# Real-time analysis at the LHC

Rosen Matev (CERN)

24 May 2019

EIC Streaming Readout IV workshop

# What is real-time analysis?

- Online we have finite time to decide what data to keep (forever)
- Here, RTA means to efficiently reduce data online
- If we are reducing, what do we keep?
  - a paper is probably too extreme, but may be useful for a preliminary result!

- Briefly show the real-time analysis landscape at the LHC
- Delve a bit deeper into LHCb
- Focus on the software part

# Motivation

- Triggering is expensive; must fit within computing constraints

$$\text{Bandwidth [GB/s]} \propto \text{Accept rate [kHz]} \times \text{Event size [kB]}$$

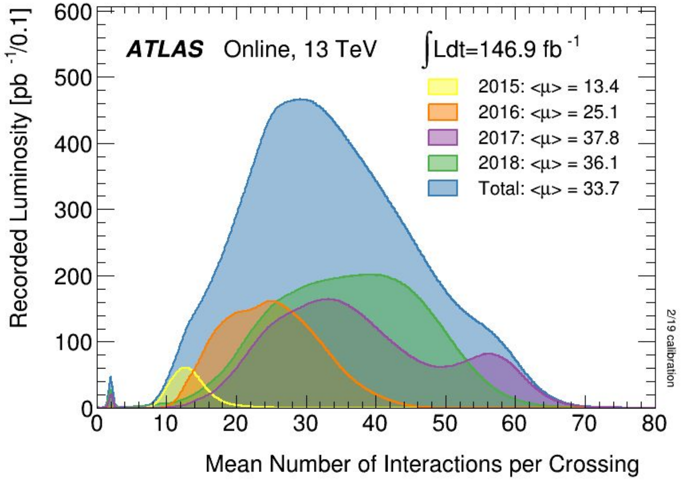
- Want highest accept rate high to maximise $\varepsilon_{Sig.}$ and reduce bias
    - Balanced against maximising $1 - \varepsilon_{Bkg.}$
- Typically, can't do much to reduce the raw event size*; it's all or nothing!

If event size is reduced, there's room for more physics!

# Ever increasing pile-up

- Traditionally, we keep all raw data for events that contain signal
- Problem is, **raw data bandwidth scales quadratically** with luminosity
  - more signal events, but much more bgr. data!
- The question is becoming less "Is this event/frame interesting?"
  - instead, "Which part of this event containing signal should we save?"
  - and how do we do it efficiently

LuminosityPublicResultsRun2

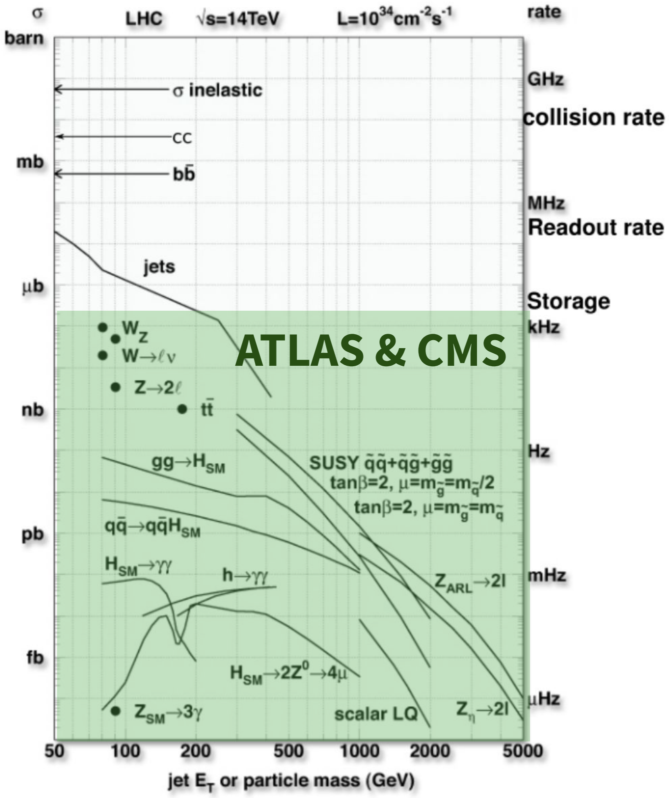


# ALICE

See talk by
P. Vande Vyvre

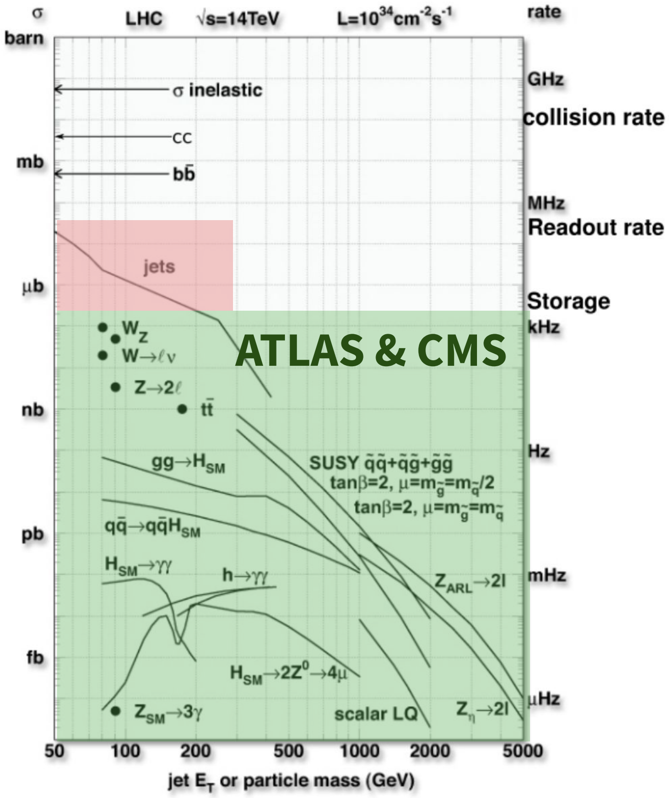
- Almost all minimum bias events contain physics
  - LHC can deliver up to 50 kHz
- Zero suppression
  - is this considered analysis? yes, non trivial, needs real-time calibration
- Compression with Huffman/ANS coding
  - save track parametrization + residuals
  - needs tracking
  - needs calibration! $\Rightarrow$ feedback loop
- Discard clusters not part of tracks
- Big buffer that accumulates data
  - asynchronously processed 1-2 times in the following months of no beam period
  - archived

# High mass physics




- A trigger is needed to reduce storage and readout costs
- A good trigger does so by keeping more signal than background
- General purpose LHC experiments are interested in signatures in the kHz region
  - Readout at 100 kHz is efficient with reasonably straightforward ET requirements

# High mass physics

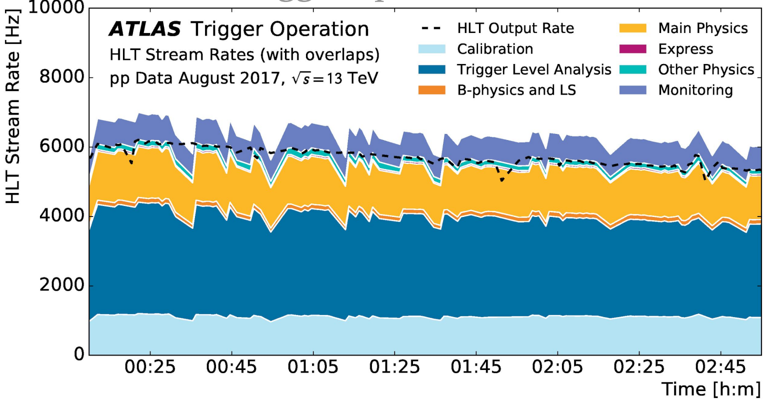



- A trigger is needed to reduce storage and readout costs
- A good trigger does so by keeping more signal than background
- General purpose LHC experiments are interested in signatures in the kHz region
    - Readout at 100 kHz is efficient with reasonably straightforward ET requirements
- What about that bit?

