LHCb trigger strategy in Run 3 and beyond

Carla Marin

2nd COMCHA workshop

A Coruña, October 2024

Universitat de Barcelona

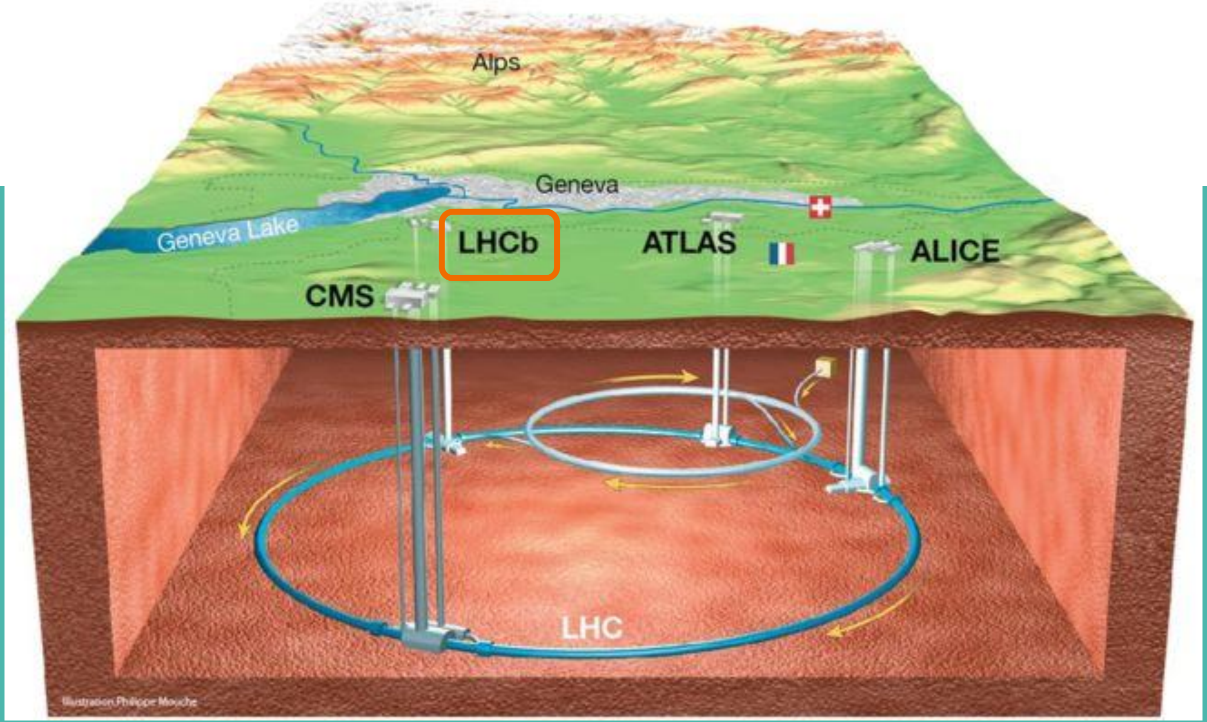
European Research Council

Established by the European Commission

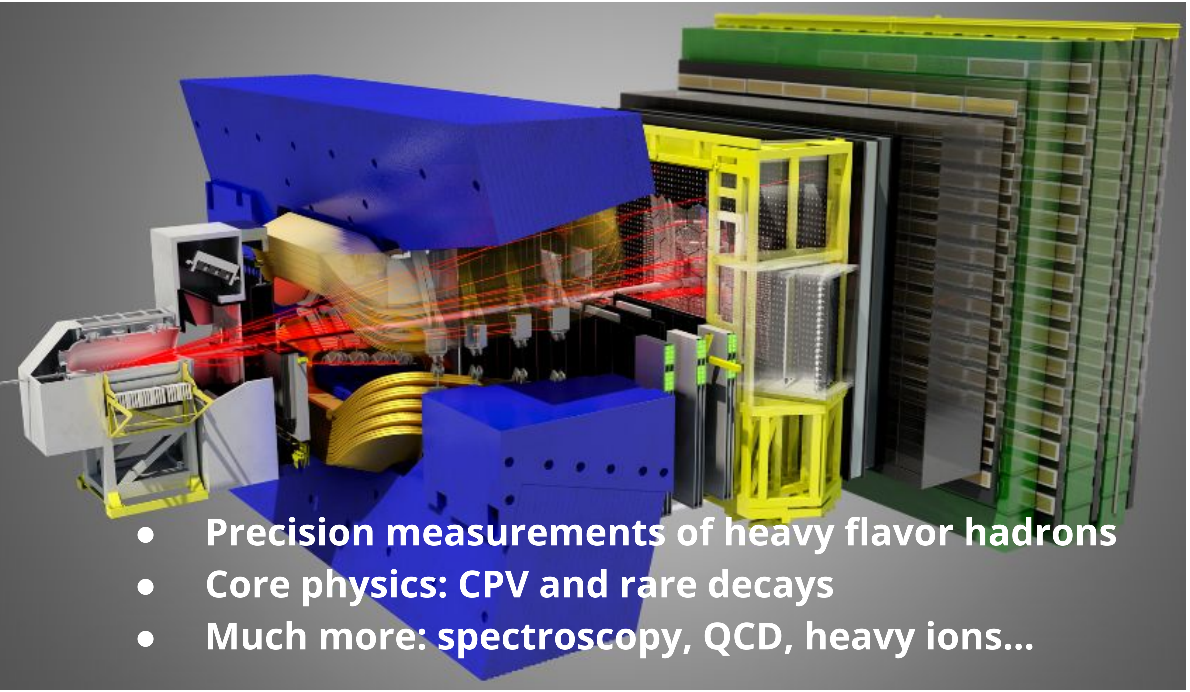
Outline

- The LHCb experiment and the Upgrade I
- The LHCb trigger system in Run 3
- Performance in 2024
- Plans for Run 4 and beyond

The LHCb experiment and the Upgrade I

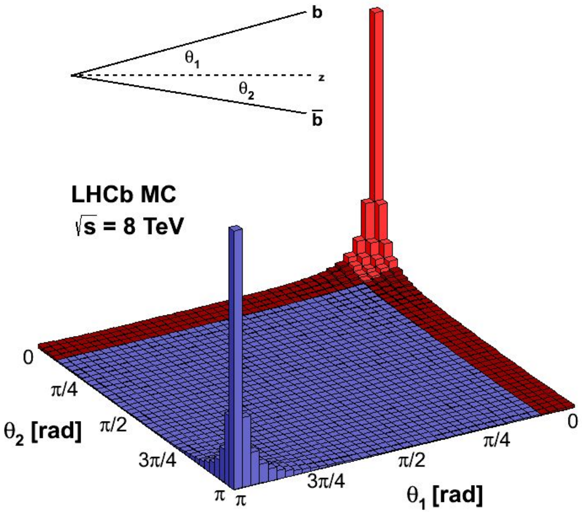


LHCb: Large Hadron Collider Beauty experiment



- Precision measurements of heavy flavor hadrons
- Core physics: CPV and rare decays
- Much more: spectroscopy, QCD, heavy ions...

Distribution of produced b-quarks

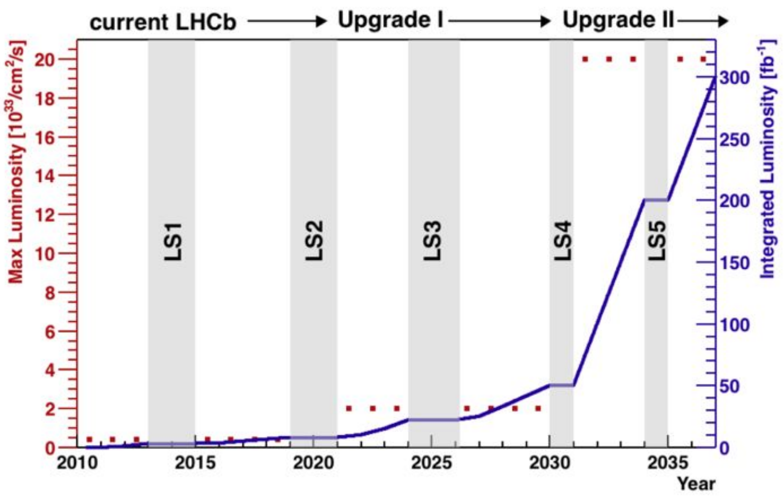


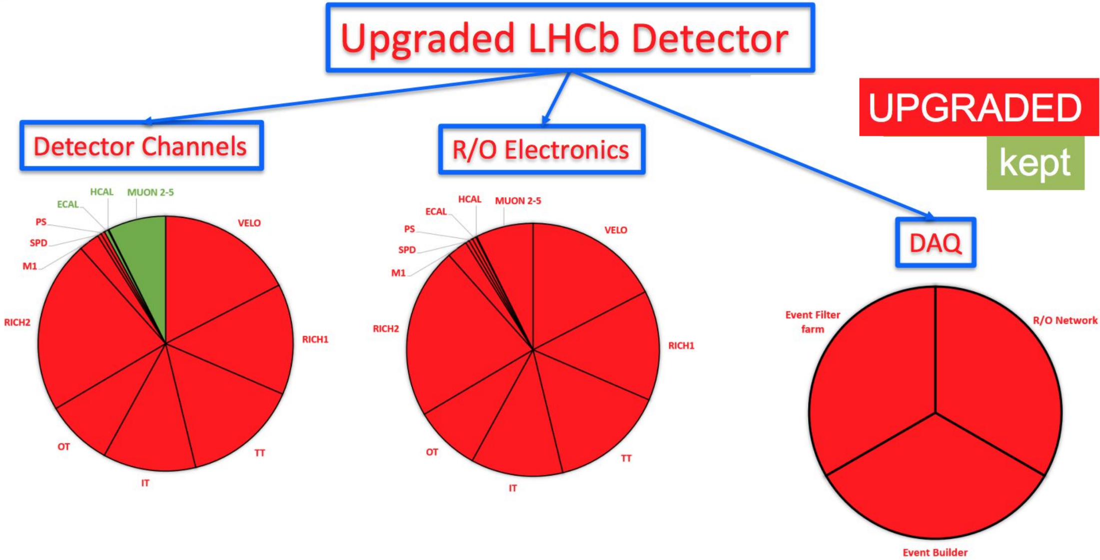
JINST 3 (2008) S08005

JINST 19 (2024) P05065

LHCb Upgrade: a quasi-new detector

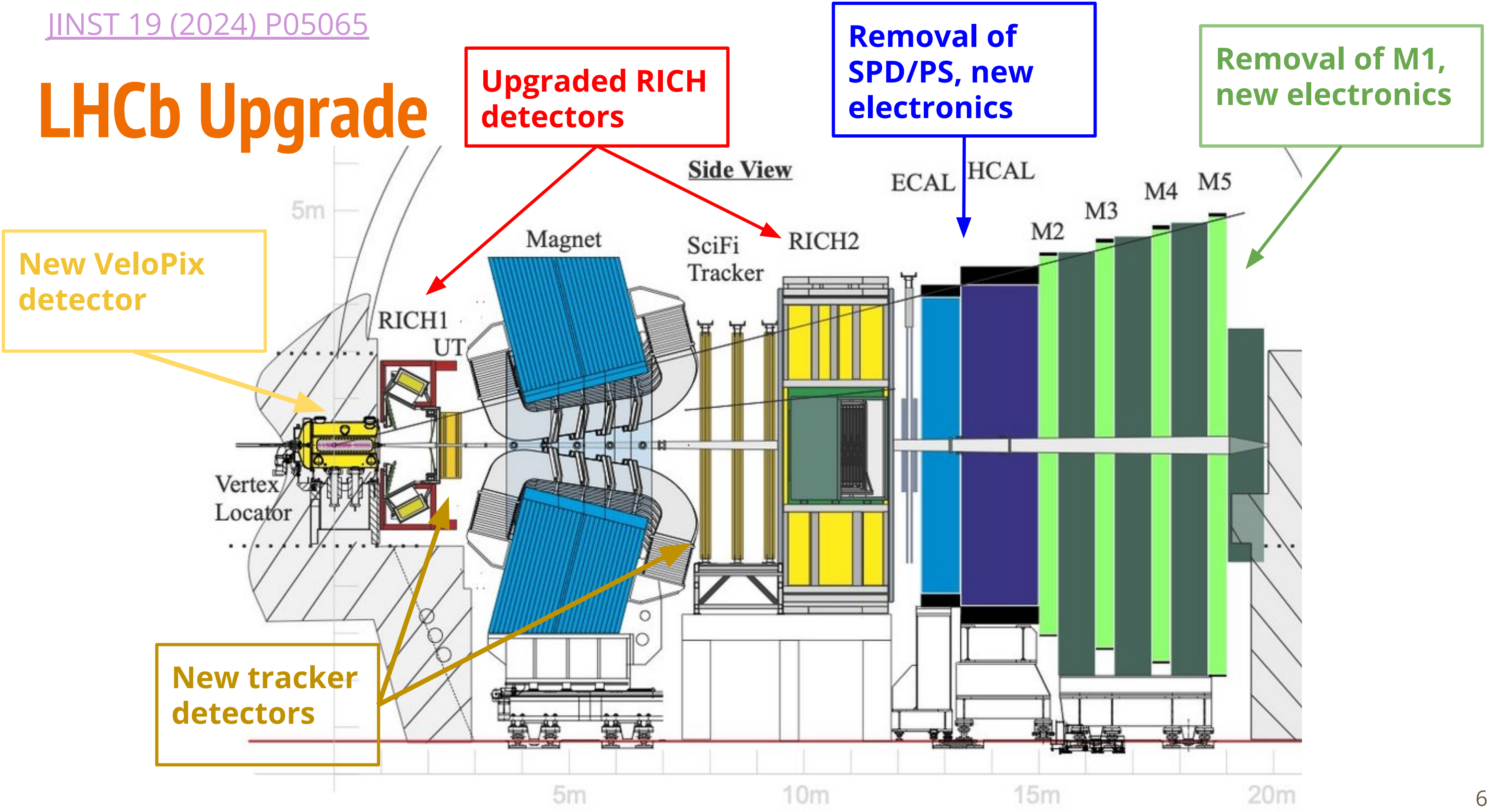
Goal: run at x5 instantaneous luminosity $L_{inst} = 0.4 \rightarrow 2 \cdot 10^{33}$ $cm^{-1}s^{-1}$





