


ECAL DigiHits Studies in ND-GAr using AnaTrees: Using my locally made sample

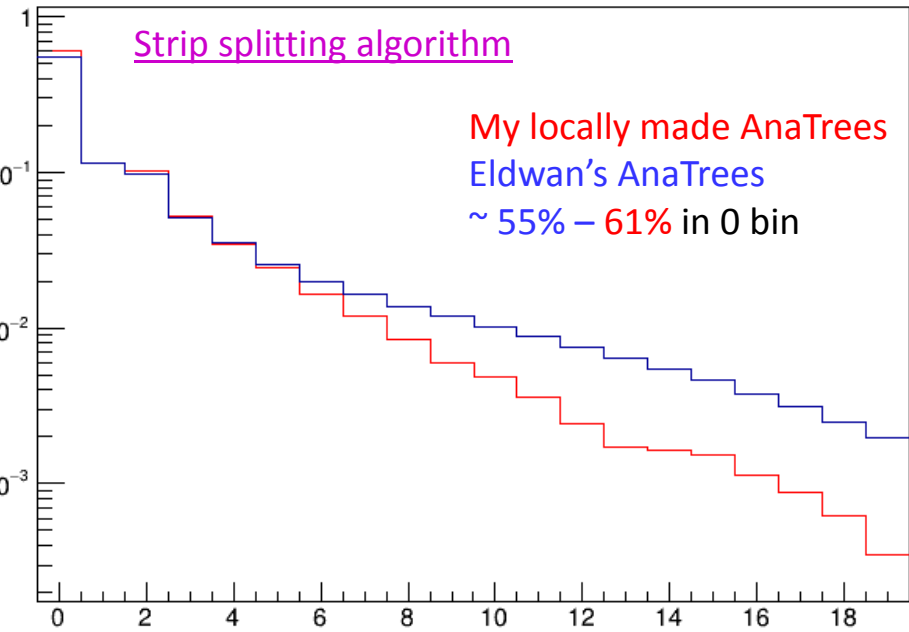
Lorenz Emberger, MPP
Vivek Jain, SUNY Albany
Aug 1, 2021

Many thanks to Eldwan Brianne for implementing the geometry, Tom/Leo for helping me generate my test sample

Plots in this talk are mainly from histograms embedded in the Digitization code – but some are from AnaTrees

Introduction

- In the past, we had noticed that many Reco Hits were matching a single Sim Hit 
 - We now understand this – it is coming from the Strip Splitting algorithm. If we use the output of the standard sipmhit finding algorithm, then we have either 0 or 1 Reco Hit matching a Sim Hit.
 - However, both tiles and strips seem to be affected. Why?? This is weird
 - Running the clustering algorithm with the standard sipmhit algorithm, number of clusters/event is a bit smaller than before (~13.4 v. 14.9), but now have fewer clusters with $E < 5$ MeV (4.5% vs. 36%) – sort of makes sense - have not explored this any further
- In the process of understanding these issues, we also looked at Digi Hits
 - Here we found one oddity, which we don't understand as yet.
 - Essentially, ~9% of Endcap strip SimHits are not being digitized, because the code seems to get “confused” as to where exactly the hit is
 - According to Lorenz, this should never happen



Reco hits matching a Sim Hit – based on CellID

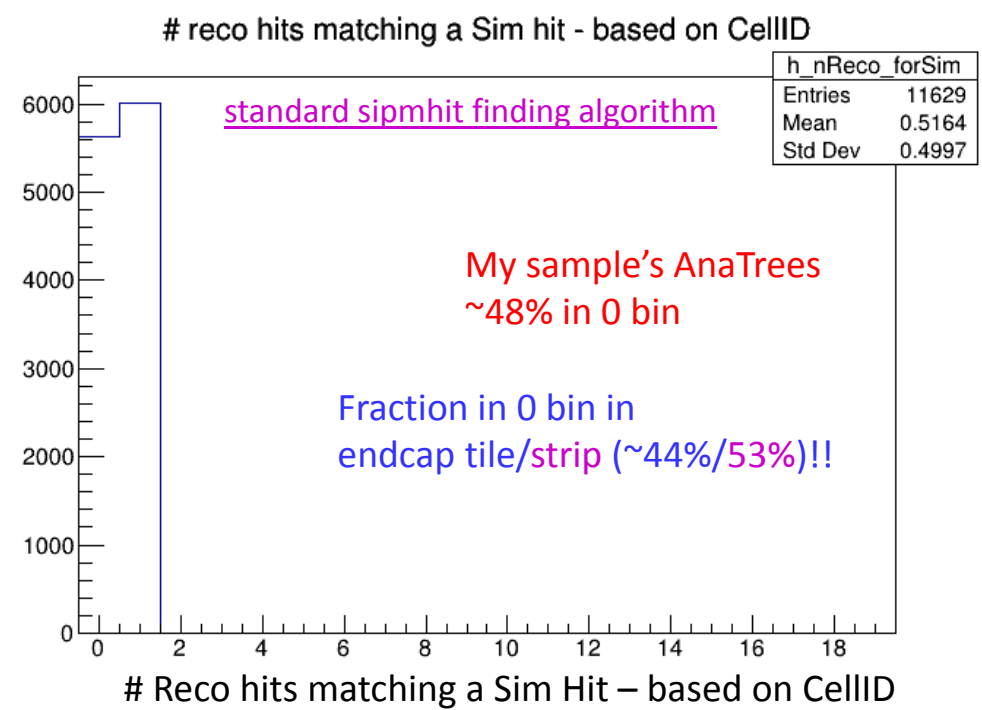
What is weird is that both endcap tiles and strips have multiple Reco matches to 1 Sim Hit. Why should strip splitting algorithm affect tiles???

Fraction in 0 bin in endcap tile/strip (~55%/83%)!!

(Clean separation of tiles and strips in the barrel requires extra geometric cuts – didn't do it here)

According to Lorenz, the 0 bin is due to a cut made during reco, i.e., the 0.25 MIP cut on hit energy

Events in TPC fiducial volume
(different number of events in different plots)



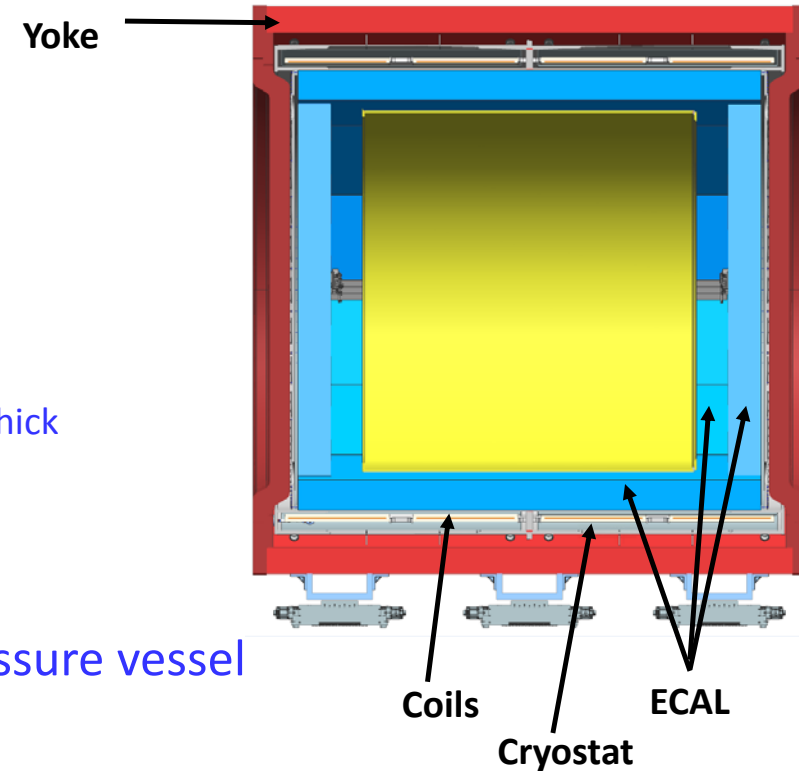
Reco hits matching a Sim Hit – based on CellID

My sample's AnaTrees
~48% in 0 bin

Fraction in 0 bin in
endcap tile/strip (~44%/53%)!!

Latest Geometry

- New ECAL geometry –
 - 42 layers of Pb-Scintillator sandwich ($\sim 10.5 X_0$)
 - Scintillator: 8 layers, each 0.5 cm thick, and 34 layers, each 1 cm thick
 - Pb – 8 layers, each 0.7 mm thick, 34 layers, each 1.4 mm thick
 - I believe endcaps are 6 + 36 layers
 - Barrel has 12 fold symmetry
 - Newly optimized SPY magnet and cryostat as the pressure vessel
 - No extra material between the ECAL and TPC



TPCRadius = 273 cm, TPCLength = 259

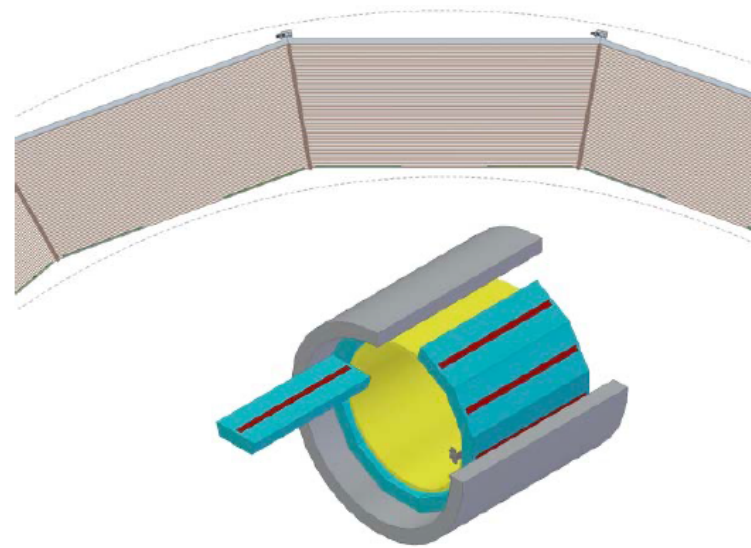
ECALInnerRadius = 278, ECALOuterRadius = 334 cm

