

The NextGen Triggers Project

Concept and initial activities

Axel Naumann on behalf of the NextGen Project Management Committee

CHEP 2024



NexTGen
Next Generation Triggers

IMPROVE PHYSICS REACH

THROUGH TRIGGERS



Efficiency: more relevant data given data rate



More cycles / seconds means smarter selection



Exploring trigger updates for new physics sensitivity, with improved models, generators, and simulations



The Project

NEXT GENERATION TRIGGERS



Five years: 2024-2028 enabled by external donation, combining

- ATLAS, CMS; limited participation of ALICE, LHCb
- CERN's Theory & IT departments
- CERN's Exp Physics Software group



Project goals ([proposal](#))

- opportunity for wider R&D
- improve LHC experiments in 2028+
- invest in community



HOW

Common R&D + Training

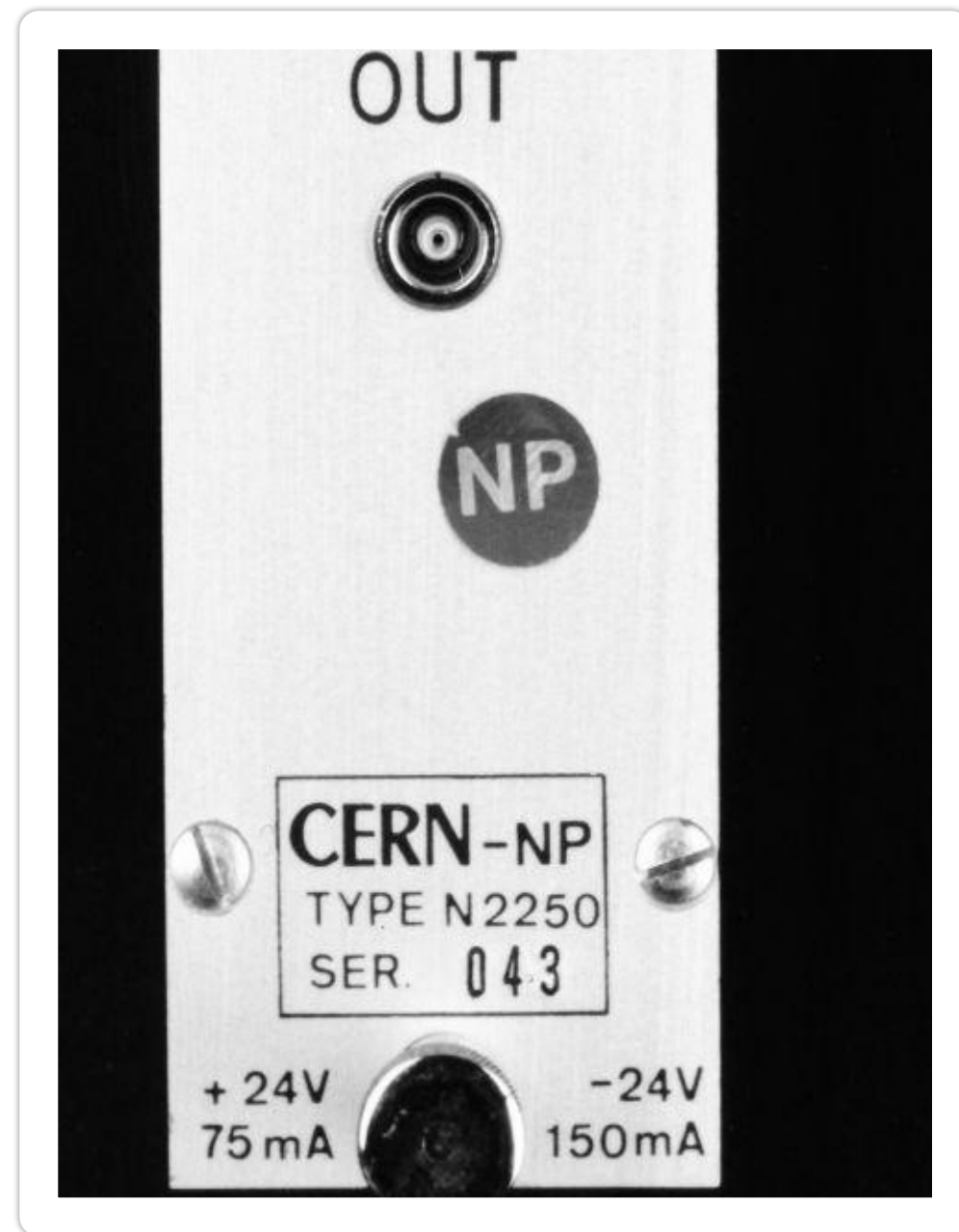
- ▶ Combining all parties: 2 (+2) experiments, IT, theory, experimental physics software

