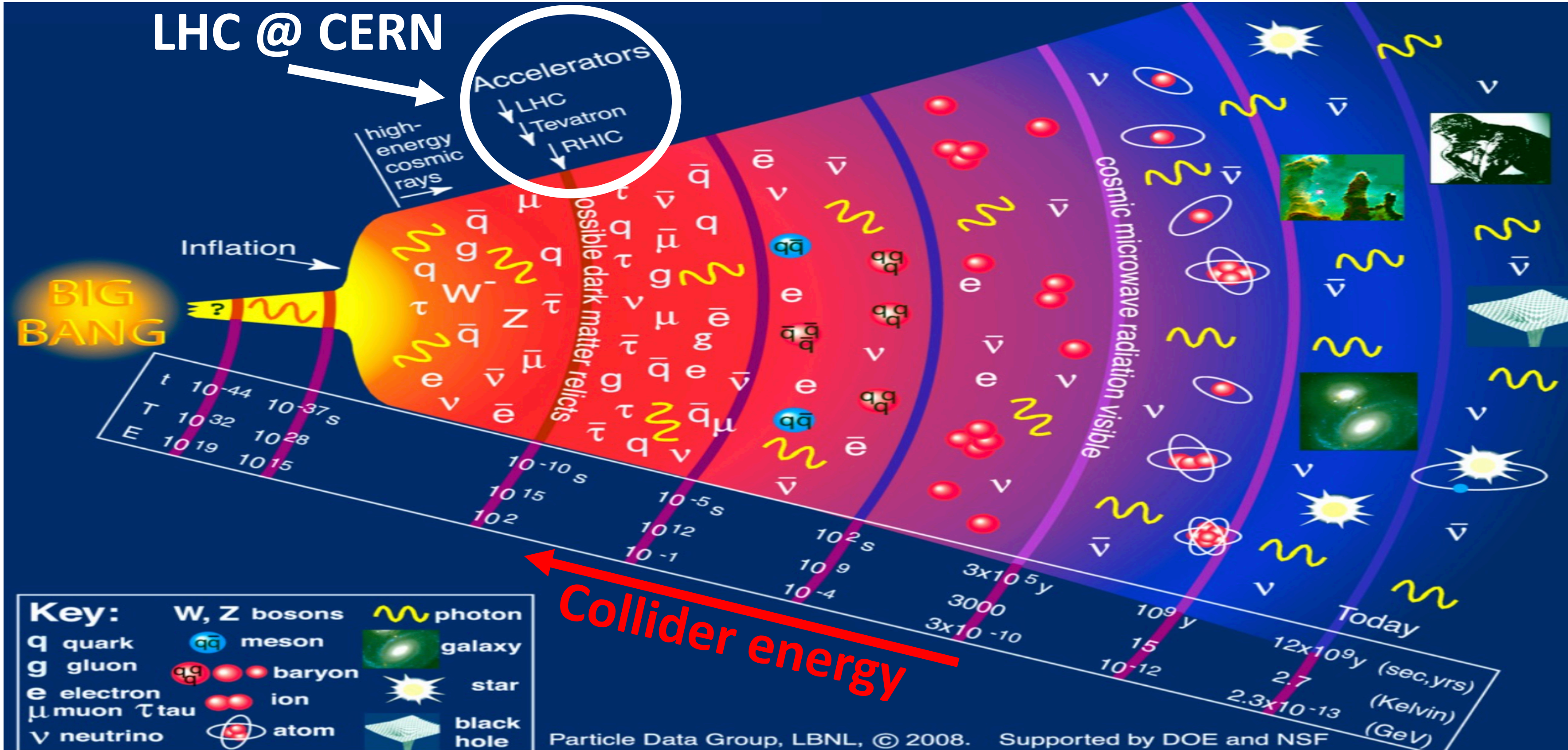
Highlights of ML in particle physics @LHC [for non-physicists]

Loukas Gouskos
CERN/Brown

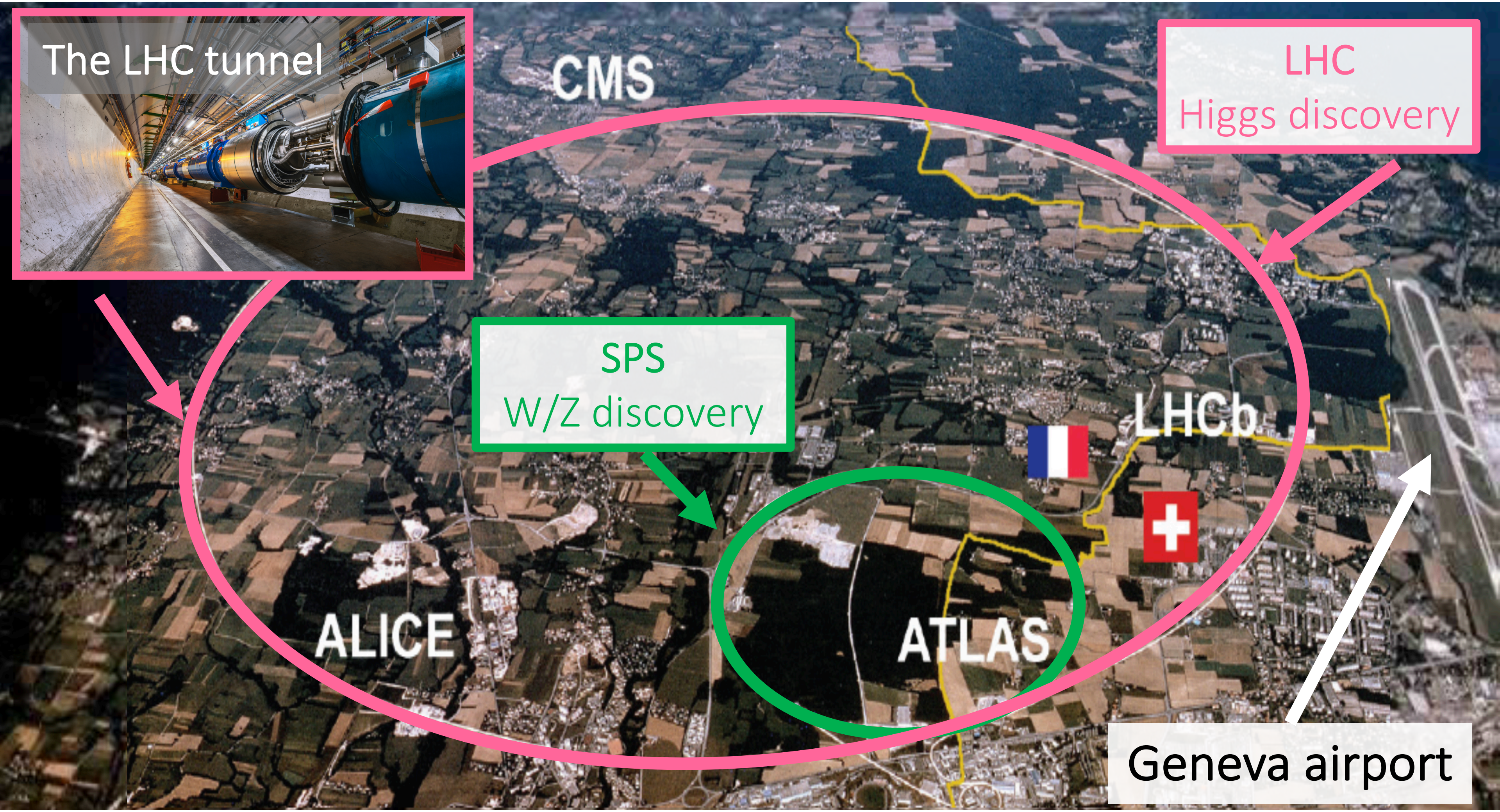
Hammers & Nails 2023

Introduction



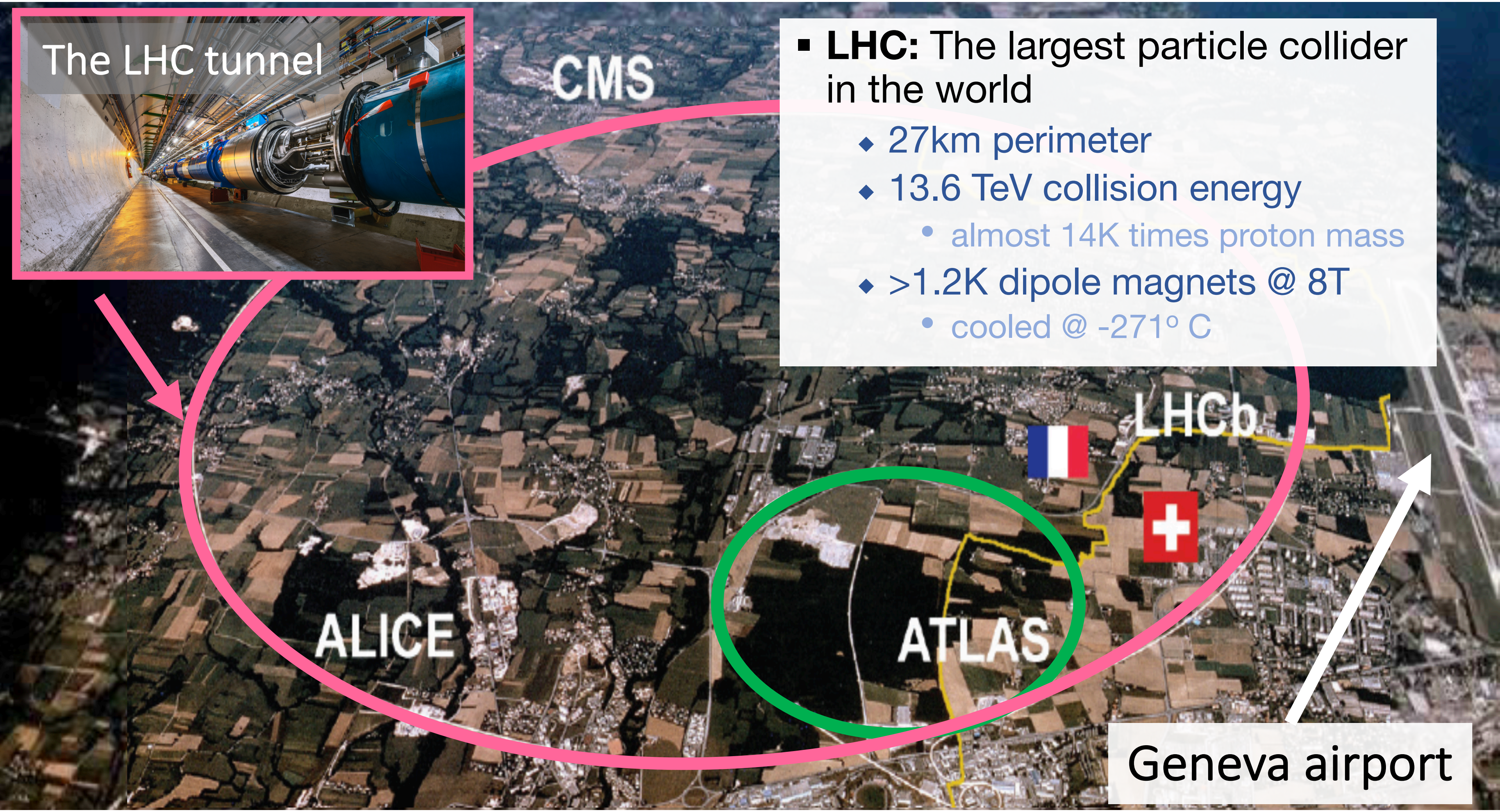
- Particle colliders: a powerful probe of the early universe
  - Study fundamental interactions and search for new

LHC @ CERN

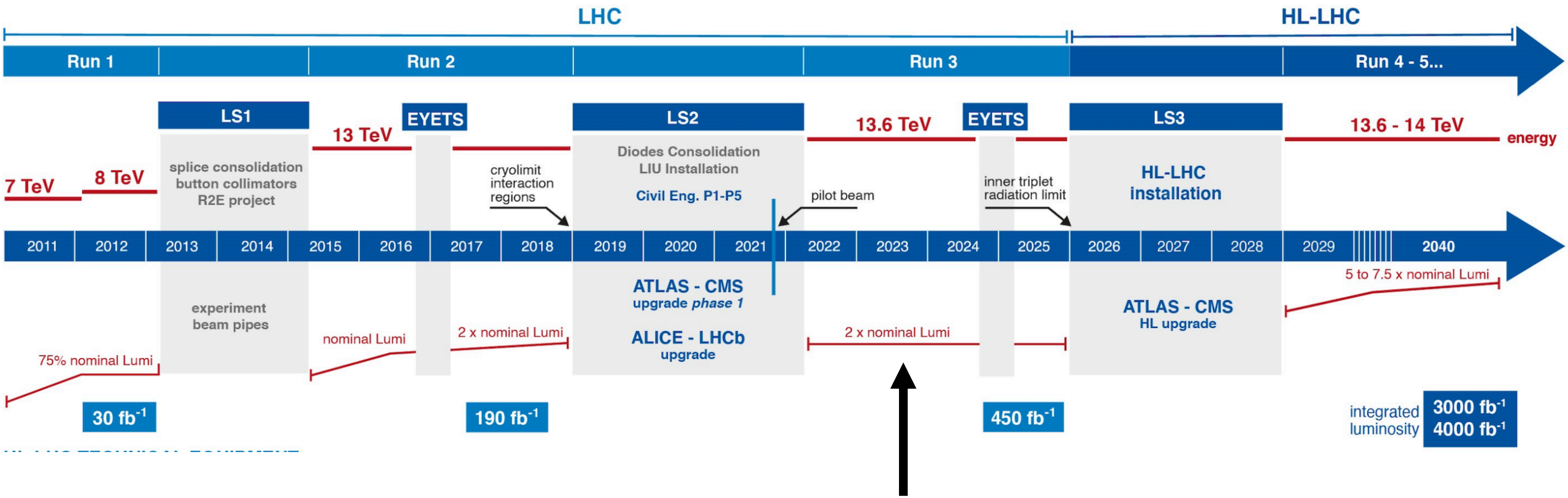

The LHC tunnel
CMS
LHC
Higgs discovery
SPS
W/Z discovery
LHCb
ALICE
ATLAS
Geneva airport

LHC @ CERN

- **LHC:** The largest particle collider in the world
  - 27km perimeter
  - 13.6 TeV collision energy
    - almost 14K times proton mass
  - >1.2K dipole magnets @ 8T
    - cooled @ -271° C



LHC timeline

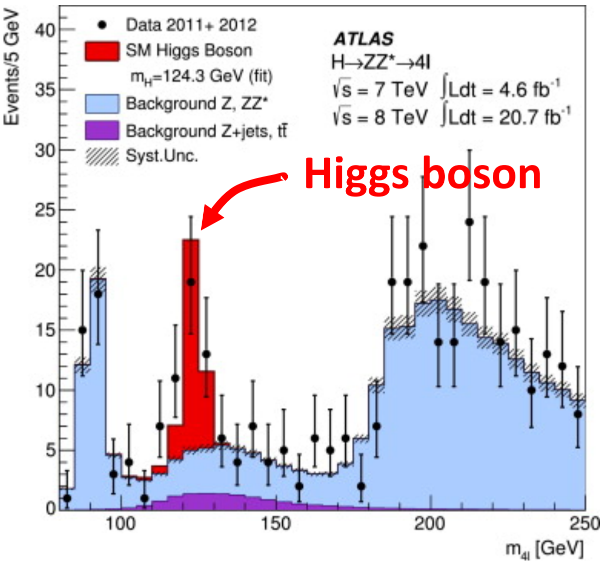


Present

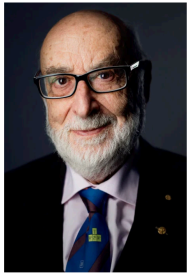
~10y of operation

Executive summary of a decade

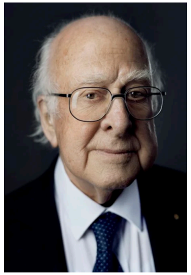
Higgs discovery



Nobel Prize (2013)



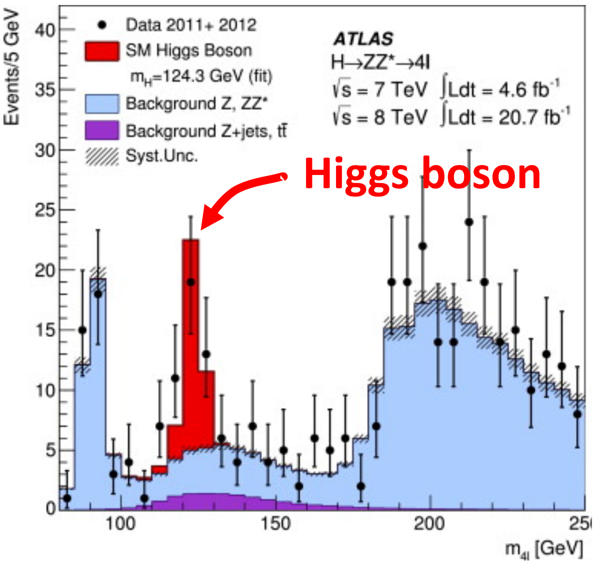
Englert



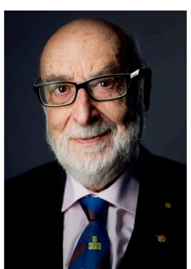
Higgs

Executive summary of a decade

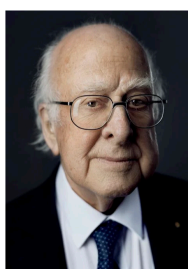
Higgs discovery



Nobel Prize (2013)

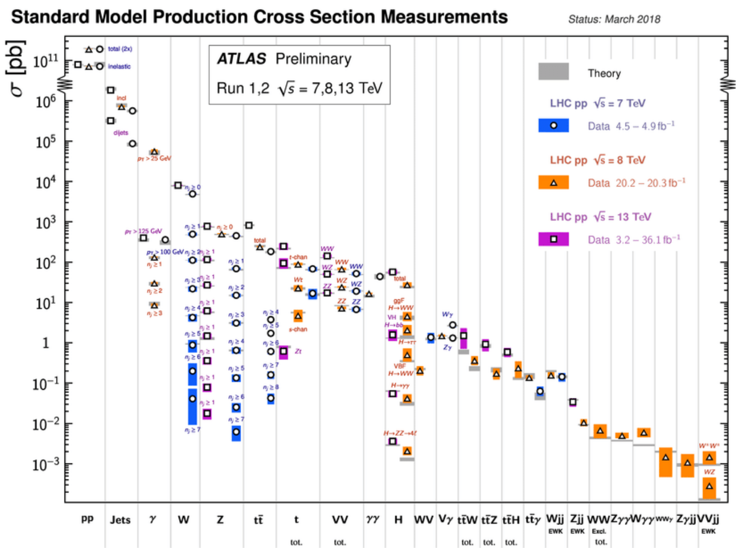


Englert



Higgs

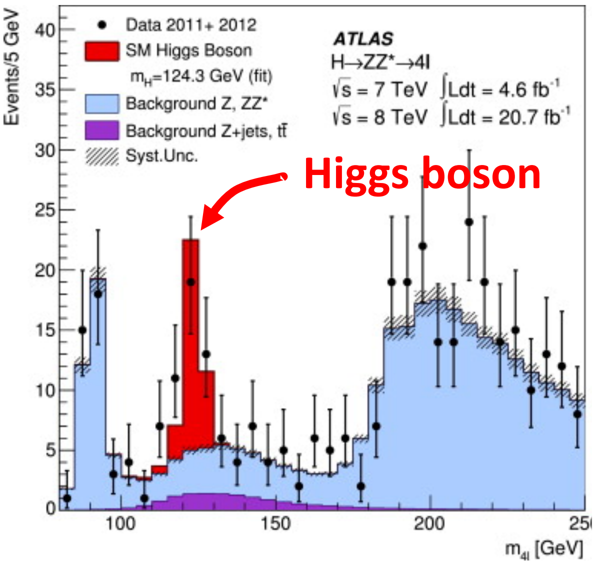
Challenge "Standard Model"



