

ProtoDUNE HD Offline Data Processing

Barnali Chowdhury, Wenqiang Gu, Jay Jo, Tom Junk, Michael Kirby
Argonne National Laboratory, Fermilab & Brookhaven National Laboratory

Offline Data Processing for ProtoDUNE HD

- The ProtoDUNE SP reco strategy runs “Dataprep” module which writes full (i.e. no ROI filtering) processed waveforms as recob:Wire to the event data store
 - “larwirecell” provides recob:Wire to WireCell
 - which are used as input to the WCT (wirecell) module. WCT does deconvolution and ROI finding (and more) writing another recob::Wire container to be used in hit finding.
- What changed between ProtoDUNE SP and now?
 - DAQ started taking data in HDF5 format
 - We want to move away from using DataPrep
- Goal is
 - to replace DataPrep with WireCell
 - to move Dataprep functionality into Wire-Cell. This means communication with the tpc decoder and all ADC mitigations, noise removal, and pedestal finding etc.
 - to develop WCT Noise Filter for HD and integrate it with tpc raw decoder.
 - to develop PDHD reconstruction chain and put into production.

Offline Data Processing for ProtoDUNE HD

- Introducing HD TPC decoder as the first stage of reconstruction chain
- Fhicl file aimed to run officially for upcoming PDHD data reconstruction
 - `/exp/dune/app/users/barnali/ProtoDUNE2/srcs/dunesw/fcl/protodunhd/reco/testing_for_standard_reco_protoduneHD.fcl`

```
producers: {  
  # TPC wire signals  
  tpcrawdecoder: @local::PDHDTPCReaderDefaults  
  # WireCell  
  wclsdatahd: @local::protodunehd_nf  
  # Hit finder
```

```
reco: [ #ophit,  
        #opflash,  
        #opslicer,  
        tpcrawdecoder,  
        wclsdatahd  
        #gaushit,  
        #nhitsfilter,  
        #reco3d,
```

```
PDHDTPCReaderDefaults:  
{  
  module_type: "PDHDTPCReader"  
  InputLabel: "tpcrawdecoder:daq"  
  OutputInstance: "daq"  
  APAList: [ 1, 2, 3, 4 ]  
  DecoderToolParams:  
    @local::PDHDDDataInterfaceWIB3Defaults  
}
```

The snippet above locates the input `raw::RawDigit` collection in the `art::Event` by `raw_input_label: "tpcrawdecoder:daq"`.

Offline Data Processing for ProtoDUNE HD

- Wirecell configuration for reading “RawDigits”
- Configuration “protodunehd_nf” is added in https://github.com/DUNE/dunereco/blob/develop/dunereco/DUNEWireCell/wirecell_dunehd_nf.fcl#L143
 - It assumes a data product of RawDigit with label tpcrawdecoder:daq.
 - Only configured for noise filtering (nf)

```
producers: {  
  # TPC wire signals  
  tpcrawdecoder: @local::PDHDTPCReaderDefaults  
  # WireCell  
  wclsdatahd: @local::protodunehd_nf  
  # Hit finder
```

```
protodunehd_nf : {  
  module_type : WireCellToolkit  
  wcls_main: {  
    tool_type: WCLS  
    apps: ["Pgrapher"]  
    logsinks: ["stdout"]  
    logLevels: ["debug", "pgraph:info"]  
    plugins: ["WireCellGen", "WireCellSigProc", "WireCellRoot", "WireCellPgraph", "WireCellLarsoft"]  
    inputers: ["wclsRawFrameSource"]  
    outputers: ["wclsFrameSaver:nfsaver"]  
    configs: ["pgrapher/experiment/pdhd/wcls-nf.jsonnet"]  
    params : {  
      raw_input_label: "tpcrawdecoder:daq"  
      reality: "data"  
      signal_output_form: "sparse"  
    }  
  }  
  structs: {  
    clock_speed: @local::protodunehd_services.DetectorClocksService.ClockSpeedTPC  
  }  
}
```


Offline Data Processing for ProtoDUNE HD

- We also modify nf.jsonnet

`/exp/dune/app/users/barnali/ProtoDUNE2_HD/srcs/dunereco/
dunereco/DUNEWireCell/pdhd/nf.jsonnet`

```
function(params, anode, chndbobj, n, name='', dft=default_dft) {  
  local single = {  
    type: 'PDHOneChannelNoise',  
    name: name,  
    uses: [dft, chndbobj, anode],  
    data: {  
      noisedb: wc.tn(chndbobj),  
      anode: wc.tn(anode),  
      dft: wc.tn(dft),  
    },  
  },  
  local grouped = {  
    type: 'PDHCoherentNoiseSub',  
    name: name,  
    uses: [dft, chndbobj, anode],  
    data: {  
      noisedb: wc.tn(chndbobj),  
      anode: wc.tn(anode),  
      dft: wc.tn(dft),  
      rms_threshold: 0.0,  
    },  
  },  
}
```

- np04_hd Data Selection : Run no. 24720, 24726, 19337, and 16751
 - 24720 and 24726 are both mid-filling (taken in last couple of weeks) and noise runs
 - 19337 is a pulsar run (taken in February 2023)

np04_hd Data Selection

Relatively old : Run 19337

Fairly recent data: Run 24720/24726

```
-bash-4.2$ h5dump-shared -A /exp/dune/data/users/barnali/np04hd_raw_run024720_0418_dataflow0_datawriter_0_20240404T080920.hdf5 | head -1000
```

```
-bash-4.2$ h5dump-shared -A /exp/dune/data/users/barnali/np04_hd_run019337_0001_dataflow0_datawriter_0_20230213T153224.hdf5 | head -1000
```

```
GROUP "RawData" {  
  DATASET "Detector_Readout_0x00000064_WIBeth" {  
    DATATYPE H5T_STD_I8LE  
    DATASPACE SIMPLE { ( 928872, 1 ) / ( 928872, 1 ) }  
  }  
  DATASET "Detector_Readout_0x00000065_WIBeth" {  
    DATATYPE H5T_STD_I8LE  
    DATASPACE SIMPLE { ( 928872, 1 ) / ( 928872, 1 ) }  
  }  
  DATASET "Detector_Readout_0x00000066_WIBeth" {  
    DATATYPE H5T_STD_I8LE  
    DATASPACE SIMPLE { ( 928872, 1 ) / ( 928872, 1 ) }  
  }  
  DATASET "Detector_Readout_0x00000067_WIBeth" {  
    DATATYPE H5T_STD_I8LE  
    DATASPACE SIMPLE { ( 928872, 1 ) / ( 928872, 1 ) }  
  }  
  DATASET "Detector_Readout_0x00000068_WIBeth" {  
    DATATYPE H5T_STD_I8LE  
    DATASPACE SIMPLE { ( 928872, 1 ) / ( 928872, 1 ) }  
  }  
}
```

```
GROUP "RawData" {  
  DATASET "Detector_Readout_0x00000000_WIB" {  
    DATATYPE H5T_STD_I8LE  
    DATASPACE SIMPLE { ( 3866696, 1 ) / ( 3866696, 1 ) }  
  }  
  DATASET "Detector_Readout_0x00000001_WIB" {  
    DATATYPE H5T_STD_I8LE  
    DATASPACE SIMPLE { ( 3866696, 1 ) / ( 3866696, 1 ) }  
  }  
  DATASET "Detector_Readout_0x00000002_WIB" {  
    DATATYPE H5T_STD_I8LE  
    DATASPACE SIMPLE { ( 3866696, 1 ) / ( 3866696, 1 ) }  
  }  
  DATASET "Detector_Readout_0x00000003_WIB" {  
    DATATYPE H5T_STD_I8LE  
    DATASPACE SIMPLE { ( 3866696, 1 ) / ( 3866696, 1 ) }  
  }  
  DATASET "Detector_Readout_0x00000004_WIB" {  
    DATATYPE H5T_STD_I8LE  
    DATASPACE SIMPLE { ( 3866696, 1 ) / ( 3866696, 1 ) }  
  }  
}
```

➤ Run successful test with

- `lar -n l -c testing_for_standard_reco_protoduneHD.fcl -s /exp/dune/data/users/barnali/np04hd_raw_run024720_0418_dataflow0_datawriter_0_20240404T080920.hdf5`
- `lar -n l -c testing_for_standard_reco_protoduneHD.fcl -s /exp/dune/data/users/barnali/np04_hd_run019337_0001_dataflow0_datawriter_0_20230213T153224.hdf5`