JINST 19 (2024) P05065

LHCb Upgrade



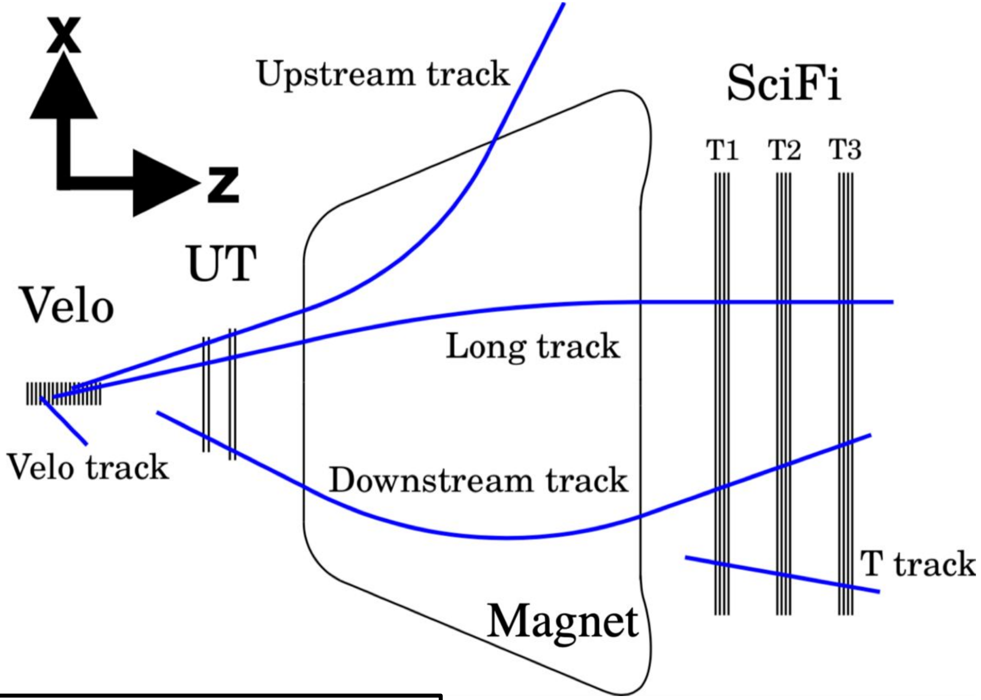
Int. J. Mod. Phys. A 30 (2015) 1530022

Tracking system

Reconstruct trajectories of charged particles

Identify pp and b-decay vertex

Measure particle momentum from bending in magnetic field



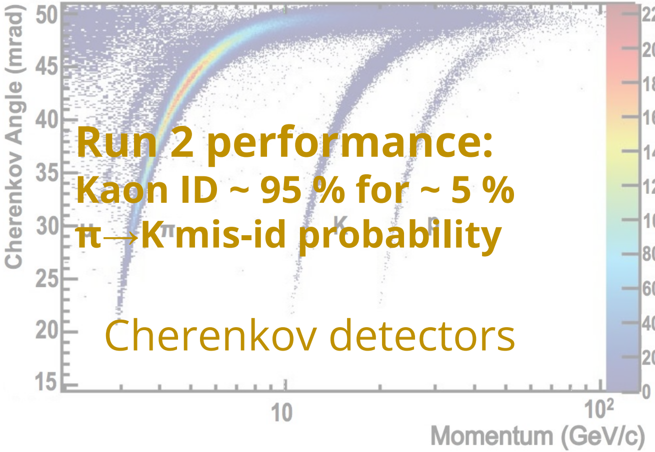
Run 2 performance:

- Long track efficiency > 96%
- Δp / p = 0.5 - 1.0%
- ΔIP = (15 + 29/p_T[GeV]) μm

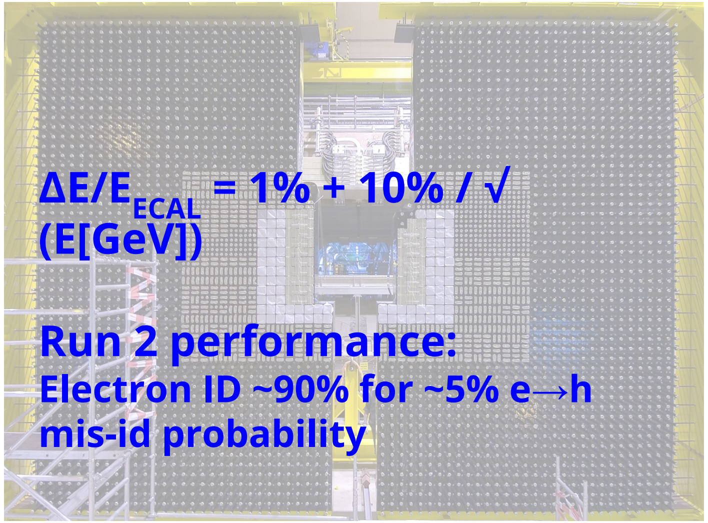
[Int. J. Mod. Phys. A 30 (2015) 1530022]

Particle identification system

- Cherenkov detectors: identify $\pi^{\pm}$, $K^{\pm}$, p
- Calorimeters: identify γ, π^0, $e^{\pm}$
- Muon chambers: identify $\mu^{\pm}$



Cherenkov detectors



Electromagnetic calorimeter

Muon chambers

The LHCb trigger system in Run 3

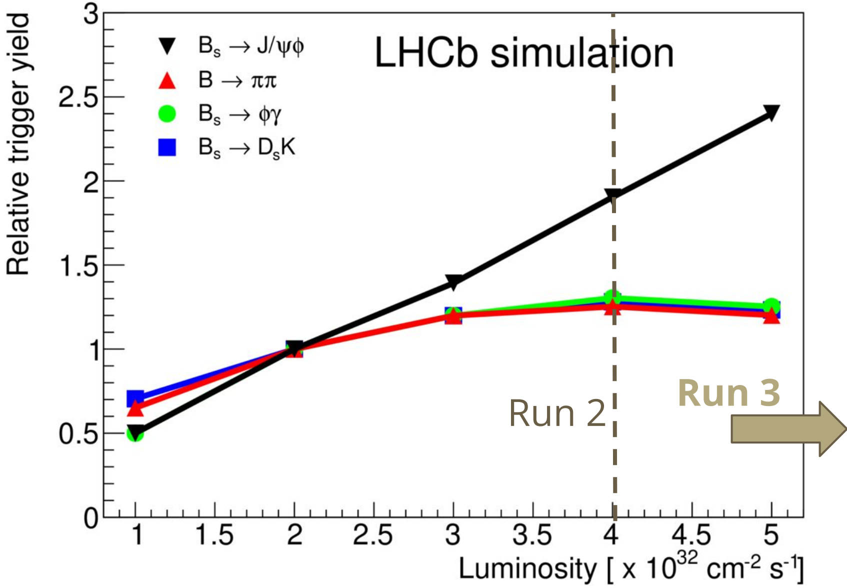
JINST 19 (2024) P05065

The Run 2 trigger limitation

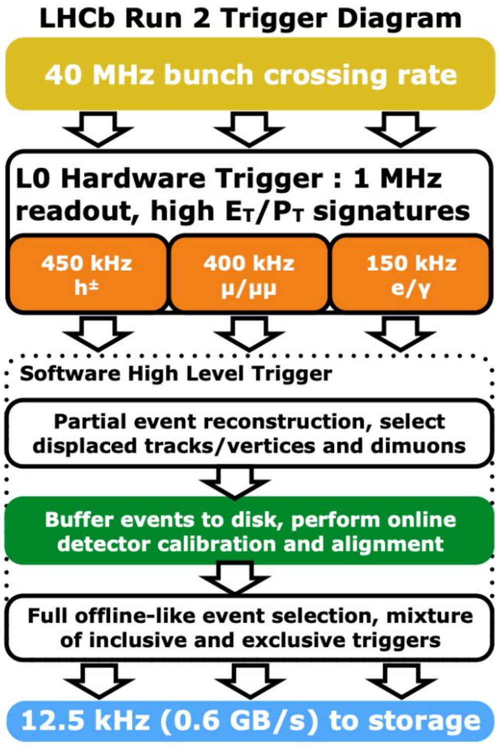
Classical trigger: hardware + software stages

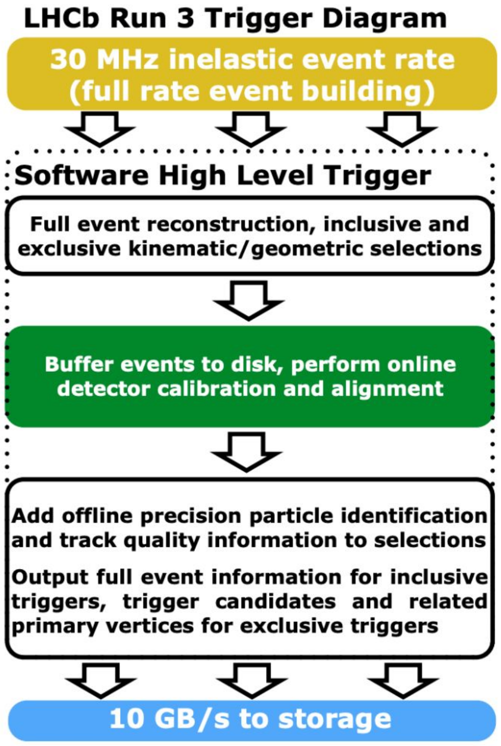
- Limitation: HW trigger rate limit (1 MHz) saturates fully hadronic & $e^{\pm}/\gamma$ modes
  - huge b and c production at LHC energy

- Solution: read full detector at 30 MHz and apply all selections in software
- Constraint: offline storage BW ≤ 10 GB/s



