HPS Cluster Reconstruction Studies (2021 Data)

Rory O'Dwyer
Cameron Bravo
Tuesday Analysis Meeting 10/31/23

Introduction

Cam and I performed extensive Hit Reconstruction optimization studies

We determined that there were ~5 effects accounting for <1 percent each misconstructed hits in these studies; we also arrived at parameter optimization and DT tree optimization strategies that could address these effects.

We moved onto baseline optimization; this, alongside alignment, was determined to be (by far) the most necessary change to hit reconstruction. This has been addressed with submissions to the database.

We now hope to perform a complete, exhaustive analysis of cluster reconstruction; the first step will be to plot clustering variables and implement dead channels.

How the NearestNeighborRMSClusterer Works

```
For hit in rawhits:
        Put hit and channel into map
        if signal/noiseRMS>nghbrThresh:
                Add to clusterable set
        if signal/noiseTMS>seedThresh:
                Add to seeding set
```

What We Change w/ Dead Channels:
Don't add to either of these if the channels in question are dead

```
for seed hit in seed_clusters:
        Remove seed from clusterable
        Add seed to unchecked
        Instantiate a cluster object

        while(unchecked unempty):
                pop unchecked;
                if(unchecked>thresh, is in neighbor, and in time):
                        addtocluster
                        put its neighbors into unchecked
```

What We Change w/ Dead Channels:
In this step, we just skip over the dead channels and add the channel 1 or 2 over to this list.

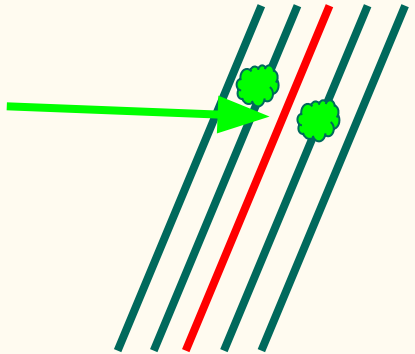
Variables we explore in Preliminary Studies

Hit Number in Cluster, On, Off Track, and NTD

Cluster Charge, On, Off Track, and NTD

Cluster Position On, Off Track, and NTD

Cluster Distances on Same Sensor All and NTD



Cluster Splitting Across Dead Channels

NTD Cuts:

For NTD we search left and right of a seed for a dead channel. As implemented, we expect NTD clusters to have significantly less cluster width and a peak at small cluster distances. This would arise from charge sharing on the sides of a dead strip.

Study Source Files for Cluster Study.

We performed studies on low lumi run 14166 for data and MC, and also a high lumi data study on 14552.

All the plots, separated by plot type, layer, or all put together, are located in the following websites:

High Lumi Low Lumi Low Lumi MC

We used HPS_Run2021Pass0_v0 for Low Lumi MC and HPS_Run2021Pass1_v3 for both data runs. The steering file for the evio to lcio conversion was PhysRun2021_pass0_recon_evio. We then kept only raw hits, and SiClusters on and off Kalman Tracks.

I can provide the realpaths to the MC file location upon request.

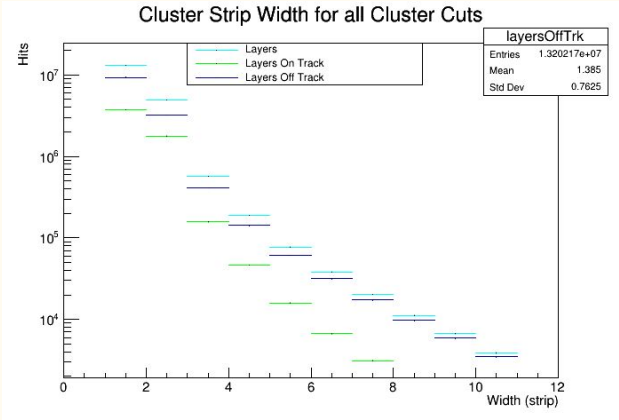
Cluster Widths (Number of Raw Hits in a Cluster) 14552

Here are the distribution of cluster widths.

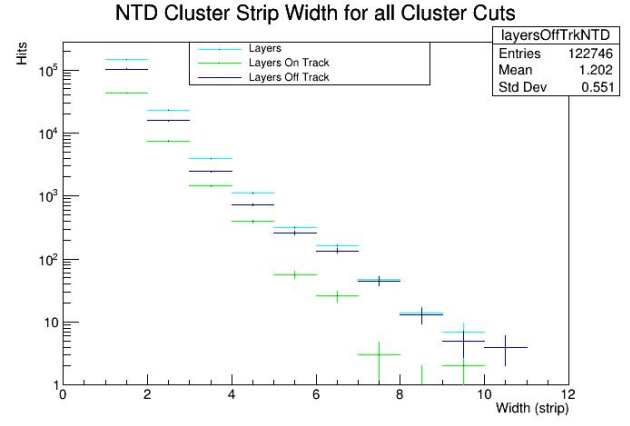
On top we have cyan for all clusters, blue for off track and green for on.

