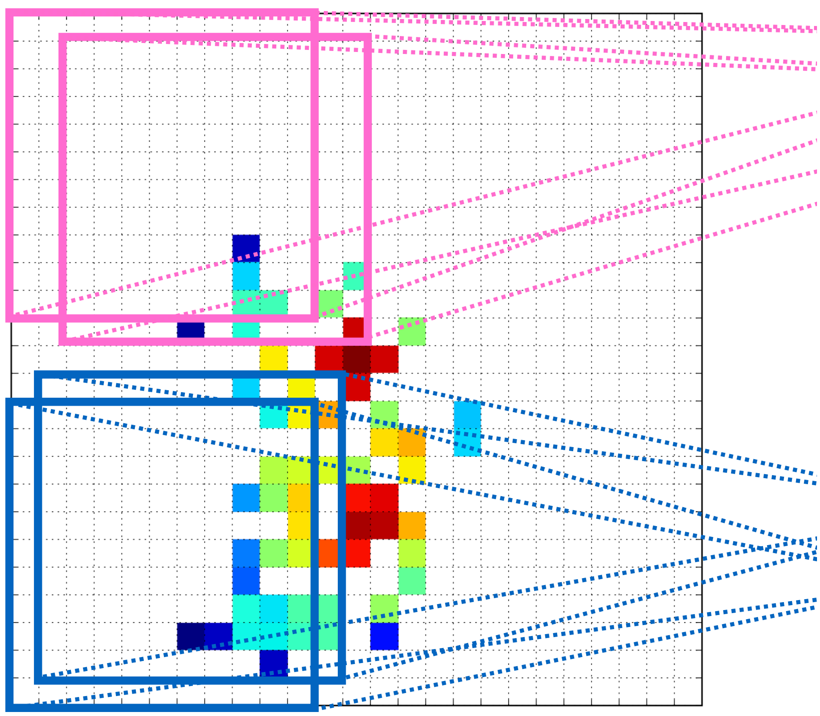
High dimensional jet measurements (at the EIC)

Benjamin Nachman

Lawrence Berkeley National Laboratory

bpnachman.com
bpnachman@lbl.gov

@bpnachman bnachman

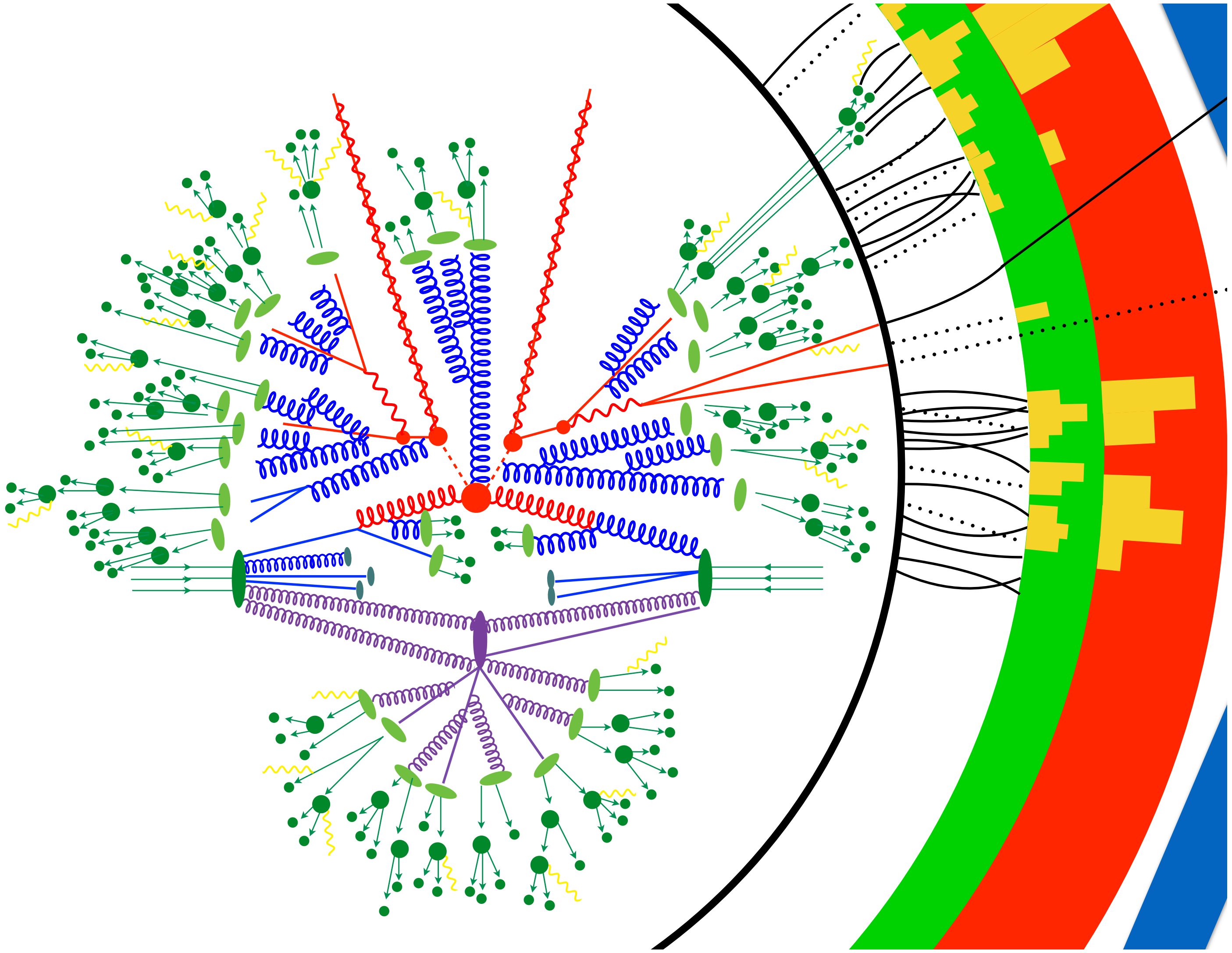


BERKELEY LAB

BIDS
BERKELEY INSTITUTE FOR DATA SCIENCE

EIC opportunities for Snowmass
Jan. 27, 2021

Measurements at the EIC



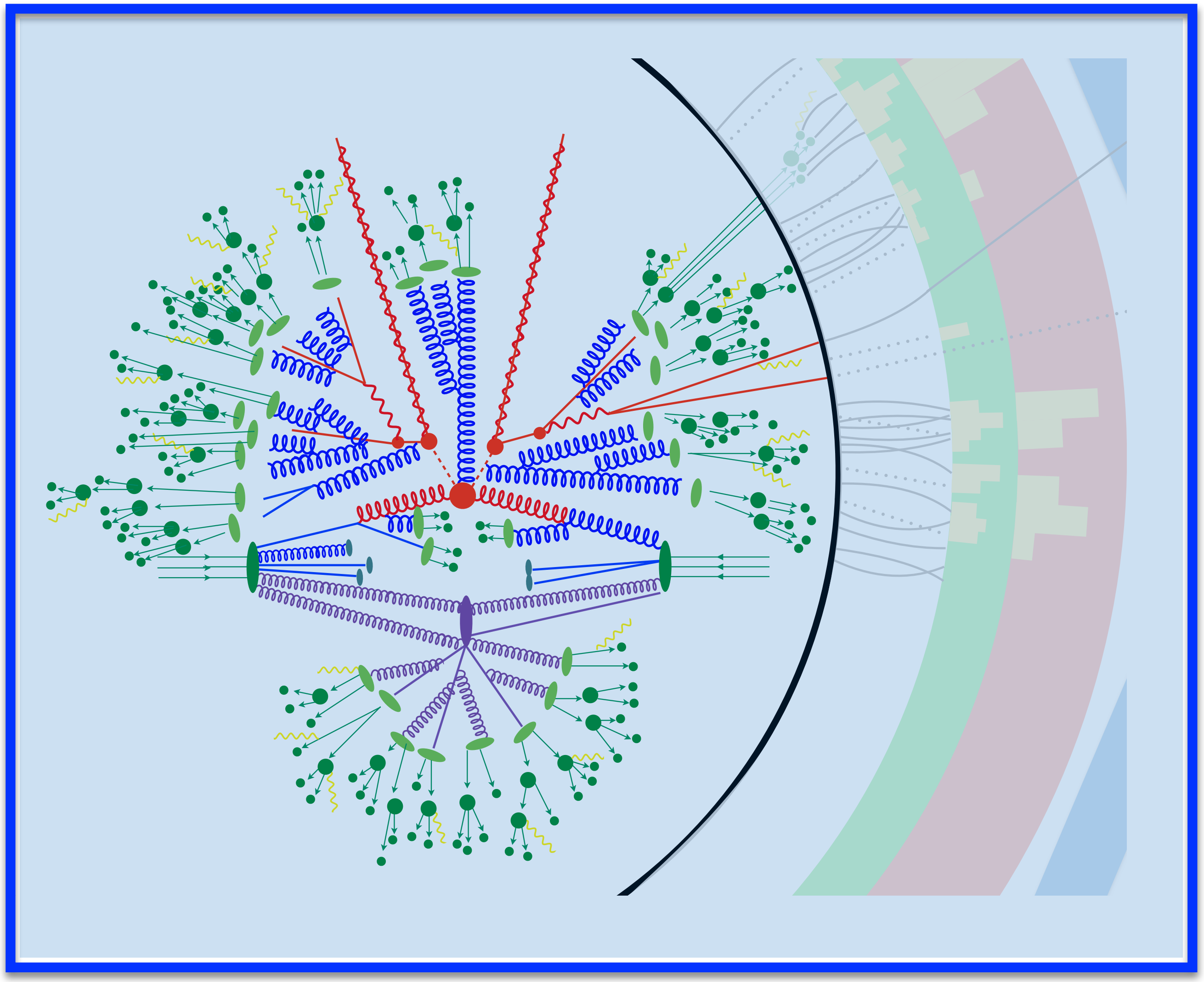
To extract the most QCD (+ more) from the EIC, we will want to make highly differential measurements.

This will be possible with the high statistics and fine detector precision.

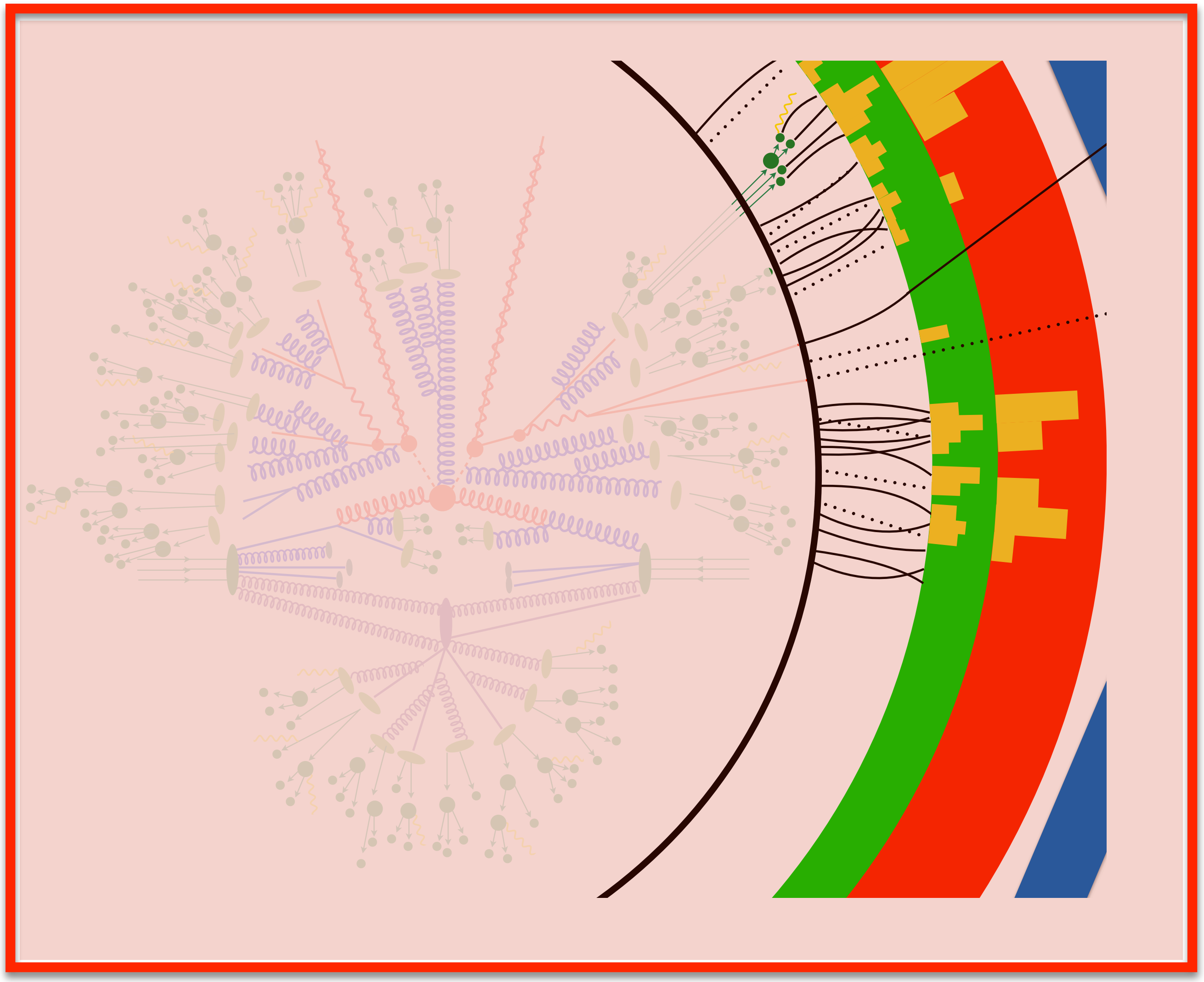
Sorry this is *pp* … I will soon make an *ep* version !