Testing np04_hd Data

- `lar -nl -c eventdump.fcl`
`np04hd_raw_run024720_0418_dataflow0_datawriter_0_20240404T0809`
`20_reco.root`

```
Begin processing the 1st record. run: 24720 subRun: 1 event: 11709 at 04-Apr-2024 15:57:45 CDT
PRINCIPAL TYPE: Event
PROCESS NAME | MODULE LABEL.. | PRODUCT INSTANCE NAME | DATA PRODUCT TYPE..... | .SIZE
Reco..... | TriggerResults | ..... | art::TriggerResults..... | ....1
Reco..... | tpcrawdecoder. | daq..... | art::Assns<raw::RawDigit,raw::RDTimeStamp,void> | 40960
Reco..... | tpcrawdecoder. | daq..... | std::vector<raw::RawDigit>..... | 40960
Reco..... | wclsdatahd.... | raw..... | std::vector<recob::Wire>..... | 10240
Reco..... | tpcrawdecoder. | daq..... | std::vector<raw::RDTimeStamp>..... | 40960
Reco..... | wclsdatahd.... | badmasks..... | std::vector<int>..... | ....0
Reco..... | tpcrawdecoder. | daq..... | std::vector<raw::RDStatus>..... | ....1
Reco..... | daq..... | ..... | raw::DUNEHDF5FileInfo2..... | ....-
Reco..... | daq..... | trigger..... | raw::RDTimeStamp..... | ....-
Reco..... | wclsdatahd.... | badchannels..... | std::vector<int>..... | ....0

Total products (present, not present): 10 (10, 0).
```

- `lar -nl -c eventdump.fcl`
`np04_hd_run019337_0001_dataflow0_datawriter_0_20230213T153224_r`
`eco.root`

```
PROCESS NAME | MODULE LABEL.. | PRODUCT INSTANCE NAME | DATA PRODUCT TYPE..... | .SIZE
Reco..... | TriggerResults | ..... | art::TriggerResults..... | ....1
Reco..... | tpcrawdecoder. | daq..... | art::Assns<raw::RawDigit,raw::RDTimeStamp,void> | .2560
Reco..... | tpcrawdecoder. | daq..... | std::vector<raw::RawDigit>..... | .2560
Reco..... | wclsdatahd.... | raw..... | std::vector<recob::Wire>..... | 10240
Reco..... | tpcrawdecoder. | daq..... | std::vector<raw::RDTimeStamp>..... | .2560
Reco..... | wclsdatahd.... | badmasks..... | std::vector<int>..... | ....0
Reco..... | tpcrawdecoder. | daq..... | std::vector<raw::RDStatus>..... | ....1
Reco..... | daq..... | ..... | raw::DUNEHDF5FileInfo2..... | ....-
Reco..... | daq..... | trigger..... | raw::RDTimeStamp..... | ....-
Reco..... | wclsdatahd.... | badchannels..... | std::vector<int>..... | ....0
```

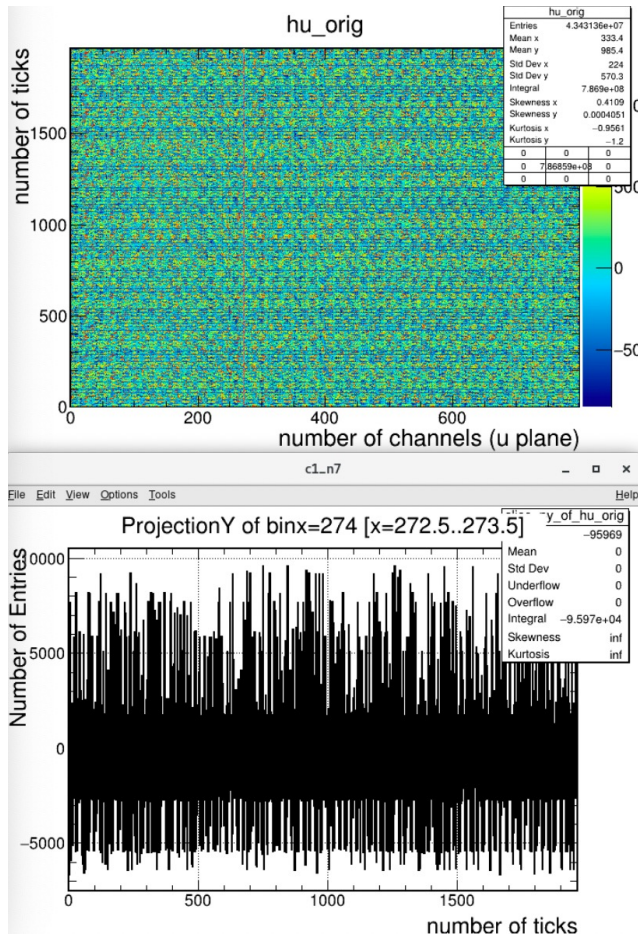
Implementation of HD Noise Filter

- There are two types of noise filters:
 - Single-channel : RC undershoot correction (aka “tail removal”) and the pedestal removal.
 - Group-channel : Coherent noise removal
 - **channel calibration for electronics response is not implemented**
- First one is implemented in
 - WireCell::Waveform::ChannelMaskMap PDHD::OneChannelNoise::apply(int ch, signal_t& signal) const
 - <https://github.com/WireCell/wire-cell-toolkit/blob/master/sigproc/src/ProtoduneHD.cxx#L726>
- Coherent noise removal is implemented in
 - WireCell::Waveform::ChannelMaskMap PDHD::CoherentNoiseSub::apply(channel_signals_t& chansig) const
 - <https://github.com/WireCell/wire-cell-toolkit/blob/master/sigproc/src/ProtoduneHD.cxx#L809C1-L809C99>
- There could be some other noise removal components we need to work on when protoDUNE HD data arrives

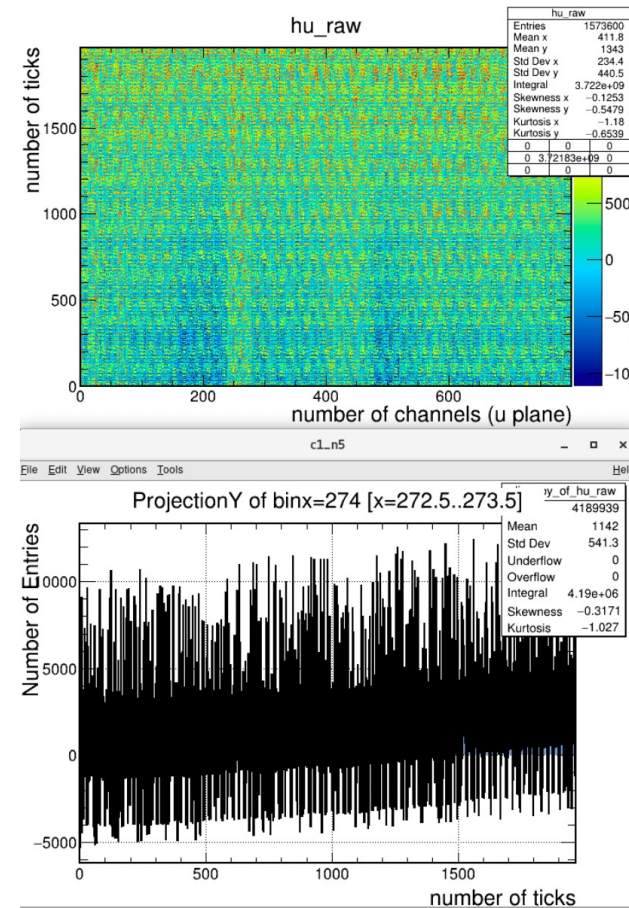
First look at ProtoDUNE HD noise data with WireCell noise filter

