COMET Phase-I Hit Selection for Helix Fitting

Jingyuan Gao
Osaka University Kuno Lab

September 16, 2018

Abstract:

An hit selection algorithm was developed for the data preparation before the hit results proceed to helix fitting and GENFIT, since 28.0% of the charged particle trajectories recorded have multiple turns. Three patterns were identified for the incident hits, while the outgoing hits could be mostly classified into four pattern categories, with three of the patterns the same as the incident side. The percentage of events with maximum purity above 80% for both the incident and outgoing side have been calculated to be 92.9% and 84.3% respectively. The results indicate that the patterns found can be used as the hit result for helix fitting. The purity of each layer for both sides were also investigated for finding more specific, layer dependent patterns to improve purity.

1 Introduction

For COMET Phase-I, the software preparations for helix fitting and GENFIT are examined using the events generated by GENT4. For normal high energy physics experiments, the hit fitting procedure is straight forward; feed the single turn hit results into helix fitting to obtain the initial values, then transfer all information to GENFIT for a more rigorous track fitting. However, since 28.0% of the COMET events, out of all the simulated events with hits, have recorded charged particles trajectories with multiple turns, the hit results cannot be passed on to helix fitting immediately. Instead, hit selection algorithms have to be developed for the trajectory turn separation.

The hit selection algorithm used in this project was developed by the author with the aid of Chen Wu, through manual pattern spotting by looking at events with multiple turns. First, three patterns were developed for the separation of the first turns on the incident side of the hit map. The purities of the three patterns, which are generally defined as the number of correctly selected hits over total number of selected hits, were analysed with the plots generated. Towards the end of the project, the three patterns for the incident side were also used for the outgoing hit selection with the addition of one extra pattern. The total purity for each of the four patterns were studied, as well as the purity for each layer, in the hope of discovering a more detailed pattern on the layer bases. Lastly, the relationship between the maximum purity out of all patterns and the maximum layer ID were explored briefly.

This project also serves as a cross-check for the hit selection algorithm developed at IHEP. Other algorithms have been under development for pattern spotting, including using neural networks method.

2 Method

2.1 General

The data set used in this project, *ana.signal_100K.root*, was provided by Chen Wu, using his framework conversion algorithm. This set of hit results was generated by simulating electrons with 105 MeV of energy, with the beam noise (background) excluded from the simulation.

Hit selection for the incident side was initially considered, since Tianyu Xing had shown with his helix fitting code that the fit result was the most accurate when only the incident data used as the input.

Using manual pattern recognition method, three major patterns were identified. In addition, a special scenario of the hit results was recognised. Figure 1, 2 and 3 exhibits the key features for the incident first, second and third pattern respectively. As one can see from the figures, pattern 1 has a simple feature with all the first turn incident hits on the right most of every layer; pattern 2 includes all events with the incident first turn hits on the left side of the right most cluster for even layers, and at the right most of the layer for odd layers; while pattern 3 and 2 are the exact opposite.

An example for the special scenario can be found in Figure 4. This event display shows a clear difference from the previous displays, with the incident hits recoded with smaller wire ID comparing to the exit ones, thus being on the left side of the display. This is caused by the labelling system for the wire ID. As the Cylindrical Drift Chamber (CDC) is of a circular cross section, the start and the end of the wire ID number are adjacent.

Towards the end of the project, hit selection algorithm was also extended to cover the patterns revealed from the outgoing hits. For the exit hits, it could be roughly categorised into four patterns, with three of them identical to the incident patterns and are named following the same order. The additional pattern is the exact opposite to pattern 1, with all the exit first turn hits being on the left most of each layer, in other words, they have the smallest wire ID in each layer.

Purity of the selected results was calculated with the aid of the truth information obtained from the simulation. The selected hits' turn ID were checked individually, and the ones which equal to 1, i.e. the first turn, are recorded as a correct selection. The overall purity for each pattern was calculated by

$$\text{purity} = \frac{\text{Number of correctly selected hits}}{\text{Total number of selected hits}}. \tag{1}$$

The result of the hit selection algorithm was evaluated by the calculated percentage of multi-turn events with maximum purity over 80%, where the maximum purity is defined as the largest purity value out of all patterns.

