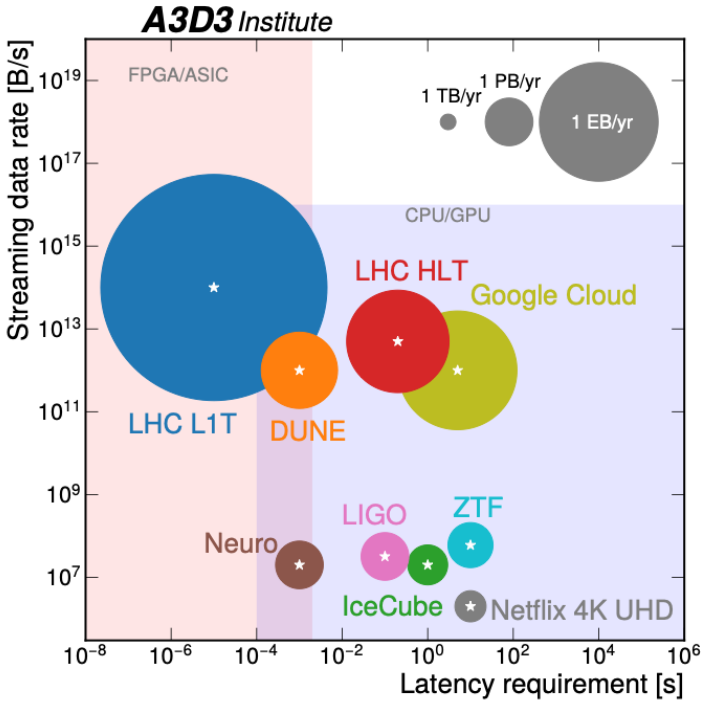
Triggering Tb/s of data: CMS perspective

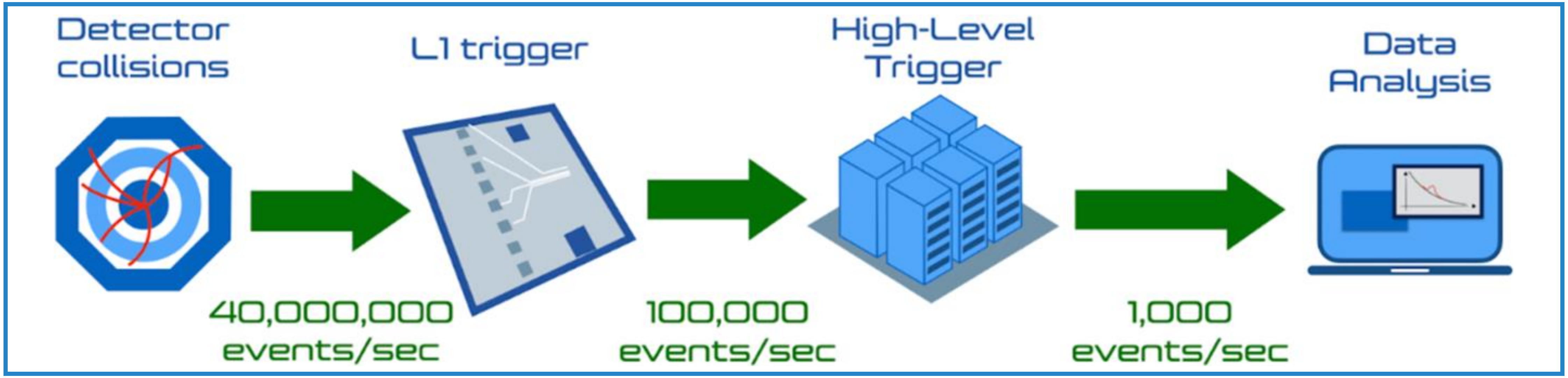
Santiago Folgueras on behalf of the CMS Collaboration

Funded by the European Union
European Research Council
Established by the European Commission
INTREPID
Universidad de Oviedo

Introduction: how a trigger system works?

- Data explosion and AI applications: our world needs **higher throughput** and **real-time computing** capabilities
- LHC provides ideal benchmark to explore real-time data processing technologies
- Only a handful of the collisions contain **interesting physics**
- **Trigger system** decides, in **real time** if a collision is saved or lost forever

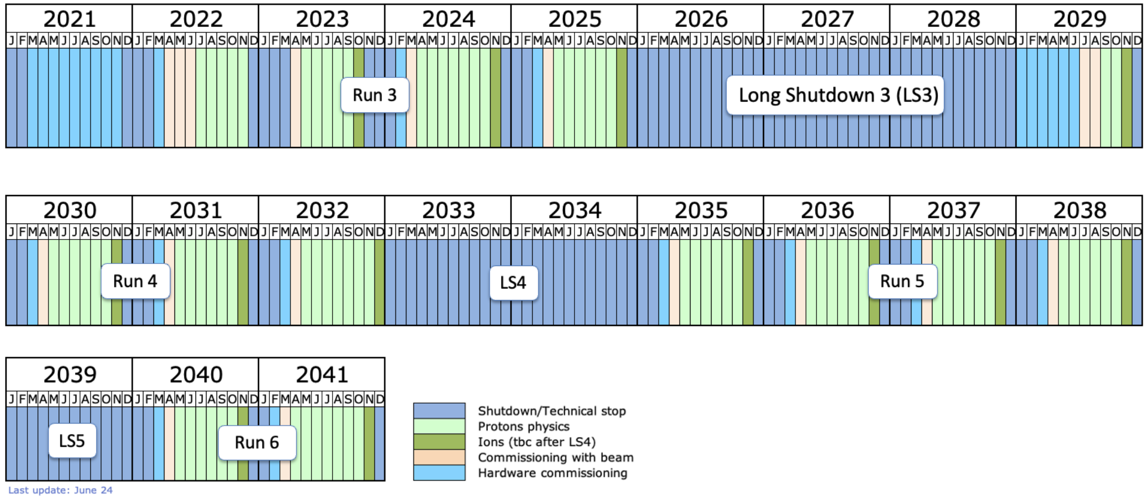


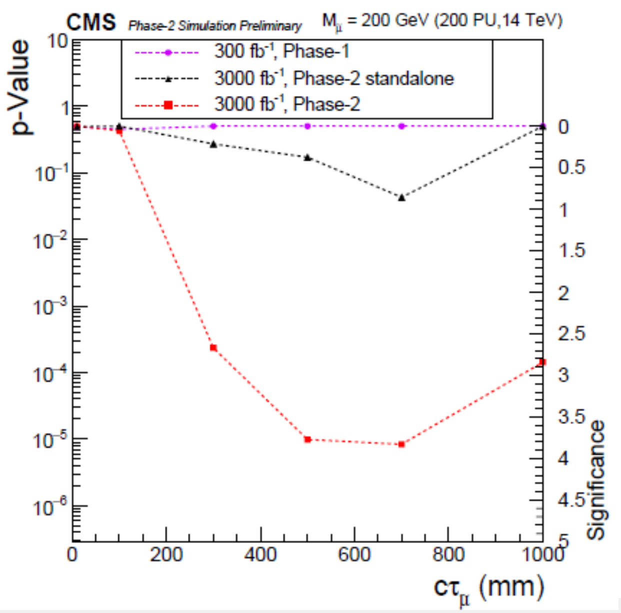


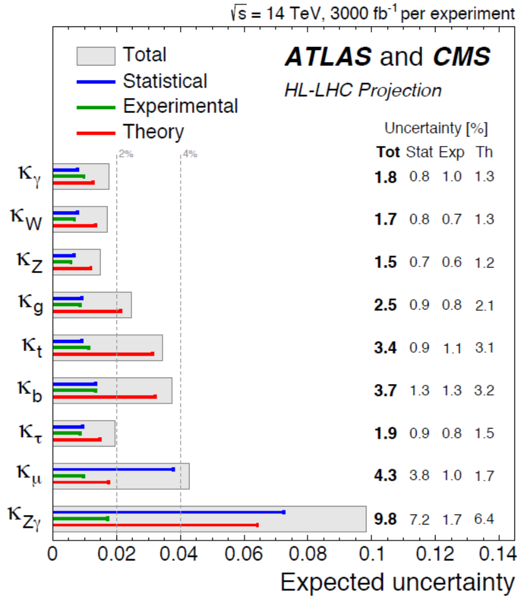
[CERN-LPCC-2019-01]

Towards the HL-LHC

- **Preparing for the big upgrade** of the LHC detectors, starting 2030.
- HL-LHC upgrade offers an **unprecedented opportunity** to explore uncharted lands and achieve scientific progress.
  - 10 times more data to what we will have by the end of Run 3 will facilitate a rich physics program.
- **Extend reach of new physics searches**: unexplored signatures (LLPs, HSCPs…) or regions of the phase-space will be within reach.
- **Improve current understanding of the SM and Higgs** sector by improving existing precision measurements and accessing rare decays ($H \rightarrow \mu\mu$) or production modes (HH) previously unseen at the LHC.
- However, this physics program will have to overcome **significant challenges** to succeed.



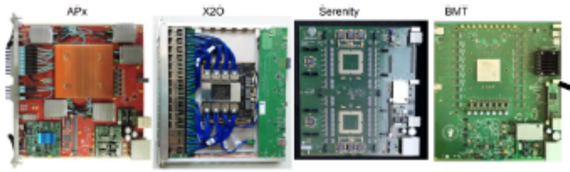




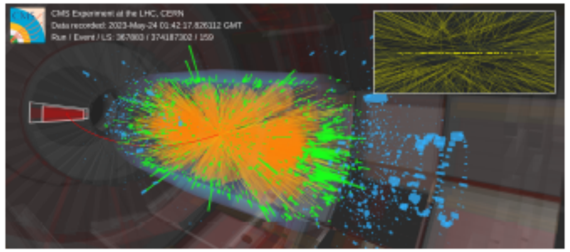
The Phase-2 Trigger Upgrade: Strategy

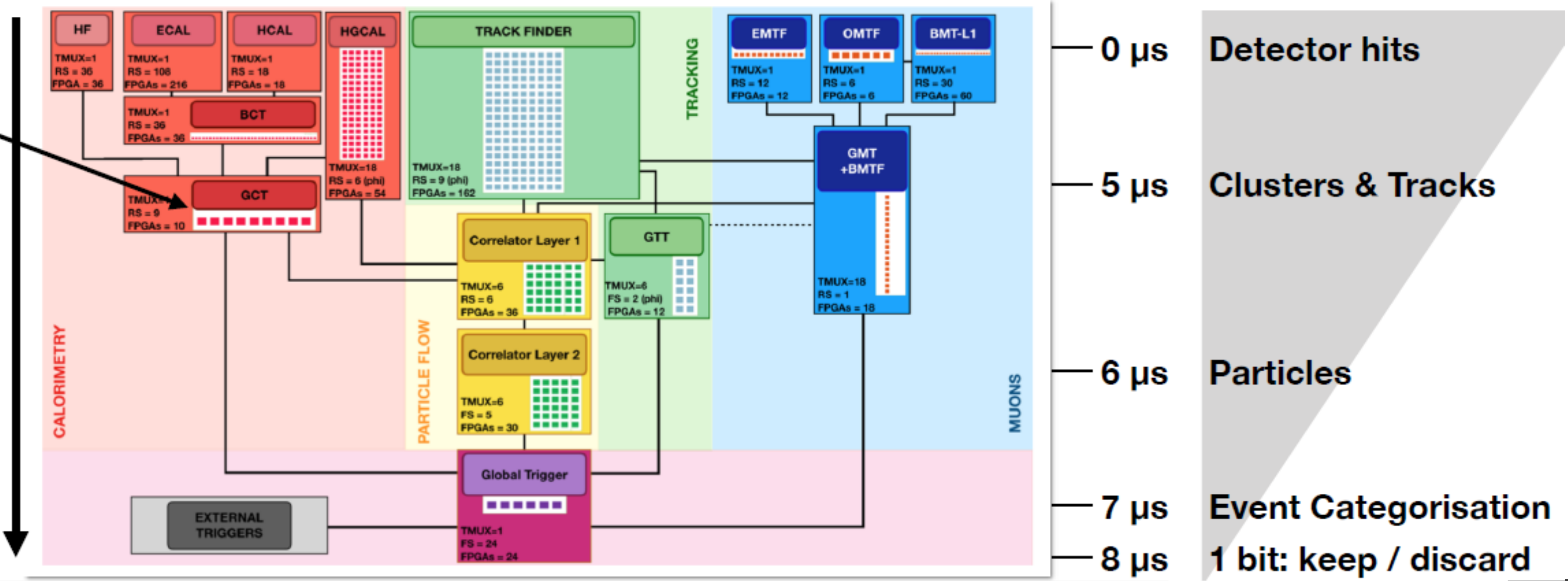
- **Benefit from the upgrade of the CMS detector:** high granularity information and tracking information
- The system allows a **throughput** of > **+64 Tb/s** using top-of-the-line FPGAs and ultra-fast optical links (25 Gbps).
  - Adapt and evolve as needs of experiment change.
  - Increased bandwidth to **750 kHz** at increased latency of **< 12.5 μs**
- Incorporate **sophisticated algorithms and advanced techniques** to extend CMS physics acceptance

t = 0



1 small box = 1 FPGA board with AMD VU13P FPGA

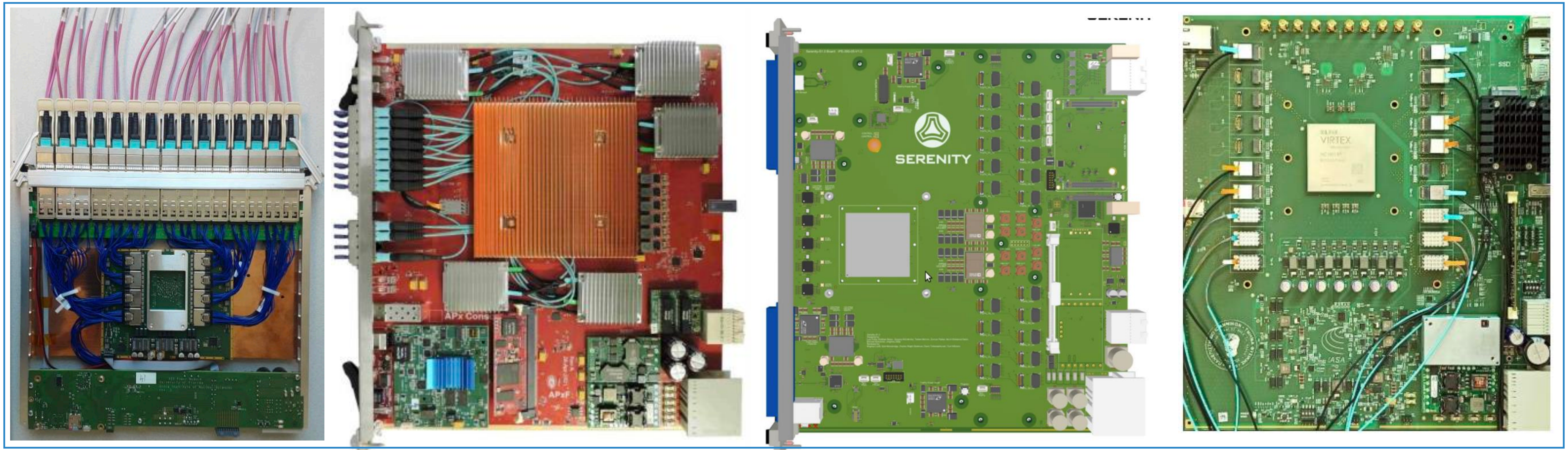




t < 12.5 μs

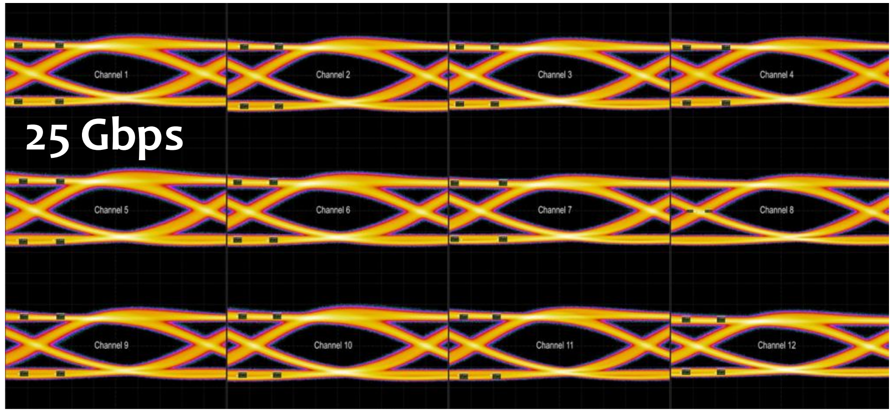
Hardware prototypes

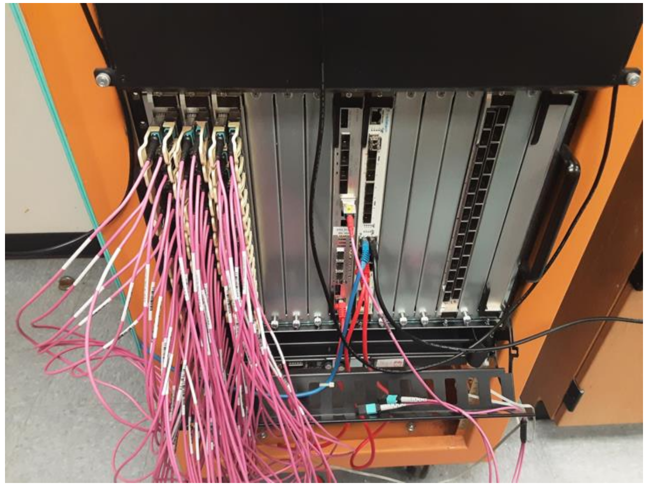
- **Design philosophy:**
  - Custom ATCA-boards. Generic Processing Engines → I/O, FPGA → sophisticated algo, arch flexibility
- **Design evolution:** increased I/O and computing power
  - FPGA : larger A2577 pin package, Xilinx Virtex Ultrascale VU13P
  - Optics : New denser version of on-board fly over Samtec Firefly & QSFP
  - Processors on board running commercial linux for flexible configuration and monitoring

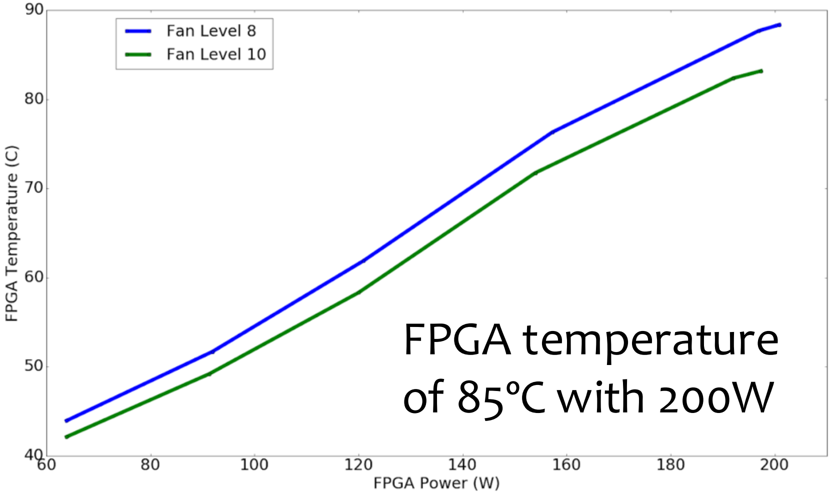


Hardware optics and thermal tests

- **Optical requirements:**
  - Support sufficient signal integrity in both the electrical and optical domains by demonstrating a bit error rate (BER) much better that 10^{-12}
  - Optics should provide sufficient optical margin with a receiver sensitivity better than -6 dBm to ensure operability at end of life (as laser degrades)
  - Tested Samtec Firefly x12 and QSFP (single and double density)
- **Thermal performance**
  - Integrity of the optics and FPGAs

