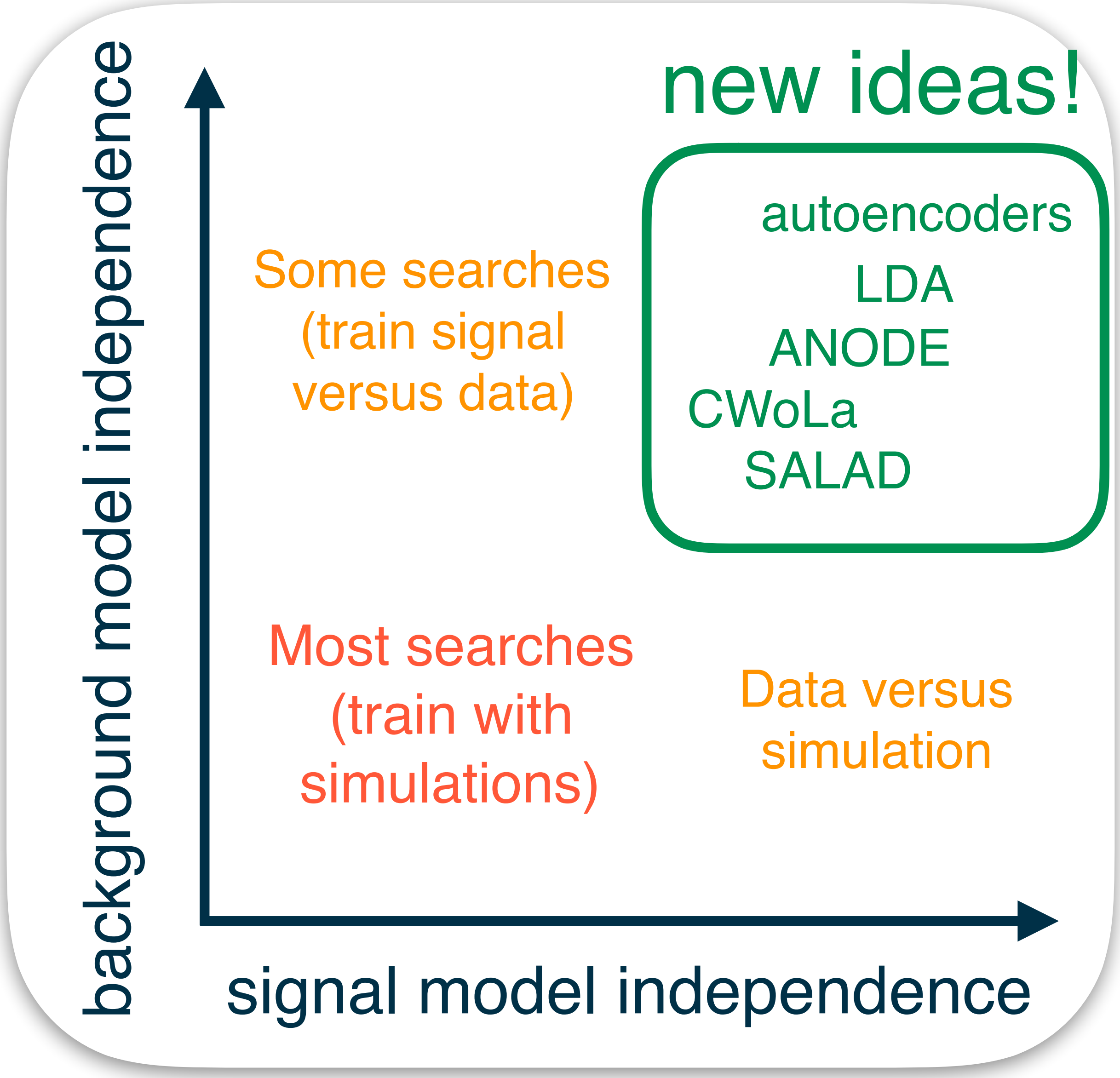
Signal sensitivity

Standard Model

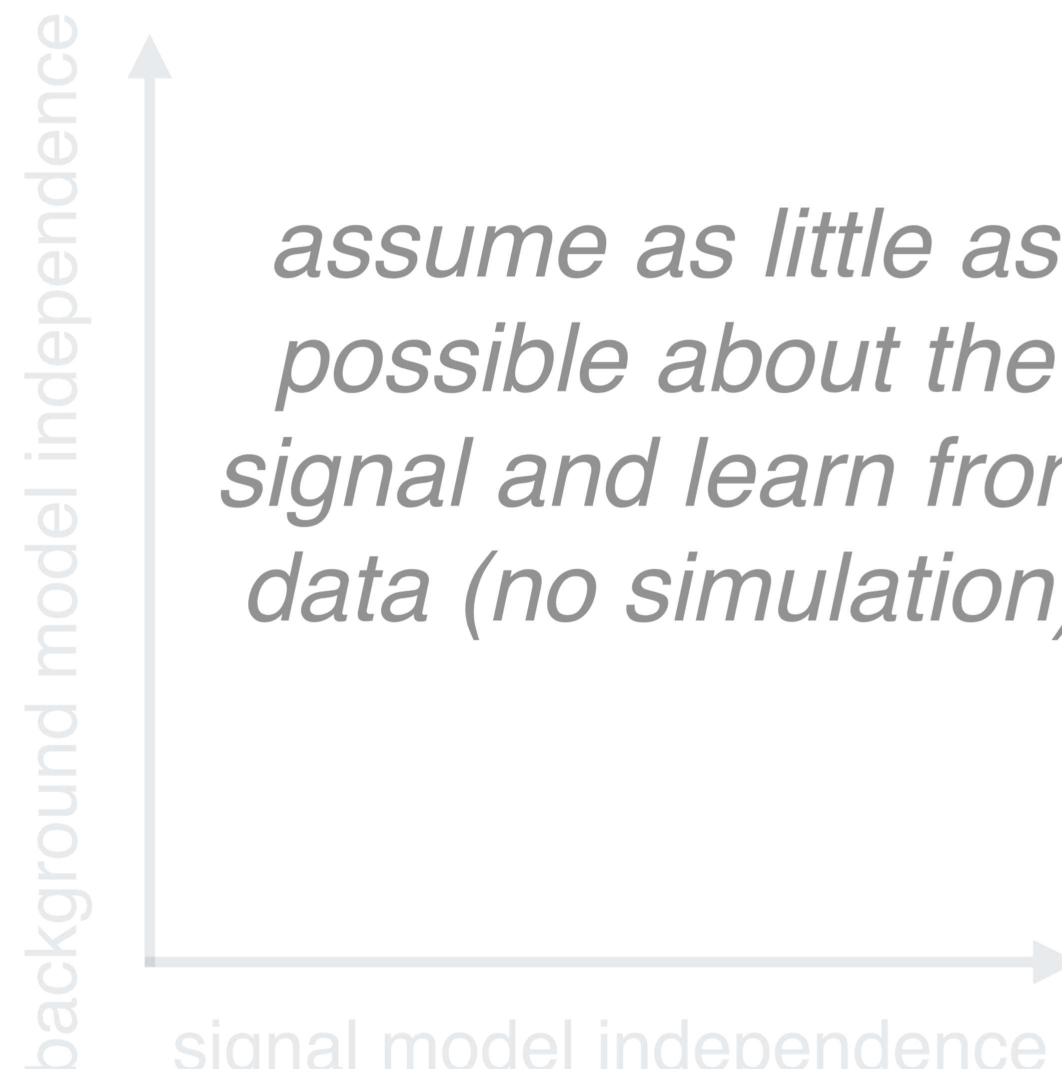
Background specificity

BPN, D. Shih, 2001.04990

Model dependence



Signal sensitivity



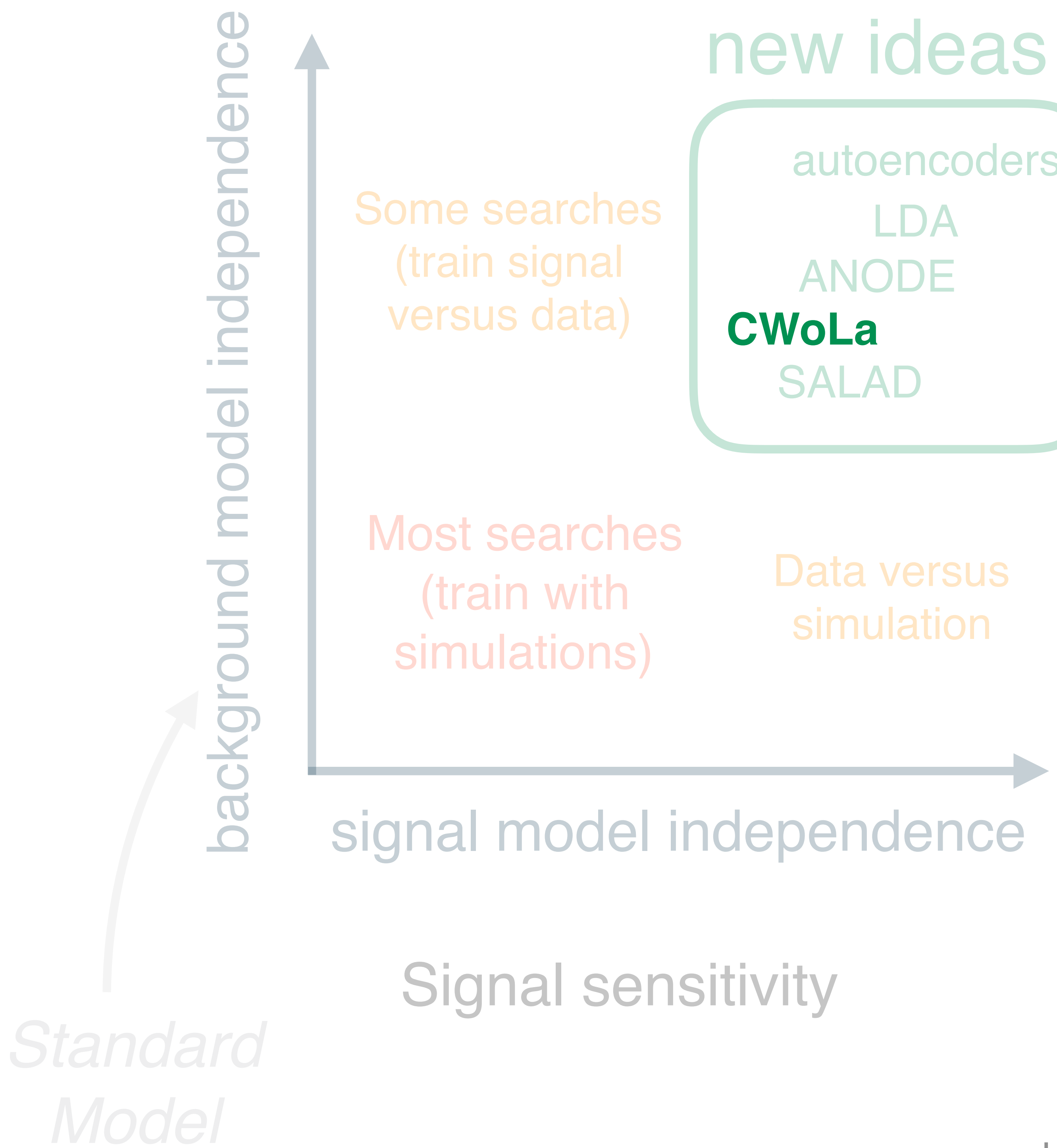
assume as little as possible about the signal and learn from data (no simulation)

Background specificity

M. Farina, Y, Nakai, D. Shih, 1808.08992 + others
B. Dillon et al., PRD 100 (2019) 056002
BPN, D. Shih, 2001.04990

J. Collins, K. Howe, **BPN**, PRL 121 (2018) 241803
A. Andreassen, **BPN**, D. Shih, 2001.05001

Model dependence



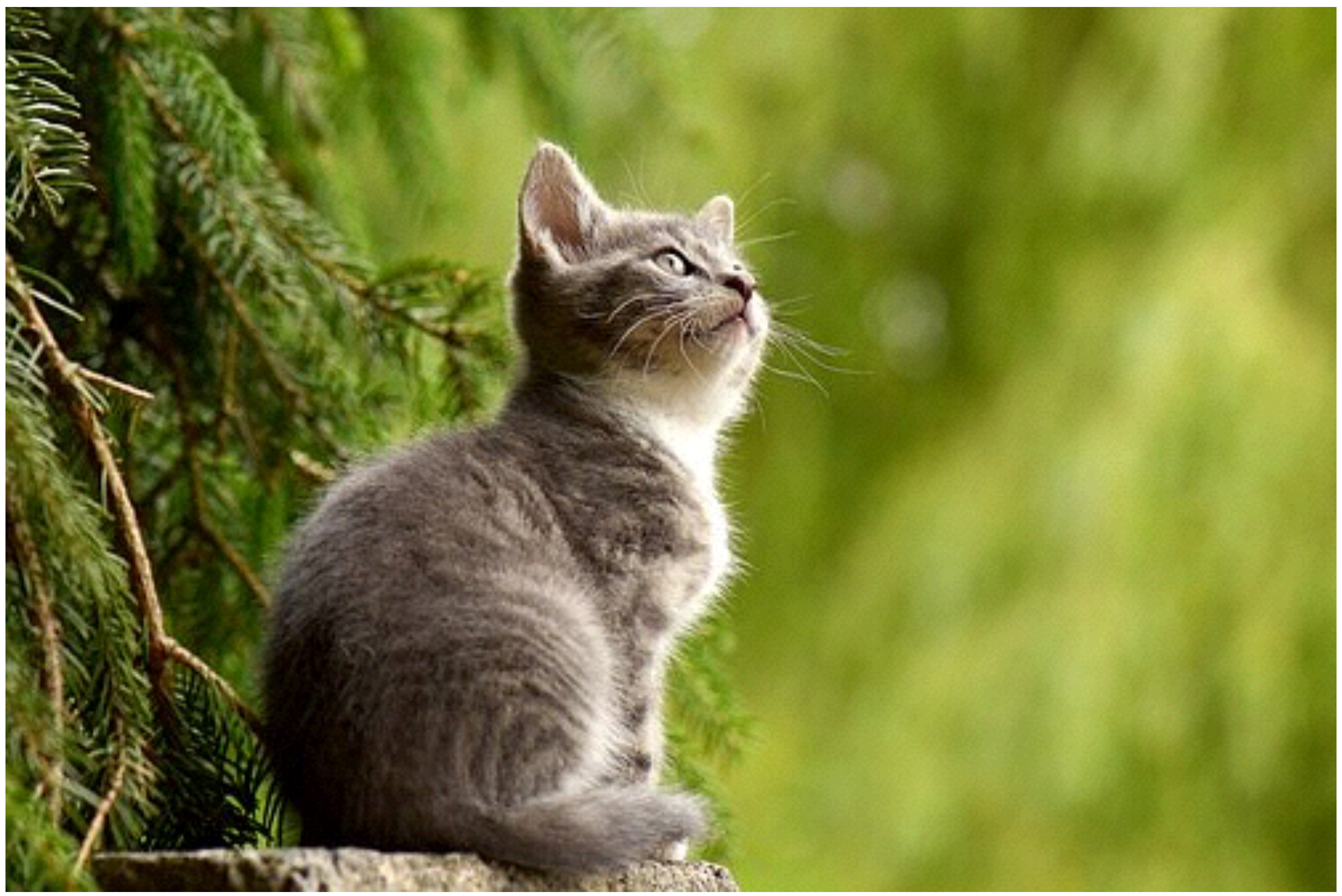
Signal sensitivity

Background specificity

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

What is the problem?

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



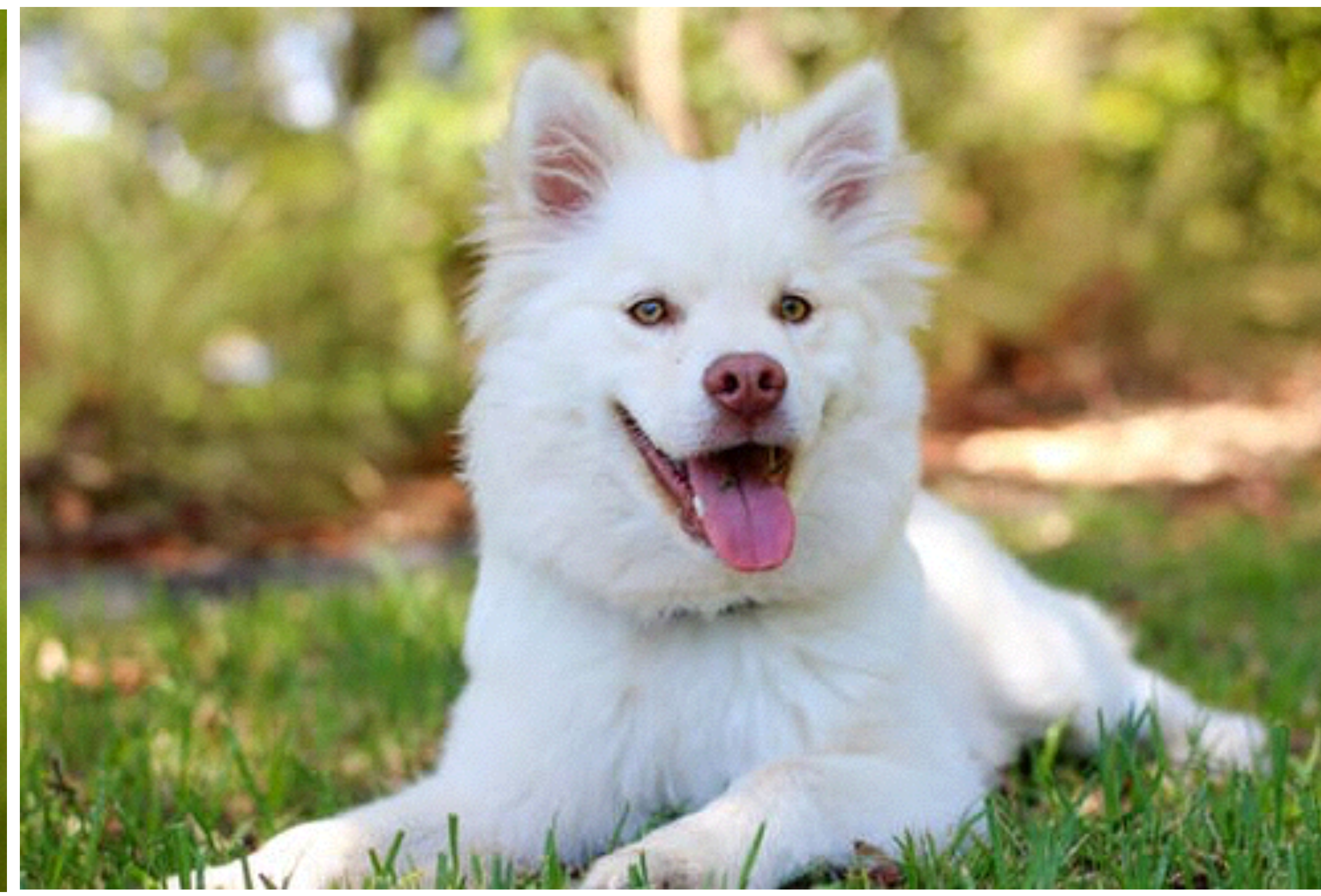


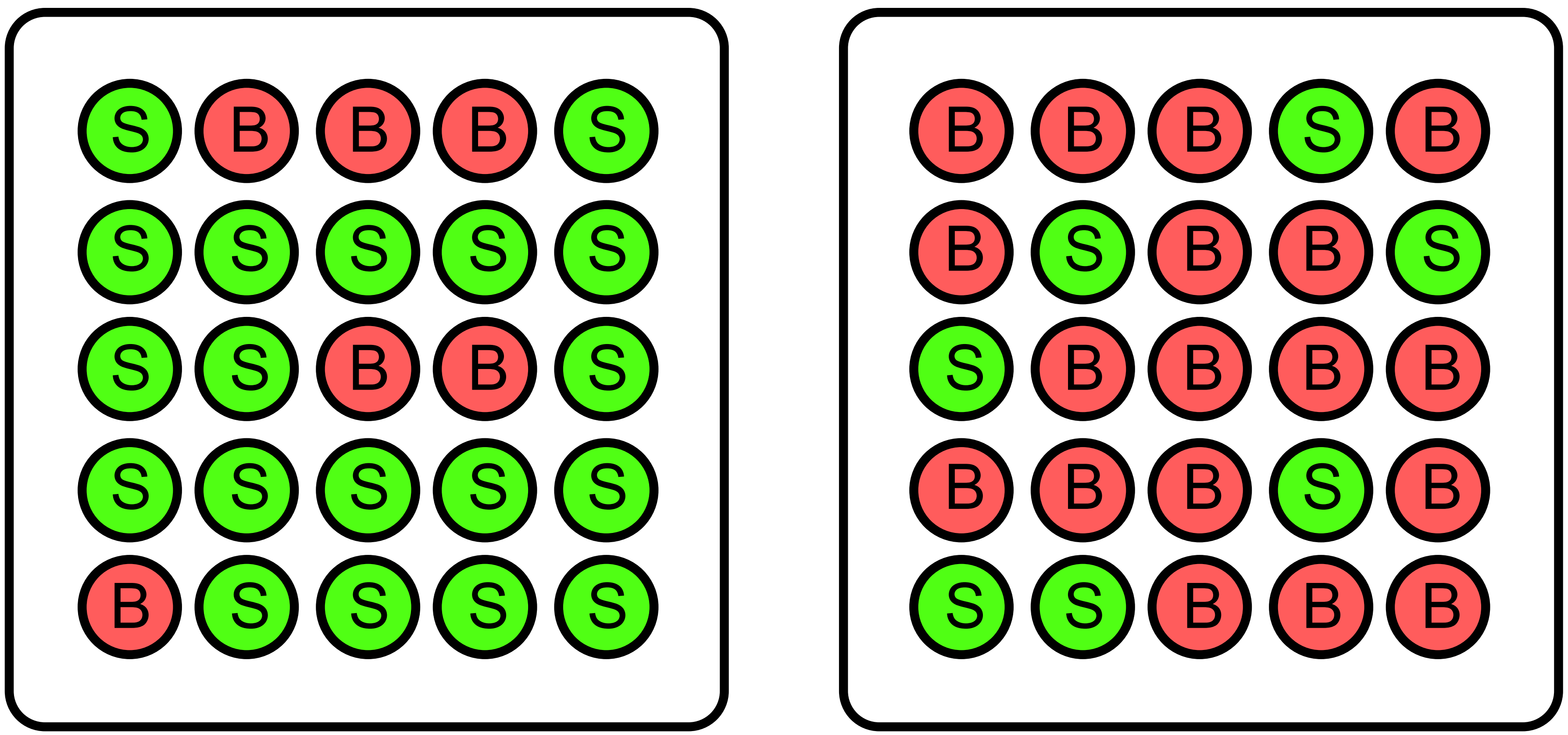
Image credit: pixabay.com

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

What is the problem?

The data are unlabeled and in the best case, come to us as mixtures of two classes (“signal” and “background”).

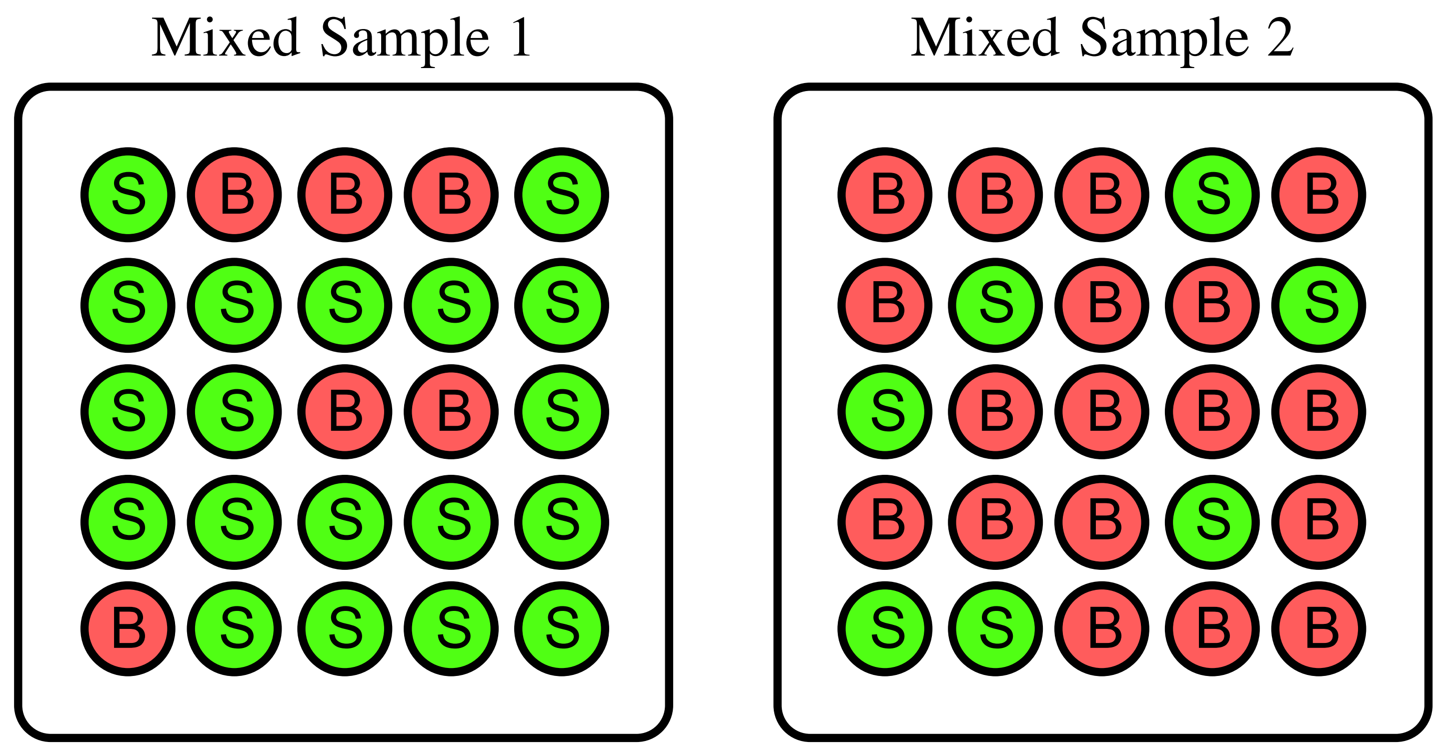
Mixed Sample 1 Mixed Sample 2



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

Weak supervision: *Classification Without Labels*

Can we learn without any label information?



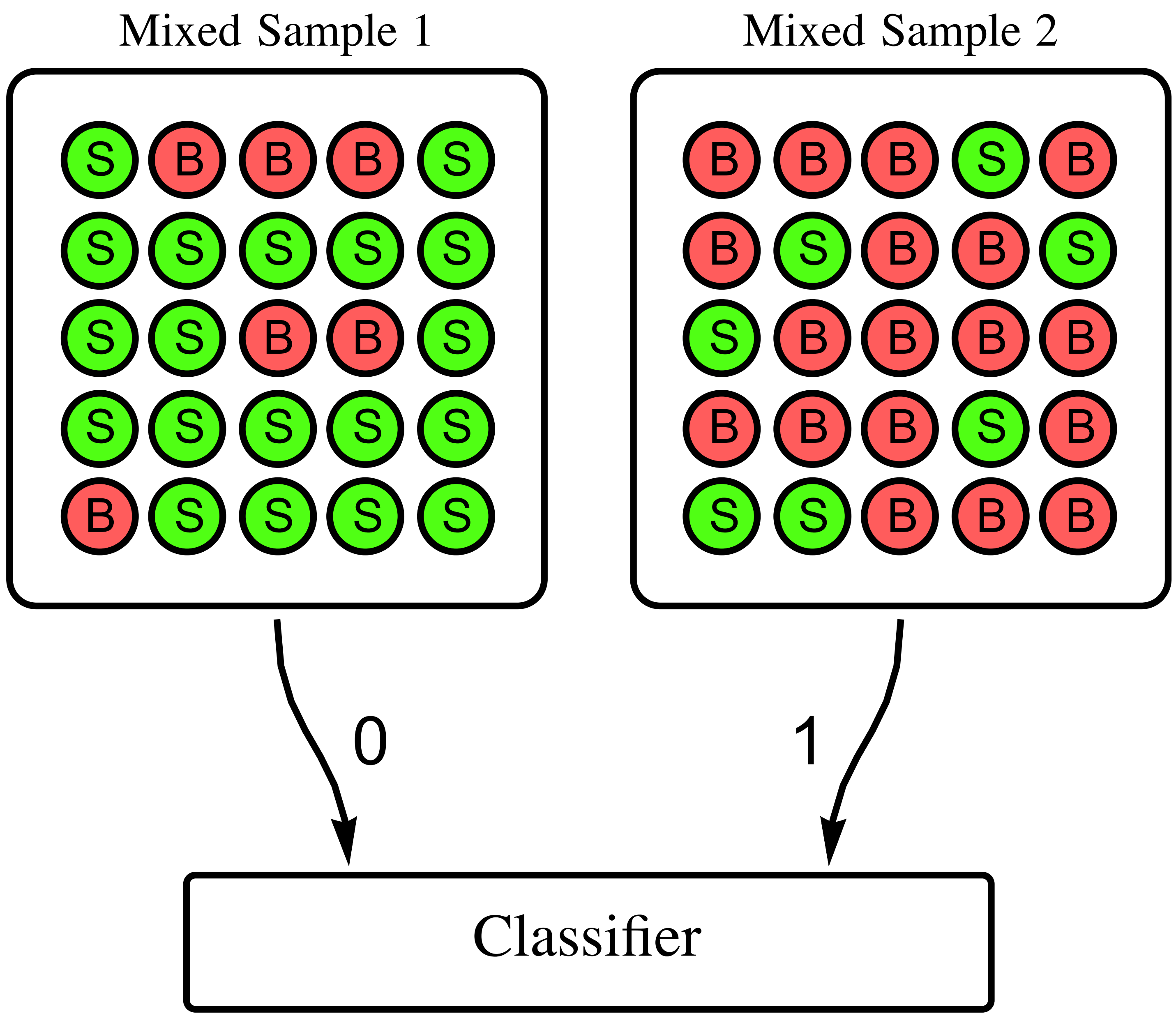
E. Metodiev, **BPN**, J. Thaler, JHEP 10 (2017) 51

Weak supervision:
***C**lassification **W**ith**o**ut **La**bels*

Can we learn without any label information?

Yes !

Training on impure samples is (asymptotically) equivalent to training on pure samples



E. Metodiev, **BPN**, J. Thaler, JHEP 10 (2017) 51

How can we use CWoLa to **hunt for new** particles?

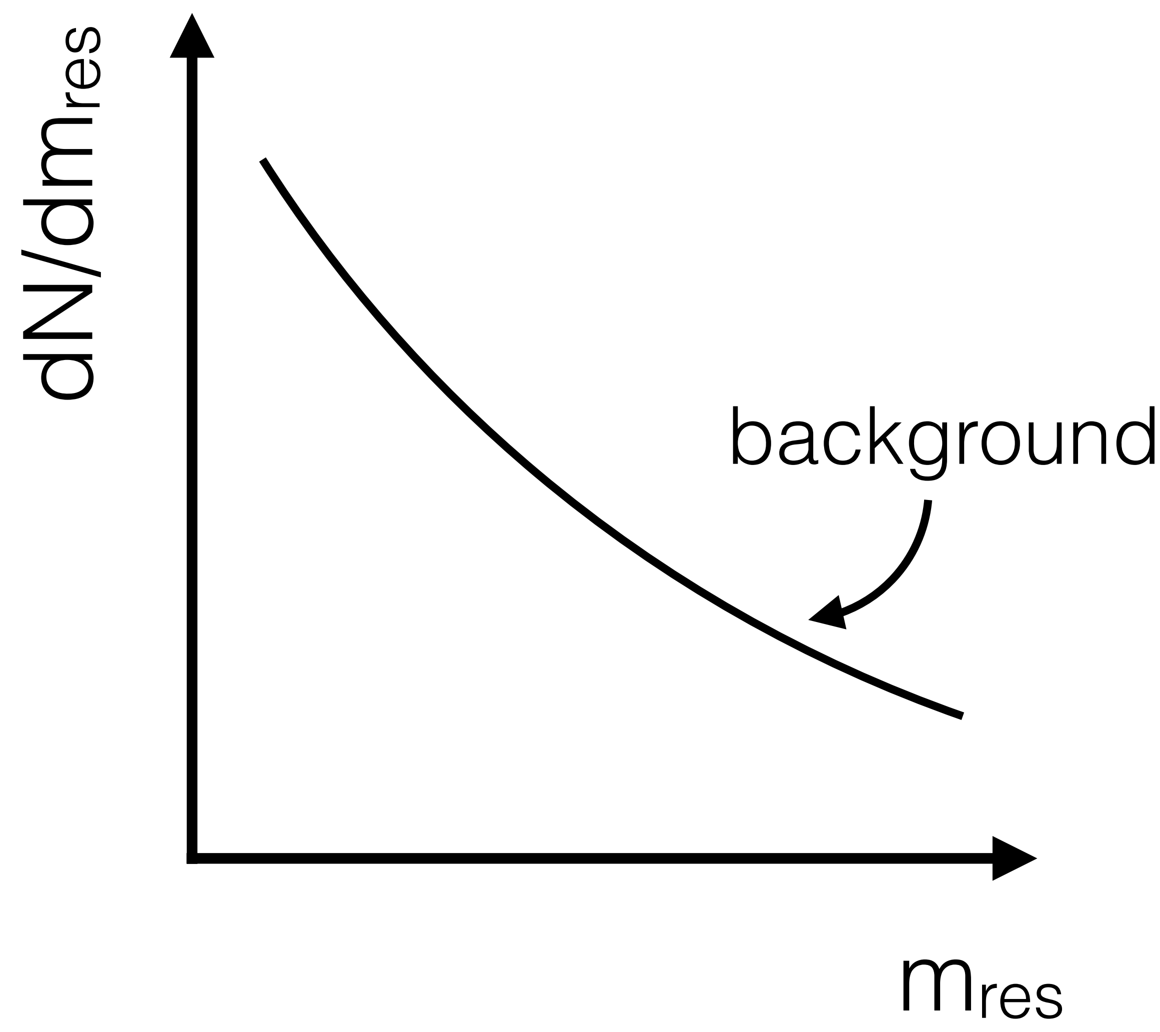
CWoLa Hunting

How can we use CWoLa to **hunt for new** particles?

*Image from *The Courier Mail*. Koala is actually being freed - I do not condone violence against these animals!

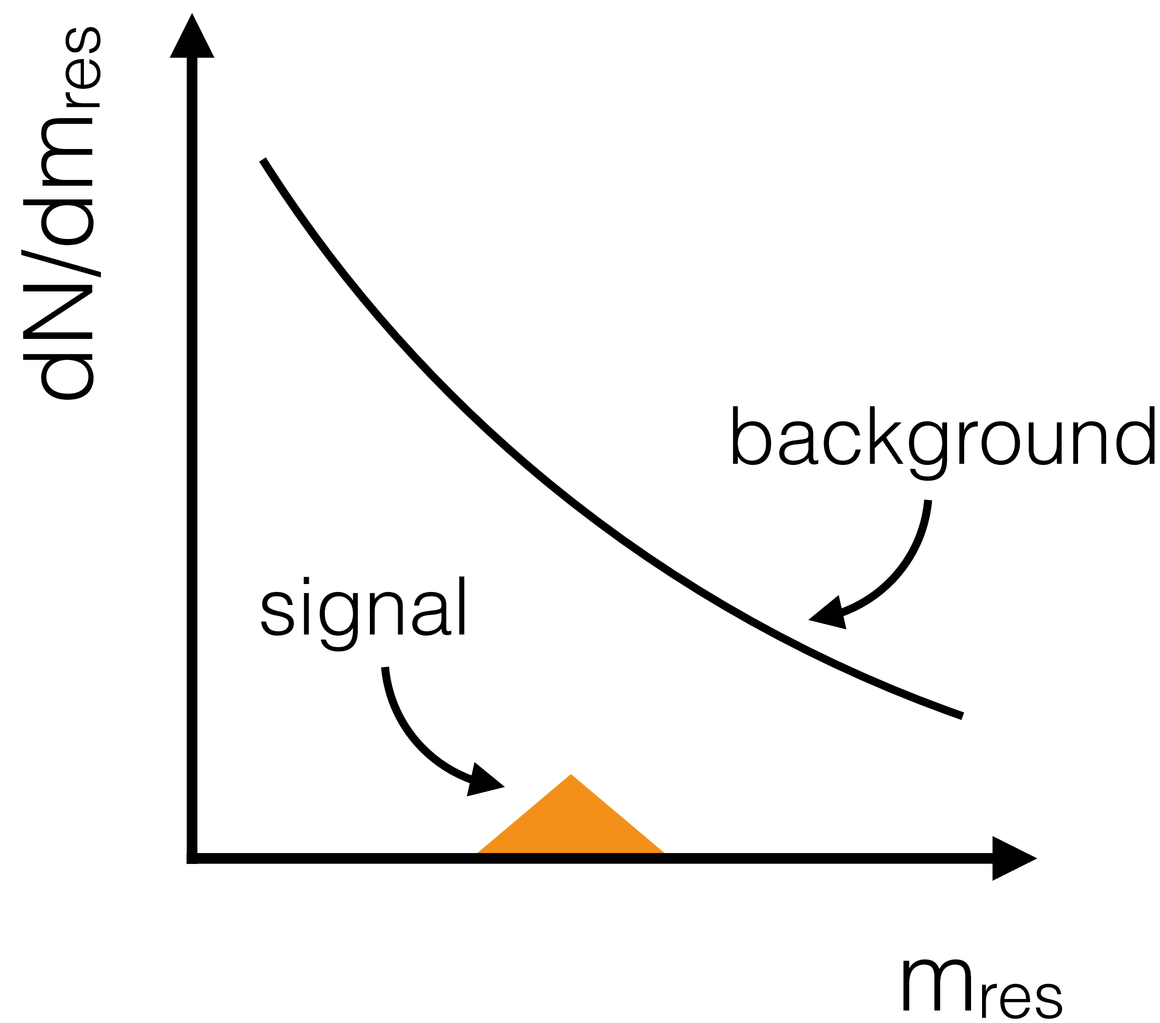
CWoLa Hunting

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



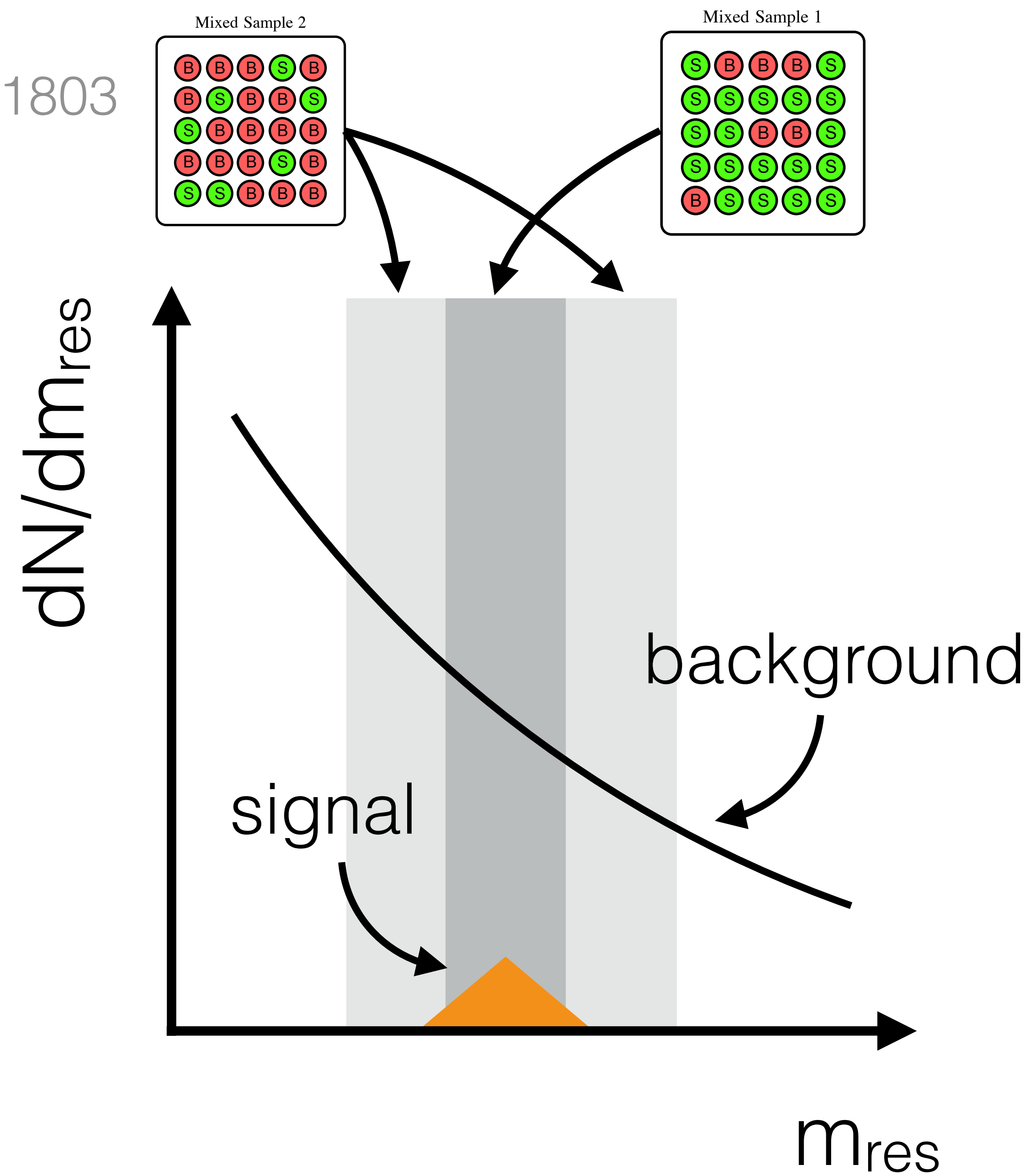
CWoLa Hunting

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



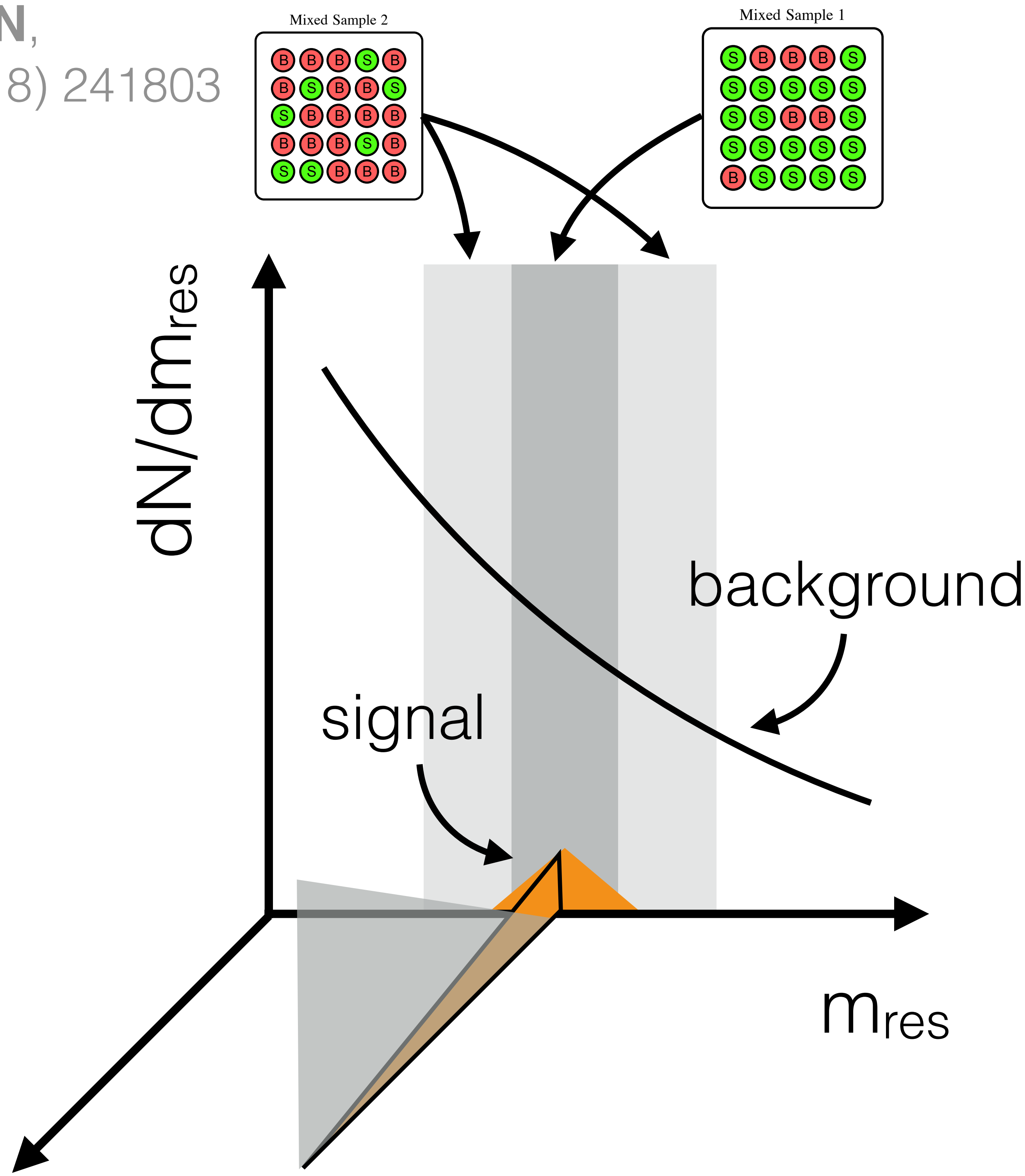
CWoLa Hunting

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



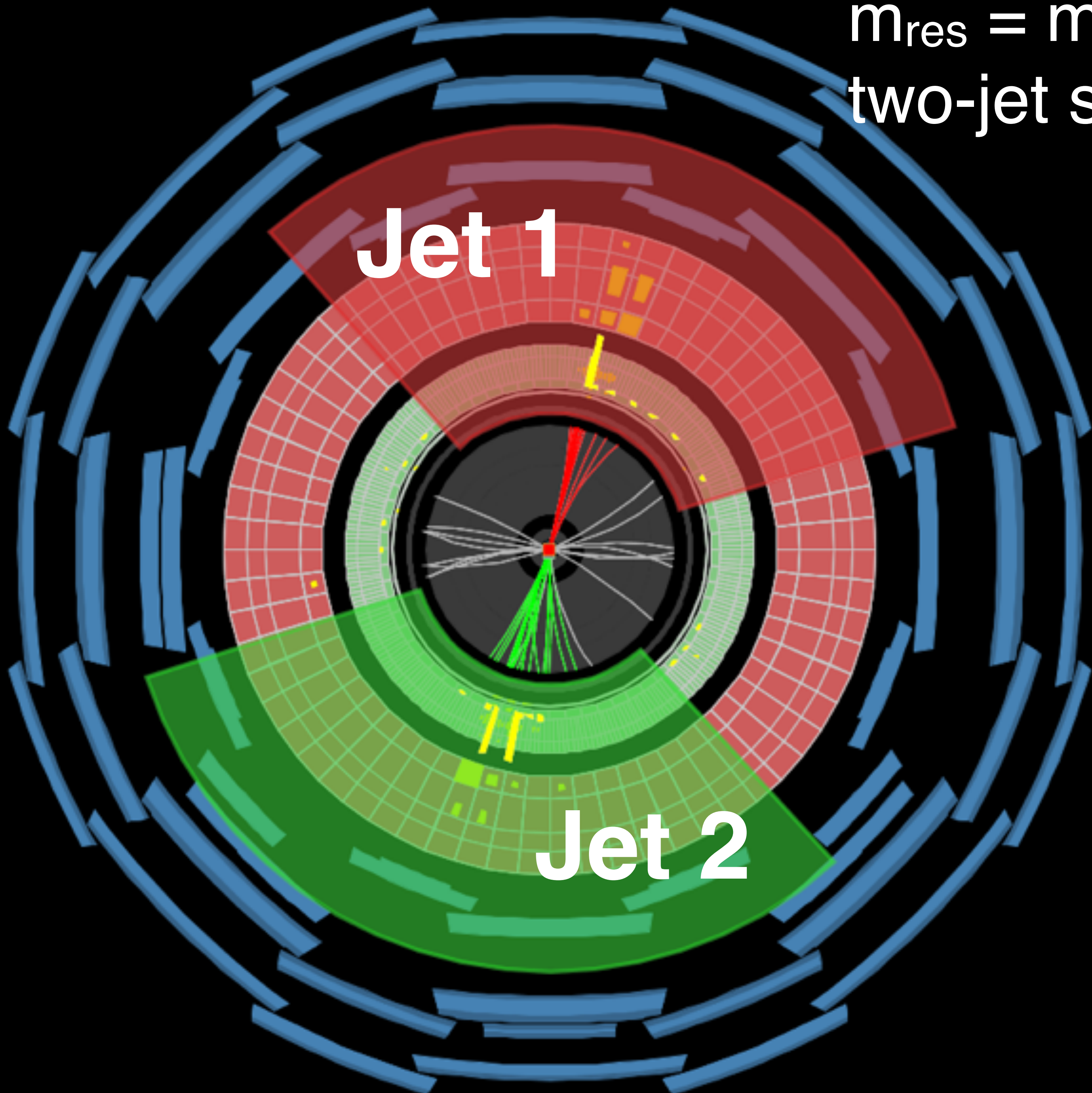
CWoLa Hunting

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

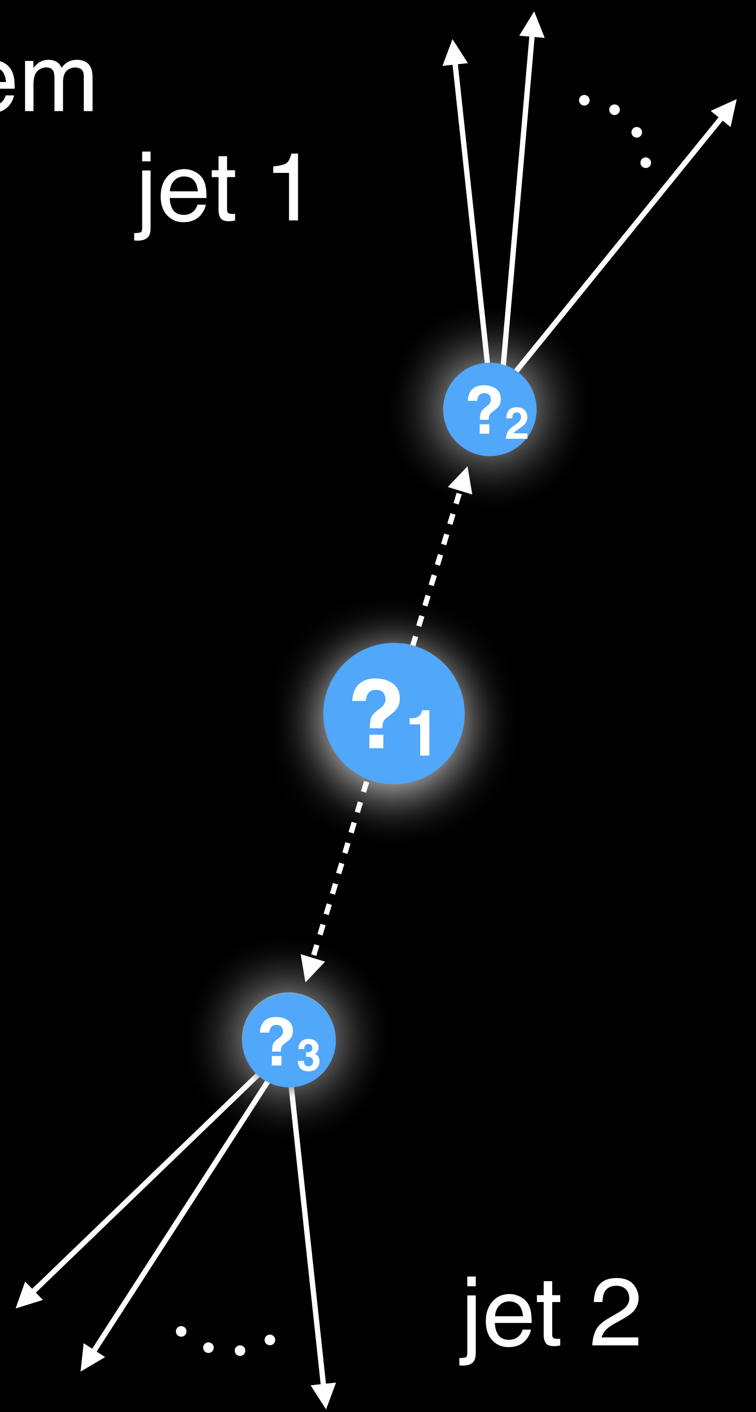


+ be careful to not pay a big trails factor
(ask if interested)

