An overview of the CMS high level trigger upgrade for the HL-LHC era

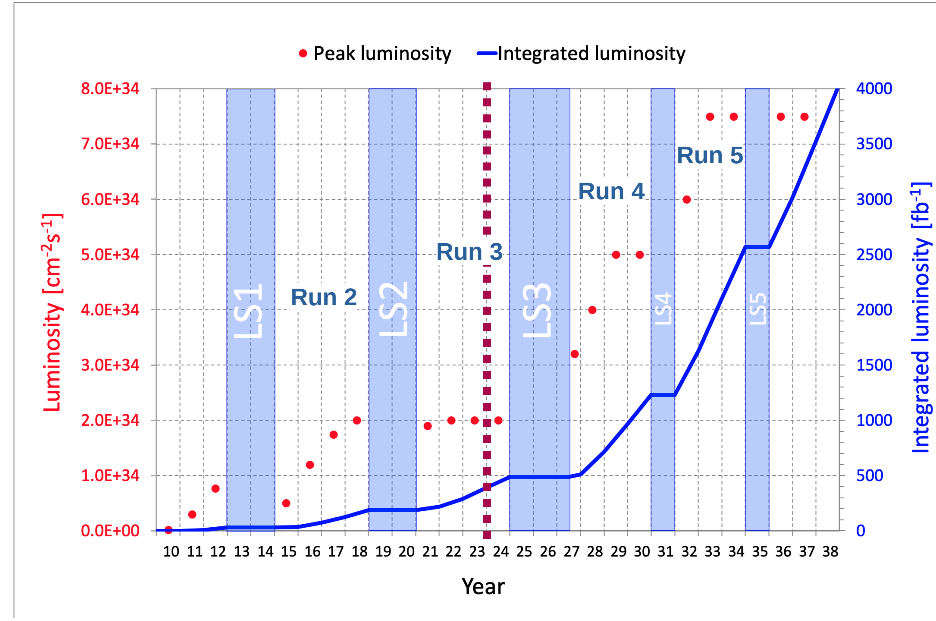
Soham Bhattacharya (DESY, Hamburg)

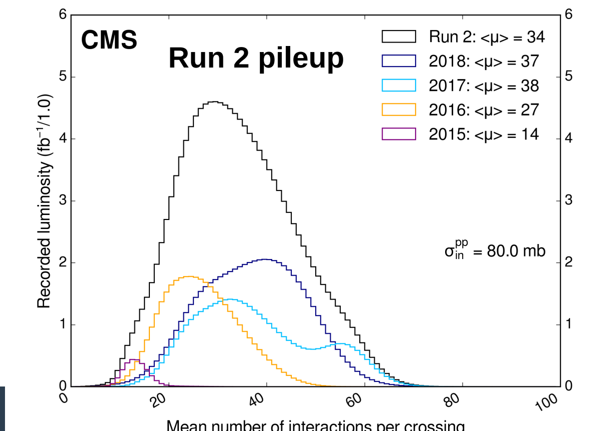
28th November, 2023

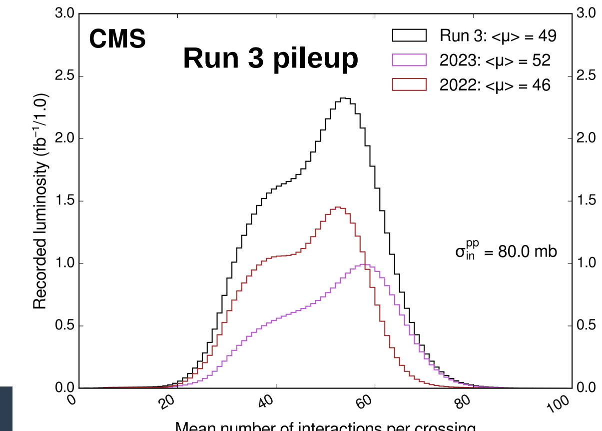
CP3, UC Louvain

The high luminosity LHC (HL-LHC) era

- **Significant increase in beam luminosity (intensity)**
  - Higher pileup (pp interactions per bunch crossing)
  - Higher particle flux
  - Increased radiation damage (detector ageing)
- **New challenges for data acquisition and triggering**
  - Redesign to cope with data rate





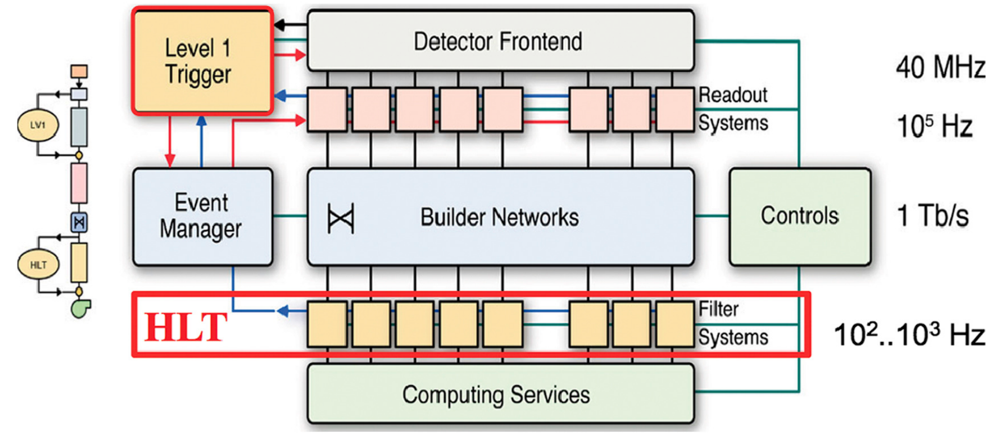


Phase-2 pileup

		$\mathcal{L}$	$\langle$ PU $\rangle$
Run 4	Baseline	$5 \cdot 10^{34}\,\mathrm{cm}^{-2}\,\mathrm{s}^{-1}$	140
Run 5	Ultimate	$7.5 \cdot 10^{34}\,\mathrm{cm}^{-2}\,\mathrm{s}^{-1}$	200

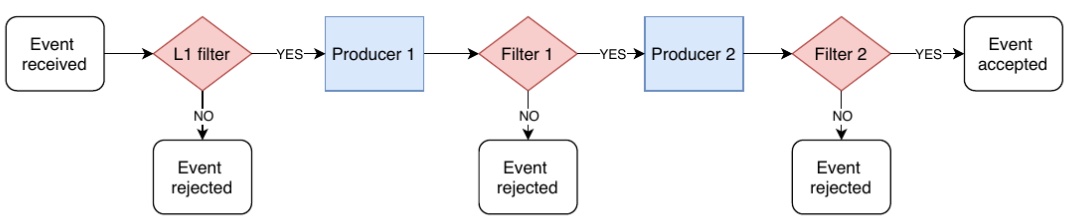
The CMS trigger system

- **Two-level trigger system**
- **Level-1 trigger (L1T)**
  - Hardware level
  - Implemented (mostly) using FPGAs
  - 40 MHz → 100 kHz (Phase-1)
- **High level trigger (HLT)**
  - Software level – computing farms
  - 100 kHz → 1 kHz (Phase-1)
- **HLT path:**
  - A sequence of selections (filters) on physics objects (electrons, photons, jets, etc.) that determines whether an event is accepted or not



	LHC	HL-LHC	
CMS detector	Phase-1	Phase-2	
Peak ⟨PU⟩	60	140	200
L1 accept rate (maximum)	100 kHz	500 kHz	750 kHz
Event Size at HLT input	2.0 MB [a]	6.1 MB	8.4 MB
Event Network throughput	1.6 Tb/s	24 Tb/s	51 Tb/s
Event Network buffer (60 s)	12 TB	182 TB	379 TB
HLT accept rate	1 kHz	5 kHz	7.5 kHz
HLT computing power [b]	0.7 MHS06	17 MHS06	37 MHS06
Event Size at HLT output [c]	1.4 MB	4.3 MB	5.9 MB
Storage throughput [d]	2 GB/s	24 GB/s	51 GB/s
Storage throughput (Heavy-Ion)	12 GB/s	51 GB/s	51 GB/s
Storage capacity needed (1 day [e])	0.2 PB	1.6 PB	3.3 PB

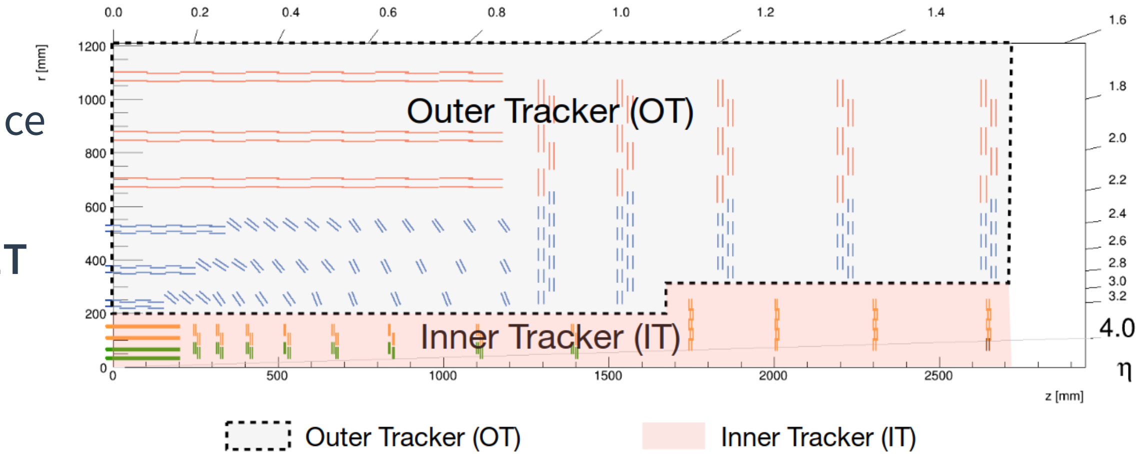
[HLT TDR]

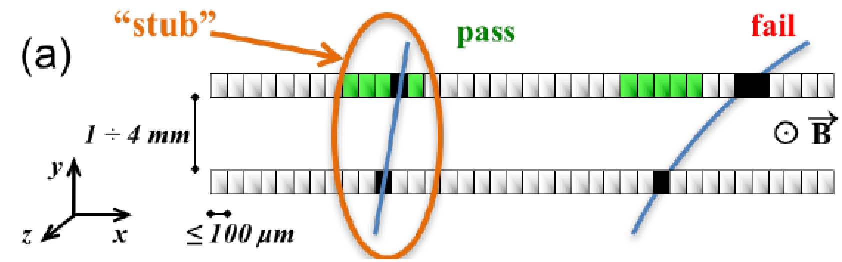


This talk will focus on the algorithmic developments for Phase-2 HLT

Physics object reconstruction:
Tracking

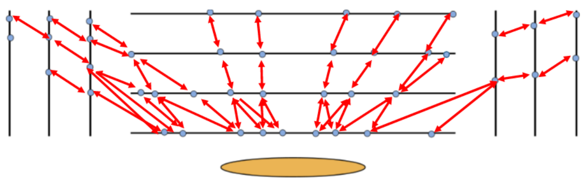
- **New tracker for Phase-2**
  - OT equipped with p_T modules that produce valid “stubs”
- **Tracker information (stubs) to be used for L1T**
  - Unlike Phase-1
- **Run-2 (legacy):**
  - Kalman Filter based: iterative procedure
- **Need improved (faster) algorithms for Phase-2**

