System Design and Prototyping for the CMS Level-1 Trigger at the High-Luminosity LHC

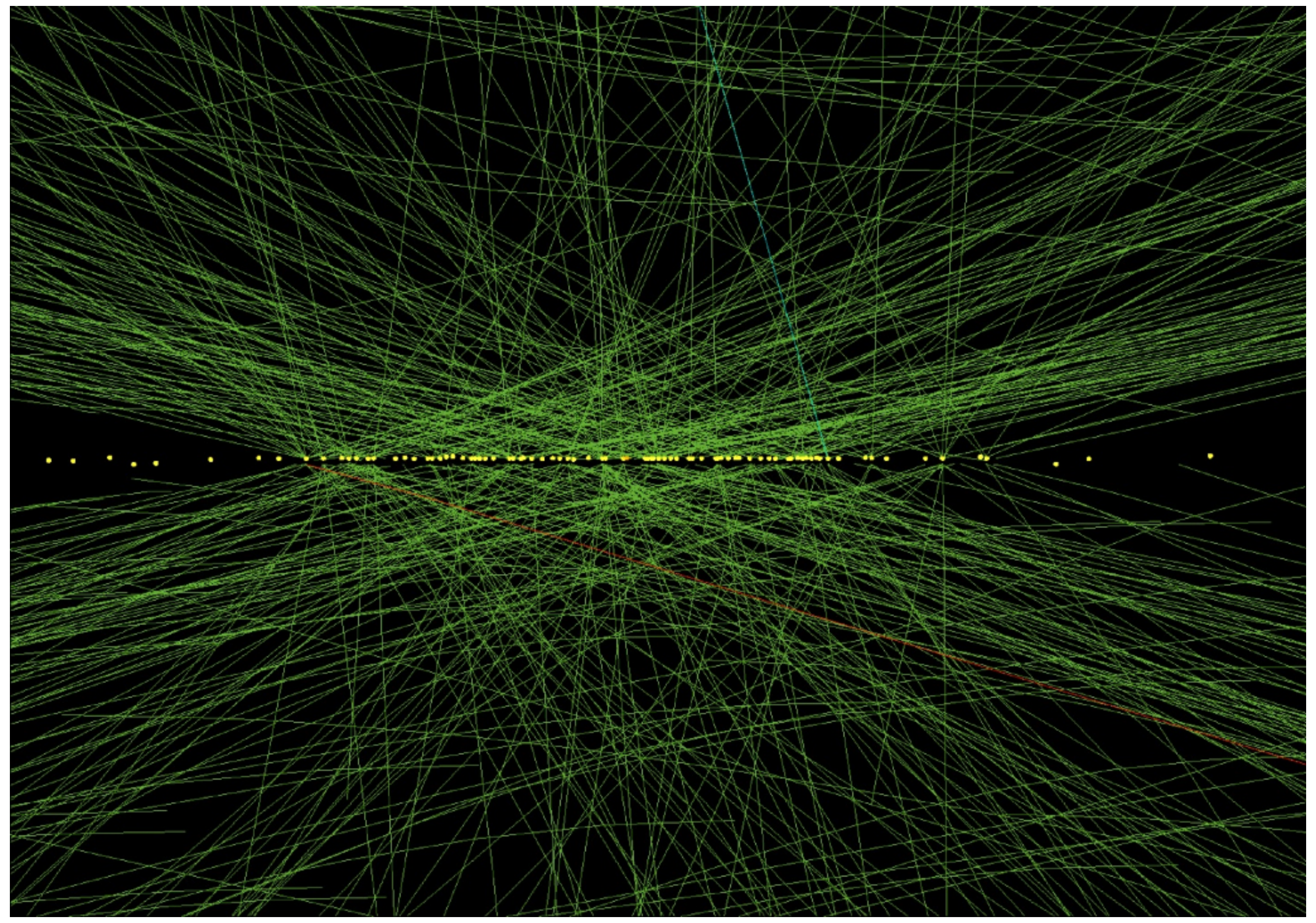
Sioni Summers for the CMS Collaboration

TWEPP, Glasgow
October 1 2024

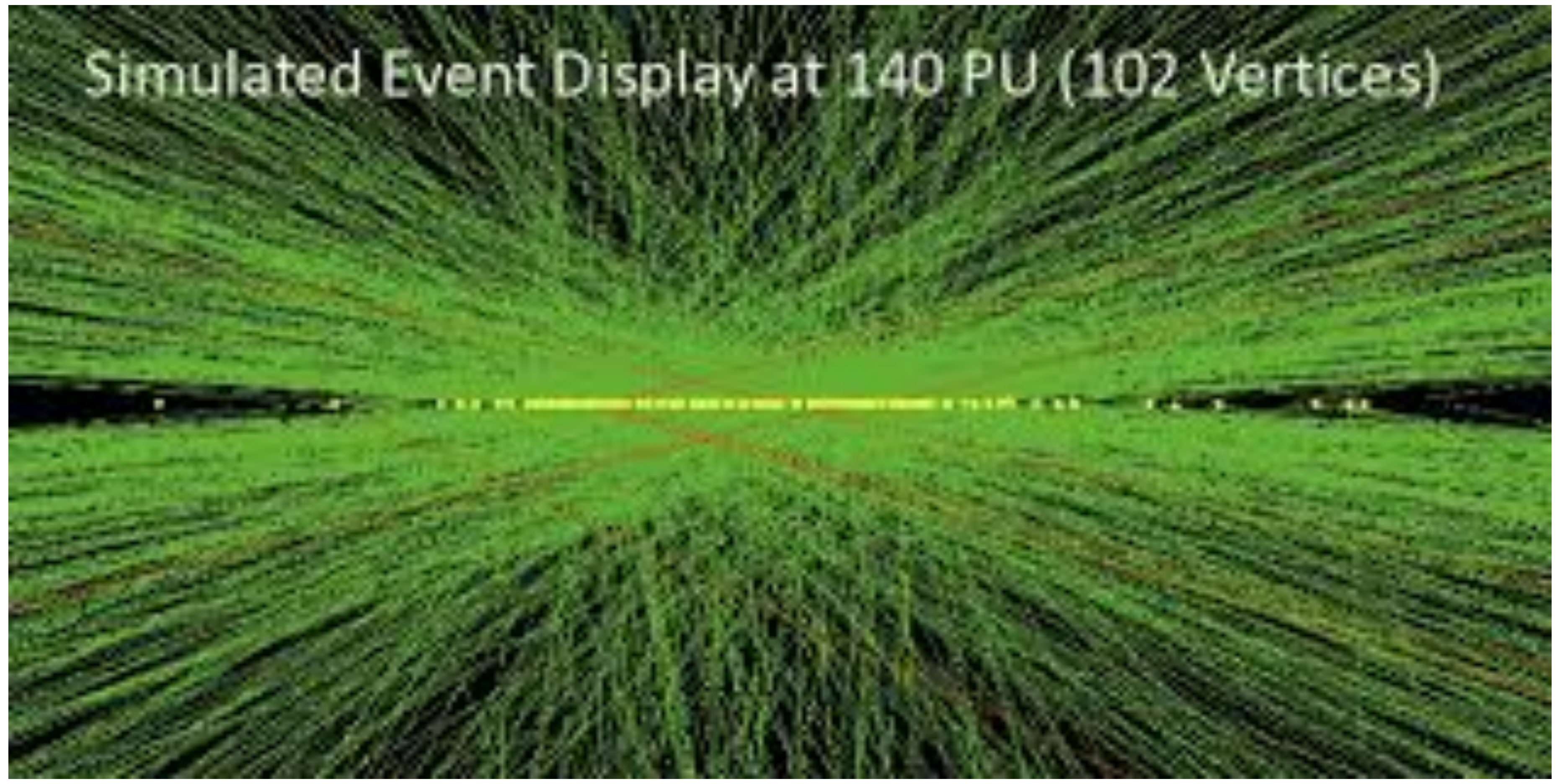
Introduction

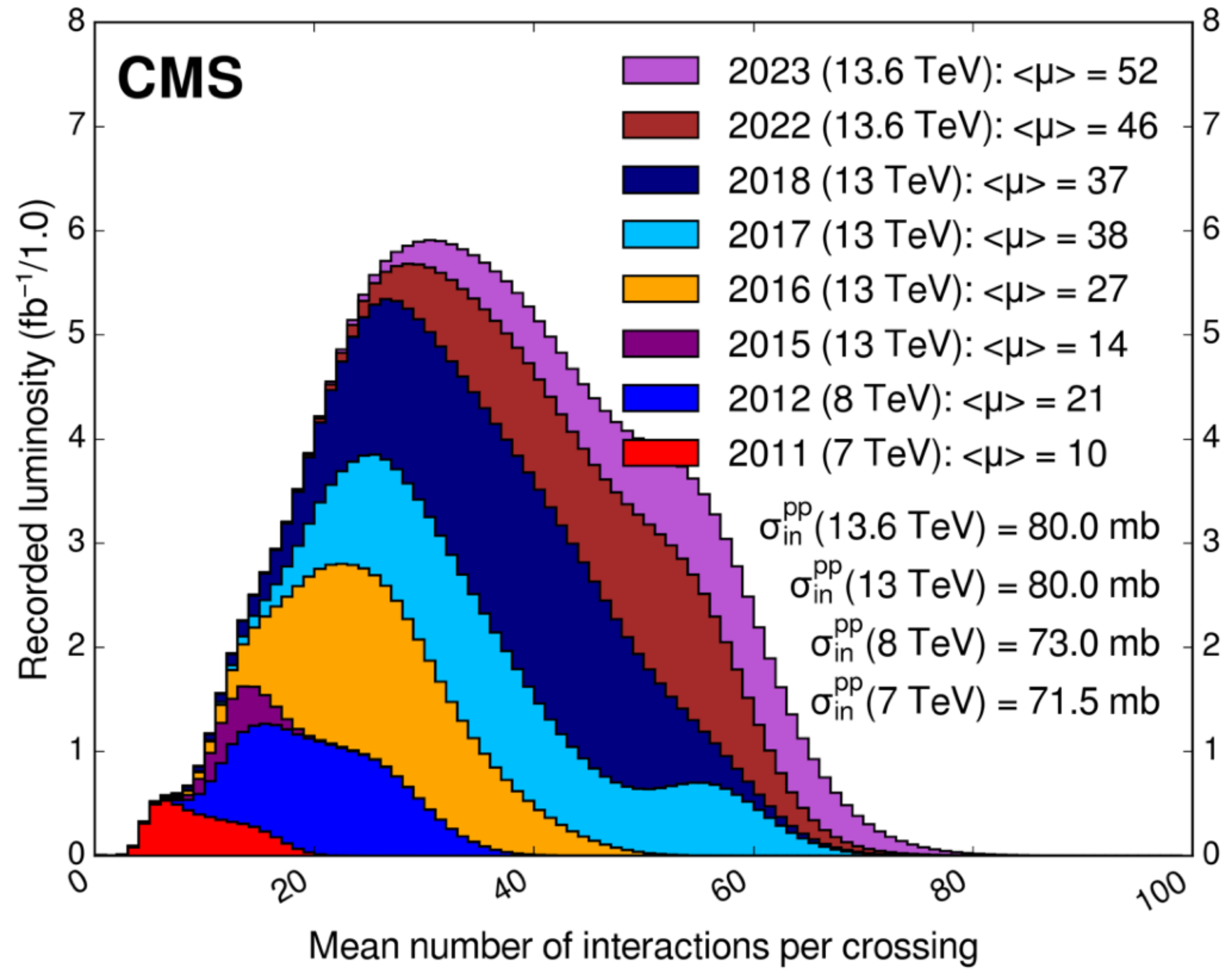
- High-Luminosity LHC will increase the number of simultaneous p-p collisions (pileup) **up to 200**
- CMS will **upgrade detectors** for the HL-LHC era: new tracker with tracking at L1T for first time, new endcap High Granularity Calorimeter (HGCAL), upgrades to muon detectors
- Phase 2 Upgrade of CMS L1T will **select 750 kHz events from 40 MHz** for further reconstruction and selection at High Level Trigger
    - **Maintain Run-3-like trigger thresholds** for standard single/double-object triggers (jets, electrons, muons, taus, missing transverse momentum)
    - **Add new algorithms and techniques** to extend CMS physics acceptance compared to Phase 1
    - Adapt and evolve as needs of experiment change

⬇ 78 pileup vertices



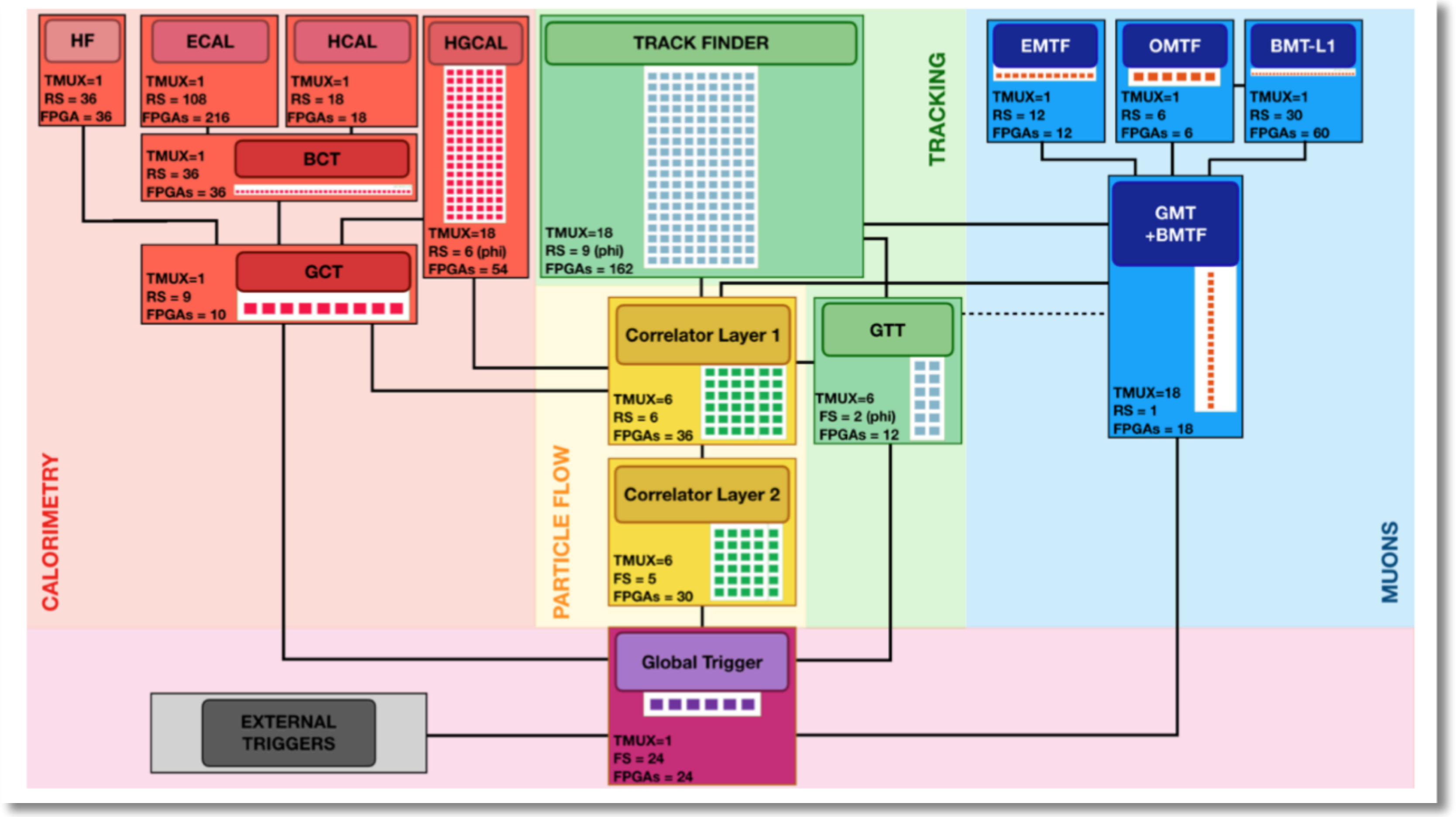
⬇ 140 pileup vertices





CMS Level 1 Trigger - System

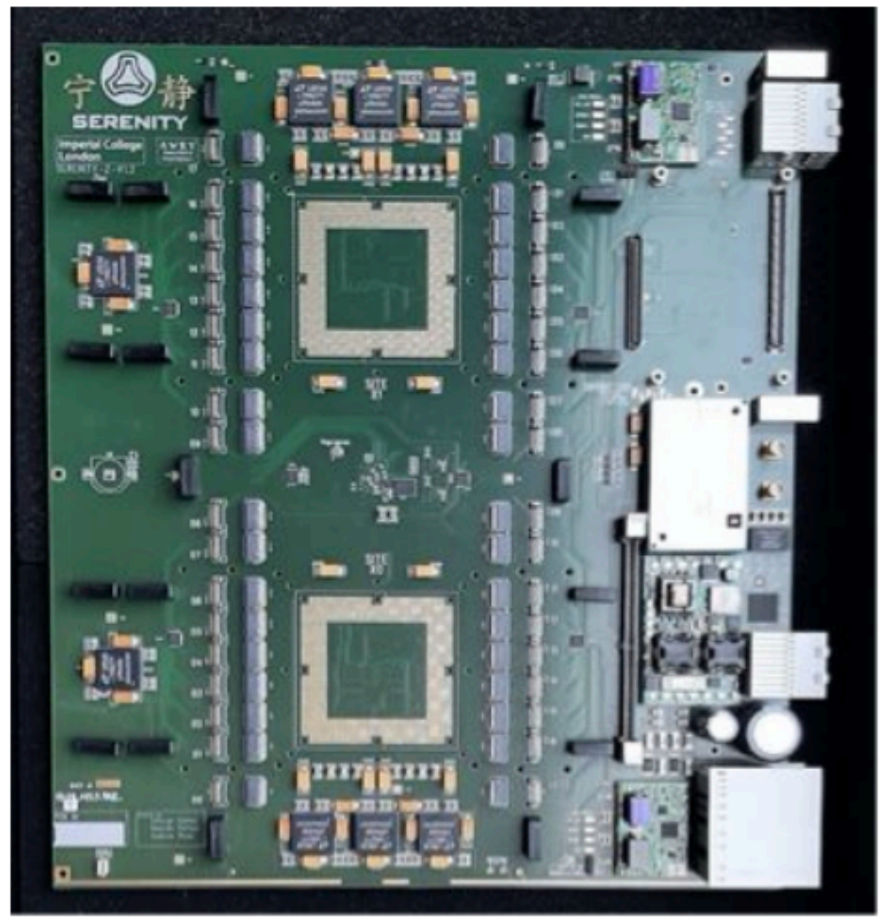
- Phase 2 Upgrade of CMS L1T will select **750 kHz events from 40 MHz** for further reconstruction and selection at High Level Trigger - up from 100 kHz
  - Trigger decision for one event must arrive within **12.5 μs latency**
- System will comprise hundreds of custom electronics boards with powerful FPGA processors (shown below)
  - ATCA platform
  - High-speed optical data transmission between boards, up to 4 Tb/s per board



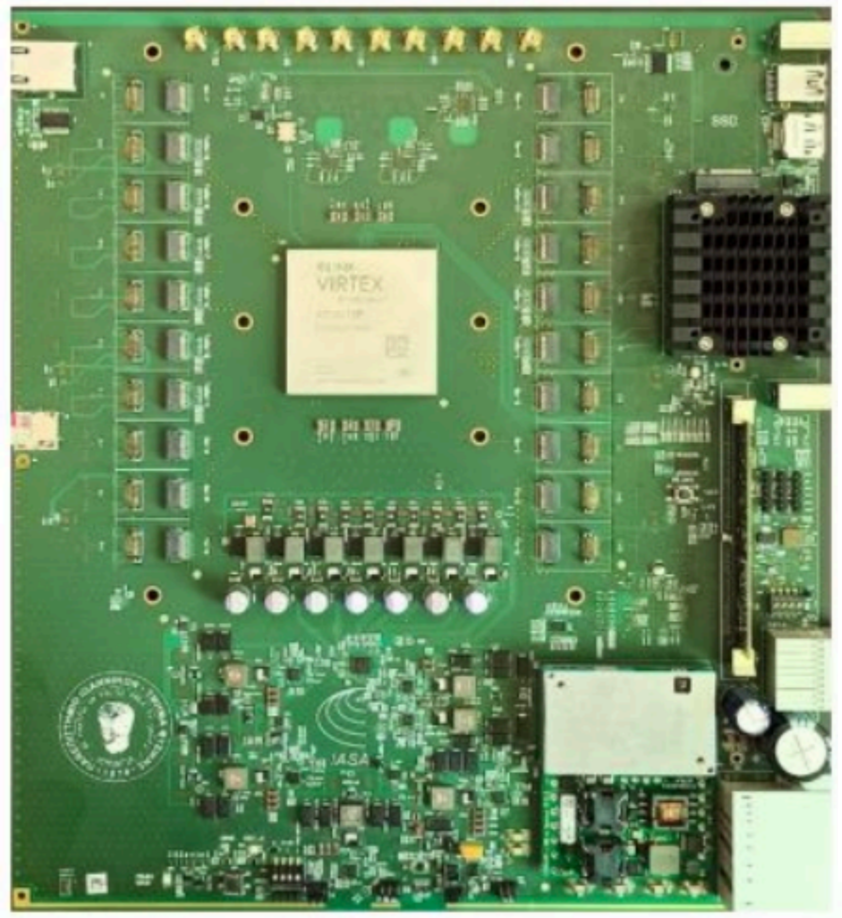
APx

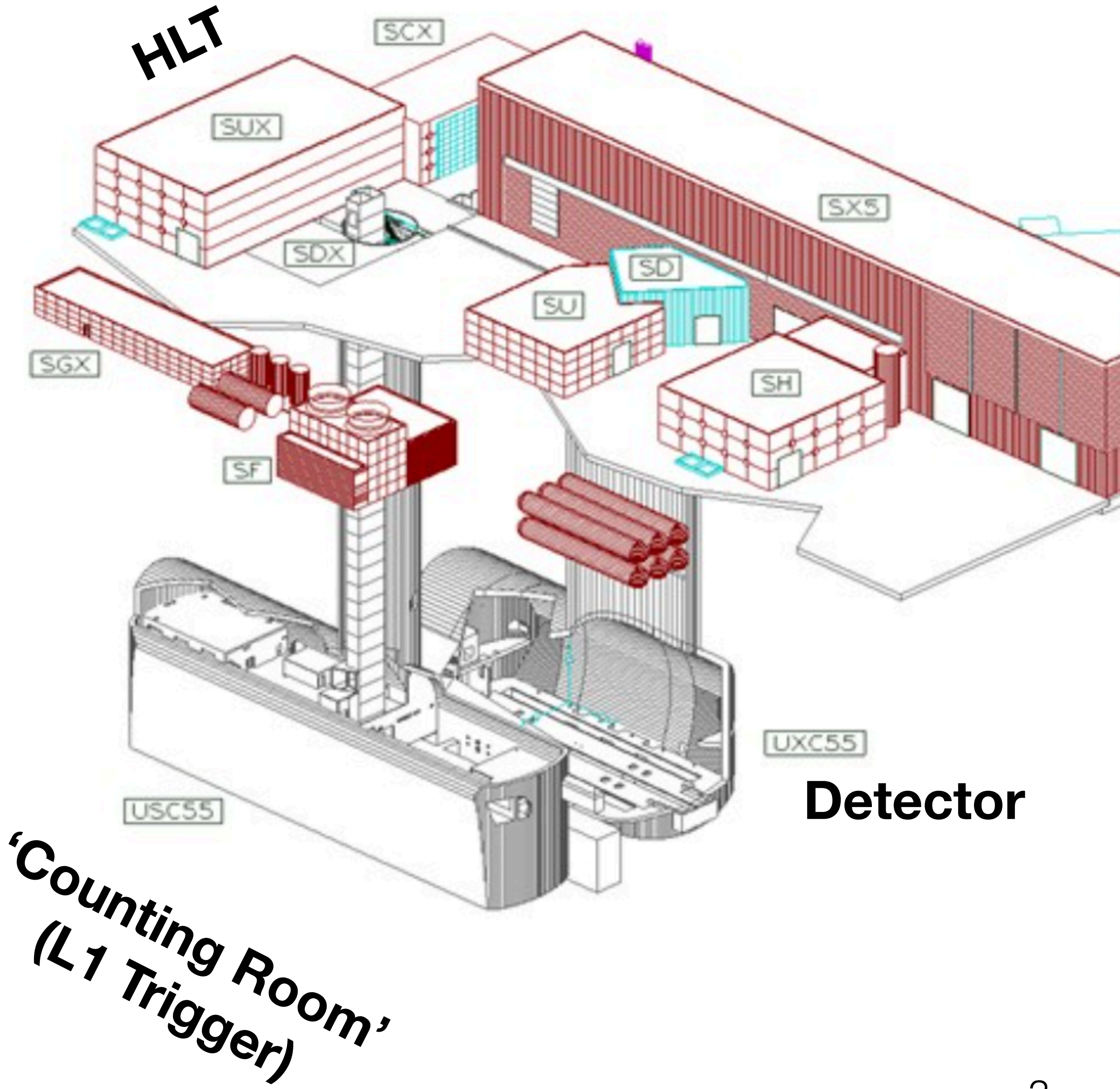
X2O

Serenity



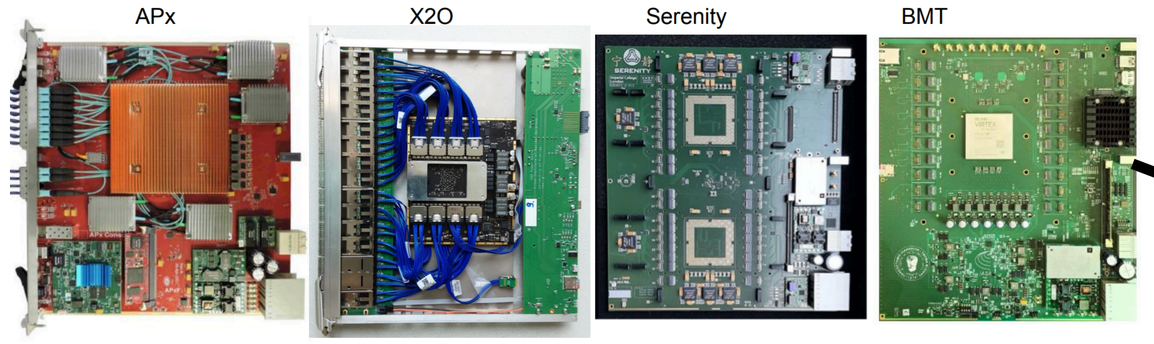
BMT



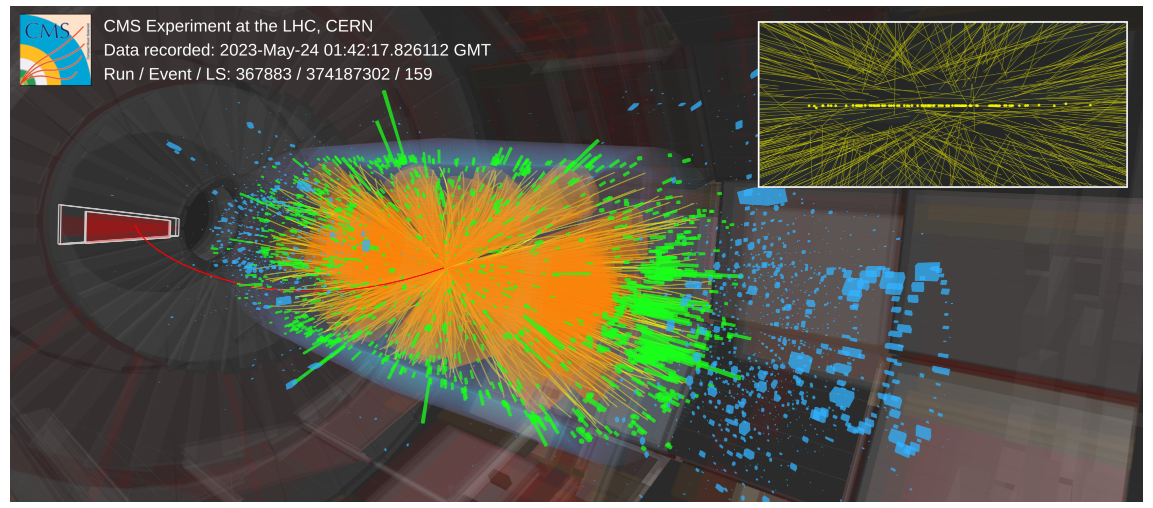


Trigger Processing

- How do we decide which events to keep? reconstructing high level information from low level detector information
  - Low level: raw detector hits (digitised measurements from sensors)
  - High level: particles, event-level quantities like total energy, jets (sprays of particles)
- Final decision compares the high level quantities with a “menu” of conditions to accept
- Processing mostly uses **physics algorithms** for reconstruction, and **Machine Learning**
  - We’ll discuss some of the algorithms, their implementations and prototyping