Experiment-specific R&D

- ▶ ATLAS+CMS define their R&D requirements
- ▶ Benefitting from common R&D, training

Results are open

- ▶ Open access, open source, including training
- ▶ Embedded in experiments



THE KEY OBJECTIVES

More than technical work



Technologies

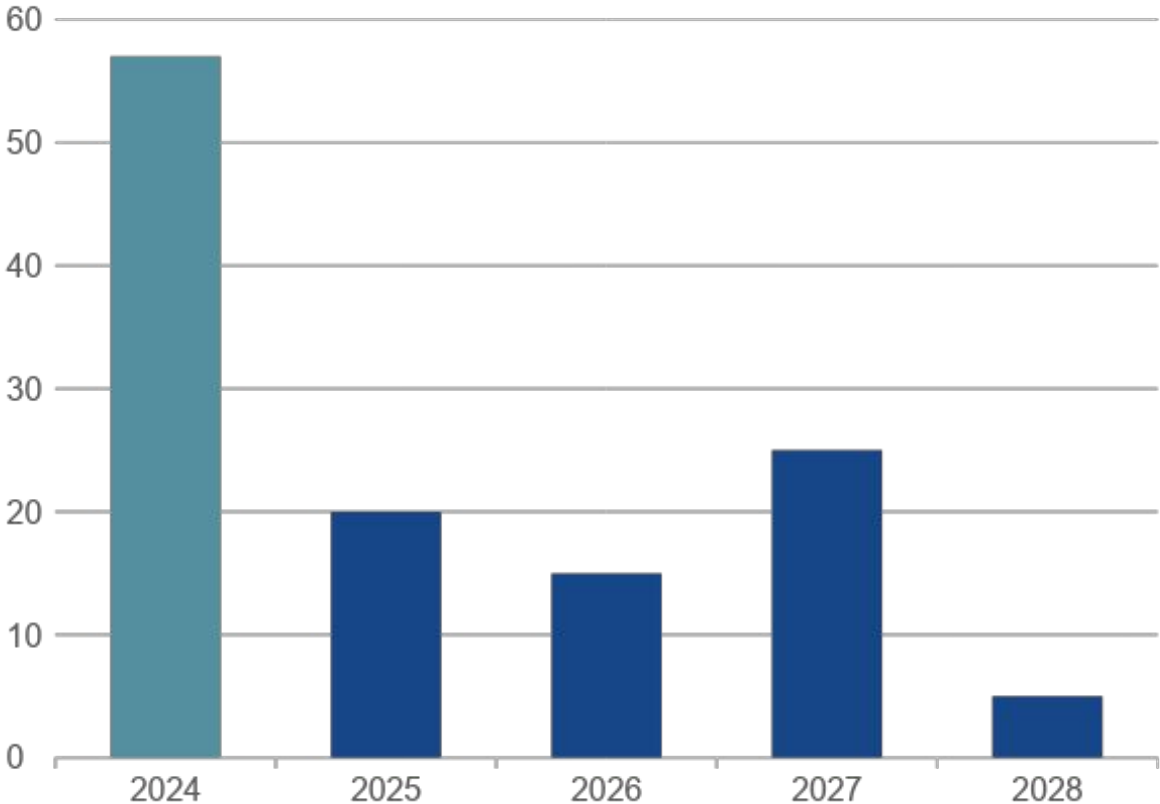
- ▶ ML and classical algos: invent, optimize, benchmark
- ▶ Deploy on FPGAs: latency + throughput
- ▶ Employ GPUs, with focus on open source



Community

- ▶ Training to stay
- ▶ Educating the next generation
- ▶ Invest in open source to reduce vendor lock-in

Results? HIRES!



Target effort

- ▶ ~280 FTEs planned in total across 5 years
- ▶ 57 new hires for 2024
- ▶ Mixed profiles: master, PhD, post-doc, staff

