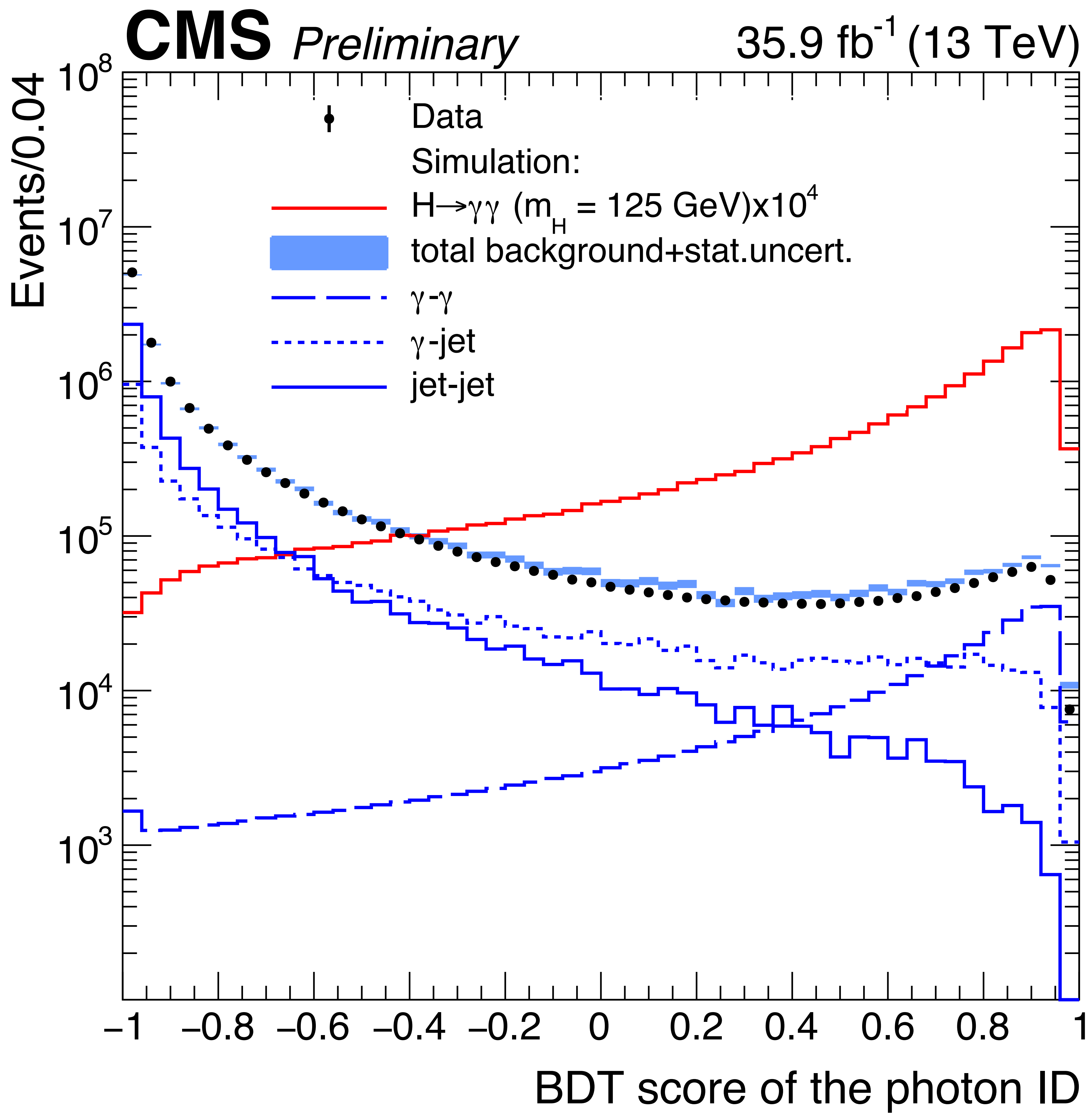
Machine learning for data taking, monitoring, and processing

Maurizio Pierini

European Research Council

What we do with ML today

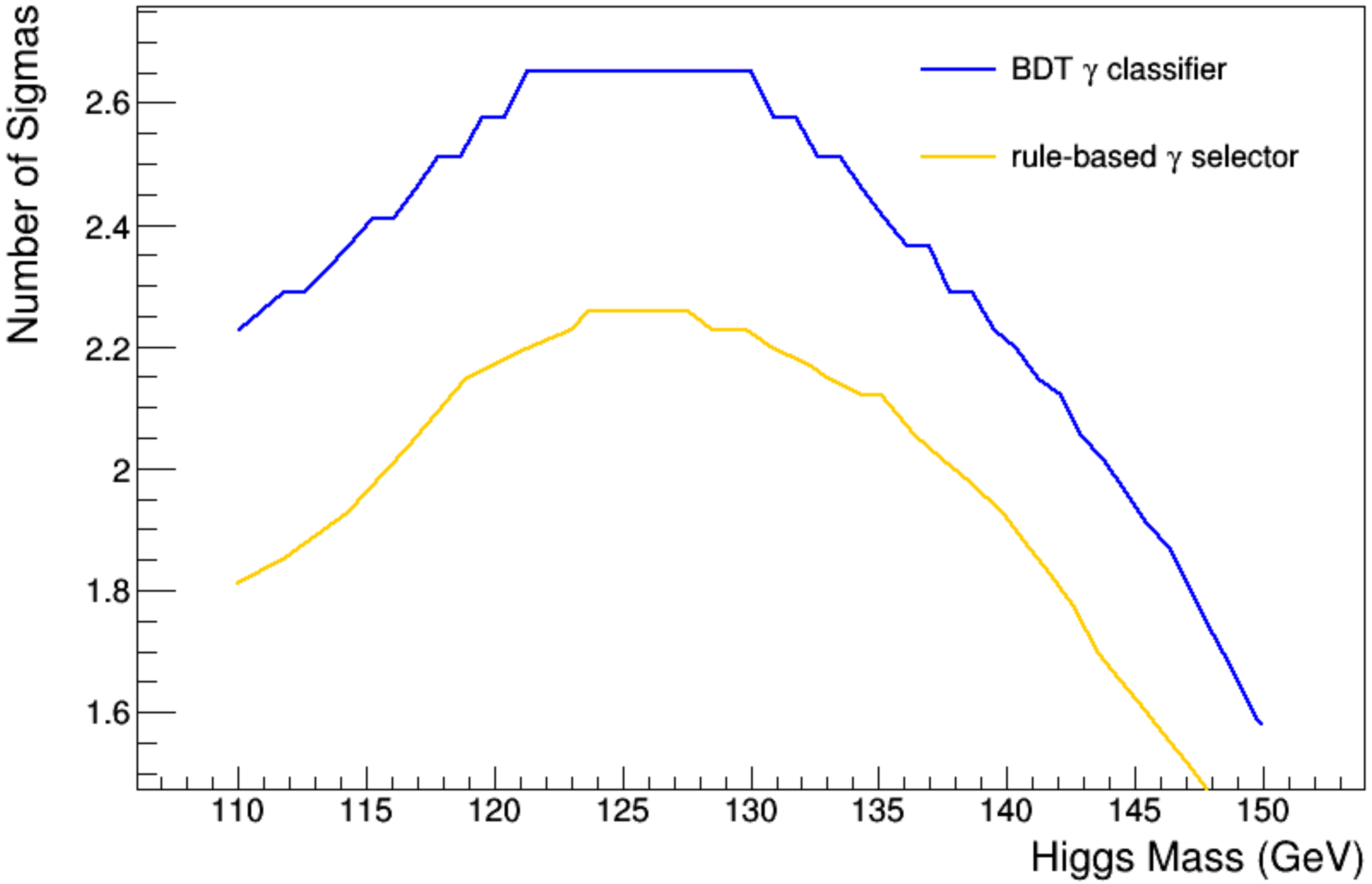
- Classification:
  - identify a particle & reject fakes
  - identify signal events & reject background
- Regression:
  - Measure energy of a particle
- We typically use BDTs for these task
  - moved to Deep Learning for analysis-specific tasks
  - same will happen for centralised tasks (eventually)

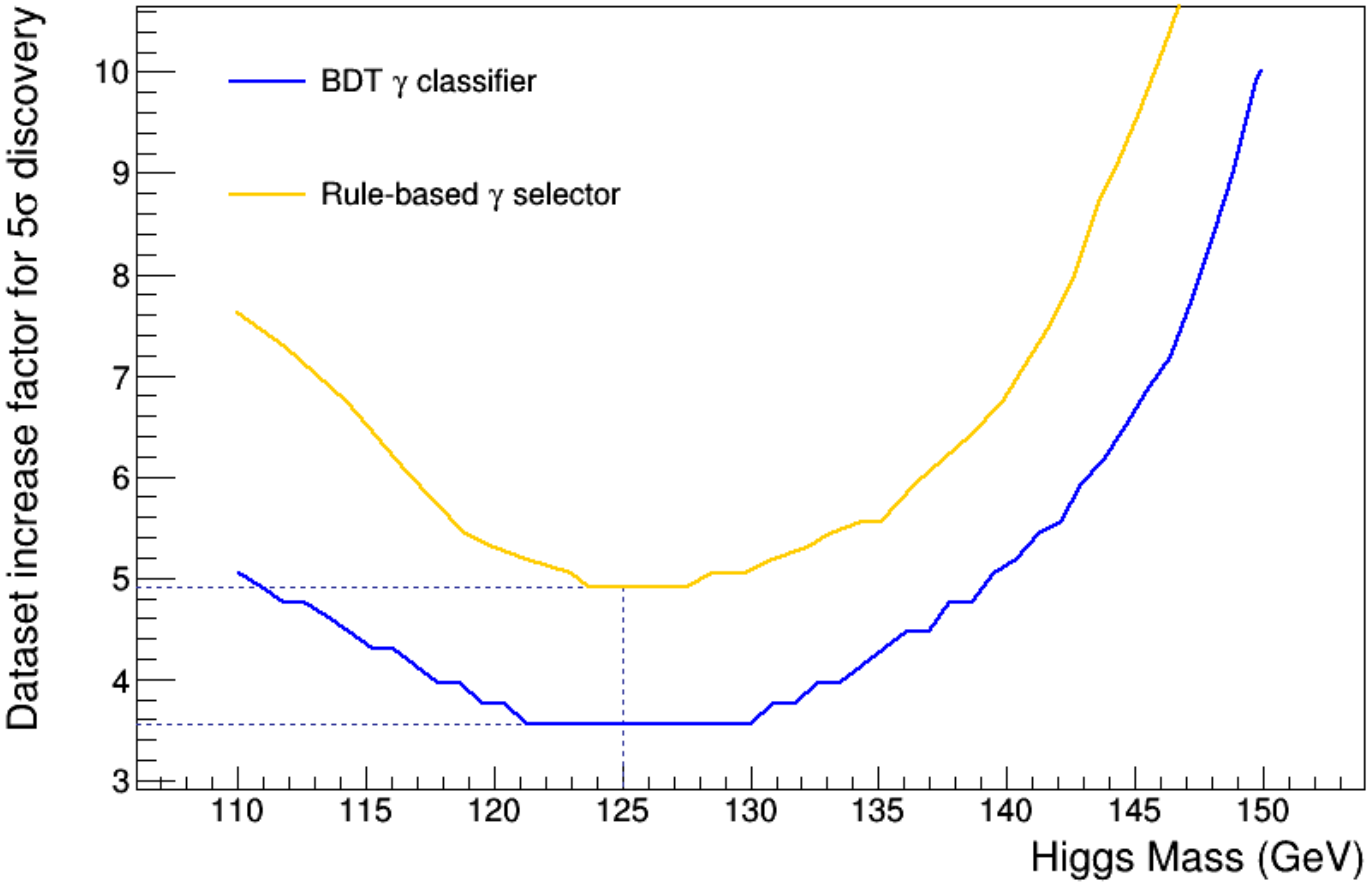


Centralised task (in online or offline reconstruction)
Analysis-specific task (by users on local computing infrastructures)

Example: ML for Higgs discovery

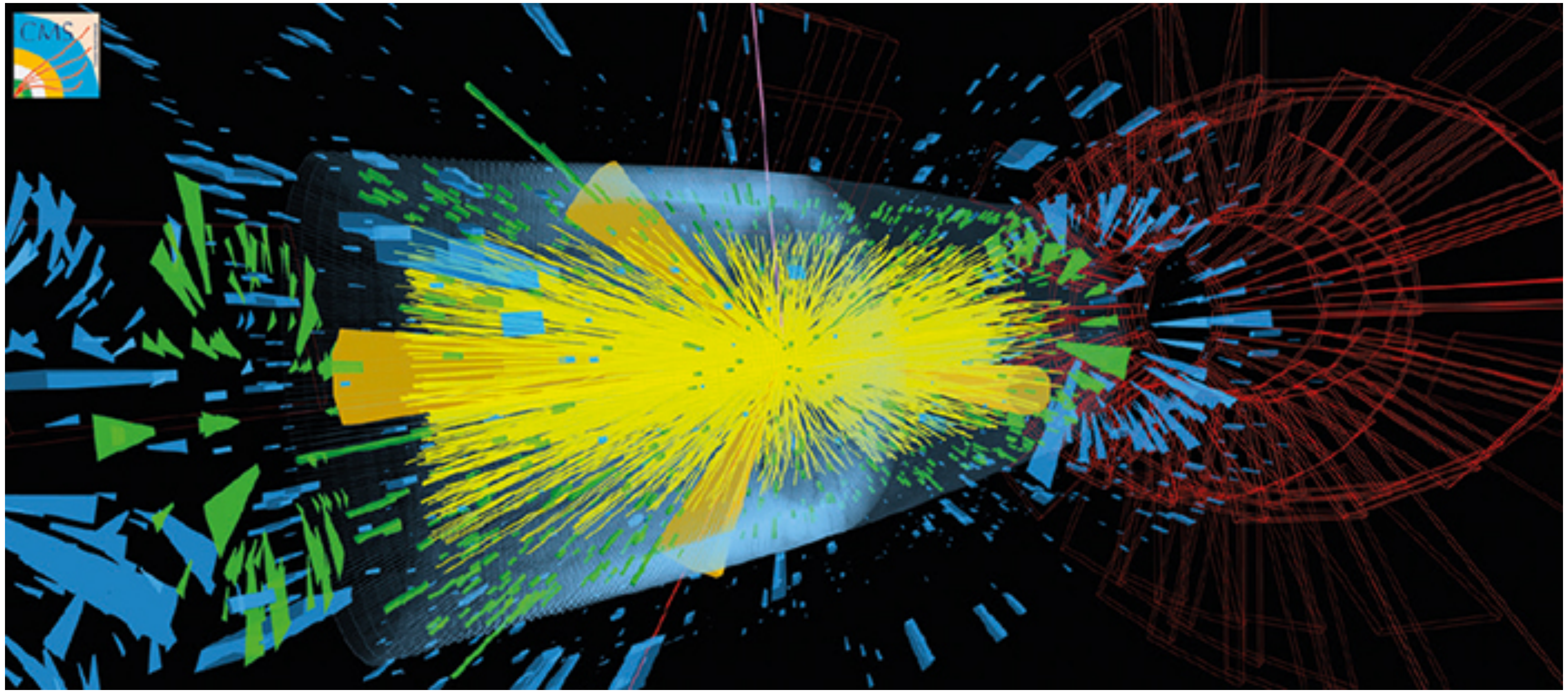
- *We were not supposed to discover the Higgs boson as early as 2012*
- *Given how the machine progressed, we expected discovery by end 2015 /mid 2016*
- *We made it earlier thanks (also) to Machine Learning*



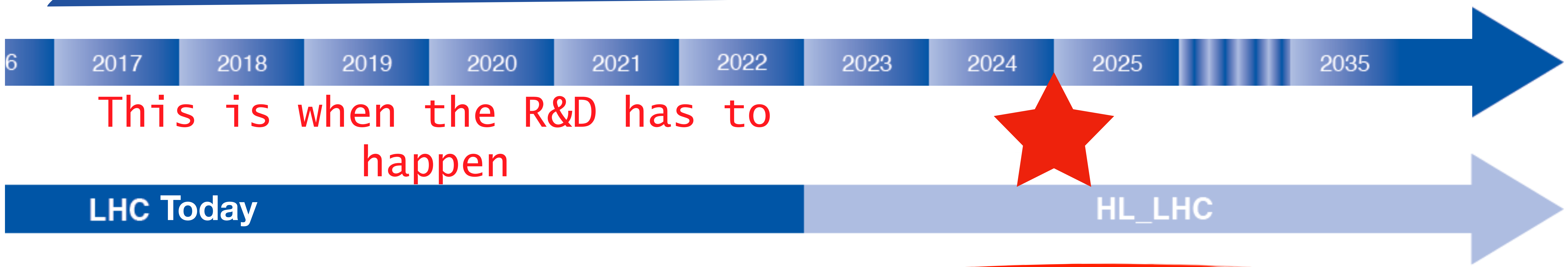


What is ahead of us

- *Deep Learning will be more and more central*
  - *Analysis-specific applications poses no problem in terms of latency/memory/etc*
- *Challenges ahead will force us (willing or not) to use DL in many centralised tasks*
  - *but we are still far from being ready to a systematic usage of DL in production*



HL-LHC: elephant in the room

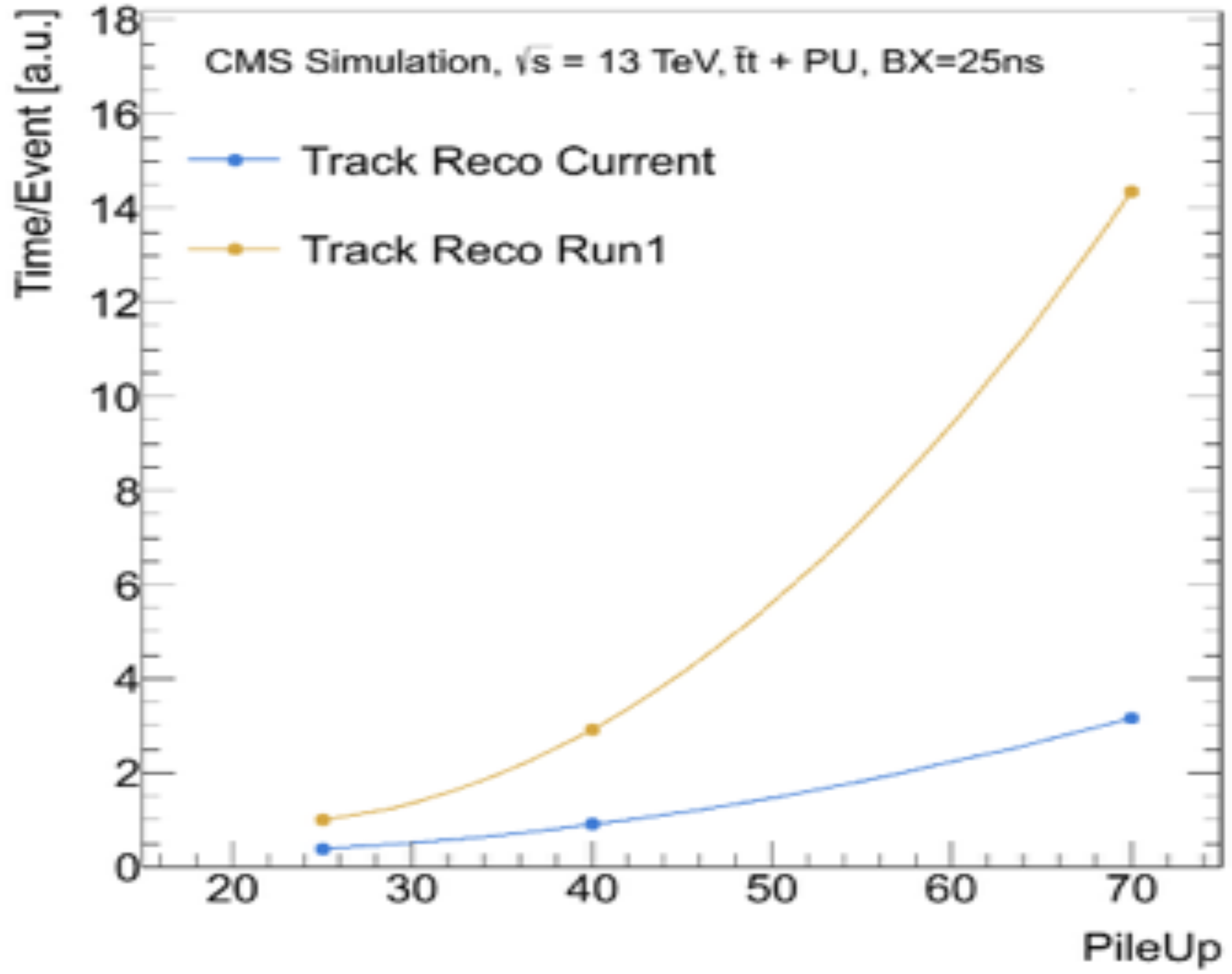


This is when the R&D has to happen

- ~40 collisions/event
- ~10 sec/event processing time
- (at best)Same computing resources as today

