Emulsion Reconstruction Analysis in SND@LHC

T. Krishnan
CERN Summer Student 2023
Harvey Mudd College, Claremont, CA, USA and CERN, Geneva, CH

Abstract
Improving the emulsion reconstruction efficiency in SND@LHC, a compact neutrino experiment in the LHC, is important to ensuring faster data processing and more efficient physics analyses. In this report, I outline different methods to improve emulsion reconstruction efficiency and the signal-to-background ratio, analyzed by Monte Carlo studies as well as tests with real emulsion data collected in Emulsion Target 1. Monte Carlo tests have shown that improving the angular alignment of emulsion films can have a significant improvement in efficiency. Additionally, reducing track density from 10^5 to 10^4 tracks/cm^2 also results in a drastic improvement of efficiency. Removing tracks very close in angle to the collision axis is a potential method to reduce track density while improving the signal-to-background ratio as well. From analysis of the muon Monte Carlo sample, omitting base tracks less than 0.005 rad from the collision axis, I was able to reduce the proportion of background base tracks from passing muons to 78.54% while maintaining 90.83% of the signal base tracks from neutrino interactions. Further analysis with real emulsion data demonstrated it is possible to reduce background by 10.04% while retaining 96.48% of signal base tracks.

1 Introduction

SND@LHC (Scattering and Neutrino Detector at the Large Hadron Collider) is a compact, standalone neutrino experiment at the LHC, located 480m downstream of the ATLAS interaction point. as shown in Figure 1.

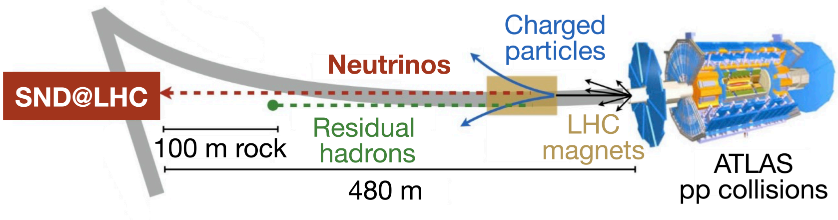

Figure 1: Location of SND@LHC 480m downstream of the ATLAS interaction point, in the TI18 tunnel. [2]

The experiment is optimized for the identification of the three neutrino flavors and detection of feebly interacting particles. It performs measurements in the pseudo-rapidity range of $7.2 < \eta < 8.4$. SND@LHC plans to probe the physics of heavy flavor production at the LHC in the very forward region, unexplored by the larger LHC experiments. This could provide useful information for future circular colliders and for predictions of very high-energy atmospheric neutrinos. The detector can also be used to search for the scattering of Feebly Interacting Particles [3]. In its first phase, the experiment will be operated throughout LHC Run 3 to collect 250 fb^{-1} of data, which corresponds to roughly 2000 neutrino events.

SND@LHC is a hybrid detector which collects both electronic data via the online system and emulsion data via the offline system. The detector is a hybrid system, with a target region made up of 800 kg of tungsten plates alternated with emulsion and electronic trackers, followed downstream by a hadronic calorimeter and muon system, as shown in Figure 2.

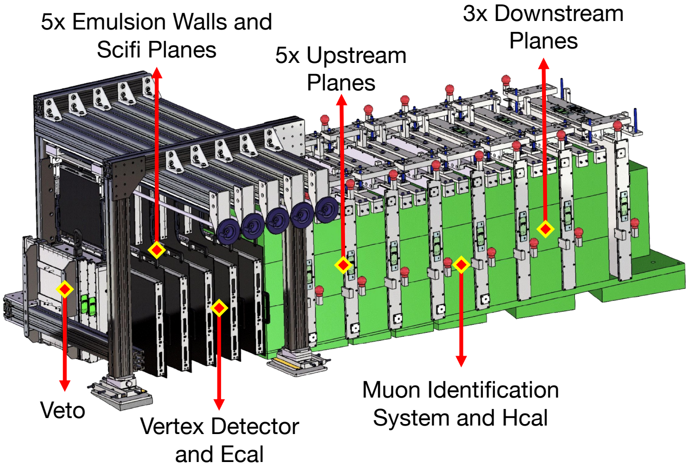

Figure 2: Schematic of SND@LHC detector [3]

My work is focused on optimizing the reconstruction efficiency of the emulsion data in SND@LHC, in order to maximize the speed and quality of emulsion data reconstruction. The aim of this optimization is to obtain physics results from the experiment more quickly and efficiently.