Example: two-“jet” search



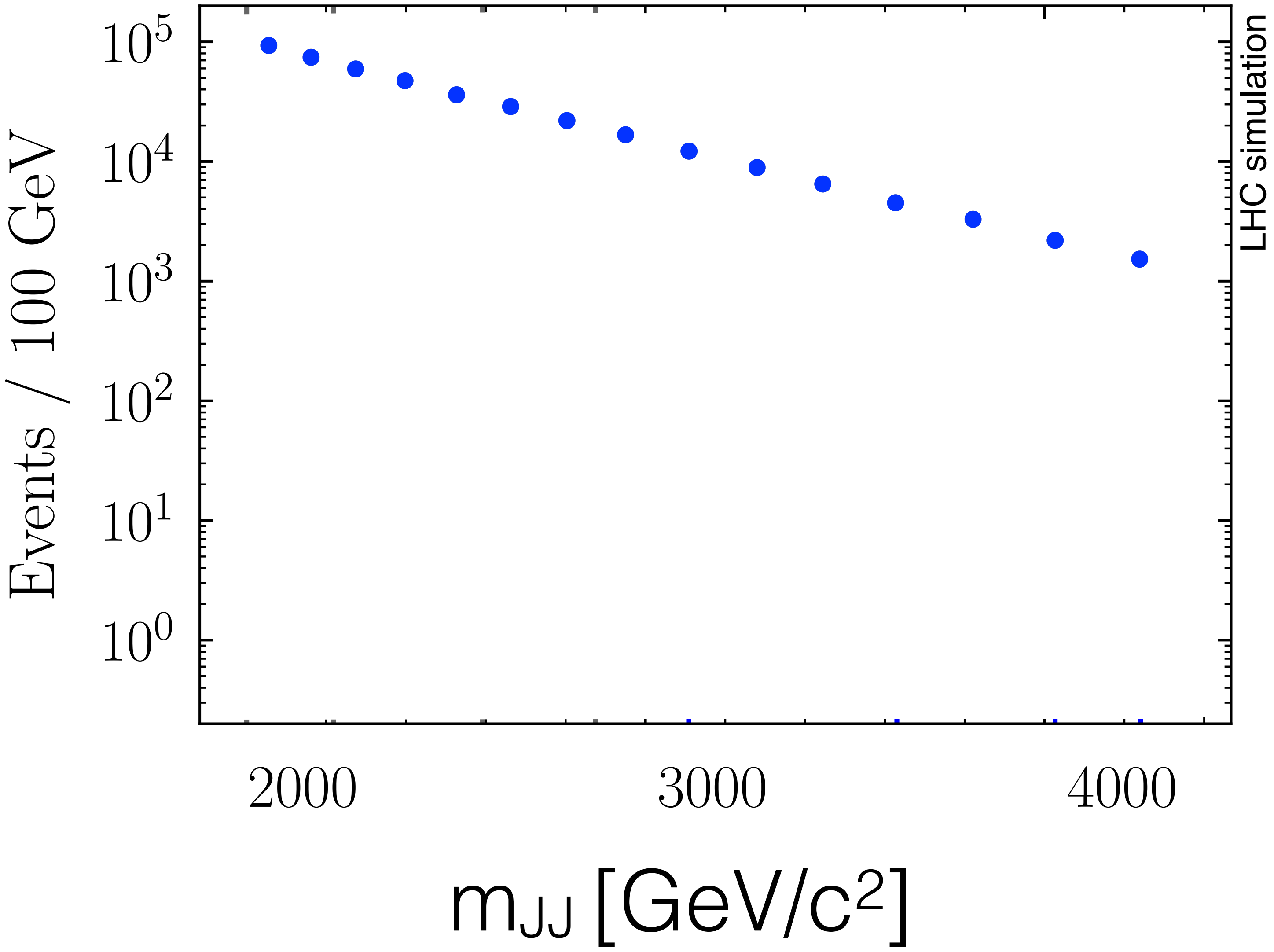
m_{res} = mass of two-jet system



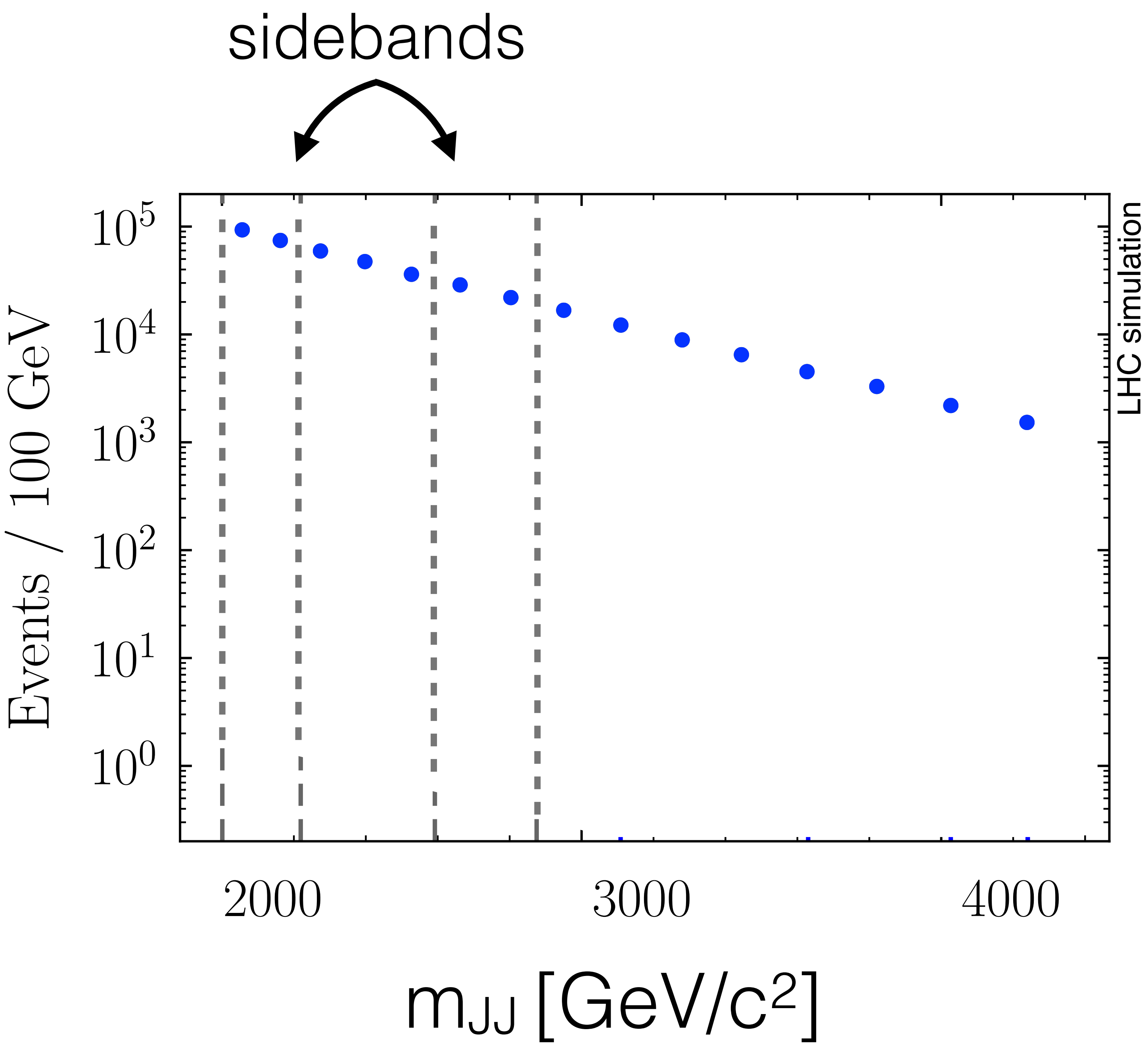
collisions in/out of page

y = many features of the two jets

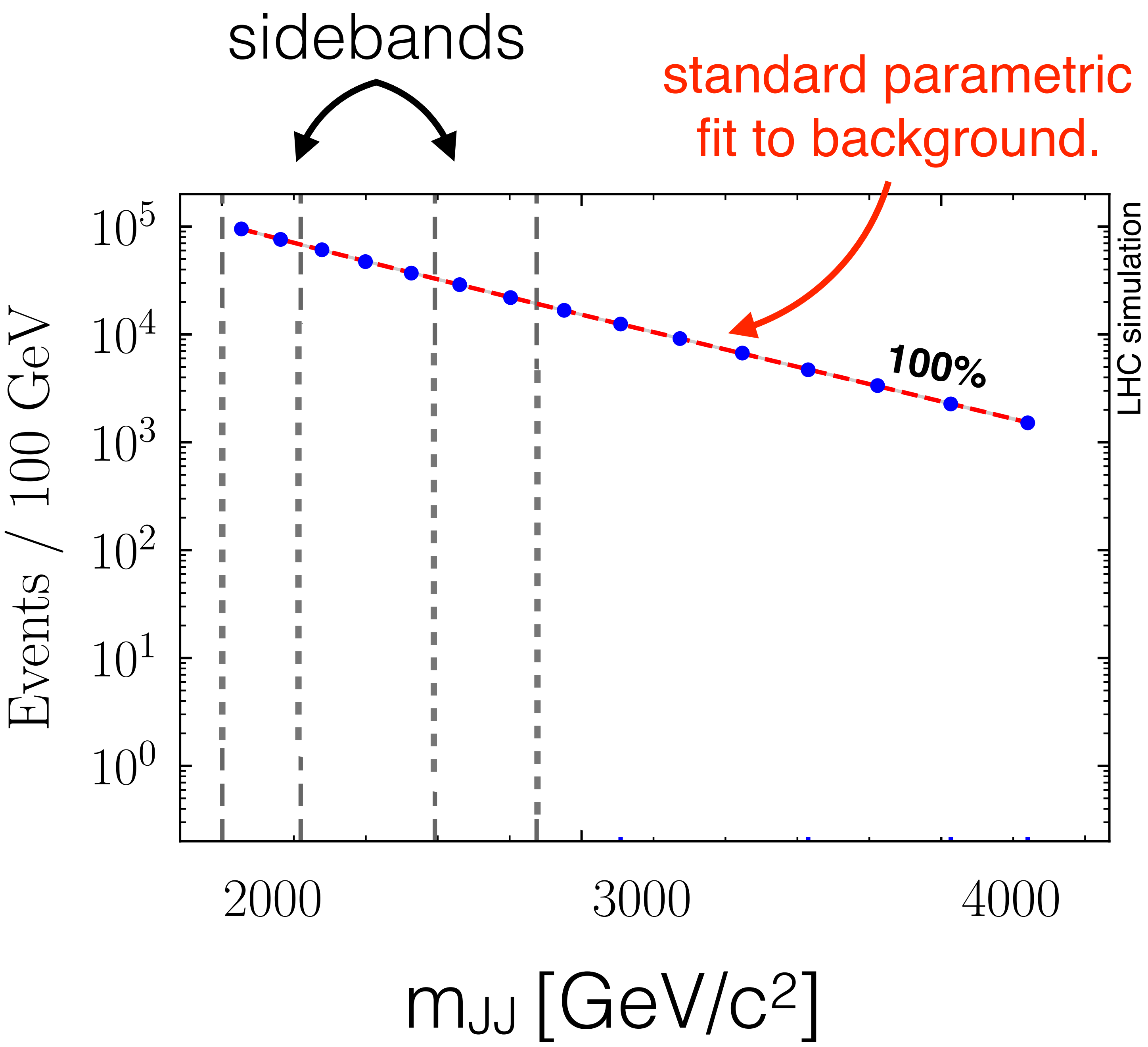
Example: two-jet search



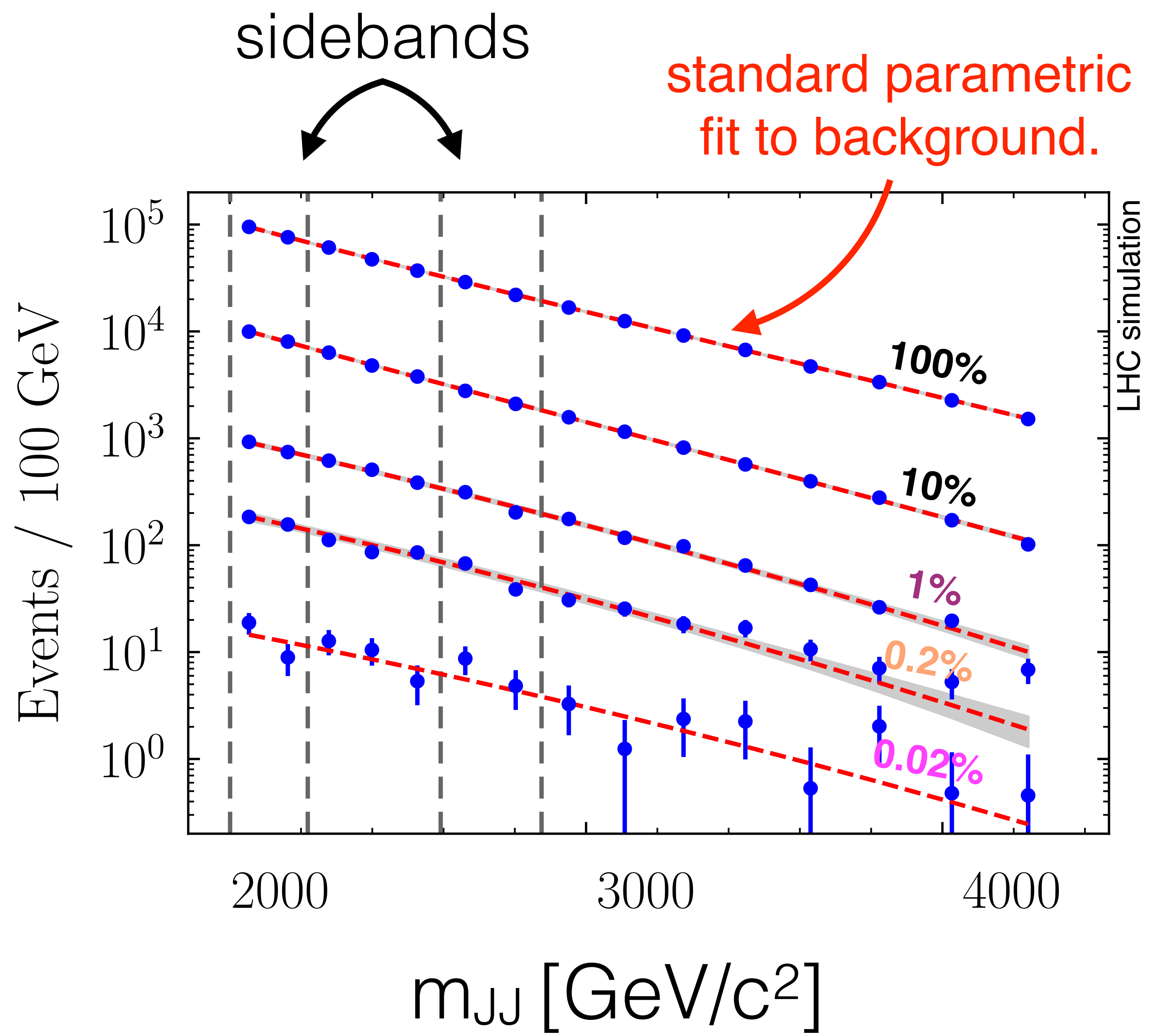
Example: two-jet search



Example: two-jet search



Example: two-jet search



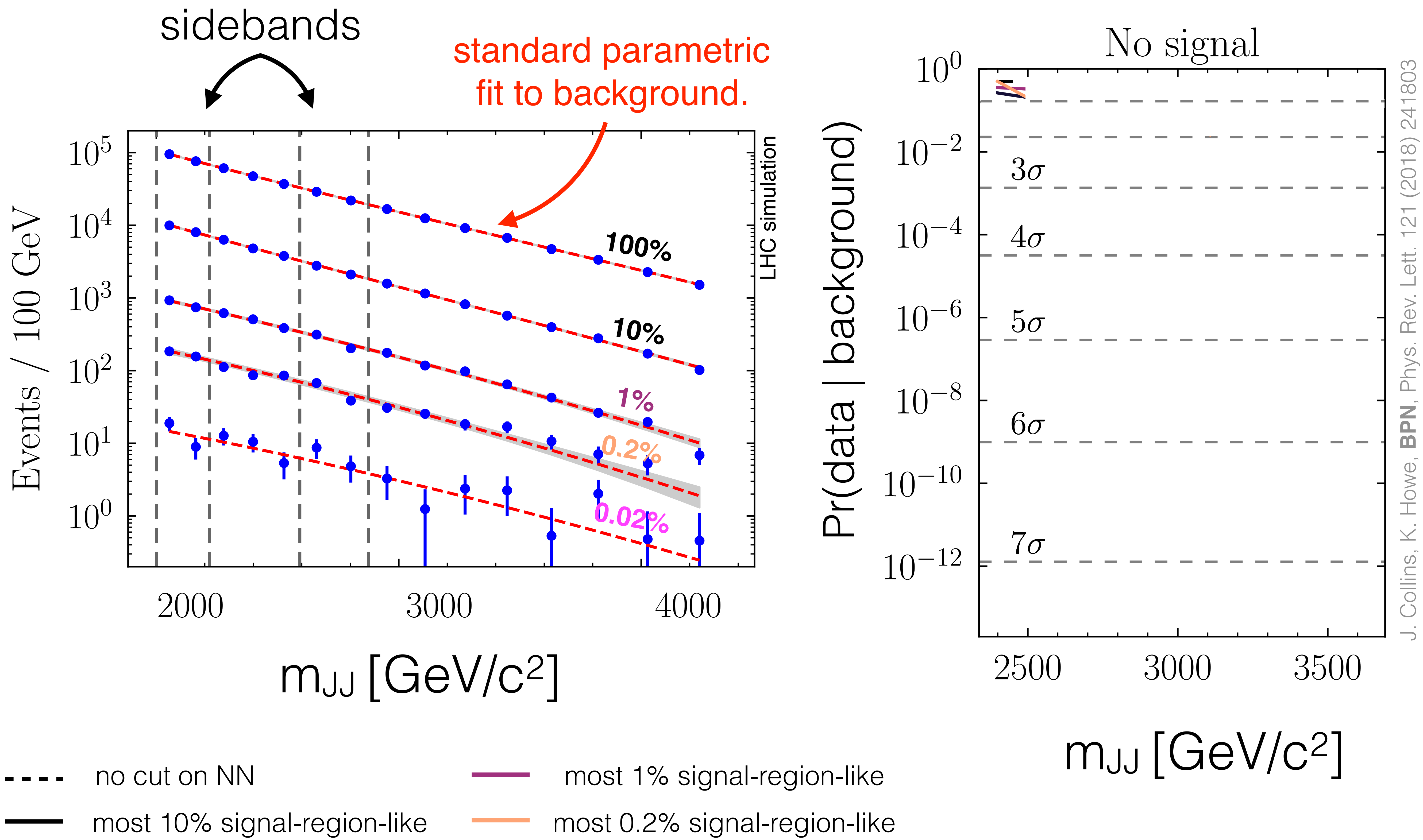
no cut on NN

most 10% signal-region-like

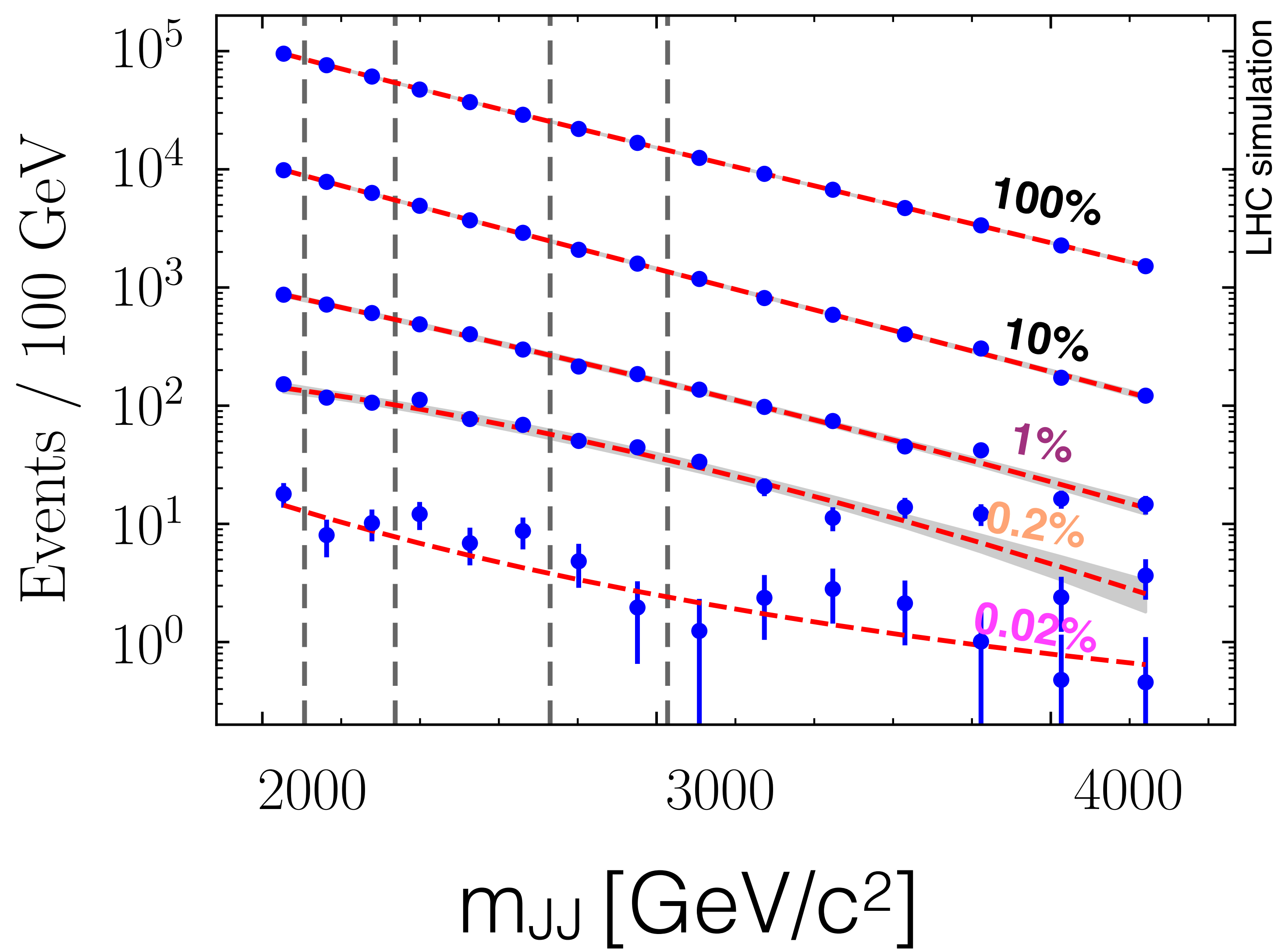
most 1% signal-region-like

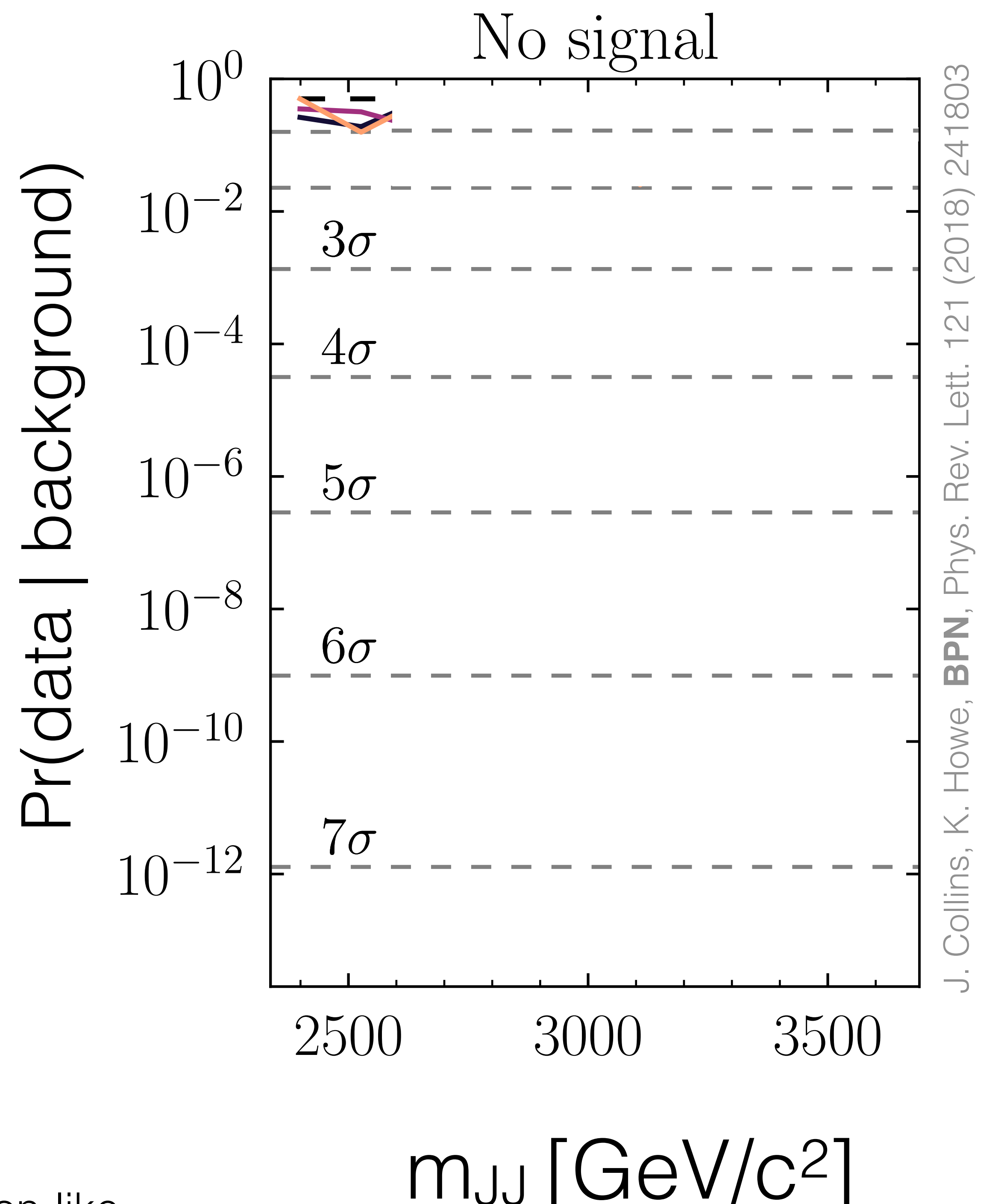
most 0.2% signal-region-like

Example: two-jet search



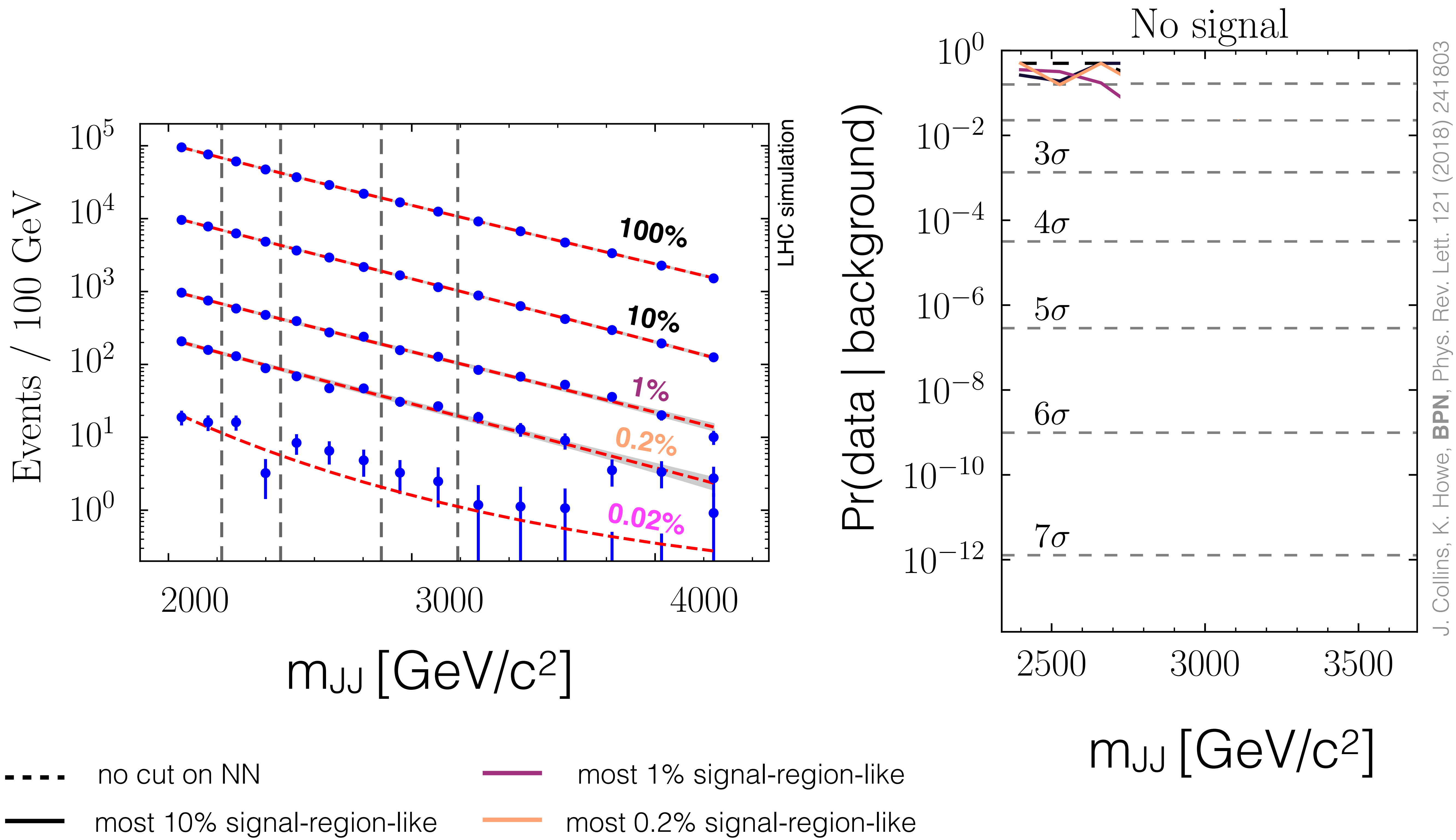
Example: two-jet search



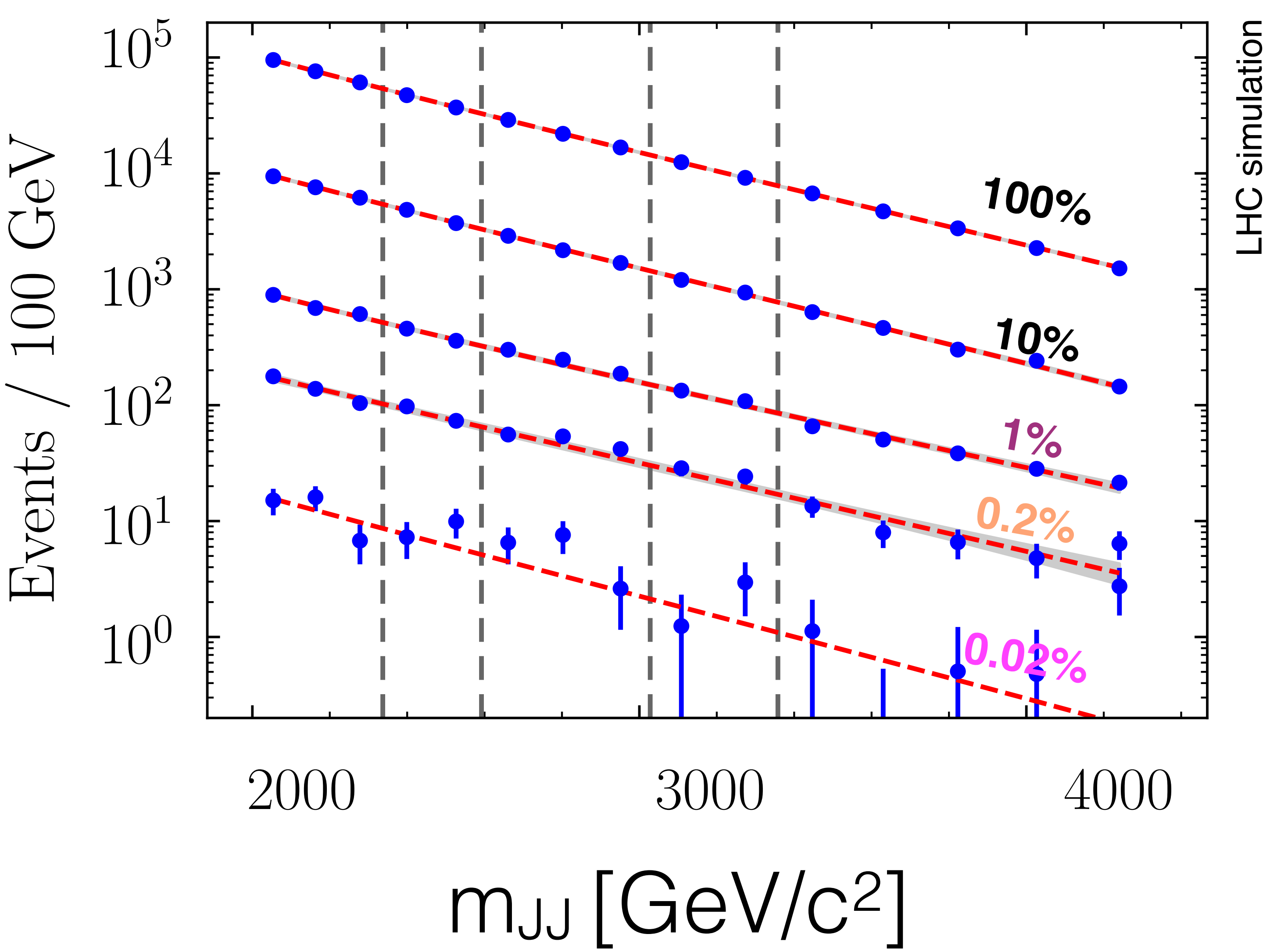


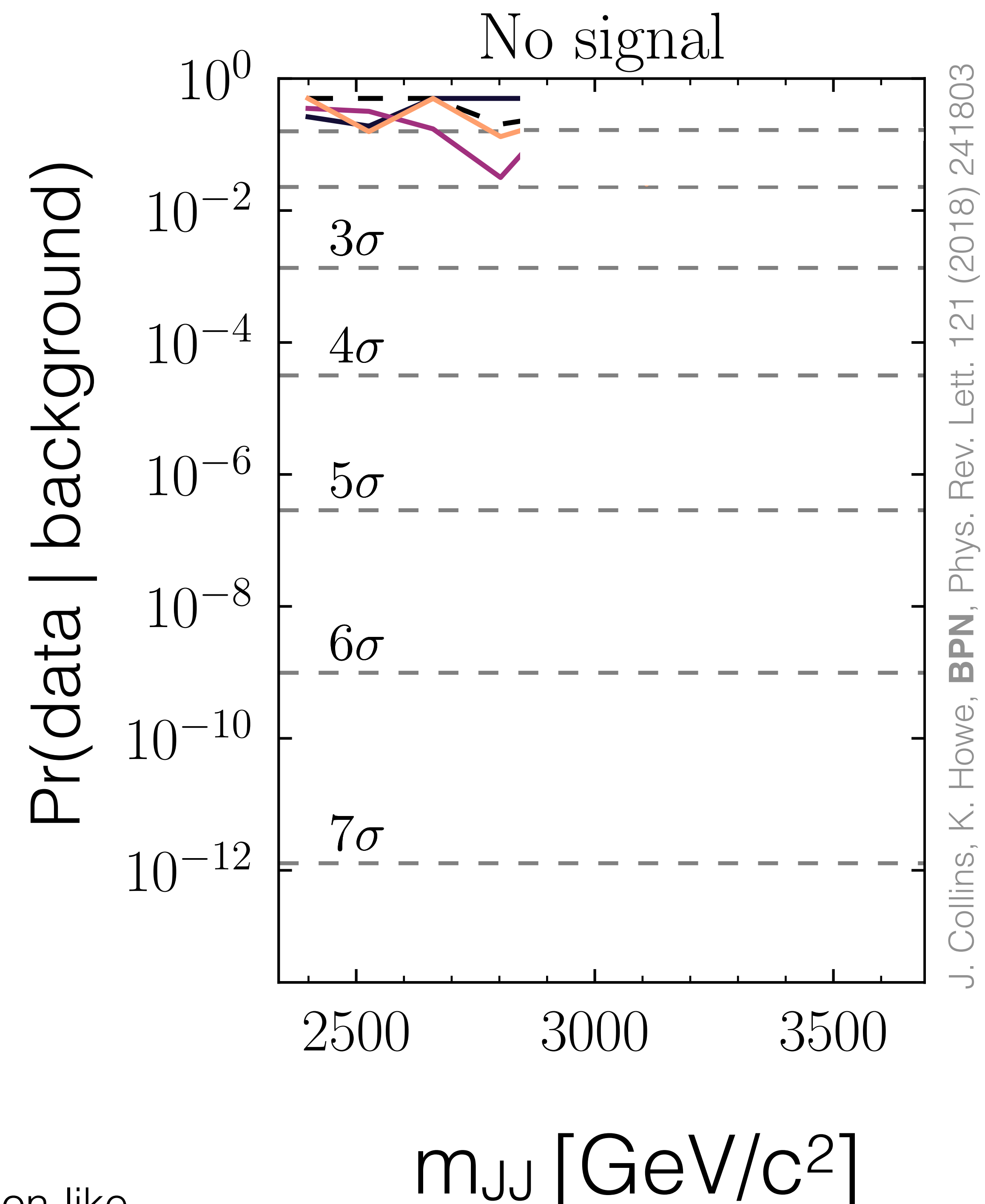
no cut on NN
most 10% signal-region-like
most 1% signal-region-like
most 0.2% signal-region-like

Example: two-jet search



Example: two-jet search





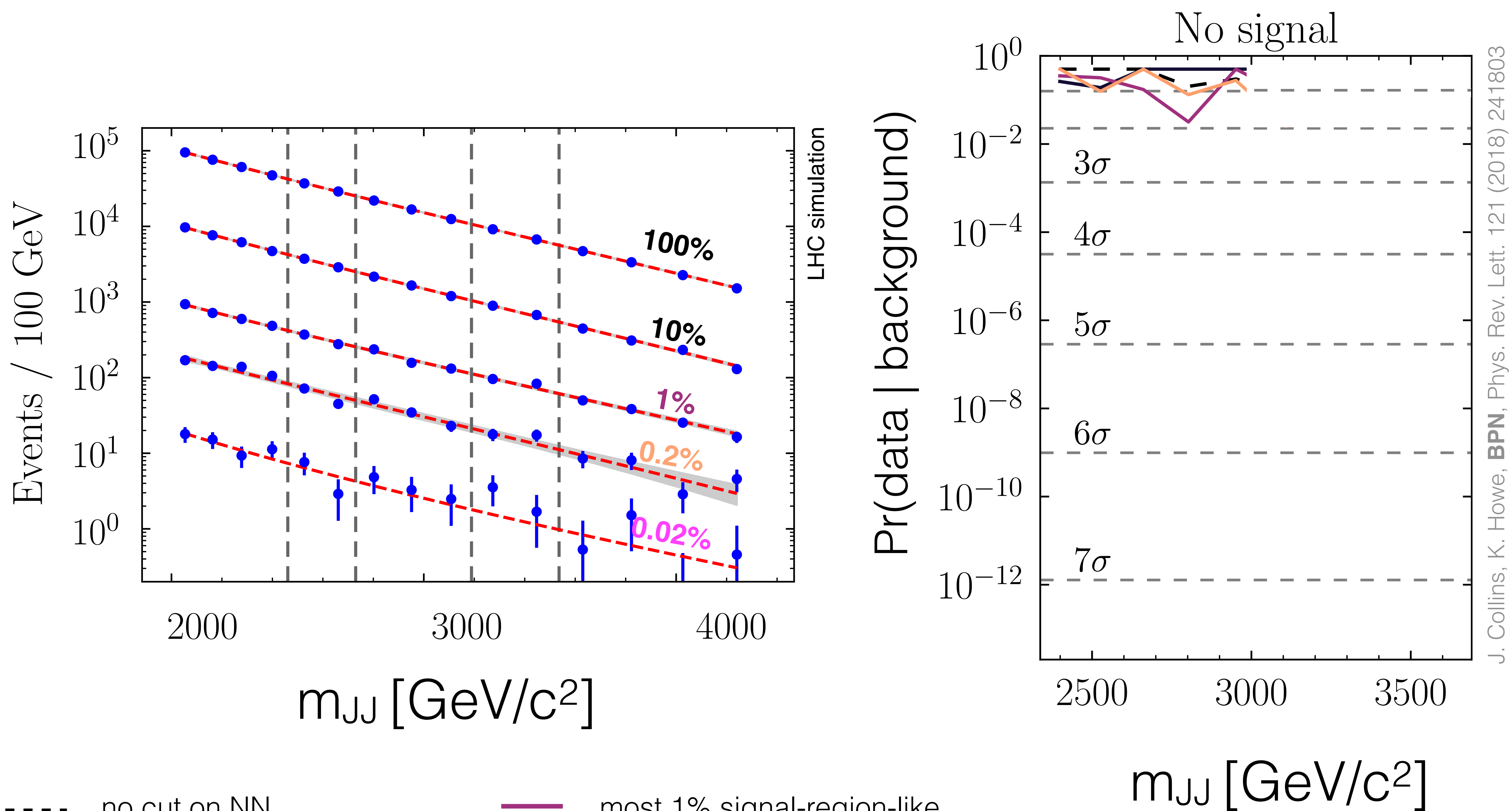
no cut on NN

most 10% signal-region-like

most 1% signal-region-like

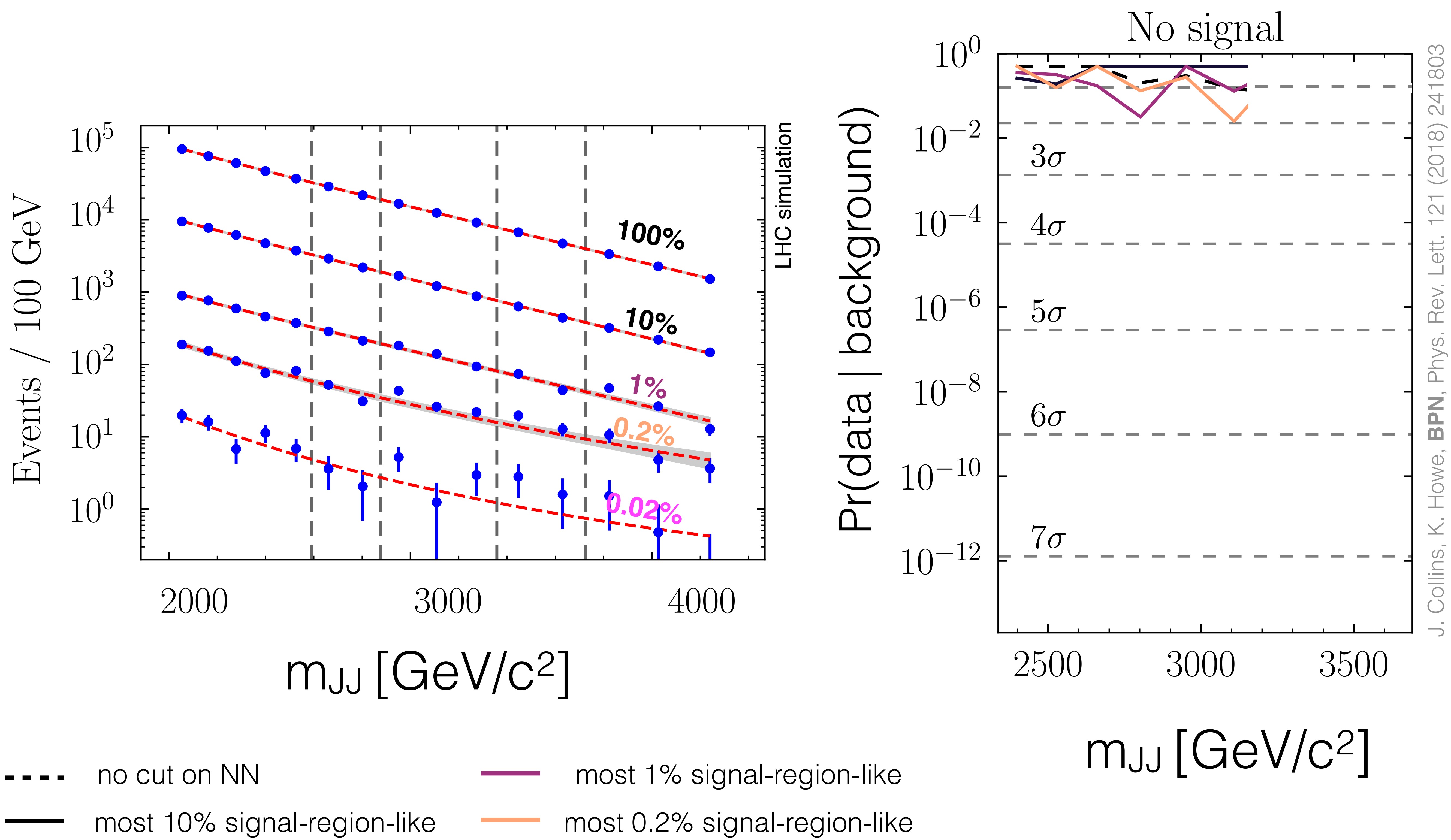
most 0.2% signal-region-like

Example: two-jet search

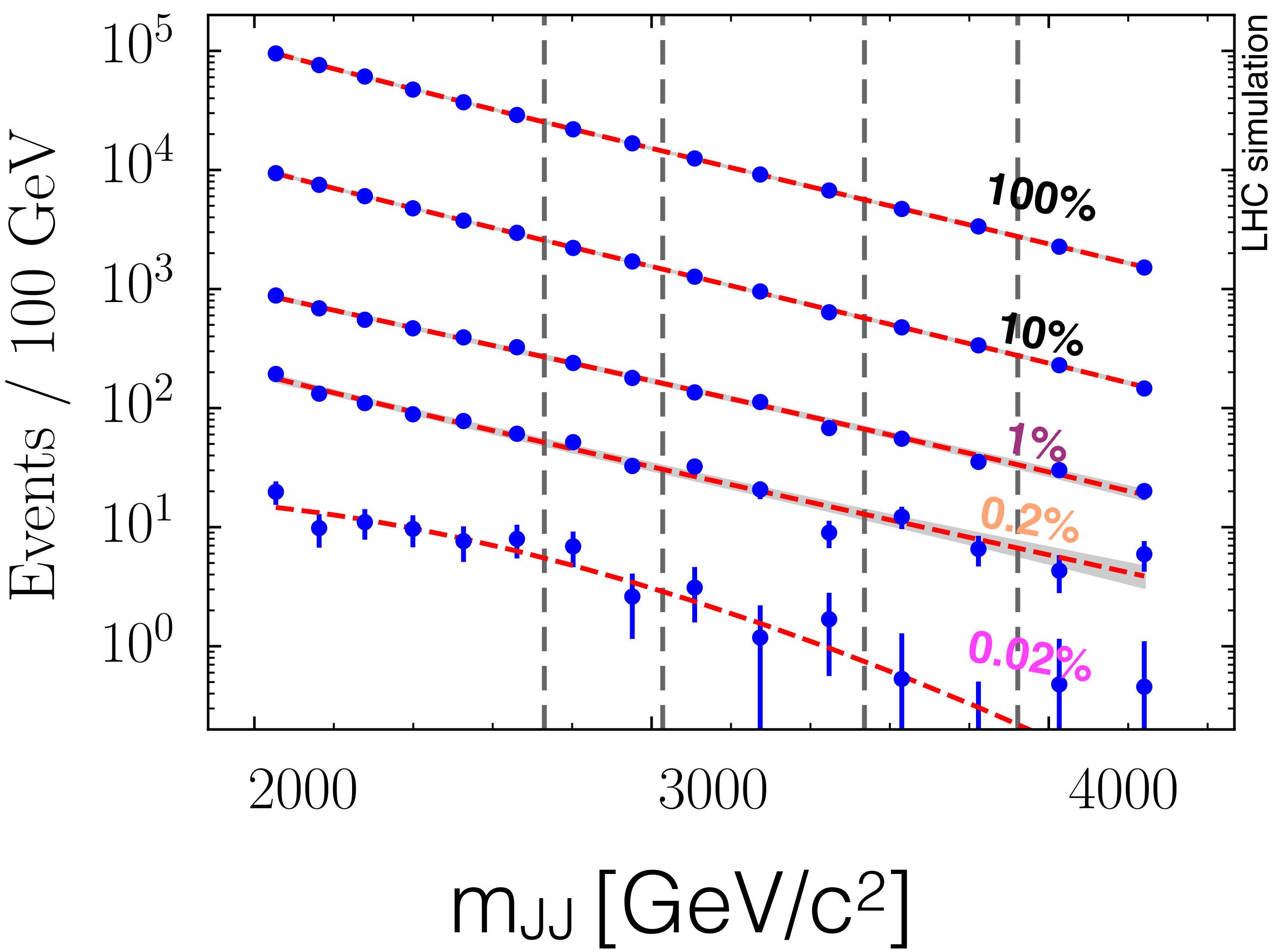


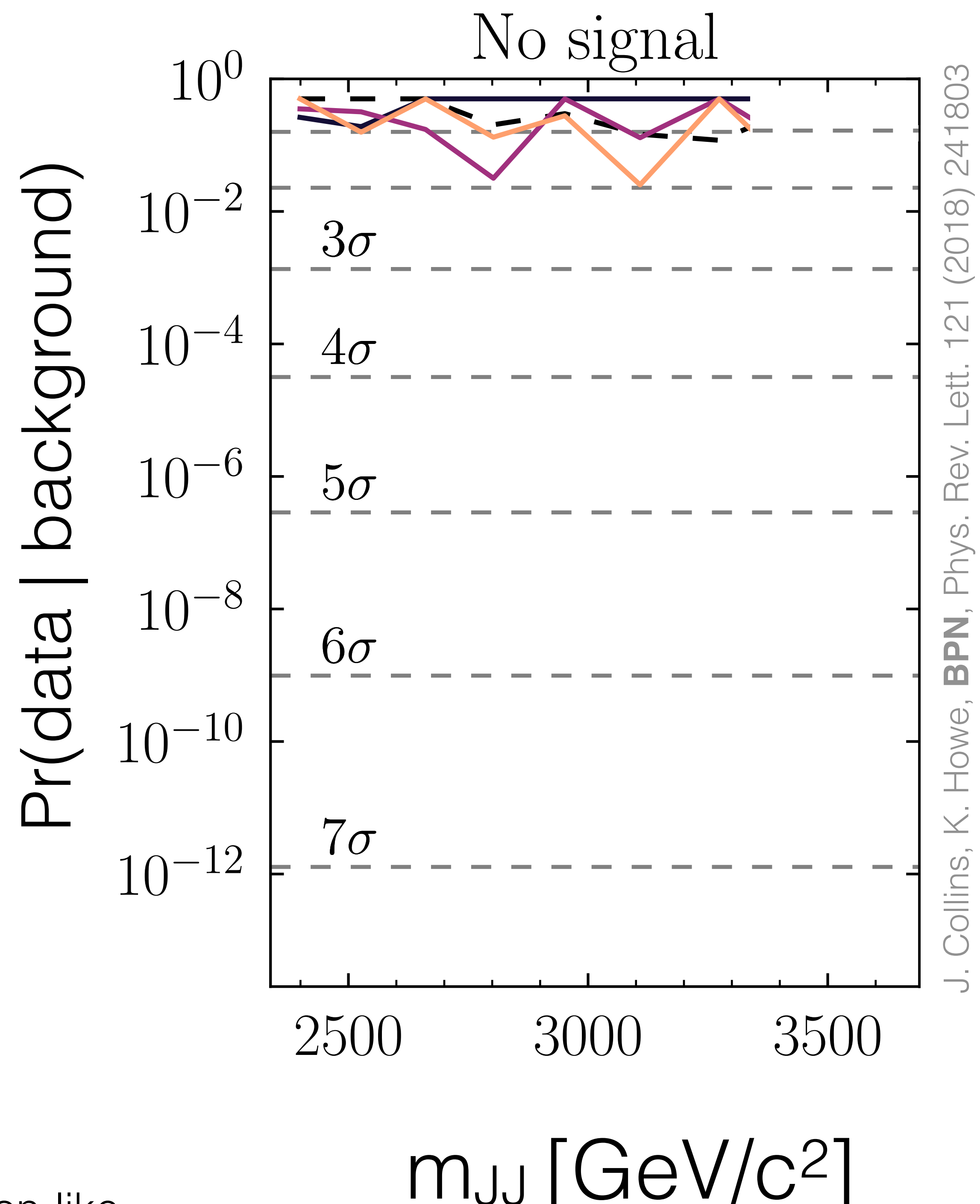
- - - no cut on NN
— most 10% signal-region-like
— most 1% signal-region-like
— most 0.2% signal-region-like

