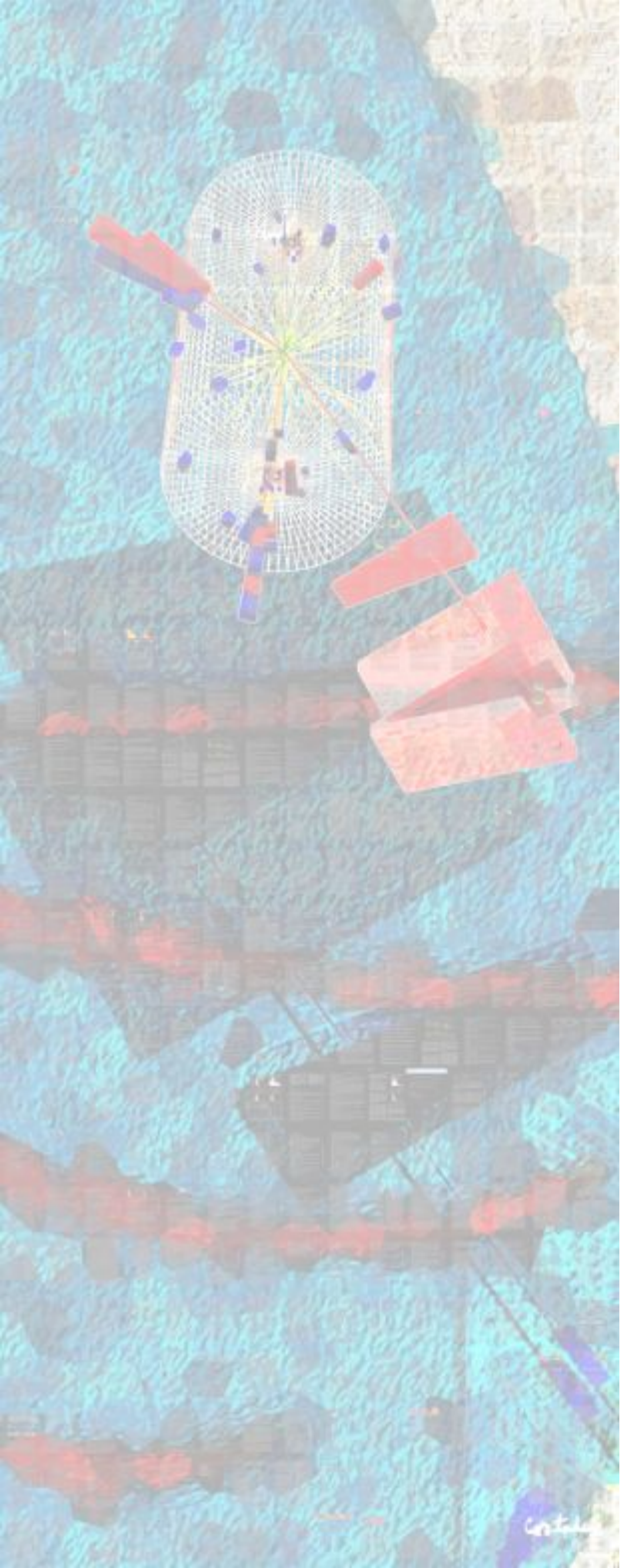


Types of ML in Particle Physics

PHYSTAT: Statistics meets Machine Learning

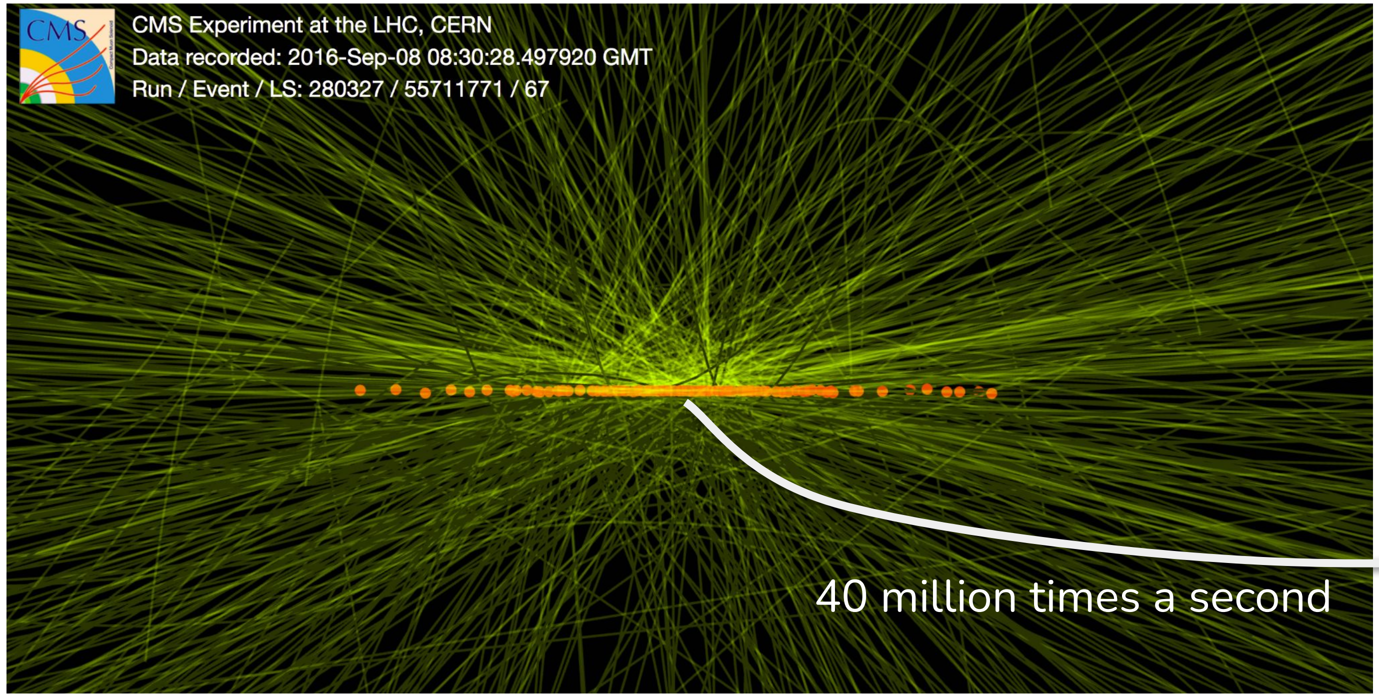
Dr. Jonathon Langford

9th September 2024

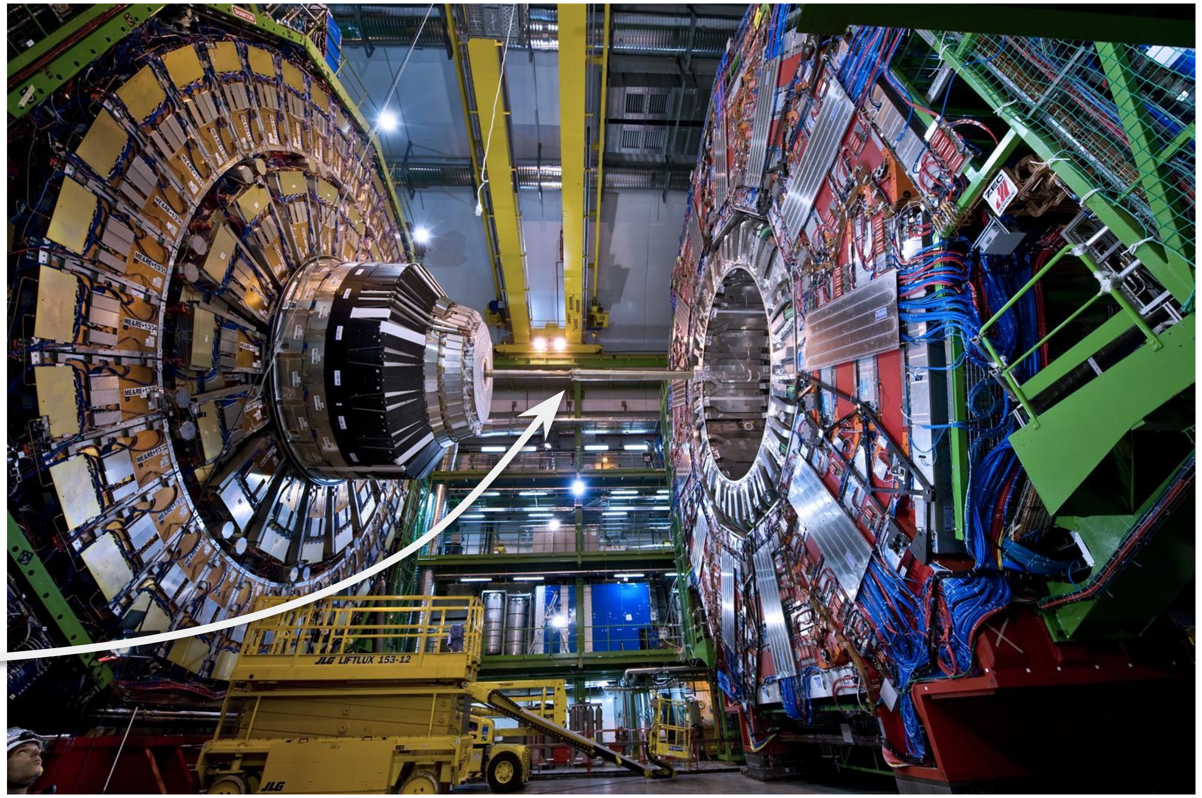
Imperial-X

CMS

Particle physics and big data



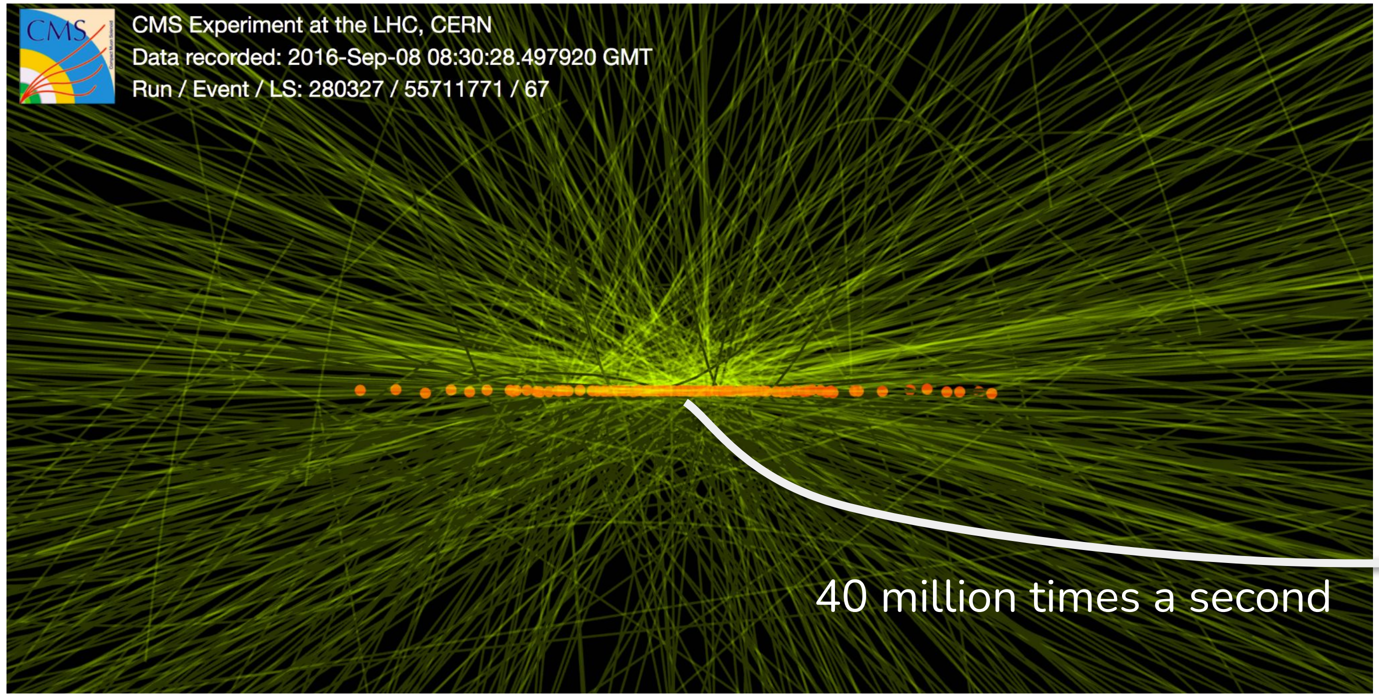
LHC proton-proton collision



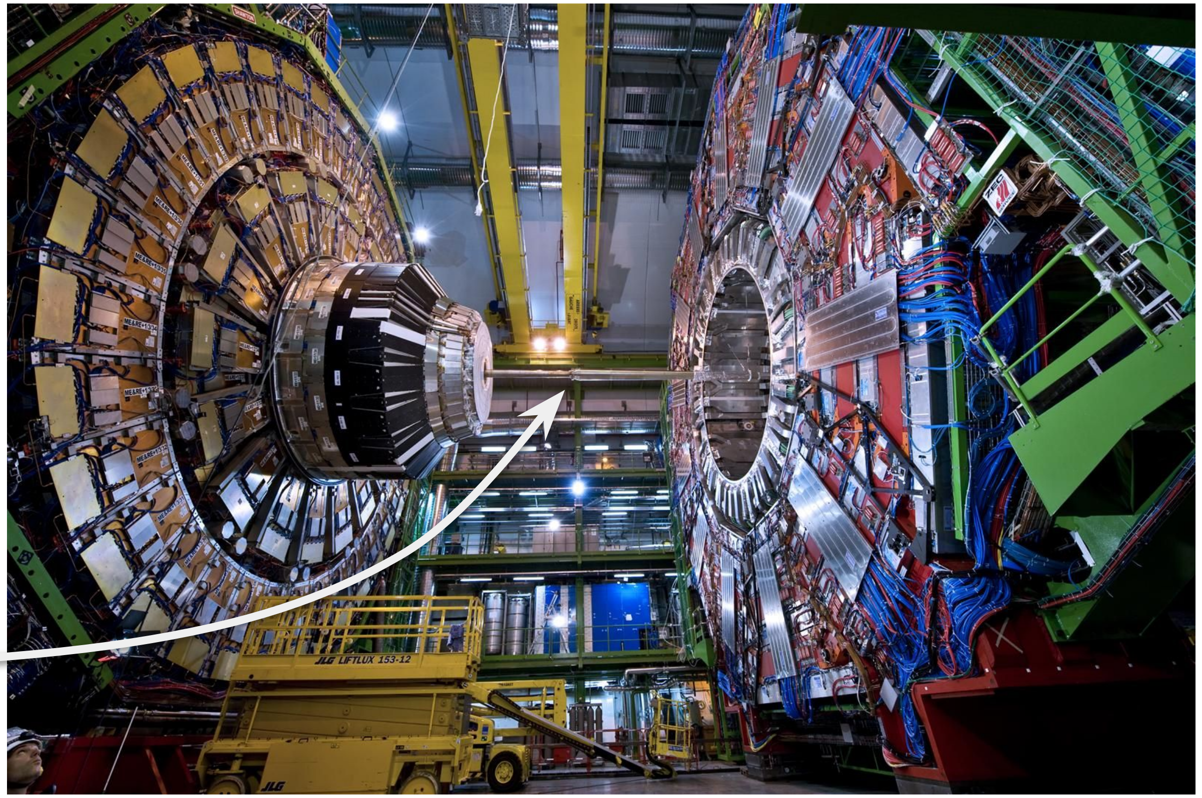
CMS detector with O(100 million) readout channels

- **Astronomically large:** ~500 Tb of data produced by CMS per-second
  - After real-time filtering of collisions (trigger) → Tens of Pb per-year saved offline for further analysis
- **Extremely diverse:** plethora of detector technologies with different geometry/readout
- **Well understood:** small uncertainty in the data
- **Well structured:** significant effort in making datasets easier to work with
- **High-fidelity/quality simulation:** provides “truth”

Particle physics and big data



LHC proton-proton collision

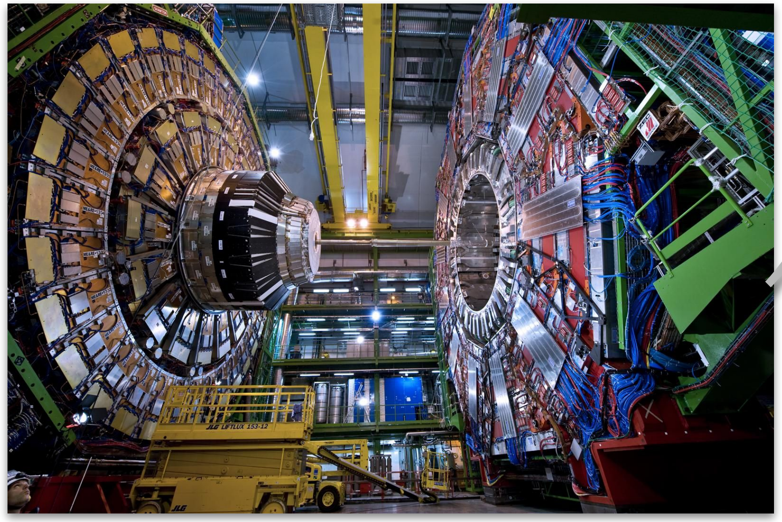


CMS detector with O(100 million) readout channels

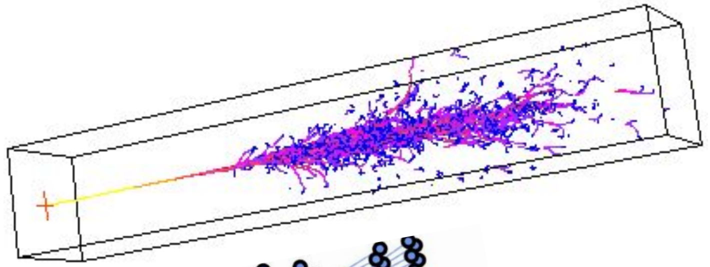
Ideal playground for Machine Learning initiatives

Analysis chain

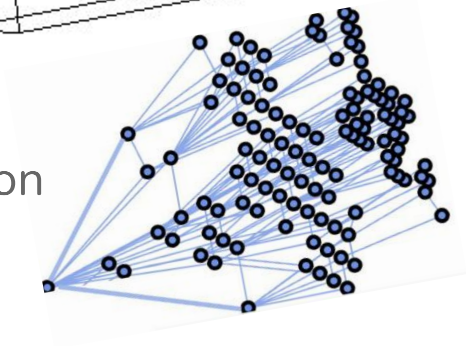
Data-taking



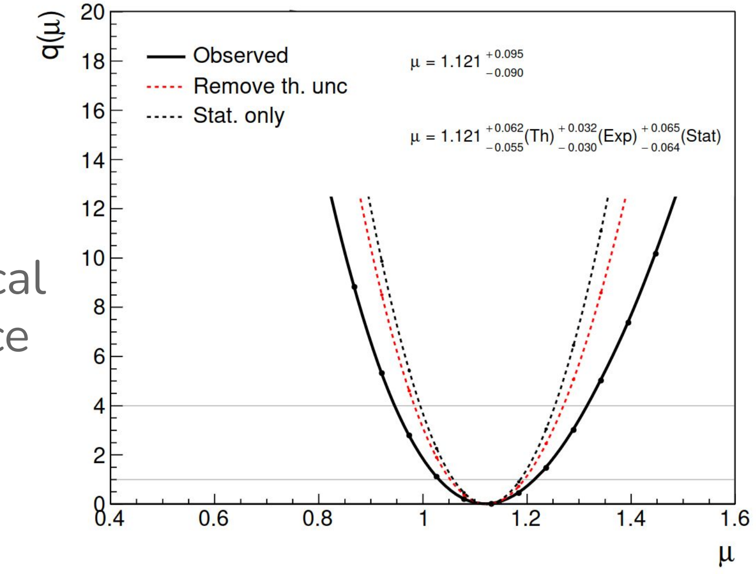
Online filtering
(Trigger)



Event reconstruction

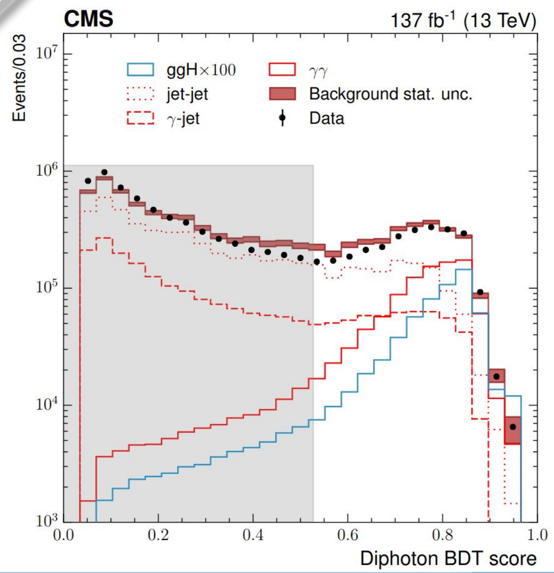


Statistical
inference

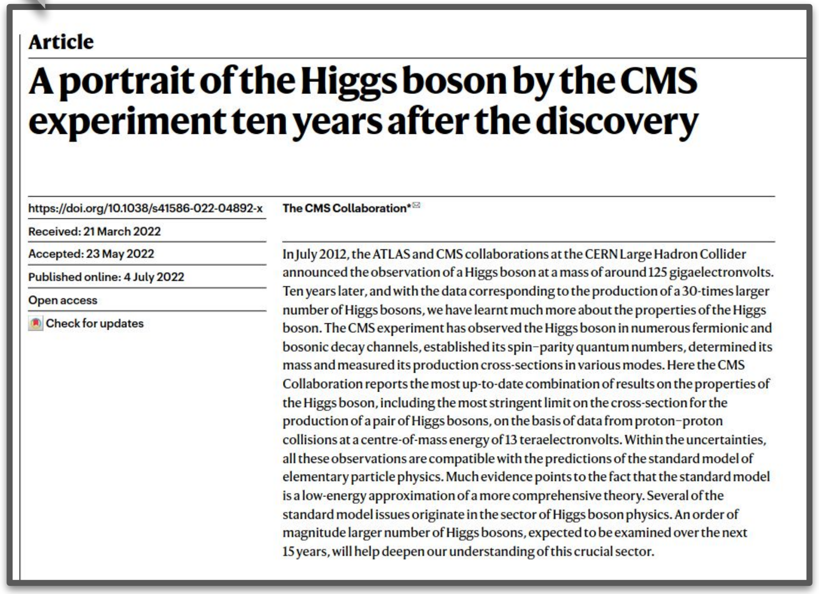


Computing
(dataset storage/retrieval)

Event
classification



Publication



Article

A portrait of the Higgs boson by the CMS experiment ten years after the discovery

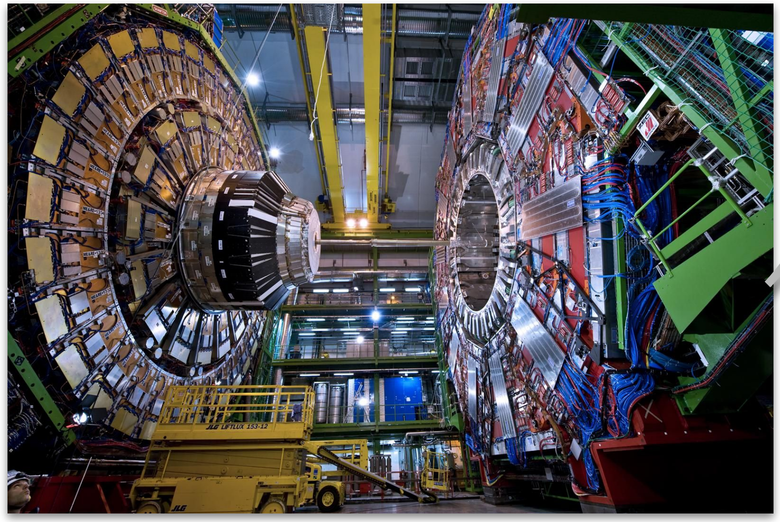
https://doi.org/10.1038/s41586-022-04892-x
Received: 21 March 2022
Accepted: 23 May 2022
Published online: 4 July 2022
Open access
Check for updates

The CMS Collaboration*

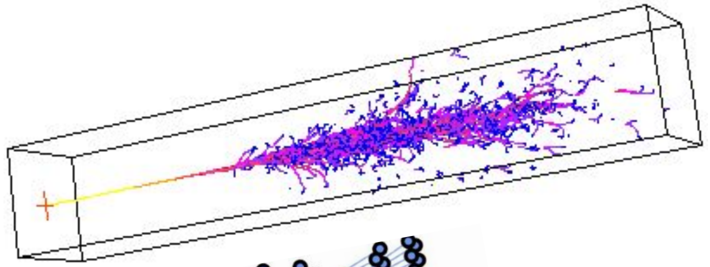
In July 2012, the ATLAS and CMS collaborations at the CERN Large Hadron Collider announced the observation of a Higgs boson at a mass of around 125 gigaelectronvolts. Ten years later, and with the data corresponding to the production of a 30-times larger number of Higgs bosons, we have learnt much more about the properties of the Higgs boson. The CMS experiment has observed the Higgs boson in numerous fermionic and bosonic decay channels, established its spin–parity quantum numbers, determined its mass and measured its production cross-sections in various modes. Here the CMS Collaboration reports the most up-to-date combination of results on the properties of the Higgs boson, including the most stringent limit on the cross-section for the production of a pair of Higgs bosons, on the basis of data from proton–proton collisions at a centre-of-mass energy of 13 teraelectronvolts. Within the uncertainties, all these observations are compatible with the predictions of the standard model of elementary particle physics. Much evidence points to the fact that the standard model is a low-energy approximation of a more comprehensive theory. Several of the standard model issues originate in the sector of Higgs boson physics. An order of magnitude larger number of Higgs bosons, expected to be examined over the next 15 years, will help deepen our understanding of this crucial sector.

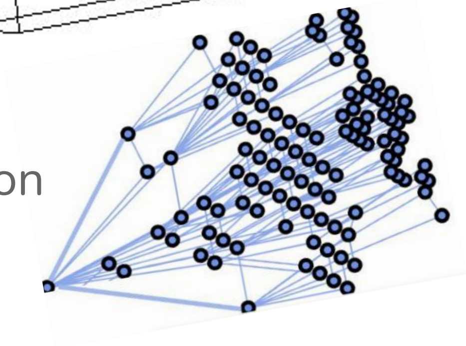
Analysis chain

Data-taking



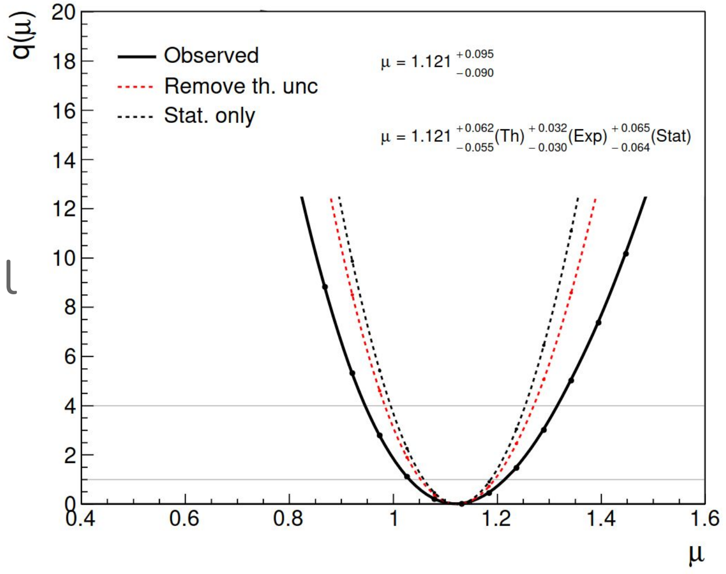
Online filtering (Trigger)





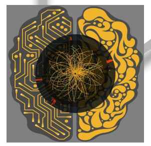
Event reconstruction

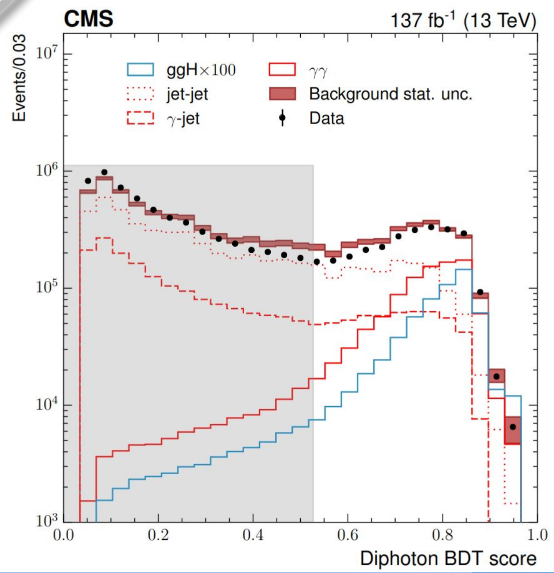
Statistical inference



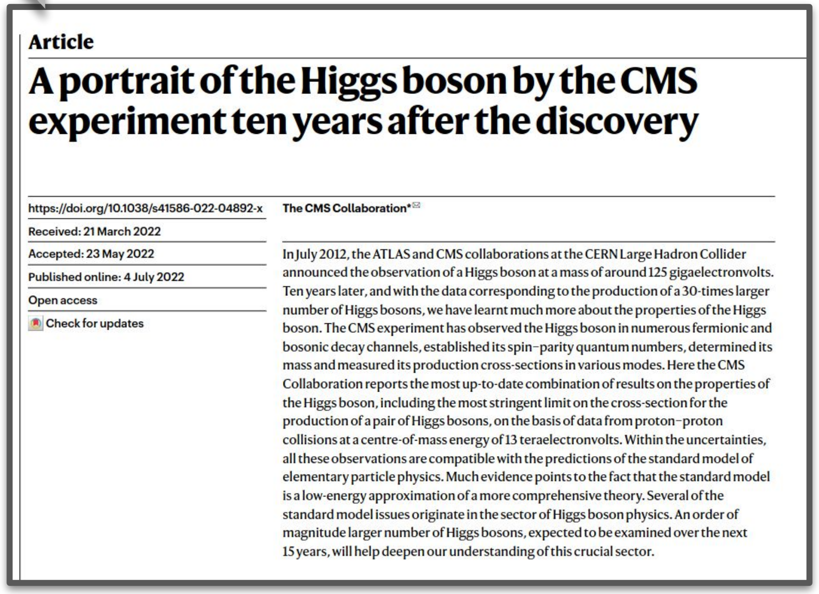
Computing (dataset storage/retrieval)

Event classification





Publication



Article

A portrait of the Higgs boson by the CMS experiment ten years after the discovery

https://doi.org/10.1038/s41586-022-04892-x

Received: 21 March 2022

Accepted: 23 May 2022

Published online: 4 July 2022

Open access

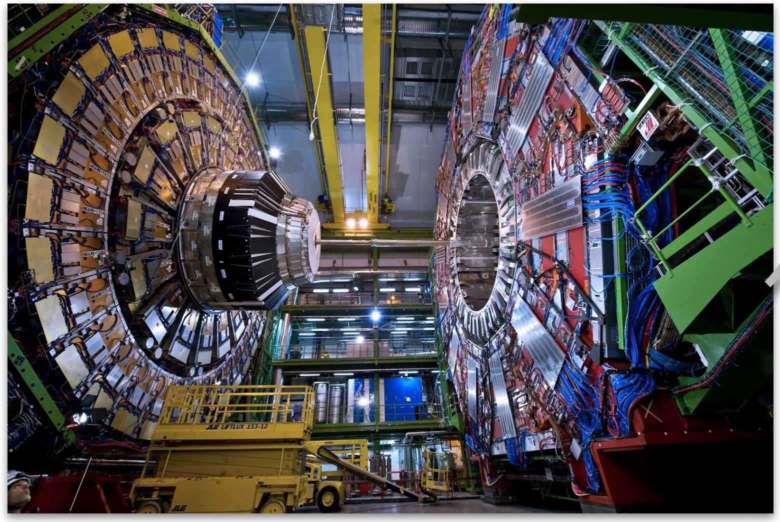
Check for updates

The CMS Collaboration*

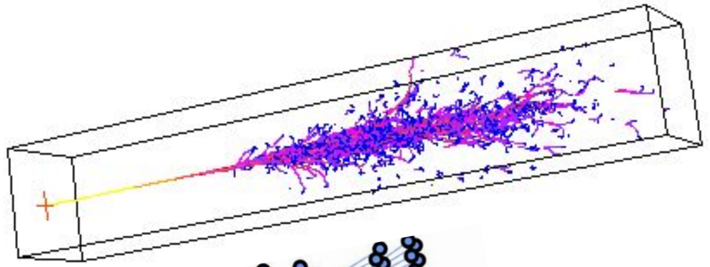
In July 2012, the ATLAS and CMS collaborations at the CERN Large Hadron Collider announced the observation of a Higgs boson at a mass of around 125 gigaelectronvolts. Ten years later, and with the data corresponding to the production of a 30-times larger number of Higgs bosons, we have learnt much more about the properties of the Higgs boson. The CMS experiment has observed the Higgs boson in numerous fermionic and bosonic decay channels, established its spin–parity quantum numbers, determined its mass and measured its production cross-sections in various modes. Here the CMS Collaboration reports the most up-to-date combination of results on the properties of the Higgs boson, including the most stringent limit on the cross-section for the production of a pair of Higgs bosons, on the basis of data from proton–proton collisions at a centre-of-mass energy of 13 teraelectronvolts. Within the uncertainties, all these observations are compatible with the predictions of the standard model of elementary particle physics. Much evidence points to the fact that the standard model is a low-energy approximation of a more comprehensive theory. Several of the standard model issues originate in the sector of Higgs boson physics. An order of magnitude larger number of Higgs bosons, expected to be examined over the next 15 years, will help deepen our understanding of this crucial sector.

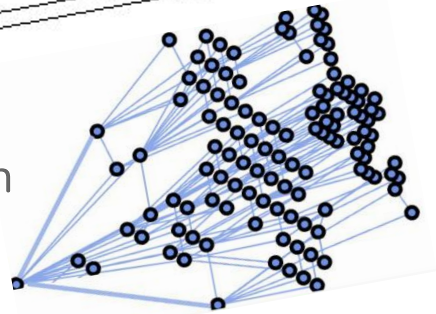
Analysis chain

Data-taking



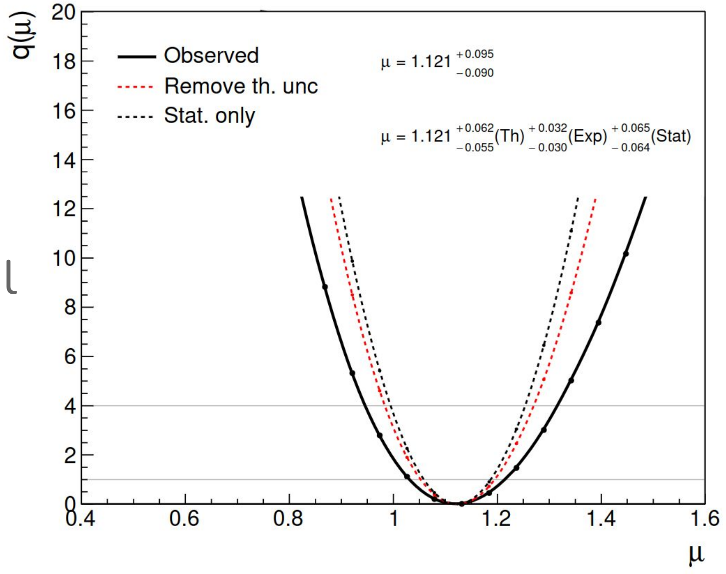
Online filtering
(Trigger)





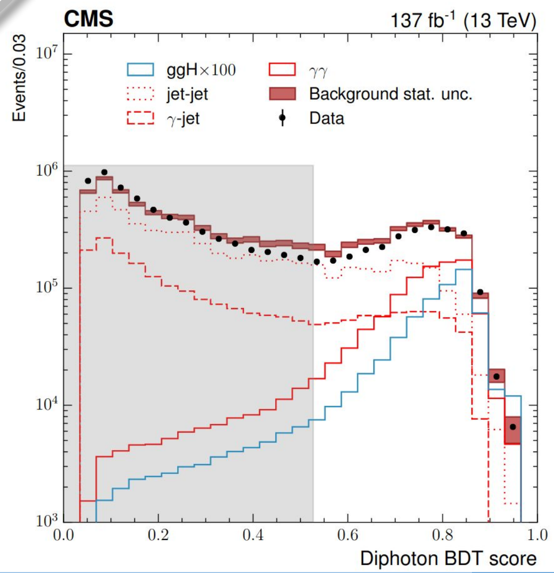
Event reconstruction

Statistical
inference

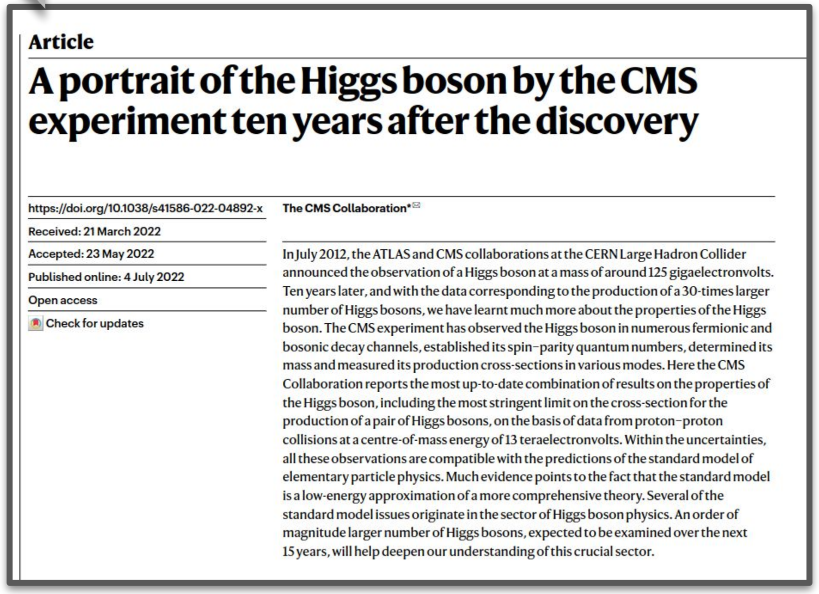


Computing
(dataset storage/retrieval)

Event
classification



Publication



Article

A portrait of the Higgs boson by the CMS experiment ten years after the discovery

https://doi.org/10.1038/s41586-022-04892-x

Received: 21 March 2022

Accepted: 23 May 2022

Published online: 4 July 2022

Open access

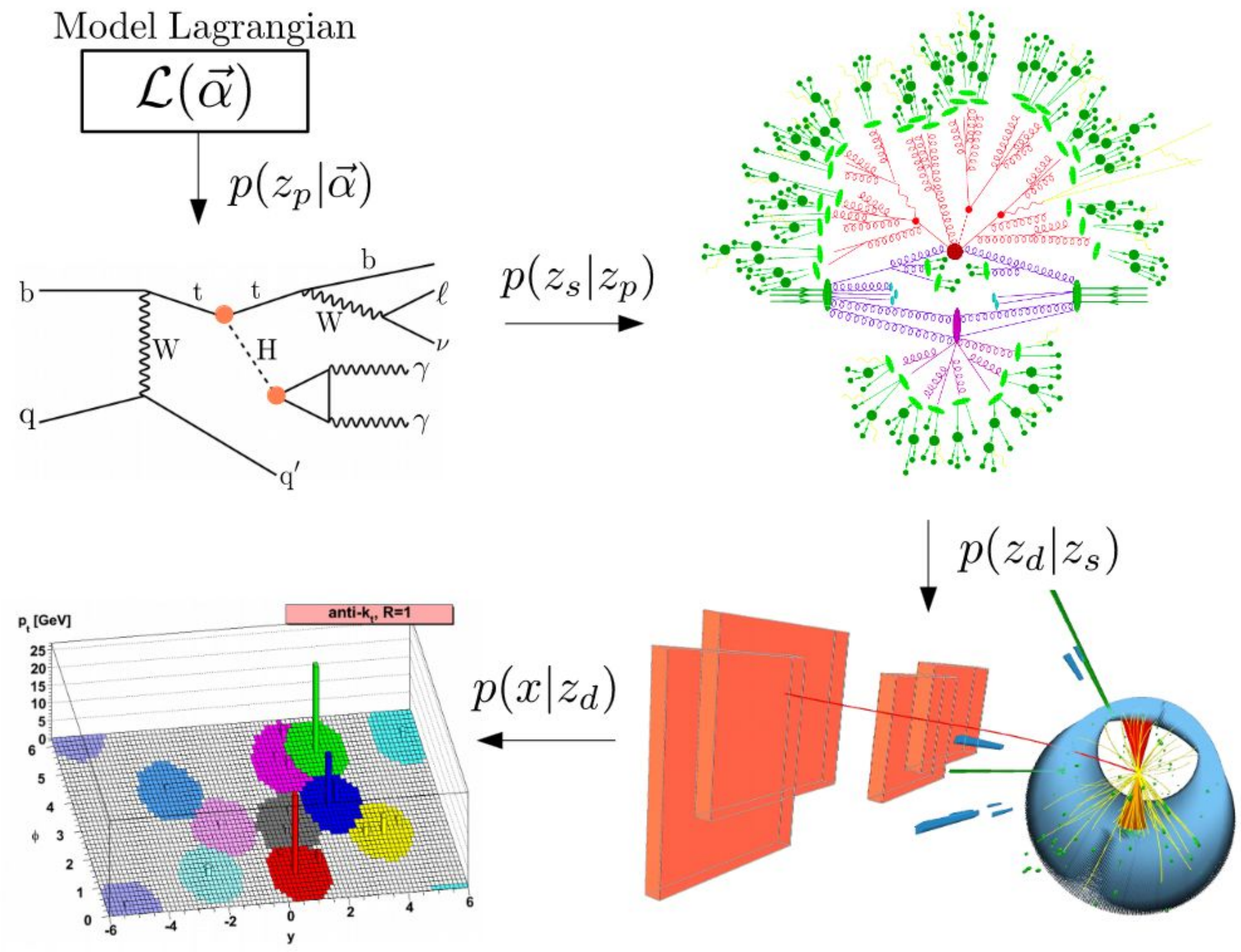
Check for updates

The CMS Collaboration*

In July 2012, the ATLAS and CMS collaborations at the CERN Large Hadron Collider announced the observation of a Higgs boson at a mass of around 125 gigaelectronvolts. Ten years later, and with the data corresponding to the production of a 30-times larger number of Higgs bosons, we have learnt much more about the properties of the Higgs boson. The CMS experiment has observed the Higgs boson in numerous fermionic and bosonic decay channels, established its spin–parity quantum numbers, determined its mass and measured its production cross-sections in various modes. Here the CMS Collaboration reports the most up-to-date combination of results on the properties of the Higgs boson, including the most stringent limit on the cross-section for the production of a pair of Higgs bosons, on the basis of data from proton–proton collisions at a centre-of-mass energy of 13 teraelectronvolts. Within the uncertainties, all these observations are compatible with the predictions of the standard model of elementary particle physics. Much evidence points to the fact that the standard model is a low-energy approximation of a more comprehensive theory. Several of the standard model issues originate in the sector of Higgs boson physics. An order of magnitude larger number of Higgs bosons, expected to be examined over the next 15 years, will help deepen our understanding of this crucial sector.