On the bottom we include the constraint that these clusters must have a seed adjacent to a dead channel.

You can see from the green histograms that the relative abundance of two hit clusters (vs. 1) is markedly different if you are next to a dead channel.



ALL OF THEM



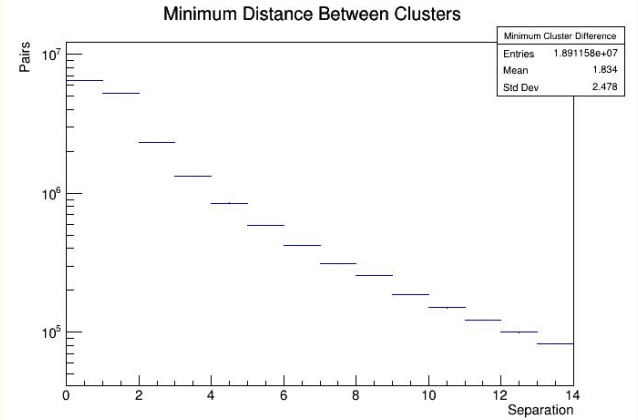
NEXT TO DEAD

Cluster Distances (between Clusters) All and NTD 14552

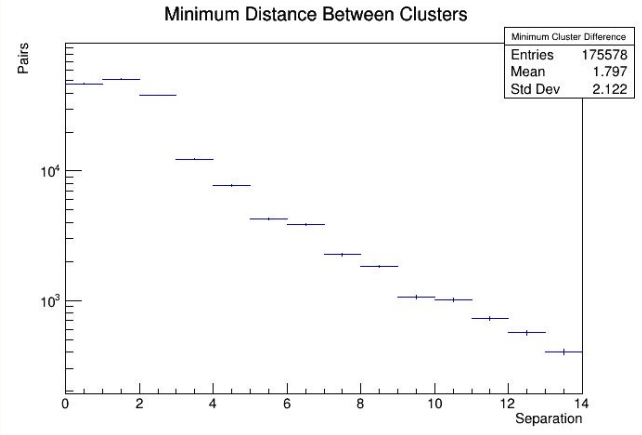
In these plots, we plot (for clusters on the same sensor) the distribution of spacings between clusters.

The Lower plot conditions on one of the sensors being next to a dead channel.

There is a clear peak at 3 for this lower plot (and some clear substructure). This large increase from the background is consistent with cluster splitting.



ALL OF THEM



NEXT TO DEAD

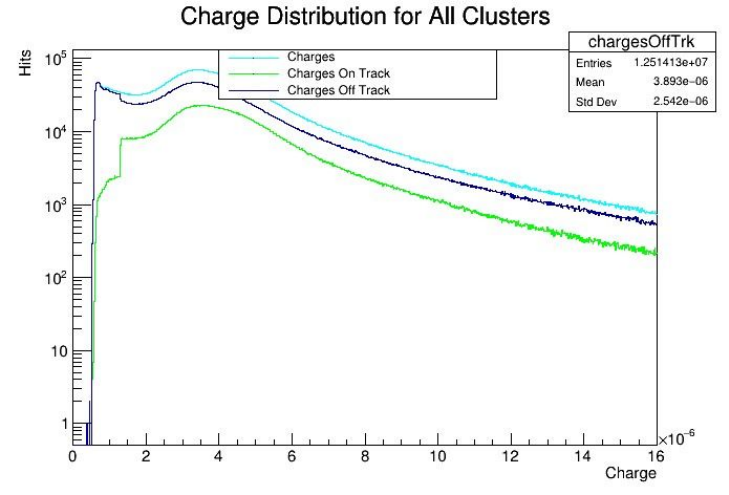
Cluster Charge Distributions 14552

With the same color scheme, we now plot charge distributions for the clusters.