Example: two-jet search



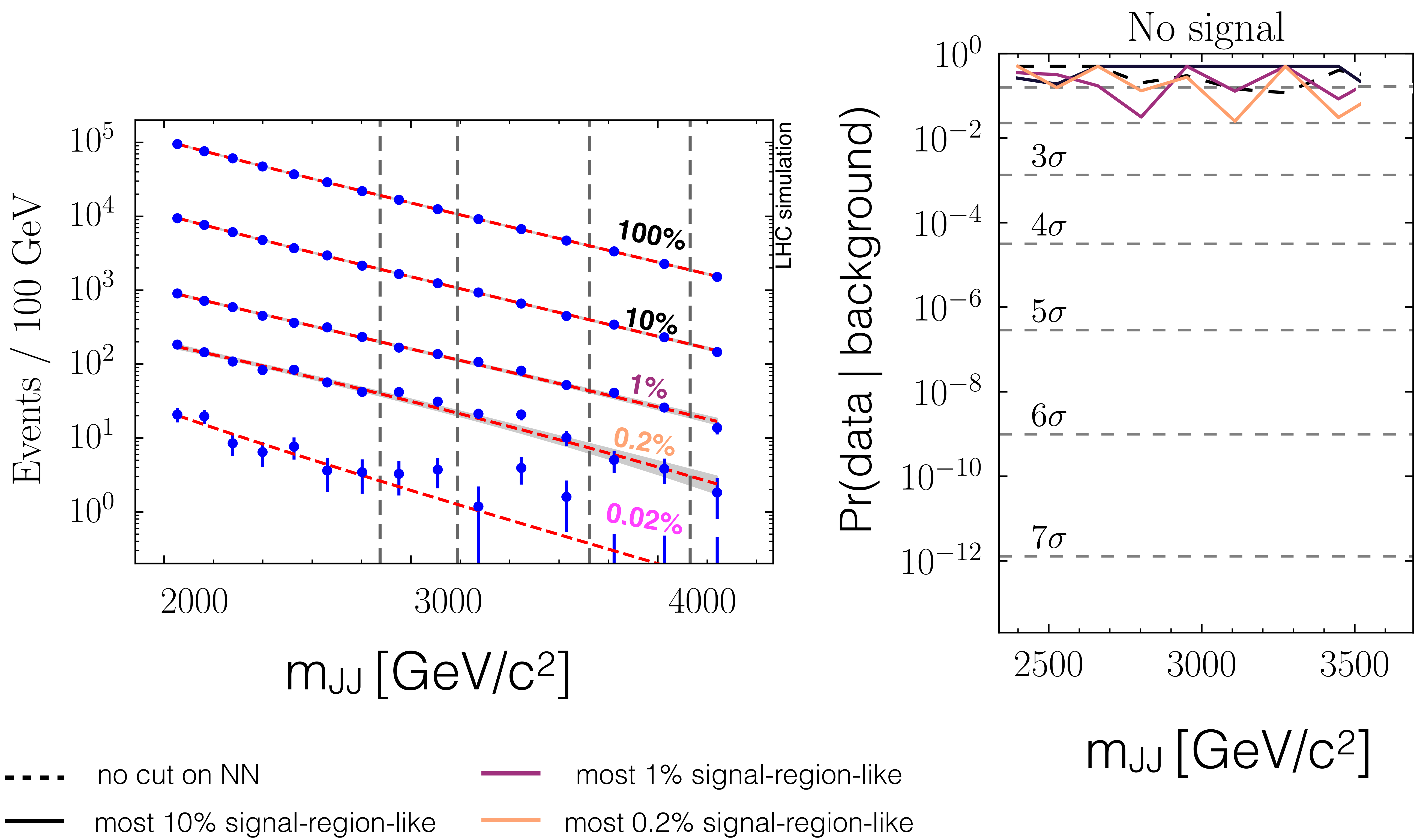
Example: two-jet search





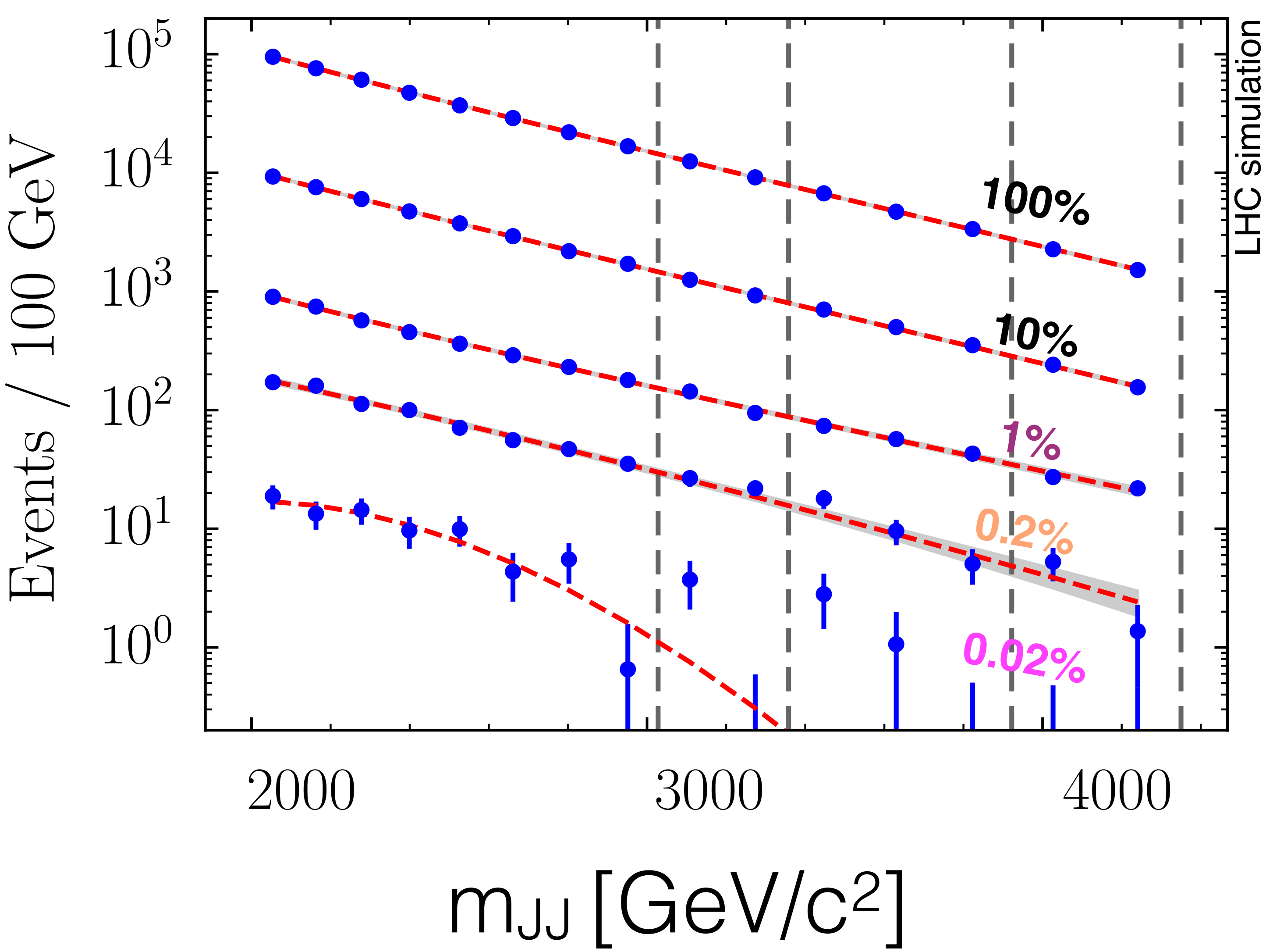
- - - - no cut on NN
— most 10% signal-region-like
— most 1% signal-region-like
— most 0.2% signal-region-like

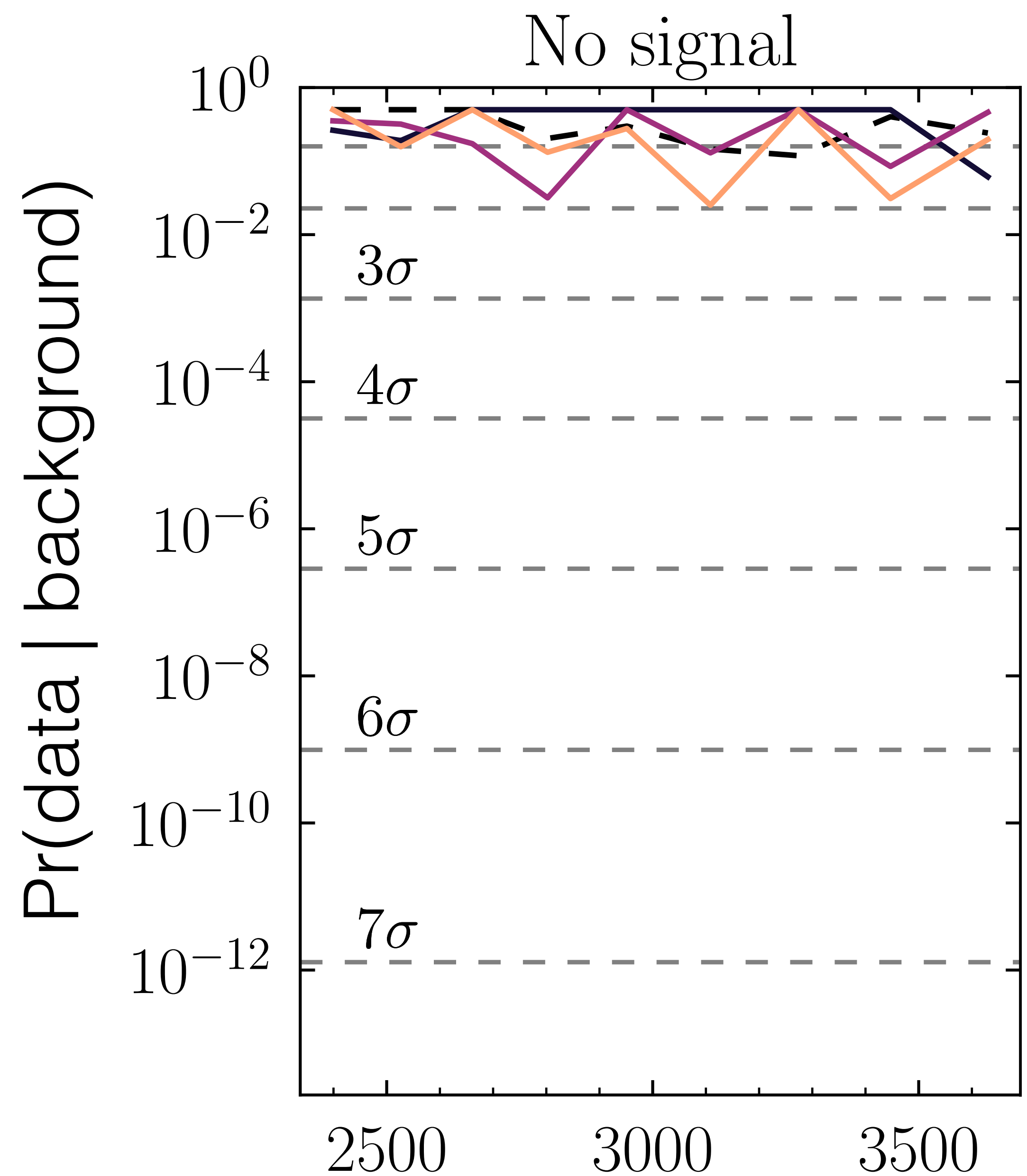
Example: two-jet search



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

Example: two-jet search





m_{JJ} [GeV/c^2]

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

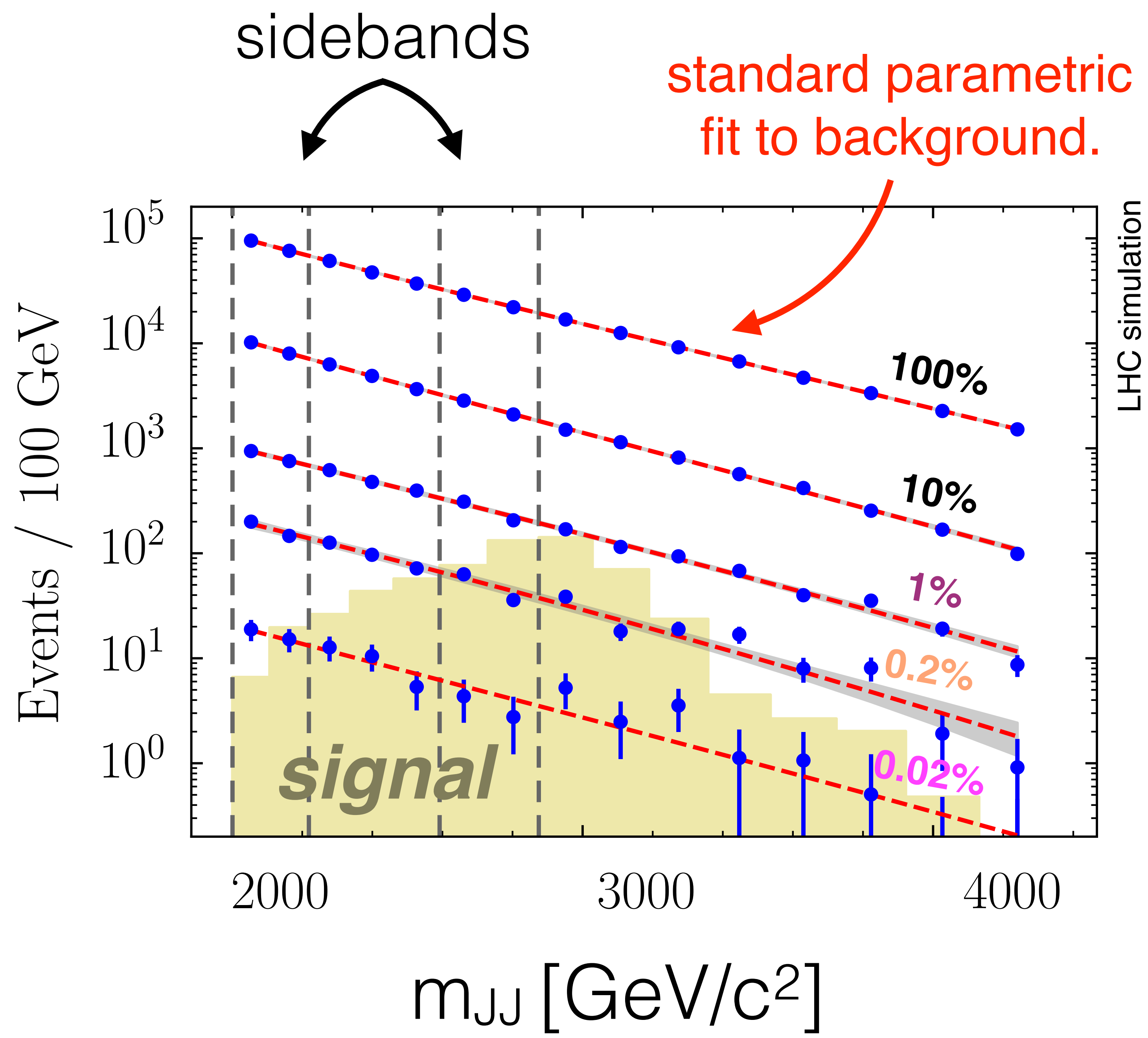
no cut on NN

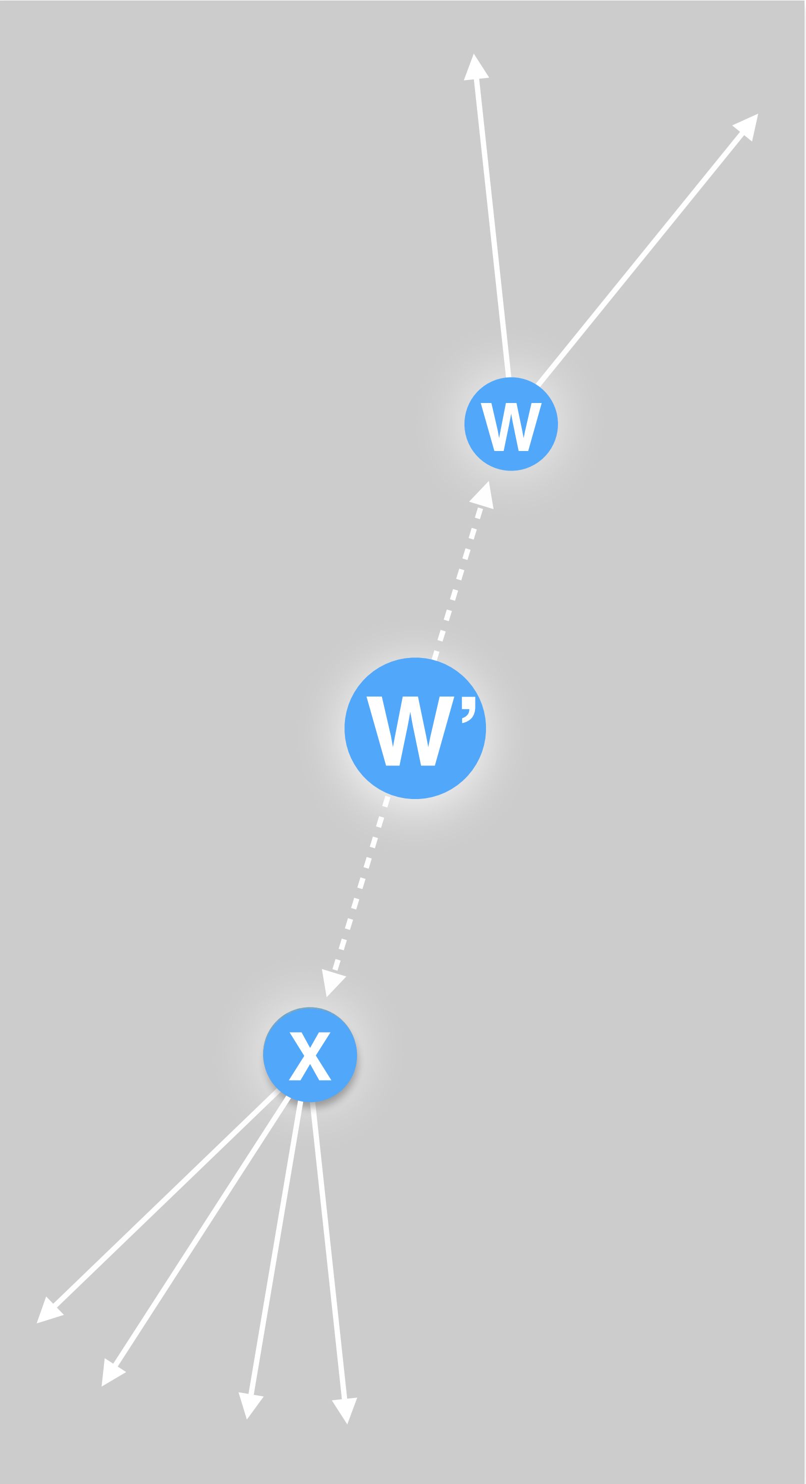
most 10% signal-region-like

most 1% signal-region-like

most 0.2% signal-region-like

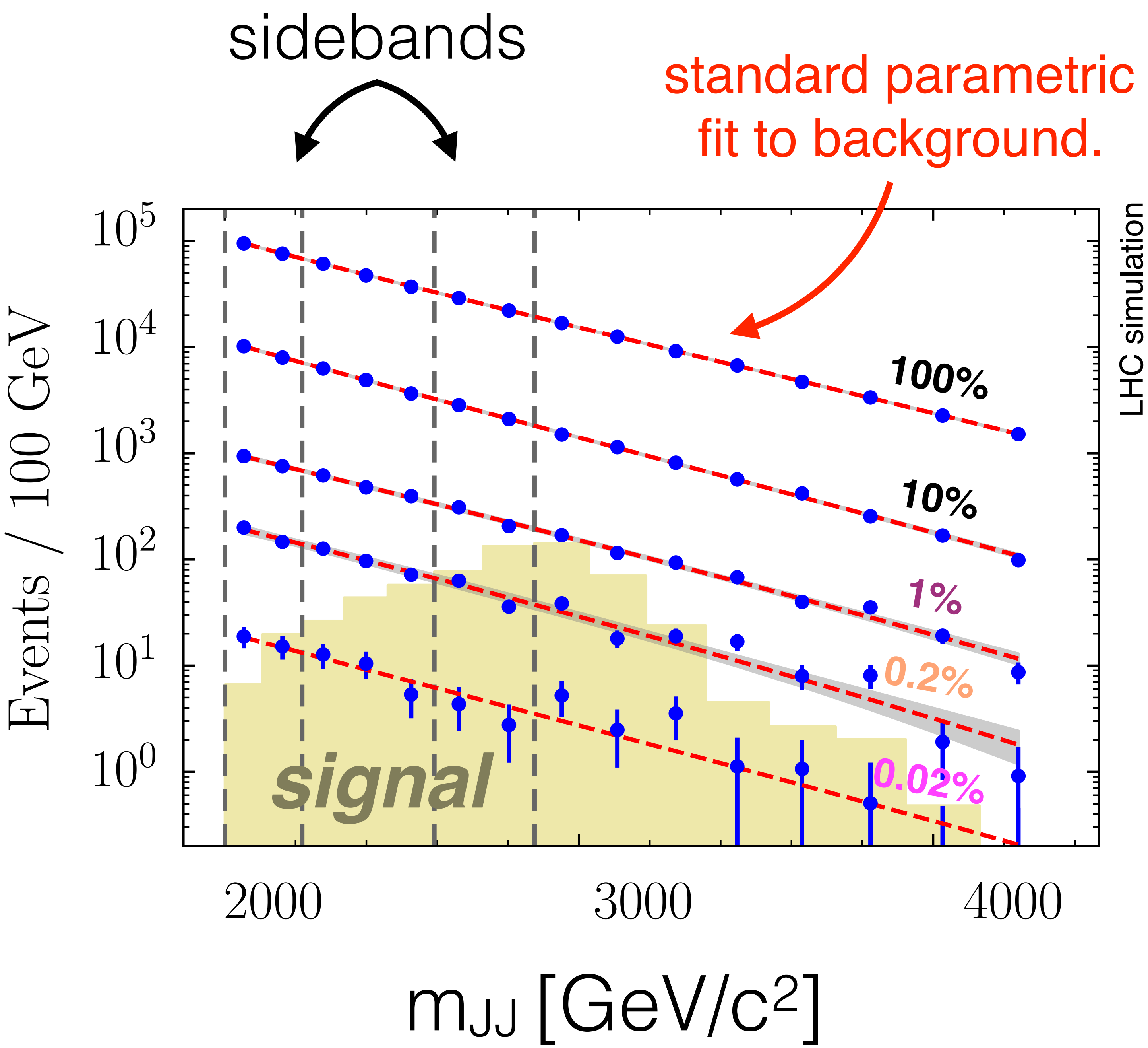
…and when there is a signal?

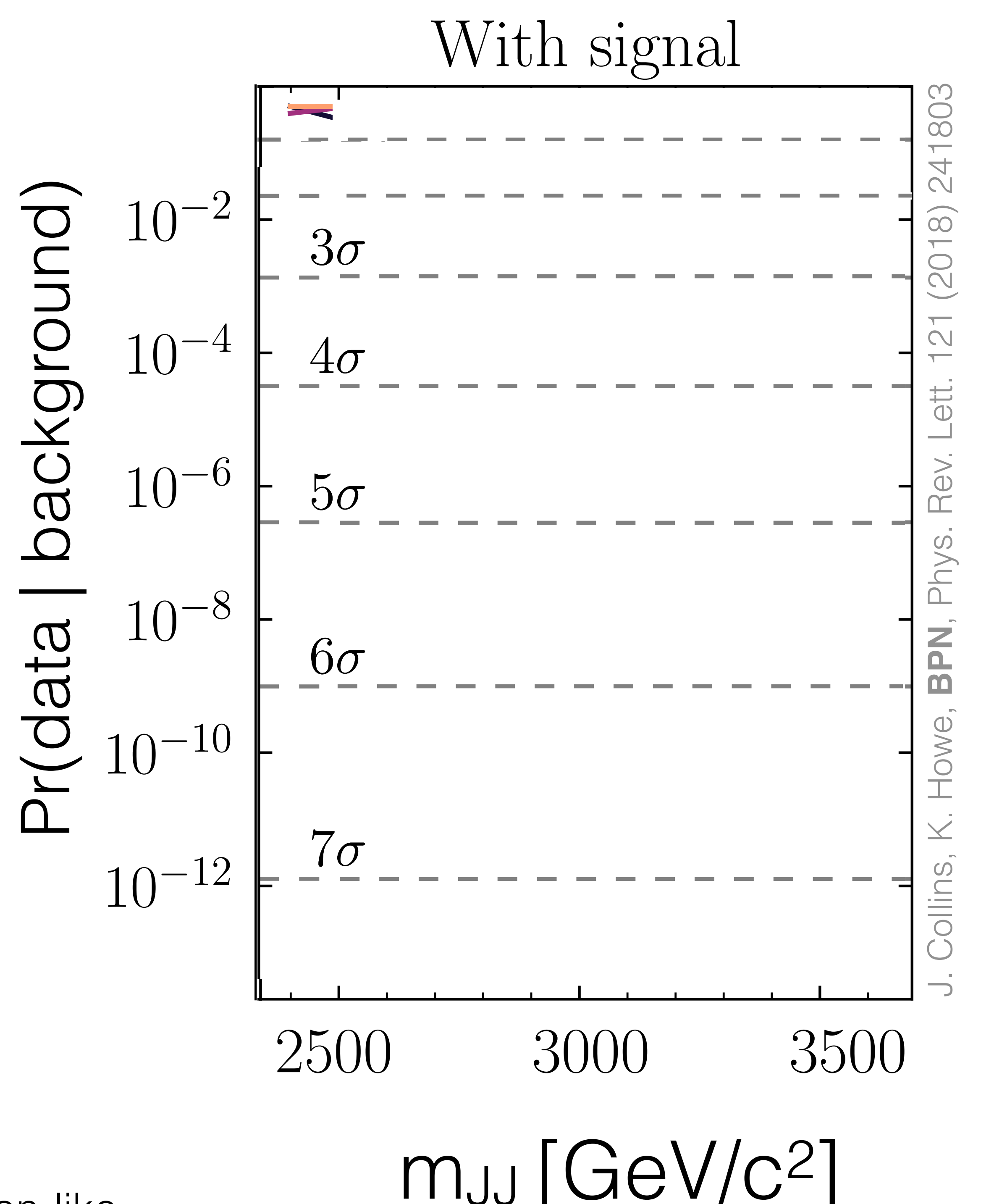




- - - - no cut on NN
—— most 10% signal-region-like
—— most 1% signal-region-like
—— most 0.2% signal-region-like

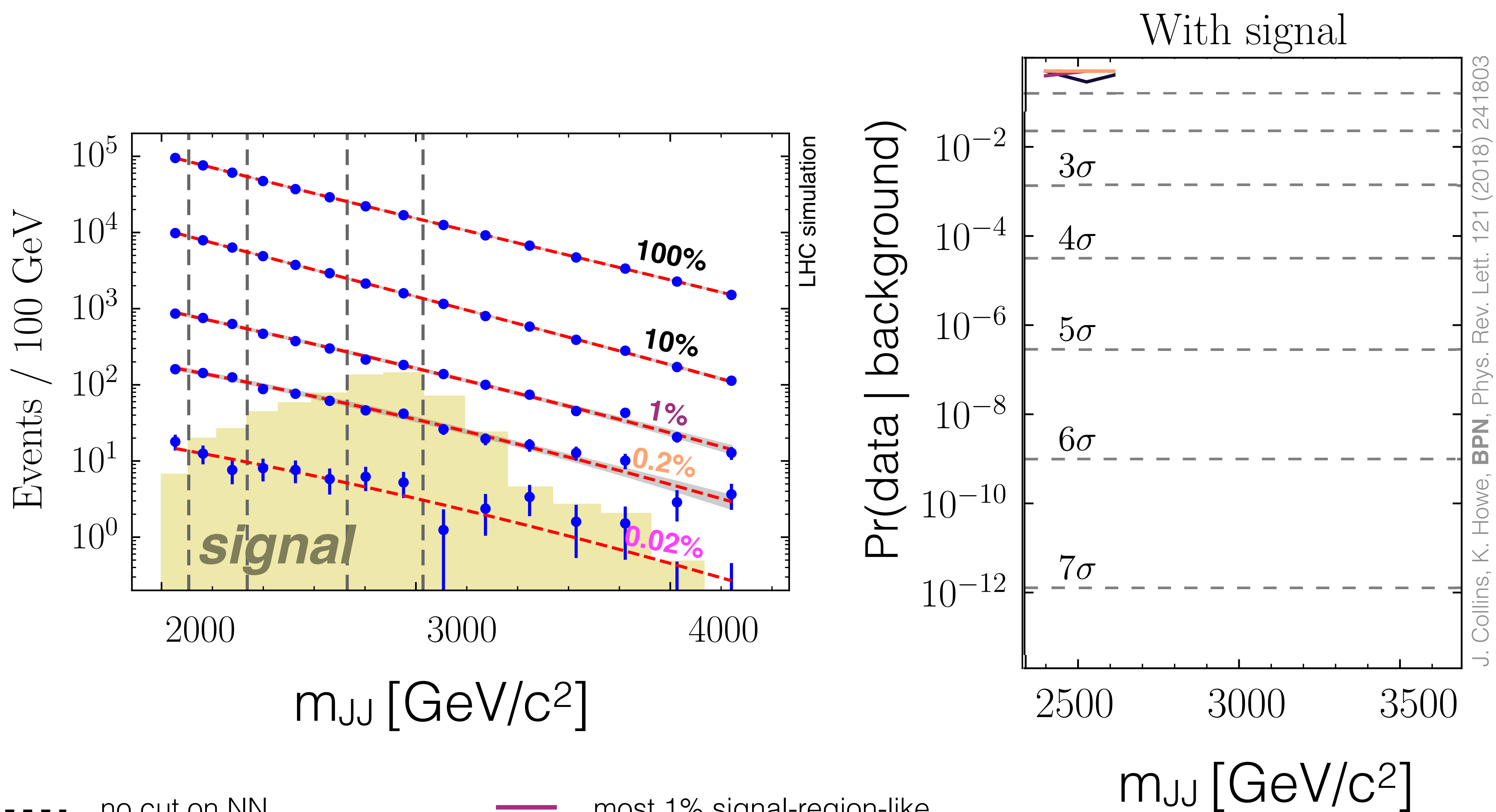
…and when there is a signal?





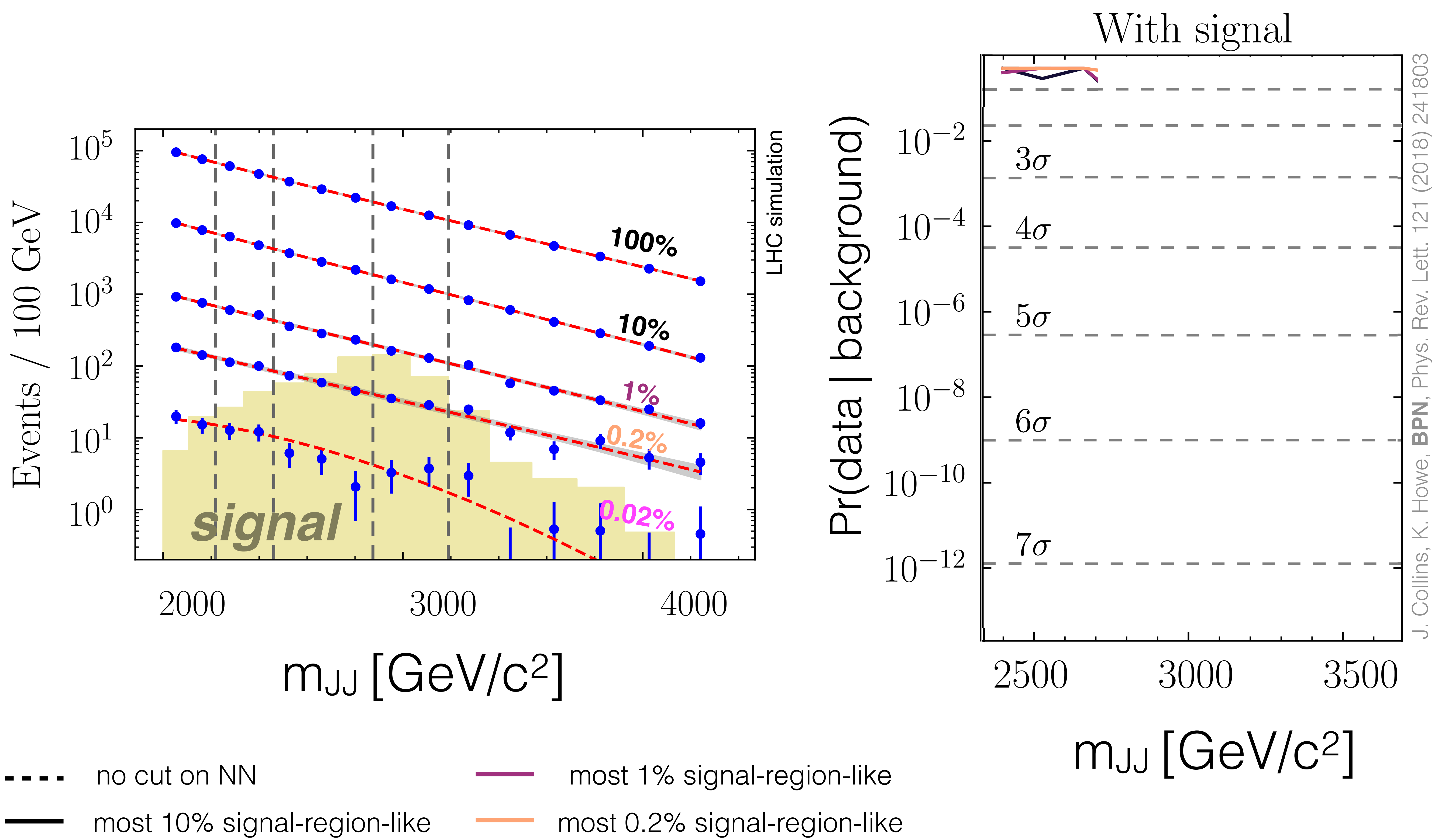
no cut on NN
most 10% signal-region-like
most 1% signal-region-like
most 0.2% signal-region-like

…and when there is a signal?

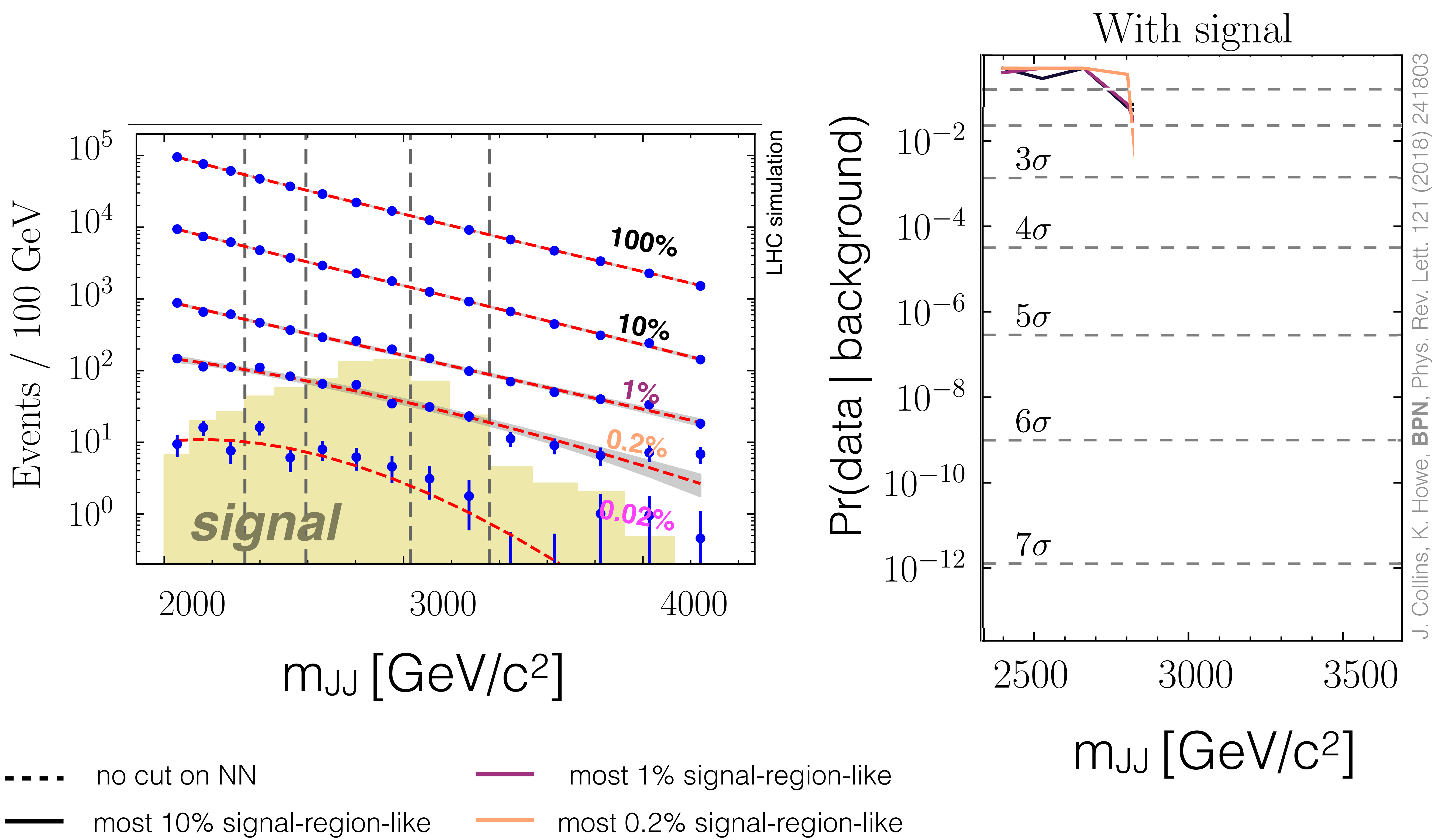


- - - no cut on NN
— most 10% signal-region-like
— most 1% signal-region-like
— most 0.2% signal-region-like

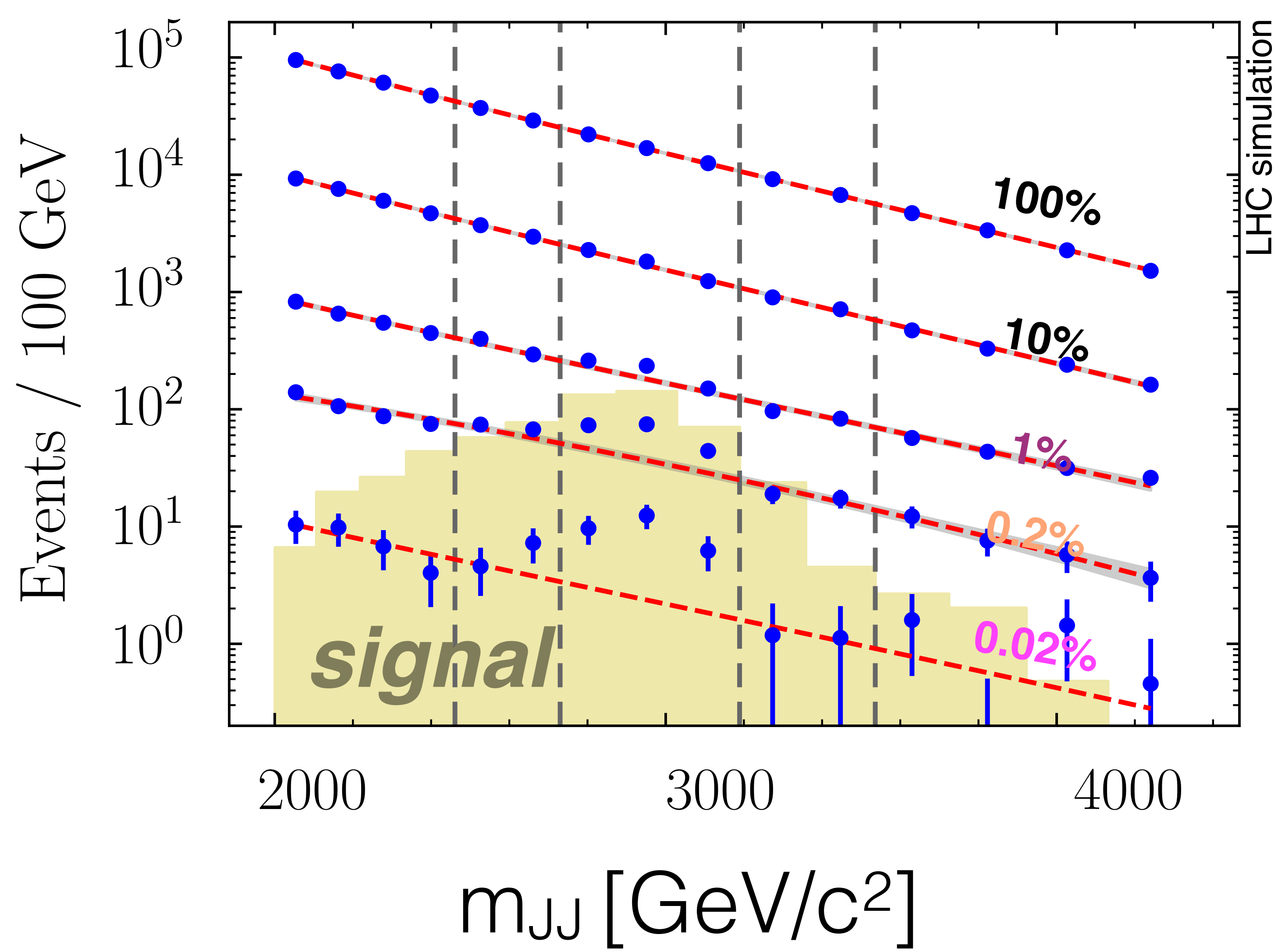
…and when there is a signal?

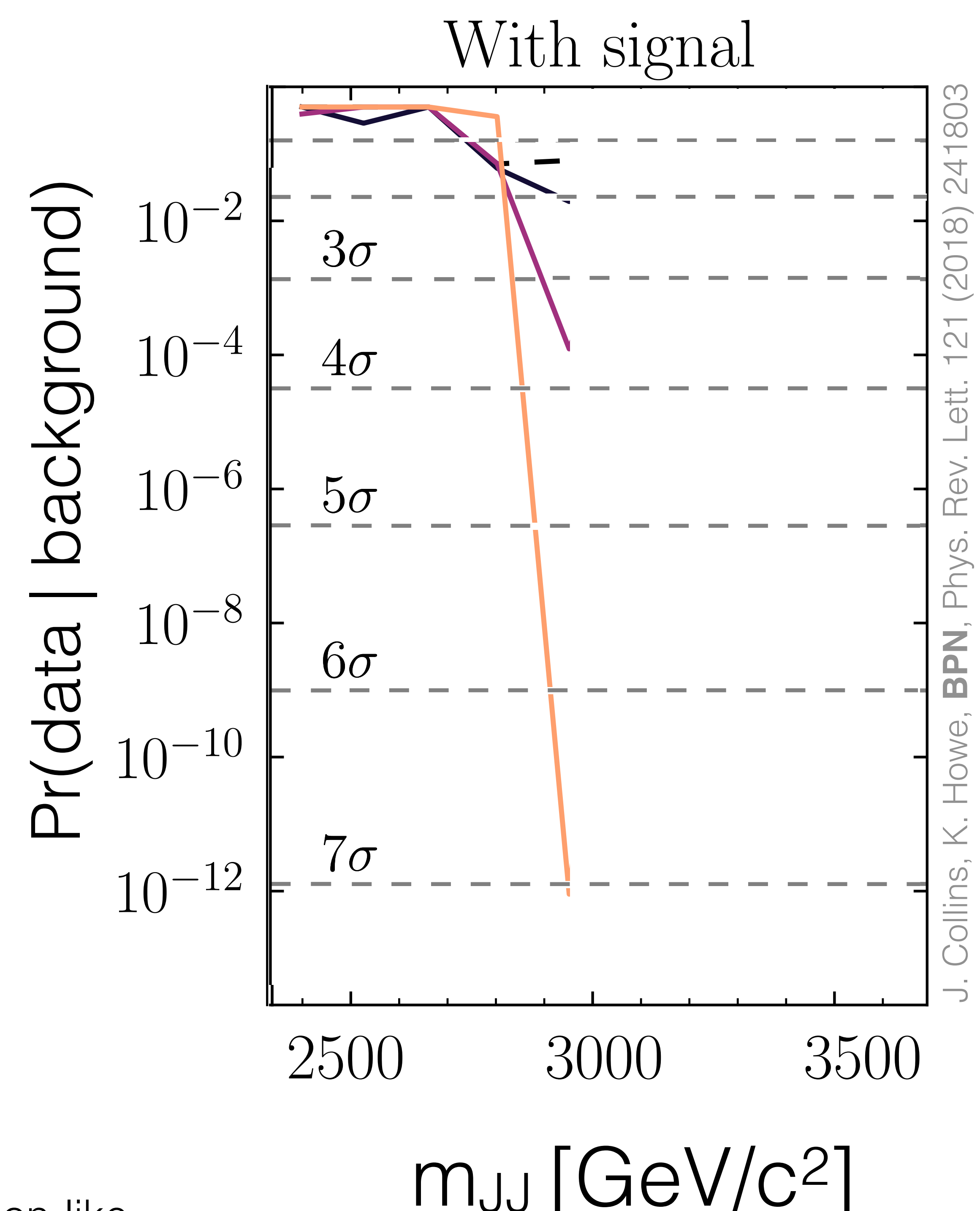


…and when there is a signal?



…and when there is a signal?





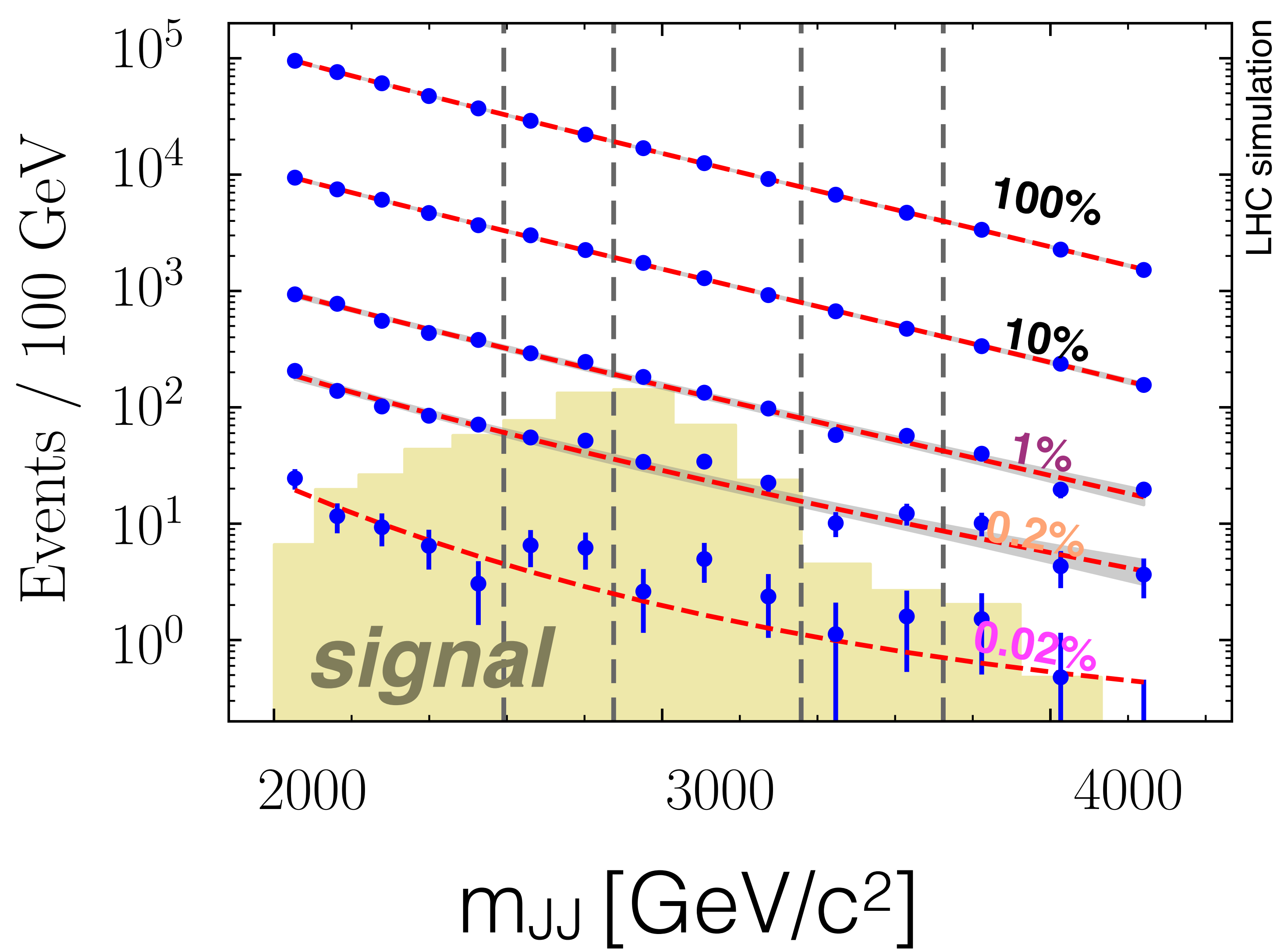
no cut on NN

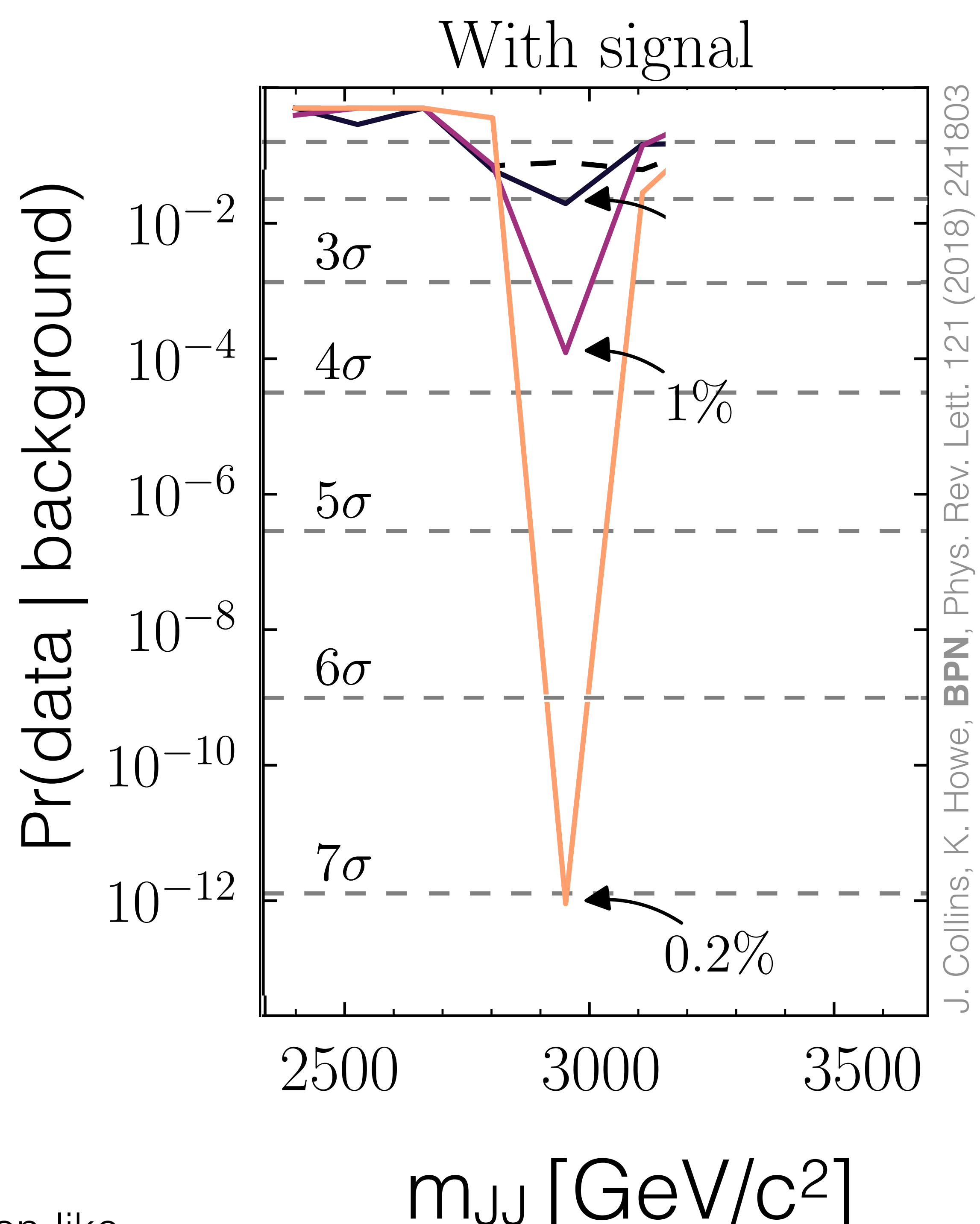
most 10% signal-region-like

most 1% signal-region-like

most 0.2% signal-region-like

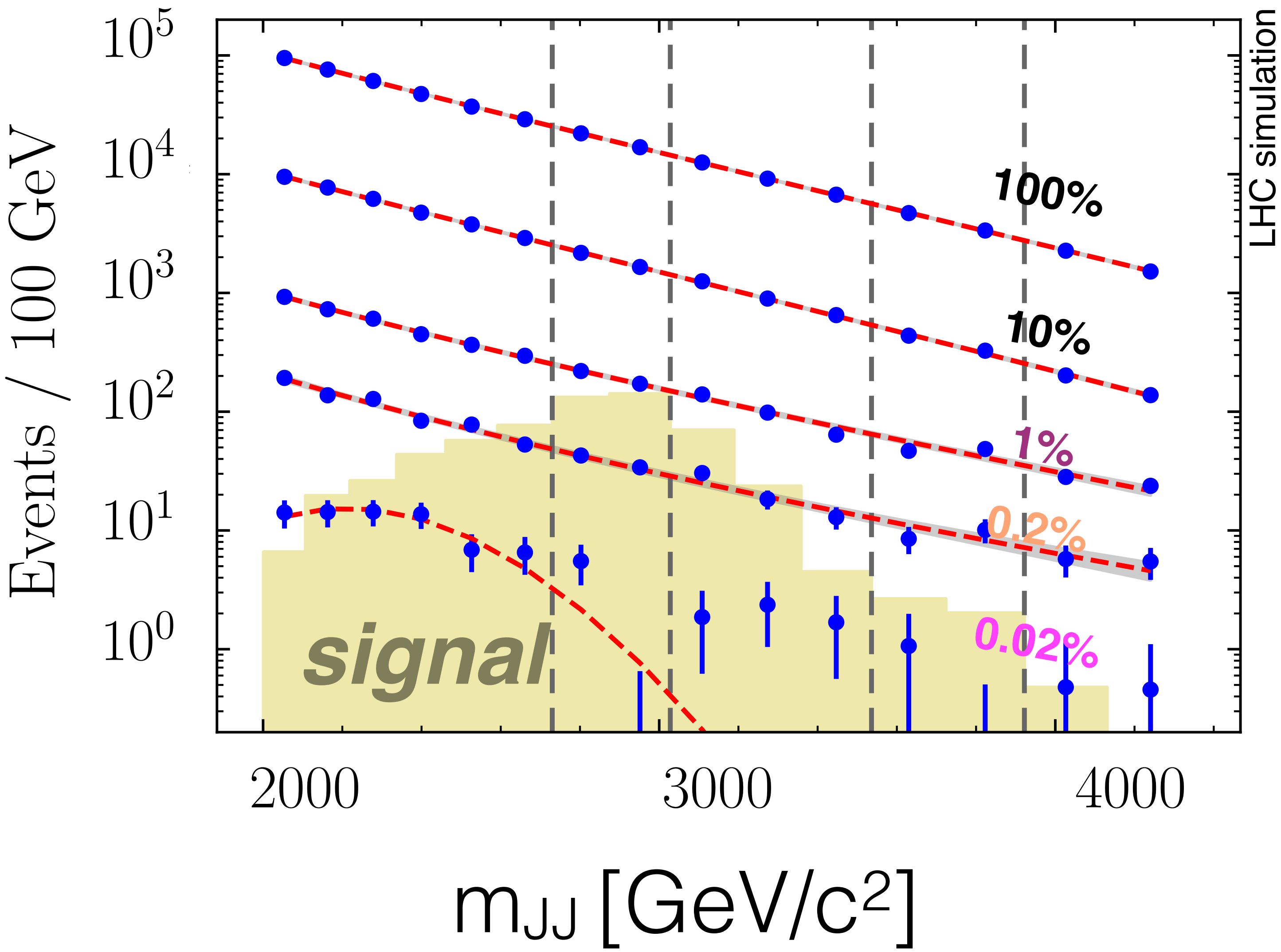
…and when there is a signal?