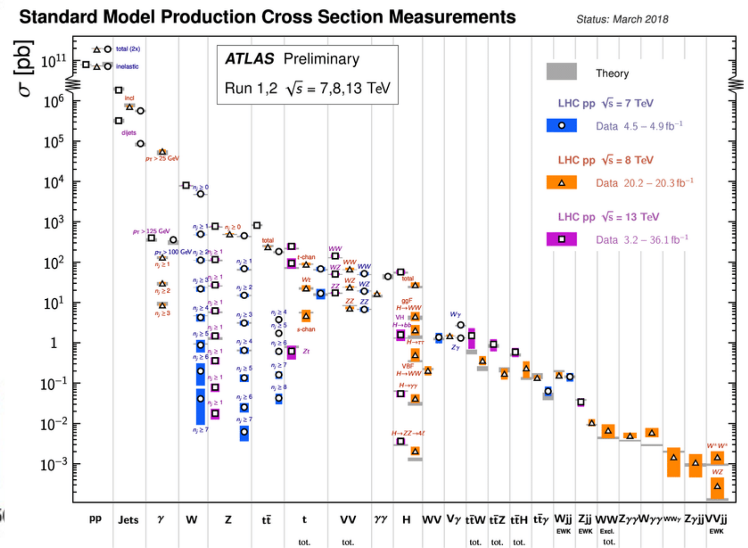
+ plethora of direct searches for New physics signals

Executive summary of a decade

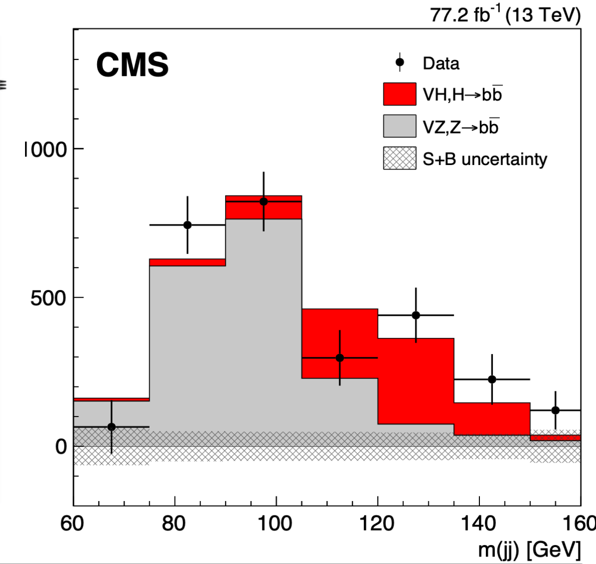
Higgs discovery



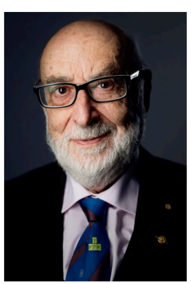
Challenge "Standard Model"



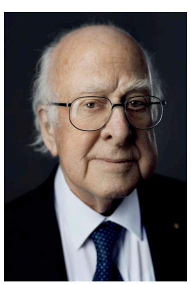
… often exceeding expectations



Nobel Prize (2013)



Englert



Higgs

+ plethora of direct searches for New physics signals

> "In conclusion, the extraction of a signal from H→bb decays in the WH channel will be very difficult at the LHC, even under the most optimistic assumptions for the b-tagging performance and calibration of the shape and magnitude of the various background sources from the data itself."

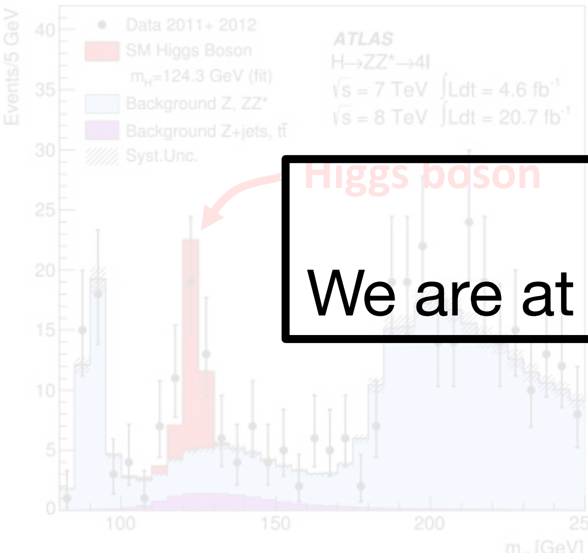
*ATLAS/CMS Technical Design report (1999)

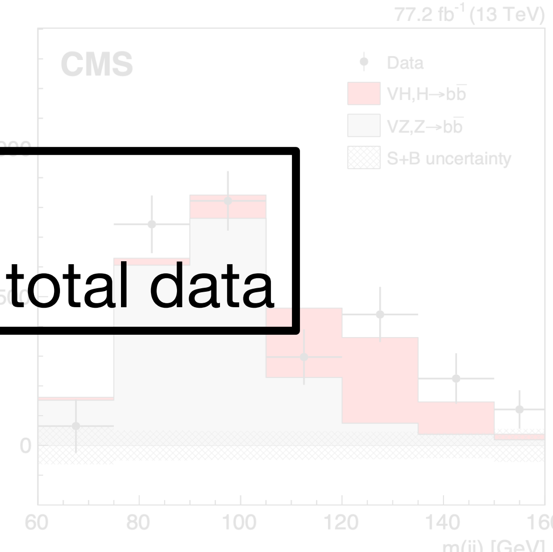
Executive summary of a decade

Higgs discovery

Challenge "Standard Model"

… often exceeding expectations





Impressive results

We are at the beginning: just ~5% of total data

Nobel Prize (2013)

Englert

Higgs

+ plethora of direct searches for New physics signals

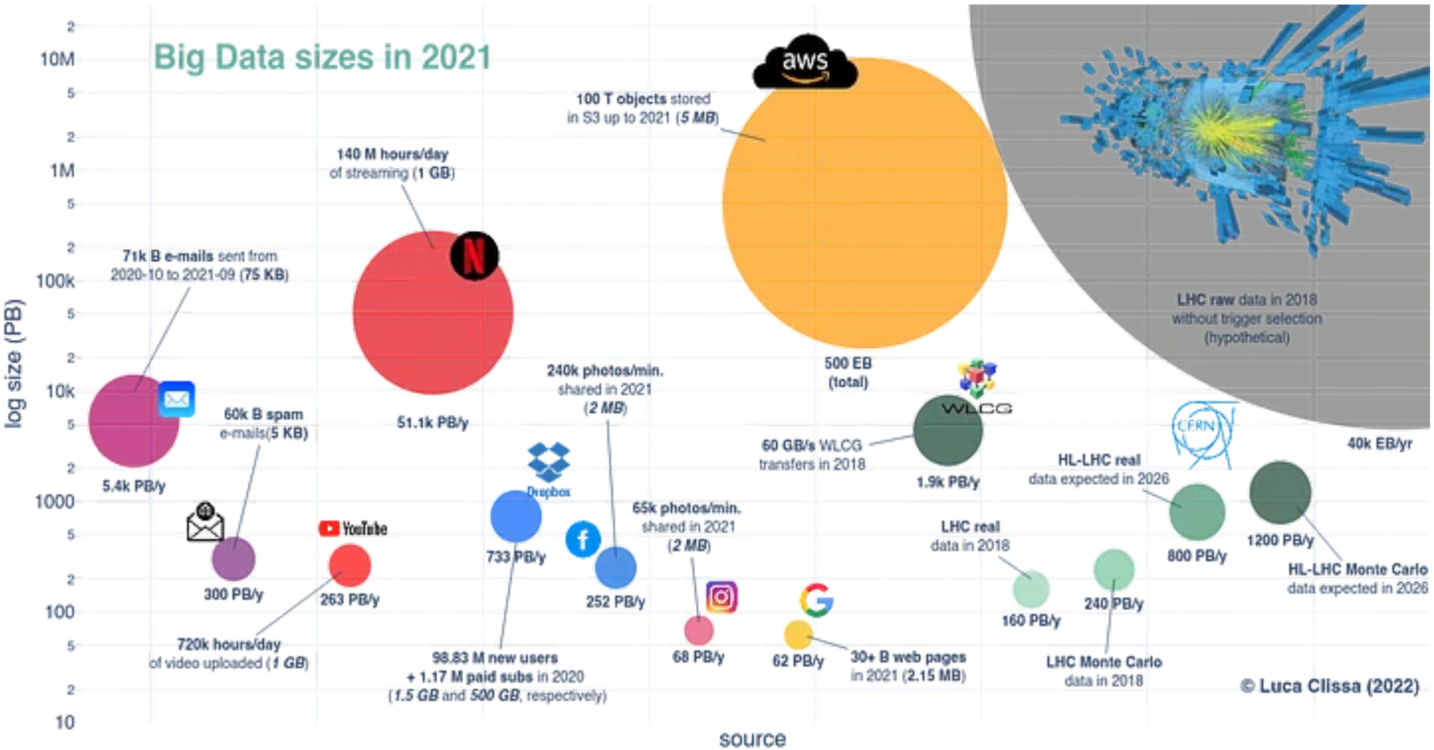
"In conclusion, the extraction of a signal from H→bb decays in the WH channel will be very difficult at the LHC, even under the most optimistic assumptions for the b-tagging performance and calibration of the shape and magnitude of the various background sources from the data itself."

*ATLAS/CMS Technical Design report (1999)

Road full of challenges

Road full of challenges

- Big Data:
  - proton-proton collisions every 25 ns; 40M collisions/sec
  - Event size: O (1-2MB)/collision

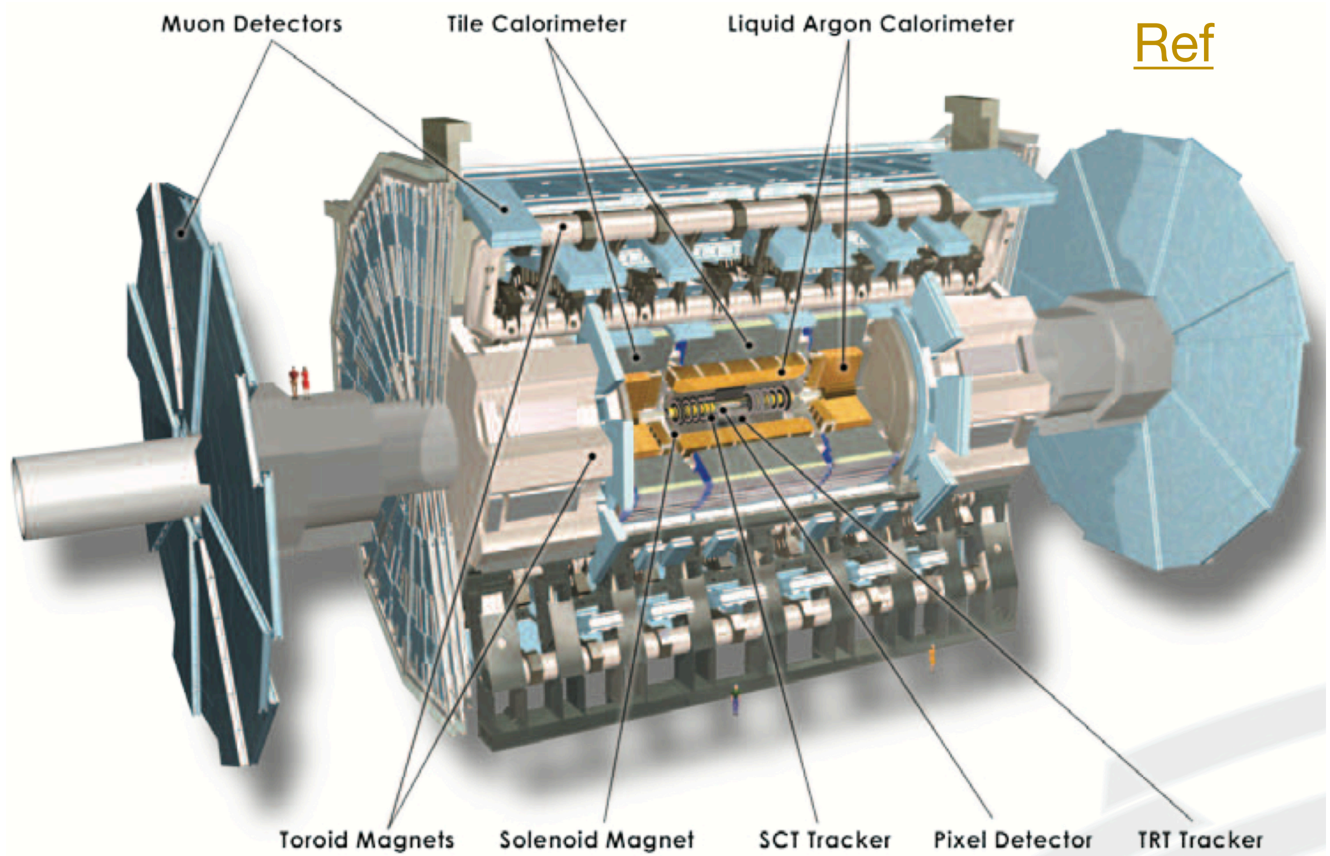


Learn how to process them

Road full of challenges (II)

- Complex & heterogeneous detectors

Ref

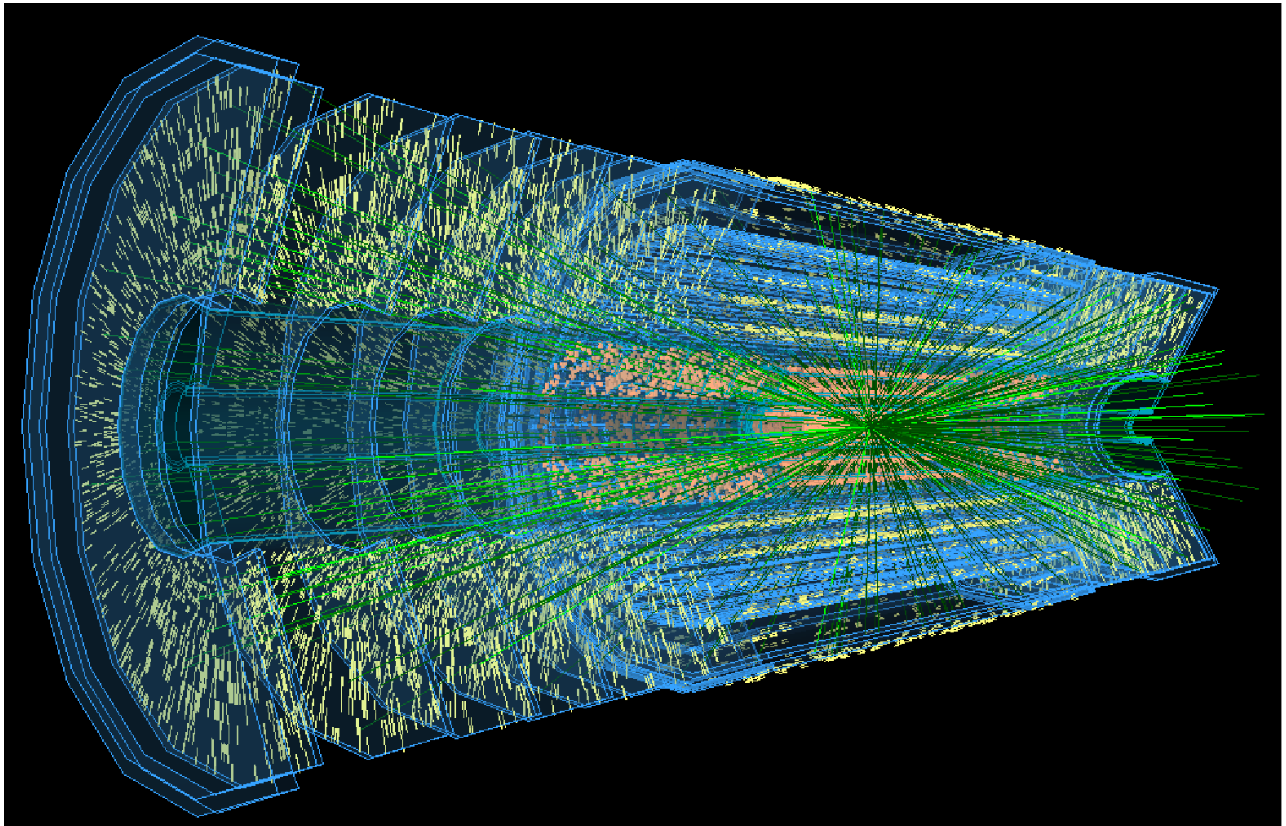


- Diameter 25 m -- Length : 46 m
- Barrel toroid length 26 m
- Overall weight 7 000 tonnes
- ~ 100 million electronic channels
- ~ 3 000 km of cables

Learn how to operate them and monitor them

Road full of challenges (III)

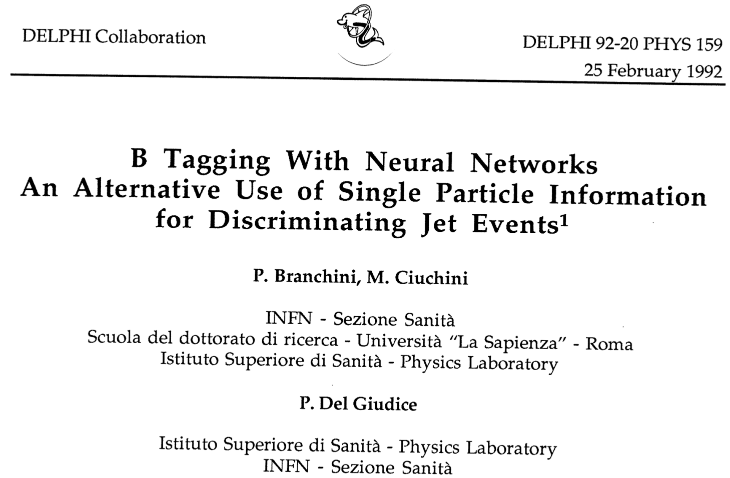
- Event reconstruction and simulation
- O(1K) particles/collision; LHC Run 3: ~O(60) collisions/event



Learn how to reconstruct and [efficiently] simulate events

How ML fits in all this

- ML it's in the game since 90s
  - classification and regression tasks using simple ML
- Modern experiments
  - Much more data [efficient training]
  - Much more channels [dimensionality]
  - Obviously: Computing resources



DELPHI Collaboration

DELPHI 92-20 PHYS 159
25 February 1992

B Tagging With Neural Networks
An Alternative Use of Single Particle Information for Discriminating Jet Events[1]

P. Branchini, M. Ciuchini

INFN - Sezione Sanità
Scuola del dottorato di ricerca - Università "La Sapienza" - Roma
Istituto Superiore di Sanità - Physics Laboratory

P. Del Giudice

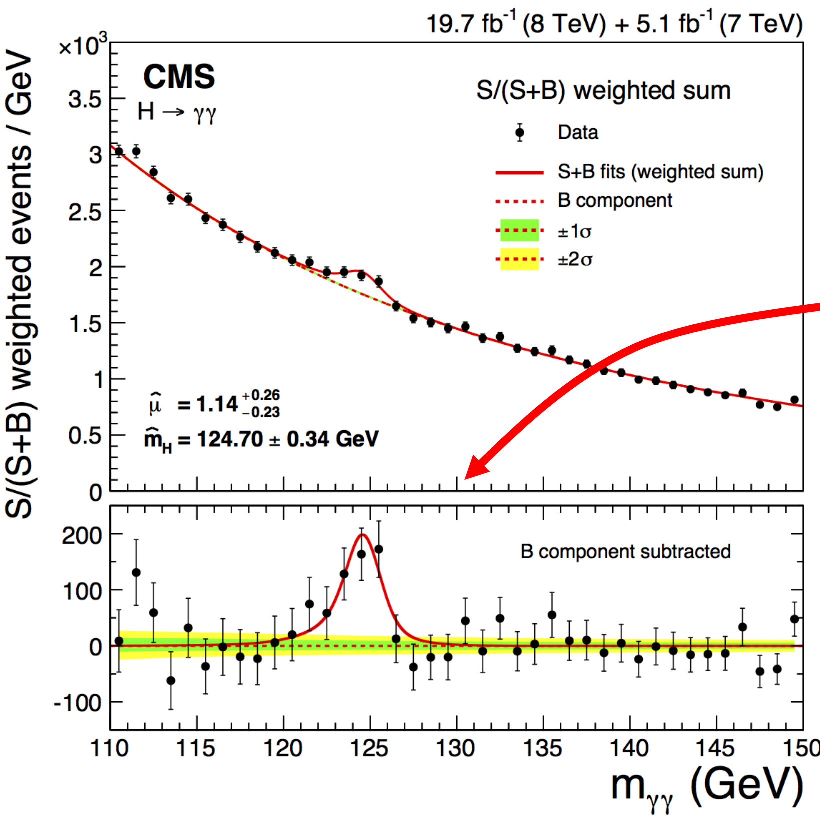
Istituto Superiore di Sanità - Physics Laboratory
INFN - Sezione Sanità

