Development of an HLT2 line for rare semi-tauonic B decays

Simon Calò

October 2020

Contents

1 Introduction

Over the past 60 years, particle physics has undergone one of the most rapid and surprising developments in the history of science, especially for what concerns the experimental side. Although multiple countries and groups have developed interesting experimental proposals, CERN remains at the forefront of modern particle physics research. Thanks to the current upgrade of both the LHC and the individual experiments, CERN is likely to hold this position for the coming years, especially due to the increased luminosity, which will lead to more precise measurements.

Among the many experiments that are part of CERN, LHCb remains the only one mainly focused on an extremely interesting subfield of particle physics: B-physics.

B-physics (i.e. the investigation of matter containing b quarks) represents a promising route towards answering some of the most crucial and still open questions in particle physics. Historically, research in this field started with the goal of understanding the mechanism of CP violation: still one of the greatest unsolved puzzles in fundamental physics today.

Throughout the years, the interest in B-physics has grown and, today, it is believed to be a feasible ground for new physics (NP) discoveries [3].

Specifically, multiple NP models predict a special kind of decay for the B meson that is forbidden in the SM and that could prove to be a definitive breakthrough for physics beyond the SM. These are fully leptonic B decays, which violate lepton flavour conservation [1].

According to most theories, these yet unobserved interactions should be particularly visible when studying semi-tauonic processes, due to an enhancement of NP effects when third generation leptons are involved [2] [5].

Hence, the purpose of the research here presented was to develop and efficient Hlt2 trigger line for the detection and storage of rare semi-tauonic B decays at the LHCb experiment, in run 3. By semi-tauonic B decays are meant the following modes:

- $B_s \to \tau^+\mu^-$
- $B_s \to X\tau^+\mu^-$
- $B_s \to \tau^+\tau^-$
- $B_s \to X\tau^+\tau^-$

where X represents any hadronic system of one or more particles.

Although multiple modes could have been chosen for these studies, we decided to pursue a study on the $B_s \to \tau^+\mu^-$ mode, with $\tau^+ \to \pi^+\pi^-\pi^+\bar{\nu}_\tau$. Previous research [5] had already determined an upper limit on the branching fraction of this mode by analysing run 1 and run 2 LHCb data. With the increased luminosity that LHCb is expected to achieve in run 3, it will be possible to obtain a tighter limit on this branching fraction, or, perhaps, to finally discover this NP process, leading to a revolution in particle physics.

The purpose of this study was to develop a selection for an Hlt2 line for the decay $B_s \to \tau^+\mu^-$. Such line should achieve a high retention rate for signal events, while still significantly reducing the number of background events stored. Specifically, in order to facilitate the offline analysis, the signal rate should be of a comparable magnitude to the background rate (of stored events), given an expected bunch-bunch crossing of about 30 MHz after the upgrade.

Moreover, it is important that the selection applied does not bias any of the distributions used for offline analysis, especially mass ones, which are of high importance for most studies.

2 Methods

Since the purpose of this research was to develop a trigger line for run 3, we needed to rely on MC simulation for the signal sample. Similarly, to study background, we decided to rely on minimum bias (minbias) run 3 MC samples, which were considered to be suitable simulations of standard LHC running conditions.

In order to prepare a provisional Hlt2 line, it was initially decided to build a framework in Moore: the standard framework used for Hlt lines.

Unfortunately, it soon became evident that through Moore, a series of interesting tools would not be available for our study, due to the ongoing effort of configuring Moore for run 3.

Hence, the preparation of the files was carried out mainly via DaVinci (v51r0).

The input Final State Particles (FSPs) used to reconstruct the candidates were the *StdLooseMuons* and the *StdNoPIDsPions* which already presented some loose selections for the respective tracks.

Finally, Ganga was employed to prepare the larger files, since it allowed for an efficient combination of multiple input files in parallel.

To perform studies on the simulated properties of the signal candidates, we included the *MCTupleTool* which provided this information in a separate decay chain. These properties were mainly used to categorise the tracks of the FSPs originating from signal B's (MC particles in this case) and to obtain an approximate estimate of the total number of simulated signal candidates. On these MC candidates, no selections were applied.

To ensure that when studying data distributions, only signal events were being considered, the candidates from the $B_s \to \tau^+\mu^-$ simulation were always matched. This was done by restricting the background category of the B candidate to include low-mass background only (i.e. BKGCAT = 50) and also by matching the true ID of each of the B daughters (and granddaughters).

Throughout the rest of this paper, when the expression "signal" is used, this is intended to mean matched candidates from the MC simulations of $B_s \to \tau^+\mu^-$ events.

