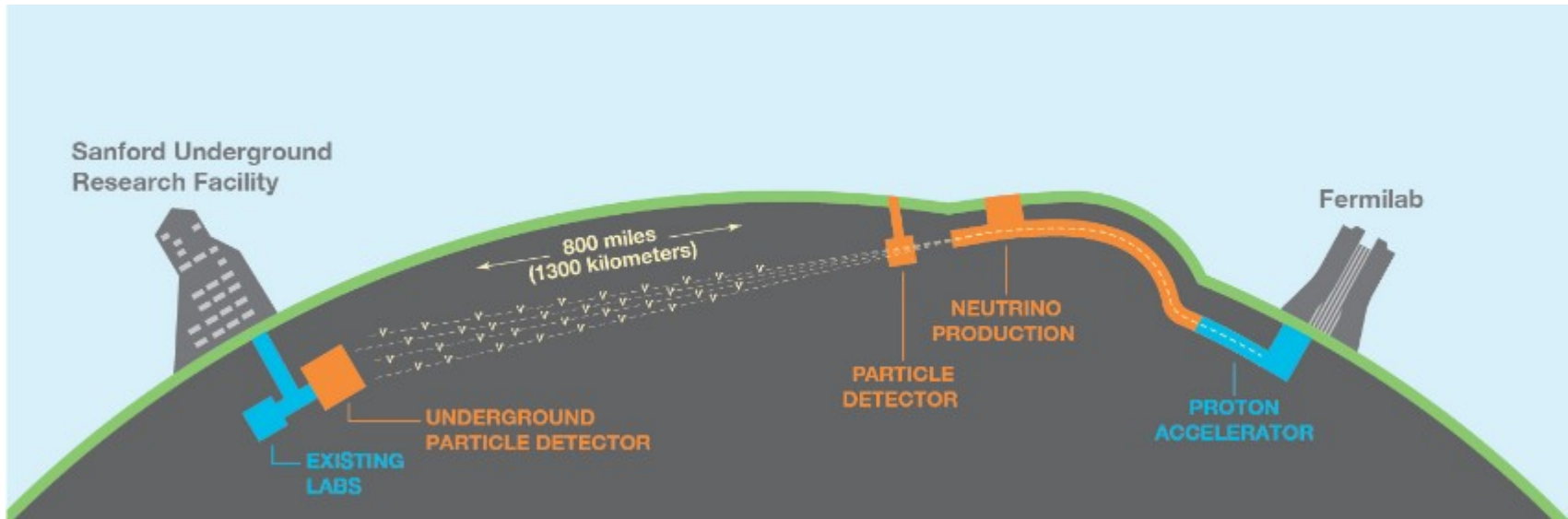




# Event reconstruction in ProtoDUNE

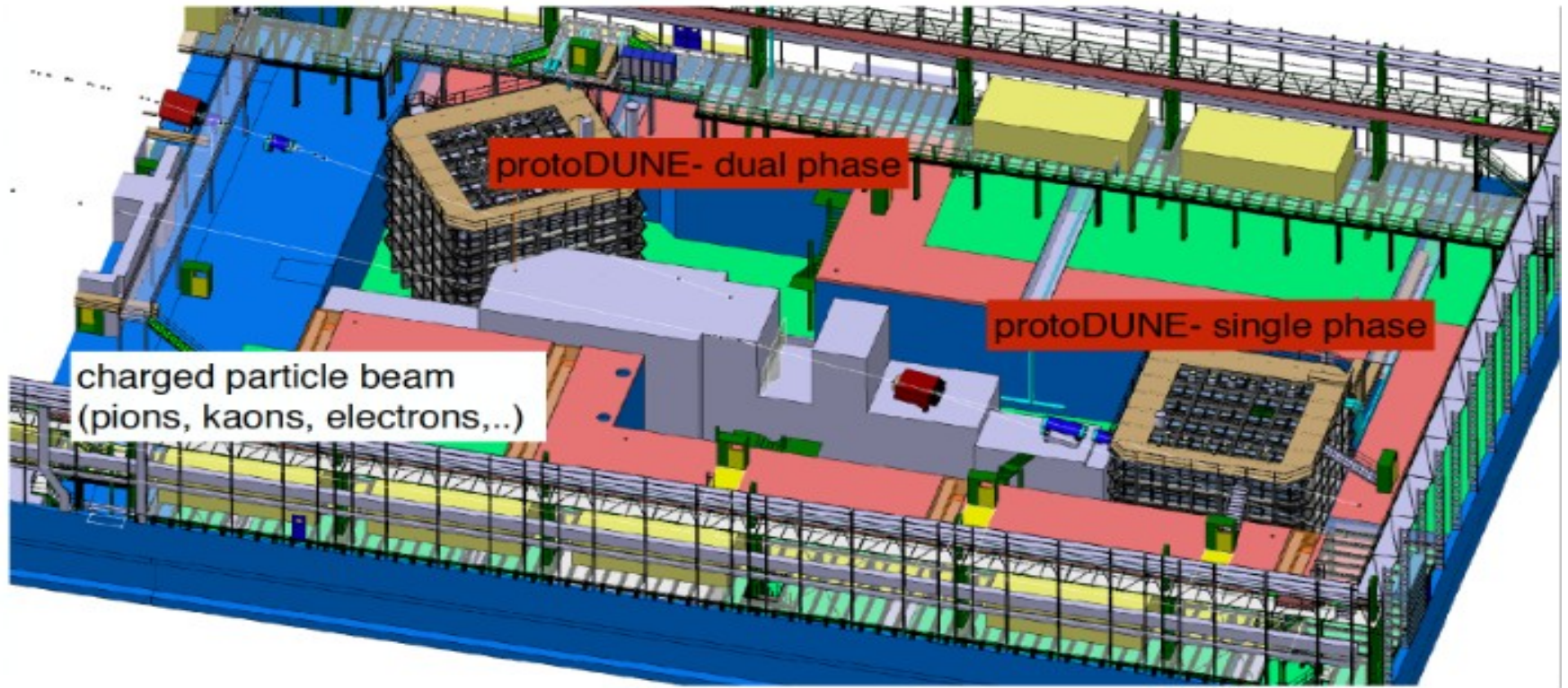
Roberto Schiattarella  
Supervisor Ken Herner  
Final report  
24 September 2019

# DUNE



- **New neutrino beam facility at Fermilab**
- **A highly capable Near Detector at Fermilab to measure the unoscillated neutrino spectrum and flux constraints**
- **4 x 10 kton (fiducial) LArTPC modules (single and/or dual-phase) deep underground at SURF (Lead, SD, 1300 km baseline) to measure oscillations, SN burst neutrinos, nucleon decay, atmospheric neutrinos...first data in 2024**

# ProtoDUNE detectors at CERN



- Each protoDUNE contains 800t of LAr—biggest ever to date!
- Validate LArTPC technologies
- Demonstrate long-term performance and stability
- Charged particle test-beams to characterise detector response with particle energies in the region of interest for DUNE ( of the order 500 MeV to 7 GeV)

# The computing problem

- **ProtoDUNE can write up to 2-3 GB/sec of data when running at 25 Hz beam.**
- **3.2 PB of raw data from ProtoDUNE SP + more from the future ProtoDUNE DP measure**
- **Big DUNE will write 10-30 PB/year from the far detectors and probably more than that from the near detector.**

# My work in the production group

- **Code development for autorelease of held jobs on the computers grid.**
- **Reprocessing protoDUNE Single Phase data**
- **Montecarlo simulation campaign for Supernova samples of elastic scattering events**
- **LarSoft module to validate the reprocessing protoDune Single Phase data**

# Autorelease for held jobs

# What is POMS?



- ✓ **A service to assist production and analysis of experiments in their MC production and DATA processing.**

*As the quantity of data originated by the running experiments greatly increases, the ability of simplifying the steps in data processing and management has become more and more appealing to the users*

- ✓ **What does POMS do ?**

*Submits grid jobs;*

*Tracks job submissions through Landscape;*

*Organizes job submissions into Campaigns with Stages;*

*Has a Graphical Editor for Campaigns and their Stages.*

*Assists in analyzing job failures with plots, charts, and easy access to log files.*

