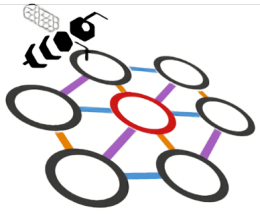
Deep-Learning Signal Processing with Geometry Information

Haiwang Yu (BNL)
for the Wire-Cell team

ProtoDUNE Sim/Reco Meeting

Jan. 8, 2020

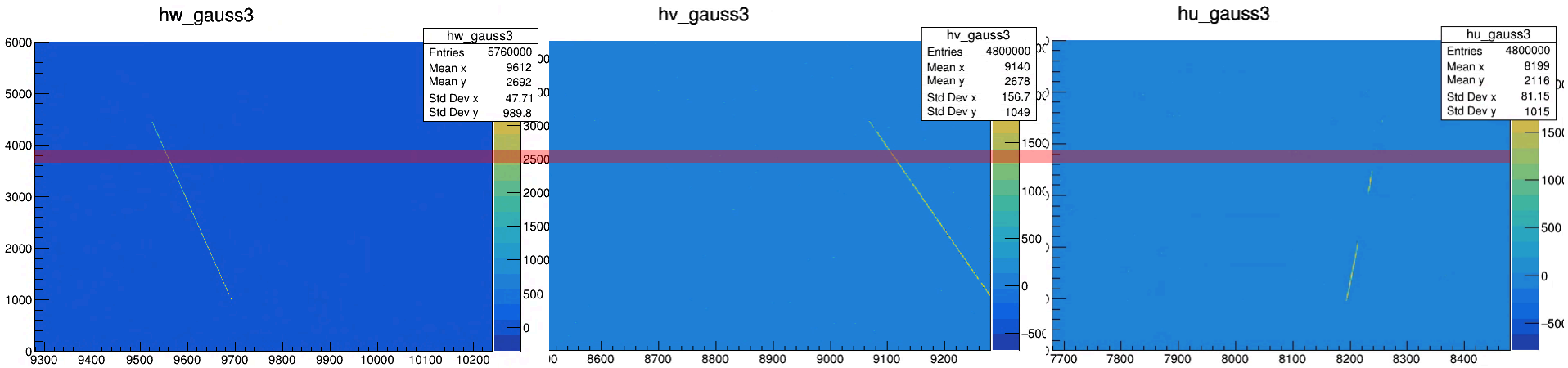


BROOKHAVEN NATIONAL LABORATORY

DUNE DEEP UNDERGROUND NEUTRINO EXPERIMENT

Multi-Plane (3-Plane Matching) Signal Processing – MP-SP

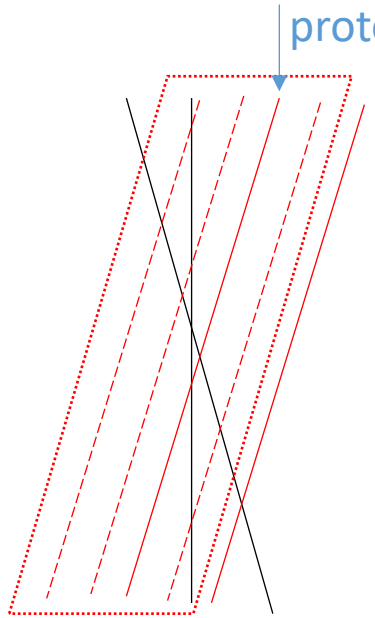
1, make time slices



Proposed by Xin Qian
- inspired by 'Imaging'

Fit for cases where signal initially observed but got removed as LF noise

2, Matching active (with ROI) wires in multiple planes



active wire in the time-slice :
ref. plane, target plane

in-active wire in the time-slice

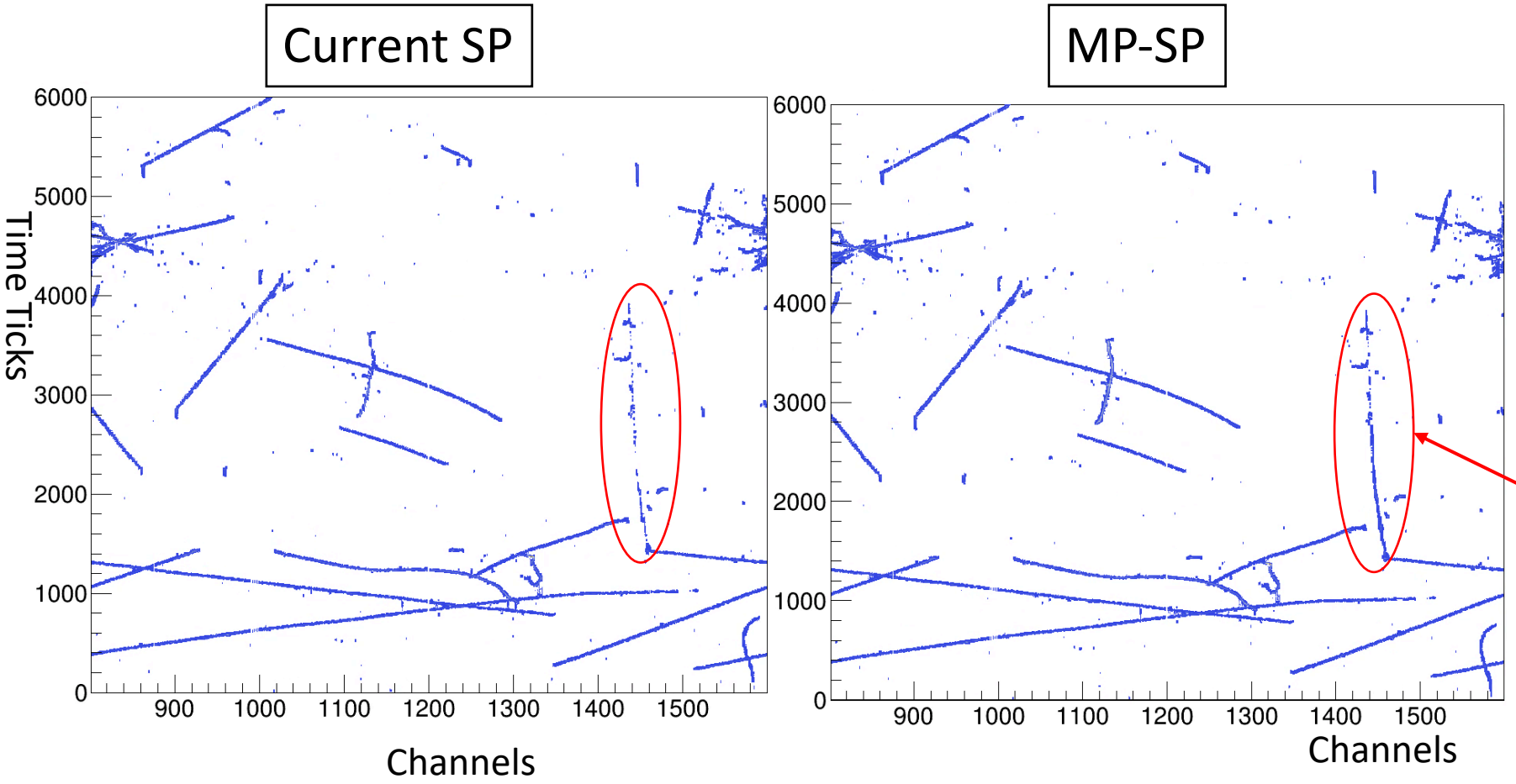
3, mark matched ROIs as 'protected'

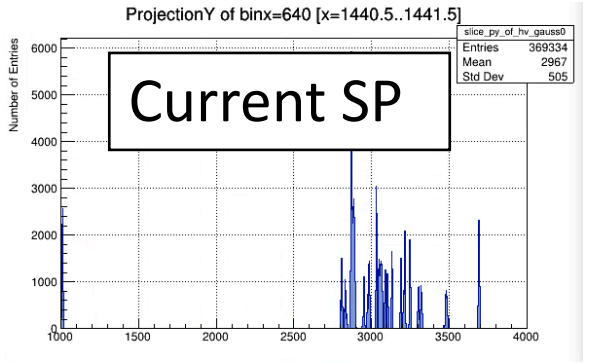
Fast projection realized using "RayGrid" tools developed by Brett:
https://github.com/WireCell/wire-cell-docs/blob/master/presentations/updates/20190321/latexmk-out/img.pdf