- ~200 collisions/event
- ~minute/event processing time(*)
- (at best)Same computing resources as today

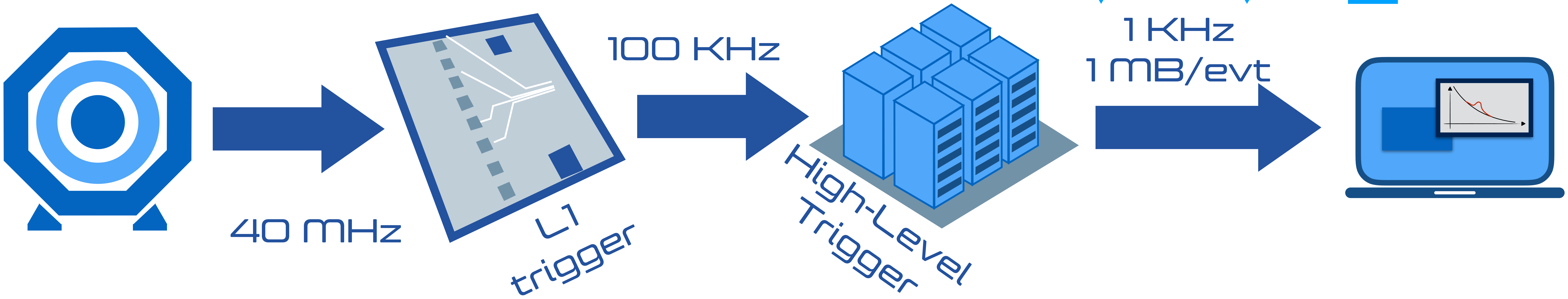
- *Flat budget vs. more needs =* ***current rule-based reconstruction algorithms will not be sustainable***
- ***Adopted solution:*** *more granular and complex detectors → more computing resources needed → more problems*
- ***Modern Machine Learning might be the way out***



(*)With nowadays software development

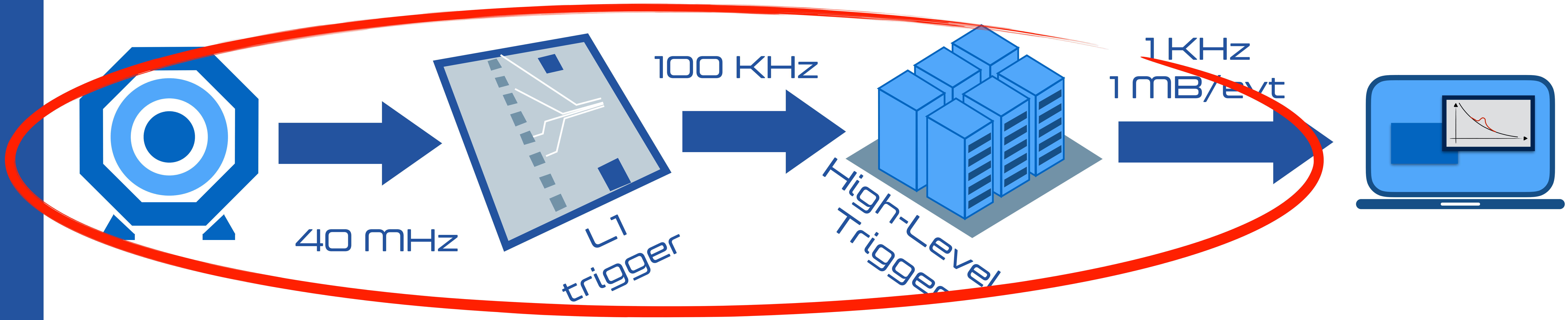
Three layers of reconstruction

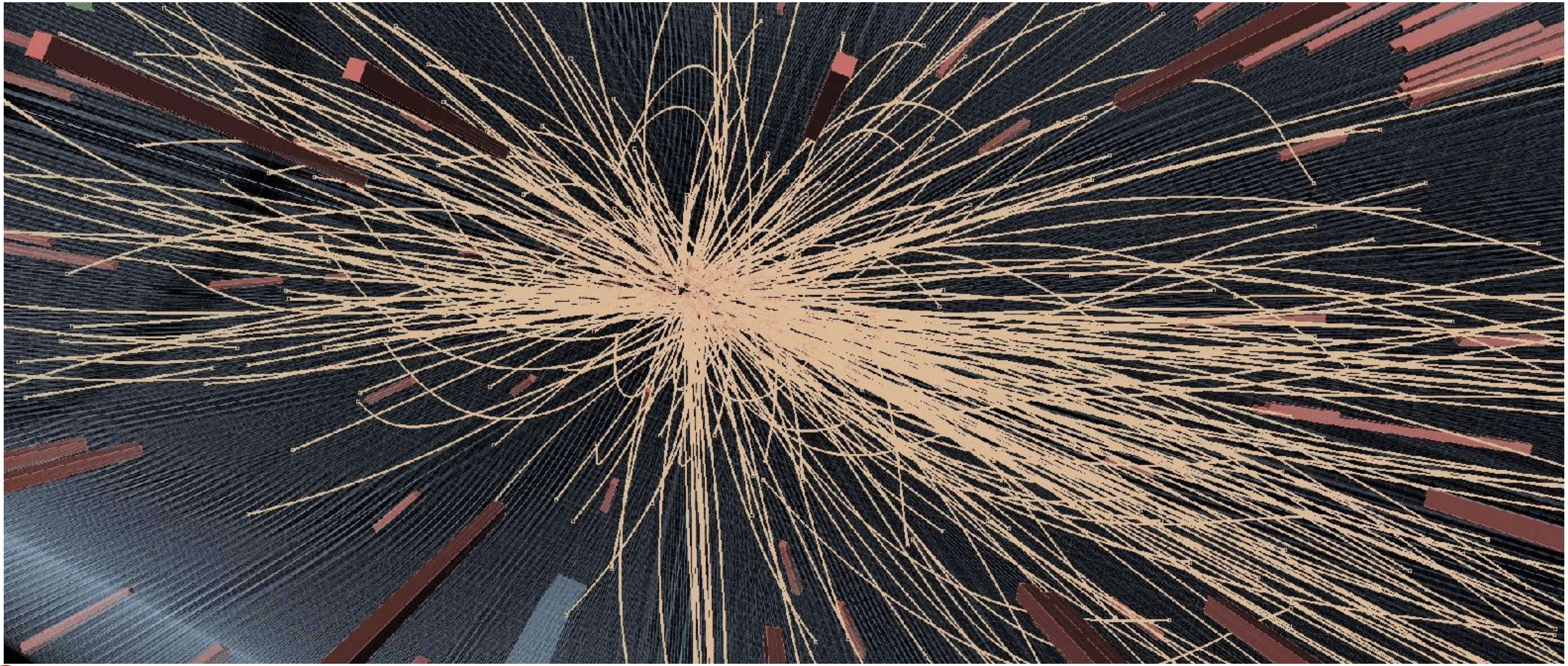
- *A typical reconstruction chain has 4 steps (*)*
  - *L1 trigger: local, hardware based, on FPGA, @experiment site*
  - *HLT: local/global, software based, on CPU, @experiment site*
  - *Offline: global, software based, on CPU, @CERN T0*
  - *Analysis: user-specific applications running on the grid*



What DL could do for us

- *The solution to the HL-LHC problem: modern Machine Learning …*
  - *… to be faster*
  - *… to do better*
  - *… to do more*
- *And this is a NEED for what happens in between data taking and data analysis (trigger, reconstruction, simulation, …)*

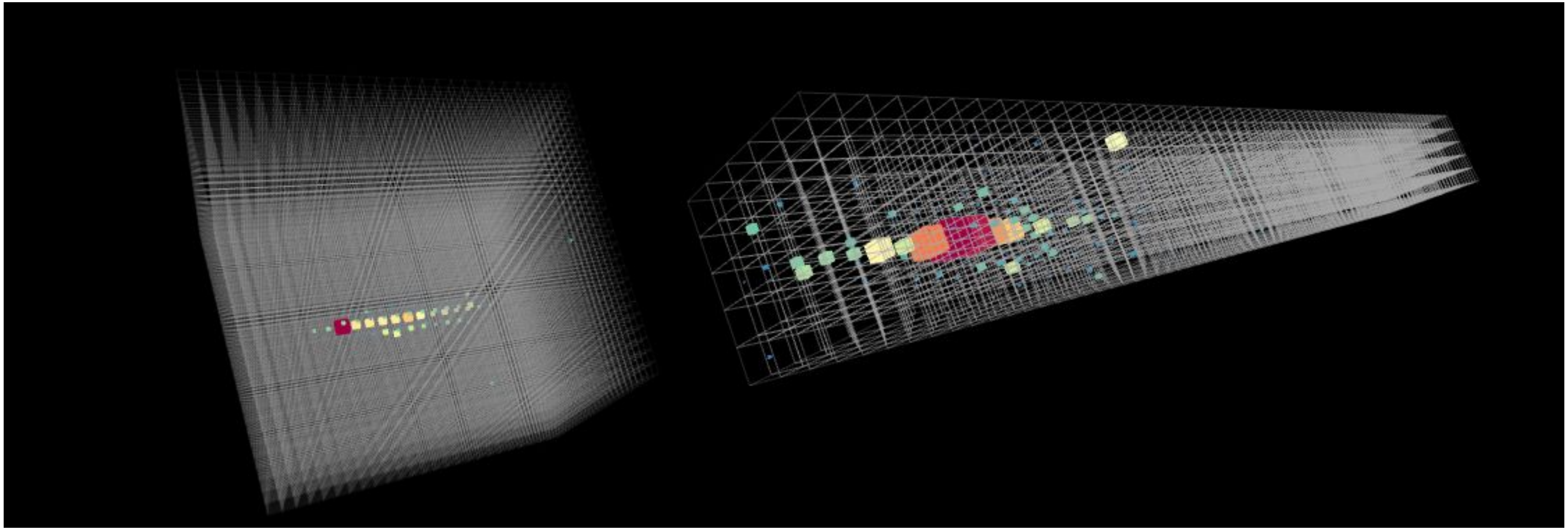




Deep ~~Machine~~ Learning for event processing

Particle reconstruction as image detection

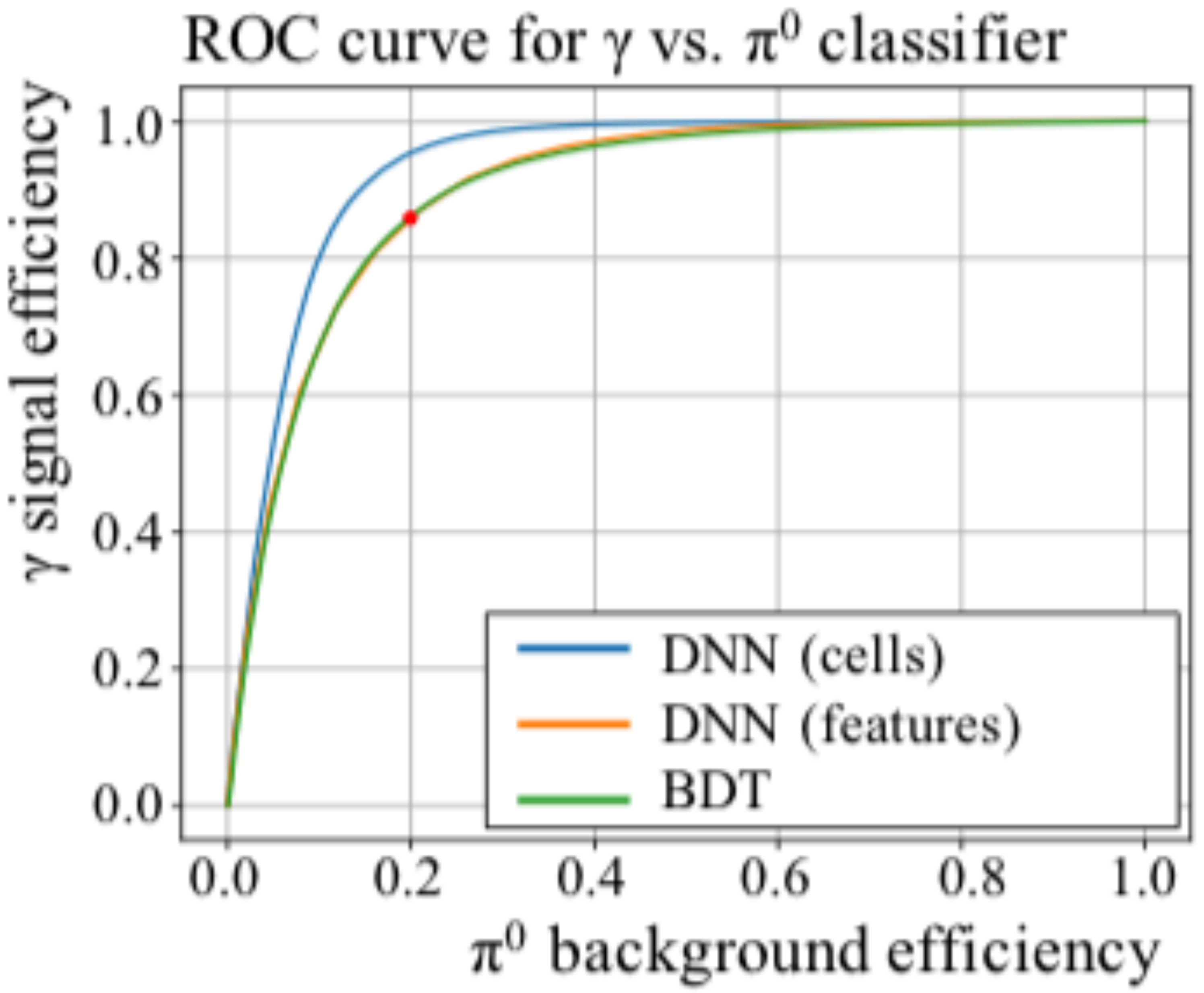
- *Future detectors will be 3D arrays of sensors with regular geometry*
- *It would be ideal to quickly reconstruct particles directly from the image (which is what Deep Learning became famous for)*

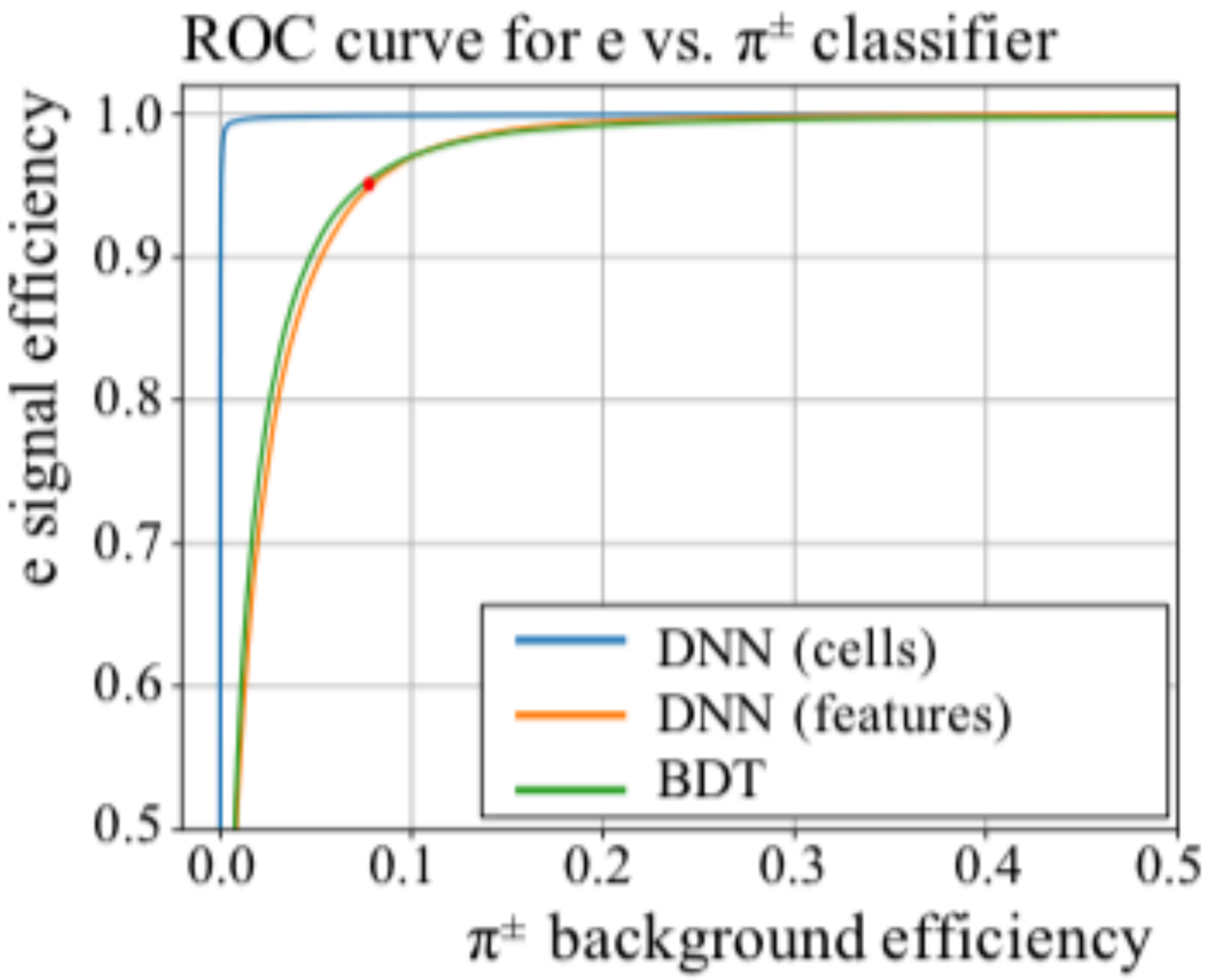


See contribution to NIPS workshop

Proof of Principle: Particle ID

- *We tried particle ID on a sample of simulated events*
  - *one particle/event (e, γ, π^0, π)*
- *Different event representations*
  - *high-level features related to event shape (moments of X,Y, and Z projections, etc)*
  - *raw data (energy recorded in each cell)*
- *Pre-filtered pion events to select the nasty ones and make the problem harder*

