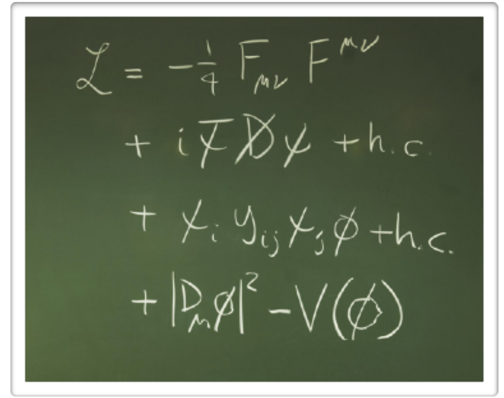
ML in Particle Physics

Prof. Gregor Kasieczka
Email: gregor.kasieczka@uni-hamburg.de
Twitter/X: @GregorKasieczka
PHYSTAT — Statistics meets ML, Imperial
10.9.2024

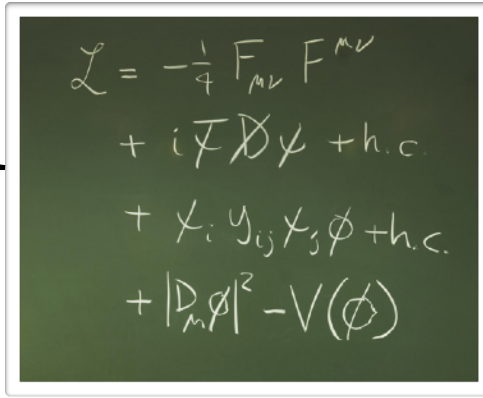
CLUSTER OF EXCELLENCE
QUANTUM UNIVERSE
KISS
FSP CMS
PUNCH 4NFDI
DASHH
GEFÖRDERT VOM
Bundesministerium für Bildung und Forschung
Universität Hamburg
DER FORSCHUNG | DER LEHRE | DER BILDUNG
CDCS
CENTER FOR DATA AND COMPUTING IN NATURAL SCIENCES
PIER
Partnership of Universität Hamburg and DESY
Emmy Noether-Programm
Deutsche Forschungsgemeinschaft
DFG

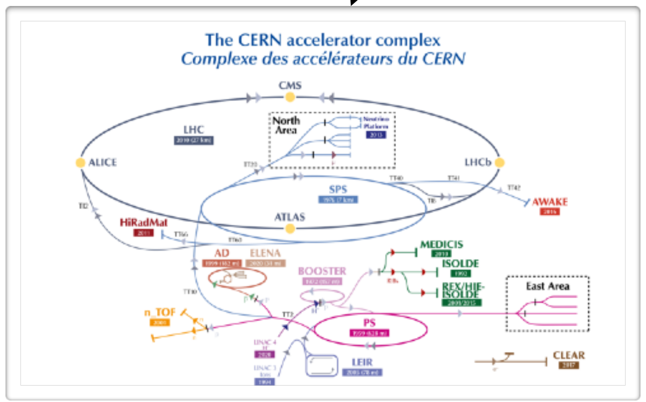
To consult the statistician after an experiment is finished is often merely to ask him to conduct a post mortem examination. He can perhaps say what the experiment died of.

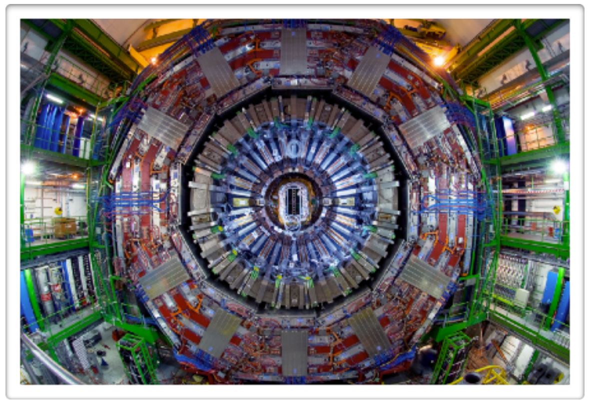
Ronald Fisher



First principle, quantum theoretical model

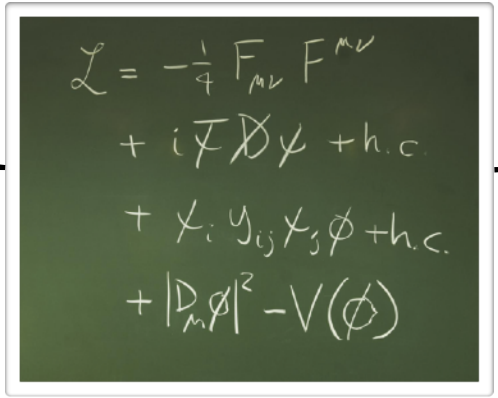


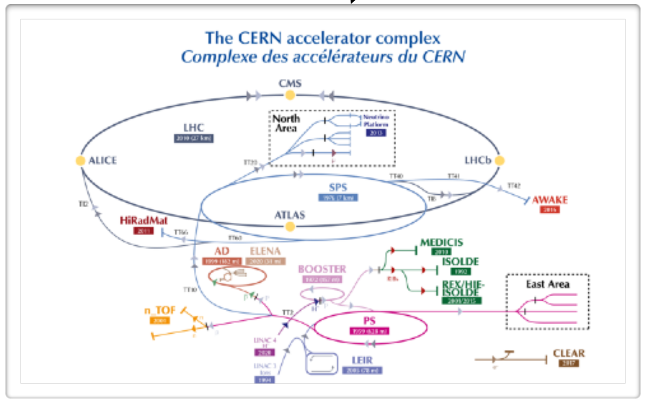


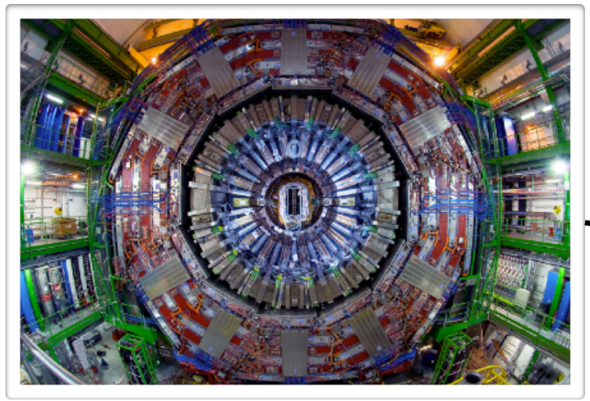


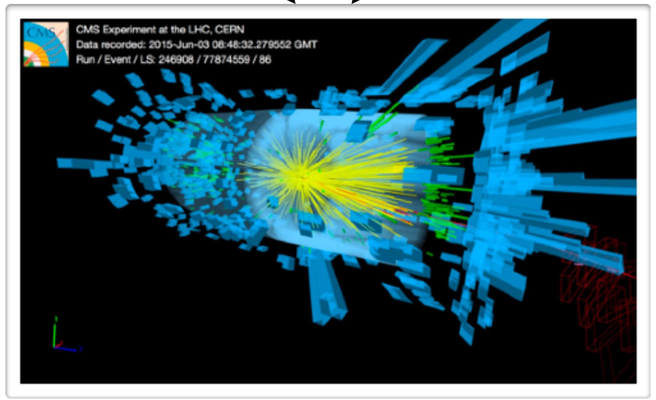
Colliders with
40 million events/second,
detectors with
100 million read-outs,

and massive theory-driven
simulation codes

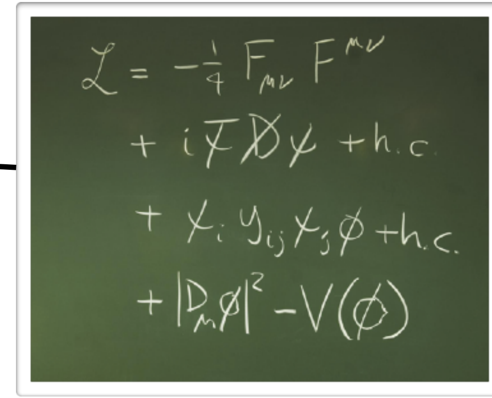


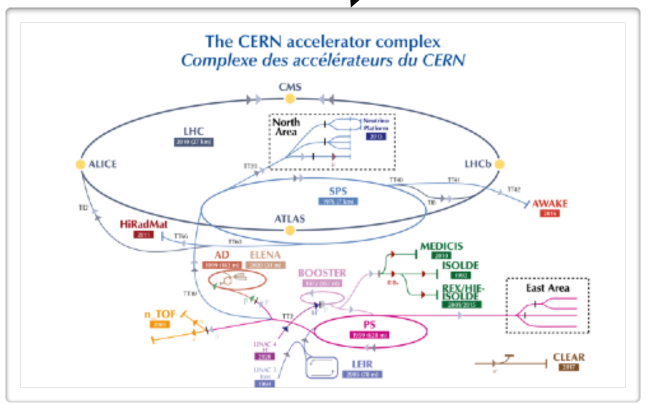


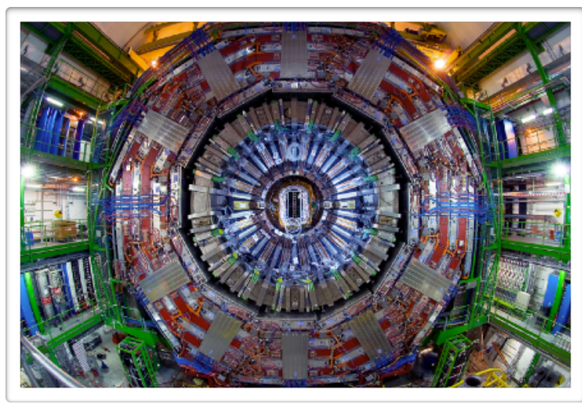


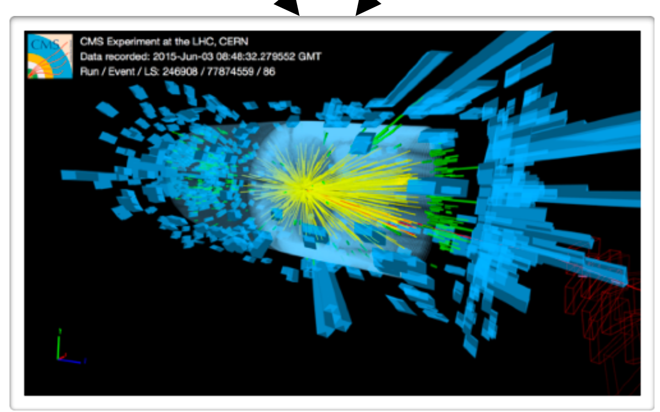


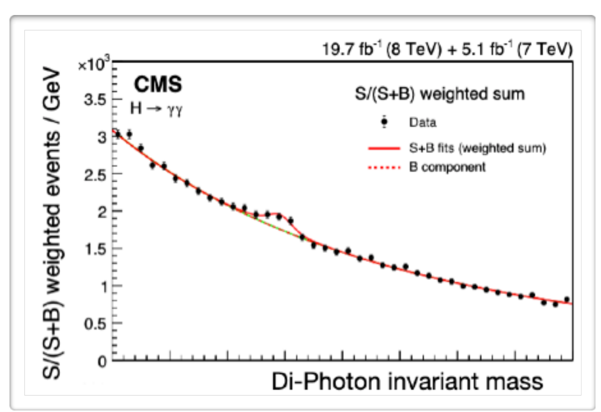
Complex reconstruction chain to turn low-level read-outs into high-level physics objects



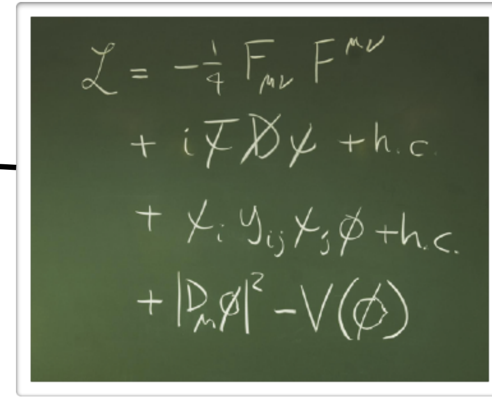


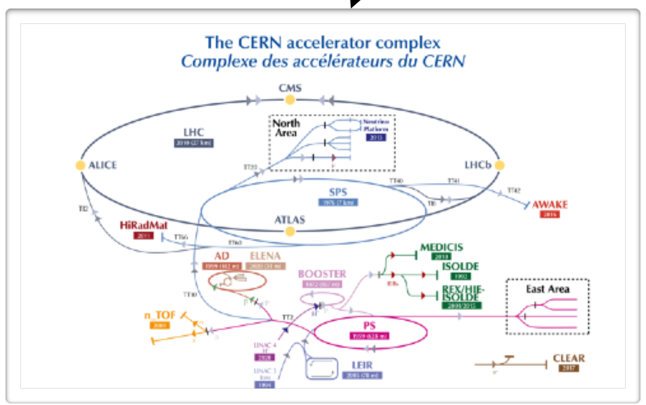


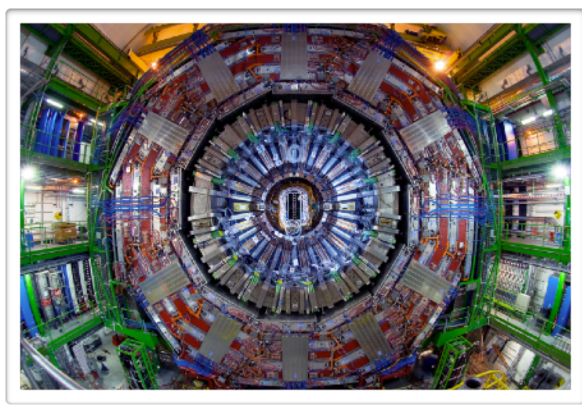


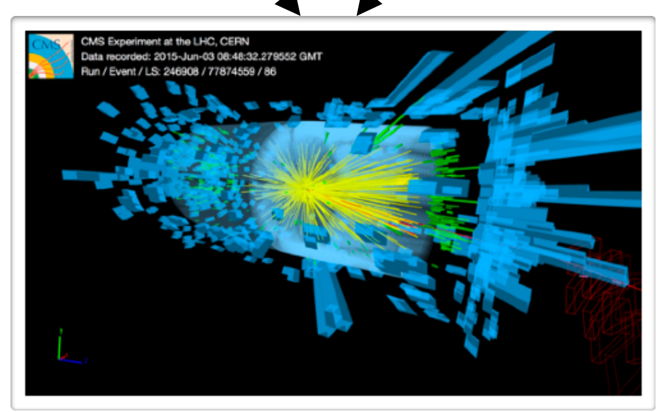


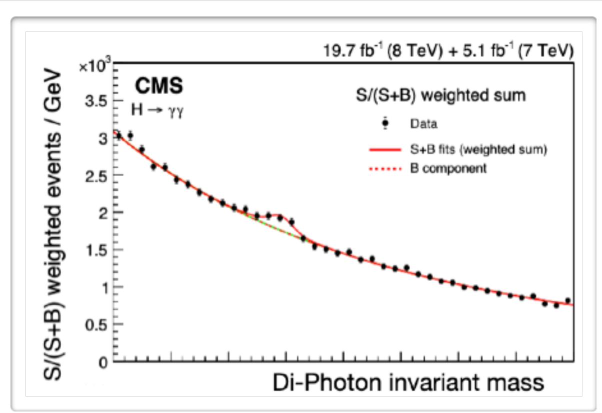
Aggregate events, define phase space regions of interest & relevant summary statistics followed by sophisticated final statistical analysis



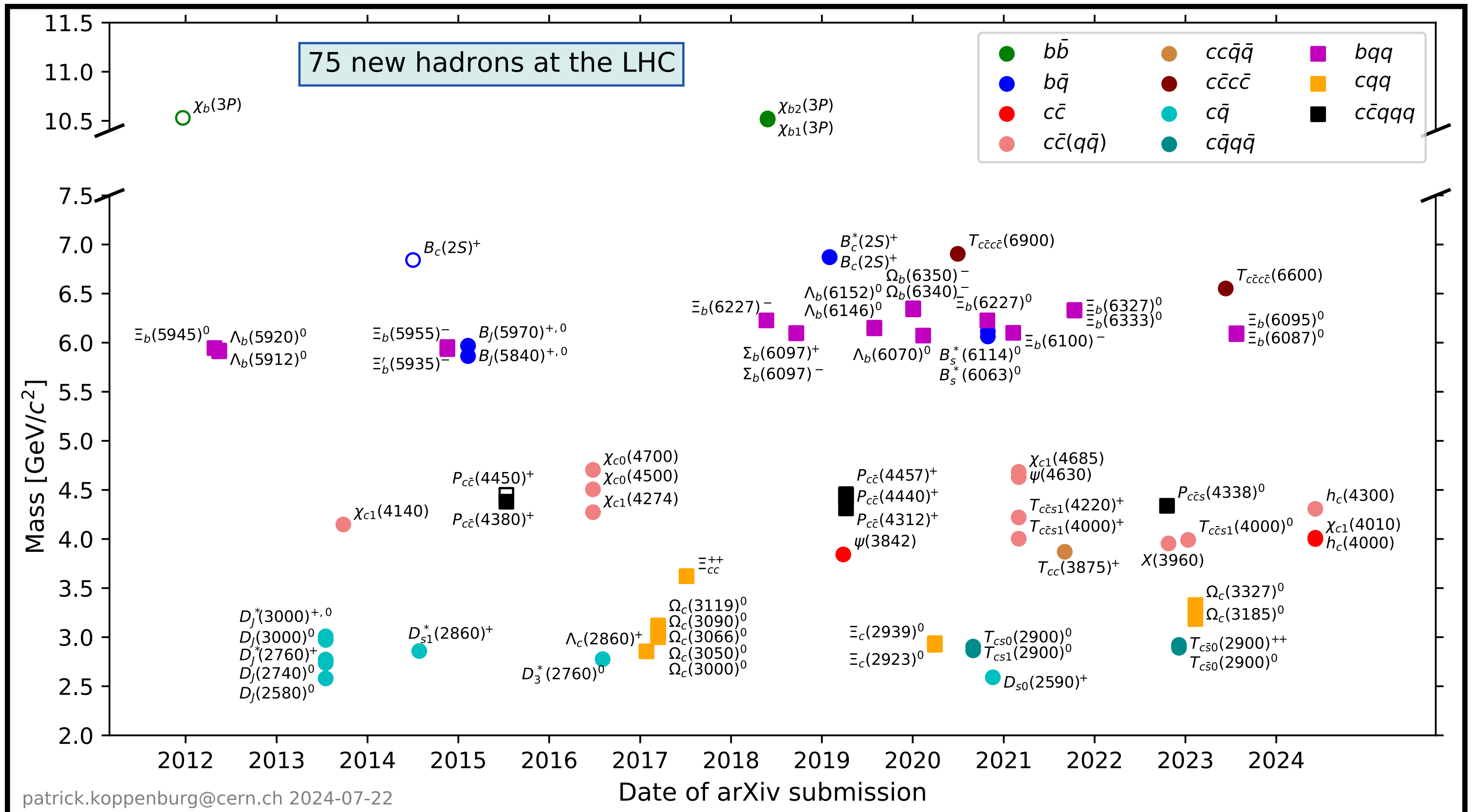


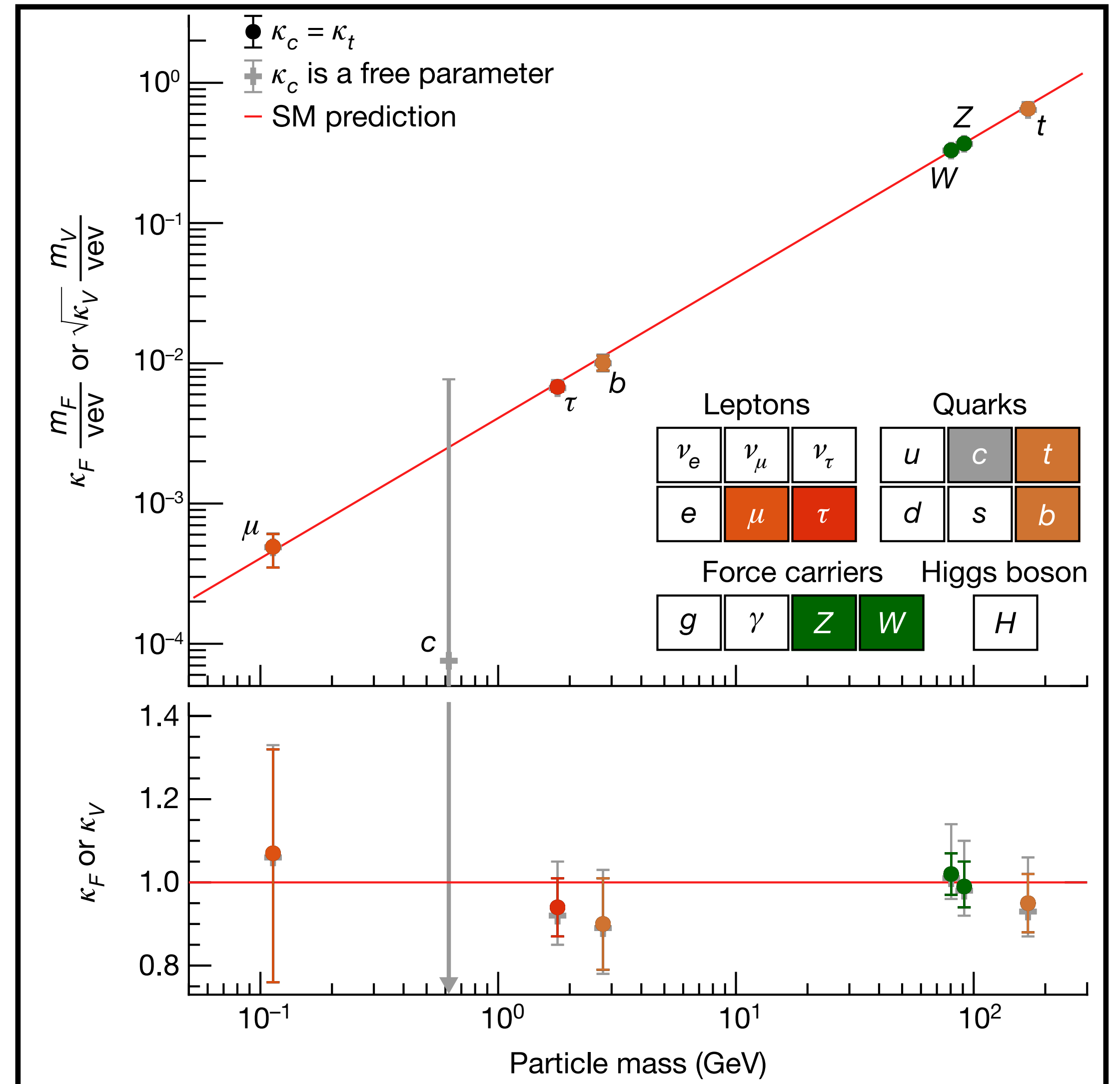






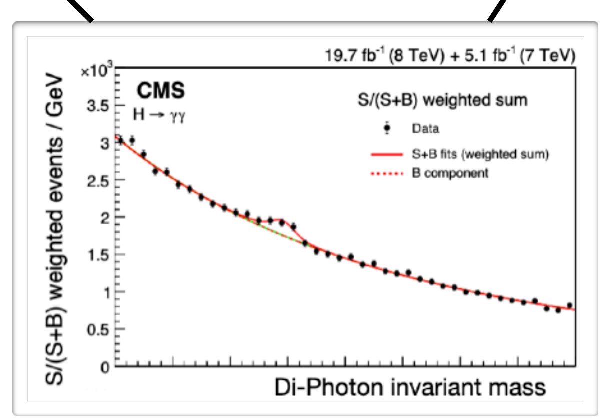
Sophisticated final statistical analysis



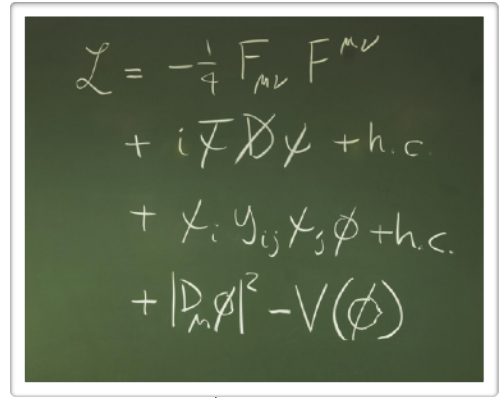


Discover new particles

Measure fundamental properties of nature

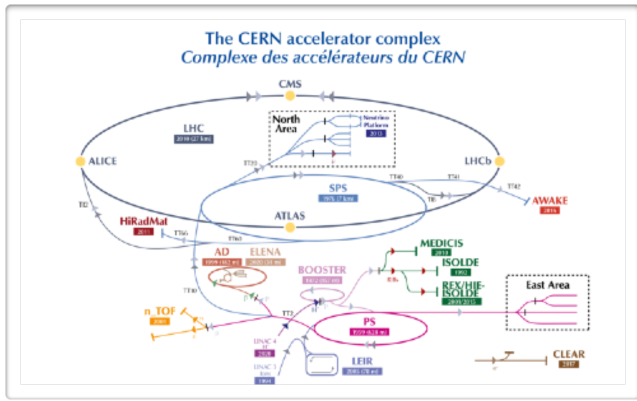


Sophisticated final statistical analysis



Inference

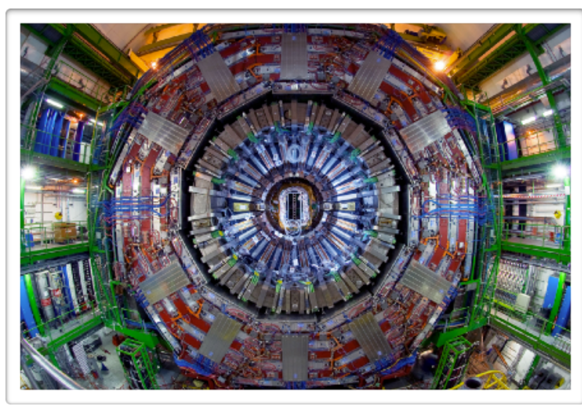
Experiment Design



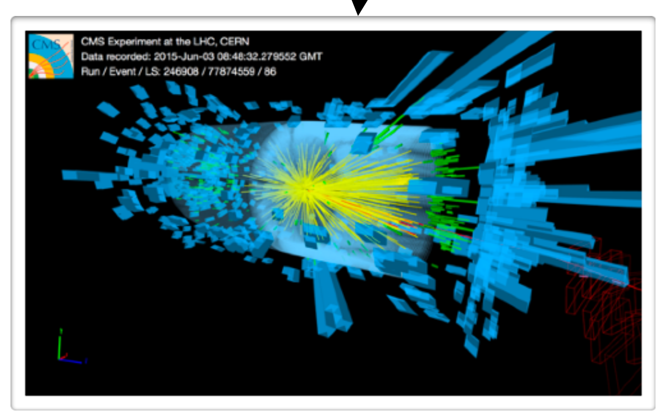
AI

Simulation

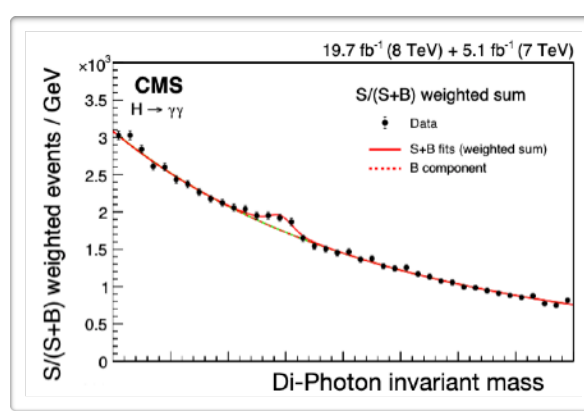
Triggers



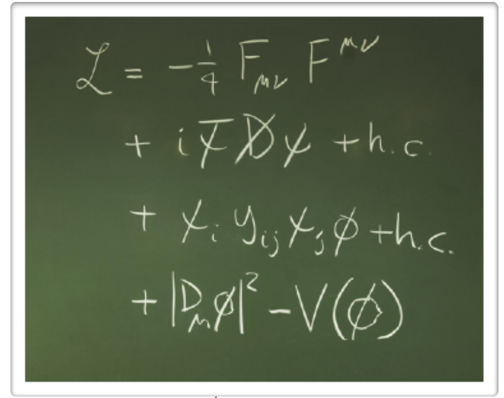
Tagging
Reconstruction



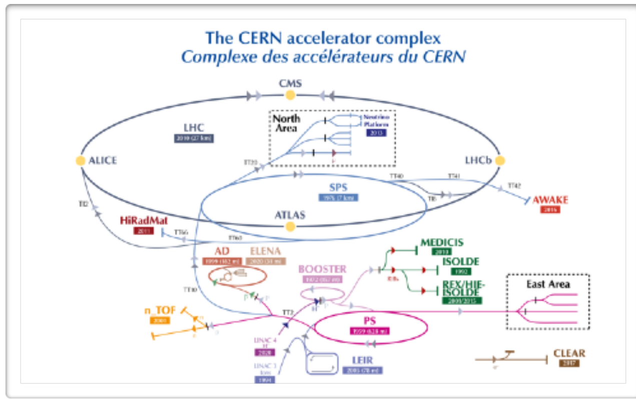
Unfolding
Anomaly Detection



AI/ML models are studied for all aspects of experimental particle physics



Inference

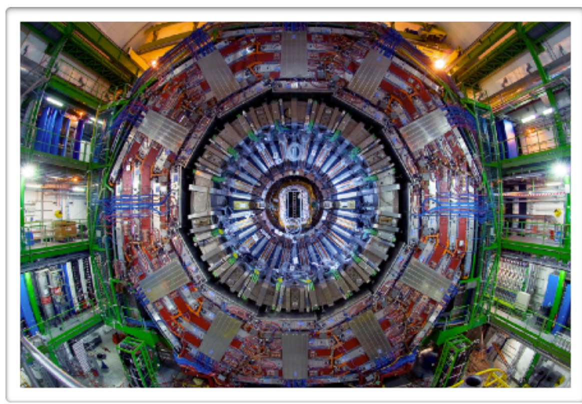


Experiment Design

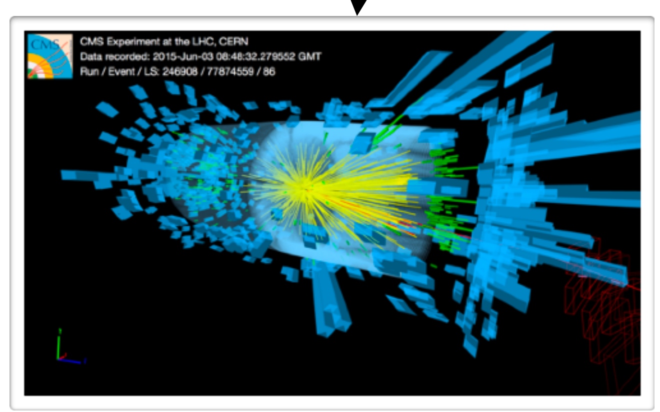
AI

Simulation

Triggers

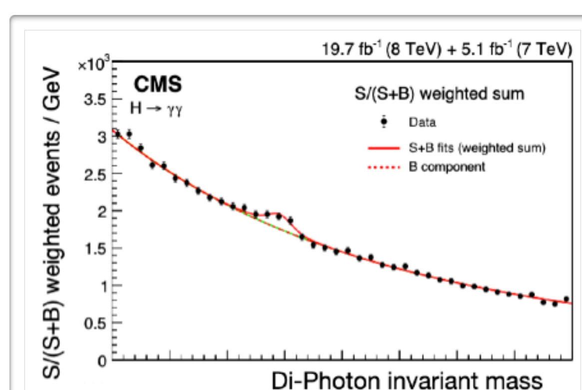


Tagging
Reconstruction

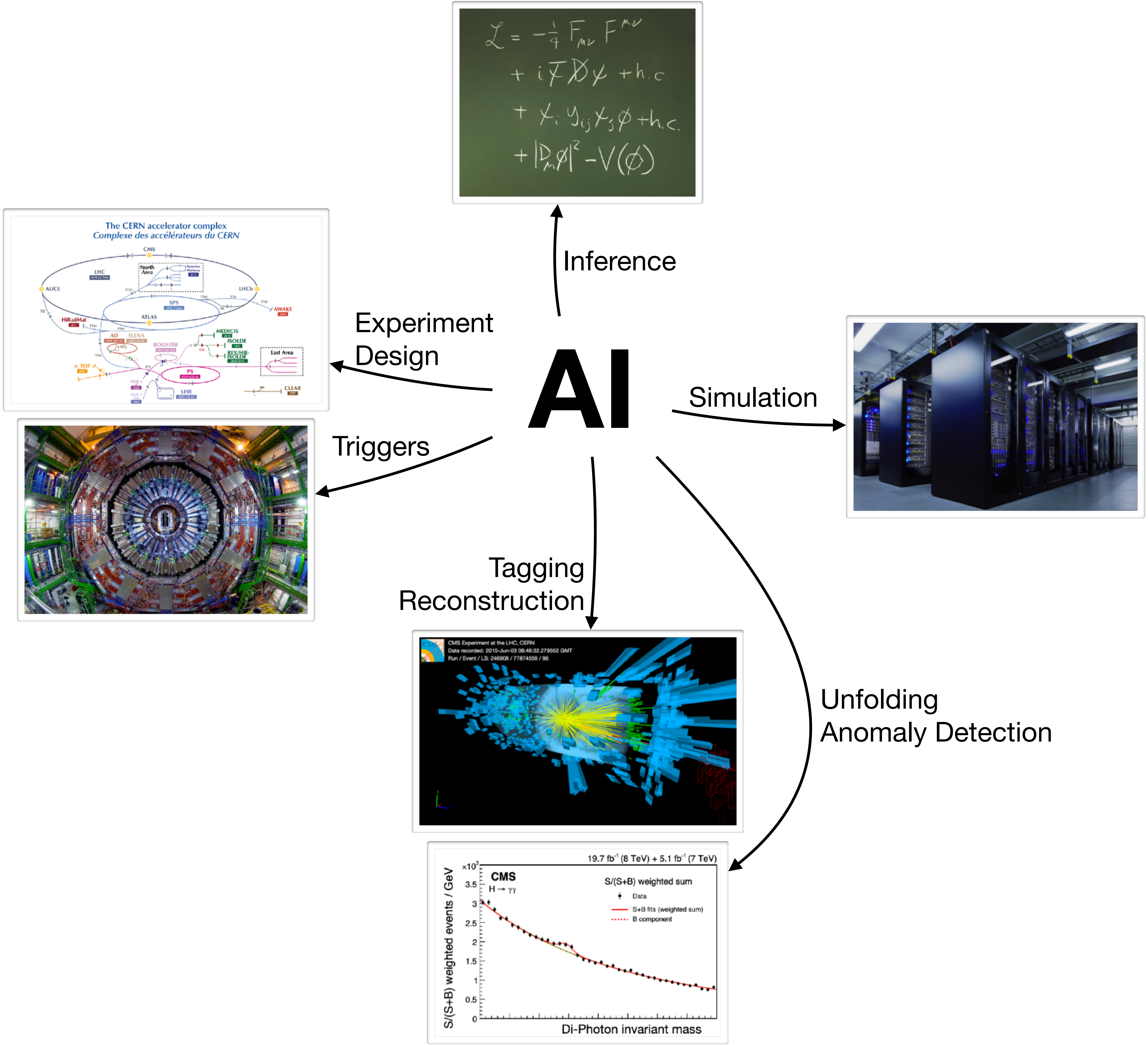


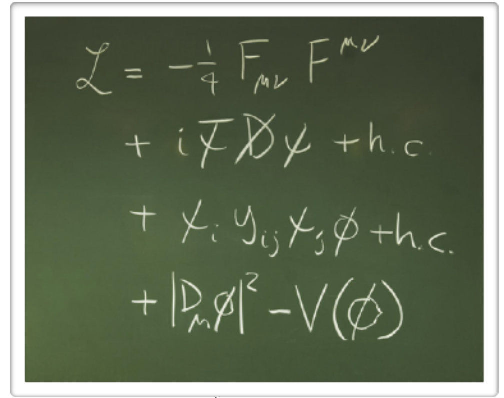
Unfolding
Anomaly Detection

AI/ML models are studied for all aspects of experimental particle physics



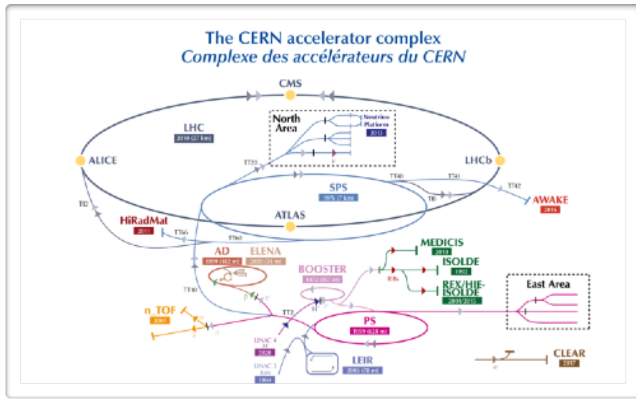
and interesting statistical problems lurk everywhere


AI
Inference
Experiment
Design
Simulation
Triggers
Tagging
Reconstruction
Unfolding
Anomaly Detection
The CERN accelerator complex
Complexe des accélérateurs du CERN
CMS Experiment at the LHC, CERN
Data recorded: 2015-Jun-03 08:48:32.279552 GMT
Run / Event / LS: 246908 / 77874559 / 86
19.7 fb⁻¹ (8 TeV) + 5.1 fb⁻¹ (7 TeV)
CMS
H → γγ
S/(S+B) weighted sum
Data
S+B fits (weighted sum)
B component
S/(S+B) weighted events / GeV
Di-Photon invariant mass



Inference

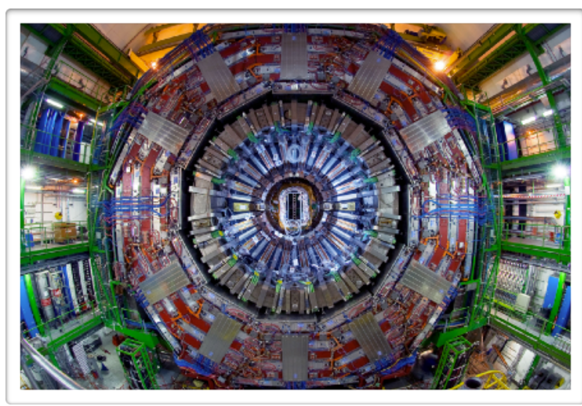
Experiment Design



AI

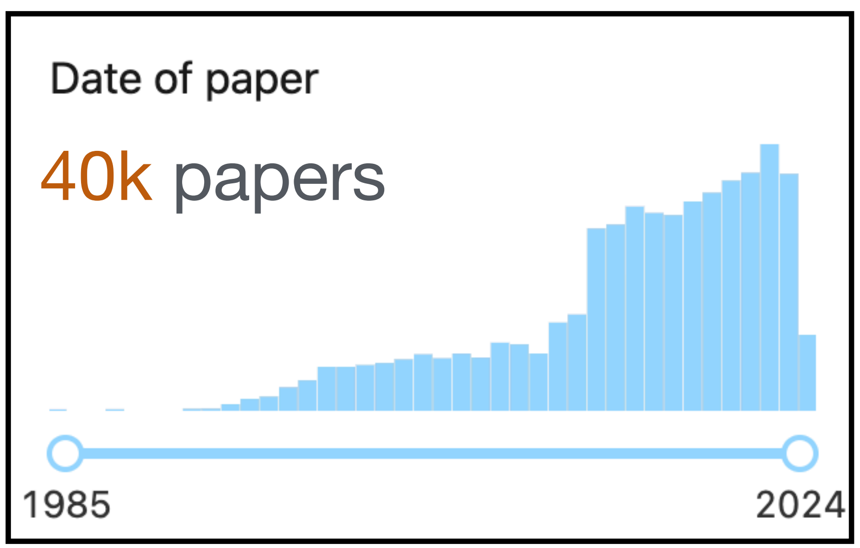
Simulation

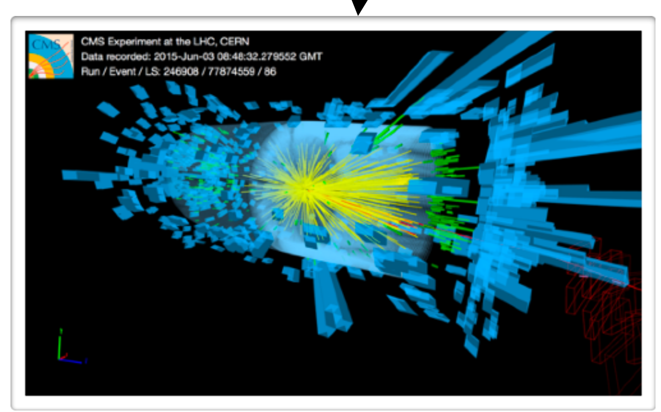
Triggers

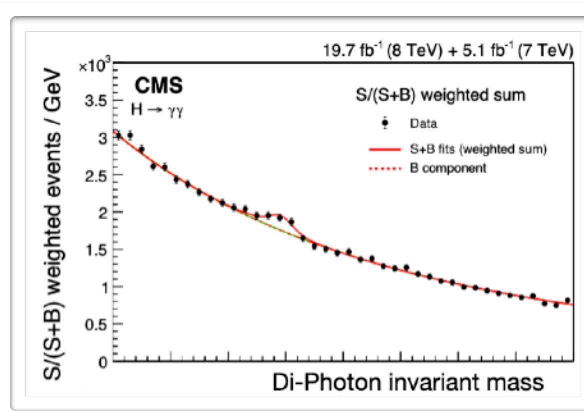


Tagging
Reconstruction

Unfolding
Anomaly Detection







Apologies - there will be some selection bias.