NextGen Workshops

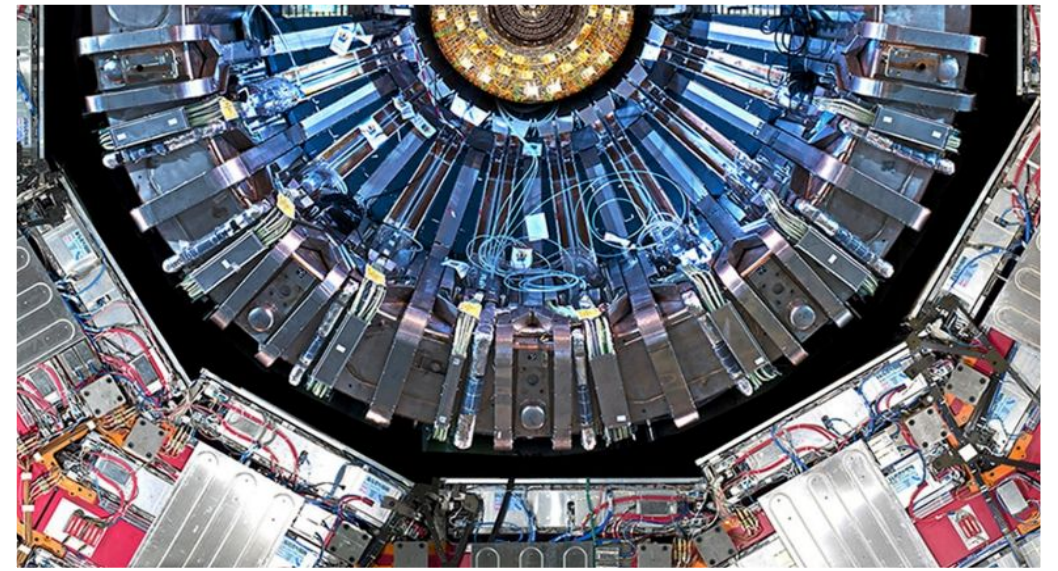
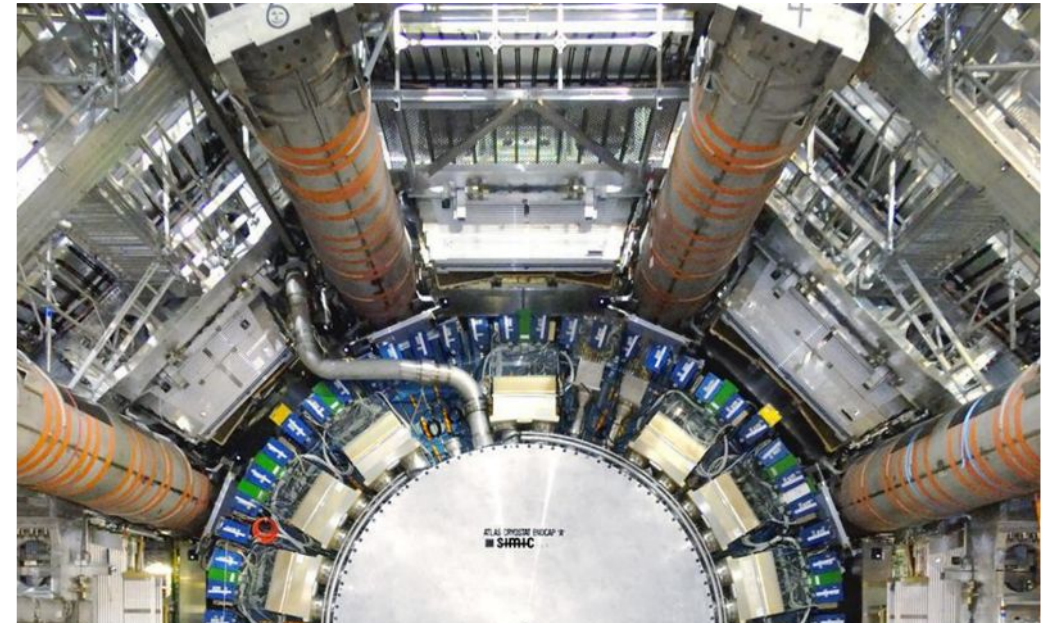
1st Technical Workshop

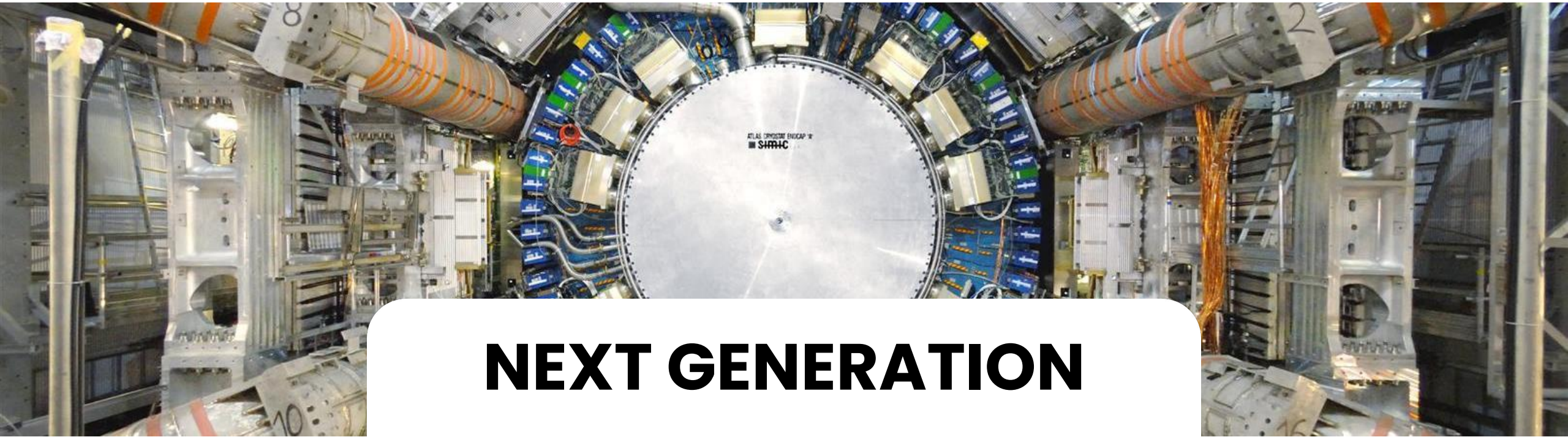
- ▶ Discussing technical progress across tasks
- ▶ November 25-27, 2024 @ CERN
- ▶ <https://indico.cern.ch/event/1421629>

