Wire-Cell for DUNE-FD

Haiwang Yu (BNL) on behalf of the DUNE collaboration

The Second Wire-Cell Reconstruction Summit

Hosted by Brookhaven National Laboratory
The workshop will held as a hybrid event on April 10–12, 2024
https://www.bnl.gov/wirecellsummit/

Wire-Cell

Brookhaven National Laboratory

U.S. DEPARTMENT OF ENERGY

@BrookhavenLab

Outline

Current Wire-Cell components in DUNE-FD

- Sim, SigProc
    - In dev: ~~GPU acceleration,~~ DNN-ROI, LS4GAN
- Imaging, clustering
    - In dev: IO

Specific needs for DUNE-FD:

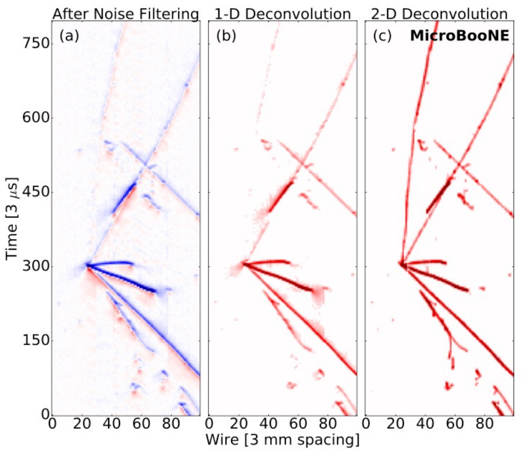
- VD field response calculation
- APA-wise sparse signal
- data/MC handling for the full 10kt geometries
- the challenges with radiological backgrounds
- multiple configurations (HD, VD, workspaces)

Discussion

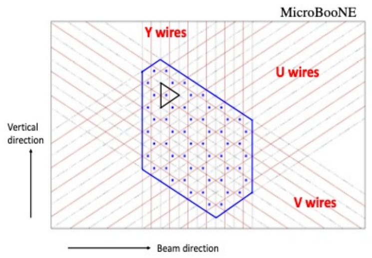
Brookhaven National Laboratory

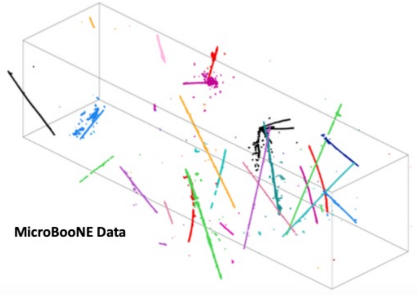
Wire-Cell Event Reconstruction

TPC simulation noise filtering signal processing	3D imaging clustering charge-light matching	3D trajectory & dQ/dx fitting cosmic muon tagger	multi-track fitting DL-3D vertexing particle identification

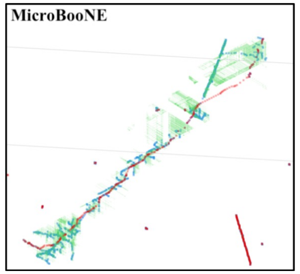


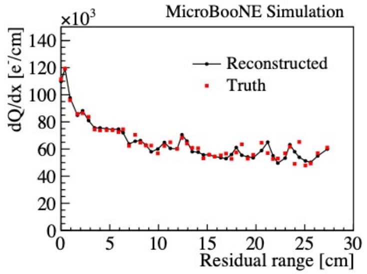
JINST 12 P08003 (2017)
JINST 13 P07006 (2018)
JINST 13 P07007 (2018)
JINST 16 P01036 (2020)



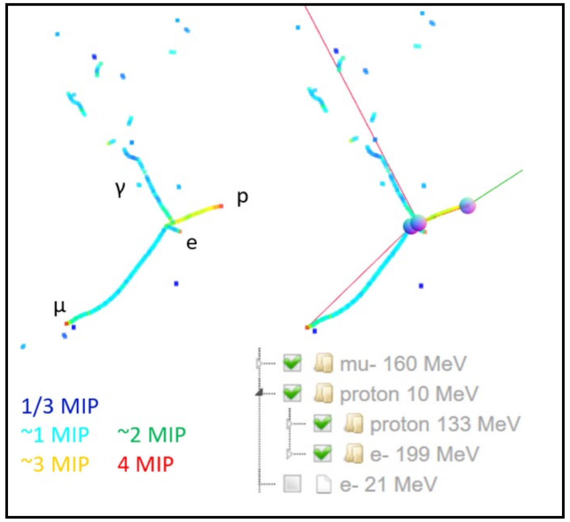


JINST 13 P05032 (2018)
JINST 16 P06043 (2021)





Phys. Rev. Applied 15, 064071 (2021)



JINST 17 P01037 (2022)