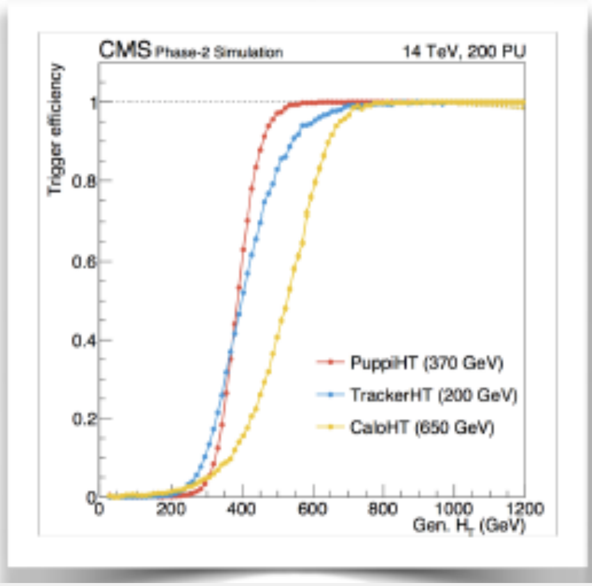


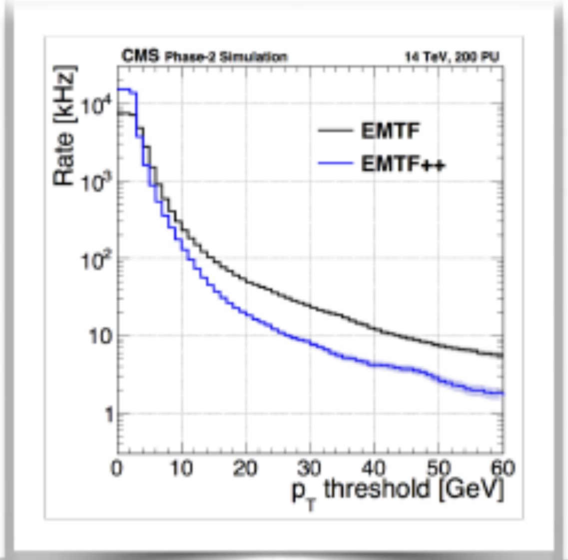


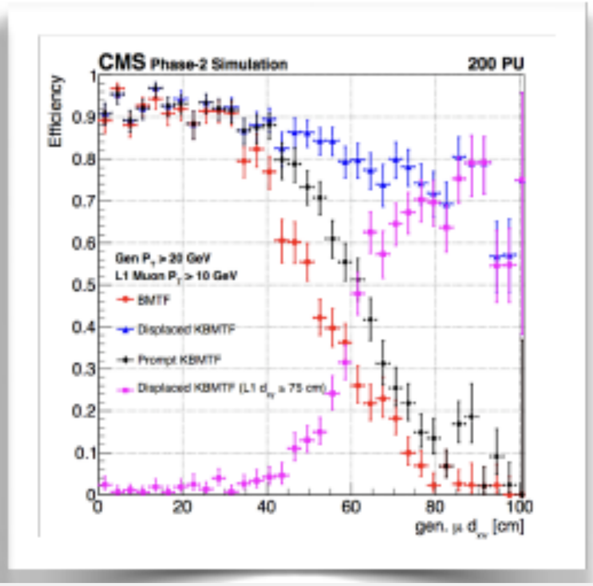


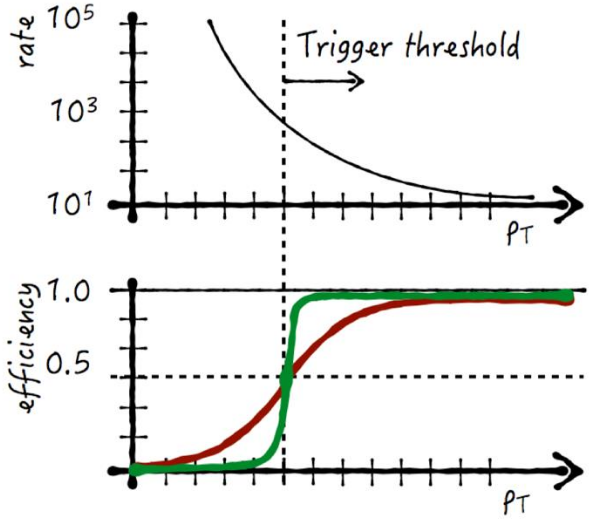
Algorithms for the Level-1 trigger

- Extensive use of tracking to reach near offline performance (sharper efficiency turn-on curves) + reconstruction of Primary Vertex.
- Exploit complementarity of different object flavor:
  - Standalone objects: robust triggers based on independent sub-detectors
  - Track-matched objects: tracking used to confirm standalone Muon and Calo objects, significant improvement with simple design
  - Particle-flow objects: ultimate performance improvement, combine all information to match offline algorithms, require most processing time and resources for calculation





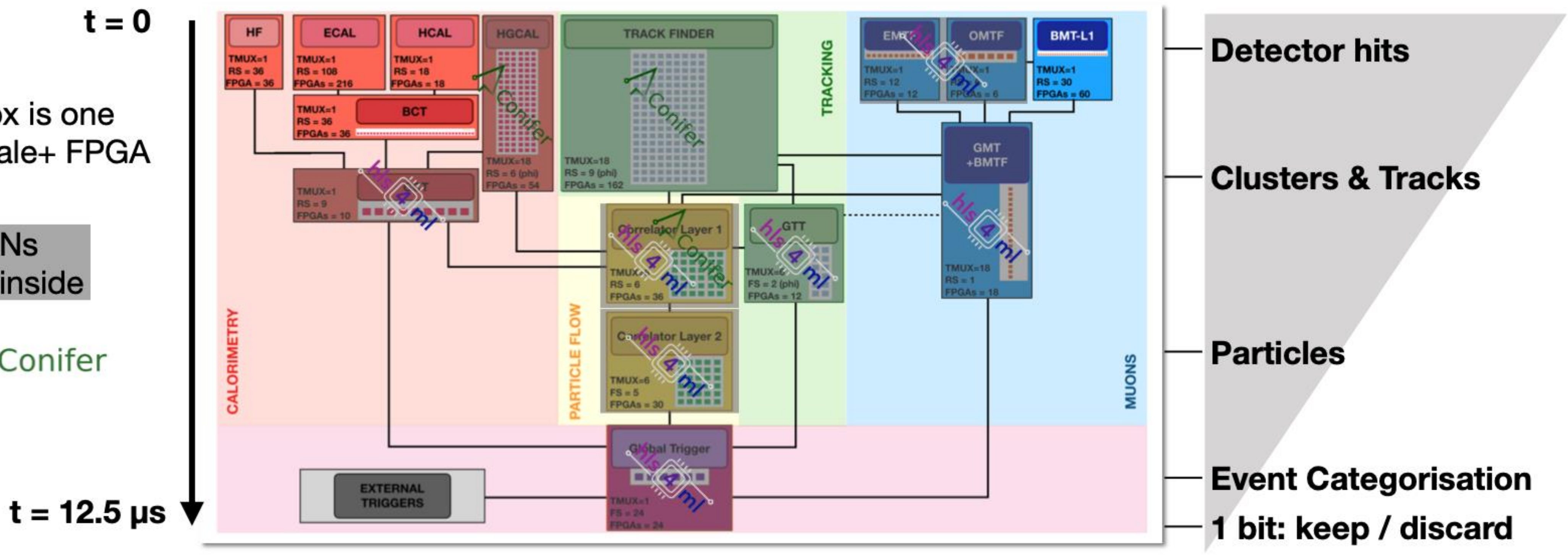




Extensive use of ML algorithms

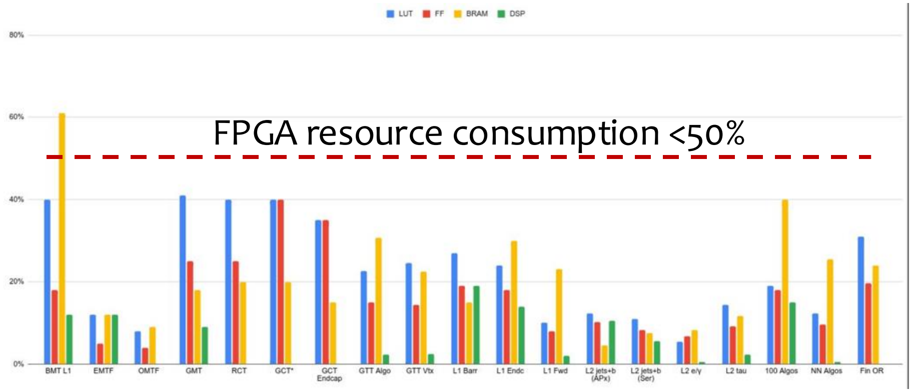
Each small box is one Xilinx Ultrascale+ FPGA

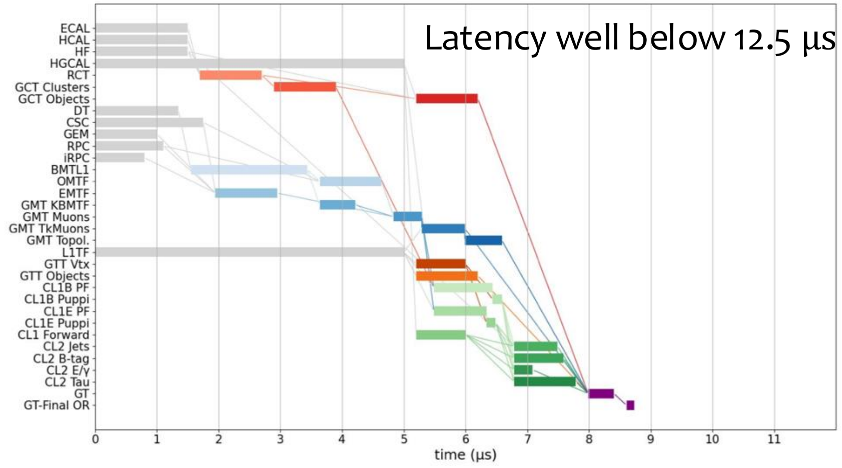
These have NNs and/or BDTs inside

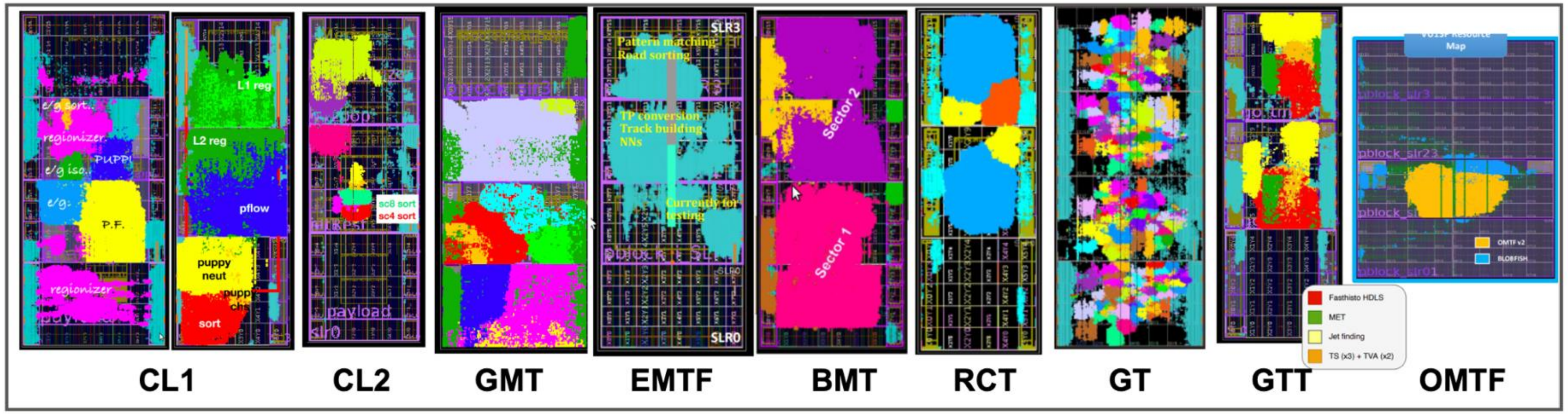


Algorithm into firmware: latency and resource utilization

- **Firmware design and integration:**
  - Algorithm developed mostly in C → High Level Synthesis (HLS). Using Vivado HLS, Vitis HLS
  - Many tools available for Machine Learning inference: hls4ml, Conifer for BDT evaluation
  - New fixed-point arithmetic in C++ [taken from Xilinx libraries] → emulator firmware
  - Continuous integration of the firmware in repository
  - Verify timing, resources utilization & latency: all using less than 50% resources, whole system evaluated to 8.6 μs

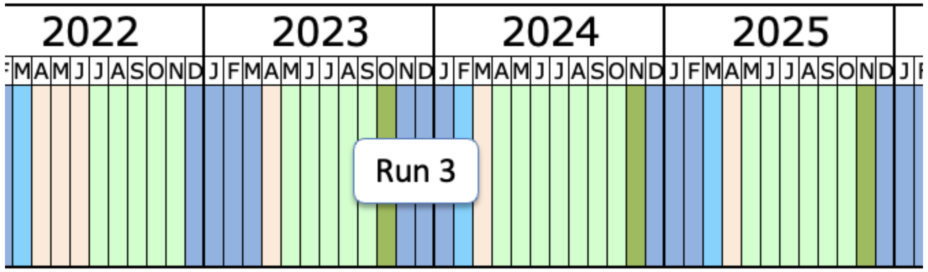


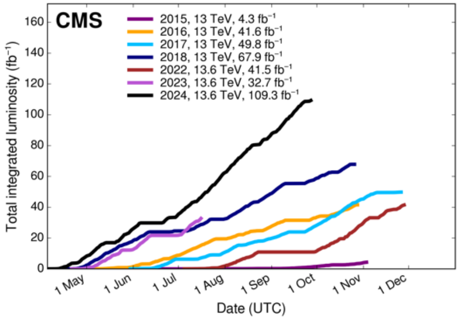


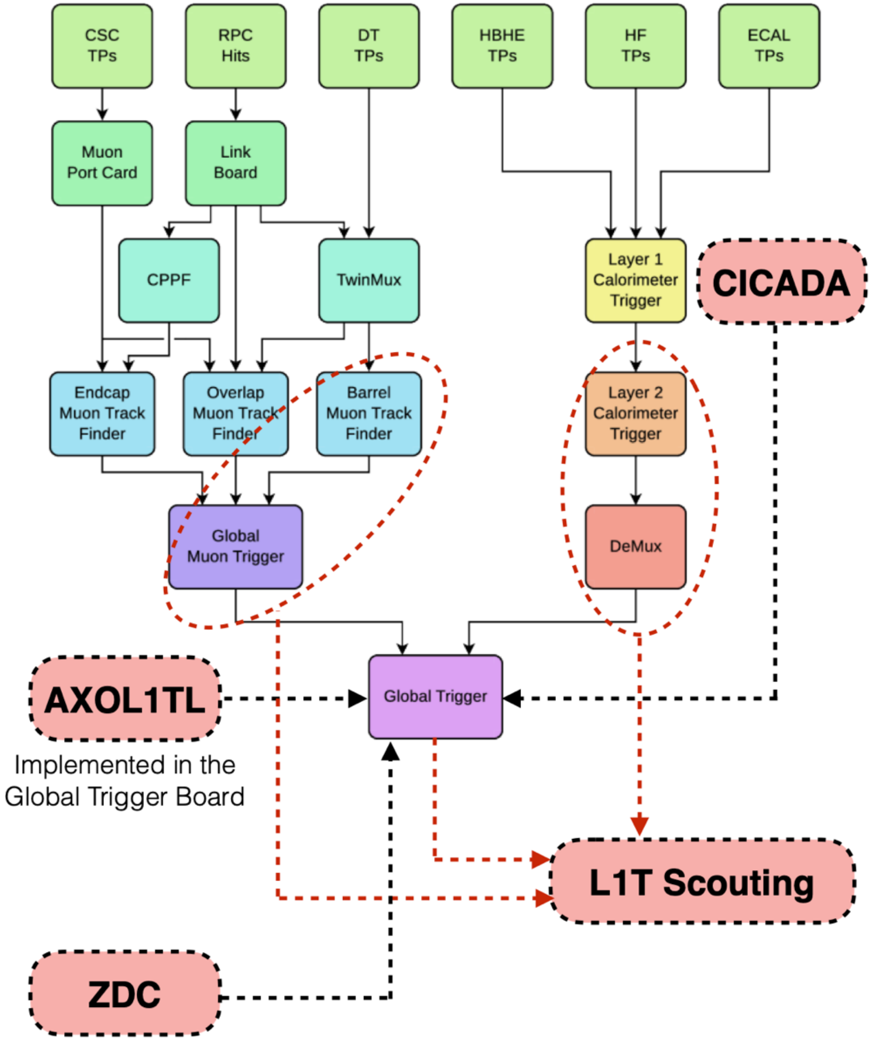


Testing new ideas during Run 3

- With almost one and half year to go, Run-3 has already surpassed Run-2 luminosity
  - **Almost 170 pb^{-1} recorded**
- Successful feedback loop into the current system: the Run-3 system now features new algorithms, optimisation techniques, hardware, inspired from the phase-2 upgrade project:
  - LLPs triggers: displaced muons, muon showers, delayed jets…
  - 40 MHz scouting (real-time data analysis)
  - Inclusion of the first anomaly detection trigger on live data: AXOL1TL and CICADA
- System exceeding original design. **Having a flexible design is an advantage!**

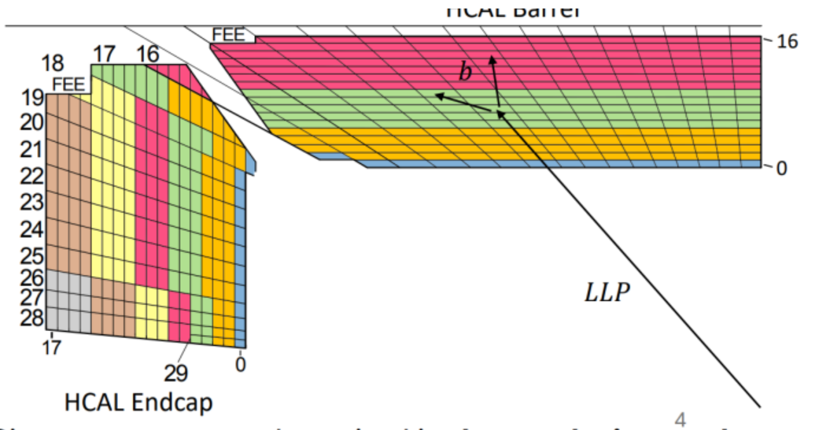






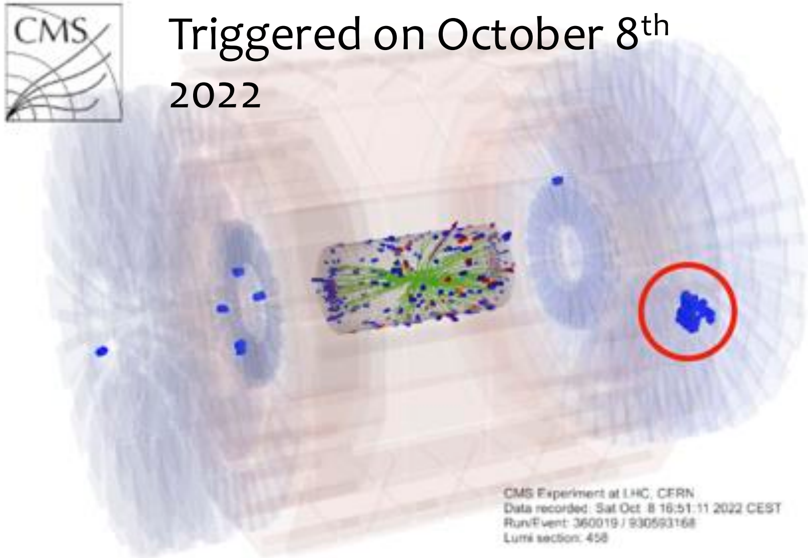
Displaced/delayed jets

- **ECAL** measures arrival time of objects with precision of ~200 ps (for energy deposits >50 GeV). Tau seeding at L1 and trackless jets at HLT
- Use **HCAL** time information at the L1 trigger level to identify delayed jets (>6ns). Prompt veto applied
- High multiplicity at the **muon system** for long-lifetimes

