Data Analysis of SoLID Detector Beam Test with High Rate

Darren Upton

May 15, 2023

SoLID

SOLENOIDAL LARGE INTENSITY DEVICE

Abstract

After the conclusion of the low-rate portion of the SoLID detector beam test from Fall 2022, the high-rate beam test began in January 2023. Here we present preliminary analyses of the beam test data collected with the test-stand located at 18° off of the beamline, with a particular emphasis on the Cherenkov detector and the Electromagnetic Calorimeter (ECal). Special attention is paid to position determination of events in the ECal, Cherenkov detector, and GEM detectors as well as characterization of the Cherenkov detector. This work lays the groundwork for more precise detector analysis and detailed Particle Identification (PID) analysis in future iterations.

Contents

1 Introduction

The Solenoidal Large Intensity Device (SoLID) detector is a large acceptance forward scattering spectrometer with full azimuthal-angular coverage capable of handling high luminosities (10^{37}-10^{39} cm^{-2}s^{-1}) with a variety of polarized and unpolarized targets. The detector will exploit the full potential of the JLab 12 GeV upgrade with programs in 3D imaging of the nucleon, parity violating electron scattering in search of beyond standard-model physics, and exploration of gluonic forces. To accomplish these lofty physics objectives, the SoLID detector design must be well-understood and validated before the construction begins. A series of beam tests were carried out from August 2022 through March 2023 to study the detector performance, with an emphasis on the Electromagnetic Calorimeter (ECal) and the Scintillator Pad Detector (SPD), under a high rate, high radiation and high background environment.

1.1 SoLID Beam Test Configuration

The hardware configuration for the SoLID detector beam test is shown in Fig. 1. There were four various-sized scintillators, two pairs of GEMs, Large-Angle Scintillator Pad Detector (LASPD), Cherenkov detector, and Electromagnetic calorimeter (ECal) composed of the Preshower and Shower Modules. All of these were attached to a test stand in Experimental Hall C at Jefferson Lab (see Fig. 2).

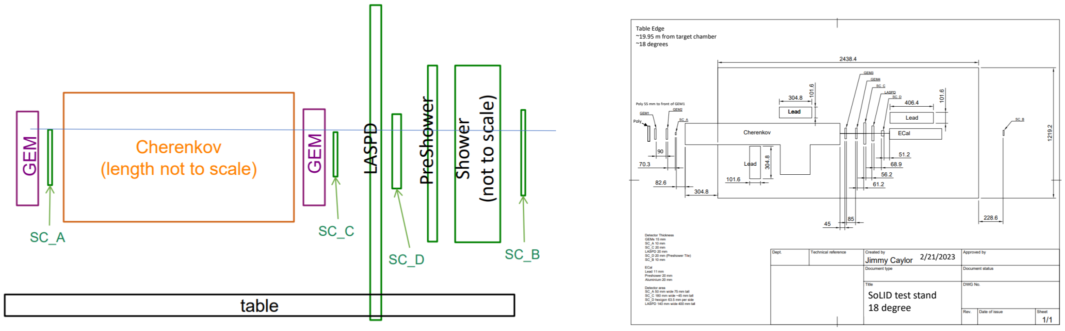


Figure 1: Layout of detectors (left) and measurements of the relative positions of the components (right).

The hardware for the SoLID (shown in Fig. 2) was located at 7° and 18 ° from the Hall C beam line for the high-rate portion of the beam test. At these angles, the stand was located near the High Momentum Spectrometer (HMS). Since the beam test was "parasitic" in nature, we needed to conform to the needs of the running experiment(s) in the Hall. Thus our locations and access to the experimental Hall were constrained by the running experiments.

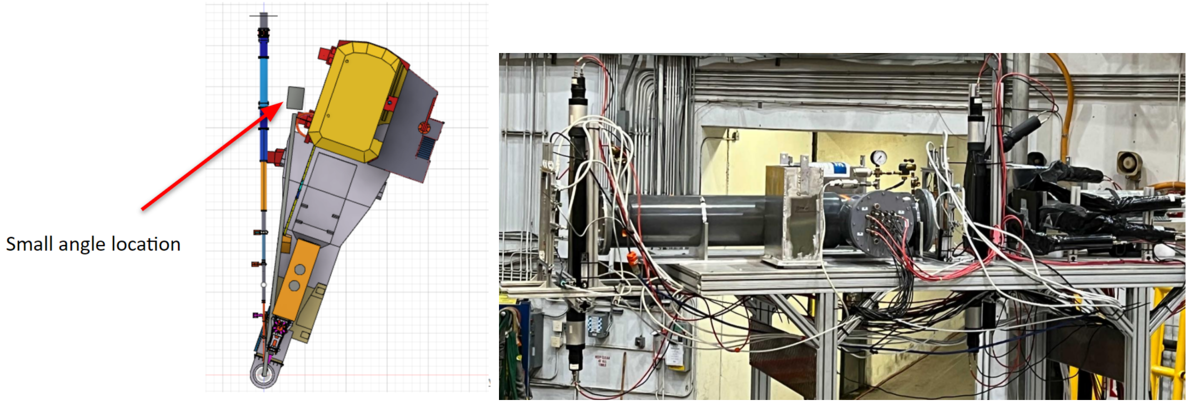


Figure 2: Diagram for SoLID at 18° (left) and SoLID hardware on the test stand in Hall C (right).

1.2 Triggers

During the high-rate test, data were taken at 7° and 18° with each having a set of hardware triggers. Since an objective of this beam test is studying particle identification (PID) methods for SoLID, we require adequate samples for electrons, pions, and photons. Therefore, we determined an array of triggers at both angles to satisfy these needs. Table 1 and 2 provide the key points for the triggers used for the high-rate beam test. The trigger "condition" implies the logical "AND" of the detector signals listed, formed by the output from sending each to a discriminator channel.

Trigger	Bit	TrigType	Aim	Condition
TS1	000001	1	e	Č Sum, ShSum
TS2	000010	2	π	SC-B, ShSum, SC-D
TS3	000100	4	2/3	Č, SC-D, ShSum
TS4	001000	8	high-energy e	ShSum
TS5	010000	16	Clean π	SC-B
TS6	100000	32	Good Č	2/15 FADC Č coincidence

Table 1: Triggers for runs at 7°

Trigger	Bit	TrigType	Aim	Condition
TS1	000001	1	e	Č Sum
TS2	000010	2	π	SC-B, SC-D
TS3	000100	4	MIP	SC-A, SC-D
TS4	001000	8	high-energy e, or γ	ShSum

Table 2: Triggers for runs at 18°

These tables provide a general summary of the triggers for the high-rate test, however more specific details on high-voltage settings of the photomultiplier tubes (PMTs), discriminator thresholds, as well as target and beam current information can be found in a more detailed spreadsheet (see e.g. 18° run list). These run-specific details are critical to properly investigating the triggers and have been taken into account for this analysis.

1.3 Data Processing

For the SoLID beam test reported in this work, all detector signals were read out by JLab 250MHz Flash-Analog-to-Digital-Converters (FADCs). The data are processed and are primarily analyzed in the form of a level0 and a level1 ROOT tree. The level0 tree contains the waveforms from each recorded event, thus the FADC sampling values for a window of time. This would rely on timing information from all of the detectors, where we require peaks in specific detectors occur within a given time window and any offset be corrected. For instance, TS2 requires SC-B and SC-D (see Fig. 2), so we expect to see coincidence peaks in these two detectors, and we look for peaks in other detector within the same time window to keep with this event. The power of the level0 tree comes when considering one waveform may contain multiple events in the same window known as event pile-up. Since the waveform is recorded, we could in principle separate these peaks through use of robust peak-finding and peak-deconvolution algorithms. While these analysis technique continue to be improved, we can for now place baseline requirements to find simple events that contain a single peak.

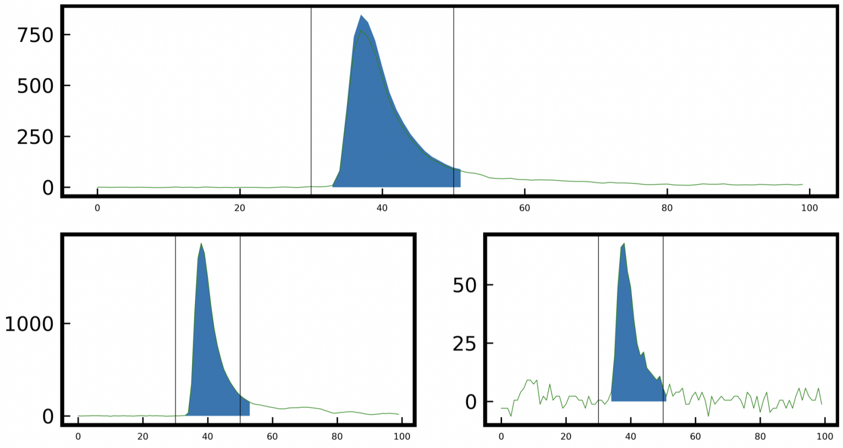


Figure 3: Subplots show the waveform from level0 tree for each of the Shower modules (left, top, and right). From this, we see pedestal seems to fluctuate on the order of 20 FADC units, thus large values can easily be distinguished from noise while small signals become indistinguishable. This factors into the peak-finding algorithm, which uses different user-specified (software) thresholds to best determine peaks and integrate over the appropriate timing-windows.

