Correcting for Detector Effects (Unfolding/Deconvolution) with Machine Learning

Benjamin Nachman

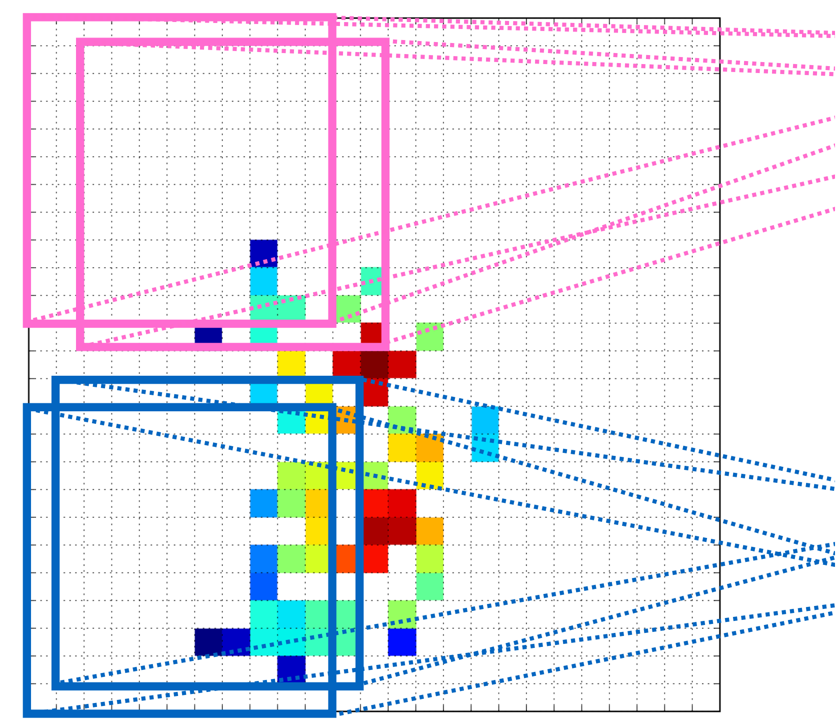
Lawrence Berkeley National Laboratory

cern.ch/bnachman @bpnachman bnachman

bpnachman@lbl.gov

BERKELEY LAB

BERKELEY EXPERIMENTAL PARTICLE PHYSICS

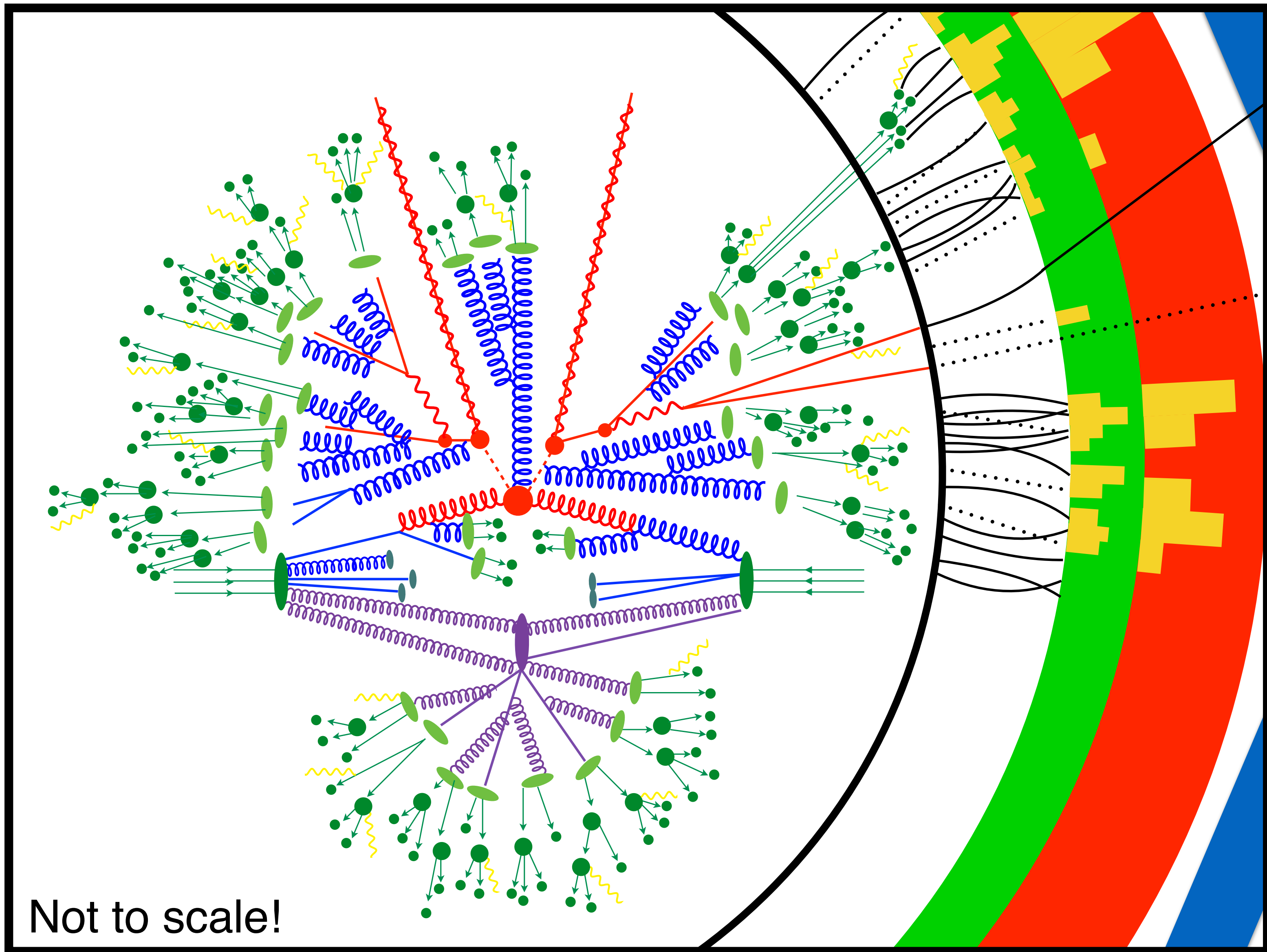


EPE Seminar
University of Washington
May 21, 2020

Today's talk

1. Brief unfolding primer
2. The hyper challenge
3. Reweighting (DCTR)
4. OmniFold
5. Plans for the future

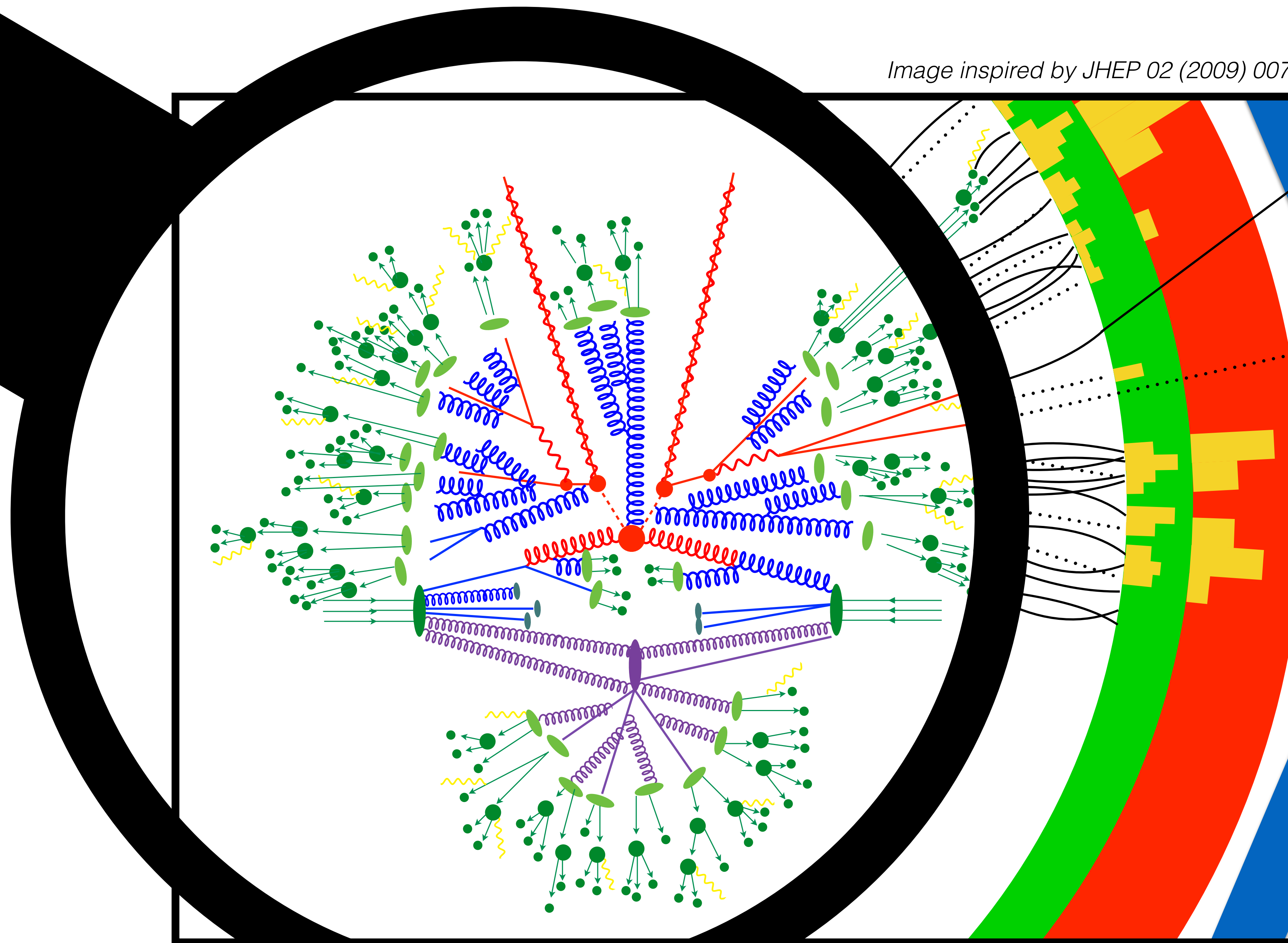
Image inspired by JHEP 02 (2009) 007



Measurements in HEP

The key challenge is that there is a detector in the way!

Image inspired by JHEP 02 (2009) 007



We need to remove detector effects in order to compare with theory. We call this ***Unfolding***.

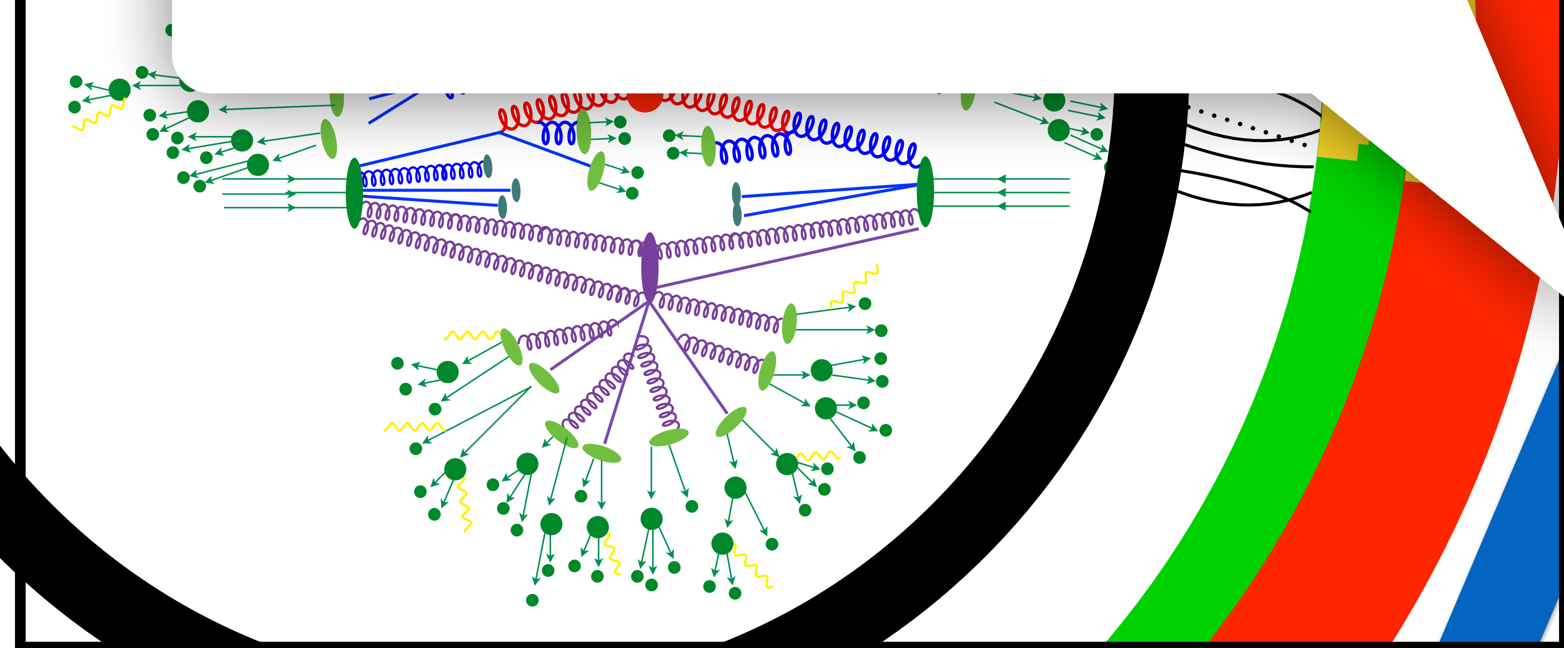
other people call it *deconvolution*

Measurements in HEP

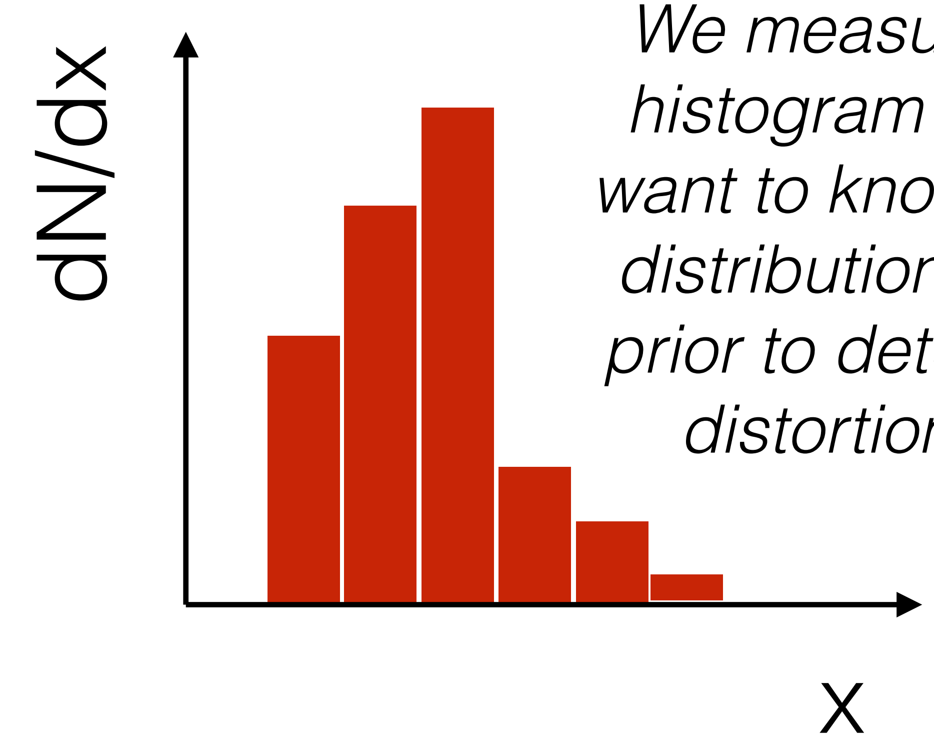
The

is

09) 007



Typical situation:



We measure a histogram and want to know the distribution of x prior to detector distortions.

We need to remove detector effects in order to compare with theory. We call this ***Unfolding***.

other people call it *deconvolution*

What does unfolding do?

In general, unfolding needs to correct for interrelated effects:

- Acceptance and efficiency
  - ➡ Particles produced may not be measured
- Detector noise
  - ➡ Particles measured may not be from real particles
- Background processes
  - ➡ If you want to measure process X, need to remove Y
- Combinatorics
  - ➡ If N particles, chance that detector can change order
- Detector distortions
  - ➡ Bias and resolution effects

Illustrative toy example

$$m = Rt$$

m **= measured;** ***t*** **= true**

We usually call ***R*** the “response matrix” because ***m*** and ***t*** are binned (and thus vectors).

In HEP, we (usually) get ***R*** from extremely detailed detector simulations.

Illustrative toy example

$$m = Rt$$

m = measured; *t* = true

What you want to do is to define $t = R^{-1} m$.

Illustrative toy example

$$m = Rt$$

m = measured; *t* = true

What you want to do is to define $t = R^{-1} m$.

In the next slides, I hope to convince you that this is not usually a good idea.

Illustrative toy example

$$m = Rt$$

m = measured; *t* = true

$$R = \begin{pmatrix} 1-\epsilon & \epsilon \\ \epsilon & 1-\epsilon \end{pmatrix}$$

Consider this case, where $0 \leq \epsilon \leq 0.5$

Illustrative toy example

$$m = Rt$$

m = measured; *t* = true

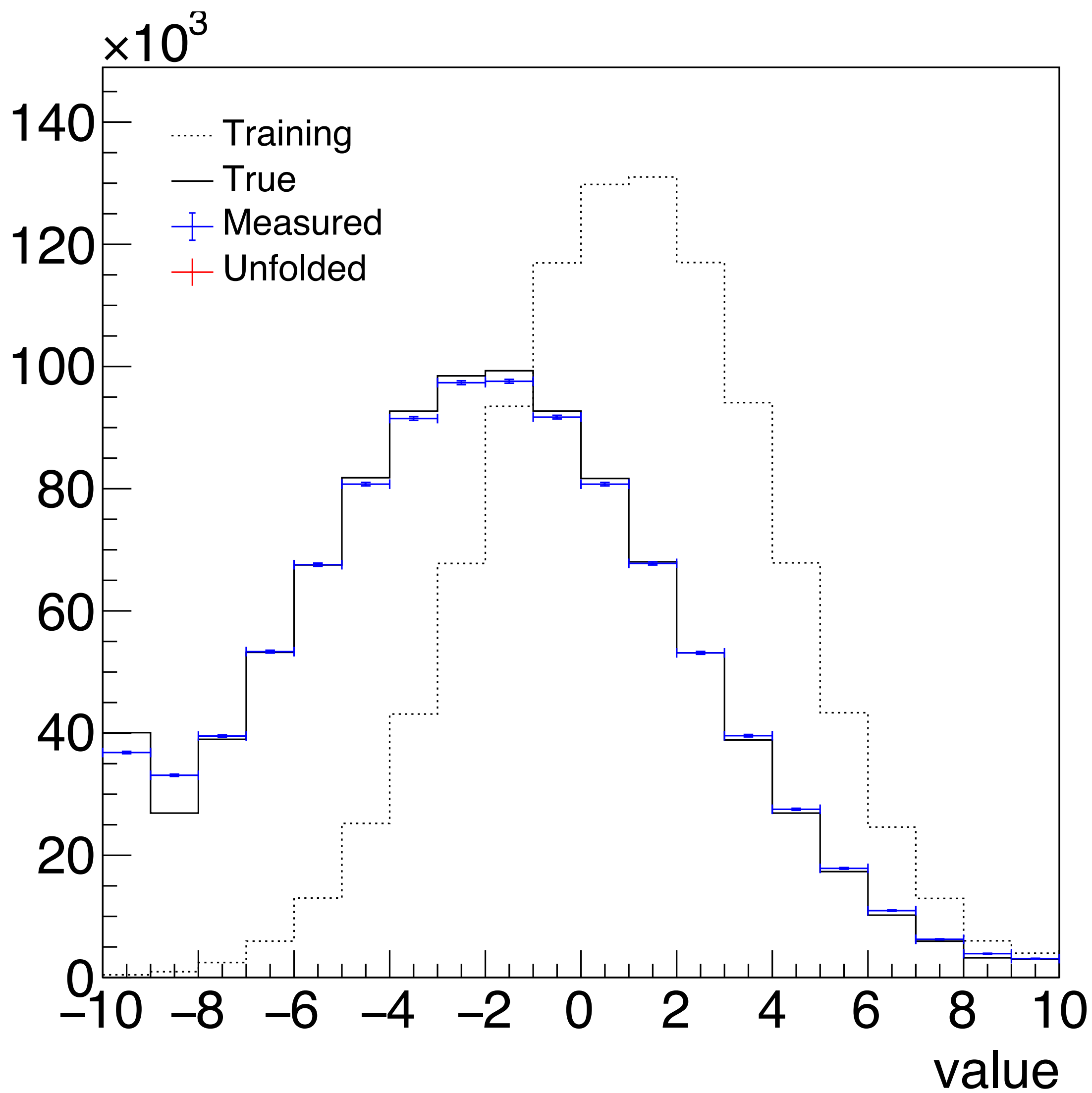
$$R = \begin{pmatrix} 1-\epsilon & \epsilon \\ \epsilon & 1-\epsilon \end{pmatrix}$$

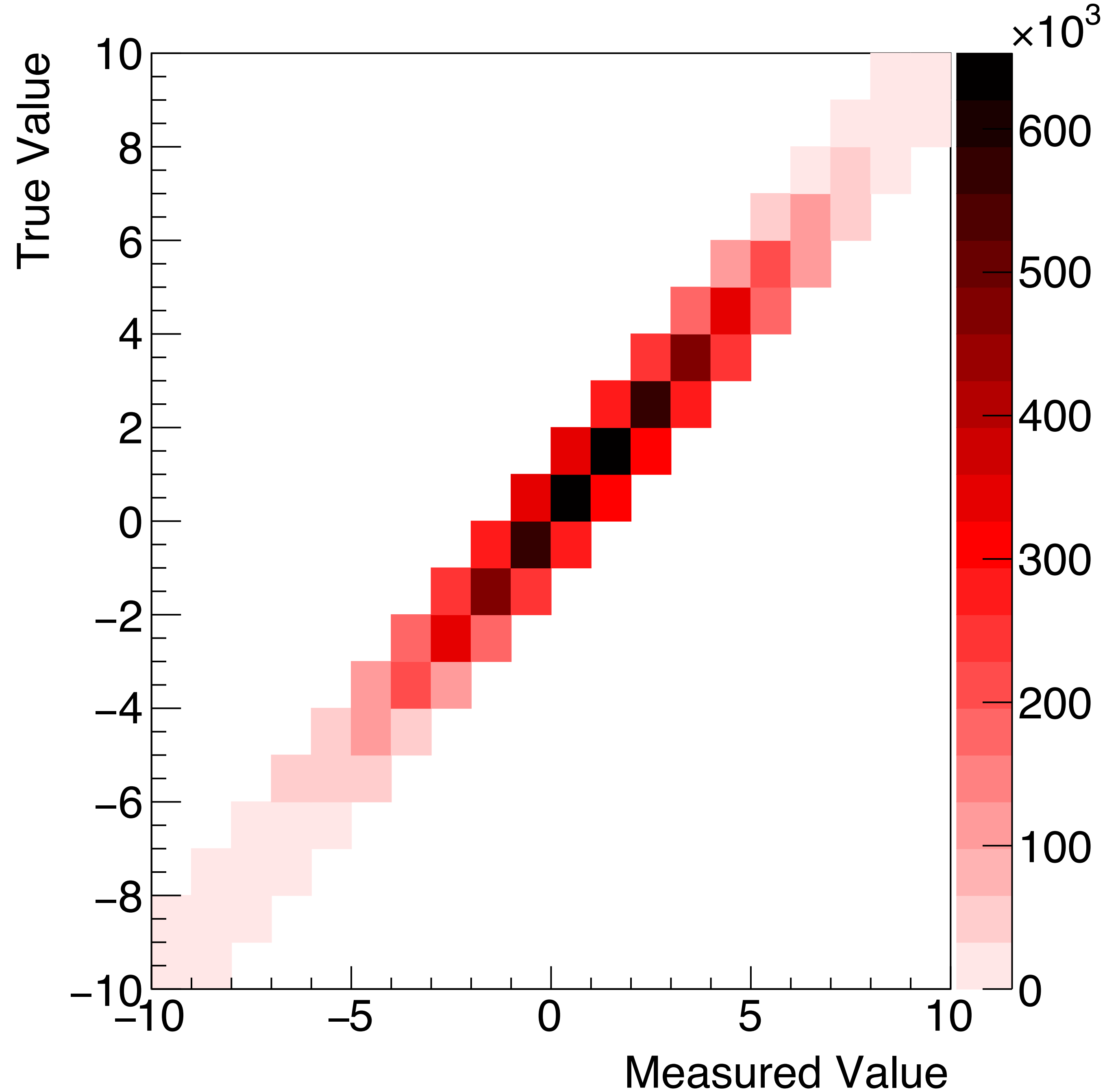
$$\mathrm{Var}(R^{-1}m) \propto 1/\mathrm{Det}(R) = 1-2\epsilon$$