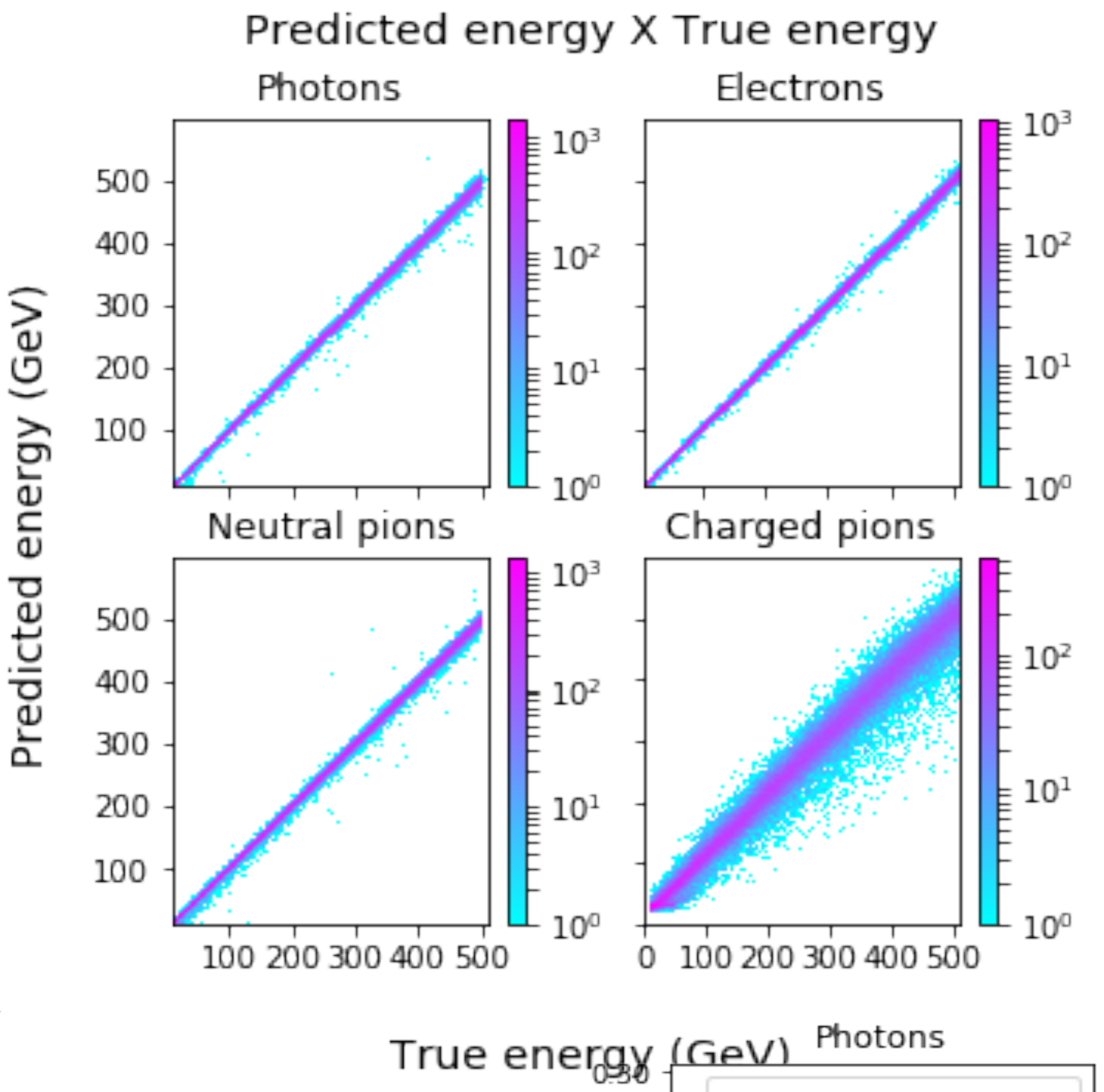


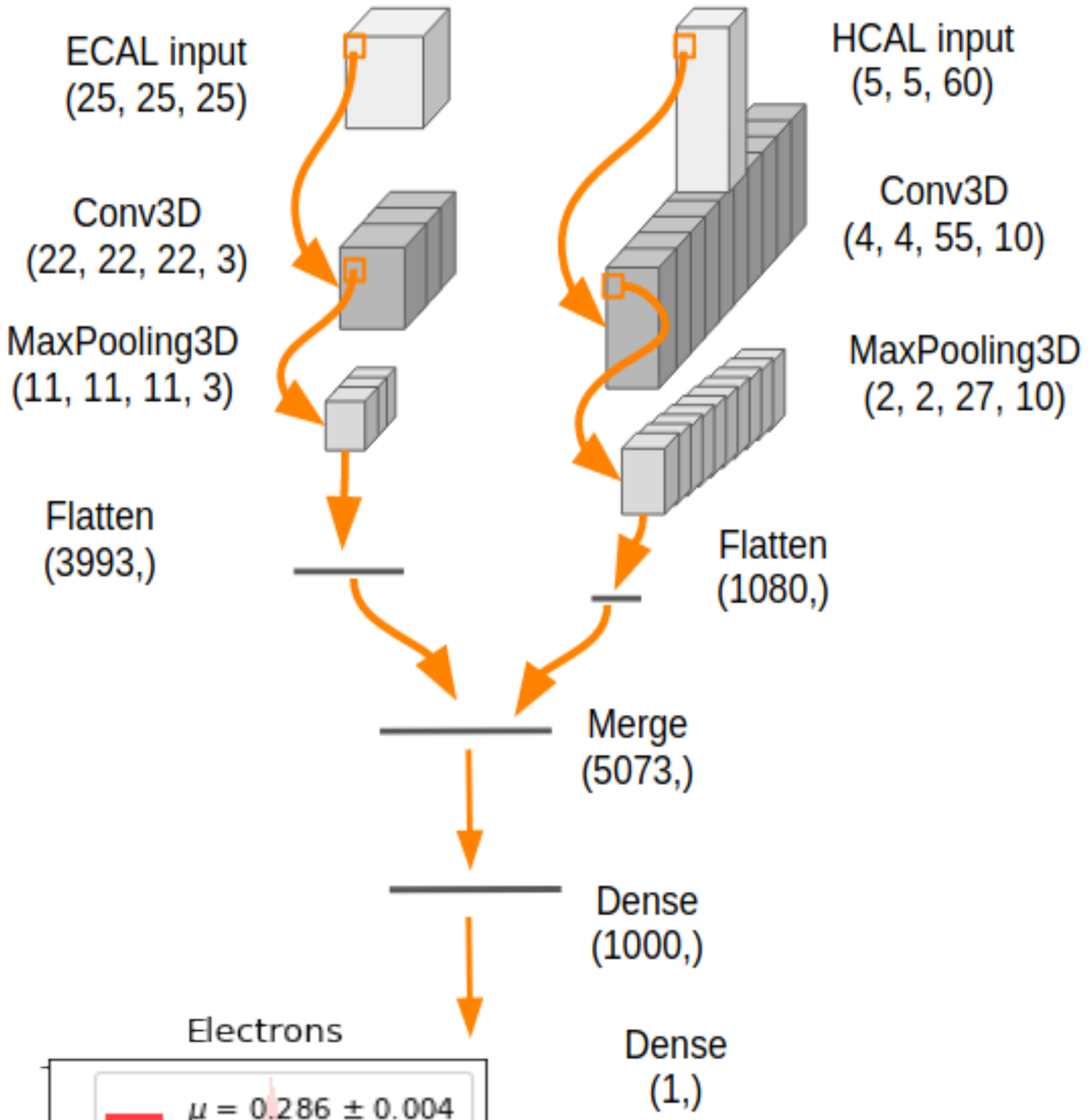


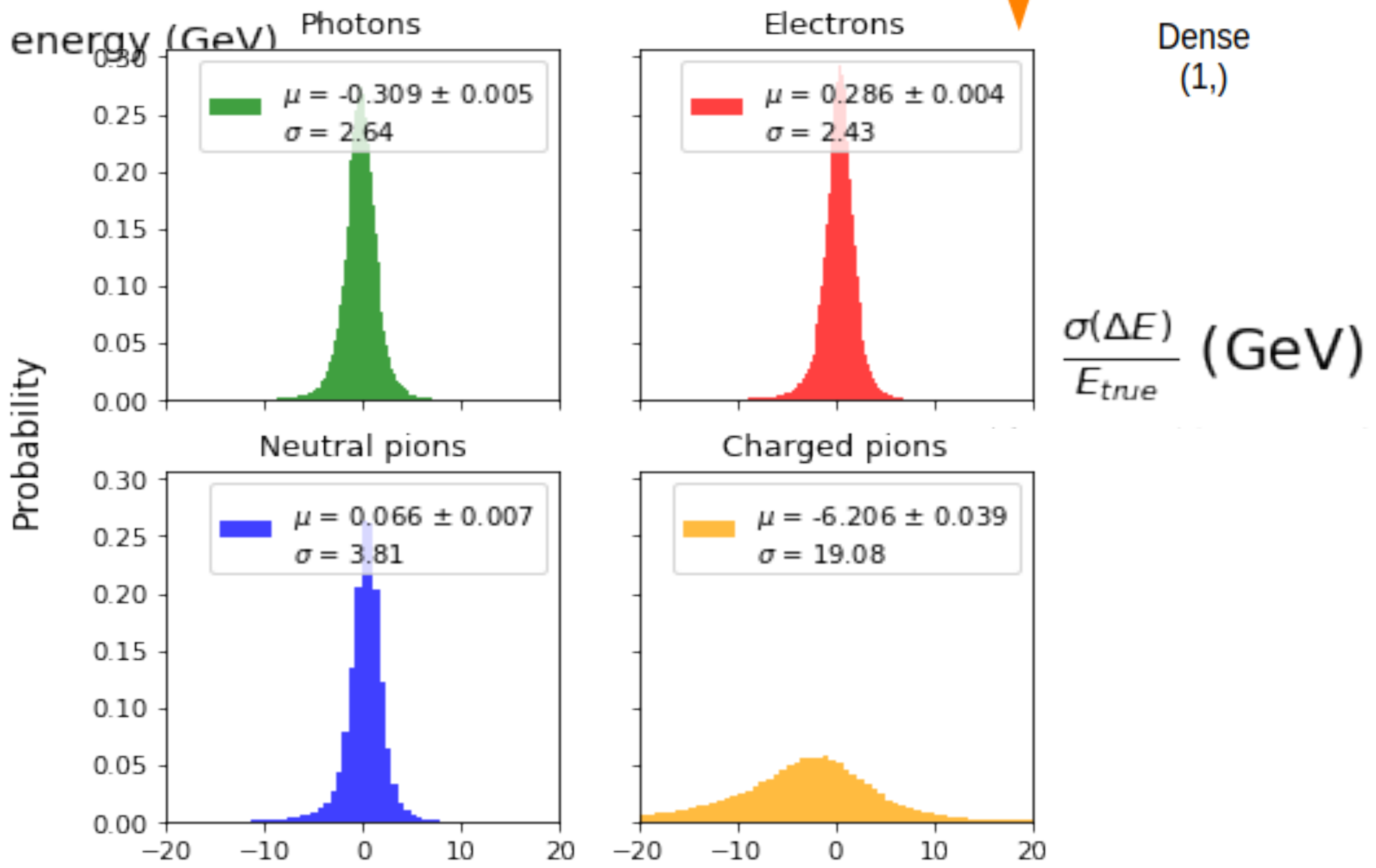
See contribution to NIPS workshop

Proof of Principle: Energy Regression

- *3D Convolution NN can learn true energy of an incoming particle from the recorded hit pattern*
  - *Correctly reconstruct energy*
  - *ECAL performances better than HCAL (as expected)*
  - *π^0 resolution ~ $\sqrt{2}$ γ resolution (as expected)*
- *No high-level knowledge of physics and/or detector features*
  - *used only RAW data as inputs*
- In real life, this could be used offline, at HLT, and (maybe) even at L1



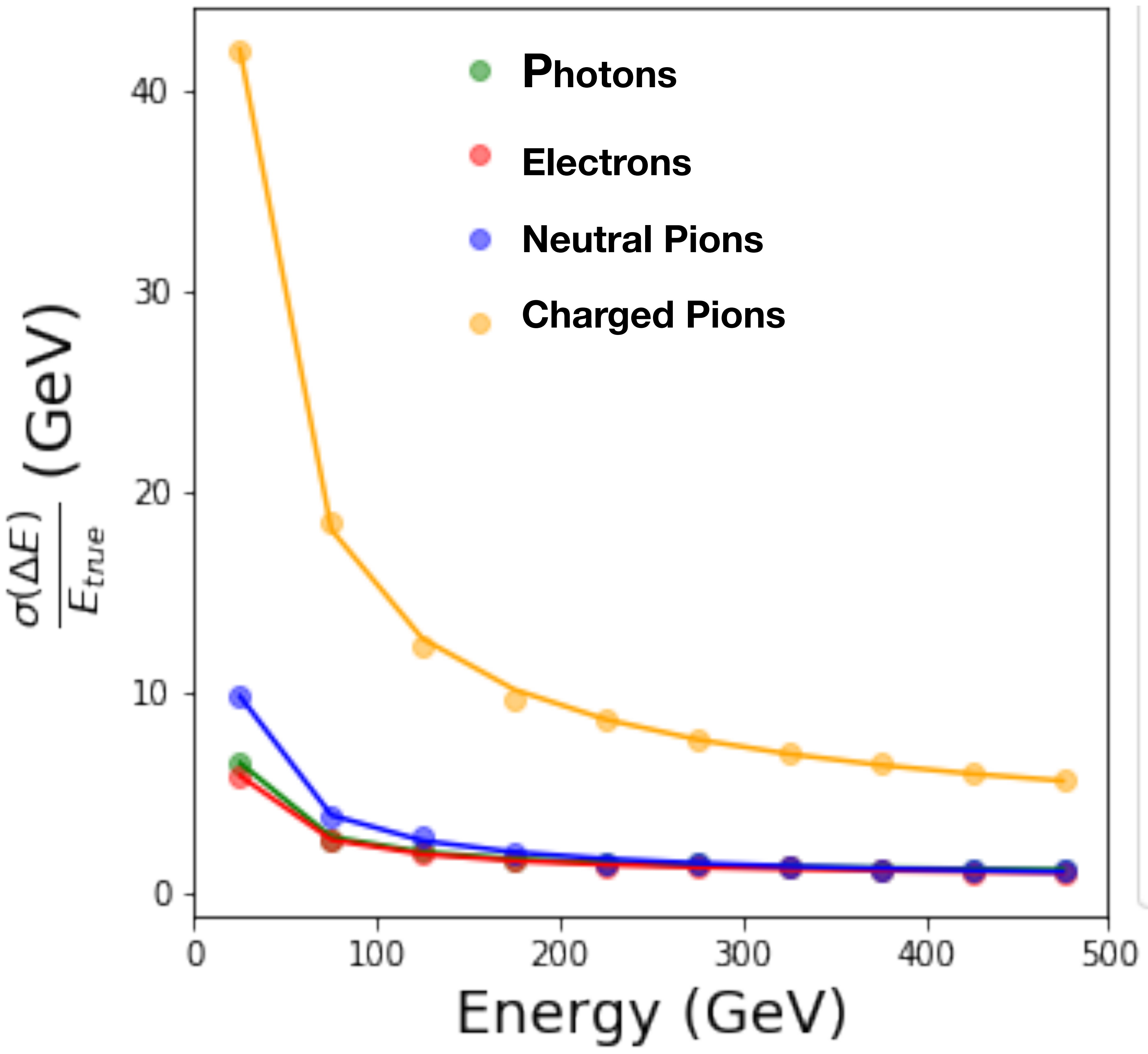




Proof of Principle: Energy Regression

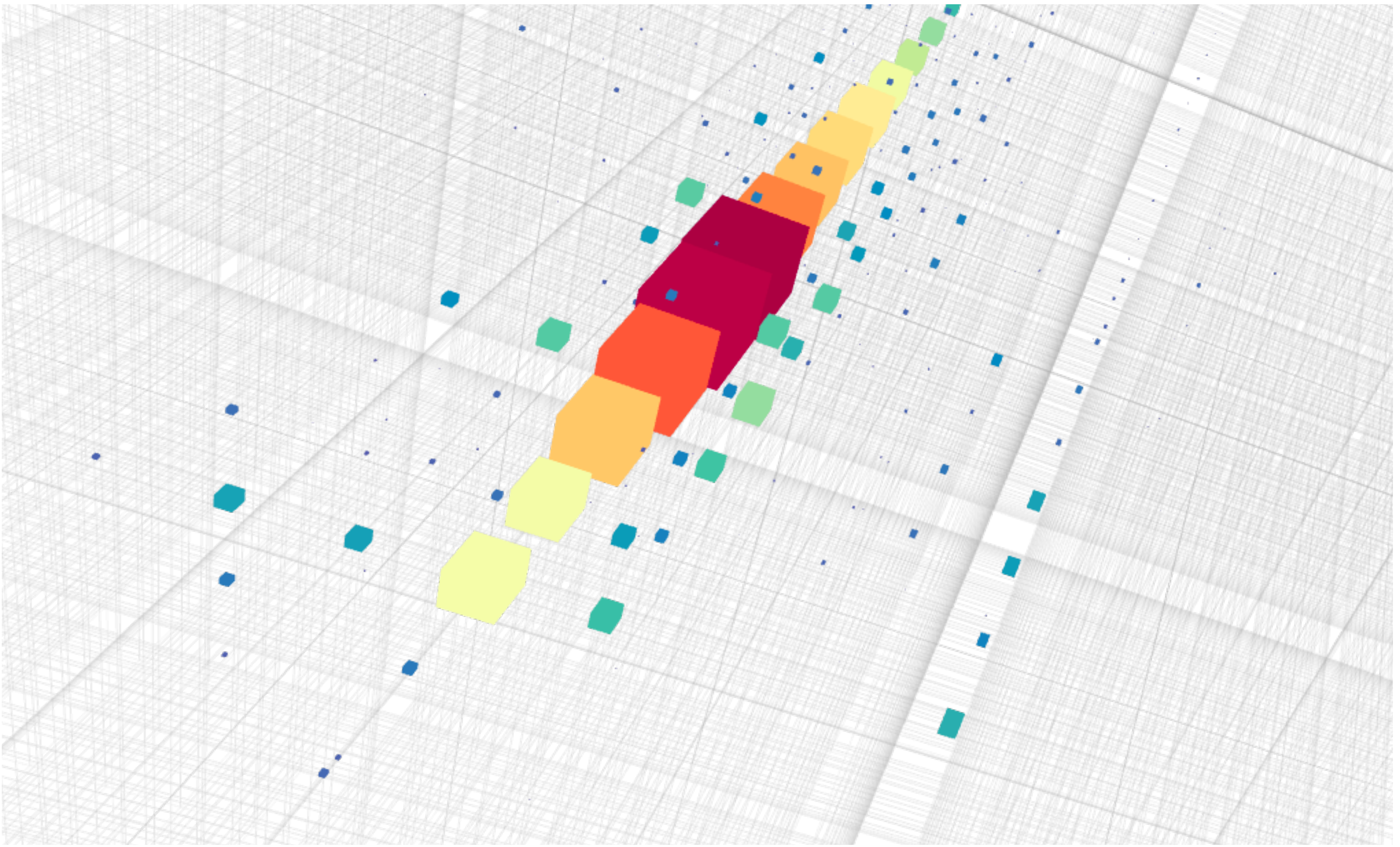
- *Competitive and meaningful results*
- *Processing time reduced by 10^3 wrt traditional approaches*

