POSSIBLE SKIM CRITERIA FOR 2016 HPS DATA

SEBOUH PAUL

1. Introduction

One way of speeding up analyses of a dataset is to create smaller data-subsets called skims, containing only the events with some specific property, whether it be kinematics or a particular type of trigger, and then run the analysis on the skimmed dataset. This reduces the amount of time spent copying files from the tape to disk for large-scale analyses, since the file sizes are smaller, as well as the amount of time spent reading the files themselves. If the skimmed files are small enough, all of the files in a given run for a given skim may be tarred together before being placed on the tape, to reduce the total number of read/write operations on the tape.

For a skim to be useful, its criteria must not be too loose (or else it would not significantly speed up the analyses done on it), nor too tight (or else it will be cutting out too many good events). The latter was the case for the 2015 dataset's v0 skim, which we did not end up using for our bump-hunt nor vertexing analyses. To avoid making the skims too complicated, a general rule of thumb is that no criterion should be used in a skim unless it significantly reduces the file size, even if it is a cut that everyone is using.

Traditionally, we have been using 5 skims for HPS data: trident/A' candidates (v0), Moller scattering candidates (moller), full-energy electron (fee), single-0 trigger (s0) and pulser trigger (pulser). The s0 and pulser skims require only that that events have a single-0 trigger or a pulser trigger, respectively. Originally, the fee skim criteria required an Ecal cluster with certain properties (seed hit energy greater than some threshold, total energy greater than some other threshold). This was useful for calibrating the Ecal gains, but it was not very helpful for an analysis comparing measured FEE cross-sections to their predicted values, since it had a cut on the amount of energy in the seed hit of the cluster. For the reconstruction pass0 of the 2016 dataset, the fee skim was changed to being simply a single-1 trigger skim.

The moller and v0 skims for the 2015 datasets had complicated cuts on the kinematics and removed many good events, and thus we abandoned using these skims. In the next two sections, I present several different versions of the skims that can be used for the moller and v0 skims for the 2016 pass1. I leave it up to the readers of this document to decide which version(s) of the moller and v0 skims we should end up using.

2. Moller skim

The Pair0 trigger was designed specifically to find Moller events, but the majority of the events that had this trigger did not have a pair of electron tracks on opposite sides of the

detector, and therefore were not useful events for studying Mollers. Our pass0 Moller skim therefore required both the Pair0 trigger and the presence of a pair of electron tracks on opposite sides of the detector. This left us with a file size that is 4.27% of the original lcio file size. We have three categories of options for the pass1 moller skim:

- keep the same skim criteria as before
- remove the Pair0 trigger criterion and use other criteria instead of it.
- keep the Pair0 trigger criterion, but add other criteria in addition to it.

We see from Monte-Carlo in Figures 1 and 2 that if one or both moller electrons hit the Ecal, they generally hit a specific part of the Ecal. That is, their positions in x are almost always between -160 mm and -30 mm. Additionally, if both hit the Ecal, then the sum of their positions in x is usually greater than -190 mm. Therefore, if we use the cluster positions to make cuts on the events used in the skims, my recommendation would be to use $x_{1,2} < -30$ mm and $x_1 + x_2 > -190$ mm

Additionally, one may cut on momenta, either of individual particles (ie, remove FEEs) or the sum of the momenta of the pair. These cuts could be similar to those used in the tuple maker, which has a loose FEE cut at 90% of the beam energy and momentum sum cuts at 70% and 130% of the beam energy.