The Unfolding Challenge

Want this

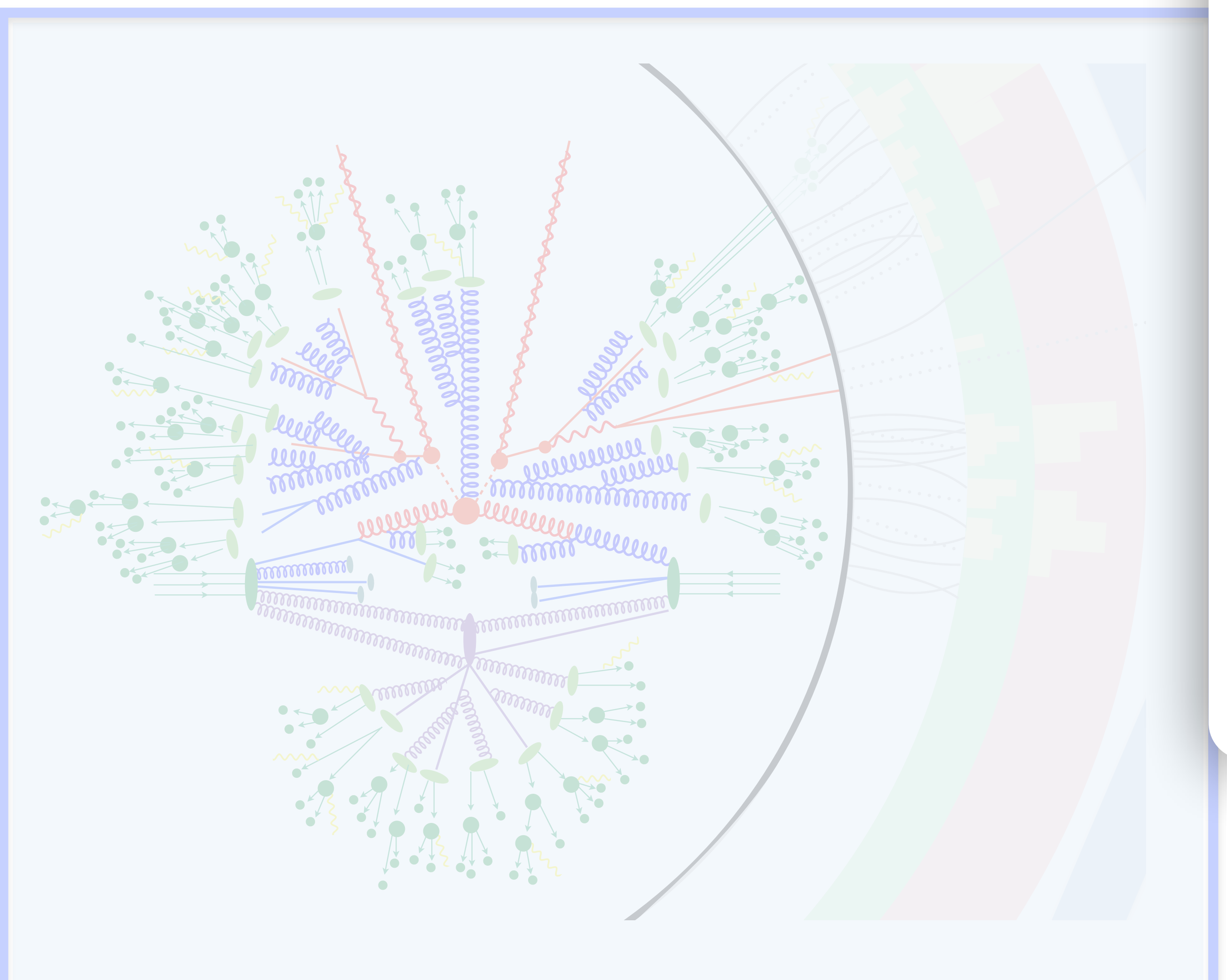


Measure this

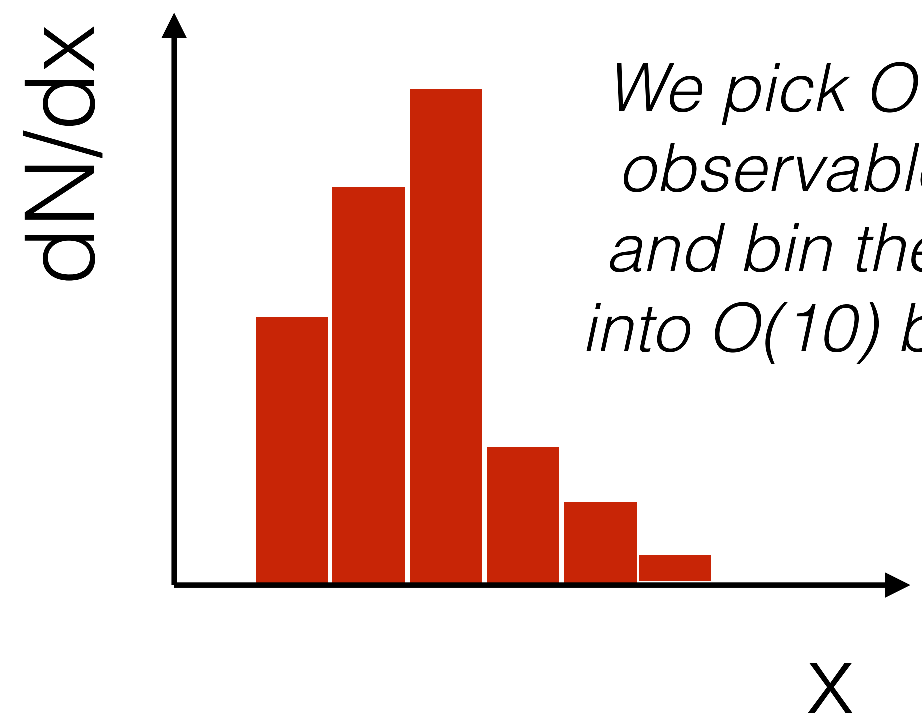


The Unfolding Challenge

Want this



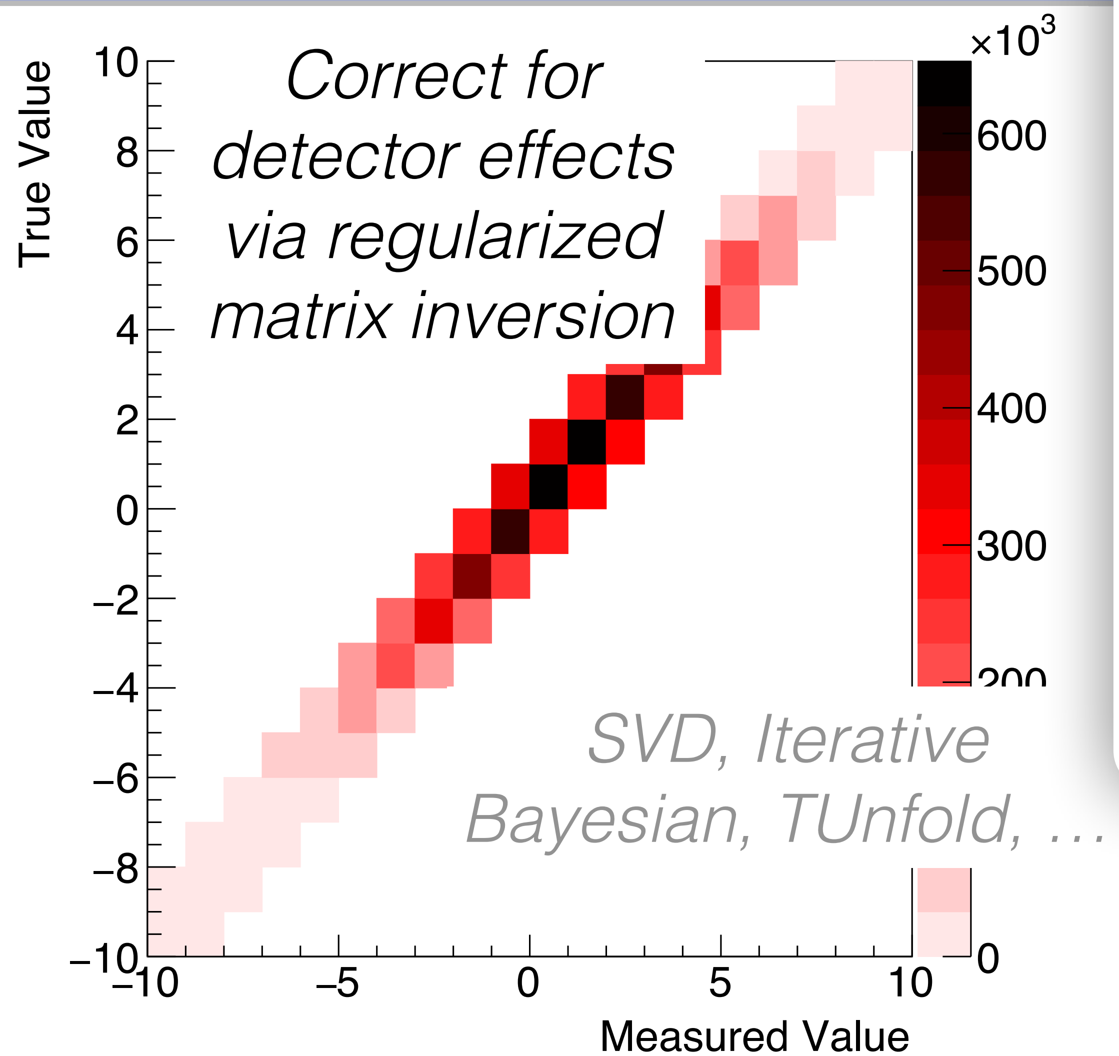
Usual solution:



The Unfolding Challenge

Want this

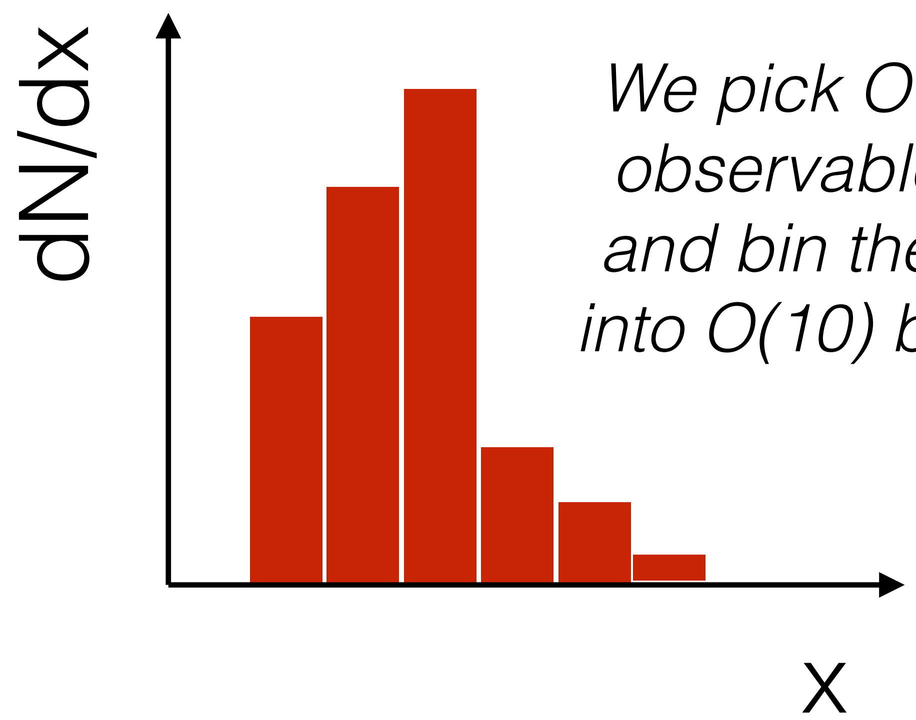
Measure this



Correct for detector effects via regularized matrix inversion

SVD, Iterative Bayesian, TUnfold, …

Usual solution:



We pick O(1) observables and bin them into O(10) bins

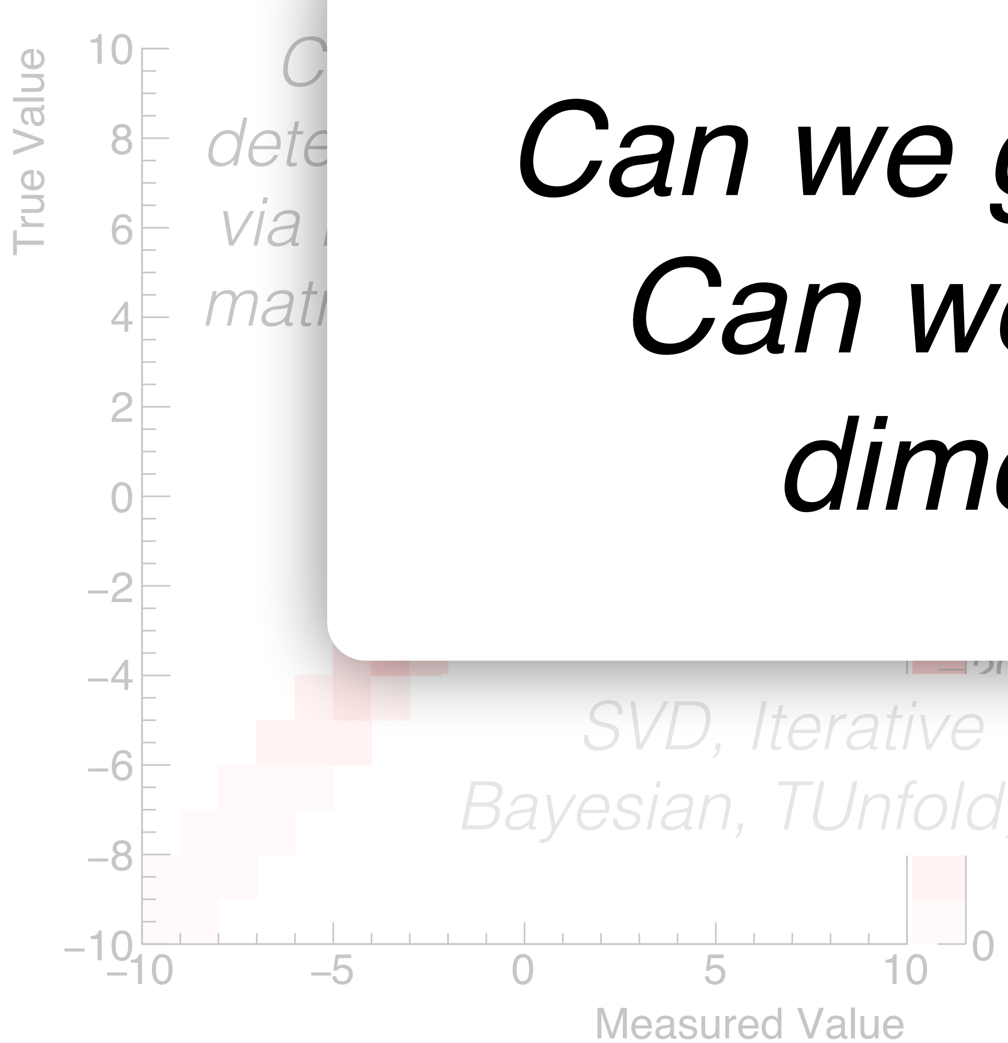
The Unfolding Challenge

Want this

Usual solution:

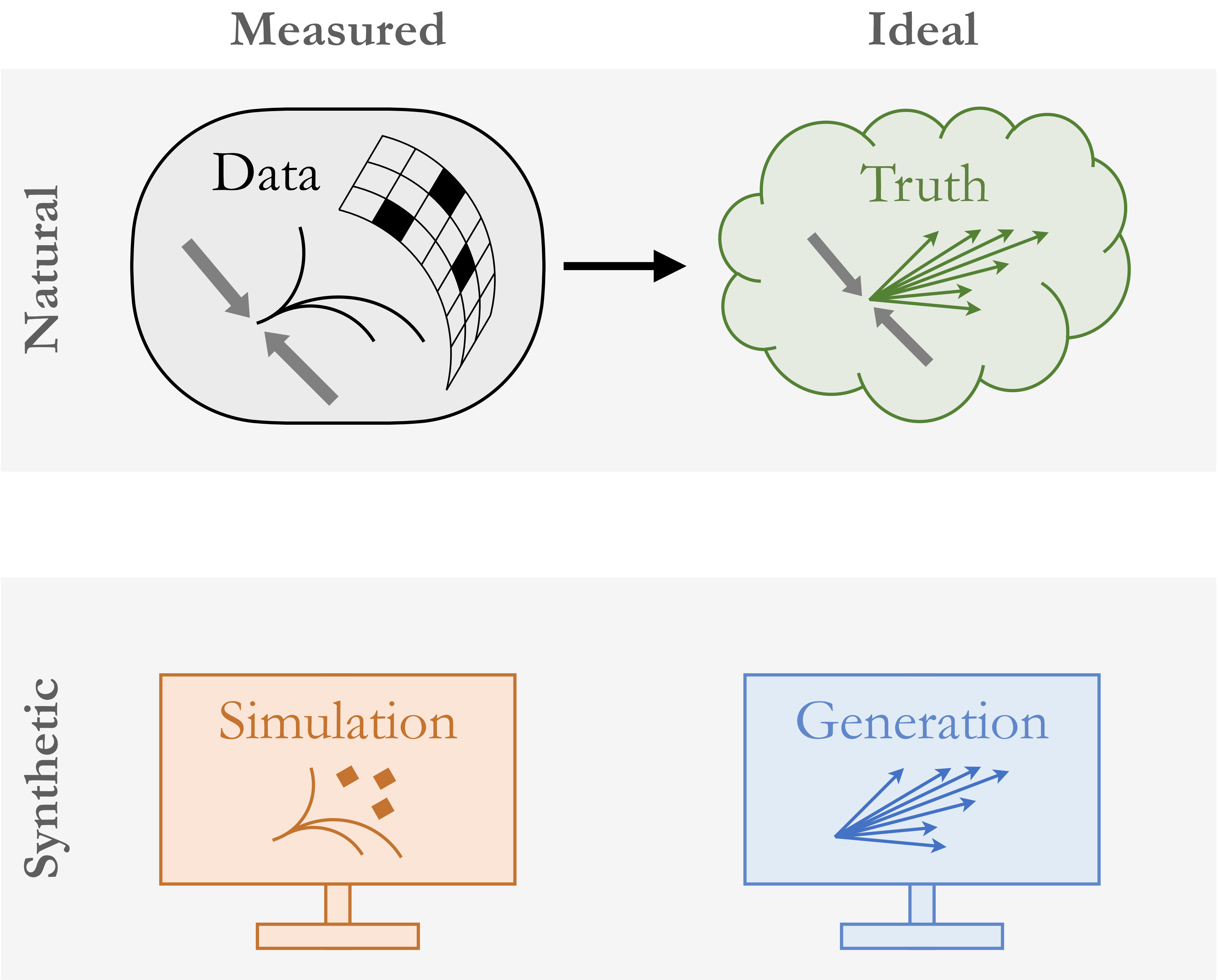
...pick O(1) ...ervables ...bin them ...O(10) bins

Can we go unbinned?
Can we use many dimensions?