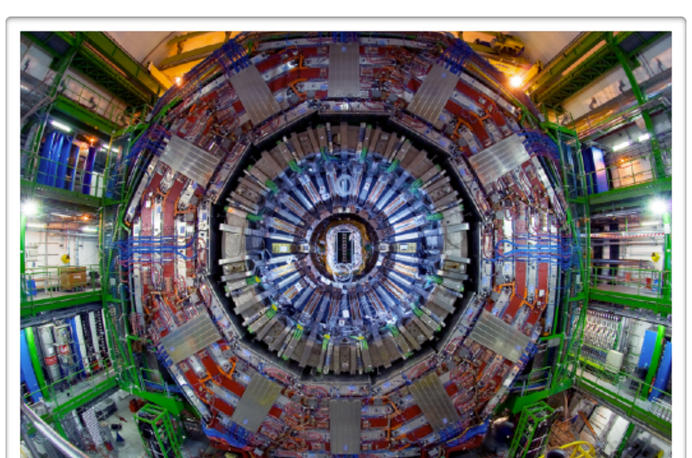
Particle Data Primer

"Data Data"

Particle collisions with ~1 MB/event happen at a rate of 40 MHz

Distill to ~1 kHz via lossy, irreversible filtering algorithms (Trigger)

Samples i.i.d. from physics distribution (e.g. the Standard Model + potential new physics)

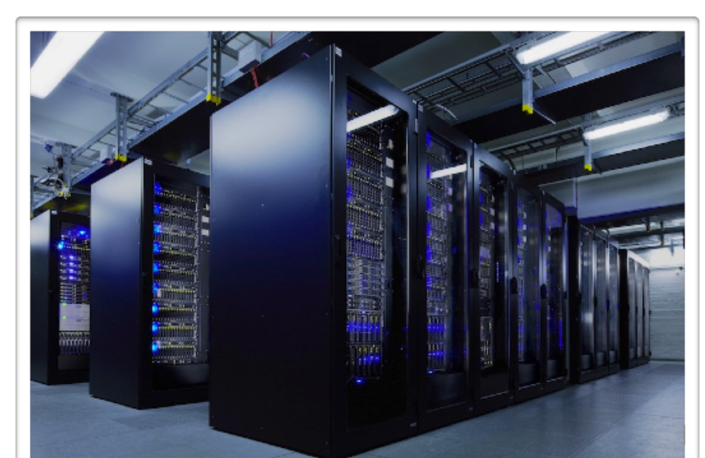


"Simulated Data"

Based on first-principle mathematical model of physical theory and detector interaction

Full control over which process occurs

Uncertainty in detector calibration and natural constants encoded by performing multiple simulations, one per set of parameter values



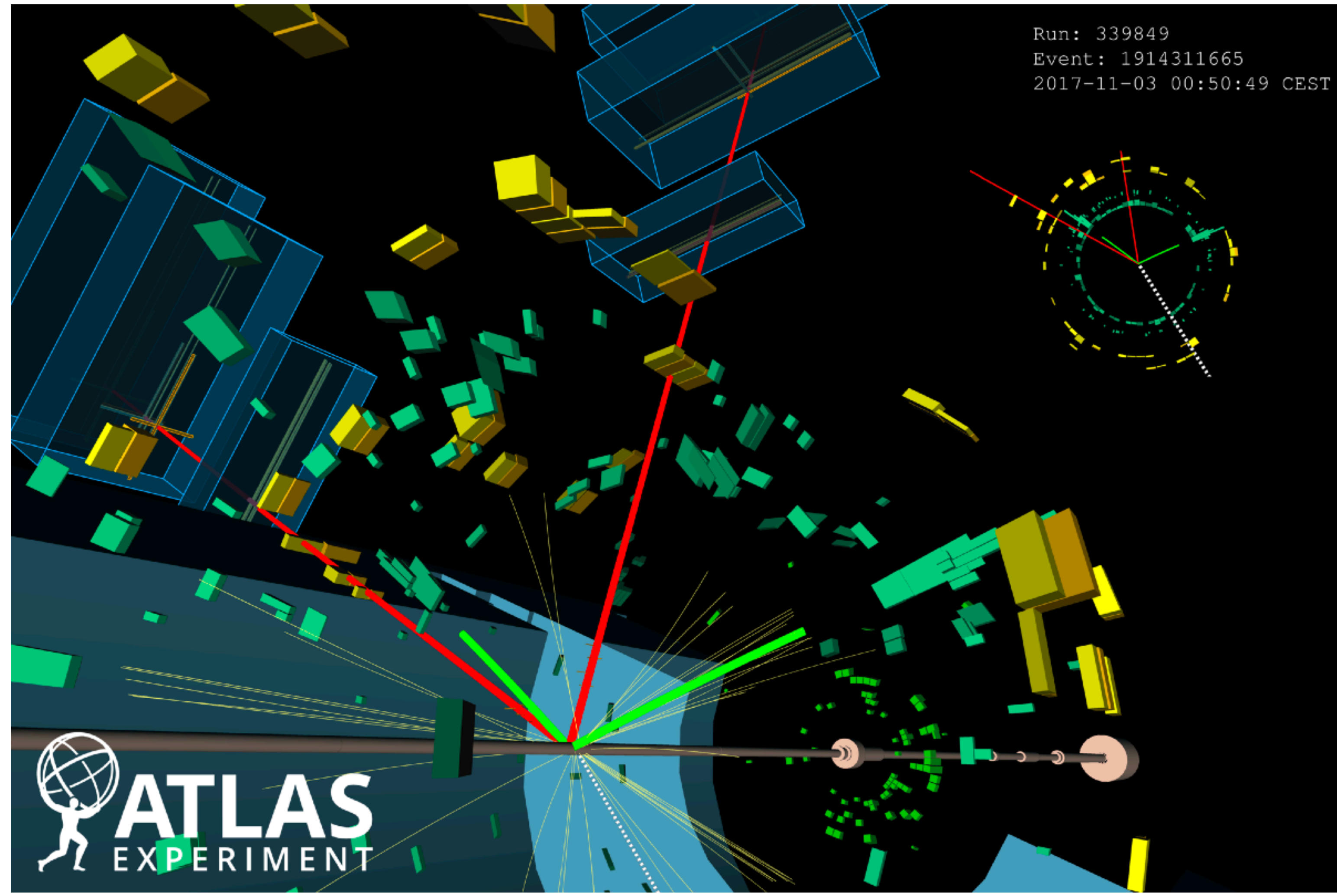
Particle Data Primer

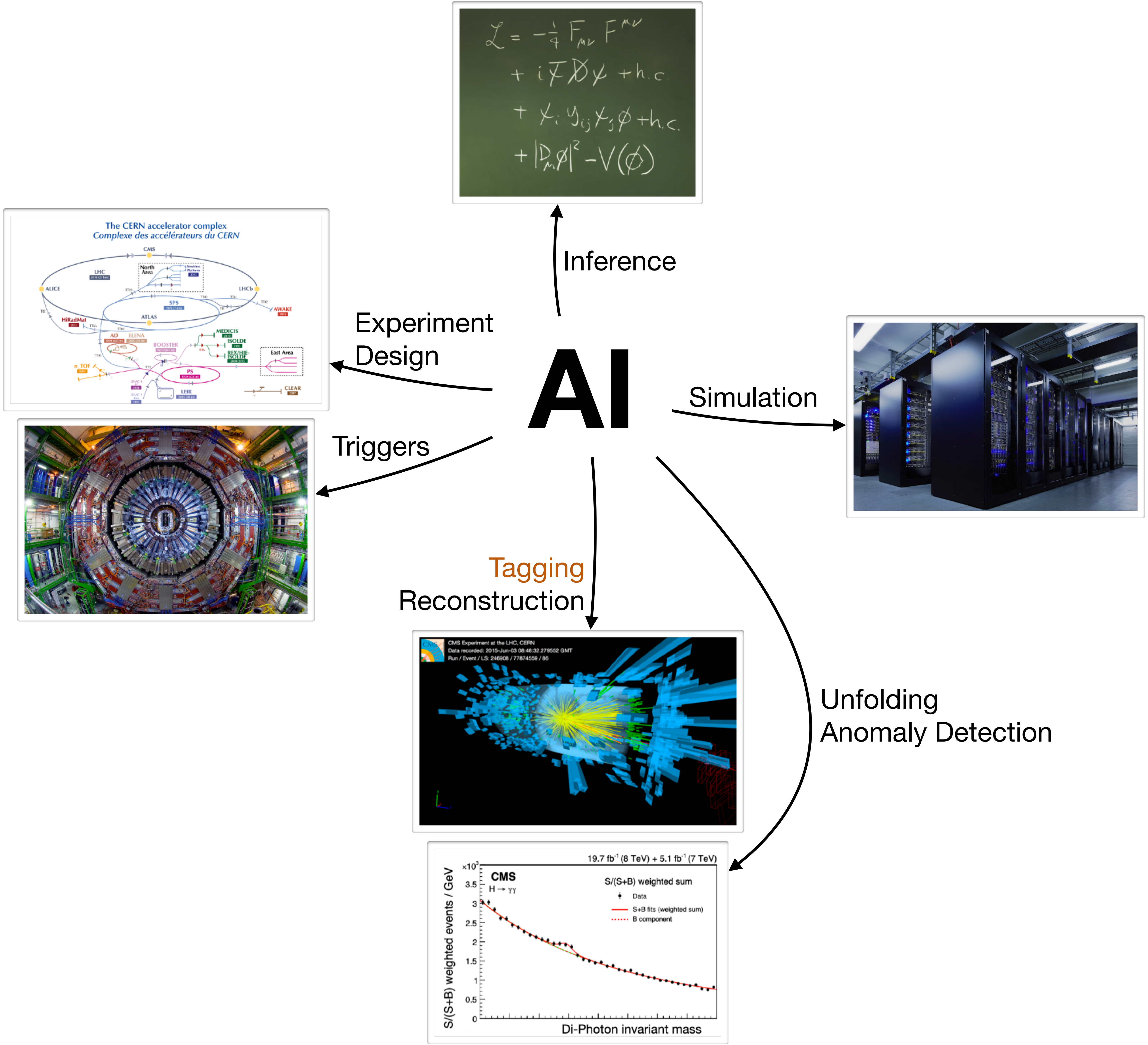
Both types of data share common format and reconstruction software

Different ways of representation (all aligned with physical interpretation):
-low-level readouts in ~100M channels
-10s-100s of intermediate four-vectors
-O(10) high-level features

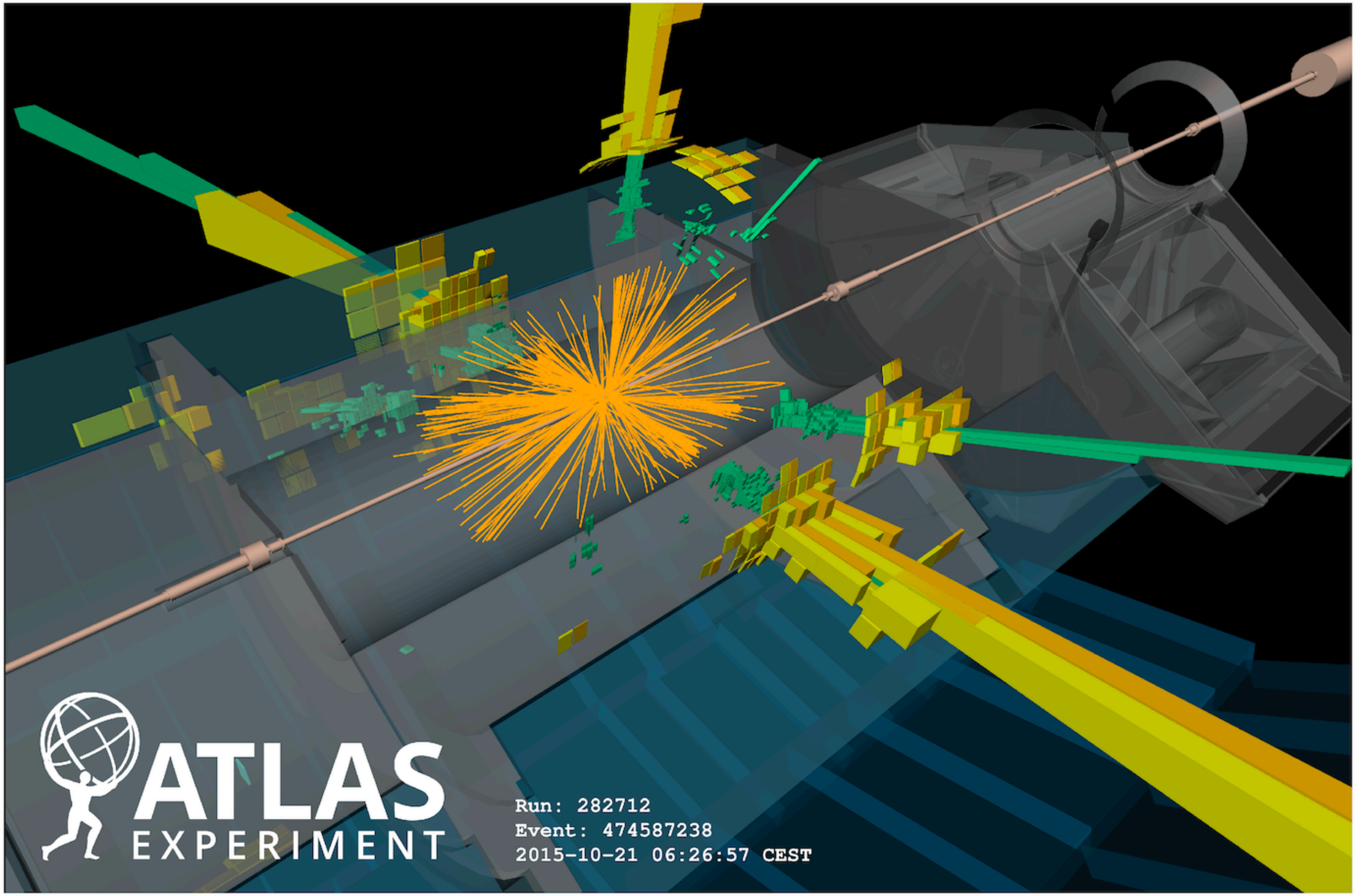
Message passing/attention-based models

One collision/event = “one image”

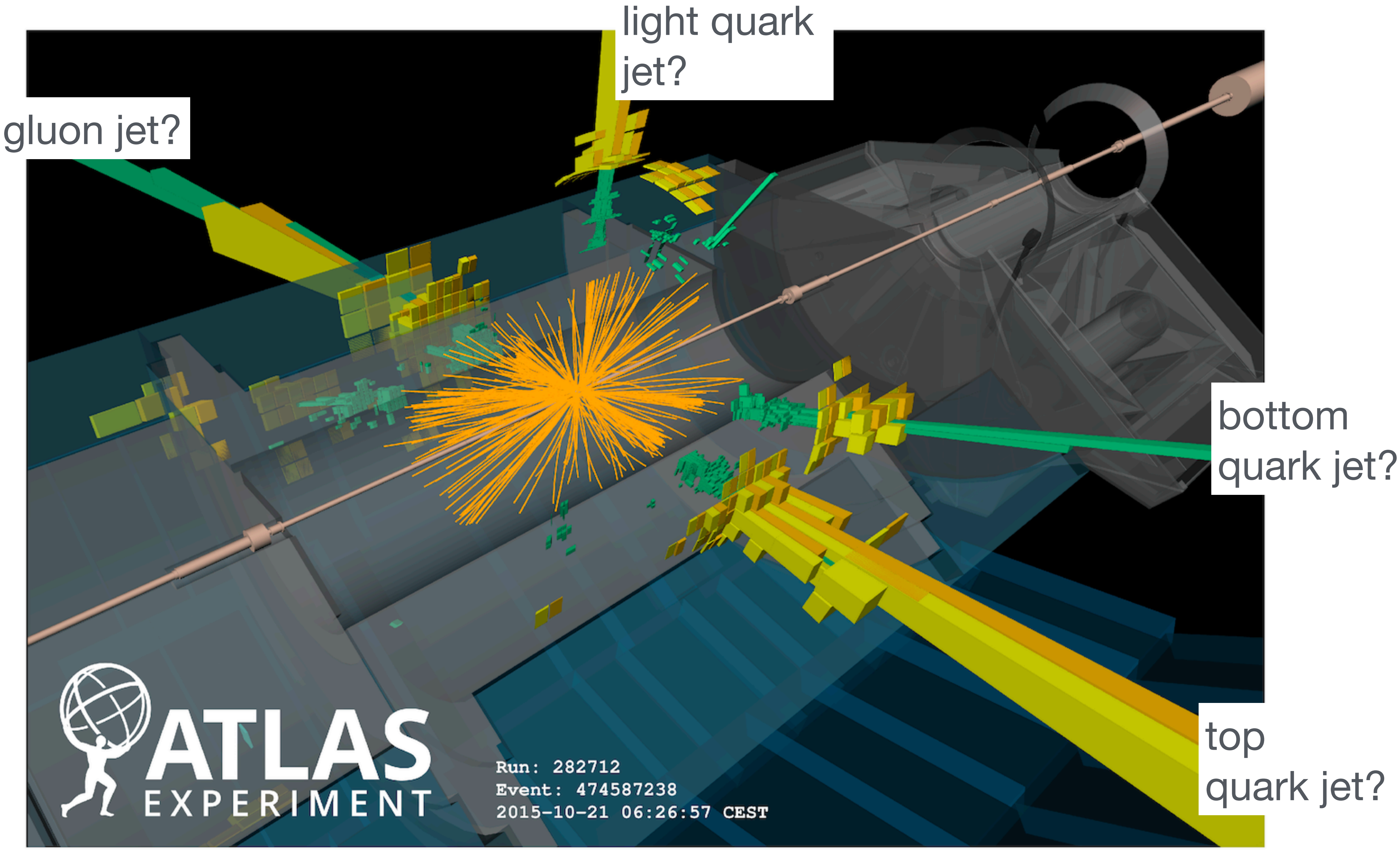




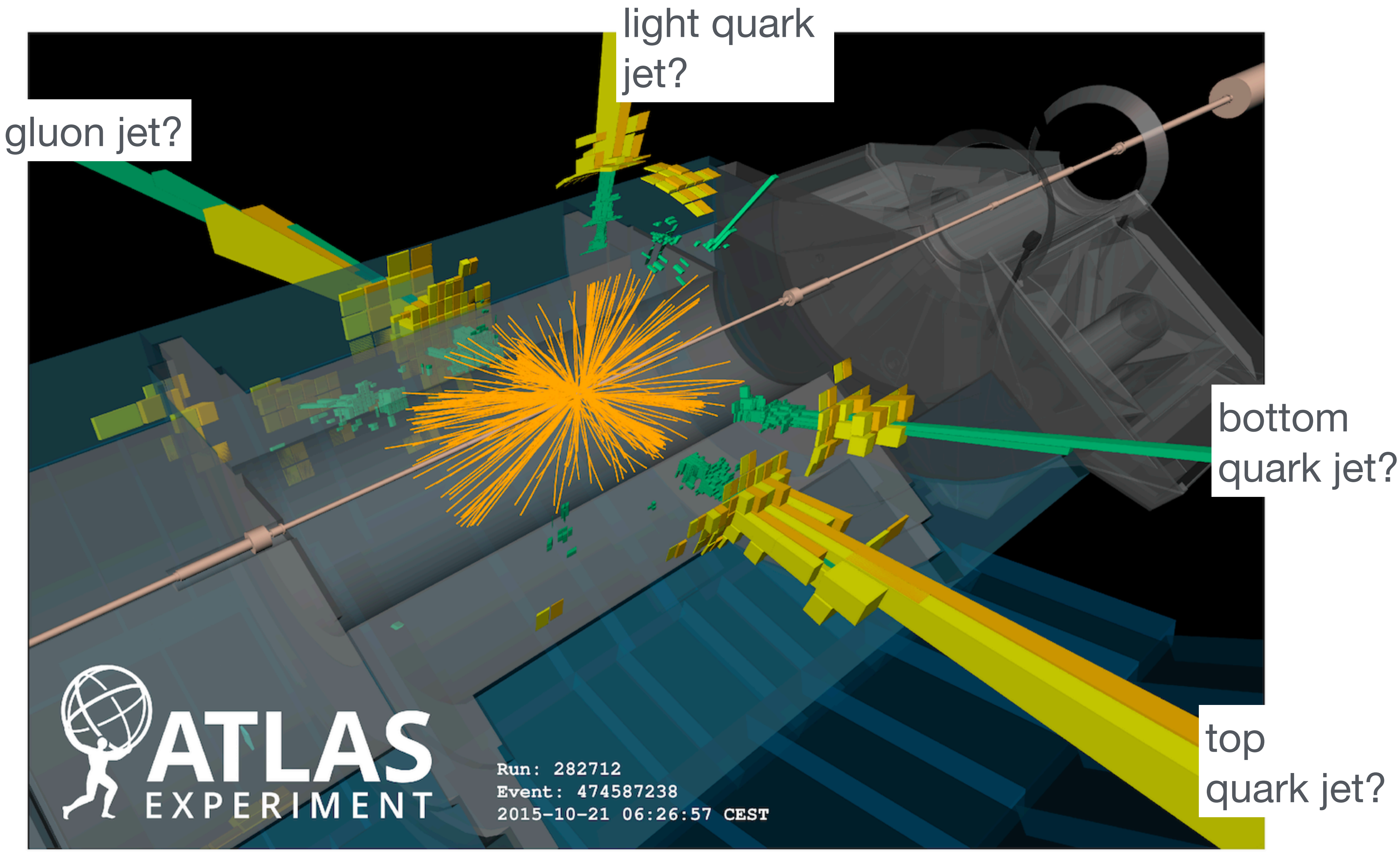
AI
Inference
Experiment Design
Simulation
Triggers
Tagging
Reconstruction
Unfolding
Anomaly Detection
The CERN accelerator complex
Complexe des accélérateurs du CERN
CMS Experiment at the LHC, CERN
Data recorded: 2015-Jun-03 08:48:32.279552 GMT
Run / Event / LS: 246908 / 77874559 / 86
CMS
H → γγ
S/(S+B) weighted sum
Data
S+B fits (weighted sum)
B component
19.7 fb⁻¹ (8 TeV) + 5.1 fb⁻¹ (7 TeV)
S/(S+B) weighted events / GeV
Di-Photon invariant mass



A jet is a
collimated shower of particles in the detector

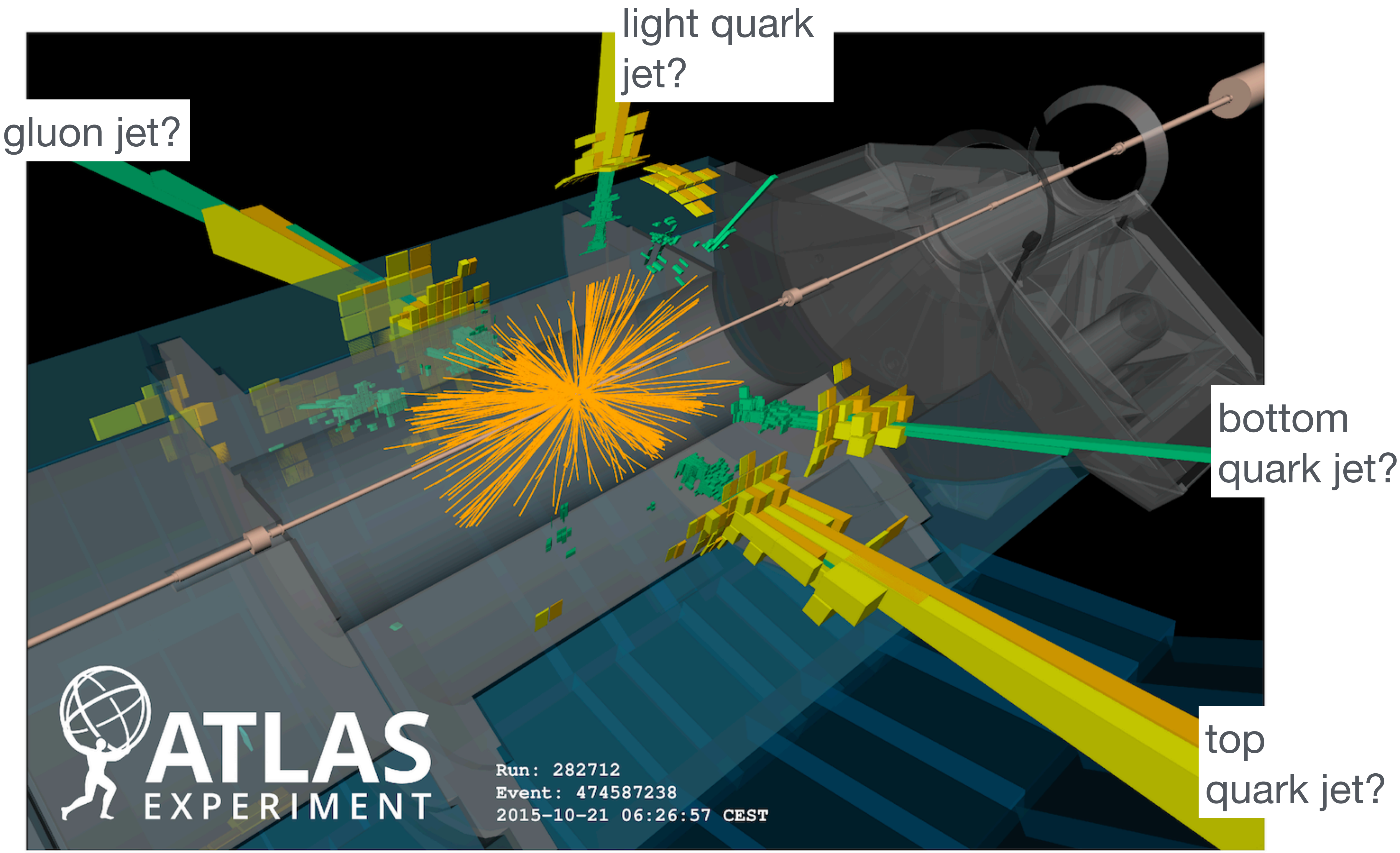


We want to know
which particle produced a jet

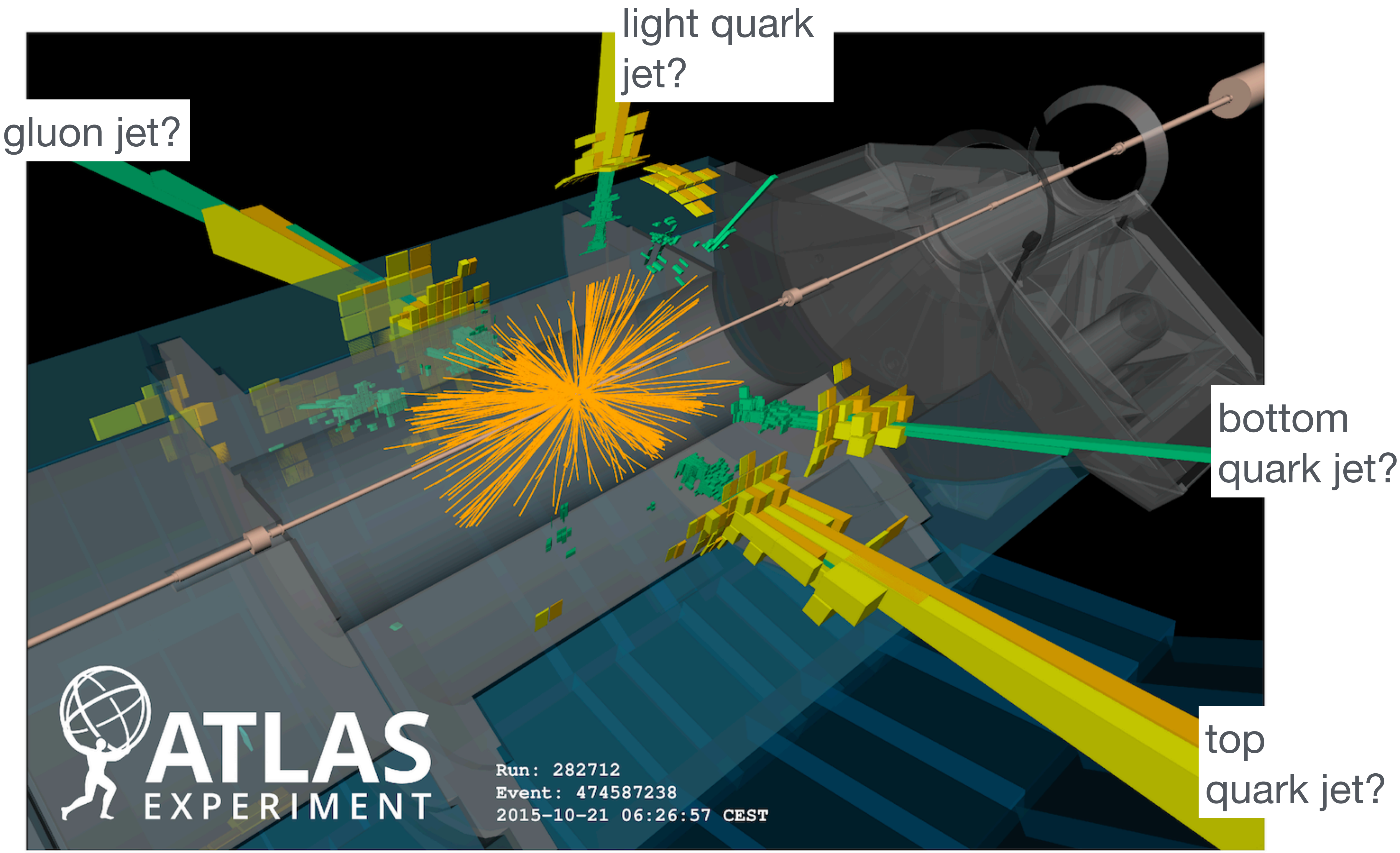


Why?

- Discover new particles
- Measure the Standard Model



Let’s focus on top quarks
(Modern taggers are multi-class)



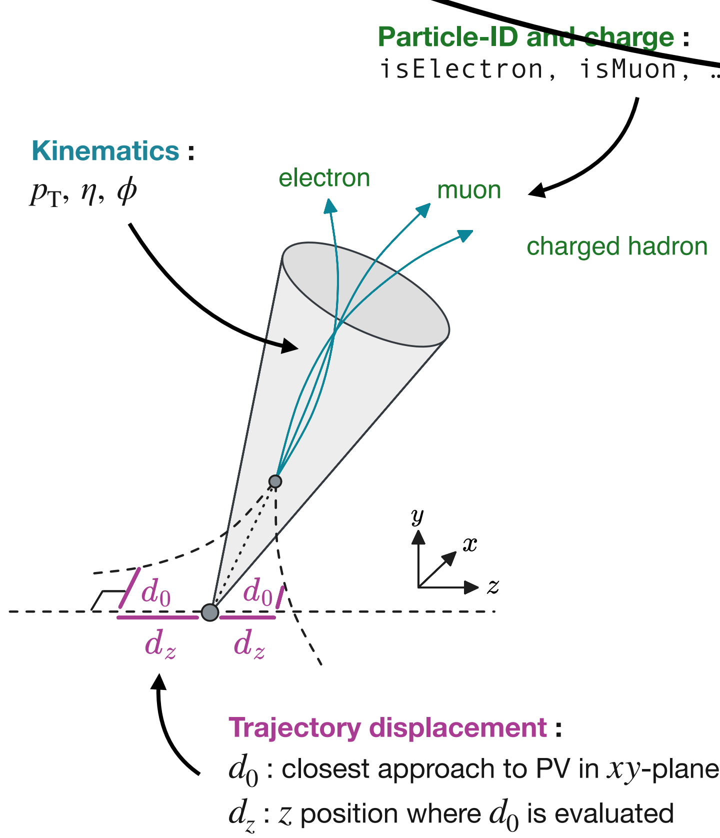
How to build ML algorithms for complex, heterogenous data?

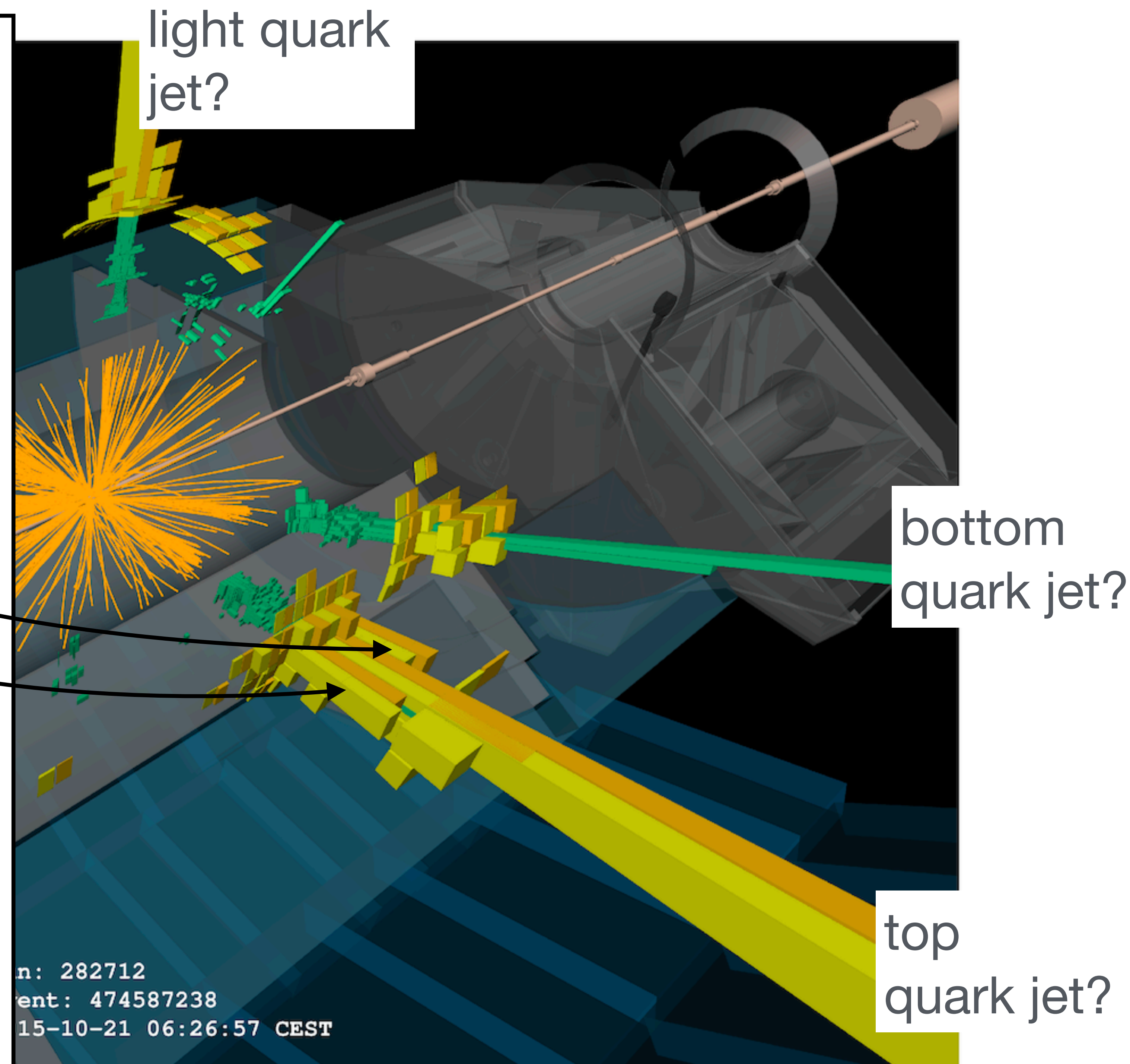
Data most naturally viewed as point cloud:

Each input (e.g. jet, event, ..) is a set of k-dimensional vectors (individual particles, hits, ..)

$$J_i = \{\vec{p_1}, \ldots, \vec{p_n}\}$$

Example per-particle features

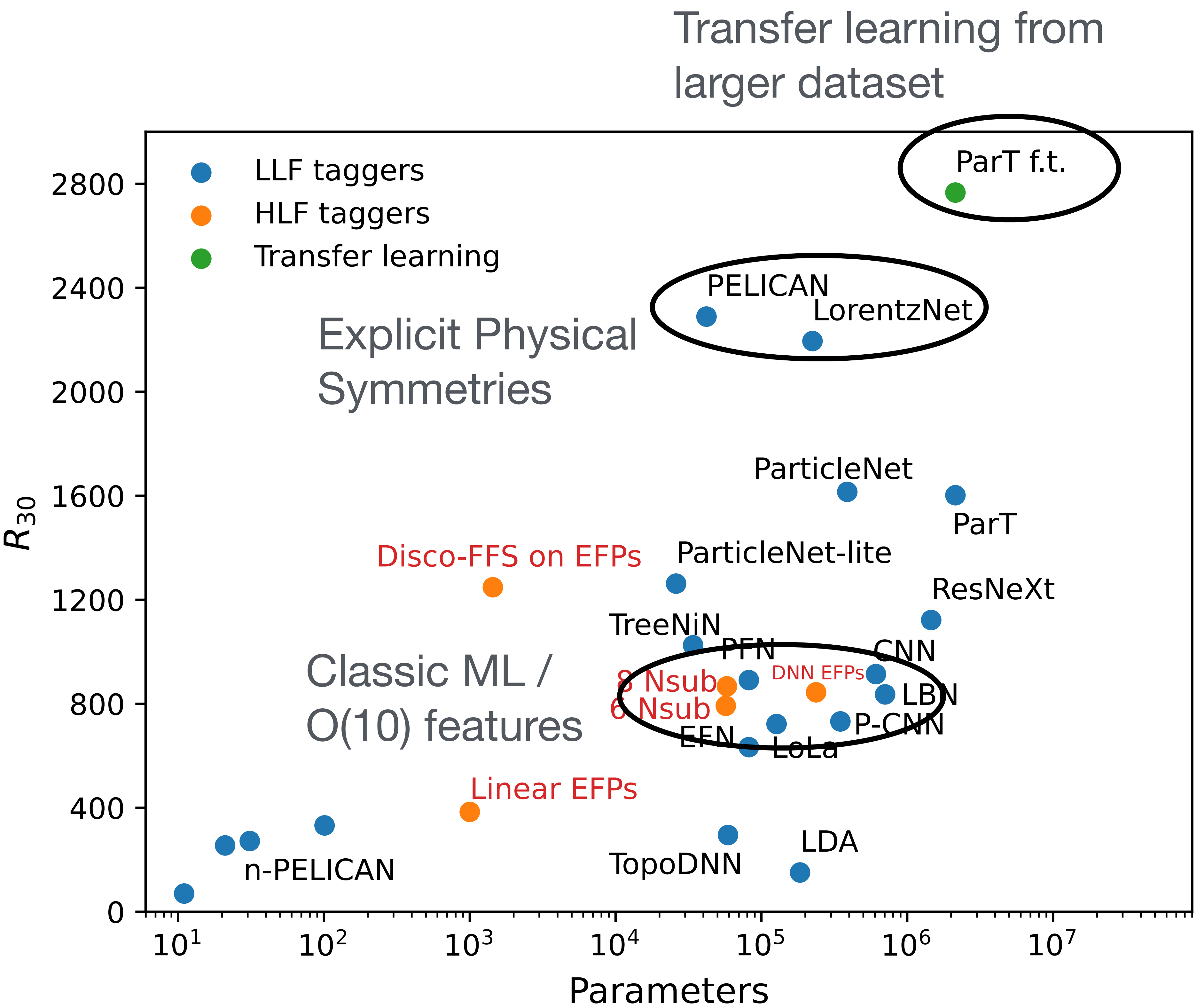




Komiske, Metodiev, Thaler 1810.05165; Birk, **GK**, et al 2312.00123;

Status

3x+ gain in background rejection by going from classical ML on few low level features to deep architectures that know physics (either built-in or transfer learned)

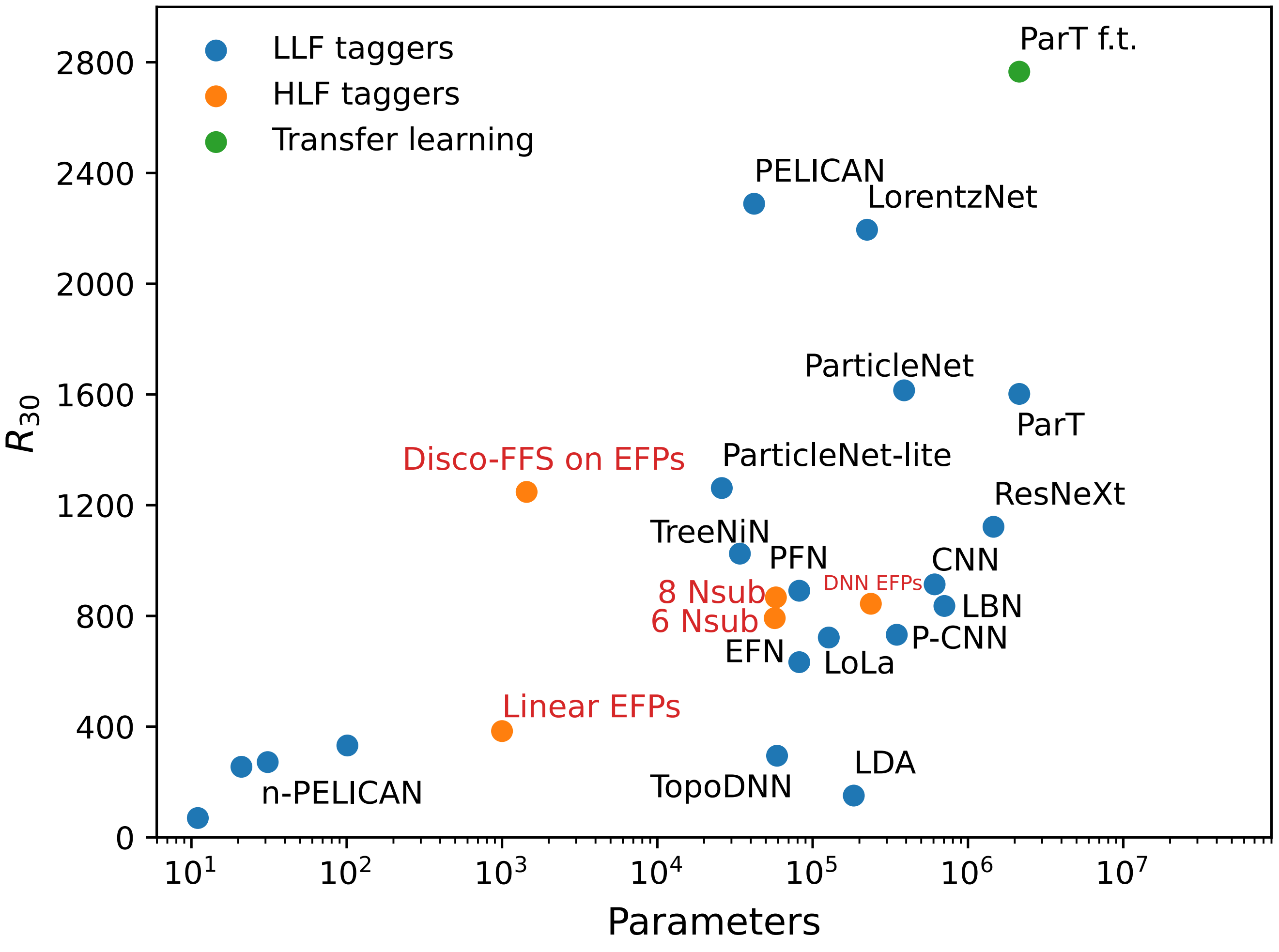


See talk by **Jonas**

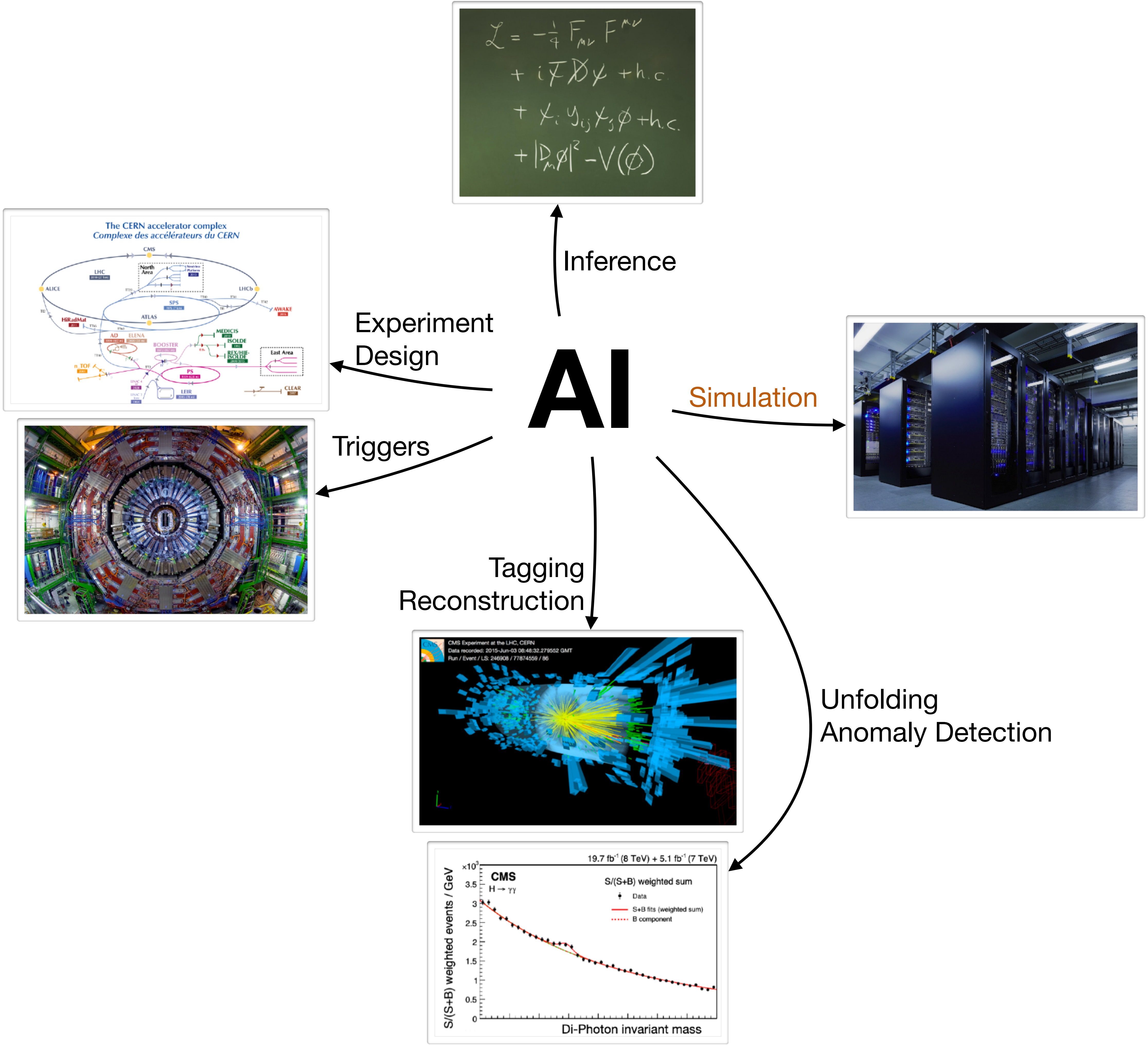
(Some) Current challenges

- "Calibration": Domain adaptation between simulation and collider data
- Local vs global optimisation
- Uncertainty aware training
- Interpretability