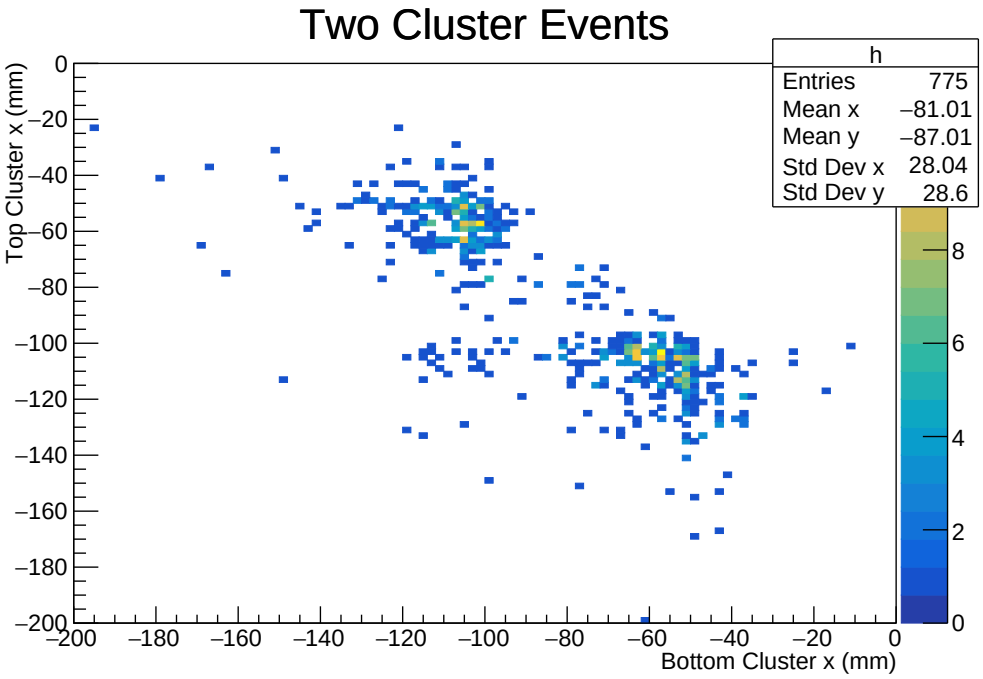

FIGURE 1. Positions of pairs of clusters in Moller monte-carlo. Events with large momentum loss ($p_{sum} < 2.0$ GeV) are excluded from this plot.

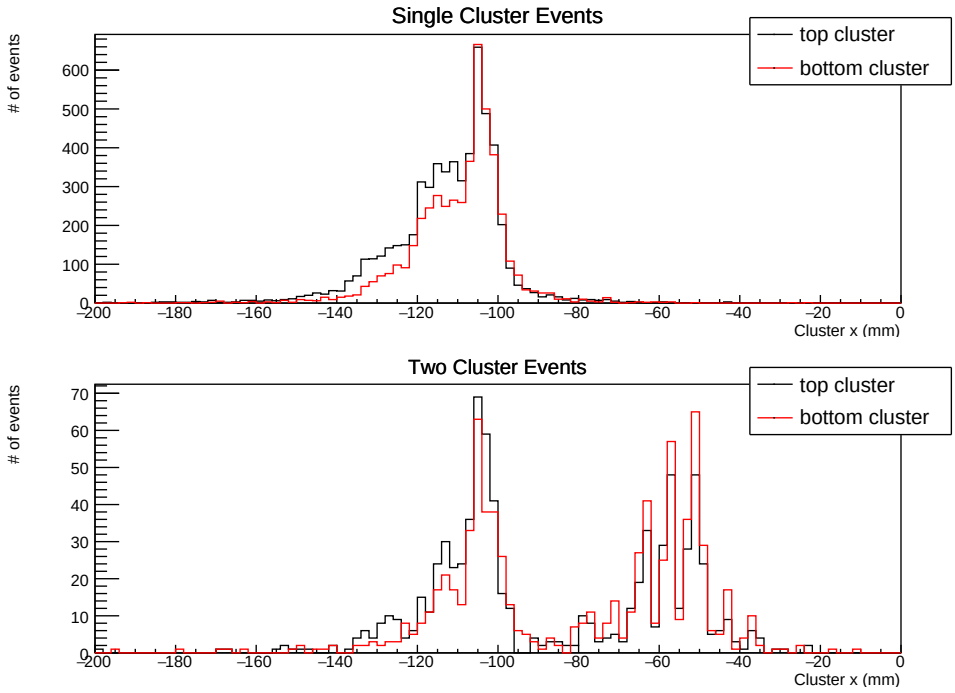

FIGURE 2. Positions of individual clusters in Moller monte-carlo. The top plot shows events with only one cluster, whereas the bottom plot shows the events with two clusters (one on top and the other on the bottom). Events with large momentum loss ($p_{sum} < 2.0$ GeV) are excluded from this plot.