# POMS Landscape



## Campaign Stages ?

Roberto Schiattarella

dune production

Campaign Stage  
protoDUNE\_SP\_cosmics\_3ms\_autorelease\_test\_v07\_08\_00\_05

3961

in Campaign protodune-  
sp\_keepup\_cosmics\_autorelease\_test\_v07\_08\_00\_05

3576



*POMS is well integrated with Landscape/Grafana with many plots, reports and status pages available*

**Job Status: 1) Idle 2) Running 3) Removed 4) Completed 5) Held**



# Autorelease for Memory and Run Time

- Job was held for Memory
  - Manual Release
    - Waste of time
      - Autorelease with memory
- Job was held for run time
  - Manual Release
    - Waste of time
      - Autorelease with Run Time

Next step: Try to combine the two autorelease codes into a single code

# Code lines for autorelease with Run Time

*lines\_1 = +OriginalRunTime=300*

*lines\_2 = +GraceRunTime=10000*

*lines\_3 = +IncreaseJobLifeTime=(NumJobStarts>0)&&(!  
isUndefined(LastHoldReasonCode))&&(LastHoldReasonCode=?  
=26&&LastHoldReasonSubCode=?=8)*

*lines\_4 = +ShouldRelease=(JobStatus=?=5)&&(HoldReasonCode=?  
=26&&HoldReasonSubCode=?=8)&&((EnteredCurrentStatus-  
JobStartDate)<(OriginalRunTime+GraceRunTime))*

*lines\_5 =  
+JOB\_EXPECTED\_MAX\_LIFETIME=ifthenelse(IncreaseJobLifeTime,OriginalRunTime+Grac  
eRunTime,OriginalRunTime)*

*lines\_6 = periodic\_release=ShouldRelease*

# Code lines for autorelease with Run Time

```
...
028 (22869915.000.000) 08/19 21:36:14 Job ad information event triggered.
JOB_Site = "CCIN2P3"
JOB_GLIDEIN_Name = "gfactory_instance"
Size = 2306408
Proc = 0
JOB_GLIDEIN_Entry_Name = "CDF_FR_CCIN2P3_cccreamceli10_long"
EventTime = "2019-08-19T21:36:14"
TriggerEventTypeName = "ULOG_IMAGE_SIZE"
JOB_GLIDEIN_SiteWMS_Queue = "long"
MemoryUsage = 151
TriggerEventTypeName = 6
JOB_GLIDEIN_Site = "CCIN2P3"
JOB_GLIDEIN_SiteWMS_JobId = "14706532"
MyType = "JobImageSizeEvent"
JOB_GLIDEIN_ProcId = "0"
JOB_GLIDEIN_Schedd = "schedd_glideins2@gfactory-2.opensciencegrid.org"
JOB_GLIDEIN_ClusterId = "617126"
Cluster = 22869915
JOB_GLIDEIN_Factory = "OSG"
JOB_GLIDEIN_SiteWMS_Slot = "Unknown"
Subproc = 0
ResidentSetSize = 154372
EventTypeName = 28
JOB_GLIDEIN_SiteWMS = "SGE"
...
012 (22869915.000.000) 08/19 21:36:15 Job was held.
    SYSTEM_PERIODIC_HOLD  Run Time/limit 436/300
    Code 26 Subcode 8
```

# Code lines for autorelease with Run Time

```
...
013 (22869915.000.000) 08/19 21:36:24 Job was released.
    The job attribute PeriodicRelease expression 'ShouldRelease' evaluated to TRUE
...
028 (22869915.000.000) 08/19 21:36:24 Job ad information event triggered.
JOB_GLIDEIN_Name = "$$(GLIDEIN_Name:Unknown)"
JOB_Site = "$$(GLIDEIN_Site:Unknown)"
Proc = 0
JOB_GLIDEIN_Entry_Name = "$$(GLIDEIN_Entry_Name:Unknown)"
EventTime = "2019-08-19T21:36:24"
TriggerEventTypeName = "ULOG_JOB_RELEASED"
JOB_GLIDEIN_SiteWMS_Queue = "$$(GLIDEIN_SiteWMS_Queue:Unknown)"
TriggerEventTypeName = 13
JOB_GLIDEIN_Site = "$$(GLIDEIN_Site:Unknown)"
JOB_GLIDEIN_SiteWMS_JobId = "$$(GLIDEIN_SiteWMS_JobId:Unknown)"
MyType = "JobReleaseEvent"
JOB_GLIDEIN_ProcId = "$$(GLIDEIN_ProcId:Unknown)"
JOB_GLIDEIN_Schedd = "$$(GLIDEIN_Schedd:Unknown)"
JOB_GLIDEIN_ClusterId = "$$(GLIDEIN_ClusterId:Unknown)"
Cluster = 22869915
JOB_GLIDEIN_Factory = "$$(GLIDEIN_Factory:Unknown)"
Reason = "The job attribute PeriodicRelease expression 'ShouldRelease' evaluated to TRUE"
JOB_GLIDEIN_SiteWMS_Slot = "$$(GLIDEIN_SiteWMS_Slot:Unknown)"
Subproc = 0
EventTypeName = 28
JOB_GLIDEIN_SiteWMS = "$$(GLIDEIN_SiteWMS:Unknown)"
...
```

# Code lines for autorelease with Run Time

```
...
005 (22869915.000.000) 08/19 23:26:44 Job terminated.
    (1) Normal termination (return value 1)
        Usr 0 01:45:00, Sys 0 00:00:37 - Run Remote Usage
        Usr 0 00:00:00, Sys 0 00:00:00 - Run Local Usage
        Usr 0 01:45:00, Sys 0 00:00:37 - Total Remote Usage
        Usr 0 00:00:00, Sys 0 00:00:00 - Total Local Usage
3279448 - Run Bytes Sent By Job
73026 - Run Bytes Received By Job
3279448 - Total Bytes Sent By Job
146052 - Total Bytes Received By Job
Partitionable Resources :      Usage Request Allocated
  Cpus                   :          0.00           1           1
  Disk (KB)              : 2263535      26214400  36621171
  Memory (MB)            :          1325           2500           2500
...
```

# Code lines for autorelease with Memory

*lines\_1 = +DUNE\_OriginalMemory=10*

*lines\_2 = +DUNE\_GraceMemory=3000*

*lines\_3 = +DUNE\_IncreaseReqMem=(NumJobStarts>0)&&(!  
isUndefined(LastHoldReasonCode))&&((LastHoldReasonCode=?=26&&  
LastHoldReasonSubCode=?=1)||((LastHoldReasonCode=?=34)))*

*lines\_4 = +DUNE\_ShouldRelease=(JobStatus=?=5)&&  
((HoldReasonCode=?=26&&HoldReasonSubCode=?=1)||  
(HoldReasonCode=?=34))&&(!isUndefined(MachineAttrMemory0))&&  
(MemoryUsage<(DUNE\_OriginalMemory+DUNE\_GraceMemory))*

*lines\_5 = request\_memory=ifthenelse(DUNE\_IncreaseReqMem,  
DUNE\_OriginalMemory+DUNE\_GraceMemory,DUNE\_OriginalMemory)*

*lines\_6 = periodic\_release=DUNE\_ShouldRelease*

*lines\_7 = job\_machine\_attrs=Memory*

# Code lines for autorelease with Memory

```
...
028 (22405182.000.000) 08/27 15:35:51 Job ad information event triggered.
JOB_Site = "SGridECDF"
JOB_GLIDEIN_Name = "gfactory_instance"
Size = 2088152
Proc = 0
JOB_GLIDEIN_Entry_Name = "DUNE_UK_SGridECDF_cel"
EventTime = "2019-08-27T15:35:51"
TriggerEventTypeName = "ULOG_IMAGE_SIZE"
JOB_GLIDEIN_SiteWMS_Queue = "eddie"
MemoryUsage = 936
TriggerEventTypeName = 6
JOB_GLIDEIN_Site = "SGridECDF"
JOB_GLIDEIN_SiteWMS_JobId = "2917475"
MyType = "JobImageSizeEvent"
JOB_GLIDEIN_ProcId = "0"
JOB_GLIDEIN_Schedd = "schedd_glideins7@gfactory-2.opensciencegrid.org"
JOB_GLIDEIN_ClusterId = "596963"
Cluster = 22405182
JOB_GLIDEIN_Factory = "OSG"
JOB_GLIDEIN_SiteWMS_Slot = "Unknown"
Subproc = 0
ResidentSetSize = 957560
EventTypeName = 28
JOB_GLIDEIN_SiteWMS = "SGE"
...
012 (22405182.000.000) 08/27 15:35:51 Job was held.
        SYSTEM_PERIODIC_HOLD Memory/limit 936/1.0000000000000000E+01
        Code 26 Subcode 1
...
```

