Machine Learning in HEP

mPP

hls4ml

Maurizio Pierini

CMS

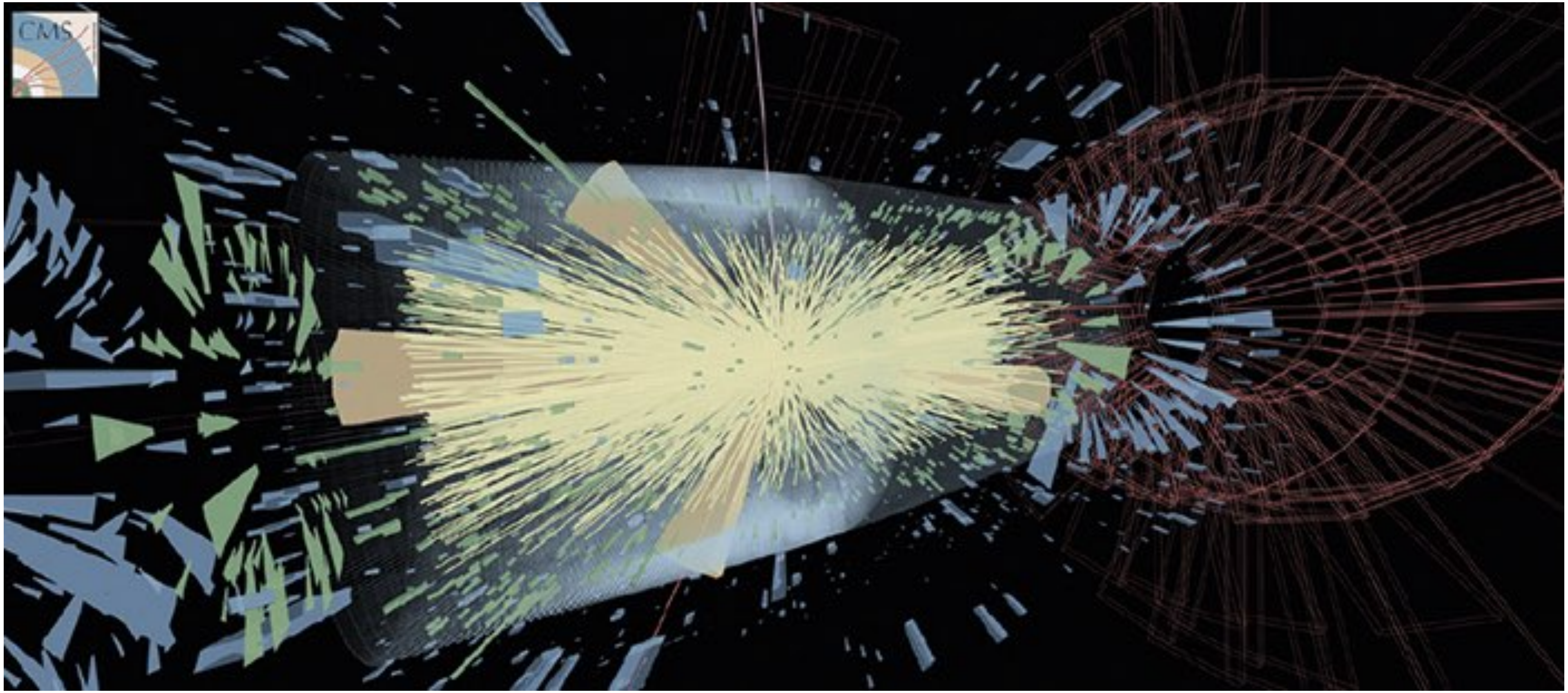
CERN

European Research Council

Outline

- *The LHC challenge*
- *A little bit of history of ML in HEP*
- *The CERN ML-LHC community*
- *Deep Learning at LHC by a few examples (biased selection)*
  - *Local reconstruction: clustering in calorimeters*
  - *Supervised Learning: jet tagging*
  - *Unsupervised searches: (re)discovering particles*
  - *Real time inference on FPGAs*
  - *Generative models: jet generation [Not Covered for lack of time]*

European Research Council

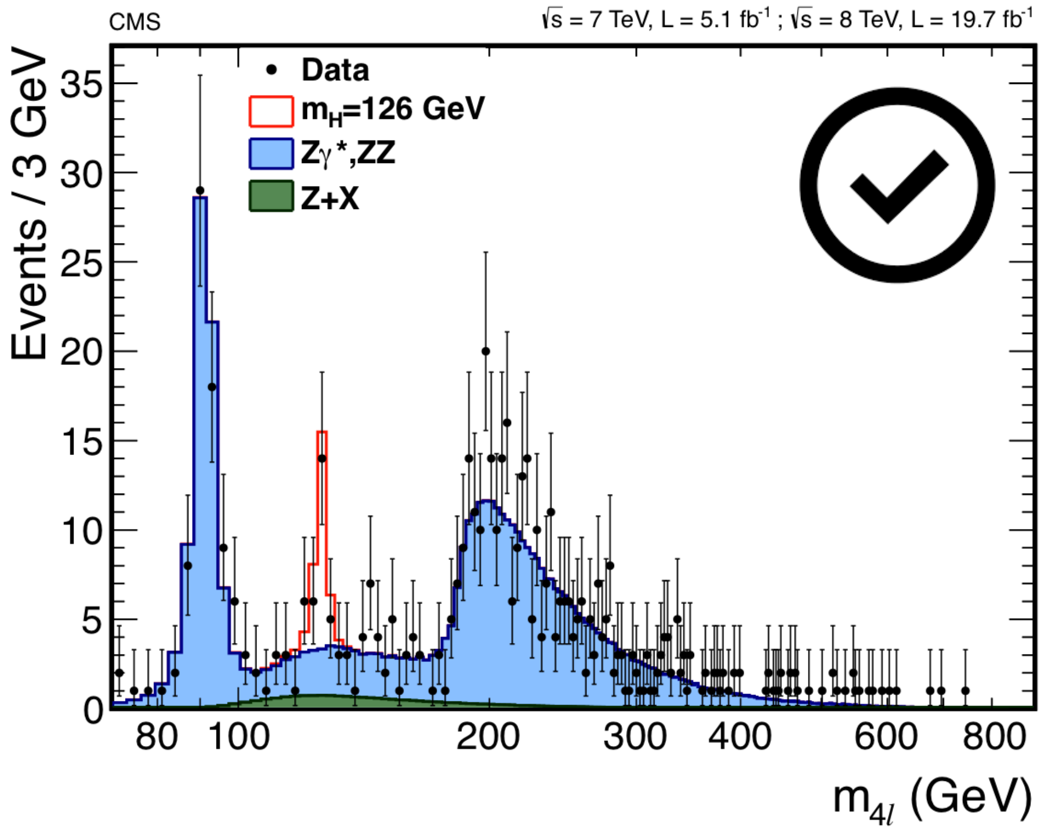


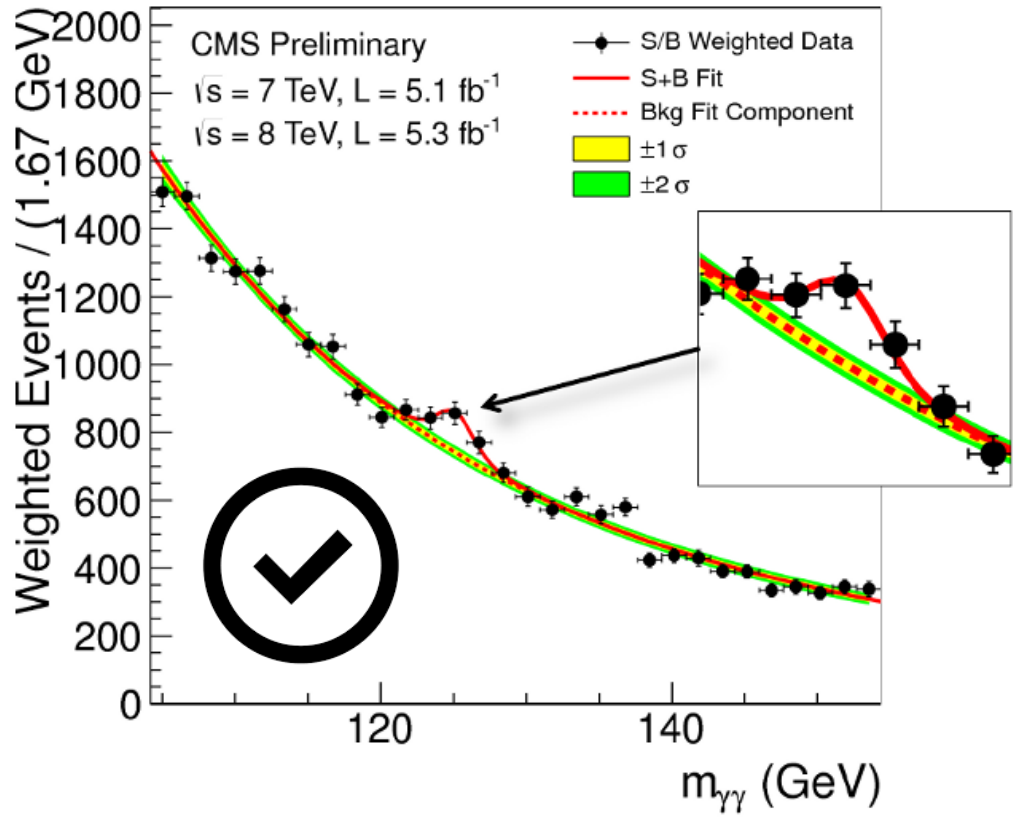
The LHC and its big-data challenge

European
Research
Council

LHC: Energy frontier exploration

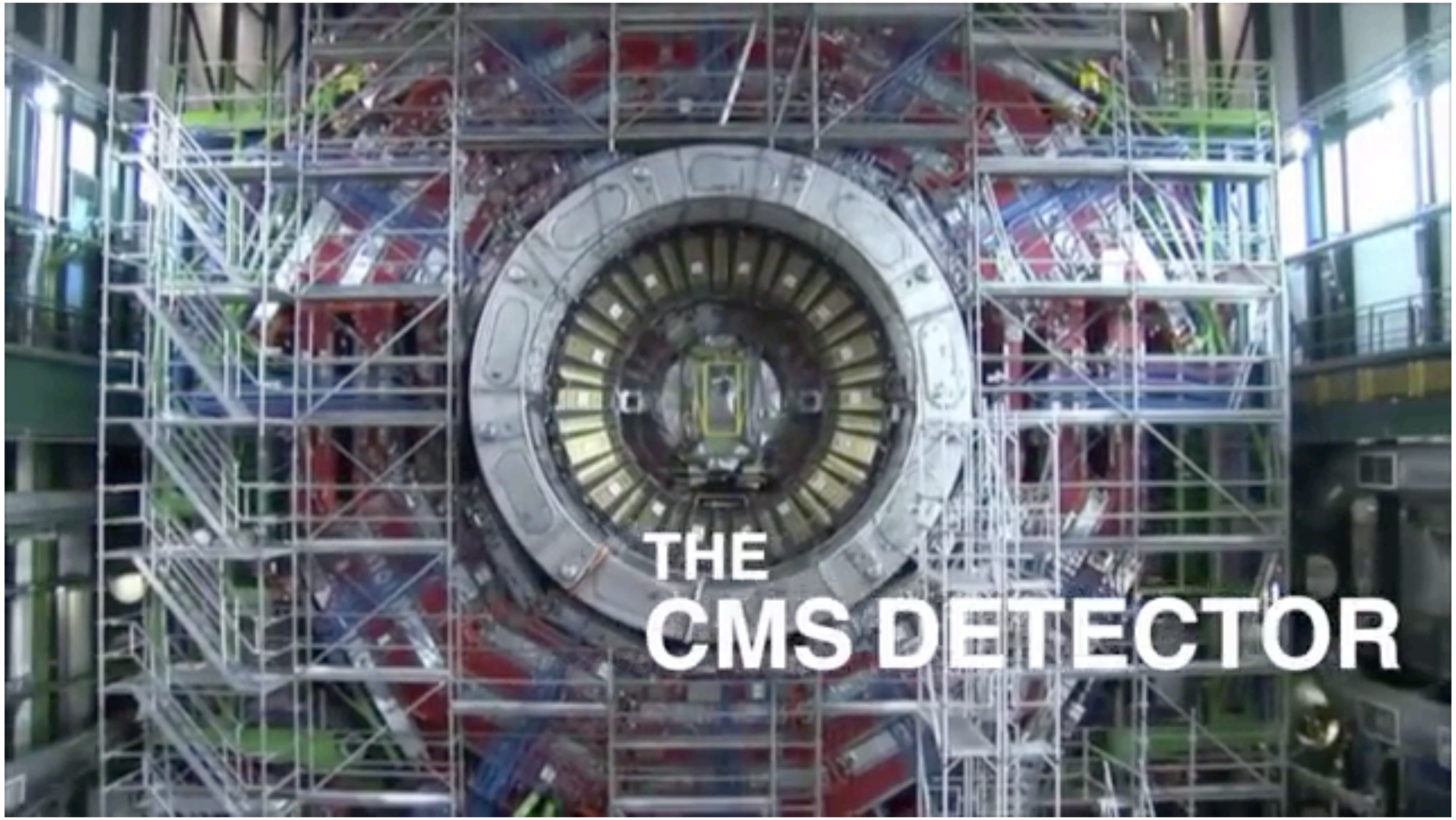
- *Discover the Higgs boson or exclude its existence*
- *Characterize the nature of EW symmetry breaking*
- *Help answering the big questions left in particle physics*
  - *What stabilises physics at EW scale?*
  - *What's the nature of Dark Matter?*
  - *Origin of cosmological matter/antimatter asymmetry*
  - *Are there unexpected phenomena at the energy frontier*



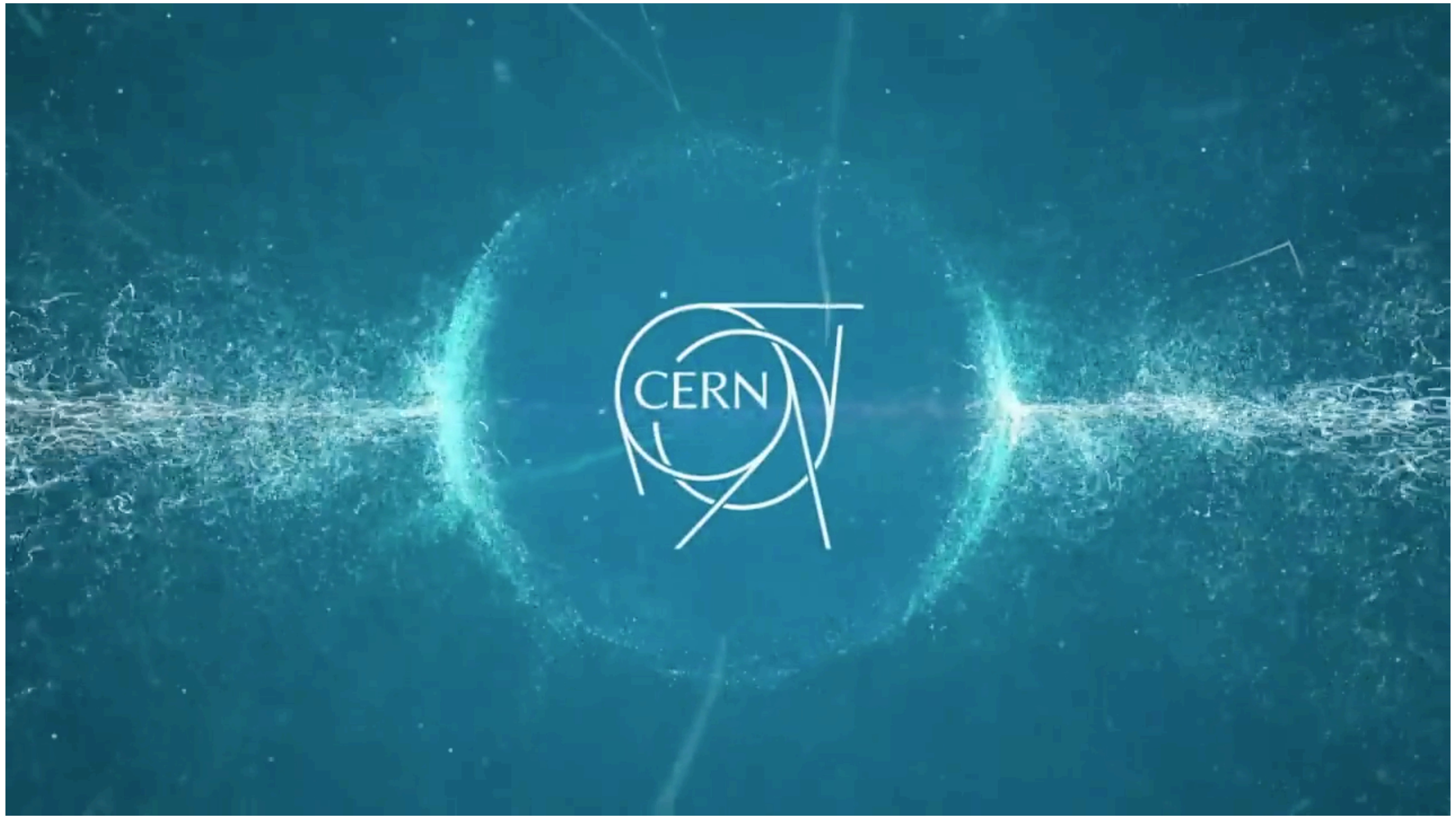


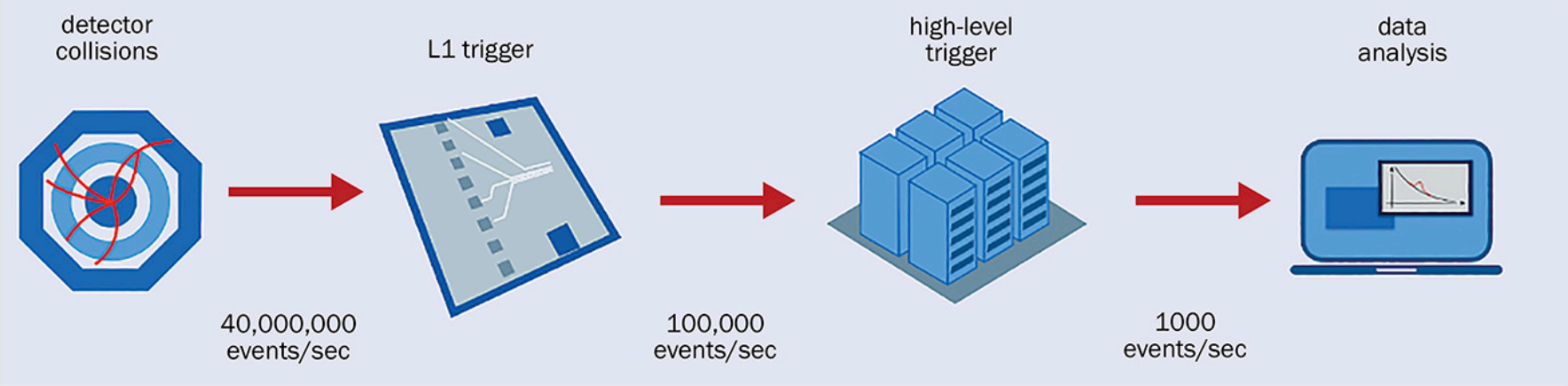
The LHC collisions

- *The LHC collides protons at unprecedented energy (equivalent to 13,000 times their mass)*
  - *one collision every 25 ns (= 40 Million collisions/sec)*
  - *Thousands of particles emerging from each collision*
  - *1 MB of data recorded at each collision by big detectors*



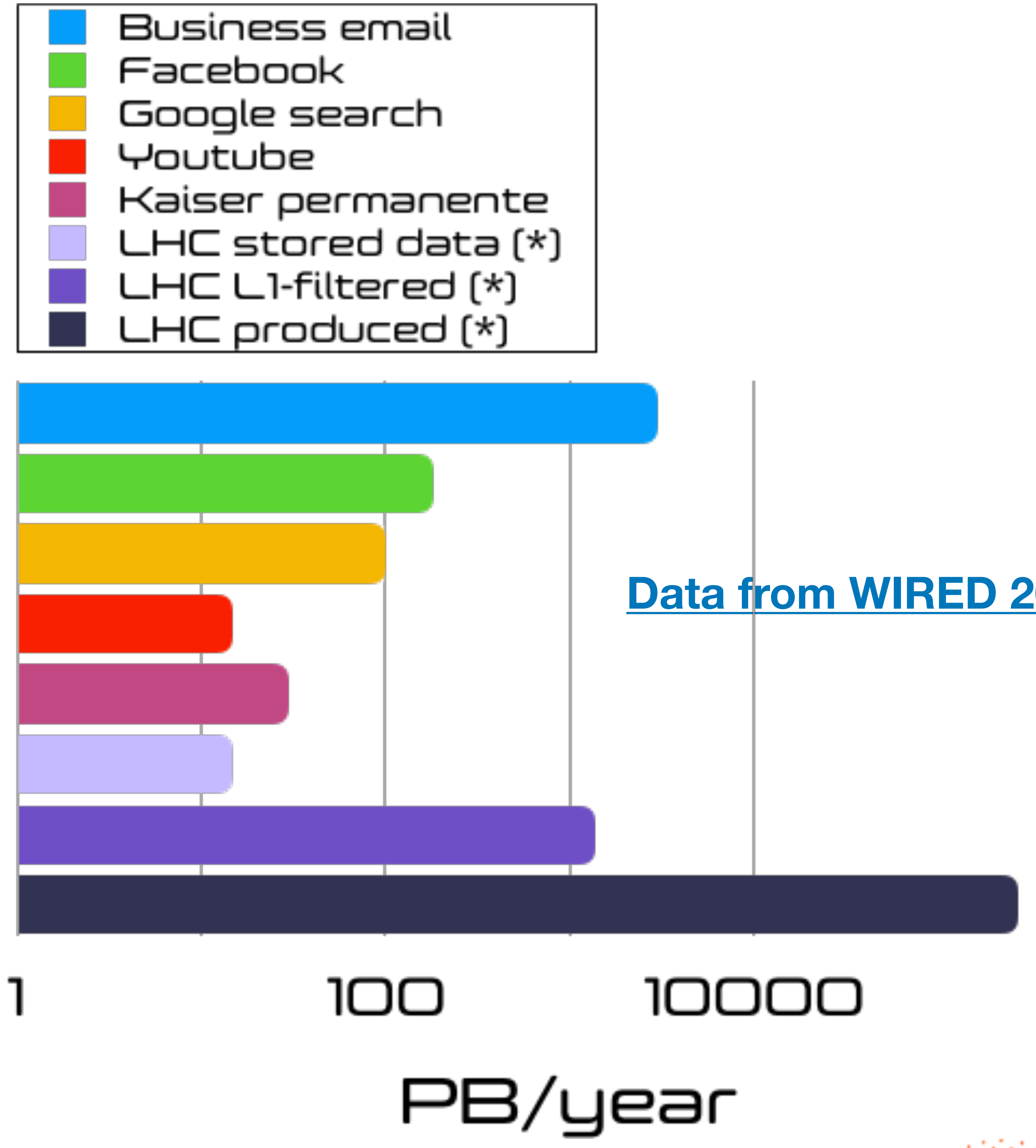
Big Data @LHC


CERN


detector collisions
L1 trigger
high-level trigger
data analysis
40,000,000 events/sec
100,000 events/sec
1000 events/sec

Big Data @LHC

- The amount of produced data is too much to be stored
  - 1,000 times the data generated by google searches+youtube+facebook back in 2013
  - Reduced to 5x(google searches+youtube+facebook) after first filtering
  - Can only store 5% of those

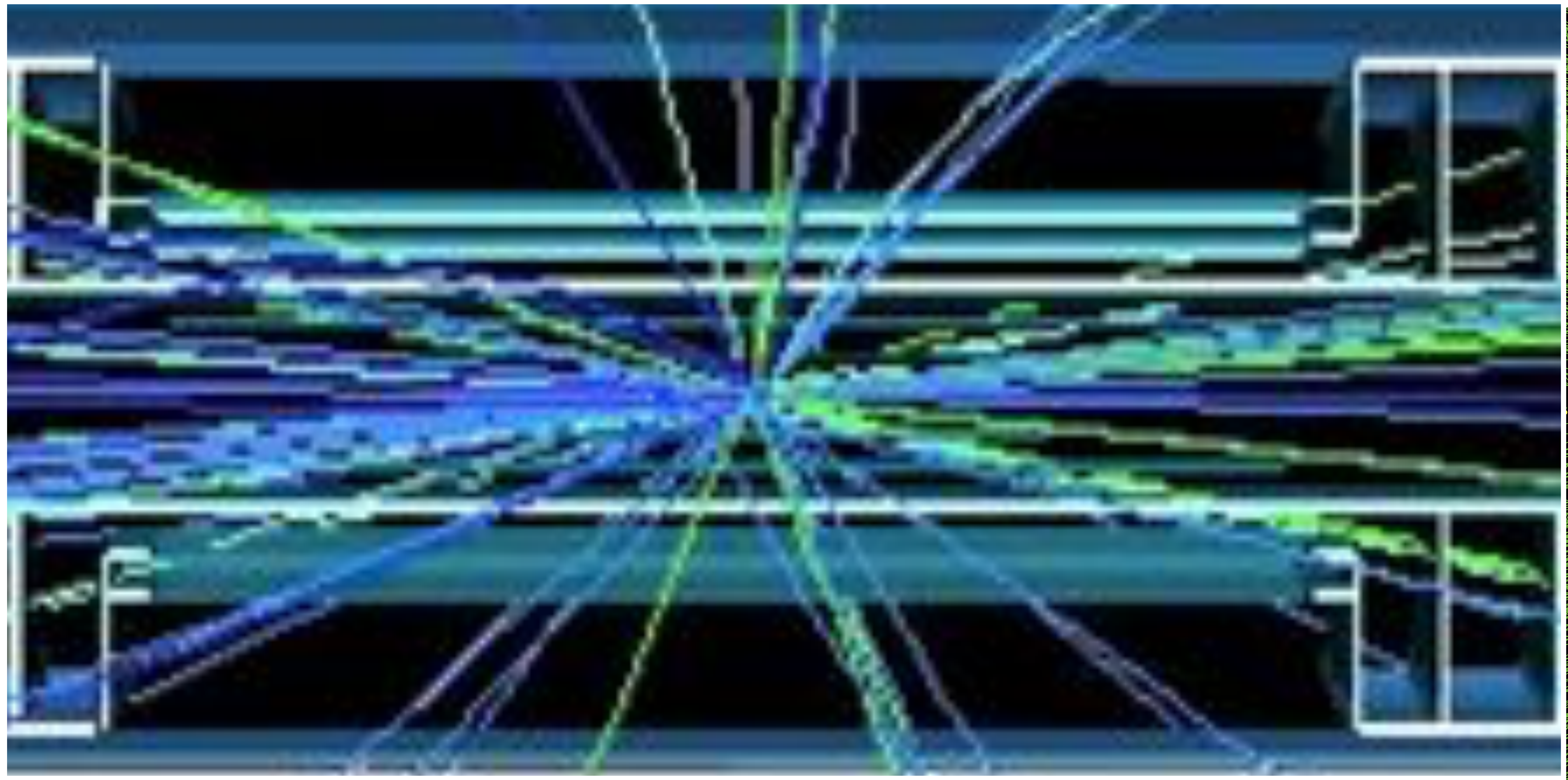


Data from WIRED 2013

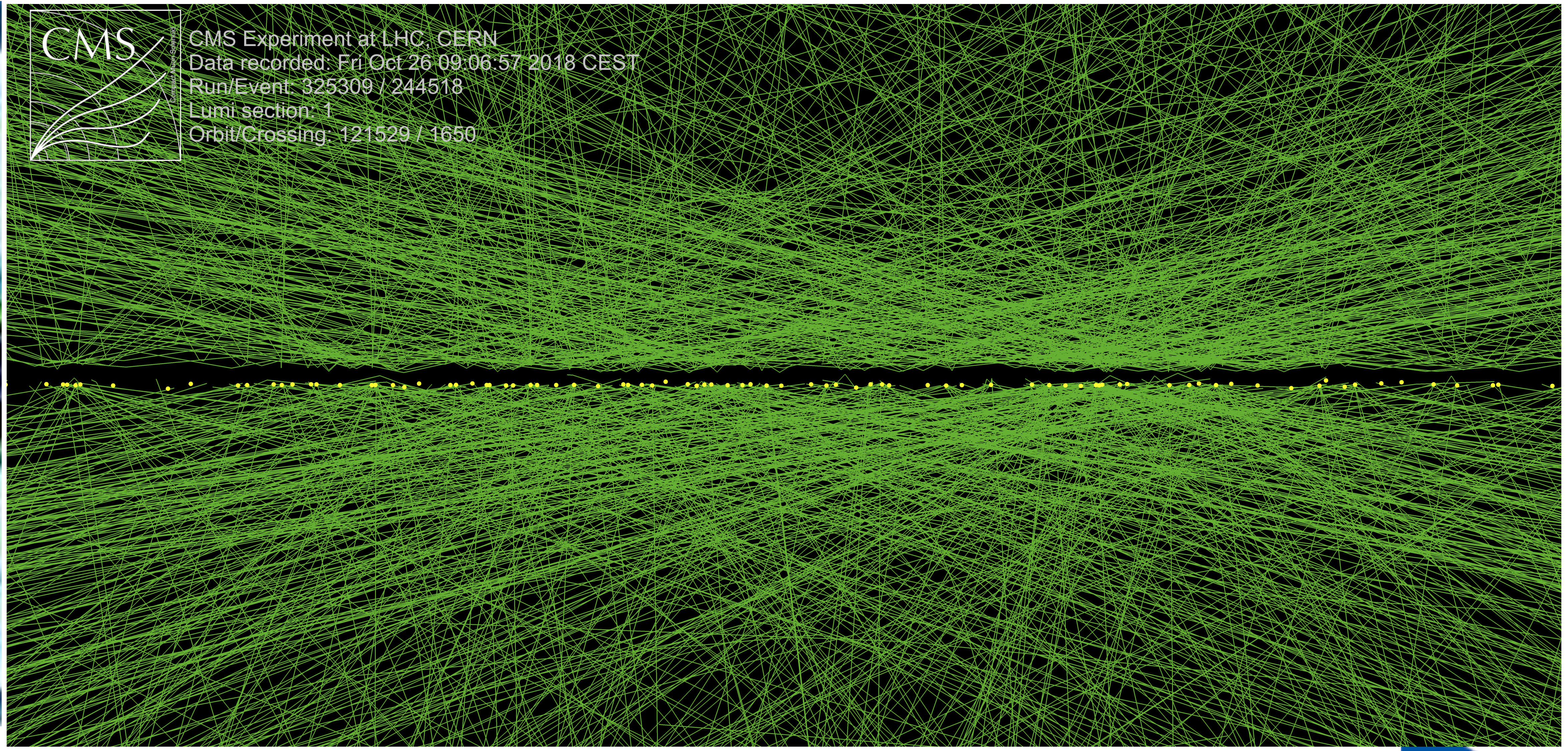
(*) Only two big experiments (ATLAS and CMS), only RAW data

Things will get worse

5 interactions/beam cross



140 interactions/beam cross



6 | 2017 | 2018 | 2019 | 2020 | 2021 | 2022 | 2023 | 2024 | 2025 | 2035

This is when the R&D has to happen

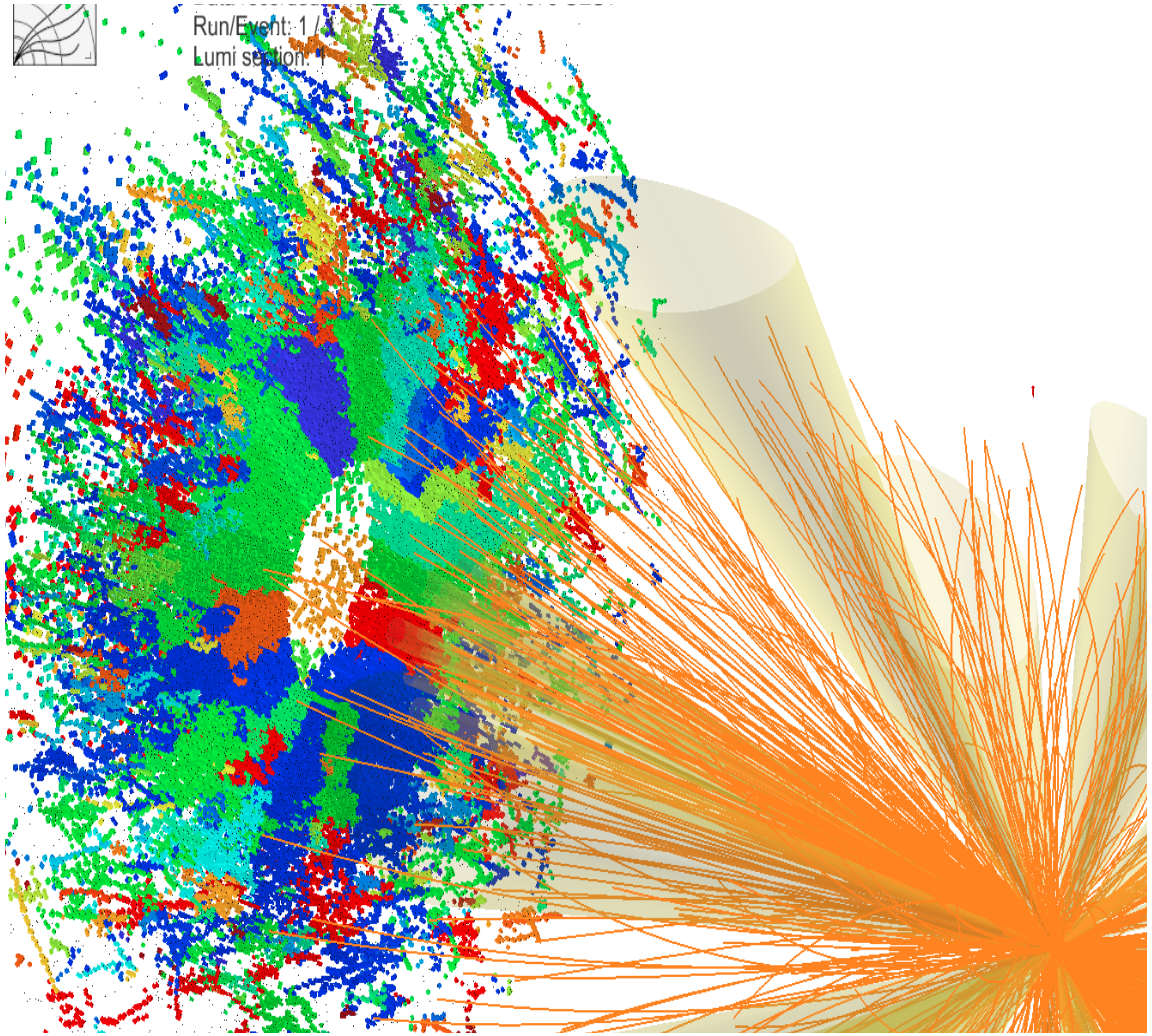
LHC Today | HL_LHC

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

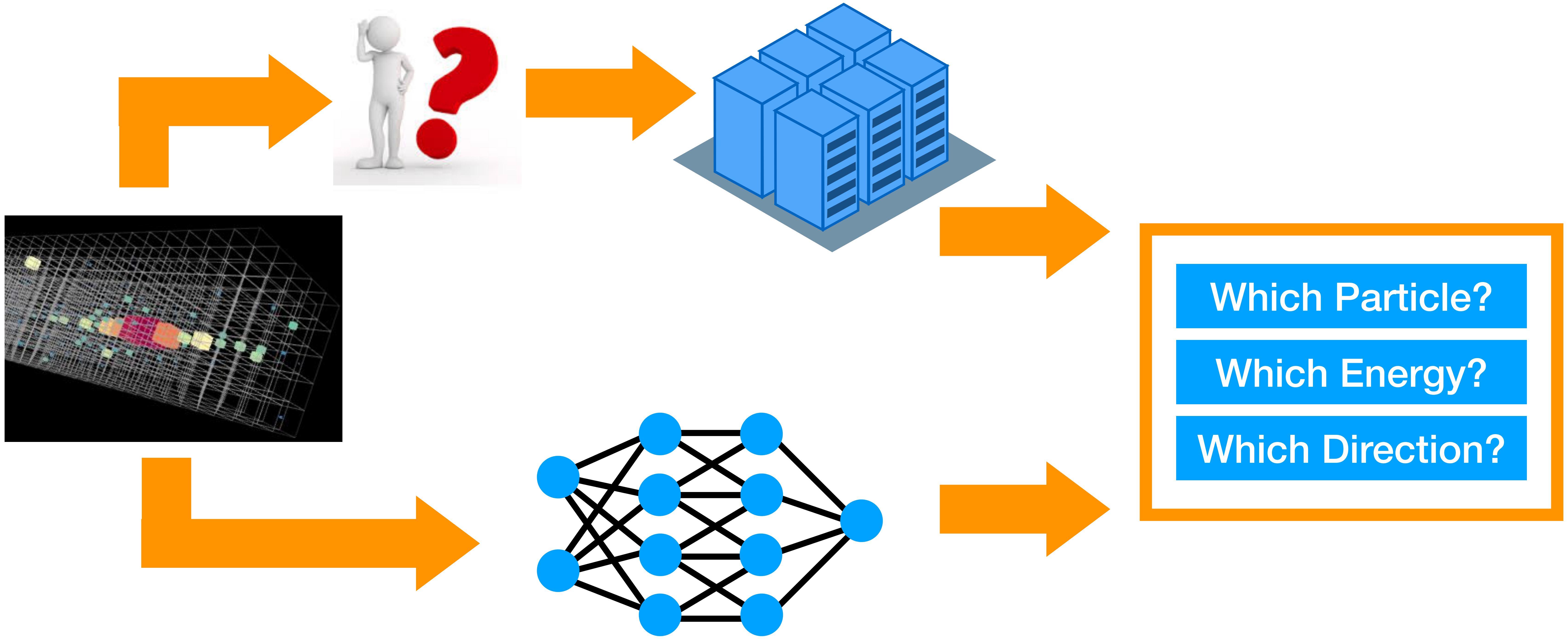
- ~200 collisions/event
- ~minute/event processing time
- (at best)Same computing resources as today