The efficiency of the selection was not explicitly calculated in this project.

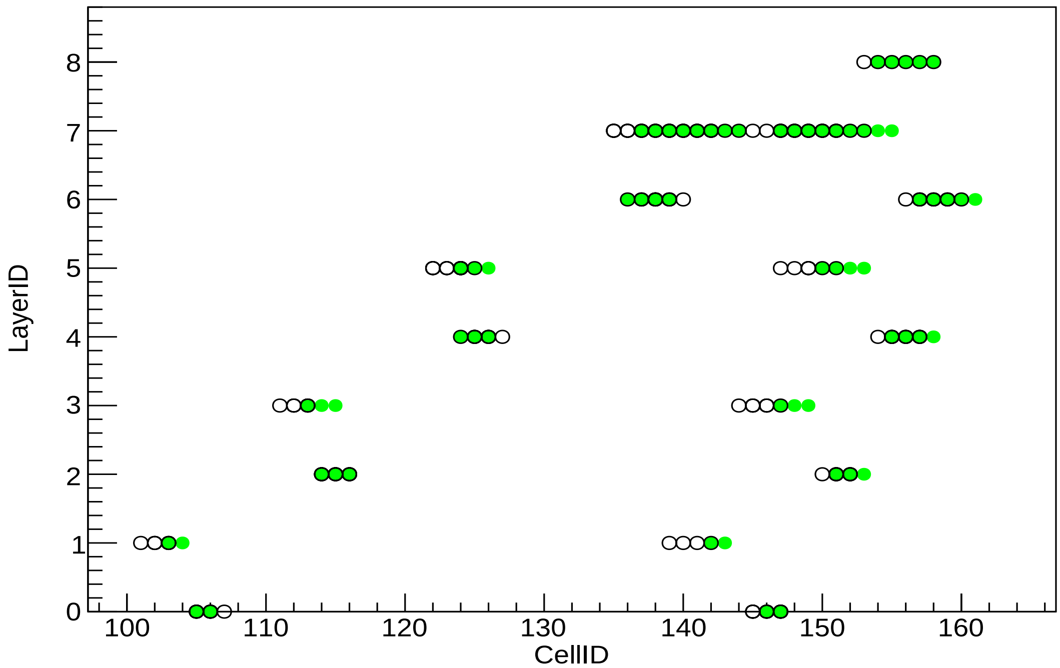

Figure 1: The incident hits of Event 24 clearly exhibits a feature of the first pattern, which has all the first turn hits at the right most of every layer. In terms of wire ID, for pattern 1, all the first turn hits have the largest wire ID within each layer.

