H1 lepton-jet correlations and ML unfolding

Benjamin Nachman
Lawrence Berkeley National Laboratory
on behalf of the **H1** Collaboration

bpnachman.com
bpnachman@lbl.gov
@bpnachman
bnachman

12th workshop on MPI at LHC

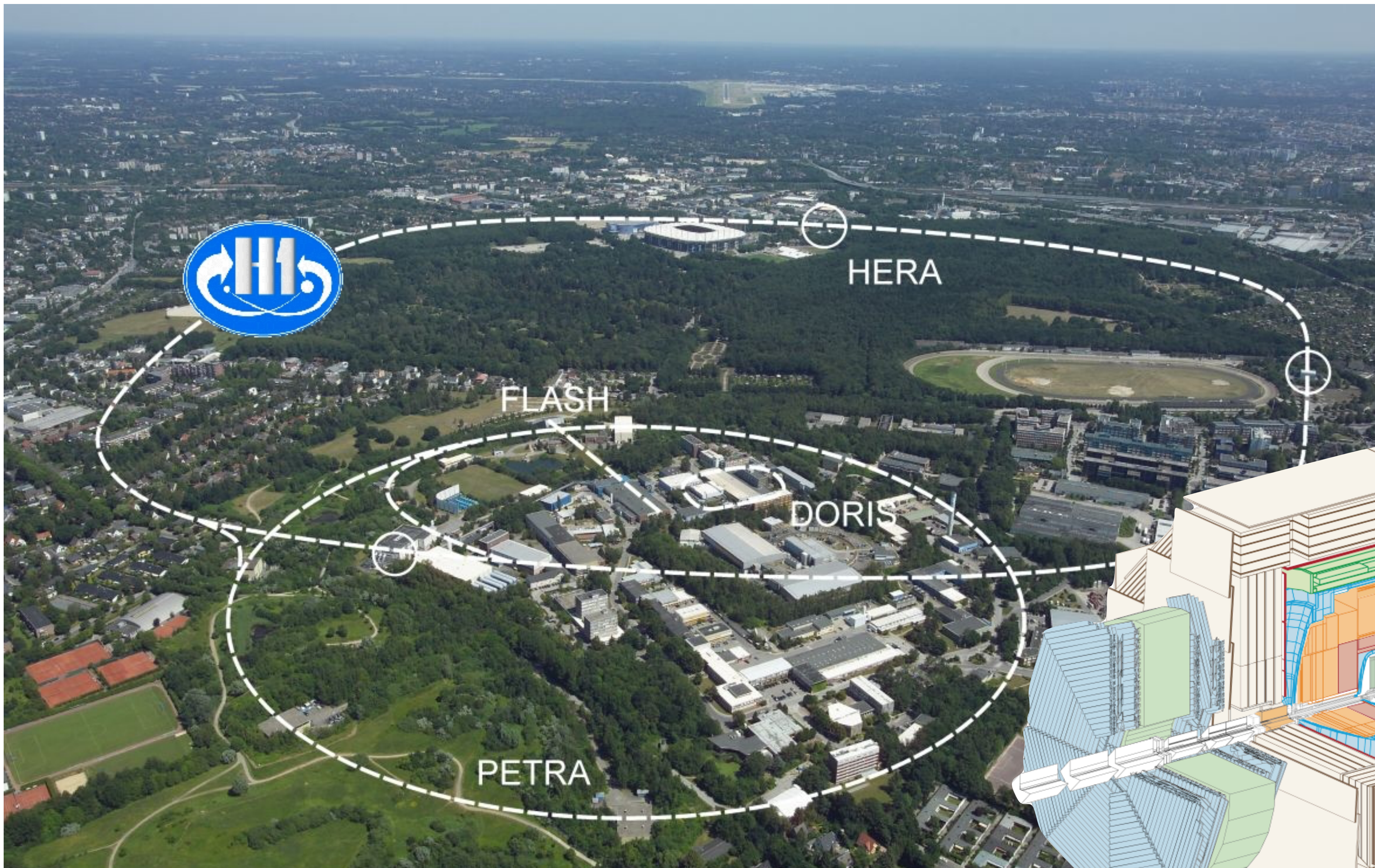
October 11, 2021

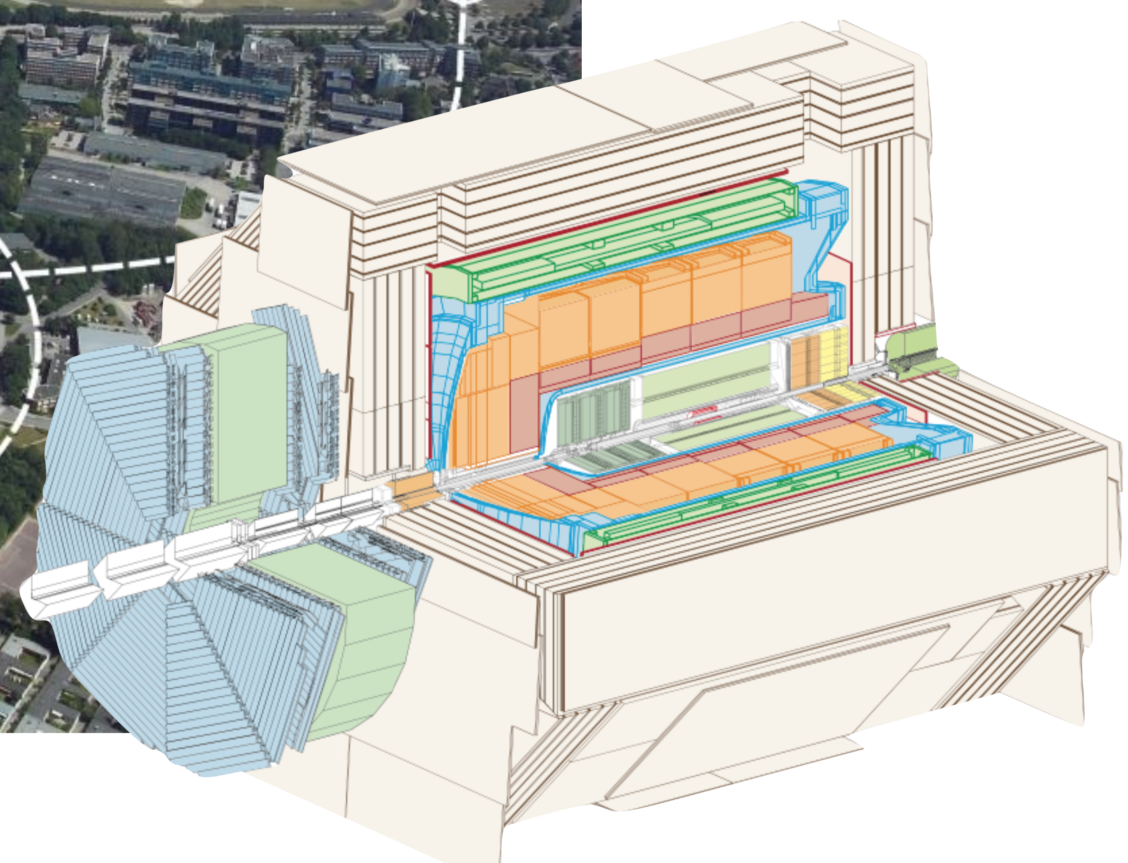
BERKELEY LAB

BIDS
BERKELEY INSTITUTE FOR DATA SCIENCE

H1 @ HERA

H1 was one of two multipurpose experiments at HERA.

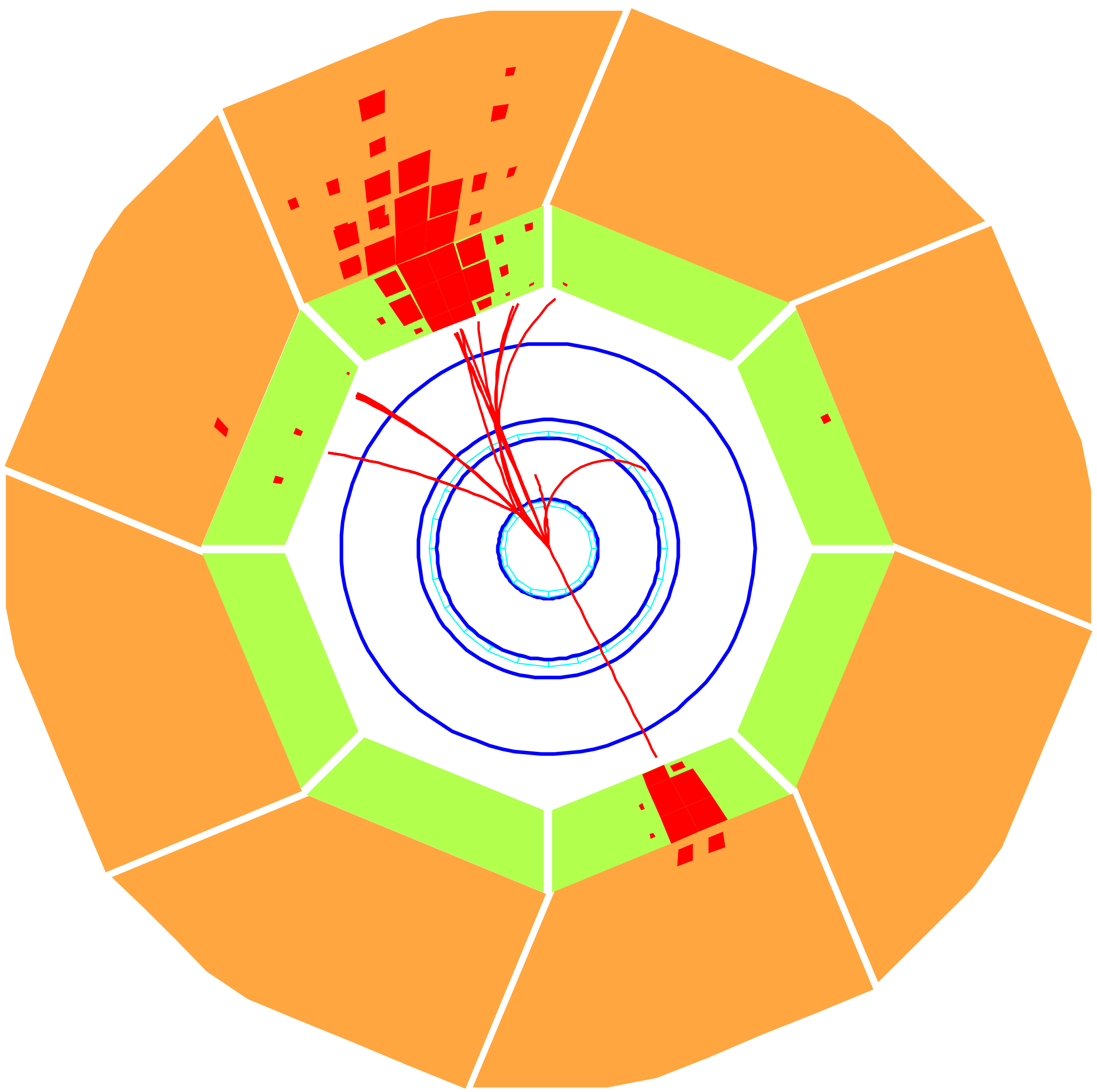




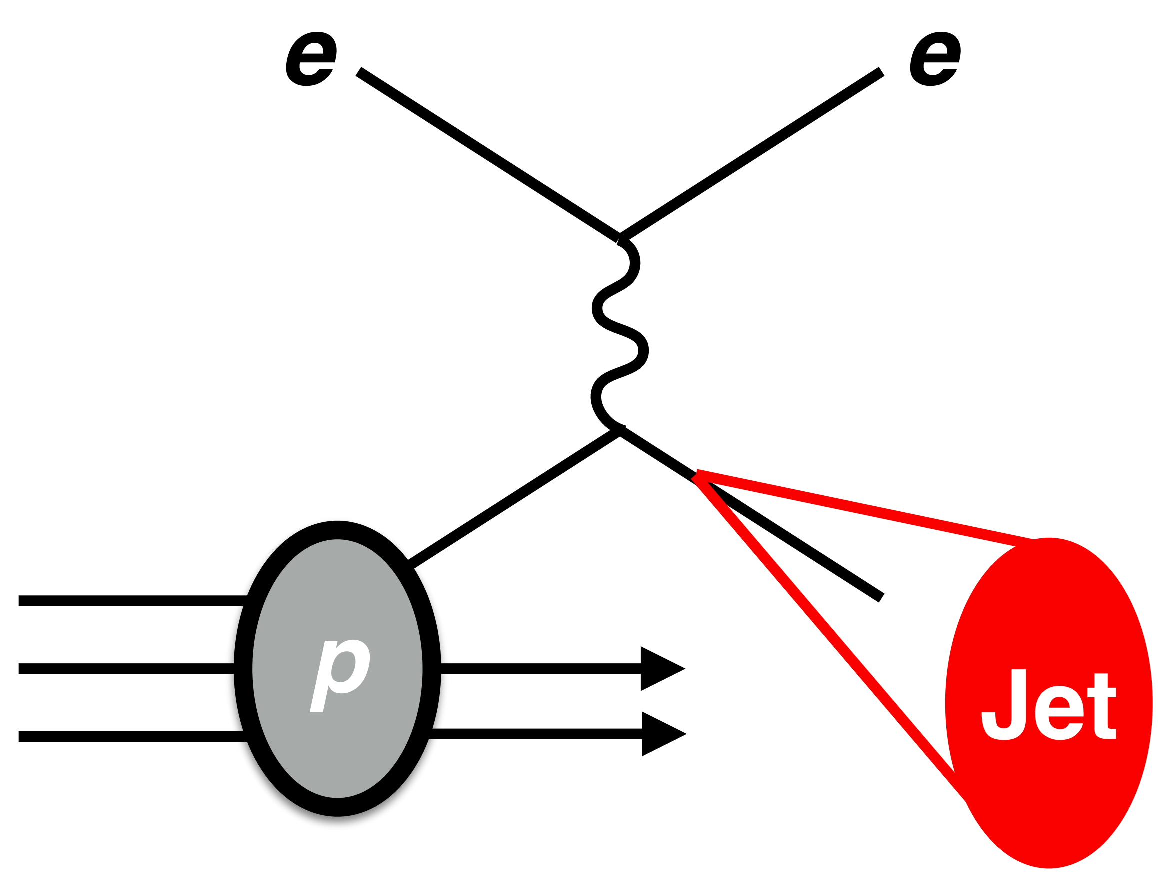
H1 @ HERA

H1 was one of two multipurpose experiments at HERA.

For this talk: 2006-2007 data, 136 pb^{-1}, 320 GeV

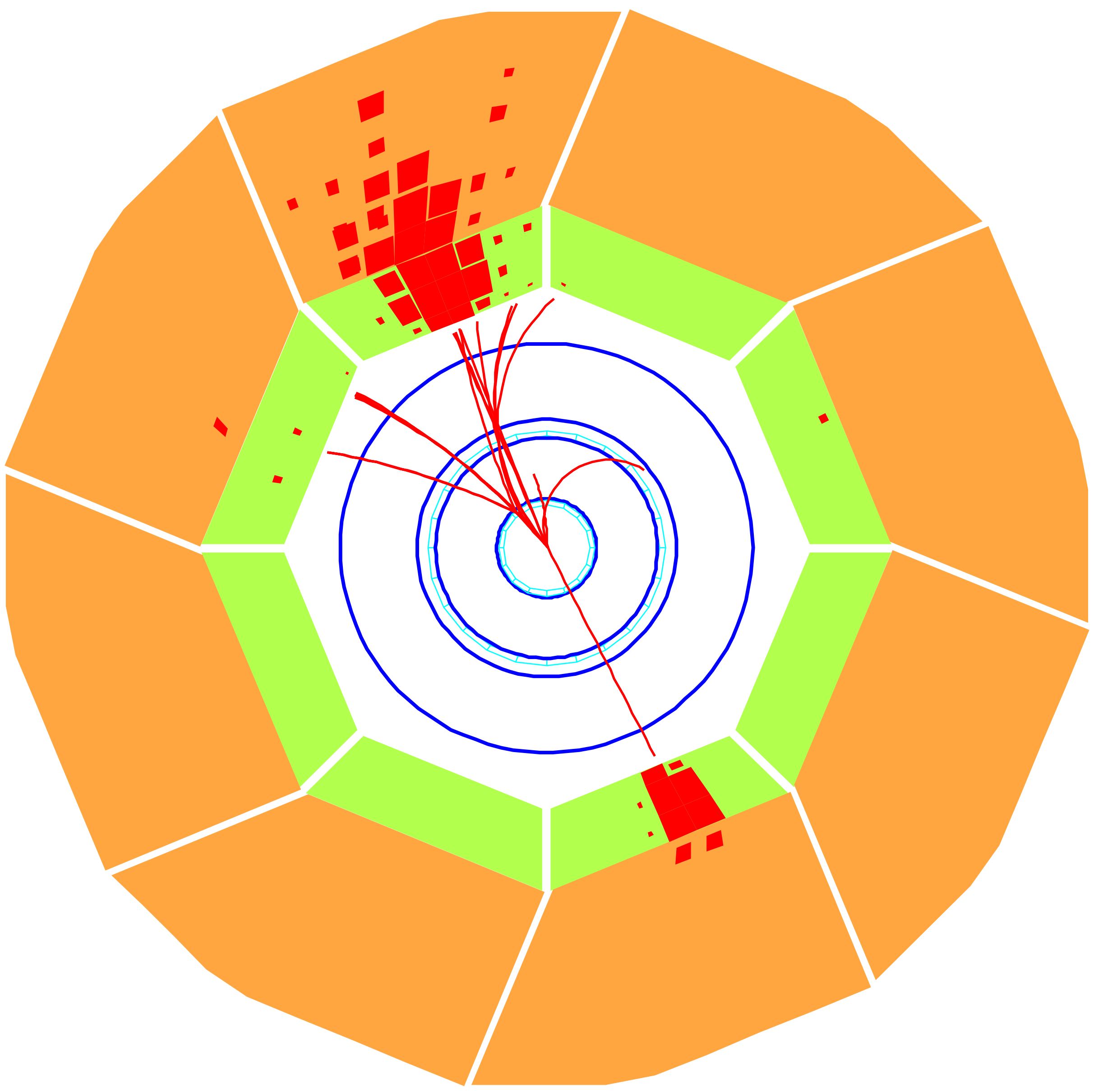


I'll present a measurement of the electron-jet inbalance



Why electron-jet imbalance?

Born-level configuration, electron and jet are back-to-back

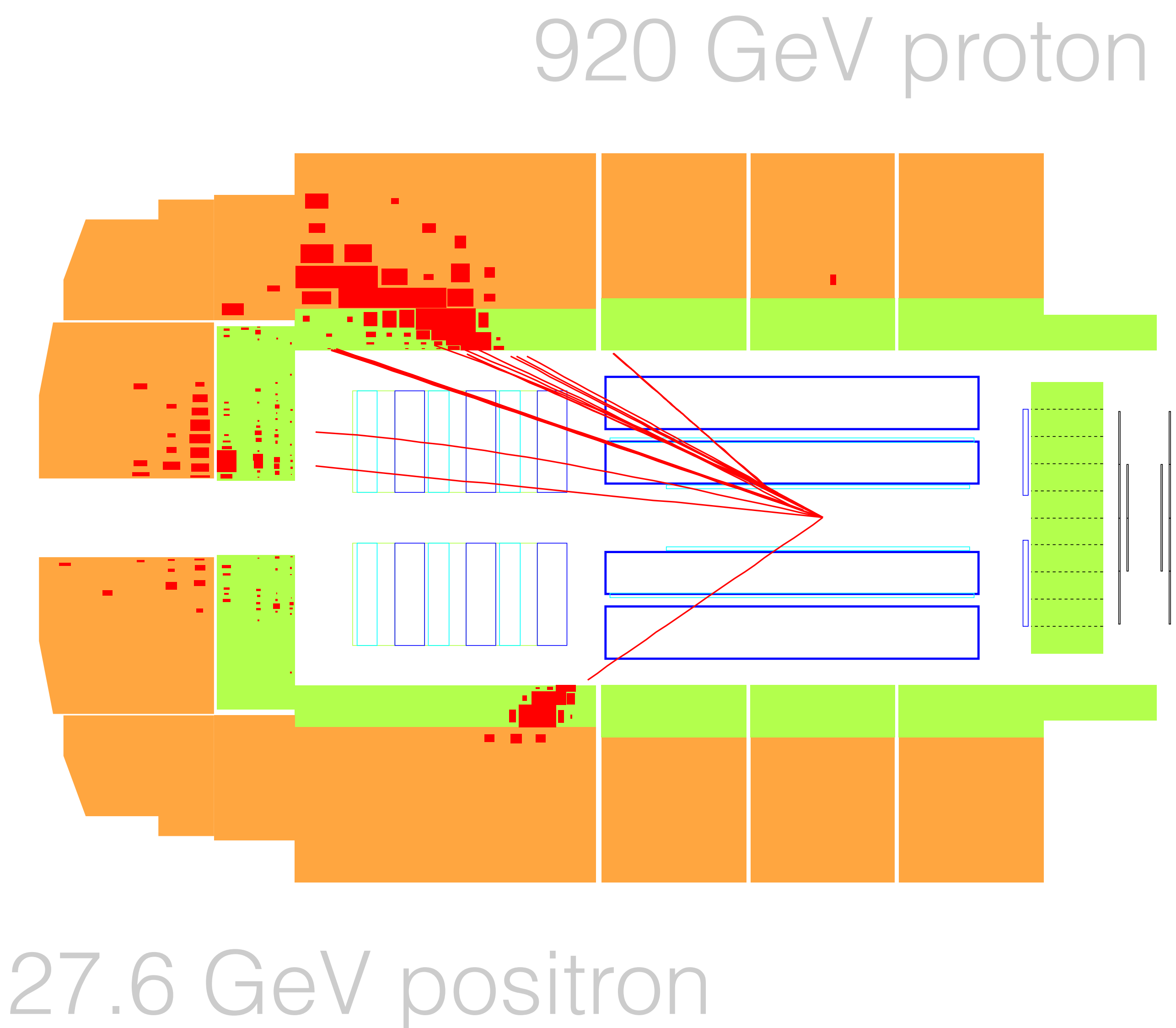


Typically, jets are studied in the Breit frame, where the Born-level configuration is discarded

However, jet production in the lab frame can be useful for probing Transverse Momentum Dependent (TMD) Parton Distribution Functions (PDFs)

