Status report on

DNNROI sigproc

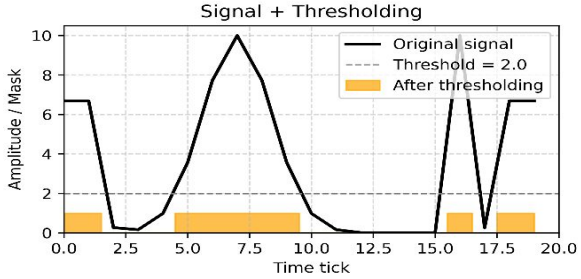
Hokyeong Nam
Chung-Ang University

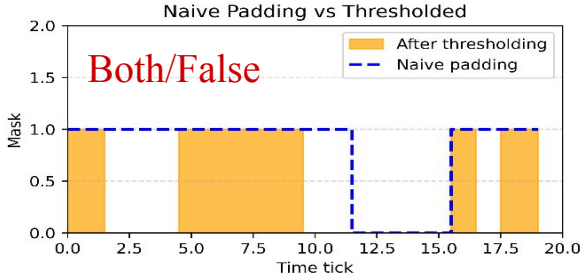
Outline

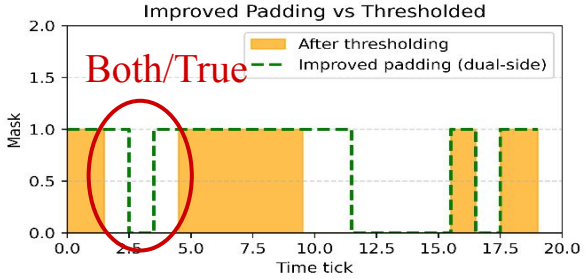
- **PD-HD**
  - **Padding Strategy**
  - **1D Waveform Study (Rebinning / truth_thr / Padding)**
  - **Performance Evaluation**

- PD-VD
  - Simulation (Preparation / Results / Issues)
  - Real Data (Preparation / Results / Issues)
  - Memory Profiling

Padding Strategy - 1D Waveform





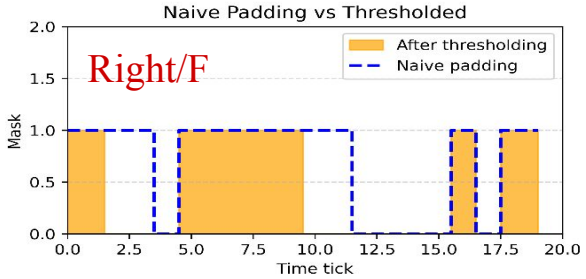


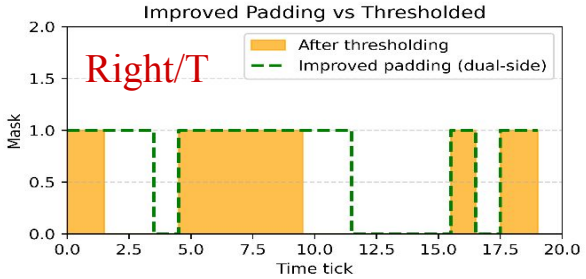
What is changed?

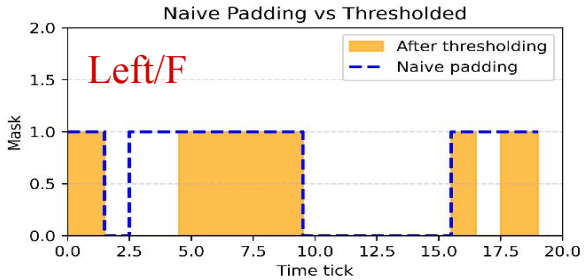
- Can choose which sides to pad
- Can avoid merging

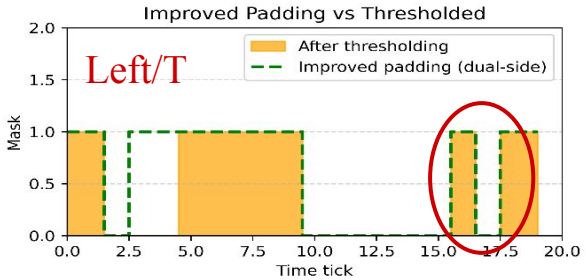
Parameters for this example

- padding = 2
- min_run = 2
- min_gap = 1
- padding_side = both/right/left
- avoid_merge = False/True

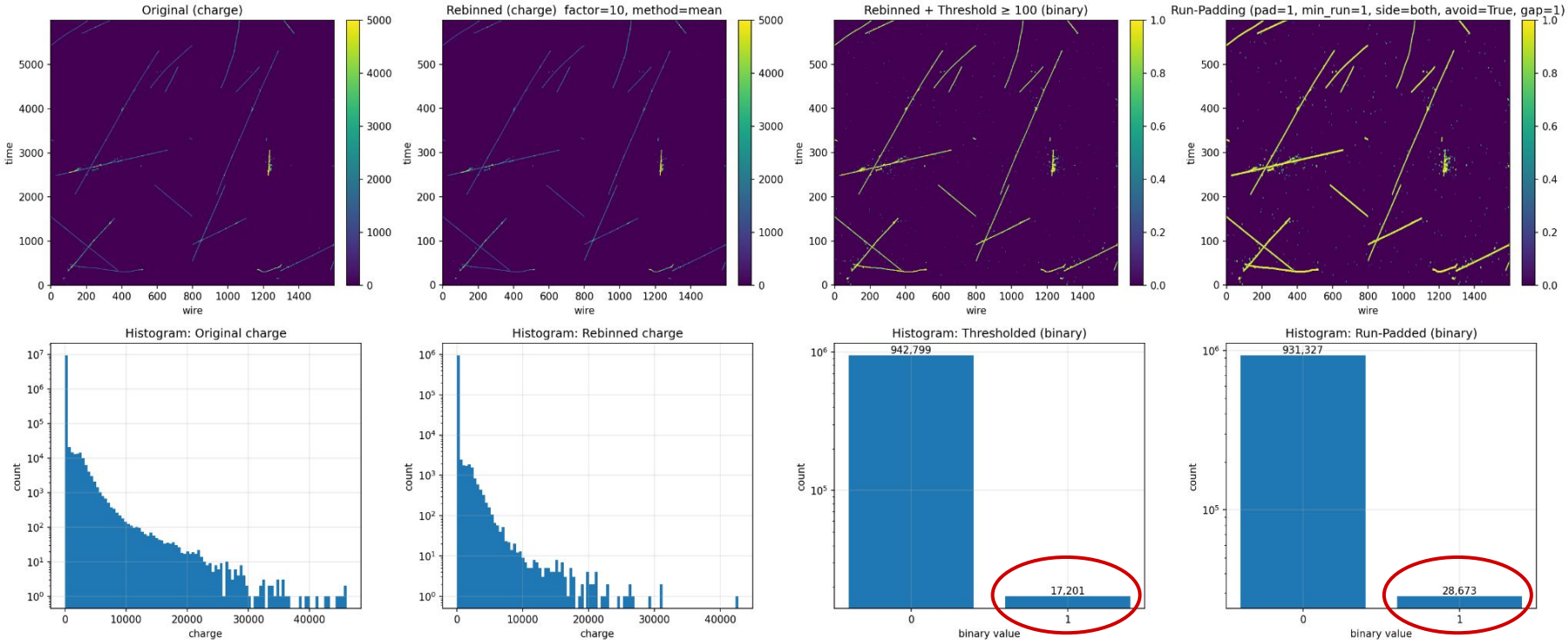






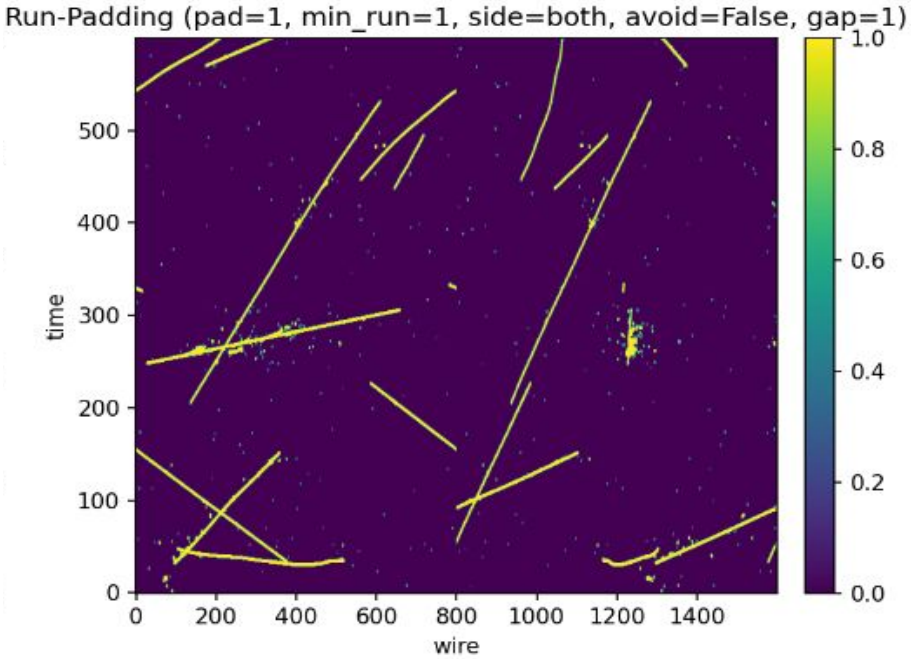


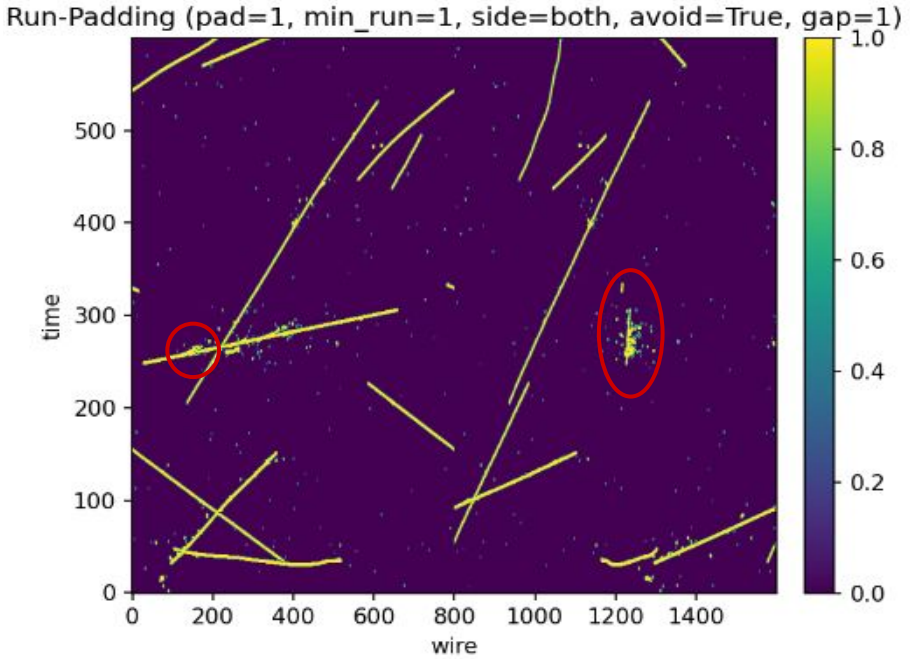
Padding Strategy - 2D Waveform



- Pre-Processing for ML: Original → Rebining → Thresholding → Padding

Padding Strategy - 2D Waveform





Option here: padding = 1, both, min_run = 1, min_gap = 1

If truth masks are clustered closely:

- avoid_merging = False: Connected
- avoid_merging = True: Separated

1D Waveform Study

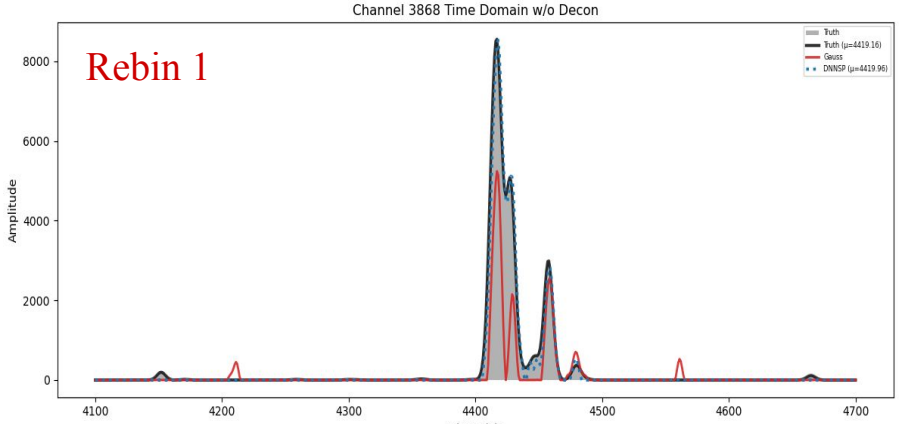
Isochronous shower

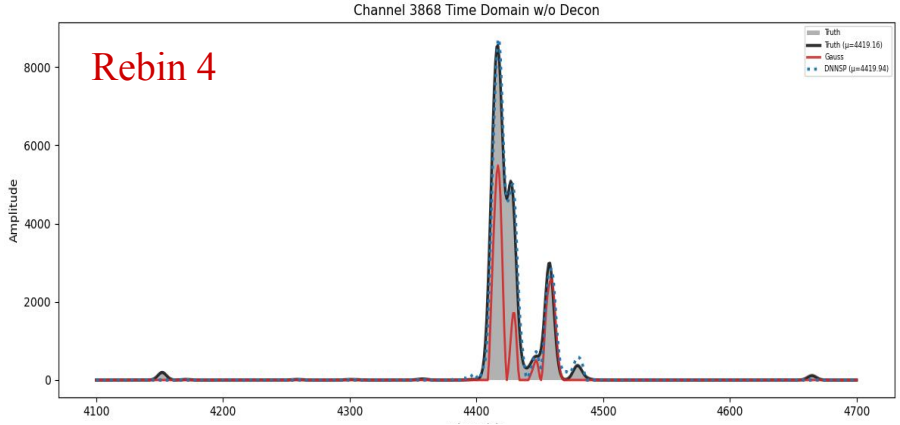
- Rebin - Done
- Truth threshold - To do
- Padding - To do

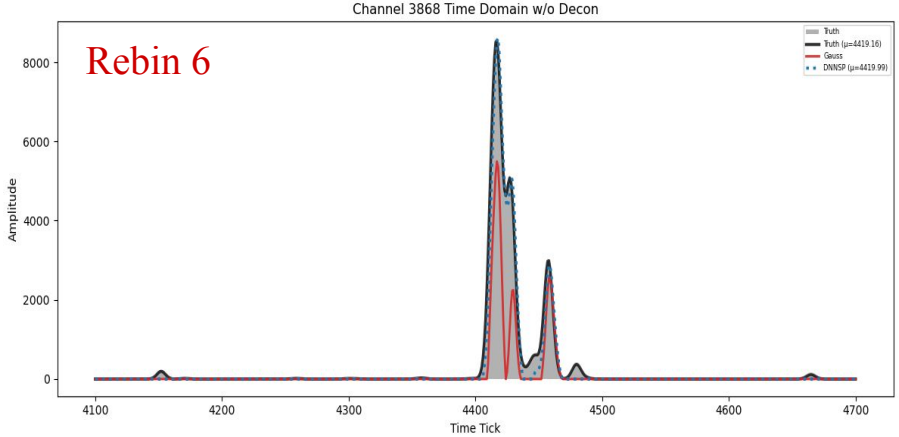
Prolonged shower

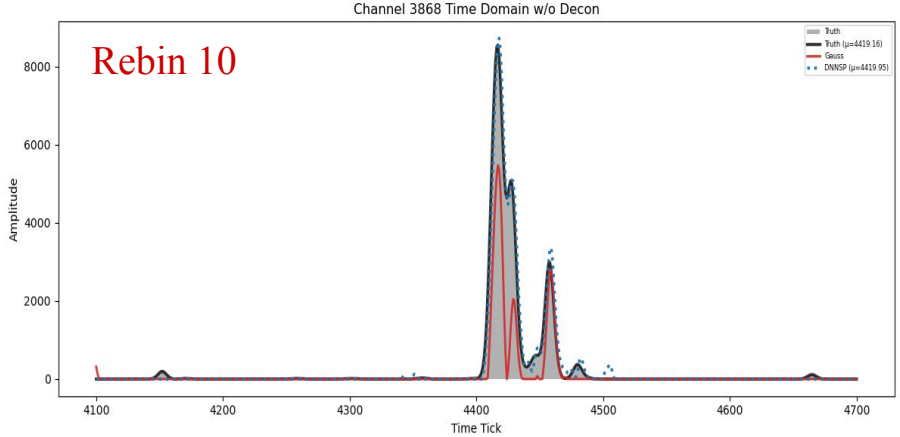
- Rebin - Done
- Truth threshold - To do
- Padding - To do