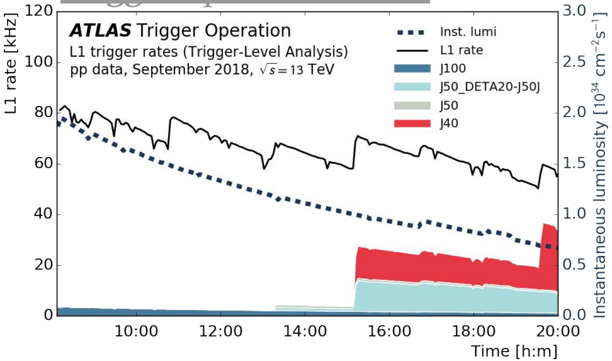
# ATLAS "Trigger-Level Analysis"

- Store only HLT jet 4-vectors and some summary info (e.g. Nconstituents)
  - event is tiny, 0.5% of full size!
  - **all 3 kHz** of relevant triggered events saved
- Profit from available L1 rate during fill
  - save **up to 25 kHz** in 2018
- Limitations
  - Parts of the jet calibration "not quite real-time"
  - Coarse L1 algorithms ⇒ bad resolution
  - No tracking available
  - Ideas to improve for Run 3 and HL-LHC

Trigger operations 2017



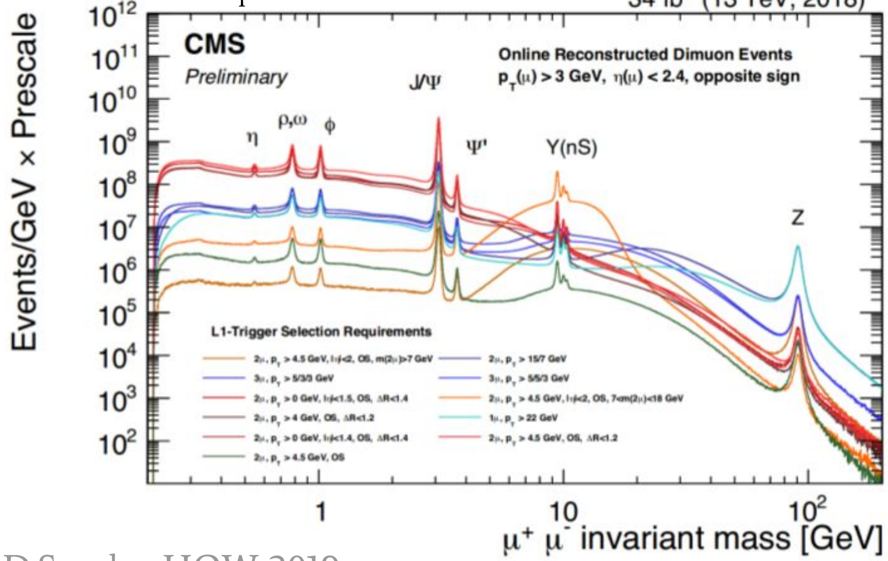

Trigger operations 2018



# CMS “Scouting”

- CaloScouting
  - vertices, muons, calo jets, MET
  - L1-limited
- ParticleFlow Scouting
  - vertices, PF muons, jets, cands, MET.
  - CPU-limited
- Possible Run 3 extensions
  - PF scouting on all L1 events?
  - or restrict on L1 input to limit CPU
- HL-LHC: 40 MHz scouting
  - tracking in L1
  - streaming readout of detectors

dimuon events using a collection of L1 muon triggers, and minimal requirements at the HLT level