ML: effectively **analyze sparse data** and **identify correlations** within the dataset

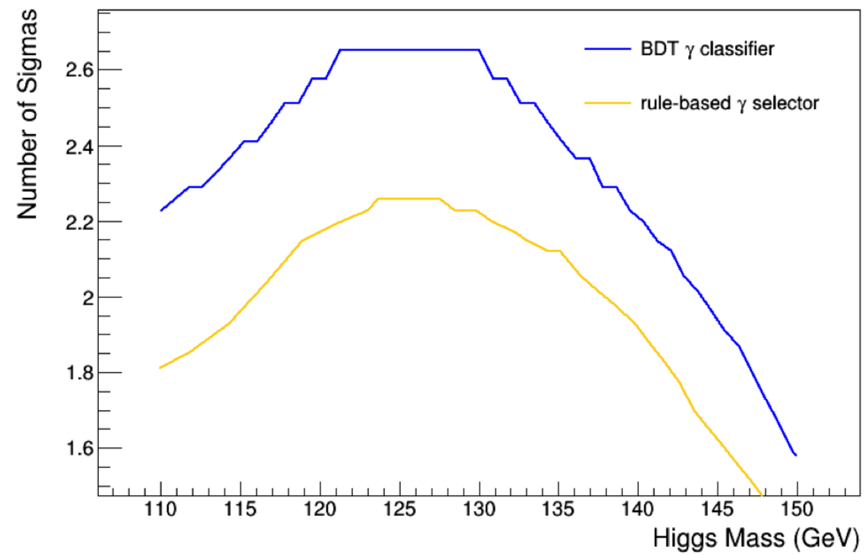
- From an experimentalists perspective
  - differences wrt to particle physics and "every-day" life
    - Particle physicists often focus on tails → Amazon/Google et. all on typical user

Example from early LHC

H→γγ observation



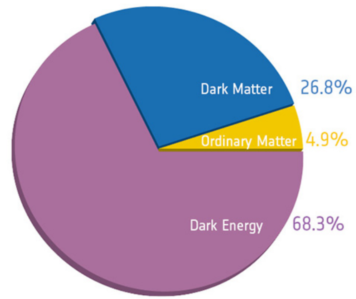
Improvement in discovery significance

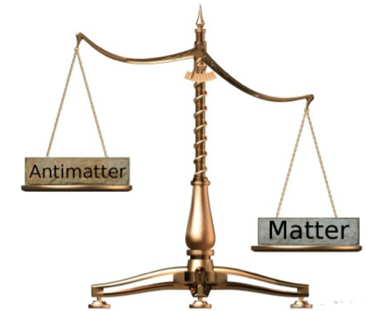


Reach similar sensitivity with ~50% less data

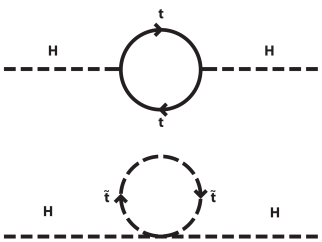
Far from done

- Many big open questions that beg for **new physics**
  - Experiment – driven
    - Dark Matter
    - Dark energy
    - Matter-Antimatter asymmetry
    - …



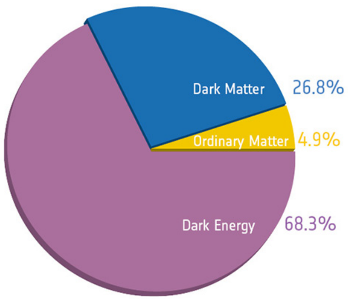


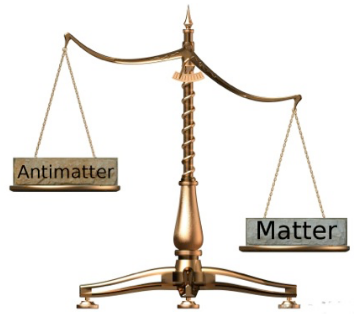
  - Theory – driven
    - Hierarchy problem & naturalness
    - Number of generations
    - Origin of fermion families
    - …



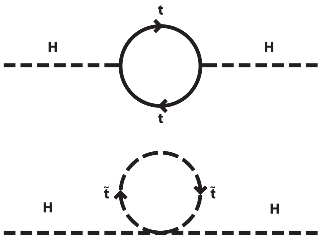
Far from done

- Many big open questions that beg for **new physics**
  - Experiment – driven
    - Dark Matter
    - Dark energy
    - Matter-Antimatter asymmetry
    - …





  - Theory – driven
    - Hierarchy problem & naturalness
    - Number of generations,
    - Origin of fermion families
    - …



- Either within the LHC reach [but small rate and/or difficult corners]
  - or beyond LHC reach
    - still: exhaustively exploit LHC physics potential
      → important for future experiments

The name of game

events for a
given process

$$\mathcal{N} = \sigma \times \mathcal{L} \times \mathcal{A} \times \epsilon$$

The name of game

cross-section:
of a given process
[∝ collision energy]

events for a given process

$$\mathcal{N} = \sigma \times \mathcal{L} \times \mathcal{A} \times \epsilon$$

The name of game

cross-section:
of a given process
[∝ collision energy]

events for a given process

$$\mathcal{N} = \sigma \times \mathcal{L} \times \mathcal{A} \times \epsilon$$

Luminosity:
- ~linear increase vs. time
- sensitivity ~ sqrt{L} ☹

The name of game

Acceptance:
- improvement mainly from new/upgraded detectors
- ML: important for the design of new detectors

cross-section:
of a given process
[∝ collision energy]

\# events for a given process

$$\mathcal{N} = \sigma \times \mathcal{L} \times \mathcal{A} \times \epsilon$$

Luminosity:
- ~linear increase vs. time
- sensitivity ~ sqrt{L} ☹

The name of game

Acceptance:
- improvement mainly from new/upgraded detectors
- ML: important for the design of new detectors

cross-section:
of a given process
[∝ collision energy]

events for a given process

$$\mathcal{N} = \sigma \times \mathcal{L} \times \mathcal{A} \times \epsilon$$

Efficiency:
- Trigger, Phys. Object reco., Sig-vs-Bks, …
- ML is revolutionizing this front

Luminosity:
- ~linear increase vs. time
- sensitivity ~ sqrt{L} ☹

The name of game

Acceptance:
- improvement mainly from new/upgraded detectors
- ML: important for the design of new detectors

cross-section:
of a given process
[∝ collision energy]

events for a given process

$$\mathcal{N} = \boldsymbol{\sigma} \times \mathcal{L} \times \mathcal{A} \times \boldsymbol{\epsilon}$$

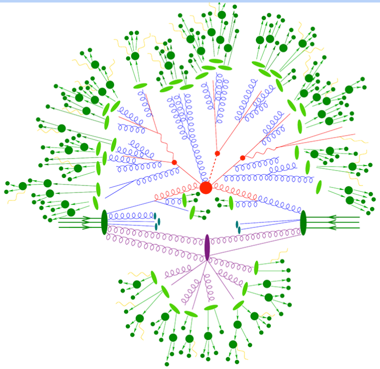
Luminosity:
- ~linear increase vs. time
- sensitivity ~ sqrt{L} ☹

Efficiency:
- Trigger, Phys. Object reco., Sig-vs-Bks, …
- ML is revolutionizing this front

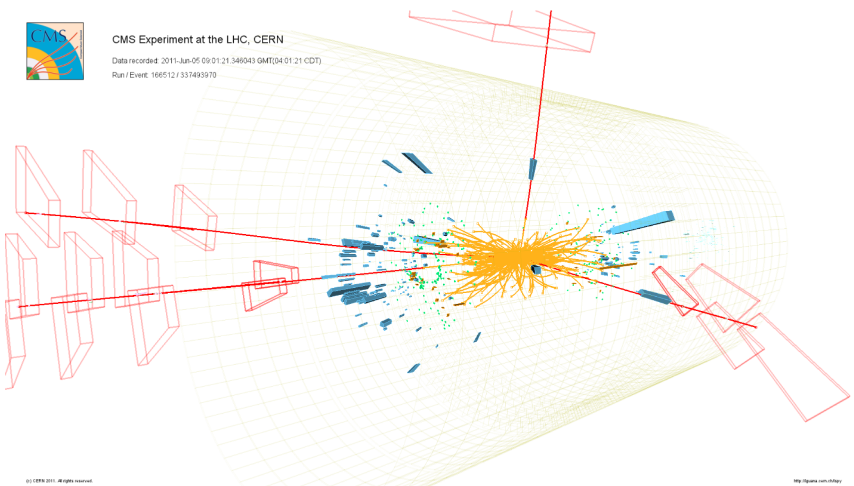
Extract as much physics out as possible
[i.e. maximize N]

Workflow: A multi-step approach

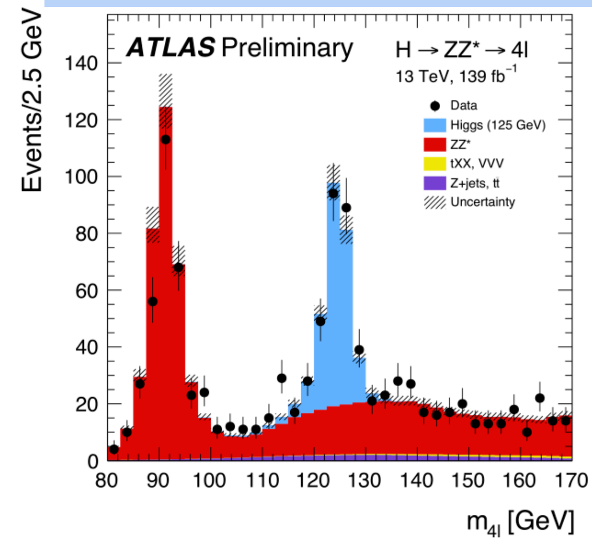
proton-proton collisions



detection, reconstruction, triggering, quality monitoring, ..



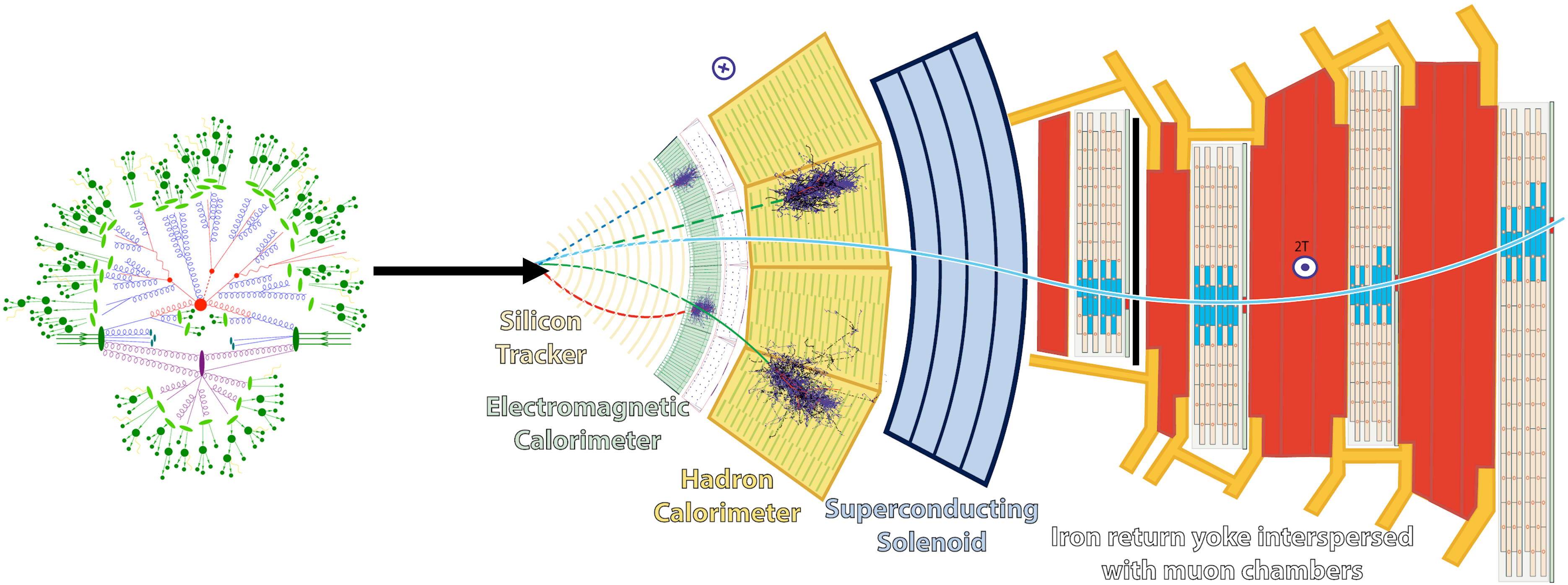
calibration, event selection, signal extraction, interpretation, …



DL-based algorithms became key ingredients in almost all parts of the chain [but it took time..]

Particle detection

- Different detectors ["colors"] designed to detect different particle species

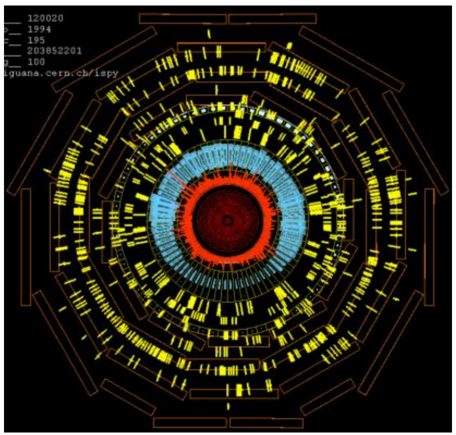


Particle Reconstruction

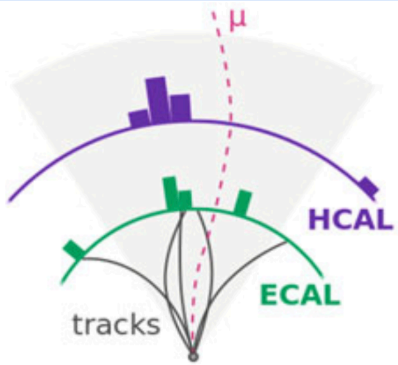
- **Particle/event reconstruction:** interpret detector signals to determine the physical process at collision

lower → **higher**

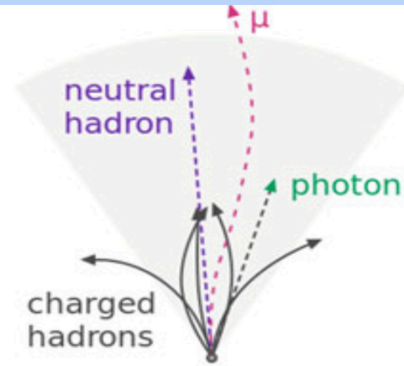
What we see:
Detector signals



Cluster signals
in each sub-detector
[Local RECO]

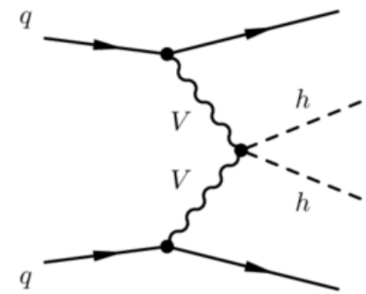


Combine info from
different detectors
[PF particles]



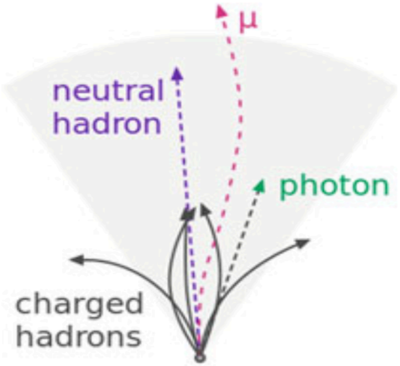
Output: γ, e, μ, charged & neutral hadrons

Target:
Physics event
@collision



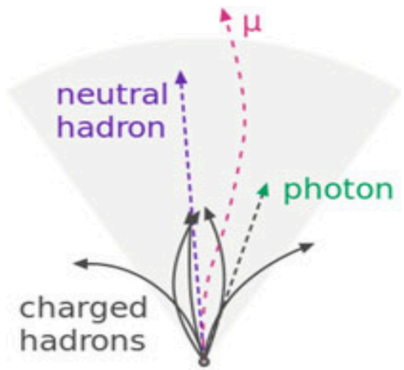
Not much ML until recently

Physics object reconstruction

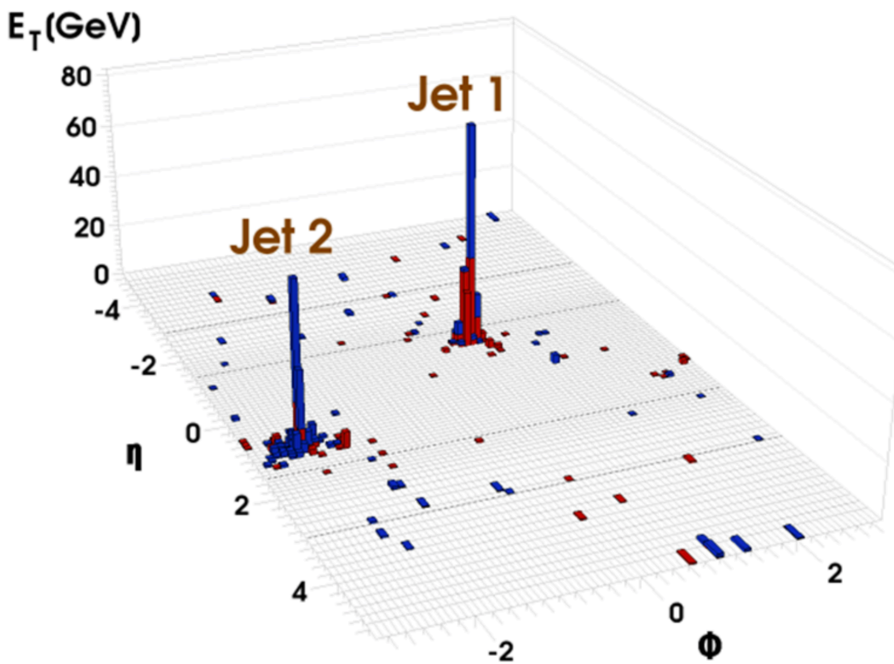


- More complicated objects are build from reconstructed particles

Physics object reconstruction

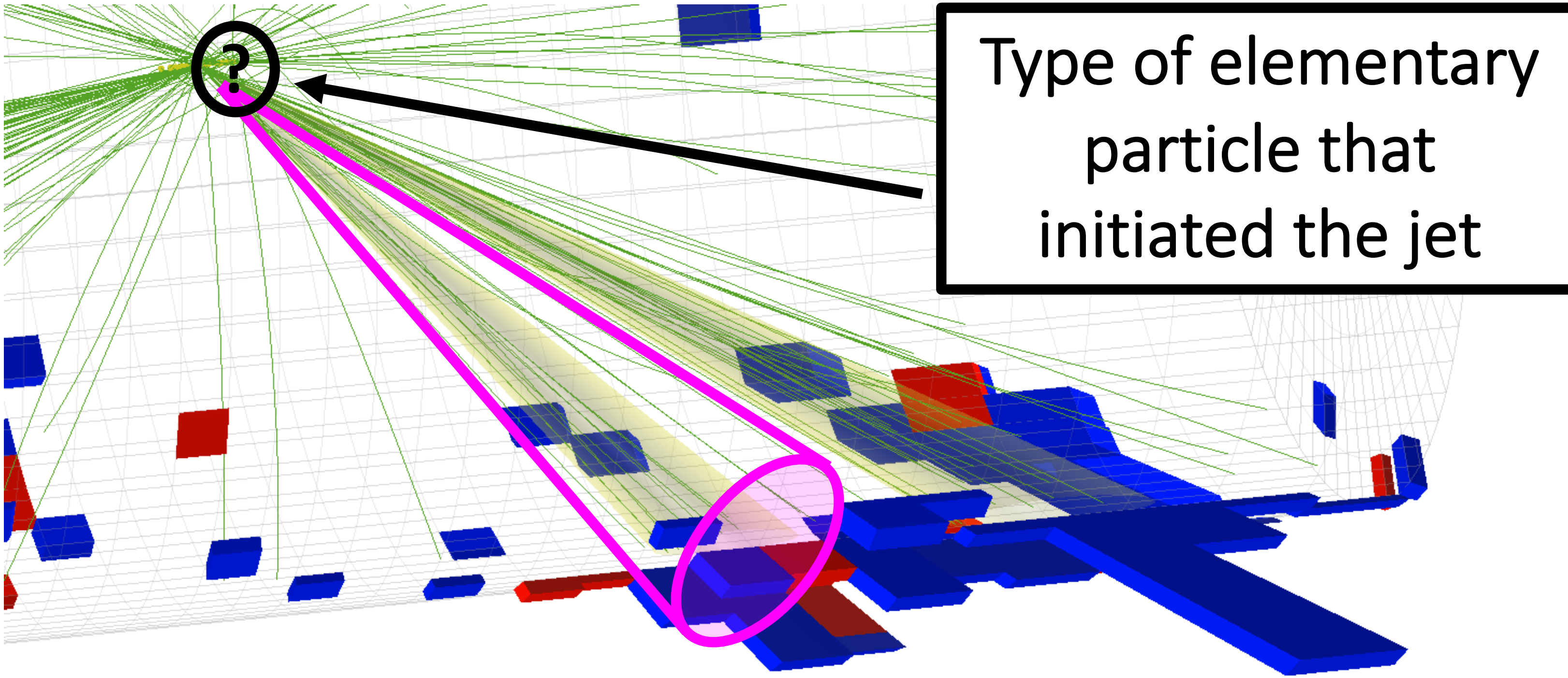


- More complicated objects are build from reconstructed particles



- **Jets:** copiously produced @ LHC
  - critical for precise description of the event
- Formed using physics-inspired rule-based algorithms

