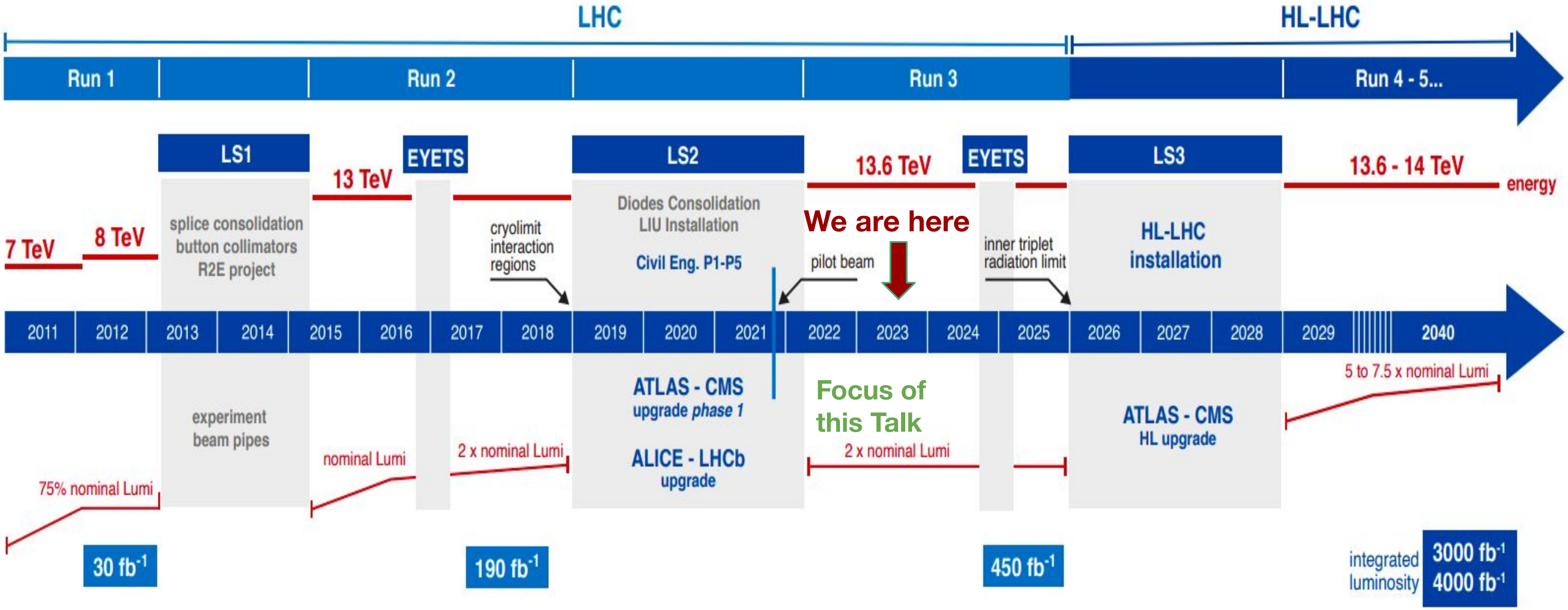
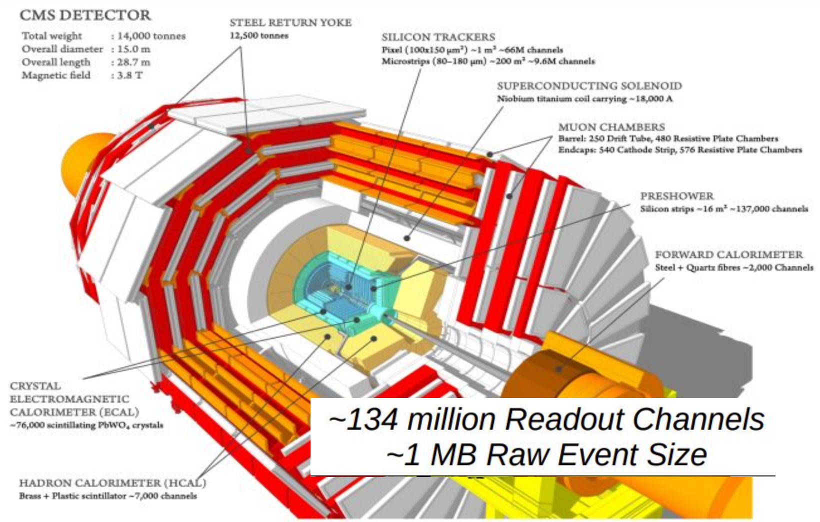
CMS High Level Trigger Performance for Run3

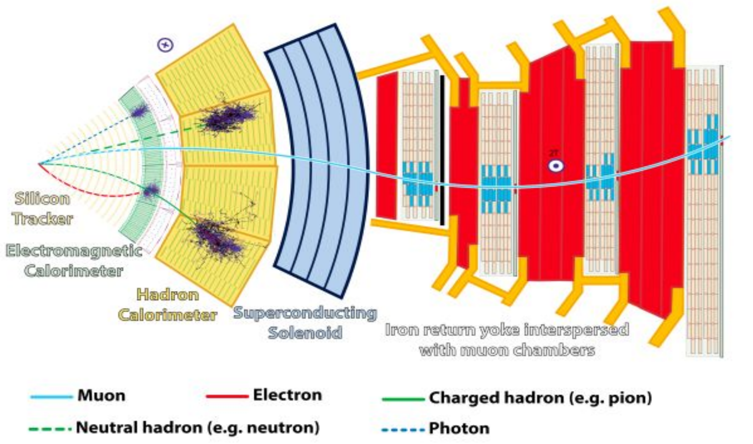
Sanu Varghese on behalf of the CMS Collaboration

LHC schedule



The CMS Detector



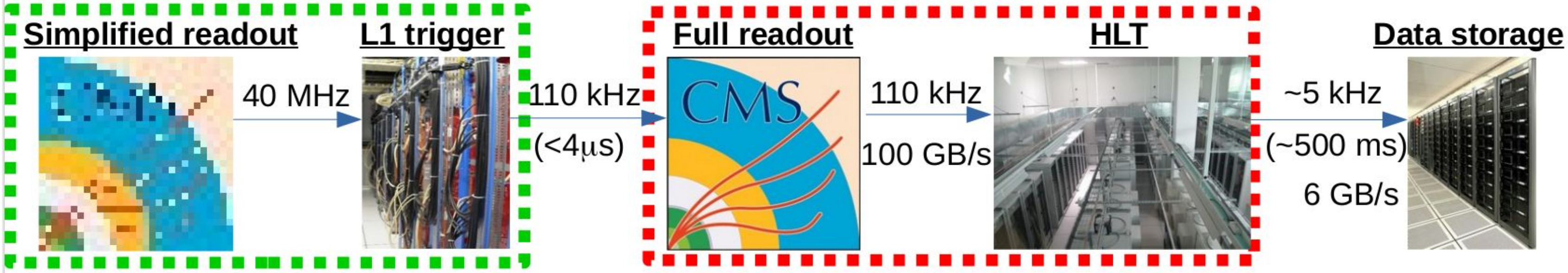


- **Global Event Reconstruction using Particle Flow (PF)**
  - Identify individually all particles
  - Combines information from all subsystems

CMS Trigger System

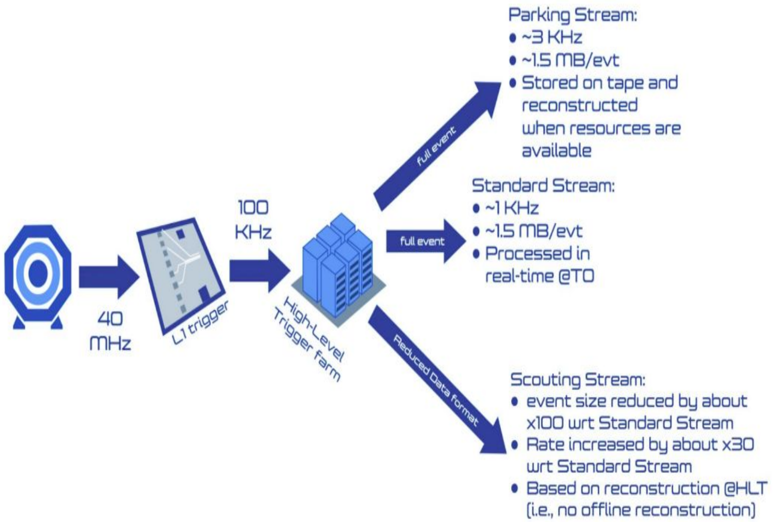
Two Level Trigger system to select events of physics interest - 40 MHz => ~5 kHz on average.

