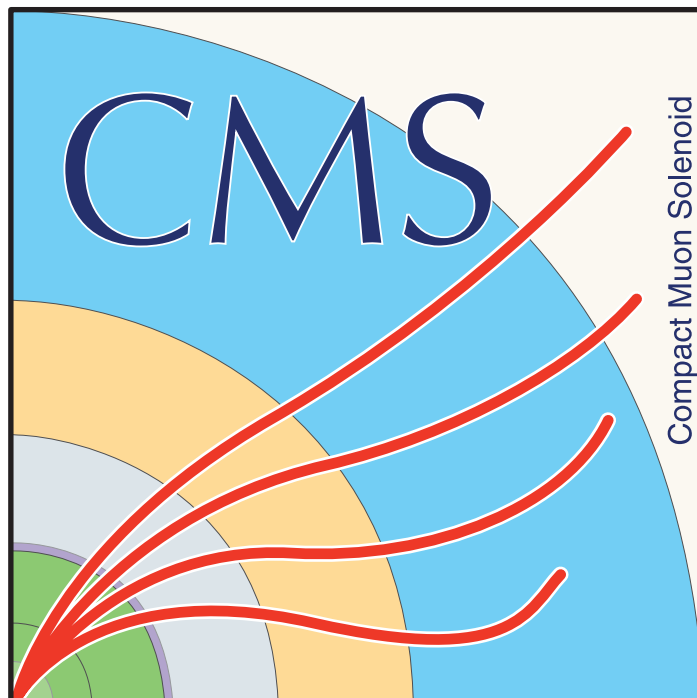


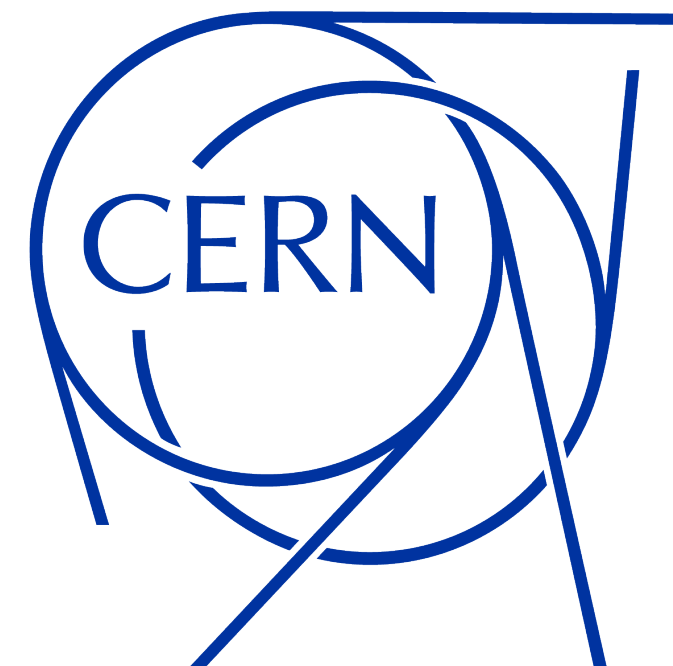
CMS Level-1 Trigger Data Scouting for HL-LHC

Efe Yiğitbaşı (Rice U.)
on behalf of CMS Collaboration

8th September 2025

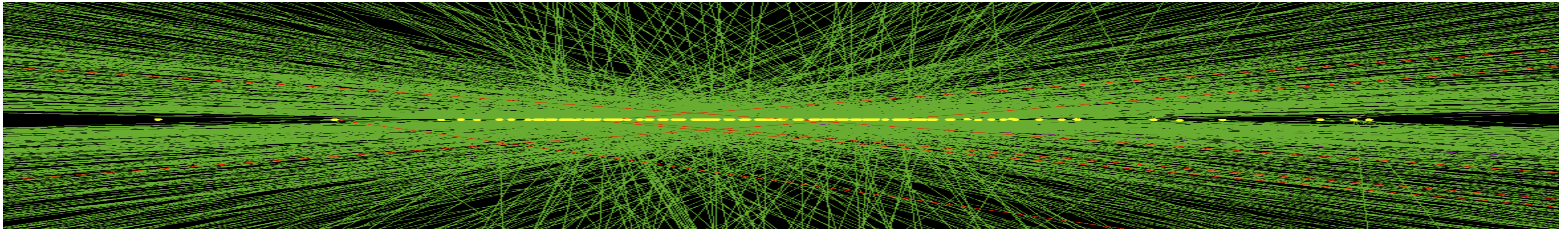


RICE

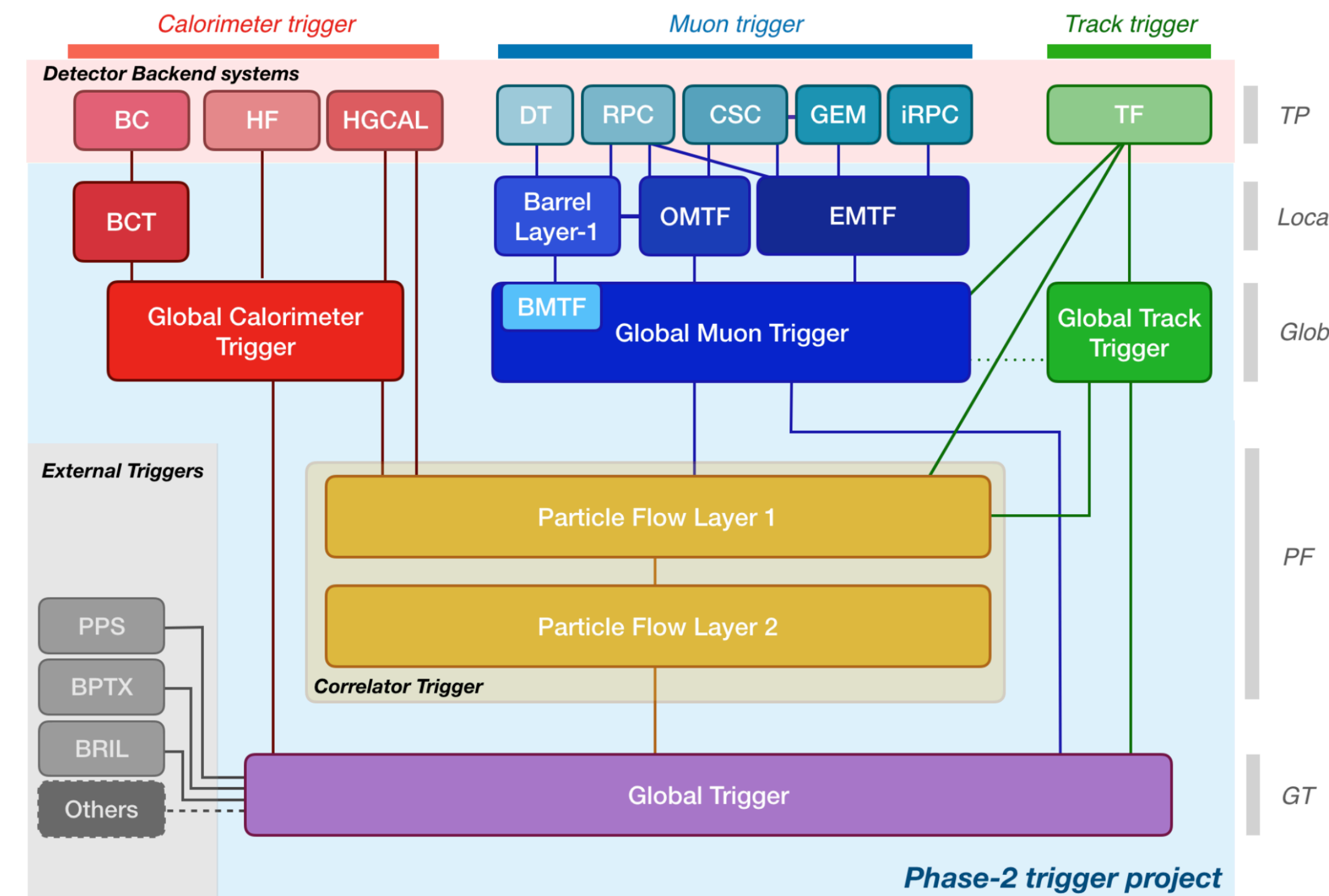


CMS Level-1 Trigger at HL-LHC

- At HL-LHC: peak instantaneous luminosity up to $7.5 \times 10^{34} \text{ cm}^{-2}\text{s}^{-1}$, up to 200 pile-up (PU) interactions
- CMS will go through major detector upgrades, including both **Level-1 Trigger (L1T)** and **High Level Trigger (HLT)**