More sensors, more RECO troubles

- To disentangle 200 collisions happening at once, we will build new detectors with more (smaller) sensors
- Event complexity grows non linearly
- To profit of that, computing resources for data processing will have to increase
- **We are off by a factor ~10 if we project to 2027**



Deep Learning at Rescue: Reco



- We know how to get from the data the answers we want
  - physics + intuition + computing
- But the process is slow
  - We can use DL solutions as a shortcut: we teach neural networks how to give us the answer we want directly from the raw data

It started with NNs & Pattern Recognition

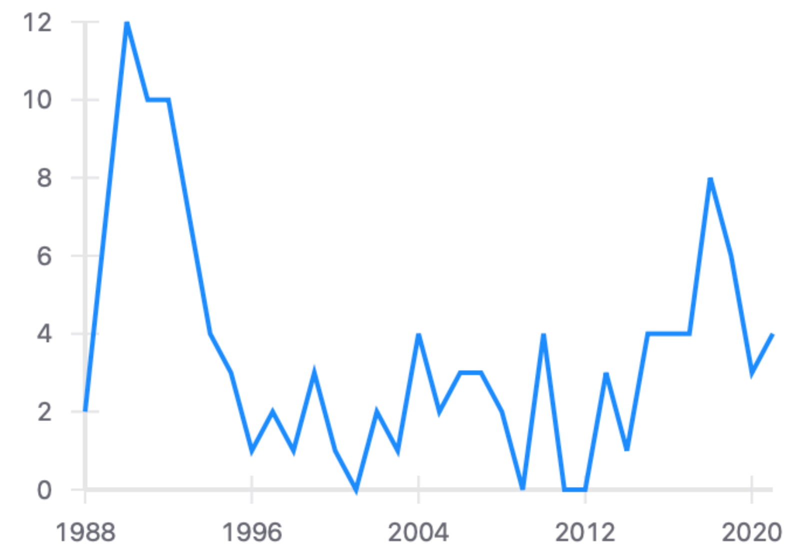
- *First papers proposing NNs applications in HEP date back to end of 80s*
  - *Pattern recognition (particle tracking)*
  - *Object identification (classification)*

Neural Networks and Cellular Automata in Experimental High-energy Physics

Bruce H. Denby (Orsay, LAL)
Nov, 1987

35 pages
Published in: *Comput.Phys.Commun.* 49 (1988) 429-448
DOI: 10.1016/0010-4655(88)90004-5
Report number: LAL-87-56
View in: ADS Abstract Service

Citations per year



Abstract: (Elsevier)
Within the past few years, two novel computing techniques, cellular automata and neural networks, have shown considerable promise in the solution of problems of a very high degree of complexity, such as turbulent fluid flow, image processing, and pattern recognition. Many of the problems faced in experimental high energy physics are also of this nature. Track reconstruction in wire chambers and cluster finding in cellular calorimeters, for instance, involve pattern recognition and high combinatorial complexity since many combinations of hits or cells must be considered in order to arrive at the final tracks or clusters. Here we examine in what way connective network methods can be applied to some of the problems of experimental high energy physics. It is found that such problems as track and cluster finding adapt naturally to these approaches. When large scale hard-wired connective networks become available, it will be possible to realize solutions to such problems in a fraction of the time required by traditional methods. For certain types of problems, faster solutions are already possible using model networks implemented on vector or other massively parallel machines. It should also be possible, using existing technology, to build simplified networks that will allow detailed reconstructed event information to be used in fast trigger decisions.

Track Finding With Neural Networks

Carsten Peterson (Lund U.)
Apr, 1988

16 pages
Published in: *Nucl.Instrum.Meth.A* 279 (1989) 537
DOI: 10.1016/0168-9002(89)91300-4
Report number: LU-TP-88-8

Citations per year

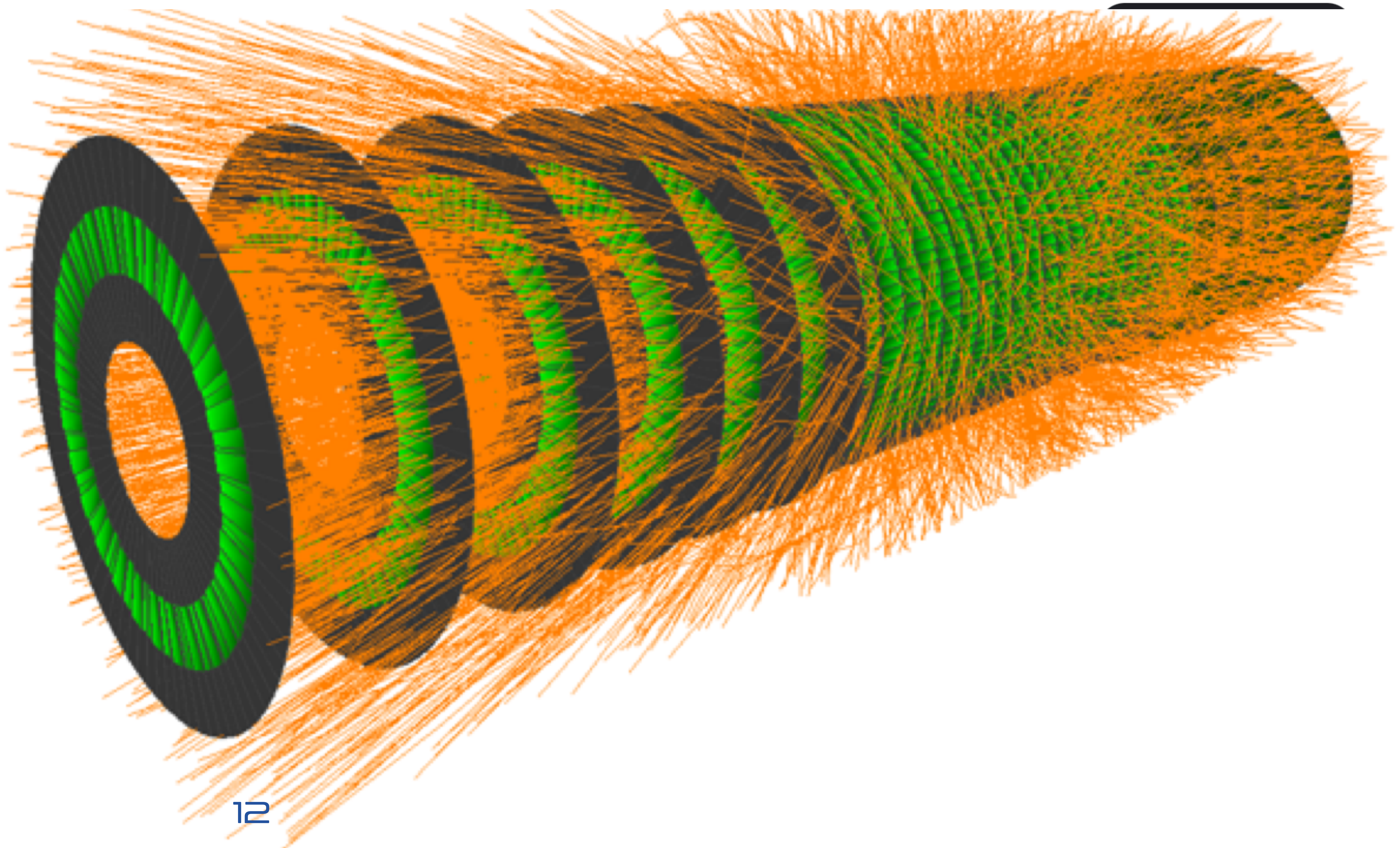
Abstract: (Elsevier)
A neural network algorithm for finding tracks in high energy physics experiments is presented. The performance of the algorithm is explored on modest size samples with encouraging results. It is inherently parallel and thus suitable for execution on a conventional SIMD architecture. More important, it naturally lends itself to direct implementations in custom made hardware, which would permit real time operations and hence facilitate fast triggers. Both VLSI and optical technology implementations are briefly discussed.

mPP

European Research Council

And it's still about NNs & Pattern Recognition

- Most of these applications are still the core of ML applications in HEP nowadays
  - but Deep Learning is broadening the use case list

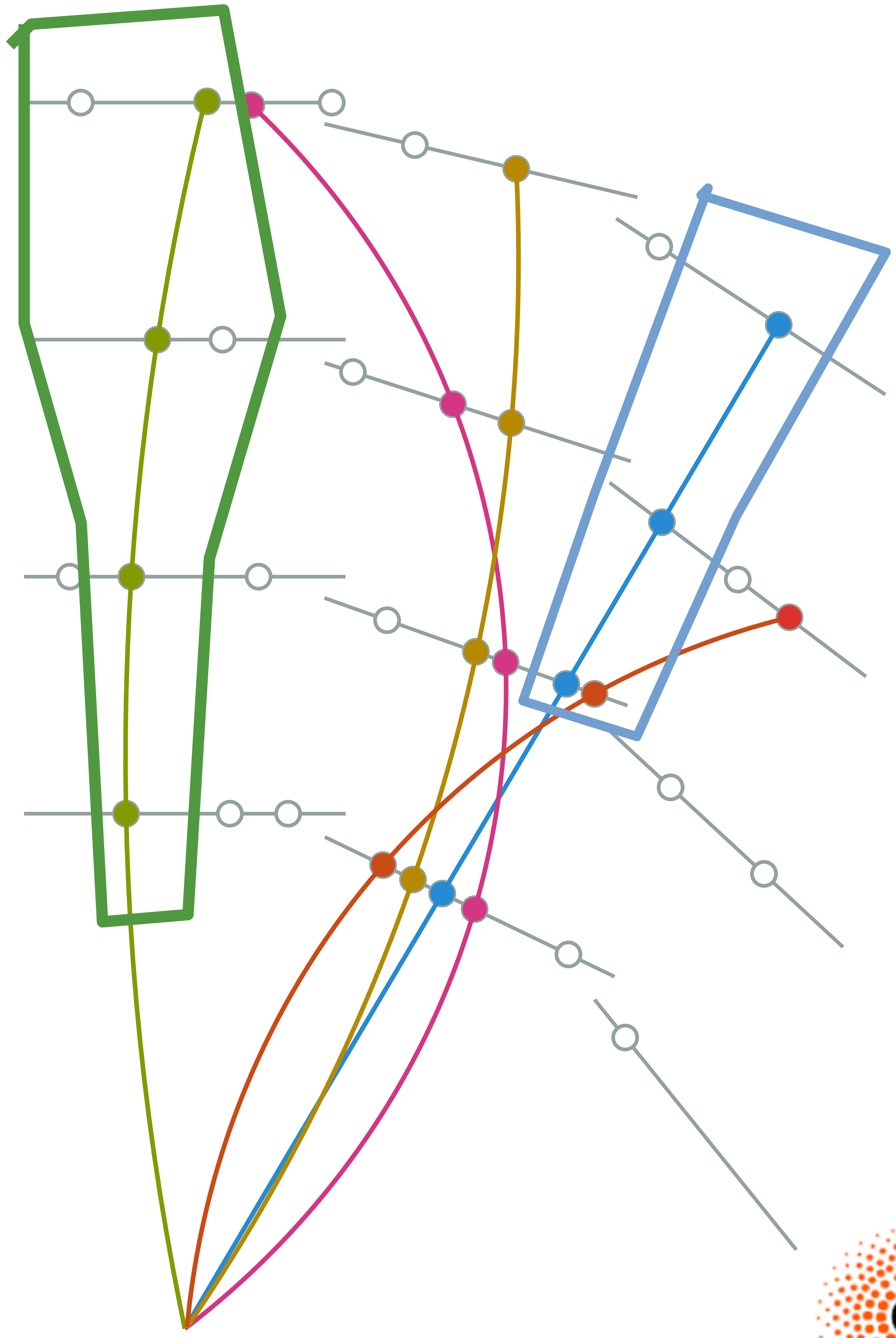


Dealing with HEP data

- ***Sparse data:*** *HEP data are sets of **detector hits**. Popular DL architectures (CNNs, RNNs) might work but with a cost (wasted memory) and could be improvable*

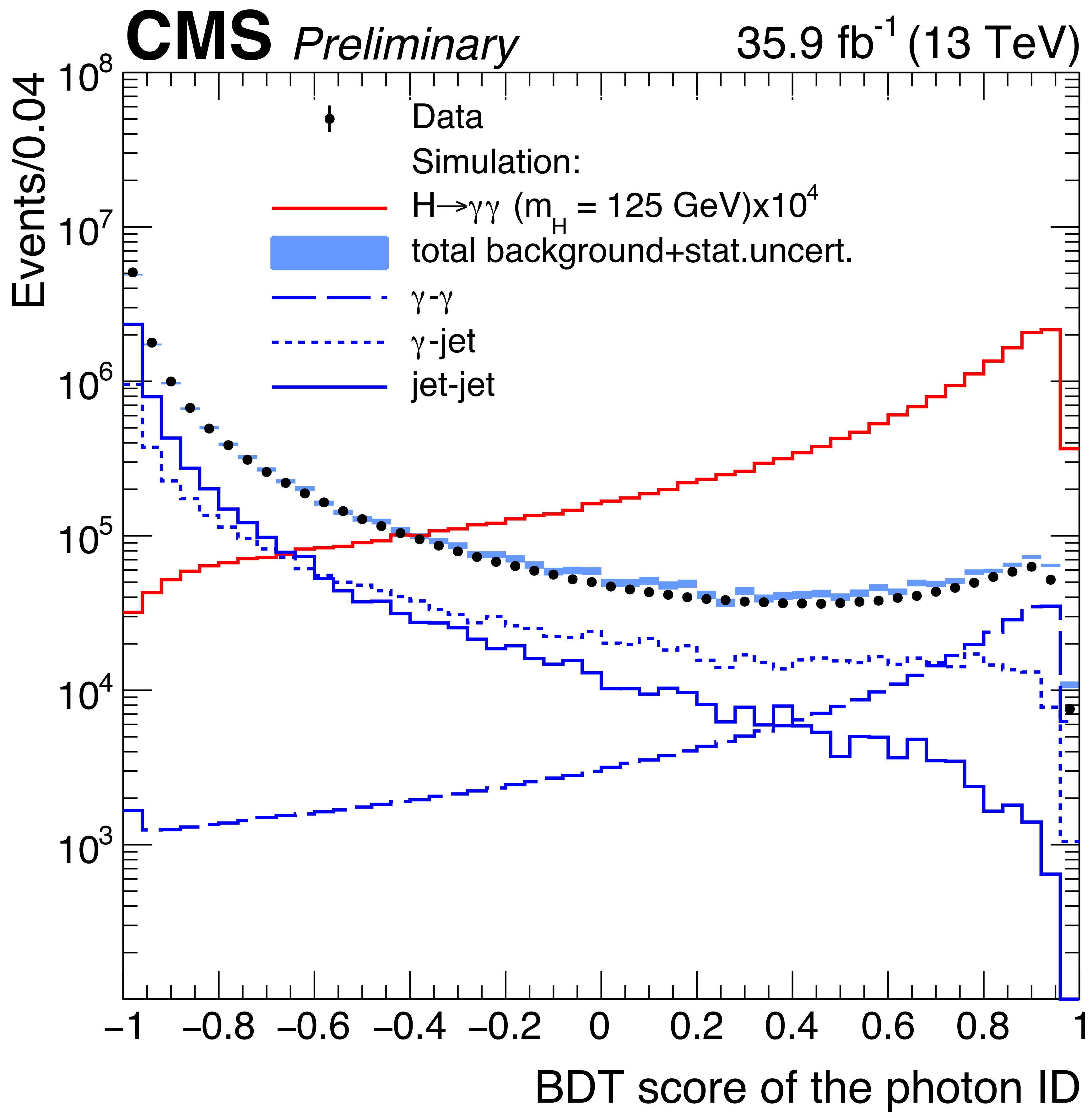
- ***Custom edge computing:*** *inference will have to run on our resources, going from front-end chips to custom electronic boards, dedicated computer centres, to the GRID (i.e., full support of site-dependent heterogenous computing)*

- ***Real-time:*** *(with real data) inference has to happen within the time boundaries of the trigger (as fast as <1 μsec)*



ML in HEP before DL

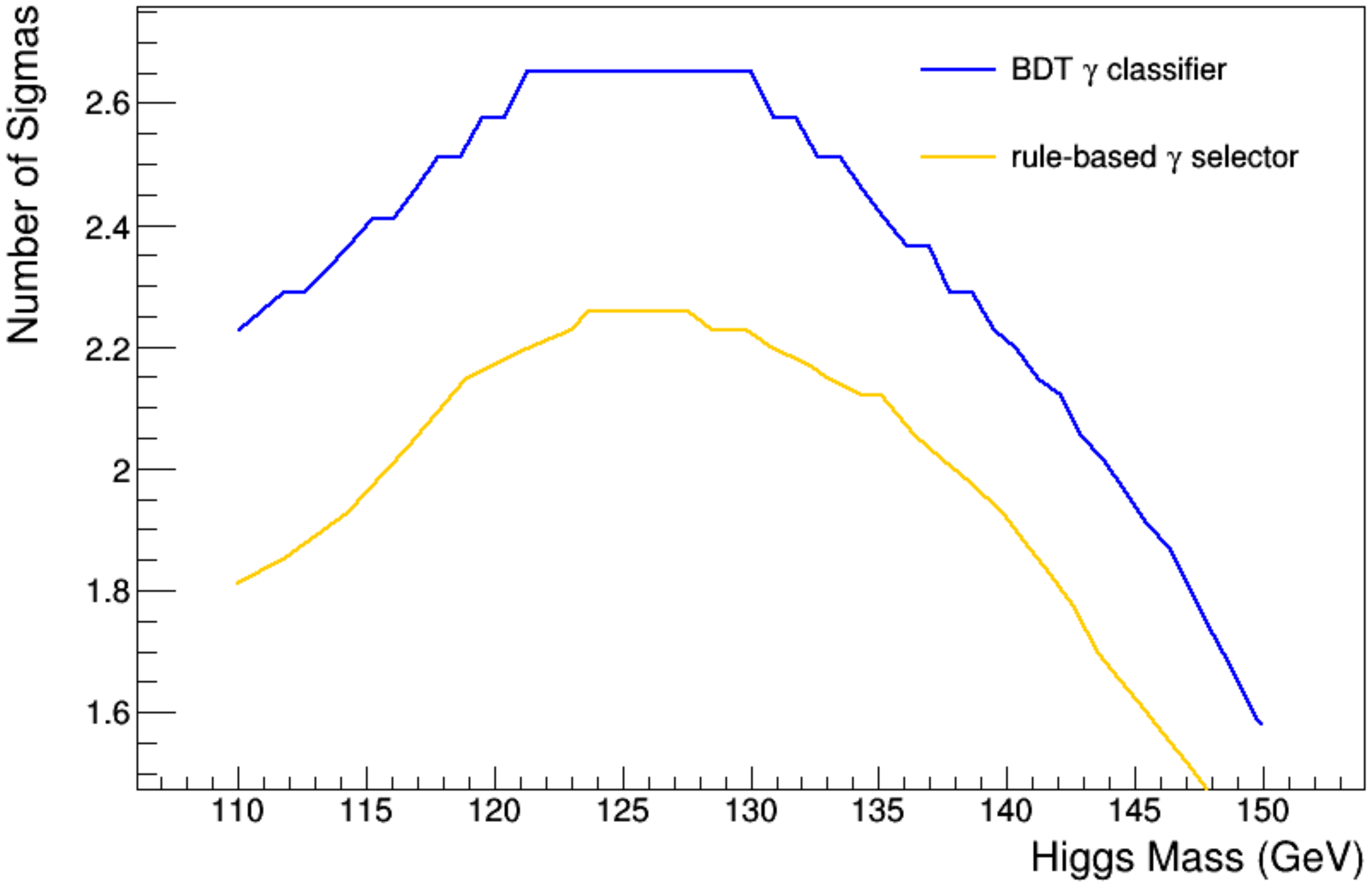
- Classification:
  - identify a particle & reject fakes
  - identify signal events & reject background
- Regression:
  - Measure energy of a particle
- We typically use BDTs for these task
  - moved to Deep Learning for analysis-specific tasks
  - same will happen for centralised tasks (eventually)

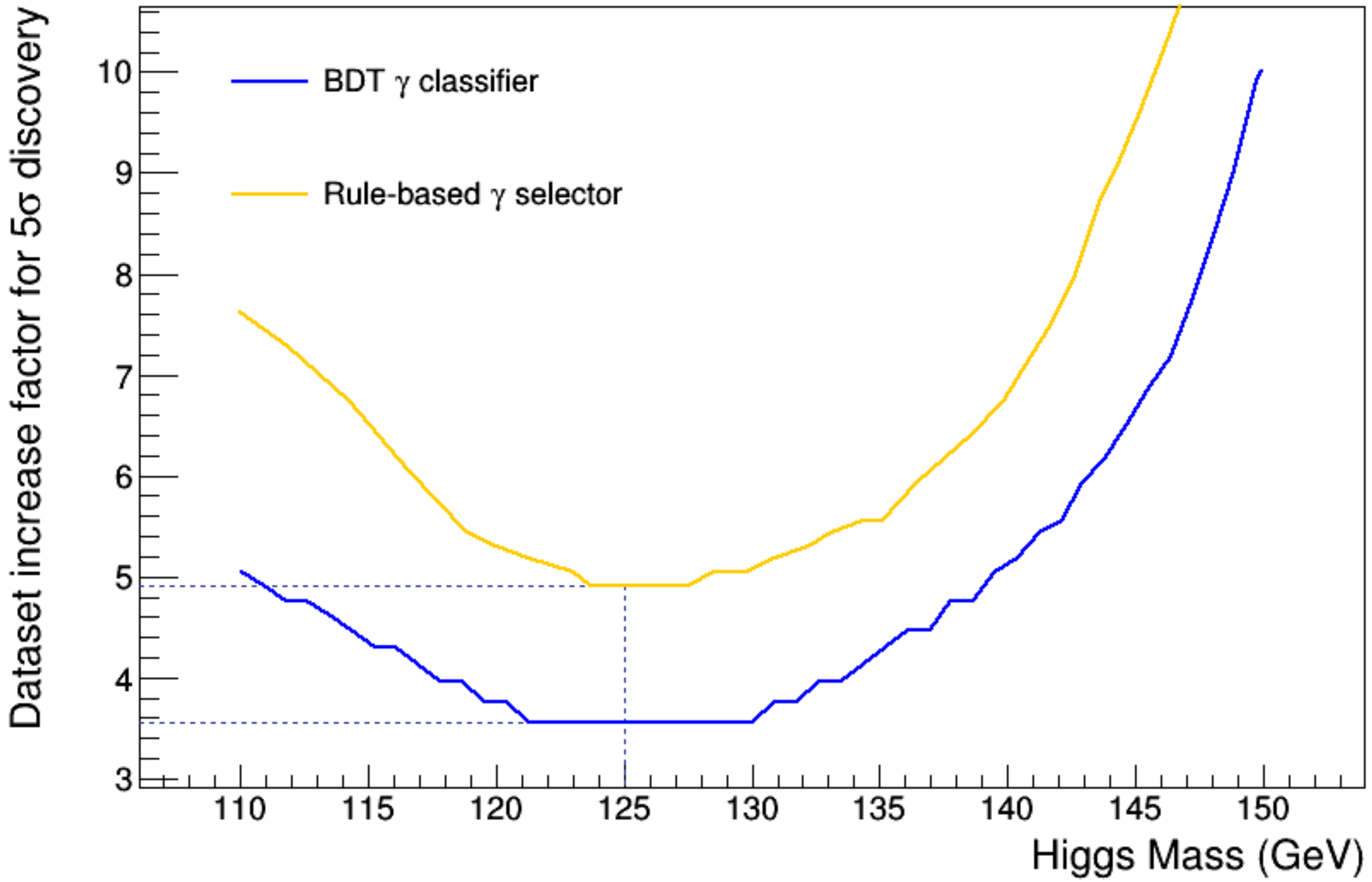


Centralised task (in online or offline reconstruction)
Analysis-specific task (by users on local computing infrastructures)

Example: ML for Higgs discovery

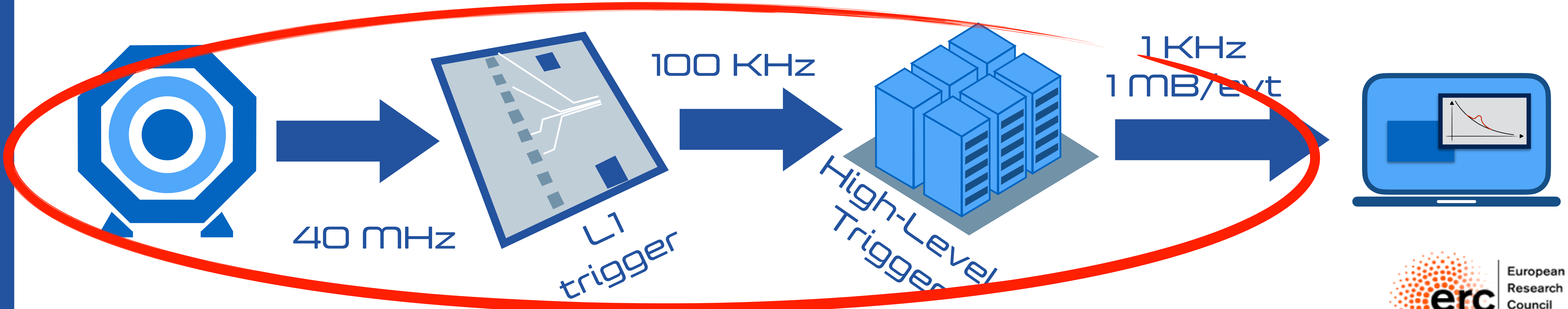
- *We were not supposed to discover the Higgs boson as early as 2012*
- *Given how the machine progressed, we expected discovery by end 2015 /mid 2016*
- *We made it earlier thanks (also) to Machine Learning*

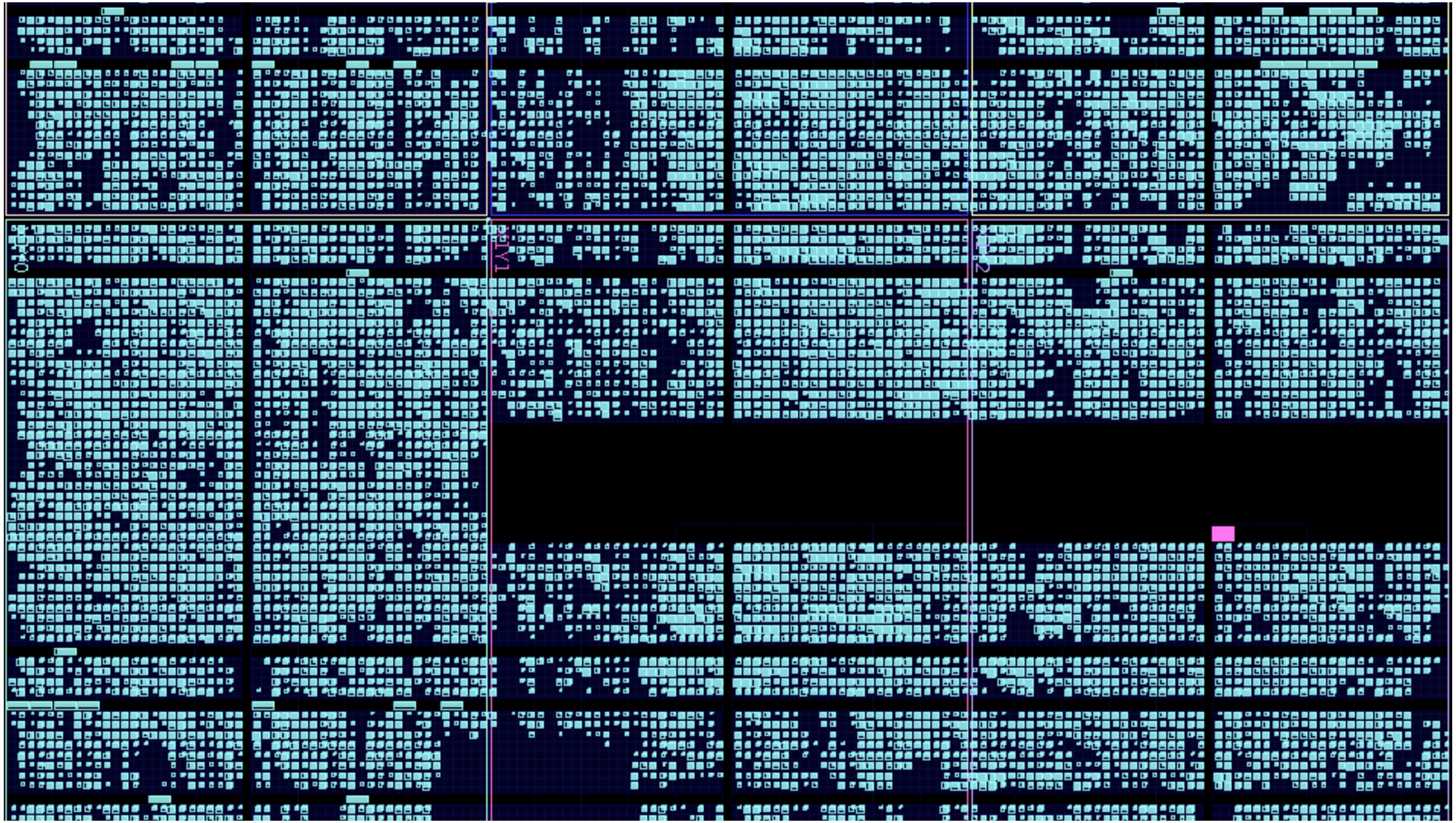




Deep Learning and LHC Big Data

- *Possible solution to the HL-LHC Big-data problem: Deep Learning to be faster and better in what we do today, freeing resources for new ideas*
- *In this seminar, I will highlight a few examples of this*
- *One BIG challenge: DL deployment needs to happen in between collisions and data analysis (trigger, reconstruction, …), where freeing resources will make a difference*
- *Other issue: our data are not mainstream Deep Learning data. Work needed to adapt techniques*