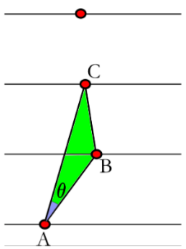


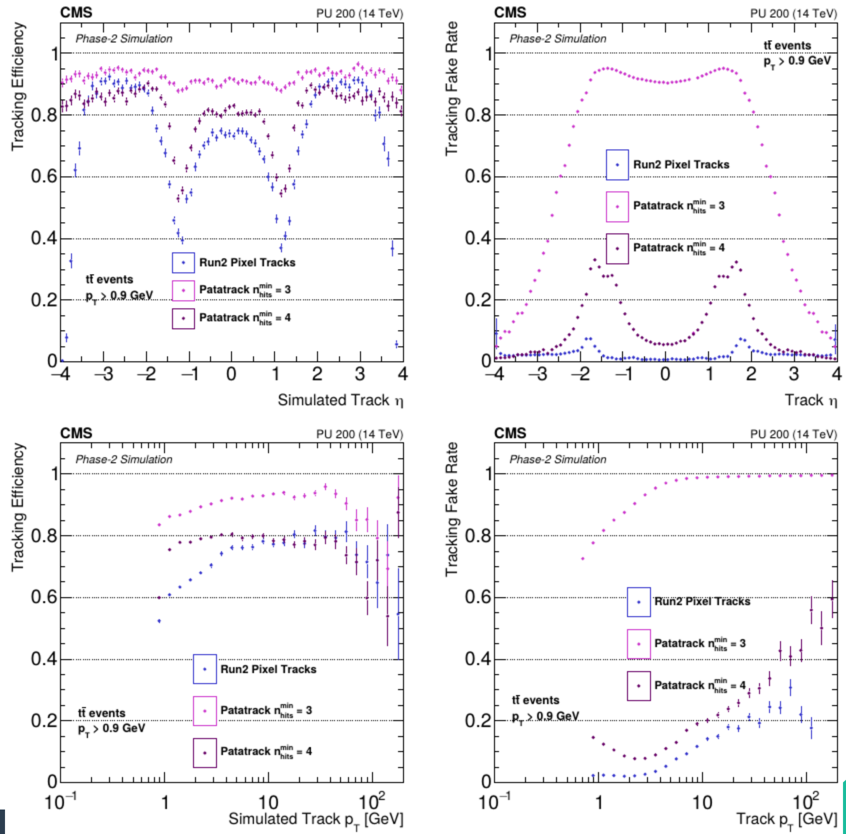


Patatrack pixel tracks

- **New procedure for reconstructing pixel tracks (IT)**
  - Cellular Automaton (CA) based
    - Create graph of all possible connections
    - Form doublets (in parallel)
    - Check compatibility of doublets if they share a hit
  - Highly parallelizable
    - Well suited for GPUs
- **Physics performance comparable to legacy (Run-2) algorithm**

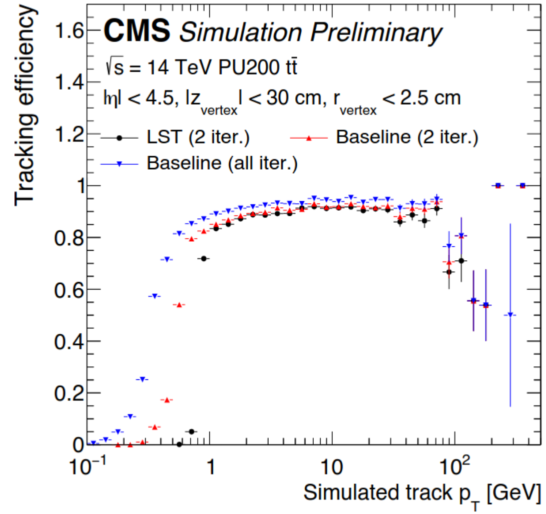


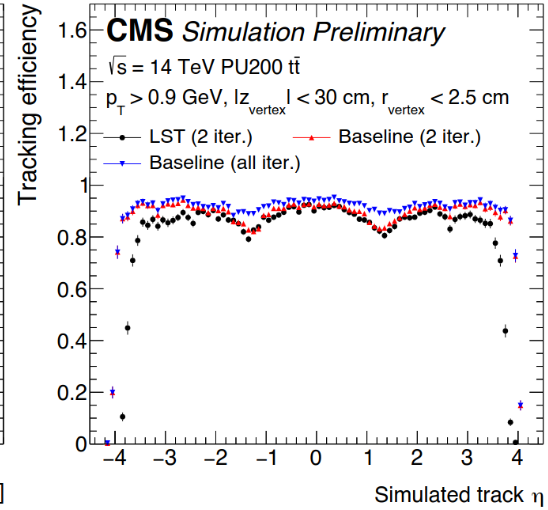




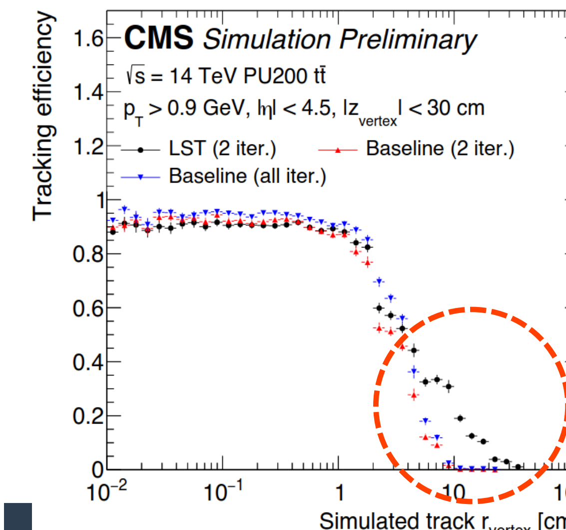
Line segment tracking (LST)

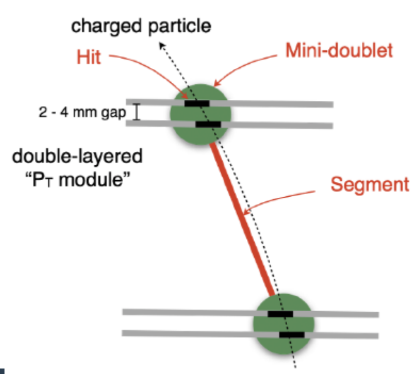
- New algorithm for reconstructing tracks in the OT
  - Form a line segment (LS) from 2 compatible doublets
  - Link line-segments to form triplets, quintuplets, etc.
- Physics performance comparable to Run-2 algorithm (baseline)
  - LST has better efficiency for displaced tracks
    - Beneficial for long-lived particle search programs

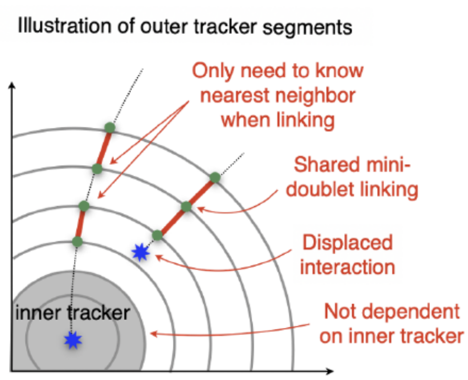


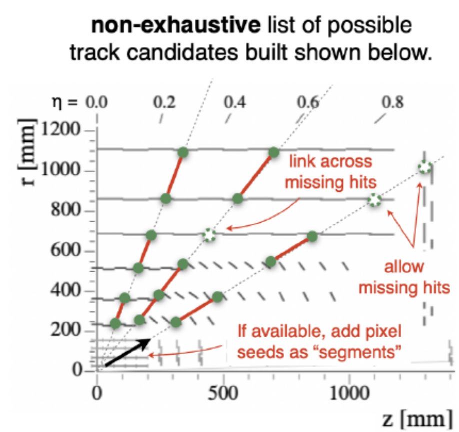


[LST]