By integrating the peaks of the level0 tree, we obtain the level1 tree with values given in integrated-FADC. Level1 was the primary one used in the analysis presented in this report. However, insight from level0 trees guided the analysis, and limitations from the peak-fitting algorithm presented many challenges during the analysis process.

In the level1 tree, there is an important distinction between the hardware and software sums. Software sums come from summing the integrated-FADC values for each component (i.e. Shower left+right+top), while the hardware sums represent the integrated-FADC value for these components summed at the hardware signal-level, typically by a summing module. Ideally, the different channels of the same detector are gain and timing matched and these values would closely agree, and they do up to a width. However, the hardware and software versus hardware sums become far more separated for the Cherenkov because it contains 16 channels that are not very-well gain matched.

2 Detector Performance

2.1 Overview of Detector Performance

As an initial step to analyzing a run, plots like Fig. 4 provide a concise summary of which detectors were operating properly, at least somewhat reasonably. Additionally, by only inspecting this plot, one can in principle determine which detectors were used in the definitions of triggers or examine if the triggers were set up correctly. Furthermore, when minimal ionization peaks (MIPs) are seen in the detectors, or when single photo-electron (SPE) peaks are seen in the Cherenkov, they can be used for calibration purposes.

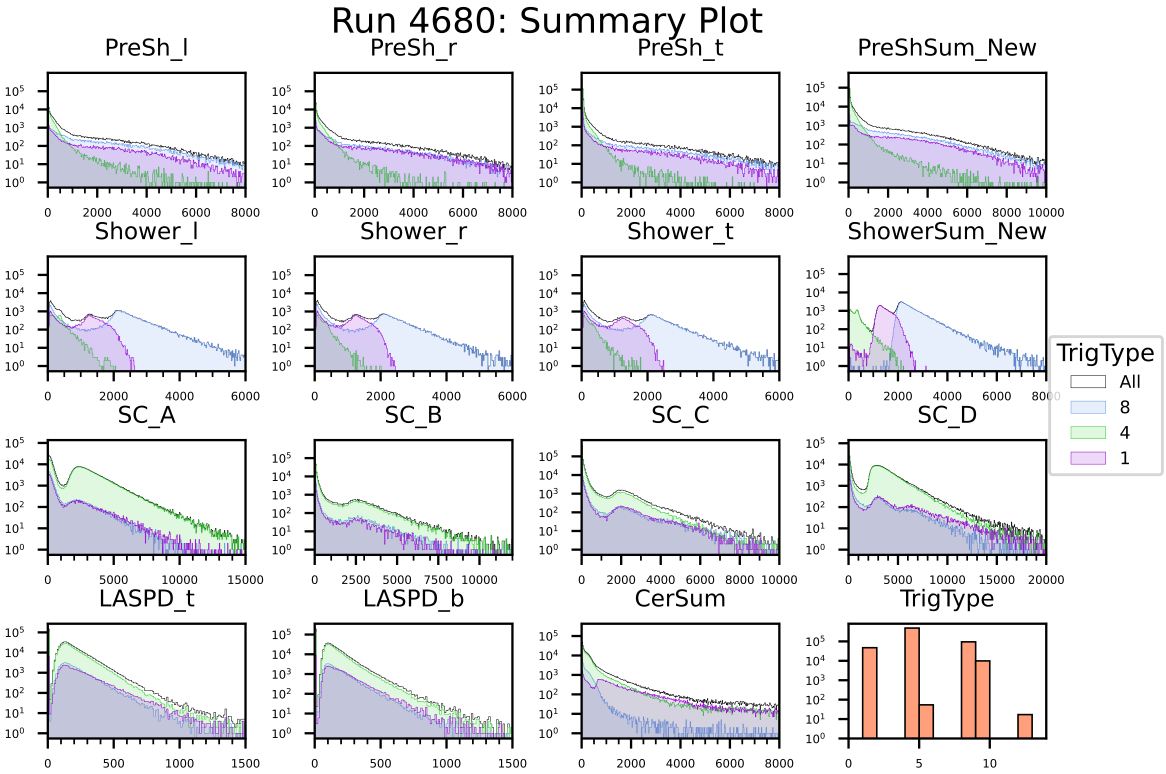


Figure 4: Overview of detector signals for Run 4680 (5 μA, Deuterium target) showing the Preshower, Shower, scintillators, LASPD, hardware Cherenkov sum, and distribution of trigger types. The primary triggers for this run are TS1, TS3, and TS4 (see Table 2), while "secondary" triggers come from coincidence between two triggers. For instance, an event triggering both Trig 1 (bit 0001) and Trig 4 (0100) at the hardware level would produce a Trig 5 (bit 0101) event.

As mentioned, the trigger configurations at both angles have an electron, pion, charged-particle, and high-energy-electron triggers to obtain samples. These will permit detailed PID and characterization of the Cherenkov detector (in addition to others). These trigger thresholds can be observed in Fig, 4. For instance, TS4 has a high ShowerSum threshold, hence one observes essentially no events until that threshold around 2200 integrated-FADC. Additionally, the scintillator thresholds for TS3, as well as the CherSum threshold for TS1 are visible from the subplots. One can determine these thresholds more precisely by setting a criterion, for instance one-tenth of maximum, and determining that position on the distribution. Additionally, these thresholds add complexity to the search for MIP peaks because the thresholds (or their effects) may create a MIP-like bump that must be carefully studied before blindly fitting ghost-MIP peaks.

2.2 ECal: Preshower & Shower

The two components to the ECal are the Preshower and Shower modules. Each consist of three, hexagonal-prism-shaped modules. The Preshower modules are 2-cm thick scintillators following a $2X_0$ (two radiation length) of lead, while the Shower modules are made using the Shashlyk design (see Fig. 5). By inspecting first two rows of Fig. 4, one finds the distributions for the Preshower and Shower integrated-FADC values separated by primary trigger type.

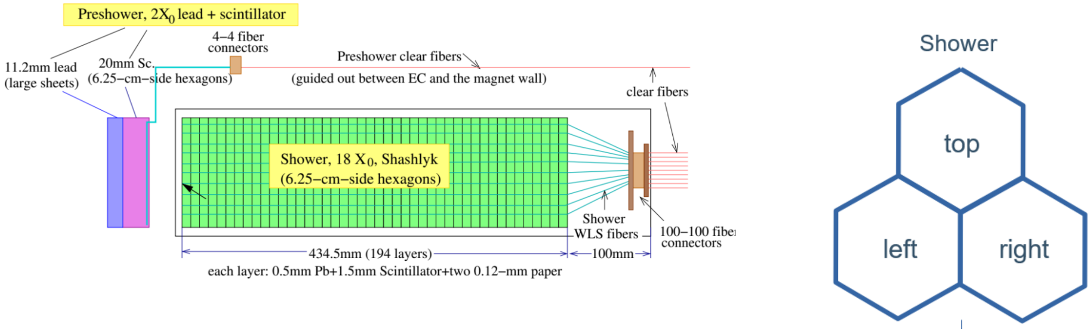

Preshower, $2X_0$ lead + scintillator
4–4 fiber connectors
Preshower clear fibers
(guided out between EC and the magnet wall)
11.2mm lead (large sheets)
20mm Sc. (6.25–cm–side hexagons)
clear fibers
Shower, 18 X_0, Shashlyk (6.25–cm–side hexagons)
Shower WLS fibers
100–100 fibe
connectors
434.5mm (194 layers)
100mm
each layer: 0.5mm Pb+1.5mm Scintillator+two 0.12–mm paper
Shower
top
left
right

Figure 5: Diagram for the Shashlyk sampling calorimeter used for SoLID (left) and hexagonal module arrangement used for the SoLID beam test ECal (right).

A vital component of the ECal analysis is the identification of MIPs. MIPs are particles whose primary mechanism of energy loss is ionization of the material through which they travel. The energy deposit (in MeV) for these peaks can be easily obtained from simulation or simple calculations, thus we can fit the peaks for data and simulation to obtain a scaling factor between integrated-FADC values and the energy deposit in the detector. The MIP finding and fitting have been a challenge at various stages of the beam test, and one of the fits is shown for the Shower in Fig. 6.

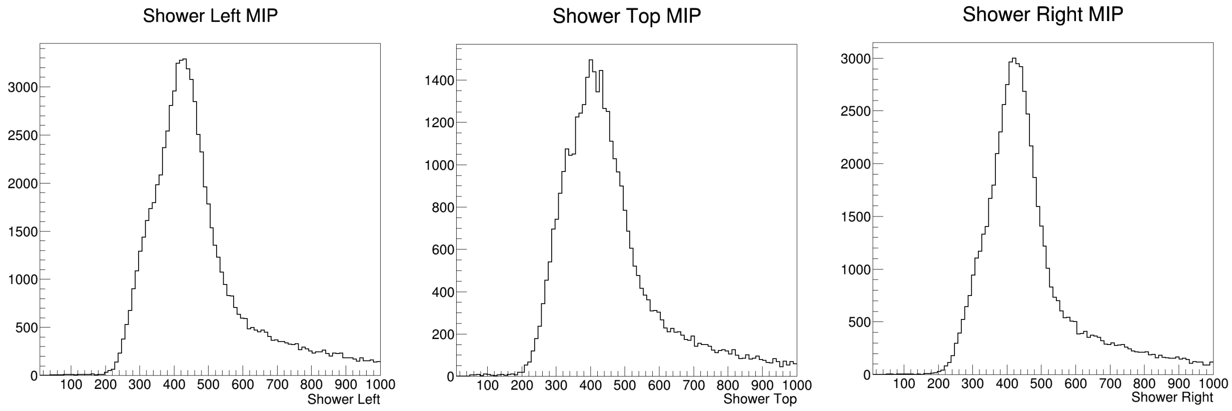


Figure 6: Shower MIP values for Run 4776 using TS3 with cuts applied such that (shower module value/shower sum)>0.75. The cut was imposed to select events made primarily of pions passing through a single module, as electrons would create electromagnetic shower and would deposit energy in all three modules.
. Note that the modules are mostly gain-matched at this point in the run list.