See e.g. Lieu et al. PRL (2019) 192003;
Gutierrez et al. PRL (2018) 162001

Jets at H1

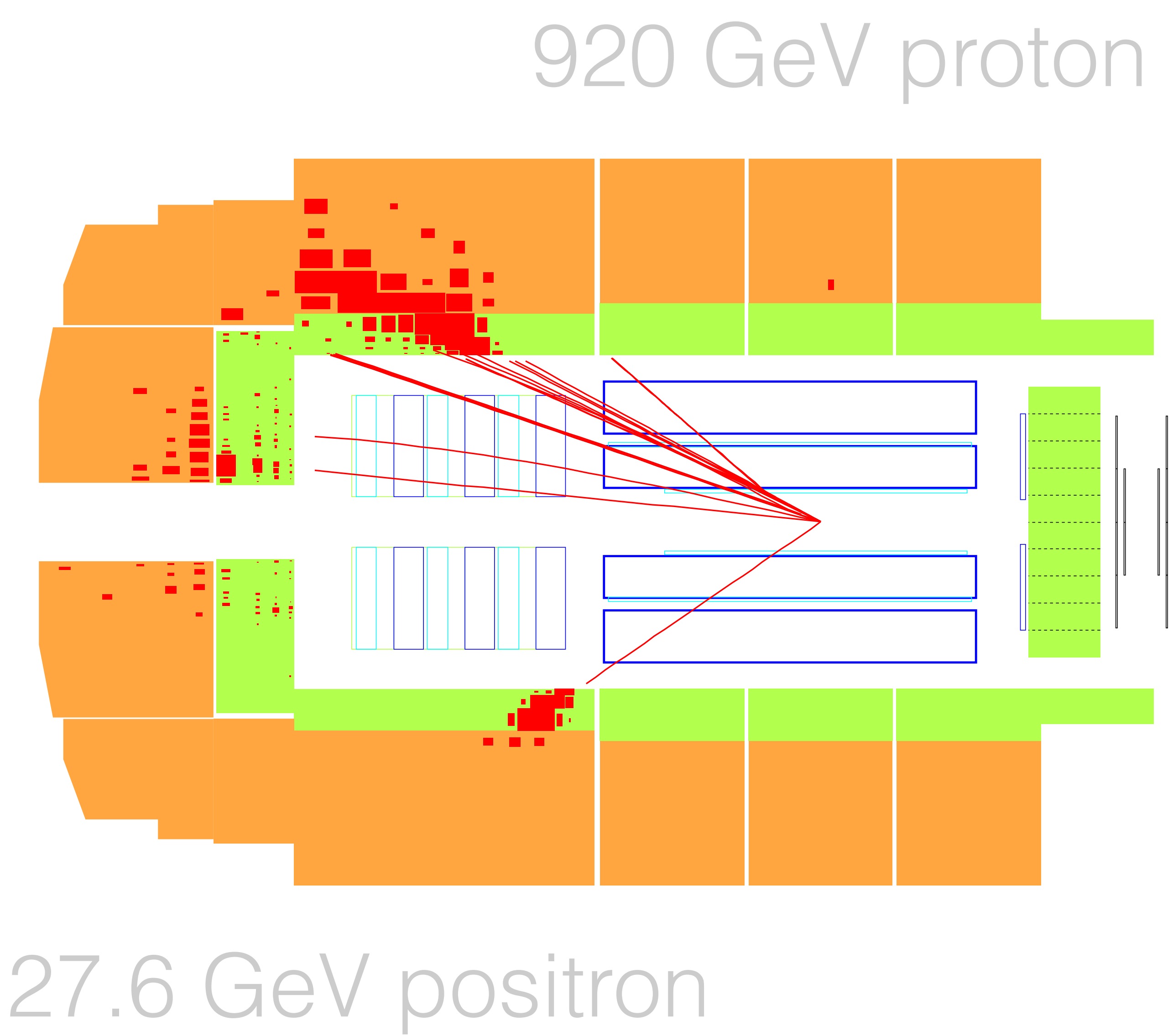


Energy flow algorithm (HFS) combines information from tracker and calorimeters

Neural network-based energy regression

1% jet energy scale uncertainty; 0.5-1% lepton energy scale uncertainty

Jets at H1



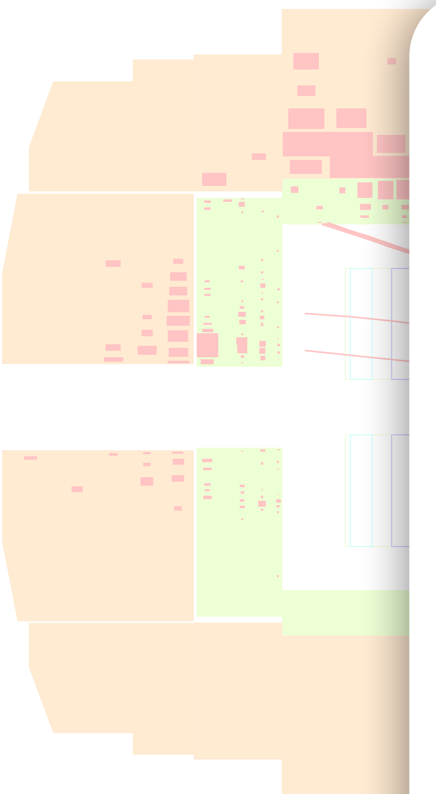
Energy flow algorithm (HFS) combines information from tracker and calorimeters

Neural network-based energy regression

1% jet energy scale uncertainty; 0.5-1% lepton energy scale uncertainty

Challenge: **unfold multidimensional phase space**

Jets at H1



Energy flow algorithm (HFS) combines information from
imeters

based
sion

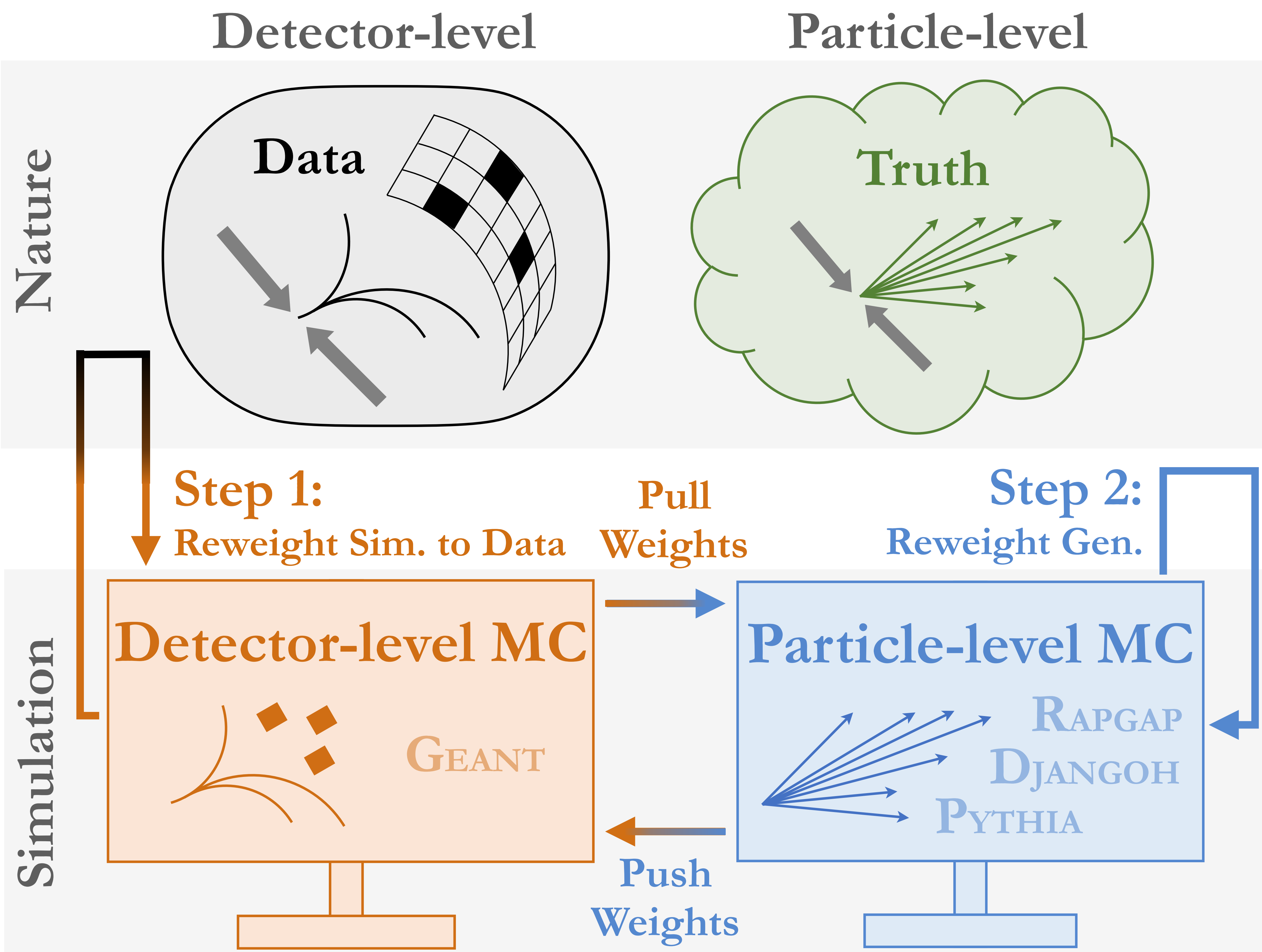
cale
lepton
ertainty

Solution: **use deep learning!**

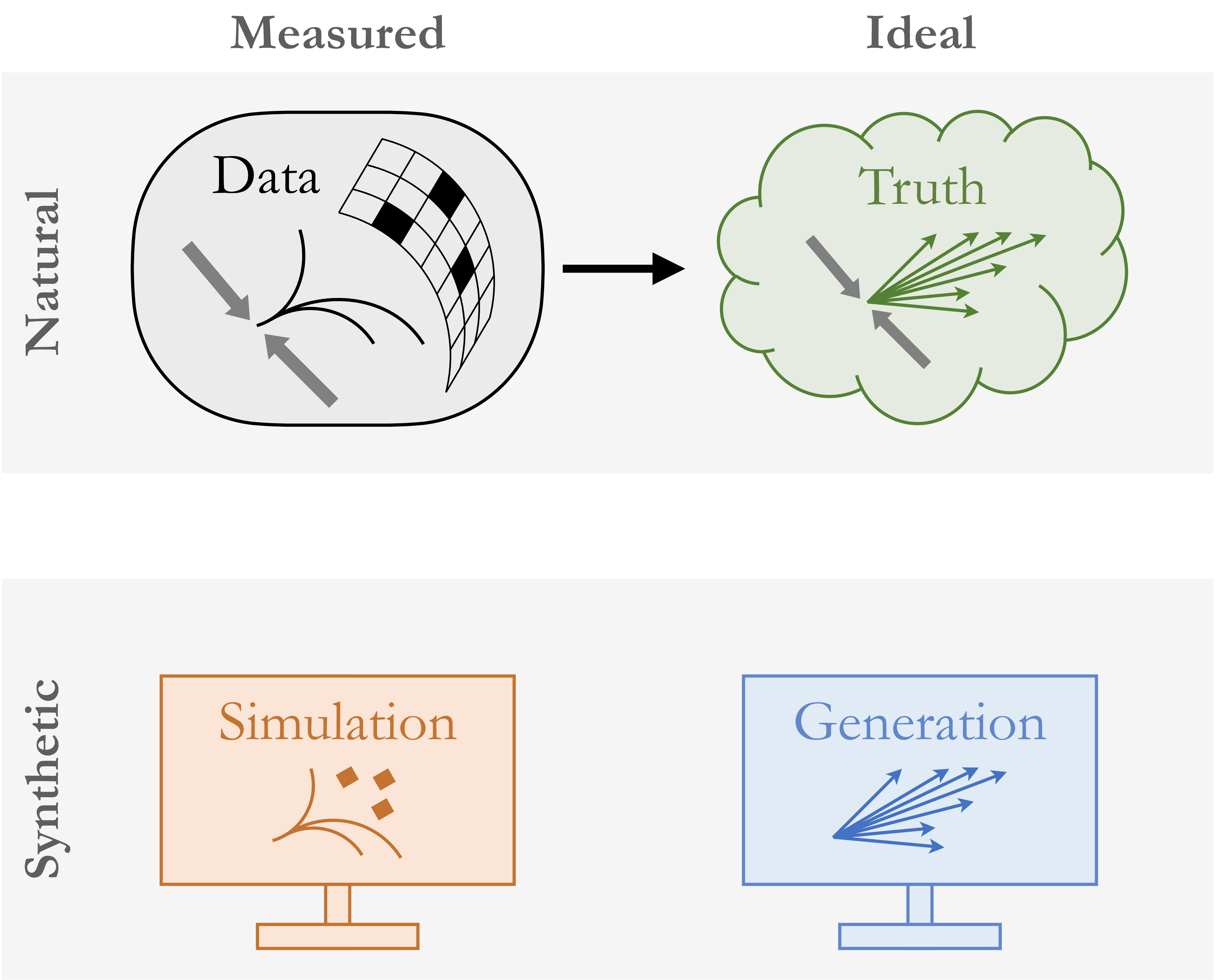
…can do unbinned, high (and variable-)dimensional unfolding

Challenge: **unfold multidimensional phase space**

Unfold by iterating: OmniFold

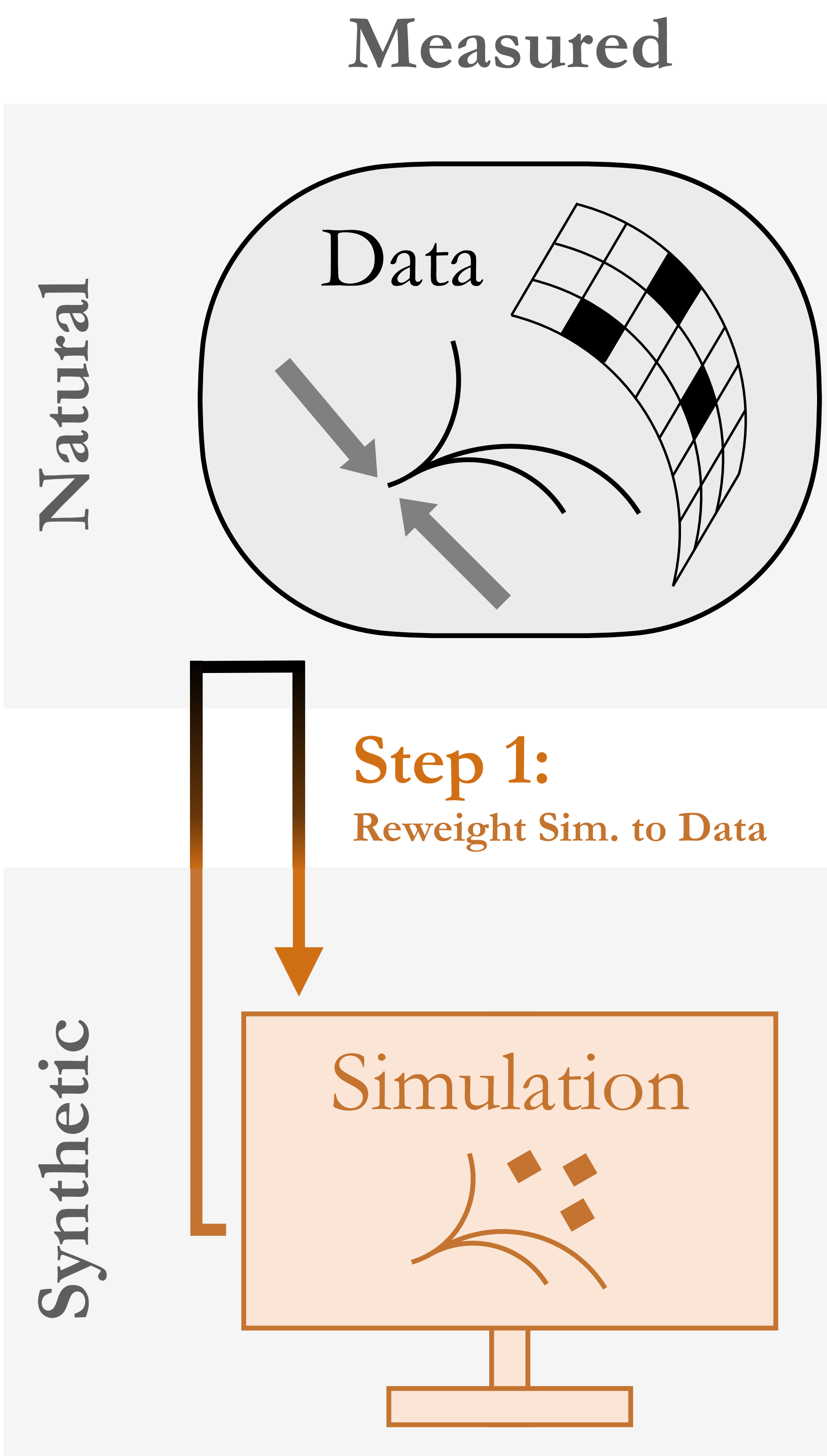


Unfold by iterating: OmniFold



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

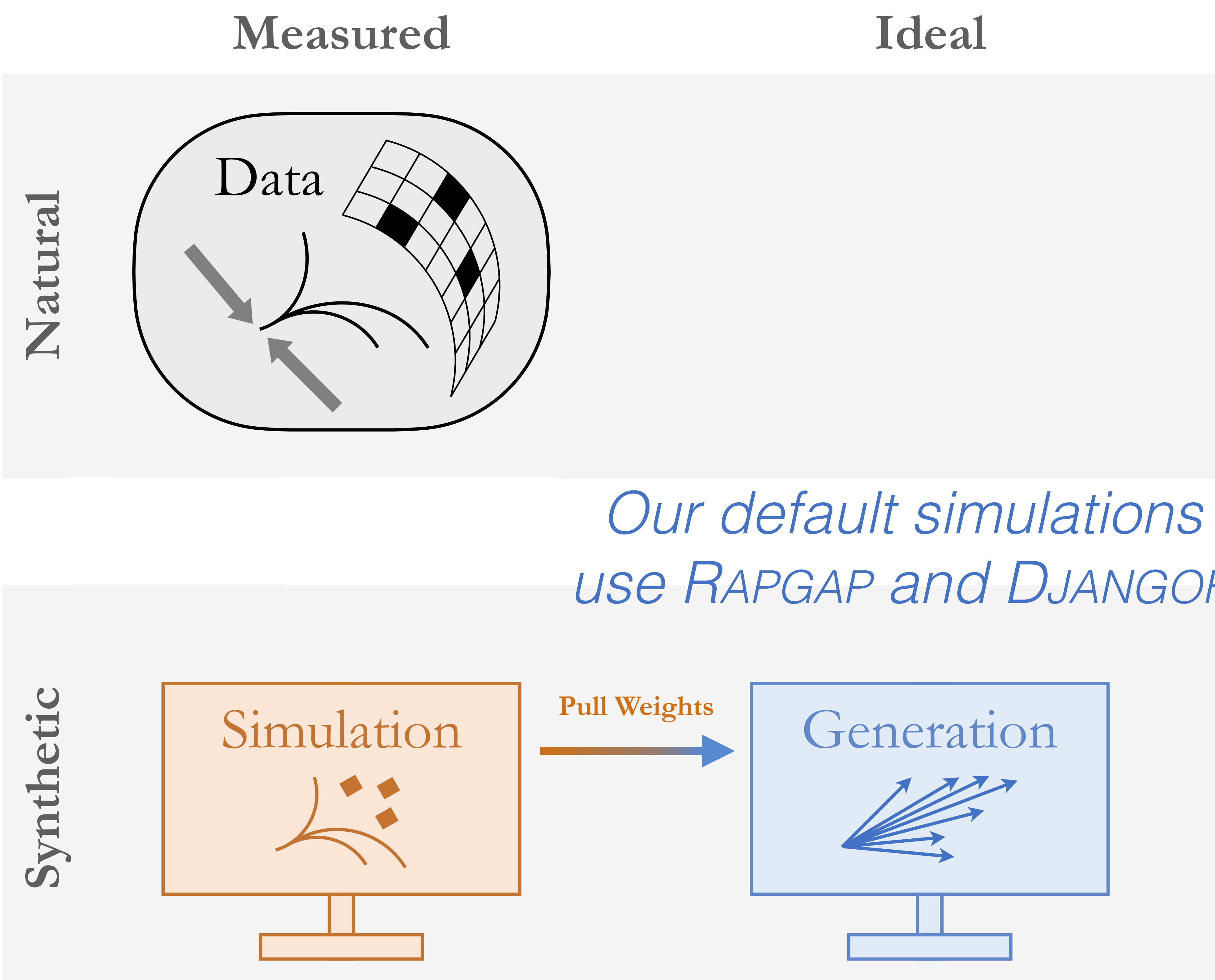
Unfold by iterating: OmniFold



Detector effects are simulated with Geant3 + H1 custom simulation code

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

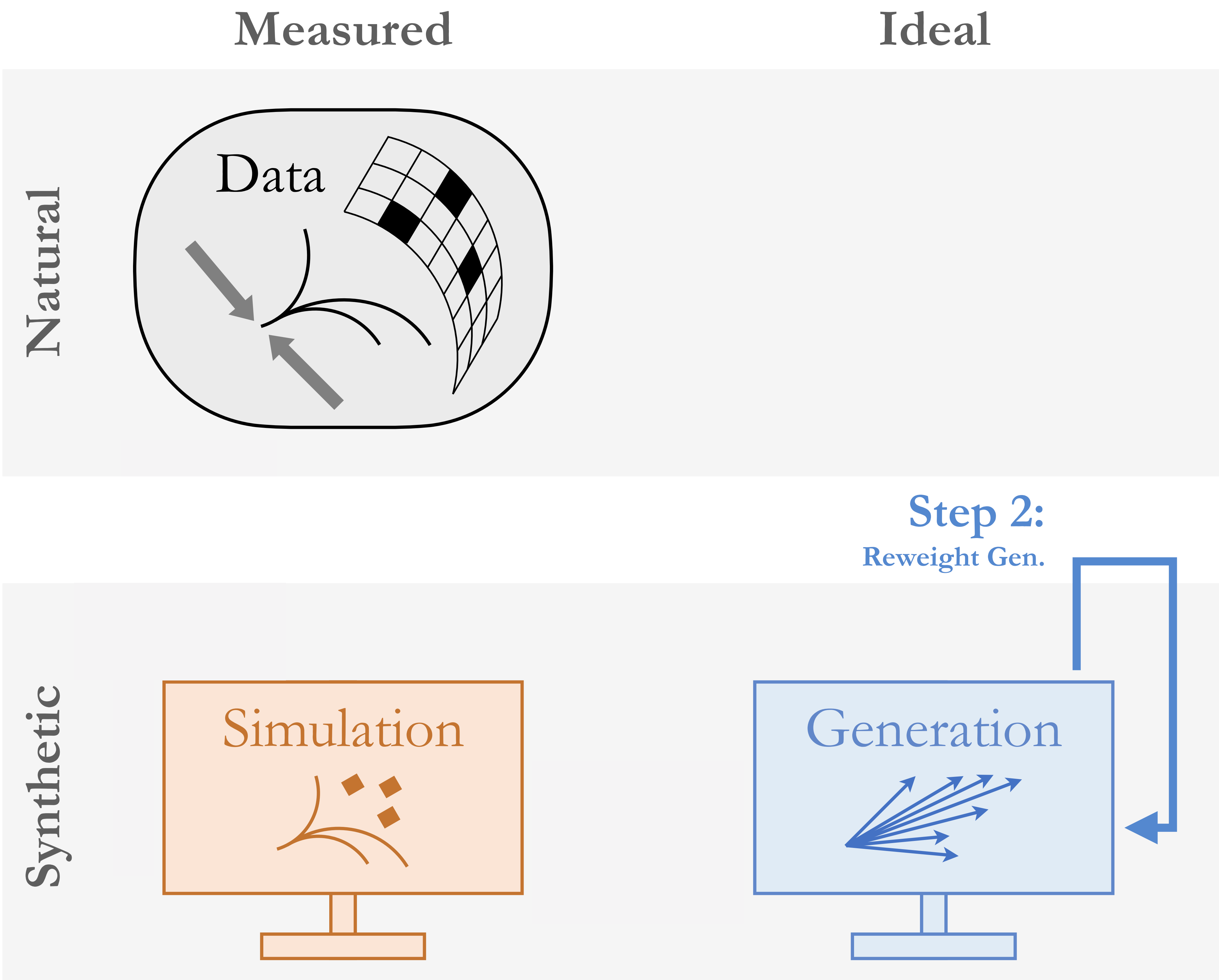
Unfold by iterating: OmniFold



Our default simulations use RAPGAP and DJANGOH

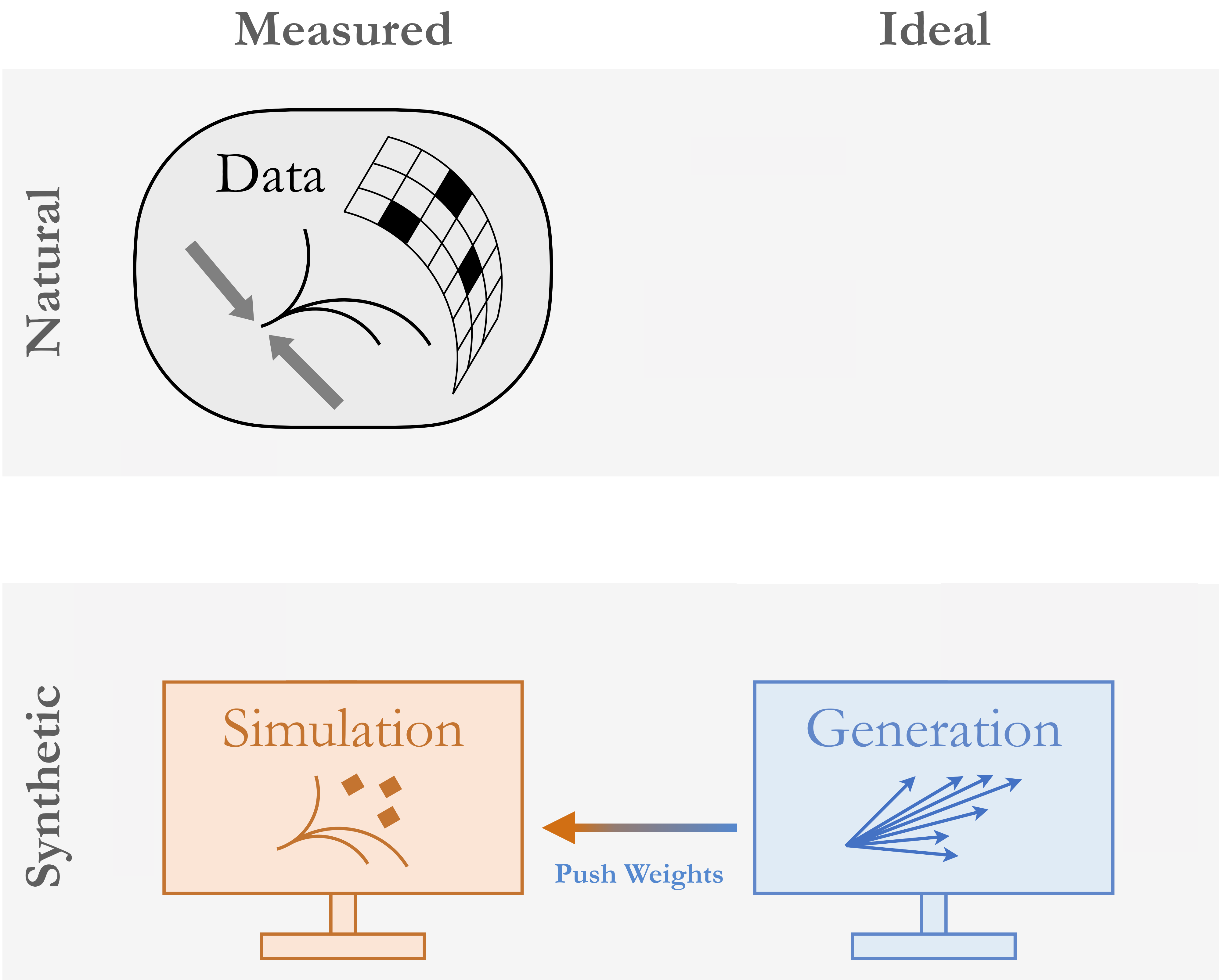
A. Andreassen, P. Komiske, E. Metodiev, BPN, J. Thaler, PRL 124 (2020) 182001