Monte-Carlo simulation

- [Theory → observables] is described by highly-intractable likelihood
- Use high-fidelity MC simulation of each stage of collision event



High dimensional integral over latent variables

$$\boxed{L(x\,|\,\vec{\alpha}) = \int \mathrm{d}z_d \int \mathrm{d}z_s \int \mathrm{d}z_p\, p(x|z_d)p(z_d|z_s)p(z_s|z_p)p(z_p|\vec{\alpha})}$$

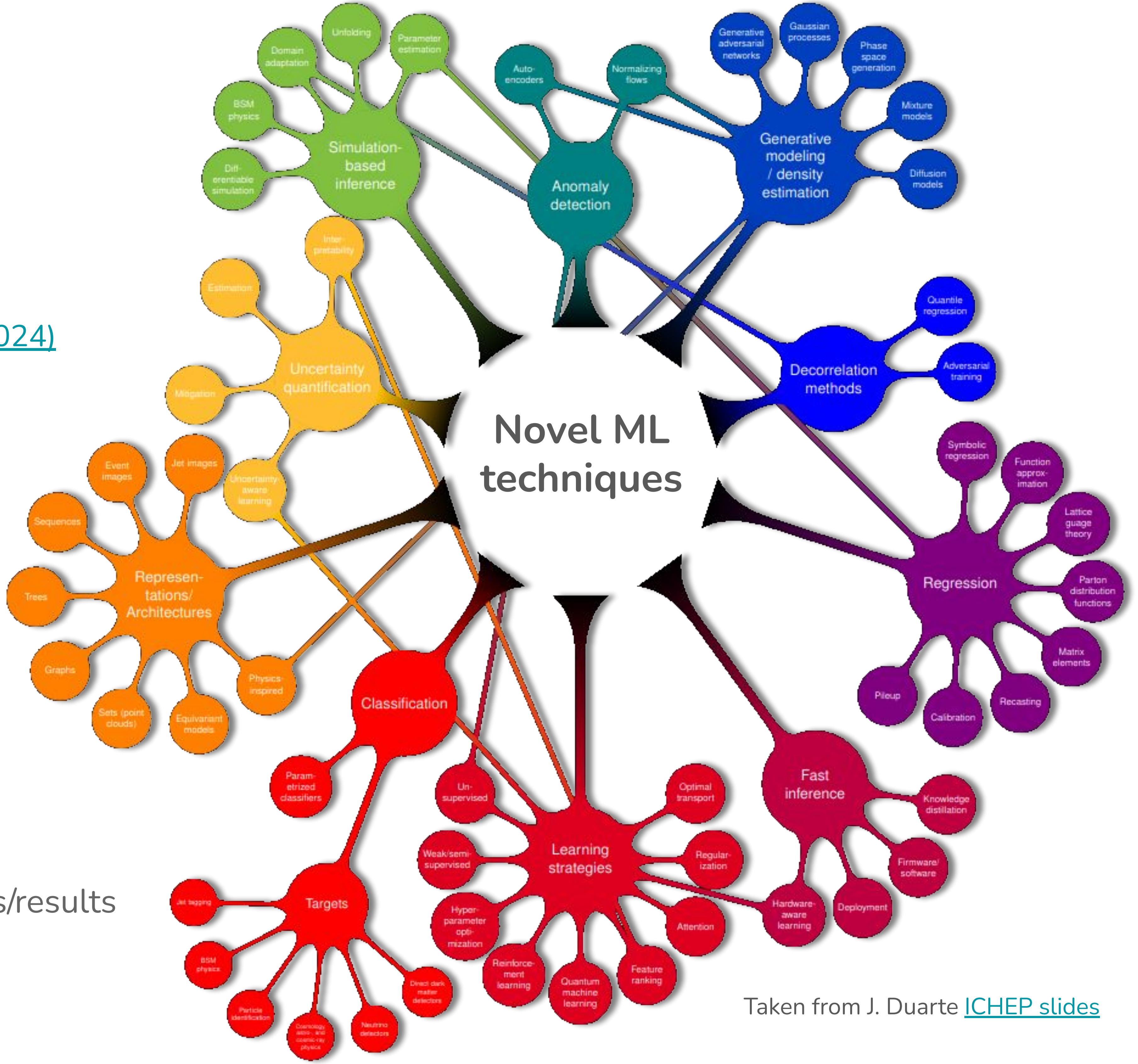
Observables e.g. reconstructed energies, momenta and angles of all final state particles

Fundamental physics **parameters of interest** e.g. Higgs boson mass

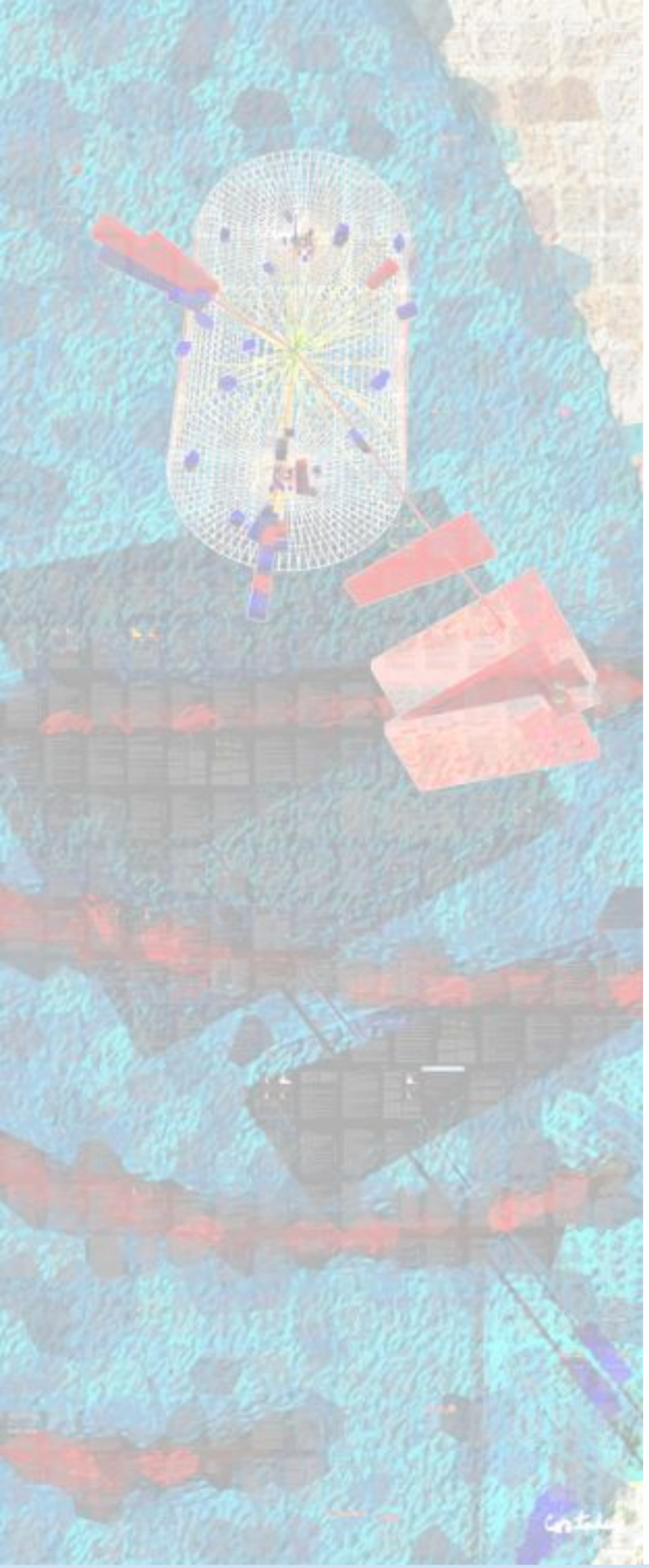
- Provides "truth" for inference on real data
  - Accurate simulation is crucial to avoid bias (calibration)
- Labelled collisions for supervised learning

ML in particle physics

- **<u>Disclaimer:</u>** collider, CMS, experimental
  - [Neutrino Physics & ML workshop, ETH (2024)]
  - [Theoretical HEP & AI talk, EuCAIFCon (2024)]
  - [Latest ML developments for LHCb, EP-IT seminar (2024)]
  - [DM direct detection [arXiv:2406.10372]]
- Topics:
  - Object identification & reconstruction
  - Event classification
  - Simulation (generative)
  - Inference
- Try to keep relevant with mostly new applications/results



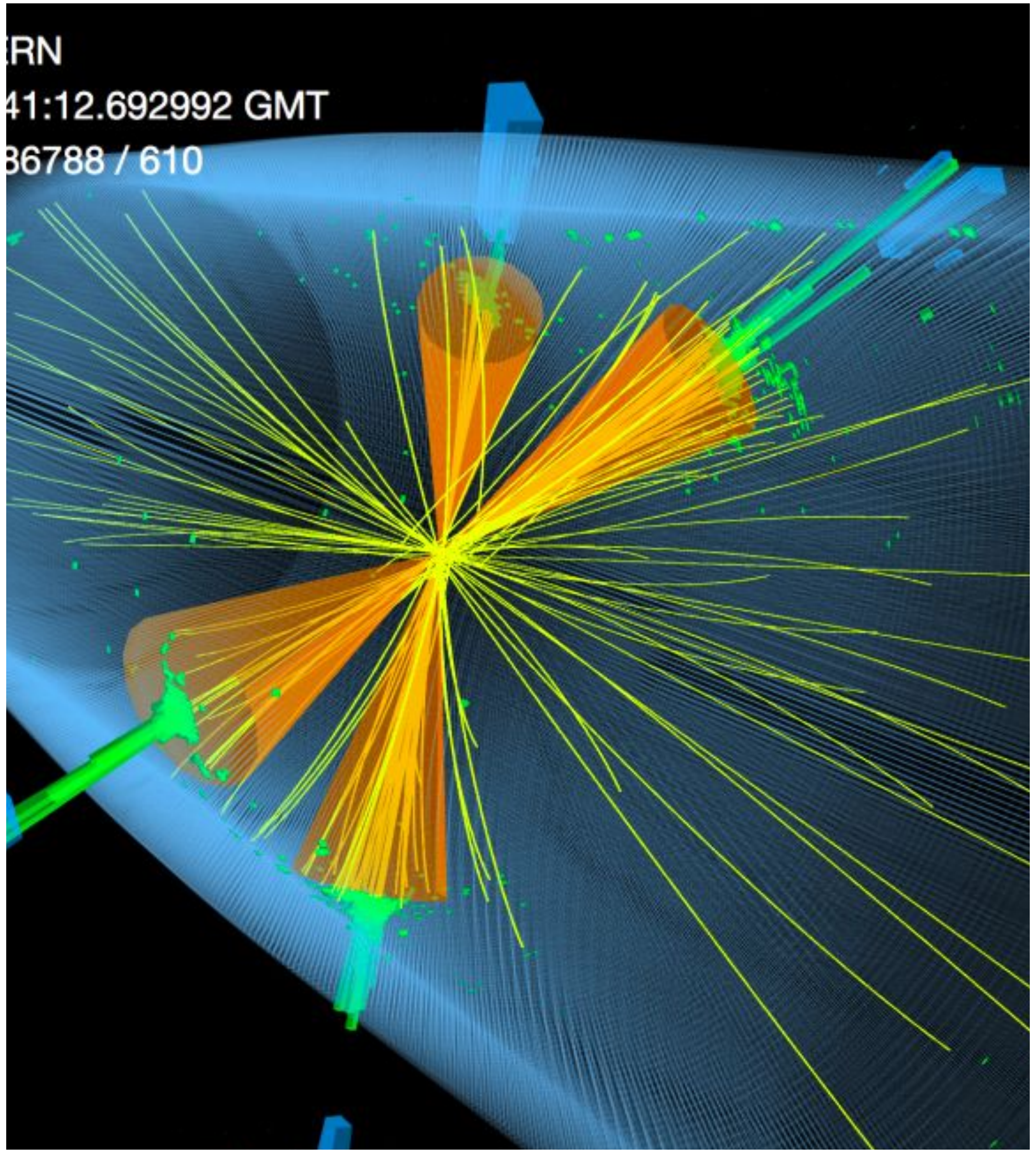
Taken from J. Duarte [ICHEP slides]



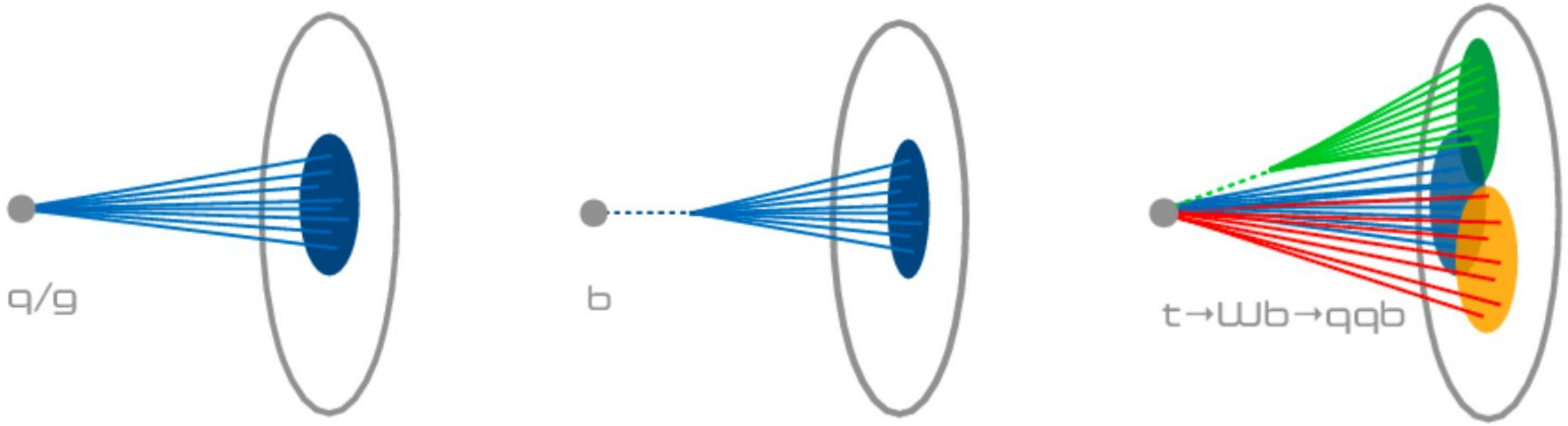
Object identification & reconstruction

Jet classification

- **Jet** = spray of particles (cone) produced by hadronization of a quark/gluon when ejected from high-energy collision



- Jets come in different "flavours" → different substructure



Light-quark jets
e.g. QCD background

b-quark jets
e.g. H→bb

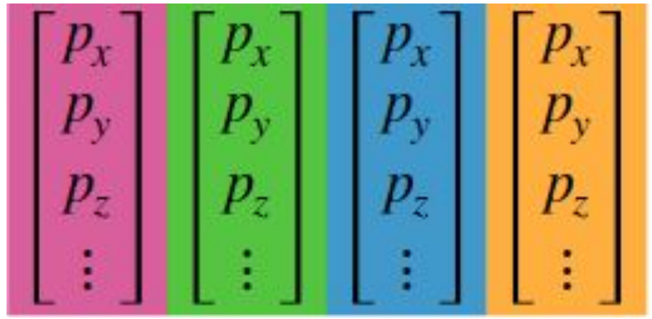
Boosted (large radius)
e.g. top decays

- Jet constituent particles produce patterns of "hits" as they traverse detector
  - Essentially a pattern recognition problem
  - Has become a huge frontier in ML over last years (see ML4Jets)

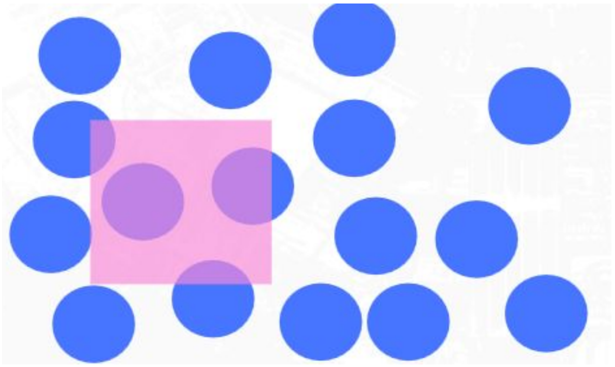
Jet representations

- Evolution of representations:

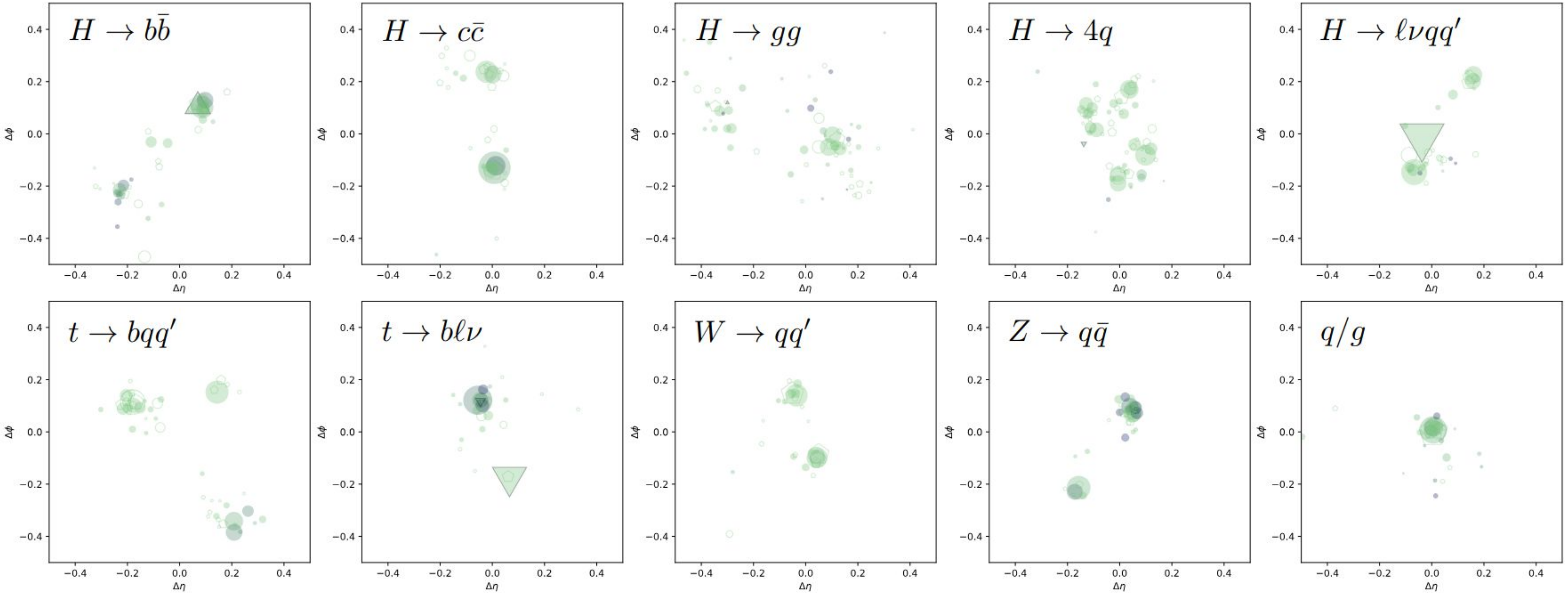
Image-based (CNN)
Difficult to combine non-additive quantities, very sparse (>90% pixels are blank)



Sequences (RNN)
Can include any kind of constituent feature, no issues with sparse data, sorted list e.g. decreasing pT



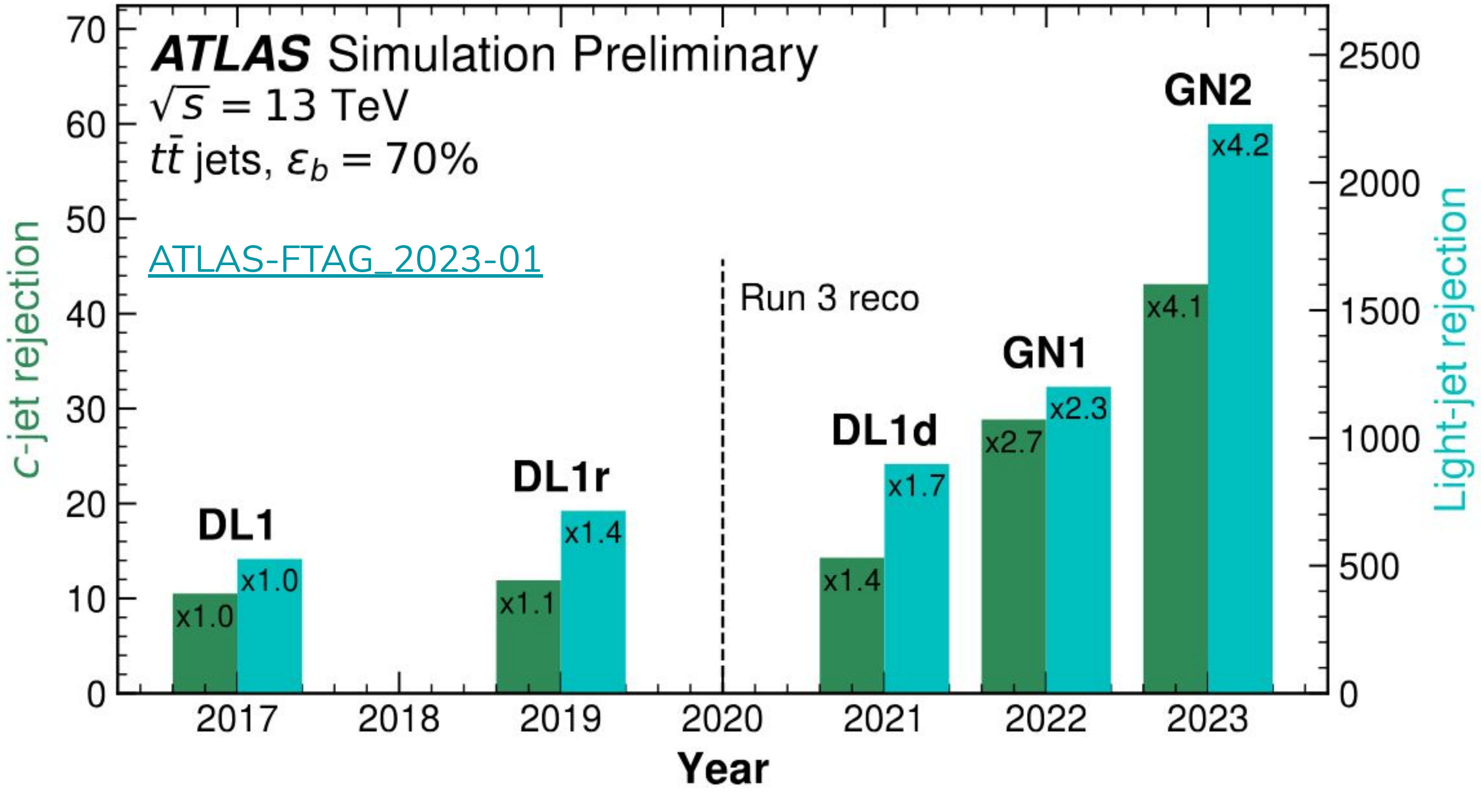
Point/particle cloud (GNN)
Unordered list is permutation invariant, no issues with sparse data

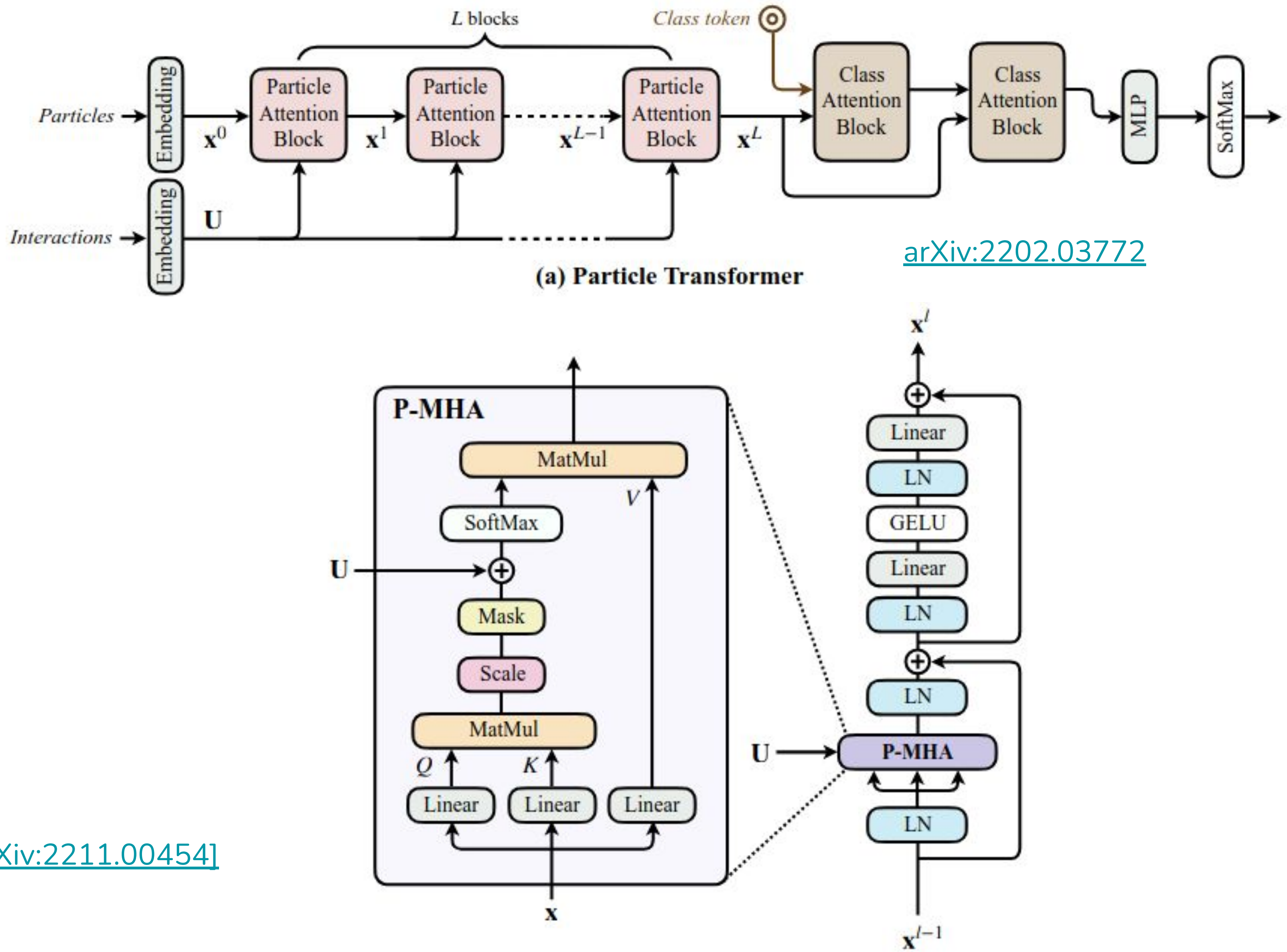


Ten types of jets viewed as particle clouds
Coordinates = Direction of flight
Size = Energy
Shape = Particle ID
Solid/Hollow = Charged/Neutral
Blueness = Displacement from IP
arXiv:2202.03772

Jet classification

- Huge advances by using low-level information with **Graph Neural Networks** (e.g. ParticleNet in CMS, GN1/GN2 in ATLAS)
- Now **Transformers** (e.g. ParT): "attention" gives more weight to certain jet constituents



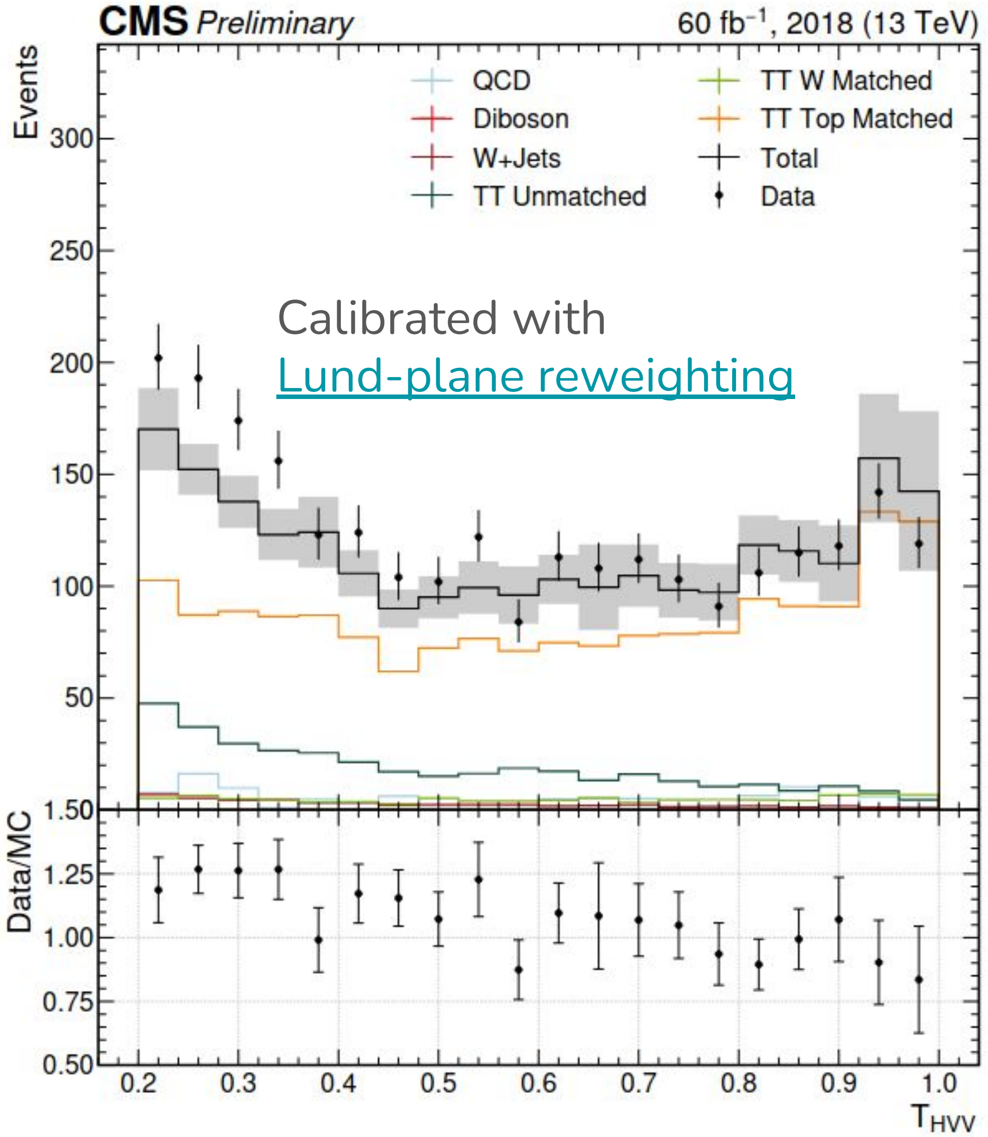


- Add domain knowledge e.g. Lorentz Invariance in Pelican [arXiv:2211.00454]
  - Competitive performance with far fewer parameters!

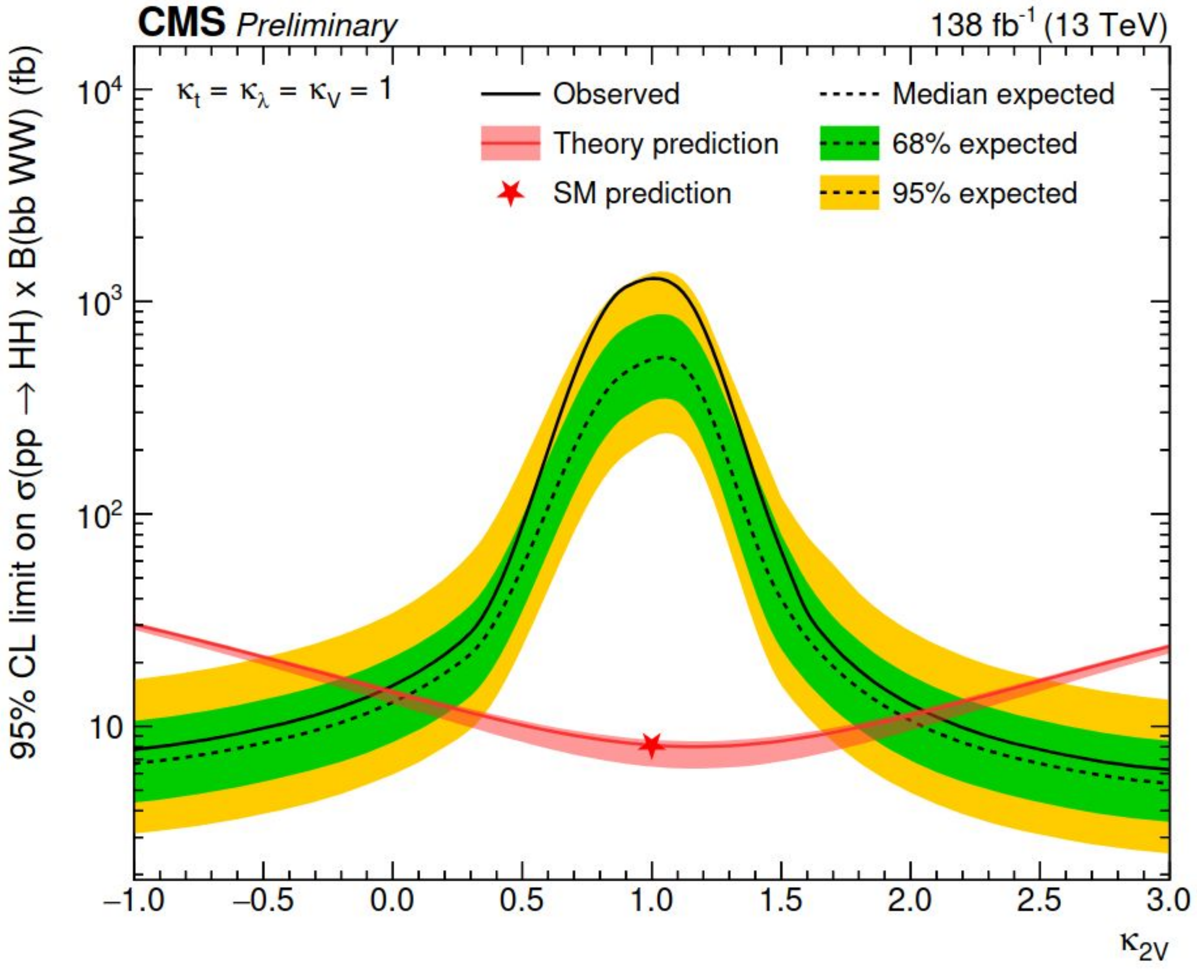
Impact of improved jet classification

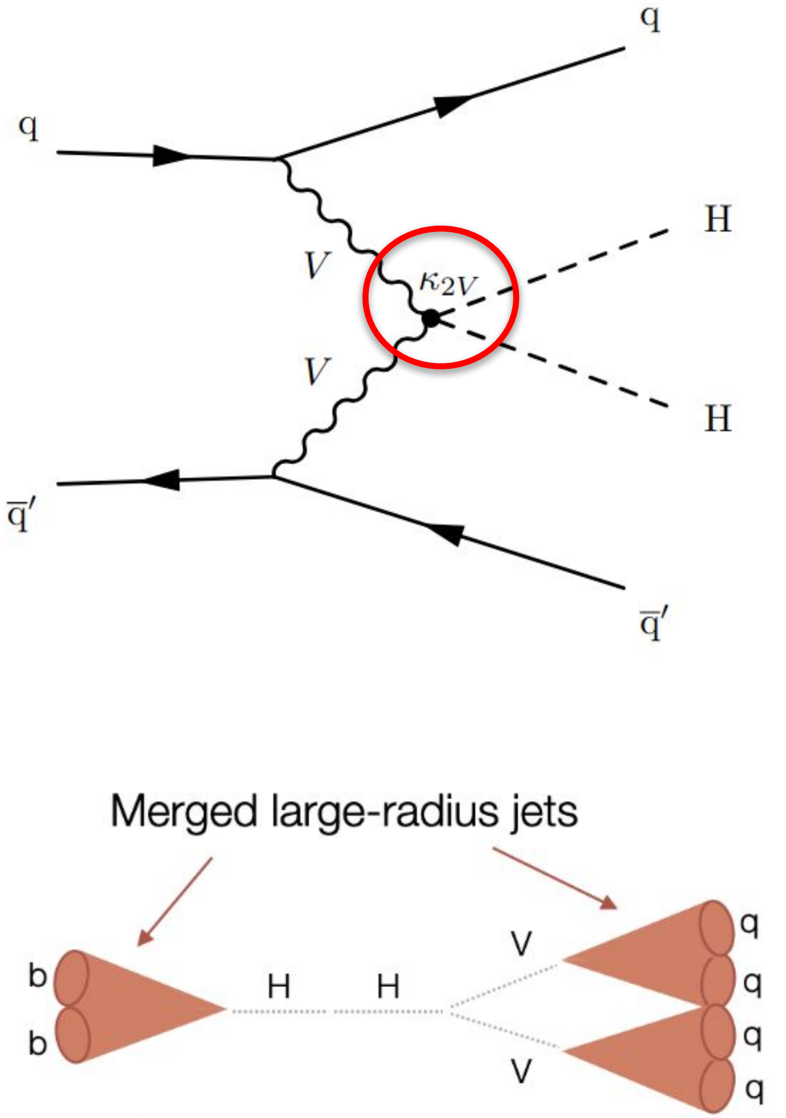
[CMS-PAS-HIG-23-012]

- Translates to significant improvements in particle physics measurements/searches
  - Search for boosted HH → bbVV → bb4q
  - Global particle transformer (GloParT) classifier to identify boosted VV→4q



Calibrated with Lund-plane reweighting





- Second best constraint on κ_{2V} from CMS to-date!