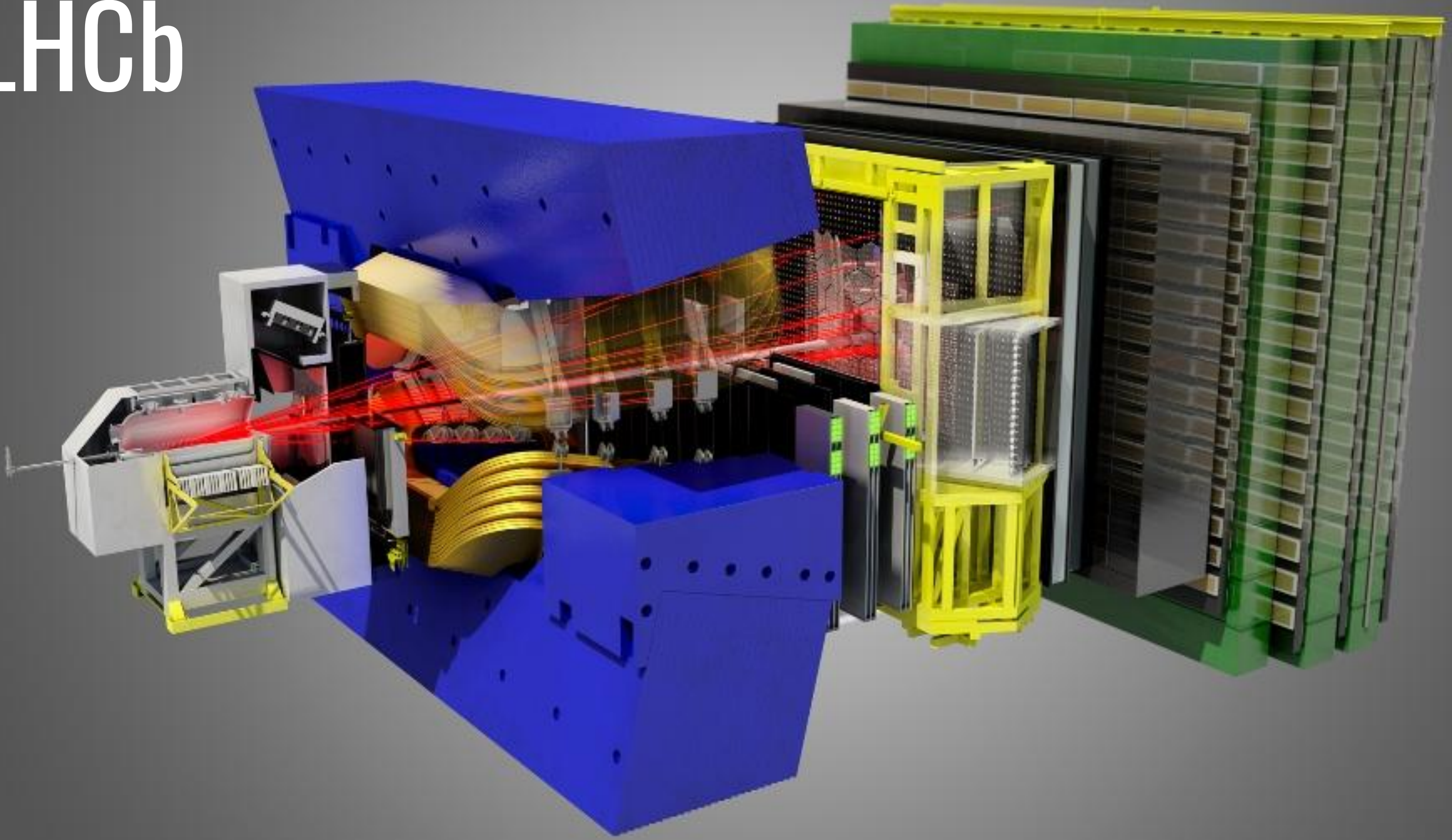



D.Sperka, HOW 2019

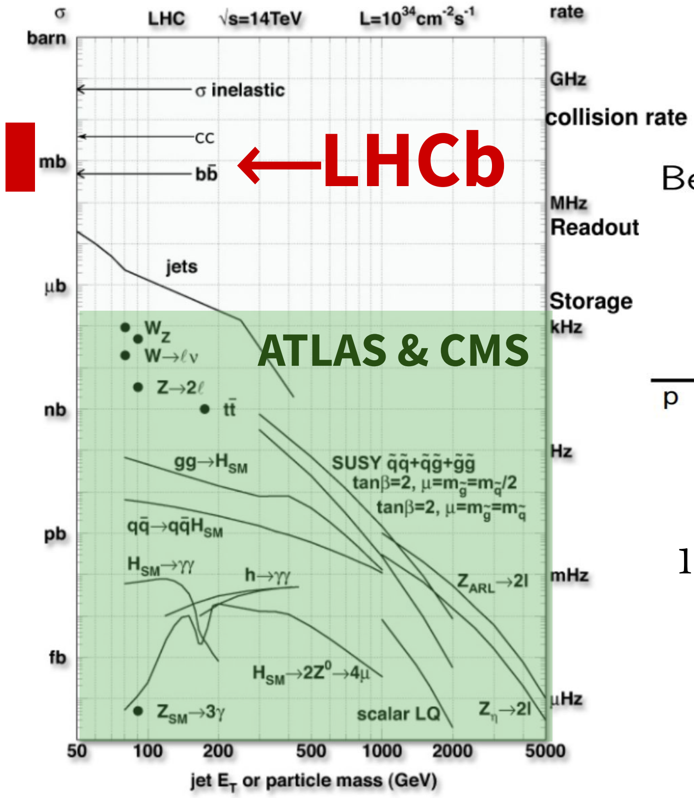
| Stream | Rate (Hz) | Event Size | Bandwidth (MB/s) |
|---|---|---|---|
| PhysicsMuons | 420 | 0.86 MB | 360 |
| PhysicsHadronsTaus | 345 | 0.87 MB | 300 |
| ScoutingCaloMuon | 4580 | 8.9 KB | 40 |
| ScoutingPF | 1380 | 14.8 KB | 20 |

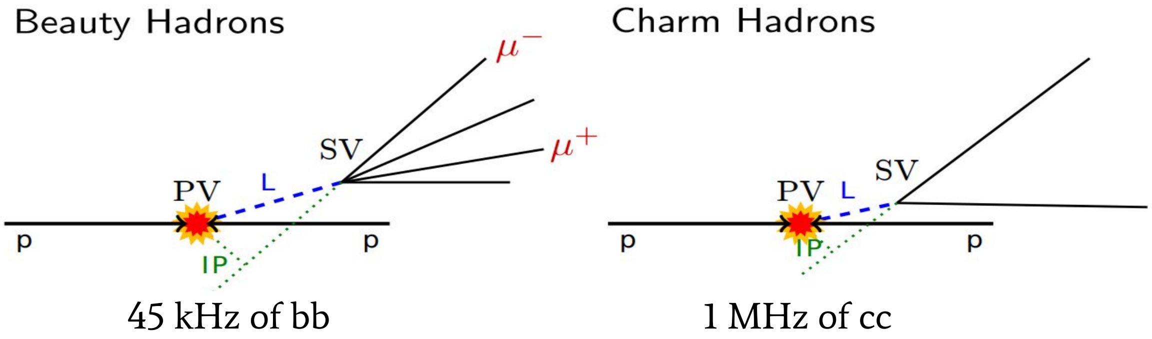
Selected CMS stream rate, event size, and bandwidth at the beginning of LHC Fill 7334 (23 Oct. 2018, L≈$1.5\times10^{34}cm^{-2}s^{-1}$)