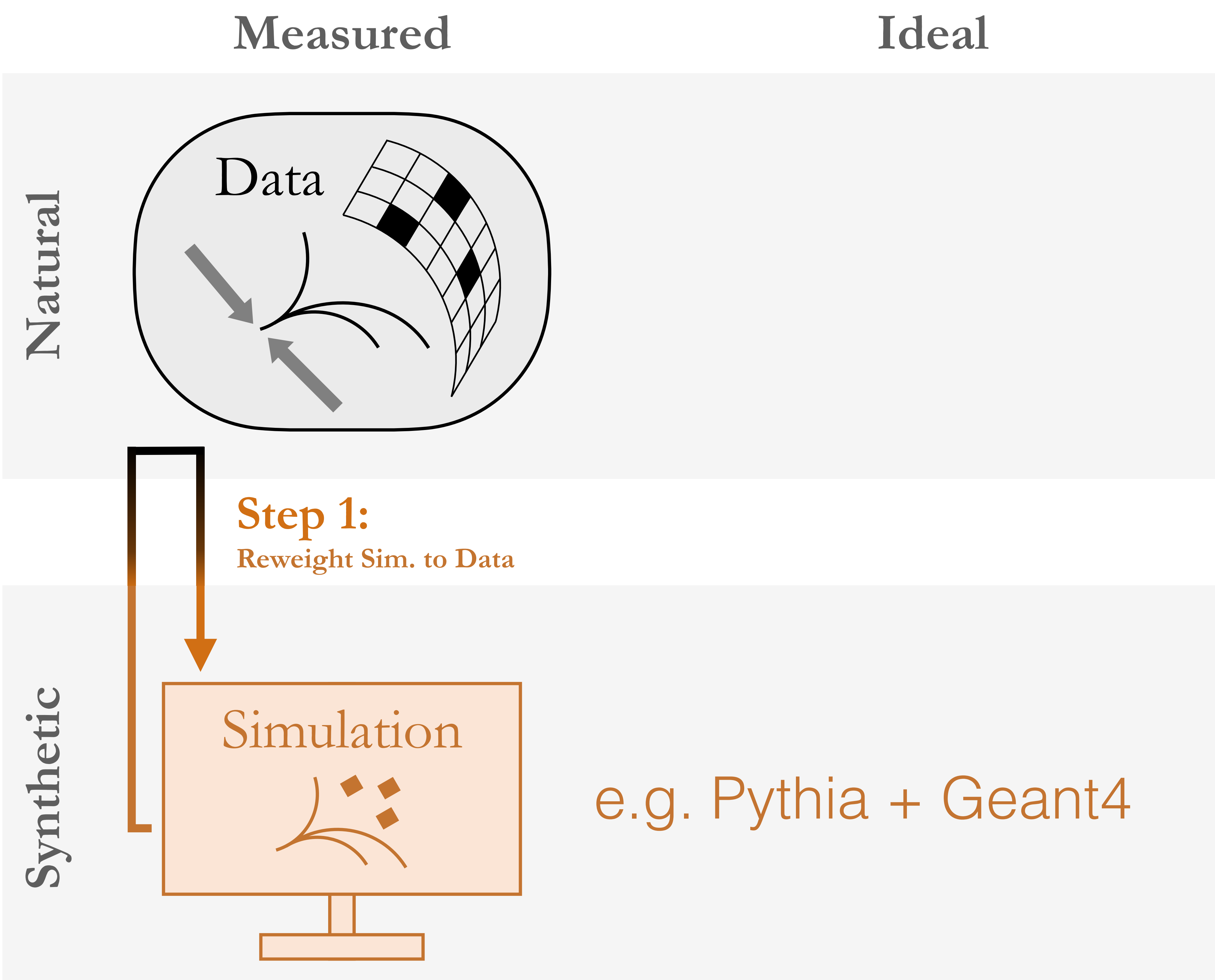
SVD, Iterative Bayesian, TUnfold, …

Unfold by iterating: OmniFold



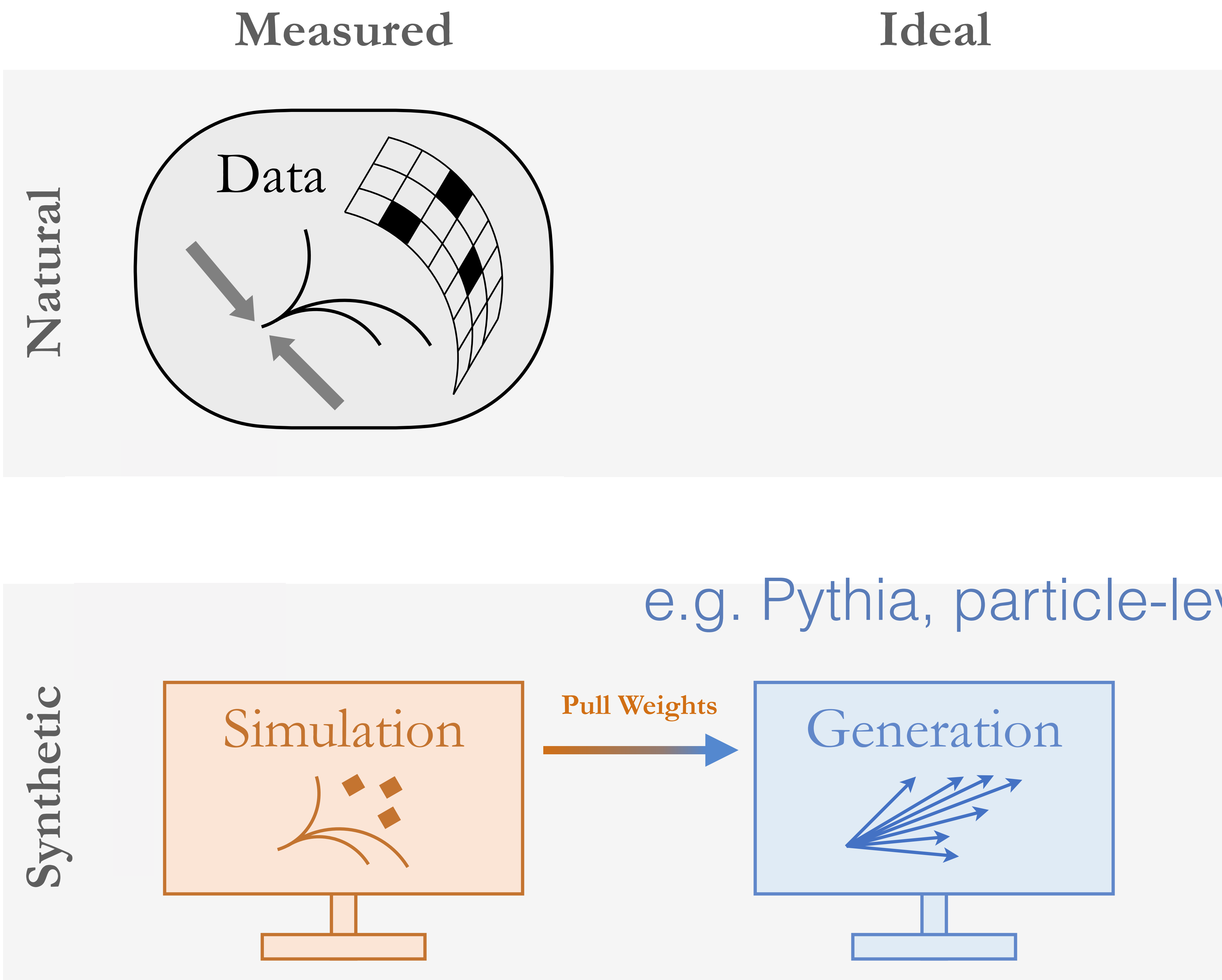
A. Andreassen, P. Komiske, E. Metodiev, BPN, J. Thaler, PRL 124 (2020) 182001

Unfold by iterating: OmniFold



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

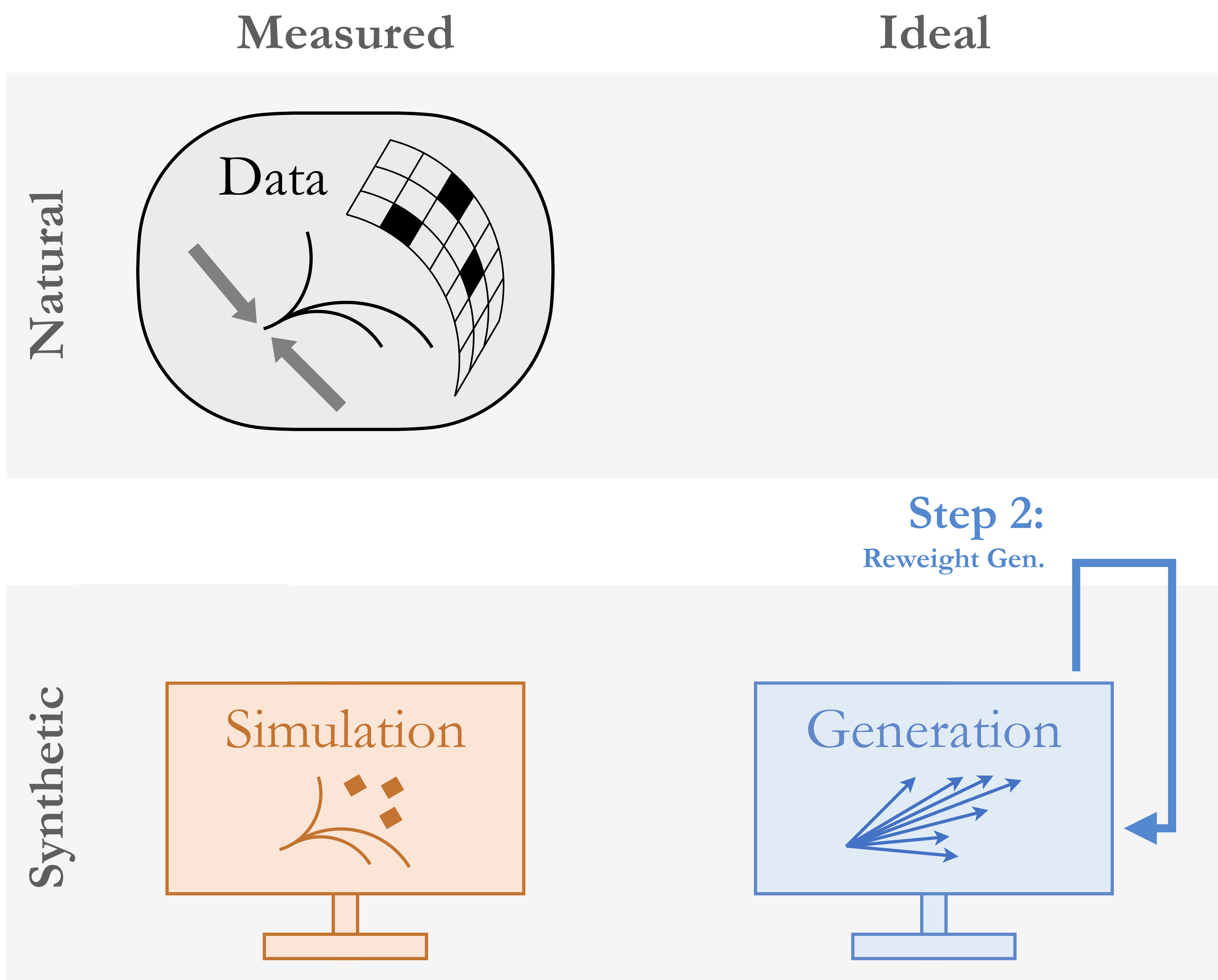
Unfold by iterating: OmniFold



e.g. Pythia, particle-level

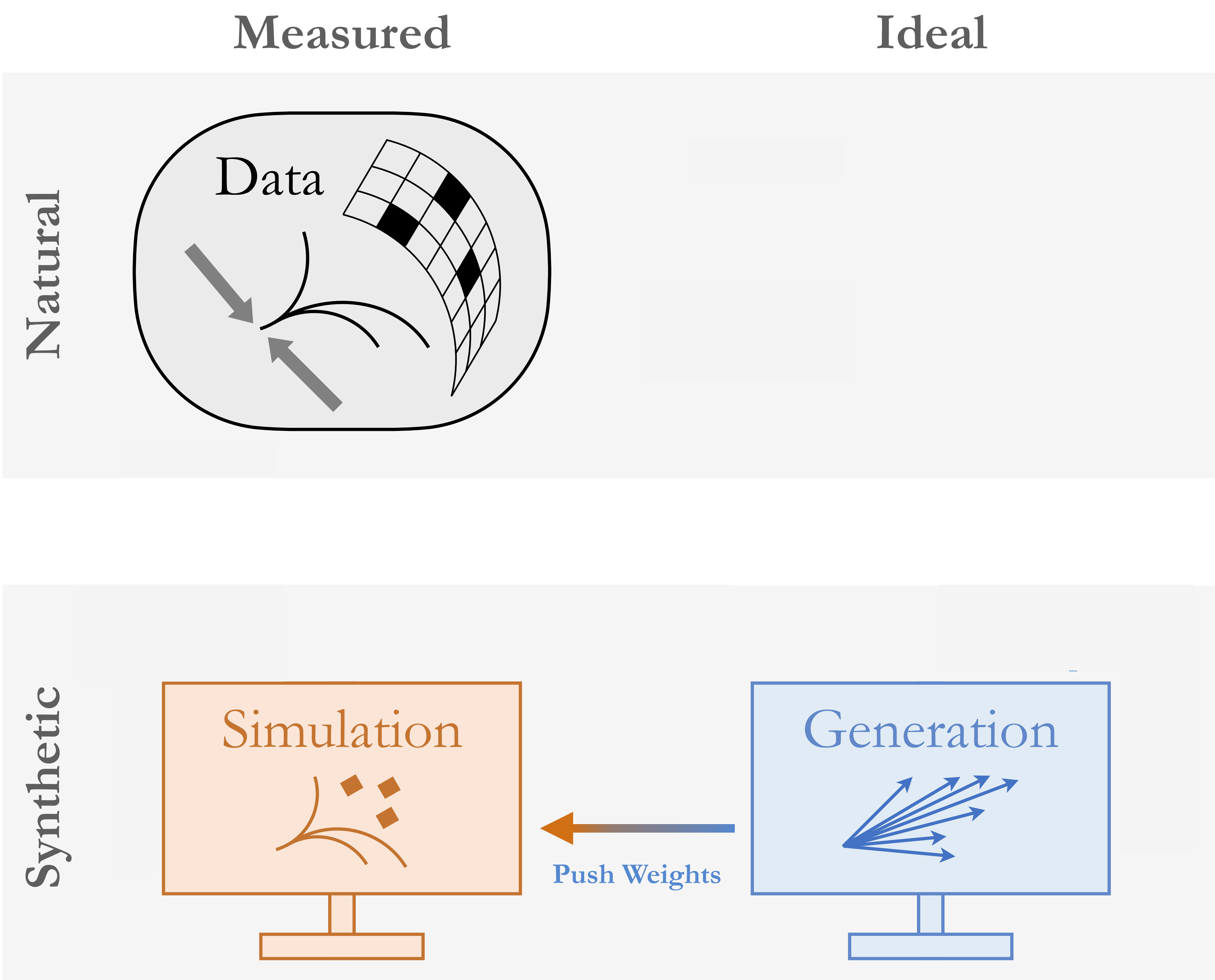
A. Andreassen, P. Komiske, E. Metodiev,
BPN, J. Thaler, PRL 124 (2020) 182001