3 High-Rate Simulation Overview

The analysis reported here involved the consistent and continuous interplay between simulation and data. We investigated the GEANT4-based simulation to gain an understanding of the expected results for the data. The simulation was produced by other SoLID collaborators that accounted for the realistic beam-on-target situation and included both electron ("eAll") and hadronic (charged and neutral pions) background ("bggen"). For this analysis, we are primarily focus on four particles from the simulation: electrons and three pions. Their simulation particle ID and generator are shown in Table 3, while basic properties of the simulated events such as momentum and scattering angle are shown in Fig. 7.

Particle	PID	Generator
e	11	eAll
π^+	211	bggn
π^0	111	bggn
π^-	-211	bggn

Table 3: Overview of PID labels (as defined in Particle Data Group or PDG) for simulation and event generators, as well as the event generators used for the simulation.

By determining the MIP position and other quantities in the simulation, we are able to more-directly compare simulation to data. The values for simulation are given in deposited energy, while the data are given in integrated-FADC.

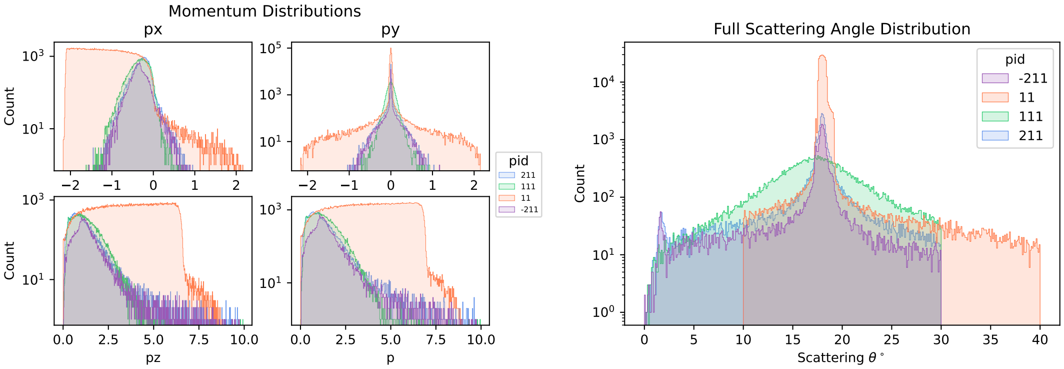


Figure 7: Distribution of momentum (left) and scattering angle (right).

By inspecting Fig. 7, we can learn quite a bit about the expected events for the 18° runs. The scattering angle for a particle in the simulation is given by:

$$\theta_S = \arccos(p_z/p) \tag{1}$$

where p_z is the momentum component in the z (beam) direction and p is the total momentum. This peaks around 18° because that is the angle where the SoLID detector is positioned in the hall. In order to have a cleaner sample, we impose a scattering-angle requirement of $|\theta_S - 18°| < 2°$ to remove some of the long tails for momenta and scattering-angle distributions of Fig. 7.

3.1 Simulation of ECal

In Fig. 8, the simulated energy deposit distribution in each of preshwoer and shower modules are shown, with the MIP peaks highlighted by blue vertical lines. From this we can extract the MIP peak values to compare simulation to the data. This allows us to make simulated trigger cuts, thus allowing for the further comparison and provide a baseline for conducting PID analysis. The units for the simulation are equal between the modules, thus the data modules must be gain matched to conduct a direct comparison between the two.

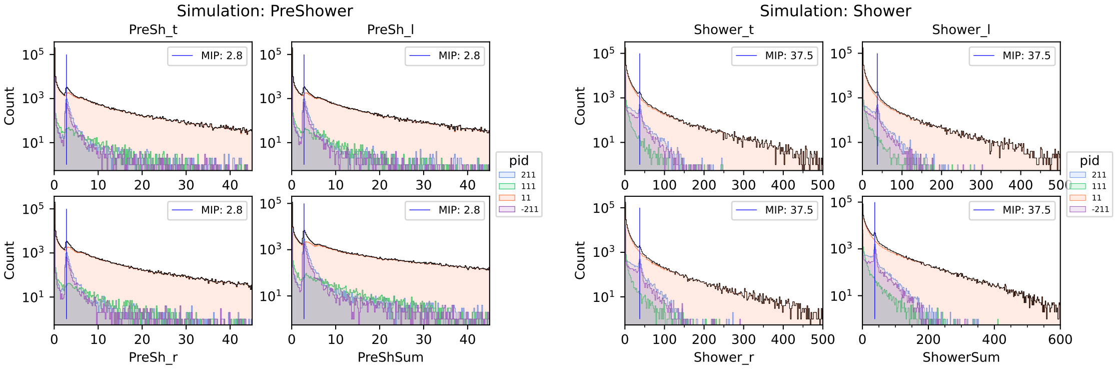


Figure 8: Distribution of energy deposited in Preshower (left) and Shower modules (right).

3.2 Simulation of Cherenkov

Figure 9 shows the Cherenkov simulated output in Number of Photo-Electrons (Npe). In the left plot, we see almost all charged pions with zero, as there should be. There is a reasonably high number of π^0 with a large NPE, however, that comes from $\pi_0 \ \rightarrow \ \gamma + \gamma$ where the photons subsequently produce $e^- e^+$ pairs.

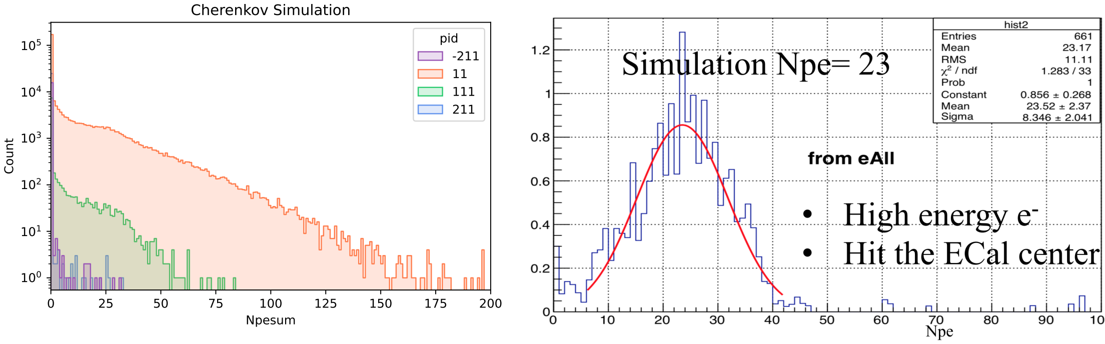


Figure 9: Simulation Cherenkov given in sum of photoelectrons (Npe) for particle type (left) and Cherenkov distribution for simulation high-energy, ECal-center electrons.

While from Fig. 9 left panel, we see that the electron signal shows a (vague) peak in NPE around 25, the peak can be highlighted if we apply cuts to select only high-energy electrons hitting the center of the 3 ECal modules. This is shown in the right panel of Fig. 9 and the peak fitted to show $N_{pe} \approx 23$. We should be able to find this peak in the data, if we can obtain a large enough sample of high-energy electrons. However, the simulation assumes all Cherenkov channels produce signal with equal efficiency. One additional step in data processing is gain matching by fitting the photo-electron peak in each channel, then scale (gain-match) the data in order to compare it to the simulation.

3.3 Simulation of Scintillators

The right plot of Fig. 7 shows the energy deposited in all four scintillators. One can easily find the MIP, double-MIP, and even triple-MIP peaks (for some). While scintillators A, C, and D look relatively similar in their general shape, scintillator B shows significantly fewer events. This is due to the fact that this scintillator is placed behind the ECal, and thus only pions (hadrons) are able to form signals in SC-B, though some high-energy electrons can also trigger it due to energy leakage. The hardware trigger utilizing scintillators typically have the discriminator threshold set at 0.5 of MIP values.