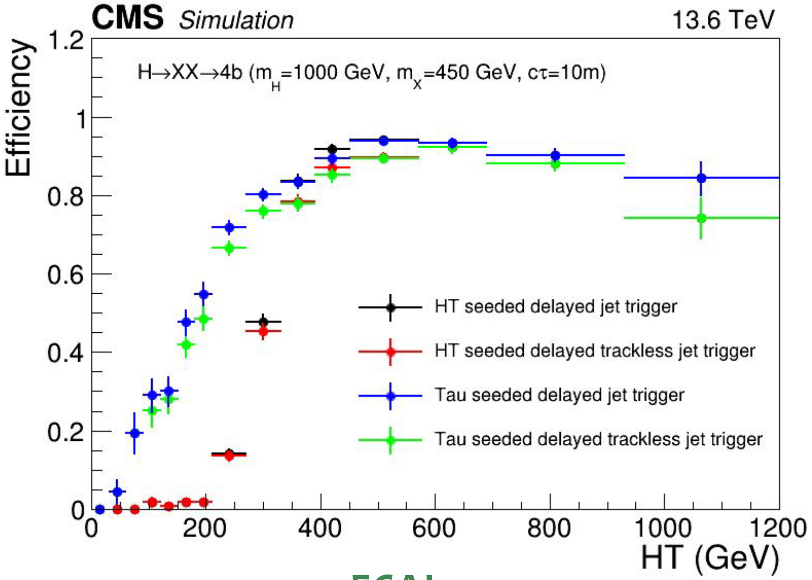


Signatures: energy deposited in **deep calorimeter layers**

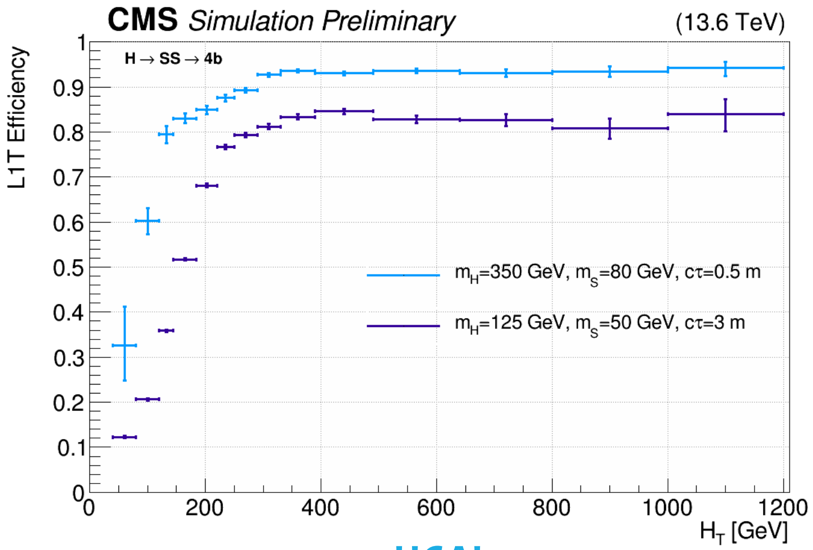
Triggered on October 8th 2022



Muon (*more later*)



ECAL

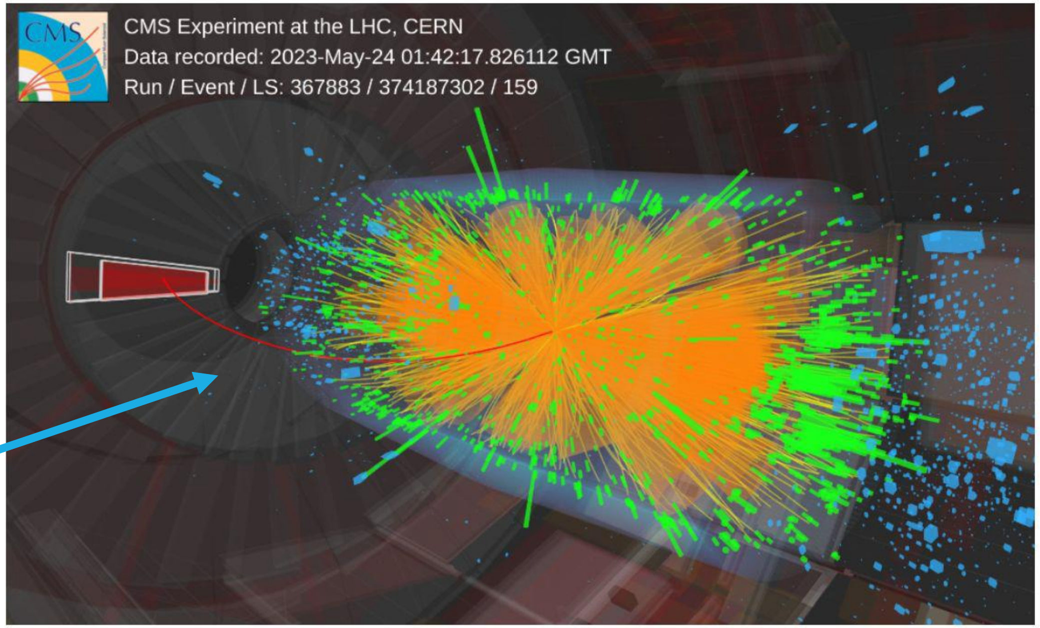


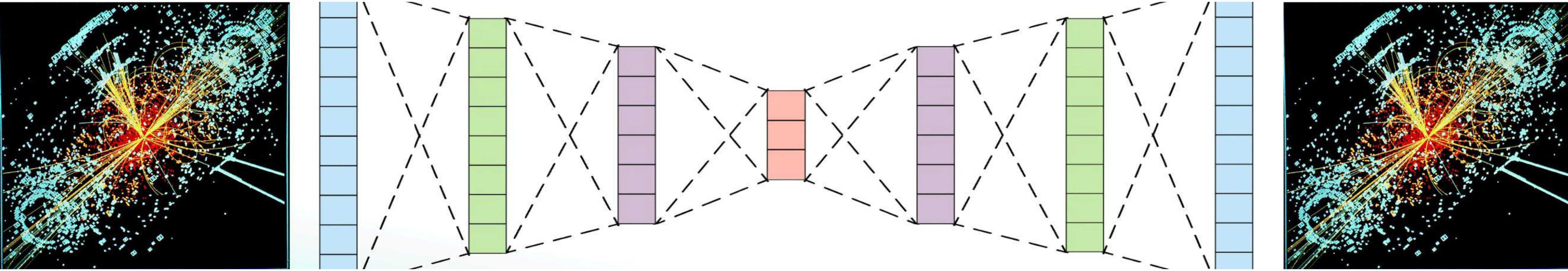
HCAL

ML at L1: Anomaly detection

- Where's the new physics? To find anything, you need a trigger
  - If we knew what we were looking for, we'd build a trigger for it!
- Cast a wide, model-independent net
  - Learn what an average event looks like, pick things that are rare
  - Autoencoder, trained on random beam events
  - Reconstruction error is a metric for anomalous-ness
- AXOL1TL & CICADA
  - Low-level variables (L1T or Calorimeter objects)
  - Outputs an anomaly metric to keep the event or not

CICADA AXOL1TL

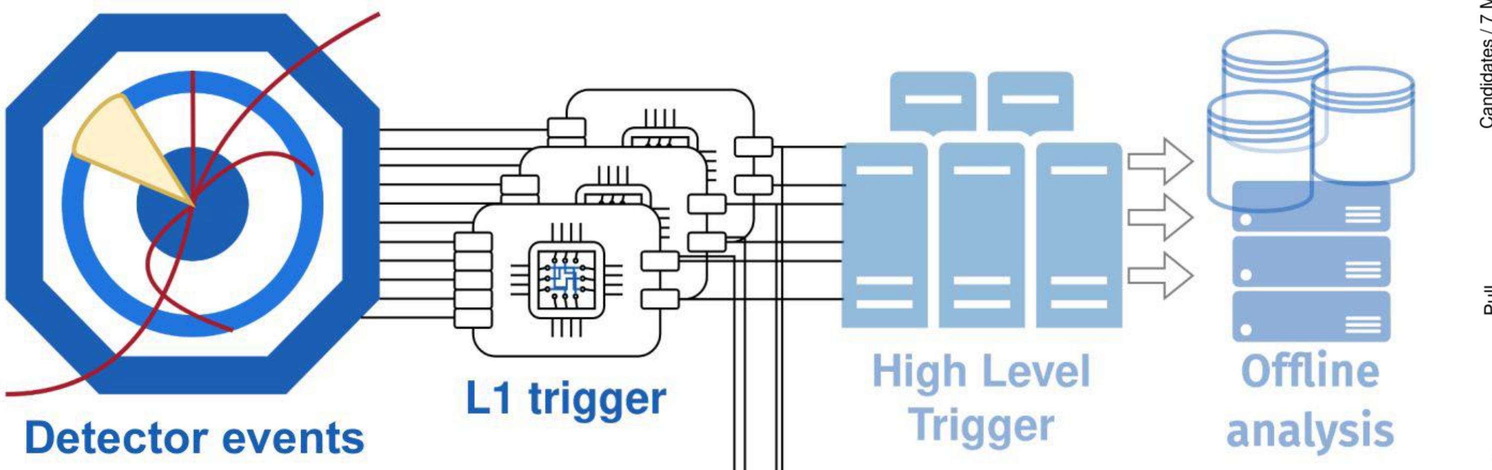


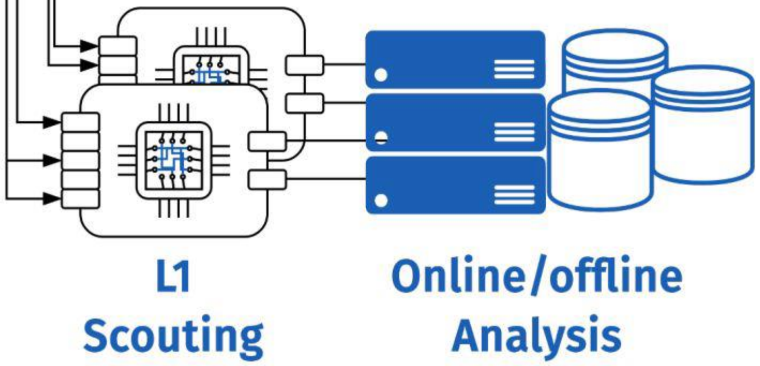


https://arxiv.org/abs/2403.16134
https://cds.cern.ch/record/2904692/

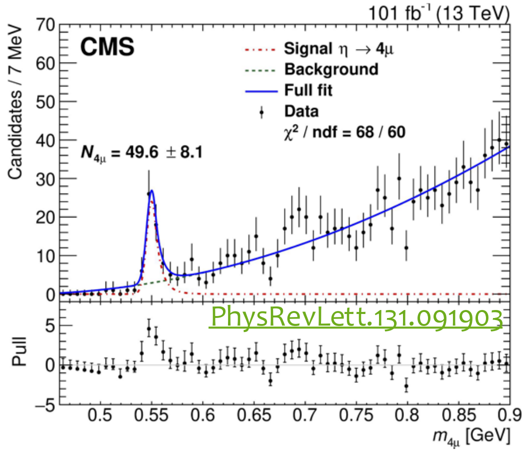
Triggerless analysis (aka scouting)

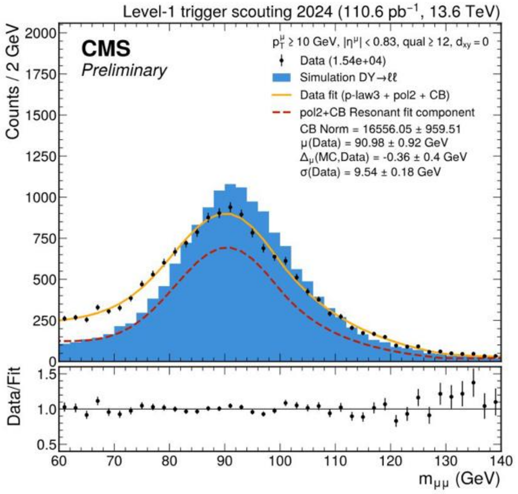
- Storing and analysing events at L1 or HLT (x100 smaller event size)
- Crucial for very low-mass bump-hunt searches, compressed spectra or b-physics





- **L1 scouting:** standard L1 rejects 99.75% events. L1 scouting will allow us to have a look at those events
- Proof of concept in Run 3





Level-1 Data Scouting rack

Universidad de Oviedo

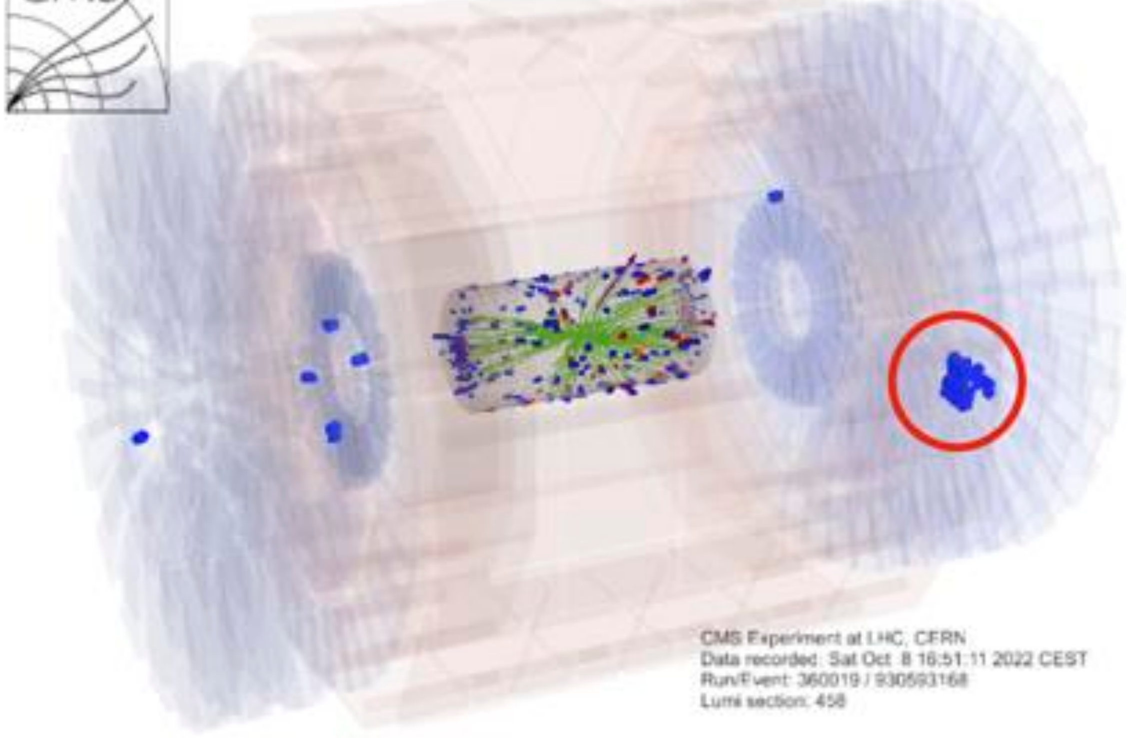
Impact of trigger design beyond HEP

- **Impact on society**
  - Massive surge of data and AI applications. The need of processing large amounts of data is an ever-increasing challenge.
  - HEP experiments provide the perfect test bed for advanced AI algorithms developments, real-time data processing and low-power solutions
- **Developing ideas for CMS trigger and beyond:** NextGen and INTREPID projects
  - Enhance the triggers and the data collection and processing, and thus the scientific potential, of ATLAS and CMS in the HL-LHC phase **beyond the currently projected scope**.
- **Driving a lot of attention**
  - from national and international funding agencies and industrial partnerships (CMS is working with Amazon, Google, Micron...)
  - Emerging applications outside HEP: data reduction onboard satellites, quantum control systems, brain implants…
  - Custom silicon for Machine Learning is big industry trend - acceleration of specific workloads

Conclusions

- The CMS trigger system for HL-LHC will process data at ~64 Tb/s using top-of-the-line FPGAs and high-speed links
- Level-1 Hardware trigger with enhanced capabilities complying with physics requirements using sophisticated ML-based algorithms
- Modular and flexible design to adapt for future ideas using custom ATCA boards
- Hardware demonstration ongoing and some tests in Run-3 data taking
- Future designs are showing exciting prospects, even beyond HEP

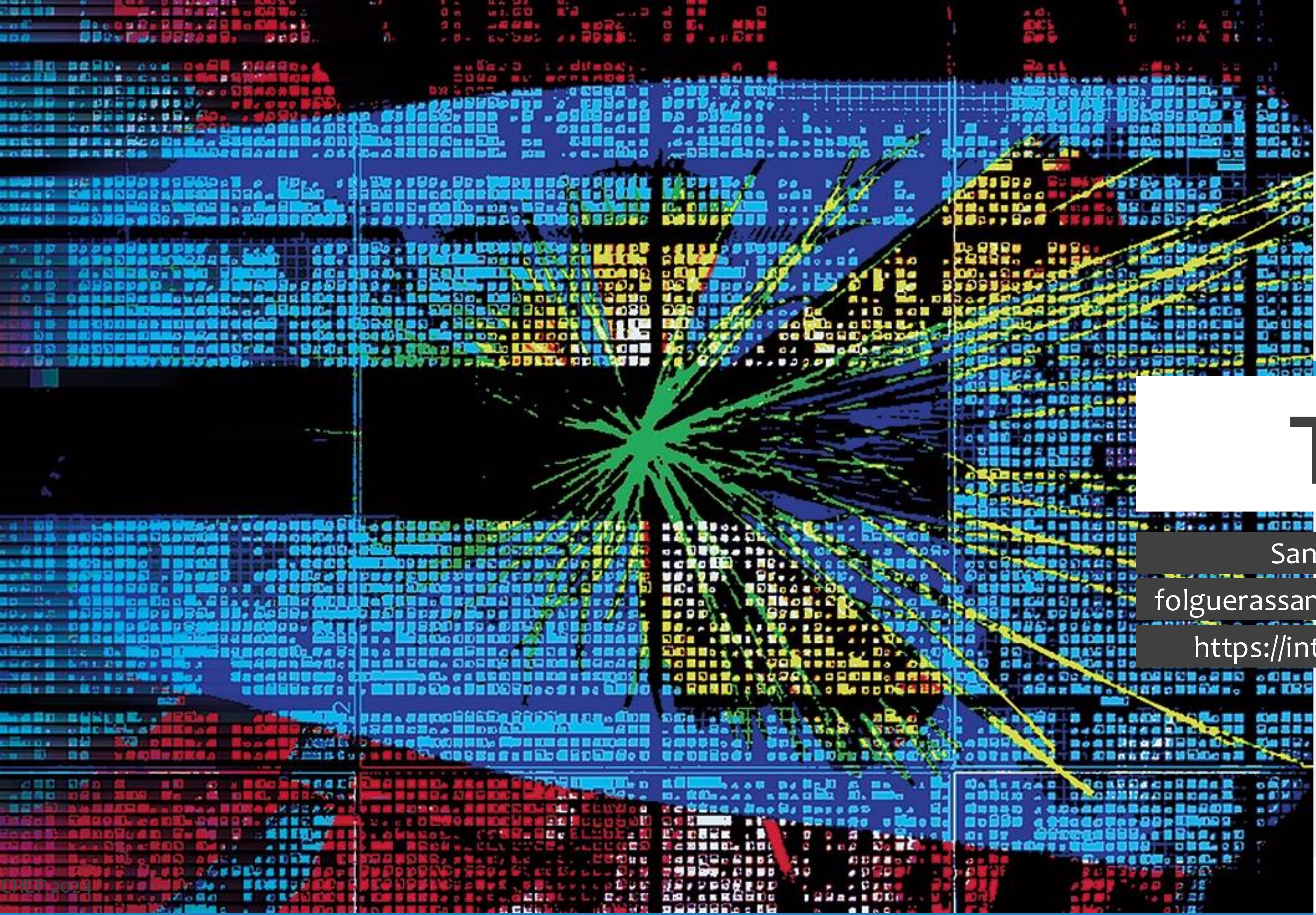
Triggered on October 8th 2022



Funded by the European Union

European Research Council
Established by the European Commission

INTREPID



Funded by the European Union

European Research Council
Established by the European Commission

INTREPID

Thanks!