Brookhaven National Laboratory

DNN-ROI – better results with easier tunning

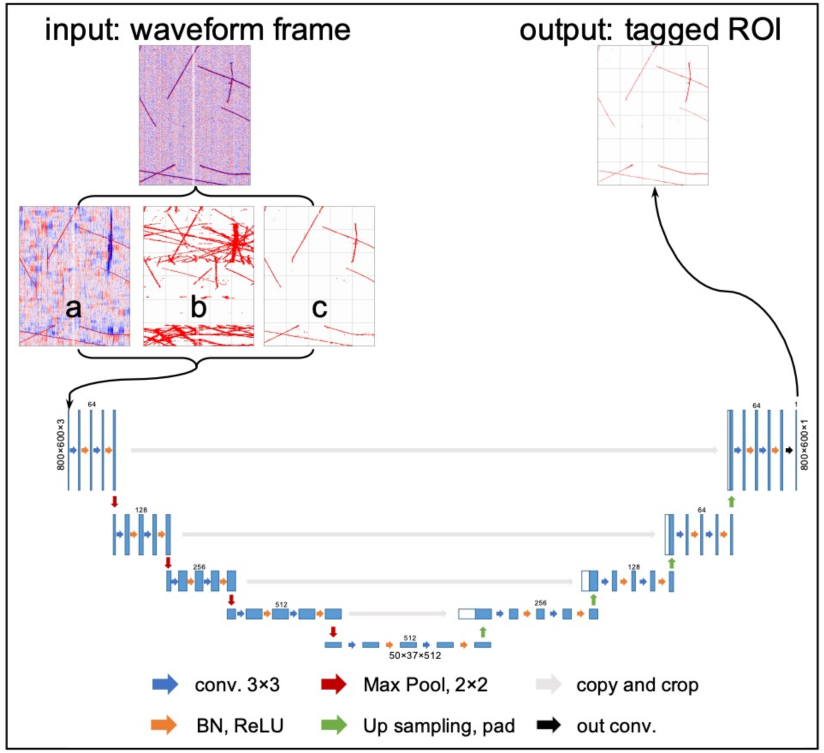
https://arxiv.org/pdf/2007.12743.pdf

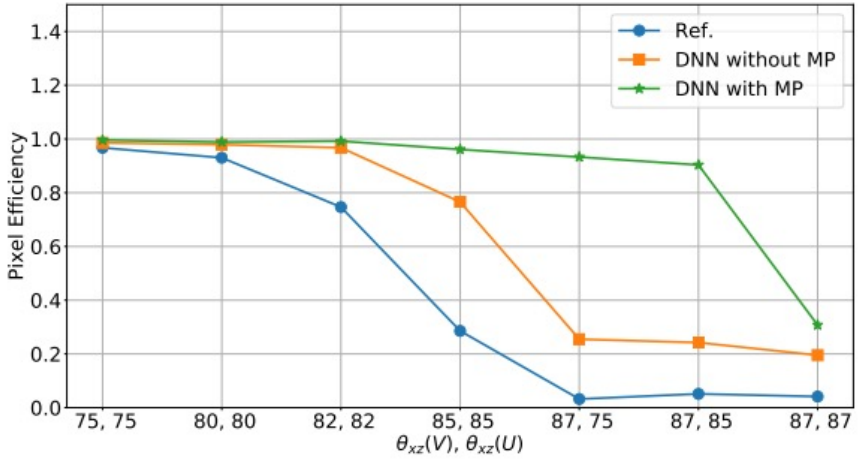
- ref. Wenqiang and Avinay/Moon's talk
  - Tests with real data in PDHD/VD and SBND

- Implementation efforts:
  - Main issue: to match truth width with decon width
    - new "DepoFluxSplat"
    - "Morse sim" to extract the extra smearing
  - Automated evaluation
    - "spdir metric": #287