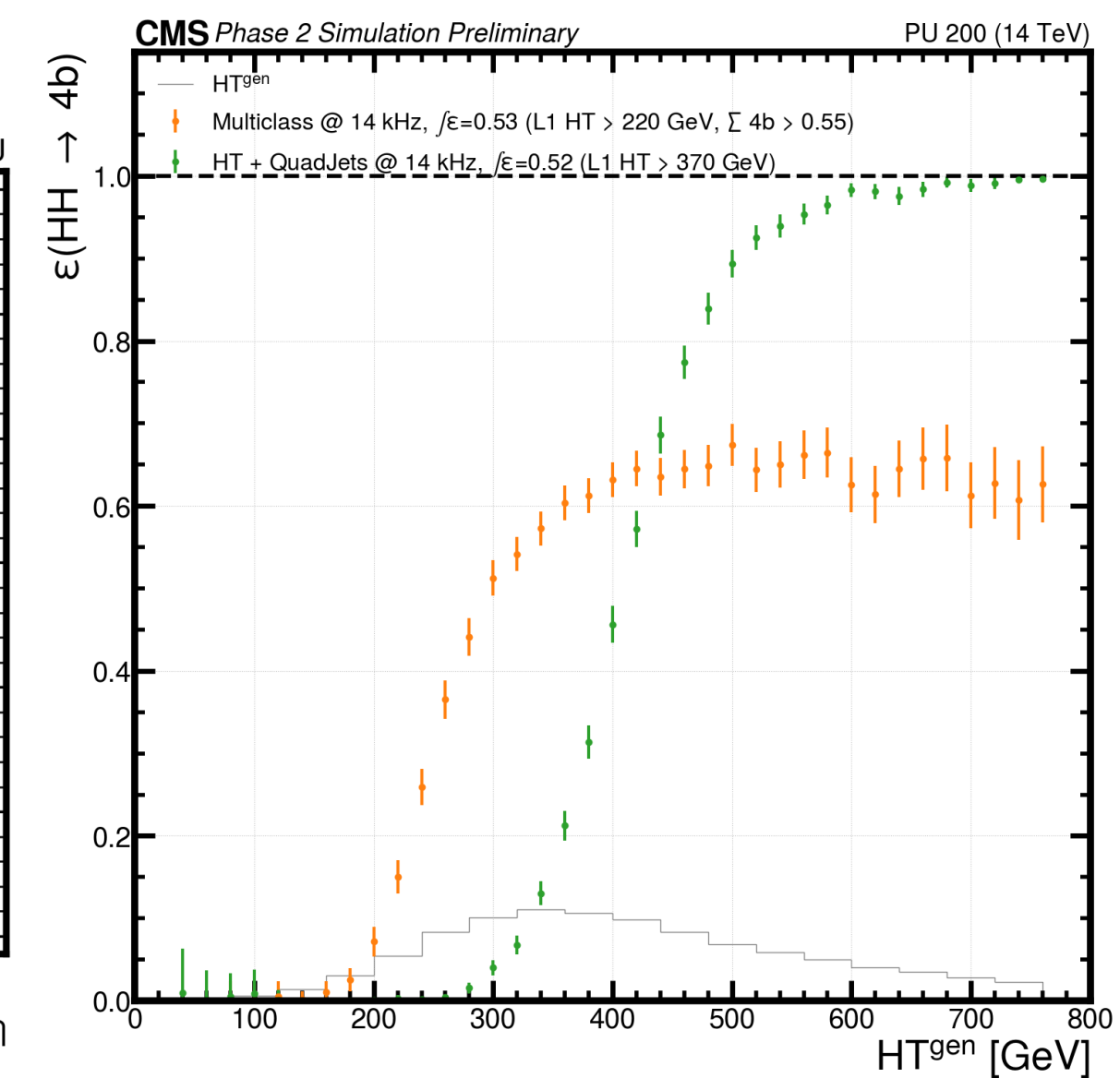
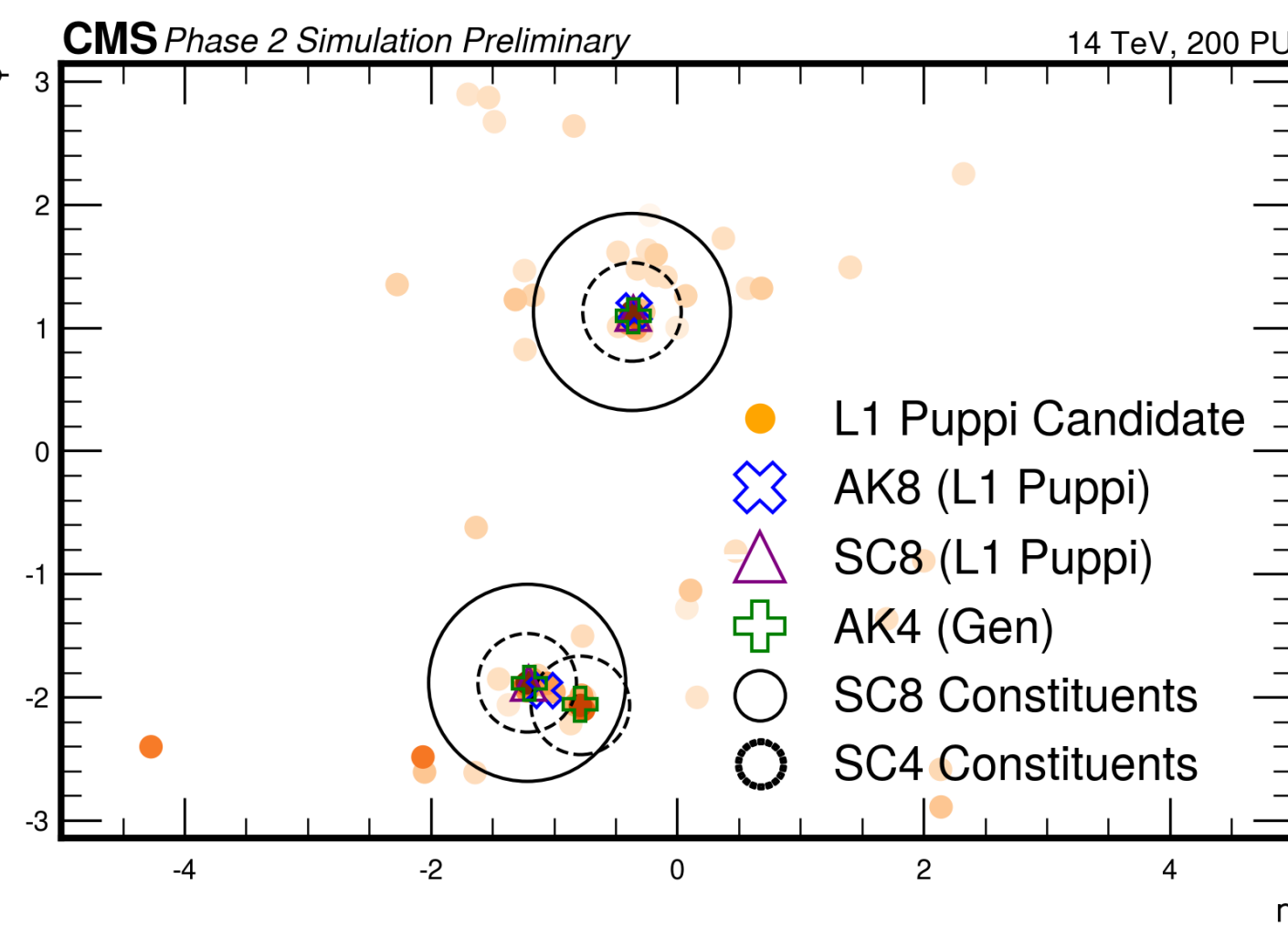
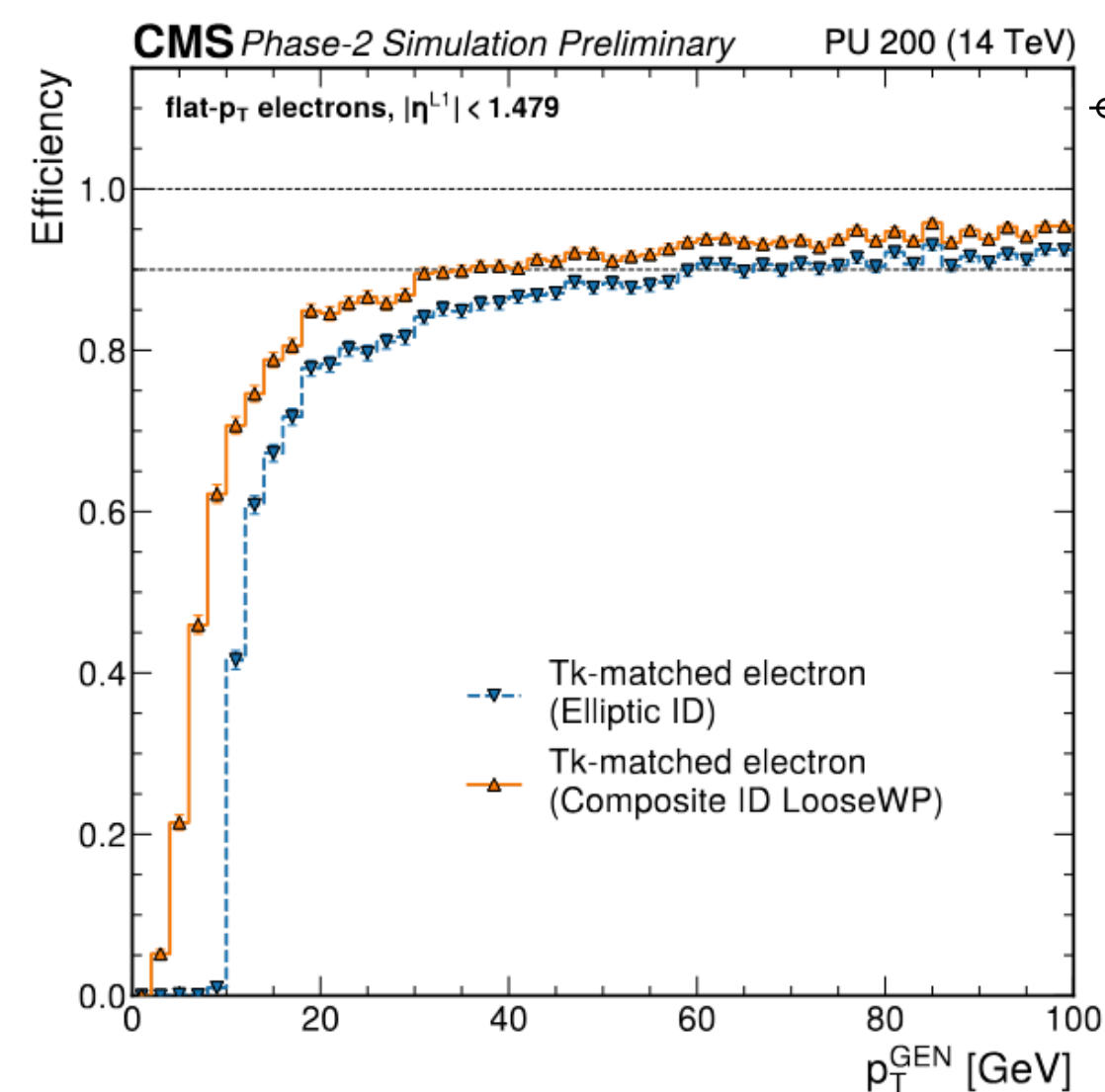
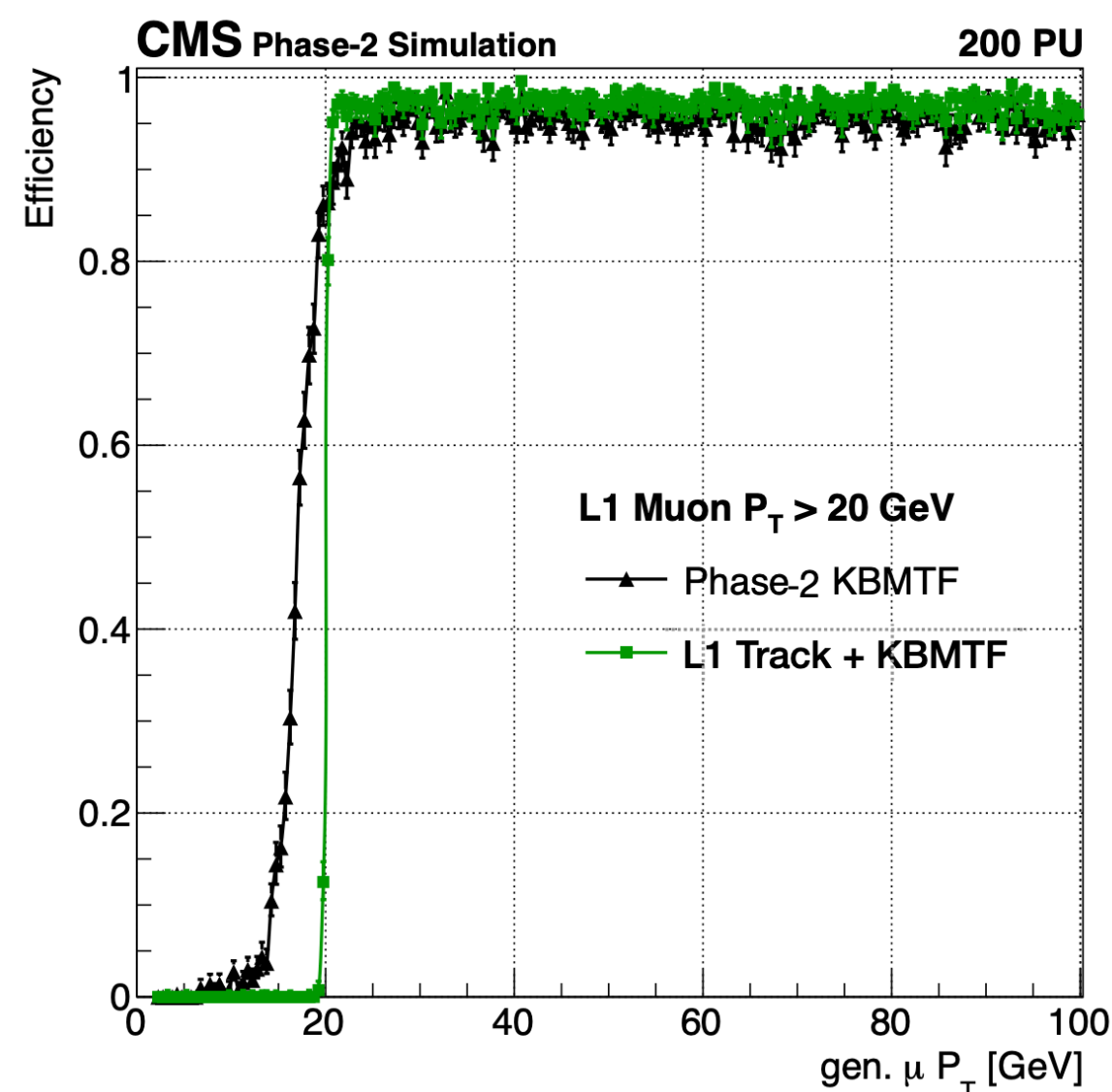
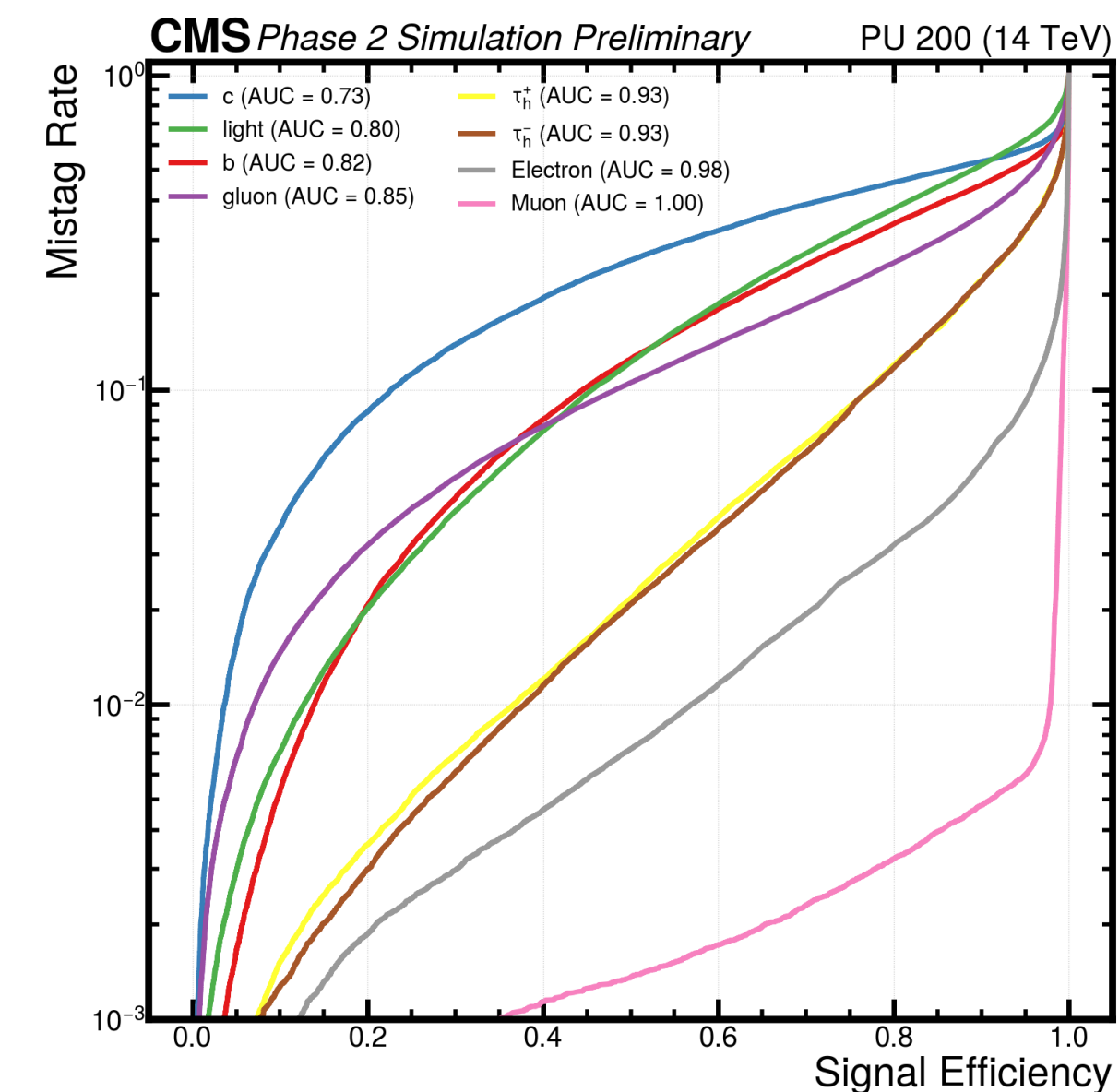


- New CMS L1T:
 - Input data rate from 2 Tb/s to 63 Tb/s (**$\sim 1/10$ of the internet traffic**)
 - First time **tracking at 40MHz** ($p_T > 2 \text{ GeV}$) & higher granularity inputs from Calorimeters and Muon detectors
 - Much more powerful FPGAs**: 7.5x more resources
 - 12.5 μ s** to take decision, **750 kHz** to HLT
- Goal:
 - Maintain the same **Run-3 physics program** in the presence of high pileup, while improving capability to **efficiently select specific signatures** and **extending reach for new physics searches**



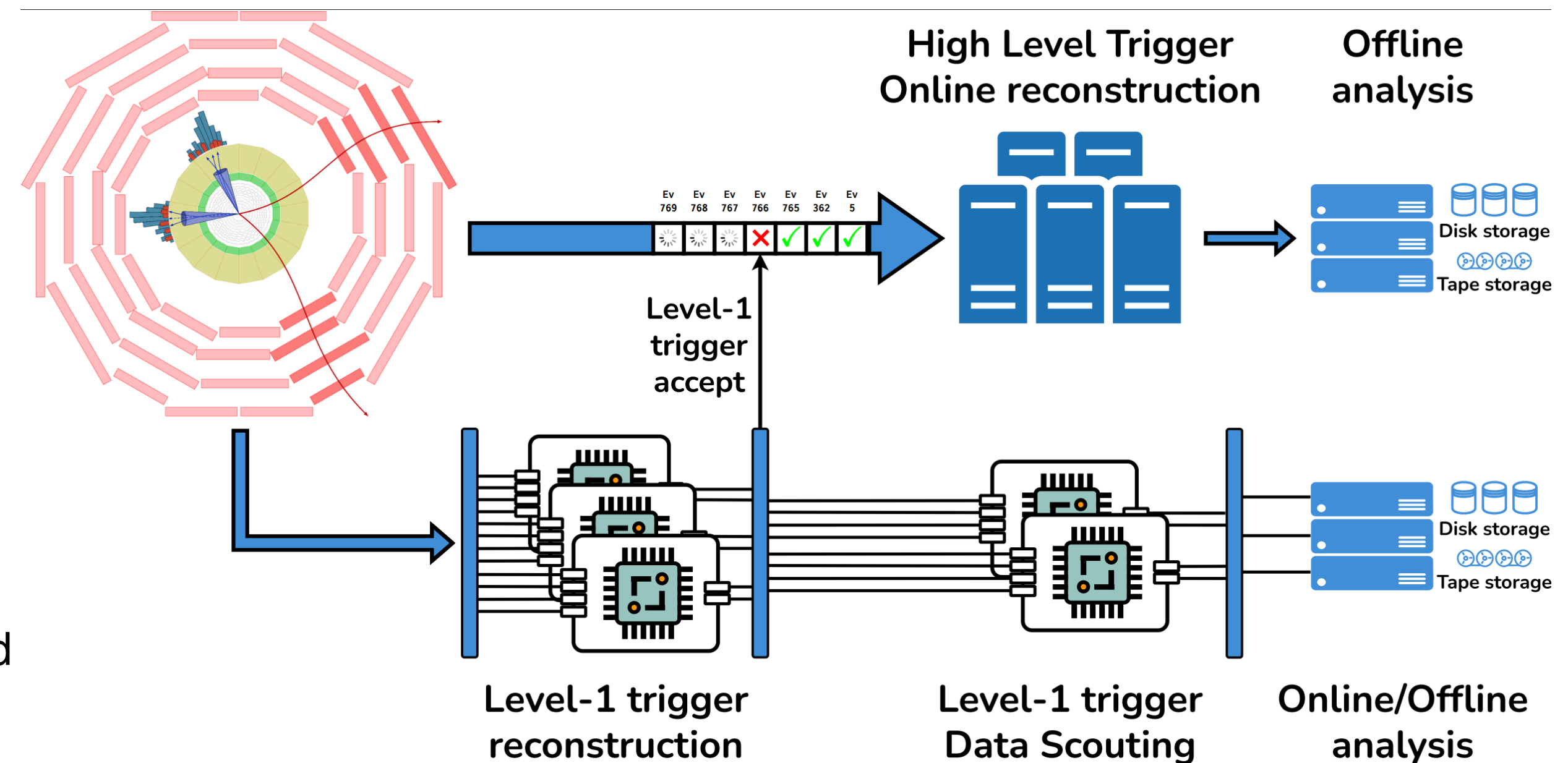
CMS Level-1 Trigger at HL-LHC

- At HL-LHC we will have **offline-like algorithms at L1T**, using the **more powerful hardware, exploiting tracks, finer granularity calorimeter, and refined muon trigger primitives**
- Track matching:** sharpened μ turn-on, electron identification, $\gamma/e/\mu$ isolation
- Particle Flow (PF):** ultimate performance for jets/ E_T^{miss}/τ_h
- L1Tracks** \rightarrow **vertexing** \rightarrow **PU mitigation** via pile-up per particle identification (PUPPI)
- FastML:** tools to synthesize NN/BDTs into FPGAs for **regression** and **event/particle classification**