2 Emulsion Reconstruction

The nuclear emulsion detector in SND@LHC employs the Emulsion Cloud Chamber (ECC) technique. This technology allows for sub-micrometric position and milliradian angular resolution [3] in the target region of the detector, enabling detection of short-lived particles.

The emulsion films are replaced every 20 fb^{-1} and are then processed and scanned with an optical microscope. A picture of the emulsion film is shown in Figure 3(a). Each emulsion film contains segments, or base tracks, and by combining several emulsion films, the tracks and vertices can be reconstructed from these base tracks, as shown in Figure 3(b).

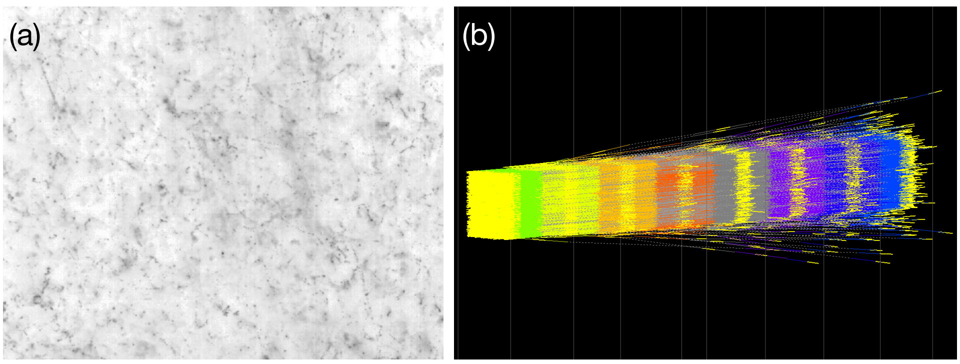

Figure 3: (a) View of single emulsion film through optical microscope. (b) Tracks from Emulsion Target 2, starting in an area of 1 mm^2

To optimize the reconstruction process, we need a measure of the performance of the reconstruction algorithm. For the Monte Carlo simulated events, the reconstruction performance for a given track was measured using the segment reconstruction efficiency defined as follows:

$$\text{Efficiency} = \frac{\text{\# of base tracks reconstructed}}{\text{\# of base tracks simulated}}$$

In this analysis, I only considered “long” Monte Carlo tracks that have more than 31 base tracks, as passing muons with many base tracks makes up most of our data, so ensuring that the longer tracks are well-reconstructed is crucial.

When analyzing emulsion data alone, we are unable to know how many events actually occurred in the target volume, as we only know what we are able to reconstruct. When we test different cuts on the data, however, we are able to compute the fraction of base tracks or vertices that are reconstructed with that cut, as compared to without any cuts. This can serve as a proxy to provide us with useful information about the quality of reconstruction under the different cuts we consider.

3 Reconstruction Optimization Studies

I performed three primary studies on Monte Carlo simulations of neutrino interactions and passing muons in order to analyze reconstruction efficiency:

1. Improving angular alignment of emulsion films in detector
2. Reducing emulsion track density
3. Omitting base tracks with shallow ZX and ZY angles

The first two studies consider a Monte Carlo simulated dataset of muon events, analyzing the segment reconstruction efficiency. The third study considers the Monte Carlo simulated datasets of muons and neutrinos and also considers a small subset of real data collected in Emulsion Target 1.

3.1 Emulsion Angular Alignment Analysis

A "smearing" procedure is applied to the Monte Carlo simulations, in order to reproduce the effects of slight angular alignment issues in the emulsion films as a Gaussian blur of a chosen radius. After the commissioning of the Emulsion Target 1, this radius was measured to be about 3mrad [3]. I tested the effects of an improvement in angular alignment, by comparing the segment reconstruction efficiency for the muon Monte Carlo sample in this case to one with smearing reduced to 1mrad.