- Becomes more important coupled with "Prompt Processing", ref. Kirby's talk

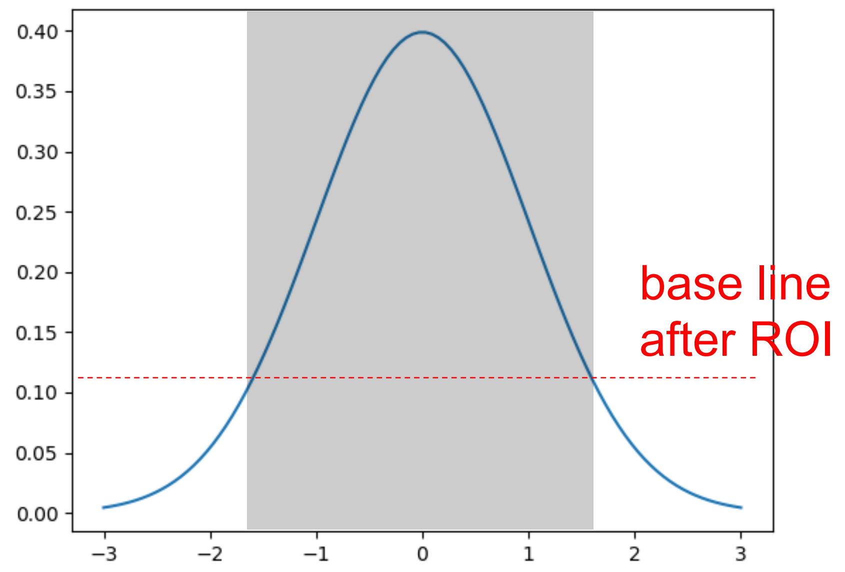
https://arxiv.org/pdf/2007.12743.pdf



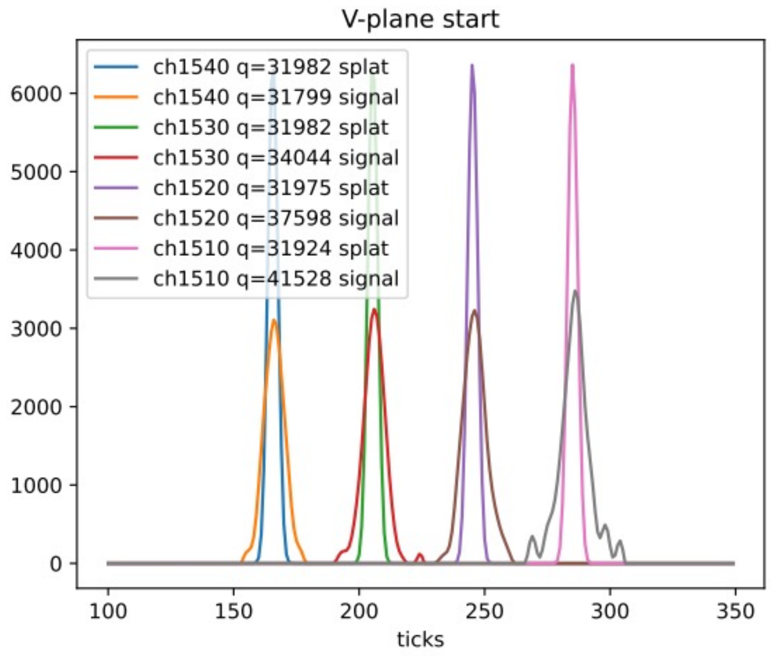


Extra truth smearing

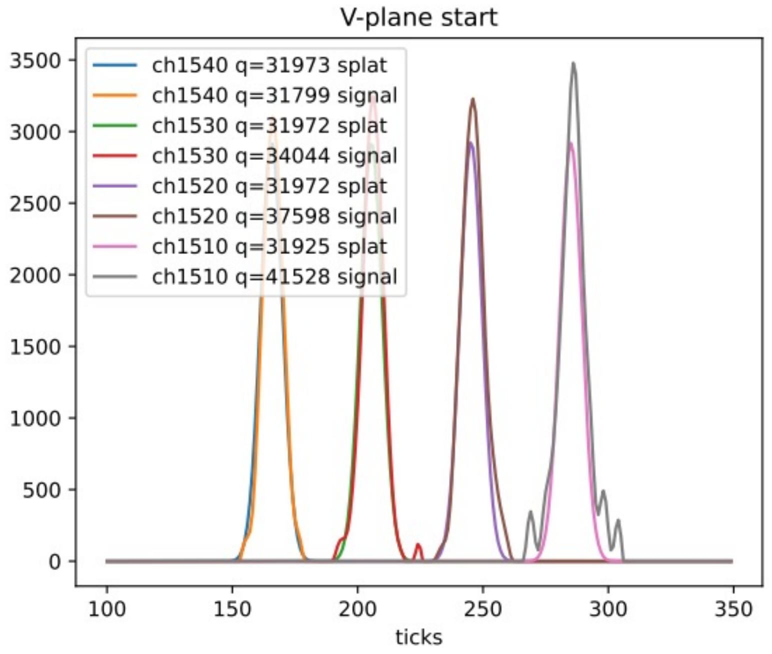
Narrower ROI leads to biased charge



No extra smearing



With extra smearing

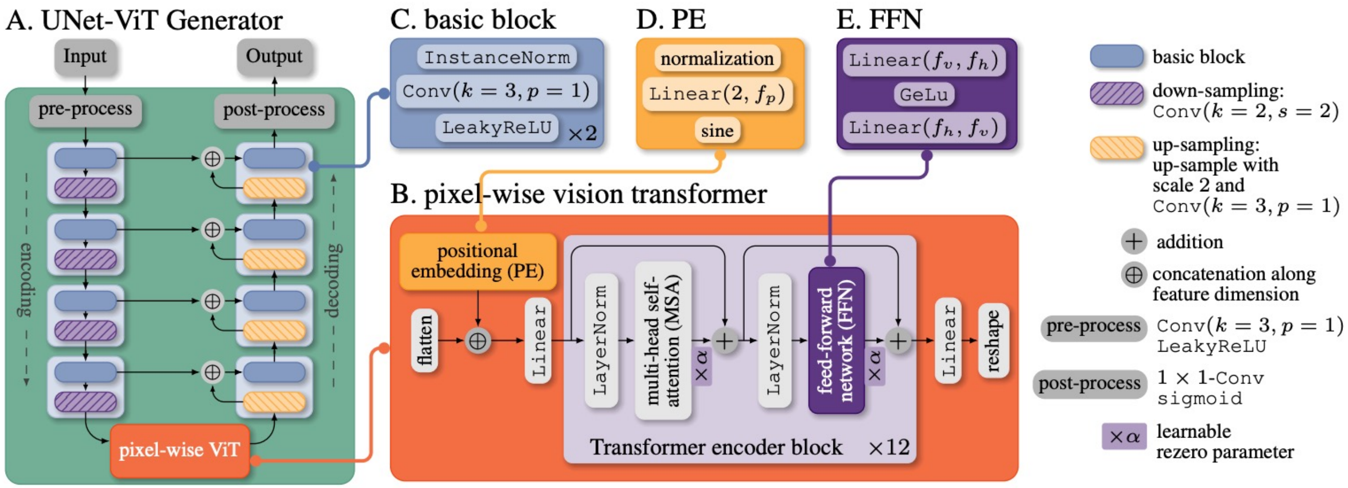


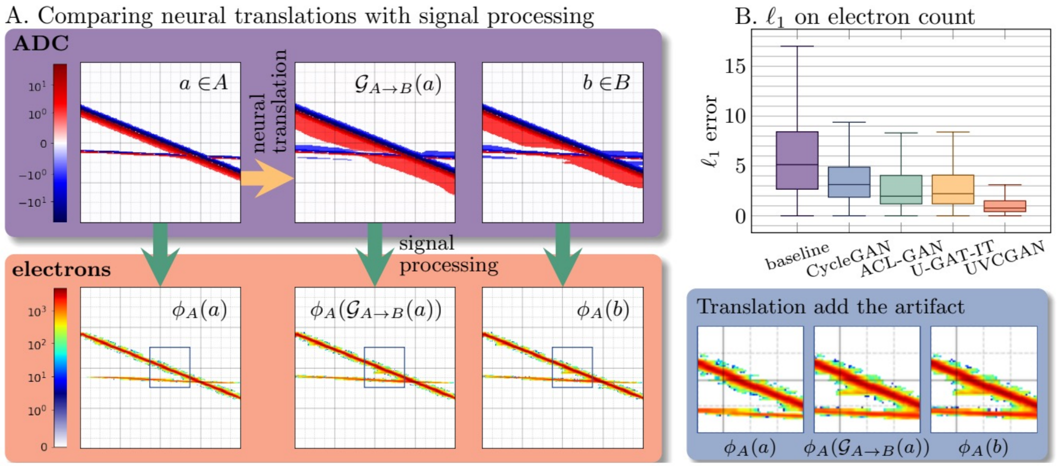
https://indico.bnl.gov/event/21684/contributions/85251/attachments/51799/88575/splat.pdf

LS4GAN – unpaired I2I translation

https://arxiv.org/abs/2203.02557
https://arxiv.org/abs/2304.12858

- ref. Dmitrii's talk
- **better results: learn features from data**