# LHCb



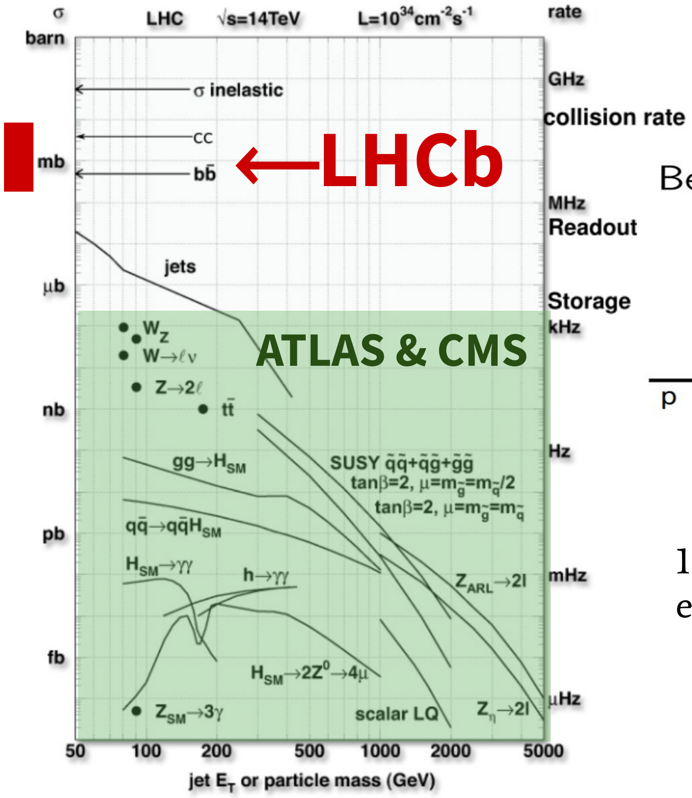
# LHCb Trigger in Run 2

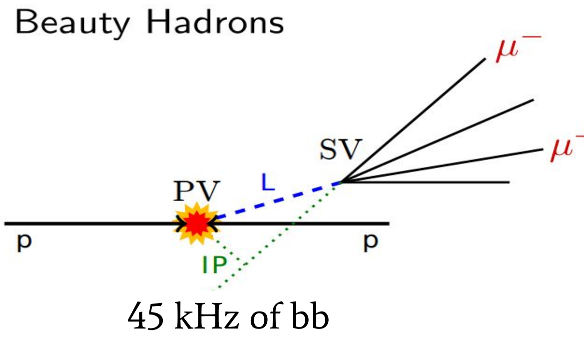


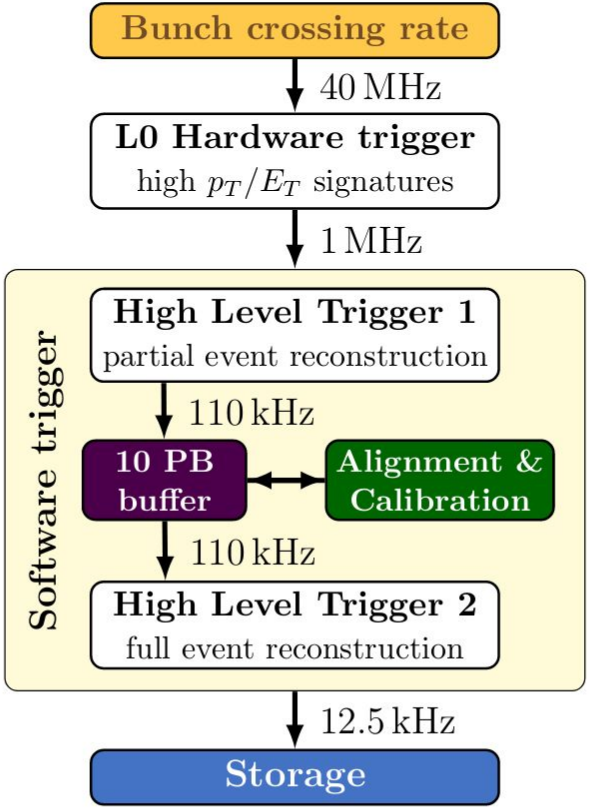

1 MHz readout is needed to stay efficient for beauty signals

# LHCb Trigger in Run 2

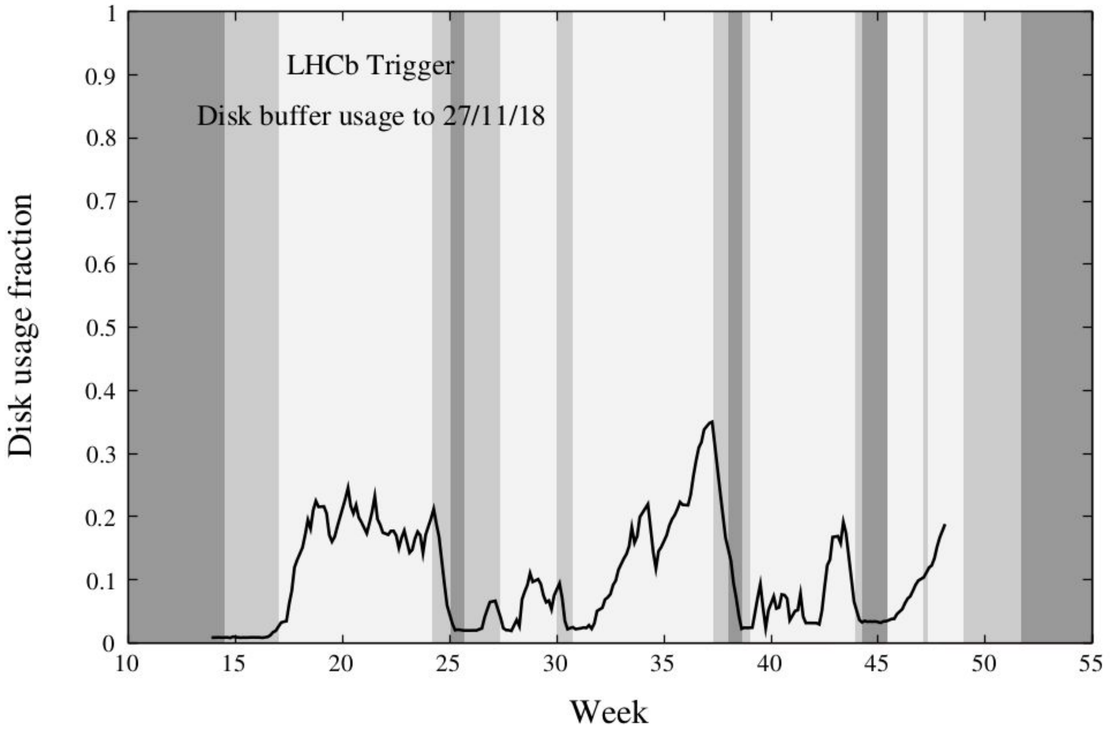



1 MHz readout is needed to stay efficient for beauty signals



# Disk buffer




11 PB of disk capacity

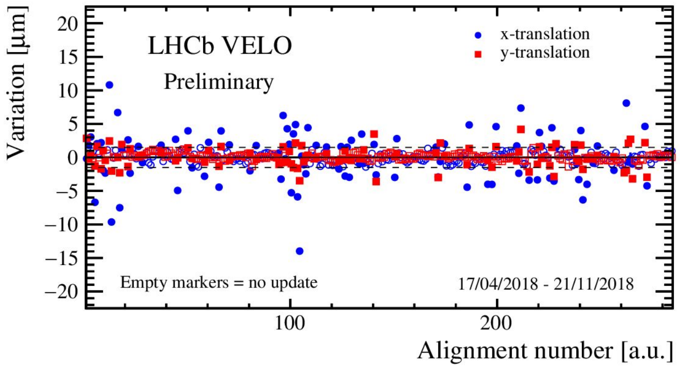
HLT1 writes at 110 kHz in fill

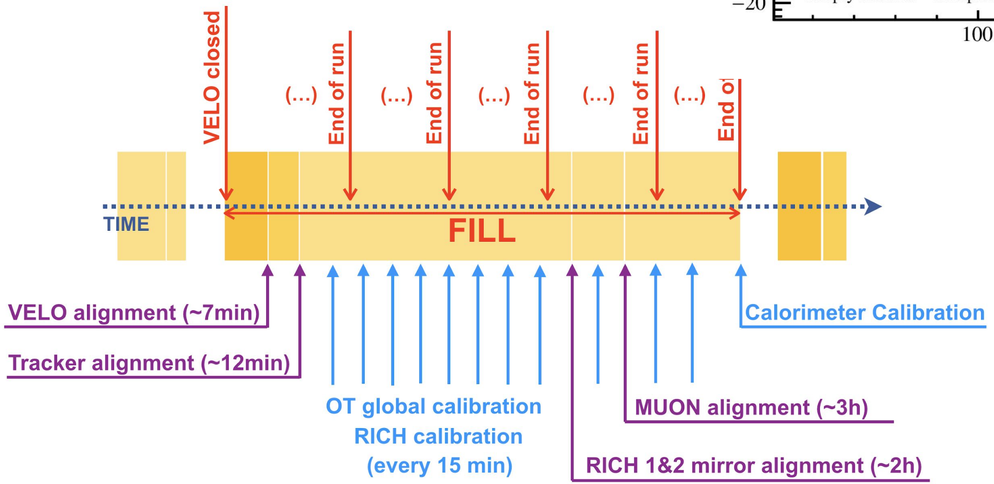
HLT2 processes at 30/90 kHz in/out-fill

Effectively **doubles the trigger CPU capacity**. Full event reconstruction becomes feasible.