Santiago Folgueras
folguerassantigo@uniovi.es
https://intrepid.uniovi.es/

Triggering TB/s of data: The LHCb perspective

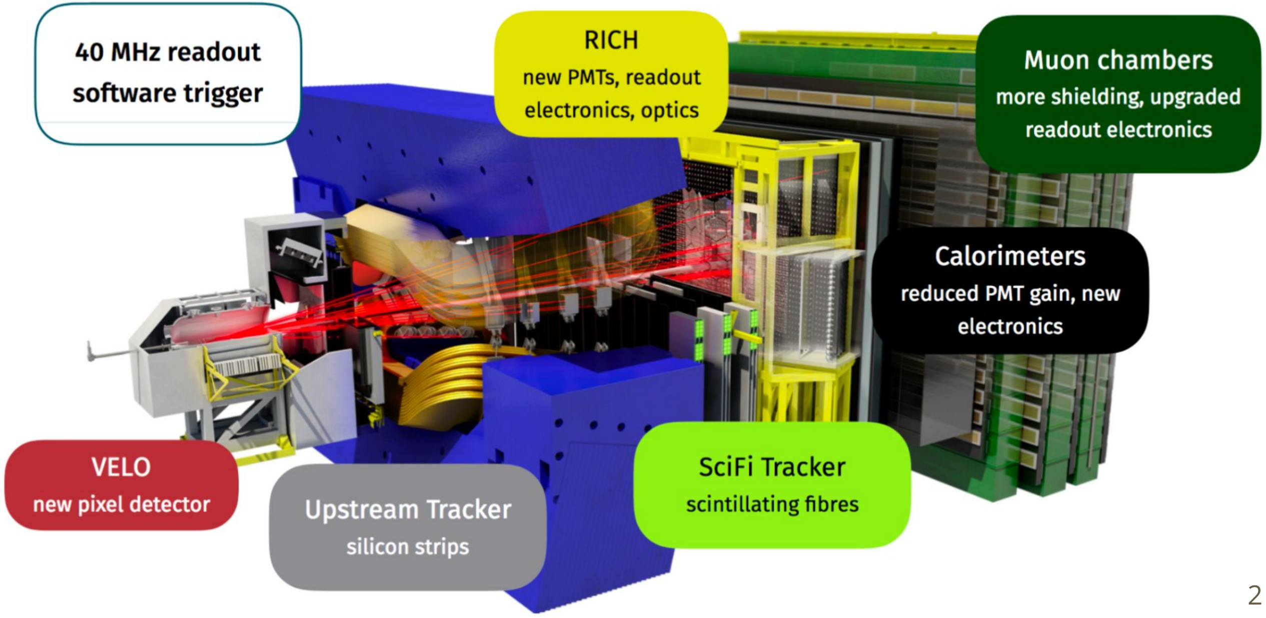
Marianna Fontana,
on behalf of LHCb
CHEP conference, 19-25 October 2025,
Krakow

The LHCb experiment

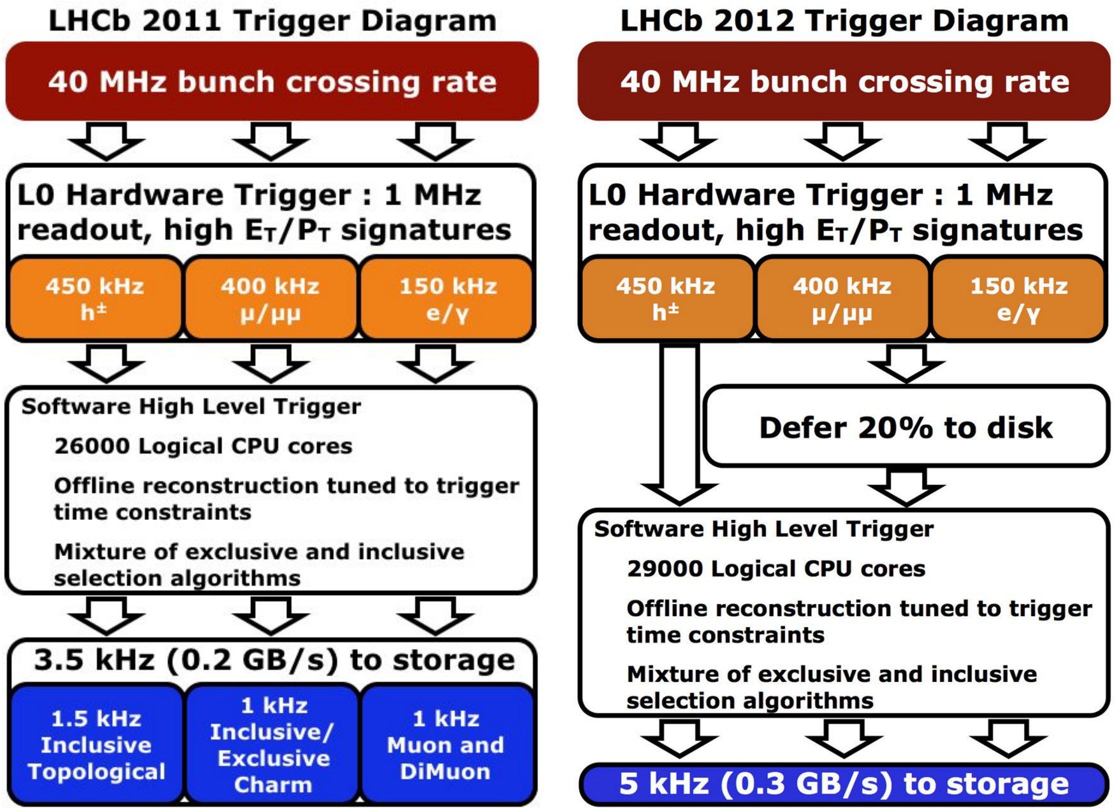
- Experiment dedicated to flavour physics
- Successfully took 9 fb^{-1} of data during Run 1-2
- **Major upgrade** of all subtectors **for Run 3**
- Factor 5 increase in instantaneous luminosity → pile-up of 5

CERN-LHCC-2012-007

- 100% of the readout electronics replaced
- **New data acquisition system** and data center



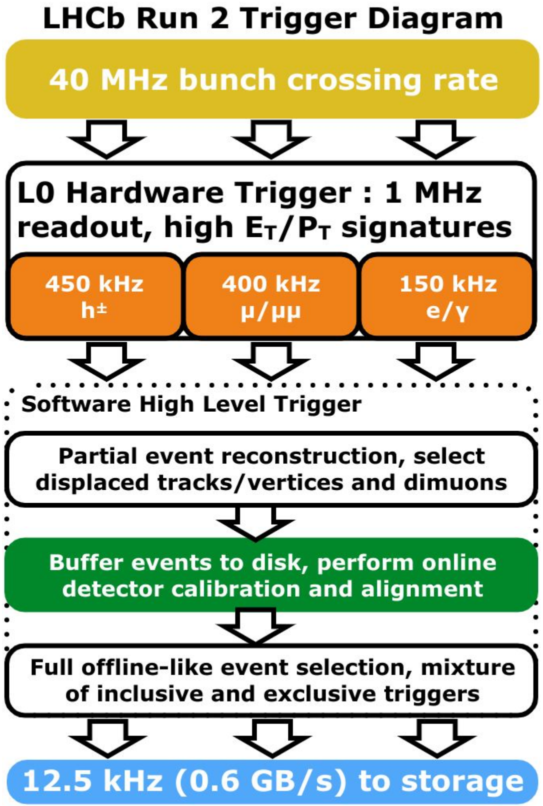
The trigger evolution: Run 1



- L0 hardware level for high Et/pt signatures
- HLT1 running tracking (for high-pt) including Kalman filter
- HLT2 *almost* full event reconstruction
- Much bigger output rate than originally foreseen
  - Inclusive selections for full beauty programme
  - The charm programme initially not foreseen became a reality

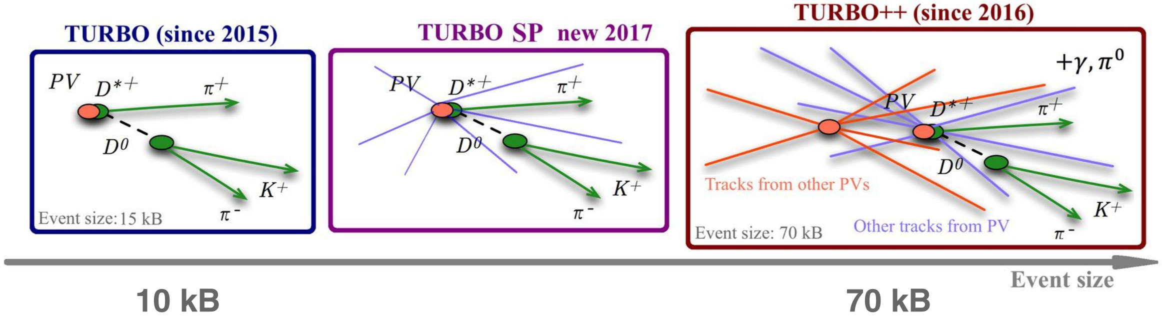
[Comput. Phys. Commun. 208 (2016) 35-42](#)
[JINST 14 P04006](#)

The trigger evolution: Run 2



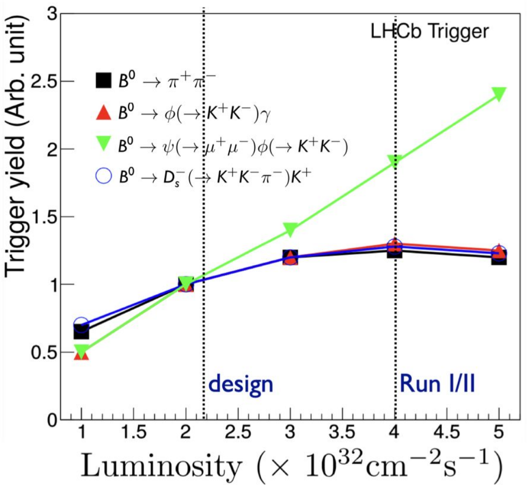
Disk buffer moved between HLT1 and HLT2 → increased number of CPUs and enabled

- Real-time alignment and calibration
- Real-time reconstruction with analysis quality reconstruction
- Ability to use trigger output for analysis and discard raw detector information in trigger (**Turbo stream**) [J. Phys.: Conf. Ser. 664 082004]
  - System fully commissioned already in 2015 with physics publications. It became the baseline for a good fraction of the Run 2 physics programme
- Adopted as the baseline approach for Run 3



The trigger (R)evolution: Run 3

J. Phys.: Conf. Ser. **878** 012012



- In Run 1-2 couldn't efficiently trigger on heavy flavour using hardware signatures
- Trigger for many hadronic channels saturated
- Solution: **fully software trigger**

https://home.web.cern.ch/news


LHCb begins using unique approach to process collision data in real-time

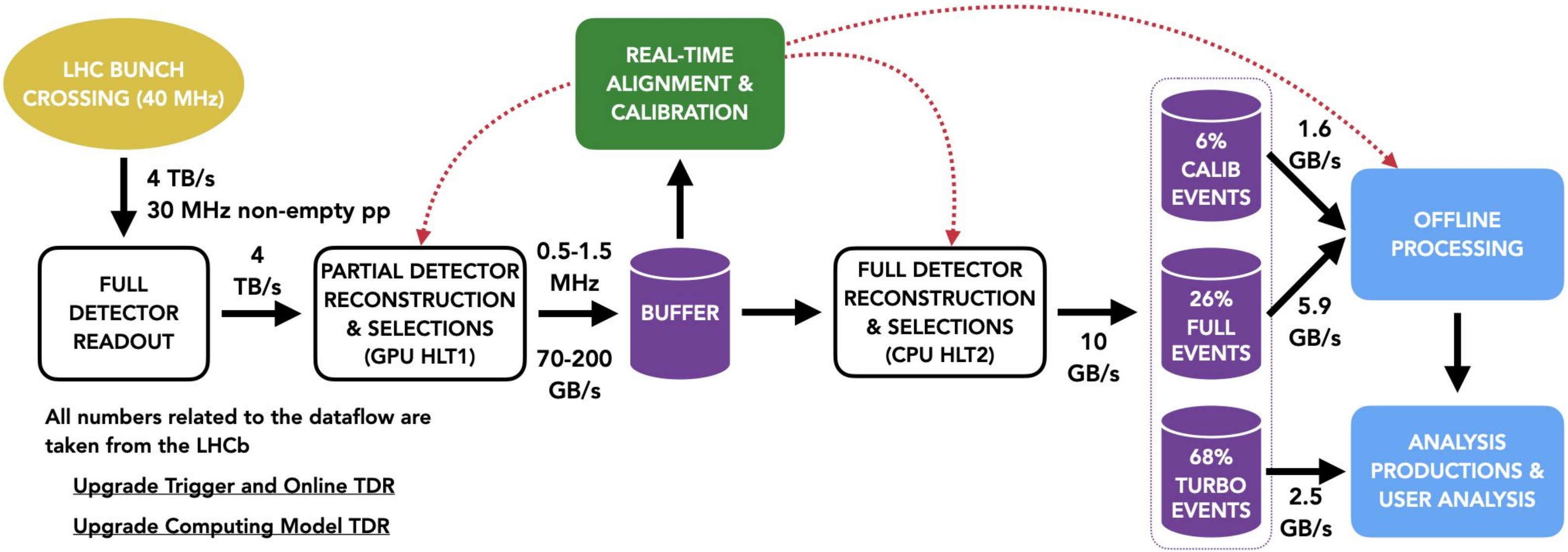
Using a new system called real-time analysis, the LHCb collaboration has made filtering and analysing experiment data simpler and faster

1 MARCH, 2023 | By LHCb collaboration

(Image: CERN)

The Run 3 data flow

LHCb-FIGURE-2020-016



- Detector data @30 MHz received by O(500) FPGAs
- 2-stage software trigger, HLT1 & HLT2
- Real-time alignment & calibration
- After HLT2, 10 GB/s of data for offline processing

HLT1 trigger

See talks from
- A. Scarabotto
- J. Horsvill

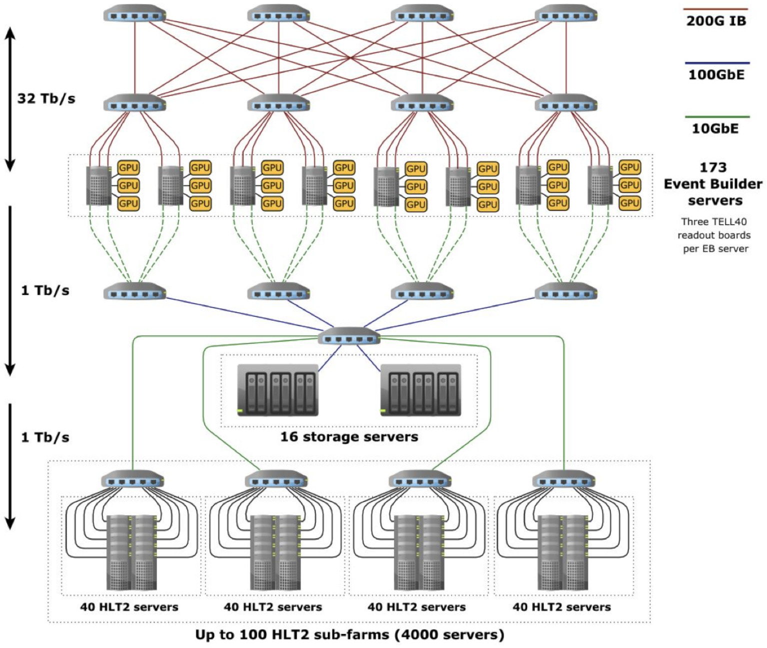
- Take as input LHCb raw data (**4 TB/s**) at 30 MHz
- Perform partial event reconstruction & coarse selection to cover the full breadth of LHCb physics
- Reduce the input rate by a factor of 30 (~1 MHz)
- **~ 500 GPUs NVIDIA RTX A5000 GPUs** installed
  - The baseline TDR design could be achieved with 300 GPUs
  - Extra GPU power used to extend the improvements beyond-TDR

The GPU choice matches the DAQ architecture of LHCb

- GPUs can be hosted by the Event Builder Nodes via PCIe slots
- reduced costs due to shared powering and cooling and smaller network

HLT1 tasks are suited for parallelisation:

- Events can be treated independently
- Objects of reconstruction (tracks, vertices, ...) are independent



Comput.Softw.Big Sci. 6 (2022) 1, 1

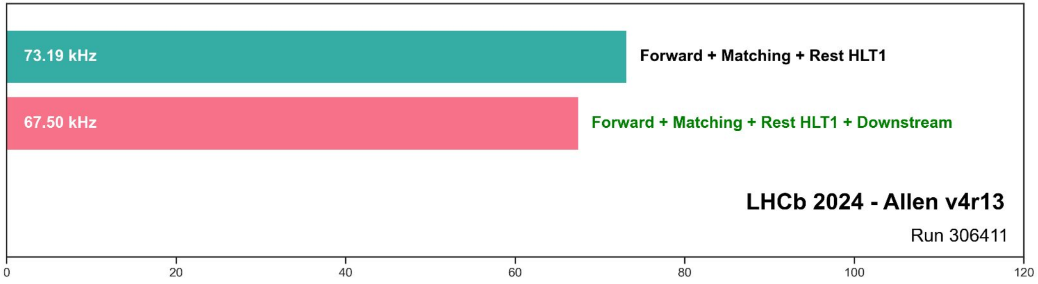
Allen: LHCb HLT1 trigger

Comput. Softw. Big. Sci. **4**, 7 (2020)

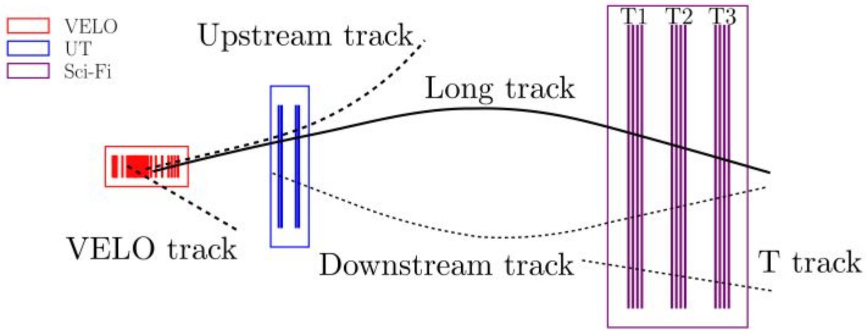
Partial event reconstruction through

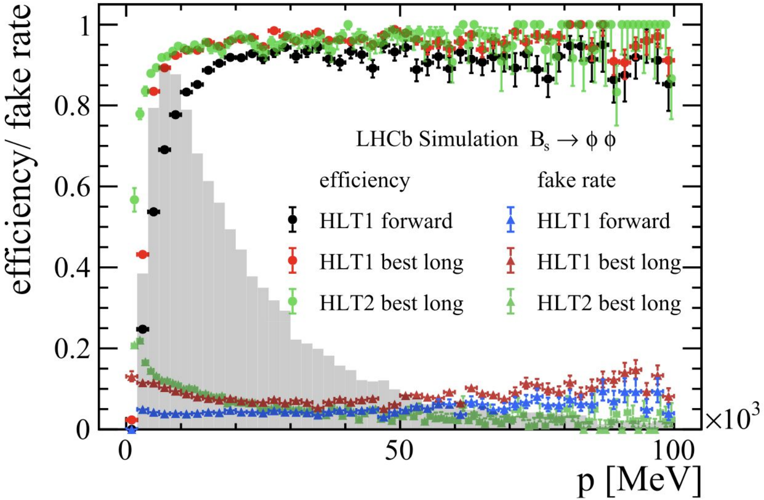
- Track reconstruction for all the track types used in physics analysis (Long and Downstream* tracks) [See talk by J. Zhuo]
- Vertex reconstruction
- Electron clustering* and bremsstrahlung recovery*
- Muon identification

HLT1 Throughput per GPU



LHCb-FIGURE-2024-035



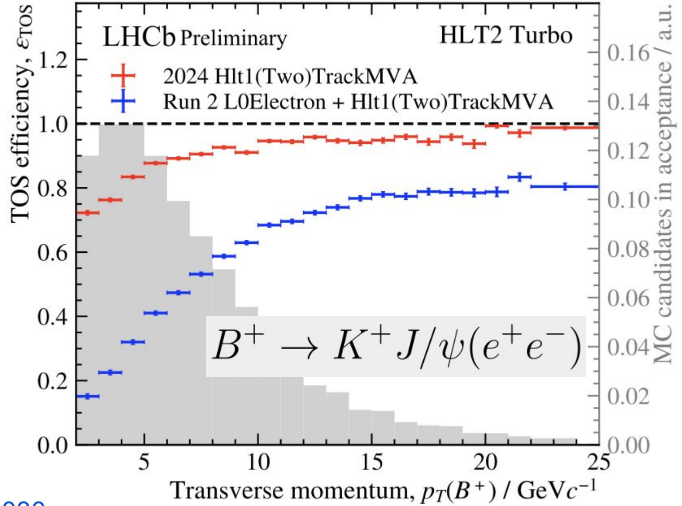


IEEE Access, vol. 12, 2024

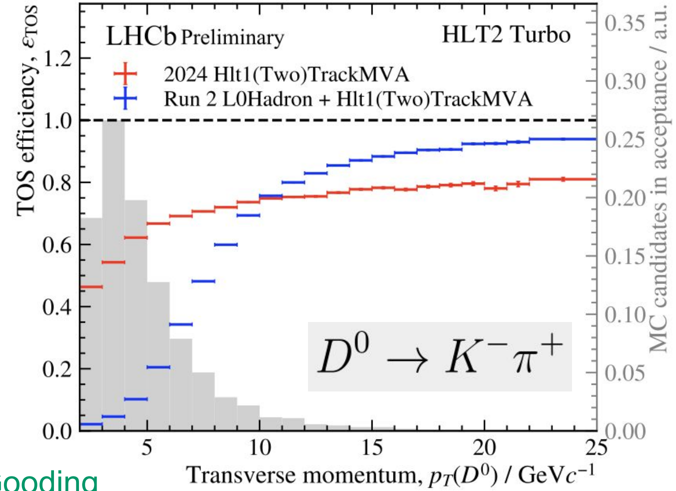
* beyond TDR

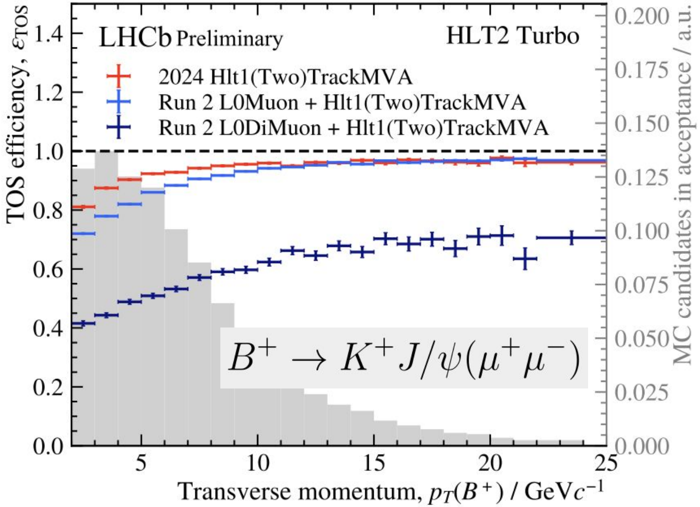
HLT1 performance

- **The real-time analysis philosophy proved to be valid**
- Significant improvements in trigger efficiencies
- Huge gain a low-pT
  - Beneficial for the charm and strange physics programme
- Large impact for electron channels
- Muon channels gained from the removal of the global event cuts