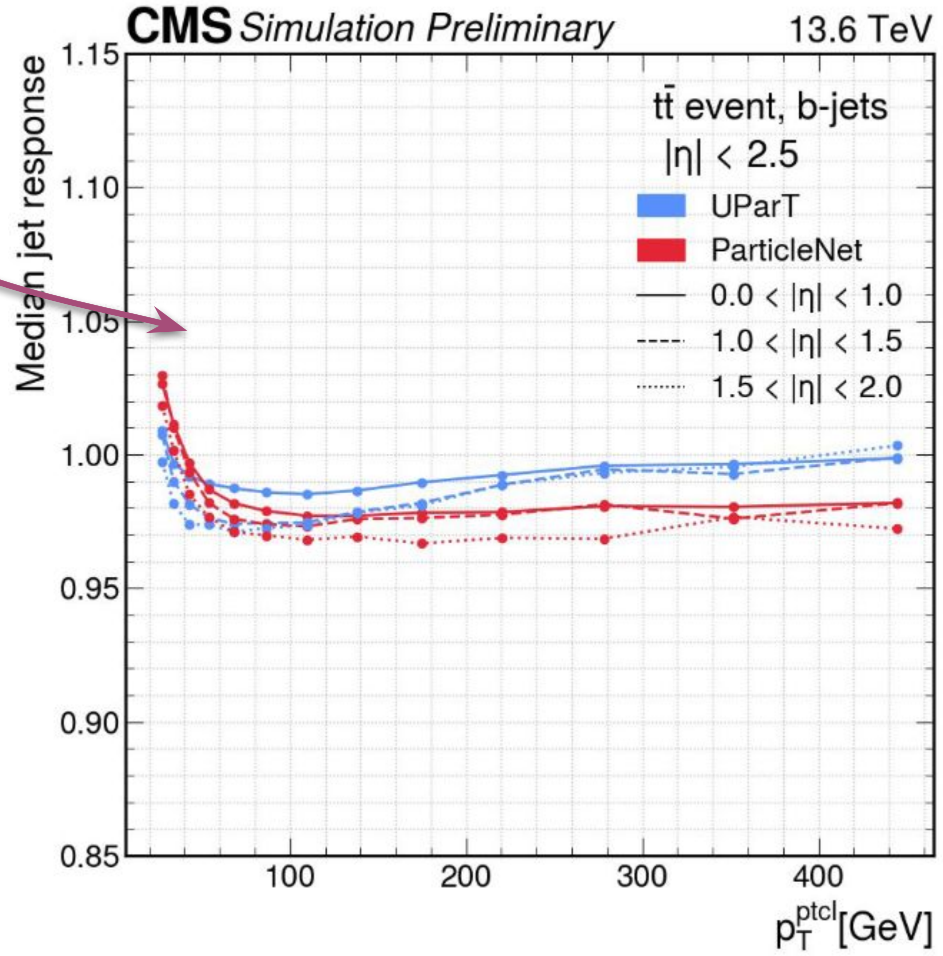
All-in-one algorithms

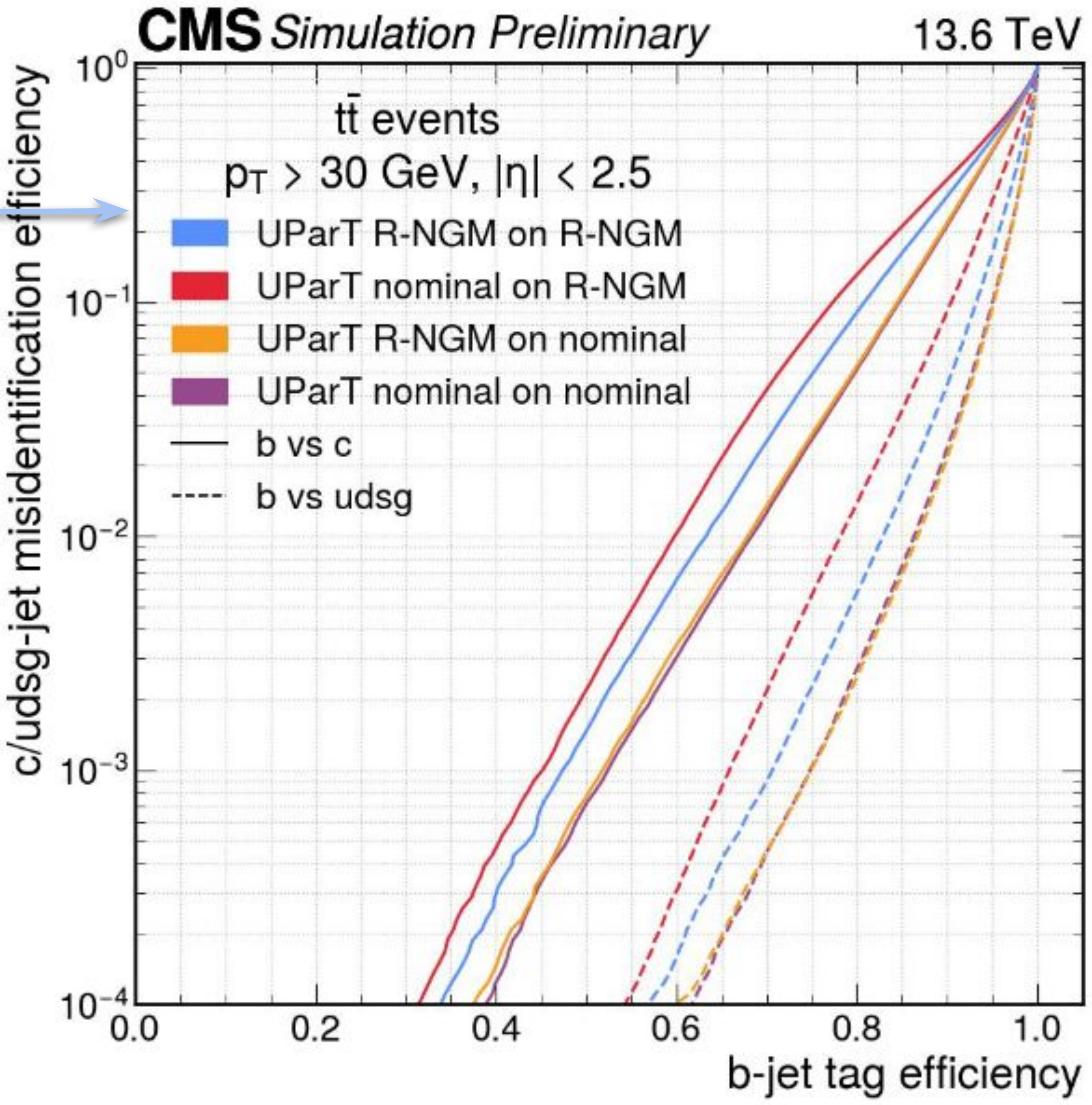
[CMS-DP-2024-066]

- Unified particle transformer for small-radius (AK4) jets: UParT
    - Simultaneously identify heavy-flavour (b, c), identify hadronically decaying tau-leptons, identify s-jets, regress jet energy, estimate jet energy resolution

$$L = \overbrace{CatEntropy(x, x_{truth})}^{\textit{Classification}} + \overbrace{\gamma_{regr} \cdot log(cosh(y - y_{truth}))}^{\textit{Regression}} + \overbrace{\gamma_{quantile} \cdot [p_{0.16}(z - z_{truth}) + p_{0.84}(z - z_{truth})]}^{\textit{Quantile regression (resolution estimation)}}$$

- Adversarial training
    - Rectified Normed Gradient Method (R-NGM)
    - Improve robustness against perturbed data



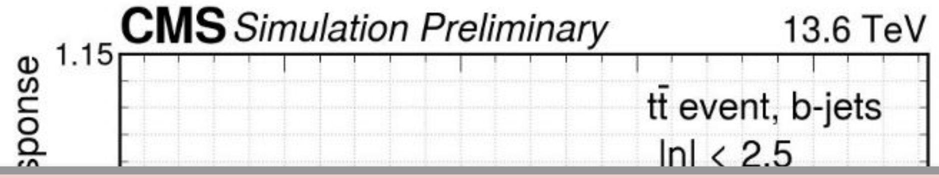


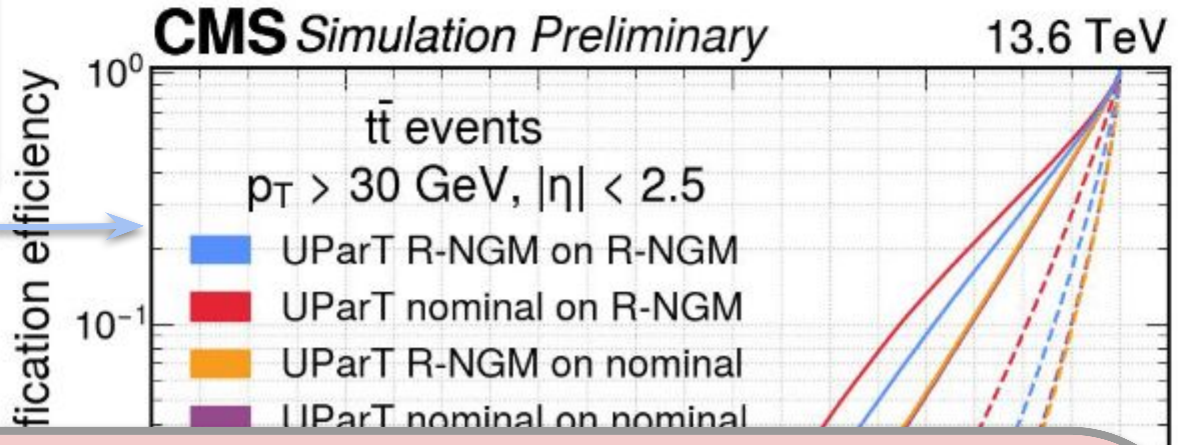
All-in-one algorithms

[CMS-DP-2024-066]

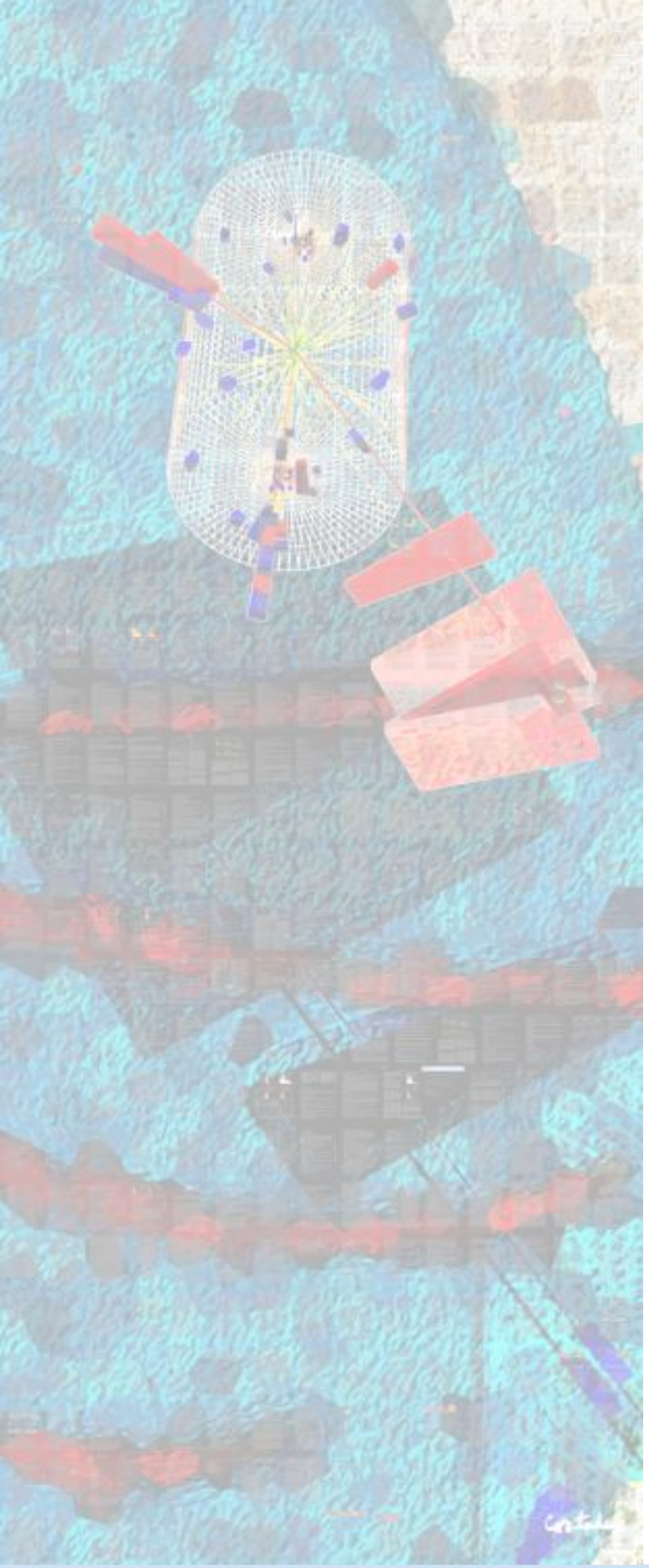
- Unified particle transformer for small-radius (AK4) jets: UParT
  - Simultaneously identify heavy-flavour (b, c), identify hadronically decaying tau-leptons, identify s-jets, regress jet energy, estimate jet energy resolution

$$L = \underbrace{CatEntropy(x, x_{truth})}_{\text{Classification}} + \underbrace{\gamma_{regr} \cdot log(cosh(y - y_{truth}))}_{\text{Regression}} + \underbrace{\gamma_{quantile} \cdot [p_{0.16}(z - z_{truth}) + p_{0.84}(z - z_{truth})]}_{\text{Quantile regression (resolution estimation)}}$$





- **A word of caution…**
  - Challenging to calibrate sophisticated jet-taggers
  - Trained with simulation → learn modeling-specific details. Systematic uncertainties!
  - Explainability/interpretability: what makes this particular jet Type-X like?
- Cover such topics this week

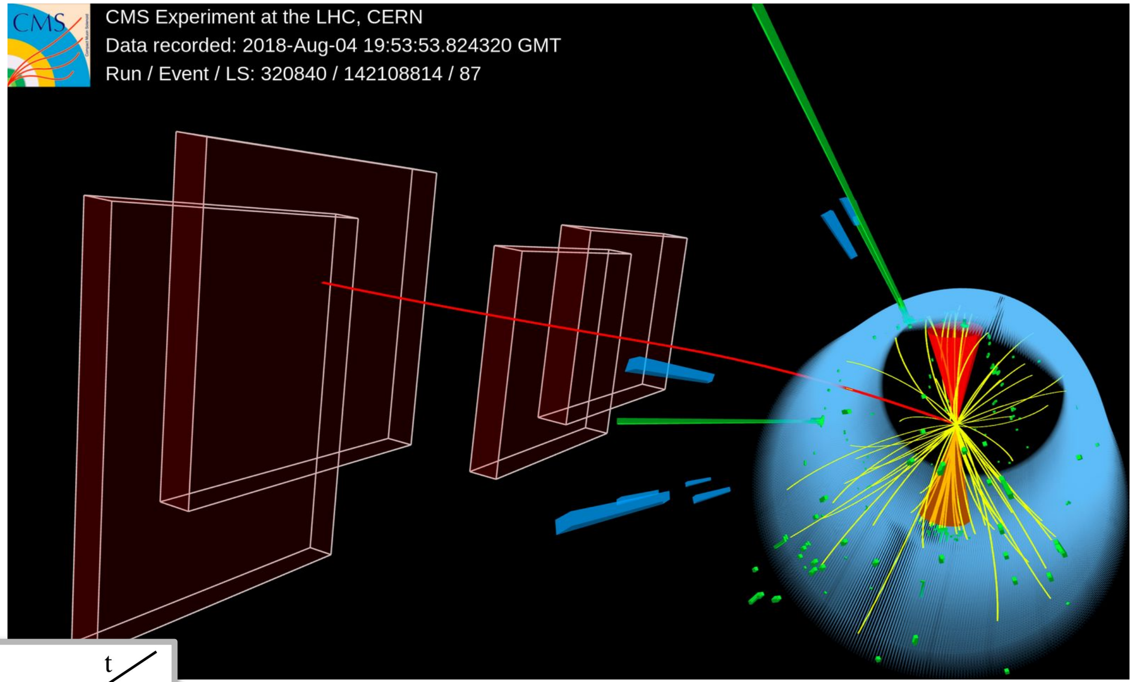


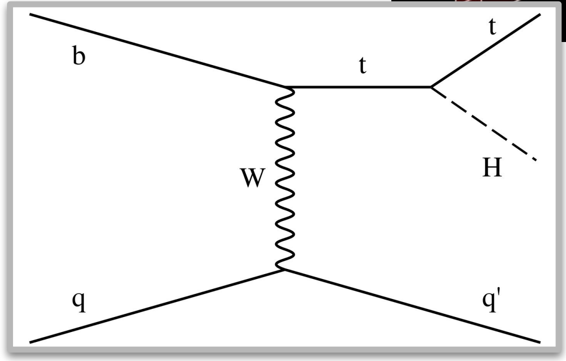
Event classification

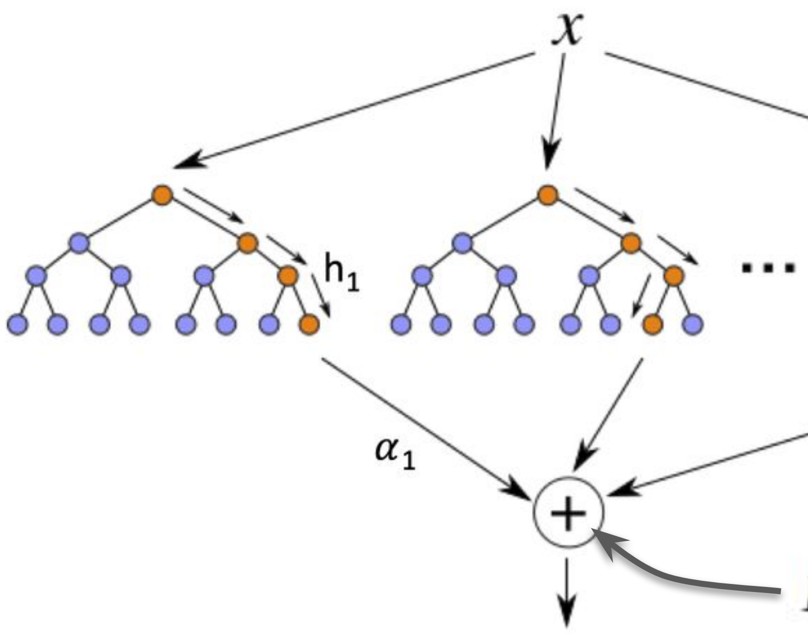
Event classification

[JHEP 07 (2021) 027]

- **Common task:** identify collisions of interest ("signal") from "background"
  - Traditionally used (sequential) selection cuts to increase signal purity
  - Now use Multivariate ML algorithms based on high-level features
  - E.g. Boosted Decision Trees (BDT), Deep Neural Network (DNN)
- Output provides powerful summary to "cut" or fit directly

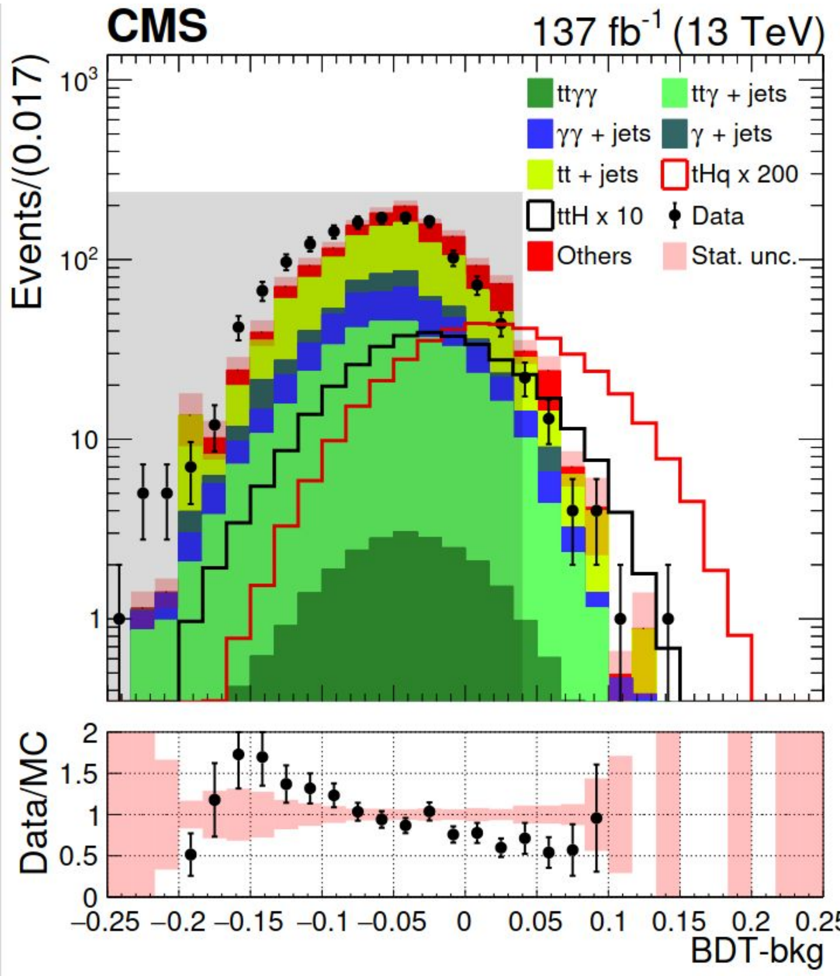






$$F(\vec{x}\,|\,\vec{\gamma},\vec{w}) = \sum_{j}^{N_{\mathrm{DT}}} \gamma_j \cdot f_j(\vec{x}\,|\,\vec{w}_j)$$

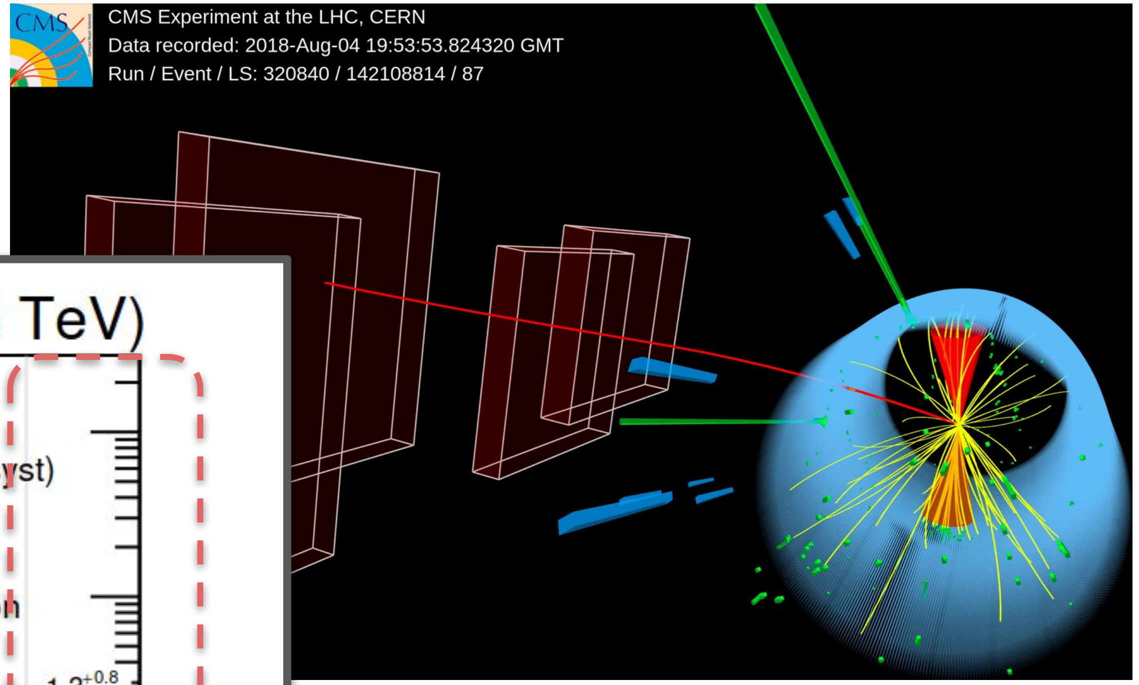
XGBoost



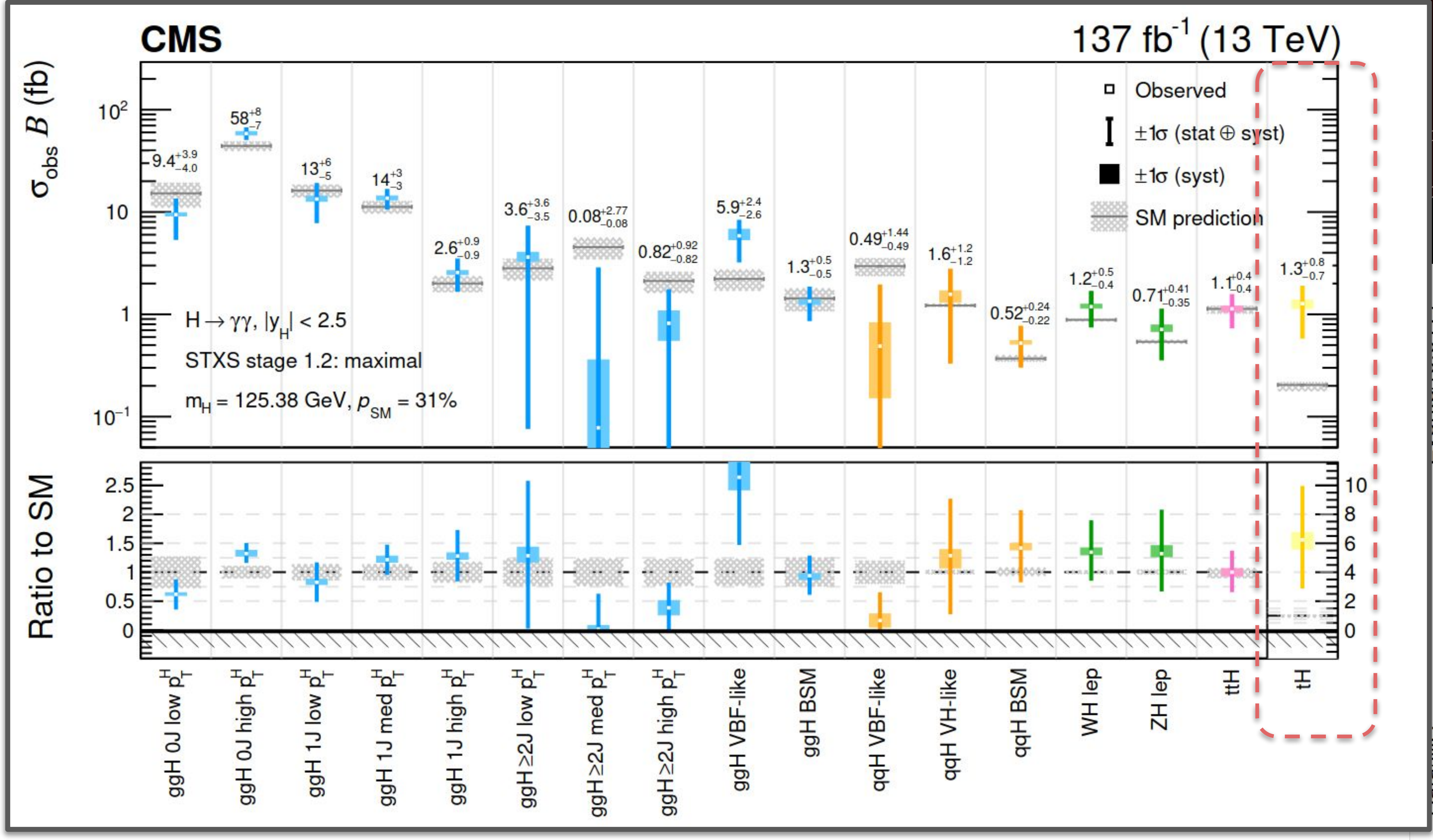
- **Tip:** XGBoost BDT typically provides most powerful, robust, calibrated classifier for "tabulated" input data

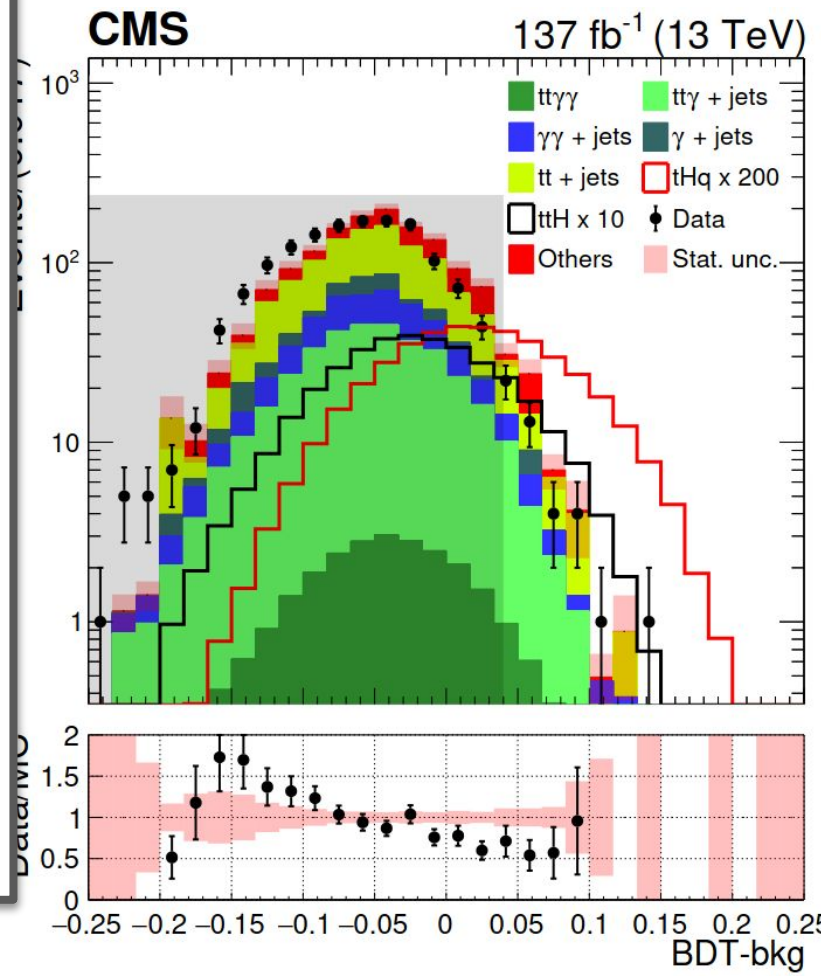
Event classification

[JHEP 07 (2021) 027]



- **Common task:** identify collisions of interest ("signal") from "background"

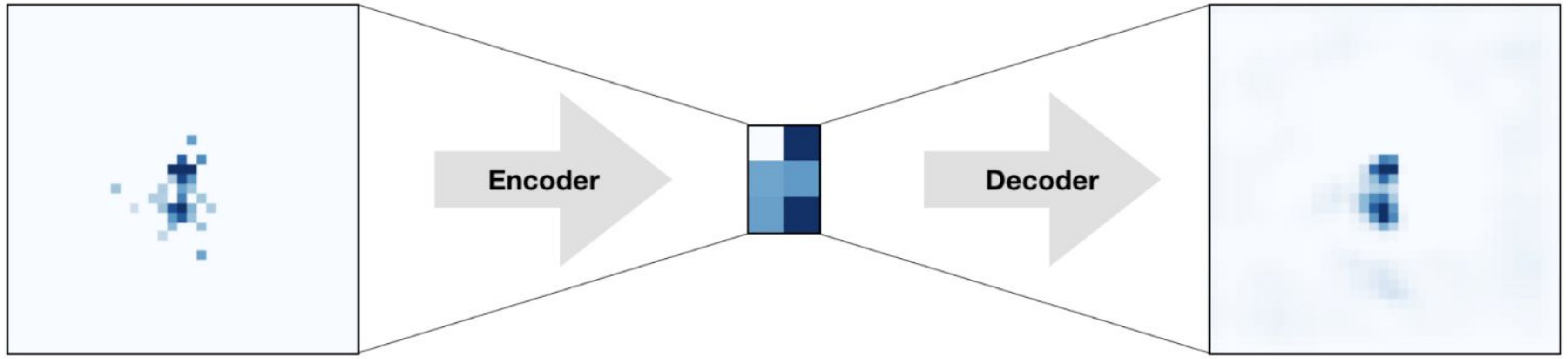




- **Tip:** XGBoost BDT typically provides most powerful, robust, calibrated classifier for "tabulated" input data

Model-agnostic searches

- What if we don't know what the signal looks like a-priori? Use **Anomaly detection algorithms**
- E.g. Unsupervised learning with (Variational) Auto-Encoders (AE)

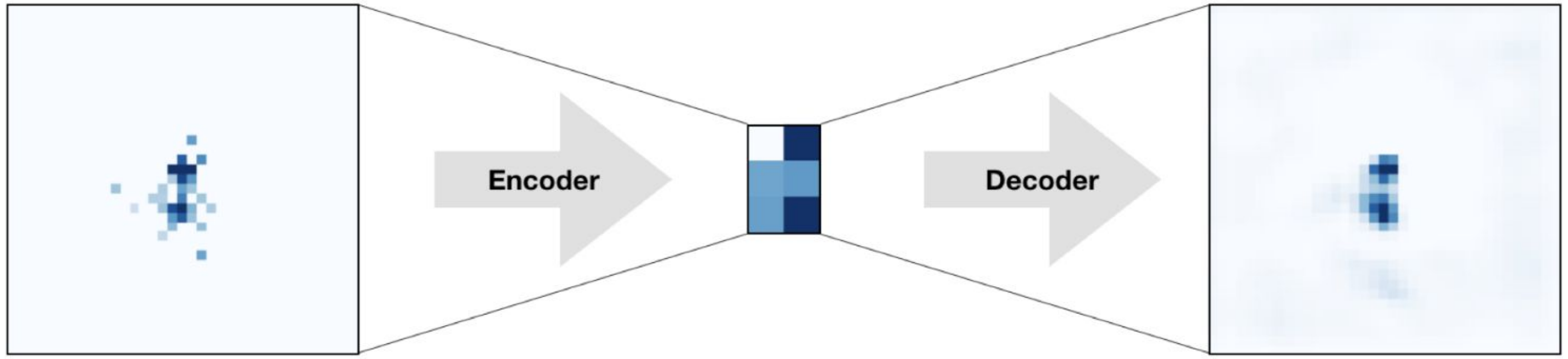


- No labels → Learn directly from data
- Anomaly metric: compare input, x, to **Decoder**(**Encoder**(x))
  - If large difference then event has low Prob(bkg)

Model-agnostic searches

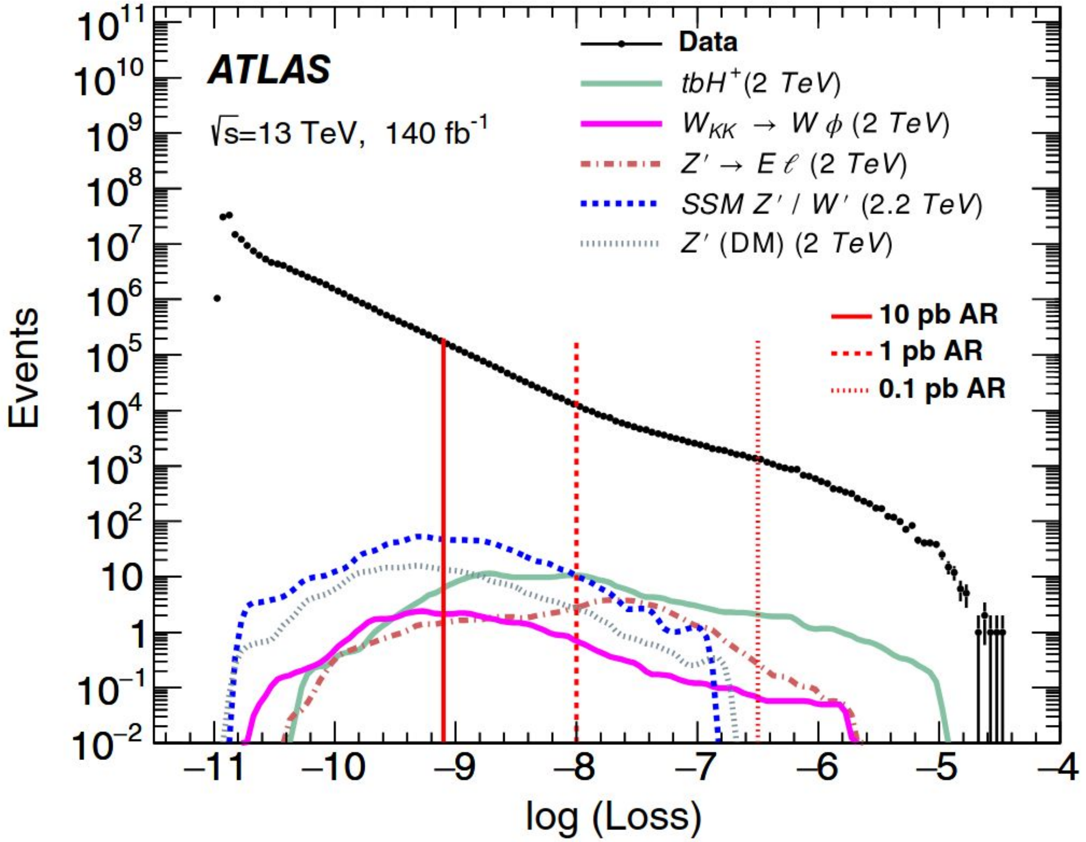
[PRL 132 (2024) 081801]