CMS L1T Data Scouting for HL-LHC

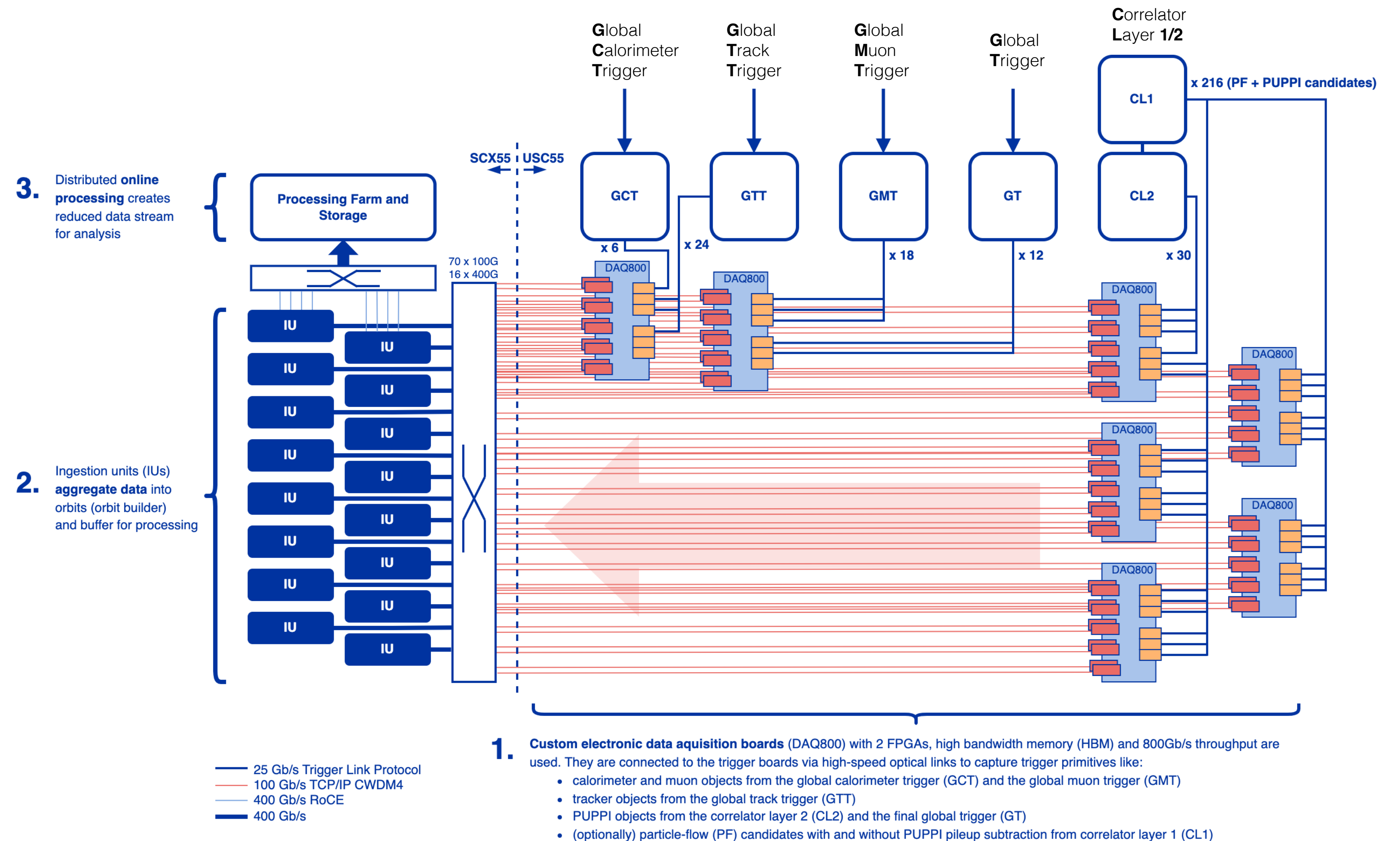
- The CMS L1T at HL-LHC will be a very capable system with close-to-offline resolution with the inclusion of tracking and particle flow
- But what if New Physics don't look as expected, and instead buried under the bulk of the background events we are throwing away due to the trigger selections?
- At L1T, **we develop an advanced Scouting workflow that**, with the increased granularity and flexible architecture of the Phase-2 CMS experiment, **can allow to perform trigger-level physics analyses at 40 MHz**
 - Look for physics signatures identifiable with **just L1T information** that would evade the L1T → HLT → Offline chain
 - Too large “irreducible” backgrounds, e.g. **narrow resonances of low and unknown mass**
 - Signal identification requires an algorithm that can't fit the L1 fixed latency and resource budget, e.g. **has too large combinatorics for some events, requires a too complex NN, ...**
 - Signal identifications requires time-correlation across several BXs, beyond what is allowed by GT, e.g. **slow or long-lived BSM**
 - **N.B. L1T objects might not be well suited for physics analyses. Detailed simulation studies are needed to explore the capabilities of a scouting system**



CMS L1T Data Scouting for HL-LHC

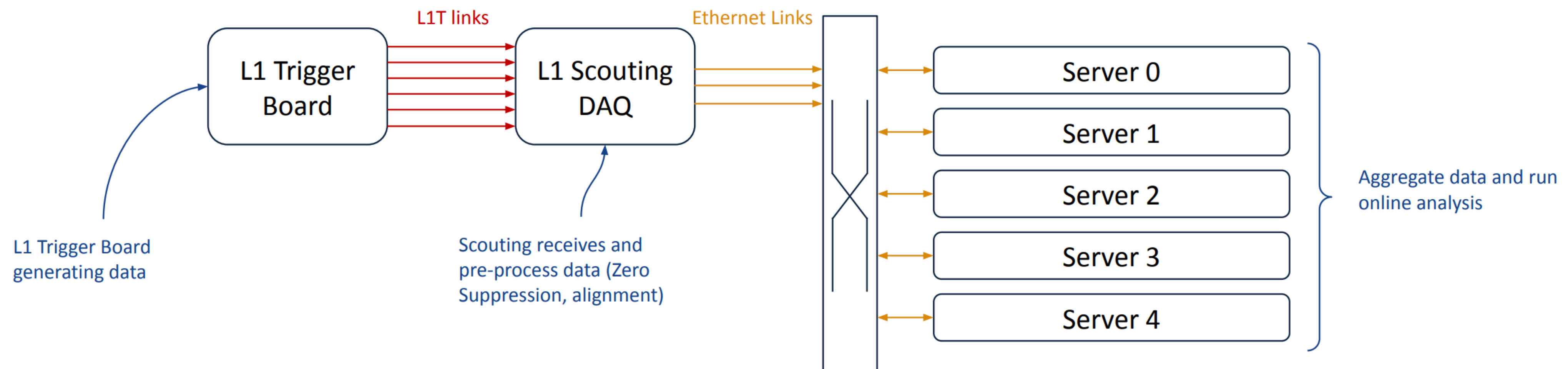
Baseline design of the L1T Scouting system for HL-LHC is designed to capture:

- **Standalone muon and calorimeter objects** from GMT and GCT
- **Tracker objects** from the GTT
- **PUPPI objects** from the CL2
- **Final decision objects** from the GT
- **Possible extension to PF candidates** with and without PUPPI pileup subtraction from CL1



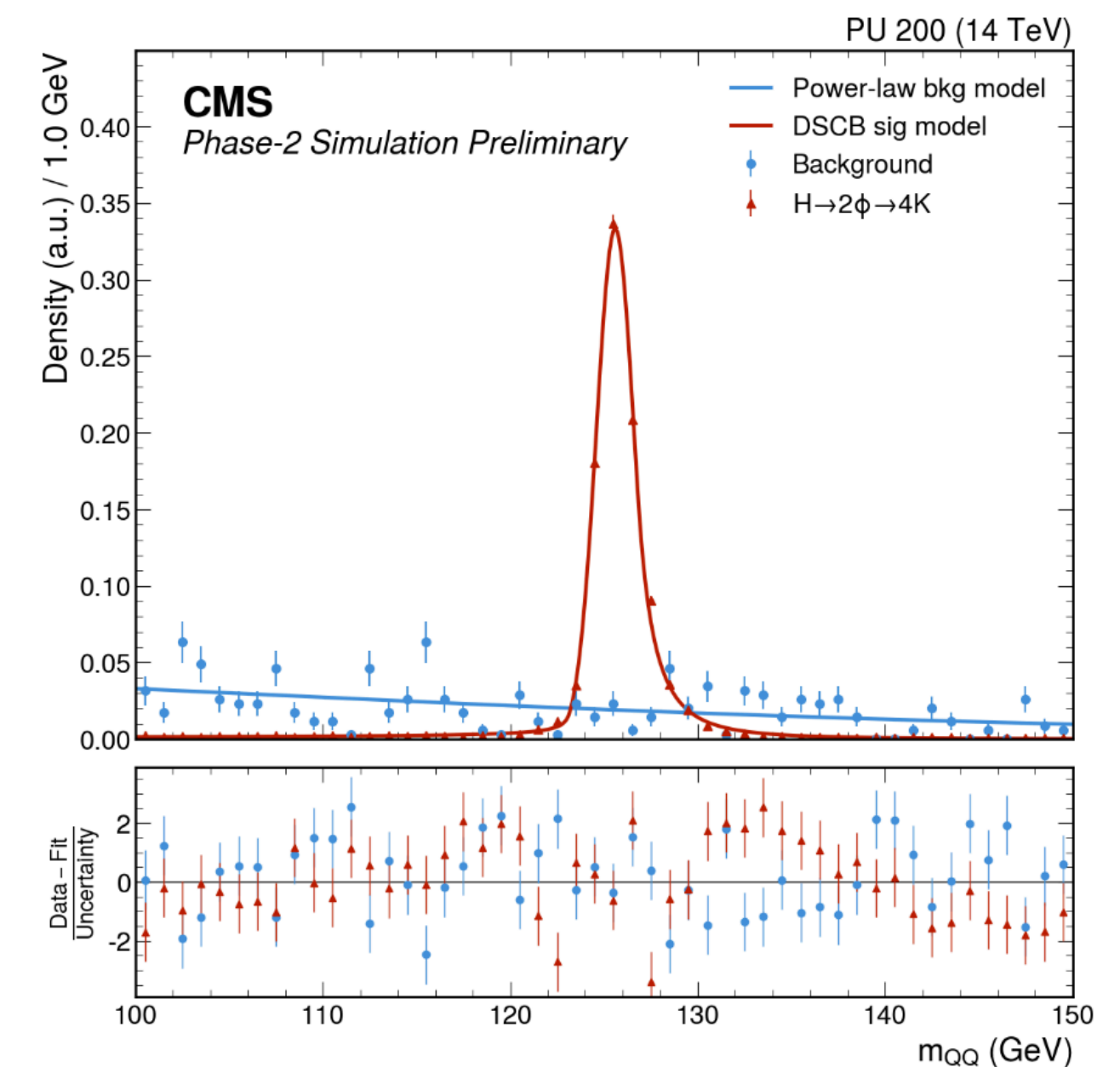
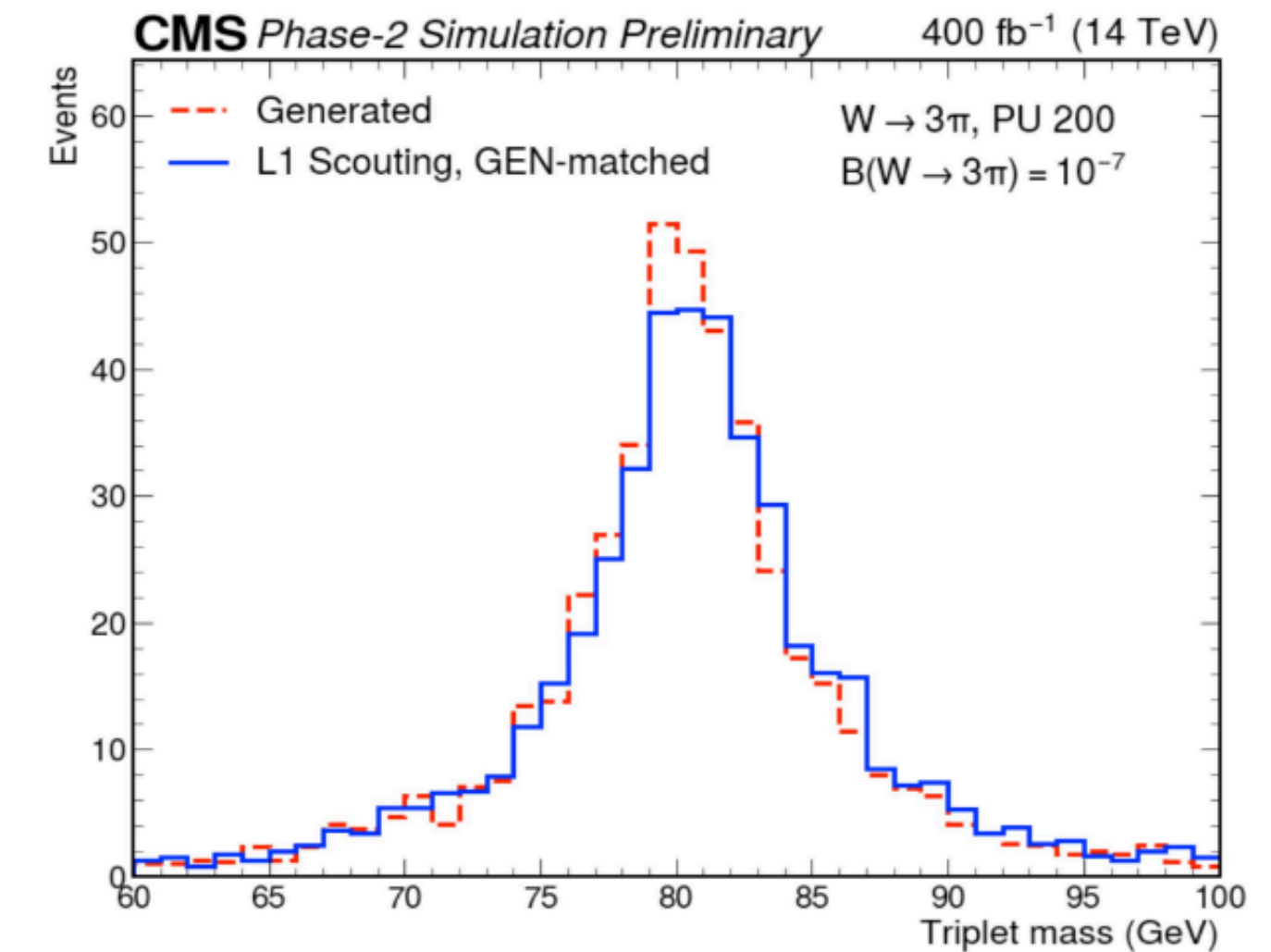
Demonstrator and Going Beyond the Baseline