LHCb-FIGURE-2024-030



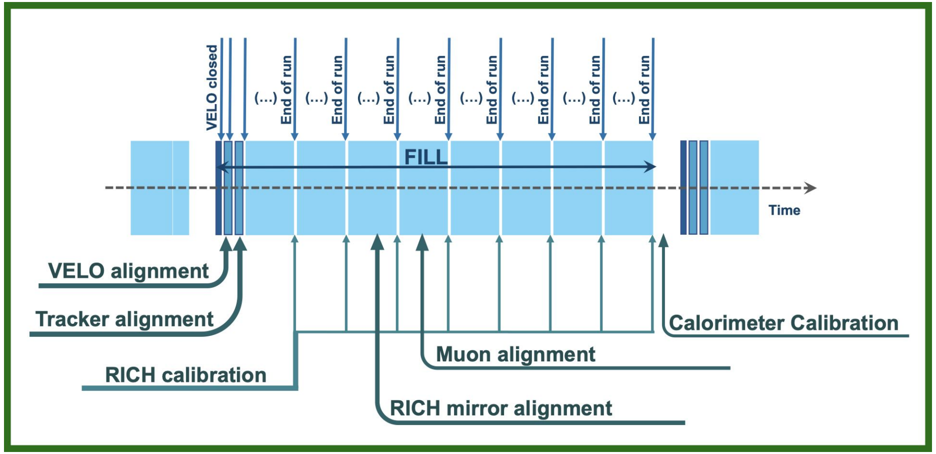


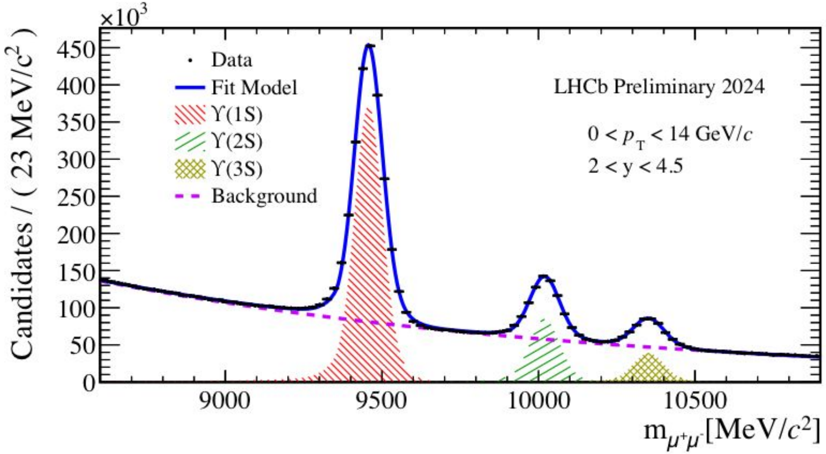
See talk by J. Gooding

Alignment and calibration

- Store data selected in HLT1 in intermediate buffer of O(30 PB) for real-time alignment and calibration
- Fully aligned and calibrated detector needed to have offline-quality reconstruction in HLT2
- Online alignment and calibration pioneered in Run 2, crucial in Run 3
- Two types of processes
  - Alignment: VELO, RICH mirrors, UT, SciFi, Muon
  - Calibration: RICH, ECAL, HCAL

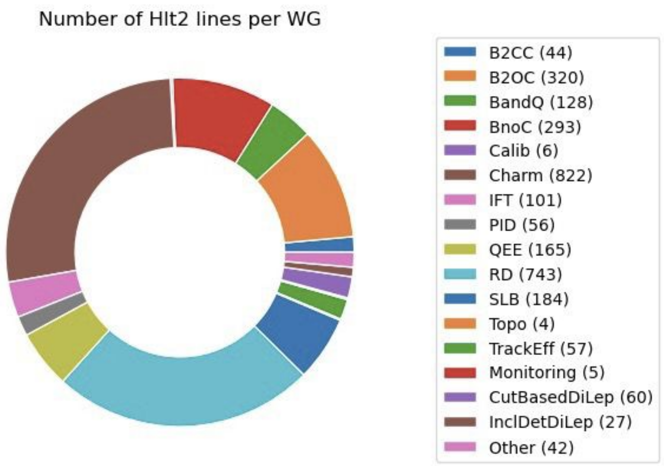
LHCb-FIGURE-2024-025

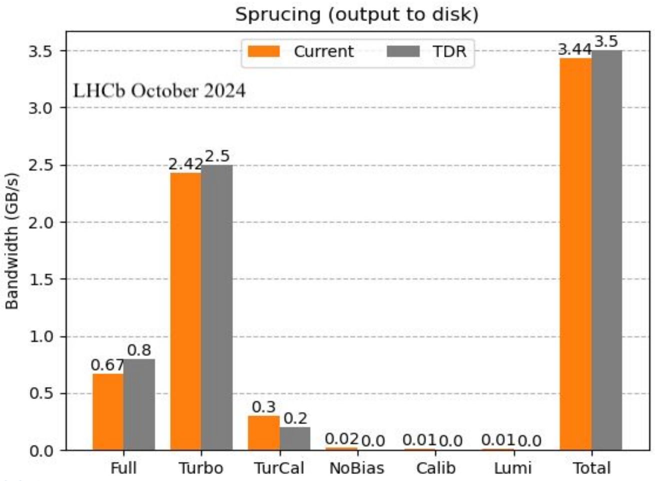




LHCb HLT2 trigger

- HLT2 runs a full reconstruction and all the necessary selections (inclusive but mostly exclusive) for the wide LHCb physics programme (~3000 lines)
- Given the hard limit on bandwidth (10 GB/s to tape and 3.5 GB/s on disk) and expected signal rate, event size is the only free parameter
- Need to "persist" all the reconstructed objects for offline analysis
- The successful strategy of the Turbo paradigm used at full speed also in Run 3

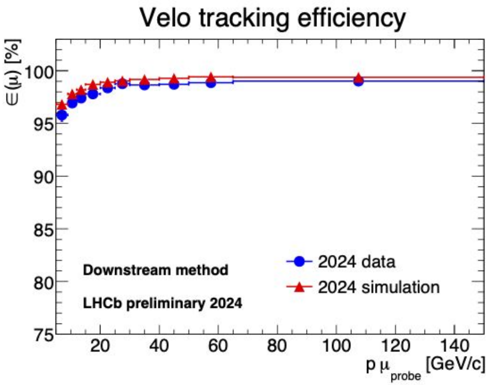




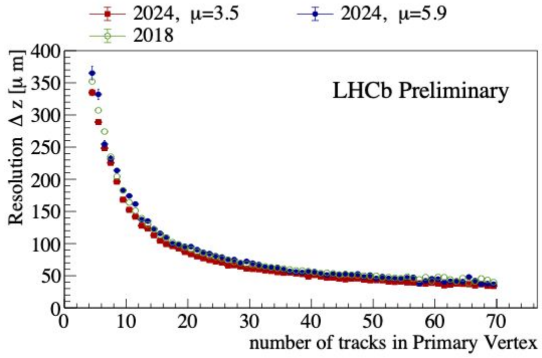
LHCb-FIGURE-2024-034

HLT2 performance

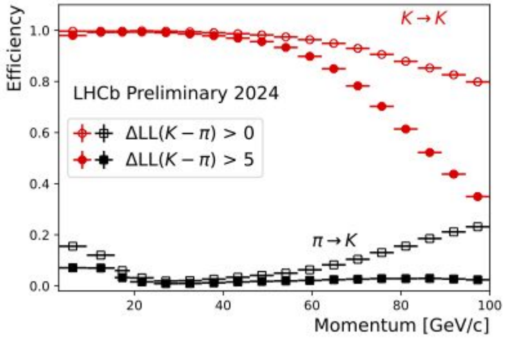
Achieving TDR performance for vertex resolutions, track reconstruction and PID performance



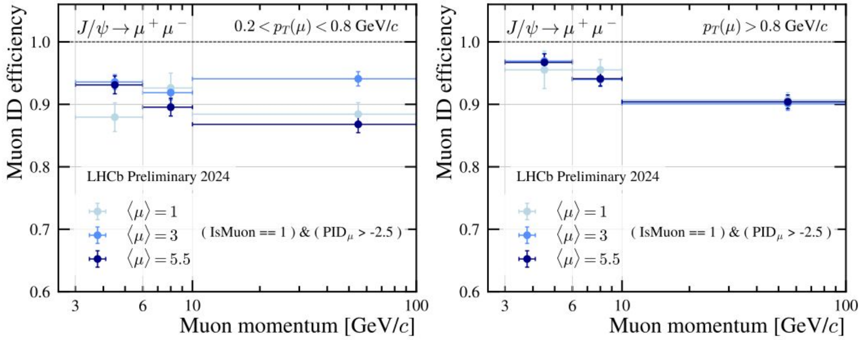
LHCb-FIGURE-2024-032



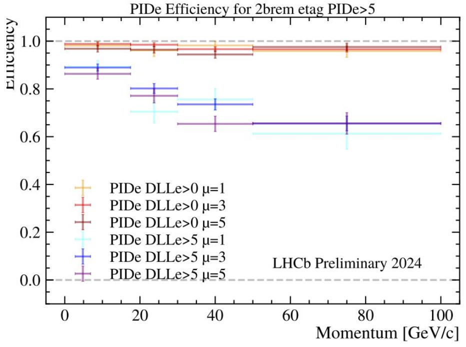
LHCb-FIGURE-2024-011



LHCb-FIGURE-2024-031



LHCb-FIGURE-2024-010

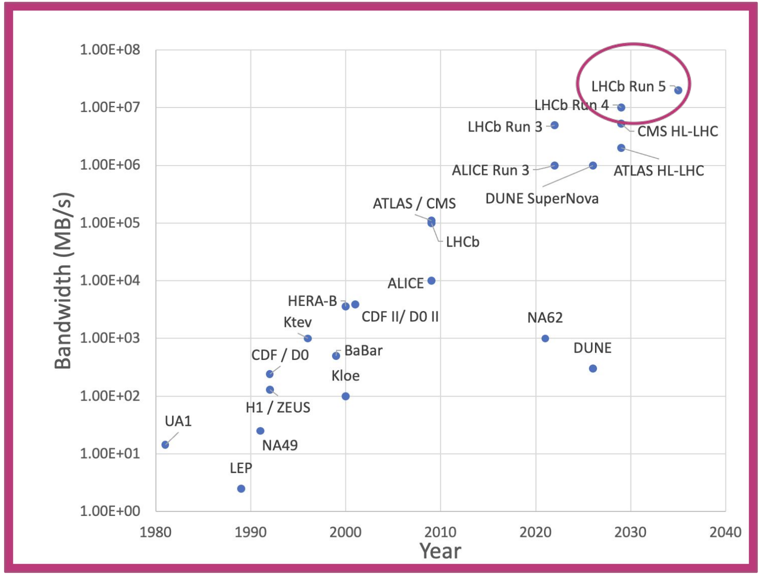


LHCb-FIGURE-2024-010

Towards the future

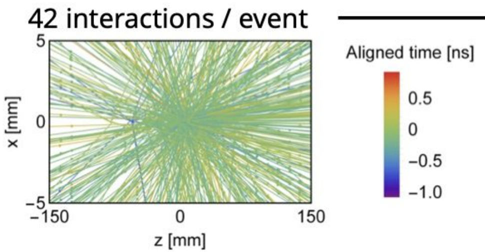
LHCb planning **Upgrade II for LS4**

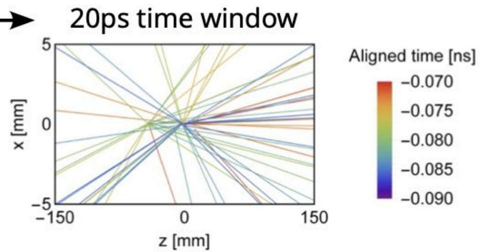
- FTDR approved in March '22 and Scoping document in preparation
- Luminosity: $(2 \times 10^{33} \rightarrow 1.5 \times 10^{34})\ cm^{-2}\ s^{-1}$
- Pile-up: $5 \rightarrow 40$
- Exciting challenges in trigger and DAQ
  - 200 TB/s of data, to be processed in real time and reduced by ~4 orders of magnitude before sending to permanent storage
  - data processing will be based around pile-up suppression
  - 4D reconstruction: timing added to tracking and ECAL detectors to better isolate signals



Courtesy of A. Cerri

The biggest data challenge in HEP!

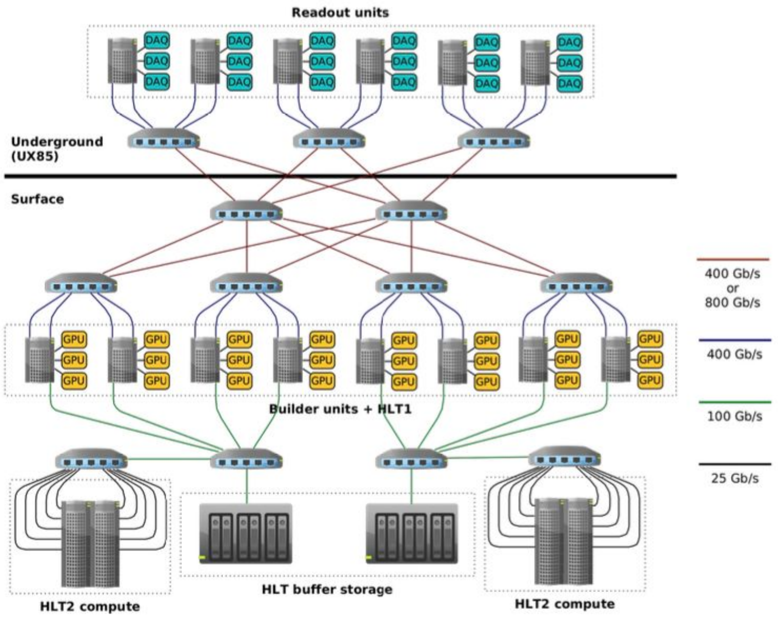


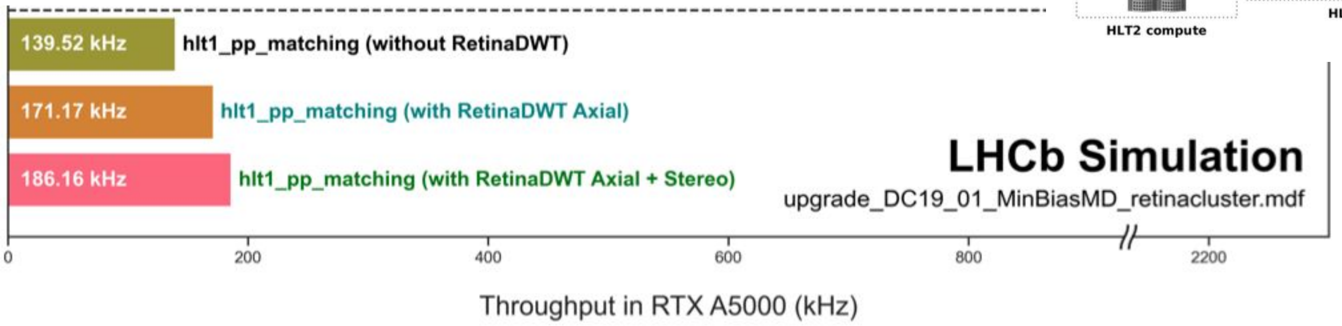


The trigger evolution: Run 5

See talk by [F. Lazzari]

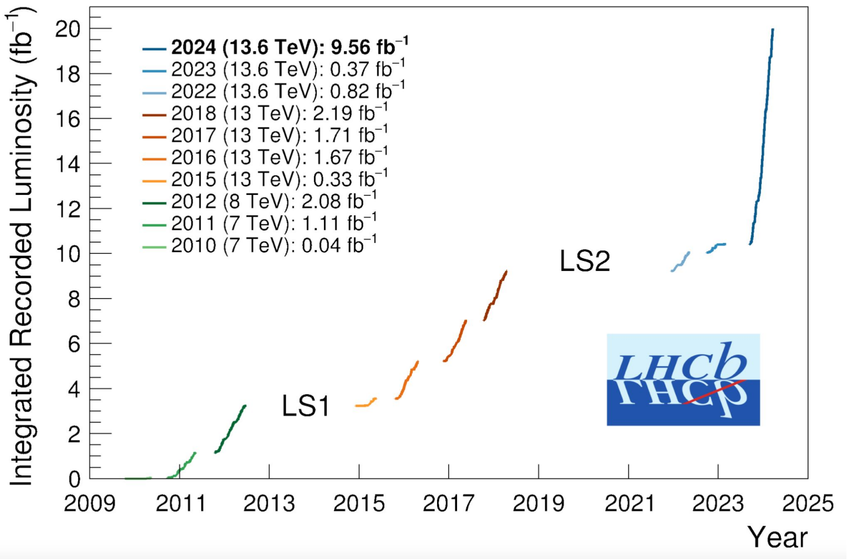
- Triggerless design philosophy will remain correct and scalable
- Partial and full detector reconstruction (and selections?) both on GPUs
- Complementary R&D activities focusing on two main areas
  - Building subdetector primitives, for example tracks or calorimeter clusters, on FPGAs [LHCb-PUB-2024-001]
  - Exploiting other architectures such as the IPU or even more exotic hardware





Conclusion

- LHCb underwent its first major upgrade in order to increase its instantaneous luminosity by x5
- Major changes in the trigger strategy:
  - Remove L0 hardware trigger, read-out full detector at 30 MHz
  - First level trigger run on GPUs
- The new trigger system has been successfully commissioned at nominal luminosity, even going beyond-expectations
- About 9.5 fb^{-1} of data have been taken and currently being analysed for a great physics outcome
- The LHCb Upgrade II is becoming a reality and this will pose very interesting challenges



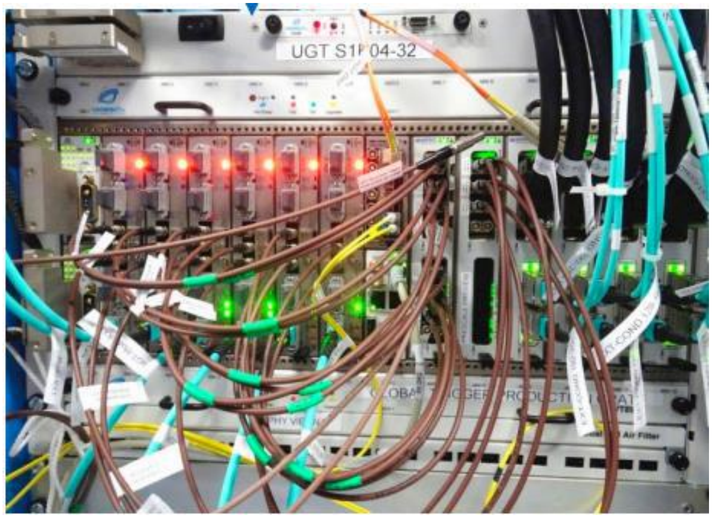
Thanks a lot for your attention!

Introduction: the CMS Trigger System

Data is selected for offline analysis 2-tiered trigger system

Level 1 Trigger (L1T)

- Hardware system run on FPGAs
- Designed to reduced rate from 40 MHz to 110 kHz
- Fixed latency of 4 μs



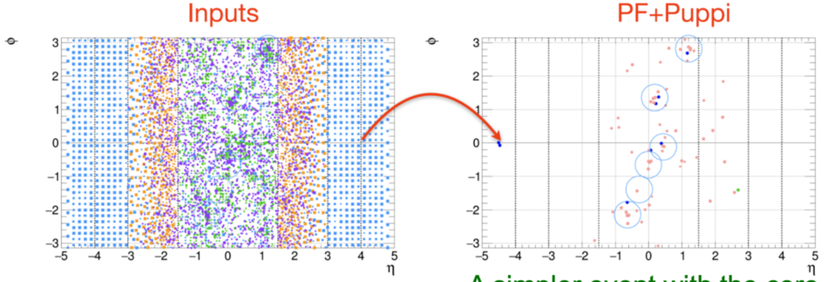
High Level Trigger (HLT)

- Software system run on CPU/GPU farm
- Designed to further reduce rate to 1-5 kHz
- Latency: 200-300 ms

Global event reconstruction (Particle-Flow) at Level-1

- Availability of tracks & high-granularity calorimetry
  - Implement global event reco @L1 and pileup mitigation
- **Challenge: can we run full PF+PUPPI at L1? YES!**
- Demonstrated a working PF+PUPPI algorithm:
  - Hugely reduces the event complexity and allows for a lot of flexibility in downstream design
  - L1 Algorithms looks like offline reconstruction
  - PF+PUPPI developed with Vivado HLS (*written by physicists + engineers*)

- What are the advantages of particle flow?