# Real-time alignment and calibration

- Data collection & analysis fully automated
- New constants automatically applied
- Shift crew verifies updates

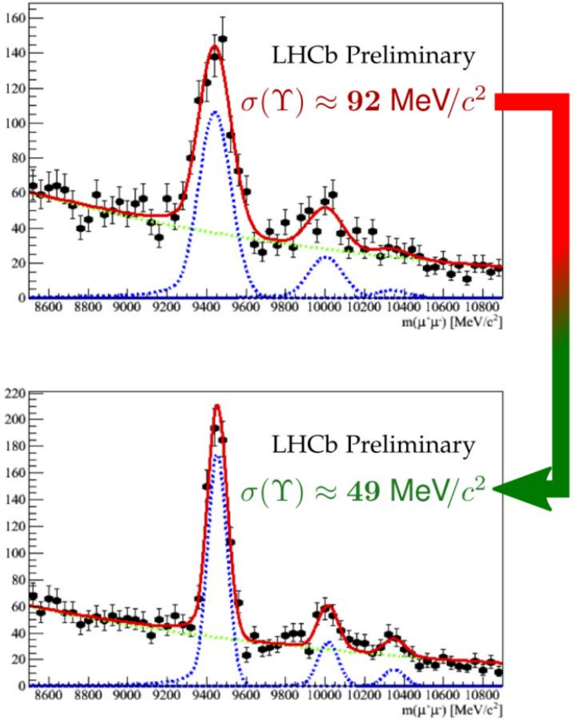



((~7min),(~12min),(~3h),(~2h)) - time needed for both data accumulation and running the task

# What this buys us

- Offline-equivalent, fully aligned and calibrated physics objects in HLT2
- Can include offline selections in the trigger with no associated systematic effects
- Offline reprocessing of the raw data is not necessary to recover information