Run 24726 : Effect of Noise Filter

Before NF

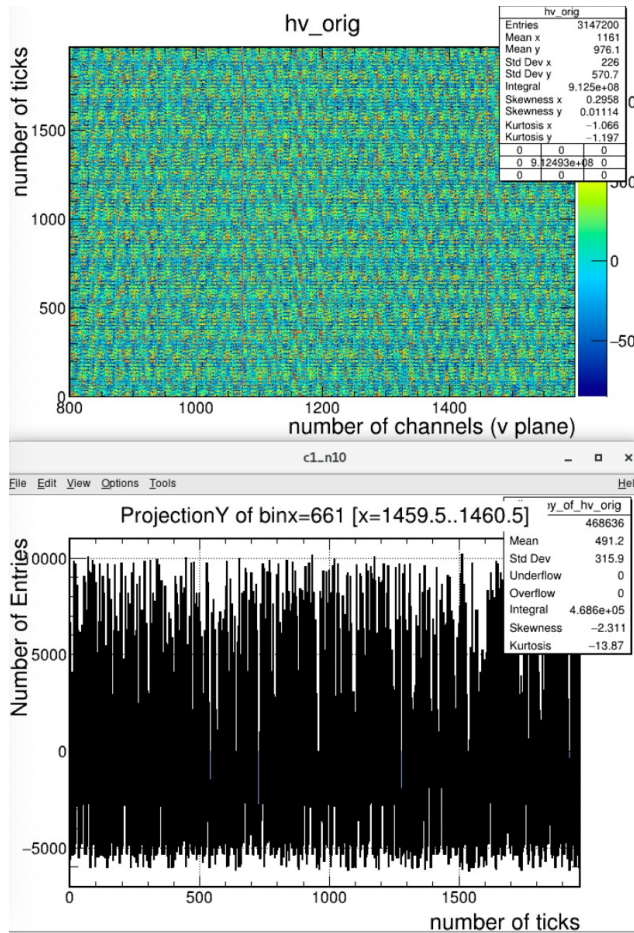


After NF

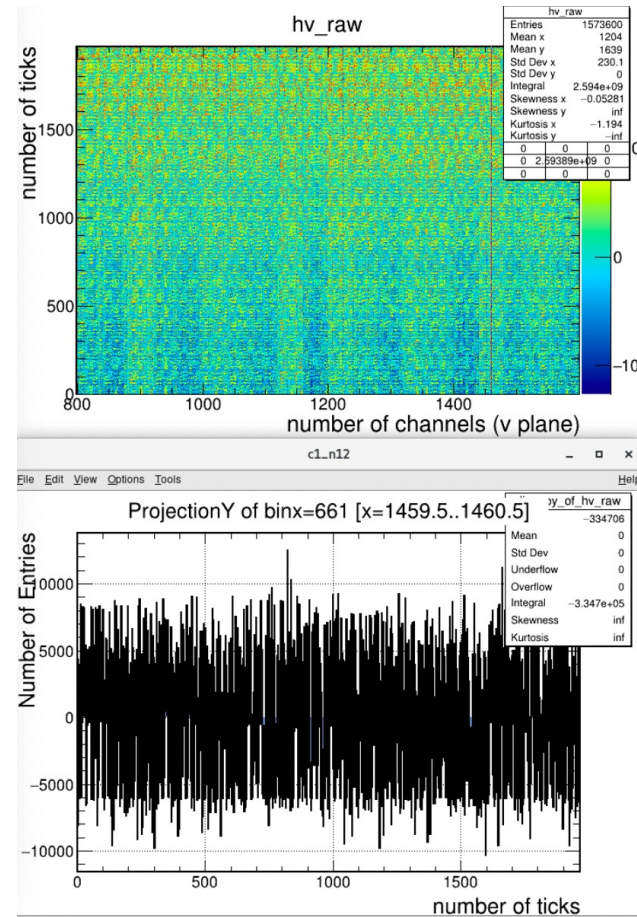


Run 24726 : Effect of Noise Filter

Before NF

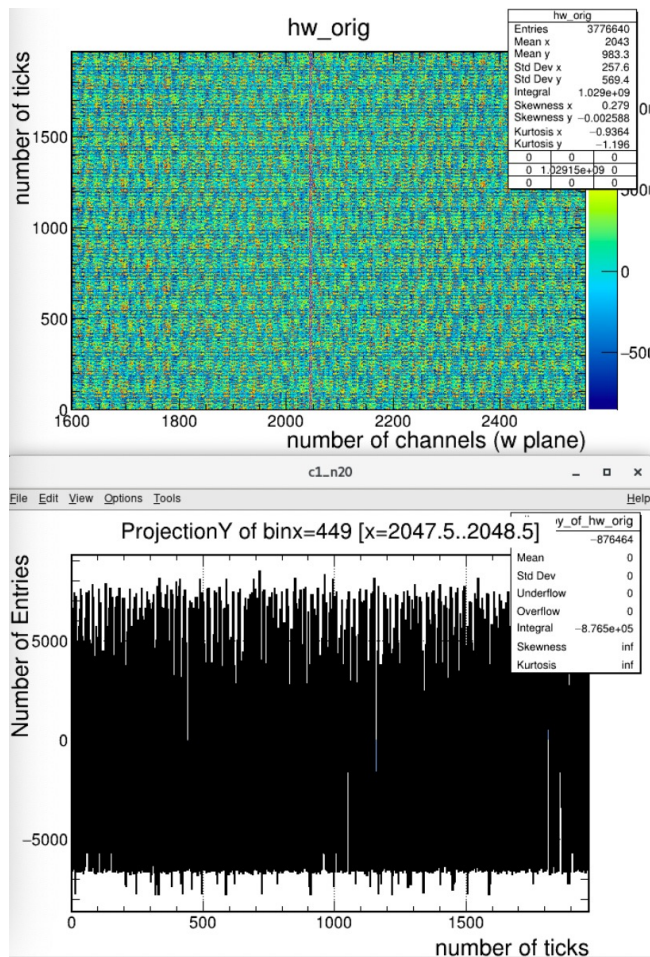


After NF

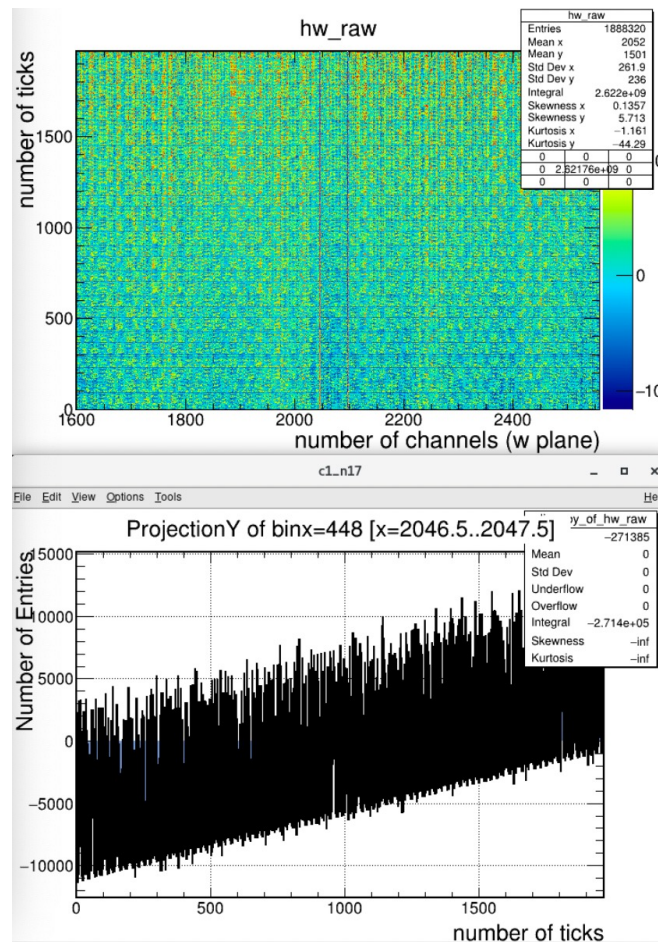


Run 24726 : Effect of Noise Filter

Before NF

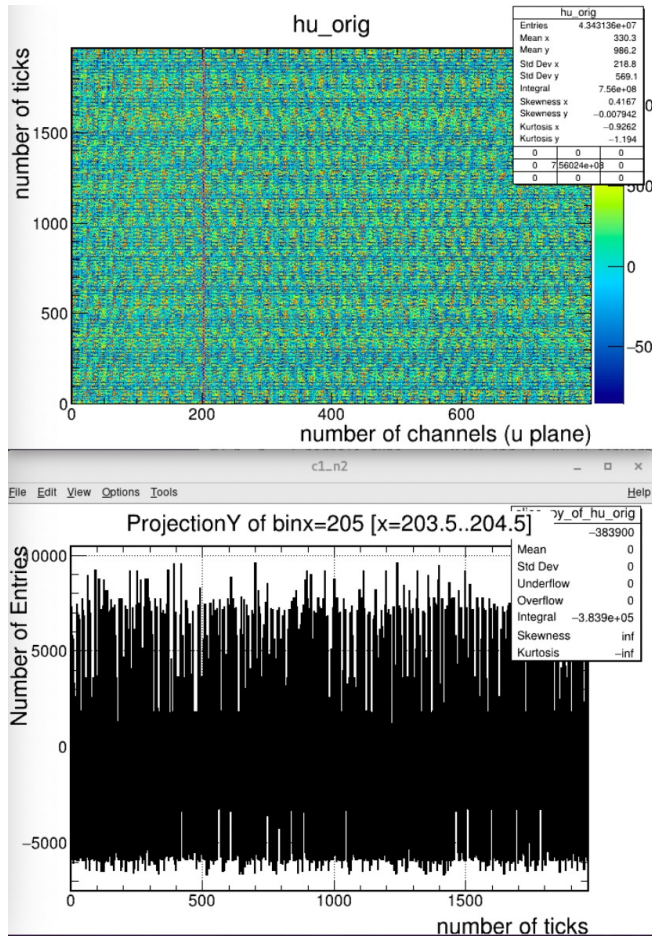


After NF

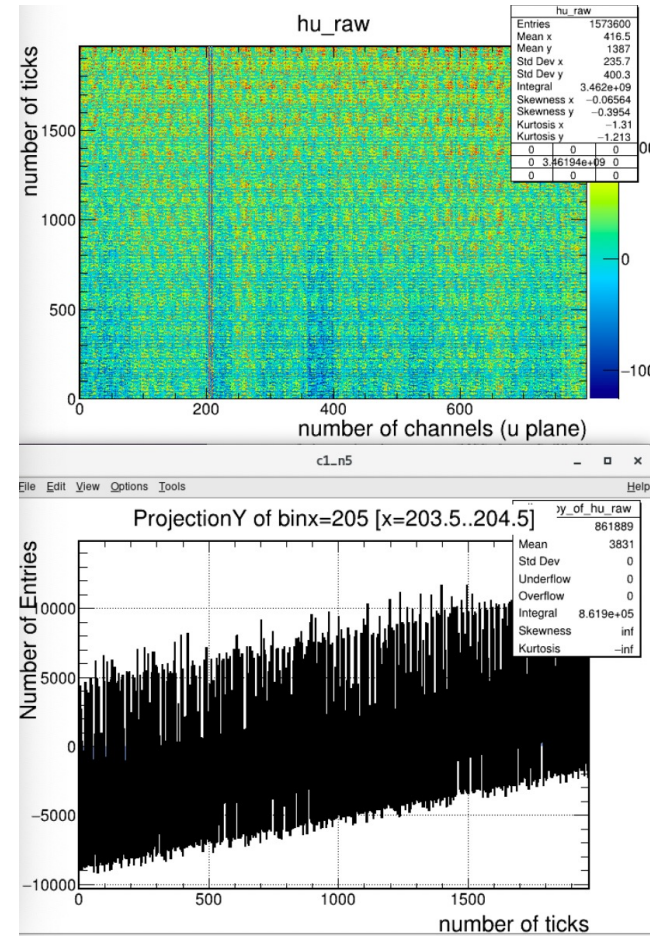


Run 24720 : Effect of Noise Filter

Before NF