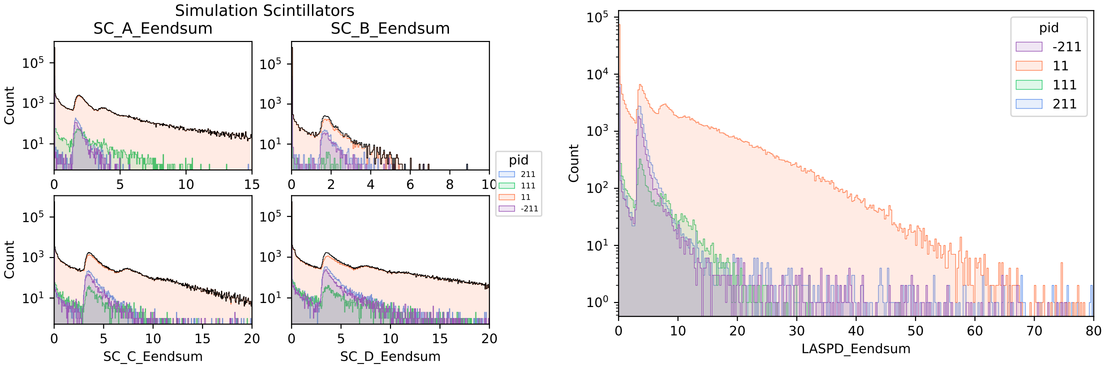


Figure 10: Distribution of energy deposited in scintillator modules (left) and LASPD (right).

3.4 Pseduo-Trigger Definitions

Using a procedure studied by Ye Tian, we can apply trigger cuts to the simulated events to create a pseudo-TS4, which will provide insight into PID for high-energy electrons.

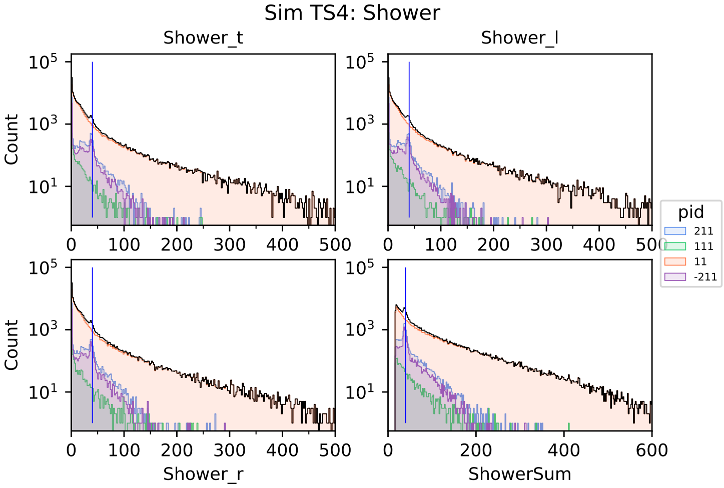


Figure 11: Application of TS4 to simulated events for 18°, which is a cut on ShowerSum.

To apply this pseudo-trigger, we determined the corresponding threshold for energy deposited in the shower modules between the simulation and data. By obtaining the MIP peaks values for data and simulation, we can then scale the simulation to the data by

$$Scaled\ Value = Sim\ Value\ \frac{Data\ MIP}{Sim\ MIP} \tag{2}$$

We scaled the simulated Shower left, right, and top by the data MIP peak integrated-FADC value (see Table 4), then place a cut on this scaled value at 200 FADC.

Shower Module	Sim MIP	Data MIP
Left	37.5	463.3
Top	37.5	502.2
Right	37.5	464.9

Table 4: Table for the Minimum Ionizing Particle (MIP) peaks for the shower modules for data and simulation. These values come from fitting the MIP peaks.

As seen in the left plot of Fig. 11, enforcing this cut creates a sharp threshold in the ShowerSum subplot. This differs from the shoulder found of the data likely because the physical trigger module uses the height of the waveform, while the level1 ROOT tree contains the integrated-FADC value.

3.5 Initial ML-Assisted PID

Using the insight provided by simulated events, we can develop particle identification (PID) methods to apply to the data. For a portion of this semester, some preliminary Machine Learning techniques were attempted for PID. The primary attempts used Multi-Layer Perceptron (MLP), a standard baseline tool in ML, to separate the four particle types. We use TensorFlow because of its increased flexibility and customizability versus scikit-learn or other packages.

3.5.1 Baseline MLP

We start with the simulated sample of $\sim$ 320,000 events, which are then split into training and validation samples at a ratio of 80:20 (creating samples of roughly 260,000:60,000). We then select the detector responses (variables) to include in the classification procedure; these include: PreShower/Shower Sums and modules, Eendsums for all 4 scintillators and LASPD, number of GEM hits for all 4 GEM modules, and photoelectron sum. These responses are then scaled using the "MinMaxScaler" to scale distributions between 0 and 1. This is standard procedure for ML because of the susceptibility for the MLP to latch onto different scales between input variables. Then the MLP model is built using six hidden layers of sizes 256, 128, 64, 32, 16, and 4 with relu activation functions. Then using the Stochastic Gradient Descent (SGD) optimizer with a learning rate of 0.01 and loss function of "SparseCategoricalCrossentropy". This model is then trained using 12 epochs and a batch size of 64, resulting in the plots of Fig. 12

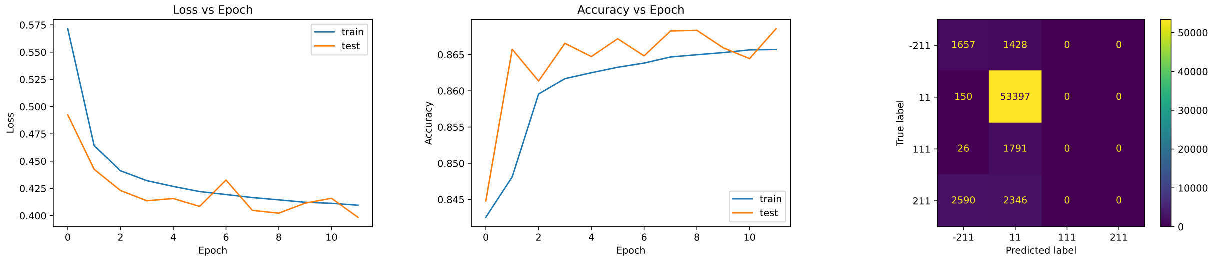


Figure 12: Loss (left) and accuracy (middle) versus training epoch for the baseline MLP approach. These show the model's improvement for the training epochs and that the model is likely not overtrained. Additionally, the Confusion Matrix (right) shows that this model has a high electron efficiency but poor pion-rejection.

As shown in Fig. 12, this initial model classifies electrons very well with an efficiency of 99.7% but a very poor pion-rejection (percent of π not classified as electrons) of 43.4%. Additionally, the Confusion Matrix shows that none of the events were classified as π^0 or π^+. This suggests that we need to reconsider our model construction.

3.5.2 MLP Model Adjustments

We then consider adjustments to the construction of the MLP model to determine a more optimal model. By including "BatchNormalization" and "Dropout" (of 0.15) between each hidden layer of the MLP model, we keep electron efficiency at 99.7% but marginally improve the classification of charged pions to get a pion-rejection of 49.2%. By noting the initial Confusion Matrix and that of Fig. 13, we observed that charged pions are either classified as all π^+ or all π^-. This makes sense because the SoLID beam test configuration has no ability to distinguish charge sign, thus we decided to try collapsing -211 and 211 into 211 for a "charged pion" sample.

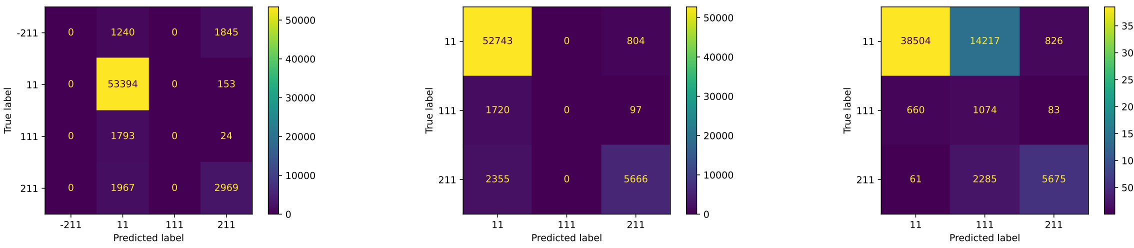


Figure 13: Confusion Matrices for various changes in configuration for MLP model construction, "data" labels, and model parameters. By including "Dropout" (left) we improve the pion-rejection slightly. By collapsing the π^+ and π^- into a single "charged pion", we improve the rejection somewhat significantly (middle). Then by adjusting the classification weights to increase the importance of correctly classifying π^0 (right), we significantly decrease electron efficiency but improve pion-rejection.

For the charged pion adjustment, we see a minor decrease in electron efficiency to 98.5% with increase in pion-rejection to 58.6%. This correction moves us past a 50:50 chance of rejecting pions with 70.6% of charged pions correctly classified. This progress is still accompanied by the complete lack of predicted π^0 events. After considering the importance of pion reject, we decided to change the event weighting to better reflect the need to reject pions and make progress in removing some of the π^0 classified as electrons.

After some initial tweaking the weights such that the model output was reasonable, we determined event type weights of e:1.5, π^0:20, and $\pi^{+/-}$:8 produced somewhat reasonable preliminary results. The right Confusion Matrix of Fig. 13 shows that the electron efficiency drops substantially to 71.9% but the pion-rejection is greatly improved to 92.7%. These results do not reach the $\sim$90% electron efficiency and $\sim$98% pion-rejection quoted by the SoLID re-Conceptual Design Report (PreCDR), however, they are in not terrible given the lack of momentum-selection for the SoLID beam test hardware configuration.