  - 3D effect
  - realistic noise
- **syst. unc. quantification**
  - extra CVN syst. unc.?





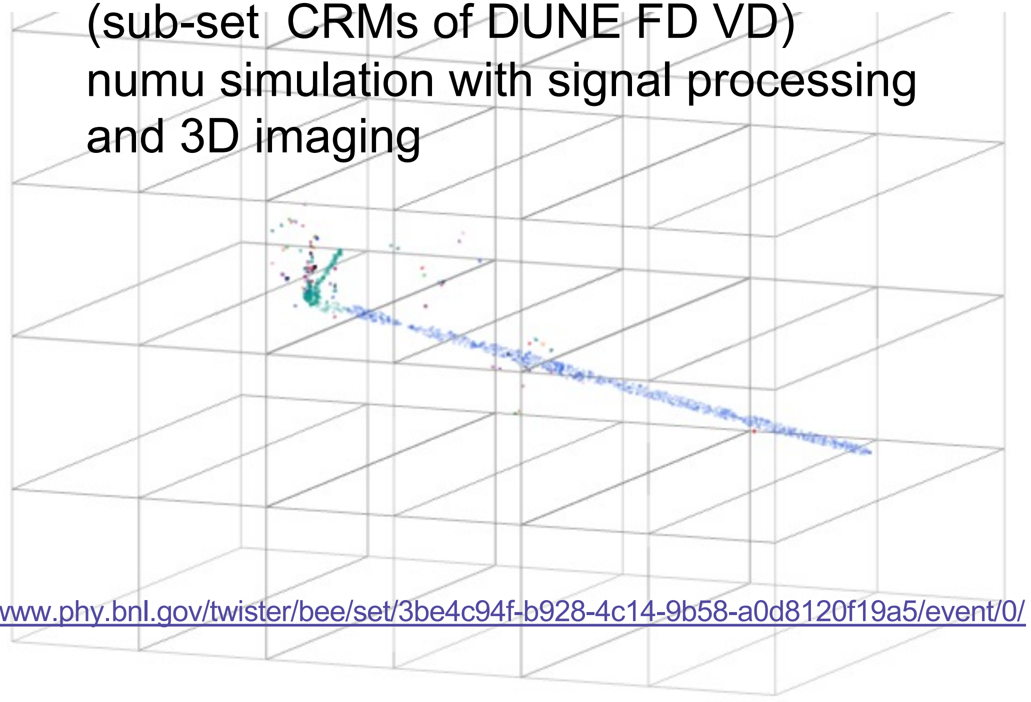
Imaging: tiling, solving, de-ghosting

https://indico.fnal.gov/event/58097/contributions/276229/
DUNE Collab. Mt. Sep. 2023
ref. to Chao, Ewerton/Lynn’s talk

- Foundation of Wire-Cell 3D reconstruction
- Potentially used by other reco. paradigms

- Some algorithms can be improved by AI/ML
- Still working on IO to LArSoft

DUNE-FD-VD-1x8x6 workspace
(sub-set CRMs of DUNE FD VD)
numu simulation with signal processing
and 3D imaging



https://www.phy.bnl.gov/twister/bee/set/3be4c94f-b928-4c14-9b58-a0d8120f19a5/event/0/