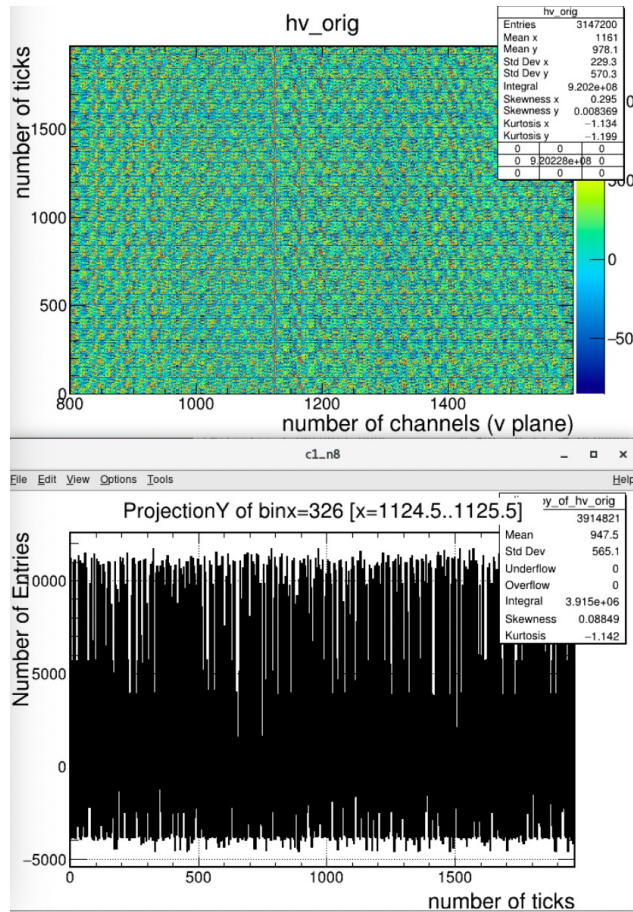


After NF

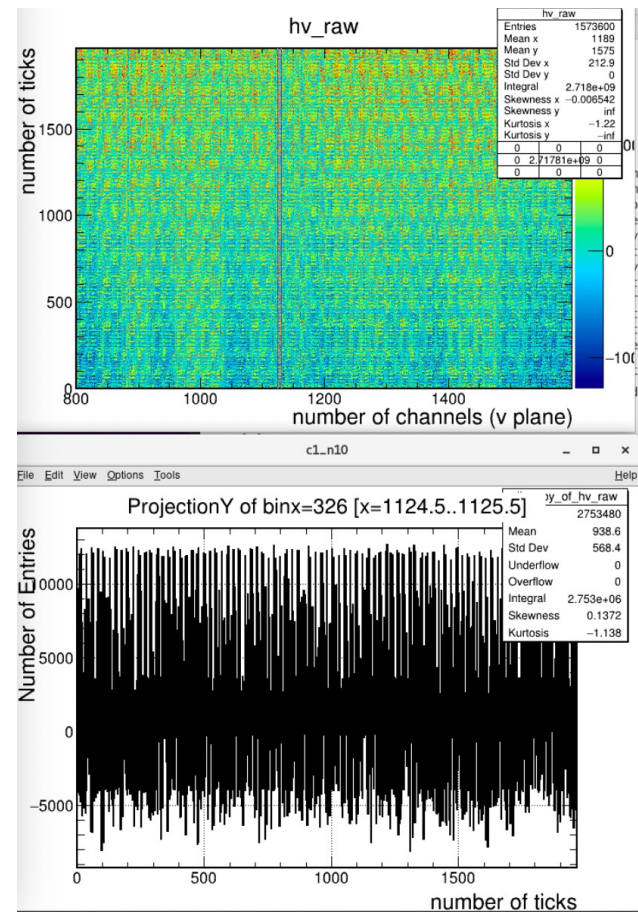


Run 24720 : Effect of Noise Filter

Before NF

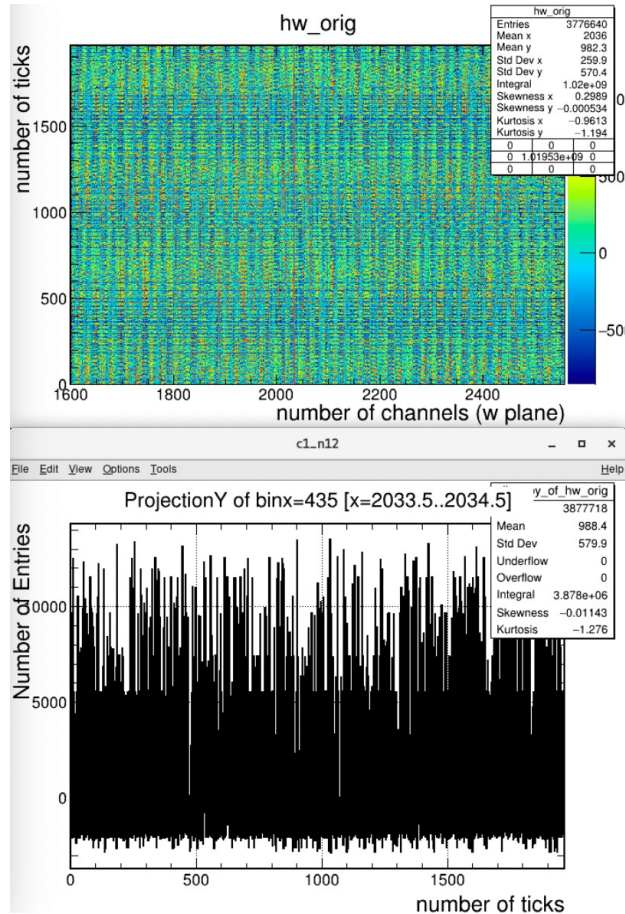


After NF

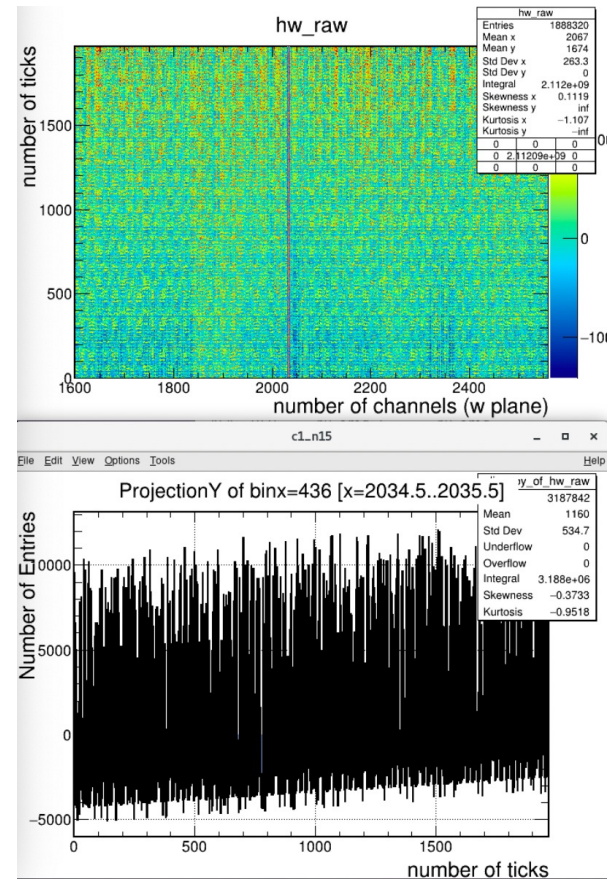


Run 24720 : Effect of Noise Filter

Before NF

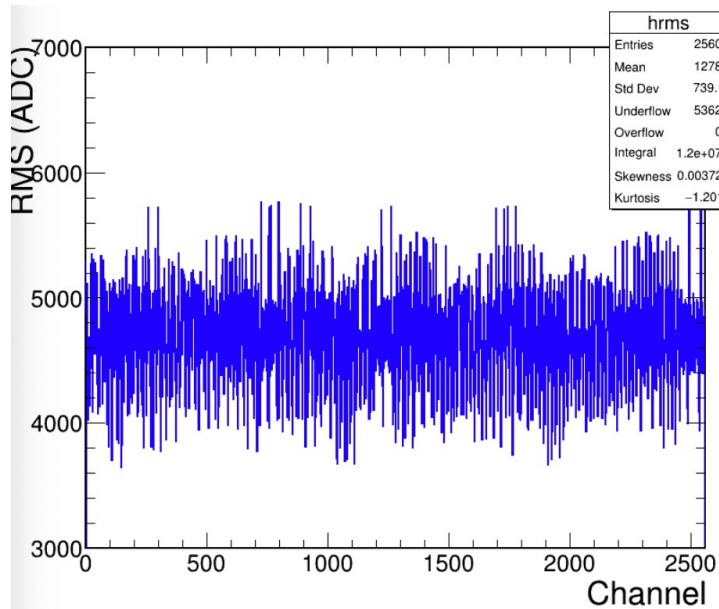


After NF

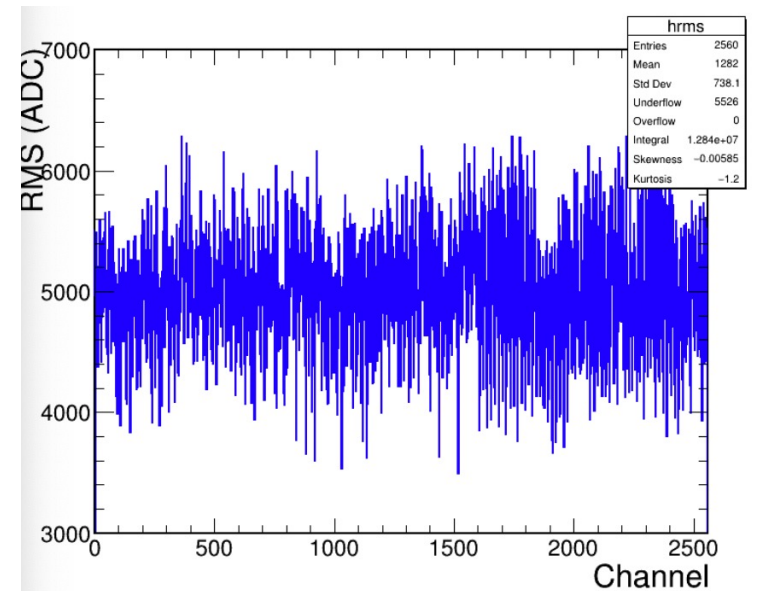


Run 24720 : Effect of Noise Filter

Before NF

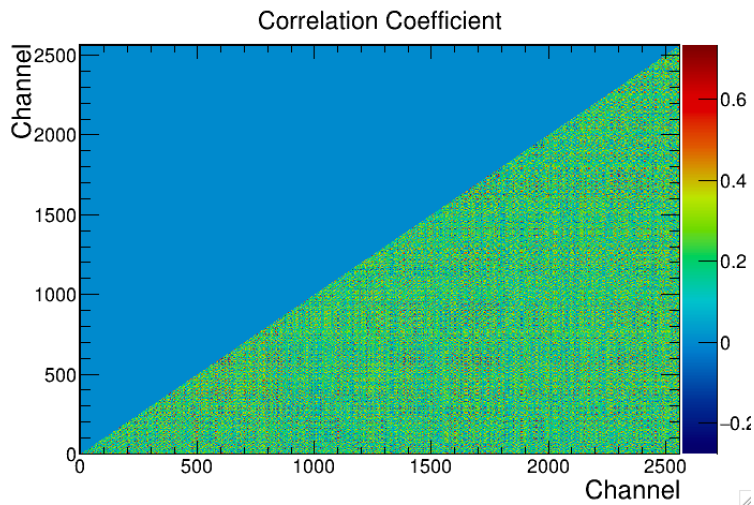