- Build a demonstrator for data acquisition and real-time processing for the baseline scouting system, collect L1 trigger objects at 40 MHz, and run online analyses
 - Also very useful to test alternative approaches for data processing and extensions to the hardware for data acquisition
- Explore the physics opportunities and technical feasibility using different L1T inputs with respect to TDR baseline and R&D to investigate different implementations
- How?
 - Investigate **different analyses of increasing complexity** in terms of data processing and bandwidth
 - **Incrementally build a test system** of increasing bandwidth & processing power to run benchmark analyses



First Prototype Analyses: Exclusive Rare Decays

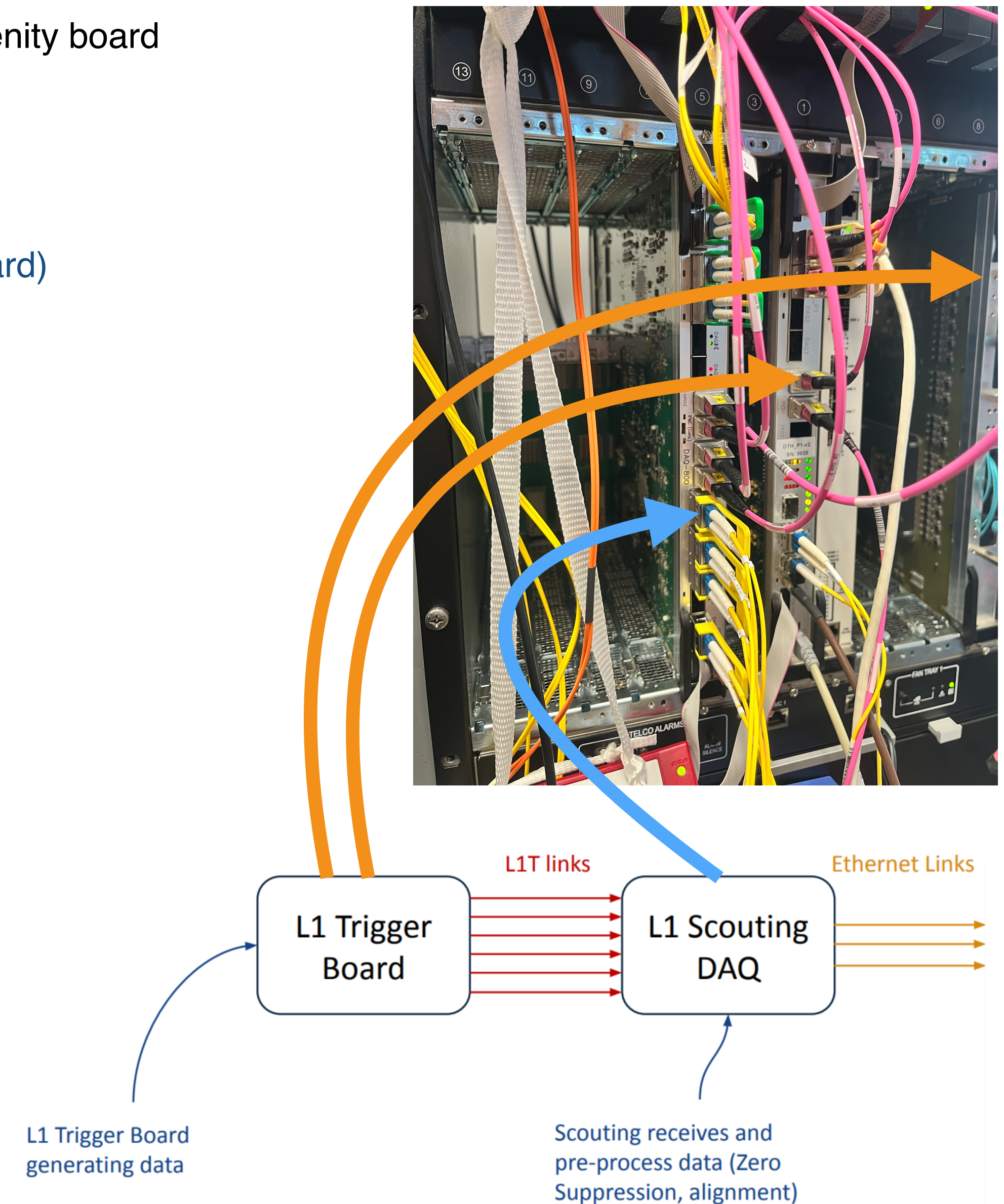
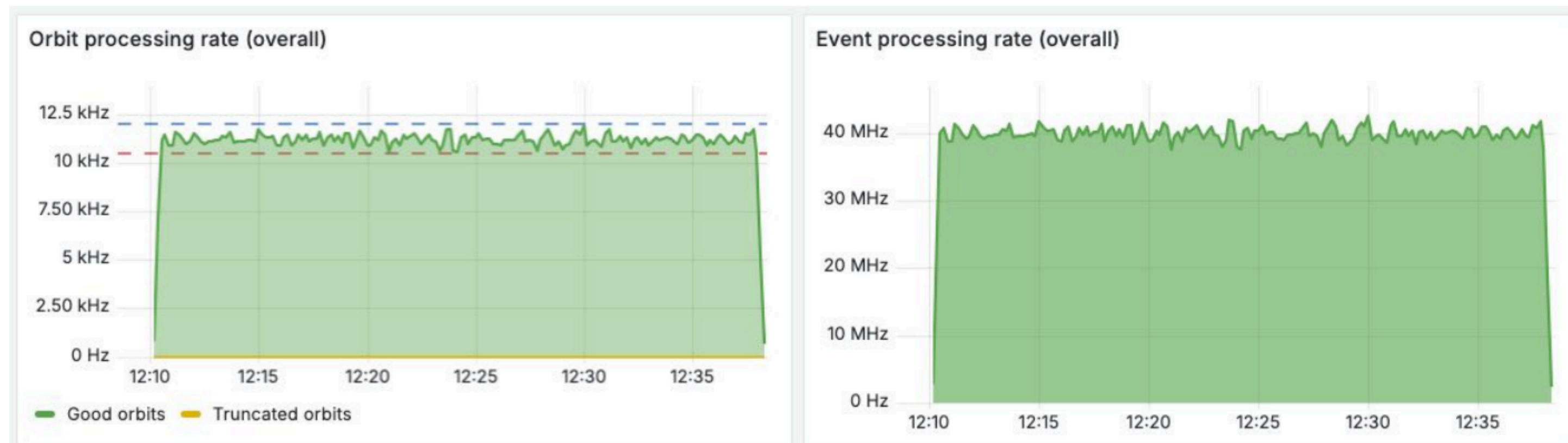
- Developed first set of physics analyses to test the capabilities of 40 MHz demonstrator
 - Soft final state**, not selected efficiently by the standard L1 Trigger
 - Small signal yield**, but visible as narrow peak over smooth background
 - Combinatorics with many objects** may not fit fixed-latency L1
- Example: **Exclusive rare decays of W & H**
 - Signatures identifiable with PUPPI candidates and e/γ objects
 - Simple selection on kinematic quantities and isolation reduce drastically the background rate by **allowing the selected events to be saved for final analysis**
 - $W \rightarrow 3\pi$** (current BR UL 10^{-6} , SM @ $10^{-8/-12}$)
 - Achieved 5×10^{-7} background reduction with 5.5% signal selection efficiency
 - $W \rightarrow Ds \gamma \rightarrow KK\pi \gamma$** (current BR UL 10^{-4} , SM @ 10^{-8}), **$W \rightarrow \pi \gamma$** (current BR UL 10^{-5} , SM @ $10^{-9/-7}$)
 - 6×10^{-6} bkg reduction 4% signal selection, 4×10^{-6} background reduction with 12% signal selection efficiency
 - $H \rightarrow \rho\rho \rightarrow 4\pi$ ($H \rightarrow \phi\phi \rightarrow 4K$, $H \rightarrow \phi J/\psi \rightarrow KK\mu\mu$)**
 - 3 (1, 0.3) $\times 10^{-5}$ background reduction & 19 (42, 31)% signal selection efficiency
 - $H \rightarrow \rho\gamma \rightarrow \pi\pi\gamma$ ($H \rightarrow \phi\gamma \rightarrow KK\gamma$, $H \rightarrow J/\psi\gamma \rightarrow \mu\mu\gamma$)**
 - 9 (3, 0.1) $\times 10^{-6}$ background reduction & 28 (44, 34)% signal selection efficiency



CMS-DP-2024-096

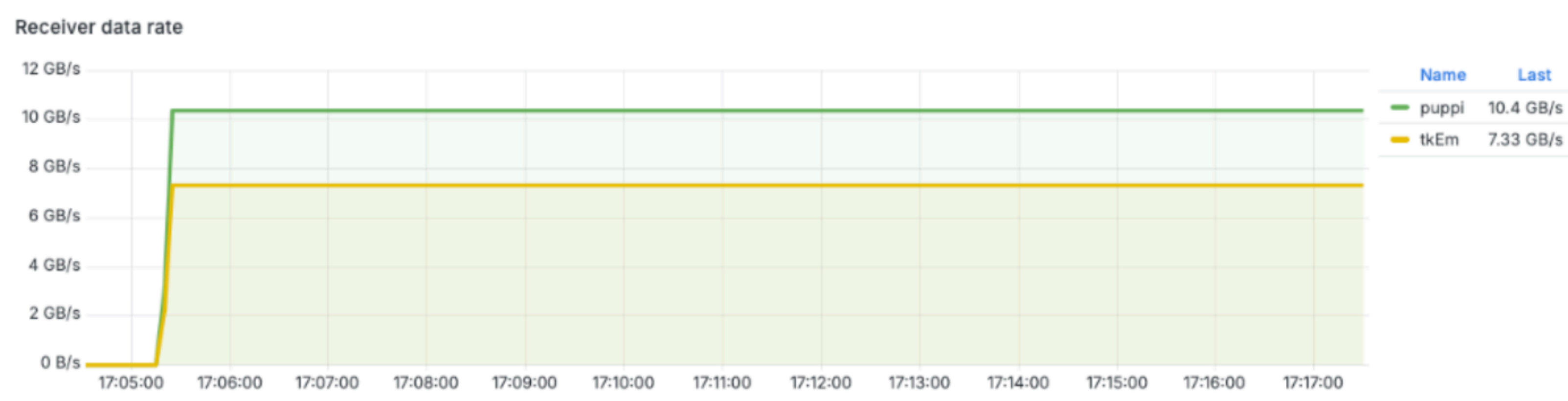
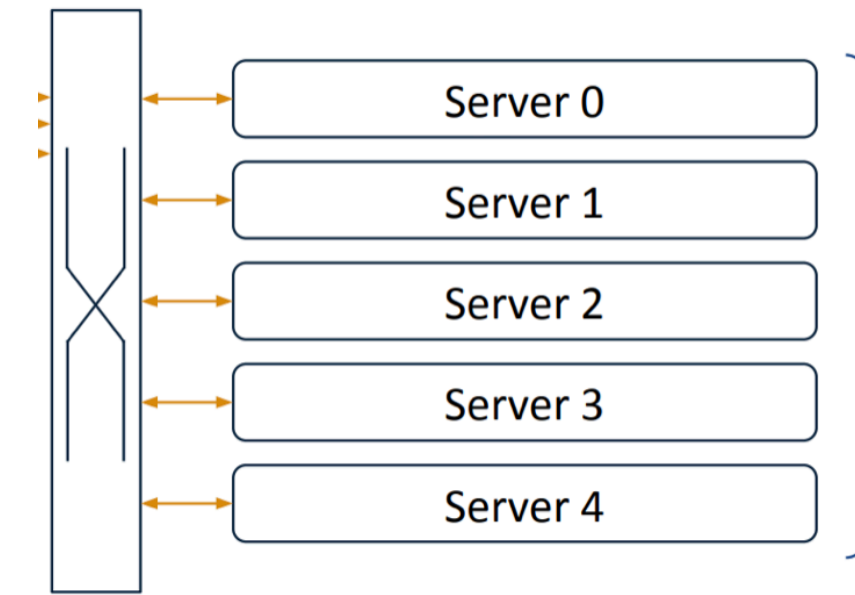
Current Demonstrator Setup

- L1T data generated in one DAQ and Timing Hub (DTH) prototype ATCA board and one Serenity board
 - Clock & TCDS2 source + L1T-like data source
 - Generate **12 data streams to simulate L1T PUPPI and e/y objects**
 - **48 links @ 25Gbps with L1 Protocol** sent to L1T Scouting DAQ (24 links from each board)
- L1T Scouting DAQ with DAQ-800
 - Two Xilinx Ultrascale+ VU35P FGPAs with High-Bandwidth Memory (HBM)
 - Up to 48 L1 input links via FireFly connectors for a total of 1.2 Tbps
 - 10 × 100 Gb Ethernet (GbE) links from the on-board QSFPs
 - Receive **48 links @ 25Gbps (1/7th of the full system)**