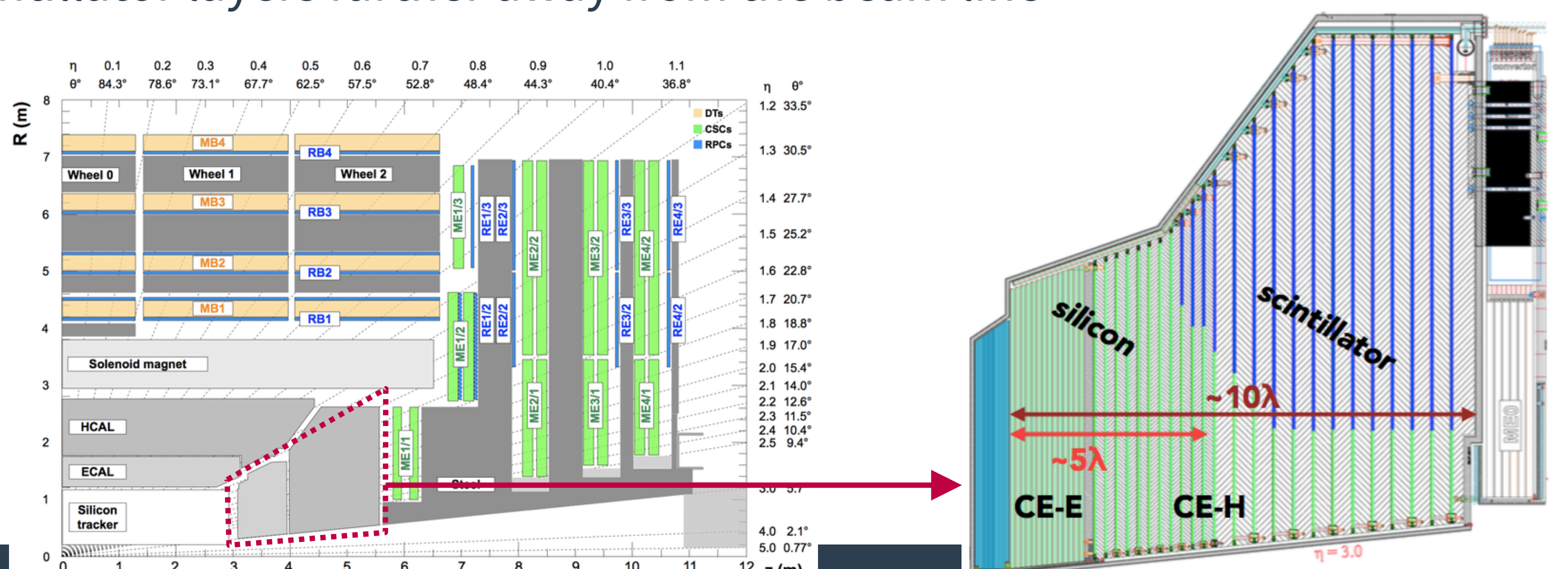




Physics object reconstruction:
Electron and photons

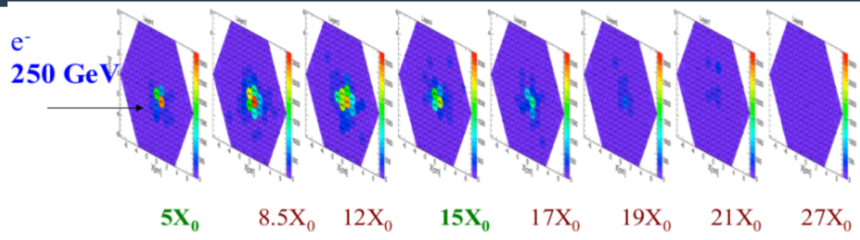
A new endcap calorimeter

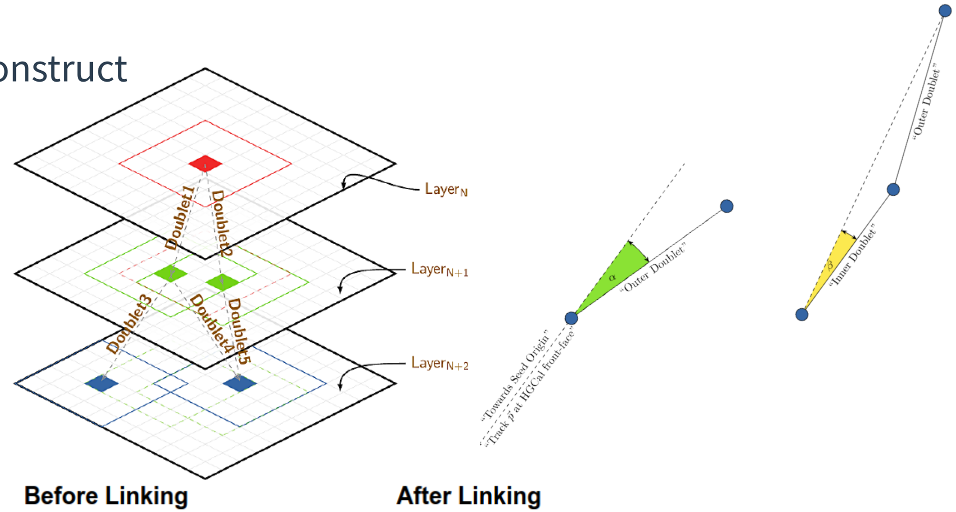
- **The endcap calorimeters (electromagnetic and hadronic) get replaced**
- **New endcap calorimeter for Phase-2**
  - High granularity calorimeter (HGCal)
  - Longitudinal segmentation
  - Silicon layers closer to the beam line (high radiation)
  - Scintillator layers further away from the beam line

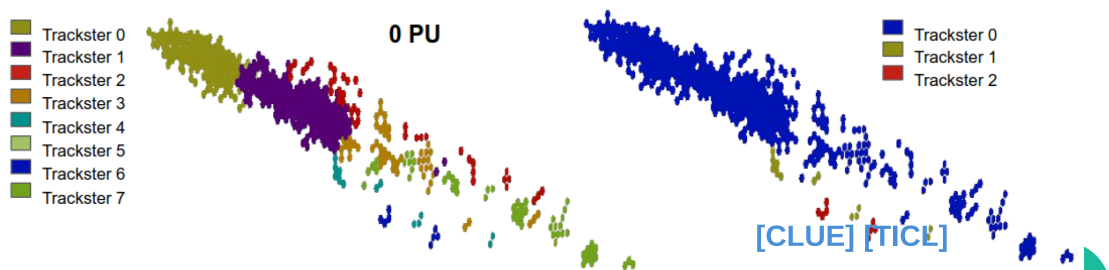


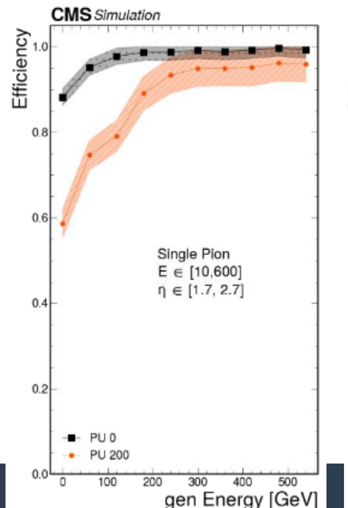
New clustering for a new calorimeter: TICL

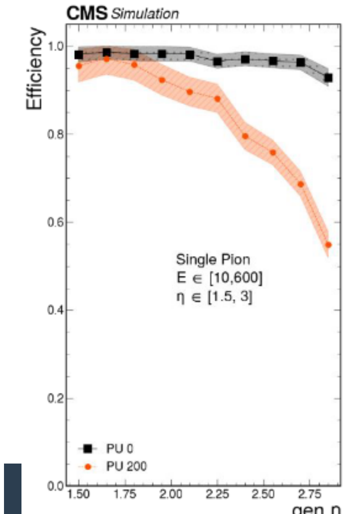
- **Clustering hits in layers**
  - CLUE algorithm (density-based, parallelizable)
  - Form layer clusters (LC)
- **Clustering in 3D**
  - CLUE3D: group LCs to form 3D energy blobs (tracksters)
  - TICL: geometrical linking of compatible tracksters to reconstruct the shower
- **Final shower classified as electromagnetic/hadronic using a deep neural network (DNN) based classifier**
- **Good reconstruction efficiency even at high pileup**