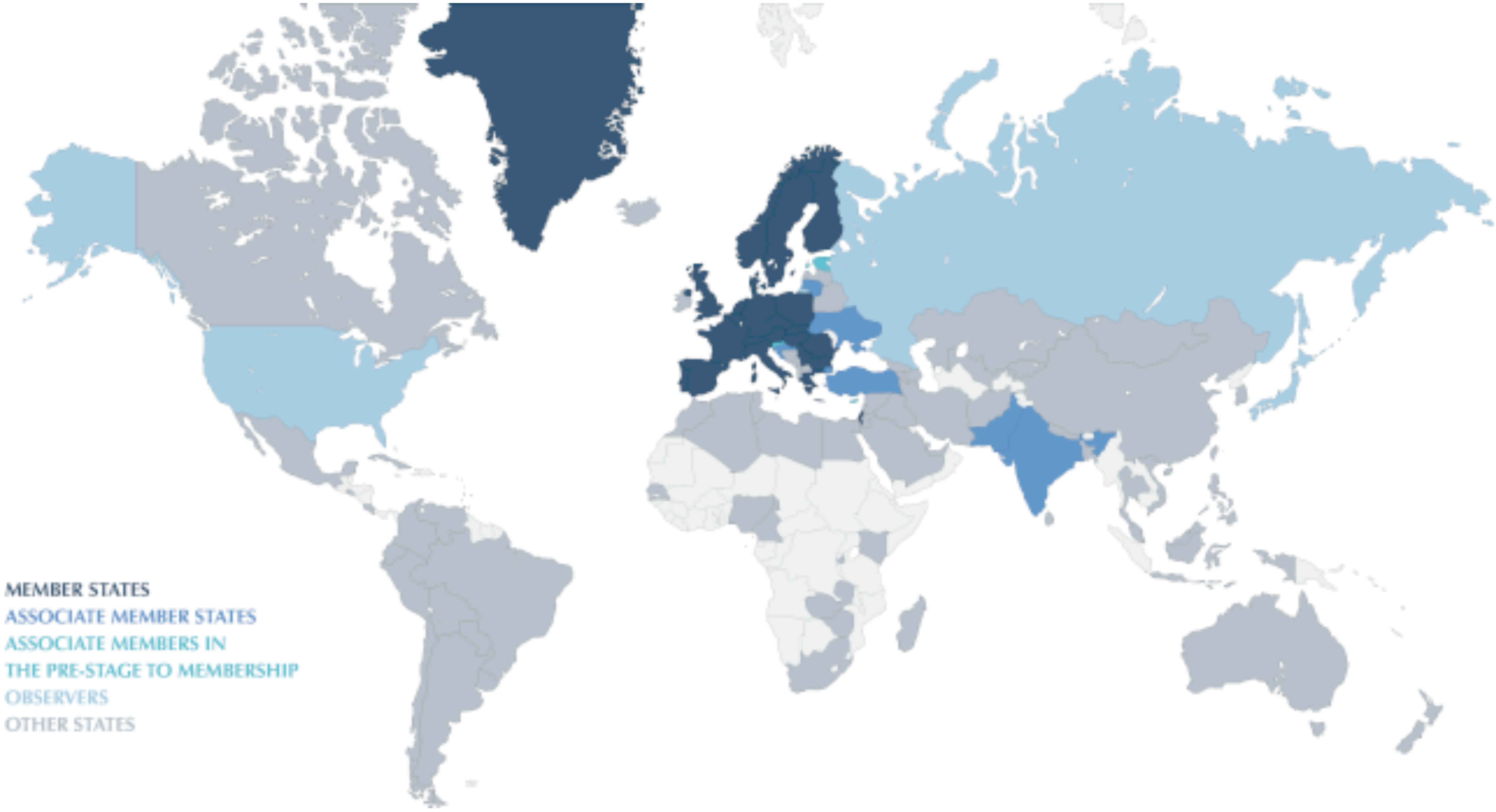


The CERN ML community

European Research Council

CERN: a worldwide community

- International community that goes beyond CERN research staff
  - We (CERN physicists) are a minority
  - Most physicists at CERN visit from other universities
  - Organized in independent scientific collaborations (the experiments: ATLAS, CMS, LHCb, ALICE, DUNE, …)
  - O(1000) people in each
  - Across experiments, a growing community of researchers applying ML to many problems

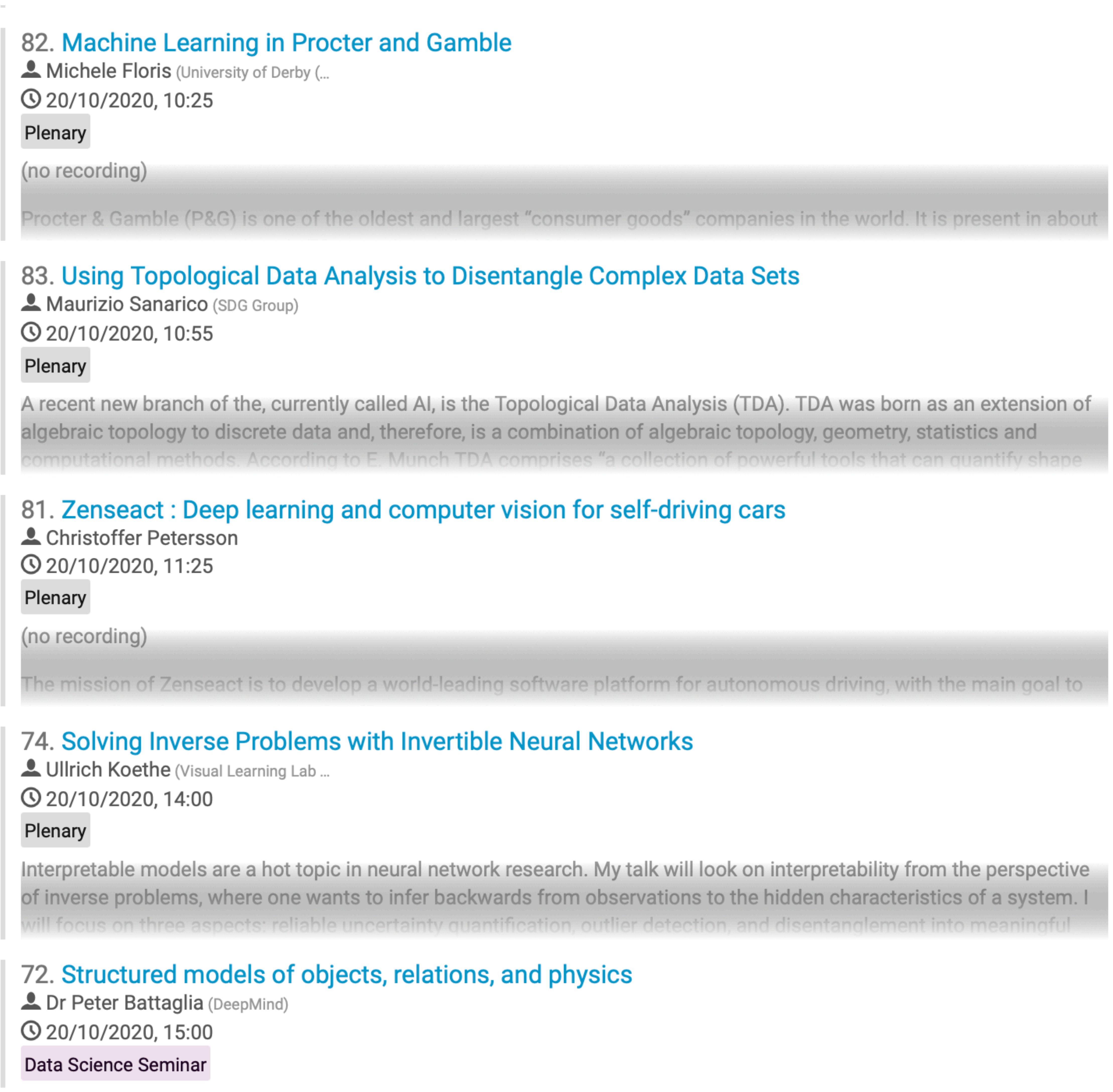


Conferences & Workshops

- Within years, DL discussion in our community has been carried on in dedicated workshops
  - DS@LHC (then DS@HEP) from 2015 to 2017
  - ML4Jets (since 2017)
  - DarkMachines (with astro)
  - …
- In special sessions at dedicated conferences
  - CHEP, ACAT, etc.
- And in workshops @ML conferences
  - ML4PS workshop @NeurIPS
  - AI & Physics @AMLD
  - …

The iML group

- The iML group is a cross-experiment forum at CERN
  - representative from all LHC experiments + Theory
  - Monthly meetings on various subjects
  - Yearly workshop with invited talks from ML community



82. Machine Learning in Procter and Gamble
Michele Floris (University of Derby (...
20/10/2020, 10:25
Plenary
(no recording)
Procter & Gamble (P&G) is one of the oldest and largest "consumer goods" companies in the world. It is present in about

83. Using Topological Data Analysis to Disentangle Complex Data Sets
Maurizio Sanarico (SDG Group)
20/10/2020, 10:55
Plenary
A recent new branch of the, currently called AI, is the Topological Data Analysis (TDA). TDA was born as an extension of algebraic topology to discrete data and, therefore, is a combination of algebraic topology, geometry, statistics and computational methods. According to E. Munch TDA comprises "a collection of powerful tools that can quantify shape

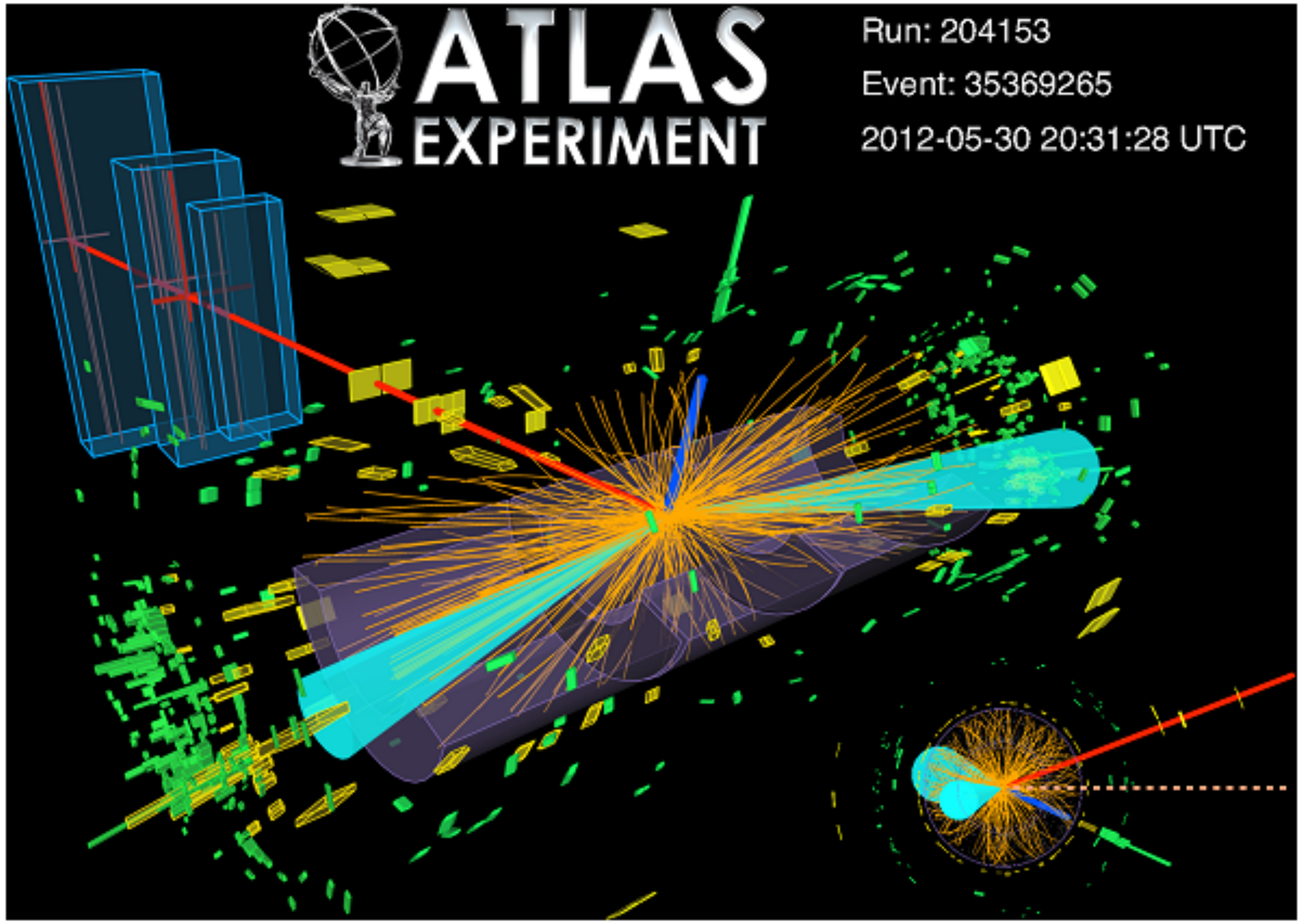
81. Zenseact : Deep learning and computer vision for self-driving cars
Christoffer Petersson
20/10/2020, 11:25
Plenary
(no recording)
The mission of Zenseact is to develop a world-leading software platform for autonomous driving, with the main goal to

74. Solving Inverse Problems with Invertible Neural Networks
Ullrich Koethe (Visual Learning Lab ...
20/10/2020, 14:00
Plenary
Interpretable models are a hot topic in neural network research. My talk will look on interpretability from the perspective of inverse problems, where one wants to infer backwards from observations to the hidden characteristics of a system. I will focus on three aspects: reliable uncertainty quantification, outlier detection, and disentanglement into meaningful

72. Structured models of objects, relations, and physics
Dr Peter Battaglia (DeepMind)
20/10/2020, 15:00
Data Science Seminar

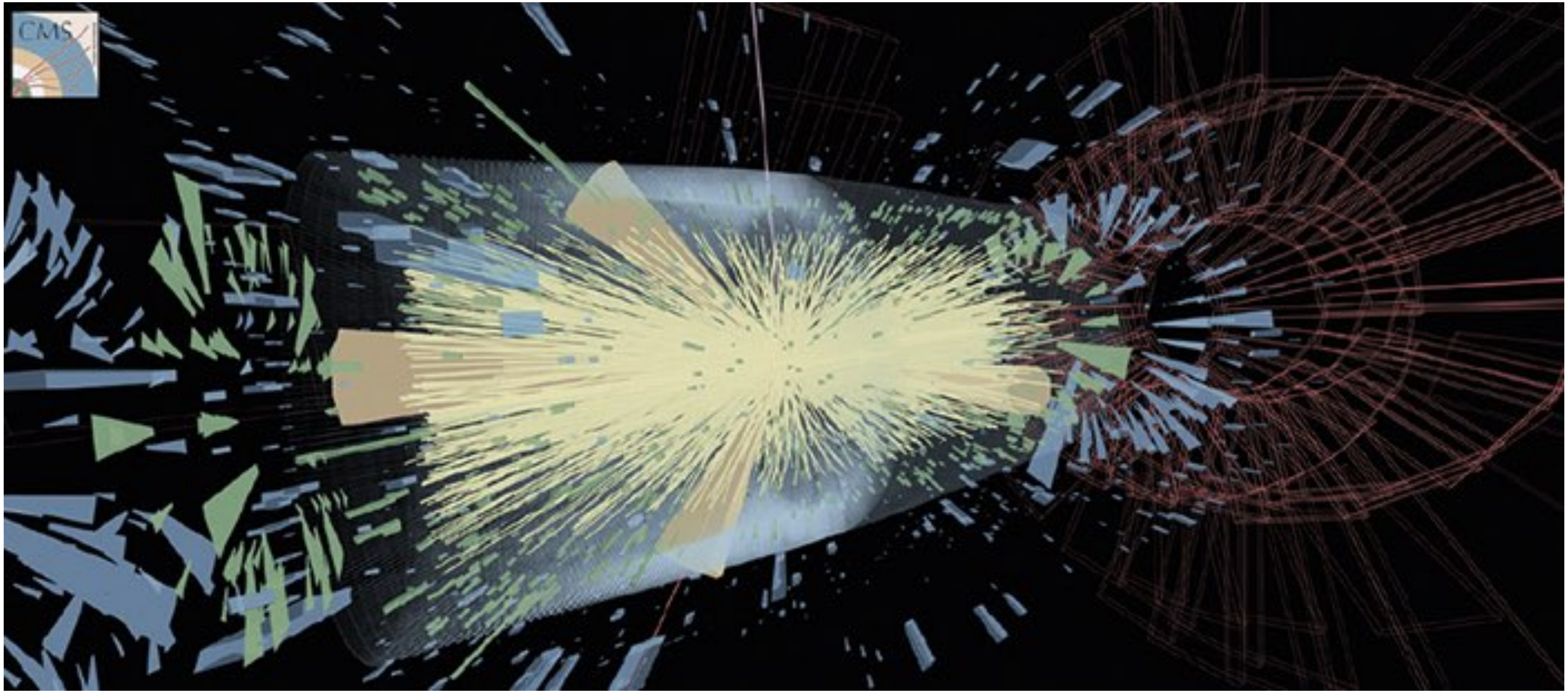
Data Challenges

- *In recent years, data challenges served as opportunity to attract attention of ML community on our problems*
  - *Higgs Kaggle challenge (classification)*
  - *TrackML Kaggle challenge (pattern recognition)*
  - *Flavor Kaggle challenge (classification)*
  - *LHC Olympics (anomaly detection)*
  - *DarkMachines (anomaly detection)*
  - *40 MHz Anomaly detection (anomaly detection)*



Disclaimer

- *This meeting is about CERN and ELLIS*
  - *For this reason, I focused on what CERN physicists in the experiments are working on*
  - *You should keep in mind that the ML effort for ML experiments goes beyond this*
    - *Many more groups all over the world developing ML-based solutions for the LHC experiments, neutrinos, etc.*
    - *By joining ELLIS, CERN would act as a bridge between the ELLIS community and this large HEP-ML worldwide community*

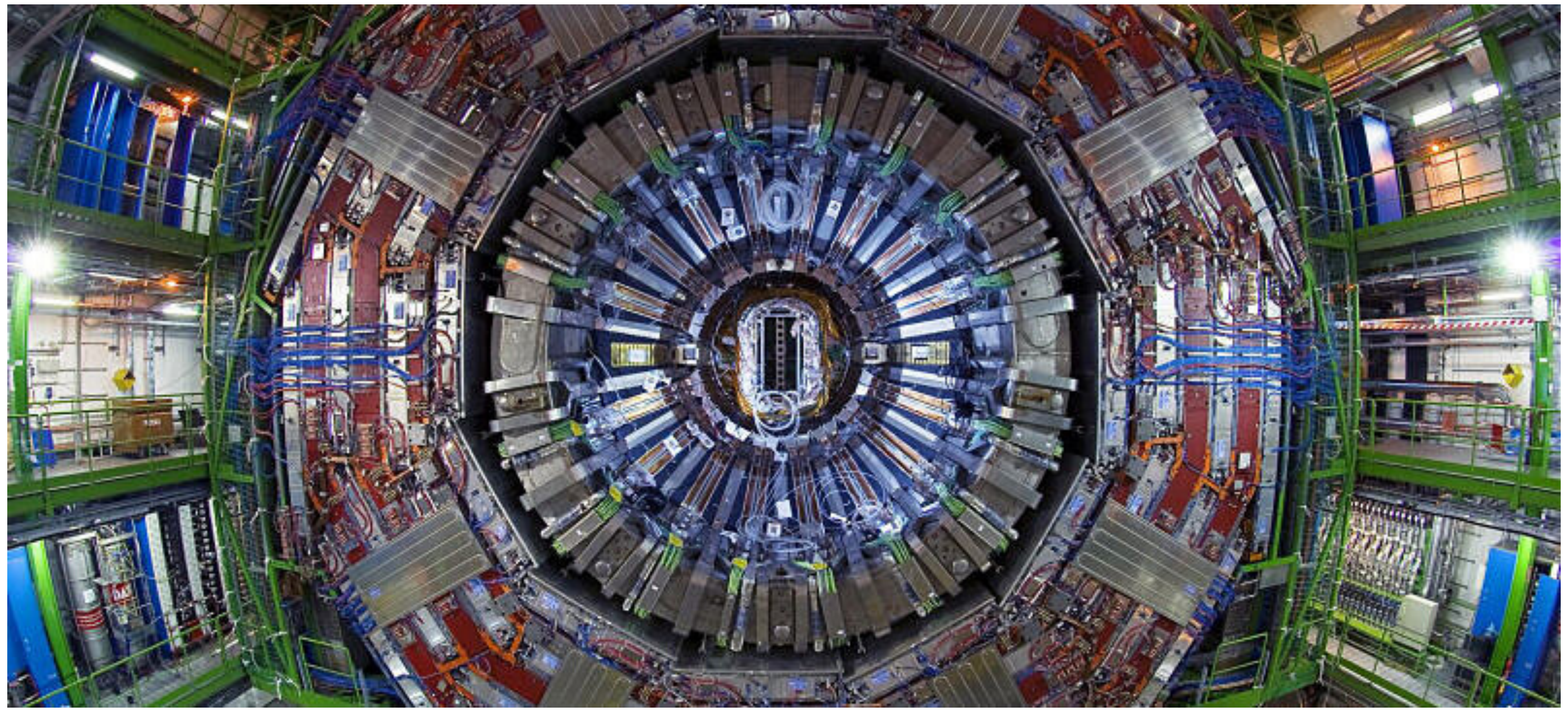


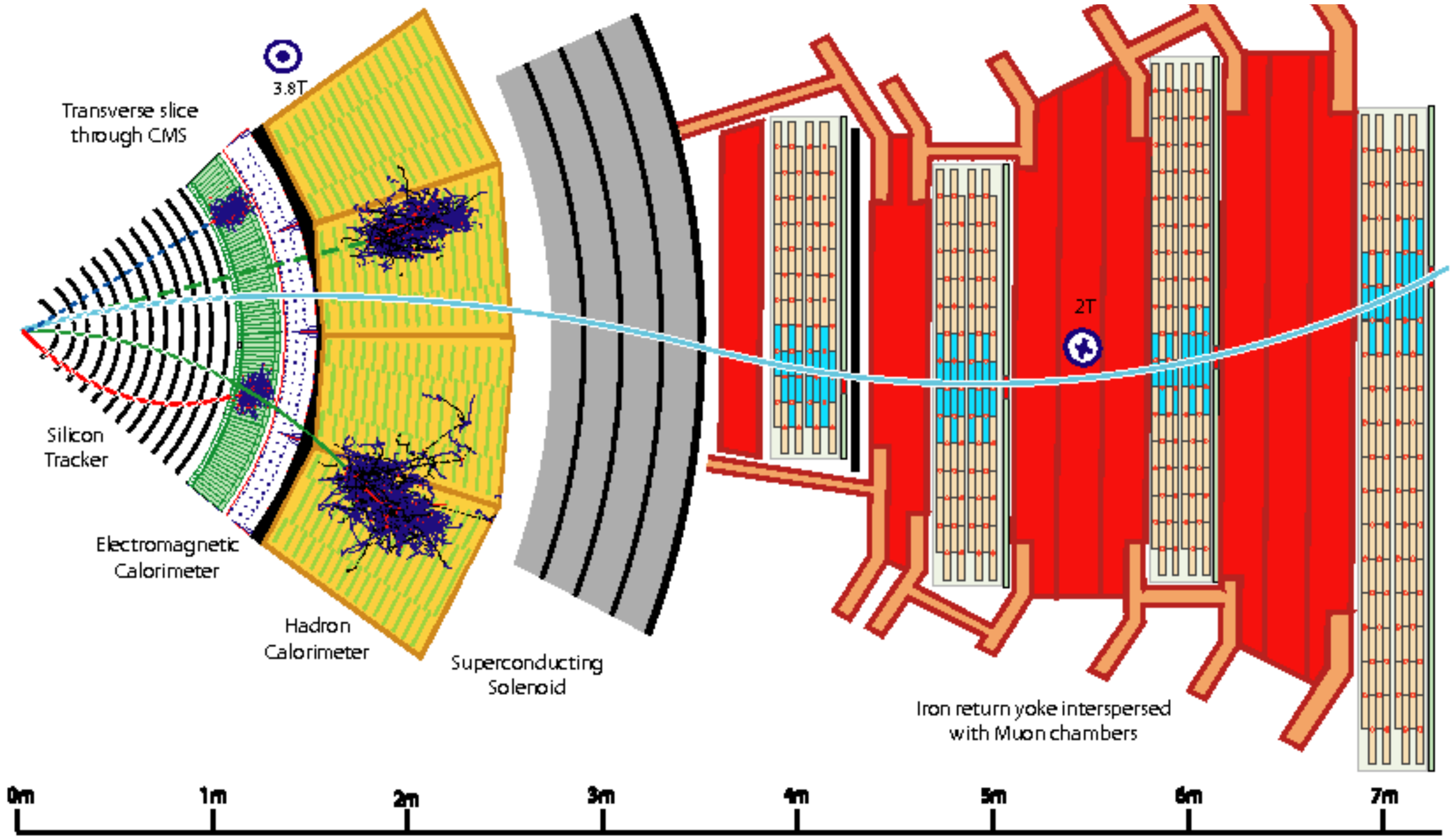
Local Reconstruction

European
Research
Council

Dealing with Real Life Detectors

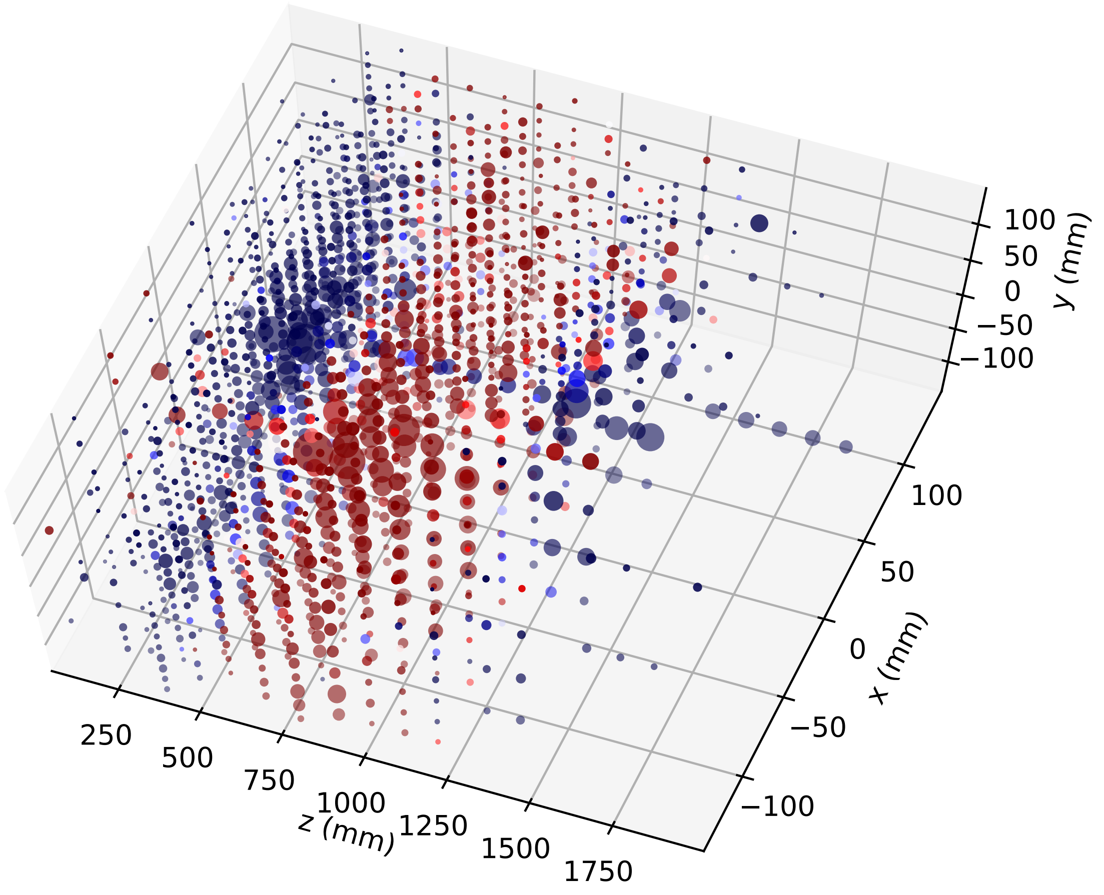
- Most of HEP-related DL literature uses ConvNNs
- In practice, little of that made it to production so far
- Main issue (IMO): difficult to fit an irregular array of sensors (unordered set of dots in some feature space) in a regular array of pixels
- Several solutions attempted
    - pixelate the data with a coarser binning
    - use recurrent networks (imposing some use-case-specific ordering criterion)
    - use graph networks

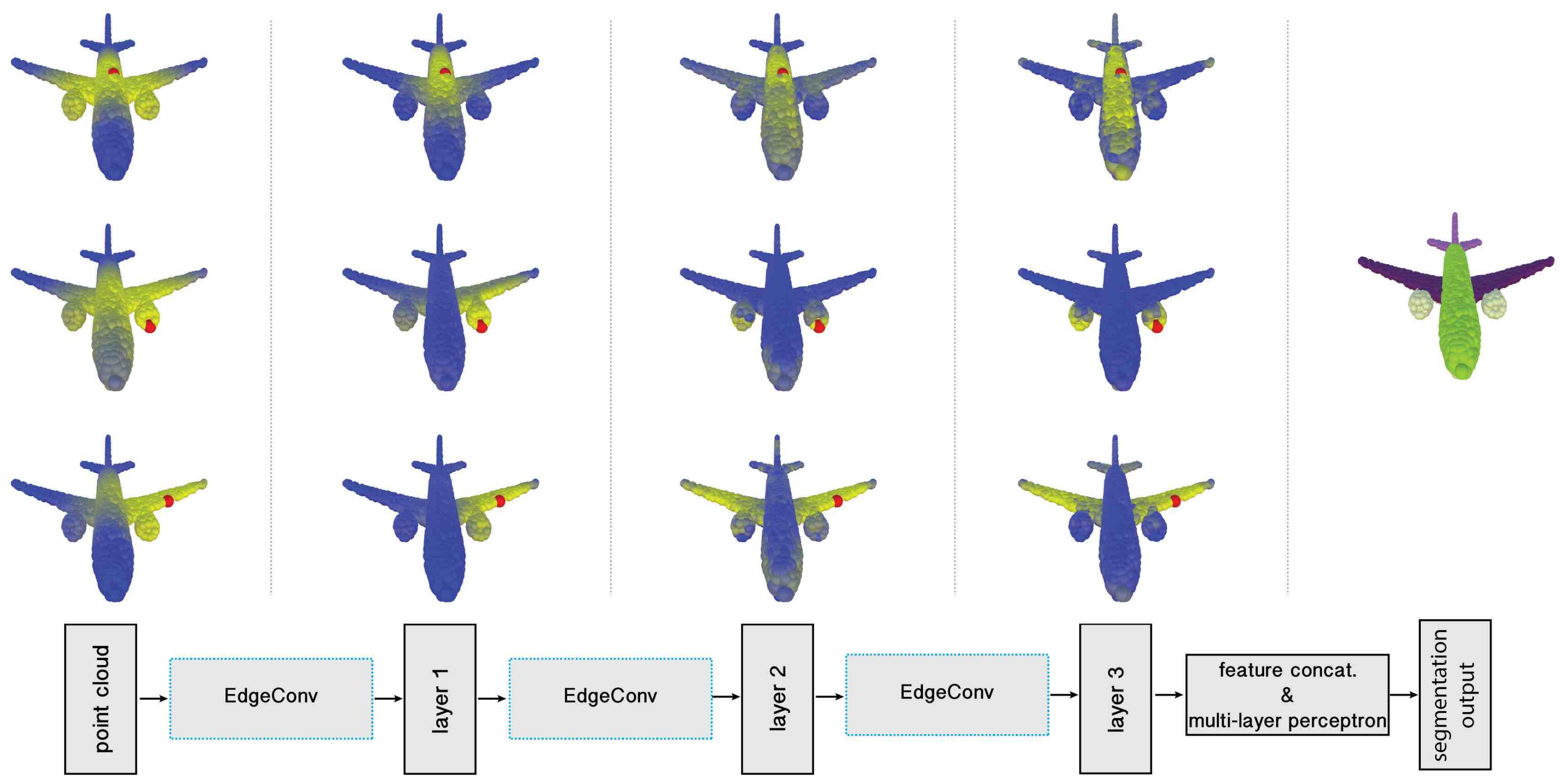


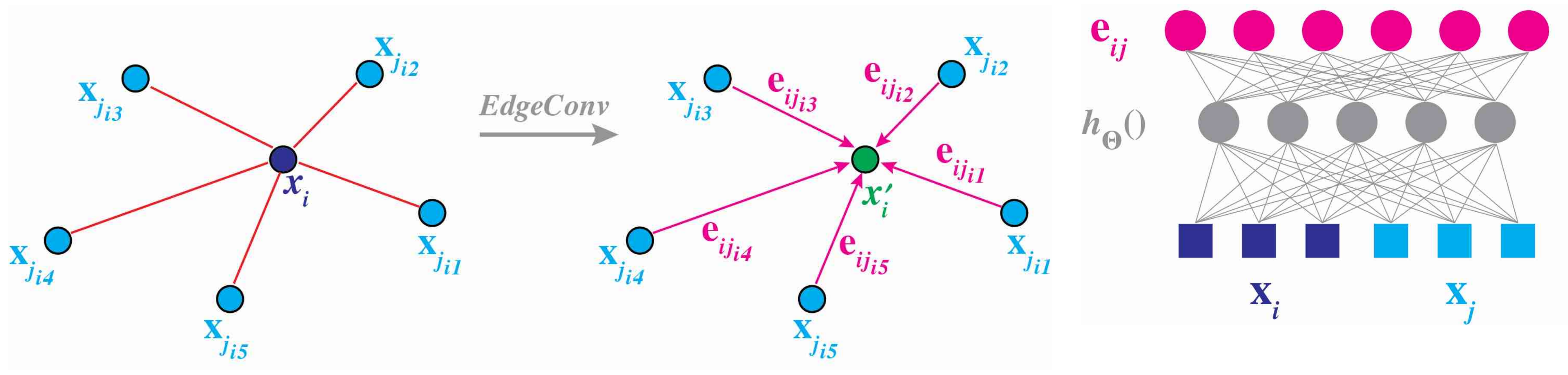


EdgeConv for Particle Physics

- *Graphs can be very functional to process raw data*
  - *each detector hit represented as coordinates + energy*
- *EdgeConv used as a baseline*
  - *Pros: no assumption on the underlying geometry*
  - *Cons: large memory consumption (large number of connections)*