Unfold by iterating: OmniFold



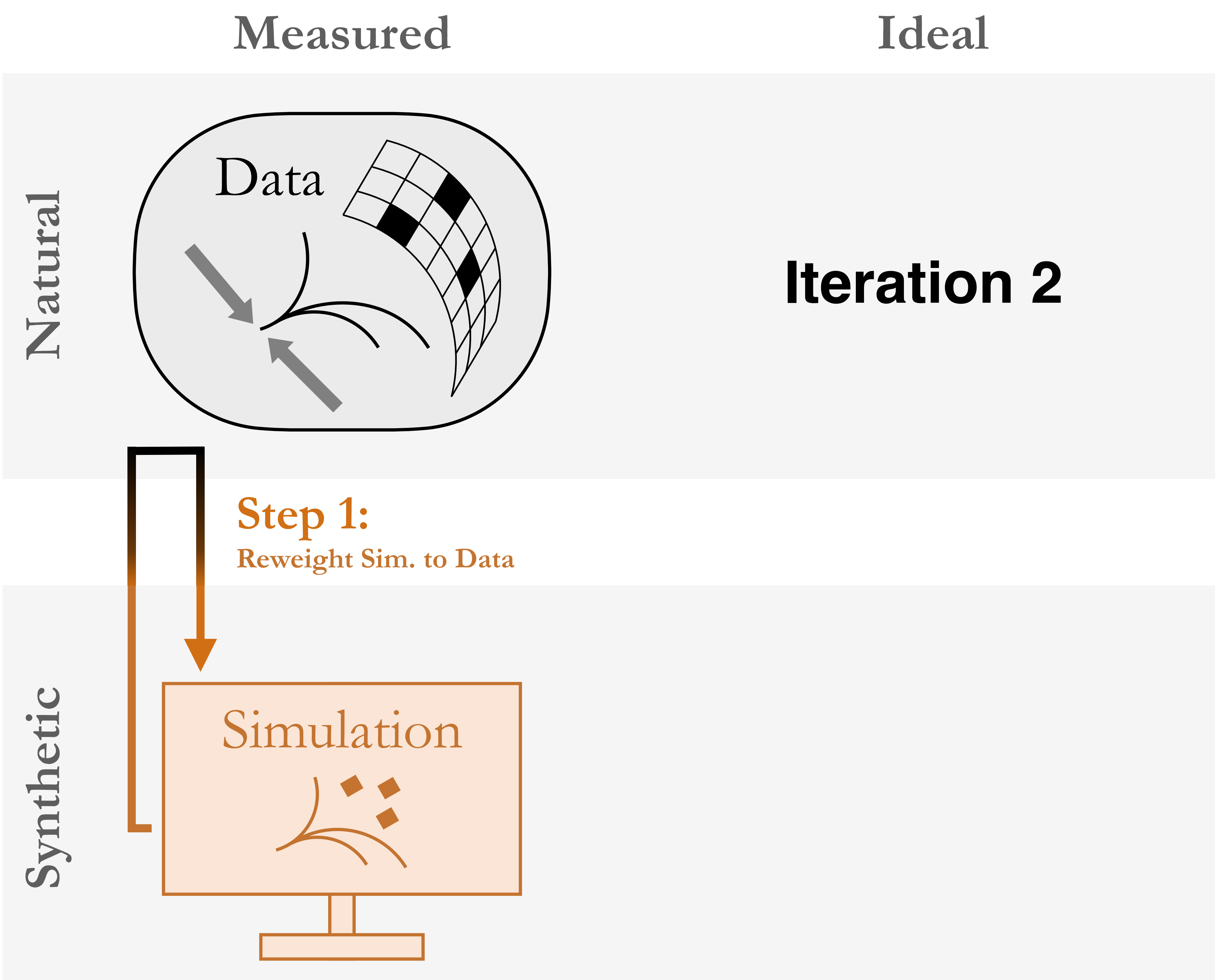
A. Andreassen, P. Komiske, E. Metodiev, BPN, J. Thaler, PRL 124 (2020) 182001

Unfold by iterating: OmniFold



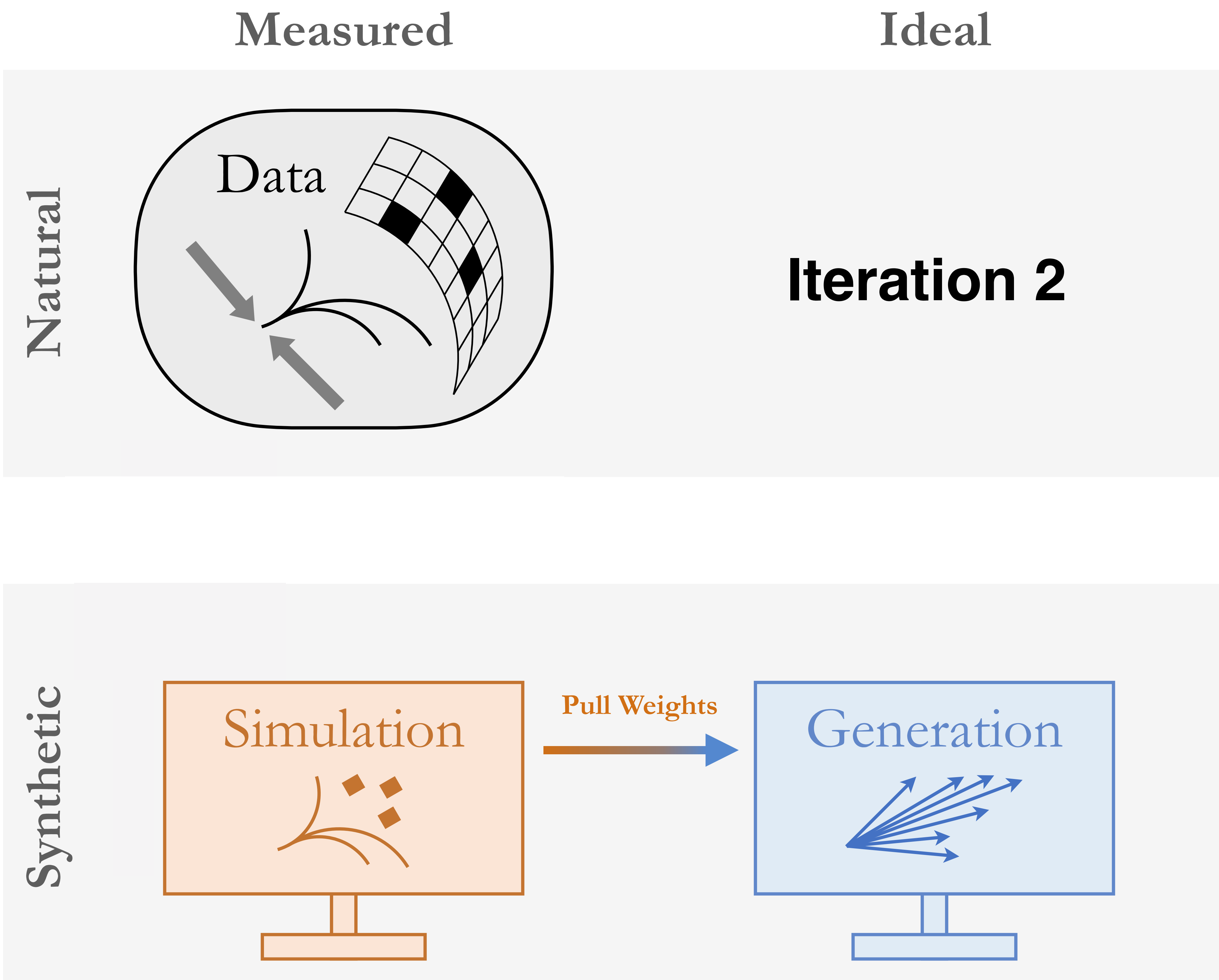
A. Andreassen, P. Komiske, E. Metodiev, BPN, J. Thaler, PRL 124 (2020) 182001

Unfold by iterating: OmniFold



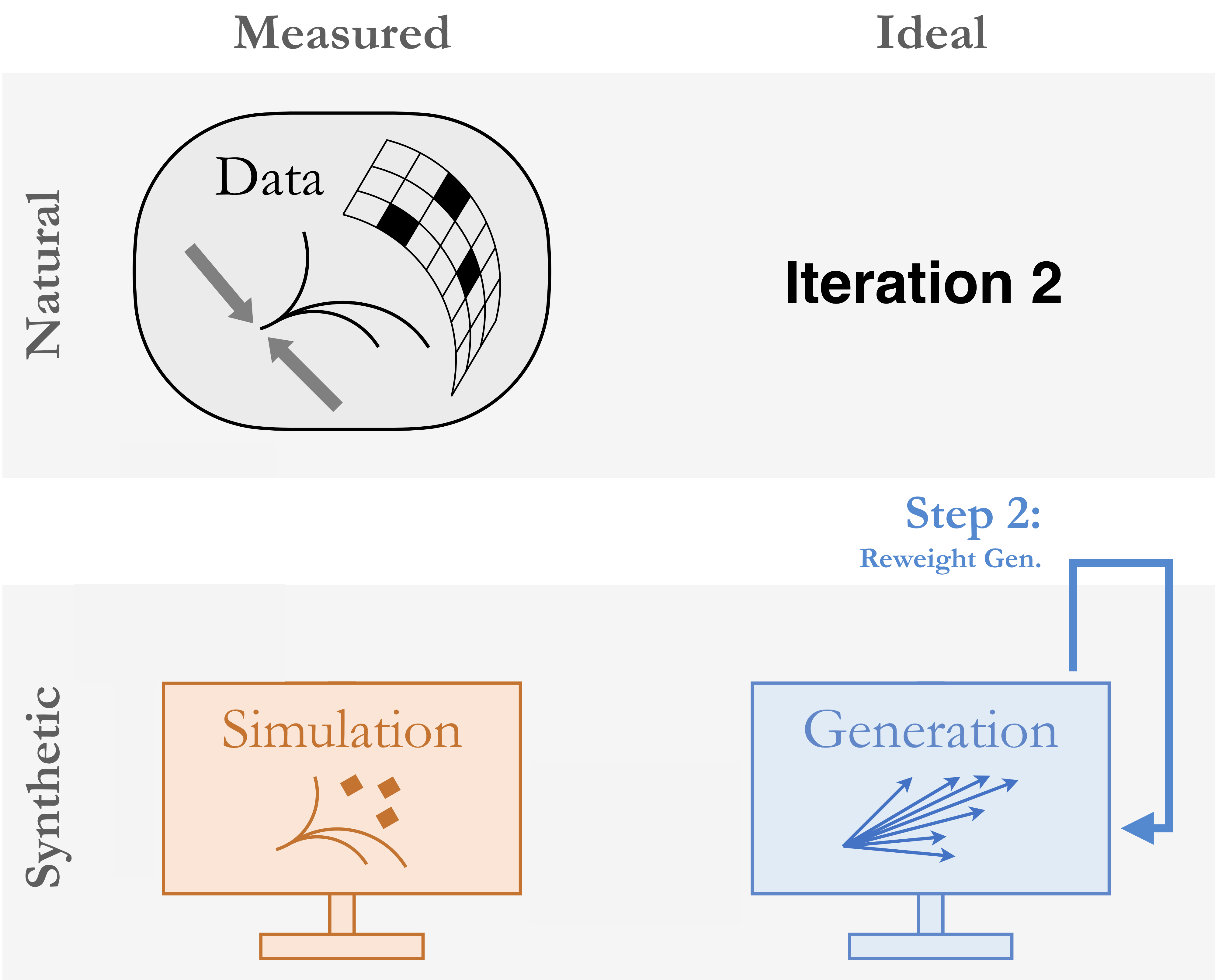
A. Andreassen, P. Komiske, E. Metodiev, BPN, J. Thaler, PRL 124 (2020) 182001

Unfold by iterating: OmniFold



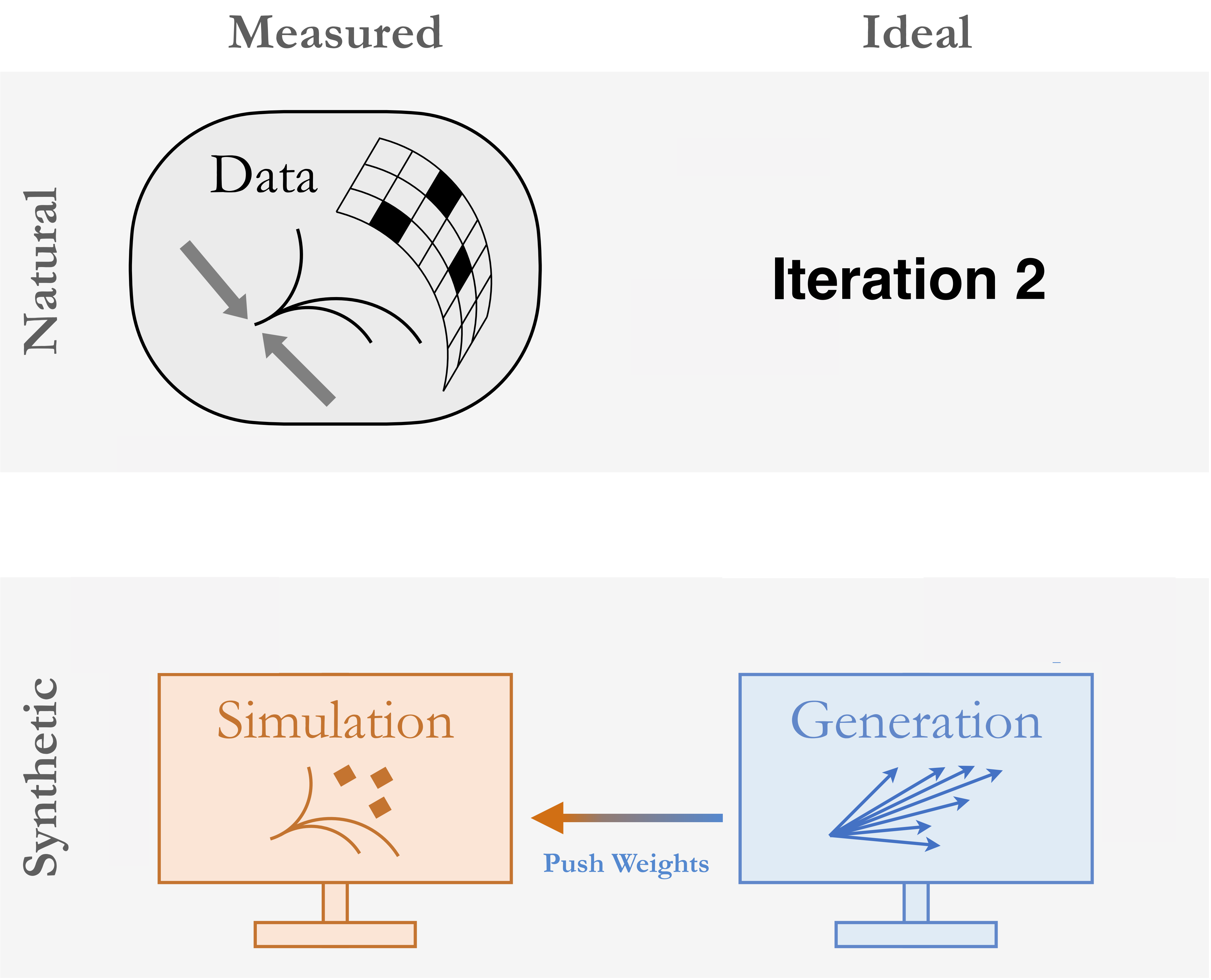
A. Andreassen, P. Komiske, E. Metodiev, BPN, J. Thaler, PRL 124 (2020) 182001

Unfold by iterating: OmniFold



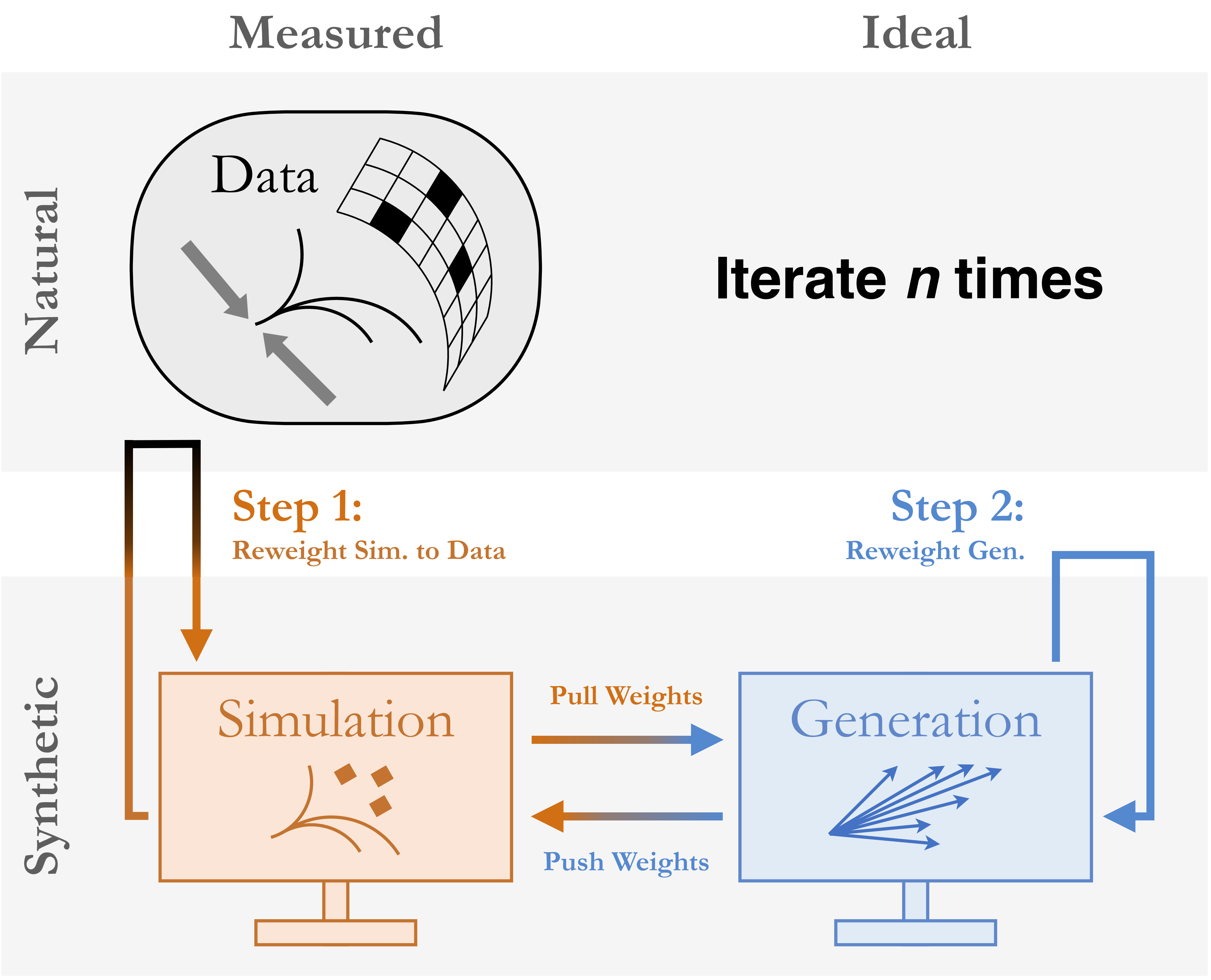
A. Andreassen, P. Komiske, E. Metodiev, BPN, J. Thaler, PRL 124 (2020) 182001

Unfold by iterating: OmniFold



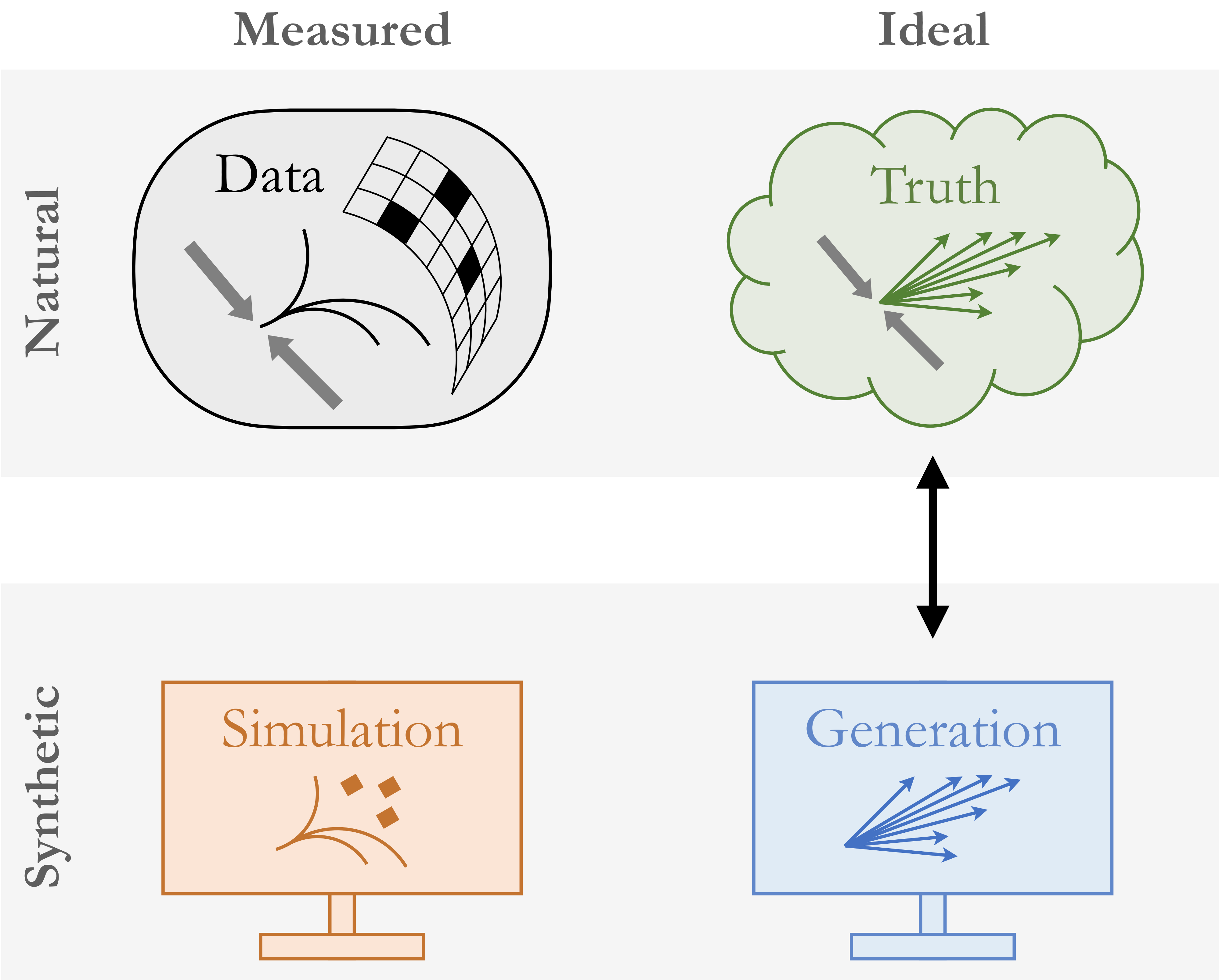
A. Andreassen, P. Komiske, E. Metodiev, BPN, J. Thaler, PRL 124 (2020) 182001