TPCFidRadius = 222.5, cm TPCFidLength = 215

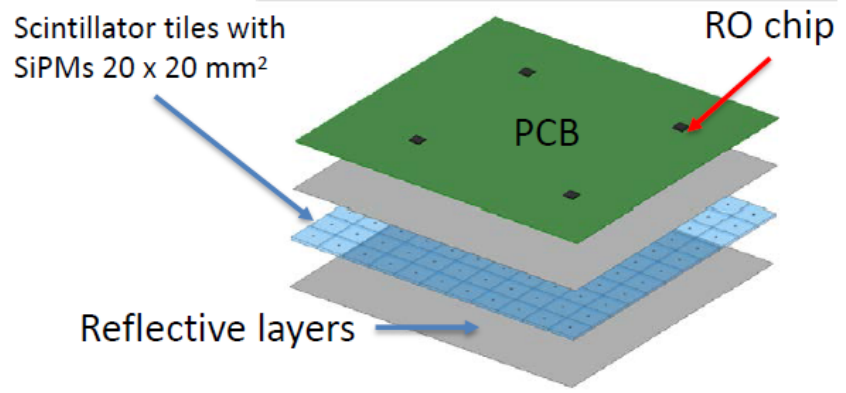
ECALStartX = 328 , ECALEndX = 375 cm

ECAL: Engineering status


- Based on the ILD design
- Preliminary designs
 - Check ECAL space between TPC/Magnet and put in realistic tolerances.
 - First ideas on how to fix the ECAL
 - Self-supporting
 - Individual rails
 - Module/Layer design: Lead too soft/toxic - most likely in a "super-module" made in carbon fiber



From a talk by either Alan or Eldwan

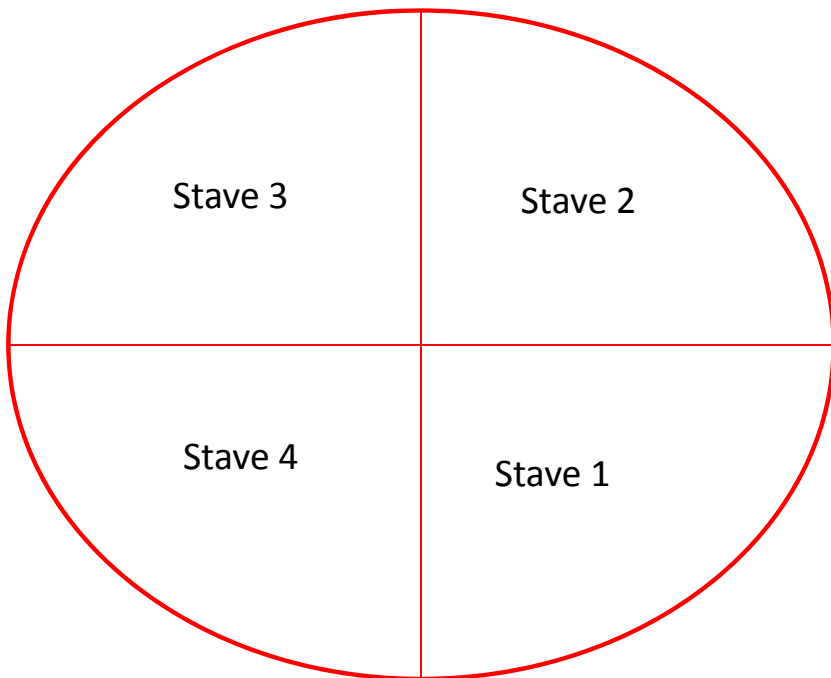


Nomenclature

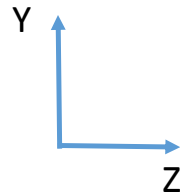
- ECAL is made up of **detector id, staves, modules, slices and layers** – these are used to encode the CellID (which is being used during digitization)
 - **det_id = 1 (Barrel ECAL), = 2 (Endcap ECAL)** [Yoke barrel appears to be 4]
 - **Barrel ECAL: has 12 staves (as you go around in phi)**
 - Module = 1 for Tile, = 2 for Strips
 - Slice = 1 for absorber, = 2 for Scintillator
 - Layers go from 1 to 42 (first 8 are tiles, remainder are strips)
 - **Endcap ECAL: has 4 staves, appear to match the four quadrants** 
 - Module: = 0 for negative X, = 3 for positive X. **So no distinction between strips/tiles?**
 - Slice = 1 for absorber, = 2 for Scintillator
 - Layers go from 1 to 42 (first 6 are tiles, remainder are strips)
- Nodenames (in the code) look like
 - **BarrelECal_stave10_module02_layer_21_slice2_vol_0**
 - **EndcapECal_stave02_module03_layer_13_slice2_vol_0**

Staves in the Endcap – results based on hand scanning output in log file

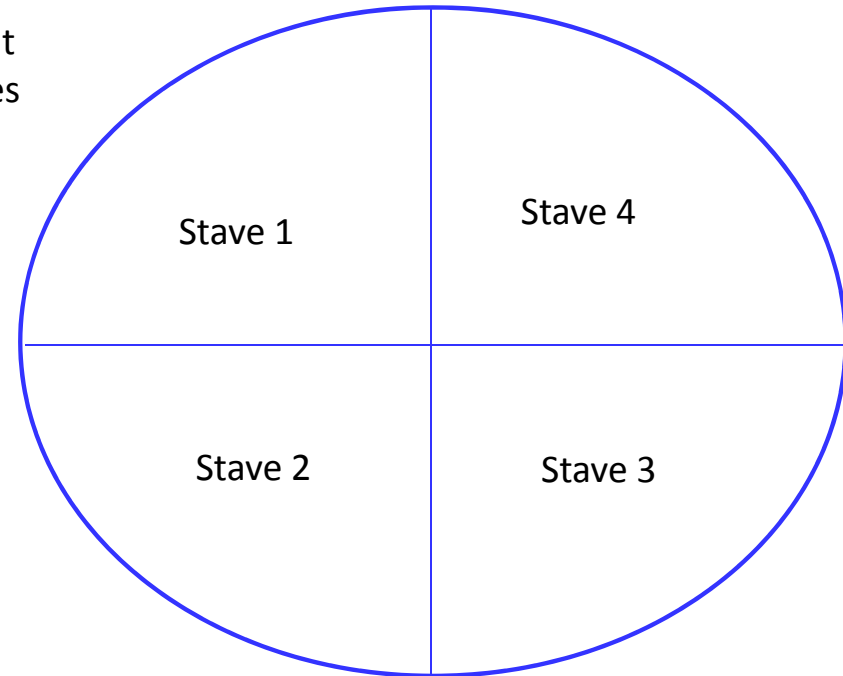
Module 0, i.e $X < 0$



Boundaries are not along the Y & Z axes
There is small amount of slop across the axes



Module 3, i.e $X > 0$

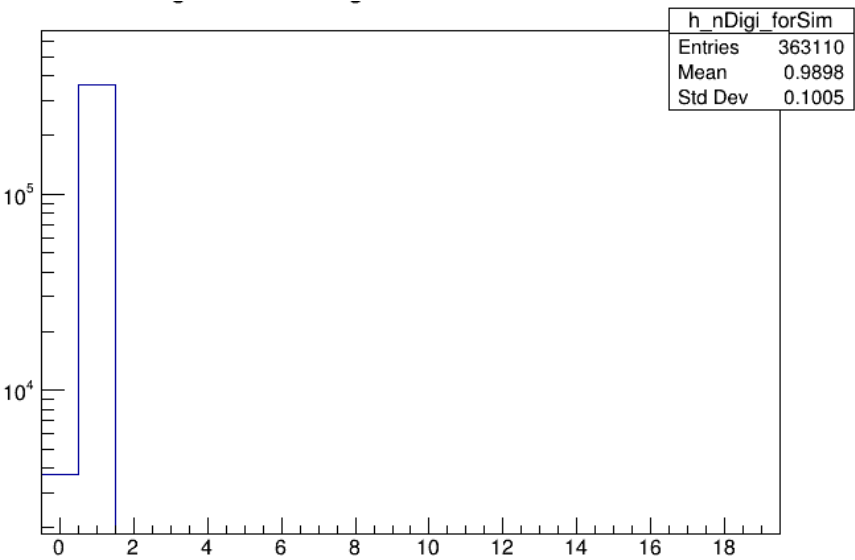


My sample*

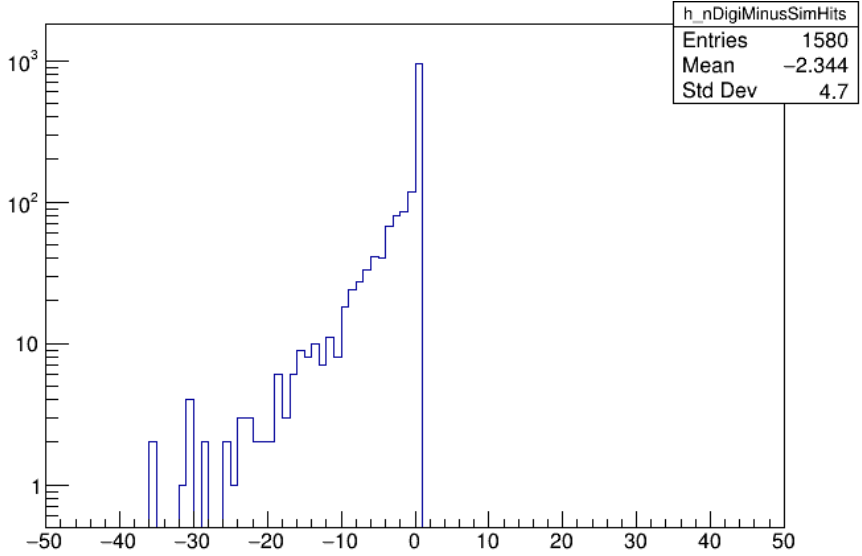
*My files are in /dune/data/users/vj/test4/work/