- **Level 1 (L1T) : Hardware, course/fast readout, < 3.8 µs , only muon/calo info , ~110 kHz (max rate)**
- **High Level (HLT) : Software based, Full event readout, < 0.5 s , ~5 kHz , ~15 GB/s output limit.**



Run 3 Trigger Strategy

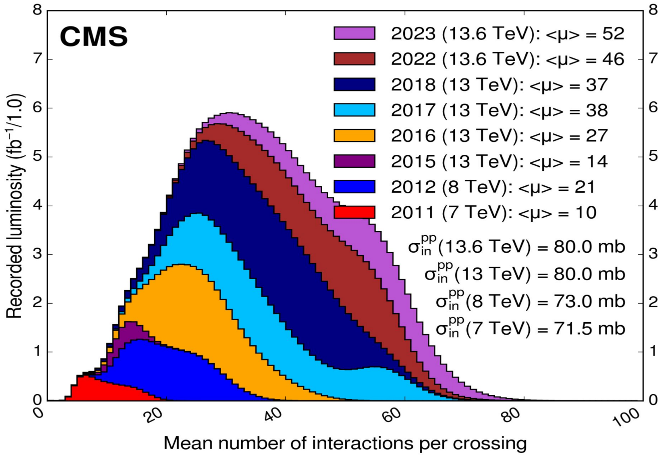
- **Standard Physics :**
  - Promptly reconstructed within 48 hrs.
  - ~2 kHz on average
- **Data Parking (Delayed Reconstruction) :**
  - Reconstructed based on resource availability
  - ~3 kHz on average
- **Data Scouting (Trigger Level Analysis) :**
  - No offline reconstruction
  - Only HLT info saved - ~7 kB vs ~1 MB for Standard
  - Analysis done with HLT objects only

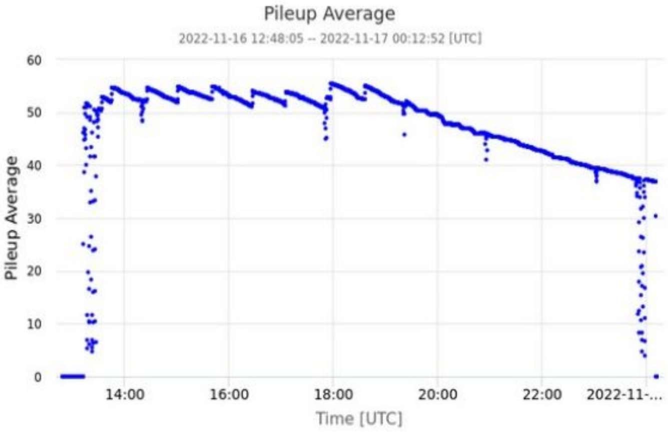


Goals: Identify better/faster, trigger on new objects, cover more phase space!

LHC Run 3 :

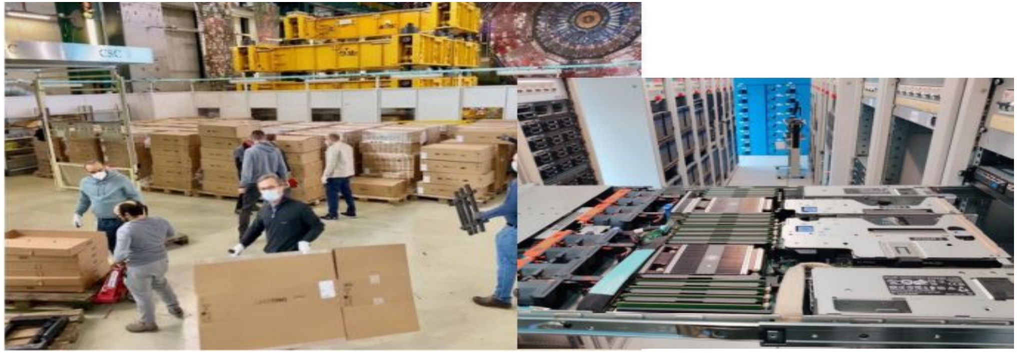
- Higher energy : 13 -> 13.6 TeV
- Higher Inst luminosity/pileup
- Lumi Levelling
    - <μ> = 63 (2023)
    - <μ> = 55 (2022)
- Inst lumi upto 2.15 x 10^{34} cm^{-2} s^{-1} (2023)





New HLT Farm

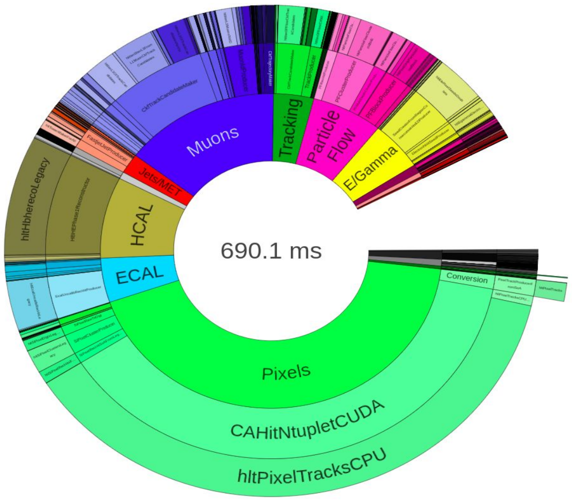
source



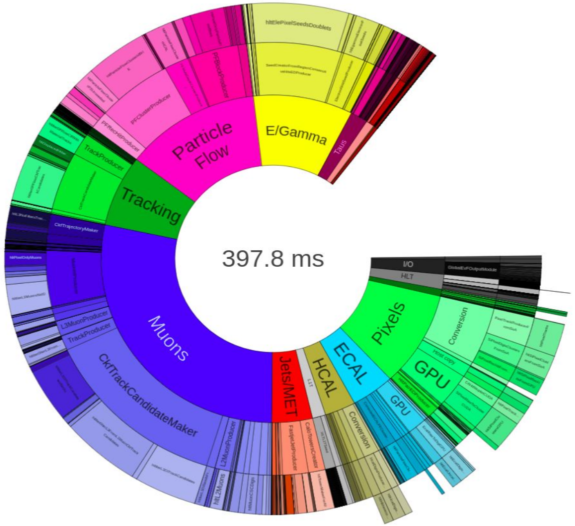
- **Brand new HLT farm in operation from start of Run 3!**
- **200 nodes each equipped with 2 GPUs (2 x NVIDIA T4)**
  - 2x AMD EPYC 7763 “Milan” 64-core processors (128 Cores, 256 threads)
  - 25600/51200 CPU cores/threads and 400 GPUs in total
  - Can sustain up to pileup 65 for a max L1 rate of 110 kHz.
  - Decommissioned Run 2 farm nodes provide additional capacity for offline processing
- **Considering expanding farm capacity by ~20% for 2024**