Unfold by iterating: OmniFold



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

Unfold by iterating: OmniFold

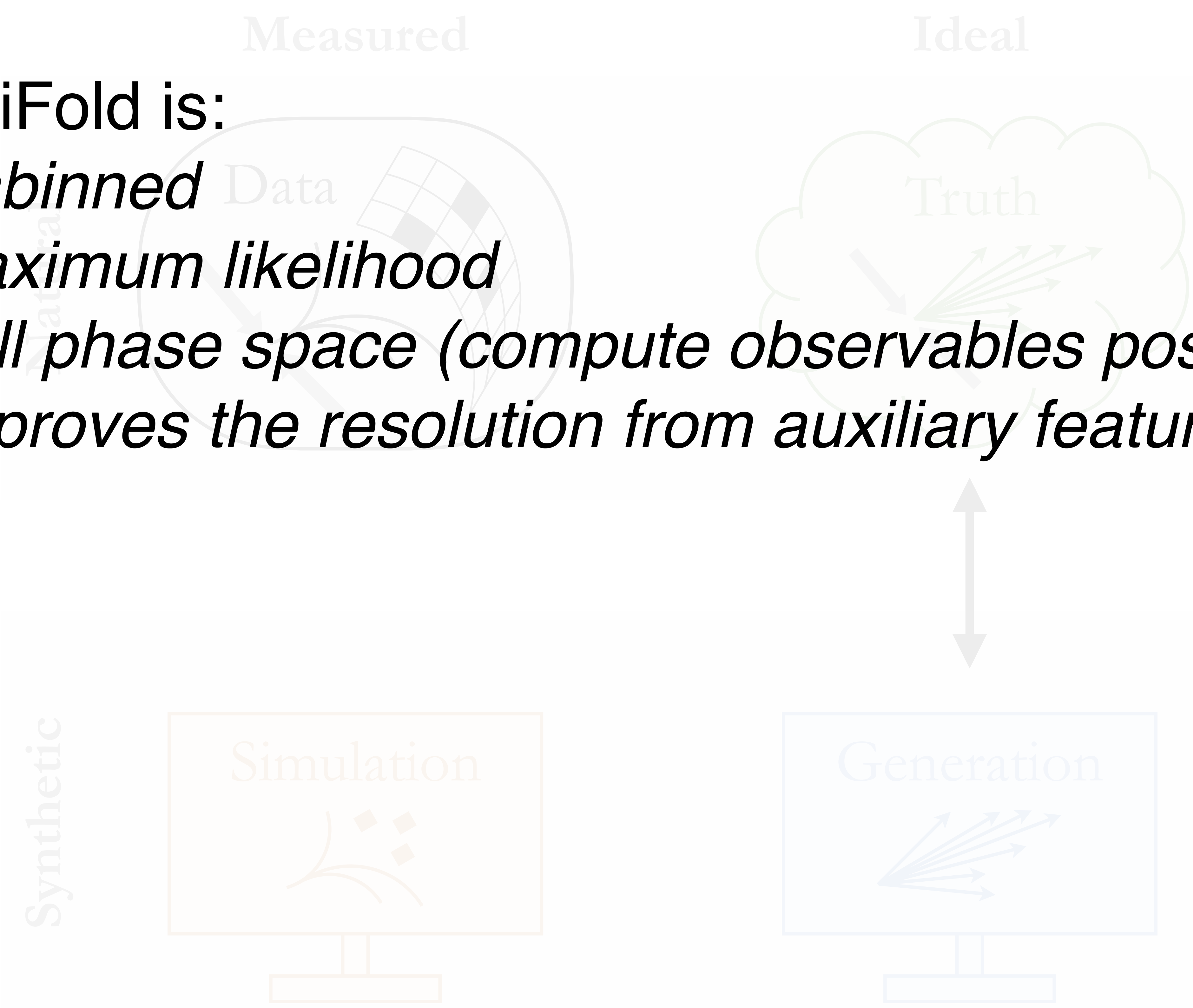


A. Andreassen, P. Komiske, E. Metodiev,
BPN, J. Thaler, PRL 124 (2020) 182001

Unfold by iterating: OmniFold

OmniFold is:

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



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

Unfold by iterating: OmniFold

OmniFold is:

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

In this measurement: simultaneously unfold lepton and jet kinematics and report binned spectra for jet p_T, $\Delta\phi$, q_T/Q, and jet η

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

Classification for reweighting

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

Classification for reweighting

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

We use a trick whereby classifiers can be repurposed as reweighters

Classification for reweighting

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

We use a trick whereby classifiers can be repurposed as reweighters

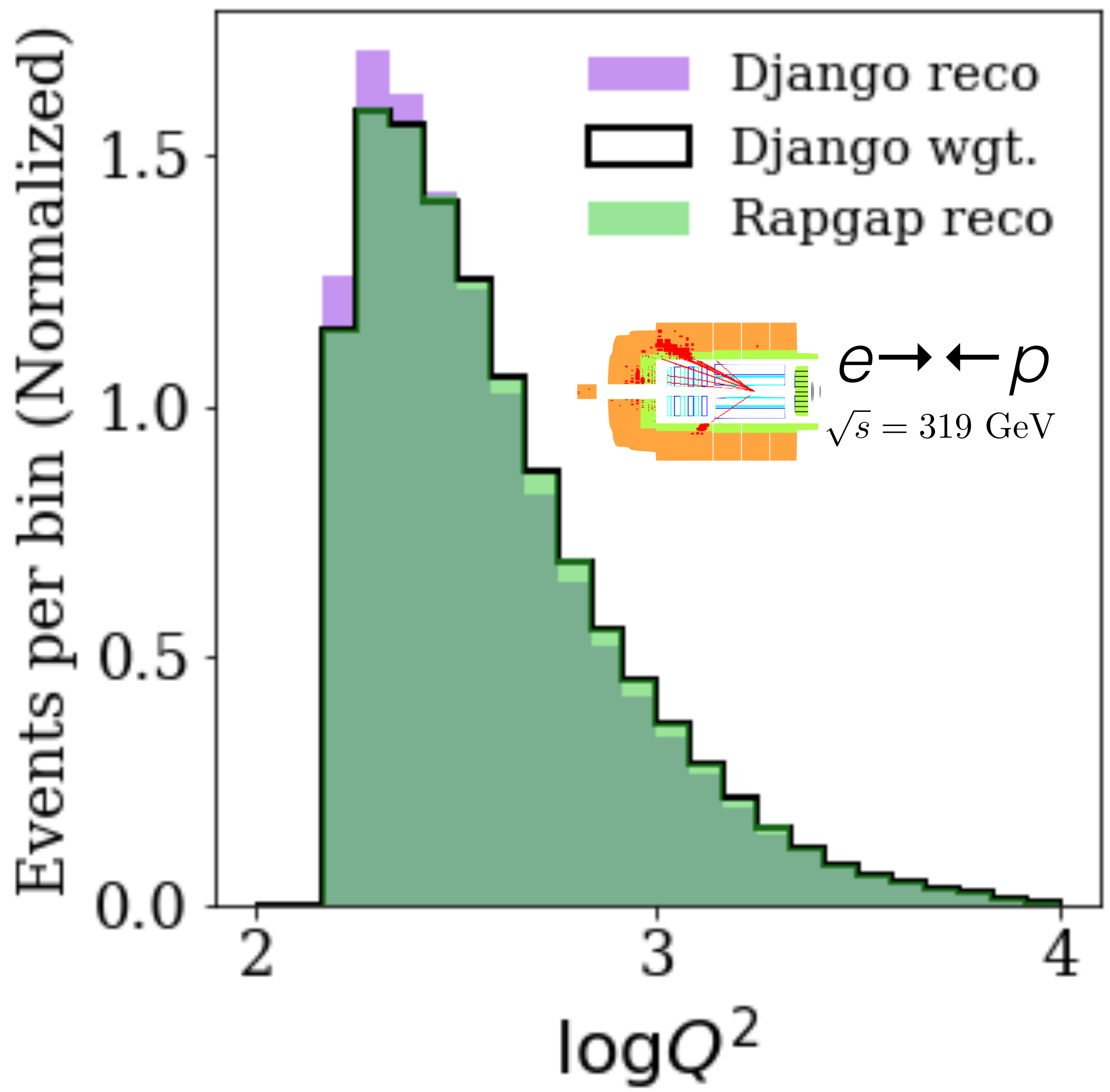
$$\frac{p_1(x)}{p_0(x)} \approx \frac{\mathrm{NN}(x)}{1 - \mathrm{NN}(x)}$$

Classifier (NN) trained to distinguish data sampled from p_1 versus p_0.

Classification for reweighting

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

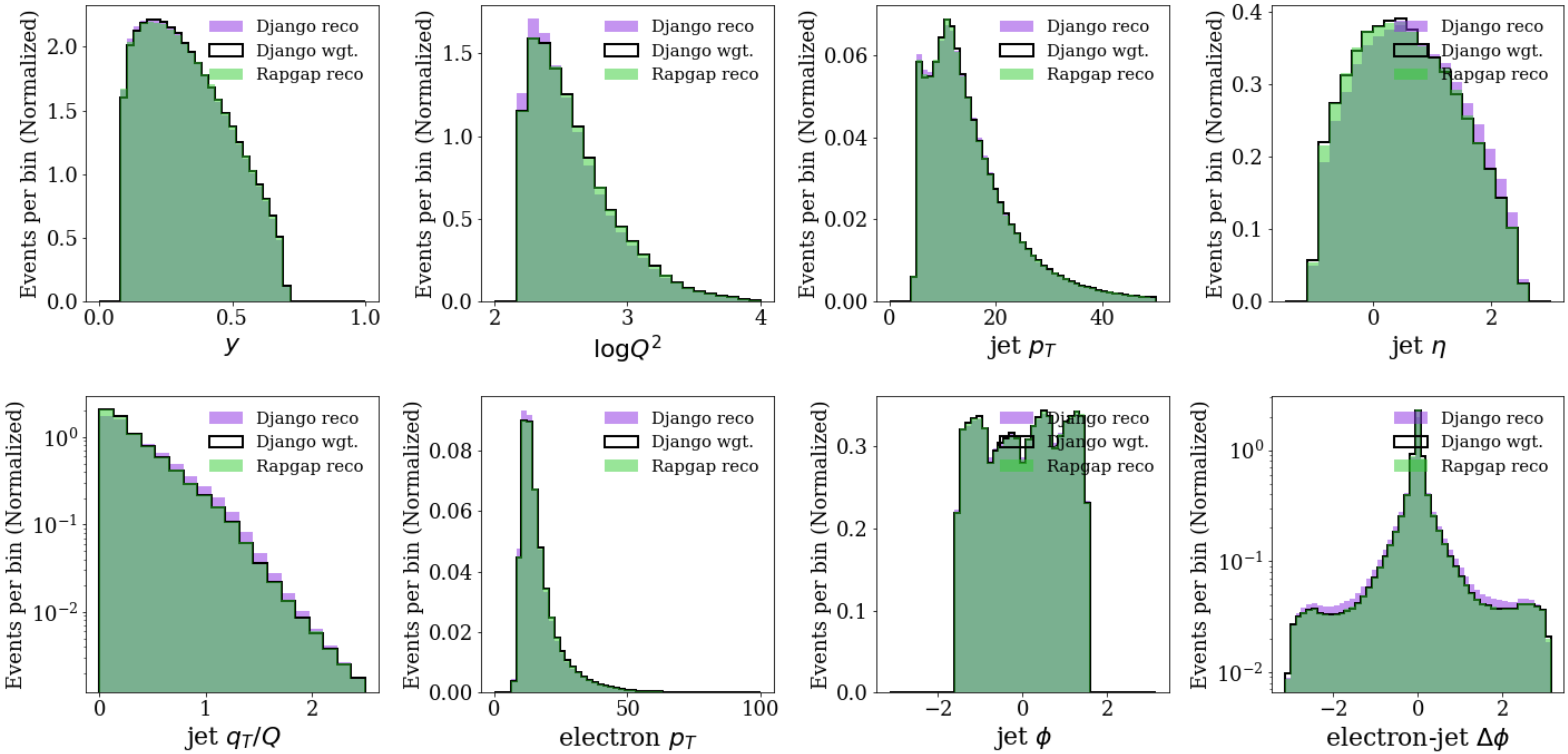
We use a trick whereby classifiers can be repurposed as reweighters



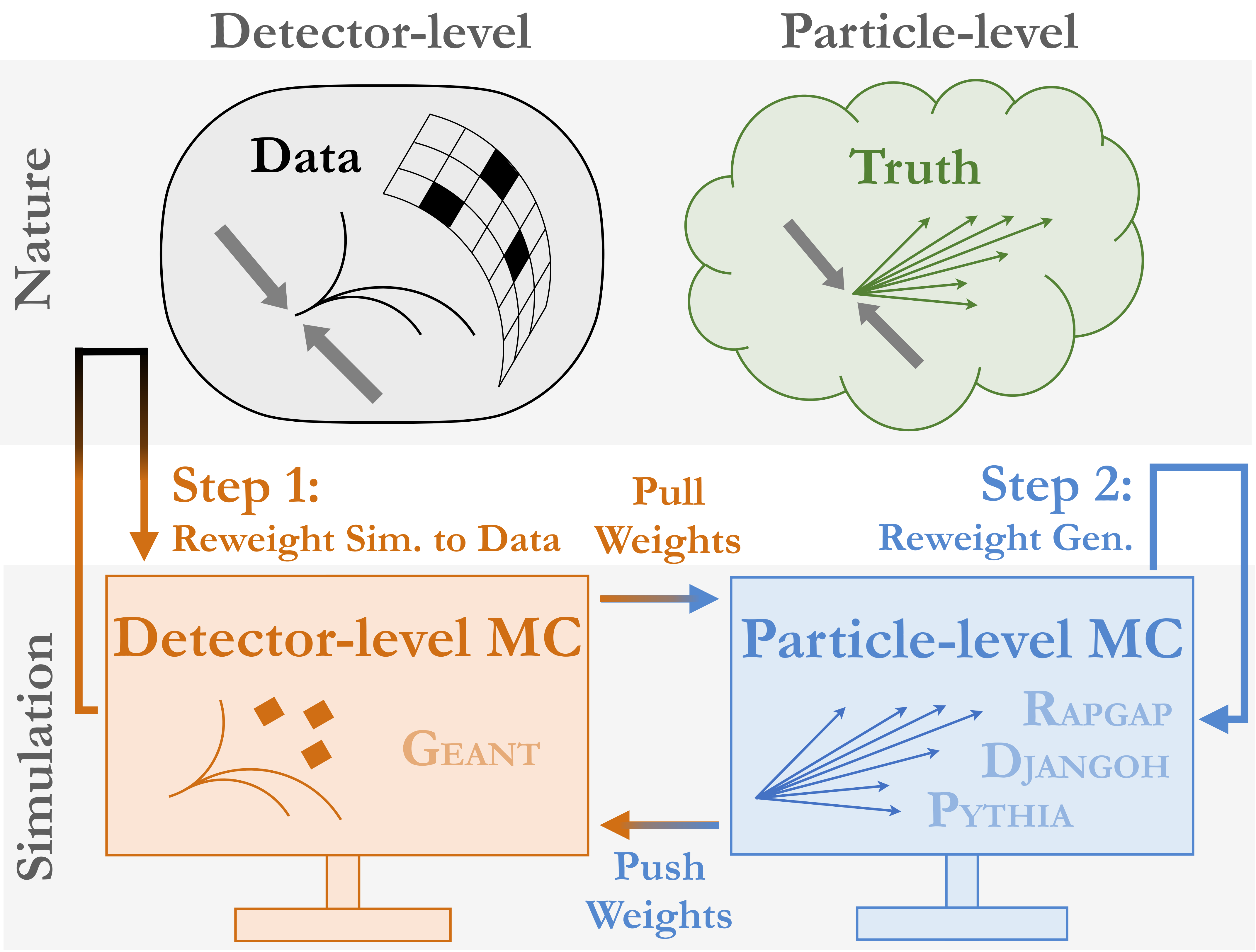
N.B. the distribution is binned for illustration, but the reweighting is unbinned.

Classification for reweighting

All of these distributions are simultaneously reweighted!

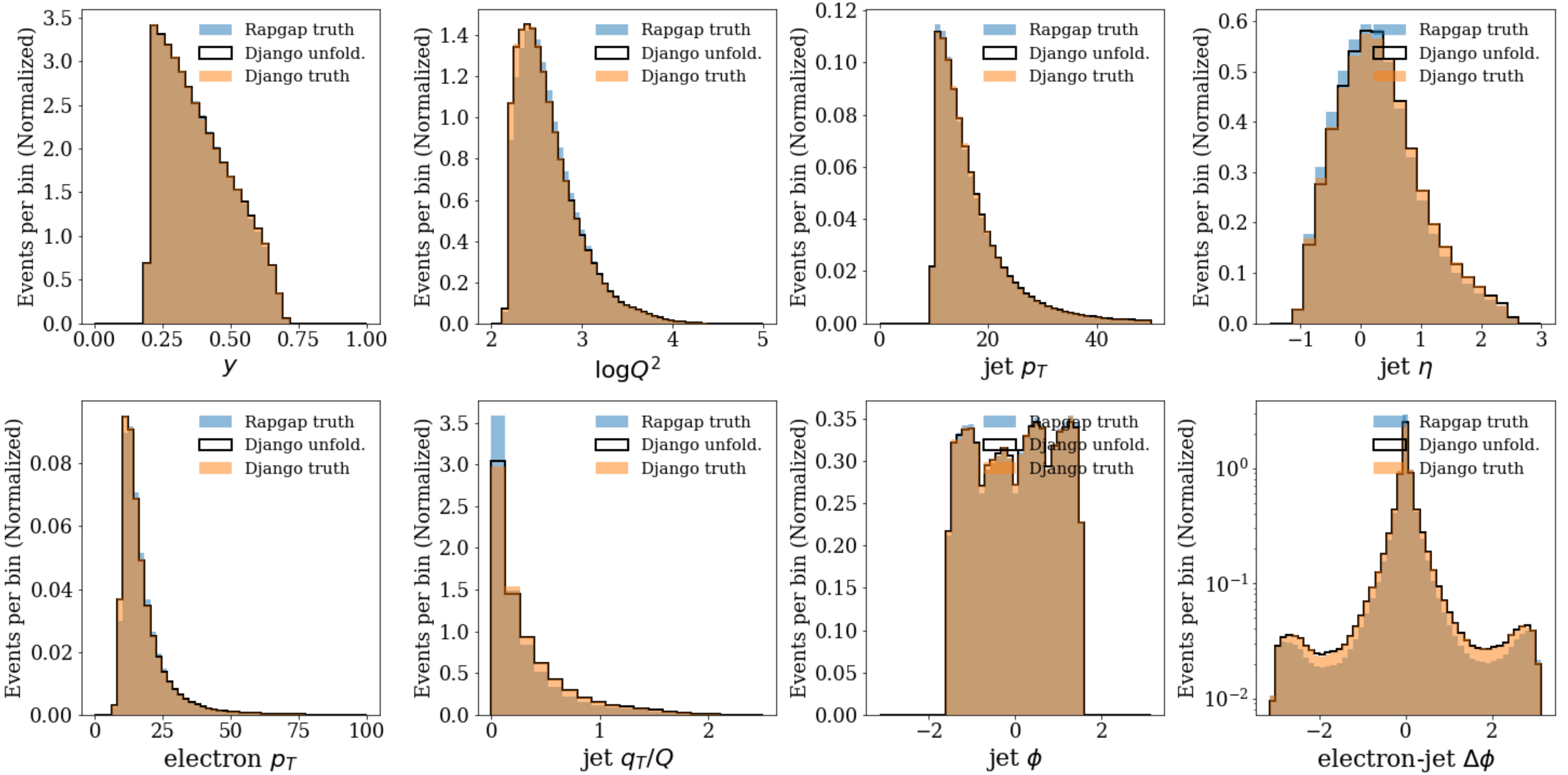


Unfold by iterating: OmniFold



OmniFolding *ep* simulations

We see excellent closure for the full phase space!



2108.12376

DESY 21-130, ISSN 0418-9833

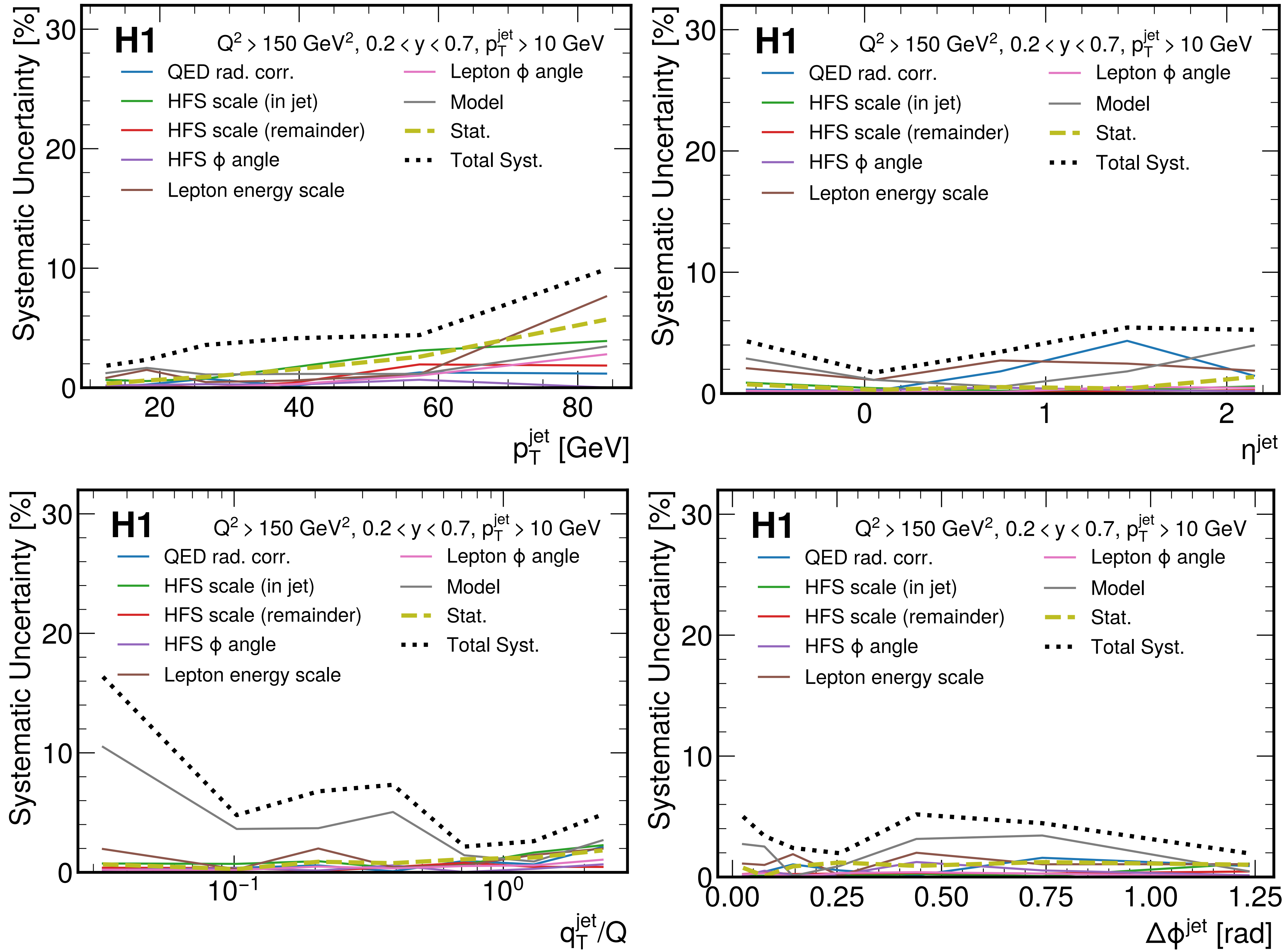
Measurement of lepton-jet correlation in deep-inelastic scattering with the H1 detector using machine learning for unfolding

H1 Collaboration*

(To be submitted to Physical Review Letters)
(Dated: August 30, 2021)