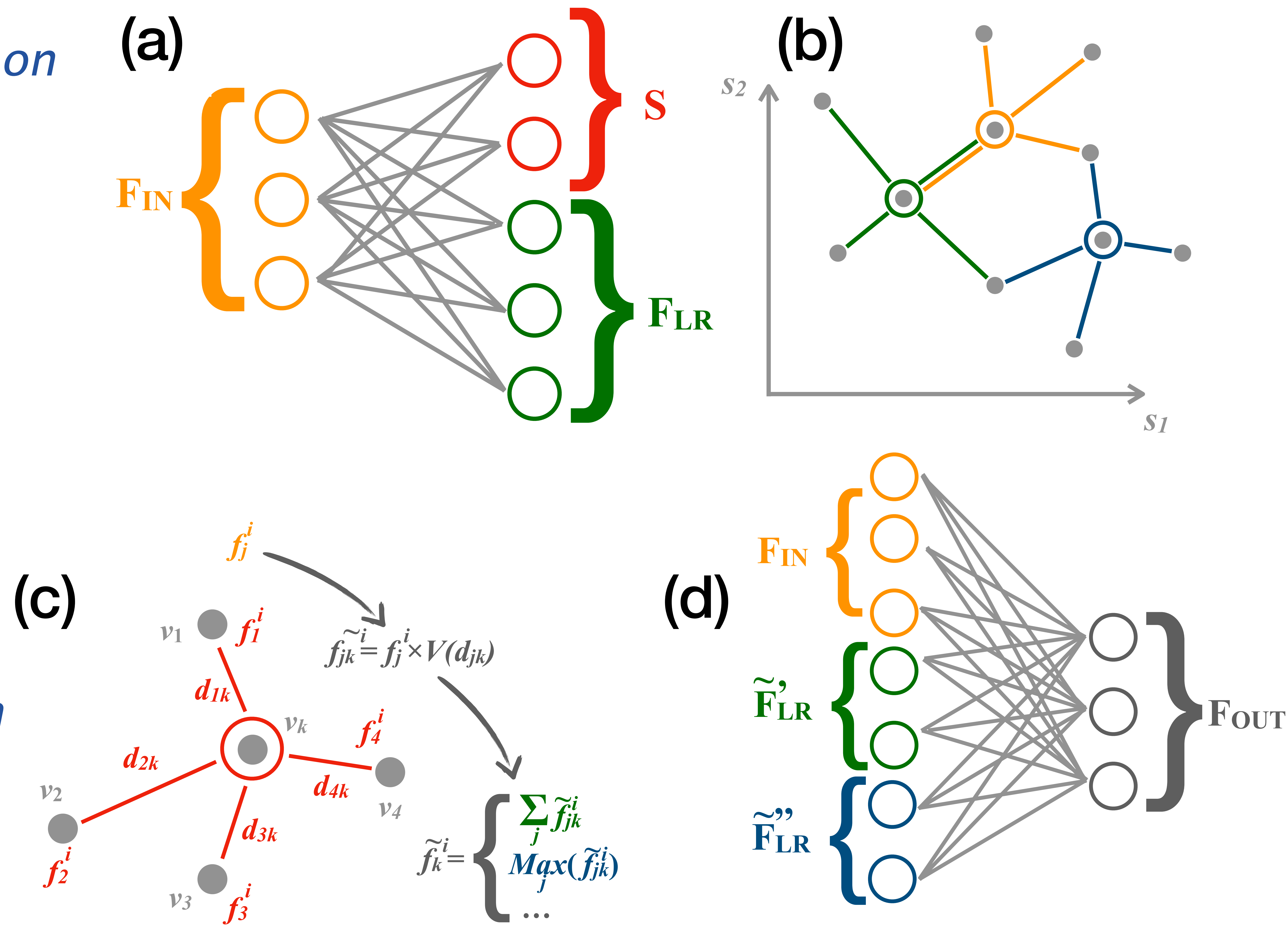






Reducing memory consumption

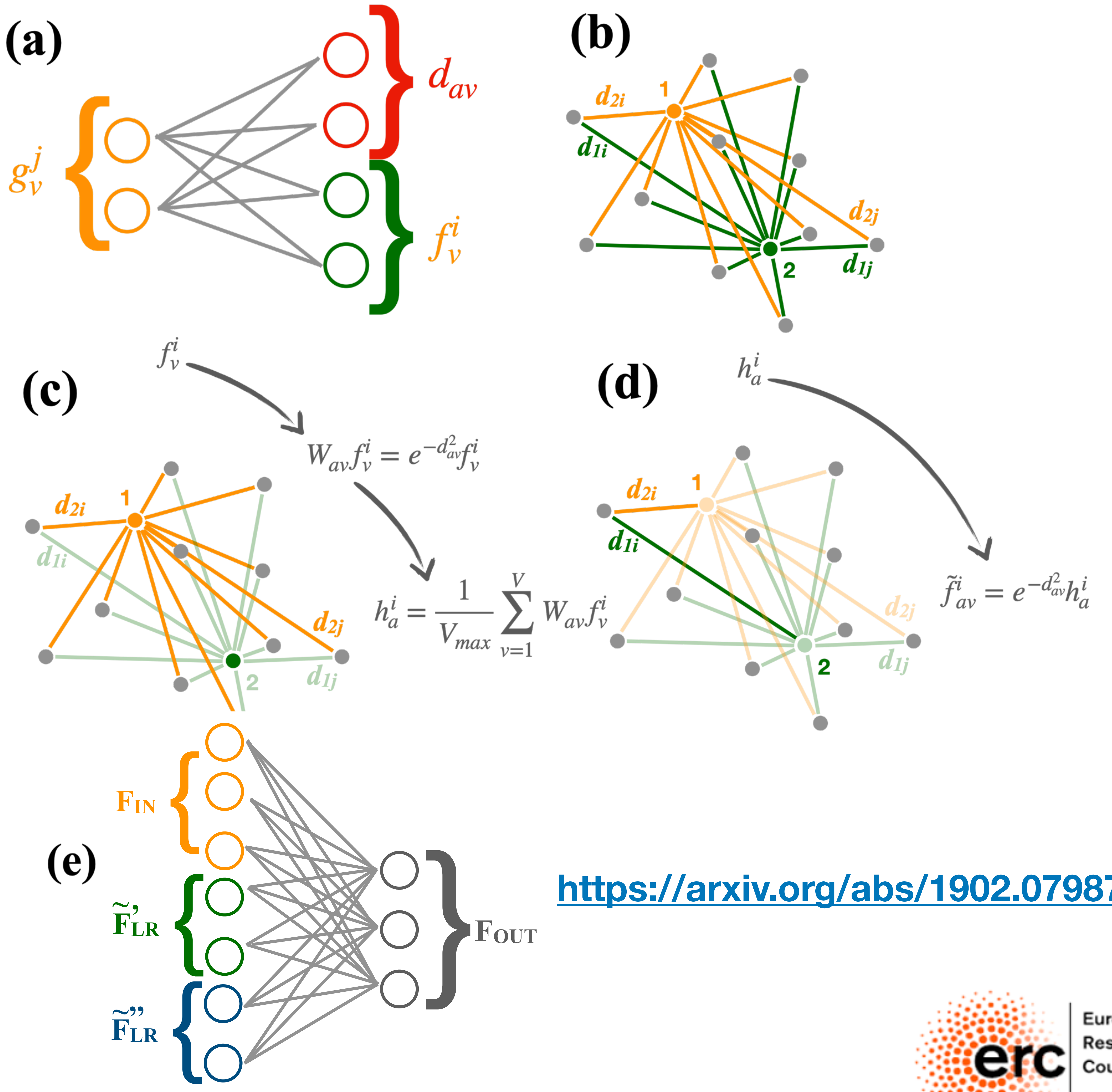
- Introduce step to learn an optimized spatial distribution
  - as coordinates of a new space (b)
  - as distances from a fixed number of aggregators (c)
- Use customised architectures to keep resources under control
  - Gravnet: weight connection by potential of euclidean distance



https://arxiv.org/abs/1902.07987

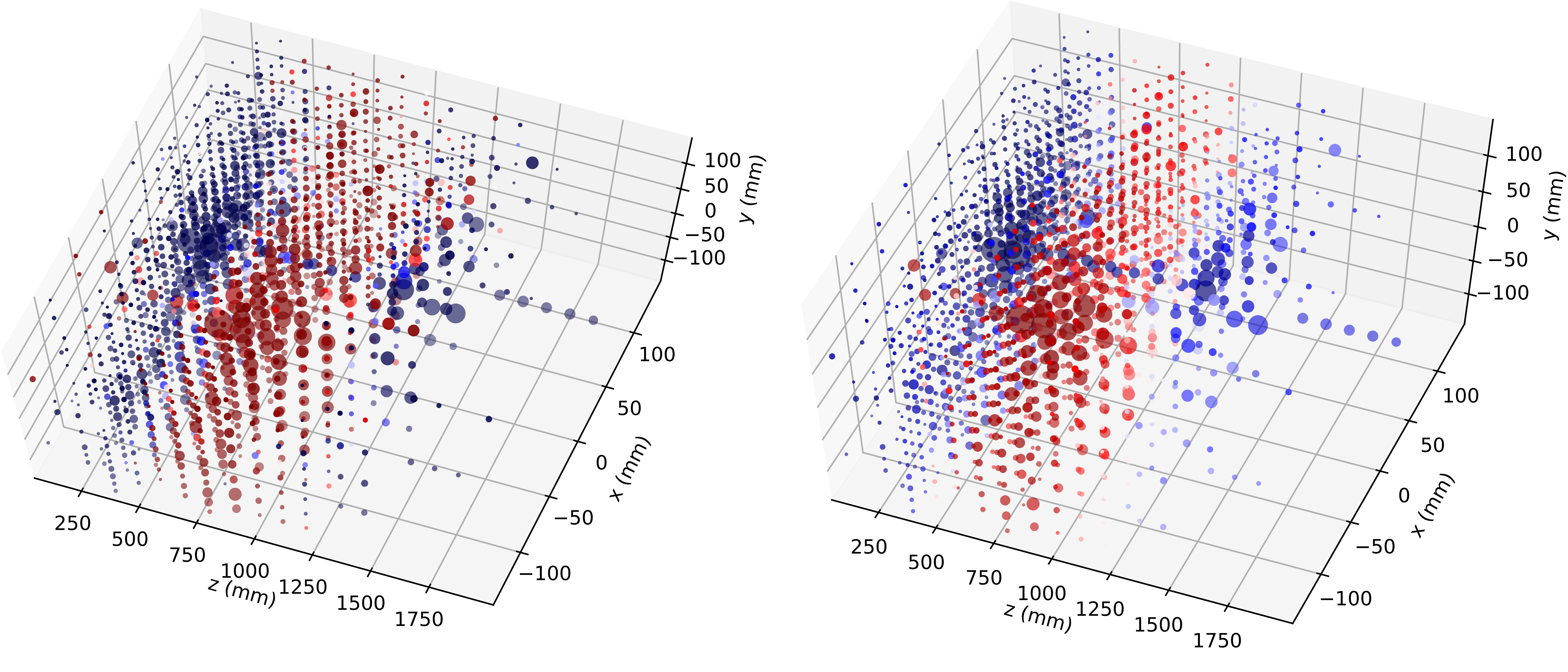
Reducing memory consumption

- *Introduce step to learn an optimized spatial distribution*
  - *as coordinates of a new space (b)*
  - *as distances from a fixed number of aggregators (c)*
- *Use customised architectures to keep resources under control*
  - *Gravnet: weight connection by potential of euclidean distance*
  - *Garnet: keep number of connections small through small number of aggregators*



https://arxiv.org/abs/1902.07987

Separating overlapping showers

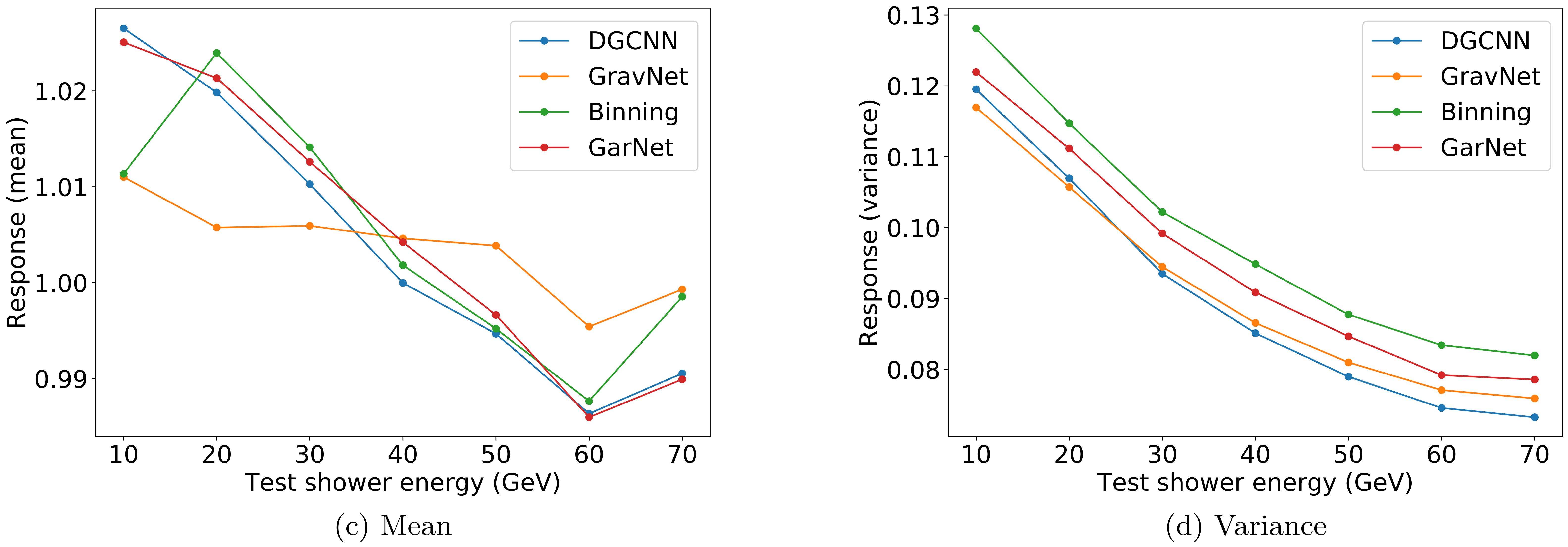


(a) Truth

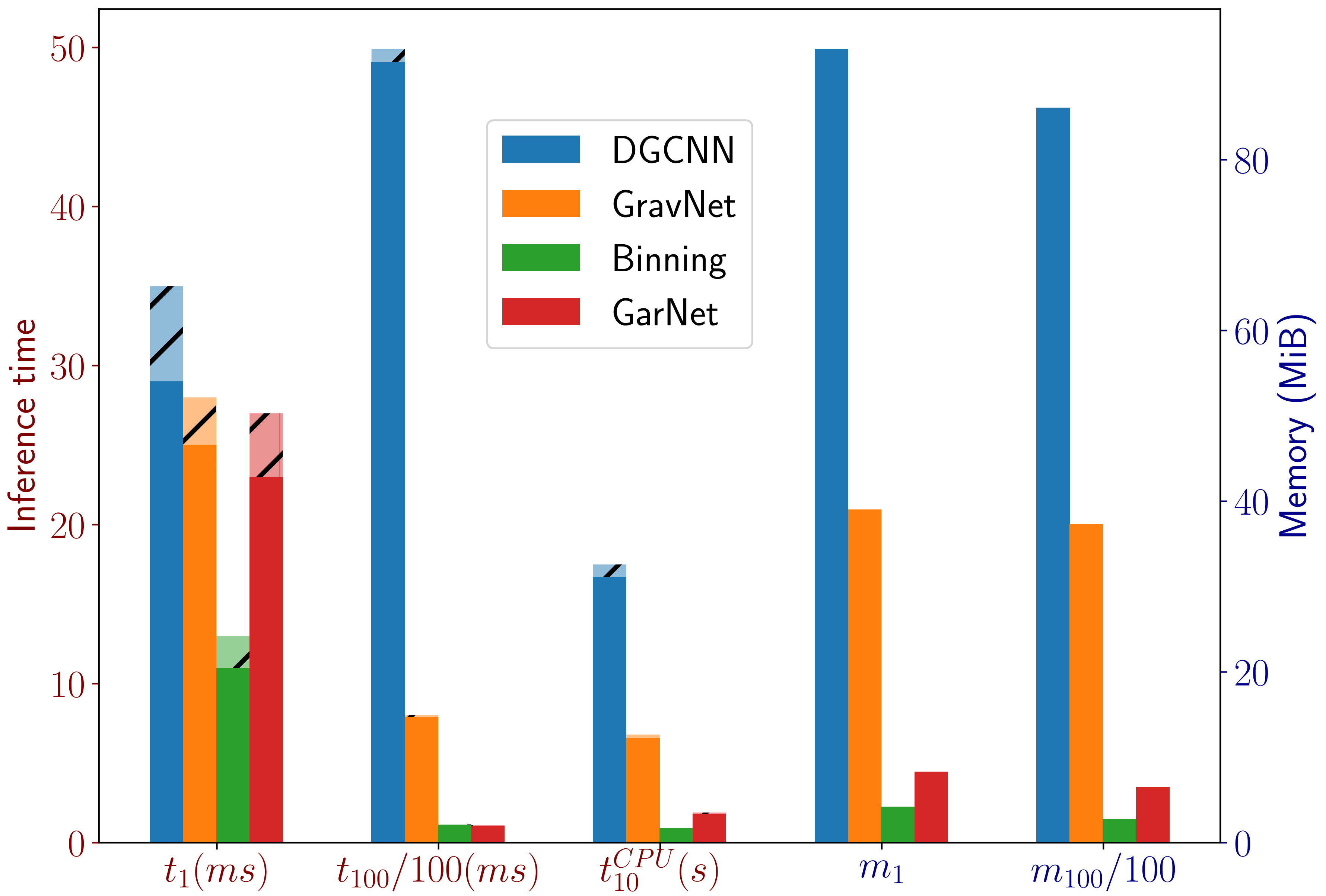
(b) Reconstructed

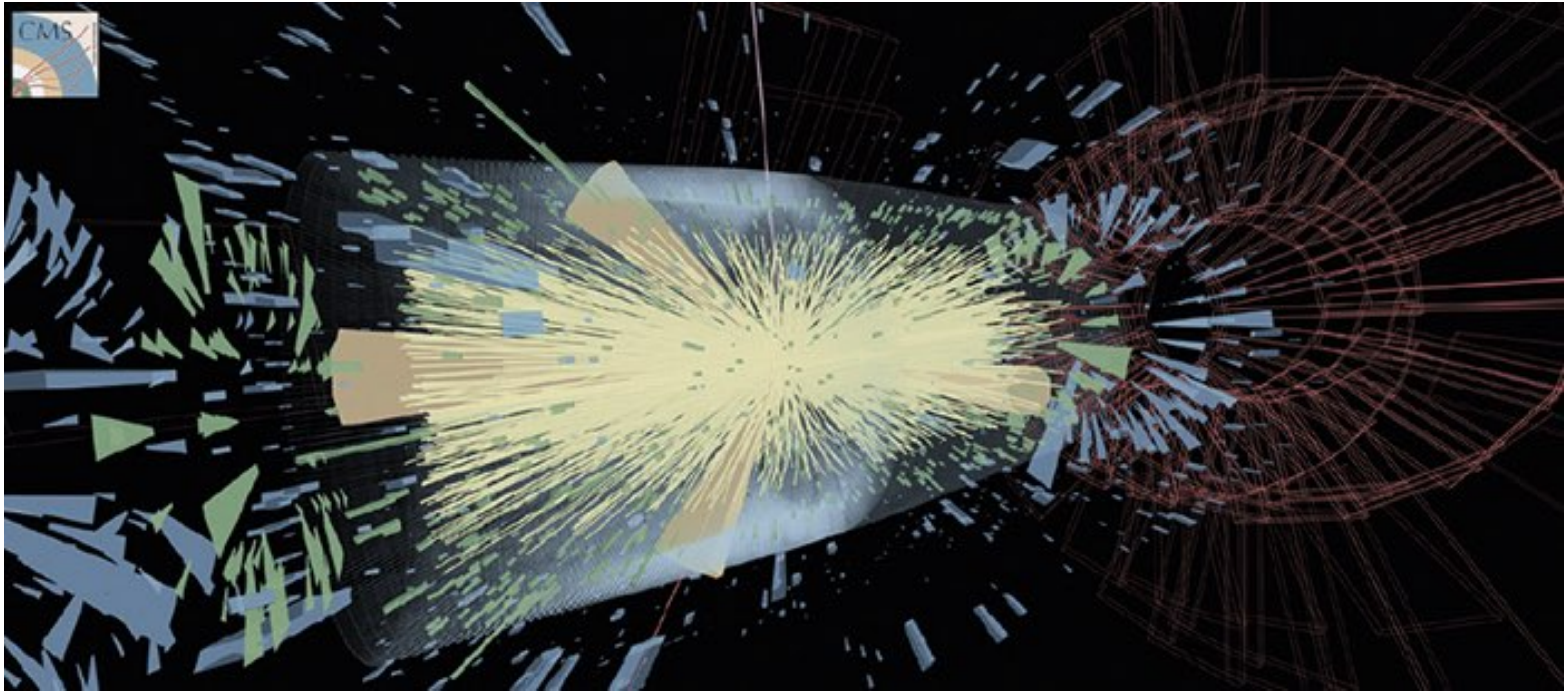
GraphNets for Calorimetry

- Good performance achieved, comparable to more traditional approaches
- Using a potential (V(d)) to weight up the near neighbours allows to keep memory footprint under control (with respect to other graph approaches)



(c) Mean
(d) Variance

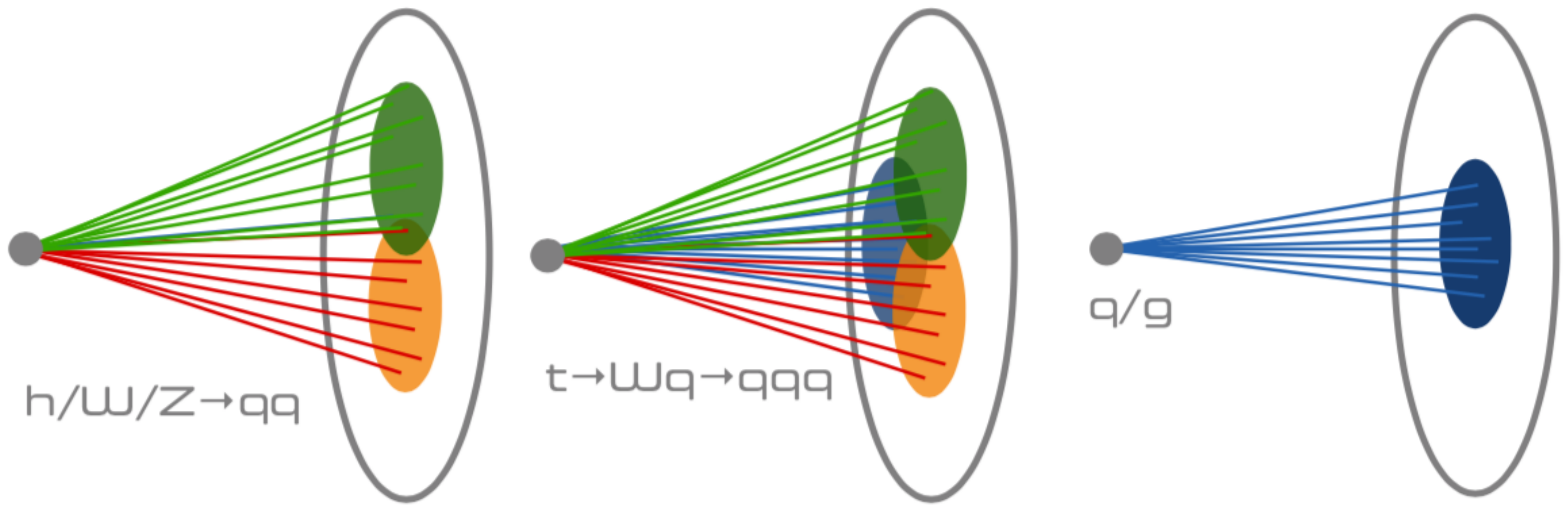


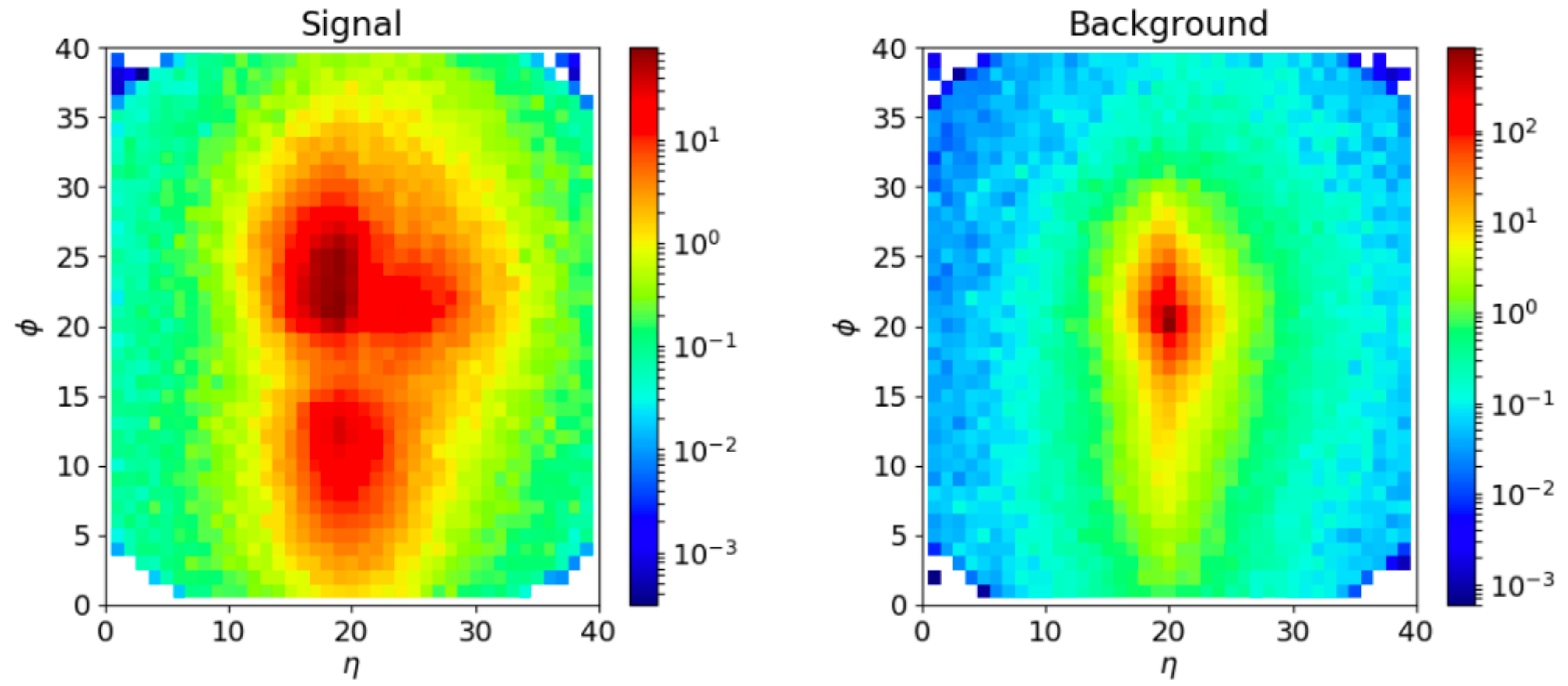


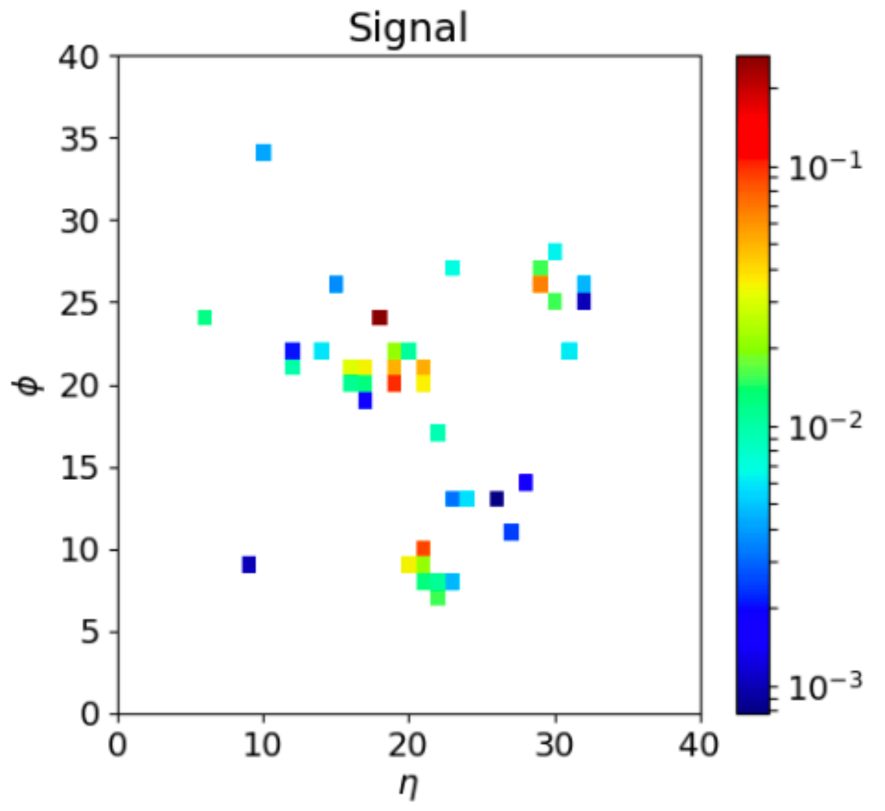
Jet Tagging

Jet tagging

- You have a jet at LHC: spray of hadrons coming from a “shower” initiated by a fundamental particle of some kind (quark, gluon, W/Z/H bosons, top quark)
- You have a set of jet features whose distribution depends on the nature of the initial particle
- You can train a network to start from the values of these quantities and guess the nature of your jet