# Code lines for autorelease with Memory

```
...
013 (22405182.000.000) 08/27 15:36:22 Job was released.
    The job attribute PeriodicRelease expression 'DUNE_ShouldRelease' evaluated to TRUE
...
028 (22405182.000.000) 08/27 15:36:22 Job ad information event triggered.
JOB_GLIDEIN_Name = "$$(GLIDEIN_Name:Unknown)"
JOB_Site = "$$(GLIDEIN_Site:Unknown)"
Proc = 0
JOB_GLIDEIN_Entry_Name = "$$(GLIDEIN_Entry_Name:Unknown)"
EventTime = "2019-08-27T15:36:22"
TriggerEventTypeName = "ULOG_JOB_RELEASED"
JOB_GLIDEIN_SiteWMS_Queue = "$$(GLIDEIN_SiteWMS_Queue:Unknown)"
TriggerEventTypeName = 13
JOB_GLIDEIN_Site = "$$(GLIDEIN_Site:Unknown)"
JOB_GLIDEIN_SiteWMS_JobId = "$$(GLIDEIN_SiteWMS_JobId:Unknown)"
MyType = "JobReleaseEvent"
JOB_GLIDEIN_ProcId = "$$(GLIDEIN_ProcId:Unknown)"
JOB_GLIDEIN_Schedd = "$$(GLIDEIN_Schedd:Unknown)"
JOB_GLIDEIN_ClusterId = "$$(GLIDEIN_ClusterId:Unknown)"
Cluster = 22405182
JOB_GLIDEIN_Factory = "$$(GLIDEIN_Factory:Unknown)"
Reason = "The job attribute PeriodicRelease expression 'DUNE_ShouldRelease' evaluated to TRUE"
JOB_GLIDEIN_SiteWMS_Slot = "$$(GLIDEIN_SiteWMS_Slot:Unknown)"
Subproc = 0
EventTypeName = 28
JOB_GLIDEIN_SiteWMS = "$$(GLIDEIN_SiteWMS:Unknown)"
...
```



# Code lines for autorelease with Memory

```
...
005 (22405182.000.000) 08/28 02:48:25 Job terminated.
    (1) Normal termination (return value 0)
        Usr 0 11:04:03, Sys 0 00:02:34 - Run Remote Usage
        Usr 0 00:00:00, Sys 0 00:00:00 - Run Local Usage
        Usr 0 11:04:03, Sys 0 00:02:34 - Total Remote Usage
        Usr 0 00:00:00, Sys 0 00:00:00 - Total Local Usage
6707190 - Run Bytes Sent By Job
73133 - Run Bytes Received By Job
6707190 - Total Bytes Sent By Job
146266 - Total Bytes Received By Job
Partitionable Resources : Usage Request Allocated
  Cpus                   :          0          1          1
  Disk (KB)              : 2477516 26214400 27981112
  Memory (MB)           :          14         3010         3010
...
```

# Montecarlo Simulation Campaign

# Supernova samples of elastic scattering events

## Campaign Stage

Name: mcc11\_RITM0858667\_sn\_gen\_g4\_detsim\_reco

Id:4019

Experiment: dune

Dataset: None

Software Version: v08\_26\_00

VO Role: Production

Param Overrides: [

```
[-c; '/dune/app/home/dunepro/poms_MCC11/RITM0858667_sn/gen_g4_template_RITM0858667_sn.cfg']
```

```
[-Oglobal.version='%(version)s']
```

```
[-Oglobal.utilquals='e17:prof']
```

```
[-Oglobal.quals='e17:prof']
```

```
[-Oglobal.basename='RITM0858667_sn']
```

```
[-Oglobal.gen_g4fcfile='prodnuescatter_dunefd_wbkg.fcl']
```

```
[-Oglobal.detsimfclfile='standard_detsim_dune10kt_1x2x6_bestop.fcl']
```

```
[-Oglobal.recofclfile='standard_reco_dune10kt_1x2x6_opslicer.fcl']
```

```
[-Oglobal.nevents='260']
```

```
[-Osubmit.N='500']
```

```
[-Osubmit.expected-lifetime='12h']
```

```
[-Osubmit.memory='2500MB']
```

```
[-Osubmit.site='
```

```
CCIN2P3,CERN,CIEMAT,Clemson,Florida,FNAL,FZU,Liverpool,London,Manchester,Michigan,Nebraska,NIKHEF,Not
```

```
[-Oglobal.tarfile='RITM0858667_sn.tar.gz']
```

```
[-Oglobal.tardir='/pnfs/dune/resilient/users/dunepro']
```

```
[-Osubmit.email-to='kherner@fnal.gov']
```

```
[-Ojob_output.dest='/pnfs/dune/scratch/dunepro/dropbox/mcc11/RITM0858667_sn']
```

```
[-Oglobal.logdir='/pnfs/dune/scratch/dunepro/test_MCC11/logs/RITM0858667_sn']]
```

Split Type: None

Last Split: None

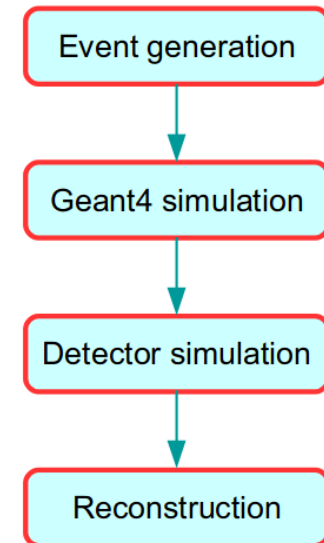
Created: 2019-08-09 18:10:55.084829-05:00

Creator: rschiatt

Updated: 2019-08-16 16:51:59.042236-05:00

Updater: rschiatt

Major processing steps are in a set of pre-defined fcl files



*A sample of 500 jobs with  
260 events per job*

# Reprocessing Campaign

# The Reprocessing Campaigns

- ✓ *Reprocessing good beam runs with dunetpc v08\_27\_01*
- ✓ *Generating new MC for 1,2,3,6,7 GeV conditions*
- ✓ *Originally expected 4-6 weeks for full campaign. On track for that*

# Progress as of 23 Sep 2019

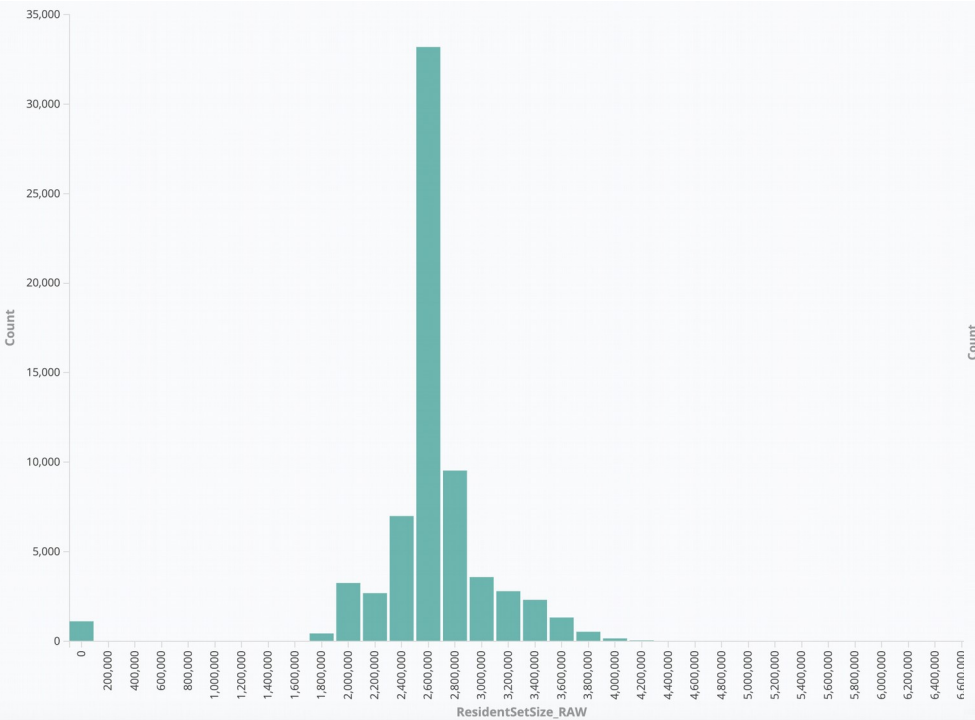
→ Began with high-priority data runs: 5387, 5809, 5770, 5834, 5826, 5432, 5786, 5204