Statistical uncertainty blows up as $\epsilon \to 0.5$

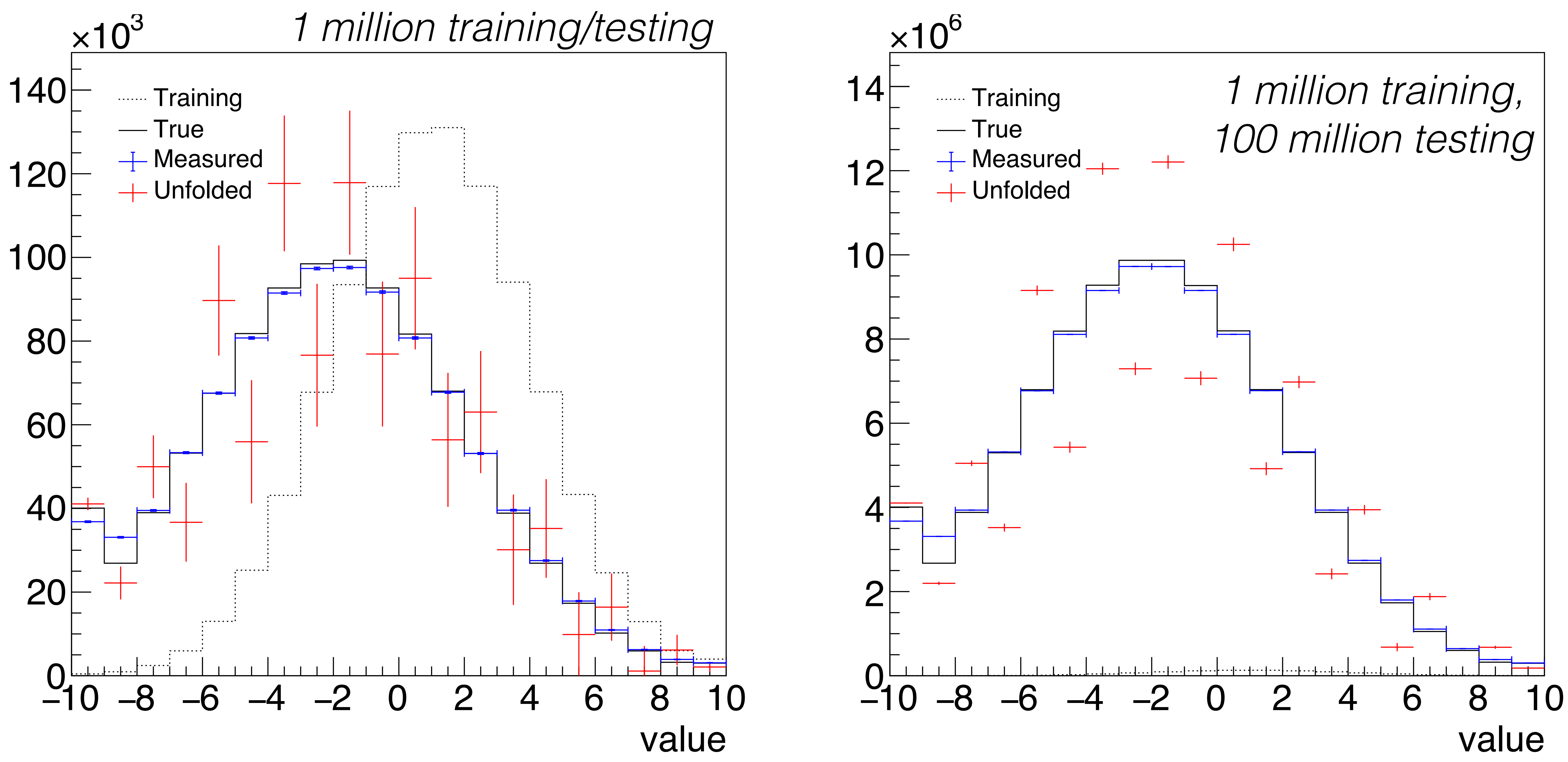
Same idea, more bins

$$R_{\text{norm}} = \begin{pmatrix} 0.75 & 0.25 & 0 & & & \\ 0.25 & 0.50 & 0.25 & 0 & & \\ 0 & 0.25 & 0.50 & 0.25 & 0 & \\ & 0 & 0.25 & 0.50 & 0.25 & \\ & & 0 & 0.25 & 0.50 & \\ & & & & & \ddots \end{pmatrix}$$





Unfolding by Matrix Inversion



Stat. uncertainty is large and there is a bias when training dataset is too small.

The HEP solution

Our solution is to do regularized matrix inversion.

There are two main techniques that we use:

"Iterative Bayesian Unfolding"

$$\theta_{ij} = \frac{\Pr(m_j|t_i)\cdot\Pr(t_i)}{\sum_i \Pr(m_j|t_i)\cdot\Pr(t_i)}$$

response matrix

regularization = number of iterations

$$\Pr_{k+1}(t_i) = \sum_j \theta_{ij}\Pr_k(t_i)$$

Nucl. Inst. Meth. A 362 (1995) 487

"Singular Value Decomposition (SVD) Unfolding"

$$R = USV^T$$

U, *V*, orthogonal, *S* diagonal & non-negative

$$d = U^T m \qquad z_i(\tau) = \frac{d_i}{s_i}\cdot\frac{s_i^2}{s_i^2+\tau}$$

$$t = Vz$$

regularization parameter

Nucl. Inst. Meth. A 372 (1995) 469

~1000 citations each

Main tool: RooUnfold (ROOT-based C++ code)

The HEP solution

Our solution is to do regularized matrix inversion.

There are two main techniques that we use:

Note: regularized matrix inversion depends on unphysical irregularization parameters

One choses parameters to tradeoff bias and uncertainty.

IBU Unfolding

- depend on prior
- depends on # of iterations

SVD Unfolding

- depend on t

Nucl. Inst. Meth. A 362 (1995) 487

Nucl. Inst. Meth. A 372 (1995) 469

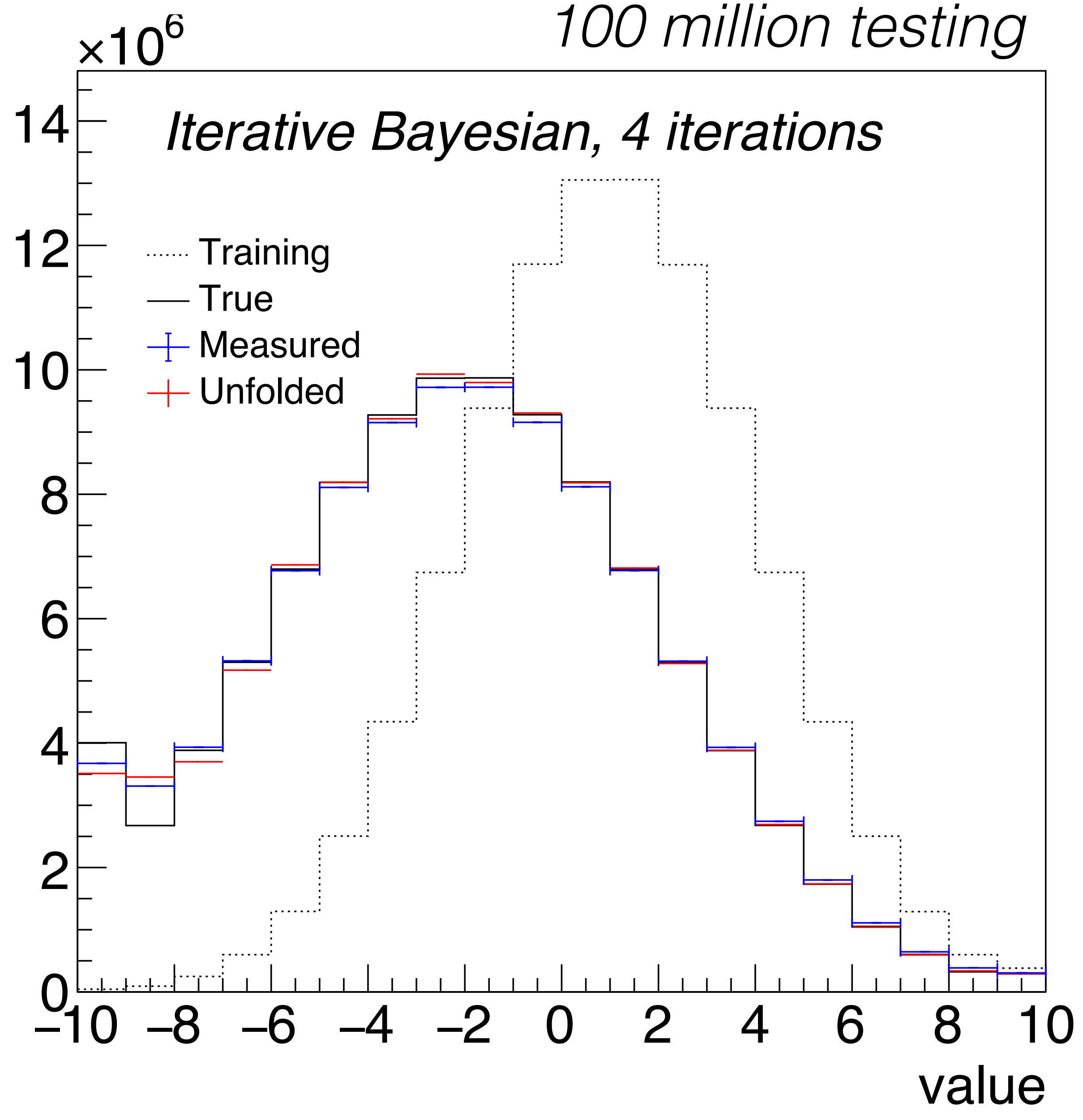
~1000 citations each

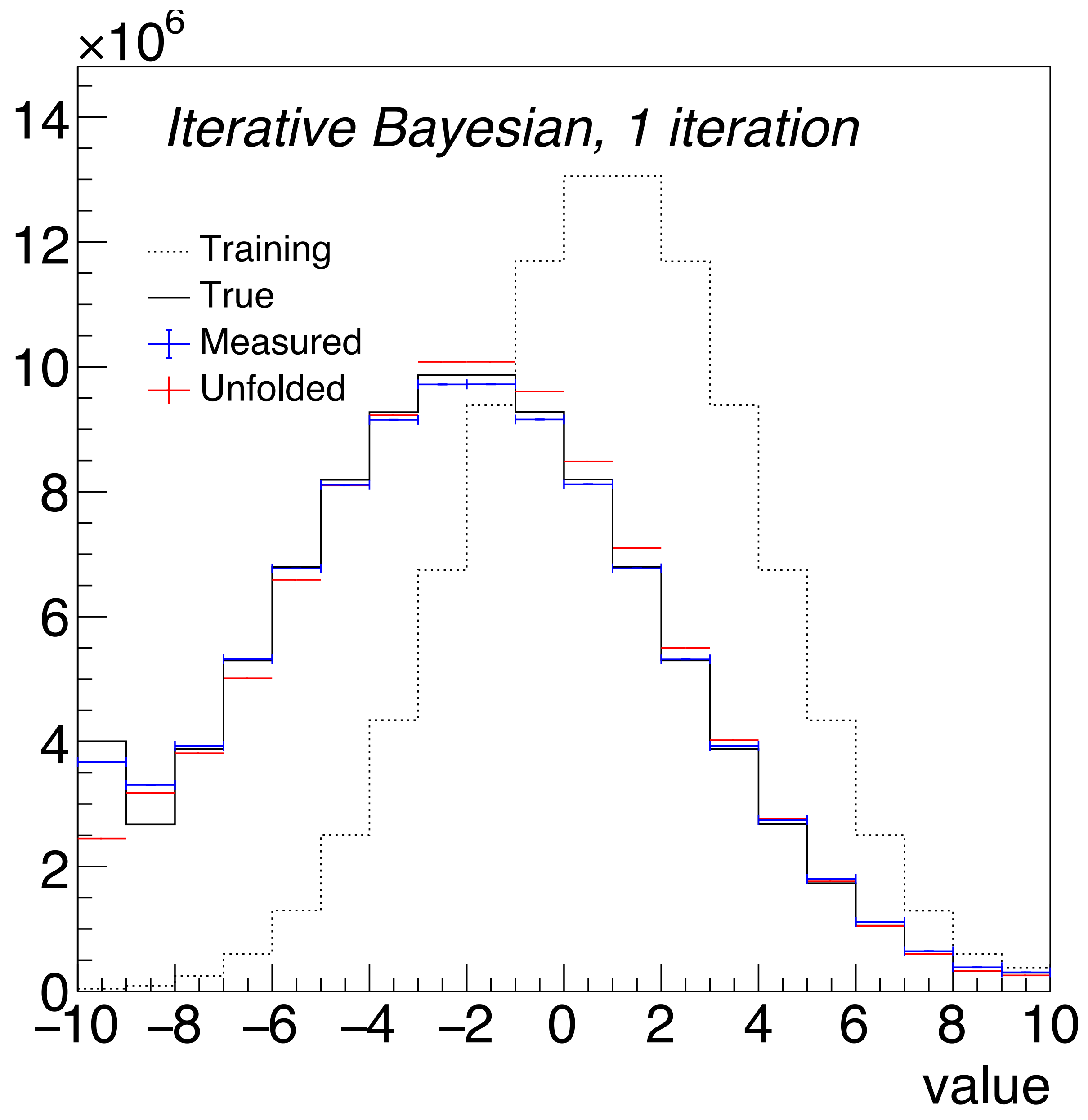
(one can show that this converged to the maximum likelihood estimator)

Main tool: RooUnfold (ROOT-based C++ code)

Example: Iterative Bayesian Unfolding

1 million training,
100 million testing

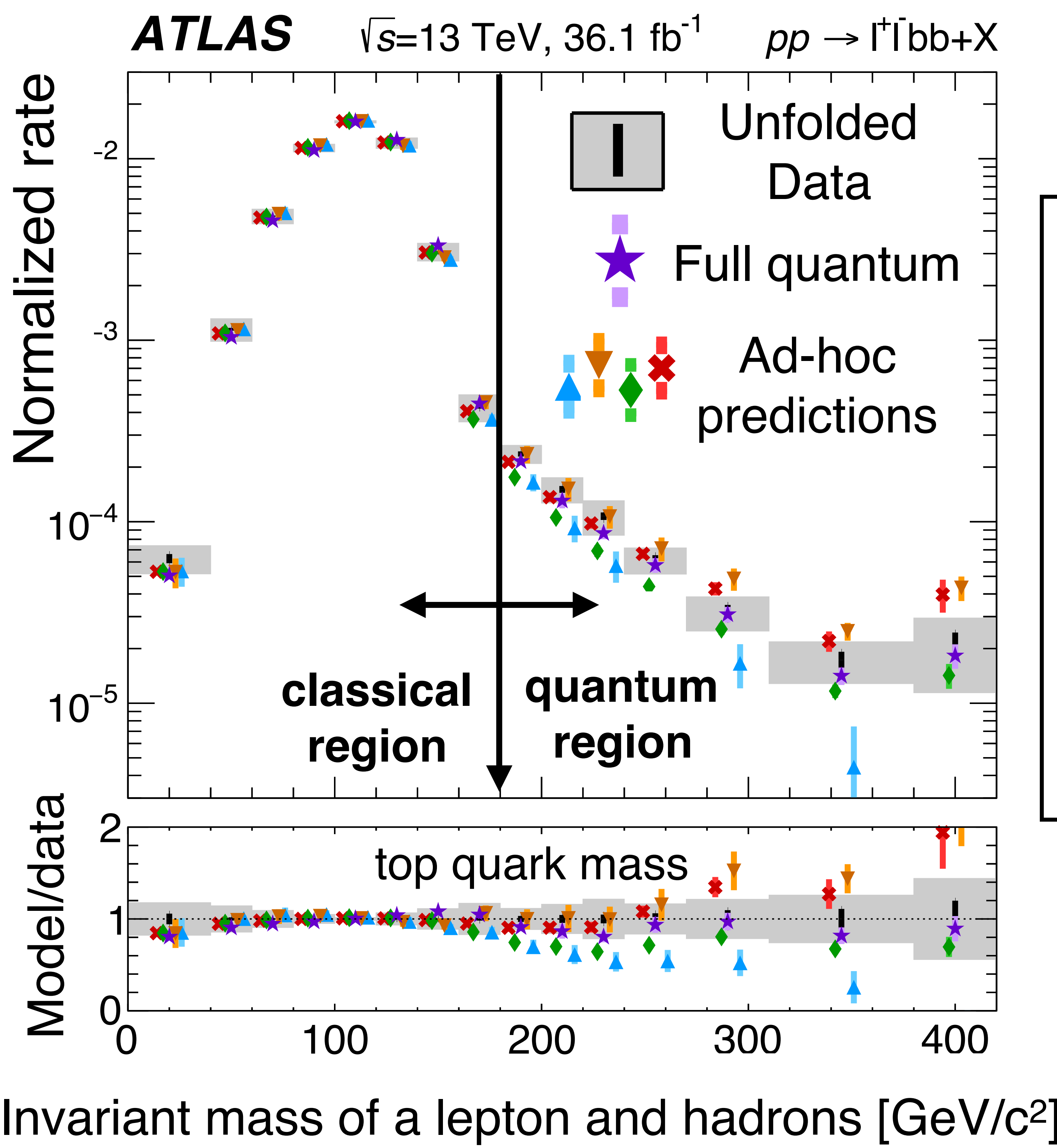




Unfolding in action

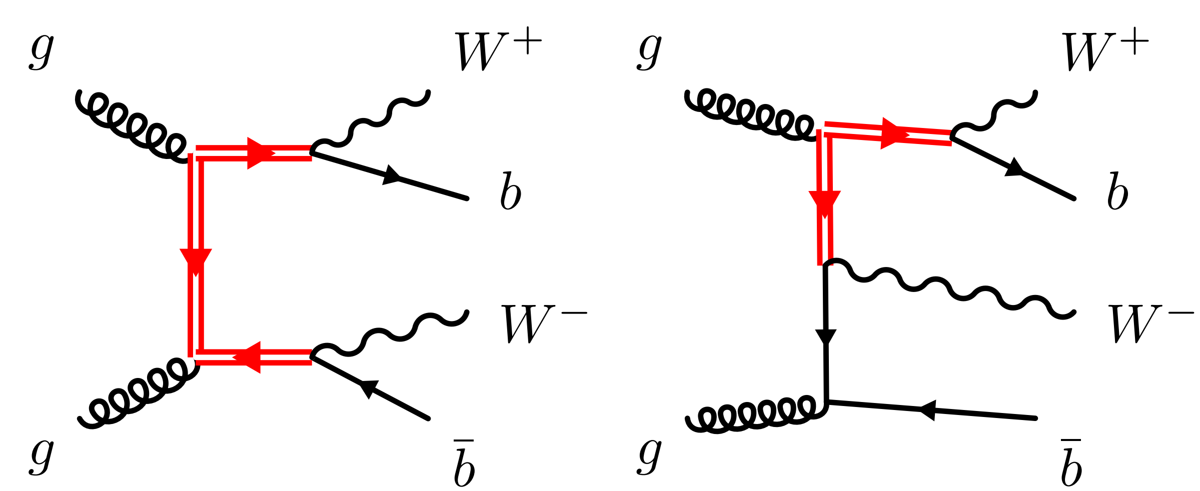
Phys. Rev. Lett. 121 (2018) 152002

Analysis Team: T. Eifert, C. Herwig, **BPN**



proton proton → two W bosons and two b-quarks

Multiple possible diagrams = interference



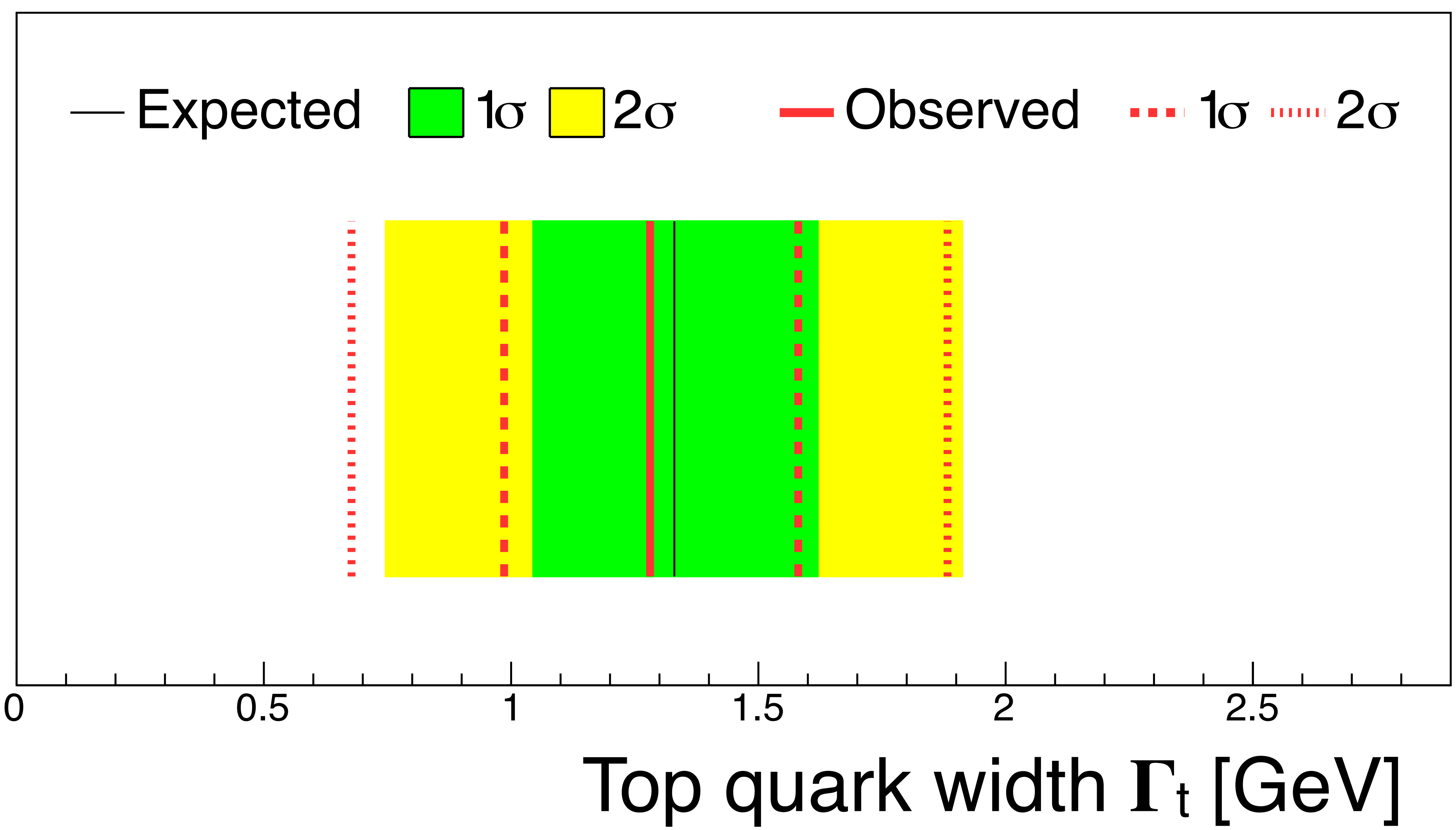
Unfolded data + state-of-the-art calculations = new insight to fundamental interactions

Unfolding in action

These quantum effects were then used to make the most precise direct measurement of the top quark width