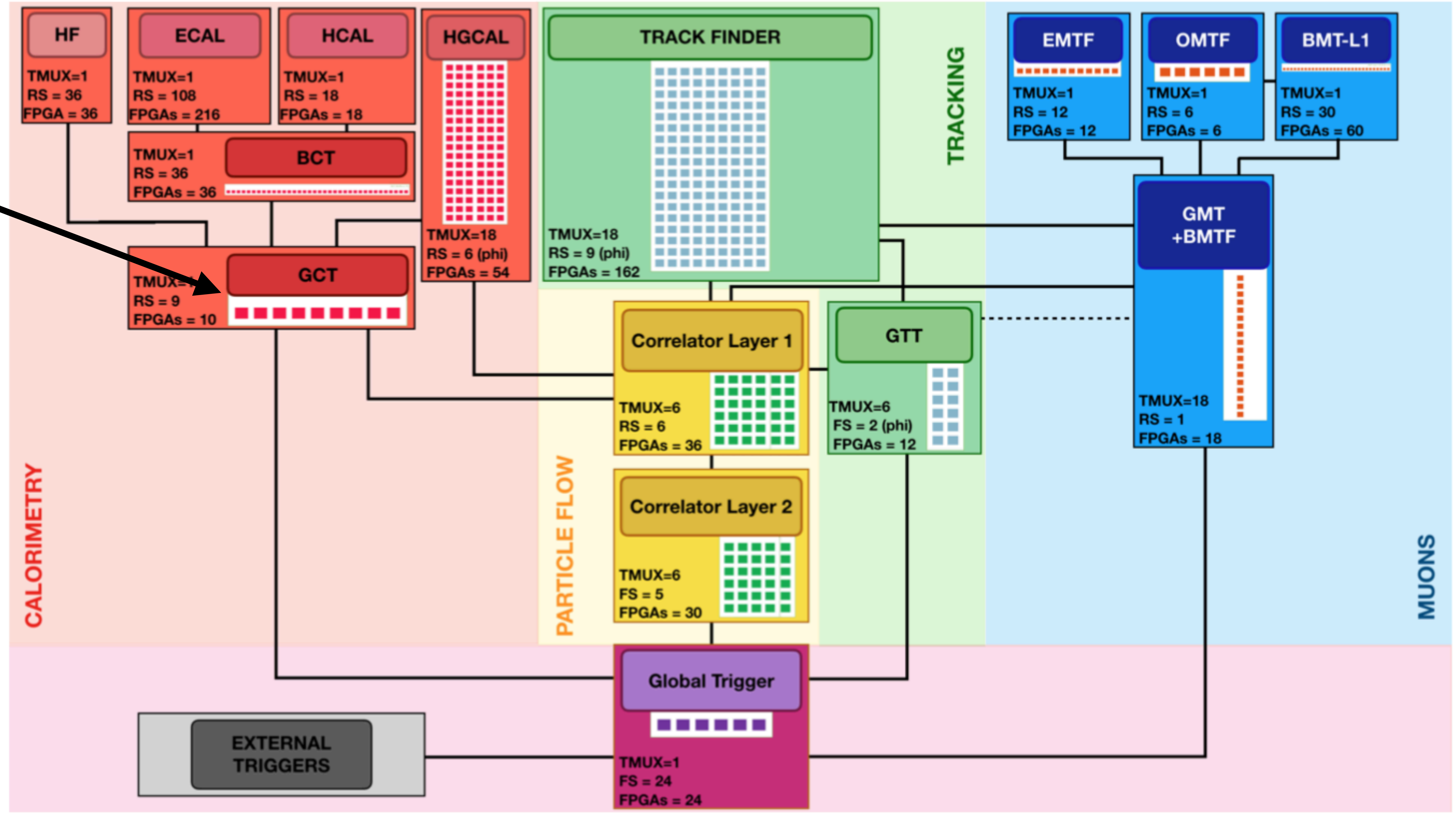


1 small box = 1 FPGA board with AMD VU13P FPGA

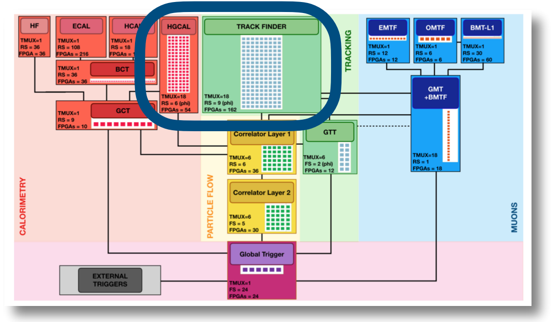


t = 0

t < 12.5 μs

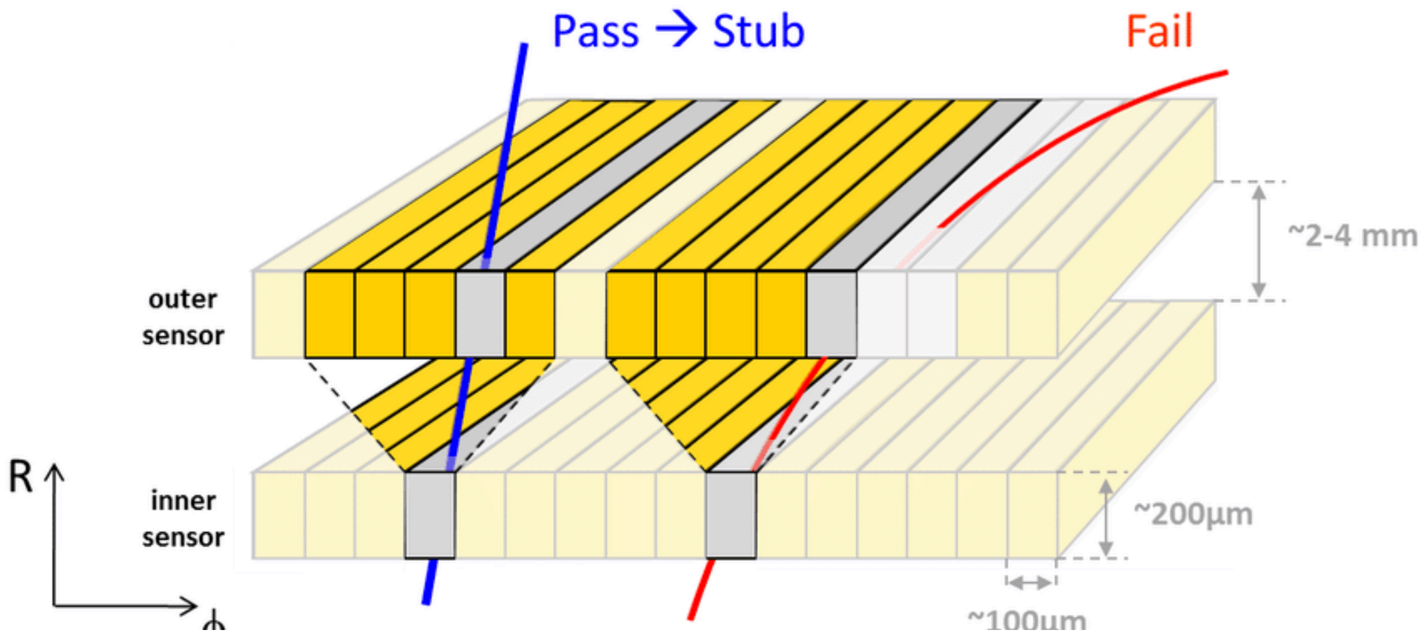


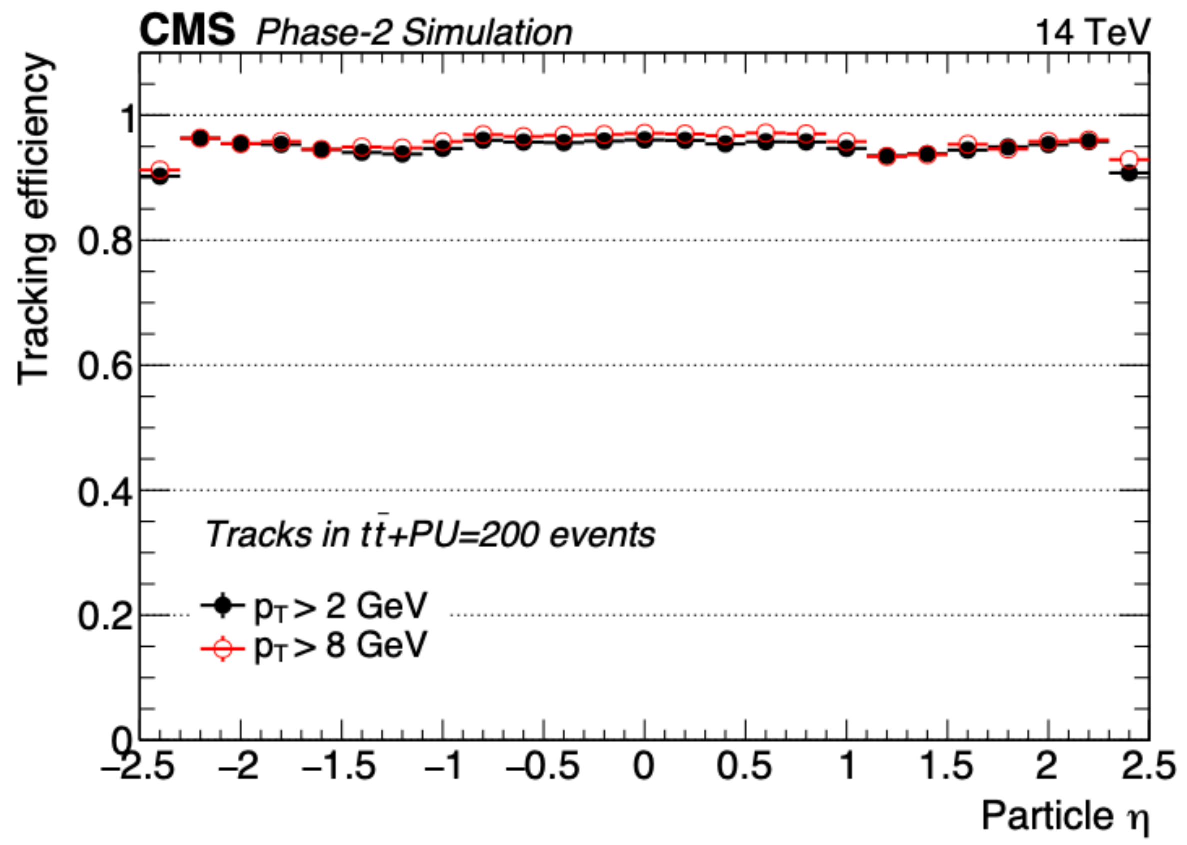
0 μs	**Detector hits**
5 μs	**Clusters & Tracks**
6 μs	**Particles**
7 μs	**Event Categorisation**
8 μs	**1 bit: keep / discard**

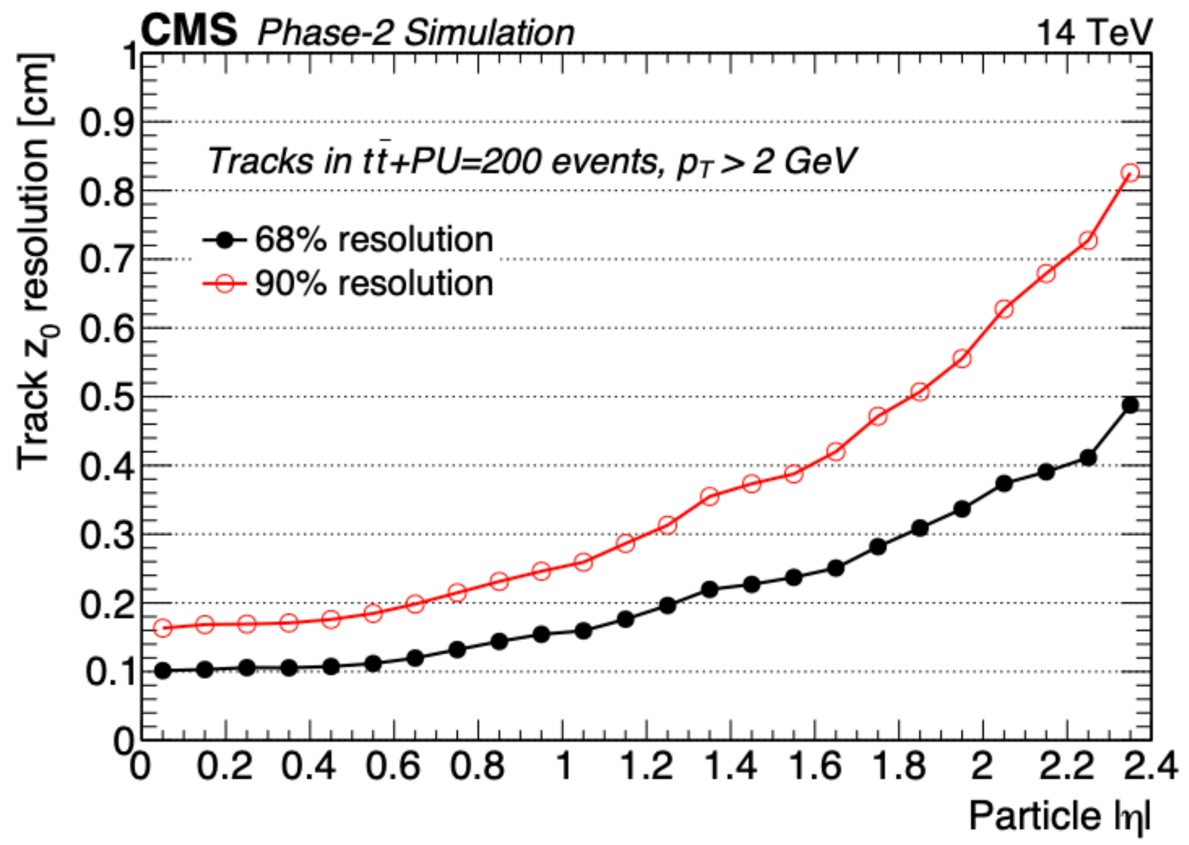


Phase 2: New Detectors

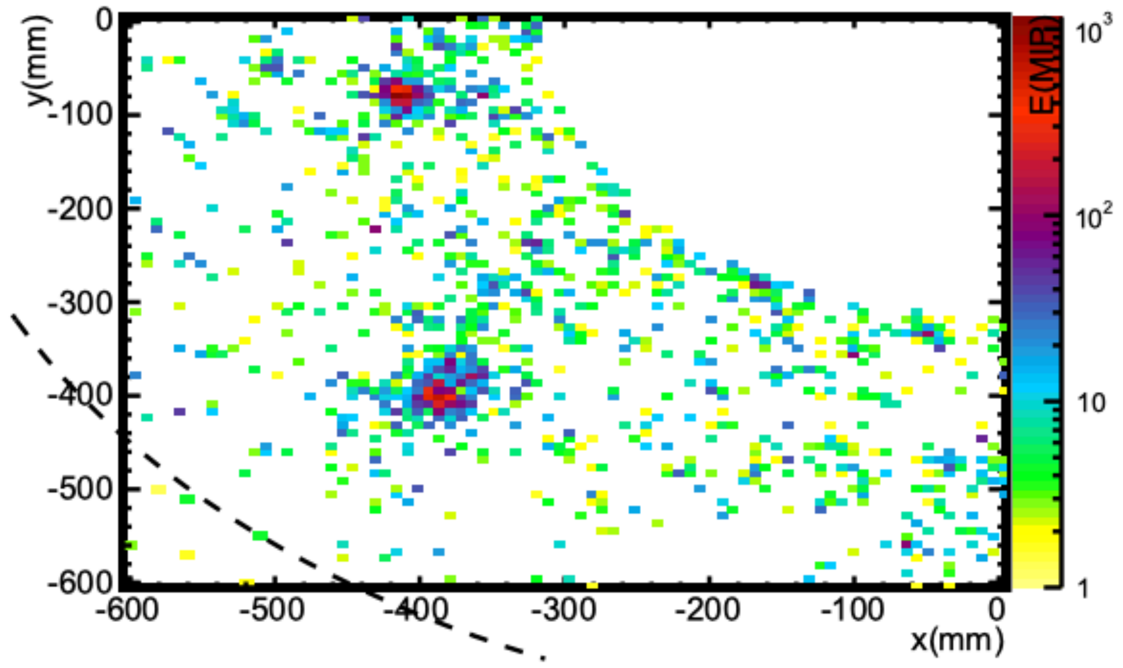
- **Track Reconstruction** at Level 1 Trigger for first time up to |η| < 2.4
- “Stubs” with p_T > 2 GeV will be sent to L1T from outer tracker
- Tracks in the Level 1 Trigger essential for 200 PU conditions
  - Primary vertex reconstruction, particle reconstruction
- L1T Track finding in around 200 FPGAs
- Seed finding, road building, track fitting

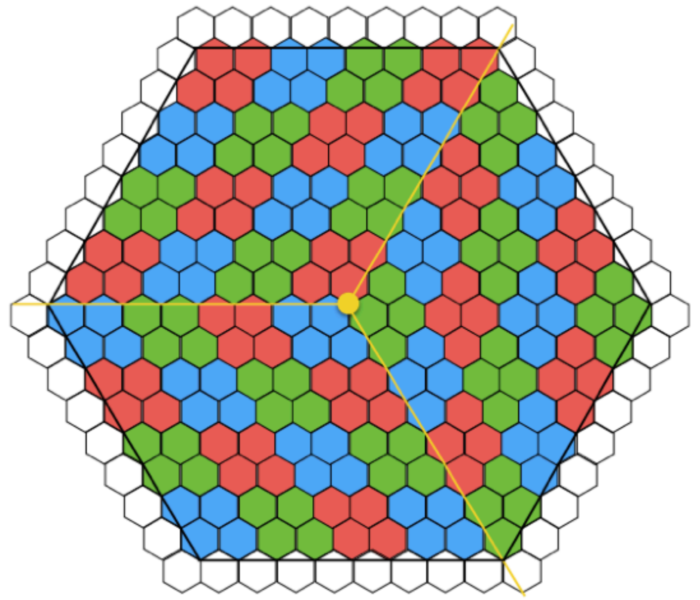


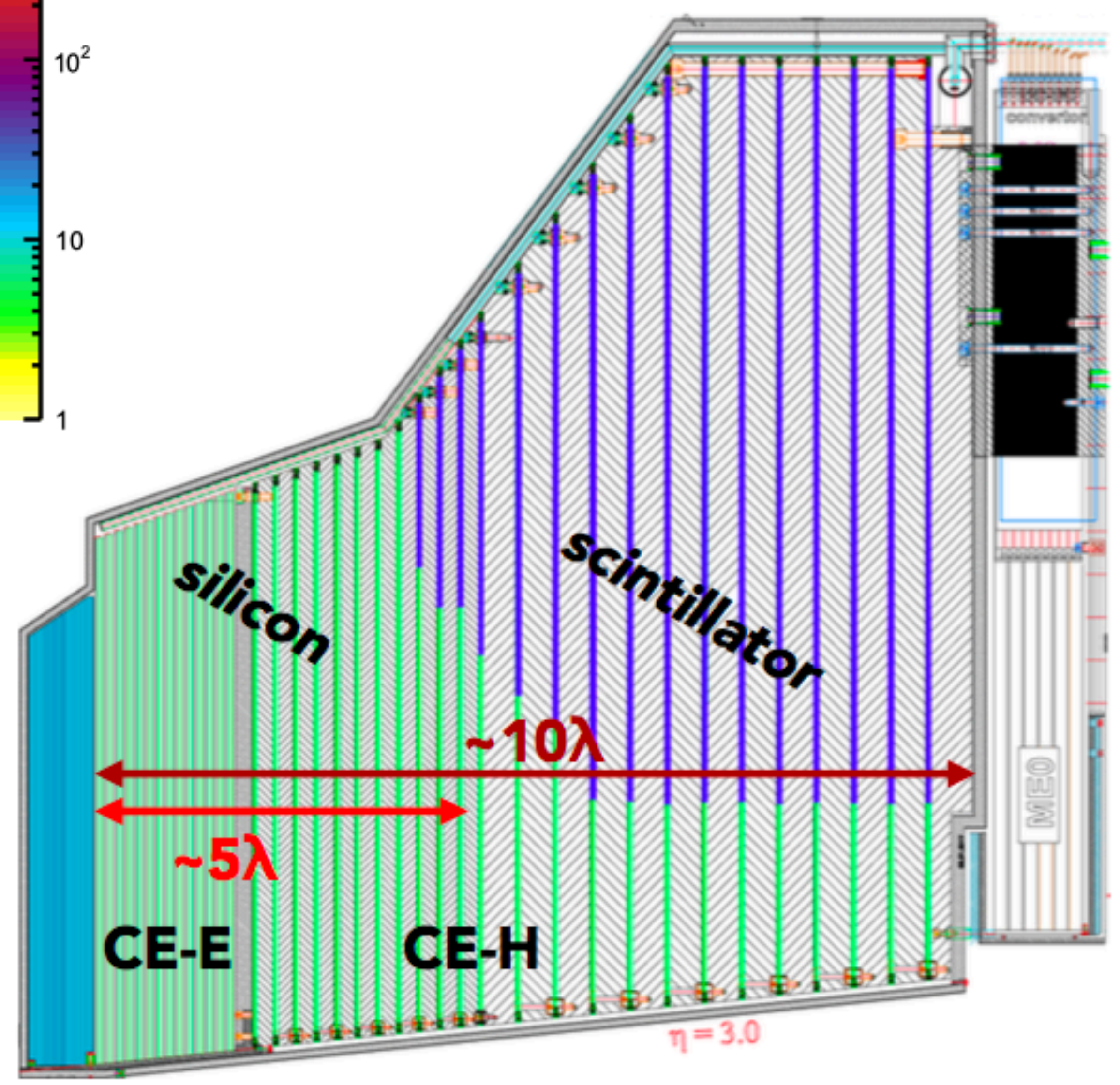




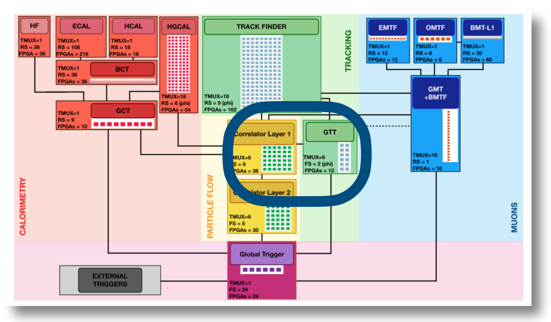
- **High granularity calorimeter**: silicon sampling calorimeter for the endcaps (1.5 < |η| < 3)
- 6.5 million channels (1 million to trigger) in 47 layers
  - Very fine transverse and longitudinal segmentation
- Around 200 FPGAs for 3D cluster reconstruction in L1T