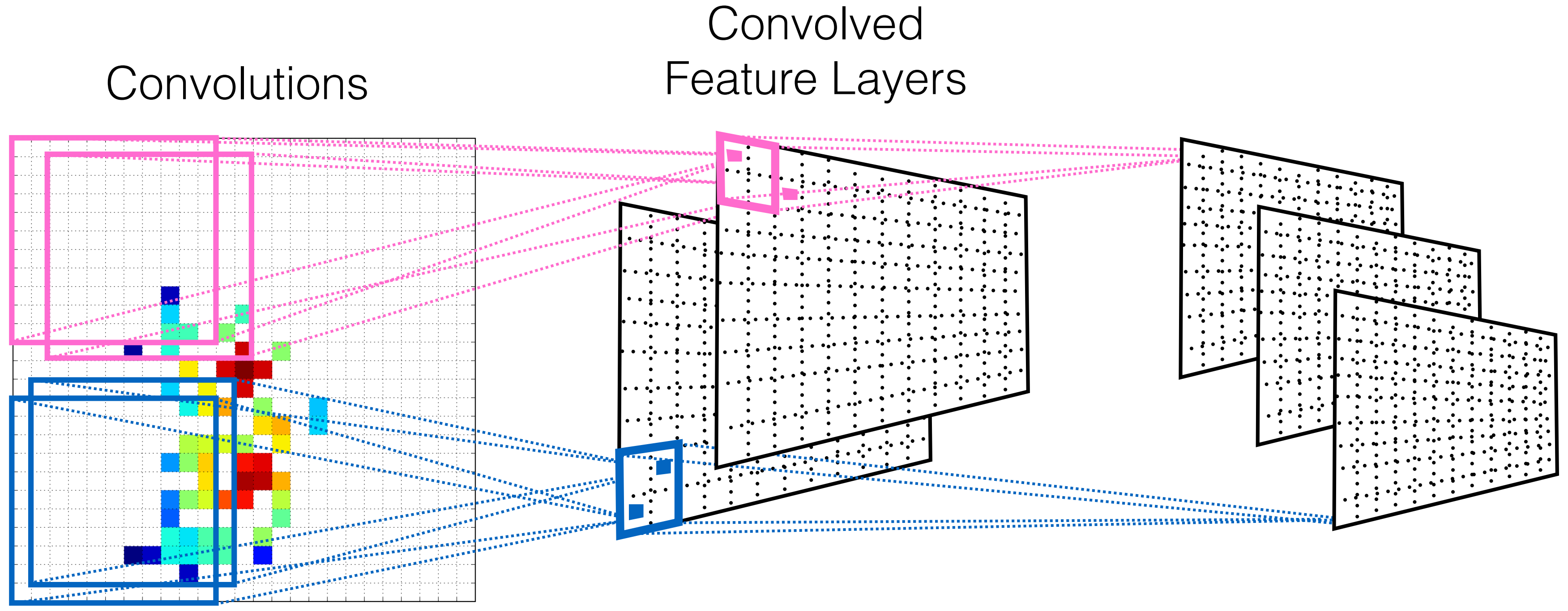




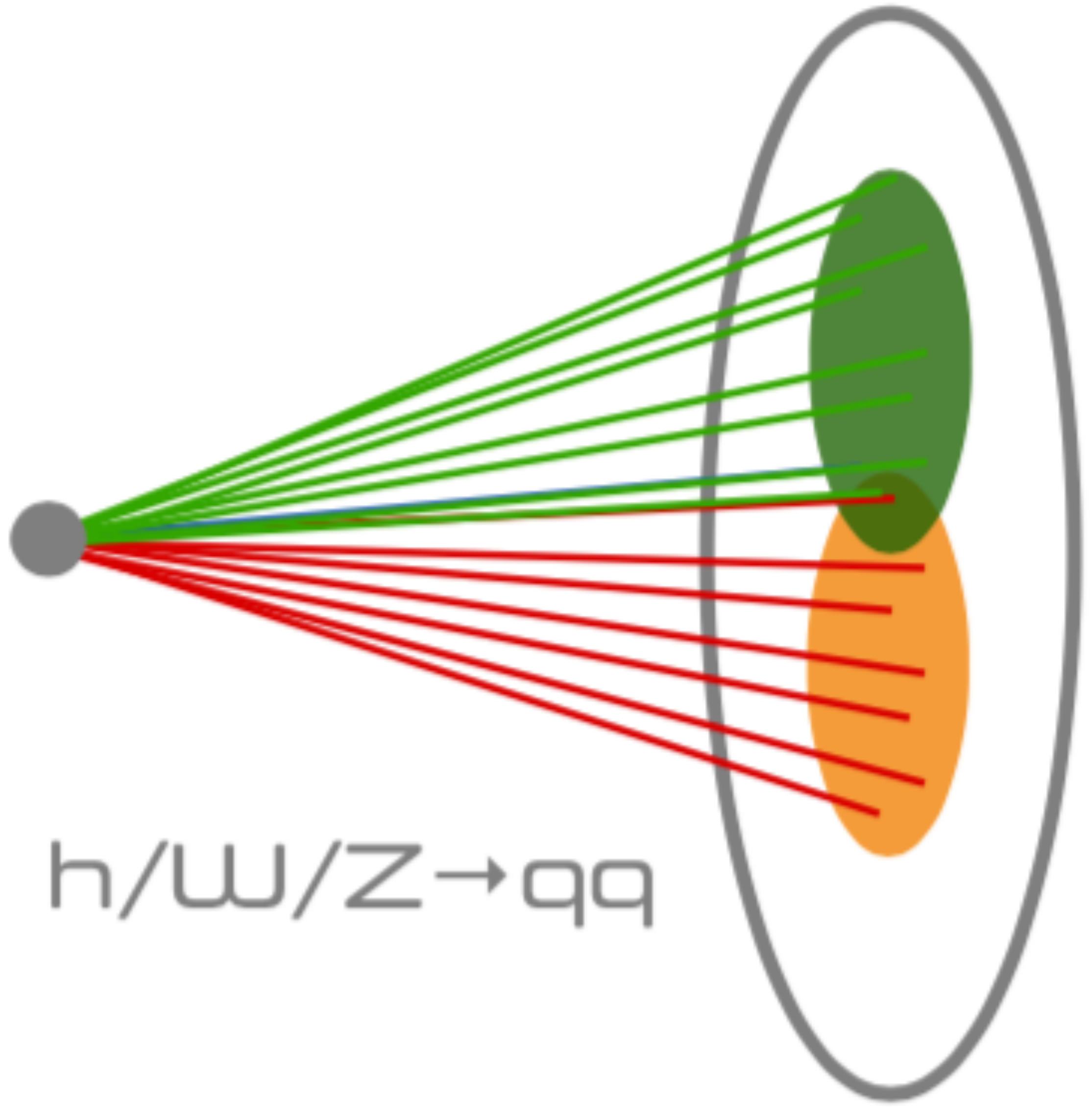


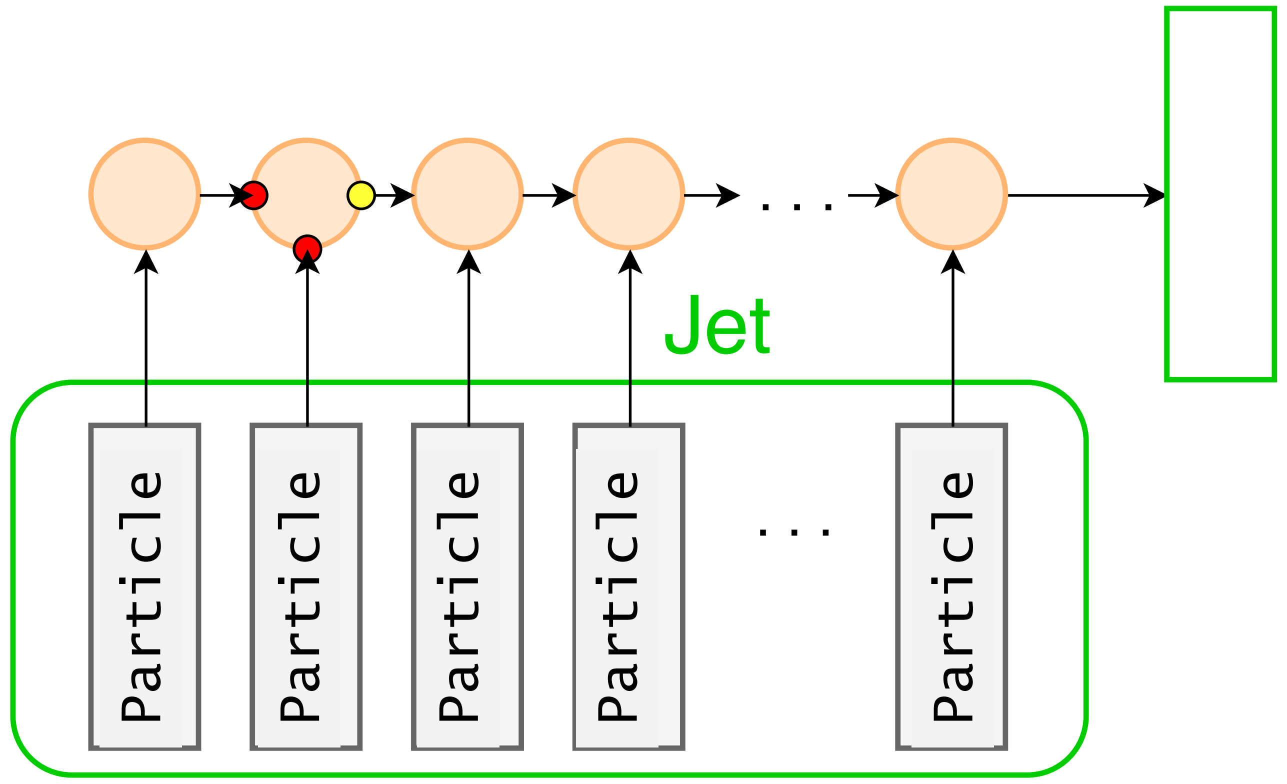
One problem, many solutions

- **Data as images**
  - *can use computing vision techniques*

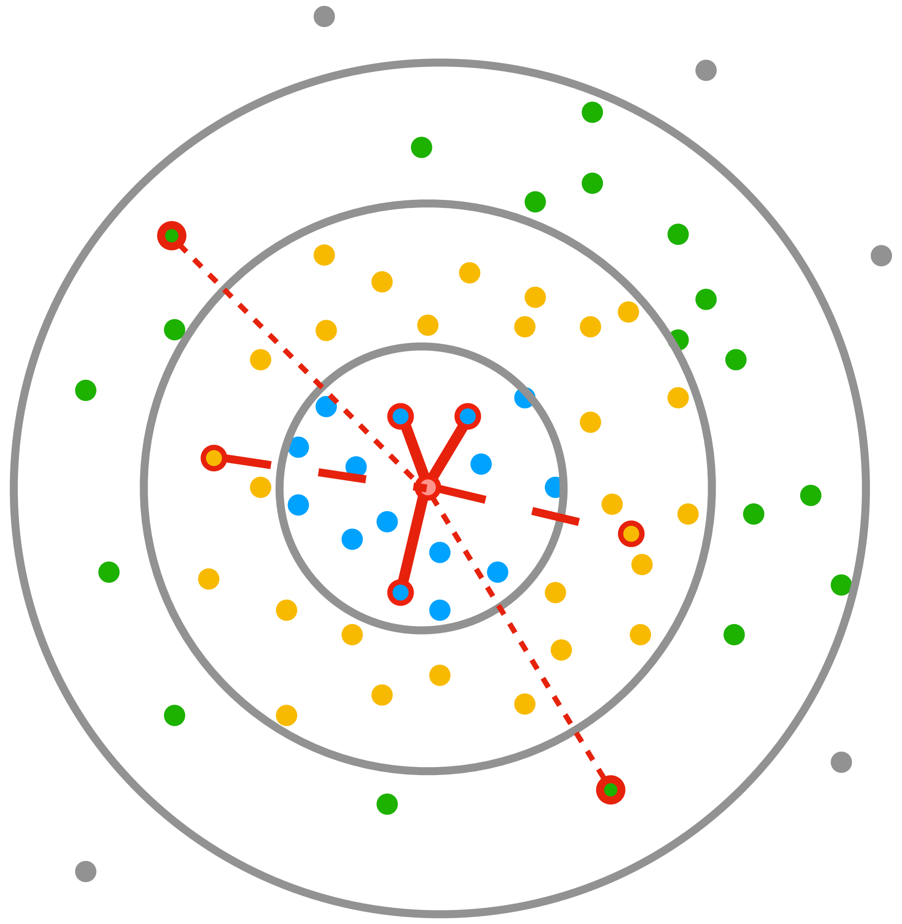


- **Data as sequences**
  - *can use text processing techniques*





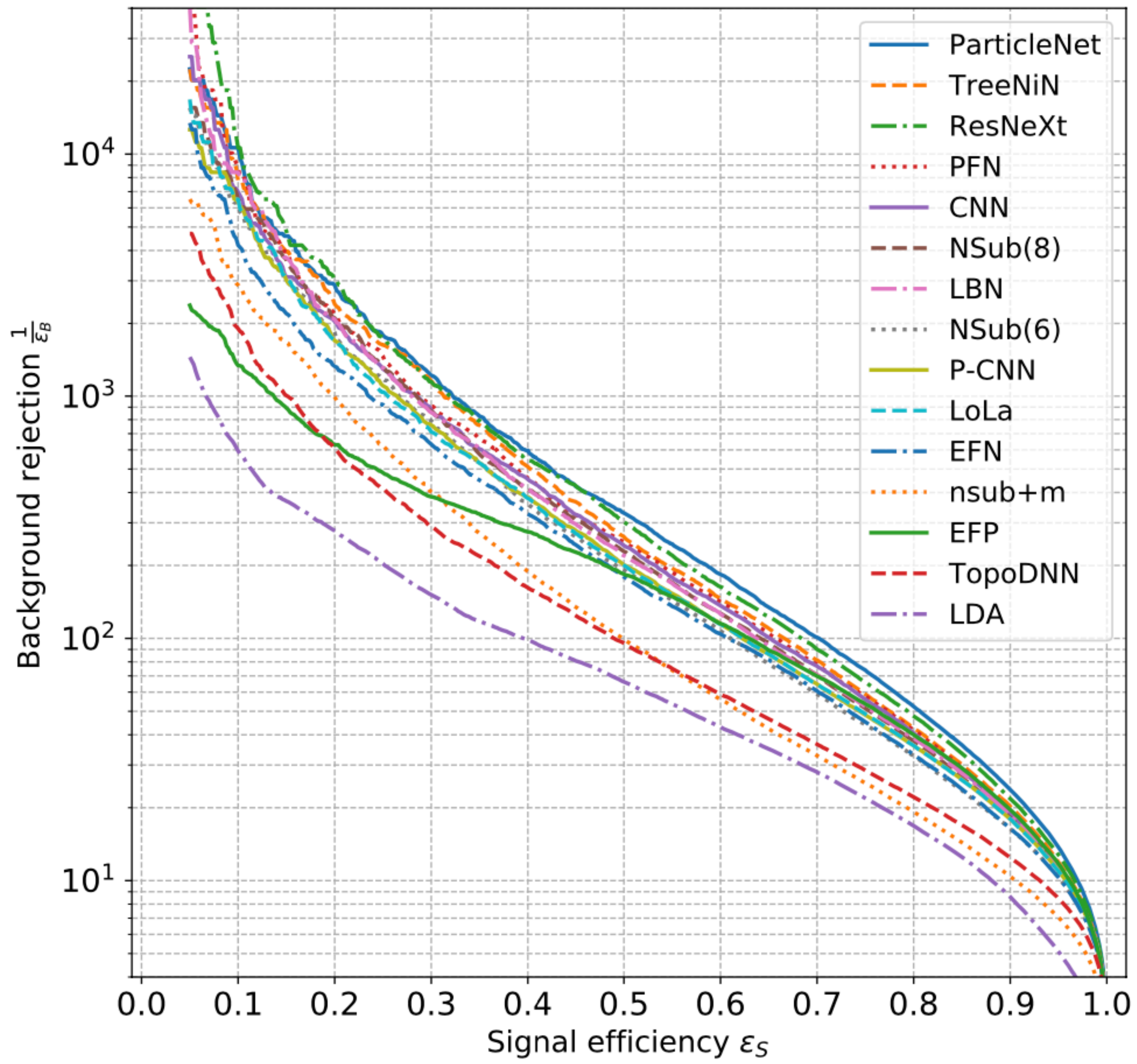
- **Data as graphs of points**
  - *can use graph networks, as in social-media analyses*



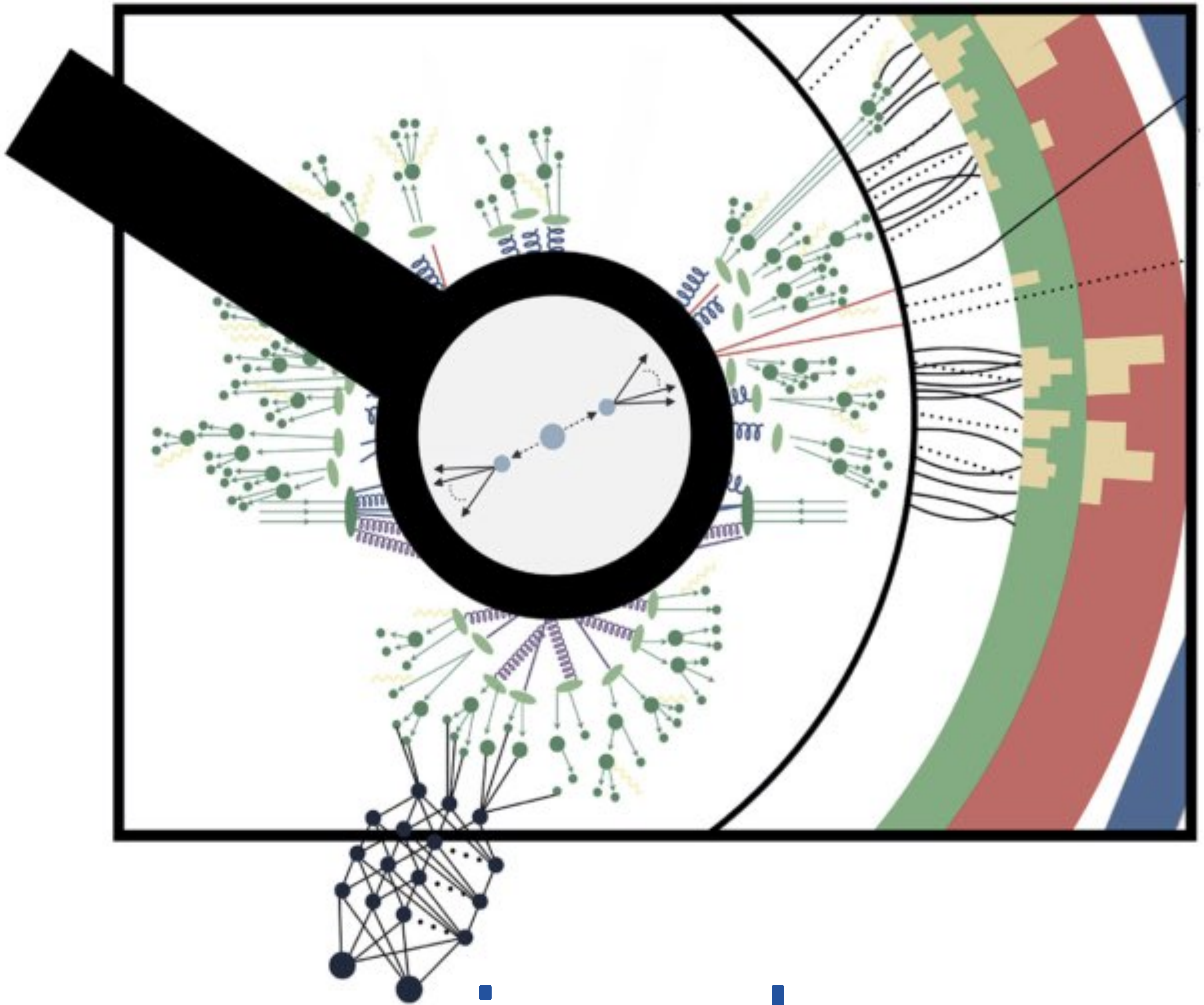
Example: Top Taggers

	AUC	Acc	$1/\epsilon_B$ ($\epsilon_S = 0.3$) single	mean	median	#Param
CNN [16]	0.981	0.930	914±14	995±15	975±18	610k
ResNeXt [31]	0.984	0.936	1122±47	1270±28	1286±31	1.46M
TopoDNN [18]	0.972	0.916	295±5	382± 5	378 ± 8	59k
Multi-body N-subjettiness 6 [24]	0.979	0.922	792±18	798±12	808±13	57k
Multi-body N-subjettiness 8 [24]	0.981	0.929	867±15	918±20	926±18	58k
TreeNiN [43]	0.982	0.933	1025±11	1202±23	1188±24	34k
P-CNN	0.980	0.930	732±24	845±13	834±14	348k
ParticleNet [47]	0.985	0.938	1298±46	1412±45	1393±41	498k
LBN [19]	0.981	0.931	836±17	859±67	966±20	705k
LoLa [22]	0.980	0.929	722±17	768±11	765±11	127k
LDA [54]	0.955	0.892	151±0.4	151.5±0.5	151.7±0.4	184k
Energy Flow Polynomials [21]	0.980	0.932	384			1k
Energy Flow Network [23]	0.979	0.927	633±31	729±13	726±11	82k
Particle Flow Network [23]	0.982	0.932	891±18	1063±21	1052±29	82k
GoaT	0.985	0.939	1368±140		1549±208	35k

https://arxiv.org/pdf/1902.09914.pdf



- *Several architectures tried on problem*
  - *CNNs, physics motivated custom architectures, PointCloud, etc.*
  - *Best results achieved with graph architectures*



Unsupervised searches

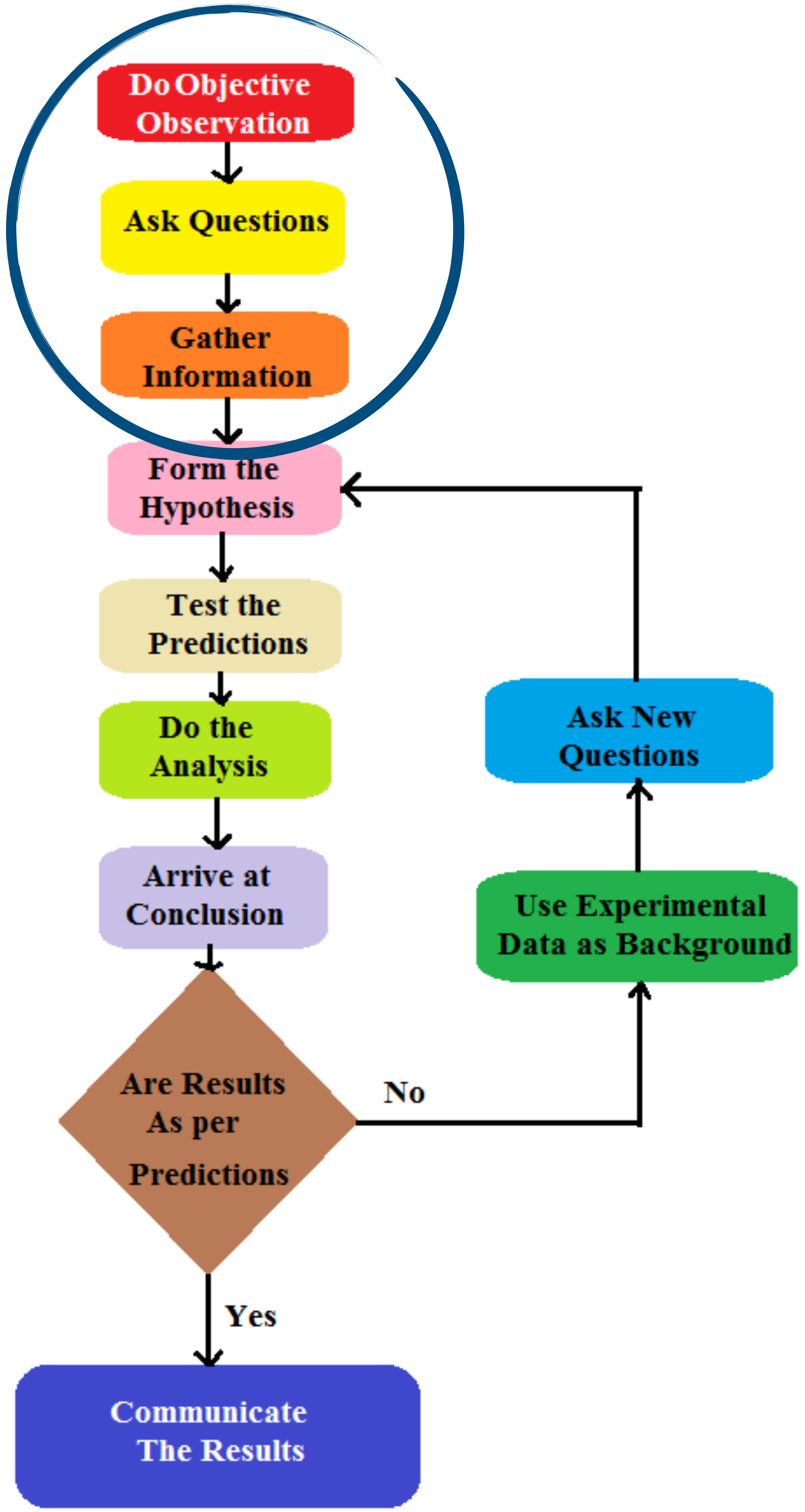
European Research Council

Learning from Data

- Rather than specifying a signal hypothesis upfront, we could start looking at our data
- Based on what we see (e.g., clustering alike objects) we could formulate a signal hypothesis
- EXAMPLE: star classification was based on observed characteristics

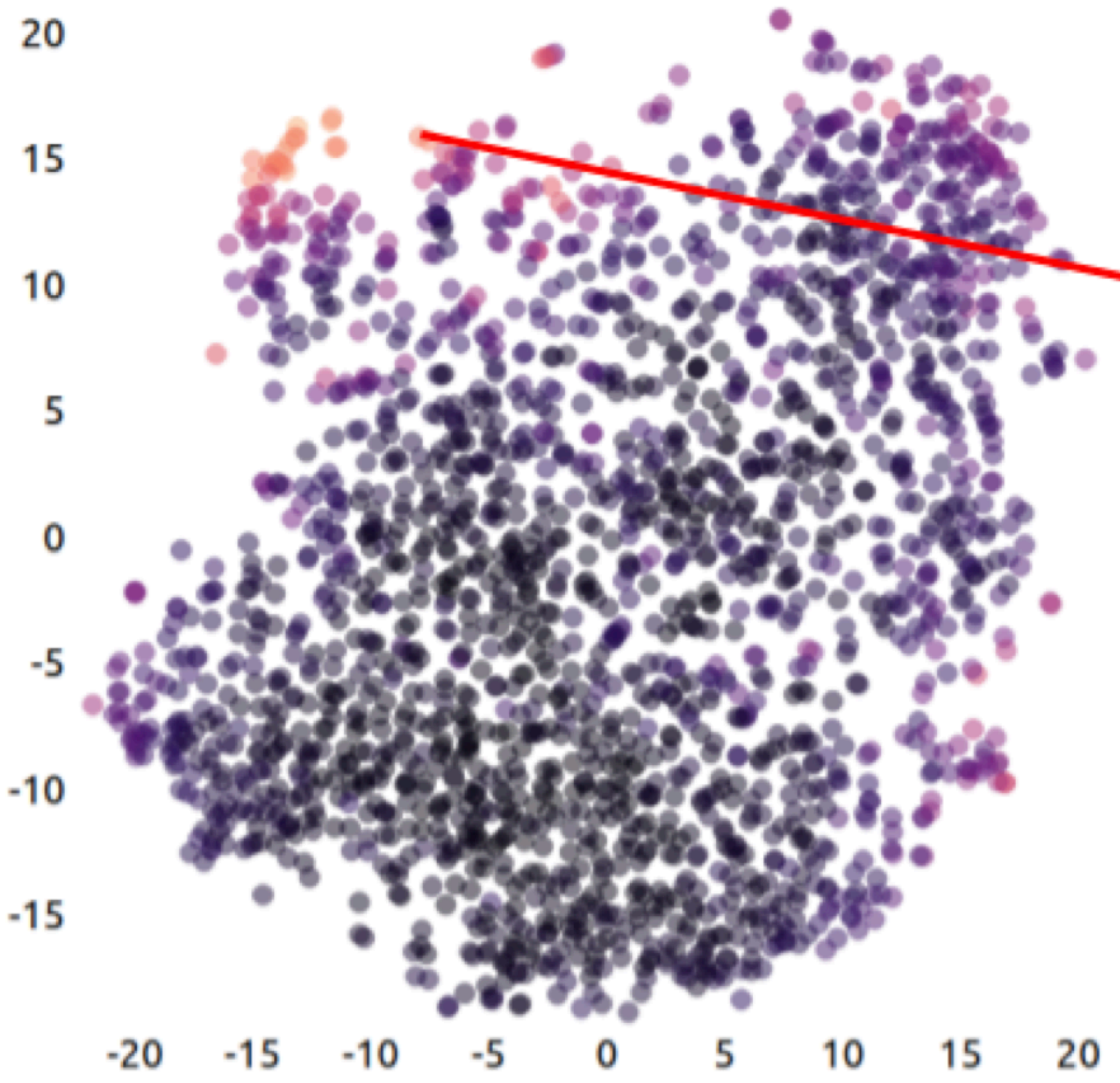
Class	Effective temperature[1][2]	Vega-relative chromaticity[3][4][a]	Chromaticity (D65)[5][6][3][b]	Main-sequence mass[1][7] (solar masses)	Main-sequence radius[1][7] (solar radii)	Main-sequence luminosity[1][7] (bolometric)	Hydrogen lines	Fraction of all main-sequence stars[8]
O	≥ 30,000 K	blue	blue	≥ 16 $M_\odot$	≥ 6.6 $R_\odot$	≥ 30,000 $L_\odot$	Weak	~0.00003%
B	10,000–30,000 K	blue white	deep blue white	2.1–16 $M_\odot$	1.8–6.6 $R_\odot$	25–30,000 $L_\odot$	Medium	0.13%
A	7,500–10,000 K	white	blue white	1.4–2.1 $M_\odot$	1.4–1.8 $R_\odot$	5–25 $L_\odot$	Strong	0.6%
F	6,000–7,500 K	yellow white	white	1.04–1.4 $M_\odot$	1.15–1.4 $R_\odot$	1.5–5 $L_\odot$	Medium	3%
G	5,200–6,000 K	yellow	yellowish white	0.8–1.04 $M_\odot$	0.96–1.15 $R_\odot$	0.6–1.5 $L_\odot$	Weak	7.6%
K	3,700–5,200 K	light orange	pale yellow orange	0.45–0.8 $M_\odot$	0.7–0.96 $R_\odot$	0.08–0.6 $L_\odot$	Very weak	12.1%
M	2,400–3,700 K	orange red	light orange red	0.08–0.45 $M_\odot$	≤ 0.7 $R_\odot$	≤ 0.08 $L_\odot$	Very weak	76.45%

- Afterwords, it was realised that different classes correspond to different temperatures

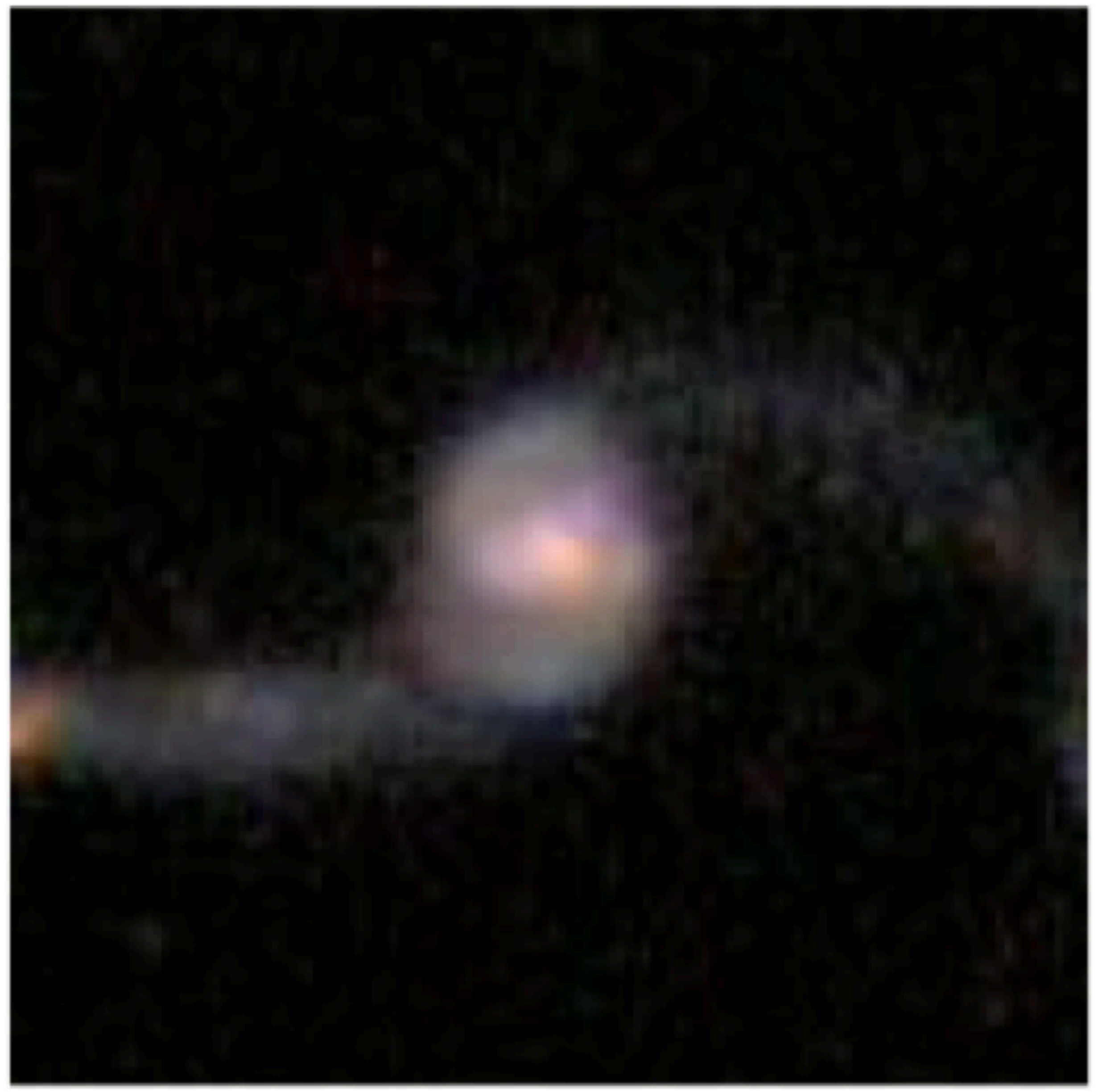


Learning from Anomalies

- Anomaly detection is one kind of data mining technique
  - One defines a metric of "typicality" to rank data samples
  - Based on this ranking, one can identify less typical events, tagging them as anomalies
  - By studying anomalies, one can make hypotheses on new physics mechanisms



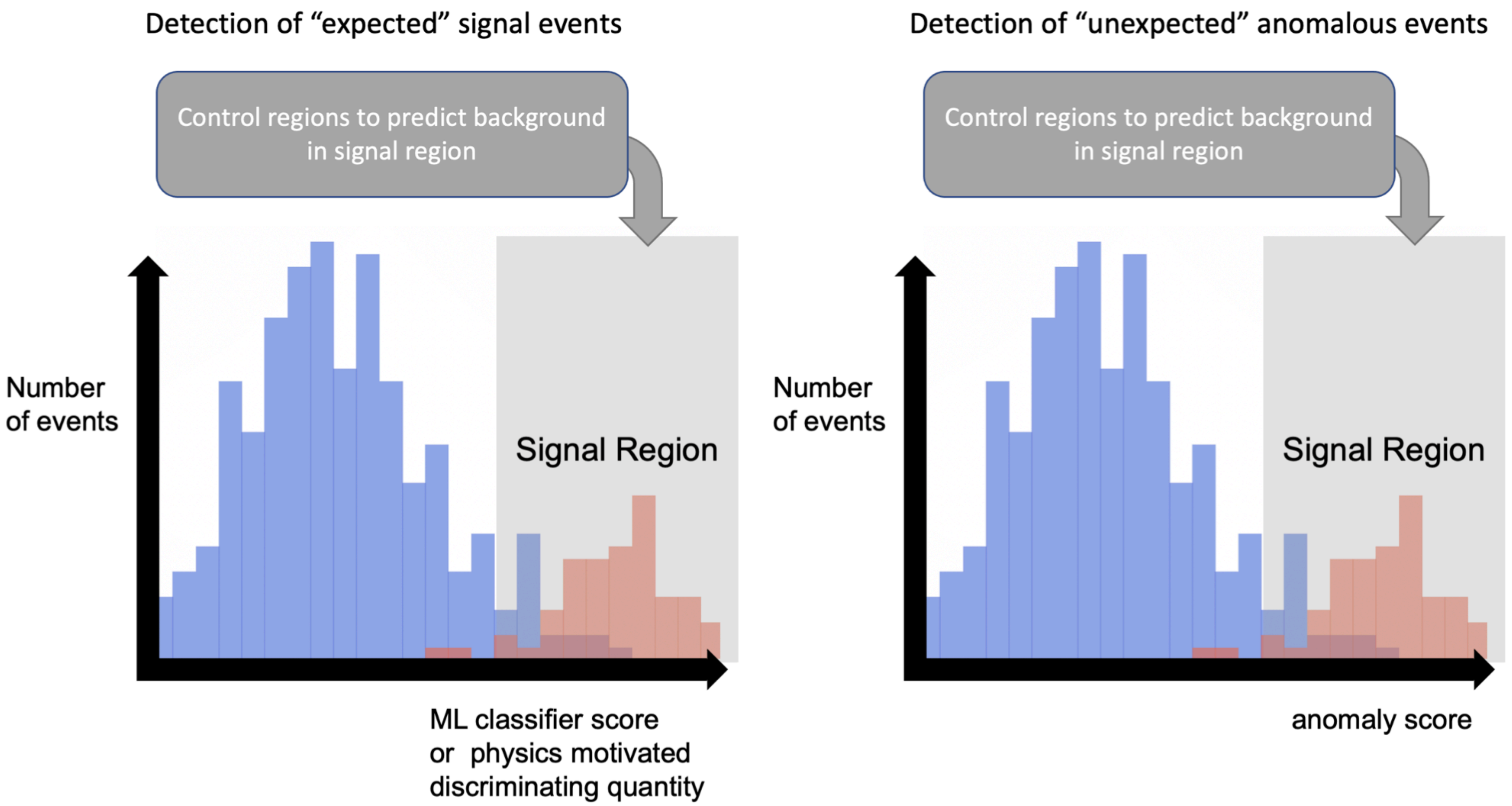
Object ID: 960415

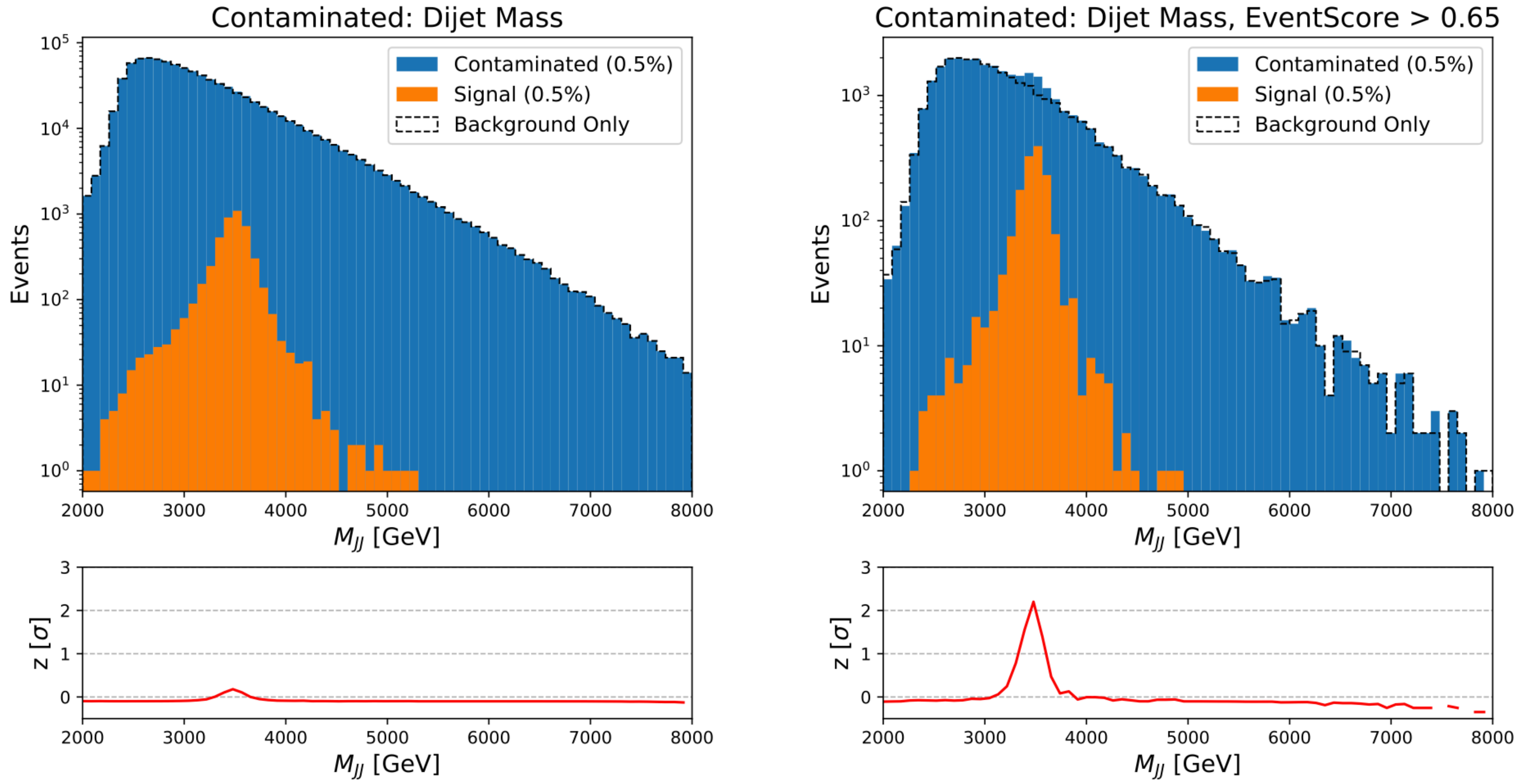


Anomaly Score: 4.470837

DeepLearning from Anomalies

- *Use semi/weakly/un supervised learning techniques to learn from data a metric*
- *Use that metric to replace physics motivate features (or supervised ML scores)*
- *Could be useful*
    - *Online, to select events that we should keep but we are not (human bias in defining what is interesting)*
    - *Offline, to enhance signals from unexpected signatures*