1D Waveform Study (Isochronous) - Rebinning

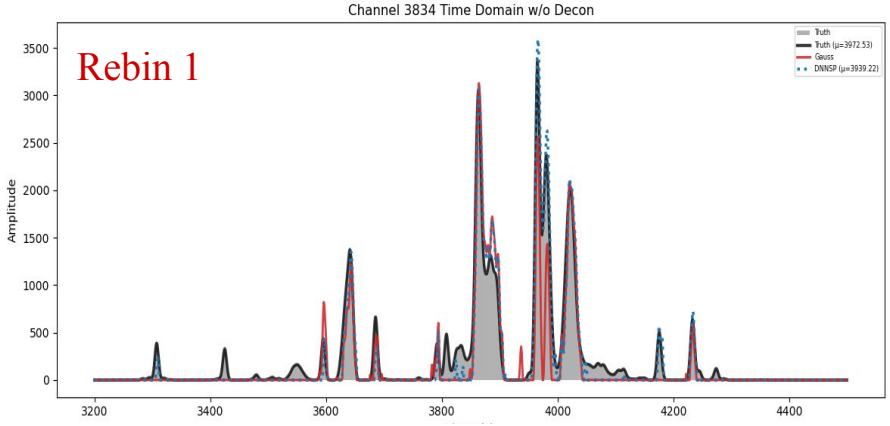


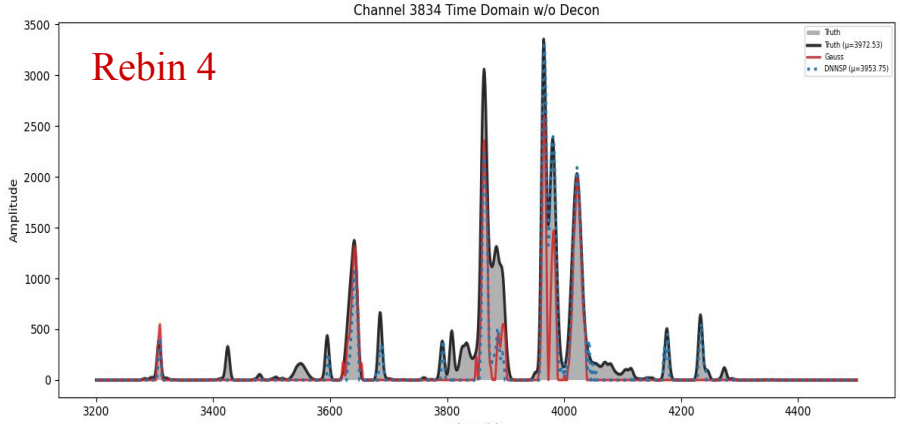


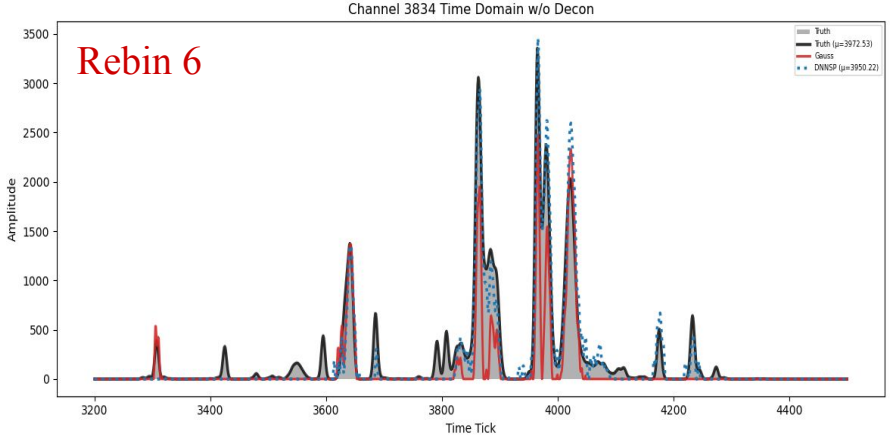


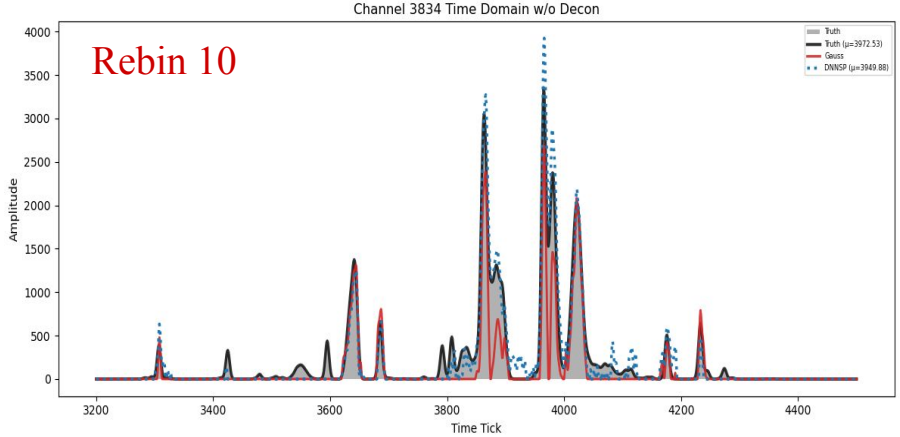


1D Waveform Study (Prolonged) - Rebinning

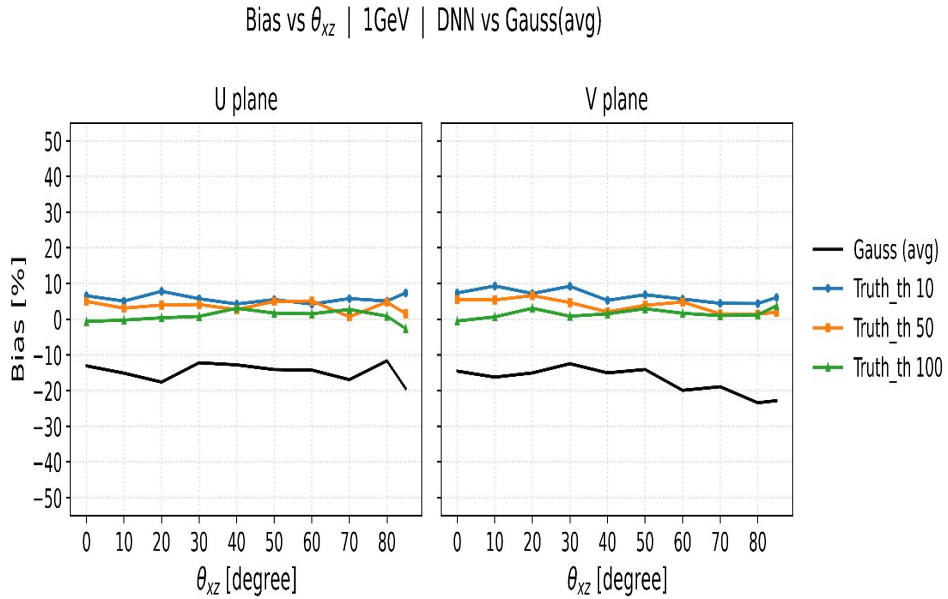


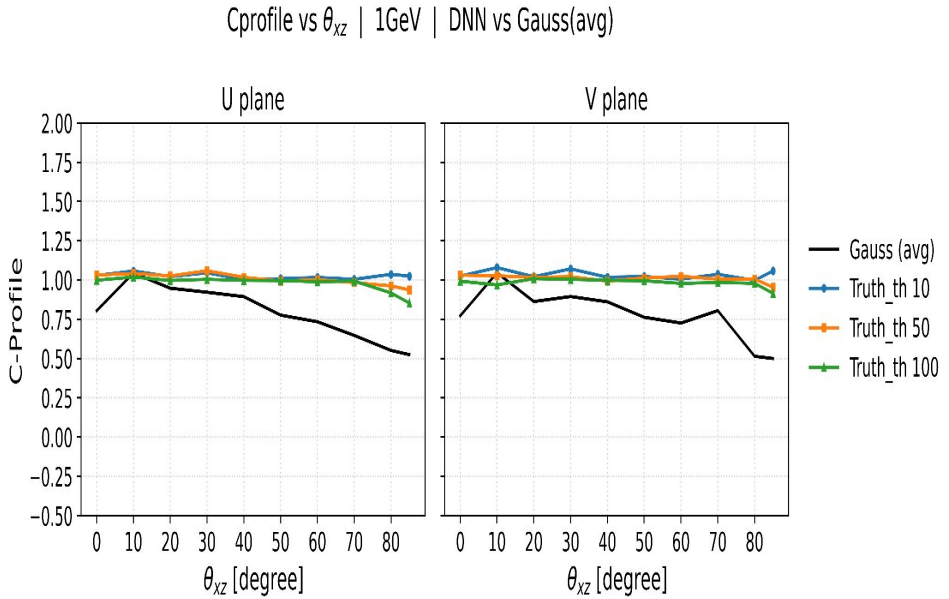






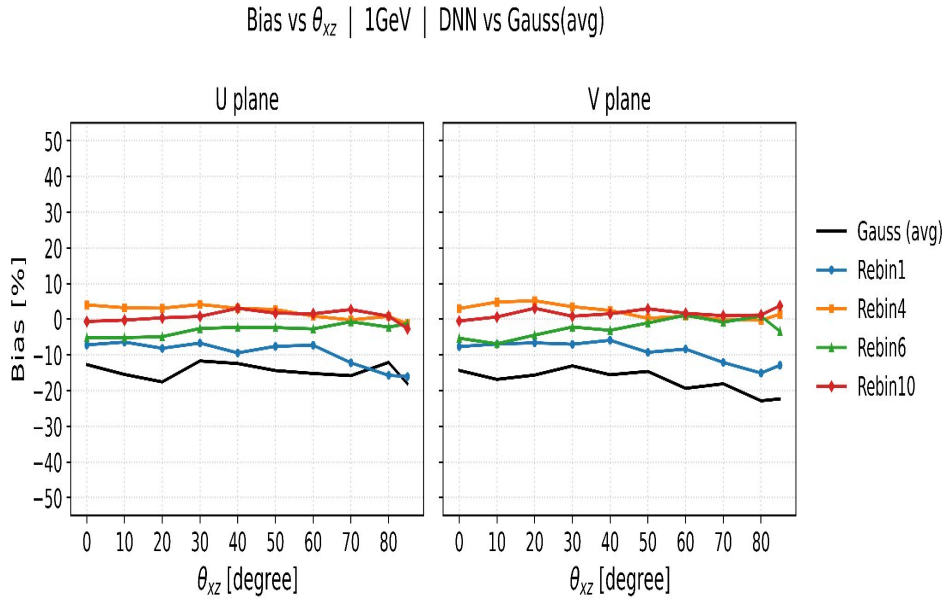
Performance Evaluation - Single Shower

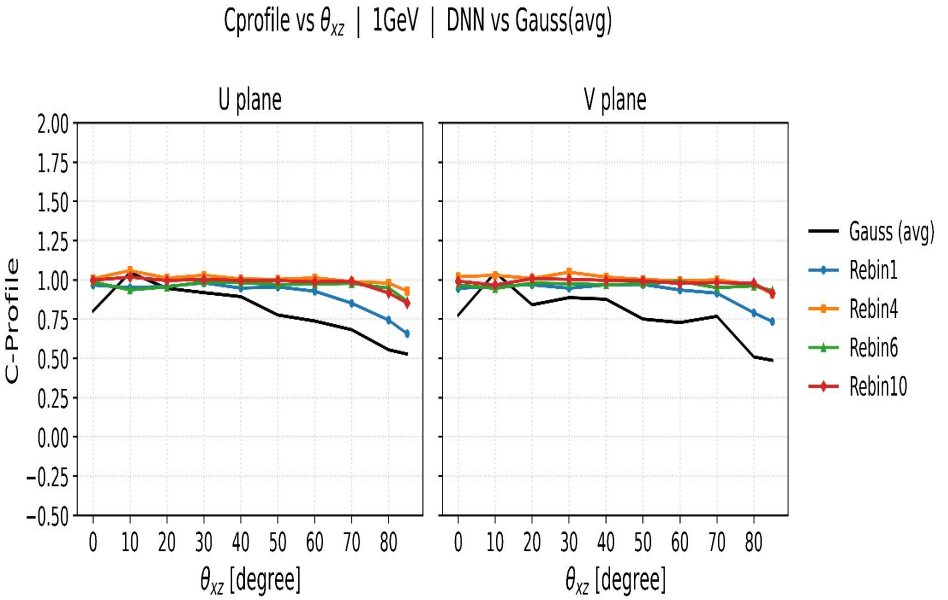




- Measured using **truth threshold** = 10, 50, 100
- Fixed: Rebin = 10, Padding = 0
- The ROI window get narrower as the truth threshold increases

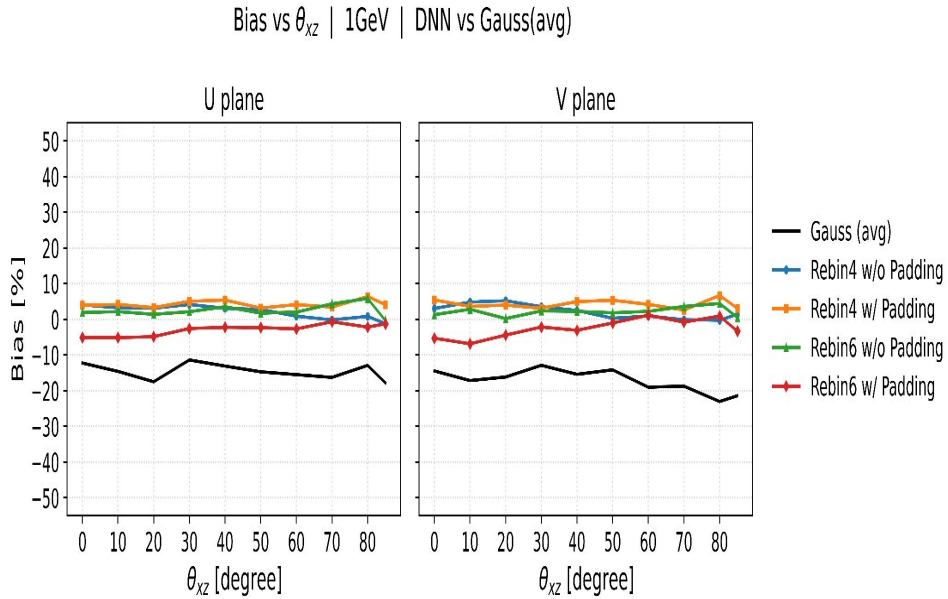
Performance Evaluation - Single Shower

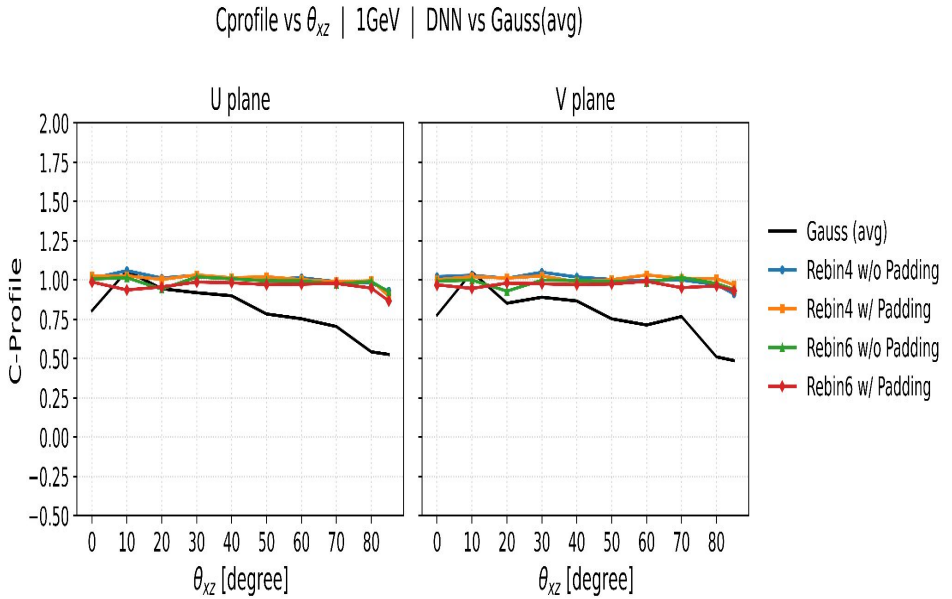




- Measured using **rebin factor** = 1, 4, 6, 10
- Fixed: Truth_th = 100, Padding = 0
- Except for rebin 4, the ROI window get narrower as the rebin factor decreases

Performance Evaluation - Single Shower





- Measured using **rebin factor** = 4, 6 & **Padding** = 1
- Fixed: Truth_th = 100
- Padding effectively broadens the ROI window

Outline

- PD-HD
  - Padding Strategy
  - 1D Waveform Study (Rebinning / truth_thr / Padding)
  - Performance Evaluation

- **PD-VD**
  - **Simulation (Preparation / Results / Issues)**
  - **Real Data (Preparation / Results / Issues)**
  - **Memory Profiling**

PDVD Simulation - Preparation

Software

- SL7 Container & dunesw: v10_05_00d00
- LArSoft: art 3.14.04
- Wire-Cell: 0.30.3

For lar execution

- FCL: pdvd_wirecell_sim_deposplat.fcl (available in /exp/dune/data/users/hnam/wire-cell-hnam/pdvd-data)
- GDML: protodunevd_v4_refactored.gdml

Modified configuration files:

- Path: /exp/dune/data/users/hnam/wire-cell-hnam/larwc/wire-cell-cfg/pgrapher/experiment/protodunevd
- List:
  - wcls-sim-drift-deposplat.jsonnet
  - magnify-sinks.jsonnet
  - dnnroi.jsonnet
  - funcs.jsonnet

PDVD Simulation - Preparation

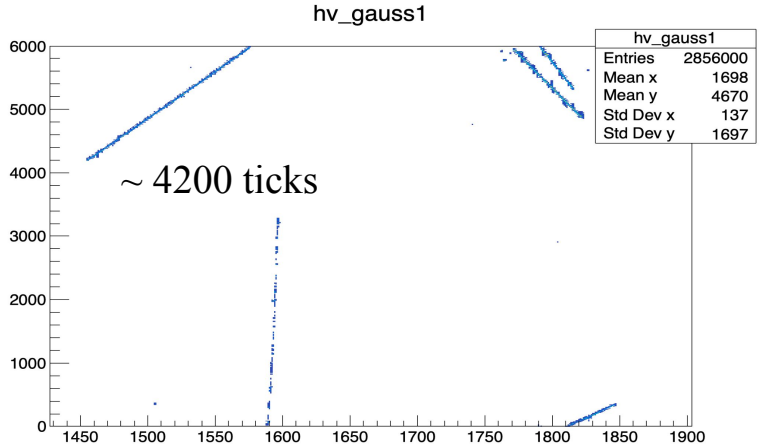
How to run:

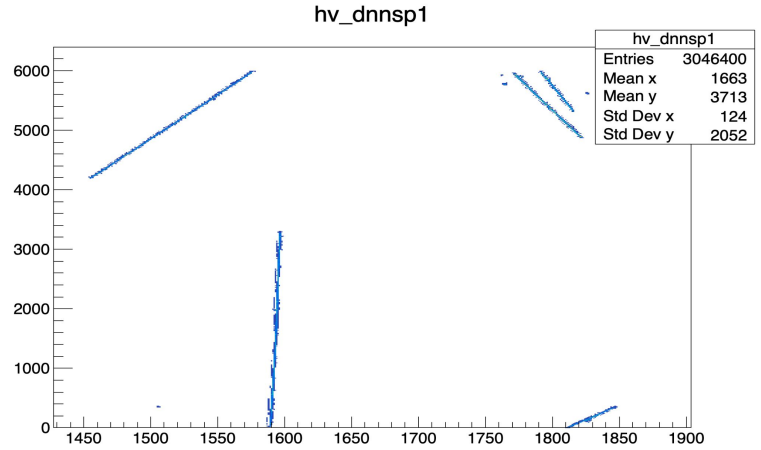
- lar -n1 -c gen_protodunevd_cosmics.fcl -o gen.root
- lar -n1 -c protodunevd_refactored_g4_stage1.fcl -o g4_stage1.root
- lar -n1 -c protodunevd_refactored_g4_stage2.fcl -o g4_stage2.root
- lar -n1 -c pdvd_wirecell_sim_deposplat.fcl g4_stage2.root

Output:

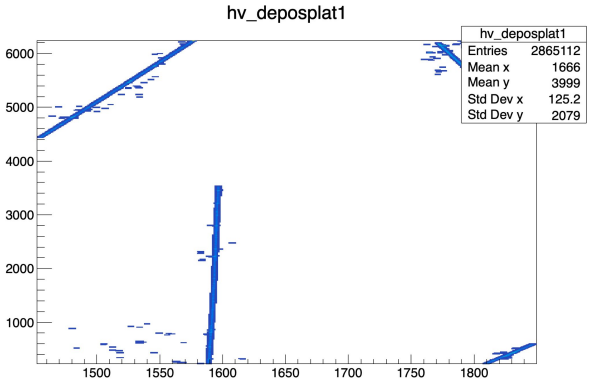
- HDF5 files:
  - g4-rec-#.h5
  - g4-tru-#.h5
- art root file: g4_stage2_deposplat.root
- plain root file: magnify-protodunevd.root

PDVD Simulation - Results & Issues





~ 4400 ticks



- Cosmic-ray event
- DNN improvement is observed
- time offset: ~ 200 ticks

PDVD Real Data - Preparation

Software

- SL7 Container & dunesw: v10_05_00d00
- LArSoft: art 3.14.04
- Wire-Cell: 0.30.3

For lar execution

- FCL: **reco.fcl** (available in /exp/dune/data/users/hnam/wire-cell-hnam/pdvd-data)
- GDML: **protodunevd_v5_gdd.gdml**
  (export FW_SEARCH_PATH=$FW_SEARCH_PATH:/cvmfs/dune.opensciencegrid.org/products/dune/dunecore/v10_10_04d00/gdml)

Modified configuration files:

- Path: /exp/dune/data/users/hnam/wire-cell-hnam/larwc/wire-cell-cfg/pgrapher/experiment/protodunevd
- List:
  - **wcls-nf-sp.jsonnet**: time ticks for TDE/BDE, TorchScript file, chunking
  - magnify-sinks.jsonnet
  - dnnroi.jsonnet

PDVD Real Data - Preparation

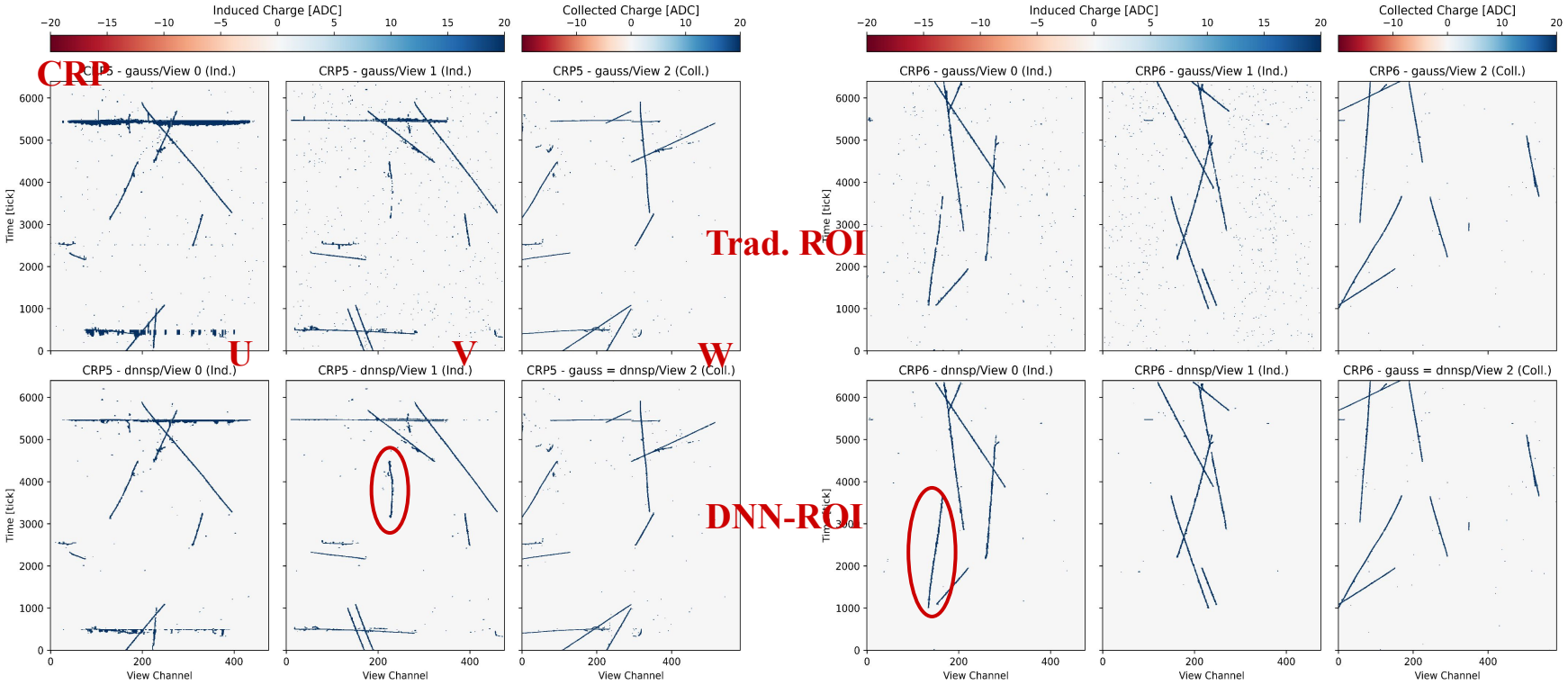
How to run:

- lar -n1 -c reco.fcl -s reco_sate1.root
  (e.g. np02vd_raw_run039300_0106_df-s03-d0_dw_0_20250904T012512_reco_stage1_20250904T063407_keepup.root)

Output:

- art root file
  (e.g. np02vd_raw_run039300_0106_df-s03-d0_dw_0_20250904T012512_reco_stage1_20250904T063407_keepup_**reco**.root)
- plain root file: protodune-data-check.root

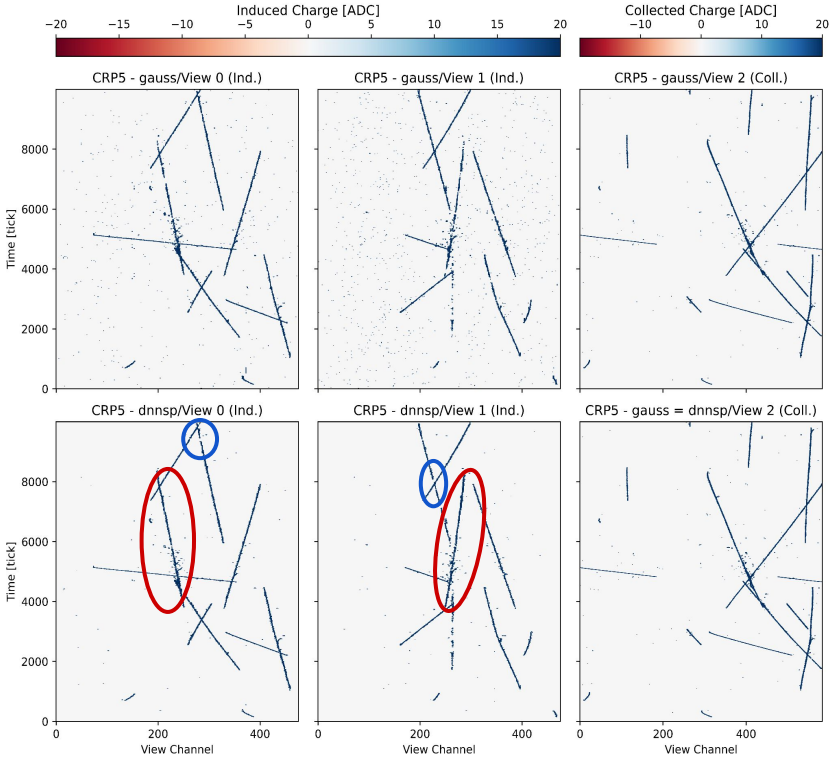
PDVD Real Data - Results

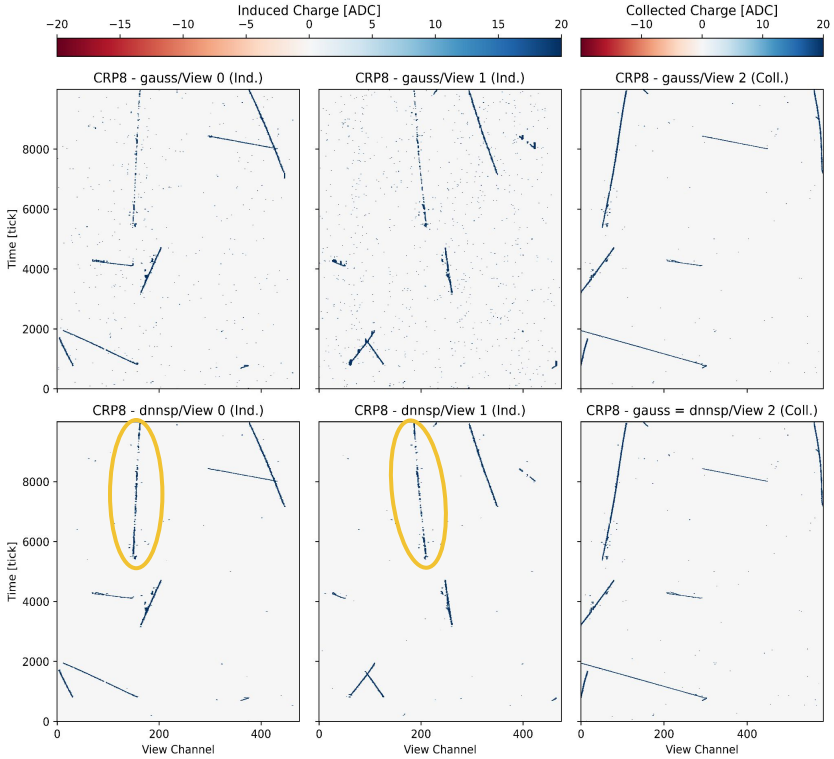


Run 039300-0106 (Better, Worse, Similar, Not understood)

- Readout window (3.2 ms, beam), CRP: 5, 6, 7, 8

PDVD Real Data - Results

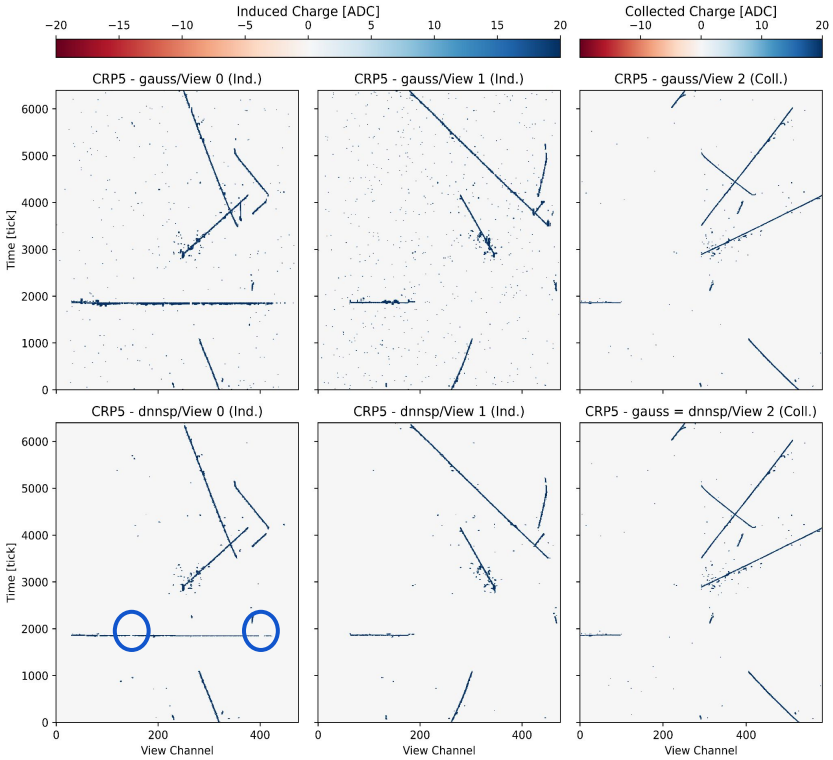


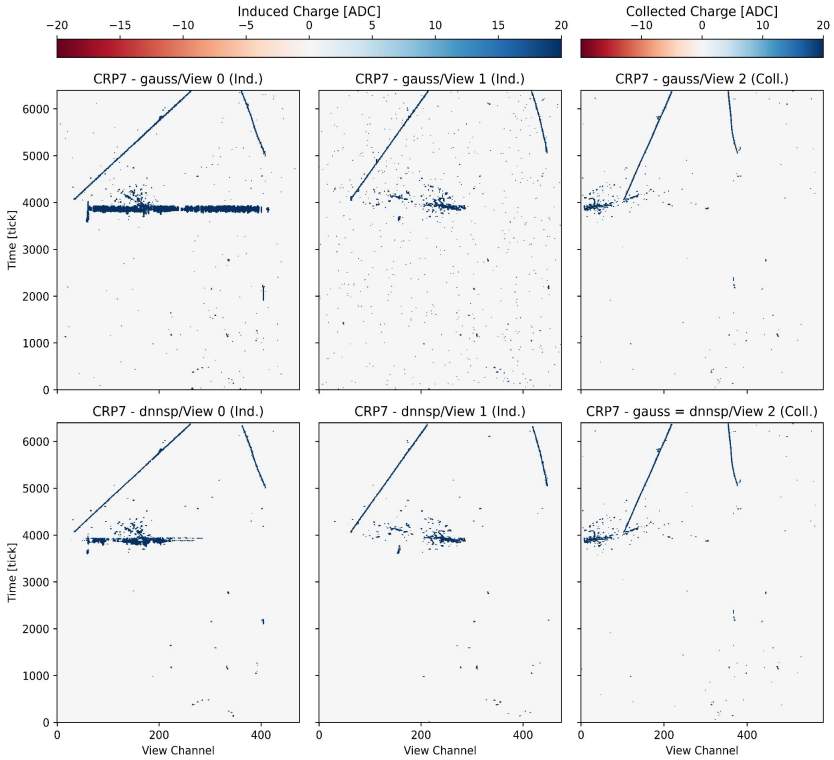


Run 039300-0115 (Better, Worse, Similar, Not understood)