Jet identification: beginning of DL@LHC



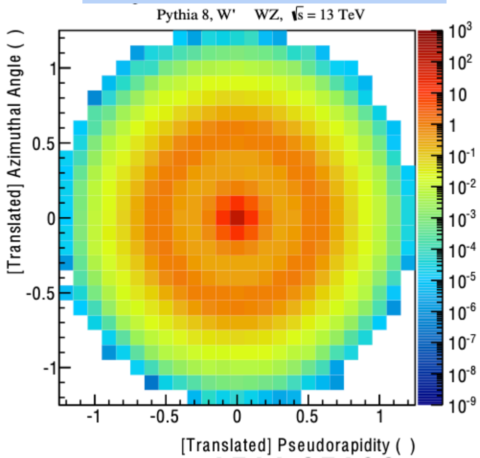
- Jets: complex objects, unordered sets, strong correlations b/w particles, sparse..
  - **Natural playground for DL-based algorithms**
- Key: Jet **representation** and DL **architecture**

Facts:
- O(50-100) particles per jet
- O (50) features per particle

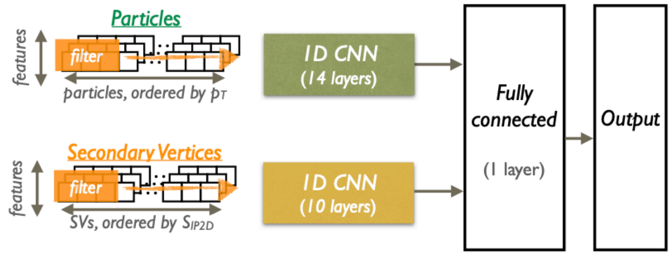
~O(1000) inputs/jet

A topic of high interest [EXP and TH]

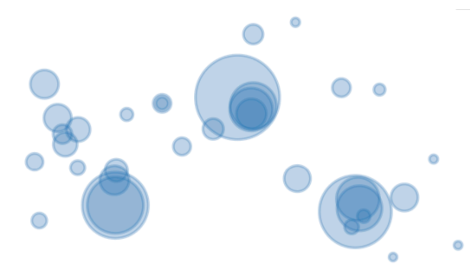
As images



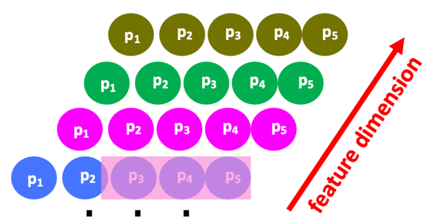
As lists of particles



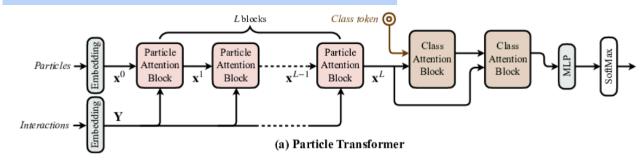
As point clouds



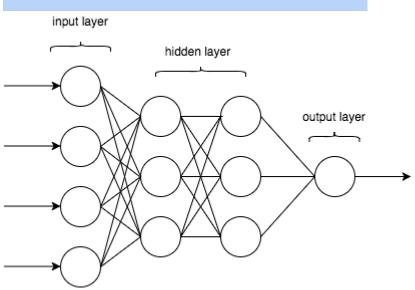
CNN-1D



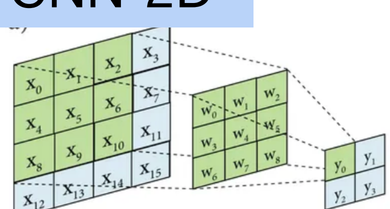
Transformers



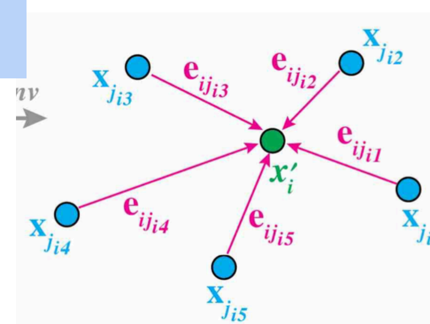
Dense NN



CNN-2D

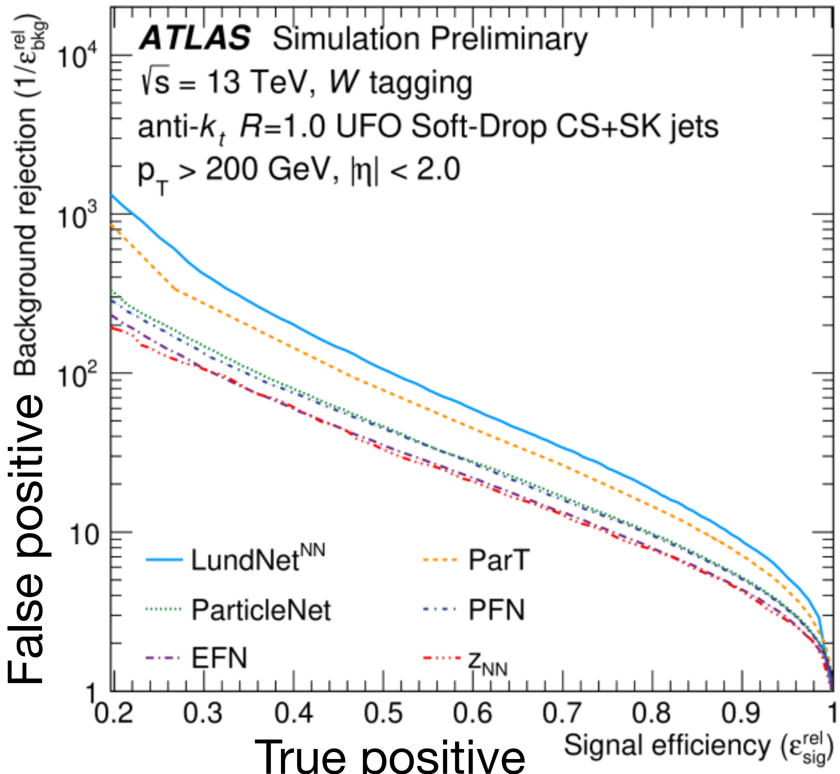


GNN

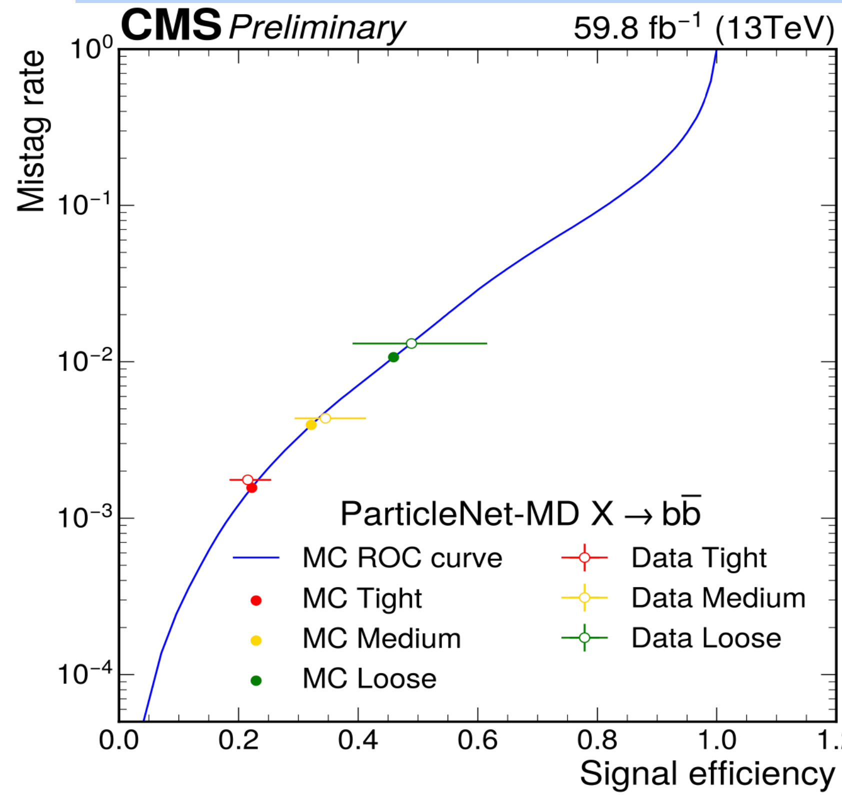


DL in particle reconstruction

Trained in simulation



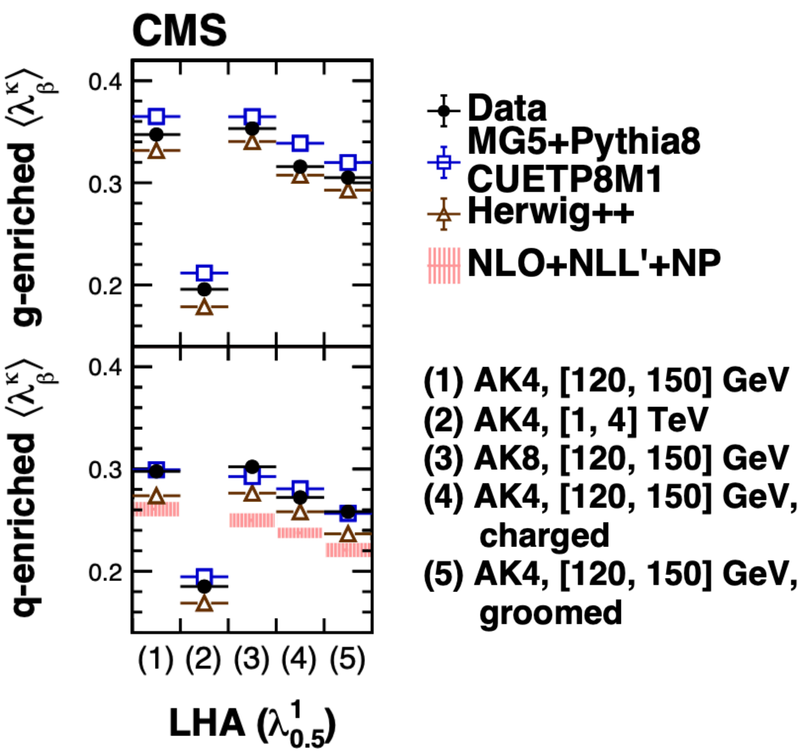
Crucial:
Performance in collision data

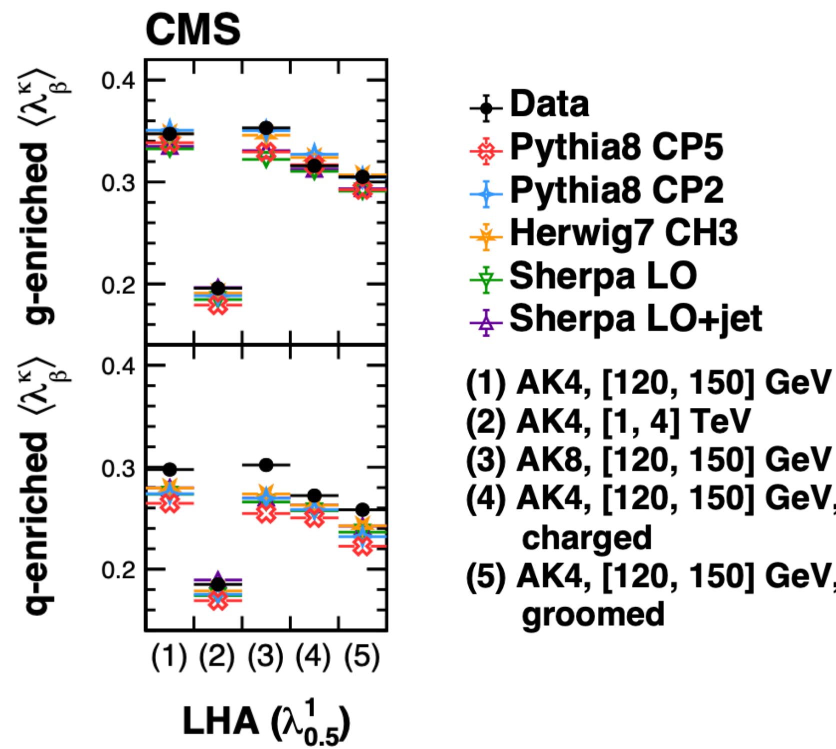


- Very good agreement b/w Data and simulation

Beyond ROC curves..

- Important: What the algorithm learns?
  - Improve physics knowledge
    → better agreement w/data, reduced syst uncertainties
  - → improve physics reach



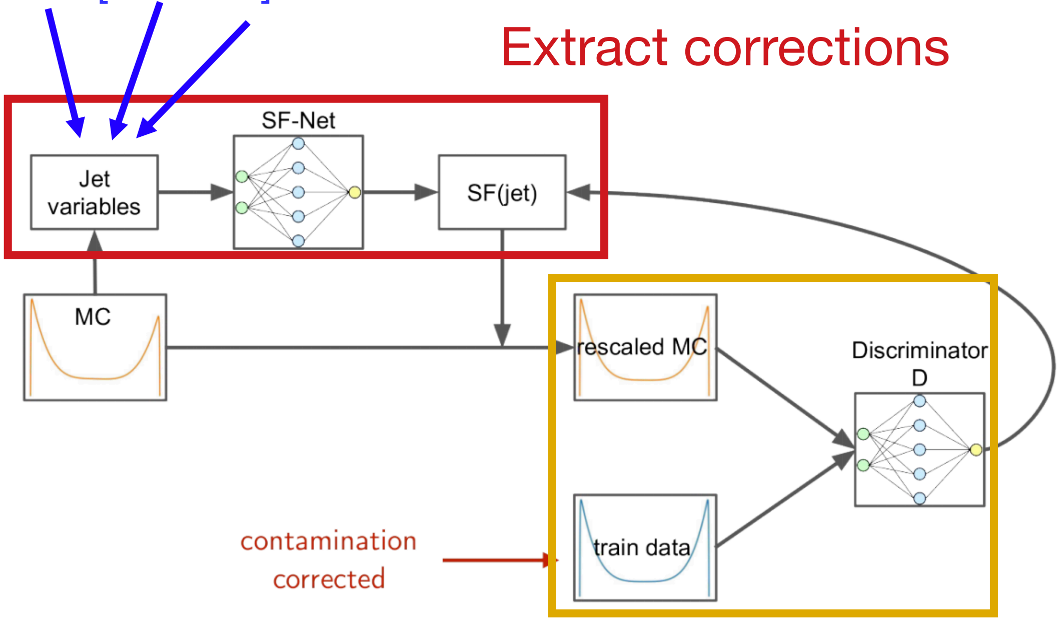


DL in object calibration

- Improve calibration strategy using DL
    - improve **corrections** to minimize **data-simulation differences**

Particle properties [features]

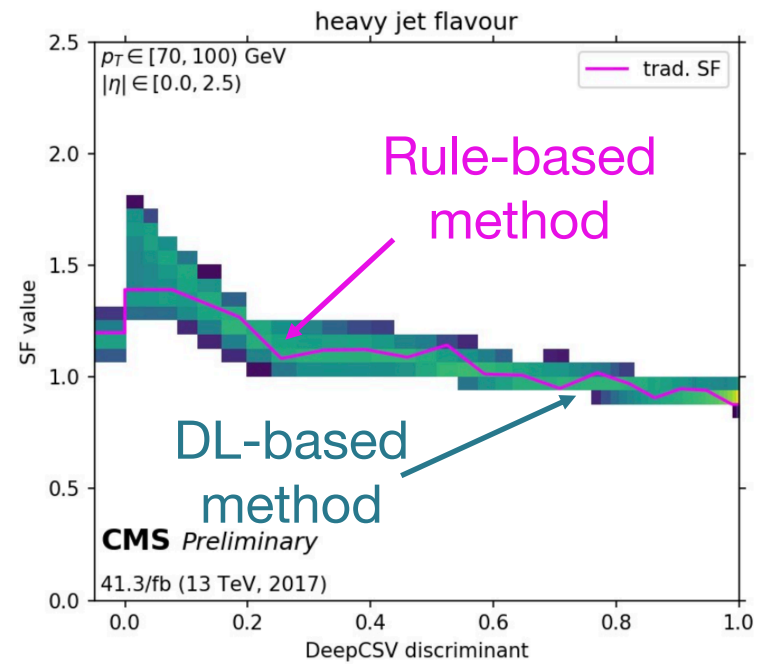
Extract corrections



Loss: $MSE\left(SF, \frac{Data}{MC} = \frac{D}{1-D}\right)$

Data-MC Discrimination

Results



DL-method: account for correlations b/w features

Game changer

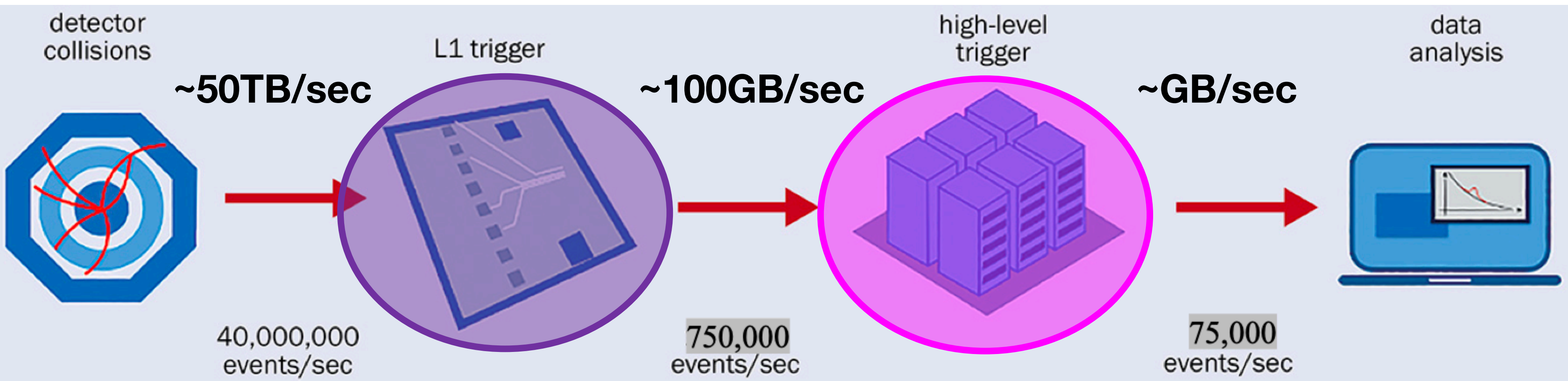
- Demonstrated that they work [and they work well]
  - extend developments to other tasks
    - eg, regression [predict particle energy, mass, …]
  - Opened new physics opportunities