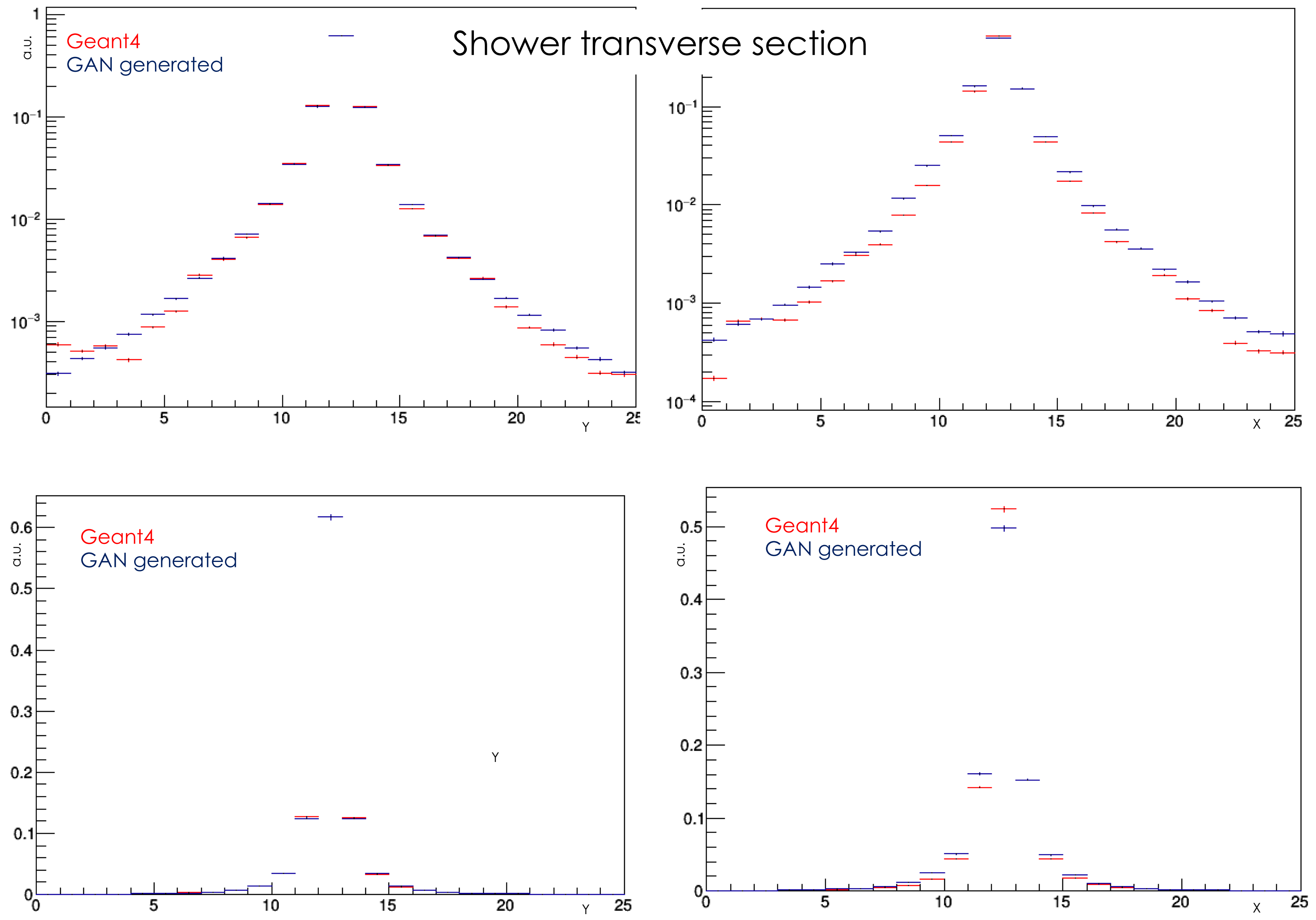
◉ In real life, this could be used while selecting events in real time ("trigger")

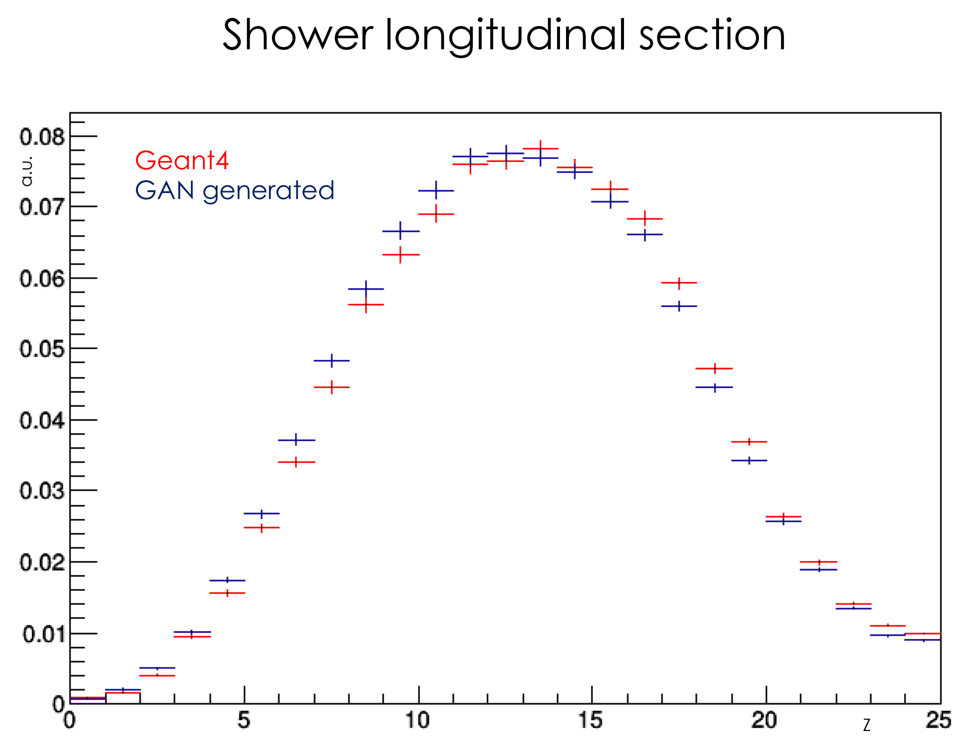


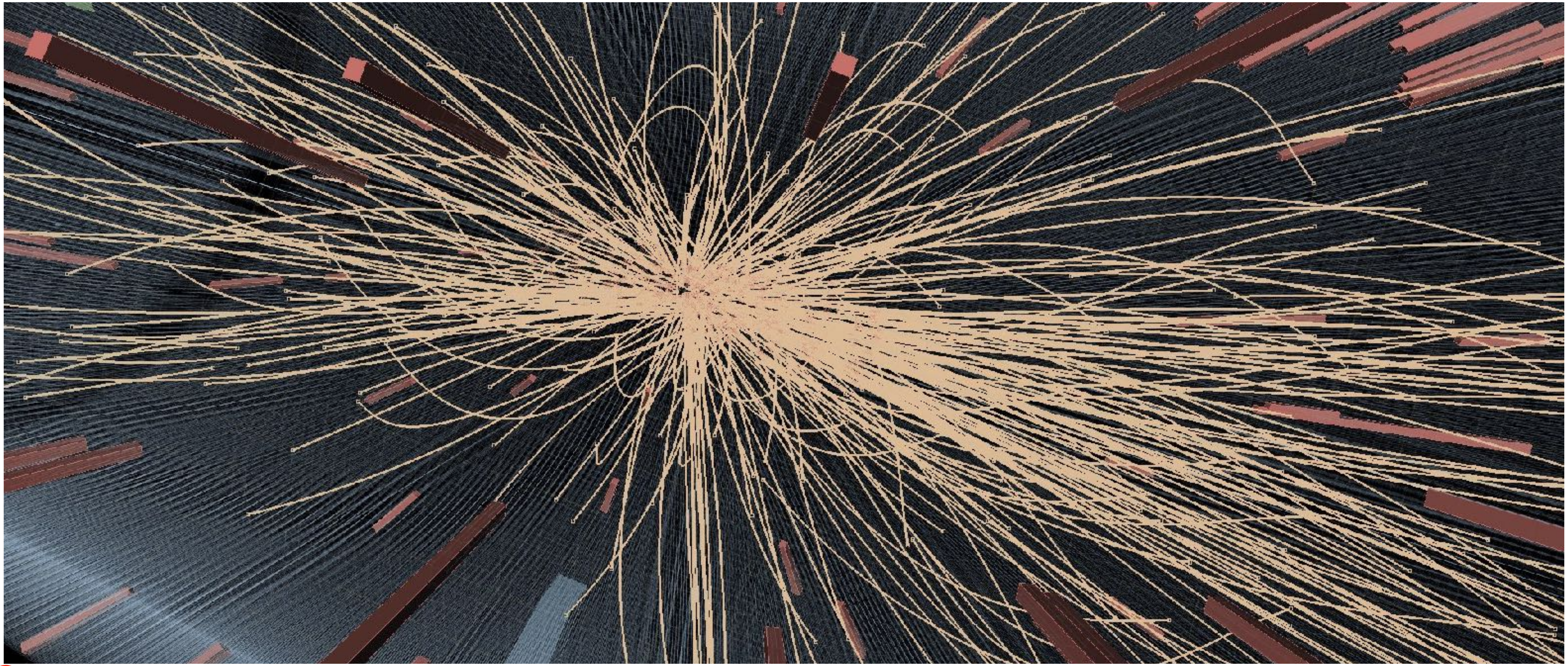
Generative Adversarial Networks

- With x10 more data being stored during HL-LHC, we will need > x10 more Monte Carlo to do precision physics
- This will not be possible with current generation techniques
- Generative models might provide a way out of this dead end







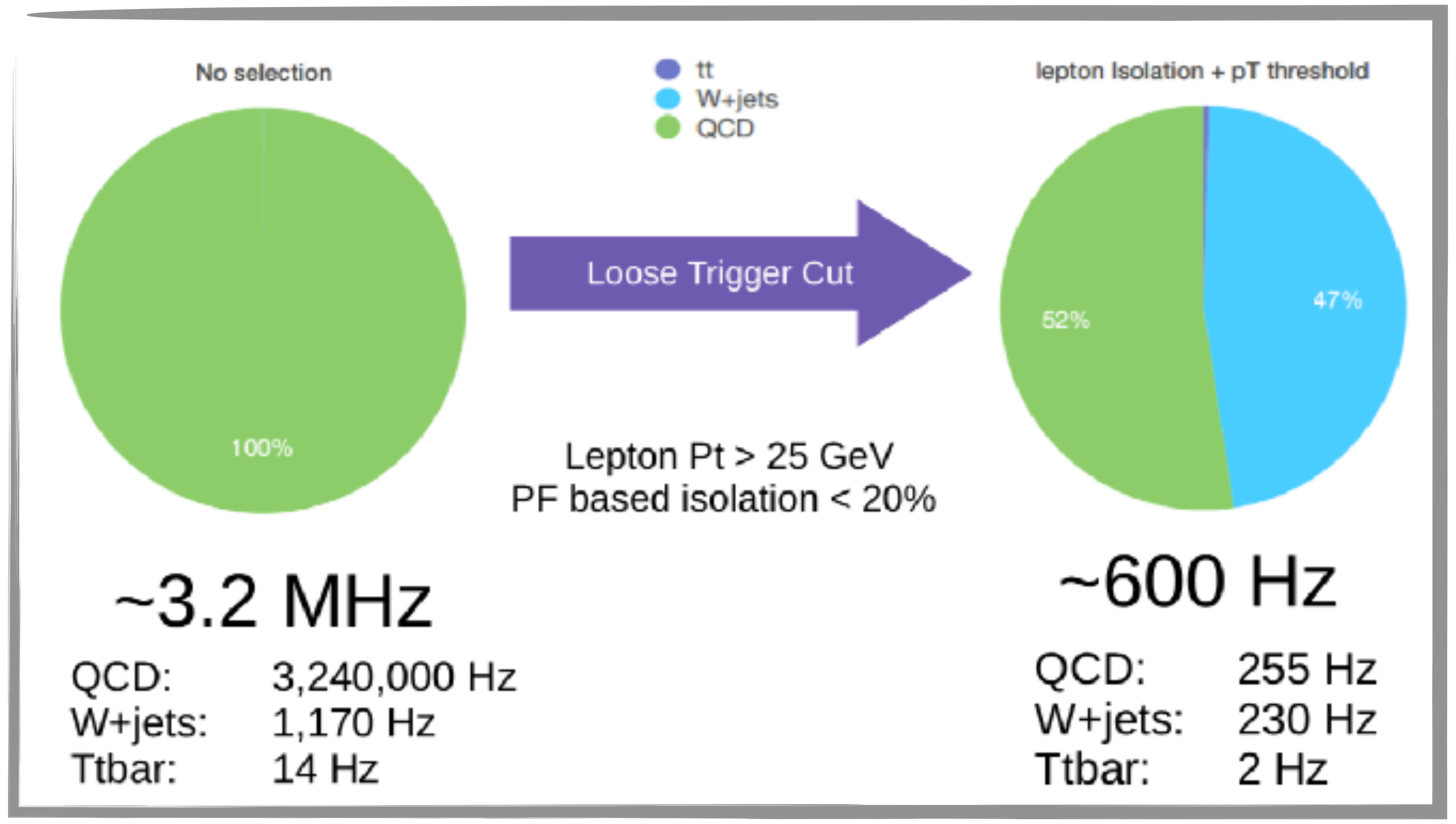


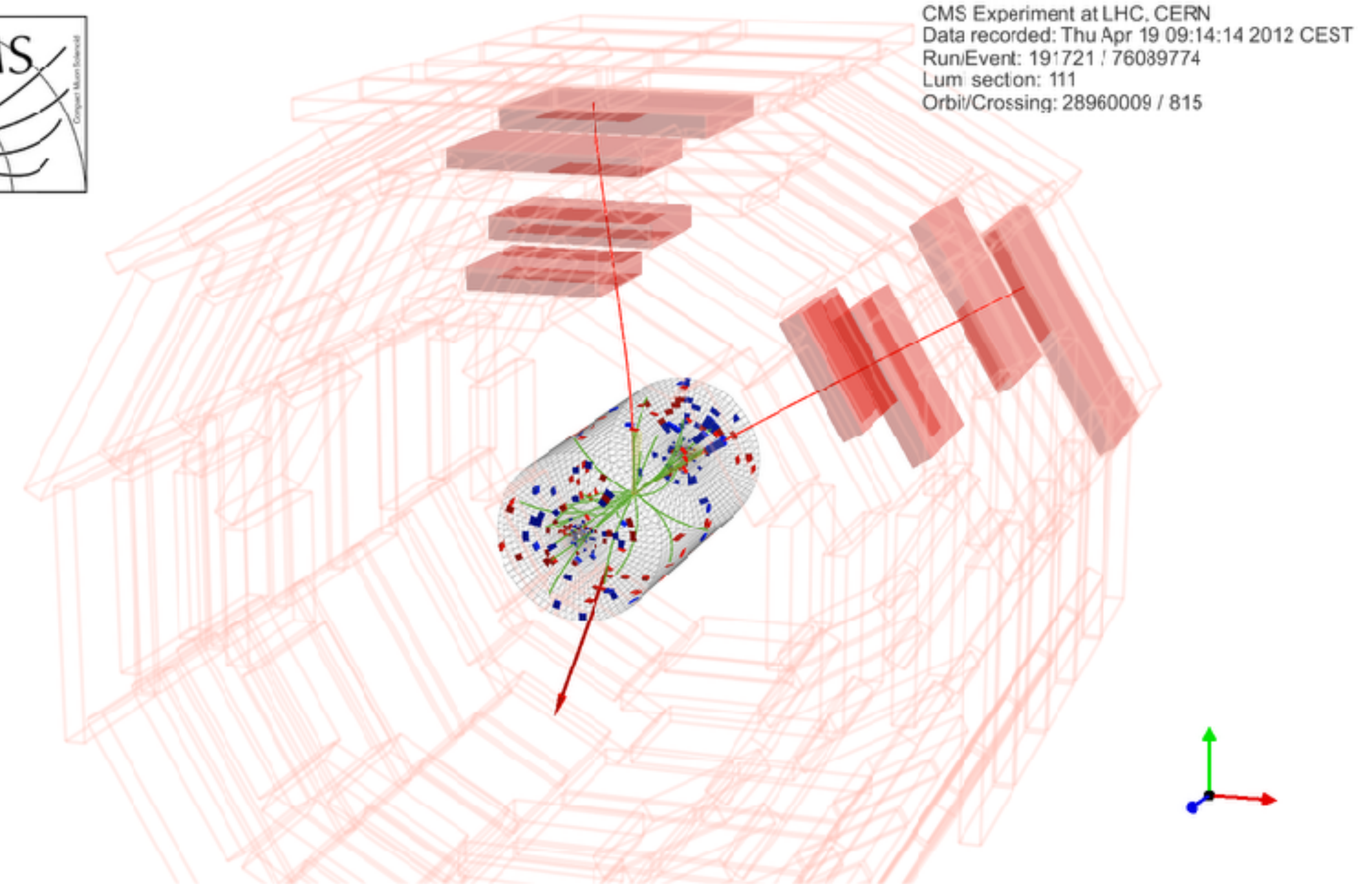
Deep ~~Machine~~ Learning for data taking

Cleaning up selected sample

A typical example: leptonic triggers

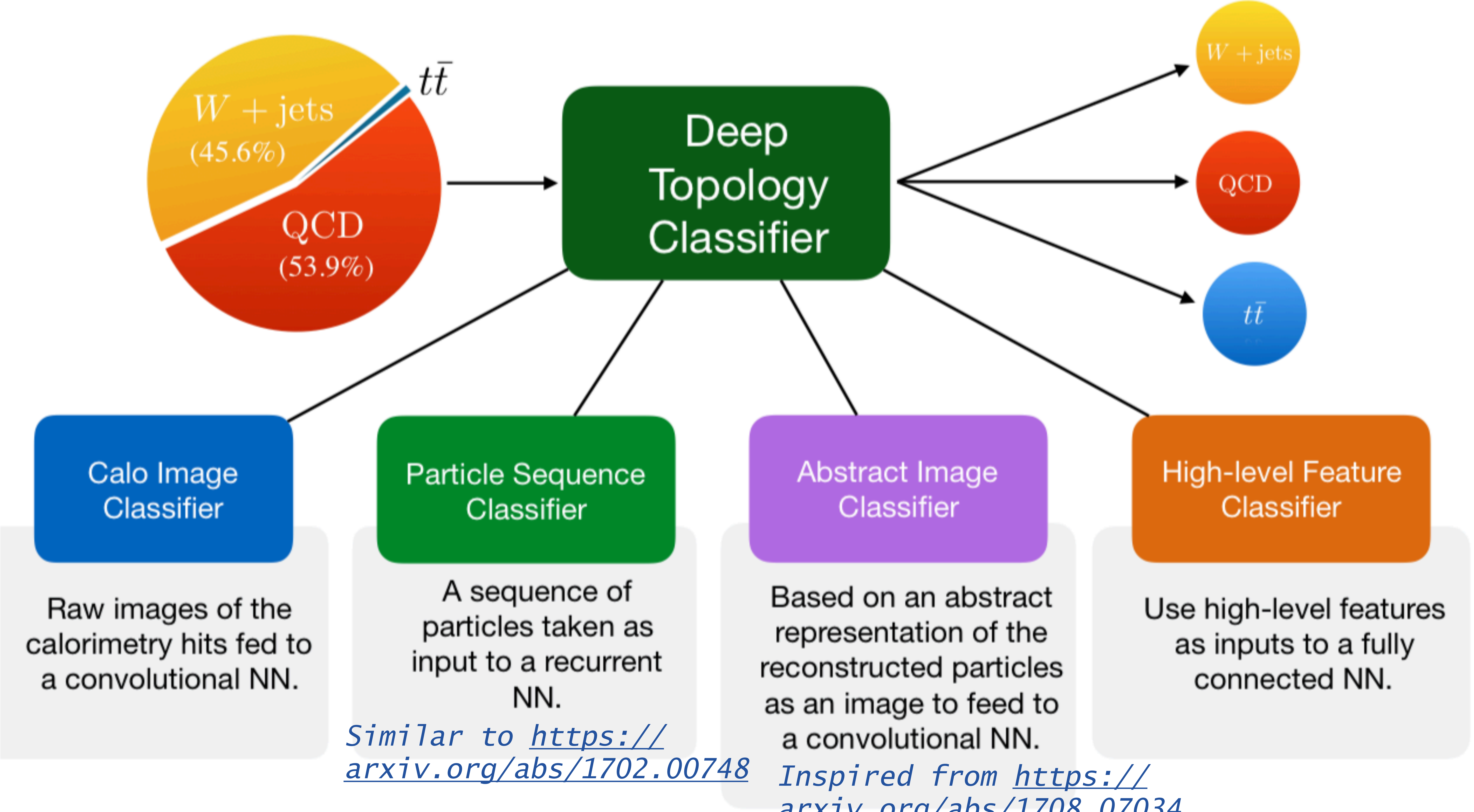
- *at the LHC, producing an isolated electron or muon is very rare. Typical smoking gun that something interesting happened (Z,W,top,H production)*
- *Triggers like those are very central to ATLAS/CMS physics*
- *The sample selected is enriched in interesting events, but still contaminated by non-interesting ones*
- *Contamination can be reduced with a DL classifier that rejects obvious false positives looking at the full event, not just at the lepton*