- Built GArSoft locally using develop branch (on July 6)
- `art -n 1000 -c prodgenie.fcl` (I don't know what .fcl/flux files Eldwan used)
 - *"...New detector geometry loaded from: /nashome/v/vj/datadir/test4/localProducts_garsoft_develop_e20_prof/garsoft/v02_10_03/gdml/nd_hall_mpd_only/nd_hall_mpd_only_ECal12sides_42l_SPY_v3_wMuID.gdml ..."*
 - *In GENIEGen.fcl, I see the following:*
 - *# Optimized flux, Nov 2017:*
 - *FluxType: "dk2nu"*
 - *FluxFiles: ["g4lbne_v3r5p4_QGSP_BERT_OptimizedEngineeredNov2017_neutrino_?????.dk2nu.root"]*
 - *FluxCopyMethod: "IFDH"*
 - *FluxSearchPaths: "/pnfs/dune/persistent/users/ljf26/fluxfiles/g4lbne/v3r5p4/QGSP_BERT/OptimizedEngineeredNov2017/neutrino/flux/dk2nu"*
- `art -c readoutsimjob.fcl -s genie_gen.root` (debug statements in Digi code)
- `art -c recojob.fcl -s readoutsim.root`
- `art -c anajob.fcl -s reco.root`
 - Changed the default anatree.fcl to store BackTracker info for CALO, as well as storing CALO DigiHits
- **I made a total of 3000 events – because the flux files are in the /neutrino/ sub-dir, I am guessing they are FHC**

DigiHit plots (events in TPC fiducial volume – with AnaTrees (3K events – my sample))



DigiHits matching a SimHit – based on CellID



DigiHits - # SimHits (1 entry/event)

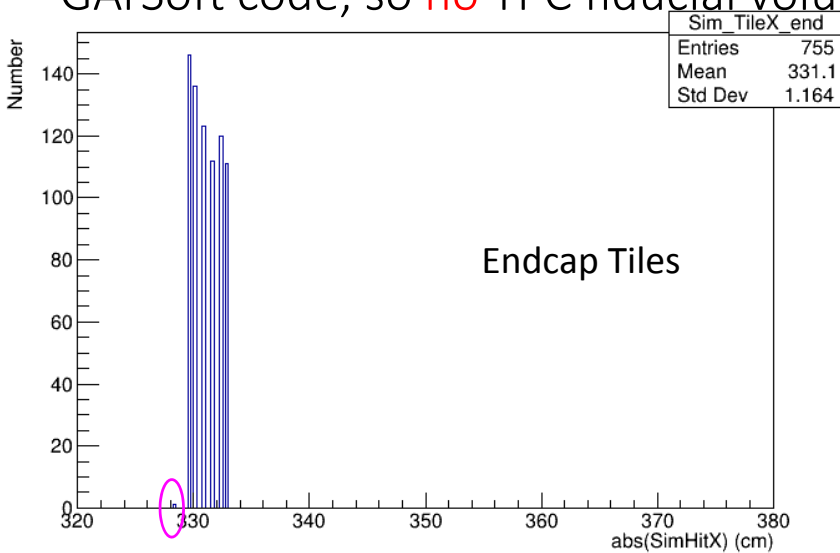
So, either 1 match or 0 matches –
 ~1% of the SimHits are not digitized
 (Undigitized fractions in Barrel/Endcap are wildly different - later)

In ~60% of events, all SimHits are Digitized (peak at 0)
 In a few events as many as 30 SimHits are not digitized

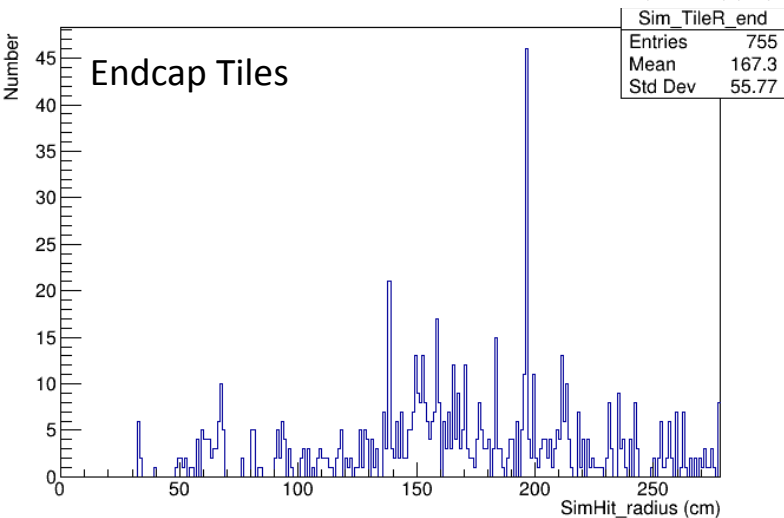
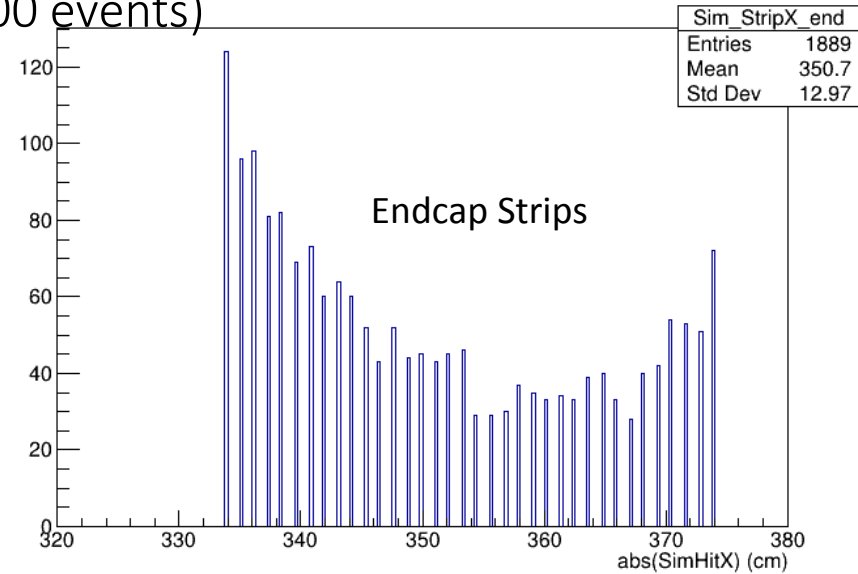
In the code, the position of the SimHit is shifted to, we think,
 to where the SiPM (very obvious to spot in the barrel)
 The nodenames of the original and the new positions are compared.
 If they are the same, hit is digitized, otherwise, it is a failed digitization.

(If you do a cumulative sum of all but the 0 bin, the total agrees with what you see in the 1st bin in the plot on the left)

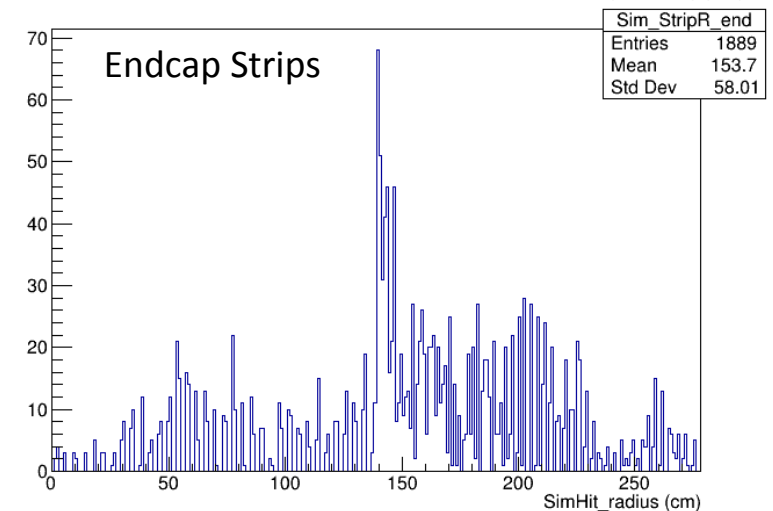
When digitization is successful – here the plots are made during digitization in GARSoft code, so **no** TPC fiducial volume cut is made (100 events)



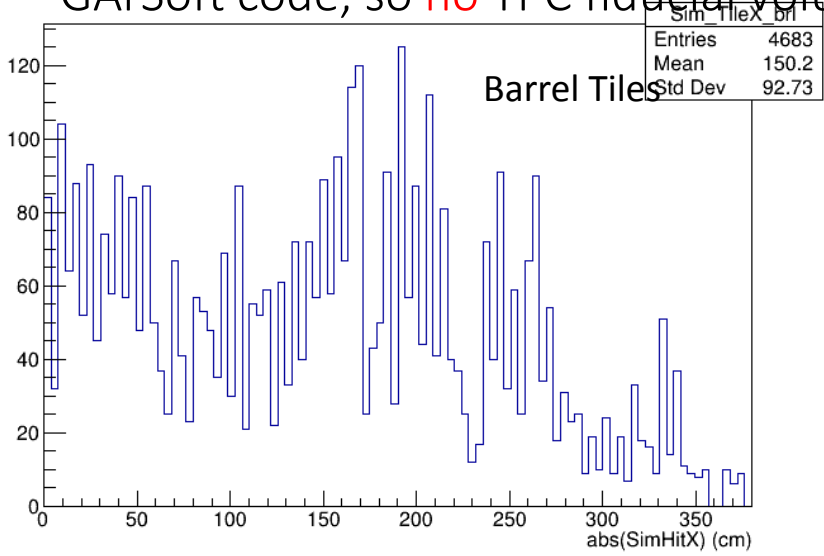
I am plotting
abs(X)