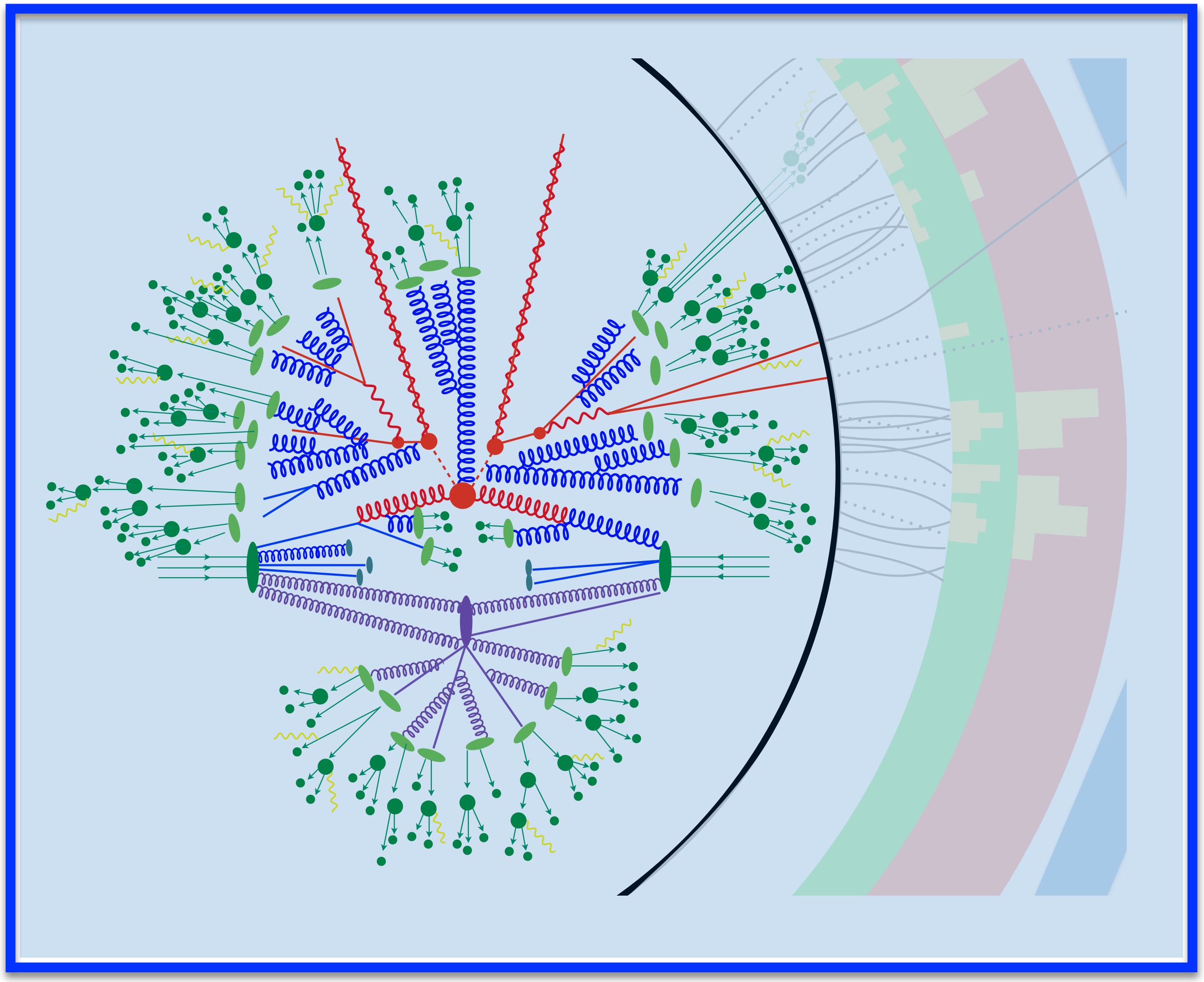
[Phys. Rev. Lett. 122 (2019) 231803](https://doi.org/10.1103/PhysRevLett.122.231803)

C. Herwig, T. Ježo, **BPN**

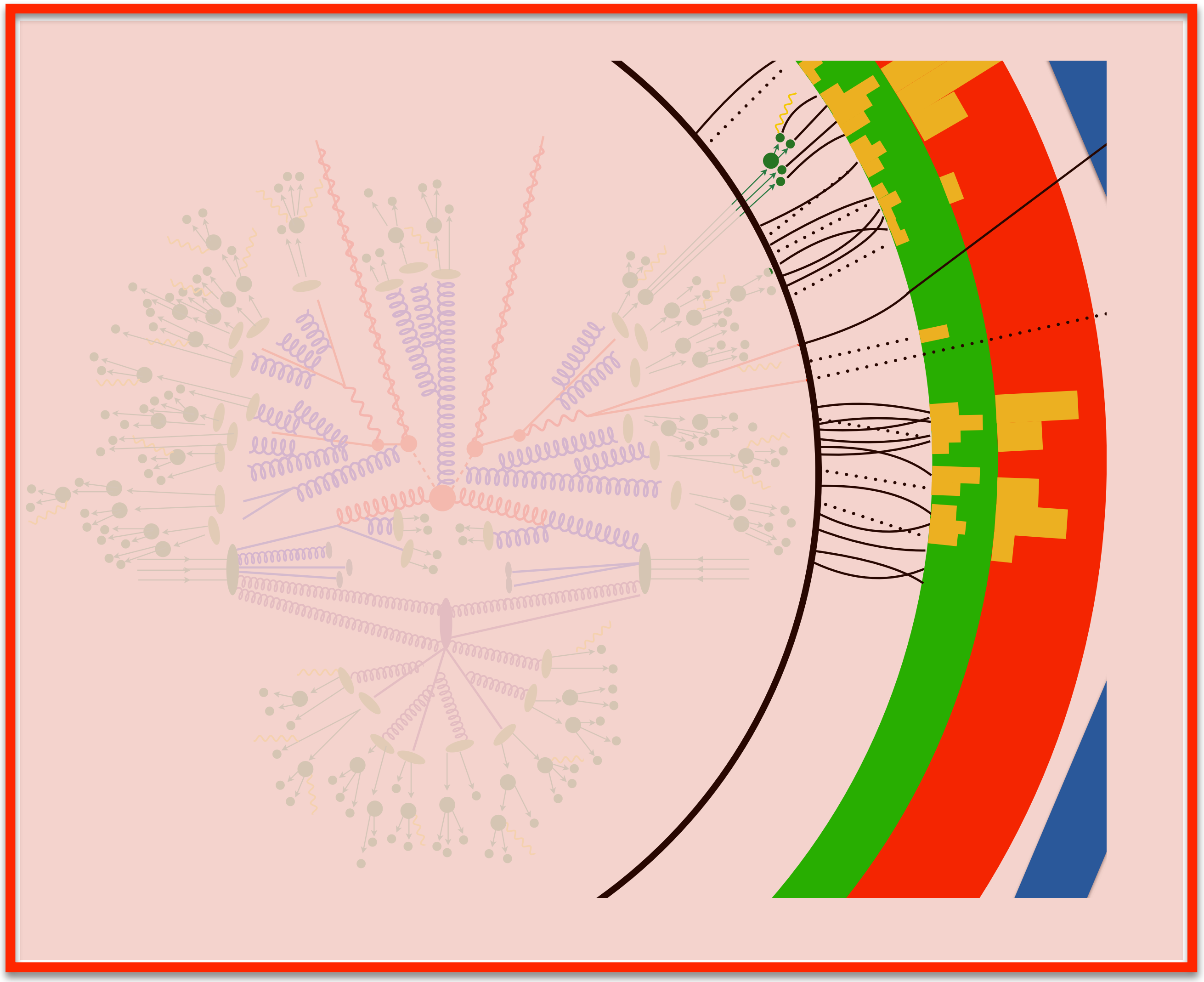


What about the full phase space?

Want this



Measure this



Can we measure the all hadrons (not just 1D histograms)?

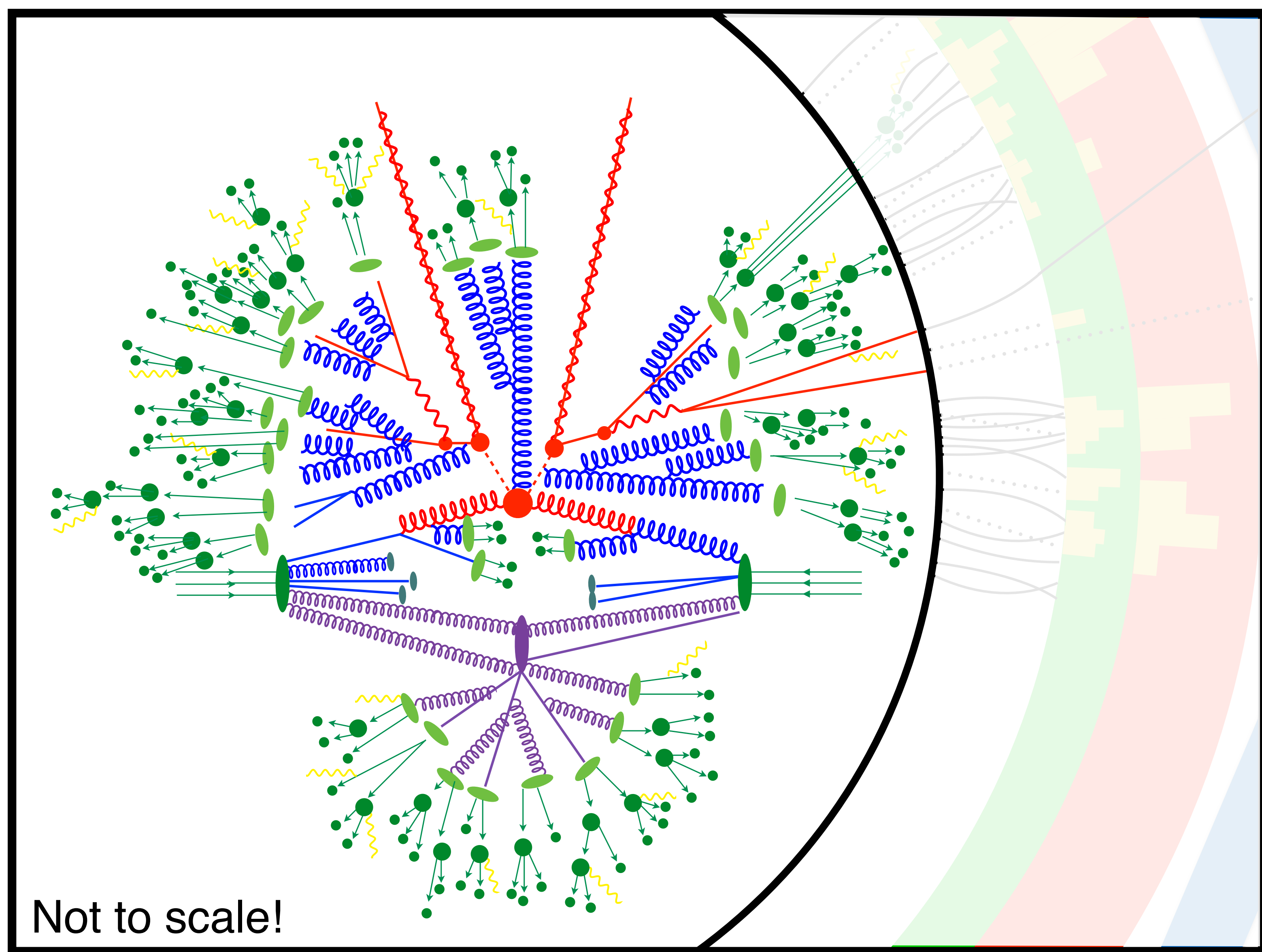
A *hyper* challenge

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



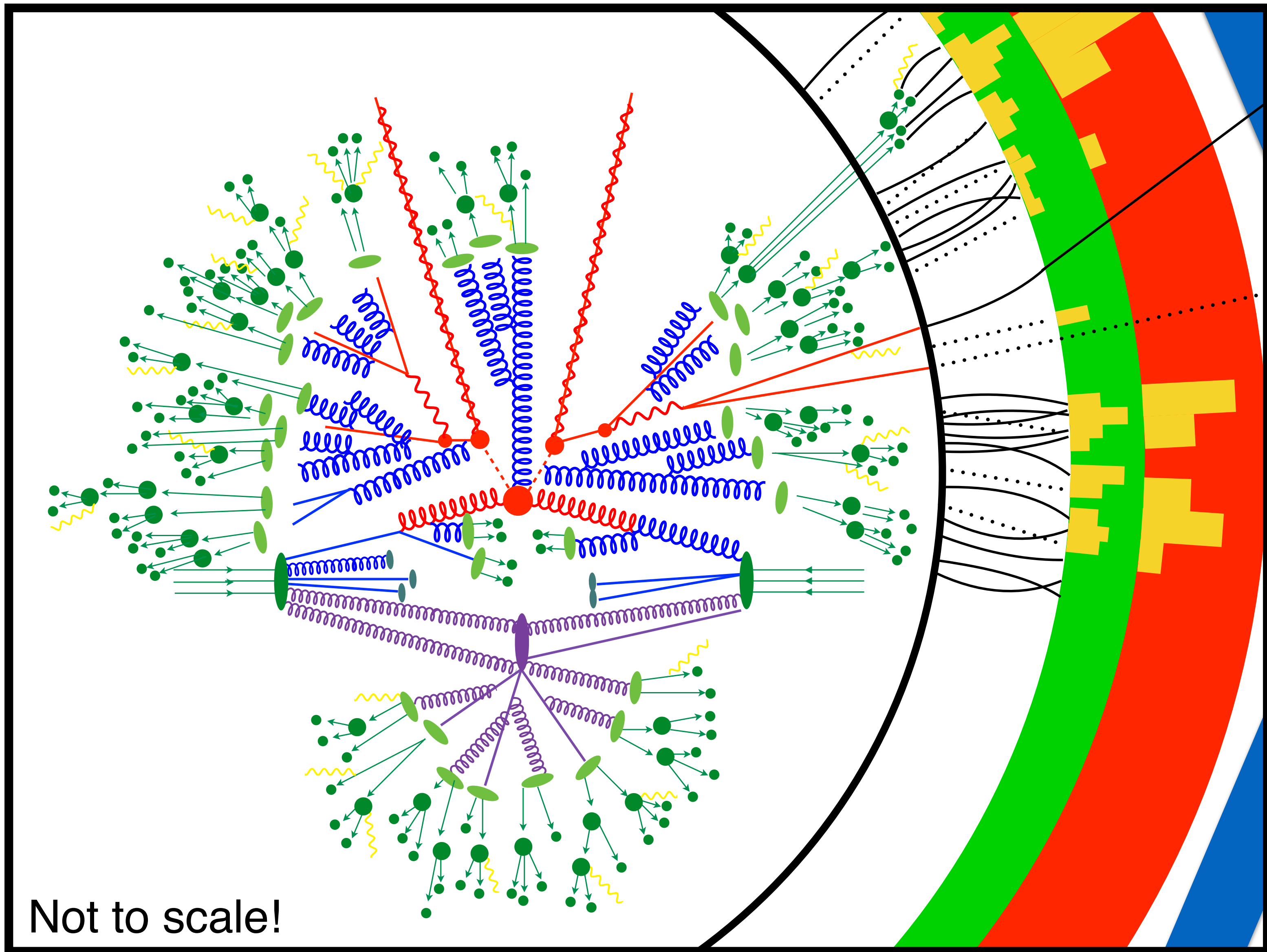
A *hyper* challenge

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

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

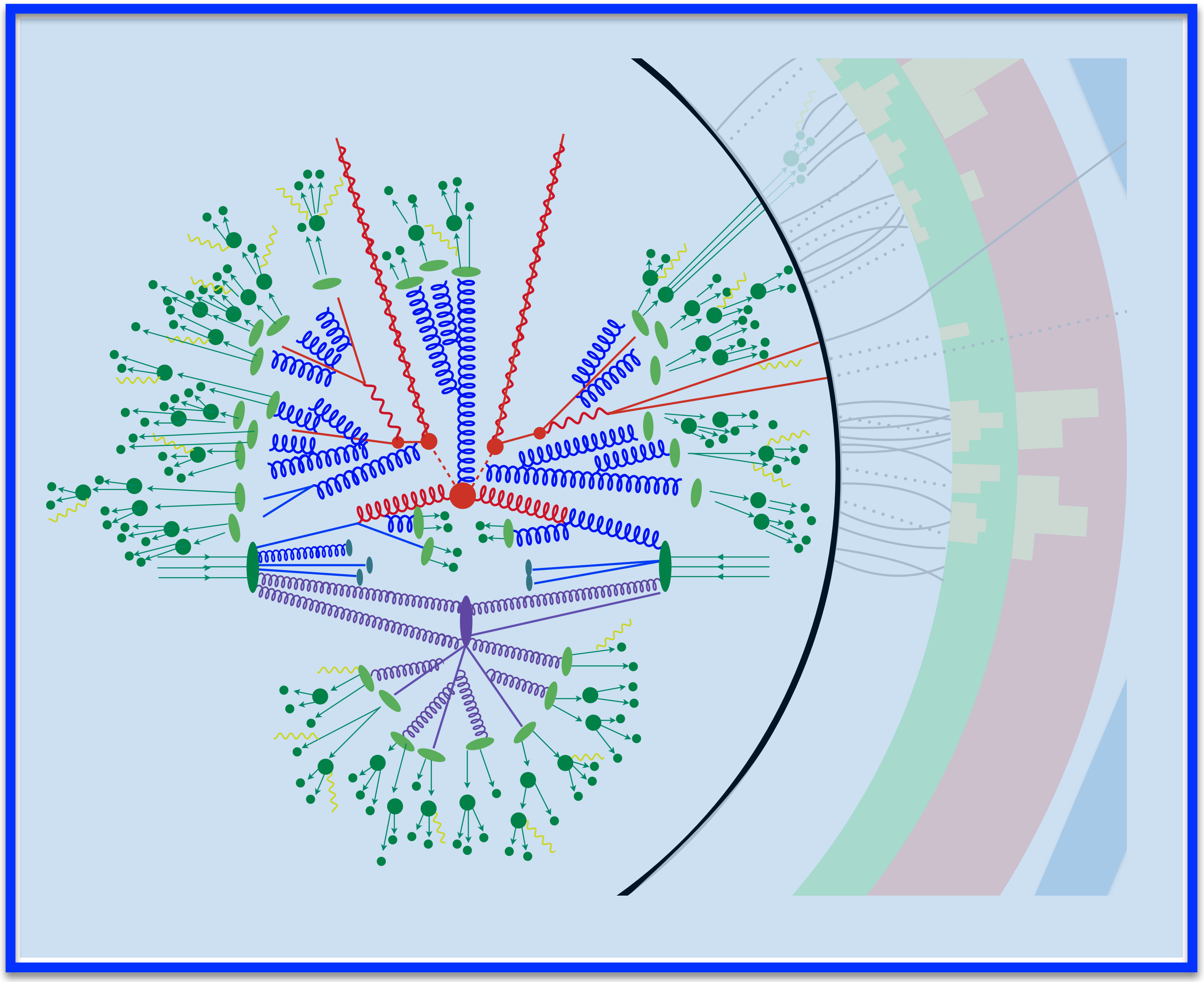
We detect these particles with **O(100 M)** readout channels

Image inspired by JHEP 02 (2009) 007

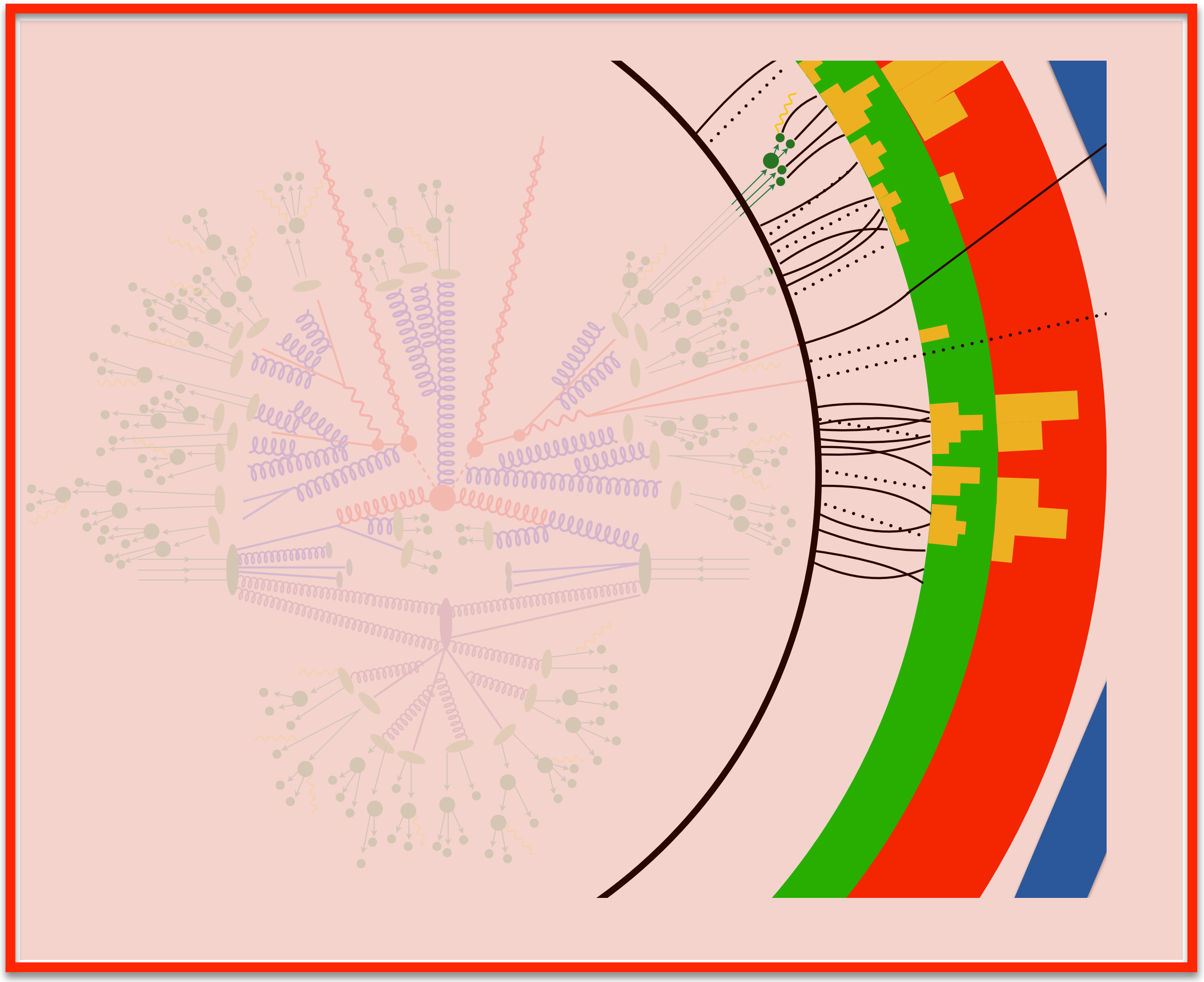


What about the full phase space?

