Impact of the high-level trigger for detecting long-lived particles at LHCb

26th International Conference on Computing in High Energy and Nuclear Physics , 08-12 May 2023, Norfolk

C. Agapopoulou, L. Calefice, A. Hennequin, V. Kholoimov, L. Henry, B. Jashal, D. Mendoza, A. Oyanguren, I. Sanderswood, V. Svintozelskyi, J. Zhuo on behalf of the LHCb RTA project



Outline

- Introduction
 - LHCb detector in Run 3
 - Tracking system and track types
 - The High Level Trigger (HLT) in LHCb upgrade
 - LLPs at LHCb
- Physics case: Sensitivity studies of long-lived particles
 - LLPs in Standard Model
 - LLPs in Beyond the Standard Model
- •
- Summary

LHCb detector in Run 3

- Single-arm spectrometer in the forward direction
- Originally designed for specialised study of beauty and charm hadrons
- Excellent secondary vertex resolution
- Particle ID: calorimeters, muon systems and Ring Imaging Cherenkov (RICH) detectors
- Upgrade I (Run3):
 - Luminosity: $4x10^{32} \text{ cm}^{-2}\text{s}^{-1} \rightarrow 2x10^{33} \text{ cm}^{-2}\text{s}^{-1}$
 - Purely software-based trigger @ 30MHz





Tracking system and track types

- Three tracking stations:
 - VELO: 26 modules of pixellated hybrid silicon detectors
 - UT: 4 planes of silicon strip trackers
 - SciFi: 12 planes of scintillating fibre trackers
- Main track types for physics analysis:
 - Long: signal in VELO and SciFi (minimum) + UT (full)
 - Downstream: signal in UT and SciFi
 - T: hits only in SciFi





Run HLT in 2 stages:

- HLT1: Partially reconstructs event and reduces bandwidth from 30MHz to 1MHz
- HLT2: Fully reconstructs event and reduces bandwidth from 70-200GB/s to 10GB/s



- Long living particles (LLPs) are particles that travel significant distances before decaying into other particles:
 - Likely to decay after VELO \Rightarrow Downstream / T tracks
 - The current HLT1 only triggers long tracks ⇒ Less sensitive to LLPs searches
- We need to add dedicated LLP triggers to HLT1: Implementing the Hybrid seeding arxiv:2007.02591



Main goal:

• Add LLP reconstruction and selection to HLT1 in Run3

Progress in reconstruction

- SciFi seeding ✓ Done
- Velo-SciFi matching 🗸 Done
- Downstream In review



LLPs at LHCb

- LLPs present in SM and many BSM extensions
- Expected track types depend on LLP flight distance
- Downstream and T tracks are NOT yet considered in HLT1 decision before → target Run3





LLPs in the Standard Model

- $\Lambda_b \rightarrow \Lambda^0 [\rightarrow p^+ \pi^-] \gamma$: photon polarization
 - Non-standard right-handed current

 $(\alpha_{\gamma}) = \frac{P(\gamma_L) - P(\gamma_R)}{P(\gamma_L) + P(\gamma_R)}$ 10.1140/epjc/s10052-019-7123-7





roderic.uv.es/handle/10550/74707

9

LLPs in the Standard Model

- $K_{s^0} \rightarrow \mu^+ \mu^-$: not yet observed
 - Suppressed by SM
 - Sensitive to BSM: SUSY and Leptonquarks





LLPs in the Standard Model

• SM LLPs: $\Lambda^{\scriptscriptstyle 0}$ and $K_{s}{}^{\scriptscriptstyle 0}$

10.3389/fdata.2022.1008737

	LL	DD	ТТ	HLT1 eff (TOS)
$ ensuremath{V}^0 $	12%	51%	37%	< 10%
K _s ⁰	46%	39%	16%	< 25%





No observed in Run2 Br($\Xi_{b}^{-} \rightarrow \Xi^{-}\gamma$) < 1.3(0.6) x 10⁻⁴ at 95%(90%) CL

(PHYS. REV. D105 (2022) L051104)

11

LLPs in Beyond the Standard Model

- Higgs portal to Dark Matter: $B^+ \rightarrow K^+ H' [\rightarrow \mu^+ \mu^-]$ arXiv:1612.07818
 - The new scalar particle, H', can mix with the SM Higgs boson (H)
 - Presence of displaced vertices
 - Sensitivity depend on H' lifetime, mass and track types.