Reconstruction using GPUs

CMS-DP-2023-004



CPU Only

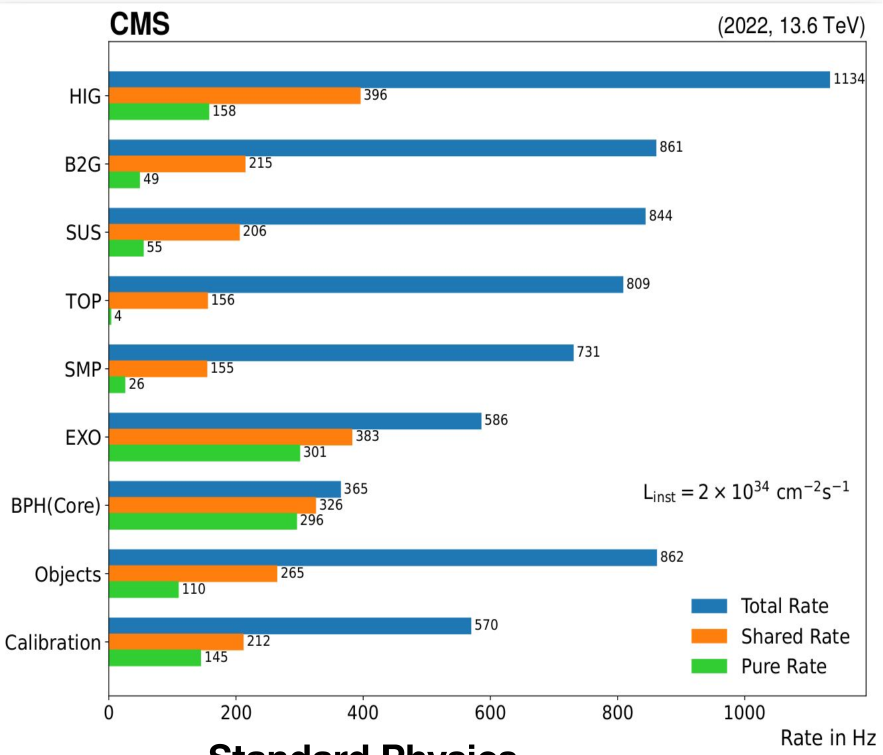


CPU + GPU

- **Currently HCAL, ECAL, Pixel Local Reconstruction, Pixel-Only Track and Vertex Reconstruction running on GPUs**
  - ~40% reduction in avg HLT timing (~70% increase in throughput)
  - ~30% reduction in power consumption for the same throughput

Run 3 Trigger Menu

CMS-DP-2023-042



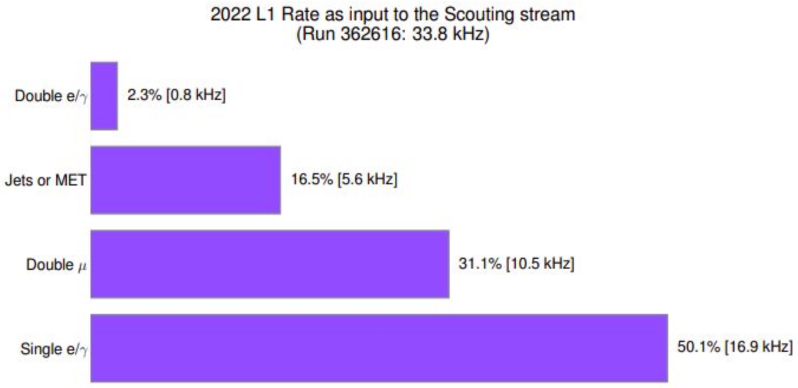
Standard Physics

Table 15: HLT thresholds and rates of some generic triggers in the Run-3 HLT menu. The rates were obtained from measurements during an LHC fill in November 2022 and have been scaled to a luminosity of $2.0 \times 10^{34}\,\mathrm{cm}^{-2}\,\mathrm{s}^{-1}$.

HLT algorithm	Rate
Isolated muon with $p_T > 24$ GeV	250 Hz
Isolated electron with $E_T > 32$ GeV	182 Hz
Particle flow (PF) based $p_T^{miss} > 110$ GeV	81 Hz
4 PF jets with $p_T > 70/50/40/35$ GeV with 2 b tag	57 Hz
Two isolated tau leptons with $p_T > 35$ GeV	54 Hz
Muon with $p_T > 50$ GeV	51 Hz
Two electrons with $E_T > 25$ GeV	21 Hz
AK4 PF jet with $p_T > 500$ GeV	16 Hz
Two same-sign muons with $p_T > 18/9$ GeV	10 Hz

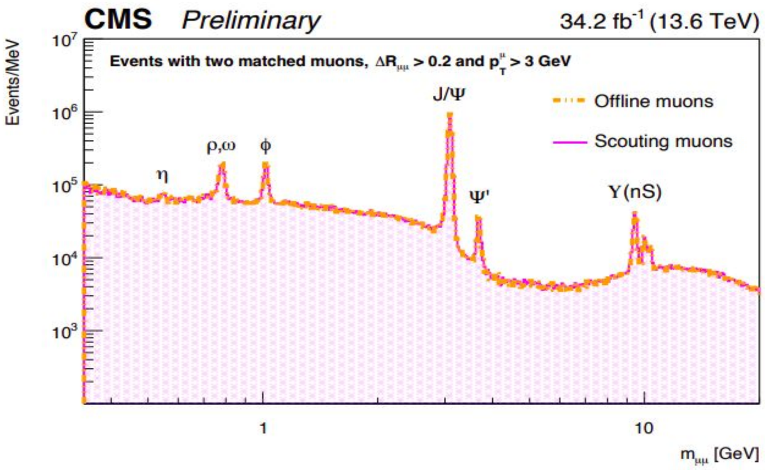
~600 HLT "paths"

Data Scouting

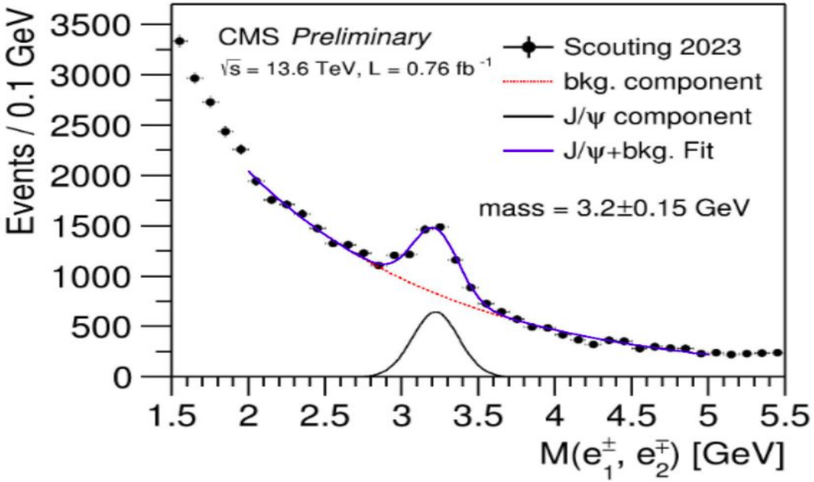


- Trigger Level Analysis using only HLT objects/info
  - No offline reconstruction
- Scouting in Run 2 using selected objects -
  - Jets/muons , ~5 kHz
- For Run-3, scouting have been extended to almost all physics objects
  - muons, electrons, photons, jets, pixel tracks
- **Rate allocation increased to ~30 kHz (~30% of the total L1 rate!) with an event size of ~7 kB (vs ~1MB for Raw Event)**
  - Special version of Particle Flow using only pixel tracks (reconstructed using GPUs!)
  - Gain in reconstruction time allows PF to be run on larger fraction of input events.