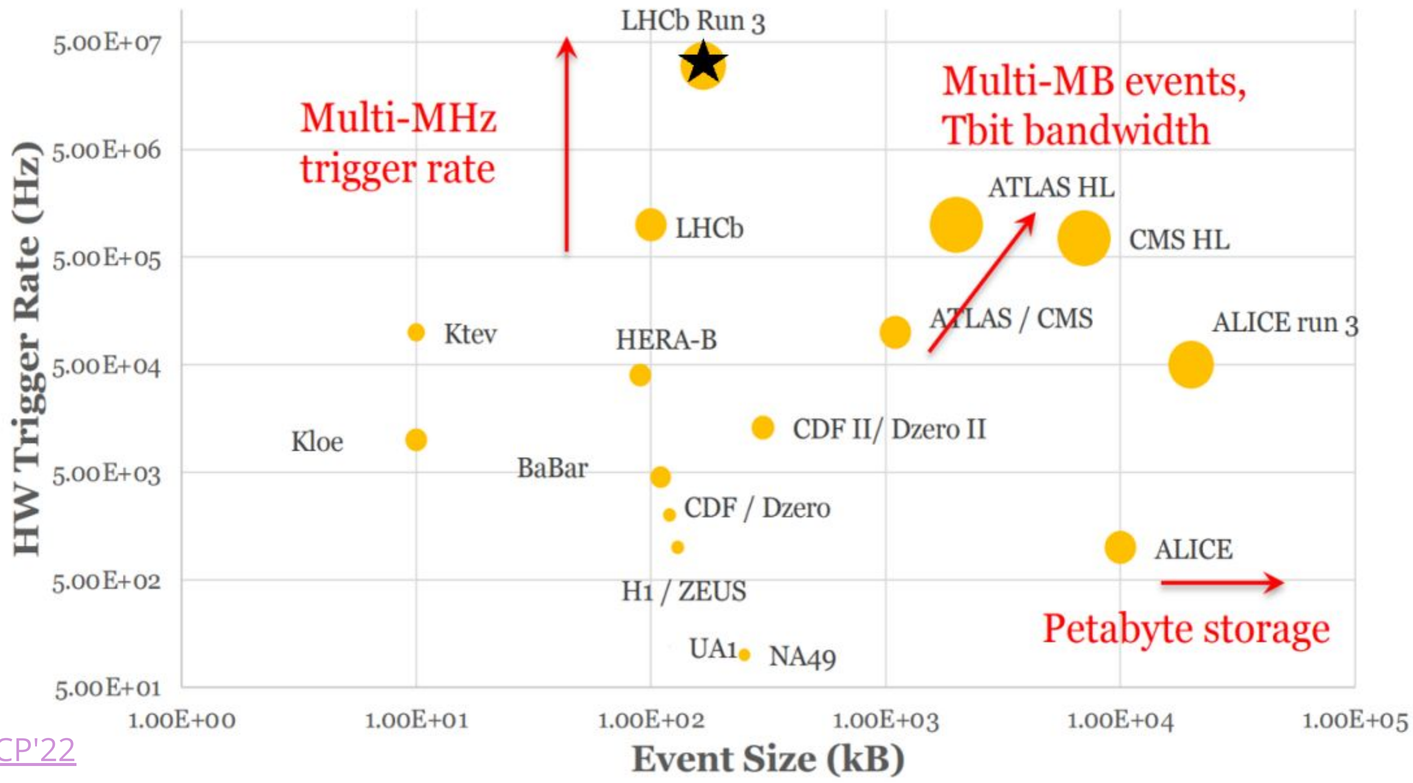
Run 3: a trigger-less readout





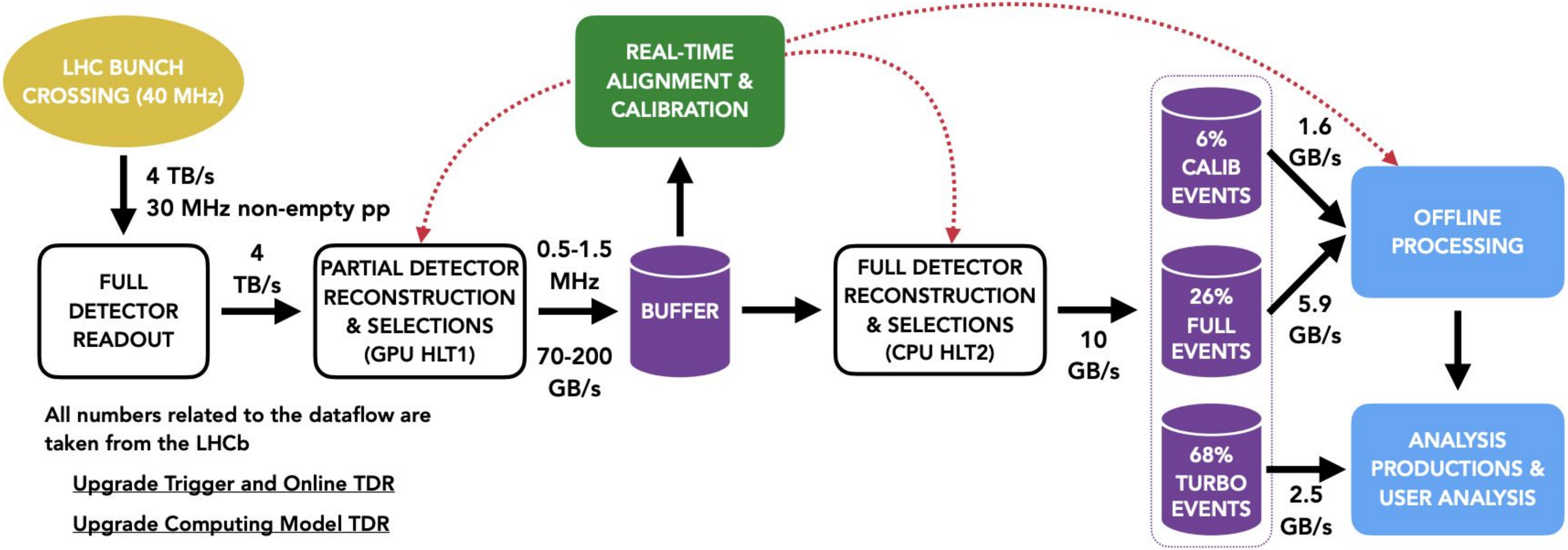
The computing challenge

- Rate: 30 MHz
- Event size: 130 kB
- BW: 4 TB/s



A. Cerri at LHCP'22

LHCb Run 3 trigger overview

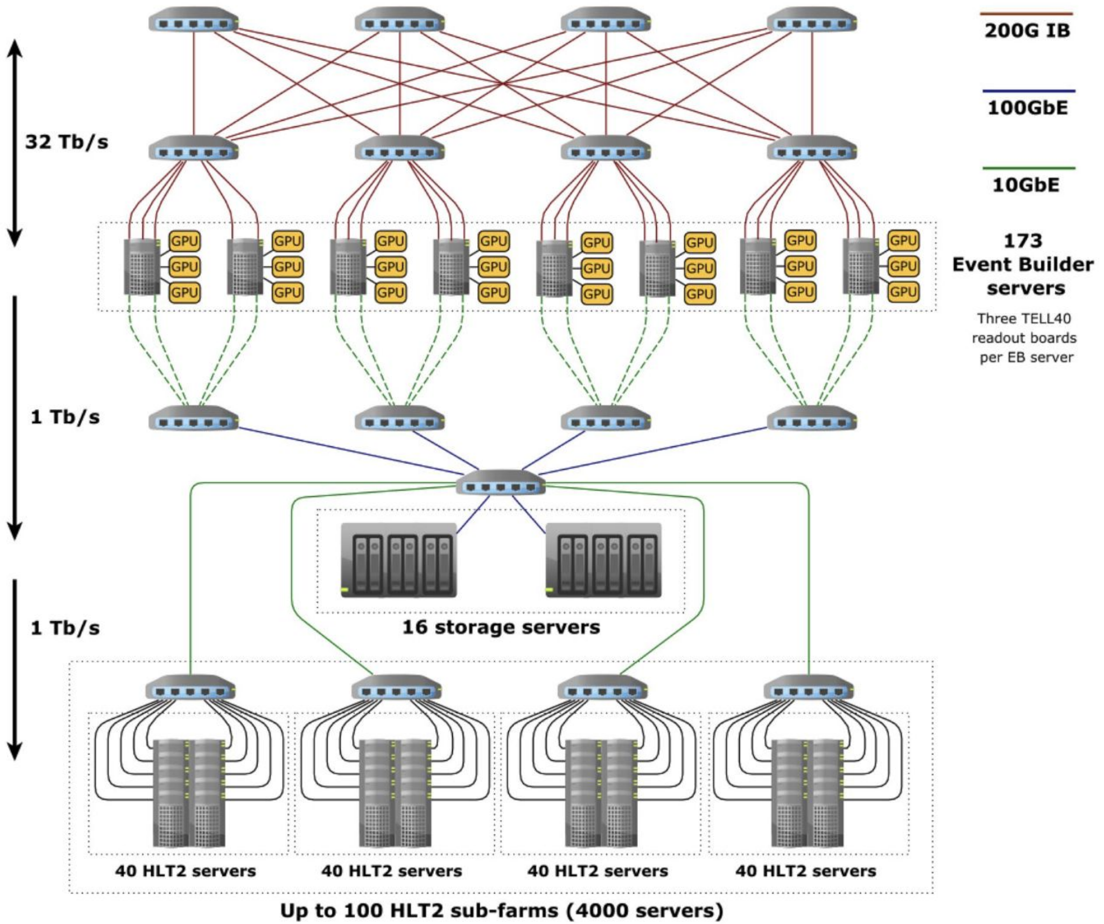


DAQ architecture

Hybrid architecture:

- HLT1: **GPUs** installed in Event Builder servers

- HLT2: **CPUs** in Event Filter Farm



HLT1

Runs at 30 MHz

Reduces rate x30

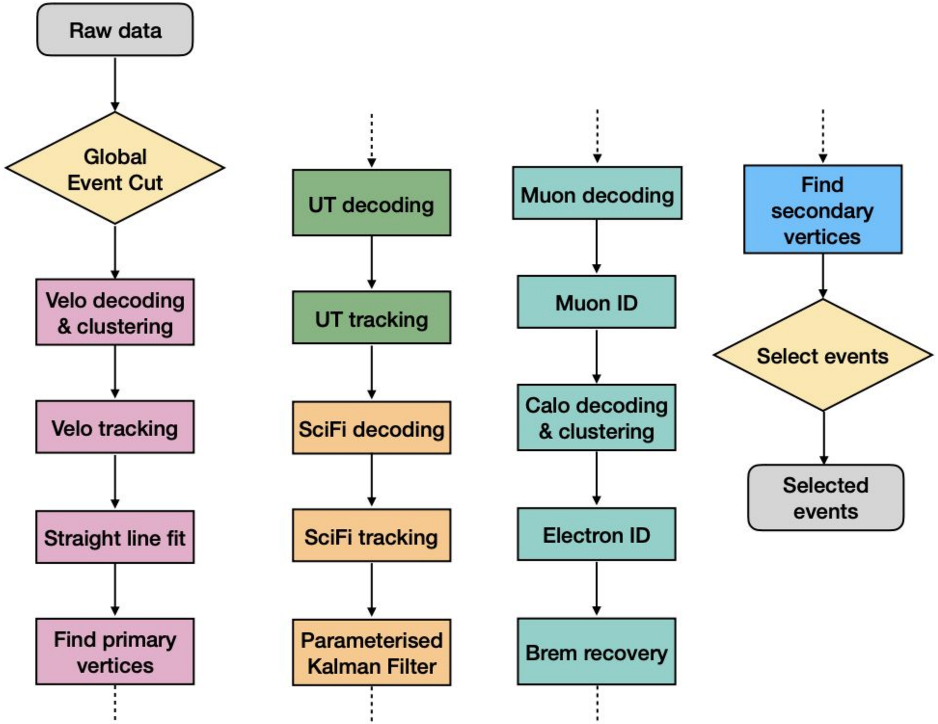
Fast reco largely based on tracking:

- VELO: tracking, vertex reconstruction
- UT: tracking, p estimate, fake rejection
- SciFi: track reconstruction, momentum measurement
- Muon: fast muon PID

Selections: inclusive and exclusive

Extra since TDR:

- ECAL: reco, $e^{\pm}$ ID and brem recovery
- Donwstream tracking → [J. Zhuo talk]



HLT1

Runs at 30 MHz

Reduces rate x30

Fast reco largely based on tracking:

- VELO: tracking, vertex reconstruction
- UT: tracking, p estimate, fake rejection
- SciFi: track reconstruction, momentum measurement
- Muon: fast muon PID

Selections: inclusive and exclusive

Extra since TDR:

- ECAL: reco, $e^{\pm}$ ID and brem recovery
- Donwstream tracking → J. Zhuo talk

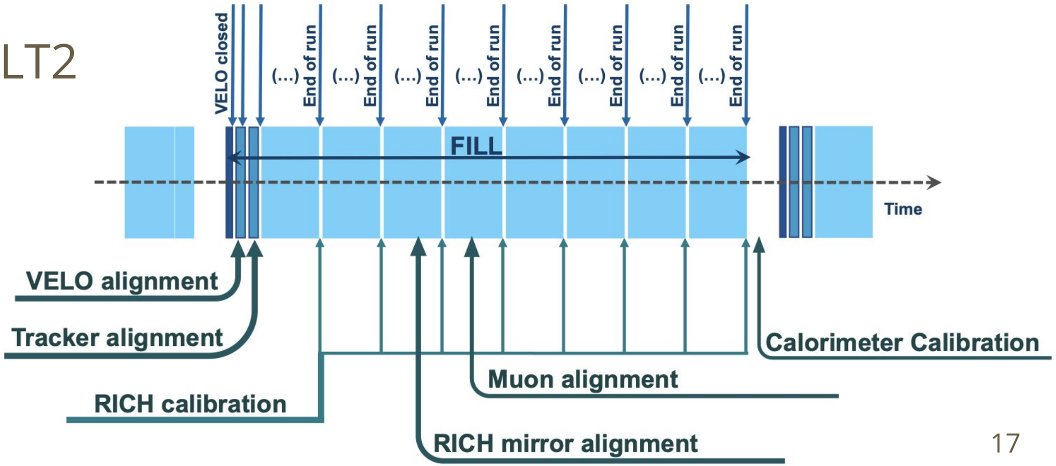
Highly parallel tasks → exploit GPUs: Nvidia RTX A5000

