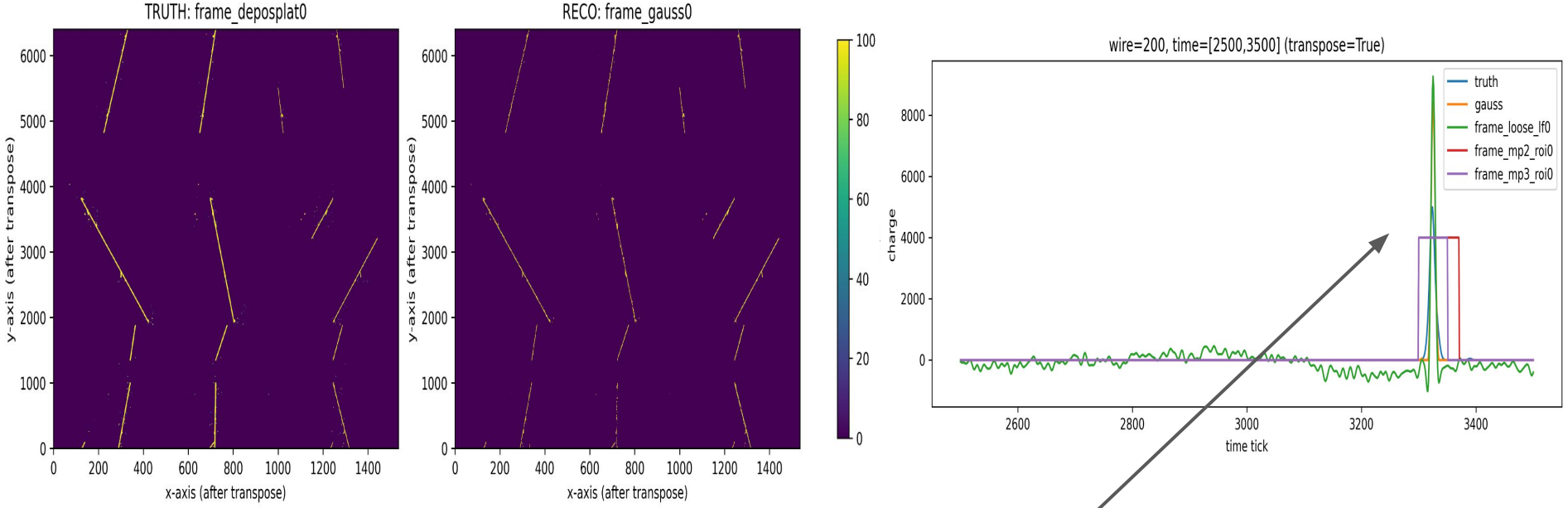
Status report on

DNNROI sigproc

Hokyeong Nam
Chung-Ang University

Brief Recap



PD-VD training sample validation

- Mapping issues & time offset issues → Resolved
- There is discrepancy between truth and reco in 1D waveform

Software Filter and Truth Smearing

- PD-HD and PD-VD are using different software filter (Wiener tight)
- For the truth smearing, they are using the same values
- PD-HD HfFilter: Optimized
- PD-VD HfFilter: parameters from PD-SP