From Figure 4, a reduction in smearing does appear to correspond to a much improved segment reconstruction efficiency on average. Therefore, if the angular alignment of the emulsion films is improved, the segment reconstruction efficiency also improves. Afterwards, I began to explore complementary methods to improve reconstruction efficiency.

3.2 Emulsion Track Density Analysis

I next considered how reconstruction efficiency might change if the density of tracks reconstructed in the emulsion films was varied. Starting from the original muon Monte Carlo sample, with 10^5 tracks/cm^2, I applied cuts based on the track ID prior to reconstruction in order to reduce the track density by 1 or 2 orders of magnitude. I then applied the reconstruction algorithm to the three different densities of muon Monte Carlo simulation: 10^5 tracks/cm^2, 10^4 tracks/cm^2, and 10^3 tracks/cm^2.

As shown in Figure 5, a lower density of tracks corresponds to a higher segment reconstruction efficiency, with a significant improvement from 10^5 to 10^4 tracks/cm^2, but a plateau beyond that. As a result, it seems that reducing the track density by an order of magnitude prior to reconstruction may

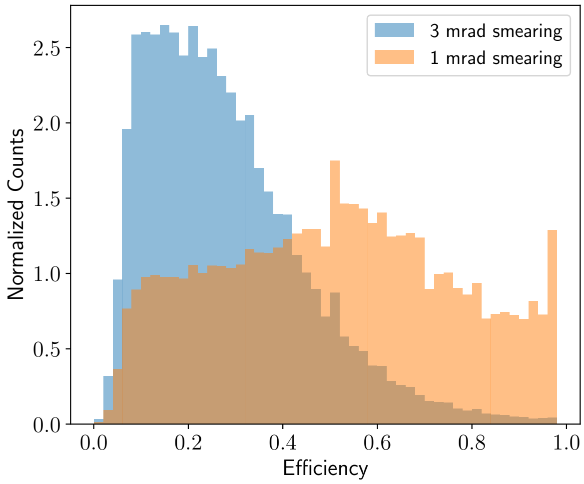

Figure 4: Segment reconstruction efficiency of muon Monte Carlo tracks using two different smearing parameters

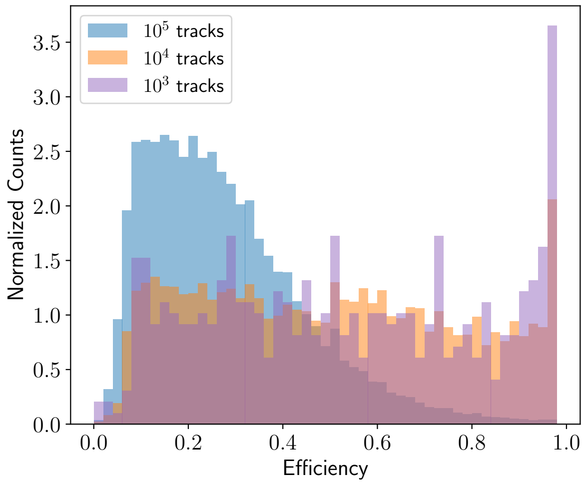

Figure 5: Segment reconstruction efficiency of muon Monte Carlo tracks with three different track densities.

allow for a much-improved reconstruction efficiency. The most feasible way to implement this would be through cuts, as it is not practical to swap the emulsion films more frequently than the current schedule.

However, we must then decide which base tracks to cut prior to reconstruction, without removing too many signal (neutrino event) base tracks. I next began exploring one possible method, by cutting base tracks that deviate significantly in angle from the collision axis.

3.3 Base Track Angular Cut Analysis

Emulsion films are oriented in the X-Y plane of the detector, stacked up along the collision axis, or Z-axis of the detector. The background events, which mostly consist of passing muons, tend to have much shallower XZ and YZ angles than the signal tracks, which are those from neutrino events. In principle, applying cuts eliminating shallow-angled base tracks could allow us to obtain a higher signal-to-background ratio. This would also help in maximizing segment reconstruction efficiency by reducing the track density.