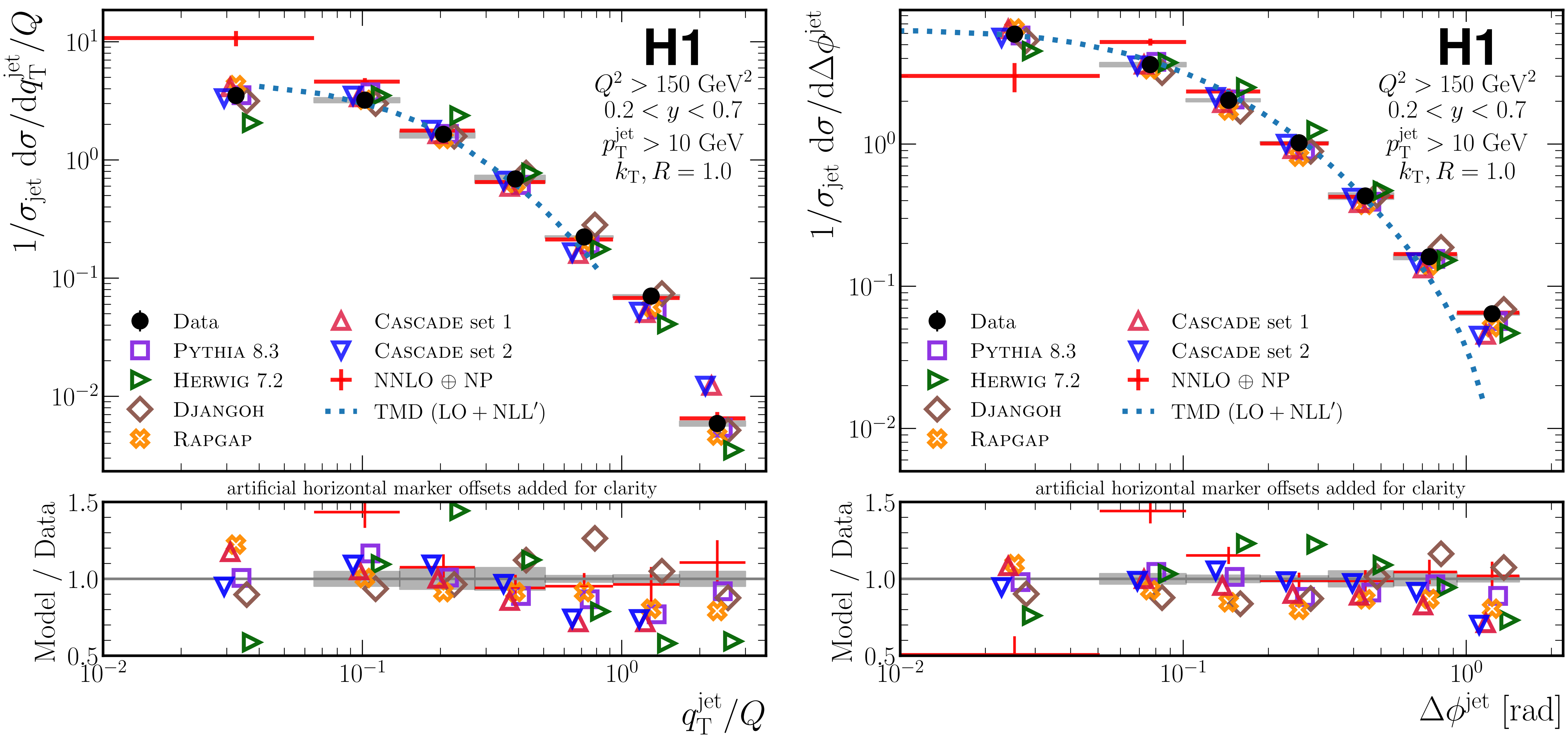
The first measurement of lepton-jet momentum imbalance and azimuthal correlation in lepton-proton scattering at high momentum transfer is presented. These data, taken with the H1 detector at HERA, are corrected for detector effects using an unbinned machine learning algorithm (OmniFold), which considers eight observables simultaneously in this first application. The unfolded cross sections are compared to calculations performed within the context of collinear or transverse-momentum-dependent (TMD) factorization in Quantum Chromodynamics (QCD) as well as Monte Carlo event generators. The measurement probes a wide range of QCD phenomena, including TMD parton distribution functions and their evolution with energy in so far unexplored kinematic regions.

Results



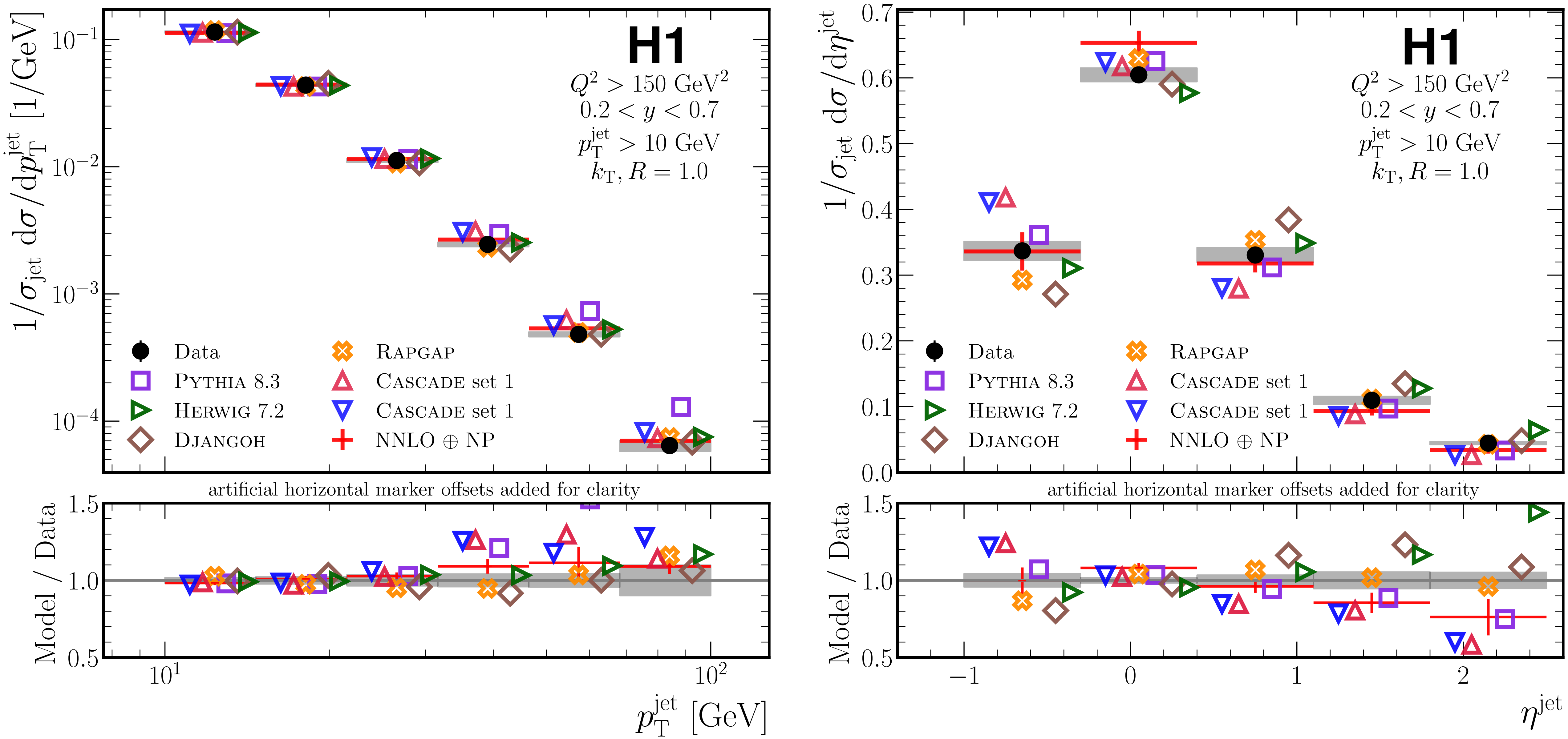
H1
Q² > 150 GeV², 0.2 < y < 0.7, p_T^{jet} > 10 GeV
QED rad. corr.
HFS scale (in jet)
HFS scale (remainder)
HFS ϕ angle
Lepton energy scale
Lepton ϕ angle
Model
Stat.
Total Syst.
Systematic Uncertainty [%]
p_T^{jet} [GeV]
η^{jet}
q_T^{jet}/Q
$\Delta\phi^{jet}$ [rad]

Results



Excellent agreement with fixed order at high q_T,
excellent agreement with TMD prediction at low q_T.

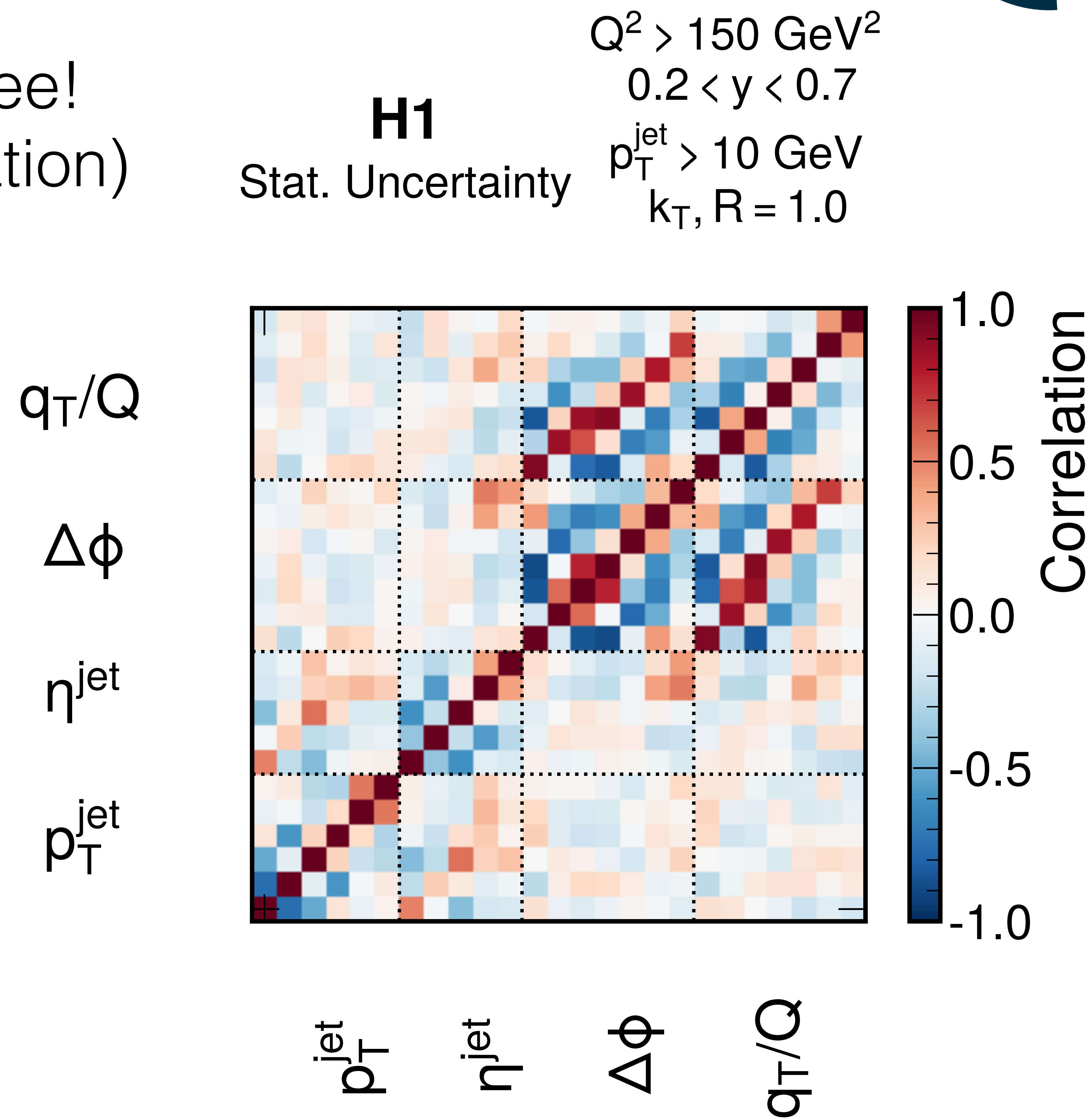
Results



Parton shower Monte Carlo programs also provide excellent agreement with the data across the spectra.

Results

Simultaneous for free!
(binning is for illustration)



Looking forward

Publishing unbinned measurements is tricky - we have started a conversation about this in a recent paper. Feedback is most welcome!

2109.13243

Presenting Unbinned Differential Cross Section Results

Miguel Arratia,[a,b] Anja Butter,[c] Mario Campanelli,[d] Vincent Croft,[e] Dag Gillberg,[f] Kristin Lohwasser,[g] Bogdan Malaescu,[h] Vinicius Mikuni,[i] Benjamin Nachman,[j,k] Juan Rojo,[l,m] Jesse Thaler,[n,o] Ramon Winterhalder[p]

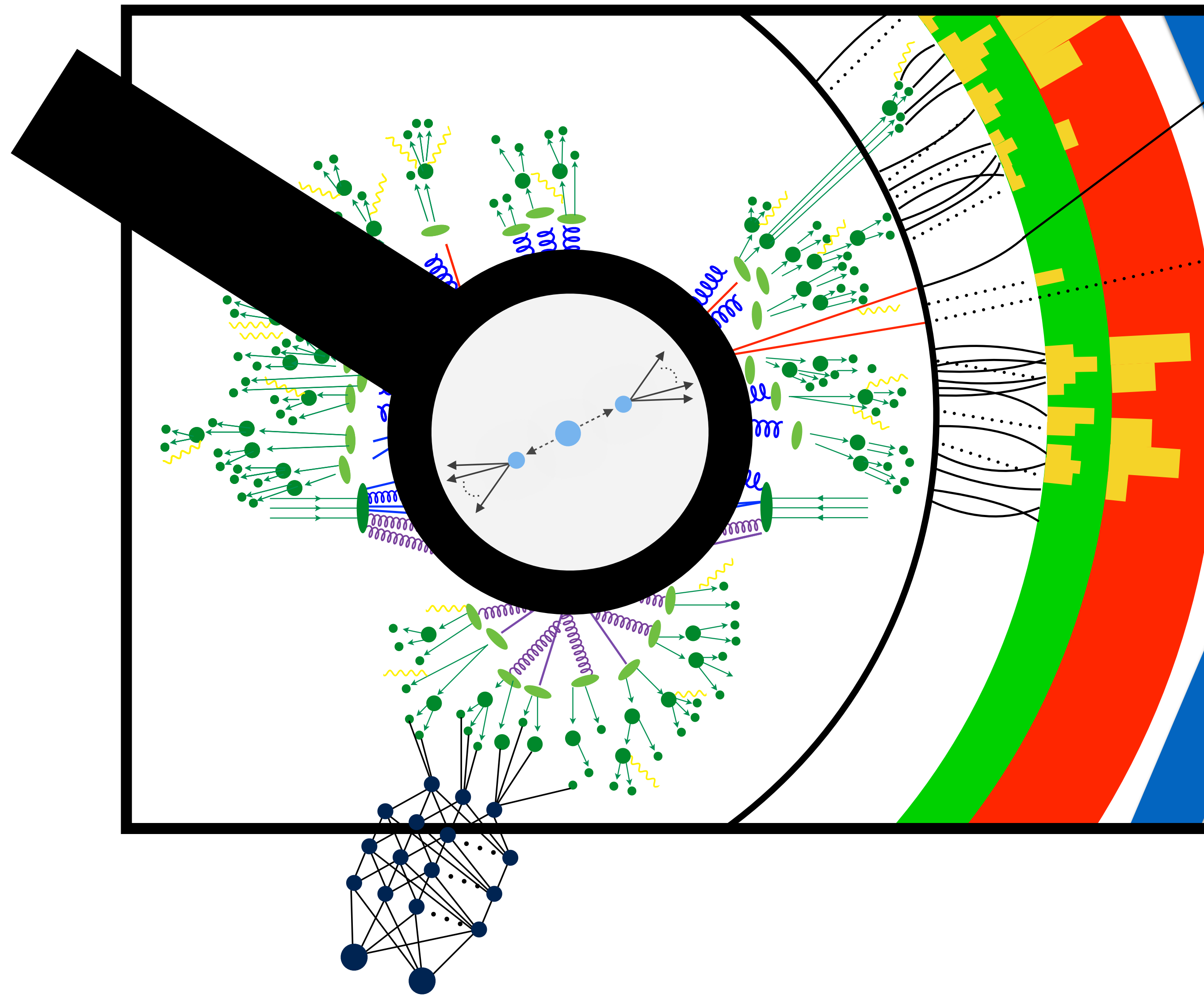
[a] *Department of Physics and Astronomy, University of California, Riverside, CA 92521, USA*
[b] *Thomas Jefferson National Accelerator Facility, Newport News, VA 23606, USA*
[c] *Institut für Theoretische Physik, Universität Heidelberg, Heidelberg, Germany*
[d] *University College London, London WC1E 6BT, UK*
[e] *Department of Physics and Astronomy, Tufts University, Boston, MA 02155, USA*
[f] *Department of Physics, Carleton University, Ottawa ON K1S 5B6, Canada*
[g] *University of Sheffield, Sheffield, S10 2TN, UK*
[h] *LPNHE, Sorbonne Université, Université de Paris, CNRS/IN2P3, Paris, France*
[i] *National Energy Research Scientific Computing Center, Berkeley, CA 94720, USA*
[j] *Physics Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA*
[k] *Berkeley Institute for Data Science, University of California, Berkeley, CA 94720, USA*
[l] *Nikhef Theory Group, Science Park 105, 1098 XG Amsterdam, The Netherlands*
[m] *Department of Physics and Astronomy, Vrije Universiteit Amsterdam, NL-1081 HV Amsterdam, The Netherlands*
[n] *Center for Theoretical Physics, Massachusetts Institute of Technology, Cambridge, MA 02139, USA*
[o] *The NSF AI Institute for Artificial Intelligence and Fundamental Interactions*
[p] *Centre for Cosmology, Particle Physics and Phenomenology (CP3), Université catholique de Louvain, 1348 Louvain-la-Neuve, Belgium*

E-mail: bpnachman@lbl.gov

Abstract: Machine learning tools have empowered a qualitatively new way to perform differential cross section measurements whereby the data are unbinned, possibly in many dimensions. Unbinned measurements can enable, improve, or at least simplify comparisons between experiments and with theoretical predictions. Furthermore, many-dimensional measurements can be used to define observables after the measurement instead of before. There is currently no community standard for publishing unbinned data. While there are also essentially no measurements of this type public, unbinned measurements are expected in the near future given recent methodological advances. The purpose of this paper is to propose a scheme for presenting and using unbinned results, which can hopefully form the basis for a community standard to allow for integration into analysis workflows. This is foreseen to be the start of an evolving community dialogue, in order to accommodate future developments in this field that is rapidly evolving.

Conclusions and outlook

Today, I have presented the first ML-based unfolding with collider data

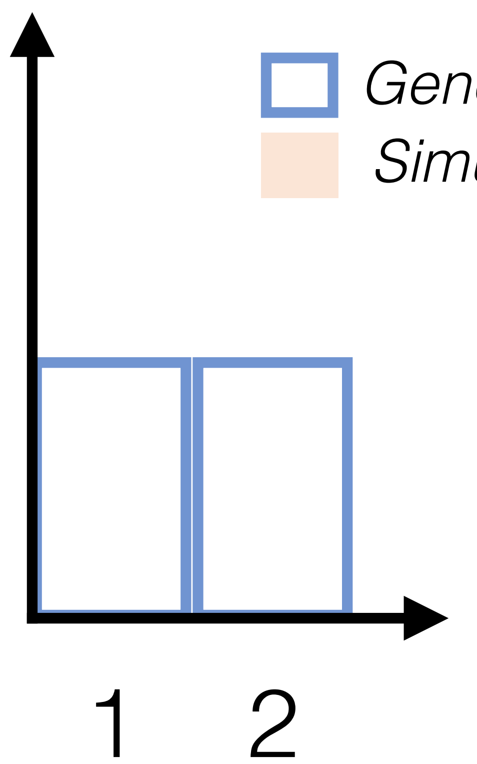


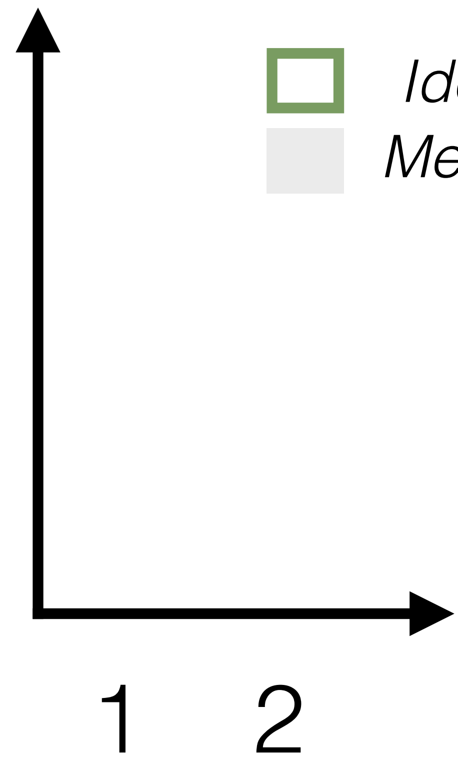
This is the start of an exciting program to advance our study of QCD into higher dimensions

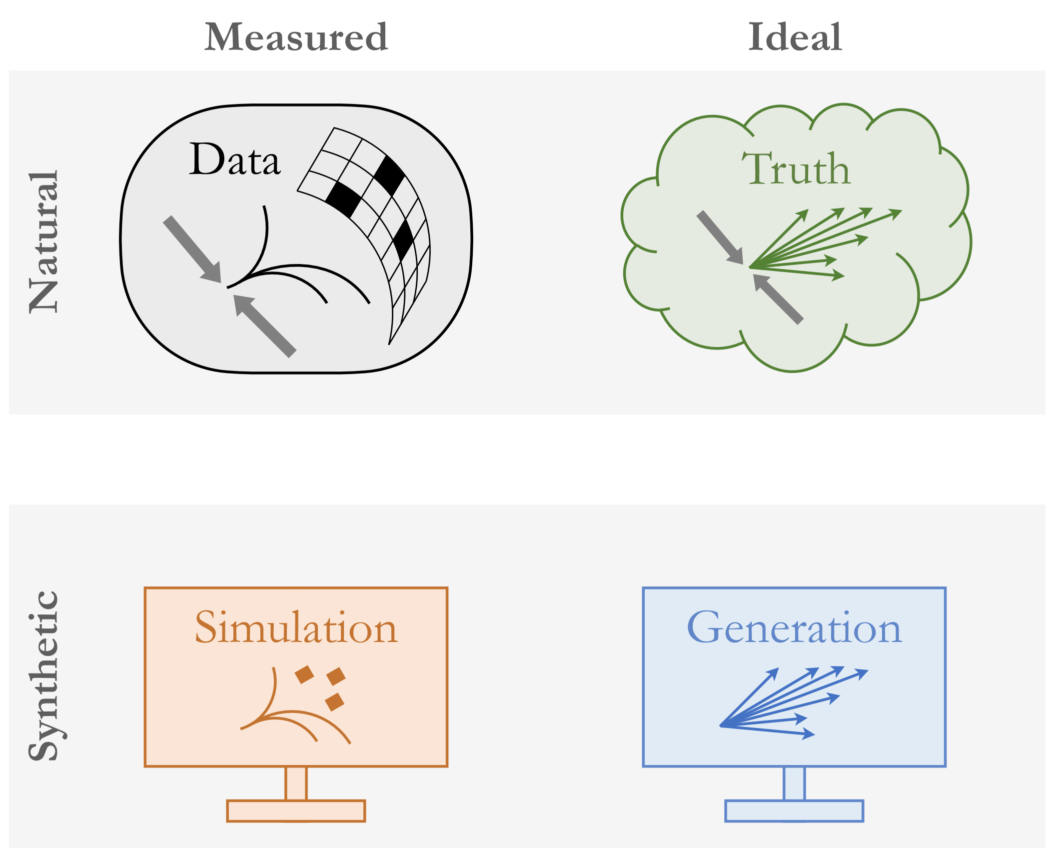
This particular measurement has important constraining power for TMD PDFs and provides important input to planning and design for the future EIC