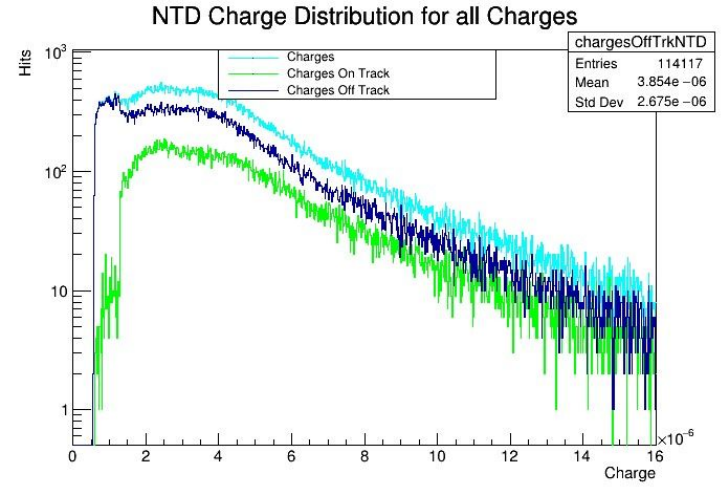
There is an abrupt drop in number of hits for charge less than ~1.5e-6 for clusters on track. We are still thinking about why this occurs.

For the NTD distribution, the peak in charge at ~4e-6 is seemingly gone.

The Units of these distributions are nano-coulombs (4e-6 nC ~ 26,000 e-)