While none of these models show particularly phenomenal results, the only cuts in these samples are the scattering angle, thus we could look to bridge the gap between ML and classical cuts by obtaining baseline "cleaning" cuts to improve the efficiency and rejection for π^0. While the PID effort remains a work-in-progress, we have a baseline high-energy electrons from the simulation, which is given by a ShowerSum > 12000 integrated-FADC. This will be used in later results.

4 Cherenkov Characterization

A primary focus of this analysis has been to characterize the Chernekov detector used in this SoLID beam test. To best accomplish this, we analyze the signal from various runs with differing targets, rates, and triggers. The diagram in Fig. 14 shows the Cherenkov detector used for the beam test with its dimensions and components.

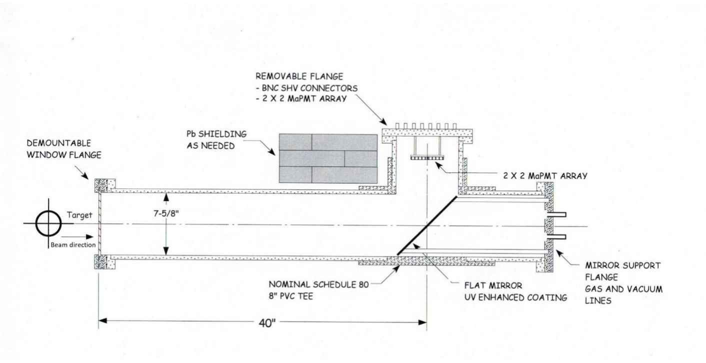


Figure 14: Diagram of the Cherenkov detector from the beam test

4.1 Principle of Cherenkov Radiation

When charged particles pass through a medium with speed $v > c/n$ with c the speed of light in vacuum and n the index of refraction of the material, they emit Cherenkov light. By careful selection of the medium, one can tune a detector such that only charged particles of specific momenta emit Cherenkov light. For the SoLID Cherenkov light gas detector, we use CO_2 to so that high-energy electrons will trigger a response in the detector, while pions will not result in Cherenkov radiation. The distribution of Cherenkov photons forms a cone (or disk) distribution with the half-angle angle of the Cherenkov cone θ_c set by

$$\cos\theta_c = \frac{1}{n(E)\beta} \tag{3}$$

where the index of refraction n, photon energy E, and $\beta = v_p/c$ with velocity of particle v_p. Following the PDG Section 35, the number of photo-electrons N_{PE} detected by a device of radiator length L is given by

$$N_{PE} = L^2 \frac{\alpha^2 z^2}{r_e m_e c^2} \int \epsilon(E) \sin^2\theta_c(E) dE \tag{4}$$

where $\epsilon(E)$ is the efficiency of Cherenkov light and the subsequent photoelectron conversion and detection by the photomultiplier tube, z is the charge of the particle being detected, and $\alpha^2/(r_e m_e c^2) = 370\text{cm}^{-1}\text{eV}^{-1}$ with r_e is the classical electron radius. For electrons of $p > 17$ MeV/c in CO_2, the value of $\theta_c \approx 1.3°$.

4.2 Cherenkov Signal

As seen in the left frame of Fig. 16, the signal distribution depends on the trigger used. For TS1 and TS4 (blue and green), we see a clear Single-Photo-Electron (SPE) peak, which contrasts to TS3 (purple), which has a much less clear peak, thus we fit the SPE peaks using TS1 and TS4. By following a similar fitting procedure as for the Preshower, Shower, and Scintillator MIPs, a fit is performed on each channel to determine the SPE peak for use in scaling the individual Cherenkov channel spectra. The function for this procedure is a linear combination of a decay and approximate Landau function:

$$f(x) = A(x-d)^n + Be^{-\frac{1}{2}(a(x+b)+e^{-a(x+b)})} \tag{5}$$

where A and B are normalization parameters, d and b shift their pieces away from 0, n is the exponent for the background, and a is a width parameter for the Landau distribution.

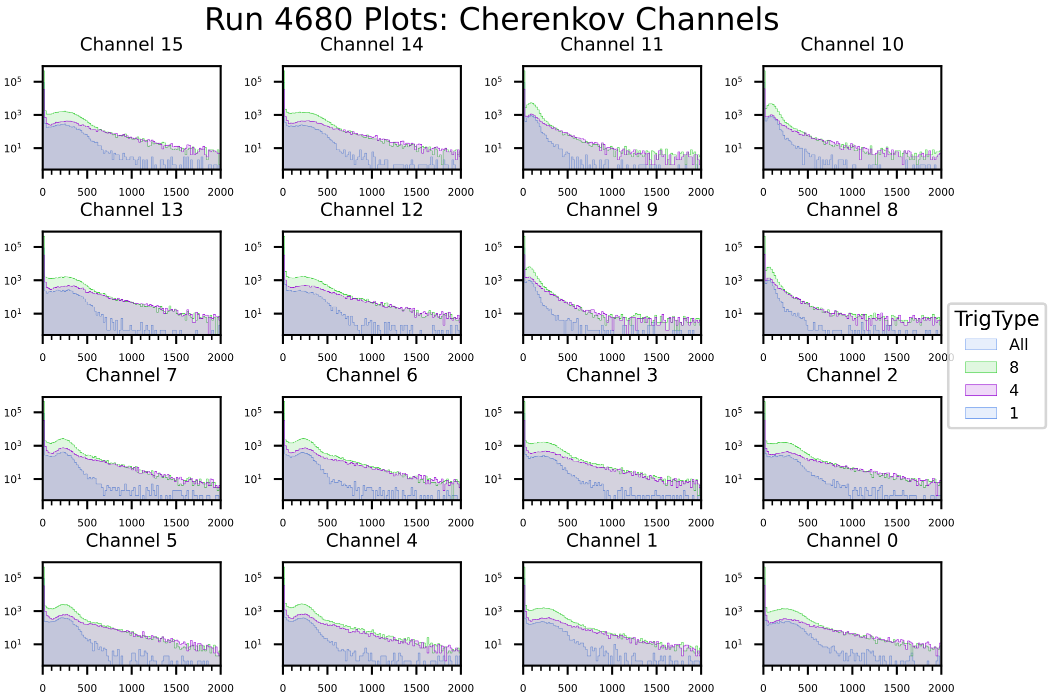


Figure 15: Here we observe the signal separated by trigger type for each of the Cherenkov readout channels.

Selecting events for TS1 and TS4 in Fig. 15, each channel is run through the initial fitting procedure to determine the most appropriate binning, range, and initialization parameters. Due to substantially different single-PE peak locations, this portion requires individual tailoring for each block. The results of the procedure for Channel 10 are shown in middle frame of Fig. 16.

After fitting the SPE peaks, we can scale the signal for each channel's SPE peak to a value of 100 using:

$$Scaled = FADC \times \frac{100}{SPE\ Peak} \tag{6}$$

This method assumes that an integral value 0 corresponds to 0 signal. Additionally, this does not perfectly account for the fact that the ratio between Landau width and position do not agree perfectly between each block. This requires more detailed study, however, the results are reasonably close enough for the scope of this analysis.

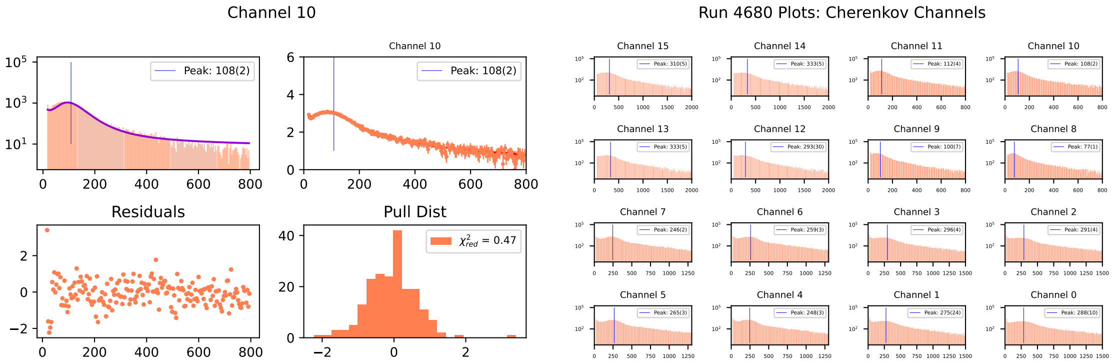


Figure 16: Fitting procedure for Single-PhotoElectron (SPE) Peak Cherenkov Channel 10 (left) and fitted SPE values for all 16 channels using triggers 1 and 8 (right).

4.3 Channel Analysis

After normalizing the Cherenkov signal for each channel, we move to more detailed channel analysis with a particular emphasis placed on applying a coincidence requirement to the Cherenkov channels. After this, we analyze the distribution of total channel signal sum as well as number of fired channels versus these channel-wise sums. Given that the electron's Cherenkov signal should consist of a cone covering 3-4 channels, thus the number of photoelectrons N_{PE} should be greater than the number of channels N_{Ch}: $N_{PE} > N_{Ch}$. While the background consists of random SPEs, thus not necessarily in adjacent channels, and furthermore $N_{PE} \approx N_{Ch}$.