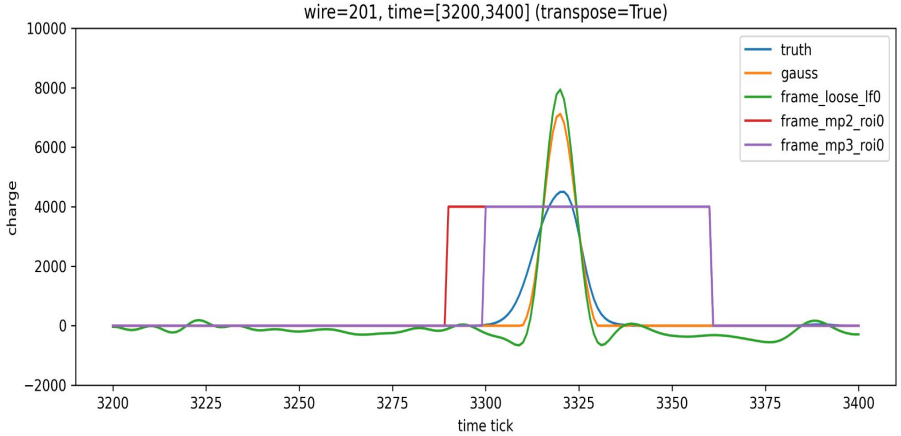
❑ PD-HD HfFilter

❑ ~~PD-VD HfFilter~~

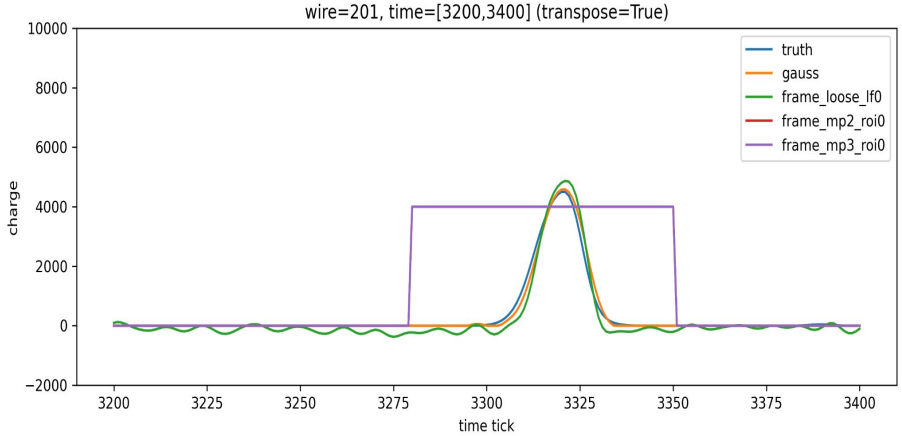
❑ PD-SP HfFilter

❑ Smearing

Software Filter Study



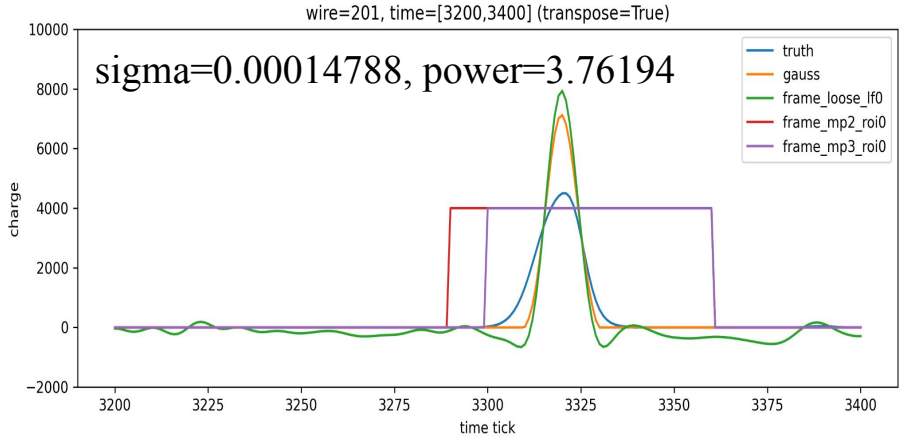
❑ PD-VD sim w/ PD-SP HfFilter



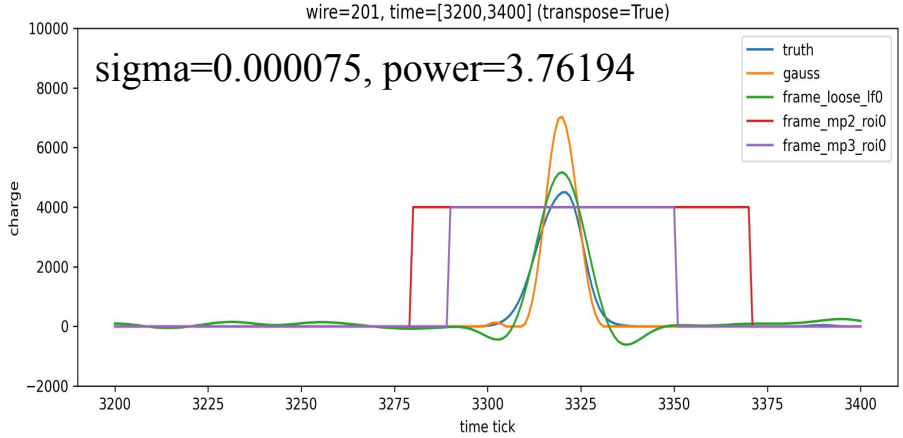
❑ PD-VD sim w/ PD-HD HfFilter

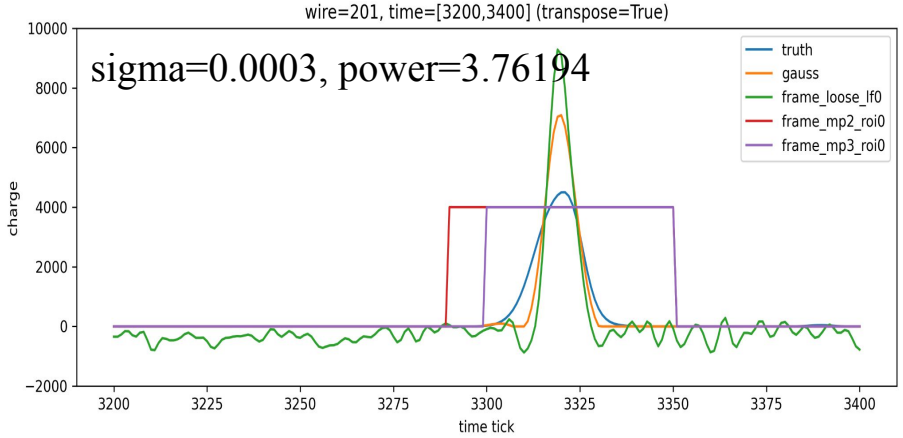
- If we apply PD-HD HfFilter to the PD-VD sim, now the waveform is well matched
- Items to study:
  - how the sigma and power affects the loose_lf result?
  - how to get optimized sigma and power parameters for the PD-VD
  - Can we keep the truth smearing parameters used for PD-HD

Software Filter Study

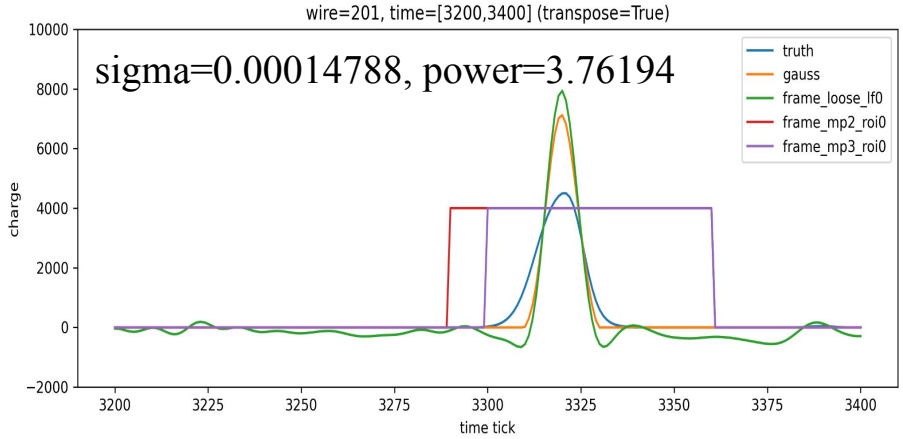


- Parameters explained:
  - Top left: PD-SP filter
  - Bottom left: ~0.5*sigma
  - bottom right: ~2*sigma
- lower sigma: peak ↓, width ↑, smoothness -
- higher sigma: peak ↑, width ↓, smoothness ↓

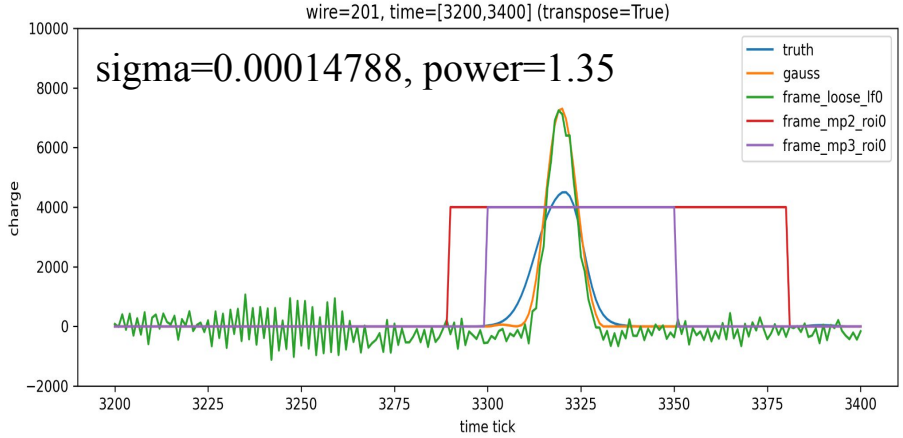


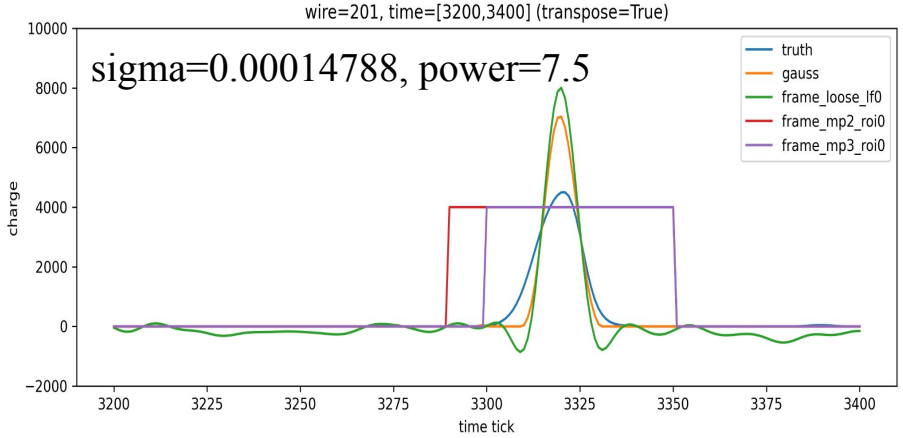


Software Filter Study



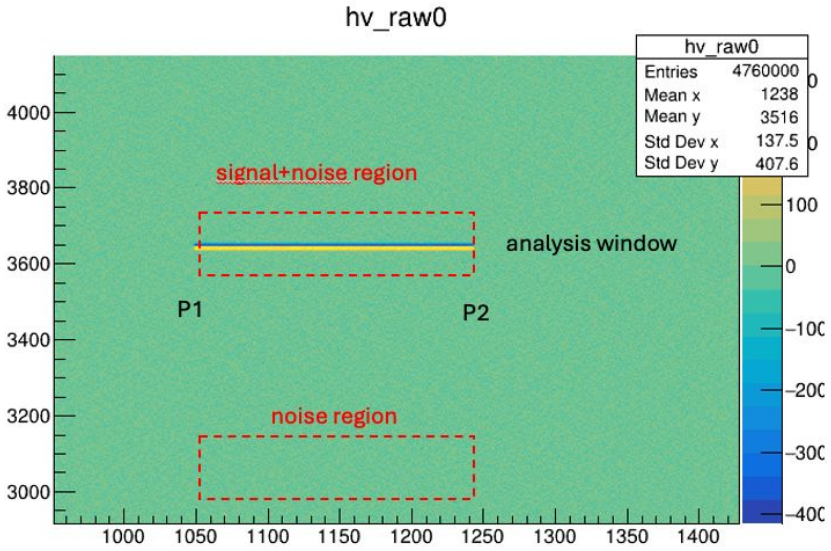
- Parameters explained:
  - Top left: PD-SP filter
  - Bottom left: ~0.5*power
  - bottom right: ~2*power
- lower sigma: peak ↓, width -, smoothness ↓
- higher sigma: peak -, width -, smoothness -