See contribution to NIPS workshop

Event Representations

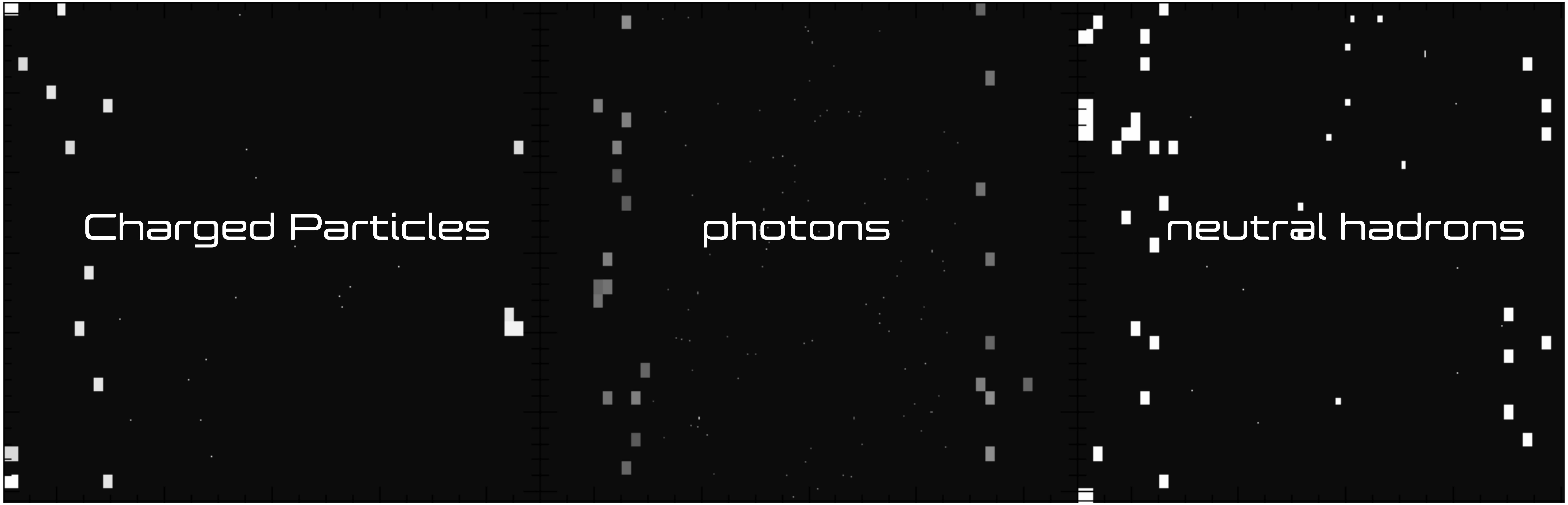


Similar to https://arxiv.org/abs/1702.00748

Inspired from https://arxiv.org/abs/1708.07034

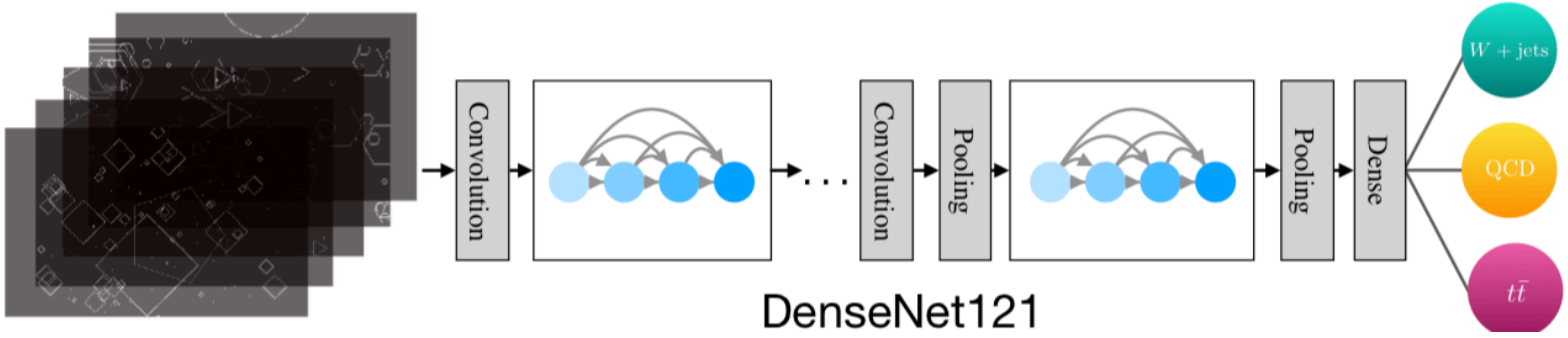
What the event looks like

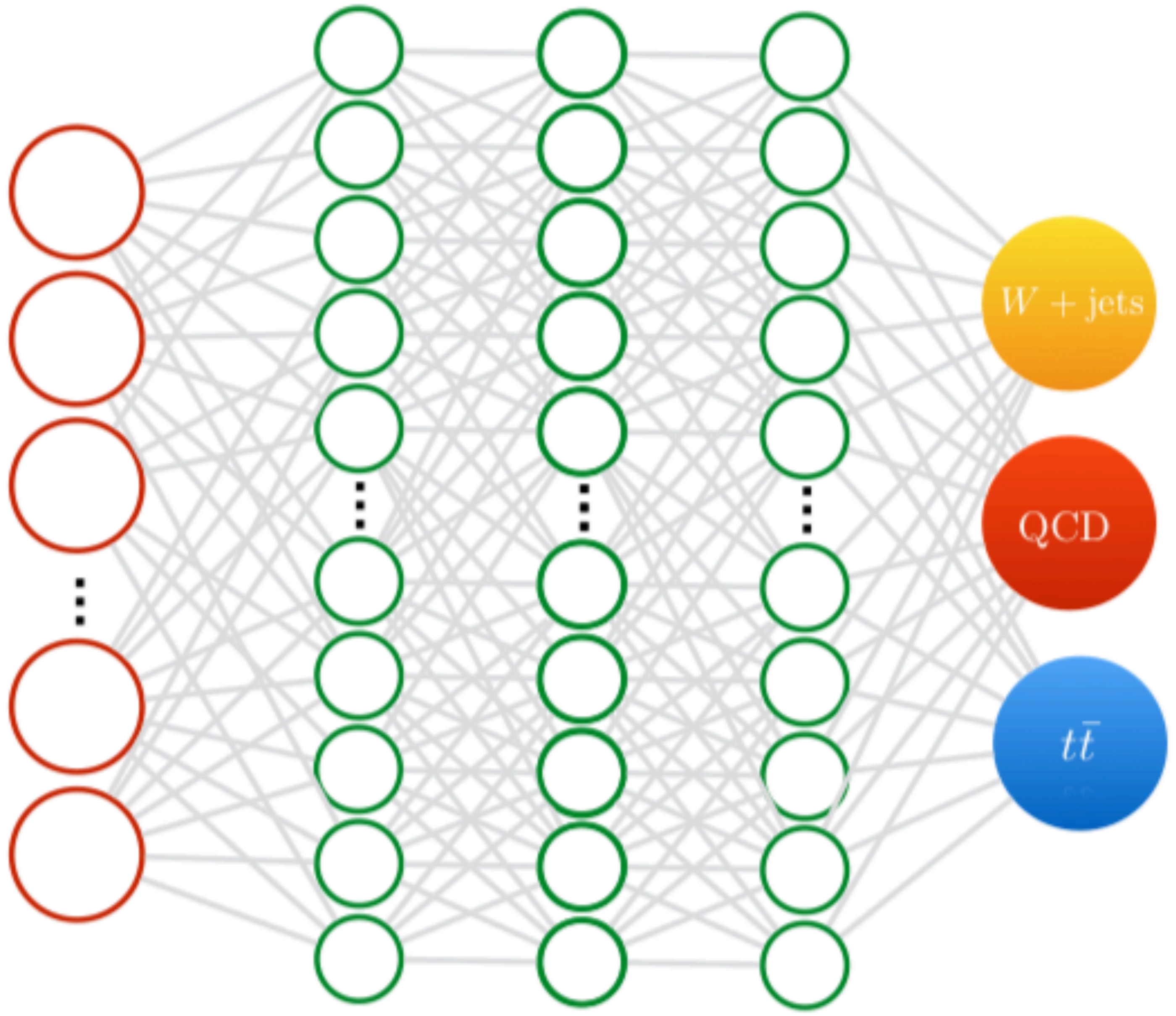
- *sparse image with many pixels*
  - *not the kind of image that CNNs usually deal with*
  - *still, reasonable performances (AUC~90%) can be obtained*



Event Representations

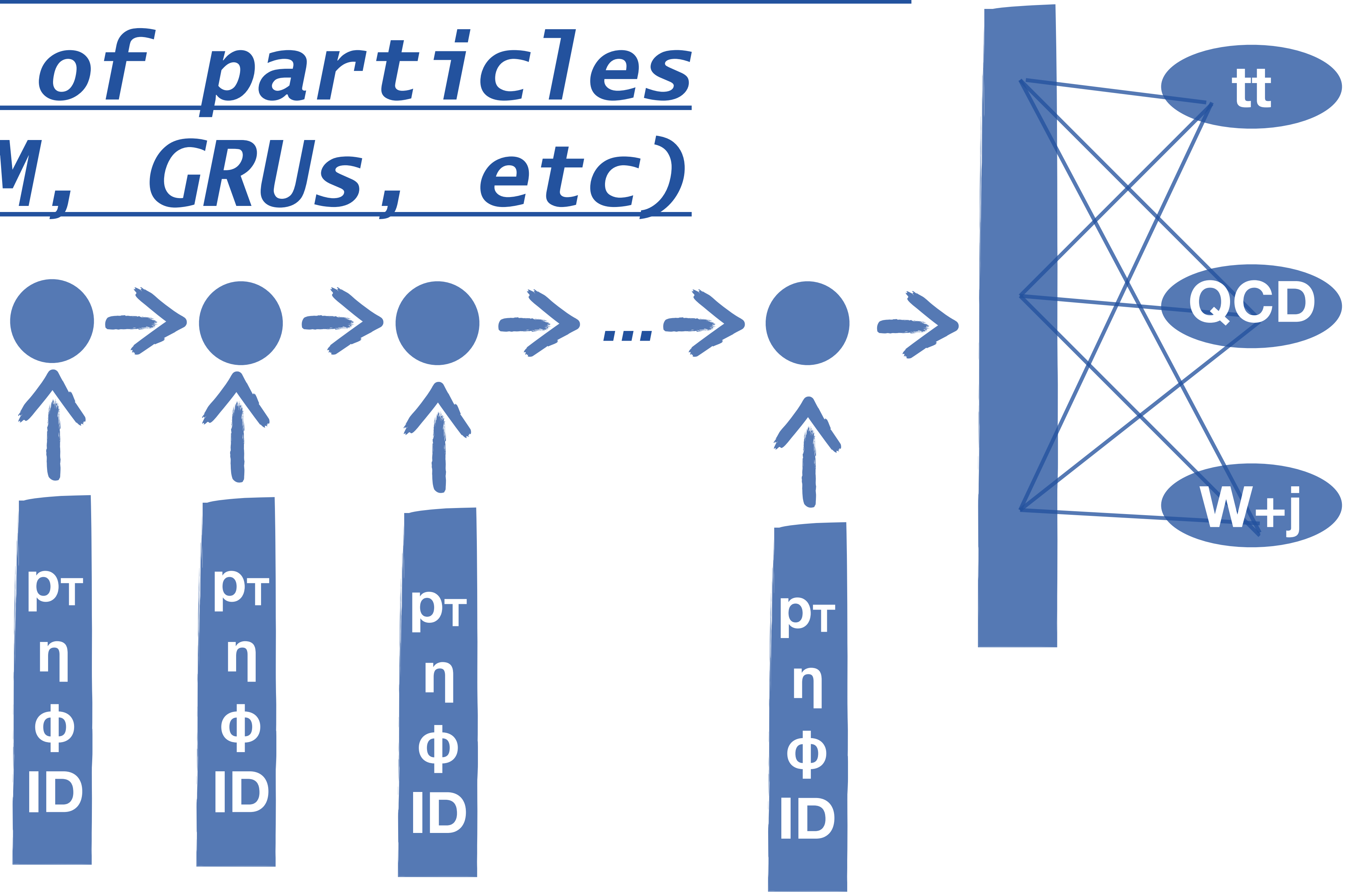
DenseNet on the abstract image





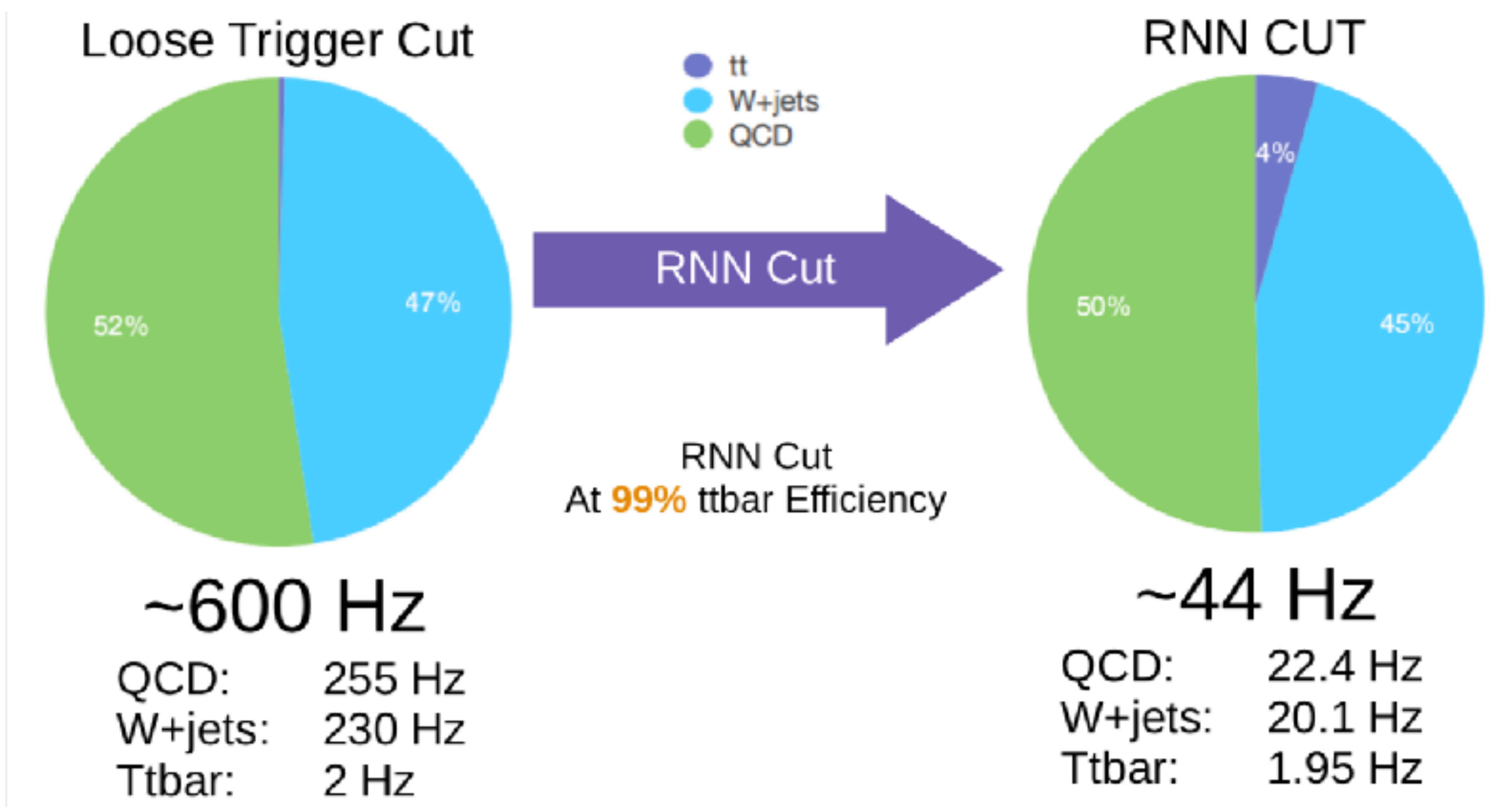
Fully-Connected classifier on physics-motivated features

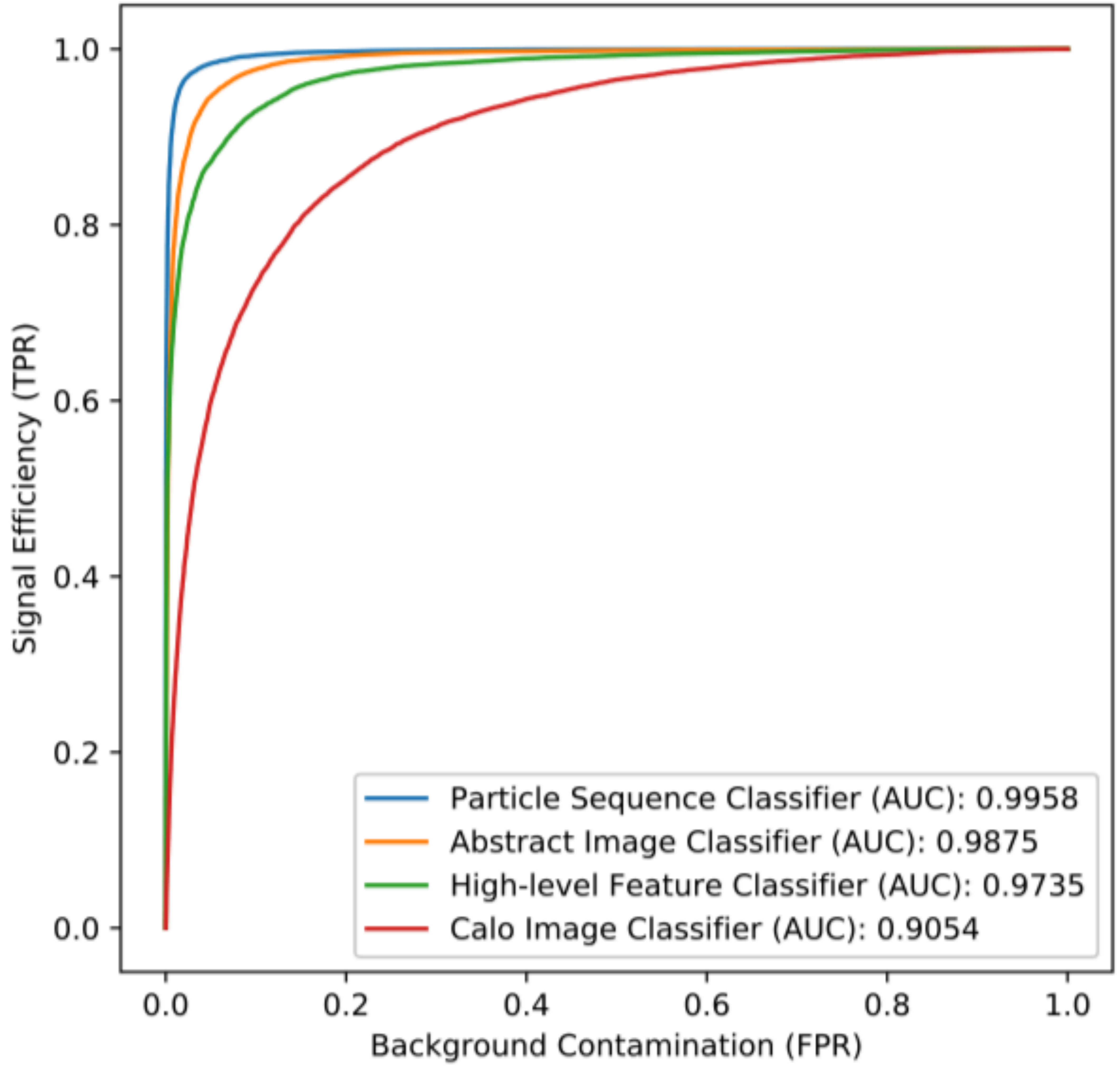
Recurrent nets on the list of particles (LSTM, GRUs, etc)