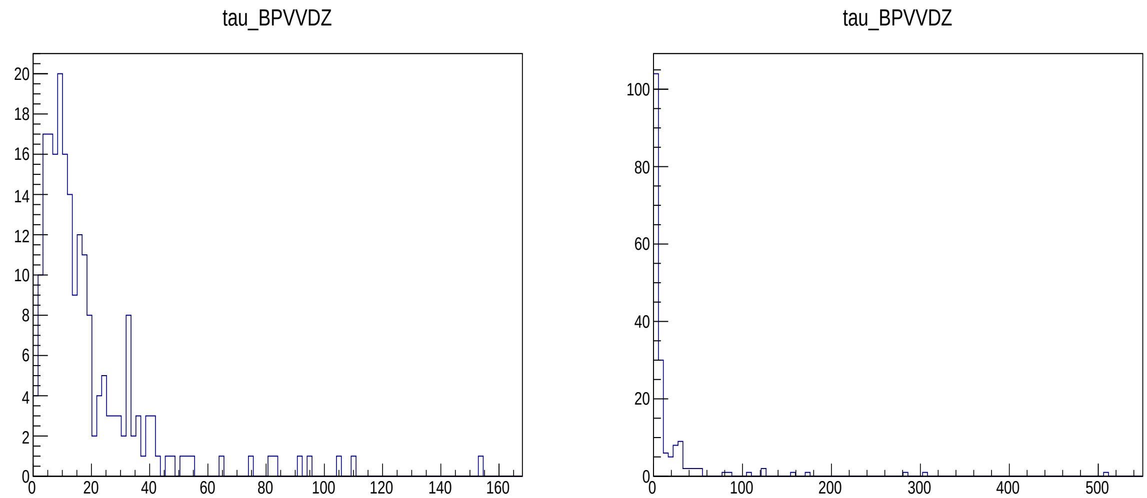
3 Results

The initial selection employed for the line was based on the one used in a previous LHCb study which focused on the same decay mode: [5].

Firstly, the distribution of the variables that were to be included in the final selection were plotted for both signal and minbias candidates, for comparison purposes.

This allowed to find suitable selections that would preserve as much signal as possible, while still causing a meaningful reduction of the background (i.e. minbias) candidates. The selection points chosen are summarised in table 1.

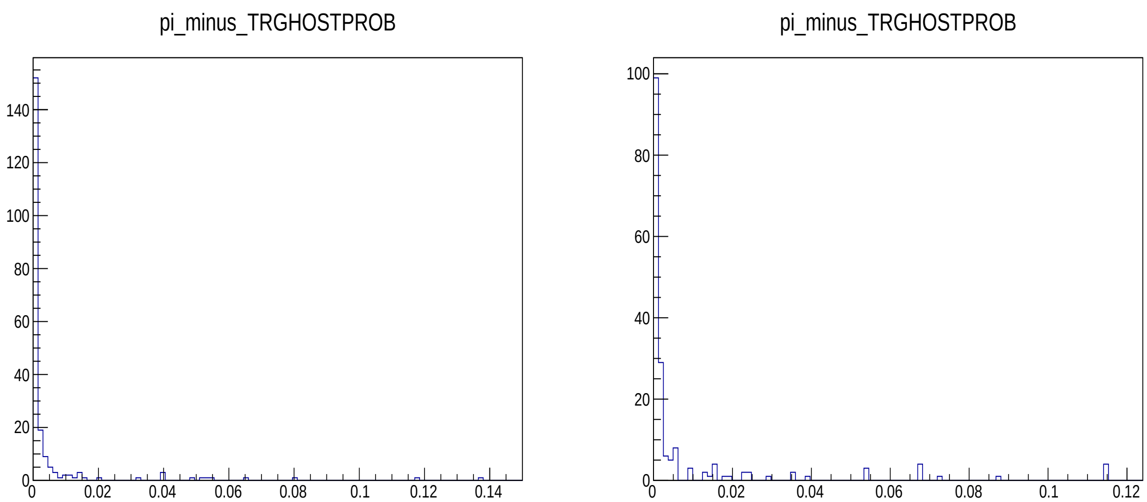
Below are some example plots displaying a variable's distribution for signal on the left, and the corresponding distribution for minbias candidates on the right.



(a) Matched signal.

(b) Minimum bias.