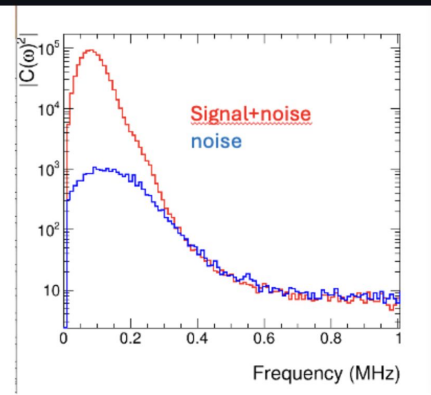


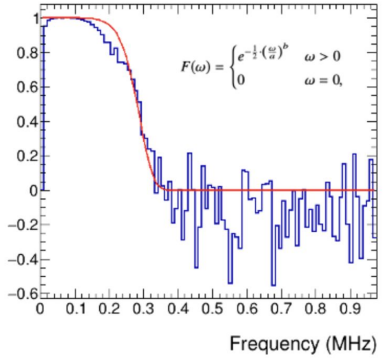
Software Filter Optimization

- There is a tool to get the optimized sigma and power for the HfFilter
  - https://github.com/WireCell/wire-cell-toolkit/blob/feature/xn_WF_fitter_script/root/test/calcFilter.C
  - https://github.com/WireCell/wire-cell-toolkit/blob/feature/xn_WF_fitter_script/root/docs/calcFilter.org
- Isochronous track sample is needed



1. **c1**: Power spectral density comparison
   - Red: Signal + noise power density
   - Blue: Noise-only power density
2. **c2**: Wiener filter response
   - Normalized to maximum value of 1
   - Fitted with function: `exp(-0.5*(x/p0)^p1)`
   - Fit parameters printed to console



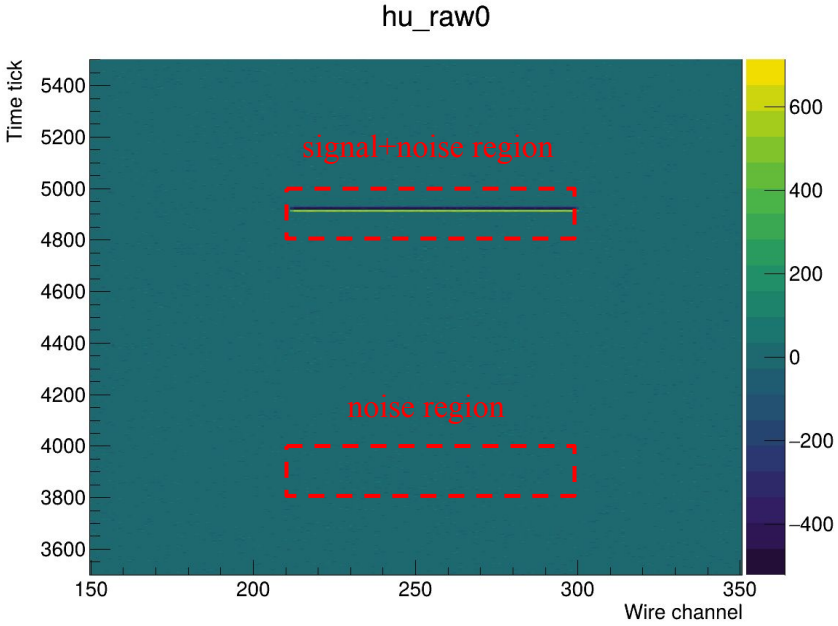


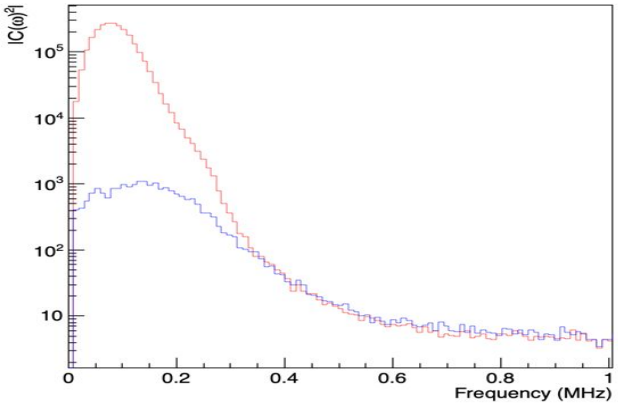

```
EXT PARAMETER                                   STEP         FIRST
NO.   NAME      VALUE            ERROR          SIZE      DERIVATIVE
 1  p0           2.69885e-01   3.74802e-02   2.98269e-05   3.55623e-02
 2  p1           8.34938e+00   6.86118e+00   5.42952e-03  -2.51530e-04
```

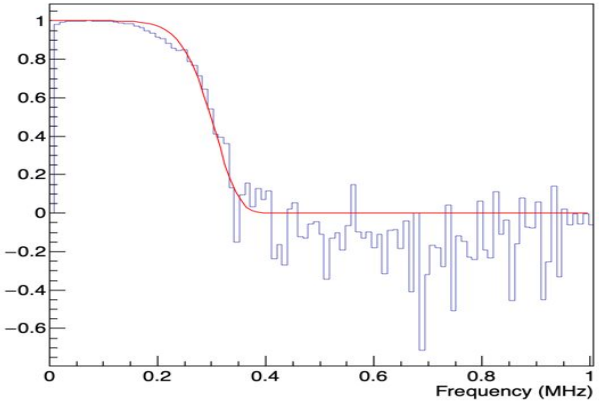
Software Filter Optimization

- Fitted values for PDVD bottom electronics (CRU0)
- u plane; p0 (sigma), p1 (power)

```
p0                      =      0.286735   +/-   0.0417082
p1                      =      8.07074    +/-   9.90182
```



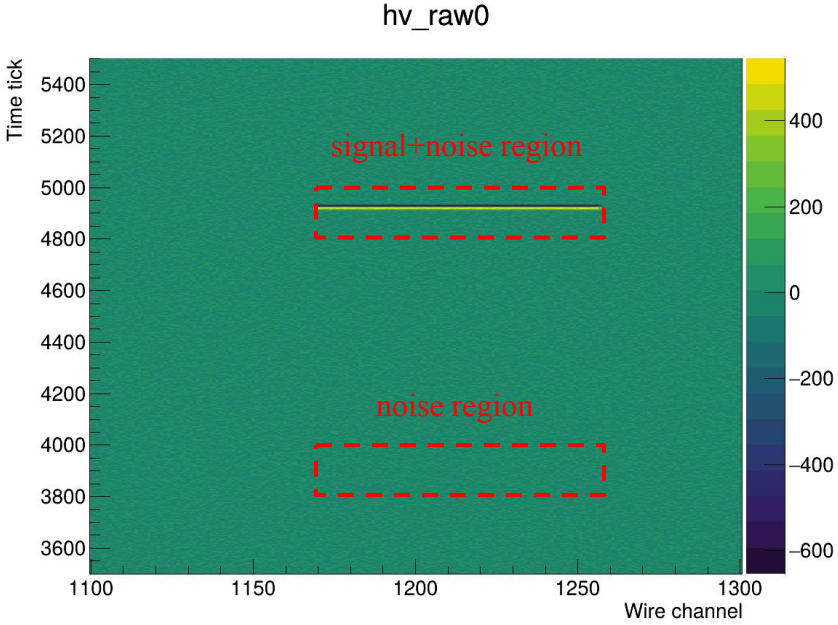


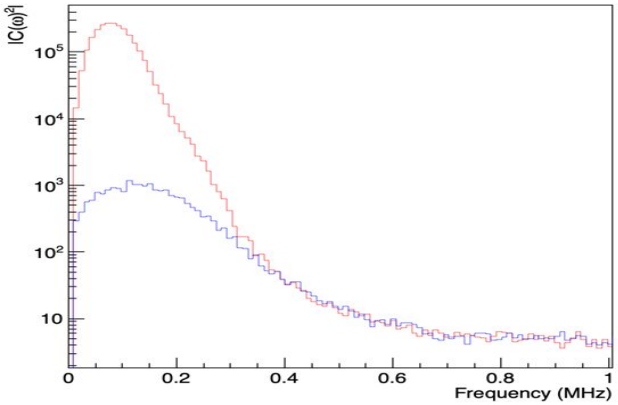


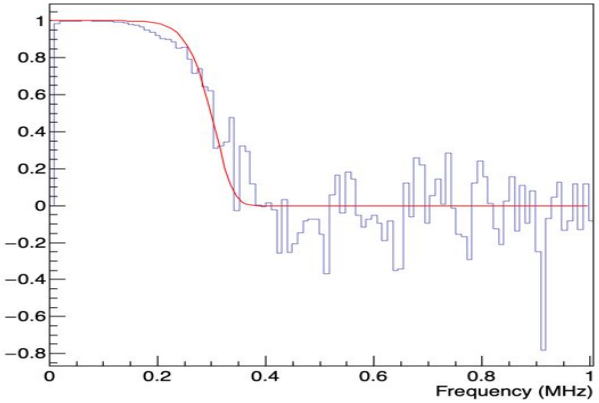
Software Filter Optimization

- Fitted values for PDVD bottom electronics (CRU0)
- v plane; p0 (sigma), p1 (power)

```
p0                      =      0.289041   +/-   0.0341532
p1                      =       9.94129   +/-   9.32916
```