Backup

Unfold by iterating: OmniFold

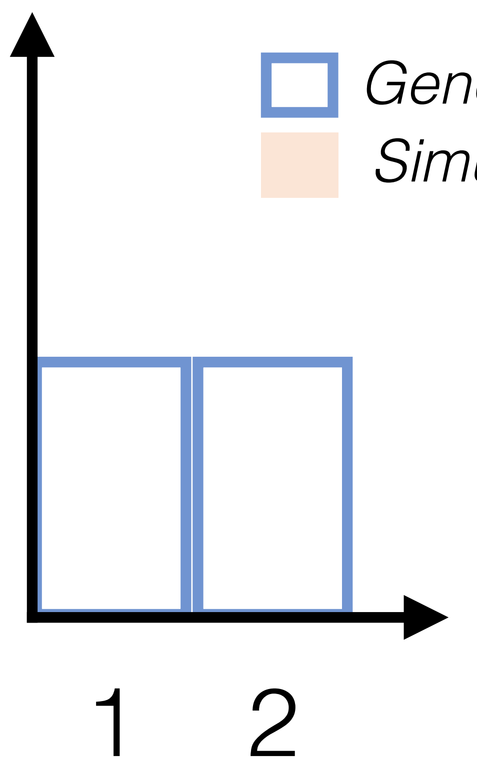


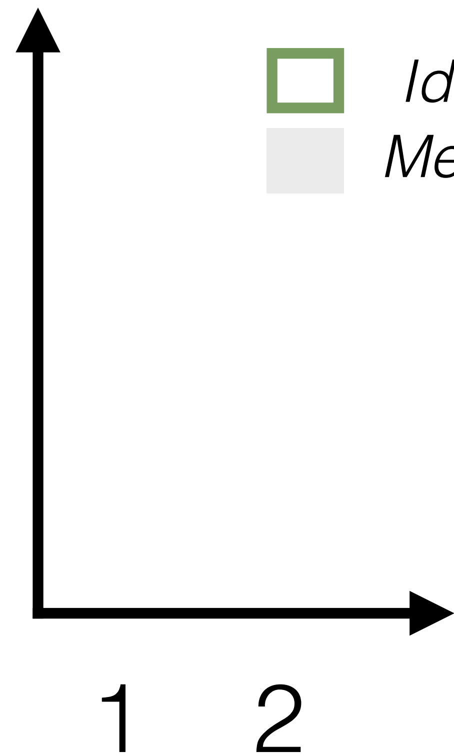


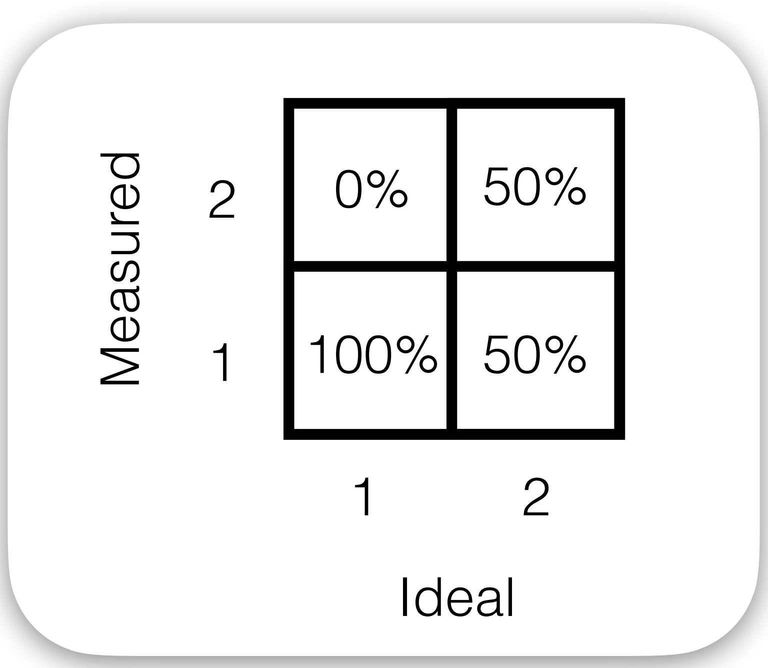


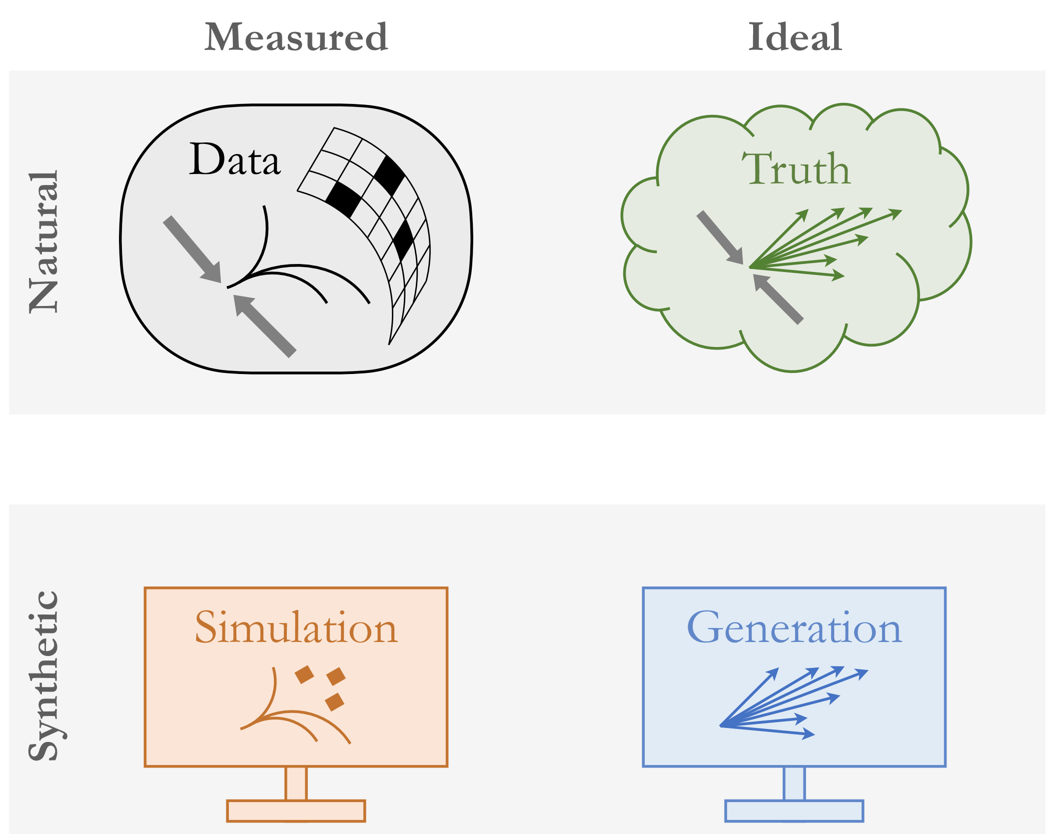
A. Andreassen, P. Komiske, E. Metodiev, BPN, J. Thaler, PRL 124 (2020) 182001

Unfold by iterating: OmniFold



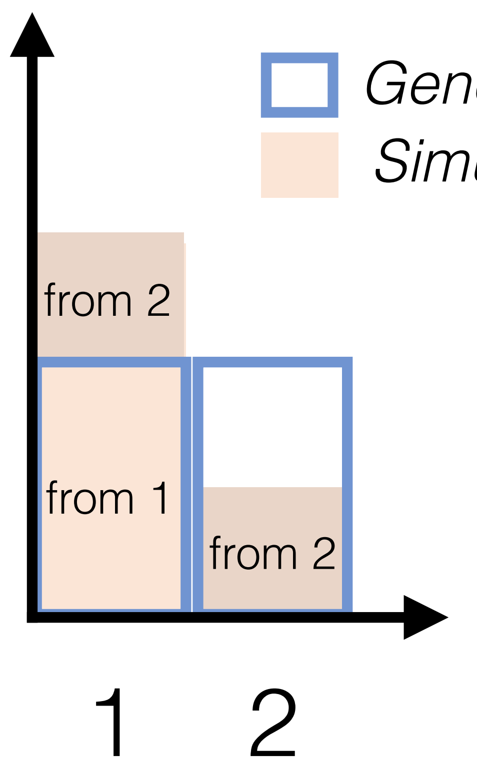


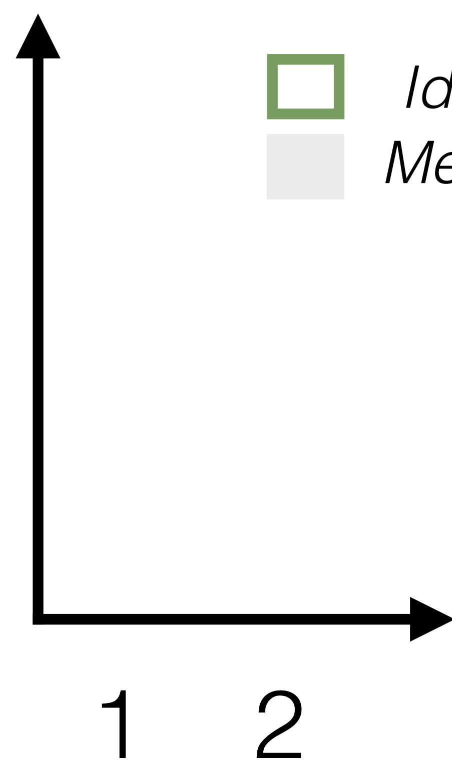


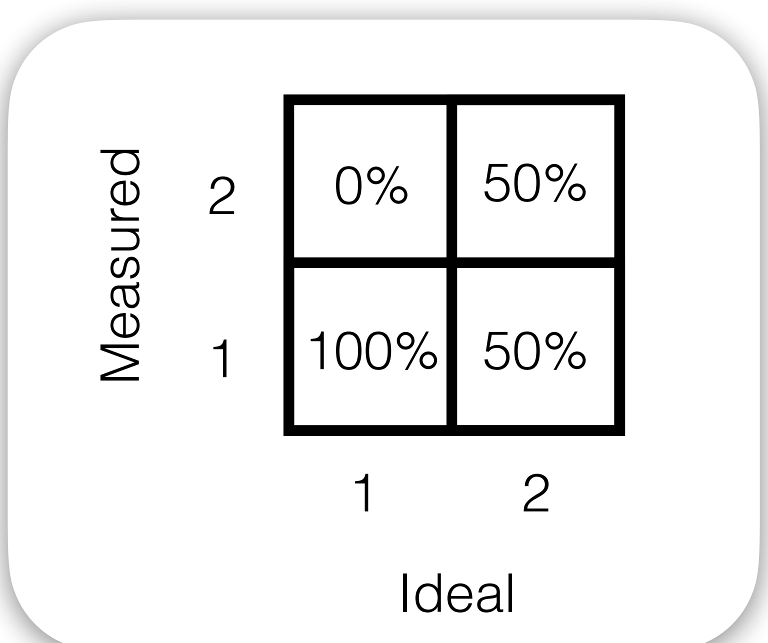


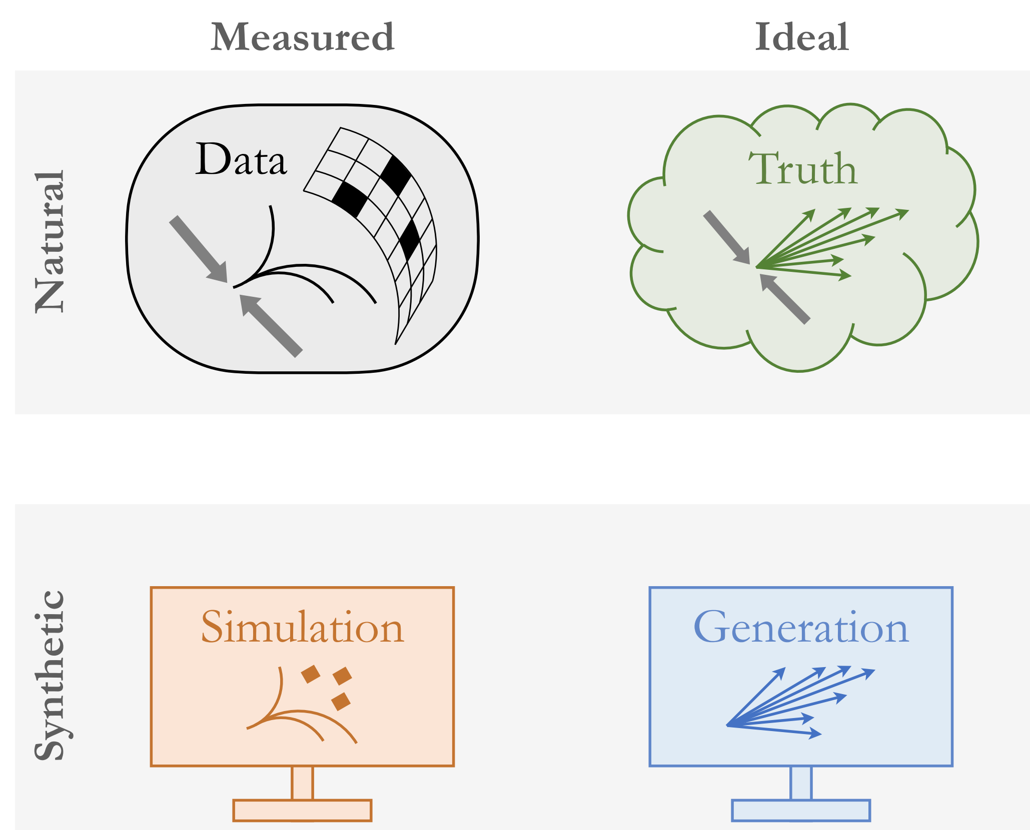
A. Andreassen, P. Komiske, E. Metodiev, BPN, J. Thaler, PRL 124 (2020) 182001

Unfold by iterating: OmniFold



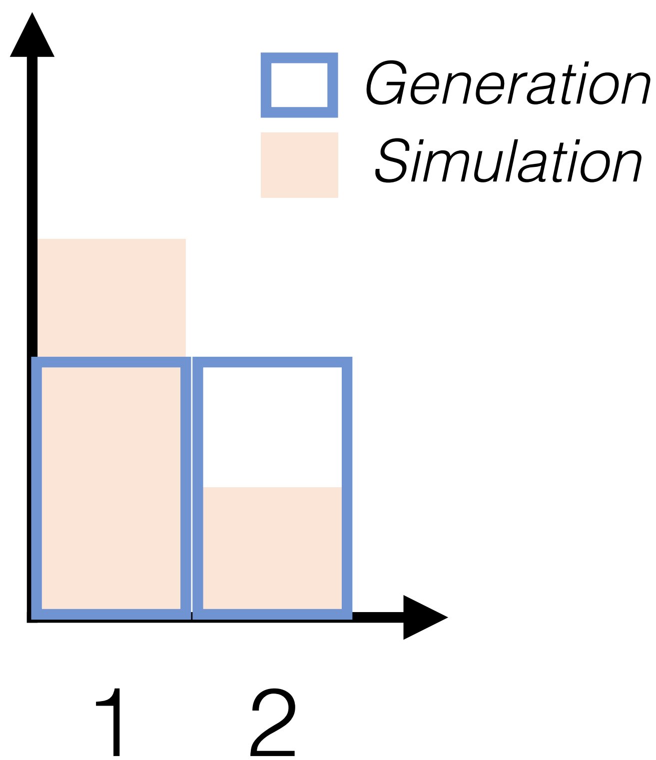


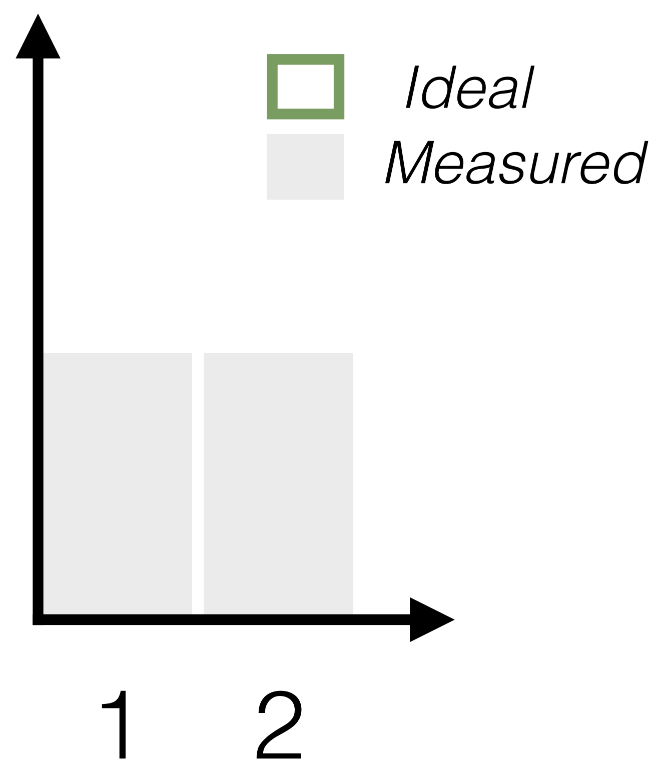




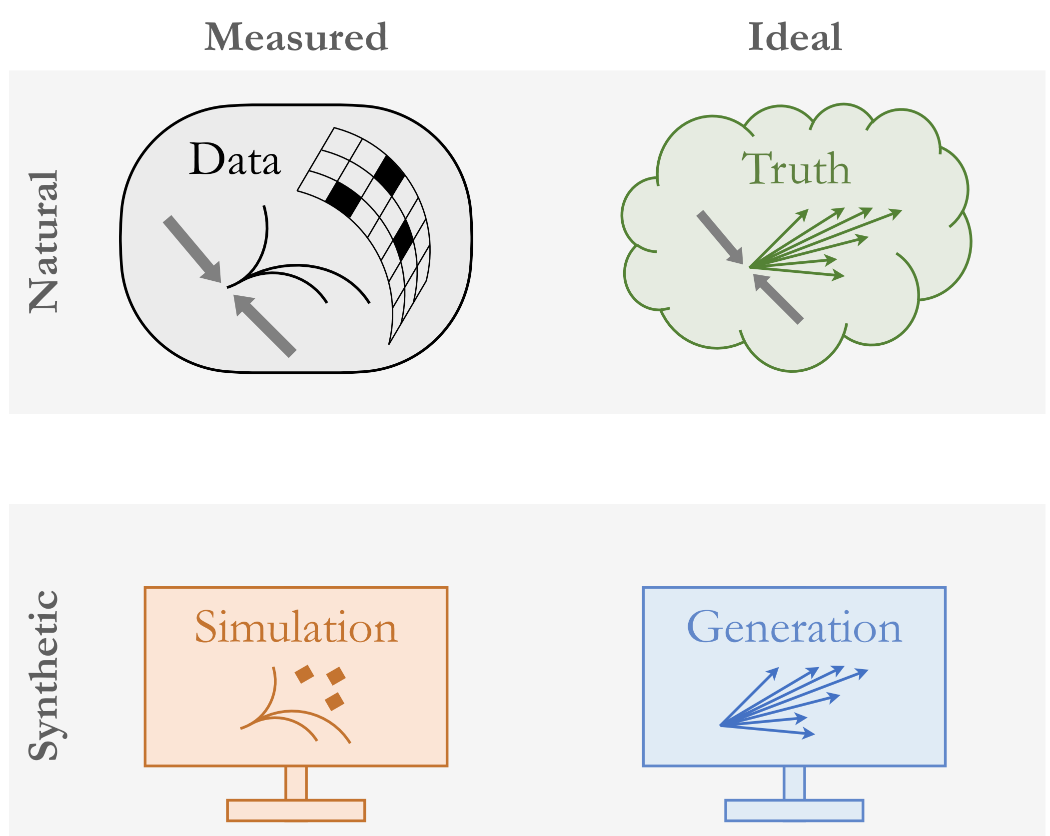
A. Andreassen, P. Komiske, E. Metodiev, BPN, J. Thaler, PRL 124 (2020) 182001

Unfold by iterating: OmniFold



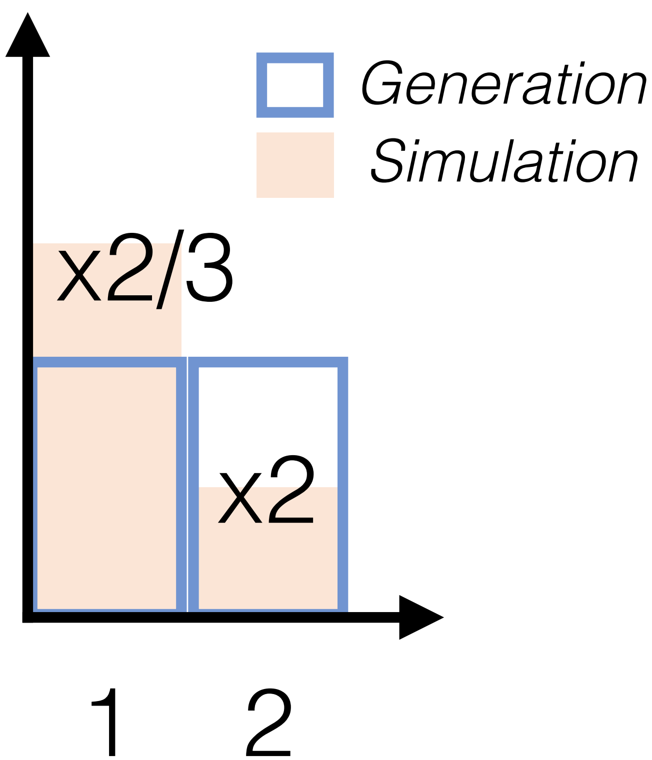


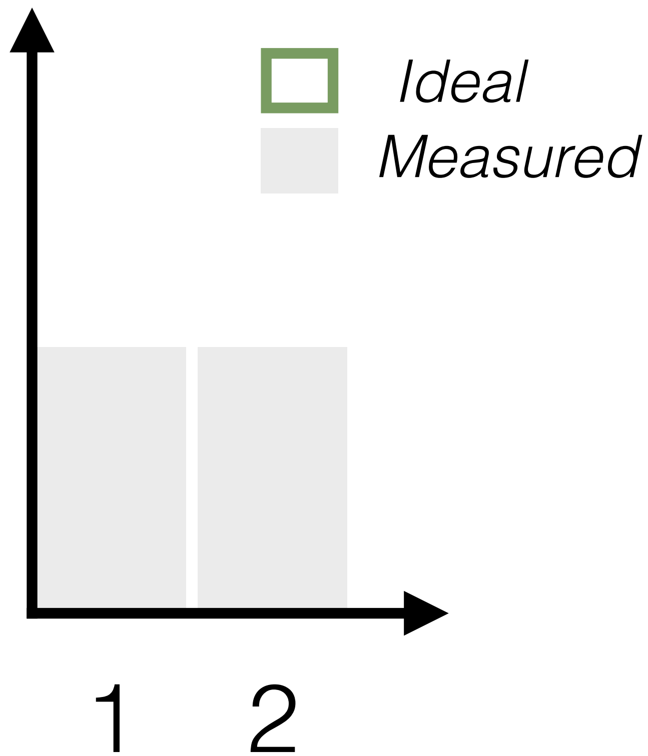
Measured \ Ideal	1	2
2	0%	50%
1	100%	50%



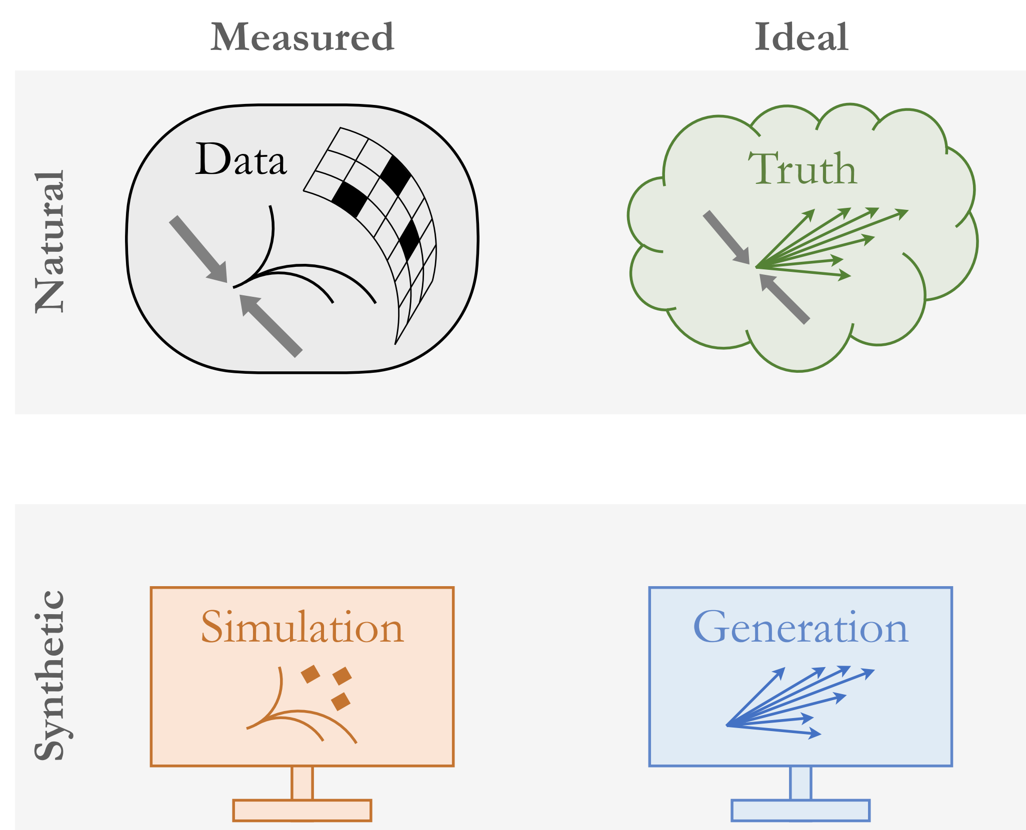
A. Andreassen, P. Komiske, E. Metodiev, BPN, J. Thaler, PRL 124 (2020) 182001