The analysis from the Cherenkov beam test mentions that a "simple coincidence requirement that at least two nearby channels fire can then drastically reduce the accidental single-photon background from random sources [π^0 decay, etc]." This sentence leaves three main points of ambiguity:

1. Does this require two other channels, in addition to the central one?
2. What does nearby mean?
3. What is the threshold for firing?

The first is self-explanatory, however, the second could mean just ones that "share a side" or also include ones diagonal to the channel in question. The third requires the distinction between non-zero entries and requiring a threshold value.

After careful discussion of these points, we concluded that coincidence requires:

1. Two channels fire
2. Any channel surrounding around a channel
3. Require a FADC (pre-scaled) threshold of 40 to be above the background.

After deciding the conditions, we implemented these methods by looping through each channel and checking that its neighbor fired (see Fig. 17). If a channel fails the coincidence requirement, then the value is set to zero for the rest of the analysis.

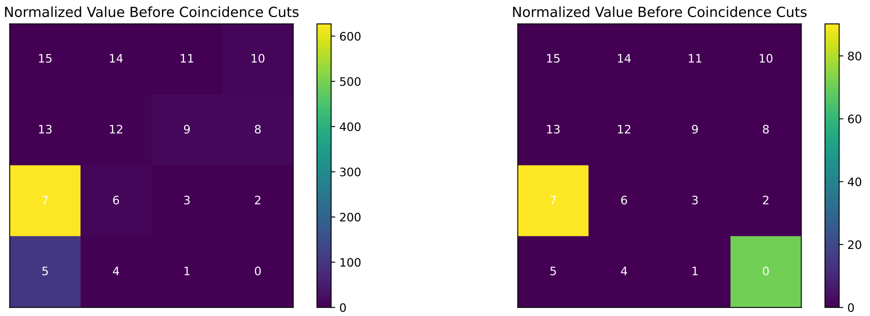


Figure 17: Overview of Cherenkov Coincidence
The left plot shows two channels (5,7) passing the coincidence requirement, while the right shows two channels failing the coincidence requirement.

4.4 Cherenkov Cuts

By following the procedure delineated, we obtain the plots of Fig. 18 and a more cleaned sample for electron candidates. We start by requiring integrated-FADC values for each channel are greater than a threshold (40 was used for these plots). Then we add the coincidence requirement, which removes single channel fires and cleans some of the remaining noise. By placing a center cut on the GEM (GEM$_x$ & GEM$_y$ < 0.03 m), we clean the sample significantly. This comes from the implicit requirement that there exists a GEM track, thus removing a significant number of charge-neutral events, as well as good events that do not have tracks. With the GEM cut, the statistics drop dramatically but the majority of the remaining events fit within the expected formation (see Fig. 18 bottom right). Using these plots, we seek to define good electron samples for ECal analysis.

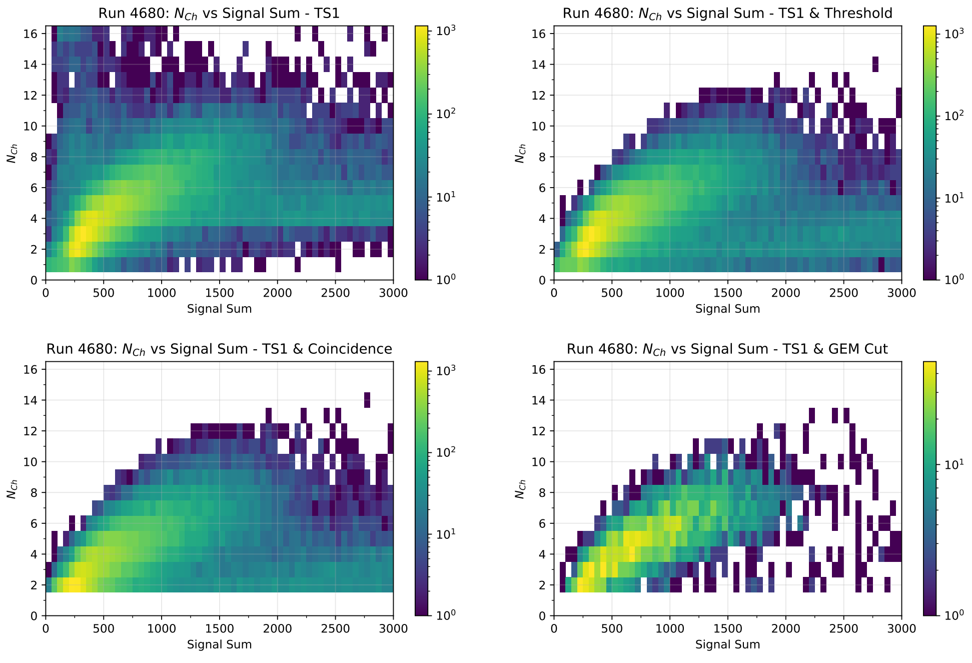


Figure 18: Overview of Cherenkov procedure

5 Position Determination of Events

An vital component of this analysis is ensuring that the detector alignments are within acceptable parameters. Therefore, we study the particle tracks and positions in multiple detectors, specifically the Shower Modules, GEM detectors, and Cherenkov detector. By comparing them to a baseline, then against each other we can more conclusively state that the detectors are aligned, thus we have a stronger footing for conducting event-level analysis.

5.1 Shower Clustering

For the Shower modules, we ensure the accuracy of the clustering algorithm by applying cuts to the three modules (left, top, and right). Given three modules packed in a triangular formation, the distribution of clustering position should look like a triangle. Then by applying high cuts in each module, we should see triangle form in each corner of original distribution.

Module	x	y	z
Left	-5.28	-3.05	3
Top	0.00	6.10	3
Right	5.28	-3.05	3

Table 5: Coordinates of the Shower Clustering centers for the three Shower modules. These are used to created a weighted-average position for the Shower to be compared against other detectors.

The clustering function works by determining a weighted-position value for each event based on the energy deposited in the Shower Modules. With three centers defined in Table 5, we follow a few steps to remove contributions from noise then find the weighted value. We obtain individual, scaled module values m_i by dividing the integrated-FADC values f_i by the MIP values s_i for each module in

$$m_i = f_i/s_i \tag{7}$$

We determine the array of weights w_i by dividing the array of module values m_i (where i iterates over the three modules) by the ShowerSum value $= \sum m_i$ for the event

$$w_i = 3.6 + \ln(m_i/\sum m_i) \tag{8}$$

Then we ensure that the weights are greater than zero by

$$W_i = w_i \times (w_i > 0) \tag{9}$$

To determine Shower Clustering array X_j (where j iterates over the x, y, and z clustering coordinates), we sum over module for the module position array M^{ij}, where element $M^{1x} = -5.28$.

$$X^j = W_i M^{ij} \tag{10}$$

This procedure essentially removes values in modules that fall below a certain threshold, then calculates the weighted position for energy deposited in the Shower. This threshold is governed by the 3.6 value given in Eqn. 8.

To test the clustering, we followed the procedure shown in Fig. 19, where a series of cuts were applied using shower module values. Given that the clustering is a weighted average, we can apply high cuts on individual shower modules, which should only allow events that deposit large amounts of energy in that module. For instance, applying a high Shower Top cut should only result in events in the top half of the triangle, which is observed in Fig. 19.

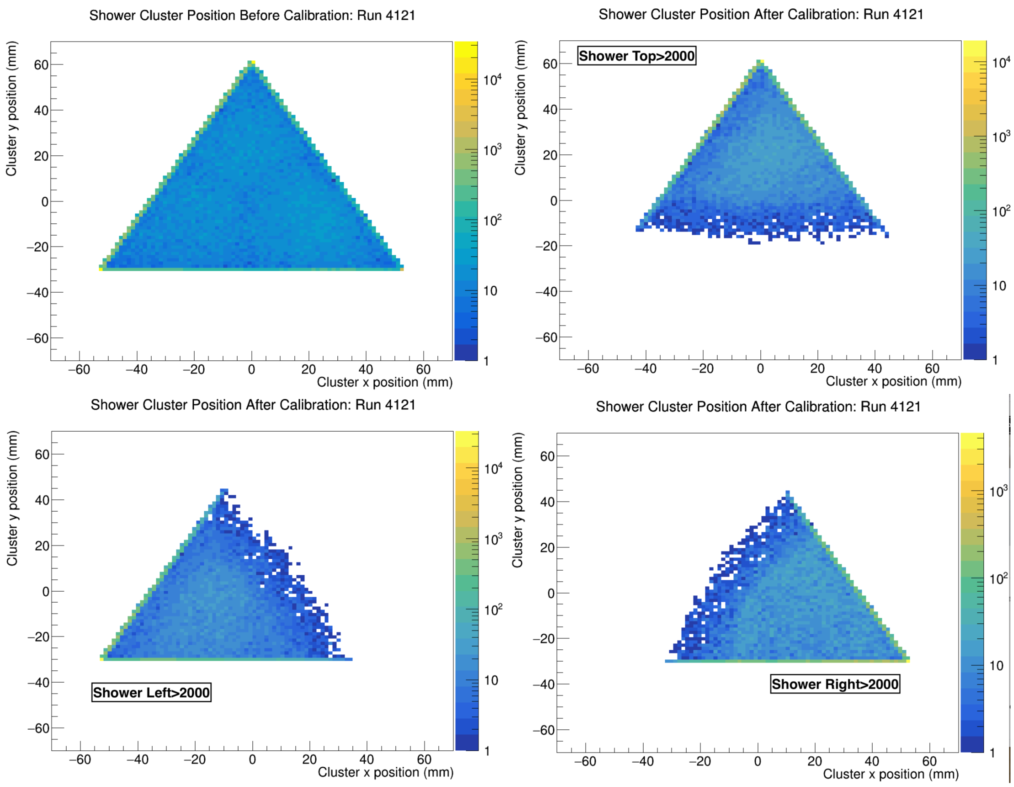


Figure 19: Shows the Shower Clustering position for the initial distribution of energy deposited in the Shower modules (top left). The other three show the the distribution after placing cuts to ensure that the clustering algorithm works as intended. For instance, the bottom left plot shows the events remaining after a high Shower left cut, where the only remaining events are near the bottom-left corner of the triangle.