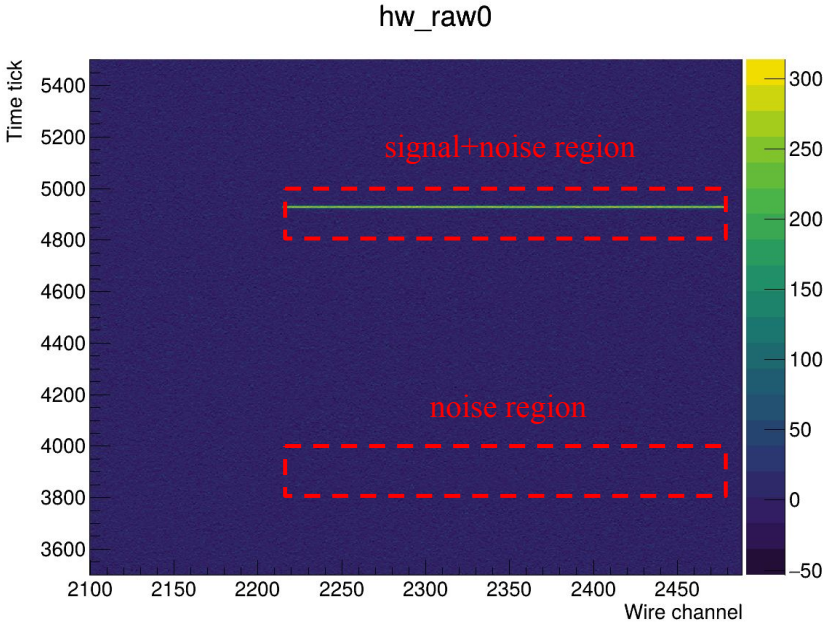


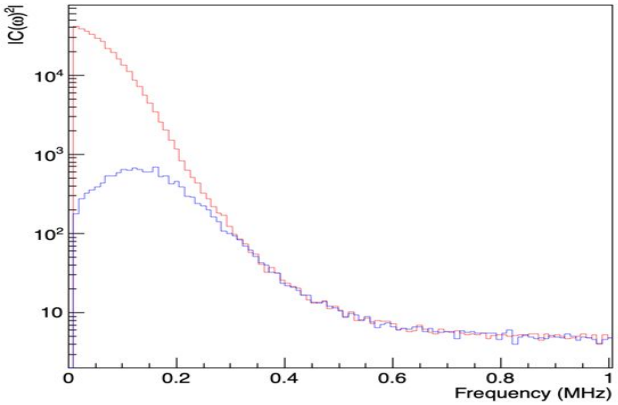


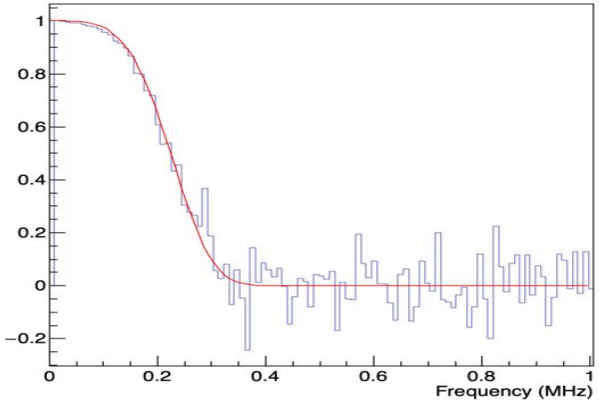
Software Filter Optimization

- Fitted values for PDVD bottom electronics (CRU0)
- w plane; p0 (sigma), p1 (power)

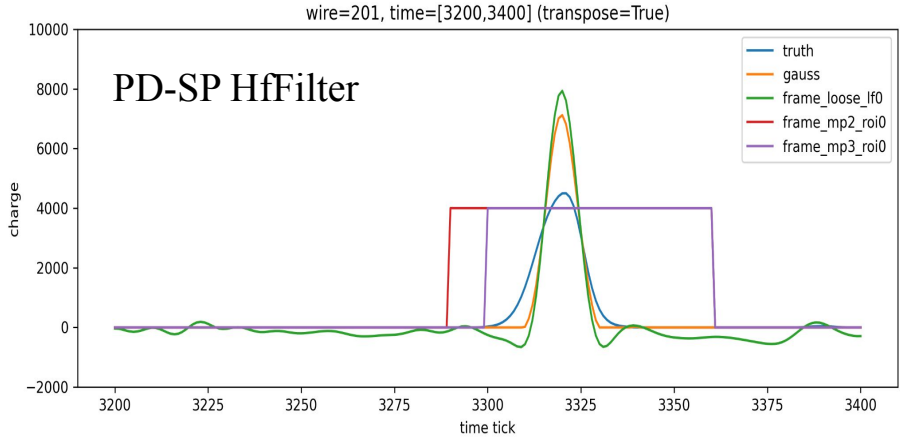
```
p0                        =      0.207127
p1                        =      4.13956
```