10.3389/fdata.2022.1008737

LLPs in Beyond the Standard Model

- Current HLT1 trigger efficiency (Trigger On Signal):
 - Only Long tracks are triggered by HLT1TrackMVA and HLT1TwoTrackMVA
 - Decent efficiency (30-50 %) for low lifetime
 - Poor efficiency (< 10 %) for τ > 100 ps
 - Loss in sensitivity for small H' mass
- Expect significant improvement from dedicated LLP trigger in HLT1 (DD or TT):
 - Throughput challenge in HLT1: 30MHz
 - Output rate challenge in HLT1: 1MHz



 $B^+ \rightarrow K^+ H' [\rightarrow \mu^+ \mu^-]$

LHCb advanced R&D on LLP triggers

LHCb advanced R&D on LLP triggers

- Target Run 4 and beyond
- FPGA-based:
 - Utilize pattern recognition for tracking through the RETINA algorithm cf. Federico Lazzari's talk
 - Performing SciFi tracking right after readout
 - The resulting tracks are passed on to the GPU as raw data
 - Saving time for more complex tasks



1525 (2020) 012101 (ACAT 2019)



- Dedicated High-Level Trigger 1 (HLT1) for Downstream and T tracks substantially contributes to long-lived particle (LLP) research in both Standard Model (SM) and Beyond the Standard Model (BSM) scenarios:
 - → SM: Increase the efficiency for Λ^0 and K_{s^0} in $\Lambda_b \rightarrow \Lambda^0 \gamma$ and $K_{s^0} \rightarrow \mu^+ \mu^-$.
 - → BSM: Improve sensitivity for large H' lifetimes and small H' masses in $B^+ \rightarrow K^+ H'$.

THANKS FOR LISTENING

ALL BUD MAN

- Run 1+2 trigger system consist of two parts:
 - L0: hardware trigger on either muon tag or transverse energy
 - HLT: software trigger on displaced vertices, high transverse momentum signatures
- In higher luminosity, L0 saturates fully hadronic and e⁺e⁻/γ modes
 - \rightarrow Run a software only trigger at 30MHz



Challenge in HLT1: Process data in 30MHz and reduce the output rate to 1MHz

(178 GPUs in >168kHz each) or (356 GPUs in >84kHz each)

Solution: Accelerate with GPU (RTX A5000)

- One event per CUDA block
- Parallel threads for track/hit processing
- Tricks:
 - Avoid dynamic allocation
 - Access global memory with coalescing
 - Use shared memory as user-managed data caches to store frequently-accessed data



Memory coalescing

- Accessing global memory (DRAM) is slow ⇒ The device coalesces global memory loads and stores issued by threads of a warp into as few transactions as possible to minimize the bandwidth.
- Coalescing the global memory accessing making continues thread accessing continuous memory addresses can help achieve memory coalescing in CUDA by merging the memory accesses into a single transaction.



Memory coalescing to 4 transaction

L0 Saturation



- L0 limit the rate to 1 MHz
- The hadronic and e⁺e⁻/γ modes are fully saturated:

⇒ randomly reject signal events

Shared memory



- Scratch-pad memory on each SM
- Compared to DRAM
 - 20-40x lower latency
 - ~15x higher bandwidth
- Shared with L1 cache, 128 KB/SM in A5000 (CUDA 8.6)
- Should avoid the bank conflict in shared memory

Bank conflict

- Shared memory are divided into equally-sized banks
- Different banks can be accessed simultaneously
- Multiple threads access the same bank causes Bank conflict



Standalone SciFi seeding in HLT1

- 12 layers: 6 x^{layer} + 3 u^{layer} + 3 v^{layer}
- 2.7 meters fibers: poor information in Y axis
- Tracking in two steps:







Standalone SciFi seeding in HLT1





Velo-SciFi matching

- Reconstruct the long tracks by matching Velo tracks with SciFi seeds
- The performance meets the baseline requirements for HLT1 long track reconstruction.



