# Optimizing the Stripping Procedure for LHCb


Rachel Richardson

Laurentian University
Sudbury, Ontario, Canada
*rrichardson@laurentian.ca*

Supervisors:

Dr. Mark Whitehead & Dr. Marco Bettler

CERN Summer Student Program 2017
Meyrin, Switzerland

# 1 Introduction

The Large Hadron Collider Beauty (LHCb) experiment at CERN receives a very large amount of data during the running of the LHC. The ability to deal with this information in an efficient manner is a key part of the success of the experiment. The data received following a collision goes through a hardware trigger, known as the L0 trigger, and then two software triggers, HLT1 and HLT2. The L0 trigger reduces the data rate from around 40 MHz down to about 1 MHz and then the two hardware triggers bring it down even further to 5 to 10 kHz. Once the data has been through the various triggers it is stored and waits to undergo the stripping process, something that can take several months to run. The stripping process sorts the data into streams, such as Leptonic (events that involve leptons), and then further into more specific stripping lines that are used for analysis projects. This process is illustrated in Figure 1. Each stream can have anywhere from 21 to 987 stripping lines. The various stripping lines are created by individual members of the LHCb collaboration and as such there isn't a systematic way of creating them. This implies that there is the potential for there to be significant overlap, meaning the same entry is saved multiple times between the stripping lines making this process inefficient. This is inefficient in both the CPU being used to run the process as well as the larger amount of memory needed to save the information. The purpose of this project is to determine the extent of this overlap between stripping lines and then create a way to find the lines contributing to this inefficiency so that they can be dealt with appropriately.

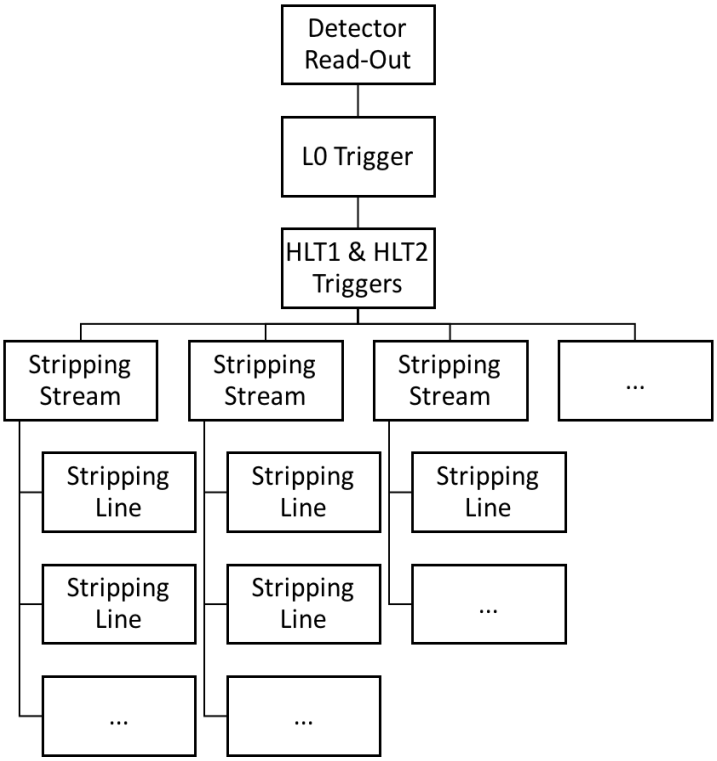



Figure 1: Diagram showing the flow of data for the LHCb experiment

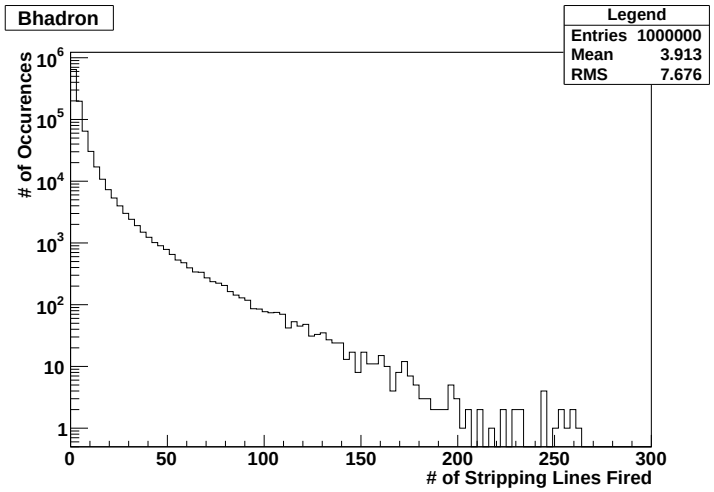



Figure 2: Plot of the Bhadron stream which has a total of 987 stripping lines. The graph shows the number of times each event is saved by a stripping line. Over 260 stripping lines are saving the same event in some cases and on average an event is being saved by approximately 4 stripping lines.