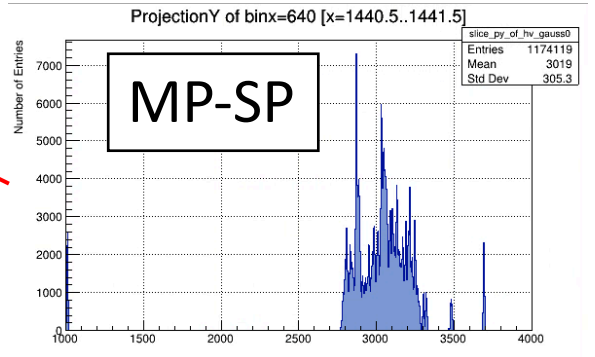
MP-SP Tested on ProtoDUNE Data

https://indico.fnal.gov/event/22240/contribution/6

Previous report in ProtoDUNE Sim/Reco Meeting

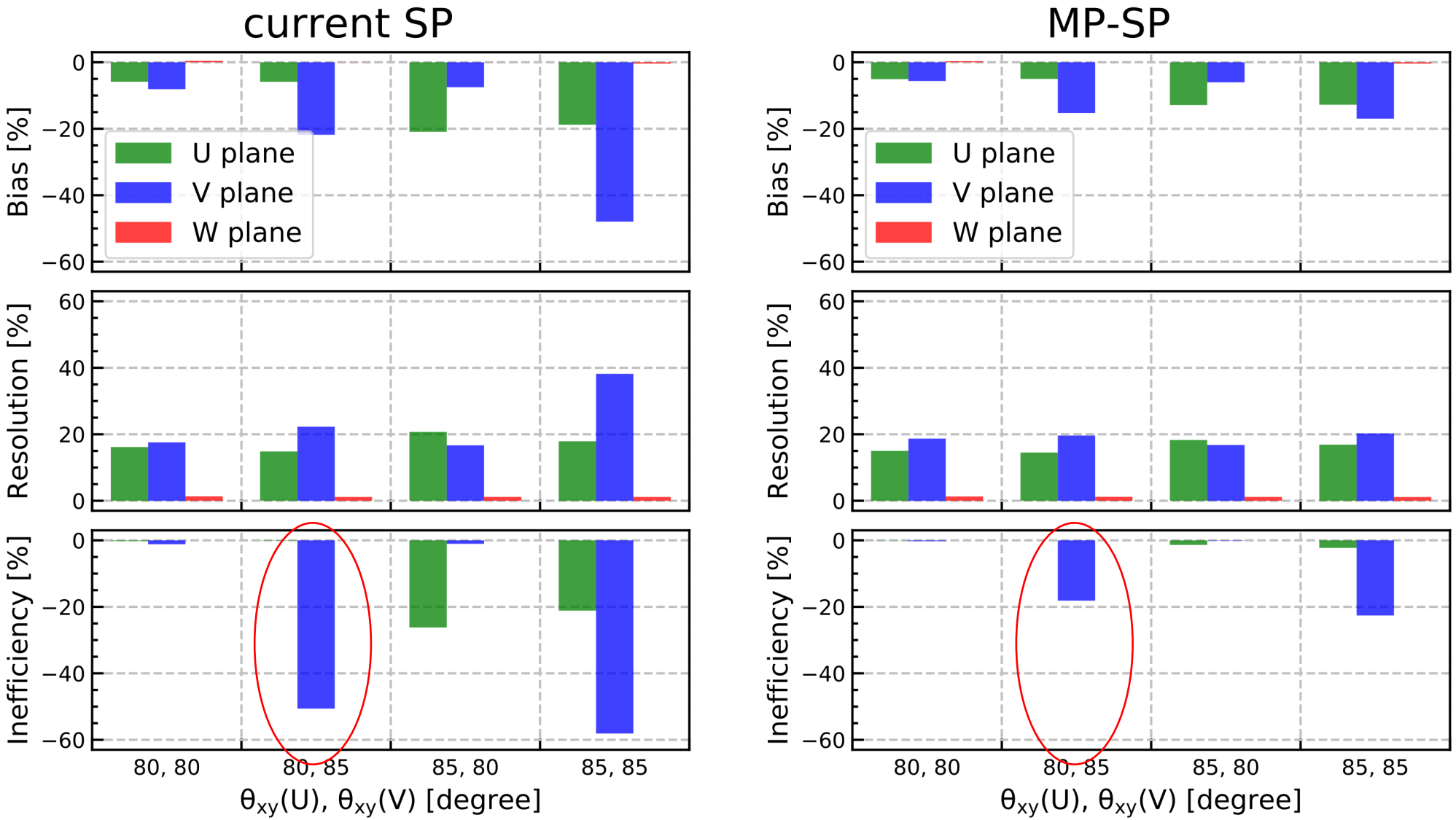




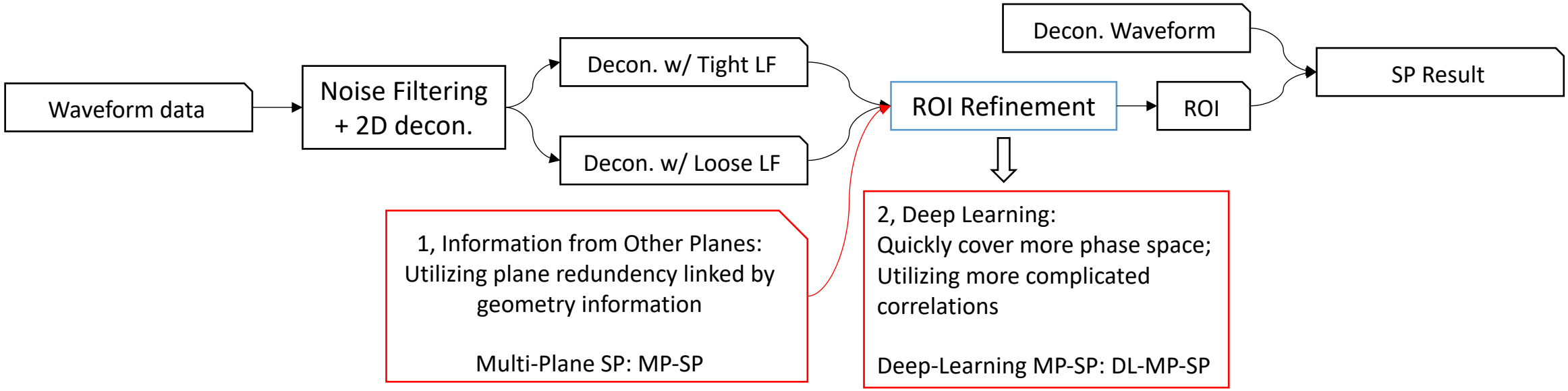


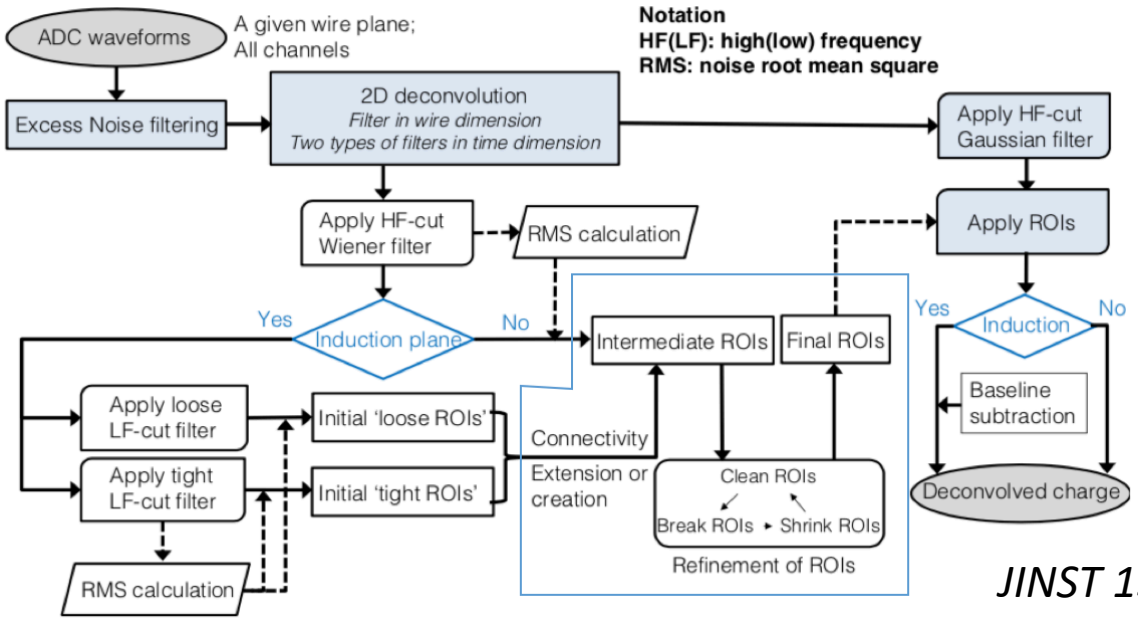
MP-SP Evaluated with Simulation

Work from Hongzhao Yu (SYSU)



Deep Learning "ROI Refinement" with Geometry Info.



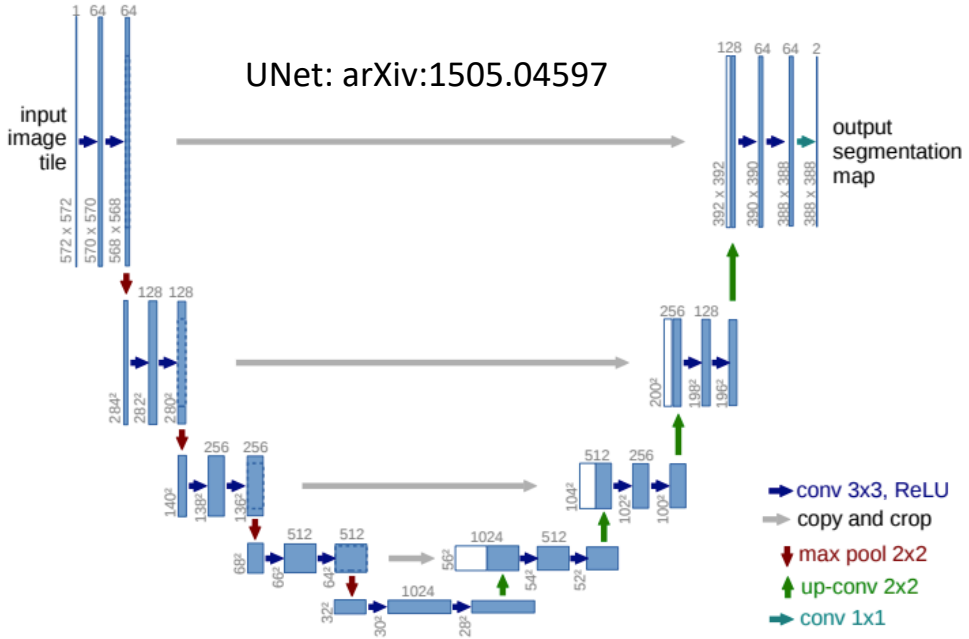


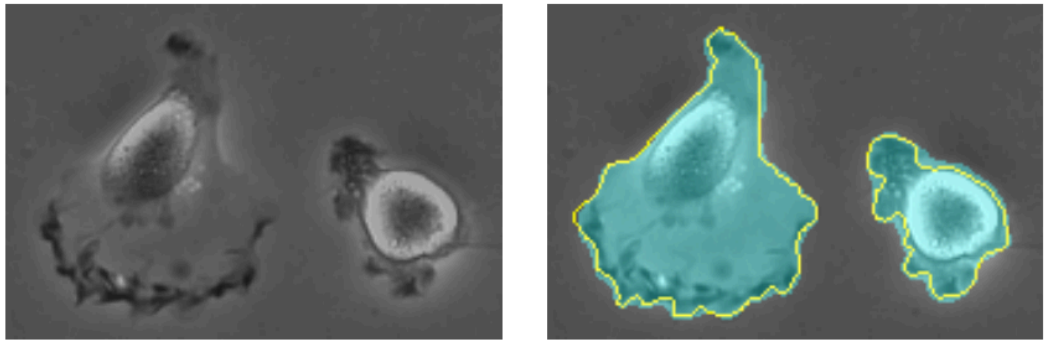
JINST 13 P07006 (2018)

DL-MP-SP: ROI Finding as Image Segmentation

- ROI finding as image segmentation – Idea emerged from discussions with Xin, Chao, Brett, Kazu, etc.