Data Scouting



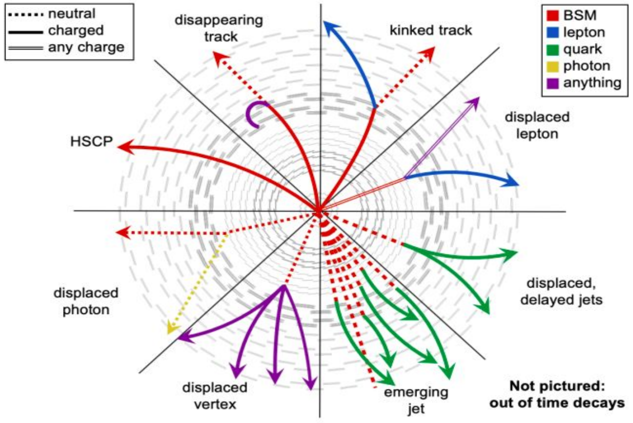
Dimuon spectrum



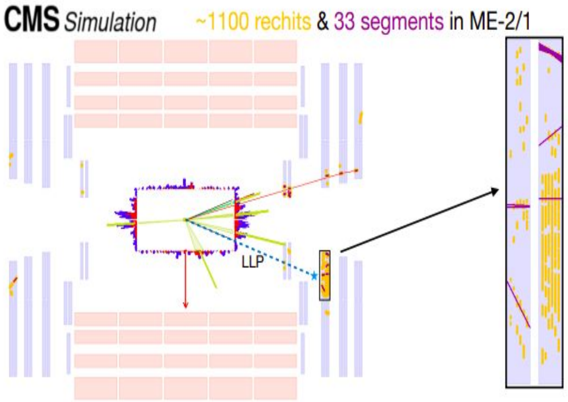
Di electron spectrum (J/Ψ Peak)

- Comparable performance for scouting and offline muons.
- Improved reconstruction of low-pT electrons in scouting

Long Lived Particle Triggers

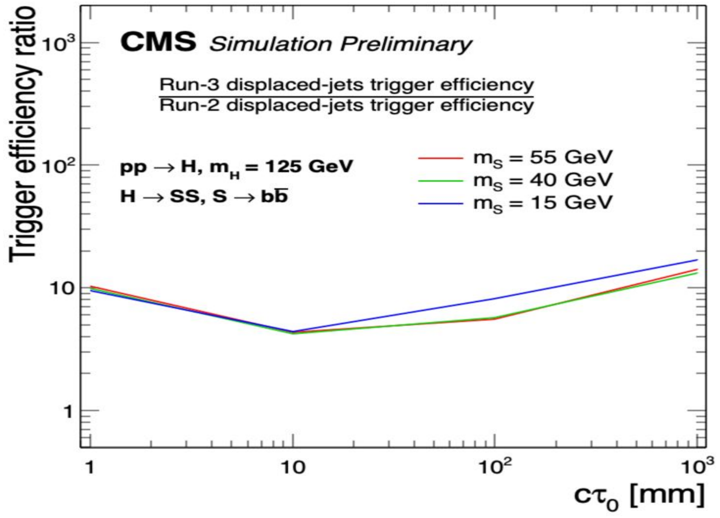


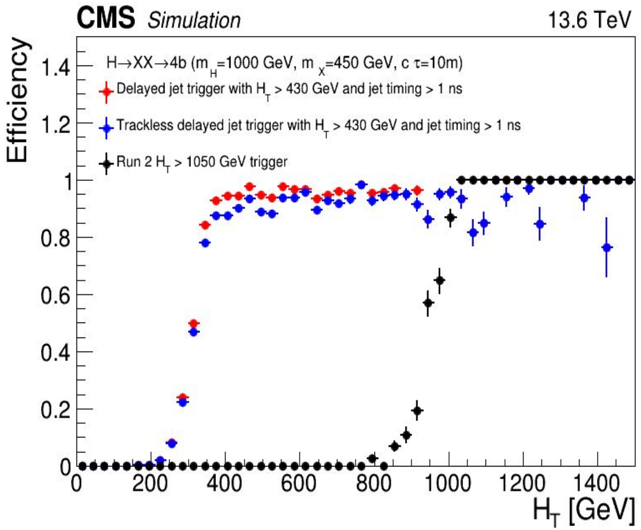
- Run 2 LLP triggers had limited scope
  - few dedicated L1 triggers
  - High kinematic thresholds
  - Mostly reliant on pixel/tracker or MET (eg disappearing tracks)
- Increased focus on developing new dedicated triggers targeting a wide range of LLP signatures like
  - Using calorimeter (Hcal/Ecal) timing info to tag delayed Jets /electons/photons
  - Trigger on Particle showers in the muon chambers (ie use the muon system as a sampling calorimeter!).
  - Improved Displaced Jet triggers with more optimised requirements.
- Several of these HLT paths seeded by dedicated L1 triggers
- Lowered thresholds using Data Parking Strategy (2023)



Long Lived Particle Triggers

CMS-DP-2023-043



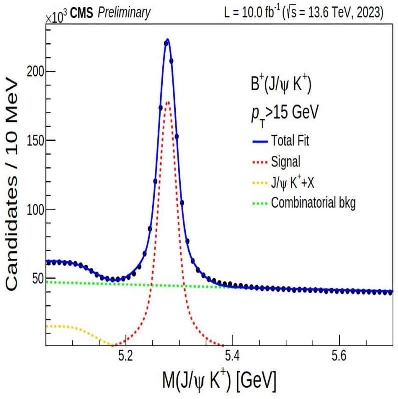


Huge gain in acceptance on benchmark signal models with LLP signatures!

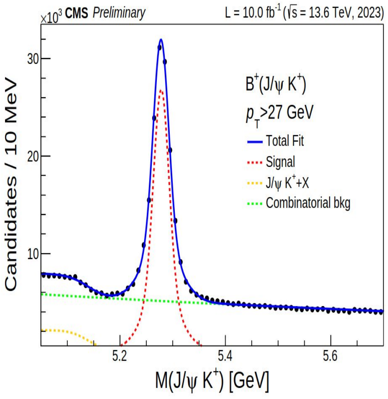
B Physics Triggers

- In Run-3 we increased the rate of B-physics triggers using delayed reconstruction strategy
    - New Inclusive Di-muon trigger, triggers on opposite sign di-muons for $m_{\mu\mu}$ = [0.2 , 8.5] GeV
    - ~7x the yields of old Run 2 trigger for $B^+ \rightarrow J/\psi K^+$
- New low mass Di-electron triggers for Rk/Rk* measurement
    - Dynamic prescaling - enabling lower threshold triggers during lumi-decay phase of the LHC fill

	Run 3	Run 2
Invariant mass range	0.2 – 8.5 GeV	2.9 – 3.3 GeV
Dimuon p_T	>4.9 GeV	>25 GeV
Vertex fit probability	>0.5%	>0.5%
Leading muon p_T	>4 GeV	-
Trailing muon p_T	>3 GeV	-



Run 3 Trigger

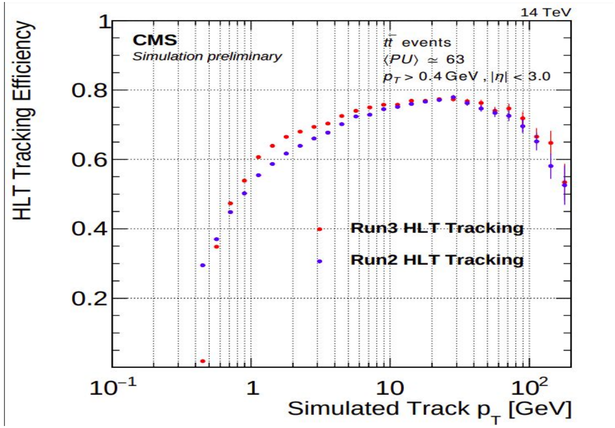


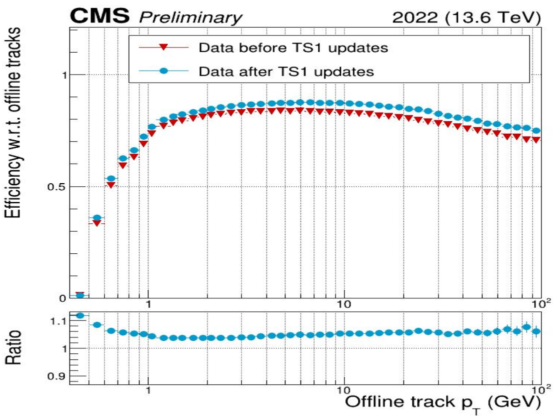
Run 2 Trigger

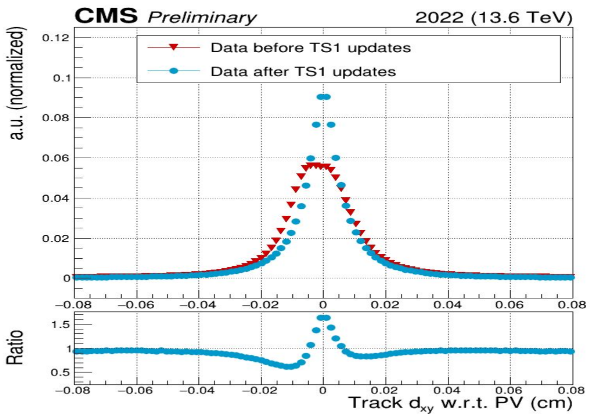
Tracking

CMS-DP-2022/014 , CMS-DP-2023/028 ,

- New Tracking Algo based on optimised pixel tracks (Patatrack)
  - Allows tracking to be reduced to a single iteration. - faster reconstruction
  - Better performance than the previous Run 2 tracking
  - Good performance in 2022 despite high instantaneous luminosity/pileup.