5.2 Cherenkov Position & Alignment

To check the positioning of the Cherenkov detectors, a series of cuts were applied to Run 1421 (8°) to ensure that the mirror was properly aligned. Using the methods described in the previous section for the heatmap, we enforce the coincidence requirement and sum over the number of hits (not signal sum) to create the baseline heatmap shown in Fig. 20.

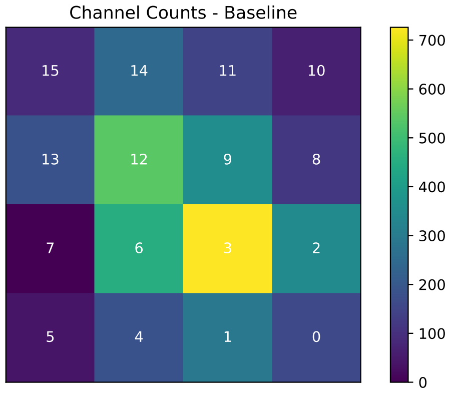


Figure 20: Baseline for Cherenkov Quadrant tests using Run 1421 using cuts of ShowerSum>18000, SC-D>200, and Shower Cluster x & y between [-2,2].

By inspecting the plot in Fig. 20, one finds that the central channels have the highest values, while the outer ones have a lower count. This makes sense because the small angle of the Cherenkov cone. However, the point of this plot is to create a baseline for studying with cuts.

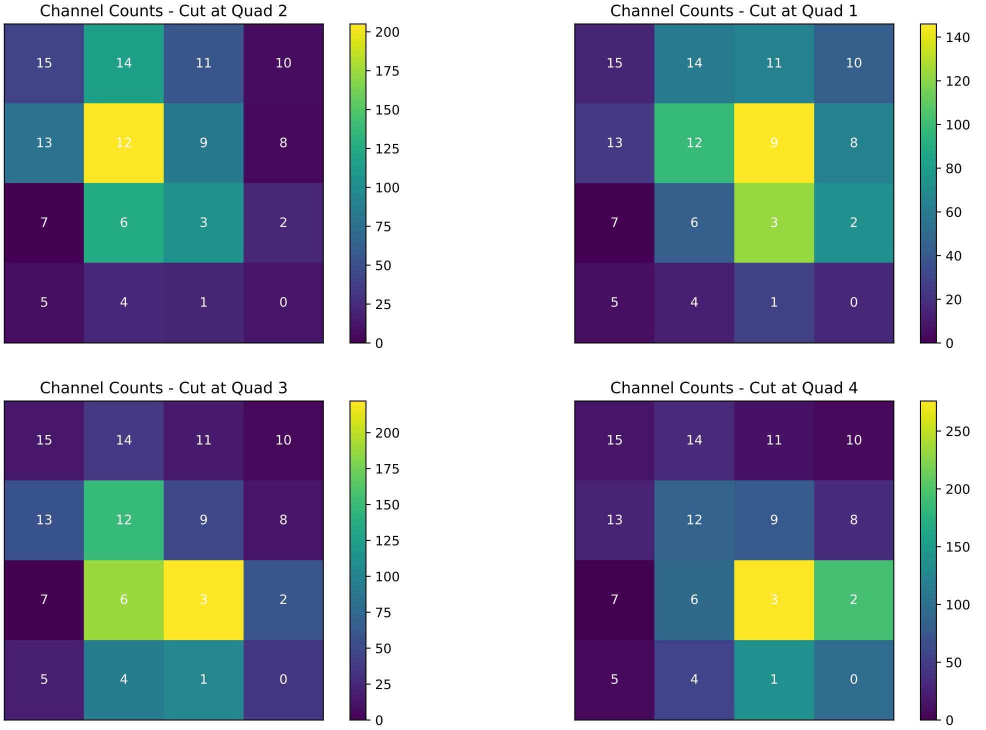


Figure 21: Test of applying ”quadrant cuts” in the Shower Clustering (see Fig. 19) to observe the pattern in the Cherenkov heatmap. This shows that the quadrants of the Shower Clustering match with those of the Cherenkov channels. Quadrant 3 suffers from limited statistics, as well as the shape of the Shower Clustering distribution.

Since we have already establish that the Shower clustering method works reasonably-well, we can use the clustering position to check the Cherenkov position. Requiring the events to have a Shower Clustering within a certain value, we can try to correlate the shower clustering position to that of the Cherenkov. By applying ”quadrant cuts” on the Shower clustering, or requiring the position to span only one of the Cartesian quadrants at a time (i.e. $x > 0$ & $y > 0$ for quadrant 1), we create the four plots shown in Fig. 21. Here we can show that the quadrants of the Shower Clustering and Cherenkov read-out match. One observes that as we move the quadrant cuts, the hot-spot moves towards the center of that quadrant. This is very apparent for quadrants 1, 2, and 4, however, limitations from statistics result in the hot-spot not shifting fully into the center of quadrant 3 for its cut. Additionally, channel 7 was not working during these runs.

After determining that the clustering position and Cherenkov position match, we extended this method to investigate the alignment of the Cherenkov mirror. Given that the previous beam test suffered from a misalignment, we wanted to ensure that the anomaly with quadrant cut 3 of Fig. 21 was not due to a mirror misalignment. To accomplish this, we decided to make a square cut around the center of the Shower Clustering distribution starting from -2 cm to 2 cm. This should roughly correspond to tracks through the center of the Cherenkov detector, thus we should observe a clustering in the center of heatmap as we make this square smaller.

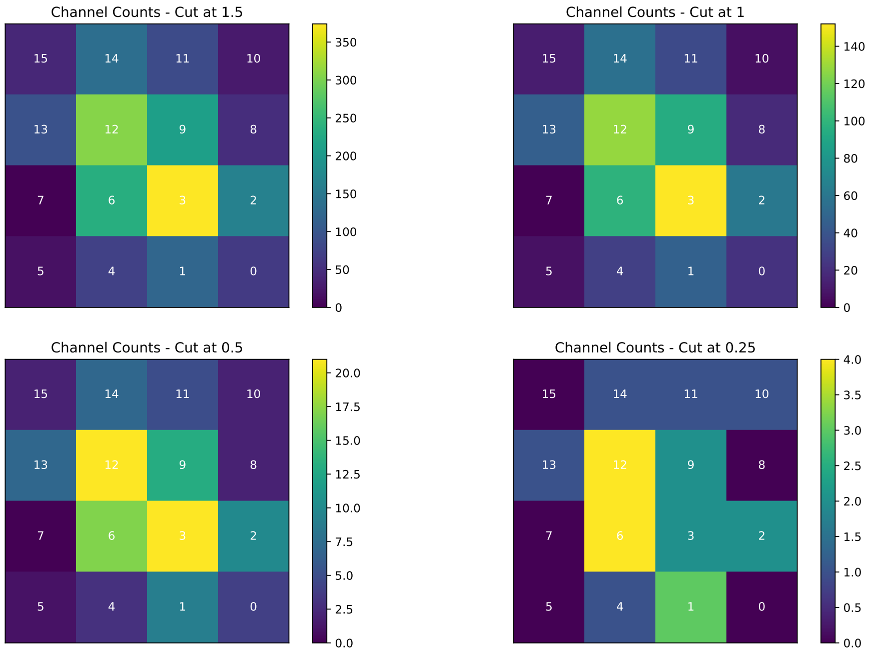


Figure 22: Test of applying a series of increasingly-finer cuts to the Shower Clustering requirement where the value starts at 2 cm for the baseline then decreases to 1.5, 1, 0.5, & 0.25 cm. For example, “ Cut at 1.5” means the shower cluster positions $|x| < 1.5$ and $|y| < 1.5$ cm.

As shown in Fig. 22, by making this square smaller and smaller, we were still able to show that the center of the events stays in the center of the map. If this center were to shift substantially, we might need to more carefully consider mirror-alignment issues. However, this ”shrinking cut” lends evidence to the mirror being reasonably-well-aligned.

5.3 GEM Tracking

While the specifics of the GEM detector and tracking associated with it are not the focus of this report, we verify the results and use the tracking information as a check for other detectors. The GEM tracking algorithm is an ongoing component of this analysis, however we present the current results.