See talk by **Jesse, Oliver**



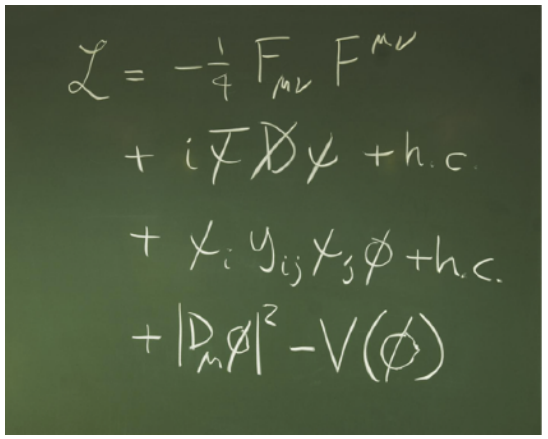
ATL-PHYS-PUB-2022-039

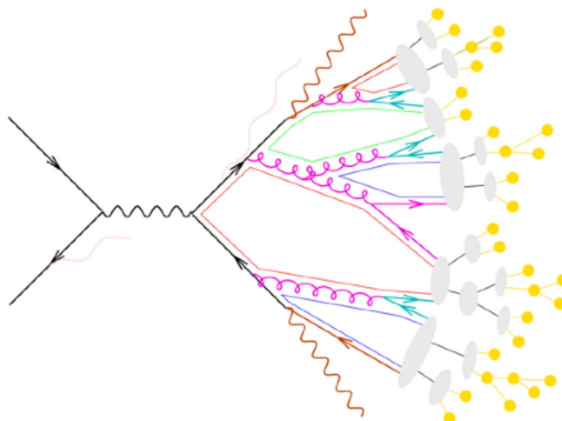


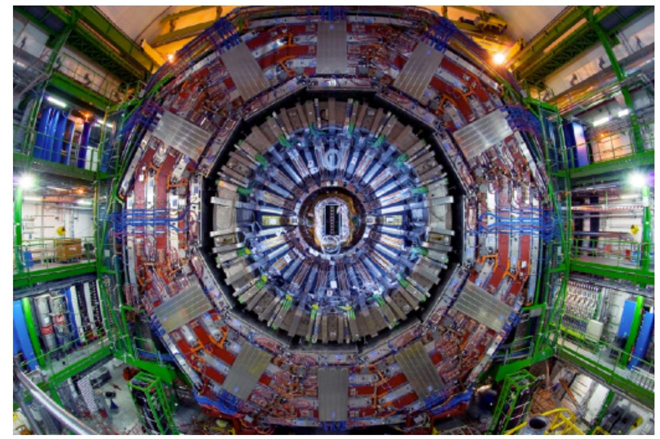
$\mathcal{L} = -\frac{1}{4} F_{\mu\nu} F^{\mu\nu} + i \bar{\psi} \not{D} \psi + h.c. + \psi_i y_{ij} \psi_j \phi + h.c. + |D_\mu \phi|^2 - V(\phi)$
Inference
The CERN accelerator complex
Complexe des accélérateurs du CERN
Experiment Design
AI
Simulation
Triggers
Tagging
Reconstruction
CMS Experiment at the LHC, CERN
Data recorded: 2015-Jun-03 08:48:32.279552 GMT
Run / Event / LS: 246908 / 77874559 / 86
Unfolding
Anomaly Detection
19.7 fb^{-1} (8 TeV) + 5.1 fb^{-1} (7 TeV)
CMS
$H \rightarrow \gamma\gamma$
S/(S+B) weighted sum
Data
S+B fits (weighted sum)
B component
S/(S+B) weighted events / GeV
Di-Photon invariant mass

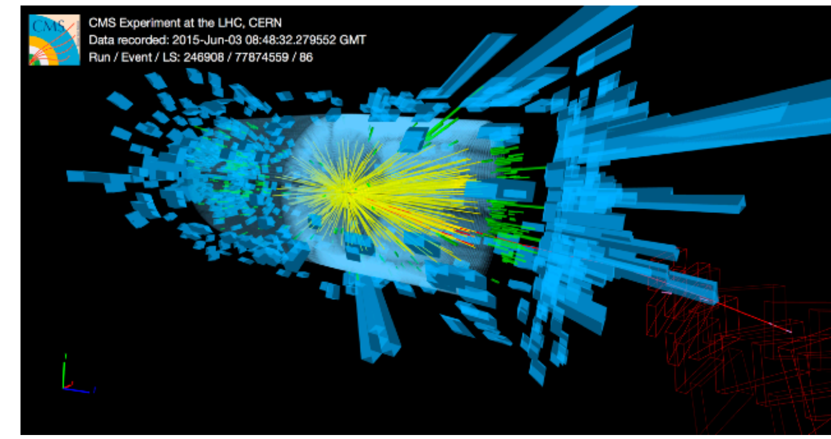
Generative Models

This happens in the experiment







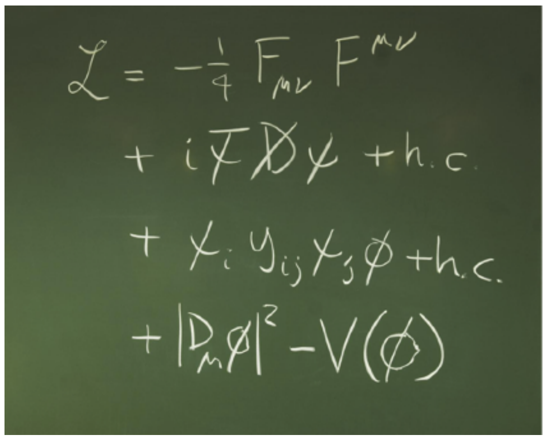


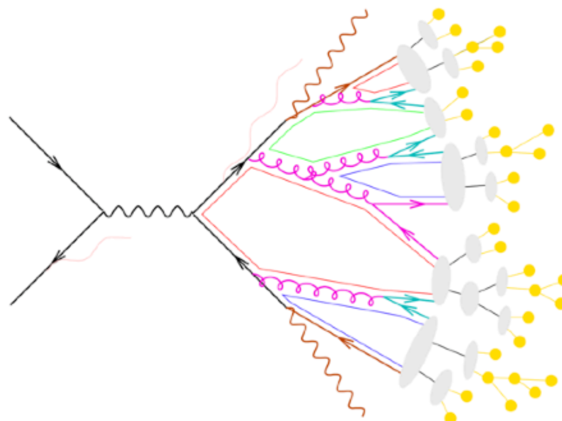
This is what we want to know

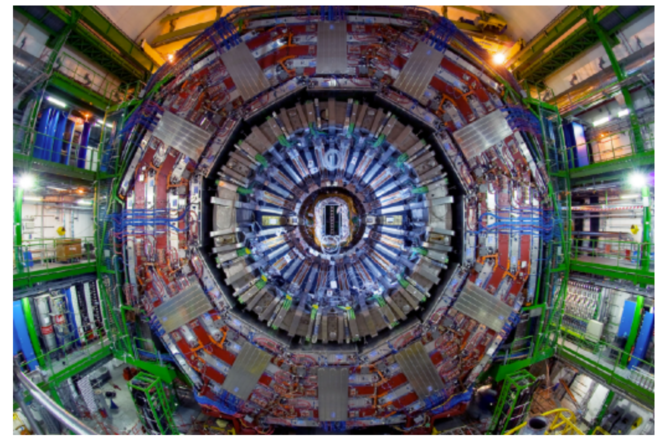
Simulation is crucial to connect experimental data with theory predictions

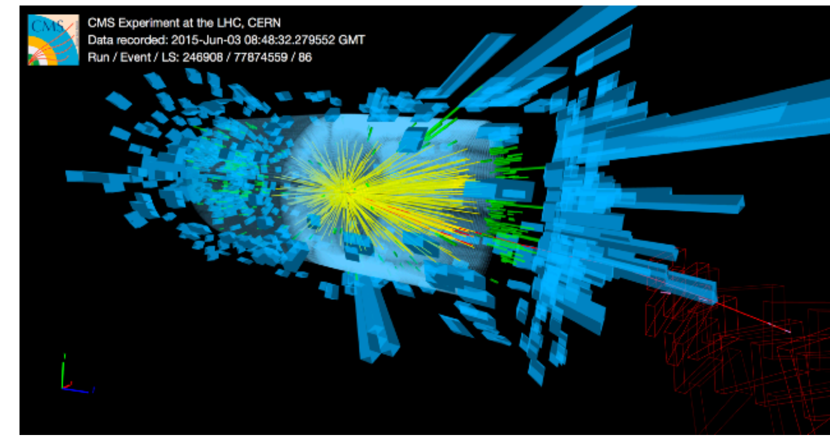
Generative Models

This happens in the experiment



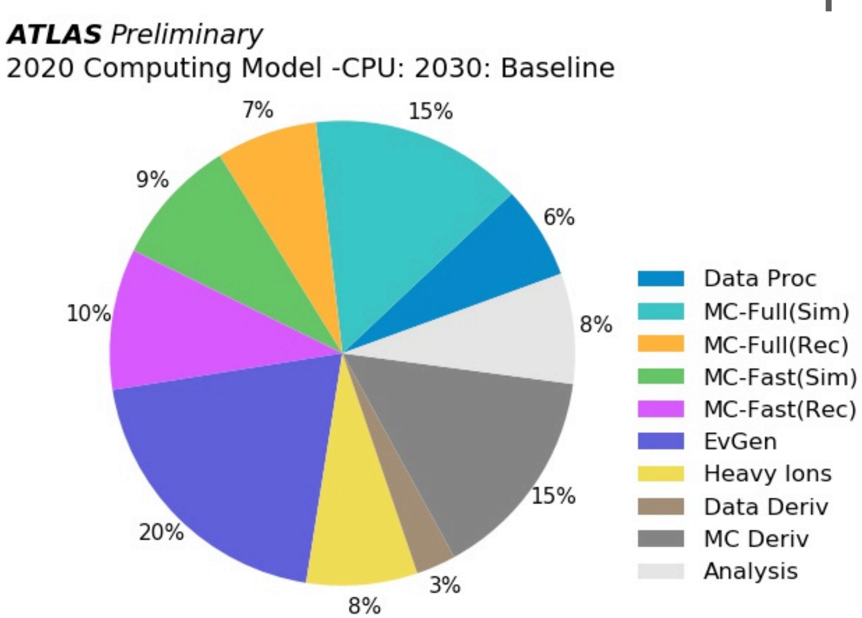


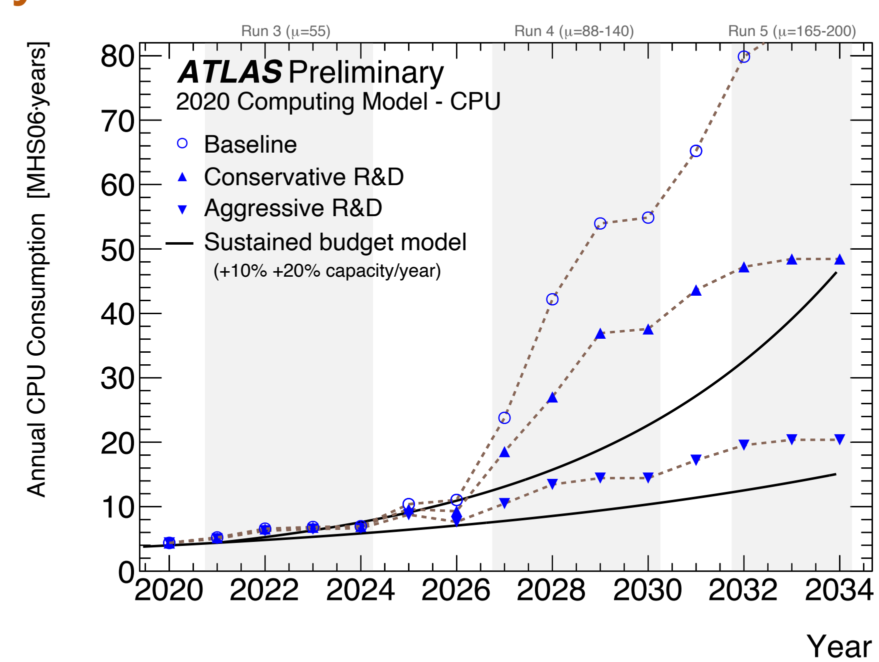




This is what we want to know

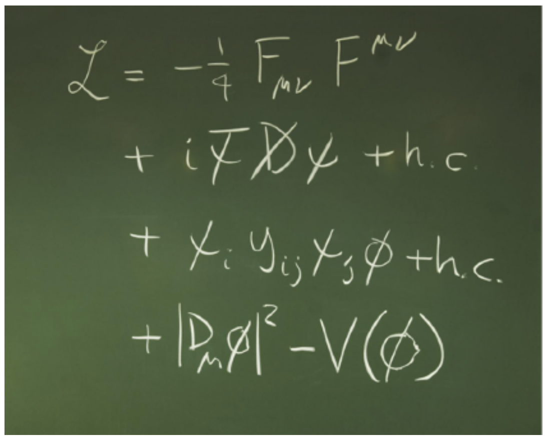
Simulation is crucial to connect experimental data with theory predictions, but computationally very costly

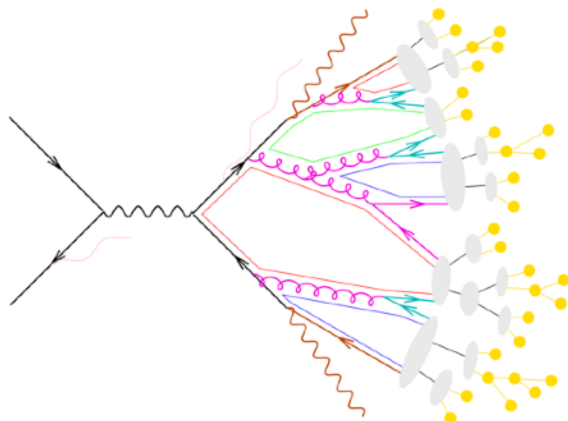


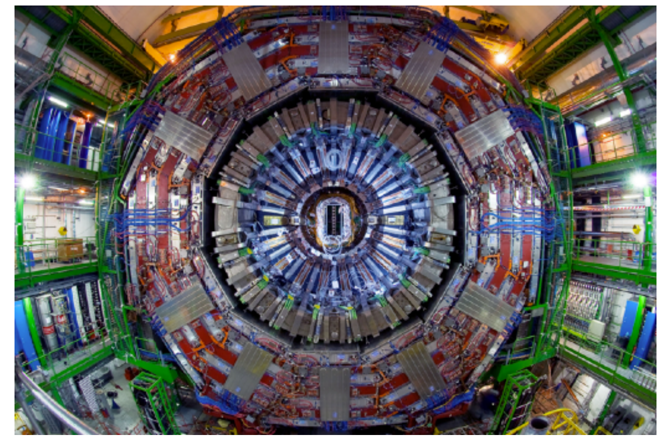


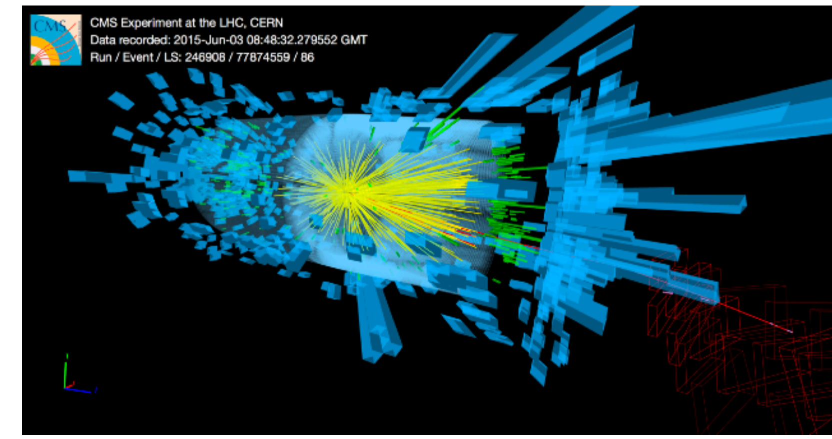
Generative Models

This happens in the experiment









This is what we want to know

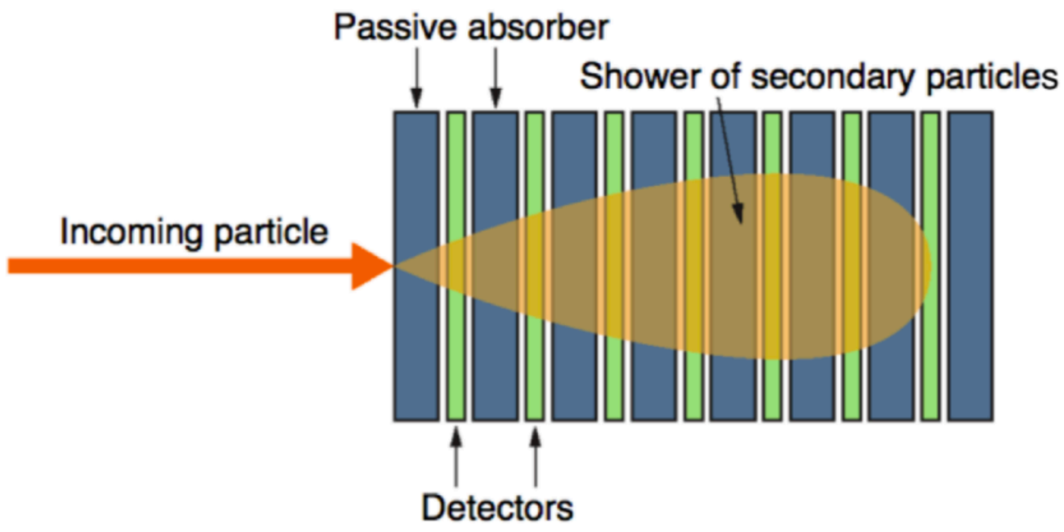
Simulation is crucial to connect experimental data with theory predictions, but computationally very costly

Use AI to improve efficiency of simulation codes or learn surrogates

Calorimeter Simulation

Interaction of particles with multi-layer detectors to determine their initial energy (and type)

Measurement of energy, position, (and time) of secondary particle hits

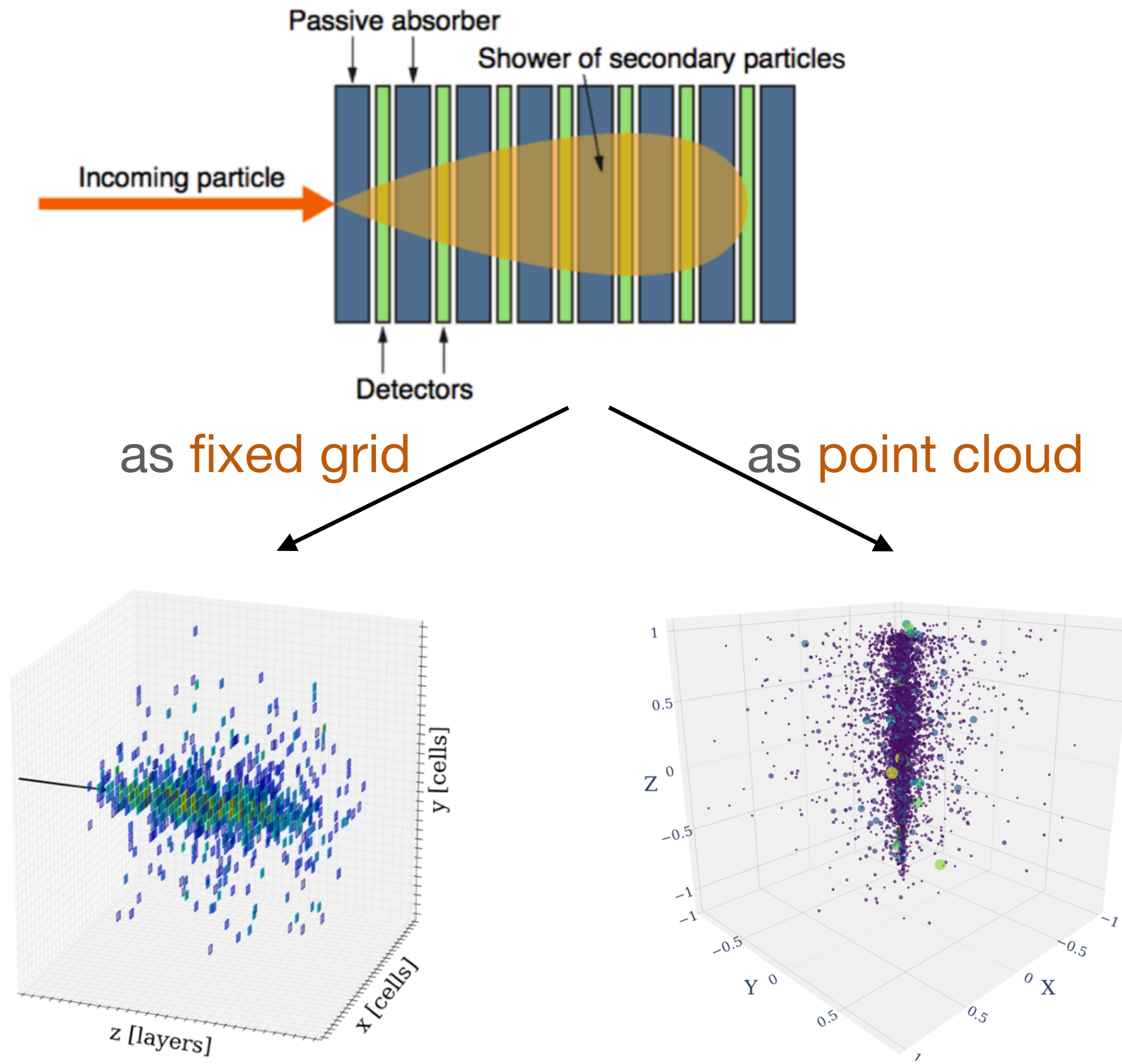


Calorimeter Simulation

Interaction of particles with multi-layer detectors to determine their initial energy (and type)

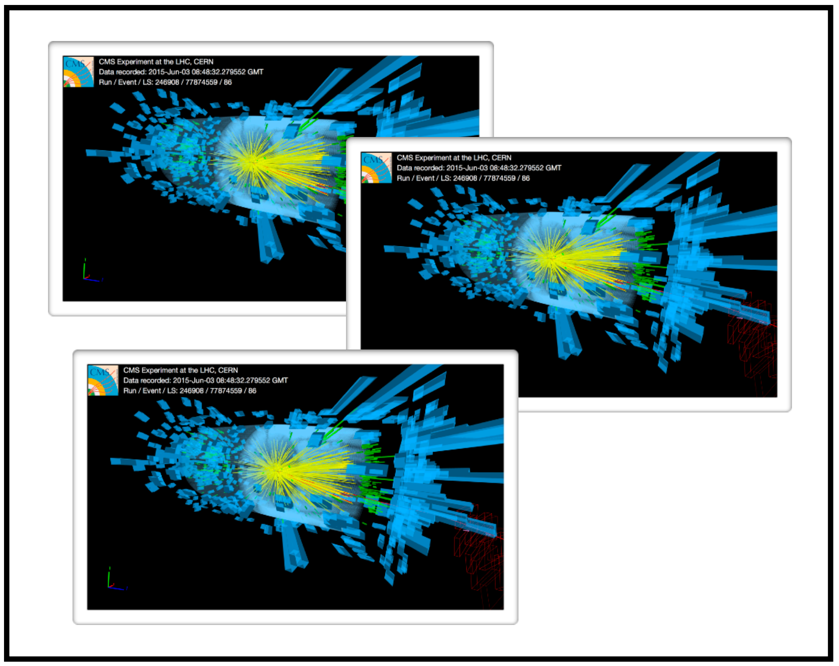
Measurement of energy, position, (and time) of secondary particle hits

Represent data as
-fixed grid
(3d matrix of detector elements, value=energy)
-point cloud
(set of hits, each hit is a 3d vector with position+energy)

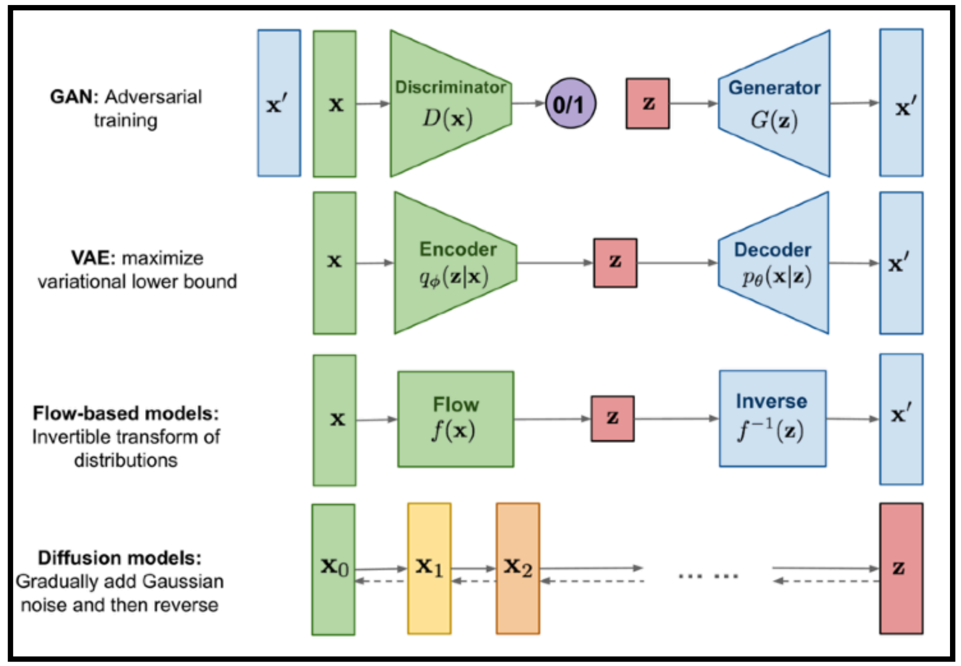


Strategy

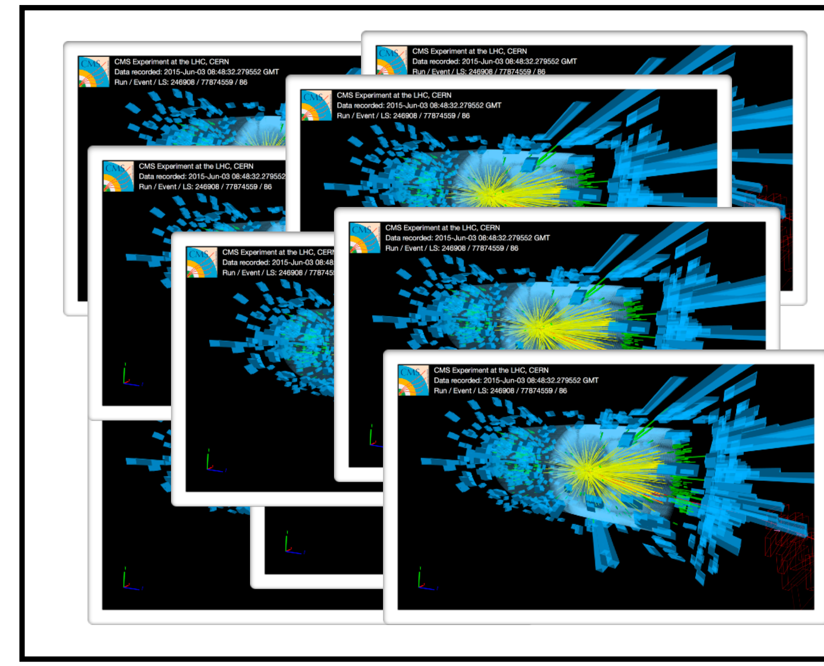
1. Use classical simulation or collider data as input



2. Train generative surrogate



3. Oversample



See talk by **Ramon, Tobias**

Paganini, Oliveira, Nachman 1705.02355; Butter, Diefenbacher, **GK**, et al 2008.06545;

Surrogate Models

Collider or classical simulation

Slow

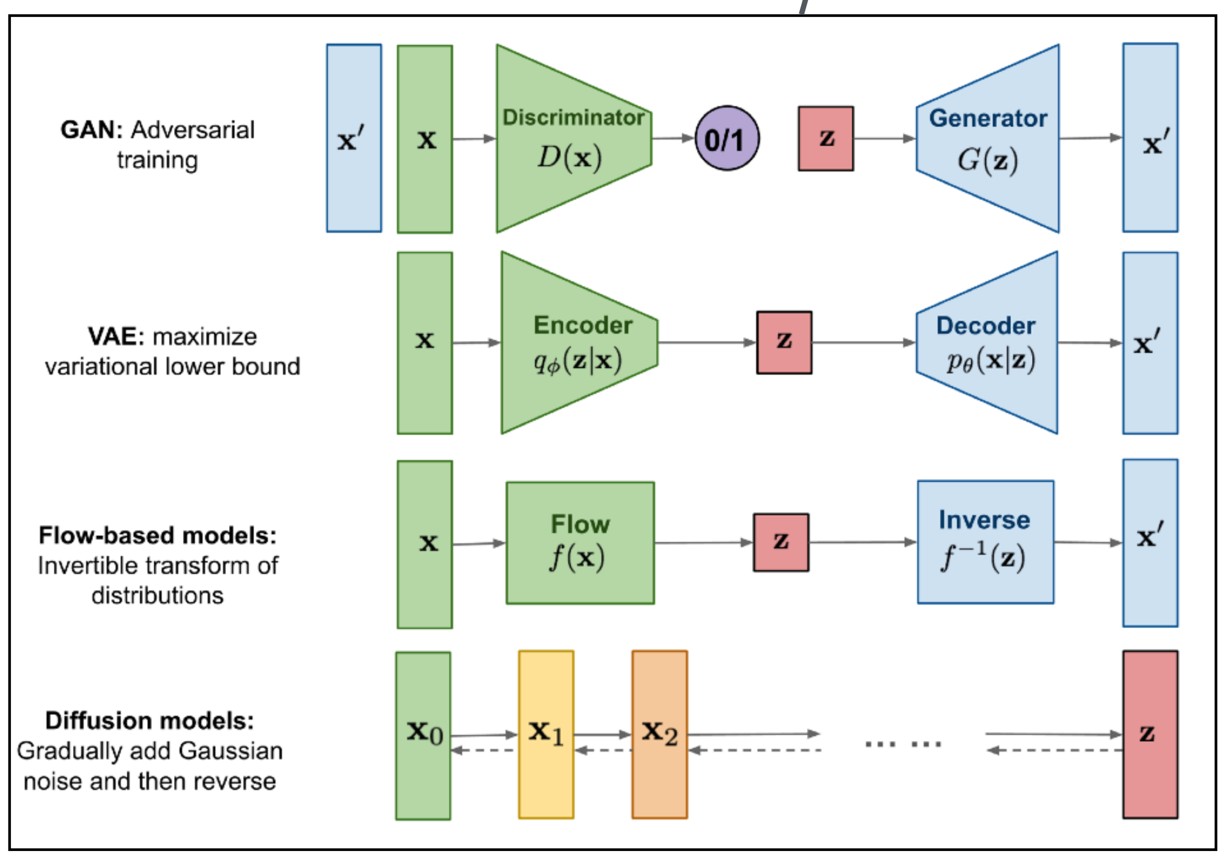
Training Data Set

Surrogate Model

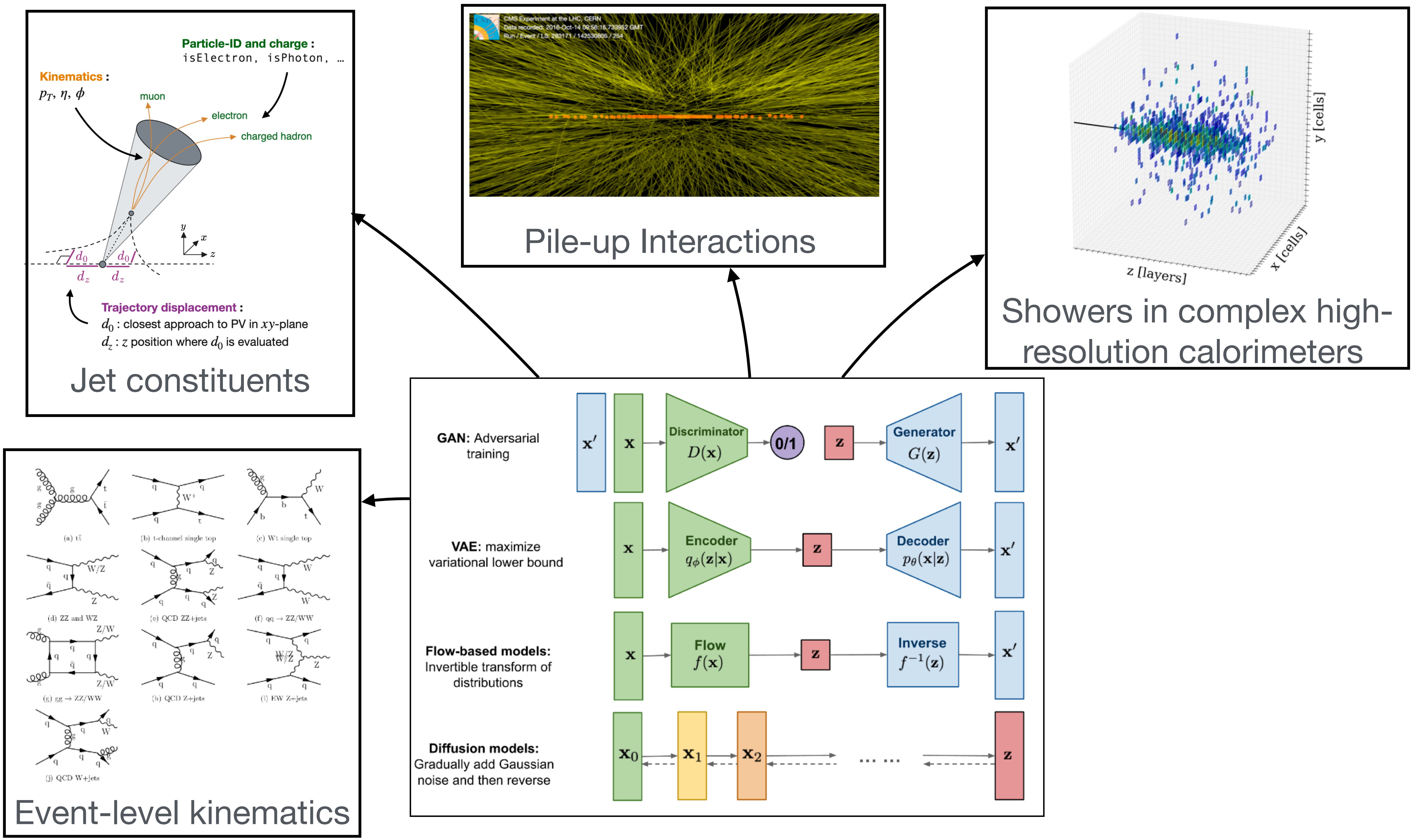
Fast

Synthetic data used in analysis

GEANT4
A SIMULATION TOOLKIT



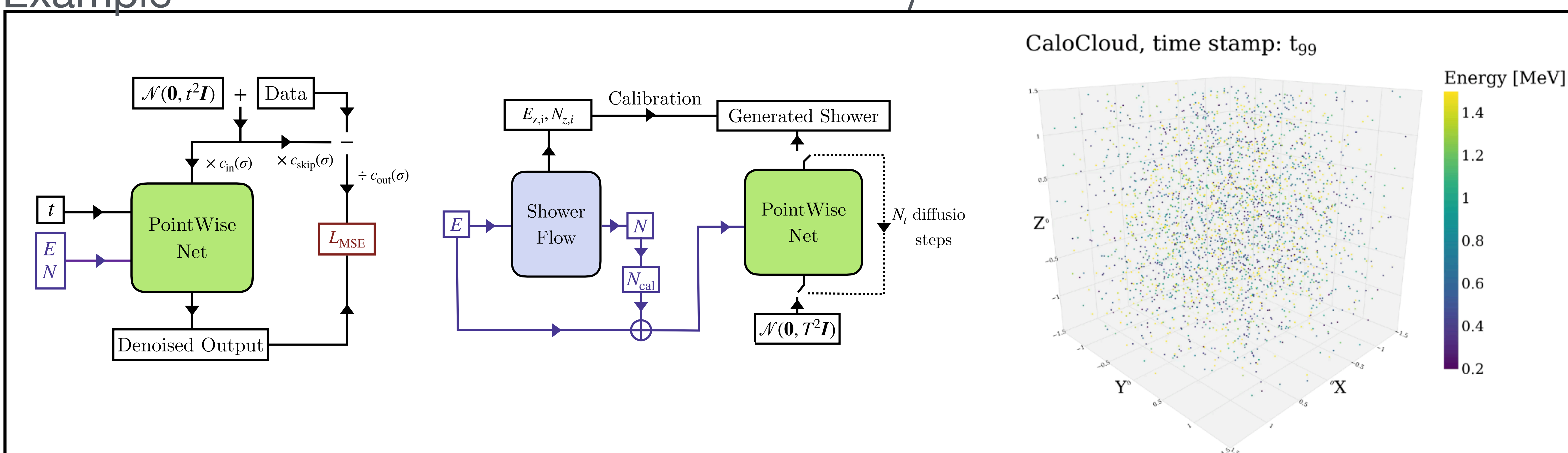
Surrogate Models



Surrogate Models

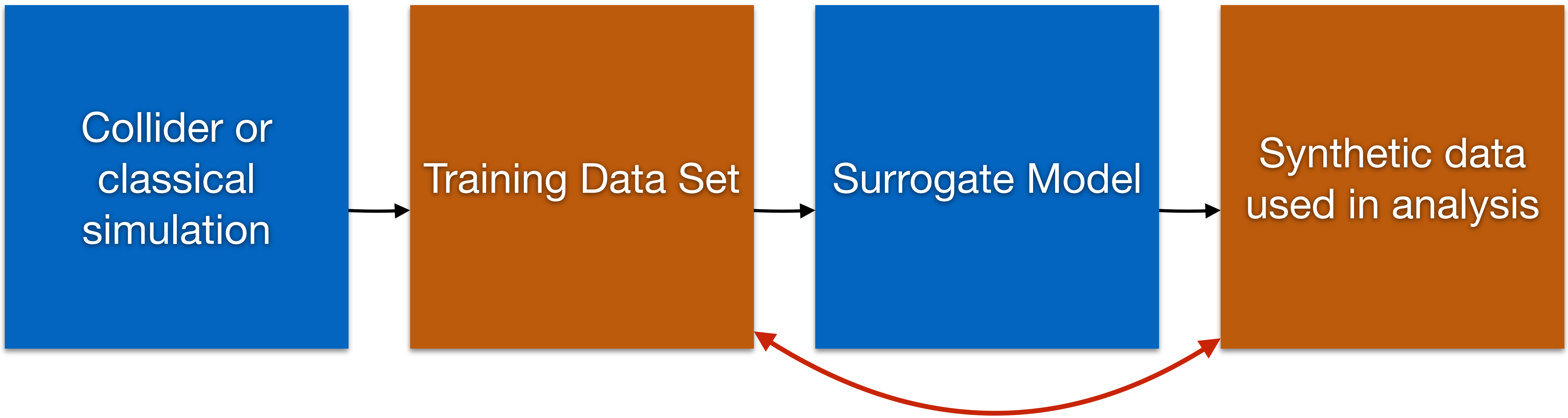
Collider or classical simulation → Training Data Set → Surrogate Model → Synthetic data used in analysis

Example



Buhmann, .., GK, et al 2305.04847; Buhmann, .., GK et al 2309.05704; see e.g. CaloChallenge for comprehensive results

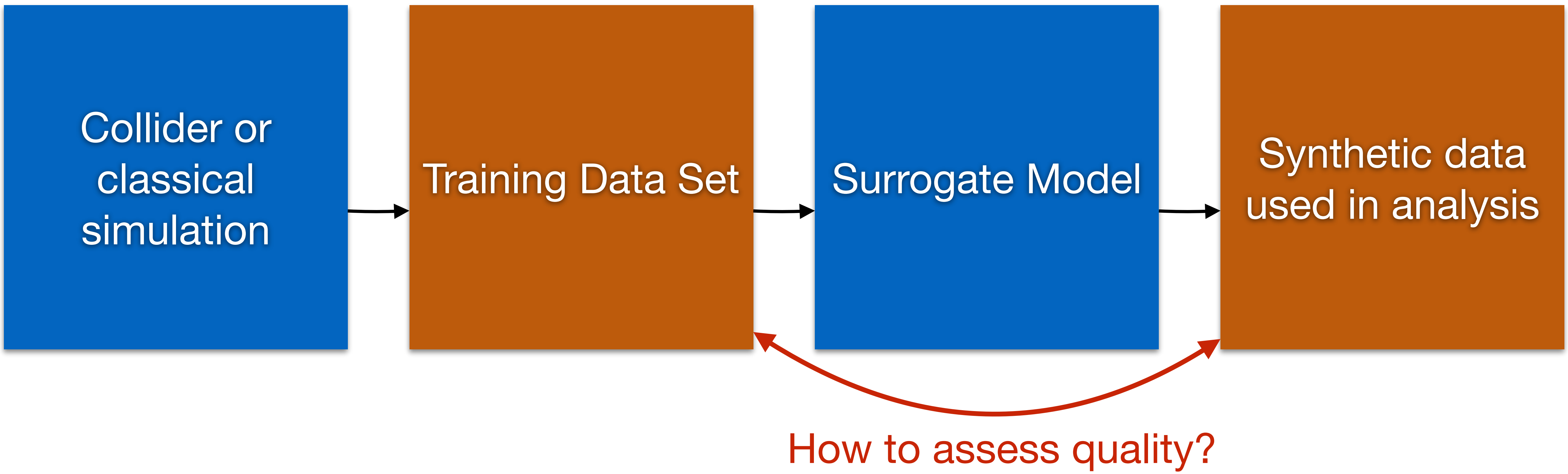
Surrogate Models

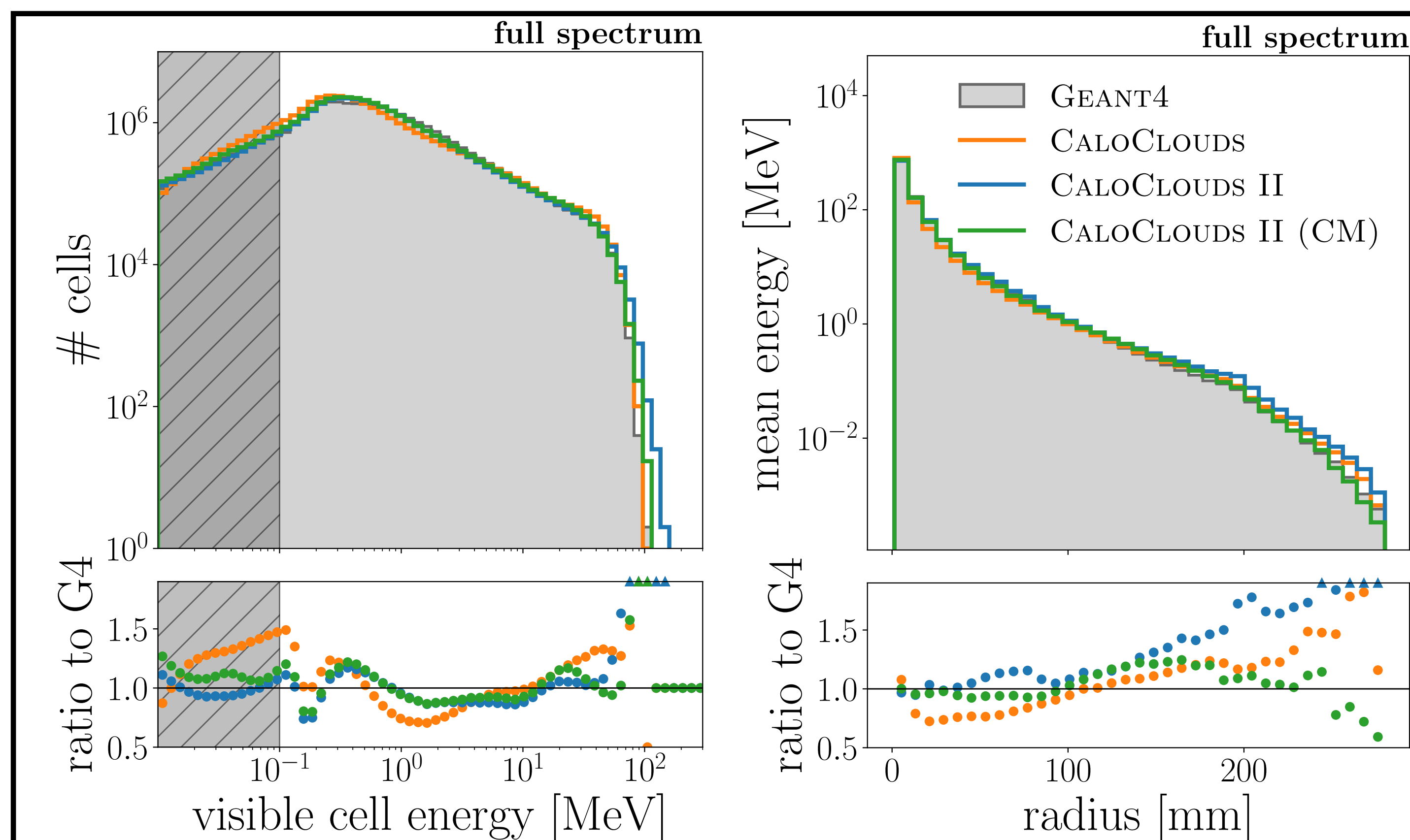


How to assess quality?