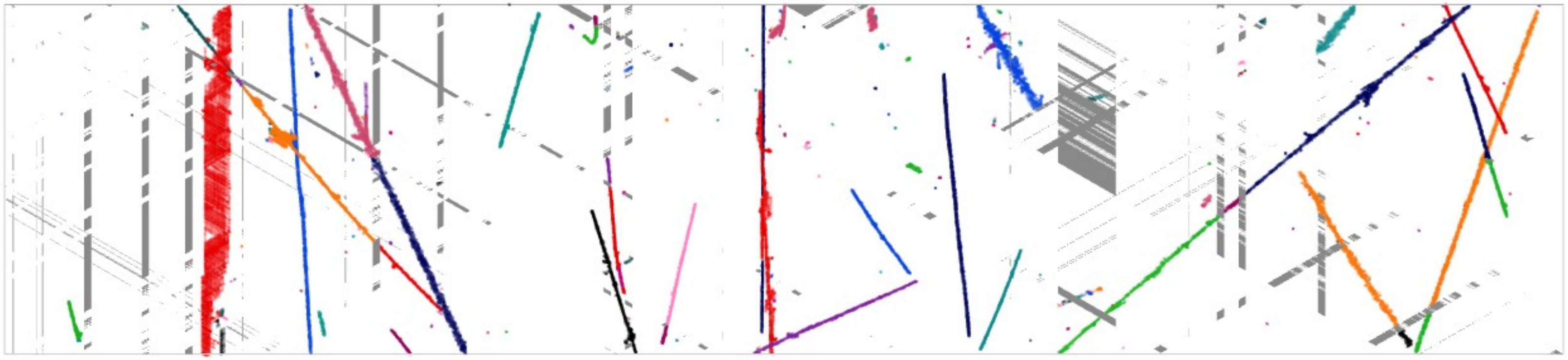
Brookhaven National Laboratory

Clustering

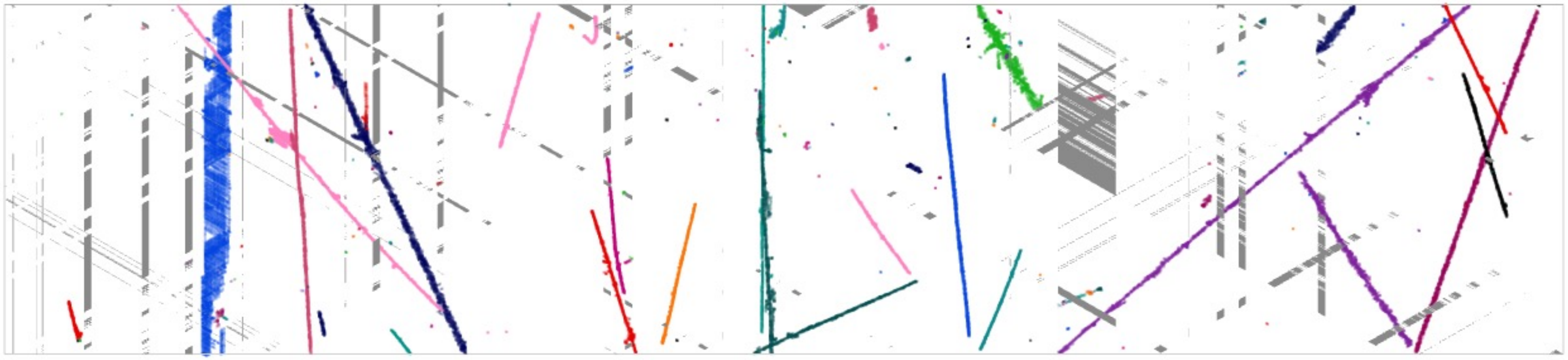
- ref. Chao's talk
- **partially available in WCT now**
- Preparing for following PatRec
  - Q-L matching, Traj. fitting, etc.
  - **After selection of the neutrino activities (e.g., in SBND), many DUNE-focused alg. could be tested.**
- Currently most heuristic -> very likely replaced/improved by AI/ML

Clustering for MicroBooNE sim. using WCT

before clustering



after clustering



https://www.phy.bnl.gov/twister/bee/set/b2dbddbe-e11a-4d7c-a1d2-a00d929ac7f4/event/6501/

IO for imaging results

- 2D images with selected neutrino activities
  - surface detectors
- sampled points -> space points
  - reduced info, needs effort to make it useful

- **ITensorSet [arrays]**
  - Ref: Brett's talk
  - Req:
    - **(de)serialization needs to be fast**
    - **Interoperability**
      - **LArSoft**
      - **AI/ML**
  - WCP ROOT:
    - TC, TDC
      - vector<POD>
      - vector<vector<POD>>