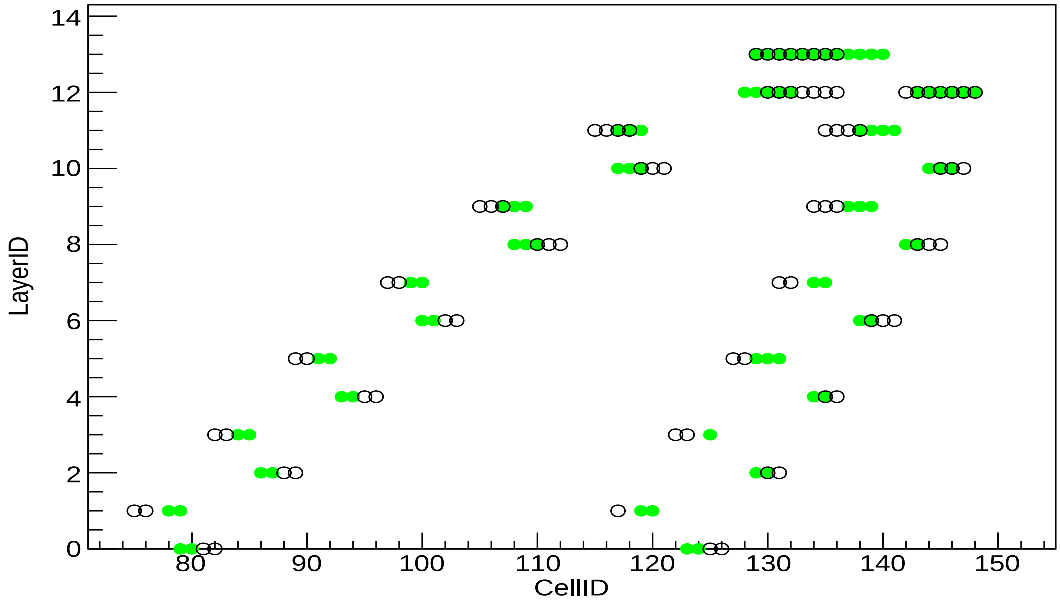

Figure 2: The incident hits of Event 40 clearly exhibits a feature of the incident second pattern, which has all the first turn hits on the left side of the right most cluster for even layers, and at the right most of the layer for odd layers. In addition, the exit hits illustrates a feature of the outgoing pattern 2.

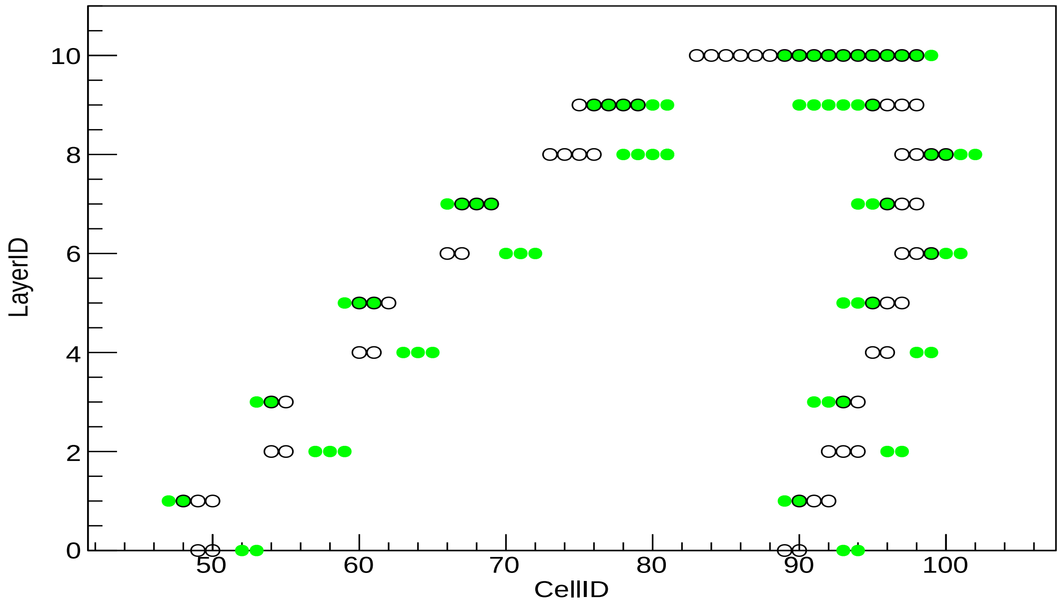

Figure 3: The incident hits of Event 36 clearly exhibits a feature of the third pattern, which has all the first turn hits at the right most of the layer for even layers, and on the left side of the right most cluster for odd layers. In addition, the exit hits illustrates a feature of the outgoing pattern 3.