- Candidate Deep-Learning Model – UNet





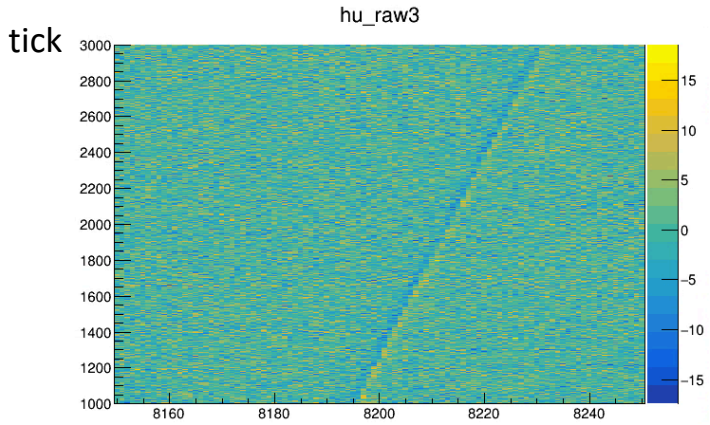
UNet: auto encoder-decoder + skip connections

- Output is sparse connected components
- Input and output are similar at leading order

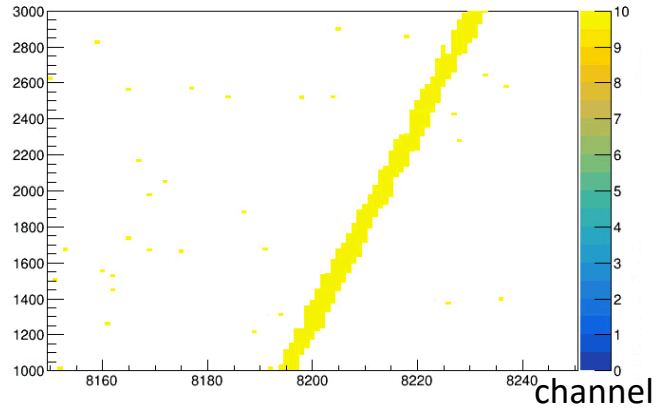
Forked a Pytorch implementation of the original U-Net

- https://github.com/HaiwangYu/Pytorch-UNet

Raw waveform vs channels

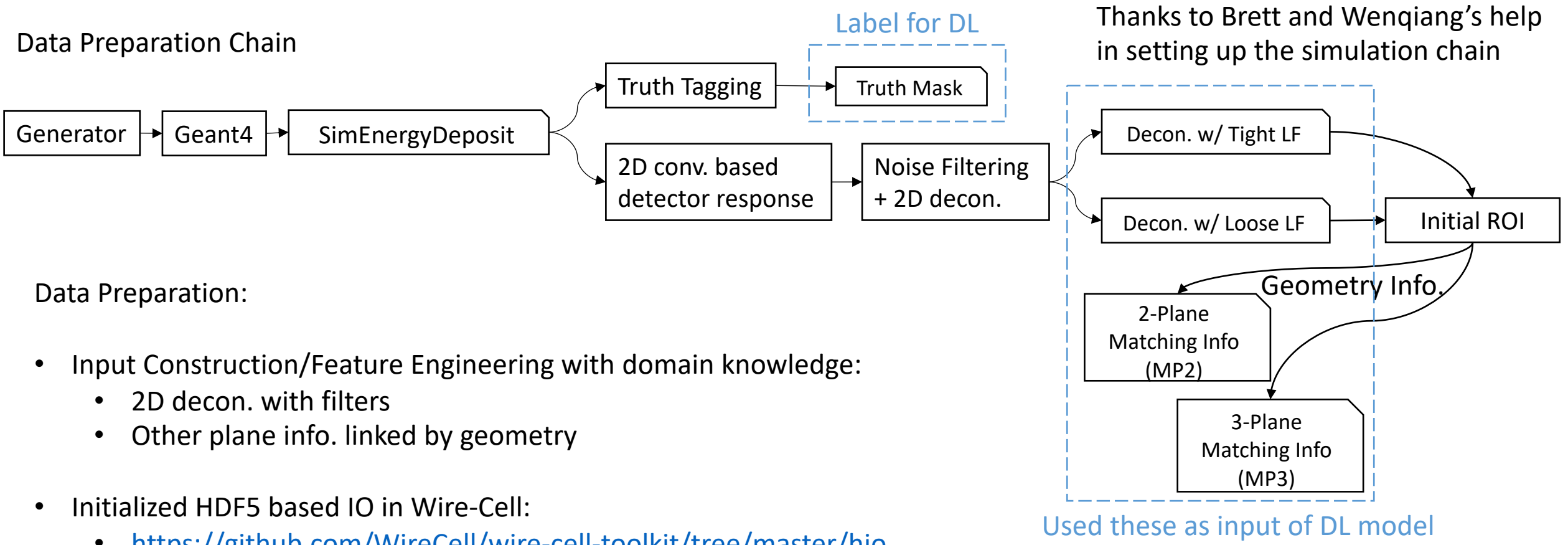


Tagged ROI



Data Preparation

Data Preparation Chain



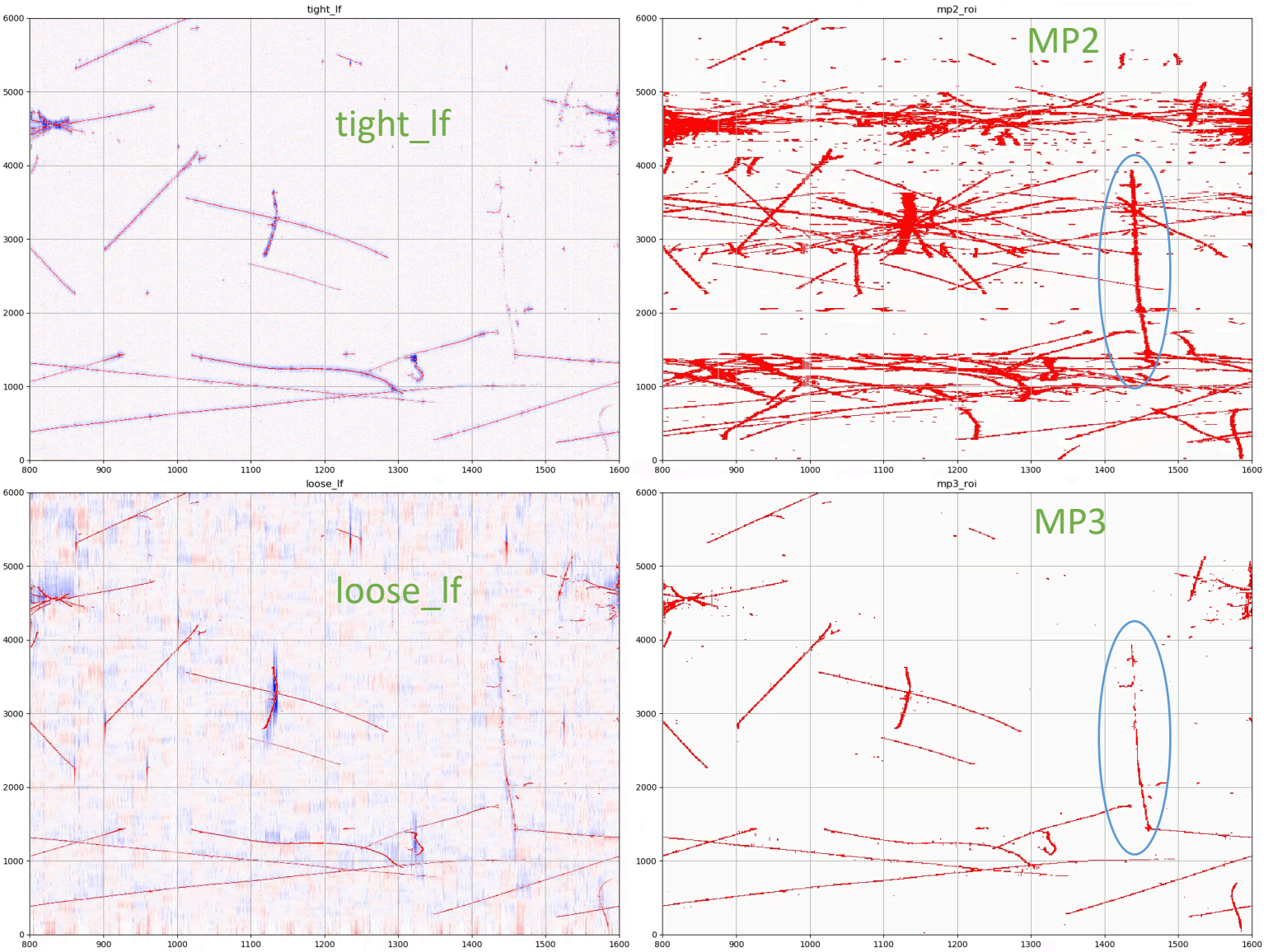
Data Preparation:

- Input Construction/Feature Engineering with domain knowledge:
  - 2D decon. with filters
  - Other plane info. linked by geometry

- Initialized HDF5 based IO in Wire-Cell:
  - https://github.com/WireCell/wire-cell-toolkit/tree/master/hio