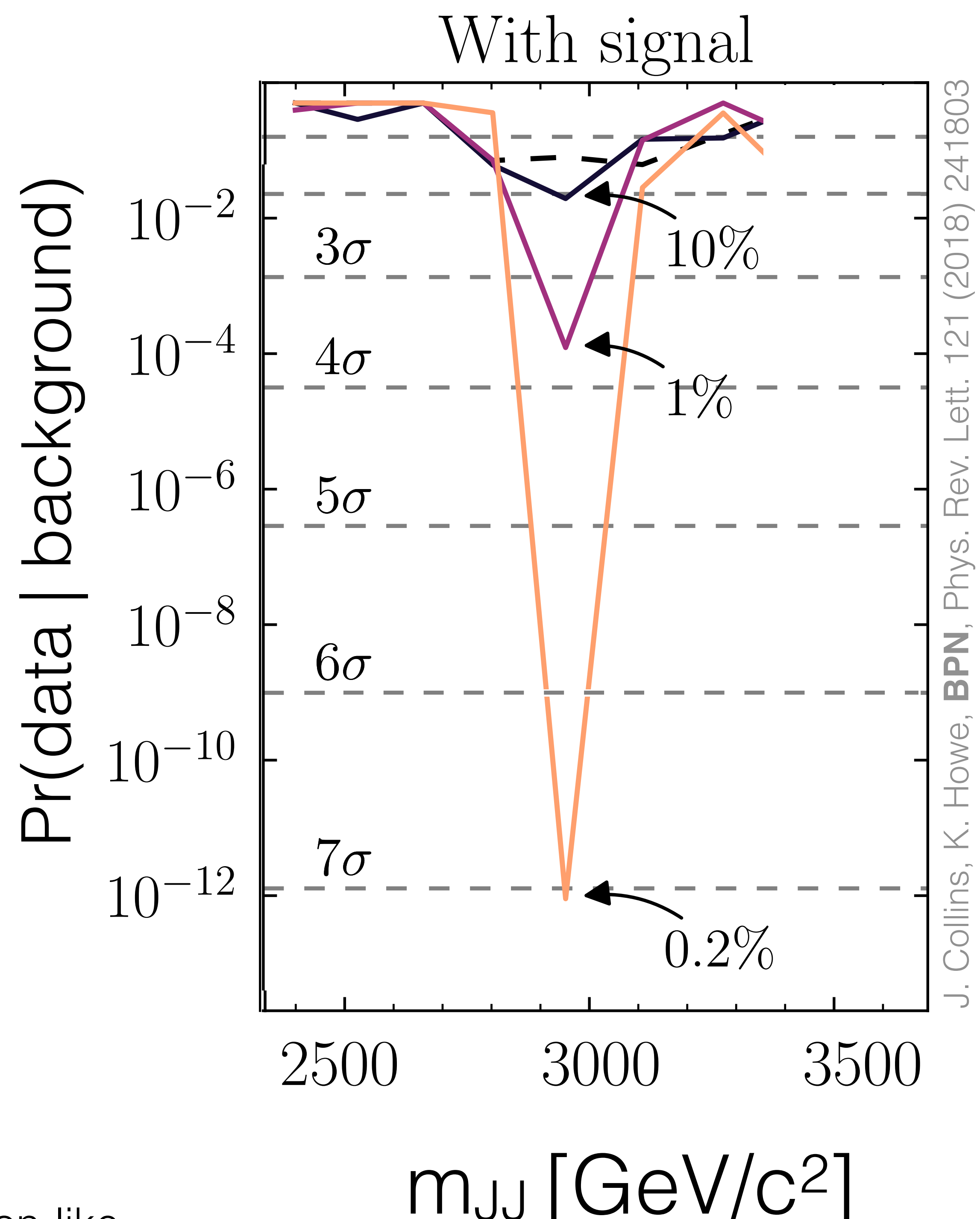




no cut on NN
most 10% signal-region-like
most 1% signal-region-like
most 0.2% signal-region-like

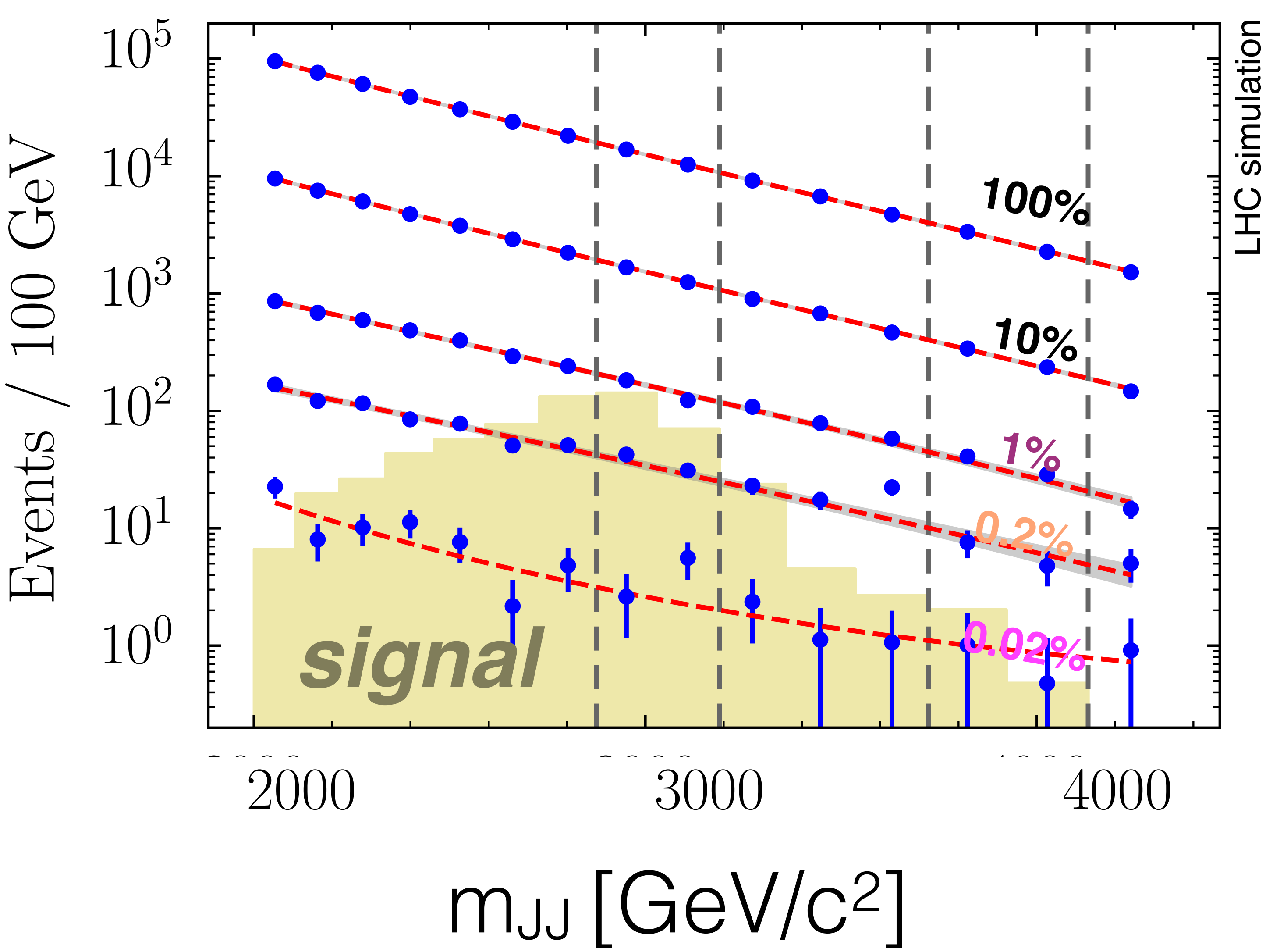
…and when there is a signal?

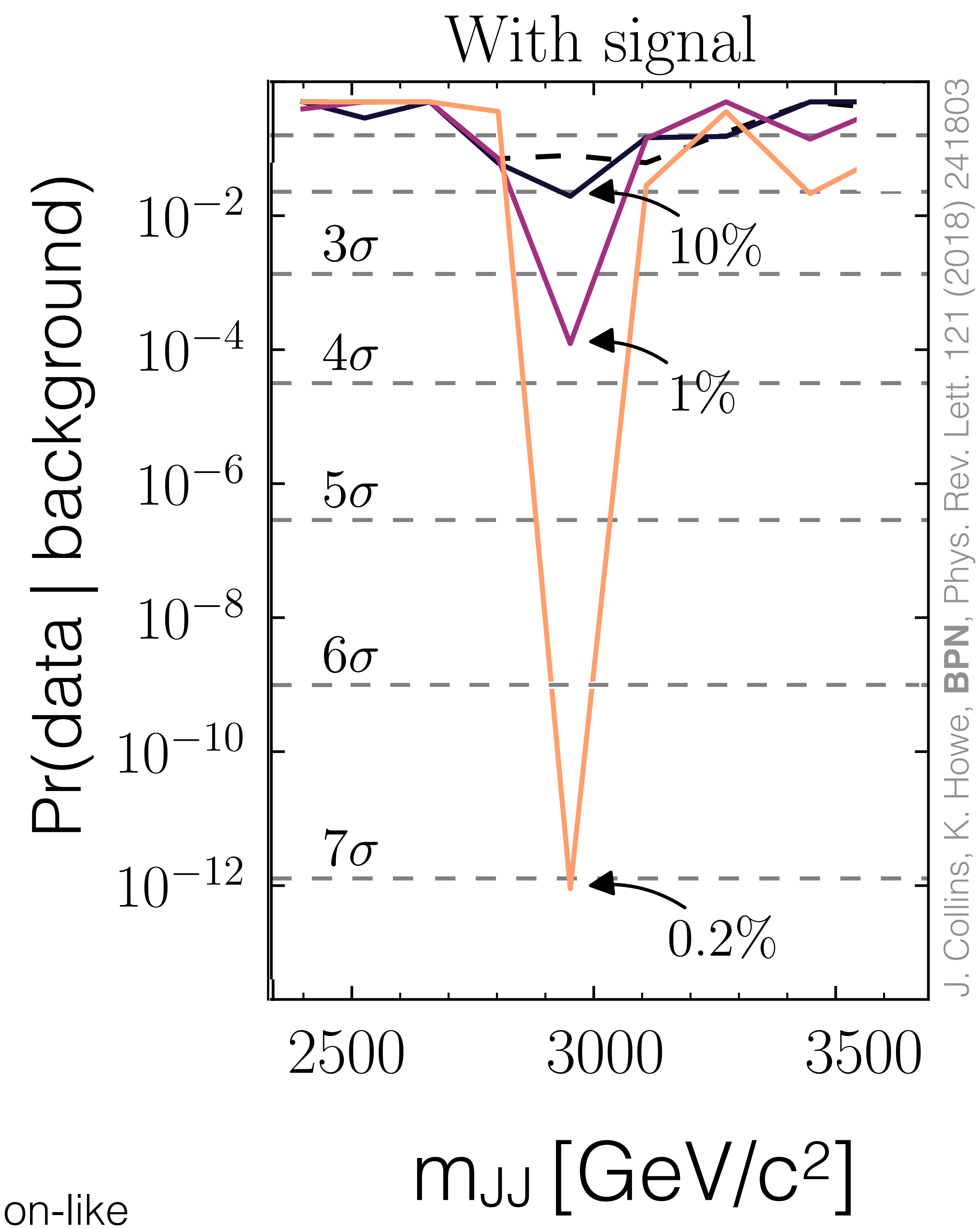




- - - no cut on NN
— most 10% signal-region-like
— most 1% signal-region-like
— most 0.2% signal-region-like

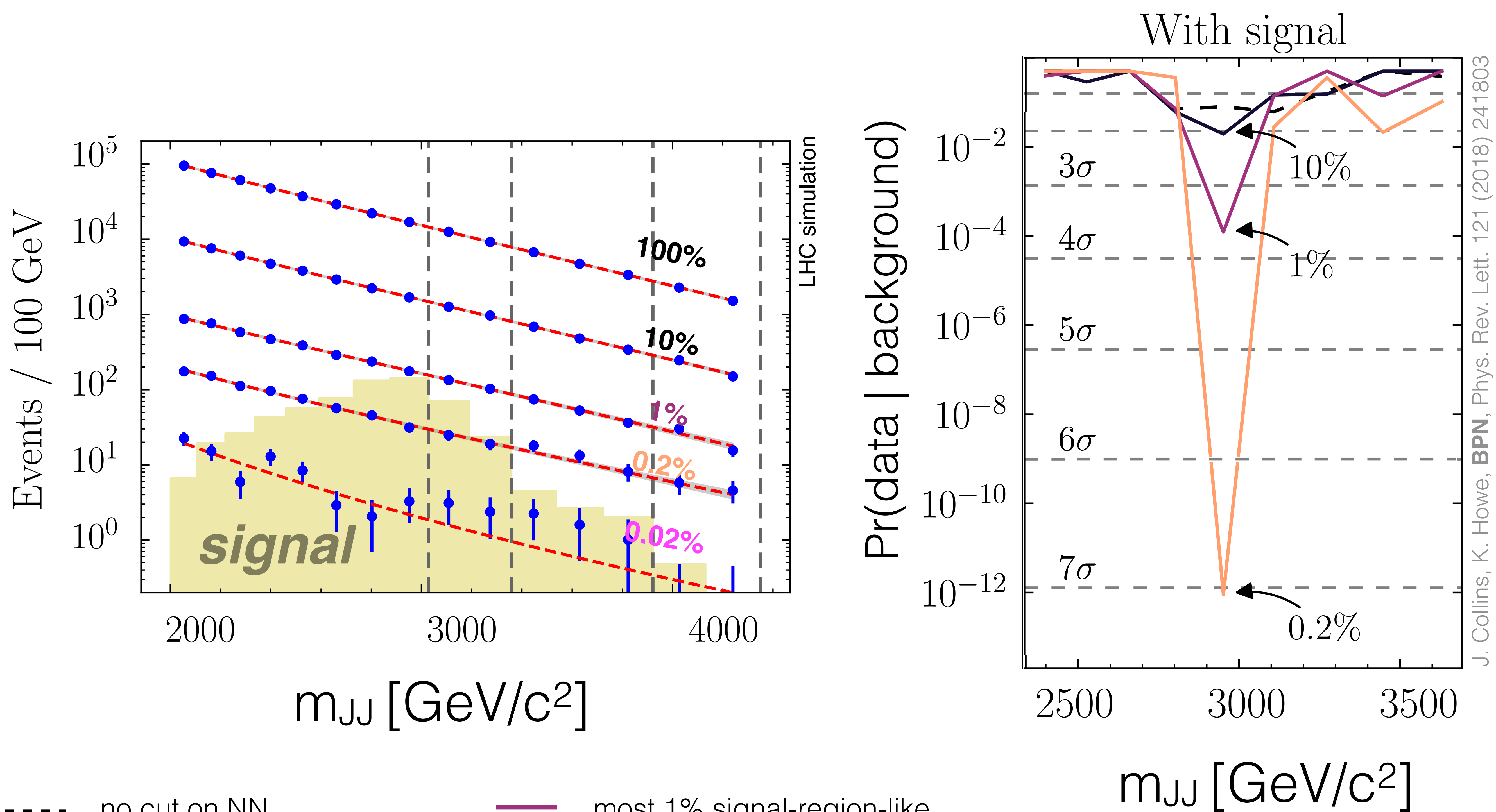
…and when there is a signal?





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?

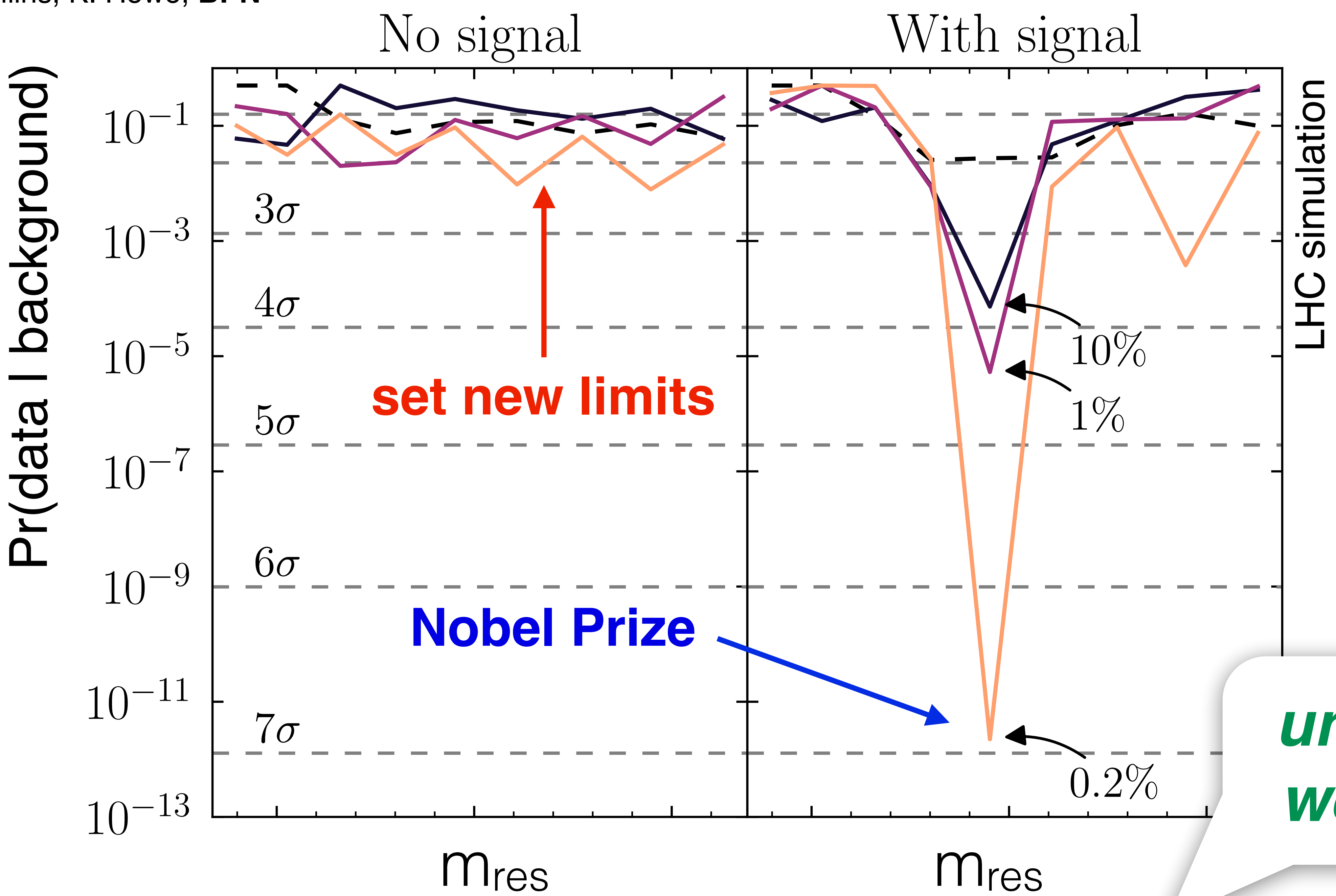


- - - - no cut on NN
— most 10% signal-region-like
— most 1% signal-region-like
— most 0.2% signal-region-like

CWoLa hunting: overview

[Phys. Rev. Lett. 121 (2018) 241803](#)
J. Collins, K. Howe, **BPN**

A. Cukierman, **BPN**, and the ATLAS Collaboration, in preparation



unblinded 2 weeks ago!

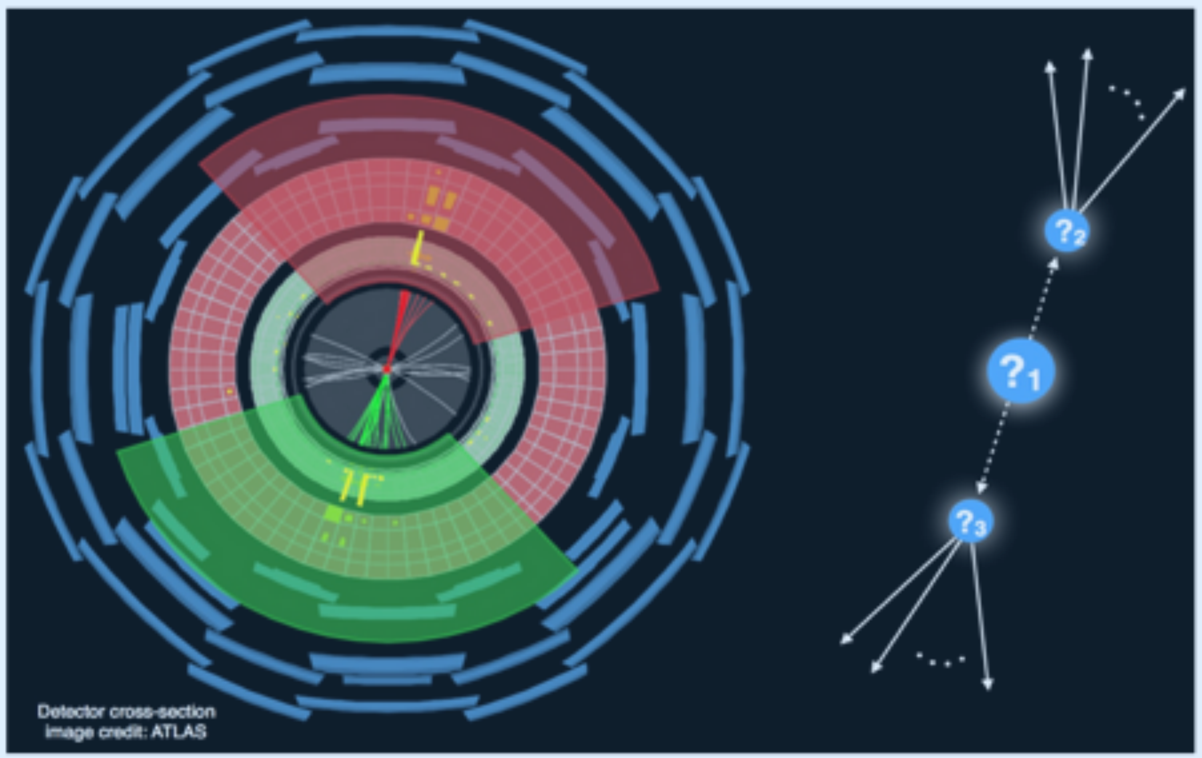
*Our first data result from **ATLAS** will come out this spring!*

The future

I believe that anomaly detection will become a significant effort at the LHC and beyond.

The future

I believe that anomaly detection will become a significant effort at the LHC and beyond.

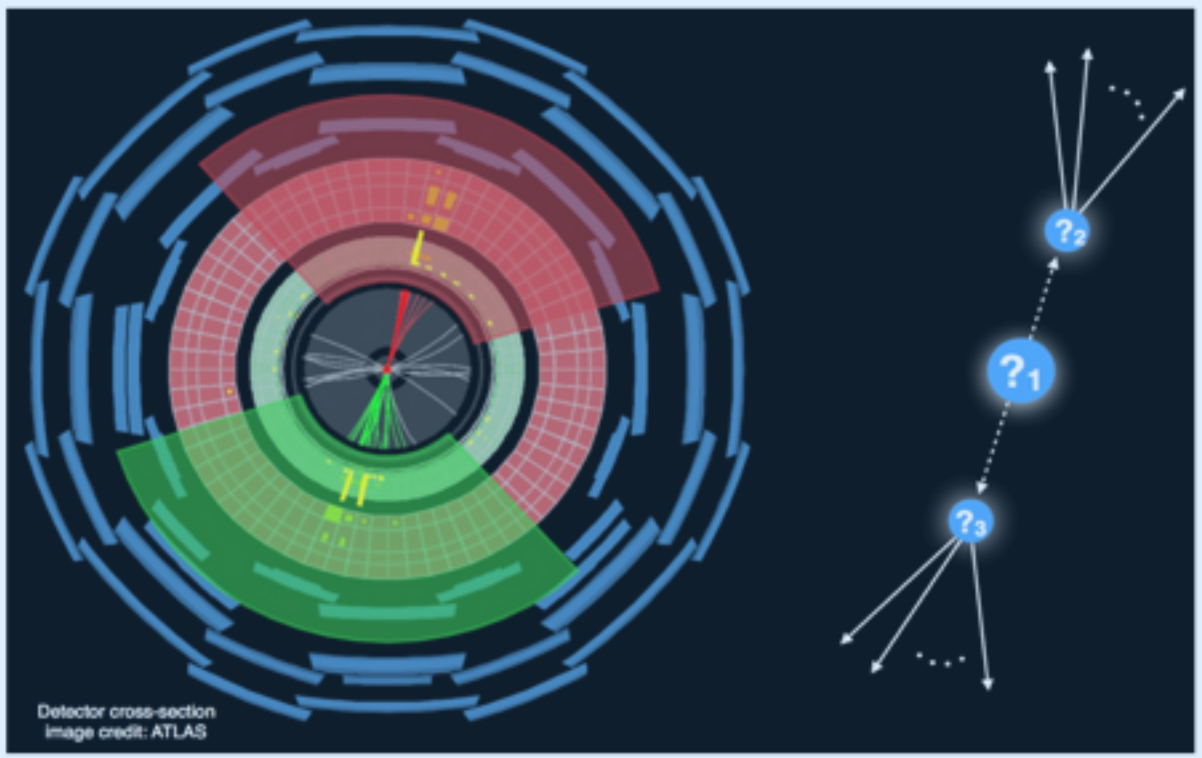


LHC Olympics 2020

https://indico.cern.ch/event/809820

The future

I believe that anomaly detection will become a significant effort at the LHC and beyond.



LHC Olympics 2020

https://indico.cern.ch/event/809820

No one algorithm will cover all possibilities - we will need a variety of methods to ensure broad coverage!

Looking forward

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

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

Likelihood-Free inference

Deep learning & **Quantum computing** for fundamental **physics**

Label-Free learning

[Deconvolution/Unfolding]

[Generative models]

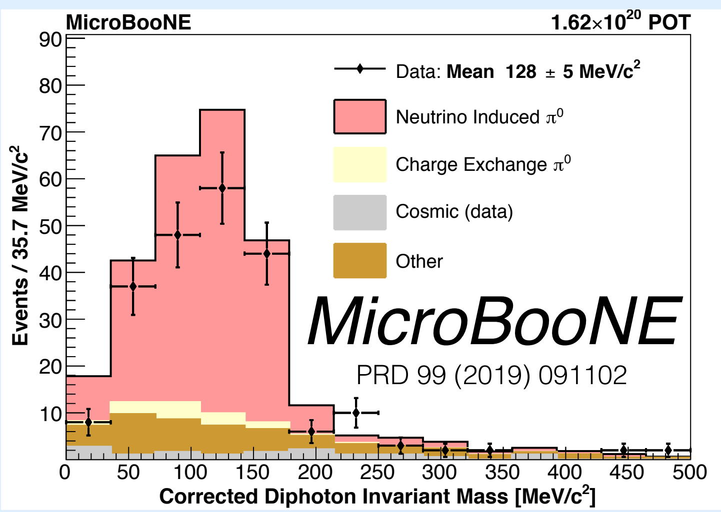
[Weak supervision]

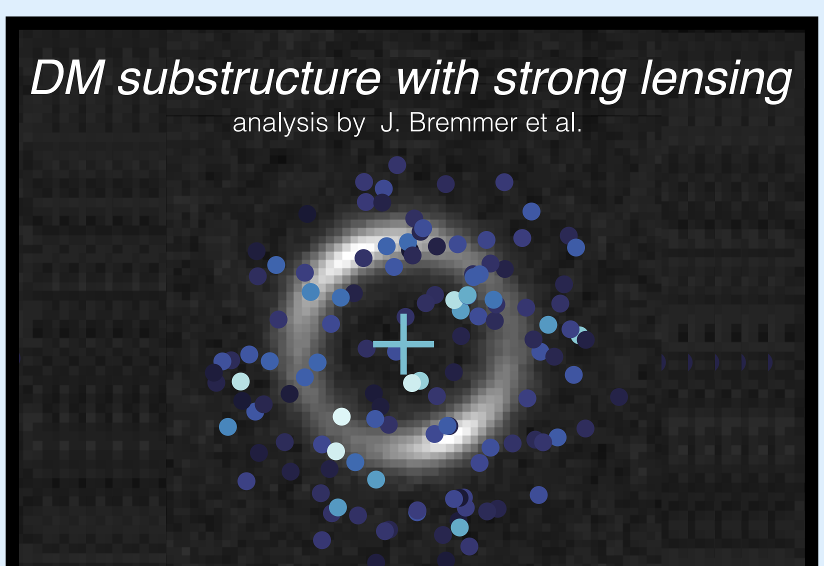
[Anomaly detection]

The future

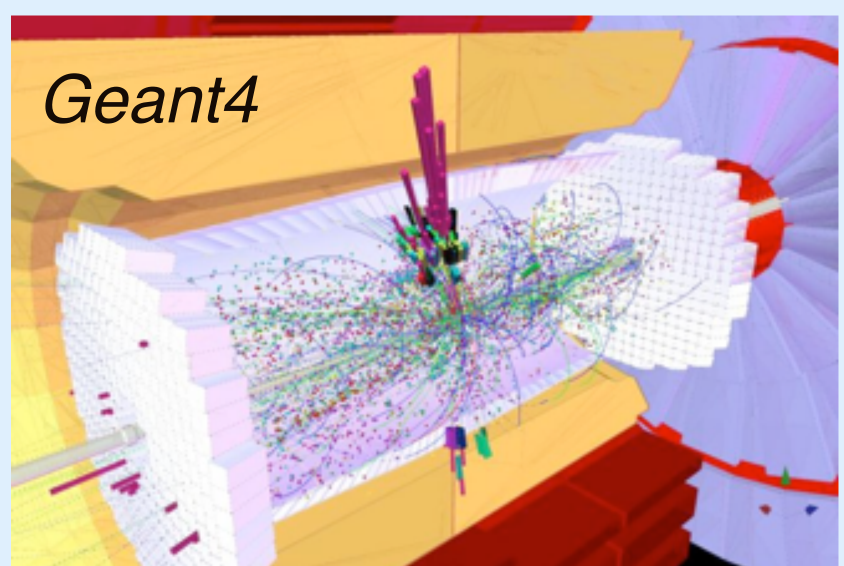
The methods I have highlighted will lead to cross-cutting R&D to further SLAC's fundamental science goals

Likelihood-free inference

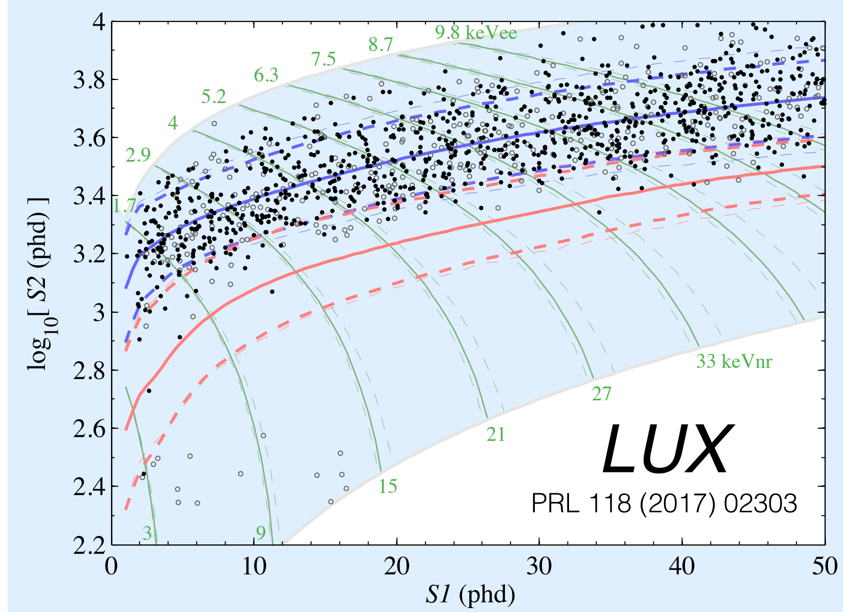


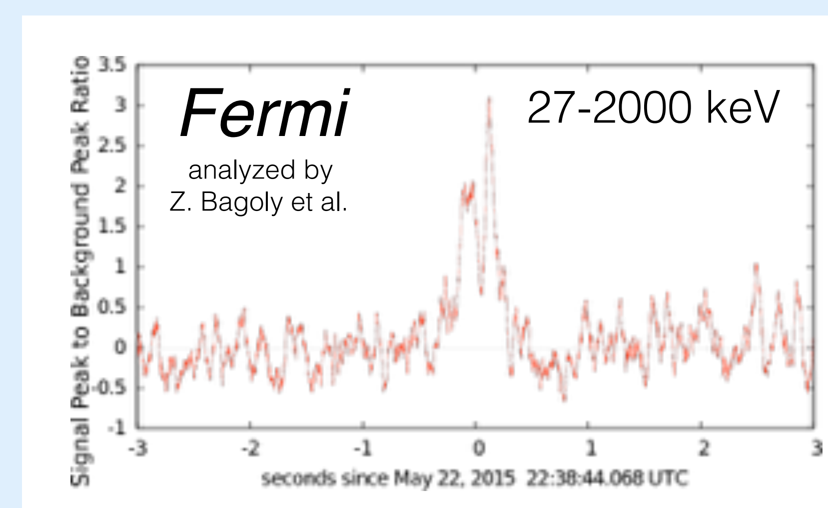


Generative Models



Anomaly Detection





The future: collaborations

snorkel

Connect with Stanford CS weak supervision experts for physics apps.

NESAP for Learning

NERSC

co-PI for distributed GAN training for accelerated calorimeter simulation on next CPU/GPU supercomputer

QuantISED HEP initiative

U.S. DEPARTMENT OF ENERGY

Continuing to study quantum computing for particle/nuclear physics

⟨AIDE|QC⟩

5-year Advanced Scientific Computing Research (ASCR) multi-institute team working on error mitigation

Q-FARM

PULSE

The future: community planning

Rare Processes, Precision Measurements and Dark Sector Production

Cosmic Frontier

Energy Frontier

Neutrino Physics Frontier

Theory Frontier

Accelerator Science and Technology Frontier

Computational Frontier

N.B. ML & QIS not part of last Snowmass!

The future: community planning

Rare Processes, Precision Measurements and Dark Sector Production

Cosmic Frontier

Energy Frontier

Neutrino Physics Frontier

Theory Frontier

Accelerator Science and Technology Frontier

Computational Frontier

N.B. ML & QIS not part of last Snowmass!

Summary

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

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

Likelihood-Free inference

Deep learning & **Quantum computing** for fundamental **physics**

Label-Free learning

[Deconvolution/Unfolding]

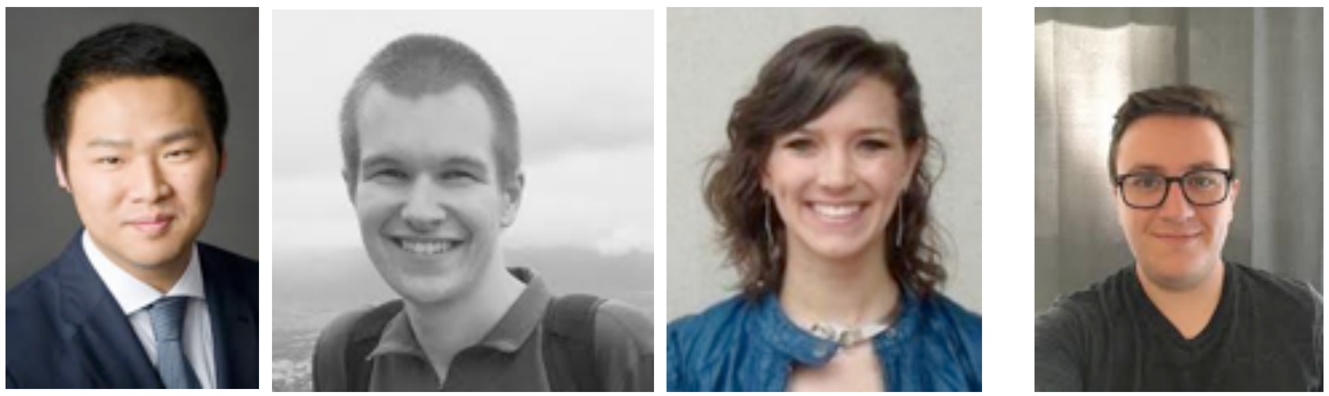
[Generative models]

[Weak supervision]

[Anomaly detection]

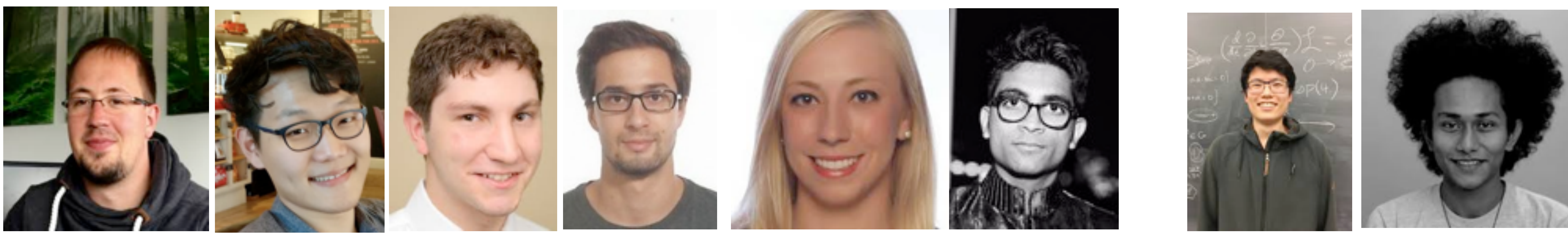
Thank you - (under)graduate students

Jet physics

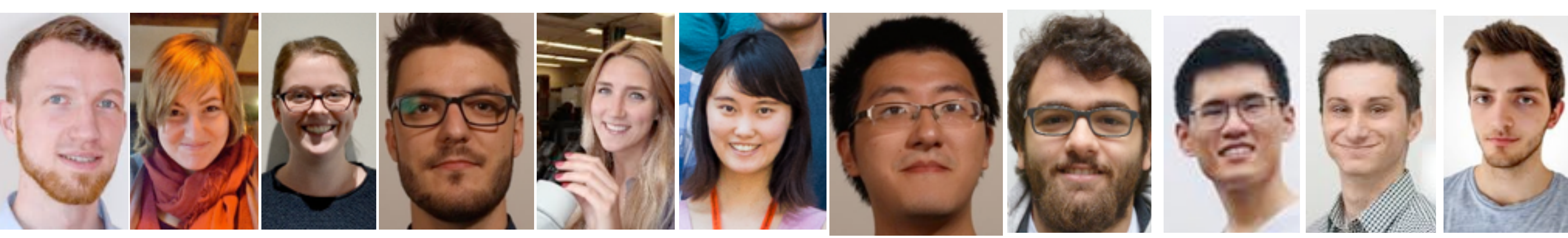


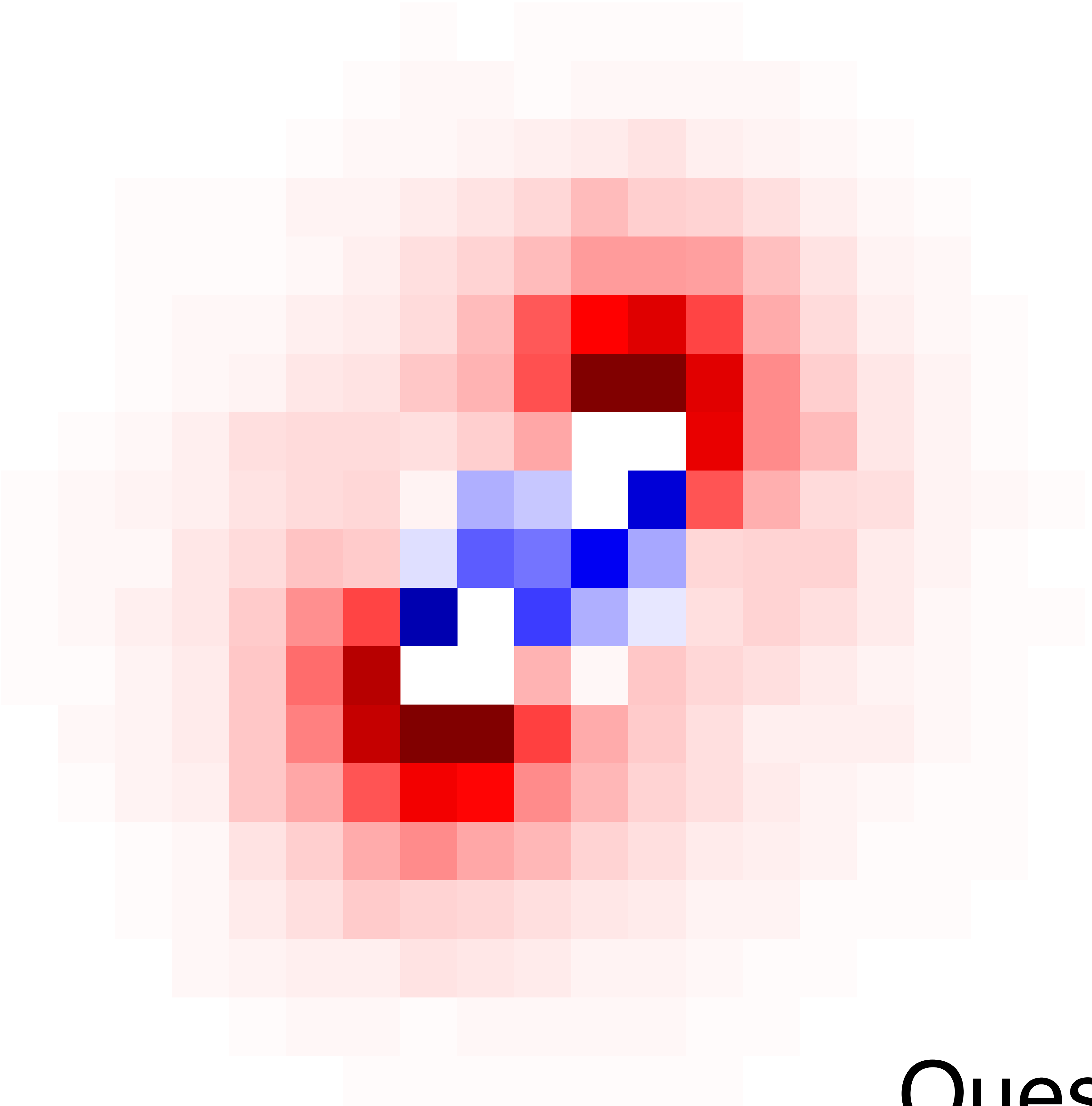
Quantum computing

Deep learning



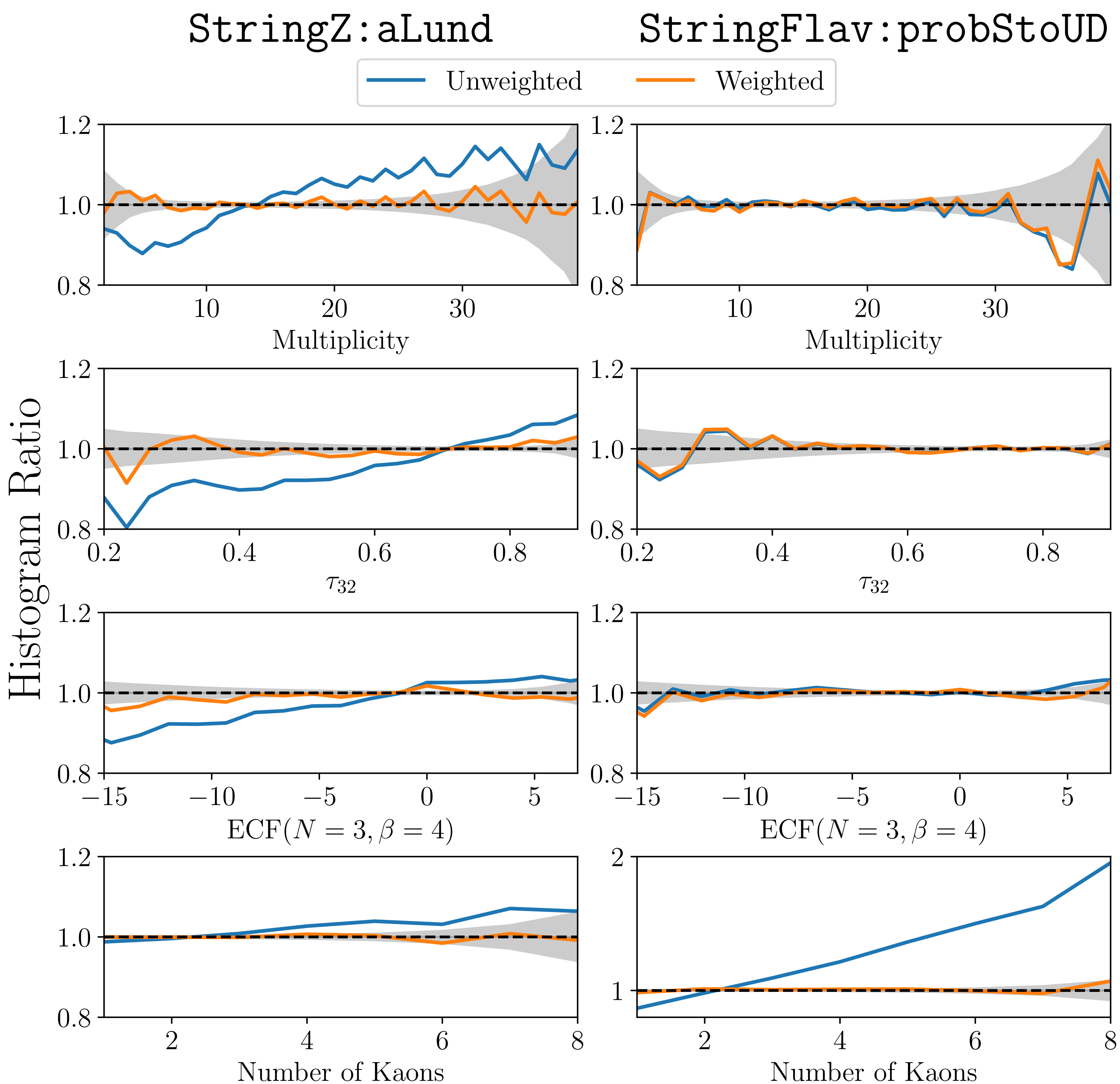
Radiation damage / Silicon detector research





Questions?

Achieving precision



Works also when the differences between the two simulations are small (left) or localized (right).

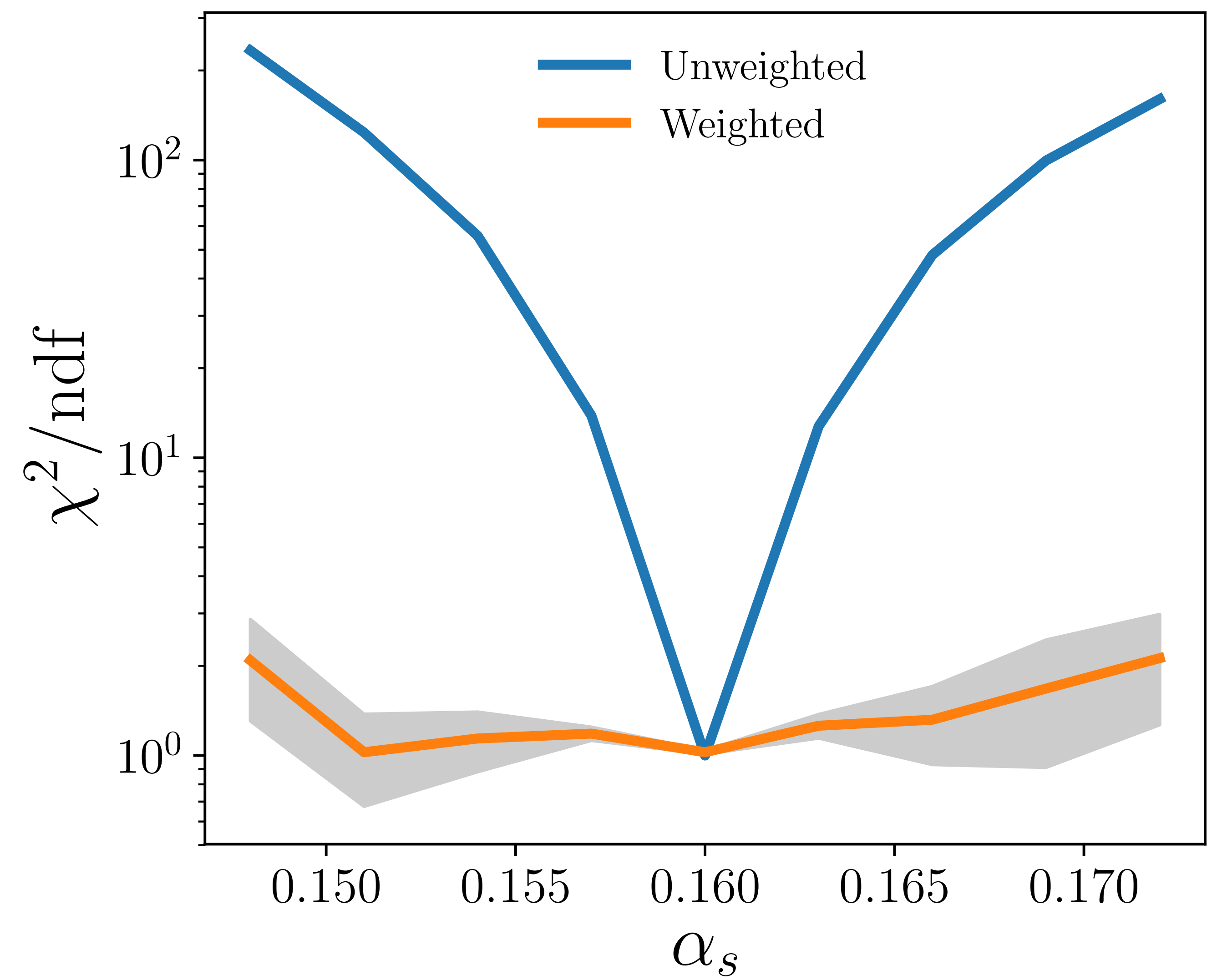
These are histogram ratios for a series of one-dimensional observables

Parameterized reweighting

What if we have a new simulation with multiple continuous parameters θ?