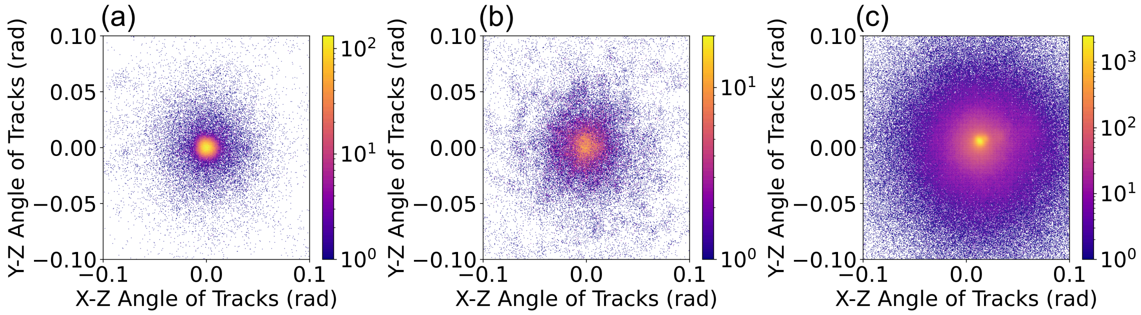

Figure 6: Example of angular distribution of base tracks for (a) muon Monte Carlo sample, (b) neutrino Monte Carlo sample, and (c) small sample of real data from Emulsion Target 1

3.3.1 Applying Cuts

I first began by testing this procedure on individual Monte Carlo samples of muons and neutrinos. As shown in Figure 6(a) and (b), the distributions of base track angles (θ_{XZ} vs θ_{YZ} for the simulated datasets are radially symmetric, so I employed a circular cut on shallow-angled base tracks of the form:

$$\sqrt{\theta_{XZ}^2 + \theta_{YZ}^2} > \theta_{min}$$

. Figure 7 shows one example of the cut, where $\theta_{min} = 0.01$rad.

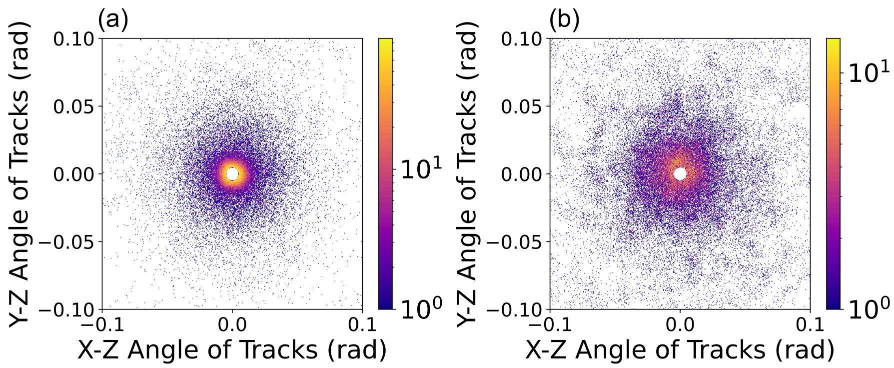

Figure 7: Example of angular cut on (a) muon and (b) neutrino MC samples with $\theta_{min} = 0.01$rad

Since the muon base tracks are much more concentrated near the origin, this cut allowed me to cut a higher proportion of the muon base tracks than the neutrino base tracks, which should improve the segment reconstruction efficiency while preserving most of the signal.

Next, I tested a similar procedure on real data collected in Emulsion Target 1. As shown in Figure 6(c), the angular distribution of base tracks in real data is not radially symmetric. This asymmetry is not reproduced in the Monte Carlo simulations, which is why a circular cut was used in that case. For the real data, I decided to apply an elliptical cut instead of a circular cut, in order to more efficiently cut around the peak of the data. In order to construct the best-fit elliptical cut around the data, I generated elliptical confidence intervals using the method outlined in [1]. I then reconstructed the equation of the ellipse from the resulting confidence interval, as outlined in Appendix A. The resulting best-fit elliptical cuts are shown overlaid on the data in Figure 8.