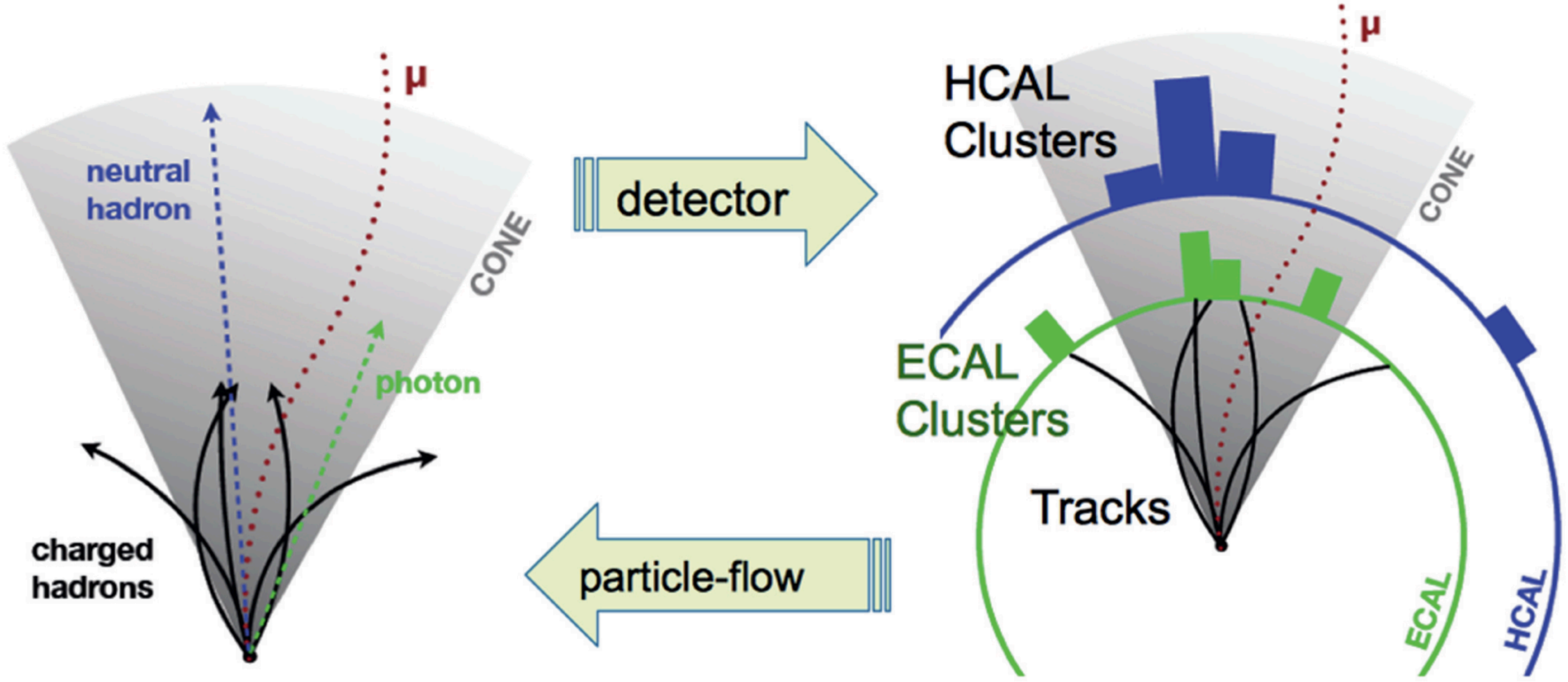


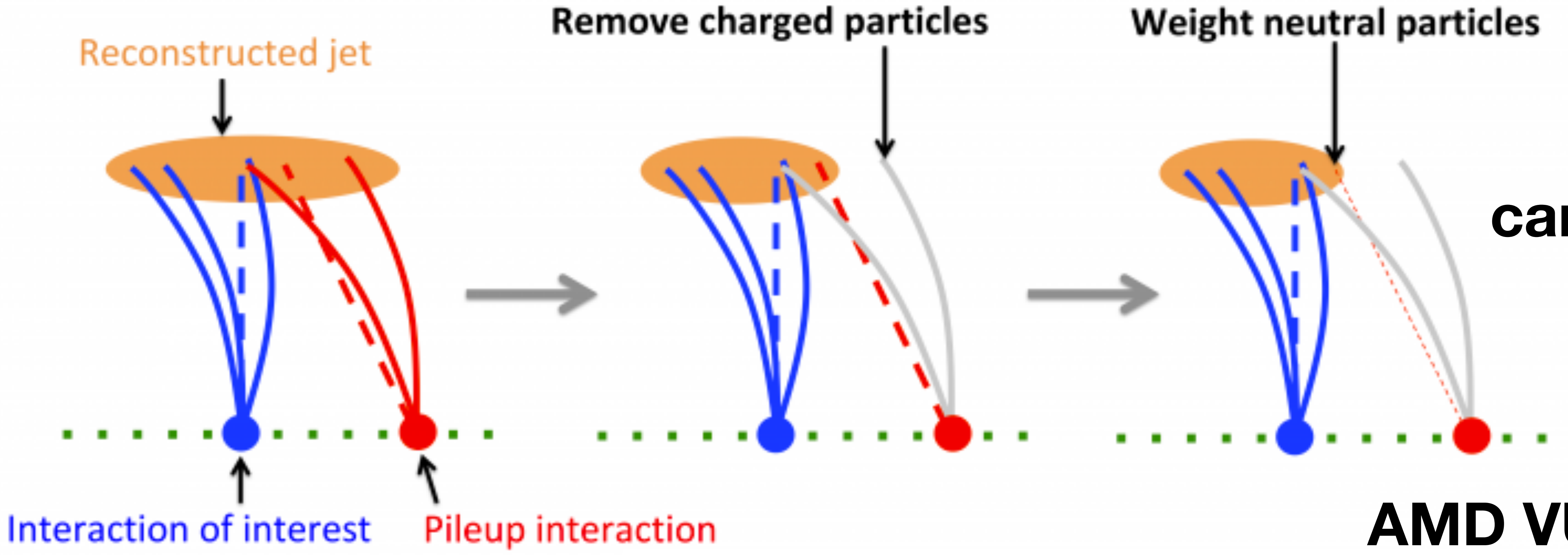
CMS-TDR-019
arXiv:1708.08234



Particle Flow, Vertexing, and PUPPI

- Particle Flow and PUPPI are principle CMS offline reconstruction algorithms, now to be at L1T
- Each sub-detector first performs **local reconstruction**
- ***Particle Flow*** links elements from different sub-detectors to reconstruct final state *particles*
  - Link tracks to calo. clusters for charged/neutral hadrons and electrons/photons; link tracks to muons
- ***Vertex Finding*** reconsts primary vertex from tracks
- ***PileUp Per Particle Identification*** (PUPPI) isolates the particles from the primary interaction
  - Using vertex association for charged particles
  - Nearby energy weighting for neutral particles
- Implementation splits detector into small regional chunks for parallelism, takes about 1 μs latency
- Algorithms primarily implemented with **HLS**, data movers written in **HDL**
- Different modules run at 360, 240, & 180 MHz

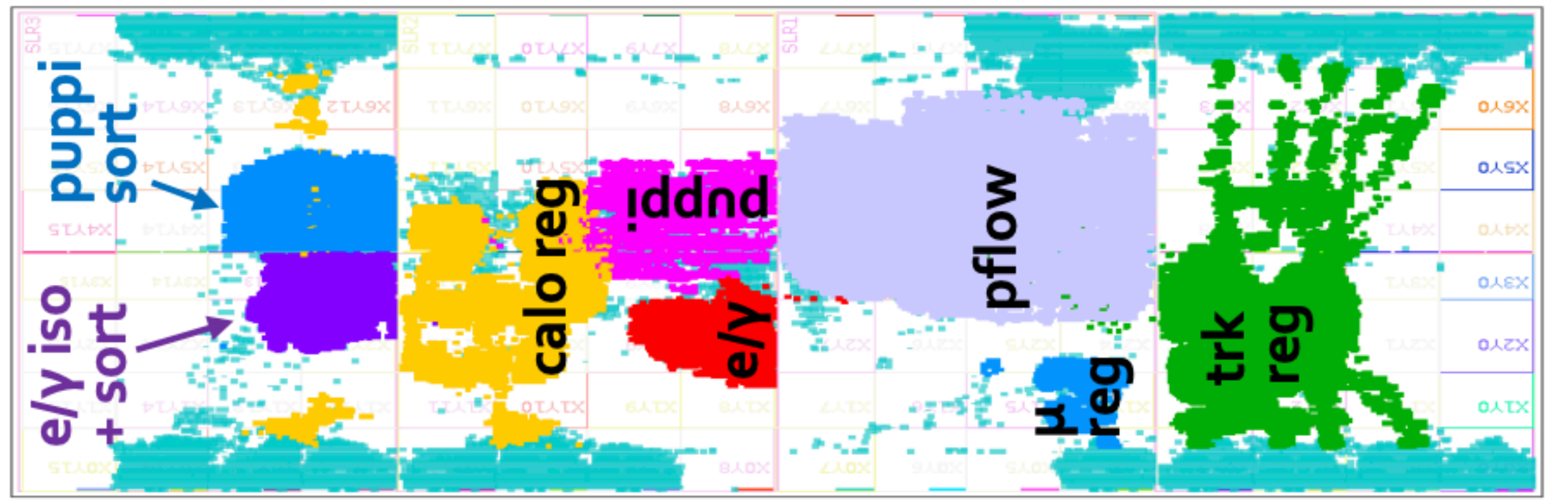


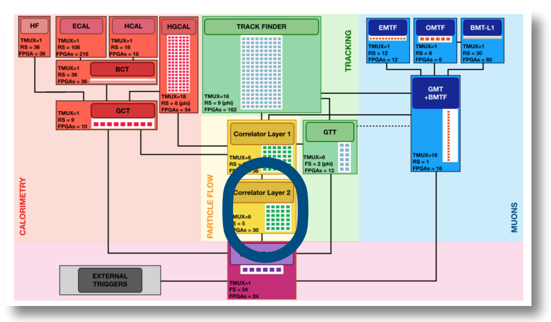


PUPPI candidates

Particle Flow candidates

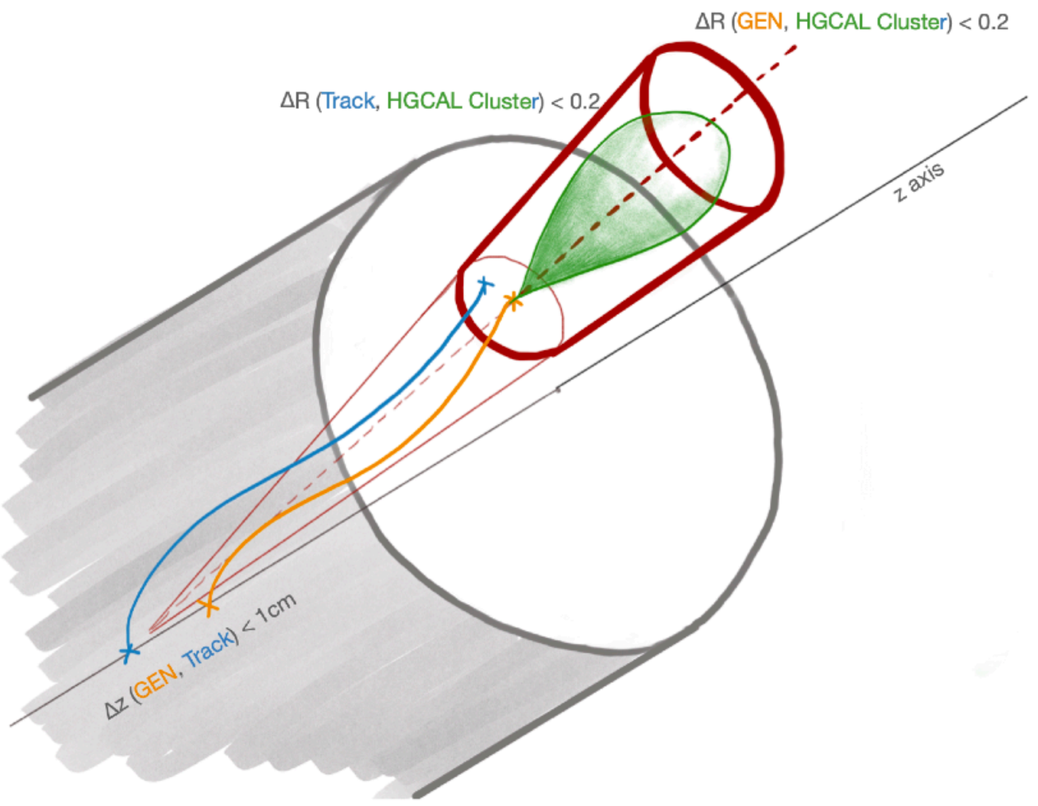
AMD VU13P



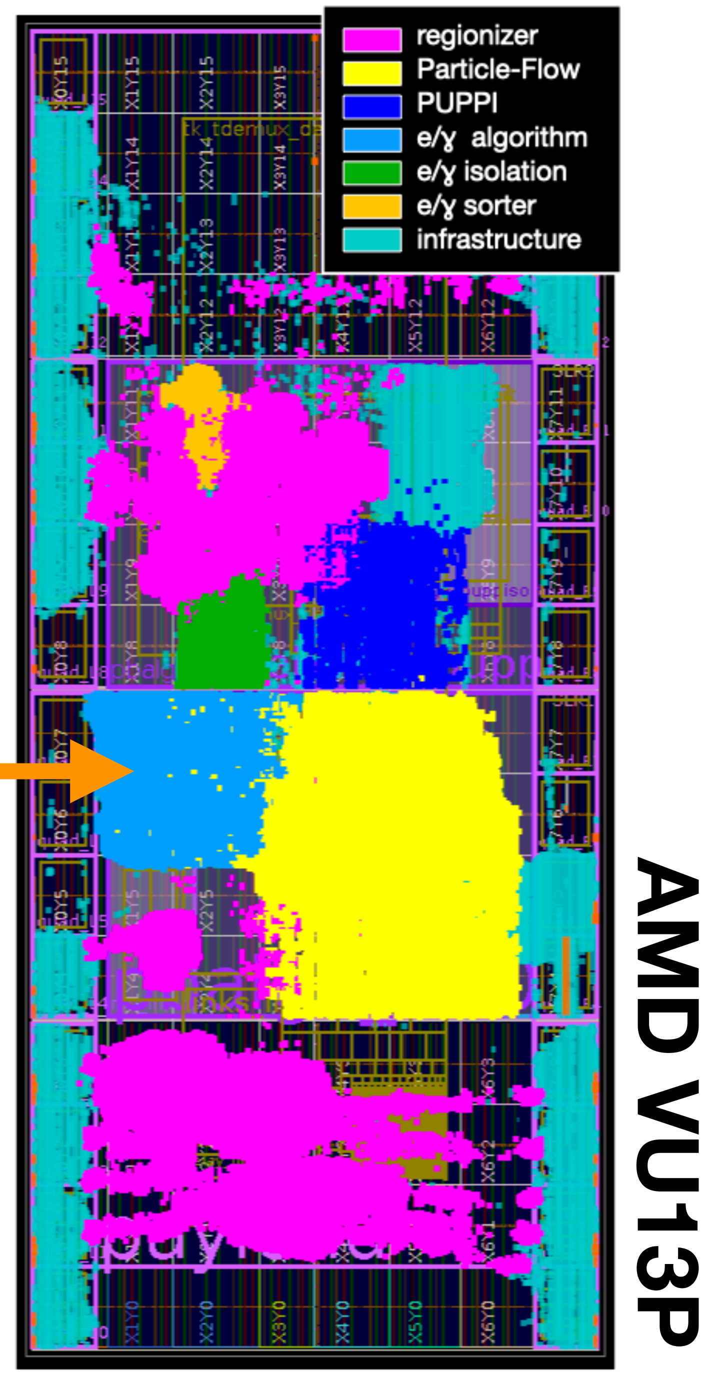


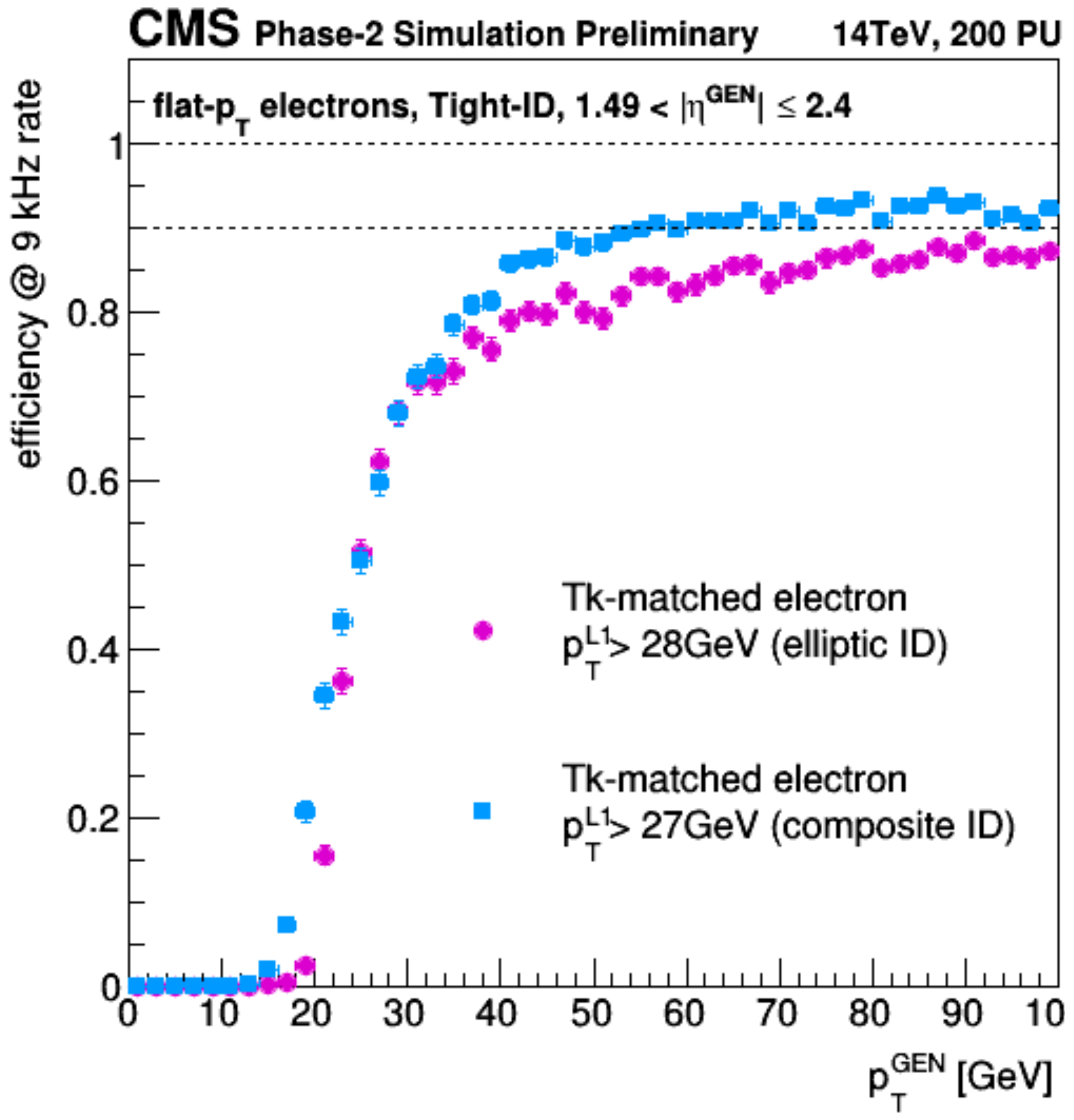
Electron ID

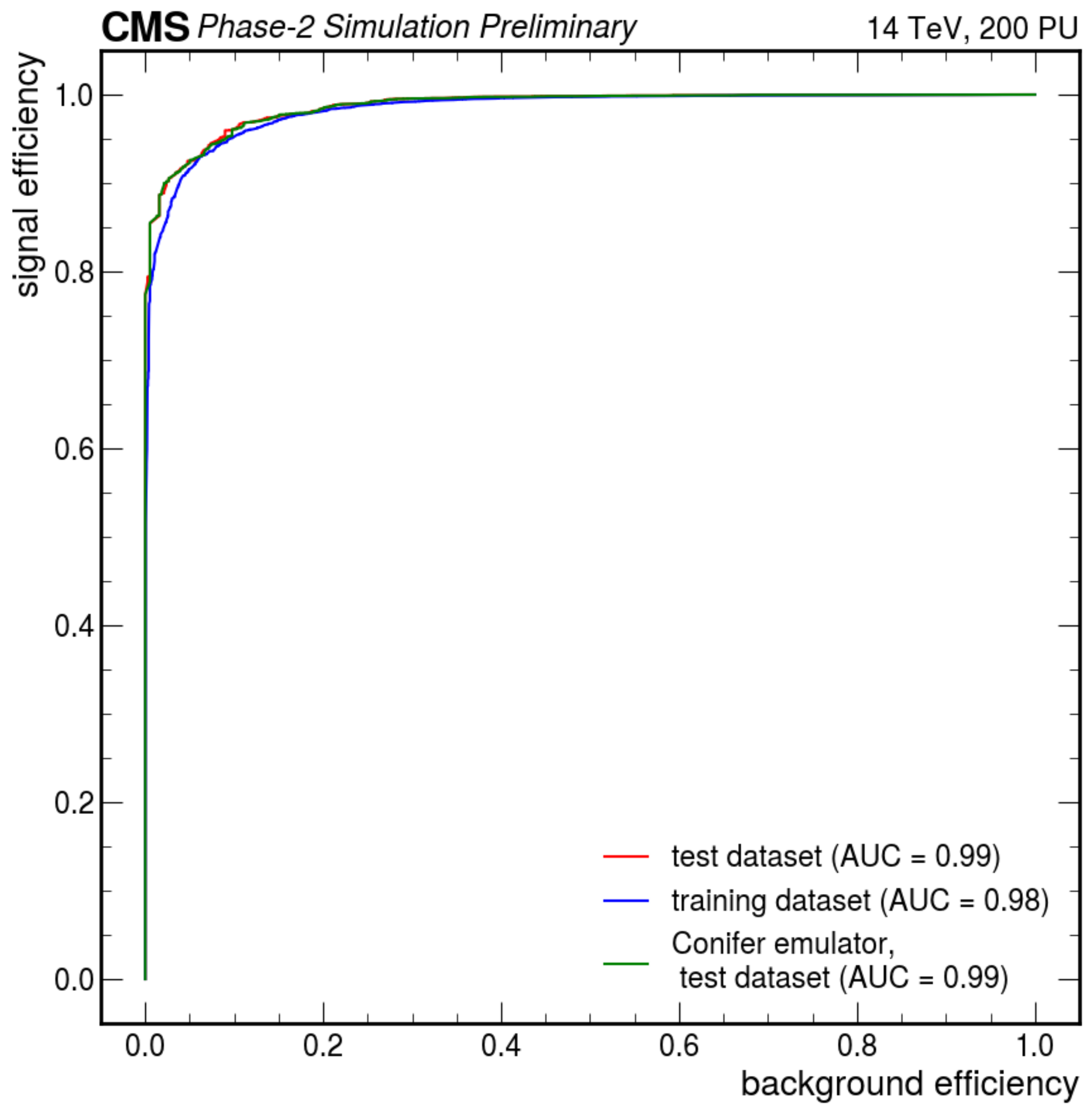
- Electrons will be reconstructed in Correlator Layer 1 by linking a track with a calorimeter cluster
- Neither reconstruction is perfect, and electrons emit bremsstrahlung
- **Baseline kinematic approach** used (η, ϕ) distance and p_T compatibility to make a link
- **New BDT approach** first makes a loose kinematic selection, then uses ML to predict probability that the track & cluster both originated from an electron
  - Using variables from both track and cluster
- **Improved** electron reconstruction **efficiency** with new method (bottom left)
- Keeps electron trigger thresholds as low as Run 3 while maintaining sustainable rate
- Tiny xgboost model, **conifer** for BDT inference in FPGAs
  - 10 instances of BDT to keep up with track/cluster rate

