LHCb GPU trigger commissioning with first data

Conor Fitzpatrick
On Behalf of the LHCb RTA Project

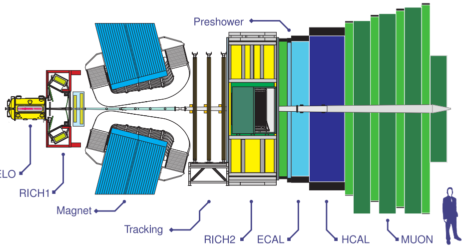
CHEP 2023, Norfolk VA

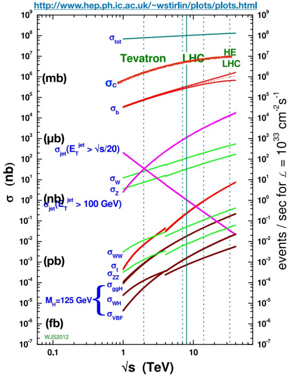
Beauty2Charm
European Research Council
Established by the European Commission

UK Research and Innovation

The LHCb detector in Run 1+2

- LHCb was built to exploit the high rates of beauty and charm at the LHC[1]:

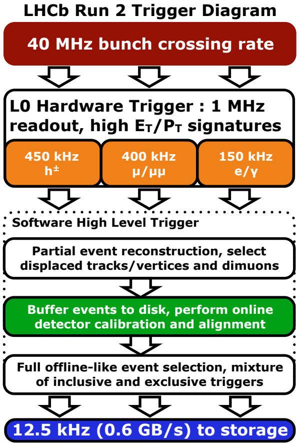




- Single arm spectrometer instrumented on $2 < \eta < 5$
  - Precise particle identification (RICH + MUON)
  - Excellent decay time resolution: $\sim$45 fs (VELO)
  - High purity + efficiency with flexible trigger and reconstruction down to low p_{T}

[1] [JINST 3 S08005 (2008)],[IJMPA 30, 1530022 (2015)]

The Run 2 LHCb Trigger

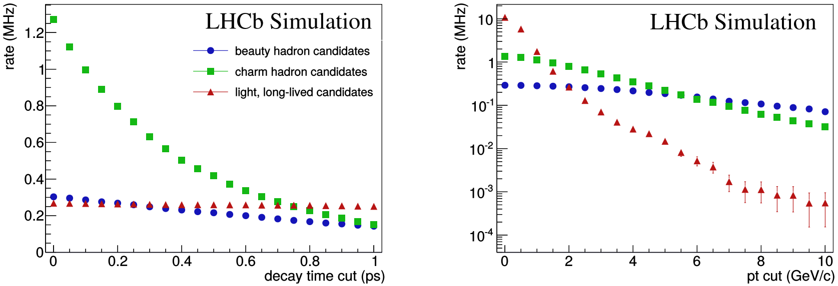


- The LHCb Run 2 trigger (2015-2019)
- Three trigger levels, with a hardware L0 stage:
  - Level-0 trigger buys time to readout the detector with Calo, Muon p_{T} thresholds: $40 \rightarrow 1$MHz
  - Events built at 1MHz, sent to HLT farm (~ 27000 physical cores)
  - HLT1 has $40\times$ more time, fast tracking followed by inclusive selections $1\text{MHz} \rightarrow 100\text{kHz}$
  - HLT2 has $400\times$ more time than L0: Full event reconstruction, inclusive + exclusive selections using whole detector
- Flexibility comes from software-centric HLT design[2]

[2] JINST 14 (2019) P04013

The MHz signal era

- For Run 3, LHCb is running at $\mathcal{L} = 2 \times 10^{33}\ \mathrm{cm}^{-2}\ \mathrm{s}^{-1}$: $5\times$ more collisions per second



- Readout becomes a bottleneck as signal rates $\rightarrow$ MHz even after simple trigger criteria [3]