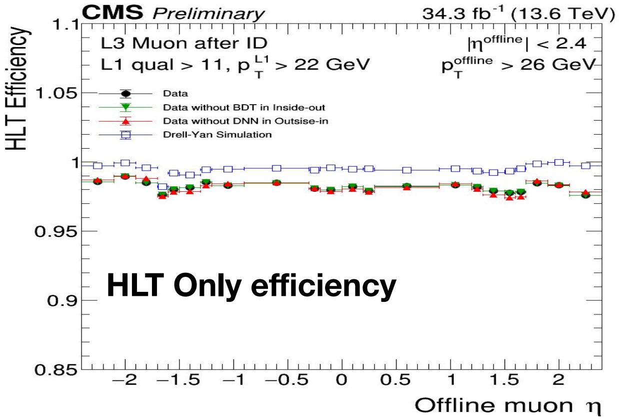


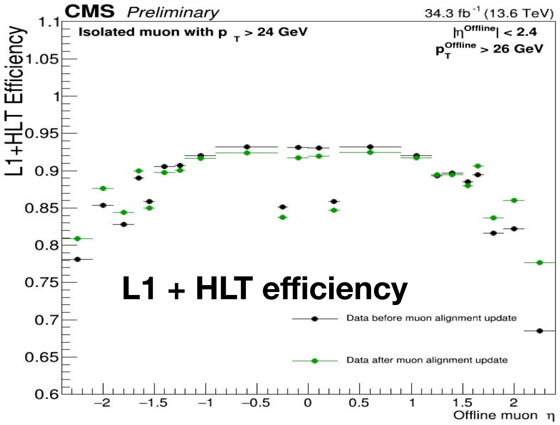


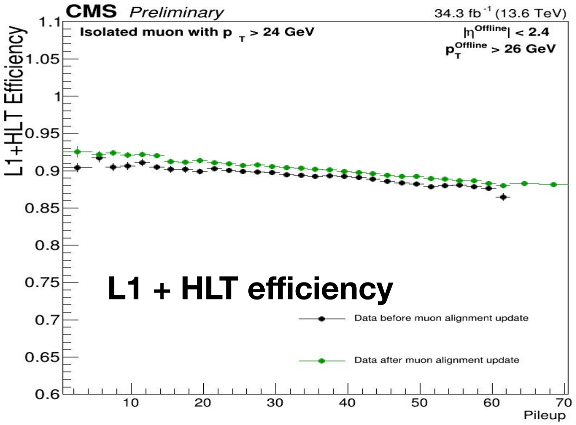
Muons

CMS-DP-2023/017

- New BDT seeding technique which speeds up HLT reconstruction by ~18%!
- Selection efficiency at HLT almost 99%.
- Efficiency almost flat with pileup.
- Improved performance after endcap muon-tracker alignment update (Oct 2022)

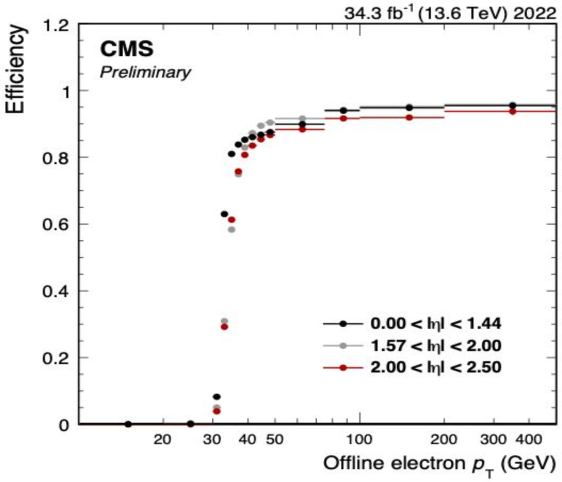


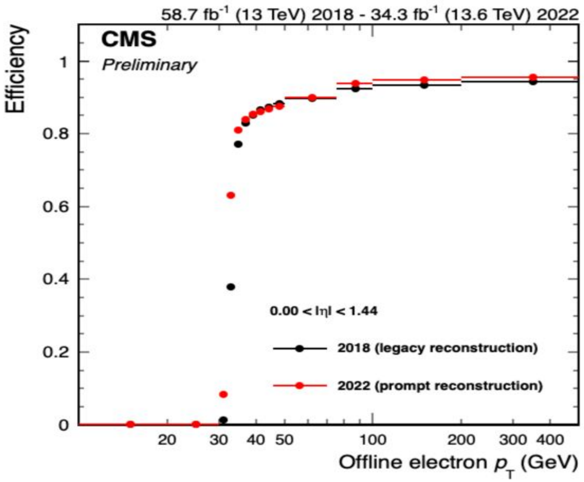


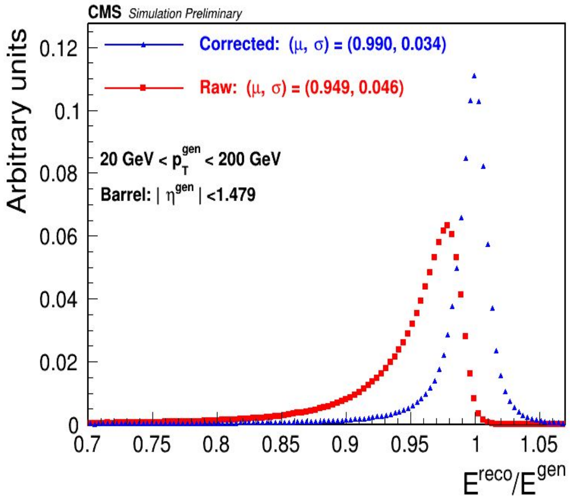


Electrons and Photons

CMS-DP-2023-015





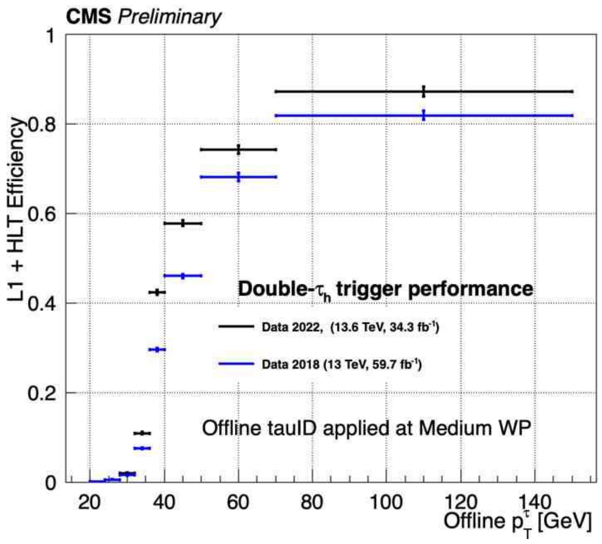


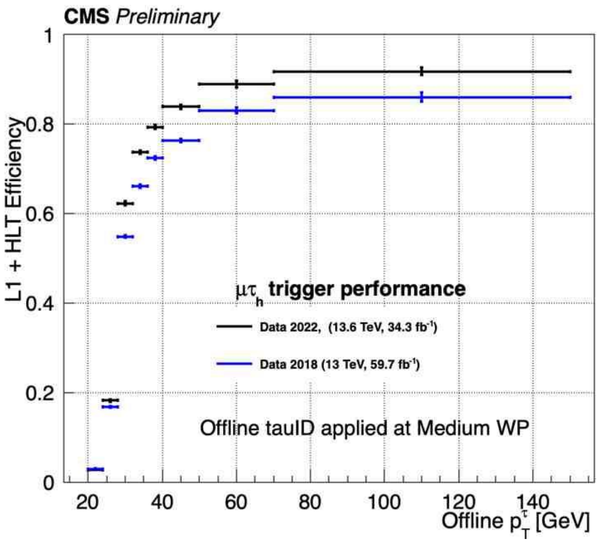
- BDT based ECAL energy correction
- Similar performance as 2018