- ITensorSet
  - meta/json
  - vector<ITensor>

- ITensor
  - meta/json
  - Boost.MultiArray

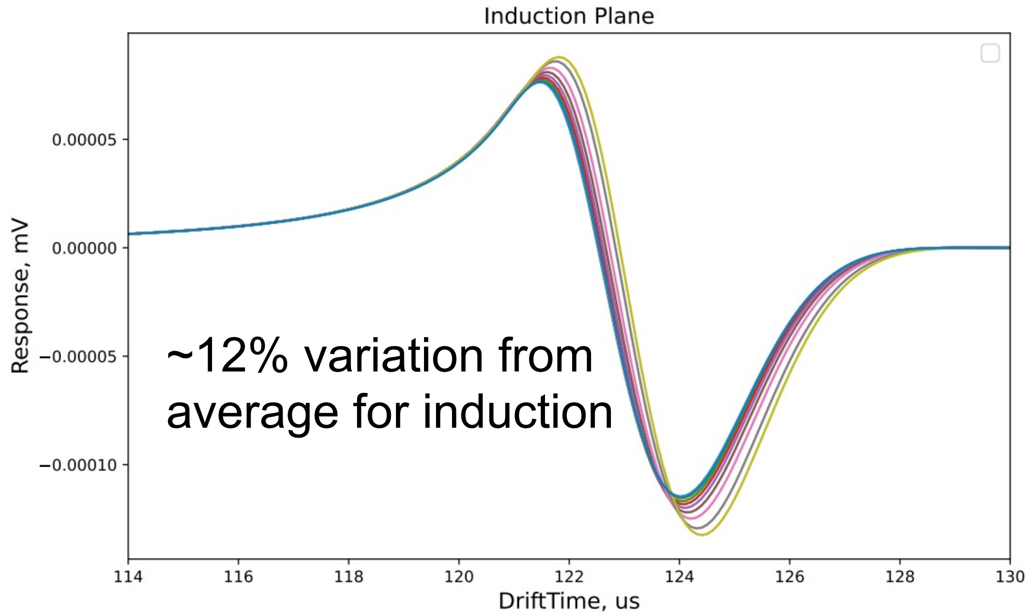
VD field response calculation

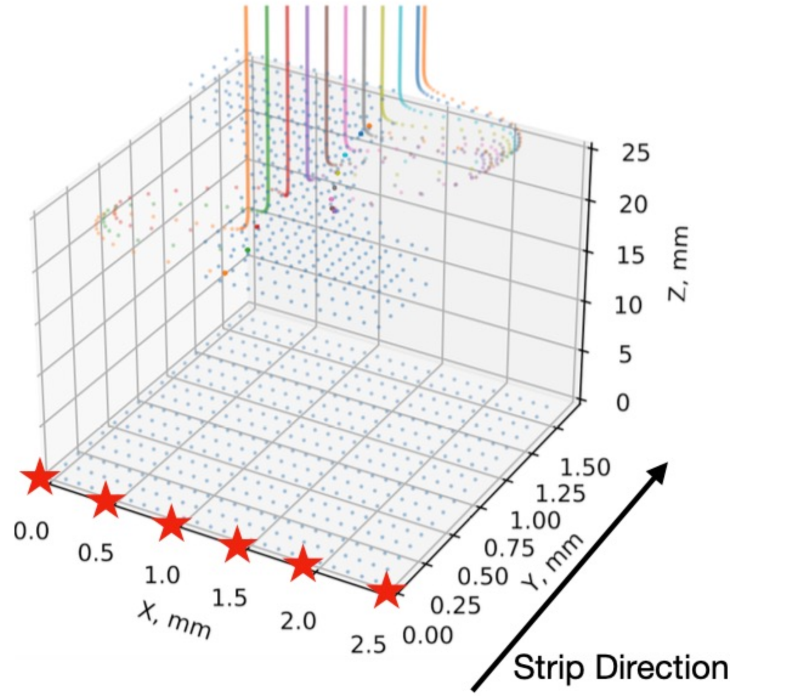
- Ref. Wenqiang’s talk
- Considering potential 3D effect
  - took Francesca’s approach
  - https://github.com/brettviren/pochoir
  - paper: *S. Martynenko et al 2023 JINST 18 P04033*
  - validated with PDVD coldbox data

- 3D + 2D FR calculation
  - drift path/speed: 3D
  - weighting field: 3D central + 2D outer region

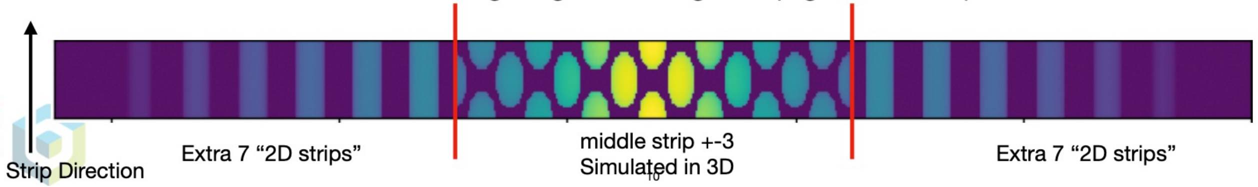
- averaged 3D for 2D Wire-Cell LArTPC simulation
  - average multiple paths

- Two field resp. available for 50L and PDVD





Slice of weighting field along drift (log color scale)

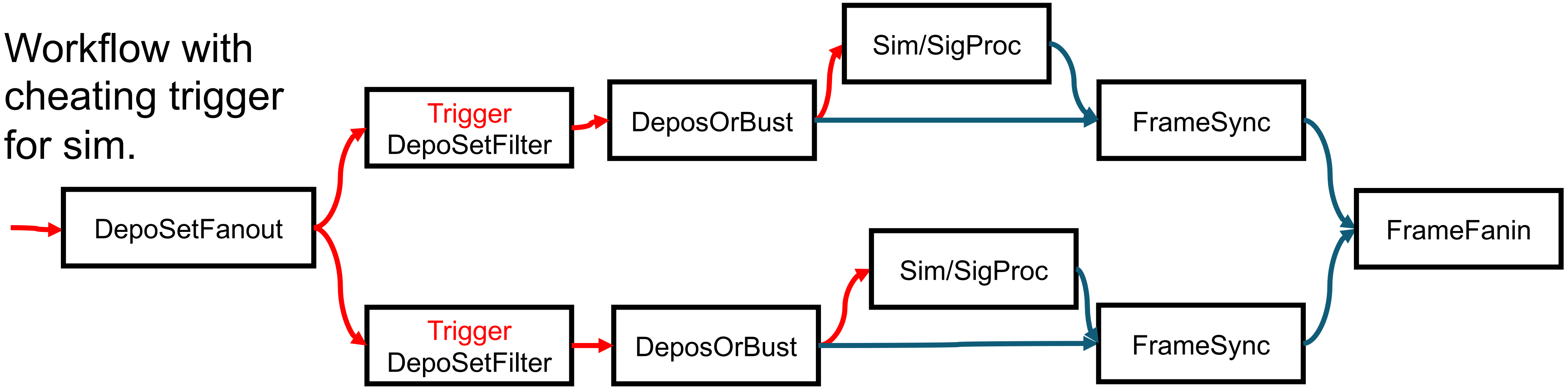


Skip processing APAs using async. WC node

ref: https://indico.fnal.gov/event/63824/

- Considering the APA-sparsity
- realized by new WC asyc. node introduced by BV
- **Critical for efficient 10kt simulation, especially beam focused ones**
- Makes it possible to keep raw digits
- **For SigProc (data), need a real APA level trigger alg.**