Proof of principle: trigger cleanup

- *With non-trivial event representation, can drop false positives by factor 10 with 99% true-positive rate*

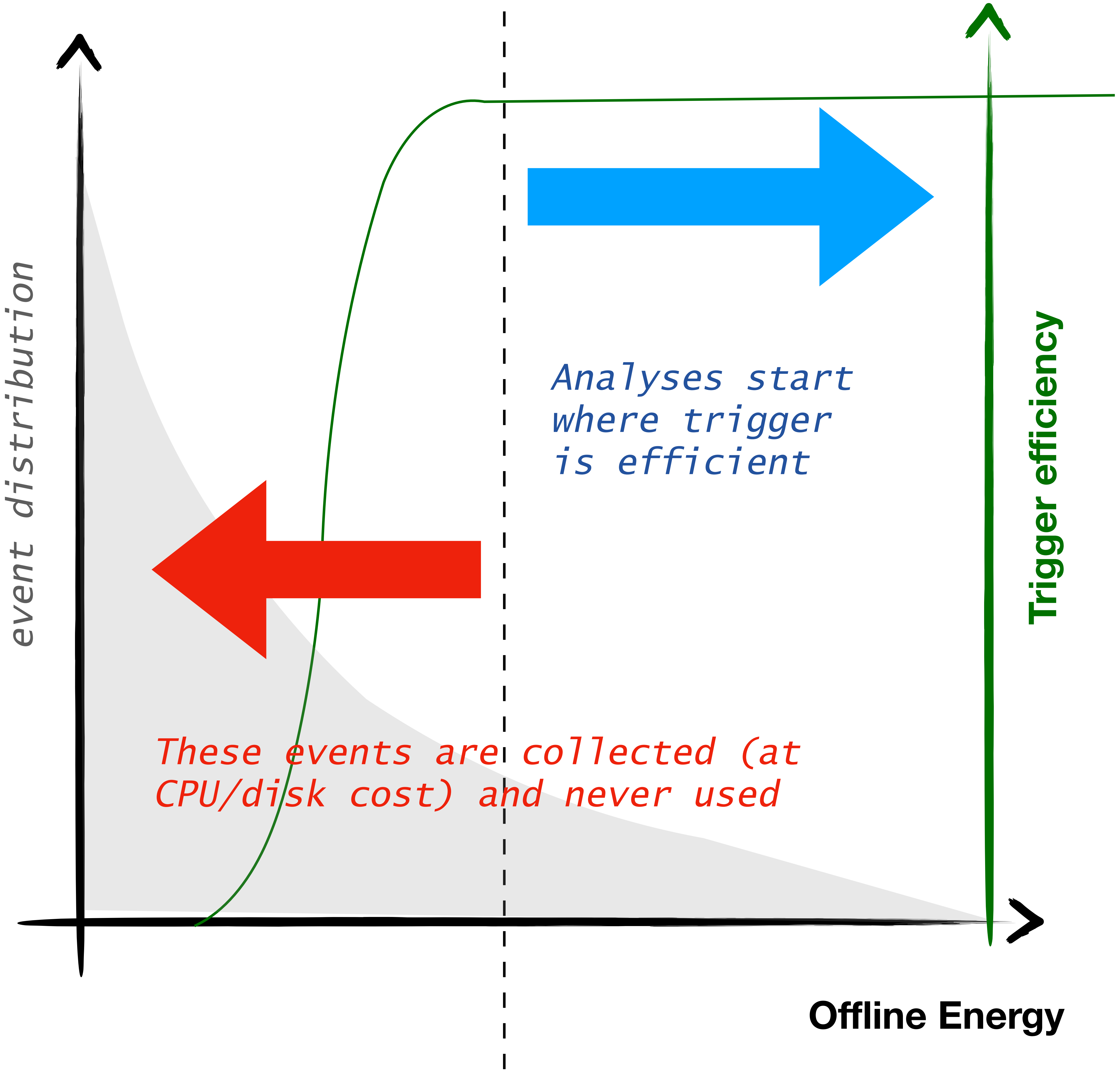




See contribution to NIPS workshop

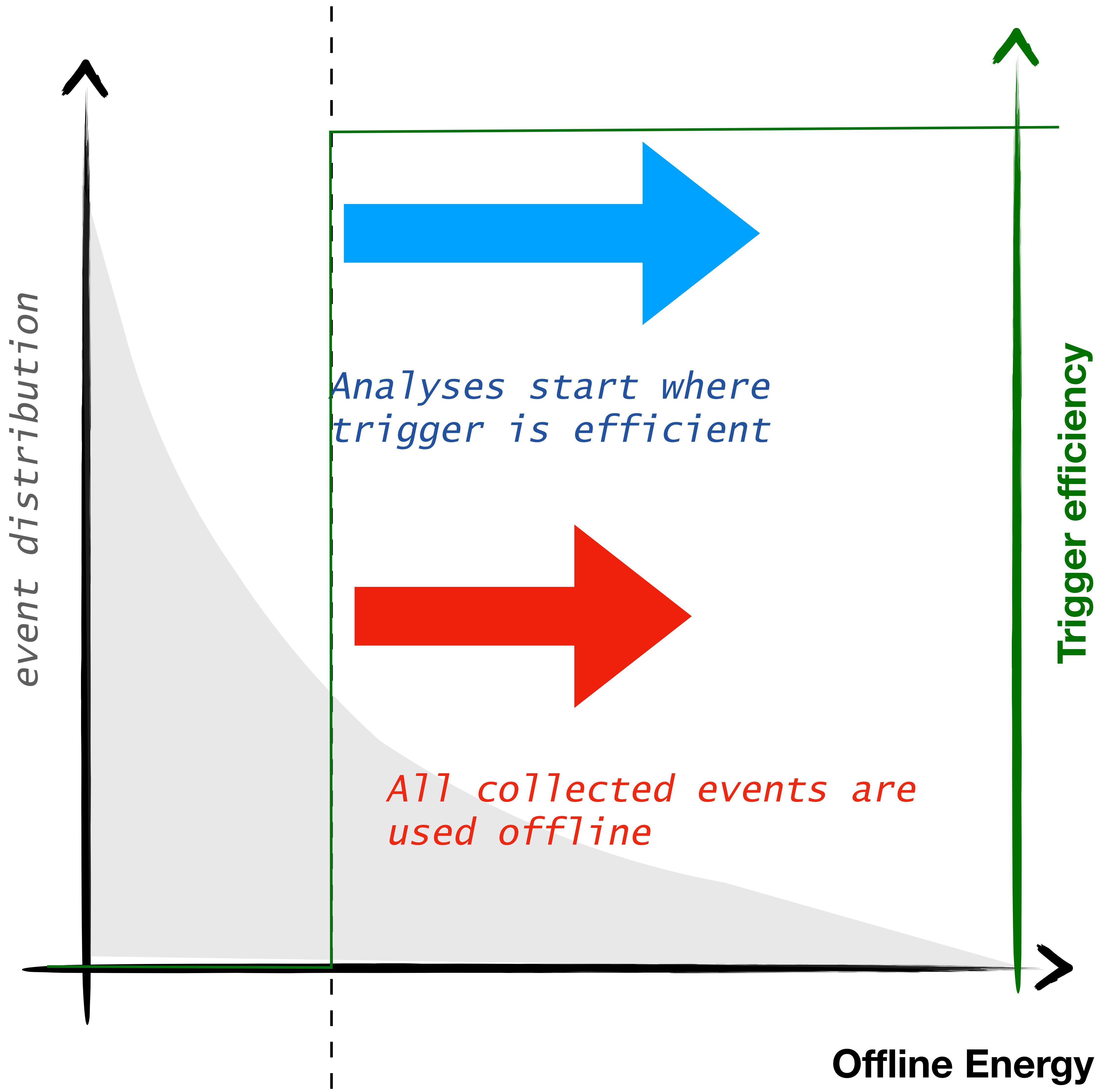
The importance of being fast

- *Online vs offline reconstruction differences are limiting our discovery reach*
  - *Seen offline, the online selection is a not-flat response function*
  - *Forces us to work on tails of event distribution, reducing sensitivity to new physics*
  - *Not optimal use of resources*



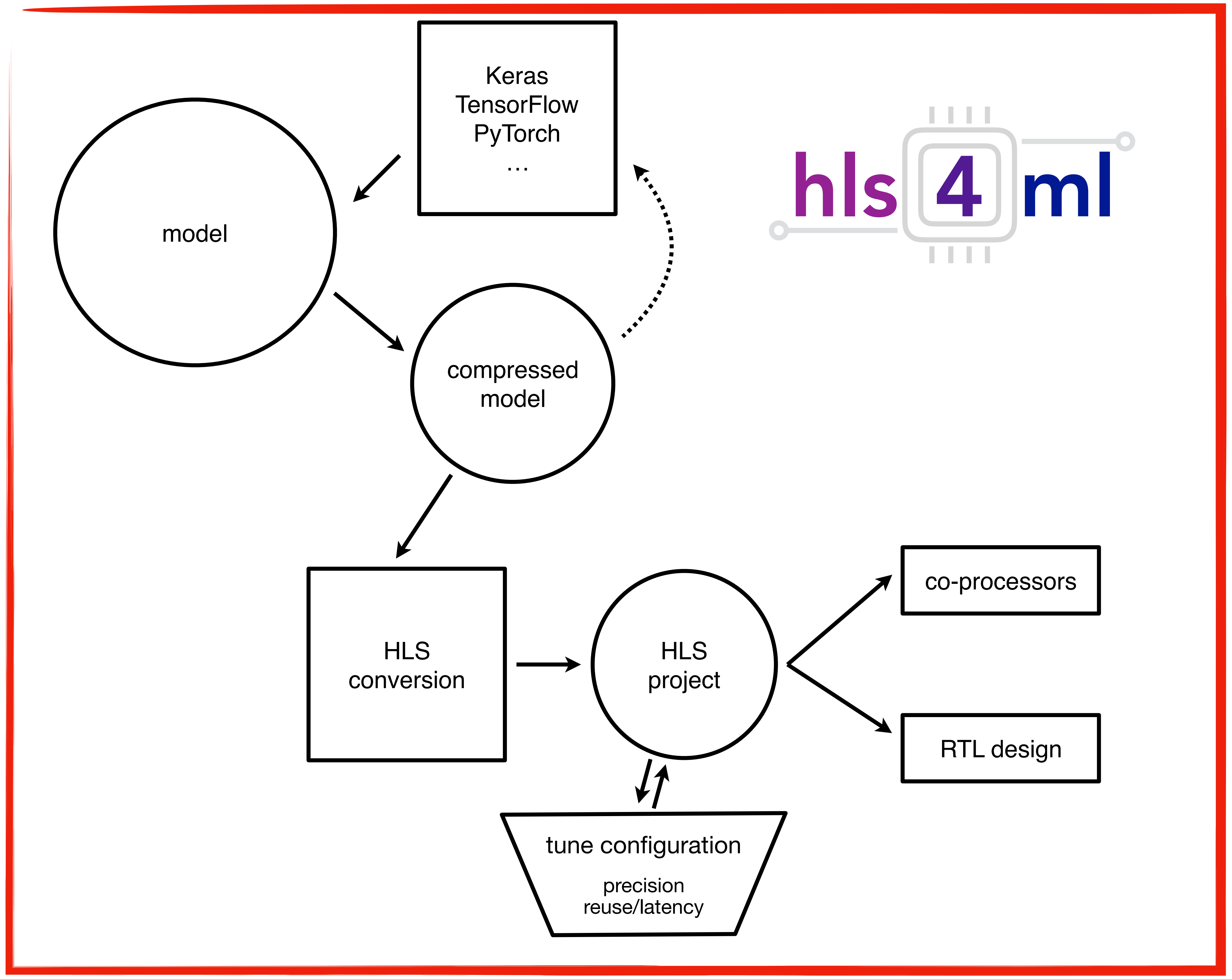
The importance of being fast

- ***Having the same reconstruction at L1/HLT/Offline*** *would help us to recover this lost sensitivity …*
- *… and to free resources that could be spent otherwise (e.g., looking for tricky new physics scenarios)*
- *This cannot be done exactly (offline code too slow)*
- *But it could be done "in average" (offline response modelled by ML algorithm)*



The frontier: bring DL to L1

- The L1 trigger is a complicated environment
  - decision to be taken in ~10 μsec
  - only access to local portions of the detector
  - processing on Xilinx FPGA, with limited memory resources
- Some ML already running @L1
  - CMS has BDT-based regressions coded as look-up tables
- Working to facilitate DL solutions @L1 with dedicated library

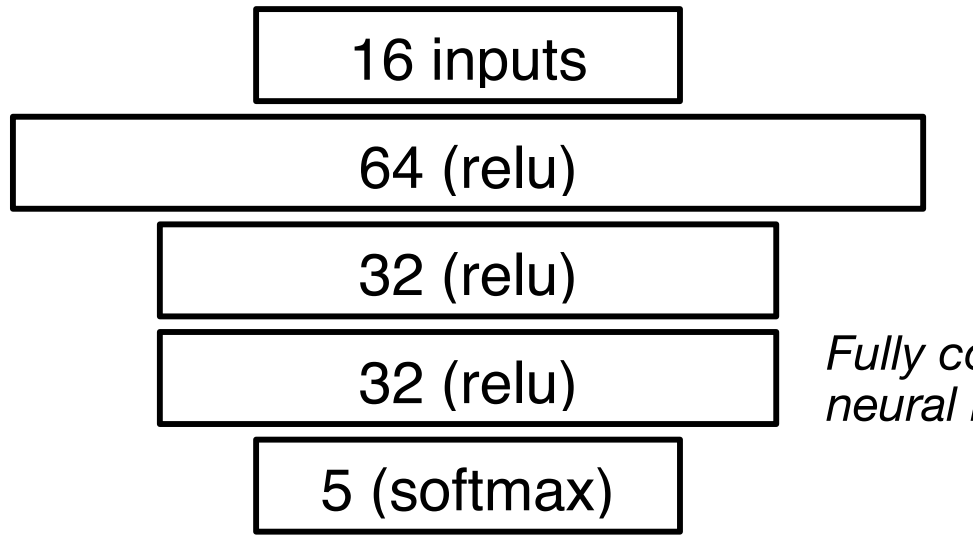


HLS4ML: CERN/FNAL/MIT joint effort

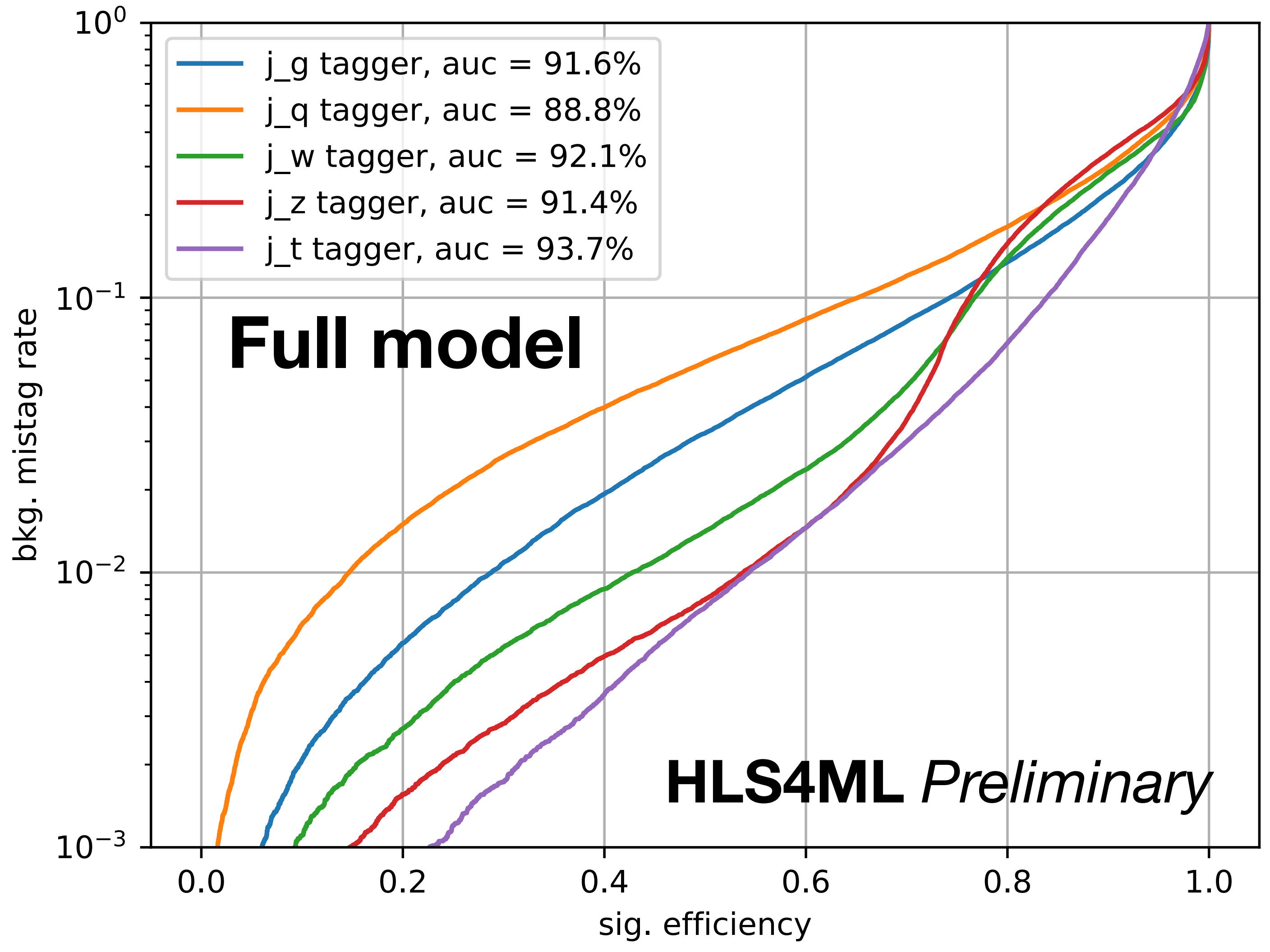
To debut at Connecting The Dots 2018 in Seattle (March 2018)