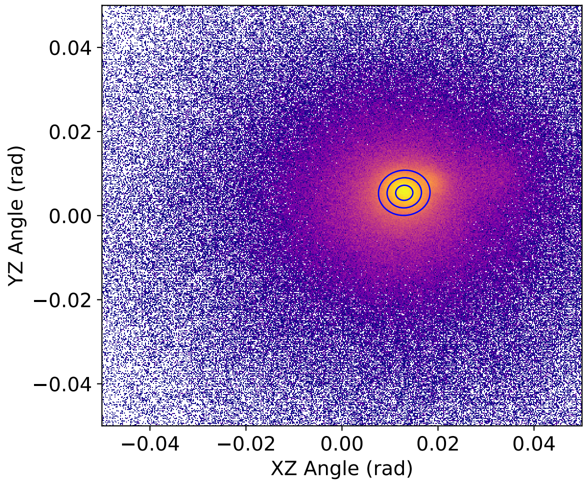

Figure 8: Best-fit elliptical cuts on data sample from Emulsion Target 1.

To compare these cuts with the circular ones, I computed the effective radius of each cut, r, as

$$\pi r^2 = \pi \frac{lw}{4}$$

where l and w are the lengths of the major and minor axes of the ellipse respectively.

3.3.2 Reconstruction Optimization

Once I selected the cuts to apply to each data set, I computed the fraction of the original reconstructed base tracks and vertices, when no cuts were applied, were reconstructed under the different selected cuts. Figure 9 shows the proportions of vertices and base tracks reconstructed with no cuts that are reconstructed under different applied cuts, defined by the cut radii.

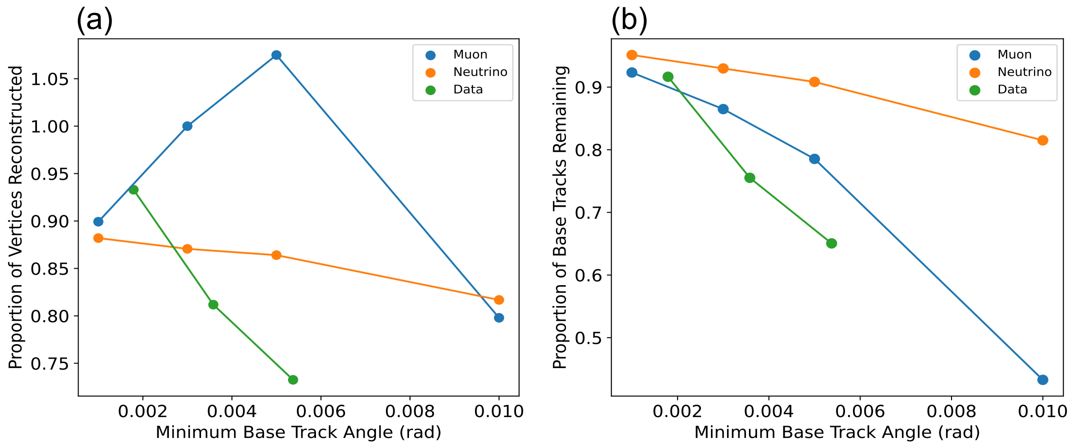

Figure 9: Proportion of (a) vertices and (b) base tracks reconstructed after applying varying thresholds of base track angular cuts compared to with no cuts.

Note that the proportion of muon Monte Carlo sample vertices reconstructed becomes larger above 1 under the 0.005rad cut. This is because as the muon density decreases up to a certain amount, vertex reconstruction becomes easier as there are fewer base tracks to connect. As a result, there is an increase in proportion of vertices reconstructed compared to without cuts. However, when the cut becomes large enough, we do end up reducing the number of reconstructed vertices, as we remove a significant number of the base tracks resulting from that vertex.