Easy - simply learn a parameterized classifier* !

…simply add the parameter as a feature to the network during training and let it learn to interpolate.



“fine structure constant” of the strong force

*see Cranmer, Pavez, Louppe, 1506.02169

Parameter estimation

What if we want to reweight with **pre-detector particles**, but fit to **detector-level objects**?

$$\theta^* = \operatorname*{argmax}_{\theta'} \min_{g} \sum_{i \in \theta_0} \log g(x_{D,i}) \quad \text{[data]}$$

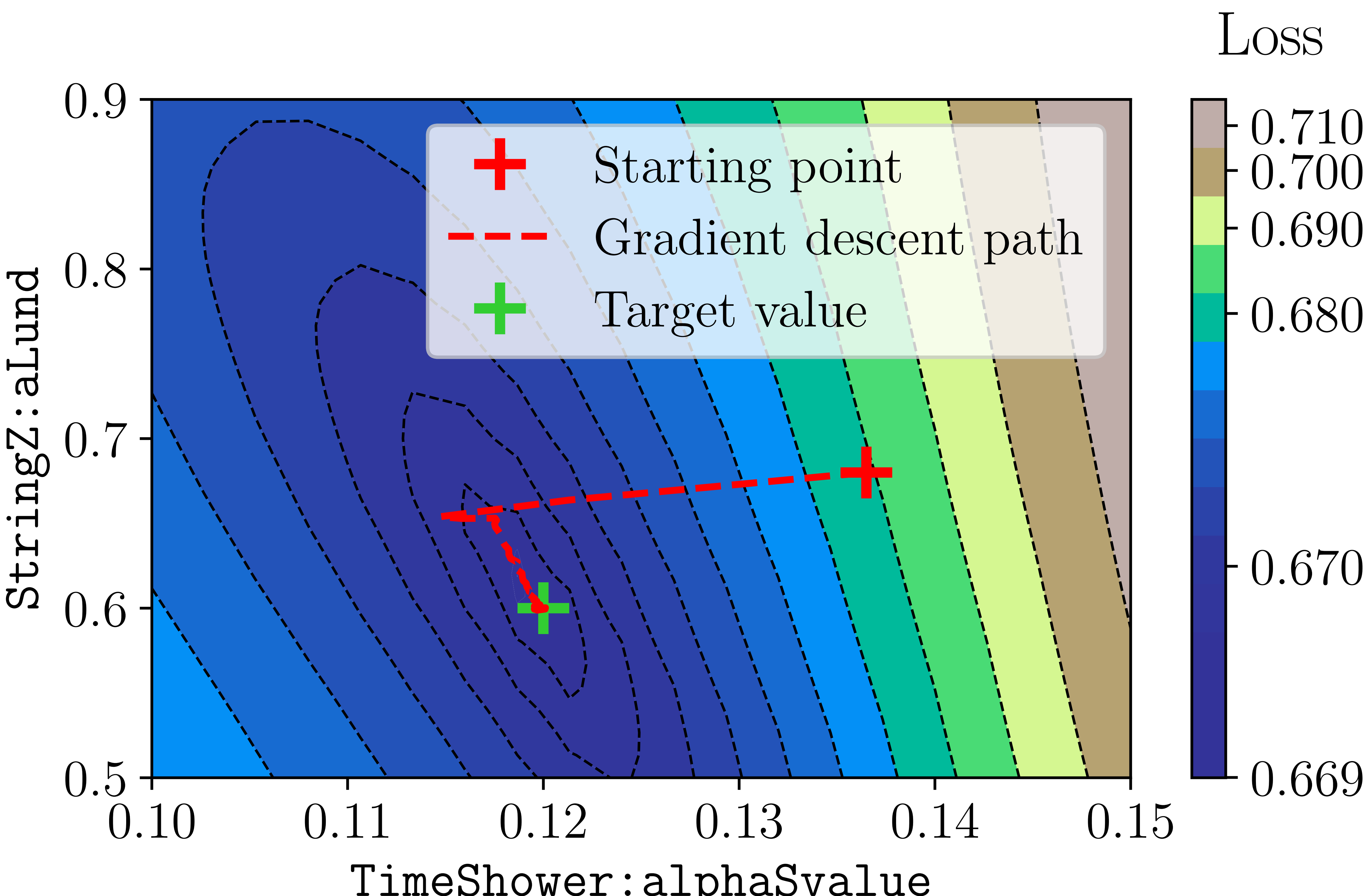
$$\text{[reweighted simulation]} \quad + \sum_{i \in \theta} w(x_{T,i}, \theta) \log(1 - g(x_{D,i}))$$

$$\frac{f(x_T, \theta)}{1 - f(x_T, \theta)}$$

Intuition: reweight until you can't distinguish the data from the (reweighted) simulation!

Parameter estimation

Fit 3 (2 shown) parameters using the full phase space!



*a different method was used for this fit - it is not a minimax procedure but doesn't work at two levels … ask later for details

Parameter estimation

Mean and standard deviation over 20 runs:

	Parameter	Target value	Fit value
Val.	TimeShower:alphaSvalue	0.1200	0.1195 ± 0.0022
	StringZ:aLund	0.6000	0.6276 ± 0.0373
	StringFlav:probStoUD	0.1200	0.1203 ± 0.0071
Blinded	TimeShower:alphaSvalue	0.1700	0.1707 ± 0.0022
	StringZ:aLund	0.7500	0.7425 ± 0.0453
	StringFlav:probStoUD	0.1400	0.1422 ± 0.0065

Similar spread

1D:

Parameter	Target value	Fit value
TimeShower:alphaSvalue	0.1600	0.1601 ± 0.0018
StringZ:aLund	0.8000	0.7980 ± 0.0257
StringFlav:probStoUD	0.2750	0.2754 ± 0.0065

The meaning of this “uncertainty” is discussed later.

Full phase space unfolding: OmniFold

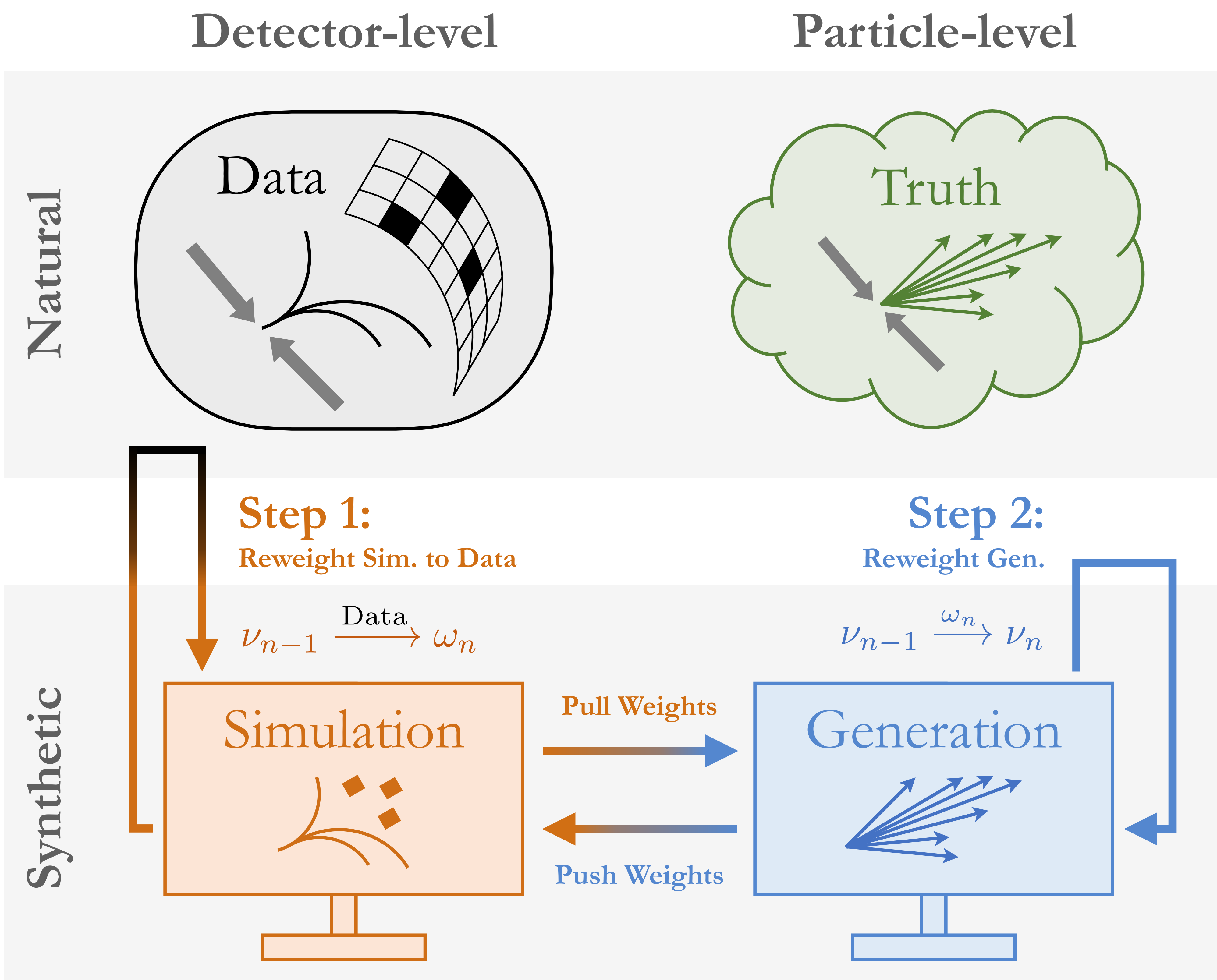
Emily Dickinson, #975

The Mountain sat upon the Plain
In his tremendous Chair –
His observation omnifold,
His inquest, everywhere –

The Seasons played around his knees
Like Children round a sire –
Grandfather of the Days is He
Of Dawn, the Ancestor –

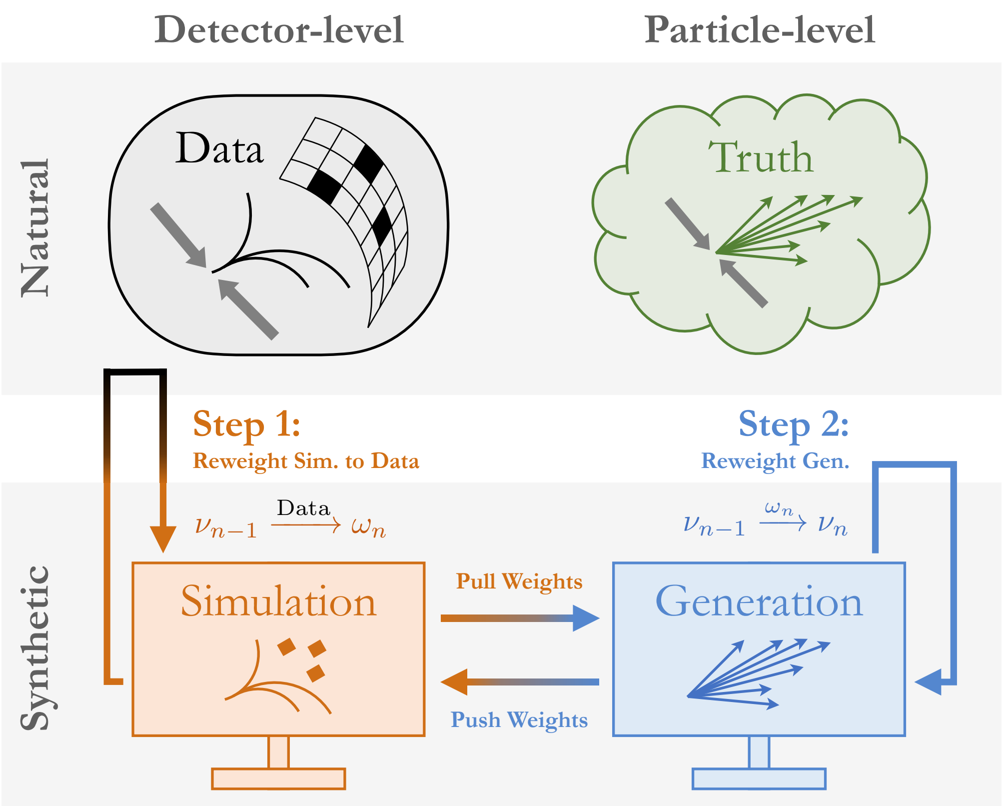
A. Andreassen, E. Metodiev, P. Komiske, **BPN**, J. Thaler, 1911.09107

Full phase space unfolding: OmniFold



A. Andreassen, E. Metodiev, P. Komiske, **BPN**, J. Thaler, 1911.09107

Full phase space unfolding: OmniFold



Notation: m = measured, t = true

$$L[(w,X),(w',X')](x) = \frac{p_{(w,X)}(x)}{p_{(w',X')}(x)}$$

(accomplish with a classifier, as before)

$$\nu_0^{\text{push}}(m) = \nu_0(t) \qquad \omega_n^{\text{pull}}(t) = \omega_n(m)$$

(these are not functions, since t → m is not 1:1)

Iterate:

1. $\omega_n(m) = \nu_{n-1}^{\text{push}}(m)\, L[(1,\text{Data}),(\nu_{n-1}^{\text{push}},\text{Sim.})](m),$
2. $\nu_n(t) = \nu_{n-1}(t)\, L[(\omega_n^{\text{pull}},\text{Gen.}),(\nu_{n-1},\text{Gen.})](t).$

Auxiliary features

extreme example: $$\mathrm{measured}|\mathrm{true} = \mathrm{true} + X$$

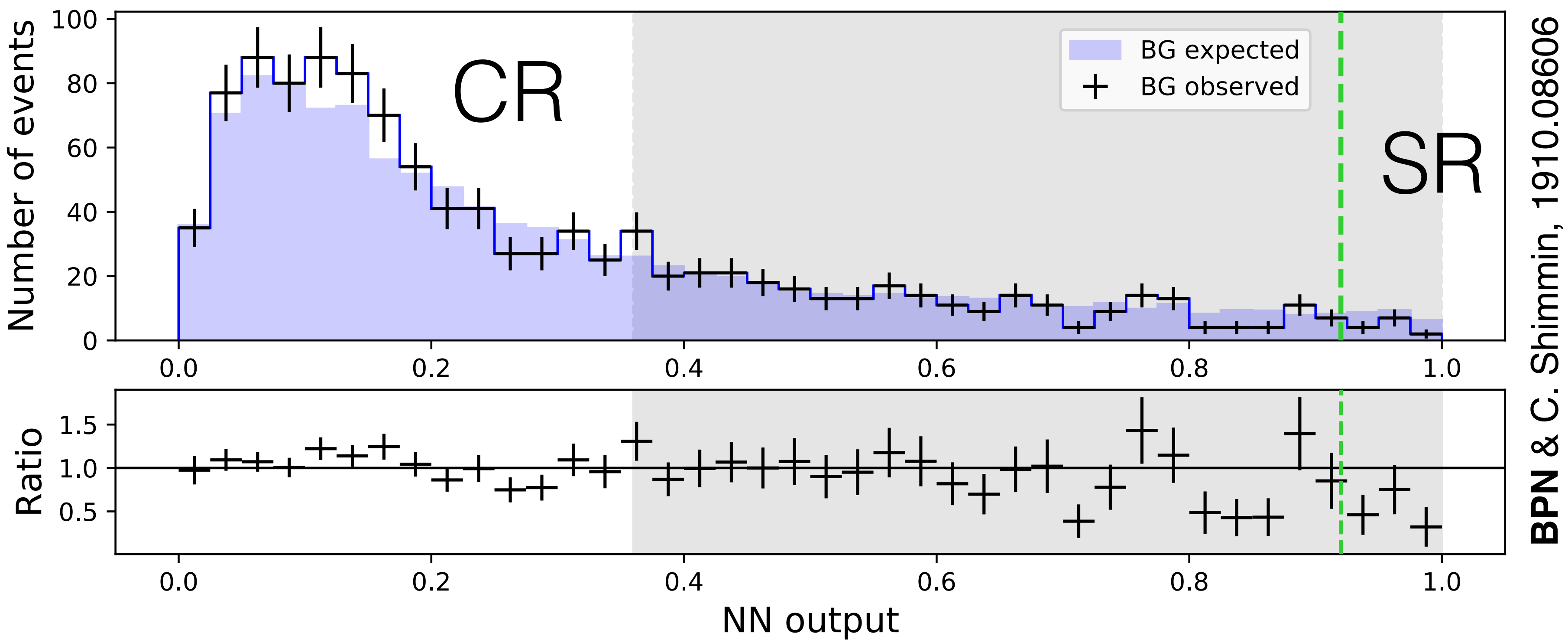
$$X \sim \mathcal{N}(\mu, \sigma)$$

If you control for X (=auxiliary feature), response is a delta-function!

Setup

To keep things simple, let's use
the following common example:

1. Train a classifier (in sim.) for signal vs. background.
2. Define a control region and a signal region using (1).
3. Normalize simulation in CR.
4. Compare data and scaled simulation in SR.
5. Significantly different? go to Stockholm; else publish limits.



Uncertainties for a NN-based analysis

Precision / Optimality

*Bad use of our data, time, money, etc. but **not wrong**.*

Accuracy / Bias

Uncertainties for a NN-based analysis

Precision / Optimality: $\mathrm{NN}(\mathrm{x}) \neq \frac{p_{\mathrm{true}}(x|\mathrm{S+B})}{p_{\mathrm{true}}(x|\mathrm{B})}$

Optimal by Neyman-Pearson

Accuracy / Bias

Note that this is not p(x|S) / p(x|B), however the two are monotonically related to each other.

Uncertainties for a NN-based analysis

Precision / Optimality: $\mathrm{NN(x)} \neq \frac{p_{\mathrm{true}}(x|\mathrm{S+B})}{p_{\mathrm{true}}(x|\mathrm{B})}$

Accuracy / Bias: $p_{\mathrm{prediction}}(\mathrm{NN}) \neq p_{\mathrm{true}}(\mathrm{NN})$

The distribution of the scaled sim. is not correct.

Uncertainties for a NN-based analysis

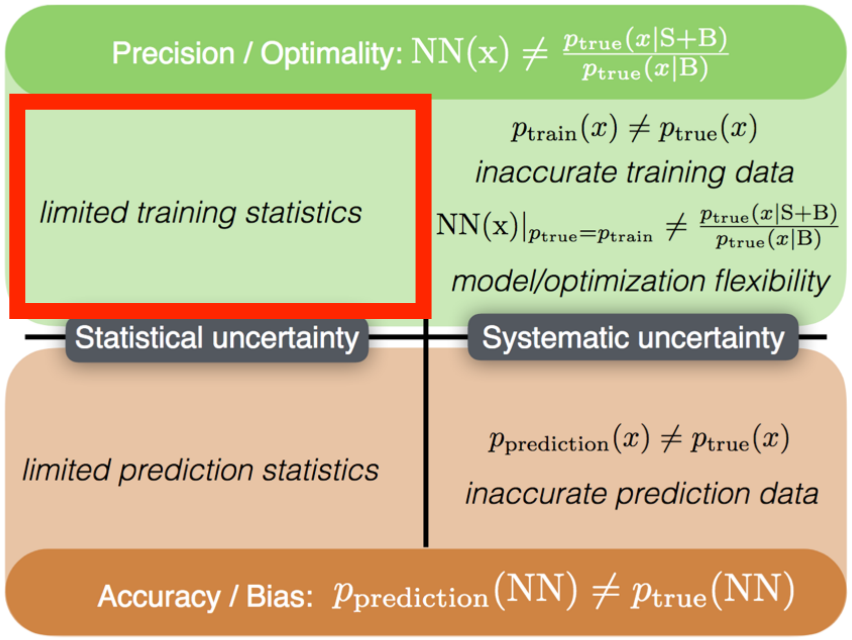
Precision / Optimality: $\mathrm{NN}(x) \neq \frac{p_{\text{true}}(x|\mathrm{S+B})}{p_{\text{true}}(x|\mathrm{B})}$

Statistical uncertainty	Systematic uncertainty
limited training statistics	$p_{\text{train}}(x) \neq p_{\text{true}}(x)$ *inaccurate training data* $\mathrm{NN}(x)\big|_{p_{\text{true}}=p_{\text{train}}} \neq \frac{p_{\text{true}}(x|\mathrm{S+B})}{p_{\text{true}}(x|\mathrm{B})}$ *model/optimization flexibility*
limited prediction statistics	$p_{\text{prediction}}(x) \neq p_{\text{true}}(x)$ *inaccurate prediction data*

Accuracy / Bias: $p_{\text{prediction}}(\mathrm{NN}) \neq p_{\text{true}}(\mathrm{NN})$

BPN, 1909.03081

How to estimate precision stat. uncerts.

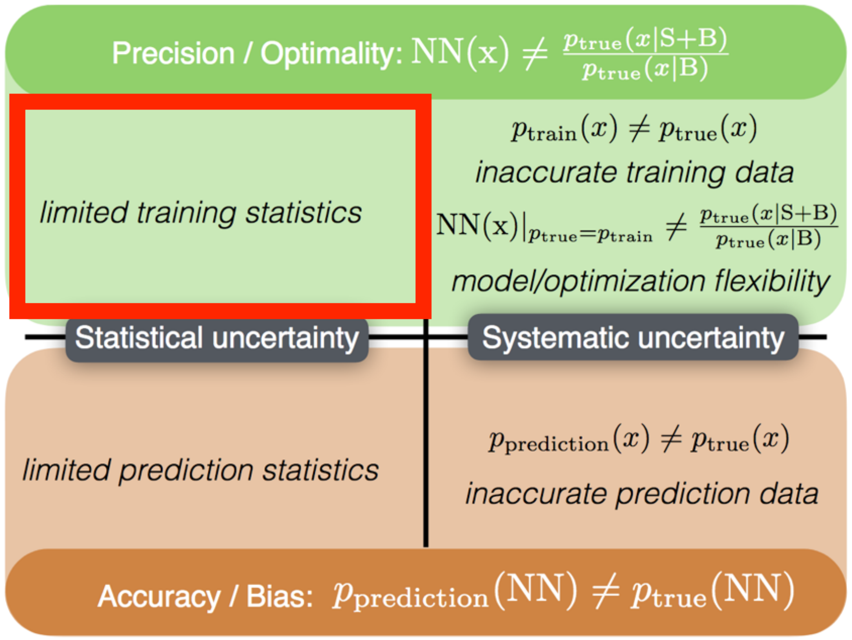


You can always accomplish this by bootstrapping: making pseudo-datasets from resampling and then retraining.

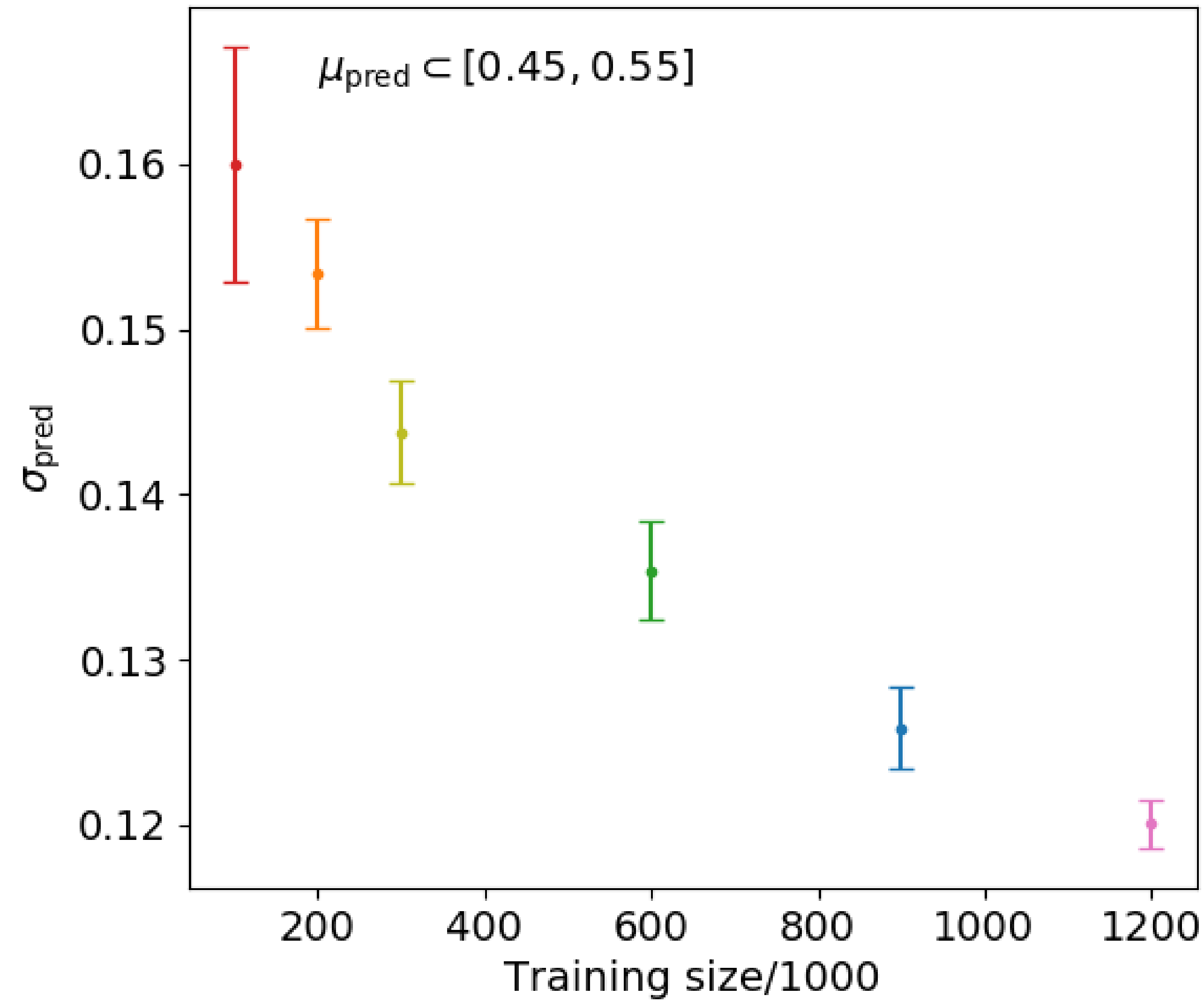
It is important to fix the NN initialization so that you are not also testing your sensitivity to that.

This can be painful because it requires retraining many NNs.

How to estimate precision stat. uncerts.

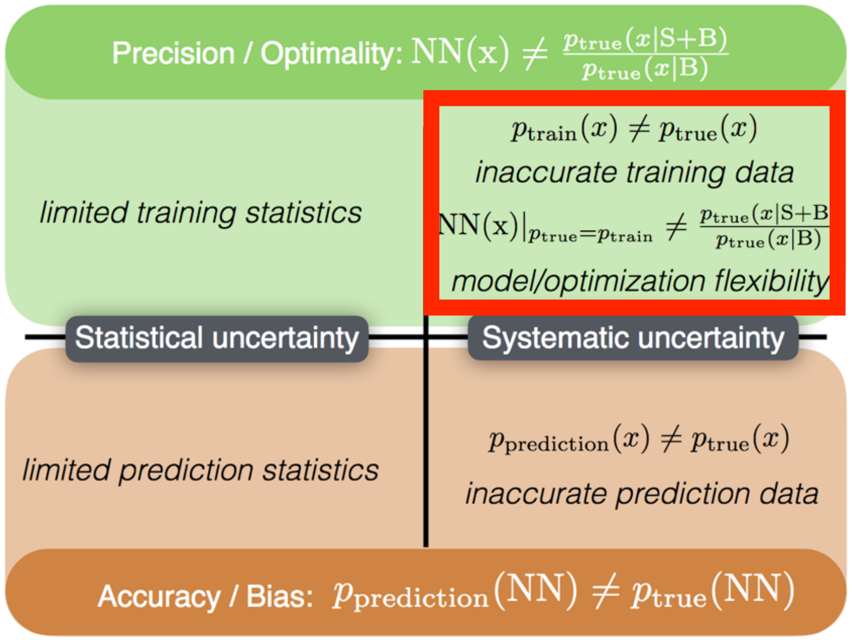


Alternative: train **one** Bayesian NN?!



S. Bollweg, M. Haußmann, G. Kasieczka, M. Luchmann, T. Plehn, J. Thompson, 1904.10004

How to estimate precision syst. uncerts.

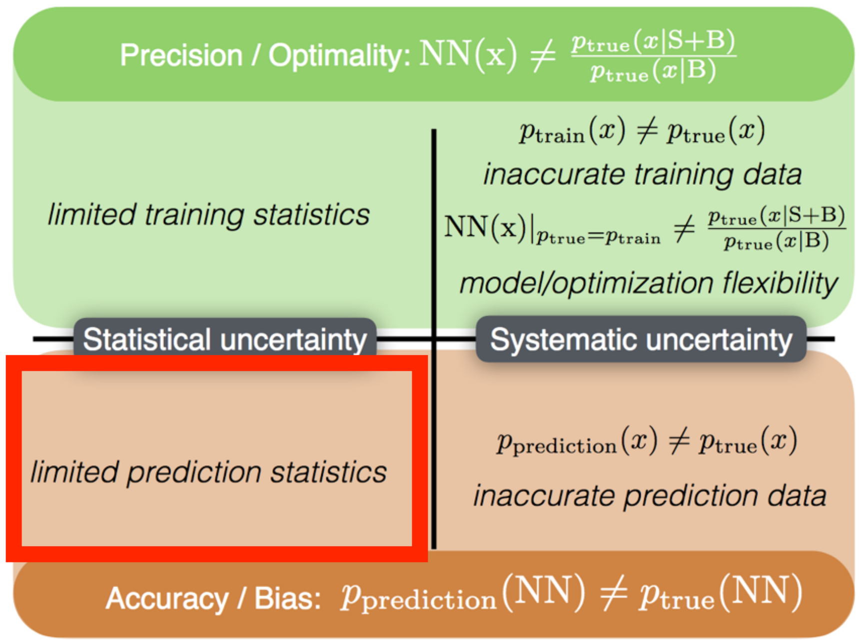


As with all systematic uncertainties, this is hard to quantify.

One component is due to the modeling of p(x) - more on this later.

Testing the flexibility of the network requires checking the sensitivity to the architecture (#layers, nodes/layer, etc.), the initialization, the training procedure (#epochs, learning rate, etc.)

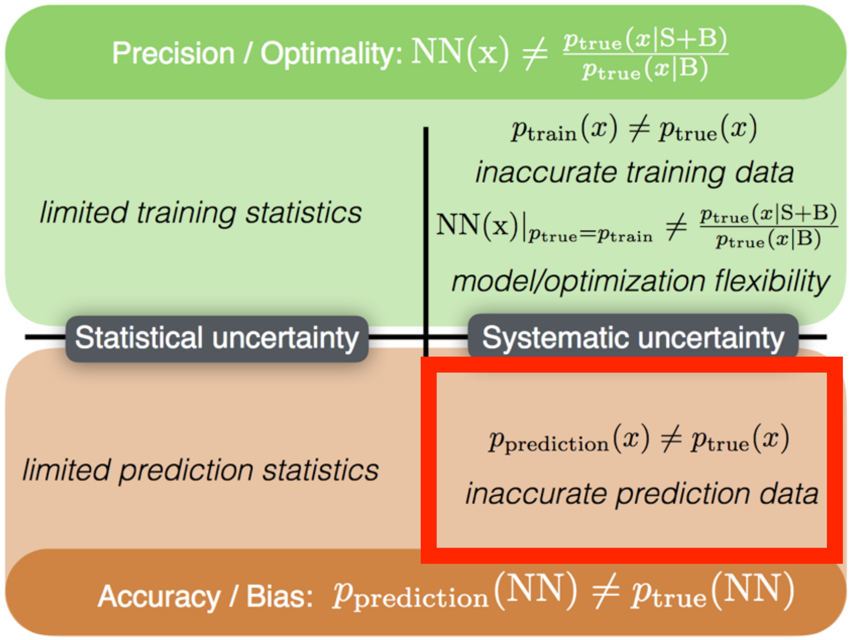
How to estimate bias stat. uncerts.



This uncertainty is post-training - Bayesian NNs unfortunately won’t help.

Unfortunately, this is often not small given that we are now probing extreme final states and have a limited computing budget.

How to estimate bias syst. uncerts.



This is the trickiest one…

…because we need the uncertainty on the modeling of x and x can be high-dimensional!

In many cases, the uncertainties factorize, e.g. the uncertainty on the jet energy is measured and evaluated per jet.

What about physics modeling uncertainties where we usually have a two-point comparison? (e.g. Pythia versus Herwig)

High-dimensional Bias Uncertainties

One word of caution: current paradigm for uncertainties may be too naive for high-dimensional analysis!

(truly end-to-end)

e.g. for some uncertainties, we often compare two different models - one nuisance parameter.

How can we even see how sensitive we are to high-dimensional effects?

High-dimensional Bias Uncertainties

One word of caution: current paradigm for uncertainties may be too naive for high-dimensional analysis!

(truly end-to-end)

e.g. for some uncertainties, we often compare two different models - one nuisance parameter.

How can we even see how sensitive we are to high-dimensional effects?

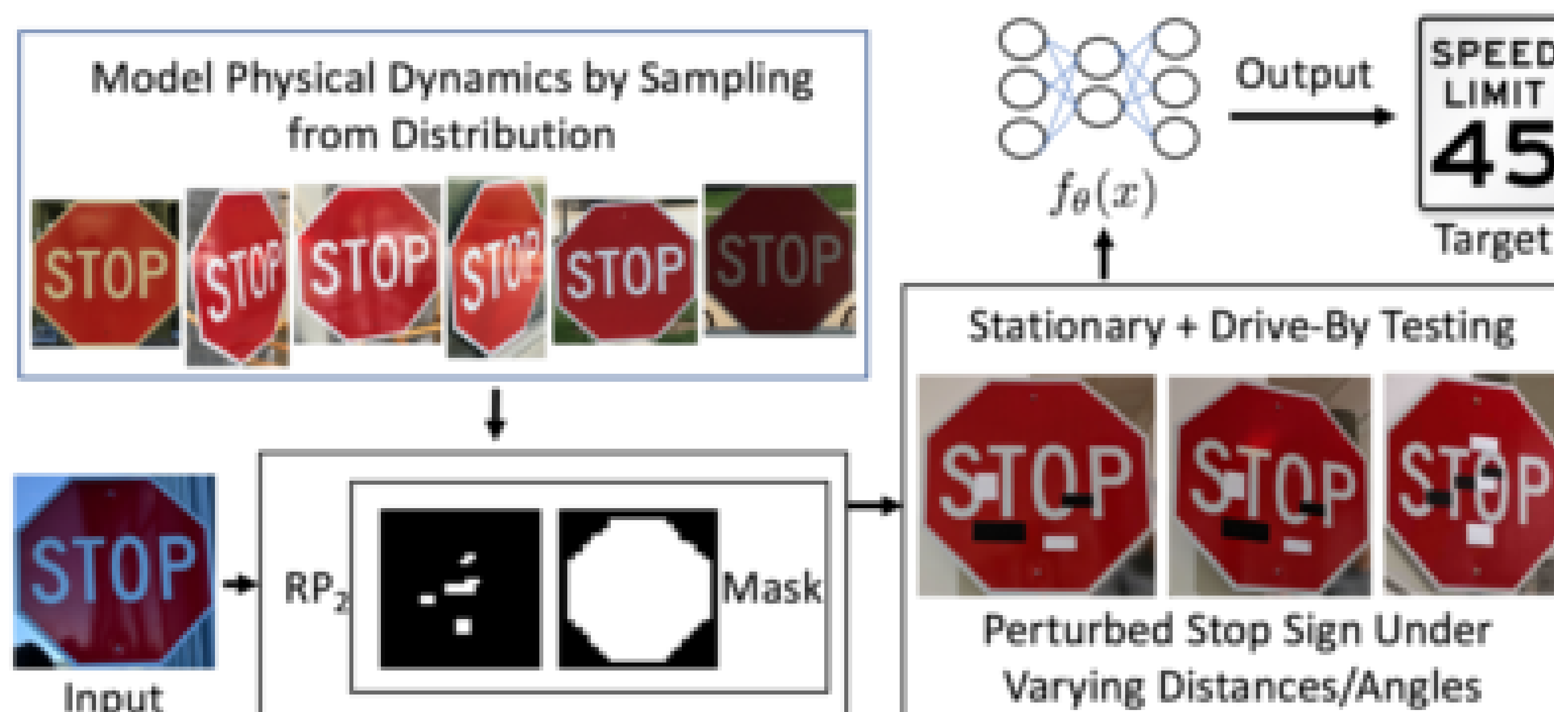
Answer: borrow tools from AI Safety

AI Safety

There is a vast literature on how easy it is to "attack" a NN.

They want to know: how subtle can an attack be and still significantly impact the output.

We know (hope?!) that nature is not evil, but these tools can help us probe the high-dimensional sensitivity of our NNs.



K. Eykholt et. al, 1707.08945

Bounding high-dim. uncerts: strategy

J = jet (in all of its high-dimensional glory)

f = fixed classifier for signal vs. background

Loss

$$\mathcal{L}_{\mathrm{sig}} = \log(1 - f(g(J))),$$

$$\mathcal{L}_{\mathrm{bg}} = \lambda_{\mathrm{cls}}(f(J) - f(g(J)))^2 + \sum_i \lambda_{\mathrm{obs}}^{(i)}(\mathcal{O}^{(i)}(J) - \mathcal{O}^{(i)}(g(J))^2$$

BPN & C. Shimmin, 1910.08606

g is a learned NN that maps J to J + δJ.

O(J) are observables that will be validated in the CR.

Bounding high-dim. uncerts: strategy

J = jet (in all of its high-di

f = fixed classifier for signa

$$\mathcal{L}_{\text{sig}} = \log(1 - f(g(J)))$$

$$\mathcal{L}_{\text{bg}} = \lambda_{\text{cls}}(f(J) - f(g \quad))) + \sum_i \lambda_{\text{obs}}^{(i)} (\mathcal{O}^{(i)}(J) - \mathcal{O}^{(i)}(g(J))^2$$

Fun fact: this requires computing ***O*** in the NN - now have GPU implementations of many standard observables!

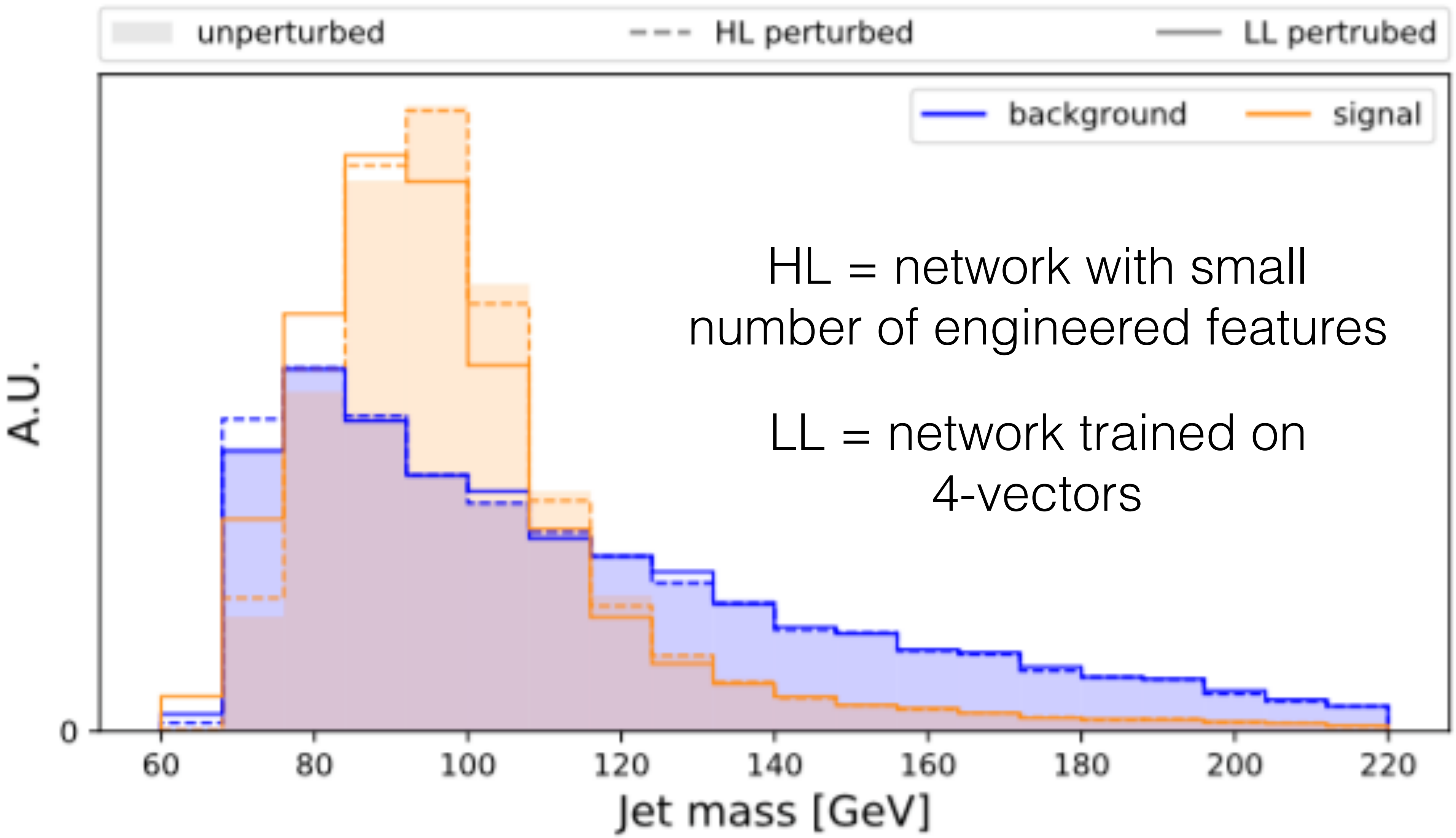
BPN & C. Shimmin, 191

g is a learned NN that maps J to J + δJ.

O(J) are observables that will be validated in the CR.

High-dimensional Uncertainty

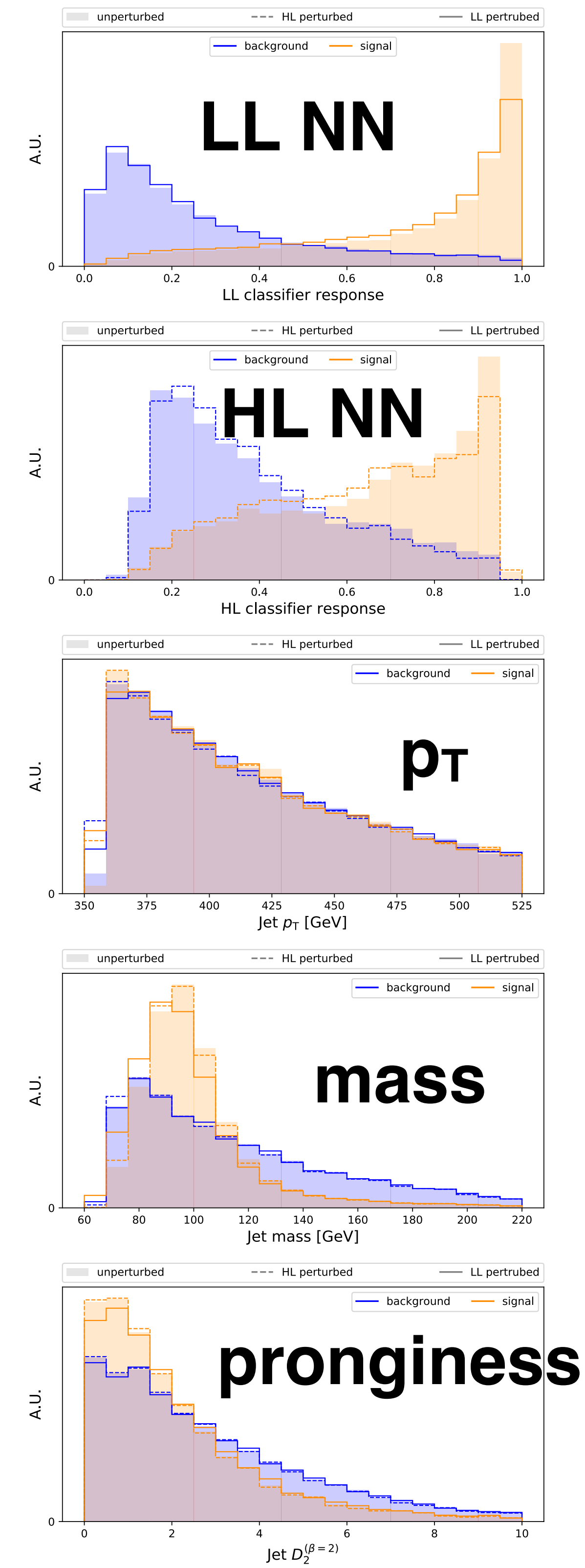
Example case: Boosted Z's versus QCD



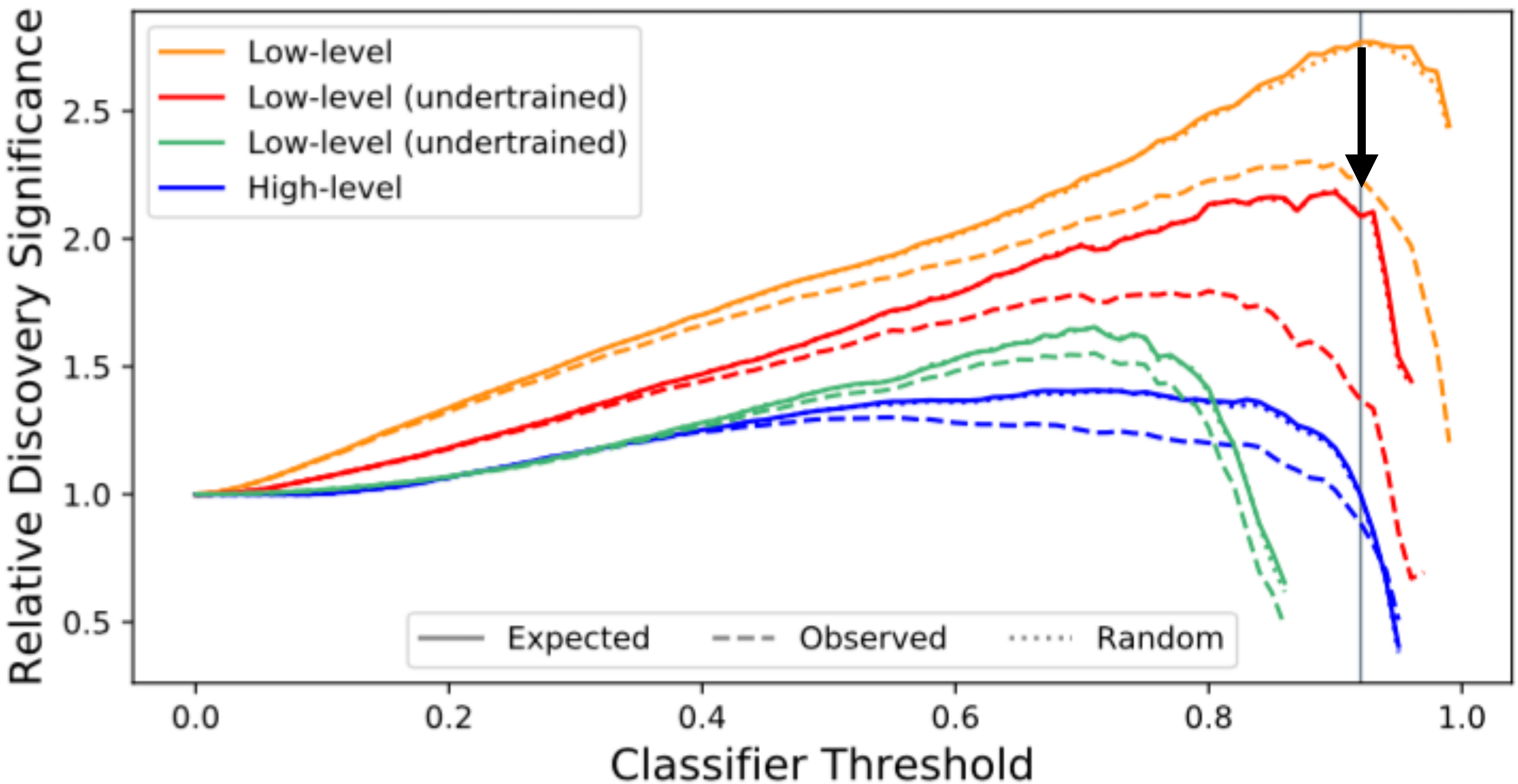
HL = network with small number of engineered features

LL = network trained on 4-vectors

It is "easy" to preserve the blue and modify the orange.