Natural next step: Explore DL in other areas of the experiment

Real-time selection

- 40M events / sec; O(1-2MB)/event → **impossible** to store
  - Real-time filtering [i.e. triggers] → O(10^3) reduction
  - … while keeping “interesting” events

- Factorized approach:



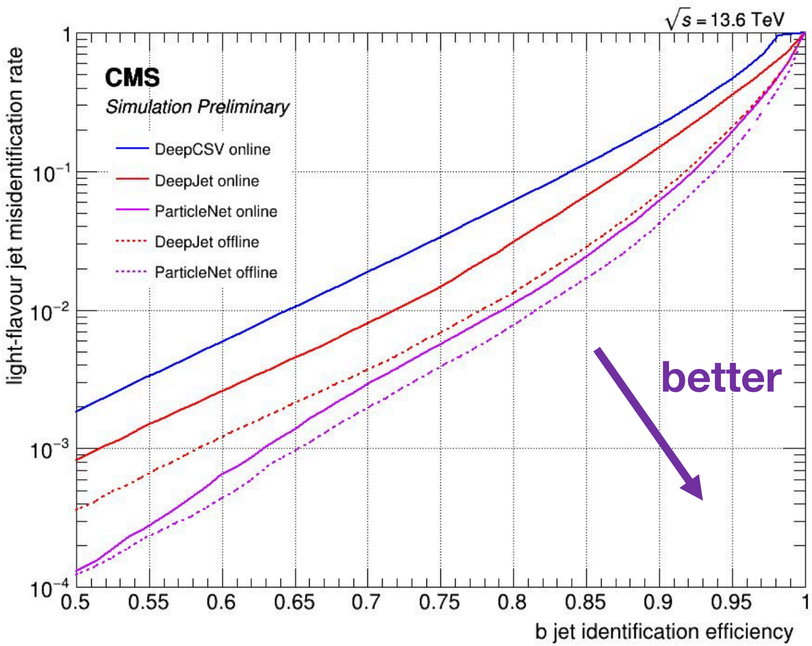
Custom HW/FPGAs
latency < O(μsec)

Commercial CPUs/GPUs
Latency < O(100)msec

Real-time selection: HLT

- Build on DL success in [offline] event reconstruction → Exploit at HLT
  - Similar event information but reduced precision
  - Use same concepts; respect tightest computing constraints

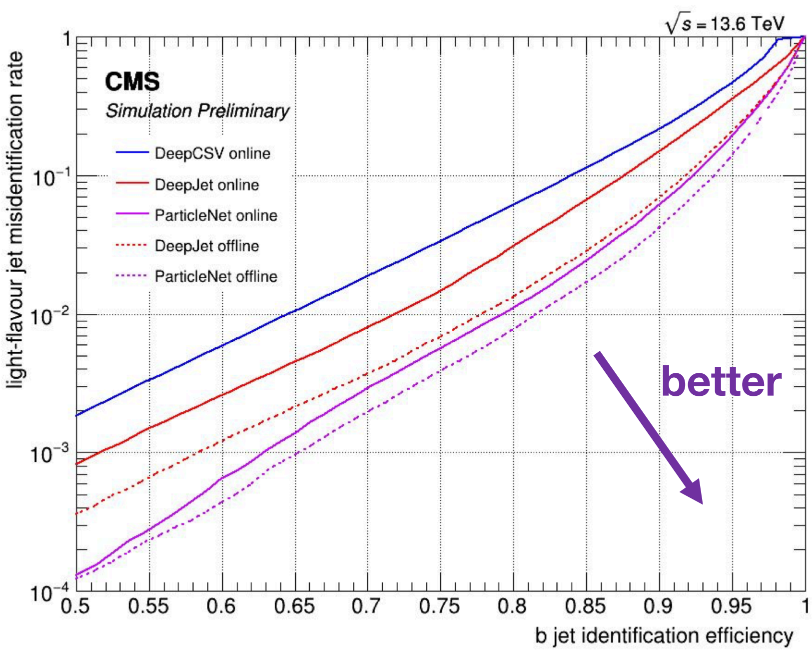
Particle reconstruction using GraphNN @ HLT



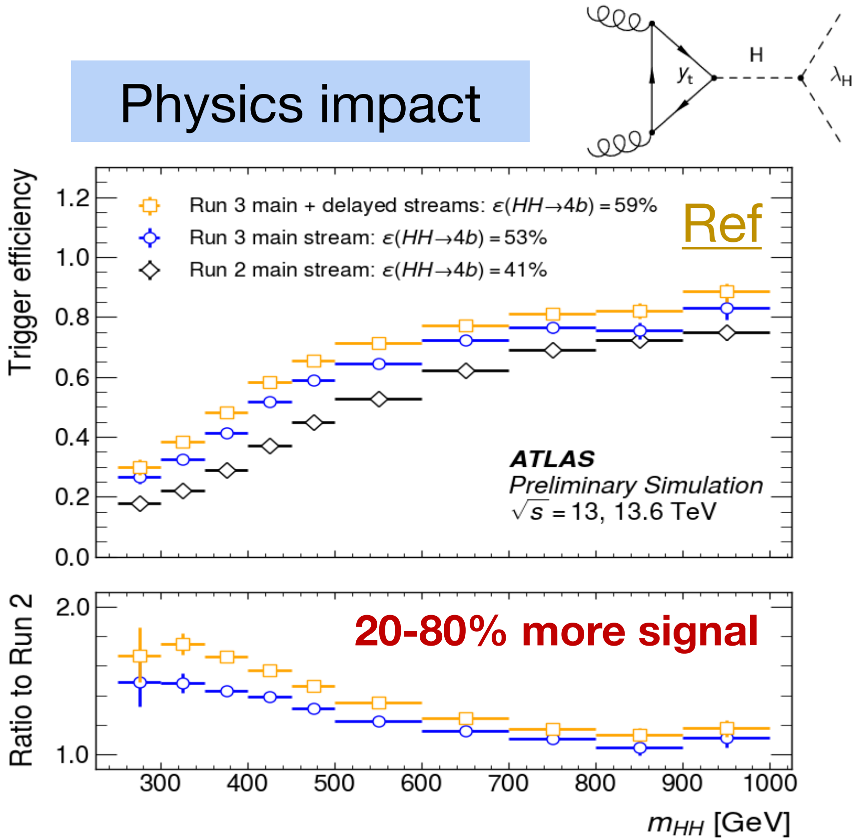
Real-time selection: HLT

- Build on DL success in [offline] event reconstruction → Exploit at HLT
  - Similar event information but reduced precision
  - Use same concepts; respect tightest computing constraints

Particle reconstruction using GraphNN @ HLT



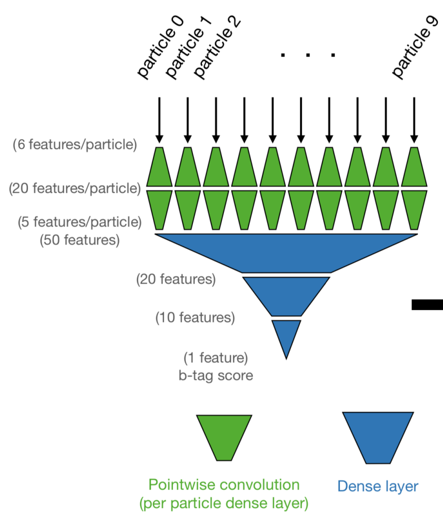
Physics impact

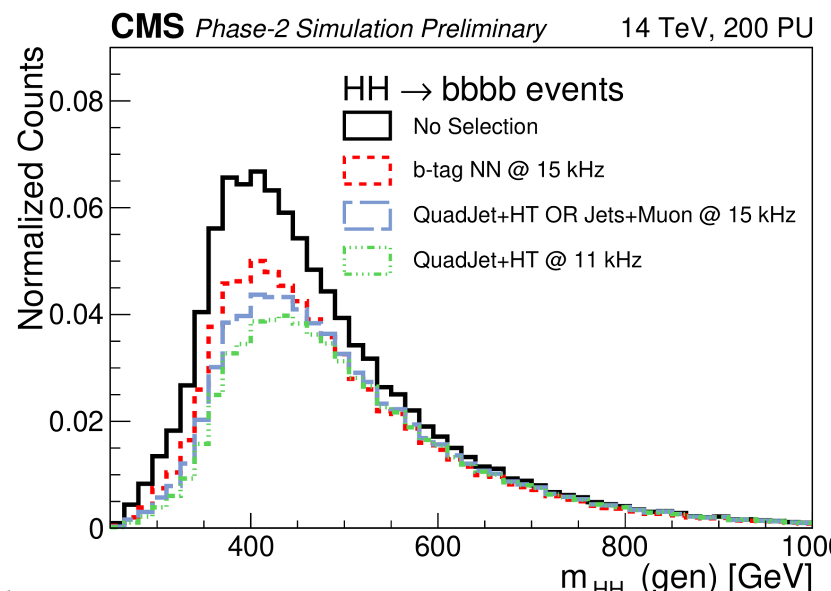


Real-time selection: L1T

- Even more challenging:
  - Less information [only parts of the detector available, coarser granularity]
  - Take decisions < O(1μsec)
  - Limited computing resources/memory
- Shallow ML [i.e. BDTs] exist
  - Next step: more advanced ML?

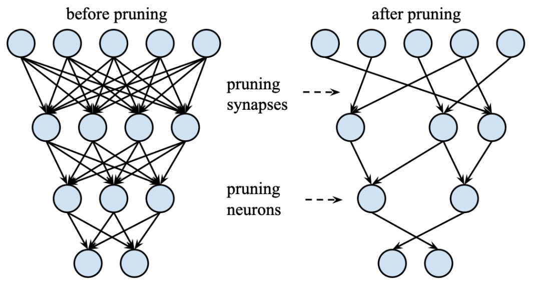
e.g. b-tagging:

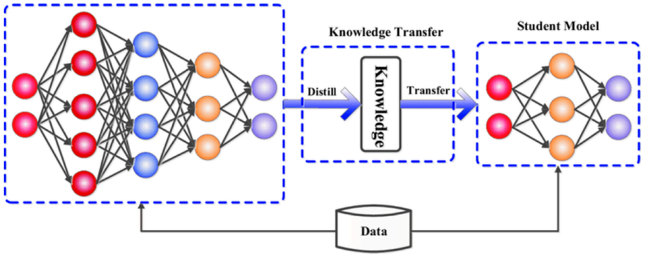




Real-time selection: L1T (II)

- Latency
  - Even very simple DL models cannot fit in FPGAs
  - Defining a compression strategy is a **need**
- Several approaches under exploration
  - **Quantization:**
    - Reduce # bits to represent a number
  - **Pruning:**
    - Drop parameters that do not impact performance
  - **Knowledge distillation:**
    - Transfer "knowledge" from a larger model to a smaller one
  - ....



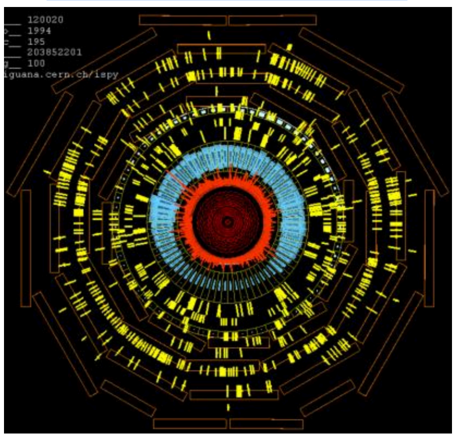


Extend to lower-level tasks

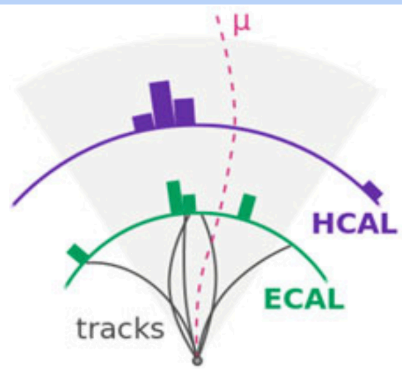
DL for PF Reconstruction

- Improvements in performance translate to improvement in higher-level reco

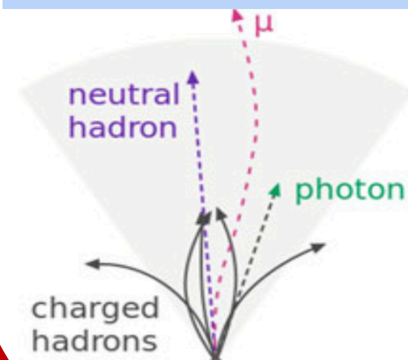
What we see:
Detector signals



Cluster signals
in each sub-detector
[Local RECO]

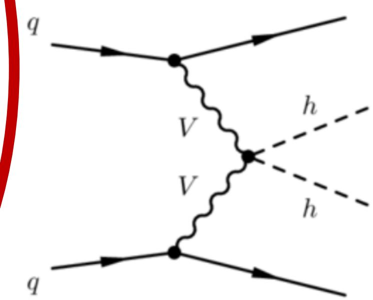


Combine info from
different detectors
[PF particles]



Output: γ, e, μ, charged &
neutral hadrons

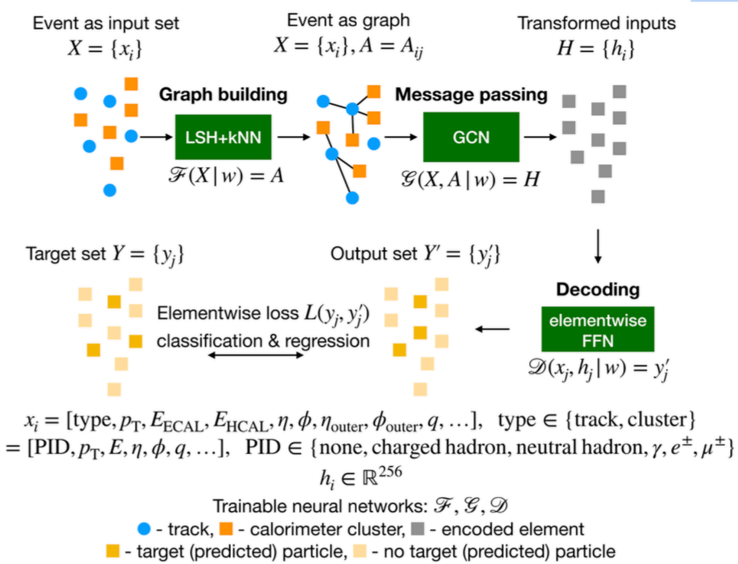
Target:
Physics event
@collision



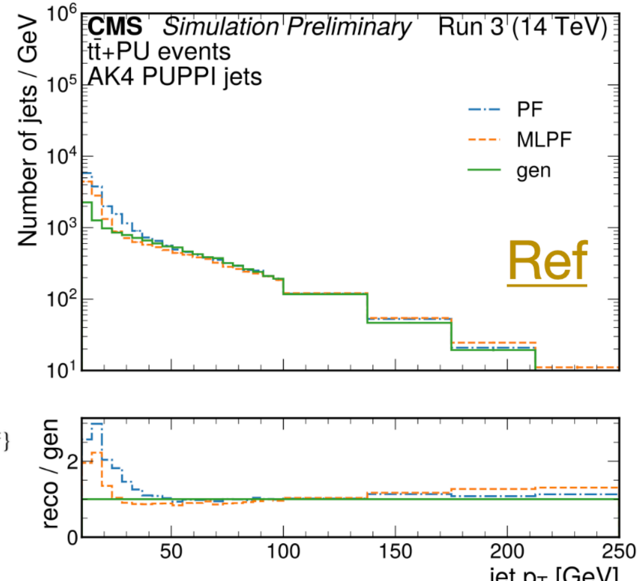
ML -PF

ML-PF reconstruction

- **Traditional [current] PF:** Rule-based algorithm using info from all sub-detectors
  - **Output:** mutually exclusive list of particles
- **ML-PF:**
  - Start from same inputs as PF [ie. tracks, clusters]
  - Particles: Sparse data → Point-could representation + GraphNN
  - Targets: Truth particles, particle ID and energy



Impact on high-level observables



Encouraging results

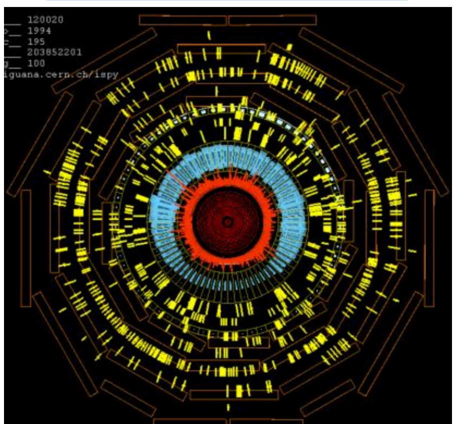
Obviously: WIP

- improving algorithm
- study robusteness/performance w/ collision data

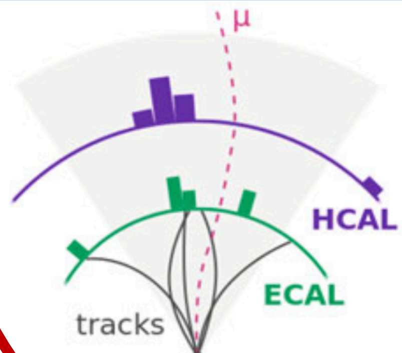
DL for Local Reconstruction

- Improvements in performance translate to improvement in higher-level reco

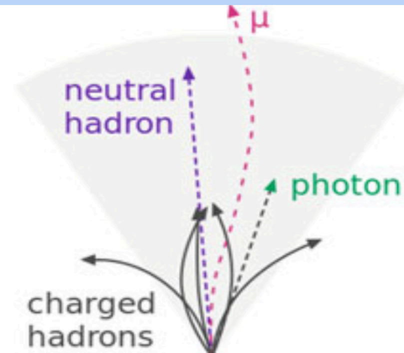
What we see:
Detector signals



Cluster signals
in each sub-detector
[Local RECO]

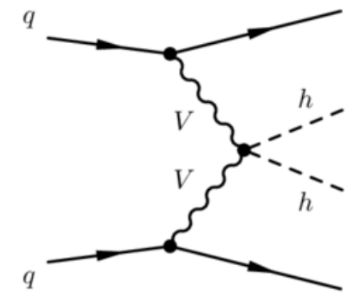


Combine info from
different detectors
[PF particles]



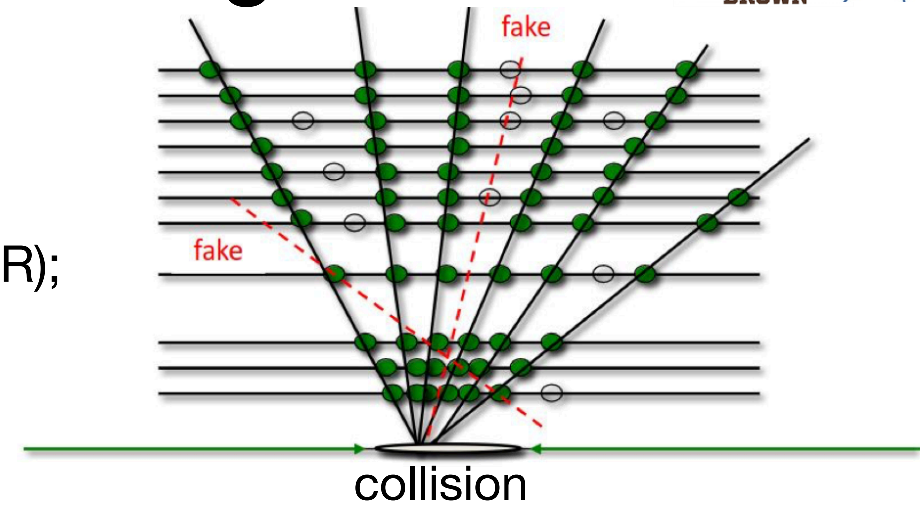
Output: γ, e, μ, charged & neutral hadrons