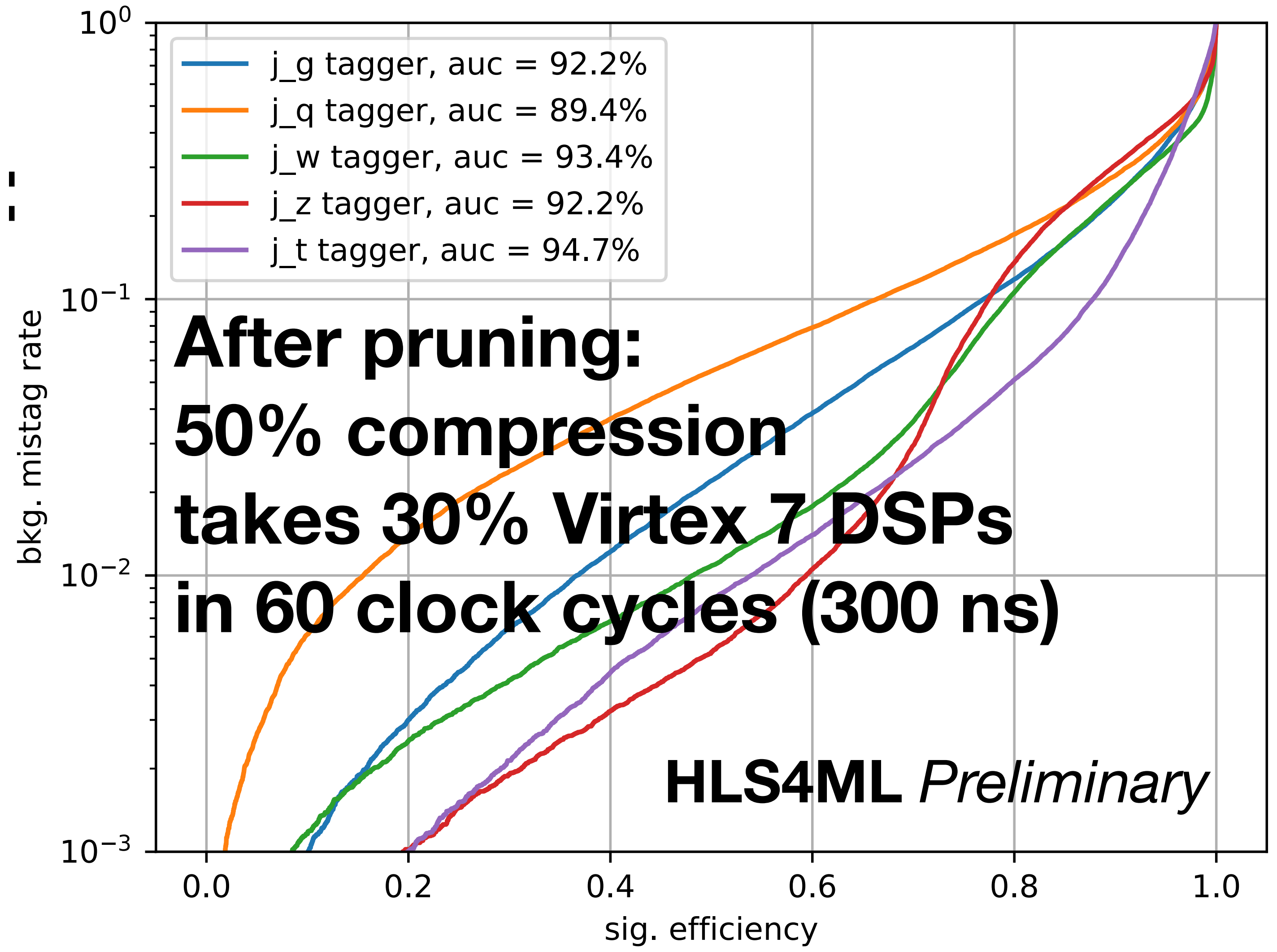
The frontier: bring DL to L1

- *Work is still preliminary*
- *Take as use case jet tagging*
  - *get a jet @L1*
  - *from its shape (jet substructure) tell which jet it is*
- *Problem solved with large network (cannot fit FPGA)*
- *Implementing pruning solutions to keep only relevant neurons*
- *Further decrease resources with practical tricks & thumb rules (e.g., divisions are expensive)*



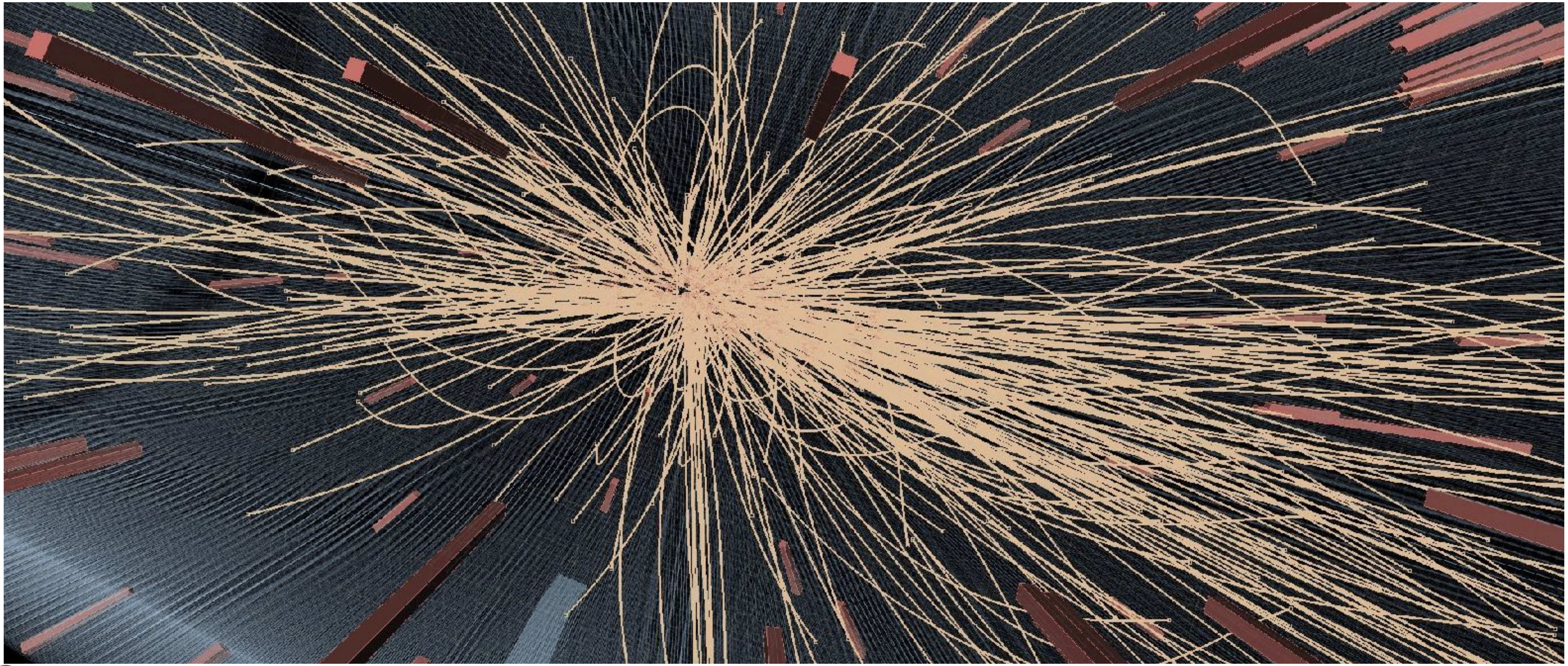
Fully connected deep neural network





HLS4ML: CERN/FNAL/MIT joint effort

To debut at Connecting The Dots 2018 in Seattle (March 2018)

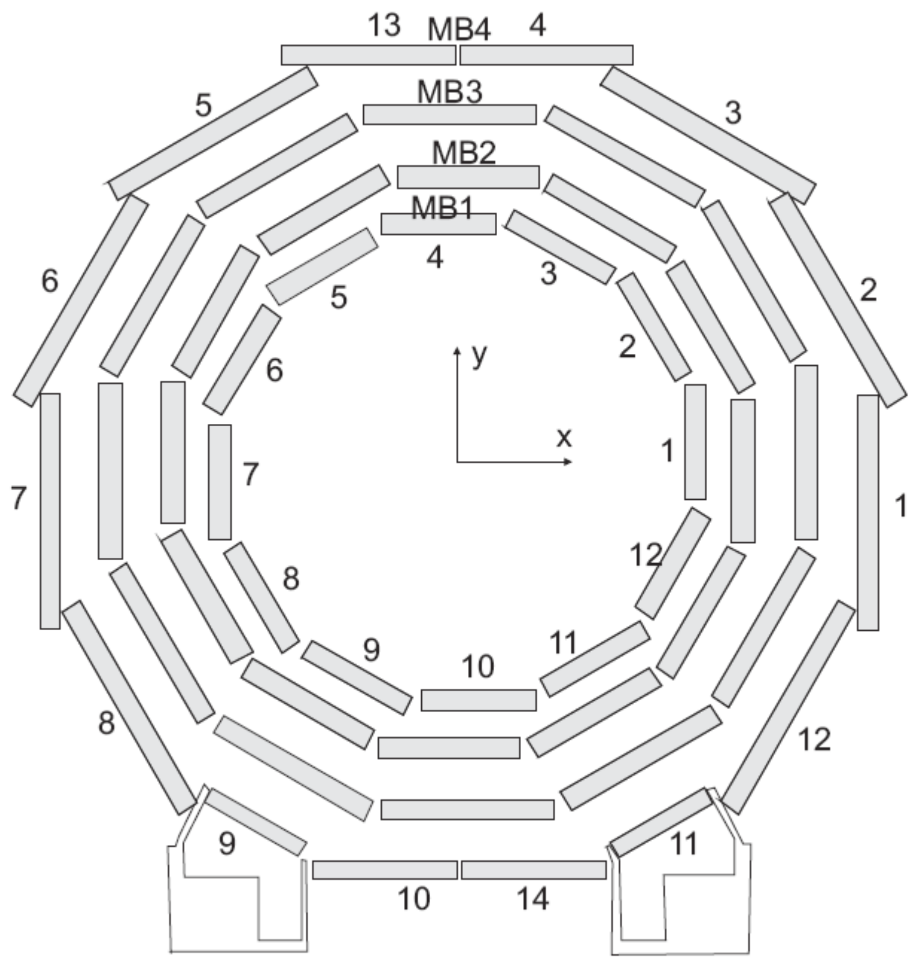


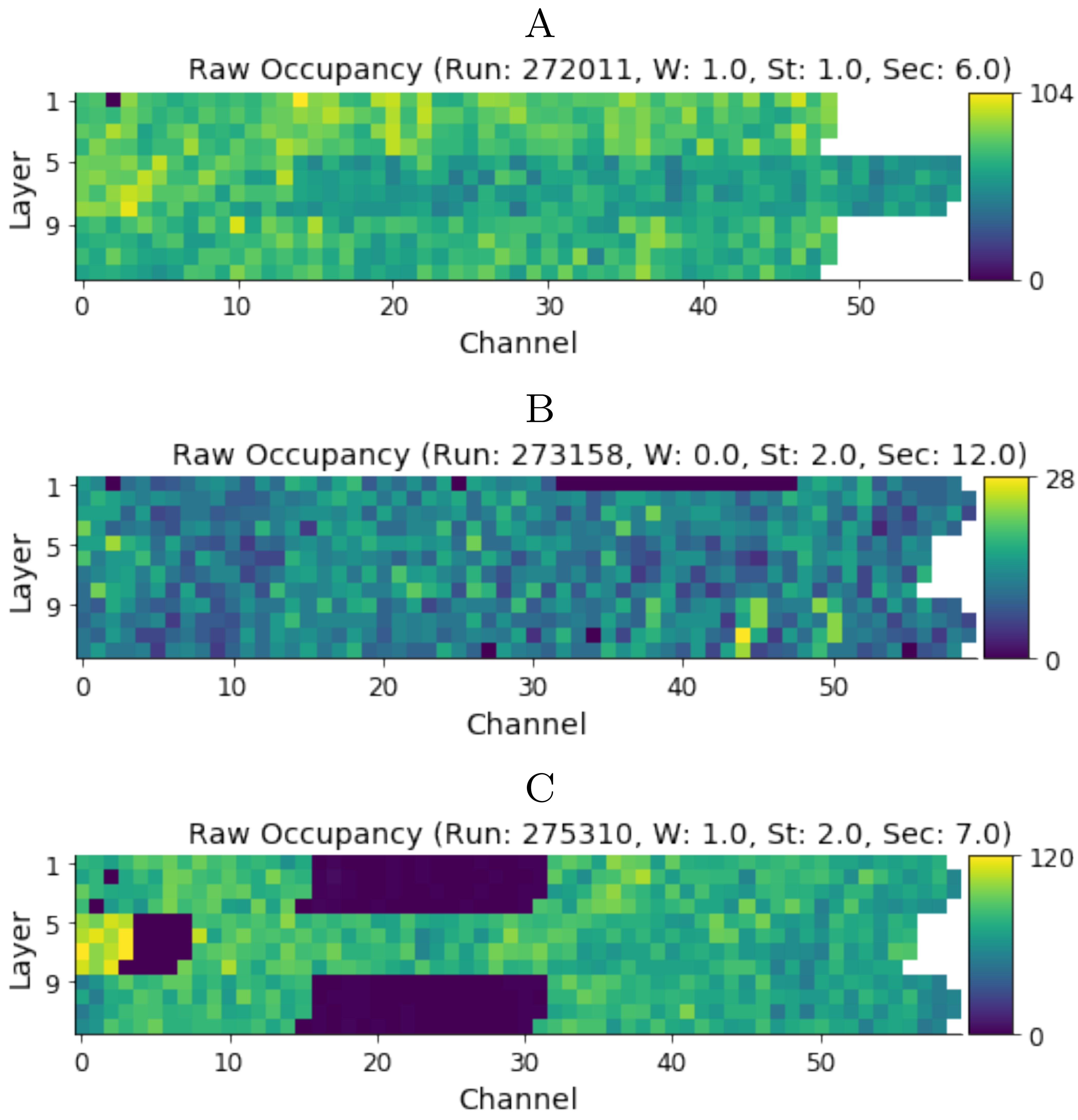
Deep

~~Machine~~ Learning for monitoring

Data Quality Monitoring

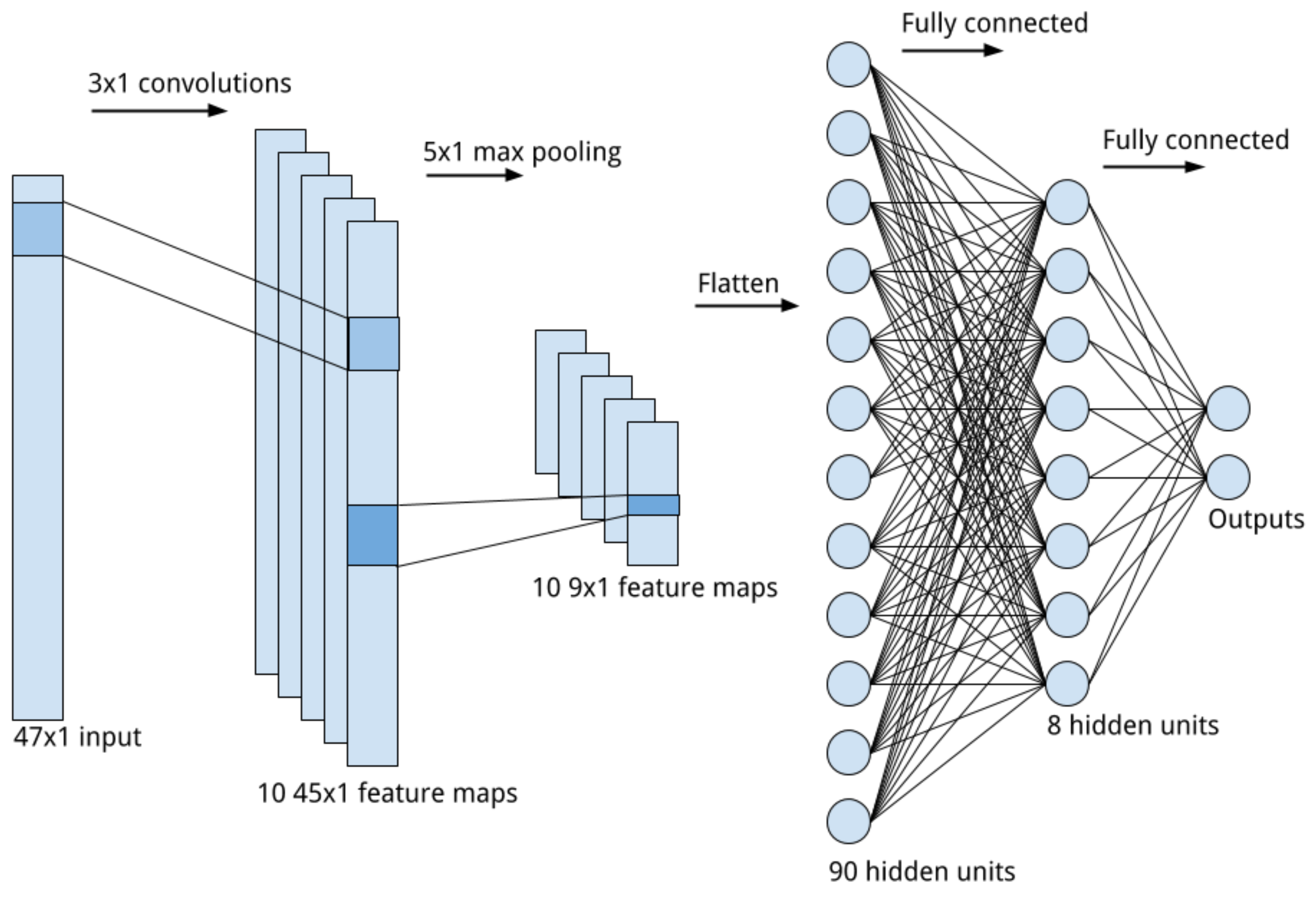
- When taking data, >1 person watches for anomalies in the detector 24/7
- At this stage no global processing of the event
- Instead, local information from detector components available (e.g., detector occupancy in a certain time window)

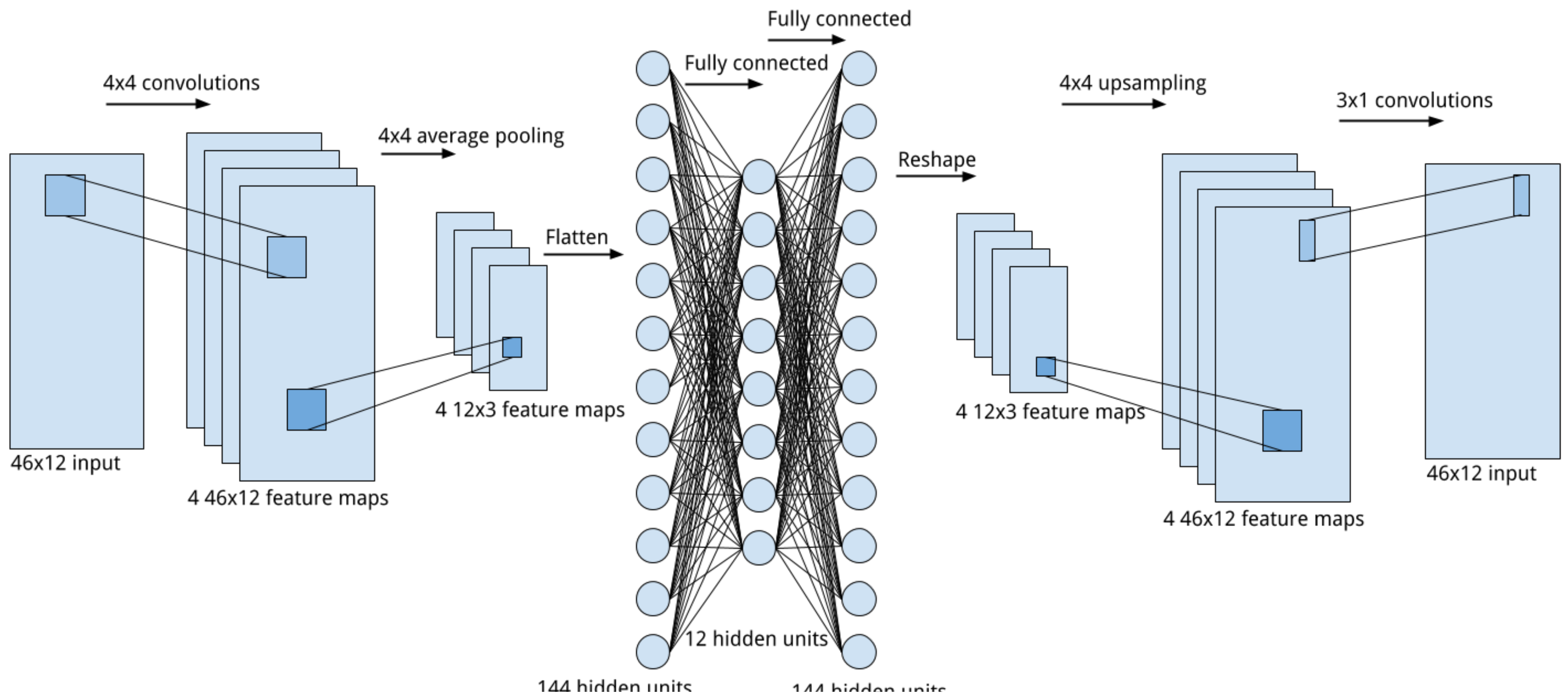




Two approaches

- Given the nature of these data, ConvNN are a natural analysis tool. Two approaches pursued
  - Classify good vs bad data. Works if failure mode is known
  - Use autoencoders to assess data “typicality”. Generalises to unknown failure modes