ML Input from ProtoDUNE data

Input Candidates



`tight_lf`, `loose_lf`: 2D decon. with tight/loose low frequency filter

`MP2`, `MP3`: multi-plane 2 plane/3plane

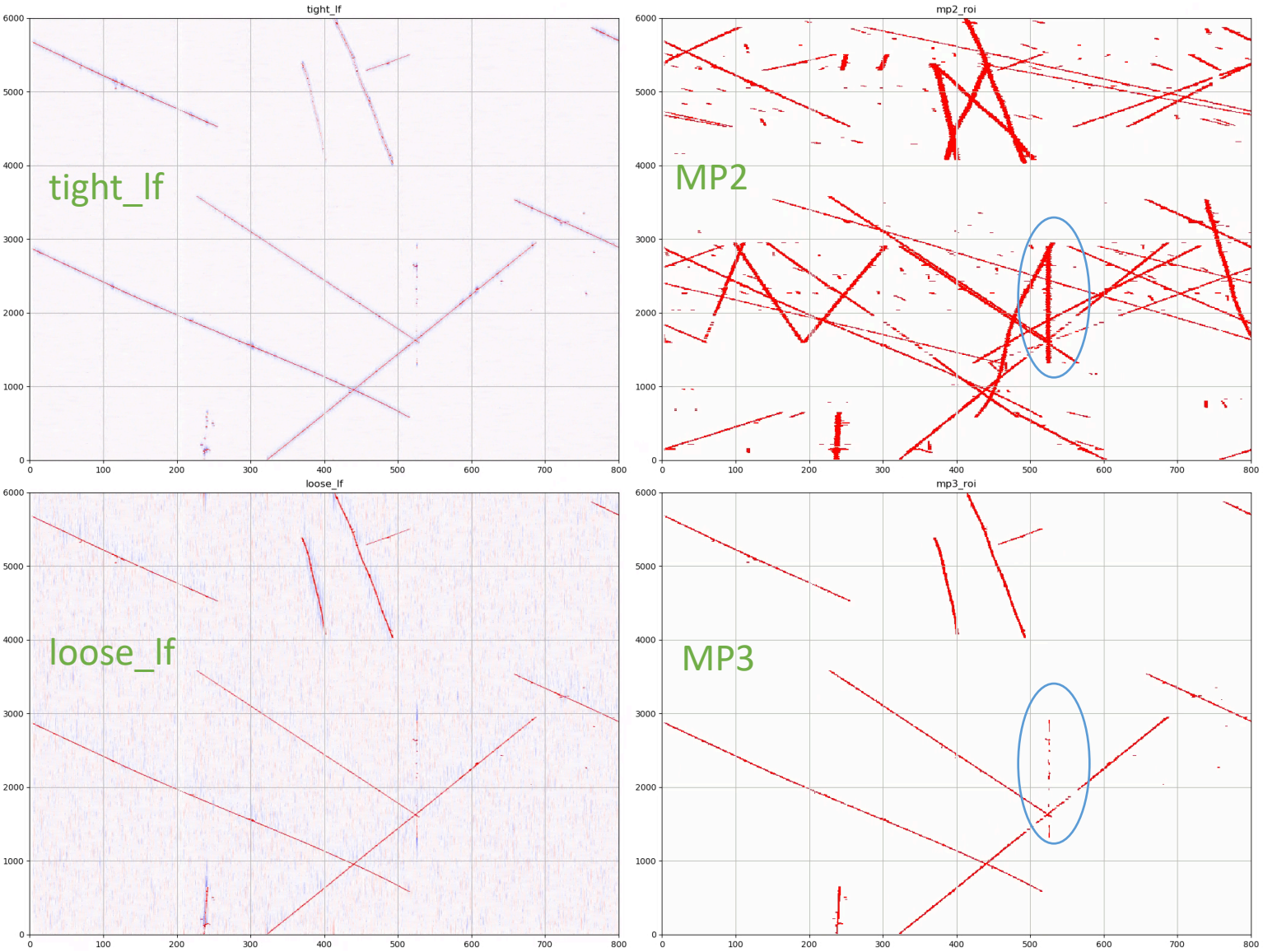
- 2 plane: for a given tick and wire check if any wires pairs from other 2 planes crossing – many ghosts
- 3plane: for a given tick and wire check if any wire trios from all 3 plane crossing – much fewer ghosts

Current best results are given by `loose_lf`, `mp2_roi` and `mp3_roi`. More experiments ongoing.

- loose_lf: decon. charge info
- MP2: low-purity high-efficiency info
- MP3: high-purity low-efficiency info

ML Input and label from ProtoDUNE simulation

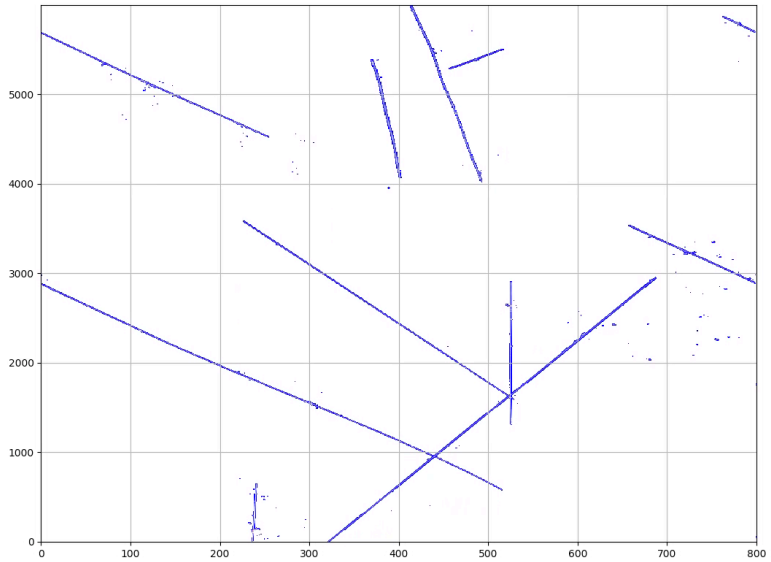
Input Candidates



Current noise simulation was based on real detector noise spectrum and random walk in complex frequency domain https://www.phy.bnl.gov/~diwan/talks/pedagogy/diwan-random-noise-basics-2.pdf

Truth labelling was done by rasterization of charge depos – 2D gaussian response

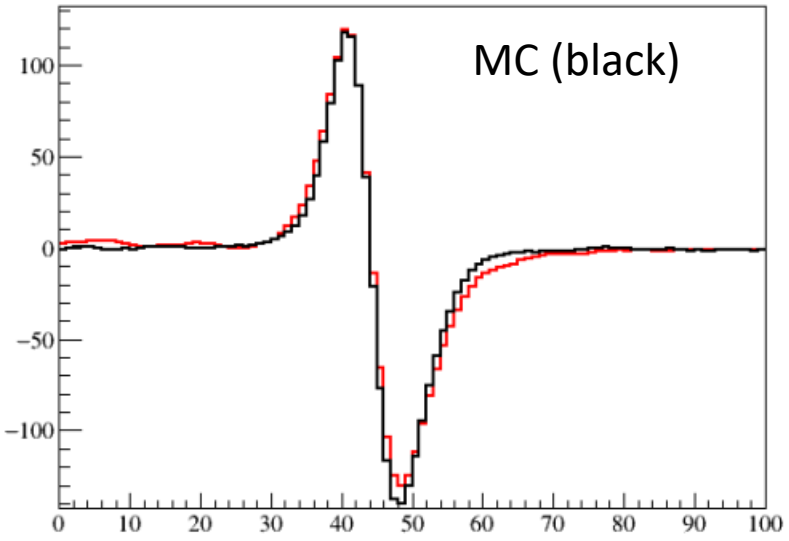
Label



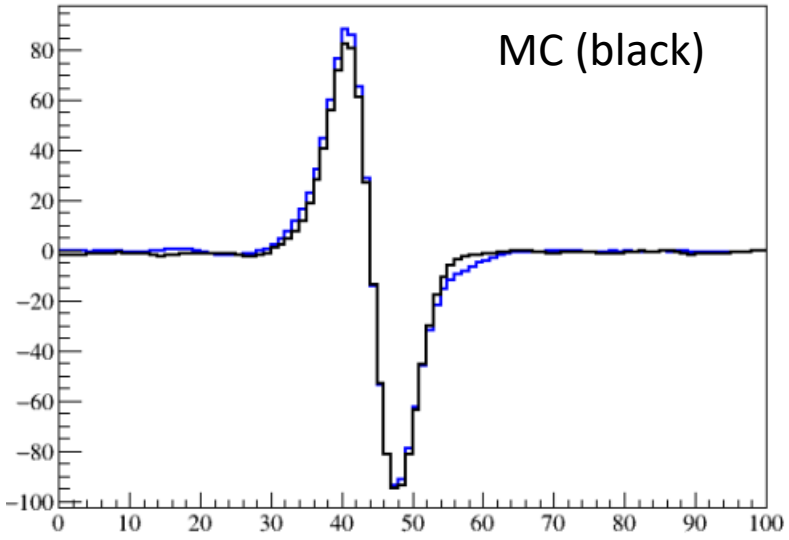
Signal Simulation Validation

From Wenqiang Gu's talk:
https://indico.bnl.gov/event/7024/contributions/32749/

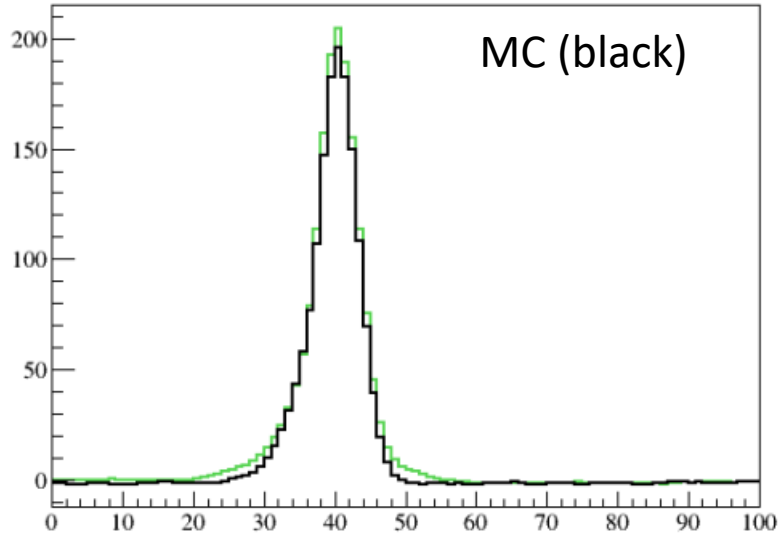
Average waveform: data vs. MC



U plane
MC scaling: 1.6
Tick: [570, 970]
Ch: [8050, 8200]



V plane
MC scaling: 1.6
Tick: [570, 970]
Ch: [8950, 9200]