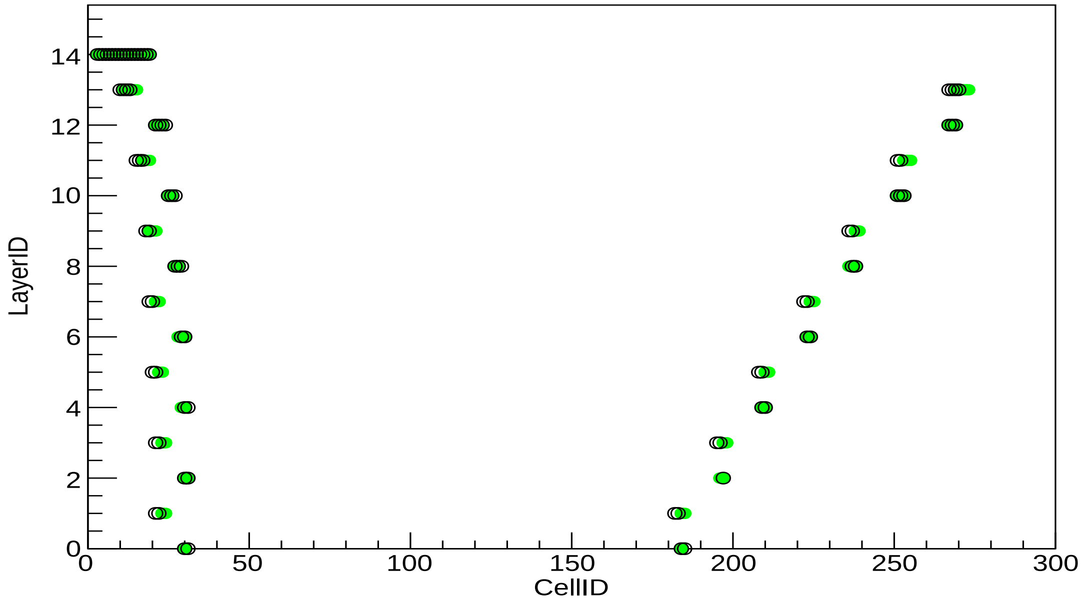

Figure 4: Event 82 has a clear feature which is classified as the special scenario. The trajectory enters the Cylindrical Drift Chamber from a special position such that the recorded wire ID for the incident hits are smaller than that of the exit ones.

2.2 Algorithm

The code for hit selection first excludes all the event with no hits recorded, recorded turn ID less than 2 or the maximum layer ID smaller than 7. The third condition was imposed since the helix fitting algorithm only functions properly for events with hits scattered across more than 8 layers. Then, in a loop going through the entire ordered hit list, the algorithm checks the layer ID of every hit to pick out the right most hit of each layer, which is used as the upper bound for selection in every layer. At this step, it also takes the special scenario into account by checking the wire ID of the adjacent hits. If the difference between wire IDs is bigger than 120, it is said that the event is a special case.

The 120 threshold was set based on a plot of the distribution of maximum distances between the left most and right most hit's wire ID, as shown in Figure 5. The graph clearly illustrated that the maximum wire ID distance clusters around two values, one centred around 45 and the other peaks at 160. By deduction, the bigger value cluster represents the special scenario; and 120 threshold was chosen since there are no tails from either distance distribution.

After the upper bound is found, the hits with adjacent indices were selected if the pattern requires to select the right most hits. For some events, an individual detection wire would detect hits from multiple turns. Hence, if hits have the same wire and layer ID, their drift distances are compared and the hit with the smallest drift distance would be selected.

To select the left most hits of the cluster, the hit list is looped back starting from the upper bound found earlier. The lower bound for the cluster is identified using a check for the difference between the adjacent hits. If the difference is bigger than 1, that very hit is the end of the cluster. After the lower bound is found the selection logic is the same as selecting from right.