Unfold by iterating: OmniFold



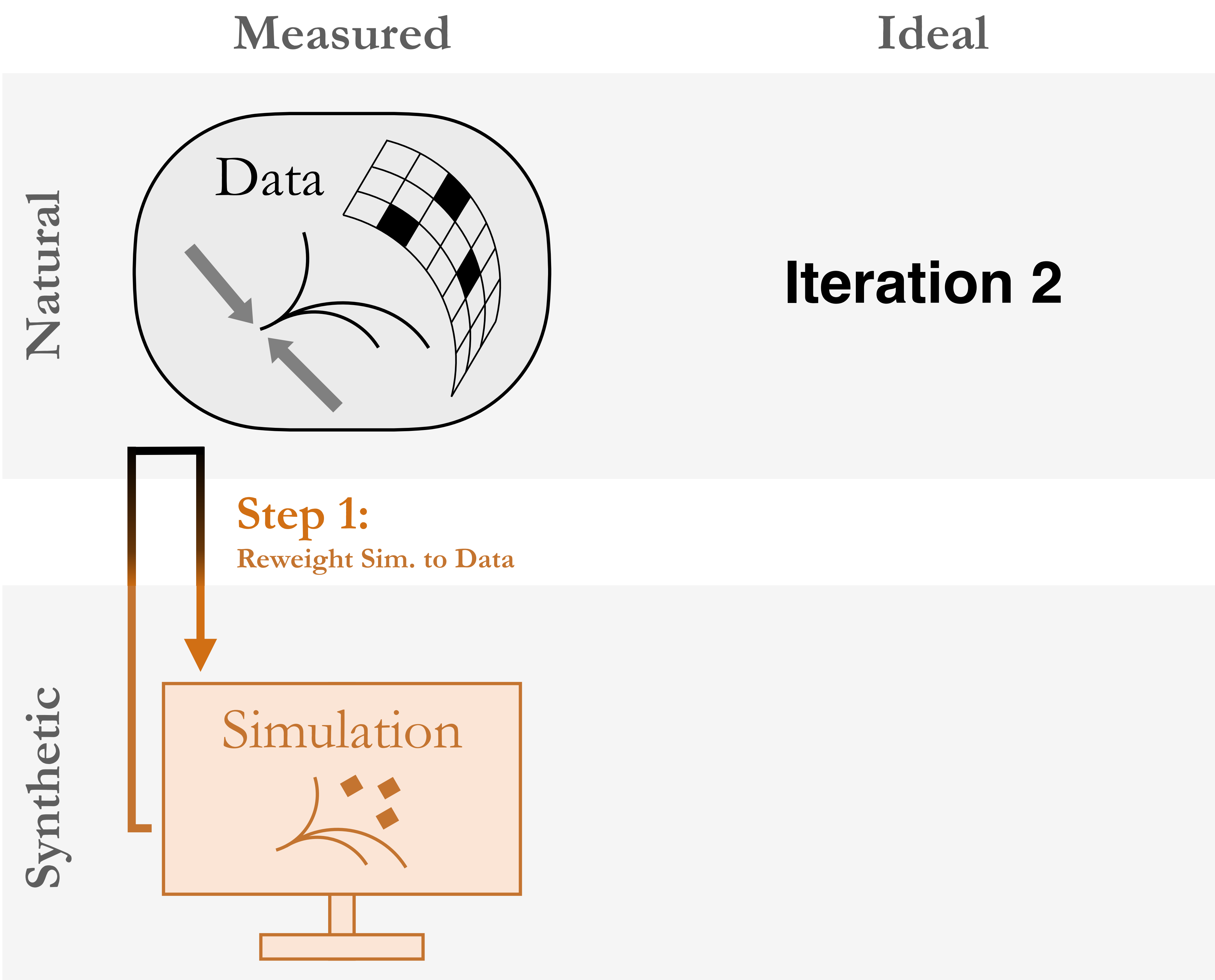
A. Andreassen, P. Komiske, E. Metodiev, BPN, J. Thaler, PRL 124 (2020) 182001

Unfold by iterating: OmniFold



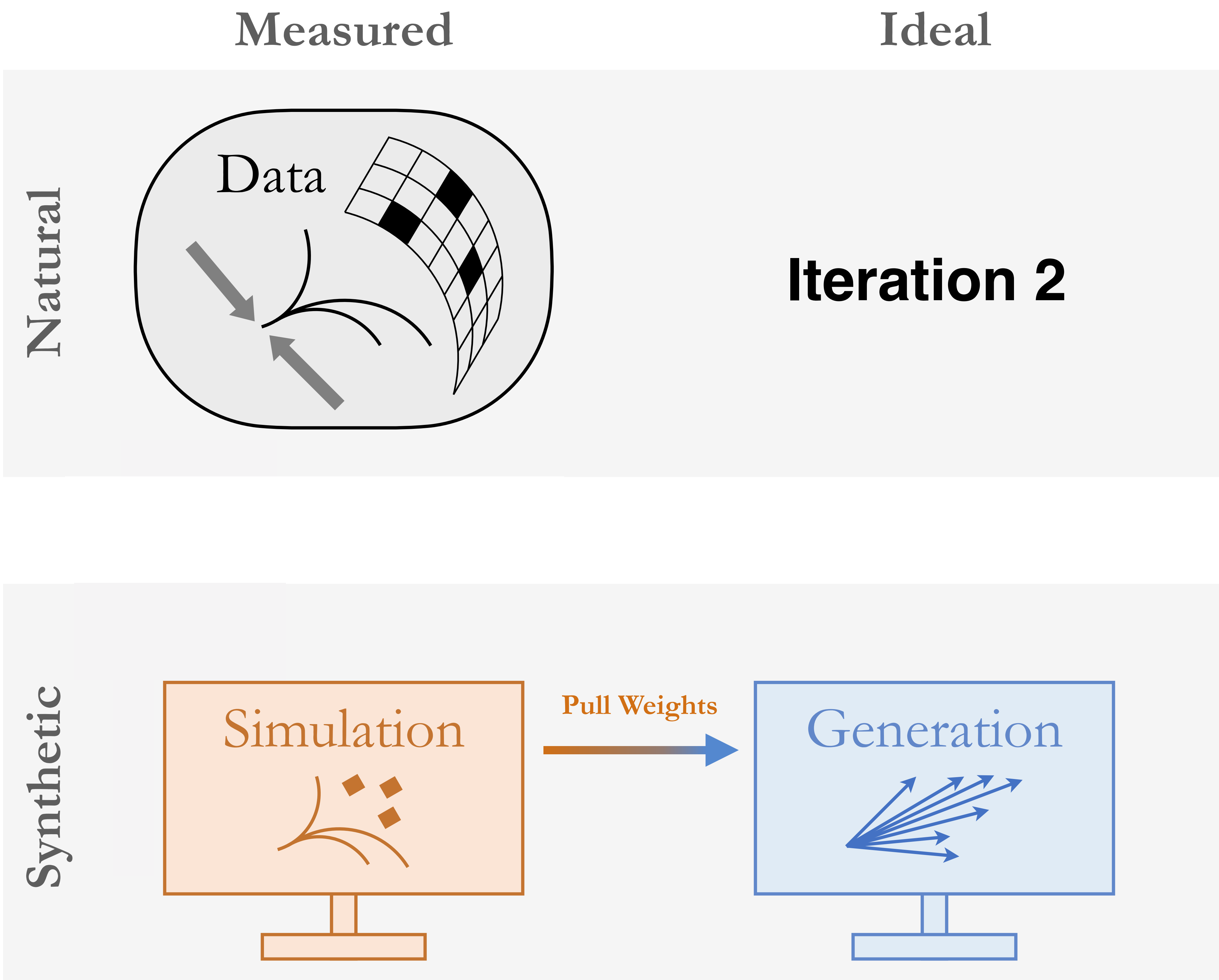
A. Andreassen, P. Komiske, E. Metodiev, BPN, J. Thaler, PRL 124 (2020) 182001

Unfold by iterating: OmniFold



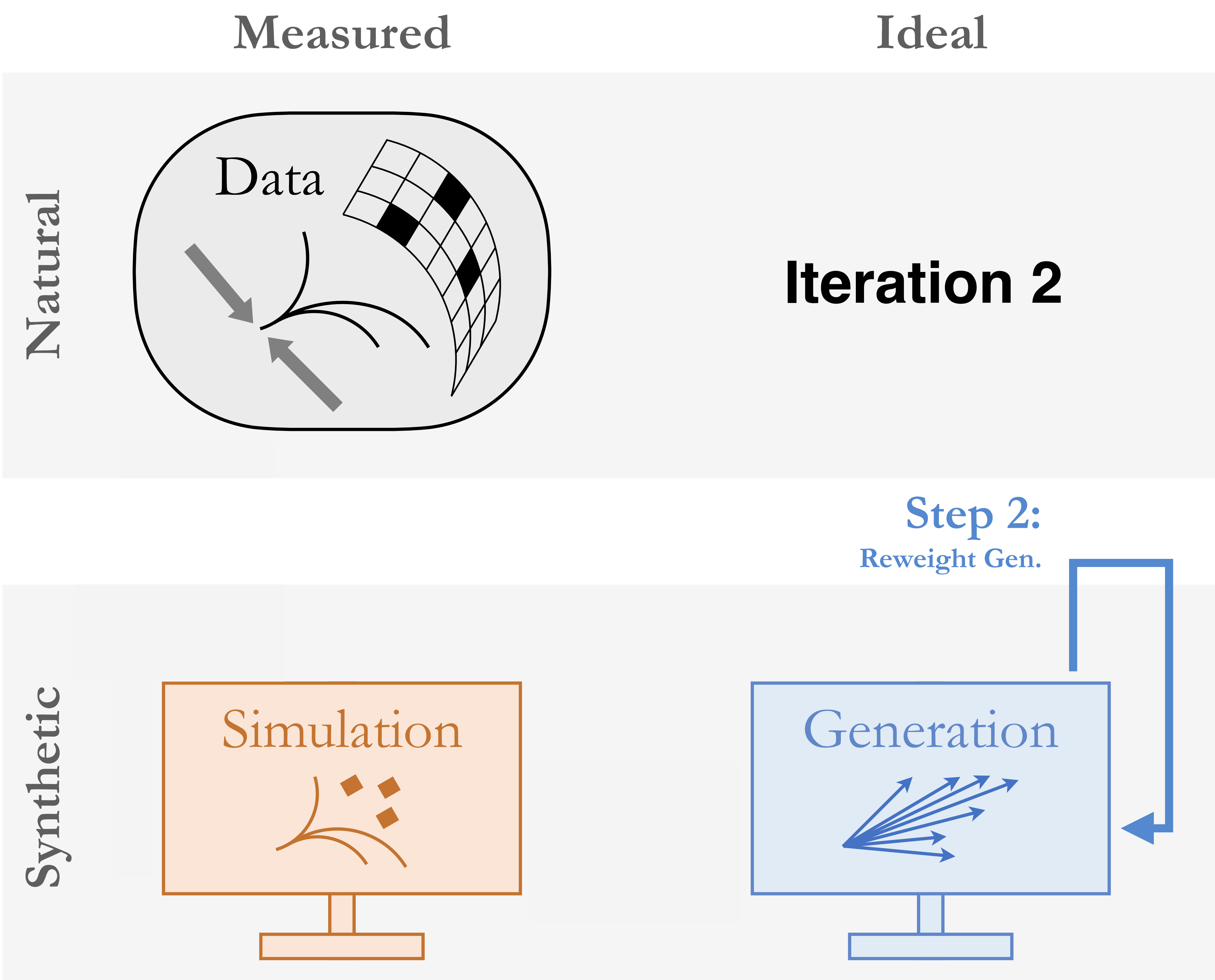
A. Andreassen, P. Komiske, E. Metodiev, BPN, J. Thaler, PRL 124 (2020) 182001

Unfold by iterating: OmniFold



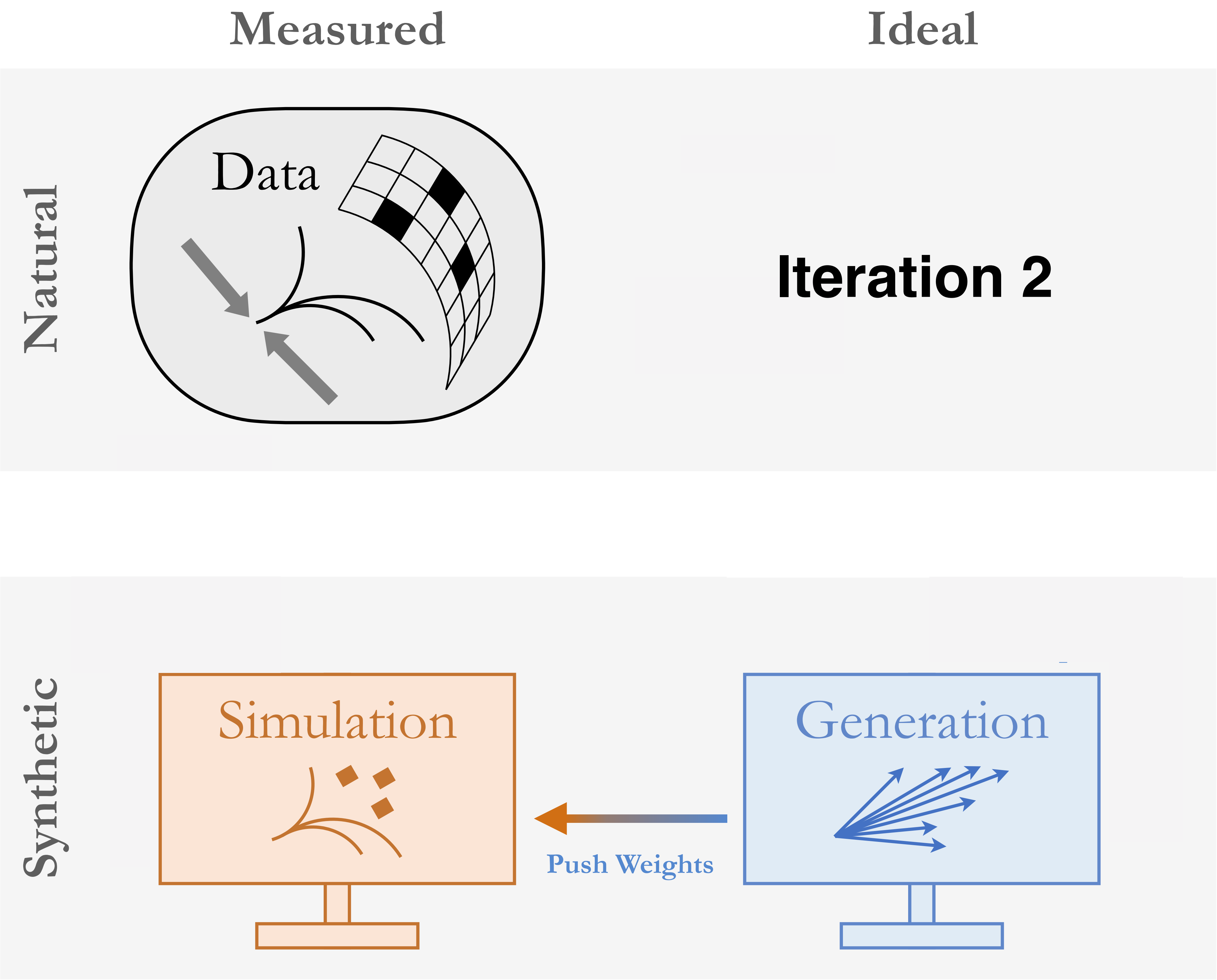
A. Andreassen, P. Komiske, E. Metodiev, BPN, J. Thaler, PRL 124 (2020) 182001