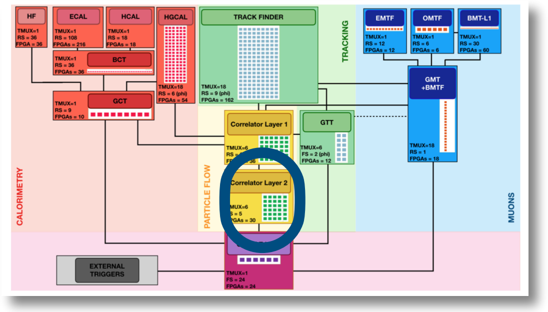


CMS-DP-2023-047



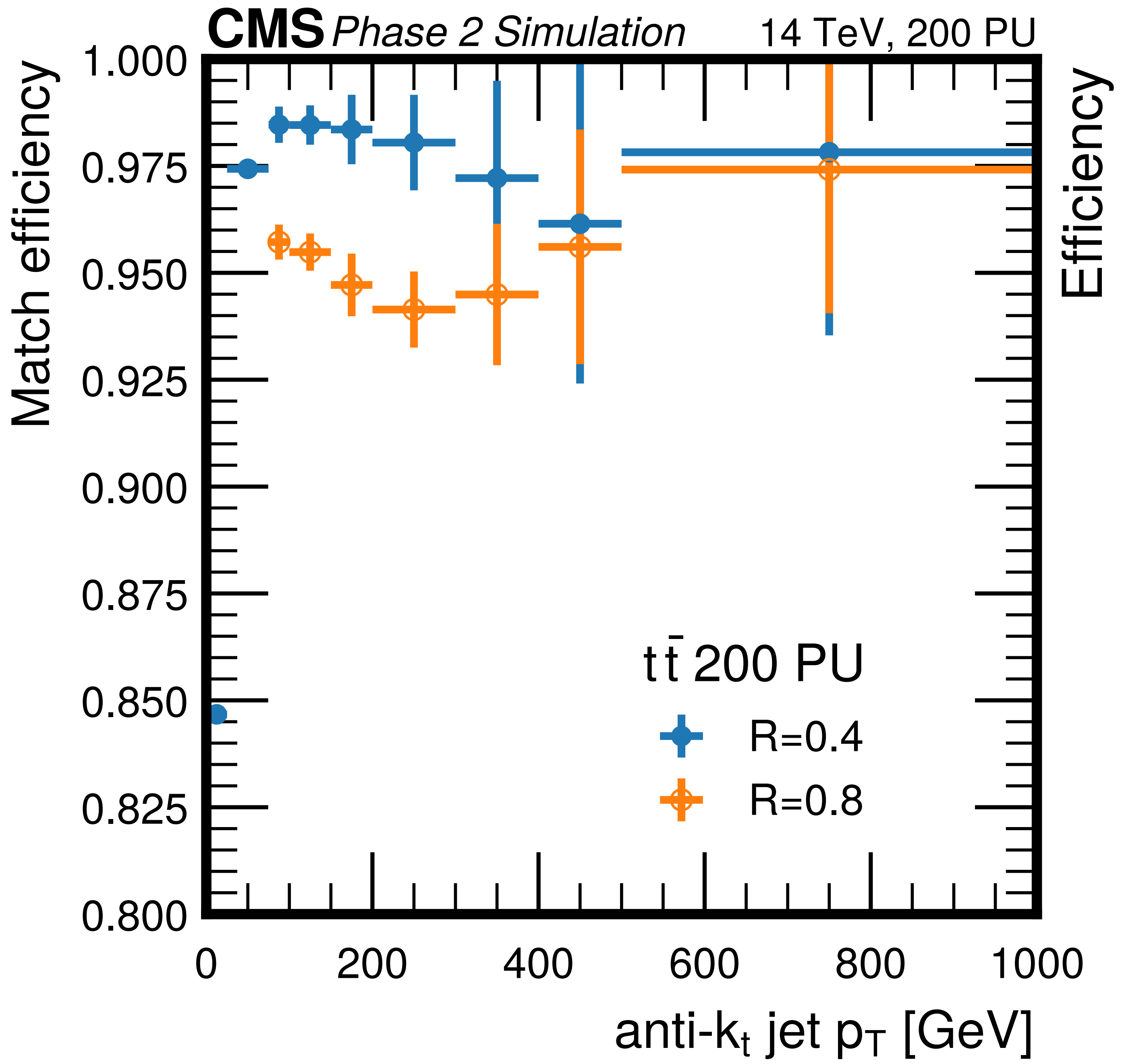


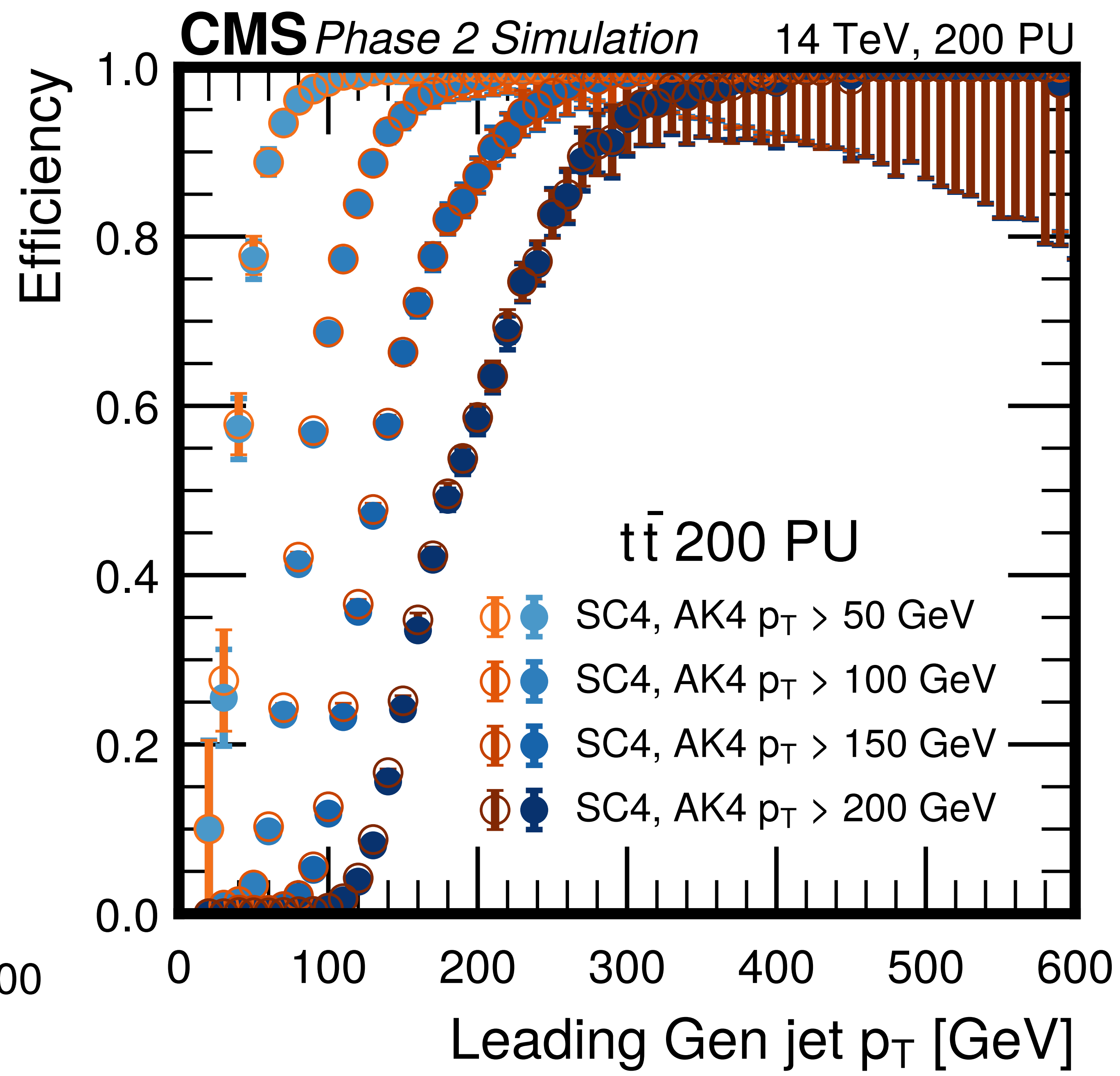


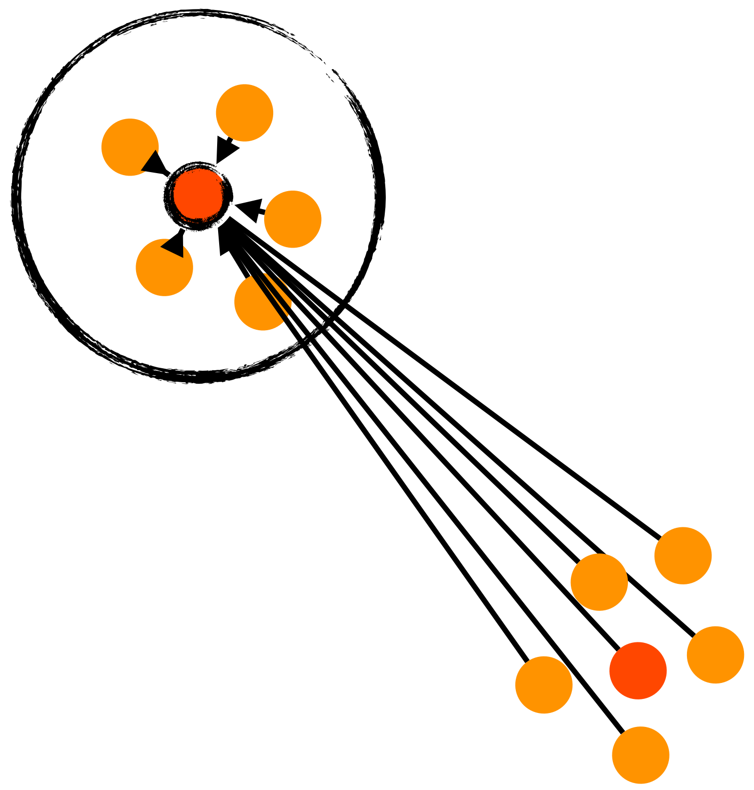


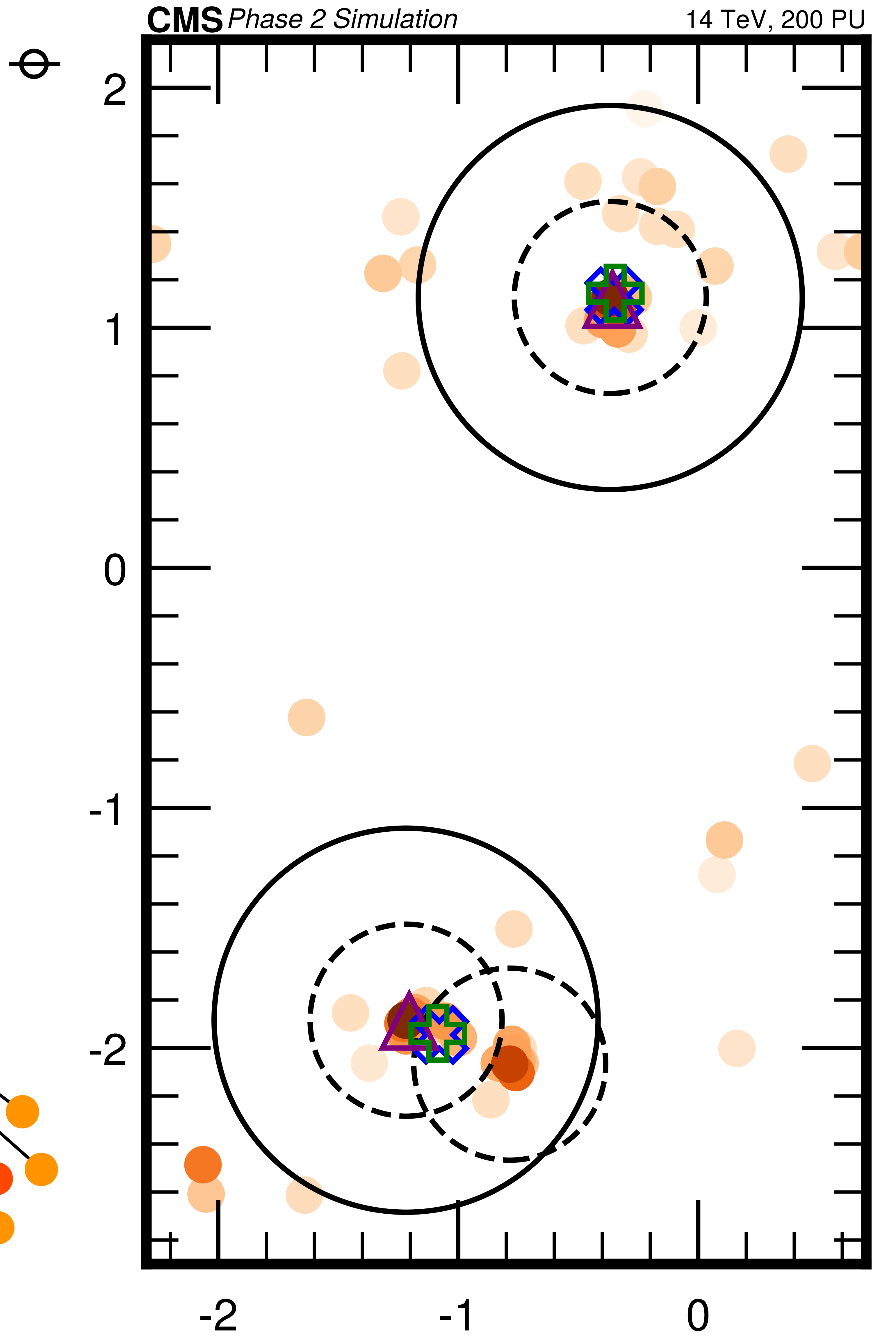
Jet Reconstruction

- Particle Flow and PUPPI at L1T, plus powerful FPGAs and **new techniques** allow us to push what can be done at L1T further
- Now we can cluster particles into jets and **tag the flavour** of those jets
- First we develop a fast and performant jet reconstruction for FPGA
  - Very **simple cone algorithm**: choose a high-p_T seed, cluster in a cone around it
  - Latency 750 ns for 12 jets, **performance close to anti-k_T**
  - Reconstruction of jets with both R=0.4 and R=0.8 enabling **access to substructure**
  - Mix of HLS (arithmetic) and HDL (data-flow, control-flow, data movement) modules

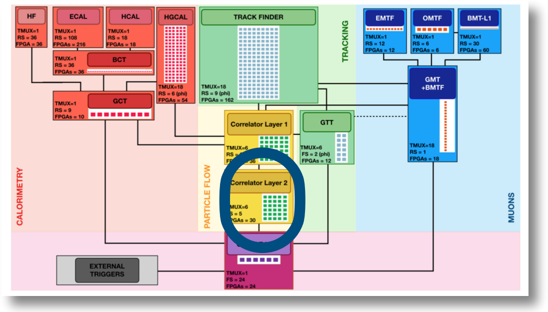






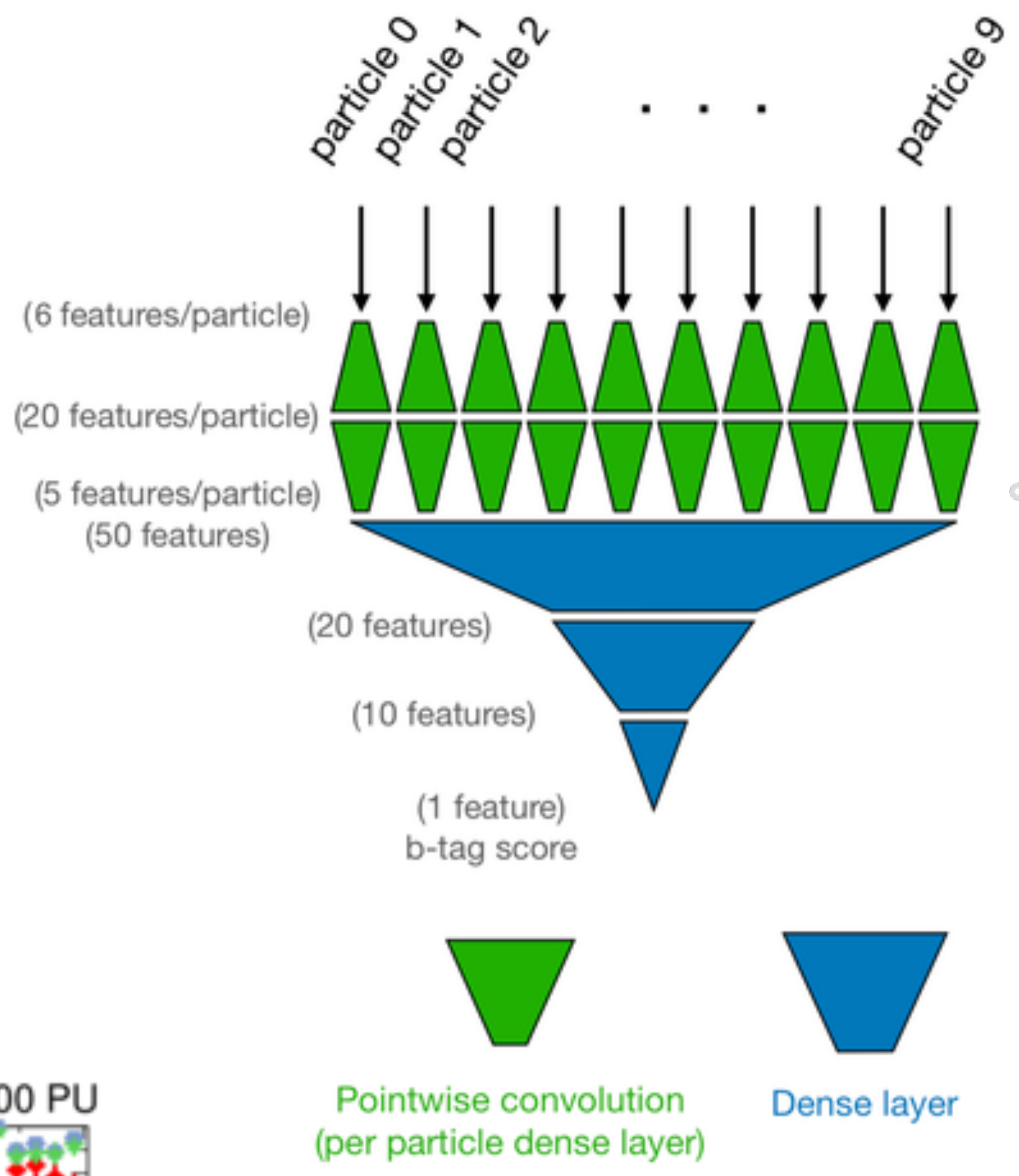


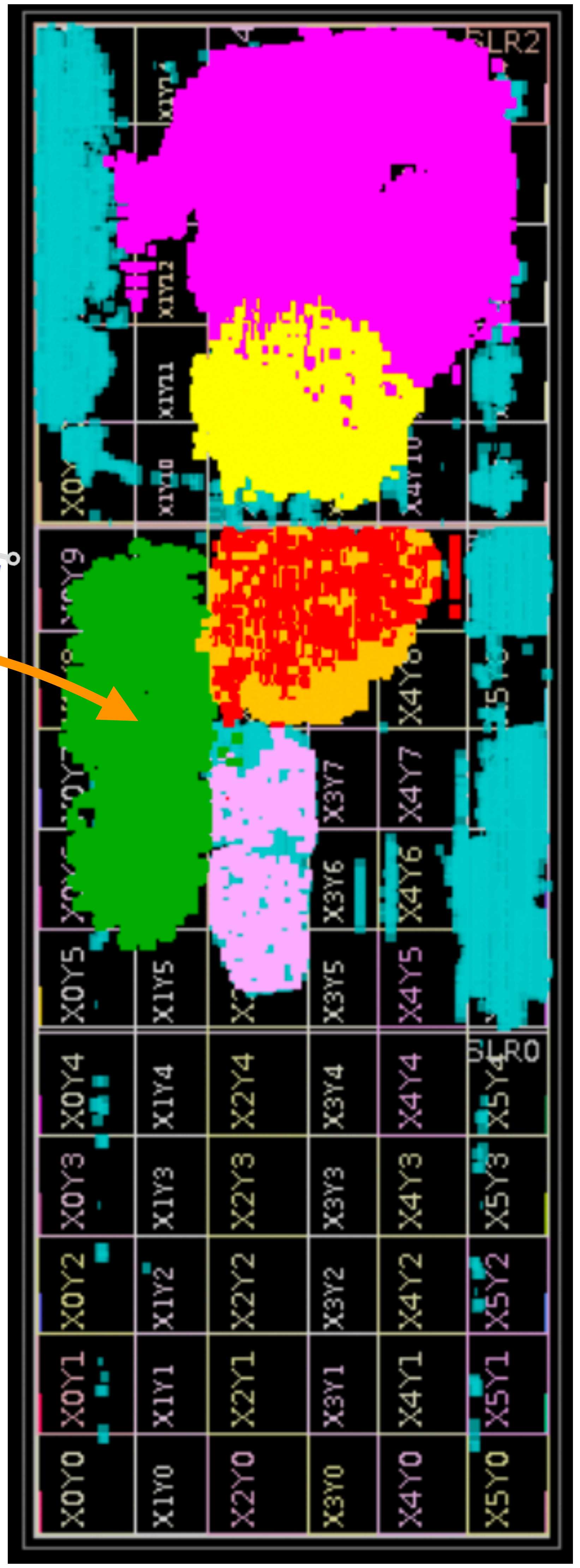
arXiv:2310.08062



Jet Tagging

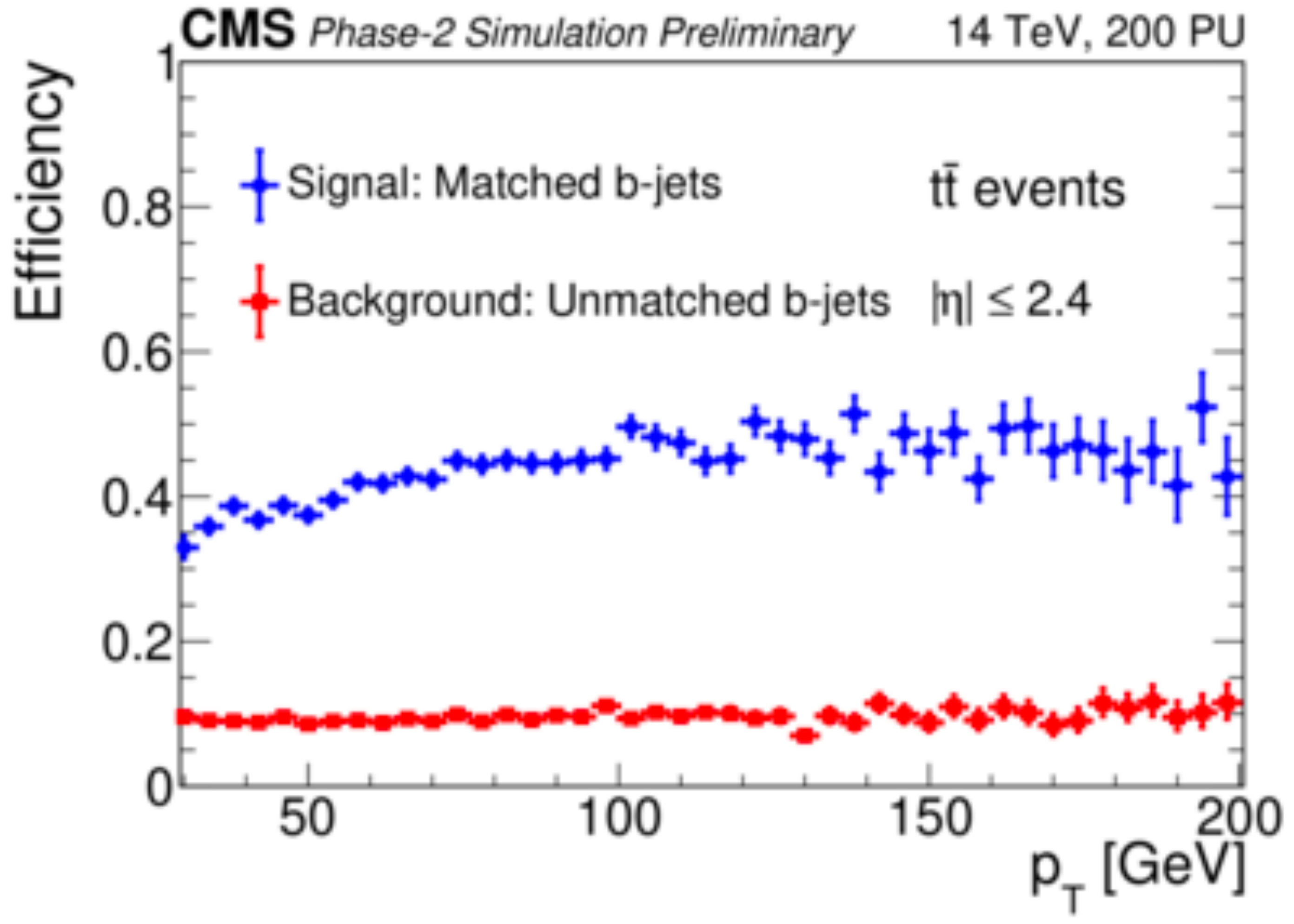
- After jet reconstruction we retain the list of **jet constituents**
- We can use **Machine Learning** for jet tagging
- Using Deep Neural Network for **b-tagging**
  - Relies on track displacement measurement from L1 track finder
- **Tiny model** improves trigger reach to important final states (HH→bbbb shown)
- Fits in FPGA (right) and total latency (jet reco + tagging) less than 1 μs, **hls4ml** for NN inference in FPGAs