Is one surrogate better than the other?
Is the synthetic data good enough for usage?
What additional uncertainties to take into account?

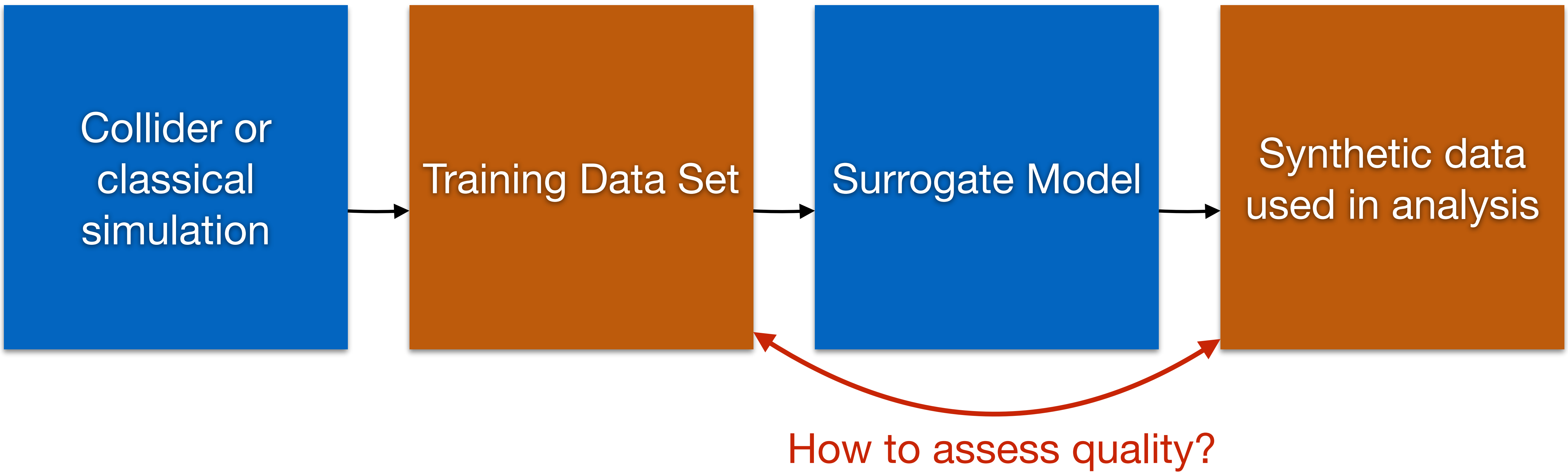
Surrogate Models





Define and manually inspect relevant distributions

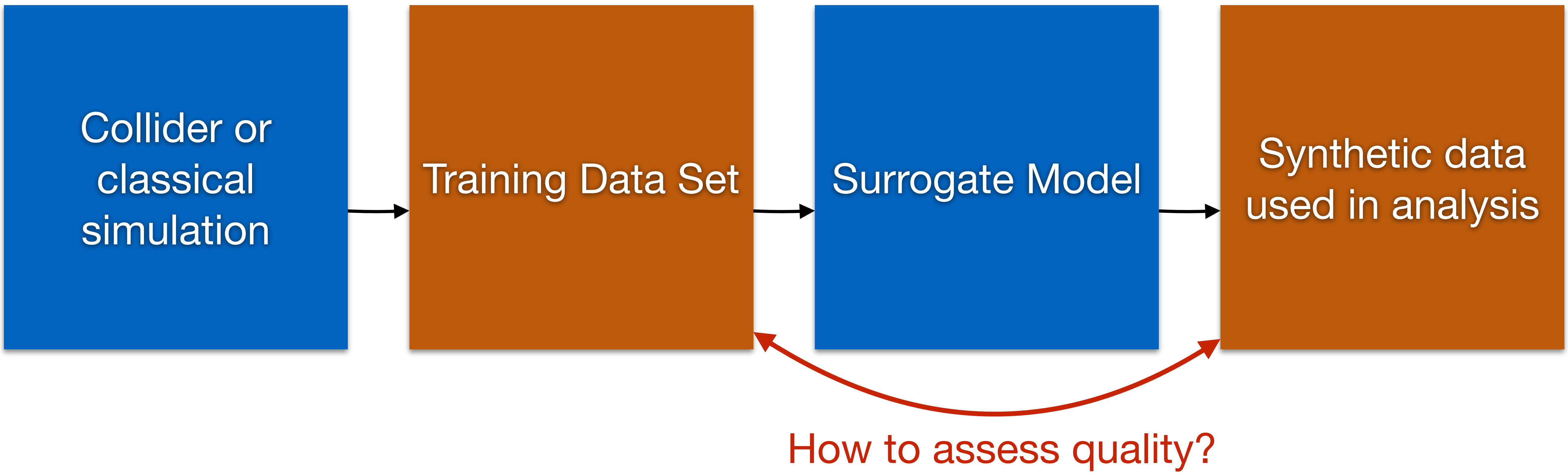
Surrogate Models



Simulator	$W_1^{N_{\text{hits}}}$ ($\times 10^{-3}$)	$W_1^{E_{\text{vis}}/E_{\text{inc}}}$ ($\times 10^{-3}$)	$W_1^{E_{\text{cell}}}$ ($\times 10^{-3}$)	$W_1^{E_{\text{long}}}$ ($\times 10^{-3}$)	$W_1^{E_{\text{radial}}}$ ($\times 10^{-3}$)	$W_1^{m_{1,X}}$ ($\times 10^{-3}$)	$W_1^{m_{1,Y}}$ ($\times 10^{-3}$)	$W_1^{m_{1,Z}}$ ($\times 10^{-3}$)
GEANT4	0.7 ± 0.2	0.8 ± 0.2	0.9 ± 0.4	0.7 ± 0.8	0.7 ± 0.1	0.9 ± 0.1	1.1 ± 0.3	0.9 ± 0.3
CALOCLOUDS	**2.5 ± 0.3**	11.4 ± 0.4	15.9 ± 0.7	**2.0 ± 1.3**	38.8 ± 1.4	4.0 ± 0.4	8.7 ± 0.3	1.4 ± 0.5
CALOCLOUDS II	3.6 ± 0.5	26.4 ± 0.4	**15.3 ± 0.6**	3.7 ± 1.6	11.6 ± 1.5	**2.4 ± 0.4**	**7.6 ± 0.2**	3.9 ± 0.4
CALOCLOUDS II (CM)	6.1 ± 0.7	**9.8 ± 0.5**	16.0 ± 0.7	**2.0 ± 1.4**	**8.3 ± 1.9**	3.0 ± 0.4	9.5 ± 0.6	**1.2 ± 0.5**

Define distributions and calculate 1D Wasserstein distance or Kullback-Leibler divergence

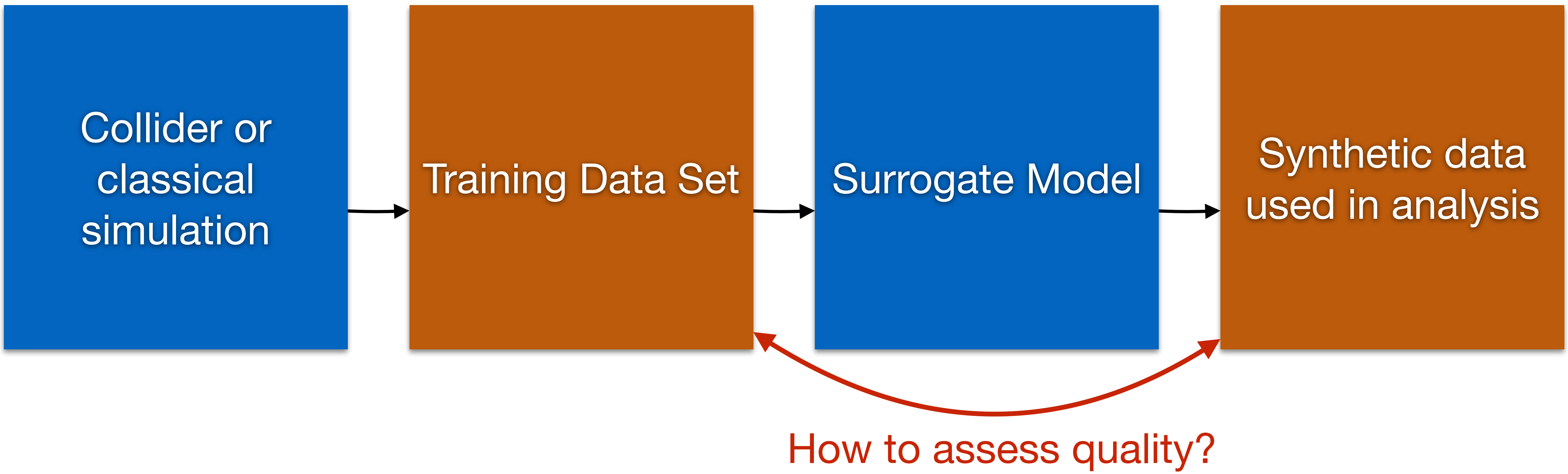
Surrogate Models



Train a classifier and look at area under curve/JSD

Dataset	Simulator	high level classifier AUC	high level classifier JSD	low level classifier AUC	low level classifier JSD
GettingHigh	L2LFlows	.634 ± .002	.047 ± .002	.905 ± .003	.438 ± .009
	BIB-AE	.903 ± .002	.436 ± .005	$\gg$ .999	.985 ± .001
CaloChallenge 3	L2LFlows	.686 ± .002	.084 ± .001	.983 ± .002	.760 ± .013

Surrogate Models



Room for a principled, holistic (for high-D data) quality measure that works for low amounts of reference data

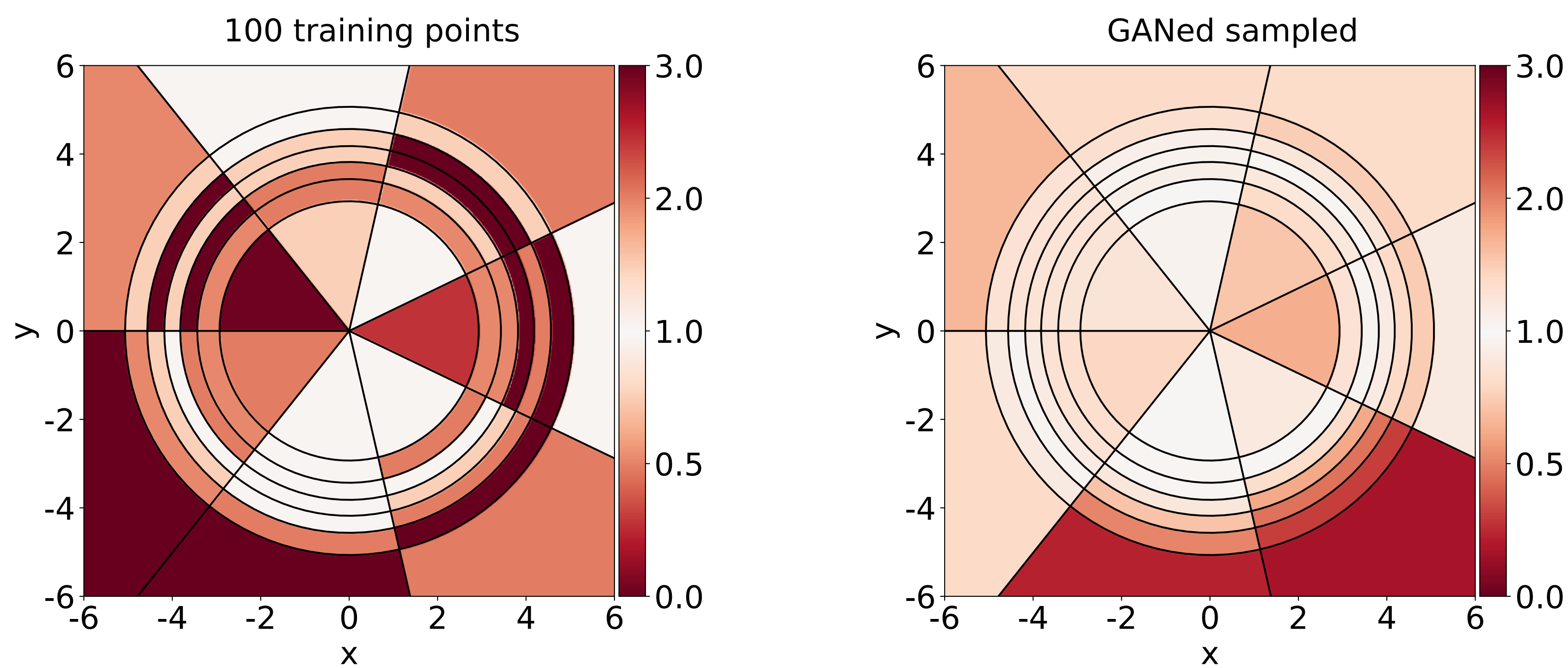
Surrogate Models

Collider or classical simulation → Training Data Set → Surrogate Model → Synthetic data used in analysis

Few samples

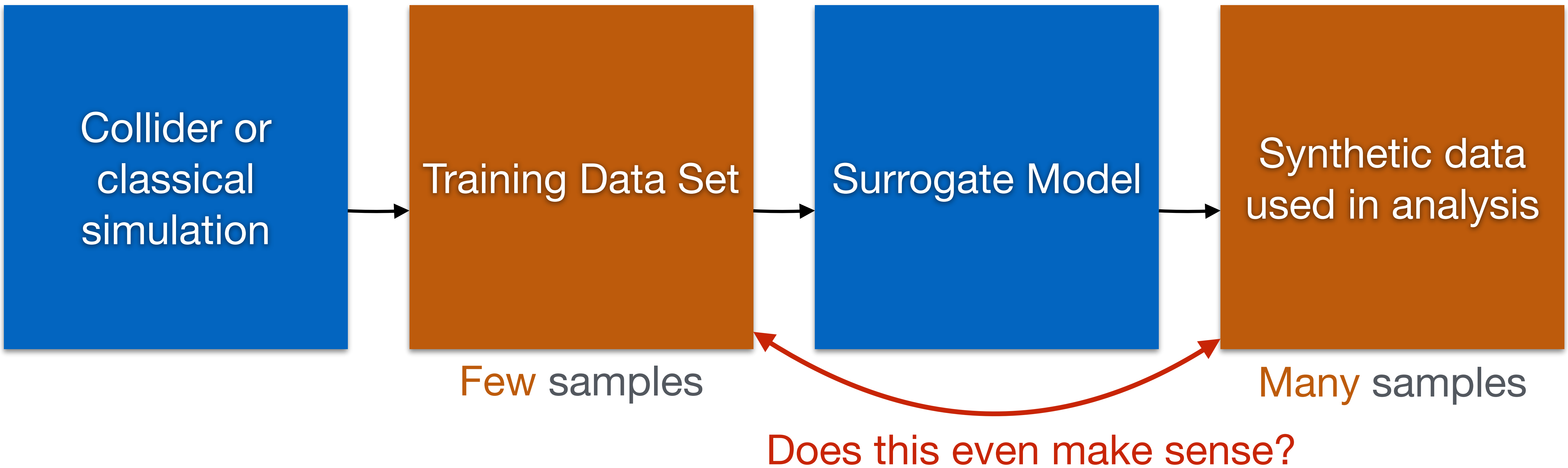
Many samples

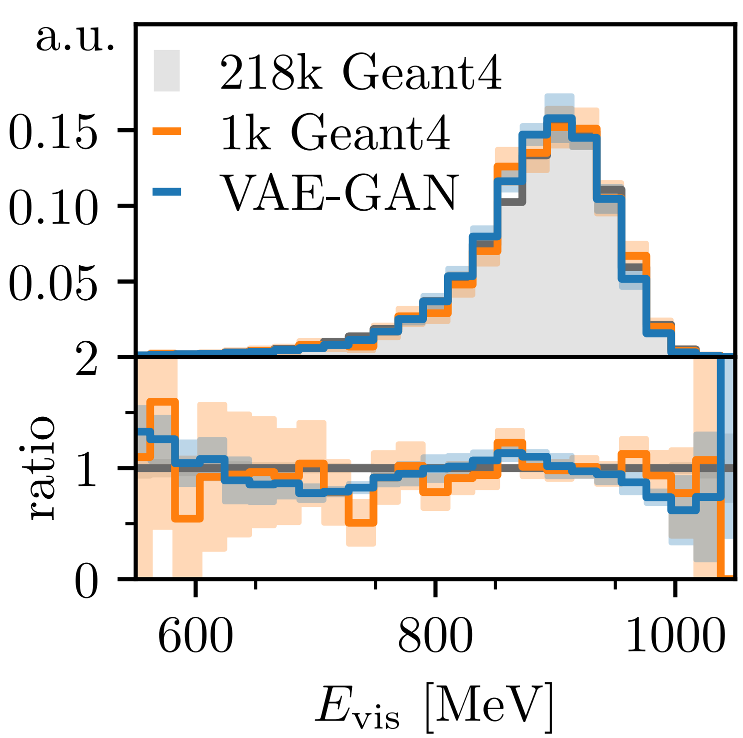
Does this even make sense?

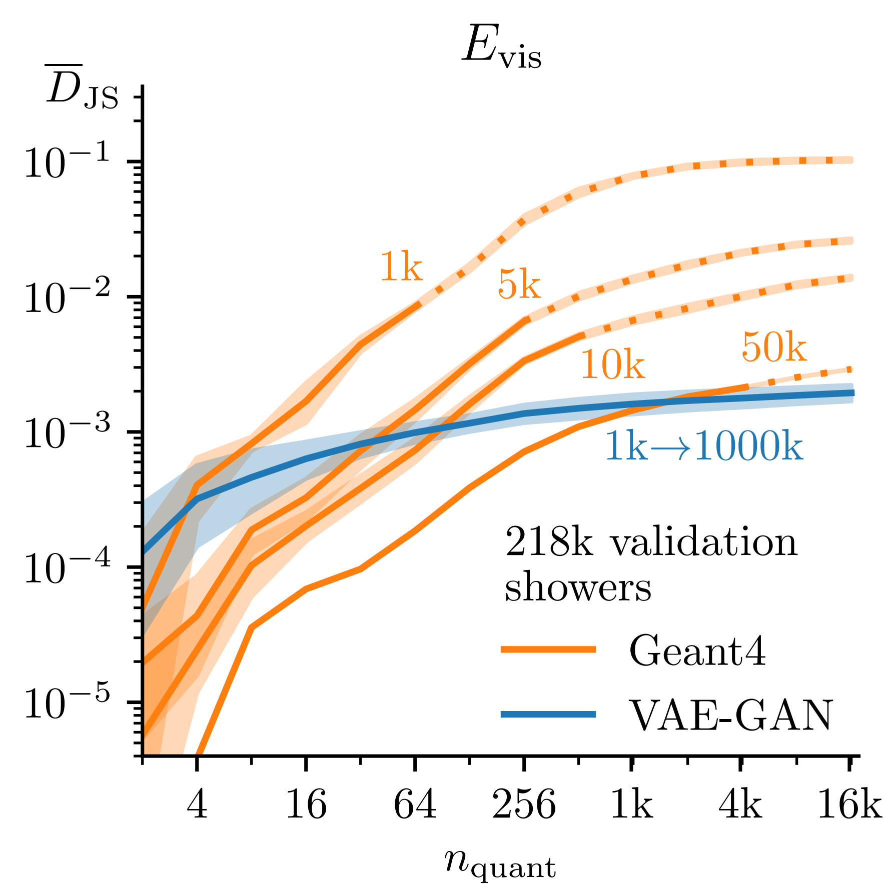


Diefenbacher, .., GK et al 2008.06545

Surrogate Models



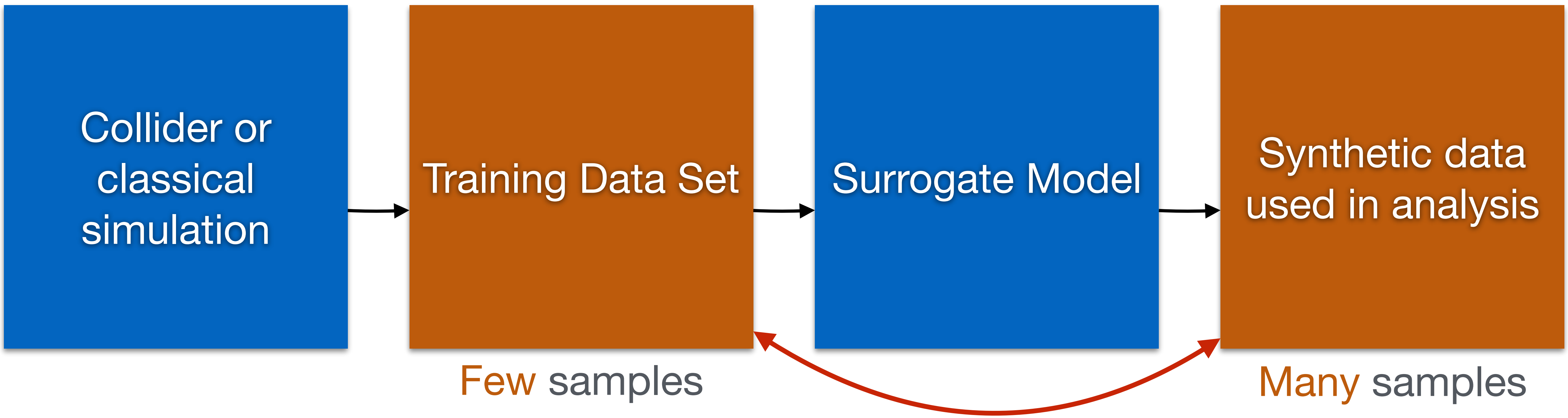




Scaling of difference to ground truth with resolution again better for the generative model.

Bieringer, .., GK et al 2202.07352

Surrogate Models



Does this even make sense?

Seems so, especially if we include mixing (multiple showers/event) and interpolation

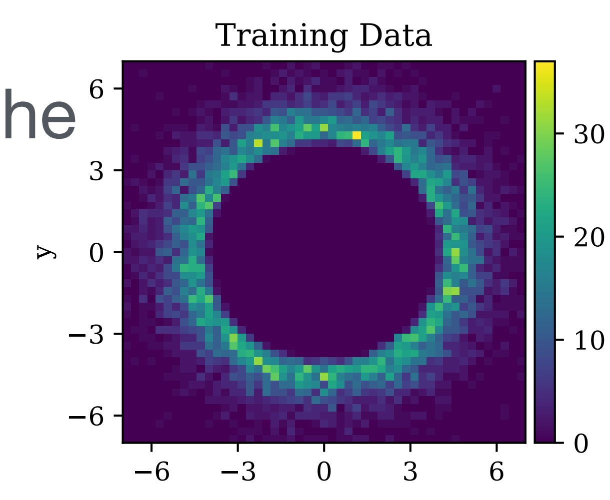
Surrogate Models

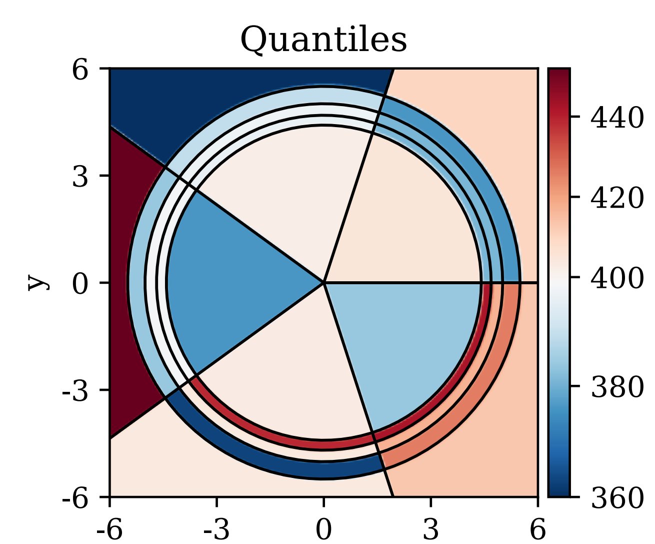
Collider or classical simulation → Training Data Set → Surrogate Model → Synthetic data used in analysis

Many samples

Can we determine statistical power of a sample without truth knowledge/ large reference?

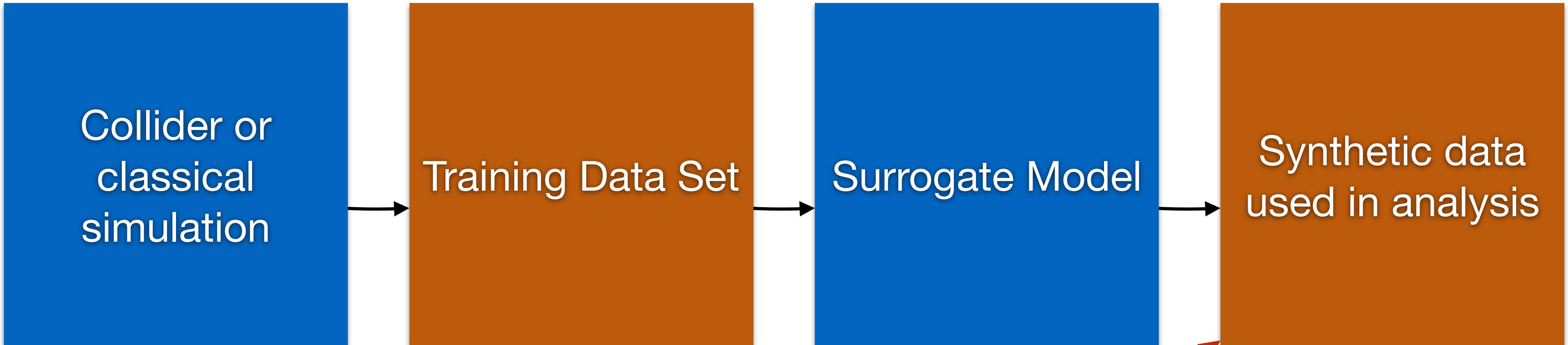
- Train a generative Bayesian (e.g. Bayes-by-backprop or AdamMCMC) network (e.g. continous normalising flow)
- Sample weights (i.e. individual models) from the Bayes model
- Sample examples from the individual models
- Compare predicted uncertainty to truth





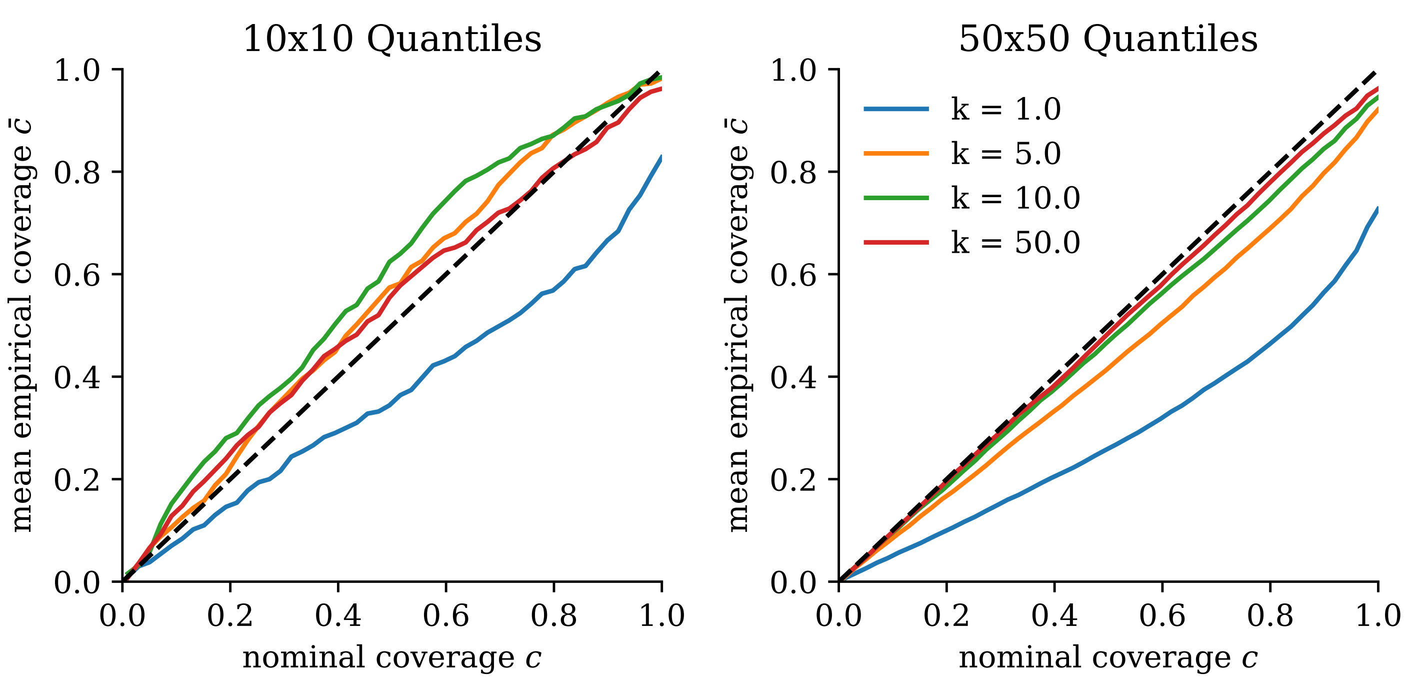
Bieringer, .., GK et al 2202.07352

Surrogate Models



Many samples

Can we determine statistical power of a sample without truth knowledge/ large reference?



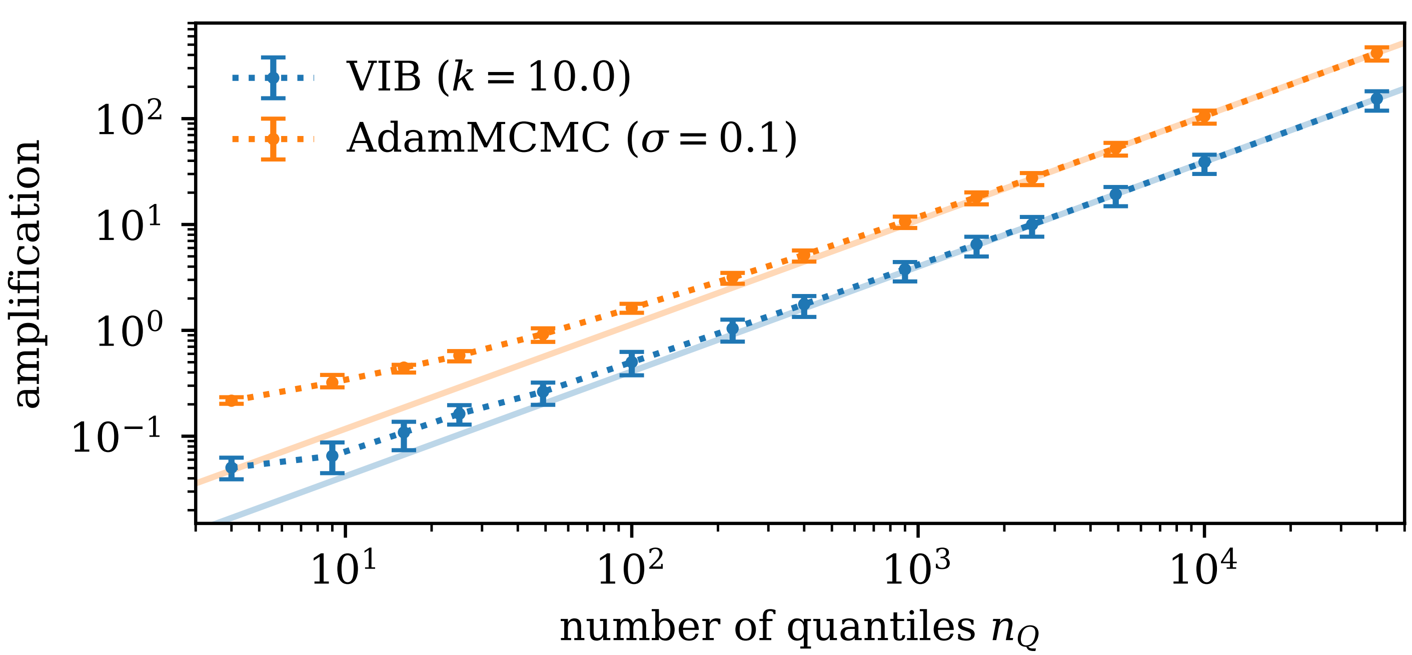
Bayesian model is well calibrated

Surrogate Models

Collider or classical simulation → Training Data Set → Surrogate Model → Synthetic data used in analysis

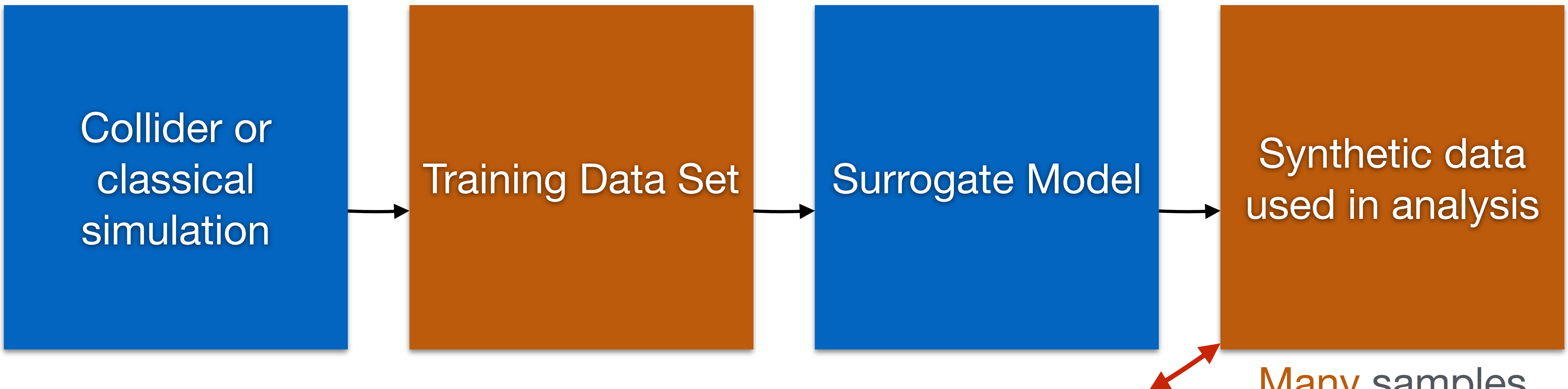
Many samples

Can we determine statistical power of a sample without truth knowledge/ large reference?



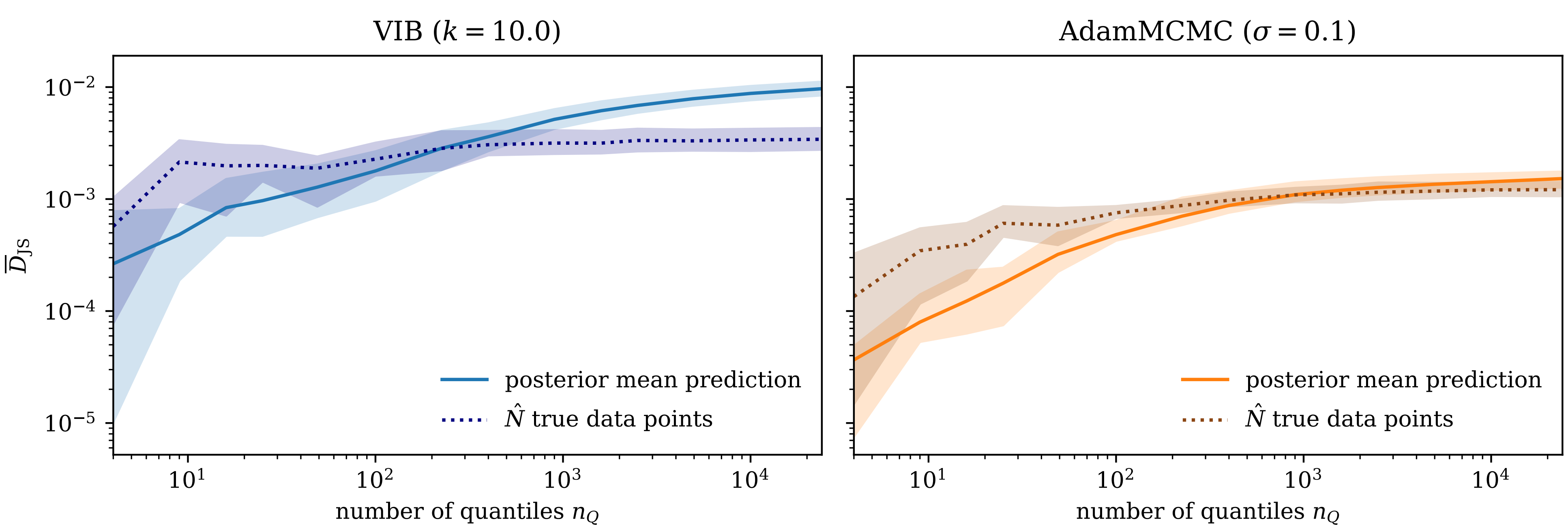
Equate Bayesian uncertainty to Poisson error to effective sample size

Surrogate Models



Many samples

Can we determine statistical power of a sample without truth knowledge/ large reference?