European Research Council

LHC Olympics

- *Many "boxes" with X->YZ topology*
- *A few given for strategy design*
- *A few kept "black" and opened after submissions collected*
- *Several methods designed, now being considered for real LHC analyses*

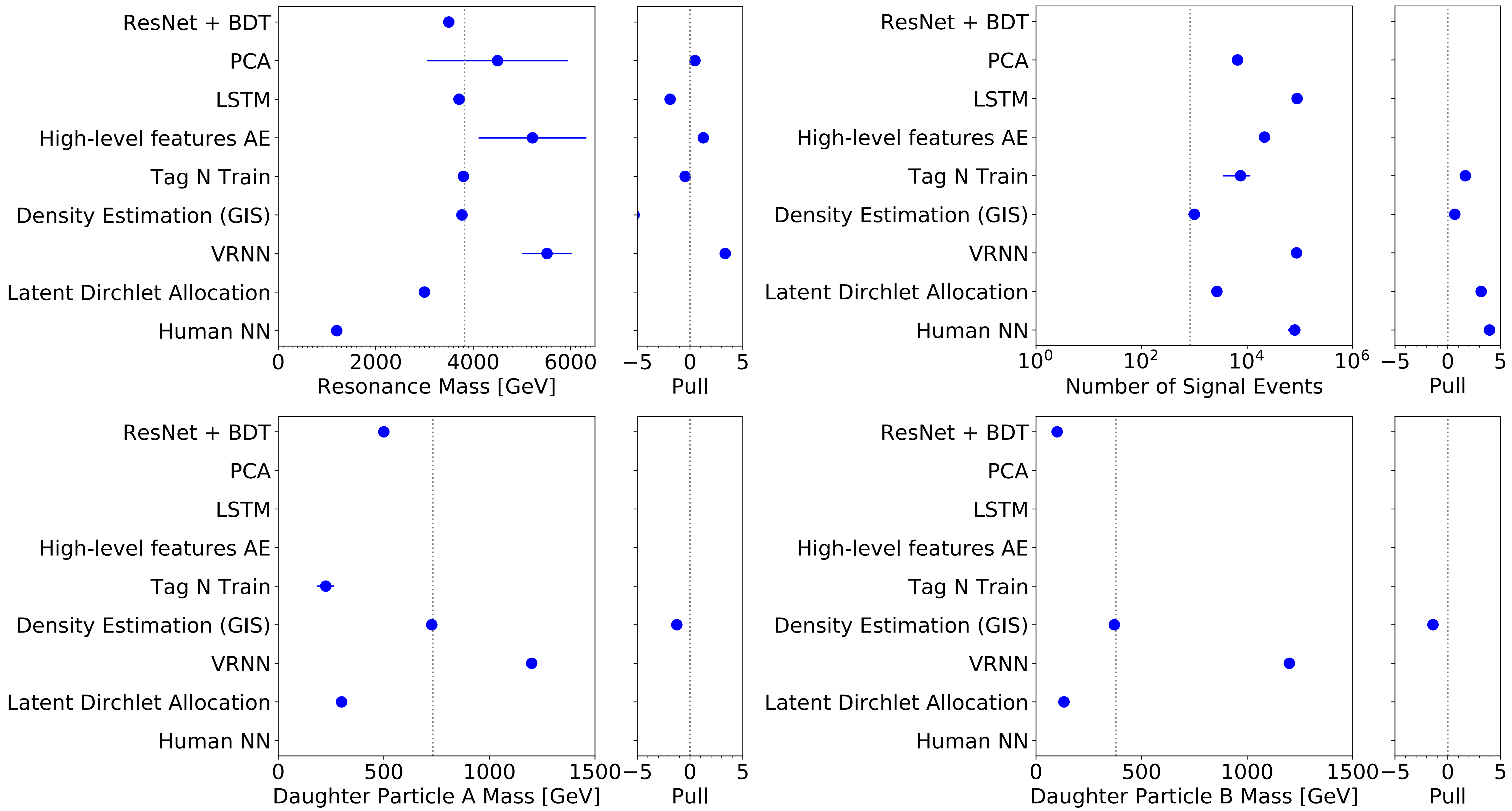


Figure 51. Results of unblinding the first black box. Shown are the predicted resonance mass (top left), the number of signal events (top right), the mass of the first daughter particle (bottom left), and the mass of the second daughter particle (bottom right). Horizontal bars indicate the uncertainty (only if provided by the submitting groups). In a smaller panel the pull (answer-true)/uncertainty is given. Descriptions of the tested models are provided in the text.

DarkMachine Challenge

- Similar scope, different setup
  - no specific event topology
  - generic event representation (list of reconstructed particles)
- Use unsupervised algorithms to define anomaly score
  - mainly autoencoders, with various architecture and training setup
  - High performance on benchmark examples, not always generalising to black boxes (optimization is an issue)

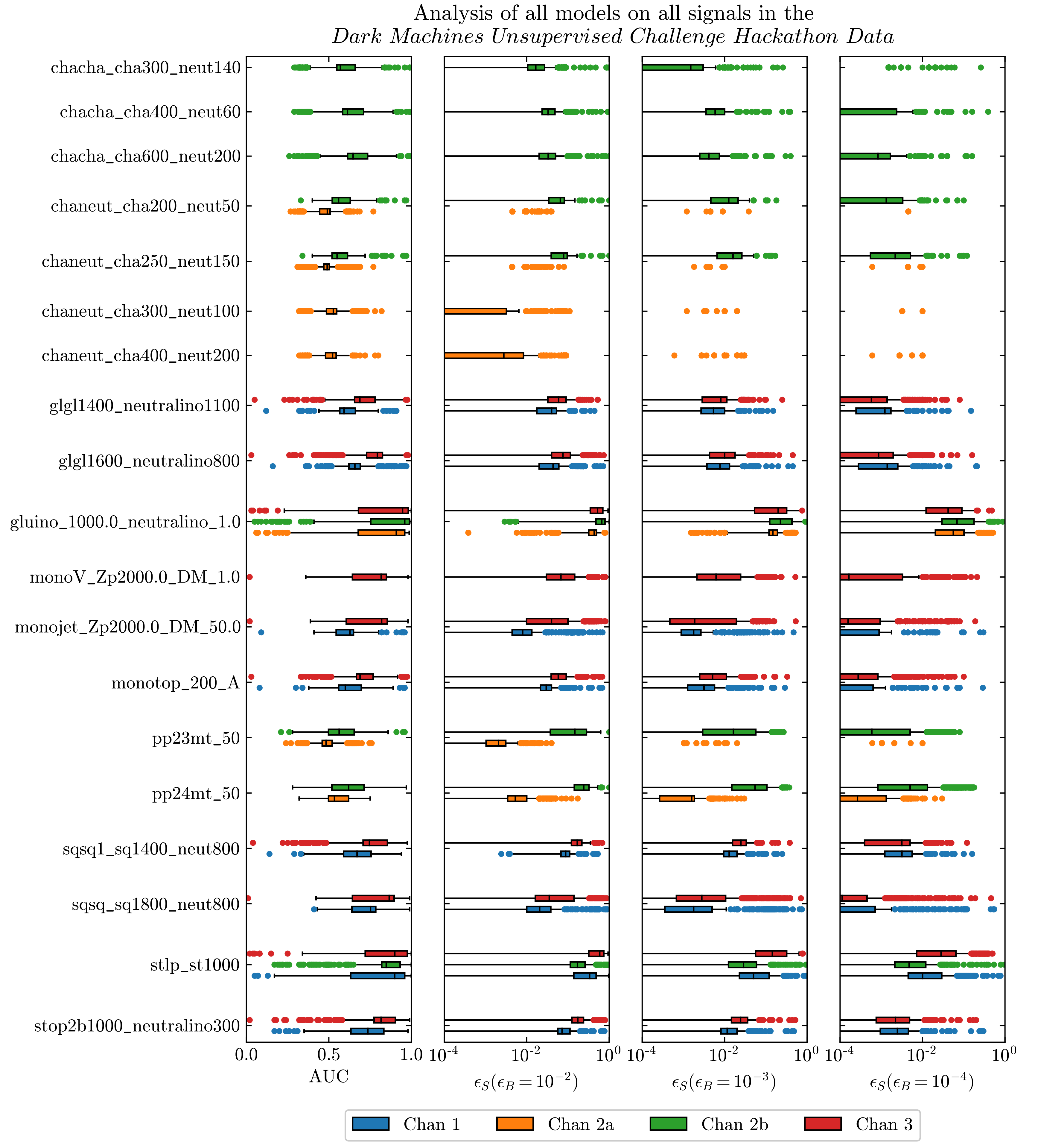
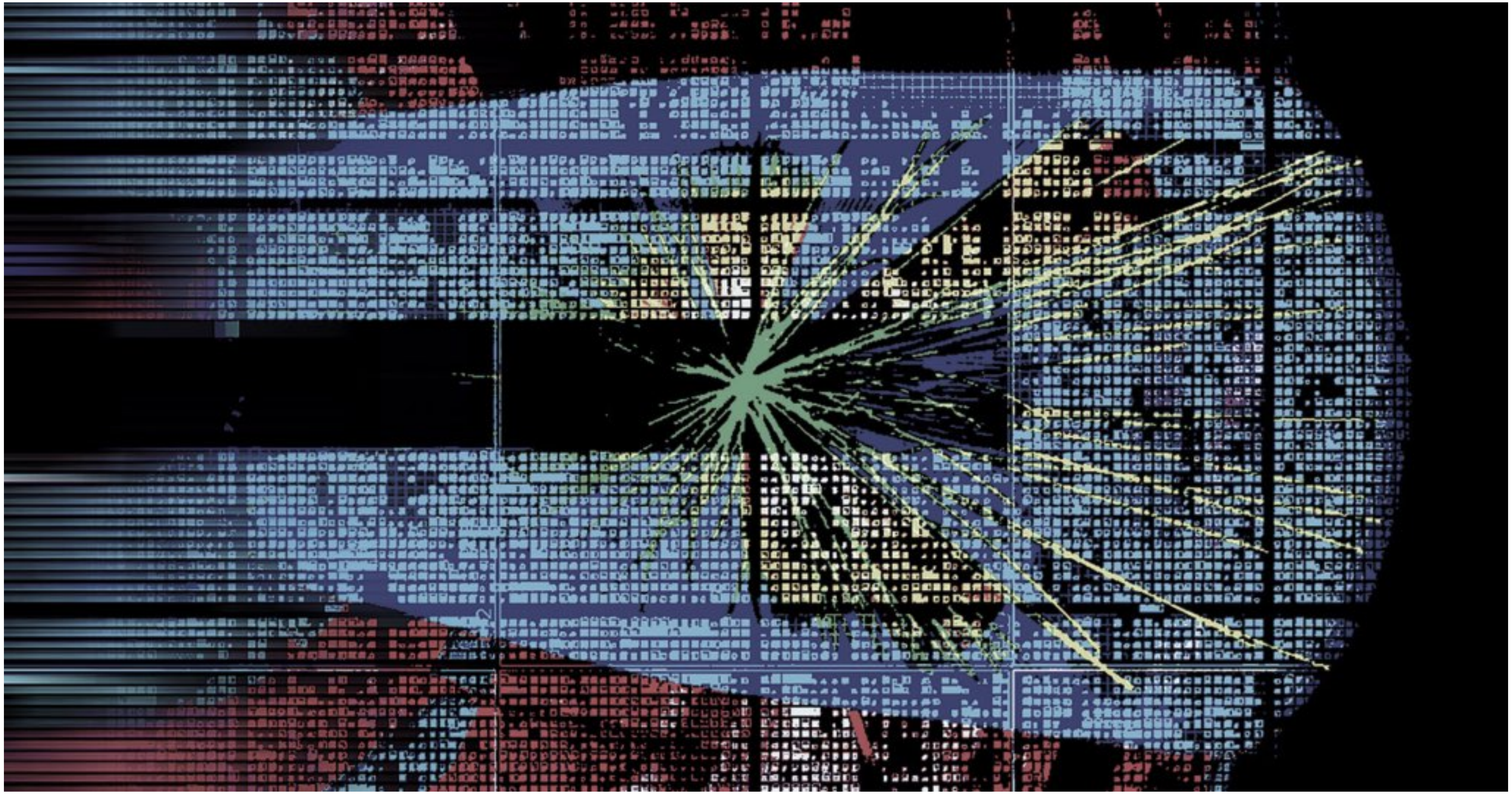


Figure 11: Box plots for each of the physics signals in the hackathon dataset. These summarize the span of results for the many anomaly detection models trained on background only samples. Channel 2a has the tightest pre-selection cuts, and therefore less data, which leads to the signals looking less anomalous.

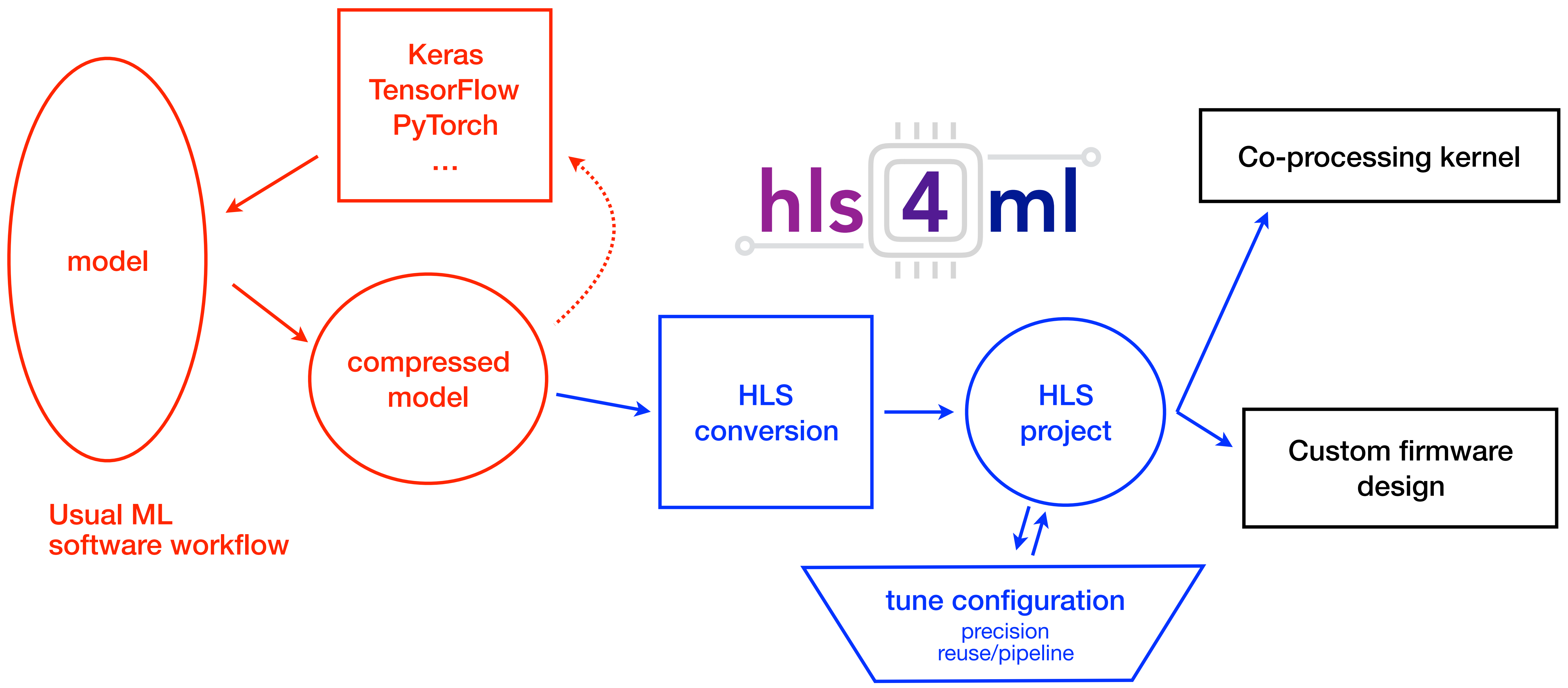


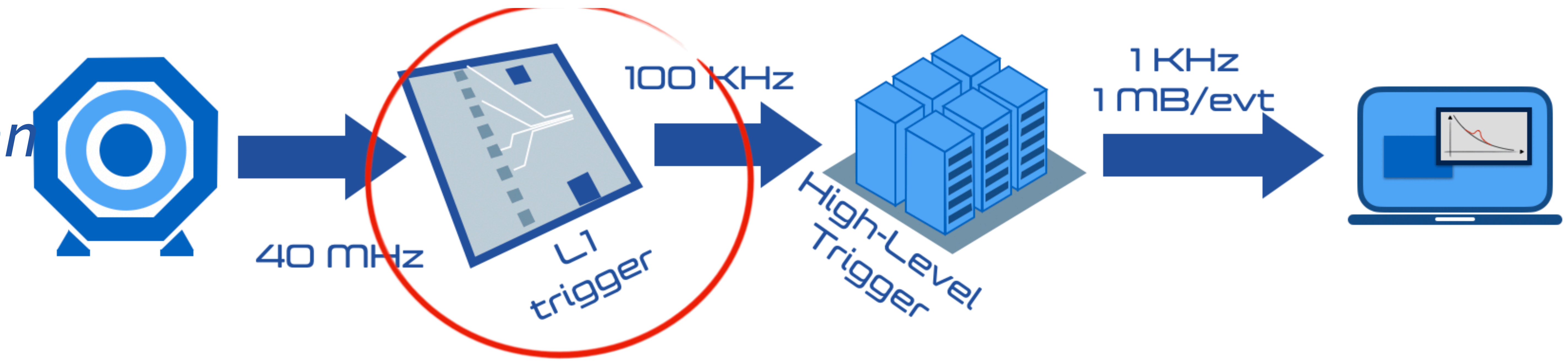
DL as an electronic circuit

European Research Council

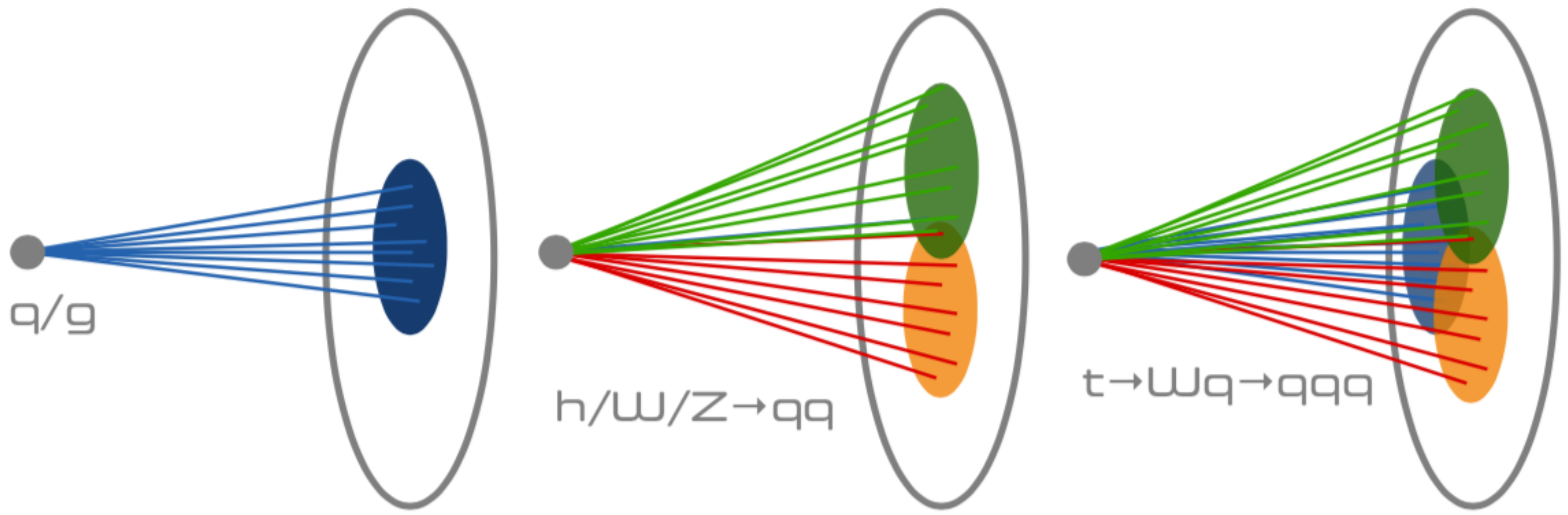
HLS4ML

- Tool to deploy NNs to FPGA
  - reads as input models trained on standard DeepLearning libraries
  - comes with implementation of common ingredients (layers, activation functions, etc)
  - Uses HLS libraries to deliver a firmware implementation of a given network on FPGA
  - Could also be used to design AI-specific ASICs for future experiments