Topical Workshop Examples

Bringing the tasks' communities together

- ▶ [NGT Algorithm Workshop](#) – Lattice QCD at large scale on exascale computing facilities
- ▶ [hls4ml HEP Community Forum](#)





NEXT GENERATION TRIGGERS

R&D Topics

PROGRESS YOU MIGHT CARE ABOUT

WP1: Main goals

Overarching rationale

- ▶ Providing constituent R&D to NextGen themes
- ▶ Interleaving experiments, IT, theory, software engineering, ML, accelerators

Infra O(100) GPUs, including wide range of GPUs (vendors, specs) for benchmarking workflows

ML Optimization of ML models for inference hardware; offering training + optimization service

Lattice-quantum-field theory with HPC optimization and novel algos (ML, quantum-inspired)

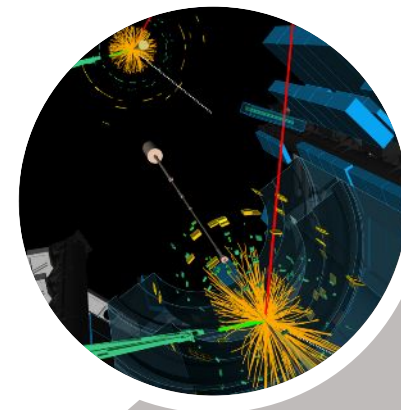
Generators Wider physics reach, applicability to experiments' triggers

Heterogeneous compute Novel approaches for accelerators, reducing vendor lock-in

EXAMPLE: OPTIMIZING MEMORY LAYOUT FOR INFERENCE AND ACCELERATOR USE

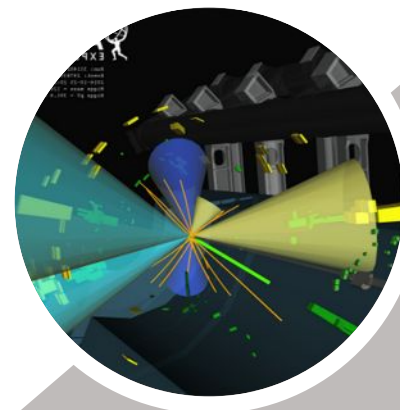
Array-of-structs
`vec<jet{pT, eta, phi}>`

Struct-of-arrays
`vec<pT>, vec<eta>, vec<phi>`



All LHC experiments implement this
transformation - differently!

[Rely on C++ reflection?](#)



Decouple experiments' frameworks
from vendors' inference libraries,
optimized for **inference from GPU**