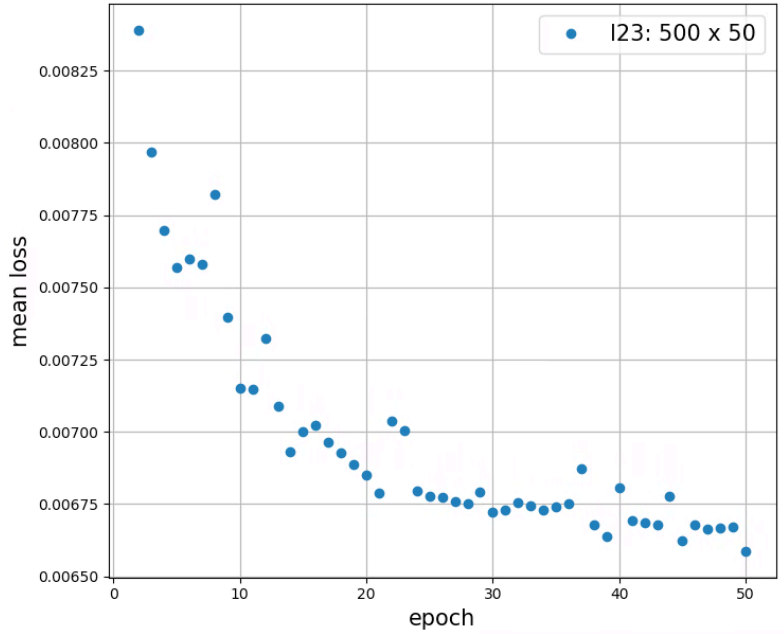


W plane
MC scaling: 1.6
Tick: [570, 970]
Ch: [9500, 9750]

Training

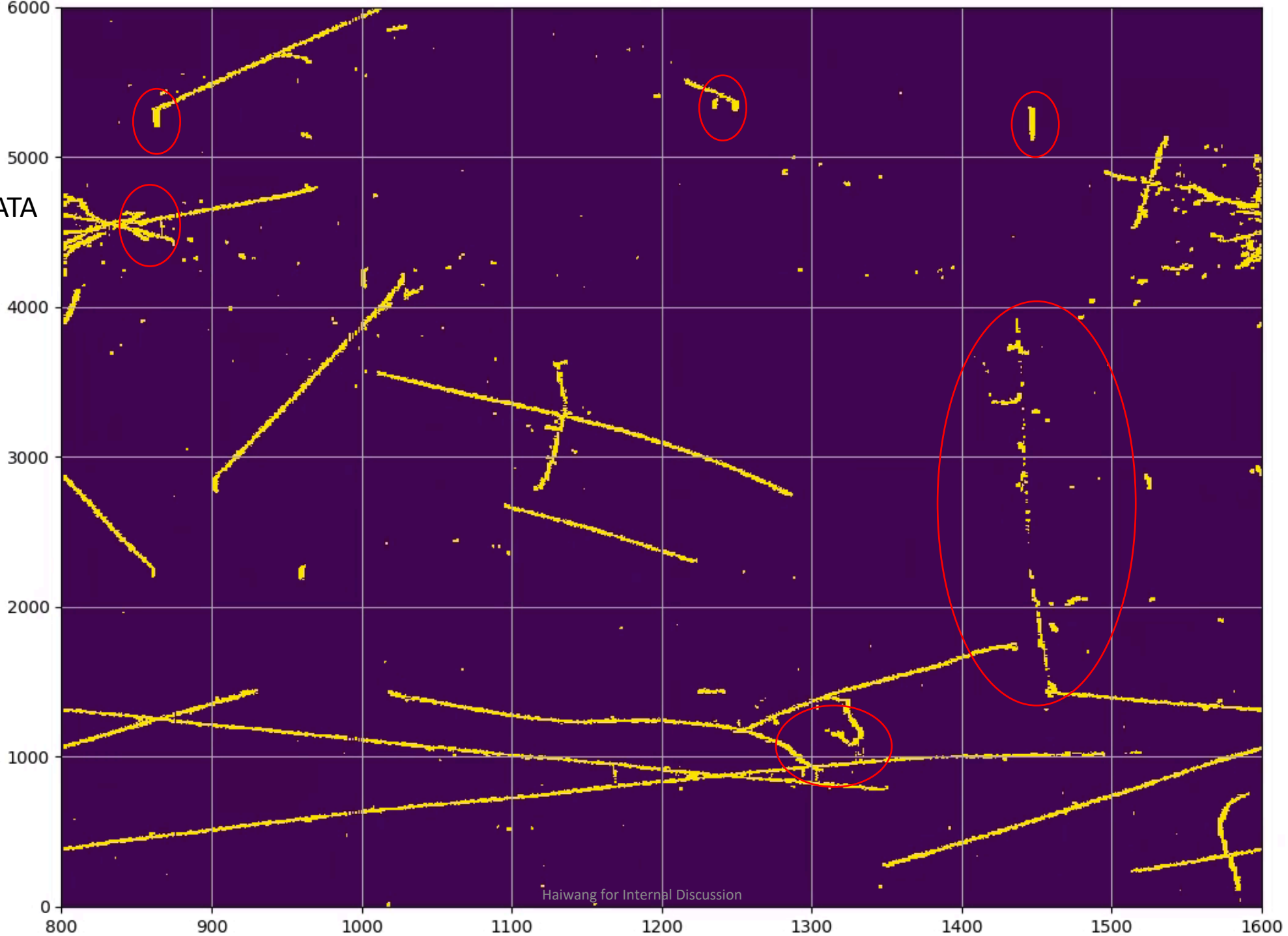
Platform: I9-9900K, 32 GB memory, Nvidia GTX 2080 Ti 11GB, Samsung 970 500GB NVMe SSD
500 APA samples using cosmic generator

	Unit Time	Total Time
Generator	2 sec/event	0.3 hour/500events
G4	23 sec/event	3.2 hour/500events
detector response, truth tagging and waveform preprocessing	68 sec/APA	9.4 hour/500 APA
Network training	6 min/epoch (1 epoch: 1 iteration of 500 APA sample)	5 hour/50epoch
Sum		17.9 hour