Unfold by iterating: OmniFold



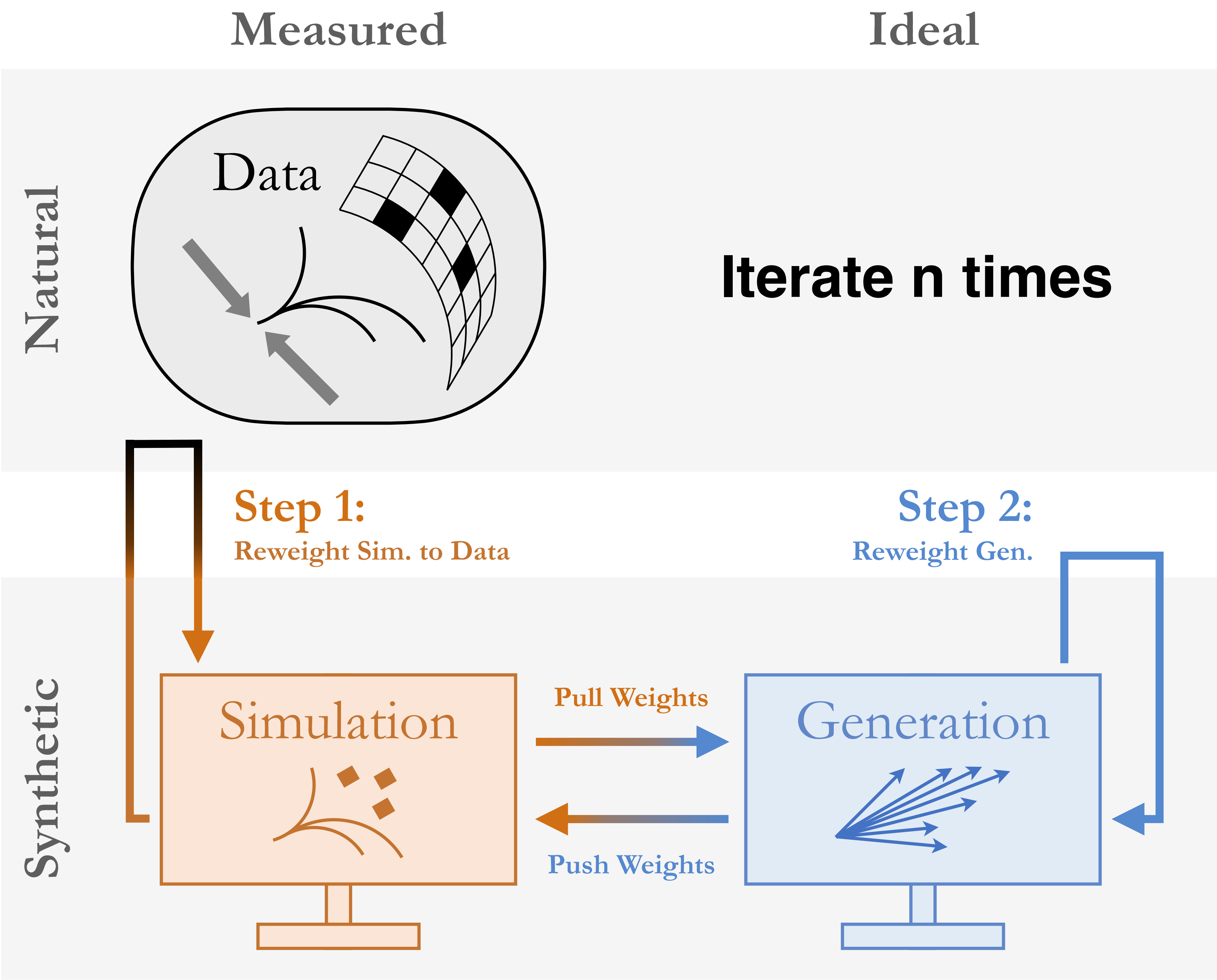
A. Andreassen, P. Komiske, E. Metodiev, BPN, J. Thaler, PRL 124 (2020) 182001

Unfold by iterating: OmniFold



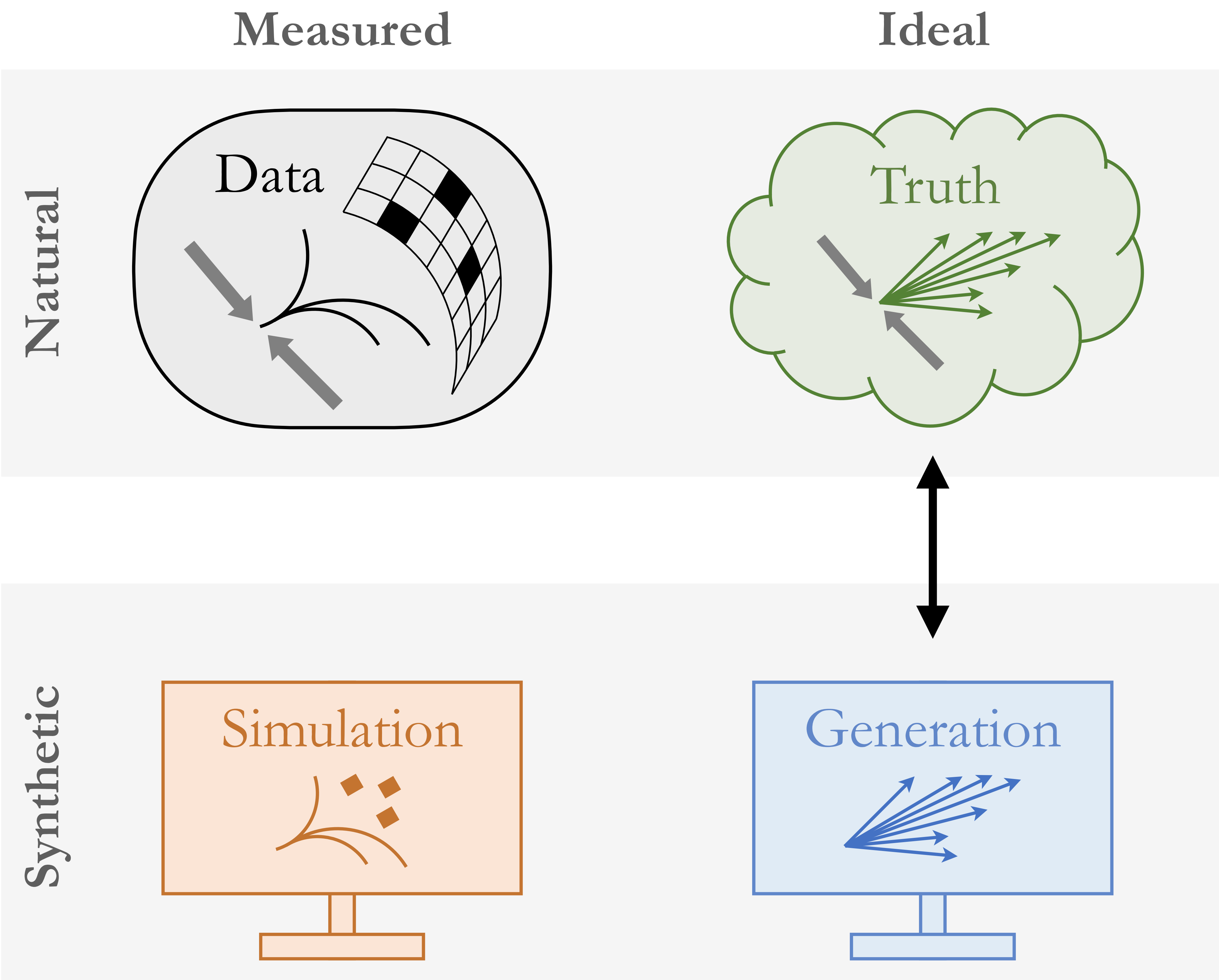
A. Andreassen, P. Komiske, E. Metodiev, BPN, J. Thaler, PRL 124 (2020) 182001

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

Reweighting

How do to the reweighting?

Reweighting

How do to the 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***?

(and don't want to estimate them by binning)

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**)

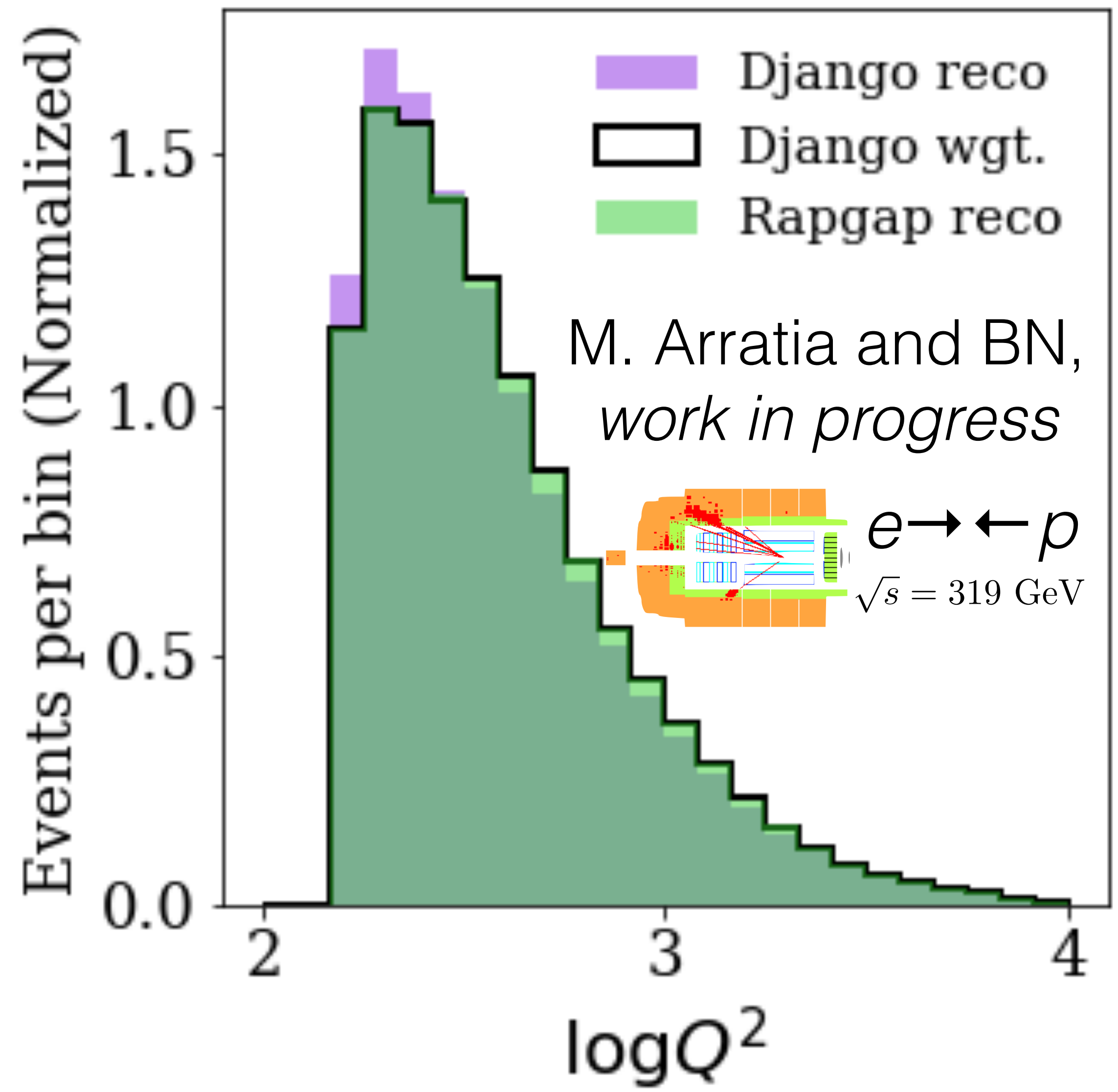
This is a well-known fact, but you can read about how this can be used for many tasks in HEP in this paper called DCTR: A. Andreassen, BPN, PRD RC 101 (2020) 091901, 1907.08209

Classification for reweighting

Neural networks are naturally unbinned and readily process high-dimensional data.