| Data set  | Data % Complete | MC % complete |
|---|-----------------|---------------|
| Runs 5387, 5809, 5770, 5834, 5826, 5432, 5786, 5204 | 96.6%           | --            |
| 0.3 GeV   | 74.4%           | --            |
| 0.5 GeV   | 36.0%           | --            |
| 1 GeV   | 94.8%           | 99.5%         |
| 2 GeV   | 97.4%           | 105%          |
| 3 GeV   | 85.7%           | 92.4%         |
| 6 GeV   | 77.3%           | 97.8%         |
| 7 GeV   | 60.9%           | 103%          |

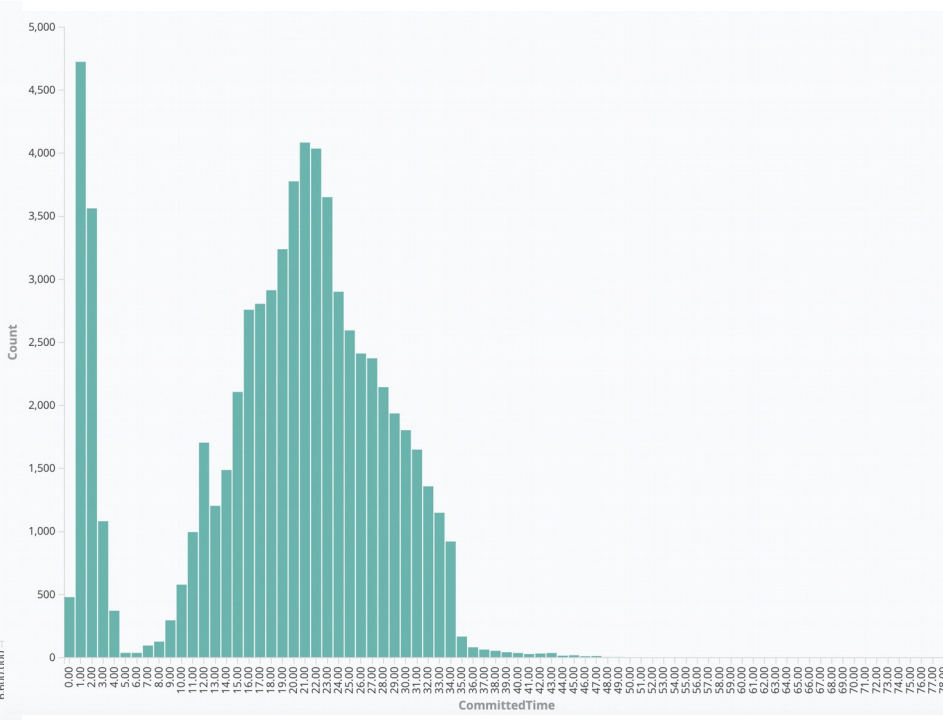
# Job Resource Requirements

- Data

Initial run time req is 35h;  
Longer recovery pass still running



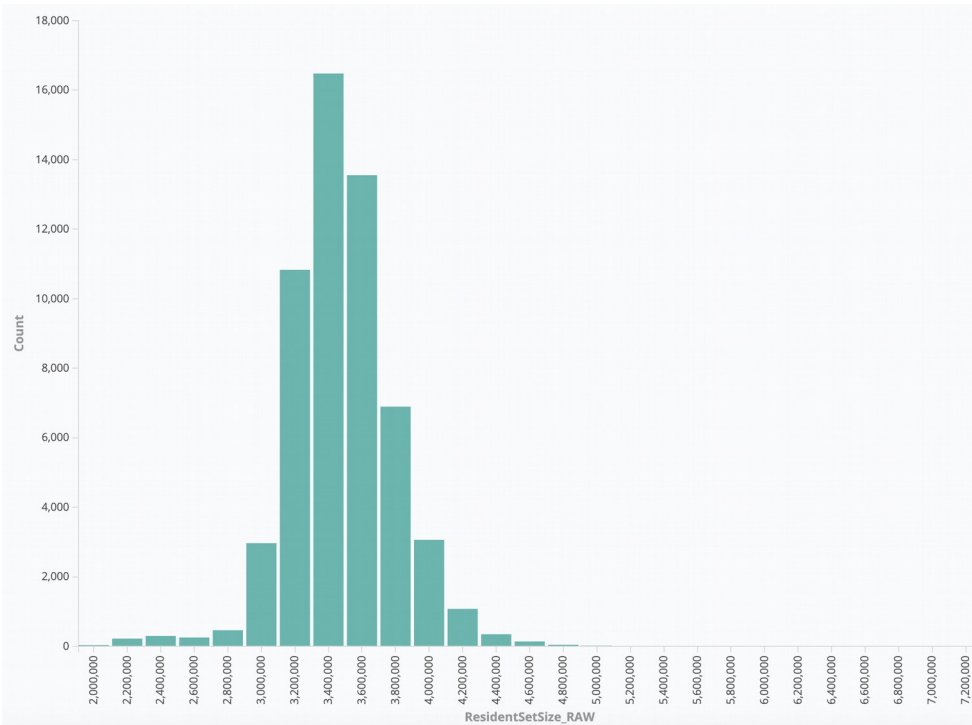
Memory (KB)



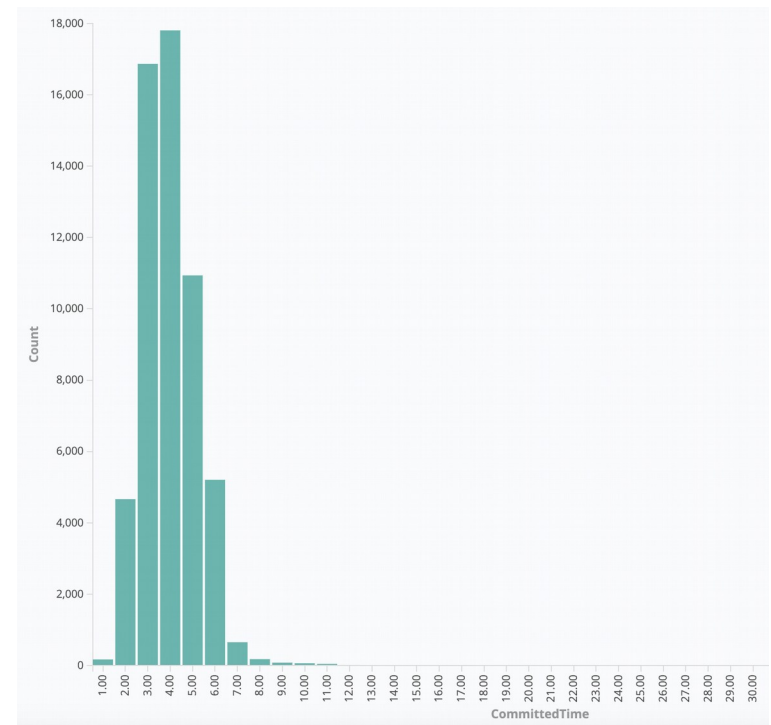
Run time(h)