# 2 Initial Investigation

As already mentioned the first step of the project is to determine the extent of the overlap between the stripping lines in a stream. This is done by getting the number of times an event is saved by any stripping line. If the same event is saved multiple times it is assumed that there is possible overlap between the stripping lines. Two examples of the plots obtained from a sample data set[1] are shown here with the rest being located in the appendix. From the Bhadron plot, Figure 2, it can be seen that some events are saved by more than 260 stripping lines, with the average being around 4 stripping lines per event. This points to the possibility of significant overlap between lines if the same event is being used by more then one stripping line. It should be noted that the same event may be needed by multiple stripping lines, but this is something to be determined later by the experts. Again for the second plot of the Leptonic stream, Figure 3, it can be seen that up to 40 lines are saving the same event, with the average being approximately 1.5. This once again shows the potential for a problem with overlap between stripping lines.

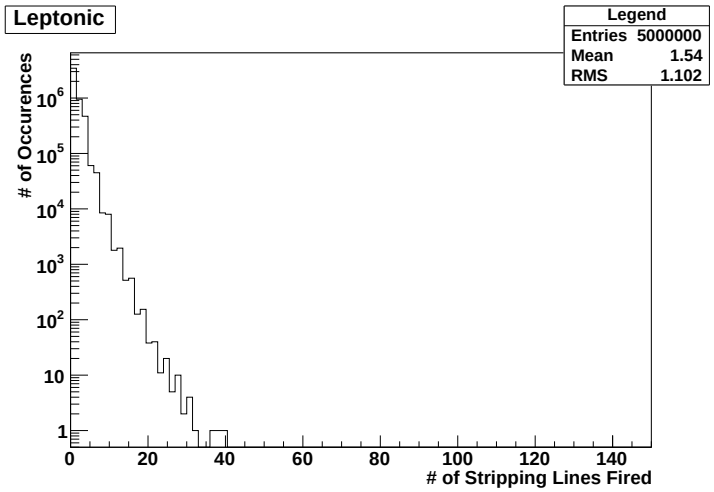



Figure 3: Plot of the Leptonic stream which has a total of 212 stripping lines. The graph shows the number of times each event is saved by a stripping line. Up to 40 stripping lines are saving the same event and on average an event is being saved by approximately 1.5 stripping lines.

[1]Data sample used: Stripping 28, September 2016. Thank you to Michael Alexander and Stefanie Reichert for providing the data sample.

# 3 Analysis

Since it is clear there is a possible overlapping issue a more detailed investigation is resumed to look at the different streams. By constructing a 2D array of events and stripping lines it is possible to compare stripping lines with each other. In doing so the goal is to find lines that are subsets of another line or lines with strong correlations. This could lead to two lines being merged or even one being removed. By systematically moving through all the pair combinations of stripping lines and comparing if an event is saved by one stripping line or the other, both or neither it is possible to see how similar two stripping lines are. If there are a lot of times that the two lines are both saving the same event it can point to a subset candidate. Other considerations in determining a subset are the total number of events saved by the line and how many times a line saves an event when the other line doesn't. Equation 1 shows the formula used to determine if two stripping lines are subsets or how close they are to being a subset. The variables are defined as follows: *bothSave* is the number of times both stripping lines save the same event and *oneSaves* is the number of times just one of the stripping lines saves the event. Depending on which stripping line is used to calculate *oneSaves* determines which is the subset of the other.

$$\%\,\text{subset} = \frac{bothSave}{bothSave + oneSaves} \times 100\% \tag{1}$$

All this information is written out to a file if the percentage match for a subset is greater than 50 percent so that it can be looked at in more detail. Figure 4 shows the first few results in the Dimuon stream. This process is then used to run over the other streams to continue looking for subsets in them.

```
Number of Entries: 5e+06
Number of Relevant Stripping Lines: 58

Number of 1s in stripping line StrippingHeavyBaryonXibzero2JpsiXistarDecision is 8218
Number of 1s in stripping line StrippingHeavyBaryonXib2JpsiXiDecision is 3607
Number of 1-1 matches between StrippingHeavyBaryonXibzero2JpsiXistarDecision and StrippingHeavyBaryonXib2JpsiXiDecision is 2222
Percentage match for StrippingHeavyBaryonXibzero2JpsiXistarDecision being a subset of StrippingHeavyBaryonXib2JpsiXiDecision: 27.0382
Percentage match for StrippingHeavyBaryonXib2JpsiXiDecision being a subset of StrippingHeavyBaryonXibzero2JpsiXistarDecision: 61.6024

Number of 1s in stripping line StrippingExoticaDisplDiMuonNoPointLineDecision is 1290331
Number of 1s in stripping line StrippingExoticaDisplDiMuonLineDecision is 298888
Number of 1-1 matches between StrippingExoticaDisplDiMuonNoPointLineDecision and StrippingExoticaDisplDiMuonLineDecision is 252829
Percentage match for StrippingExoticaDisplDiMuonNoPointLineDecision being a subset of StrippingExoticaDisplDiMuonLineDecision: 19.5941
Percentage match for StrippingExoticaDisplDiMuonLineDecision being a subset of StrippingExoticaDisplDiMuonNoPointLineDecision: 84.5899
```