- Readout window (5 ms, random), CRP: 5, 6, 7, 8

PDVD Real Data - Results

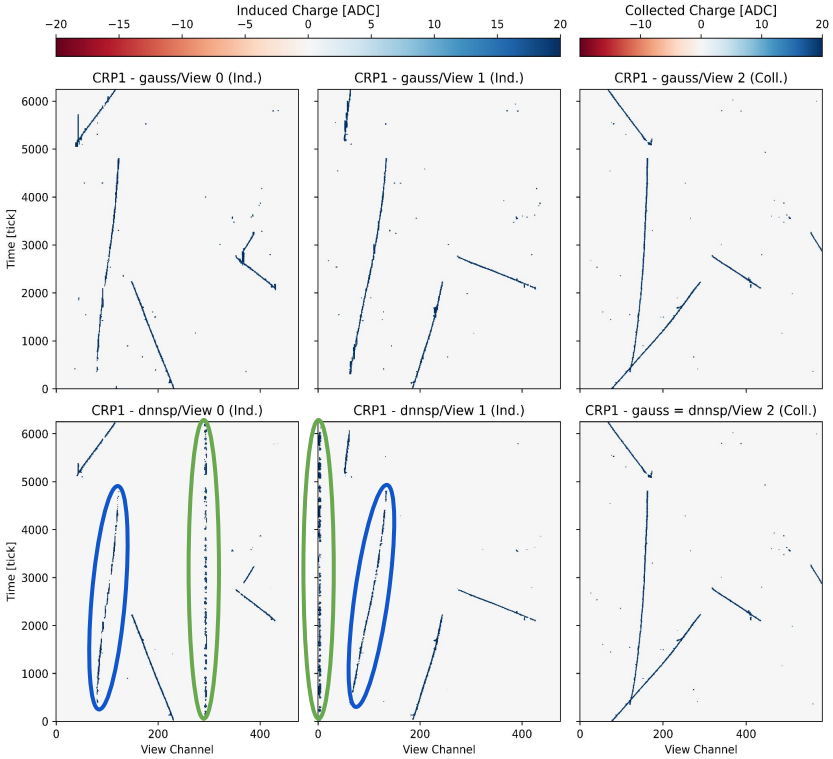


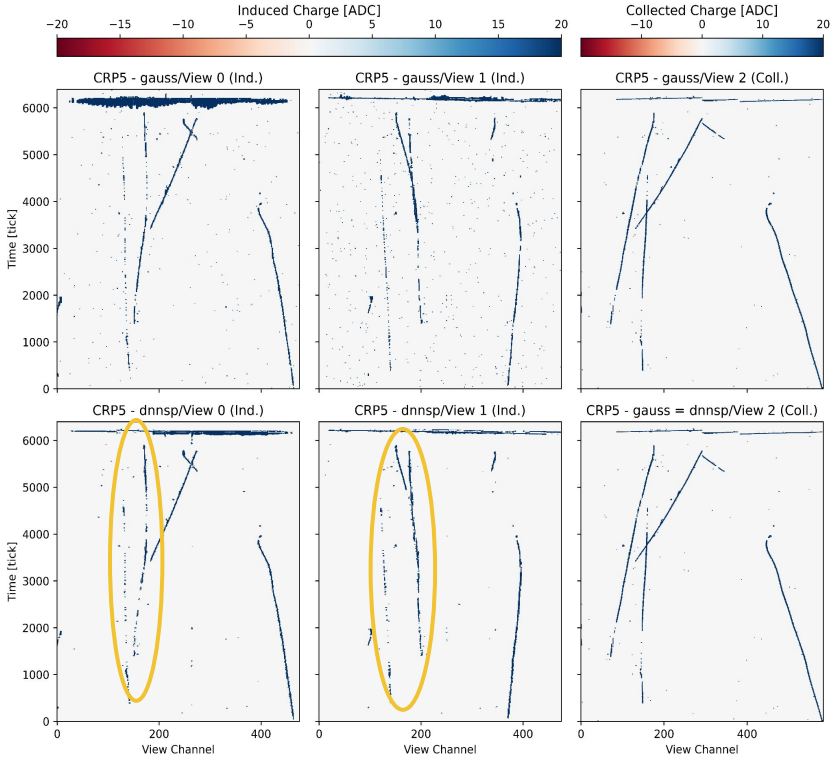


Run 039300-0141 (Better, Worse, Similar, Not understood)

- Readout window (3.2 ms, beam), CRP: 5, 6, 7, 8

PDVD Real Data - Results

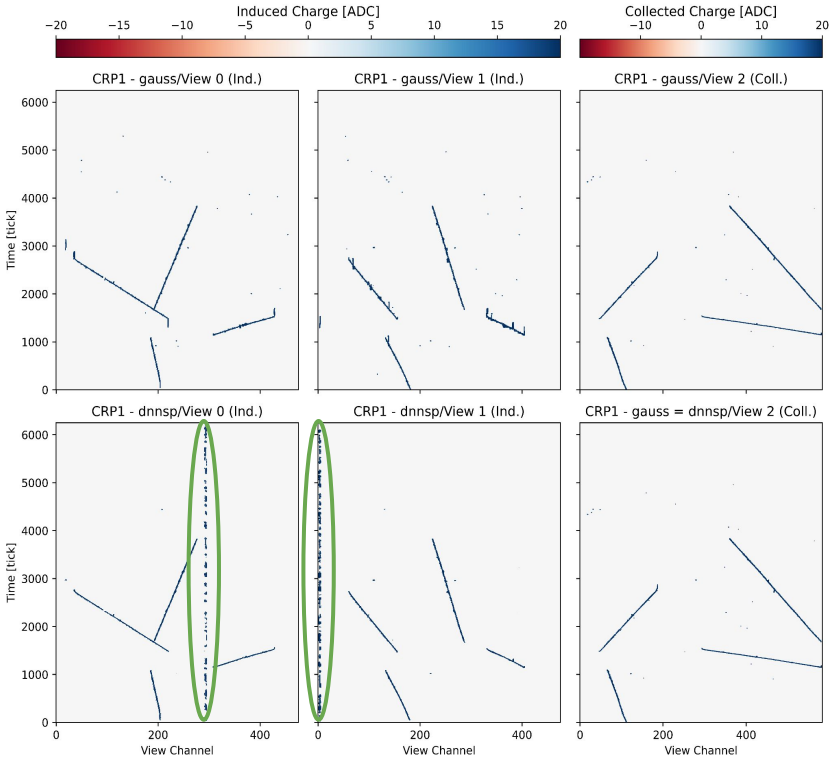


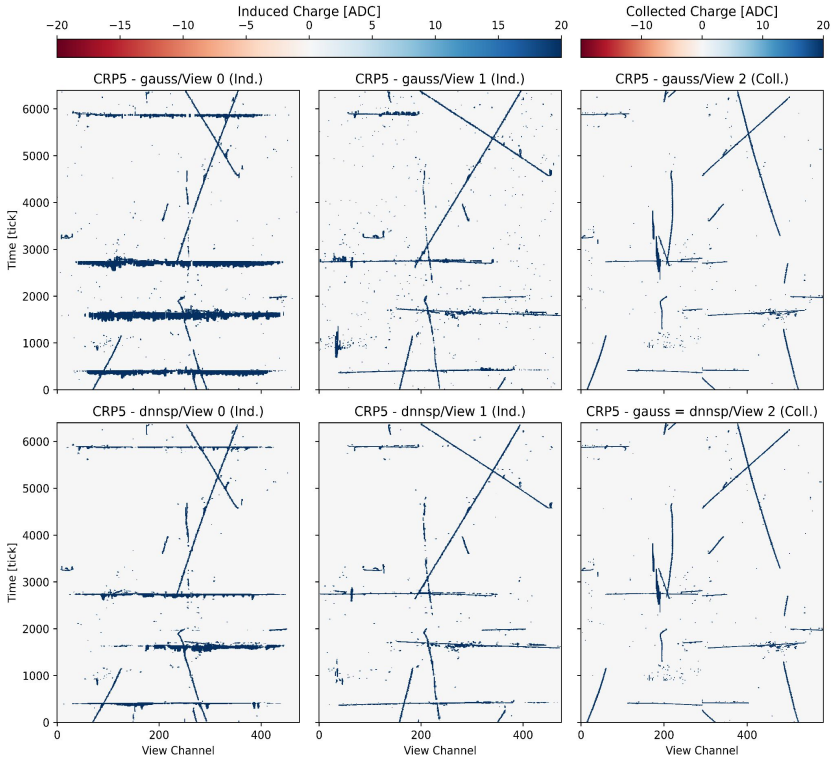


Run 039324-0647 (Better, Worse, Similar, Not understood)

- Readout window (3.2 ms, beam), CRP: 1, 2, 3, 4, 5, 6, 7, 8

PDVD Real Data - Results





Run 039324-0649 (Better, Worse, Similar, Not understood)

- Readout window (3.2 ms, beam), CRP: 1, 2, 3, 4, 5, 6, 7, 8

PDVD Real Data - Issues

In jsonnet files, are detector configuration for PDVD validated enough?

Time ticks (y-axis)

- y-axis readout window depends on run configuration: beam/random, TDE/BDE, resampling
- Downsampling along y-axis for ML inference could be issues (e.g. 6250/6 is not an integer → rebin 6 X)
- For now, we need to manually put the time bins for DNN-ROI → how to automize?

Channels (x-axis)

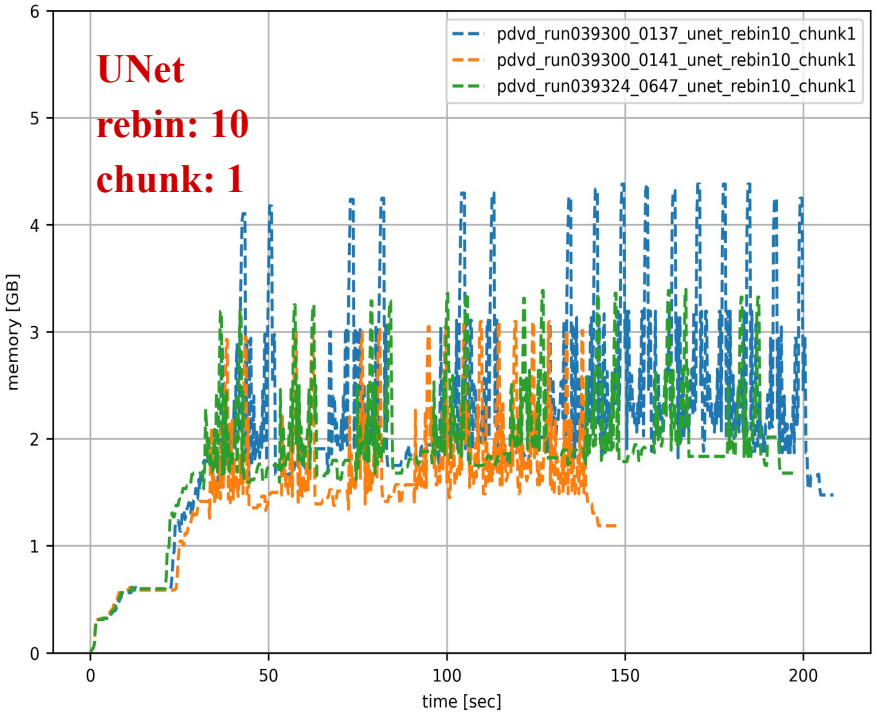
- Wire channels should be (Laura Zambelli's talk on NuFact 2023):