Workflow with cheating trigger for sim.



Test with one numu event:
Execution time (4032sec/259sec) ~ 15.5 times faster than baseline (no skip)

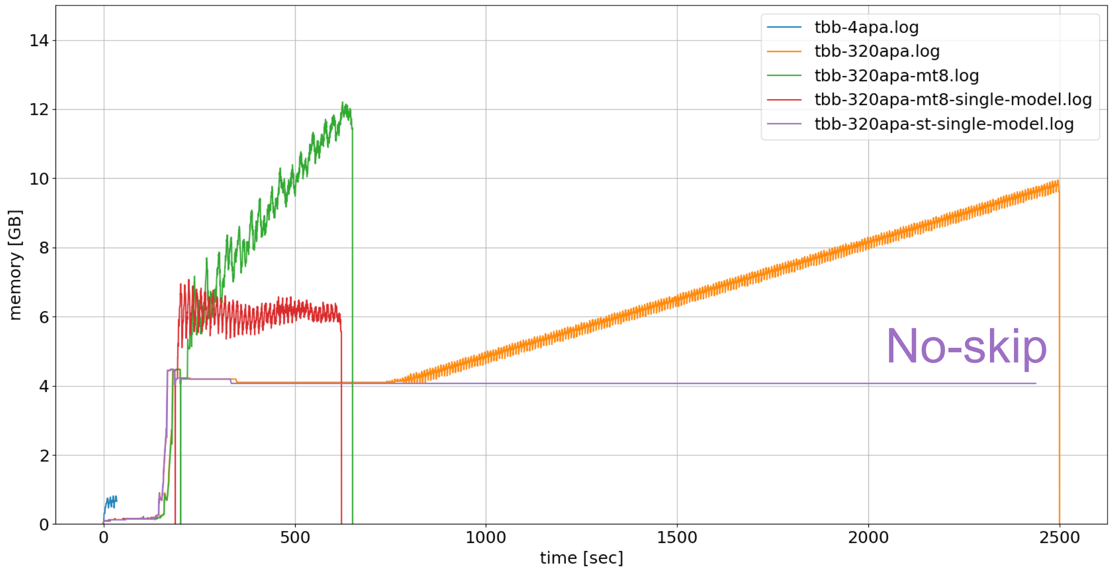
- **This ratio depends on the event activity**
    - Processed CRUs (ref/skip: 320/15) ~ 21 times
        - some overhead compared to 15.5

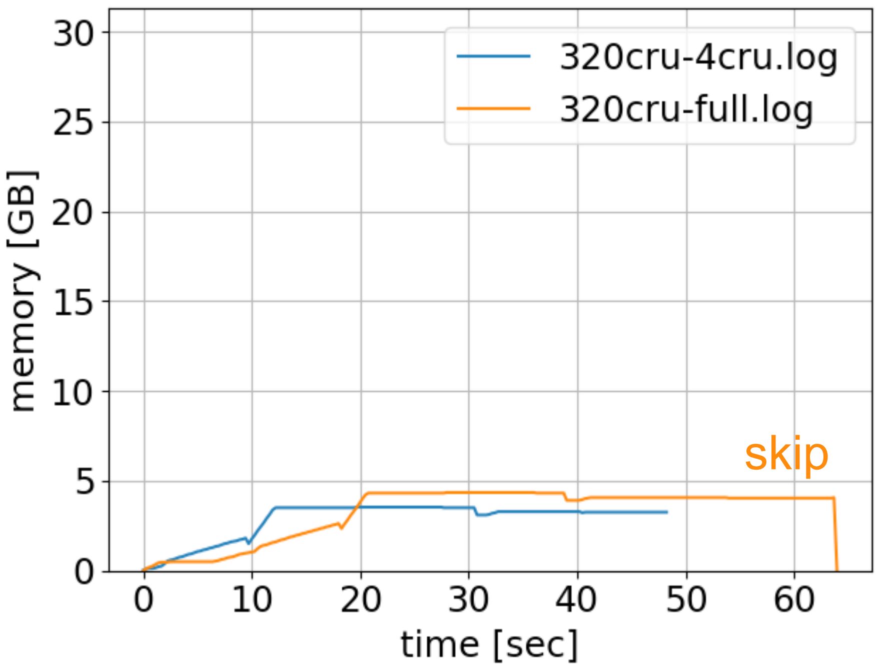
Brookhaven National Laboratory

Earlier skip test with an ideal track in 2-CRUs

Ref: https://indico.fnal.gov/event/60987/contributions/282811/

Initial tests for full 320CRU geom, ideal depo tracks in CRU 0, 4
- process all with shortcut: **~25sec (skip) vs. ~2400sec (ref)**





Radiological backgrounds

- Adding more realistic radiological backgrounds would break the simple yes/no APA level (cheating) filter of the skipping
    - Some better cheating alg. is needed, e.g., **depo->process == neutrino?**
- Compared to FDHD, FDVD non-bridged needs more resources, can bridged channels help?