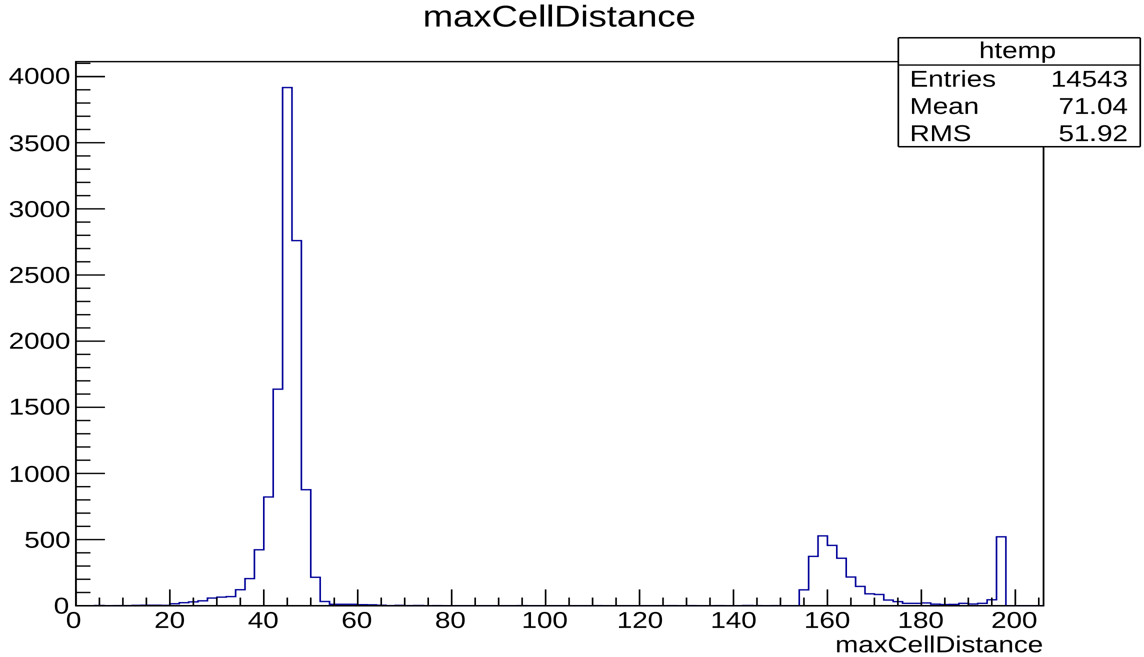

Figure 5: The distribution of maximum distance between the left most and right most hits' wire ID.

3 Results and Discussions

For incident side, the percentage of events with purity over 80% for the three patterns are 62.6%, 24.1%, and 28.4% respectively, with an overall percentage found with maximum purity being 92.9%. Meanwhile, on the exit side, the individual percentages of events for the four patterns are 45.3%, 28.3%, 27.3% and 8.3%, which leads to 84.3% of the event have max purity above 80%. Figure 6 and 7 illustrate the event maximum purity distribution. It is clear that most of the multi-turn events have max purity being 100%, which is consistent with the numerical results.

The above result is a promising find, as it is assumed that the helix fitting is able to exclude some of the false selections to an extend (purity above 80%), the fitted result is most likely to be consistent with the truth.

There exists a small amount of events with strange hit distributions, the exact number was found by comparing the number of events with turn ID greater than 1 form the source file and the number of events undergone hit selection. Some of the strange distributions are caused by the secondary particles generated in the detector.

From the calculated results, It is obvious that the performance of the outgoing hit selection is not as good as that for the incident despite having one more pattern. This was investigated by plotting the purity of each pattern for individual layers. The purity of a layer is 1 if all the selections in the layer are correct, 0.5 if one out of two is correct,