3. V0 SKIM

the simplest version of the v0 skim is to require at least one member of the v0 candidates collection (that is any electron positron pair). However, if this is too loose, a tighter version of the skim would require that each of the two particles be matched to a cluster with goodness of track-cluster matching n_σ less than some threshold (default in recon is 30), and that both particles be on opposite sides of the detector. We could also cut on the two track fit χ^2ss. There are a few cuts that I intentionally decided not to use in this skim:

- momentum sum (radiative cut): If we want to compare the spectrum of summed momenta from data to monte-carlo, it would be useful to be able to do so using the full available range of momentum sums.
- paired cluster-time difference: The events with time-mismatched pairs of cluster are useful for estimating the fraction of events that are mismatched background within the same beam-bucket. Making a cut on this variable would make the skim useless for making this sort of analysis.

- WAB reduction cuts: It would be useful to be able to use this skim to analyze both converted WABs and tridents.

4. Additional Skims

In addition to the traditional skims, we could also implement a few of the following types of skims, which range in plausibility from not-so-crazy to insane. If anyone wants me to implement them, let me know.

- Not so-crazy ideas:
  - three prong events.
  - high zed vertex events.
  - converted WABs (v0 with additional cuts)
  - unconverted WABs ($e^-\gamma$, maybe with additional cuts)
- more complicated physics (just for fun):
  - double positron: ($A' \to e^+e^-e^+e^-$?)
  - unmatched tracks that extrapolate to the Ecal, but the cluster is missing: would be useful to get an estimate on the muon rate
  - v0 skim with cuts on mass at $211\text{MeV} \pm 5\sigma_m$ where 211 MeV is the theoretical mass of true-muonium (2 muon masses, minus a little bit of binding energy).

5. Results

To test the various versions of the skims, I ran the skims on the blinded 10% of run # 8098, and calculated the ratio of the average size of the skimmed files to that of the original reconstructed files, which I present in Table 5 . The abbreviations used in the table for the various cuts on mollers are as follows:

- moller cand. = Moller candidate: an electron track on the top half of the SVT and another on the bottom half.
- FEE = Full Energy Electron cut: both electrons have less than 1.9 GeV momentum
- p_{sum} = momentum sum cut: sum of particle momenta is between 1.5 and 3.0 GeV.
- EcalX: Two clusters in the Ecal (one top, one bottom). Each of them has their x coordinate less than -30 mm, and the sum of the x coordinates is greater than -190 mm.

6. Conclusions

As I mentioned, I allow the reader to draw their own conclusions regarding what type of cuts to use in the

skim type	average file size (MB)	% of original
original	6429	100
single1 (fee)	463	7.2
single0	190	2.8
pulser	182	4.3
pair0	1071	16.7
pair0; moller cand.	275	4.3
pair0; moller cand.; EcalX	45.6	.71
pair0; moller cand.; EcalX; FEE	30.1	.47
pair0; moller cand.; EcalX; FEE; Psum	27.7	.43
pair0; moller cand.; FEE	153	2.4
pair0; moller cand.; FEE; Psum	108	1.68
moller cand.	734	11.4
moller cand.; FEE	394	6.1
moller cand.; FEE; Psum	281	4.4
e^+e^-	543	8.4
e^+e^-; clusters on top/bottom	441	6.9
e^+e^-; track fit $\chi^2 < 50$	431	6.7
e^+e^-; clusters on top/bottom; track fit $\chi^2 < 50$	396	6.2

TABLE 1. Sizes of various skims. The abbreviations for the Moller cuts are described in the text.