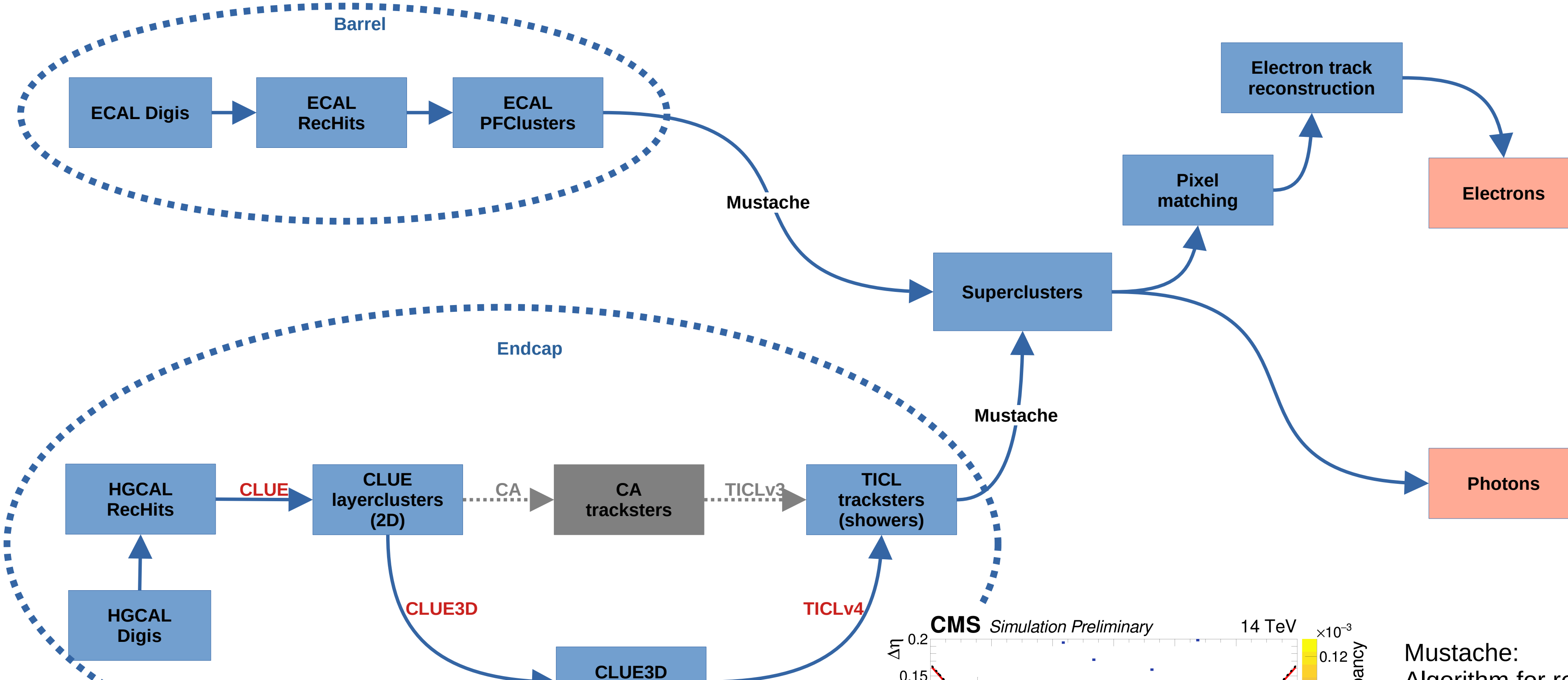


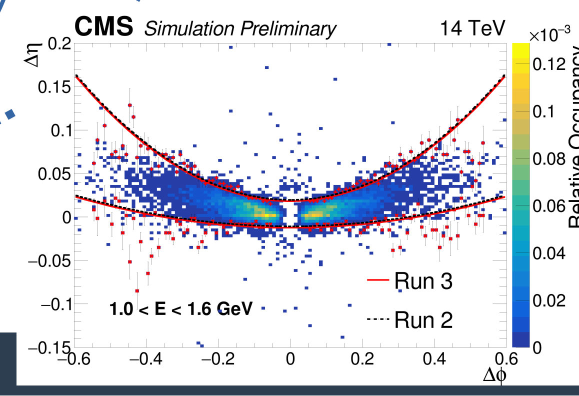






Electron and photon reconstruction at HLT

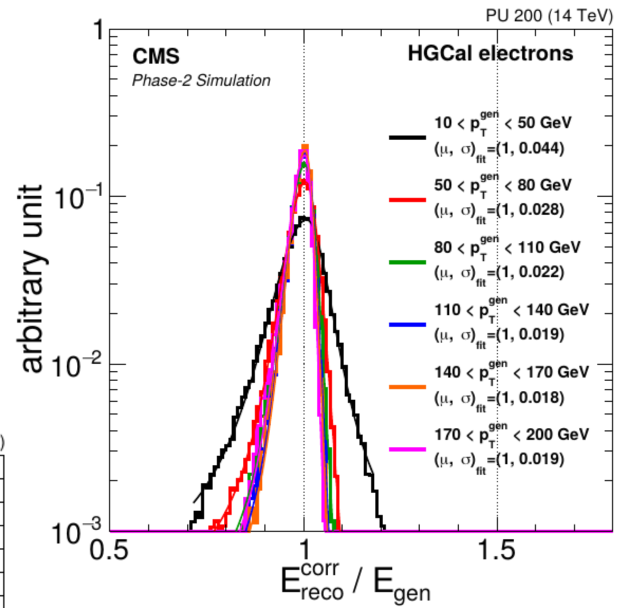


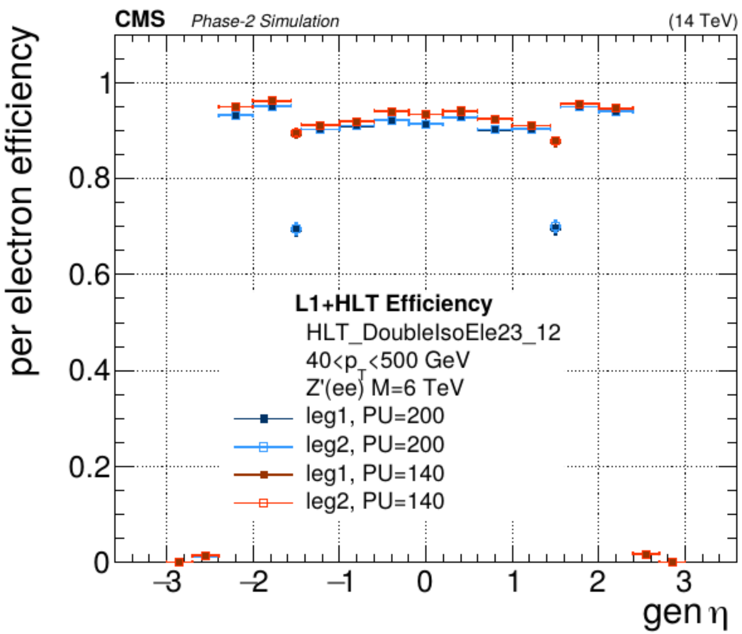


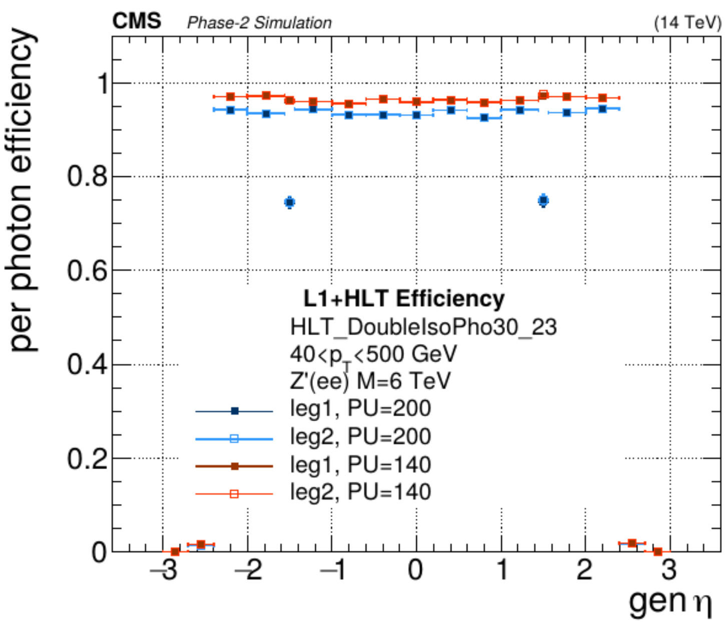
Mustache:
Algorithm for recovering Bremsstrahlung clusters

Electron/photon performance at HLT

- **Dedicated regression to correct the energy**
  - Boosted decision tree (BDT) based – lightweight and fast
  - Obtains good response and resolution over a wide energy range
- **Good performance for electron and photon HLT paths**



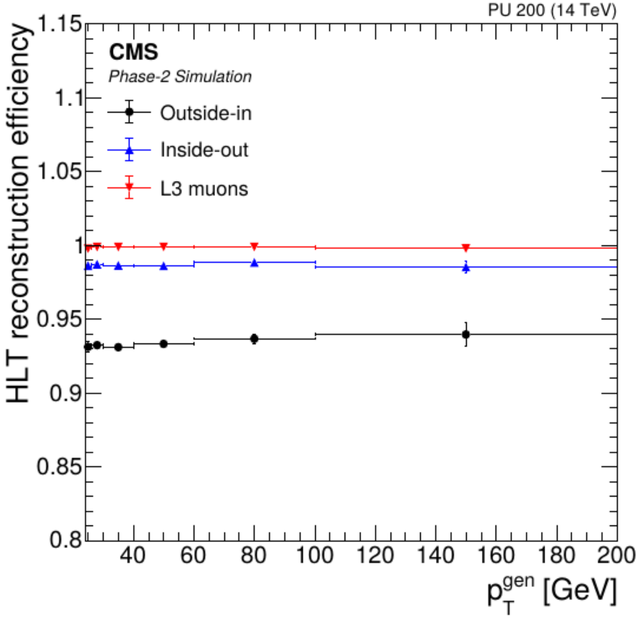




Physics object reconstruction:
Muons

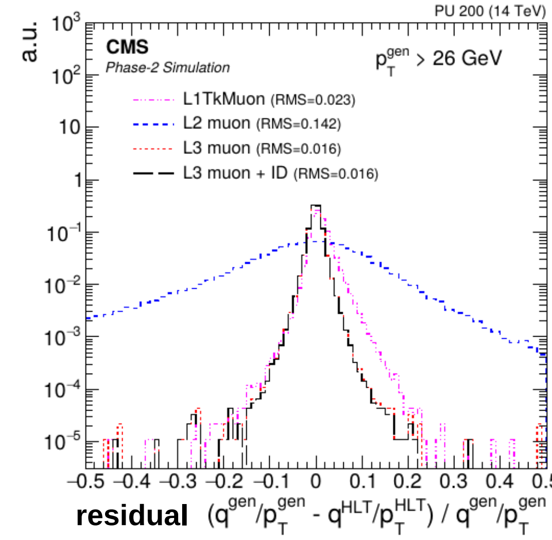
Muon reconstruction at HLT

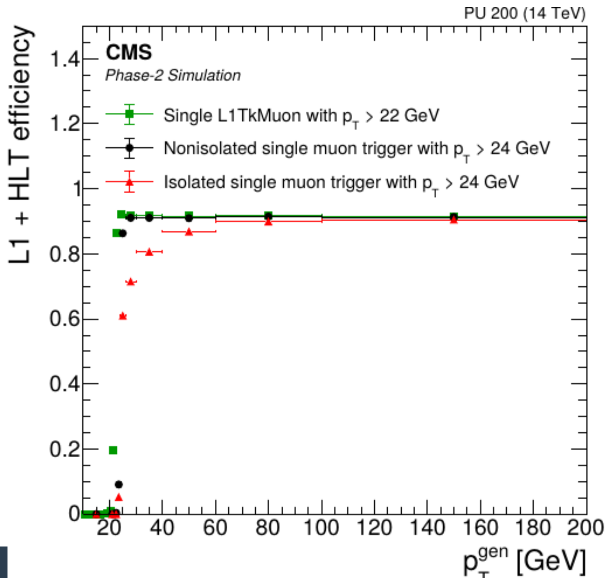
- **Starts from muons reconstructedat the Level-1 trig**
- **Level-2 (L2) reconstruction:**
  - Standalone: only using the muon system
  - Refines the L1 muon trajectory in the muon system
  - Seeds the track reconstruction in the outer layers of OT (for the outside-in tracking)
- **Level-3 (L3) reconstruction**
  - Combines information from tracker as well as muon system
  - Computationally heavy – cannot perform over entire tracker
  - Outside-in: seeded by L2 muons
  - Inside-out: seeded by L1 muon tracks

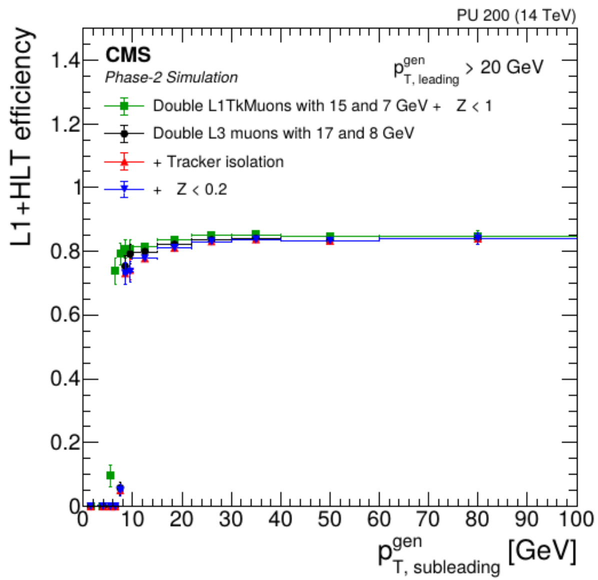


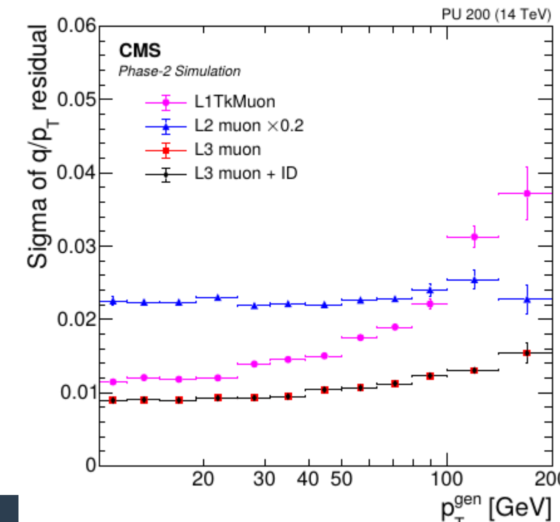
Muon performance at HLT

- **Muon momentum resolution critical for sharp efficiency turn-on**
  - Good resolution (low residual) over a large momentum range
- **Muon HLT paths show a sharp turn-on**