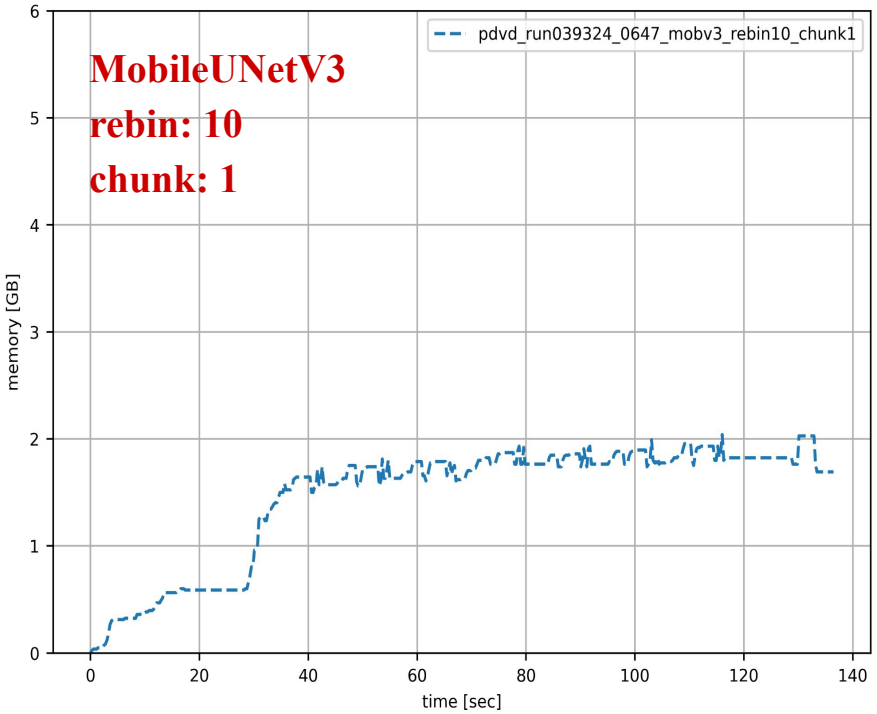
> ▸ Induction views : 7.65 mm wide strips, total of 952/view/CRP
> Collection view : 5.1 mm wide strips, total of 1168 strips/CRP

- The below is log from LArWC execution, the number is the half of the expected channels

```
[12:51:45.623] D [magnify ] MagnifySink: cbin:476[9691.5,10167.5] tbin:10000[0,10000]
[12:51:45.685] D [magnify ] MagnifySink: cbin:476[10643.5,11119.5] tbin:10000[0,10000]
[12:51:45.749] D [magnify ] MagnifySink: cbin:584[11703.5,12287.5] tbin:10000[0,10000]
```

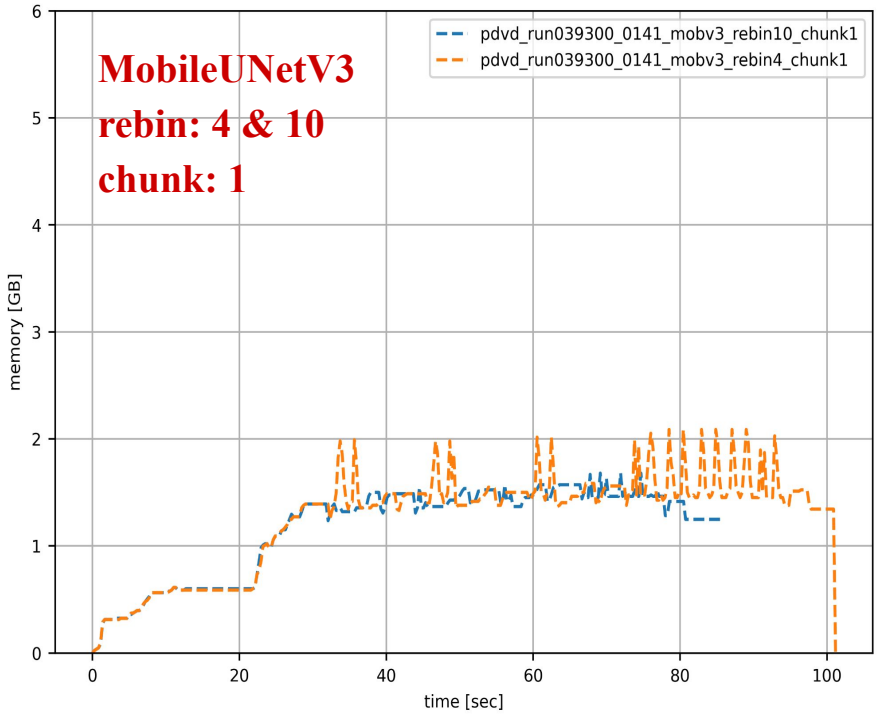
PDVD Memory Profiling - Activity Logger

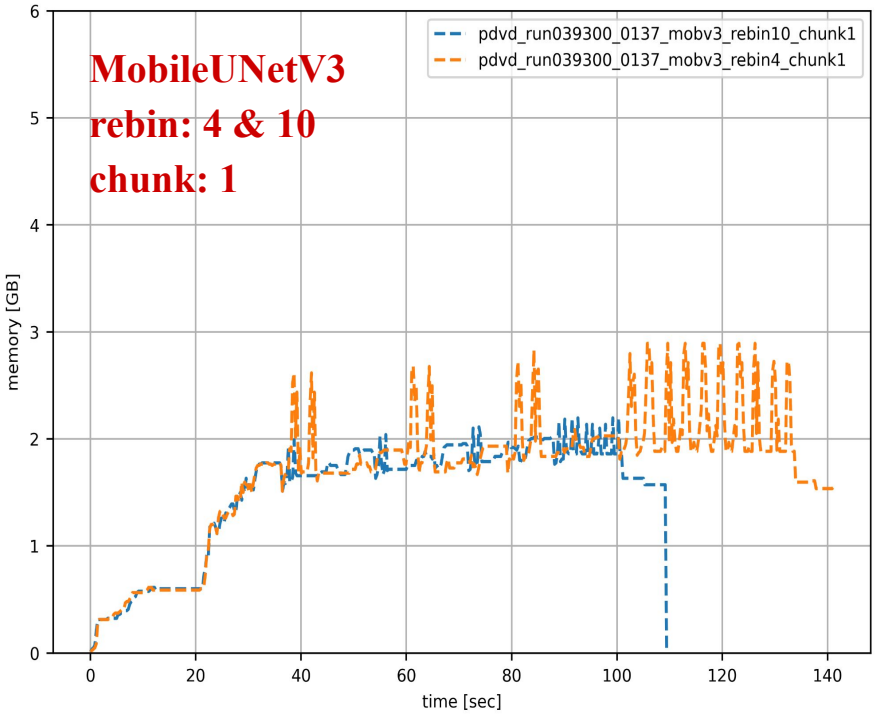




- Later peaks from Blue and Orange are clustered because the data is empty (CRP 1, 2, 3, 4)
- Green has consistent gap along the time
- Blue has highest peak because of input image size for ML inference (476 channel x 10000 tick)

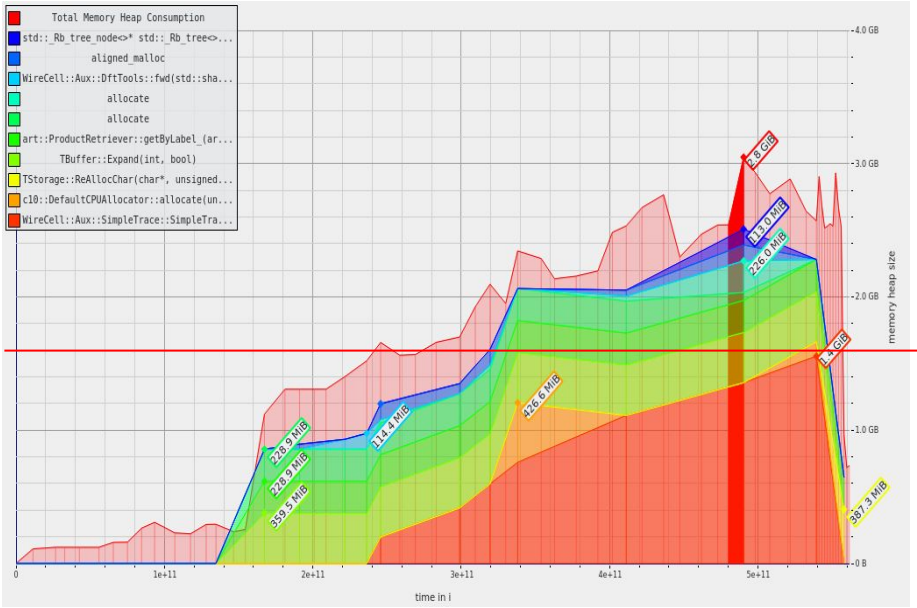
PDVD Memory Profiling - Activity Logger

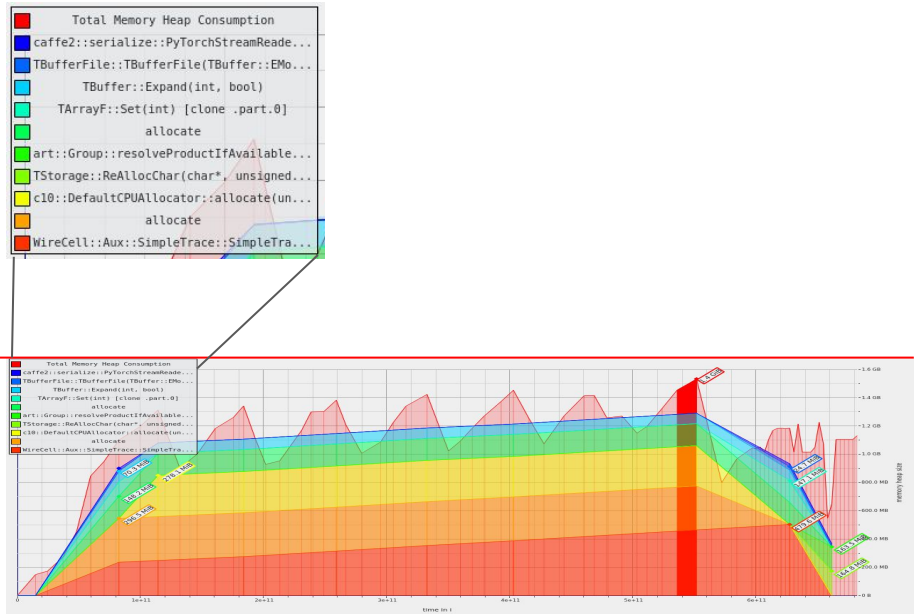




- Left - beam (476 x 6400) vs Right - random (476 x 10000)
- Orange - rebin 4 vs. Blue rebin - 10
- **Accumulated memory in PD-VD is MUCH smaller than PD-HD**

Memory Profiling - Valgrind Massif





- ❑ PD-HD real data processing w/ DNN-ROI
- ❑ PD-VD real data processing w/ DNN-ROI

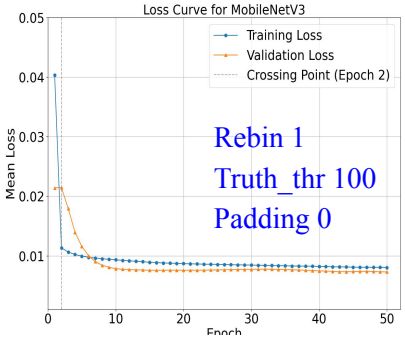
- Heap profiling confirmed accumulated memory is lower in PD-VD case
- The massif output files are available (/exp/dune/data/users/hnam/wire-cell-hnam/share/massif)
  - massif_(pdvd or pdhd)_mobileunetv3_timei1.out
- How to investigate this further?