# Job Resource Requirements

- MC
- Started with gen+g4+detsim stage followed by reco stage; Now running all stages in same job



Memory (KB)



Run time(h)



# New Production Data Validation

# Data Validation

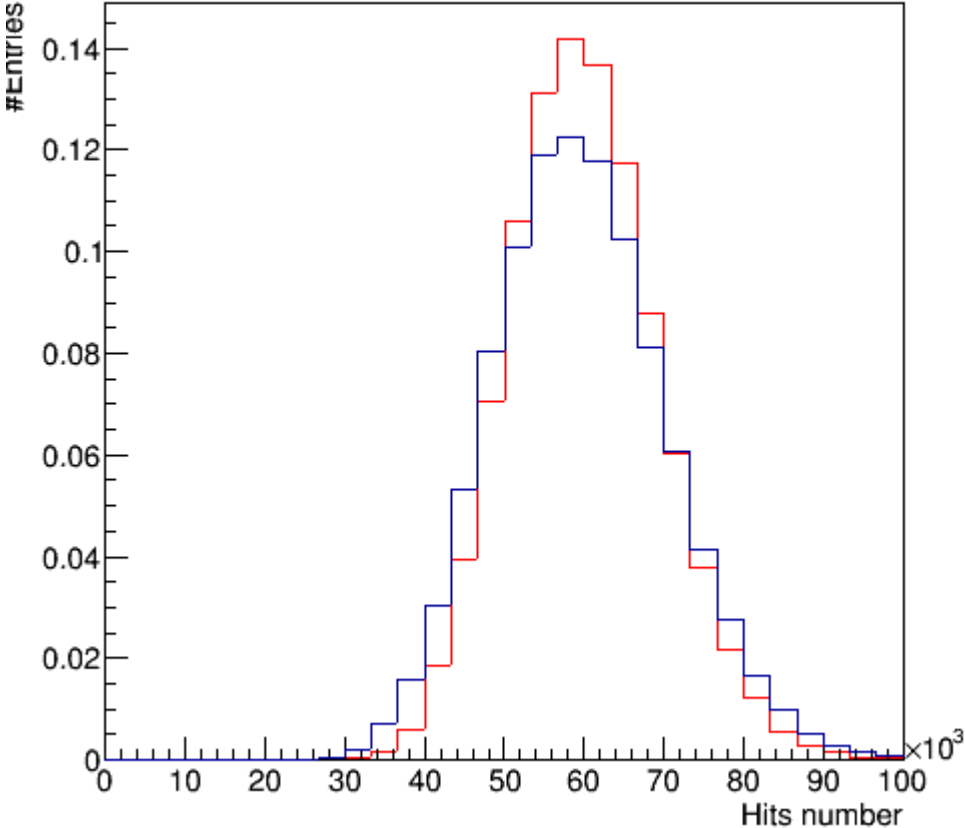
- ✓ *Comparison of the latest production files with the previous production sample with dunetpc version v07\_08\_00\_ %*
- ✓ *Starting from data run 5387.*
- ✓ *Important information about analyzable quantities as Hits, PFParticles, Tracks and Showers*