Here I am using
the location of hits
after they are
shifted to where the
SiPM is

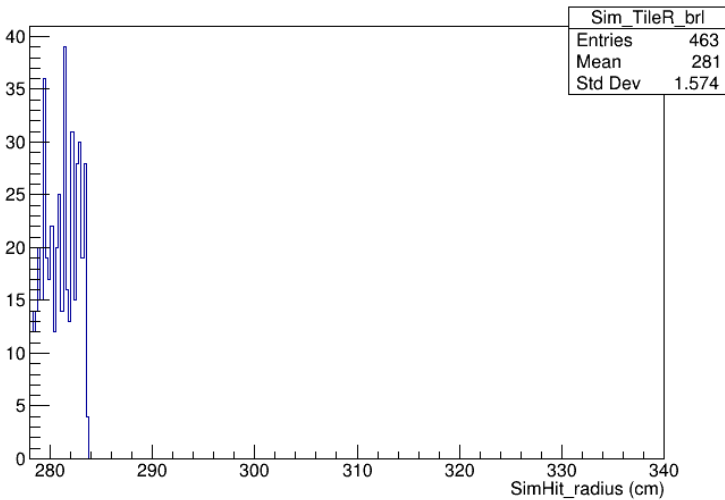
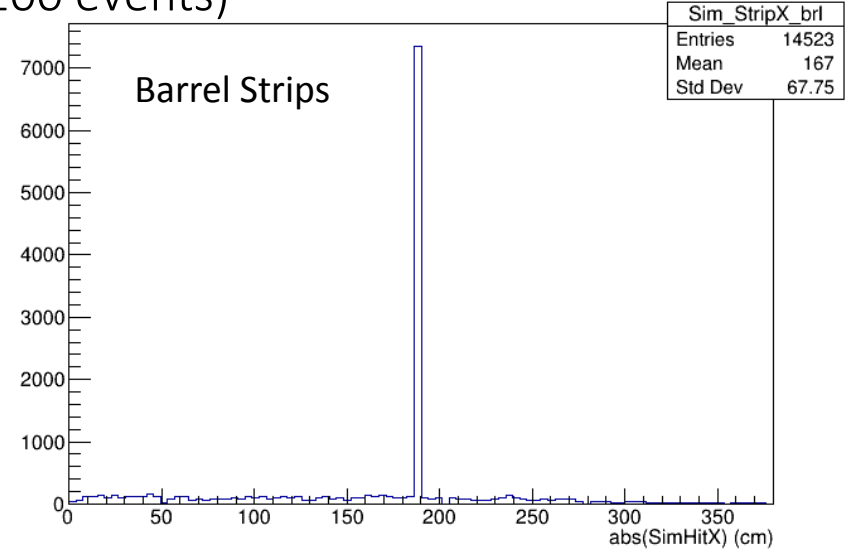


When digitization is successful – here the plots are made during digitization in GARSoft code, so **no** TPC fiducial volume cut is made (100 events)

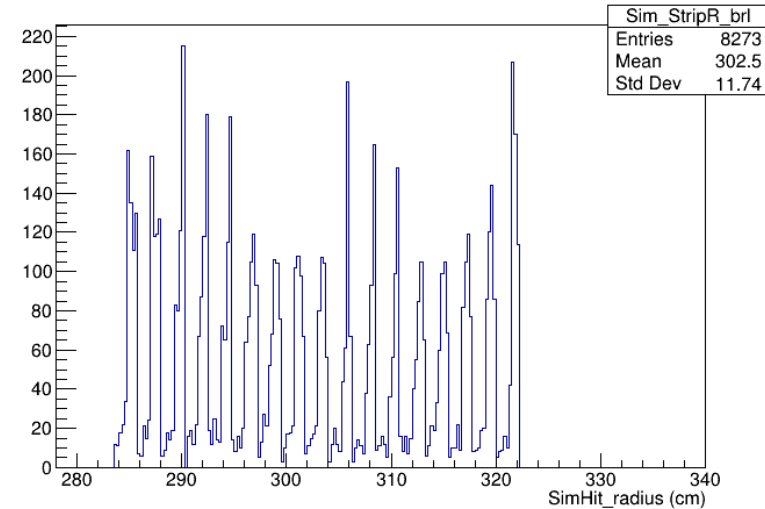


I am plotting
abs(X)

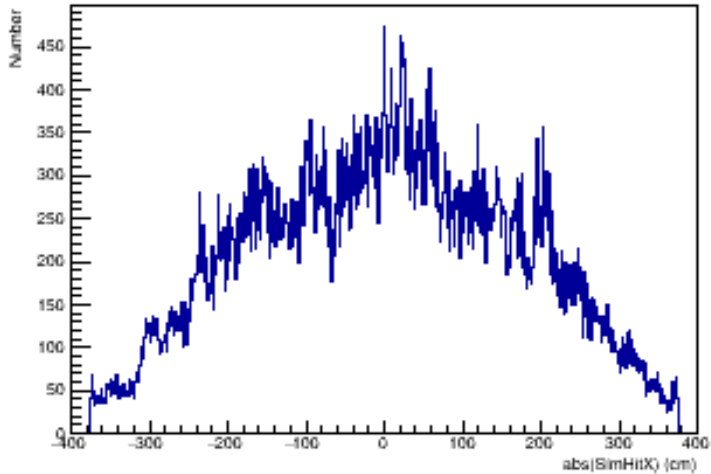
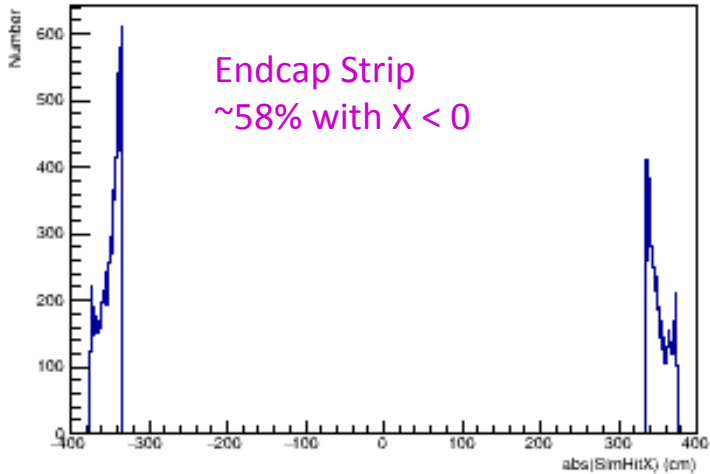
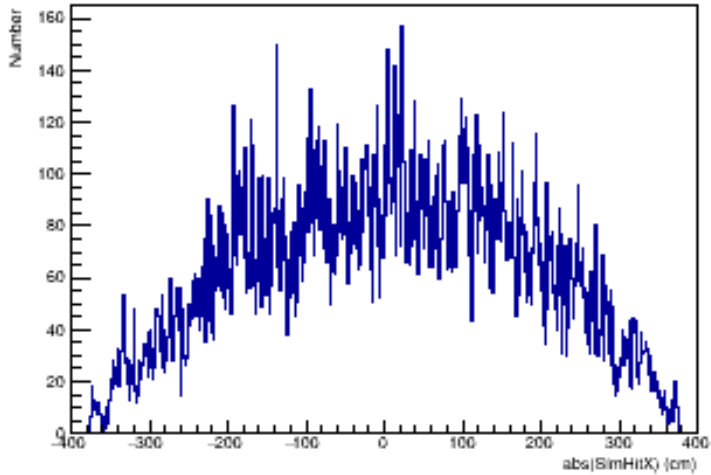
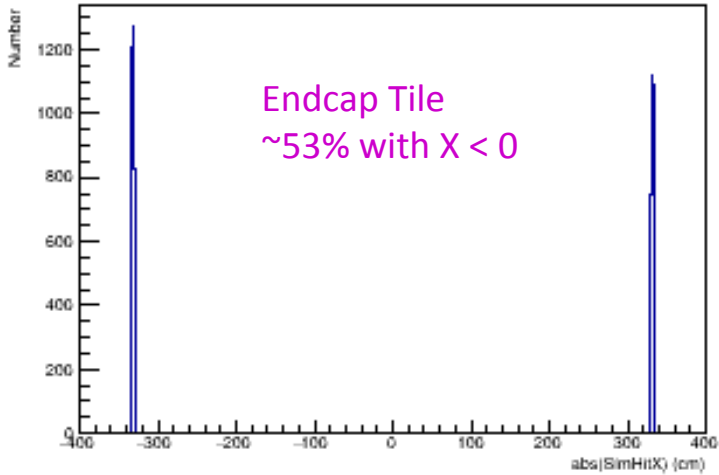
Here I am using
the location of hits
after they are
shifted to where the
SiPM is



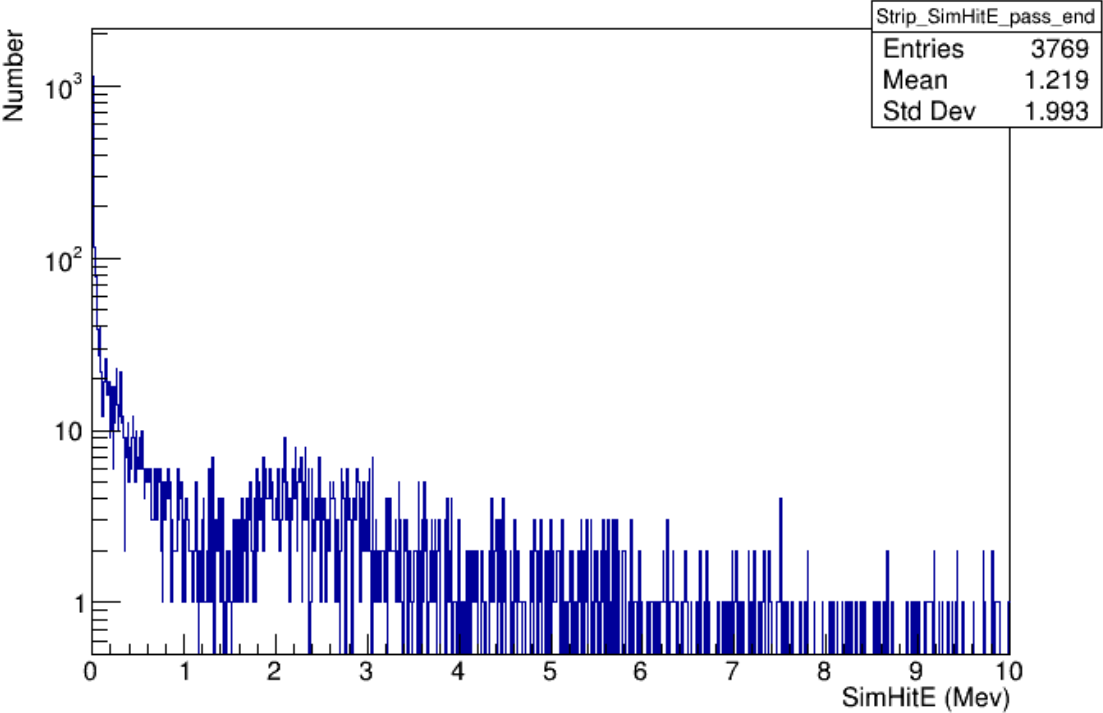
Since the barrel is 12-sided,
to get separation between
Tiles and Strips in
the radial direction, I only
looked at cases where the
phi of the hit was within a small
range of a multiple of 30°