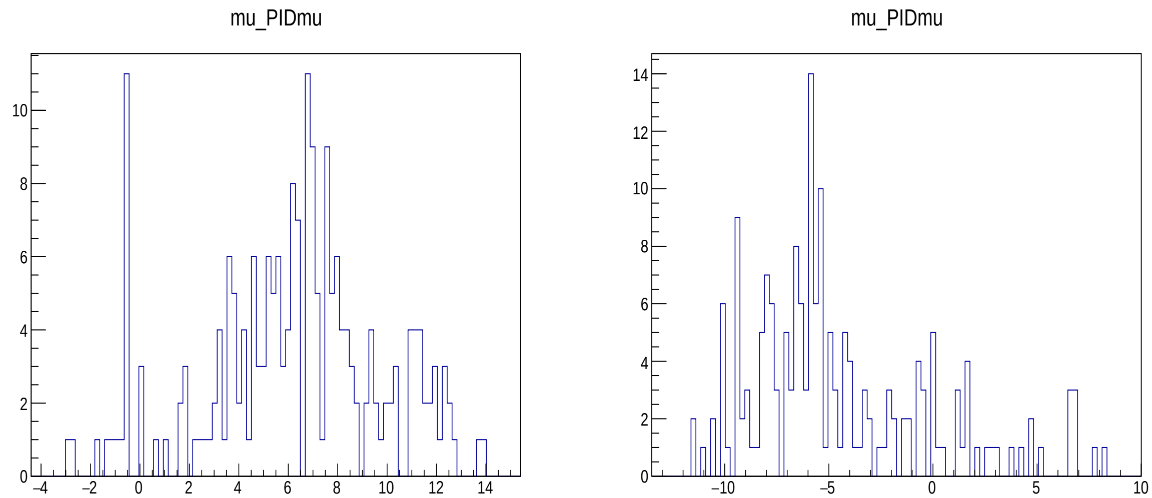
Figure 1: Comparison of τ BPVVDZ.



(a) Matched signal.

(b) Minimum bias.

Figure 2: Comparison of π^- TRGHOSTPROB.



(a) Matched signal.　　(b) Minimum bias.

Figure 3: Comparison of PID μ.

After closely analysing these and other such plots, the final selections to be applied in the Hlt2 line were chosen. The selections are summarised in the following table:

Table 1: Selection chosen for the (loose) Hlt2 line.

Particle	Cuts applied
π, μ	PT $>$ 250 MeV TRGHOSTPROB $<$ 0.25
μ	PIDmu $>$ -2 IPχ^2 $>$ 7.5
τ	400 MeV $<$ M $<$ 2100 MeV DOCAMAX $<$ 0.5 mm BPVDIRA $>$ 0.99 0.1 mm $<$ BPVVDRHO $<$ 10 mm 0.1 mm $<$ BPVVDZ $<$ 160 mm ENDVERTEXχ^2 $<$ 150 IPχ^2 $>$ 3 BPVVDχ^2 $>$ 13
B	2 GeV $<$ M $<$7 GeV BPVCORRM $<$ 10 GeV BPVVD $<$ 115 mm ENDVERTEX χ^2 $<$ 150 $(\tau PT + \mu PT) >$ 5 GeV

The above selections formed the (pseudo) Hlt2 line that we used to filter the events that were then added to the data files we analysed. The following information and efficiencies mentioned were relative to such data files. //

As previously mentioned, we decided to investigate the MC properties of the simulated FSP tracks. Out of a total of 4438 B MC candidates:

- 37 % of the candidates only presented long tracks (i.e. Reconstructible = 2) for the FSPs

- 16.6 % presented one upstream pion (i.e. Reconstructible = 4), two long pions and one long muon
- and 12.4 % displayed two long pions, one Velo (i.e. Reconstructible = 6) pion and a long muon.

We then decided to evaluate the signal efficiency for the above (loose) selections (table 1). In order to do this, we calculated the number of matched reconstructed candidates and divided it by the number of corresponding MC candidates in which all the FSPs were reconstructible as long tracks (i.e. Reconstructible = 2), as such:

$$eff = \frac{\text{n(matched candidates)}}{\text{n(MC candidates LLLL)}} = \frac{683}{1640}$$

This was done since only long tracks were used to reconstruct our candidates as these contained more refined information (i.e. higher momentum resolution). This yielded an overall efficiency of approximately 41.6% for our selection.

Another important measurement that was taken was relative to the efficiency of our selection combined with that of a sample Hlt1 line.

The Hlt1 line employed consisted of the following selection:

$$(TR\chi^2 DOF < 2.5) \quad \text{and} \quad (TRGHOSTPROB < 0.2) \quad \text{and} \quad \{(PT > 26 GeV and IP\chi^2 > 7.4)$$

$$\text{or} \quad [PT > 2 \quad \text{GeV} \quad \text{and} \quad PT < 26\text{GeV} \quad \text{and}$$

$$log(IP\chi^2) > (\frac{1}{(PT/1000-2)^2 + (1.248/26000) * (26000 - PT) + 2})]\}$$

where the PT was measured in MeV. Since this was an emulation of an Hlt1 line, the candidate passed the line by requiring that any of the four FSP tracks fulfilled the above selection.