Agrees with JSD from independent dataset

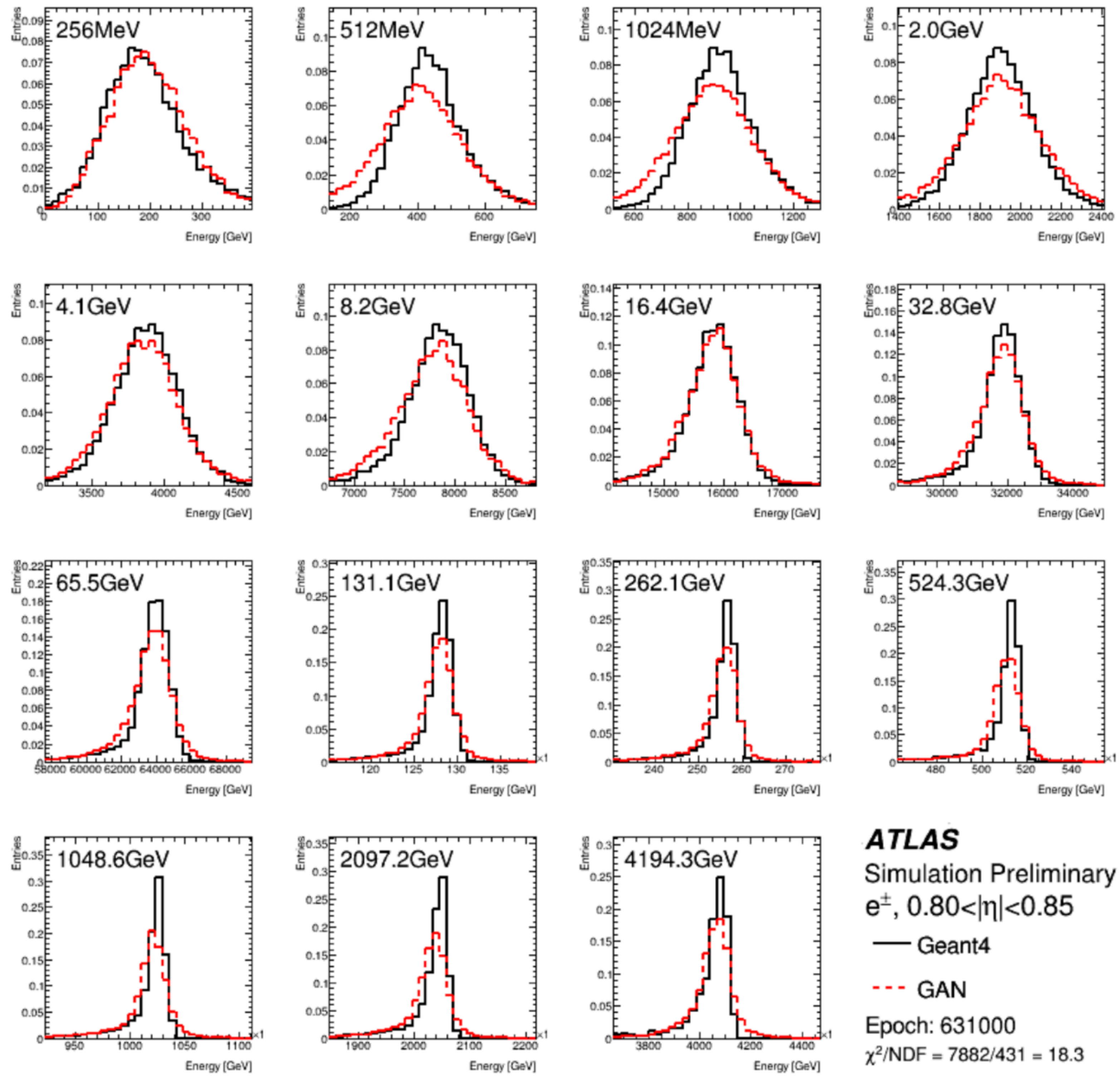
Bieringer, .., GK et al 2202.07352

Application

Not only theoretical development: e.g. ATLAS includes FastCaloGAN in ATLFAST3

100 networks (slices in η)

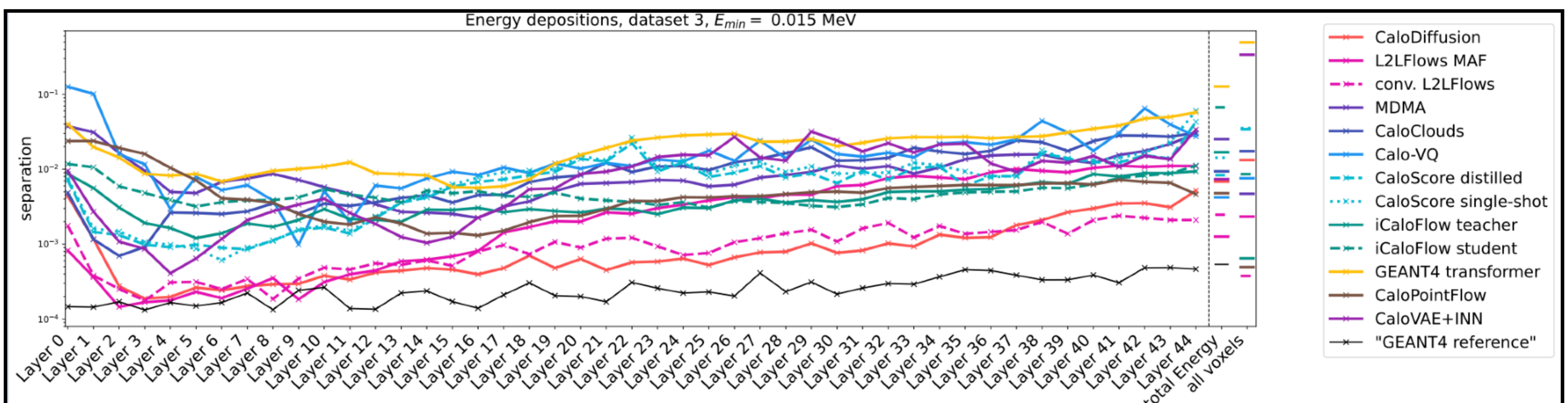
O(500) voxels



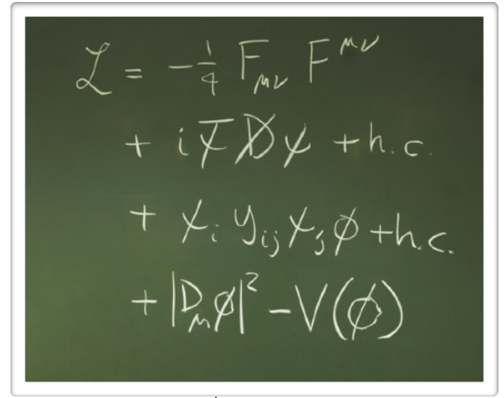
ATL-SOFT-PUB-2020-006; ATLAS 2109.02551

Moving forward

- 3 Public datasets to compare simulation techniques
  - Simplest: ATLAS dataset (see prev. page)
  - Most complex: Future detector with 40k voxels
  - Write-up currently ongoing



https://calochallenge.github.io/homepage/; Krause at ML4Jets '23



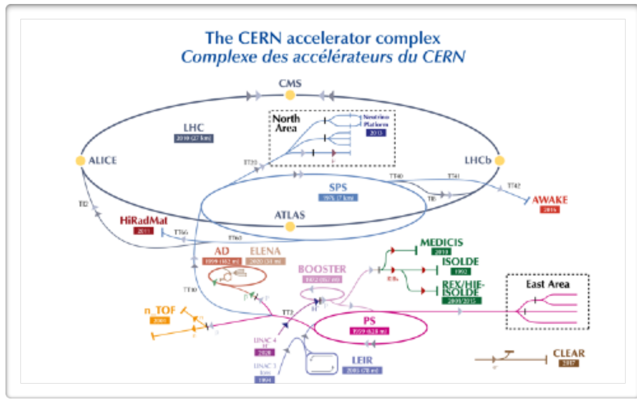
Inference

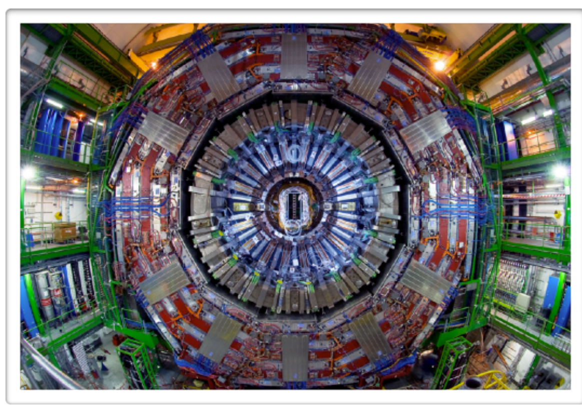
Experiment Design

AI

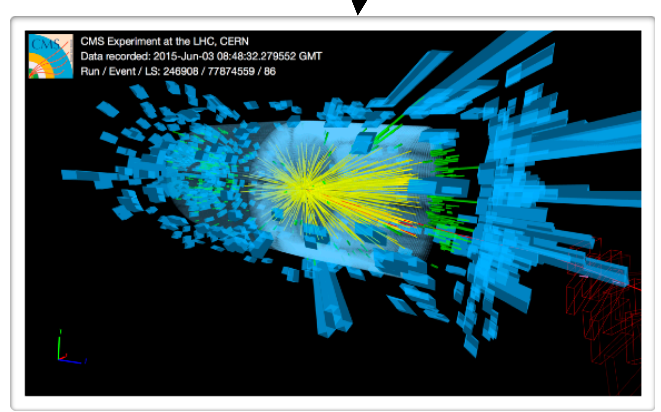
Simulation

Triggers

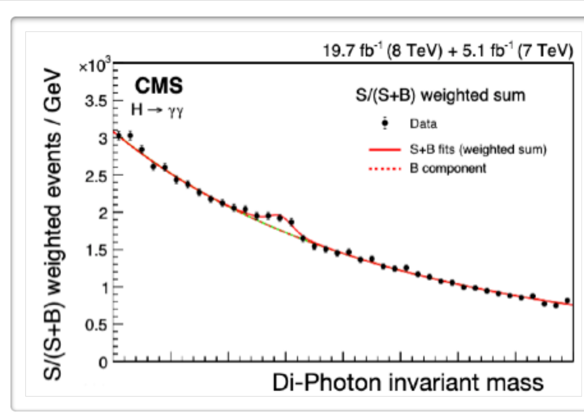




Tagging
Reconstruction



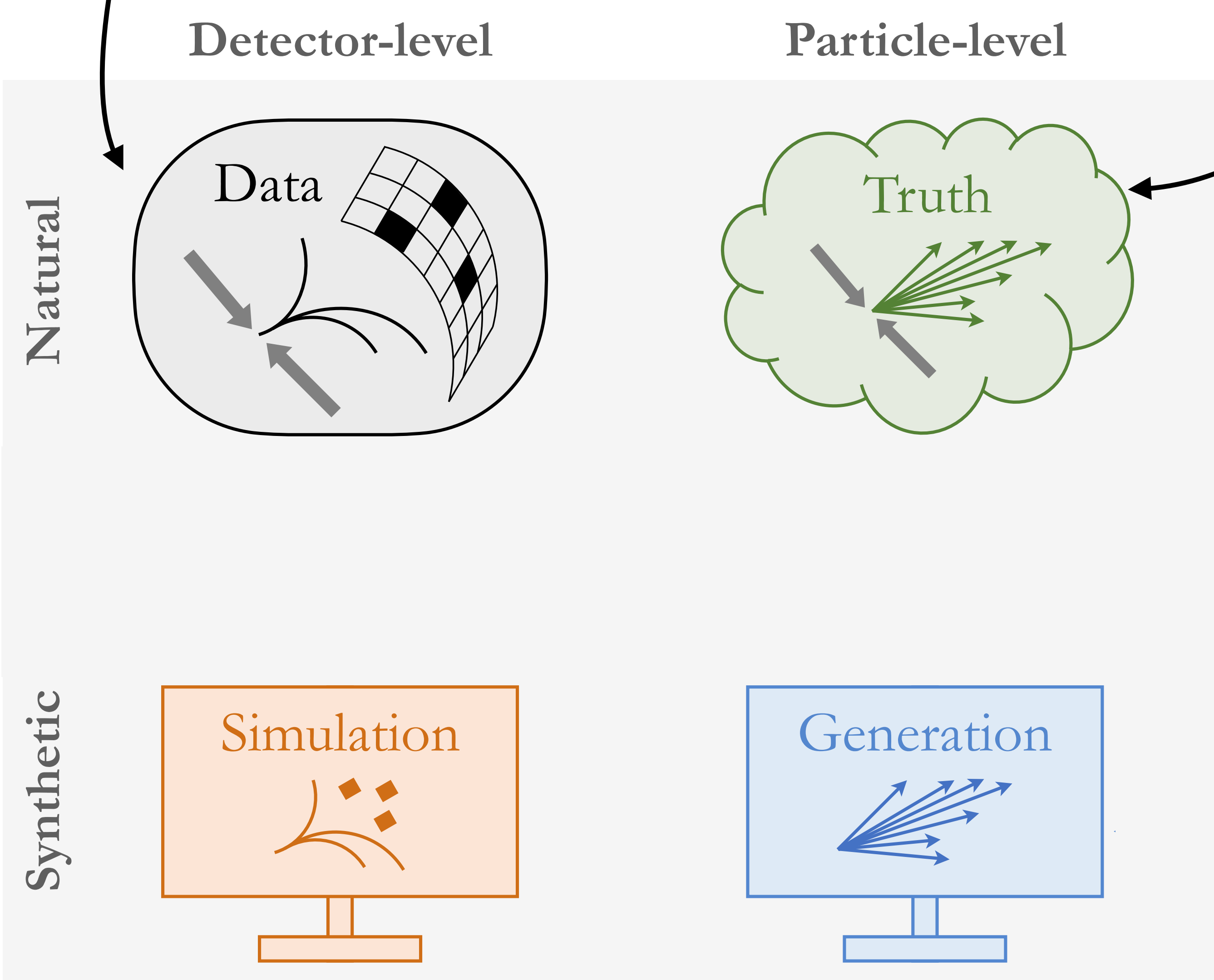
Unfolding
Anomaly Detection



See talk by **Vinicius**

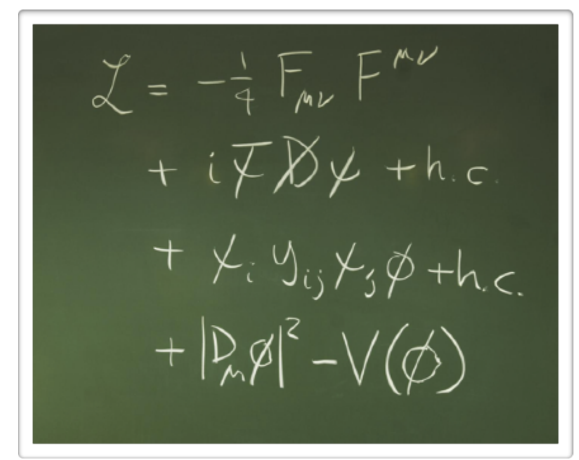
Unfolding

Observe



Want to know

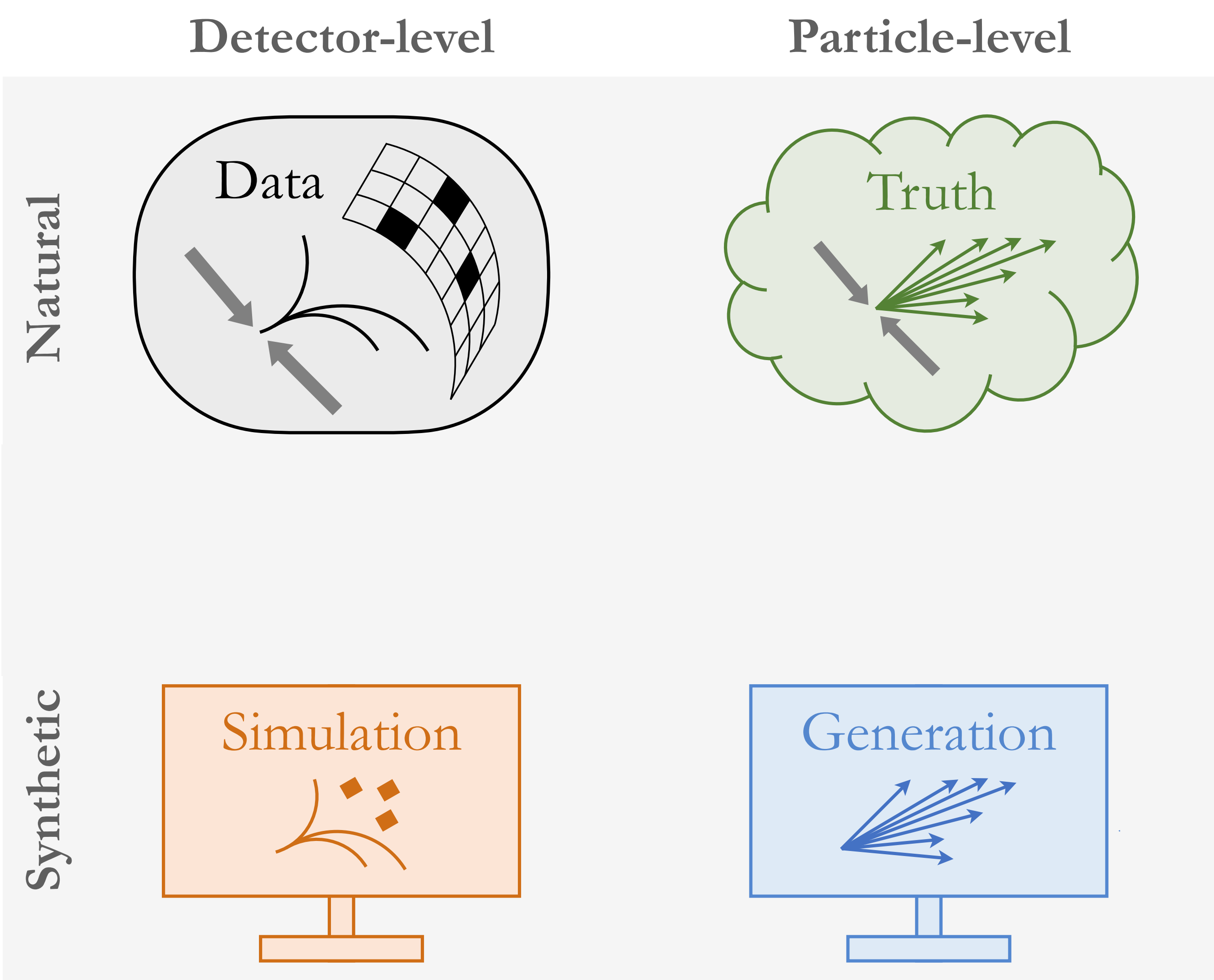
(and relate to fundamental theory)



Synthetic data provides both views: How to use?

Andreassen et el 1911.09107;

Unfolding

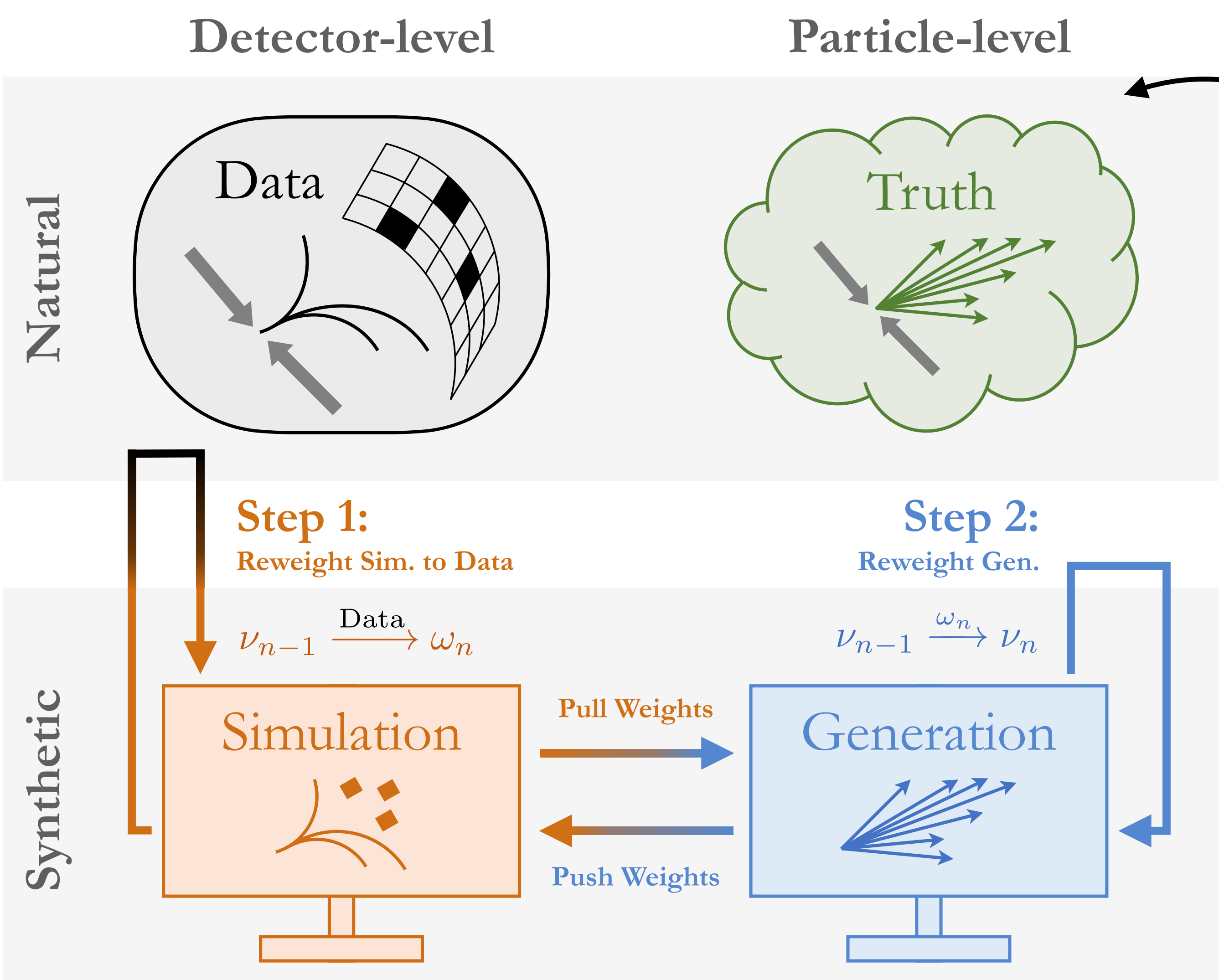


2 key approaches:

- Reweighting based on classifiers
- Morphing based on diffusion or generative models

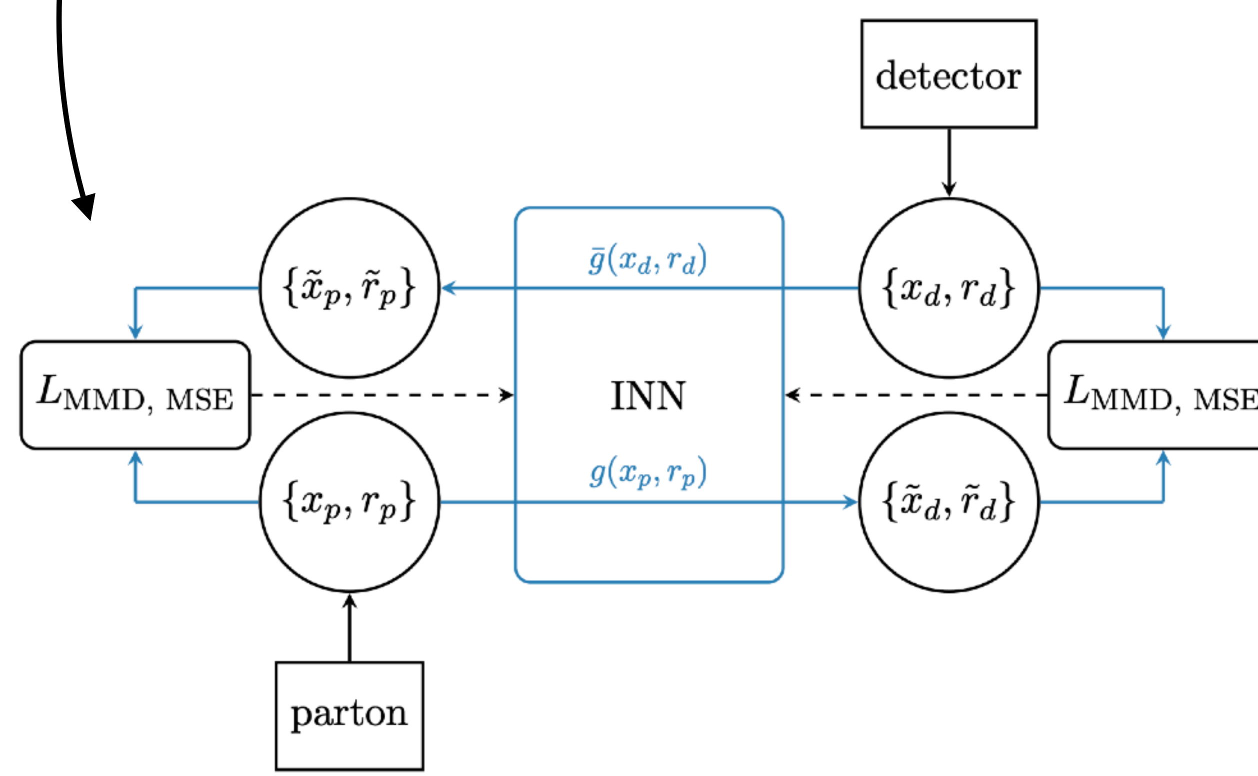
Andreassen et el 1911.09107; Huetsch et al 2404.18807

Unfolding



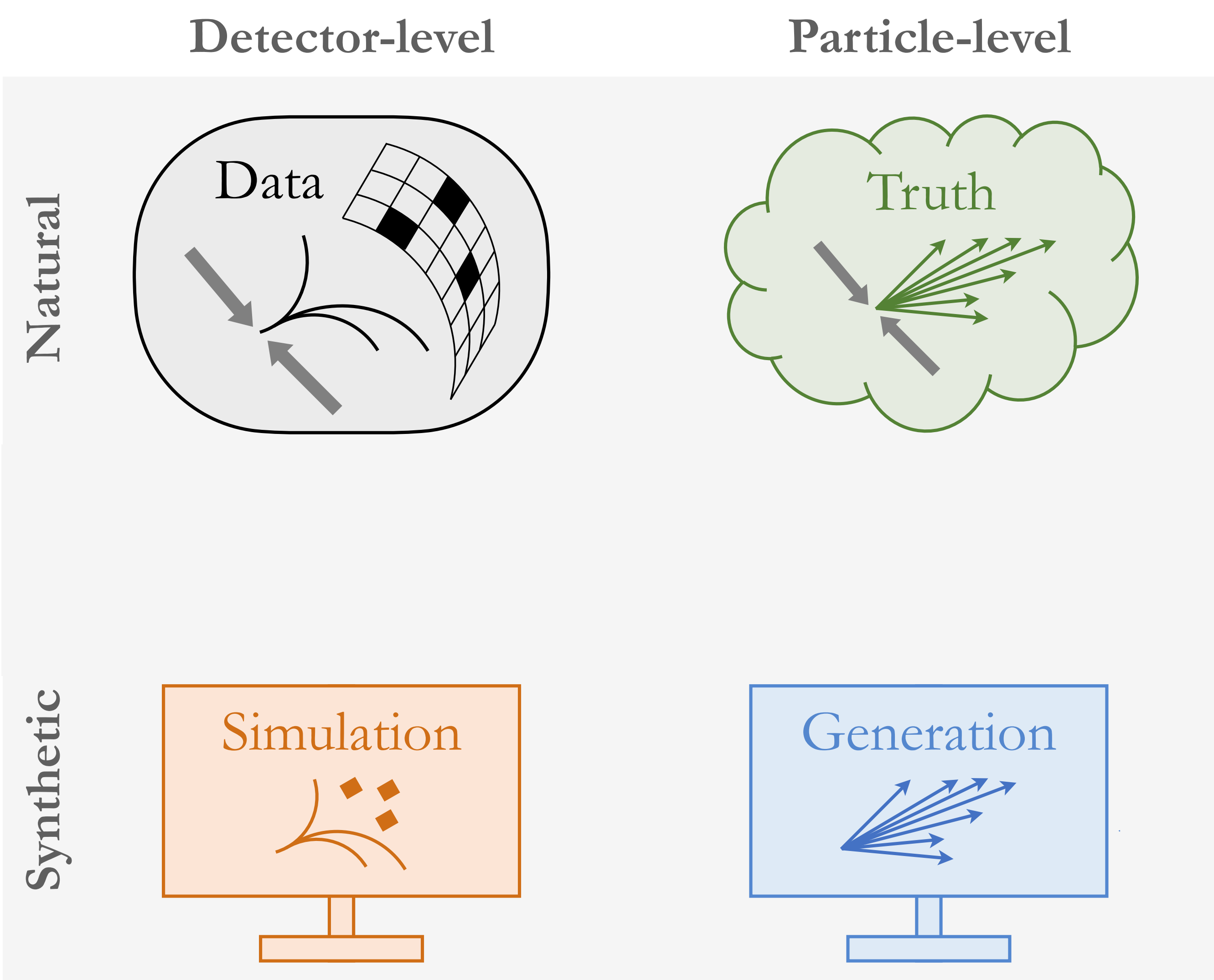
2 key approaches:

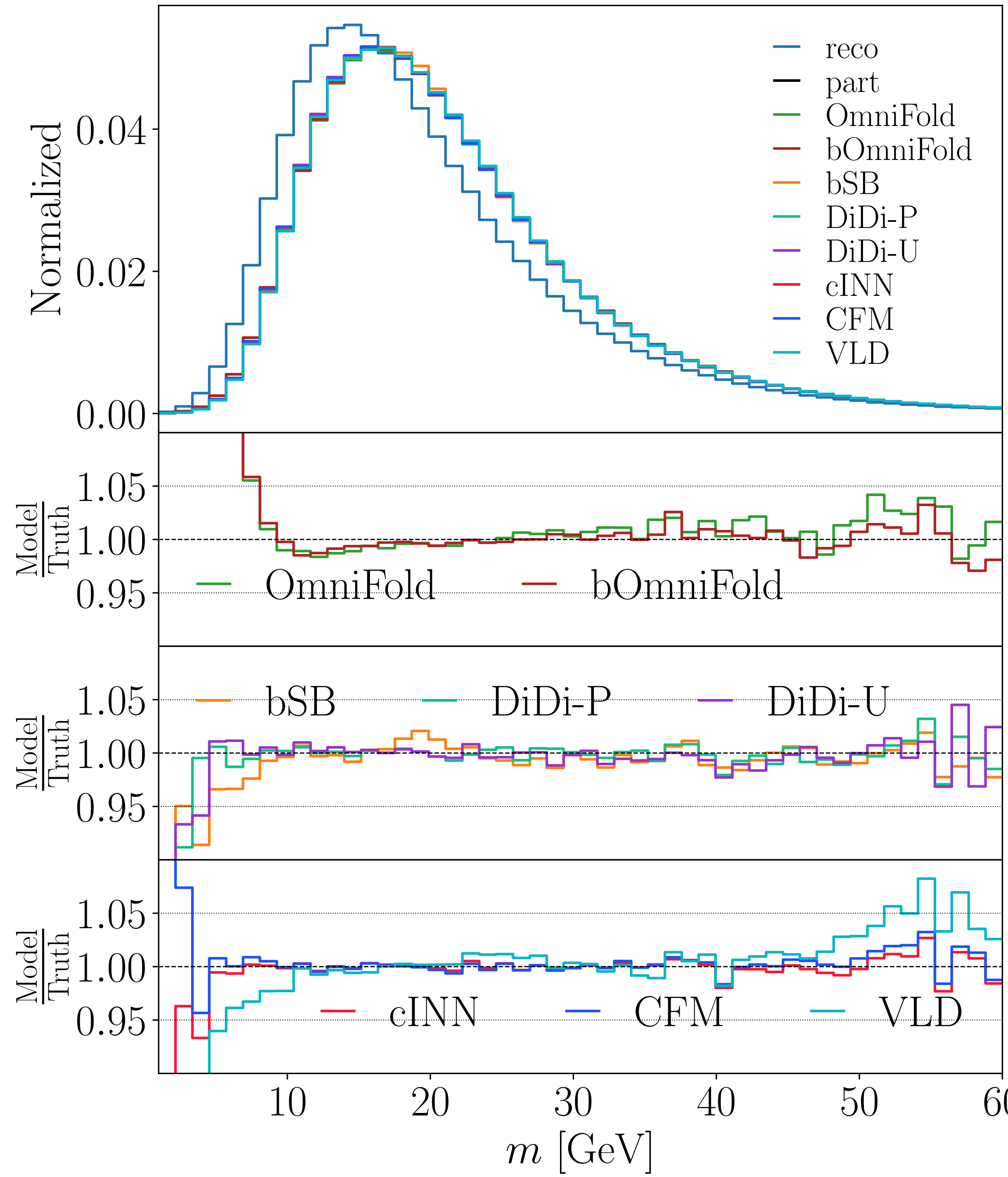
- Reweighting based on classifiers
- Morphing based on diffusion or generative models



Andreassen et el 1911.09107; Bellagente, .., GK, et al 2006.06685; Huetsch et al 2404.18807

Unfolding



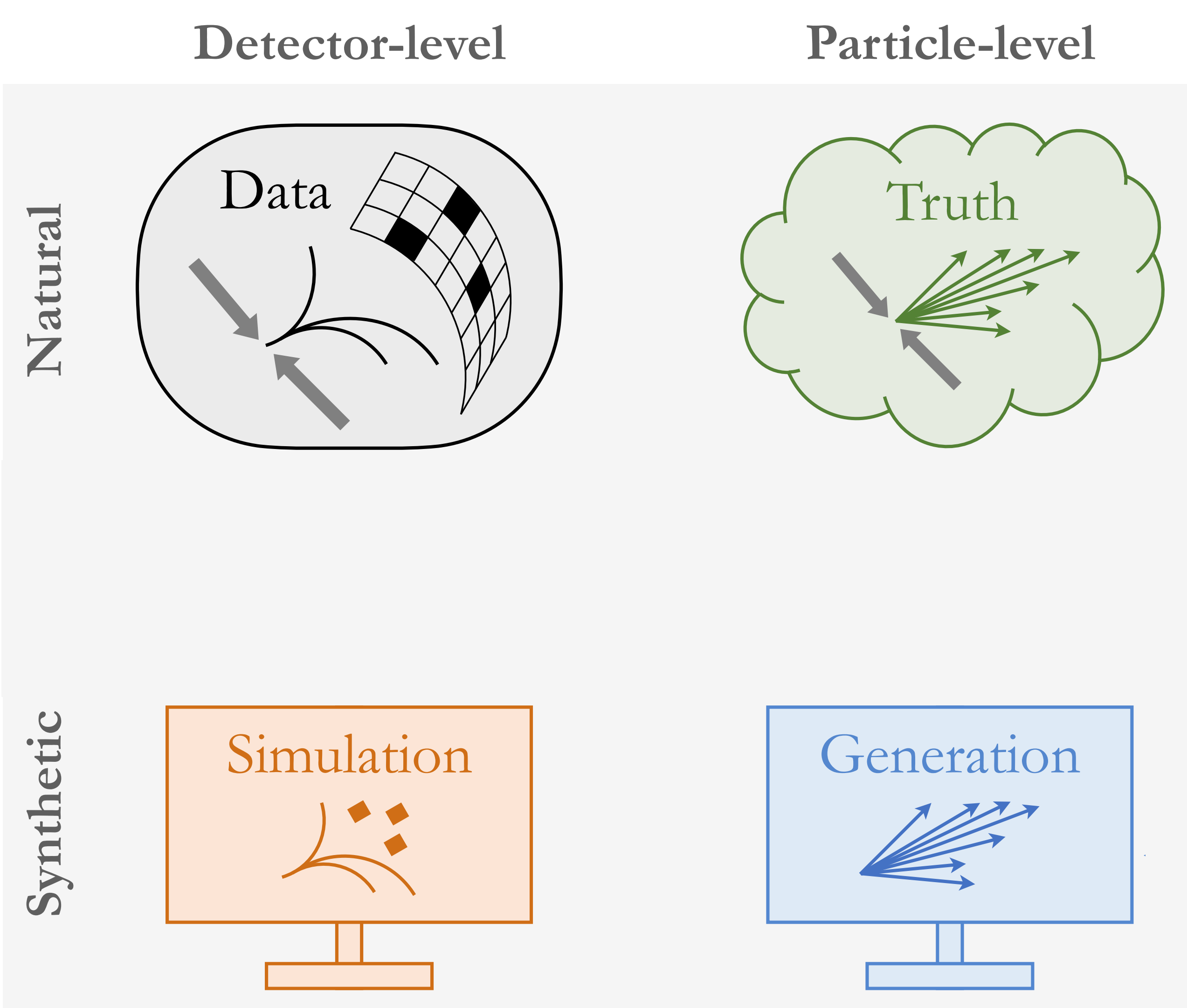


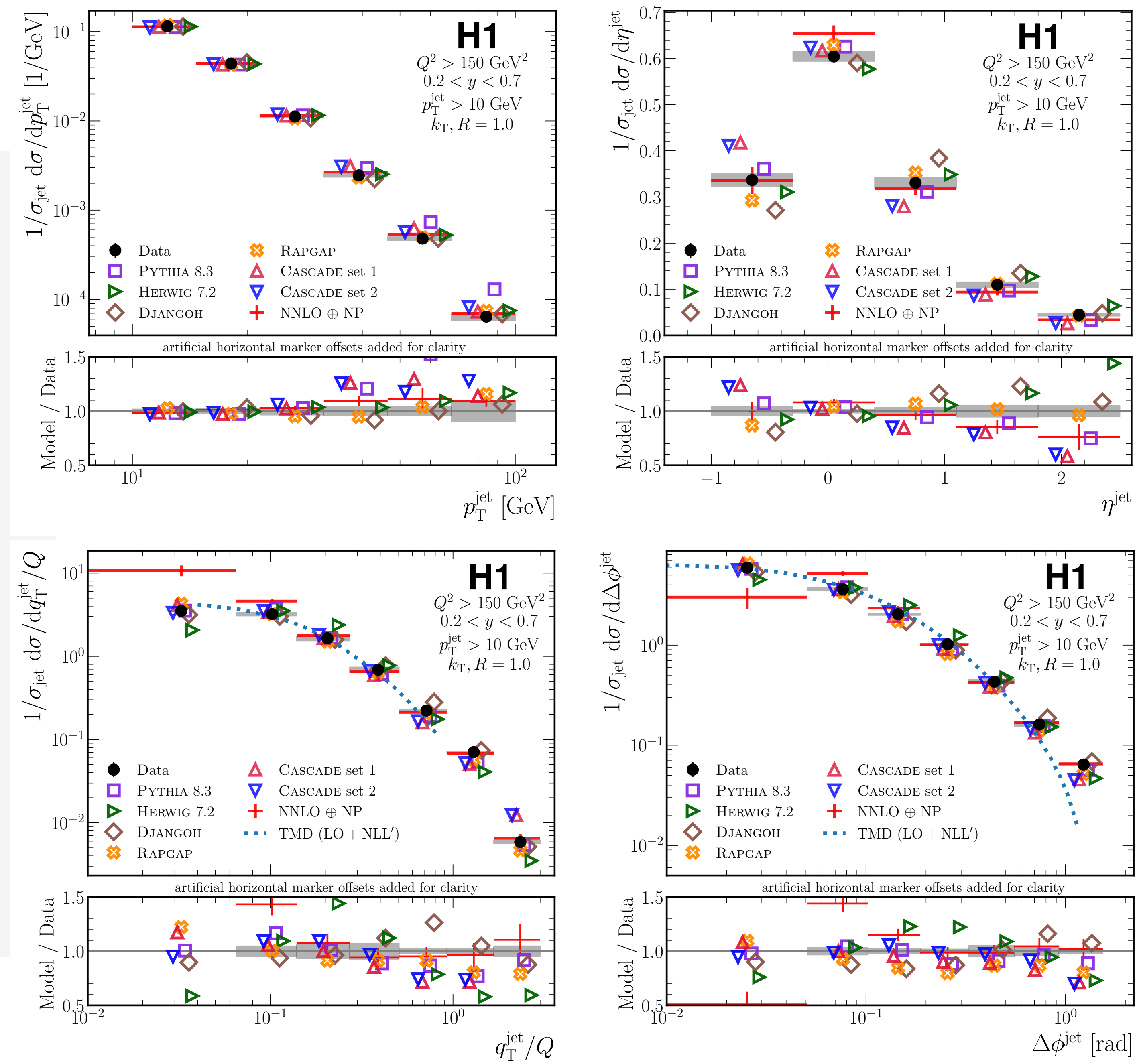
See talk by **Tilman** and **Vinicius**

Example: Unfold Z+jets distributions in six dimensions

Andreassen et el 1911.09107; Huetsch et al 2404.18807

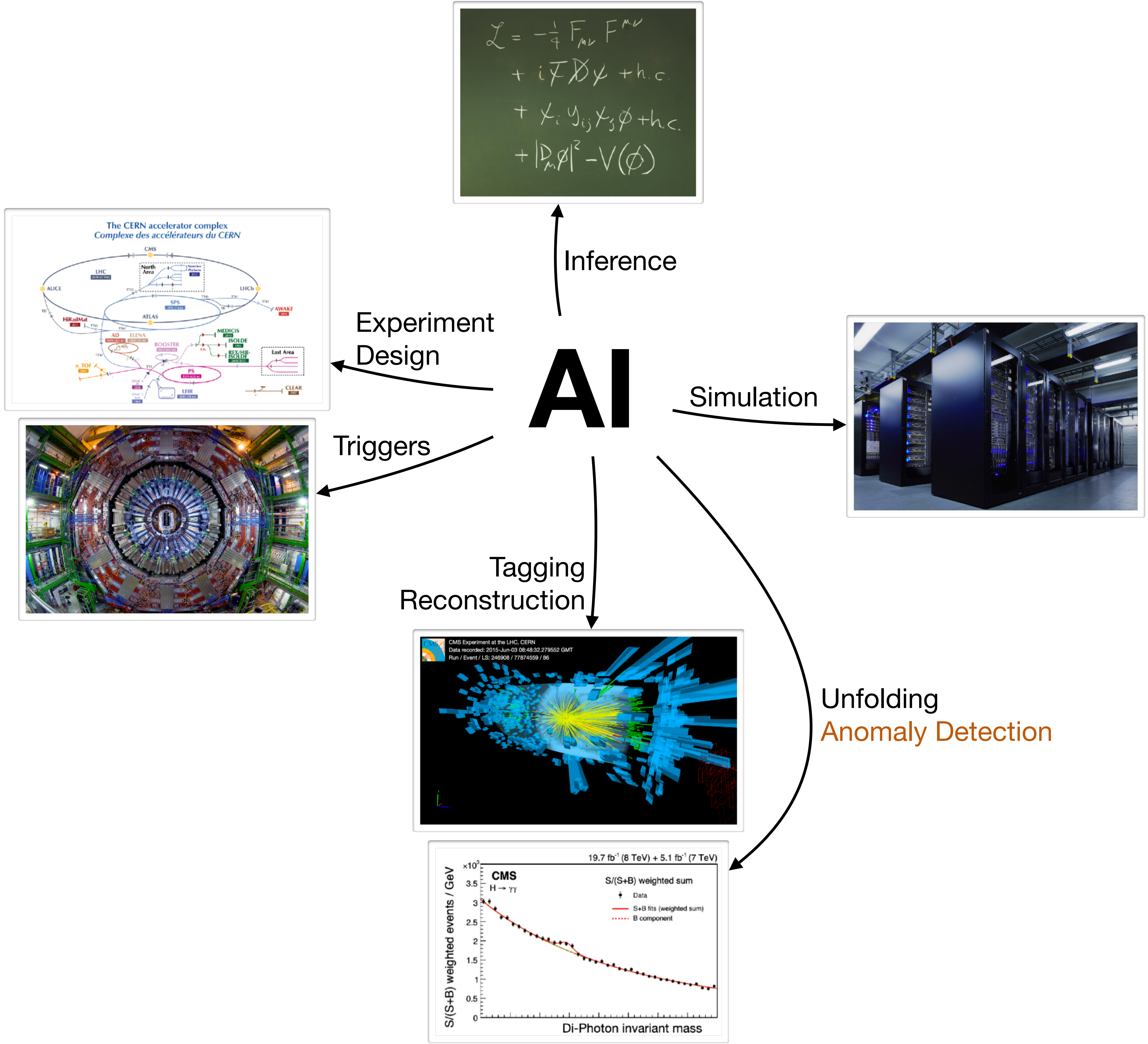
Unfolding





Already applied to collider data:
Multifold on lepton/jet events at H1

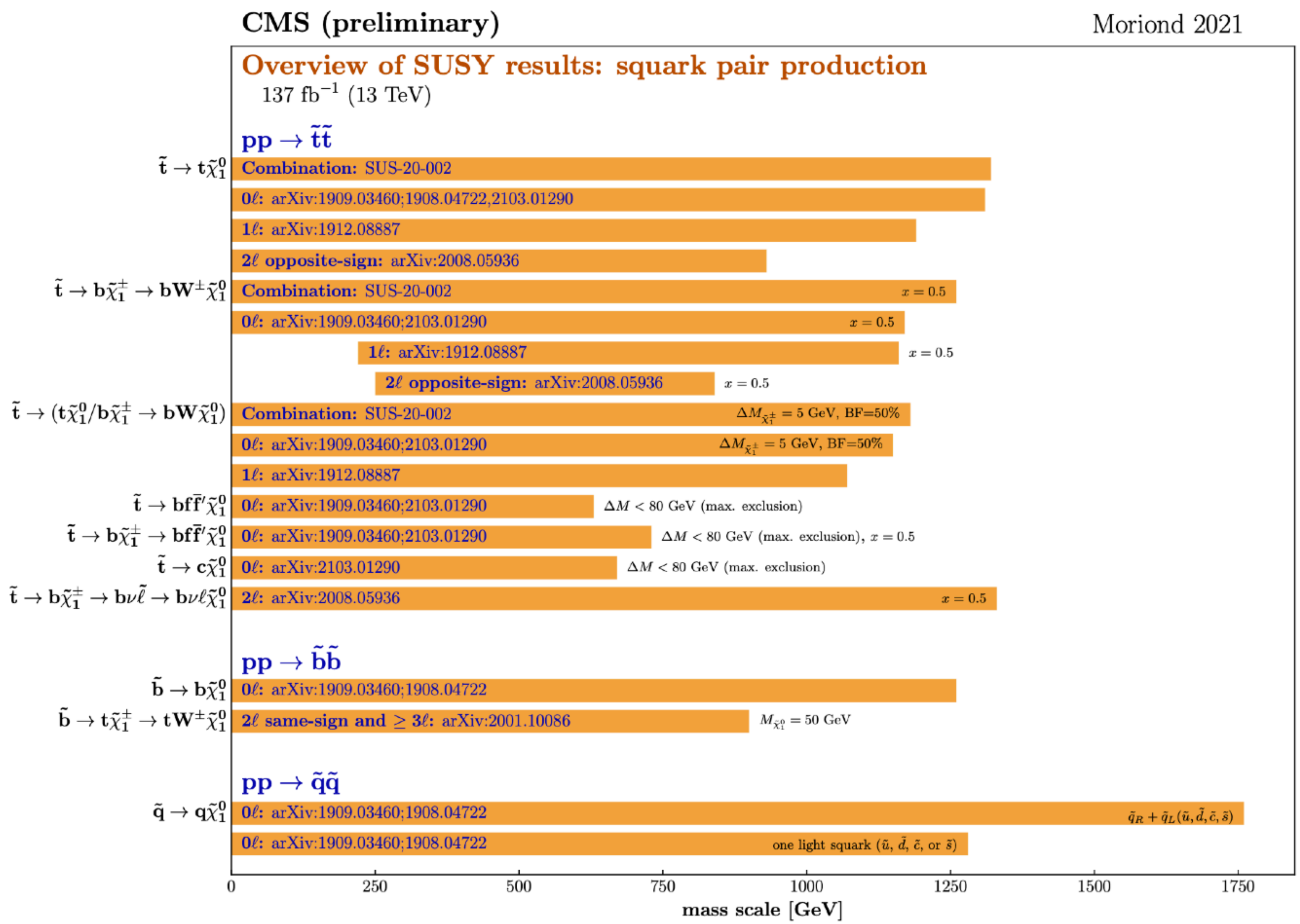
Andreassen et el 1911.09107; H1 2108.12376



AI
Inference
Experiment Design
Simulation
Triggers
Tagging
Reconstruction
Unfolding
Anomaly Detection
The CERN accelerator complex
Complexe des accélérateurs du CERN
CMS Experiment at the LHC, CERN
Data recorded: 2015-Jun-03 08:48:32.279552 GMT
Run / Event / LS: 246908 / 77874559 / 86
19.7 fb^{-1} (8 TeV) + 5.1 fb^{-1} (7 TeV)
CMS
H → γγ
S/(S+B) weighted sum
Data
S+B fits (weighted sum)
B component
S/(S+B) weighted events / GeV
Di-Photon invariant mass

Motivation

- Expect physics beyond the Standard Model
- Only negative results in searches
- Two discovery strategies:
  - Model-specific
  - Model independent
- Trade off:
  Sensitivity to
  specific model vs
  broad coverage



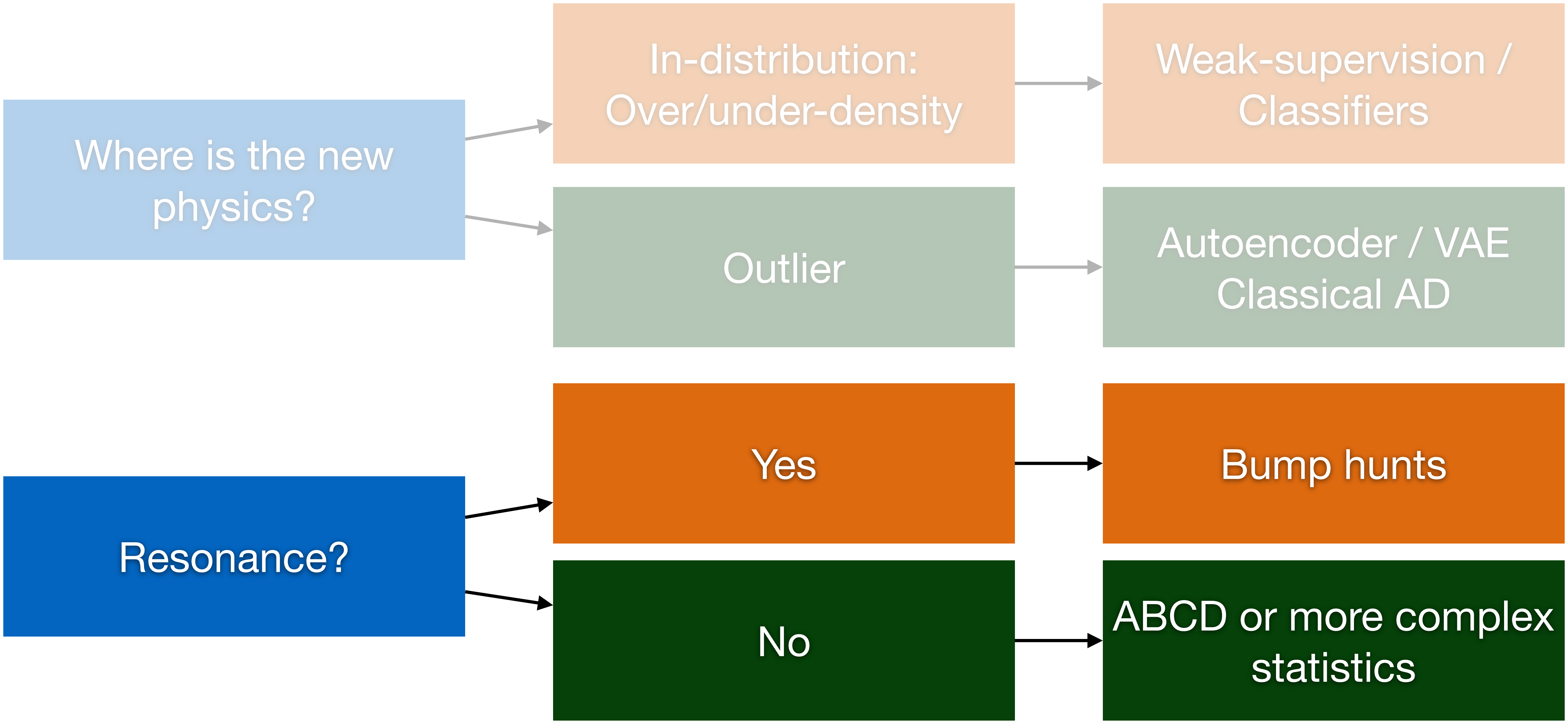
Selection of observed limits at 95% C.L. (theory uncertainties are not included). Probe **up to** the quoted mass limit for light LSPs unless stated otherwise. The quantities ΔM and x represent the absolute mass difference between the primary sparticle and the LSP, and the difference between the intermediate sparticle and the LSP relative to ΔM, respectively, unless indicated otherwise.