Example: jet tagging



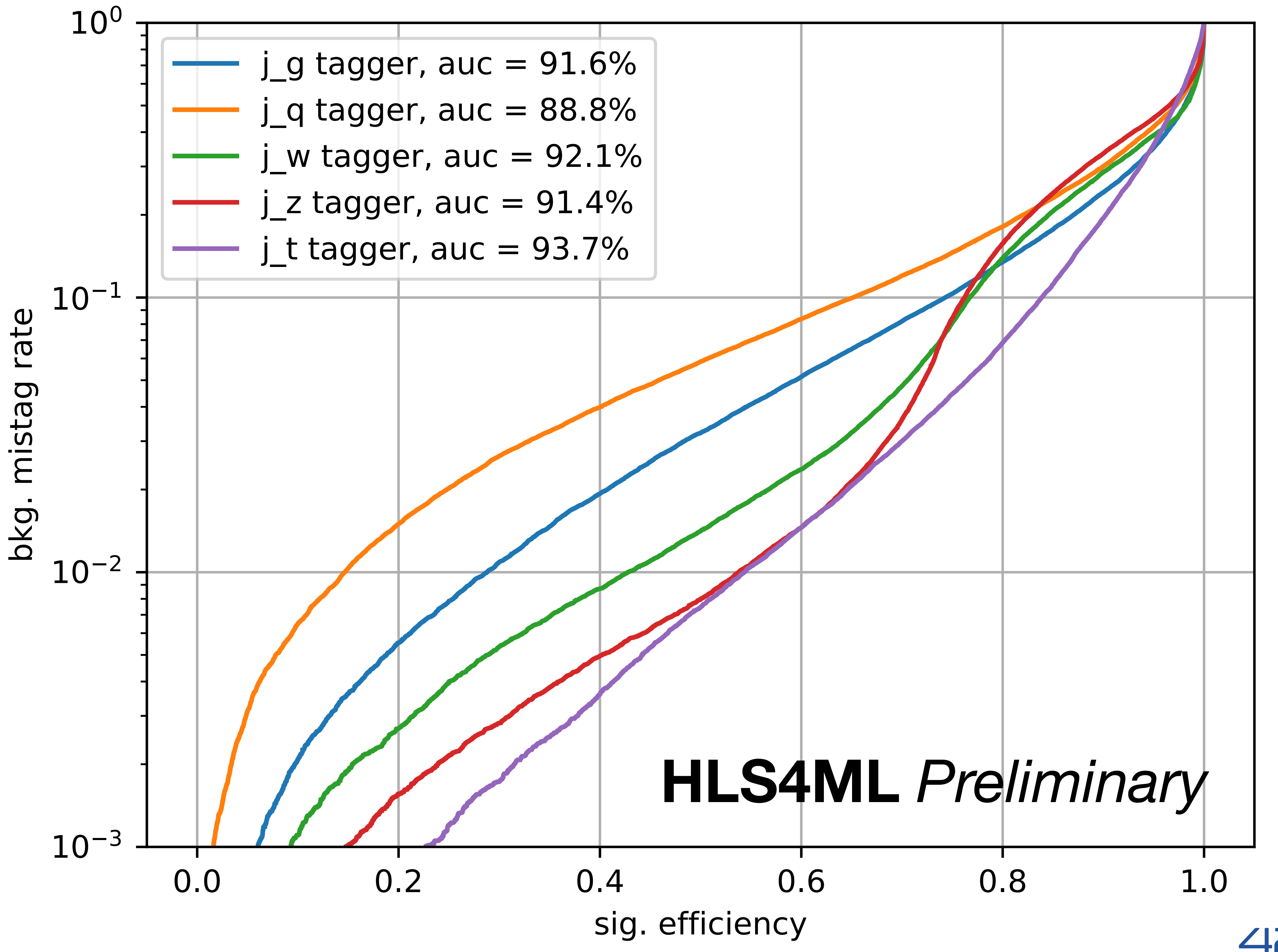
16 inputs

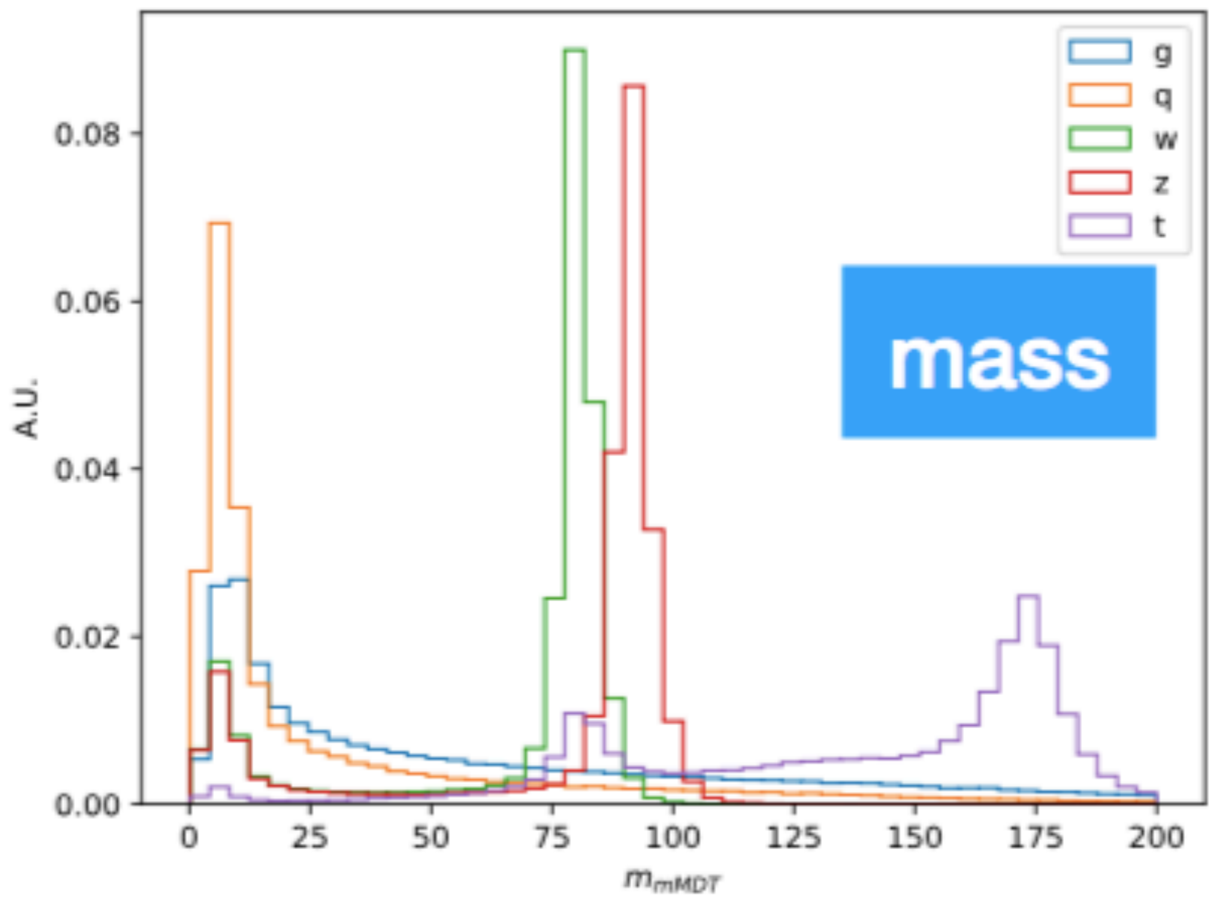
64 nodes
activation: ReLU

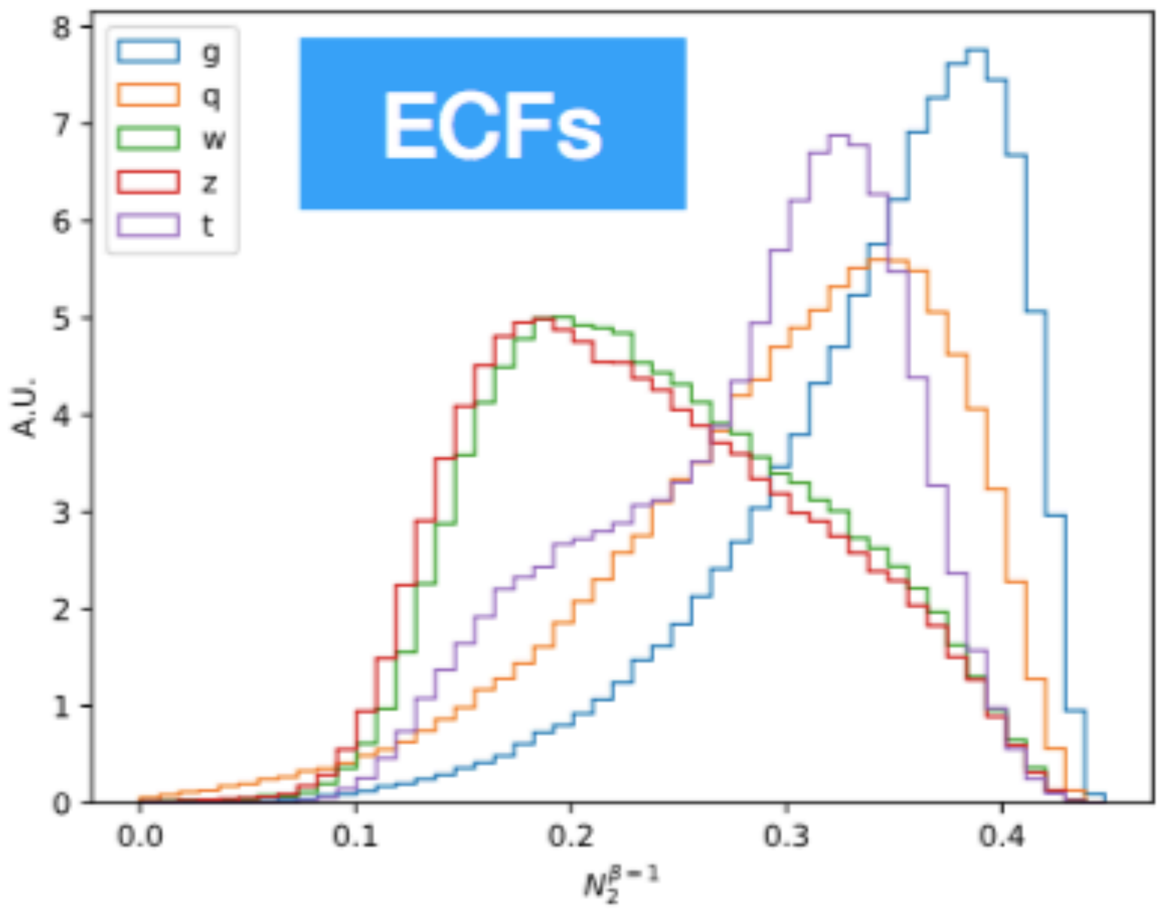
32 nodes
activation: ReLU

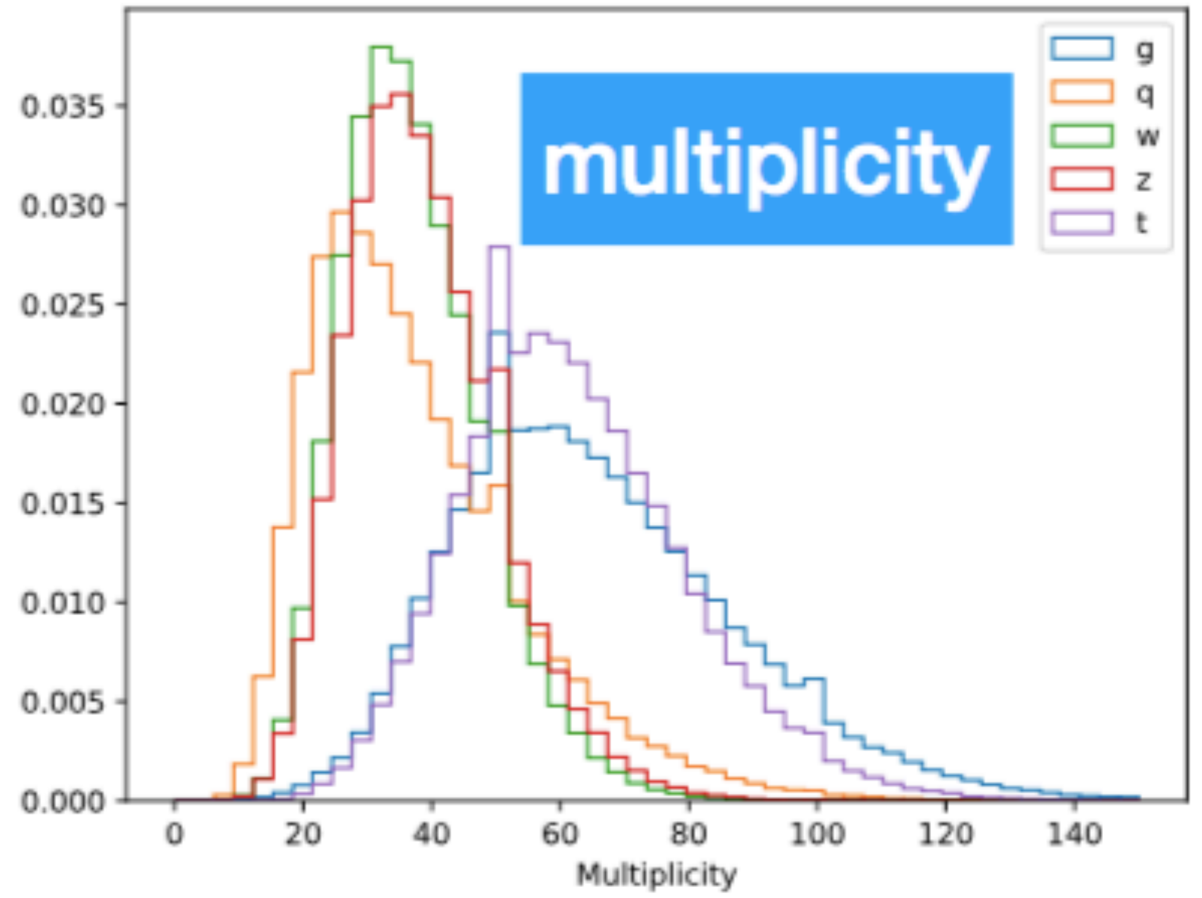
32 nodes
activation: ReLU

5 outputs
activation: SoftMax



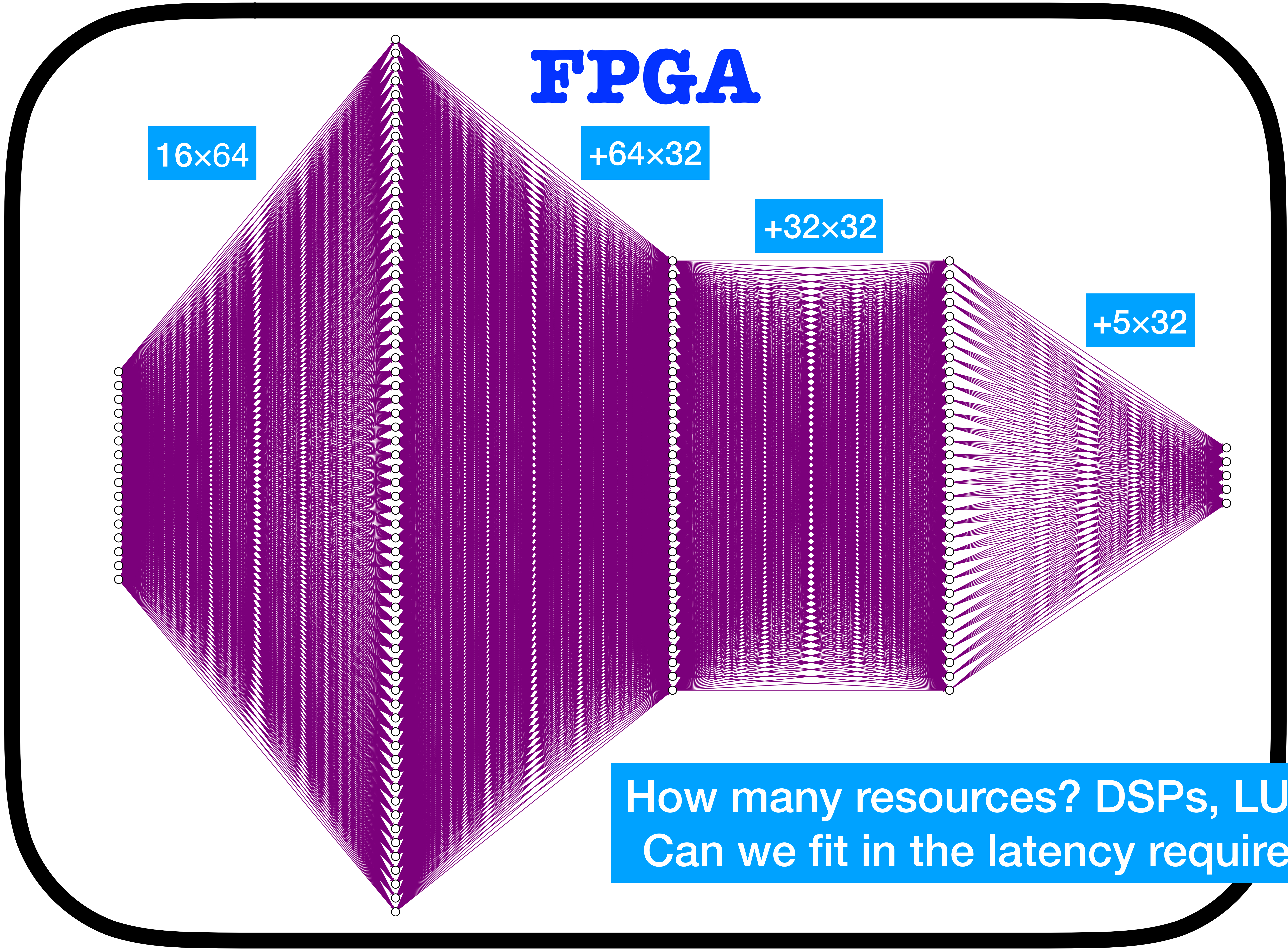






European Research Council

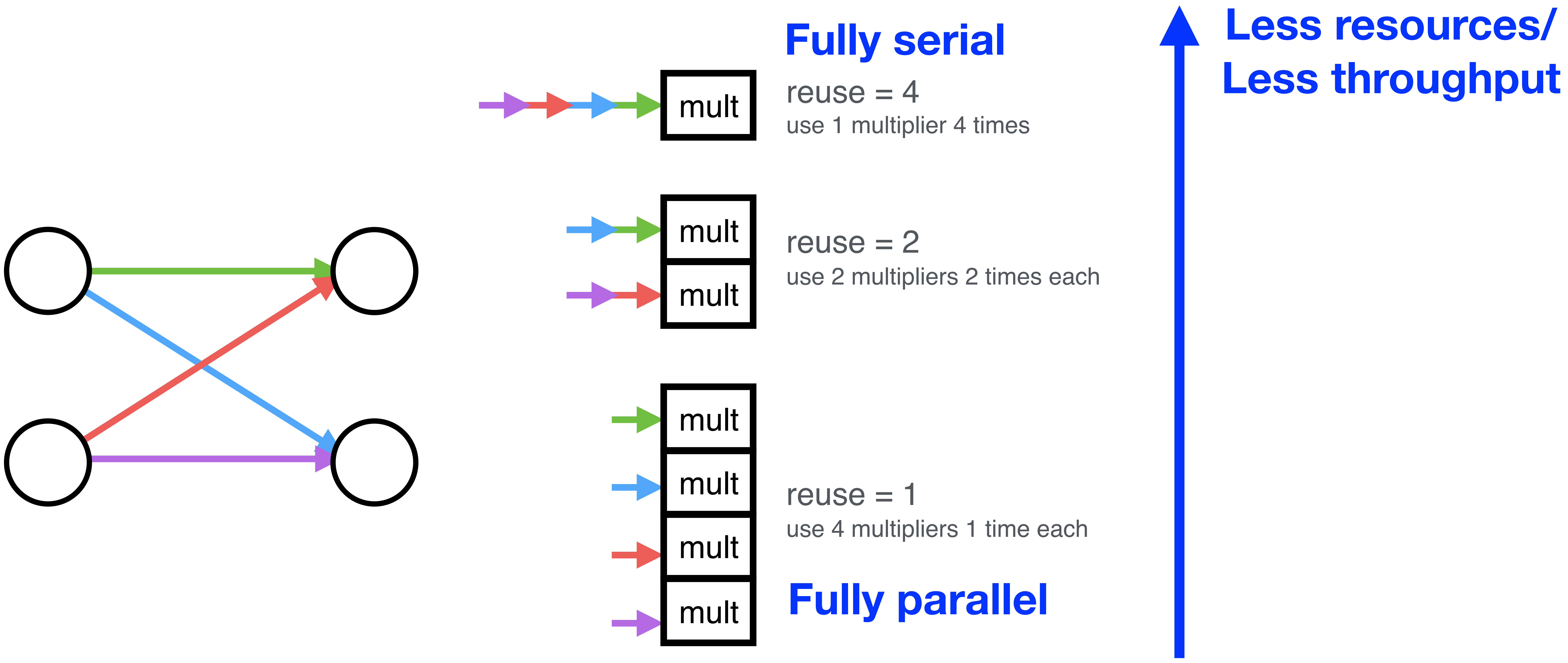
The full model



= 4,256 synapses / mult.

How many resources? DSPs, LUTs, FFs?
Can we fit in the latency requirements?

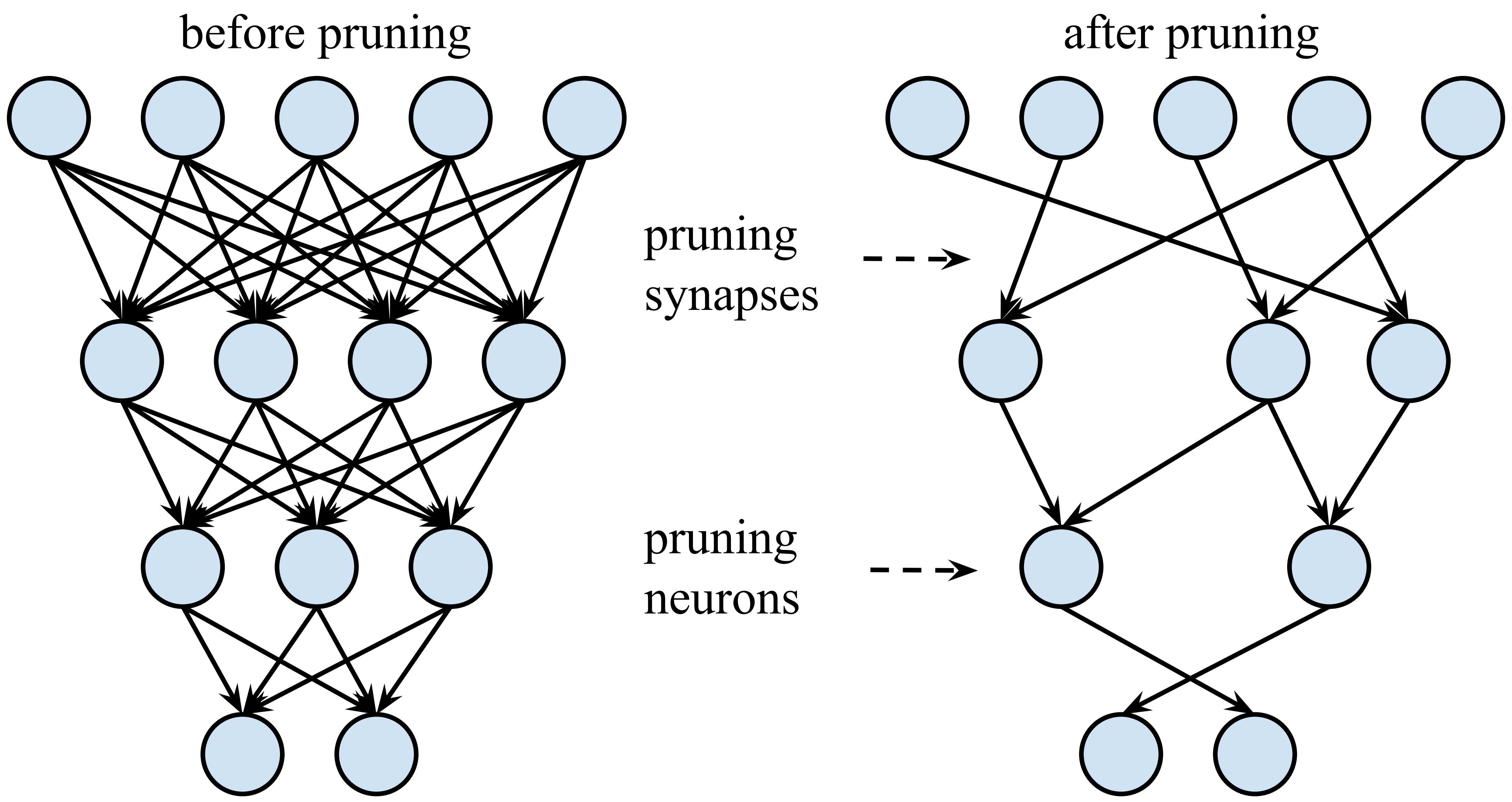
Model Compression: reuse



Reuse factor: how much to parallelize operations in a hidden layer

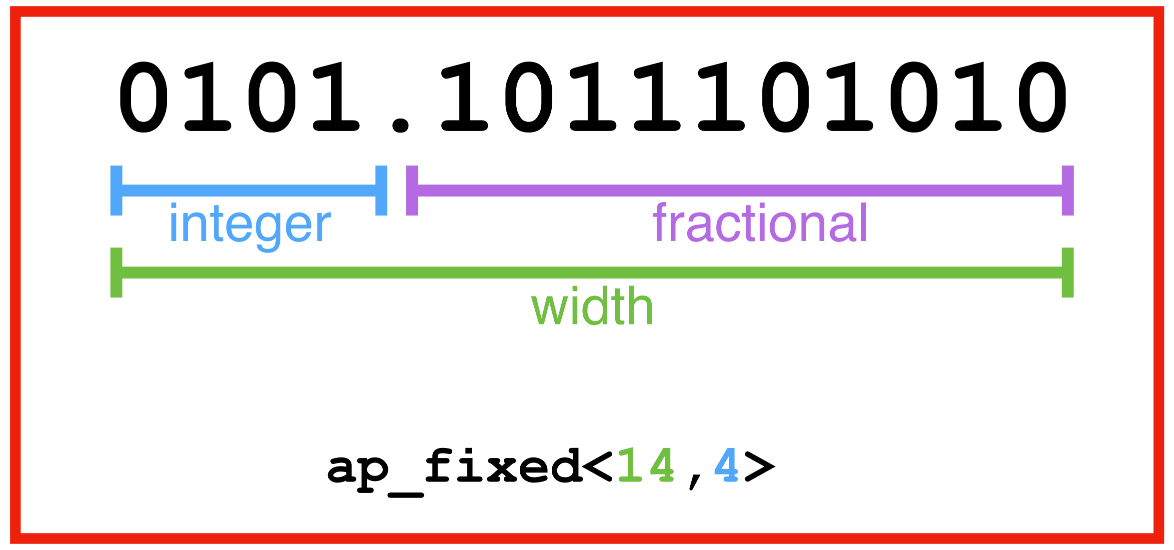
Model Compression: pruning

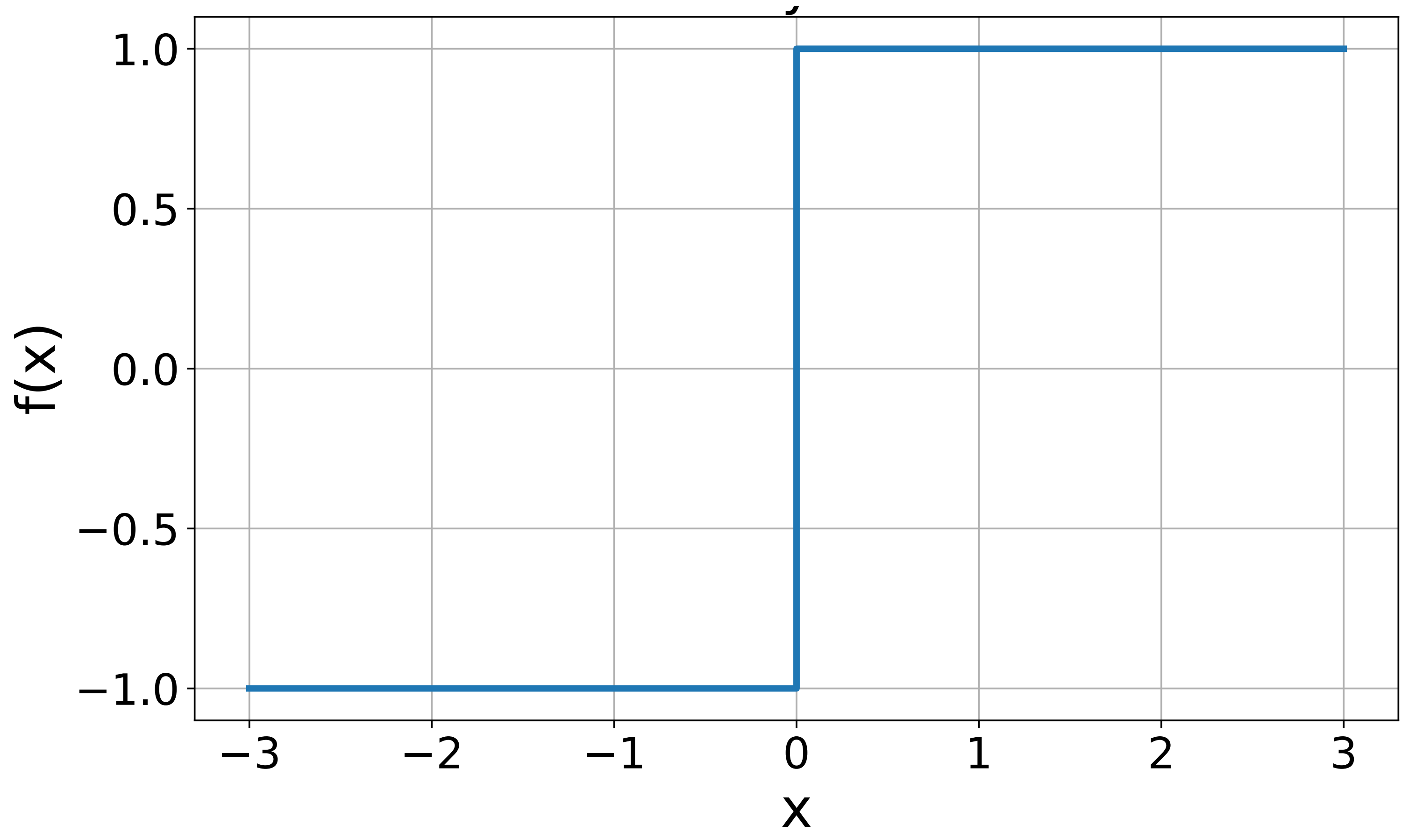
- Remove parameters that don't really contribute to performances
- For DNN, can remove up to 70% of a network with little impact on performance
- Resources saving exploited easily at HLS conversion
- More complicated with other architectures (requires dedicated pruning strategies)

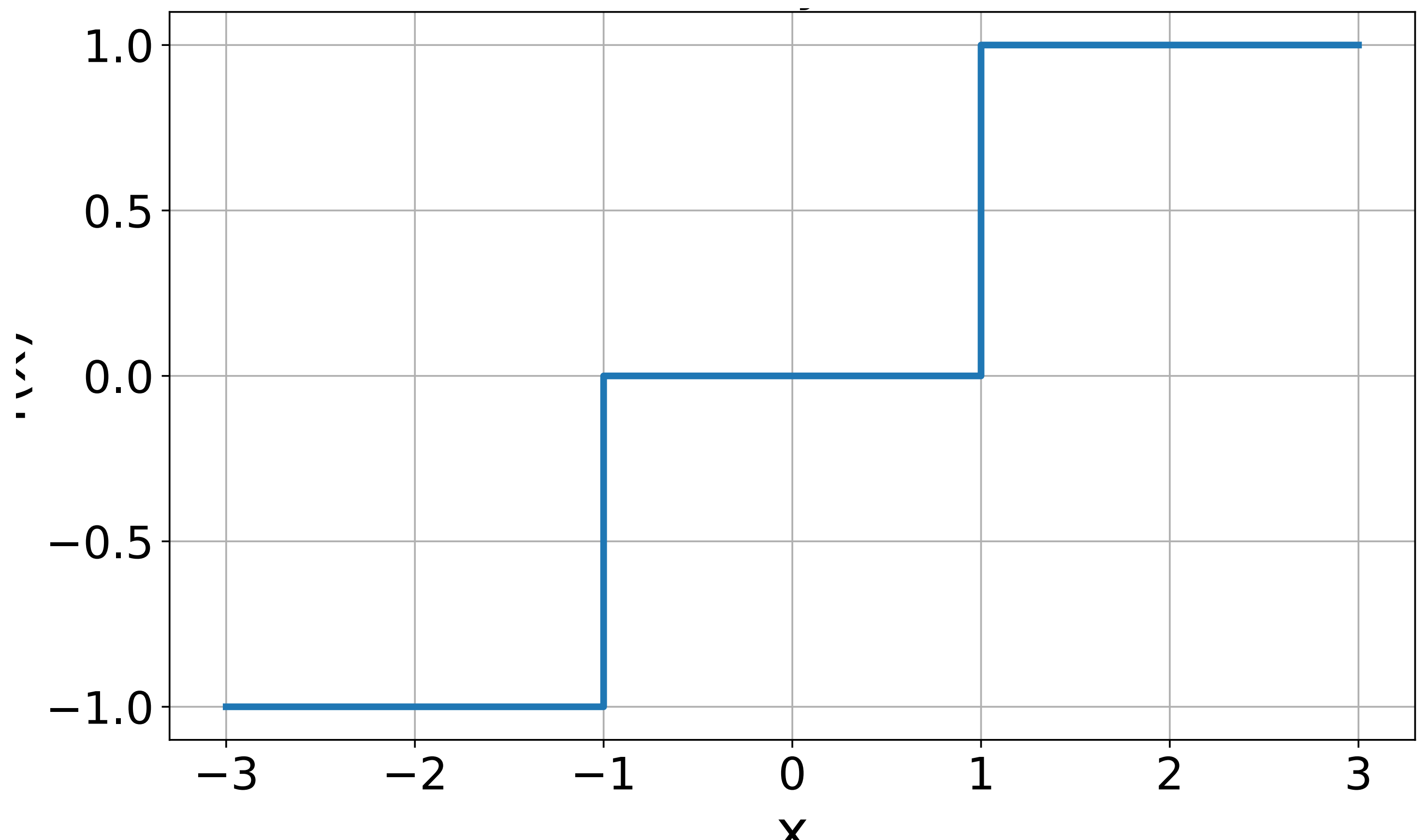


Model Compression: quantization

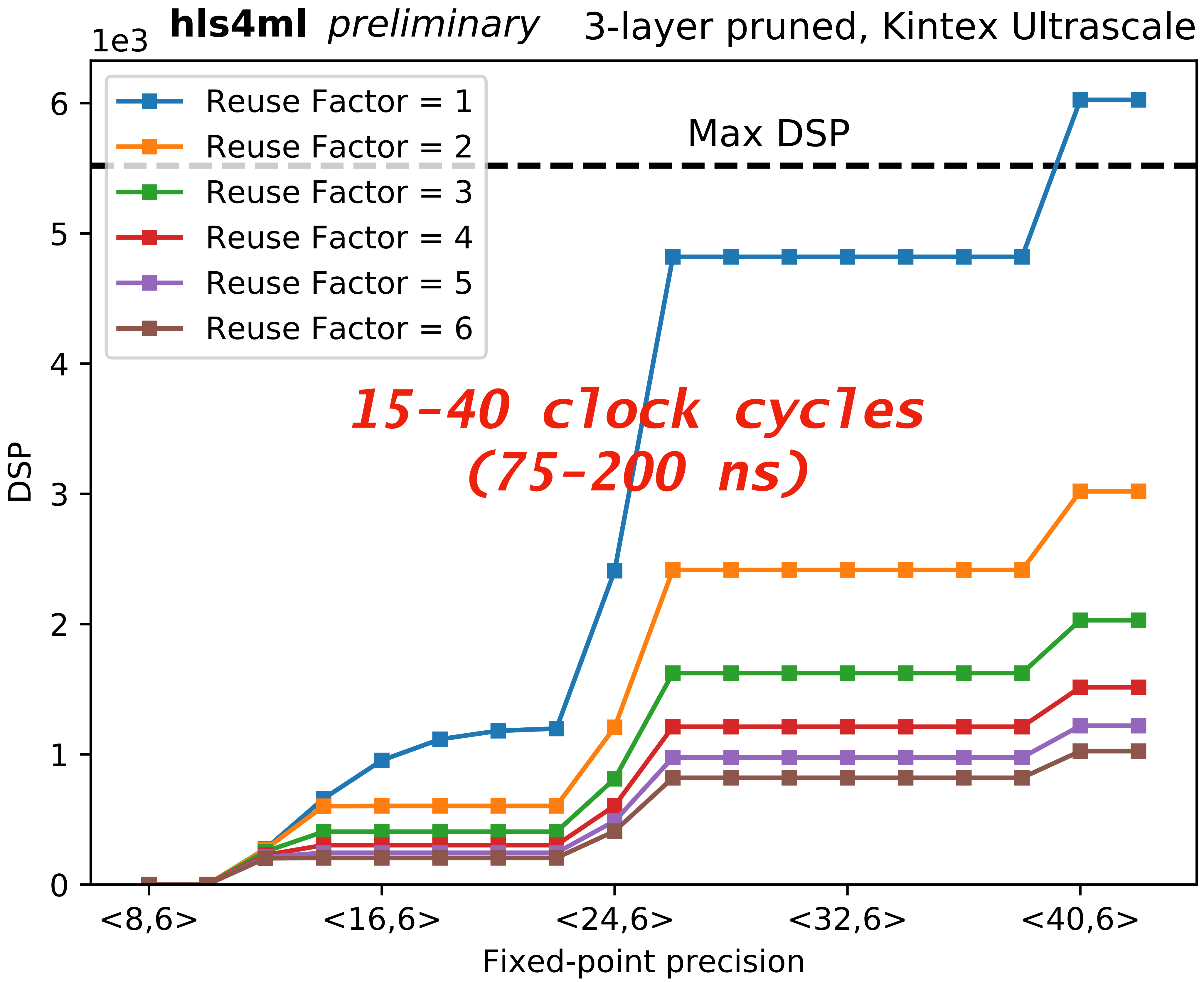
- ***Quantization:*** *reduce the number of bits used to represent numbers (i.e., reduce used memory)*
  - *models are usually trained at 64 or 32 bits*
  - *this is not necessarily needed in real life*
  - *one can go down to 16 bits w/o performance loss*
  - *one can do more quantising the model WHILE training*
  - *One could go as down as binary/ ternary precision with further computational advantage*

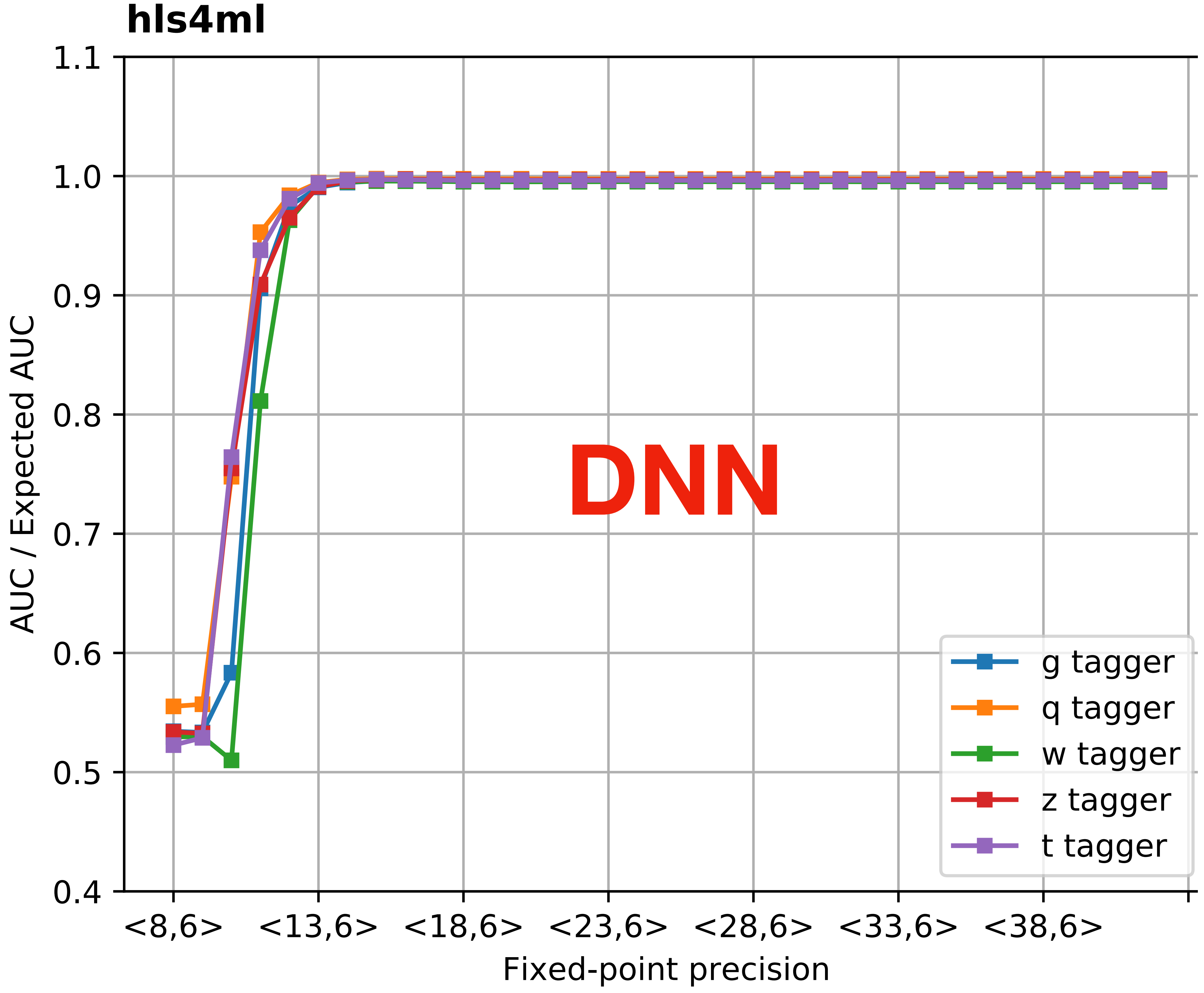






DNN Fast-inference

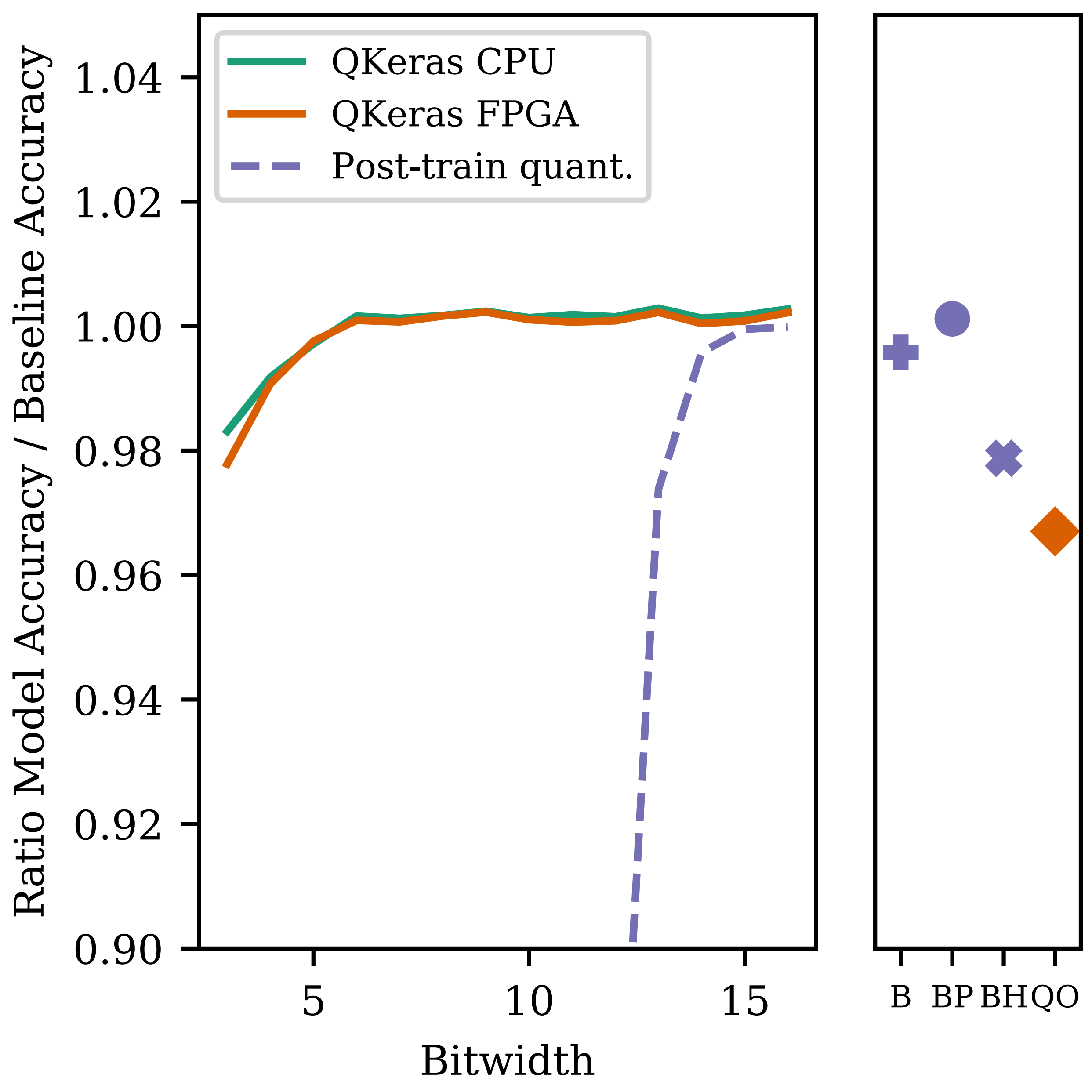


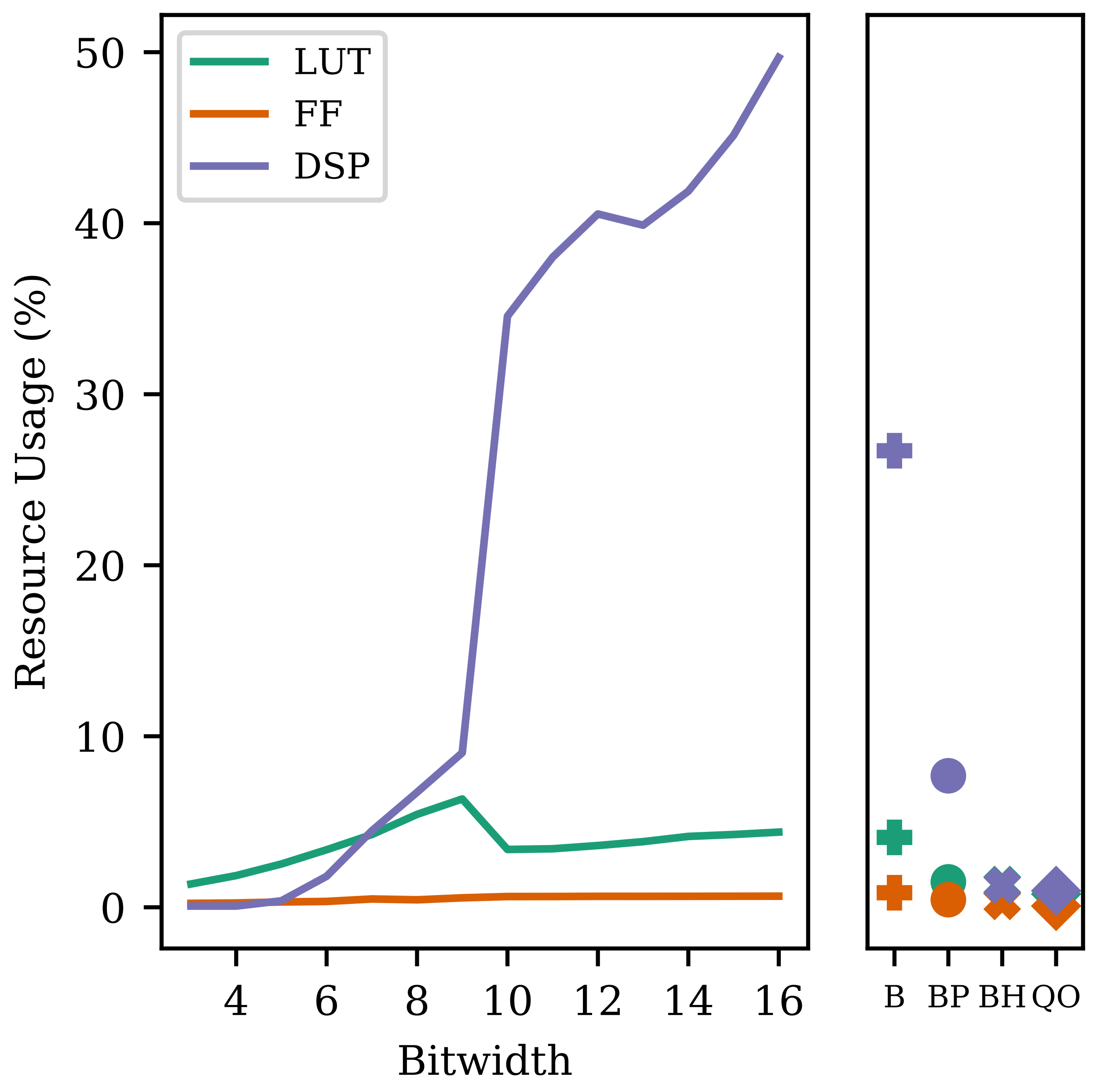


Accuracy stable vs quantization down to ~ 13 bits for DNNs
Latency well within the constraints