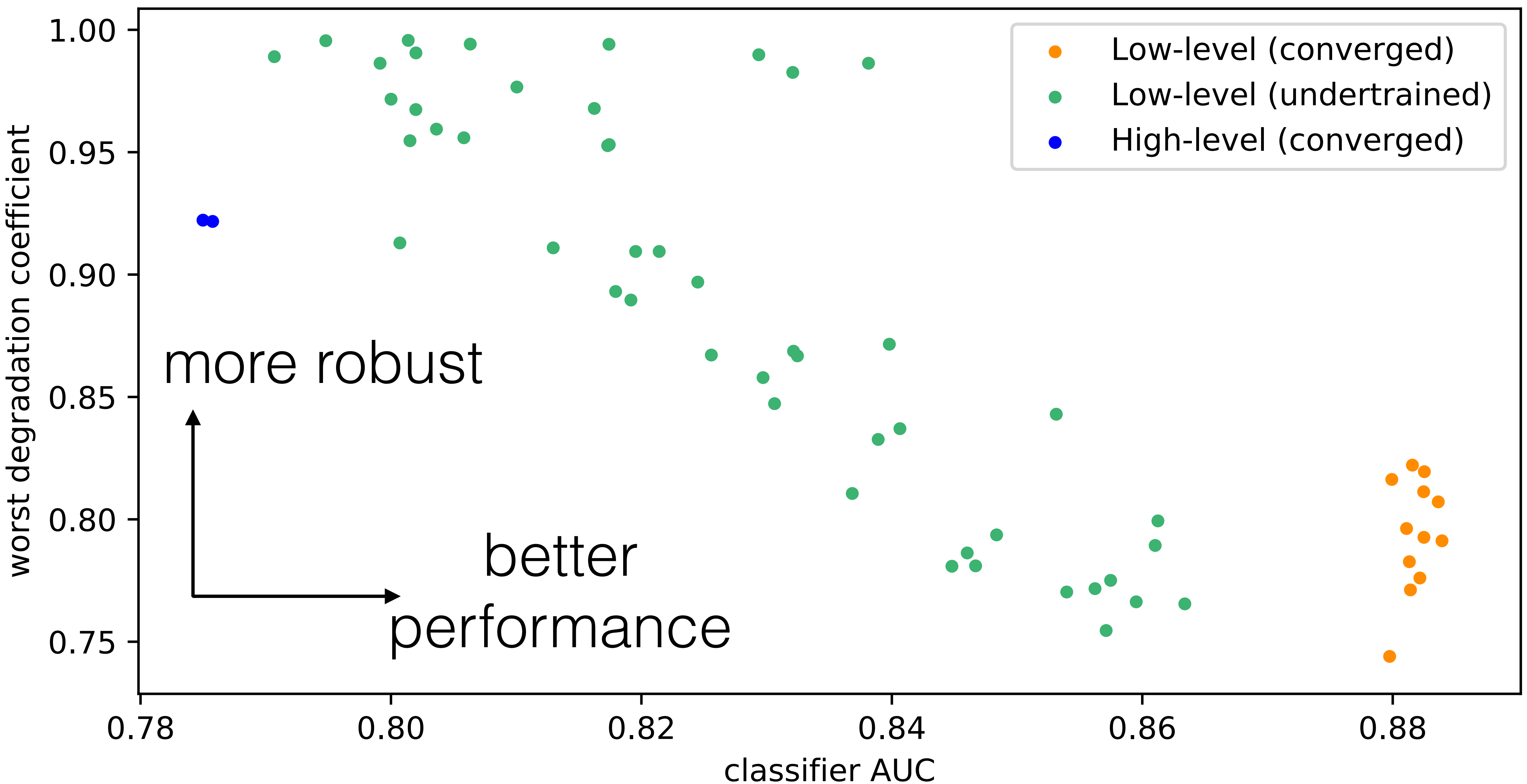


High-dimensional Uncertainty



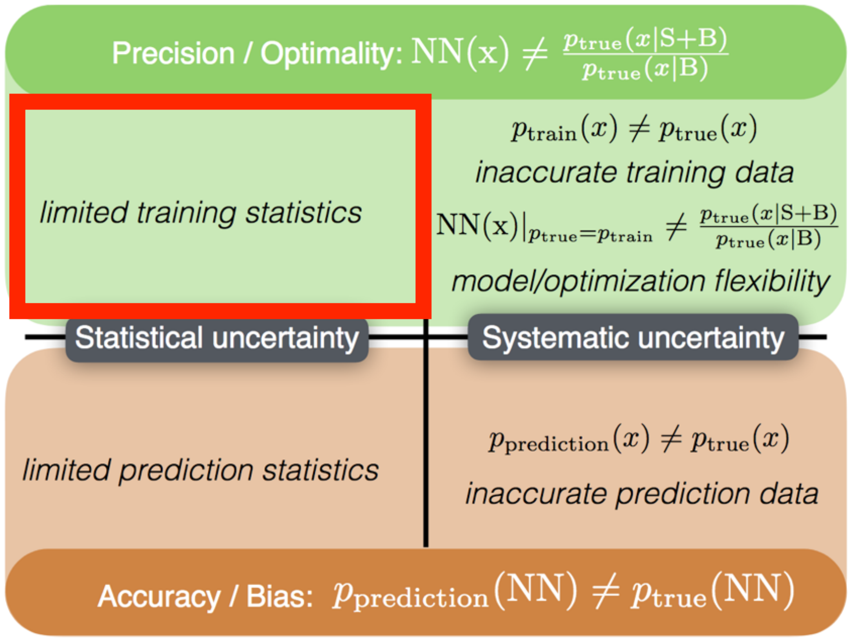
High-dimensional Uncertainty

Under training may help with robustness:



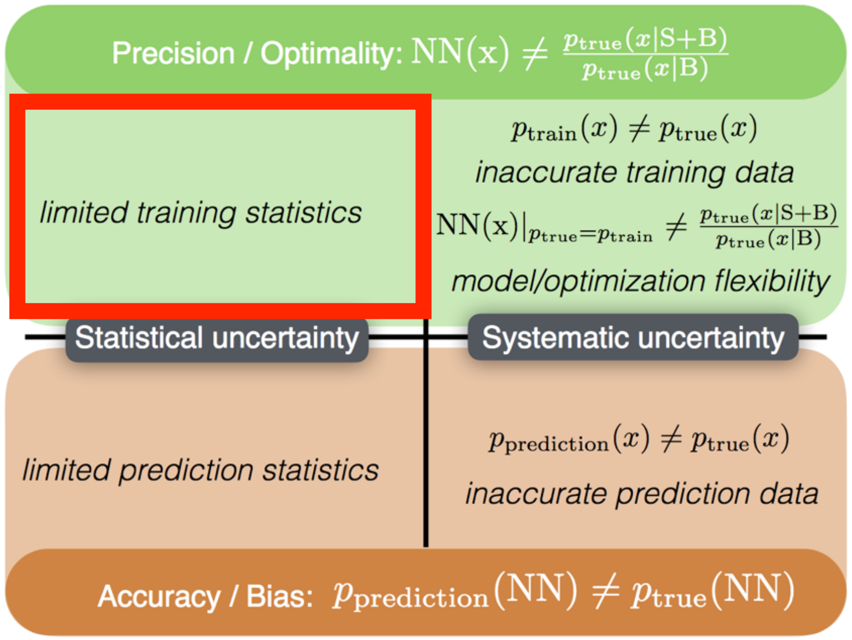
Perturbed significance improvement / nom. sig. improvement

How to reduce precision stat. uncerts.



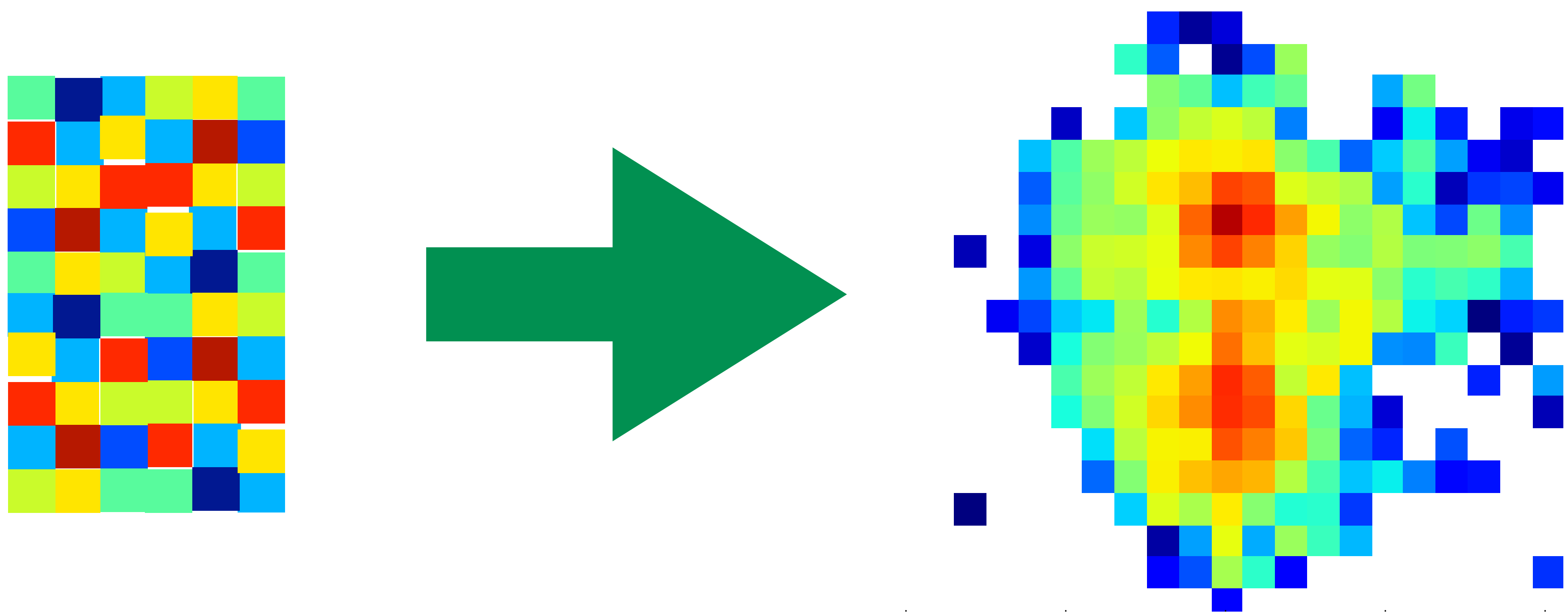
Train with more events!

How to reduce precision stat. uncerts.



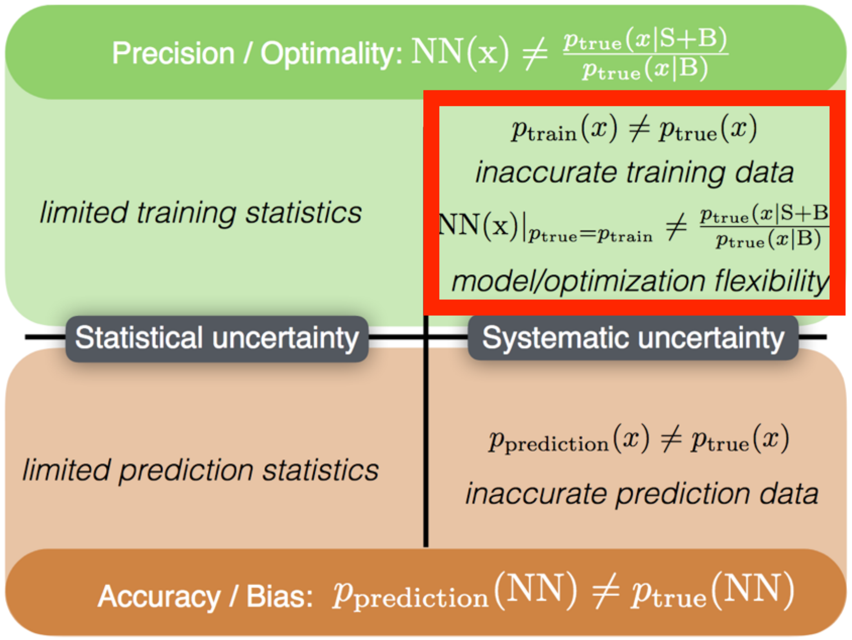
Train with more events!

…maybe use NN's to help with that



M. Paganini, L. de Oliveira, **BPN**, PRL 120 (2018) 042003 + many others (including some from other people here!)

How to reduce precision syst. uncerts.



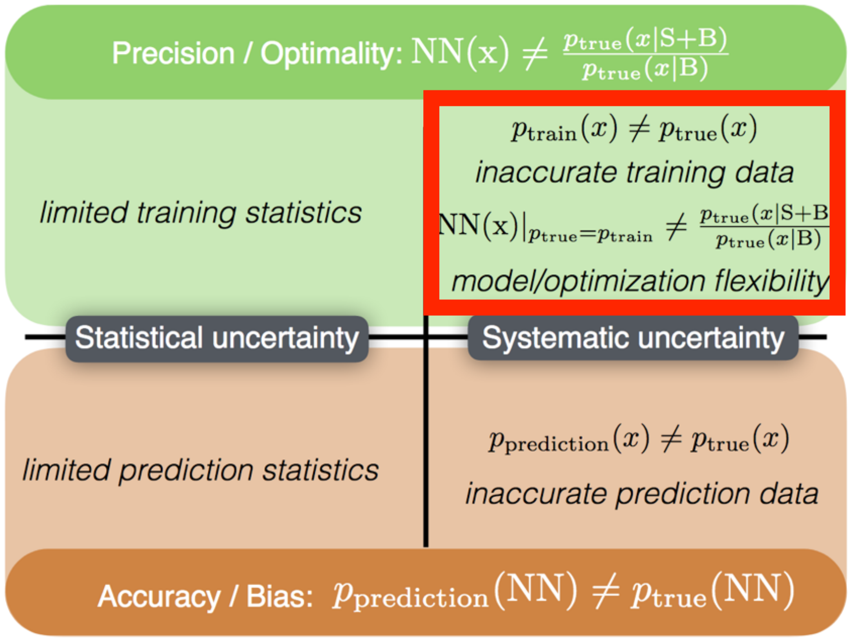
You might be tempted to force the NN to not depend on some uncertain parameters.

There are many ways to do this, e.g. adversarial techniques[1] or DisCo[2]

Unless this is needed to estimate the background[3], this is usually suboptimal and may not even reduce the uncertainty.

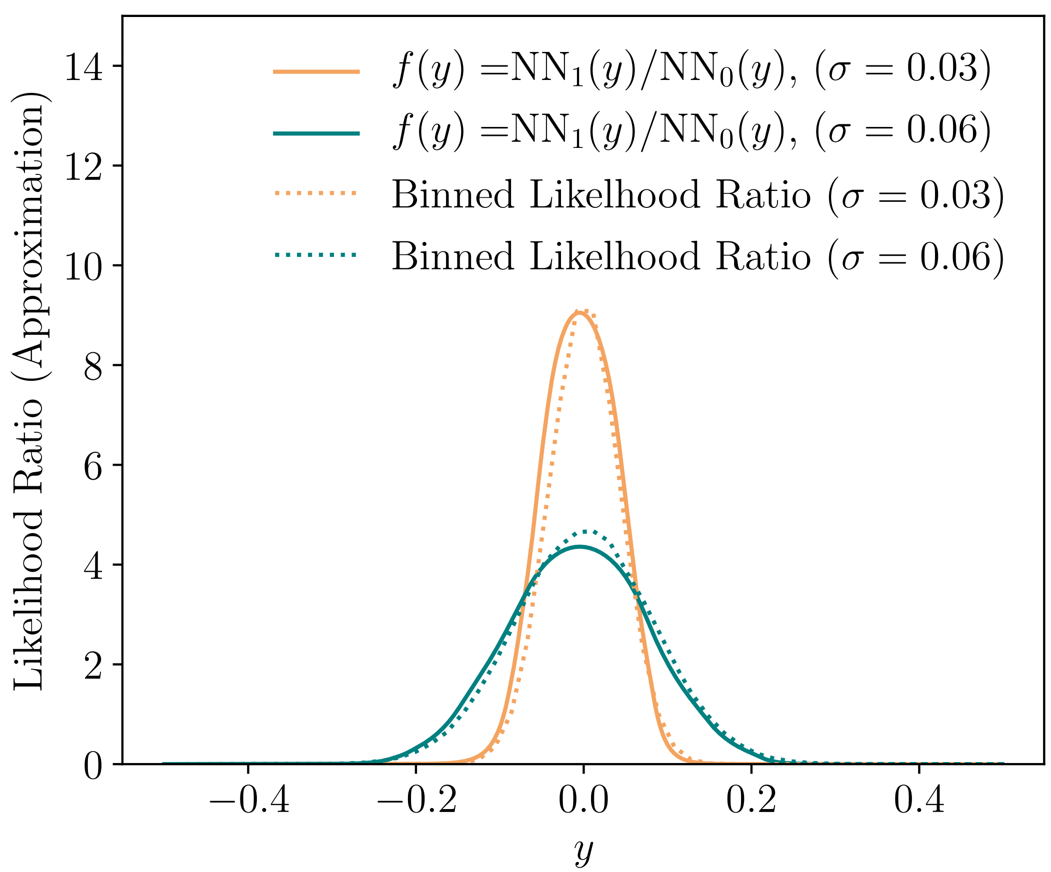
1. G. Louppe, M. Kagan, K. Cranmer, 1611.01046
2. Gregor Kasieczka and D. Shih, 2001.05310
3. C. Shimmin et al. Phys. Rev. D 96, 074034 (2017), and many others including (2)

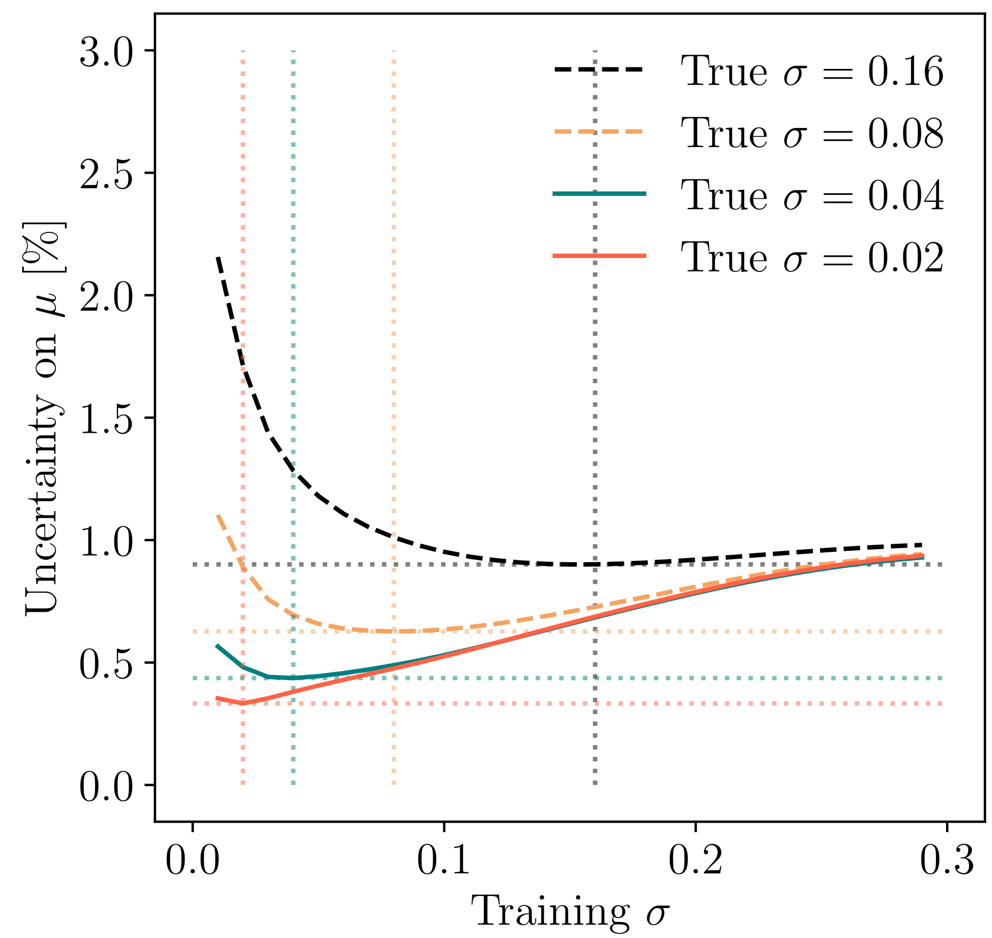
How to reduce precision syst. uncerts.



Profiling instead of pivoting:

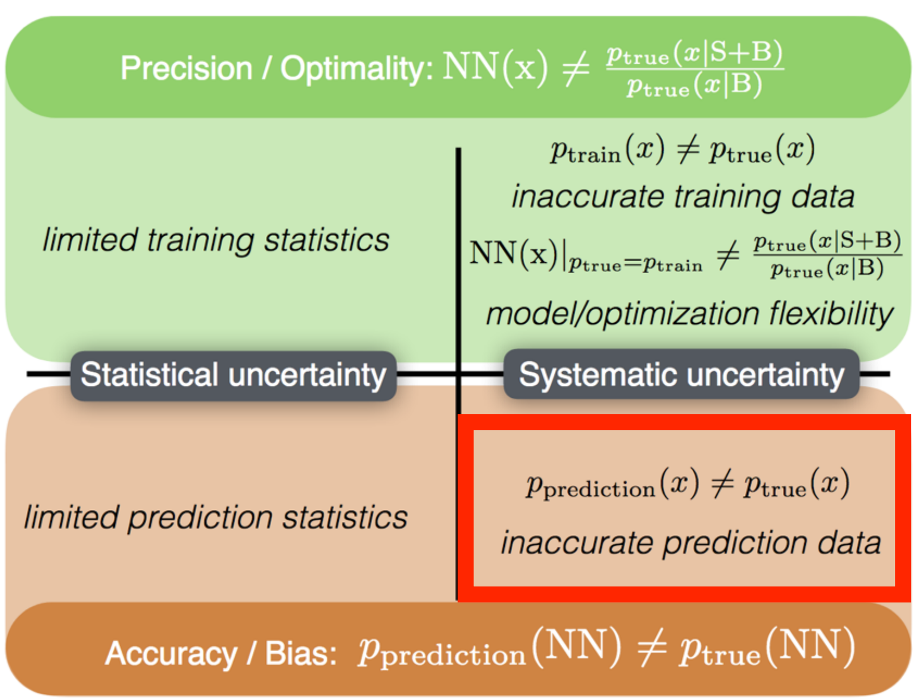
Better to do the opposite: let your NN **depend explicitly** on uncertainty quantities and then **profile them**!





How to get around high-D bias uncerts?

Work hard to understand the true nuisance parameters in the hypervariate parameter space.

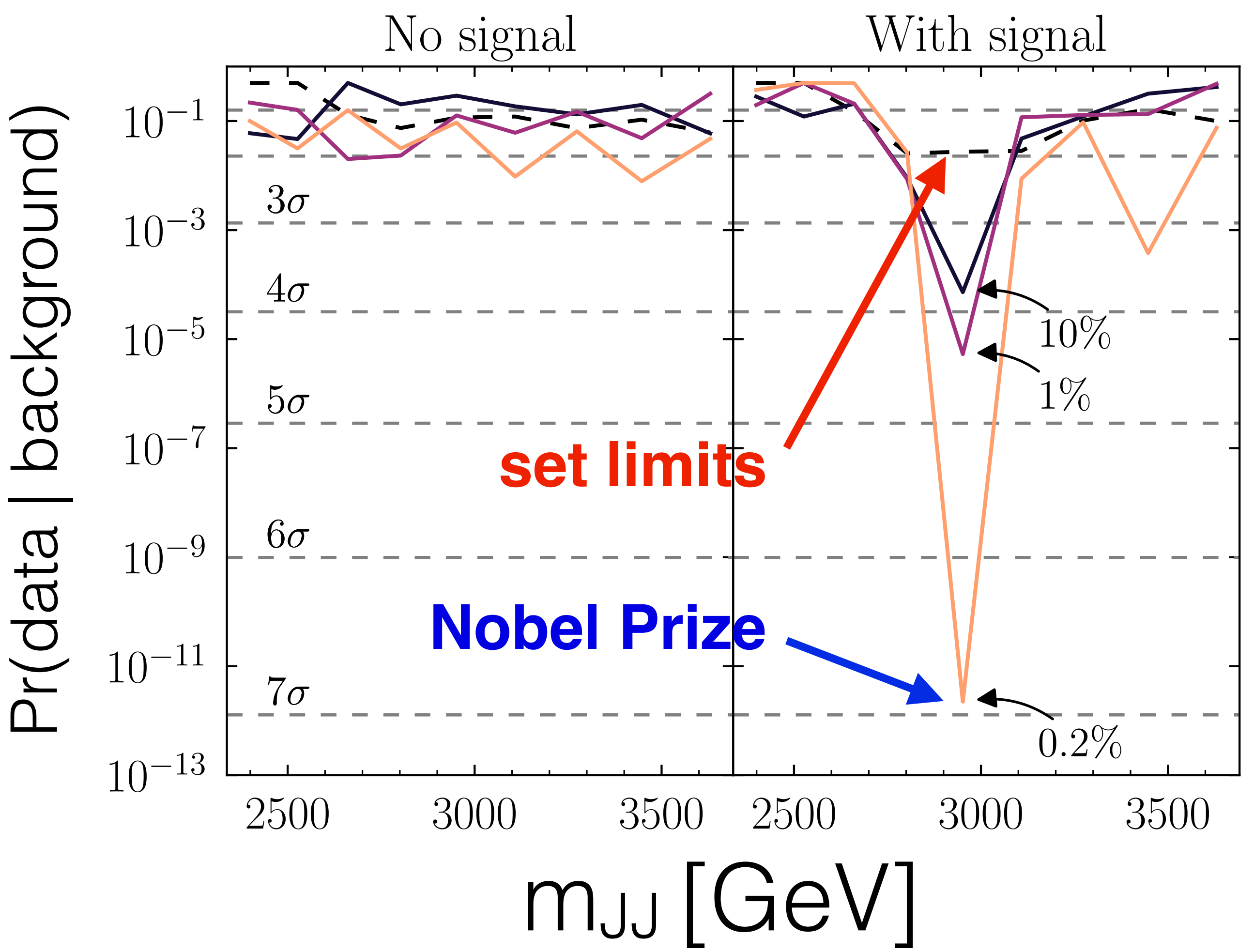


In my opinion, this is **THE** biggest challenge with deploying NN-based analyses … solving it will require hard physics work.

How to get around high-D bias uncerts?

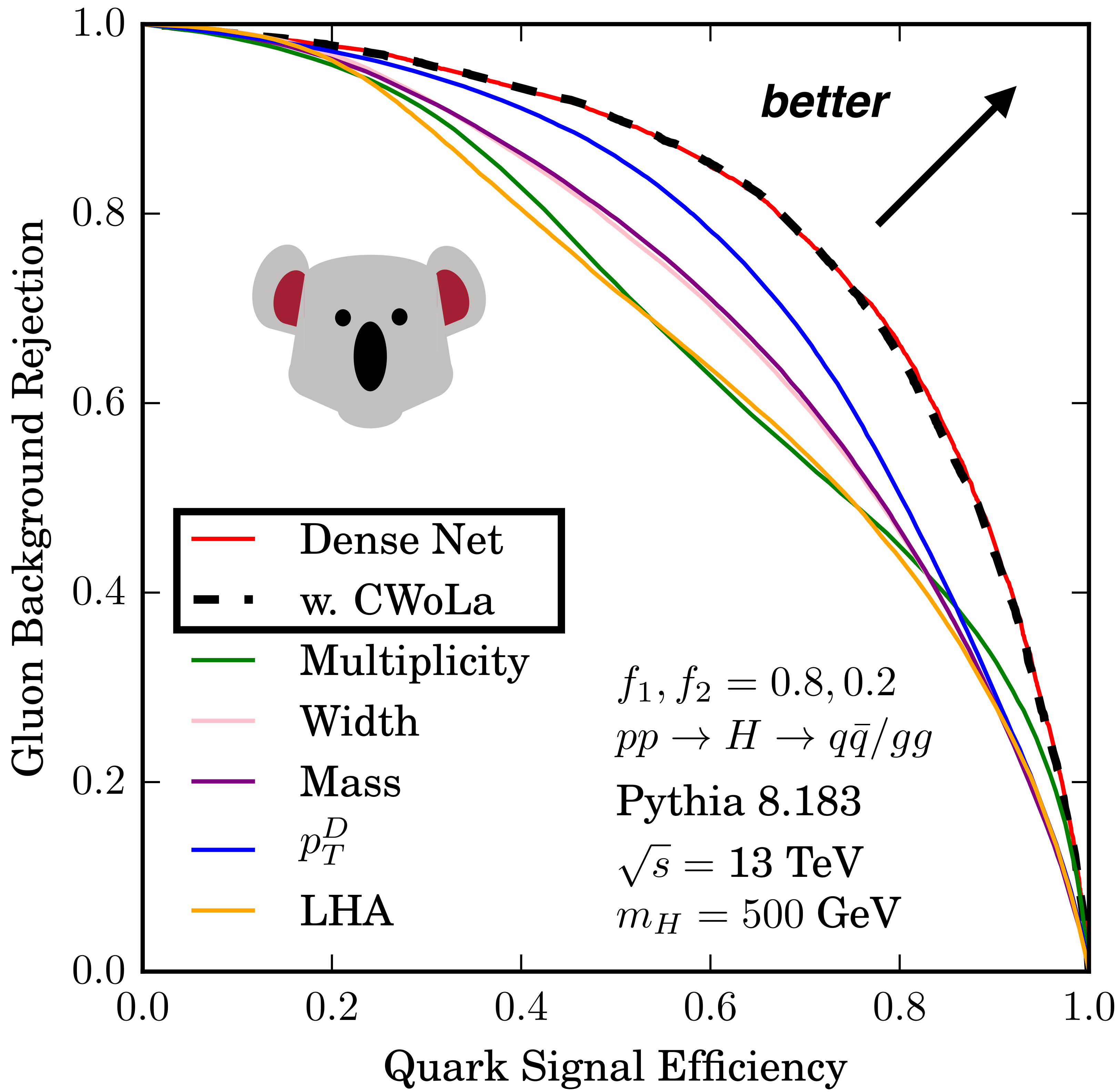
Work hard to understand the true nuisance parameters in the hypervariate parameter space.

Don't use simulation!
(not always possible!)



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

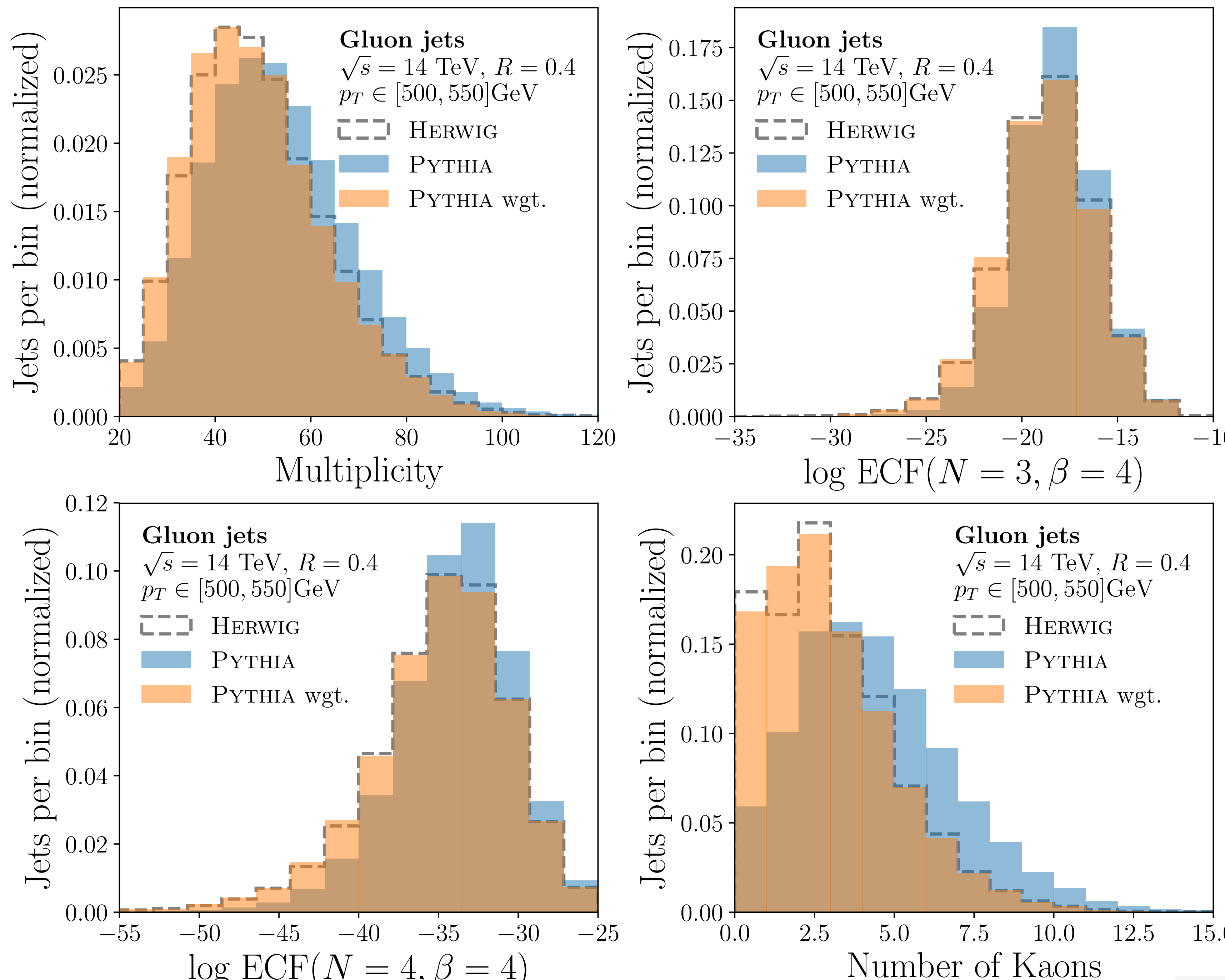
Works!



1.0
0.8
0.6
0.4
0.2
0.0
Gluon Background Rejection
0.0
0.2
0.4
0.6
0.8
1.0
Quark Signal Efficiency
better
Dense Net
w. CWoLa
Multiplicity
Width
Mass
p_T^D
LHA
$f_1, f_2 = 0.8, 0.2$
$pp \to H \to q\bar{q}/gg$
Pythia 8.183
$\sqrt{s} = 13$ TeV
$m_H = 500$ GeV

Pythia versus Herwig

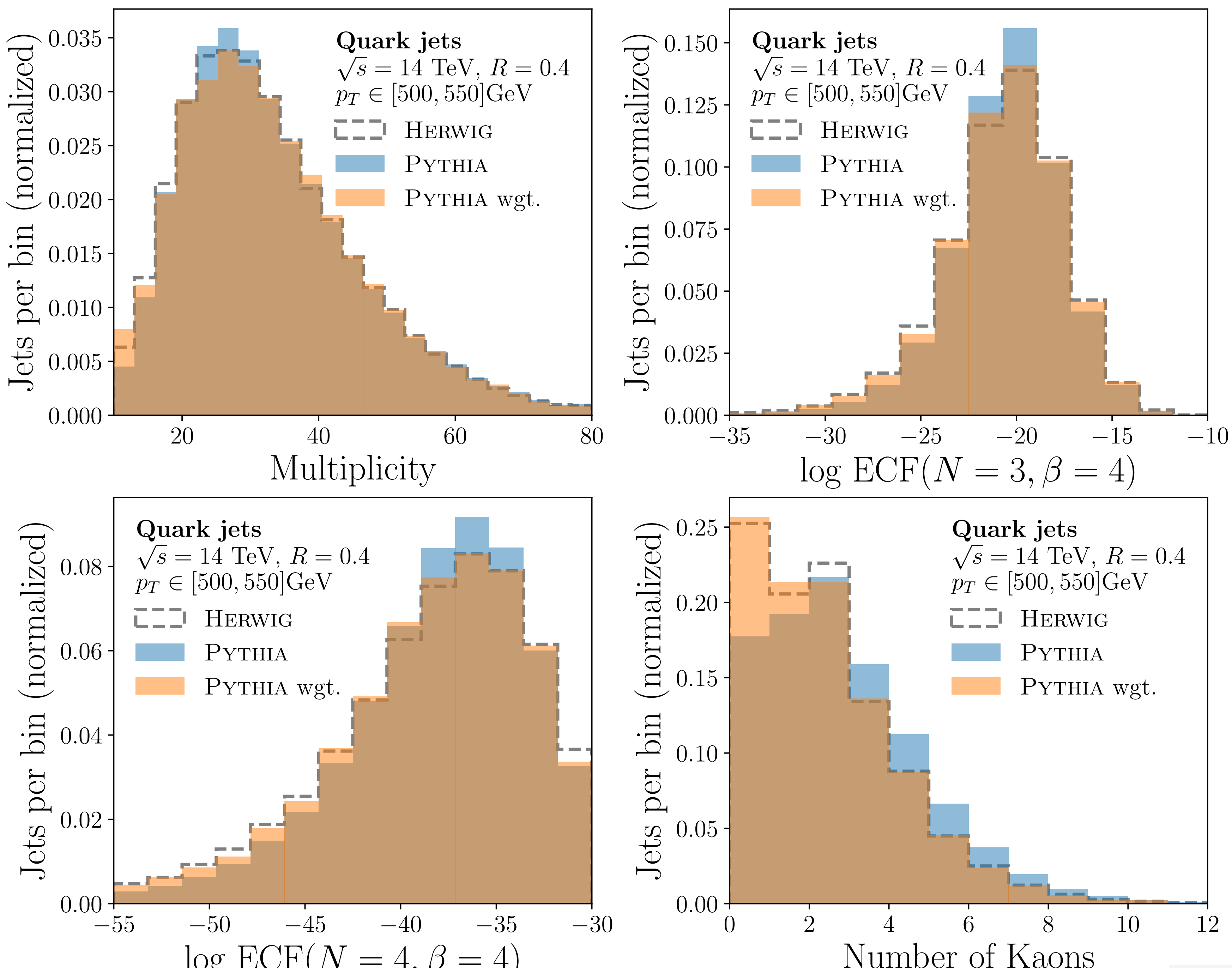
No hyper-parameter tuning - out of the box!



Samples from
10.5281/zenodo.2658764
10.5281/zenodo.3164691

Pythia versus Herwig

No hyper-parameter tuning - out of the box!



Samples from
10.5281/zenodo.2658764
10.5281/zenodo.3164691

More weak supervision: *Learn with proportions*

This is essentially a generalization of the template method.

Standard supervised learning

# of examples for training

loss fcn.

labels

$$f_{\text{full}} = \text{argmin}_{f':\mathbb{R}^n \to [0,1]} \sum_{i=1}^{N} \ell\left(f'(x_i) - t_i\right)$$

neural network

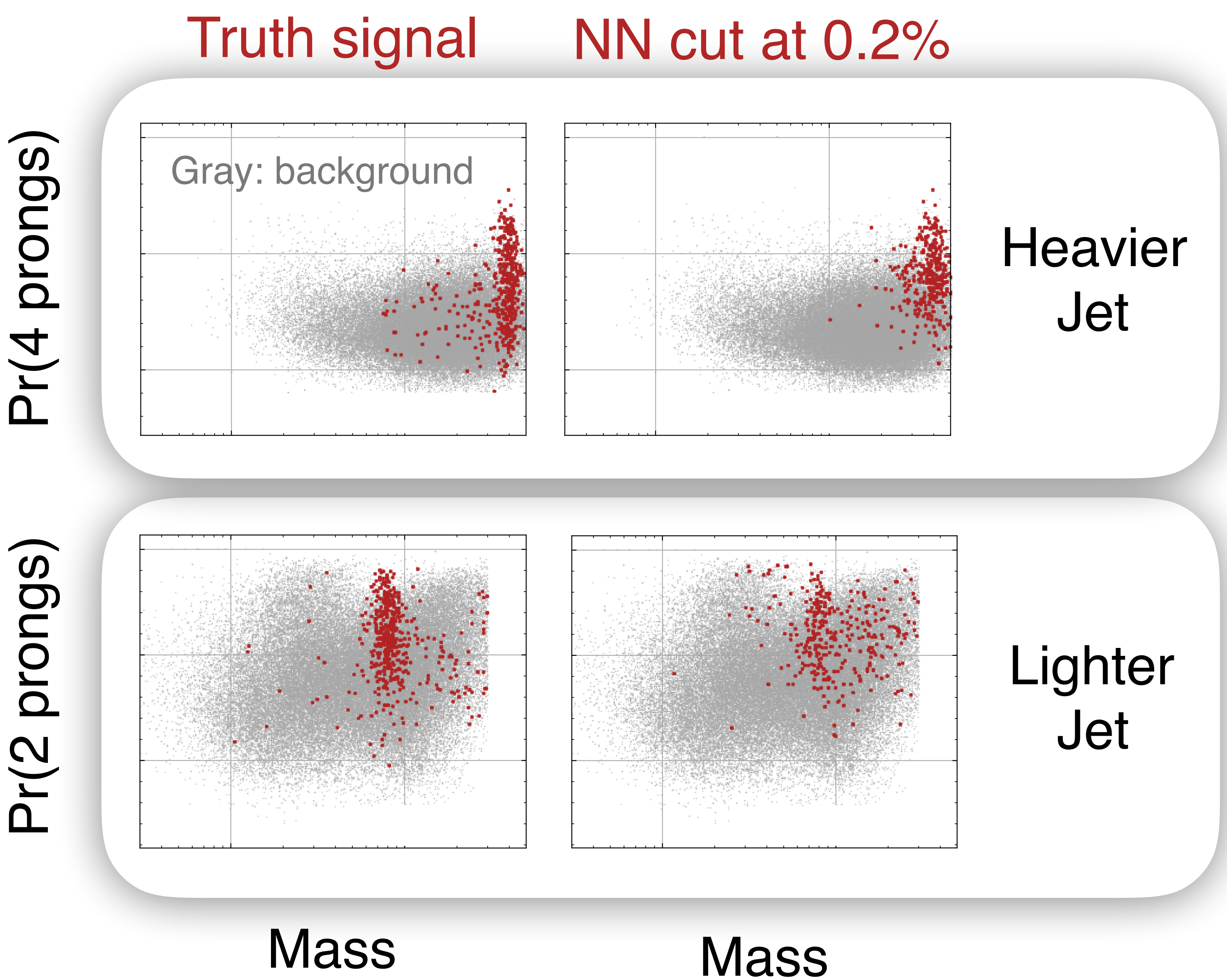
Weakly supervised learning

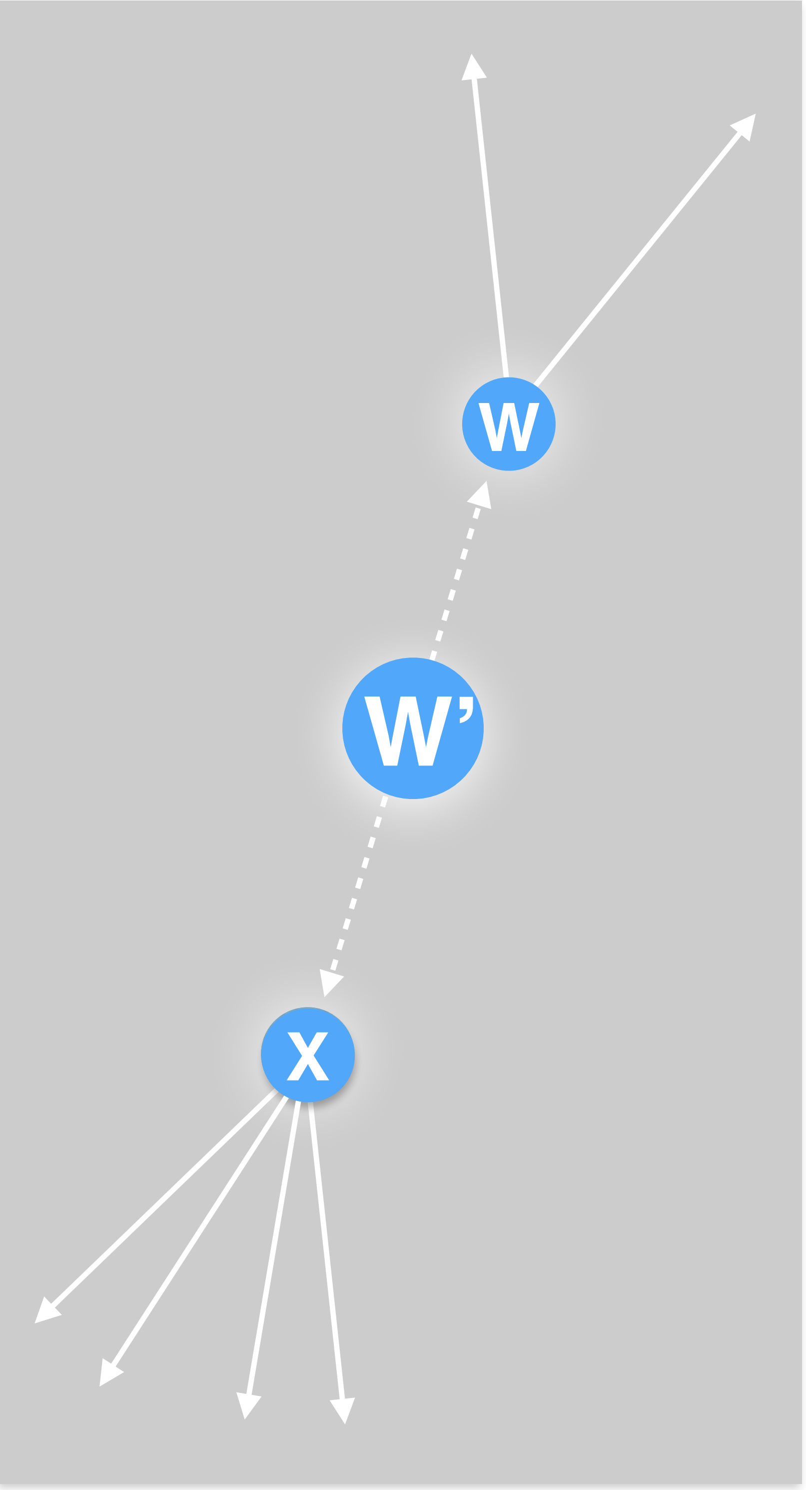
$$f_{\text{weak}} = \text{argmin}_{f':\mathbb{R}^n \to [0,1]} \ell\left(\sum_{i=1}^{N} \frac{f'(x_i)}{N} - y\right)$$

proportions

L. Dery, **BPN**, F. Rubbo, A. Schwartzman, JHEP 05 (2017) 145

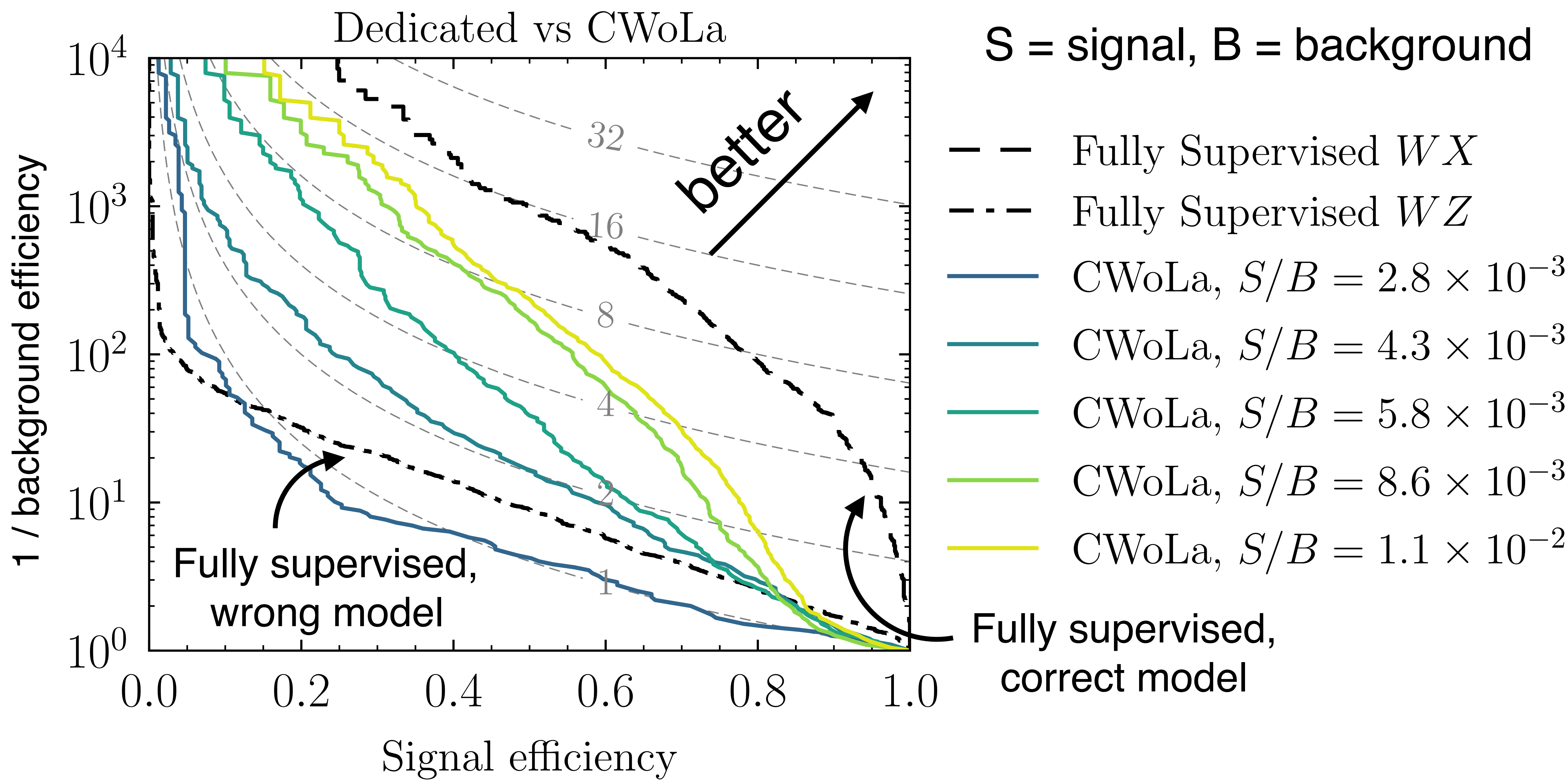
What is the network learning?





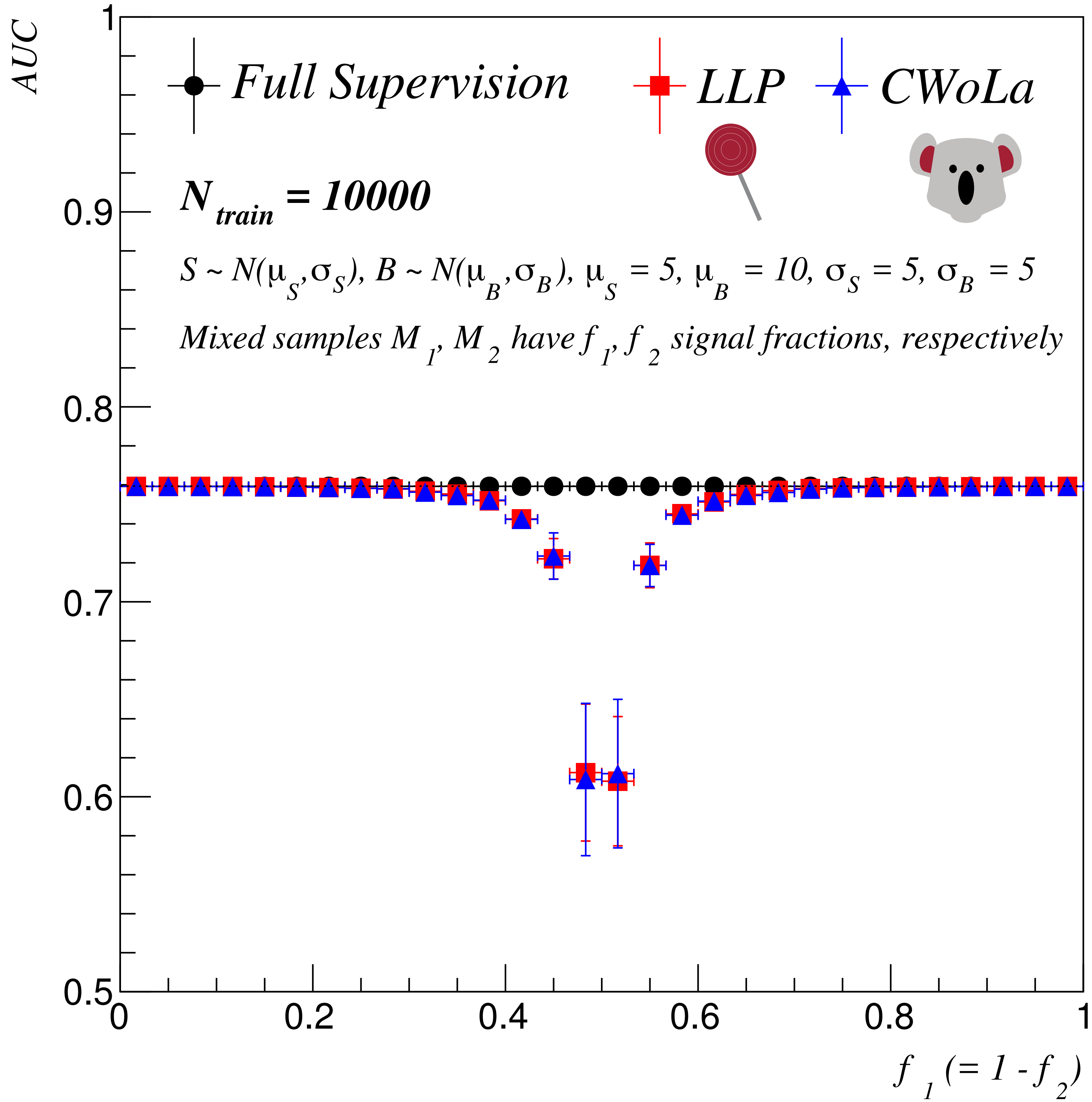
Learns to find the signal !

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!

A note about training statistics



signal fraction of mixed sample 1

Can't learn when the two proportions are the same.