After NF

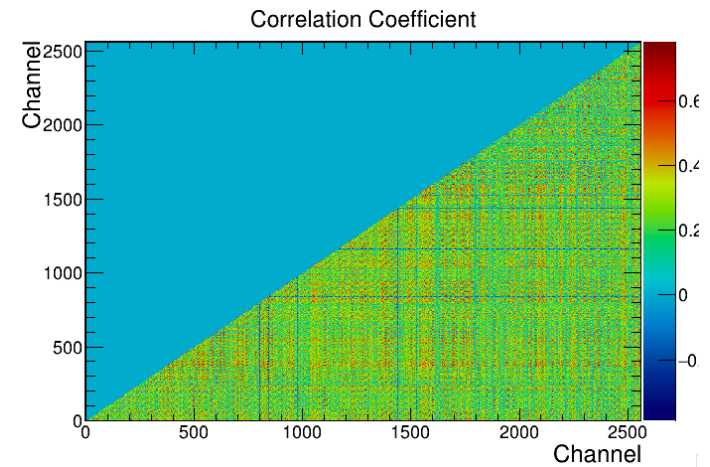


Run 24720 : Linear Correlation Between Channels

Before NF



After NF

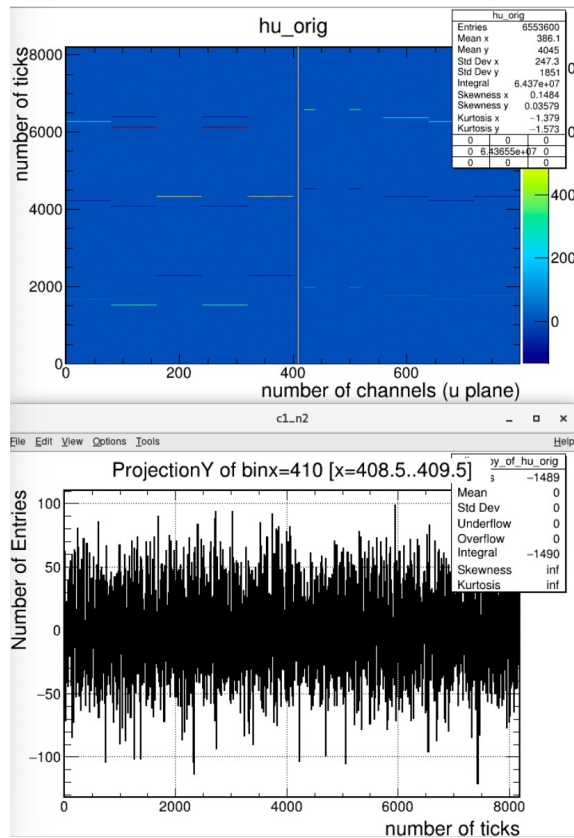


Reason for observing unexpected behavior for Run no. 24720 & 24726

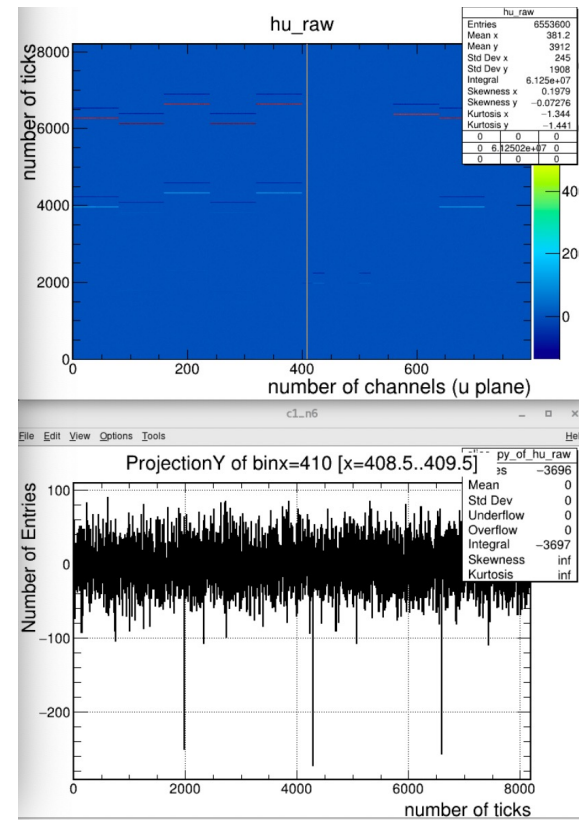
- Chatting with Tom
 - Hi Barnali -- Kurt sent me a message saying that they changed the data and file format (again!) But that the software will not be finalized until April 17. I've been holding off