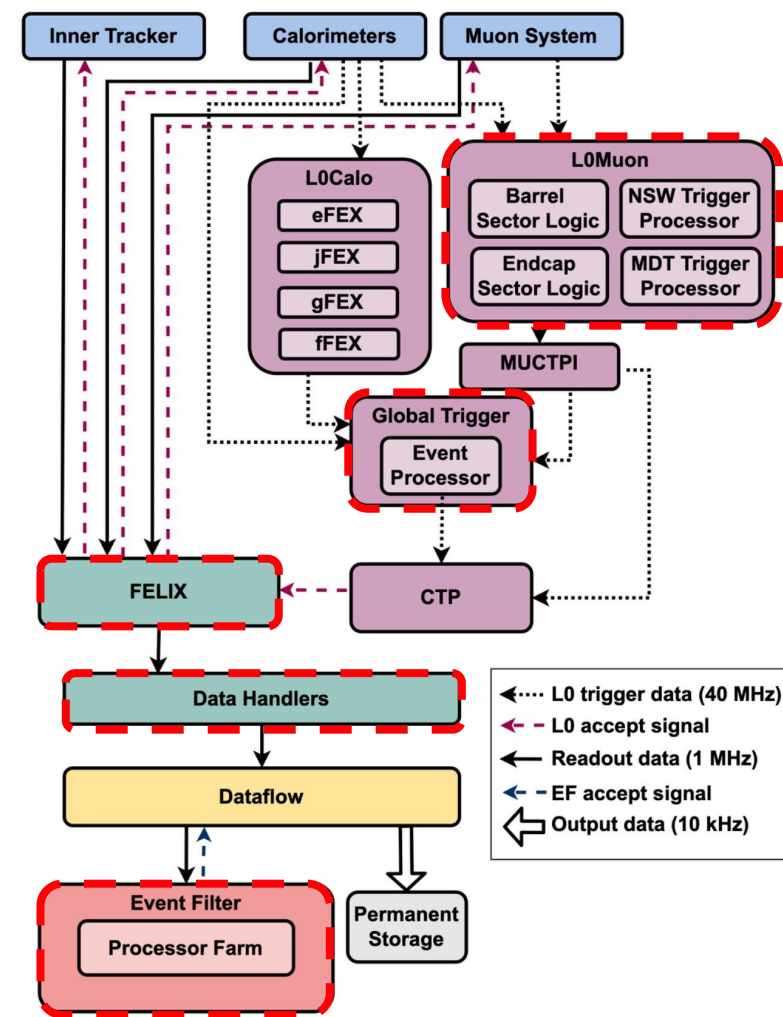
WP2: Main goals

Diagram: overall T-DAQ architecture for Phase-II

- 1 MHz L0 rate
- 4.6 TB/s to Event filter
- Event filter data reduction 1 MHz -> 10 kHz

Main goal for the ATLAS NextGen work package

- ATLAS R&D, novel (AI) approaches and innovation for several sub-systems (dotted boxes)
- Level-0 hardware trigger: L0Muon and Global Trigger
- Event Filter event processing: Track reconstruction in the Inner Tracker and Muon detectors, plus ACTS tracking software infrastructure
- Novel trigger signatures and physics optimisation



NextGen R&D

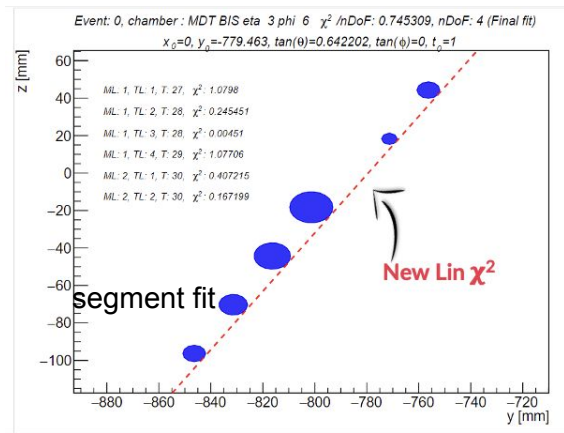
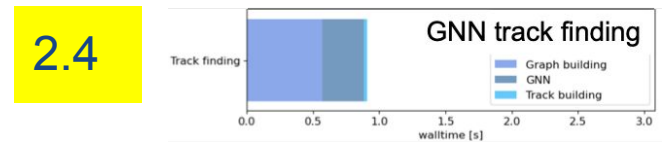
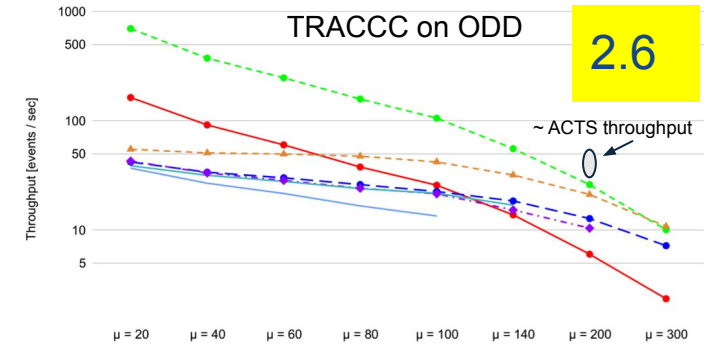
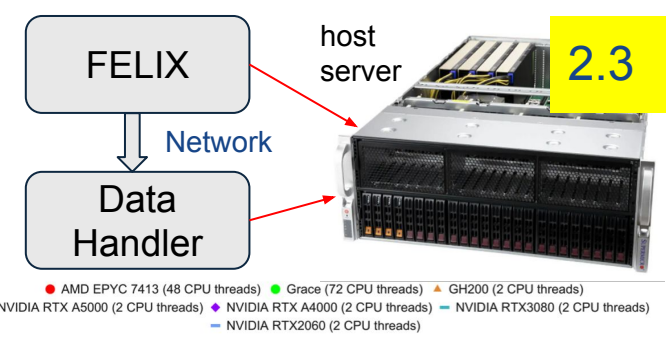
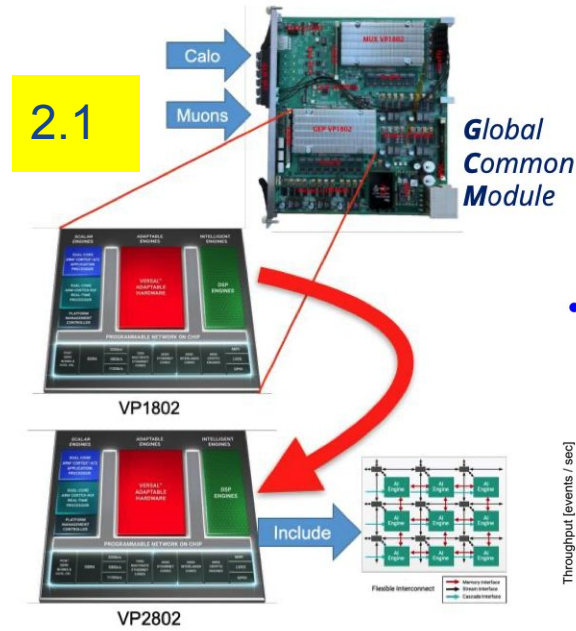
WP2 - Highlights