Alignment & calibration

Use HLT1 selected data to derive alignment and calibration constants

- dedicated task for each detector, running on HLT2 farm
- jobs run automatically when enough events collected
- results stored in conditions database

⇒ enables offline-level quality in HLT2



HLT2

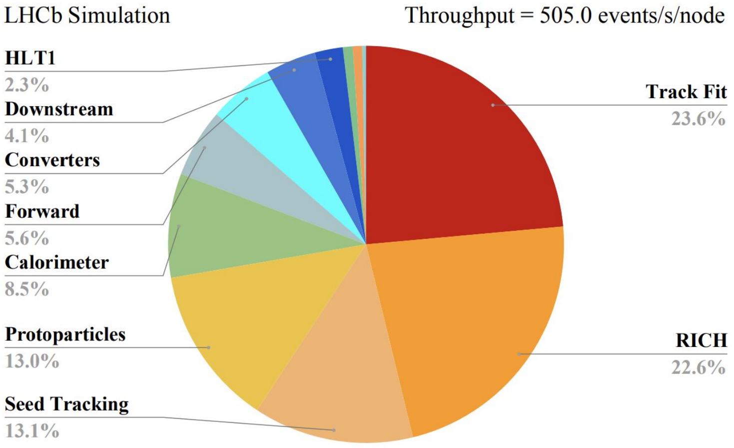
Runs at 500 kHz

Reduces BW to 10 GB/s

LHCb-FIGURE-2022-005

Full event reconstruction with offline quality:

- More complex tracking algorithms, using all alignment conditions
- Full Kalman fit for precise track momentum determination
- Full calorimeter reconstruction, including cluster corrections, neutral PID, jets, etc.
- RICH reconstruction and combined PID variables
- Selections for full physics programme O(10^3)
- Persistence of reco'ed objects for offline usage



HLT2: the turbo model

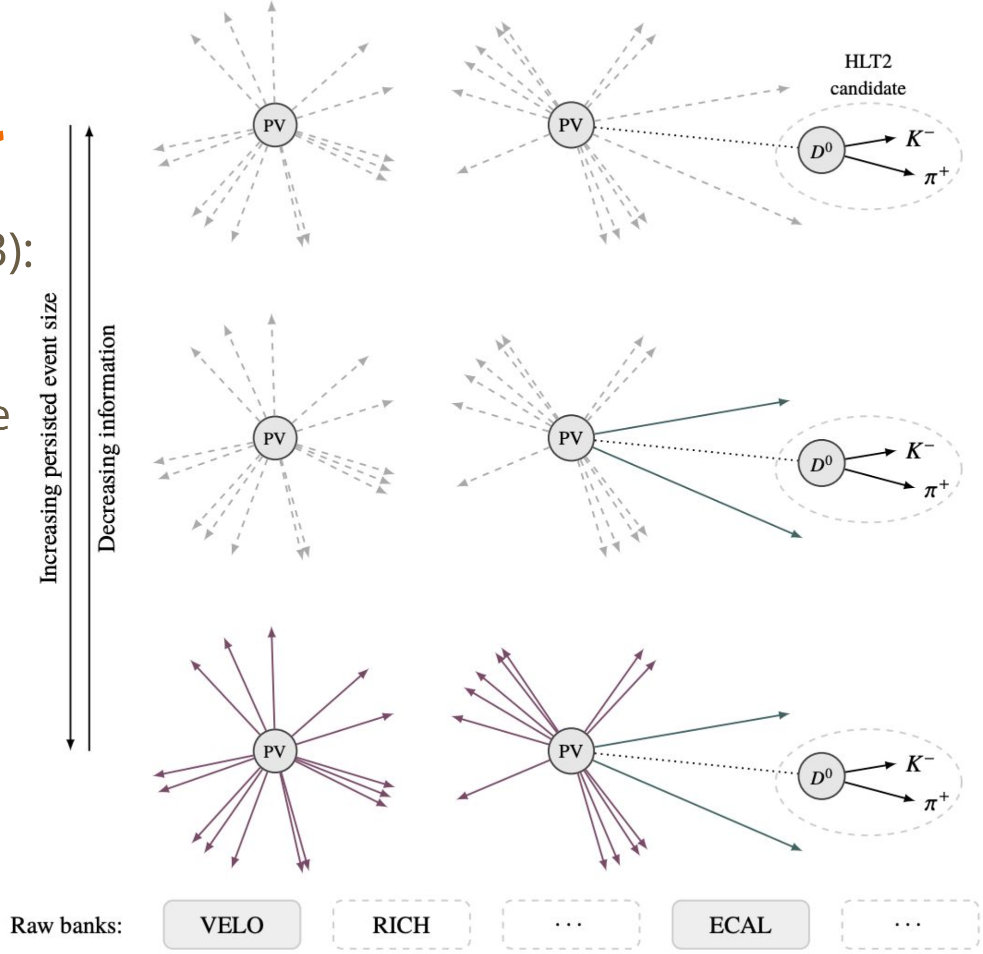
BW(kB/s) = rate(Hz) x event size(kB):
4 TB/s input → 10 GB/s offline limit

- huge signal rate → reduce evt size
- rare signal → can write more info

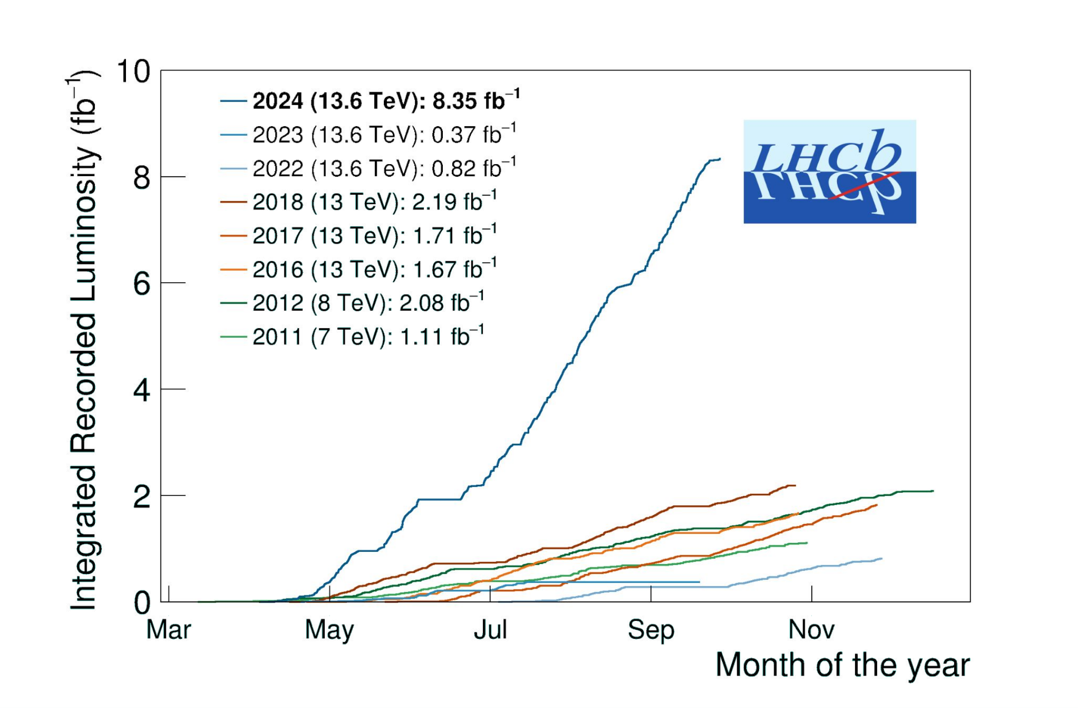
Flexible persistence model:

- **Turbo** (35 kB): signal only
- **Full** (70 kB): all reco'ed objects
- **Selective**: signal + selection of reconstructed objects and raw banks

JINST 14 P04006



Performance in 2024



The challenges in 2024

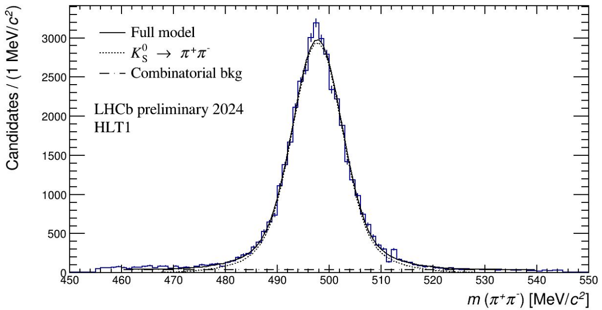
- Pile-up increase: started at 1 and increased in steps until 5.3
  - HLT1 thresholds optimised at each step for max physics output (BW division)
  - HLT2 selections monitored and tuned when needed for different pile-ups
  - HLT1 & HLT2 throughput continuously monitored and improved
- UT detector included in data taking since June
  - tracking without UT delibered first, optimised for low and high pile-up
  - tracking with UT delibered since then in both HLT2 and HLT1 → background reduction
  - full alignment of tracking system first without UT, then largely improved with UT

The challenges in 2024

- Pile-up increase: started at 1 and increased in steps until 5.3
  - HLT1 thresholds optimised at each step for max physics output (BW division)
  - HLT2 selections monitored and tuned when needed for different pile-ups
  - HLT1 & HLT2 throughput continuously monitored and improved
- UT detector included in data taking since June
  - tracking without UT delibered first, optimised for low and high pile-up
  - tracking with UT delibered since then in both HLT2 and HLT1 → background reduction
  - full alignment of tracking system first without UT, then largely improved with UT

Online monitoring of low and high-level metrics has proven critical!

LHCb-FIGURE-2024-013



LHCb-FIGURE-2024-006
LHCb-FIGURE-2024-007
LHCb-FIGURE-2024-014

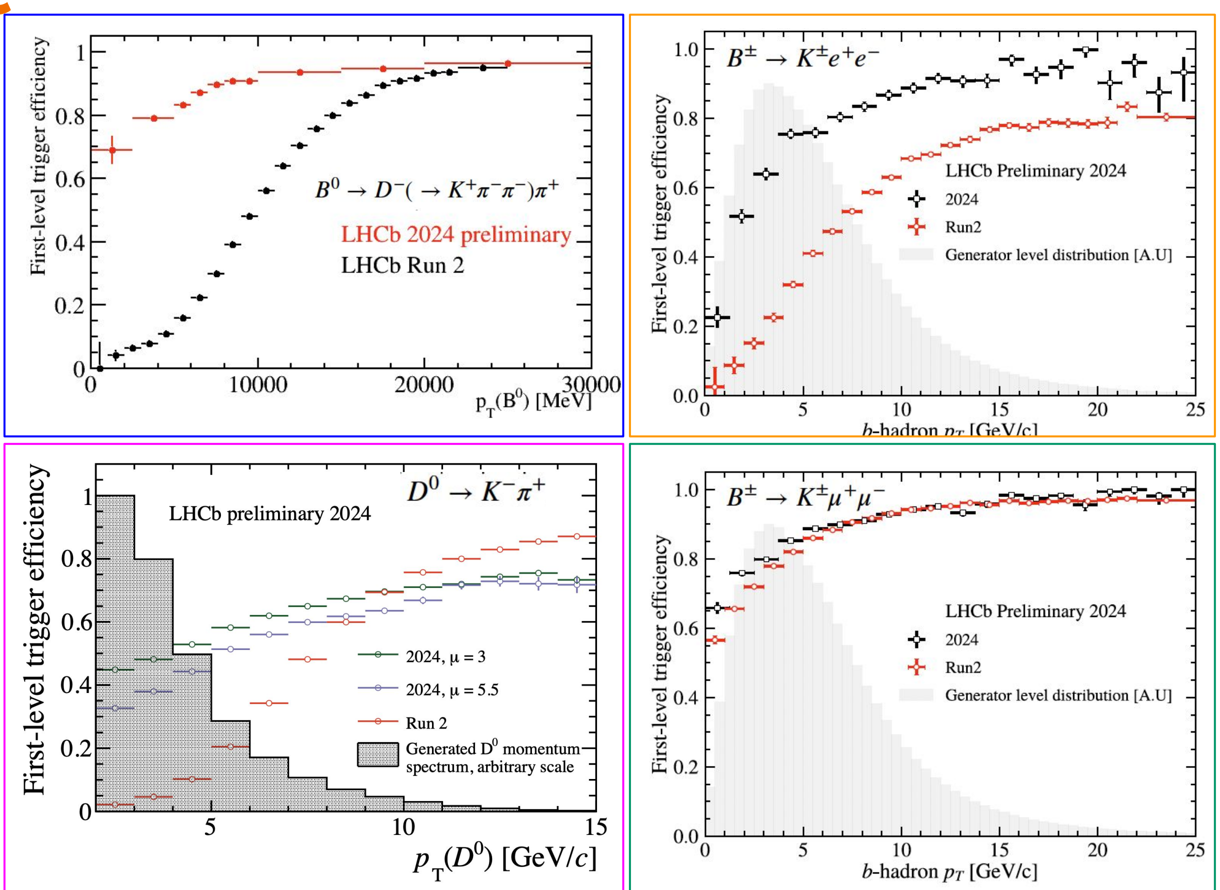
HLT1 performance

Successfully processing 30 MHz at pile-up of 5.3.