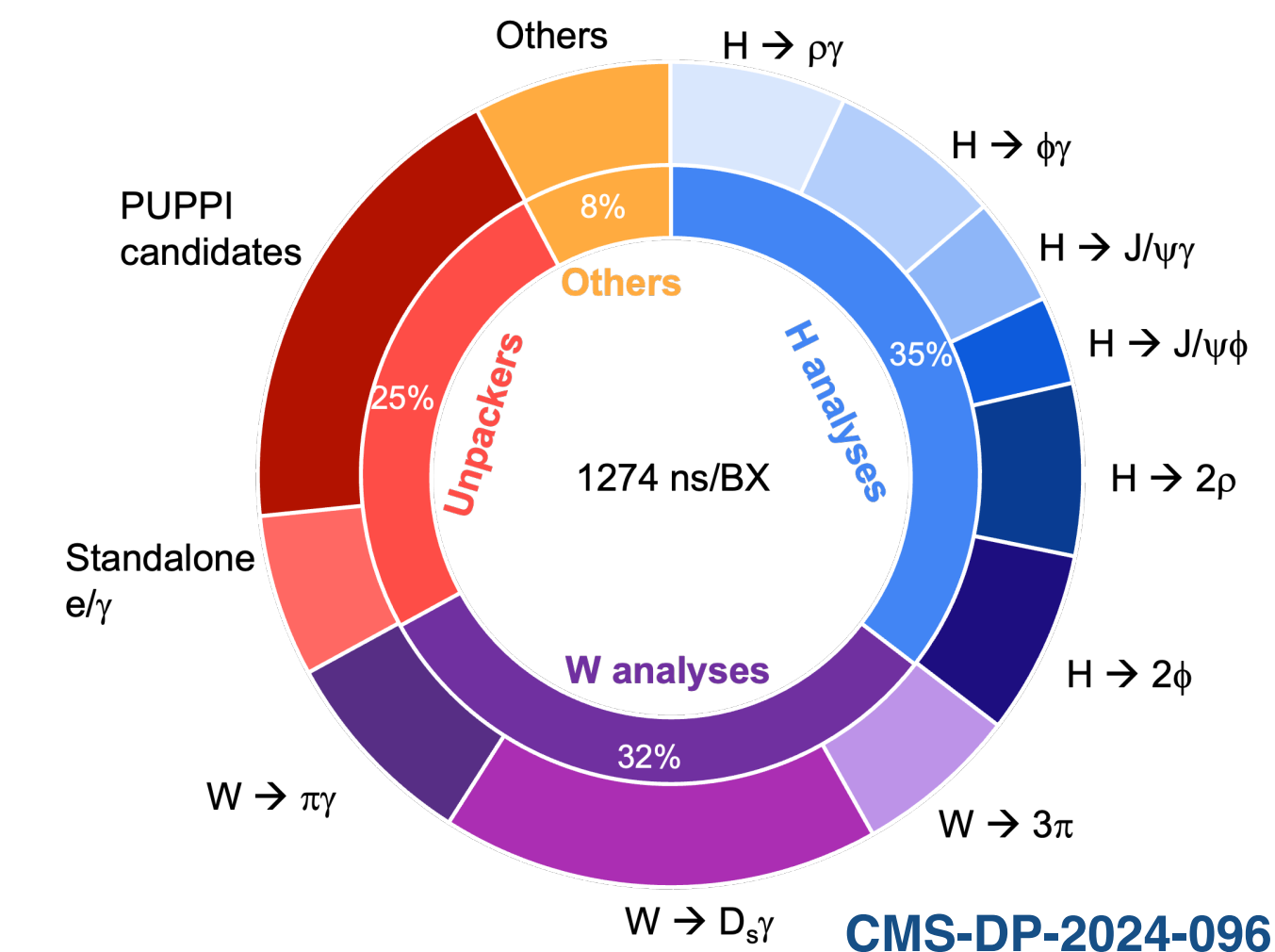


Current Demonstrator Setup

- L1T Scouting online processing servers
 - 5 servers with AMD EPYC 9654, 768 GB of RAM and 2x100G NIC
 - Receivers writing data on ramdisk shared over NFS
 - Aggregation of the stream and processing implemented in standard CMS software
 - Processing in batches of 1 LHC orbit, or 3564 bunch crossings (BX), taking on average **$\sim 1.3\mu\text{s}$ per BX (CPU time)**
 - Switched to Kubernetes to run the setup and software services on the servers
- **Successfully demonstrated receiving of 6 streams of PUPPI and 6 streams of e/γ objects and processing them in 10 online analyses!**



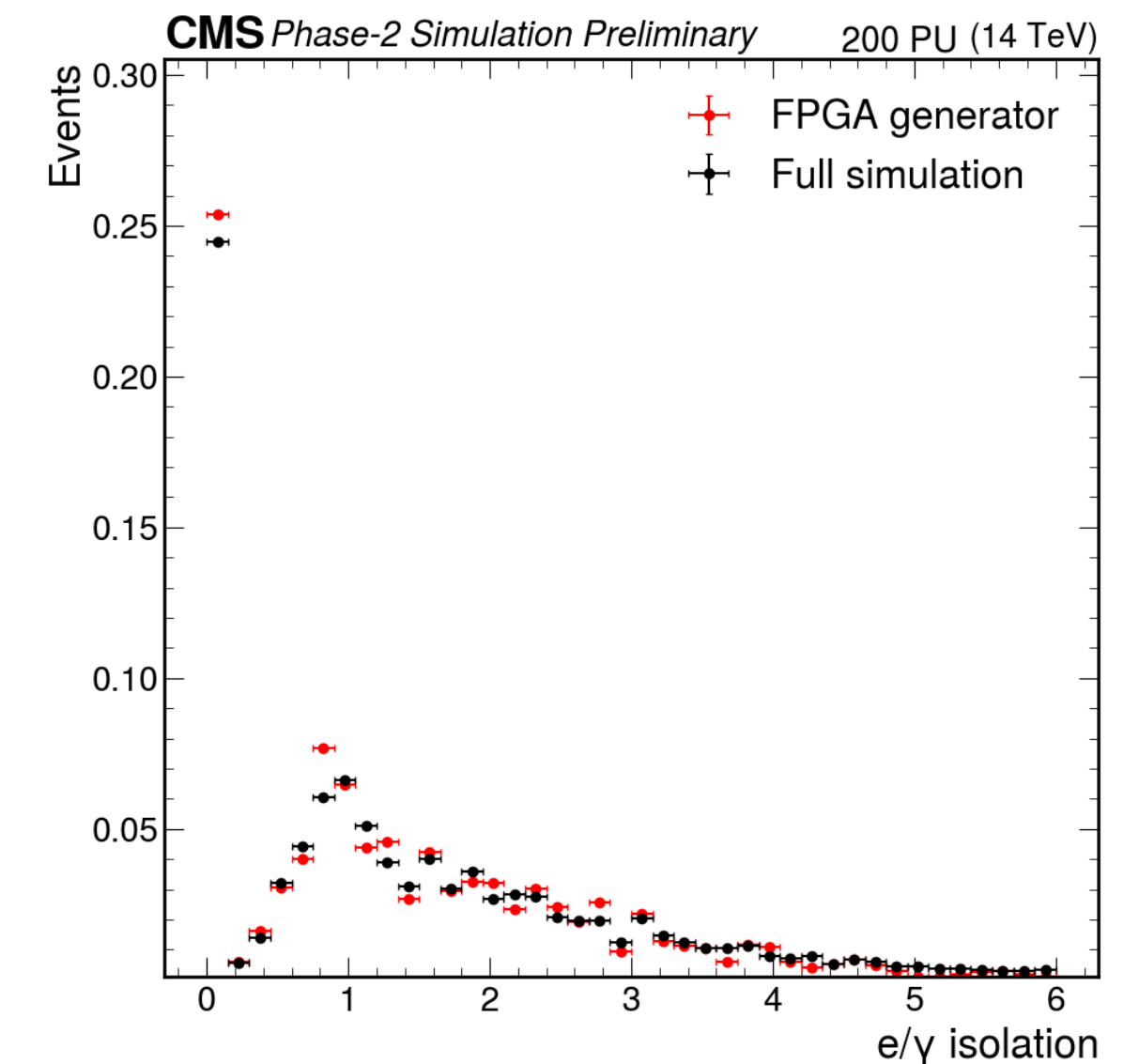
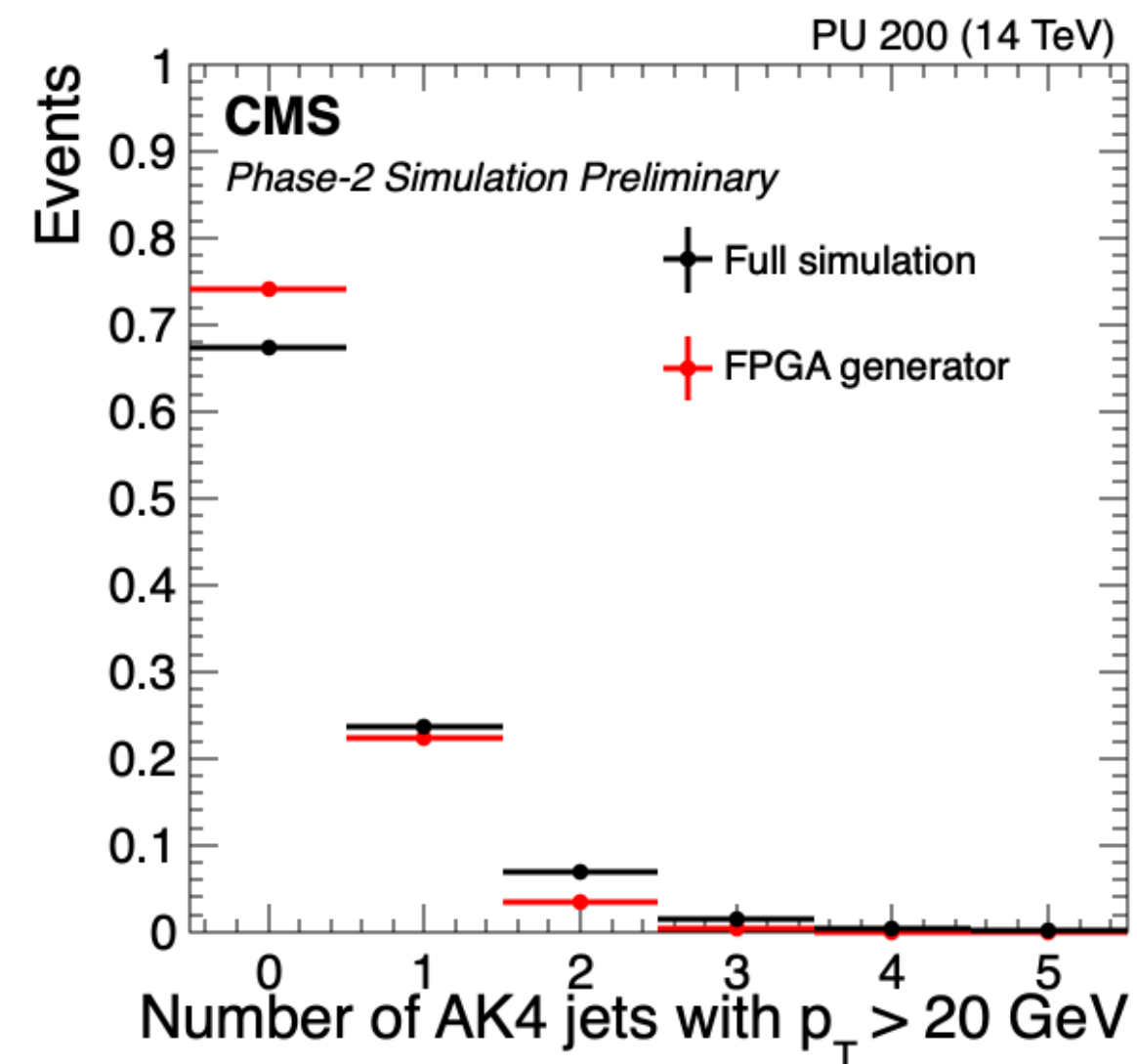
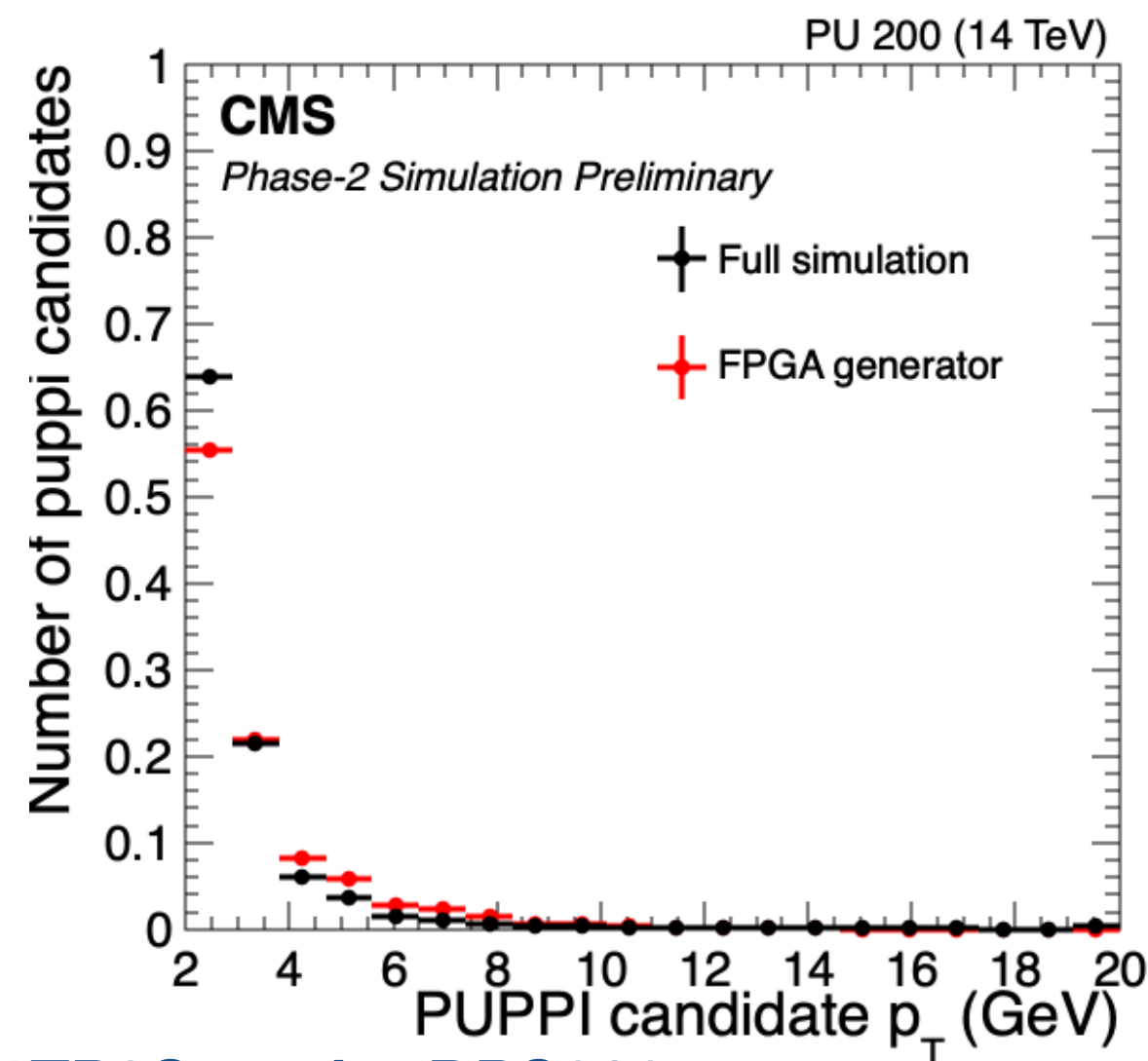
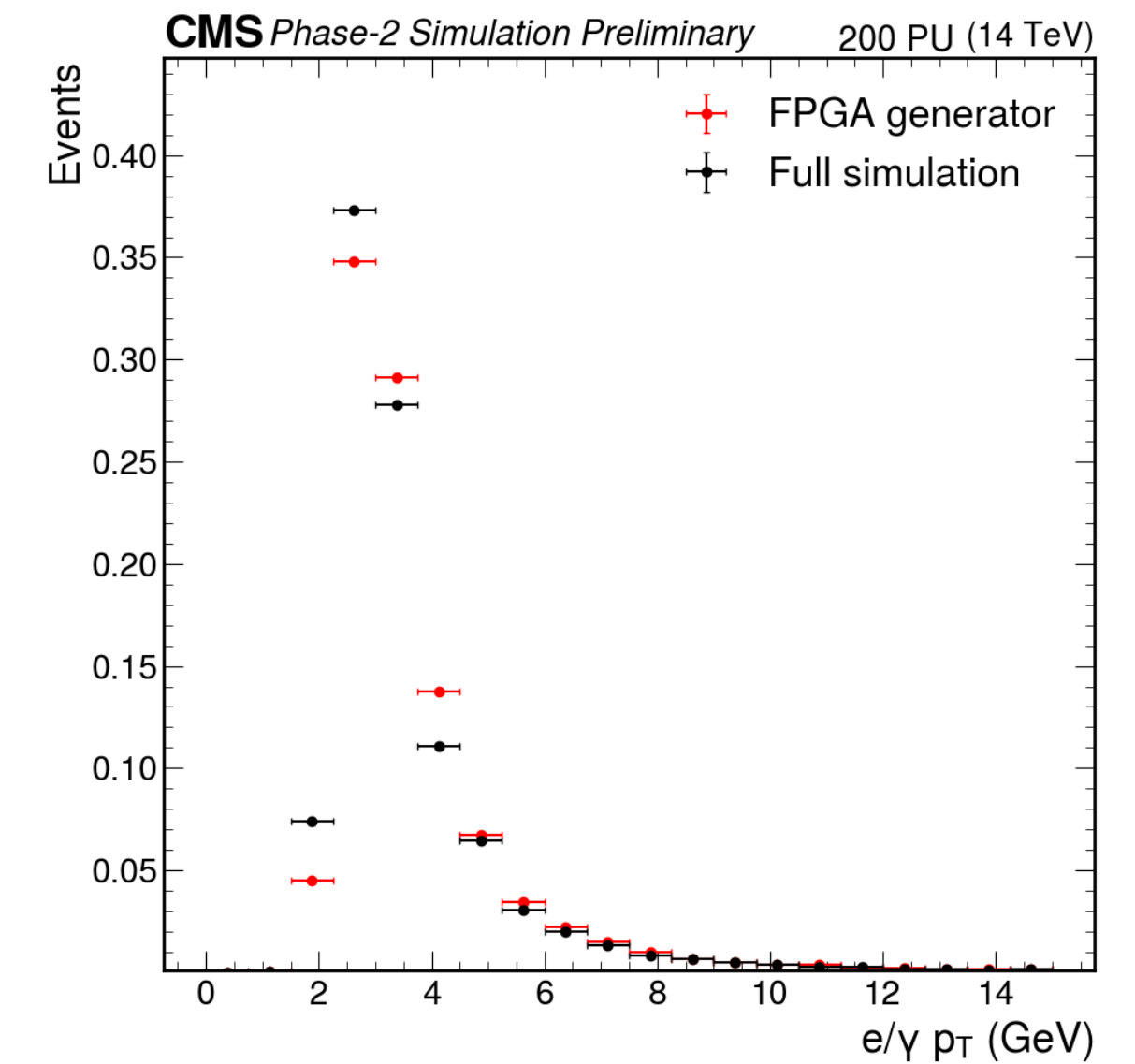
~ 18 GB/s input data rate across all streams, more than the Run 3 CMS HLT output rate