When digitization is successful, X-distribution is reasonably symmetric



SimHit energy when digitization is successful (200 events – no FV cut – in garsoft code) – **only look at Endcap Strips** (to compare with failed Digi cases)



30% of hits have $E < 0.01$ MeV

18% have E between 0.01 and 0.1 MeV

52% have $E > 0.1$ MeV

Where are the missing DigiHits? –based on 1000 events (no FV cut)

Number of SimHits in file = 230216 (printout in garsoft code)

| | Barrel Tile | Barrel Strip | Endcap Tile | Endcap Strip |
|---------------------------|------------------|--------------|-------------|--------------|
| # SimHits | 46605 | 152867 | 8380 | 22364 |
| # Successful digitization | 46605 | 152864 | 8363 | 20300 |
| # Failed digitization | 0 | 3 | 17 | <u>2064</u> |
| % failed | <0.005% (@90%CL) | ~0.002% | ~0.2% | <u>9.2%</u> |

Problem is in CalculatePosition() method in ReadoutSimulation/ECALReadoutSimStandardAlg.cxx (in backup)

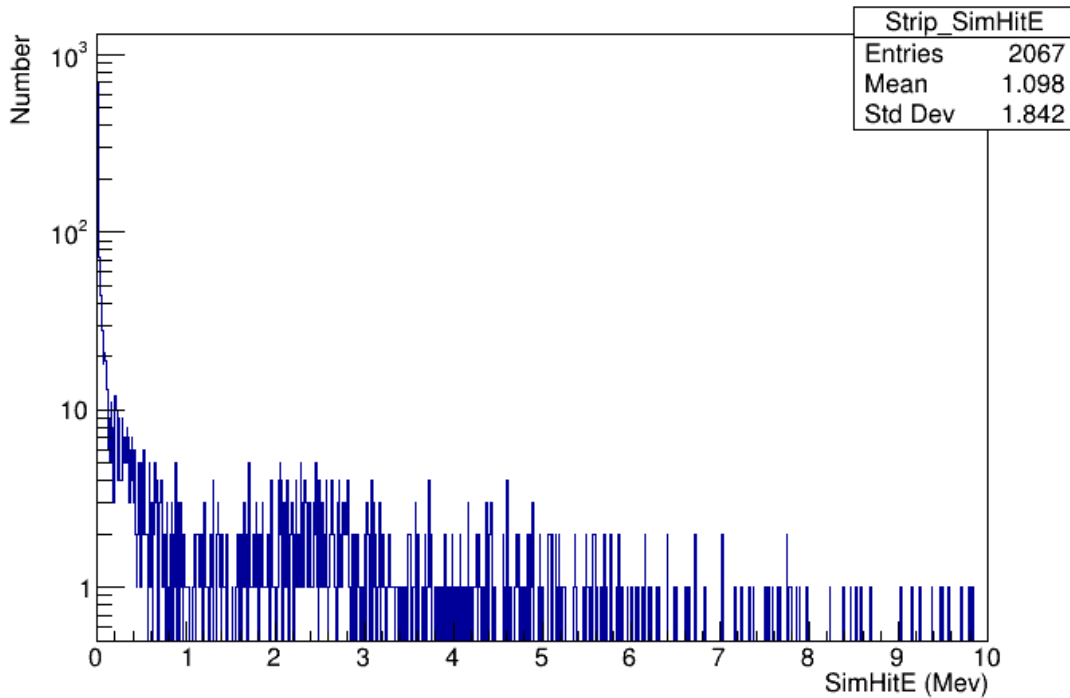
As per the code, the process is as follows:

- 1) The “World” location of the SimHit is used to get the node/nodename of the volume it is in
- 2) The “World” location is converted to a “local” position – this call appears to fill a transformation matrix
- 3) A “new local” position is determined using CellID and the original node of volume
- 4) This “new Local” position is transformed back into a “new World” position, using the matrix from (2)
- 5) A new node/nodename is obtained from the “new World position”
- 6) Nodenames of the old and new World locations are compared. If they are different, the hit is not digitized

Some examples of changed node names

- ND_Gar X, Y, Z 334.792, -124.208, 117.419
 - Before local->world->local: EndcapECal_stave03_module03_layer_08_slice2_vol_0
 - After local->world->local: EndcapECal_stave02_module03_layer_08_slice2_vol_0
- ND-GAr X,Y,Z 343.708 , 46.6549 , -39.4362
 - Before: EndcapECal_stave01_module03_layer_16_slice2_vol_0
 - After: EndcapECal_stave04_module03_layer_16_slice2_vol_0
- ND-GAr X,Y,Z -340.336 , 110.43 , 254.805
 - Before: EndcapECal_stave02_module00_layer_13_slice2_vol_0
 - After: volIMPD_0 ????
- ND-GAr X,Y,Z -358.072 , 274.529 , -15.1259 (radius = 275 cm, so why the confusion?)
 - Before: EndcapECal_stave02_module00_layer_28_slice2_vol_0
 - After: BarrelECal_stave09_module01_layer_07_slice1_vol_0 (moved to the absorber??)
- ...

When digitization fails – plots made in digitization code, so **no** TPC fiducial volume cut (1000 events) – only look at Strips (essentially all in Endcaps)

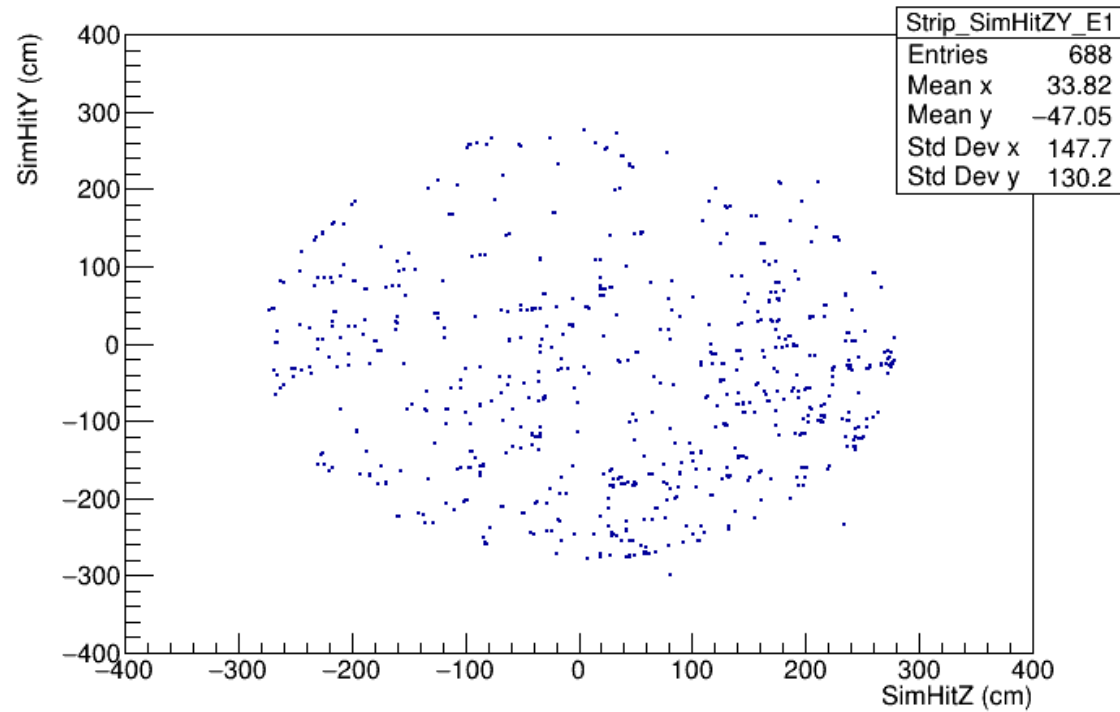
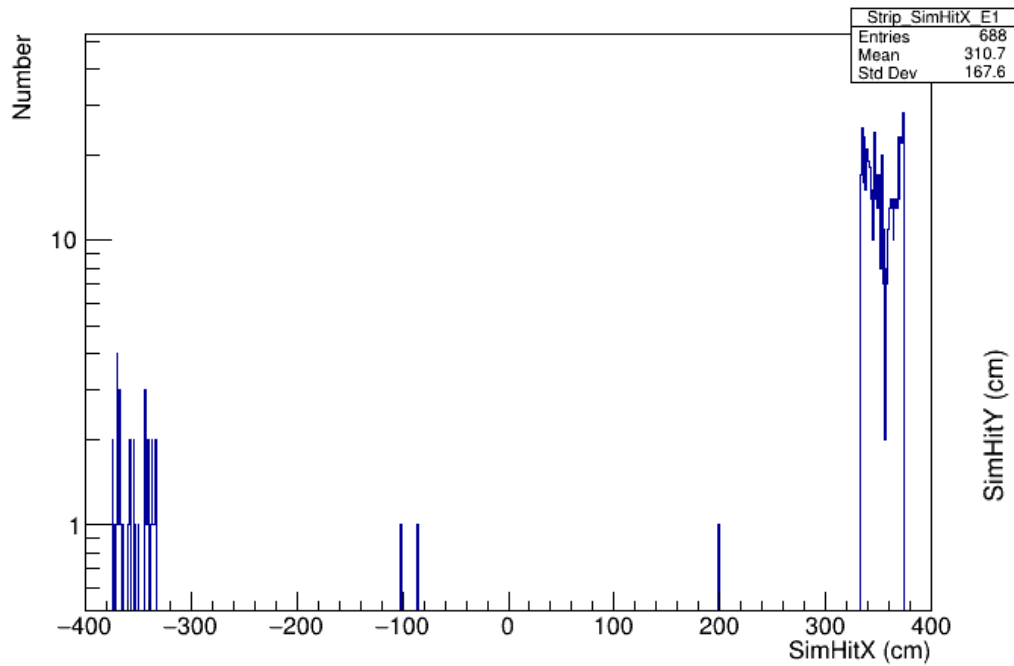