[3] LHCb-PUB-2014-027

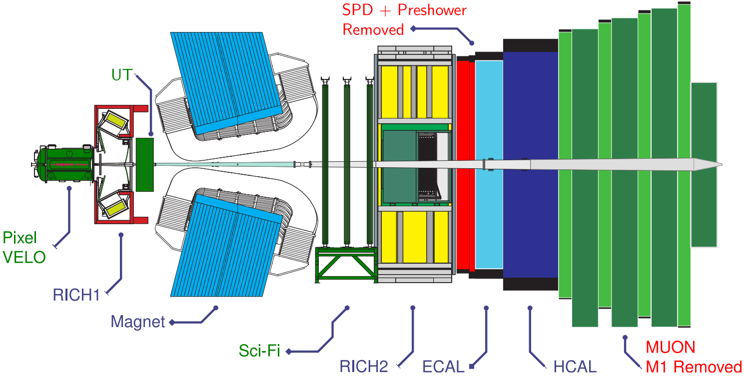
C. Fitzpatrick
May 9, 2023

So what 'stuff' can we throw away?

- The problem is no longer one of rejecting (trivial) background
- Fundamentally changes what it means to trigger

- Instead, we need to categorise different 'signals'
  - Requires access to as much of the event as possible, as early as possible
  - Solution: Drop the L0 trigger, reconstruct 30 MHz of events before making trigger decisions!

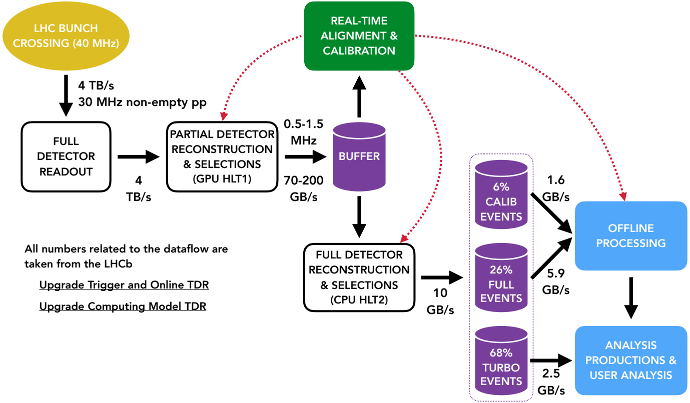
Upgrade I



- VELO has moved from r, ϕ strips to pixels: LHCb-TDR-013
- RICH replaced photon detectors, SPD, PRS, M1 removed: LHCb-TDR-014
- Trackers replaced: scintillating fibers + silicon microstrips: LHCb-TDR-015
- The readout & trigger is upgraded: LHCb-TDR-016,

Trigger & Reconstruction

- RTA: Real-Time Analysis (or Reconstruction, Trigger, Alignment)



- Builds on successful hybrid strategy for Run 2.
- In 2022 the GPU HLT1 (Allen[4]) was commissioned and took decisions for the first time

[4] [Comput Softw Big Sci 4, 7 (2020)] LHCb-TDR-021

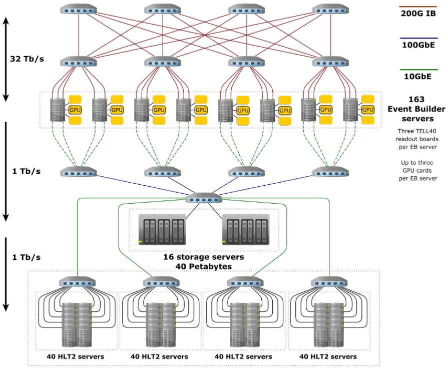
Why GPUs?

- The Allen team identified that GPUs are well suited to HEP reconstruction and trigger workloads:

Characteristics of LHCb HLT1	Characteristics of GPUs
Intrinsically parallel problem: - Run events in parallel - Reconstruct tracks in parallel	Good for - Data-intensive parallelizable applications - High throughput applications
Huge compute load	Many TFLOPS
Full data stream from all detectors is read out → no stringent latency requirements	Higher latency than CPUs, not as predictable as FPGAs
Small raw event data (~100 kB)	Connection via PCIe → limited I/O bandwidth
Small event raw data (~100 kB)	Thousands of events fit into O(10) GB of memory

- Key insight: **Using GPUs as processors instead of coprocessors avoids overheads**
- But where to place them?