A simpler event with the core physics preserved

- Better resolution
- Ease of use → build particle level algorithms
- Reduction in bandwidth and reduction of resources

PF takes in everything

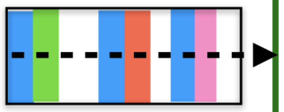
Assemble All detectors

PF is local

Assemble Things in local regions η x φ

PF Links

Particle flow combines everything together to Particls



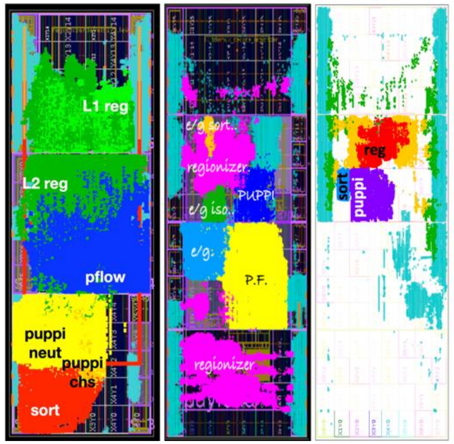
Can we run a local PU Algo?

PUPPI is local
Can parallelize it by region

Also its the CMS default PU algo

HW demostrator

Barrel *APx VU9P* | **Endcap** *Serenity VU13P* | **Forward** *Serenity VU13P*

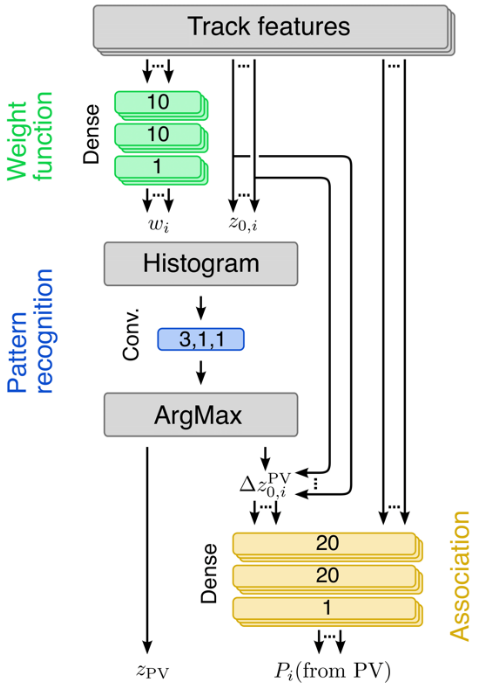


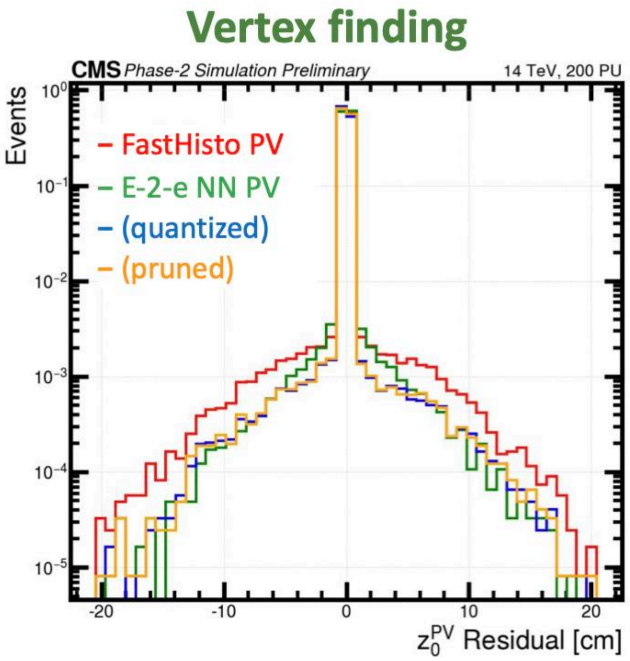
Universidad de Oviedo

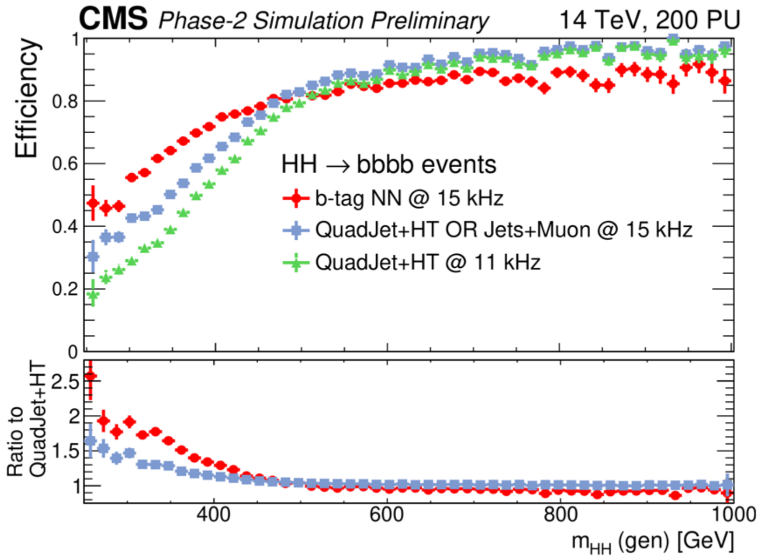
Recent development highlights (with ML)

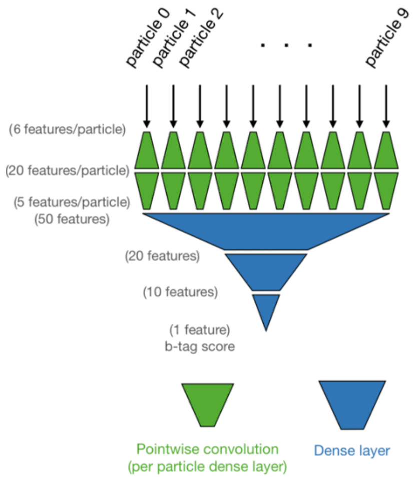
- **NN Vertex Finding:**
  - Combination of dense BDTs and CNN to perform Vertex Finding and Track-to-Vertex association
  - Firmware quantised and pruned to fit within FPGA
  - ***Improved performance wrt to baseline (reduction in the tails of the residual by 50%)***

- **b-tagging:**
  - Training NN to ID jets from b-quarks
  - Runs on PUPPI particles within each jet and discriminate between b-quark jets and those from light quarks and gluons



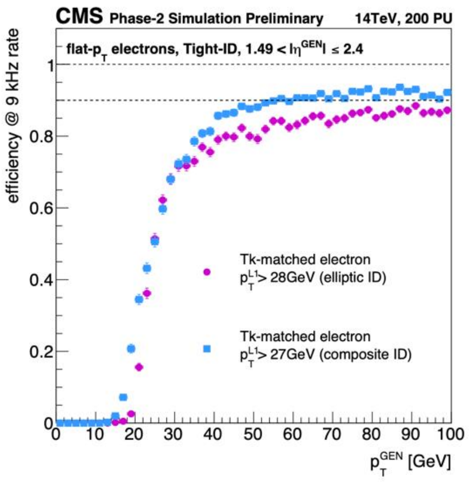


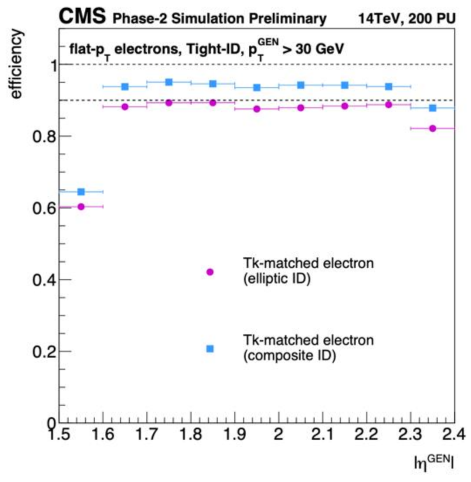




Recent development highlights (with ML)

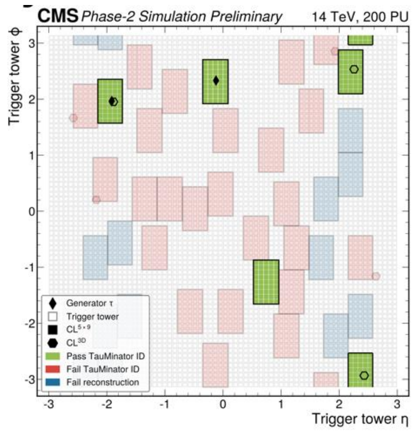
- **Electron-ID**
  - New Composite-ID, combines information about tracks and clusters in the HGCAL into a single model for matching and identification
  - A single BDT model: controlling the identification of track and calorimeter deposit and the tightness of the matching. ▸ 10% more efficiency for the same rate

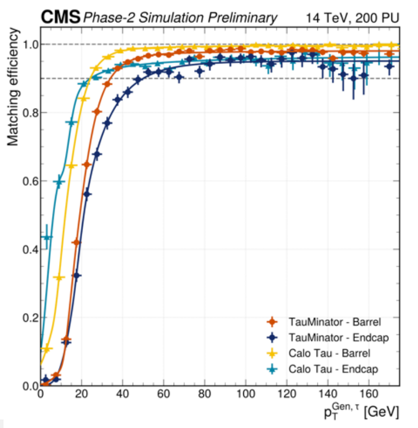


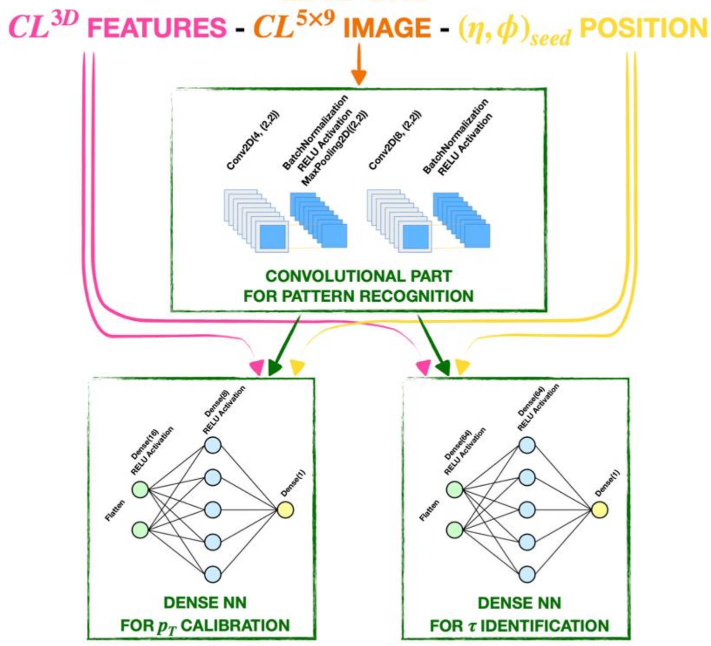


	LUT	FF	BRAM	DSP
e/γ IP	3.1%	0.4%	0.0%	1.6%
Total	24.4%	17.6%	29.5%	14.3%

- **Tau reconstruction: Tauminator**
  - Training dedicated CNN to reconstruct and identify Tau-induced signal in calorimeters (5x9)
  - Elegant way to deal with different geometries in Barrel (Crystals) and EndCap (HGCAL 3D clusters).

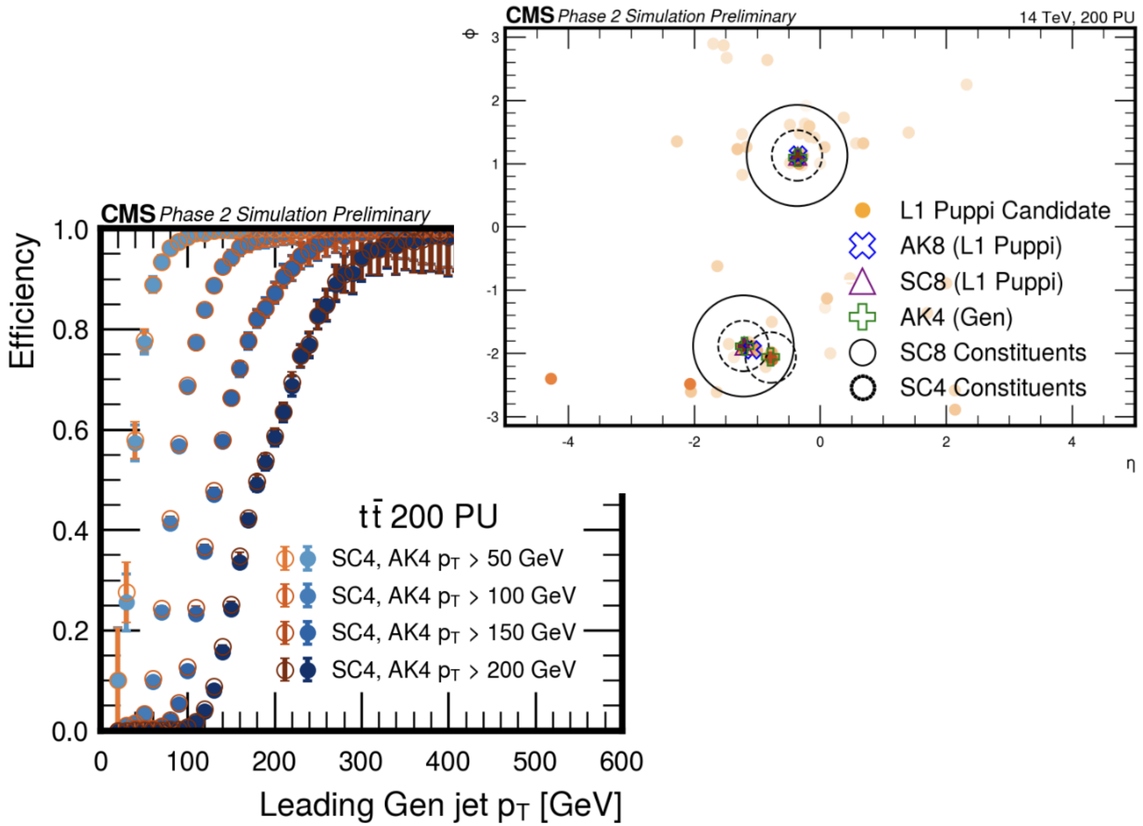




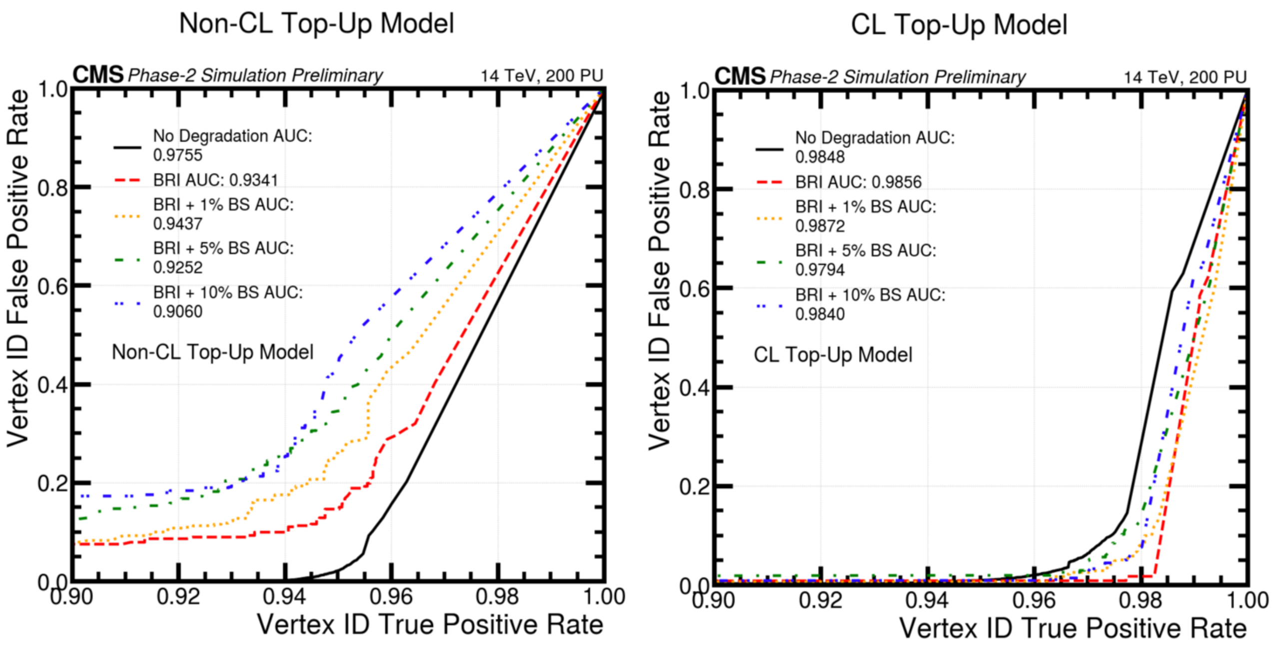


Recent development highlights (with ML)

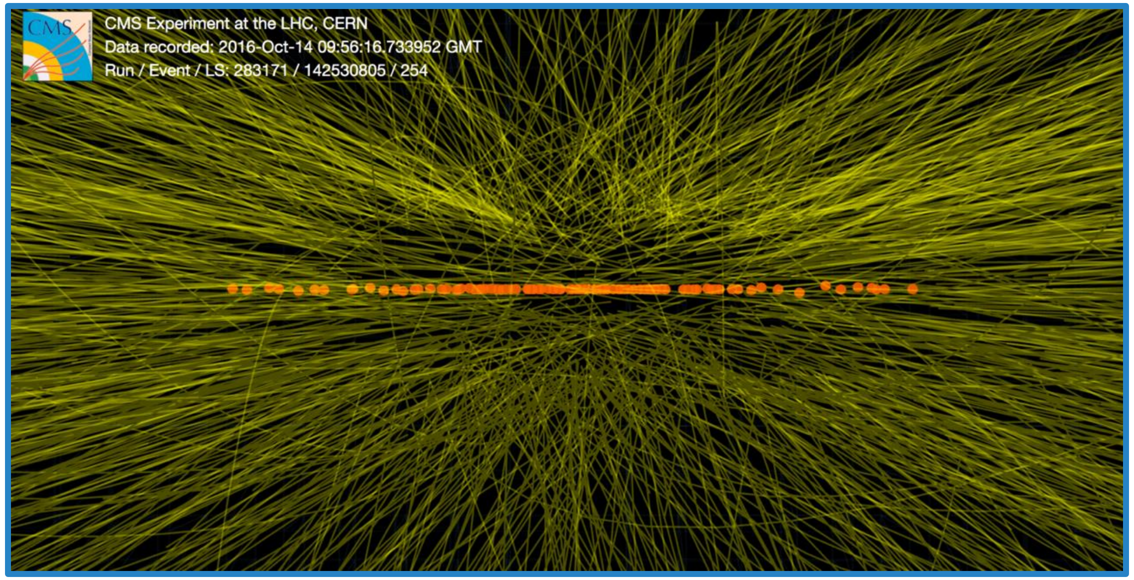
- **SeededConeJets:**
  - Jet finding based on PF candidates
  - Iterative approach computing distance between each particle and jet radius (SC4 or 8), compute jet axis and energy.
  - Jet matching anti-kt jets



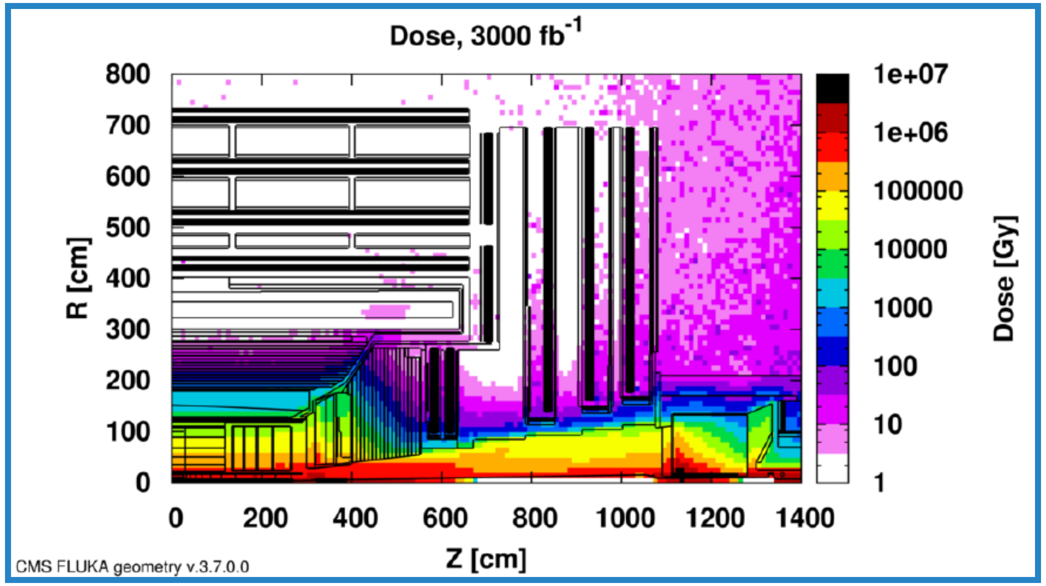
- **Continual learning:**
  - Elegant way to deal with changing detector conditions (ageing, noise, LHC interfill, etc.)
  - Train a model with a continuous stream of data. Learns from a sequence of partial experiences rather than all the data at once.
  - Update model to changing conditions without large MC production.
  - Method tested on Vtx reconstruction



HL-LHC: challenges



- **Expected pileup** (PU): ~140 (nominal HL-LHC lumi)
- Motivates/requires:
  - Improved granularity wherever possible
  - Novel approaches to in-time Pile Up mitigation: Precision Timing detectors (30ps)
  - A complete renovation of the Trigger and DAQ systems for better selectiveness, despite the high PU.