Taus

[CMS-DP-2023-024]

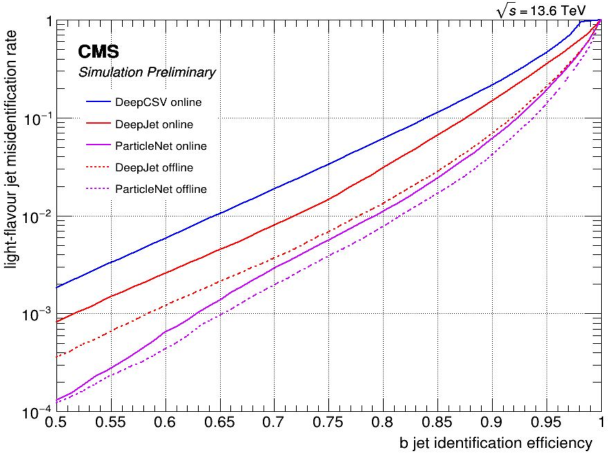
- New ML based algorithms introduced in trigger paths, replacing the cut based ones (Run 2)
- CNN based filter at L2 stage using only calo and pixel track info
- DeepTau algorithm at L3, using PF information
- improved performance w.r.t. 2018

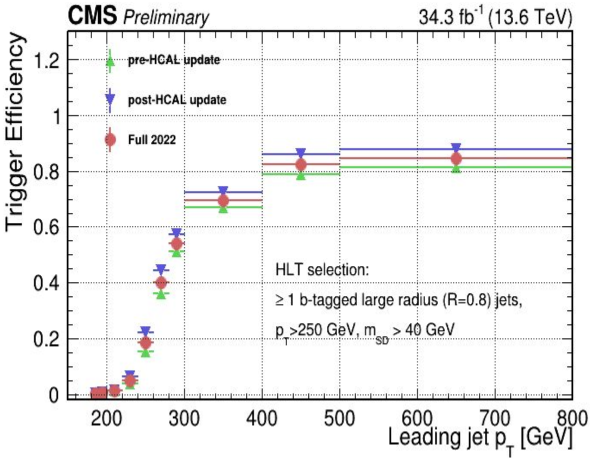


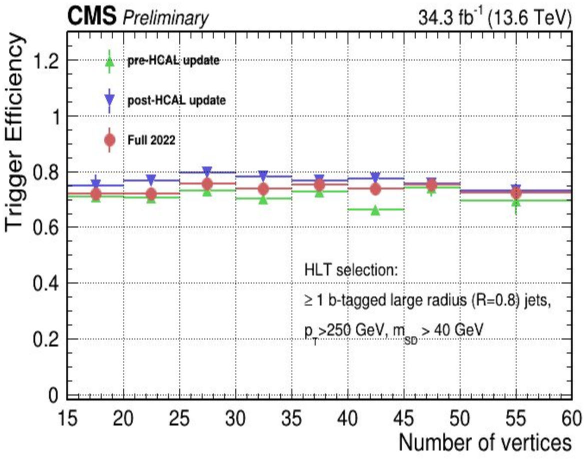


B-Tagging

CMS-DP-2023-021



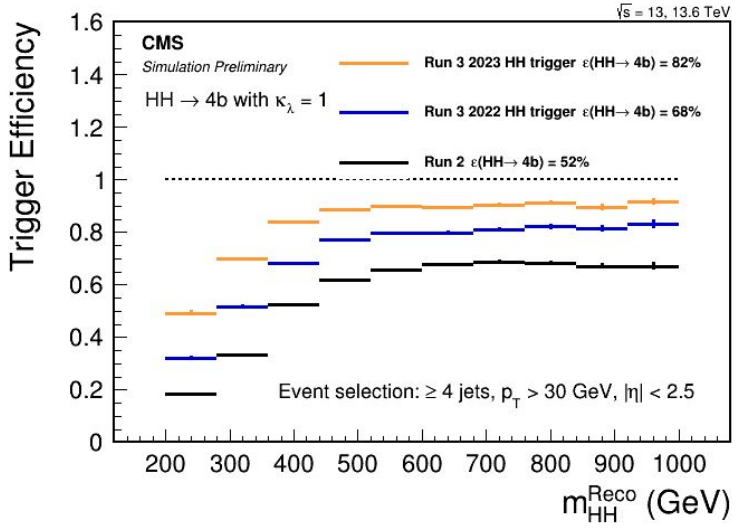


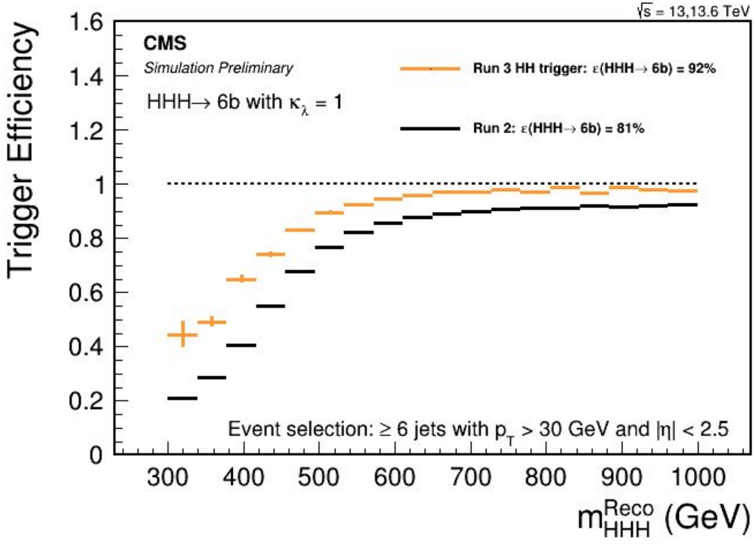


- New CNN (Deep Jet) and Graph Network (ParticleNet) based algorithms replacing the DNN based DeepCSV of Run 2.
- Much lower misidentication rate at same b-jet identification efficiency