Want this



Measure this

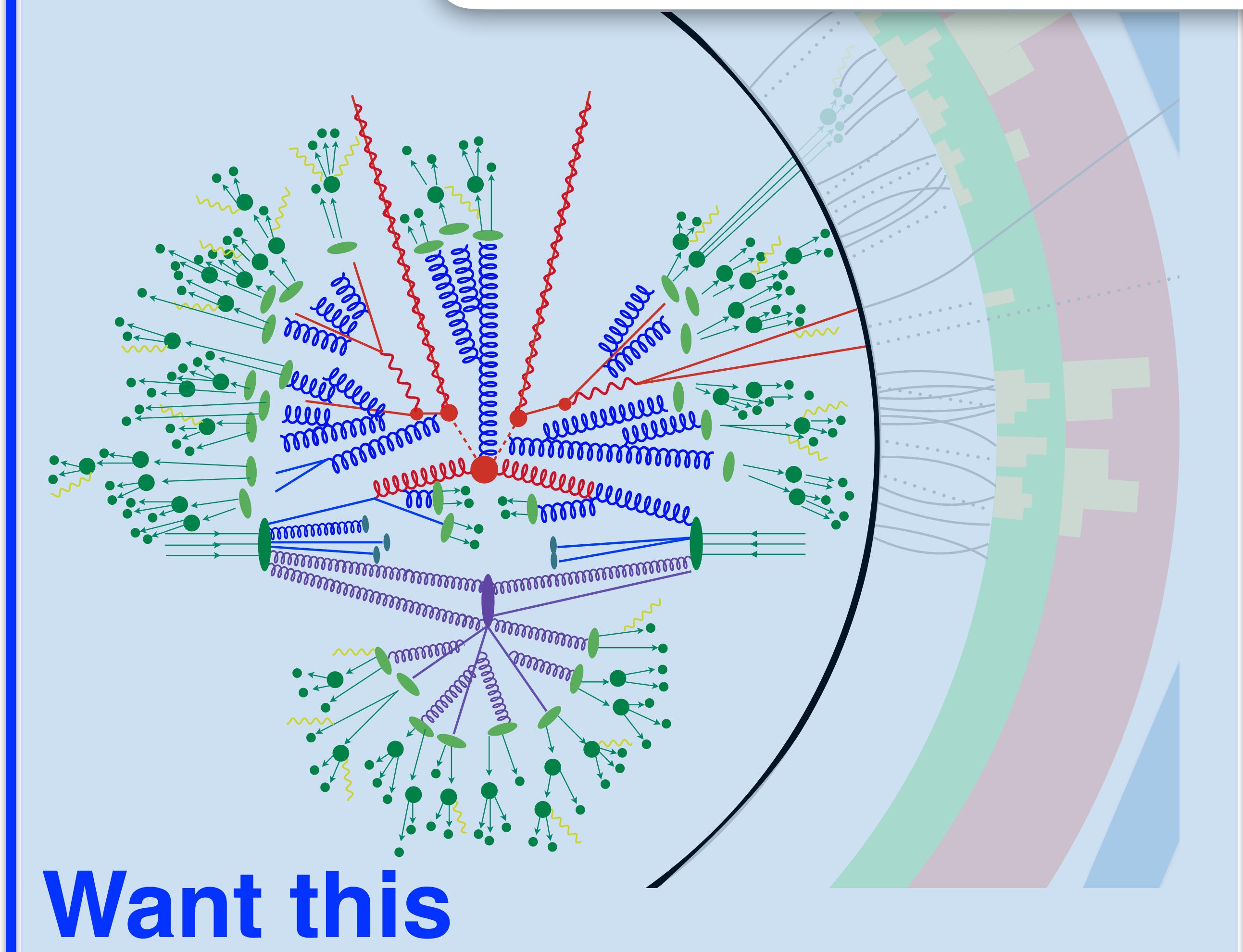


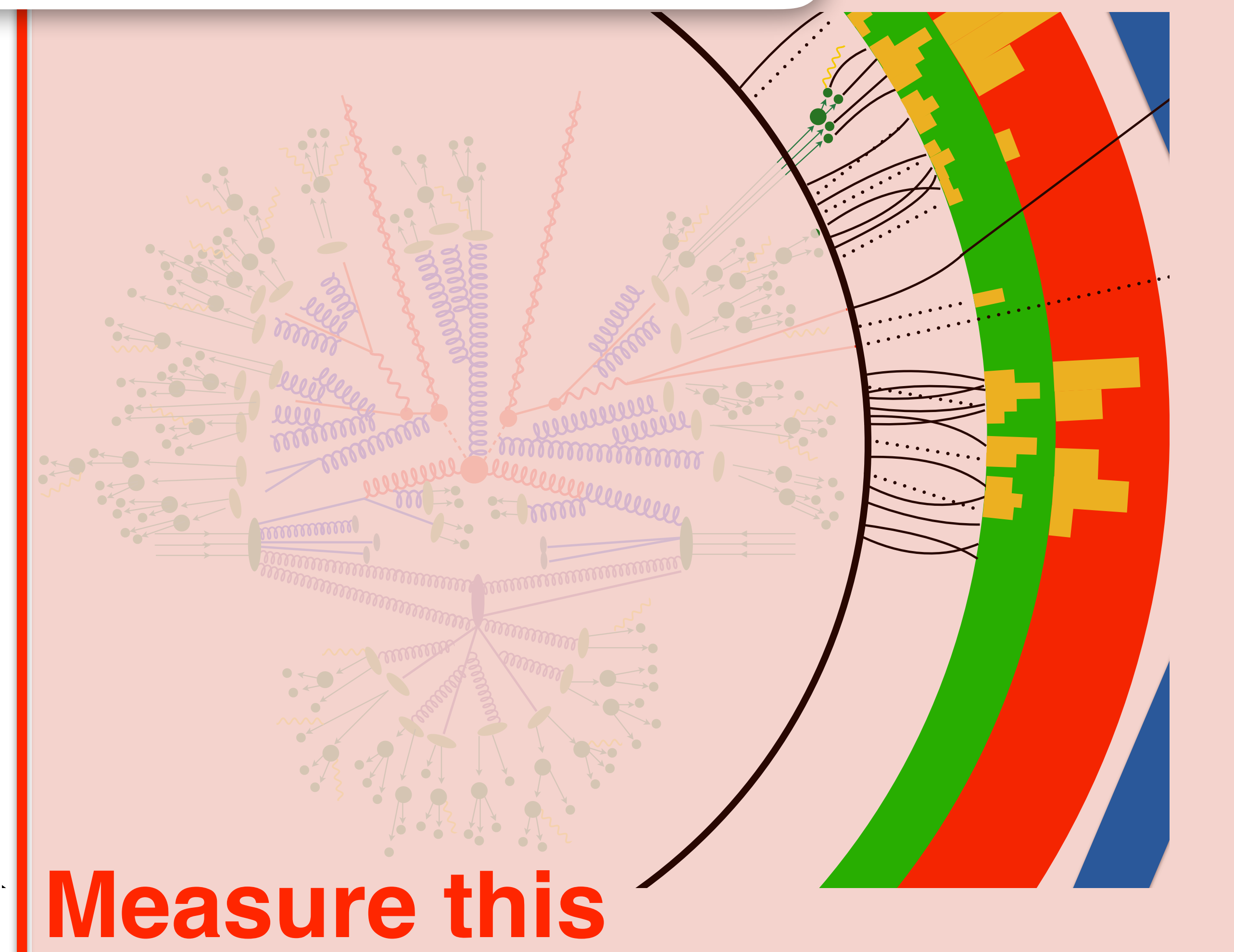
Can we measure the all hadrons (not just 1D histograms)?

Full phase space unfolding

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

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





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

Full phase space unfolding

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

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

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

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

Full phase space unfolding

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

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

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

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

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

*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 = $q(x)/p(x)$

(or something monotonically related to it in a known way)

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

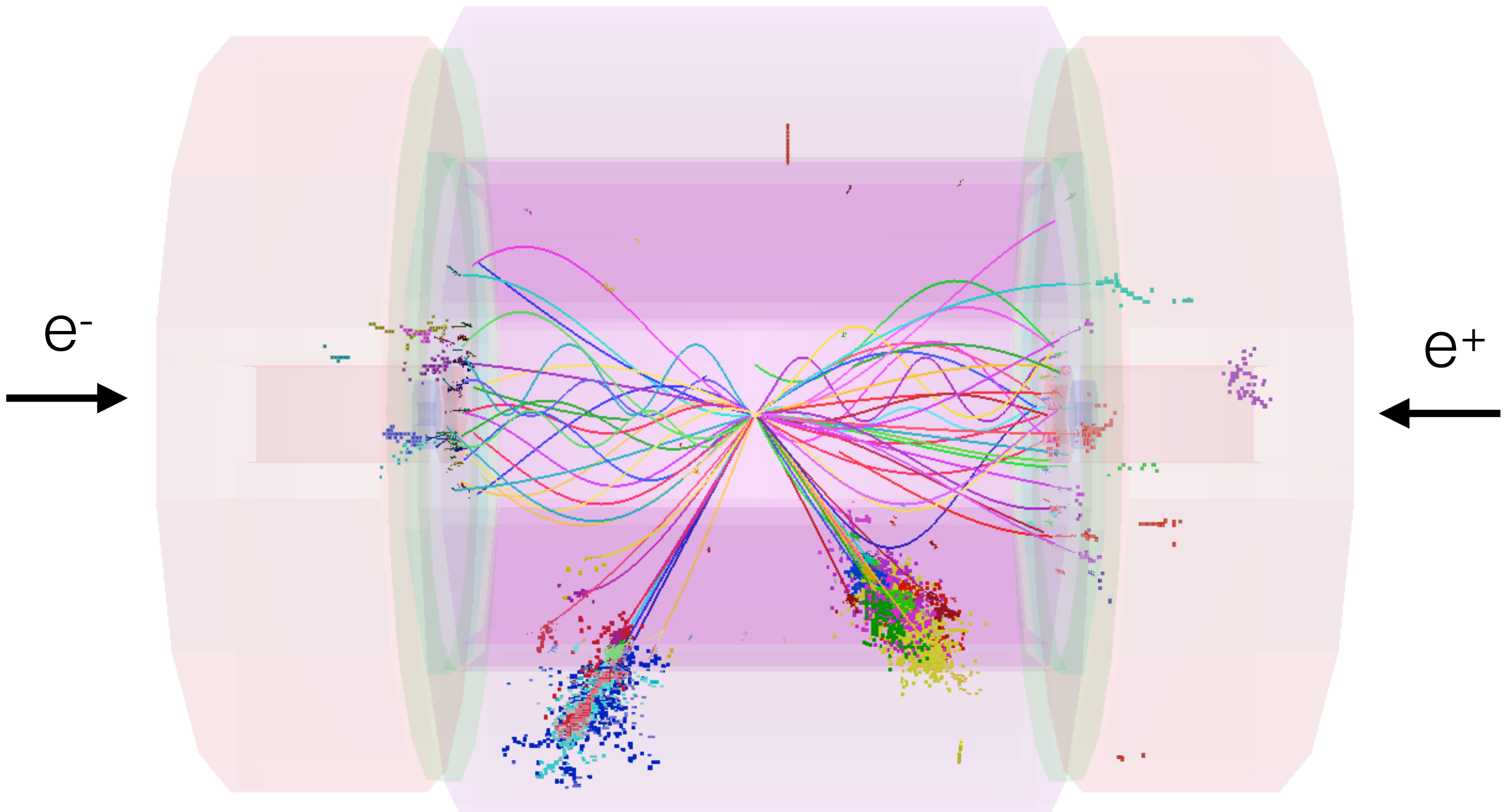


Image: Linear Collider Detector Project

Example: electron-positron collisions

Learn a classifier on the full observable phase space (momenta + particle flavor) and then check with some standard observables.

Our events have a variable number of particles & due to quantum mechanics, are permutation invariant. Thus, we use a deep-sets variant called **particle flow networks**.

PFNs: Komiske, Metodiev, Thaler, JHEP 01 (2019) 121

Deep sets: Zaheer et al., NIPS 2017

Example: electron-positron collisions

Learn a classifier on the full observable phase space (momenta + particle flavor) and then check with some standard observables.

Our events have a variable number of particles & due to quantum mechanics, a ... ve use a deep-sets varia ... .

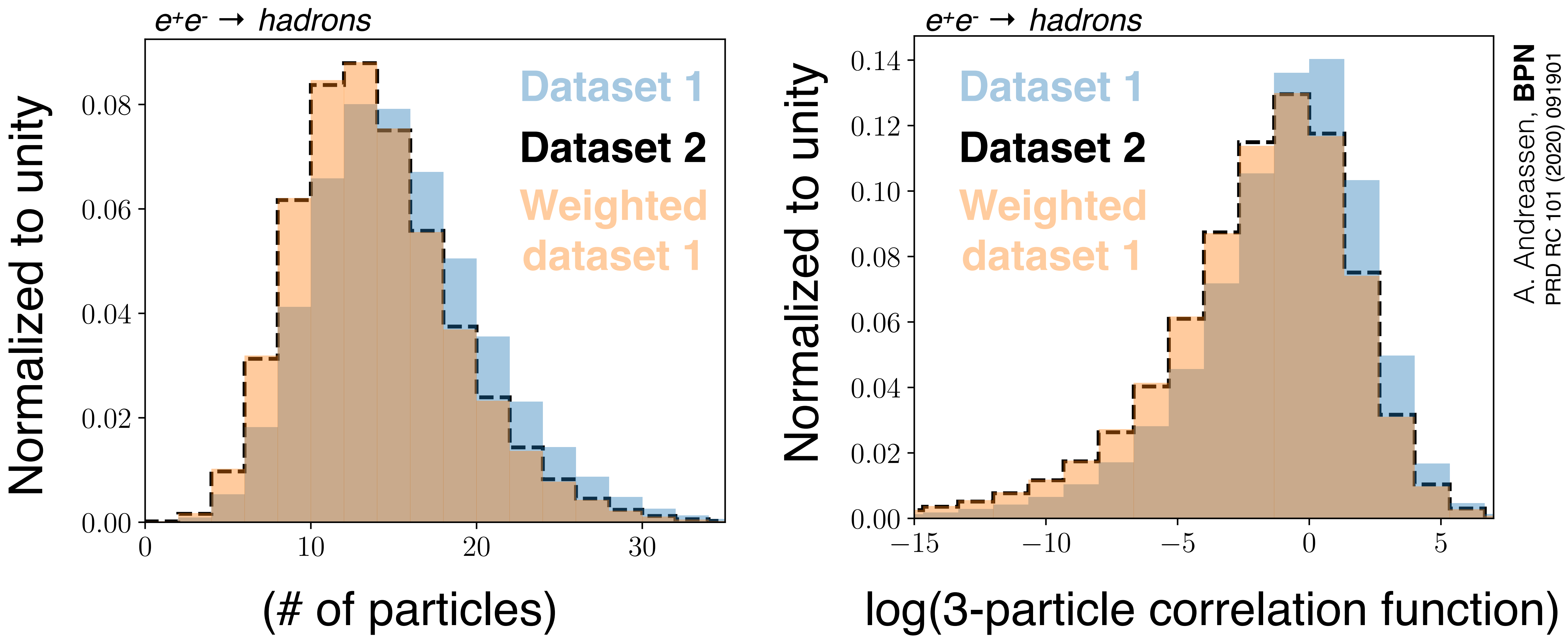
Just to stress: this gives you a new simulation with all the 4-vectors that is statistically indistinguishable.

PFNs: Komiske, Metodiev, Thaler, JHEP 01 (2019) 121

Deep sets: Zaheer et al., NIPS 2017

Classification for reweighting

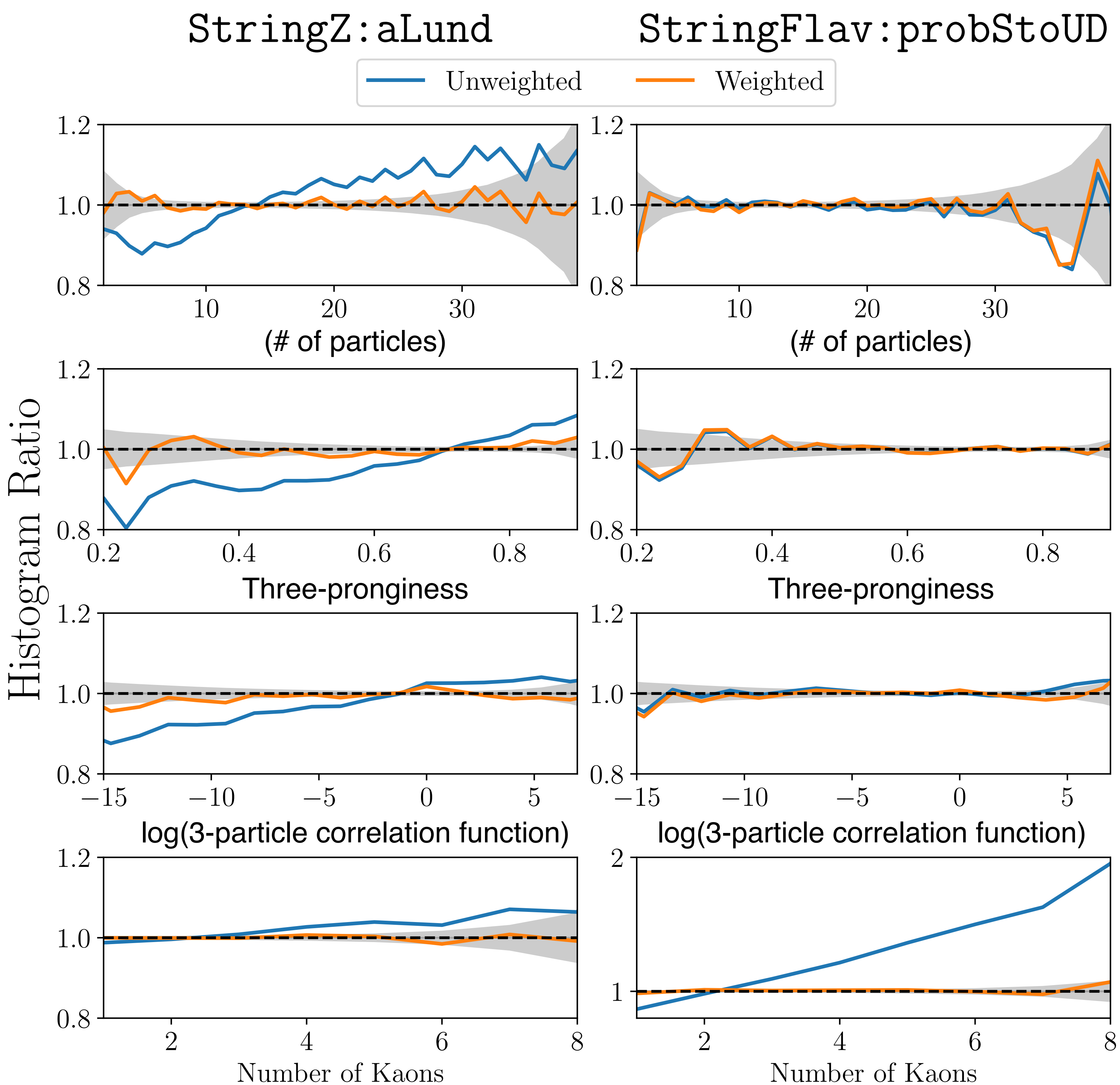
Reweight the **full phase space** and then check for various binned 1D observables.



A. Andreassen, **BPN**
PRD RC 101 (2020) 091901

We call this Deep neural networks using Classification for Tuning and Reweighting or "**DCTR**"

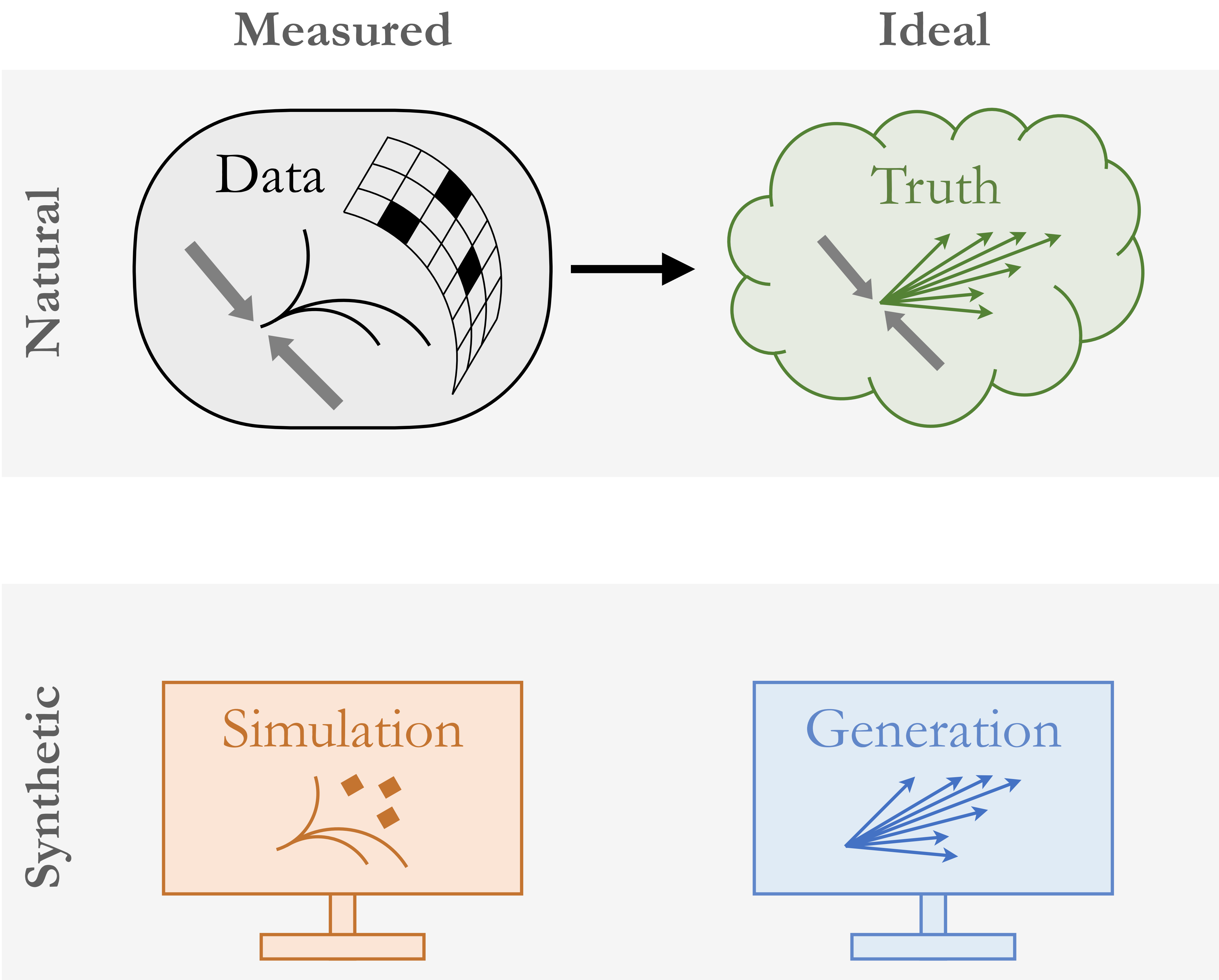
Achieving precision



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

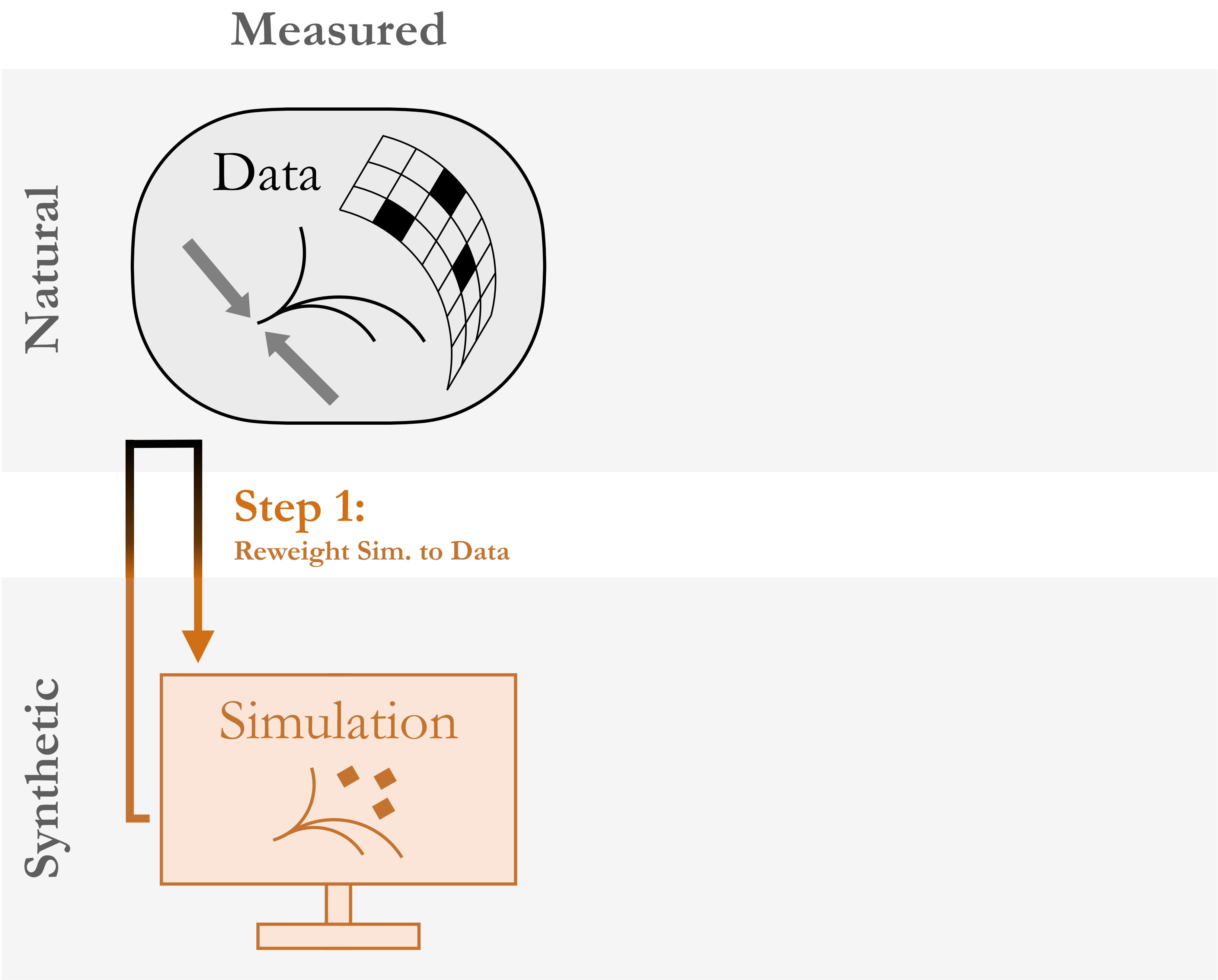
These are histogram ratios for a series of one-dimensional observables