- What if we don't know what the signal looks like a-priori? Use **Anomaly detection algorithms**
- E.g. Unsupervised learning with (Variational) Auto-Encoders (AE)



$$\begin{pmatrix}
e_T^{miss} & m_T(j_1) & m_T(j_2) & \ldots m_T(j_N) & m_T(\mu_1) & m_T(\mu_2) & \ldots m_T(\mu_N) \\
h_L(j_1) & e_T(j_1) & m(j_1,j_2) & \ldots m(j_1,j_N) & m(j_1,\mu_1) & m(j_1,\mu_2) & \ldots m(j_1,\mu_N) \\
h_L(j_2) & h(j_1,j_2) & \delta e_T(j_2) & \ldots m(j_2,j_N) & m(j_2,\mu_1) & m(j_2,\mu_2) & \ldots m(j_2,\mu_N) \\
\ldots & \ldots & \ldots & \ldots, & \ldots & \ldots & \ldots \\
h_L(j_N) & h(j_1,j_N) & \ldots & \ldots \delta e_T(j_N) & m(j_N,\mu_1) & m(j_N,\mu_2) & \ldots m(j_N,\mu_N) \\
h_L(\mu_1) & h(\mu_1,j_1) & h(\mu_1,j_2) & \ldots h(\mu_1,j_N) & e_T(\mu_1) & m(\mu_1,\mu_2) & m(\mu_1,\mu_N) \\
h_L(\mu_2) & h(\mu_2,j_1) & h(\mu_1,j_2) & \ldots h(\mu_2,j_N) & h(\mu_1,\mu_2) & \delta e_T(\mu_2) & m(\mu_2,\mu_N) \\
\ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \\
h_L(\mu_N) & h(\mu_N,j_1) & h(\mu_N,j_2) & \ldots h(\mu_N,j_N) & h(\mu_N,\mu_1) & h(\mu_N,\mu_2) & \delta e_T(\mu_N)
\end{pmatrix}$$

- No labels → Learn directly from data
- Anomaly metric: compare input, x, to **Decoder**(**Encoder**(x))
  - If large difference then event has low Prob(bkg)
- ATLAS apply AE to physics-informed representation (rapidity-mass matrix)
  - For searches involving different object pairs: j+j, j+b, b+b, j+e, b+e, j+γ, j+μ, b+μ, b+γ

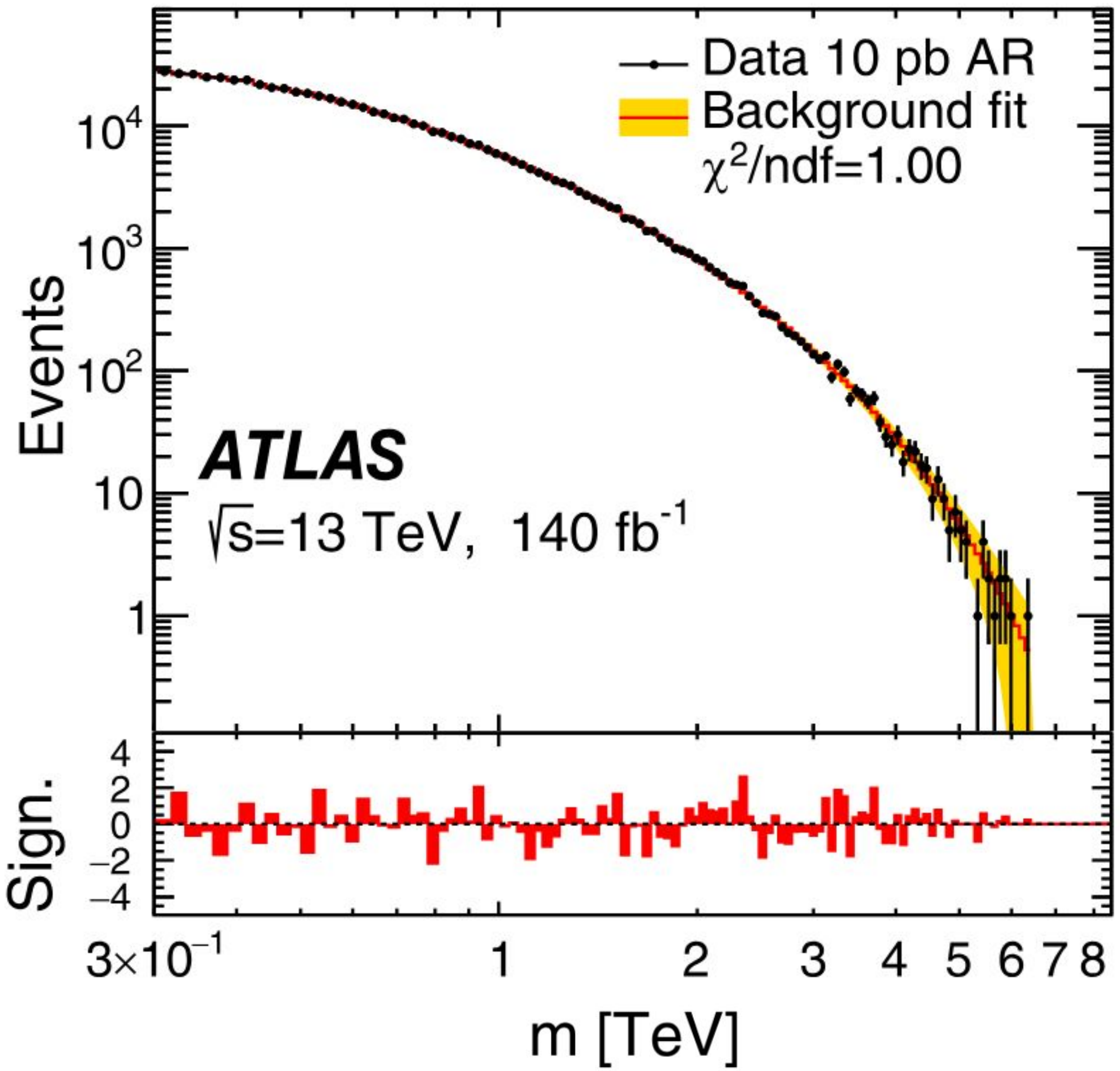


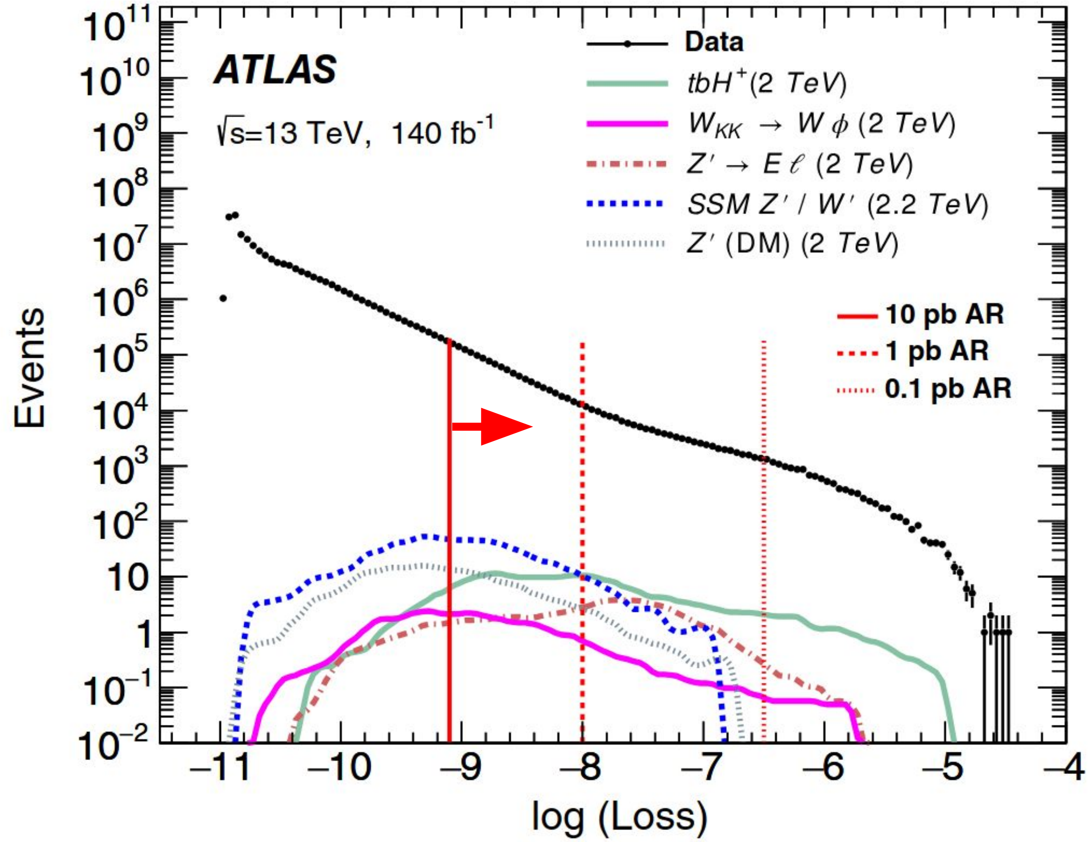
Model-agnostic searches

[PRL 132 (2024) 081801]

- What if we don't know what the signal looks like a-priori? Use **Anomaly detection algorithms**
- E.g. Unsupervised learning with (Variational) Auto-Encoders (AE)

$$\begin{pmatrix}
e_T^{miss} & m_T(j_1) & m_T(j_2) & \ldots m_T(j_N) & m_T(\mu_1) & m_T(\mu_2) & \ldots m_T(\mu_N) \\
h_L(j_1) & e_T(j_1) & m(j_1,j_2) & \ldots m(j_1,j_N) & m(j_1,\mu_1) & m(j_1,\mu_2) & \ldots m(j_1,\mu_N) \\
h_L(j_2) & h(j_1,j_2) & \delta e_T(j_2) & \ldots m(j_2,j_N) & m(j_2,\mu_1) & m(j_2,\mu_2) & \ldots m(j_2,\mu_N) \\
\ldots & \ldots & \ldots & \ldots, & \ldots & \ldots & \ldots \\
h_L(j_N) & h(j_1,j_N) & \ldots & \ldots \delta e_T(j_N) & m(j_N,\mu_1) & m(j_N,\mu_2) & \ldots m(j_N,\mu_N) \\
h_L(\mu_1) & h(\mu_1,j_1) & h(\mu_1,j_2) & \ldots h(\mu_1,j_N) & e_T(\mu_1) & m(\mu_1,\mu_2) & m(\mu_1,\mu_N) \\
h_L(\mu_2) & h(\mu_2,j_1) & h(\mu_1,j_2) & \ldots h(\mu_2,j_N) & h(\mu_1,\mu_2) & \delta e_T(\mu_2) & m(\mu_2,\mu_N) \\
\ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \\
h_L(\mu_N) & h(\mu_N,j_1) & h(\mu_N,j_2) & \ldots h(\mu_N,j_N) & h(\mu_N,\mu_1) & h(\mu_N,\mu_2) & \delta e_T(\mu_N)
\end{pmatrix}$$

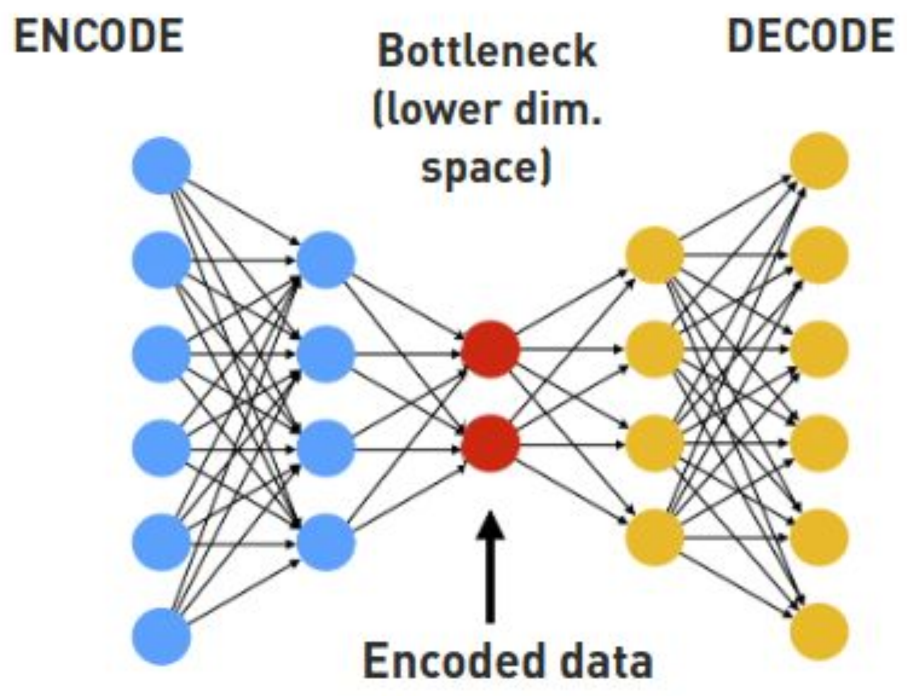




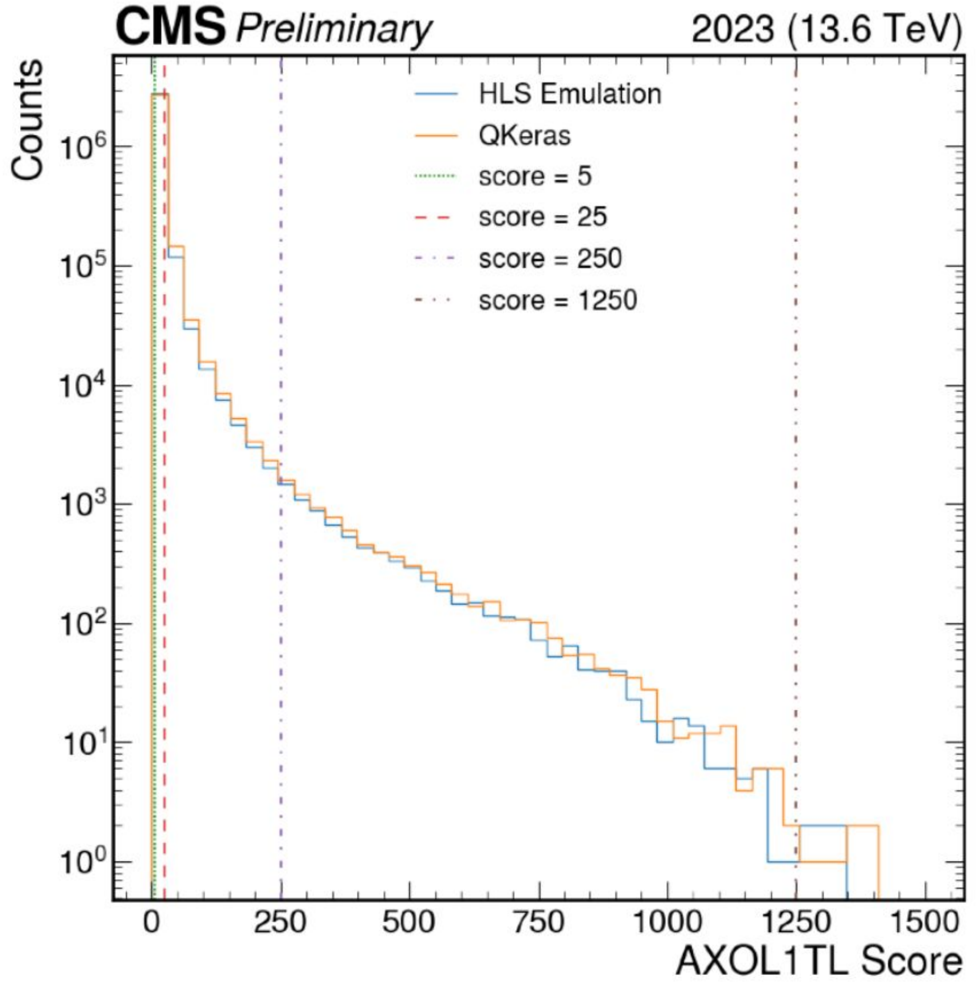
Anomaly detection in real-time

[CMS-DP-2023-079]

- What if we don't know what the signal looks like a-priori?
    - If we don't consider this in the trigger (online filter), **we lose data before we even begin**
    - Apply anomaly detection algorithms online e.g. AXOL1TL

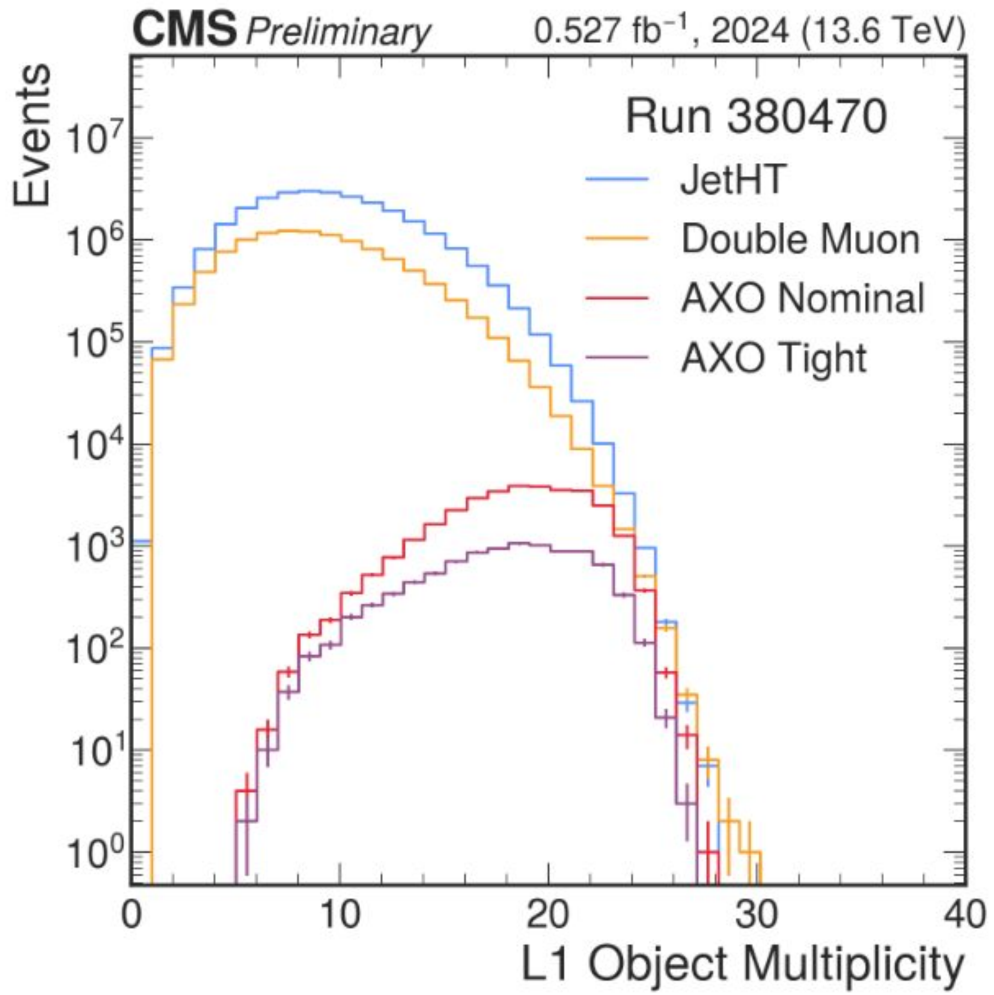


Variational Auto-Encoder
Trained to compress and reconstruct collision data



Trigger on events with high anomaly score

Selects unique events, preference for high multiplicity

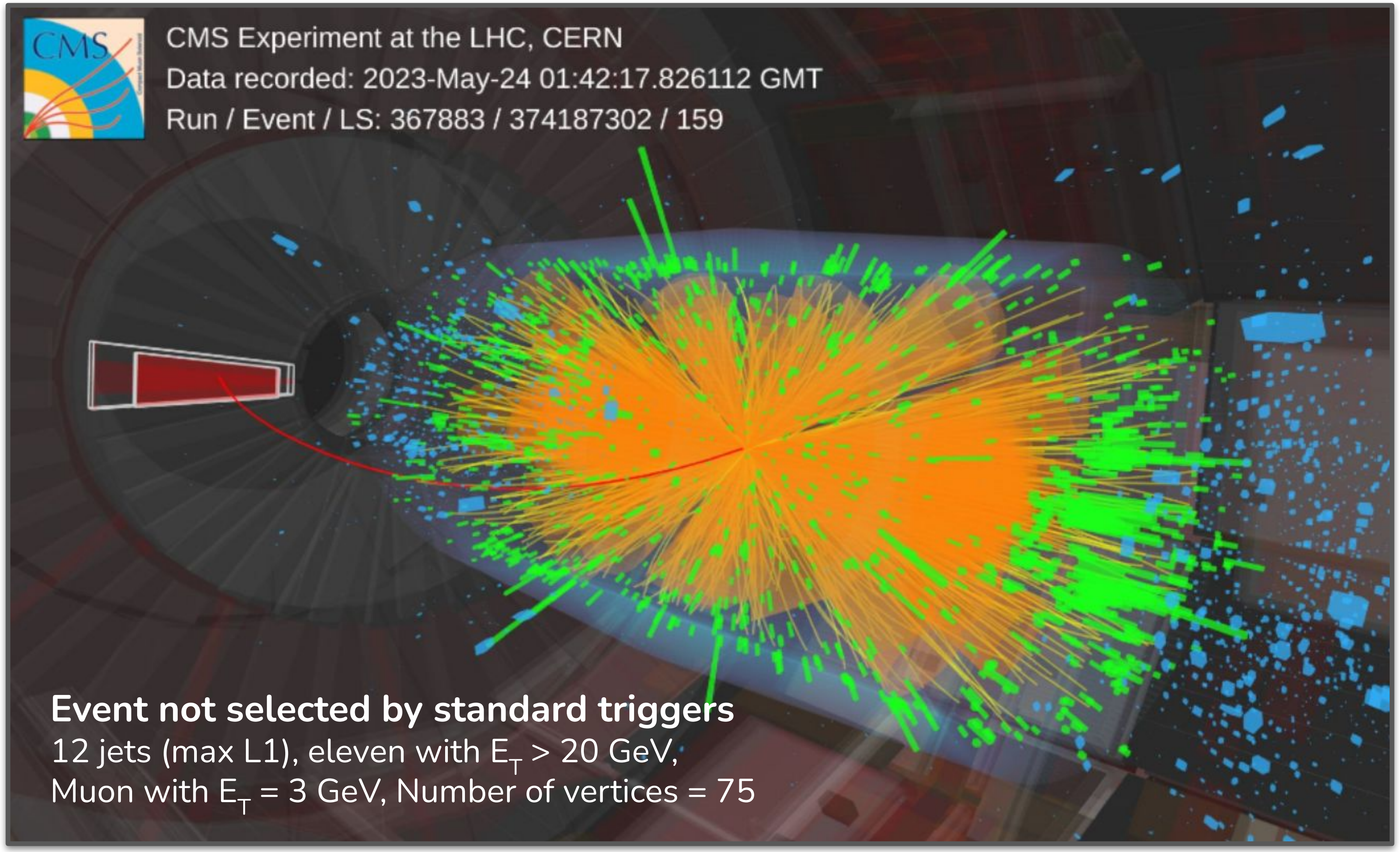


Anomaly detection in real-time

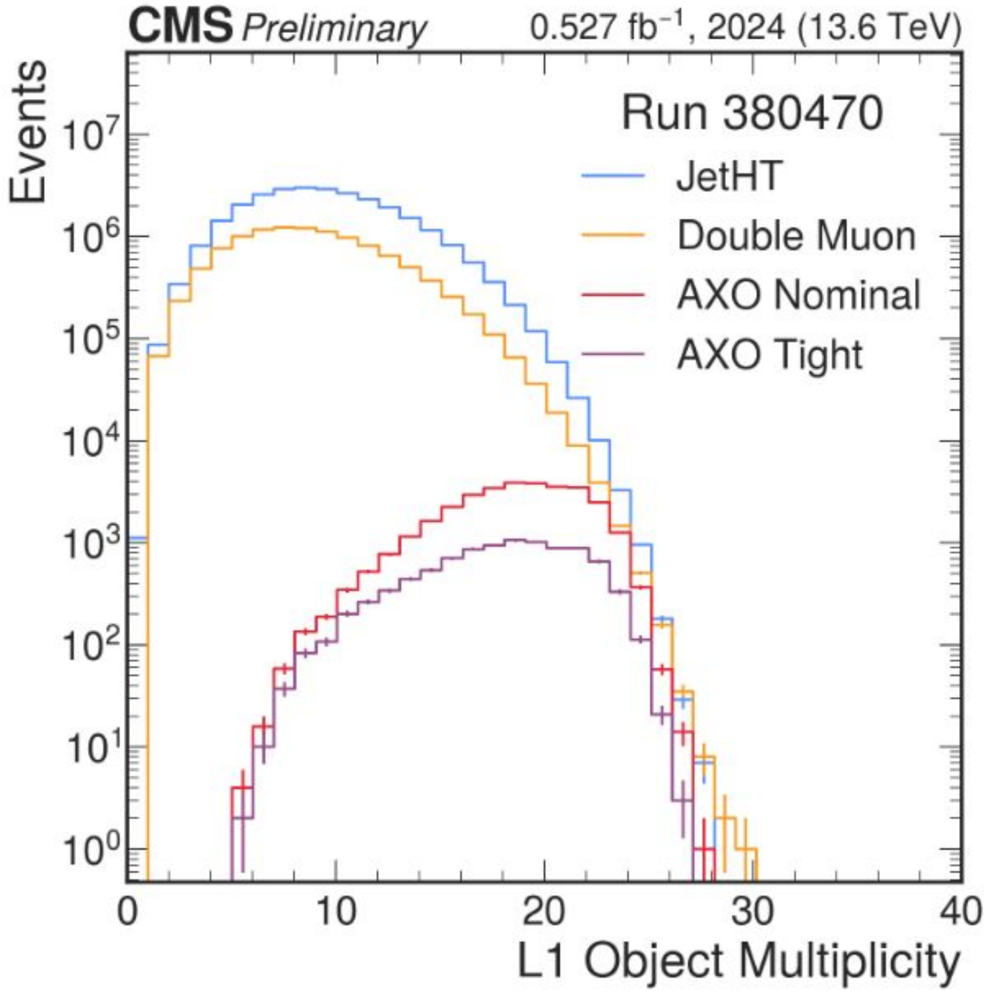
AXOL1TL

[CMS-DP-2023-079]

- What if we don't know what the signal looks like a-priori?
  - If we don't consider this in the trigger (online filter), **we lose data before we even begin**
  - Apply anomaly detection algorithms online e.g. AXOL1TL



Selects unique events,
preference for high multiplicity

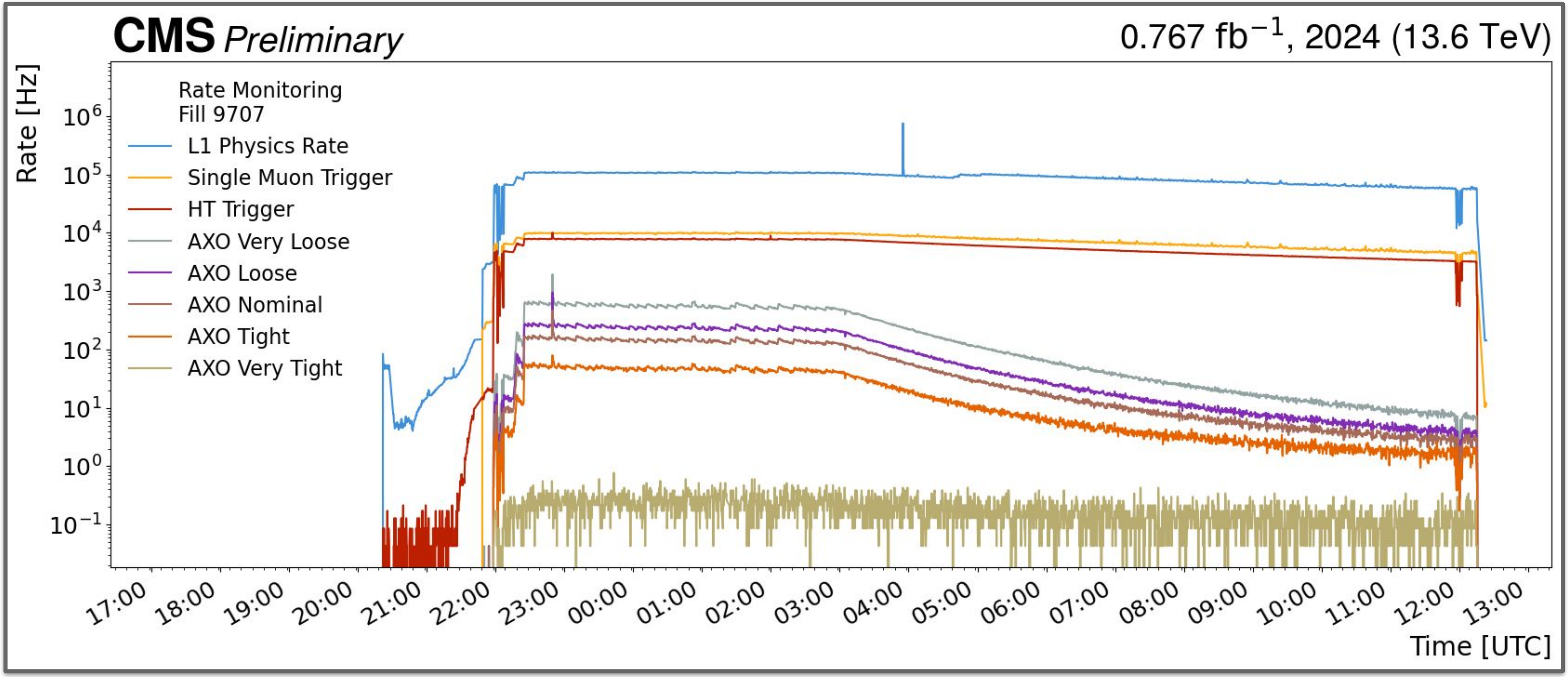


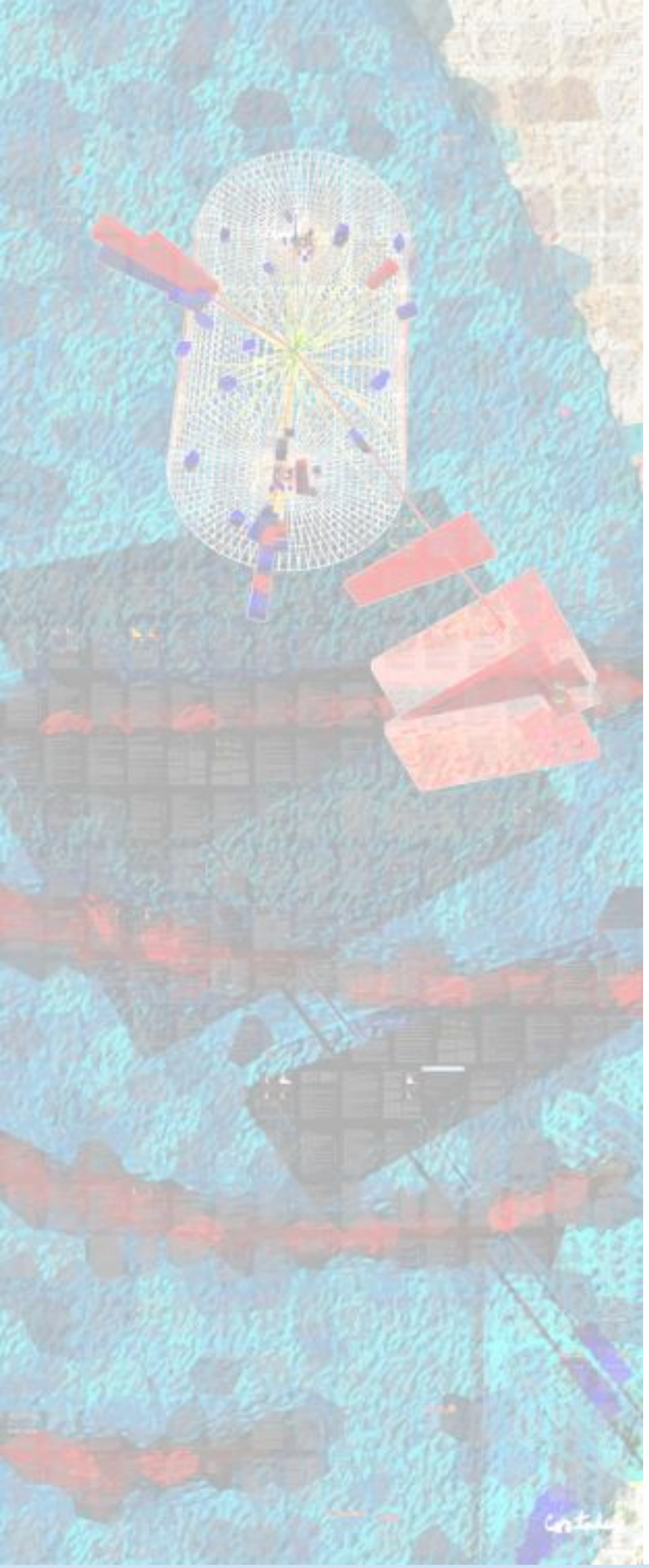
Anomaly detection in real-time

AXOL1TL

[CMS-DP-2023-079]

- Demonstrated successful running in L1T (2024)

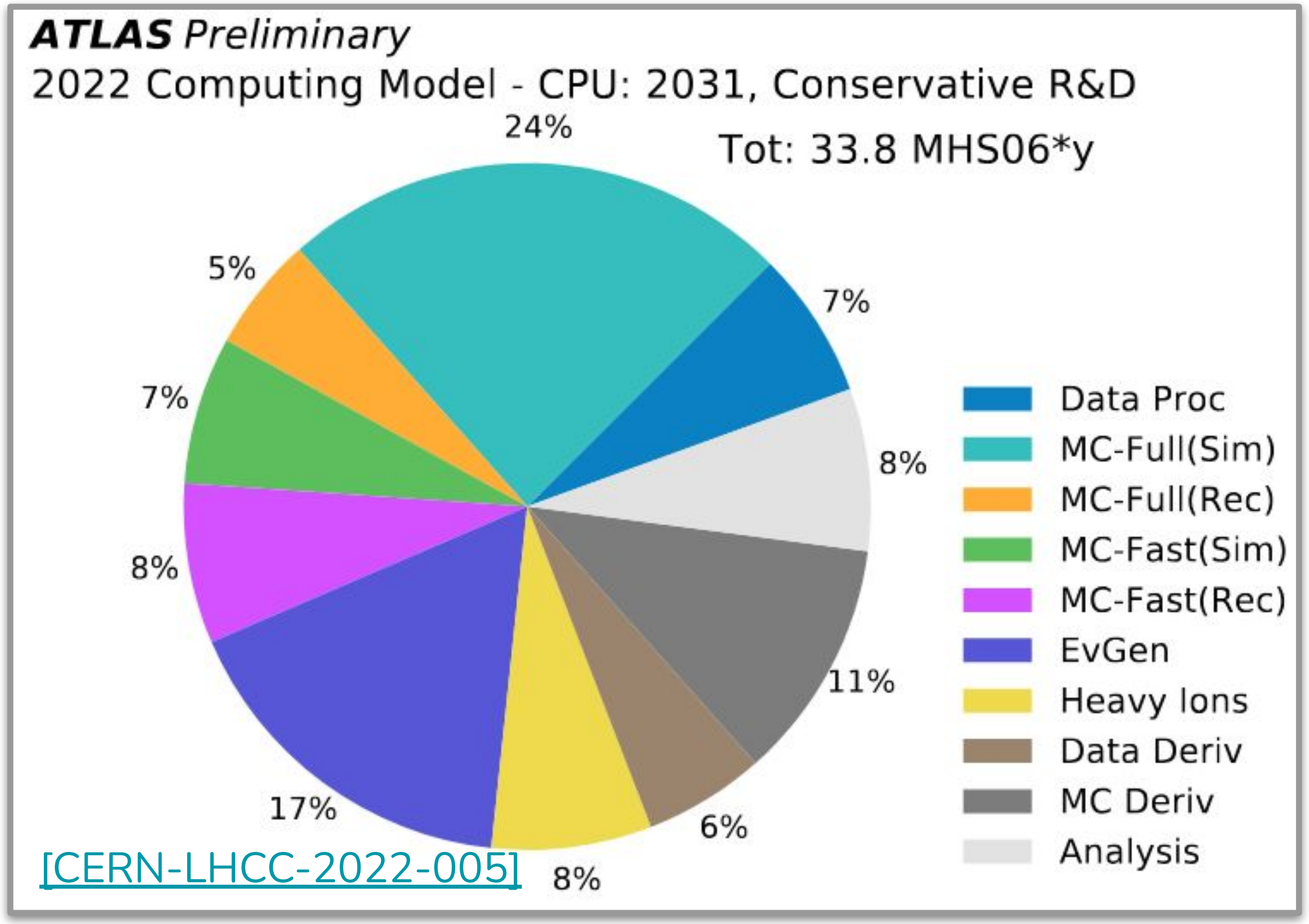


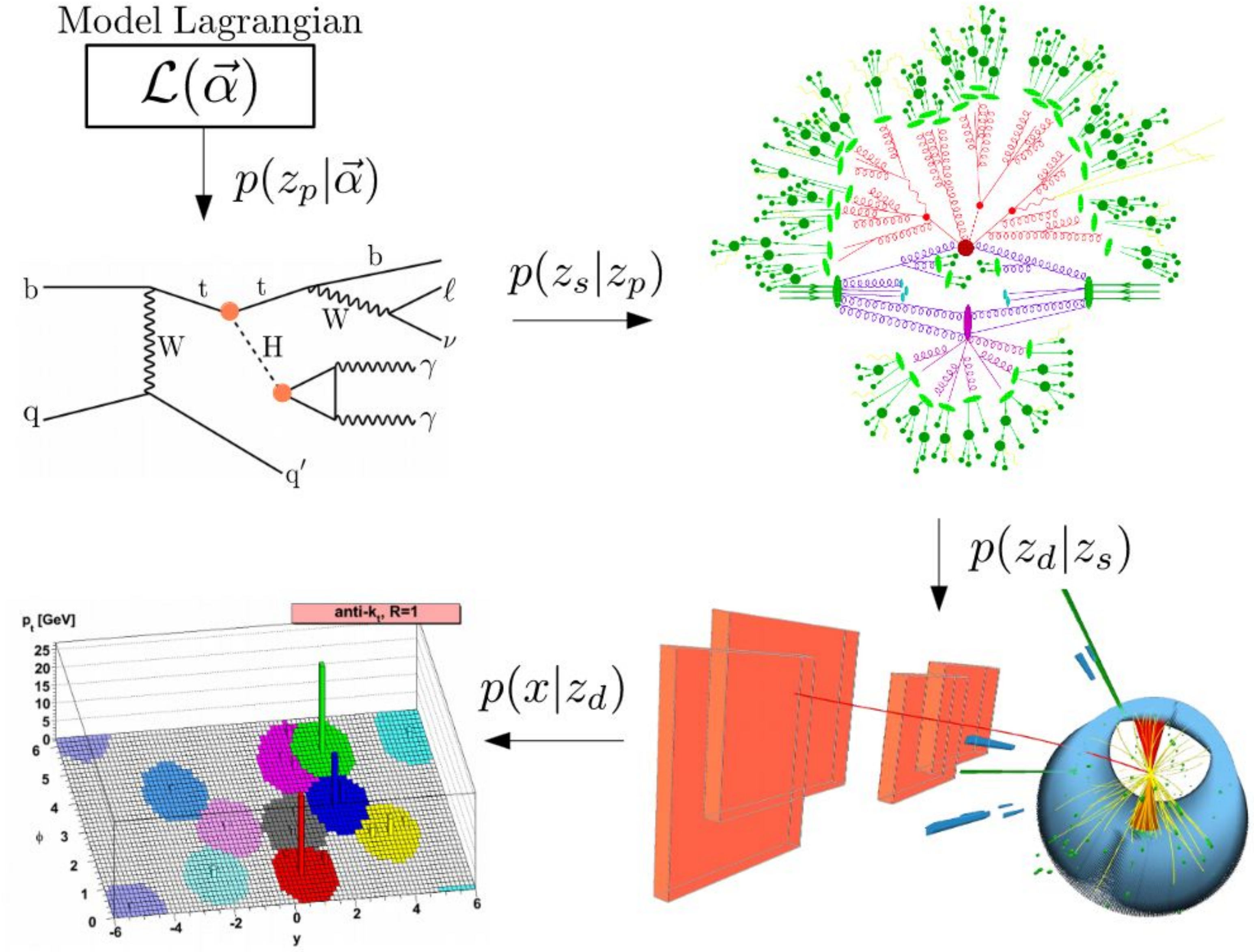


Simulation
(generative)

Simulation is painful!