Large gains in efficiency compared to Run 2 for:

- hadronic modes
- electron modes
- charm decays

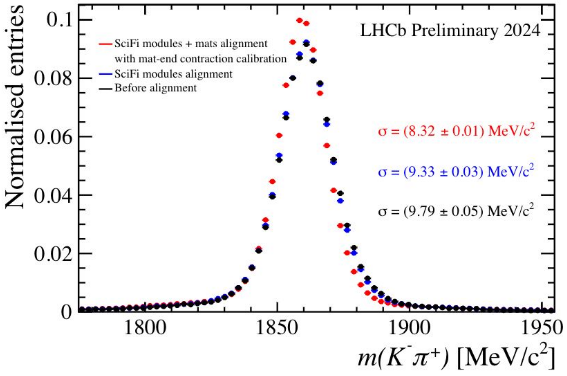
Even a bit for muon modes



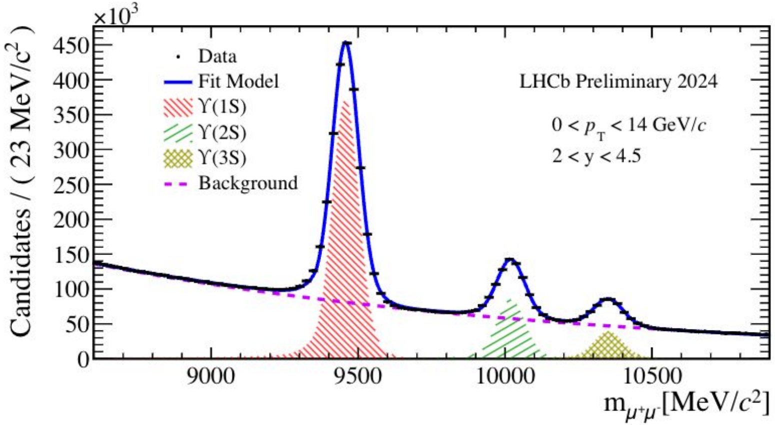
Alignment & Calibration performance

Several alignment iterations improving momentum and mass resolution.

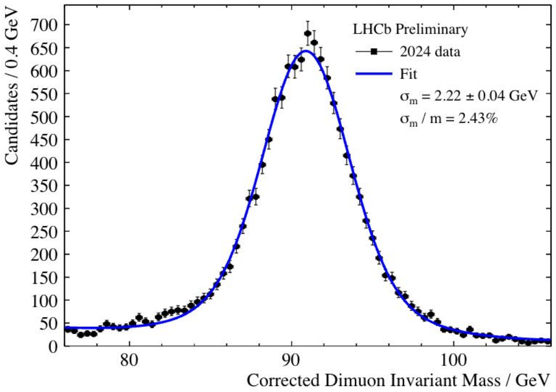
Ultimate performance after UT inclusion.



LHCb-FIGURE-2024-009



LHCb-FIGURE-2024-025

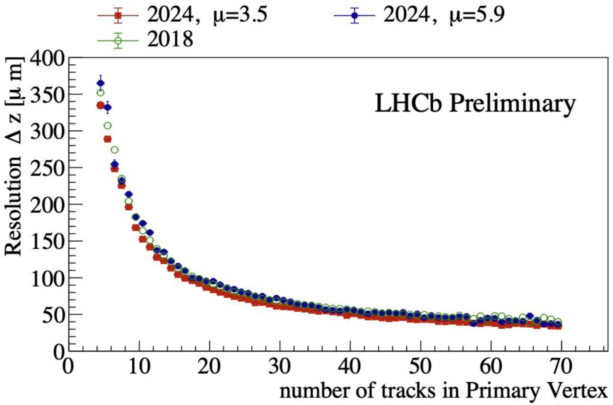


LHCb-FIGURE-2024-020

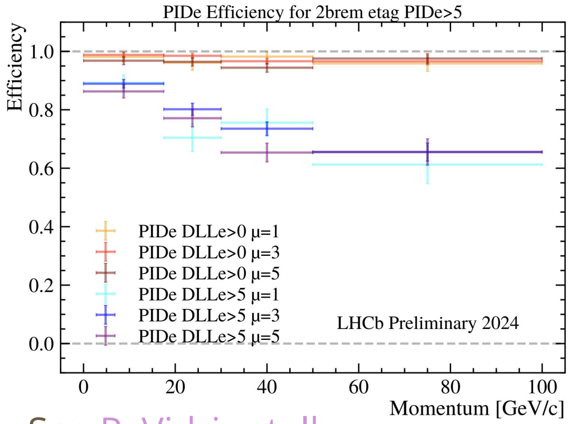
HLT2 performance

Good track reconstruction and excellent vertex resolution.

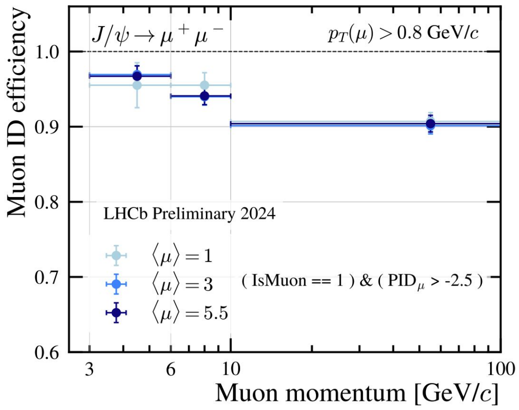
Stable PID performance with pile-up.



LHCb-FIGURE-2024-011

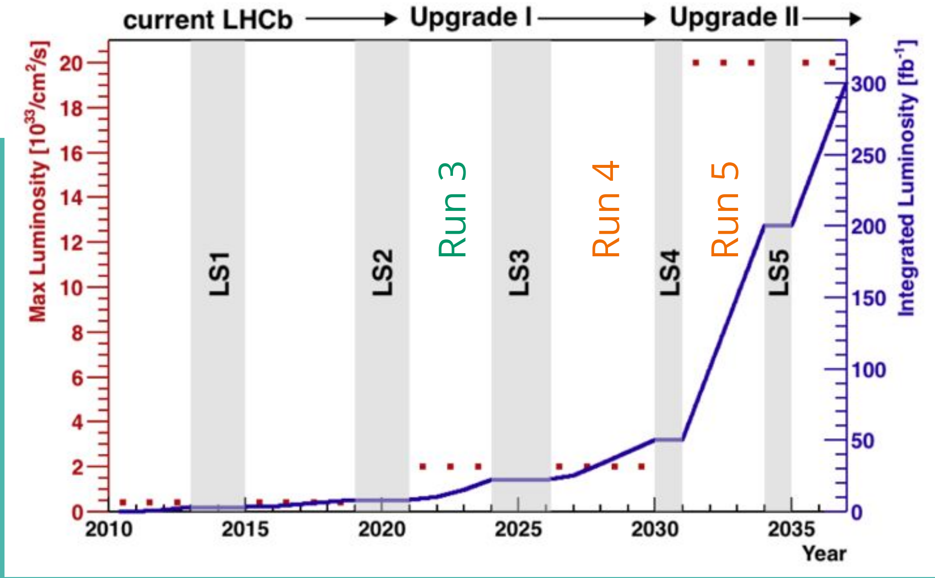


See P. Vidrier talk



LHCb-FIGURE-2024-010

Plans for Run 4 and beyond

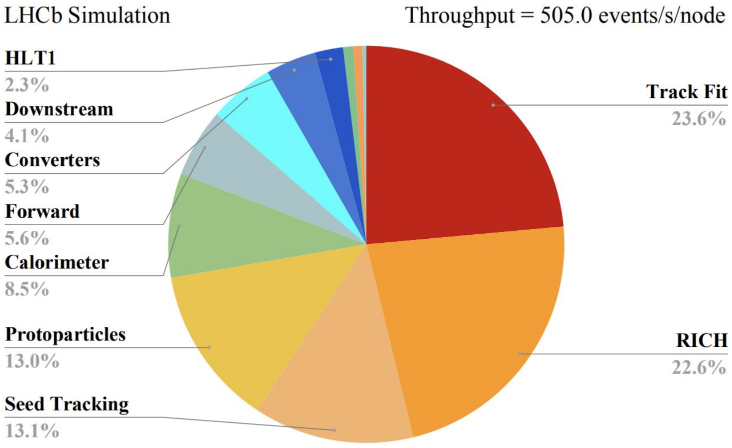


LHCb trigger in Run 4

New prototype detectors in Run 4, but no major upgrade for LHCb. Also, same instantenous luminosity than Run 3 → similar trigger strategy to Run 3.

R&D prototypes to prepare for Run 5, focusing on most consuming algorithms:

- downstream tracker in FPGAs: pre-build Scifi track in FPGAs to speed up HLT1 & HLT2
- RICH reconstruction in GPUs: would enable much cleaner exclusive selections in HLT1
- full track fit in GPUs: would enable more precise momentum determination in HLT1

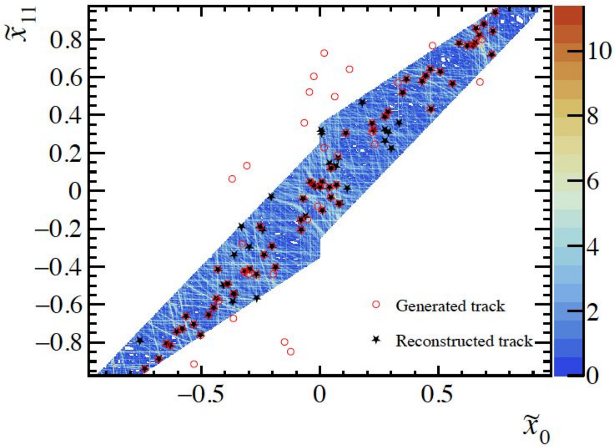


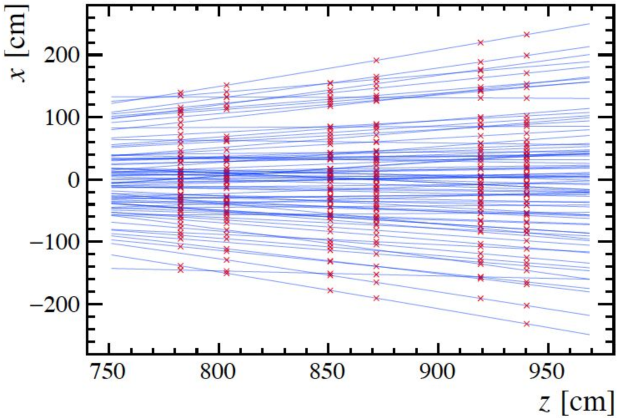
Run 4: downstream tracker in FPGAs

CERN-LHCC-2024-001

Use RETINA architecture based on human vision

- used in Run3 for Velo clustering
- developing Scifi seeding for Run 4 (~10% of HLT2)
- promising results on MC and first tests with real data on Run 3 testbench