Unfold by iterating: OmniFold



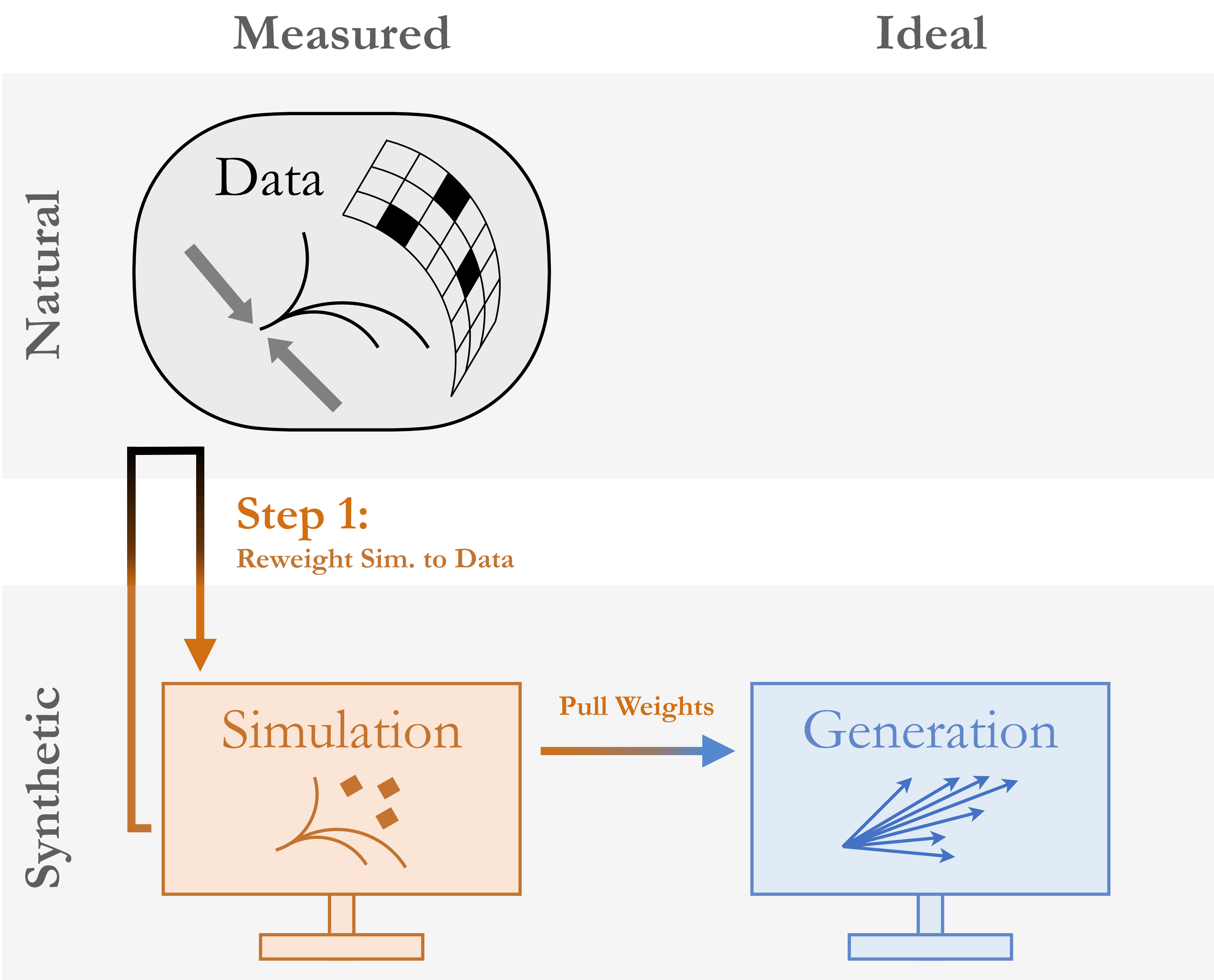
A. Andreassen, P. Komiske, E. Metodiev, **BPN**, J. Thaler, PRL 124 (2020) 182001

Unfold by iterating: OmniFold



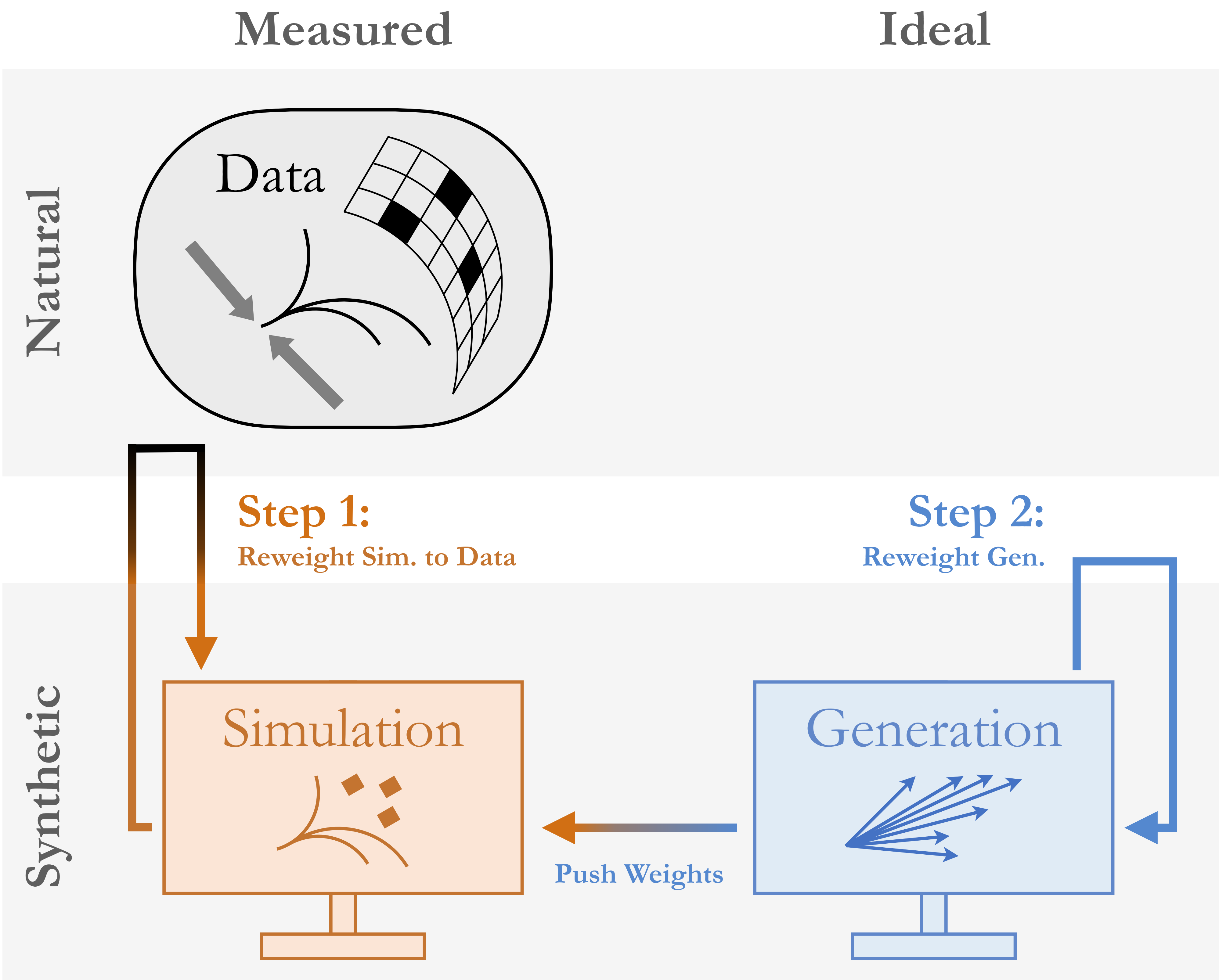
A. Andreassen, P. Komiske, E. Metodiev, **BPN**, J. Thaler, PRL 124 (2020) 182001

Unfold by iterating: OmniFold



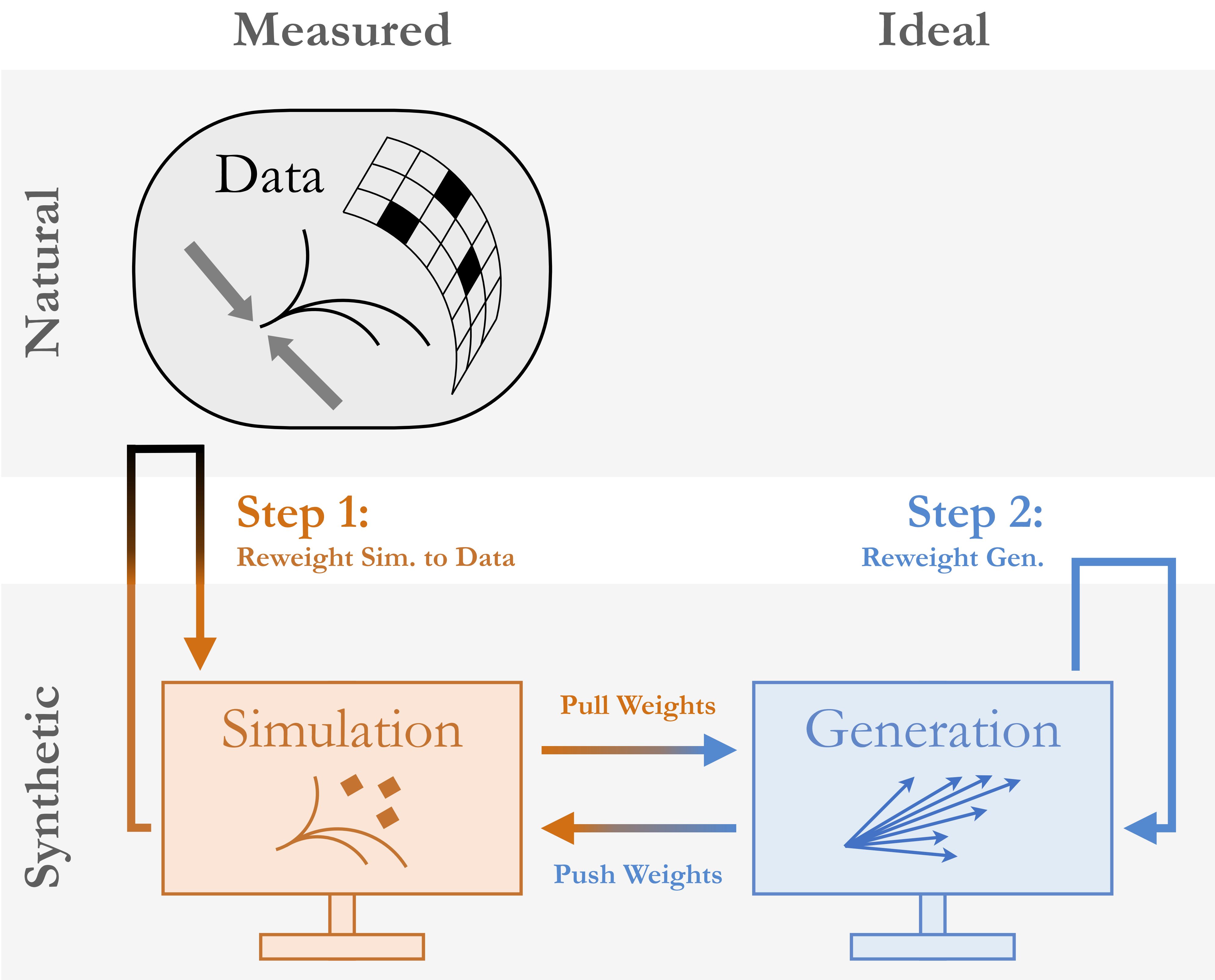
A. Andreassen, P. Komiske, E. Metodiev, **BPN**, J. Thaler, PRL 124 (2020) 182001

Unfold by iterating: OmniFold



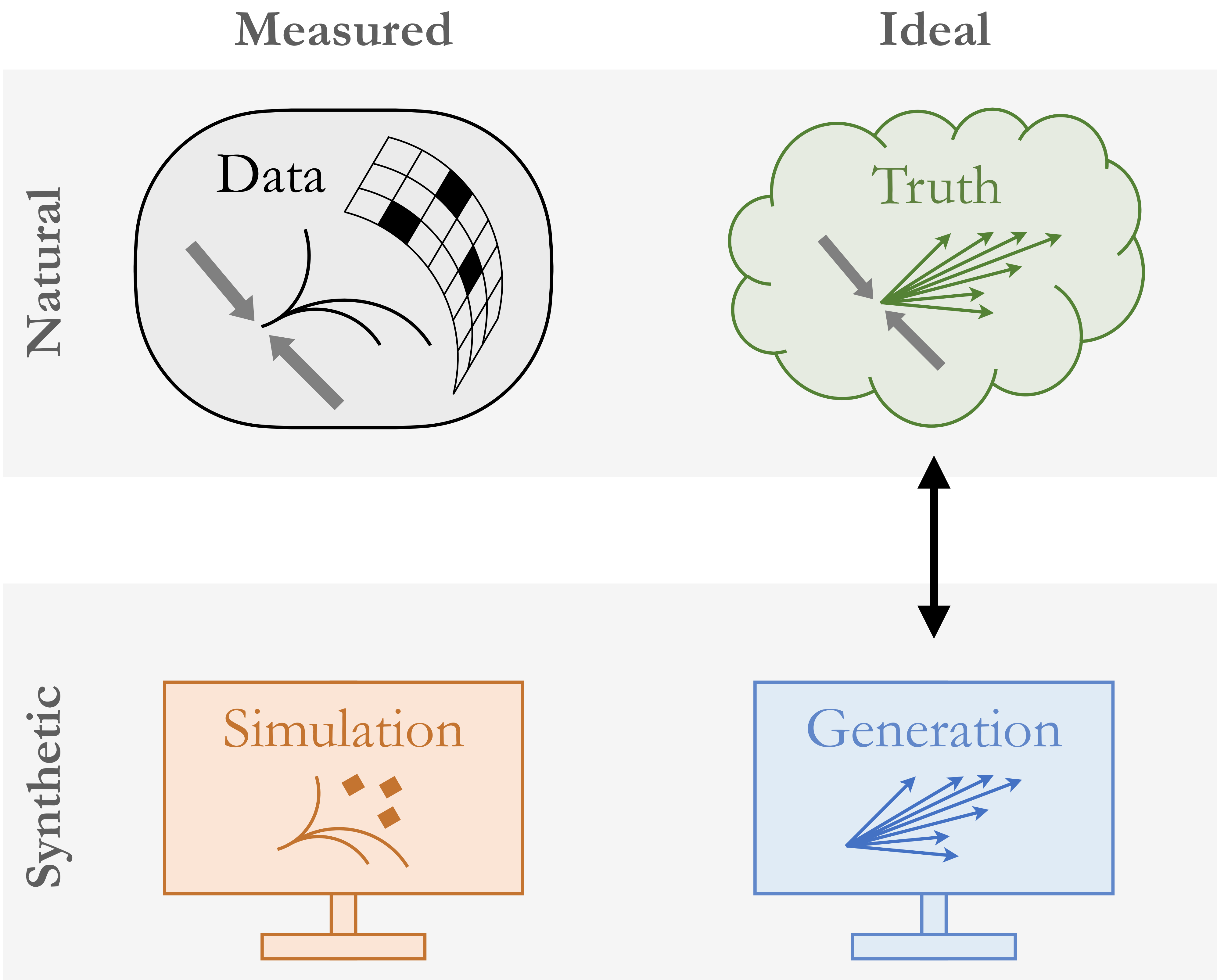
A. Andreassen, P. Komiske, E. Metodiev, **BPN**, J. Thaler, PRL 124 (2020) 182001

Unfold by iterating: OmniFold



A. Andreassen, P. Komiske, E. Metodiev, **BPN**, J. Thaler, PRL 124 (2020) 182001

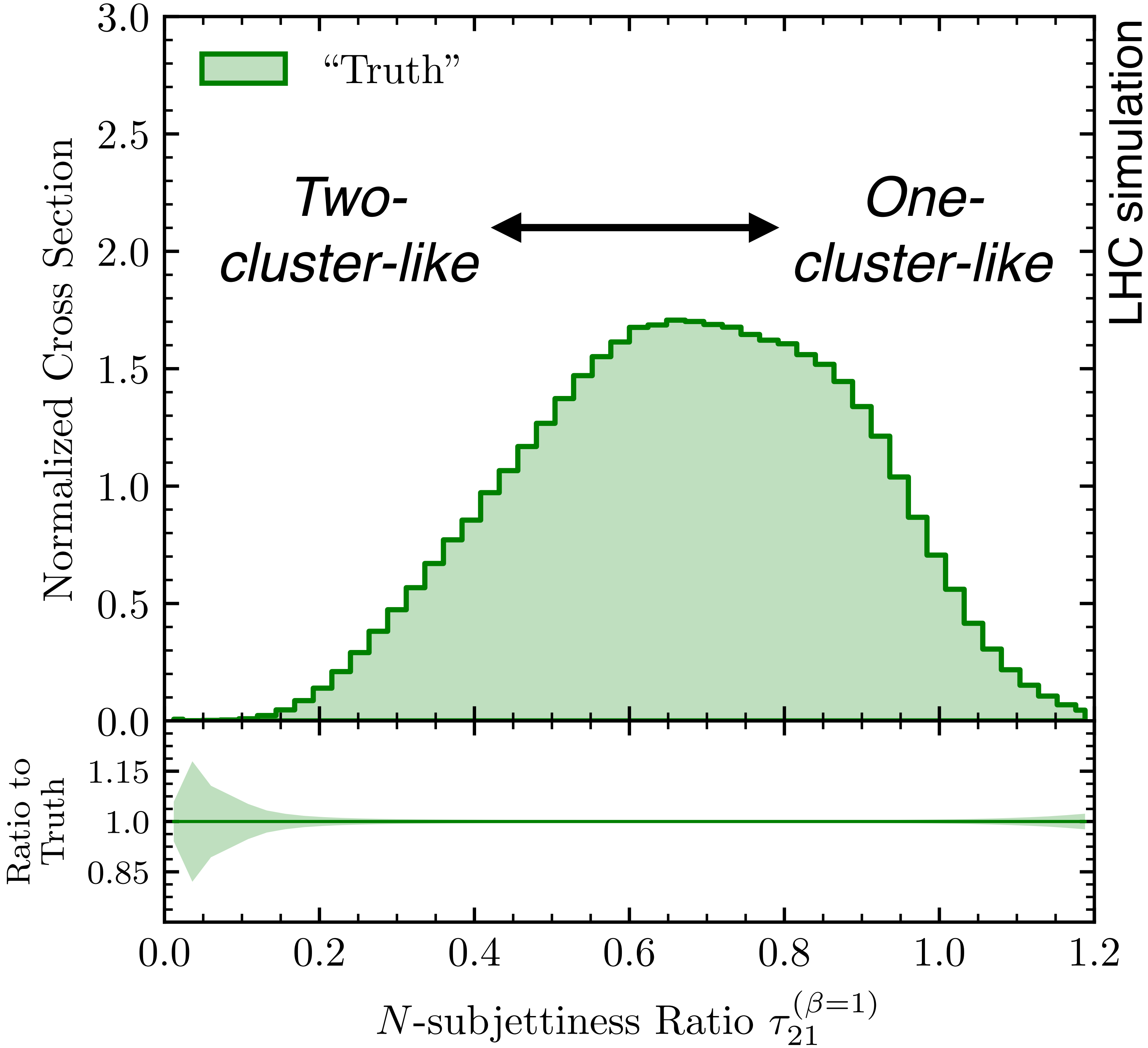
Unfold by iterating: OmniFold



A. Andreassen, P. Komiske, E. Metodiev, **BPN**, J. Thaler, PRL 124 (2020) 182001

Results

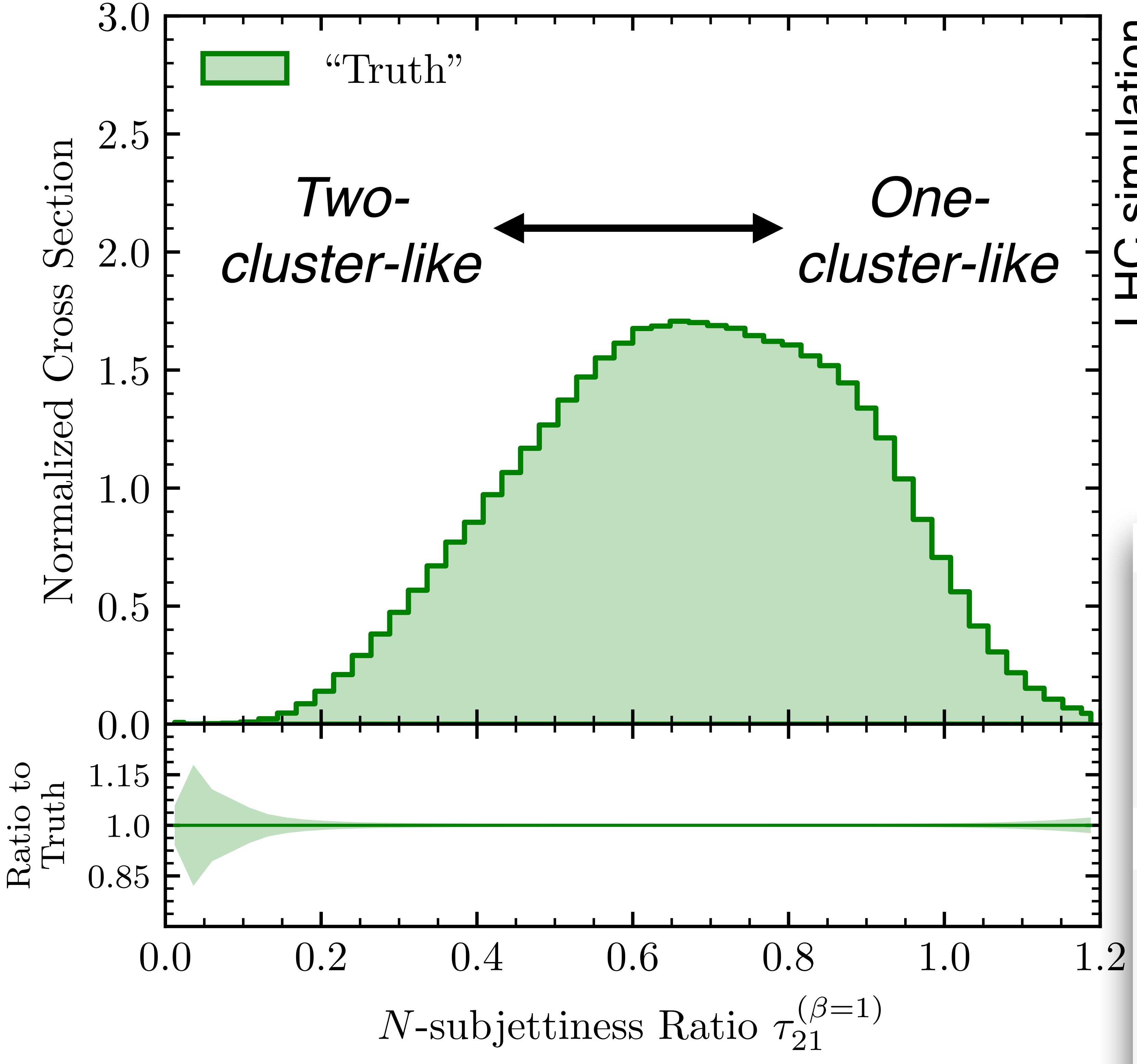
A. Andreassen, P. Komiske, E. Metodiev, **BPN**, J. Thaler, PRL 124 (2020) 182001