# About Hits..

----- Old Production data

----- New Production data

Old Processing Hits Number vs New Processing Hits Number

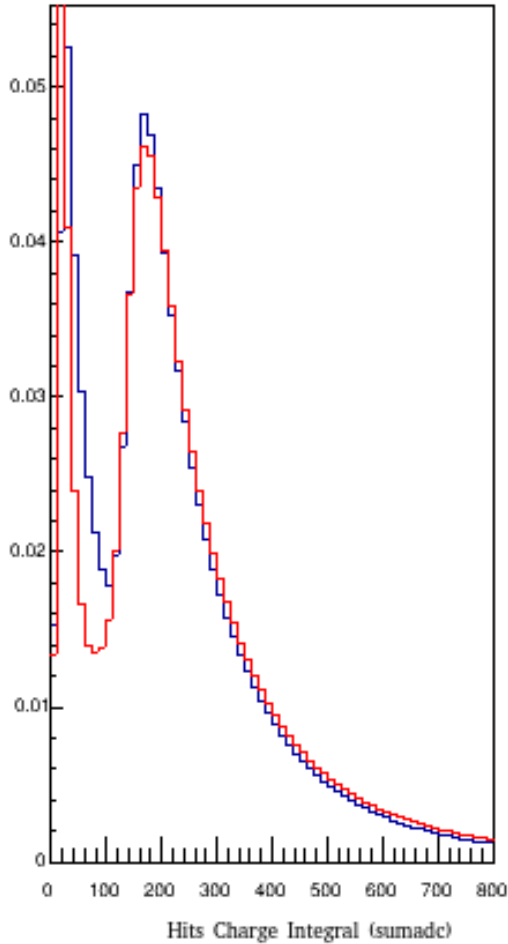


# About Hits..

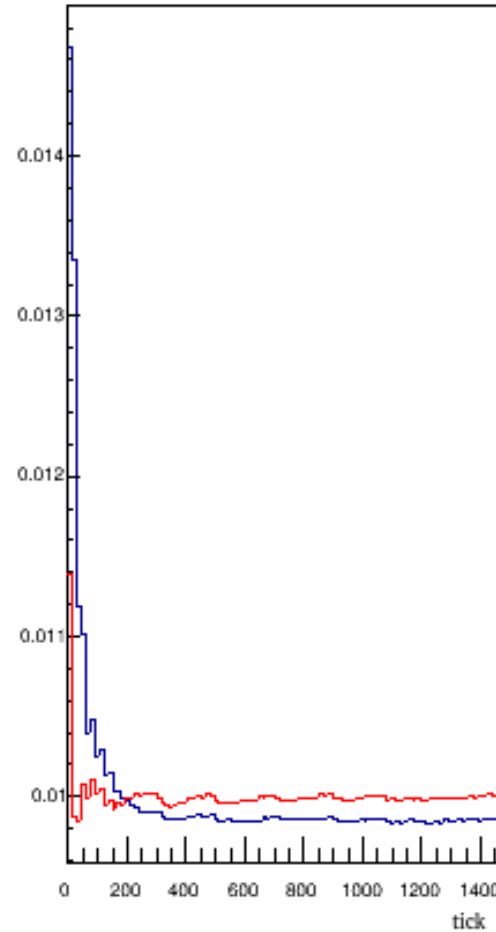
----- Old Production data

----- New Production data

Old Hits Charge Integral vs  
New HitsCharge Integral



Old Hits Peak Time vs  
New Hits Peak Time



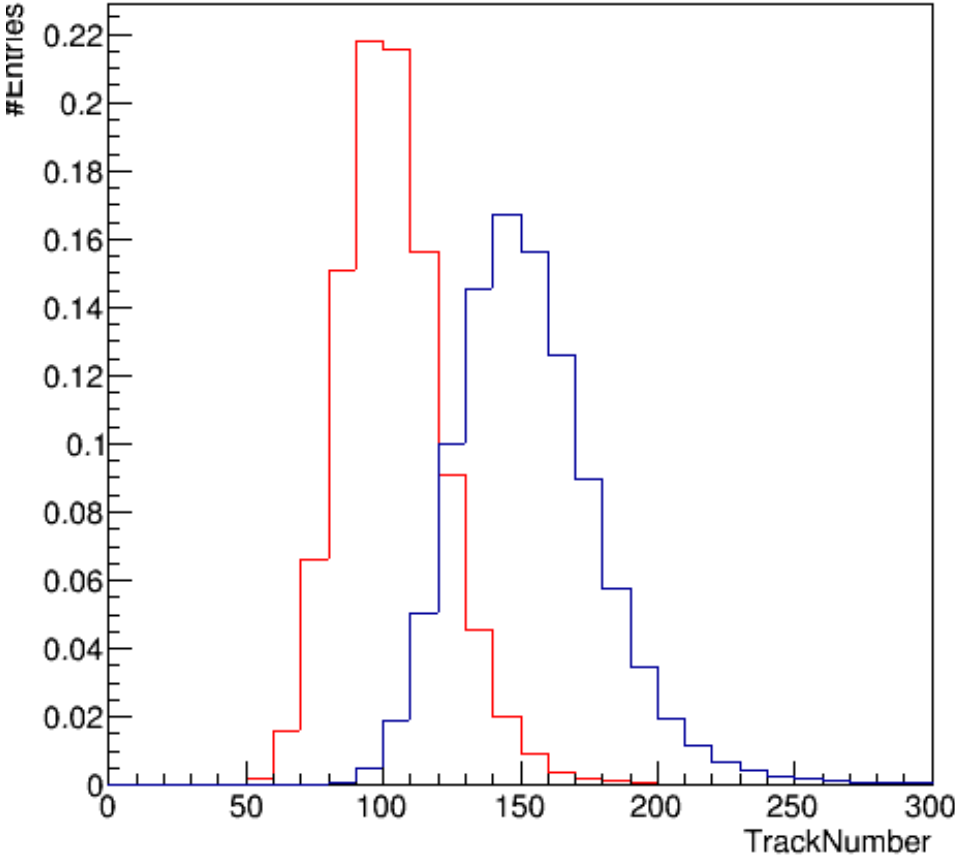
*Hits Charge Integral  
And Peak Time*

# About Tracks..

----- Old Production data

----- New Production data

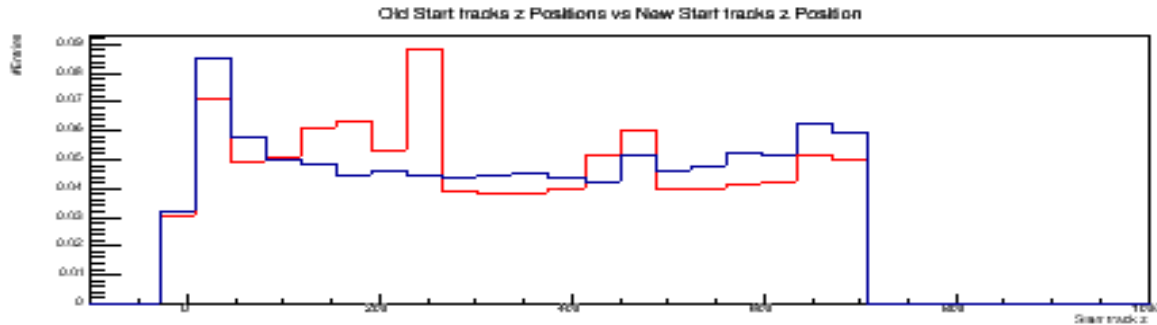
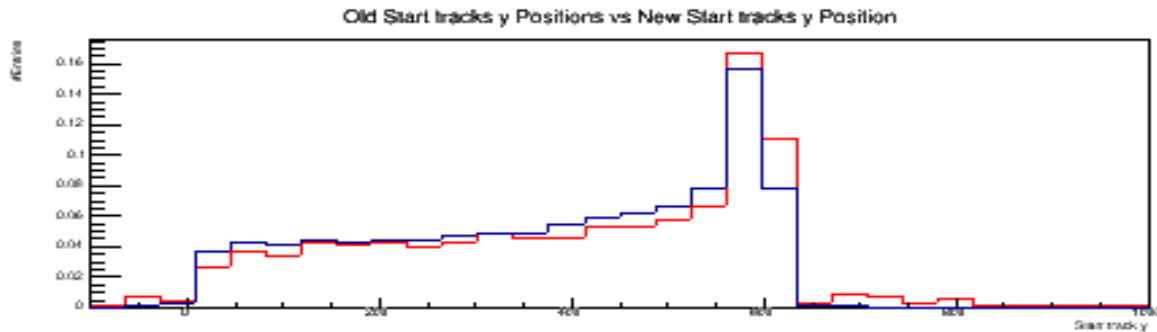
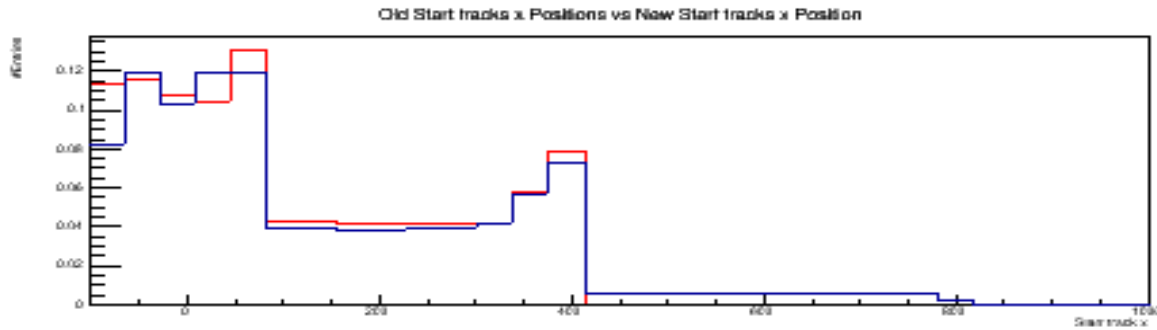
Old Processing Tracks Number vs New Processing Tracks Number



# About Tracks..

----- Old Production data

---- New Production data

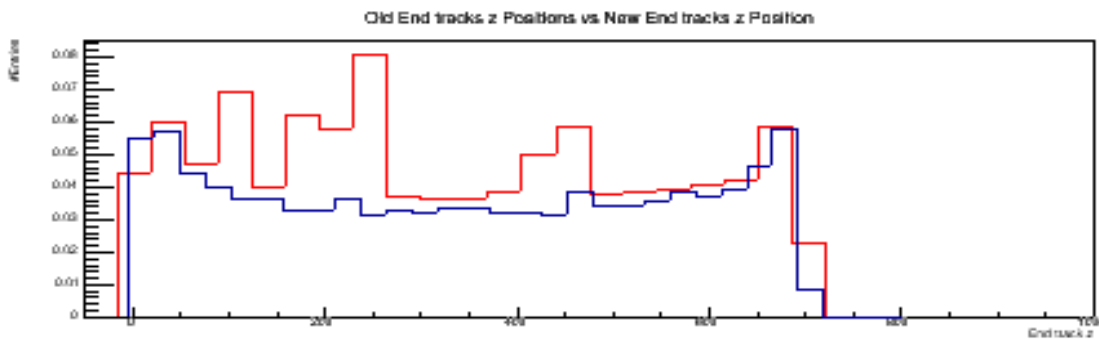
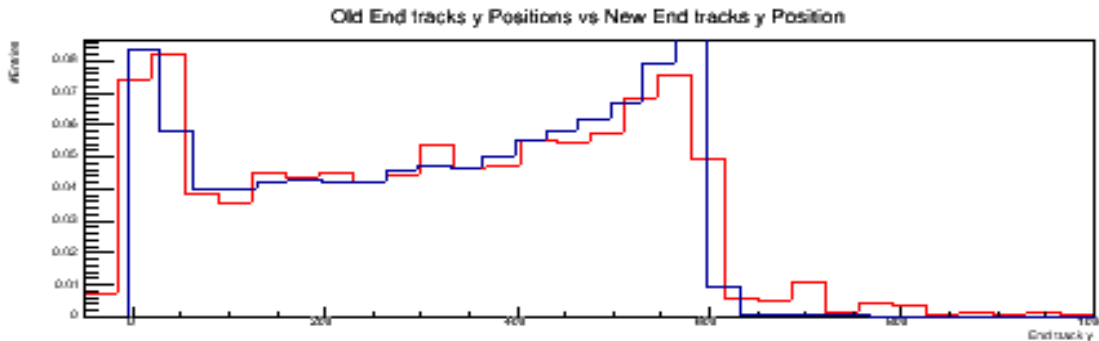
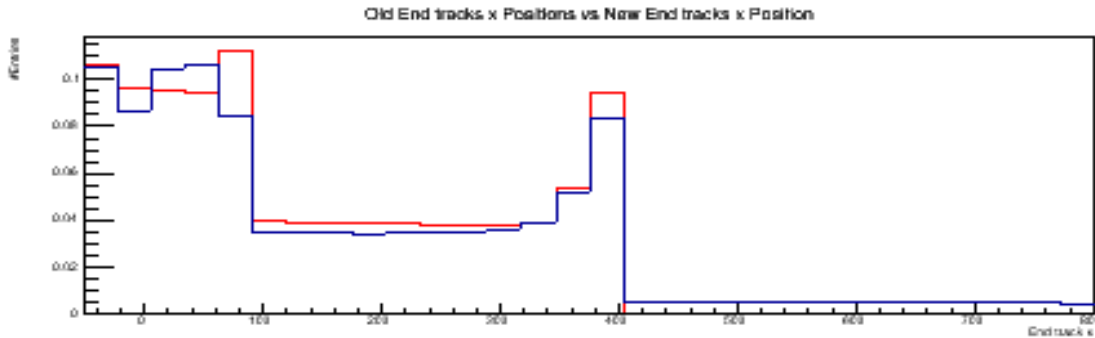