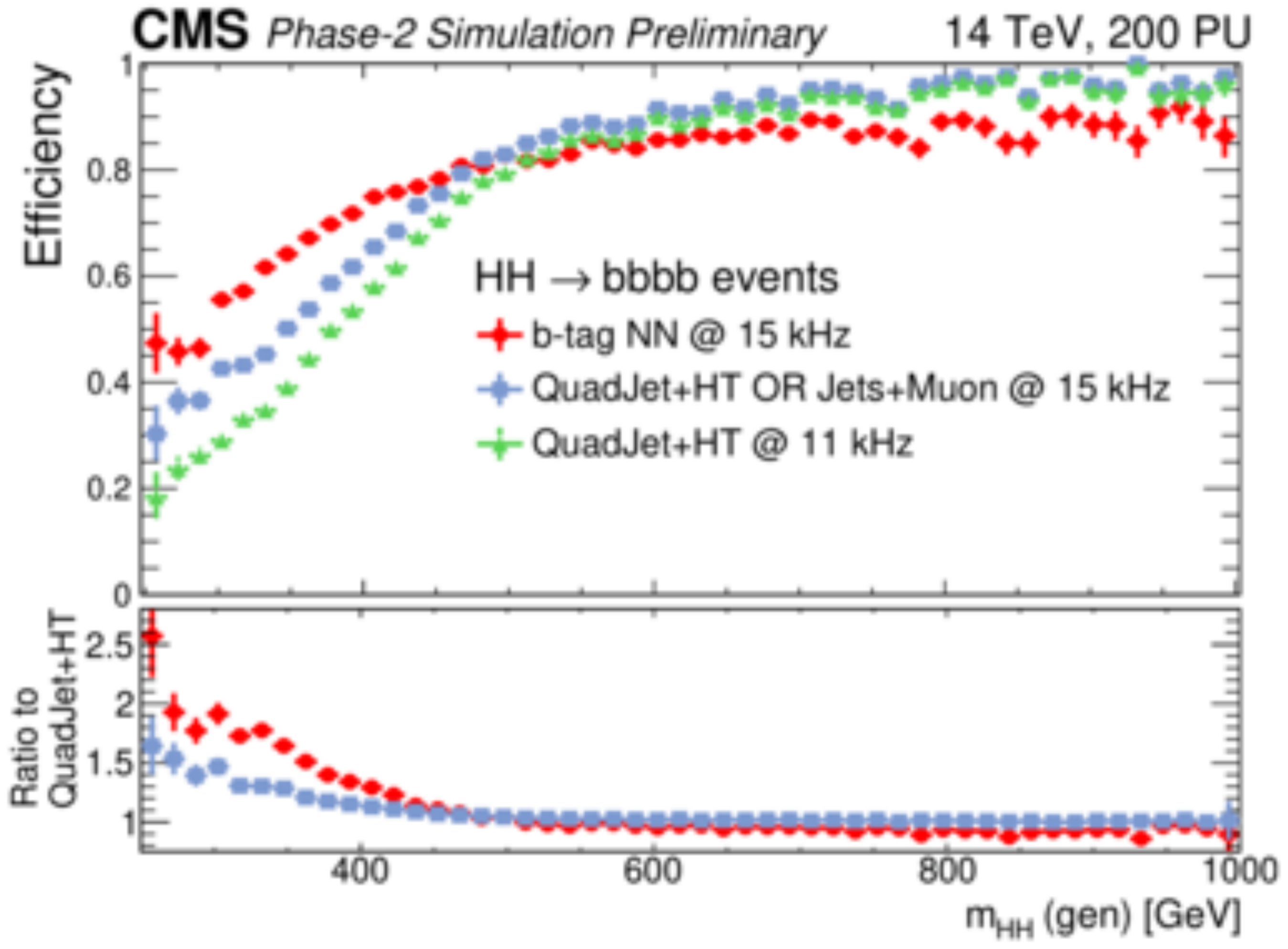




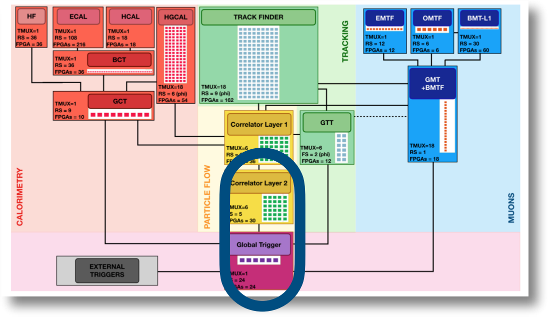
Particle Receiving
Jet Constituent Finding
Jet Axis Computation
Sorting, Buffering
B tagging Neural Network

AMD VU9P Serenity



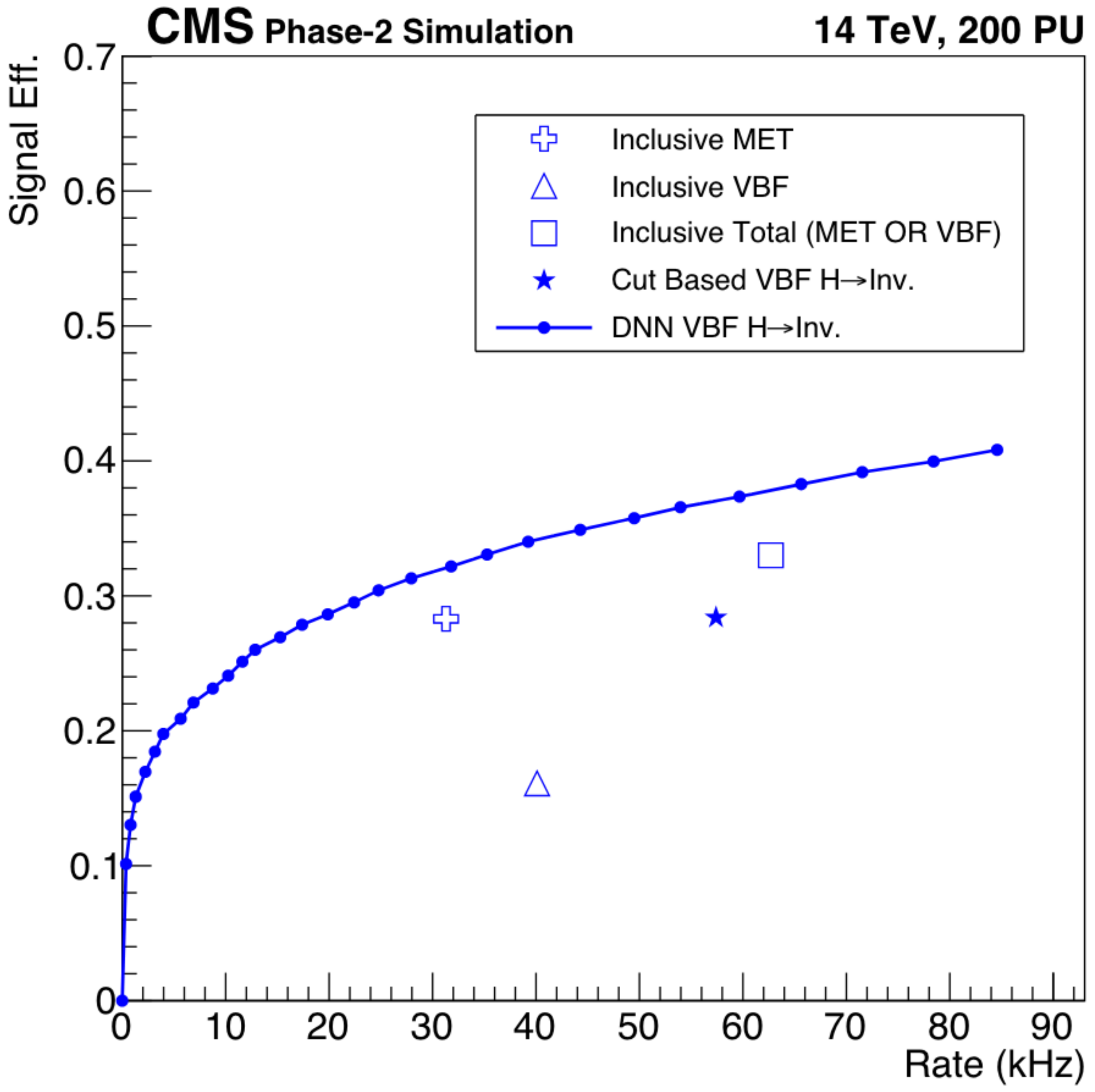


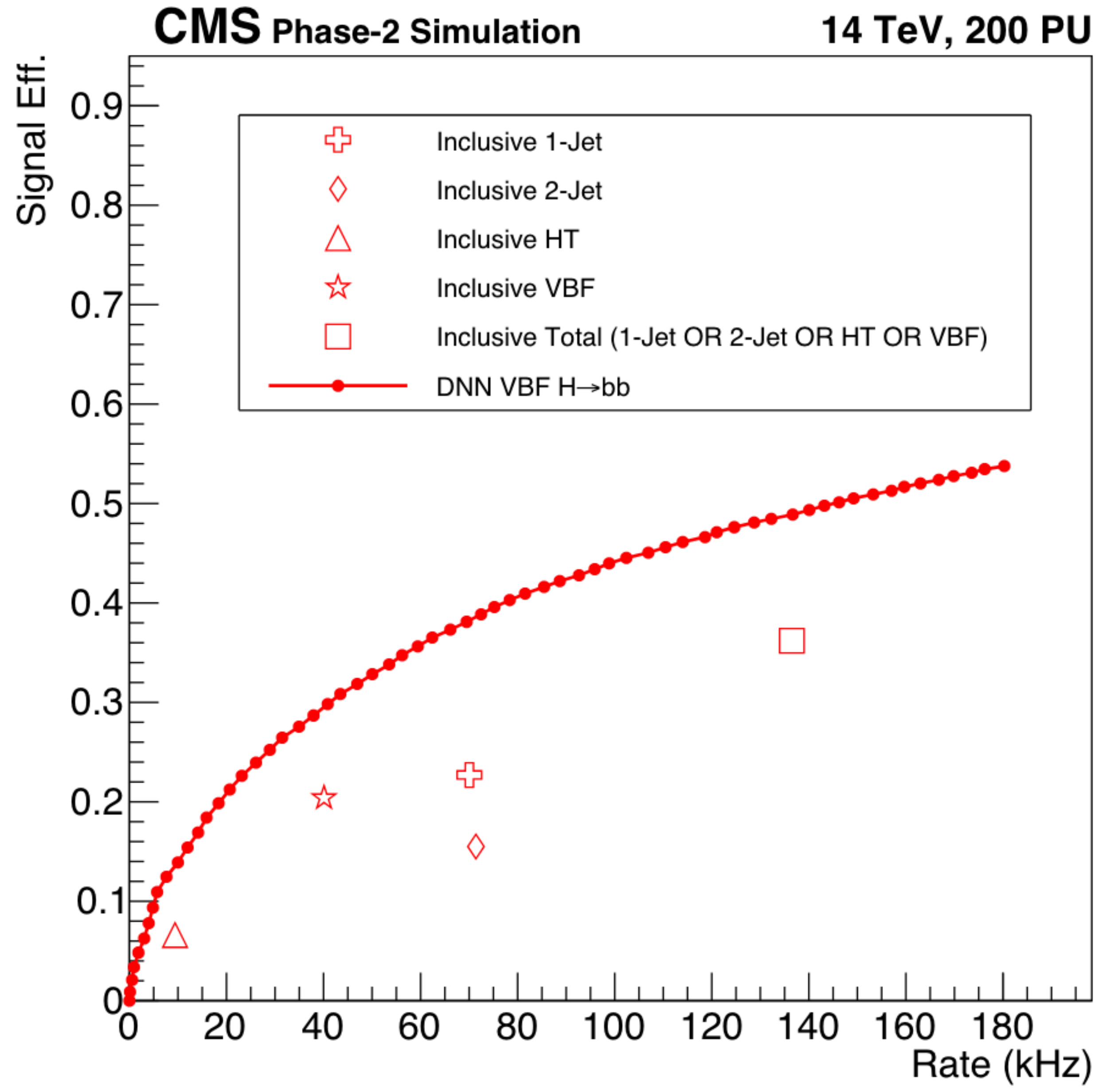
CMS-DP-2022-021

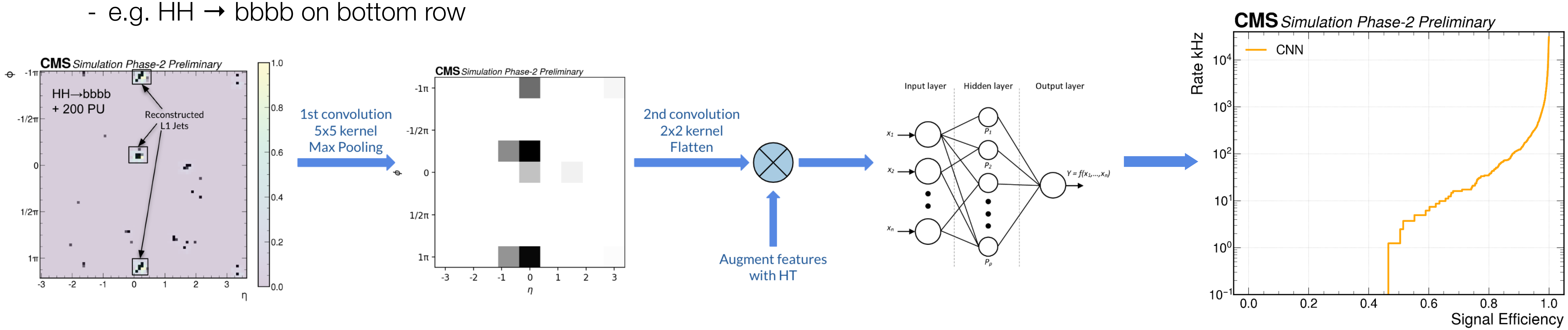


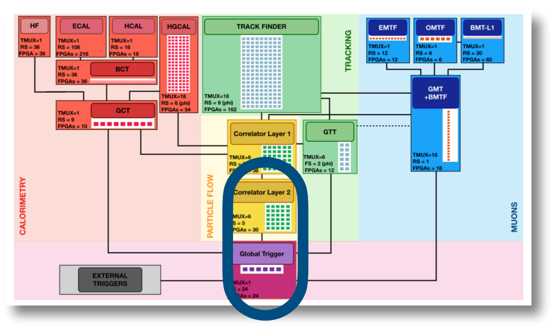
Final State Selection with ML

- **Global Trigger** makes the final event selection decision comparing event objects against "menu" of conditions
  - jets, sums (missing transverse energy), electrons, photons, muons, taus
- We will **exploit ML extensively** to improve reconstruction, and also to make trigger decisions
- In Global Trigger we can train ML models on event objects
  - e.g. VBF Higgs in top row plots
- In Correlator we can use all of the PUPPI candidates to **classify topologies**
  - e.g. HH → bbbb on bottom row



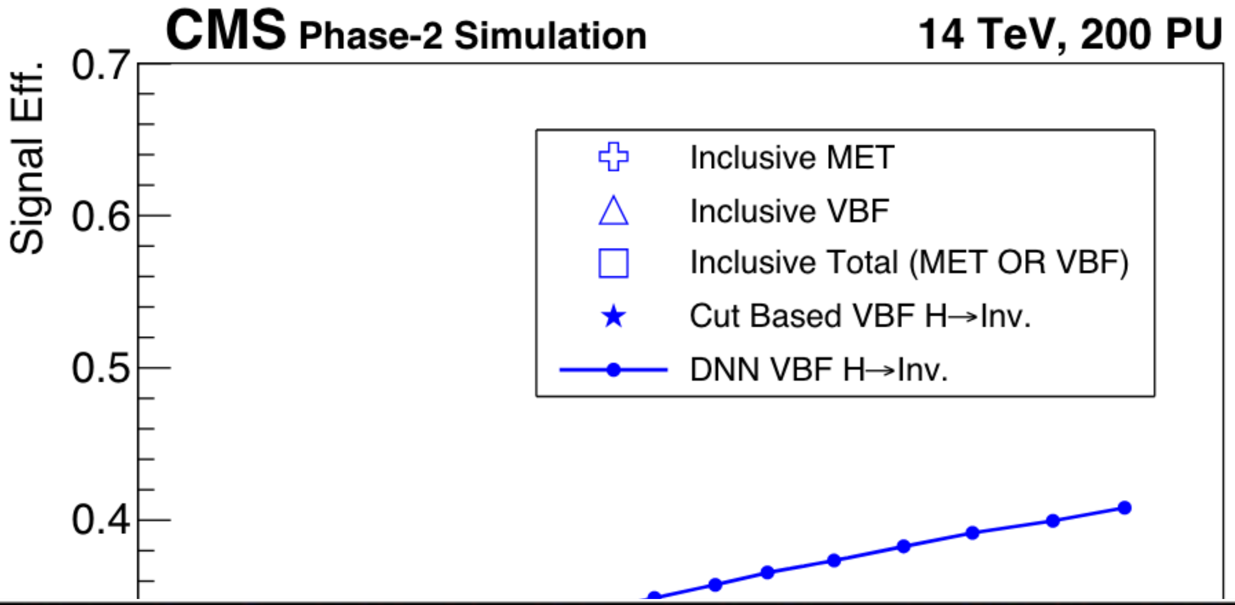


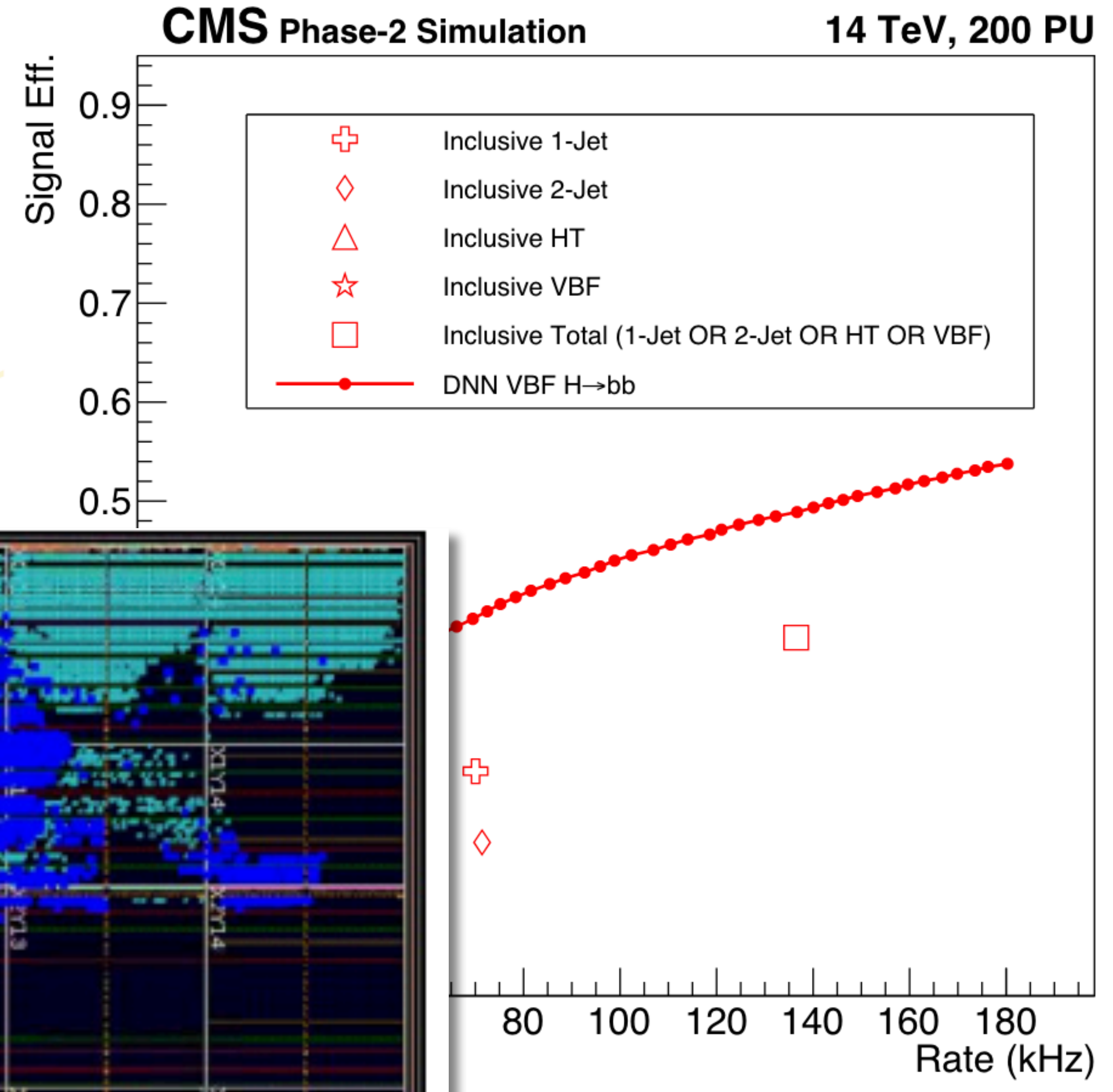


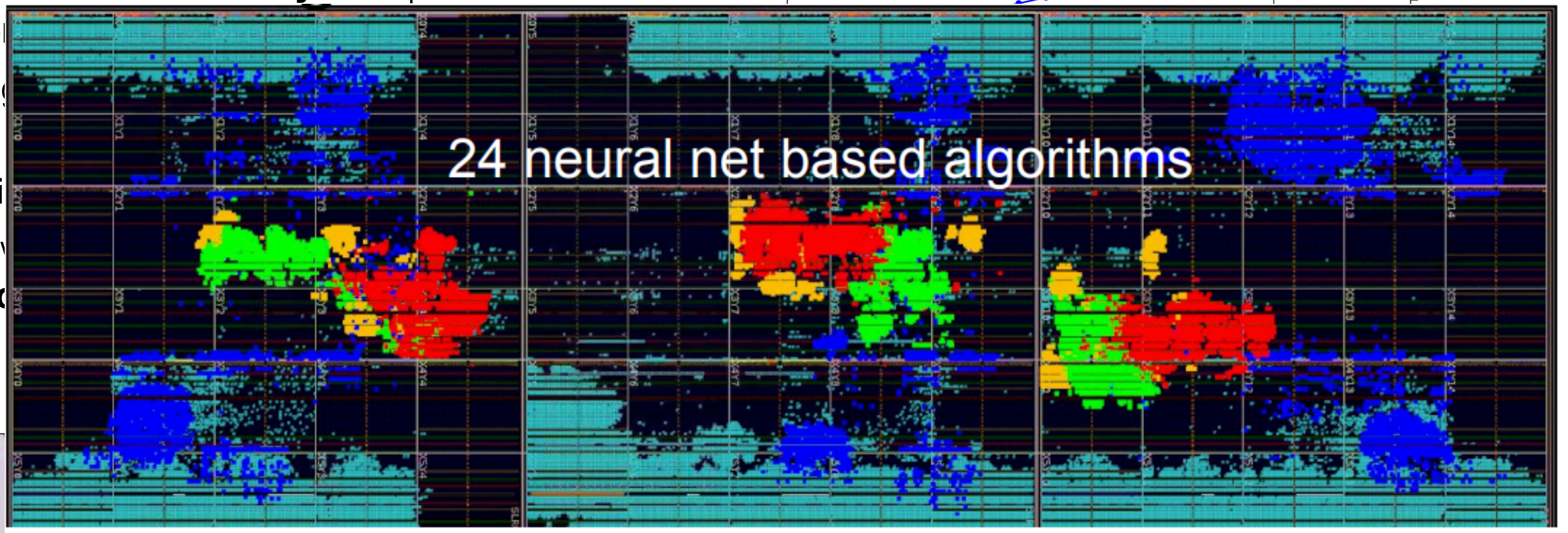


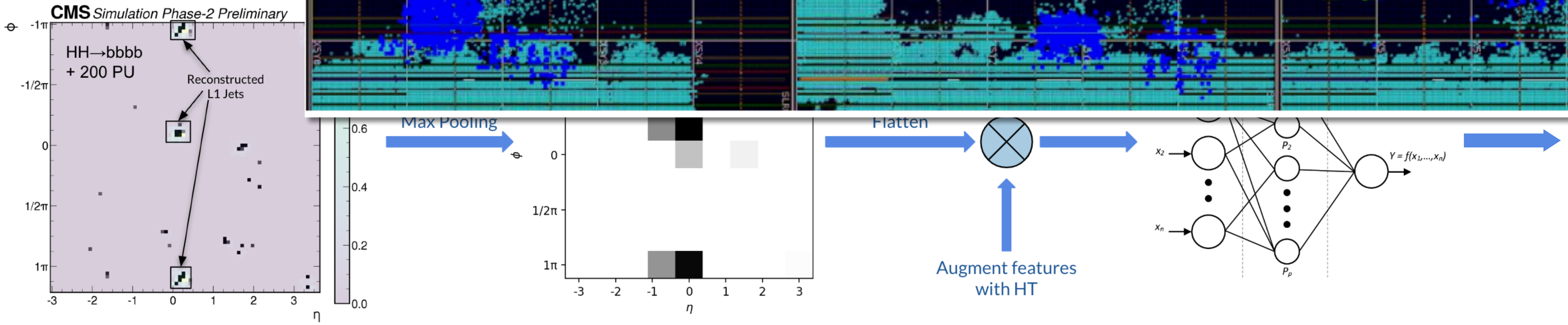
Final State Selection with ML

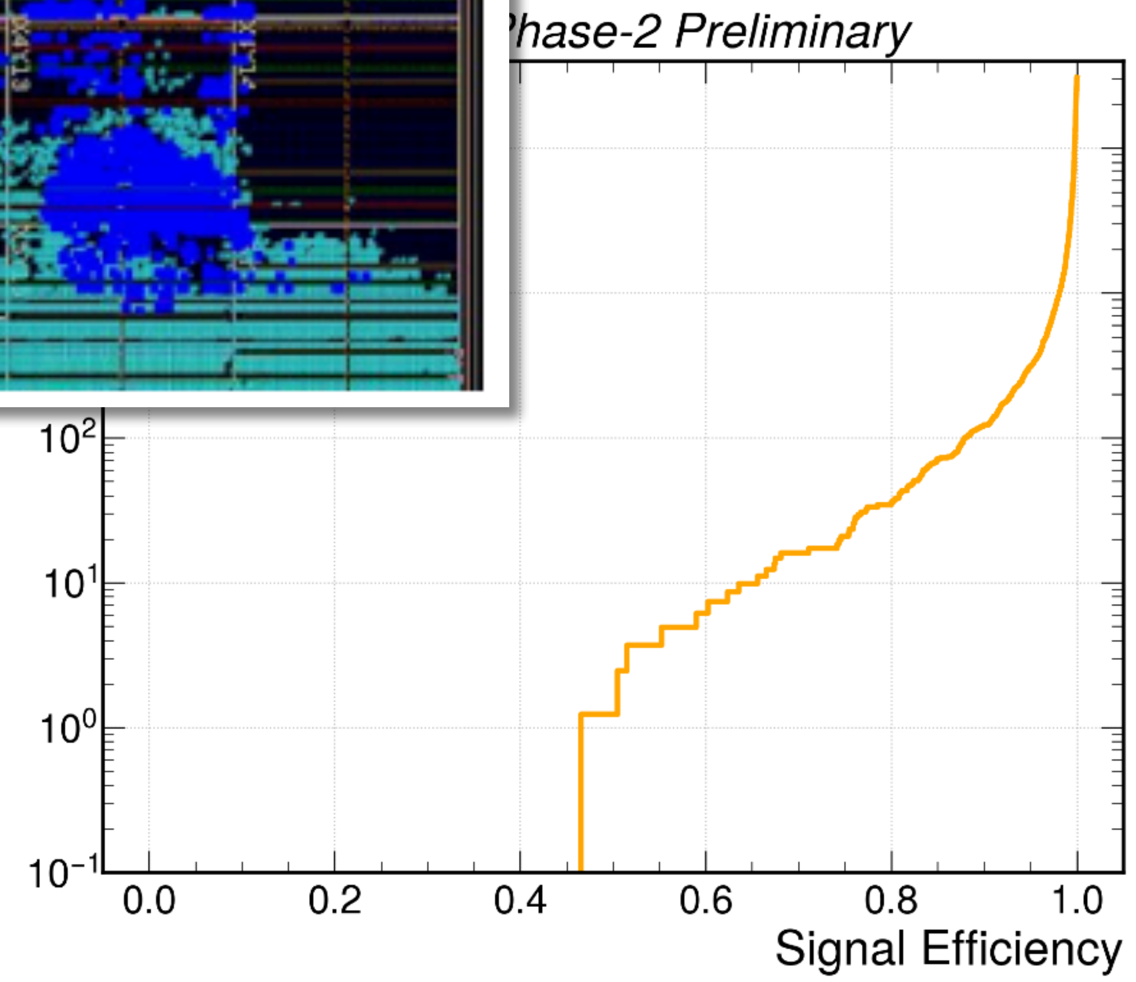
- **Global Trigger** makes the final event selection decision comparing event objects against “menu” of conditions
  - jets, sums (missing transverse energy), electrons, photons, muons, taus
- We will **exploit ML extensively** to improve reconstructio
- In Global Trig objects
  - e.g. VBF Hi
- In Correlator **classify topo**
  - e.g. HH →