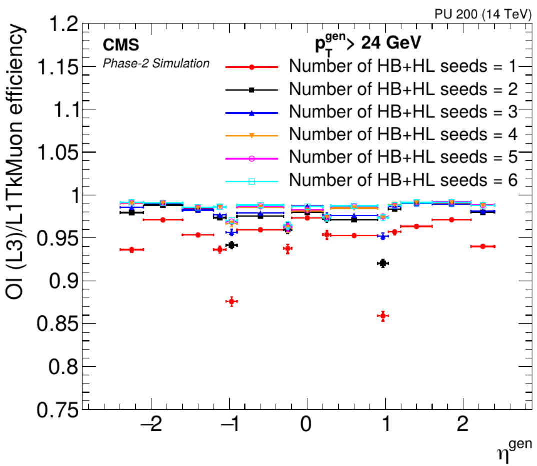


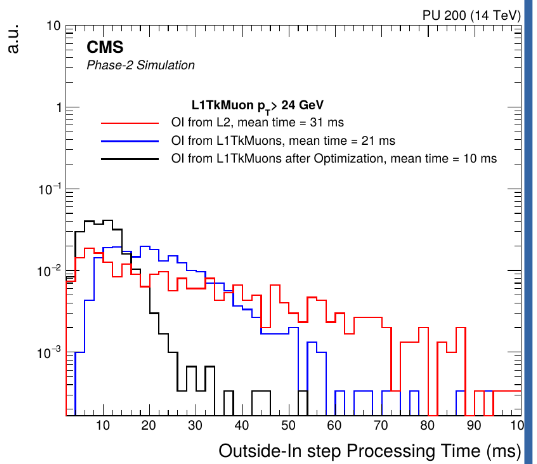




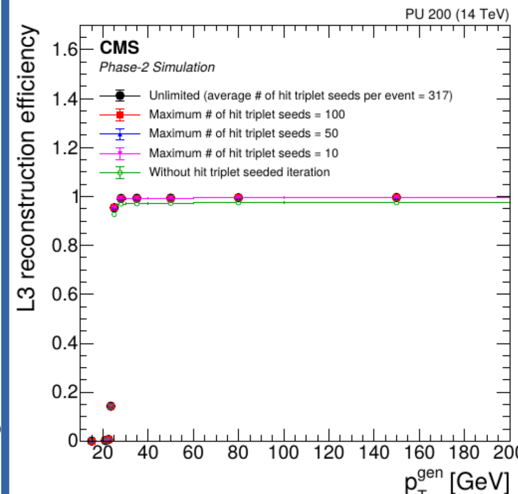
Improvements to muon reconstruction

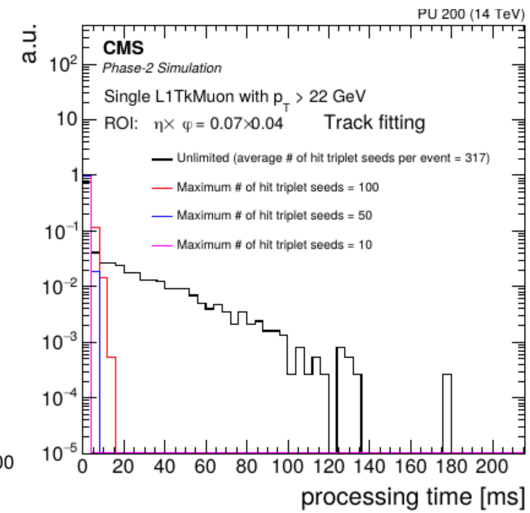
- **Outside-in reconstruction**
  - DNN to optimize seed selection
  - Significantly reduces processing time w/o efficiency loss





- **Inside-out reconstruction**
  - BDT to select tracker seeds
  - Significantly reduces processing time w/o efficiency loss





Physics object reconstruction:
Tau leptons

Tau reconstruction and identification at HLT

- **Hadronic tau decays reconstructed with the hadrons-plus-strips (HPS) algorithm**
  - Hadrons: (mostly) charged pions
  - Strips: photons from neutral pion decays form a strip like pattern in the ECAL
  - Select combinations of hadrons and strips that are compatible with tau decay
- **Two approaches for identification (against quark/gluon jets)**
  - Isolation based
    - Measure energy deposits in a cone around the HPS tau
  - DNN based (DeepTau)
    - Use a convolutional neural network (CNN) ~1M trainable parameters
    - Each cell contains properties of charged/neutral hadrons, electrons, photons and muons

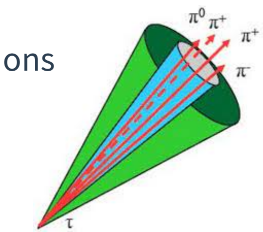
Decay mode	Meson resonance	$\mathcal{B}$ [%]
$\tau^- \to e^- \overline{\nu}_e \nu_\tau$		17.8
$\tau^- \to \mu^- \overline{\nu}_\mu \nu_\tau$		17.4
$\tau^- \to h^- \nu_\tau$		11.5
$\tau^- \to h^- \pi^0 \nu_\tau$	$\rho(770)$	26.0
$\tau^- \to h^- \pi^0 \pi^0 \nu_\tau$	$a_1(1260)$	9.5
$\tau^- \to h^- h^+ h^- \nu_\tau$	$a_1(1260)$	9.8
$\tau^- \to h^- h^+ h^- \pi^0 \nu_\tau$		4.8
Other modes with hadrons		3.2
All modes containing hadrons		64.8

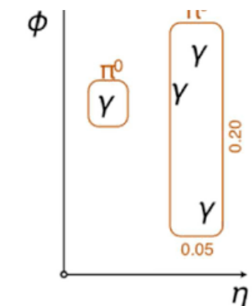
$h^\pm\pi^0$: $0.3 < m_{\tau_h} < 1.72\sqrt{p_T/100\,\text{GeV}}$. The upper limit on the mass window is constrained to be at least 1.72 and at most 4.2 GeV.

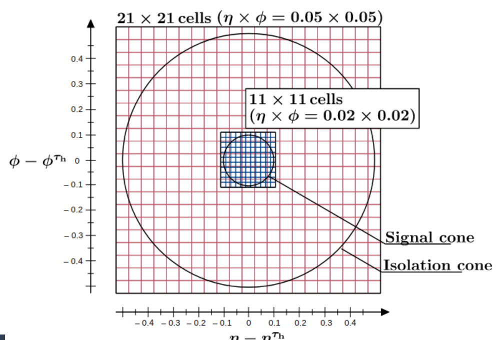
$h^\pm\pi^0\pi^0$: $0.4 < m_{\tau_h} < 1.72\sqrt{p_T/100\,\text{GeV}}$. The upper limit on the mass window is constrained to be at least 1.72 and at most 4.0 GeV.

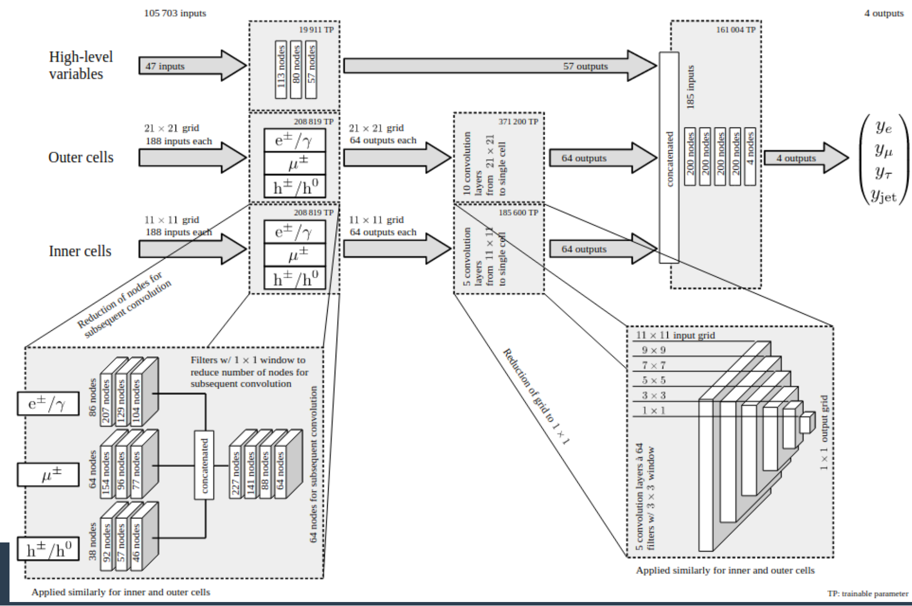
$h^\pm h^\mp h^\pm$: $0.8 < m_{\tau_h} < 1.6\,\text{GeV}$.

$h^\pm h^\mp h^\pm \pi^0$: $0.9 < m_{\tau_h} < 1.6\,\text{GeV}$.