copying the DAQ versions for the offline until it is tagged and released, as Kurt says further changes are possible. He said some files have "snuck through" with the new format.

Run 19337 : Effect of Noise Filter

Before NF



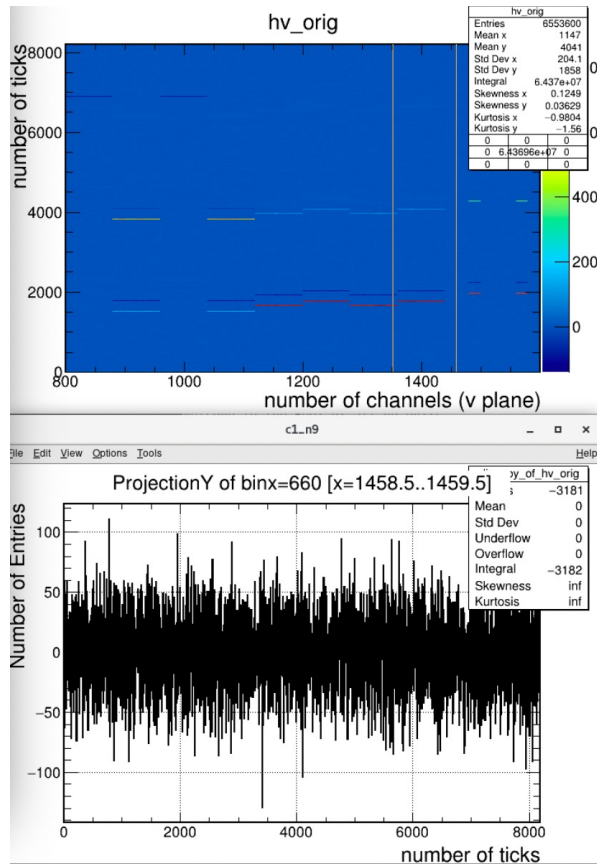
After NF



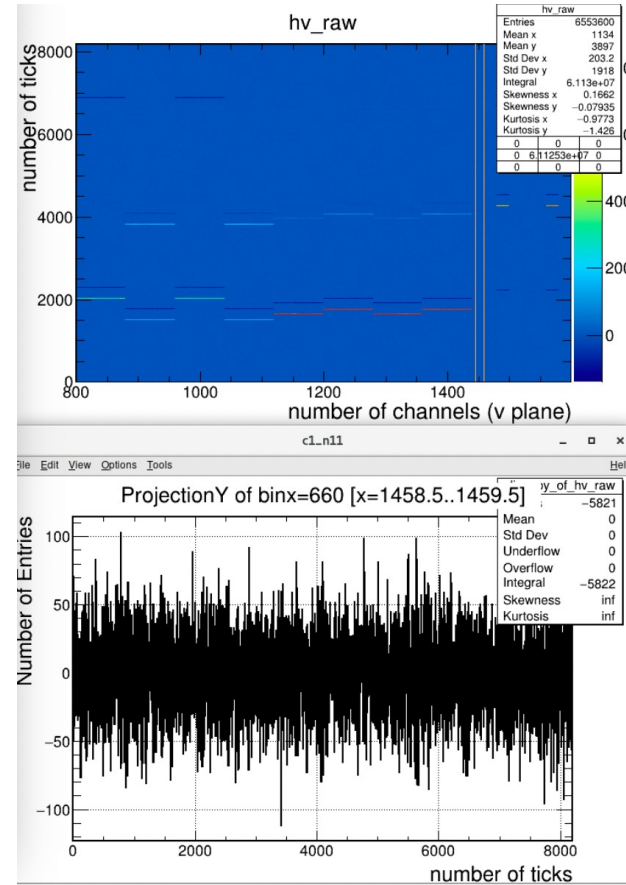
- This is pulsar data, but selected channel is seeing only noise
- The negative peak after noise filter is interesting --- need further investigation