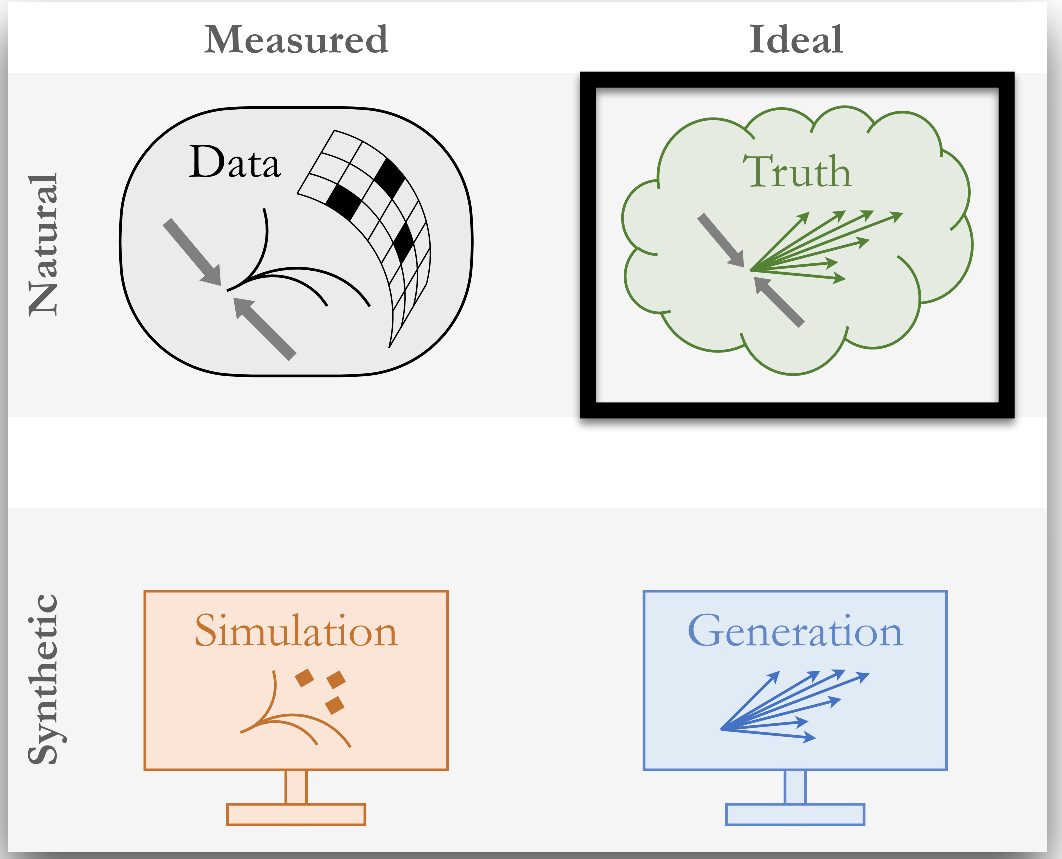
Consider this observable, which characterizes the substructure