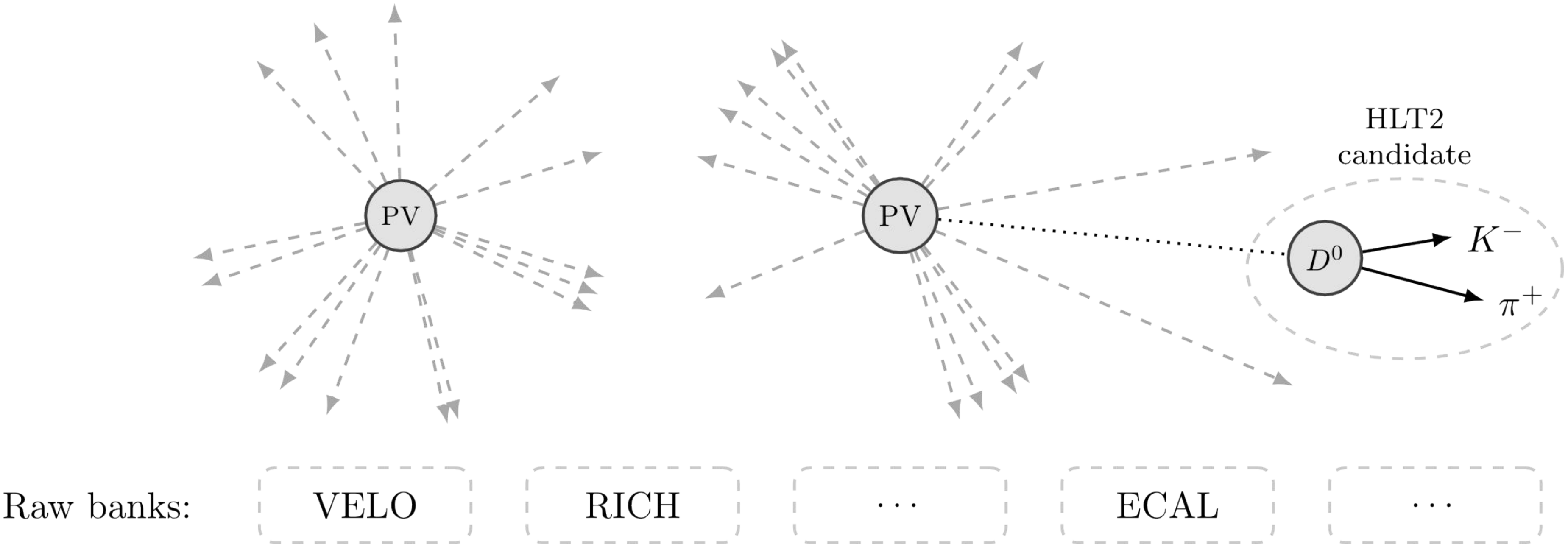



Real-time analysis with offline-quality physics objects

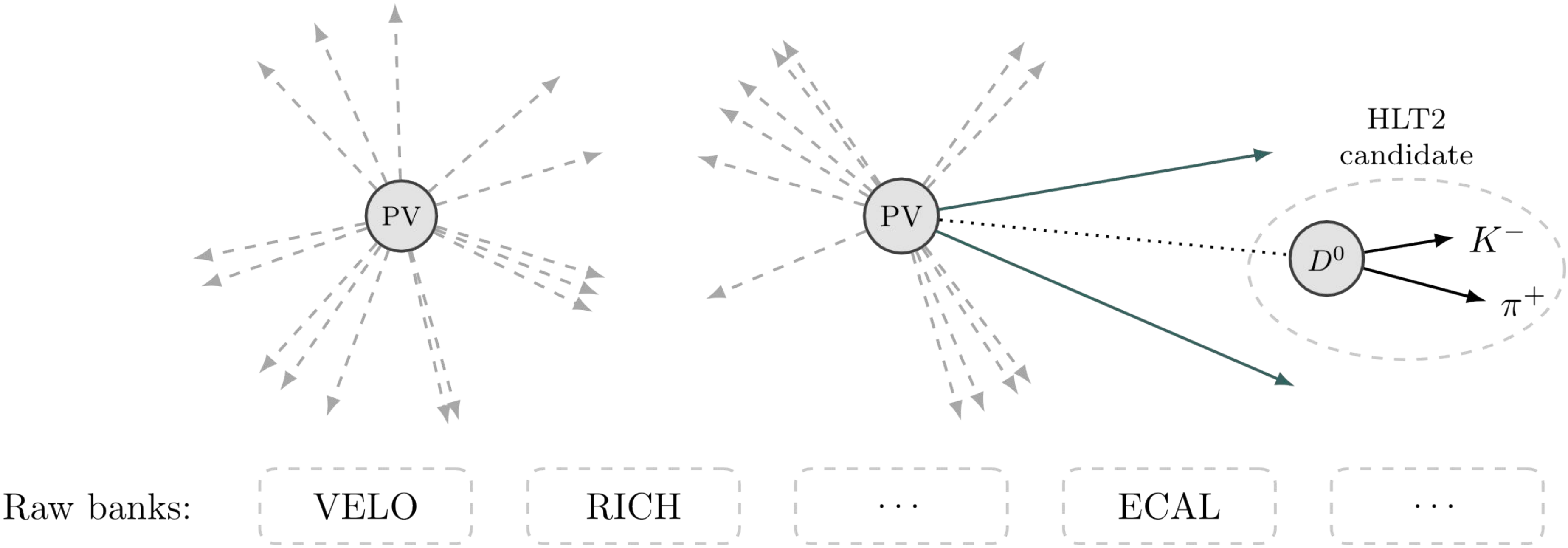
# Turbo

- Persist objects from HLT2 directly, analyse only these offline
- Each trigger selection has complete control over what objects are saved
- Evolved over time to meet increasing needs