*Start Track Position (cm)*

# About Tracks..

----- Old Production data

----- New Production data

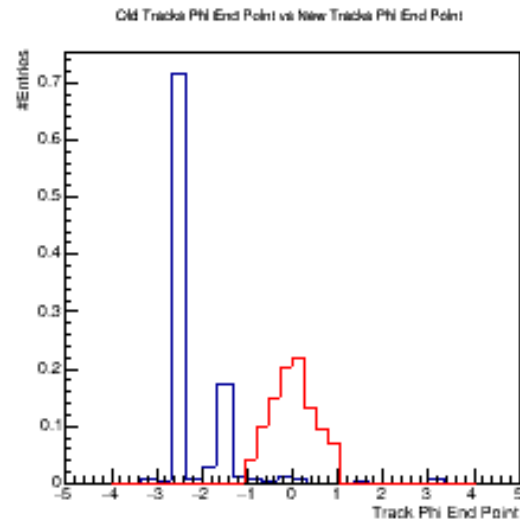
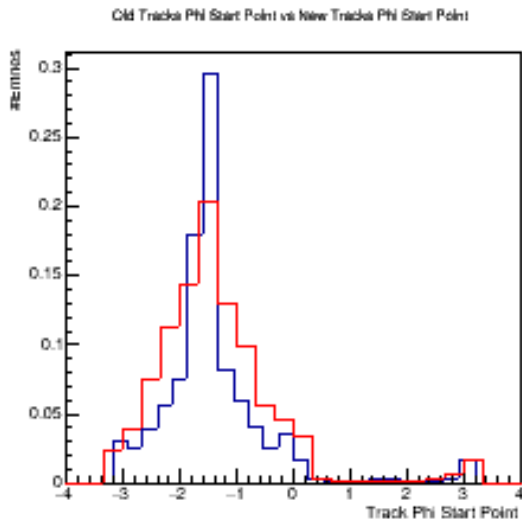
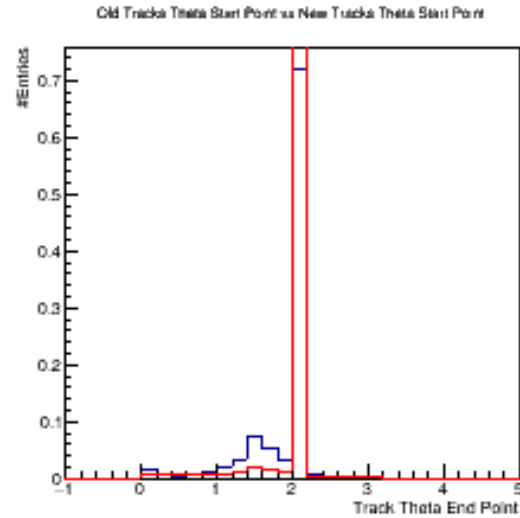
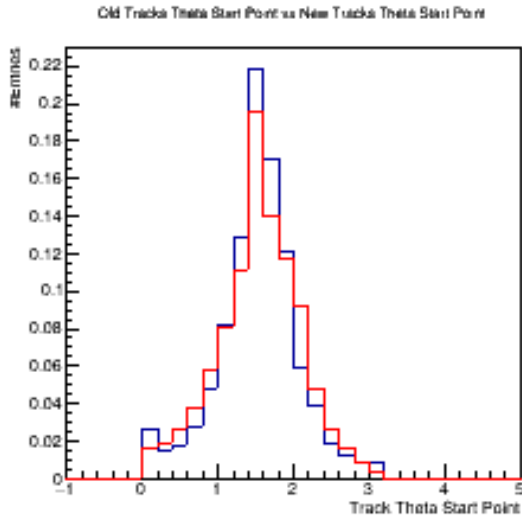