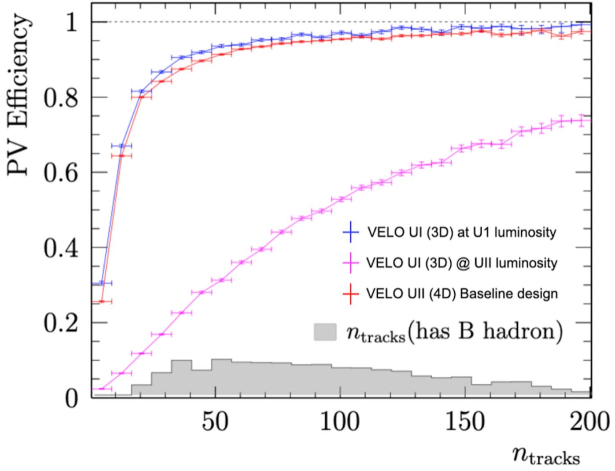
LHCb trigger in Run 5

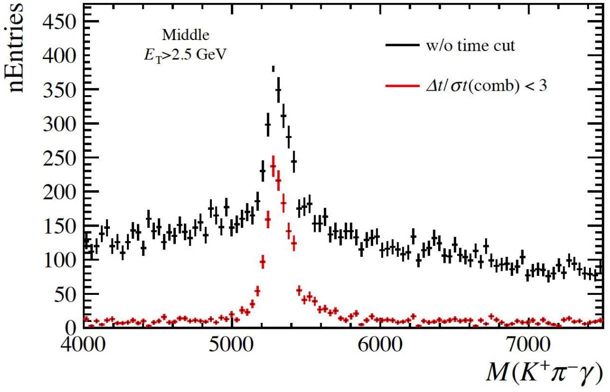
CERN-LHCC-2024-010

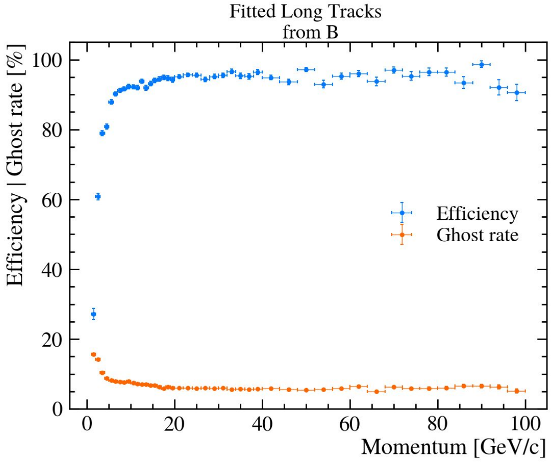
Major upgrade proposed for LHCb Run 5: x7.5 L_{inst} → 1.5 x10^{34} cm^{-2} s^{-1}

Timing information critical → 4D reconstruction

- demonstrated performance with proposed detectors







LHCb trigger in Run 5

CERN-LHCC-2024-010

Major upgrade proposed for LHCb Run 5: x7.5 L_{inst} → 1.5×10^{34} cm^{-2} s^{-1}

Baseline trigger design follows Run 3 approach:

- triggerless readout: HLT1 + disk buffer (Ali & Cali) + HLT2
  - both would need to run on GPUs to achieve necessary throughputs
- alternatives being considered:
  - run reconstruction in GPUs but selection in CPUs
  - open to new technologies (IPUs, ARM, RISC-V, etc)
  - energy consumption should be kept in mind

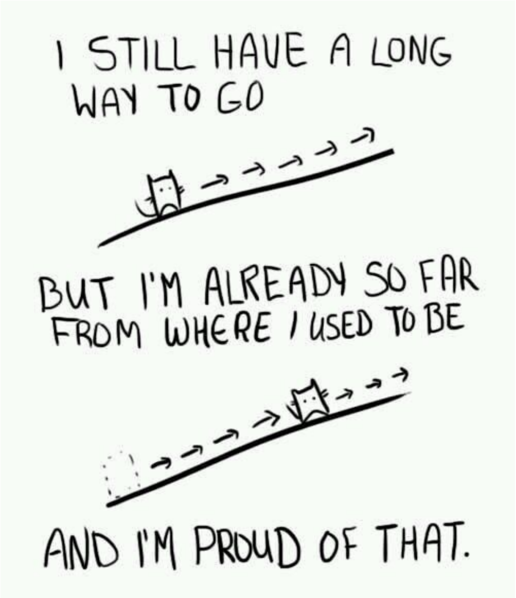
Conclusions

Wide range of physics at LHCb → flexible trigger

Unique trigger-less readout and Turbo model allow to benefit from x5 lumi → 2024 data ≈ Run 1 + Run 2

Strong progress in 2024 reaching nominal pile-up and anticipated trigger gains, specially for:

- hadronic and electron final states
- charm decays
- long-lived particles

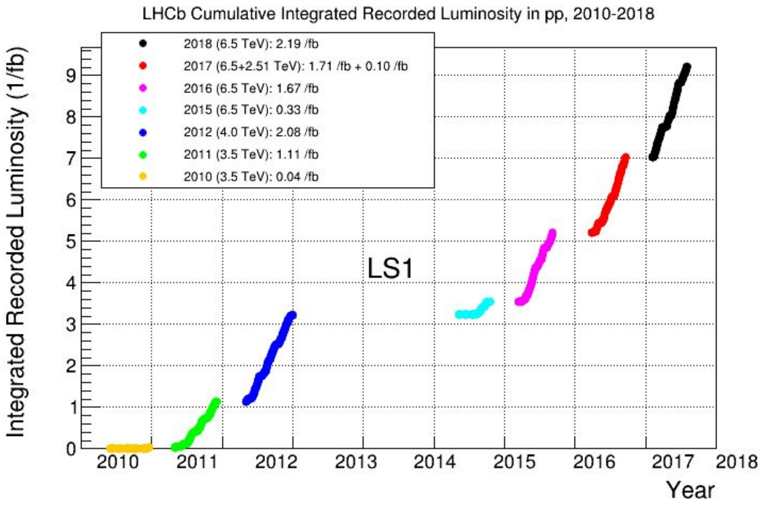


Stayed tuned for first physics results!

Thanks for the attention

BACK-UP

LHCb dataset in Run 1 and 2



All b-hadron species! [PRD100(2019)031102]

- B_s: $\frac{f_s}{f_d+f_u} = 0.122 \pm 0.006$
- Λ_b: $\frac{f_{\Lambda_b}}{f_d+f_u} = 0.259 \pm 0.018$

and more: Ξ_b, Ω_b, B_c, B^* ...

Total recorded luminosity ~9 fb^{-1}:

- Run 1 (2010-2012) ~ 3 fb^{-1}
- Run 2 (2015-2018) ~ 6 fb^{-1}

x2 b-quark production from 7 to 13 TeV pp collisions
→ around x4 b-hadrons in Run 2

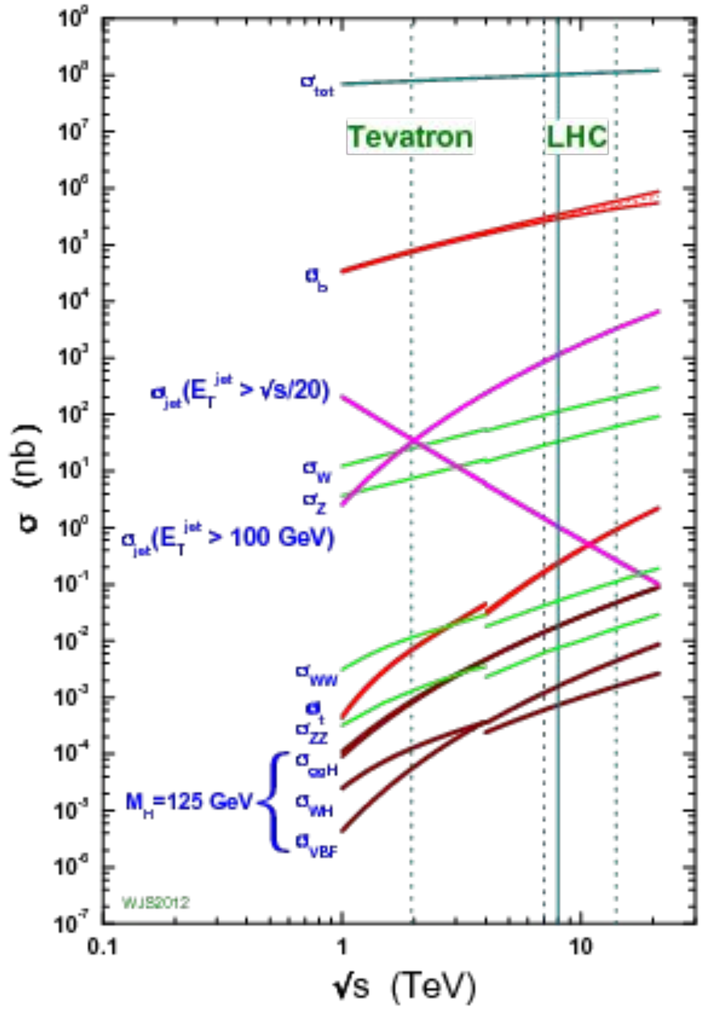
Run 3: physics & constraints

Heavy flavour physics at LHC:

- $L_{inst} = 2 \cdot 10^{33}$, x5 that of Run 2
- huge b and c productions

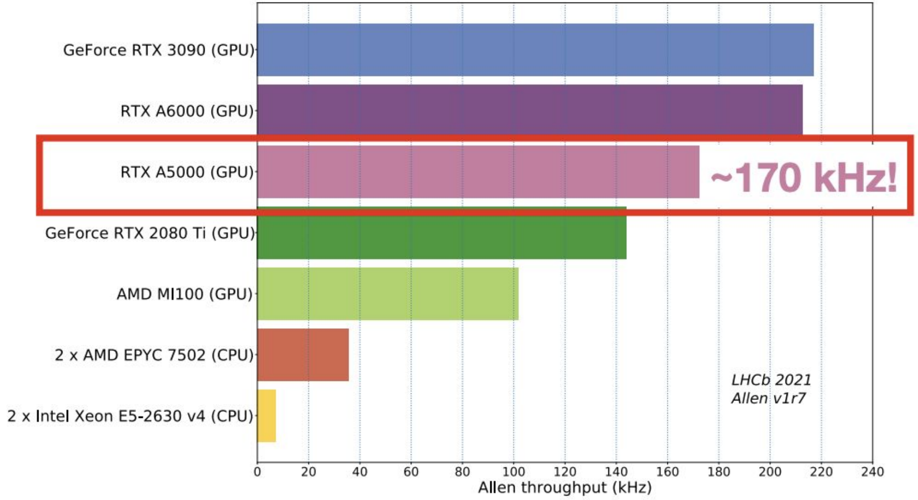
Trigger goals:

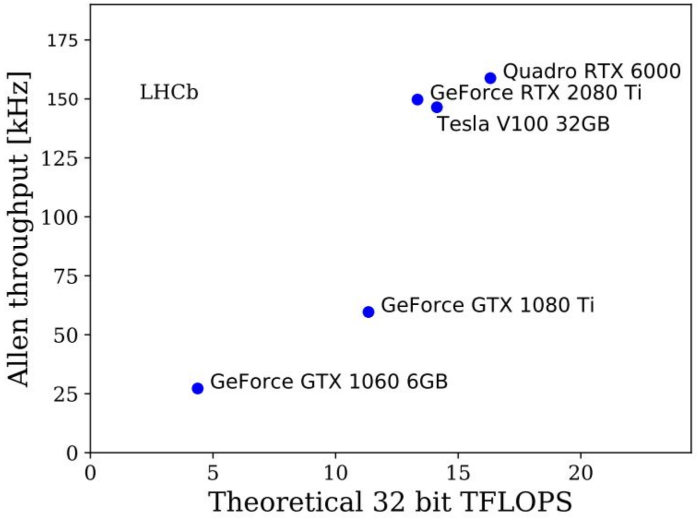
- Select interesting events: $O(10^{5\text{-}6}$ Hz)
- Reduce bandwidth: 4 TB/s → 10 GB/s