In order to reduce the density of base tracks and improve reconstruction efficiency, we can focus on Figure 9(b). Comparing the muon and neutrino Monte Carlo sample results, it appears that we can reduce the muon background quite significantly while retaining the majority of the signal neutrino base tracks. By omitting base tracks at X-Y angles less than 0.005 rad, it is possible to reduce the proportion of background base tracks from passing muons to 78.54% while maintaining 90.83% of the signal base tracks from neutrino interactions. As a result, this method of omitting shallow-angled base tracks appears promising in improving the signal-to-background ratio while reducing track density to allow for improved reconstruction efficiency.

However, it is also clear from Figure 9(a) that the vertex reconstruction of the muon Monte Carlo sample does not line up very well with the sample of real emulsion data, although the base track reconstruction agrees fairly well. Since we expect muon background to constitute the majority of the real data, it is odd that they differ significantly. It appears that there are other complicating factors influencing reconstruction of real data that are not accounted for entirely by the passing muons.

3.3.3 BDT Analysis of Data

As a result of the discrepancy between the data and the muon Monte Carlo simulation, I decided to separate the signal and background in the real data sample using the BDT (Boosted Decision Tree) output from an existing TMVA classification algorithm to verify that the signal-to-background ratio in the data can be improved through these cuts, just as we saw with the Monte Carlo samples. I cut the data based on the BDT output at the optimal value, which was previously determined by the multivariate analysis. This cut leads to a signal efficiency of 92.48% and background rejection at 97.43%.

After separating signal and background with this cut, I looked at the fraction of vertices and base tracks that were reconstructed under each cut, compared to the tracks reconstructed without any pre-selection, to check if there is a change in the pattern previously reported using the Monte Carlo sample.

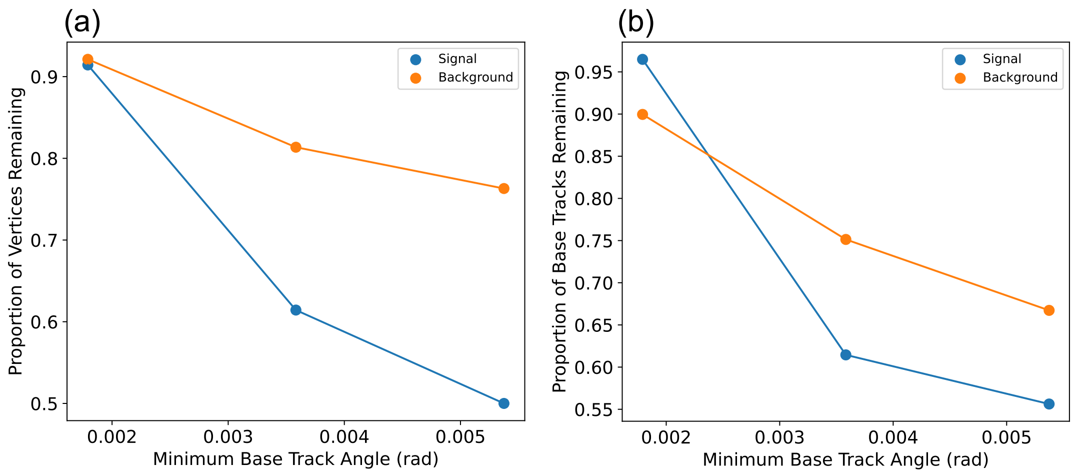

Figure 10: Proportion of (a) vertices and (b) base tracks reconstructed after applying varying thresholds of base track angular cuts compared to the tracks reconstructed without any pre-selection.

The results are shown in Figure 10, which looks quite different from the Monte Carlo test. For the larger cuts, it appears that we do end up losing more signal base tracks and vertices compared to background, which is not acceptable for our analysis. However, with the smallest cut, we can retain 96.48% of the signal while keeping only 89.96% of the background base tracks, which is promising. We have yet to analyze whether this cut alone would provide enough of a reduction in base track density to significantly improve segment reconstruction efficiency, or determine further methods to optimize the cuts selected, but it appears to be a step in the right direction.