+ $X \rightarrow ee$ analysis which is not shown here

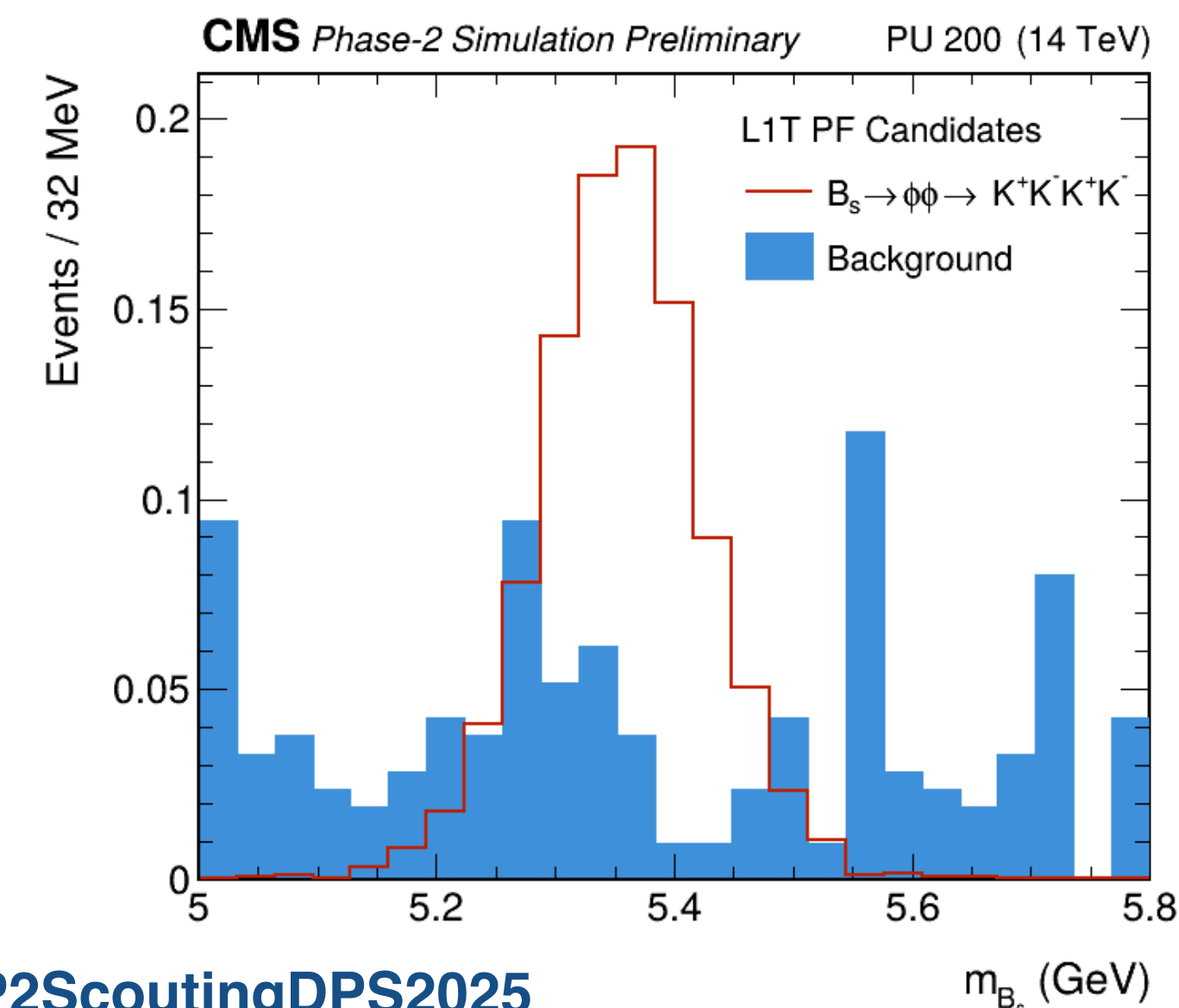
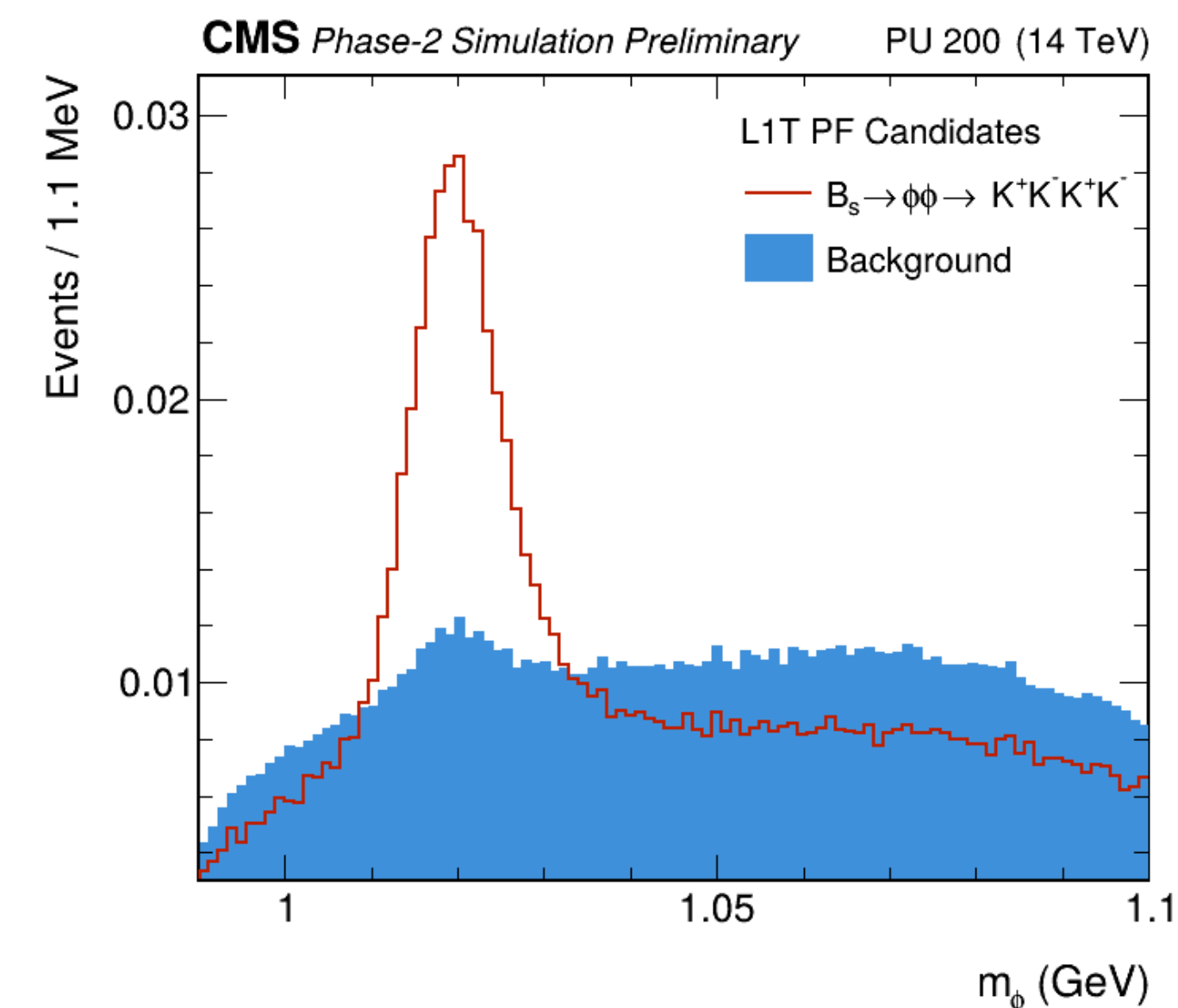
Improving the Demonstrator: 40 MHz Event Generators

- To test the L1T scouting system, two **40 MHz event generators** have been implemented on FPGA generating events resembling minimum bias events at PU = 200
 - One generator simulates the **production of inclusive particles after PF + PUPPI reconstruction** up to the maximum number candidates that L1T can output.
 - One generator simulates the **production of up to 12 inclusive electron or photon (e/ γ) candidates**
- The generators are implemented in HLS and VHDL, and fit hardware constraints in a XCKU15P Kintex UltraScale FPGA
- Kinematic properties of events produced with the FPGA generators are compared with full simulation Minimum Bias samples generated with Pythia8 and Geant4 at PU = 200
 - Generally good agreement is observed between the two samples
- These generators will be implemented in the demonstrator setup and will replace the current repeated loop or software based event generation.



Going Beyond the Baseline: Analyses

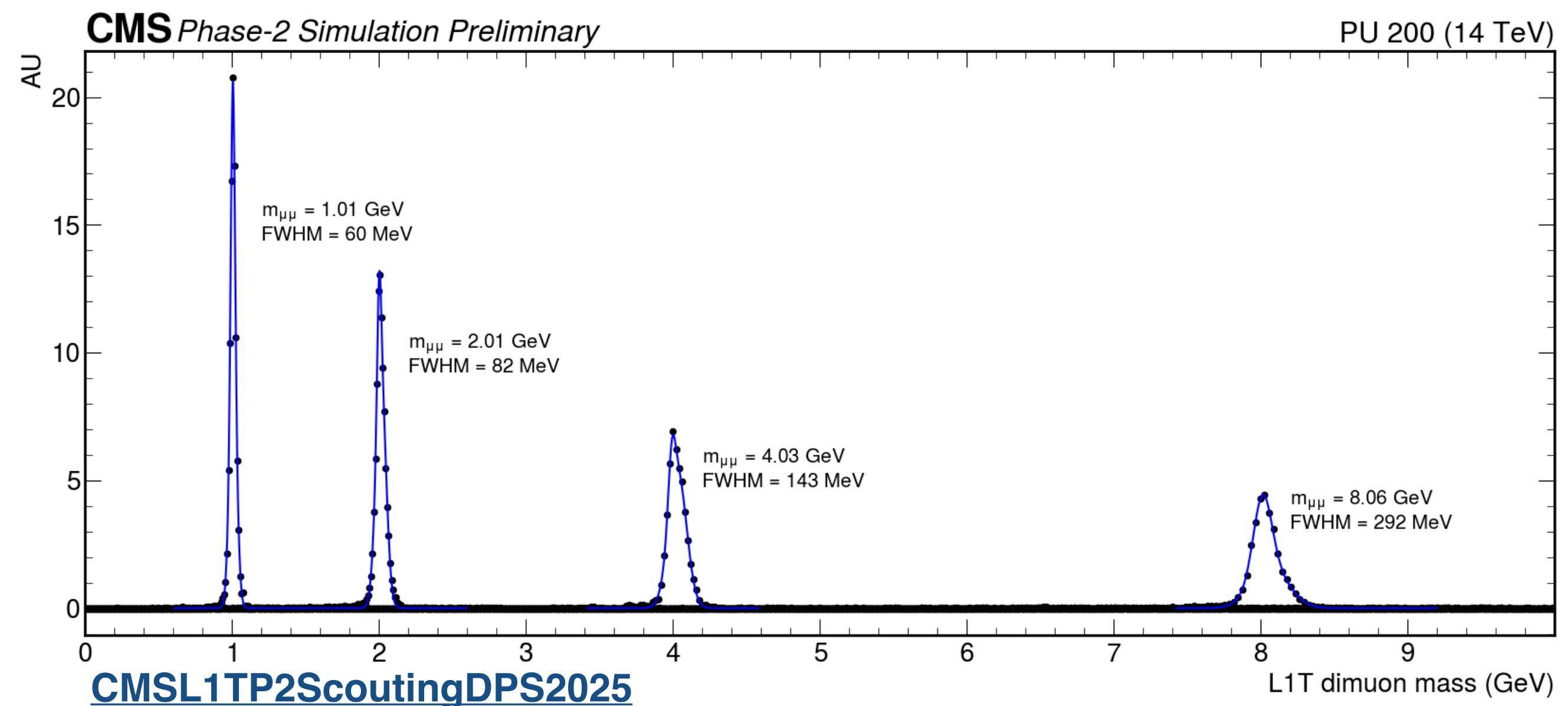
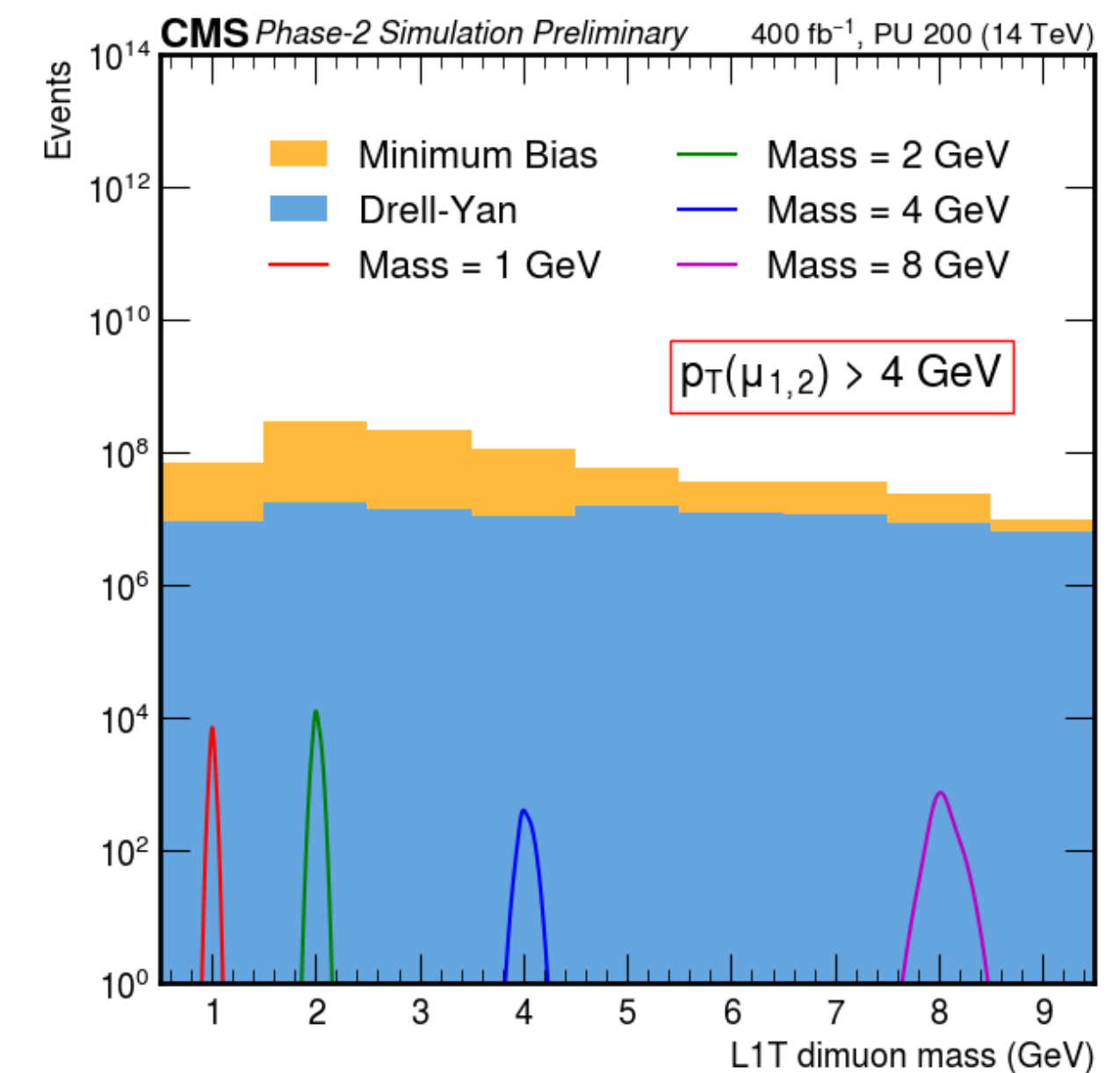
- Investigating final state with multiple soft hadronic objects
 - Soft final states**, not selected efficiently by the standard L1 Trigger
 - Small signal yield**, but high multiplicity of soft b/c jets, hadronic taus, or tracks
 - Might need **dedicated soft-object reconstruction & ML to extract signal**
- $B_s \rightarrow \phi\phi \rightarrow K^+K^-K^+K^-$
 - Rare low p_T , fully hadronic SM process, with branching ratios $BR(B_s \rightarrow \phi\phi) \approx (1.85 \pm 0.14) \times 10^{-5}$ and $BR(\phi \rightarrow KK) \approx (49.1 \pm 0.5)\%$
 - The final state will be **accessible in L1T by CMS for the first time during HL-LHC, with L1T tracking**
 - Simple cut based analysis using L1T tracks and PF candidates to achieve 1.1×10^{-4} background reduction with 19% signal selection efficiency
 - Further improvements are being studied: **track isolation, distance of closest approach, ML based selection...**
- Other ideas we are considering:
 - Multiple Soft Hadronic objects**, multiple new states (from 5 to 50) with compressed mass spectra, small couplings
 - SUEPs-like signatures, instantons** using L1T extended tracking
 - Slow or long-lived objects** with correlating signals **across multiple BXs**