Figure 4: Sample output of the script looking for subsets between stripping lines. It only prints if one of the lines is greater than a 50% subset of the other. Other information, including the names of the stripping lines, is printed so that an appropriate decision can be made about how to proceed with the stripping lines that contain overlap.

# 4 Conclusion

Overall, this detailed look at the stripping lines used in the data stripping process shows that there is room for improvement. Some stripping lines are found to be 100% subsets of others and even more are found to be greater than 90% of a subset. These results obtained will be passed on so that the appropriate people can evaluate the next steps for stripping lines with a high percentage match for a subset. This could lead to the merging of redundant lines or the complete removal of others. This script could also be used in the future for checking new stripping lines to make sure they add to the stripping process rather than make it less efficient. The next step for this stripping optimization is to look between the different stripping streams, with the possibility that there is overlap there, but this task is left for future work.

# 5 Acknowledgments

A very special thank you to my supervisors, Mark and Marco, for their support in working through this project, your help has been invaluable. Thank you as well to my supervisor in Canada, Dr. Christine Kraus, for encouraging me in my work over the years and especially in this experience. This work would not have been possible without the support of the Institute for Particle Physics (IPP) and the Natural Science and Engineering Research Council (NSERC) of Canada. Lastly, thank you to CERN and the Summer Student Program for making this fantastic program possible.