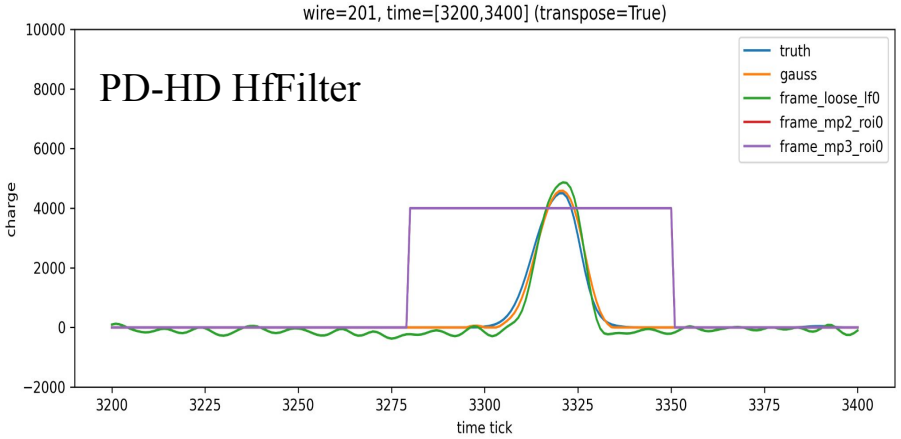


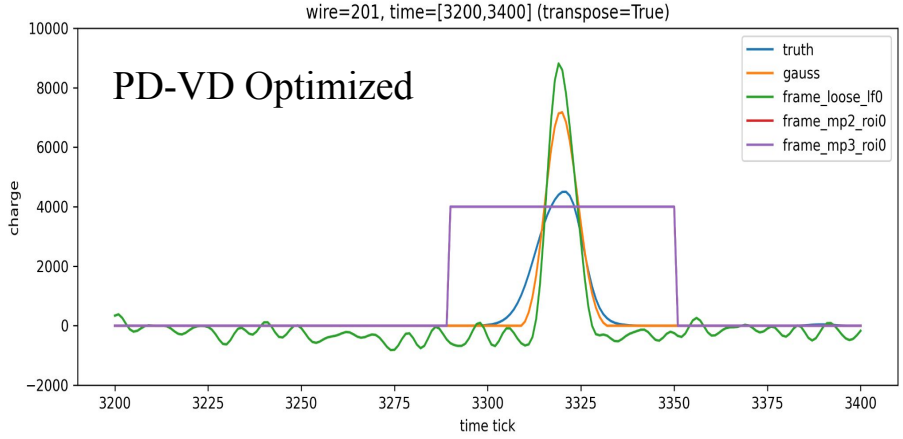




Software Filter Comparison



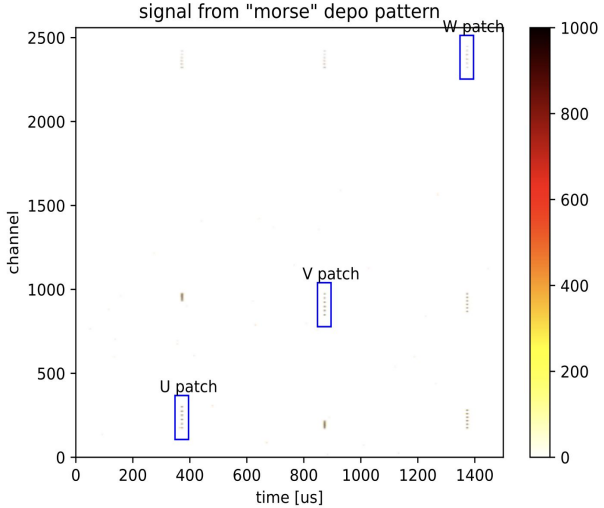


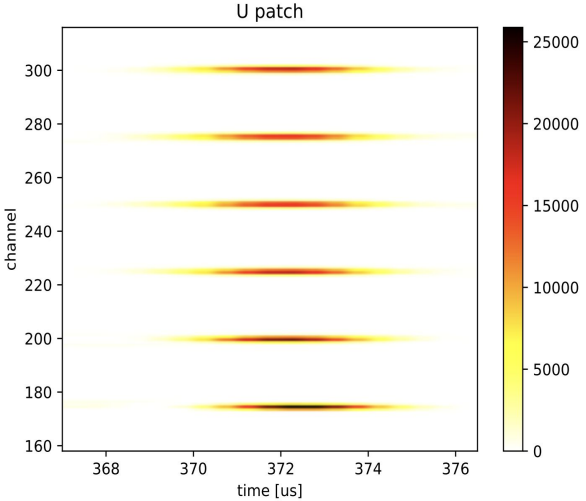


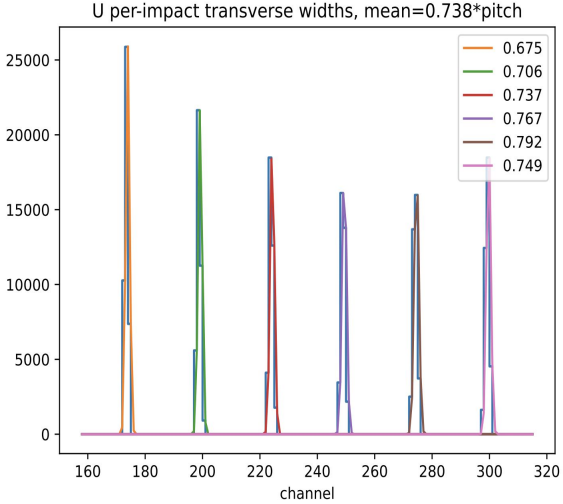
- Optimized filter gives tighter and higher loose_lf
- Also noisier than PD-SP and PD-VD on u plane
- If we keep the truth smearing value as current, the discrepancy still exists

Truth Smearing - Approach 1

- Morse test tool provided by Brett
  - https://github.com/WireCell/wire-cell-toolkit/blob/master/test/scripts/bats-debug
  - https://github.com/WireCell/wire-cell-toolkit/blob/master/test/test/test-morse-pdsp.bats



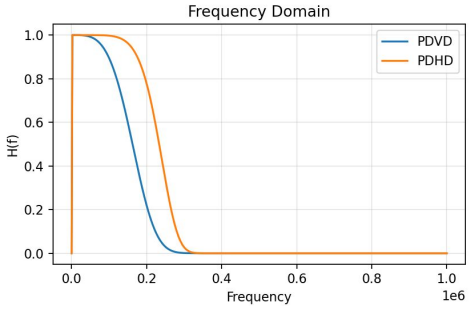


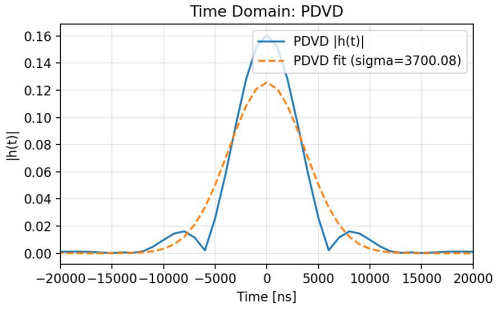


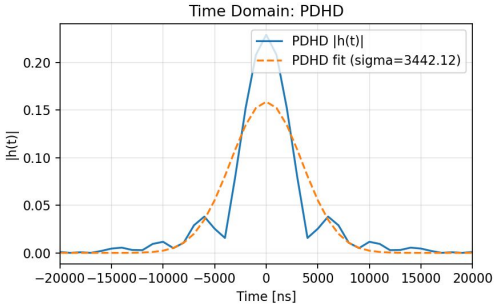
- The tool is not fully ready for PD-VD
- Would be precise, but need to understand how to use

Truth Smearing - Approach 2

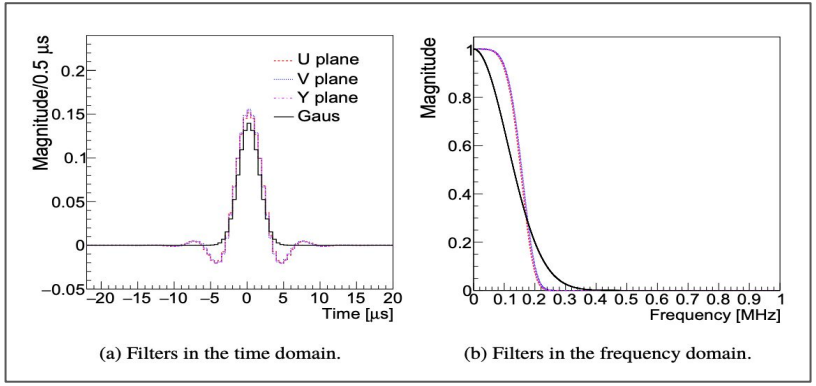
- Frequency domain Wiener filter → IFFT → Time domain signals → Gaussian fit







Gaussian fit sigmas:
PDVD: 3700.08
PDHD: 3442.12



https://arxiv.org/abs/1802.08709

- The shape looks similar with the one from MicroBooNE paper
- Relatively easy, but still need to understand how to get right numbers
- Support from Hanyu Wei may be needed