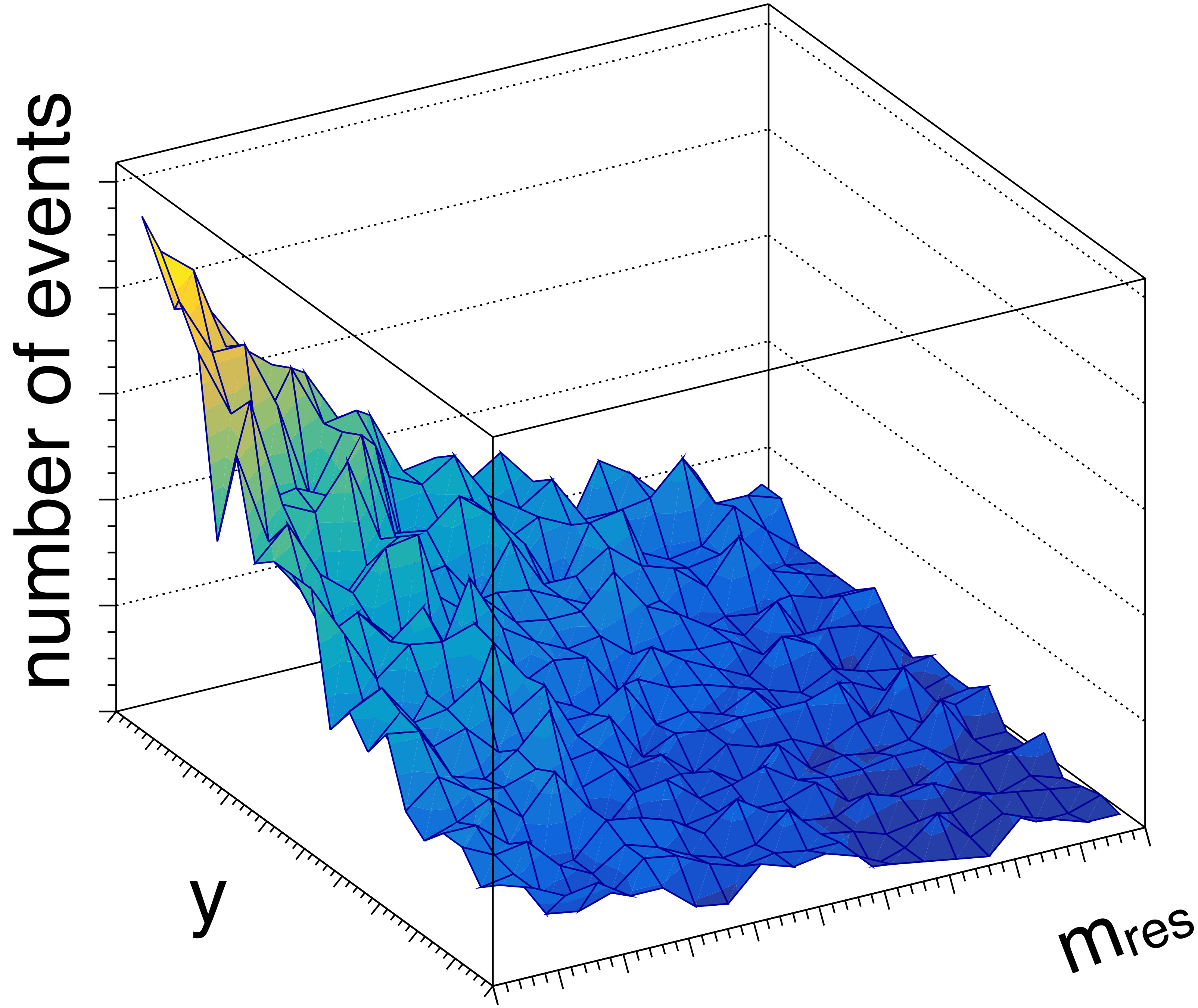
The more similar they are, the worse the performance.

(lower effective statistics)

Overtraining & Look Elsewhere Effect*

Naively, pay a huge penalty because y can be high-dimensional.

i.e. you will sculpt lots of bumps!

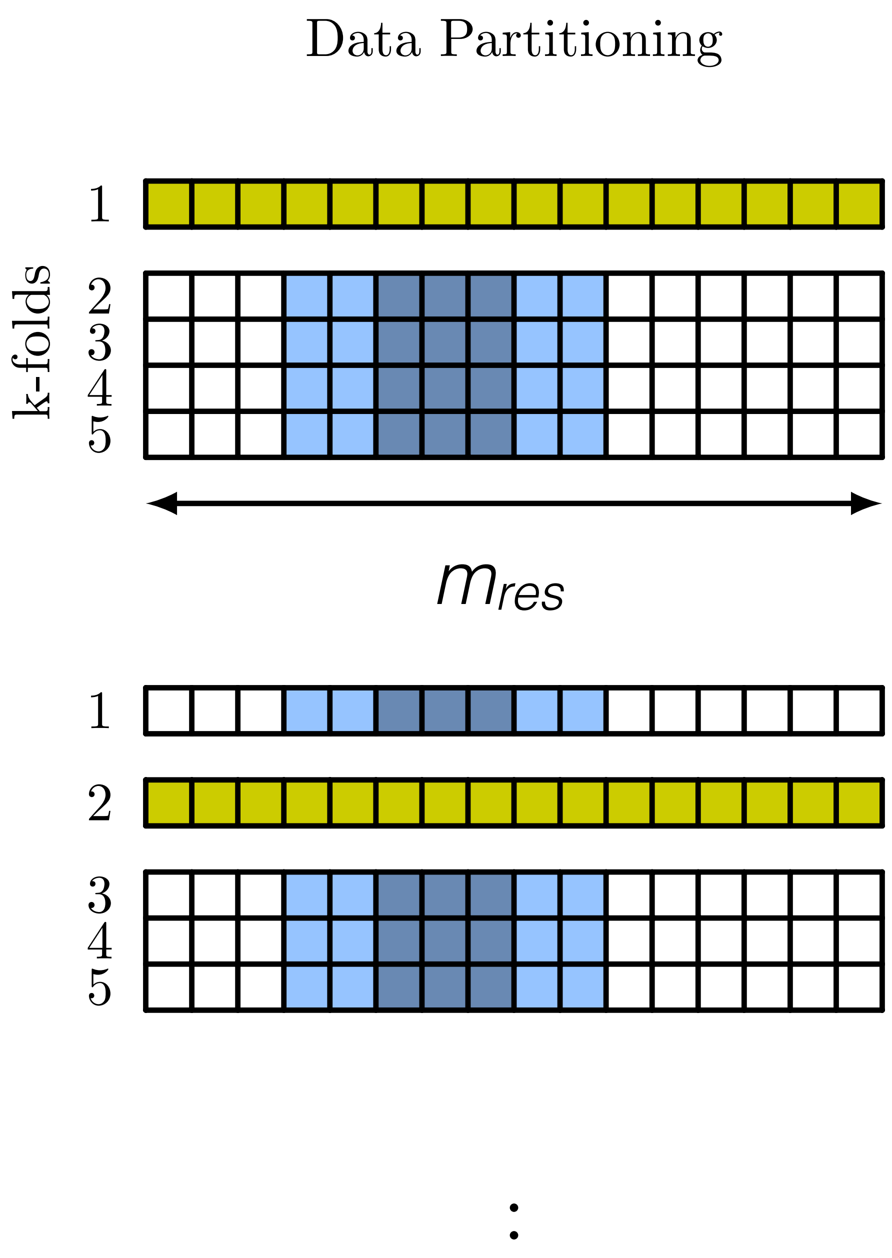


Solution: **(nested) cross-training**

*you may know this as the multiple comparisons problem

Nested cross-training

(1) Divide the entire dataset into k-folds.



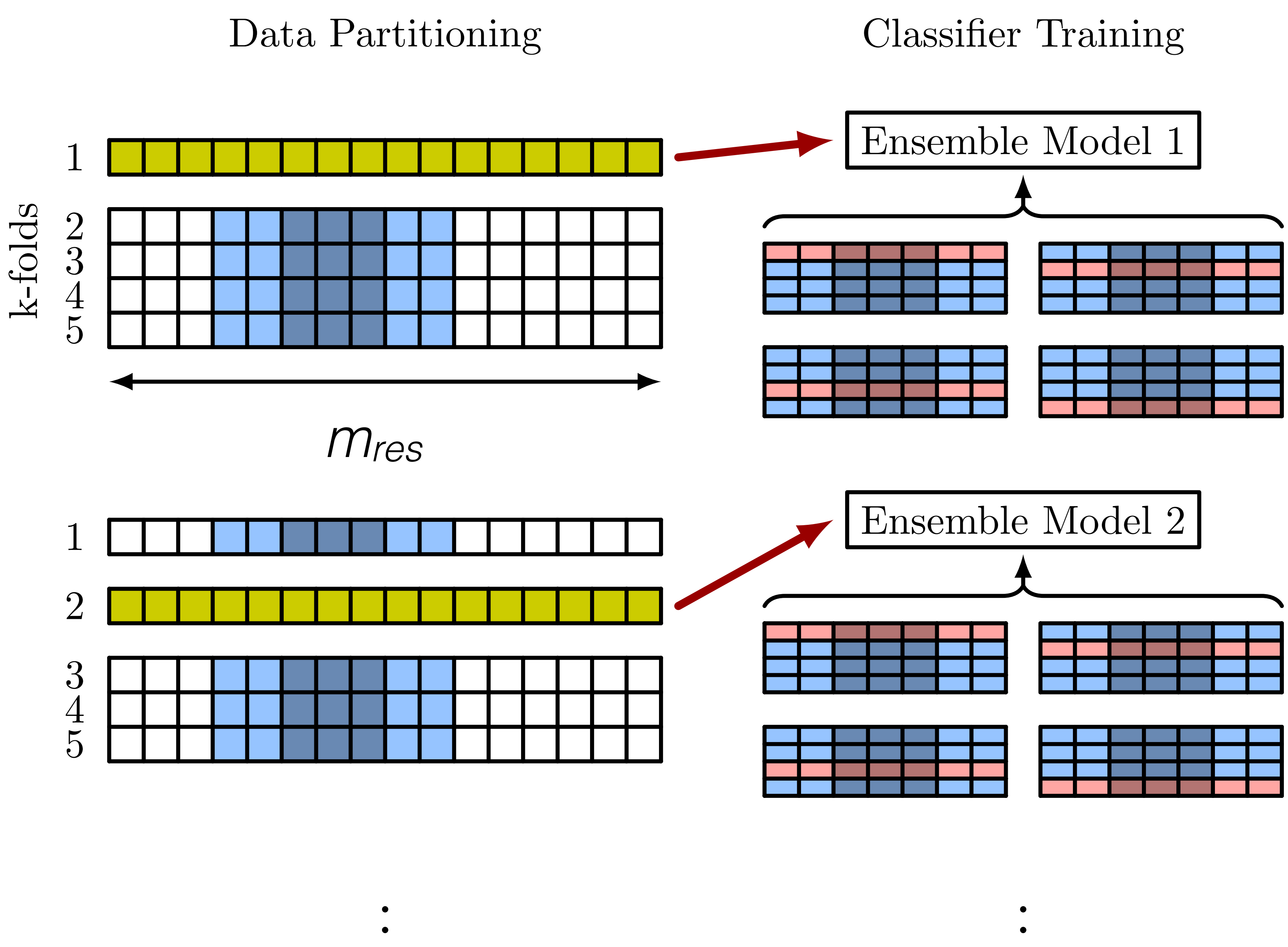
test set

signal region

sidebands

Nested cross-training

(2) Train CWoLa classifiers.

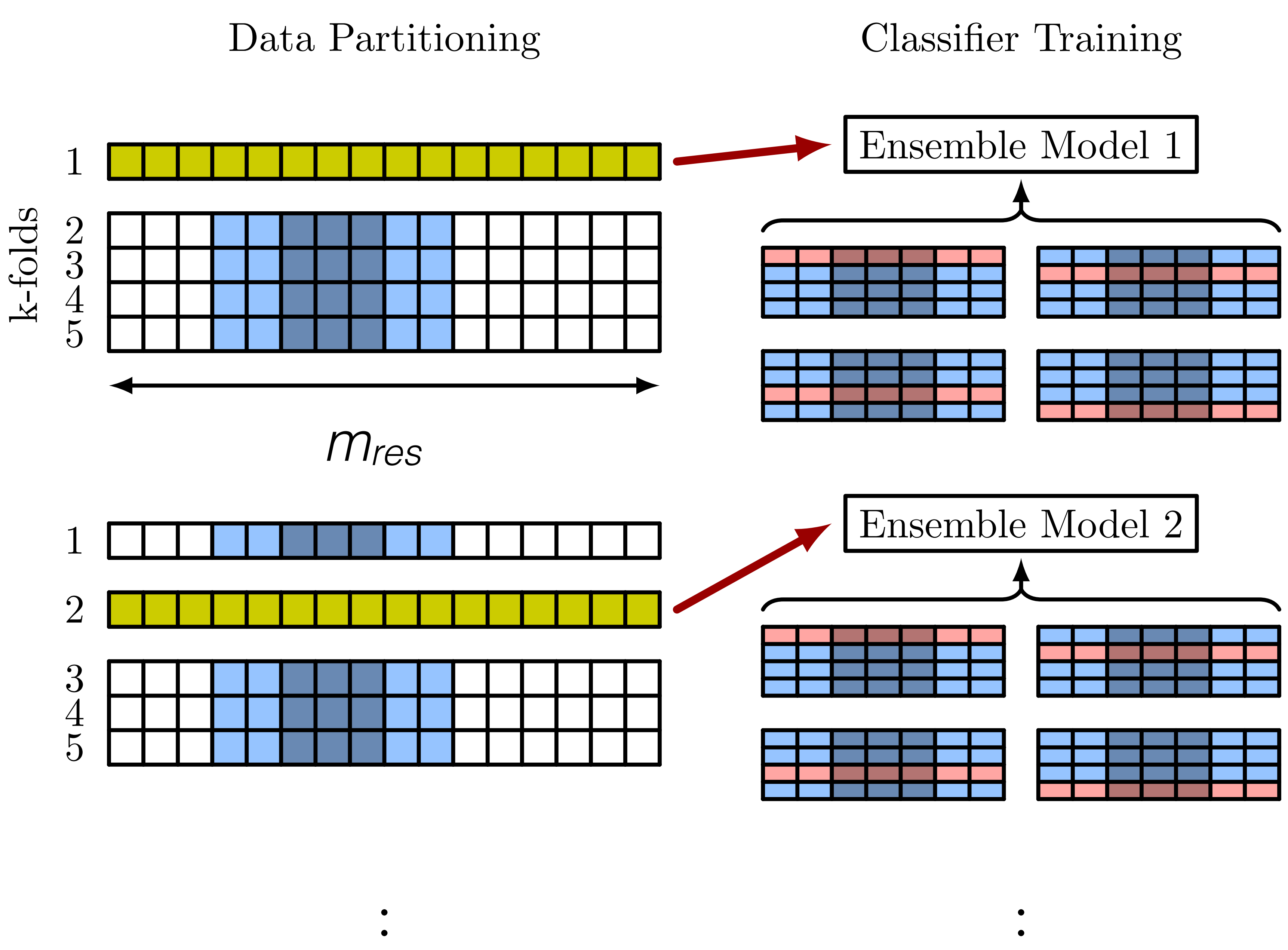


Train signal versus sideband k-1 times rotating the validation set

In practice, also train many networks with a different initialization and take the best one per k-1 rotation.

Nested cross-training

(2) Train CWoLa classifiers.

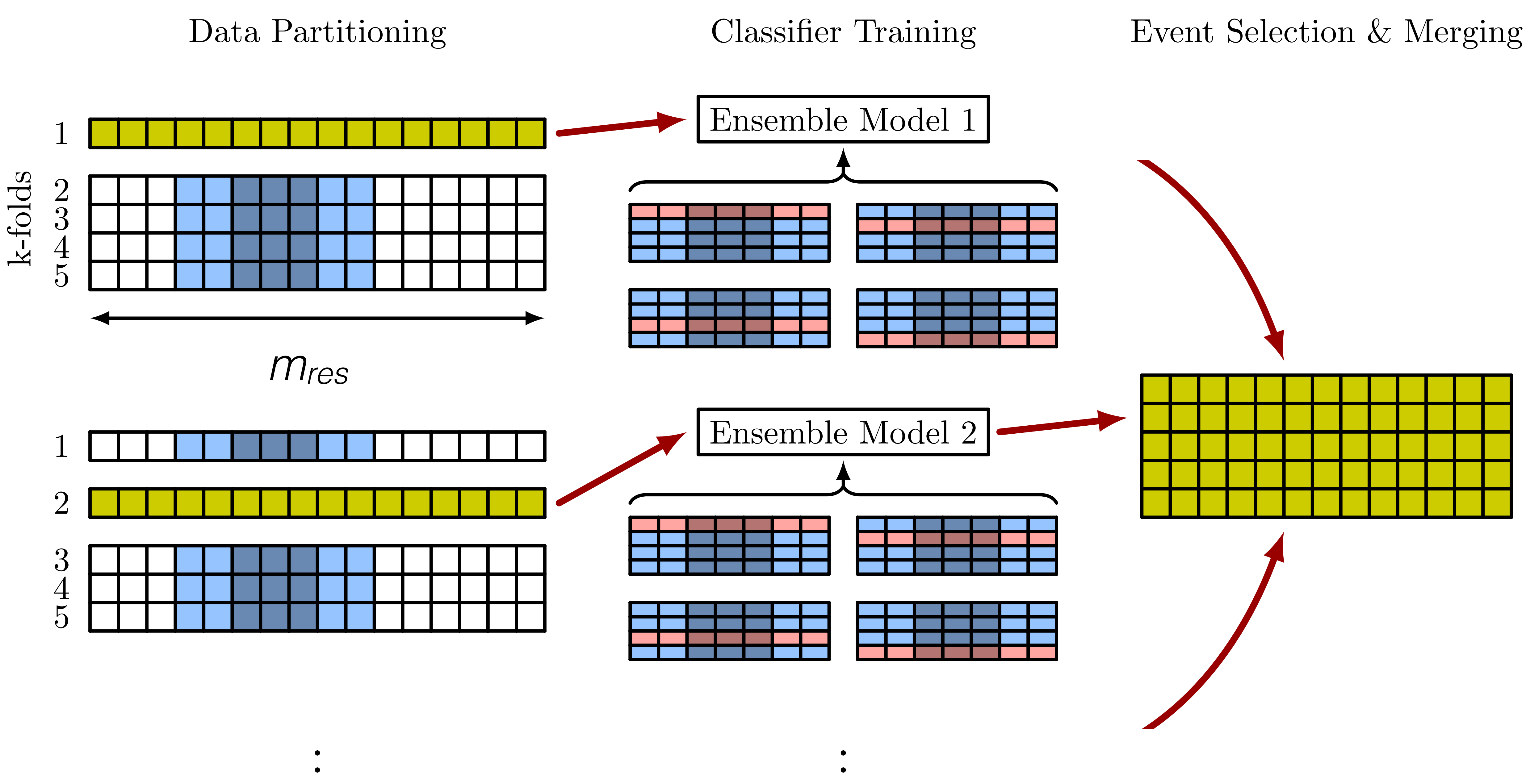


The Ensemble Model is just the average of the four networks.

Data fluctuations will cancel destructively while signal interferes constructively.

Nested cross-training

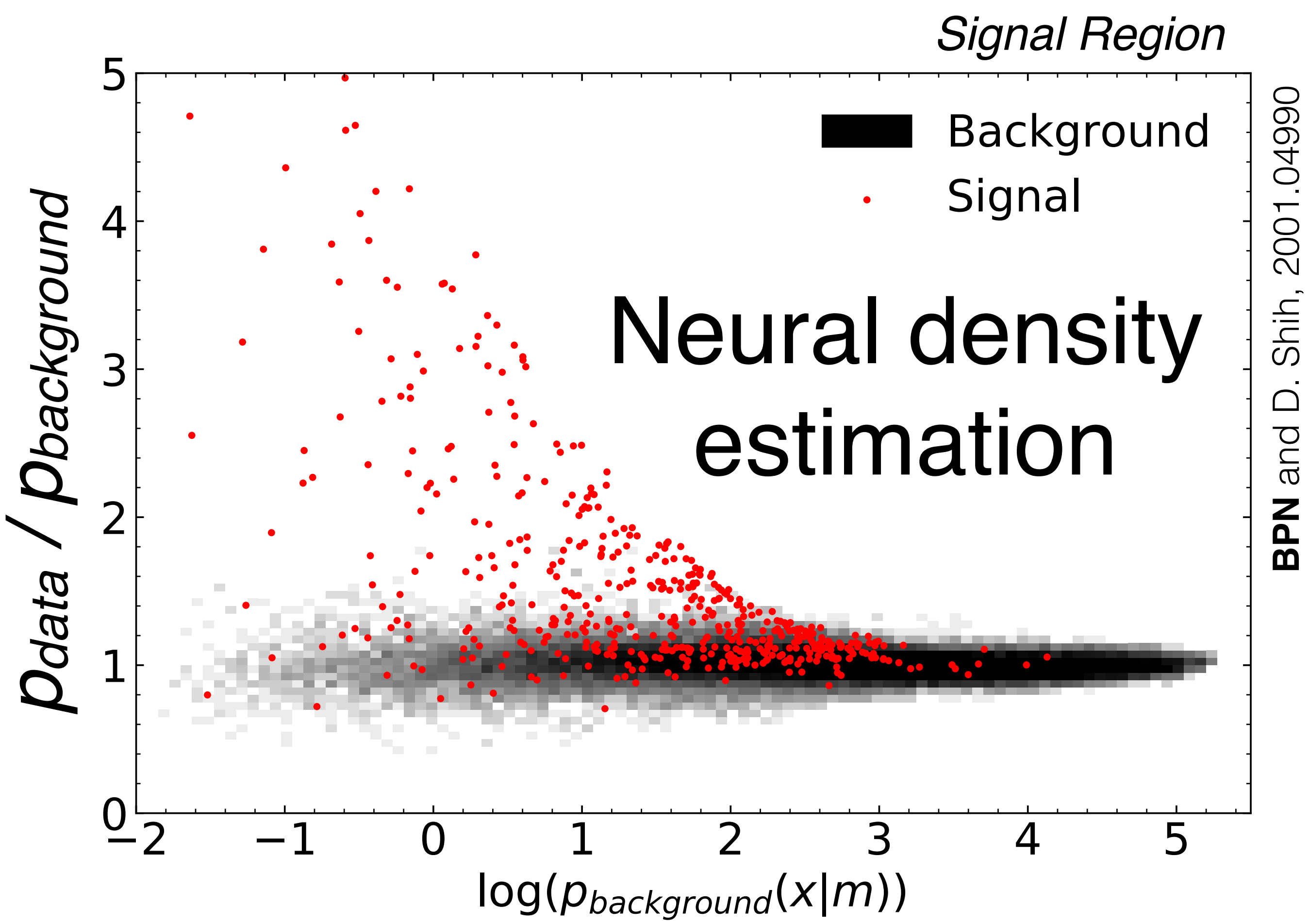
(3) Apply classifiers to holdout test sets and sum.

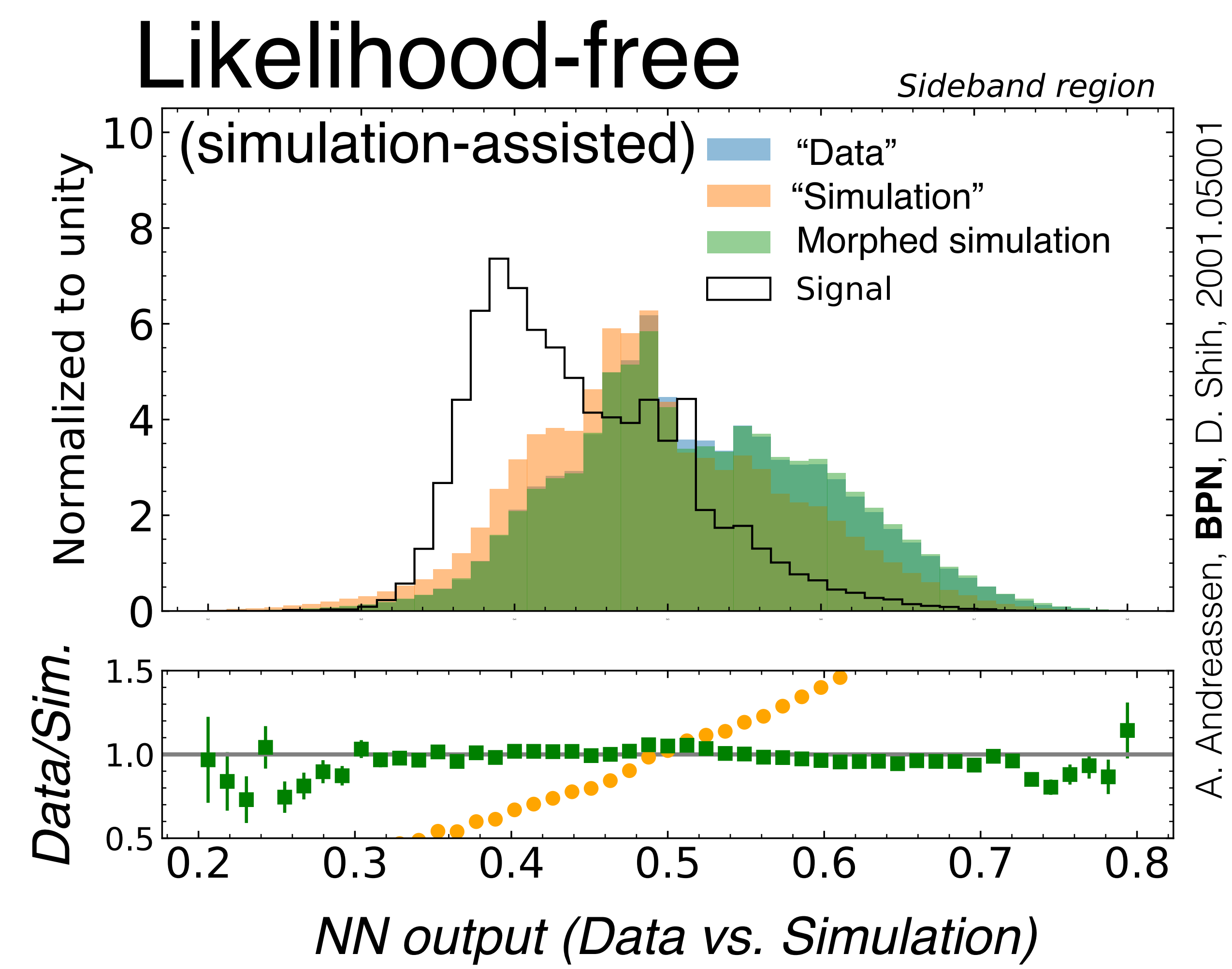


Extend techniques: Anomaly detection

CWoLa hunting was the first of a growing number of new proposals.

e.g., CWoLa hunting doesn't work when there are strong correlations between y and m_{res}. New methods solve this!





Tying it all together: DCTR + anomalies

New

We want to reduce our dependence on simulation, but we also don't want to throw away our physics priors!

Tying it all together: DCTR + anomalies

We want to reduce our dependence on simulation, but we also don't want to throw away our physics priors!

Can we use simulations in a way that is (nearly) simulation-independent?

New

Tying it all together: DCTR + anomalies

We want to reduce our dependence on simulation, but we also don't want to throw away our physics priors!

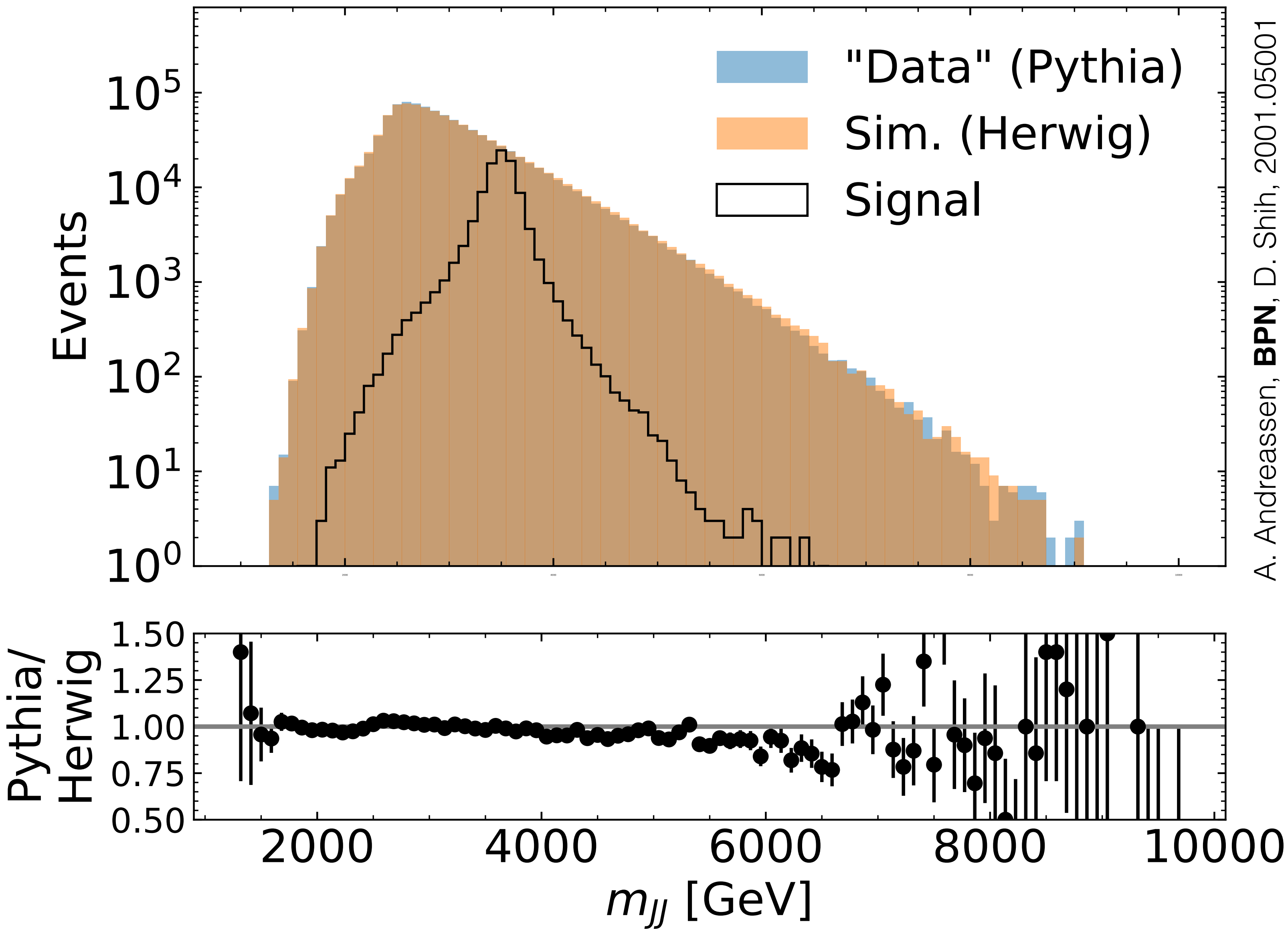
Can we use simulations in a way that is (nearly) simulation-independent?

Answer:

Simulation Assisted Likelihood-free Anomaly Detection

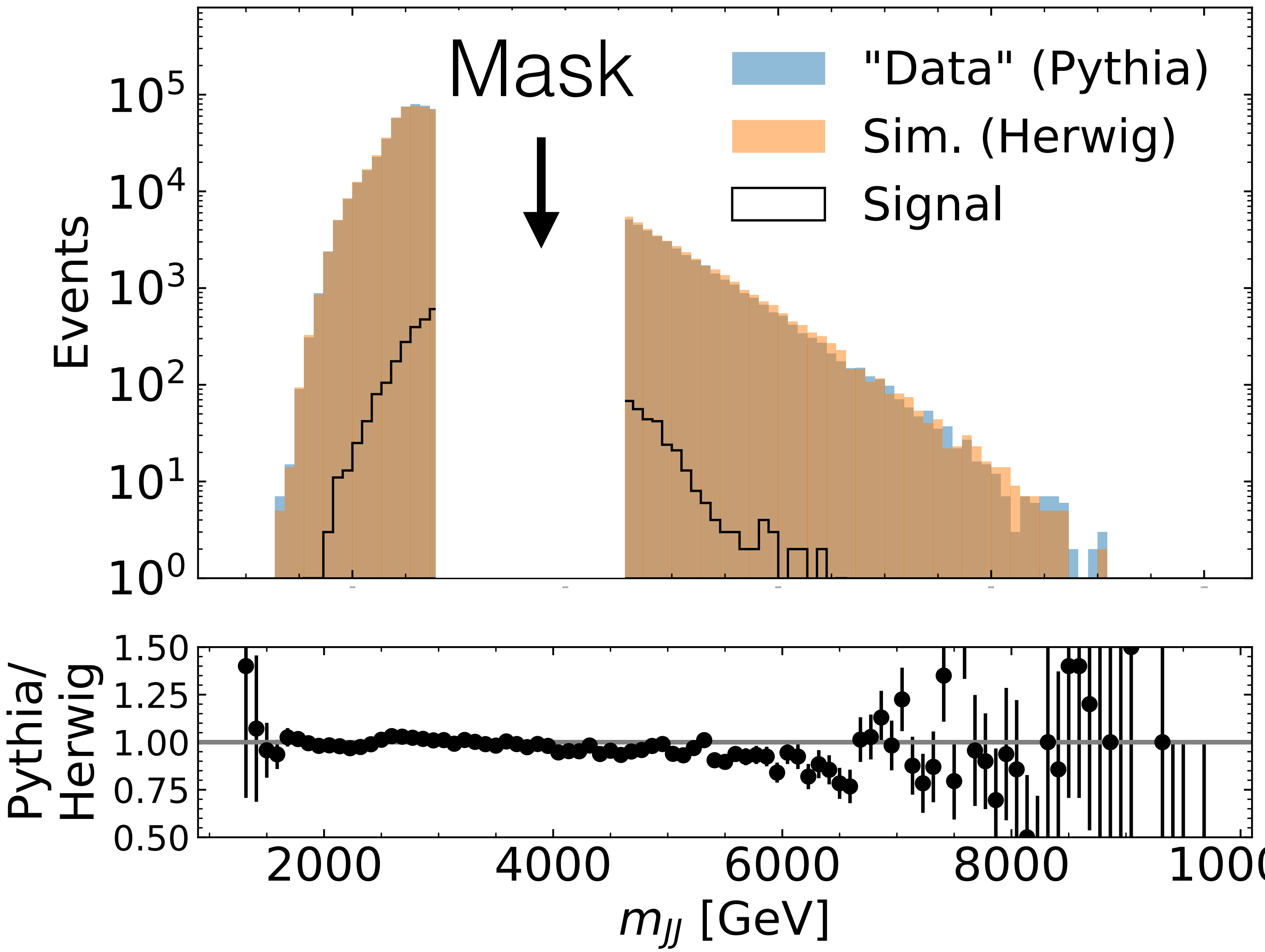
(aka SALAD)

SALAD



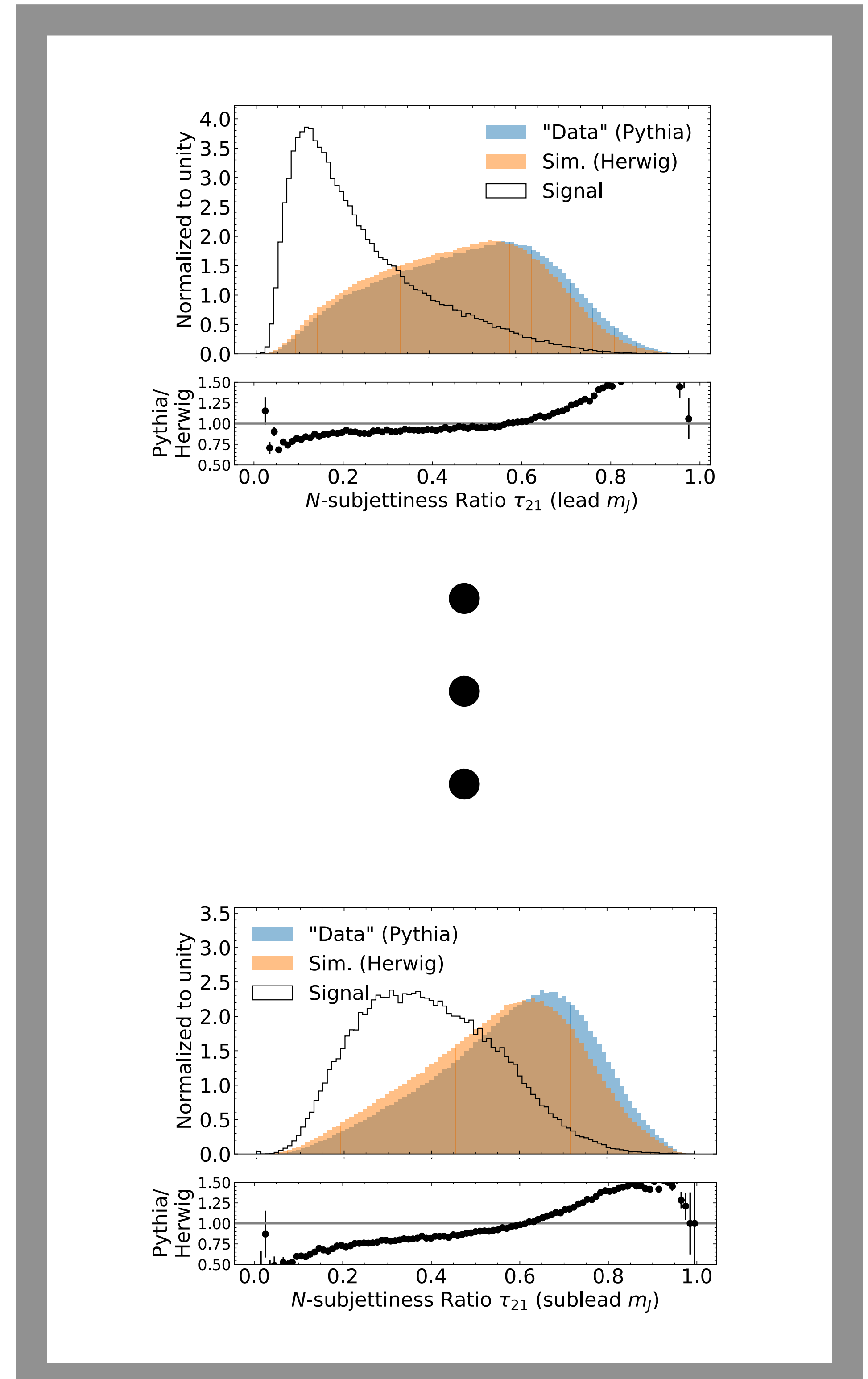
A. Andreassen, **BPN**, D. Shih, 2001.05001

SALAD



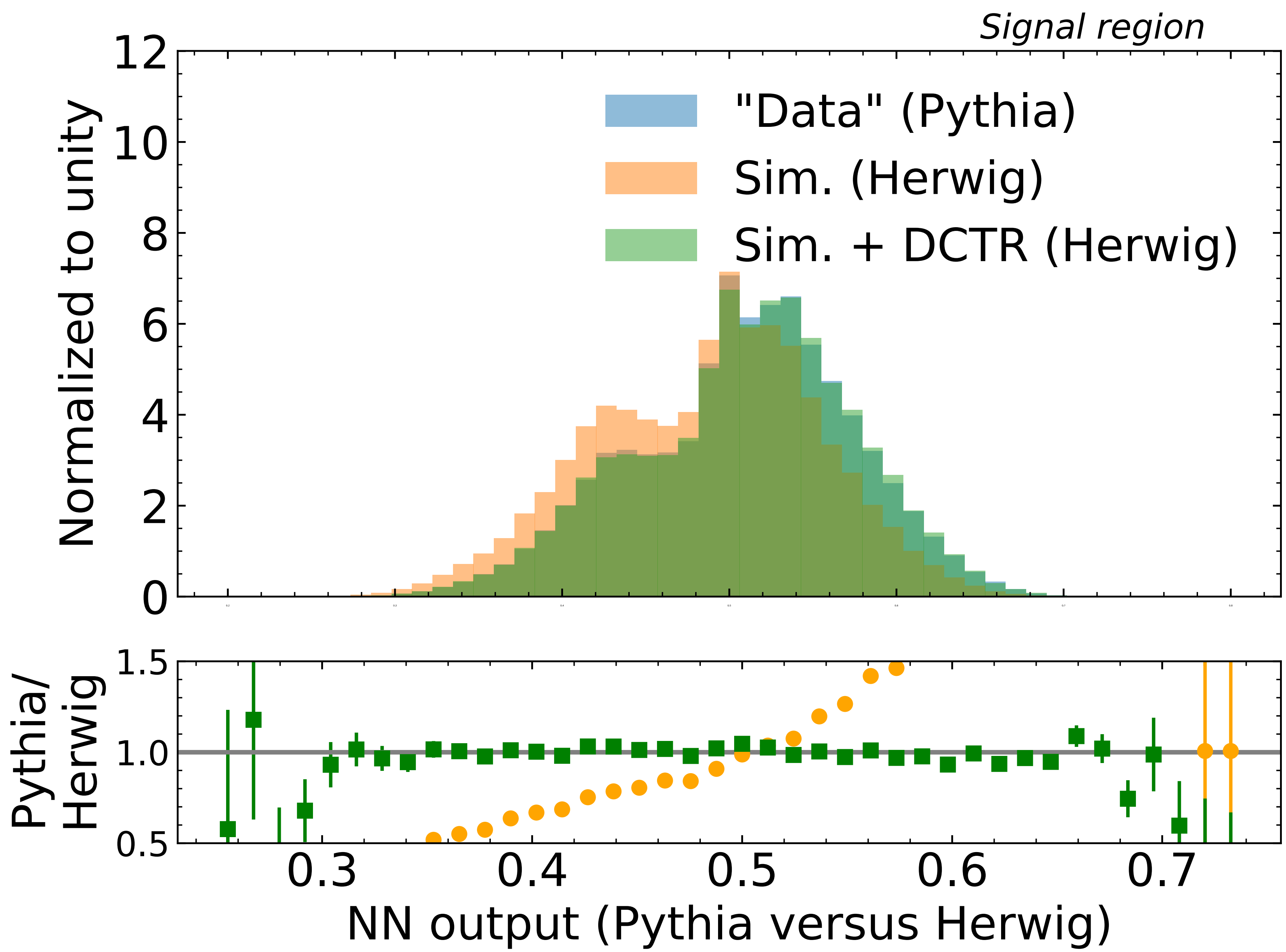
A. Andreassen, **BPN**, D. Shih, 2001.05001

Features



(1) Train DCTR in the sidebands to reweight MC to data

SALAD

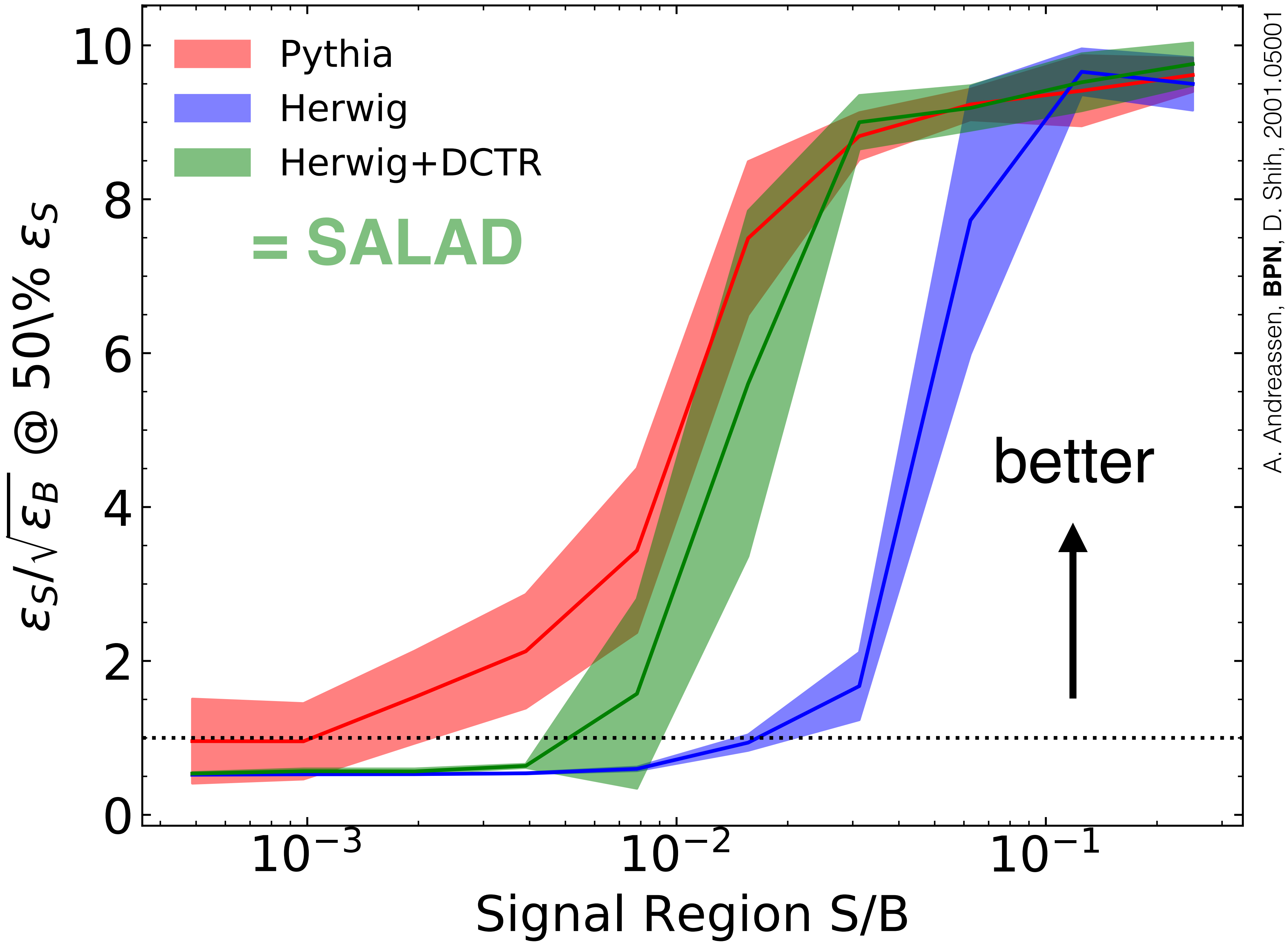


A. Andreassen, **BPN**, D. Shih, 2001.05001

(2) Interpolate DCTR to signal region

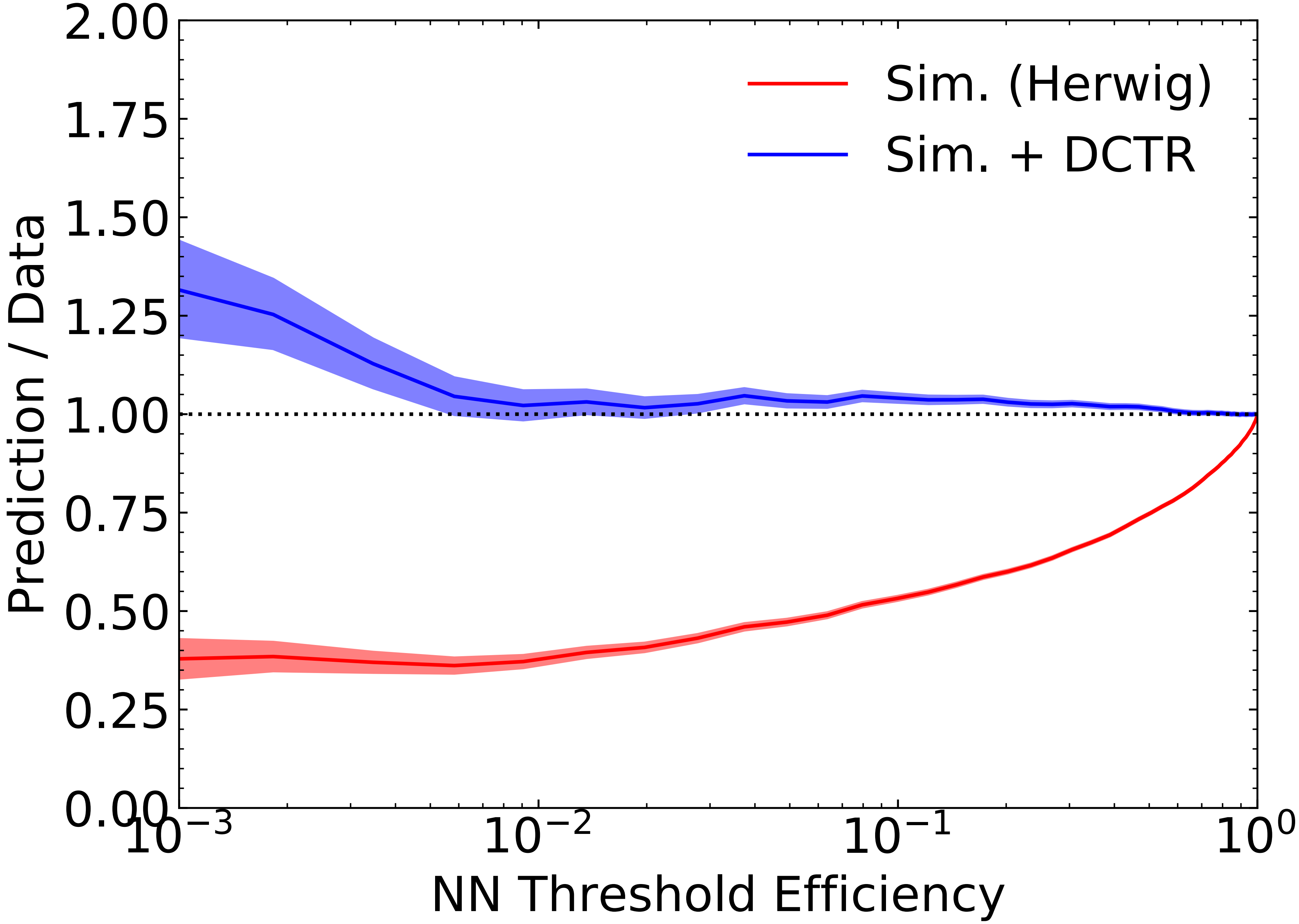
(3) Train classifier to distinguish reweighted MC from data

SALAD



Can also use SALAD for background estimation - see backup

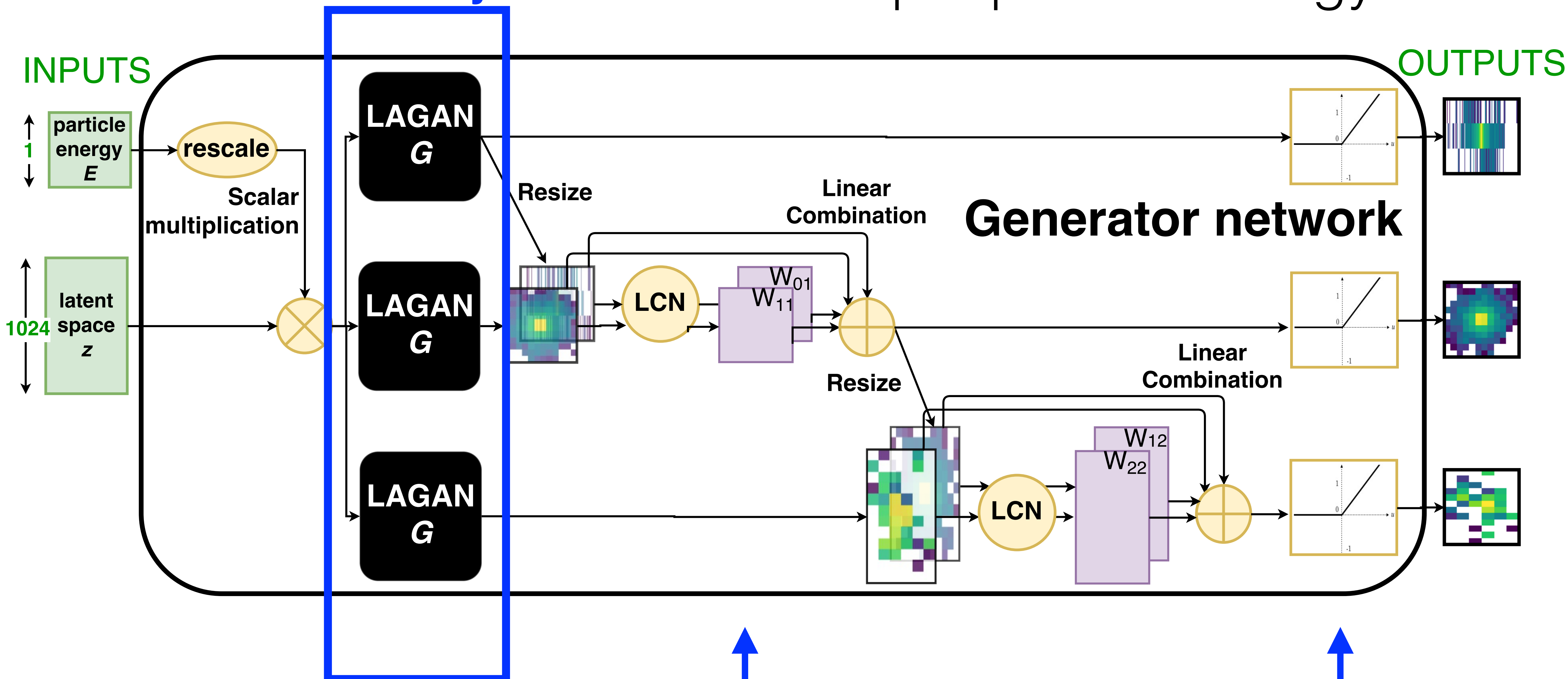
SALAD background



Introducing CaloGAN

One image per calo layer

One network per particle type; input particle energy



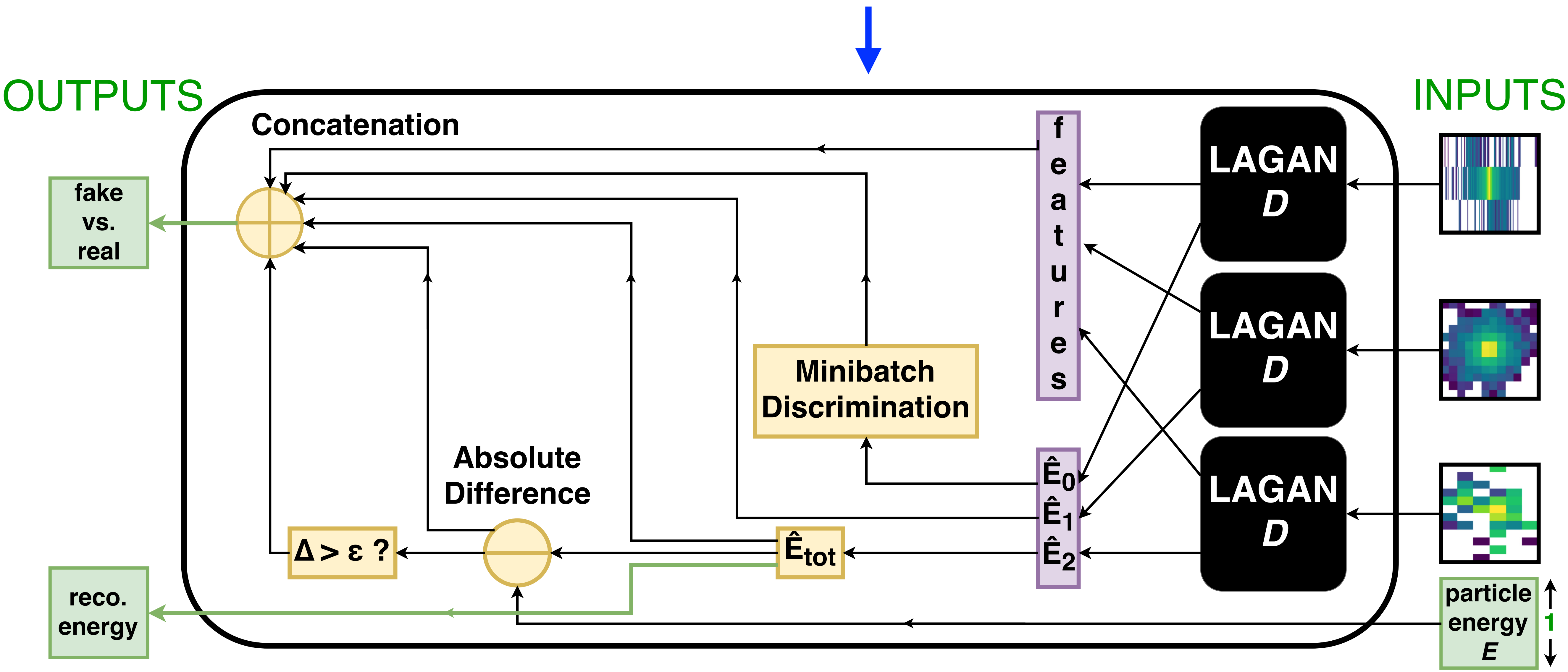
use layer i as input to layer i+1

ReLU to encourage sparsity

Introducing CaloGAN

Mode collapse: learns to generate one part of the distribution well, but leaves out other parts.