- MC simulation is extremely CPU-intensive



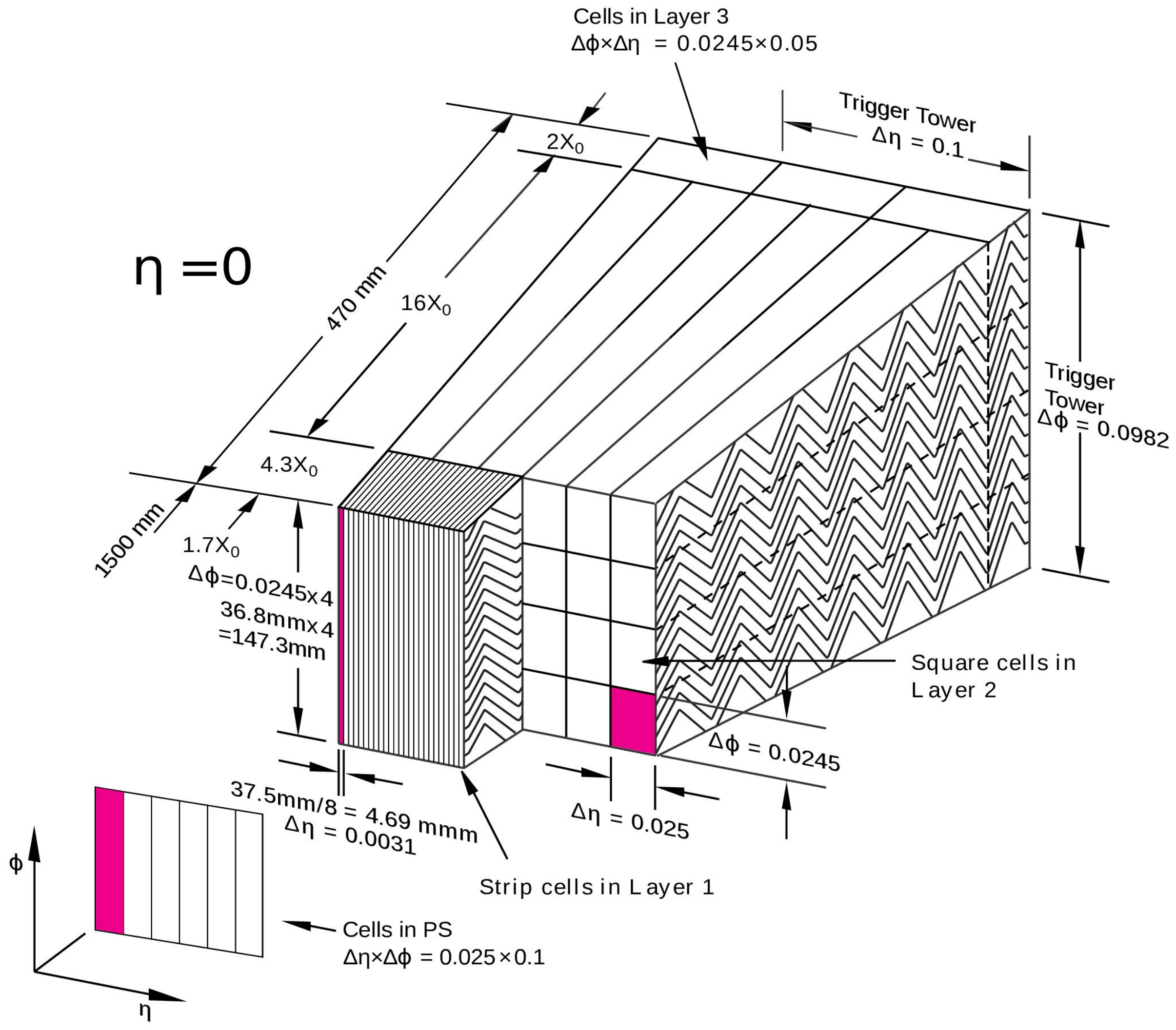


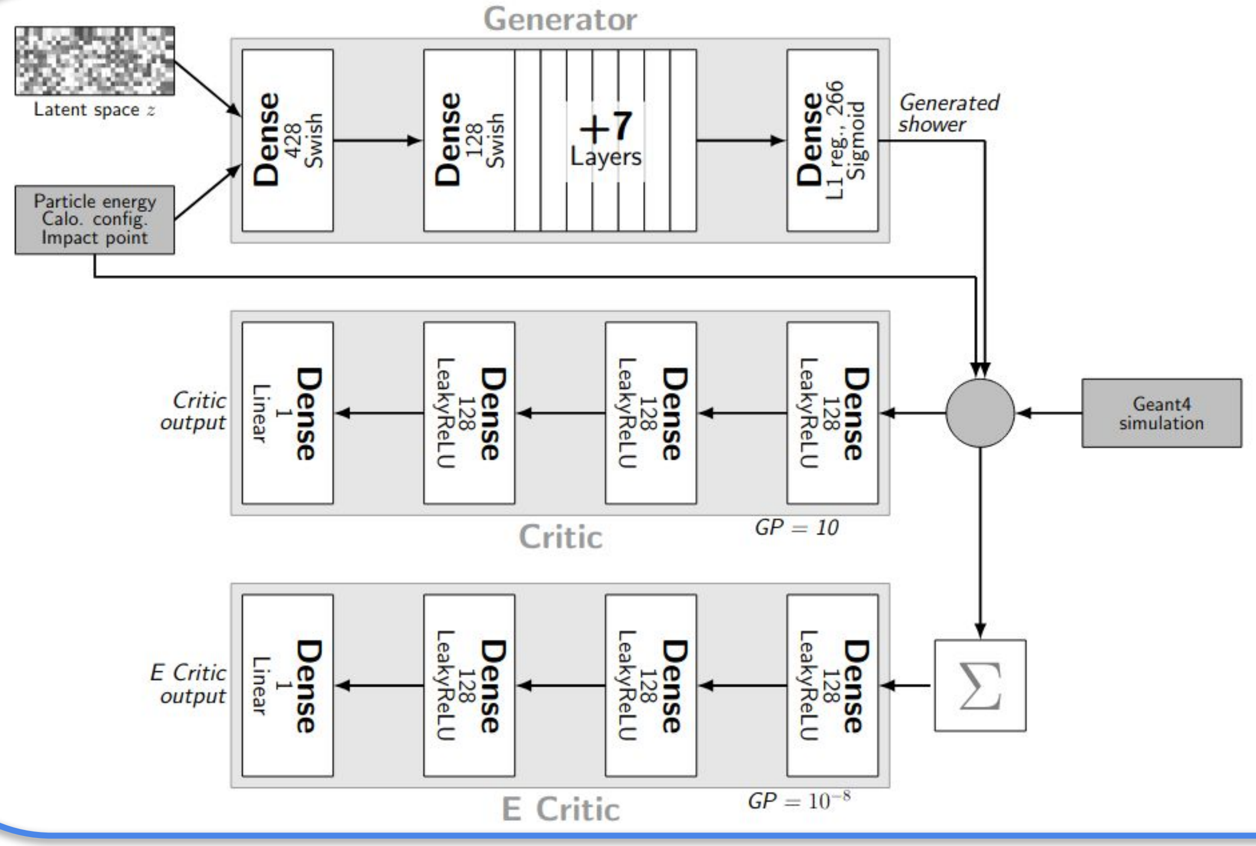
- Can we use ML to short-cut parts of the simulation chain?

Faster simulation

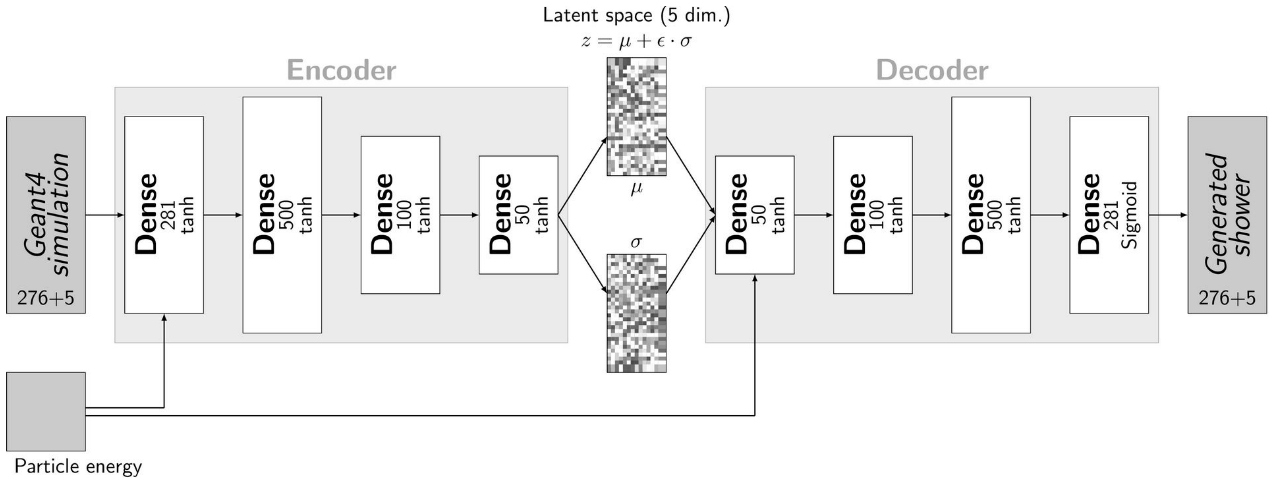
[CSBS 8, 7 (2024)]

- Deep **generative models** for fast photon shower simulation in ATLAS calorimeter to replace (slow) Geant4





Generative Adversarial Network (GAN)

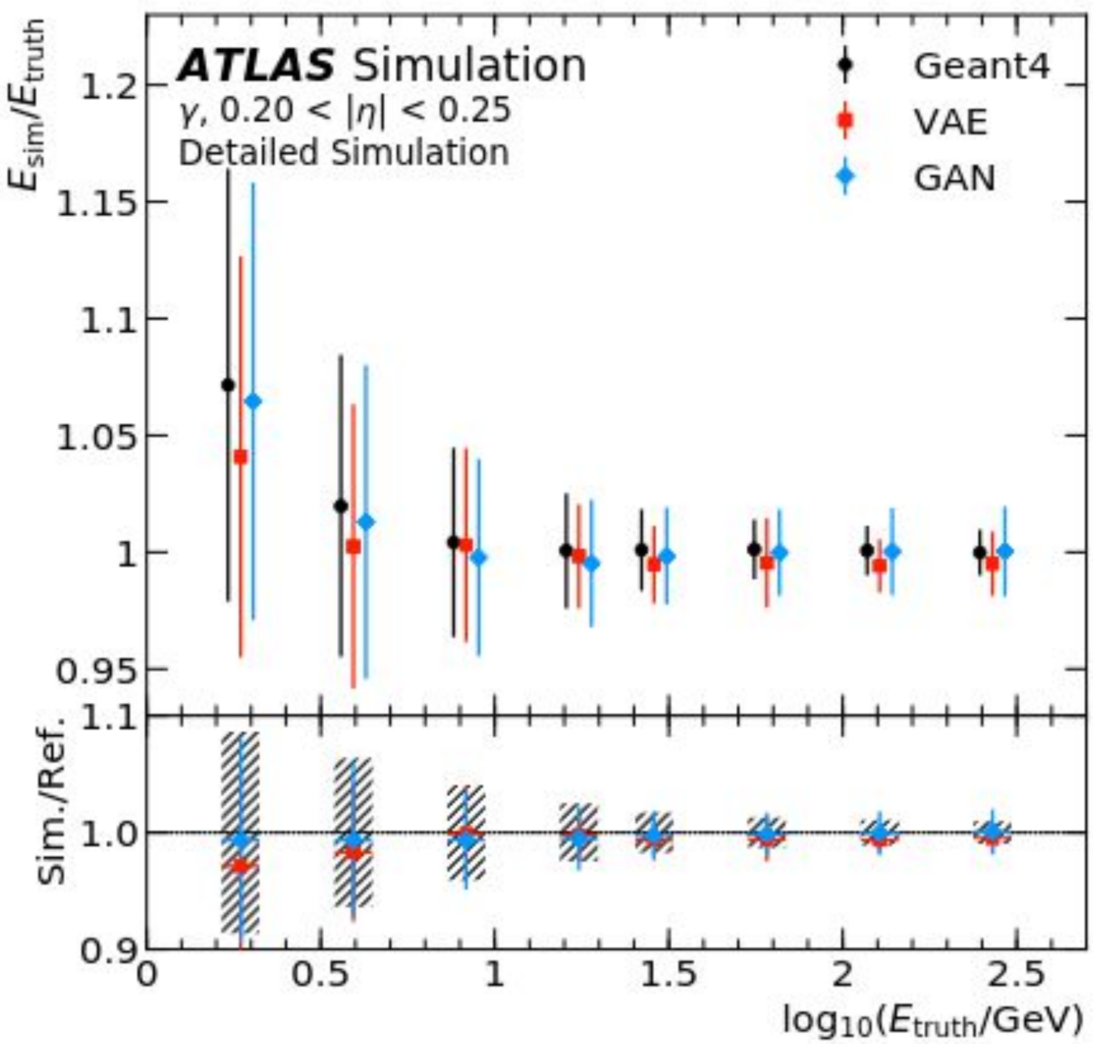


Variational Auto-Encoder (VAE)

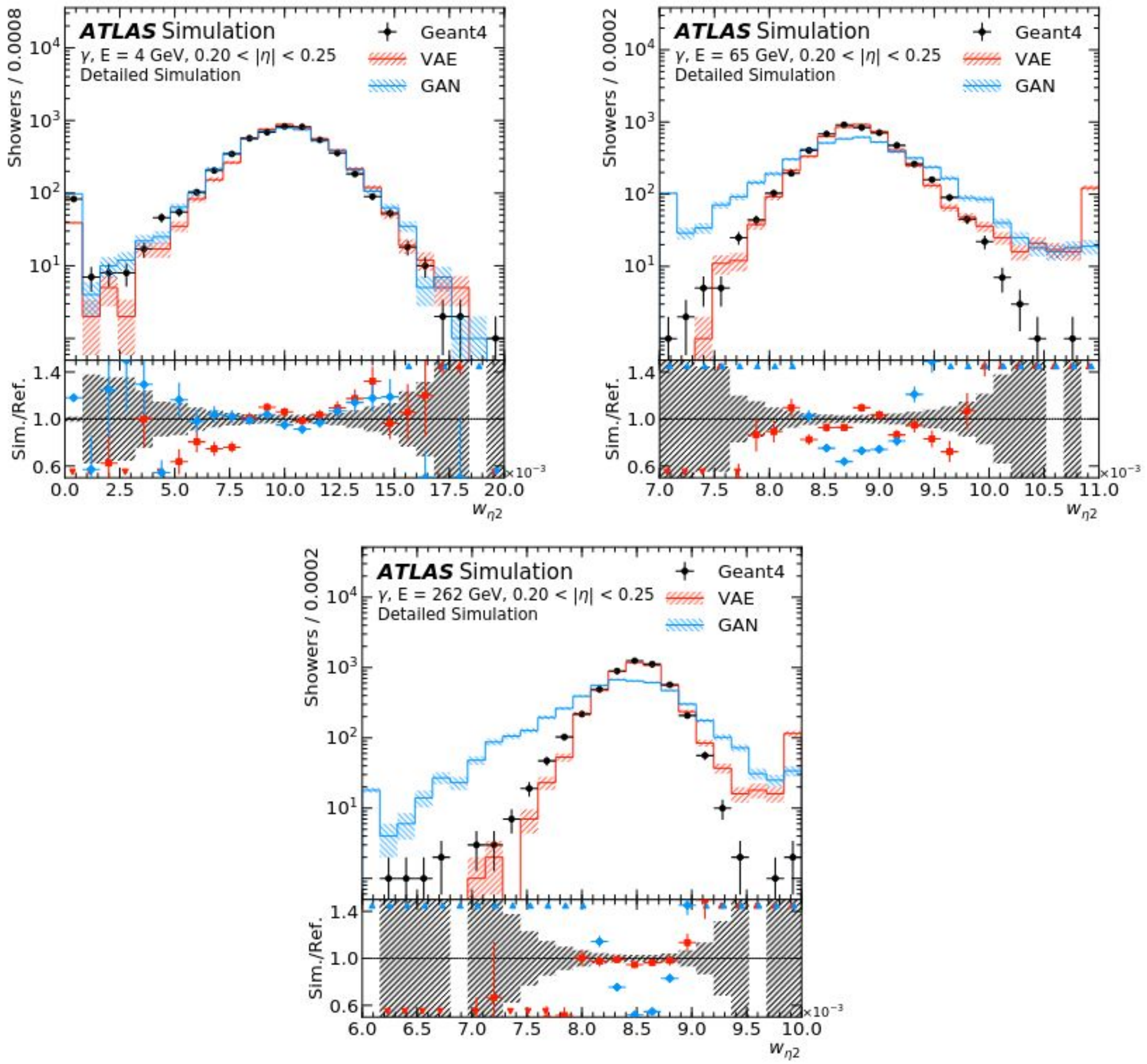
Faster simulation

[CSBS 8, 7 (2024)]

- Deep **generative models** for fast photon shower simulation in ATLAS calorimeter to replace (slow) Geant4
  - Generation time reduced by up to two orders of magnitude, very small memory footprint (5 Mb)



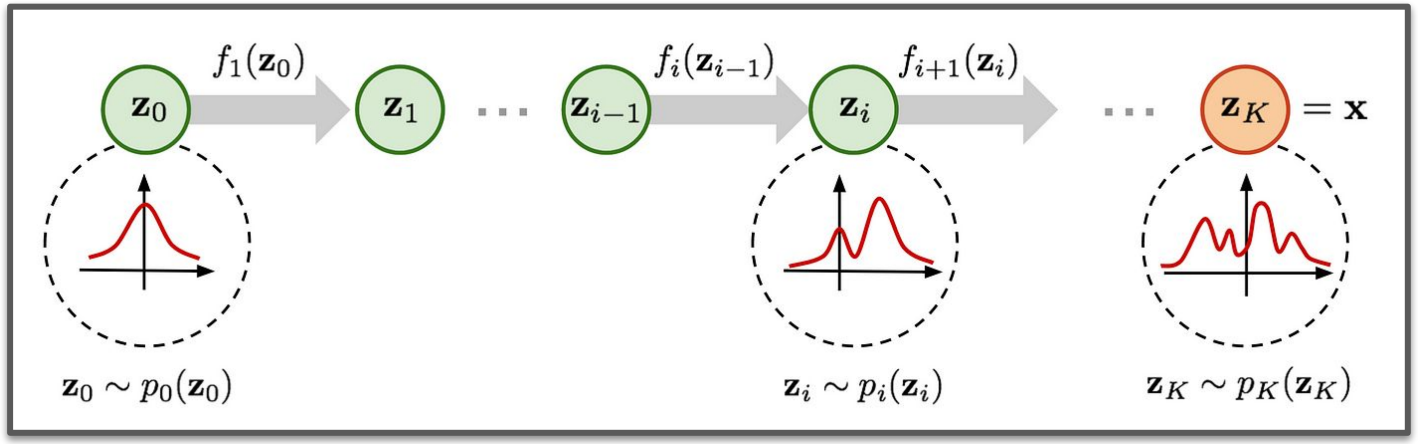
Total energy of shower: response & resolution
Decent agreement, slightly better for GANs

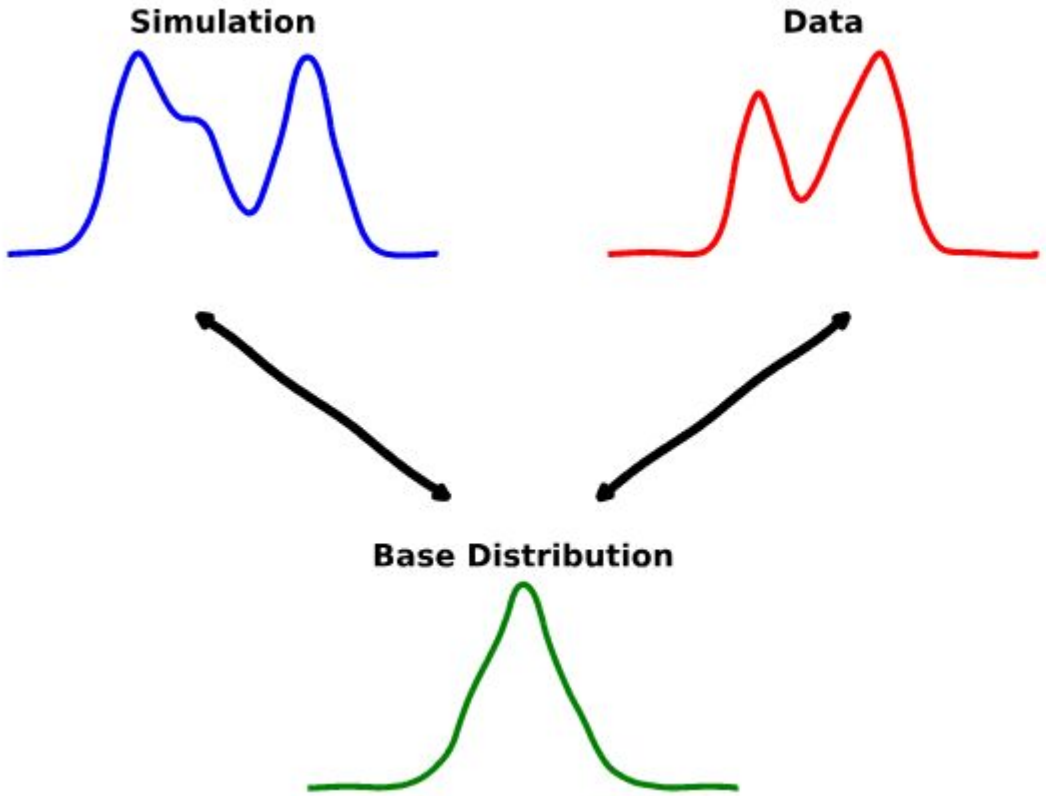


Shower shape variables (lateral shower width)
Room for improvement, VAE outperforms particularly for high pT photons

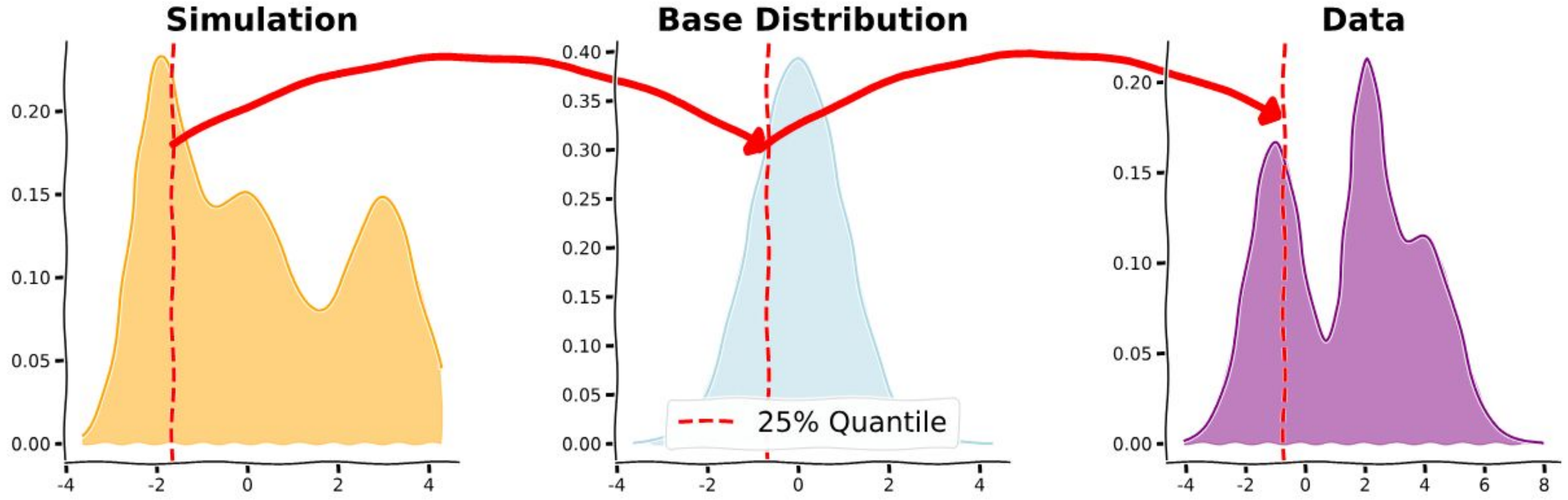
Better simulation

- Better our simulation reflects real data → more accurate inference (i.e. less bias, reduced systematic uncertainty)
  - Calibration/refinement is a crucial part of any particle physics analysis: traditionally use binned scale factor approach
  - ML approaches promise high-dimensional, unbinned calibration
- Example: “One Flow to correct them all” [arXiv: 2403.18582]





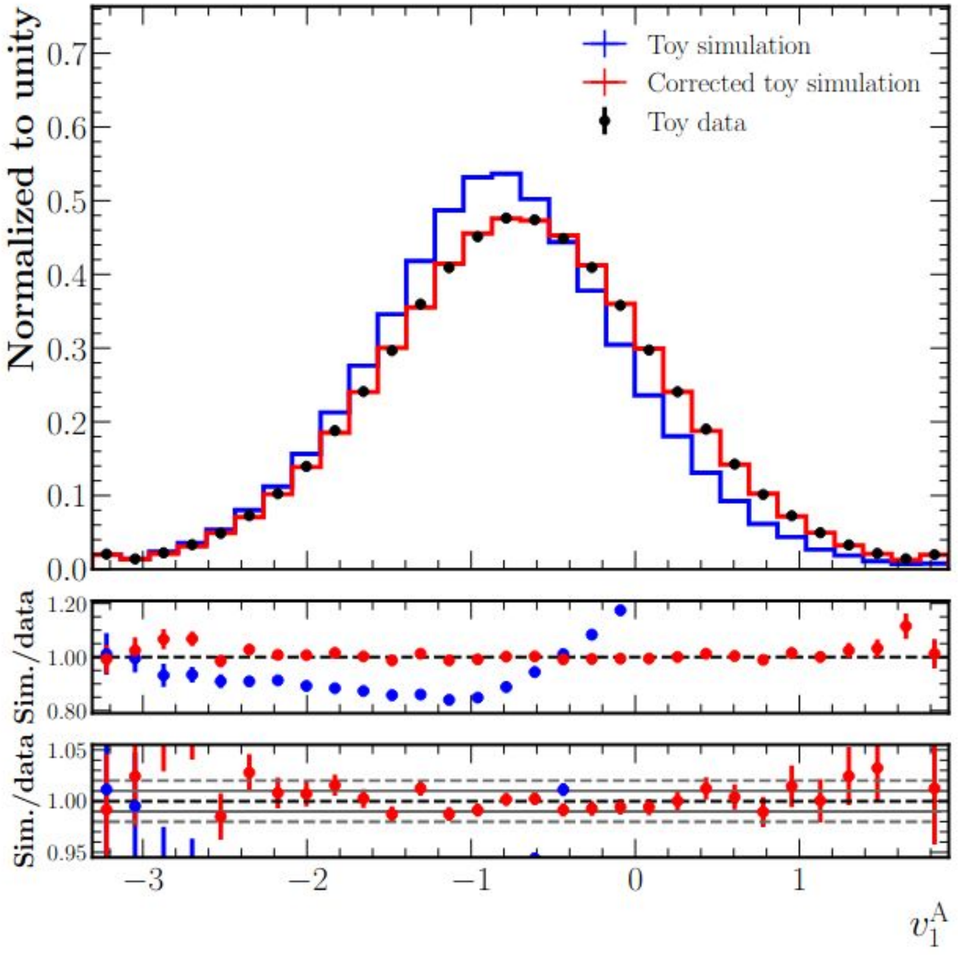
Normalising flow architecture
Map both simulation and data to share distribution, conditioned on boolean

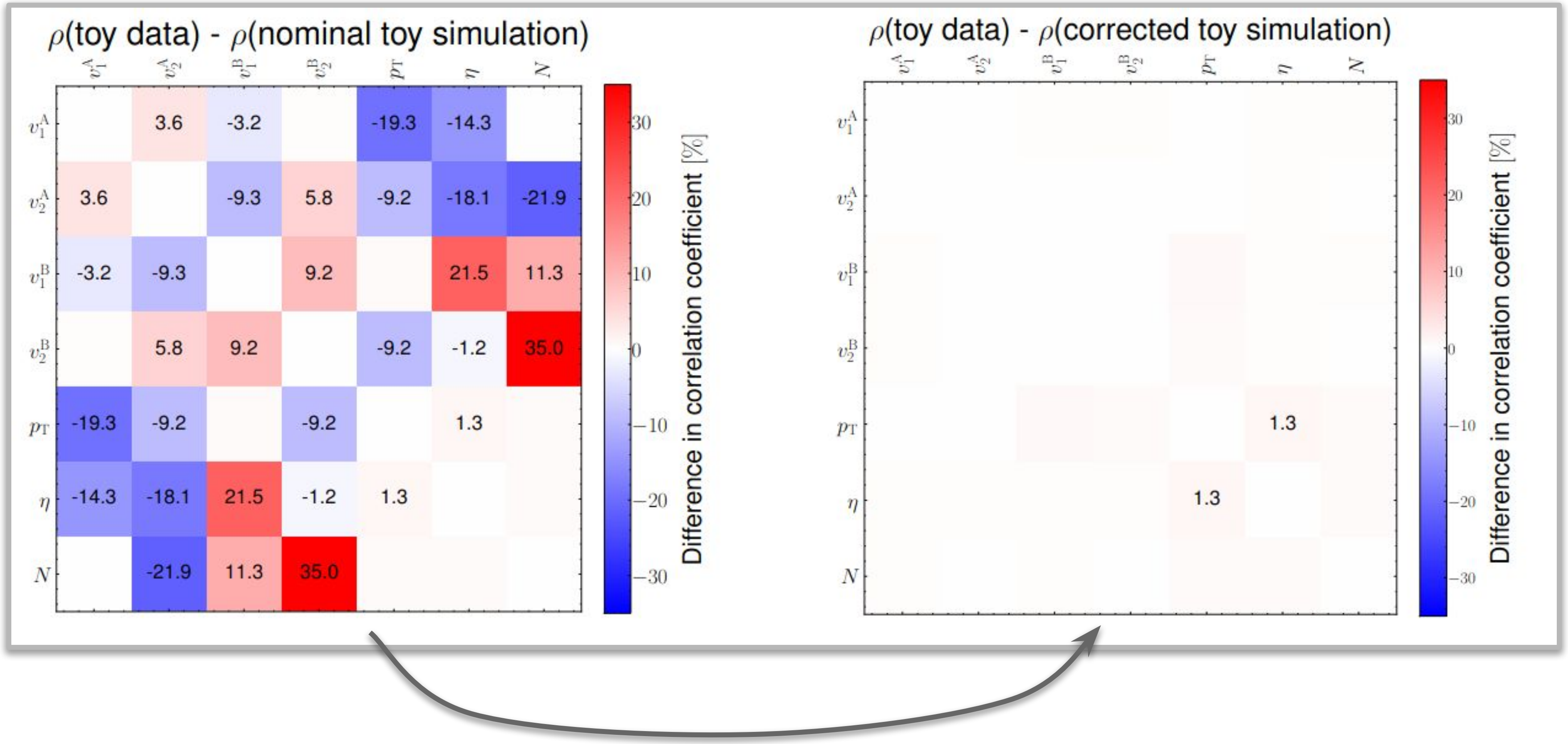


Morph simulation to data
Flip boolean switch, quantiles are preserved

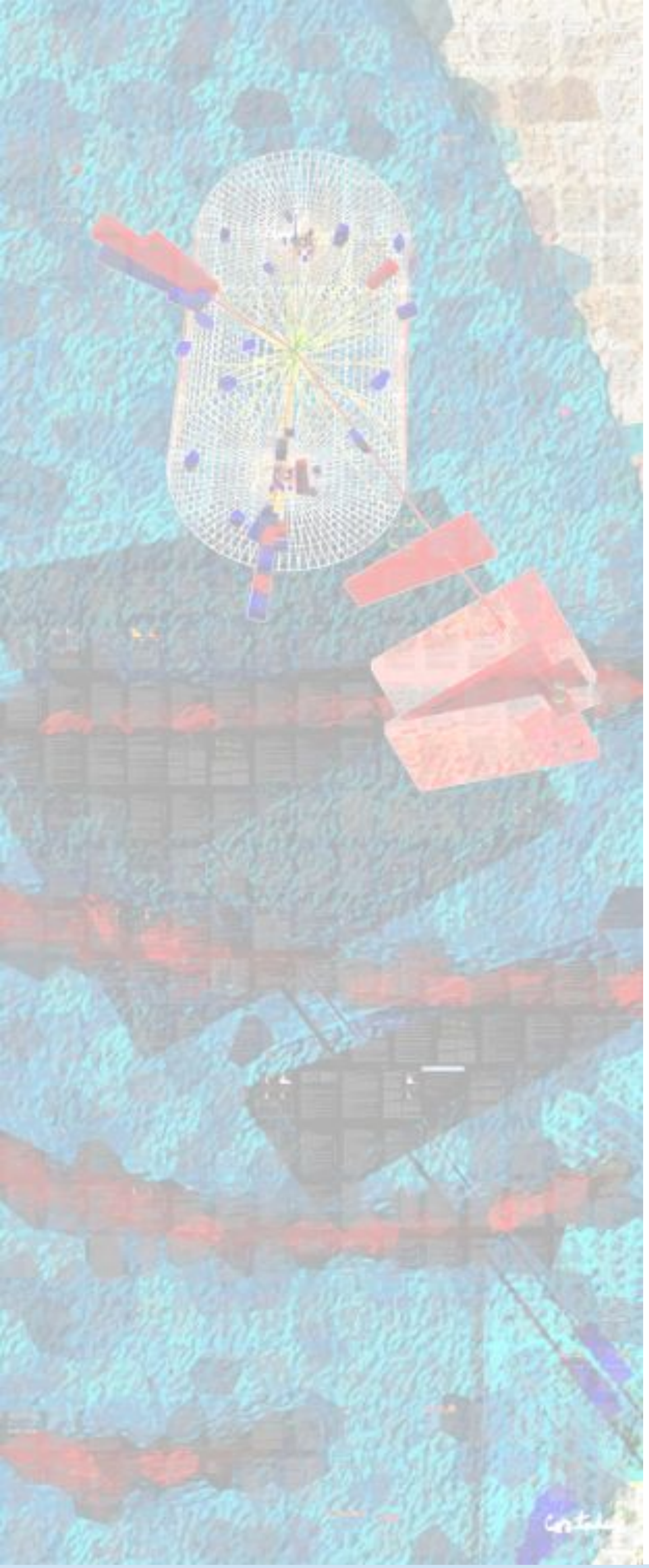
Better simulation

- Better our simulation reflects real data → more accurate inference (i.e. less bias, reduced systematic uncertainty)
  - Calibration/refinement is a crucial part of any particle physics analysis: traditionally use binned scale factor approach
  - ML approaches promise high-dimensional, unbinned calibration
- Example: “One Flow to correct them all” [arXiv: 2403.18582]