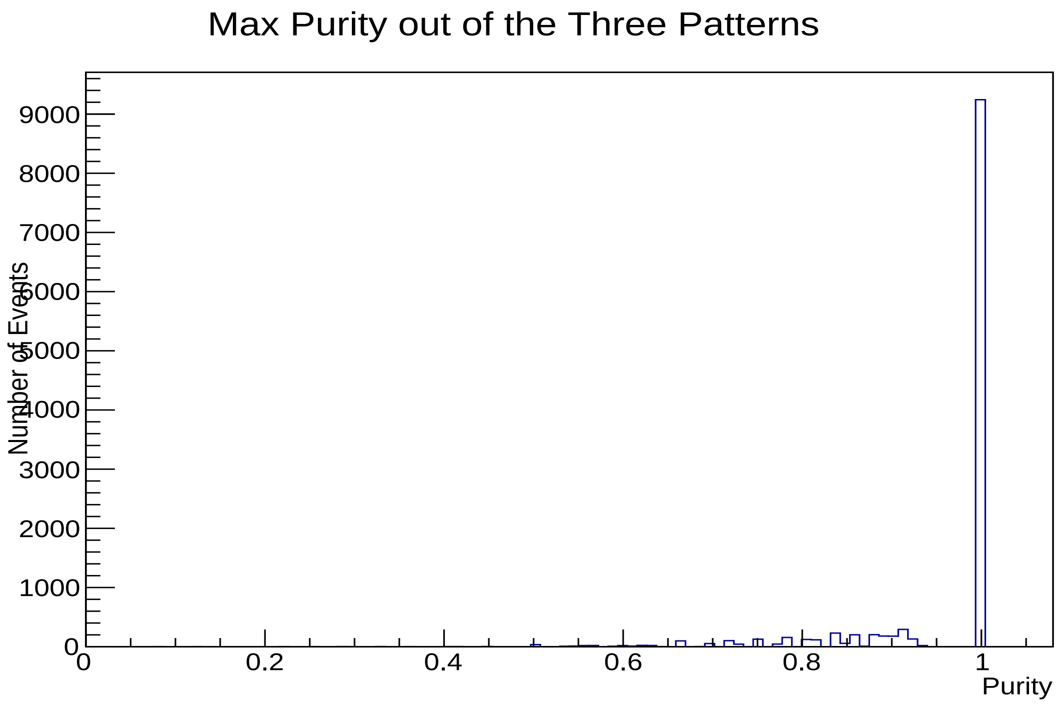

Figure 6: The maximum purity out of the three patterns identified for the incident hits. Only showed the distribution of events with more than 2 turns (chosen by cdc_turnId).

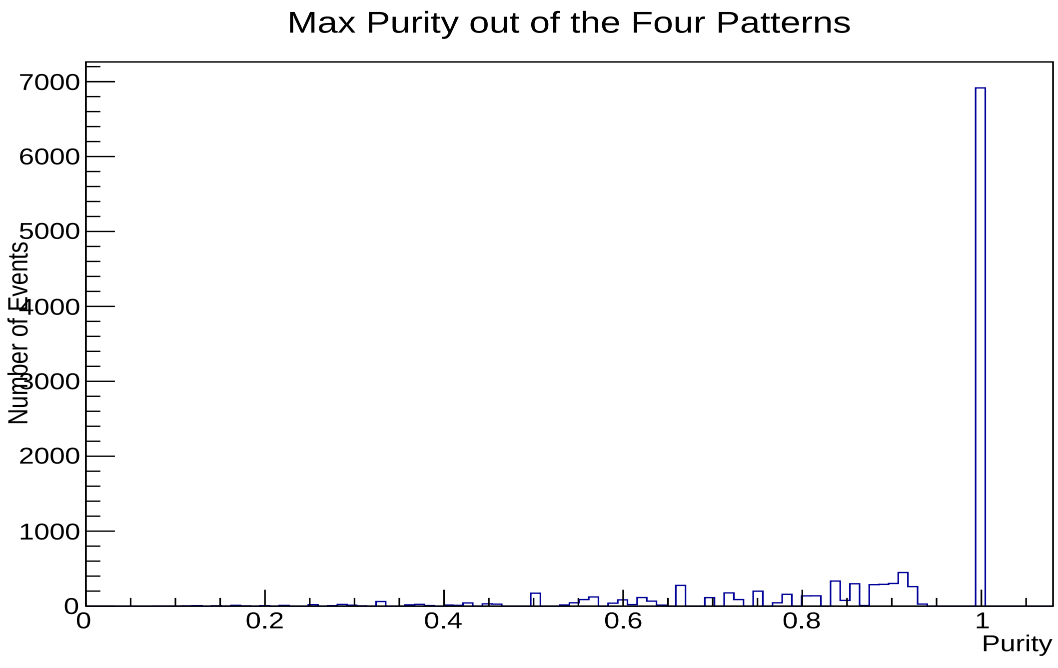

Figure 7: The maximum purity out of the four patterns identified for the outgoing hits. Only showed the distribution of events with more than 2 turns (chosen by cdc_turnId).

and 0 if no selection is correct. By plotting the purities, it was expected to find a turning point of the pattern dependence. Nevertheless, no clear pattern shifting was revealed in the plots. This indicates that it is not possible to find another simple pattern which could further improve the purity result for the exit hits.

For future developments, a sample with 85 MeV electron energy could be used to investigate the consistency of the selection algorithm. Furthermore, the code for this hit selection can be joint with the helix fitting algorithm to test its actual performance, as well as look at the fitting results for the events with max purity below 80%.

4 Conclusion

The simulated 105 MeV electron hit results were generated without beam noise using GENT4. An algorithm was developed to provide a single turn hit result for the downstream helix fitting to perform well. The algorithm looked at the incident and the exit side hits separately, with three patterns identified for the incident hits and four for exit hits. The purity of the selections were calculated, and the distributions of the maximum purity investigated. It was found that the percentage of events with maximum purity above 80% for incident and exit sides are 92.9% and 84.3% respectively. It was shown that both the patterns for selection on the incident and exit sides achieved exceptional success.