Run 19337 : Effect of Noise Filter

Before NF



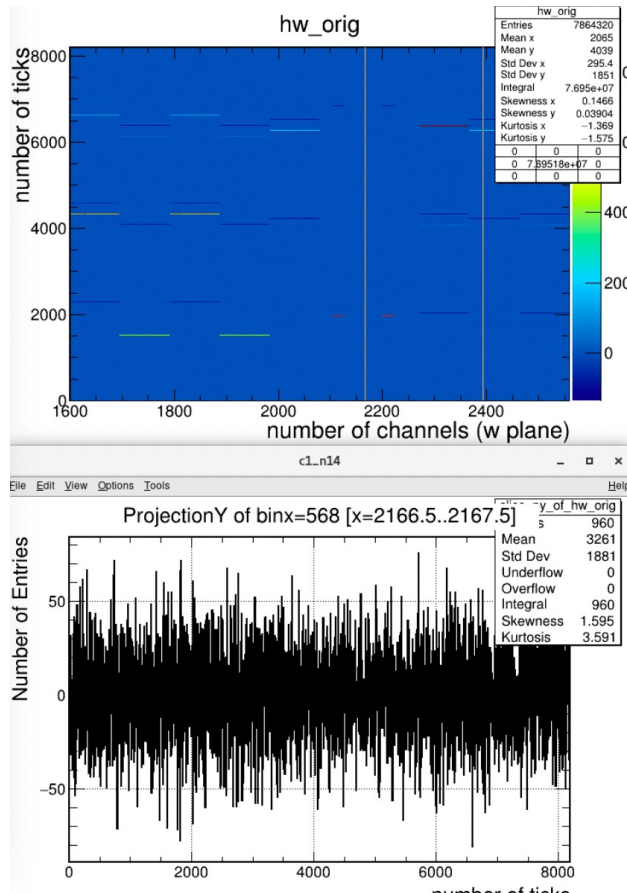
After NF



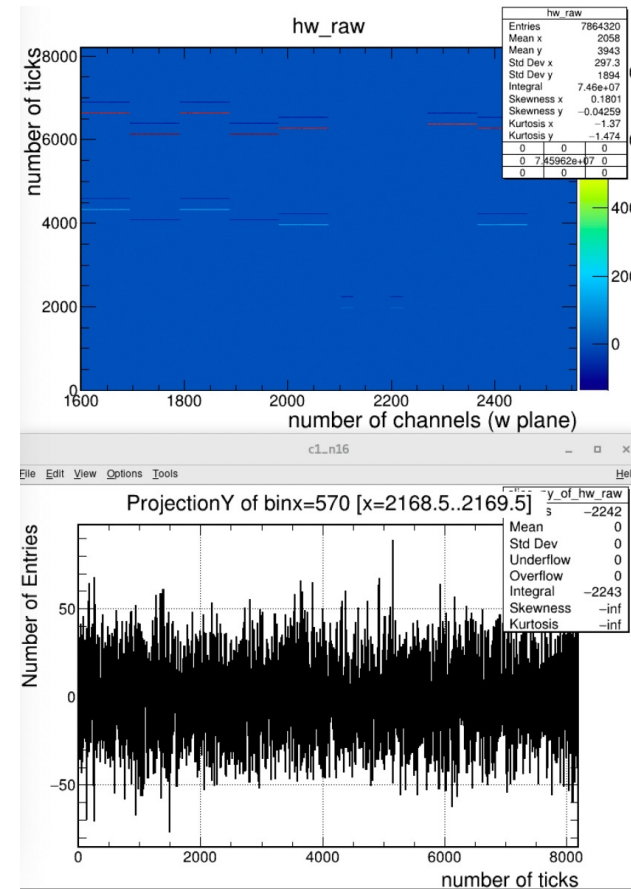
➤ The diagrams are in v plane

Run 19337 : Effect of Noise Filter

Before NF



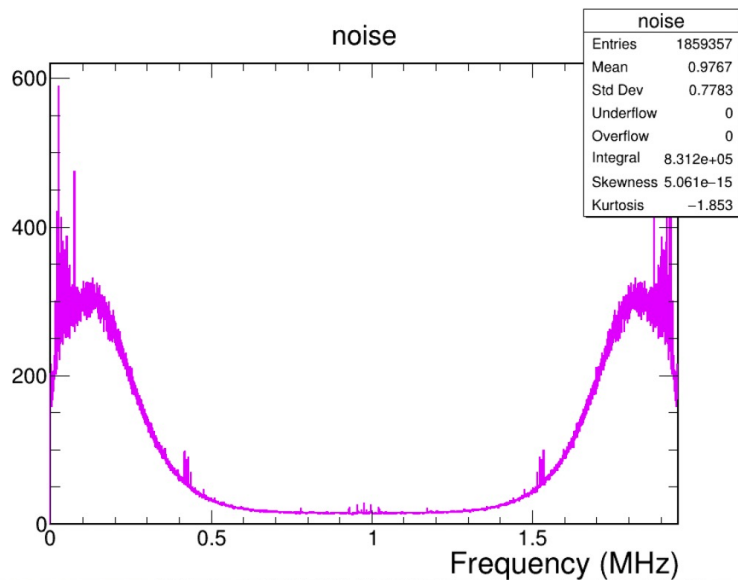
After NF



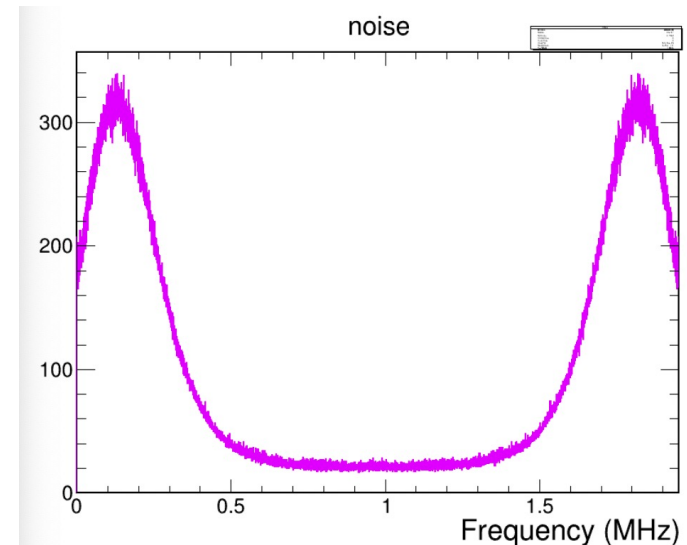
➤ The diagrams are in w plane

Run 19337 : Noise in Frequency Spectrum (Effect of Noise Filter)

Before NF



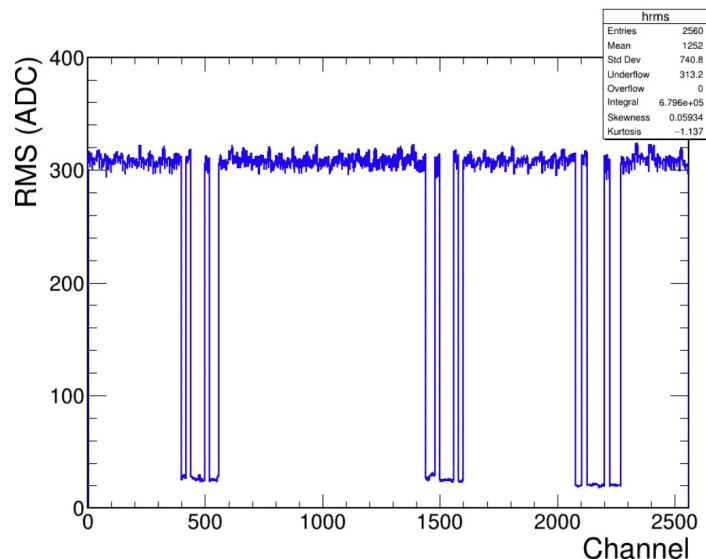
After NF



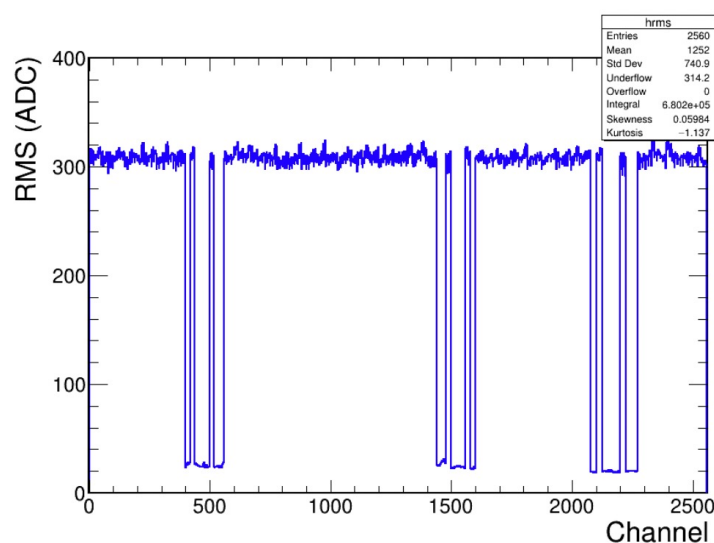
- The diagrams are in frequency spectrum
 - Frequency = a little less than 2 MHz
 - T = approximately 500 ns

Run 19337 : ADC RMS plot (Effect of Noise Filter)