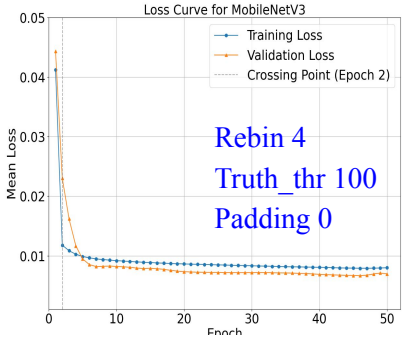
Summary

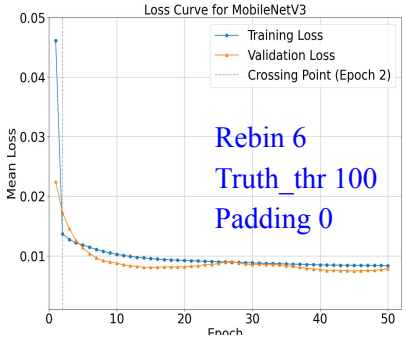
- General
  - Padding algorithm is ready for ML training
- PD-HD
  - Studied 1D waveform and checked performance with current evaluation metrics
- PD-VD
  - For both simulation and real data, LArWC runs w/o errors, but validation is needed (channel, time, gain, drift time, etc.)
  - In real data processing, DNN-ROI reconstructs dead channels → how to handle?
- Other plans
  - Refine 2D waveform evolution plots with Xuyang's image threshold
  - Full data training & compare results to Dikshant's model → Yujin is working on it

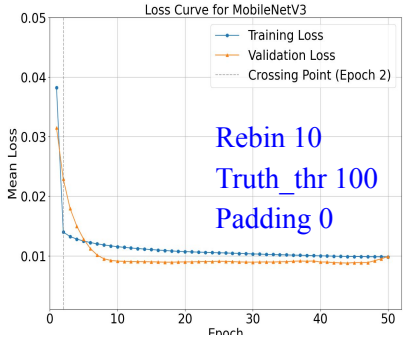
Back Up

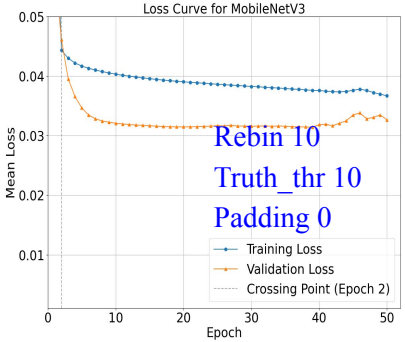
Train vs. Val. Loss Curve

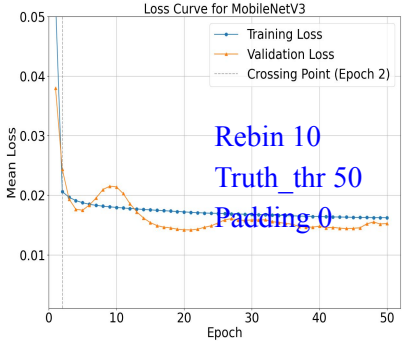


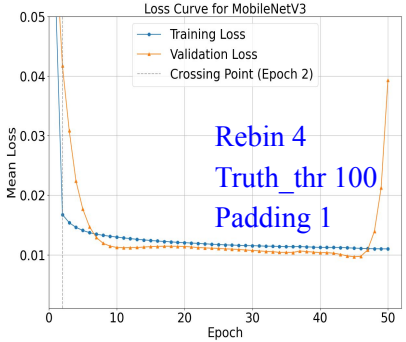


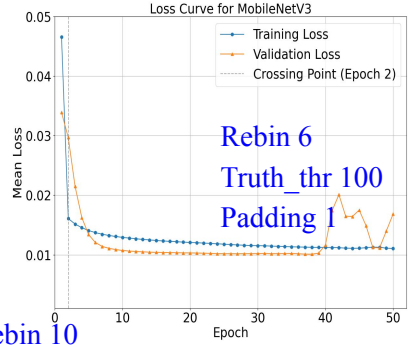


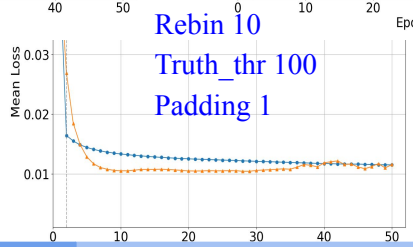






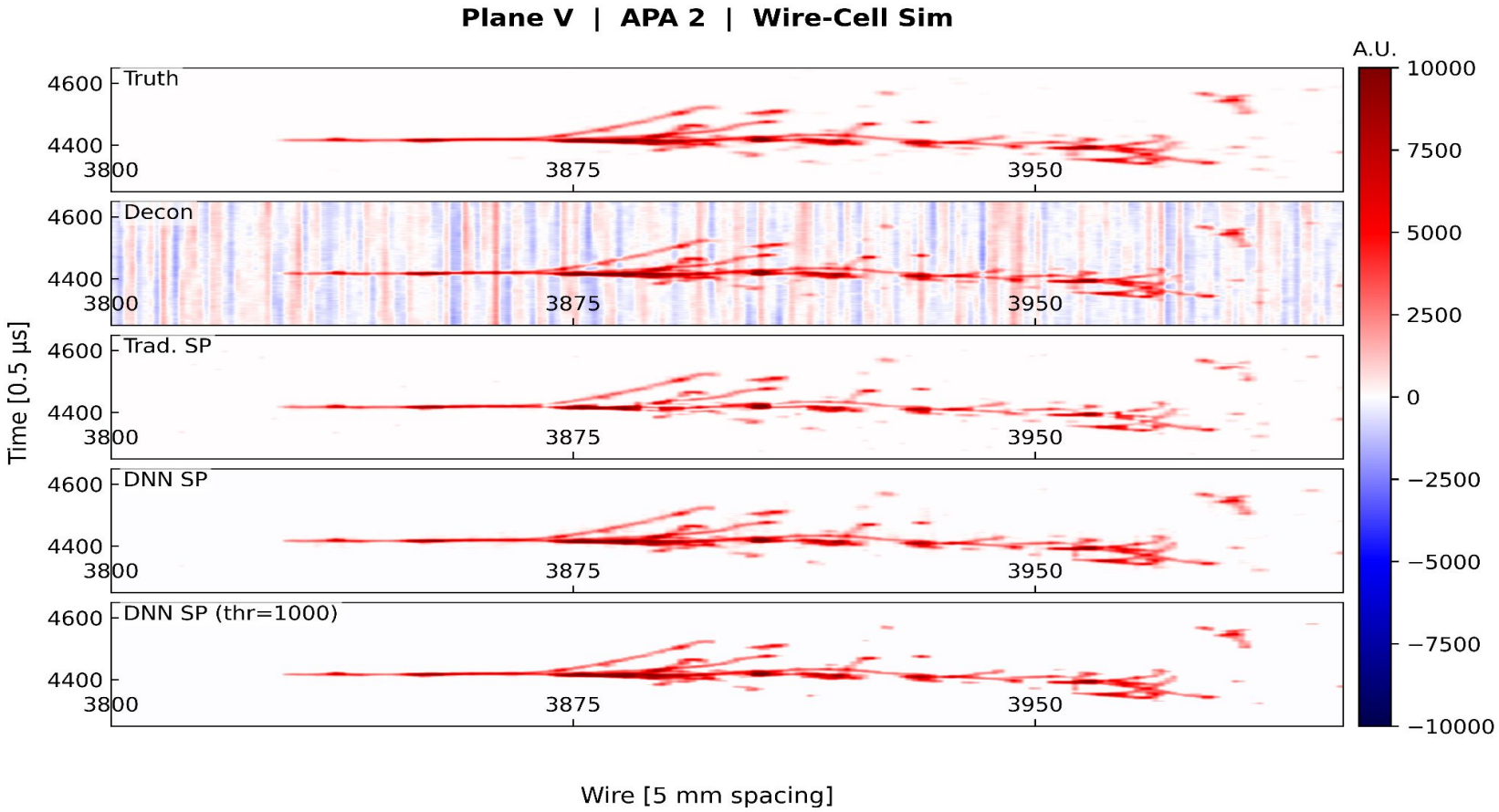




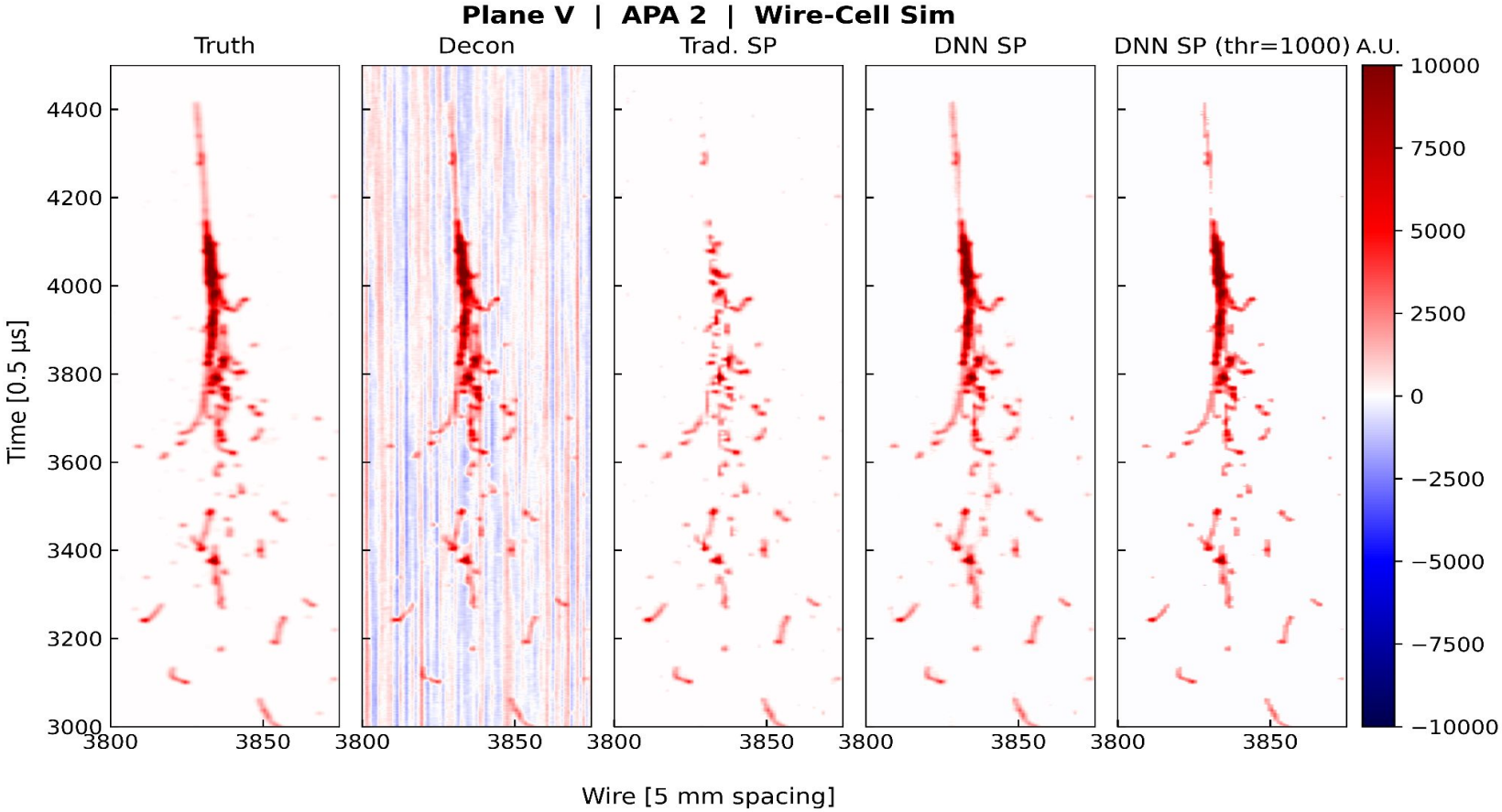


- MobileUNet V3
- Dataset: cosmic-ray 590 events

Imaging Threshold

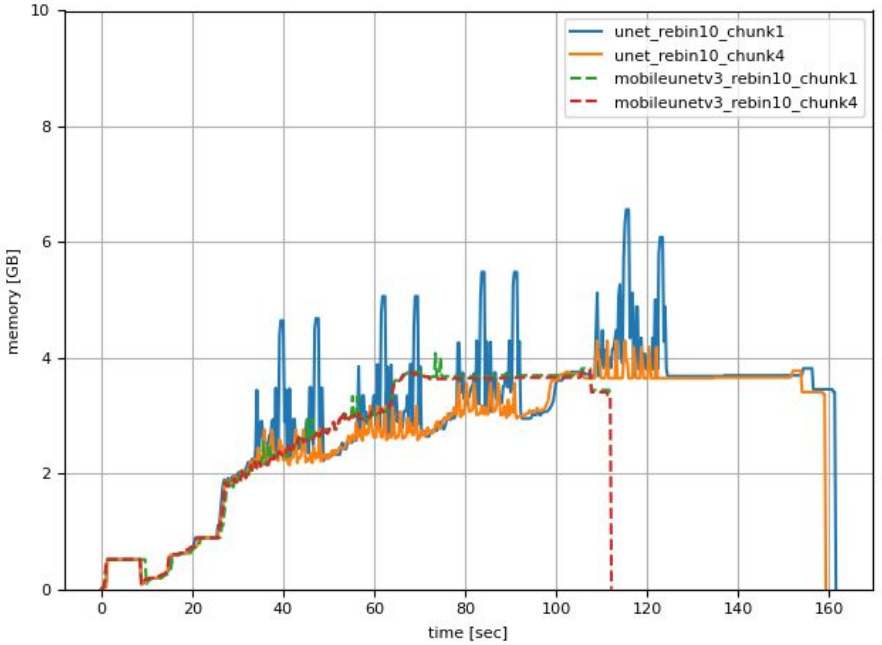


Imaging Threshold



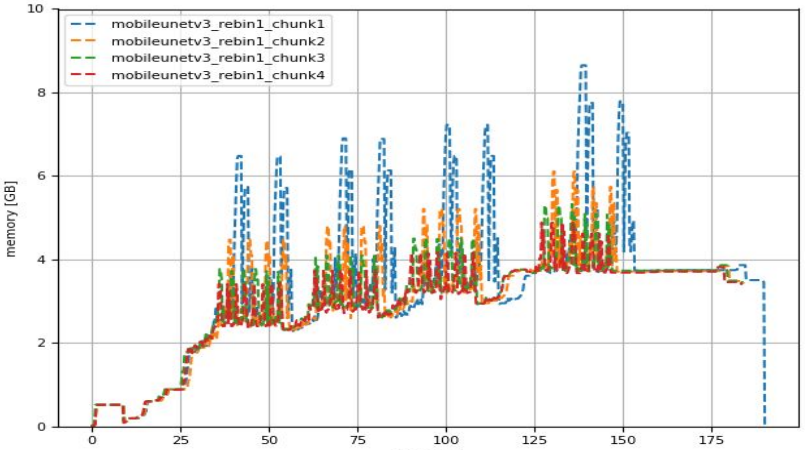
Memory Profiling - Memory Tracker & Activity Logger