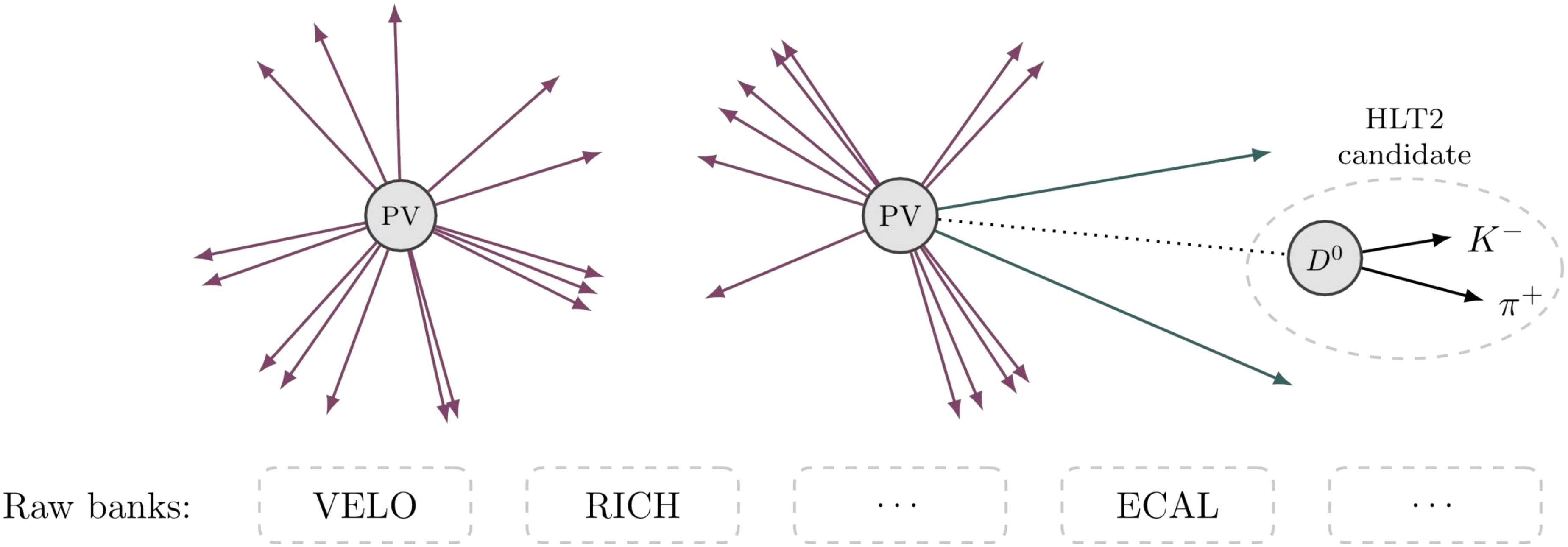
# Persistence granularity



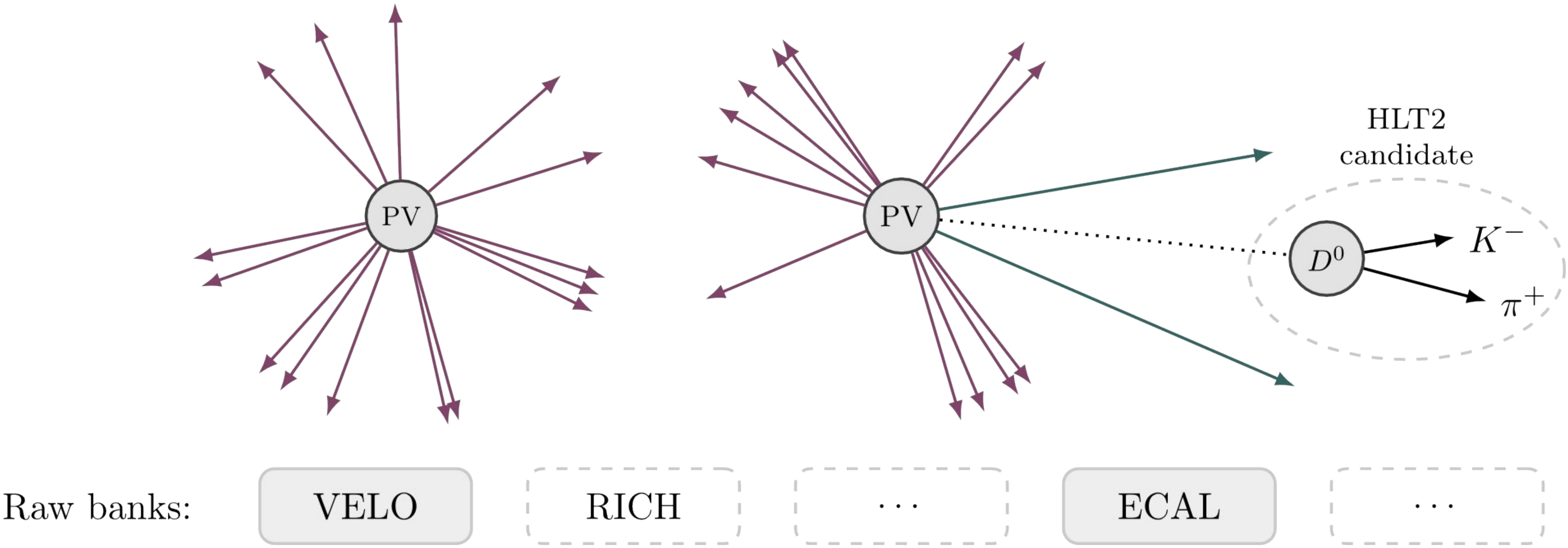
# Persistence granularity



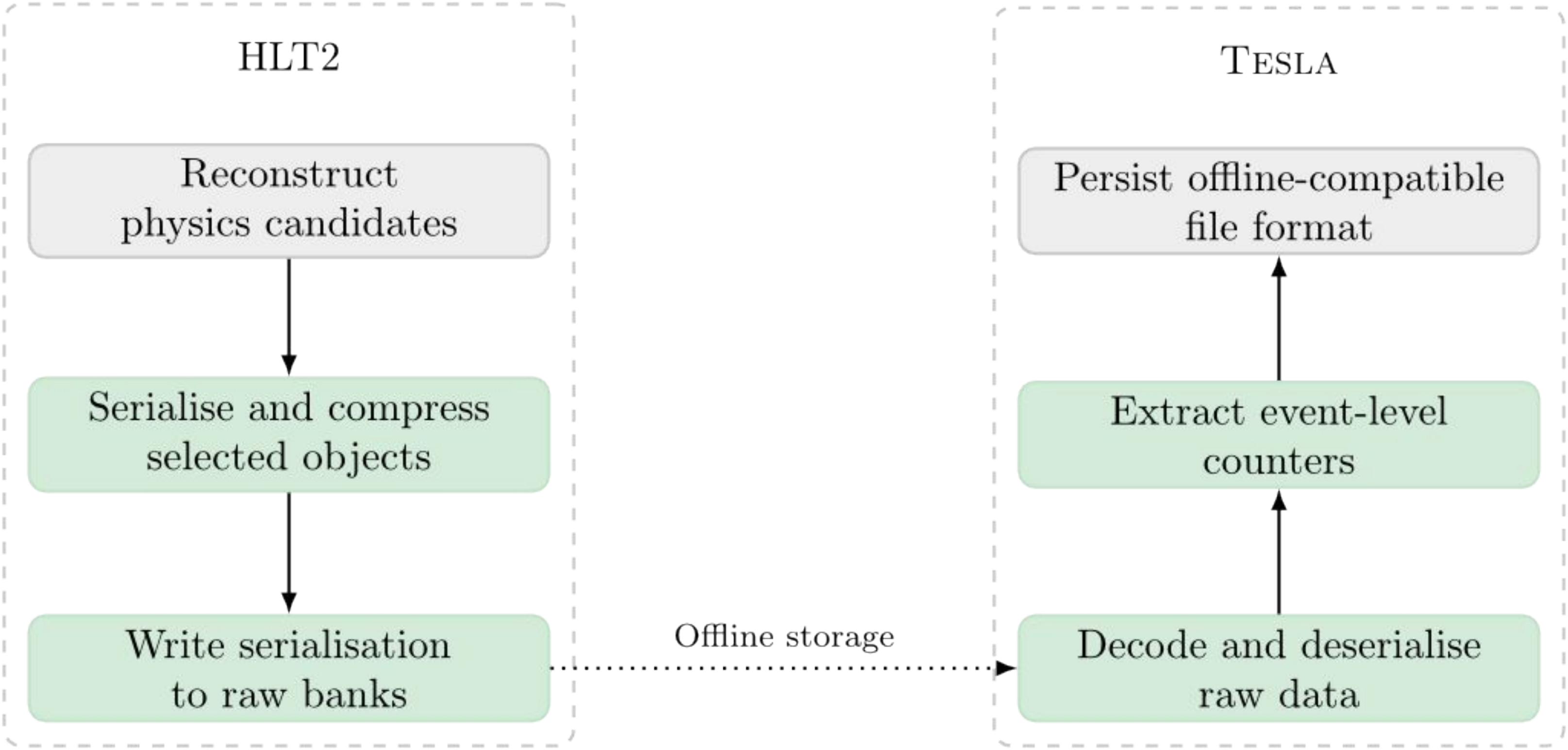
# Persistence granularity



# Persistence granularity



# Internals

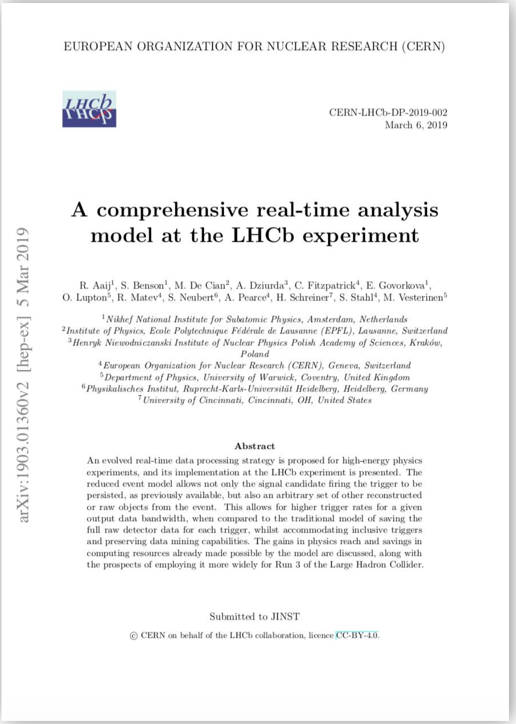

HLT2
Reconstruct physics candidates
Serialise and compress selected objects
Write serialisation to raw banks
Offline storage
TESLA
Persist offline-compatible file format
Extract event-level counters
Decode and deserialise raw data

# Rewards

Much smaller average event size
⇒ more physics within our resources

| Persistence method | Average event size (kB) |
|---|---|
| Turbo | 7 |
| Selective persistence | 16 |
| Complete persistence | 48 |
| Raw event | 69 |

Accounted for around 25% of the trigger rate in Run 2.
For 10% of the bandwidth!



EUROPEAN ORGANIZATION FOR NUCLEAR RESEARCH (CERN)

CERN-LHCb-DP-2019-002
March 6, 2019

arXiv:1903.01360v2 [hep-ex] 5 Mar 2019

**A comprehensive real-time analysis model at the LHCb experiment**

R. Aaij[1], S. Benson[1], M. De Cian[2], A. Dziurda[3], C. Fitzpatrick[4], E. Govorkova[1], O. Lupton[5], R. Matev[4], S. Neubert[6], A. Pearce[4], H. Schreiner[7], S. Stahl[4], M. Vesterinen[5]

[1] *Nikhef National Institute for Subatomic Physics, Amsterdam, Netherlands*
[2] *Institute of Physics, Ecole Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland*
[3] *Henryk Niewodniczanski Institute of Nuclear Physics Polish Academy of Sciences, Kraków, Poland*
[4] *European Organization for Nuclear Research (CERN), Geneva, Switzerland*
[5] *Department of Physics, University of Warwick, Coventry, United Kingdom*
[6] *Physikalisches Institut, Ruprecht-Karls-Universität Heidelberg, Heidelberg, Germany*
[7] *University of Cincinnati, Cincinnati, OH, United States*

**Abstract**

An evolved real-time data processing strategy is proposed for high-energy physics experiments, and its implementation at the LHCb experiment is presented. The reduced event model allows not only the signal candidate firing the trigger to be persisted, as previously available, but also an arbitrary set of other reconstructed or raw objects from the event. This allows for higher trigger rates for a given output data bandwidth, when compared to the traditional model of saving the full raw detector data for each trigger, whilst accommodating inclusive triggers and preserving data mining capabilities. The gains in physics reach and savings in computing resources already made possible by the model are discussed, along with the prospects of employing it more widely for Run 3 of the Large Hadron Collider.