LHCb Online architecture

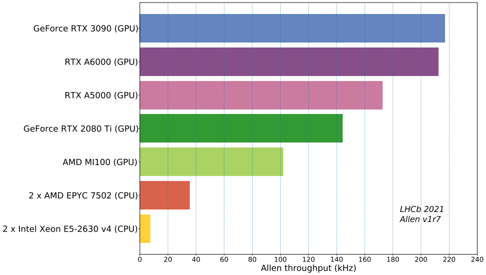


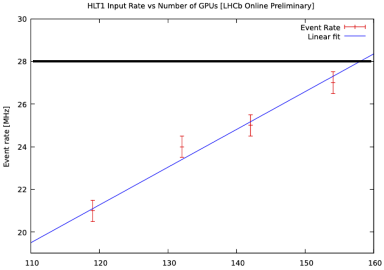
- The Online system has to build events for RTA at 30 MHz
- Commercial off-the-shelf DAQ network, event builder, buffer and event filter farm
- **Event Builder** nodes on a high-speed symmetric network
- In 2022: 40Tbits/s EB network throughput achieved, **equivalent to 4% of the internet**[5]

- 163 Event builders in total, with 3 spare PCIe slots each.
- **Reducing rate here means downstream network requirements are much lower**
- See Flavio's talk on Triggerless DAQ for more details

[5] Computer weekly, 13/09/22

Allen Performance





- The entire software HLT1 sequence has been implemented in CUDA and benchmarked on several consumer and data center GPU devices
- Allen is able to run the entire HLT1 reconstruction + trigger sequence at the LHC bunch crossing rate on 163 RTX A5000s (one per EB node)[6]

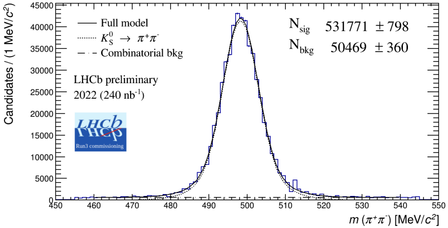
- Cost of GPUs and savings to online network mean capacity has been expanded:
- Total GPU capacity now 326 (2 per node)
- Allows margin and to expand physics scope during Run 3+4.

[6] LHCB-FIGURE-2020-014

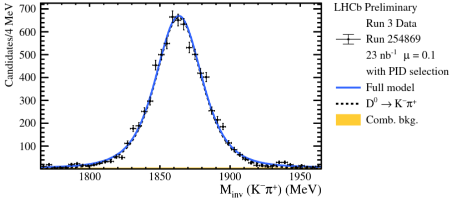
Status of commissioning

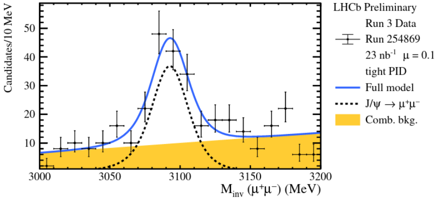
- The GPU-based HLT1 is installed and has selected signals with the first Run 3 data!

- Dedicated Ks0 trigger lines are very pure directly from HLT1
- This is a first for LHCb, enabled by Allen
- Dedicated triggers for LLPs in Jiahui's talk



- Preliminary signals[7] are very encouraging
- Expect further expansion of physics scope throughout Run 3

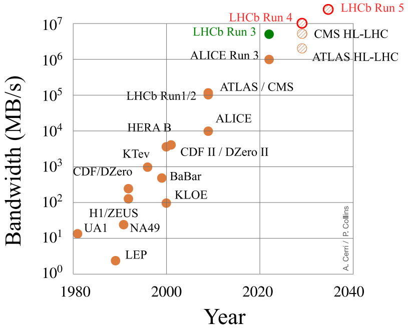




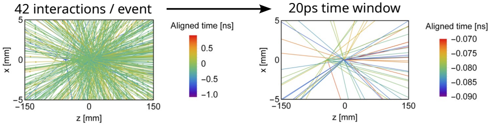
[7] [LHCb-FIGURE-2023-002], [LHCB-FIGURE-2023-005]