help avoid 'mode collapse'

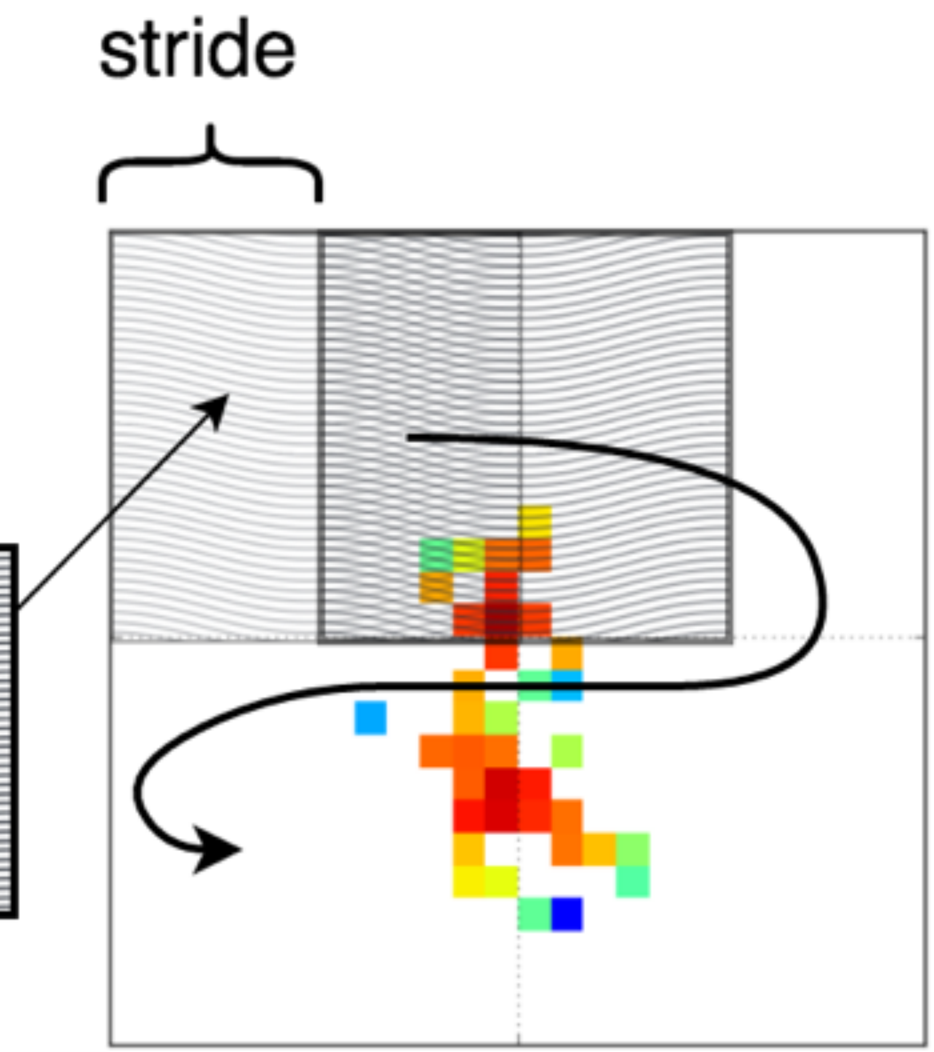


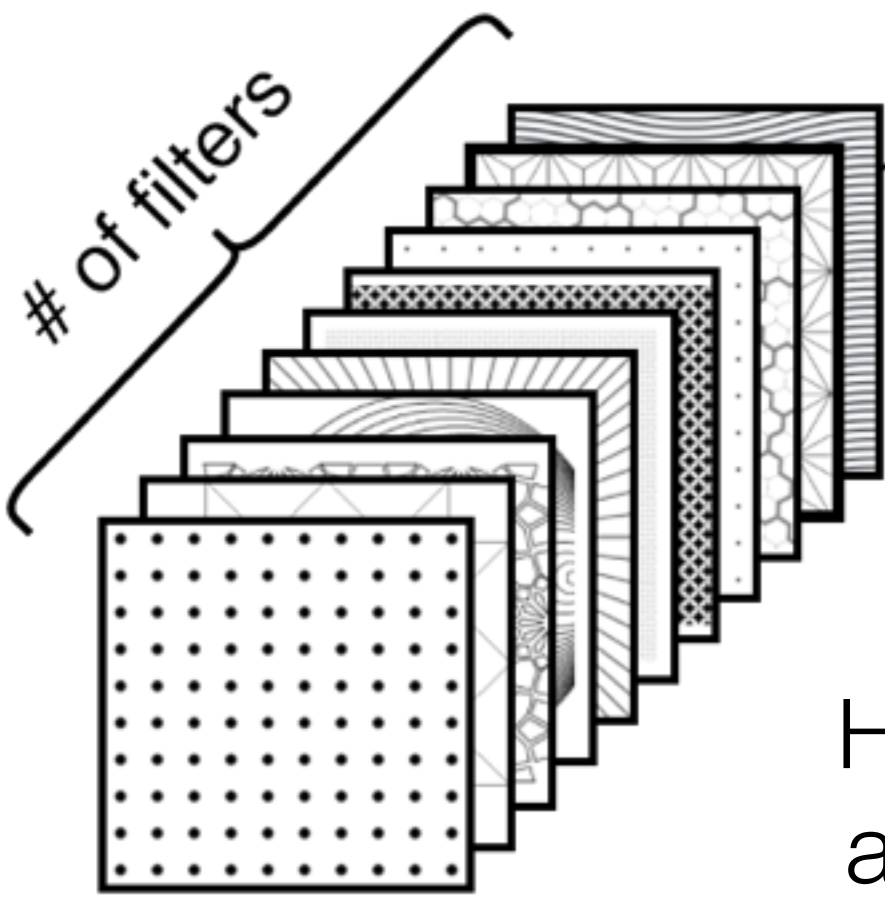
Discriminator network

Locally connected layers

Due to the structure of the problem,
we do not have translation invariance.

Classification studies found fully connected networks outperformed CNNs

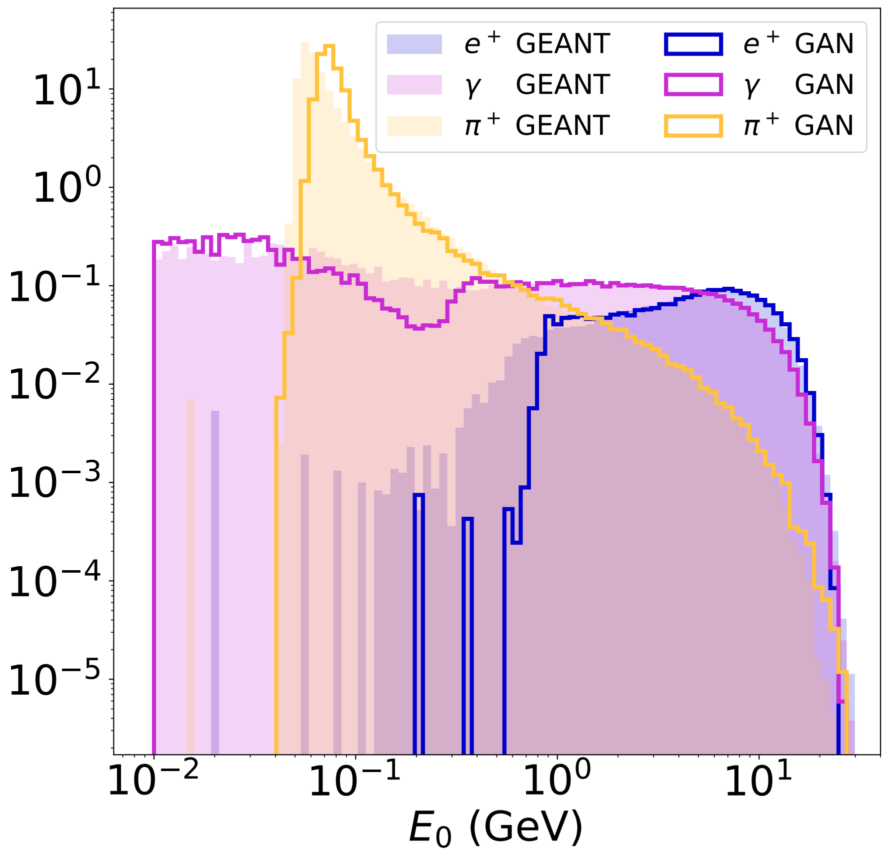


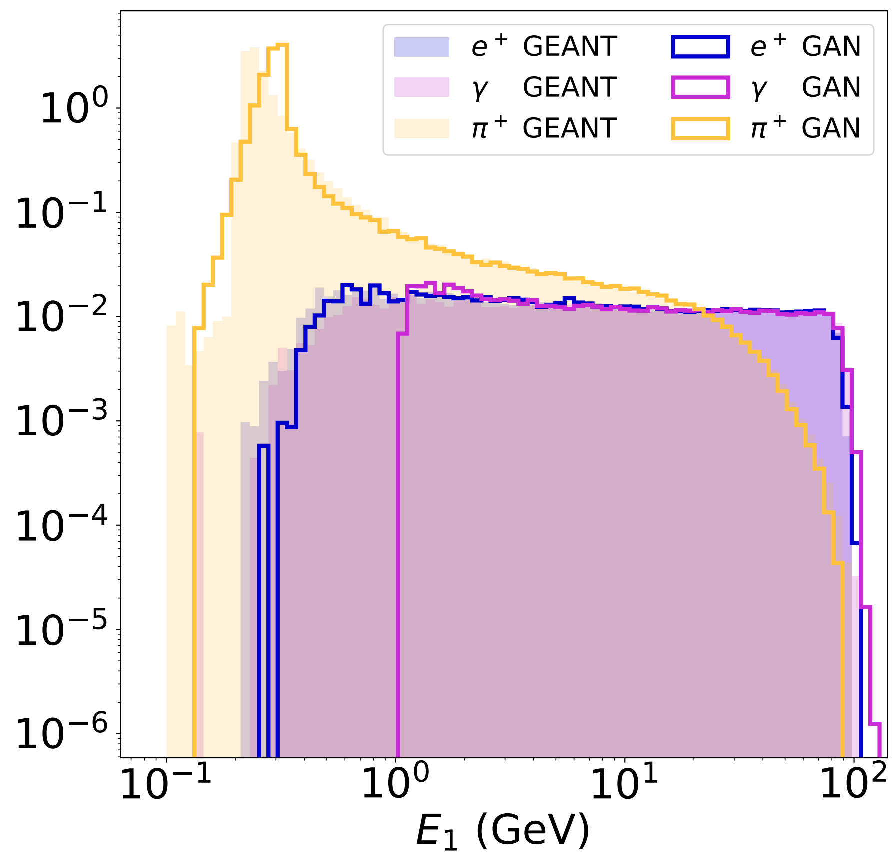


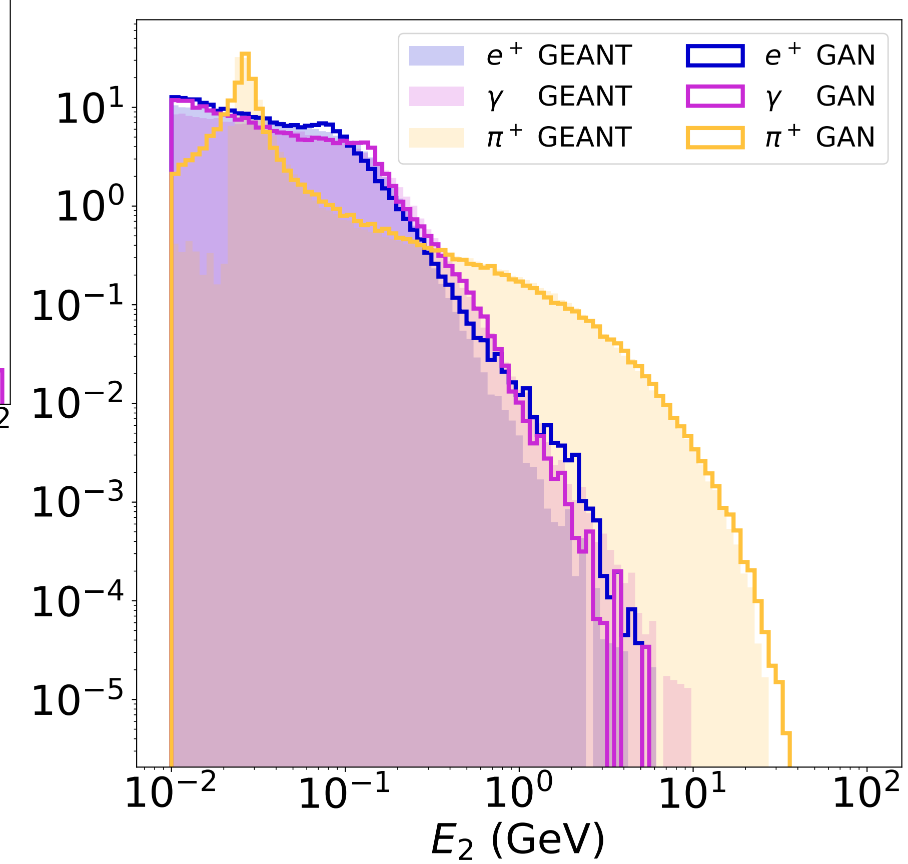
However, convolutional-like architectures
are still useful to e.g. reduce parameters

Energy per layer

Pions deposit much less energy in the first layers; leave the calorimeter with significant energy





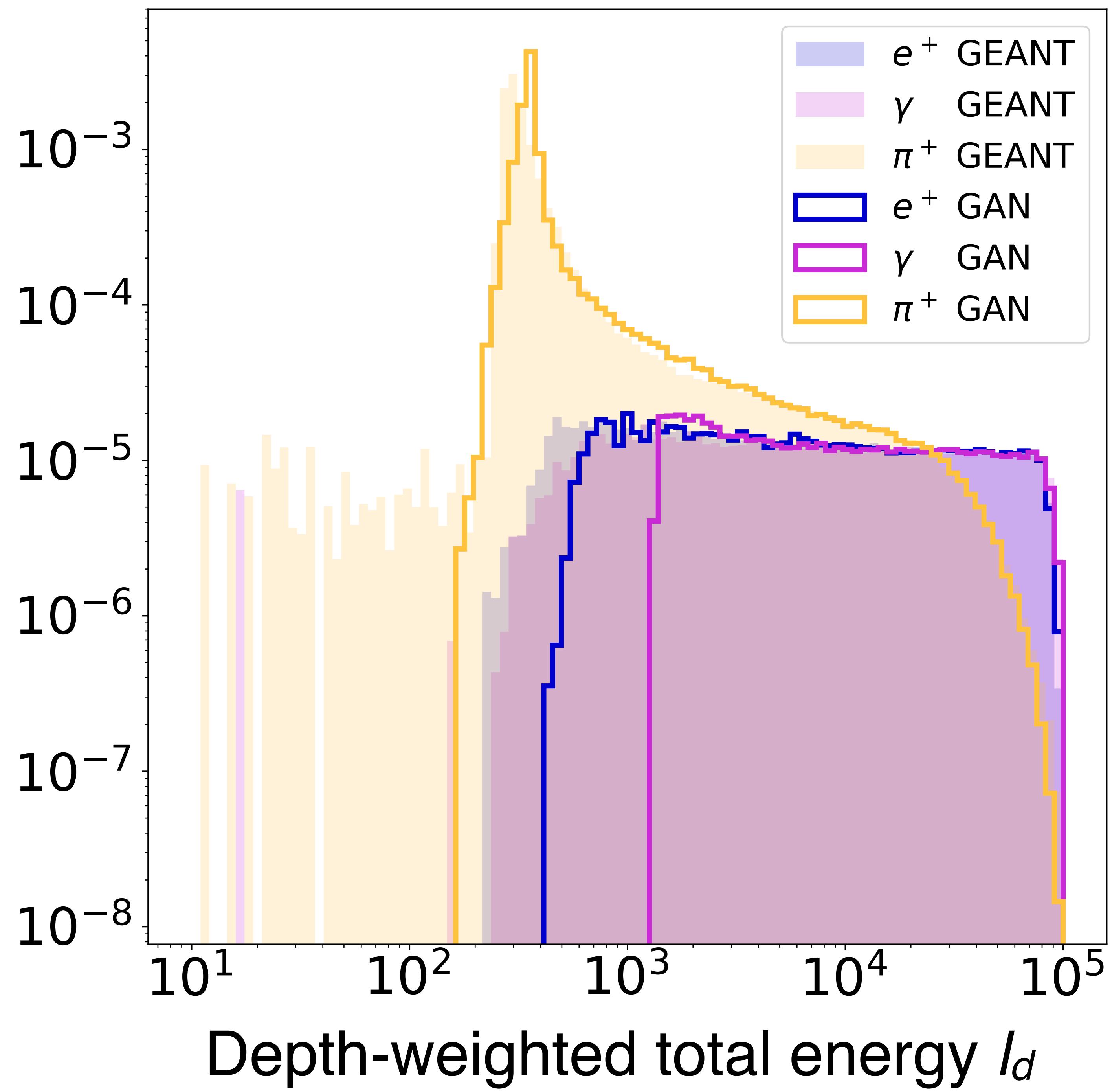


N.B. can always add these (and others) explicitly to the training

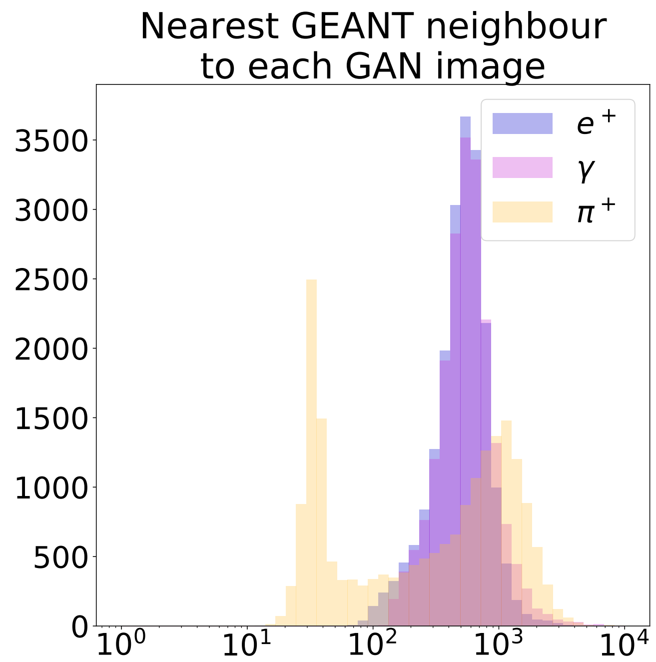
Warning: challenge with GANs

Unlike for classifiers, it is not easy to figure out which GAN is a good GAN - trying to learn a O(1000) generative model and not a single likelihood ratio!

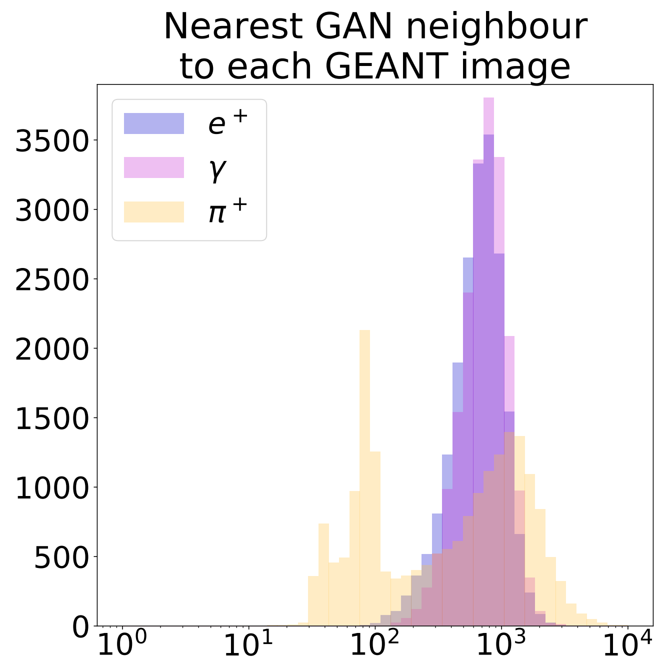
…this is a place where science applications can make a big impact on ML.



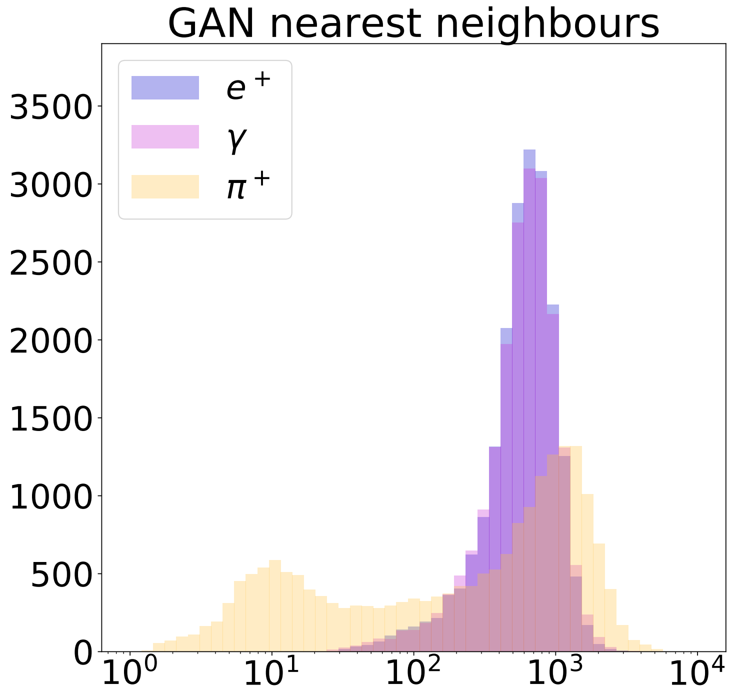
“Overtraining”



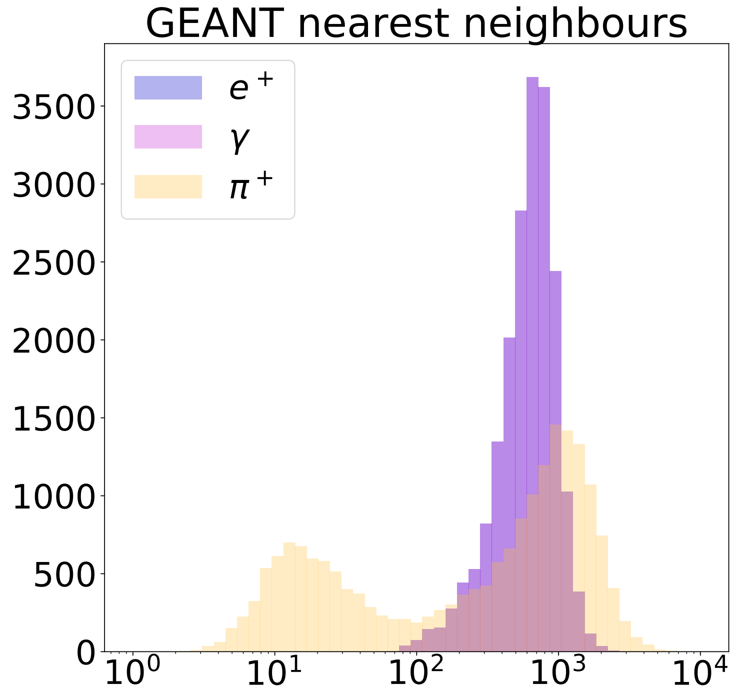
~not memorizing



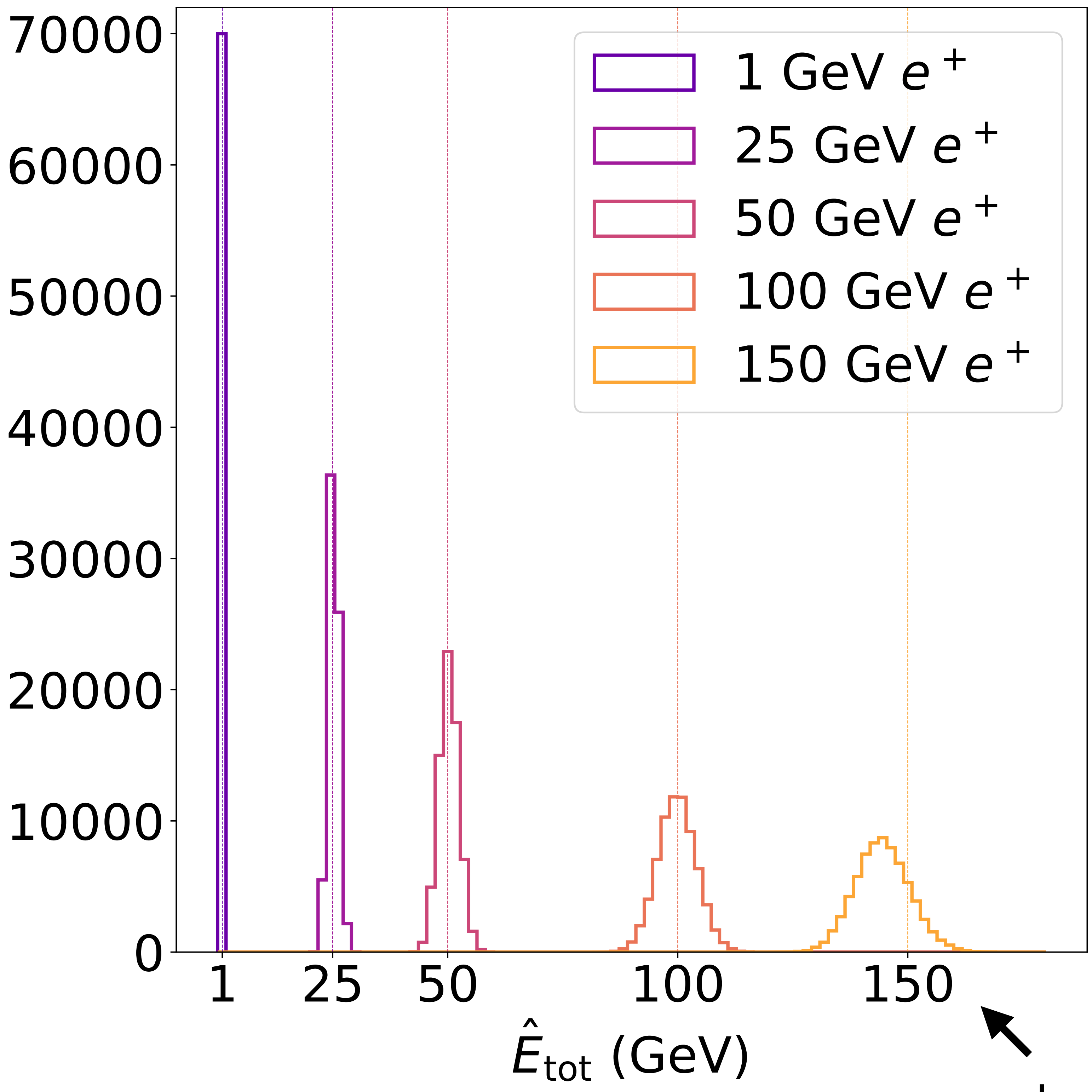
A key challenge in training GANs is the diversity of generated images. This does not seem to be a (big) problem for CaloGAN.



~no mode collapse



Extrapolating



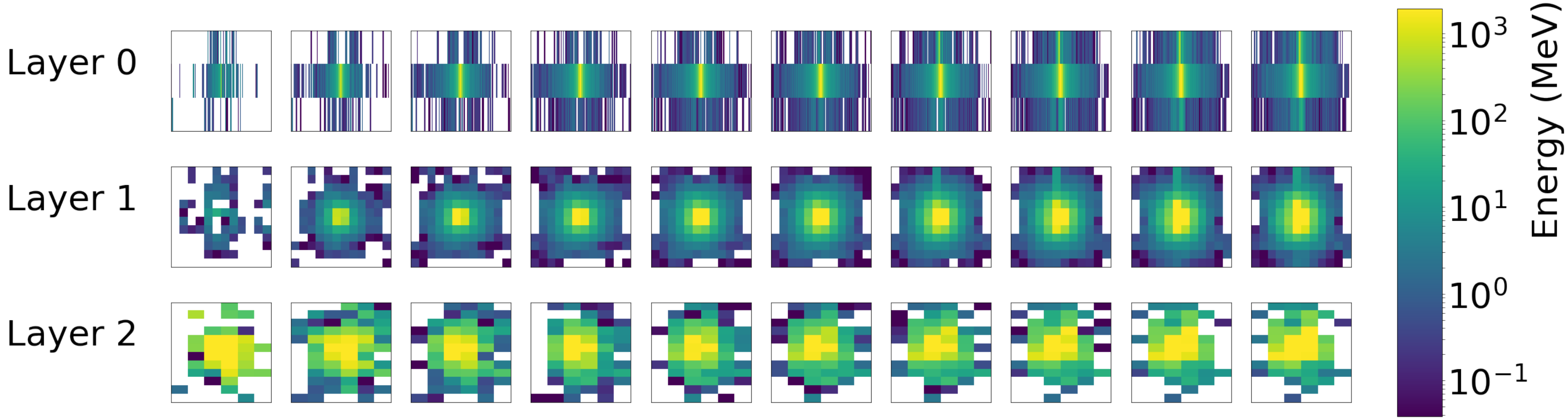
GANs are not designed to extrapolate, but in some cases, they can smoothly go on!

works here until there is no new physical principles which turn on at some energy

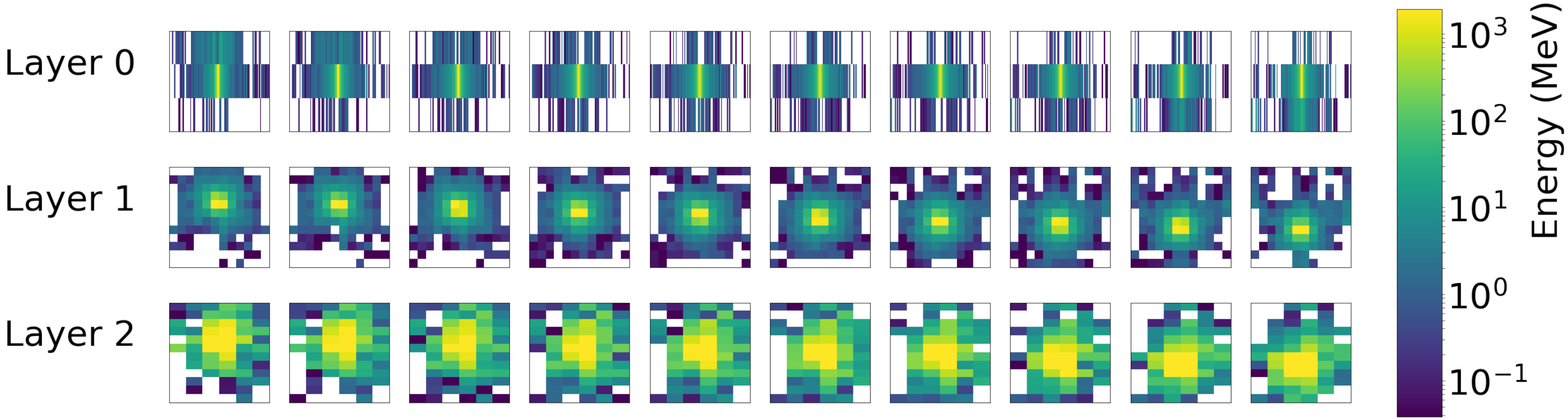
Beyond our training sample!

Conditioning

Fix noise, scan latent variable corresponding to energy



Fix noise, scan latent variable corresponding to x-position



Modern Universal Quantum Computers

There is no consensus on architecture, but most efforts for universal quantum computing use superconductors.

I'm not going to talk about hardware, though it is an exciting topic.

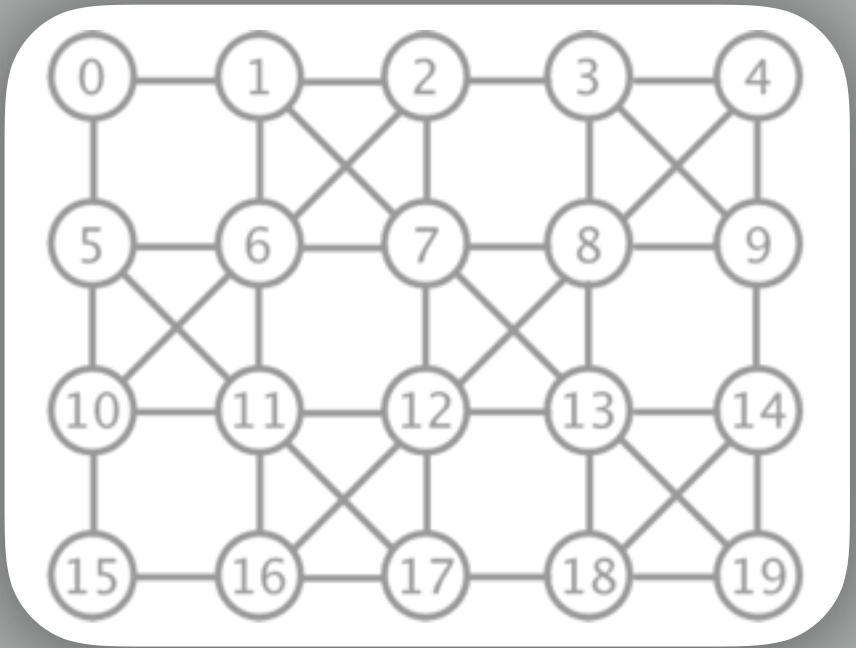
classical computing in the 1970's

quantum computing now

State-of-the-art quantum computers

The best quantum computers have ~10-20 **qubits** with ~few connections per qubit and can stay coherent for O(hundreds) of operations.

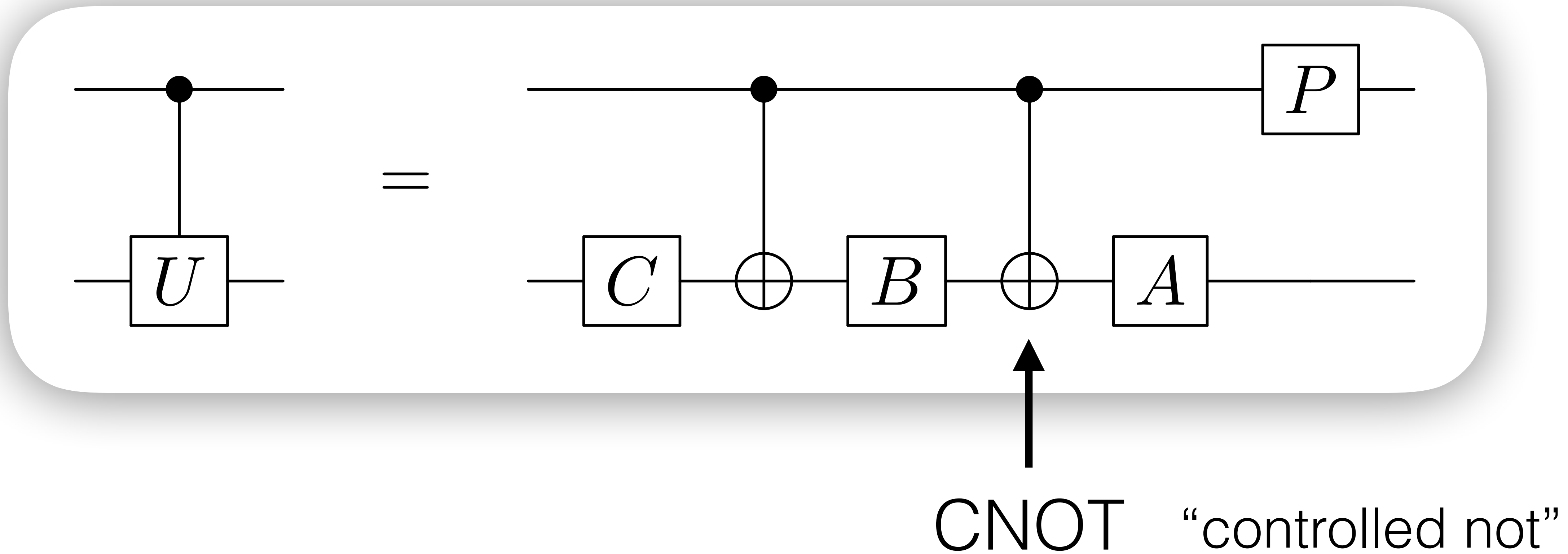
A **qubit** is an abstract representation of a quantum system that can be in a superposition of two states (often thought of as a spin)



This is one of IBM's 20-qubit quantum computers. Lines represent connections.

Challenges with current computers

In practice: only controlled operation that is allowed is CNOT (swap if 1 otherwise do nothing) … need to decompose.



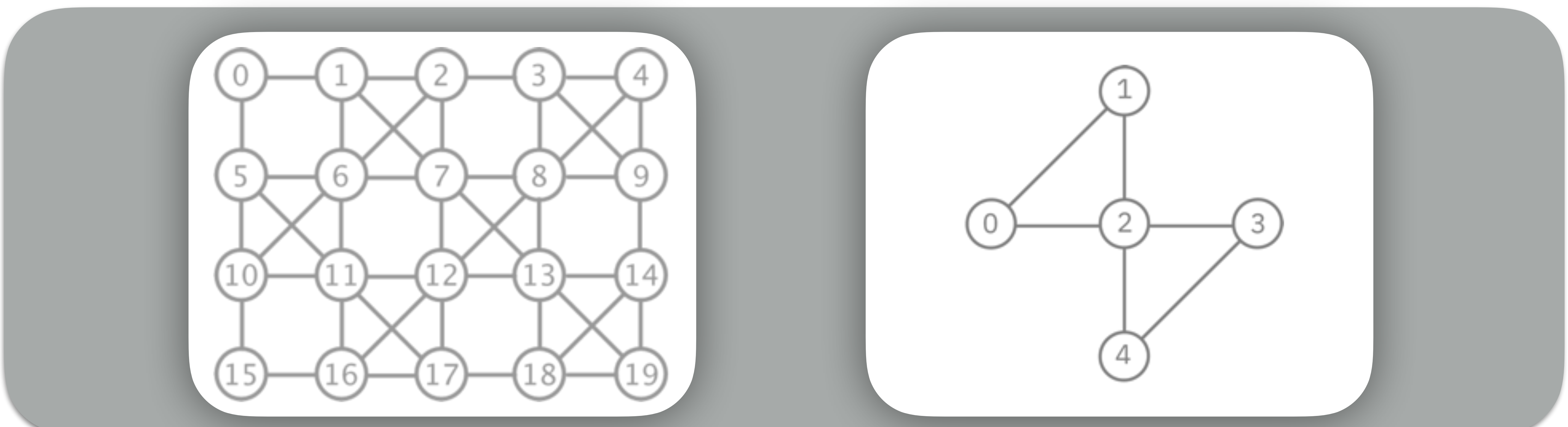
There is no compiler … need to do circuit decomposition by hand (!)

Challenges with current computers

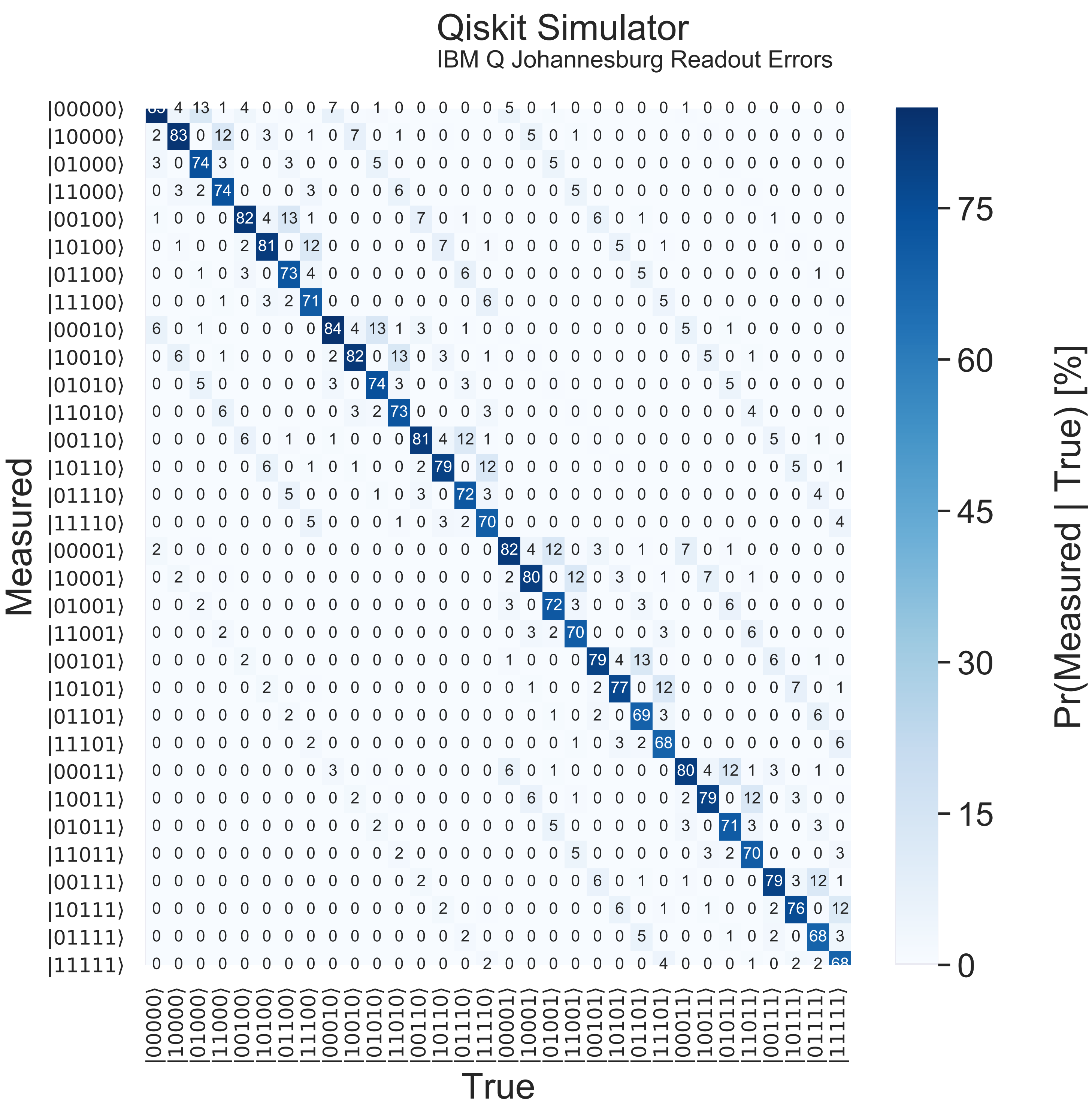
In practice: only controlled operation that is allowed is CNOT (swap if 1 otherwise do nothing) … need to decompose.

Circuit implementation is architecture-dependent
need to know what connections are available

(can swap, but CANNOT clone qubits!)



Readout error corrections



On a quantum computer, the state may be 1 but readout as a 0, etc.

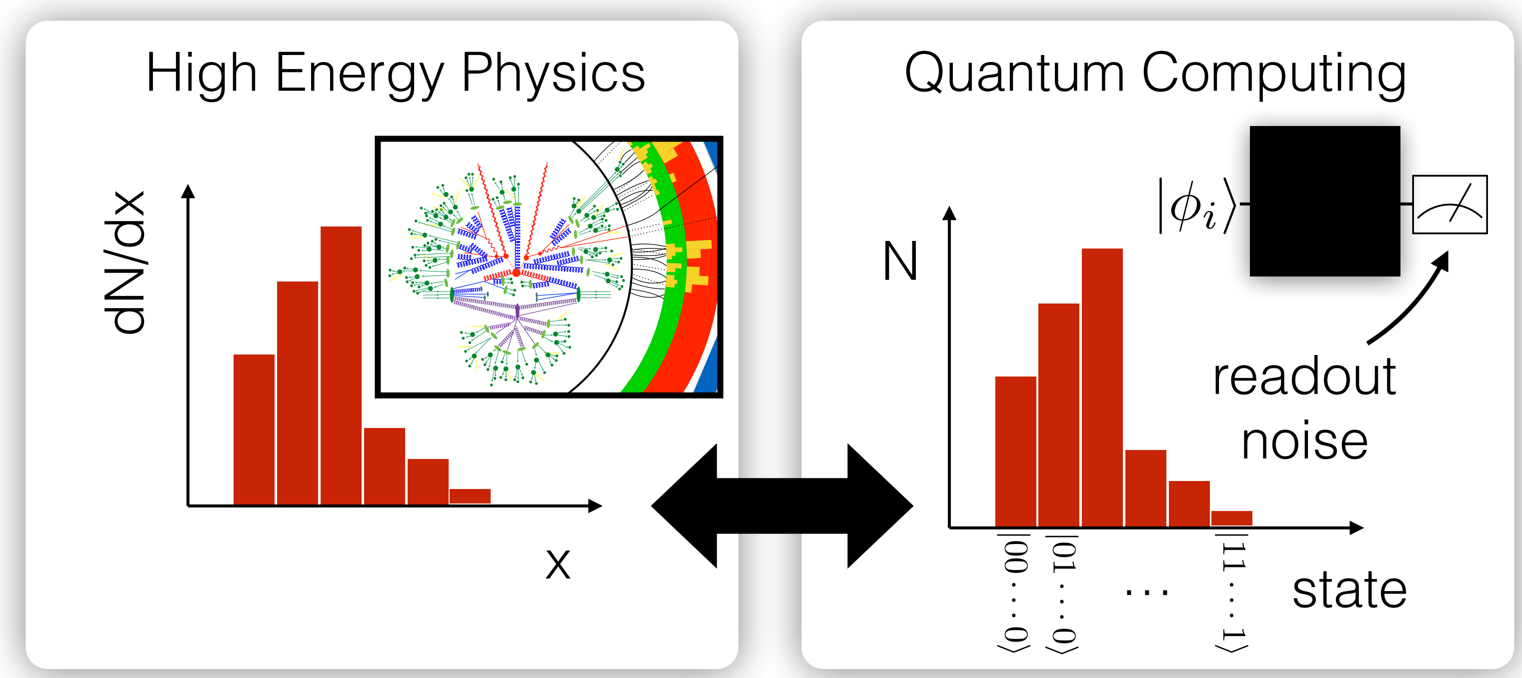
For n qubits, there is a $2^n \times 2^n$ transition matrix.

HEP has proposed many solutions to this problem!

…and we call them **unfolding**

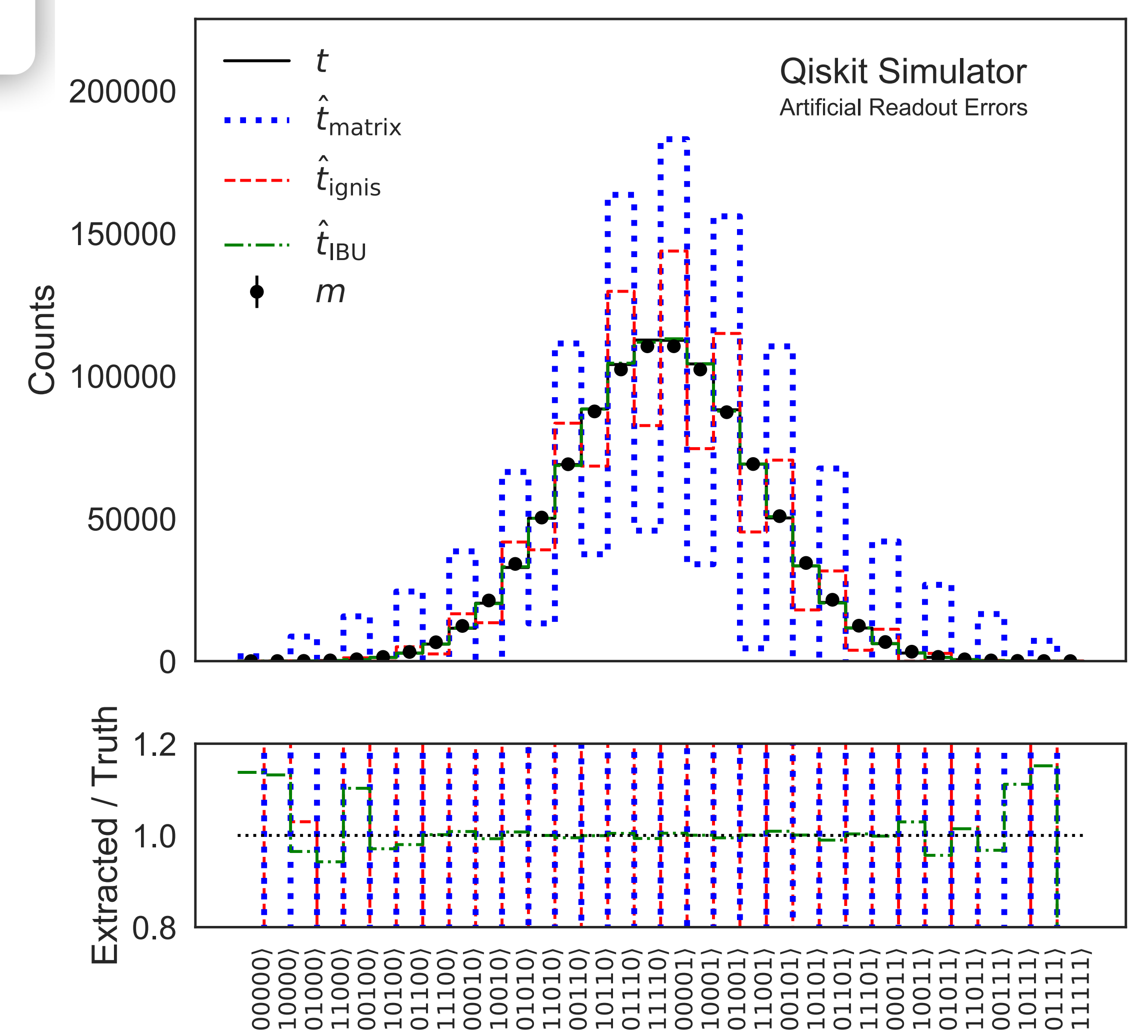
BPN, M. Urbanek, W. de Jong, C. Bauer, 1910.01969

Readout error corrections



IBM standard

HEP standard



We have proposed to use HEP unfolding techniques to correct quantum computer readout errors.

…there are other connections we are exploring in this area - stay tuned!

BPN, M. Urbanek, W. de Jong, C. Bauer, 1910.01969

Gate error corrections