CMSL1TP2ScoutingDPS2025

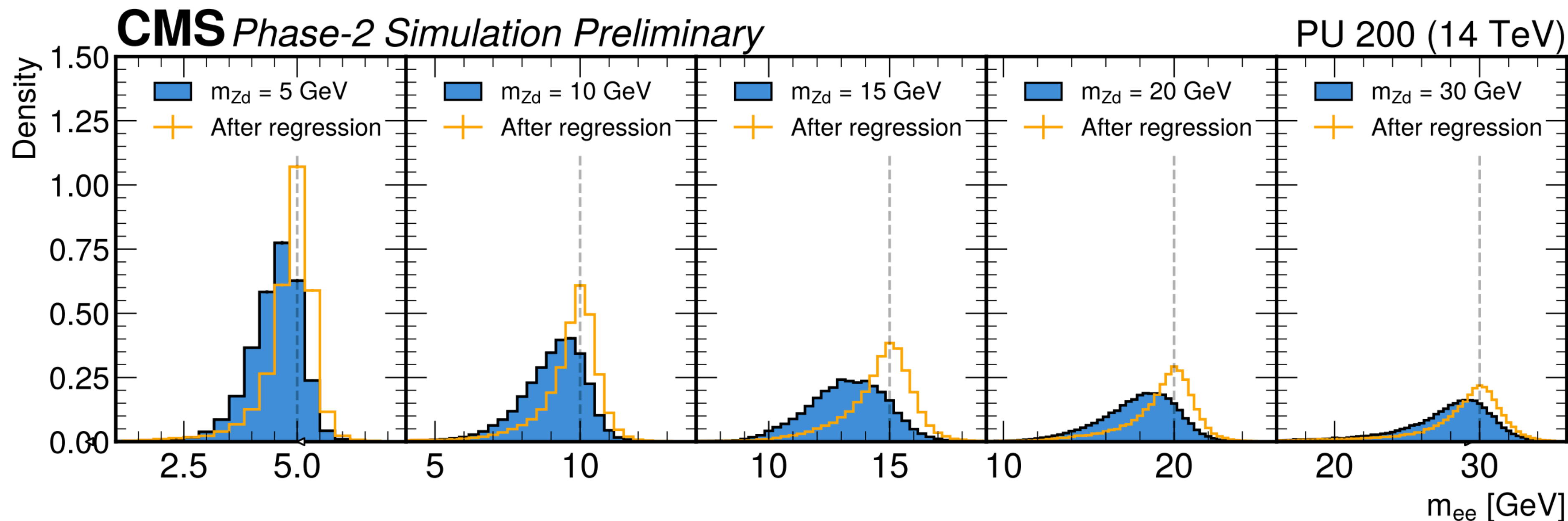
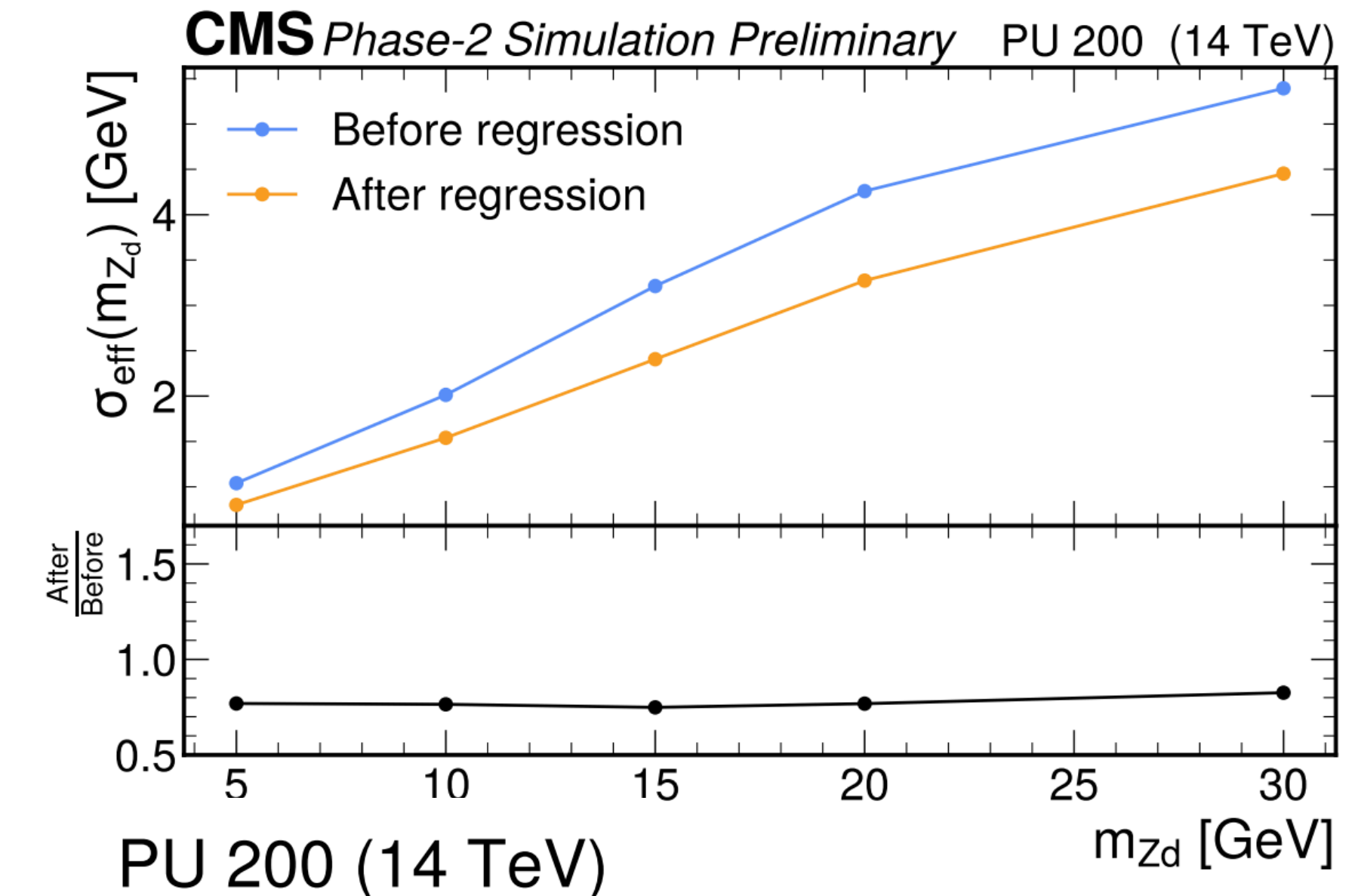
Going Beyond the Baseline: Analyses

- Investigating low-mass narrow resonances decaying into $\mu\mu$, ee , $\tau\tau$, $\gamma\gamma$, qq/bb with common characteristics
 - Soft final states**, not selected efficiently by the standard L1 Trigger
 - Small signal yield**, but visible as narrow peak over smooth background and **large “irreducible” backgrounds**
- $X \rightarrow \mu\mu$: expanding HLT Scouting acceptance ([CMS-EXO-21-005](#)) using good resolution of tk-matched muons
 - Using all L1 tk-matched muons ($p_T > 2$ GeV, $\eta < 2.4$) extending the reach to very low-mass resonances
 - Simple cut based analysis using kinematic quantities and muon isolation. **Very good dimuon mass resolution** enables large background rate suppression
 - In the pure category ($p_T(\mu_1, \mu_2) > 4$ GeV) achieved 5×10^{-6} background reduction with $> 98\%$ signal selection efficiency for all mass points when the generated muons are within L1T acceptance ($p_T > 2$ GeV, $\eta < 2.4$).
- $X \rightarrow ee$, $\gamma\gamma$ further expand HLT Scouting acceptance using all L1 tk-matched electrons and isolated photons
- $X \rightarrow \tau\tau$ further expand HLT Scouting acceptance ([CMS-EXO-24-012](#))
- Investigating di-jets resonances to extend further HLT Scouting reach
 - Can use any two PUPPI jets ($\sim 90\%$ efficient for $p_T > 30$ GeV in $t\bar{t}$ bar)
 - Can also exploit **flavour tagging** and **large radius jet reconstruction** with PUPPI jets to target specific final states
 - Can improve further in Scouting with: **jet p_T regression** (as offline for b-jets), **more complex tagging algorithm**