33% of hits have $E < 0.01$ MeV

19% have E between 0.01 and 0.1 MeV

48% have $E > 0.1$ MeV

SimHit $E < 0.01$ MeV

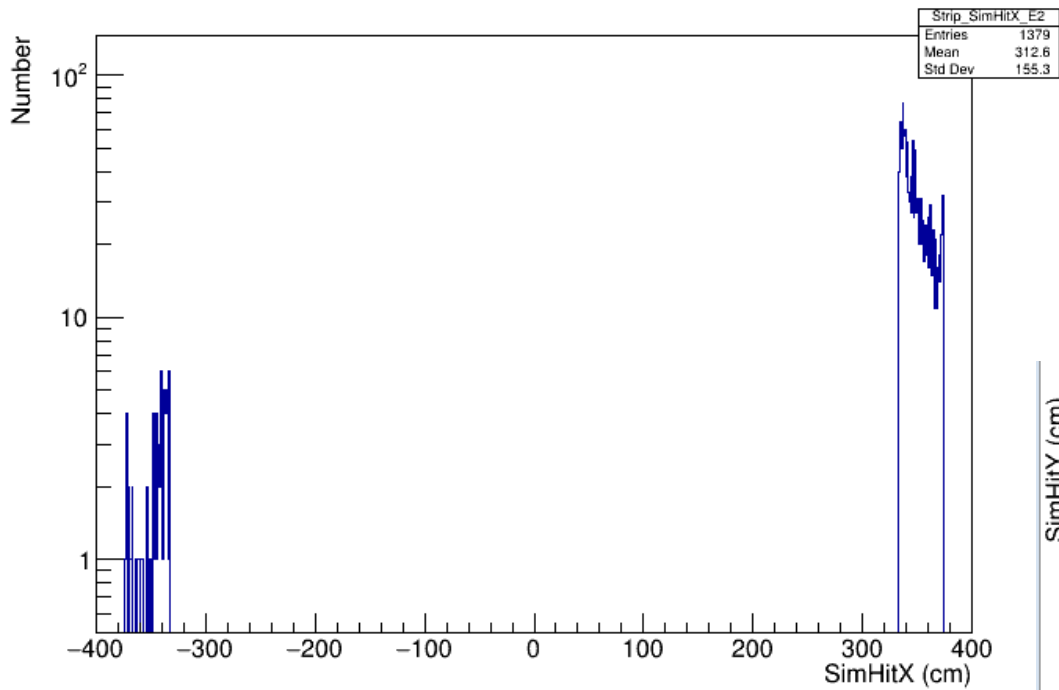


X-distribution is quite asymmetrical:

~6% entries with $X < -325$, and ~94% with $X > 325$ cm

3 entries are in barrel

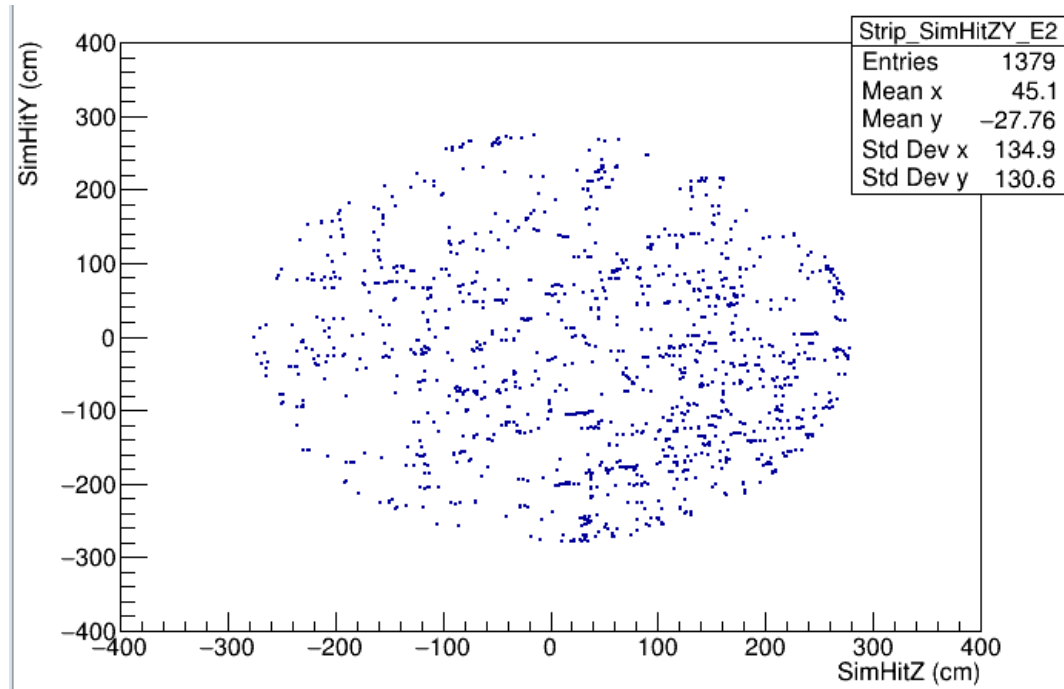
SimHit E > 0.01 MeV



X-distribution is quite asymmetrical:

~5% entries with $X < -325$, and ~95% with $X > 325$ cm

0 entries are in barrel

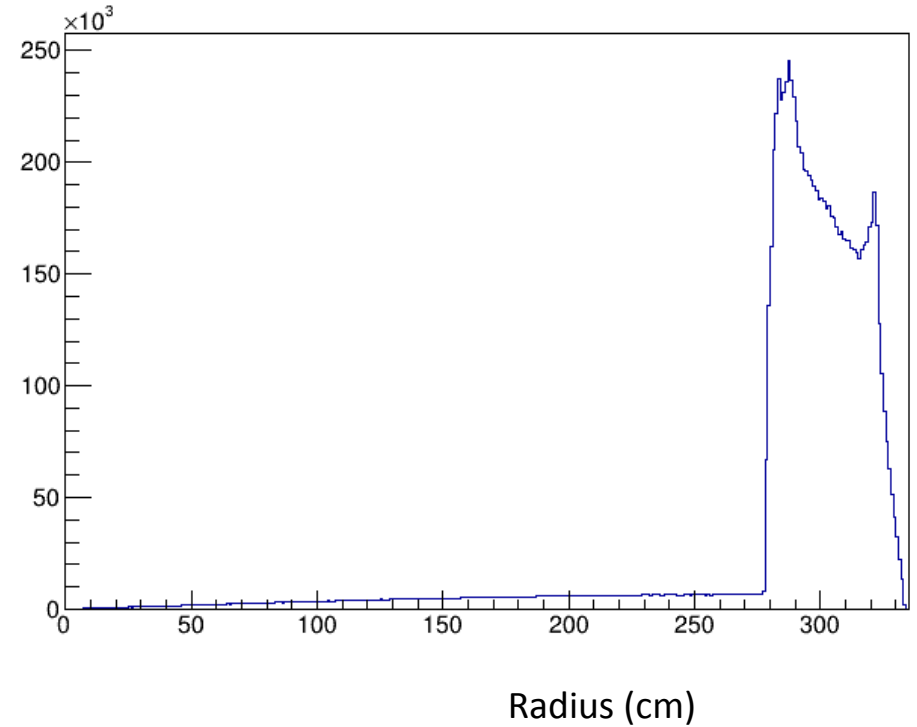
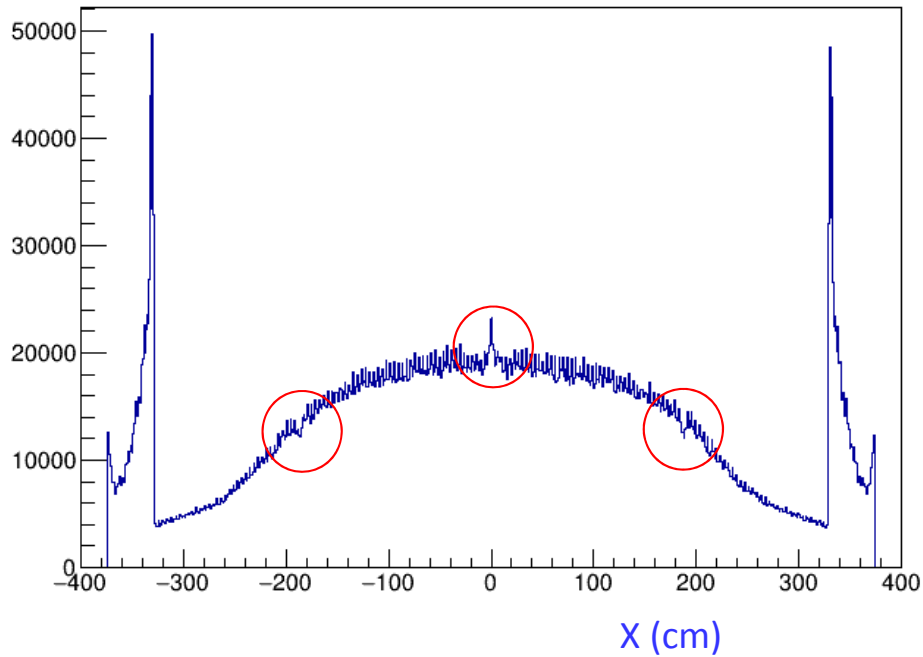