4 Conclusions

I have discussed a few methods for improving segment reconstruction efficiency of emulsion data in SND@LHC. Improving angular alignment of emulsion films and reducing the base track density prior to reconstruction are two methods that may allow for higher efficiency.

Improving angular alignment from 3mradto 1mraduncertainty provides a significant increase in segment reconstruction efficiency, but may be challenging to implement. Reducing base track density from 10^5 t0 10^4 tracks/cm^2 prior to reconstruction seems to be the most feasible option, also providing a significant improvement in segment reconstruction efficiency. By performing selections on emulsion Monte Carlo samples prior to reconstruction and cutting base tracks at X-Y angles less than 0.005 rad, I was also able to reduce the proportion of background base tracks from passing muons to 78.54% while maintaining 90.83% of the signal base tracks from neutrino interactions. This cut would also reduce track density, allowing for a higher overall segment reconstruction efficiency, and improved reconstruction quality for physics analysis. I also applied this method to a small sample of real data collected in Emulsion Target 1, and used an existing TMVA classification algorithm to separate signal from background. Thus, I was able to verify that by the application of an elliptical cut on the asymmetric angular distribution of base tracks from data, it is possible to reduce background by 10.04% while retaining 96.48% of signal base tracks, as desired. Future work might test this on a larger dataset and further optimize the selected cuts to further improve the signal-to-background ratio.

5 Acknowledgements

I would like to acknowledge Dr. Antonio Iuliano and Dr. Antonia Di Crescenzo for their guidance and mentorship throughout this program, the SND@LHC Collaboration and the Napoli Emulsion Lab for their support, the CERN Summer Student Programme Coordinators and the University of Michigan CERN REU Program Coordinators for organizing this program, and my fellow summer students for making this such a wonderful experience.

Appendix

A Derivation of Elliptical Cuts

The elliptical cuts I used are of the form:

$$\sqrt{\frac{((x-m_x)cos\theta-(y-m_y)sin\theta)^2}{(\frac{h}{2})^2}+\frac{((x-m_x)sin\theta+(y-m_y)cos\theta)^2}{(\frac{w}{2})^2}}>1$$

where m_x, m_y are the x- and y-coordinates of the peak of the 2d angular base track distribution, θ is the tilt of the major axis of the best-fit ellipse from the horizontal, and h, w are the lengths of the major and minor axes respectively. The XZ and YZ angles of a given base track are given by x and y respectively.

Since the data was asymmetric, I selected a circular subset of the data around the peak with radius 0.03radbefore applying the algorithm. This ensured that the peak was centered in the distribution. I then created confidence intervals at 0.2σ, 0.4σ, and 0.6σ.

This derivation follows from the derivation of the elliptical confidence interval provided by [1]. The output of the elliptical confidence interval algorithm provides, for each confidence interval generated, the mean of the distribution (m_x, m_y), the Pearson Correlation Coefficient (p), and the x- and y-scale factors of the ellipse (s_x, s_y). The scale factors are two times the product between the number of standard deviations desired for the confidence interval and the standard deviation of the distribution along that axis.

Then, I used the geometry of the applied transformations to calculate the length of the major and minor axes as

$$h = \sqrt{2(1+p)(s_x^2 + s_y^2)}$$

$$w = \sqrt{2(1-p)(s_x^2 + s_y^2)}$$

and the angle of the ellipse as

$$\theta = tan^{-1}\left(\frac{s_y}{s_x}\right)$$

Thus, all parameters of the cut have been determined.

Bibliography

[1] Carsten Schelp, An alternative way to plot the covariance ellipse, available at the URL https://carstenschelp.github.io/2018/09/14/Plot_Confidence_Ellipse_001.html

[2] Martina Ferillo, SND@LHC - Scattering and Neutrino Detector at the LHC, Lake Louise Winter Institute 2023 (2023, February)

[3] The SND@LHC Collaboration. SND@LHC: The Scattering and Neutrino Detector at the LHC, available at the URL https://arxiv.org/abs/2210.02784.