Target:
Physics event
@collision



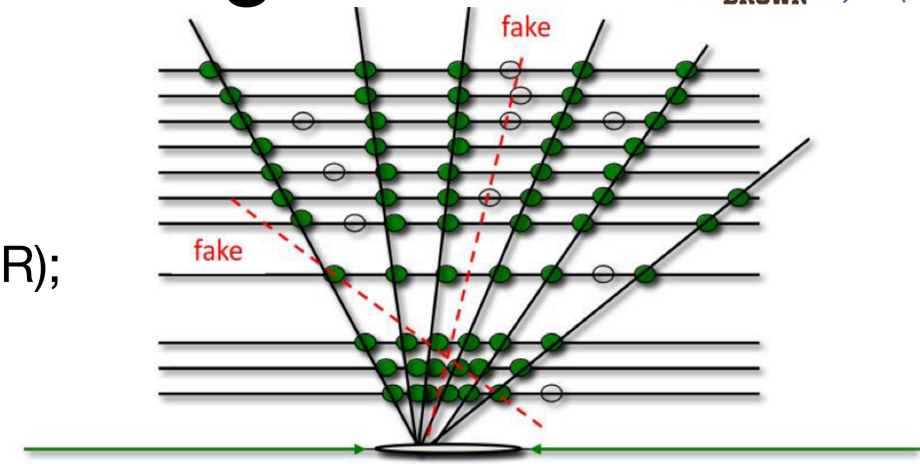
Local Reconstruction: Tracking

- eg. Tracking @LHC
  - O(30) charged particles/pp collision;
  - O(1500) charged particles/event
- Target: >90% efficiency, O(%) fake rate (FR);
  .. and fast [limited CPU resources]
- Several steps involved
  - Seed
  - Pattern Recognition
  - Fit (est. trk params)
  - Final selection (reduce FR)

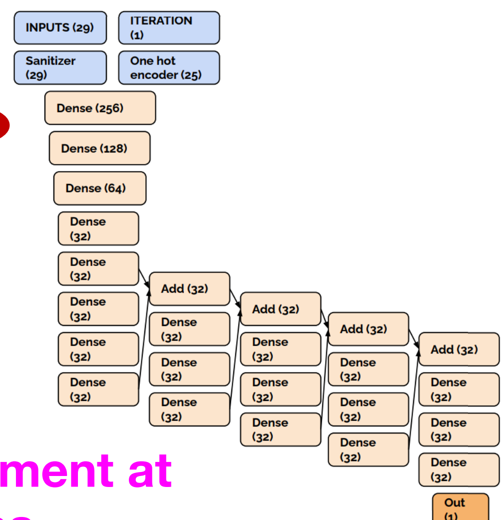


Local Reconstruction: Tracking

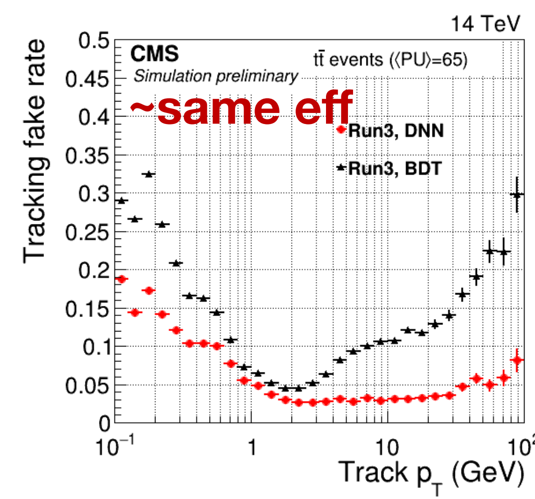
- eg. Tracking @LHC
  - O(30) charged particles/pp collision;
  - O(1500) charged particles/event
- Target: >90% efficiency, O(%) fake rate (FR);
  .. and fast [limited CPU resources]
- Several steps involved
  - Seed
  - Pattern Recognition
  - Fit (est. trk params)
  - Final selection (reduce FR)



Recent development:
DNN-based track selection



Ref



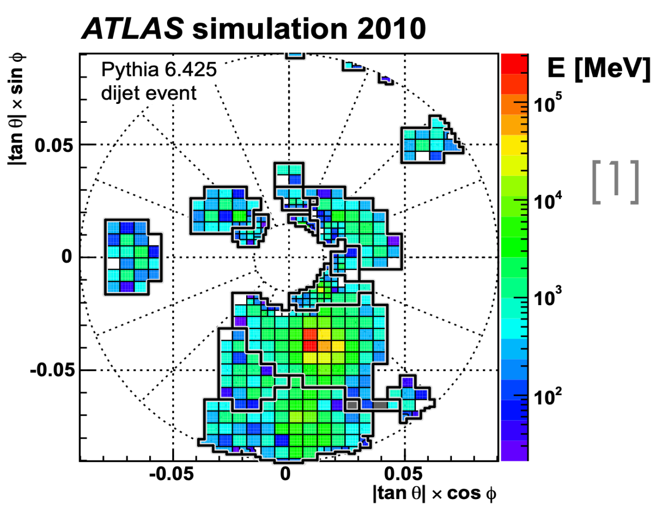
WIP: access improvement at high-level observables

2x improvement wrt BDT

Local Reconstruction: Calorimeters

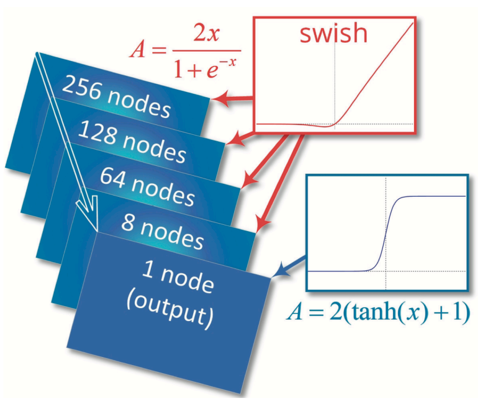
- **Calorimeter clustering:** Energy-based algorithm
  - Important for jet reconstruction
- **Energy calibration** critical: energy loss, non-compensation in hadrons,..
  - Currently: rule-based approach
    - several steps → an "average" correction

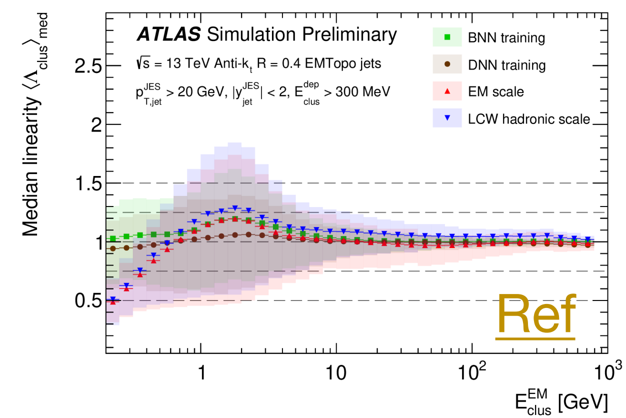
On going efforts:
DNN/Graph-based



[1]

- **Inputs:** Several cluster properties
  - better exploit correlations
- **Target:** Truth energy
- **One-step** approach





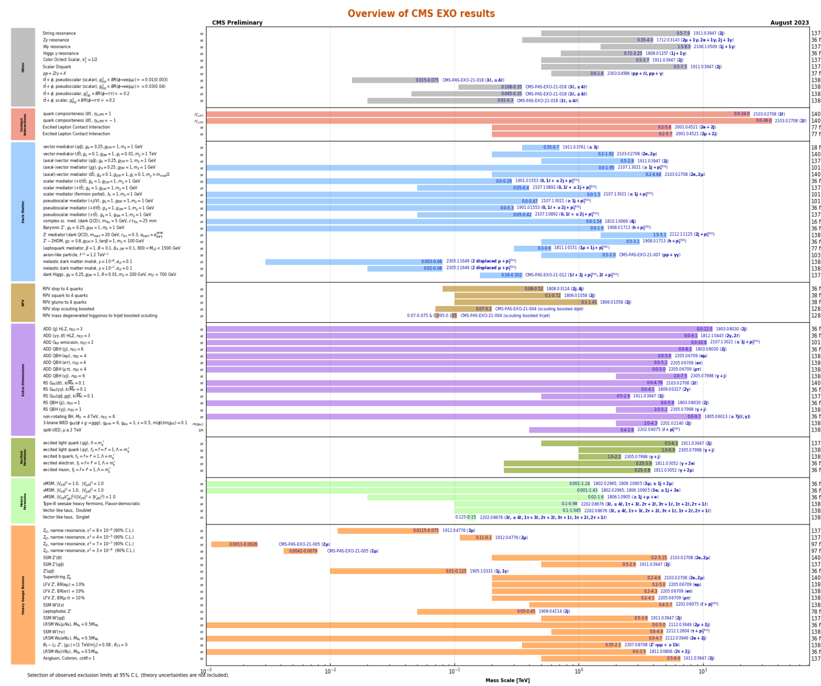
Anomaly detection [i.e. less model dependence]

Searching for the unknowns

- Typical analysis workflow:
  - Starting point: signal hypothesis
  - Design triggers and analysis strategy
  - Extract signal from data
  - Interpretation

Great!
IFF we know what you are looking for..

Summary of direct searches for new physics: [null results so far]



Searching for the unknowns

- Typical analysis workflow:
  - Starting point: signal hypothesis
  - Design triggers and analysis strategy
  - Extract signal from data
  - Interpretation

Great!
IFF we know what you are looking for..

D. Rumsfeld (2002):

"Reports that say that something hasn't happened are always interesting to me, because as we know, there are **known knowns**; there are things we know we know."

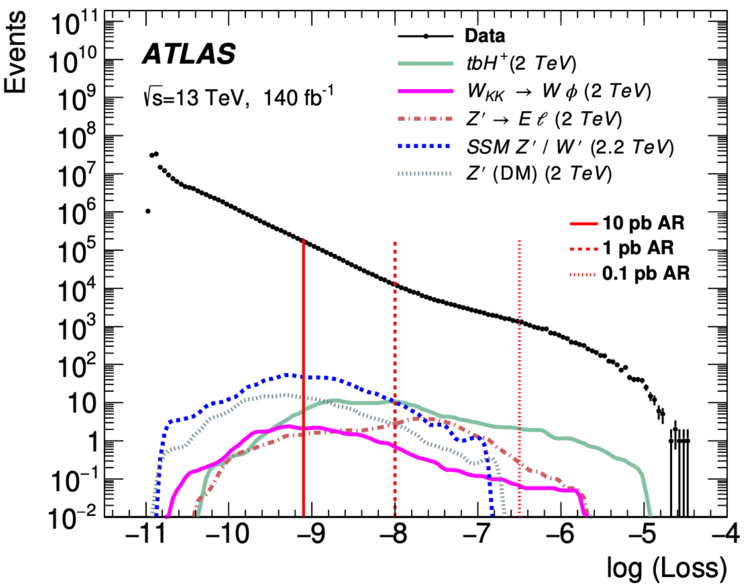
"We also know there are **known unknowns**; that is to say we know there are some things we do not know."

"But there are also **unknown unknowns**—the ones we don't know we don't know. And if one looks throughout the history of our country and other free countries, it is the latter category that tends to be the difficult ones."

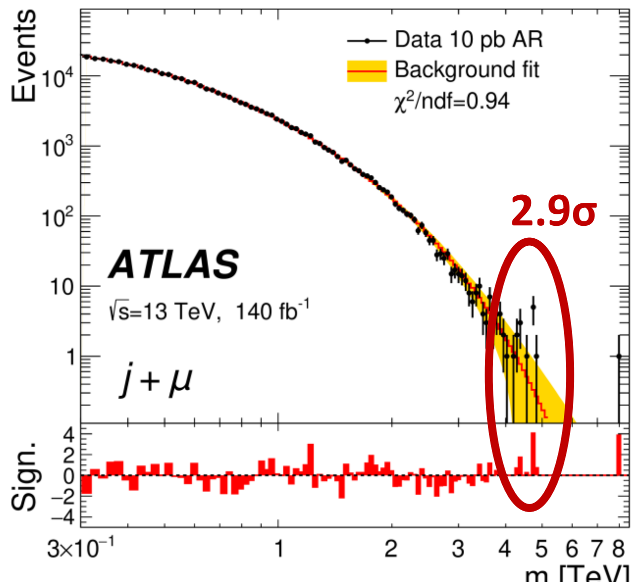
source: Wikipedia

Anomaly detection

- From **fully supervised → unsupervised** methods via autoencoders
  - **compress** input info → **learn** most important features → **decompress**
  - IF output "far" from input [eg. autoencoder loss] → **"anomaly" detection**
- In practice:
  - **Train** using only the known processes [use of collision data directly]
  - **Test** in collision data → define anomaly regions (AR) and look for deviations



Look for a "bump"

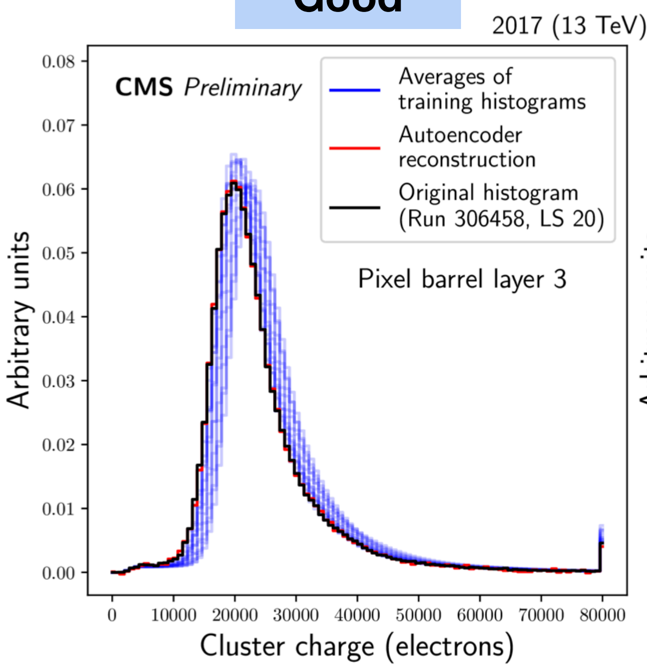


- Promising & timely development
- On going efforts: implement @trigger

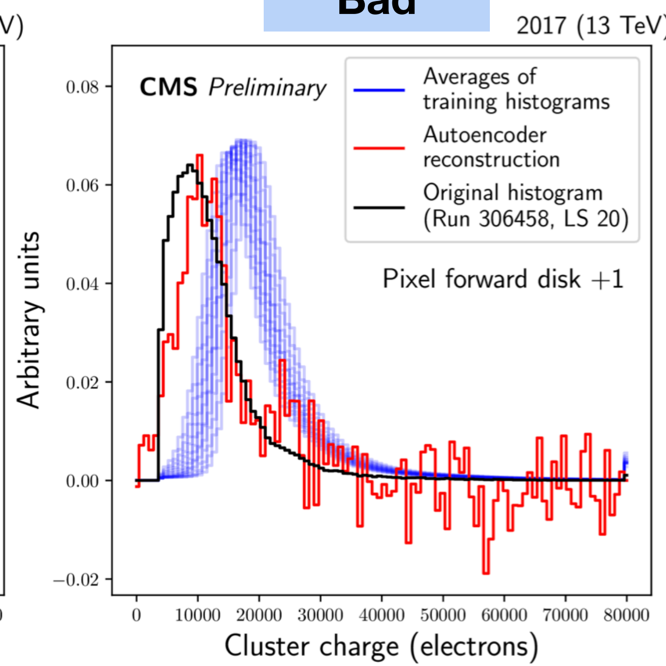
Autoencoders for detector monitoring

- Reliable and prompt evaluation of detector operation is critical
  - Human-centric validation: impossible to review all metrics [O(100M) channels]
- ML-based workflow: review huge amount of metrics, timely
  - Use autoencoders to identify distributions with large “reconstruction error”

Good



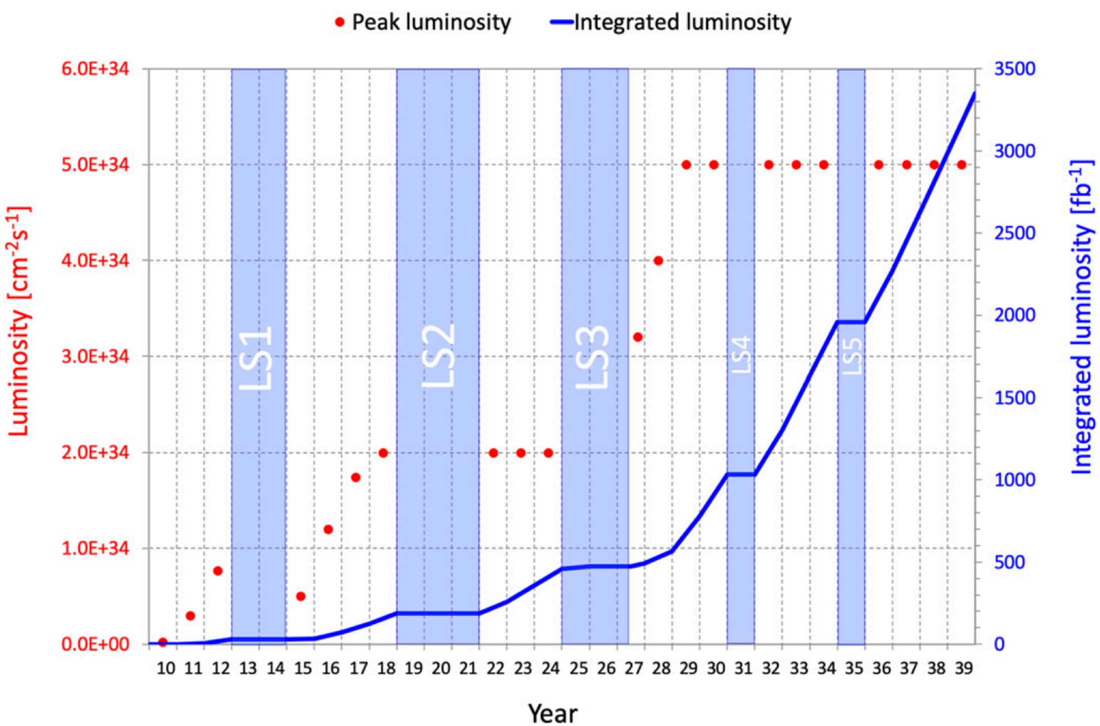
Bad



- 4x better in finding “anomalies”
- and much faster

Longer-term future: High-Luminosity LHC (HL-LHC)

LHC upgrade: HL-LHC

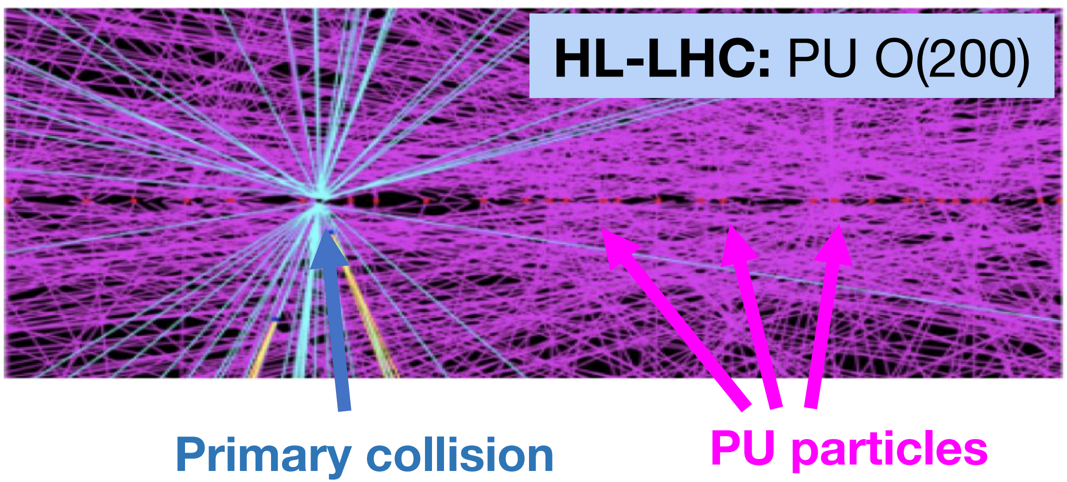


- Very rich PH program ahead
  - Unprecedented amount of data ☺
  - … at much harsher conditions ☹
    - radiation, pileup
- Major challenge for experiments
  - Requires detector upgrades
  - Ingenuity in event reco.