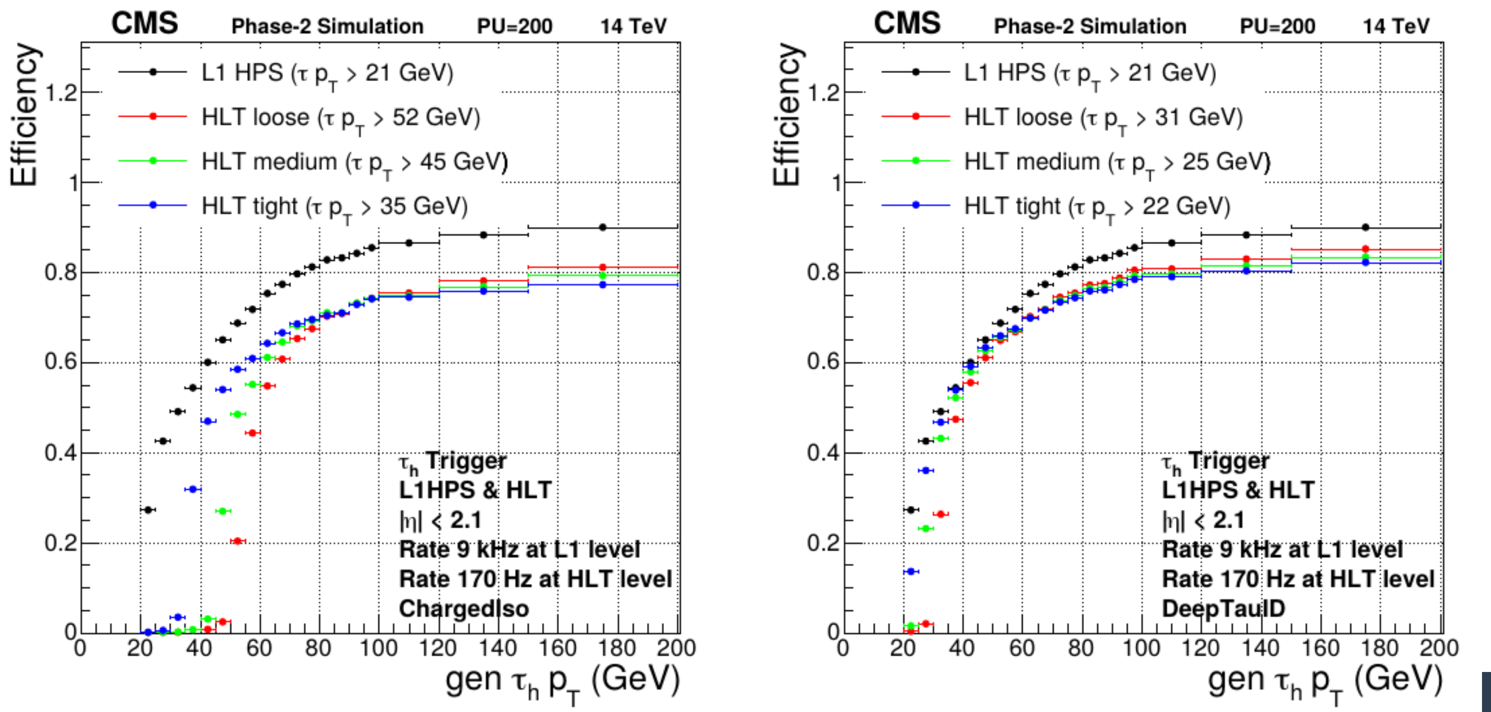






Tau performance at HLT

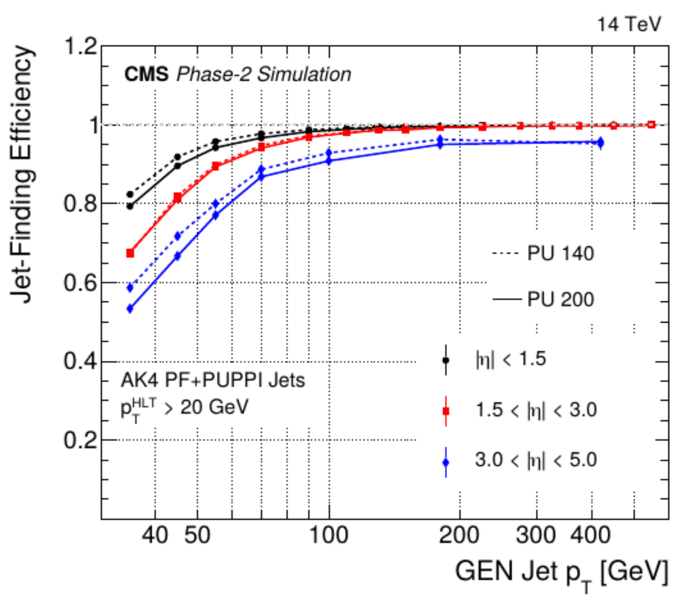
- **Better efficiency turn-on using DeepTau paths**
  - Can achieve lower thresholds for the same rate

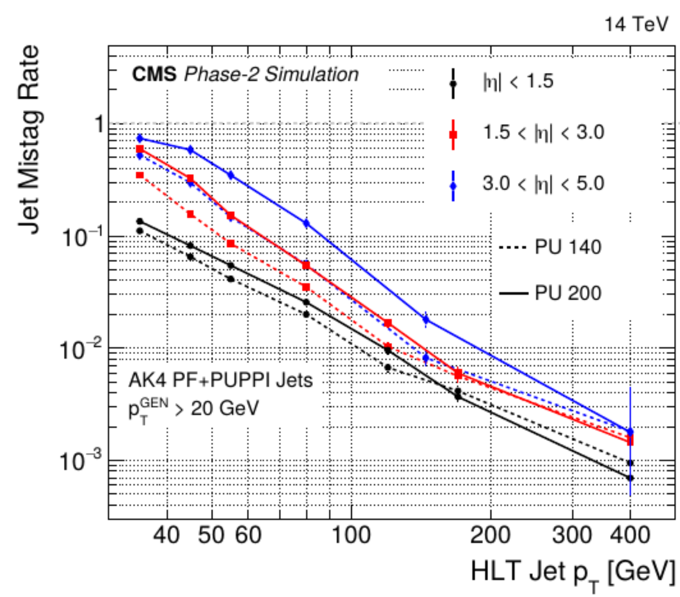


Physics object reconstruction: Hadronic jets

Hadronic jet reconstruction at HLT

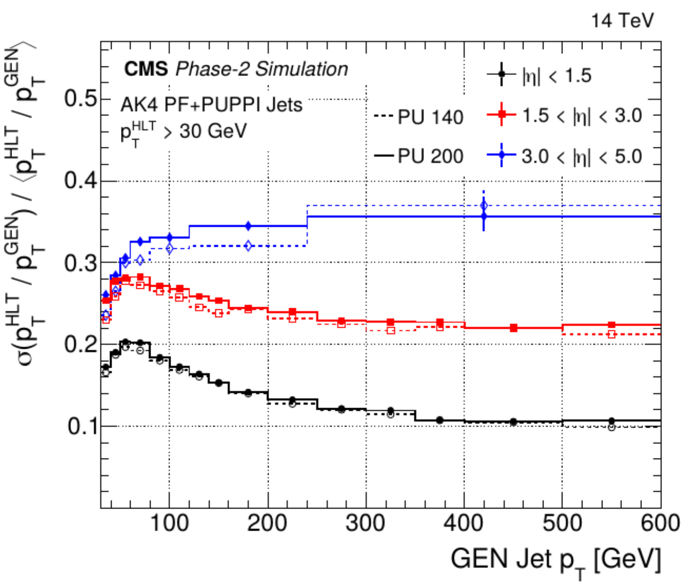
- **Quarks and gluons hadronize to form collimated beams of particles (jets)**
- **CLUE+TICL used for clustering energy deposits in the HGCal (endcap)**
  - Used in the ParticleFlow (PF) algorithm to identify charged/neutral hadrons, electrons and photons
- **PF objects clustered using the anti-k_T algorithm with a radius of 0.4 (AK4)**
- **Overall, good jet finding efficiency and low mistag rate**
  - Better in the barrel – lower pileup

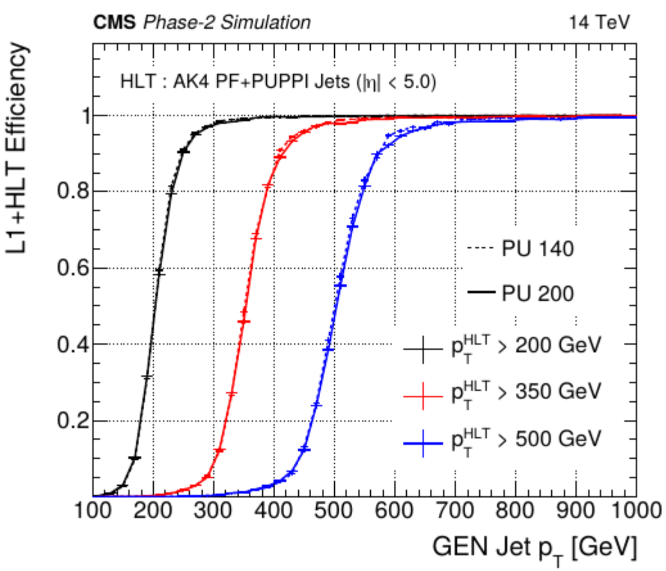


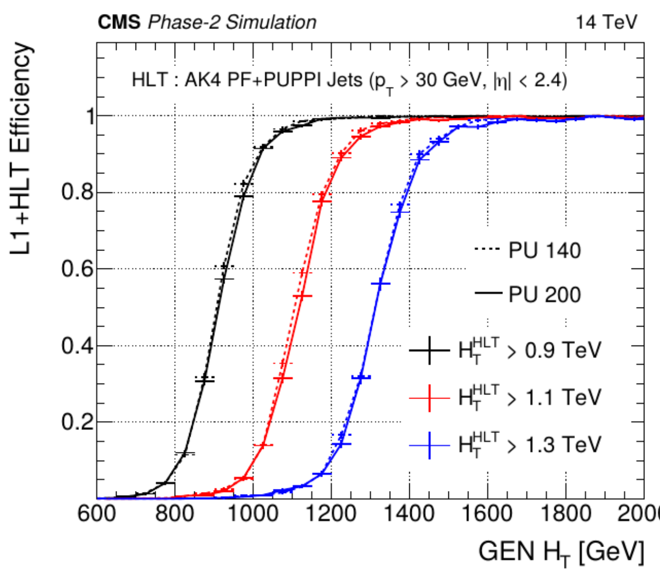


Hadronic jet performance at HLT

- **Good stability versus pileup**
  - Jet energy resolution
  - Jet HLT path efficiencies
    - Single jet
    - Scalar sum of jet transverse momenta (H_T)

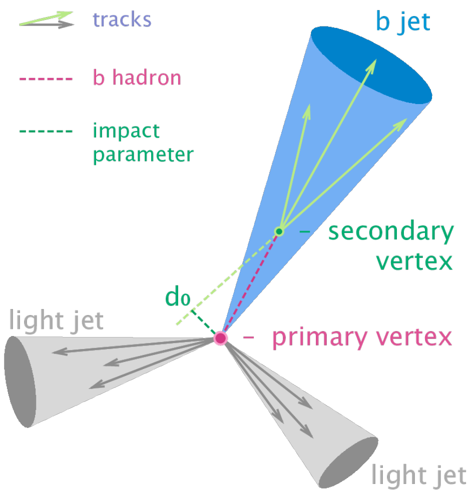


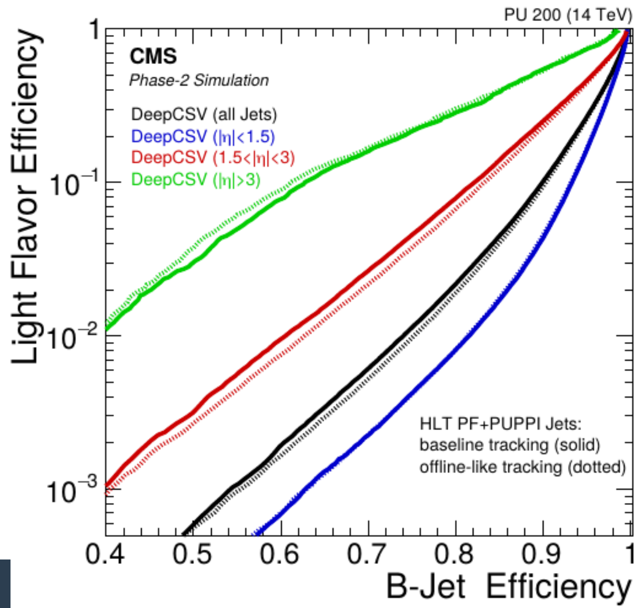




B-jet identification at HLT

- **Jets from bottom quarks differ from light quarks and gluons**
  - B hadron decay produces a secondary vertex
- **AK4 jets identfied as b-jets using a DNN (DeepCSV)**
  - Combines information from jet constituents and secondary vertices
- **Similar approach as Phase-1, but potential gain from precision timing**
  - MIP timing detector
  - HGCal
  - ECAL