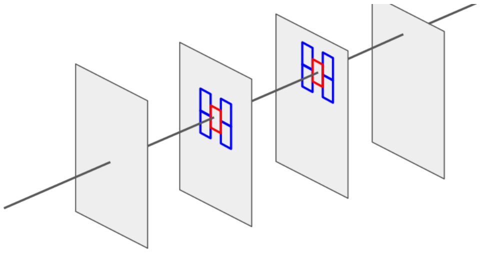


Figure 23: The GEM tracking algorithm discretizes the GEM plane, then loop through all combinations between the first layer and the last layer to brute-force all possible combinations. This results in a χ^2 value for each possible track.

As with the Shower Clustering, we need to ensure that the GEM Shower projection matches expectation. The positions for GEM Shower projection X_i (where i is the x and y coordinate) come from

$$X_i = -(x_i + 2.17 \times \tan(\theta_i)) \tag{11}$$

for the track position x_i and track angle θ_i at the first GEM detector, where 2.17 m is the drift distance from the first GEM to the front of the Shower modules.

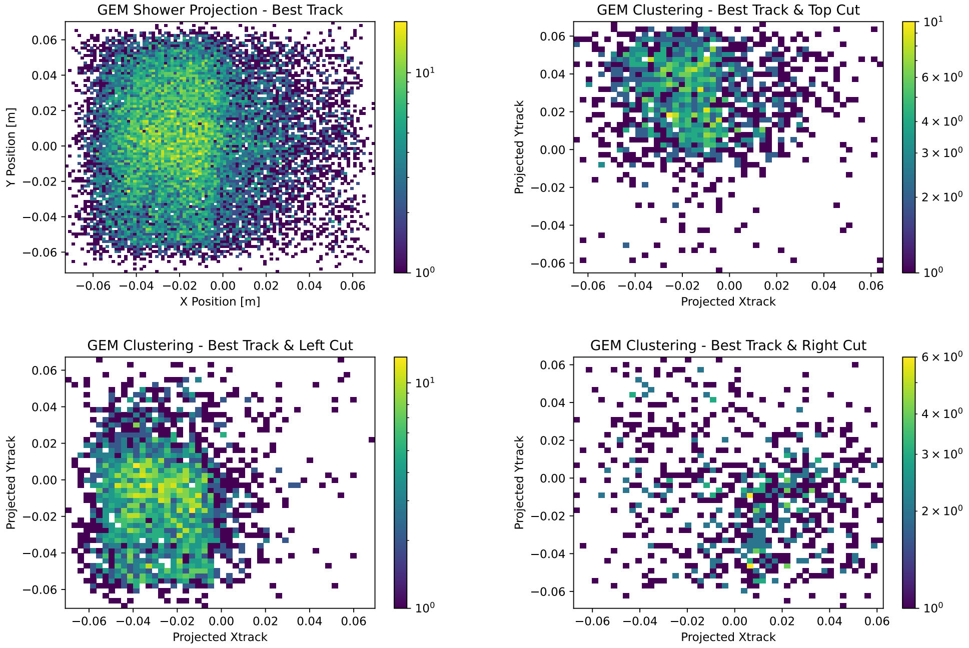


Figure 24: GEM Projection (top left) and check of GEM Projection using cuts in the Shower Modules.

Based on the results from Fig. 24, we can conclude that the GEM projection for Shower module reasonably matches expectation. The ”Right-Cut” is hindered by limited statistics in that half of the original distribution. This is likely due to imperfections in the tracking algorithm that will be improved in future iterations.

To investigate the position correlation between the GEM and Cherenkov, we apply quadrant cuts on the GEM Shower Projection to look for corresponding signals in the Cherenkov. This follows the same procedure as the Shower Clustering where we place requirements on the events for the shower clustering, then observe the results in the Cherenkov heatmap.

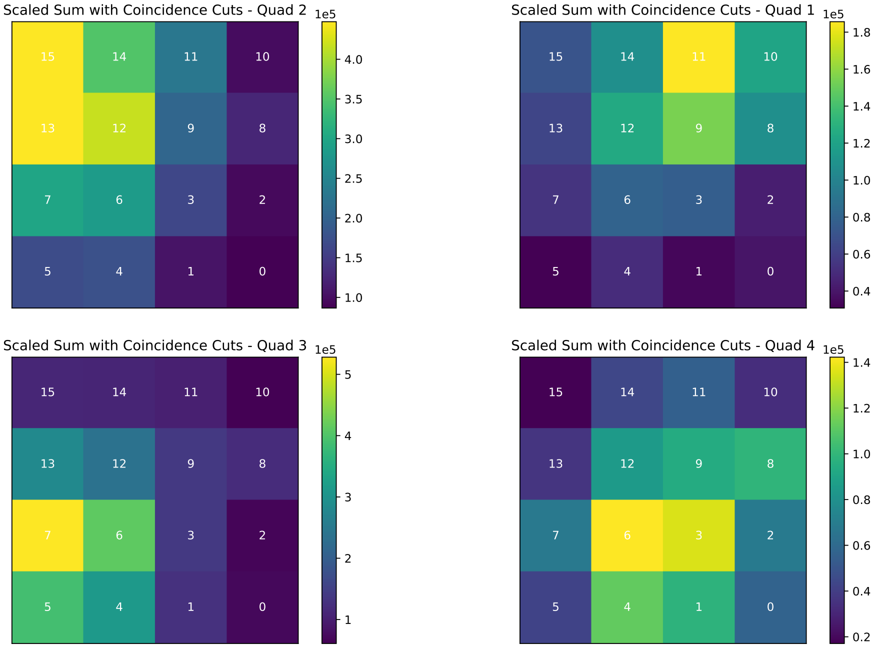


Figure 25: Quadrant Cuts for GEM Projection to demonstrate that the quadrants of GEM match those of the Cherenkov detector. Given that the hot-spot moves with the cuts, we conclude that the quadrants of the GEM Projection and Cherenkov match.

Based on Fig. 25, we conclude that our expectations align with the results because there is the associated clustering in each quadrant. Once again, quadrant four is limited by statistics and conclusive results are more difficult to draw. Support for this notion is furthered by noting the left-skewed initial GEM distribution of the top left plot in Fig. 24. Additionally, the GEM tracking still requires optimizations, thus the results are only to confirm the quadrants match.

6 Conclusion & Future Work

Now that we have established a baseline characterization for most of the detectors, our collaboration will continue to iterate through the analysis by improving GEM tracking, determine electron, pion, and photon samples, and conduct more-detailed PID analysis using simulation and data.

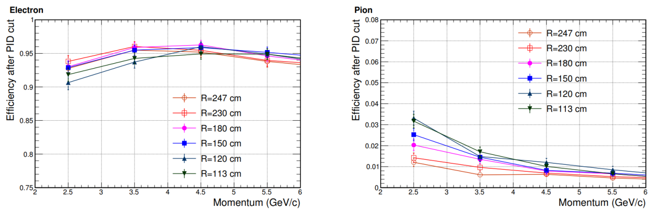
6.1 Improved Analysis

By improving the GEM tracking algorithm, we will increase the number of fully-characterized events to analyze, thus allowing us to consider more strict tracking cuts and having a larger number of events for analysis. The GEM tracking algorithm has already experienced a few iterations of fine-tuning the numerous parameters, however, the current efficiency is somewhere around 70%, depending on the beam current, trigger, target, and other run-specific variations. An increase in precision may also be possible through continued tuning of the tracking algorithm, which would permit even more careful position correlation between Shower Clustering, Cherenkov position, and GEM Shower Projection.

Determining the selection criteria for electron, pion, and photon samples remains an on-going and critical process. By careful selection of triggers, we have a starting point for making these selections. However, from checking the pseduo-trigger cuts in the simulation, we can easily see that these samples are far from pure and will require somewhat extensive cuts to provide us with anything resembling clean samples. This process will continue for the rest of May into June because clean samples are critical to proceeding towards PID analysis.

6.2 PID Analysis

After determining our samples, we will move from an exploratory to preliminary stage of PID analysis for this beam test. We are currently developing an initial phase of cuts to select the three types by manual selection of conditions. The end goal for our analysis is to make something resembling the left plot of Fig. 26, however, we do not have momentum selection. This will result in something resembling the plot collapsed in one direction. Additionally, we look to work towards PID performance metrics reported in the table of Fig. 26.



	Desired performance
π^- rejection	$\gtrsim$[50:1]
e^- efficiency	$\gtrsim$90%
Energy resolution	$< 10\%/\sqrt{E}$
Radiation resistance	$\gtrsim$400 kRad
Position resolution	$\lesssim$1 cm

Figure 26: SoLID PID curves for electron efficiency and pion-rejection (left) and overall performance goals for SoLID (right). Figures from 2019 SoLID pre-Conceptual Design Report. We plan to investigate how the beam test SoLID aligns with the projections for the full-scale detector.

We approach PID analysis by using the simulation as a guide, however, the procedure of manually checking cuts is slow and inefficient. Thus we plan to move to more advanced methods in the next iteration by employing the use of more advanced algorithms, in addition to Machine-Learning (ML) techniques. These may allow us to better identify particle type, potentially increasing electron efficiency and pion-rejection. Given that ML algorithms are more capable of working with high-dimensional data than humans, we will investigate some classification methods employed by other collaborations in high-energy and nuclear physics.