After applying the Hlt1 line on the 683 matched candidates, only 527 remained, leading to an estimated efficiency for the combined Hlt1 and Hlt2 lines of approximately 32%.

Using these two efficiencies, we set out to estimate the order of magnitude of signal retention rates in the following way:

We estimated that a $b\bar{b}$ pair would be created in $\sim 0.2\%$ of the collisions, which happen with a frequency of approximately 7.5 per bunch crossing. This factor still needs to be scaled by the 30 MHz frequency of bunch crossing (for the upgraded LHC) and by $\frac{B_s}{B_d}$, which leads to a frequency of the order of a few hundreds kHz. Finally, this last value still needs to be scaled by the branching fractions of $B_s \to \tau^+\mu^-$ and the $\tau^+ \to \pi^+\pi^-\pi^+\bar{\nu}_\tau$ and ultimately by the signal efficiency of our line. This leads to a value in the order of tens of Hz.

Alongside this, we also performed a first-order estimate of the minbias efficiency for the Hlt2 line. This was calculated by dividing the number of minbias B candidates reconstructed after passing the Hlt2 selection (39797) by the total number of events contained in the raw input files fed into Ganga (53993474). This calculation yielded and efficiency of:

$$\frac{39797}{53993474} \approx 0.074\%$$

Similarly, the combined efficiency for the Hlt1 and Hlt2 lines on minbias was approximately 0.037%.

These efficiencies, would correspond to background retention rates of approximately

$$30\text{MHz} \cdot 0.00074 \approx 22\text{KHz}$$

for the Hlt2 line alone and

$$30\text{MHz} \cdot 0.00037 \approx 11\text{KHz}$$

for the combined Hlt1 and Hlt2 lines, since background (minbias) events occur with every bunch crossing, hence with an estimated frequency of 30 MHz.

4 Discussion and outlook

Overall, the Hlt2 line developed appeared to be of decent quality, but not yet suitable for the goal we set out to reach.

The selections remained loose, allowing for a high retention of signal events (41.6%), while also significantly reducing the number of background candidates (by more than 99.9%).

This is, in fact, the role that an Hlt2 line should have, being only a filter for events considered interesting for further analysis.

However, although the signal retention rate appeared to be satisfactory, the rejection rates achieved would not be sufficient for successful offline analyses. In fact, the background retention rate remained about three orders of magnitude greater than the signal retention rate: a difference that would make it prohibitive to conduct any form of offline analysis.

The study performed on the category (i.e. long, upstream, Velo) of the FSP tracks also yielded interesting results. It demonstrated that, although they carry less information, including tracks that are not long could prove to be a successful way to increase the signal statistics. Future analyses could perform a detailed study on the B mass resolution achieved by the inclusion of the non-long tracks in the FSP used. In such a way, these analyses could make an informed choice regarding the categories to be considered.

It should be pointed out that the research was concluded while still in the process of tightening the final selections. To achieve this goal, the possibility of employing a BDT was considered. This appears to be an interesting and valid solution, that unfortunately, we were not able to properly investigate, due to time constraints.

The tool that we deemed to be a suitable one was the FastBDT package [4]: a highly efficient BDT, widely employed in the Belle II experiment and compatible with the LHCb software.

Further studies could consider performing a thorough training of a BDT for the purpose of optimising a series of selections more efficient in both retaining signal and rejecting background.

References

[1] Jason Aebischer et al. 'B-decay discrepancies after Moriond 2019'. In: *The European Physical Journal C* 80.3 (Mar. 2020). ISSN: 1434-6052. DOI: 10.1140/epjc/s10052-020-7817-x. URL: http://dx.doi.org/10.1140/epjc/s10052-020-7817-x.

[2] Javier Fuentes-Martín et al. 'With or without U(2)? Probing non-standard flavor and helicity structures in semileptonic B decays'. In: *Physics Letters B* 800 (Jan. 2020), p. 135080. ISSN: 0370-2693. DOI: 10.1016/j.physletb.2019.135080. URL: http://dx.doi.org/10.1016/j.physletb.2019.135080.

[3] S. Jäger et al. *Charming New B-Physics*. 2019. arXiv: 1910.12924 [hep-ph].

[4] Thomas Keck. *FastBDT: A speed-optimized and cache-friendly implementation of stochastic gradient-boosted decision trees for multivariate classification*. 2016. arXiv: 1609.06119 [cs.LG].

[5] J. A. Romeu, J. Cogan and G. Mancinelli. *Search for the lepton flavour violating decays $B_s \to \tau\mu$ and $B_d \to \tau\mu$*. 2019.