Unfold by iterating: OmniFold



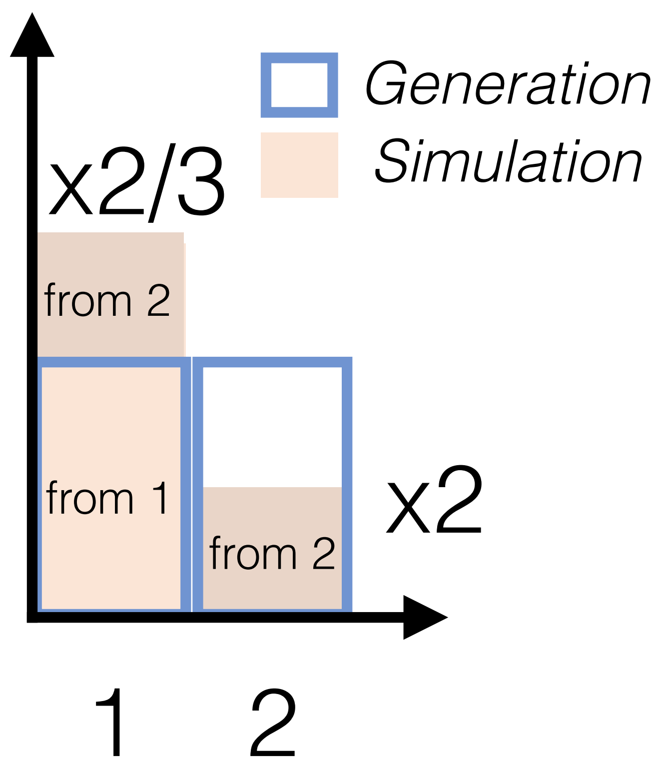


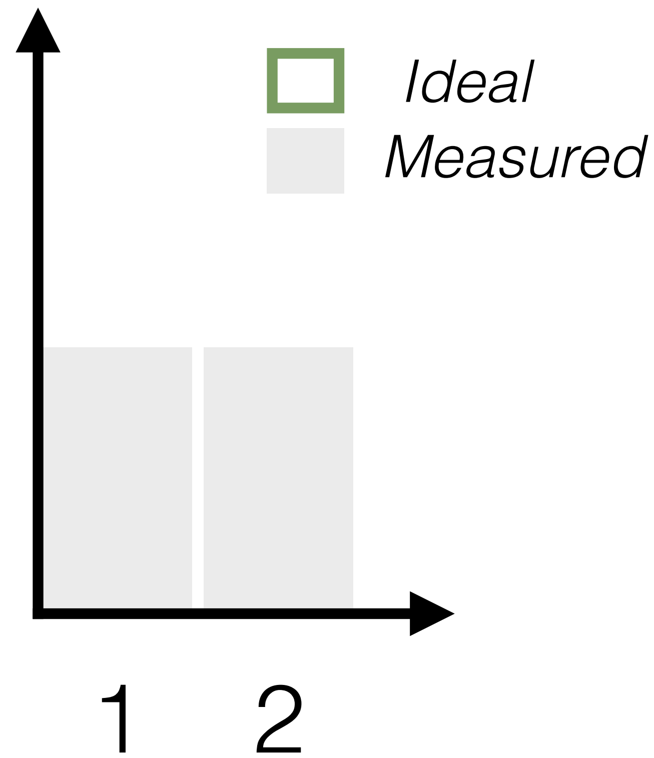
Measured \ Ideal	1	2
2	0%	50%
1	100%	50%



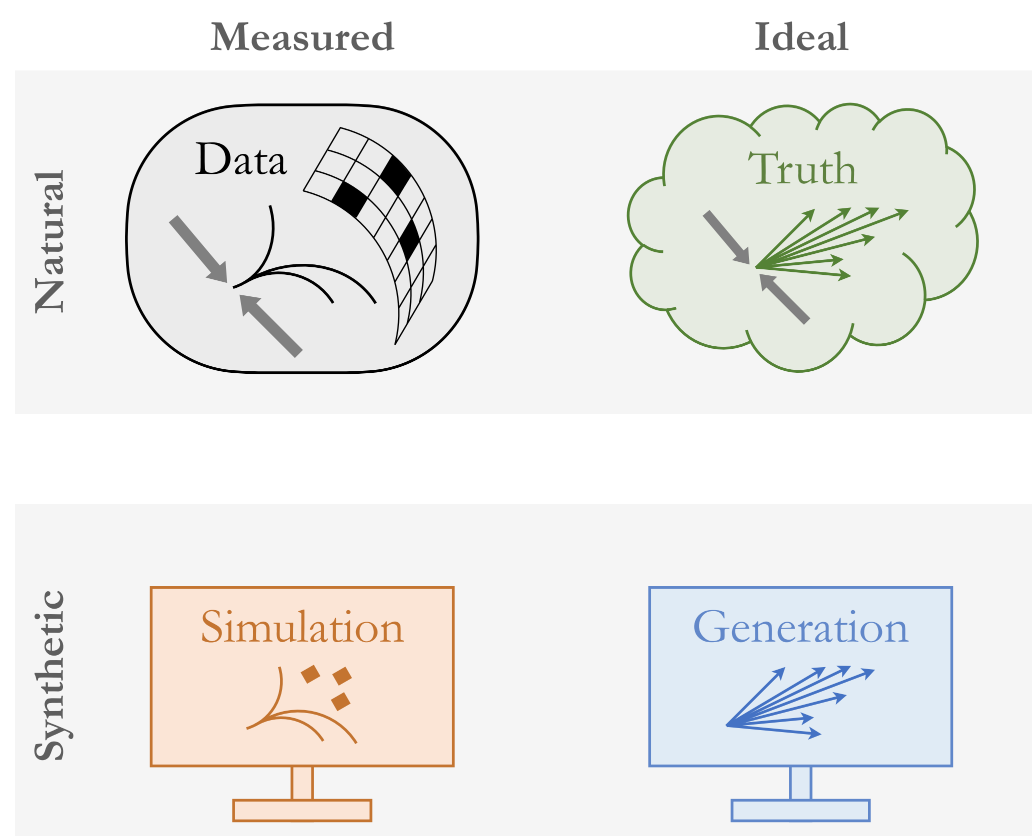
A. Andreassen, P. Komiske, E. Metodiev, BPN, J. Thaler, PRL 124 (2020) 182001

Unfold by iterating: OmniFold





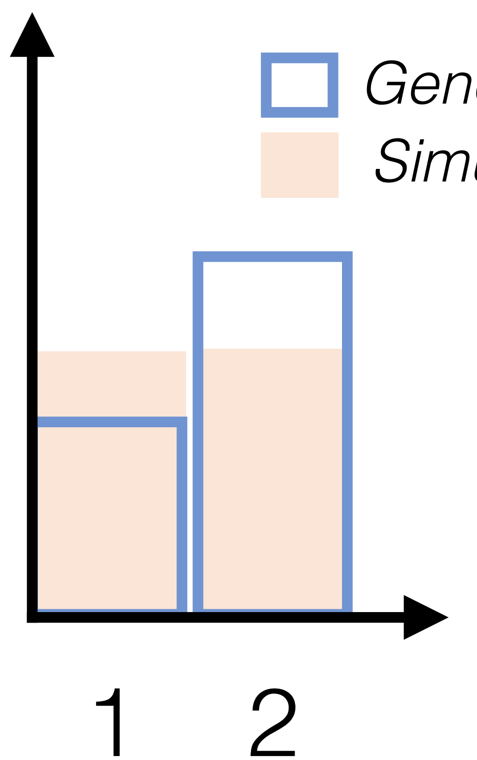
Measured \ Ideal	1	2
2	0%	50%
1	100%	50%

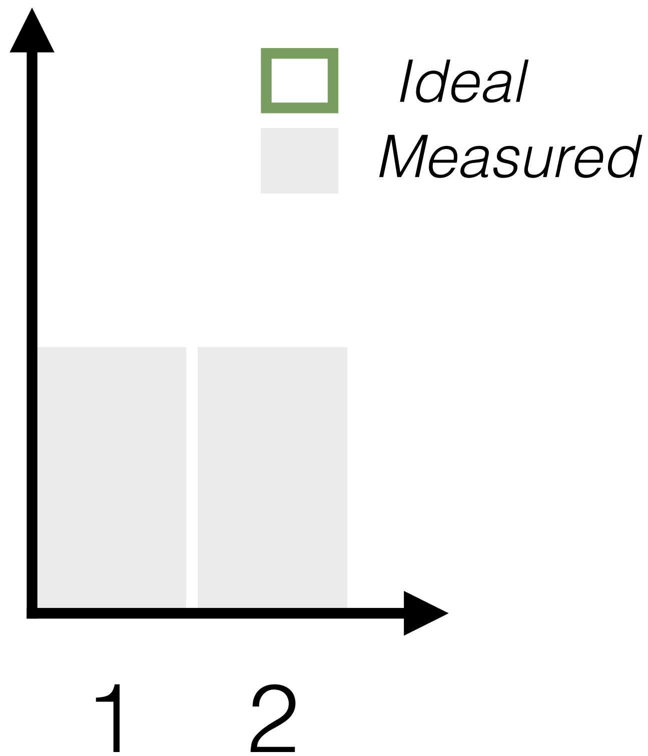


A. Andreassen, P. Komiske, E. Metodiev, BPN, J. Thaler, PRL 124 (2020) 182001

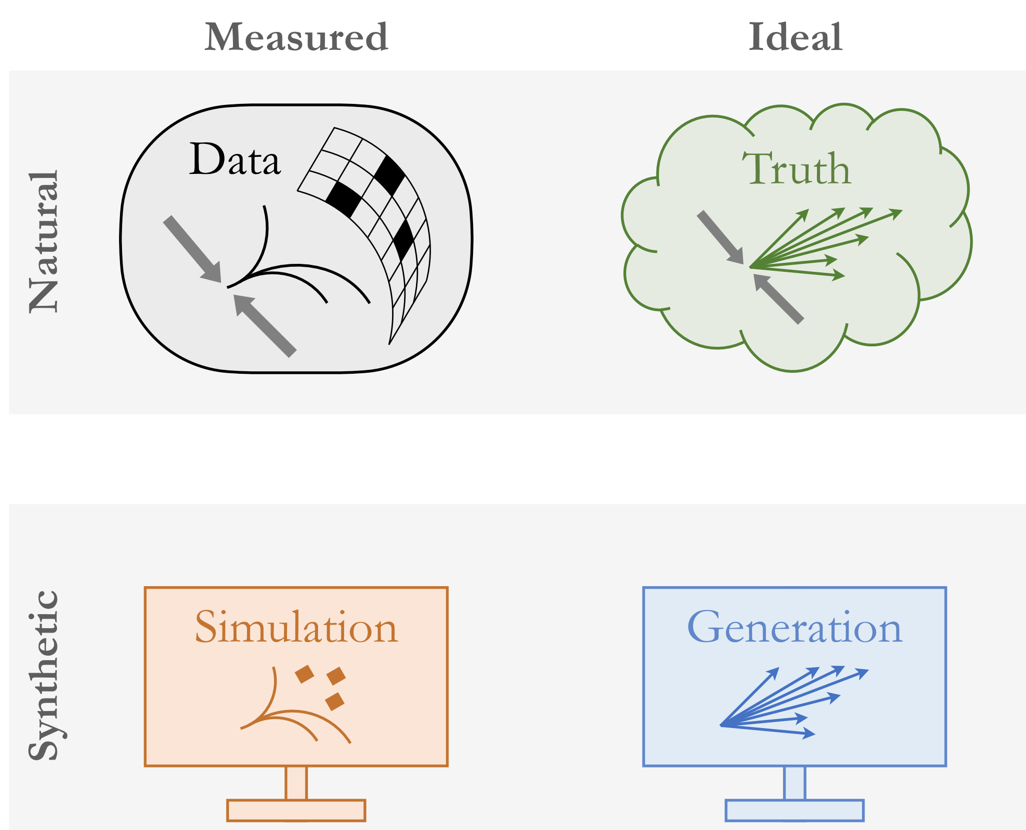
Unfold by iterating: OmniFold

After iteration 1





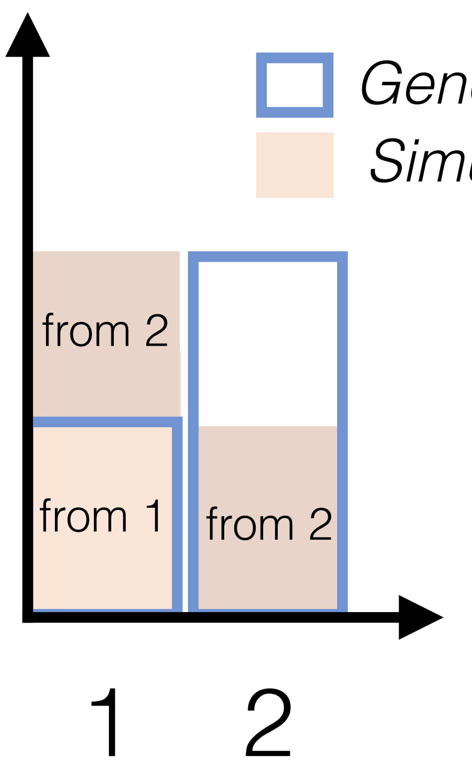
Measured \ Ideal	1	2
2	0%	50%
1	100%	50%

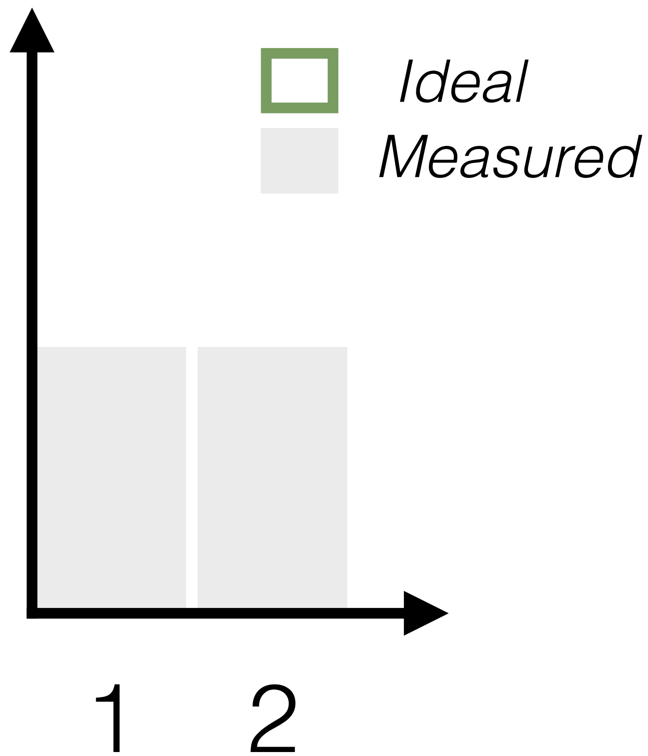


A. Andreassen, P. Komiske, E. Metodiev, BPN, J. Thaler, PRL 124 (2020) 182001

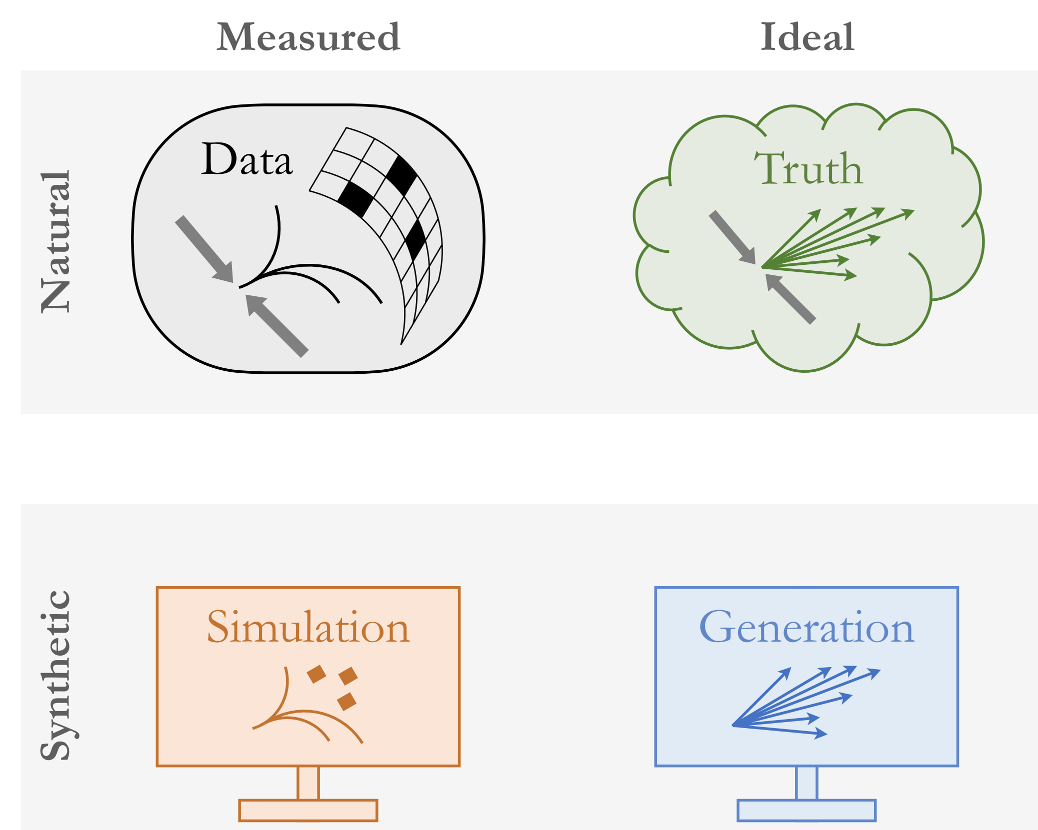
Unfold by iterating: OmniFold

After iteration 1





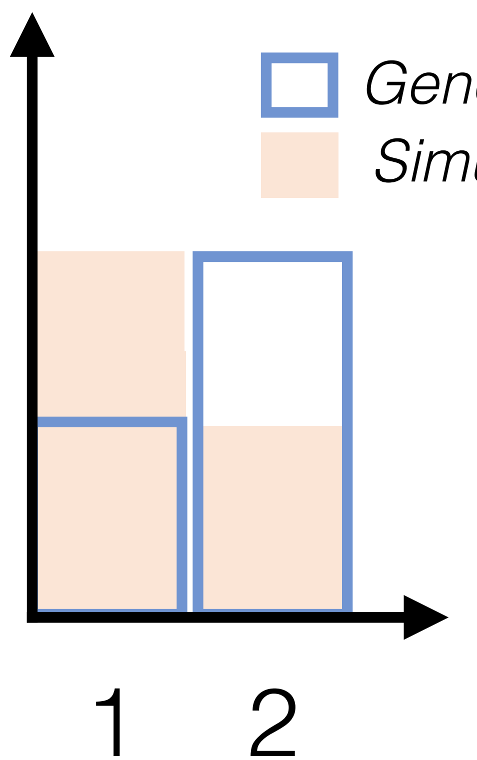
Measured		Ideal 1	Ideal 2
	2	0%	50%
	1	100%	50%

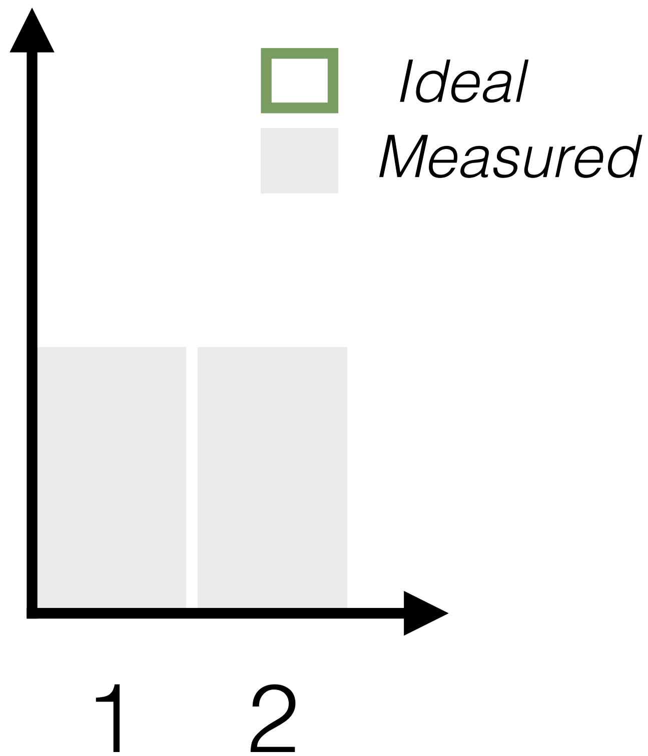


A. Andreassen, P. Komiske, E. Metodiev, BPN, J. Thaler, PRL 124 (2020) 182001

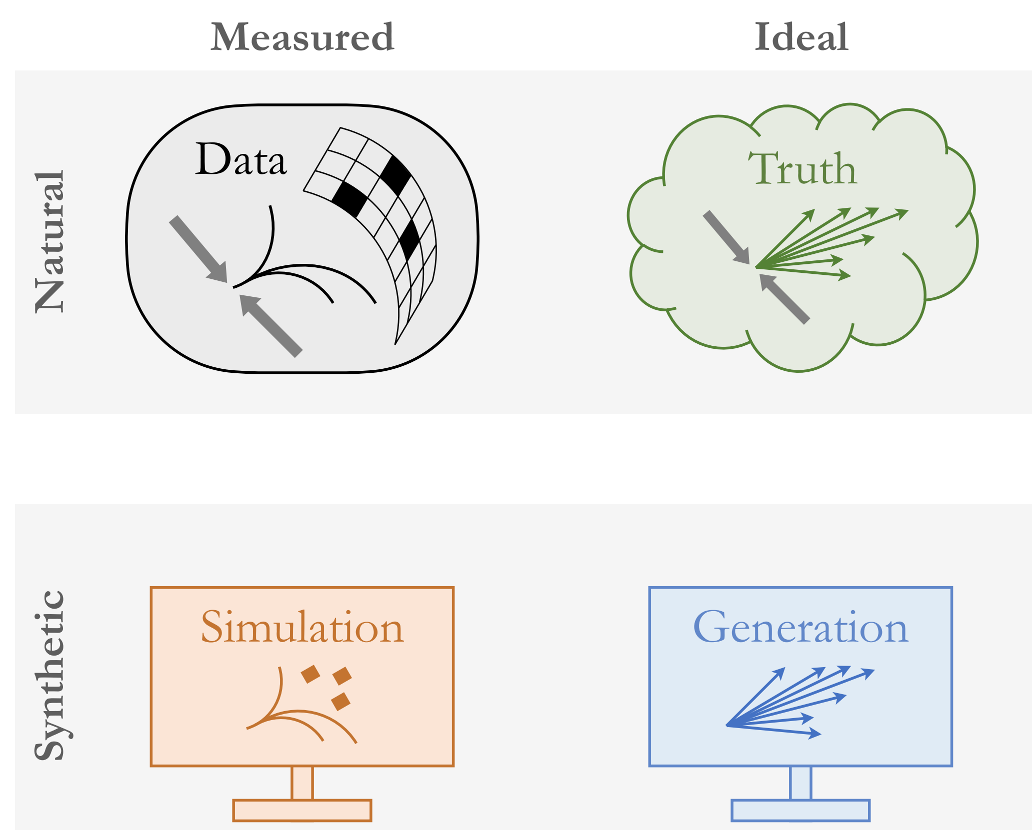
Unfold by iterating: OmniFold

After iteration 1





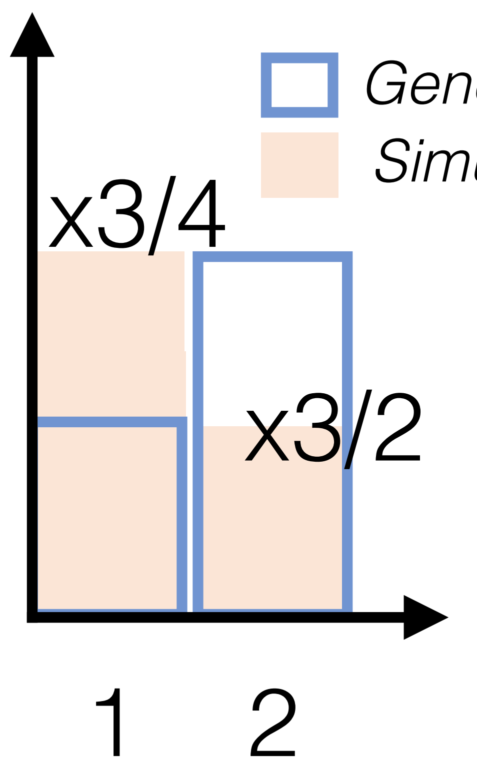
Measured \ Ideal	1	2
2	0%	50%
1	100%	50%