Apply morphing with flow

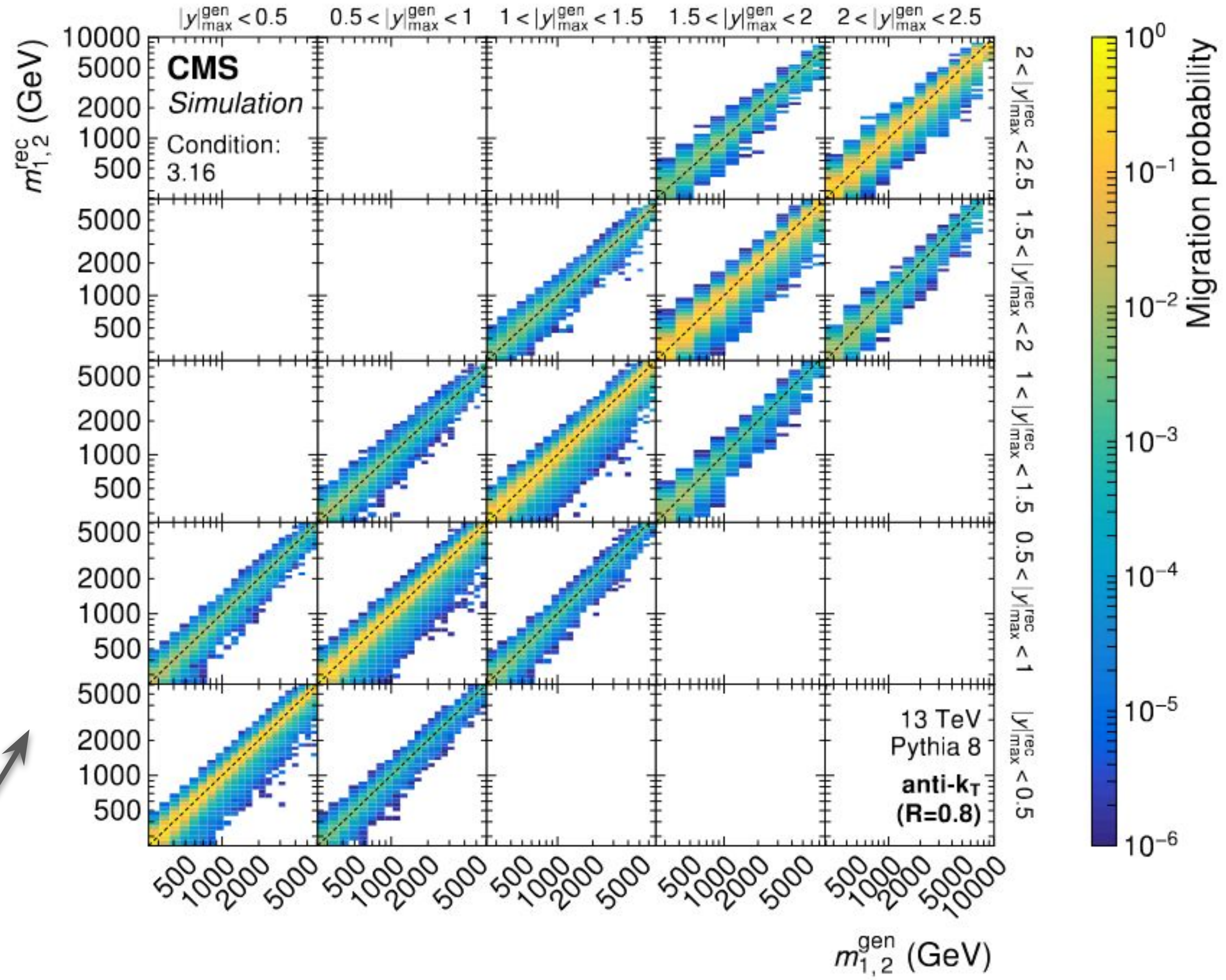


Inference

Unfolding

[Phystat Conference on Unfolding 2024]

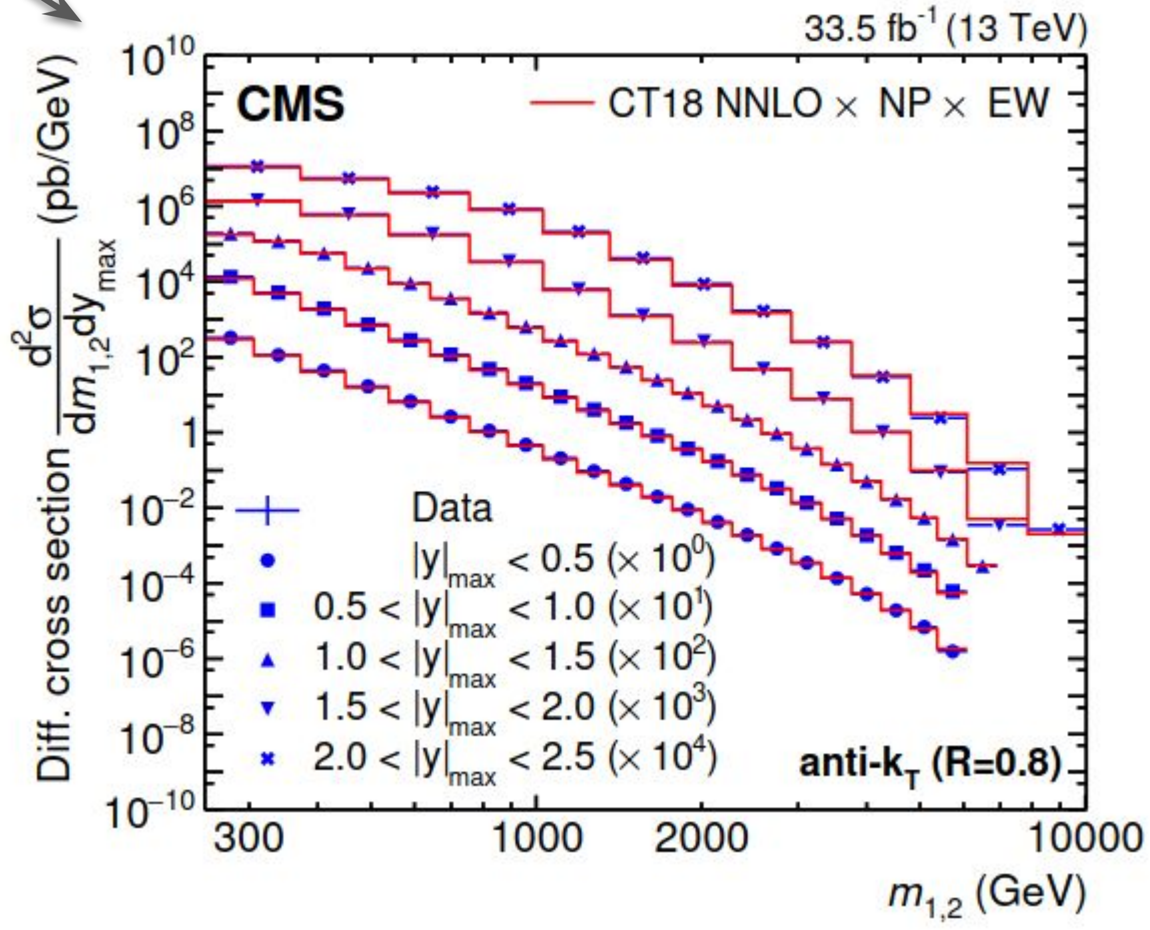
- **Unfolding:** reconstruct "true" distribution of a physical quantity from measured (i.e. smeared) data
  - Limited to small number of observables and present as differential cross section in predetermined bins



Response matrix (R): extract from MC simulation:
Fraction of "truth" bin i lands in reco bin j

Unfold by solving:
(Perhaps with regularisation)

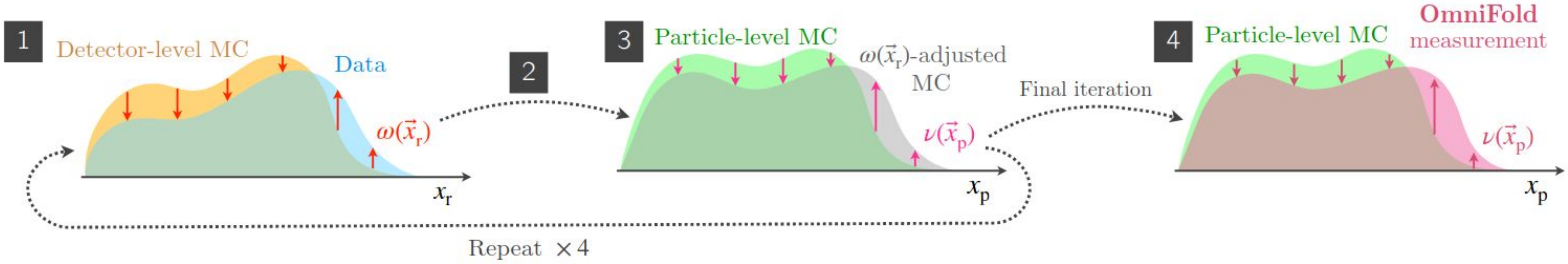
$$N_{\text{data}} = R\, N_{\text{unfold}} + N_{\text{bkg}}$$



Unfolding with omnifold

[arXiv:1911.09107]

- **Unfolding:** reconstruct "true" distribution of a physical quantity from measured (i.e. smeared) data
  - Limited to small number of observables and present as differential cross section in predetermined bins
- **OMNIFOLD:** result provided (unbinned) as dataset of particle-level events



N-dimension event observables

$$\mathcal{L}[f(\vec{x})] = -\sum_{i\in A} w_i \log(f(\vec{x}_i)) - \sum_{i\in B} w_i \log(1 - f(\vec{x}_i))$$

Output f(x) used to extract likelihood-ratio (weight)

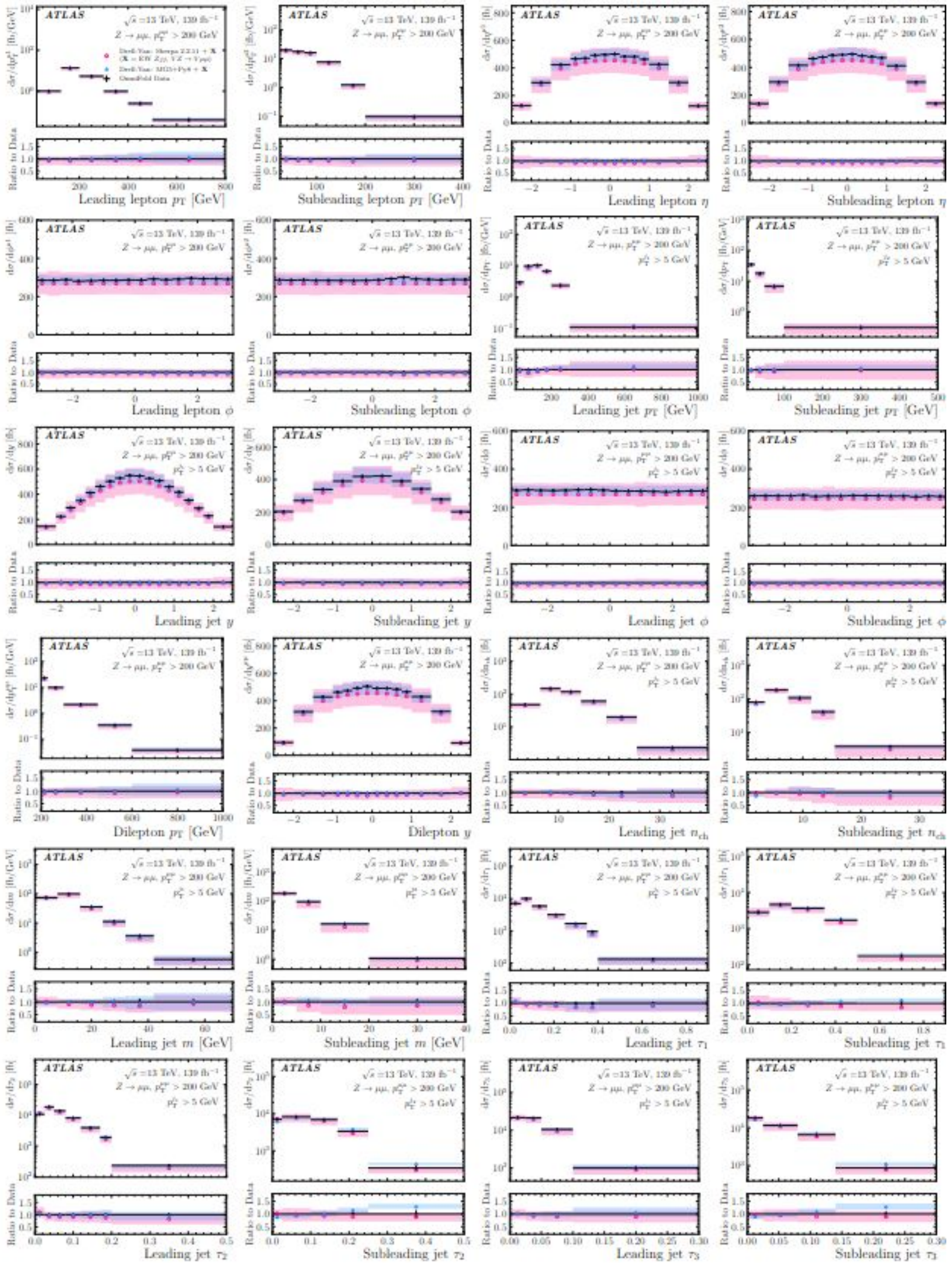
$$\boxed{w(\vec{x}) = \frac{f(\vec{x})}{1 - f(\vec{x})} = \frac{p_A(\vec{x})}{p_B(\vec{x})}}$$

Iterative NN reweighting procedure using BCE loss function over datasets A and B

Unfolding with omnifold

[arXiv:2405.20041]

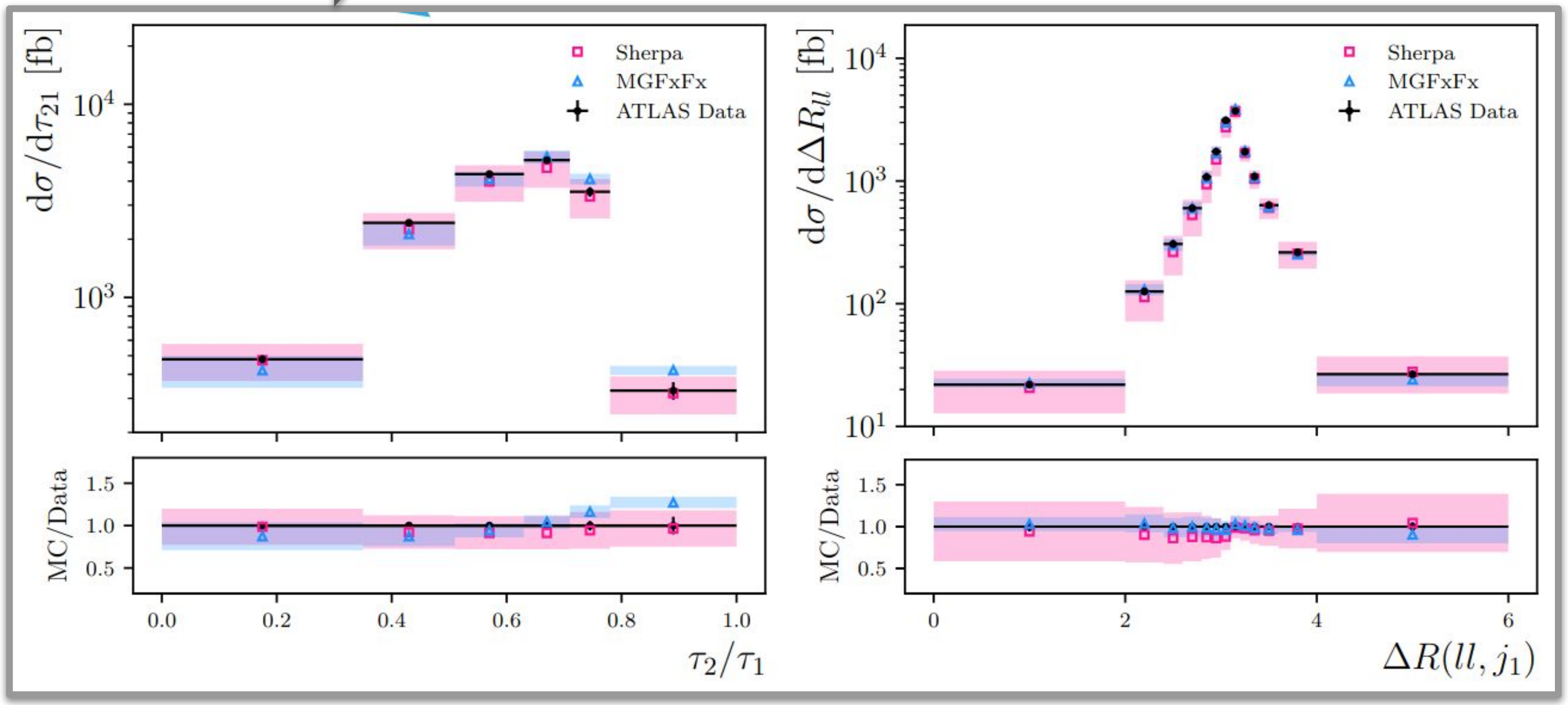
Z+jets process: x = 24 observables



Construct new observables from (reweighted) particle-level dataset: not presented in original paper

Dataset on Zenodo

Code on GitLab

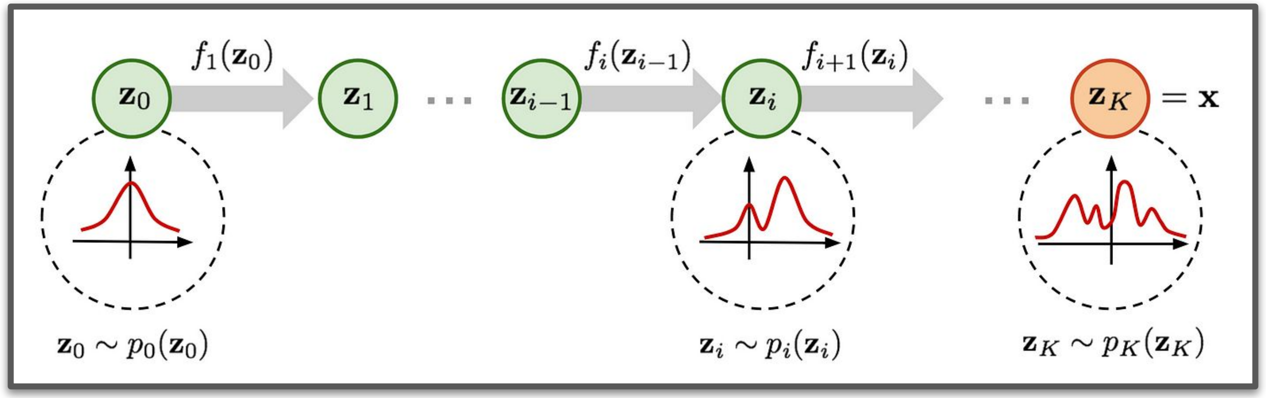


Taken from J. Duarte ICHEP slides

Invertible networks for inference

- CINN: Conditional Invertible Neural Network (e.g. Normalising Flow)
  - Map complex observable space to simple base distribution
  - <u>**Conditional on parameters we are trying to infer**</u>
  - Apply to high dimensional feature space → limited information loss
  - Learning the density, p(x|θ) !

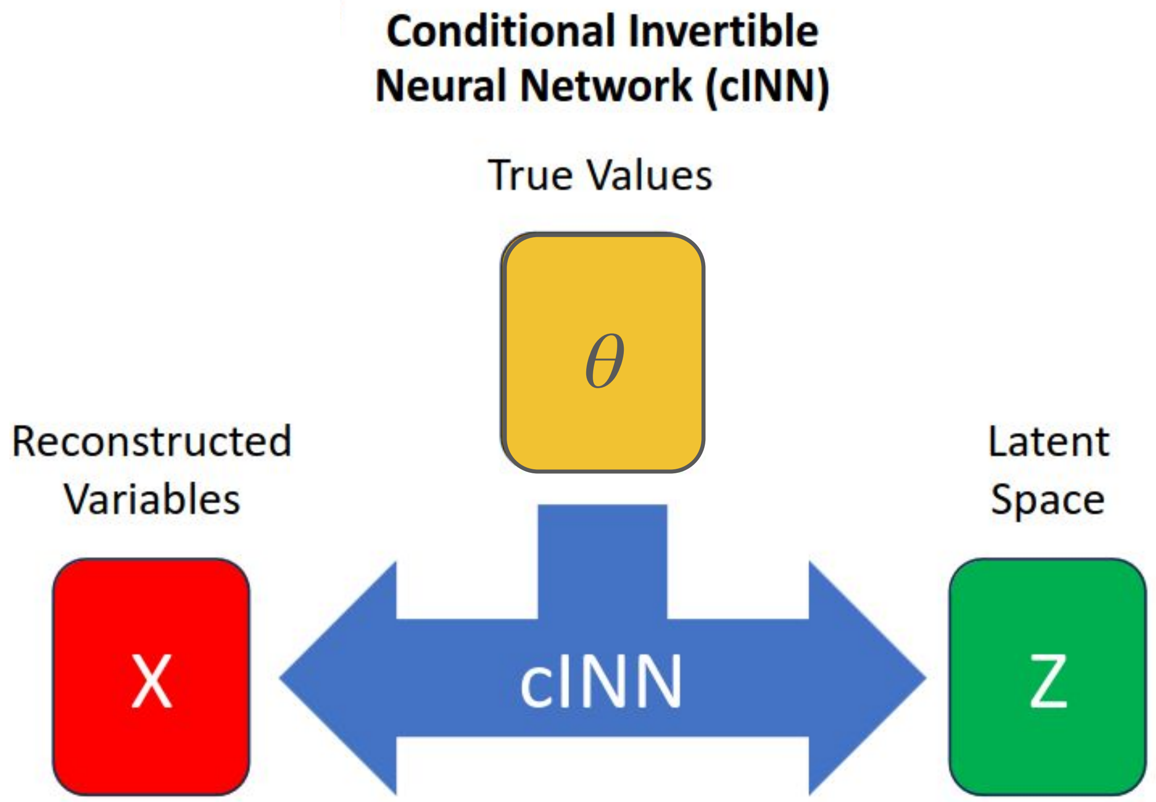
Conditional on θ



$$p(x|\theta) = p_z\left(T^{-1}(x|\theta)\right)\left|\mathrm{Det}\, J_{T^{-1}}(x|\theta)\right|$$

(Learnt) Transformations of x to latent space z
Conditional on θ. Evaluate simple base distribution density

Conserves probability mass



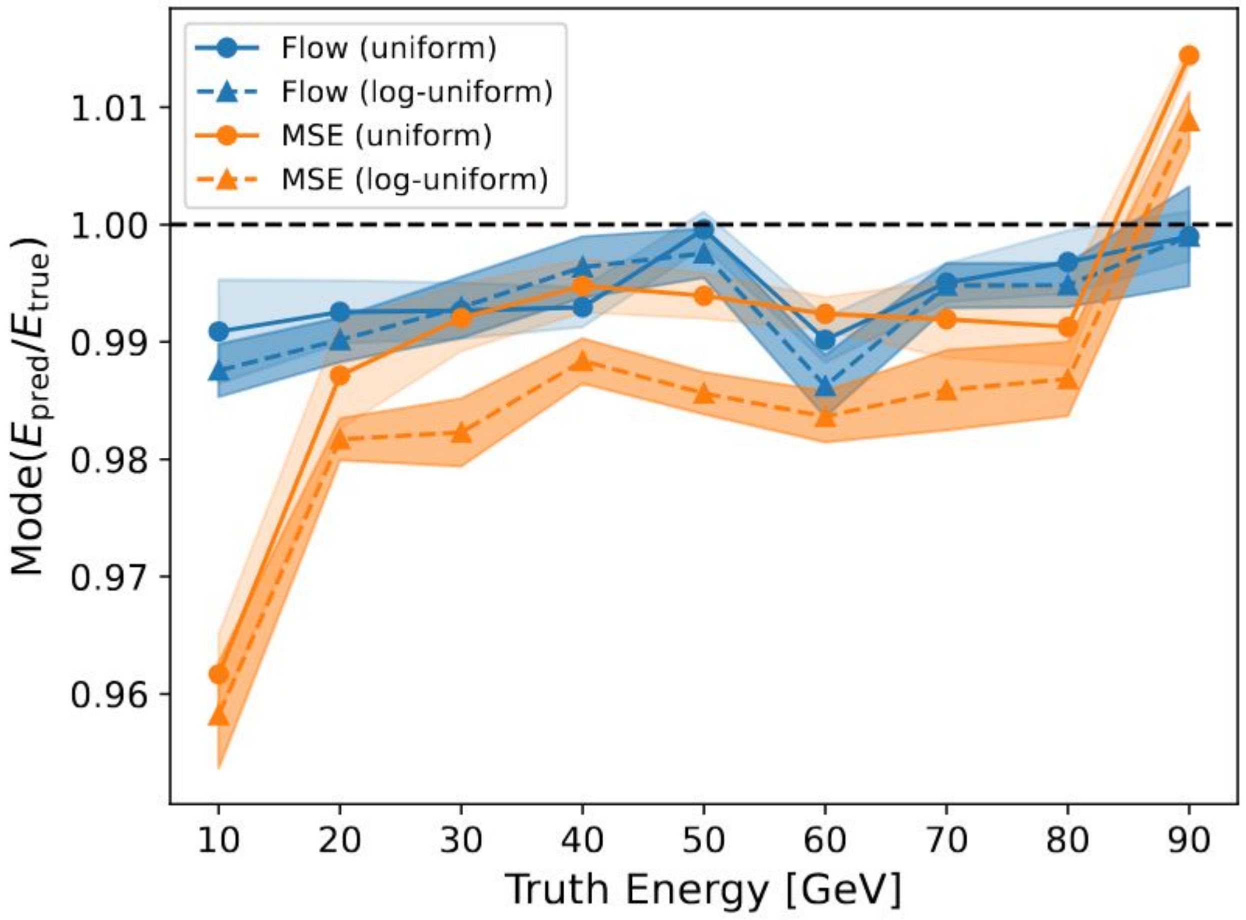
$$\boxed{L(\mathcal{D}|\theta) = \prod_{x_i \in \mathcal{D}} p(x_i|\theta)}$$

Invertible networks for inference

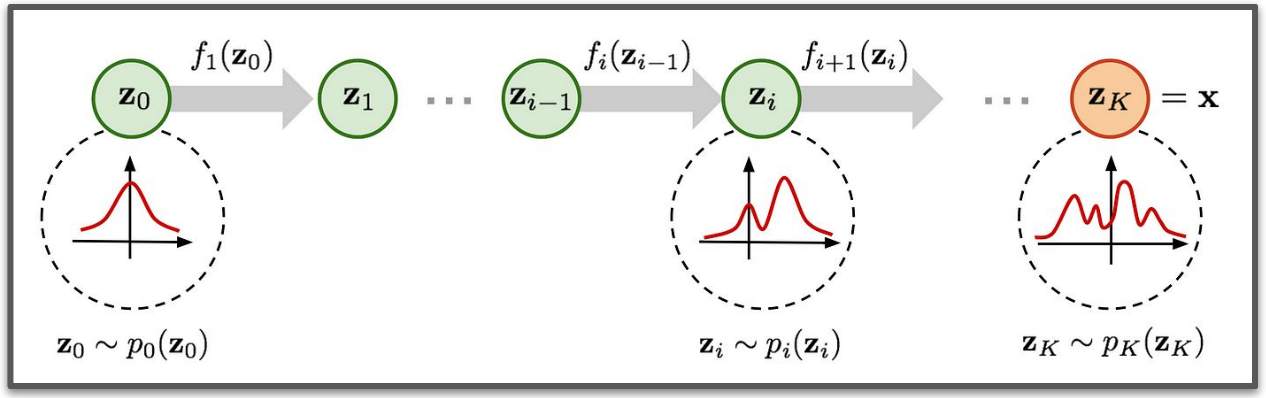
- CINN: Conditional Invertible Neural Network (e.g. Normalising Flow)

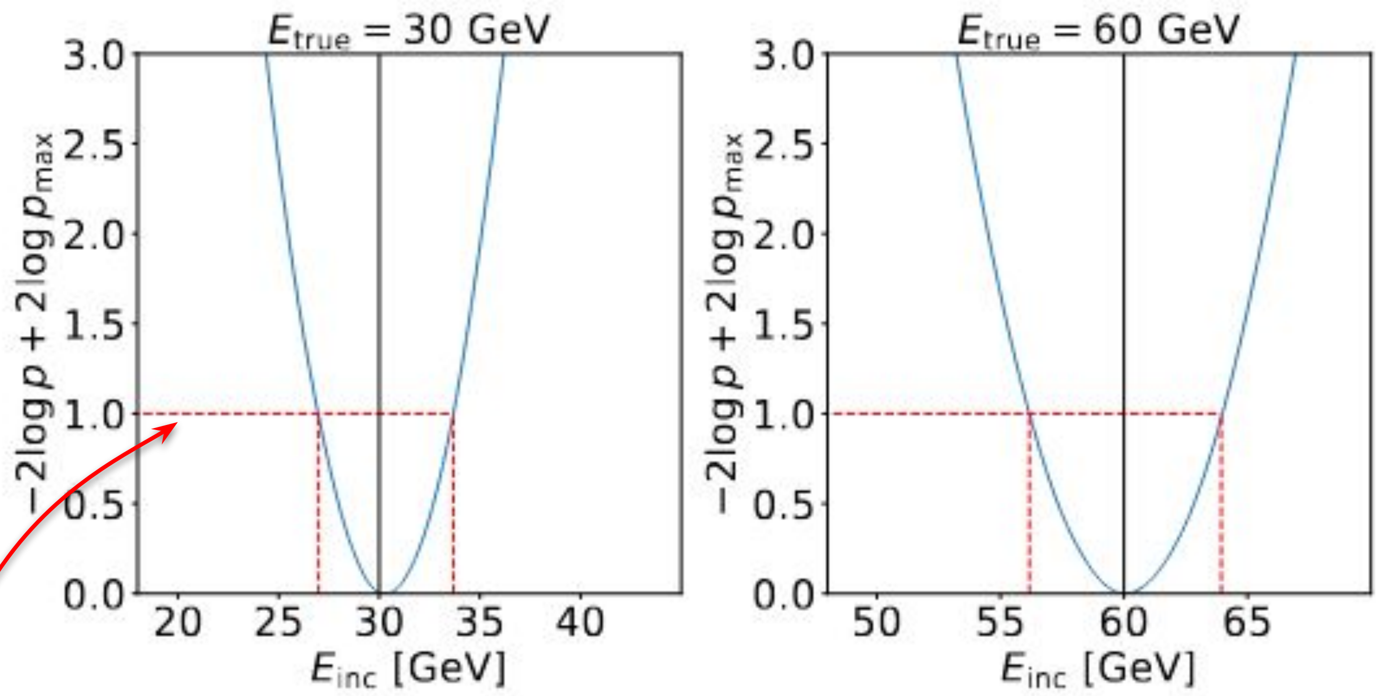
Example: CALOFLOW [arXiv:2404.18992v1]

Infer incident pion energy (θ) from measured energy in calorimeter cells (x)



Conditional on θ





$$p(x|\theta) = p_z\left(T^{-1}(x|\theta)\right)\left|\mathrm{Det}\, J_{T^{-1}}(x|\theta)\right|$$

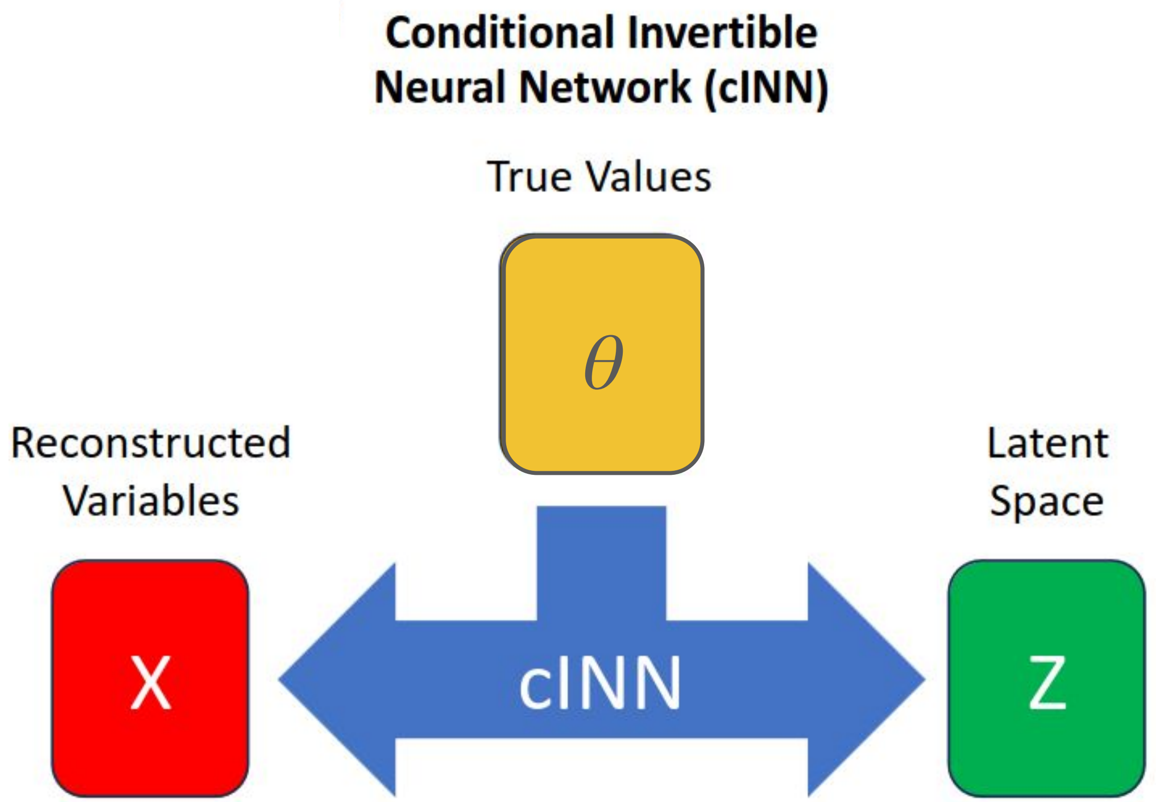
Learnt full density: per-shower resolution estimates

Invertible networks for generation

$$\{z_0\} \sim \mathcal{N}(0, \mathbb{1})$$

$$x_{\mathrm{gen}} = T(z_0|\theta) = f(z_0|\theta;\phi)$$

- Flows are invertible → use as generative model
  - Sample over base distribution, z_0
  - Obtain synthetic data $\{x_{gen}\}$ for fixed value of θ which follows learned conditional density
  - Significantly less compute than expensive MC simulation



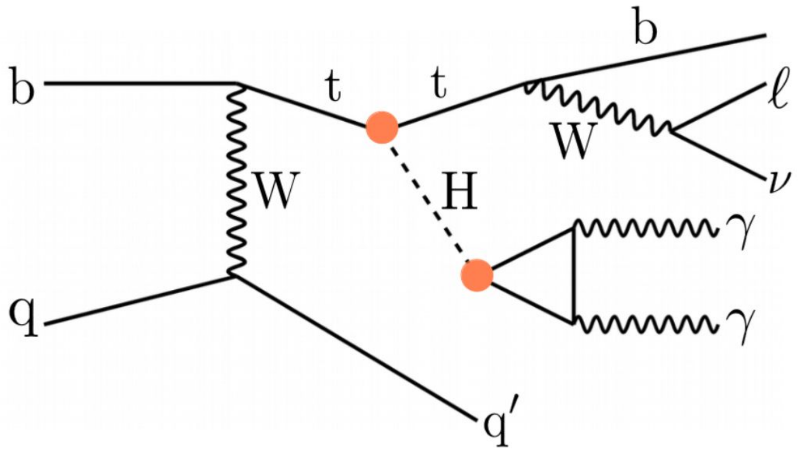
Invertible networks for generation

$$\{z_0\} \sim \mathcal{N}(0, \mathbb{1})$$

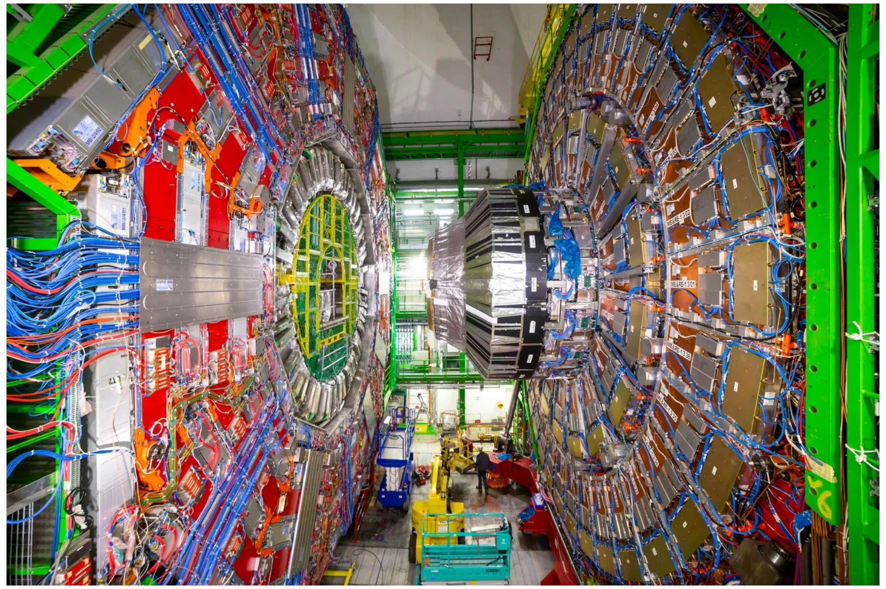
$$x_{\text{gen}} = T(z_0|\theta) = f(z_0|\theta; \phi)$$

- Flows are invertible → use as generative model
  - Sample over base distribution, z_0
  - Obtain synthetic data $\{x_{\text{gen}}\}$ for fixed value of θ which follows learned conditional density
  - Significantly less compute than expensive MC simulation e.g. **FlashSim at CMS**

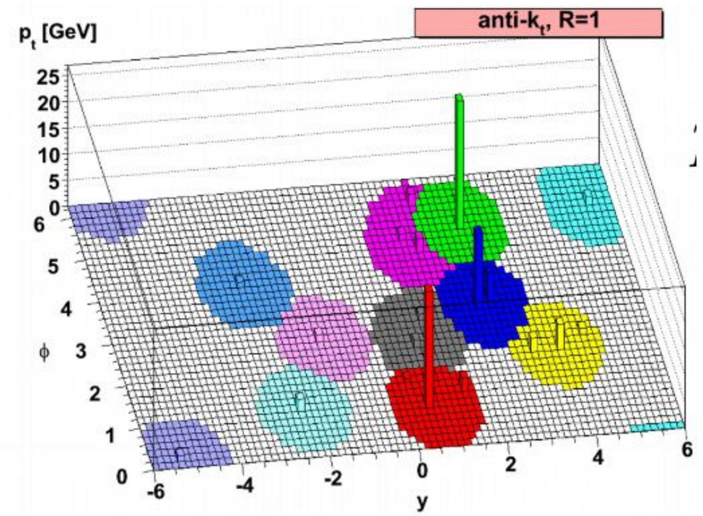
[CHEP2023 Talk]



Particle-level (truth)



Expensive CMS detector simulation (Geant4)



Detector-level (reconstructed)

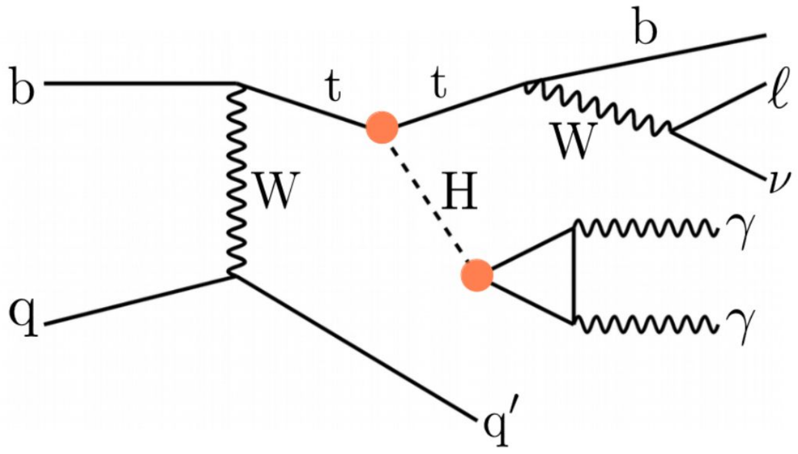
Invertible networks for generation

$$\{z_0\} \sim \mathcal{N}(0, \mathbb{1})$$

$$x_{\text{gen}} = T(z_0|\theta) = f(z_0|\theta;\phi)$$

- Flows are invertible → use as generative model
  - Sample over base distribution, z_0
  - Obtain synthetic data $\{x_{gen}\}$ for fixed value of θ which follows learned conditional density
  - Significantly less compute than expensive MC simulation e.g. **FlashSim at CMS**