Results

A. Andreassen, P. Komiske, E. Metodiev, **BPN**, J. Thaler, PRL 124 (2020) 182001

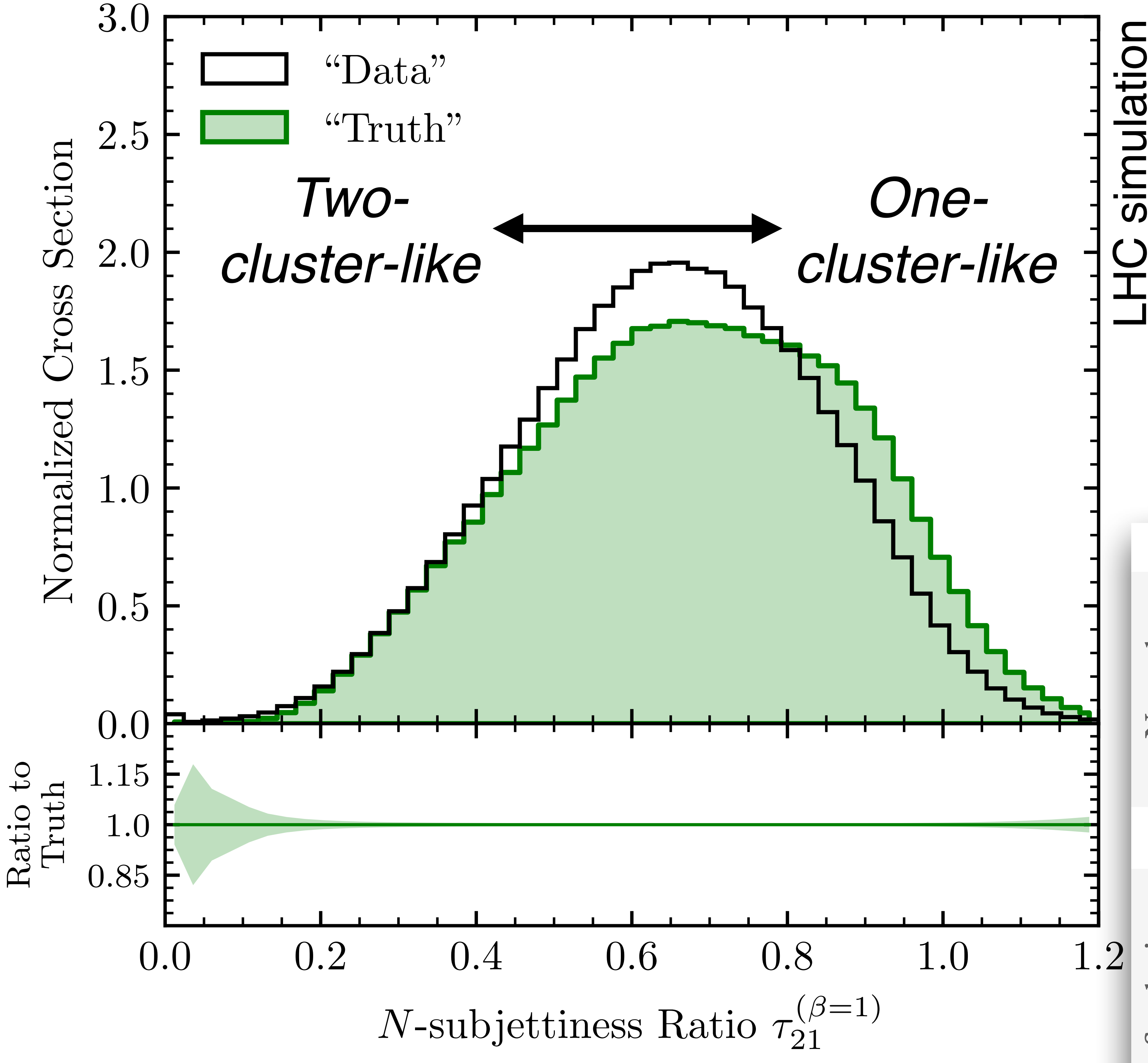


Consider this observable, which characterizes the substructure

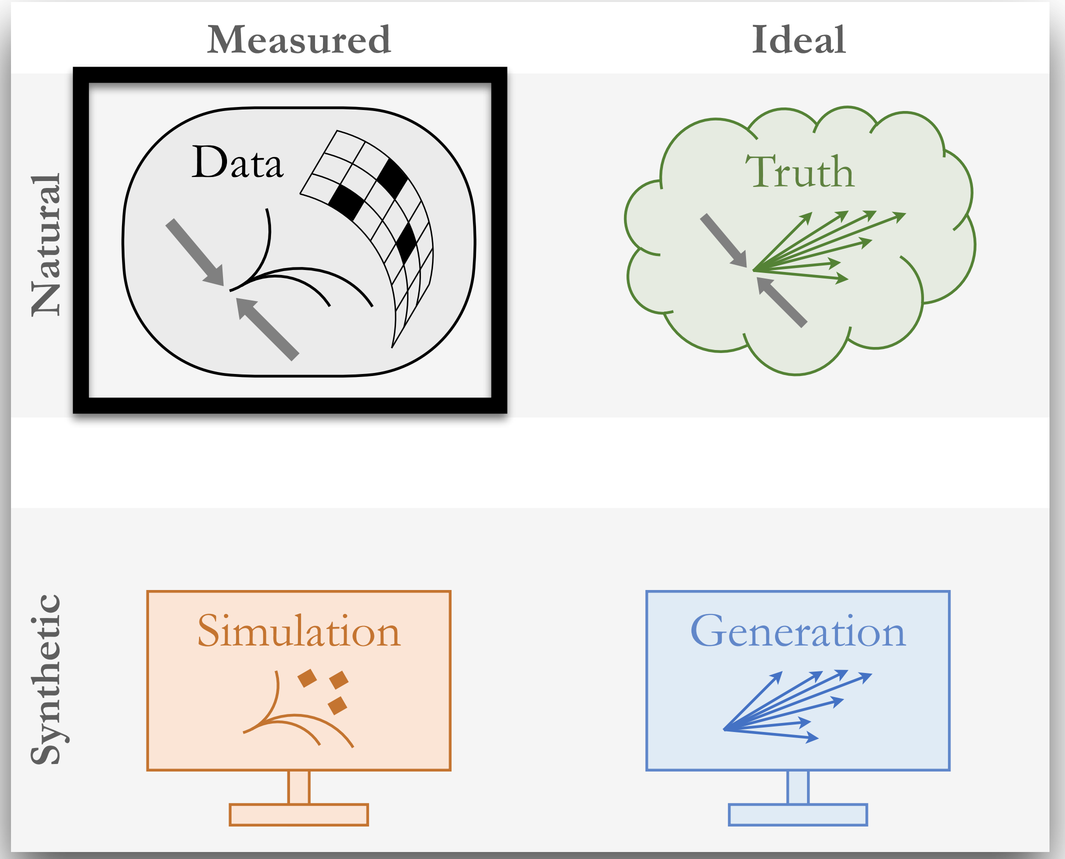


Results

A. Andreassen, P. Komiske, E. Metodiev, **BPN**, J. Thaler, PRL 124 (2020) 182001

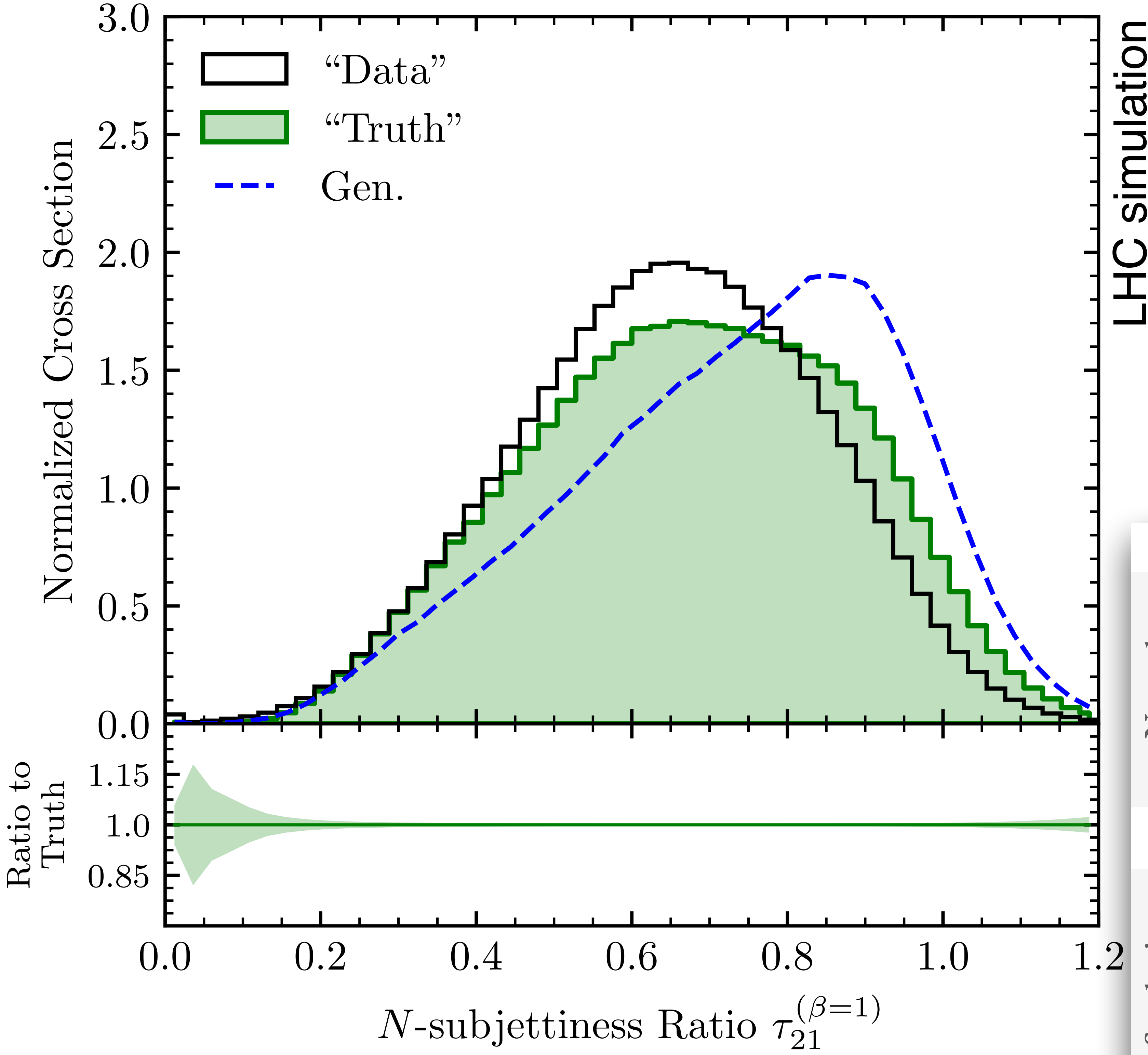


Consider this observable, which characterizes the substructure

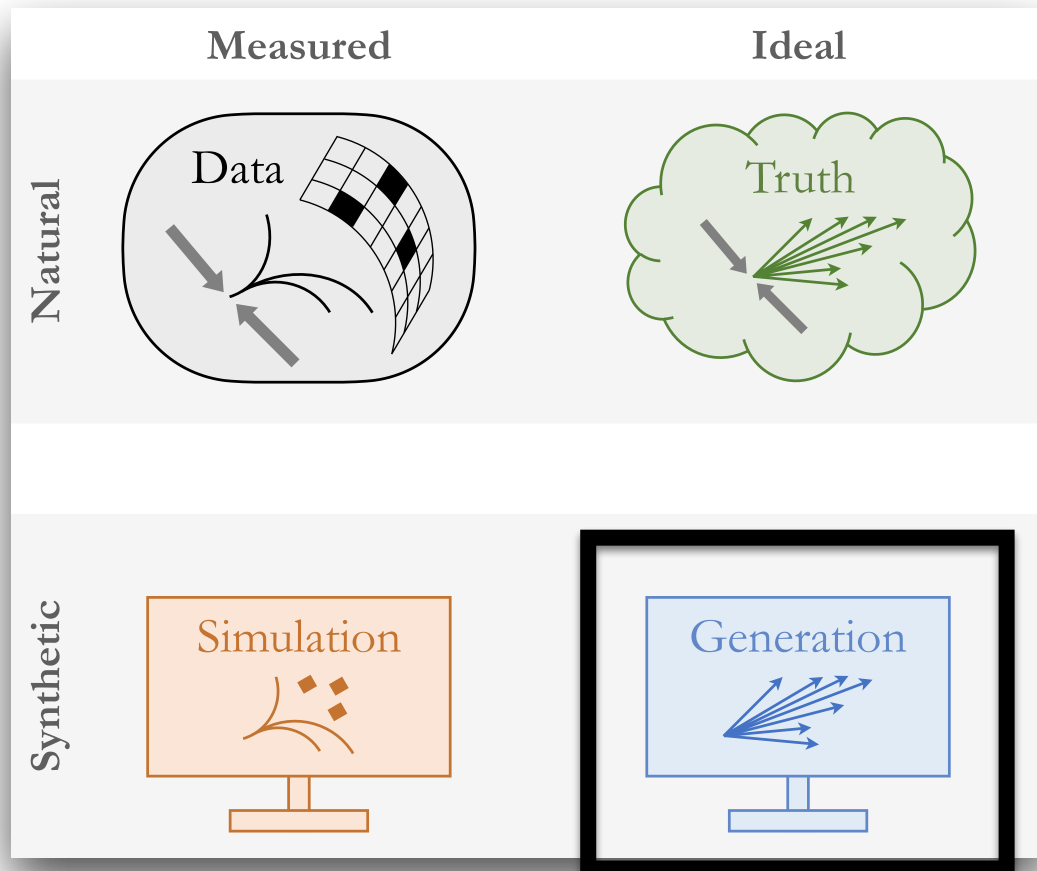


Results

A. Andreassen, P. Komiske, E. Metodiev, **BPN**, J. Thaler, PRL 124 (2020) 182001

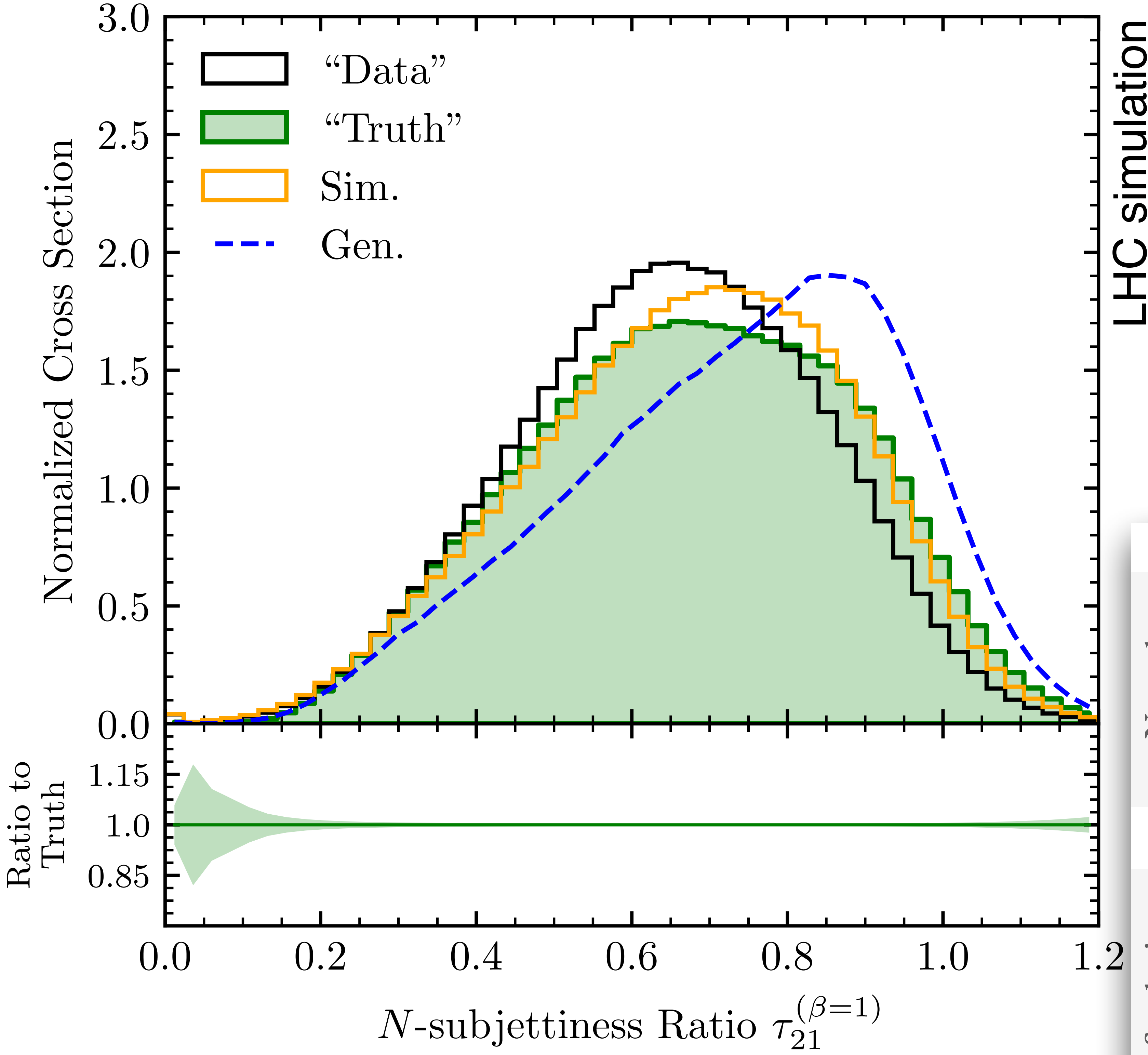


Consider this observable, which characterizes the substructure

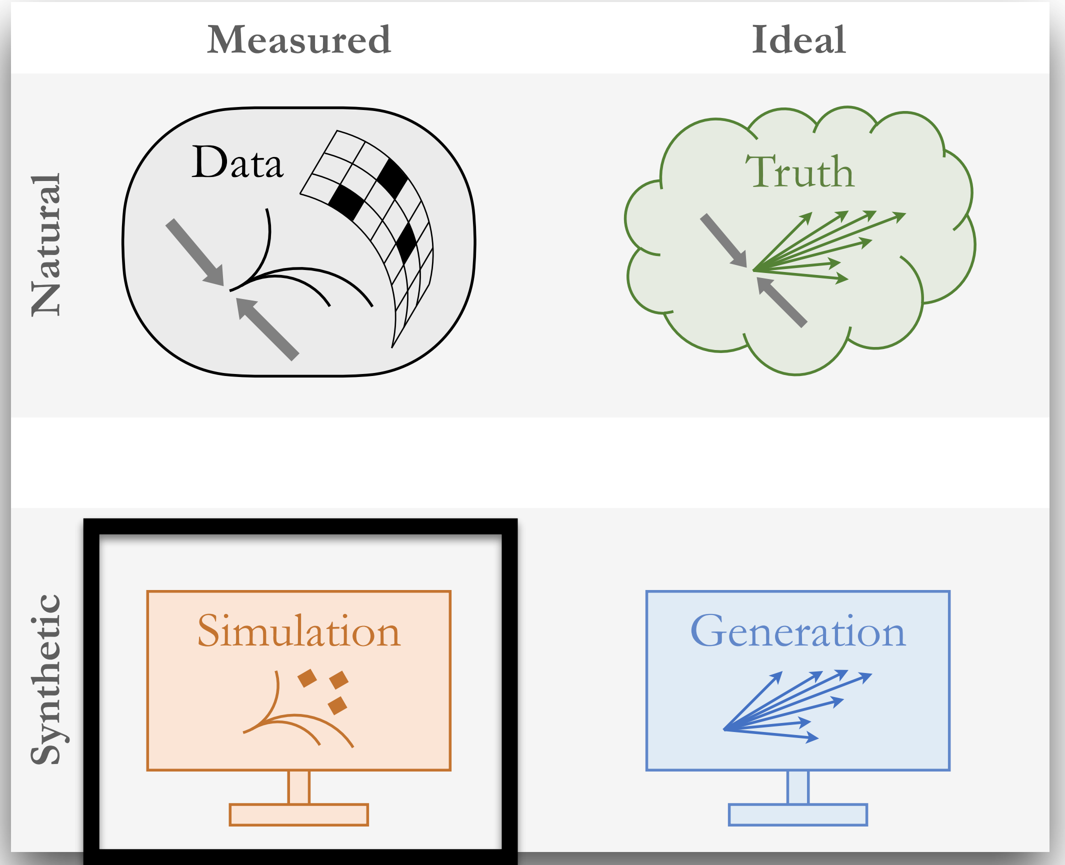


Results

A. Andreassen, P. Komiske, E. Metodiev, **BPN**, J. Thaler, PRL 124 (2020) 182001

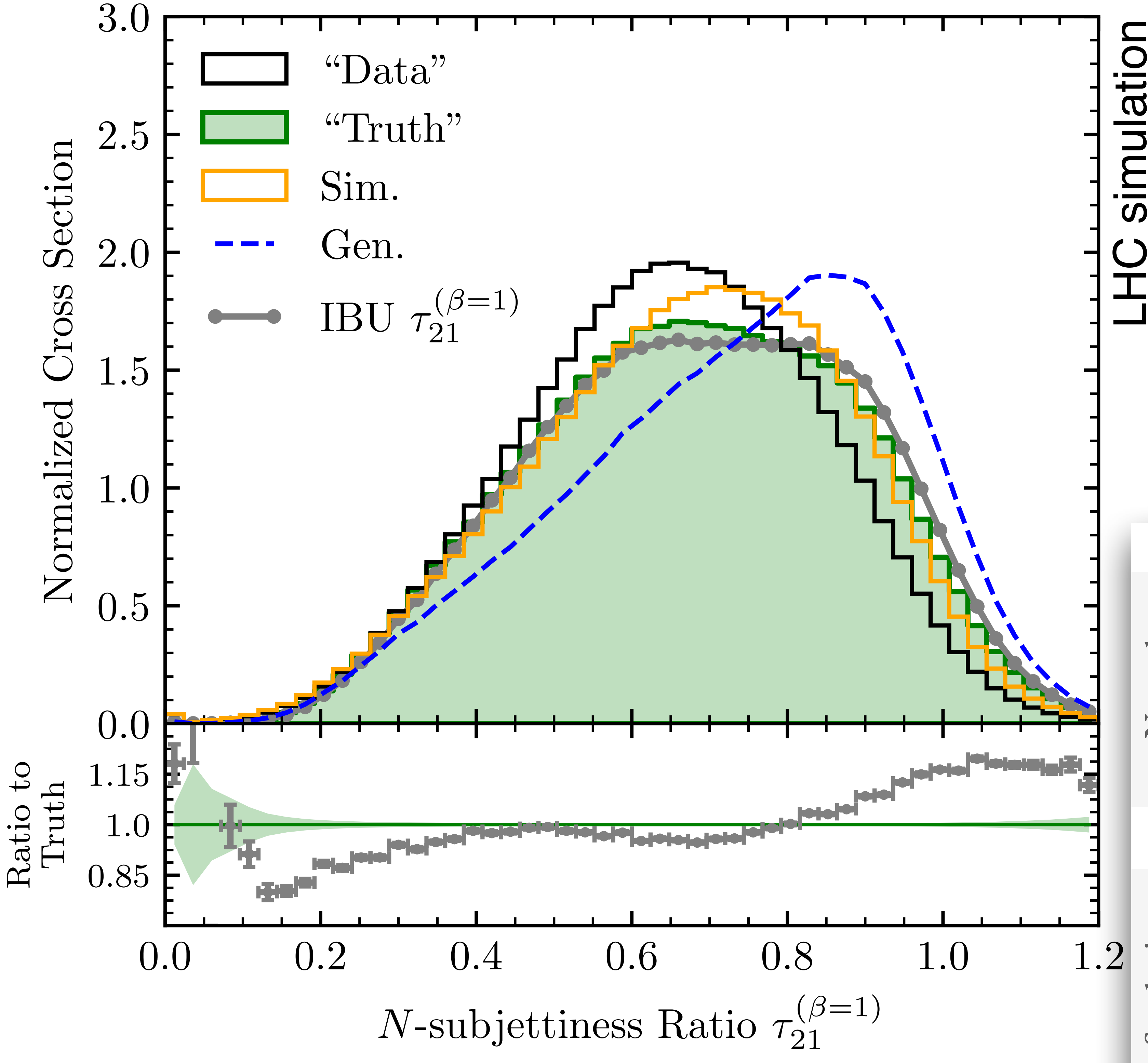


Consider this observable, which characterizes the substructure

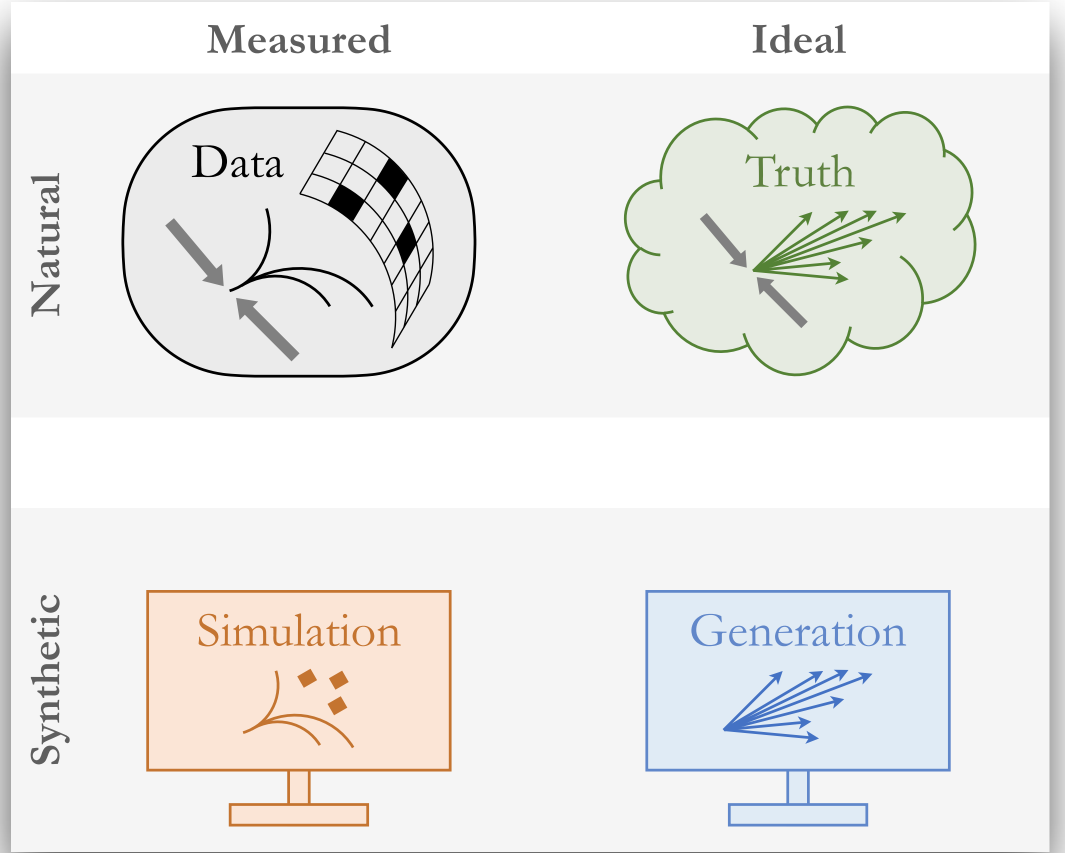


Results

A. Andreassen, P. Komiske, E. Metodiev, **BPN**, J. Thaler, PRL 124 (2020) 182001

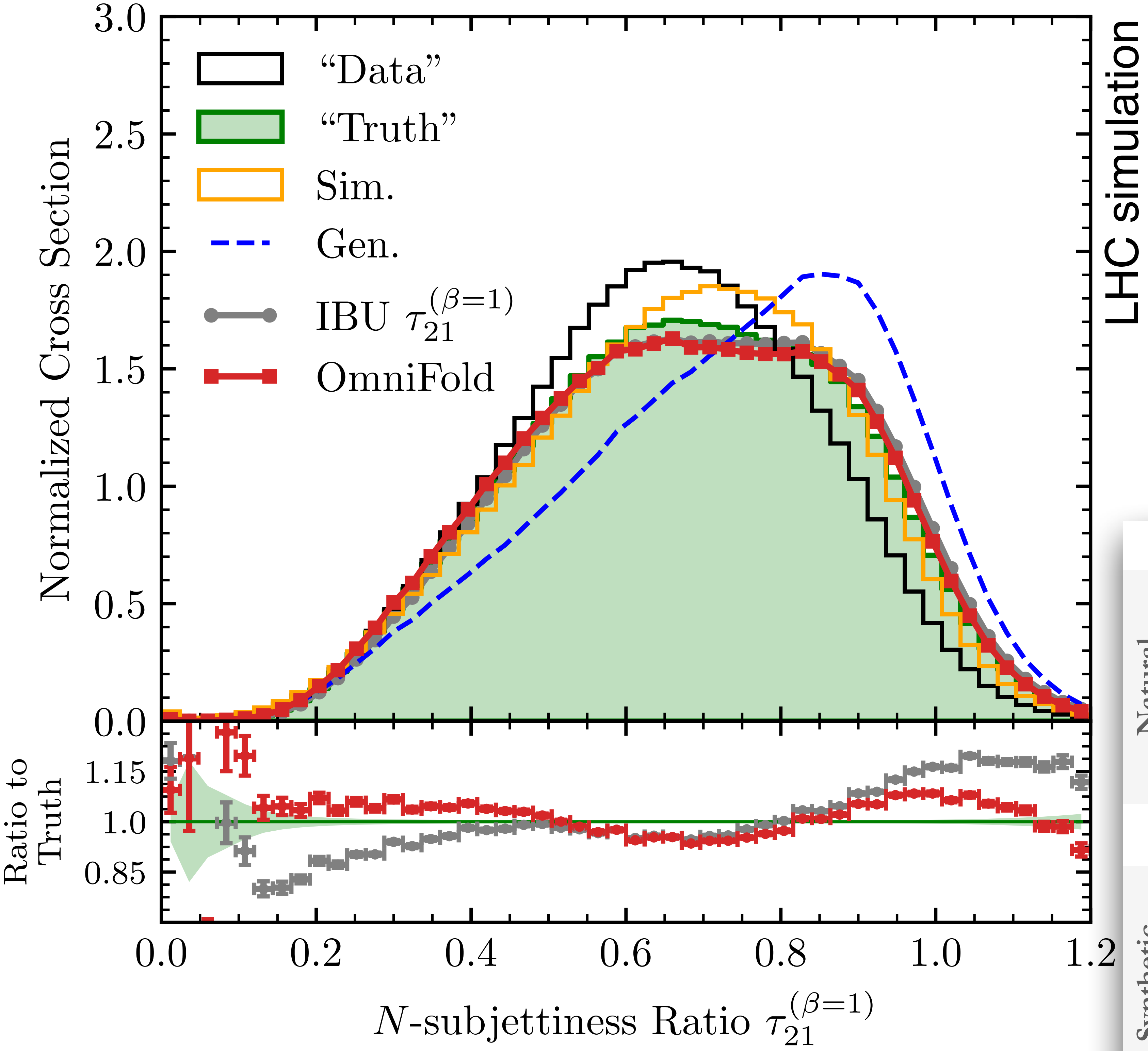


IBU is the current standard. It is a1D binned and iterative approach.

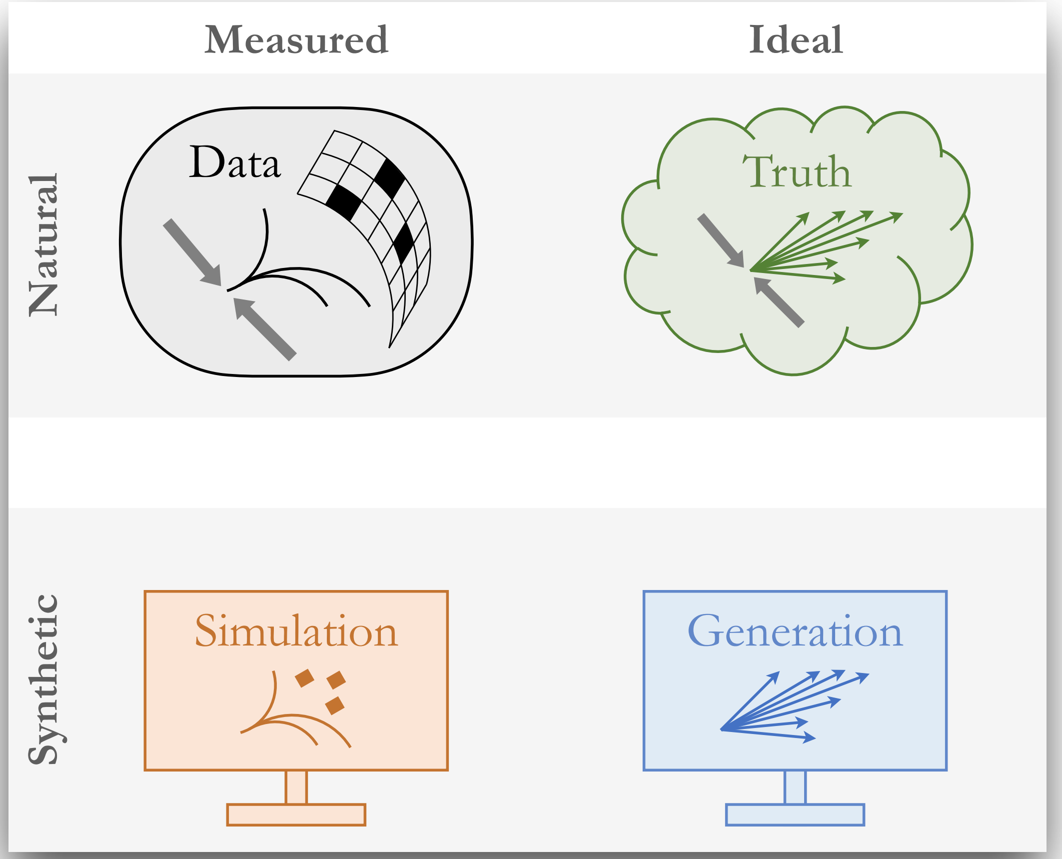


Results

A. Andreassen, P. Komiske, E. Metodiev, **BPN**, J. Thaler, PRL 124 (2020) 182001

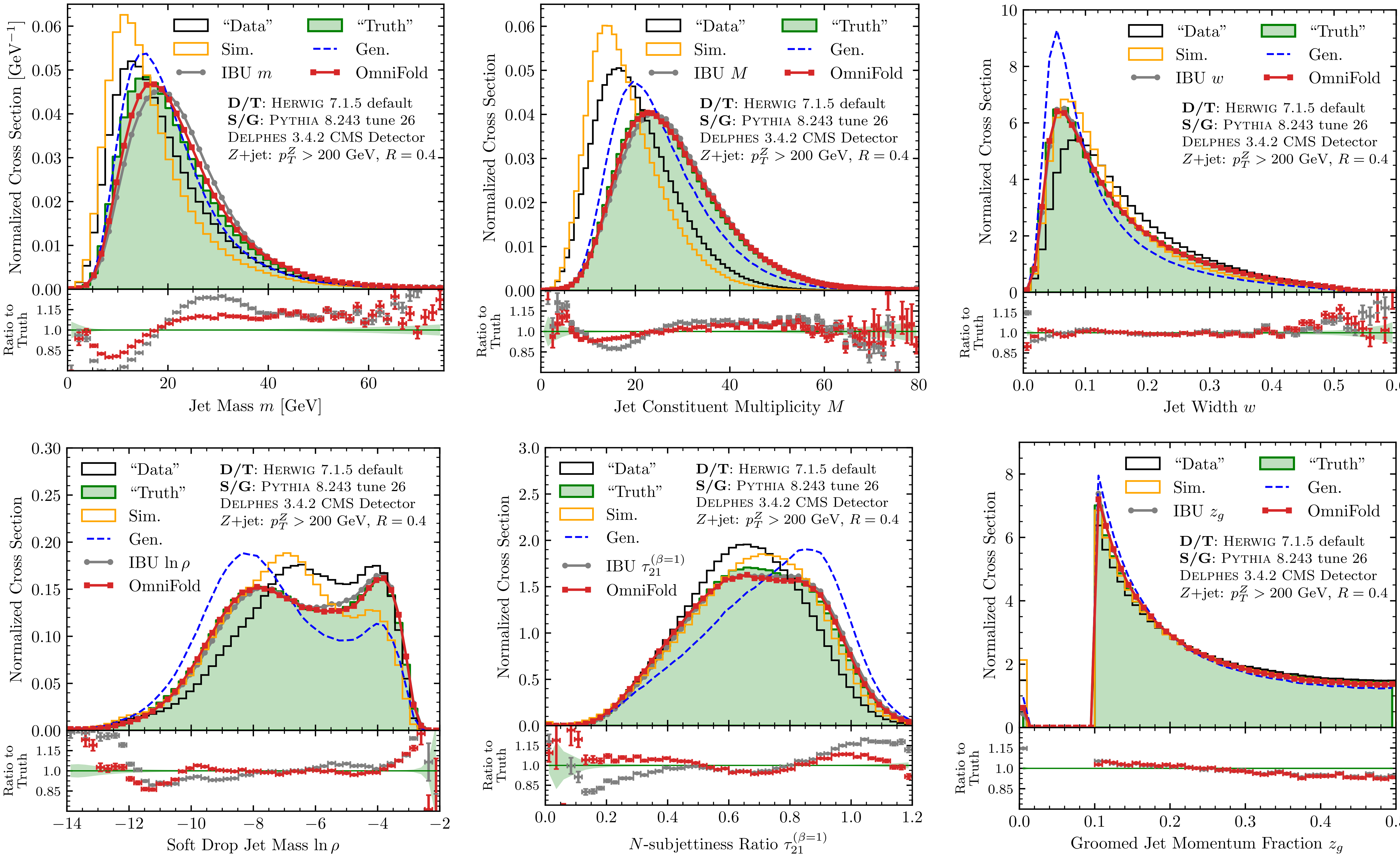


OmniFold outperforms IBU even though it is not tailored to this observable



Results

A. Andreassen, P. Komiske, E. Metodiev, **BPN**, J. Thaler, PRL 124 (2020) 182001

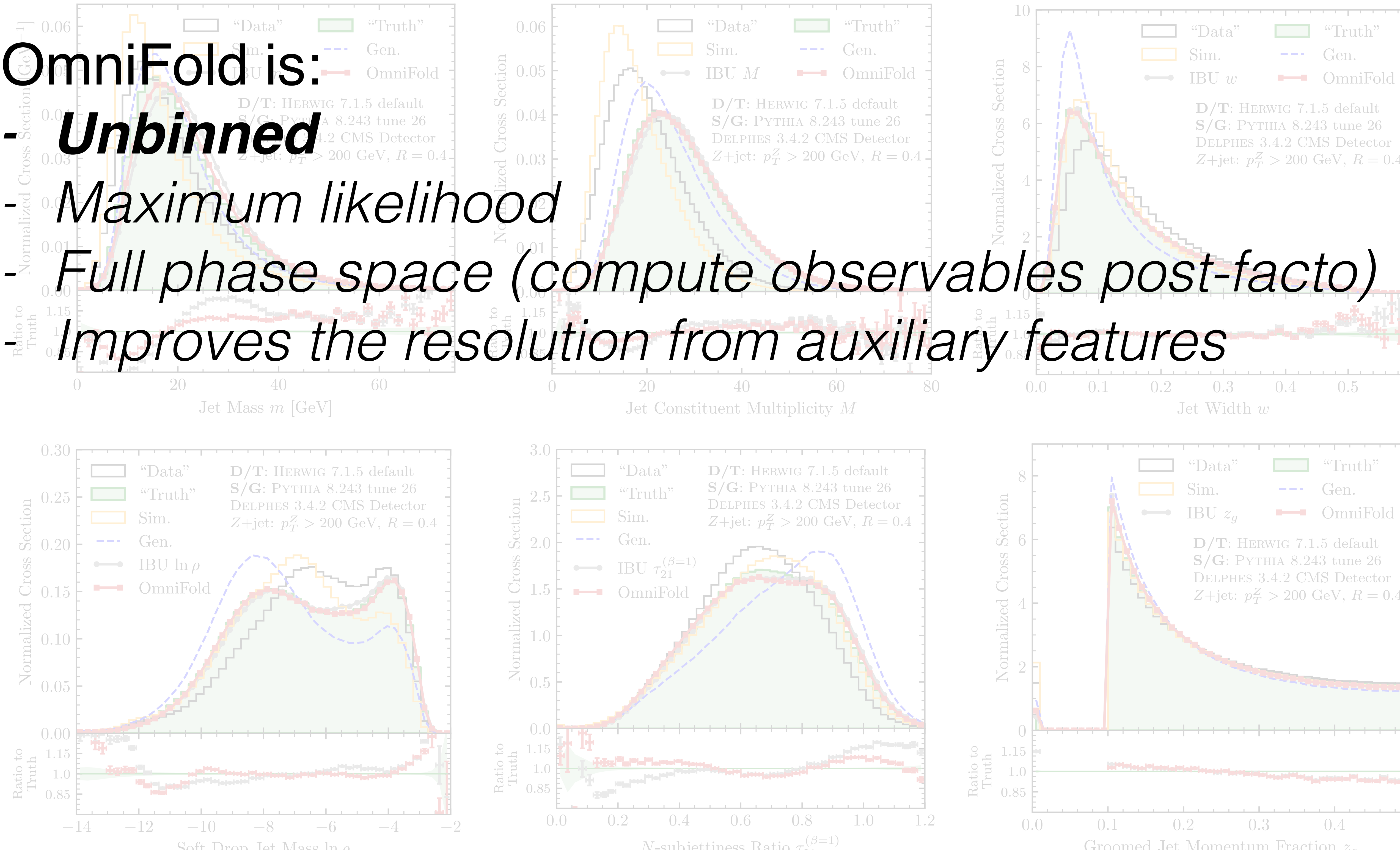


Results

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

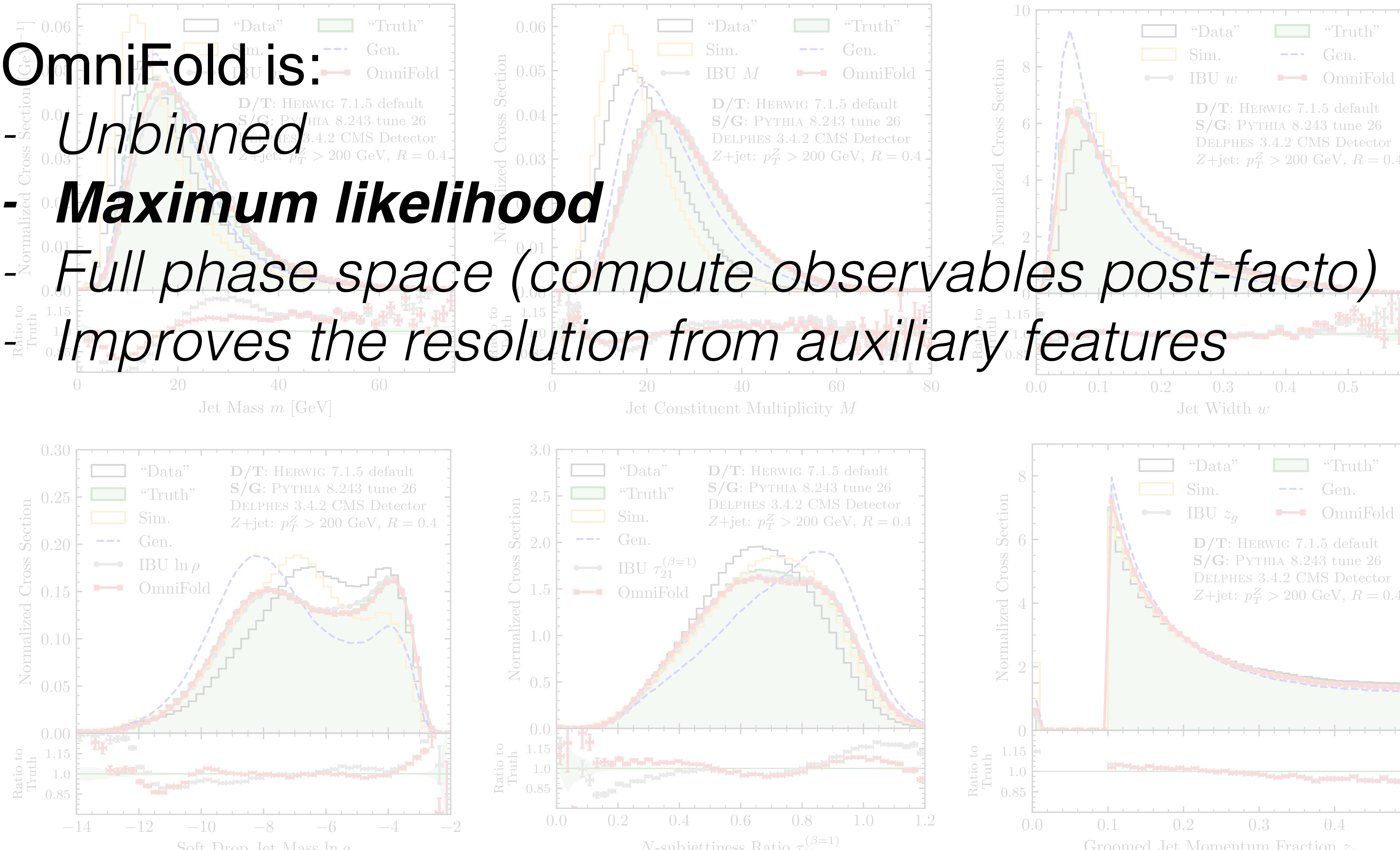
OmniFold is:

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



Results

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

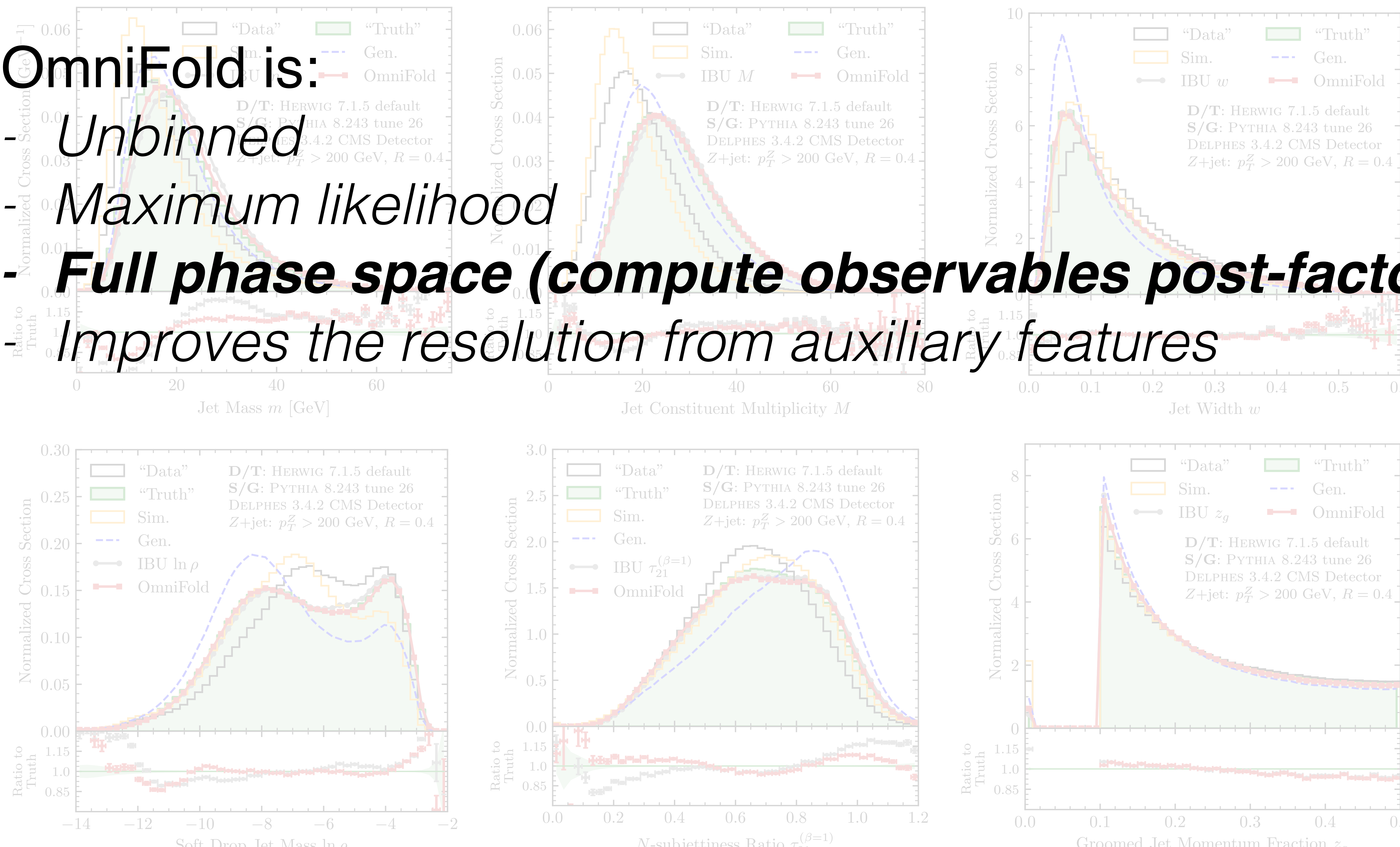


OmniFold is:

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

Results

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

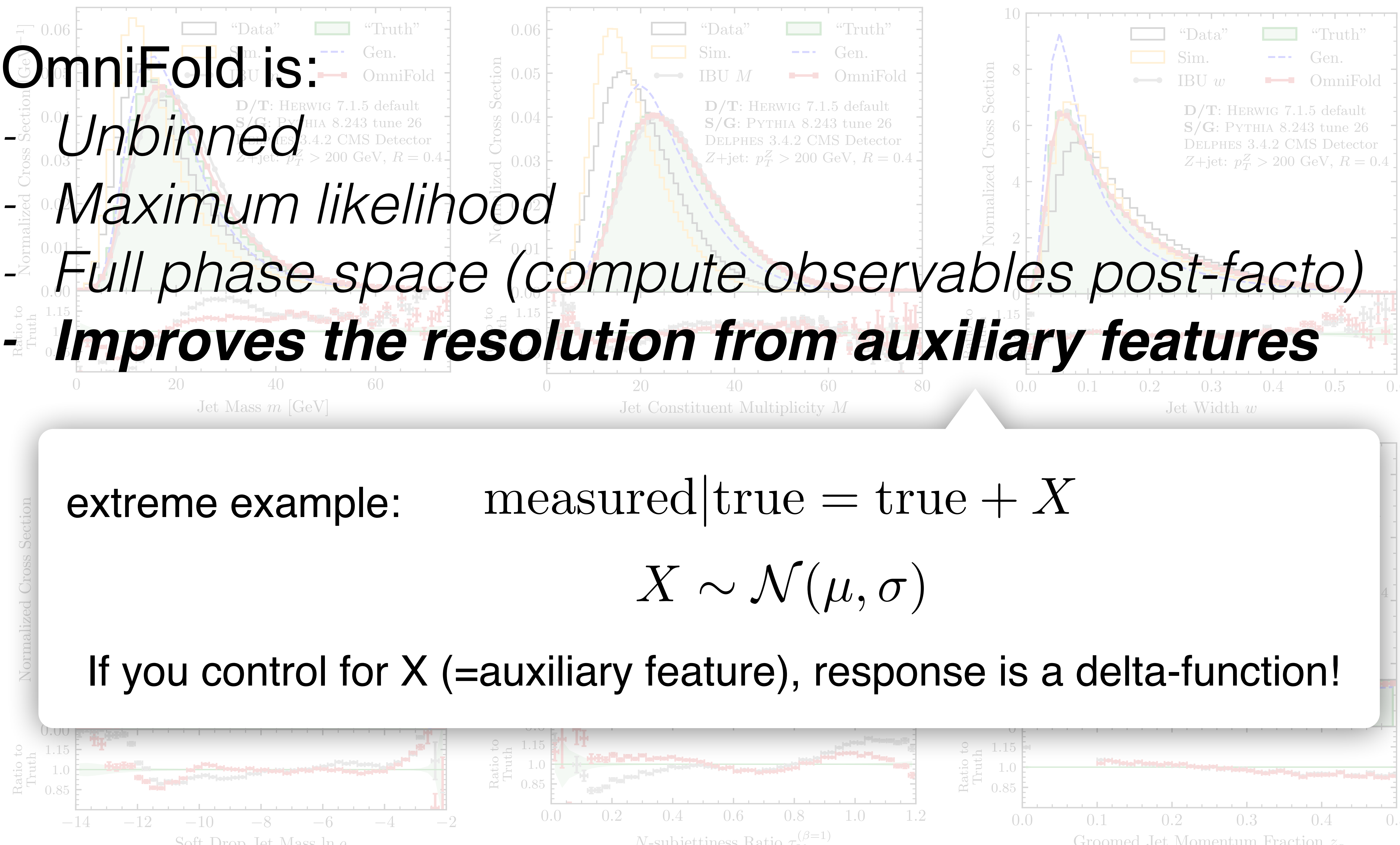


OmniFold is:

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

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***

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

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

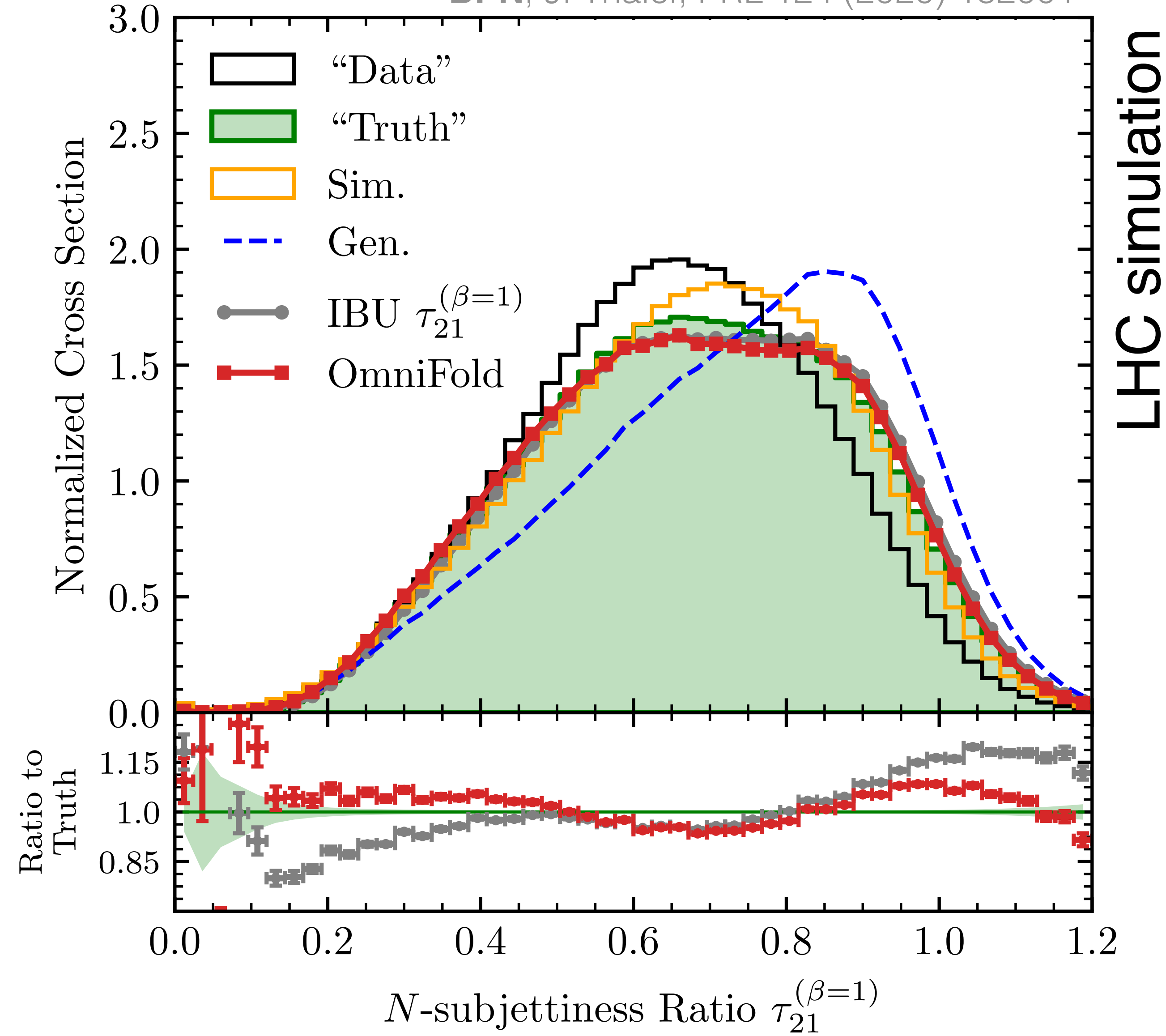
If you control for X (=auxiliary feature), response is a delta-function!

Future

We have started an effort to perform an OmniFold in the Z+jets final state.

Eric Metodiev (MIT) has joined ATLAS as a short term associate to collaborate on this project.

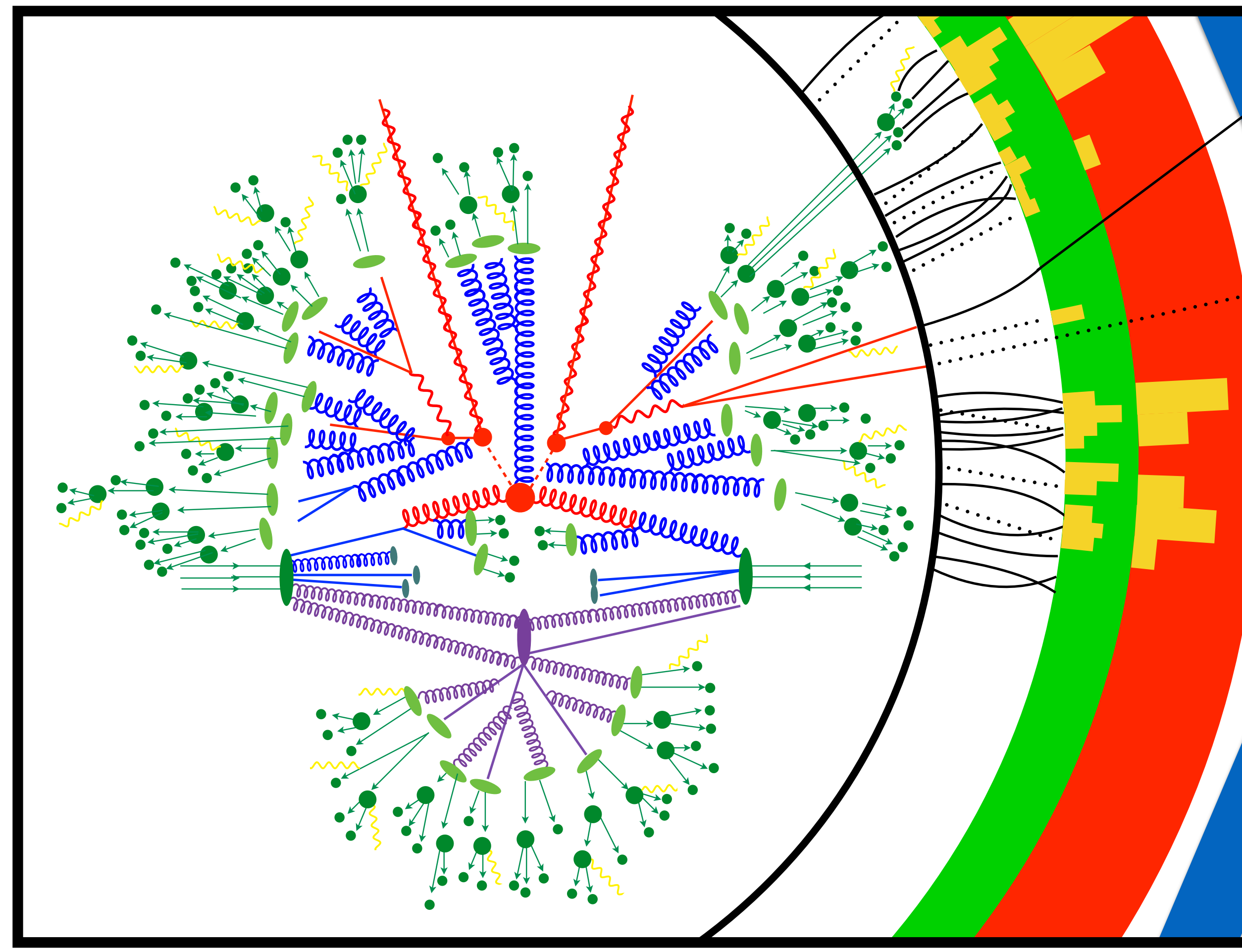
A. Andreassen, P. Komiske, E. Metodiev, **BPN**, J. Thaler, PRL 124 (2020) 182001



Exciting challenges (just to name a few): uncertainties? How to present the result? (unbinned + high-dimensional)

Conclusions and outlook

Deep learning has a great potential to **enhance**, **accelerate**, and **empower** HEP analyses



Today, I only spoke about unfolding, but there is great potential in other areas as well.

The **full phase space** of our experiments is now explorable with deep learning … it is an exciting time to be the pioneers in this hypervariate phase space !

Backup

Full phase space unfolding: OmniFold

Emily Dickinson, #975

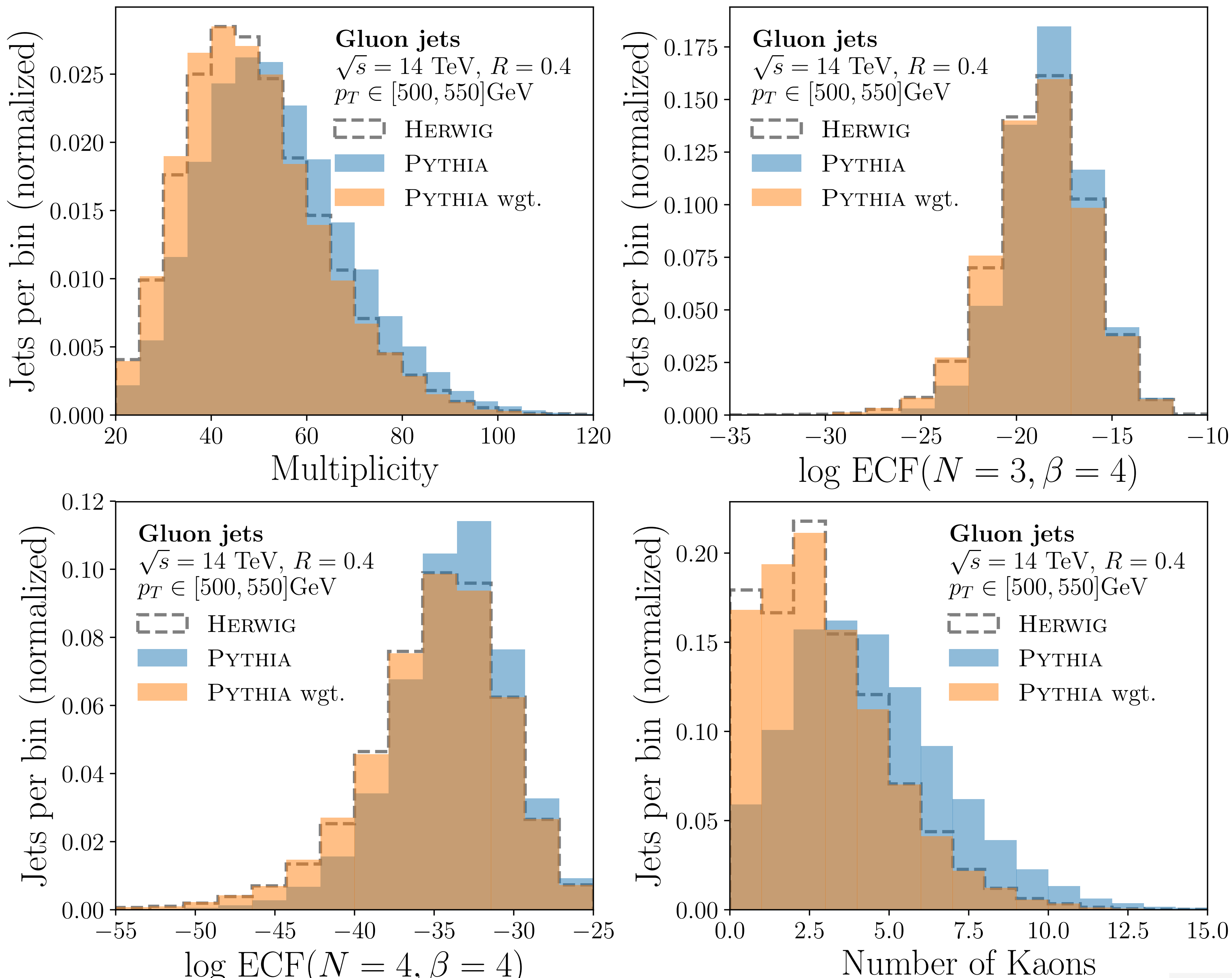
The Mountain sat upon the Plain
In his tremendous Chair –
His observation omnifold,
His inquest, everywhere –

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

Pythia versus Herwig

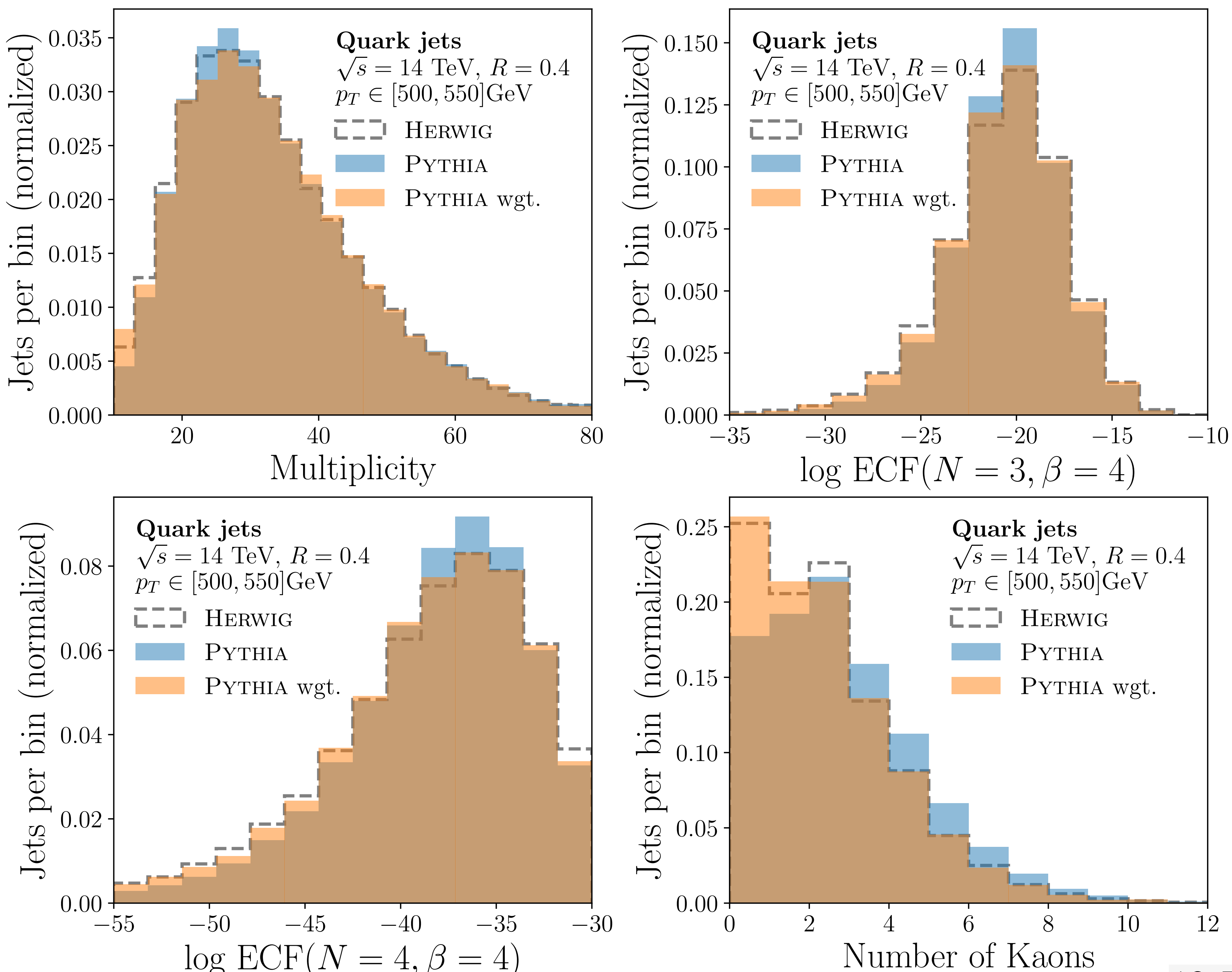
No hyper-parameter tuning - out of the box!



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