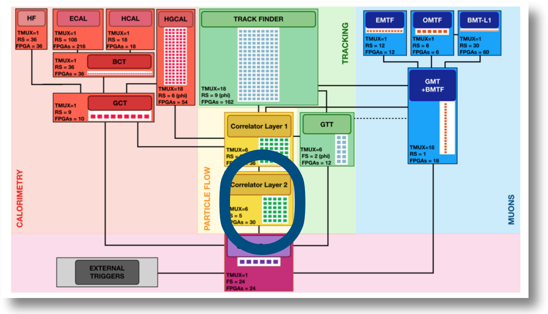






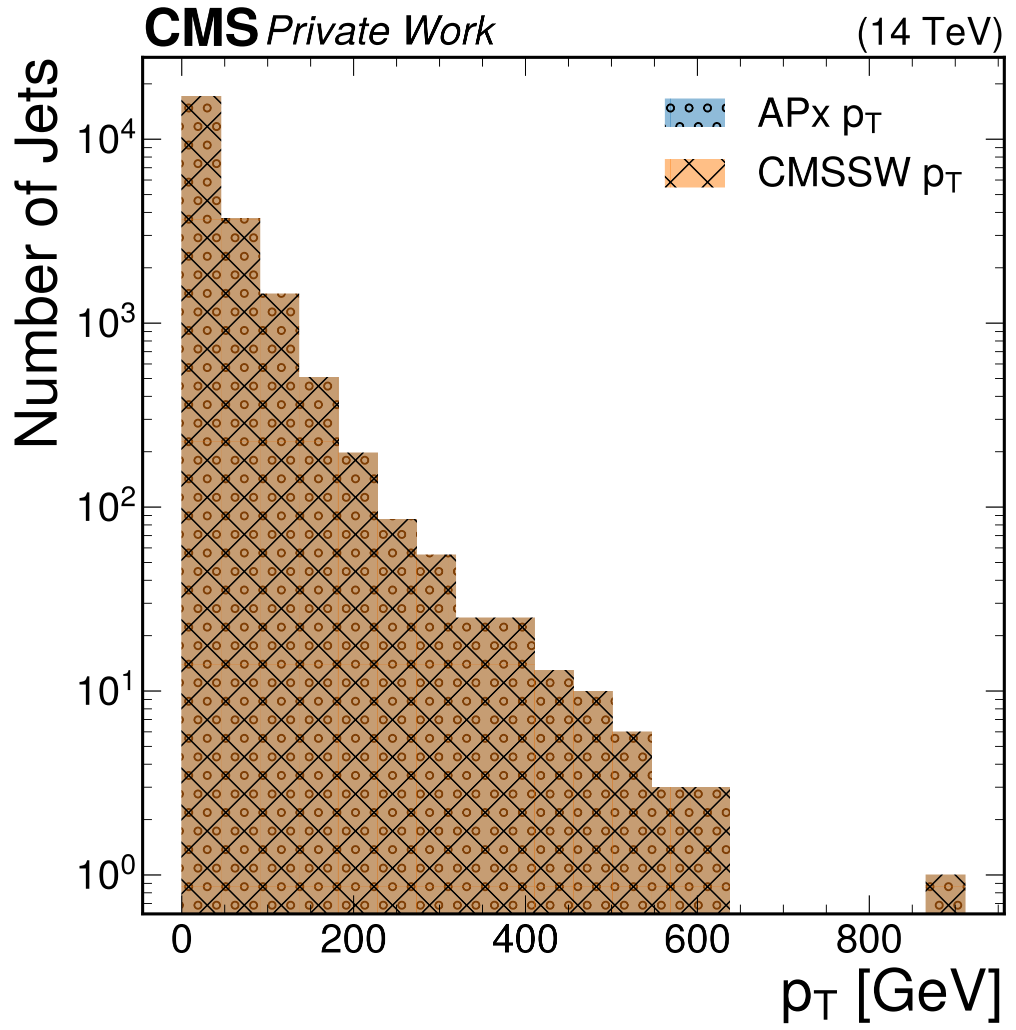


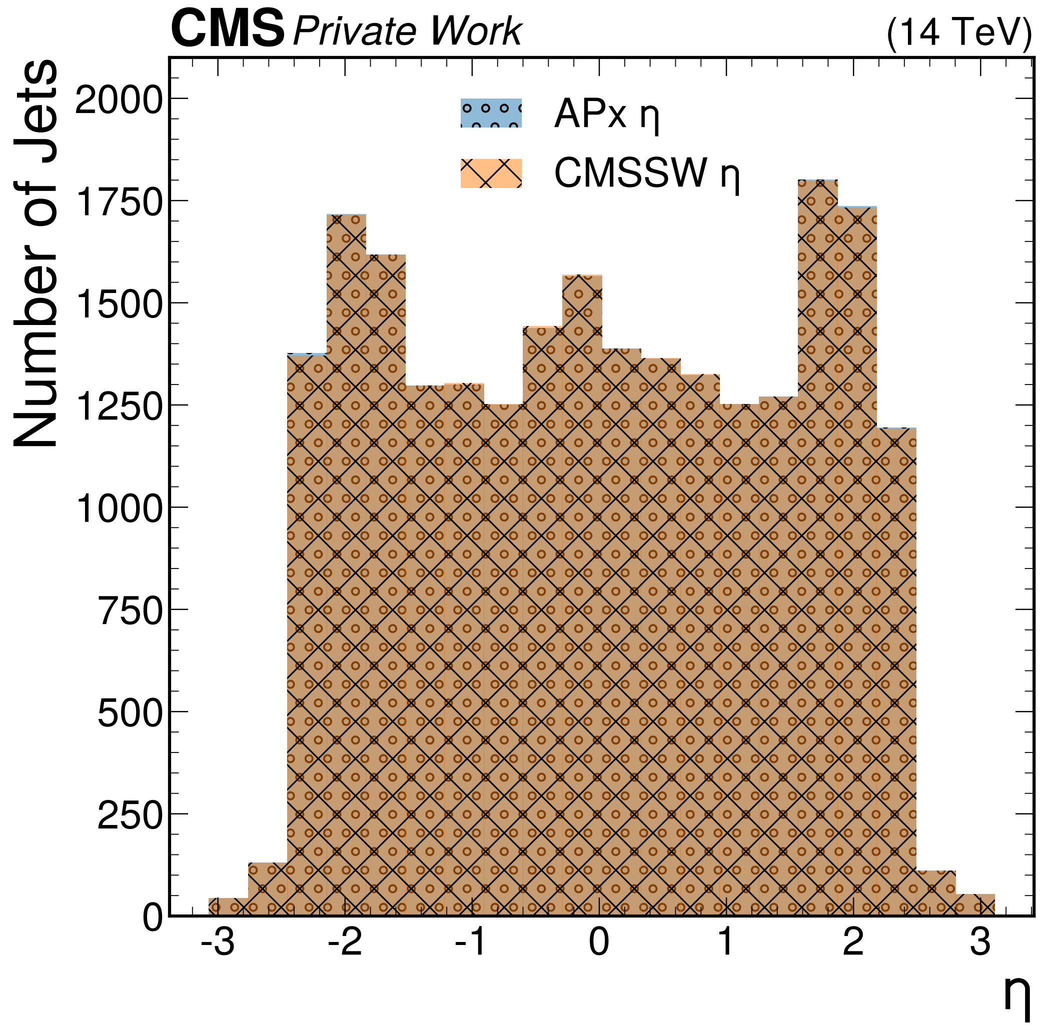


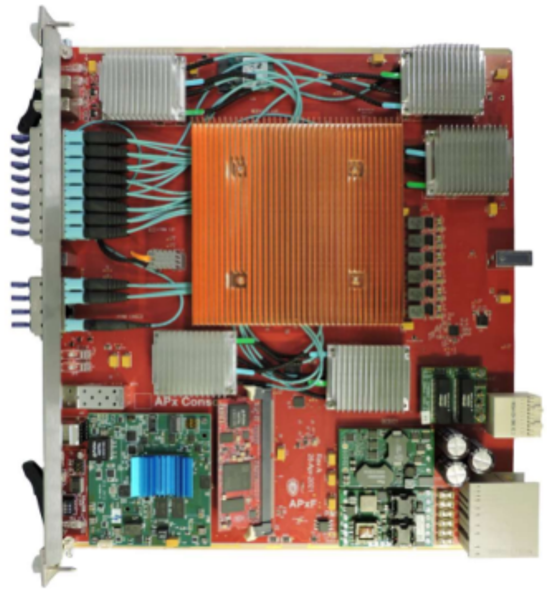


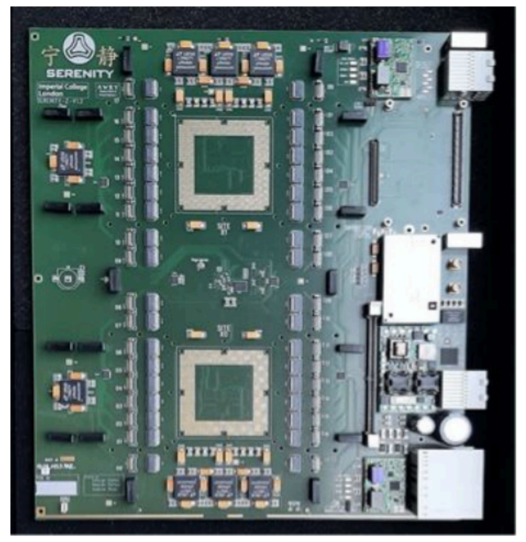
Jets on Serenity and APx

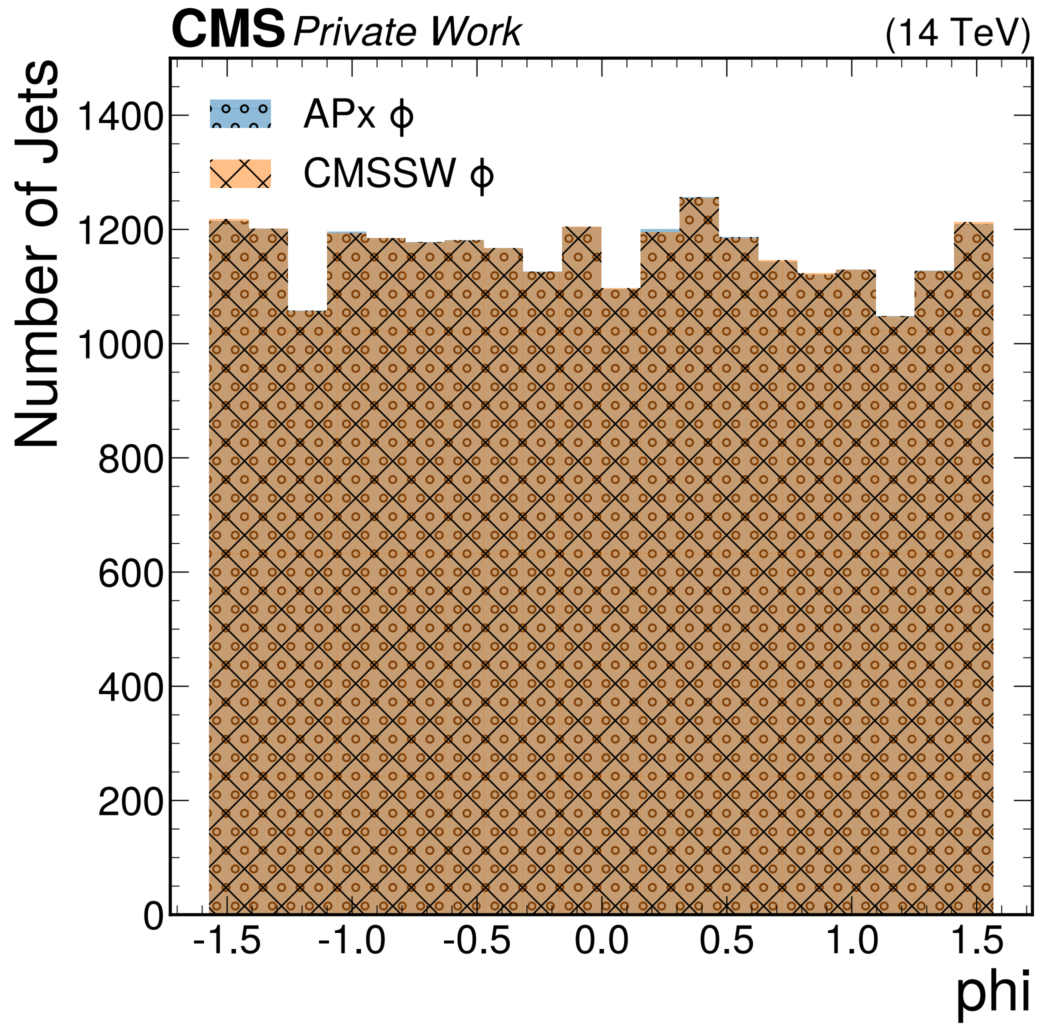
- Correlator Layer 2 will use both Serenity and APx, split by algorithm
- Jet algorithm has been demonstrated on both platforms
- Common 'adapter' interface HDL written from board framework to algorithm internals
- Internal 'token' handshaking between HLS modules allowed easy tweaking of pipeline latency to ease timing on both `-1` and `-2` speed grade parts
- Testing by writing data from simulated Phase 2 CMS events into board link input buffers → run through algorithm → capture output link buffers
- Observing 99.9% bit-exact matching between CMS Software emulation and Serenity/APx on 1000 events tt + 200 PU
  - Remaining mismatches understood to be from p_T ordering
- Successfully demonstrates the interoperability of Serenity and APx, and the flexibility of Correlator Layer 2 designs



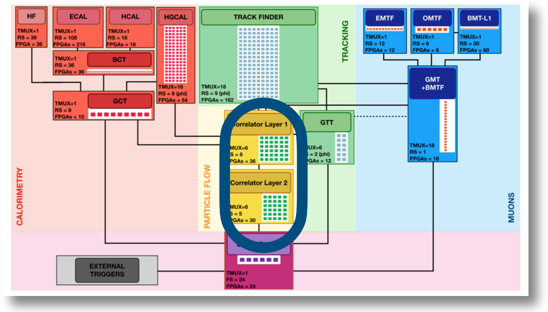






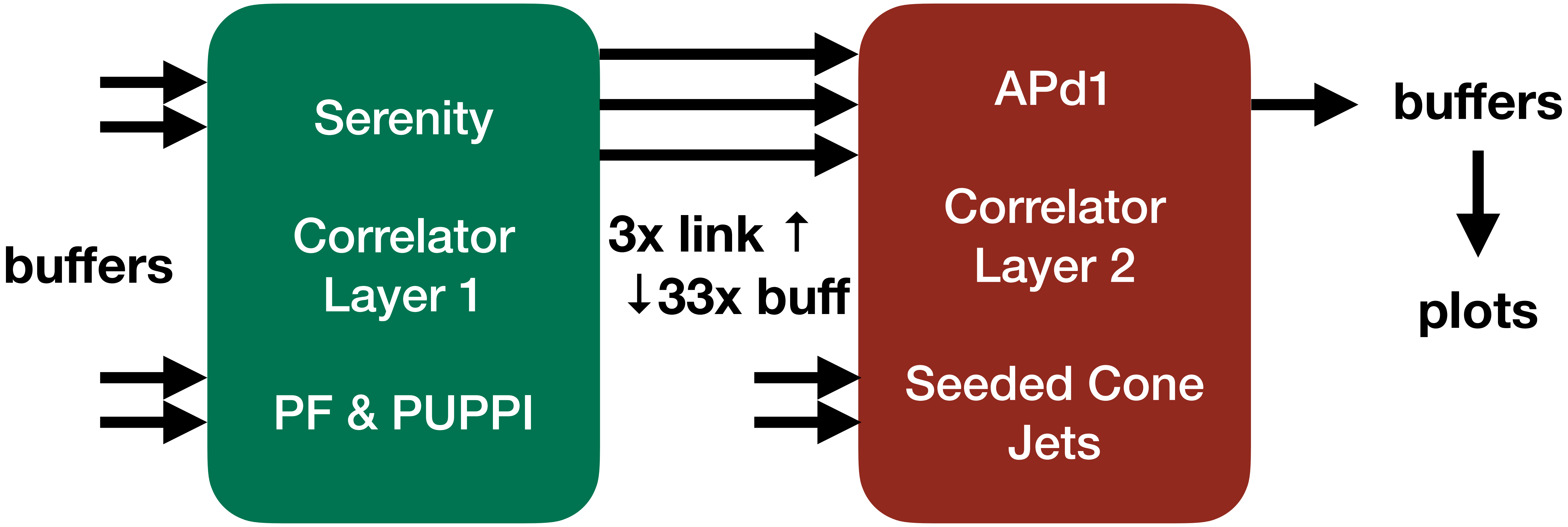


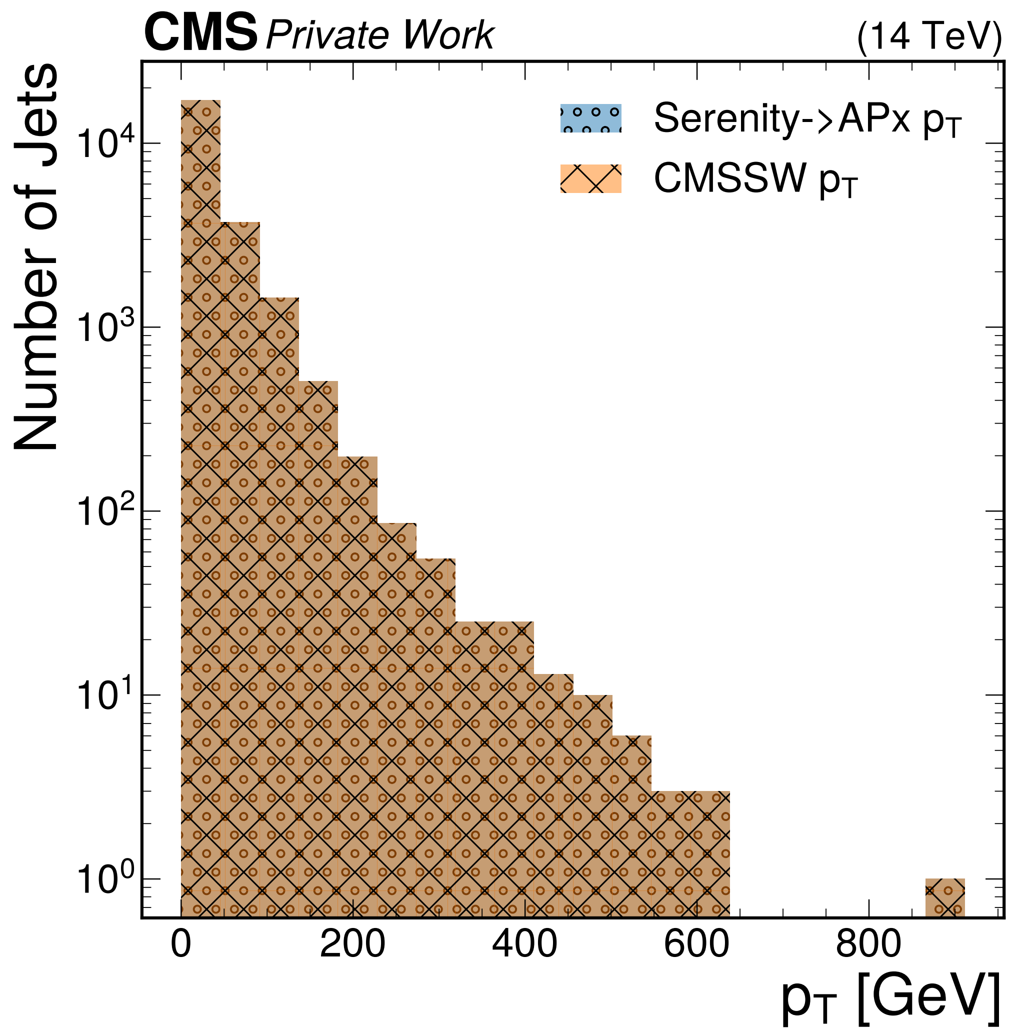
- See talk by Isobel Ojalvo later today on APx board

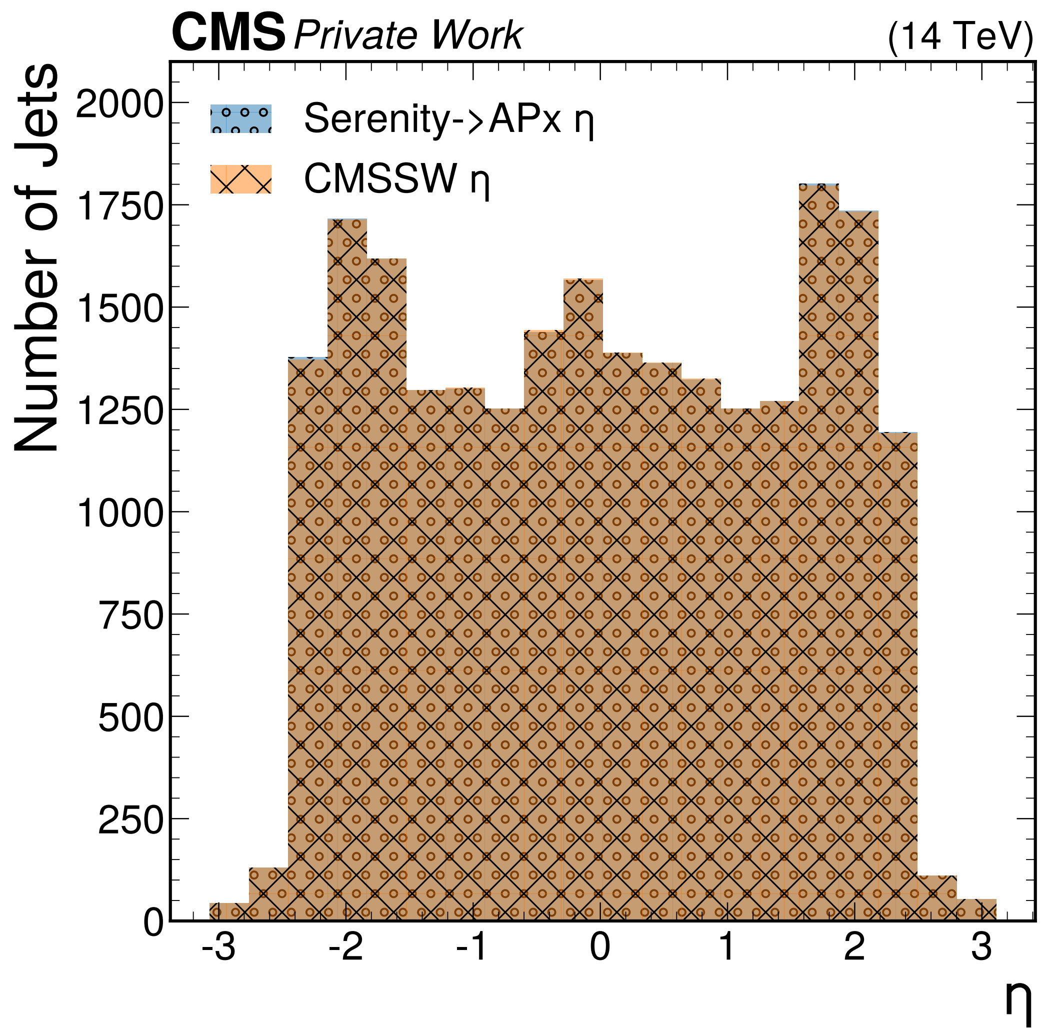


Board-to-board testing

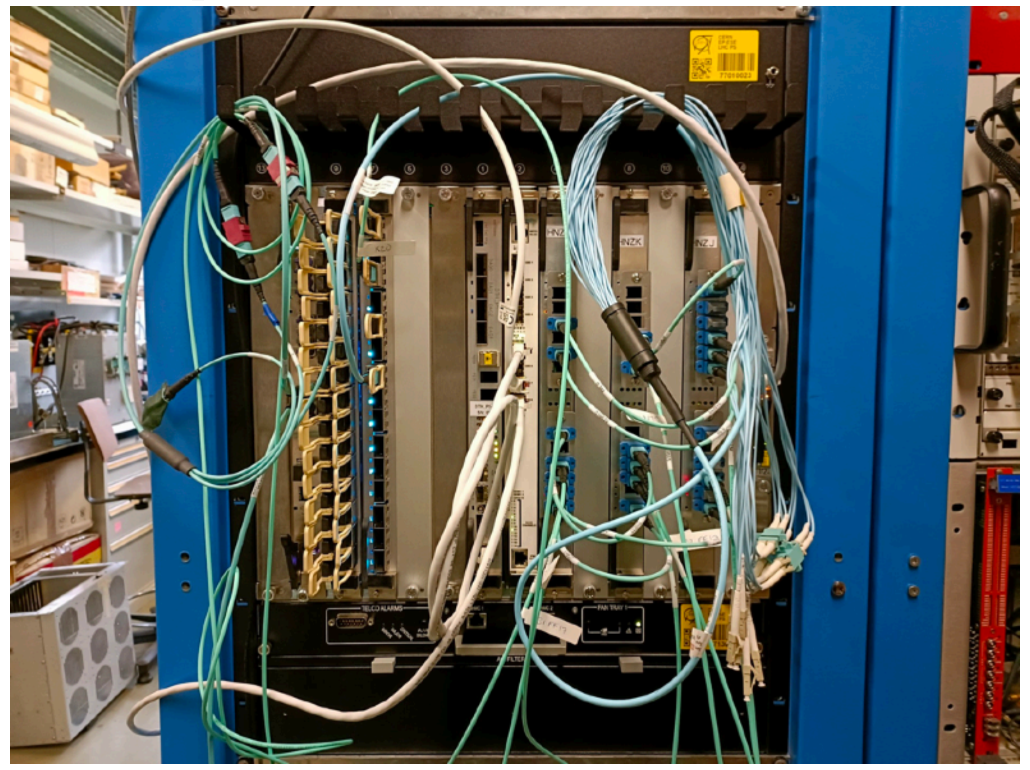
- Board-to-board communication between flavours initially established with link and CMS Protocol (CSP) testing
- Board-to-board testing between algorithm on same flavour already established
- First cross-flavour board-to-board tests with algorithms have recently taken place **Serenity**, **APd1**
  - **Correlator Layer 1 Endcap** to **Correlator Layer 2 Jets**
  - **Global Track Trigger Vertexing** to **Correlator Layer 1 Endcap**
  - **HGCal** to **Global Calorimeter Trigger**