GPU choice

C. Agapopoulou @ICHEP 22



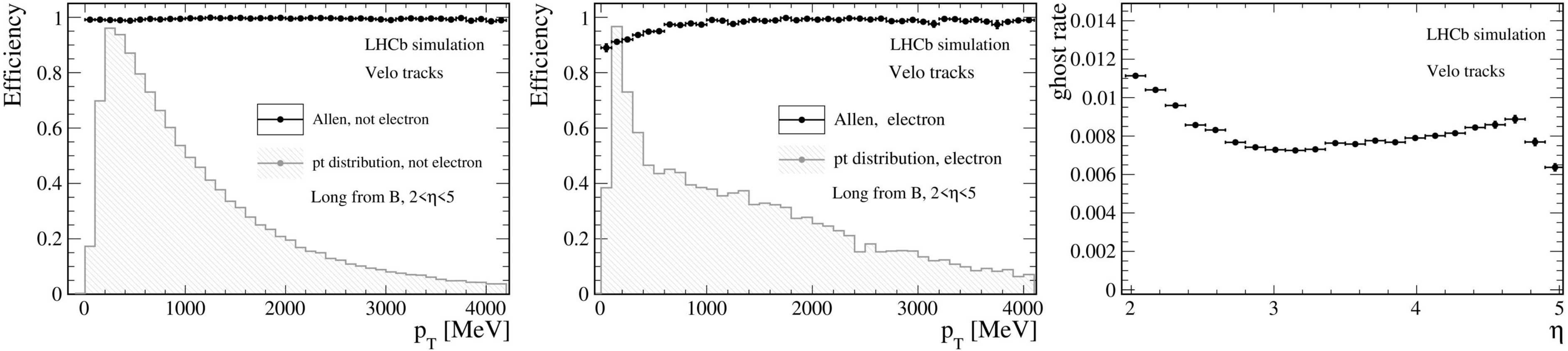


LHCb-FIGURE-2020-014

HLT1 performance: Velo

LHCB-FIGURE-2020-014

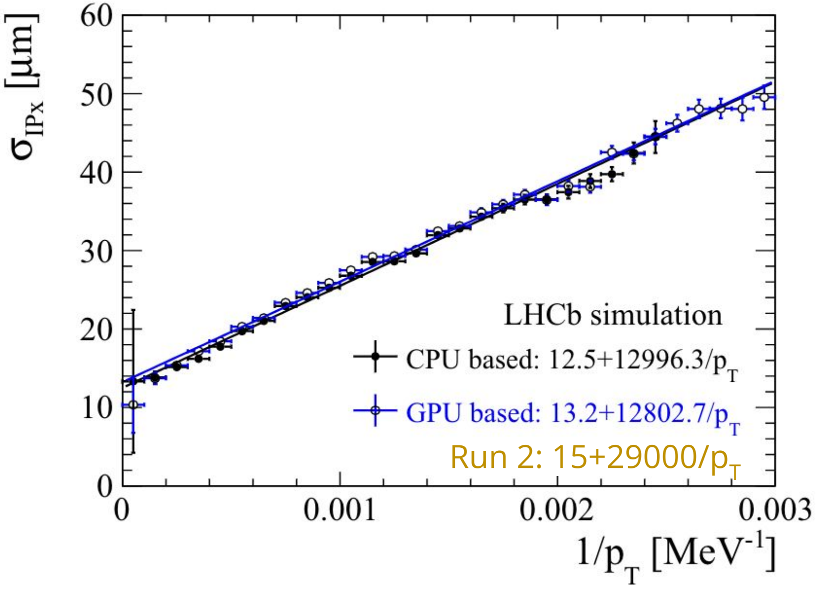
Same performance at x5 luminosity: high efficiency, good δp, low fake rate

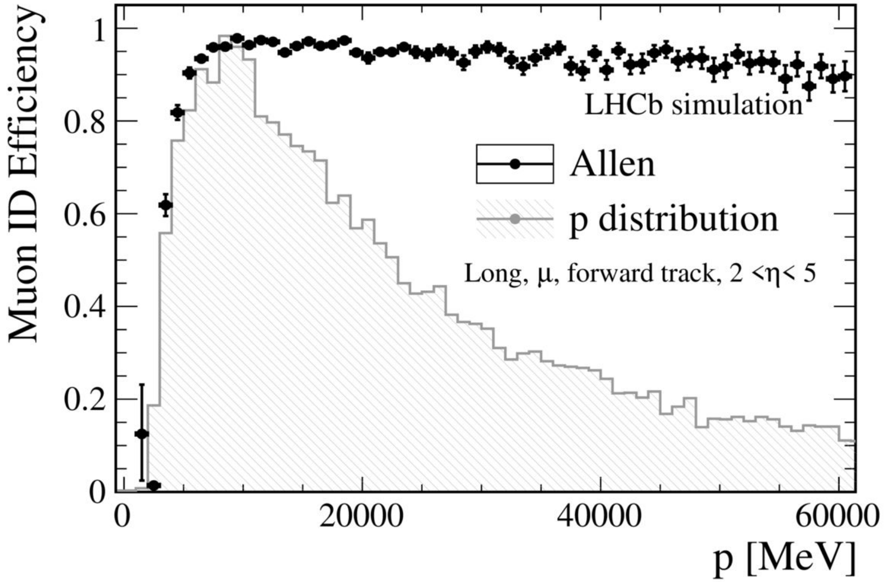


HLT1 performance

LHCB-FIGURE-2020-014

Same performance at x5 luminosity: high efficiency, good δp, low fake rate

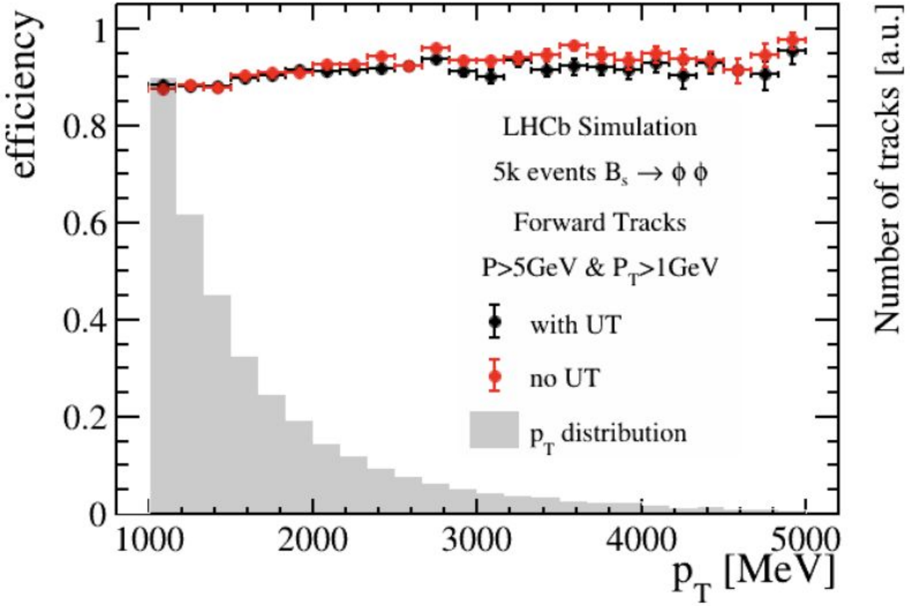


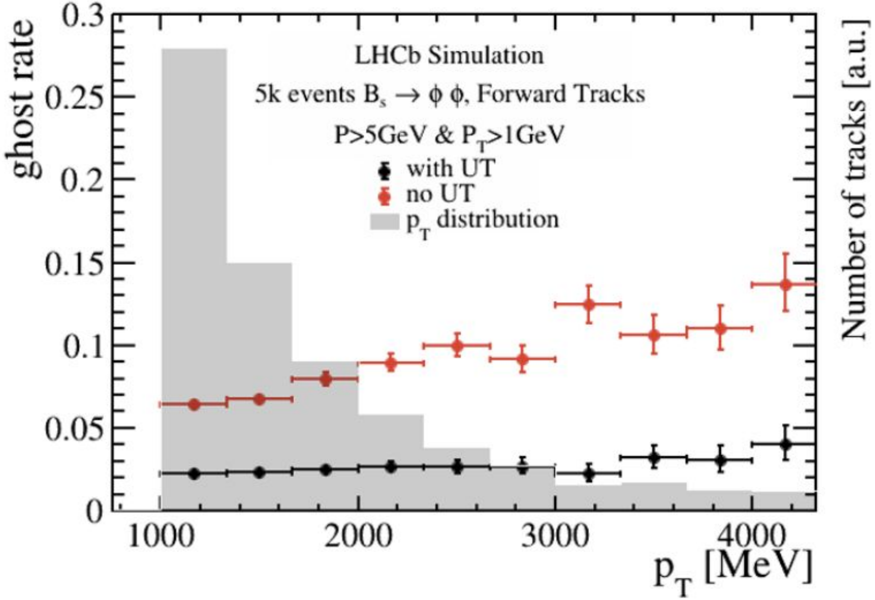


HLT1 without UT

LHCB-FIGURE-2022-007

Same signal efficiency but larger fake rate

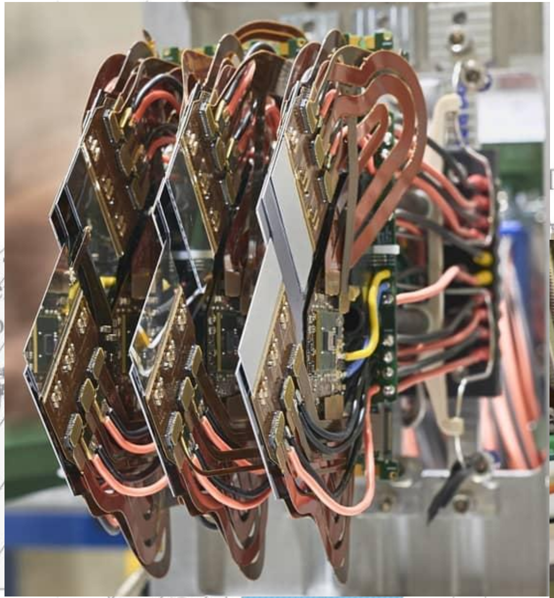




LHCb Upgrade

[CERN-LHCC-2013-021]

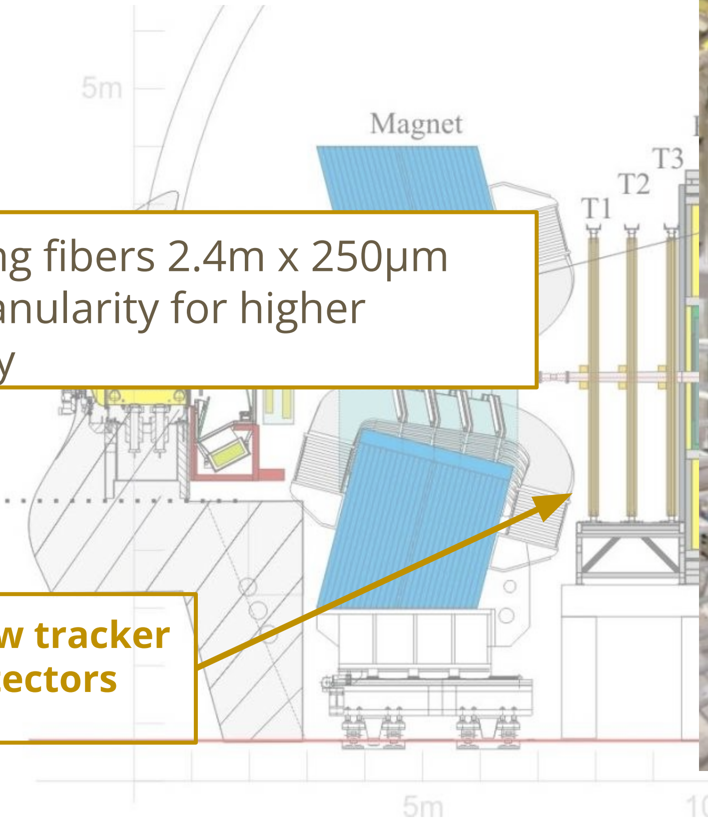
New VeloPix detector



- 3.5mm to beam (5mm Run1/2)
- 41M pixels of 55x55 μm
- improved PV and IP resolutions