The simplified HLT menu

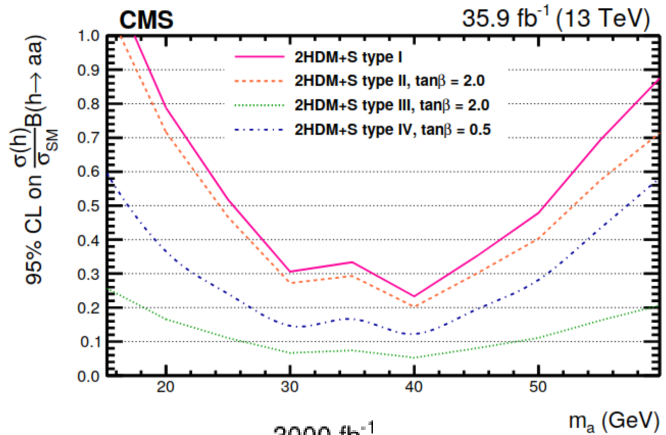
- **Current Phase-2 HLT menu is "simplified"**
  - Set of representative paths
  - Only based on basic physics objects
  - Comprised 50% of total HLT rate in 2018
- **Specialized paths to be added**

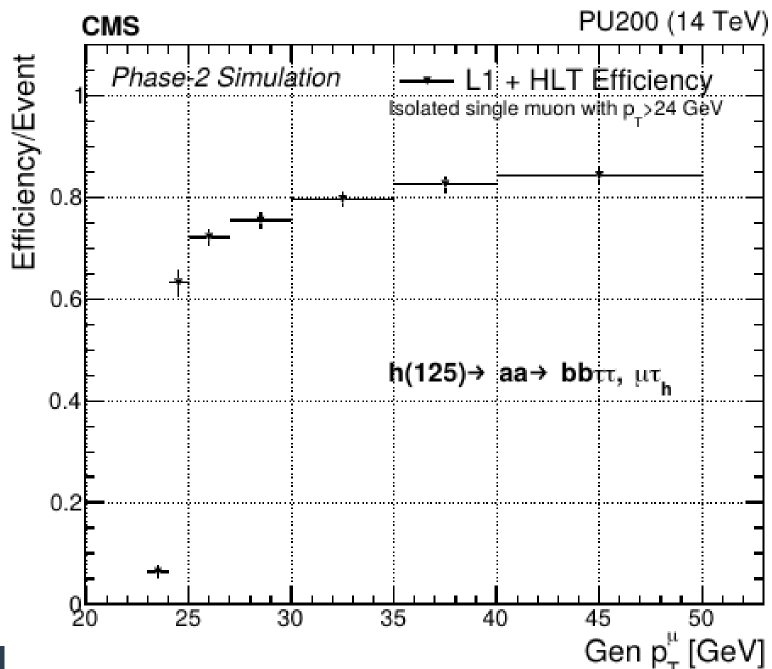
Trigger type	Phase-1 Threshold [GeV]	Phase-1 % rate	Phase-2 L1 seed	Phase-2 Threshold [GeV]	Rate at $\langle PU \rangle = 140$ [Hz]	Rate at $\langle PU \rangle = 200$ [Hz]
Single μ	50	3%	TkMu_22	50	155 ± 6	213 ± 8
Single μ (isol.)	24	14%	TkMu_22	24	943 ± 32	$1\,111 \pm 29$
Double μ	37, 27	1%	TkMu_15_7	37, 27	27 ± 1	40 ± 1
Double μ (isol.)	17, 8	2%	TkMu_15_7	17, 8	113 ± 11	143 ± 13
Triple μ	5, 3, 3	0.5%	TkMu_5_3_3	10, 5, 5	39 ± 8	48 ± 8
Single e (isol.)	28	13%	StaEG_51 OR TkEle_36 OR TkIsoEle_28	32 (WP1) 26 (WP2)	609 ± 27 664 ± 47	$1\,005 \pm 33$ $1\,012 \pm 33$
Double e	25, 25	1%	TkEle_25_12 OR StaEG_37_24	25, 25	46 ± 4	82 ± 6
Double e (isol.)	23, 12	1%	TkEle_25_12 OR StaEG_37_24 OR TkIsoEle_22_StaEG_12	23, 12	52 ± 5	104 ± 9
Single γ	200	1%	StaEG_51	187	32 ± 1	56 ± 6
Single γ (isol.)	110, EB only	1%	StaEG_51 OR TkIsoPho_36	108, EB only	35 ± 9	52 ± 7
Double γ	30, 18	2%	StaEG_37_24 OR TkIsoPho_22_12	30, 23	123 ± 12	179 ± 14
Double τ	35, 35	3%	HPSPFTau_21_21	22, 22	$106 \pm 18^{\dagger}$	159 ± 27
Single jet	500	1%	PuppiJet_230	520	53 ± 1	76 ± 1
H_T	1050	1%	PuppiHT_450	1 070	53 ± 1	74 ± 1
Missing p_T	120	3%	PuppiMET_220	140	79 ± 7	228 ± 20
Multijets with b-tagging	$H_T = 330$ jets = 75, 60, 45, 40	1%	PuppiJet_70_55_40_40_PuppiHT_328	$H_T = 330$ jets = 75, 60, 45, 40	32 ± 4	48 ± 5
Total rate		**49%**			**$2\,525 \pm 57$**	**$3\,621 \pm 62$**

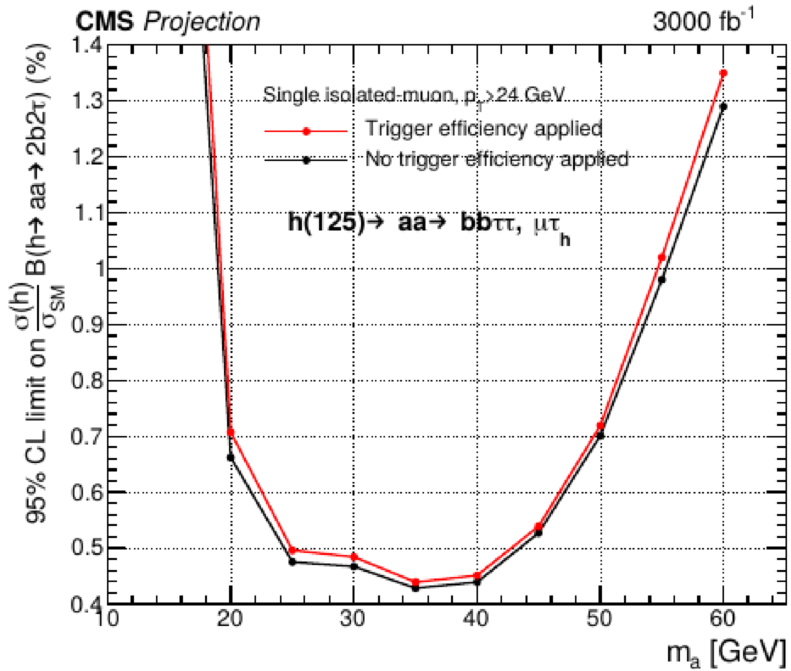
[HLT TDR]

Impact on physics results

- $h(125\ GeV) \rightarrow aa \rightarrow bb\tau\tau$
  - Most sensitive channel for $m_a > 15$ GeV
  - $\mu\tau_h$ channel offers largest sensitivity
- **Check impact of Phase-2 single muon trigger ($p_T > 25$ GeV)**
  - Extrapolated from Run-2 result
  - Sensitivity reaches 0.5% of SM Higgs cross section



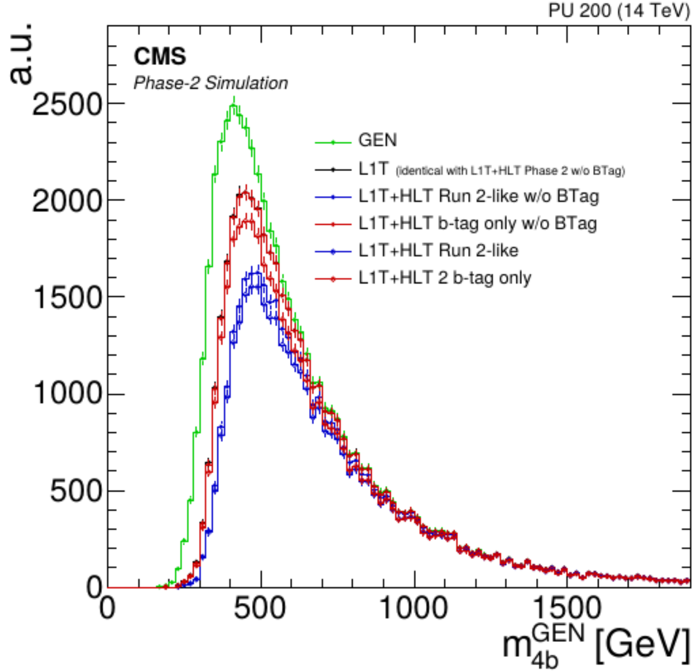


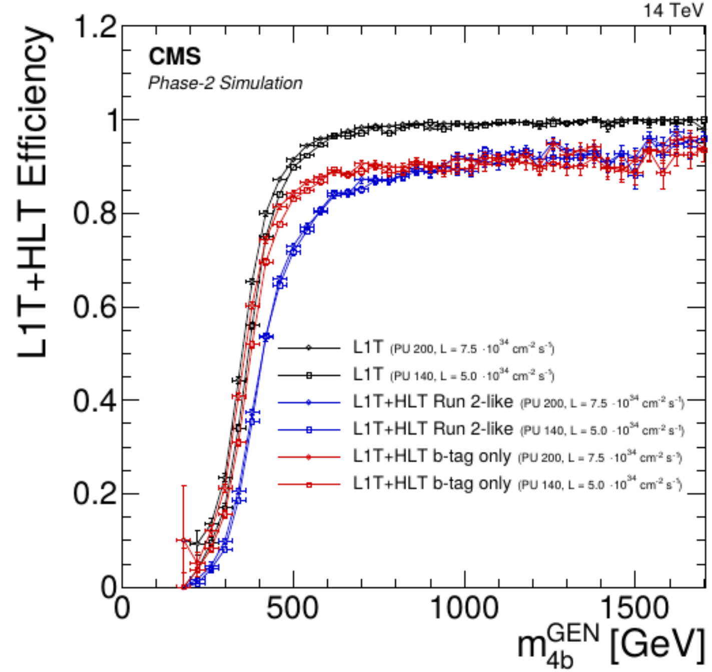


Di-higgs

Trigger	HH → 4b efficiency (inclusive)	HH → 4b efficiency ($m_{4b} < 350$ GeV)	L1 rate	b tagging rate	Output rate
b-tag only	$77.1 \pm 0.2\%$	$31.2 \pm 0.6\%$	10 kHz	10 kHz	49.5 ± 5.6 Hz
Run-2-like	$67.0 \pm 0.2\%$	$15.3 \pm 0.5\%$	10 kHz	3.2 kHz	47.8 ± 5.3 Hz