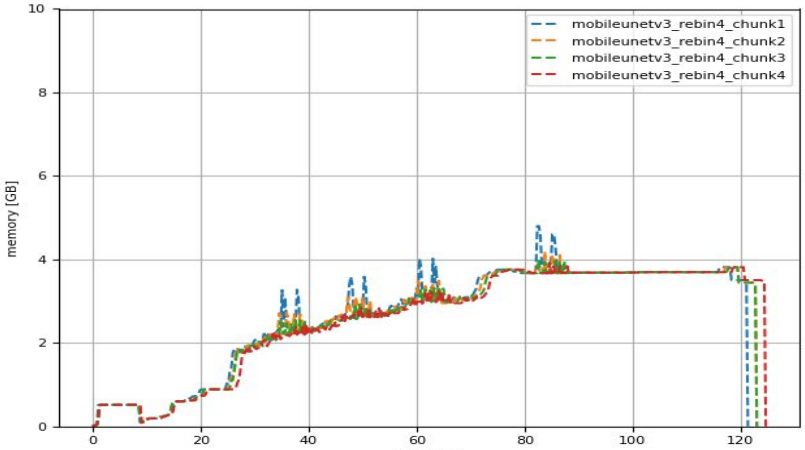
	MobileUNet V3 (MB)				UNet (MB)
chunk	Rebin 1	Rebin 4	Rebin 6	Rebin 10	
1	8848	4907	4448	**4172**	**6710**
2	6249	4278	4107	3966	5198
3	5447	4101	3990	3874	4646
4	4997	4015	3930	**3860**	**4387**

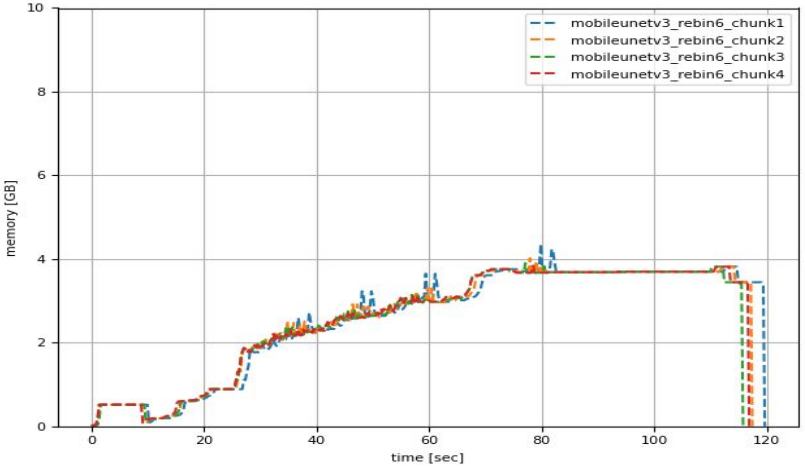


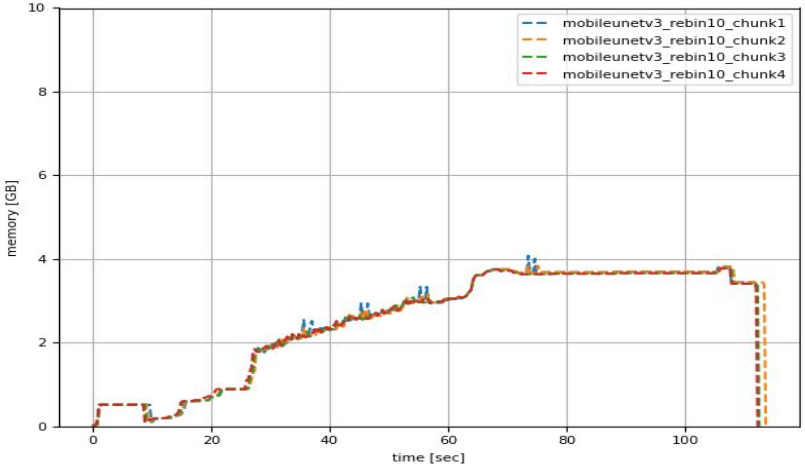
- Logged memory usage by top command (Haiwang's activity logger)
- Trad. SP reaches 2 GB → DNN SP starts with 2GB baseline
- Final accumulated memory is about ~3.9 GB, not depending on model architecture

Memory Profiling - Activity Logger

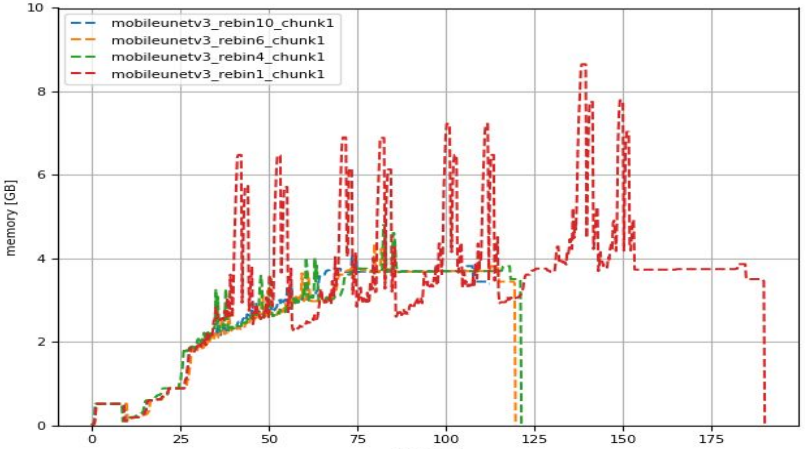


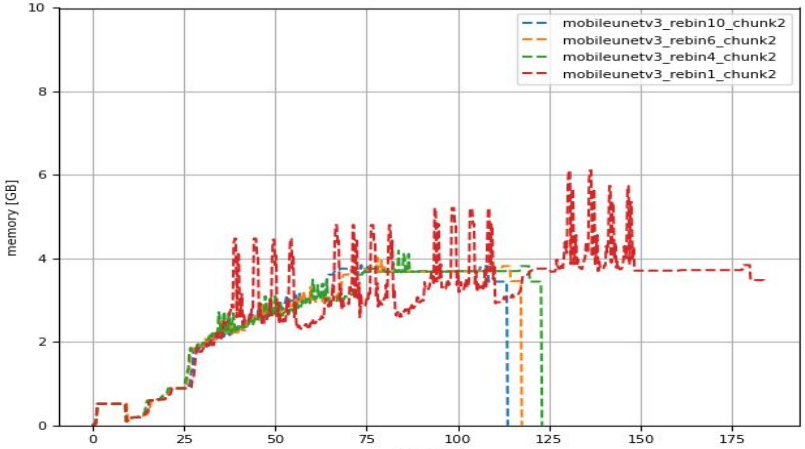


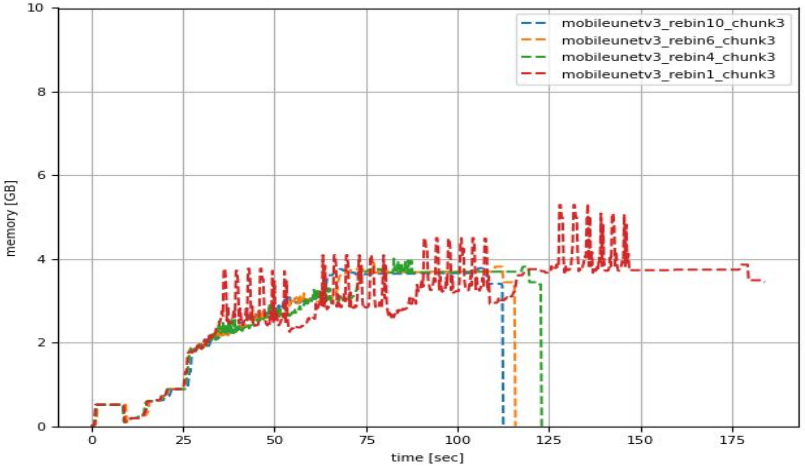


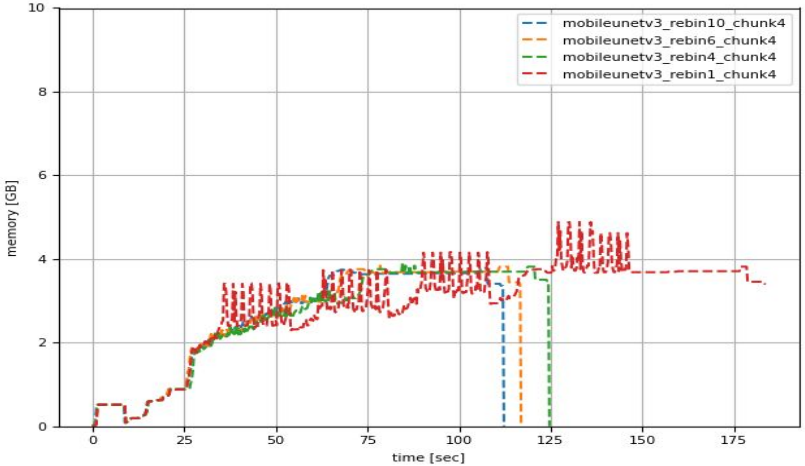


Memory Profiling - Activity Logger









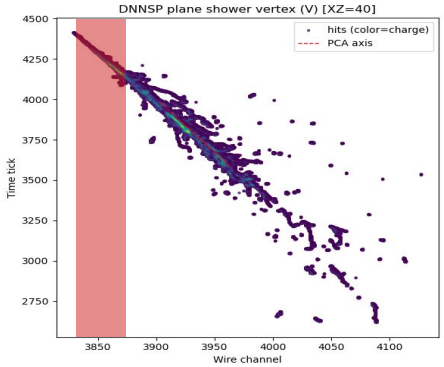
DNN-ROI Performance Evaluation

- For track events, three metrics are used: Bias, Resolution, and Inefficiency

$$Bias = 100 \times \left(\left\langle \frac{Q_{reco}}{Q_{truth}} \right\rangle - 1 \right) \qquad Resolution = 100 \times \frac{RMS\left(\frac{Q_{reco}}{Q_{truth}}\right)}{\left\langle \frac{Q_{reco}}{Q_{truth}} \right\rangle} \qquad Inefficiency = 100 \times \frac{Number\ of\ bad\ channels}{Number\ of\ valid\ truth\ channels}$$

- For shower events, a charge profile based on vertex information was added as the fourth metric:
  - Sum the charge along the shower direction up to 42 wire channels (≈ 1–2 radiation lengths)
  - Compare the reconstructed-to-truth ratio charge ratio

$$Q_{method} = \sum_{w \in W} Q_{method}(w) \qquad R_{cprofile} = \frac{Q_{reco}}{Q_{truth}}$$



- Samples were generated with
  - Detector configuration: ProtoDUNE - Horizontal Drift (PD-HD)
  - XZ angle: 0°, 10°, 20°, 30°, 40°, 50°, 60°, 70°, 80°, 85°
  - Shower energies: 100 MeV, 500 MeV, 1 GeV, 2 GeV, 3 GeV, 5 GeV
  - Software: WCT standalone (Tracks), LAr-WCT (Showers)