Often use LHC Run 3 as a prototype for HL-LHC

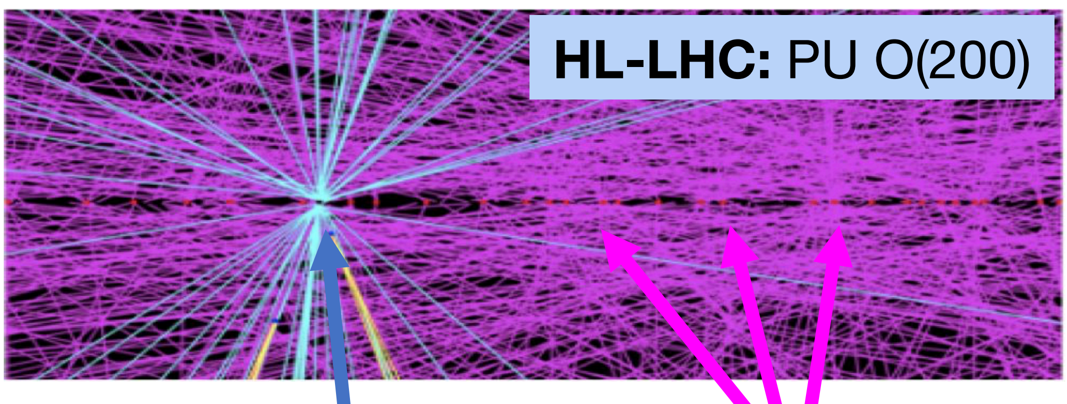
Challenge: pile-up (PU)

- PU: # simultaneous collisions to the main one



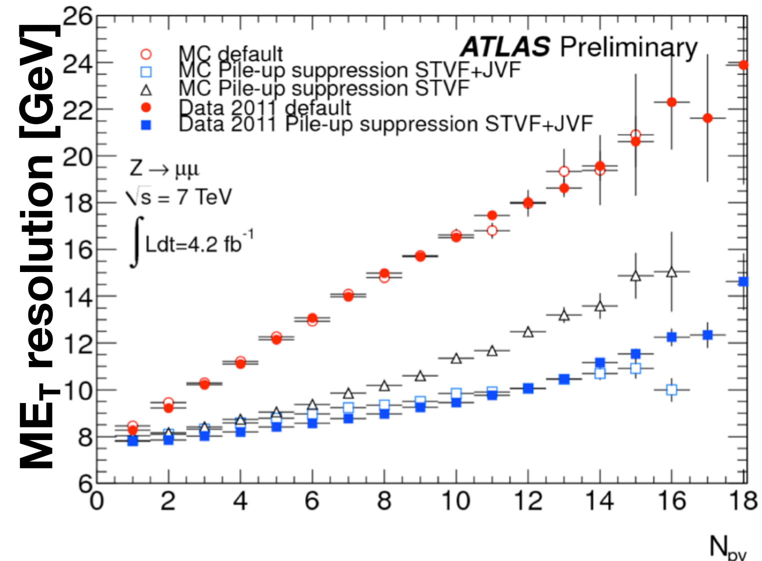
Challenge: pile-up (PU)

- PU: # simultaneous collisions to the main one



Primary collision **PU particles**

A big challenge for (HL-) LHC

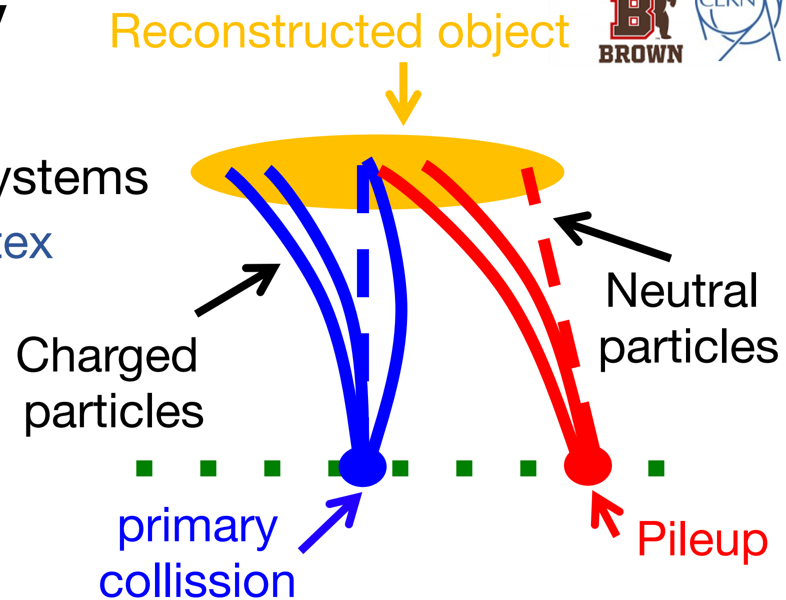


- Significant **degradation** in particle reco. performance
- **Mitigation:** remove particles from PU

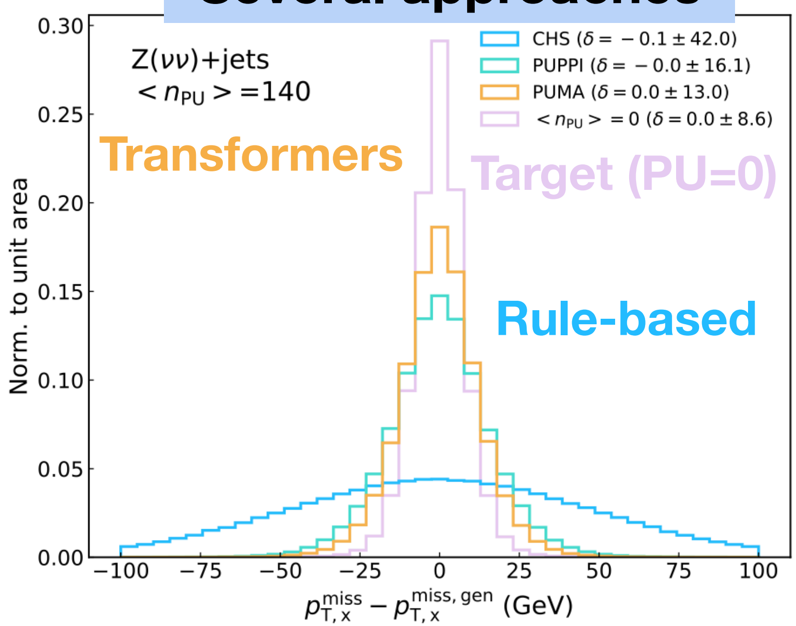
…Searching for a hay in a haystack

Tackling PU: algorithmically

- **The problem:**
  neutral particles do not interact with tracking systems
  - Impossible to directly identify the collision vertex
- **Challenges:**
  Dimensionality O(10K) particles/event, limited handles/features, truth definition



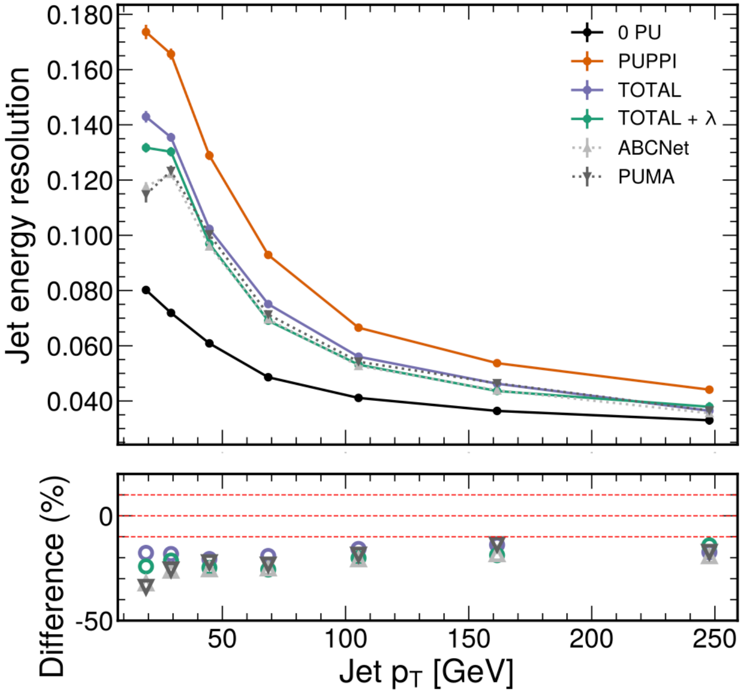
Several approaches

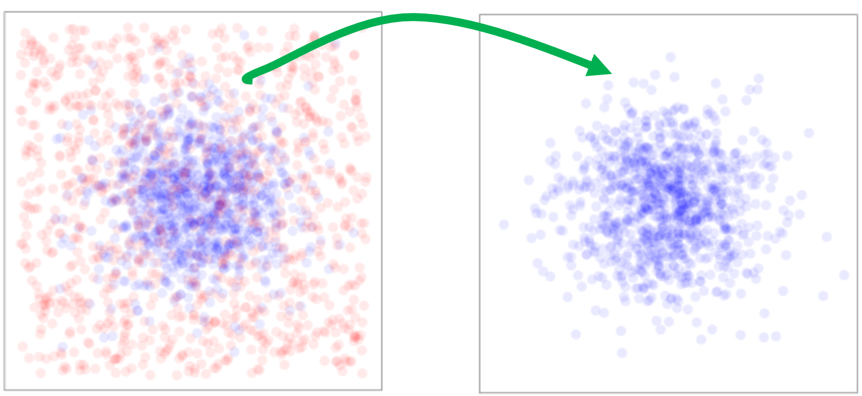


- **Transformers-based:** ~25% improved resolution wrt rule-based [ie PUPPI]
- **Still:** 80% worse than PU=0 [ultimate goal]
- **Also:** ML-based ones rely on labeled data
  → Not trivial to assign in realistic simulation

Tackling PU: algorithmically (II)

- Alternative approaches: Towards less supervised algorithms
  - Based on optimal transport
    - measure the "distance" between probability distributions
- Strategy:
  - **Input:** Same physics events [with & without PU]
  - **Train:** GNN to minimize the distance b/w the PU and no-PU sample



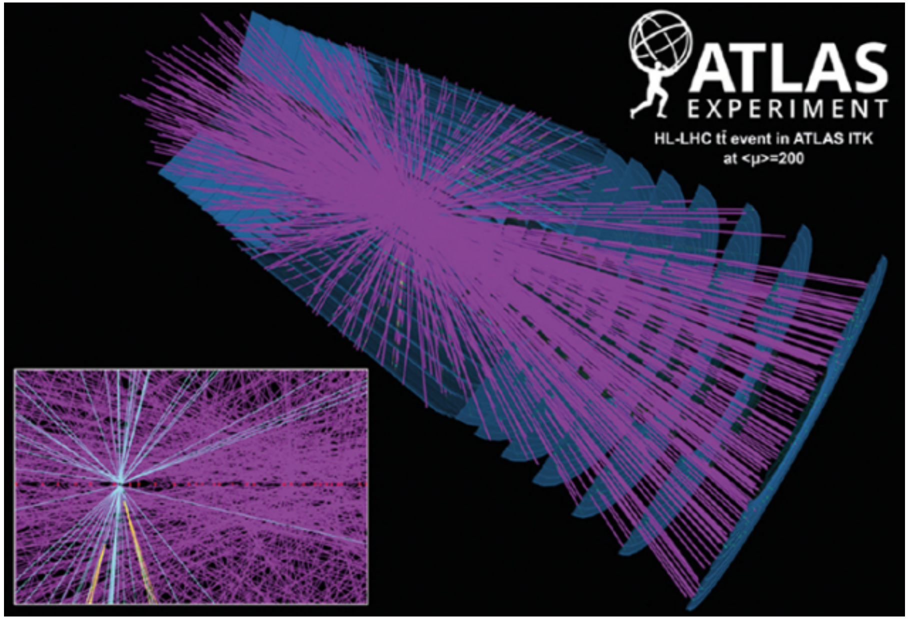


- **Similar** performance to other ML-based ones
- **Advantages:** less supervised, unlabeled [simulated] data

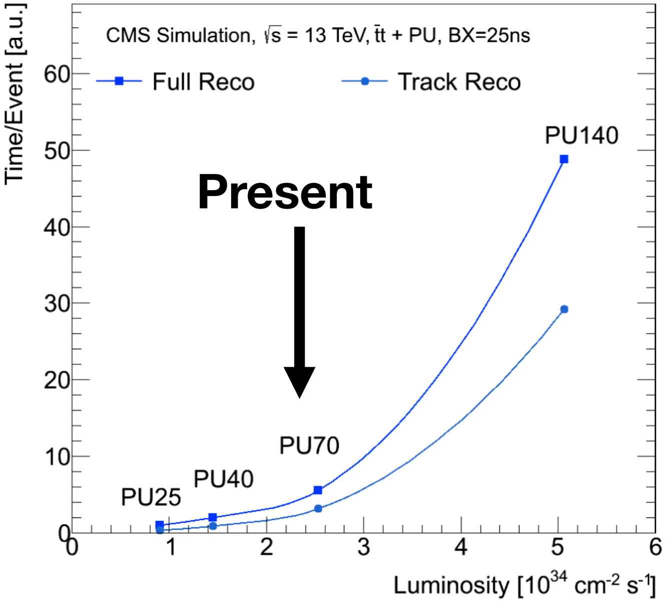
PU mitigation: still room for improvement; validation in data

Detector upgrades

- More granular, more complex, more powerful
  - Higher event rate → higher occupancy
  - More complex/granular → [much] more time to perform reconstruction
- Much more potential, but …
  - Existing techniques [factorized approach, rule-based algos] way above available budget

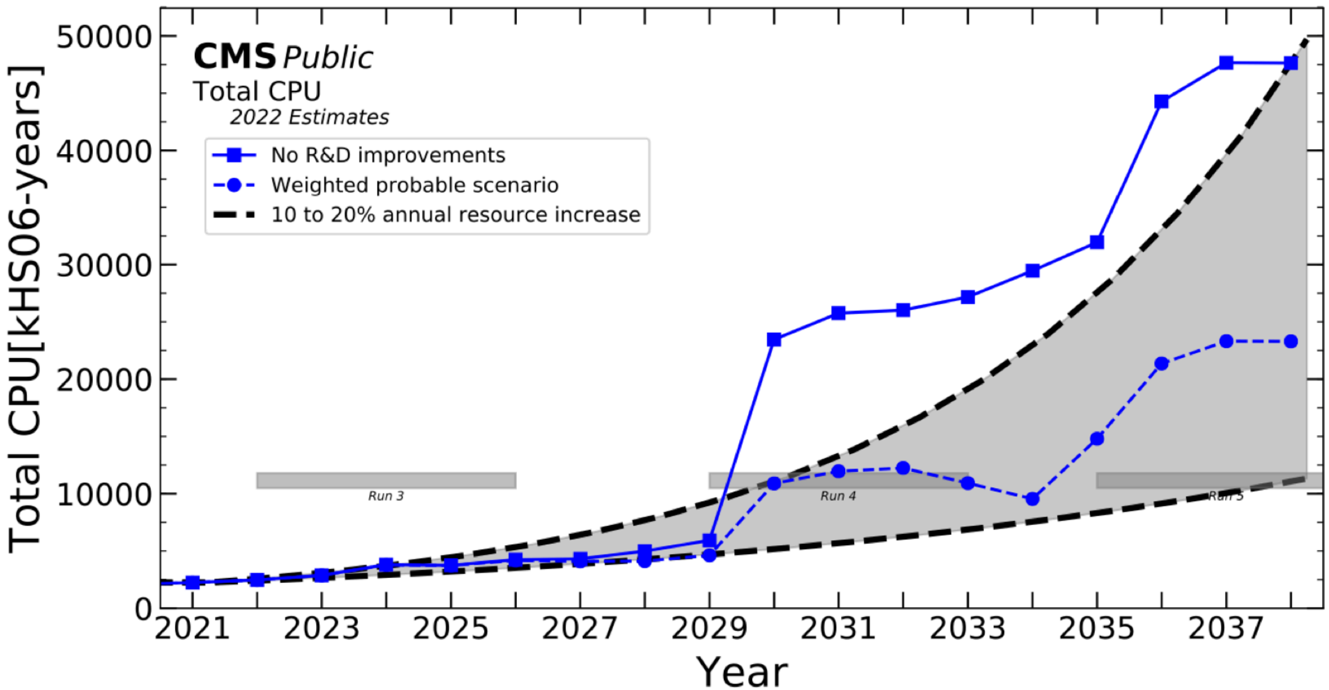


Tracking [time/event]



Detector upgrades (II)

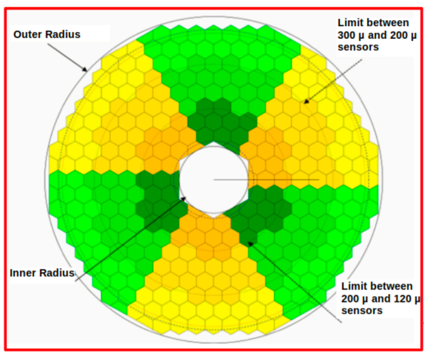
- We want to do more ..
  - **… but with modest increase** in computing resources

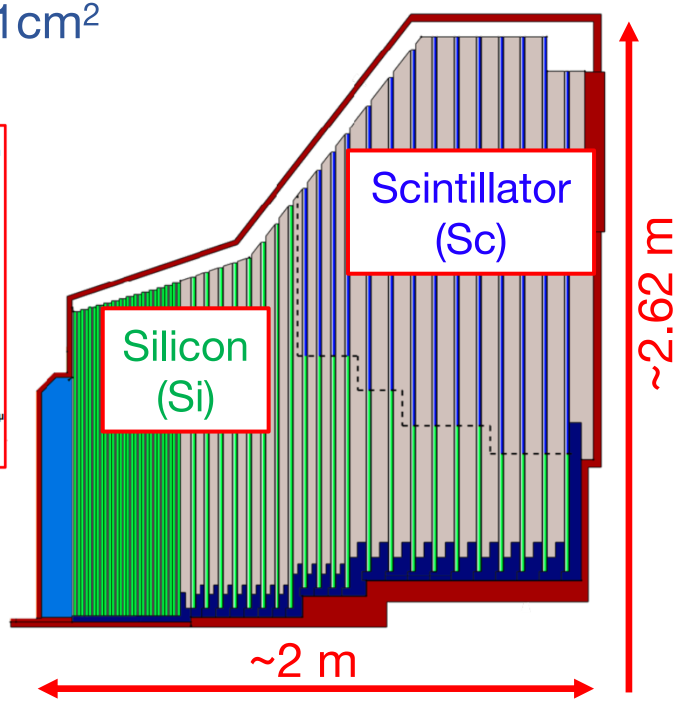


- Current algorithms [factorized approach, rule-based], beyond budget
- Need modern tools and ingenuity [in all fronts]

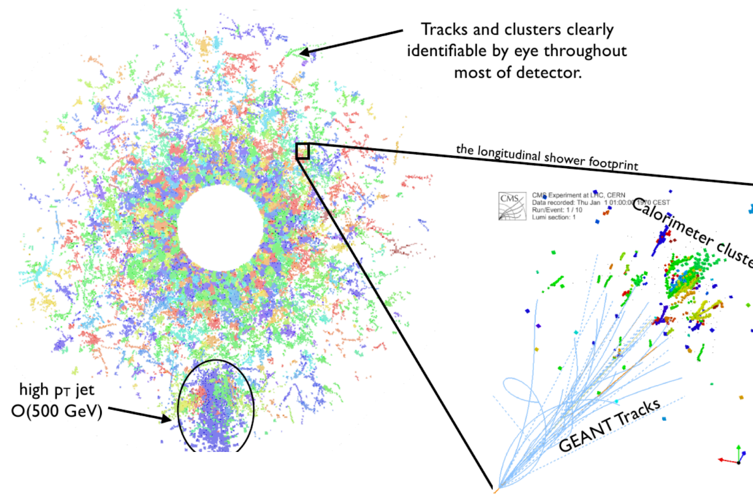
Detector upgrades

- New era in calorimetry
  - highly granular + timing capabilities [HGCAL]
    - 5D reconstruction
  - > 6M channels
  - sensor area < 1cm^2