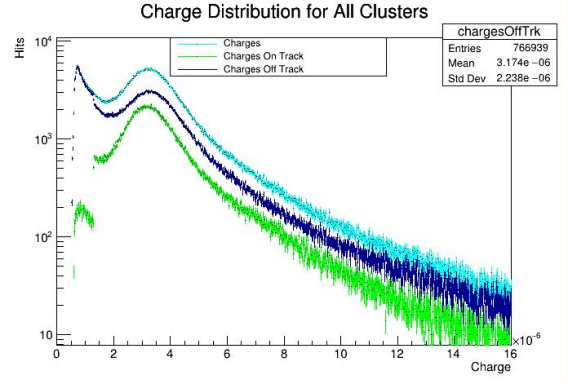


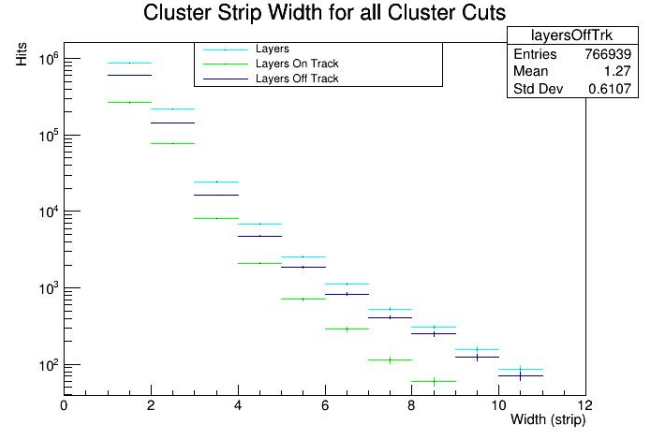
ALL OF THEM

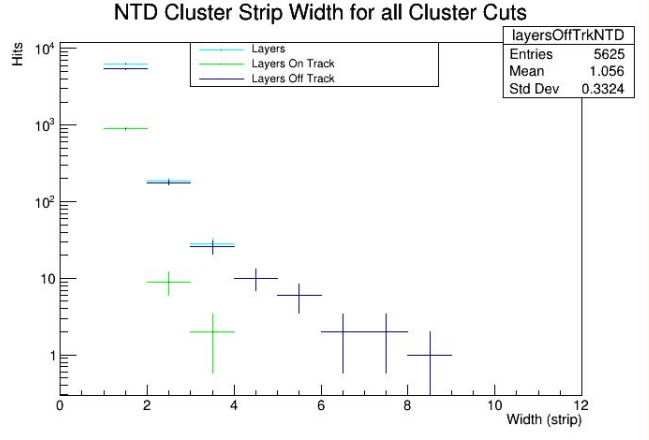


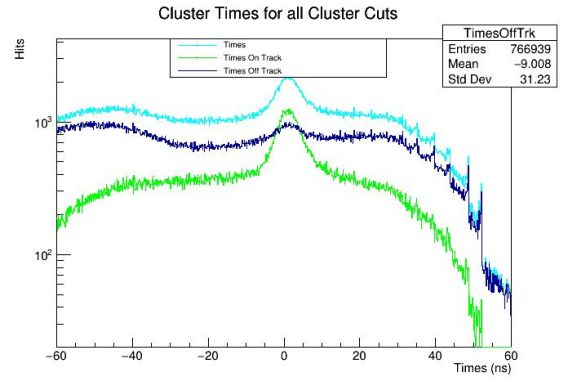
NEXT TO DEAD

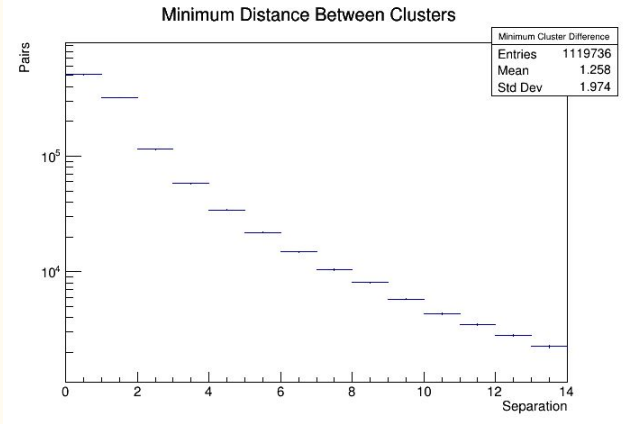
Front Detector Distributions (Namely l1m1)