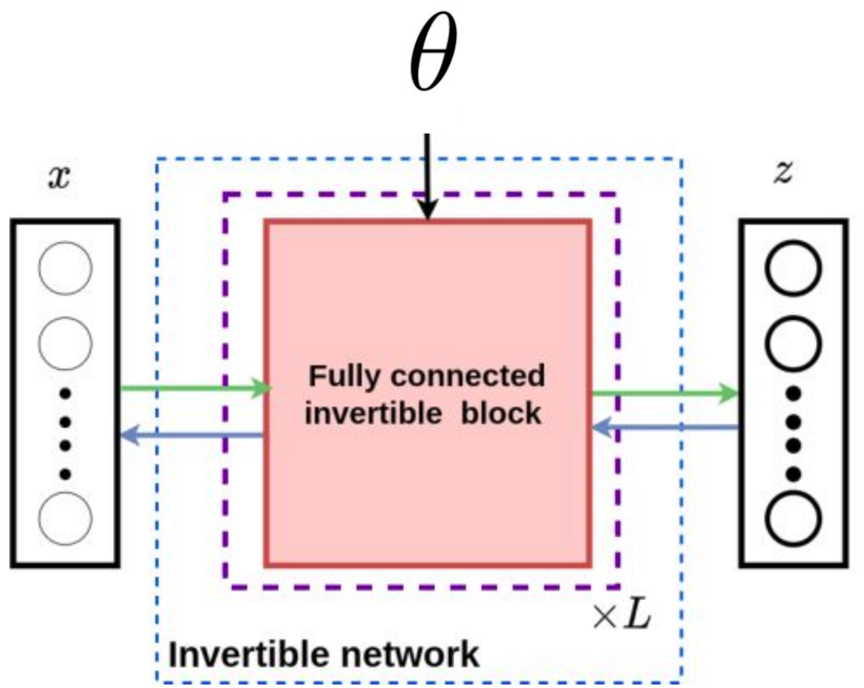
[CHEP2023 Talk]



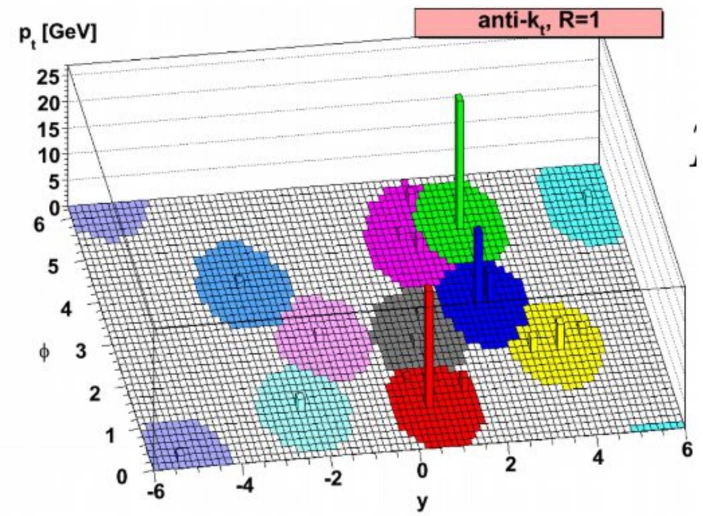
Particle-level (truth)

Conditional variables, θ

Generated particles momentum, angle, mass, flavour



Replace with surrogate of full detector simulation using flow

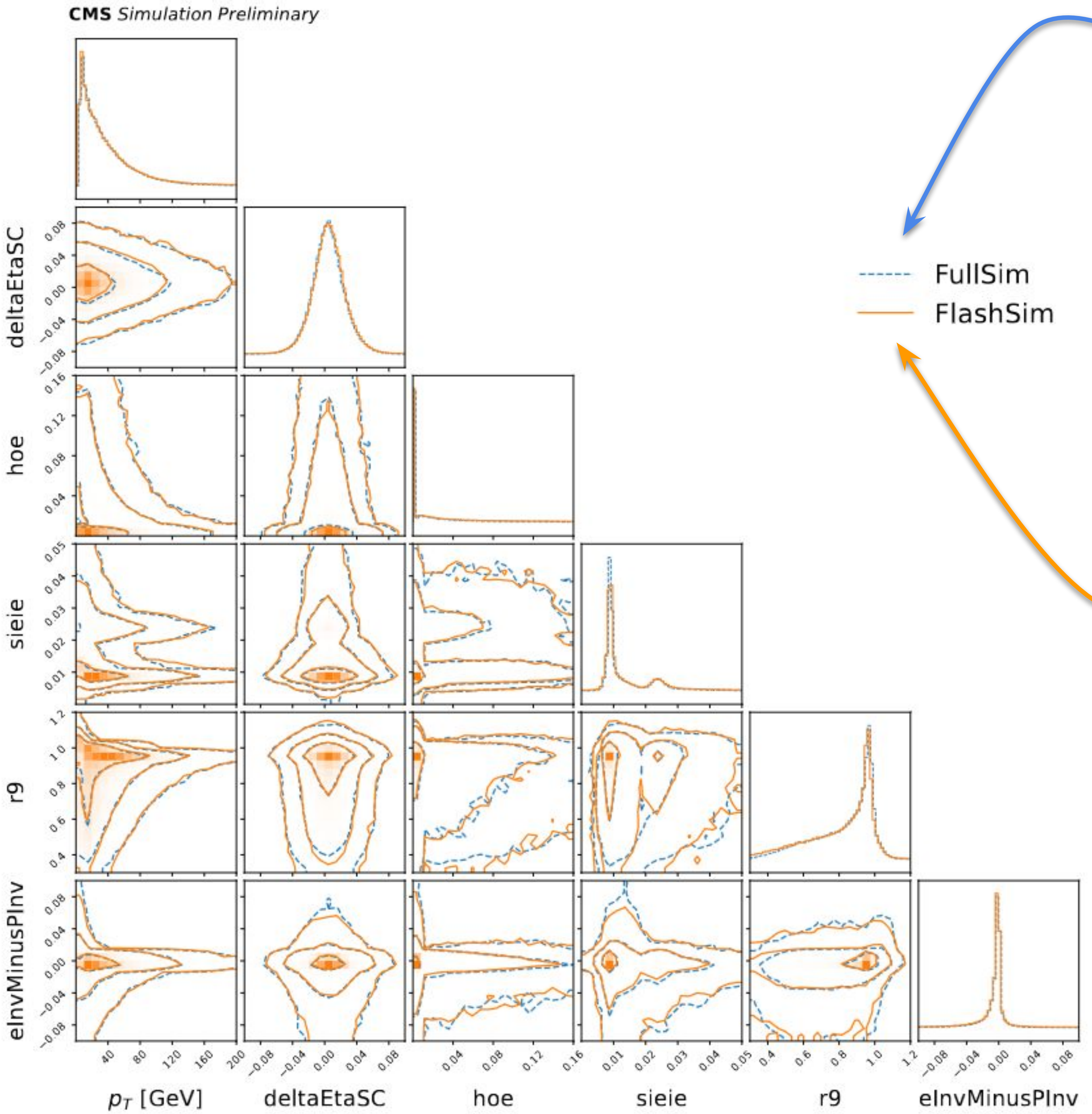


Detector-level (reconstructed)

Input data, x

Reconstructed particle momenta, angle, ID, ...

FlashSIM at CMS



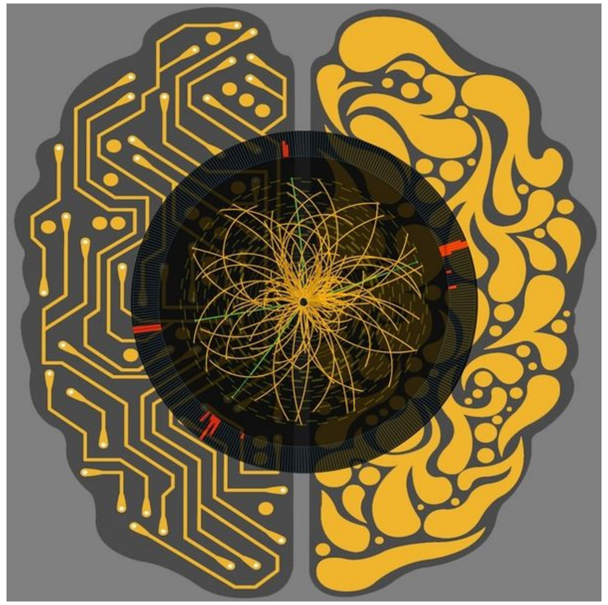
Full high-fidelity MC simulation

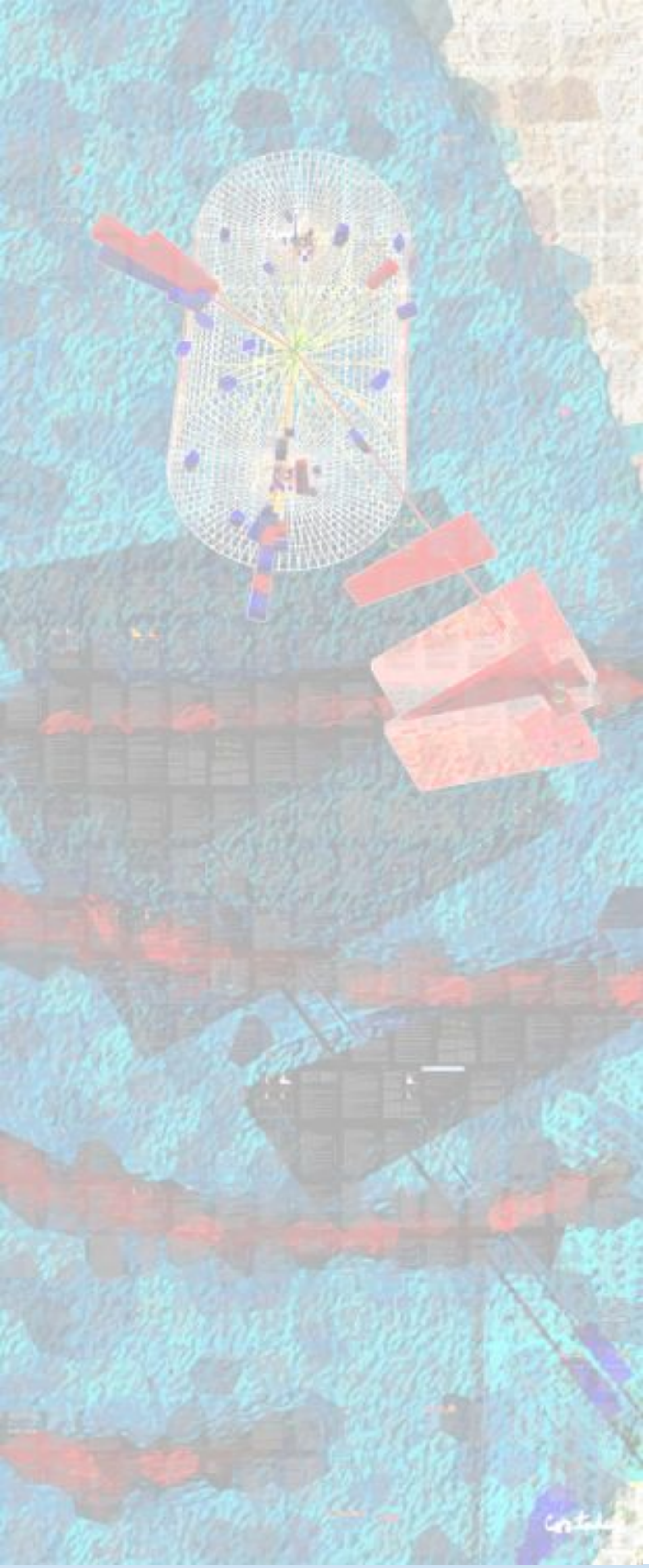
Generative flow

[CHEP2023 Talk]

Outlook

- Covered many different “**Types of ML in Particle Physics**”: BDT, DNN, CNN, GNN, Transformer, GAN, NF, ...
  - With vast array of applications: object identification/reconstruction, event classification, anomaly detection, generation, inference
  - Only a subset: diffusion models, detector design & optimisation, pileup mitigation, background prediction, ...
  - ML is clearly opening up many new possibilities in the field!
- As our dependence on ML grows → **Must ensure we use tool correctly**
  - Performance is not the only relevant metric
  - Focus on robustness, interpretability, insensitivity to modeling details, ...
  - E.g. systematic-aware learning, domain adversarial training
- We will cover these kind of topics over [Phystat: Stats meets ML]
  - Plenty of interesting discussions to come!

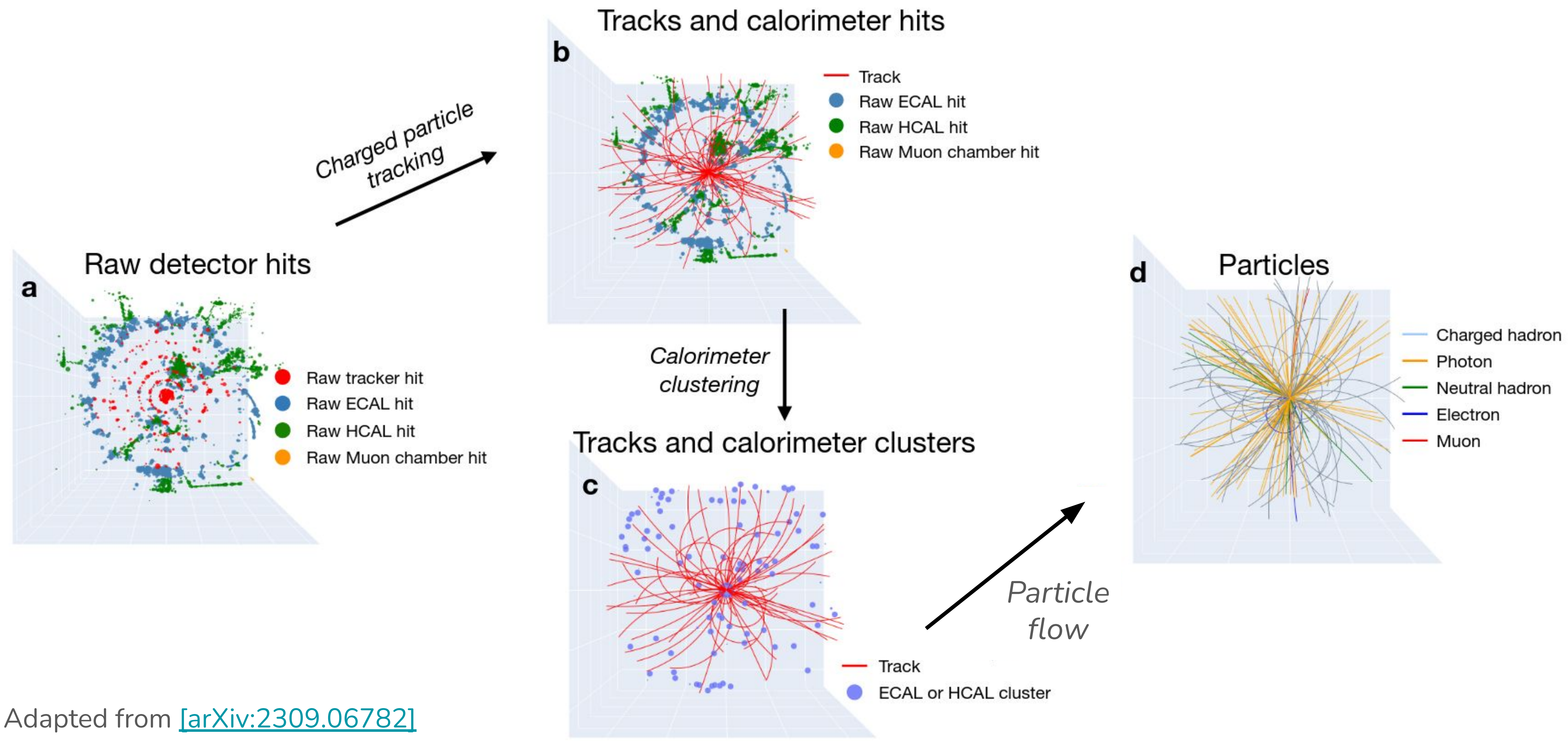




Back-Up

Object reconstruction

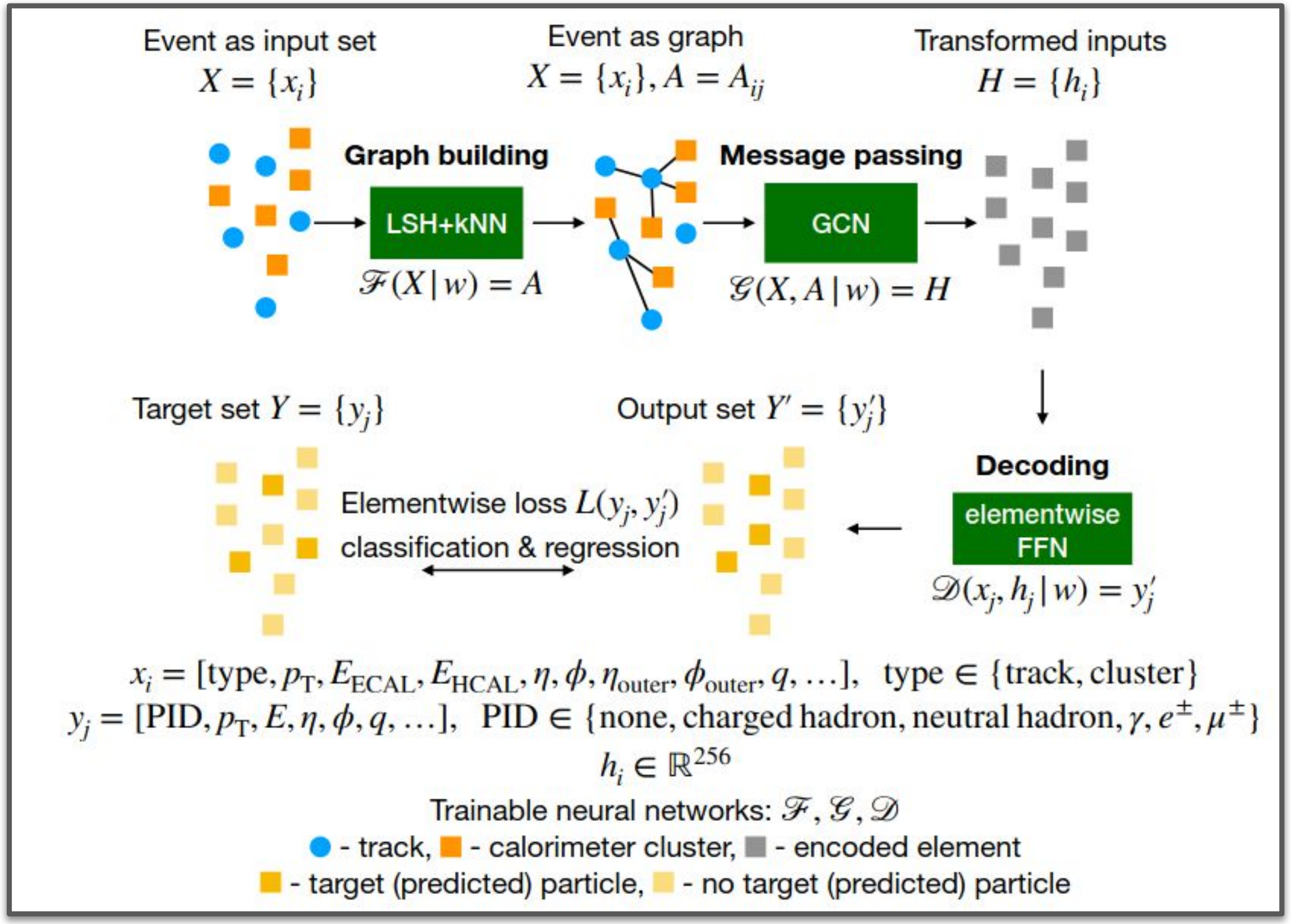
- Previous slides assume object (jet) has already been reconstructed from detector read-outs
  - Traditional object reconstruction follows rule-based algorithms (e.g. Kalman Filter, DBScan, Particle Flow)



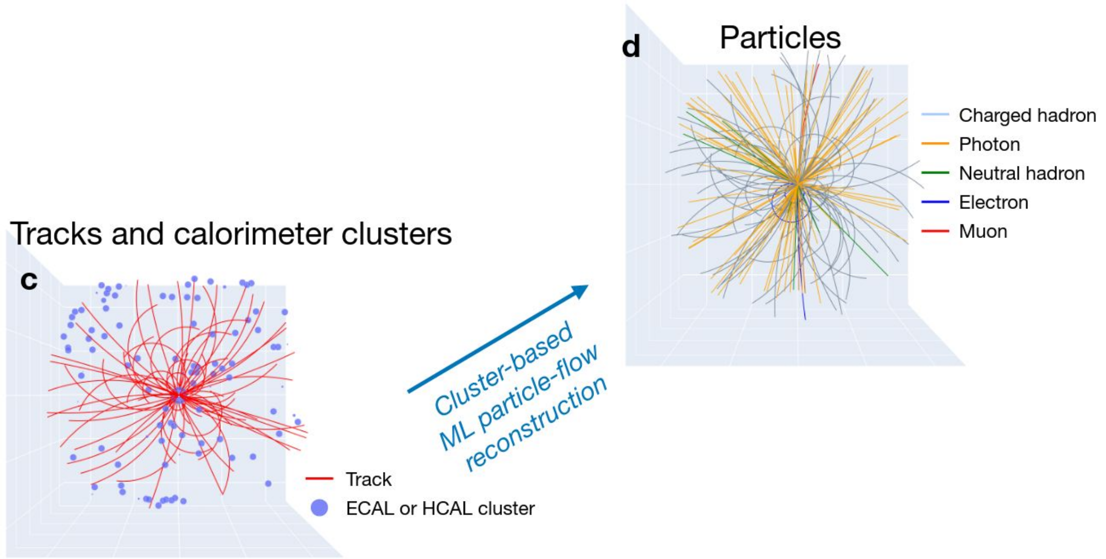
Adapted from [arXiv:2309.06782]

Object reconstruction

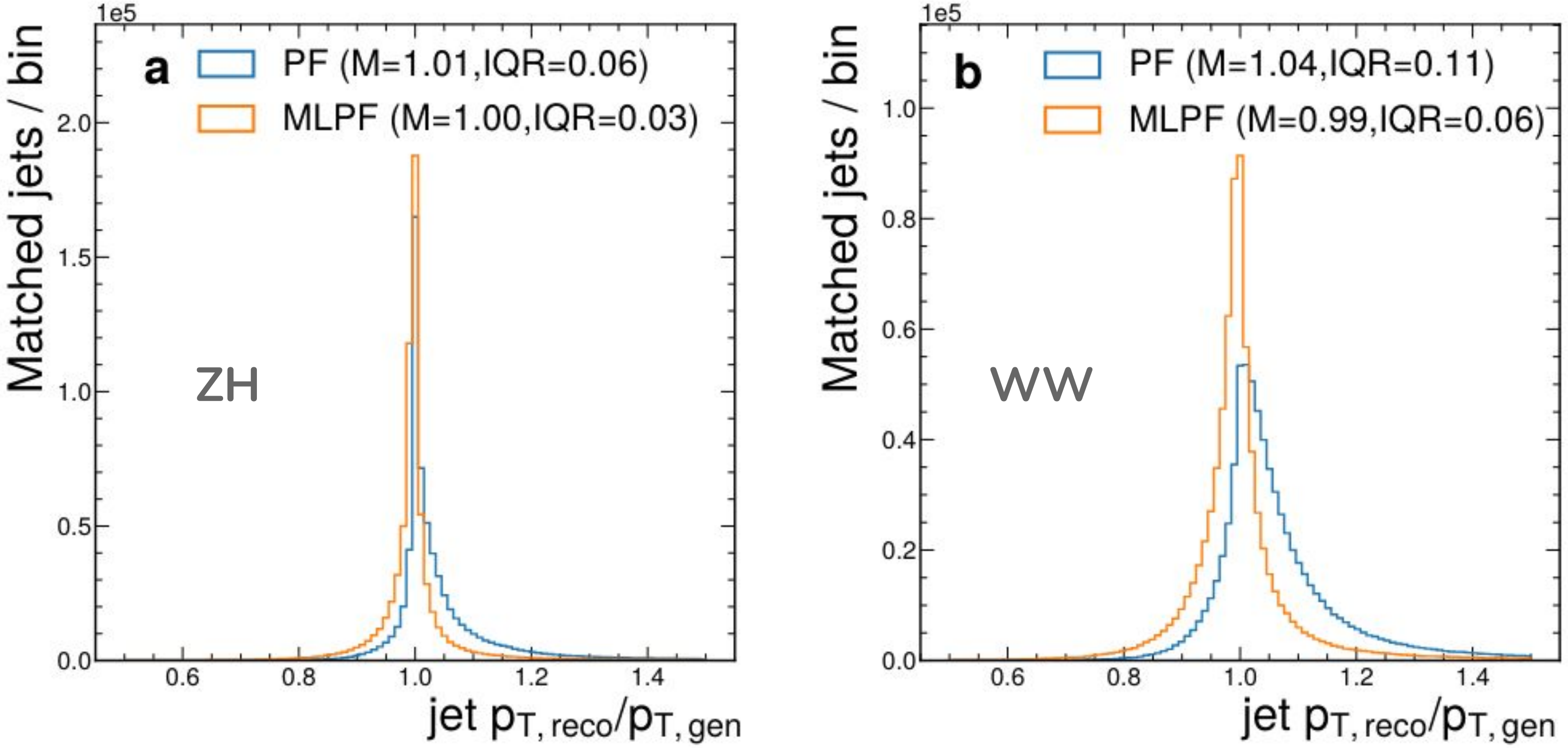
- Now investigating graph-based ML for reconstruction
  - Example: ML Particle Flow to **learn mapping** from tracks/clusters → particles



[arXiv:2101.08578]



Demonstrate improved performance over rule-based algorithms

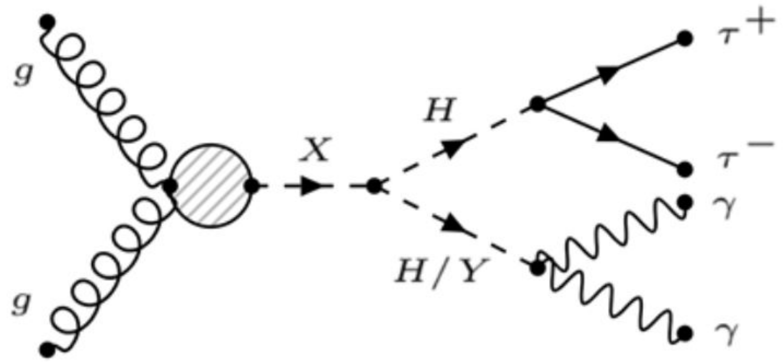


M: Median, IQR: Interquartile range

[arXiv:2309.06782]

Parametric classifiers

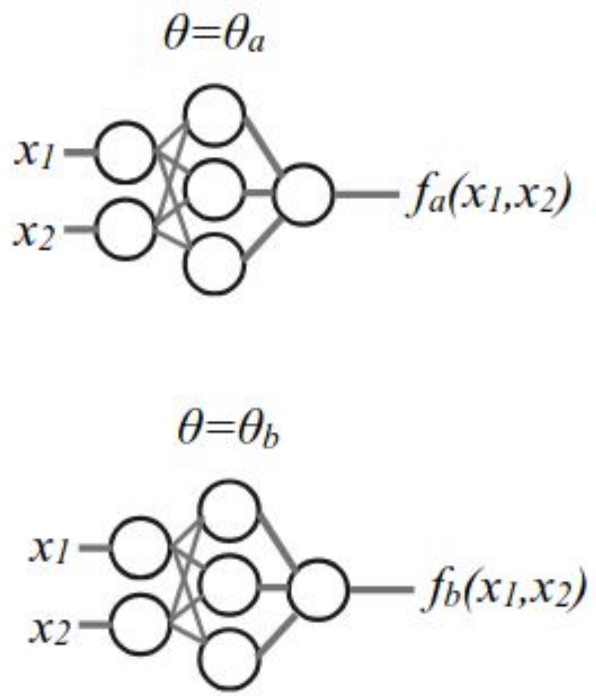
[CMS-PAS-HIG-22-012]



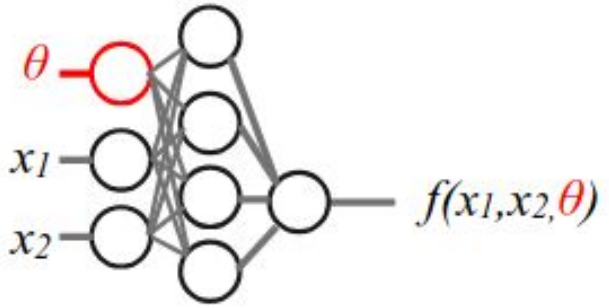
- What if we are searching for new physics (signal) over large parameter/hypothesis space?
  - Example: search for new resonant particle, X, with mass m_X in [250,1000] GeV
- Train ML classifier using MC simulation to identify signal-vs-background:

1. Train single classifier using simulation from many m_X hypotheses = sub-optimal
2. Train multiple classifiers, one at each hypothesis = unwieldy for large parameter space
3. **Parametric classifier:** output is conditional on m_X parameter

$$f(\vec{x}) \text{ to } f(\vec{x}; m_X)$$



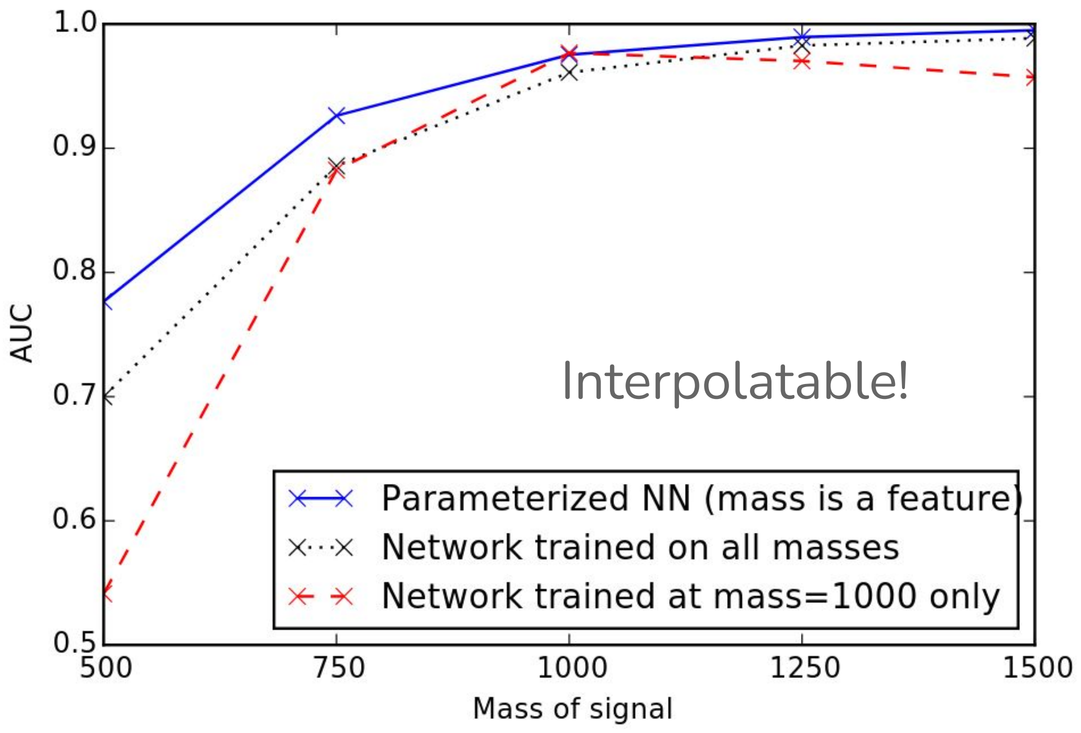
Add m_X as additional training feature (θ)



Train on all signal MC simultaneously {m1,m2,...}

Give background MC random values (in set)

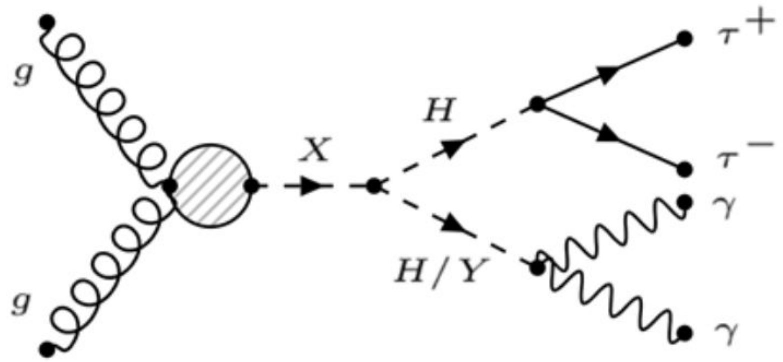
[EPJC 76 (2016) 5, 235]

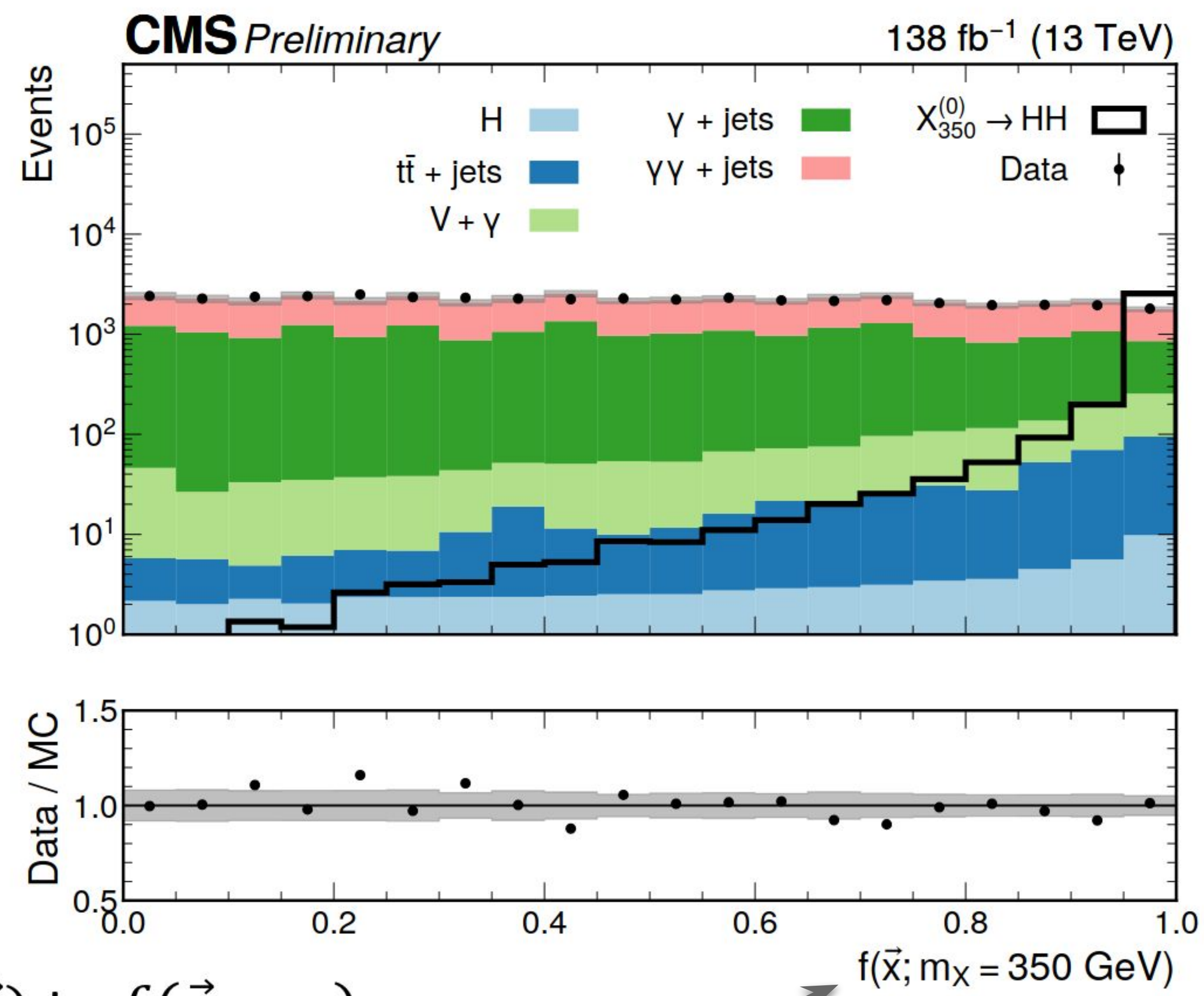


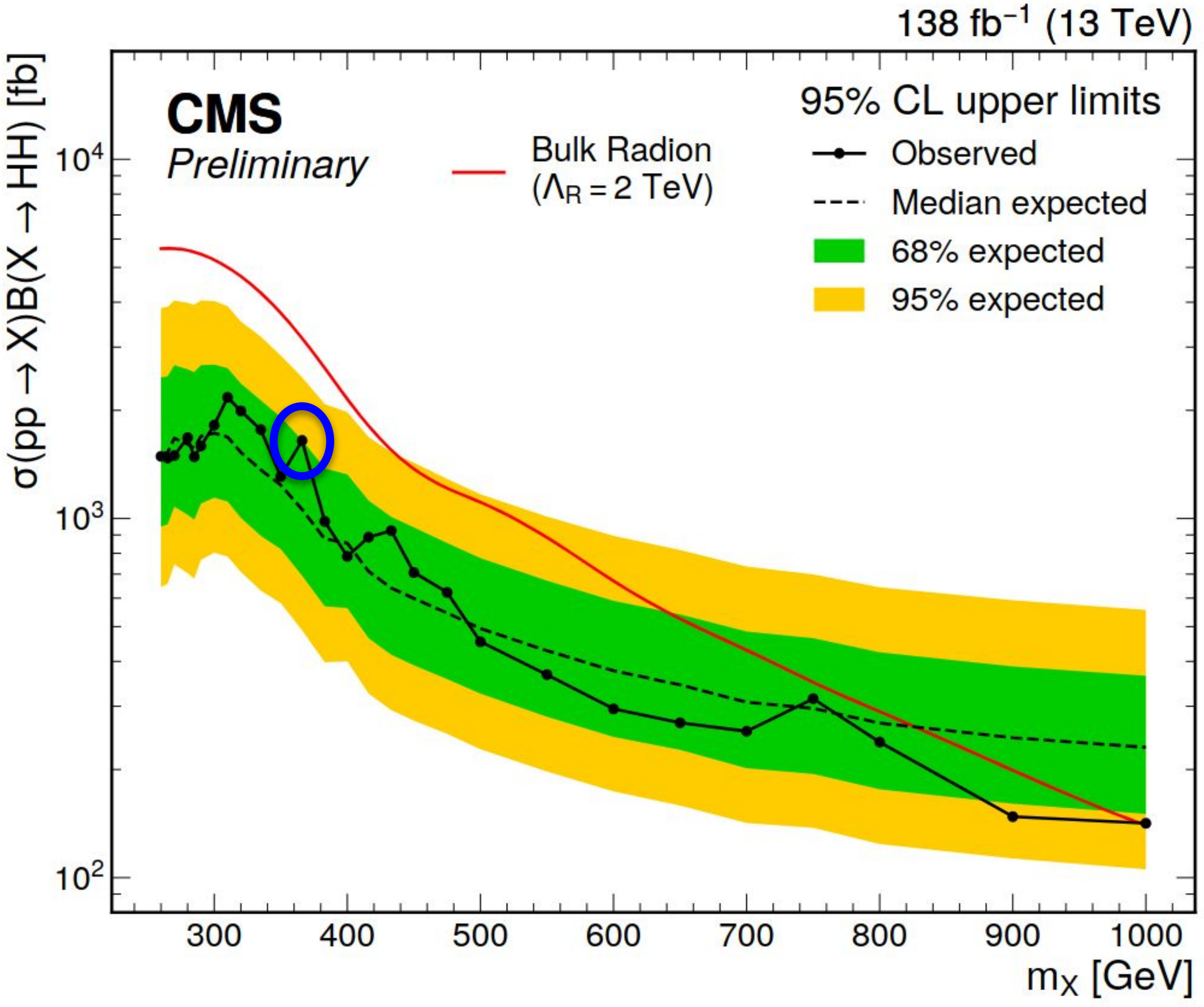
Parametric classifiers

[CMS-PAS-HIG-22-012]