HDF5 Compression Test

- Since storage on the WC cluster is limited, the PD-VD sample should be
  - storage-efficient → compression
  - stored on another server (e.g., scratch directory on the dunegpvm)
- The gzip compression option is available in the WC configuration file
- The log shows “gzip 0” is applied instead of the gzip value I set
- WC version: 0.30.3

```
local hio_truth = [g.pnode({
      type: 'HDF5FrameTap',
      name: 'hio_truth%d' % n,
      data: {
        anode: wc.tn(tools.anodes[n]),
        trace_tags: ['deposplat%d'%n],
        filename: "g4-tru-%d.h5" % n,
        chunk: [0, 0], // ncol, nrow
        gzip: 2,
        high_throughput: true,
      },
    }, nin=1, nout=1),
    for n in std.range(0, std.length(tools.anodes) - 1)
    ];
```

```
[09:09:56.230] D [  hio   ] <HDF5FrameTap:hio_sp0> gzip:0 chunks:[0,0]
[09:09:56.324] D [  hio   ] <HDF5FrameTap:hio_sp0> saved /100/frame_loose_lf0 with 1536 channels 6400 ticks @t=-0.24987847929395787 ms qtot=106350890
[09:09:56.324] W [  hio   ] <HDF5FrameTap:hio_sp0> no traces for tag: "tight_lf0"
[09:09:56.389] D [  hio   ] <HDF5FrameTap:hio_sp0> gzip:0 chunks:[0,0]
[09:09:56.478] D [  hio   ] <HDF5FrameTap:hio_sp0> saved /100/frame_cleanup_roi0 with 1536 channels 6400 ticks @t=-0.24987847929395787 ms qtot=144788380
[09:09:56.541] D [  hio   ] <HDF5FrameTap:hio_sp0> gzip:0 chunks:[0,0]
[09:09:56.646] D [  hio   ] <HDF5FrameTap:hio_sp0> saved /100/frame_break_roi_1st0 with 1536 channels 6400 ticks @t=-0.24987847929395787 ms qtot=138806370
[09:09:56.710] D [  hio   ] <HDF5FrameTap:hio_sp0> gzip:0 chunks:[0,0]
[09:09:56.805] D [  hio   ] <HDF5FrameTap:hio_sp0> saved /100/frame_break_roi_2nd0 with 1536 channels 6400 ticks @t=-0.24987847929395787 ms qtot=138496830
[09:09:56.868] D [  hio   ] <HDF5FrameTap:hio_sp0> gzip:0 chunks:[0,0]
[09:09:56.968] D [  hio   ] <HDF5FrameTap:hio_sp0> saved /100/frame_shrink_roi0 with 1536 channels 6400 ticks @t=-0.24987847929395787 ms qtot=143791580
[09:09:57.031] D [  hio   ] <HDF5FrameTap:hio_sp0> gzip:0 chunks:[0,0]
[09:09:57.129] D [  hio   ] <HDF5FrameTap:hio_sp0> saved /100/frame_extend_roi0 with 1536 channels 6400 ticks @t=-0.24987847929395787 ms qtot=805700
[09:09:57.193] D [  hio   ] <HDF5FrameTap:hio_sp0> gzip:0 chunks:[0,0]
[09:09:57.295] D [  hio   ] <HDF5FrameTap:hio_sp0> saved /100/frame_mp3_roi0 with 1536 channels 6400 ticks @t=-0.24987847929395787 ms qtot=232468000
[09:09:57.365] D [  hio   ] <HDF5FrameTap:hio_sp0> gzip:0 chunks:[0,0]
[09:09:57.465] D [  hio   ] <HDF5FrameTap:hio_sp0> saved /100/frame_mp2_roi0 with 1536 channels 6400 ticks @t=-0.24987847929395787 ms qtot=525224000
[09:09:57.530] D [  hio   ] <HDF5FrameTap:hio_sp0> gzip:0 chunks:[0,0]
[09:09:57.629] D [  hio   ] <HDF5FrameTap:hio_sp0> saved /100/frame_decon_charge0 with 1536 channels 6400 ticks @t=-0.24987847929395787 ms qtot=42769116
[09:09:57.693] D [  hio   ] <HDF5FrameTap:hio_sp0> gzip:0 chunks:[0,0]
[09:09:57.801] D [  hio   ] <HDF5FrameTap:hio_sp0> saved /100/frame_gauss0 with 1536 channels 6400 ticks @t=-0.24987847929395787 ms qtot=126555860
```

HDF5 Compression Test

- Inspected with Chao's script (inspect_h5.py)
- Confirmed the output HDF5 is not compressed

```
Inspecting: gzip6_keepall/g4-tru-0.h5

=== File Summary ===
Total OS File Size:    37.51 MB
Total Datasets:        3
Total Logical Size:    37.51 MB
Total Data on Disk:    37.51 MB
File Metadata/Overhead: 4.14 KB (0.0%)
Overall Compression:   1.00x

=== Analysis by Dataset Name (Grouped) ===
Name                  | Cnt | Type     | Shape           | Logical   | Storage   | Ratio | Comp
-----------------------------------------------------------------------------------------------
frame_deposplat0      | 1   | float32  | (1536, 6400)    | 37.50 MB  | 37.50 MB  | 1.0   | None
channels_deposplat0   | 1   | int32    | (1536,)         | 6.00 KB   | 6.00 KB   | 1.0   | None
tickinfo_deposplat0   | 1   | float64  | (3,)            | 24 B      | 24 B      | 1.0   | None
-----------------------------------------------------------------------------------------------
TOTAL                 | 3   | -        | -               | 37.51 MB  | 37.51 MB  | 1.0   | -

=== Compression Analysis ===
Poorly compressed (Ratio < 1.05):
  - frame_deposplat0: 1.00x (37.50 MB)
  - channels_deposplat0: 1.00x (6.00 KB)
```

```
Inspecting: gzip6_keepall/g4-rec-0.h5

=== File Summary ===
Total OS File Size:    389.33 MB
Total Datasets:        33
Total Logical Size:    389.32 MB
Total Data on Disk:    389.32 MB
File Metadata/Overhead: 14.14 KB (0.0%)
Overall Compression:   1.00x

=== Analysis by Dataset Name (Grouped) ===
Name                  | Cnt | Type     | Shape           | Logical   | Storage   | Ratio | Comp
-----------------------------------------------------------------------------------------------
frame_break_roi_1st0  | 1   | float32  | (1536, 6400)    | 37.50 MB  | 37.50 MB  | 1.0   | None
frame_break_roi_2nd0  | 1   | float32  | (1536, 6400)    | 37.50 MB  | 37.50 MB  | 1.0   | None
frame_cleanup_roi0    | 1   | float32  | (1536, 6400)    | 37.50 MB  | 37.50 MB  | 1.0   | None
frame_decon_charge0   | 1   | float32  | (1536, 6400)    | 37.50 MB  | 37.50 MB  | 1.0   | None
frame_extend_roi0     | 1   | float32  | (1536, 6400)    | 37.50 MB  | 37.50 MB  | 1.0   | None
frame_gauss0          | 1   | float32  | (1536, 6400)    | 37.50 MB  | 37.50 MB  | 1.0   | None
frame_loose_lf0       | 1   | float32  | (1536, 6400)    | 37.50 MB  | 37.50 MB  | 1.0   | None
frame_mp2_roi0        | 1   | float32  | (1536, 6400)    | 37.50 MB  | 37.50 MB  | 1.0   | None
frame_mp3_roi0        | 1   | float32  | (1536, 6400)    | 37.50 MB  | 37.50 MB  | 1.0   | None
frame_shrink_roi0     | 1   | float32  | (1536, 6400)    | 37.50 MB  | 37.50 MB  | 1.0   | None
frame_dnnsp0          | 1   | float32  | (584, 6400)     | 14.26 MB  | 14.26 MB  | 1.0   | None
channels_break_roi_1st | 1   | int32    | (1536,)         | 6.00 KB   | 6.00 KB   | 1.0   | None
channels_break_roi_2nd | 1   | int32    | (1536,)         | 6.00 KB   | 6.00 KB   | 1.0   | None
channels_cleanup_roi0 | 1   | int32    | (1536,)         | 6.00 KB   | 6.00 KB   | 1.0   | None
channels_decon_charge0 | 1   | int32    | (1536,)         | 6.00 KB   | 6.00 KB   | 1.0   | None
channels_extend_roi0  | 1   | int32    | (1536,)         | 6.00 KB   | 6.00 KB   | 1.0   | None
channels_gauss0       | 1   | int32    | (1536,)         | 6.00 KB   | 6.00 KB   | 1.0   | None
channels_loose_lf0    | 1   | int32    | (1536,)         | 6.00 KB   | 6.00 KB   | 1.0   | None
channels_mp2_roi0     | 1   | int32    | (1536,)         | 6.00 KB   | 6.00 KB   | 1.0   | None
channels_mp3_roi0     | 1   | int32    | (1536,)         | 6.00 KB   | 6.00 KB   | 1.0   | None
channels_shrink_roi0  | 1   | int32    | (1536,)         | 6.00 KB   | 6.00 KB   | 1.0   | None
channels_dnnsp0       | 1   | int32    | (584,)          | 2.28 KB   | 2.28 KB   | 1.0   | None
tickinfo_break_roi_1st | 1   | float64  | (3,)            | 24 B      | 24 B      | 1.0   | None
tickinfo_break_roi_2nd | 1   | float64  | (3,)            | 24 B      | 24 B      | 1.0   | None
tickinfo_cleanup_roi0 | 1   | float64  | (3,)            | 24 B      | 24 B      | 1.0   | None
tickinfo_decon_charge0 | 1   | float64  | (3,)            | 24 B      | 24 B      | 1.0   | None
tickinfo_dnnsp0       | 1   | float64  | (3,)            | 24 B      | 24 B      | 1.0   | None
tickinfo_extend_roi0  | 1   | float64  | (3,)            | 24 B      | 24 B      | 1.0   | None
tickinfo_gauss0       | 1   | float64  | (3,)            | 24 B      | 24 B      | 1.0   | None
tickinfo_loose_lf0    | 1   | float64  | (3,)            | 24 B      | 24 B      | 1.0   | None
tickinfo_mp2_roi0     | 1   | float64  | (3,)            | 24 B      | 24 B      | 1.0   | None
tickinfo_mp3_roi0     | 1   | float64  | (3,)            | 24 B      | 24 B      | 1.0   | None
tickinfo_shrink_roi0  | 1   | float64  | (3,)            | 24 B      | 24 B      | 1.0   | None
-----------------------------------------------------------------------------------------------
TOTAL                 | 33  | -        | -               | 389.32 MB | 389.32 MB | 1.0   | -
```