- Work is embedded in ATLAS T/DAQ and offline groups
- Building on ongoing ATLAS work, first new ideas and initiatives
- Actively collaborating with WP1 and WP4
- Benefit from HW provided by WP1/IT, CERN openlab and ATLAS test-beds

Some examples of ongoing work and new initiatives

- Plan to procure L0Global board with novel FPGA (VP2802) with AI engine (2.1)
- Studies of ML based L0Muon trigger techniques started (2.2)
- First developments on readout, purchased prototype host server (2.3)
- Fast ITk tracking optimisation, GNN4ITk tuning and throughput tests (2.4)
- Novel ACTS Muon segment pattern and fitting (2.5)
- Enabling ACTS EF reconstruction, first results on full TRACCC chain (2.6)
- Developing trigger analysis kit, ATLAS survey to collect input (2.7)
- Study of L1 trackless b-tagging for ITk inner pixel replacement (2.4+7)

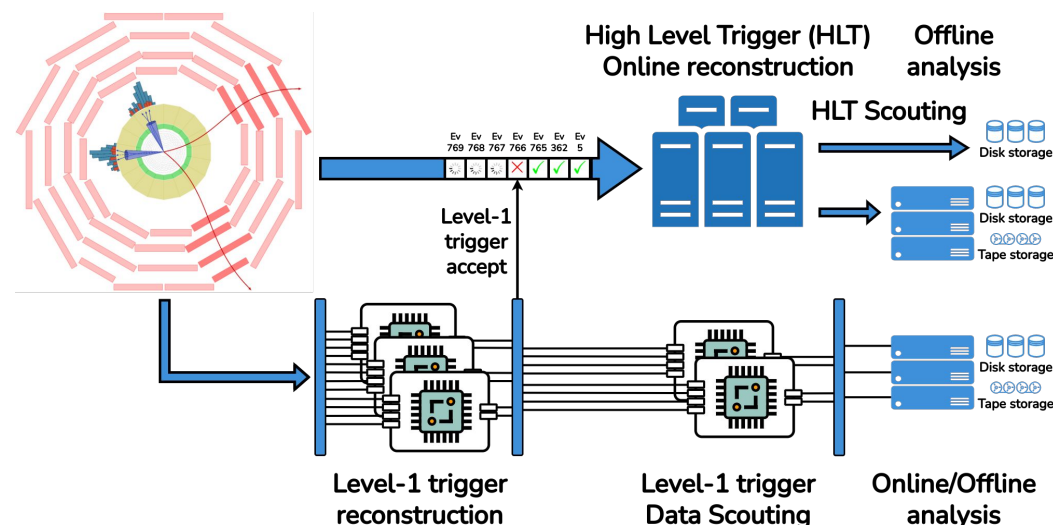
Paul Gessinger (NGT) report at CHEP on “Next Generation Geometry Modelling for Tracking”



WP3: Main goals

- At HL-LHC, up to 200 pile-up interactions: CMS is upgrading the L1T and HLT to enable the same physics program we are doing now (at 60 pile-up)
 - **L1-Trigger (L1T): ~750 kHz + Scouting, High Level Trigger: ~7.5 kHz + Scouting**
- What if New Physics is buried under the bulk of background events we are throwing away due to the trigger selections? **NGT goal is to extend the CMS discovery and precision measurement reach by:**