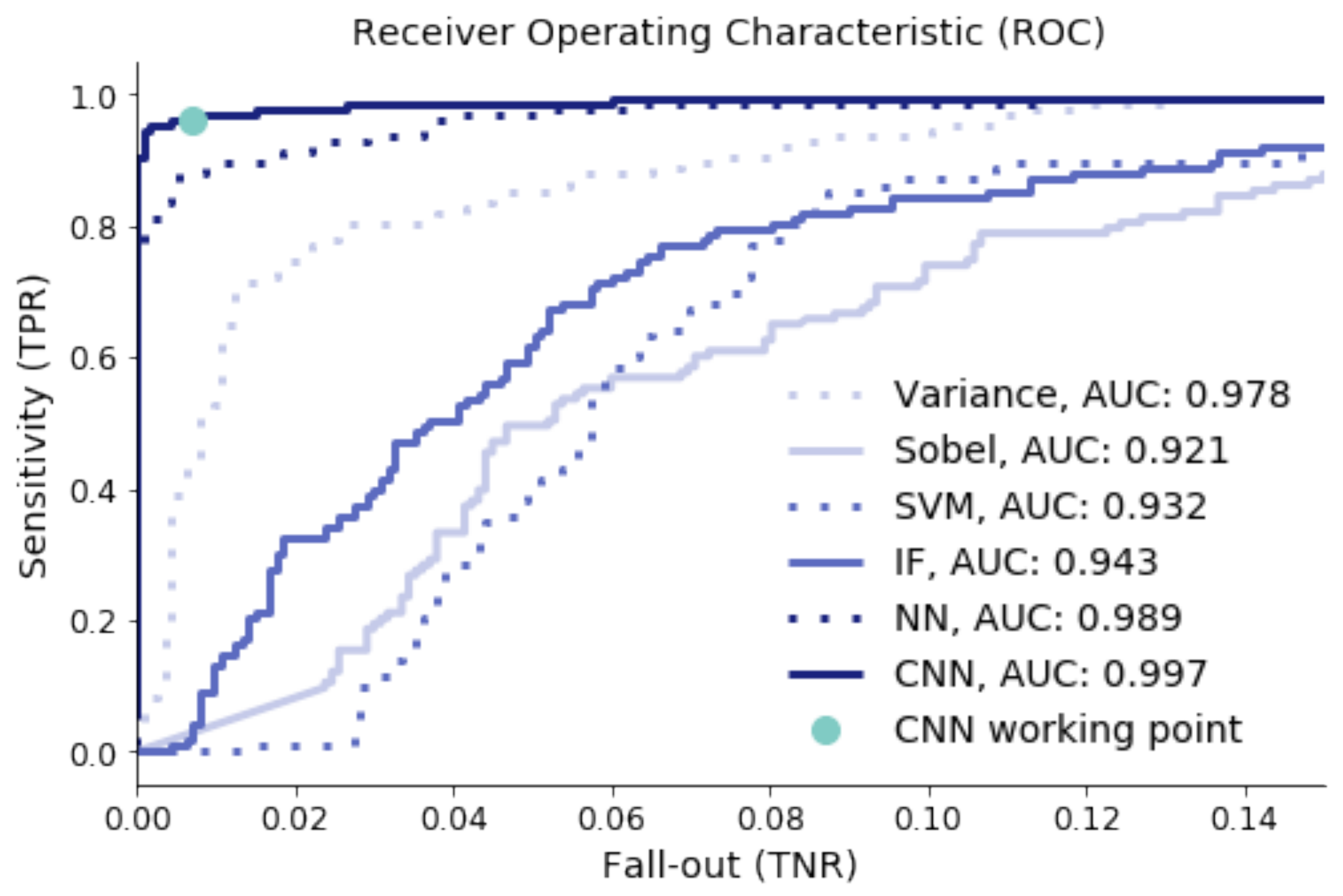


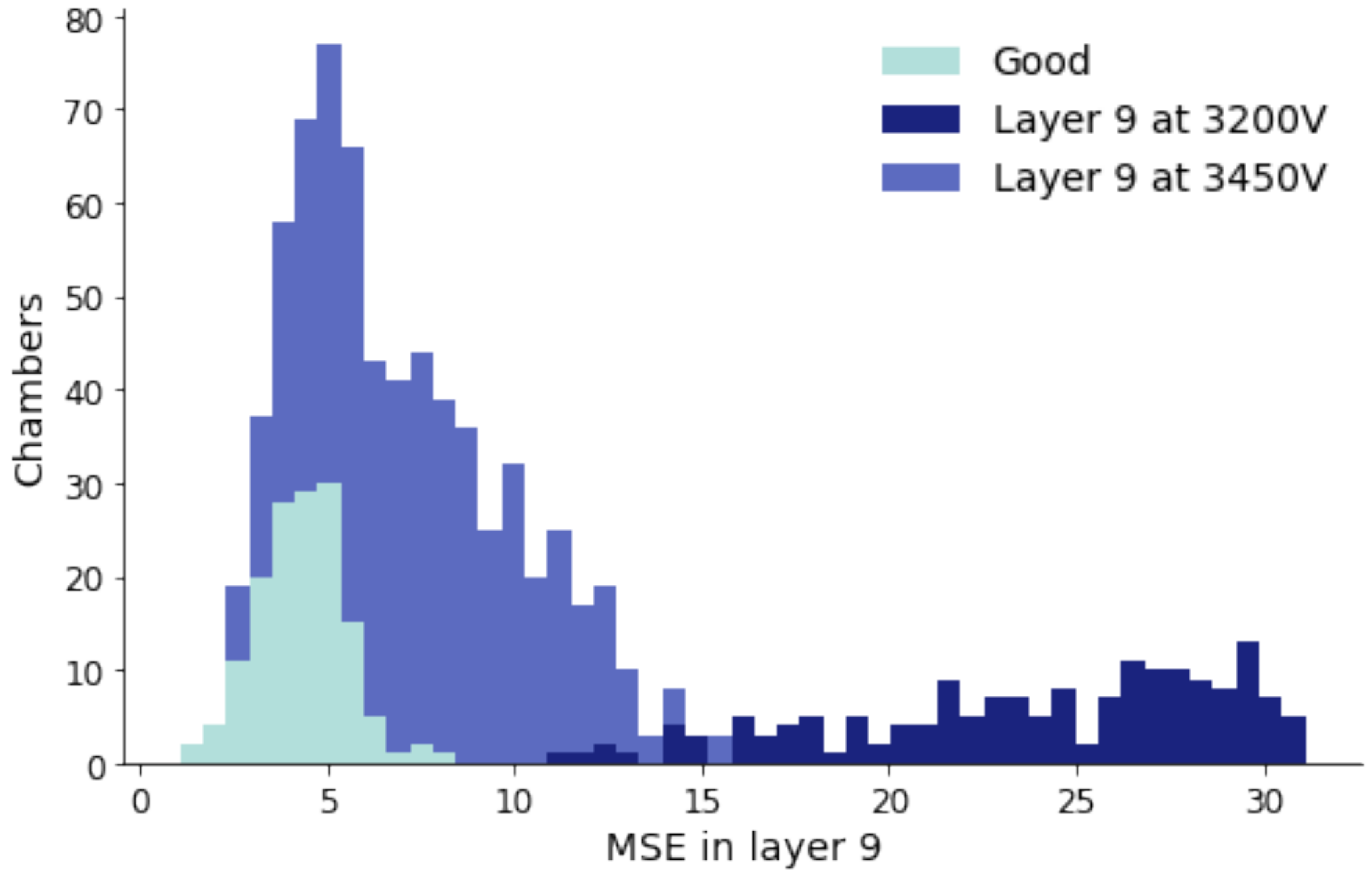
A. Pol et al., to appear soon

Two approaches

- Given the nature of these data, ConvNN are a natural analysis tool. Two approaches pursued
  - Classify good vs bad data. Works if failure mode is known
  - Use autoencoders to assess data "typicality". Generalises to unknown failure modes

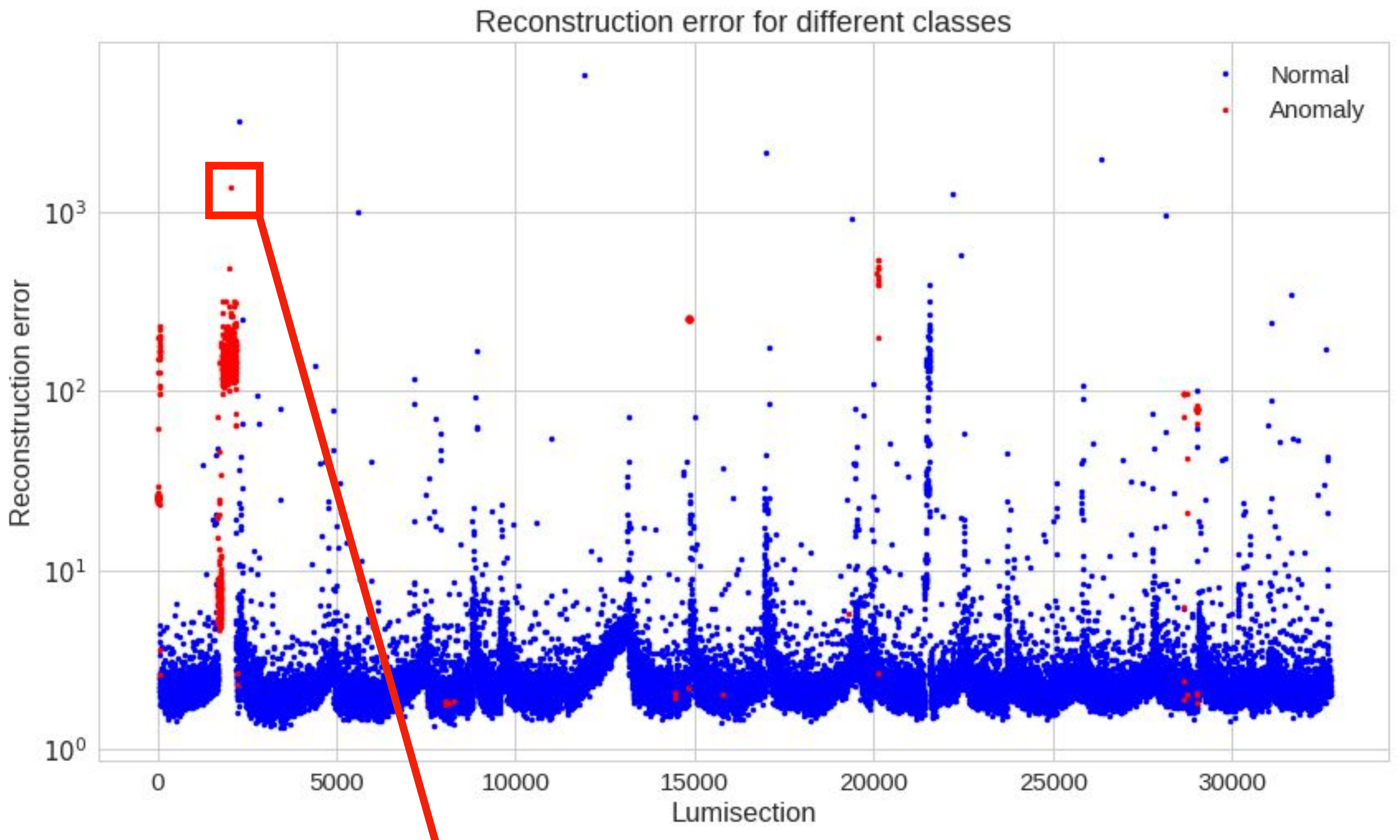
A. Pol et al., to appear soon

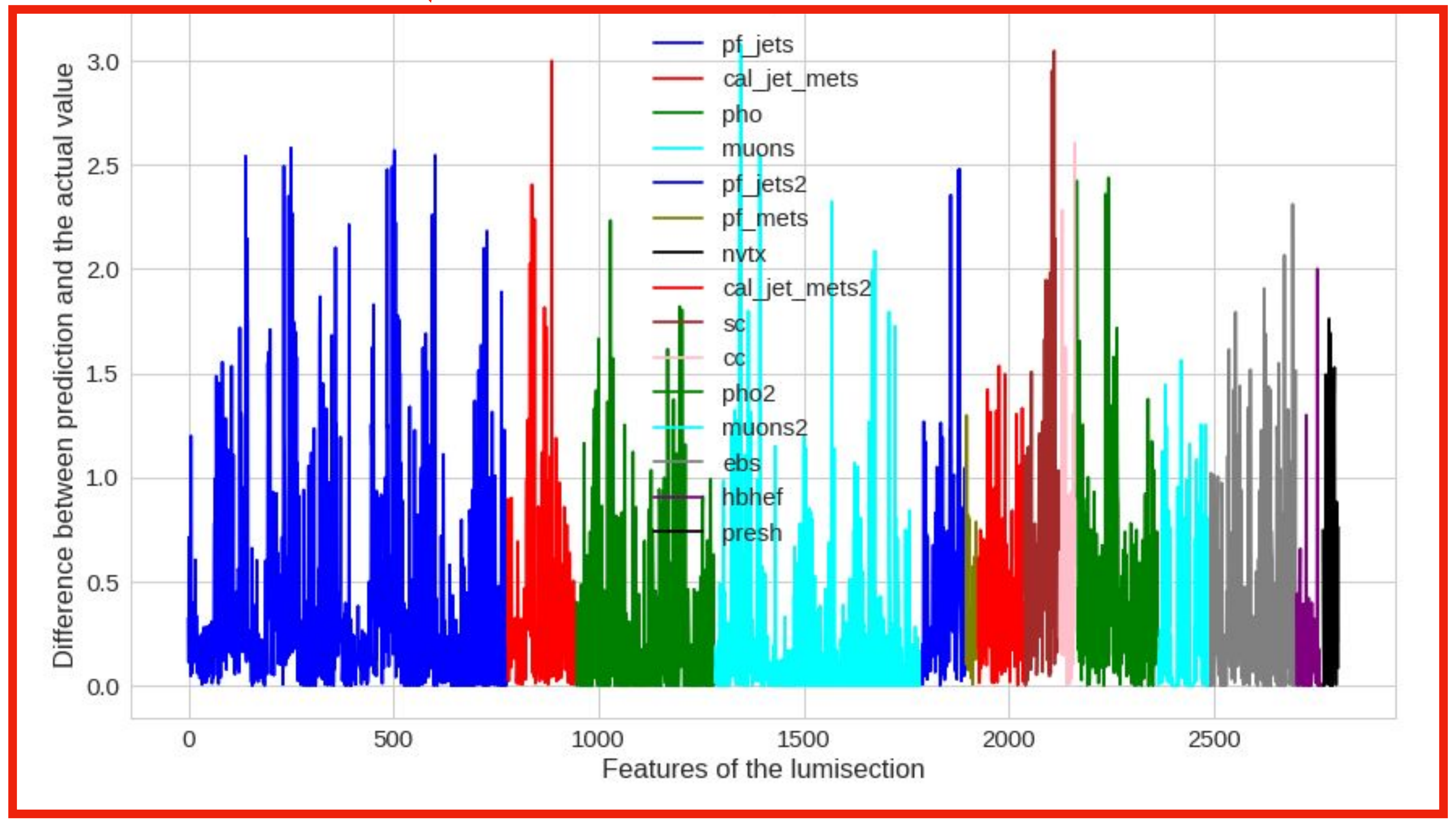




Data Quality Certification

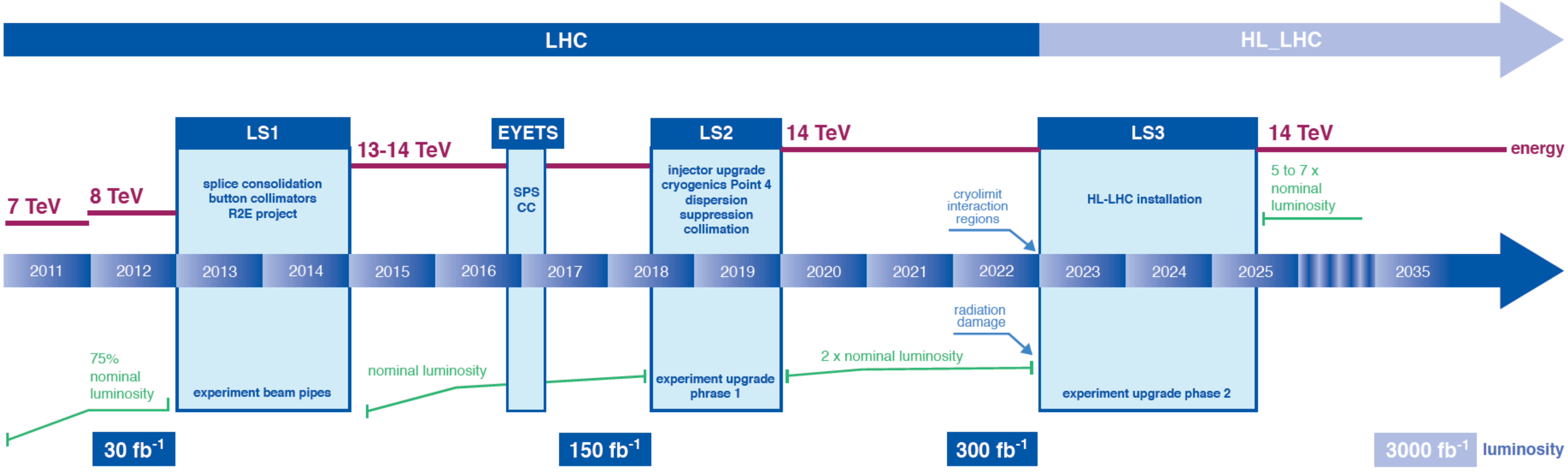
- *Autoencoder-based 1-class approach generalises to later stages of quality assessment*
  - *after reconstruction of the events, event reconstruction allows a global assessment (w.g., looking at electrons, muons, etc rather than hits in the detector)*
  - *A global autoencoder can spot all these features*
  - *Monitoring individual contributions to loss function (e.g., MSE) one can track the problem back to a specific physics object/detector component*





F. Široký et al., to appear sooner or later

A roadmap towards HL-LHC



- *We need to be ready by* ***2025*** *(High-Luminosity LHC)*
- *LHC Run 3 (2020-2022) is the ultimate demonstration opportunity*
  - *produce proof-of-principle studies on simulations and open datasets*
  - *bring ML expertise at CERN and in the experiments*
  - *within experiments, develop/test/deploy ML solutions to solve technical tasks*

Backup

Practical infos

- CERN Data Science Seminars
- LHC iML working group
- Data Science @HEP workshop series
  - CERN 2015
  - Simons Foundation (New York) 2016
  - Fermilab 2017