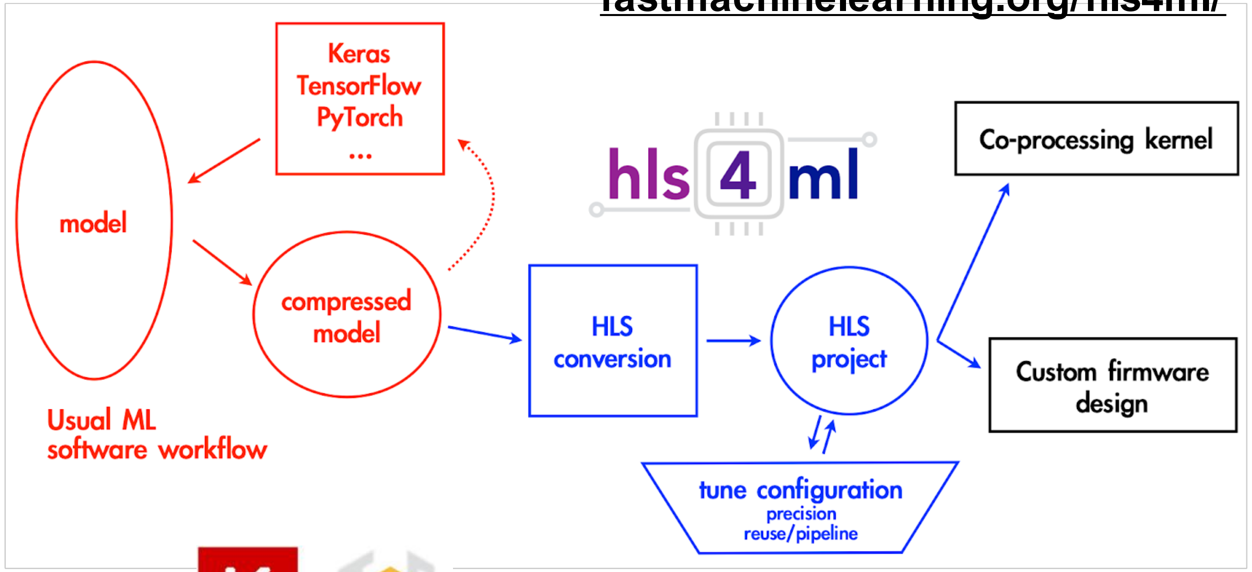


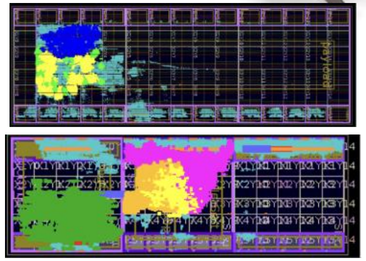
- **Radiation damage / accumulated dose** in detectors and on-board electronics may result in a progressive degradation of the performance.
- Maintain detector performance in harsh conditions:
  - The complete replacement of the Tracker and Endcap Calorimeter systems.
  - Major electronics overhaul and consolidation of the Barrel Calorimeters and Muon systems

From ML to FPGA

high level synthesis for machine learning

fastmachinelearning.org/hls4ml/





(Q) K + TensorFlow

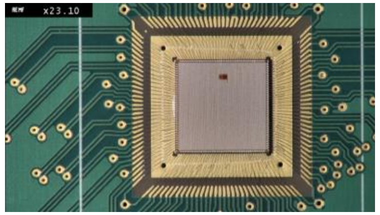
(Q) ONNX

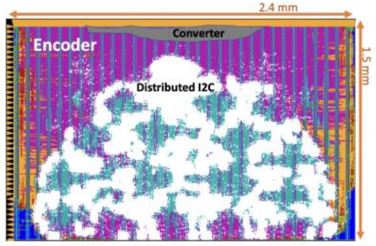
PYTORCH

Vivado™ HLS

HLS COMPILER

Mentor Catapult A Siemens Business
Coming soon





11 Sep. 2023 Fast ML - Sioni Summers 4

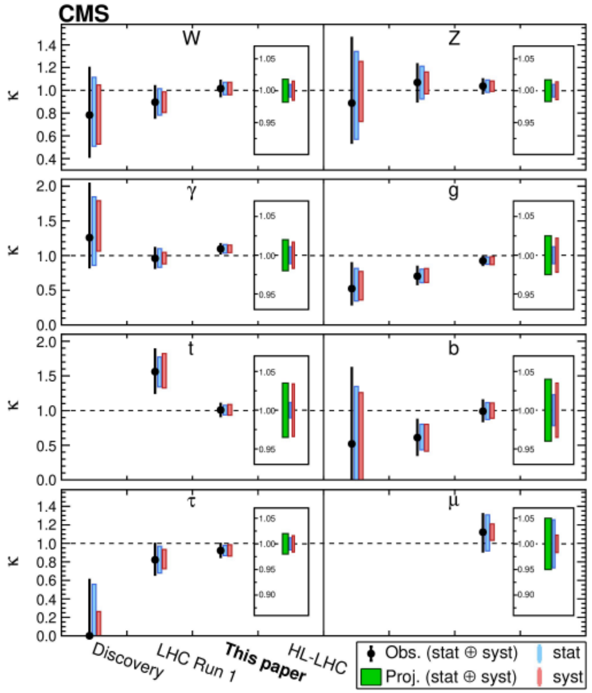
The Phase-2 Trigger Upgrade: Physics case

Improve precision of SM tests (*i.e. Higgs couplings*, m_W)

Target unobserved SM processes (*i.e.* $H \rightarrow HH; H \rightarrow cc$)

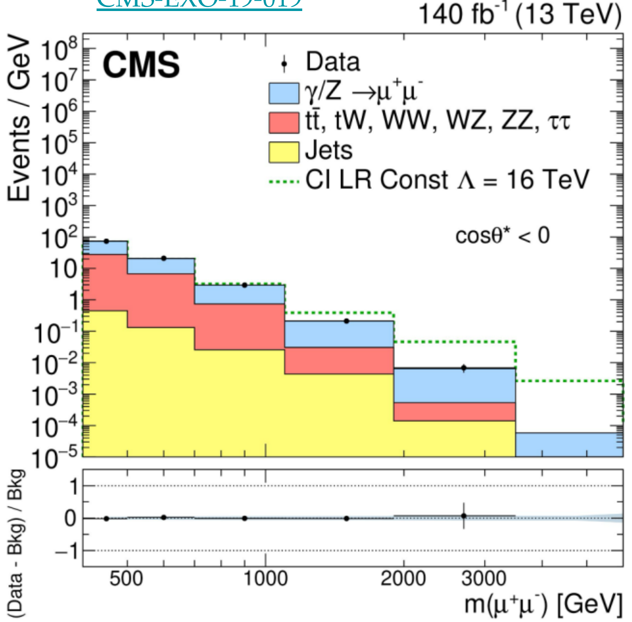
Search for deviations at high momenta (*i.e. Effective Field Theories*)

Probe new phase space (*i.e. Long-lived particles*)

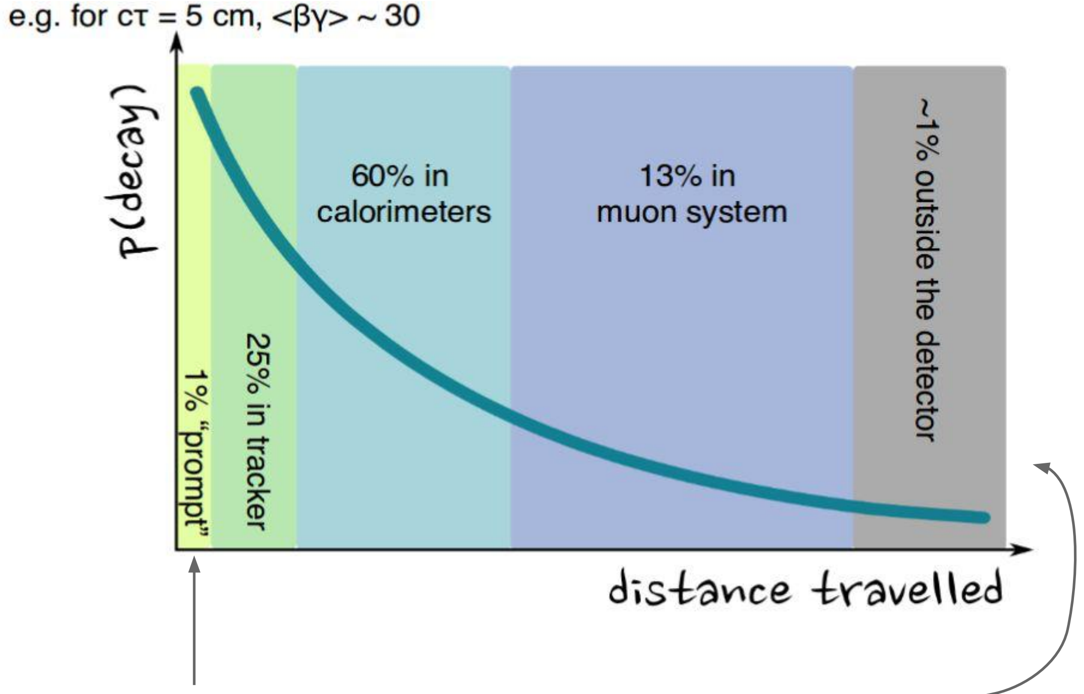


Nature 607, 60–68 (2022)

CMS-EXO-19-019



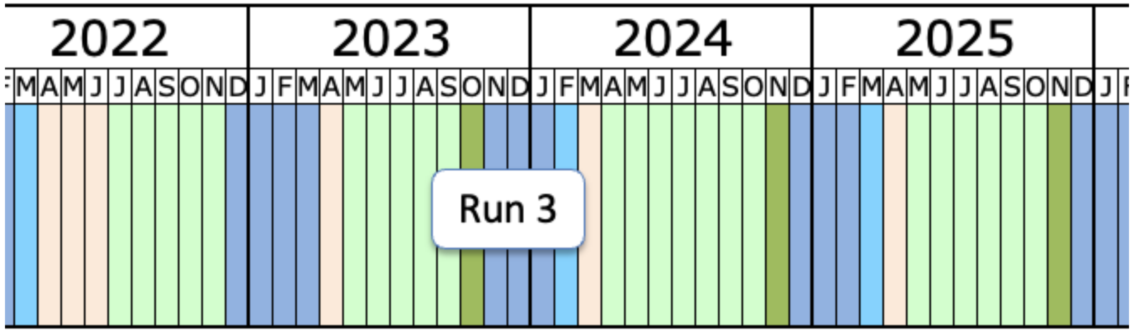
$$\mathcal{L}_{eff} = \mathcal{L}_{SM}^{(4)} + \sum \frac{C_x}{\Lambda^2} O_{6,x} + h.c.$$

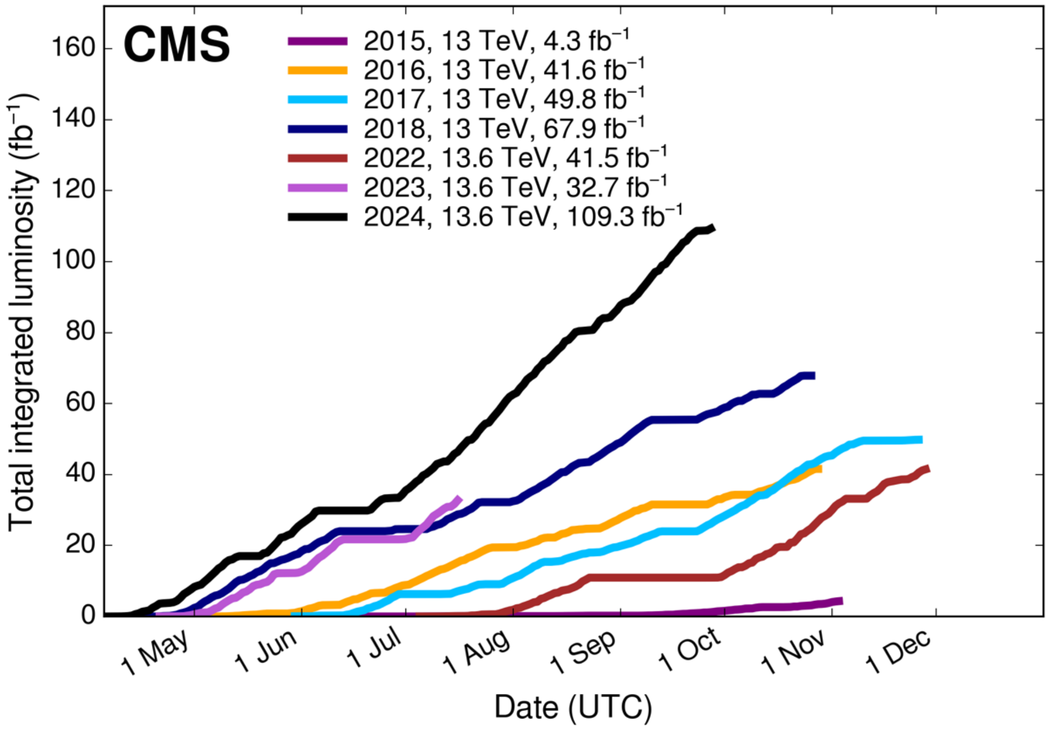


Bulk of searches

Run 3 at a glimpse

- With almost one and half year to go, Run-3 has already surpassed Run-2 luminosity
  - **Almost 170 pb^{-1} recorded**
- **New strategies** have been deployed both at L1T and HLT
- Excellent opportunity to extend physics reach and try new ideas to guide our path in the future
  - New capabilities to trigger on long-lived particles
  - Anomaly detection
  - Triggerless readout (scouting)
  - Increased GPUs usage
  - Extensive use of ML techniques

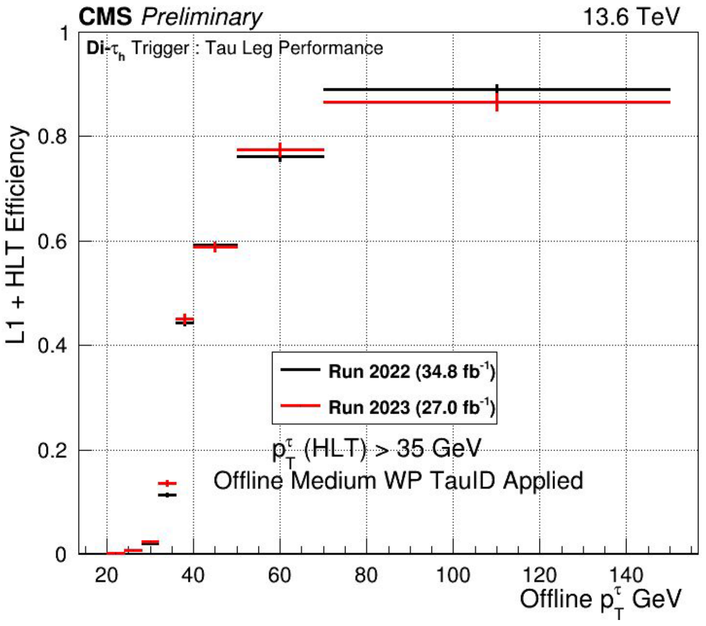




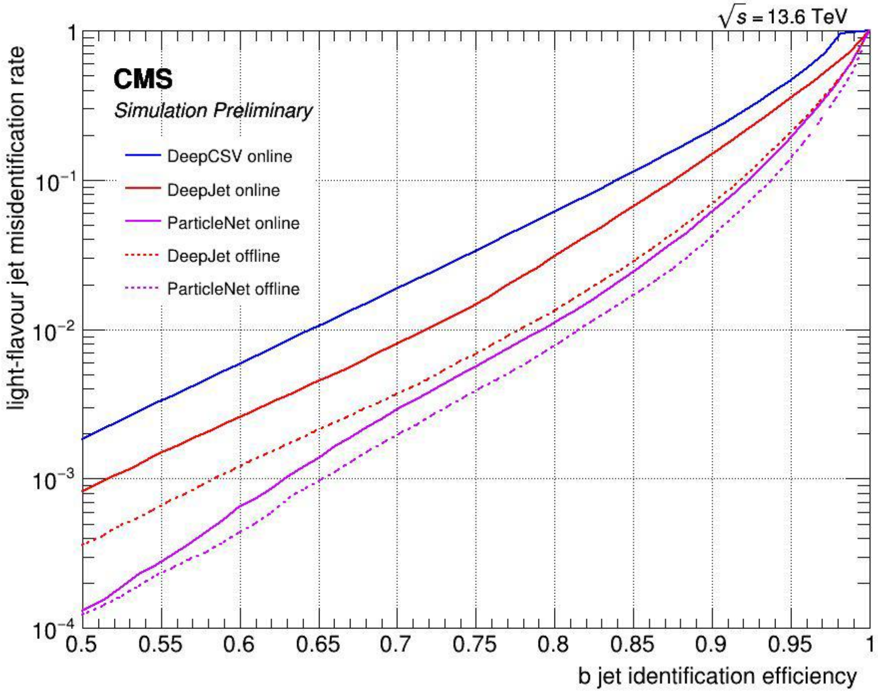
ML at HLT

- Tau @HLT
  - Reconstruction: Hadron plus strip
  - DeepTau identification: CNN+DNN based tagger
- ParticleNet b-jet tagger @HLT. GNN-based
  - Jets treated as a permutation-invariant point cloud
  - Performance gain, especially for HH processes