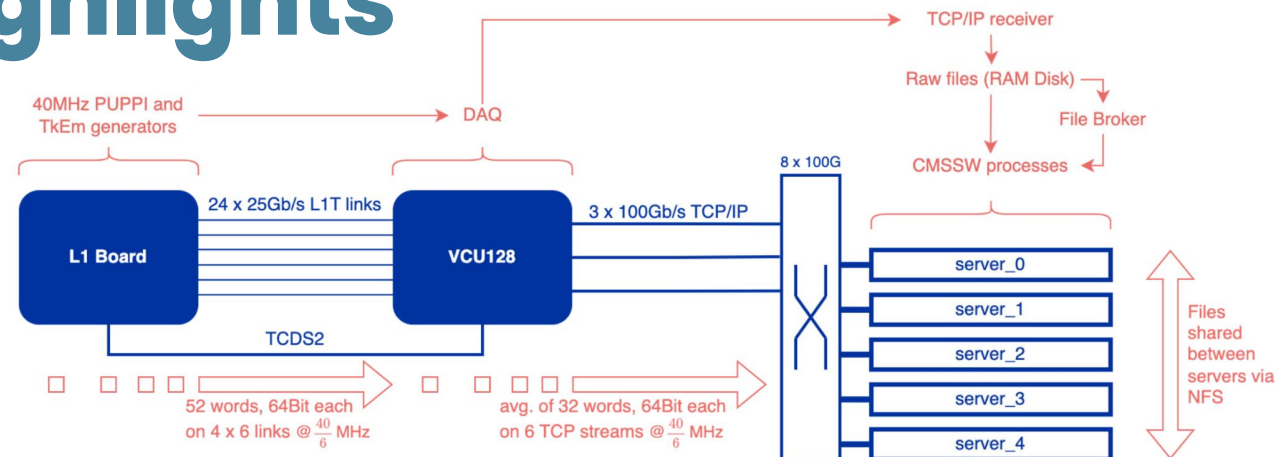
- Redesign the data collection and scouting strategy to **reduce the need to reject events in the Level-1 and High-Level CMS triggers** aiming at complementing the current workflows
- **Replace the trigger filtering task** with an event processing task **similar to what happens with offline** events stored on disk
- Achieve all this by exploiting **advanced AI solutions and heterogeneous frameworks**



WP3: Highlights

R&D has started successfully

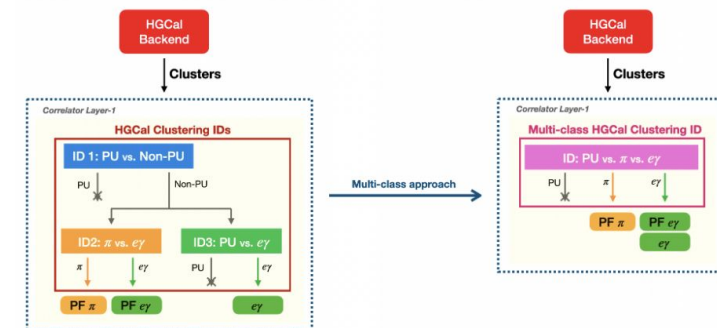
- Work embedded in CMS L1T/HLT offline groups
- Building on ongoing CMS work
 - L1T Run-3 Scouting Demonstrator, first L1T AI-models in Run-3 data taking, GPUs in production in HLT Farm since Run-3
- Actively collaborating with WP1
 - Specific needs for hls4ml/Conifer for new R&D
 - Heterogeneous frameworks and fast inference



[Overview of the new demonstrator system for the HL-LHC L1 Scouting](#)

Some examples of ongoing work and new initiatives

- **L1T Scouting for HL-LHC:** benchmark analyses defined and tested on new hardware (first demonstrator in place!)
- **AI@L1T:** ultrafast jet-taggers models, **multiclass clusters ID for $e/\gamma, \pi$, pile-up, ...**
- Operational practises for [unsupervised anomaly detection trigger for L1T Run 3](#)
- Systematic investigation of RECO performance in Phase-2 HLT, Phase-2 Offline and Run-3 Offline to identify bottlenecks
- Towards evolving CMSSW into a distributed application for Phase-2 HLT Farm: implementation of a client server prototype
- Evaluating impact of RAW data compression for HLT



[New ML-based algo in L1T FPGAs for \$e/\gamma, \pi\$, pile-up clustering ID](#)

WP4: Education & Outreach

Vision: train, sustainably

- ▶ Develop project members knowledge with an aggressive education programme and multiple training opportunities
- ▶ Reuse and enhance existing infrastructures to cover education and training needs of NextGen staff

Exchange, impact beyond NextGen

- ▶ Visiting scientists, public seminars
- ▶ Develop a coherent message for outreach and external communication, including the web site.
- ▶ Make NextGen knowledge accessible also outside the project
 - Conference participation and online presence
 - Support NextGen staff to contribute to Outreach initiatives, Schools, and Education activities



SMARTHEP
REAL-TIME ANALYSIS FOR SCIENCE AND INDUSTRY

NexTGen

EDGE ML SCHOOL



INFN
ESC
INTERNATIONAL SCHOOL

14 - 24 October 2024
CE.U.B Bertinoro (FC) Italy

Architectures, tools and methodologies for developing large scale scientific computing applications

In partnership with:
NexTGen

Already in Full Swing!