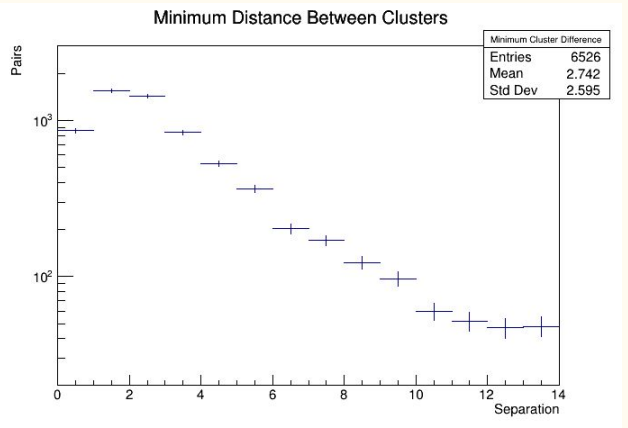




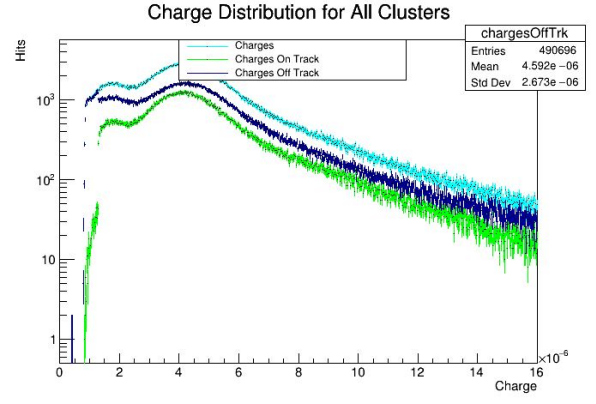


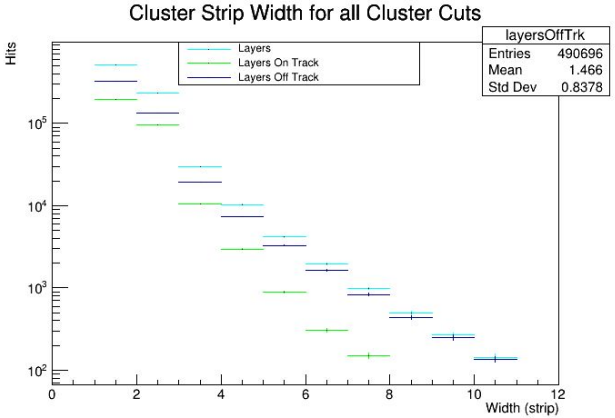


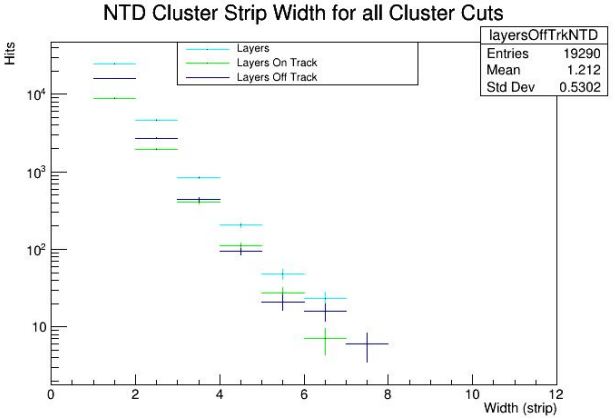


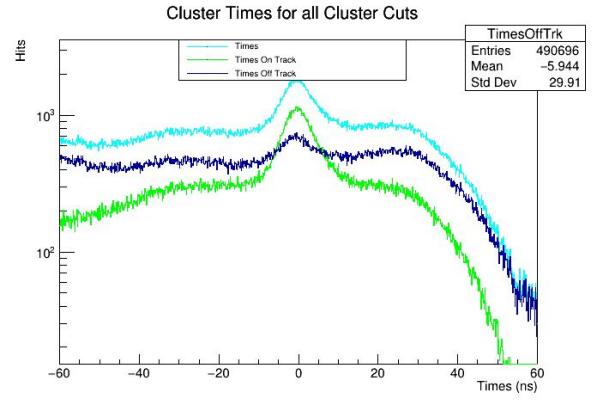


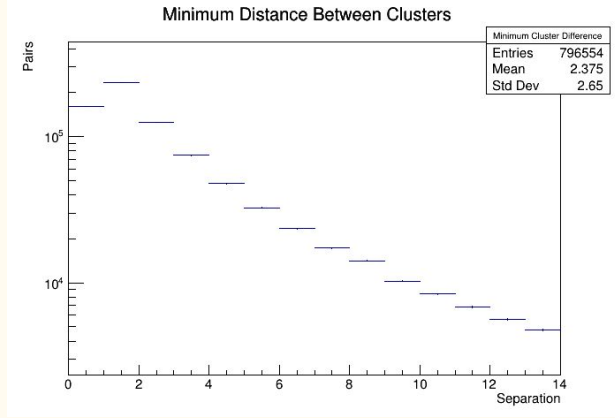
Back Detector Distributions (l4m1)

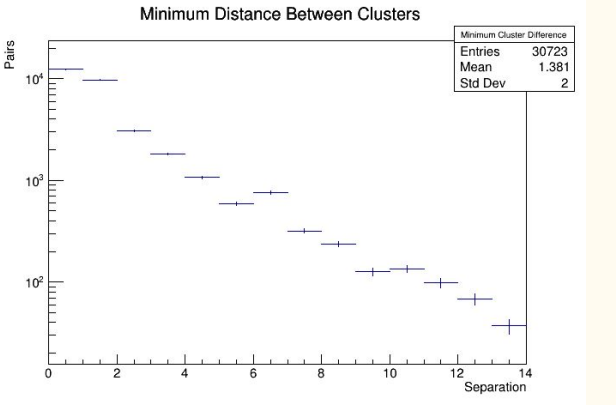












LOW LUMI DATA AND MC COMPARISONS

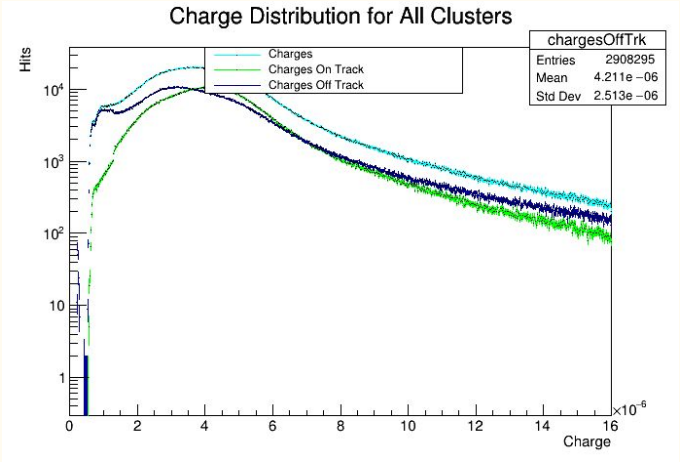
Charge Distribution for MC and Real Data

Here is the cluster distribution for all sensors with comparable statistics for MC on top and data on bottom.

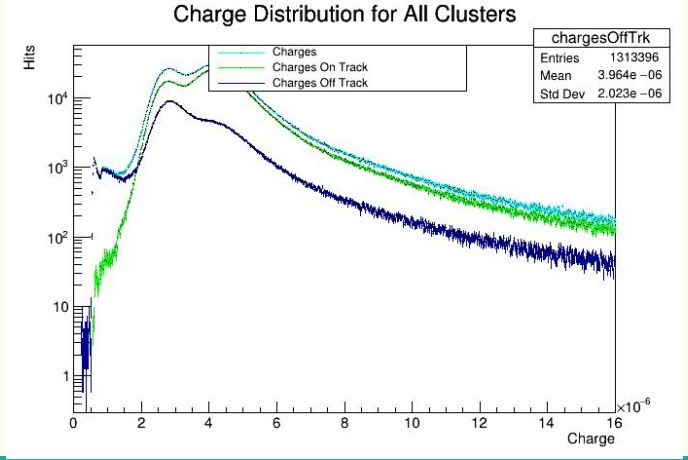
I believe the bottom distribution is bimodal because the front and back layers have different charge distributions

The MC has much more uniform charge distributions per layer.