Upgrade 2

- Upgrade 2 planning underway for LS4 (2033-2035)
  - Potential detector consolidation in LS3 (2026)
- FTDR approved in March 22 [LHCB-TDR-023]
- Exciting challenges in trigger and DAQ:



- **4D reconstruction**: timing added to tracking to better isolate signals. Potential to add timing to hadron PID in LS3
- Potential for FPGA-based tracking: See Federico's talk

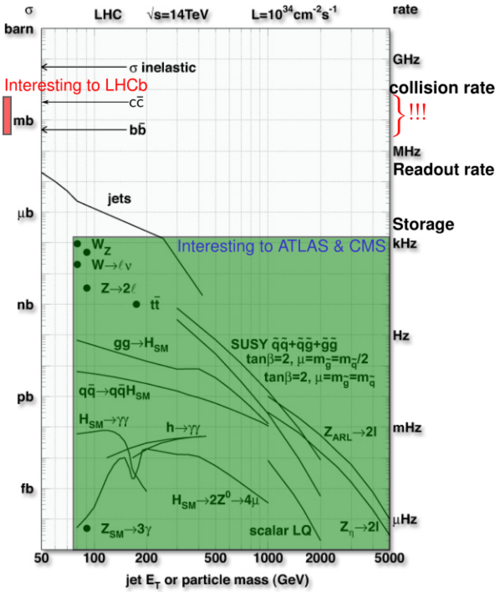


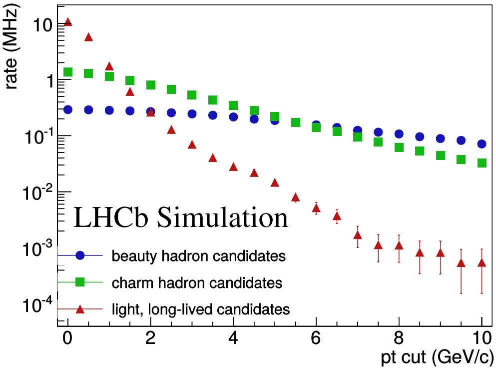
Conclusions

- LHCb has commissioned a GPU based first-level trigger and reconstruction (Allen) operating at the LHC bunch crossing rate
  - The 2022 data taking period has shown that GPUs are a cost-effective method of triggering at a hadron collider
  - Using GPUs as a complete processing solution in the Event Builder makes for a more efficient DAQ network
  - The additional capacity from 326 GPUs gives room to expand LHCb's physics scope
- Allen has taken its first steps in Run 3
- LHCb is looking forward to getting the most from this new trigger paradigm...
- ...and its expansion towards the HL-LHC era

Backups

Why read out at 30MHz?