- What if we are searching for new physics (signal) over large parameter/hypothesis space?
  - Example: search for new resonant particle, X, with mass m_X in [250,1000] GeV
- Train ML classifier using MC simulation to identify signal-vs-background:





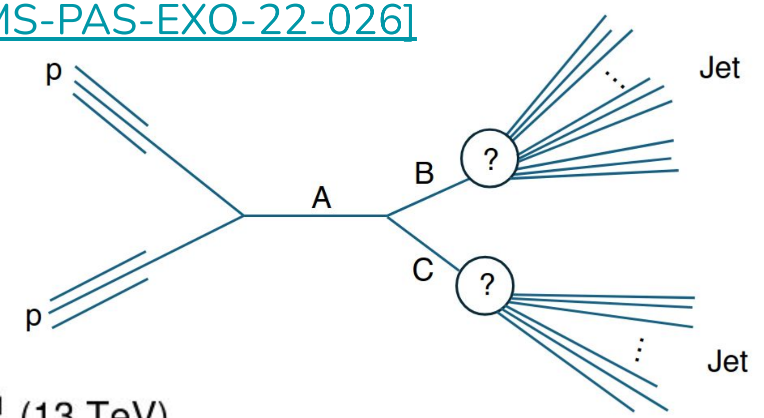


$f(\vec{x})$ to $f(\vec{x}; m_X)$

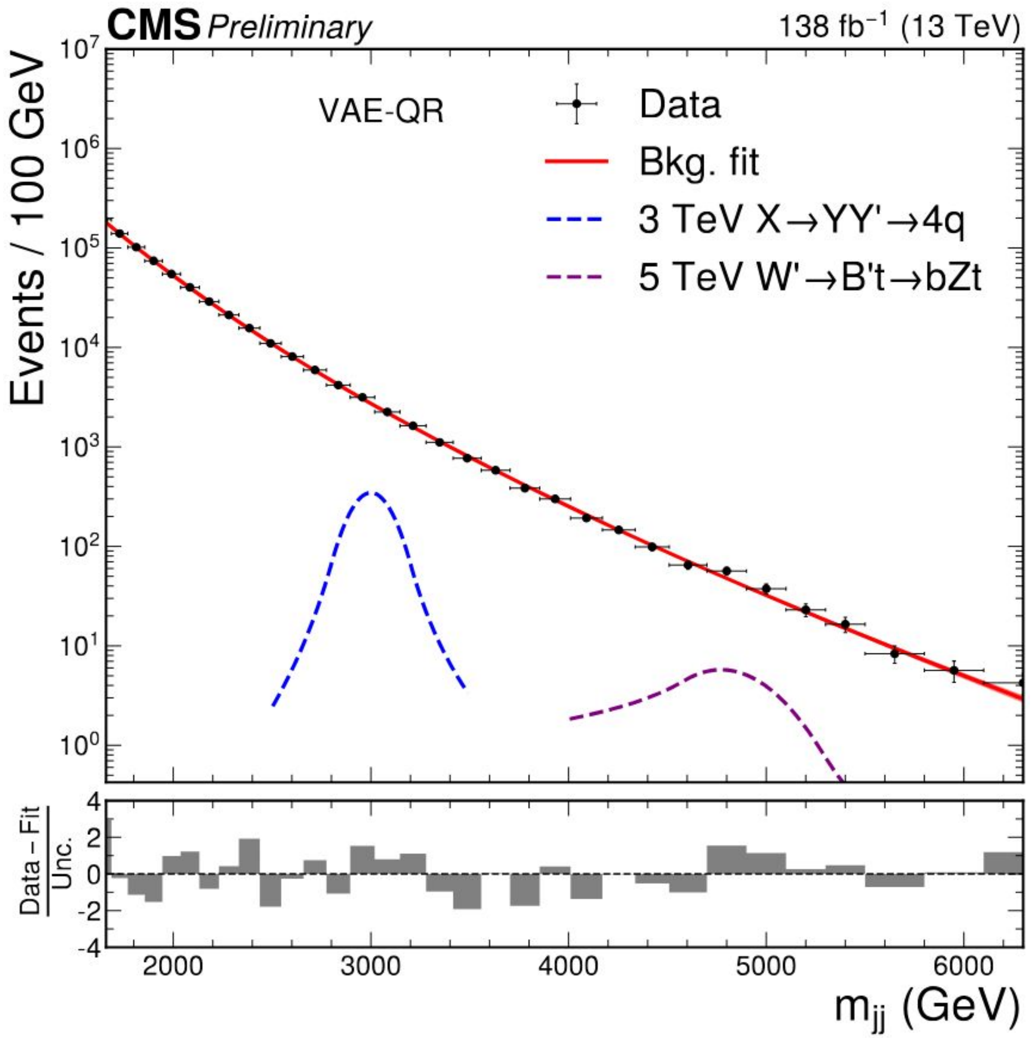
Optimised discriminant for each mass hypothesis

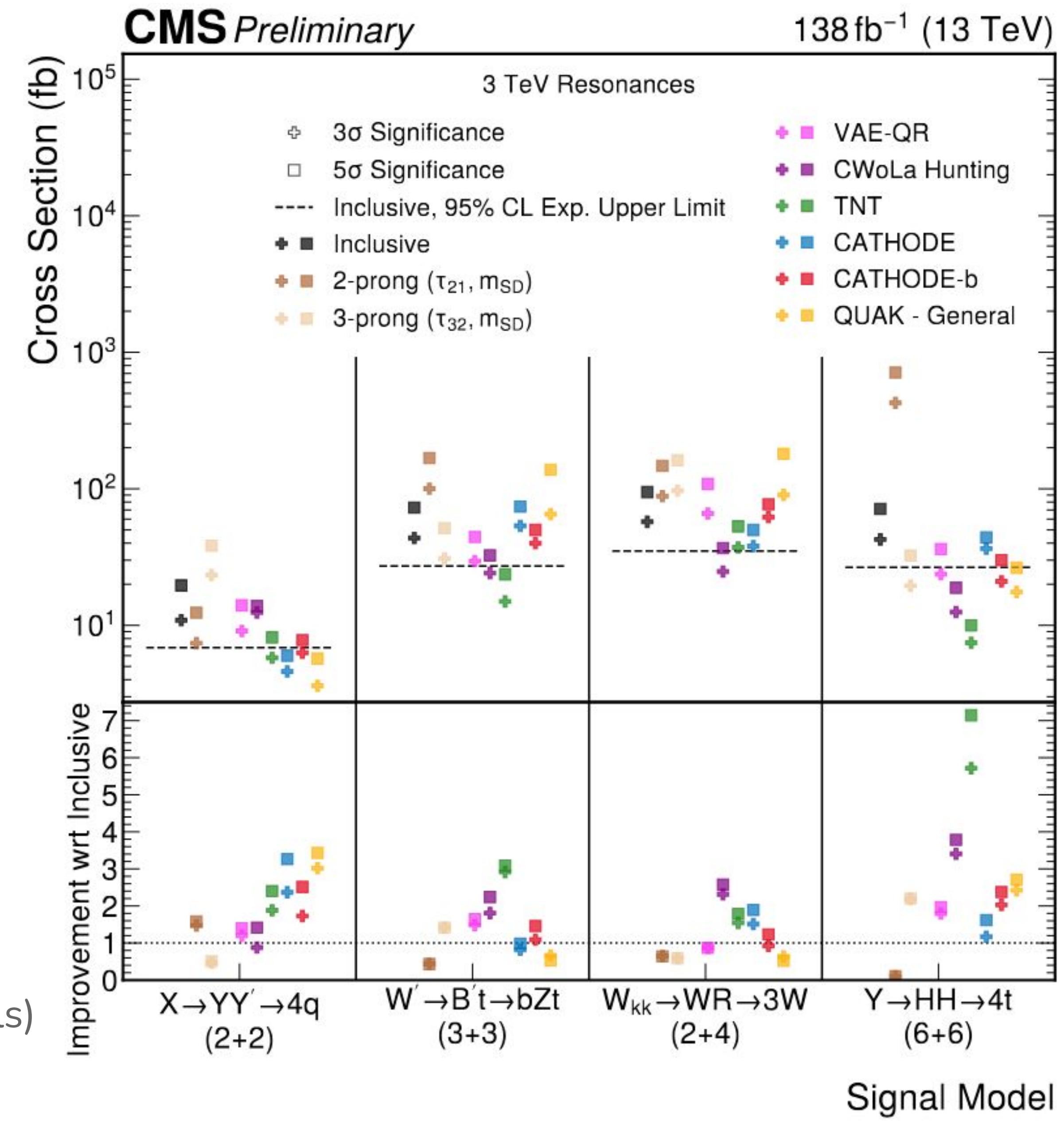
Model-agnostic searches

[CMS-PAS-EXO-22-026]



- What if we don't know what the signal looks like a-priori? Use **Anomaly detection algorithms**
- CMS apply model-agnostic approaches to dijet resonance searches with anomalous jet substructure





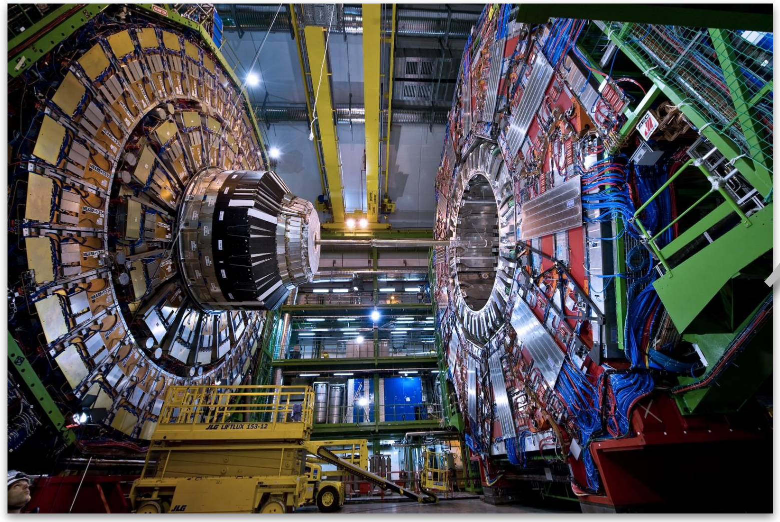
Increased discovery potential

Cross section needed for 5σ discovery reduced by up to factor of 7, compared to inclusive dijet search

Other approaches including semi-supervised learning (partial labels) and weakly-supervised learning (noisy labels)

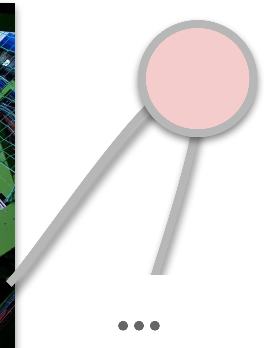
LHC triggering

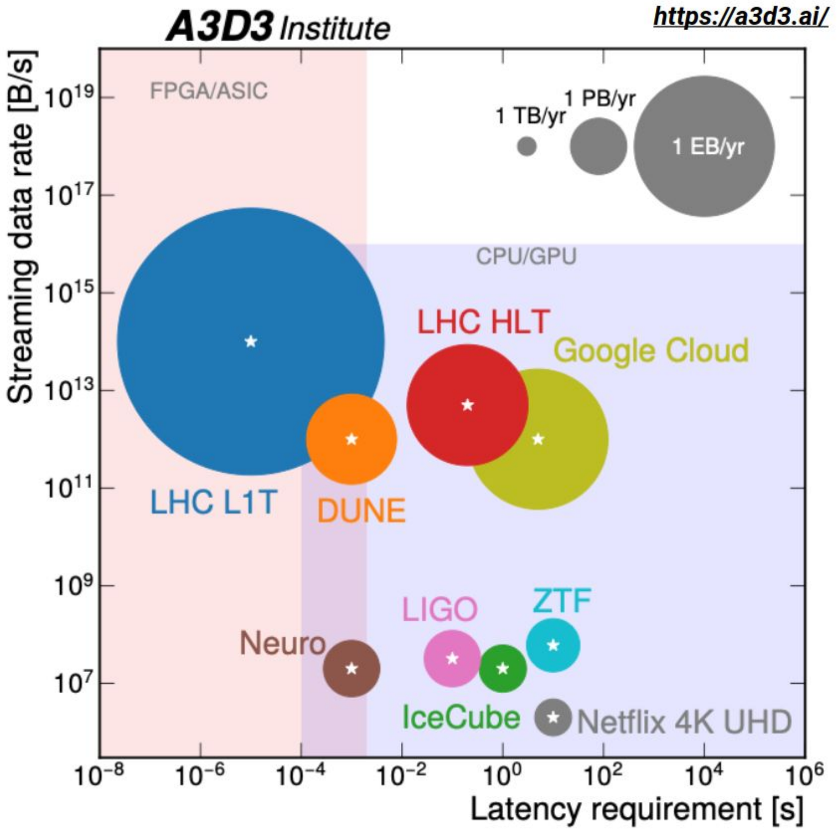
Data-taking

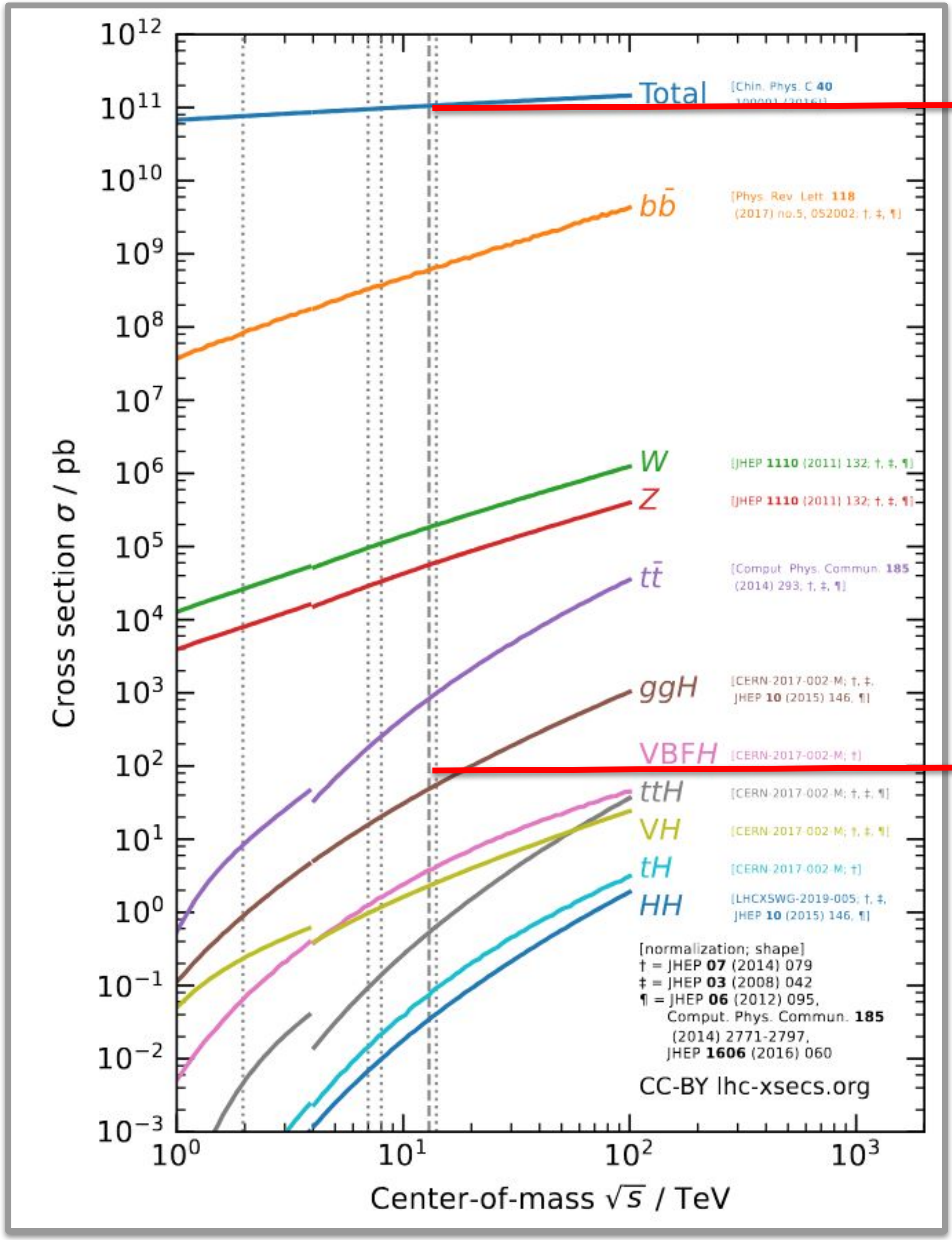


40 MHz collisions, O(100 million) readout channels

Online filtering (Trigger)







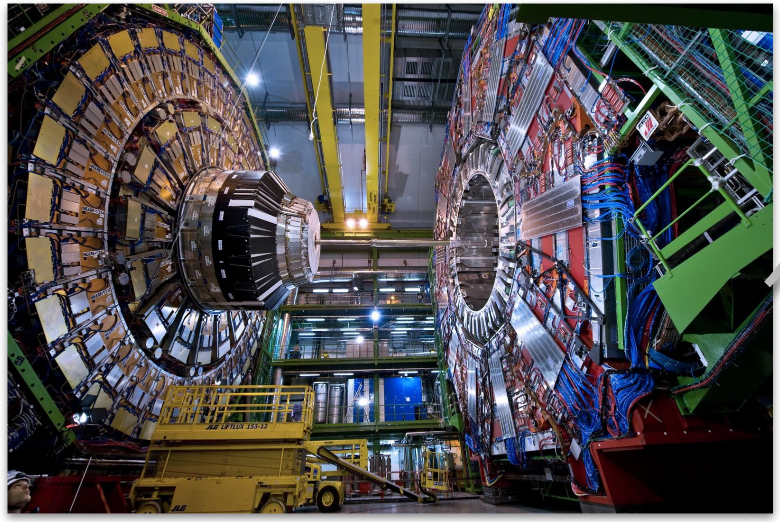
Probability to produce anything

Probability to produce a Higgs boson

~1 in a billion collisions

LHC triggering

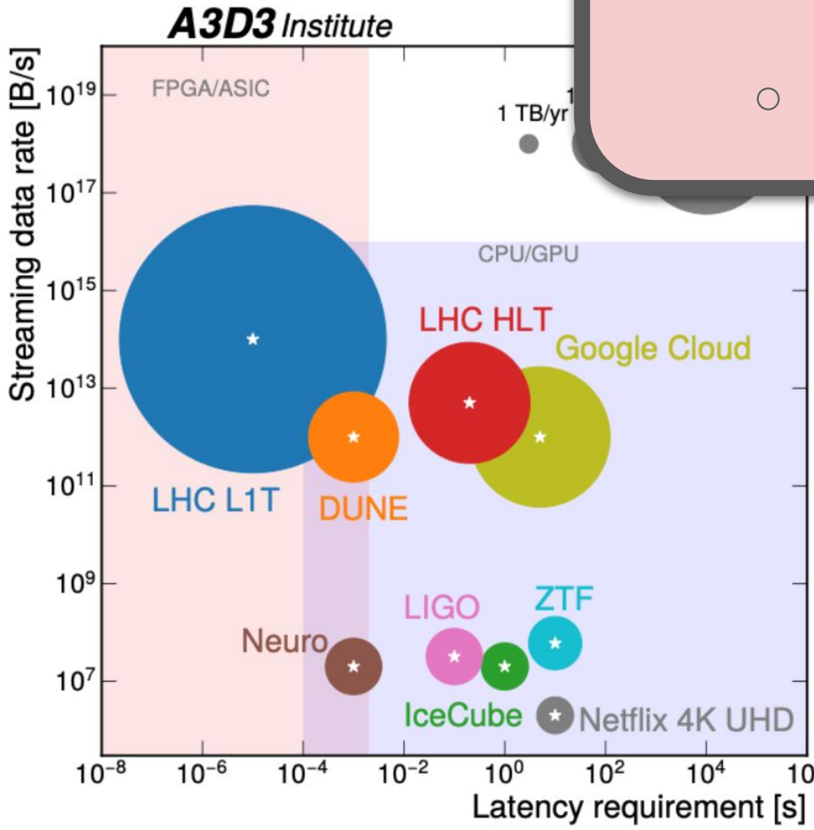
Data-taking



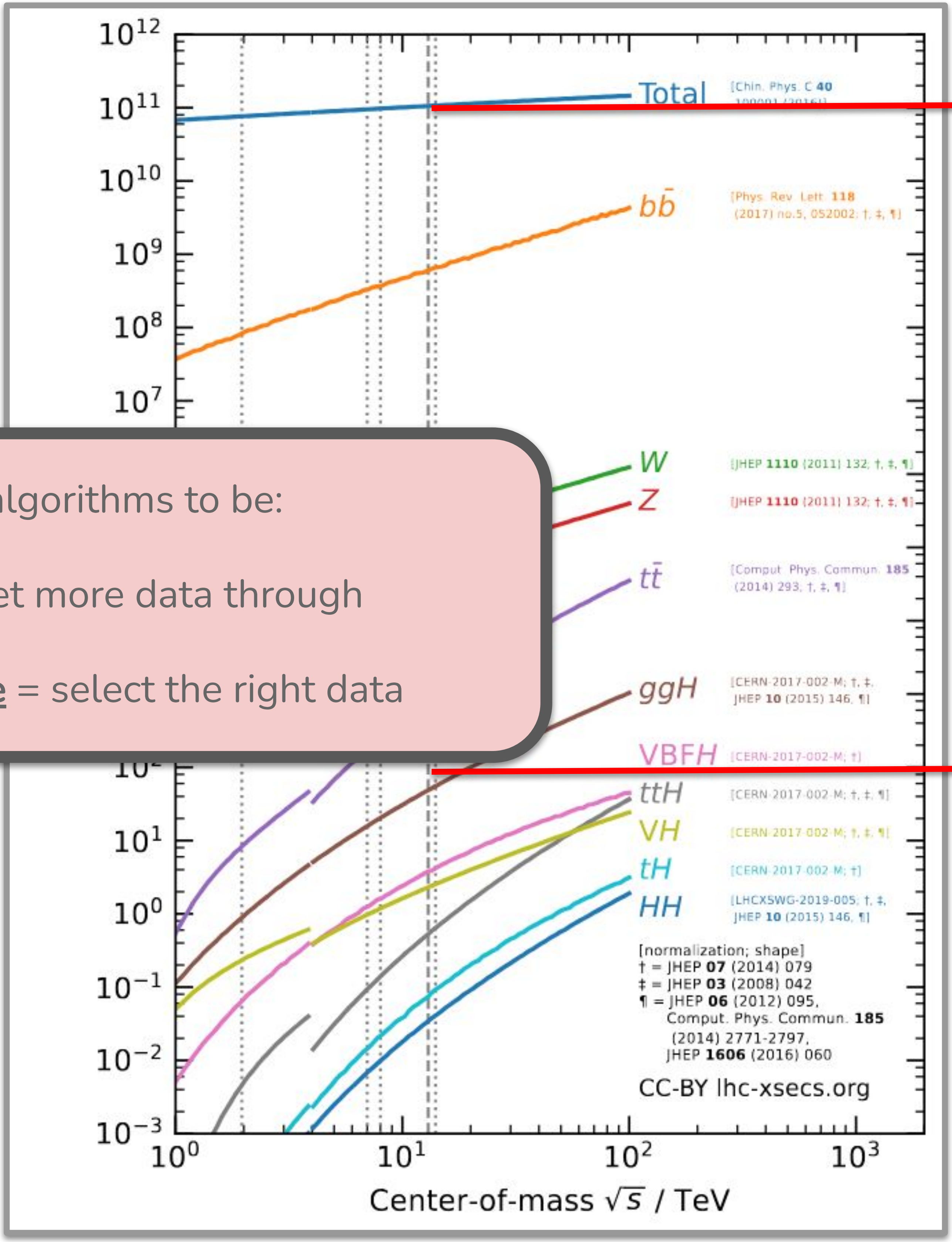
40 MHz collisions, O(100 million) readout channels

Online filtering (Trigger)

...



- We need trigger algorithms to be:
  - **Very fast** = get more data through
  - **Very accurate** = select the right data



Probability to produce anything

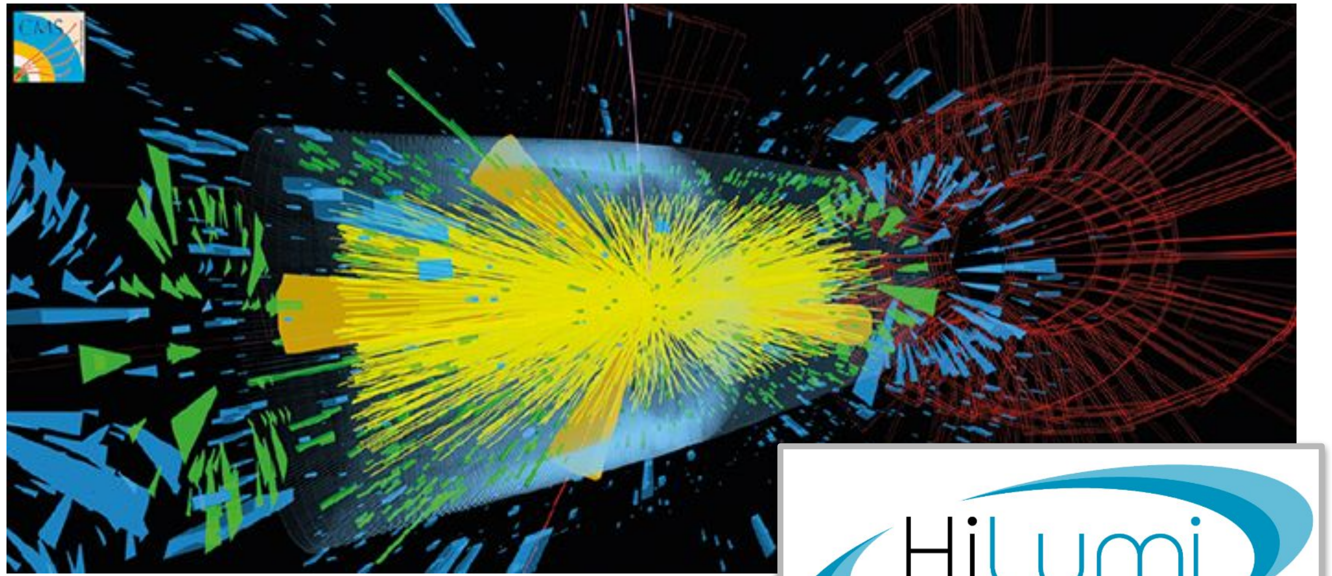
Probability to produce a Higgs boson

~1 in a billion collisions

FastML

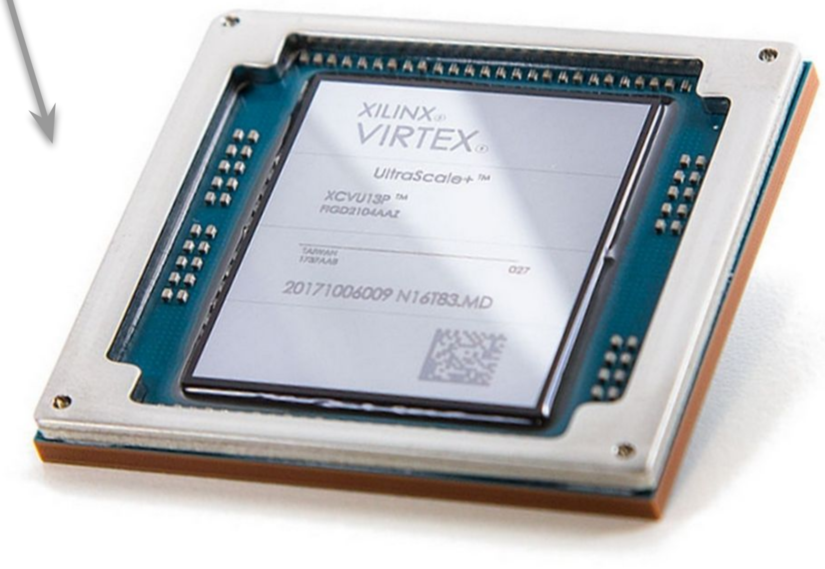
[FastML23 Workshop]

Pile-Up = number of simultaneous p-p collisions in bunch crossing

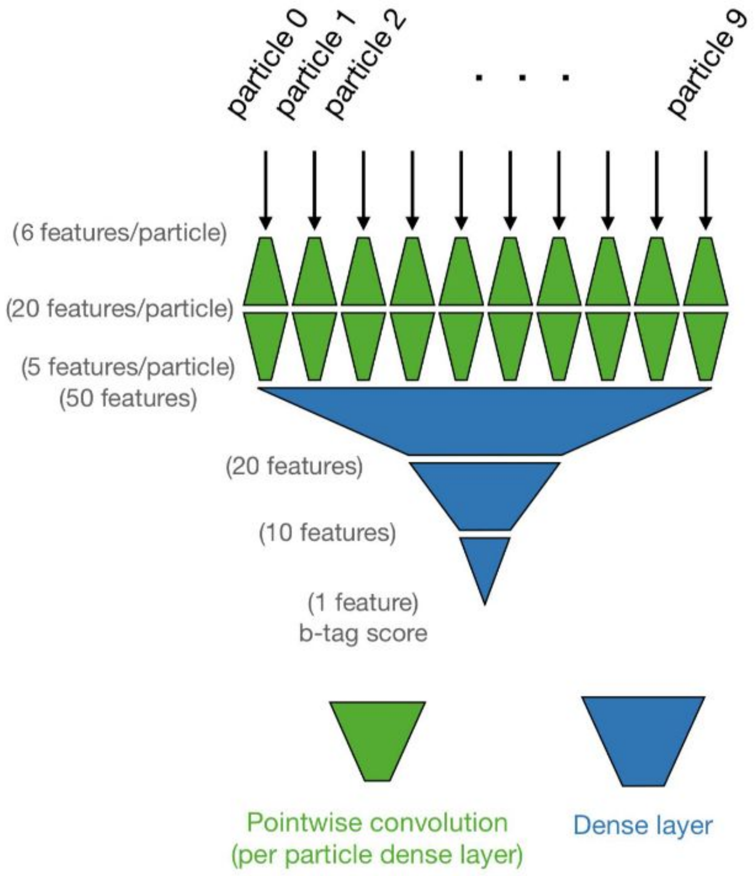


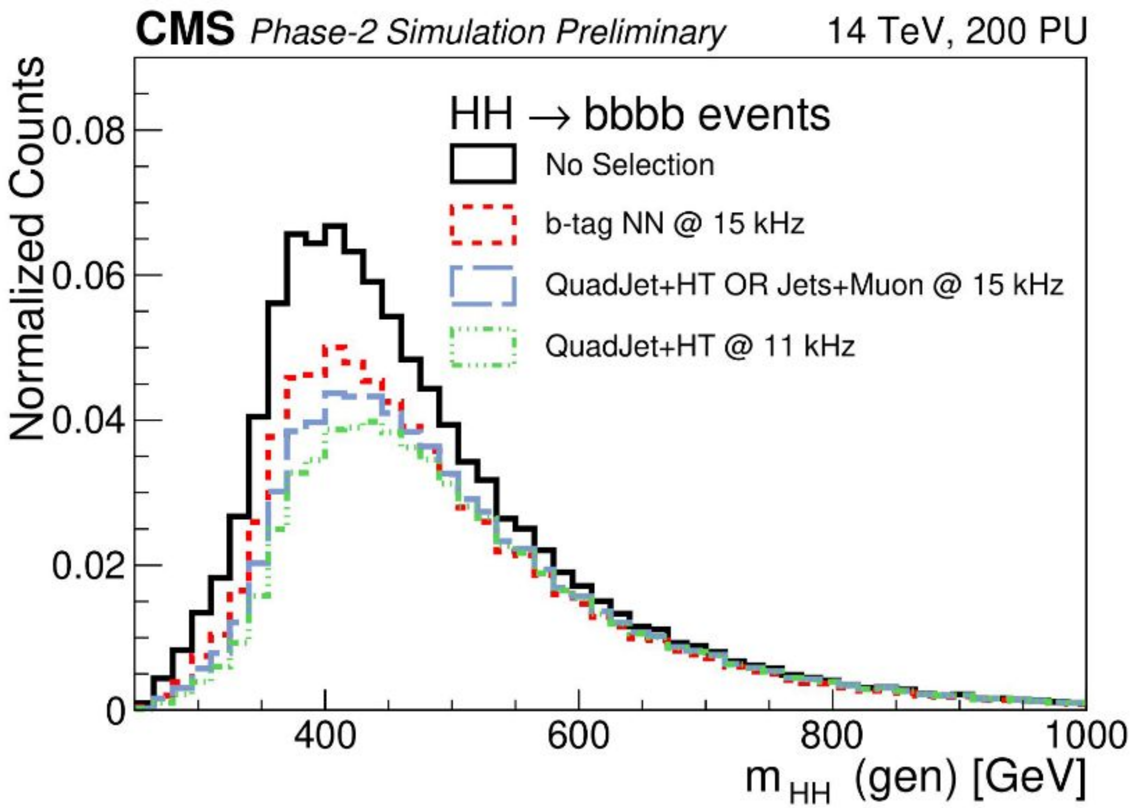
2029-2040+

- Task becomes much harder during HL-LHC due to increased Pile-Up
- **Advances in FPGA technology** facilitates ML in the ultra-low latency, high-bandwidth environment
  - Conifer hls 4 ml convert python ML model to FPGA language



Field Programmable Gate Array (FPGA)
High parallelism, high flexibility, latency deterministic, power efficient





Example: CNN to identify b-quark jets in μs domain

The inverse problem

Likelihood is integral over all possible trajectories through latent space

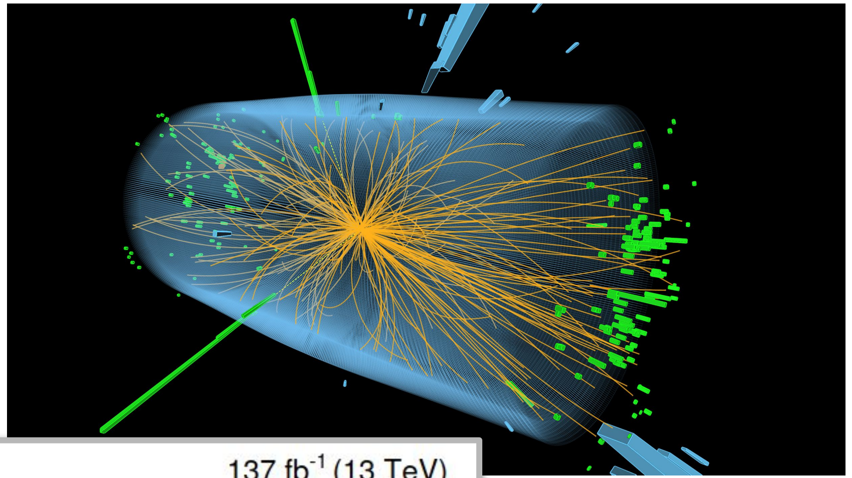
$$p(x|\theta) = \int \mathrm{d}z \, p(x, z|\theta)$$

Reconstructed four-momenta + ID of all final state particles

- Determine underlying parameters, θ, that produce observed data, x
- MC simulation: accurate density estimation in high-dim space is extremely challenging!
- Typically construct lower-dimensional summary statistic

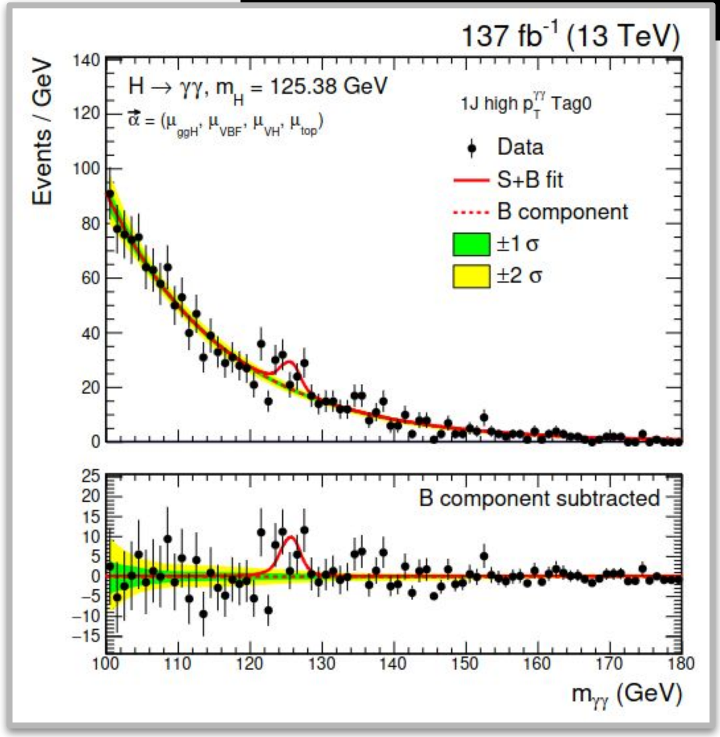
$$D \subseteq \mathbb{R}^{n \times d} \longrightarrow \mathbb{R}^{h \times k}$$

Histogram of h bins in k dimensions



- Construct Poisson-likelihood using summary statistic to infer, $\theta = \{\mu, \nu\}$
  - e.g. to extract signal rate, μ, with nuisance parameters, ν

$$\boxed{L(\mathrm{data}|\mu,\nu) = \left(\prod_r \mathrm{Pois}\left[N_r|\mu s(\nu) + b(\nu)\right]\right) \cdot \mathcal{C}(\nu)}$$



- Where can ML improve inference over traditional methods?