- **HH → 4b**
  - Critical use case for b-tagged jet triggers at HLT
- **Two options:**
  - Run-2 like: 4 jets with H_T > 330 GeV (higher jet p_T thresholds, >=3 jets b-tagged)
    - B-tagging runs at reduced rate
  - b-tag only: 4 jets with H_T > 200 GeV (lower jet p_T thresholds, >=3 jets b-tagged)
    - b-tagging runs at L1 output rate: more computationally expensive
    - But HH → 4b efficiency higher by 2x for low m_{HH} (<350 GeV)

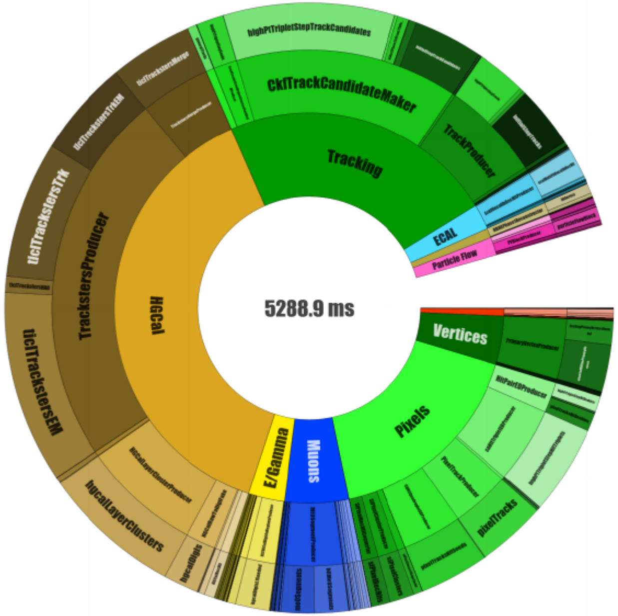




Computing

Processing time

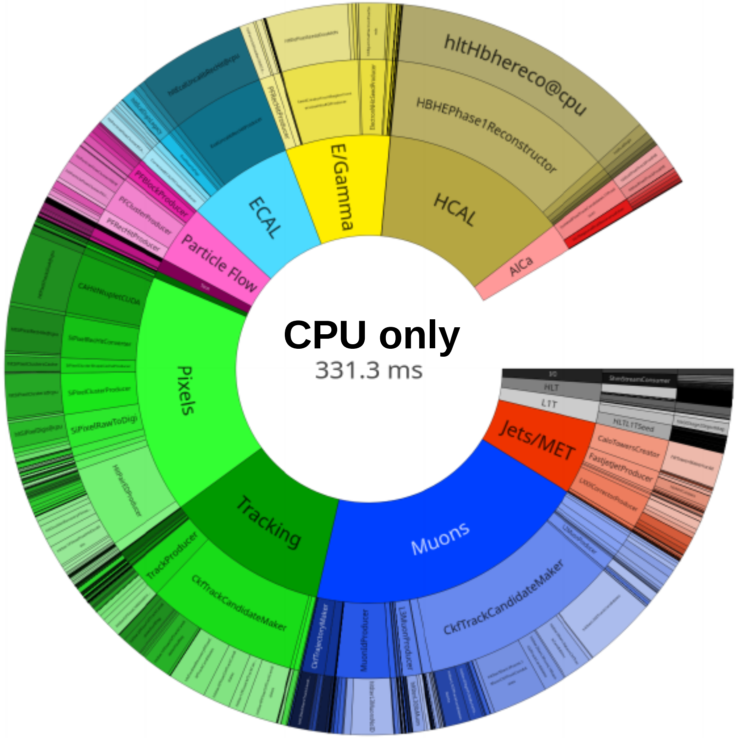
- **Projected speedup**
  - For the CPU-only parts
    - 1.6x by Run-4
    - 2.5x by Run-5
  - Heterogeneous computing (offloading to GPUs)
    - Offload 50% by Run-4
    - Offload 80% by Run-5

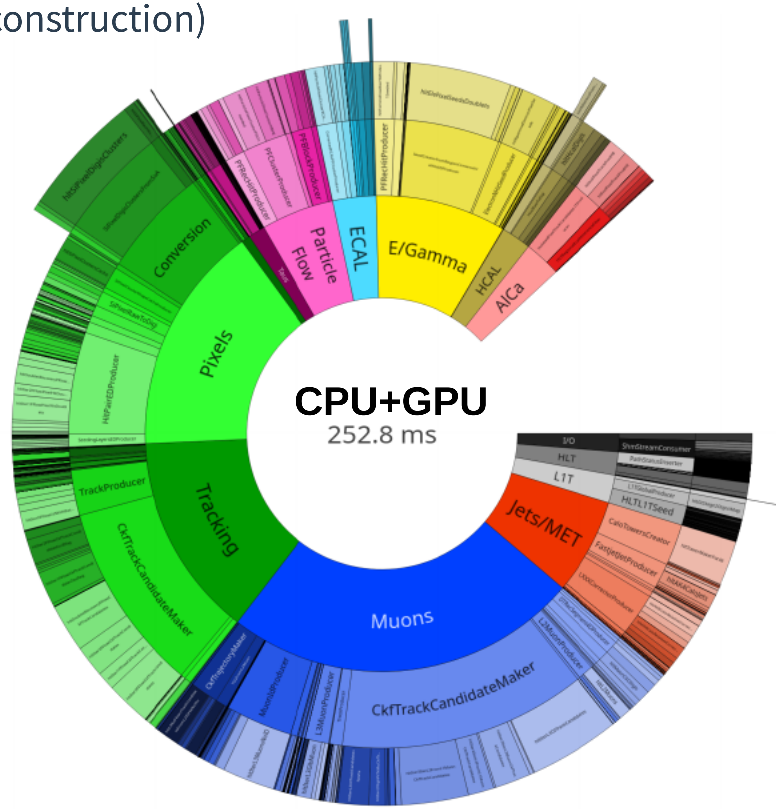


Element	Time	Fraction
B tagging	0.4 ms	0.0 %
E/Gamma	158.4 ms	3.0 %
ECAL	110.9 ms	2.1 %
Framework	0.0 ms	0.0 %
HCAL	41.6 ms	0.8 %
HGCal	2030.5 ms	38.4 %
HLT	0.7 ms	0.0 %
I/O	0.4 ms	0.0 %
Jets/MET	32.1 ms	0.6 %
L1T	2.5 ms	0.0 %
Muons	280.9 ms	5.3 %
other	232.8 ms	4.4 %
Particle Flow	78.9 ms	1.5 %
Pixels	902.3 ms	17.1 %
Tracking	1204.5 ms	22.8 %
Vertices	211.9 ms	4.0 %
total	*5288.9 ms*	*100.0 %*

Offloading to GPUs: strategy

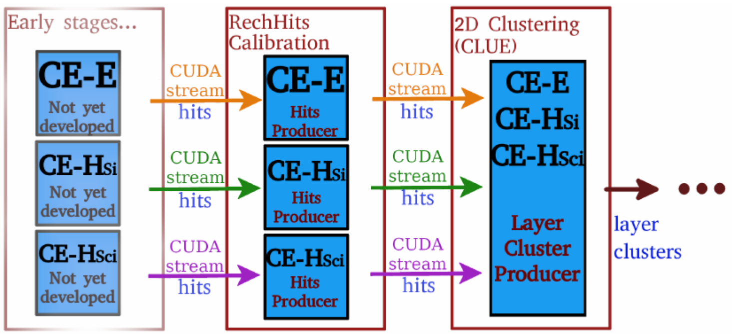
- **Develop new algorithms for Phase-1 (to be used for data-taking in 2024+25 if ready)**
- **Develop new algorithms for Phase-2**
  - Inherit/adapt algorithms from Phase-1 when possible
- **Run-3 developments guide/assist Phase-2 efforts**
  - Possible to offload ~20% already (ECAL, HCAL, and pixel reconstruction)

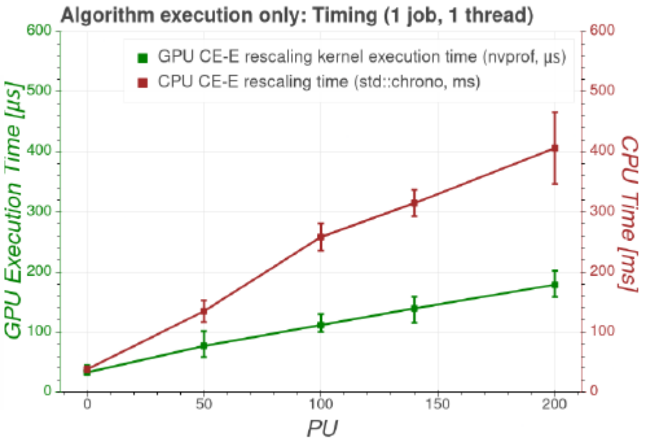


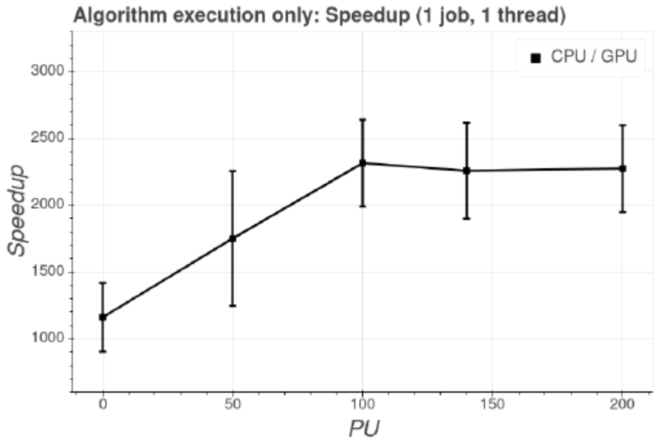


Offloading to GPUs: (some of the) ongoing developments

- Pixels: Patatrack pixel tracks
- Tracking: Line segment tracking
- Porting of ParticleFlow algorithm
- HGCal reconstruction
  - Decoding digitized hits and applying energy corrections
  - Clustering (CLUE)







Summary

Summary

- **Significant progress in developing the high level trigger for the HL-LHC era**
  - Improve existing algorithms
  - New algorithms
  - GPU friendly algorithms
  - Exploit capabilities of new subdetectors
    - Increased granularity
    - Precision timing capabilities
  - HLT processing speedup on schedule for Run-4 and Run-5 targets

Backup

NEW TECHNOLOGIES FOR THE HIGH-LUMINOSITY LHC