CMS DP-2024/042

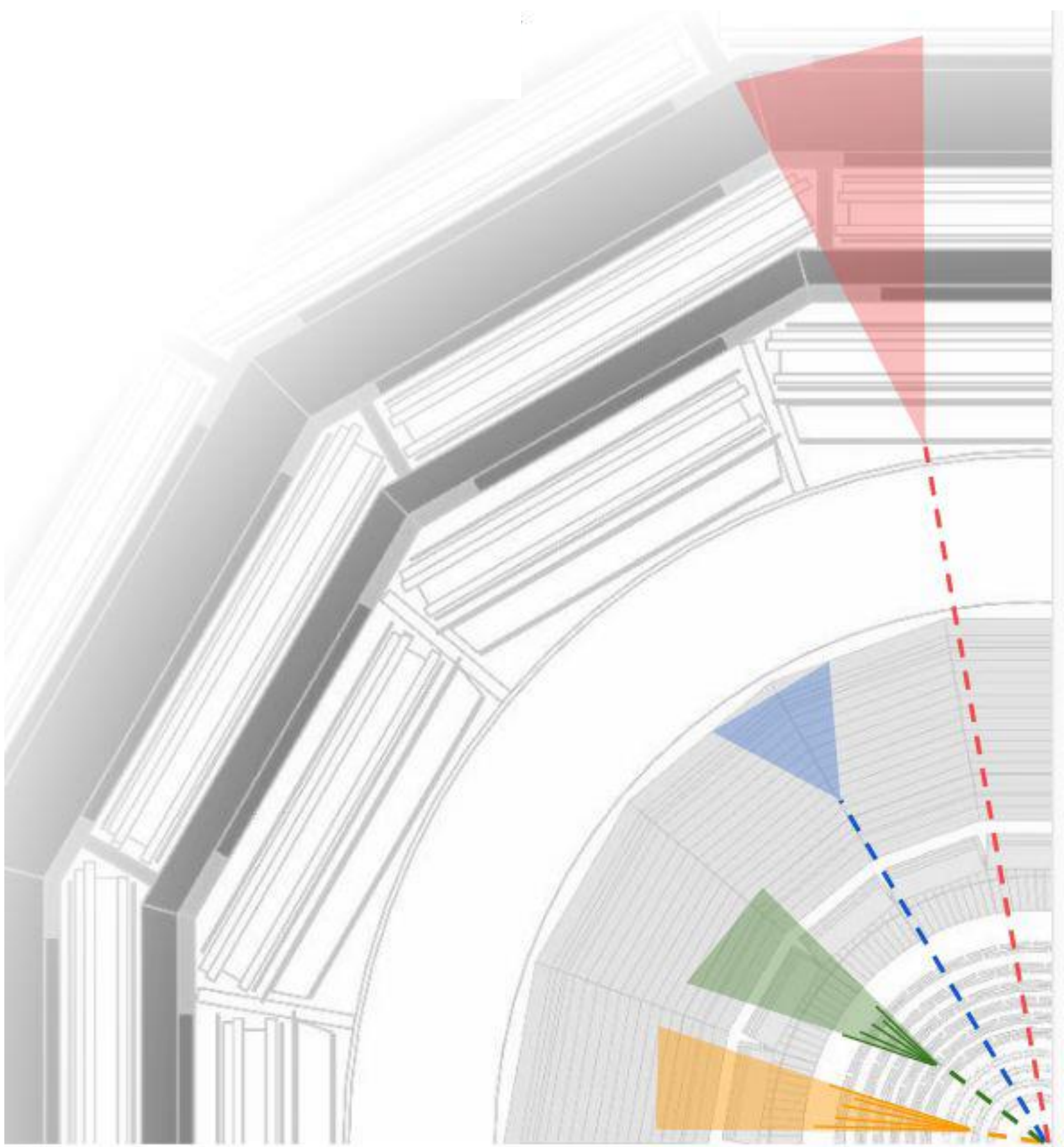


CMS DP-2023/021

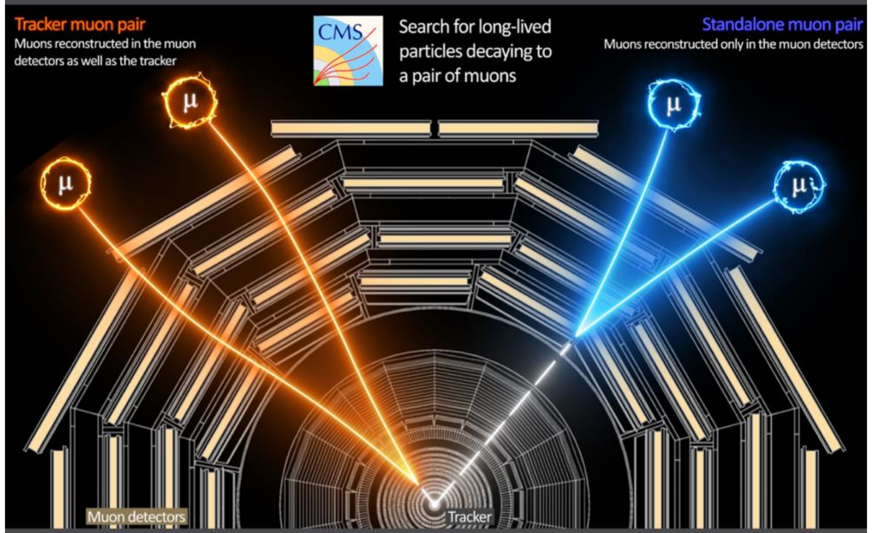


Long-lived particle triggers

- Many models predict the existence of **long-lived particles** (LLPs)
  - Many Exotic scenarios not envisioned when the trigger system was being designed!
- LLPs transit layers at later times, timing information
- LLPs decay far from the interaction point and show displaced signatures
  - Dedicated trigger paths exploiting unique features
  - Displaced jets in the tracker, calorimeters, or muon systems
- Strategies adopted mainly at HLT for Run 3
  - Some ideas already at L1
- Run 3 is the perfect benchmark for "crazy" ideas for HL-LHC



First Run 3 search: displaced dimuons at 13.6 TeV


The CMS collaboration at CERN presents its latest search for new exotic particles

The CMS experiment has presented its first search for new physics using data from Run 3 of the Large Hadron Collider. The new study looks at the possibility of "dark photon" production in the decay of Higgs bosons in the detector. Dark photons are exotic long-lived particles: "long-lived" because they have an average lifetime of more than a tenth of a billionth of a second – a very long lifetime in terms of particles produced in the LHC – and "exotic" because they

https://cms.cern/news/long-lived-particles-light-lhc-run-3-data

https://home.cern/news/news/physics/cms-collaboration-cern-presents-its-latest-search-new-exotic-particles

With a strong Spanish contribution:

Universidad de Oviedo


IFCA
Instituto de Física de Cantabria

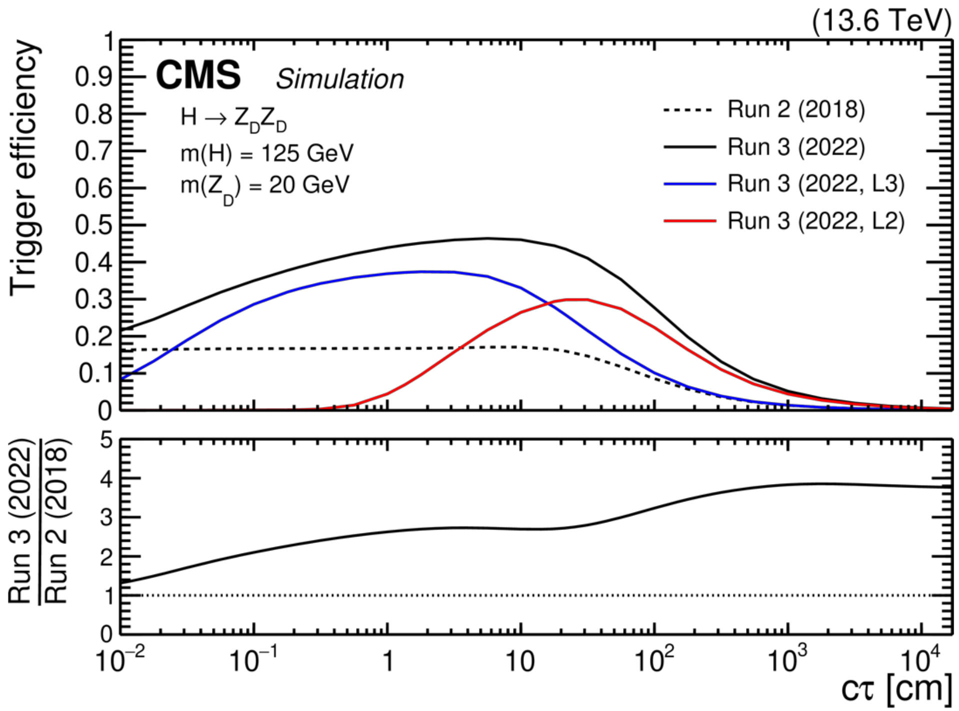

Ciemat
Centro de Investigaciones Energéticas, Medioambientales y Tecnológicas

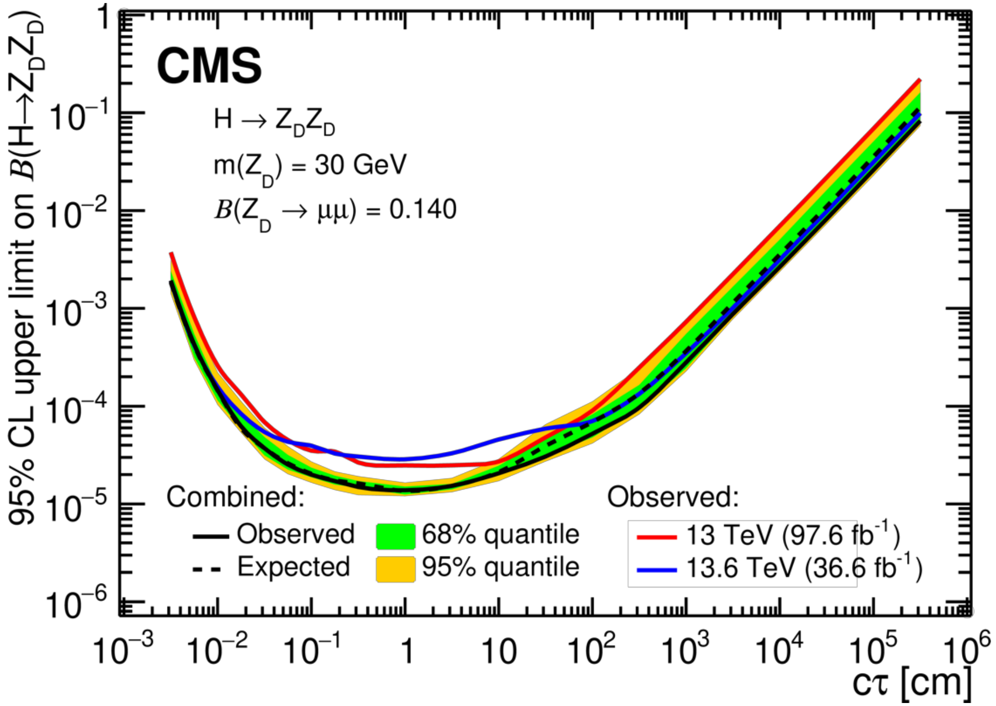


JHEP 05 (2024) 047

Displaced dimuons at 13.6 TeV. New triggers

- Use the 2022 dataset (**36.7 fb^{-1}**) recorded with **new LLP triggers** with thresholds **down to $p_T(\mu)$ > 10 GeV**
  - Re-optimized L1 triggers, including p_T without beam spot constraint, and new reconstruction algorithms.
  - Use d_{xy} information at trigger level to control the background rate.
- **Factor 2-4 more signal efficiency**
- Despite **2.5 smaller dataset**, comparable (or better) sensitivity w.r.t. 13 TeV result.

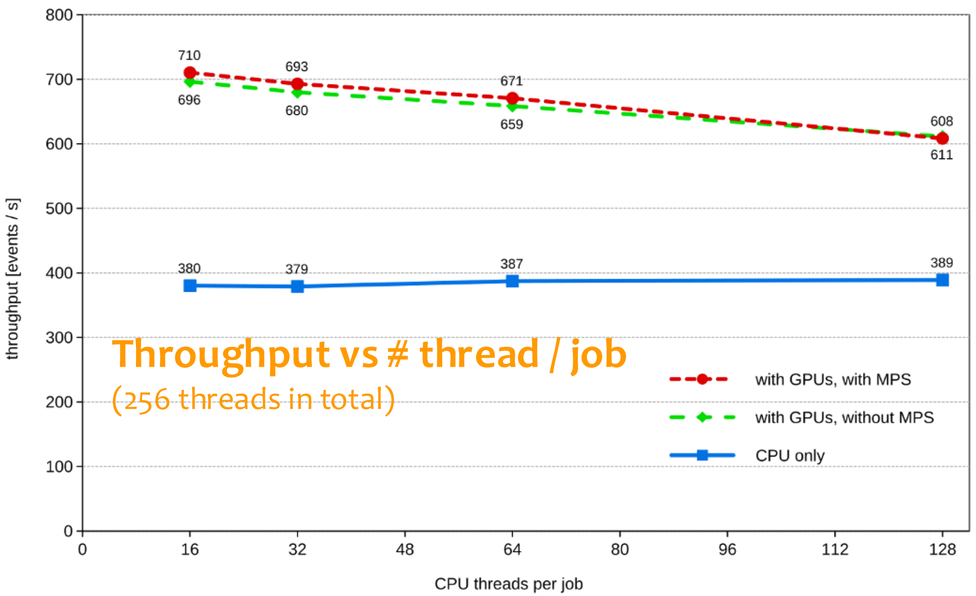




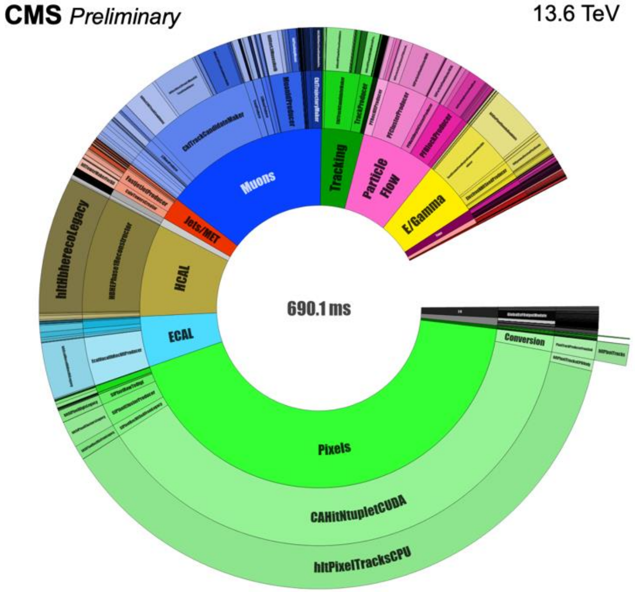
Multithreading and GPUs

- Multithreading (MT) is key to fully exploit HLT farm computational power
  - inter-event, intra-event, in-algorithm parallelism;
  - usage of "data handles" to define the data dependency among modules;
  - lower memory usage

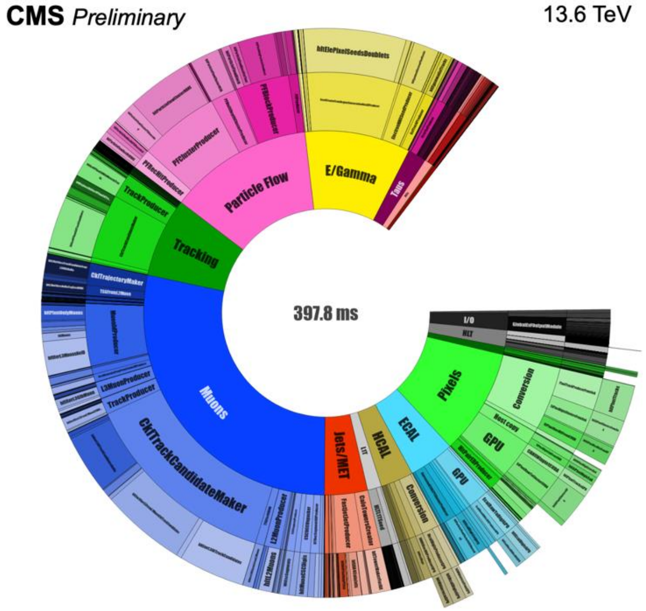
- CMS HLT farm heterogeneous since 2022 (AMD CPU + Nvidia T4):
  - 40% of HLT reconstruction ported to GPU
    - Pixel local reconstruction
    - Pixel tracking and vertexes
    - ECAL local reconstruction
    - HCAL local reconstruction



No GPUs 32 threads, 24 streams



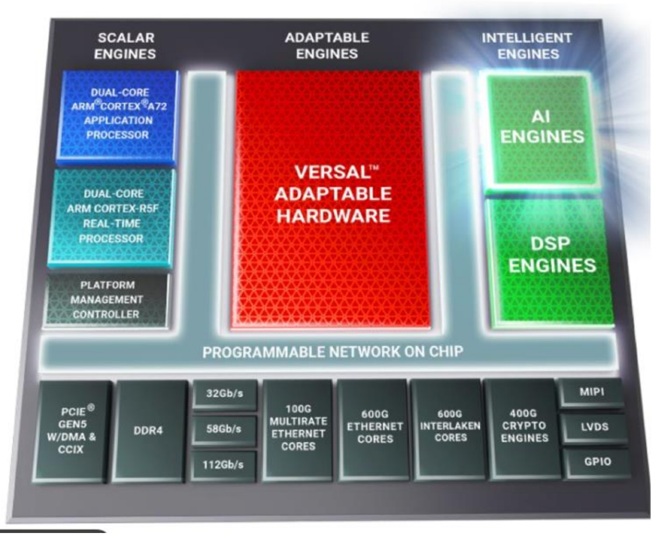
GPUs (no MPS) 32 threads, 24 streams

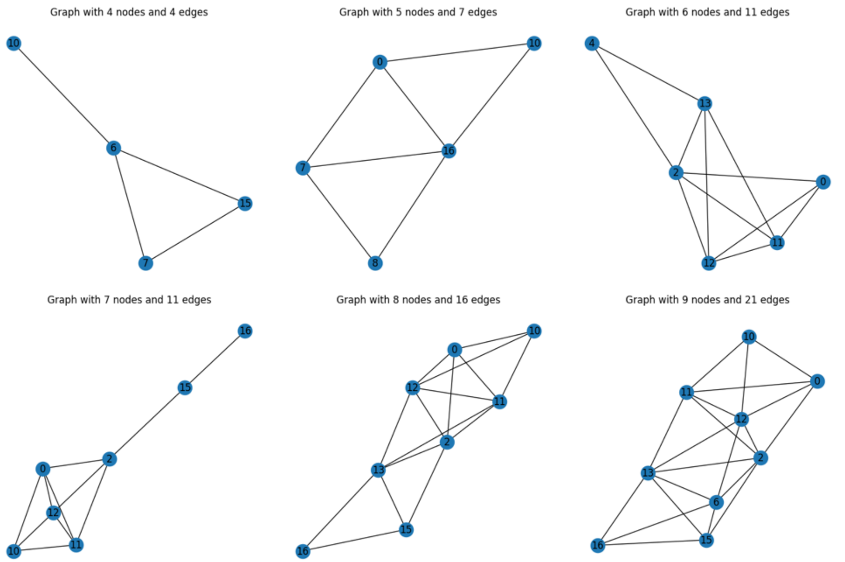


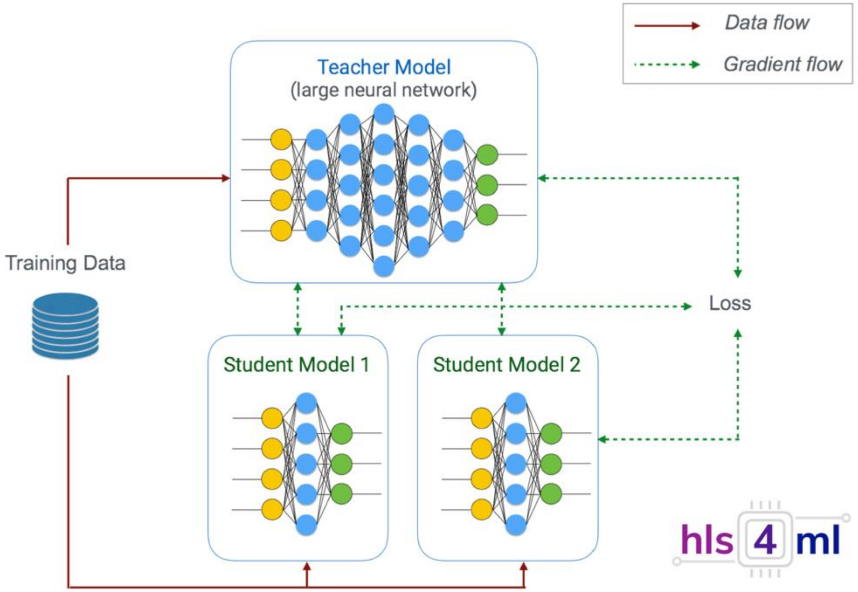
INTREPID project

INnovativeTRiggEr techniques for beyond the standard model PhysIcs Discovery at the LHC

- Improve muon trigger reconstruction with advance techniques based on machine learning: Graph Neural Network
  - Work already started with the overlap muon track-finder, first version of the network, using every detector layer as a nodes and $\Delta\phi$ and $\Delta\eta$ as edge parameters
- Considering AI accelerators (AI Xilinx Versal Chip)
  - Provide the necessary throughput and latency for triggering?







The Next-Generation Trigger Project

Innovative computing technologies for data acquisition and processing for the HL-LHC and beyond

- Enhance the triggers and the data collection and processing, and thus the scientific potential, of ATLAS and CMS in the HL-LHC phase **beyond the currently projected scope**.
  - Accelerate the evaluation and introduction of novel computing, engineering and scientific ideas already with demonstrators for Run3, but with main focus on HL-LHC
  - Provide a major push to the work already ongoing in the experiments, by enabling lines of research **currently not feasible within existing financial, human and technology constraints**
  - Provide **critical insight to develop data flows** for the even more ambitious objectives of a future collider, such as the Future Circular Collider (FCC) currently in its Feasibility Study phase
- CERN involvement to **ensure that other current & future CERN experiments benefit from the results** in terms of computing frameworks and theoretical modelling.
- All project results (IP) will belong to CERN and will be released under a valid open policy and IP generated will be released under appropriate open licenses in compliance with the **CERN Open Science Policy**.

NexTGen
Next Generation Triggers

https://nextgentriggers.web.cern.ch/