Before NF



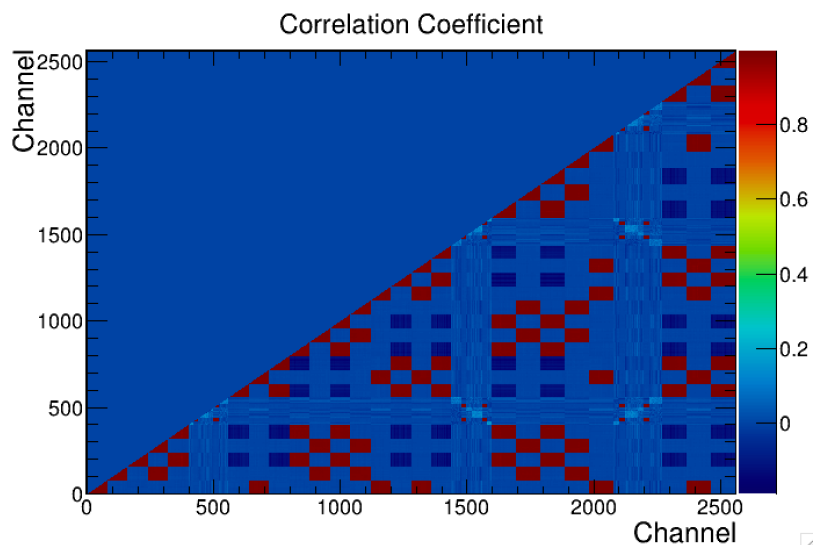
After NF



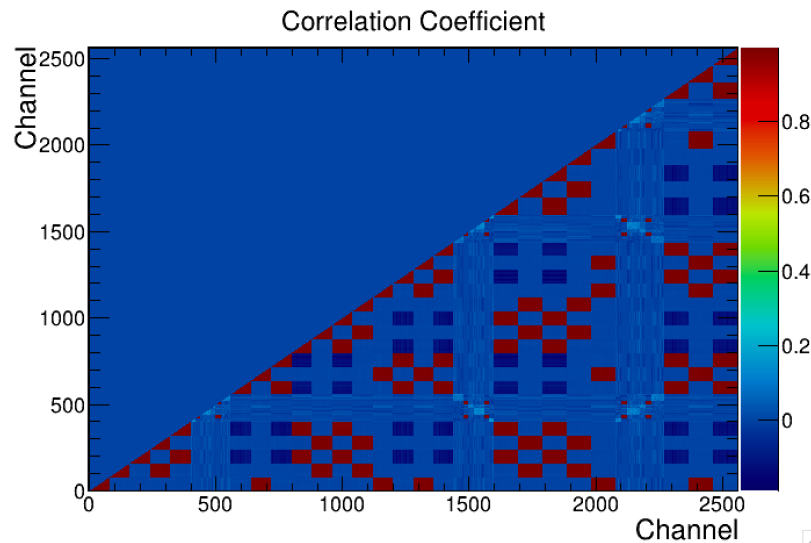
- Calculate ADC RMS before and after NF
- u, v, and w planes combined
- NF should reduce RMS in most cases
- Several dead channels cause RMS (ADC) values drop

Run 19337: Linear Correlation Between Channels

Before NF



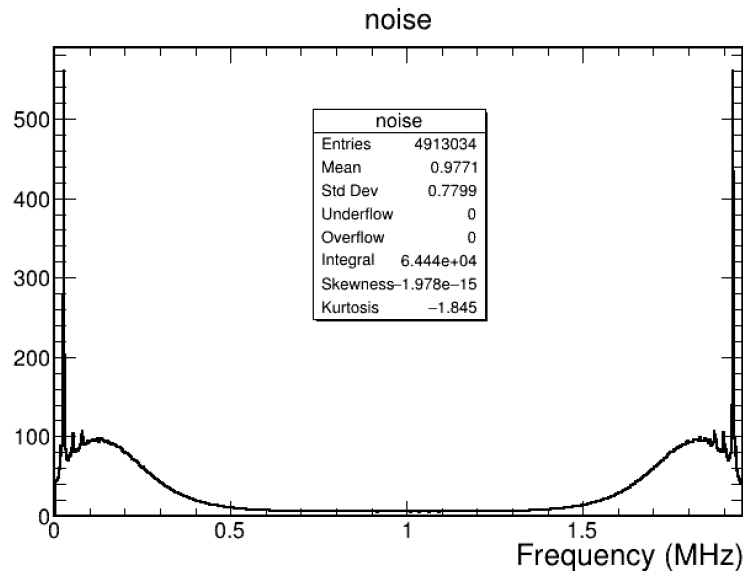
After NF



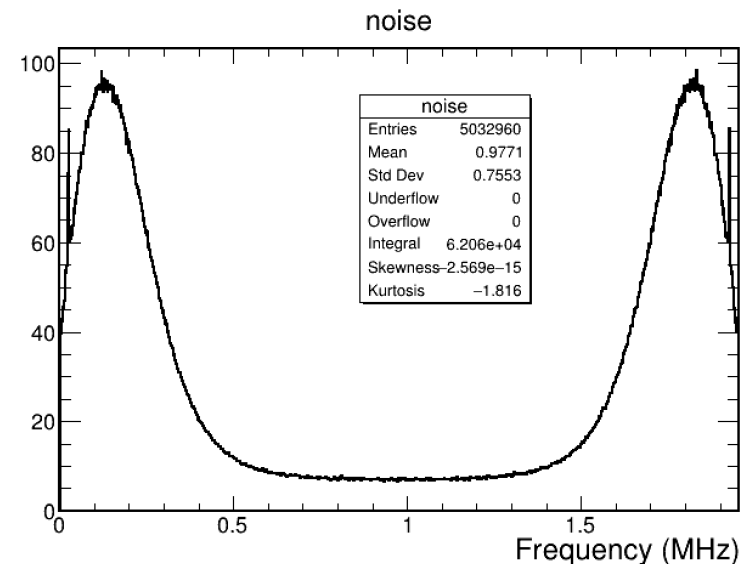
- Calculate pairwise linear correlation for all channels
- Noise removal reduces linear correlation

Run 16751 : Noise in Frequency Spectrum (Effect of Noise Filter)

Before NF



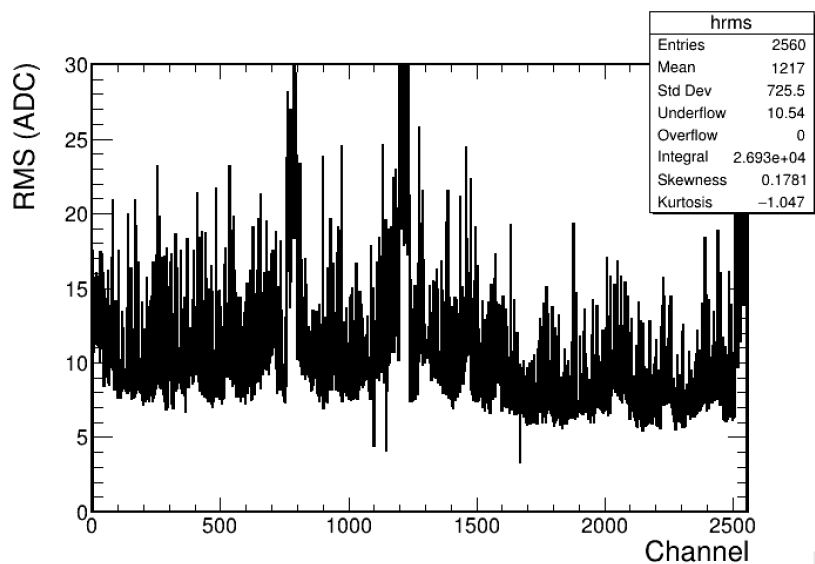
After NF



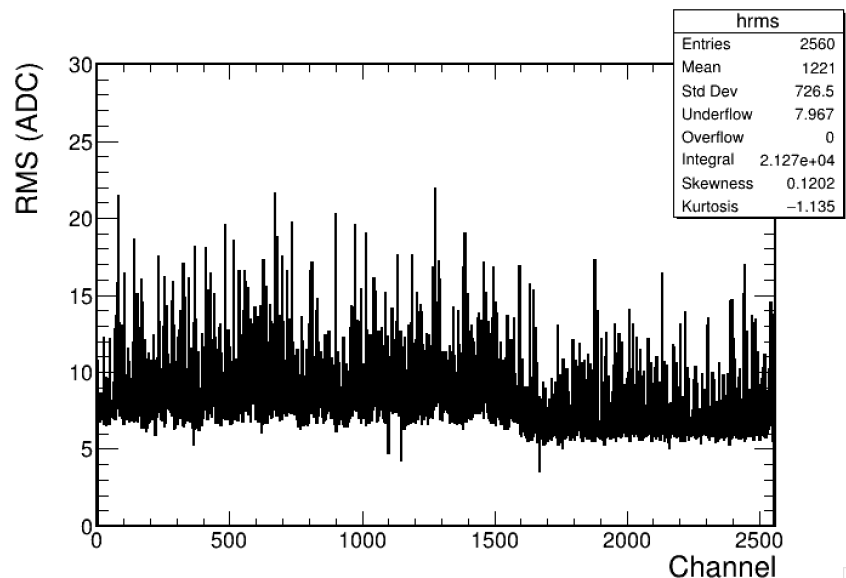
- The diagrams are in frequency spectrum
 - Frequency = a little less than 2 MHz
 - T = approximately 500 ns

Run 16751 : ADC RMS plot (Effect of Noise Filter)

Before NF



After NF



- Calculate ADC RMS before and after NF
- u, v, and w planes combined
- NF should reduce RMS in most cases
- Several dead channels cause RMS (ADC) values drop

Summary and Outlook

➤ Summary

- Code for PDHD noise filter is developed
- Consider good noise run to study from old runs ??
- Or wait until April 17th for DUNE DAQ, DUNE data format and DUNE file format to stop evolving

➤ Outlook

- Conduct analysis for other HD runs
- Investigate effect of WireCell Noise Filter (NF)
- Look for new / unknown types of noise
- Find dead / noisy channels
- Understand coherence introduced by NF
- Investigate correlation / coherence across wire planes

Thank You !