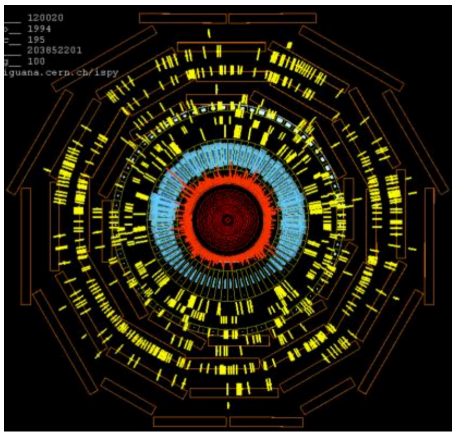
Great potential



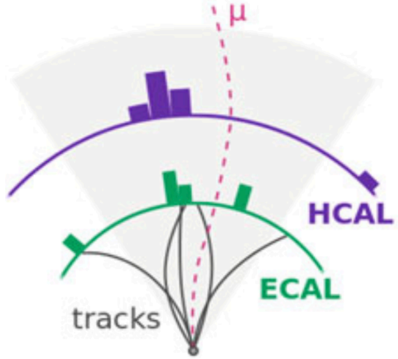
Event reconstruction @ HL-LHC

Traditional [Rule-based] approach

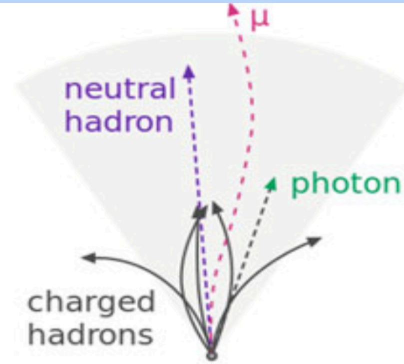
What we see:
Detector signals



Cluster signals
in each detector
[Local RECO]

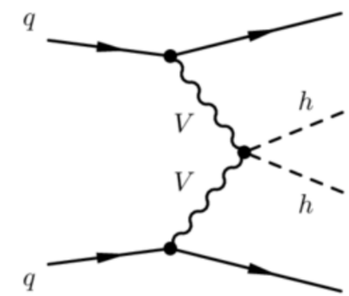


Combine info from
different detectors
[PF particles]



Output: γ, e, μ, charged & neutral hadrons

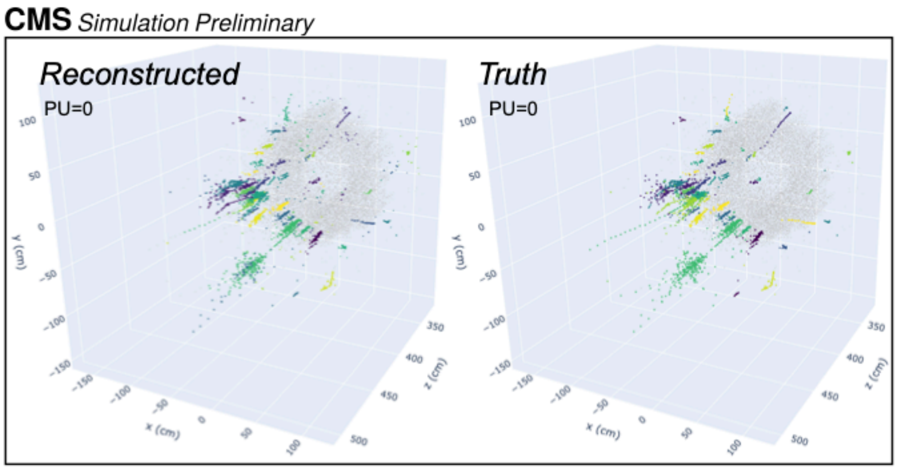
Target:
Physics event
@collision

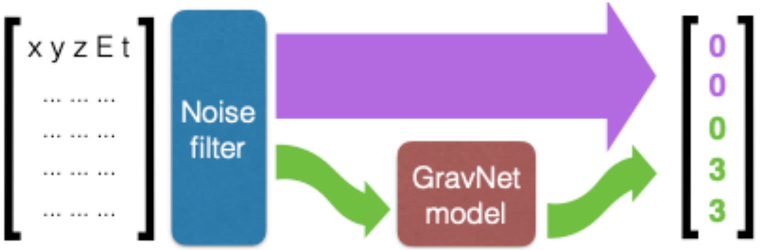


End-to-end reconstruction

ML-based event reconstruction

- (I) Fully supervised GNN-based clustering
  - start from detector hits → full shower [End-to-end approach]





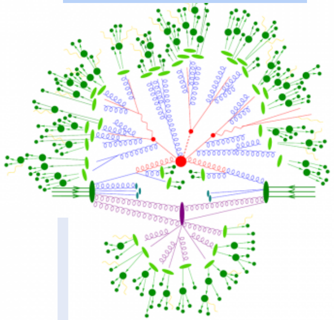
- Performance ☺
- Robustness, computing resources/inference time ☹

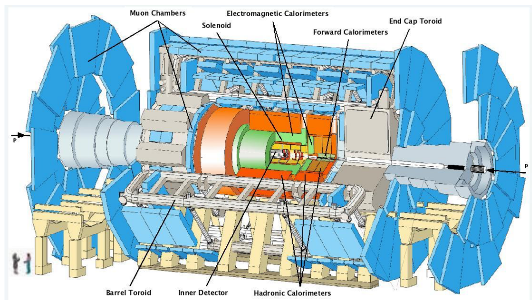
- Other efforts: “Contrastive learning”
  - **Define** a set of **positive** and **negative** connections between hits
  - **GNN-based** model to extract features → calculate cosine similarity
  - Cluster connections above some threshold

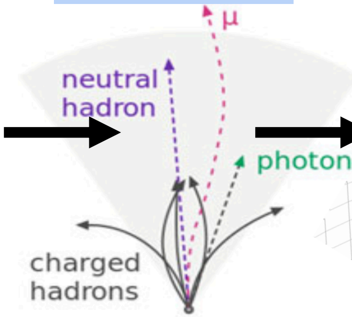
Event simulation

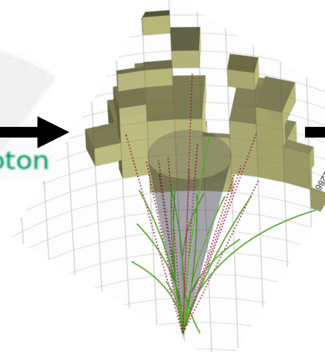
- Critical ingredient for physics analyses

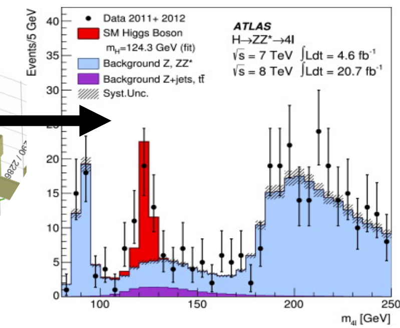
physics process → **Detector response** → **Reco particles** → **Analysis objects** → **Analysis Result**

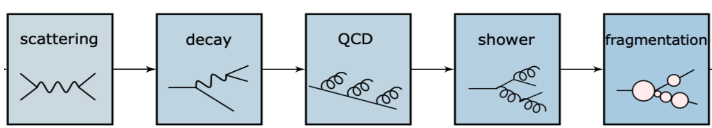






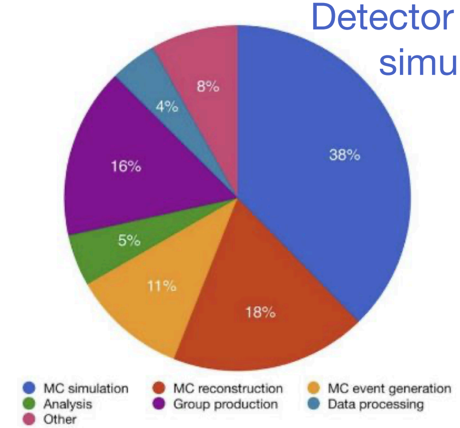






FullSim: Accurate but
computationally expensive O(10s/event)
- must be reduced to keep up w/ amount of data @HL-LHC

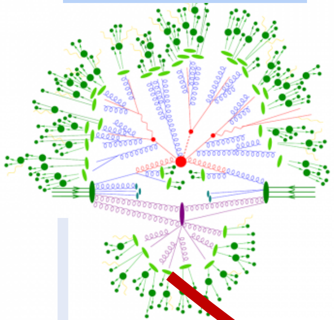
Detector response simulation

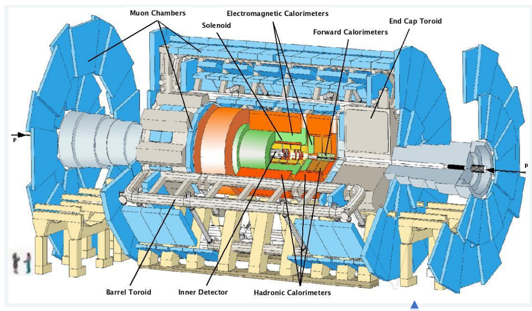


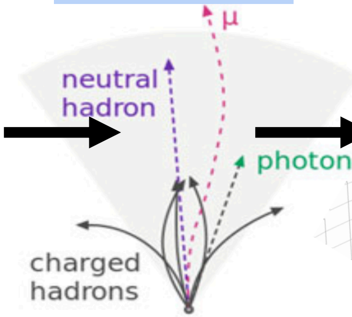
Event simulation

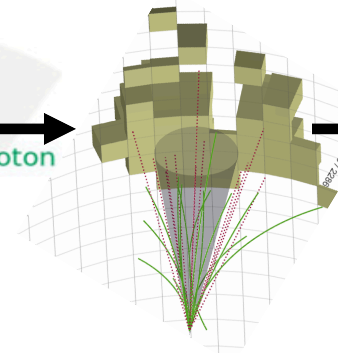
- Critical ingredient for physics analyses

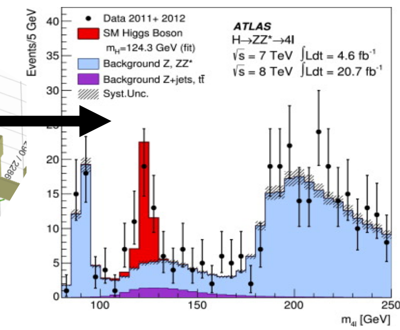
physics process → **Detector response** → **Reco particles** → **Analysis objects** → **Analysis Result**

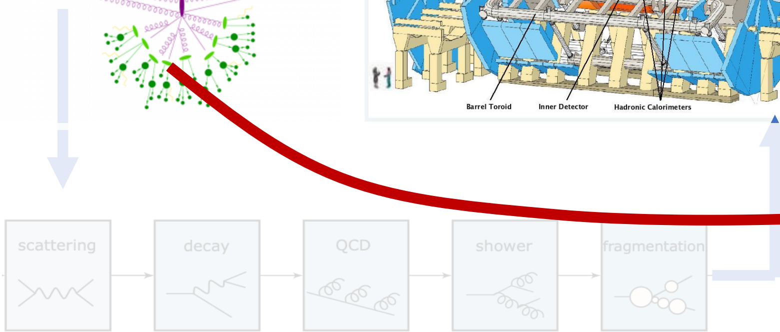








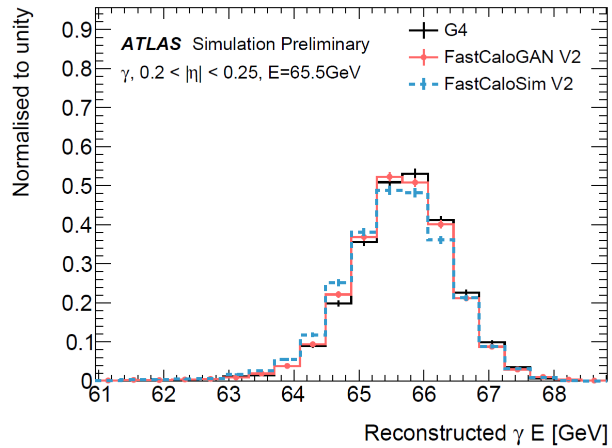




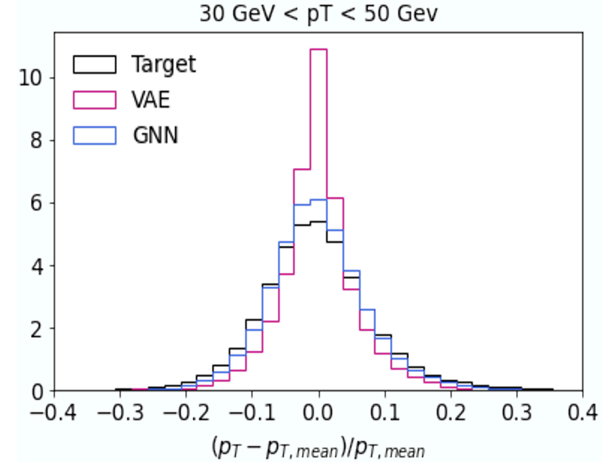
FastSim: Can we speed-up workflow by skipping steps?
[but without significant loss in accuracy?]

FastSim w/ ML ?

Calo shower generation



End-to-End Gen → Reco particles



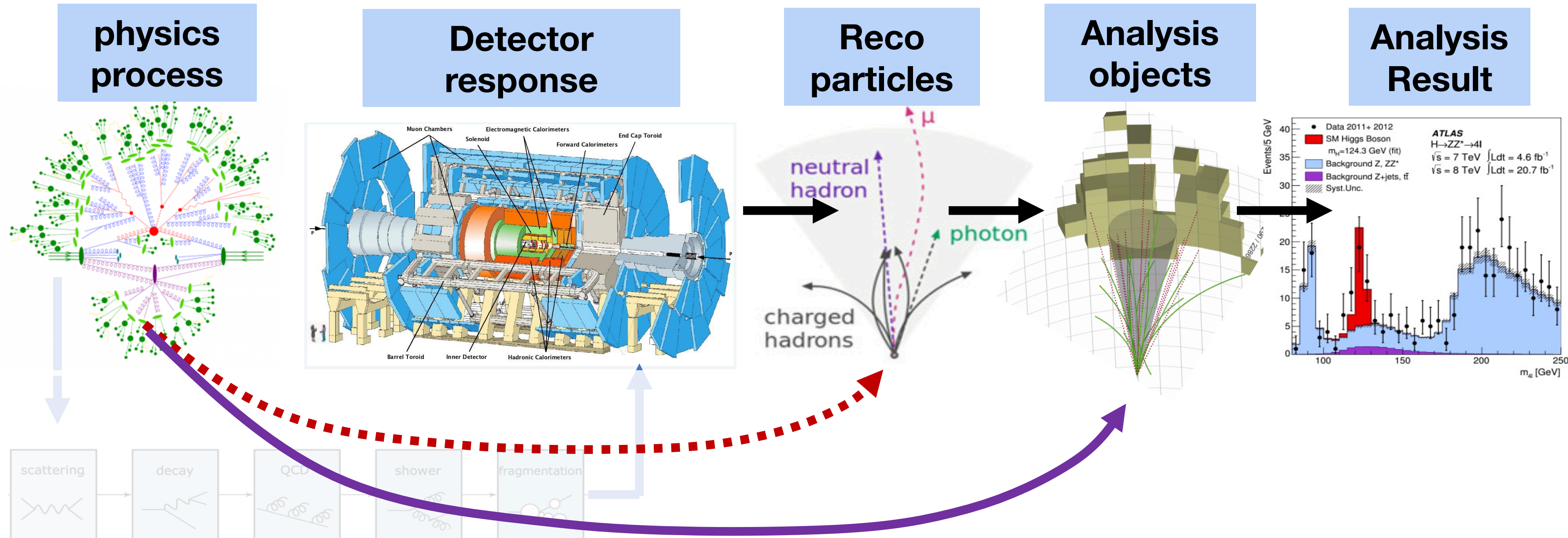
- Enormous effort; attacking from several angles
  - different networks
    - GANs, WGANs, VAE, …
  - different approaches
    - target specific parts of the chain, end-to-end..

O(100-1000)x faster

Dataset	N. of voxels	N. of weights	Time to 100 showers [s]		
			CaloScore	WGAN-GP	Geant
dataset 1	384	32M	4.0	1.3	$\mathcal{O}(10^2 - 10^3)$
dataset 2	6480	1.4M	5.8	1.33	$\mathcal{O}(10^4)$
dataset 3	46080	1.7M	33.4	2.06	$\mathcal{O}(10^4)$

Event simulation

- Critical ingredient for physics analyses



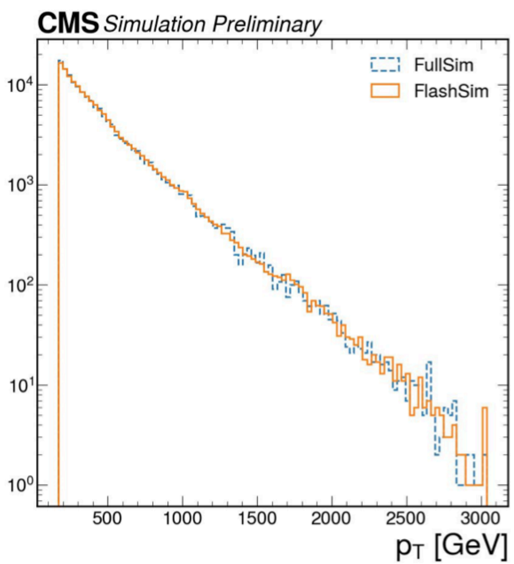
FlashSim: Skip all intermediate steps

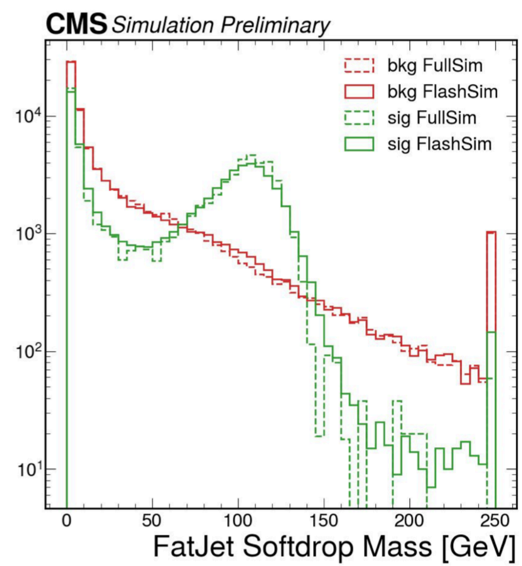
FastSim: Can we speed-up workflow by skipping steps?
[but without significant loss in accuracy?]

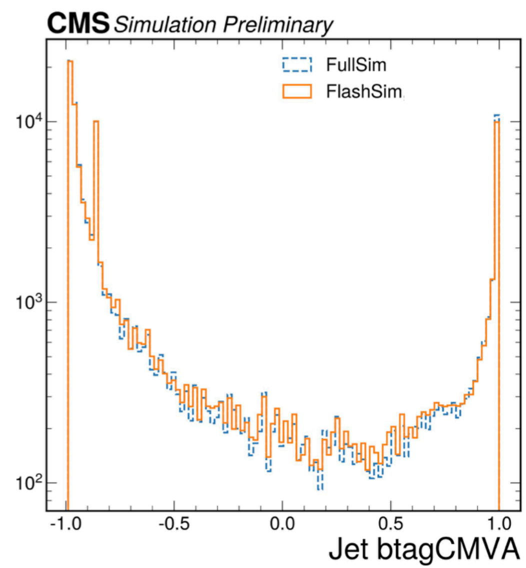
FlashSim

- GEN-level information → analysis physics objects [electrons, muons, jets…]
  - Target each physics object individually
    - instead of the whole event
  - Backbone: Normalizing Flows

complexity →







Encouraging results [accuracy, time/evt, and space]
Next step: model fakes…

Summary

- Enormous growth of ML in particle physics
  - Started with classification & regression tasks on high-level observables
- Present: Important component in almost all areas:
  - detector monitoring, real-time selection, local reconstruction, simulation, anomaly detection, …
  - Key: Interdisciplinary collaboration between Physicists and Computer Scientists
- Effort pays off: Major improvements compared to traditional techniques
  - Improved physics reach [beyond what Lumi-scaling would give us]
  - Opened up new opportunities
- There is a clear trend that ML will be the cornerstone for the success of the physics program for HL-LHC