HDF5 Dataset Pruning

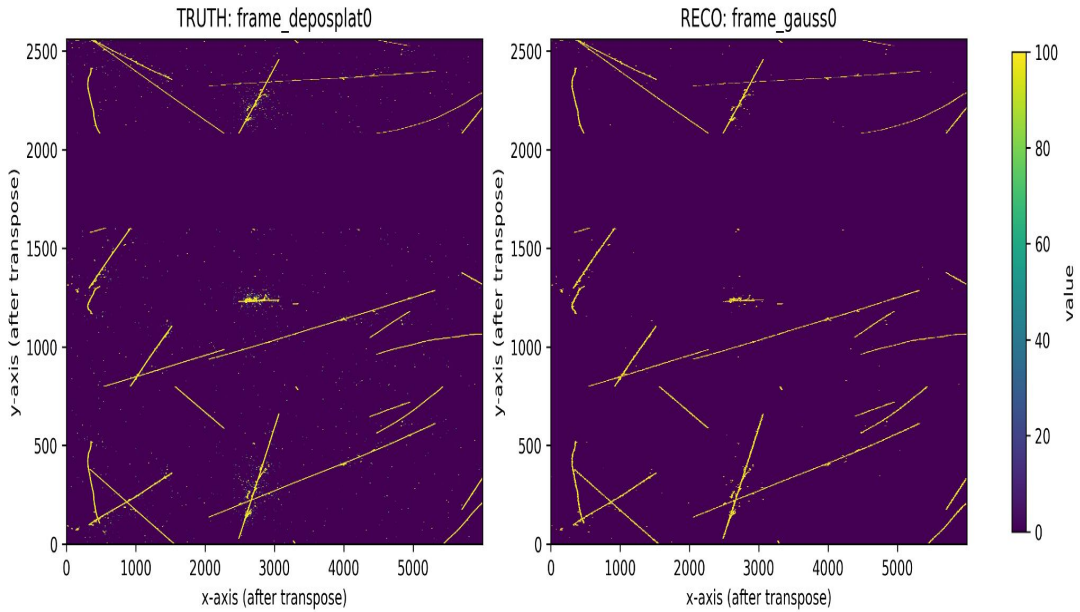
```
Inspecting: gzip2_keepall/g4-rec-0.h5

=== File Summary ===
Total OS File Size:    389.33 MB
Total Datasets:        33
Total Logical Size:    389.32 MB
Total Data on Disk:    389.32 MB
File Metadata/Overhead: 14.14 KB (0.0%)
Overall Compression:   1.00x

=== Analysis by Dataset Name (Grouped) ===
Name                   | Cnt | Type      | Shape           | Logical    | Storage    | Ratio | Comp
-------------------------------------------------------------------------------------------------
frame_break_roi_1st0   | 1   | float32   | (1536, 6400)    | 37.50 MB   | 37.50 MB   | 1.0   | None
frame_break_roi_2nd0   | 1   | float32   | (1536, 6400)    | 37.50 MB   | 37.50 MB   | 1.0   | None
frame_cleanup_roi0     | 1   | float32   | (1536, 6400)    | 37.50 MB   | 37.50 MB   | 1.0   | None
frame_decon_charge0    | 1   | float32   | (1536, 6400)    | 37.50 MB   | 37.50 MB   | 1.0   | None
frame_extend_roi0      | 1   | float32   | (1536, 6400)    | 37.50 MB   | 37.50 MB   | 1.0   | None
frame_gauss0           | 1   | float32   | (1536, 6400)    | 37.50 MB   | 37.50 MB   | 1.0   | None
frame_loose_lf0        | 1   | float32   | (1536, 6400)    | 37.50 MB   | 37.50 MB   | 1.0   | None
frame_mp2_roi0         | 1   | float32   | (1536, 6400)    | 37.50 MB   | 37.50 MB   | 1.0   | None
frame_mp3_roi0         | 1   | float32   | (1536, 6400)    | 37.50 MB   | 37.50 MB   | 1.0   | None
frame_shrink_roi0      | 1   | float32   | (1536, 6400)    | 37.50 MB   | 37.50 MB   | 1.0   | None
frame_dnnsp0           | 1   | float32   | (584, 6400)     | 14.26 MB   | 14.26 MB   | 1.0   | None
channels_break_roi_1st | 1   | int32     | (1536,)         | 6.00 KB    | 6.00 KB    | 1.0   | None
channels_break_roi_2nd | 1   | int32     | (1536,)         | 6.00 KB    | 6.00 KB    | 1.0   | None
channels_cleanup_roi0  | 1   | int32     | (1536,)         | 6.00 KB    | 6.00 KB    | 1.0   | None
channels_decon_charge0 | 1   | int32     | (1536,)         | 6.00 KB    | 6.00 KB    | 1.0   | None
channels_extend_roi0   | 1   | int32     | (1536,)         | 6.00 KB    | 6.00 KB    | 1.0   | None
channels_gauss0        | 1   | int32     | (1536,)         | 6.00 KB    | 6.00 KB    | 1.0   | None
channels_loose_lf0     | 1   | int32     | (1536,)         | 6.00 KB    | 6.00 KB    | 1.0   | None
channels_mp2_roi0      | 1   | int32     | (1536,)         | 6.00 KB    | 6.00 KB    | 1.0   | None
channels_mp3_roi0      | 1   | int32     | (1536,)         | 6.00 KB    | 6.00 KB    | 1.0   | None
channels_shrink_roi0   | 1   | int32     | (1536,)         | 6.00 KB    | 6.00 KB    | 1.0   | None
channels_dnnsp0        | 1   | int32     | (584,)          | 2.28 KB    | 2.28 KB    | 1.0   | None
tickinfo_break_roi_1st | 1   | float64   | (3,)            | 24 B       | 24 B       | 1.0   | None
tickinfo_break_roi_2nd | 1   | float64   | (3,)            | 24 B       | 24 B       | 1.0   | None
tickinfo_cleanup_roi0  | 1   | float64   | (3,)            | 24 B       | 24 B       | 1.0   | None
tickinfo_decon_charge0 | 1   | float64   | (3,)            | 24 B       | 24 B       | 1.0   | None
tickinfo_dnnsp0        | 1   | float64   | (3,)            | 24 B       | 24 B       | 1.0   | None
tickinfo_extend_roi0   | 1   | float64   | (3,)            | 24 B       | 24 B       | 1.0   | None
tickinfo_gauss0        | 1   | float64   | (3,)            | 24 B       | 24 B       | 1.0   | None
tickinfo_loose_lf0     | 1   | float64   | (3,)            | 24 B       | 24 B       | 1.0   | None
tickinfo_mp2_roi0      | 1   | float64   | (3,)            | 24 B       | 24 B       | 1.0   | None
tickinfo_mp3_roi0      | 1   | float64   | (3,)            | 24 B       | 24 B       | 1.0   | None
tickinfo_shrink_roi0   | 1   | float64   | (3,)            | 24 B       | 24 B       | 1.0   | None
-------------------------------------------------------------------------------------------------
TOTAL                  | 33  | -         | -               | 389.32 MB  | 389.32 MB  | 1.0   | -
```