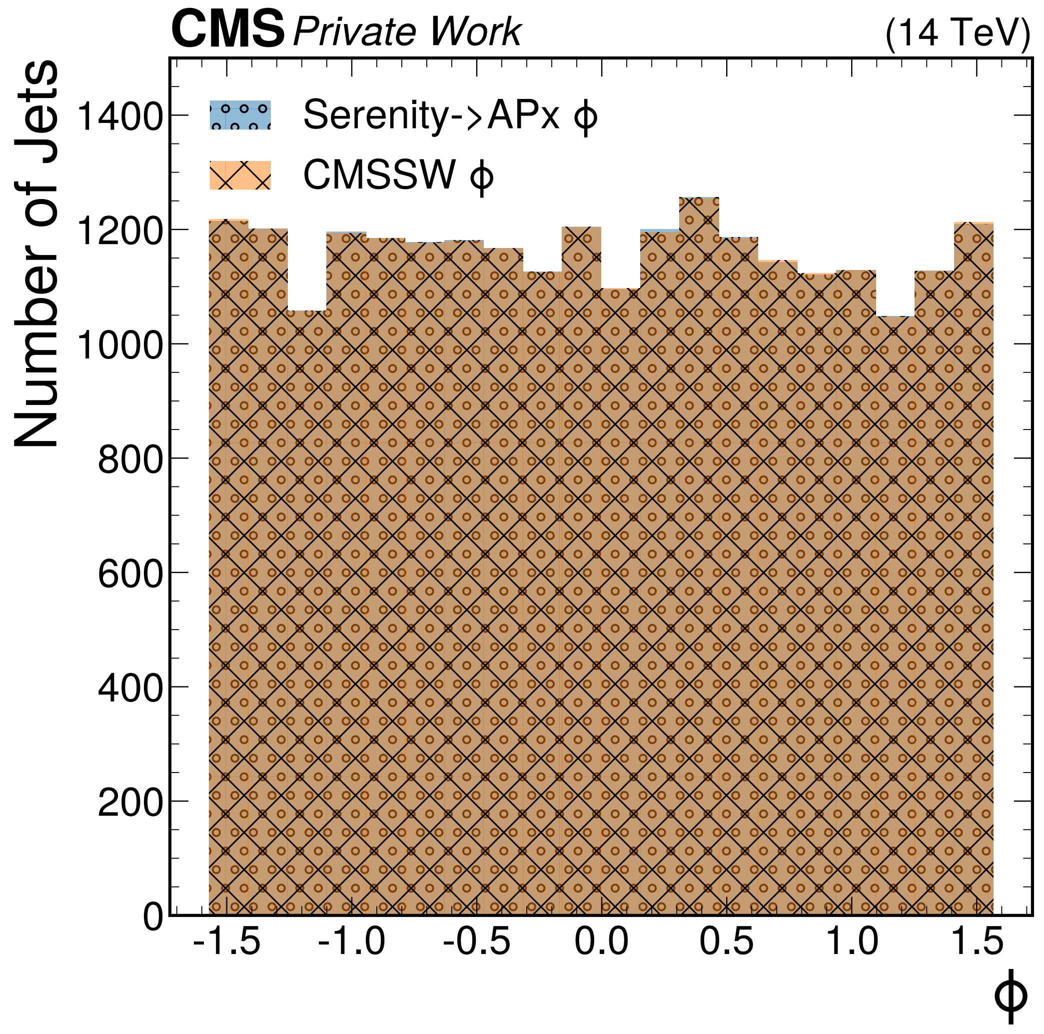






Setup at CERN Prevessin





Same test events & observed matching as prev. slide

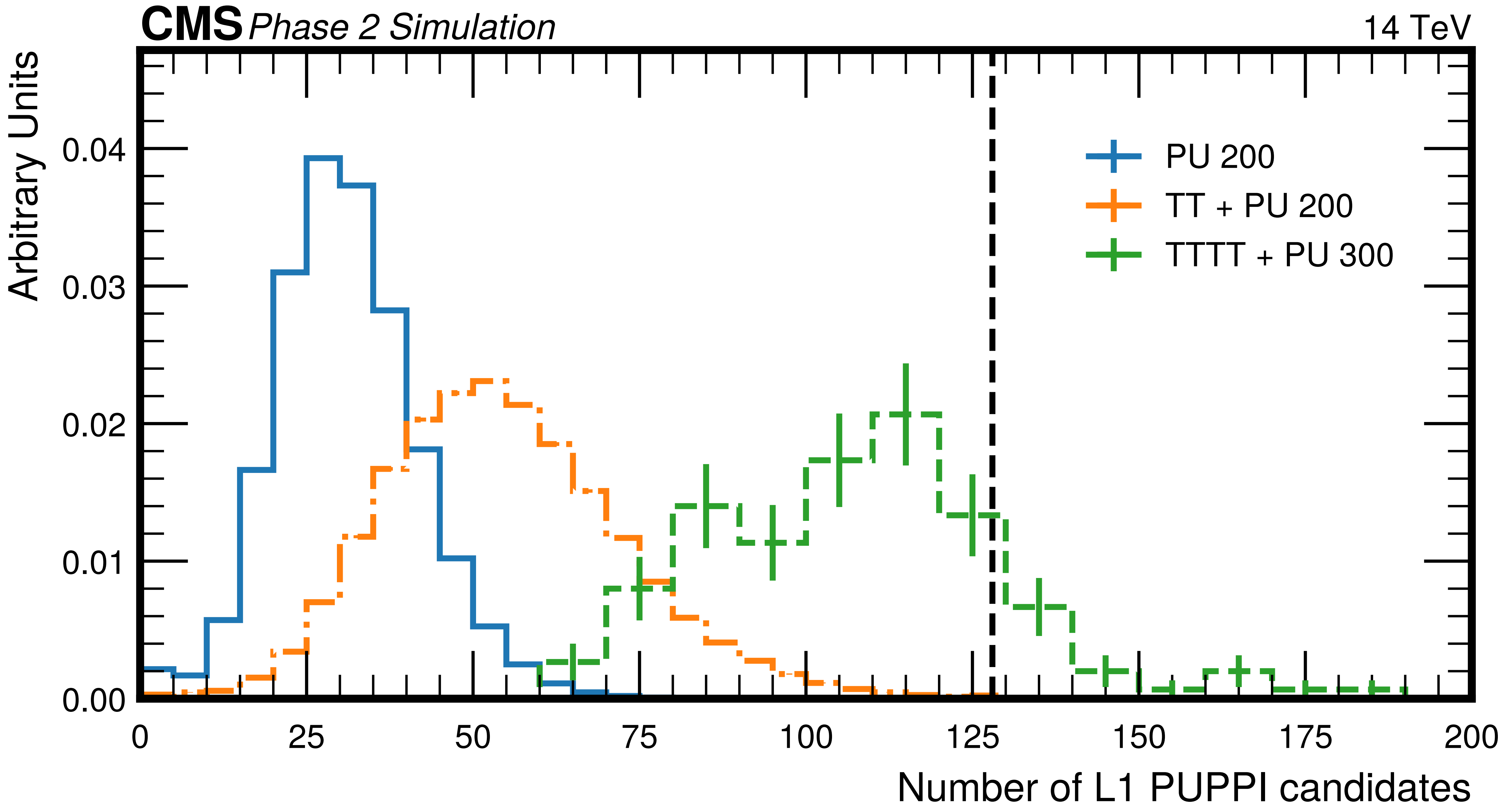
Summary

- CMS will be **extensively upgraded** for up to 200 PU conditions of High Luminosity LHC
- CMS has developed a solid solution to triggering for HL-LHC
- L1 Trigger system will be upgraded to **maintain and extend physics reach** compared to Runs 2 & 3
- New processing boards with powerful FPGA processors and high speed optics
- New algorithms exploiting **sophisticated reconstruction**:
  - Track Finding, Clustering, Particle Flow, Jet Reconstruction
- Algorithms are being successfully demonstrated in board-to-board tests across different board types
- Extensive use of **Machine Learning** for improved reconstruction and also final state selection
  - 25 billion ML predictions per second accounted for from current projects (expect this to increase!)
- Related talks at this workshop:
  - Firmware implementation of Phase-2 Overlap Muon Track Finder algorithm for CMS Level-1 trigger, Piotr Fokow
  - Testing of the Prototype CMS Global Level-1 Trigger for Phase-2, Hannes Sakulin
  - Pileup Mitigation in Hadron Forward Calorimeter at the Level-1 Trigger of the CMS experiment for the HL-LHC, Abhijeet Ghodgaonkar
  - The APx Board for the CMS Phase 2 L1 Calorimeter trigger: Testing and Performance, Isobel Ojalvo
  - Design, Construction, and Testing of the APOLLO ATCA Blades for Use at the HL-LHC, Jonathan Fulcher
  - Phase-2 CMS DAQ -- Growing from prototype boards to demonstrator systems, Jeroen Hegeman

Backup

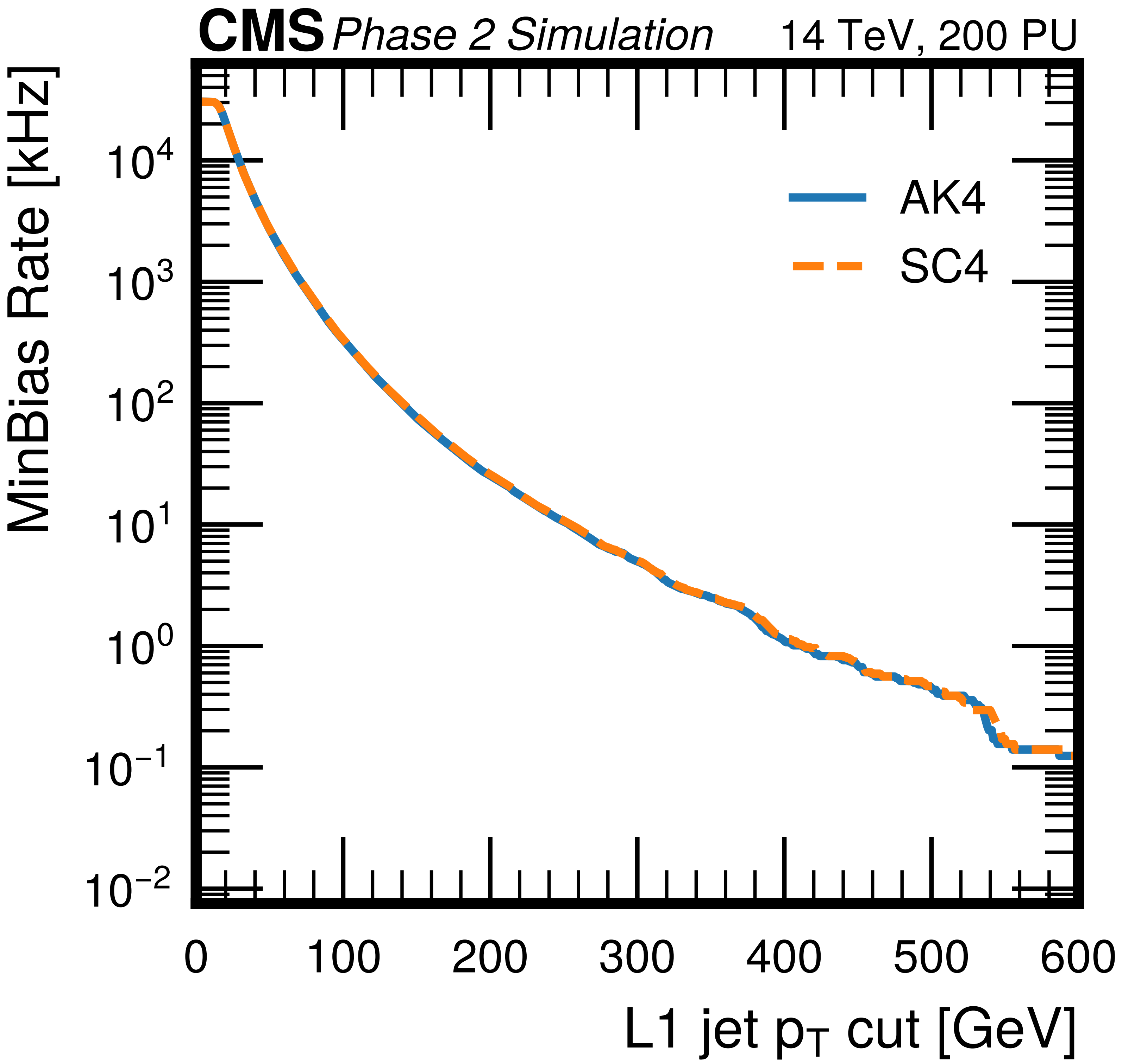
Deregionizer

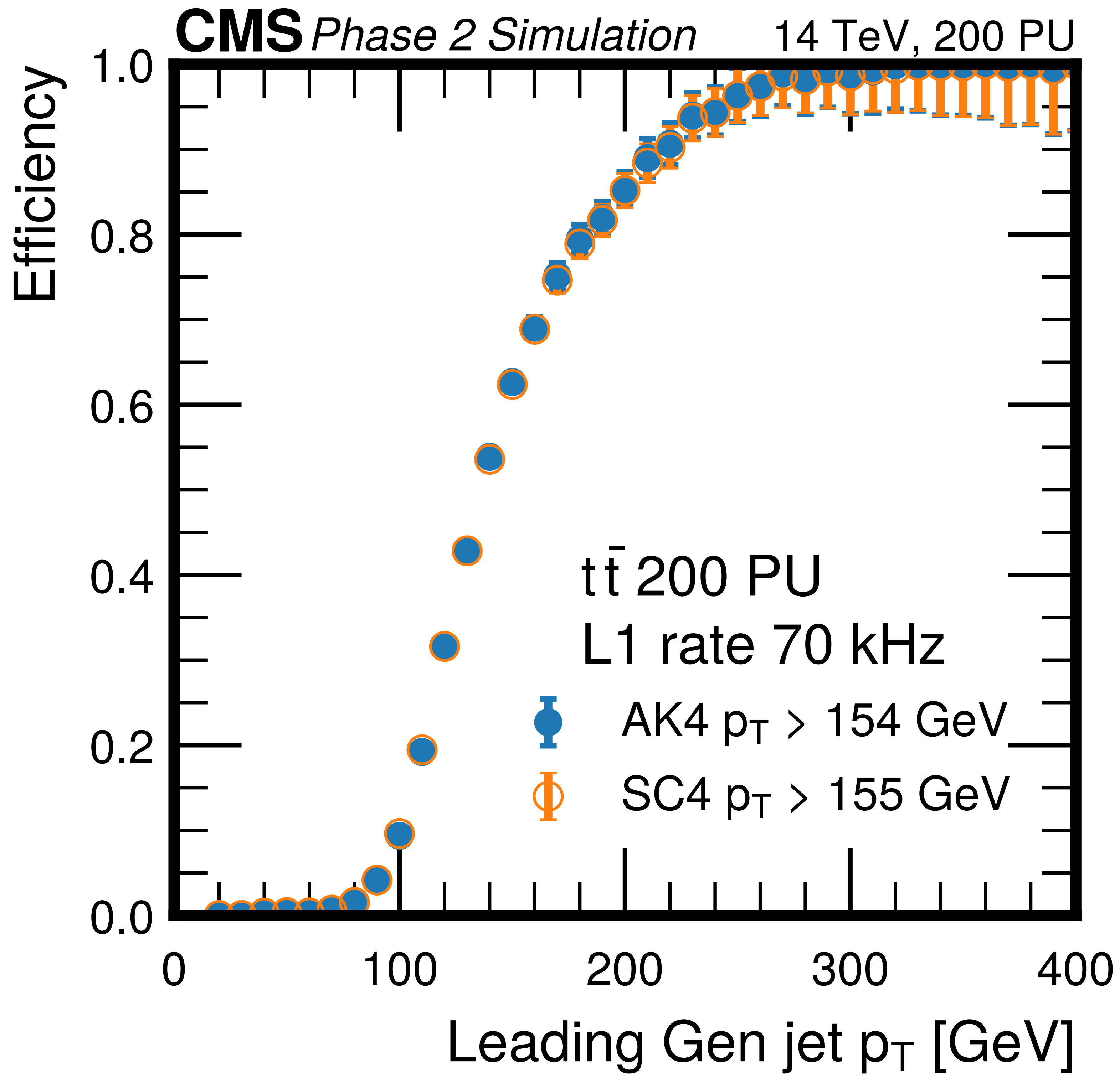
- Truncation of 128 particles in deregionizer motivated by multiplicity observed in high pileup simulations
- Typical event with no hard interaction and only pileup well below truncation limit
- High multiplicity topology tt with 200 pileup interactions has truncation of one particle for one event per thousand
- “Extreme scenario” tttt with 300 pileup has some more significant truncation, but many jets will be found regardless

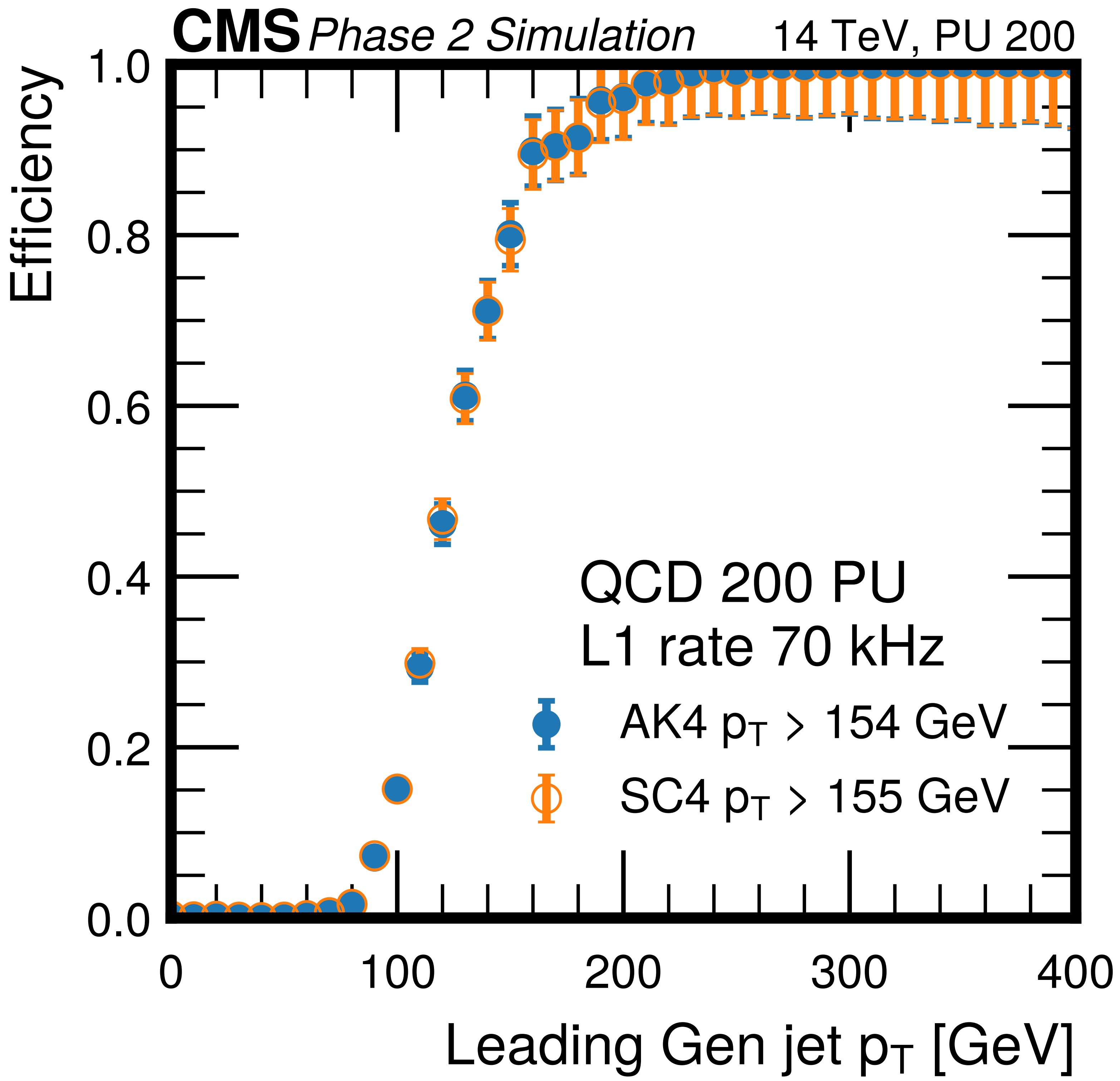


Jet performance 1 - Efficiency and Rate

- Left: online threshold vs rate in PU 200 events (no primary interaction)
- Centre: turn-on curve with thresholds chosen for a rate of 70 kHz in $t\bar{t}$ with 200 pileup
- Right: turn-on curve with thresholds chosen for a rate of 70 kHz in QCD with 200 pileup
- SC4 performance nearly identical to AK4