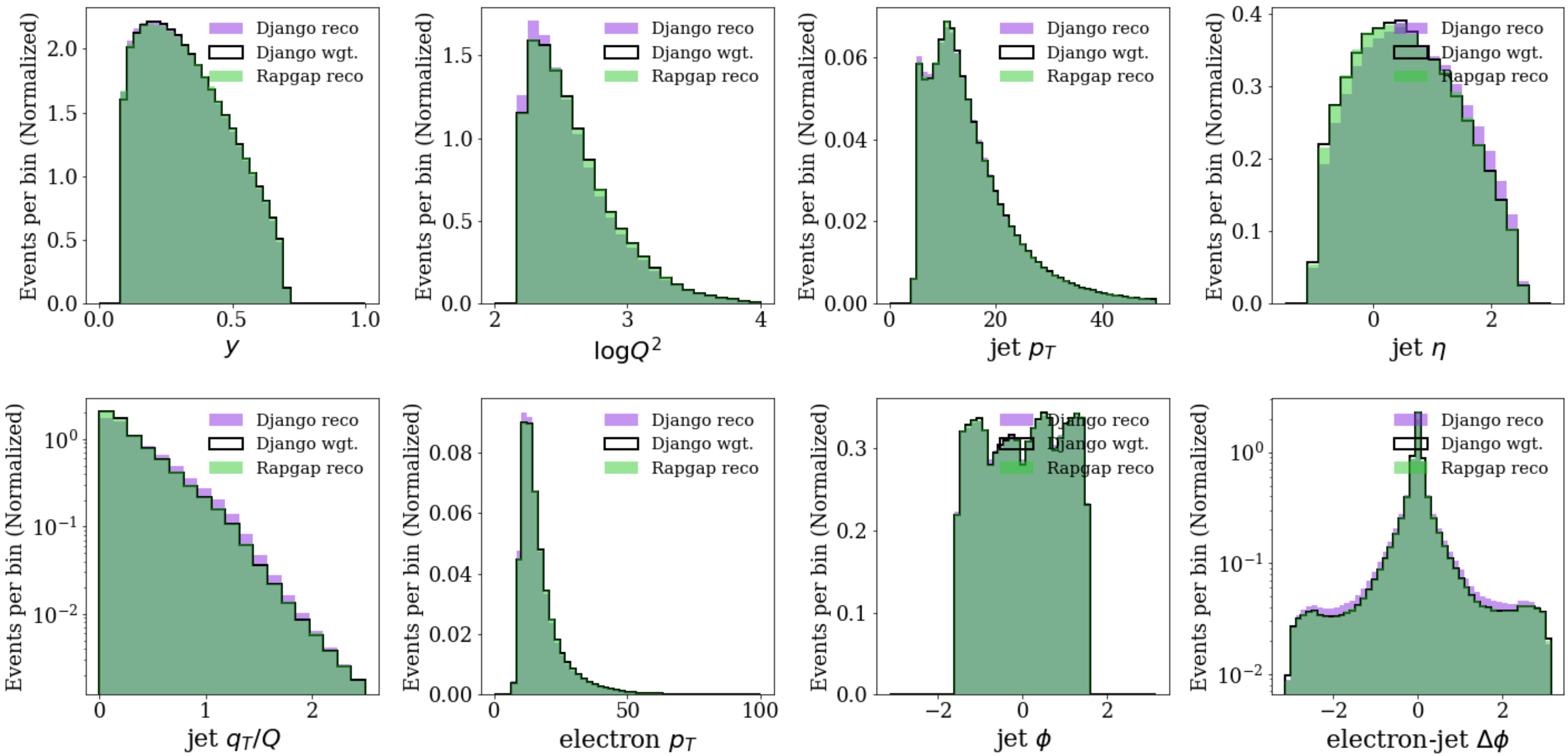
For this measurement, we use simple fully connected networks with a few hidden layers.

N.B. the distribution is binned for illustration, but the reweighting is unbinned.

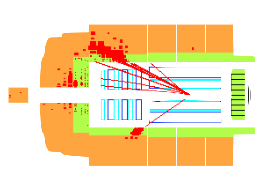


Classification for reweighting

All of these distributions are simultaneously reweighted!



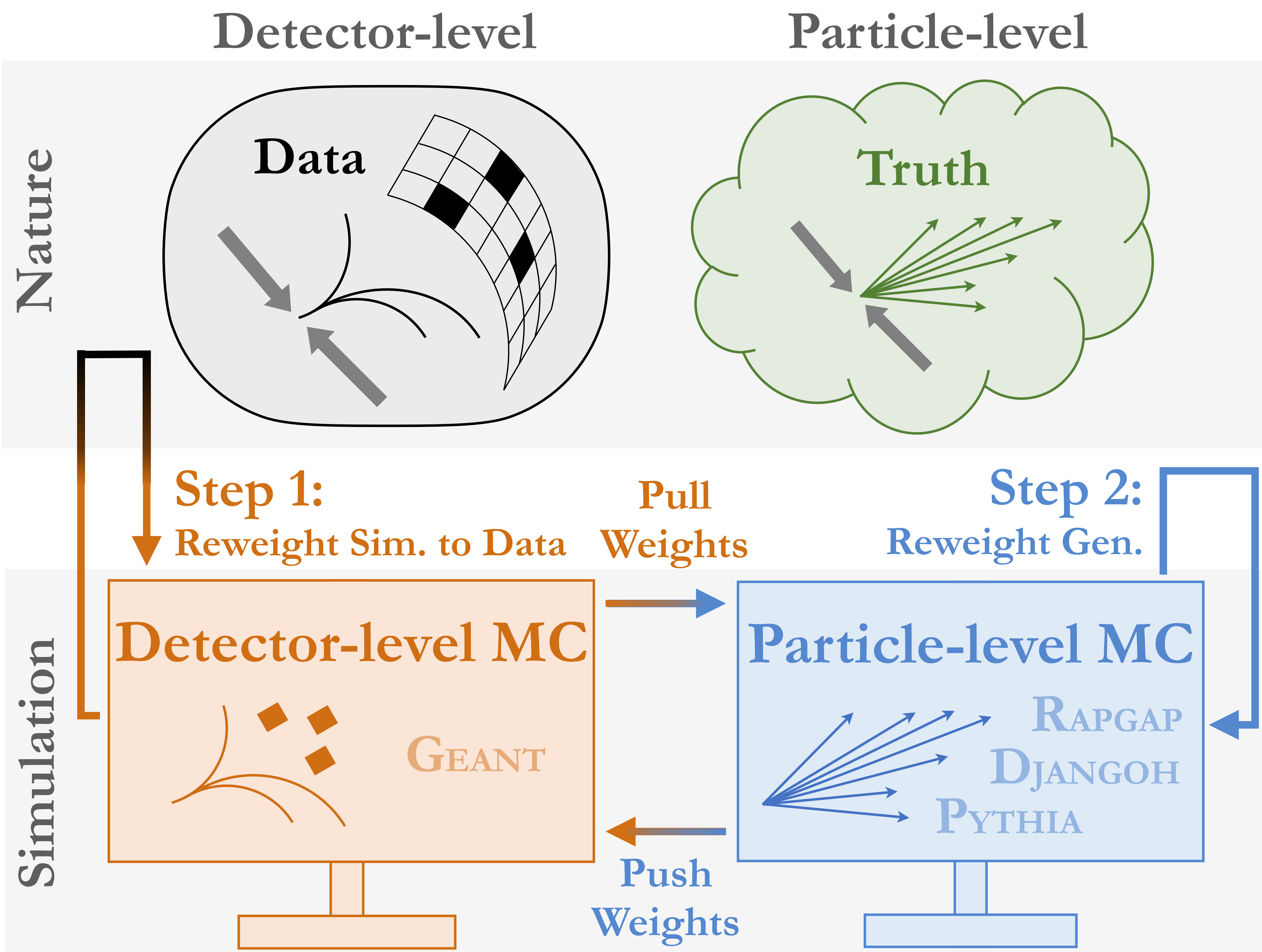
M. Arratia and BN, *work in progress*



$e\rightarrow\leftarrow p$

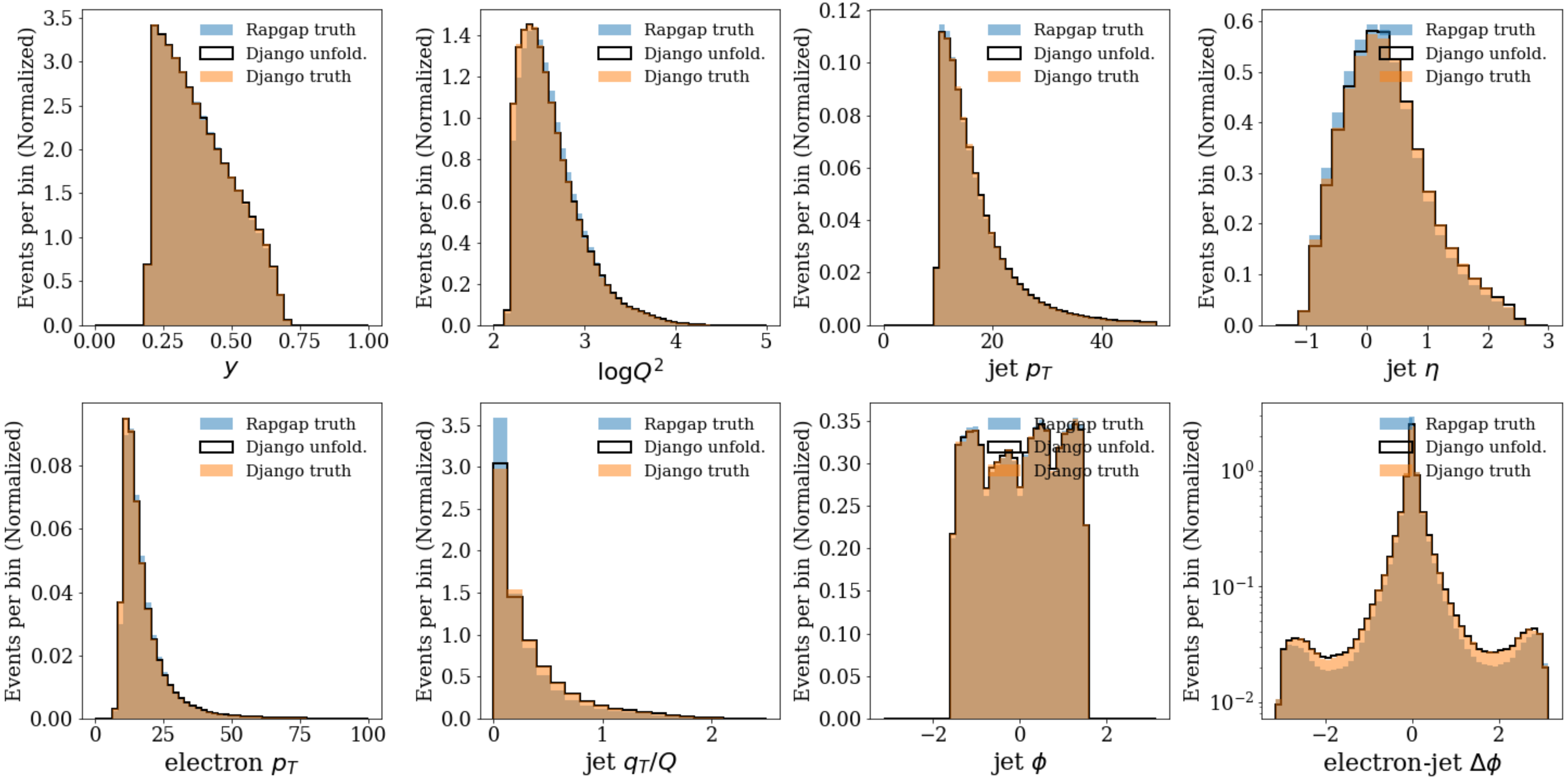
$\sqrt{s} = 319$ GeV

Unfold by iterating: OmniFold

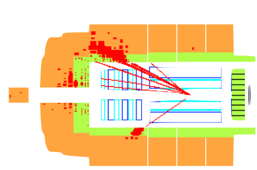


OmniFolding *ep* simulations

We see excellent closure for the full phase space!