Submitted to JINST

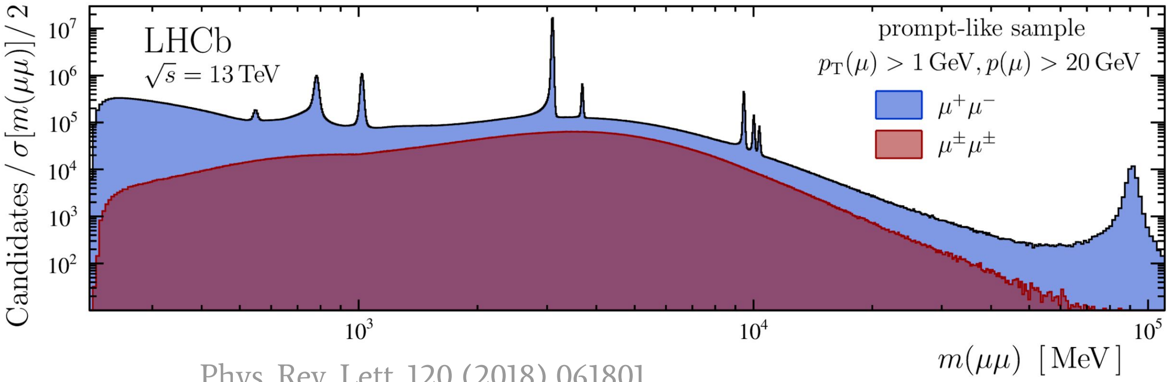
© CERN on behalf of the LHCb collaboration, licence CC-BY-4.0.

# Looking back

- Must overcome fear of losing information
- There's always room for improvement
    - Selective persistence allowed us to reduce Turbo bandwidth, then added new inclusive charm baryon lines
- Must support users in transitioning to any new features

# Looking back

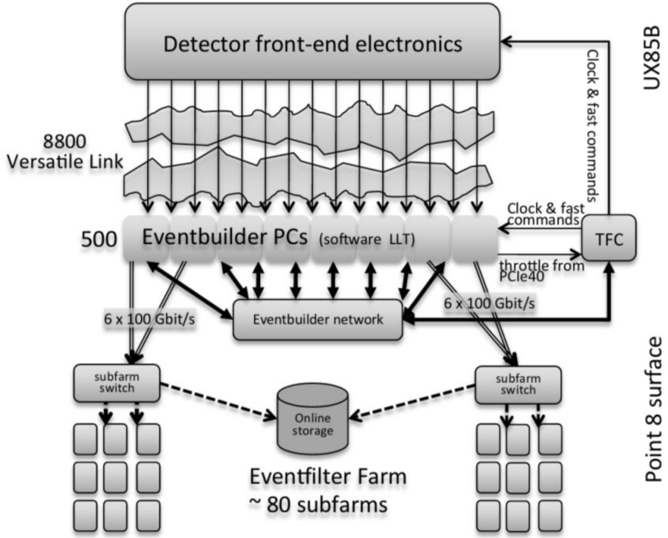
- Must overcome fear of losing information
- There's always room for improvement
  - Selective persistence allowed us to reduce Turbo bandwidth, then added new inclusive charm baryon lines
- Must support users in transitioning to any new features




Phys. Rev. Lett. 120 (2018) 061801

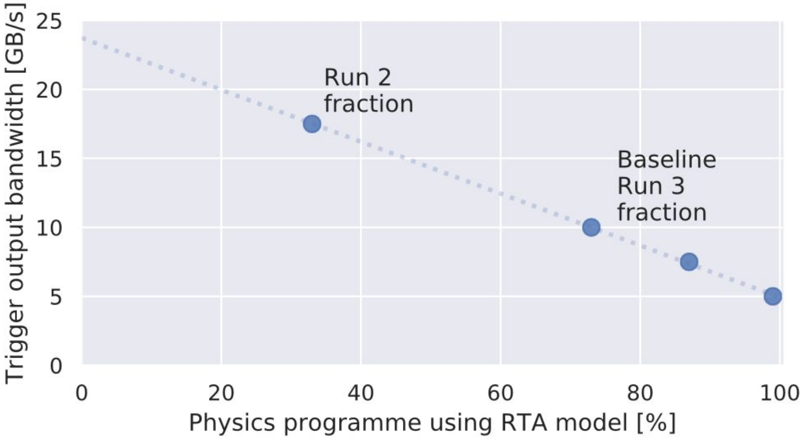
# Looking forward

- Run 3 luminosity increases 5x
- Triggerless readout, full software trigger
    - Removal of hardware trigger increases efficiency of hadronic signals >2x
    - but 4 TB/s into HLT1
- Huge increase in signal rate!




# Challenges

Run 3 physics programme is bandwidth-constrained like charm was in Run 2




- Turbo fraction must increase: baseline is 70%
- Must migrate some inclusive triggers to the RTA model
- What if we cannot achieve online/offline parity in HLT2?

# Takeaway

- Going "triggerless" helps if you have the processing power
- Align and calibrate your detector online
  - helps with improving efficiency and reducing background
- Squeeze the offline A&C and reconstruction online
  - you are sure to have the best physics objects for analysis
  - you can be much tighter on selections
- After that, it's "easy"
  - just throw away what is not necessary from the events
  - still, make sure you've convinced yourself first it's ok
  - still, make sure your QA/QC is solid as there is no going back

# References


- D. Rorh, Real-time analysis model in ALICE, HOW 2019
- W. Kalderon, Real-time analysis model in ATLAS, HOW 2019
- D. Sperka, Real-time analysis model in CMS, HOW 2019
- ATLAS collaboration, Trigger-object Level Analysis with the ATLAS detector at the Large Hadron Collider: summary and perspectives, ATL-DAQ-PUB-2017-003 (2017)
- CMS collaboration, Data Parking and Data Scouting at the CMS Experiment, CMS-DP-2012-022 (2012)
- R. Aaij et al., Performance of the LHCb trigger and full real-time reconstruction in Run 2 of the LHC, arXiv:1812.10790
- R. Aaij et al., A comprehensive real-time analysis model at the LHCb experiment, arXiv:1903.01360
- LHCb Collaboration, Computing Model of the Upgrade LHCb experiment, LHCb-TDR-018

# Example: VELO alignment

- VELO centred around the beam for each fill
  - Resolver X, Y position accuracy of 10 μm
- Kalman filter based method, minimizing the track hit residuals with PV constraints
- Automatic alignment of VELO halves in less than 10 minutes