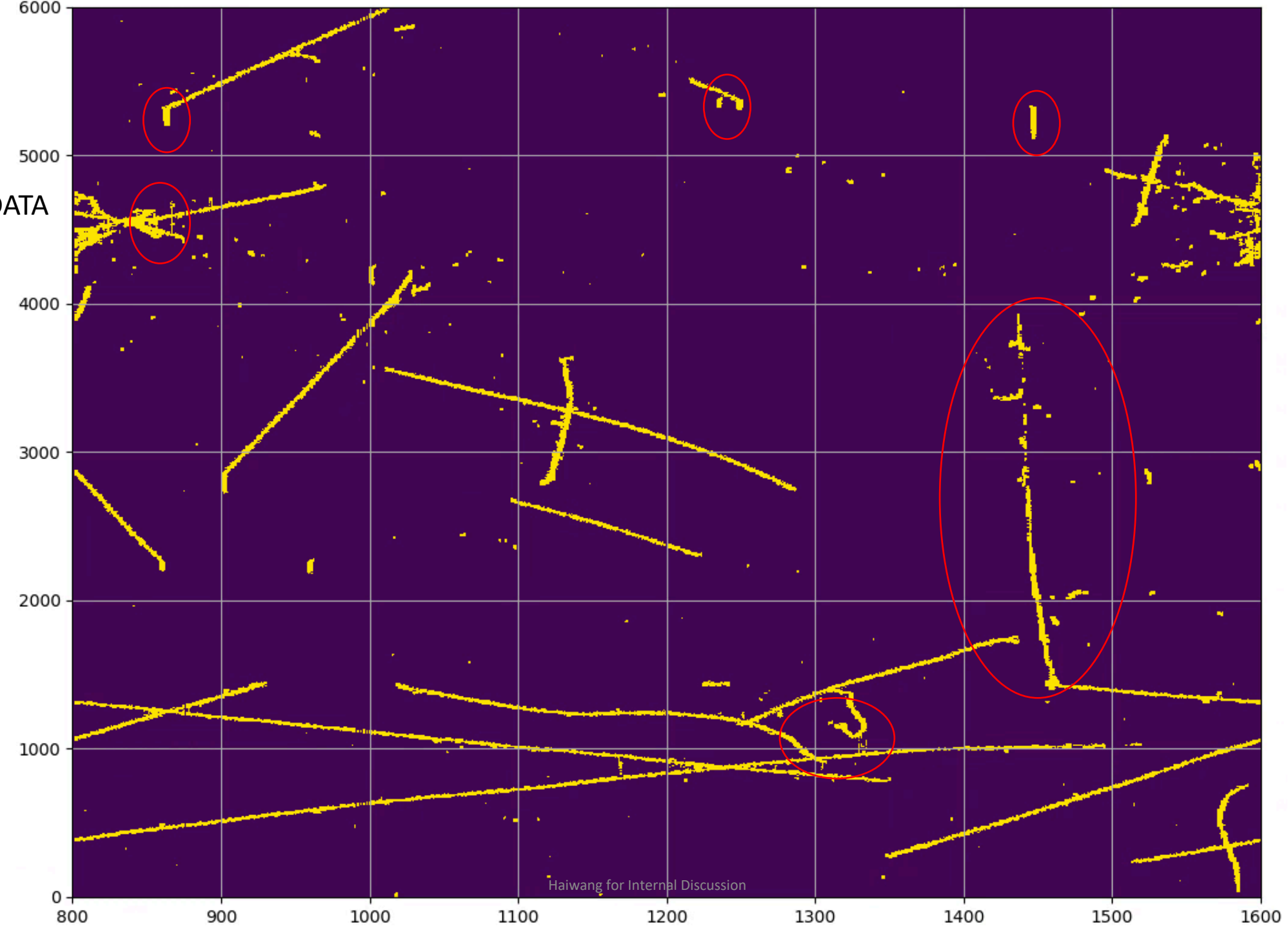
Traditional SP

ProtoDUNE DATA
run: 5145
subRun: 1
event: 26918
V plane



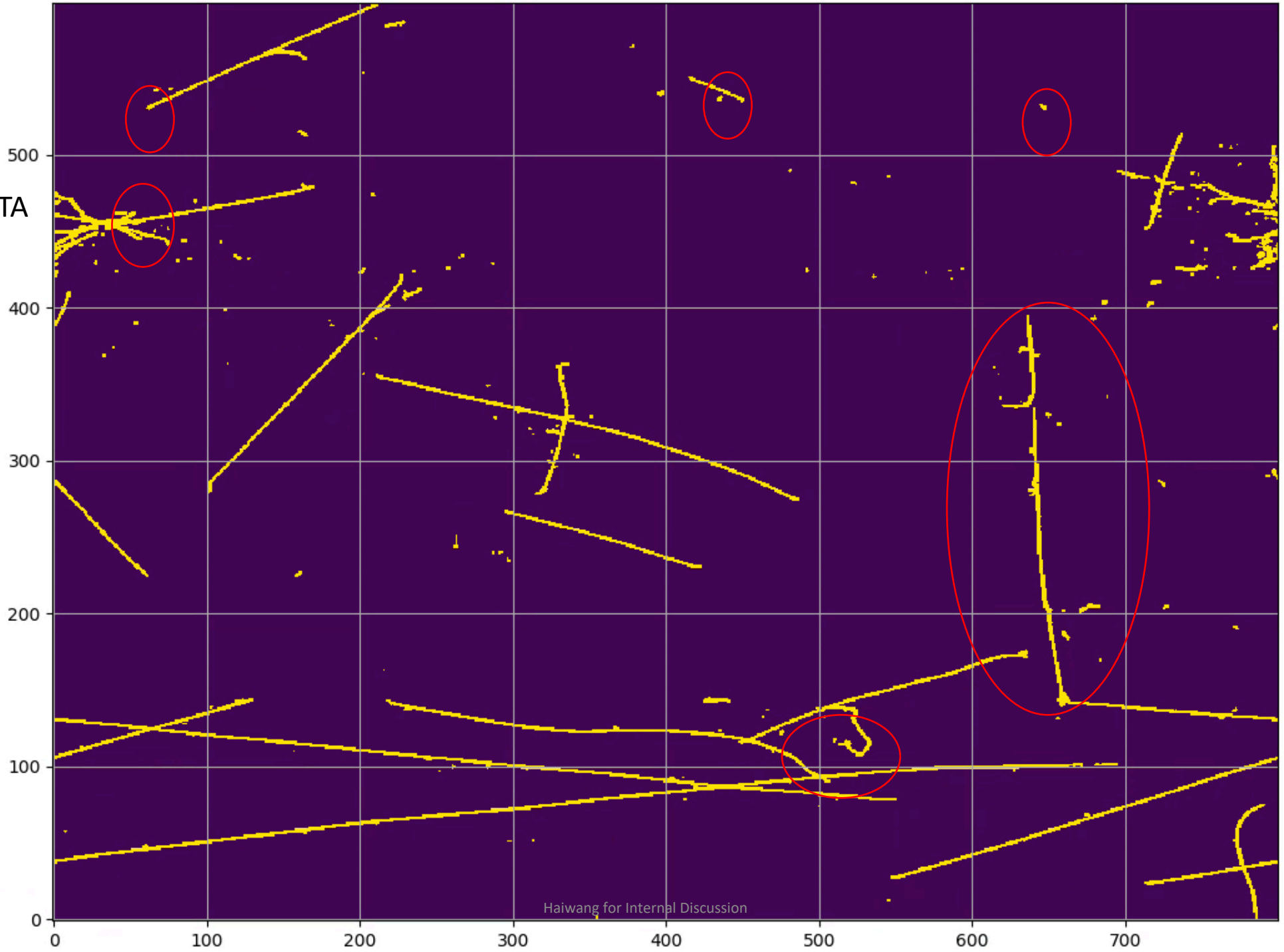
MP-SP

ProtoDUNE DATA
run: 5145
subRun: 1
event: 26918
V plane



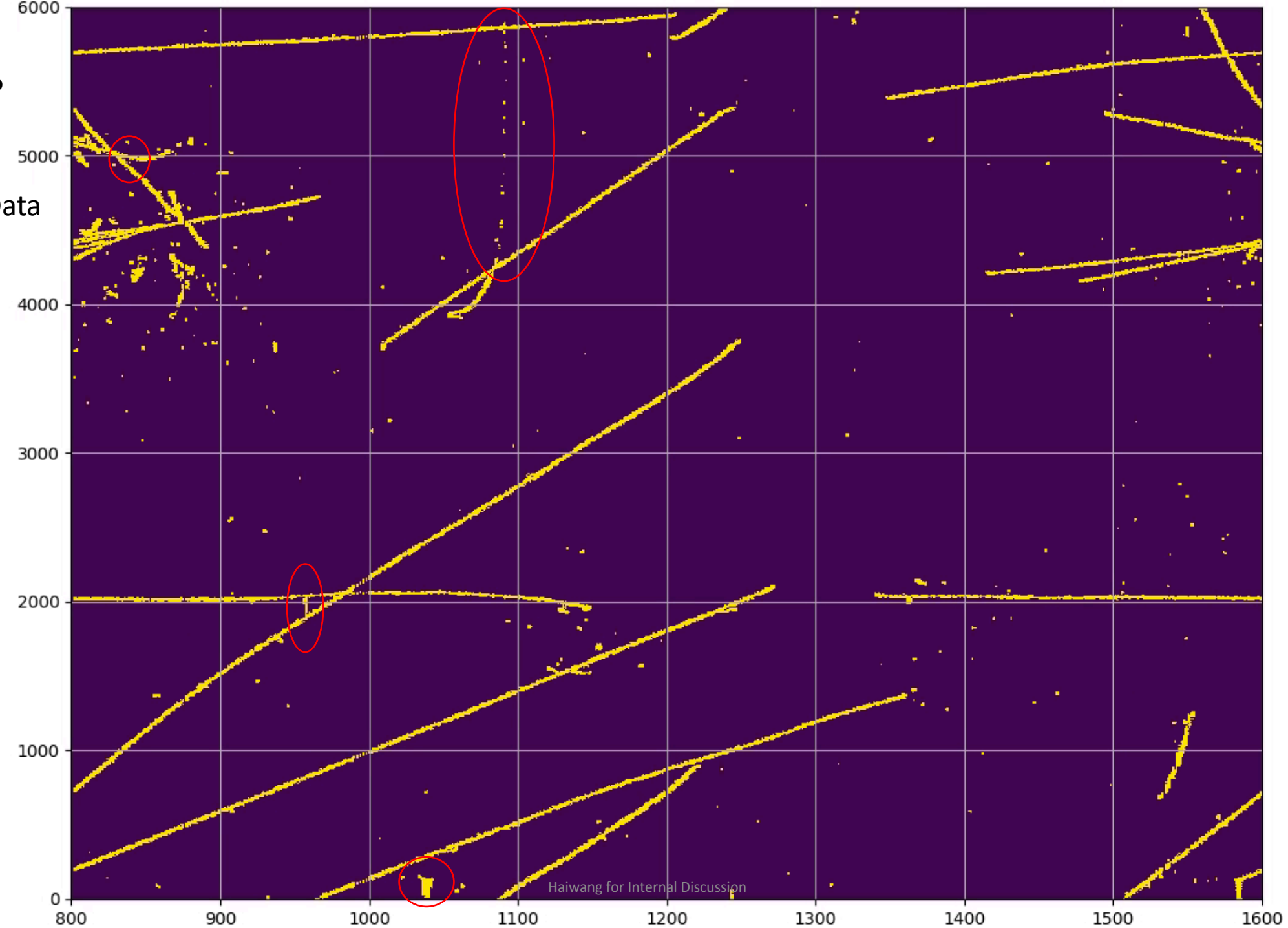
DL-MP-SP

ProtoDUNE DATA
run: 5145
subRun: 1
event: 26918
V plane



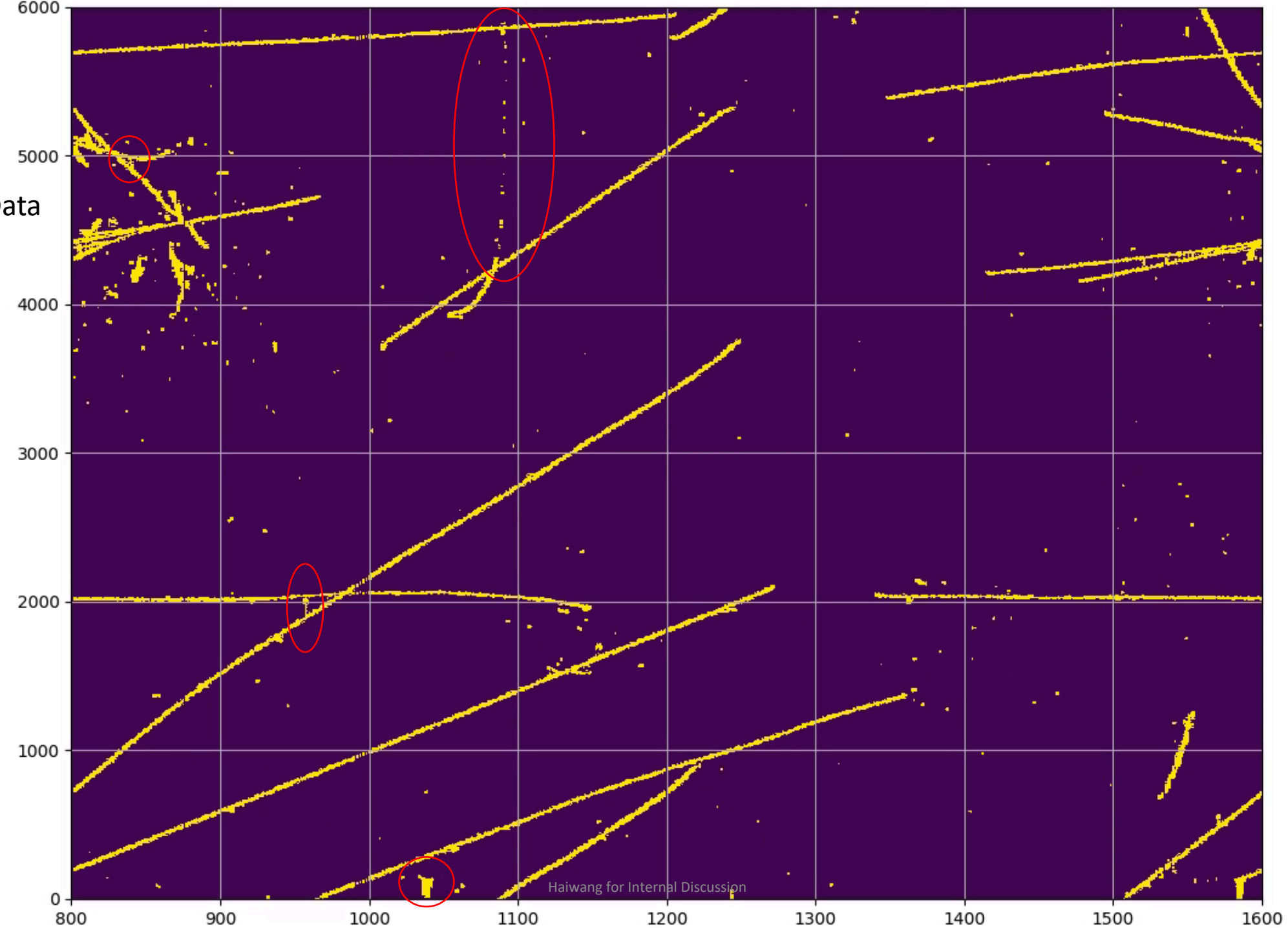
Traditional SP

ProtoDUNE Data
run: 5145
subRun: 1
event: 26925
V plane



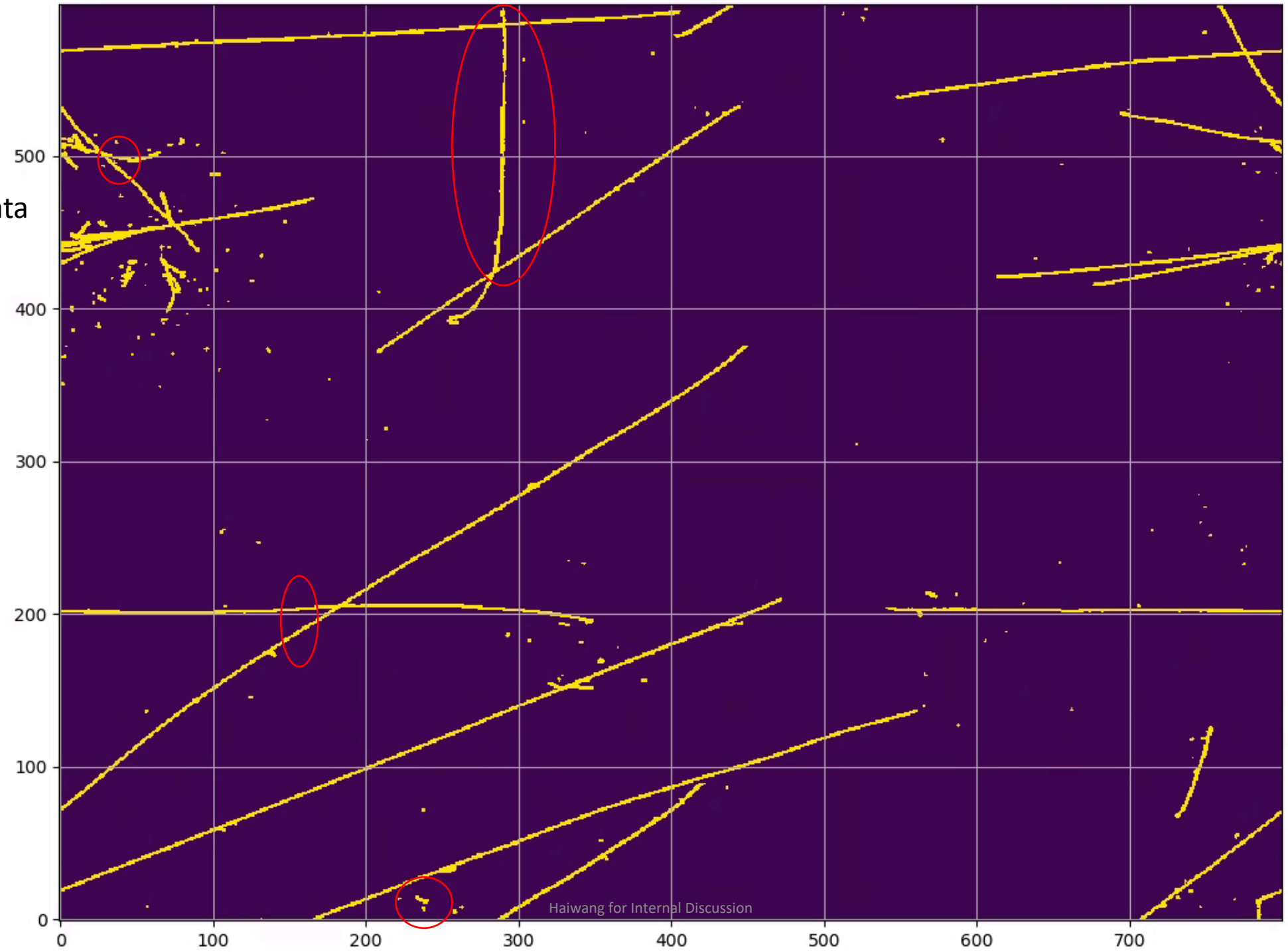
MP-SP

ProtoDUNE Data
run: 5145
subRun: 1
event: 26925
V plane



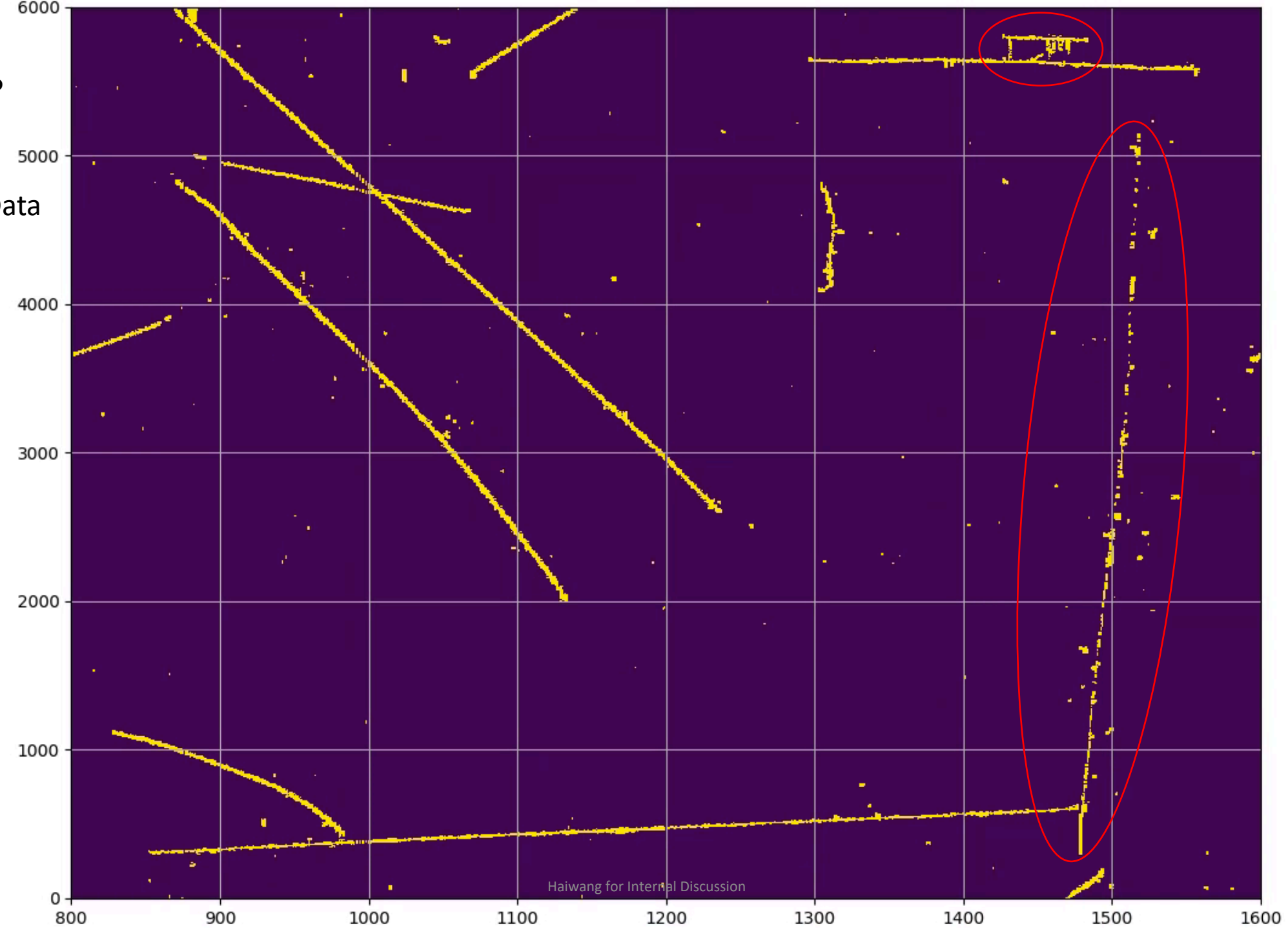