Summary

- Many issues to understand

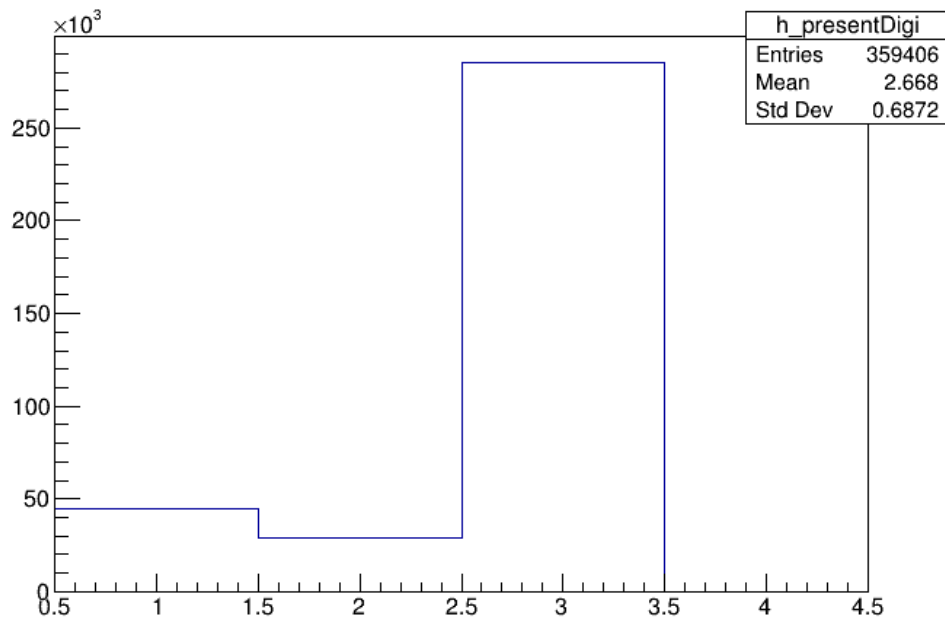
Extra –

Look at SimHits (event vertex in TPC fiducial volume) – with AnaTrees (499K events – Eldwan’s sample)

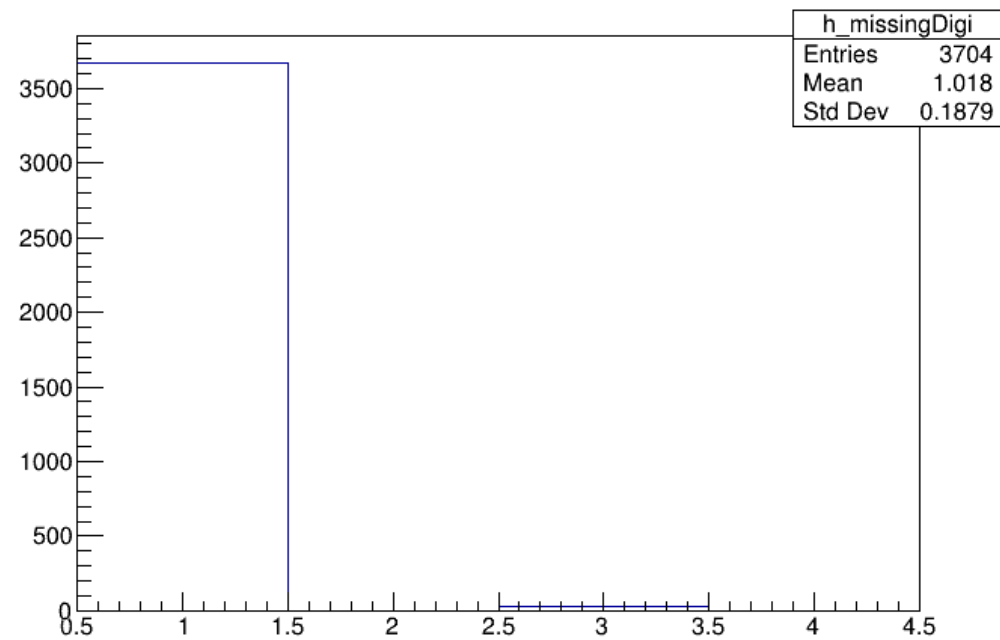


I looked at Y vs. Z and radius of hits at X: -189 to -185, +- 3 and 185 to 189, and there is nothing odd about those locations

Where are the DigiHits? (events in TPC fiducial volume) – AnaTrees (3K events – my sample)



Location of SimHit when Digitization is successful



Location of SimHit when the Digi hit fails

1: Endcap, 2: Upstream barrel, 3: Downstream barrel

```
std::pair< std::array<double, 3>, bool > ECALReadoutSimStandardAlg::CalculatePosition(float x, float y, float z, raw::CellID_t cID) const
```

```
std::array<double, 3> point = {x, y, z};  
TGeoNode *node = fGeo->FindNode(point);//Node in cm...  
std::string nodename = node->GetName();
```

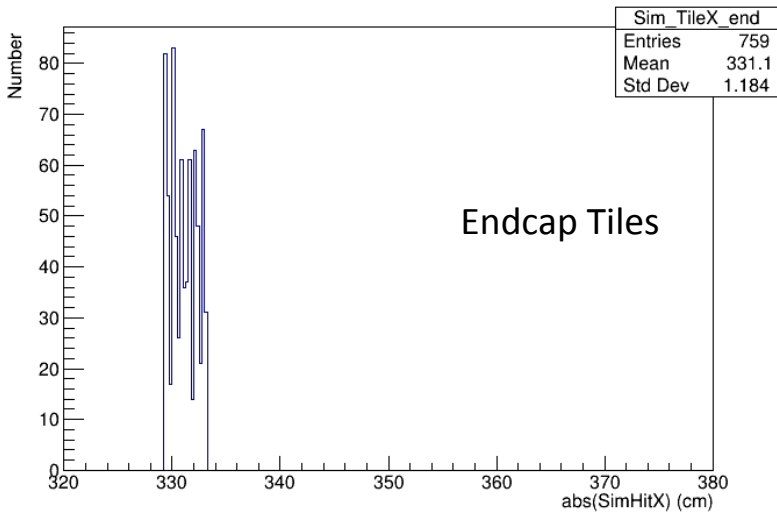
```
std::array<double, 3> pointLocal;  
gar::geo::LocalTransformation<TGeoHMatrix> trans;  
fGeo->WorldToLocal(point, pointLocal, trans);
```

```
std::array<double, 3> pointLocal_back = fGeo->GetPosition(node, cID);//returns in cm  
std::array<double, 3> point_back;  
fGeo->LocalToWorld(pointLocal_back, point_back, trans);  
TGeoNode *new_node = fGeo->FindNode(point_back);//Node in cm...  
std::string newnodename = new_node->GetName();
```

```
//get the base of the node names (slice can be added in the new node)  
std::string base_node_name = nodename.substr(0, nodename.find("_layer_") + 9);  
std::string base_new_node_name = newnodename.substr(0, newnodename.find("_layer_") + 9);
```

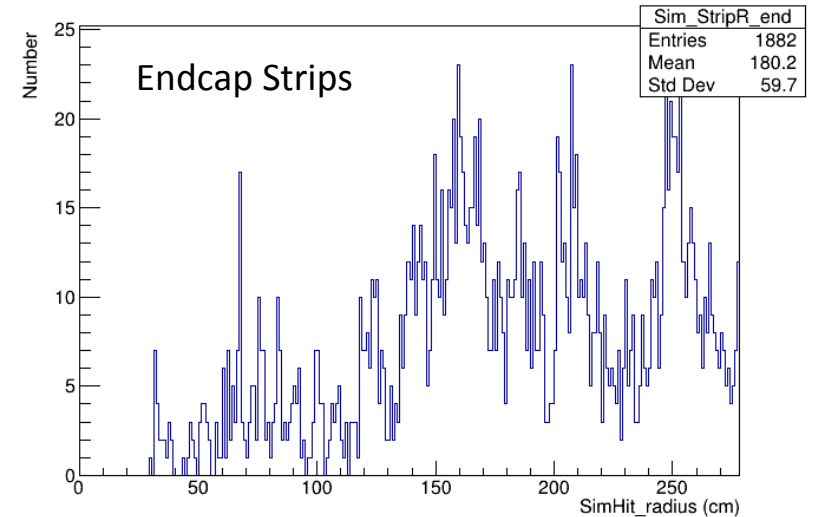
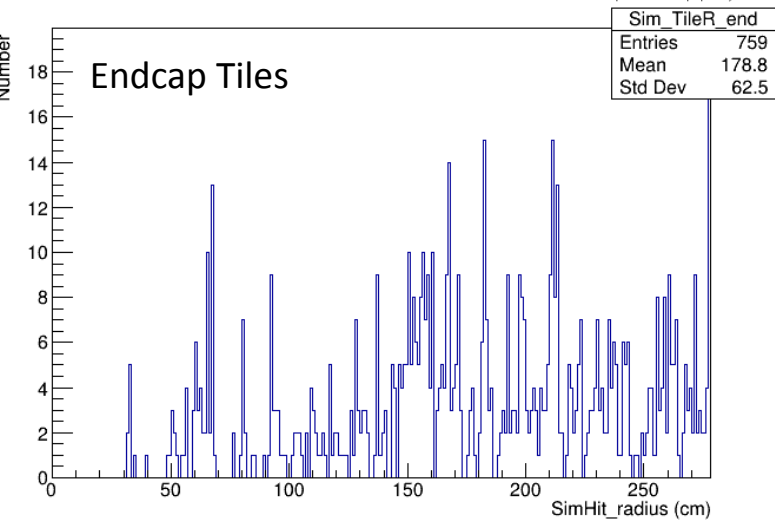
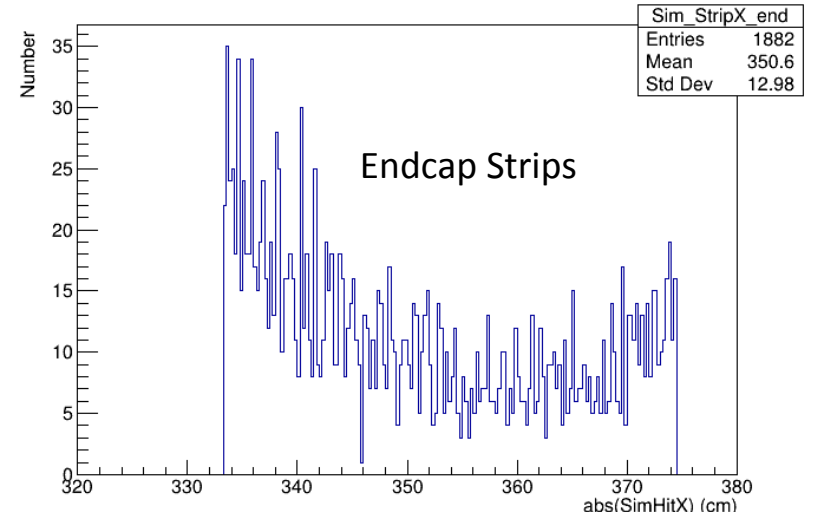
```
if( base_new_node_name != base_node_name ){  
    drop = true;  
}  
return std::make_pair(point_back, drop);
```