Quantization-aware Training

- *Post-training quantisation can affect accuracy*
  - *for a given bit allocation, the loss minimum at floating-point precision might not be the minimum anymore*
- *One could specify quantisation while look for the minimum*
  - *Maximize accuracy for minimal FPGA resources*
- *We teamed up with Google to exploit this strategy in a QKeras+hls4ml bundle*

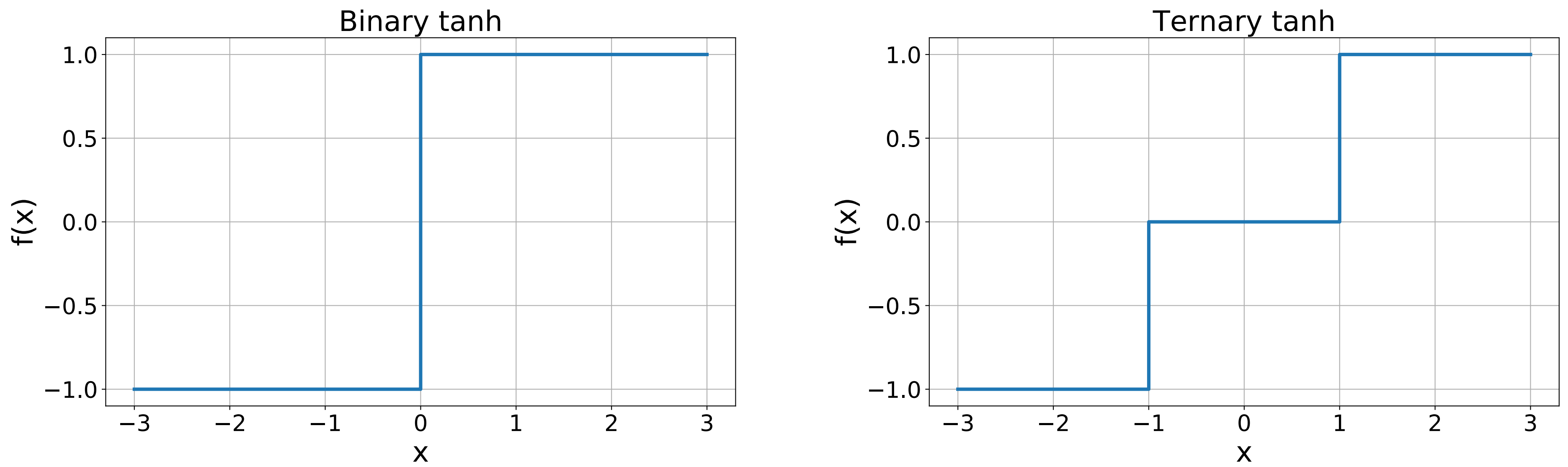




Model	Accuracy [%]	Latency [ns]	Latency [clock cycles]	DSP [%]	LUT [%]	FF [%]
Baseline	74.4	45	9	56.0 (1826)	5.2 (48321)	0.8 (20132)
Baseline pruned	74.8	70	14	7.7 (526)	1.5 (17577)	0.4 (10548)
Baseline heterogeneous	73.2	70	14	1.3 (88)	1.3 (15802)	0.3 (8108)
QKeras 6-bit	74.8	55	11	1.8 (124)	3.4 (39782)	0.3 (8128)
QKeras Optimized	72.3	55	11	1.0 (66)	0.8 (9149)	0.1 (1781)

https://arxiv.org/abs/2006.10159

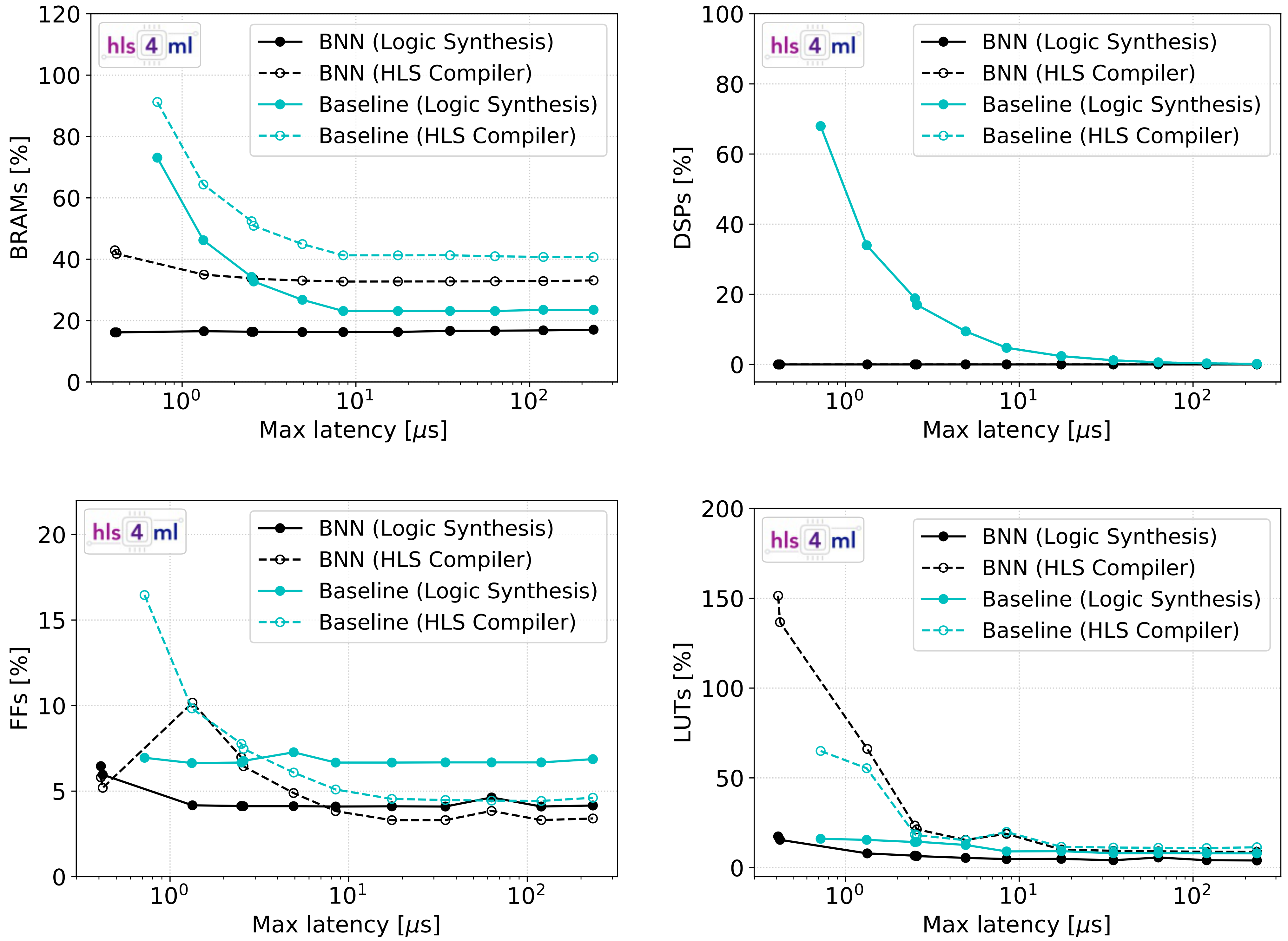
Extreme Quantization



A	B	$A \times B$
-1	-1	1
-1	1	-1
1	-1	-1
1	1	1

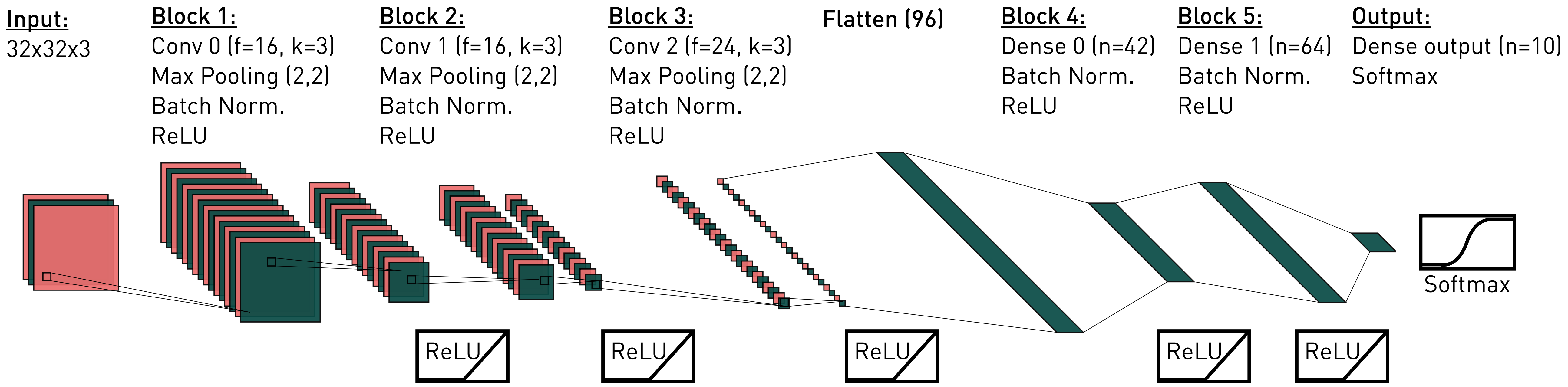
A	B	$\overline{A \oplus B}$
0	0	1
0	1	0
1	0	0
1	1	1

- *One can push quantisation to extremes*
  - *binary & ternary networks*
- *Multiplications can be replaced by bit manipulations, saving resources*
- *Can achieve low latencies at small accuracy cost and minimal resource consumption*



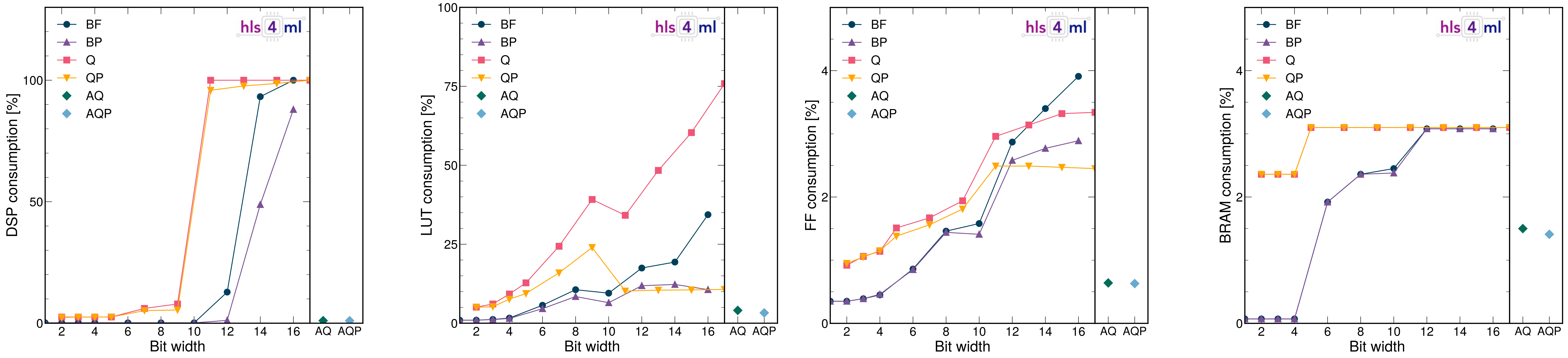
https://arxiv.org/abs/2003.06308

Fast CNN inference on FPGAs



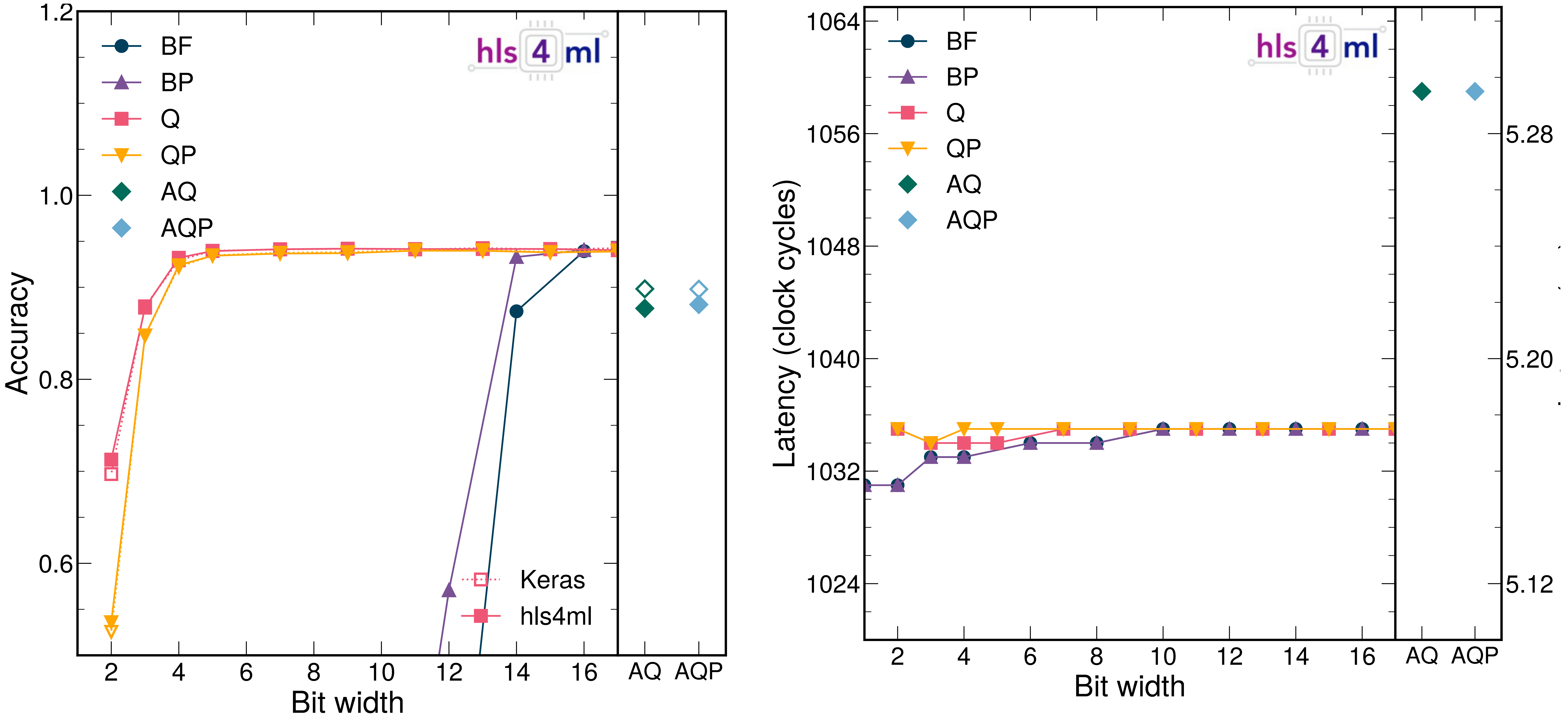
Layer name	Layer type	Input shape	Weights	MFLOPs	Energy [nJ]	Bit size
Conv 0	Conv2D	(32, 32, 3)	432	0.778	1,795	3,456
Conv 1	Conv2D	(15, 15, 16)	2,304	0.779	1,802	18,432
Conv 2	Conv2D	(6, 6, 16)	3,456	0.110	262	27,648
Dense 0	Dense	(96)	4,032	0.008	26	32,256
Dense 1	Dense	(42)	2,688	0.005	17	21,504
Output	Dense	(64)	65	0.001	4	5,200
Model total			**12,858**	**1.71**	**3,918**	**170,816**

% wrt to a Xilinx Virtex UltraScale+ VU9P



Drastic drop of DSP consumption and overall only ~ % of FPGA resources used

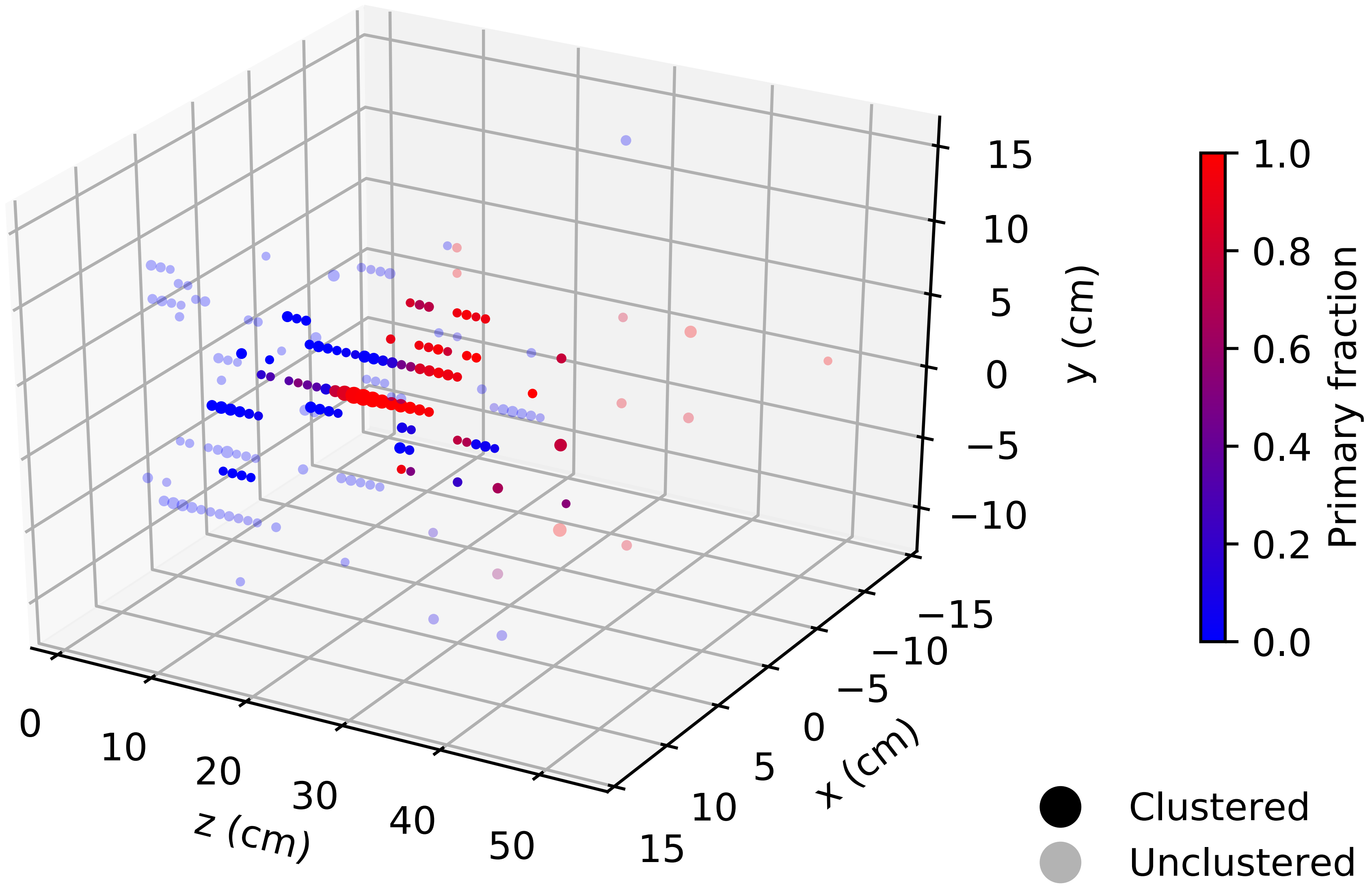
Fast CNN inference on FPGAs

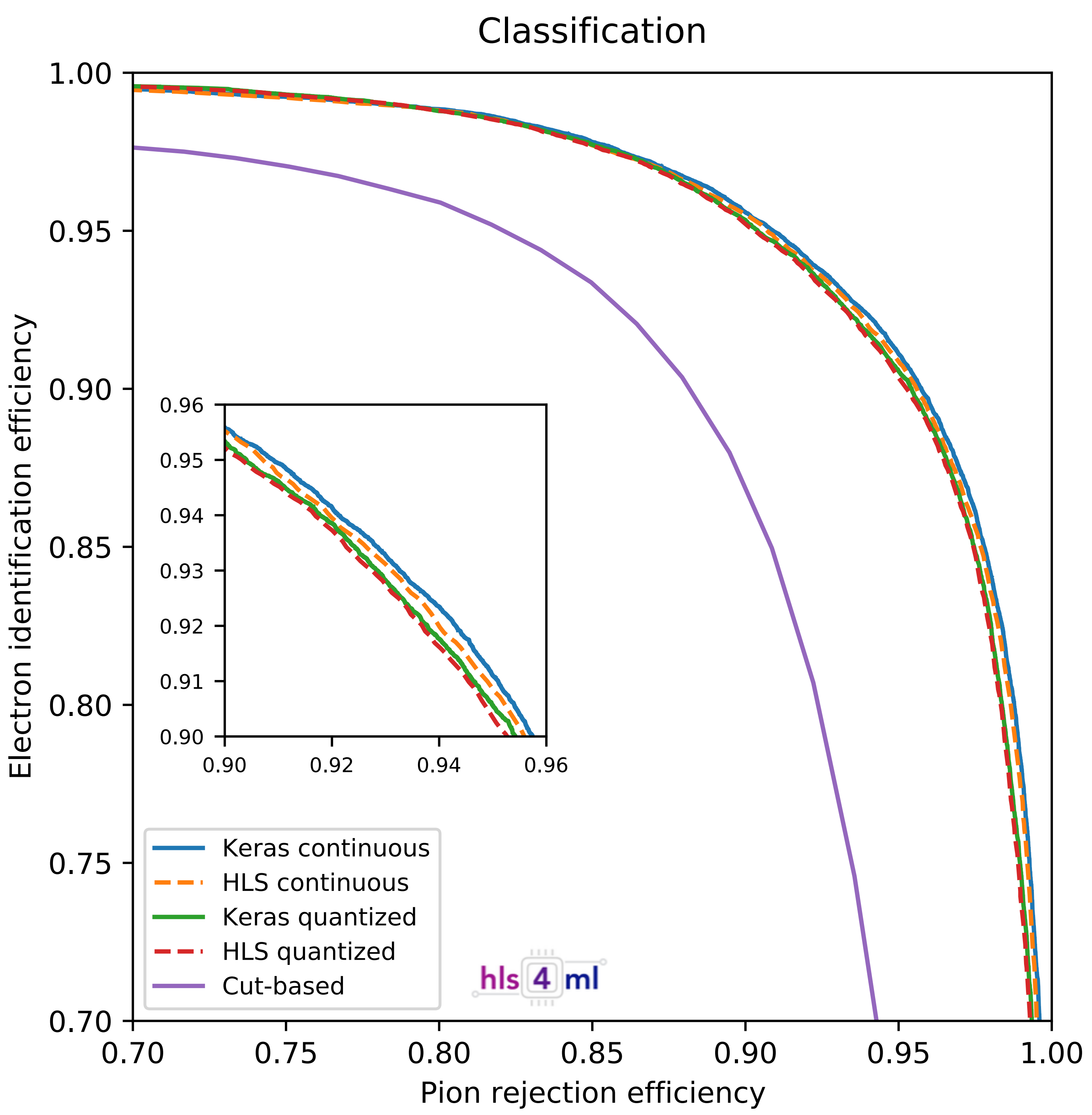


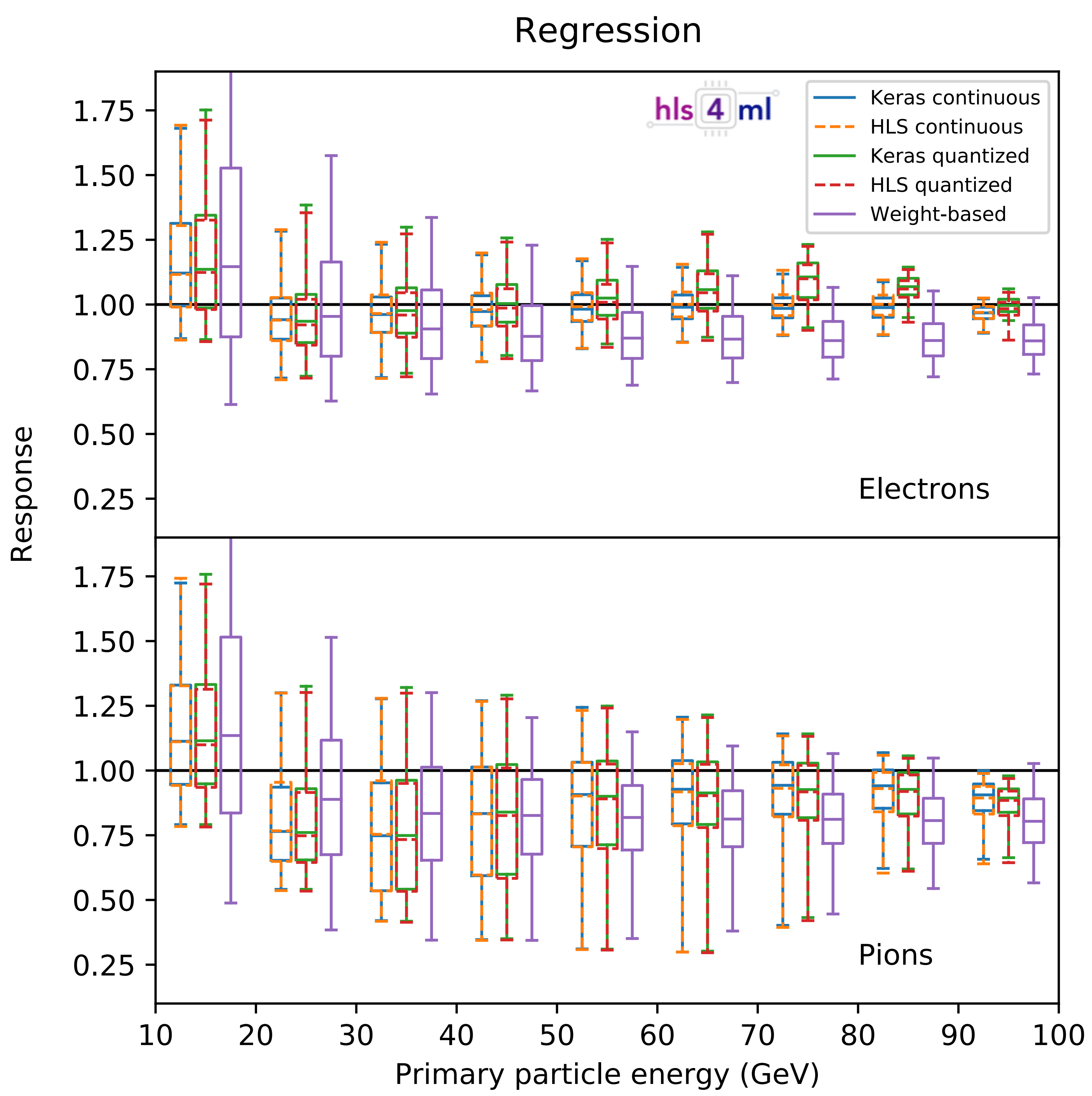
Execution time reduced to 5 µsec to basically no accuracy loss down to 6 bits

European Research Council

GraphNets on FPGAs





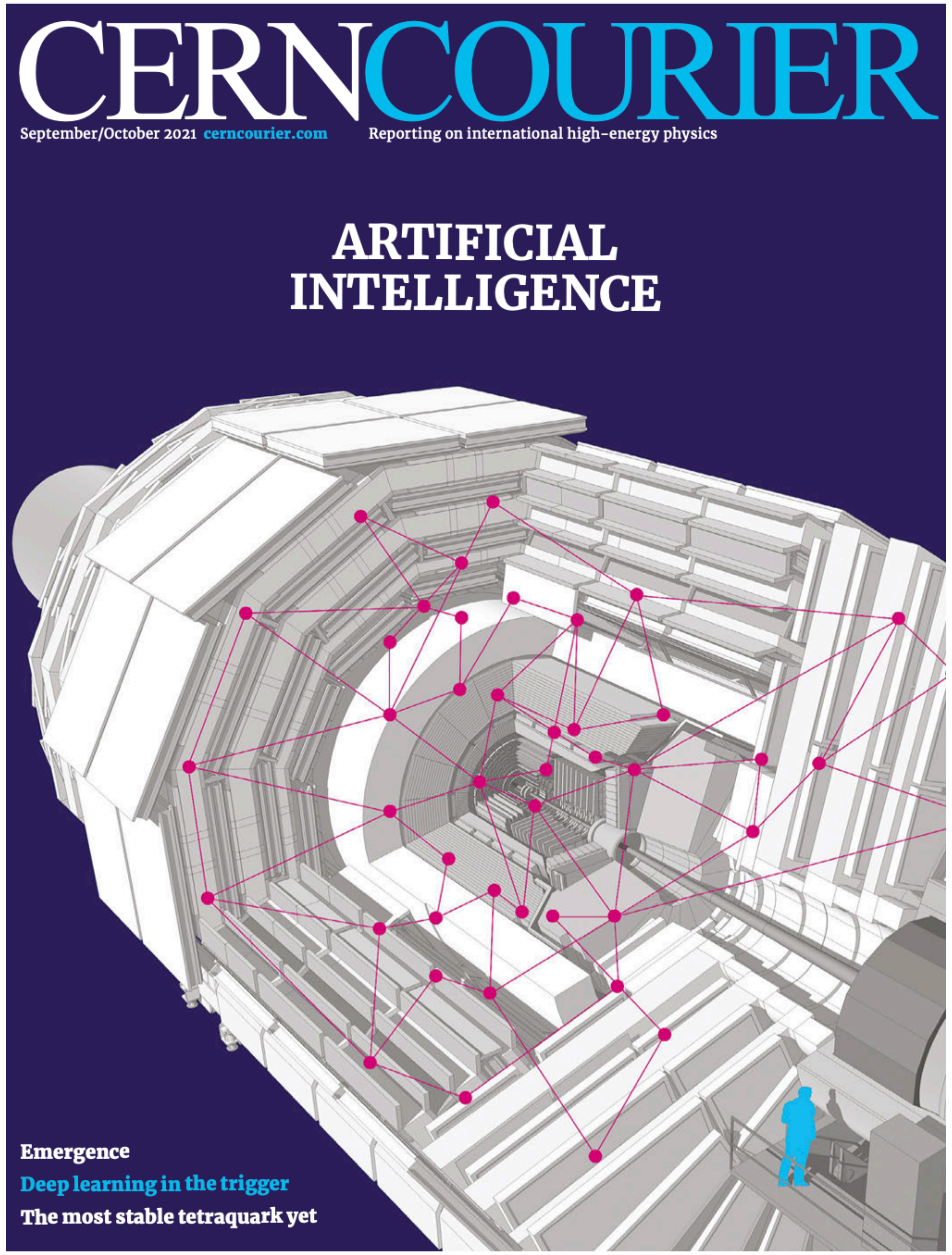


Model	$V_{\max}$	R_{reuse}	Latency (cycles)	Interval (cycles)	DSP (10^3)	LUT (10^3)	FF (10^3)	BRAM (Mb)	ROC AUC	Response RMS
Continuous	128	32	155	55	3.1 [56%]	57 [9%]	39 [2.9%]	1.8 [2.3%]	0.98	0.23
Quantized	128	32	148	50	1.6 [29%]	70 [11%]	41 [3.1%]	1.9 [2.4%]	0.98	0.24
Quantized	64	16	99	34	1.6 [29%]	63 [9%]	38 [2.9%]	1.8 [2.3%]	0.96	0.24
Quantized	32	8	75	26	1.4 [25%]	52 [8%]	33 [2.5%]	1.8 [2.3%]	0.86	0.37
Quantized	16	4	63	22	1.5 [27%]	57 [9%]	37 [2.8%]	1.8 [2.3%]	0.64	0.36

https://arxiv.org/abs/2008.03601

Conclusions

- ML has a long tradition in HEP, dating back to the end of the 80s
- ML has been functional to discoveries (e.g., Higgs but not only)
- ML popularity is increasing with Deep Learning opening new directions
  - Latest issue of CERN Courier dedicated to AI@CERN
- A community eager to learn & do more, for which CERN joining ELLIS would be a great opportunity



mPP

European Research Council