DL-MP-SP

ProtoDUNE Data
run: 5145
subRun: 1
event: 26925
V plane



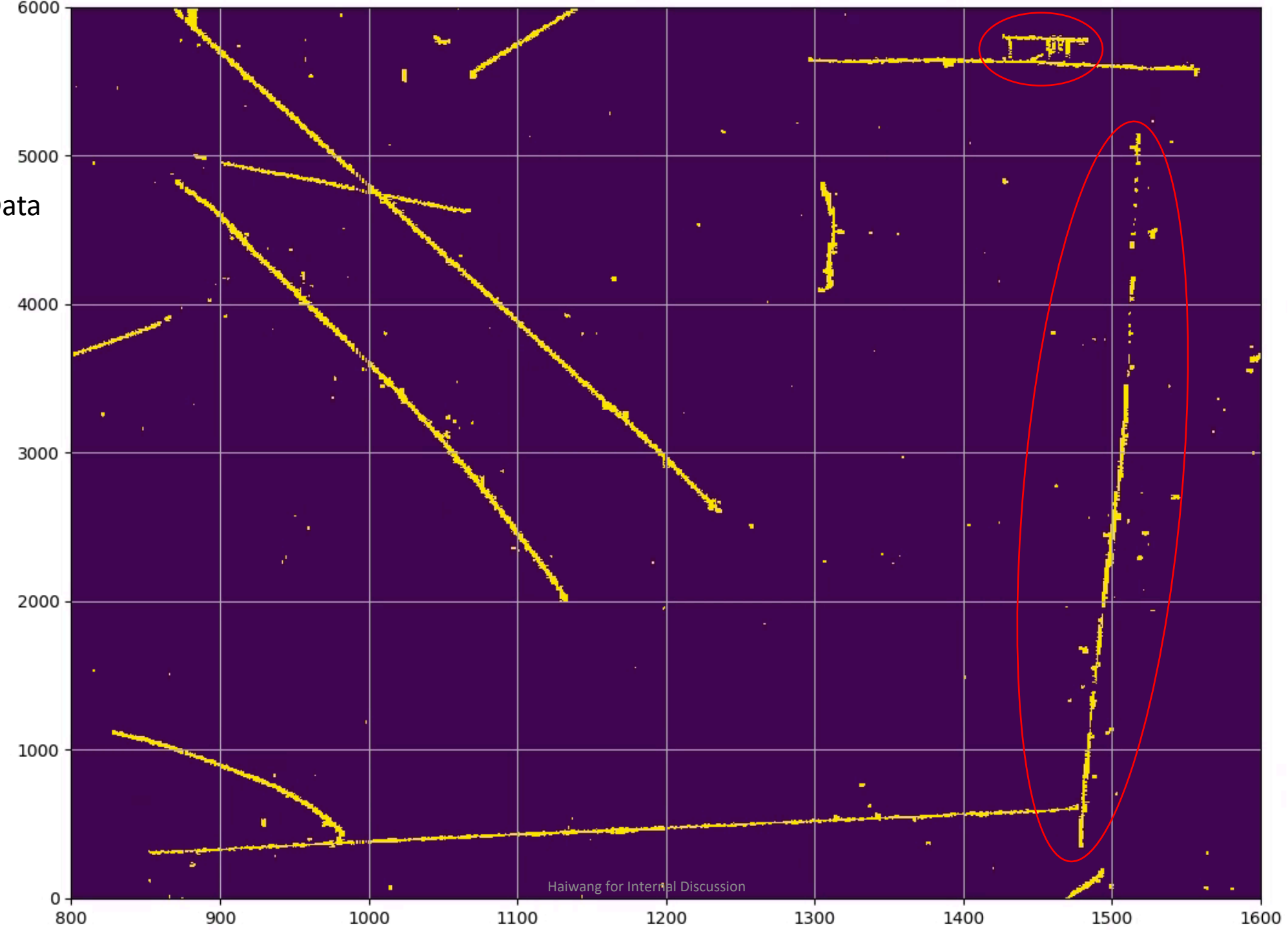
Traditional SP

ProtoDUNE Data
run: 5145
subRun: 1
event: 26945
V plane



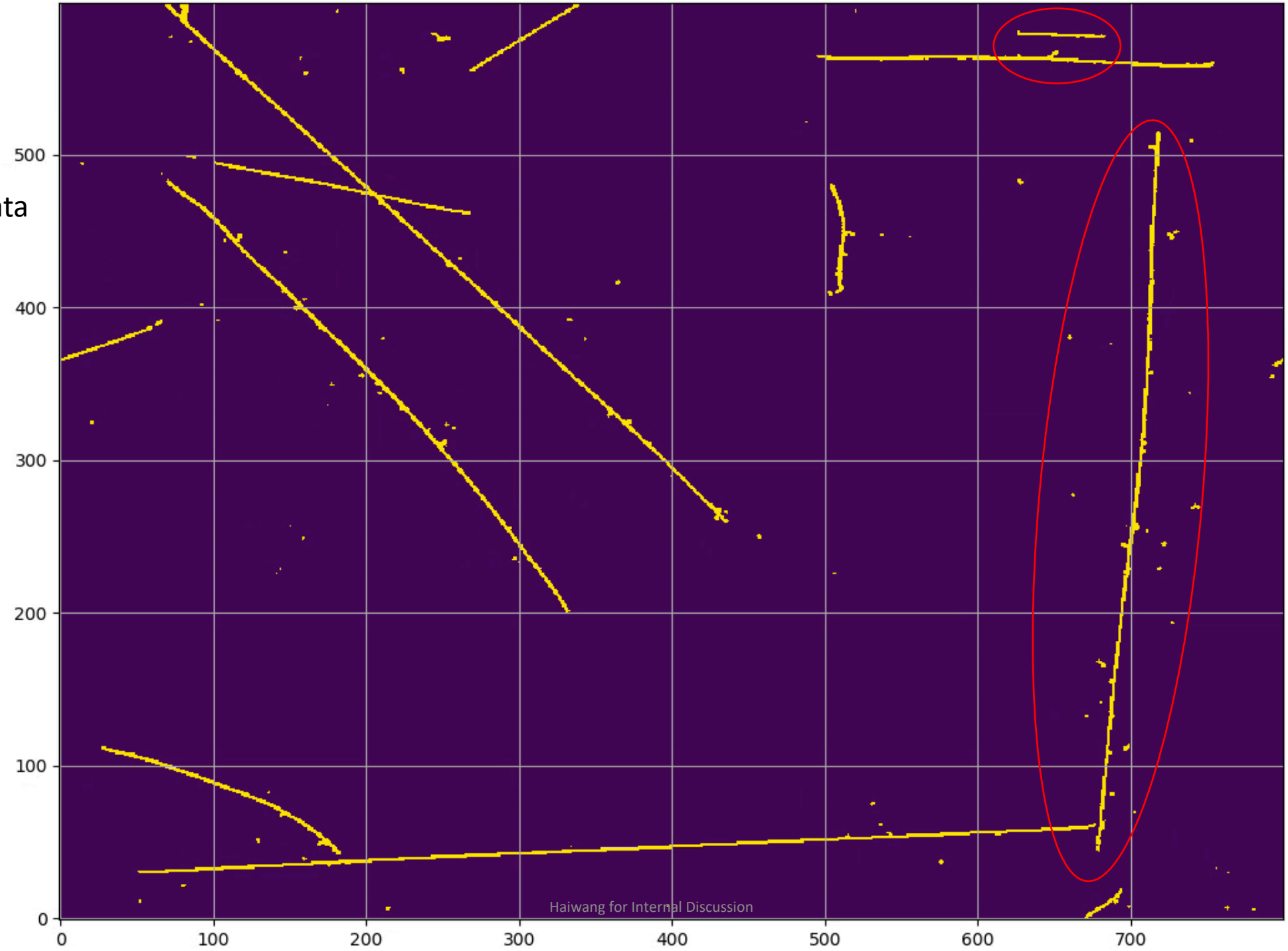
MP-SP

ProtoDUNE Data
run: 5145
subRun: 1
event: 26945
V plane



DL-MP-SP

ProtoDUNE Data
run: 5145
subRun: 1
event: 26945
V plane



DL-MP-SP Speed Estimation

Current ROI Refinement: 1.3 sec/APA

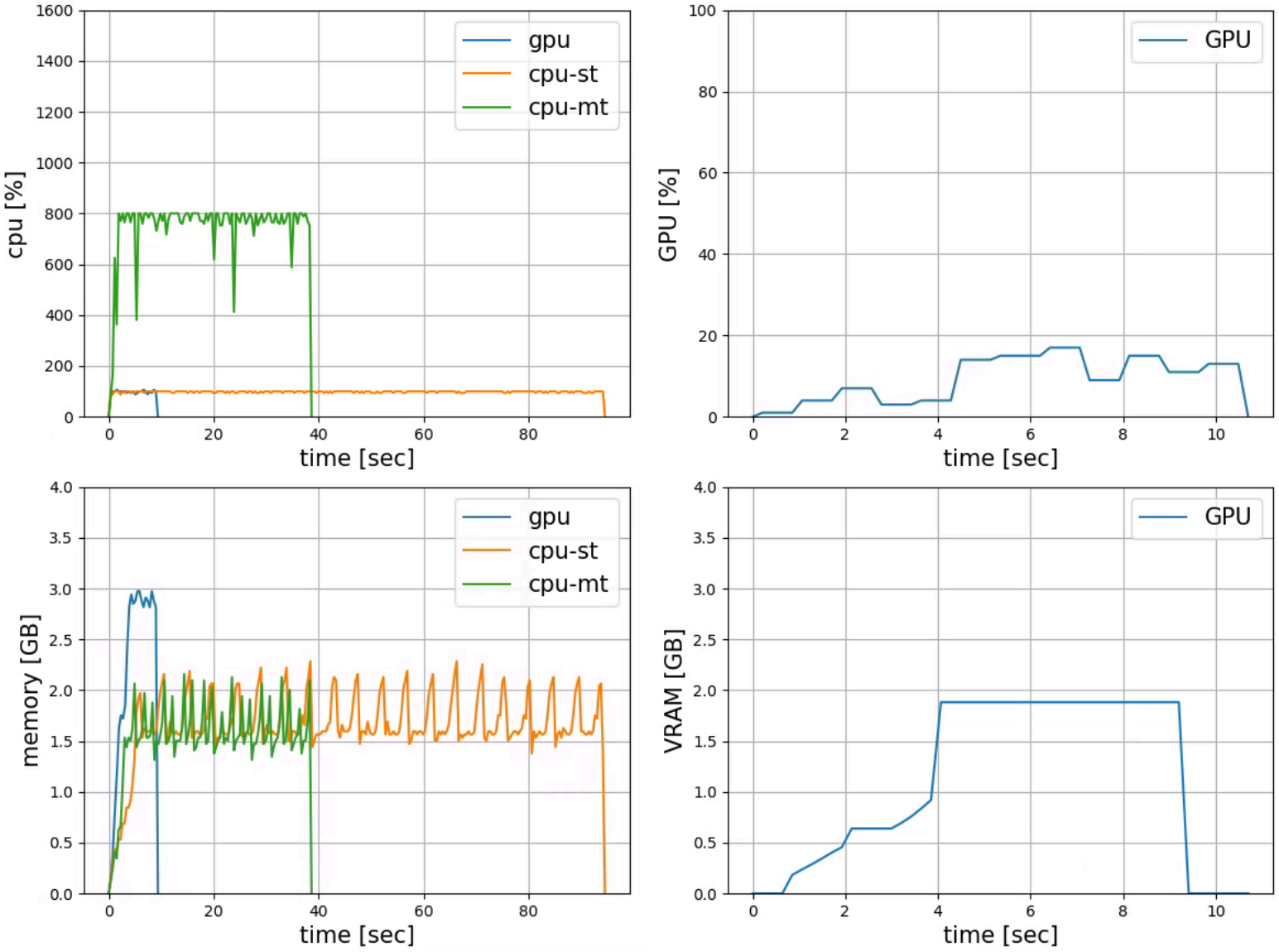
DL Model Prediction:
CPU-ST: 18 sec/APA
CPU-MT: 8 sec/APA
GPU-CPU-ST: 2 sec/APA

ST: Single Thread
MT: Multi-Thread

Current Network using Pytorch is NOT very fast on CPU

Other Neural Network? TensorFlow?

DL Model Prediction for V plane for 10 APAs



Summary

We are trying to improve LArTPC signal processing by introducing geometry information and Deep Learning

Current DL model trained with relatively small simulation sample already showed promising results on ProtoDUNE-SP real data.
More developments and evaluations are on going

Availability in Wire-Cell toolkit:

- MP-SP: ready to use in next release of Wire-Cell
- DL-MP-SP: integration with Wire-Cell ongoing