M. Arratia and BN,
work in progress

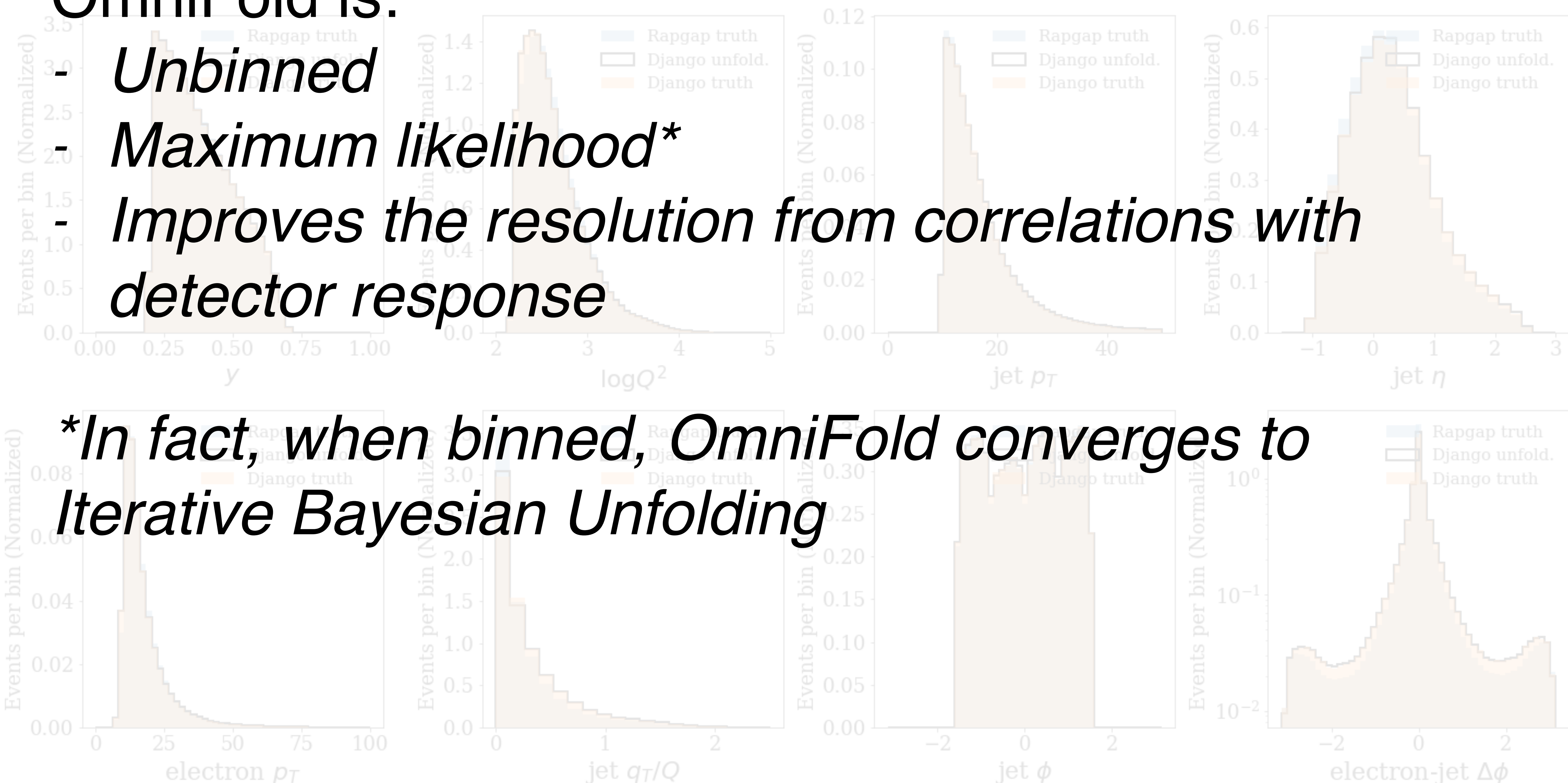


$e\rightarrow \leftarrow p$

$\sqrt{s} = 319$ GeV

We see excellent closure for the full phase space!

OmniFold is:
- *Unbinned*
- *Maximum likelihood**
- *Improves the resolution from correlations with detector response*



**In fact, when binned, OmniFold converges to Iterative Bayesian Unfolding*

→Physics details in Miguel's talk

*In fact, OmniFold can also work on low-level inputs (e.g. energy flow particles). In that case, you can construct observables **after** the measurement.*

See an example in e^+e^- and *pp* in PRL 124 (2020) 182001

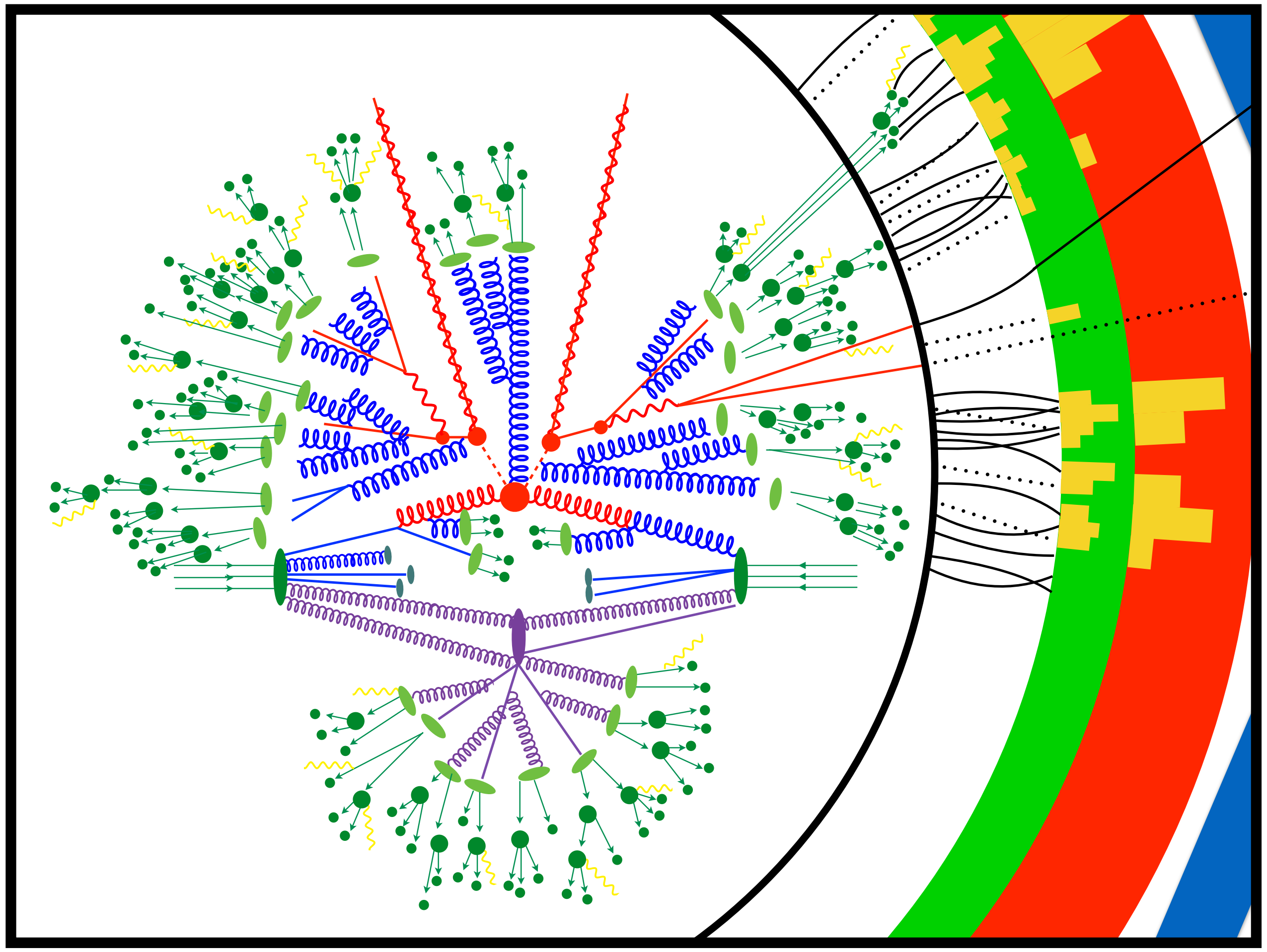
*This and the other properties of OmniFold will significantly increase the utility and **future-proofing of EIC data**.*

We need to **start thinking now** about the **computing** resources needed for this approach and also how to represent and store unbinned measurements.

e.g. good GPUs have O(10) GB of memory; what if you can't load the data into memory? Need data parallel learning.

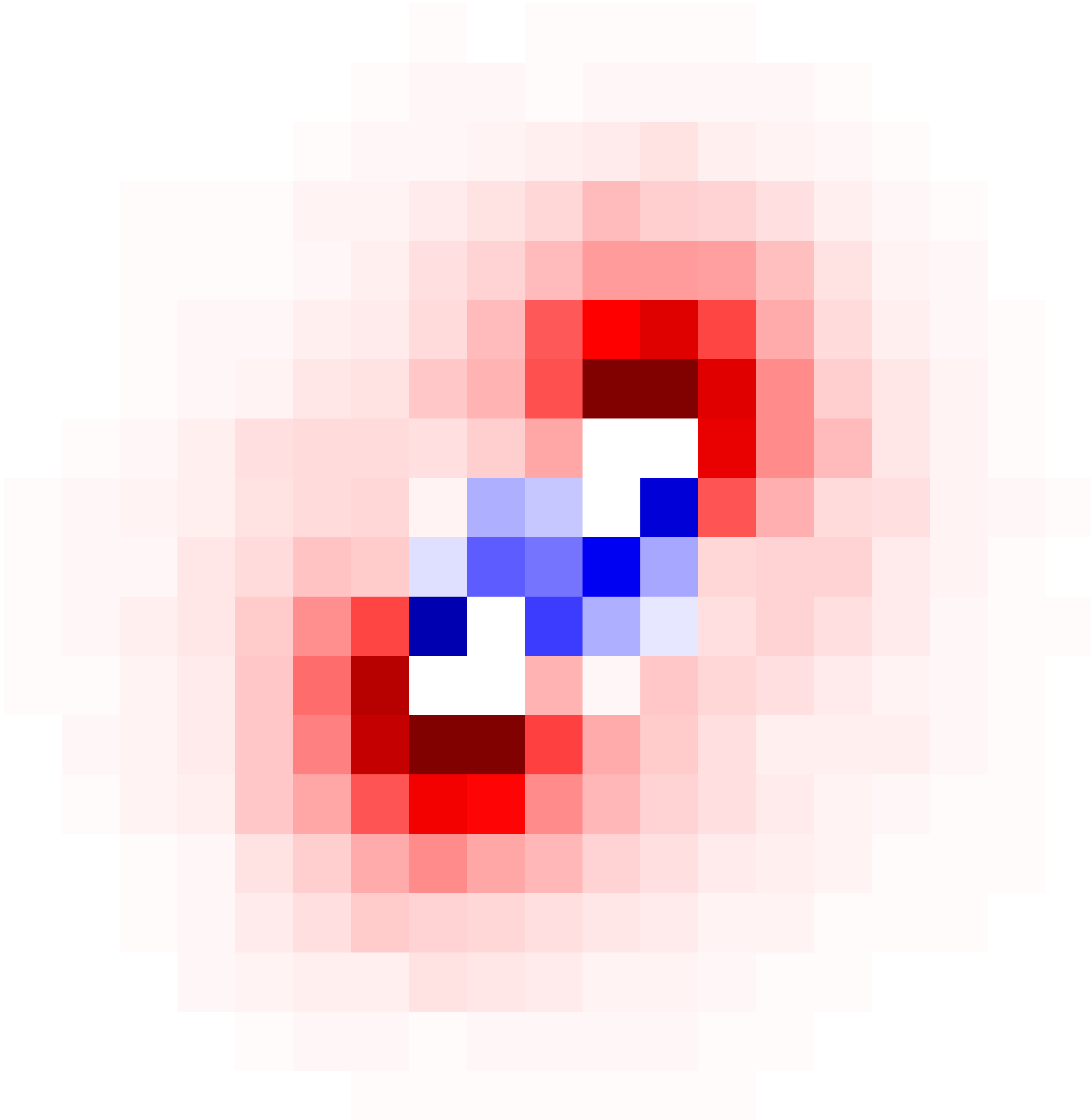
Conclusions and outlook

By using deep learning and the new OmniFold method, we are able to simultaneously unfold multiple unbinned spectra.



This will provide a wealth of information for probing QCD (including TMD, etc.) in great detail.

This methodology is general & can be applied to all measurements, in both low and and high dimensions!



Fin.