# 6 Appendix

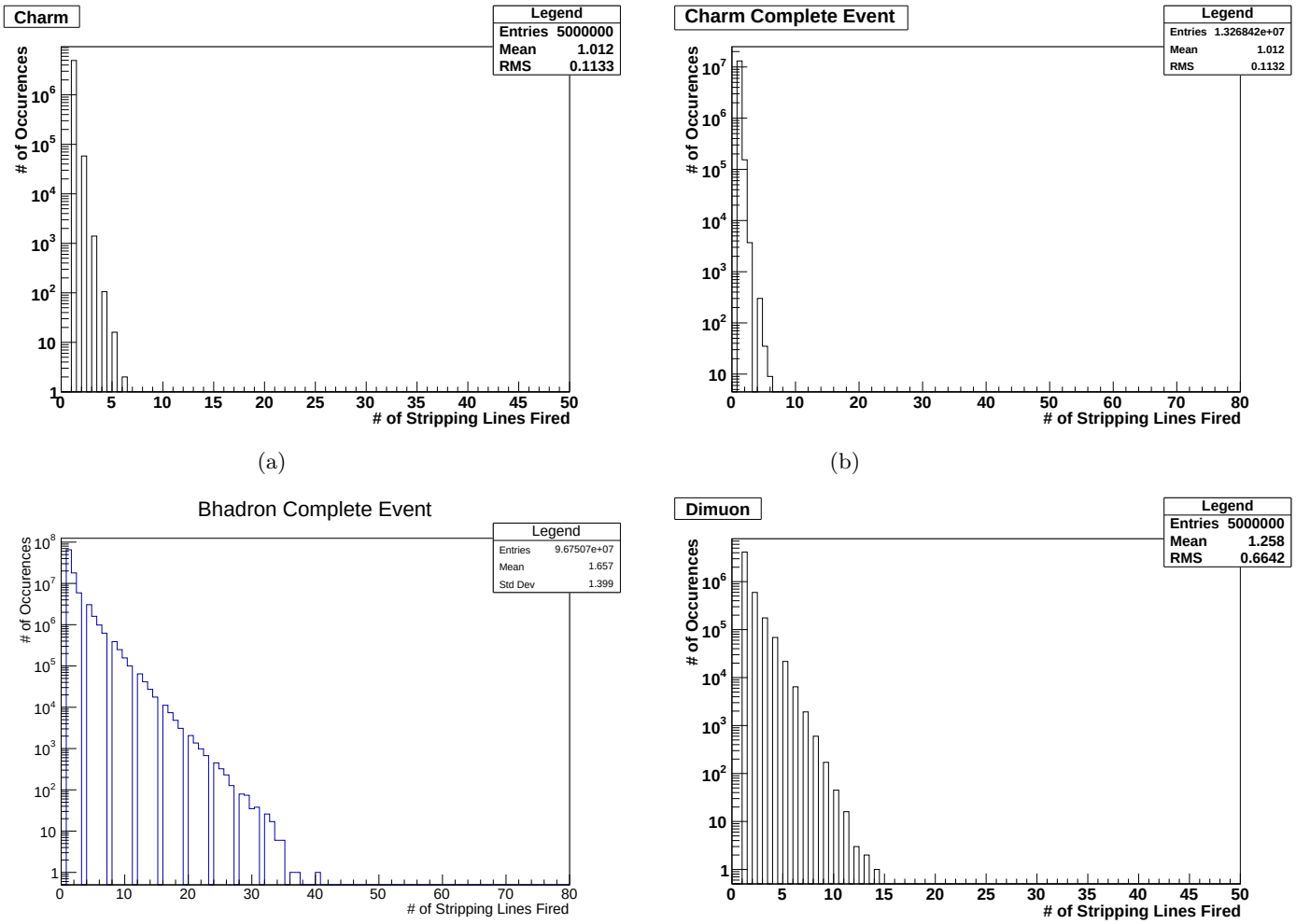
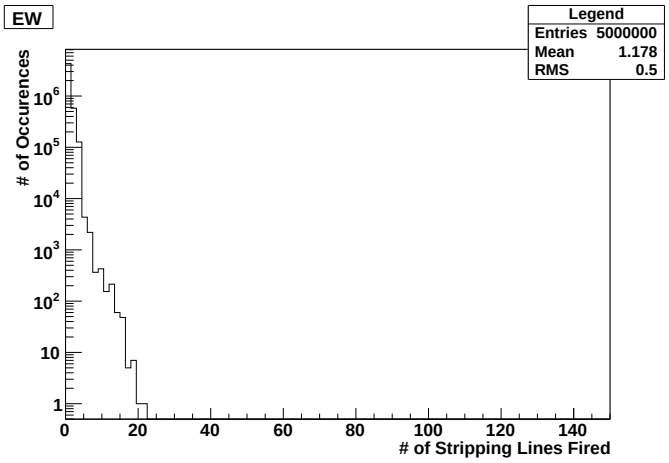
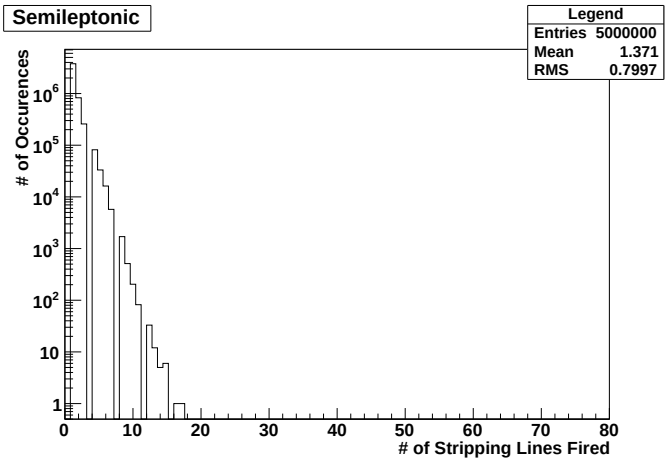


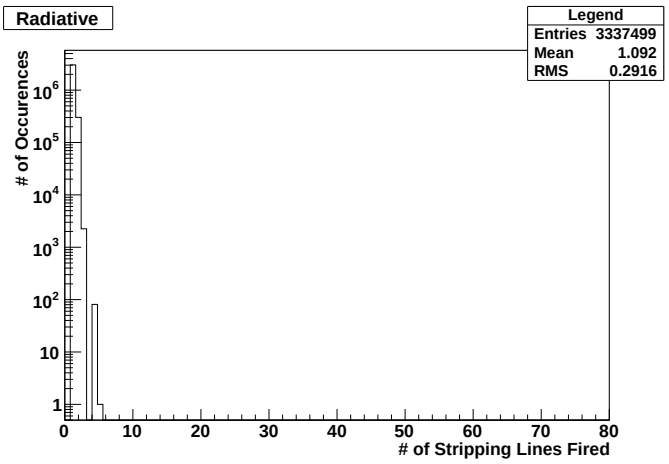

Figure 5: Plots showing the number of times each event is saved by a stripping line. (a) Charm stream, 24 stripping lines (b) Charm Complete Event stream, 23 stripping lines (c) Bhadron Complete Event stream, 89 stripping lines (d) Dimuon stream, 58 stripping lines. It should be noted that the gaps are just a binning effect




(a)




(b)




(c)

Figure 6: Plots showing the number of times each event is saved by a stripping line. (a) EW (electroweak) stream, 126 stripping lines (b) Semileptonic stream, 70 stripping lines (c) Radiative stream, 9 stripping lines. It should be noted that the gaps are just a binning effect