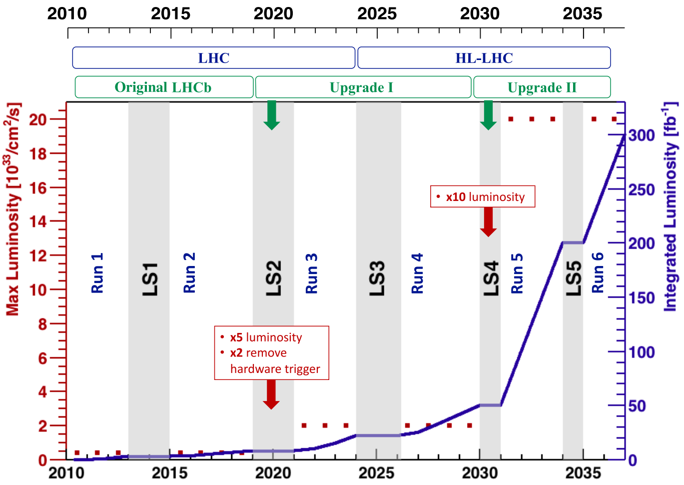


Energy efficiency

Architecture	Energy per trigger (mJ)	Gain	Total gain
E5-2630-v4 Xeon			
Before SW optimization	39.9	1.0x	
w/Physics optimizations	21.0	1.9x	1.9x
w/SIMD optimizations	8.4	2.5x	4.8x
7502 EPYC			
w/SIMD optimizations	3.2	2.6x	12.5x
Event Building Node, NR			
1 GPU	3.1	1.03x	12.9x
2 GPUs	2.4	1.29x	16.6x
3 GPUs	2.1	1.15x	19.0x
Dedicated GPU machine			
4 x 2080 Ti + 2 Network Cards	2.8	1.14x	14.3x
5 x 2080 Ti + 3 Network Cards	2.5	1.12x	16.0x
Pure GPU machine			
8 x 2080 Ti + Onboard Network	2.1	1.15x	19.0x

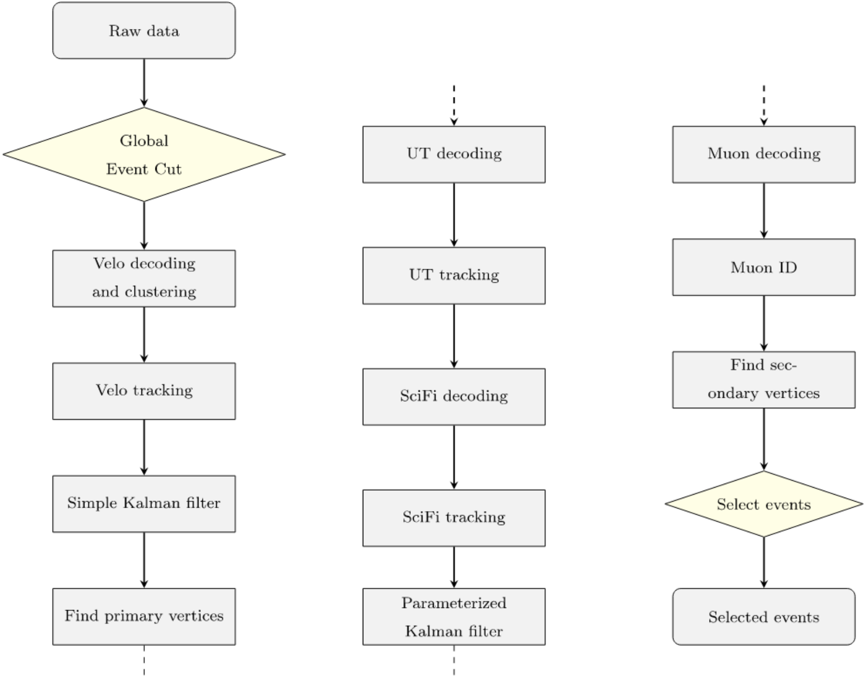
EPJ Web Conf. 251 (2021) 04009

Timeline



- LHCb collected 9fb^{-1} during Run1 + Run 2
- Upgrade I now fully installed to collect 50fb^{-1} during Run 3 + Run 4
- Upgrade II planning underway for 300fb^{-1} Run 5 onwards

HLT1 algorithms in Allen

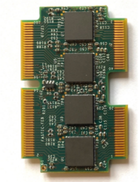


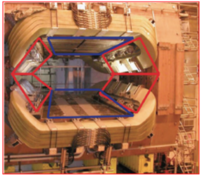
LS3 consolidation

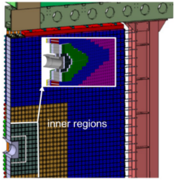
Modest consolidations with physics benefits already in Run 4 while preparing UII

driven by ageing
driven by technology
driven by physics

Detector	Proposal
SciFi consolidation	Replace inner modules (12X + 12stereo)
MAPS modules	2 layers, 1 m^2 each
Magnet Stations	full installation
RICH	new FEE electronics
ECAL	32+144 inner modules
RTA	Downstream tracking with FPGA


RICH electronics with timing


Magnet Stations



ECAL inner modules

- Consolidation & Upgrade II preparatory work
- Reused for Upgrade II
  - Costs accounted as part of Upgrade II for reused elements
- Proceed with LS3 TDRs before those for Upgrade II
  - Work already proceeding on some of these