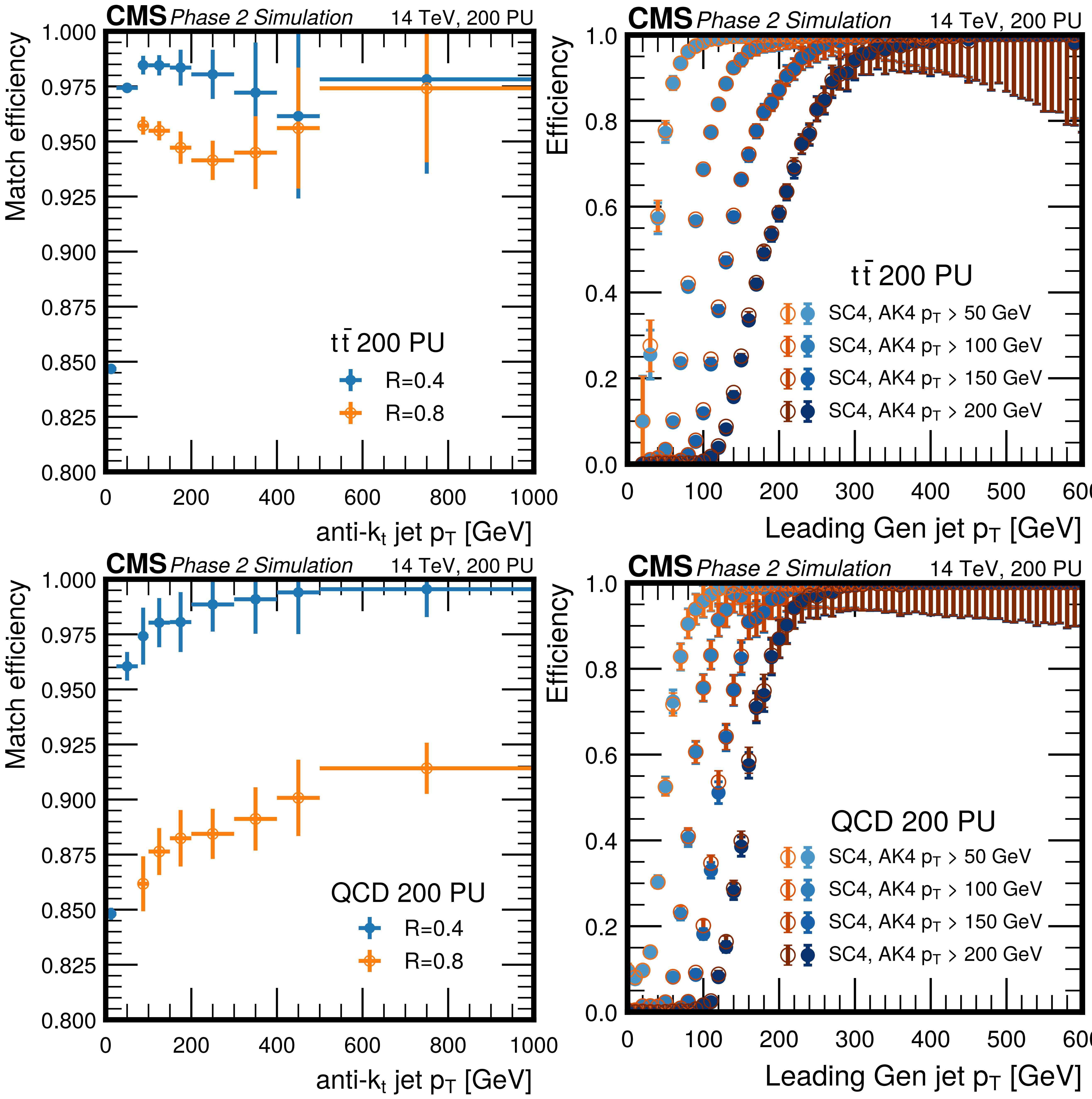






Jet performance 2

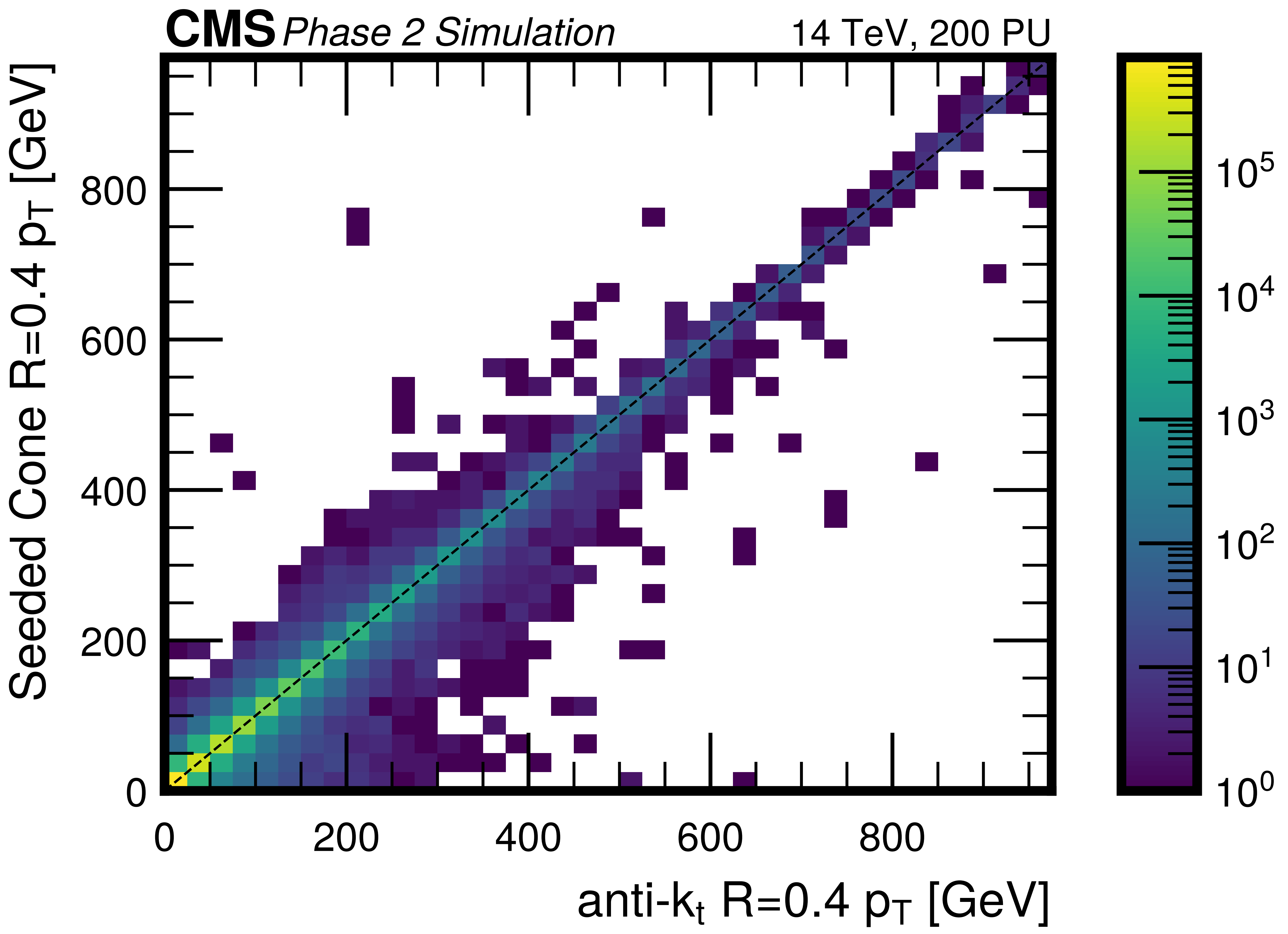
- Simulated events with tt + 200 pileup (top), QCD + 200 pileup (bottom) run through CMS detector simulation and L1T algorithm emulation
- Run Seeded Cone anti-k_t jet reconstruction on the same L1T PUPPI particles, for R=0.4 and R=0.8
- Left column: efficiency to match each anti-k_t jet to a Seeded Cone jet within $\Delta R \leq 0.2$ and p_T within 20%
- Right column: trigger efficiency as a function of simulated jet p_T for different L1T thresholds
- Seeded Cone generally matches well to anti-k_t, with some mismatches where the SC jet seeding can miss some particles / sub-jet that anti-k_t captures
- Trigger turn-ons are virtually identical for SC4 and AK4

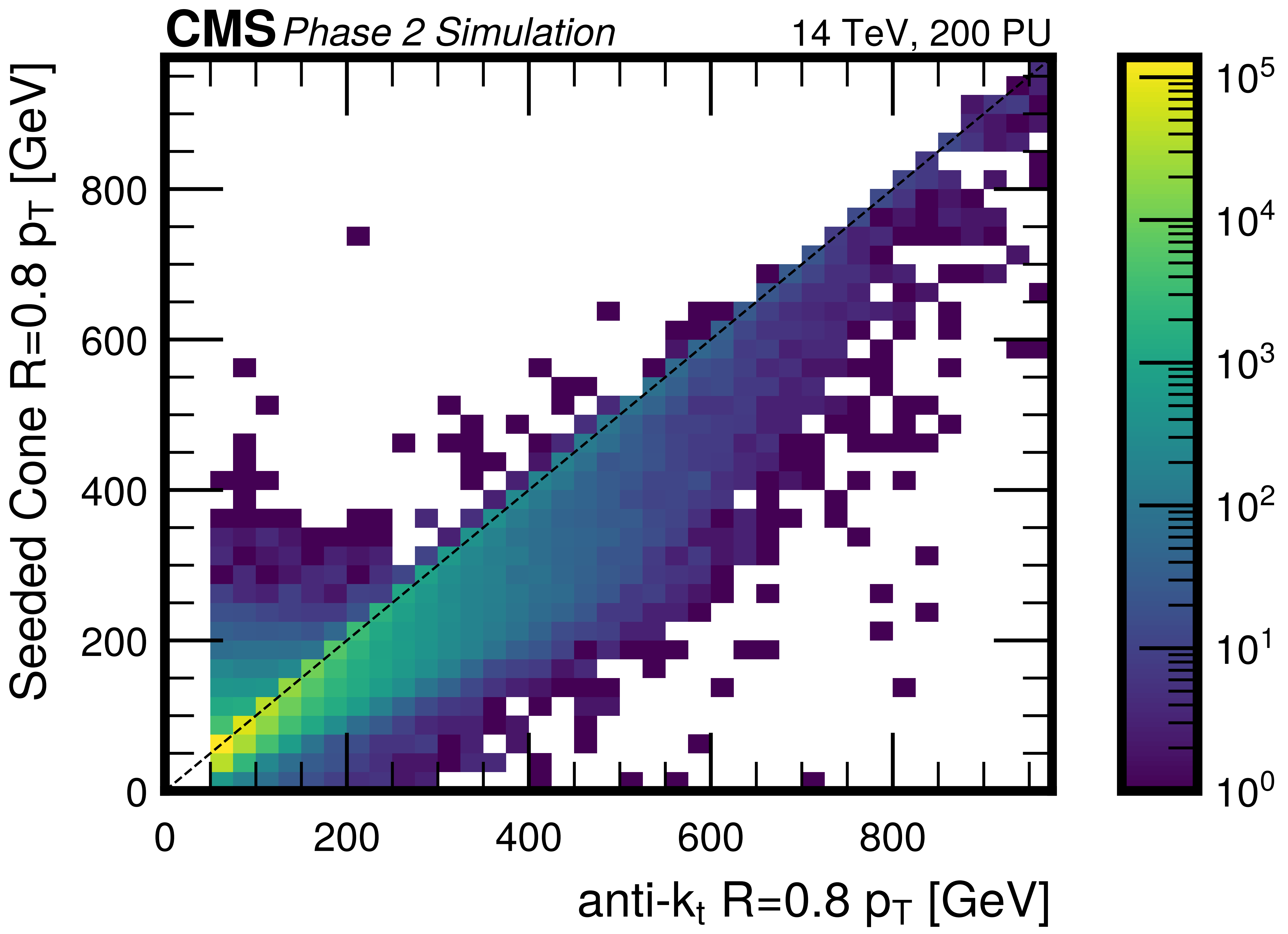


Jet performance 3 - SC to AK matching

Left: distribution of Seeded Cone p_T for jets matched within $\Delta R \leq 0.2$ of an anti-k_t jet with R=0.4 in simulated events of $\overline{t}t$ with 200 pileup.

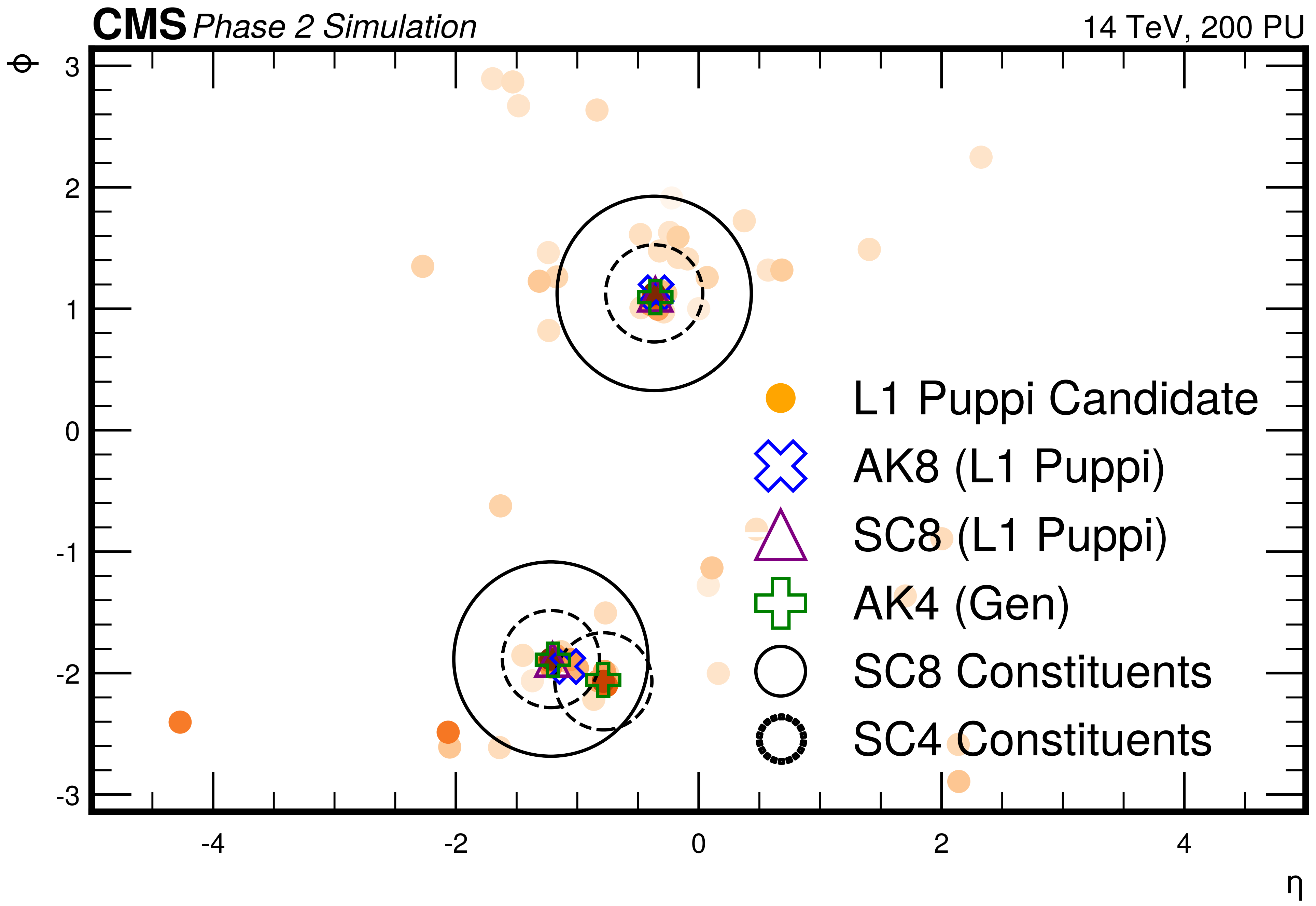
Right: distribution of Seeded Cone p_T for jets matched within $\Delta R \leq 0.2$ of an anti-k_t jet with R=0.8 in simulated events of $\overline{t}t$ with 200 pileup.

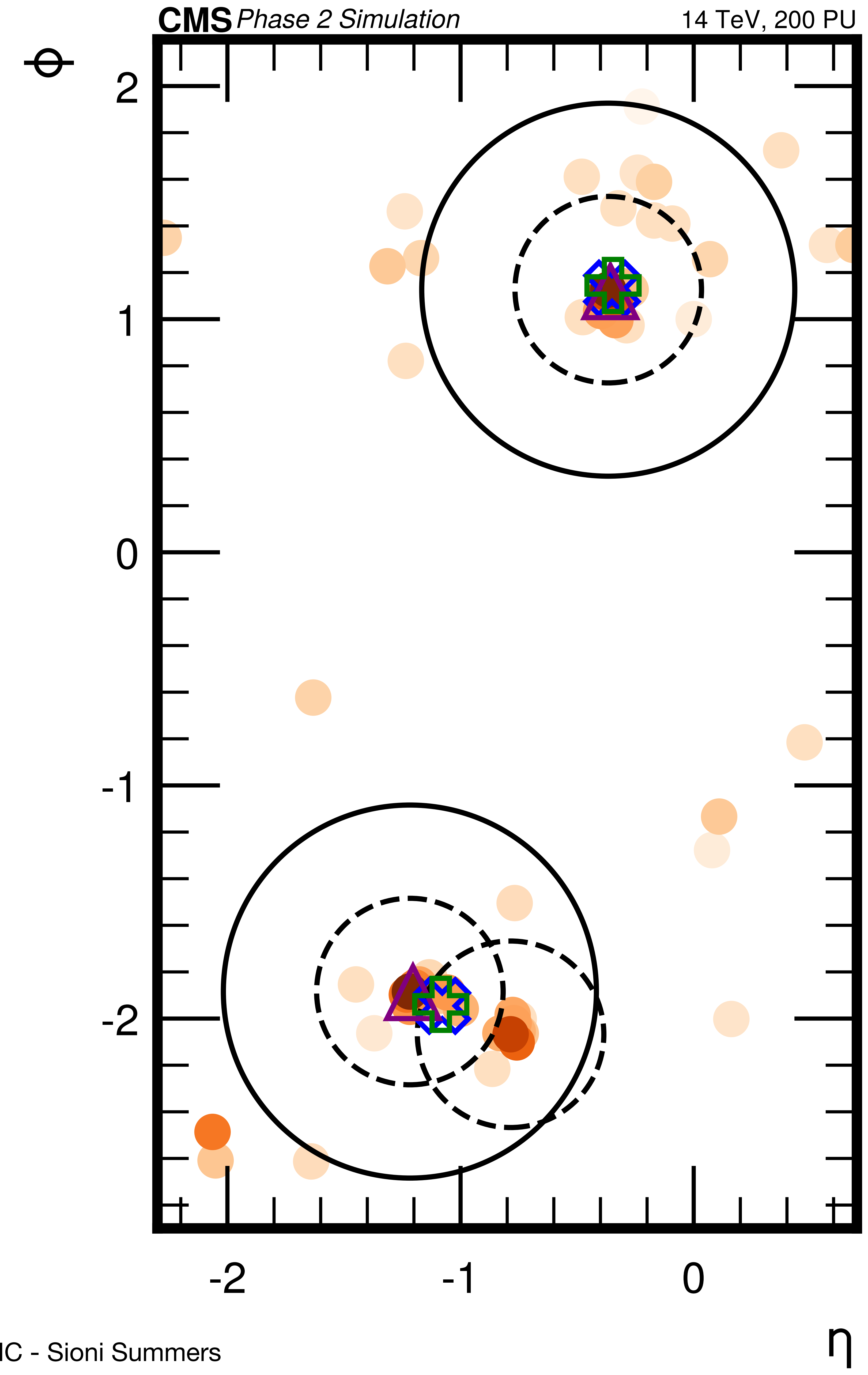




Event Displays 1

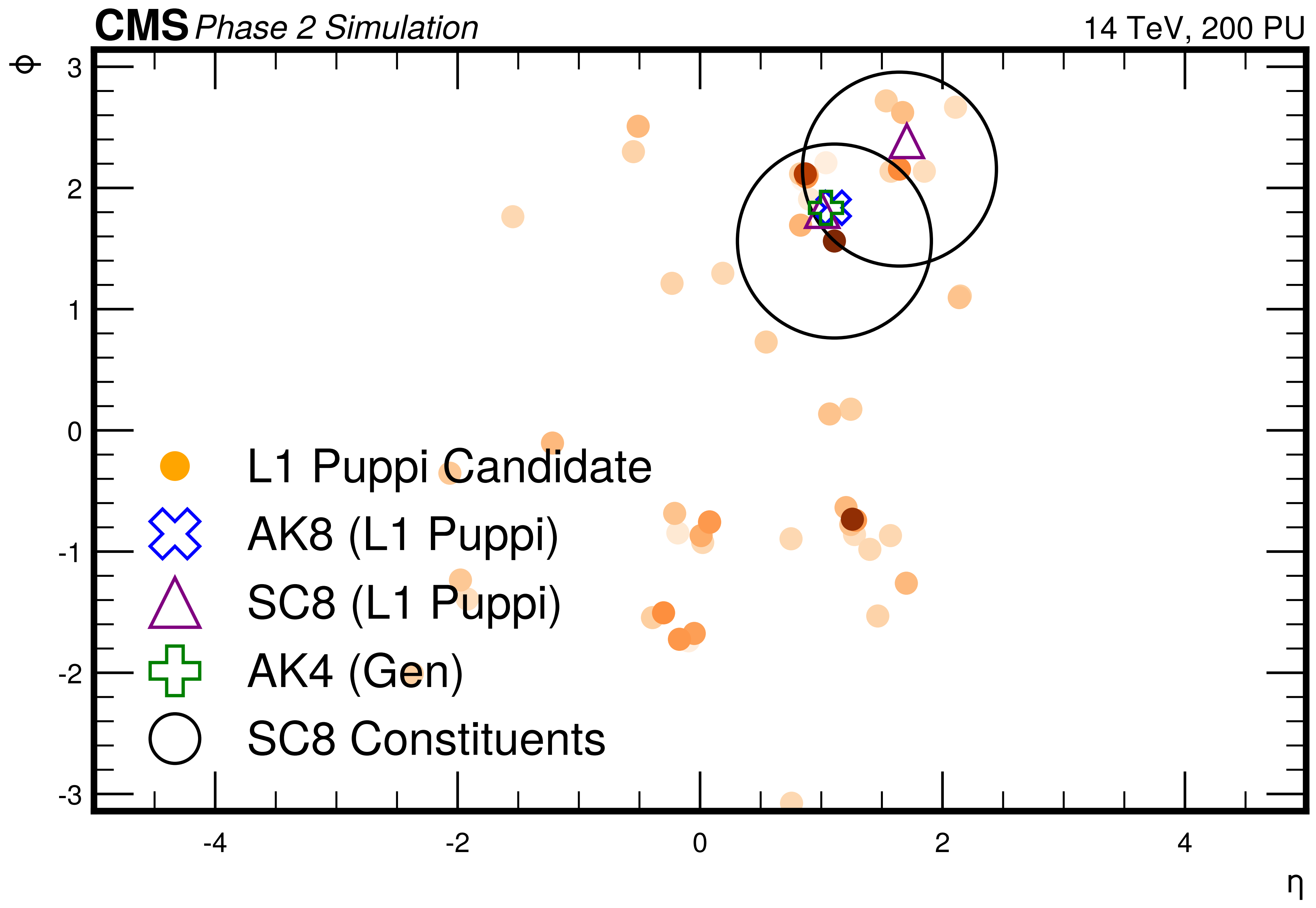
- "Event display" from sample of Higgs bosons produced in association with a W or Z boson
- Shows a Higgs with p_T of 400 GeV decaying to two b quarks around $(\eta,\phi) \simeq (0,1)$ and a W boson with p_T of 360 GeV decaying to light quarks around $(\eta,\phi) \simeq (-1,-2)$
- The W is reconstructed as two R=0.4 jets or one R=0.8 jet by both Seeded Cone and anti-k_t reconstruction.

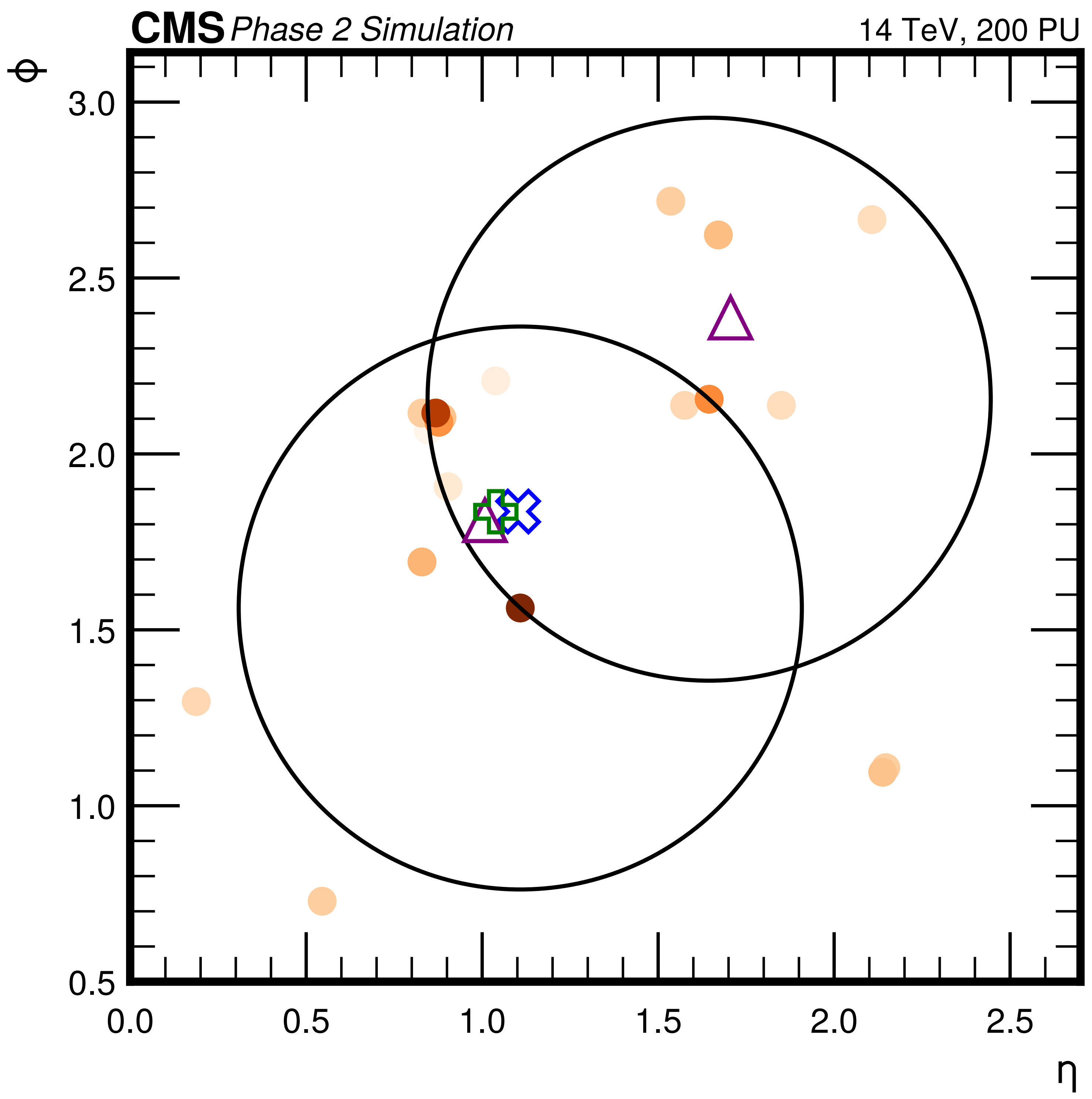




Event Displays 2

- “Event display” from sample of $t\bar{t}$ with 200 pileup
- Shows a case where one AK8 jet is reconstructed as two SC8 jets due to the limitations of casting the cone around the single highest p_T particle seed with no reclustering





Event Displays

- Event displays from $t\bar{t}$ with 200 PU where one Gen AK4 jet is resolved as two SC4/AK4 jets or one SC8/AK8 jet