Recent review: 2312.14190

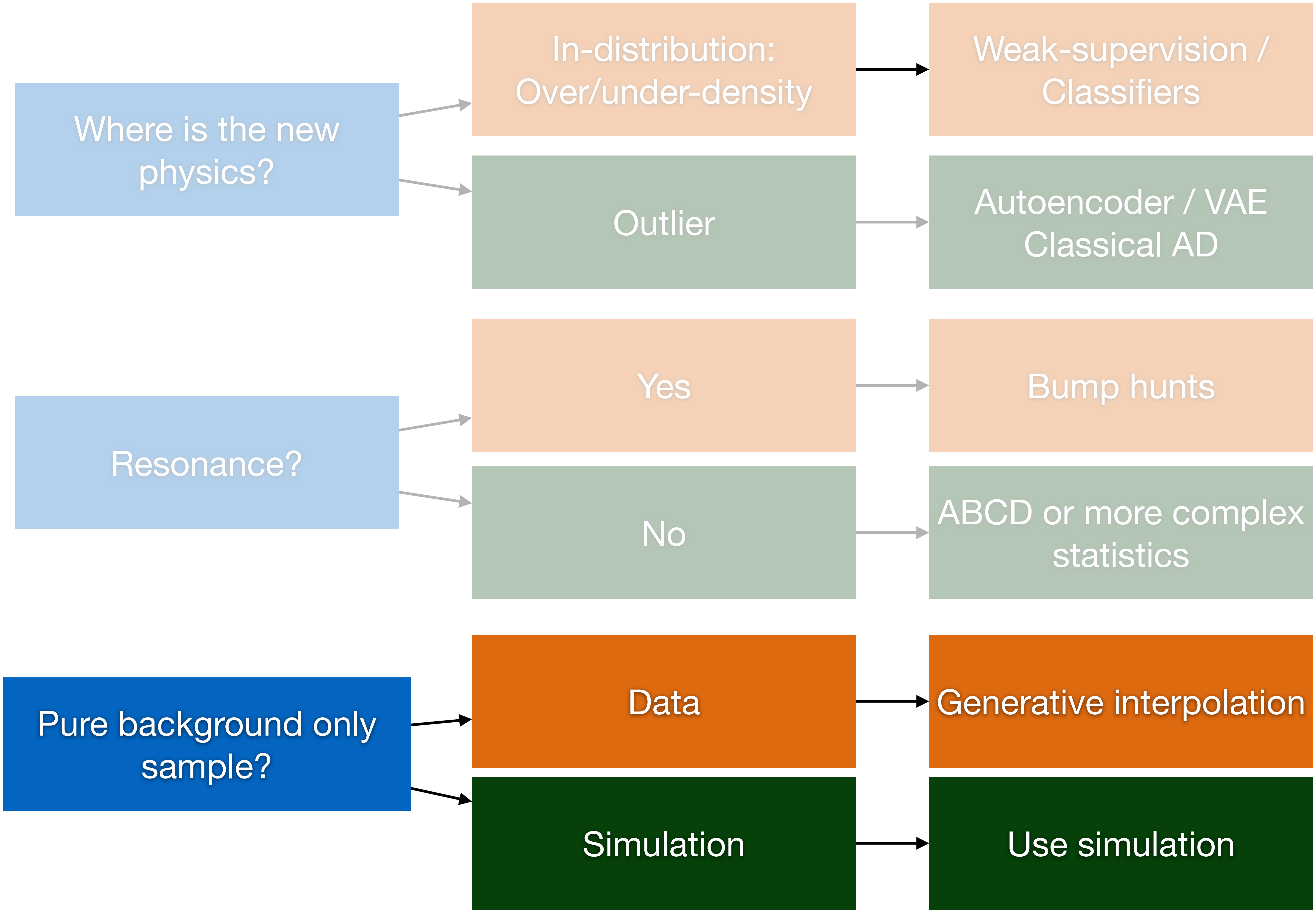
Choices


Where is the new physics?
In-distribution: Over/under-density
Outlier
Weak-supervision / Classifiers
Autoencoder / VAE Classical AD

Choices


Where is the new physics?
In-distribution: Over/under-density
Weak-supervision / Classifiers
Outlier
Autoencoder / VAE Classical AD
Resonance?
Yes
Bump hunts
No
ABCD or more complex statistics

Choices


Where is the new physics?
In-distribution: Over/under-density
Weak-supervision / Classifiers
Outlier
Autoencoder / VAE Classical AD
Resonance?
Yes
Bump hunts
No
ABCD or more complex statistics
Pure background only sample?
Data
Generative interpolation
Simulation
Use simulation

More Choices

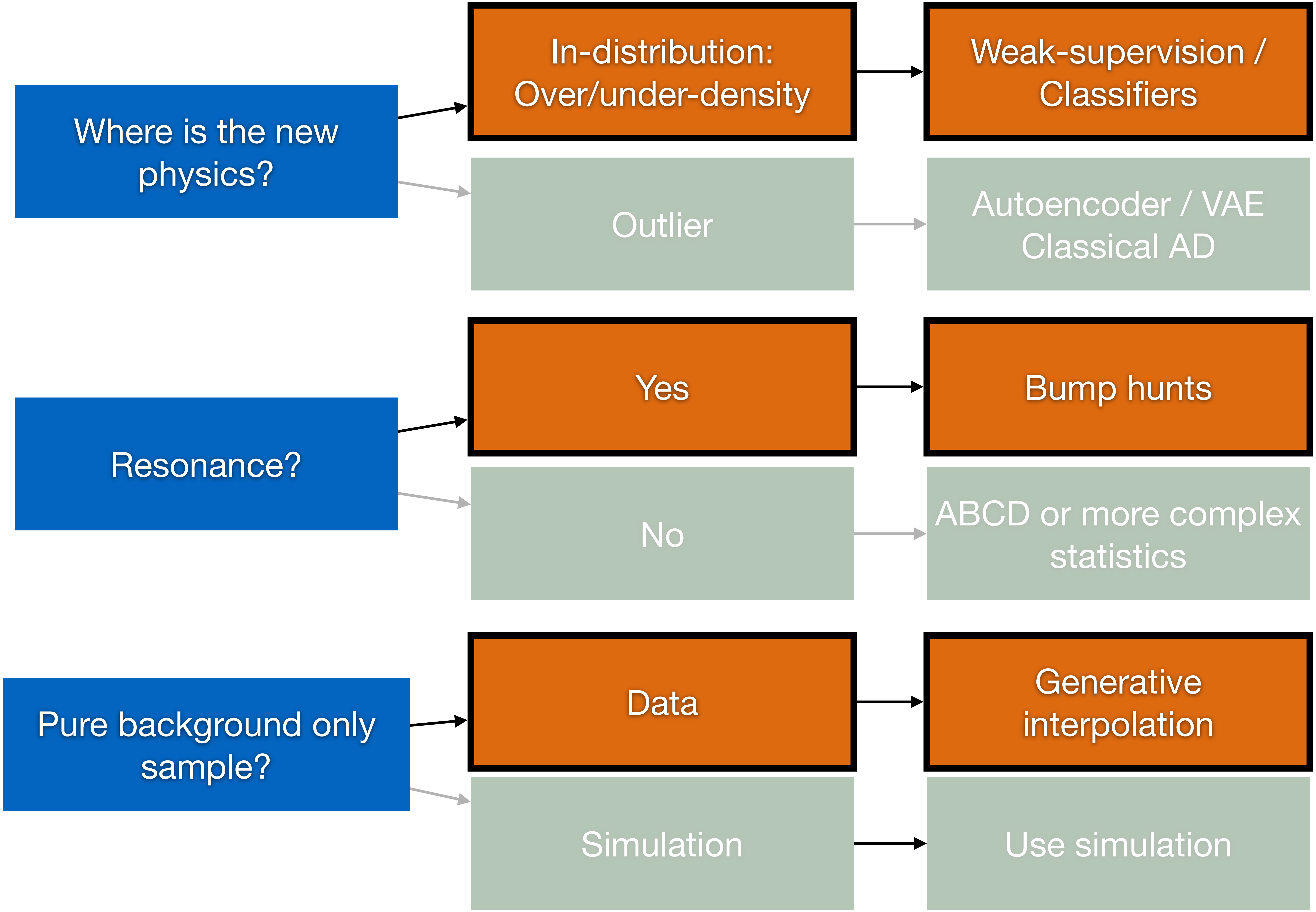
Tails?

Feature Dimension?

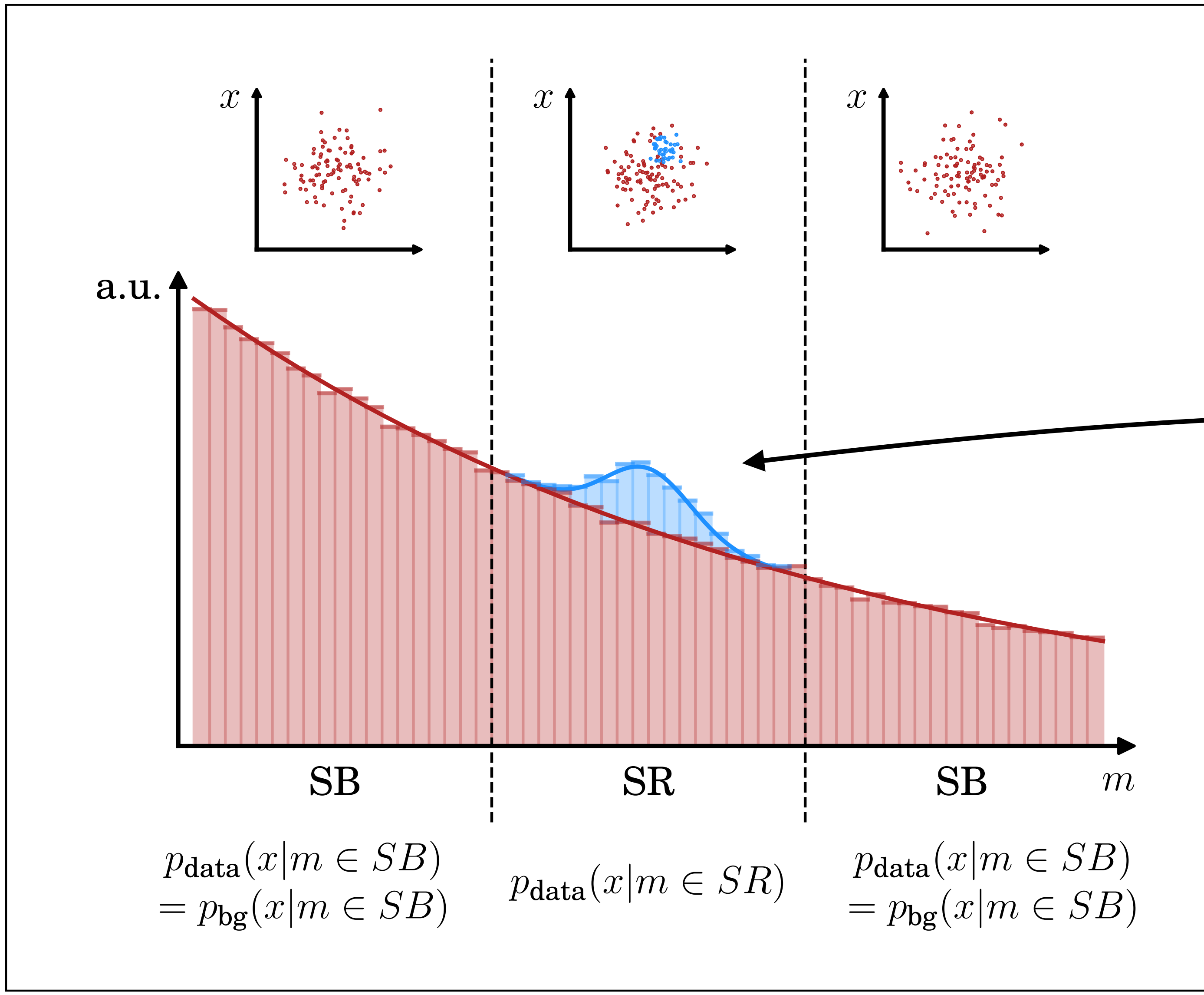
Interpolation approach?

...

Choices

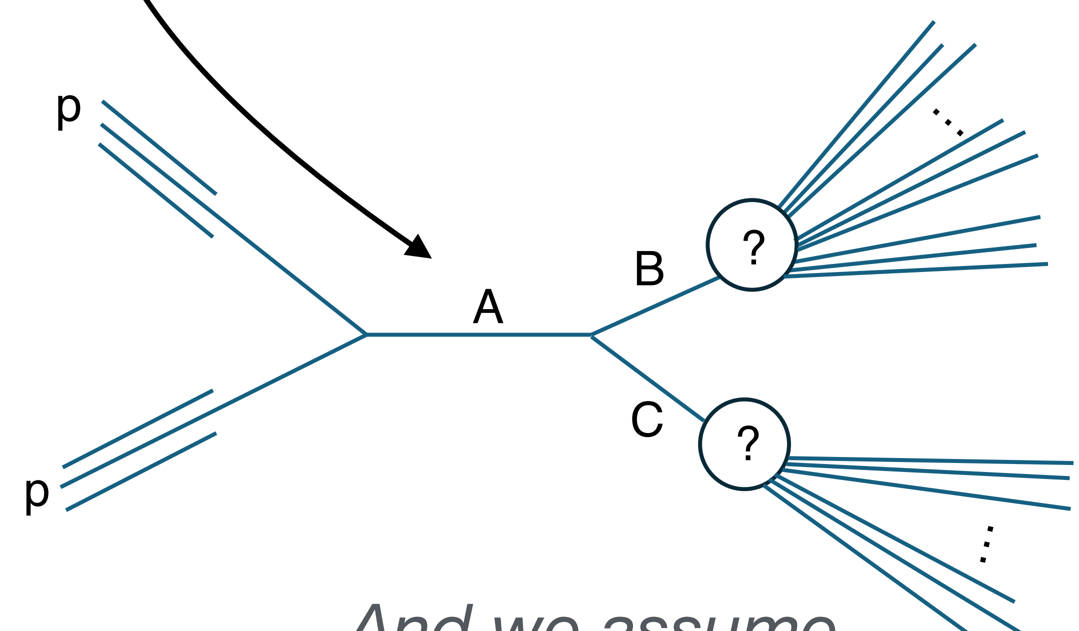

Where is the new physics?
In-distribution: Over/under-density
Weak-supervision / Classifiers
Outlier
Autoencoder / VAE Classical AD
Resonance?
Yes
Bump hunts
No
ABCD or more complex statistics
Pure background only sample?
Data
Generative interpolation
Simulation
Use simulation

CATHODE



Consider resonant anomalies: fully data-based construction of anomaly detection score

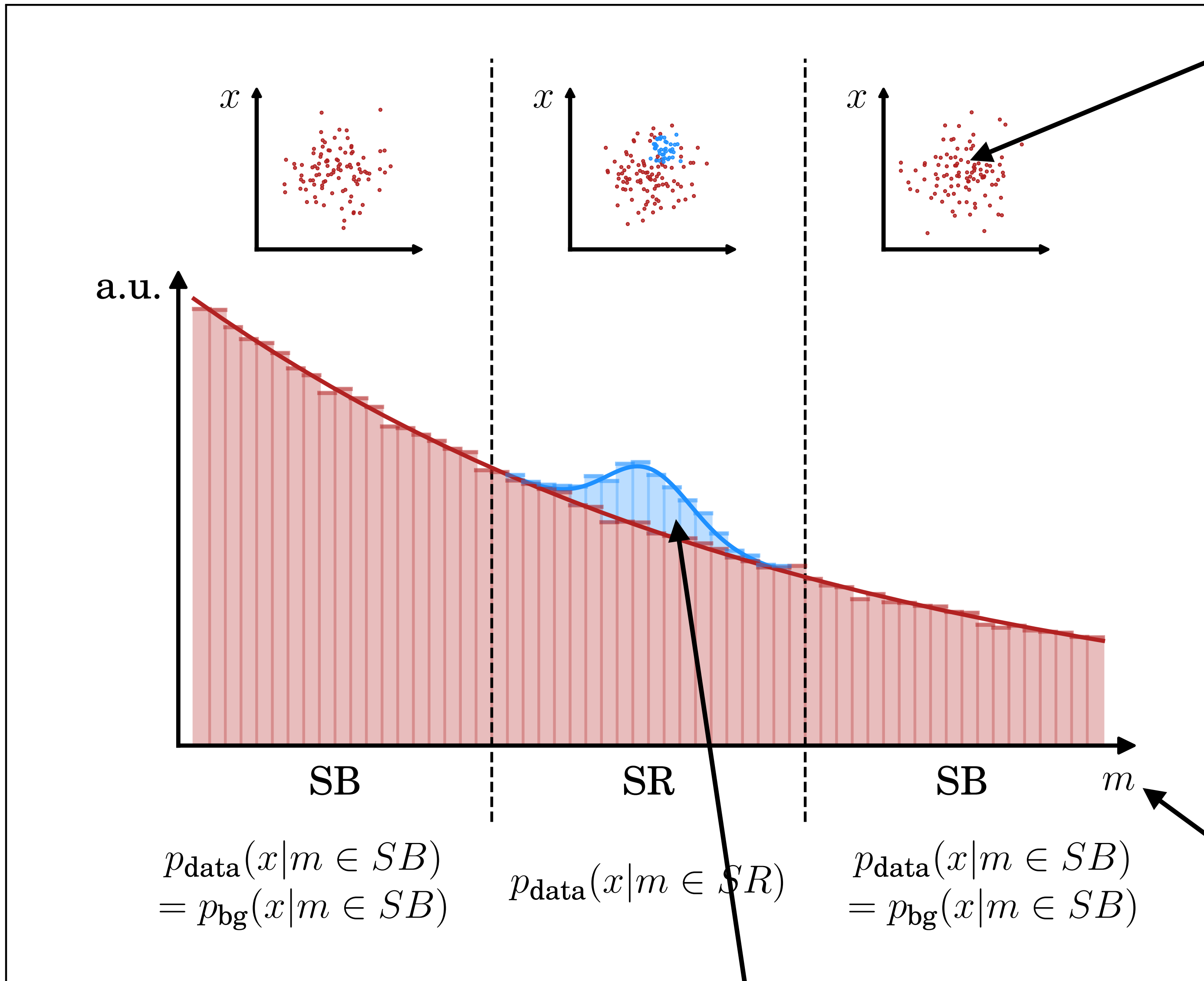
We don't assume the mass and type of the resonant particle



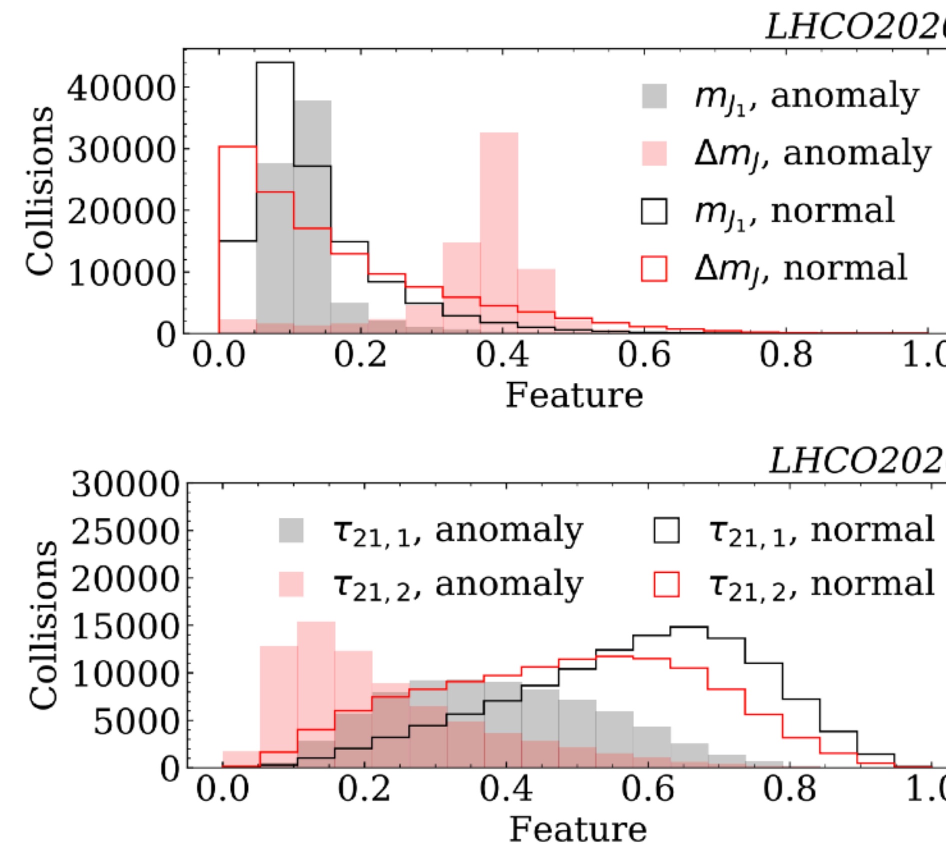
And we assume what particles it decays to

GK, Nachmann, Shih et al 2101.08320; Hallin, .., **GK** et al 2109.00546; Many similar approaches — see e.g. Golling, **GK** et al 2307.11157 for an overview

CATHODE



4 additional features

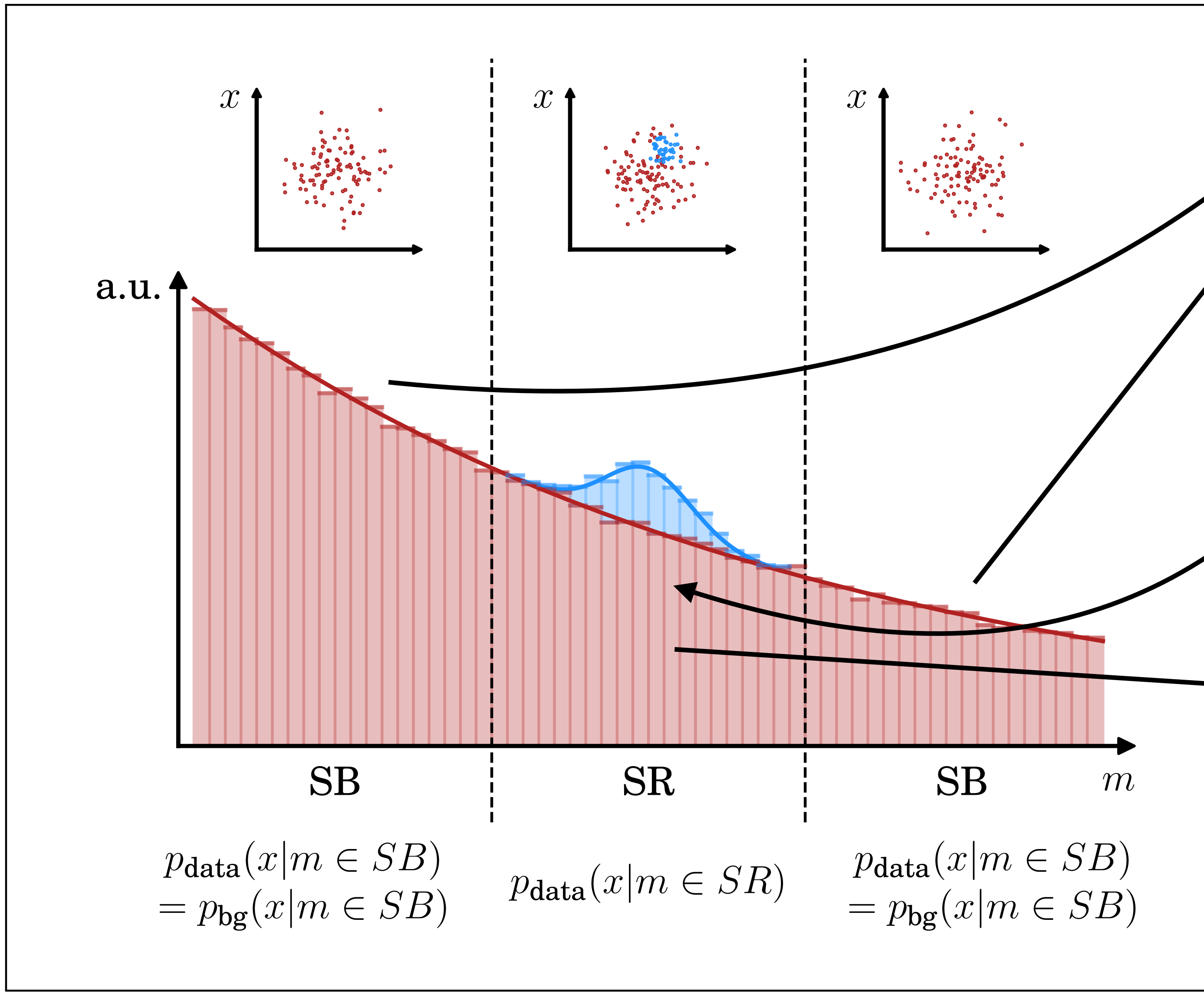


one resonant feature

Signal too small to be visible in inclusive distribution

GK, Nachmann, Shih et al 2101.08320; Hallin, .., **GK** et al 2109.00546;
LHC Olympics dataset: **GK**, Nachman, Shih, et al 2101.08320

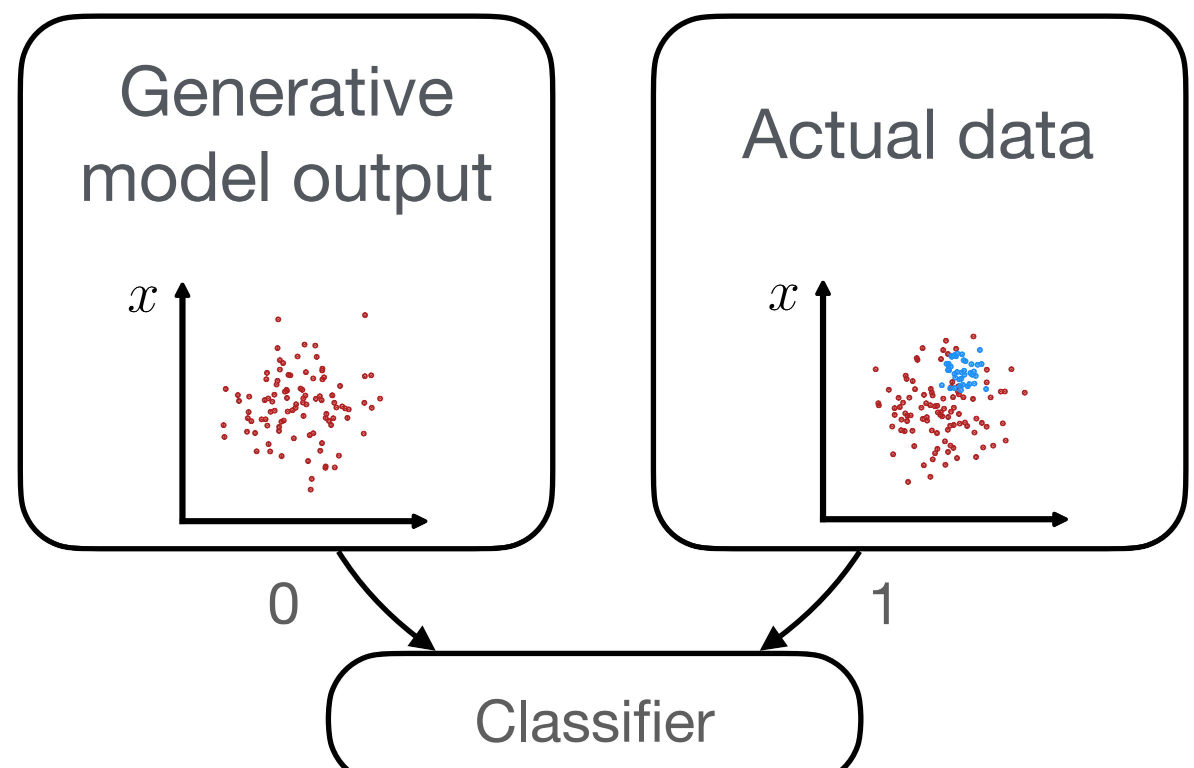
CATHODE



Train **generative** model conditional on m

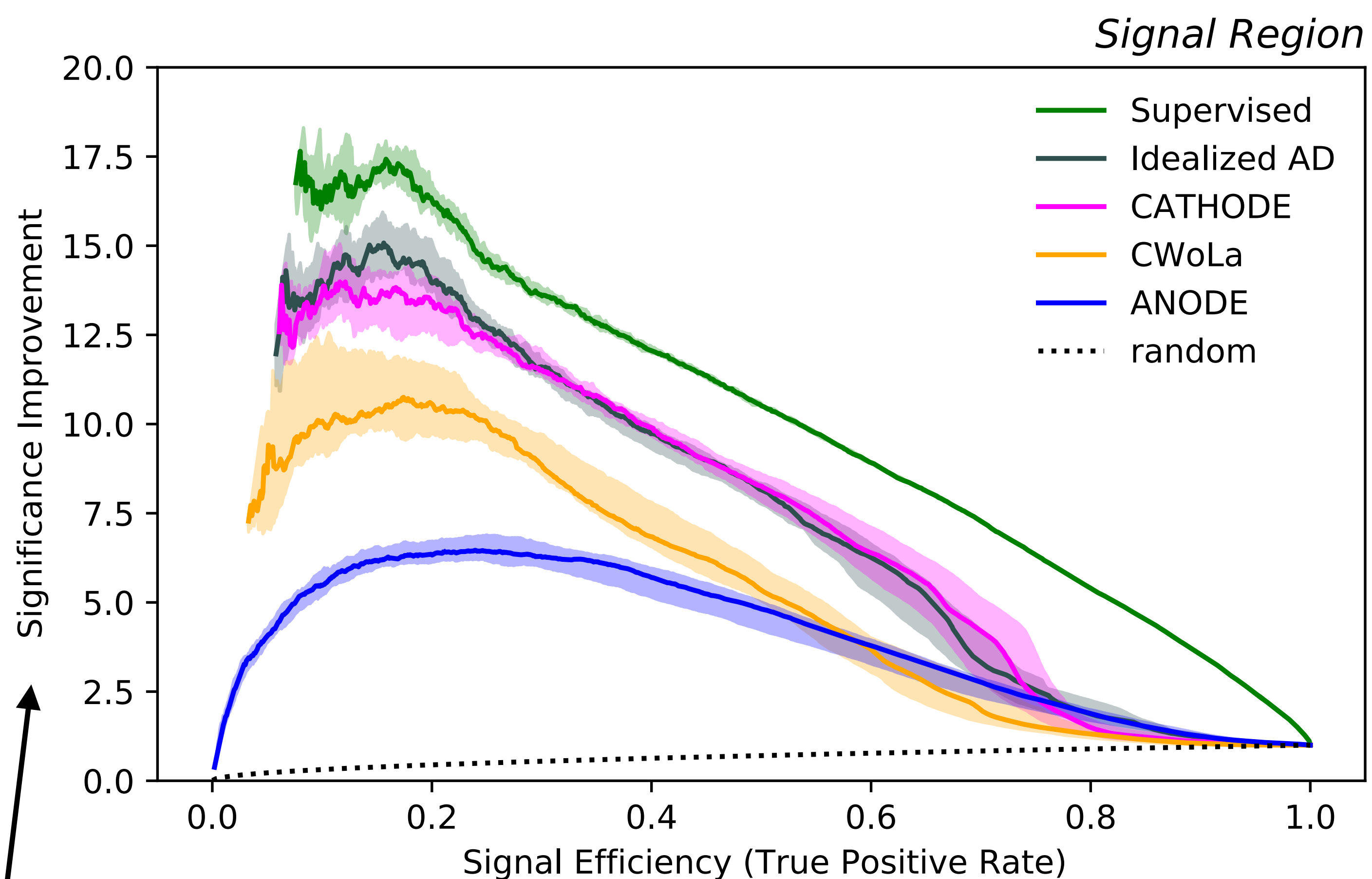
Interpolate & and **sample** here

Train a classifier between **prediction vs data**



GK, Nachmann, Shih et al 2101.08320; Hallin, ..,
GK et al 2109.00546;

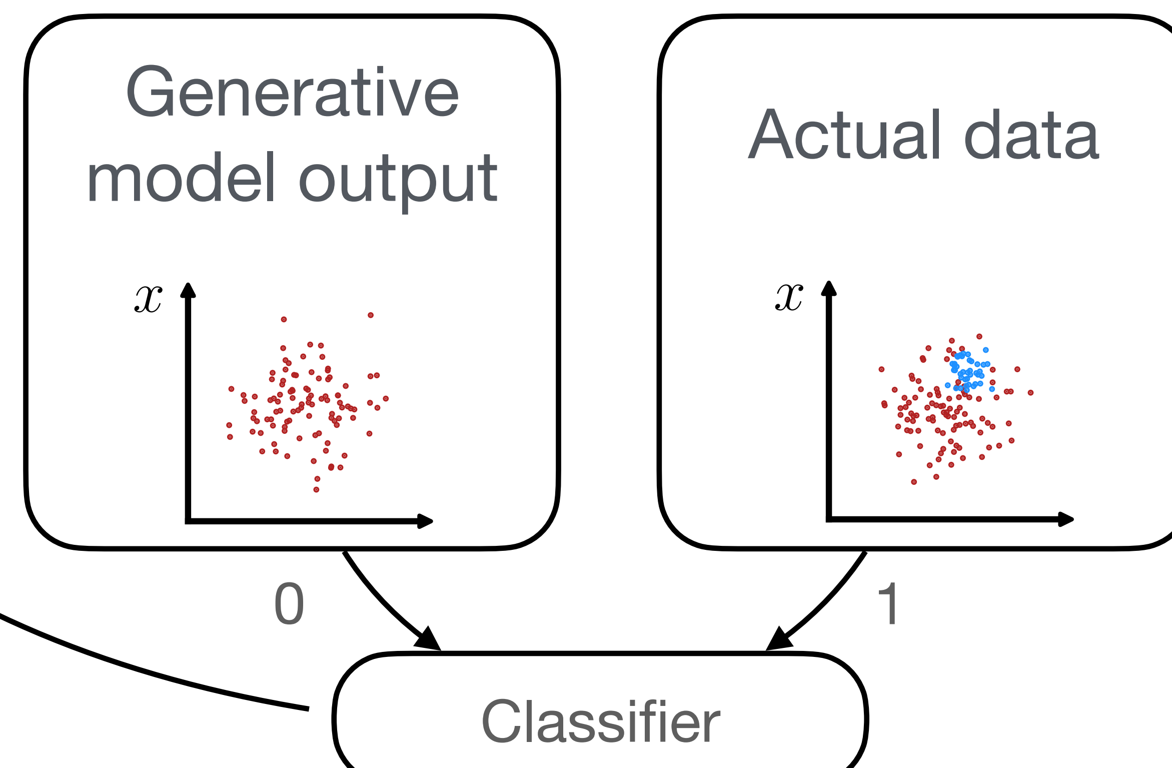
CATHODE



Use classifier to identify anomalies

Promising, but does this work on data?

$$\mathrm{SIC} = \frac{\epsilon_S}{\sqrt{\epsilon_B}}$$



GK, Nachmann, Shih et al 2101.08320; Hallin, ..,
GK et al 2109.00546;

CASE

Available on the CERN CDS information server **CMS PAS EXO-22-026**

CMS Physics Analysis Summary

Contact: cms-pag-conveners-exotica@cern.ch 2024/03/20

Model-agnostic search for dijet resonances with anomalous jet substructure in proton-proton collisions at $\sqrt{s} = 13$ TeV

The CMS Collaboration

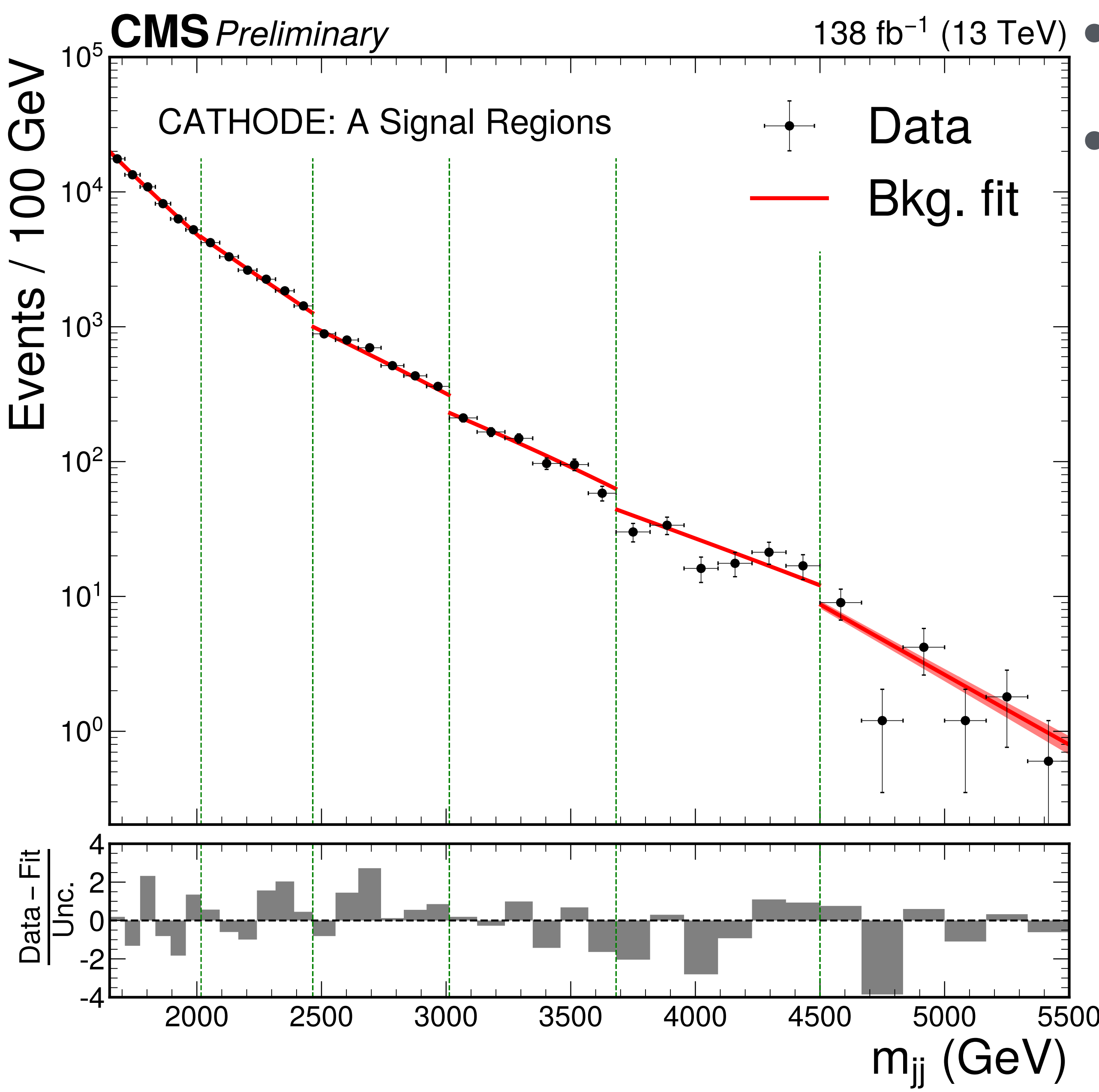
Abstract

This note introduces a model-agnostic search for new physics in the dijet final state. Other than the requirement of a narrow dijet resonance with a mass in the range of 1800-6000 GeV, minimal additional assumptions are placed on the signal hypothesis. Search regions are obtained by utilizing multivariate machine learning methods to select jets with anomalous substructure. A collection of complementary anomaly detection methods – based on unsupervised, weakly-supervised and semi-supervised algorithms – are used in order to maximize the sensitivity to unknown new physics signatures. These algorithms are applied to data corresponding to an integrated luminosity of 138 fb^{-1}, recorded in the years 2016 to 2018 by the CMS experiment at the LHC, at a centre-of-mass energy of 13 TeV. No significant excesses above background expectation are seen, and exclusion limits are derived on the production cross section of benchmark signal models varying in resonance mass, jet mass and jet substructure. Many of these signatures have not previously been searched for at the LHC, making the limits reported on the corresponding benchmark models the first ever and the most stringent to date.