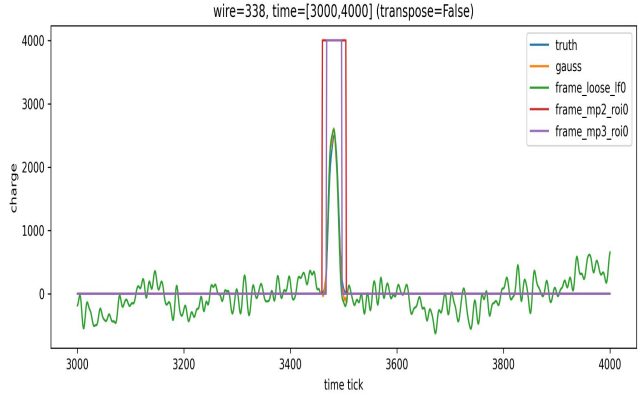
```
Inspecting: gzip2_keepless/g4-rec-0.h5

=== File Summary ===
Total OS File Size:    150.03 MB
Total Datasets:        12
Total Logical Size:    150.02 MB
Total Data on Disk:    150.02 MB
File Metadata/Overhead: 8.14 KB (0.0%)
Overall Compression:   1.00x

=== Analysis by Dataset Name (Grouped) ===
Name                   | Cnt | Type      | Shape           | Logical    | Storage    | Ratio | Comp
-------------------------------------------------------------------------------------------------
frame_gauss0           | 1   | float32   | (1536, 6400)    | 37.50 MB   | 37.50 MB   | 1.0   | None
frame_loose_lf0        | 1   | float32   | (1536, 6400)    | 37.50 MB   | 37.50 MB   | 1.0   | None
frame_mp2_roi0         | 1   | float32   | (1536, 6400)    | 37.50 MB   | 37.50 MB   | 1.0   | None
frame_mp3_roi0         | 1   | float32   | (1536, 6400)    | 37.50 MB   | 37.50 MB   | 1.0   | None
channels_gauss0        | 1   | int32     | (1536,)         | 6.00 KB    | 6.00 KB    | 1.0   | None
channels_loose_lf0     | 1   | int32     | (1536,)         | 6.00 KB    | 6.00 KB    | 1.0   | None
channels_mp2_roi0      | 1   | int32     | (1536,)         | 6.00 KB    | 6.00 KB    | 1.0   | None
channels_mp3_roi0      | 1   | int32     | (1536,)         | 6.00 KB    | 6.00 KB    | 1.0   | None
tickinfo_gauss0        | 1   | float64   | (3,)            | 24 B       | 24 B       | 1.0   | None
tickinfo_loose_lf0     | 1   | float64   | (3,)            | 24 B       | 24 B       | 1.0   | None
tickinfo_mp2_roi0      | 1   | float64   | (3,)            | 24 B       | 24 B       | 1.0   | None
tickinfo_mp3_roi0      | 1   | float64   | (3,)            | 24 B       | 24 B       | 1.0   | None
-------------------------------------------------------------------------------------------------
TOTAL                  | 12  | -         | -               | 150.02 MB  | 150.02 MB  | 1.0   | -
```

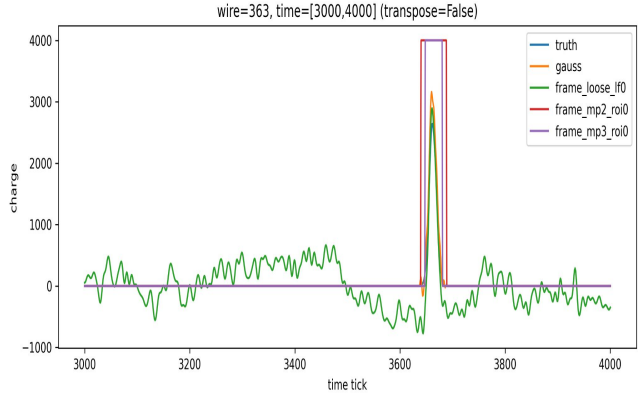
- Per event (1 CRU) file size changes as follows:
  - g4-tru-#.h5: 37.51 MB → 37.51 MB
  - g4-rec-#.h5: 389.33 MB → 150.3 MB
- Approximated required storage for:
  - 600 events:112 GB
  - 1200 events: 224 GB

Back Up

PDHD training sample

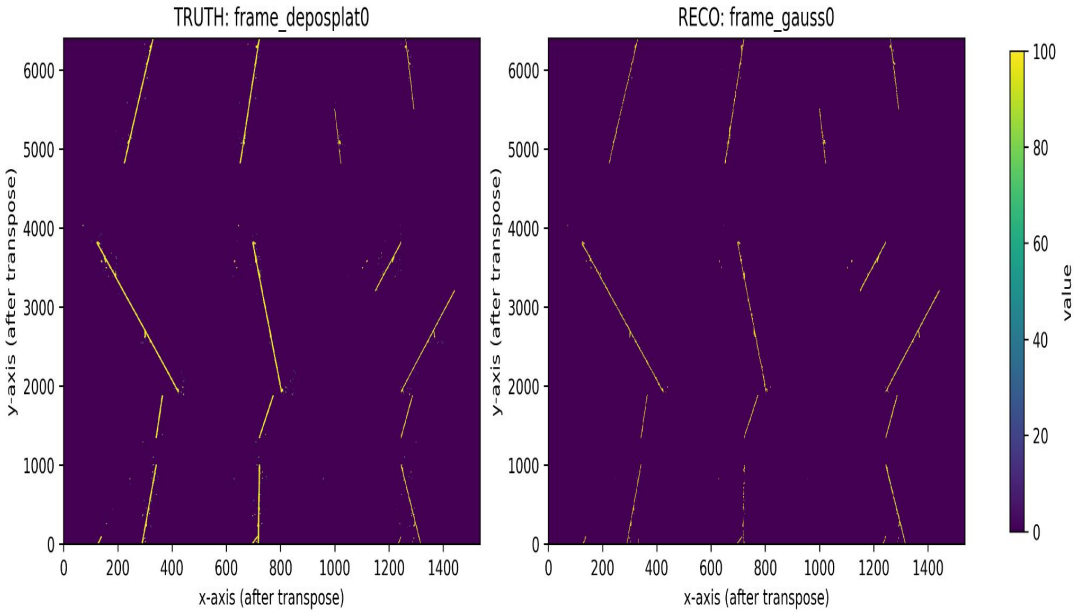


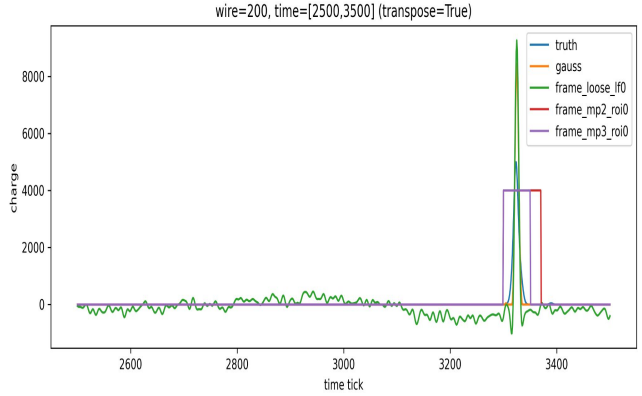


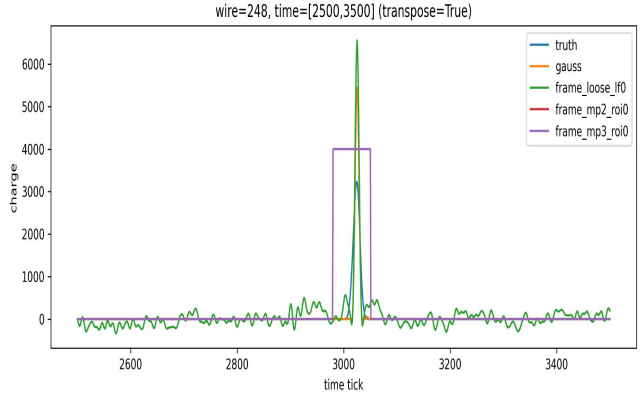


- TBD

PDVD training sample







- Truth (DepoFluxSplat) vs. Gauss
- 2D and 1D waveform are well aligned (measured time offset ~ 1 tick)

Future tasks

- Memory check
  - Is the high peak at the last really due to the APA1?
  - Tag Selector test and validation
  - Remove the unneeded datasets from the HDF5 and check the file size
- PDVD
  - Sample validation
  - Need to check estimated time, scratch storage, WC cluster availability
- Comments from the 2026 DUNE CM @ CERN
  - Jake: noise level (PDHD vs PDVD)
  - Xin: generalize the model, train the model with loose_lf only → performance