Thematic School on Machine Learning

CERN School of Computing



13 - 19 October 2024
Split, Croatia



45th CERN School of Computing



8 - 21 September 2024
Hamburg, Germany



Conclusion

Introducing the NextGen Triggers Project



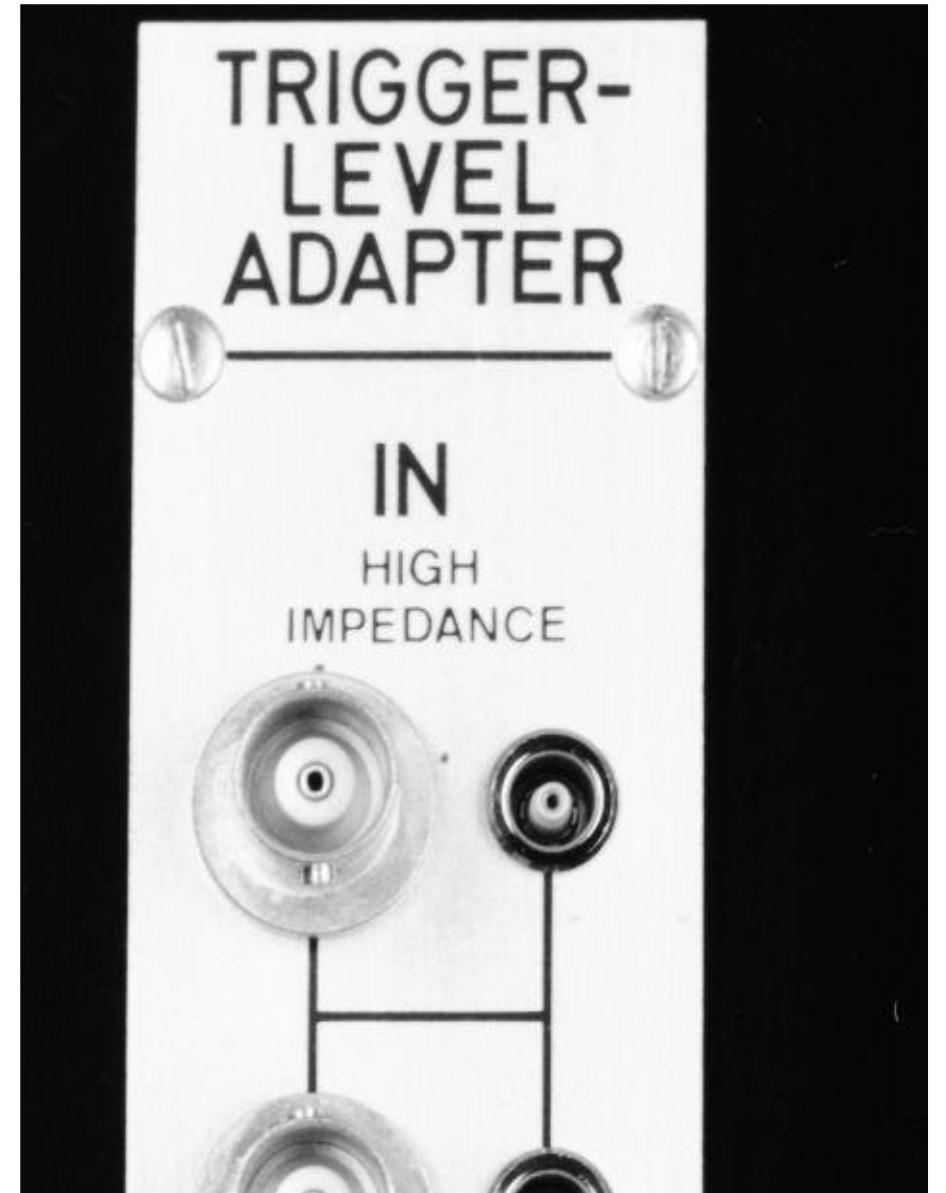
Fusing computing, algorithms, theory, ML:
future solutions for the experiments



Embedded in the experiments, significant
influx of community expertise: building with
the community, for the community



Now advancing in full steam: [technical workshop on Nov 25-27](#) for first results!



RESOURCES

Beyond 15mins



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



[Project "Proposal"](#) and other resources



1st Technical Workshop Nov 25-27

<https://indico.cern.ch/event/1421629>



NextGen
Next Generation Triggers