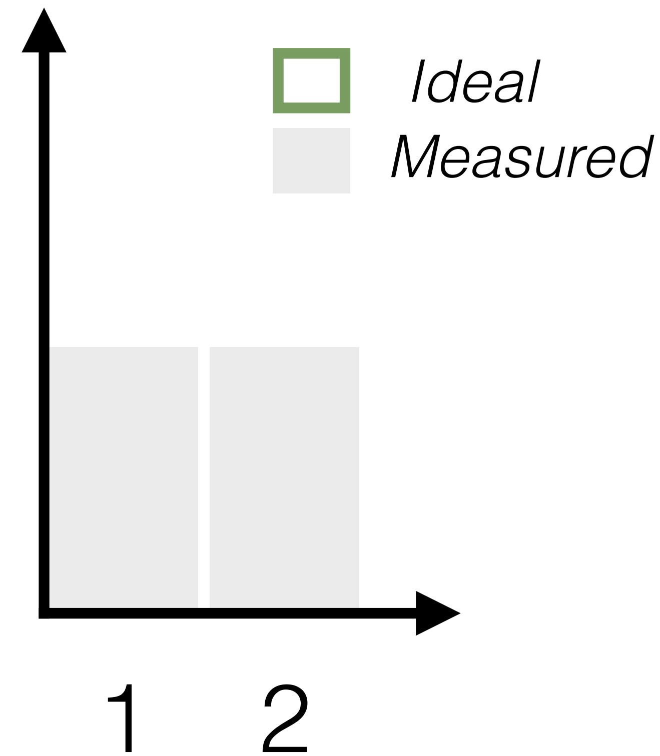


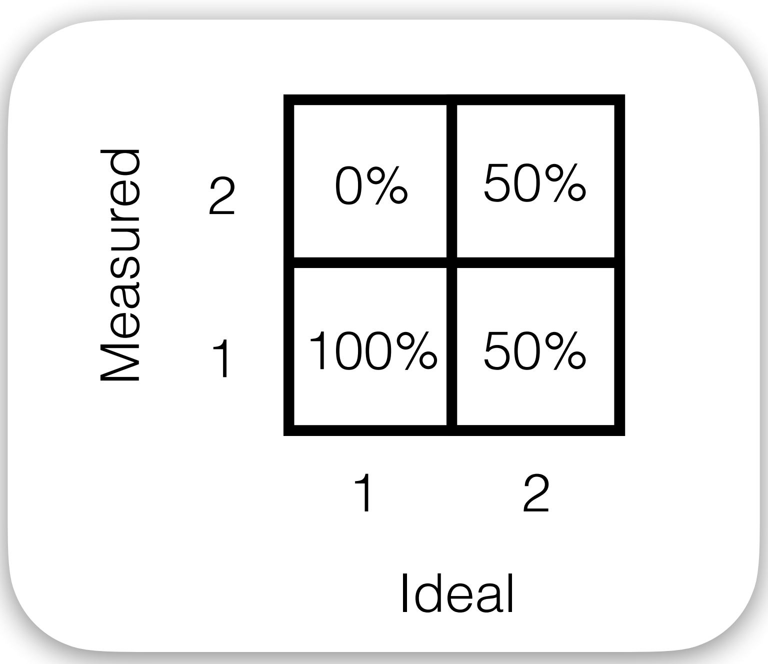
A. Andreassen, P. Komiske, E. Metodiev, BPN, J. Thaler, PRL 124 (2020) 182001

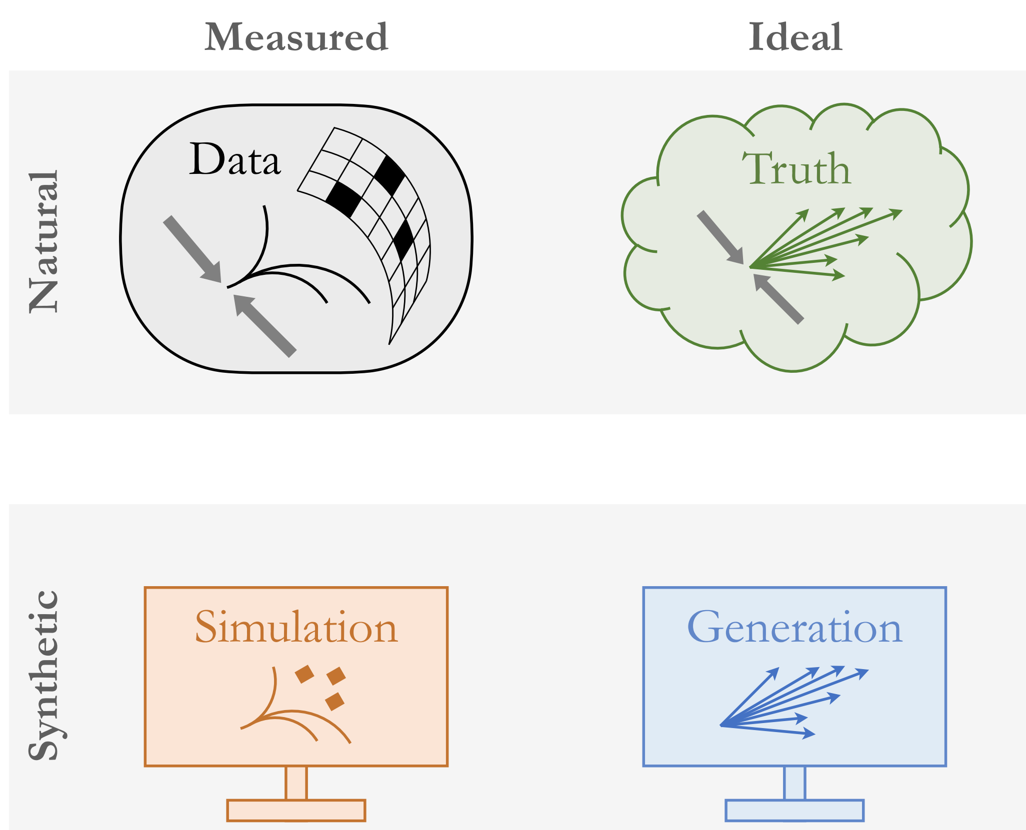
Unfold by iterating: OmniFold

After iteration 1





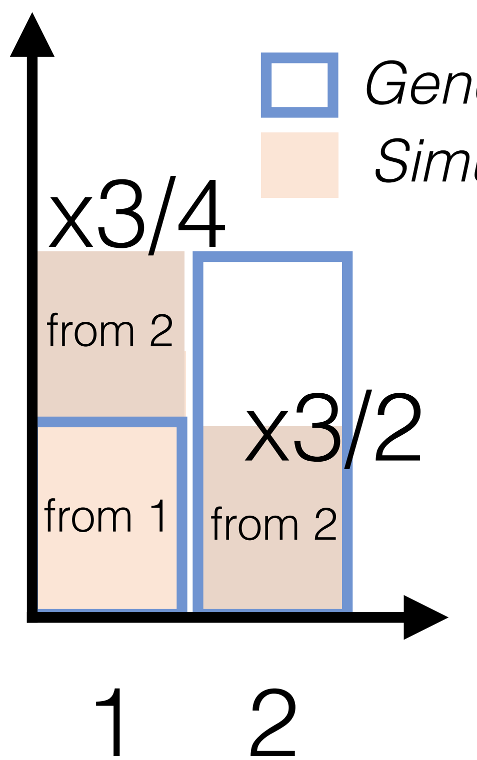


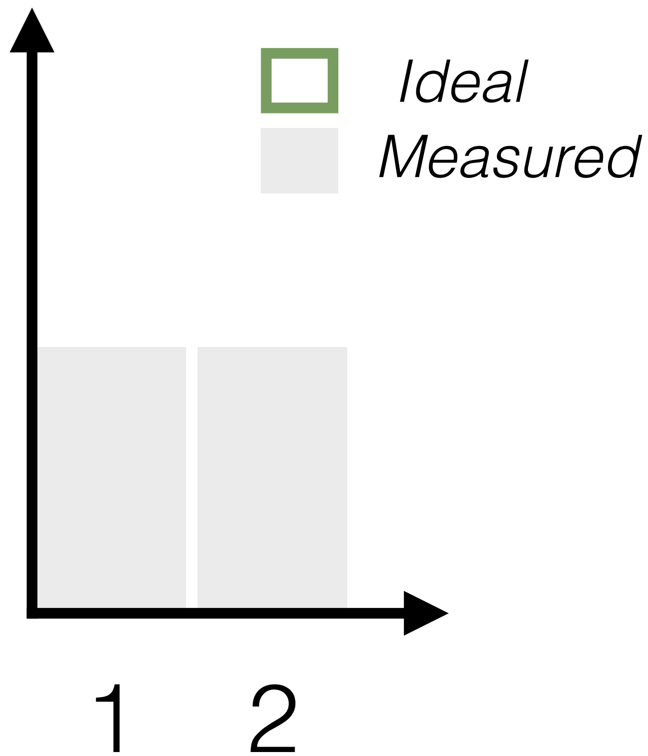


A. Andreassen, P. Komiske, E. Metodiev,
BPN, J. Thaler, PRL 124 (2020) 182001

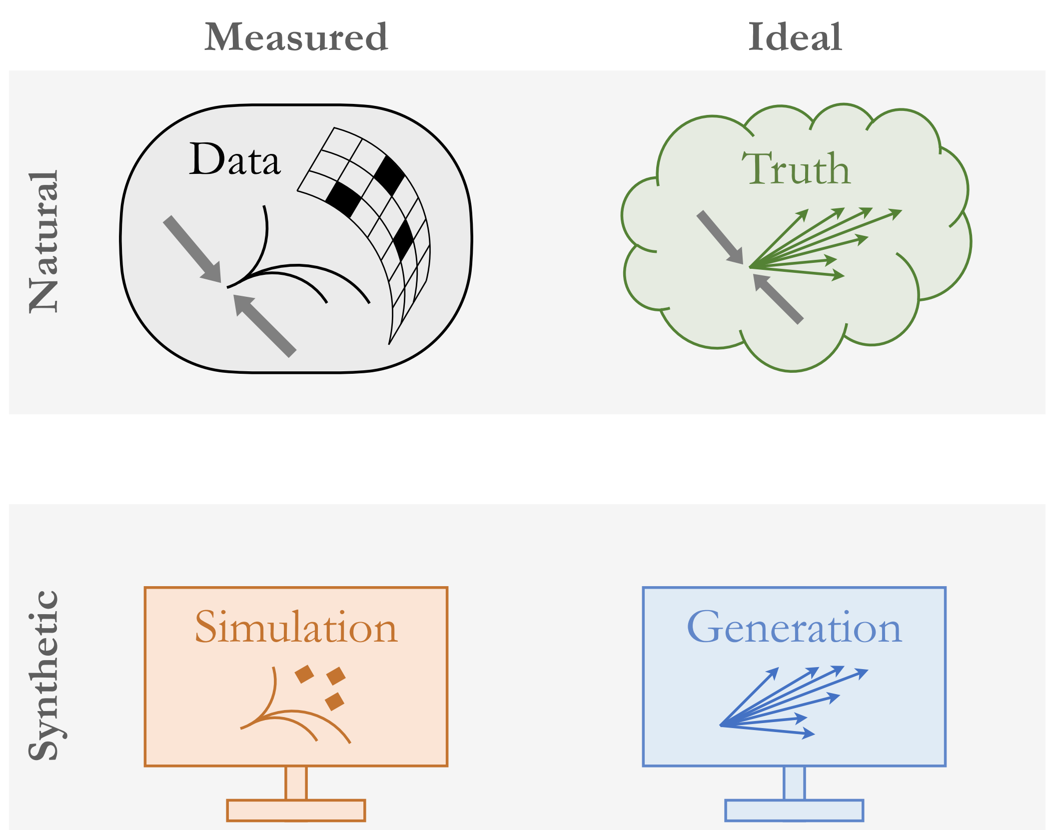
Unfold by iterating: OmniFold

After iteration 1





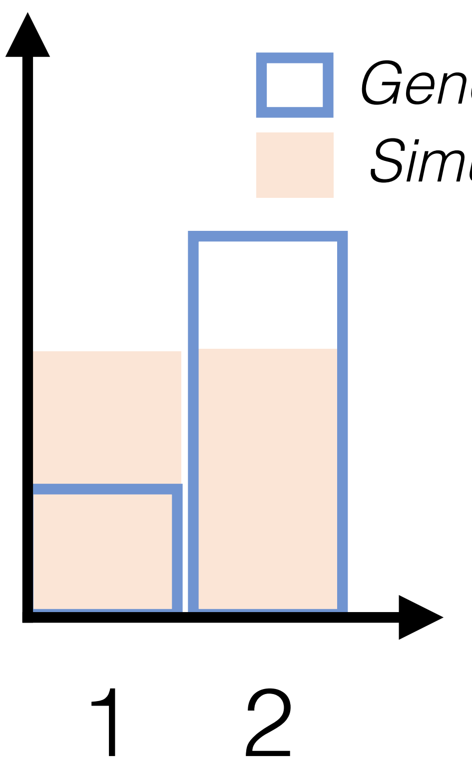
Measured \ Ideal	1	2
2	0%	50%
1	100%	50%

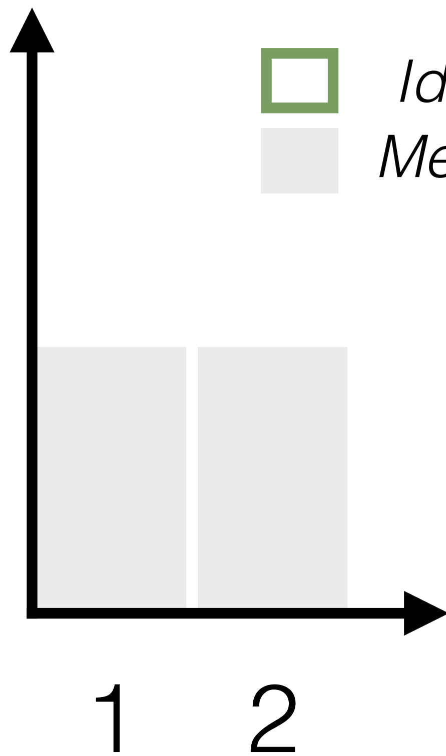


A. Andreassen, P. Komiske, E. Metodiev, BPN, J. Thaler, PRL 124 (2020) 182001

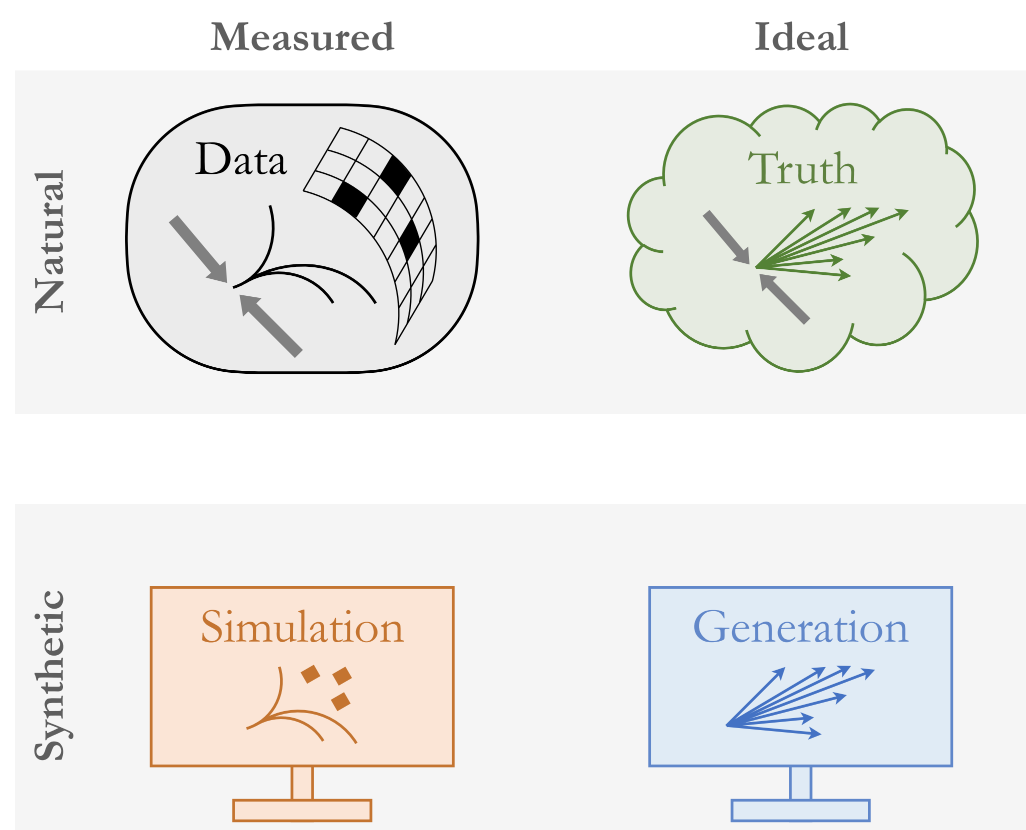
Unfold by iterating: OmniFold

After iteration 2





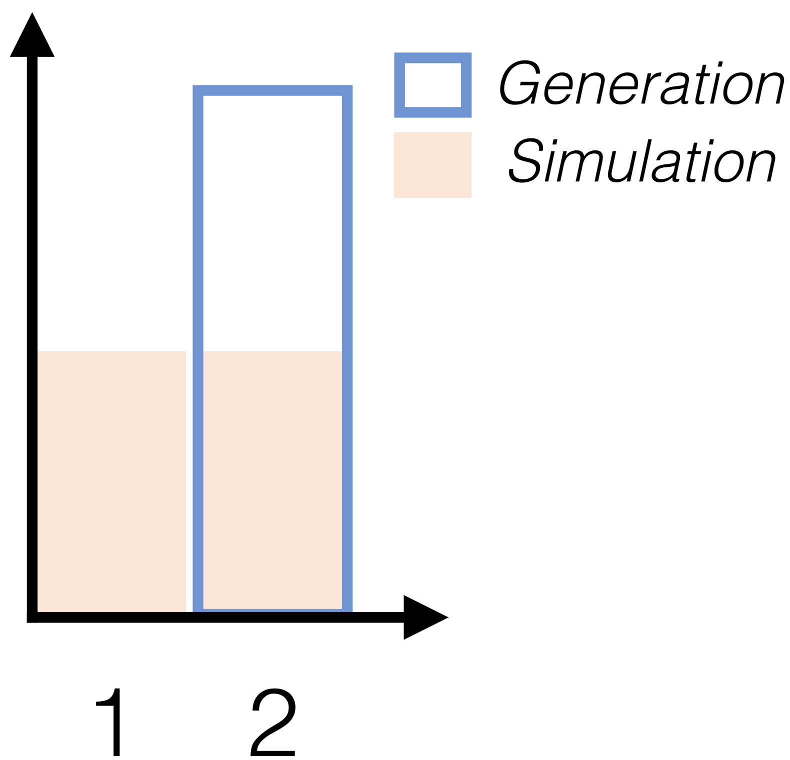
Measured \ Ideal	1	2
2	0%	50%
1	100%	50%

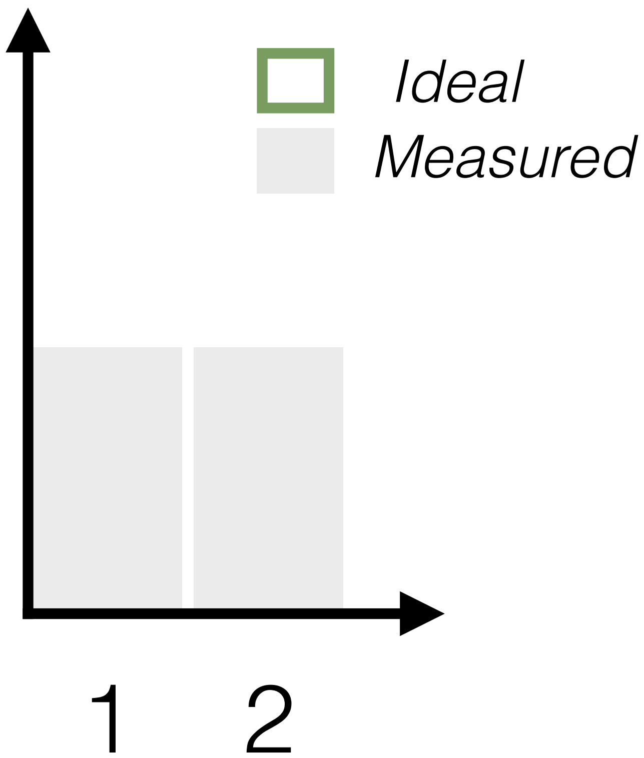


A. Andreassen, P. Komiske, E. Metodiev, BPN, J. Thaler, PRL 124 (2020) 182001

Unfold by iterating: OmniFold

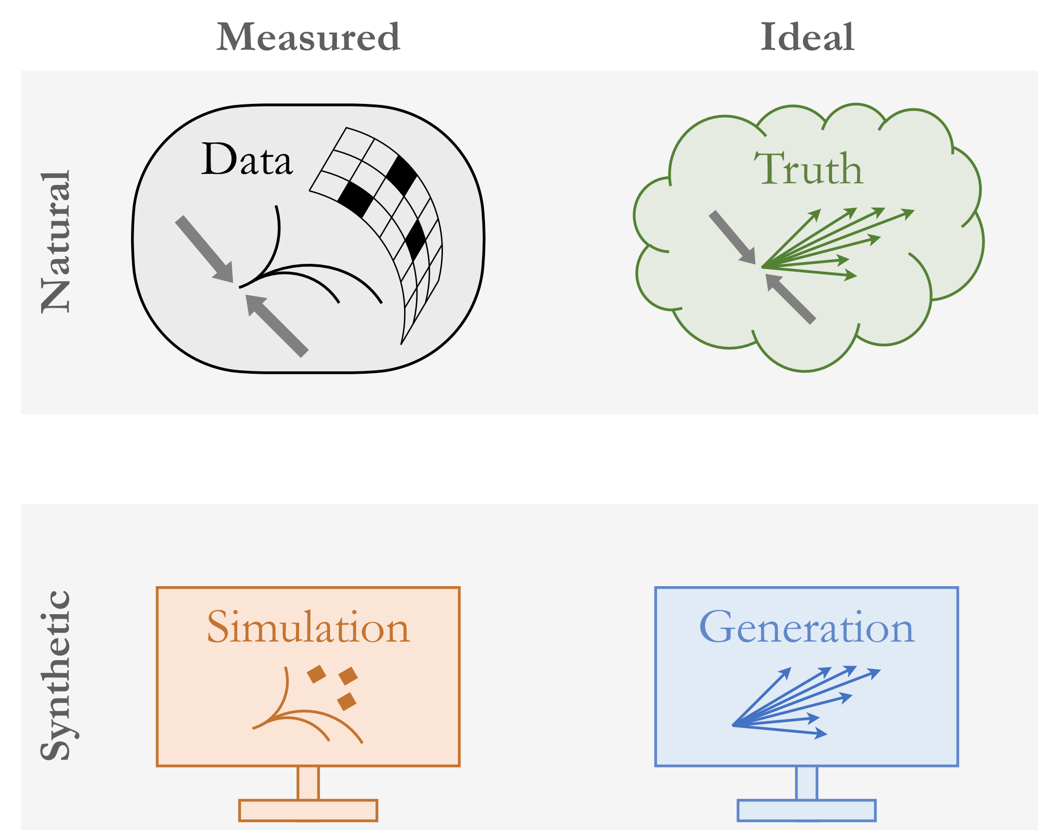
After iteration ∞





N.B. if you just apply p(ideal | measured), *you would have gotten the wrong answer!*

Measured \ Ideal	1	2
2	0%	50%
1	100%	50%



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