MONTE CARLO:

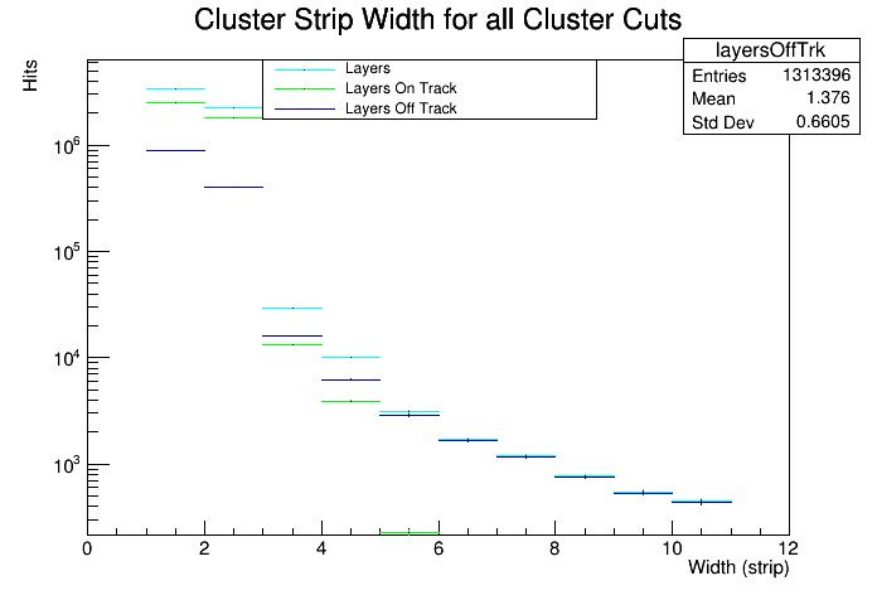


LOW LUMI DATA:

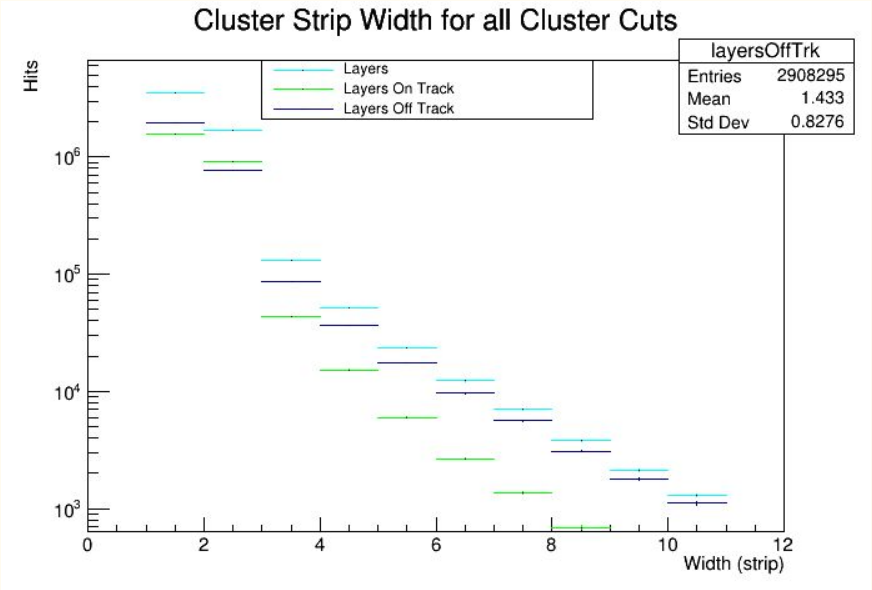


Cluster Strip Width for MC and Real Data

MONTE CARLO:

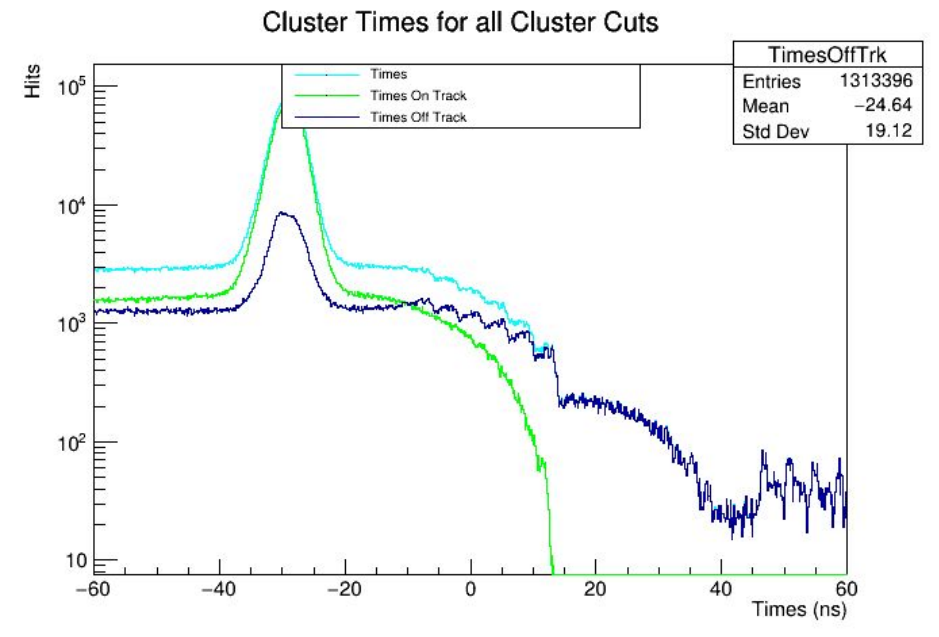


LOW LUMI DATA:

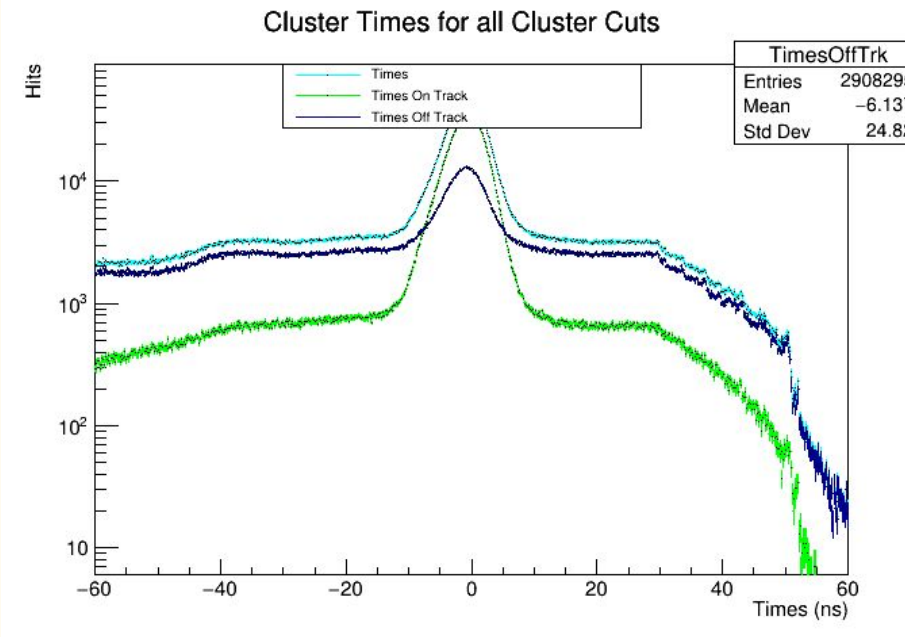


Cluster Time for MC and Real Data

MONTE CARLO:



LOW LUMI DATA:



Next Steps

We have developed an initial cluster analysis processor suited towards the quick production of plots; it will require work to be properly integrated into the hpstr github.

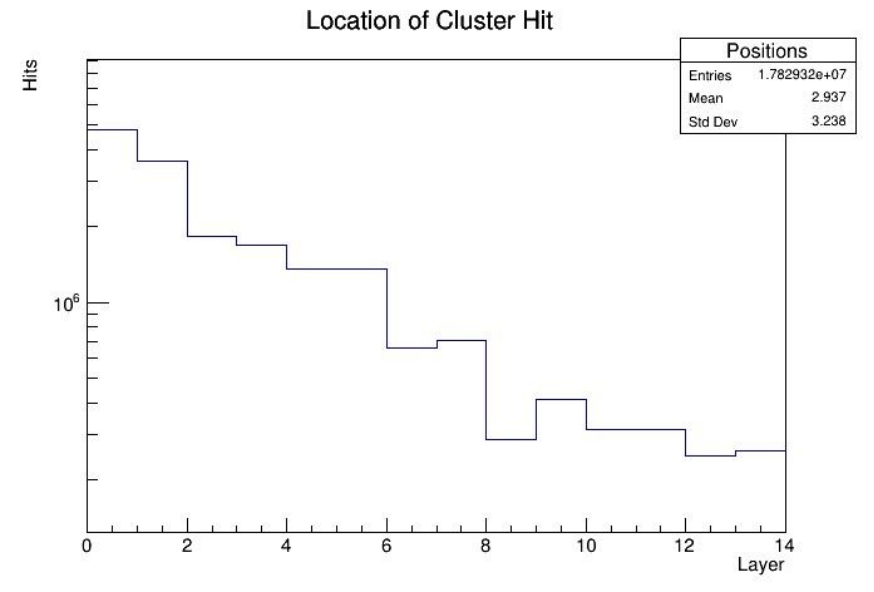
We have changes in mind that could implement dead channels into clustering reconstruction. Further work is required to see how this would affect things like cluster weighted times, etc.