HH => 4b CMS-DP-2023/050

- Particle Net based HH -> 4b trigger in Run 3

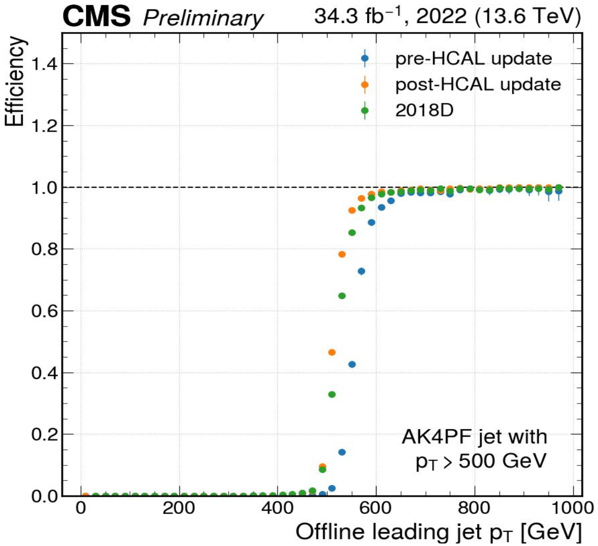


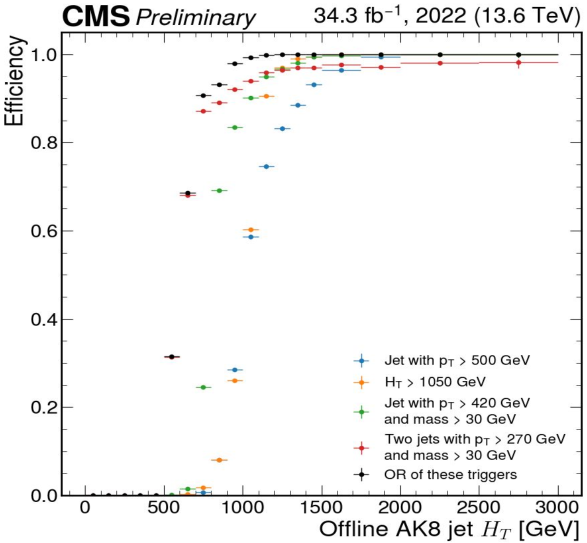


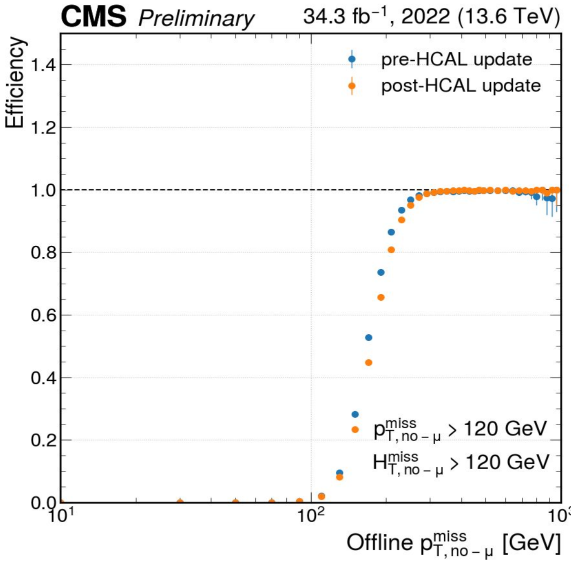
- Large efficiency increase in Run 3
- Lowered thresholds in Run 3 thanks to delayed reconstruction (~180 Hz)

Jet and Missing E_T

CMS-DP-2023-016







- New Boosted AK8 Jet grooming algorithm : TrimMass => Soft-Drop
- Good performance in Run 3

Conclusion

- Many improvements have been implemented in the CMS HLT after Run-2
  - Usage of GPUs
  - New Long Lived Particle signatures
  - Better reconstruction algorithms
  - Larger rate for scouting and delayed reconstruction
- See also Efe's talk on improvements to Level 1 Trigger
- The first Run-3 data confirm the good performance of the CMS Trigger
- Expect much more than just luminosity gains from Run 3
- **Stay tuned!**

Additional References

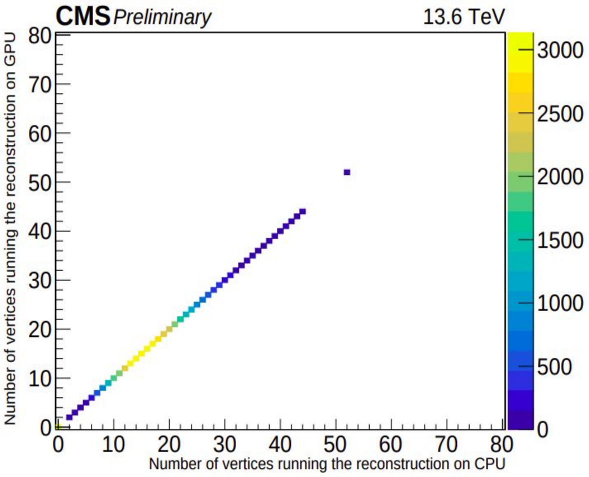
- https://twiki.cern.ch/twiki/bin/view/CMSPublic/HighLevelTriggerRunIIIResults
- Patatrack Paper
- https://cms.cern/news/first-collisions-reconstructed-gpus-cms
- https://twiki.cern.ch/twiki/bin/view/CMSPublic/LumiPublicResults

Backup

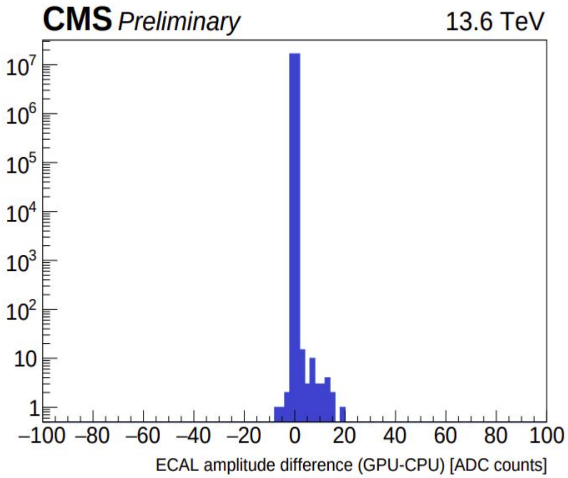
Major detector/condition updates in 2022

- **LHC Technical Stop 1 (TS1) updates** (deployed end of Sep 2022)**:**
  - BPix Layer 1 reverse bias voltage increased from 150 -> 300 V to cope with the radiation damage damage effects
  - Corresponding tracker alignment updates
- **HCAL scale update** (deployed mid Oct 2022):
  - Hcal barrel ($|eta| < 1.3$) respone increased by ~17%