- New result by the CMS collaboration: CMS Anomaly Search Effort (CASE)
- Full Run 2 dataset
- 6 anomaly detectors in parallel

CMS PAS-EXO-23-006

CASE



- New result by the CMS collaboration: CMS Anomaly Search Effort (CASE)
- Full Run 2 dataset
- 6 anomaly detectors in parallel

Fully train CATHODE on data

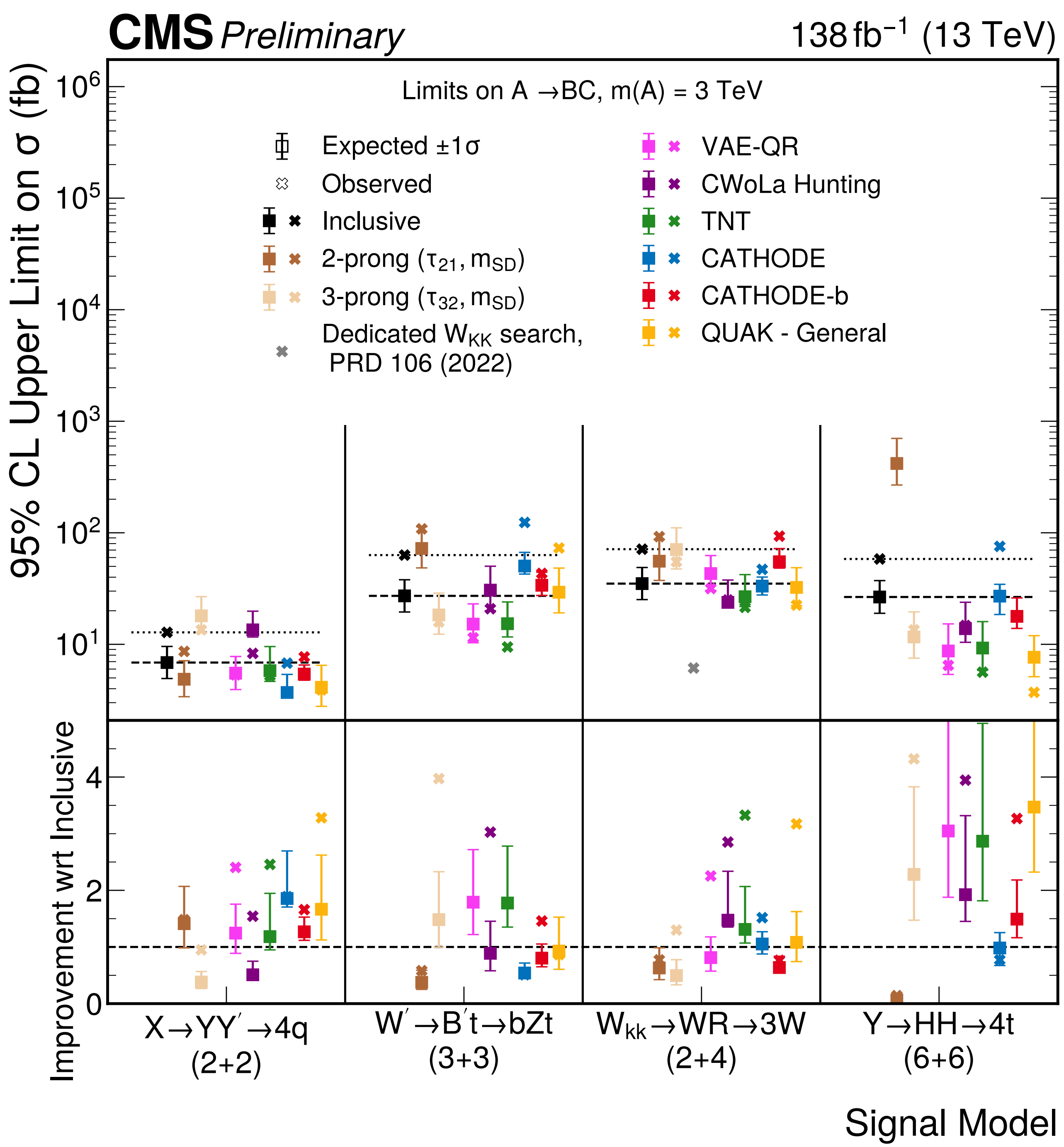
Select top 1% most anomalous events, perform bump-hunt

No signal-like outlier: set limits

CMS PAS-EXO-23-006

CASE

- New result by the CMS collaboration: CMS Anomaly Search Effort (CASE)
- Full Run 2 dataset
- 6 anomaly detectors in parallel



For limits: inject potential signals

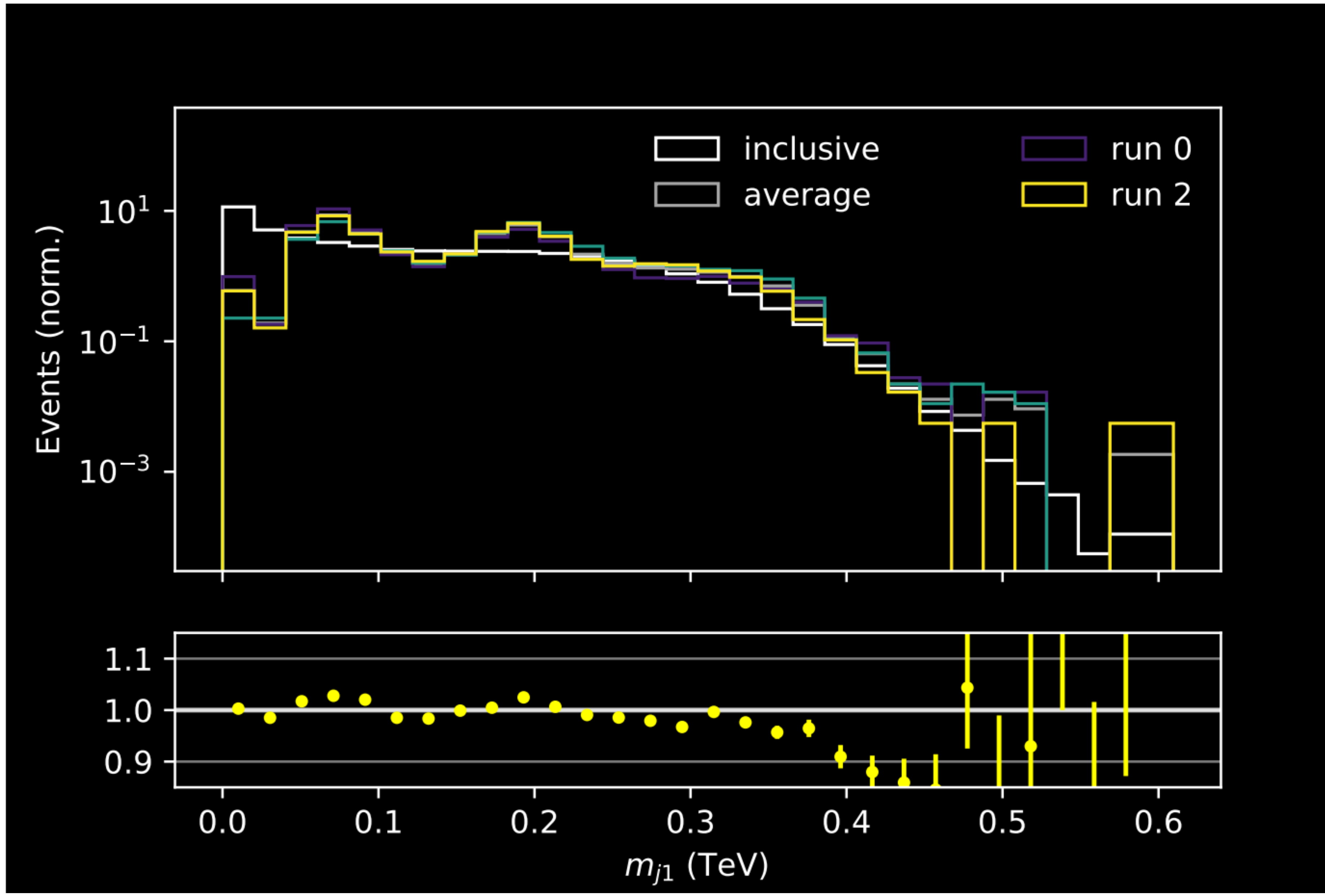
Includes uncertainties, e.g. multi-prong jets modelling

Altogether set limits for 43 different signal scenarios

CMS PAS-EXO-23-006

Generative non-closure

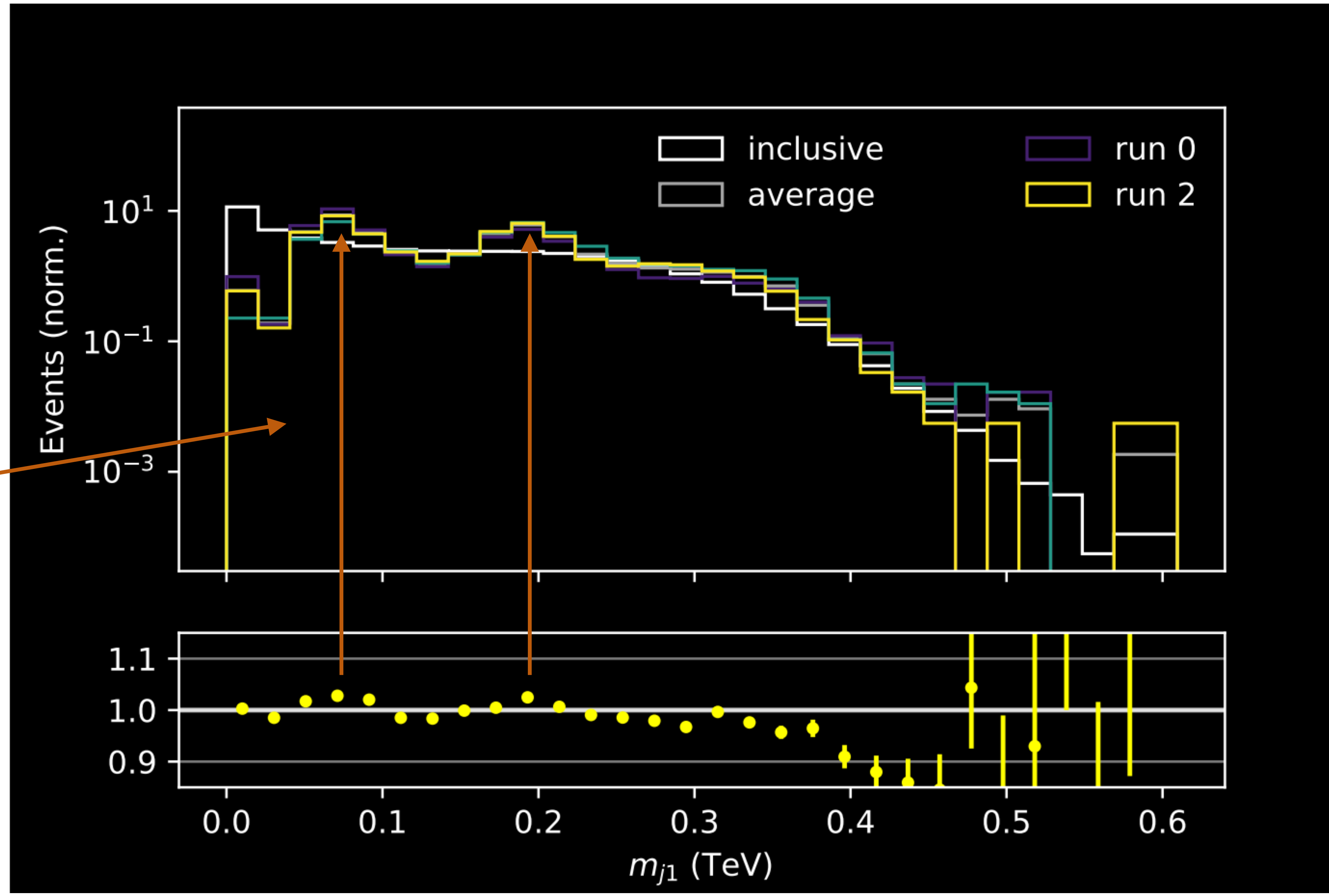
Inclusive is sideband; other distributions after classifier cut



True sideband/ Generated sideband

Figure by Sommerhalder

Generative non-closure

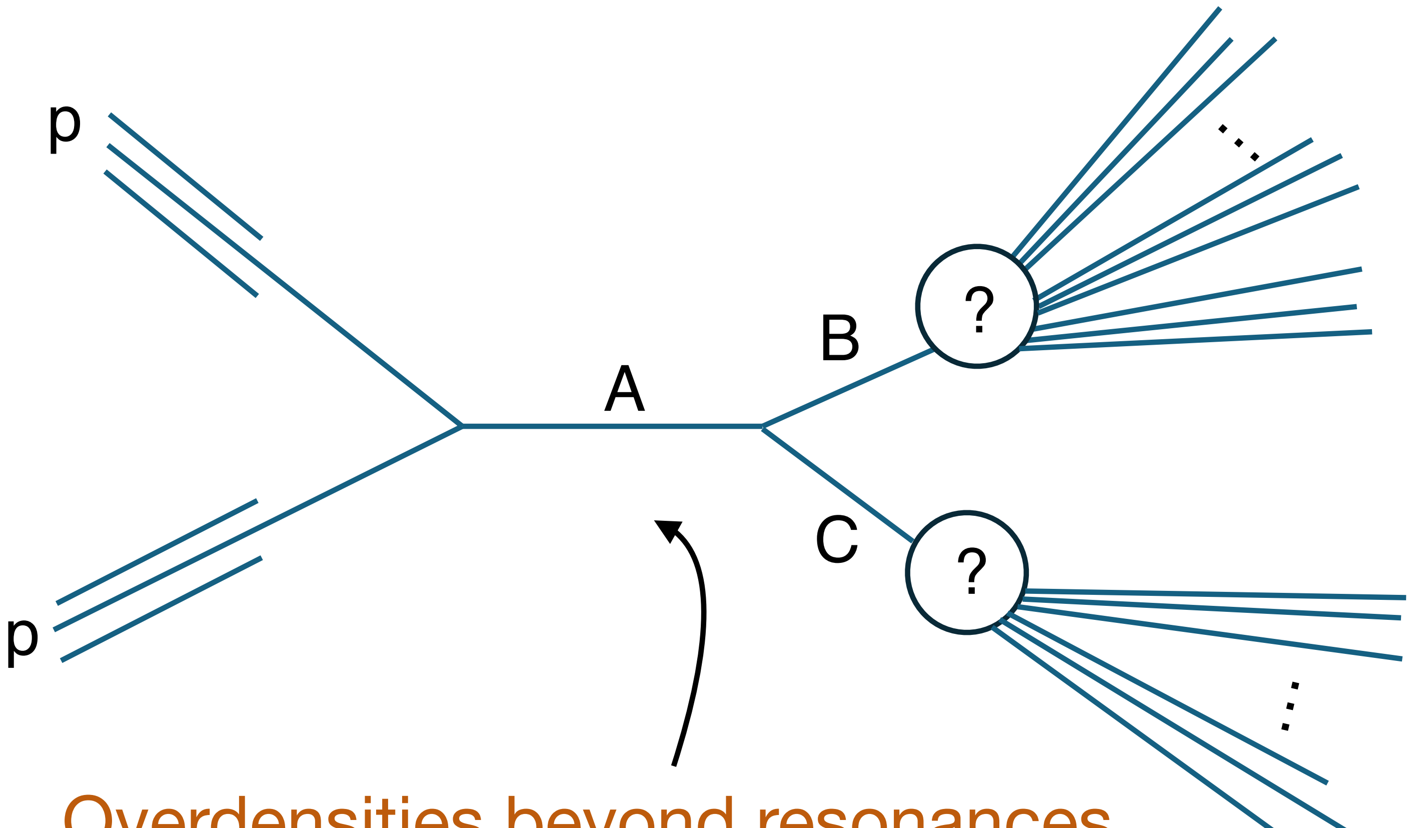


Sensitive to percent-level differences

Might benefit (highly) from clever uncertainty ideas

Figure by Sommerhalder

Other developments & issues



More features per jet
(e.g. 2309.13111)

Low-level input data
(e.g. 2310.06897)

Overdensities beyond resonances
(e.g. 2404.07258, 2311.12924)

Better sensitivity for weak signals (e.g. 2312.11629)

Reduce shaping of distributions

More topologies

Anomalies as outliers (e.g. substantial literature on auto encoder based methods)

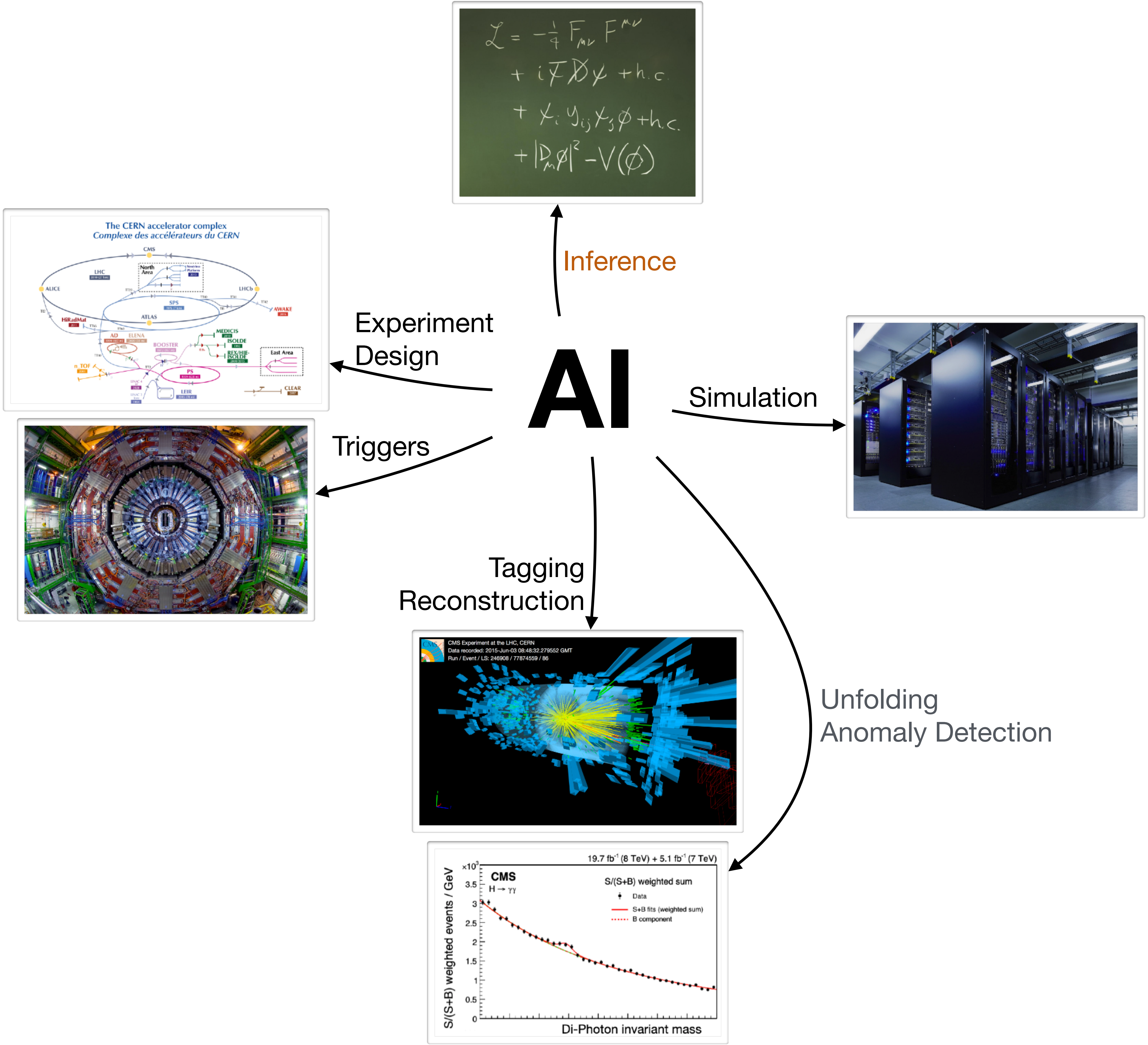
Robust statistical treatment beyond bump-hunts (e.g. 2111.13633)

Treatment of uncertainties

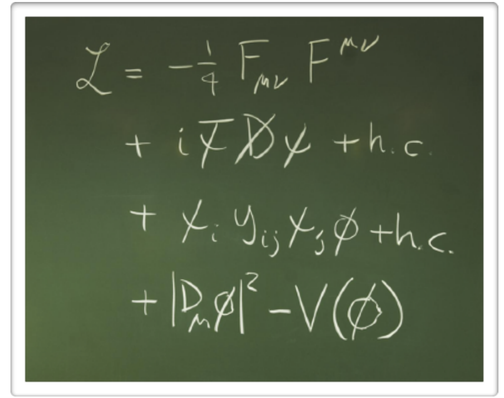
Applications to data monitoring

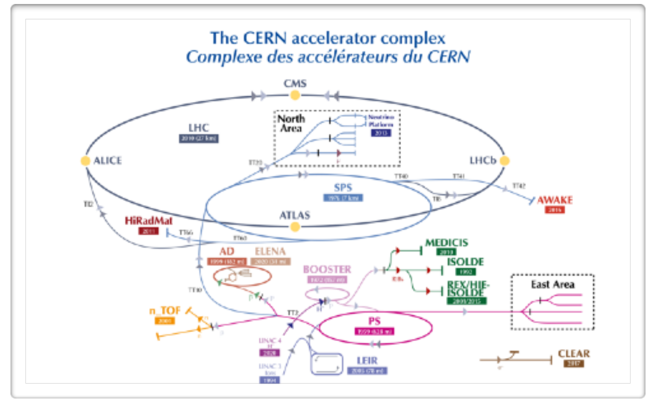
Sharing of results

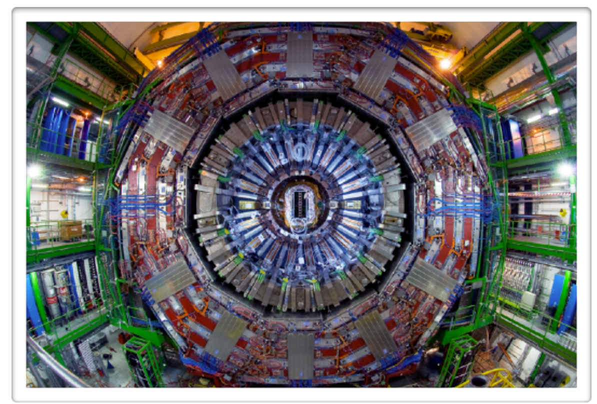
See talks by **Thea, Lily, Gaia, and Mikael**

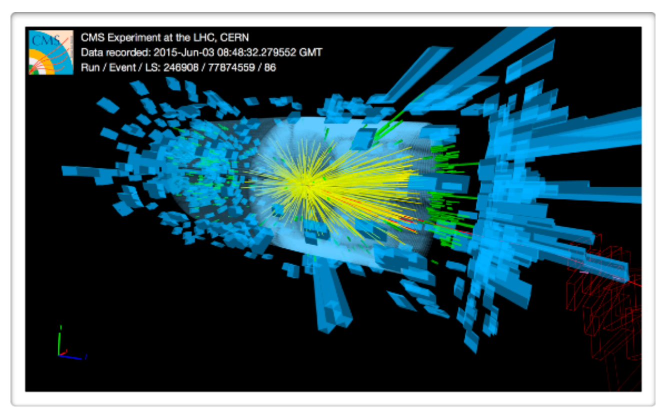

AI
Inference
Experiment
Design
Triggers
Simulation
Tagging
Reconstruction
Unfolding
Anomaly Detection
The CERN accelerator complex
Complexe des accélérateurs du CERN
CMS Experiment at the LHC, CERN
Data recorded: 2015-Jun-03 08:48:32.279552 GMT
Run / Event / LS: 246908 / 77874559 / 86
19.7 fb⁻¹ (8 TeV) + 5.1 fb⁻¹ (7 TeV)
CMS
H → γγ
S/(S+B) weighted sum
Data
S+B fits (weighted sum)
B component
S/(S+B) weighted events / GeV
Di-Photon invariant mass

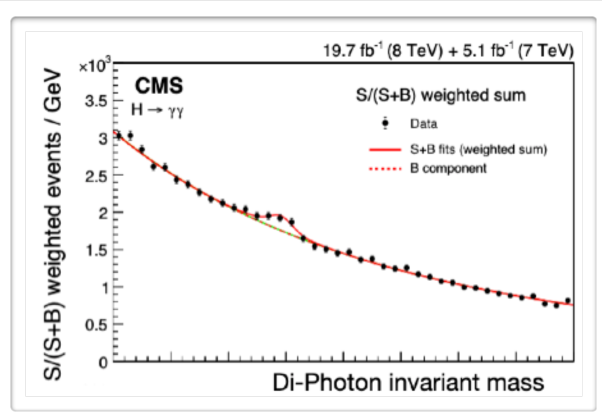
Differentiable versions of all steps in the particle physics processing chain







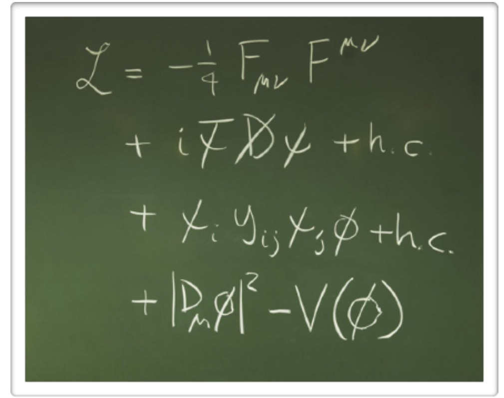


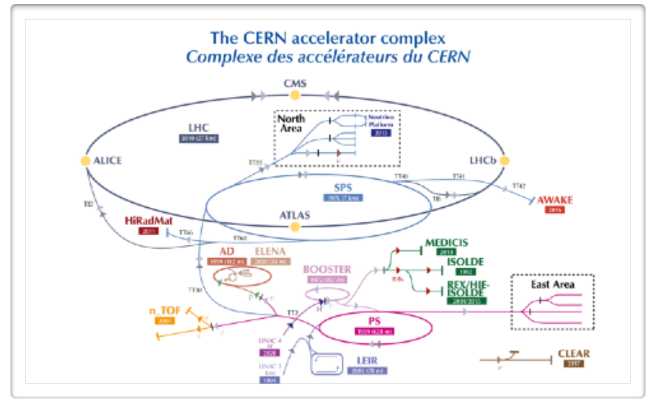


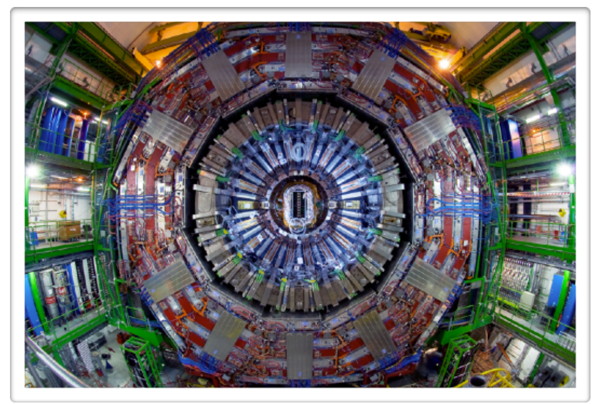
Differentiable versions of all steps in the particle physics processing chain

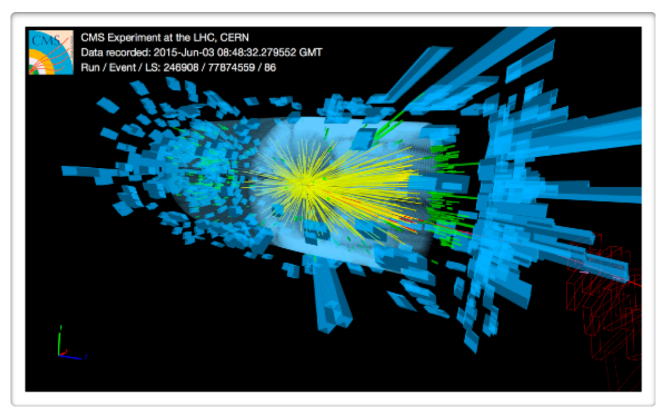
Either as ML-based surrogate models

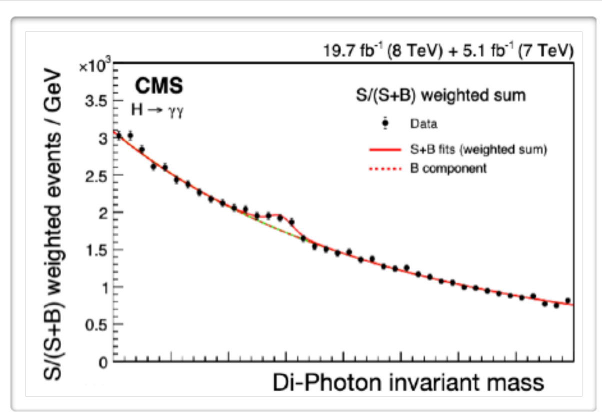
Or via e.g. differentiable programming











Heinrich, Kagan 2308.16680

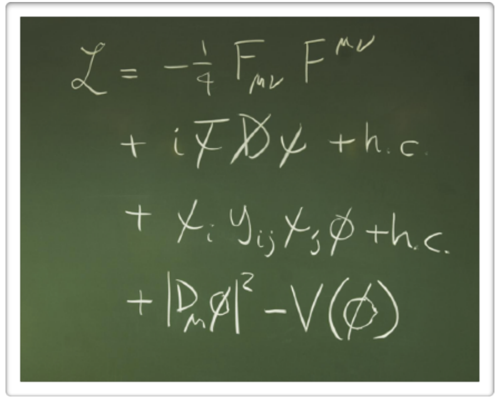
Differentiable versions of all steps in the particle physics processing chain

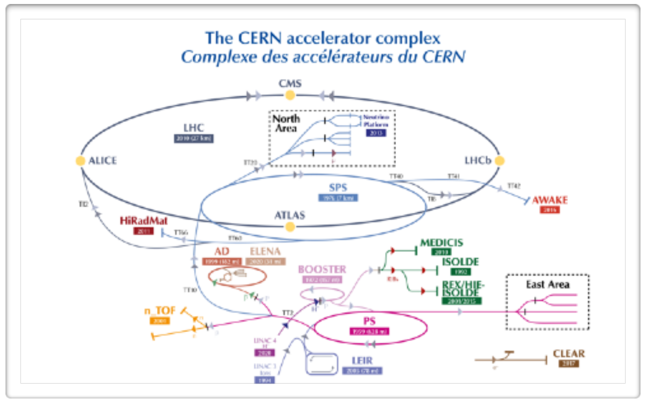
Either as ML-based surrogate models

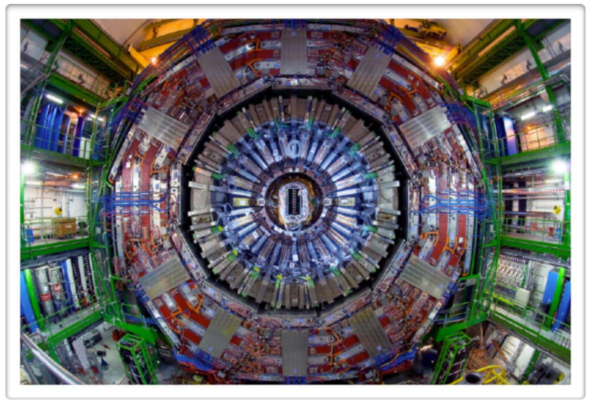
Or via e.g. differentiable programming

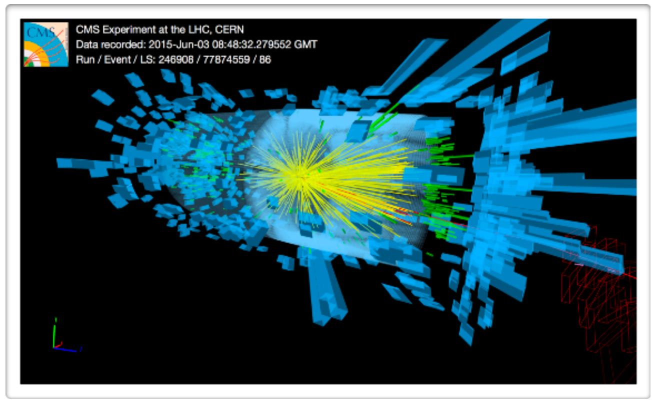
What can we do with this?

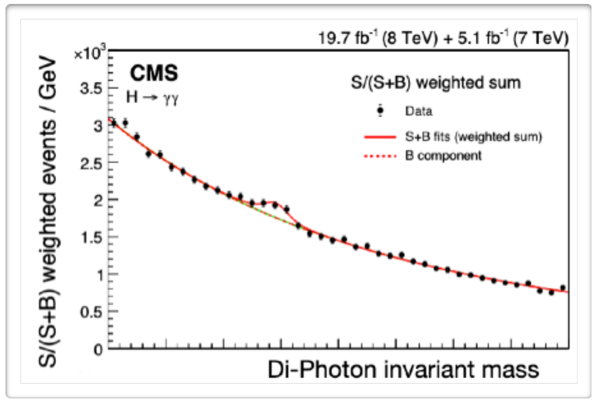
Heinrich, Kagan 2308.16680



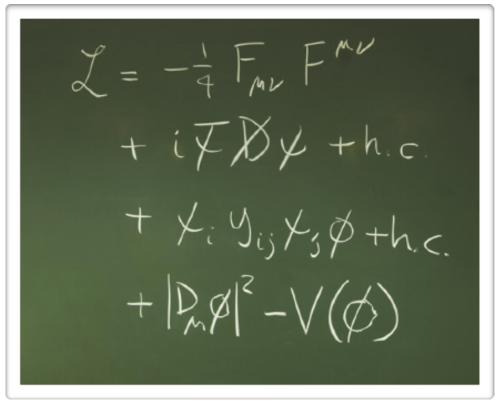


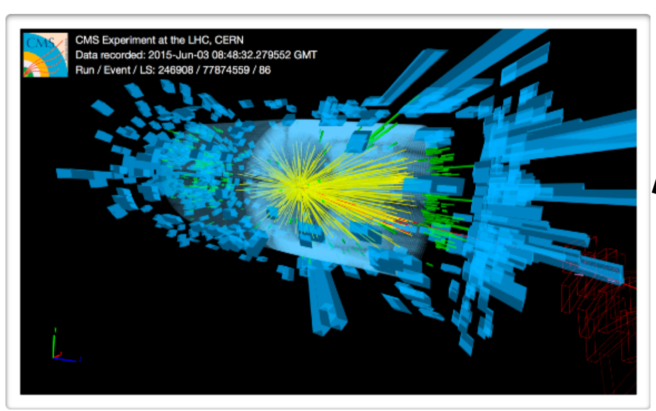






Inference





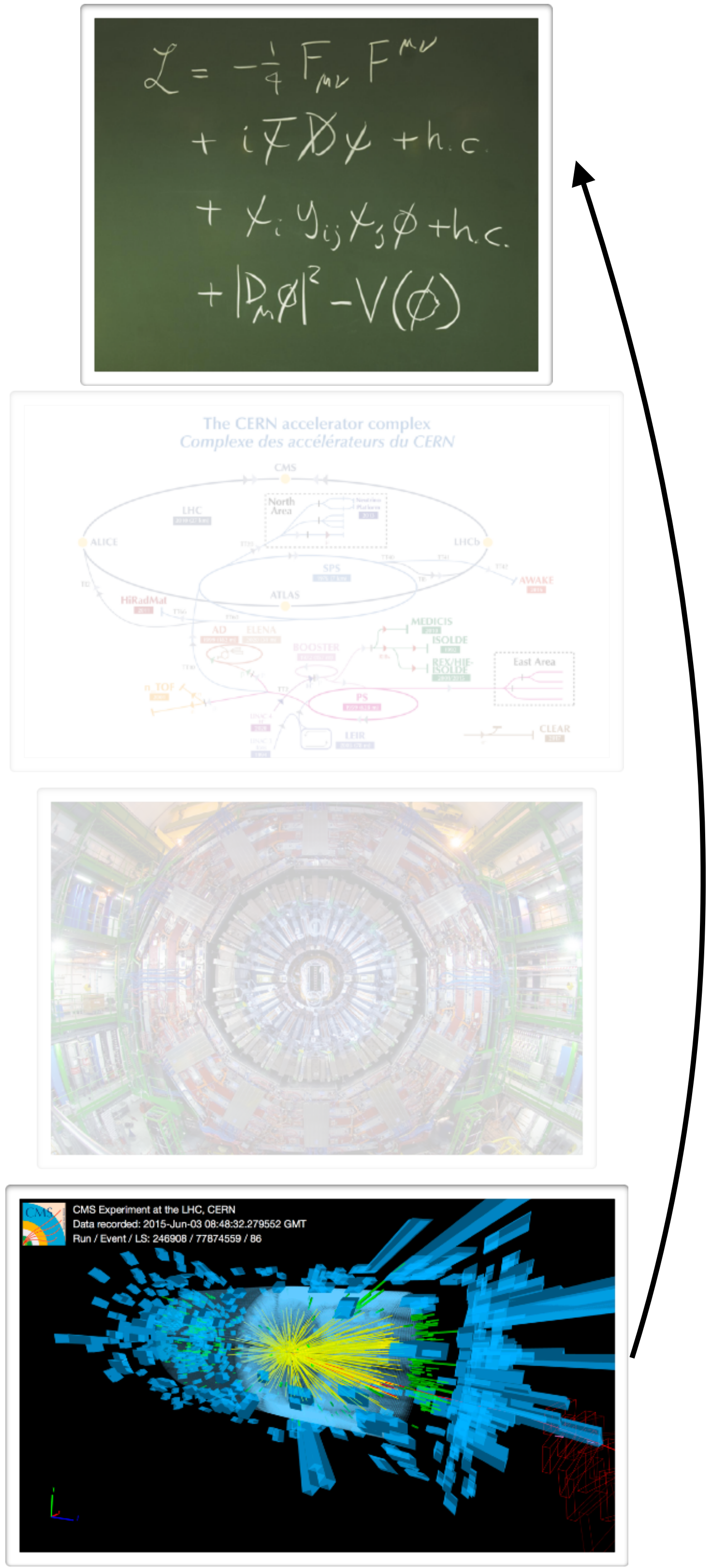
Inference

Goal: Learn parameters of theory (e.g. couplings) directly from high-dimensional data

No exact likelihood, but forward simulations available: likelihood-free / simulation based inference

Cranmer, Brehmer, Louppe 1911.01429

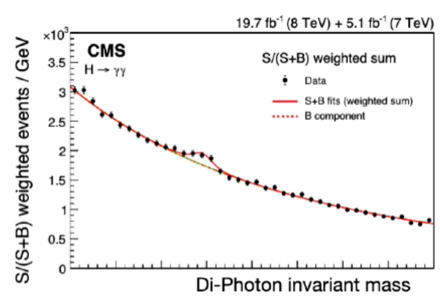
Inference



Goal: Learn parameters of theory (e.g. couplings) directly from high-dimensional data

No exact likelihood, but forward simulations available: likelihood-free / simulation based inference

Inference



Summary Statistics

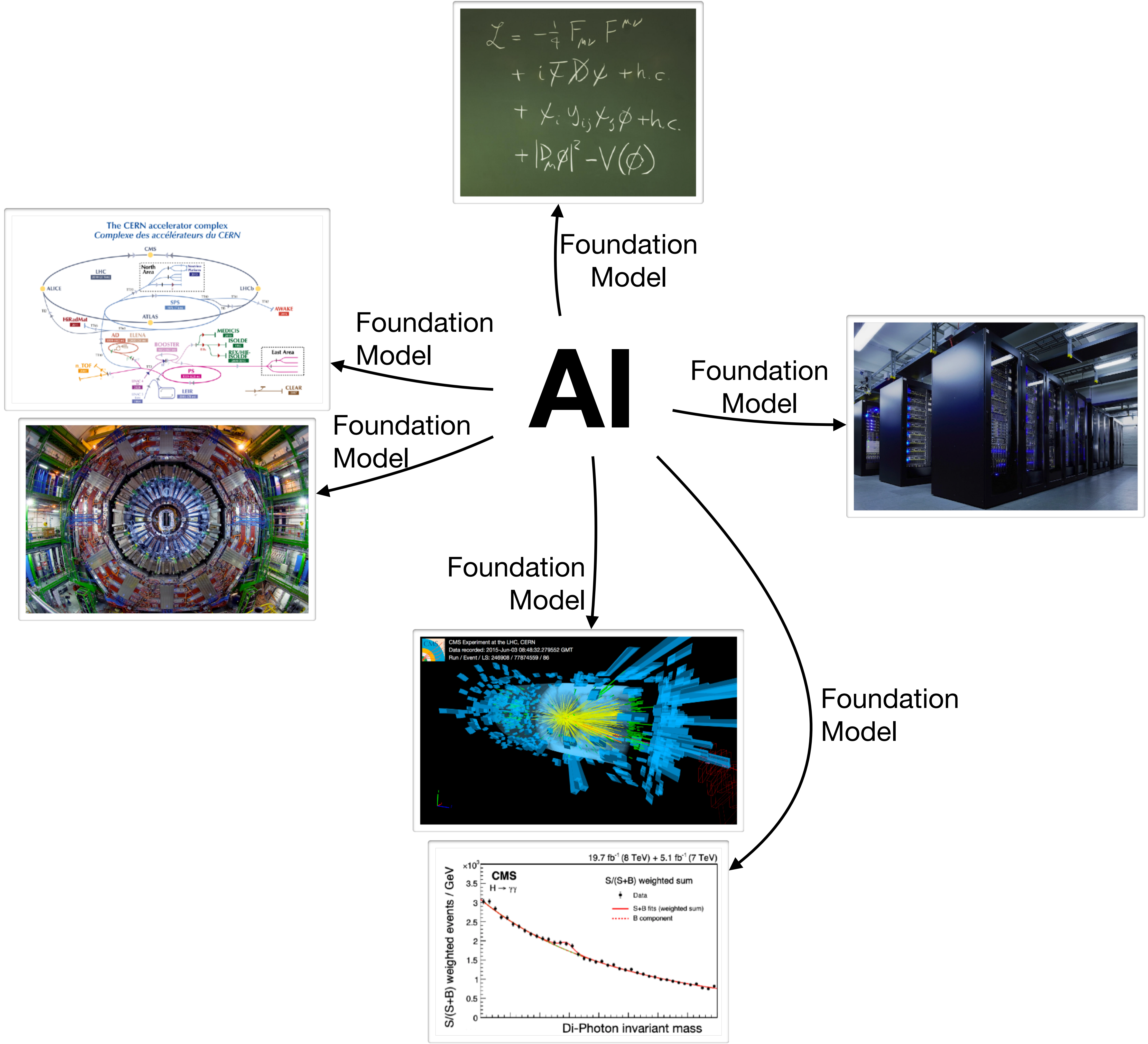
Likelihood Learning (e.g. flows or cINNs)

Likelihood ratio trick (e.g. CARL, swyft)