Any plots you would like to see in these studies and on the html I would appreciate. I think I have a nice assortment already, but advice is always welcome.

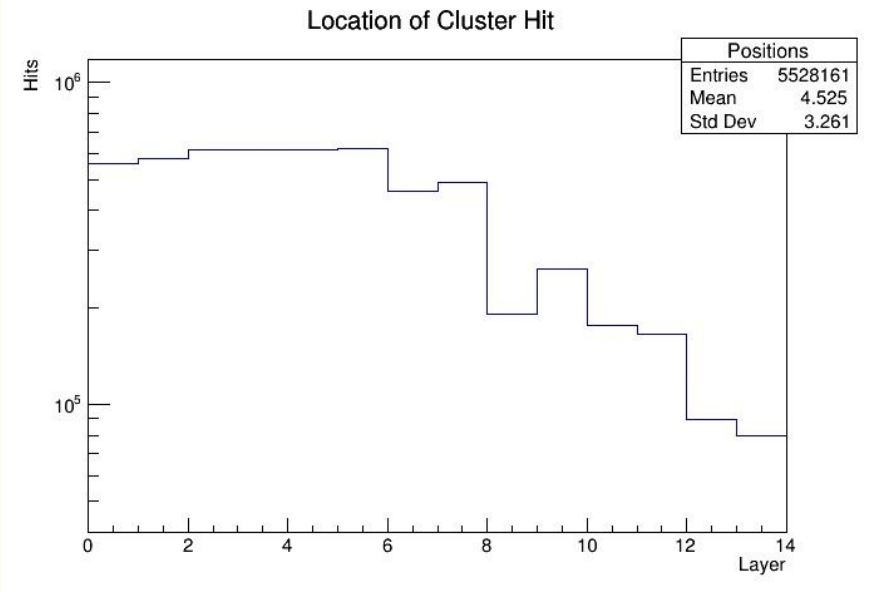
BACKUP

Cluster Position Distributions

ALL CLUSTERS (OFF AND ON TRK)

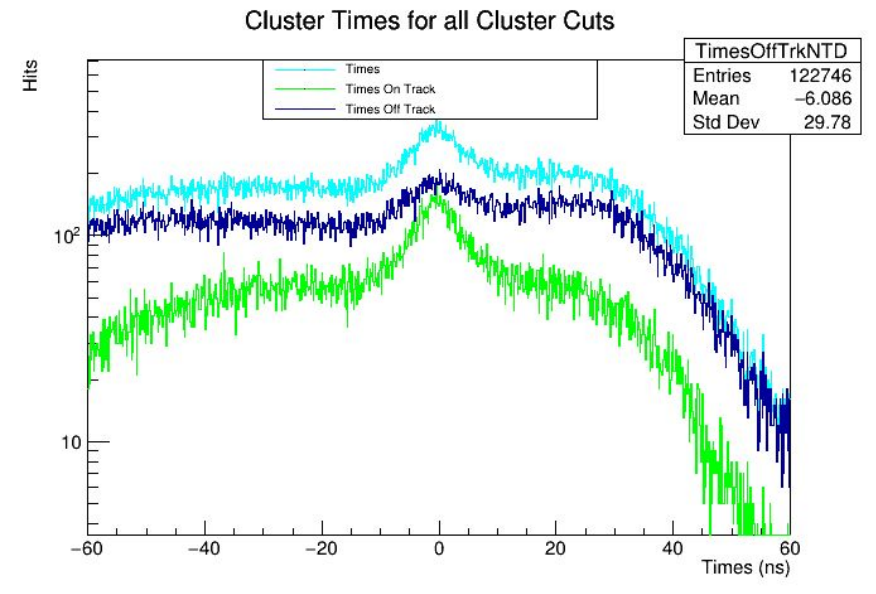


ON TRACK CLUSTERS



High Lumi 14552 Time Distributions.

NTD CLUSTERS



ALL CLUSTERS