*End Track Position (cm)*

# About Tracks..

----- Old Production data

----- New Production data

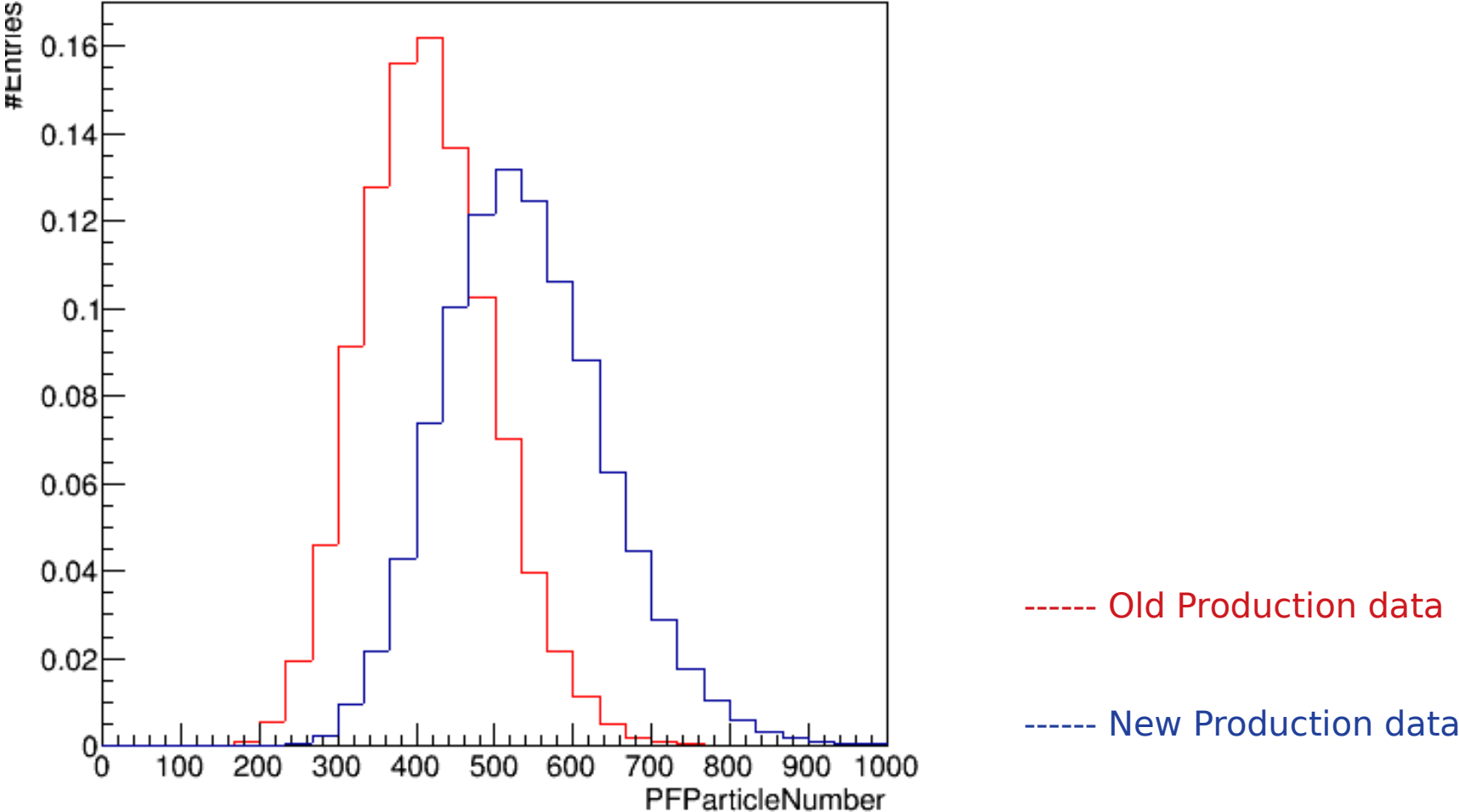


*Angular Distributions at Track Start and End Point*



# About PFParticles (Particle Flow Particles)..

Old Processing PFParticles Number vs New Processing PFParticles Number

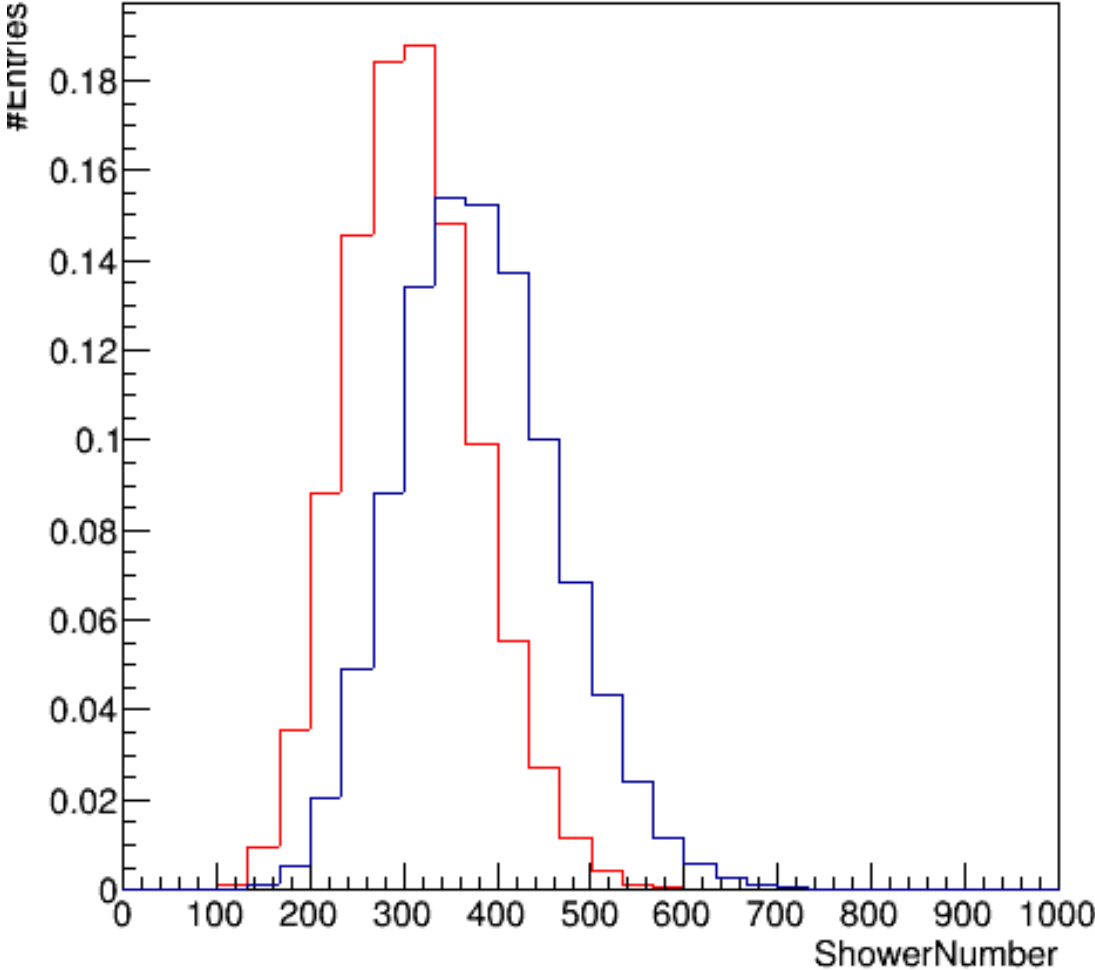


# About Showers..

----- Old Production data

----- New Production data

Old Processing Showers Number vs New Processing Showers Number

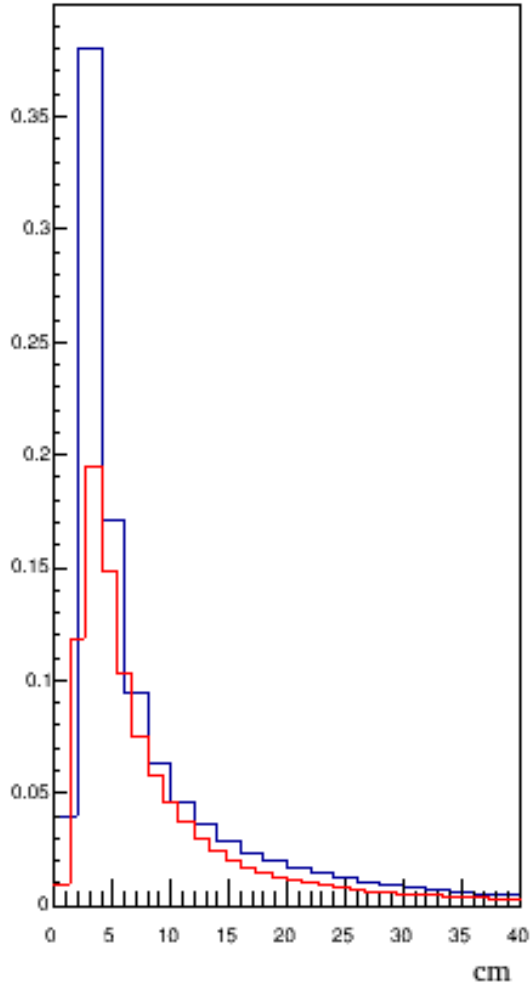


# About Showers..

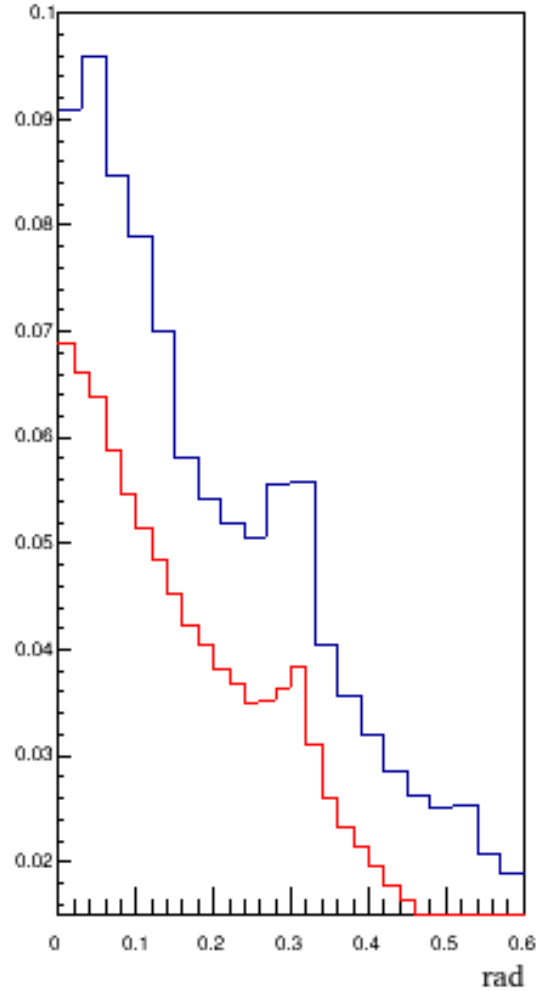
----- Old Production data

----- New Production data

Old Showers Length vs  
New Showers Length



Old Showers Open Angle vs  
New Showers Open Angle



# Other Distributions

- ✓ **About Tracks :**

  - Number trajectory points*

  - Momentum at track start point*

  - Momentum at track end points*

- ✓ **About Showers:**

  - Showers start point*

THANK YOU !