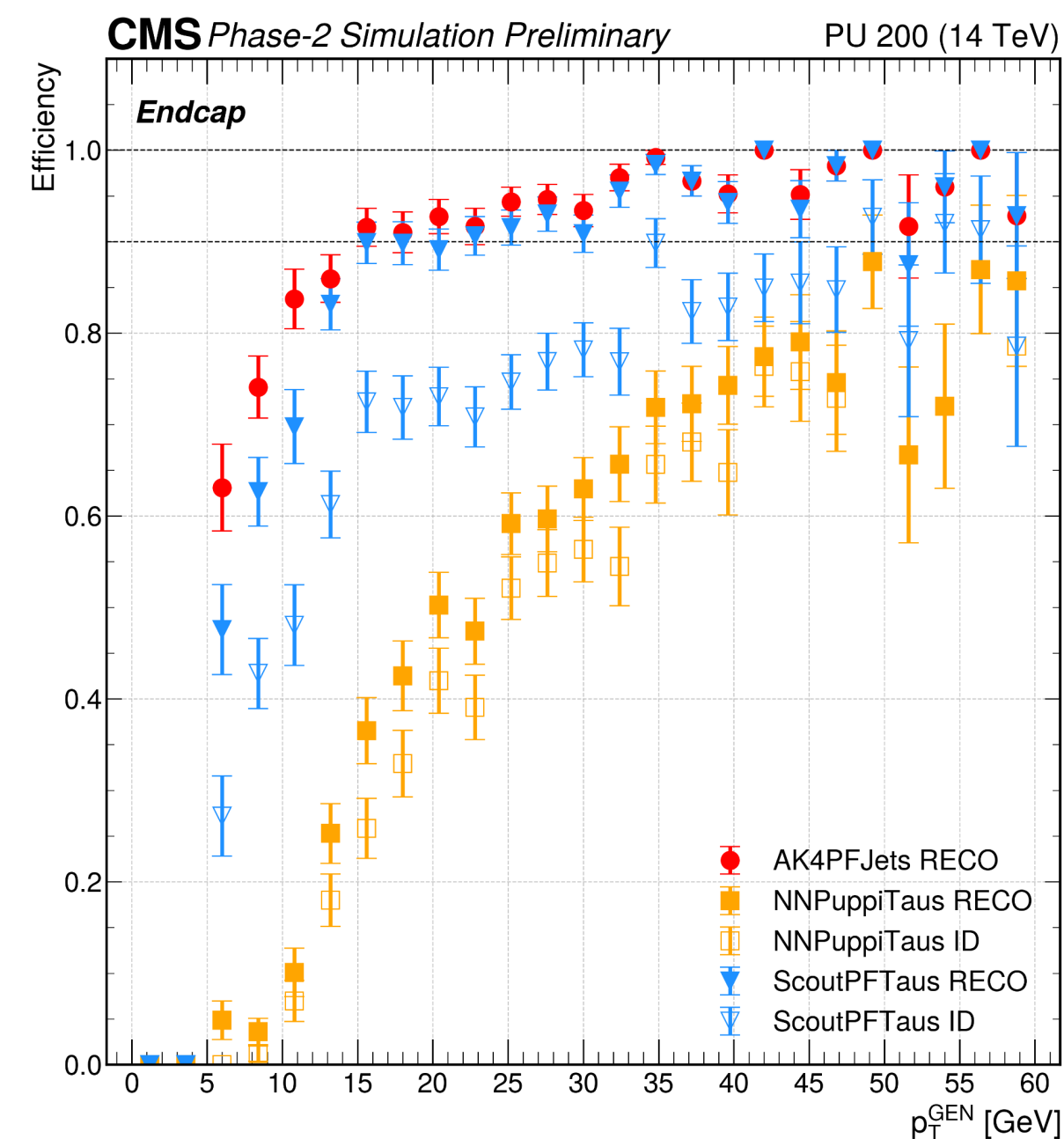
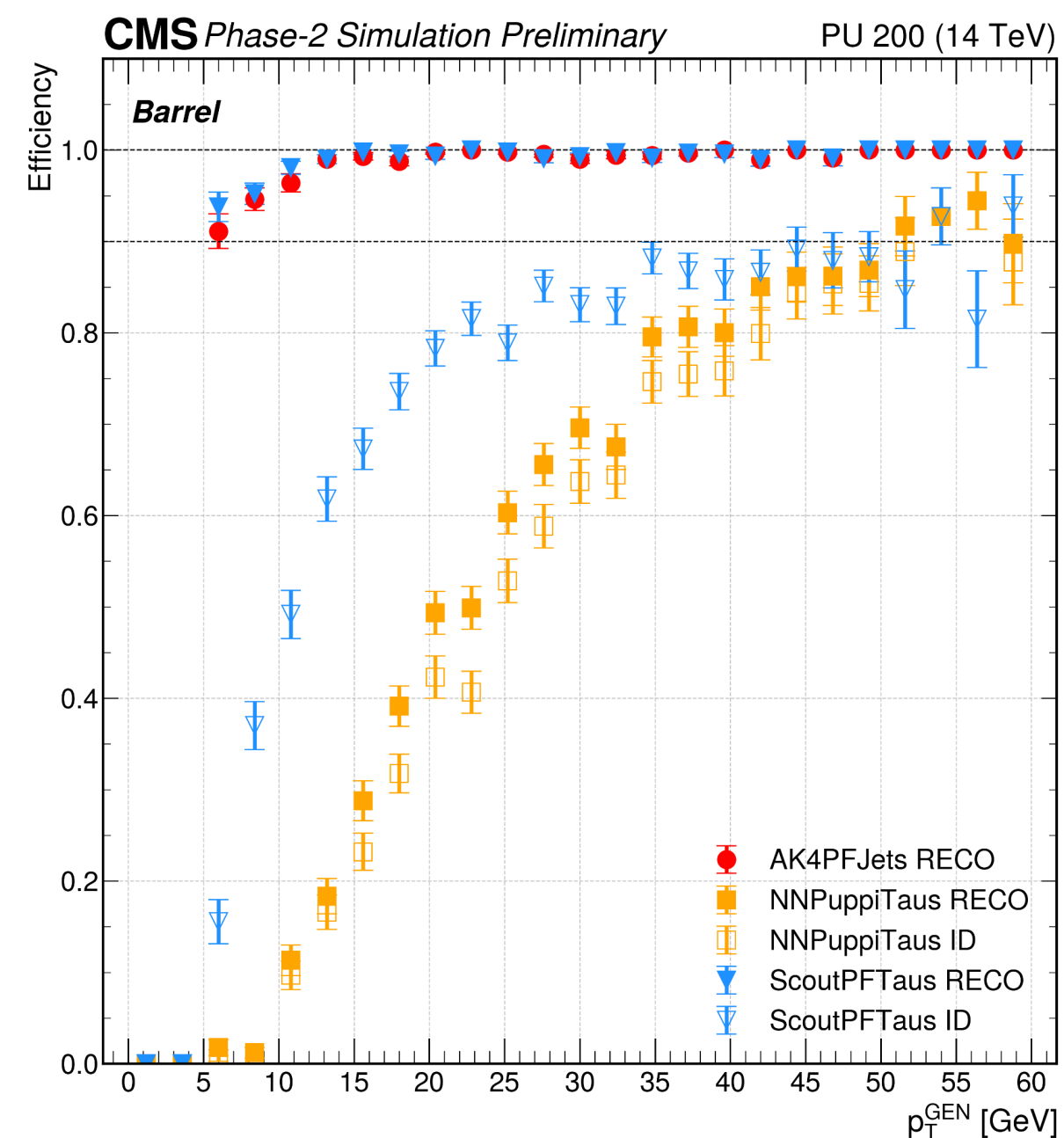
Going Beyond the Baseline: Improving Objects

- **A novel algorithm to improve the transverse momentum reconstruction of tk-matched electrons** in the barrel region was developed
- The algorithm regresses the p_T of tk-matched electrons by running **ML inference on FPGAs**
 - The different detector technology used in the calorimeter endcaps requires a dedicated model that will be explored in the future
- By training the model using signal + background we can **improve the low-mass dielectron invariant mass for $X \rightarrow ee$ searches** while keeping the trigger rates for low- p_T electrons the same



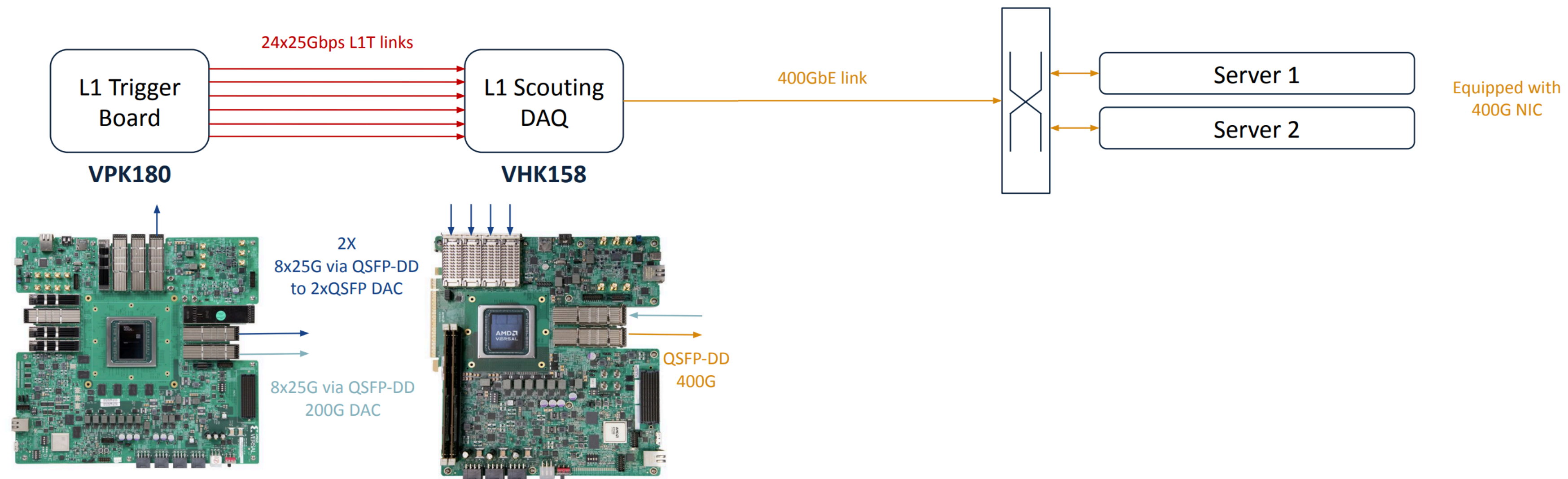
Going Beyond the Baseline: Extending the Inputs

- PF-based inputs can provide higher efficiency for soft taus from low-mass resonances as the selection of PUPPI candidates is heavily dependent on the quality of primary vertex reconstruction.
- We explore **a new clustering algorithm, CLUE**, as a replacement for the *anti- k_t* , since CLUE is specifically designed for **parallelism and being GPU-friendly**, which can be used to take advantage of the heterogeneous computing of the Phase-2 L1T Scouting online processing step
- ScoutPFTaus, reconstructed by CLUE, are shown to have **similar efficiency to AK4PFJets while having the advantage of parallelization**.
- In addition to CLUE based clustering, **a new transformer-based NN ID was developed** to separate genuine taus from fake taus.
- The improvement from using PF candidates instead of PUPPI for tau reconstruction is very clear throughout the plot, in particular at lower p_T , improving software taus that will be used by low-mass $X \rightarrow \tau\tau$ searches.



Going Beyond the Baseline: DAQ

- Building a new DAQ system/board based on newer technologies: AMD Versal, 400Gbps networking, ...
 - Testing generating the 24 trigger links from the VPK180 and receiving them on the VHK158
 - Exploring new protocols for the transmission, e.g. RoCEv2
 - Working implementation of 400G on VPK180 and VHK158



Generate the 24 trigger links from the VPK180 and receive on the VHK158

Going Beyond the Baseline: Accelerators

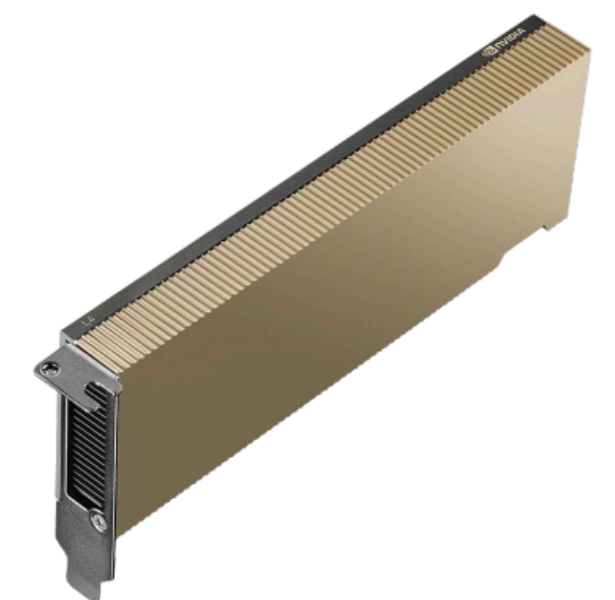
- Baseline system processes data on CPUs
- We are exploring the benefits of using **accelerators for data processing and physics analyses**
 - **AMD AI Engines, FPGA accelerator cards, and GPUs to offload complex tasks and ML models**
- Preliminary results are obtained from implementing benchmark physics analysis, $W \rightarrow 3\pi$, with **unpacking PUPPI objects, first level filtering and particles isolation, and combinatorics to identify the correct triplet**
- **Results look good on accelerators**, there is no clear advantage between different accelerator types for now



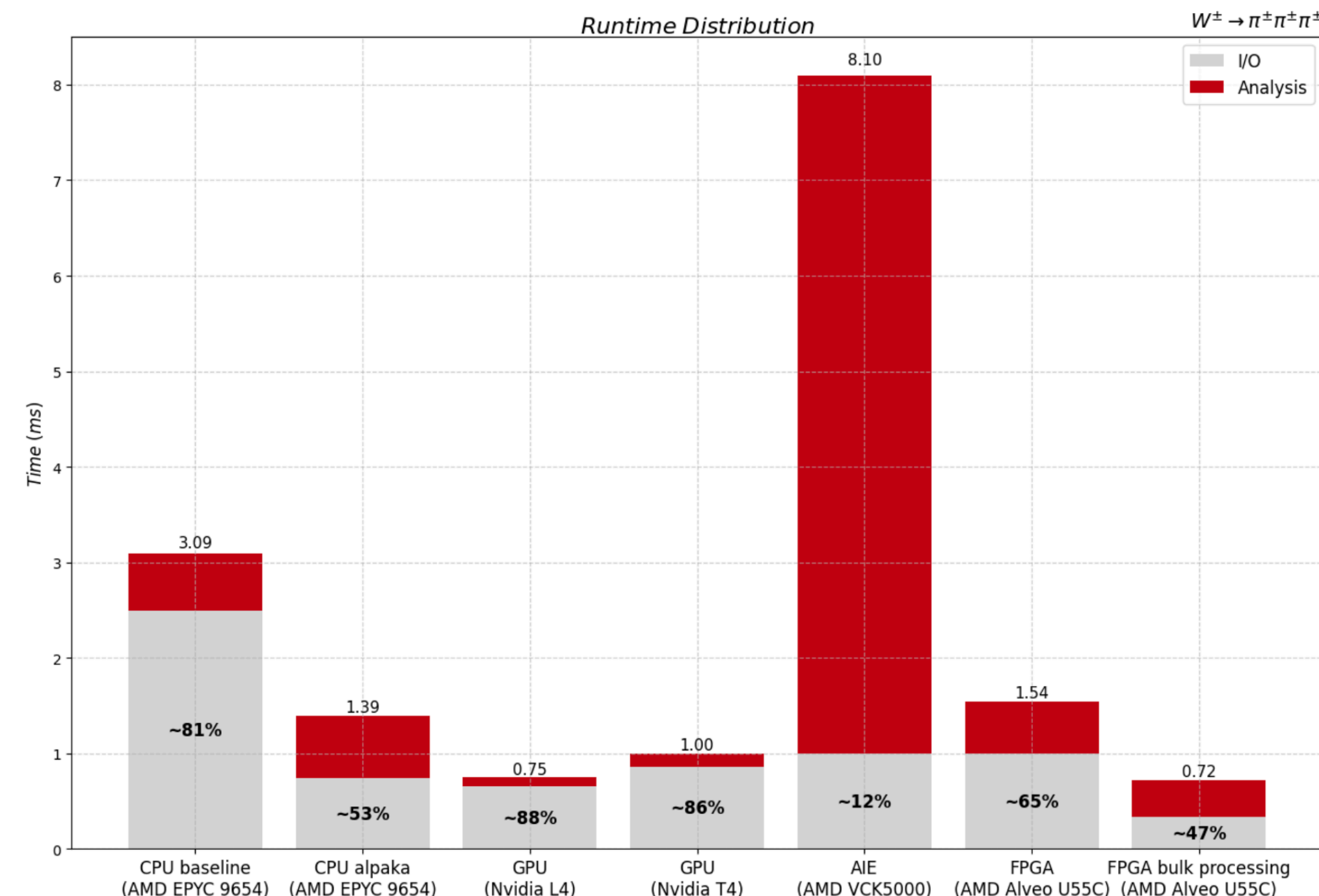
AMD VCK5000



Alveo U55C



NVIDIA L4 GPU



[G. Zago's Thesis](#)

[L. Sieder's Thesis](#)

Conclusions

- CMS L1T for HL-LHC will be a very capable system **having close-to-offline resolution with the inclusion of tracking and particle flow**
- L1T Data Scouting for HL-LHC promises to **complement the standard CMS HL-LHC physics program** by enabling **trigger-level physics analyses at 40 MHz**
- The “**baseline**” **demonstrator of the scouting system is developed** that performs collection of L1 trigger objects and running first set of prototype analyses online
 - Already capable of receiving **1/7th of the inputs of the full system, running 10 analyses online, with an input rate of ~18 GB/s** to the servers which is more than the Run 3 CMS HLT output rate.
- We are now exploring the possibility of expanding the current system
 - Working on **new analyses to explore possible new physics opportunities**
 - Can also benefit from **improved object performances and extending the inputs to the L1T data scouting** (e.g. PF candidates)
 - Exploring the use of **Versal devkits** capable of handling higher link count and speed, and **new data transmission protocols**
 - Working on **accelerators for physics analyses and running complex algorithms such as ML models**
- Many more studies are in progress to explore the full potential of L1T Data Scouting for HL-LHC