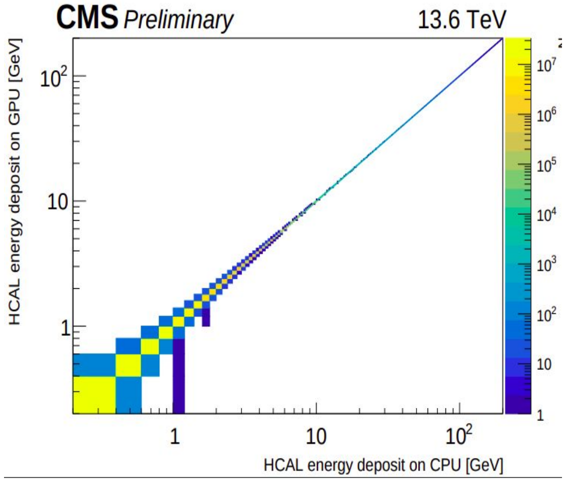
GPU-CPU comparison



# of pixel vertices



ECAL Amplitude differnce



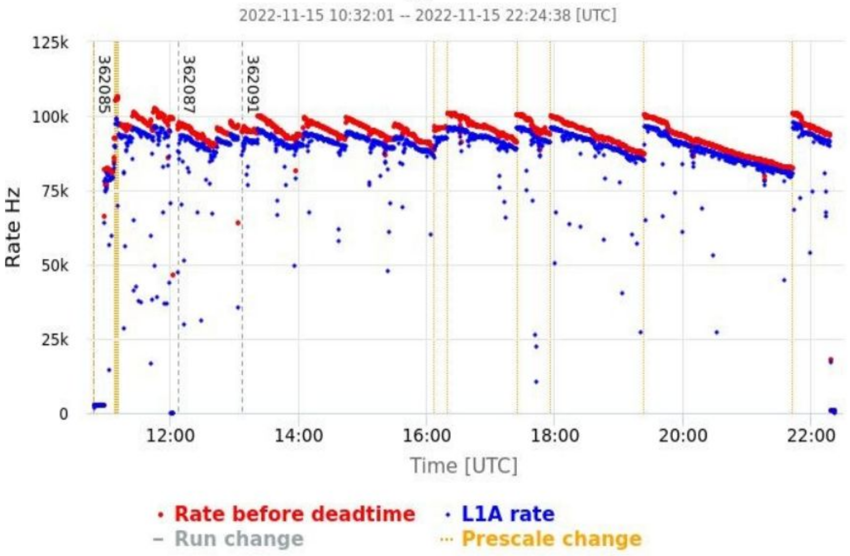
HCAL energy responce

Some representative triggers used in HH Analyses

Trigger	Requirement	Rates at HLT at 2x10^34 cm-2s-1
2023 HH trigger	HT > 280 GeV, 4 jets with pT > 30 GeV, PNet@AK4(mean 2 highest b-tag score) > 0.55	180 Hz
2022 HH trigger	4 jets pT > 70, 50, 40, 35 GeV, PNet@AK4(mean 2 highest b-tag score) > 0.65	60 Hz
2018 triple b-tag [2,3]	HT > 340 GeV, 4 jets pT > 75, 60, 45, 40 GeV, 3 b-tags with DeepCSV > 0.24	8 Hz
Run 3 tau-triggers [4]	Double medium DeepTau taus with pT > 35 GeV $\|\eta\| < 2.1$ Double medium DeepTau taus with pT > 30 GeV $\|\eta\| < 2.1$, PFJet 60 GeV Single loose DeepTau on hadronic tau with pT > 180 GeV $\|\eta\| < 2.1$	50 Hz 20 Hz 17 Hz
Run 3 MET-trigger [5]	Missing transverse energy (MET) (no muon) > 120 GeV, HT (no muon) > 120 GeV	42 Hz

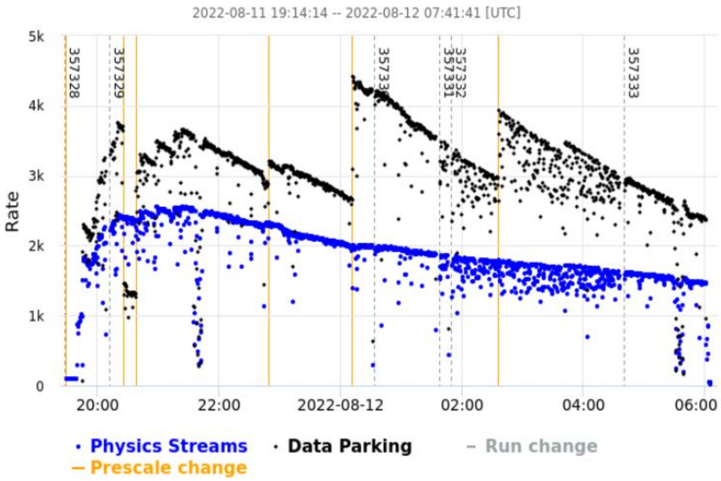
Rates (2022)

L1 trigger rate



~100 kHz L1T

HLT rate



~2 kHz in prompt reconstruction

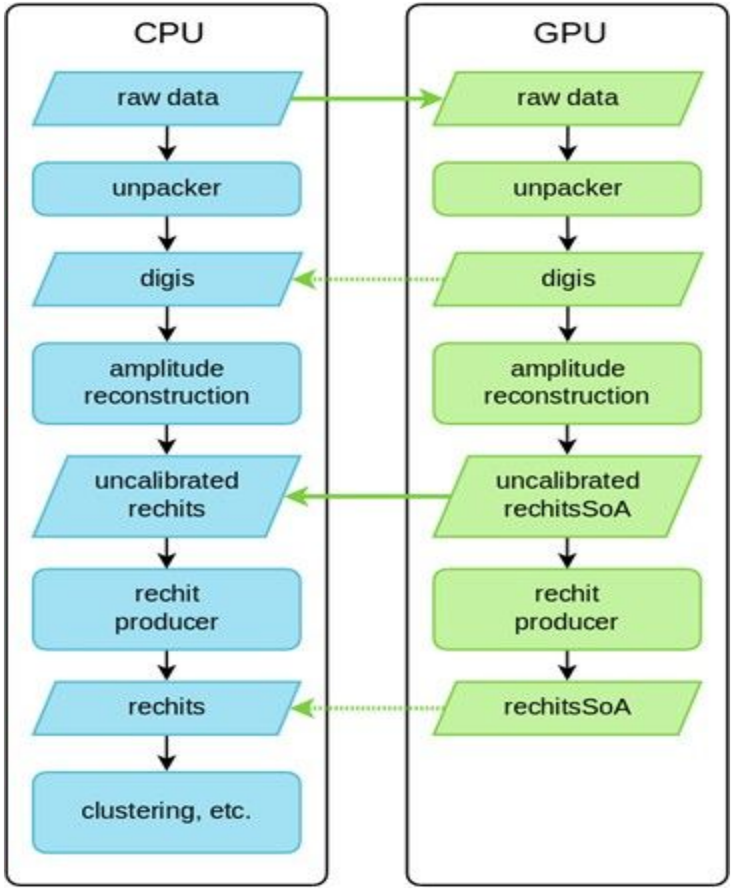
~3 kHz in delayed reconstruction

ECAL Reconstruction using GPU

- Unpacking of the raw data from the detector into digitized hits (digis) - packed raw data from the 54 ECAL front end drivers (FEDs) is accumulated on the host in one piece of memory that is then transferred to the device

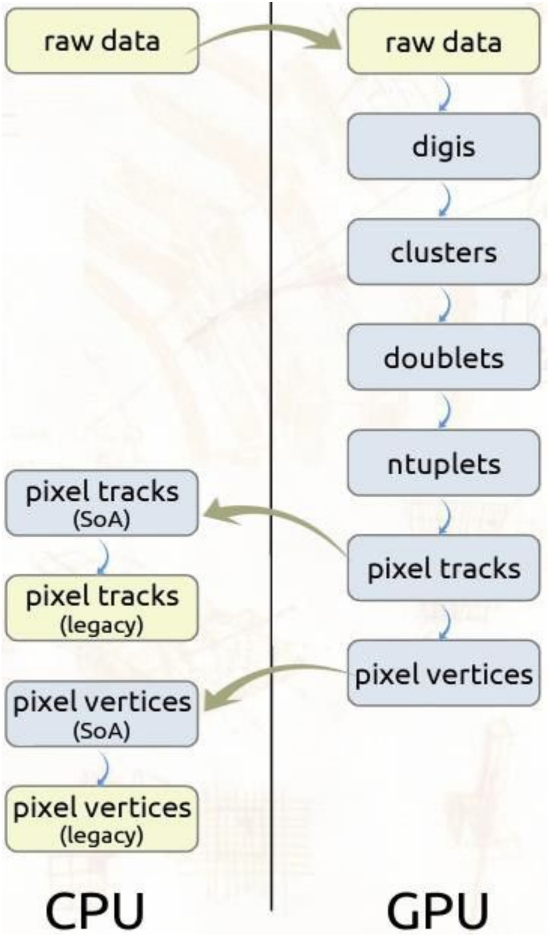
- Amplitude reconstruction from uncalibrated reconstructed - minimization and amplitude reconstruction is split into several kernels, taking care to avoid computation of common variables.

- Calculation of energies from the uncalibrated RecHits by applying various scale factors and calibrations – currently not implemented fully on GPUs



Pixel Reconstruction using GPU

- Copying over the RAW data to GPU
- Kernal launch to perform tasks in parallel
- unpack the raw data
- cluster the pixel hits
- form hit doublets
- form hit ntuplets (triplets or quadruplets)
- fit the track parameters and apply quality cuts
- reconstruct vertices
- Move the results back to the host to be used further downstream



Run 3 HLT Tracking

The CMS High-Level Trigger (HLT) runs a version of the full event reconstruction optimized for fast processing. Since the start of Run 3, the HLT makes use of a heterogeneous computing farm. In Run 3, HLT tracking is based on a single iteration of the Combinational Kalman Filter, seeded by pixel tracks reconstructed by the Patatrack algorithm, which can be offloaded to GPUs

To be used as seeds, Patatrack pixel tracks are required to:

- Be built with at least three pixel hits
- Have transverse momentum $p_T > 0.3$ GeV
- Be consistent with a leading pixel vertex Pixel vertices from primary interactions are reconstructed at the HLT from pixel tracks with at least four hits and $p_T > 0.5$ GeV.