When digitization is successful – here the plots are made during digitization in GARSoft code, so **no** TPC fiducial volume cut is made (running over 100 events)

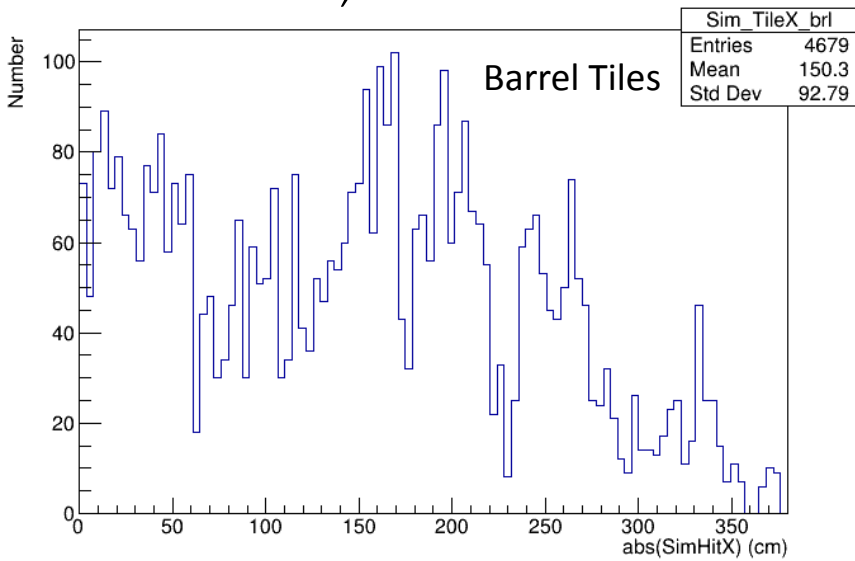


I am plotting
abs(X)

Here I am using
the location of hits
before they are
shifted to where the
SiPM is

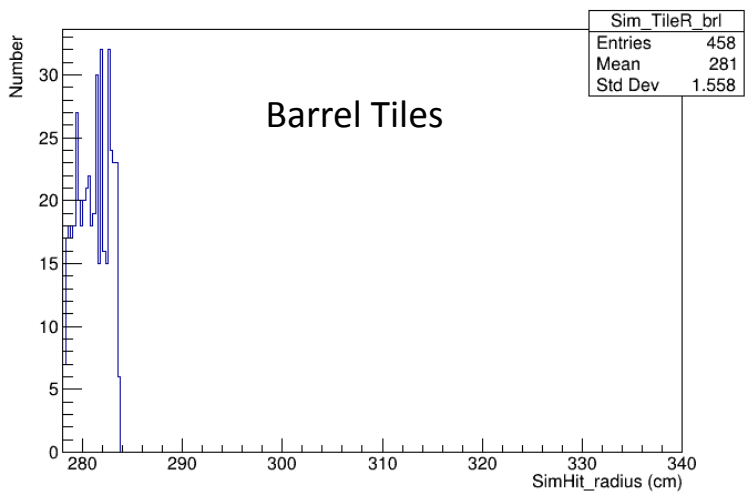
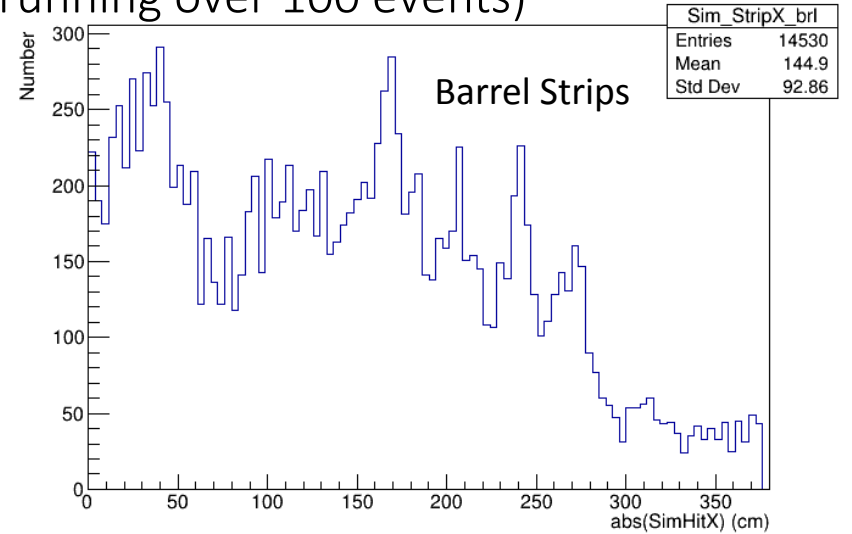


When digitization is successful – here the plots are made during digitization in GARSoft code, so **no** TPC fiducial volume cut is made (running over 100 events)

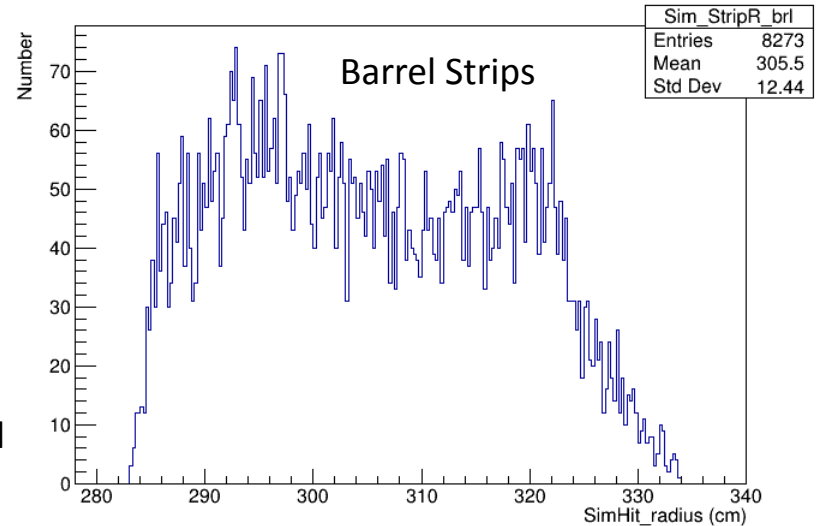


I am plotting $abs(X)$

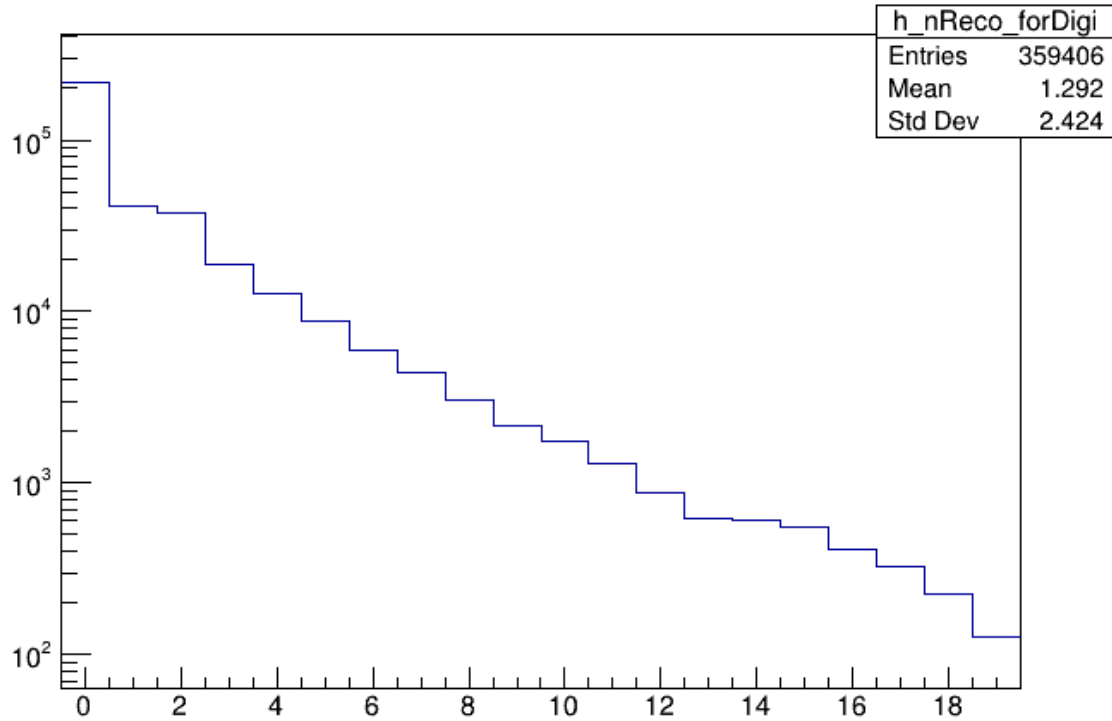
Here I am using the location of hits before they are shifted to where the SiPM is



Since the barrel is 12-sided, to get a separation between Tiles and Strips in the radial direction, I only looked at cases where the ϕ of the hit was within a small range of a multiple of 30°



DigiHit plots



RecoHits matching a DigiHit – based on CellID
~1% of the SimHits are not digitized

So, the problem of multiple matches is at the Reco stage – Digi stage looks OK
Reco is with the Strip splitting algorithm

extras

Structure of AnaTrees

- Although you have arrays like CCNC, Interaction Type, etc., whose length is the number of interactions within a spill, the reconstructed information in CALO, TPC is aggregated over all of them.
 - Current files have 1 interaction/spill, so it's a moot point
- In these studies, I use the variables **SimHitCellID** and **RecoHitCellID** to compare **Sim** and **Reco** information
 - I still need to trace CellEncoding in the code, but hits with the same cellID need not be at the same location in space (later)

/pnfs/dune/persistent/users/ebrianne/ProductionSamples/ND-GAr/ nd_hall_mpd_only_ECal12sides_42I_SPY_v3_wMuID/

- Contain Reco files, Anatree files, CAFMaker files and ParamSim files