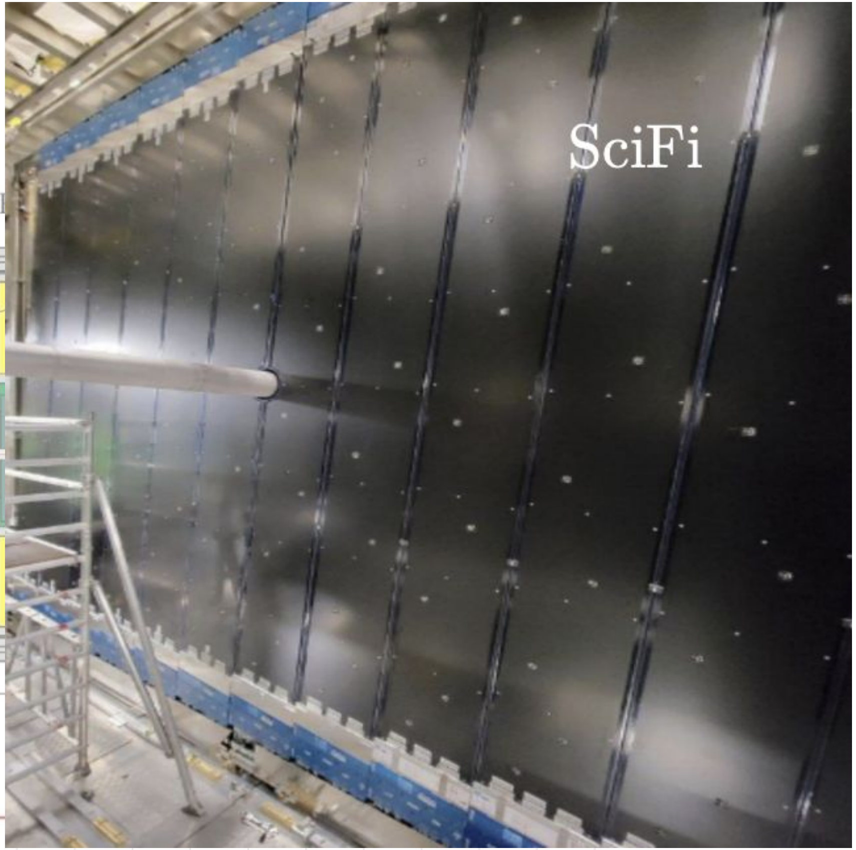
CERN-LHCC-2014-001

LHCb Upgrade



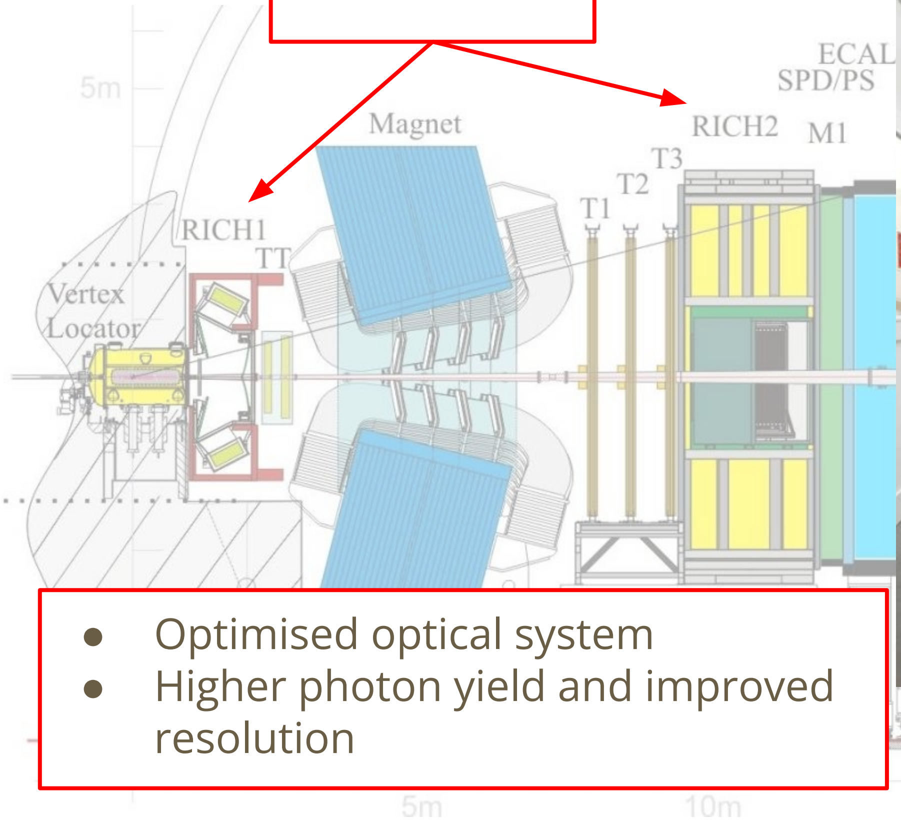
- Scintillating fibers 2.4m x 250µm
- higher granularity for higher occupancy

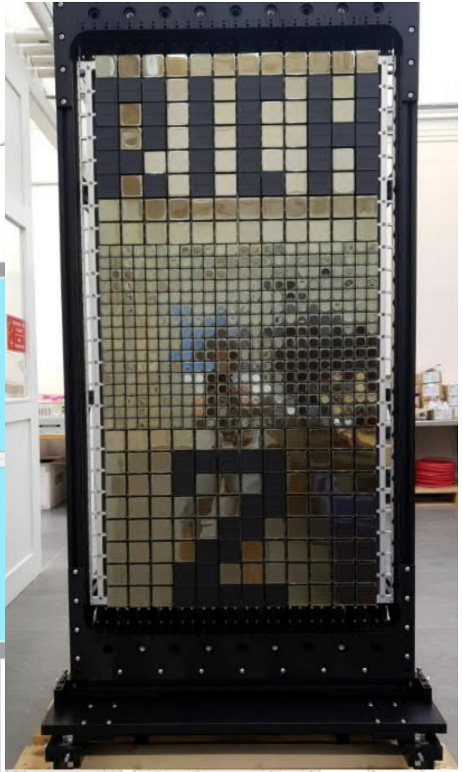
New tracker detectors



CERN-LHCC-2013-022

LHCb Upgrade





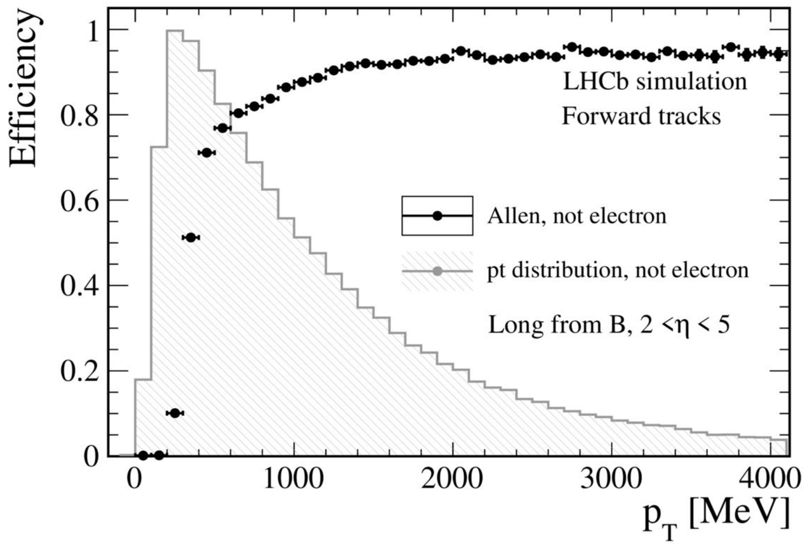
Upgraded RICH detectors

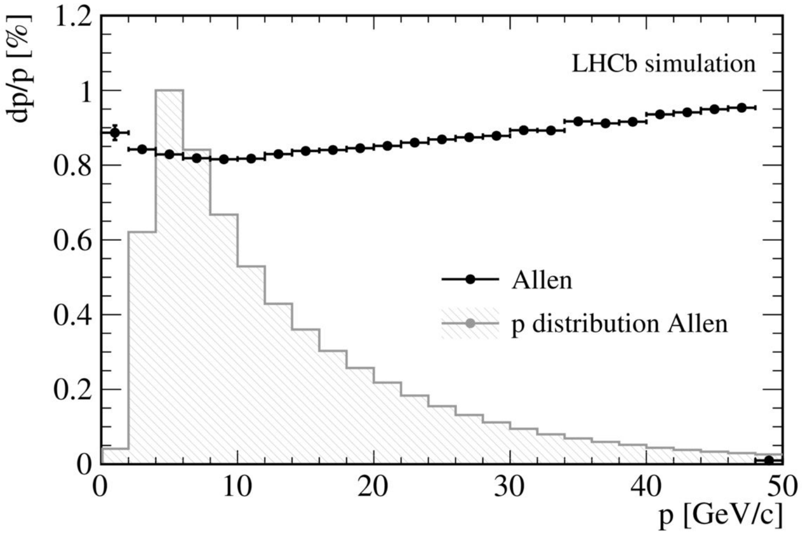
- Optimised optical system
- Higher photon yield and improved resolution

HLT1 performance on MC:

LHCB-FIGURE-2020-014

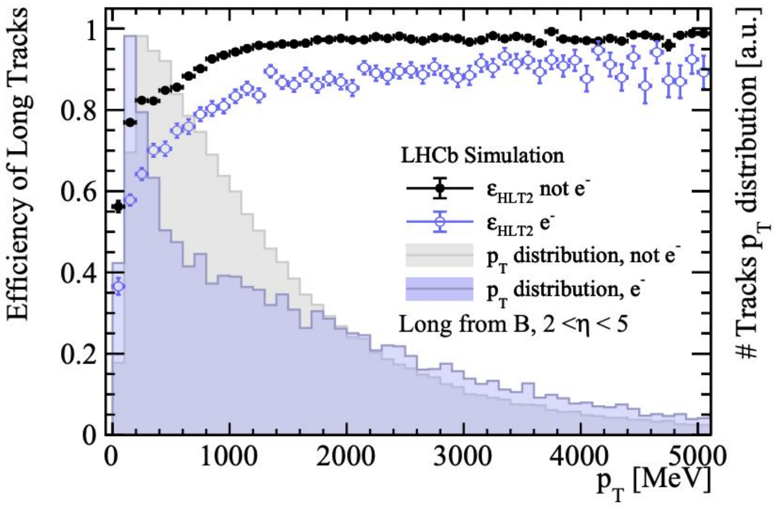
Same performance at x5 luminosity: high efficiency, good δp, low fake rate



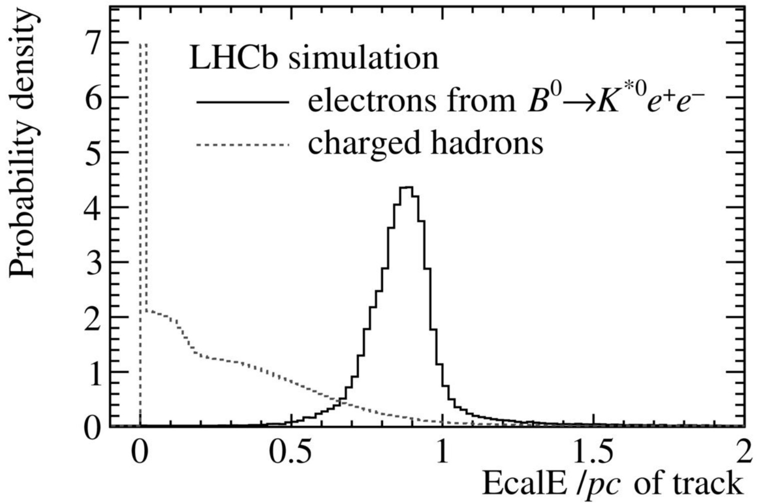


HLT2 performance on MC

Full reconstruction of tracks and neutrals, and PID with offline-quality



LHCB-FIGURE-2022-005

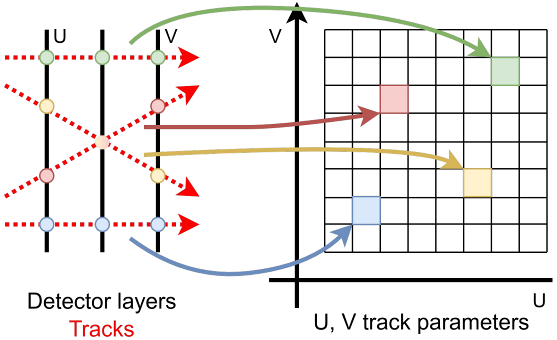


LHCB-FIGURE-2021-003

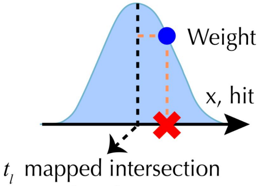
Run 4: downstream tracker in FPGAs

Use RETINA architecture based on human vision (used in Run3 for Velo clustering)

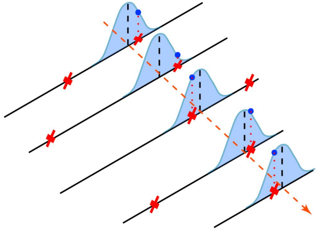
Step 1: Track space mapping



Step 2: Accumulating weights (each cell)

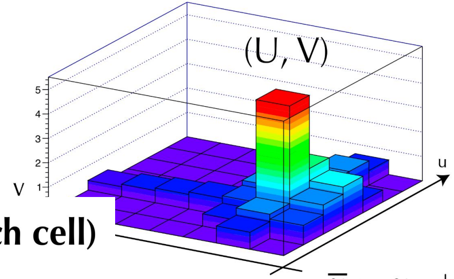


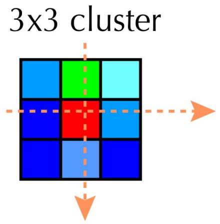
$$R = \sum^{\text{hits}} e^{-\frac{(x_l - t_l)^2}{2\sigma}}$$



R is close to N (# of layers) only if we have a set of hits near the mapped track

Step 3: Find the local maxima and compute centroid





$$\overline{u} = u_0 + \frac{\sum_{ij} i R_{i,j}}{\sum_{ij} R_{i,j}}$$

$$\overline{v} = v_0 + \frac{\sum_{ij} j R_{i,j}}{\sum_{ij} R_{i,j}}$$

$i = U-1, U, U+1$

$j = V-1, V, V+1$