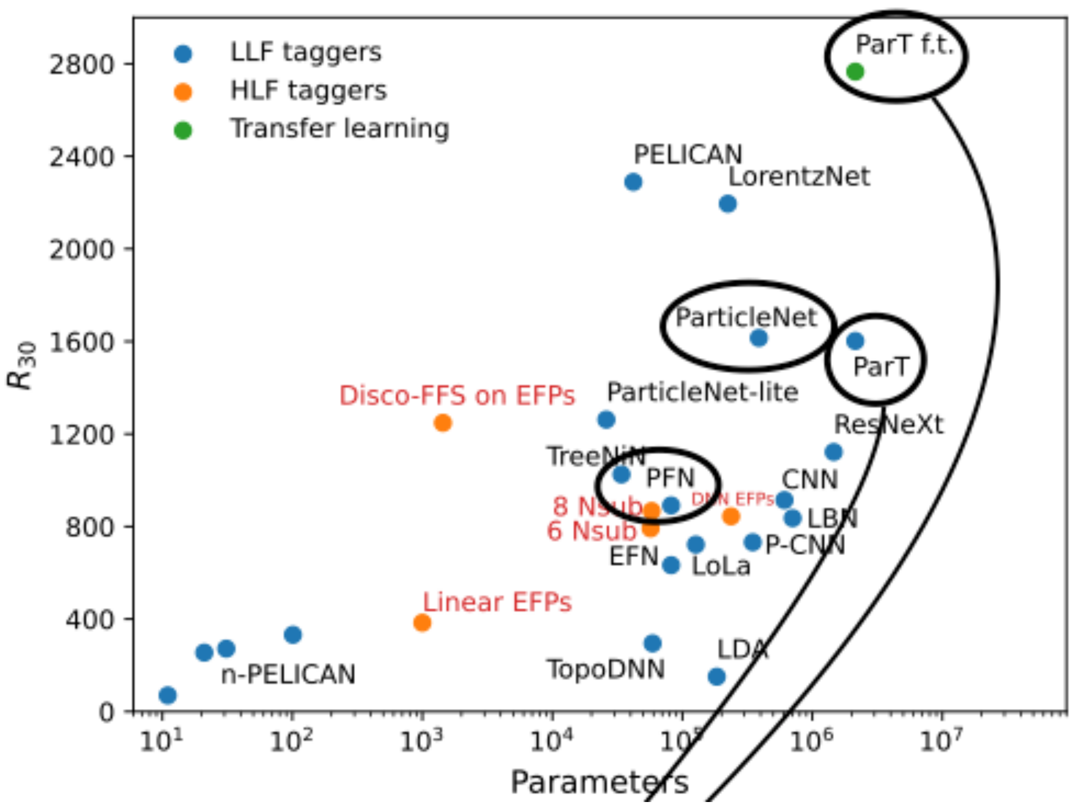
Integration (e.g. MadMiner)

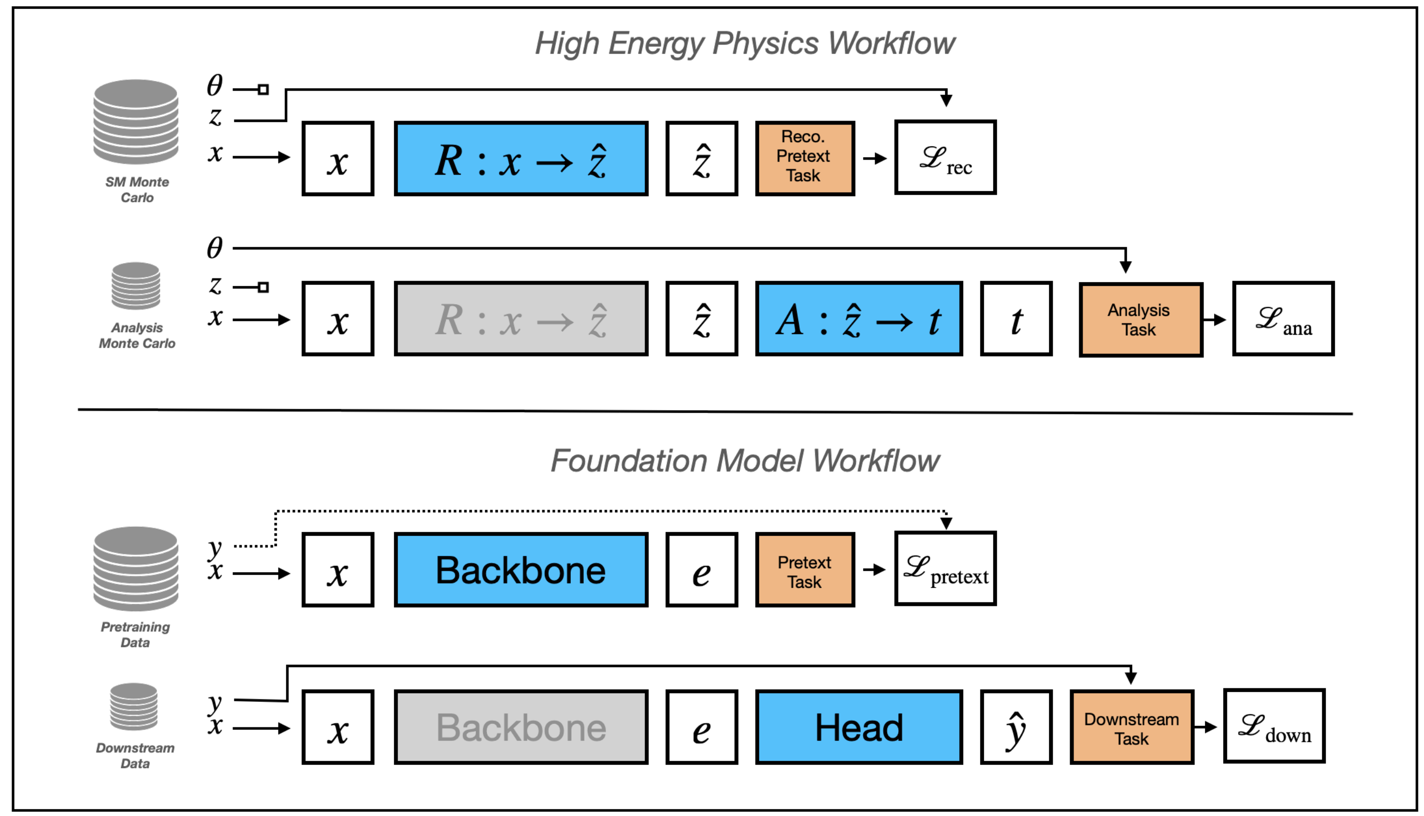
See talk by **Aishik, Kyle, Artur**

Cranmer, Brehmer, Louppe 1911.01429; Cranmer, Brehmer 2010.06439


AI
Foundation Model
Foundation Model
Foundation Model
Foundation Model
Foundation Model
Foundation Model
The CERN accelerator complex
Complexe des accélérateurs du CERN
CMS Experiment at the LHC, CERN
Data recorded: 2015-Jun-03 08:48:32.279552 GMT
Run / Event / LS: 246908 / 77874559 / 86
19.7 fb⁻¹ (8 TeV) + 5.1 fb⁻¹ (7 TeV)
CMS
H → γγ
S/(S+B) weighted sum
Data
S+B fits (weighted sum)
B component
S/(S+B) weighted events / GeV
Di-Photon invariant mass

Foundation models for physics data



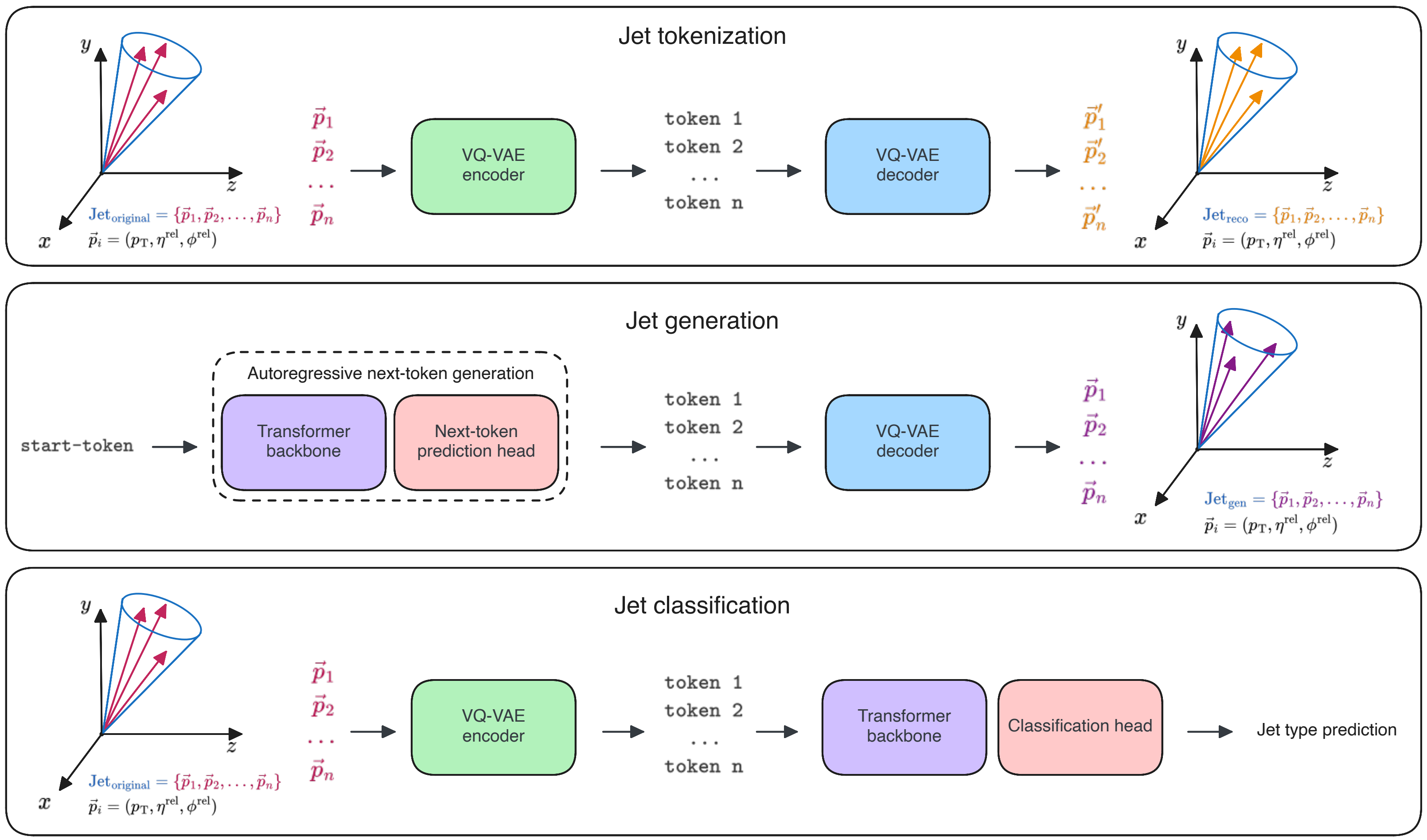


Already observed best performance in supervised classification by transfer learning

Foundation models extend transfer more broadly and centralise and re-use training

Qu et al 2202.03772; Vigl et al 2401.13536

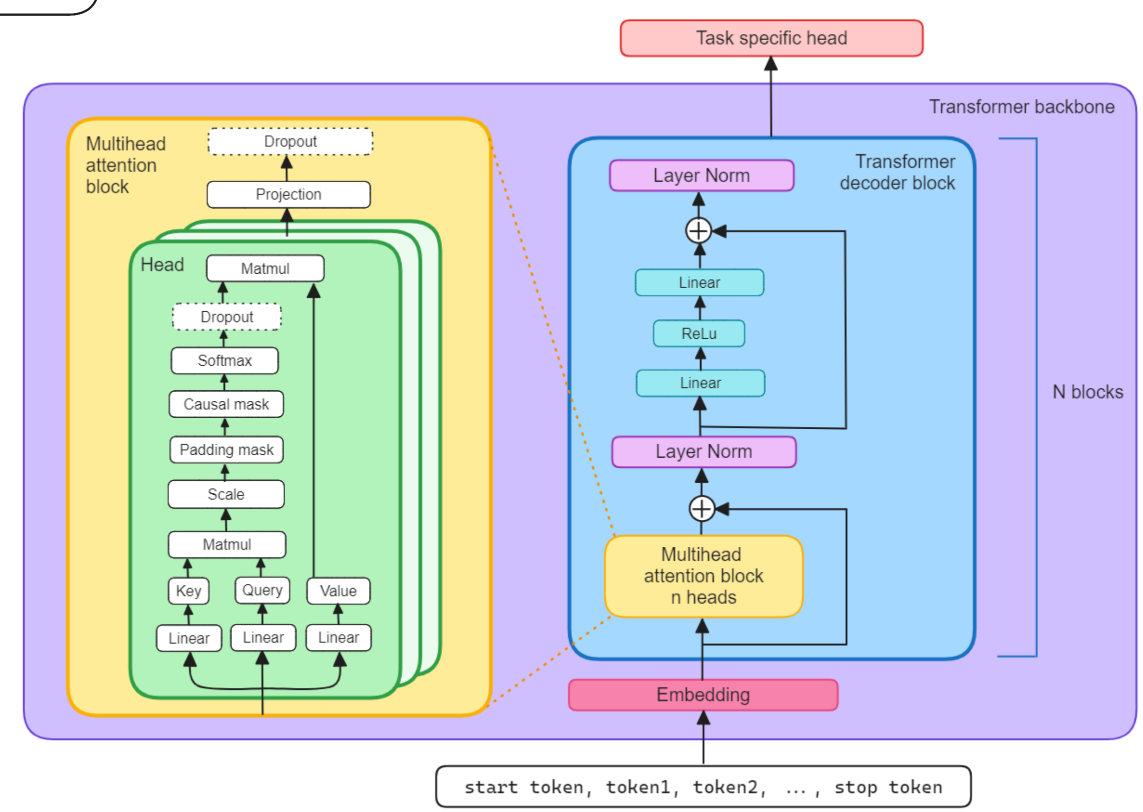
OmniJet-α



GPT for jets
using tokenised inputs

Train on
unsupervised generation

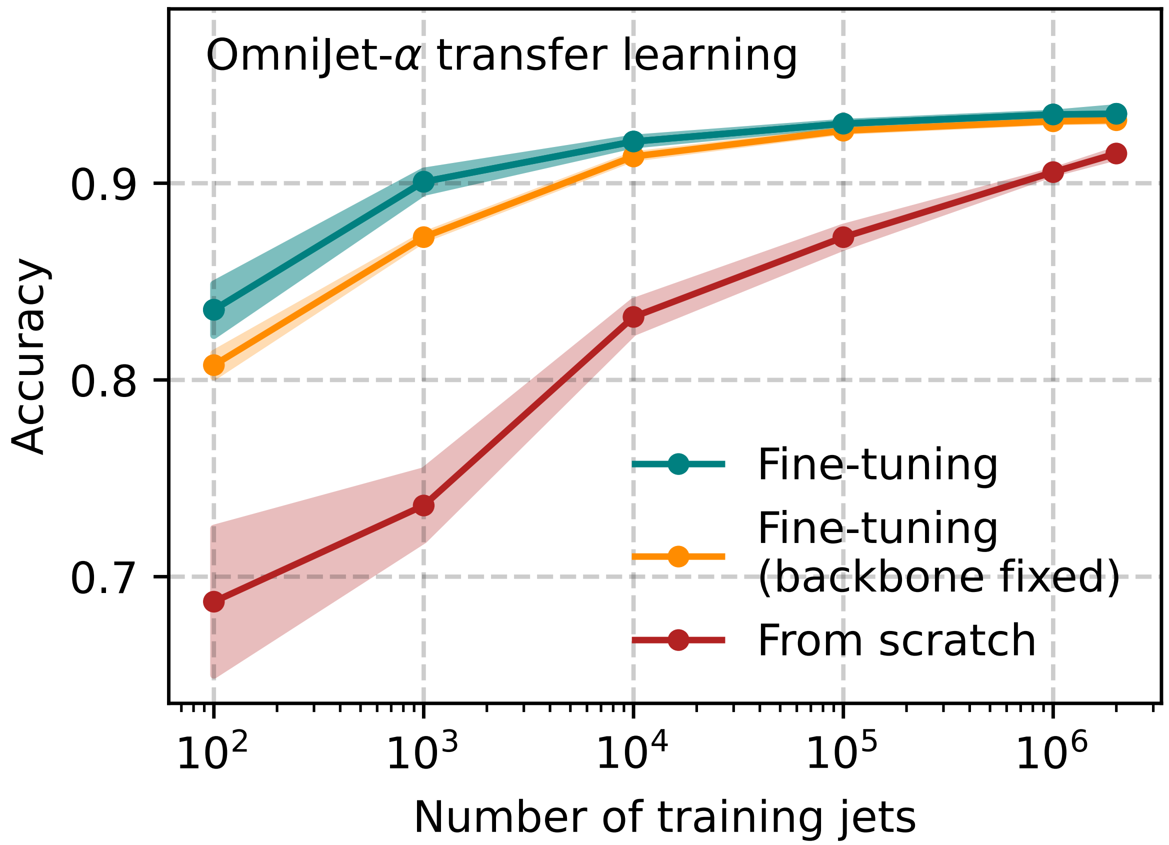
Evaluate for
supervised classification



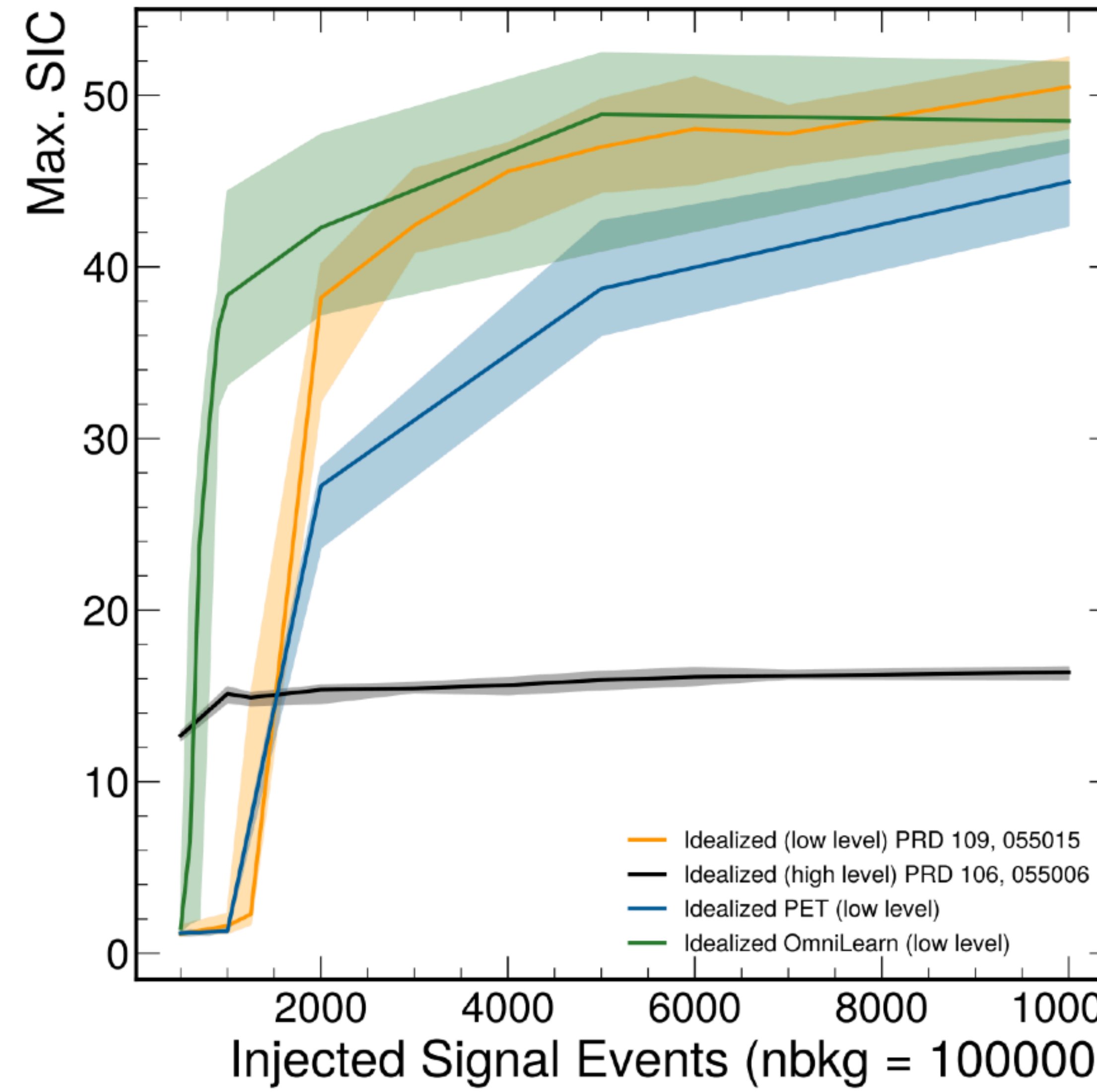
Transformer
backbone

Birk, Hallin, **GK** 2403.05618; Butter et al 2305.10475; Finke et al 2303.07364; Vigl et al 2401.13536; Heinrich et al 2401.13537; Mikuni, Nachman 2404.16091

Generalisation



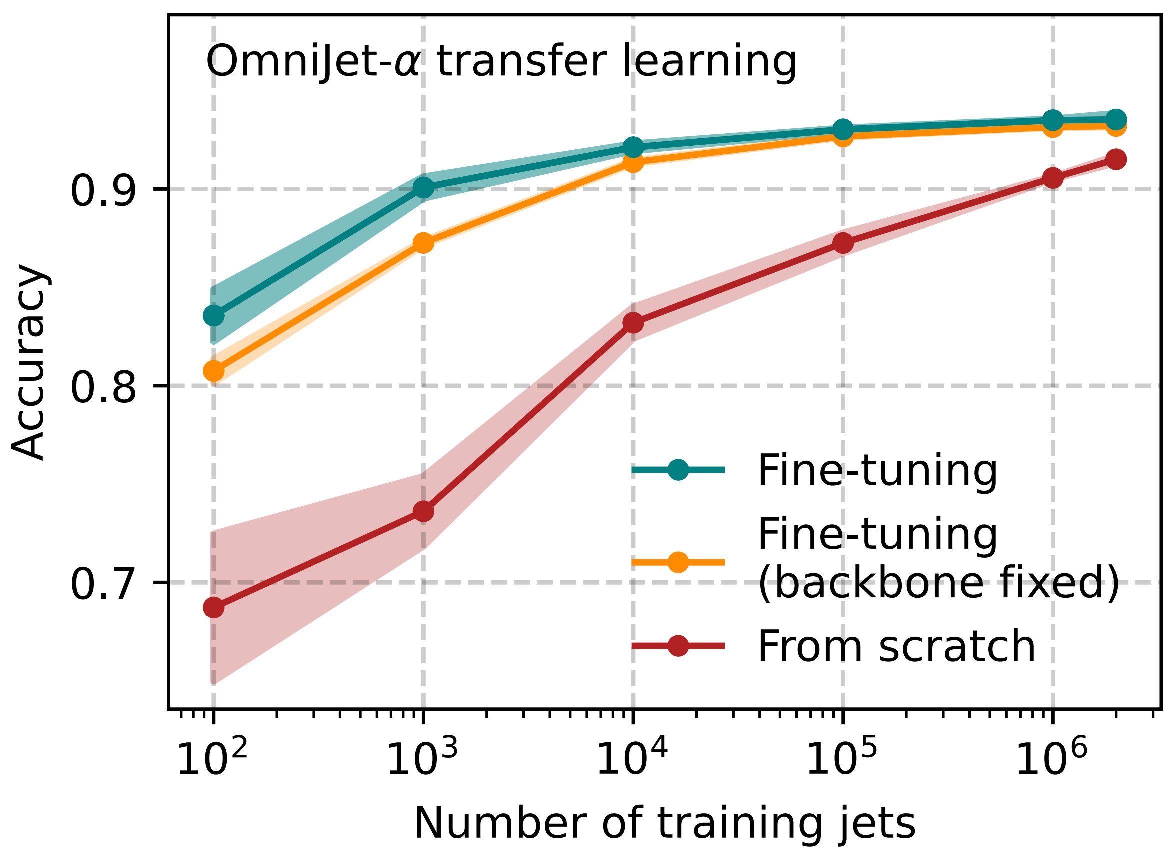
Pre-training on "cheap" unlabelled examples improves supervised classification data efficiency up to 100-1000x

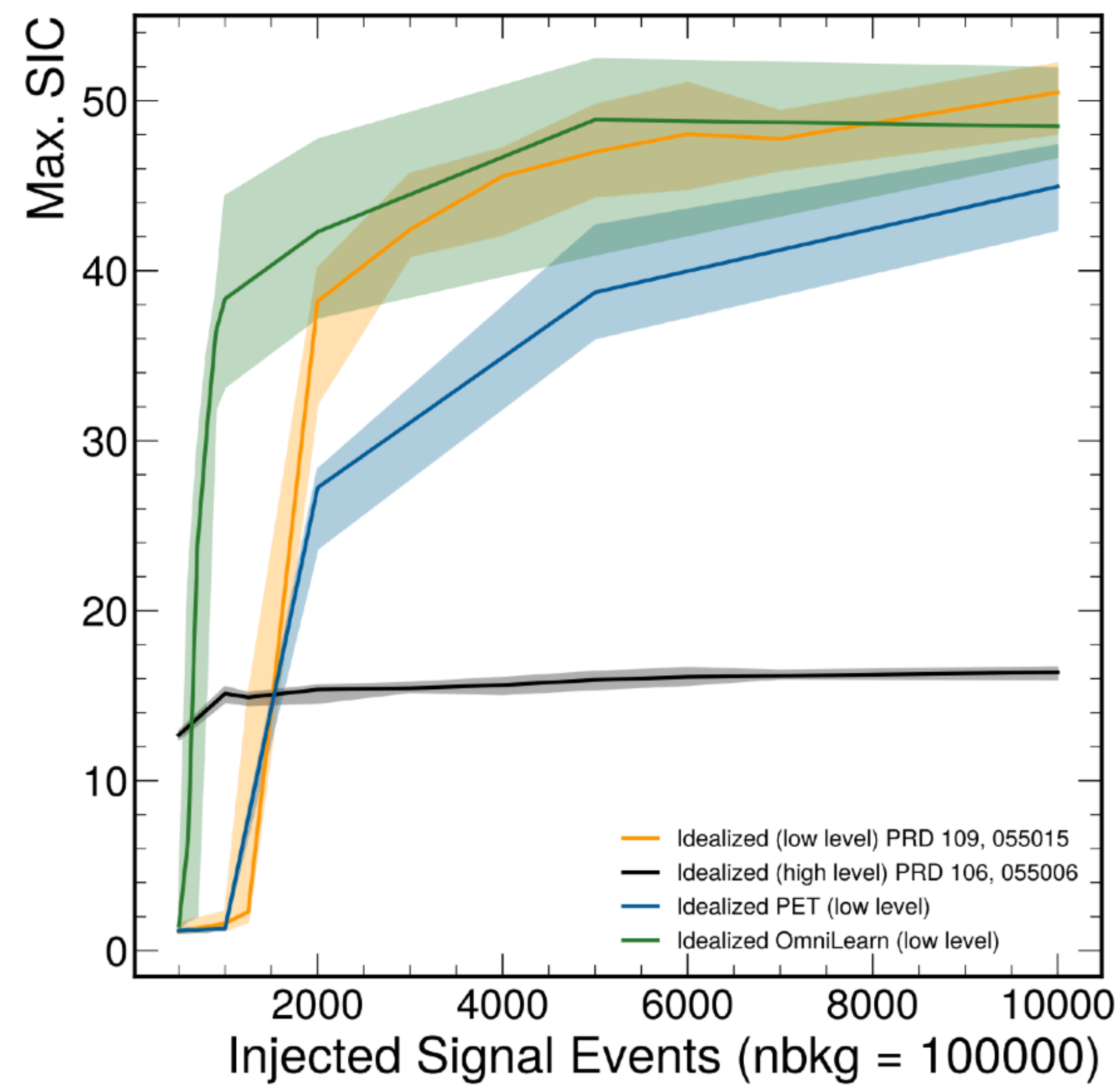


Recent OmniLearn generalises across broad range of tasks including anomaly detection

Birk, Hallin, **GK** 2403.05618; Mikuni, Nachman 2404.16091

Open questions





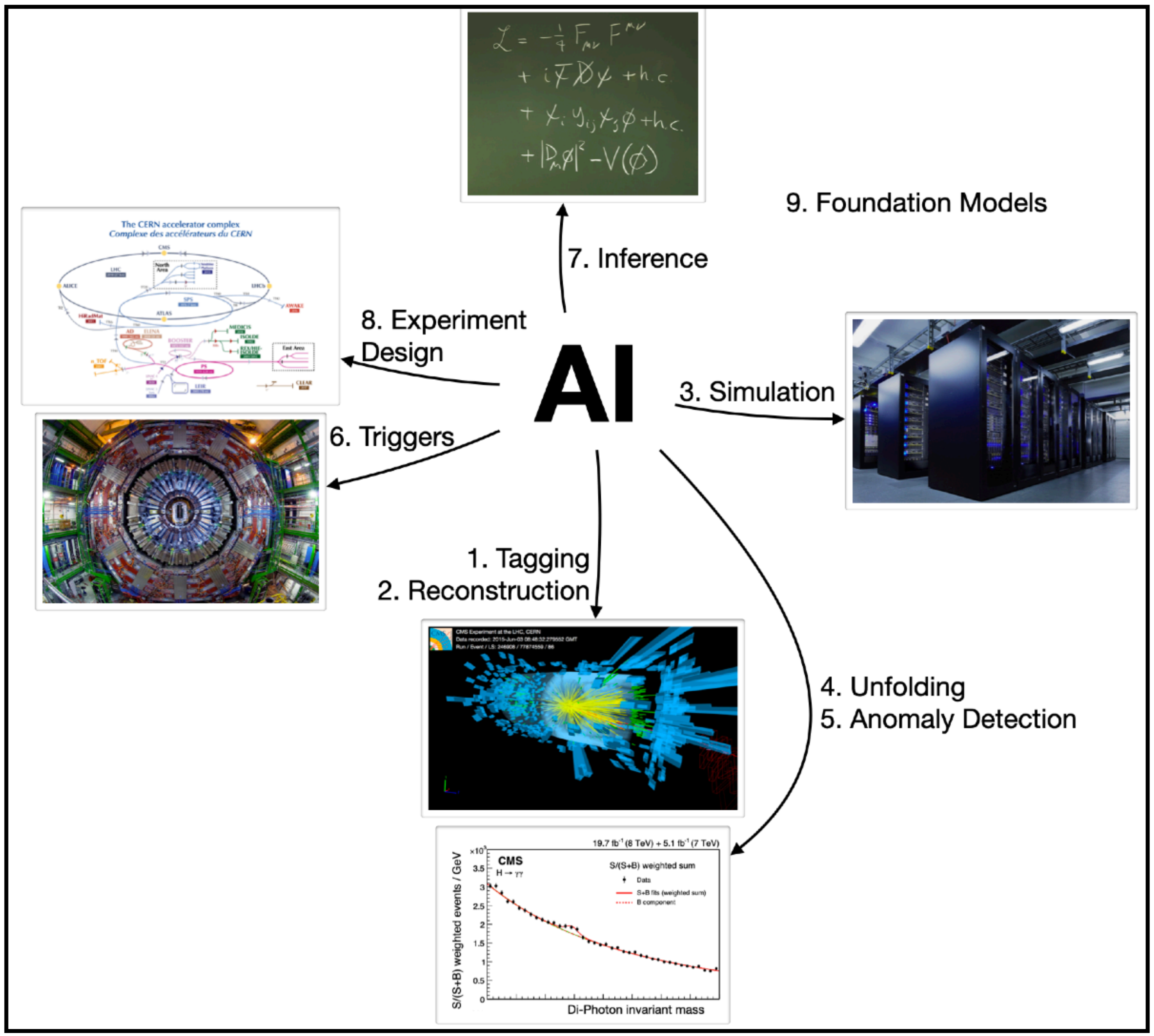
How far can transfer go

New issues from a widely shared model (correlations across experiment?)

How to re-using training data in analysis

Closing

Conclusions



Extremely broad range of application for AI in particle physics

Way beyond concept studies: Modern tools are making a real impact in data analysis

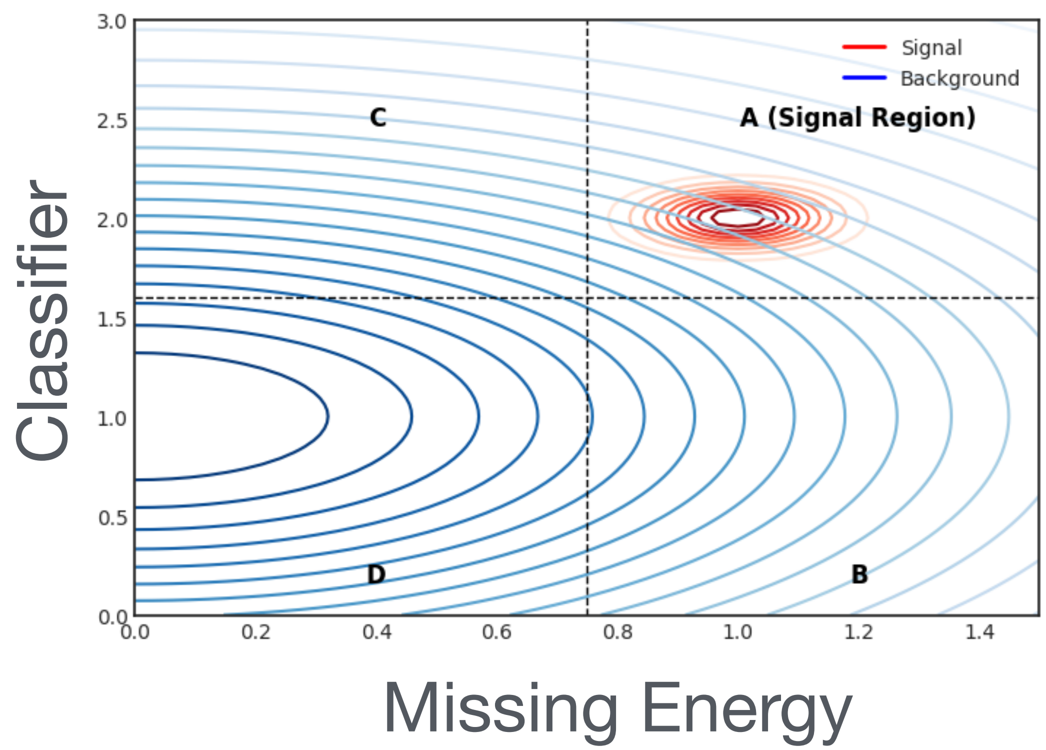
Start to realise fully AI-based processing chains

Significant compute effort: Efficient models and sharing with foundation models matter

Thank you!

Backup

CONRAD

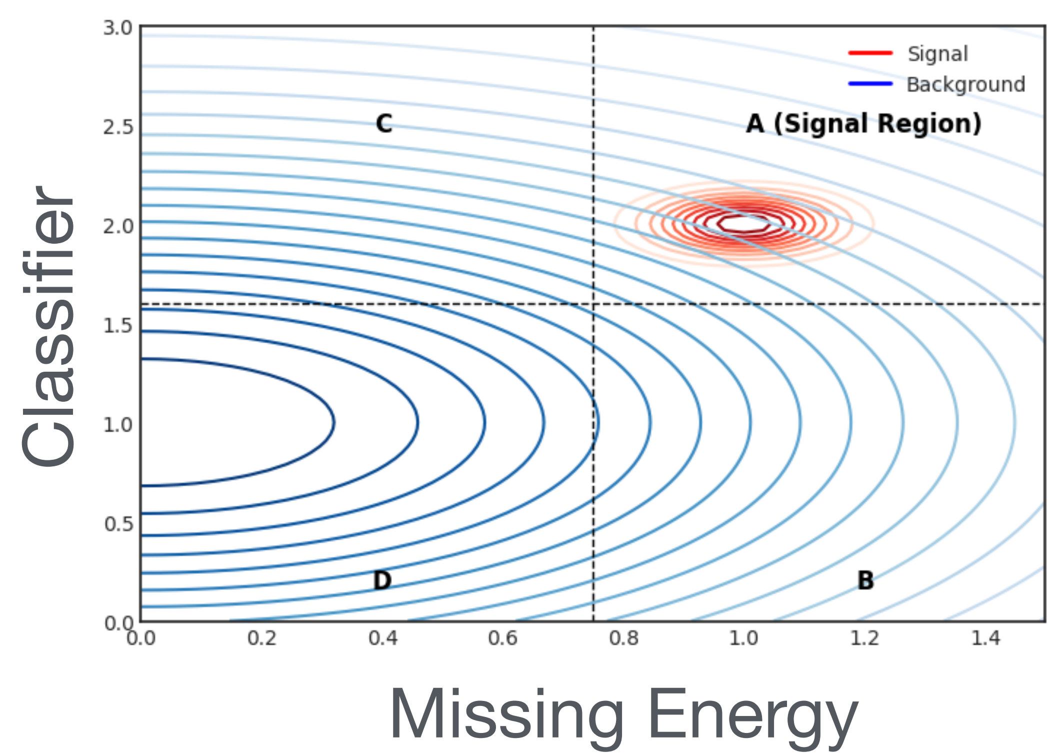


Given a classifier independent of missing energy, could use ABCD method

$$N_{A,bg}^{pred} = \frac{N_{B,bg} \cdot N_{C,bg}}{N_{D,bg}}.$$

GK et al 2404.07258

CONRAD



1. Train NF: conditional mapping for background-only sample to latent-space

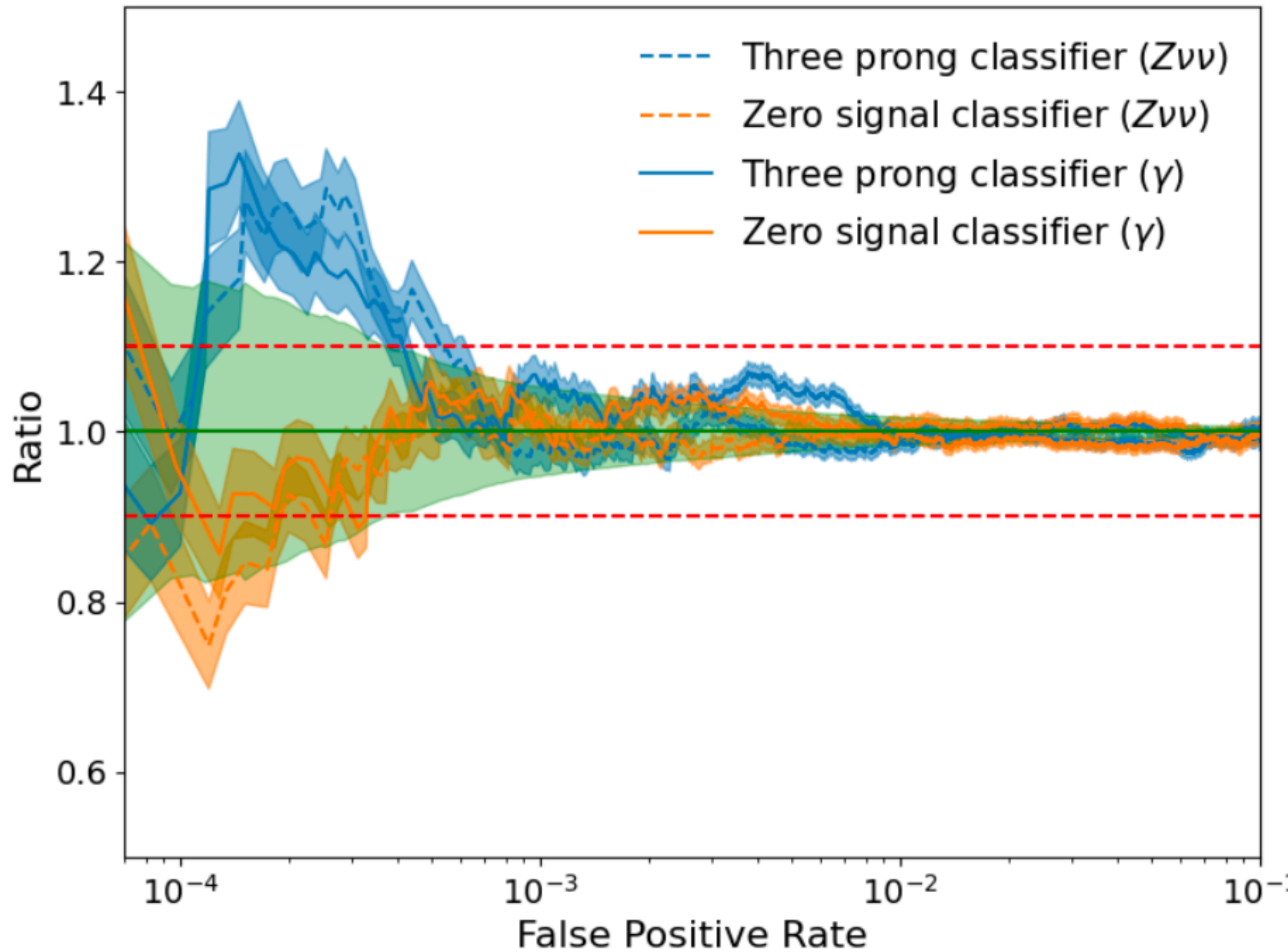
2. Train classifier: Map data to latent space & train classifier vs normal

3. Evaluate: Construct ABCD plane and analyse

GK et al 2404.07258

CONRAD

Background predicted/true



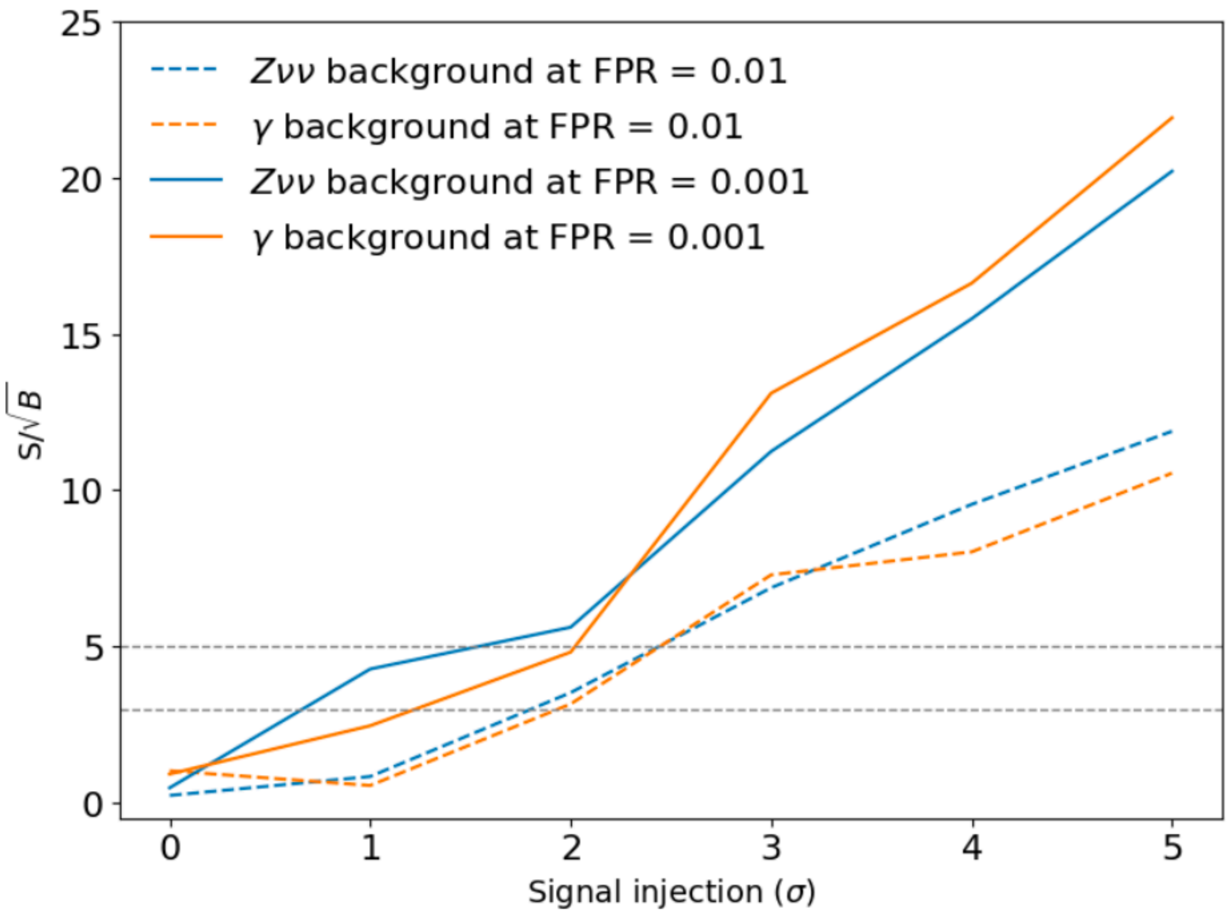
1. Train NF: conditional mapping for background-only sample to latent-space

2. Train classifier: Map data to latent space & train classifier vs normal

3. Evaluate: Construct ABCD plane and analyse

GK et al 2404.07258

CONRAD



1. Train NF: conditional mapping for background-only sample to latent-space

2. Train classifier: Map data to latent space & train classifier vs normal

3. Evaluate: Construct ABCD plane and analyse

GK et al 2404.07258