From L. Paulucci

HD 12 APA workspace | VD 112CRM workspace

	HD 1x2x6		VD 1x8x14	
	No EM children	**With EM children**	**No EM children**	**With EM children**
Gen	2.6646s	2.92583s	30.8643s	27.7531s
	2085.11 MB	2117.41 MB	1853.91 MB	1882.64 MB
G4	95.9081s	107.752s	118.236s + 88.9054s	143.669s + 85.7495s
	2616.97 MB	3307.76 MB	6863.03 MB + 4905.57 MB	10675.5 MB + 7524.54 MB
Detsim	477.651s	506.957s	1226.02s	1030.05s
	2092.95 MB	2997.5 MB	5423.43 MB	5749.58 MB
Reco1	0.11843s	0.123185s	1.30019s	1.04877s
	976.384 MB	1629.86 MB	3178.41 MB	4956.22 MB
File size	1.4GB	1.8GB	2.5 GB	3.6GB

Brookhaven National Laboratory

Discussion

Working on applying DNN-ROI for multiple experiments

- ref. Wenqiang, Avinay/Moon's talk

Skip processing for DUNE-FD almost ready for production tests

Better FR ready

Major discussion focus is the Wire-Cell->LArSoft IO

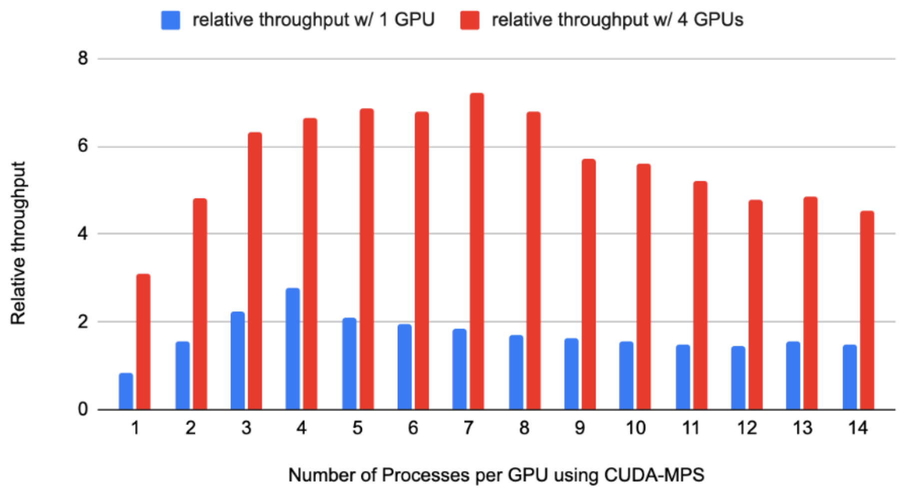
In addition, Wire-Cell has the potential to directly read in HDF5 DAQ files, but this may be discussed in the IO session. Ref. BV's talk

Thanks!

GPU based simulation – More efficient sim. for AI/ML?

https://indico.cern.ch/event/948465/contributions/4323675/
https://arxiv.org/pdf/2203.02479.pdf

- Need to be coupled with computing facility
- Needed?

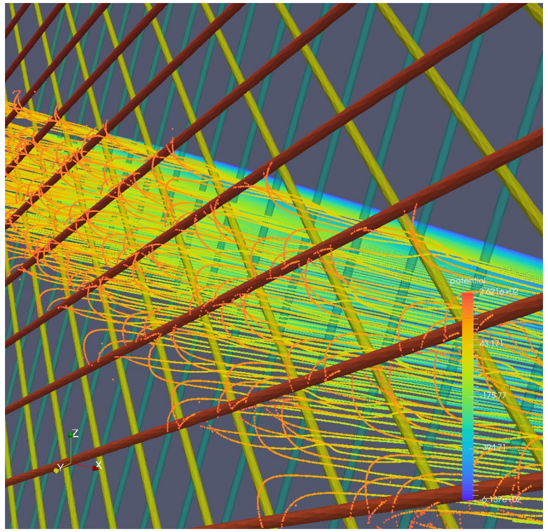


Brookhaven National Laboratory

2D-Convolution based LArTPC Simulation

$$\text{Ramo's theorem: } i = -q\,\vec{E}_w \cdot \vec{v}_q$$

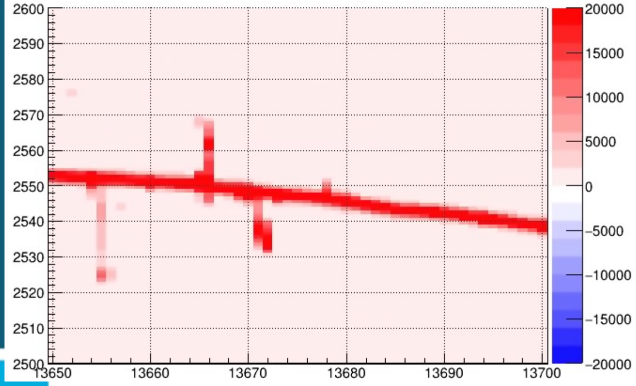
2D: approximate translational symmetry along the wire direction



LArTPC wire-readout measures induced charge ⊗ response

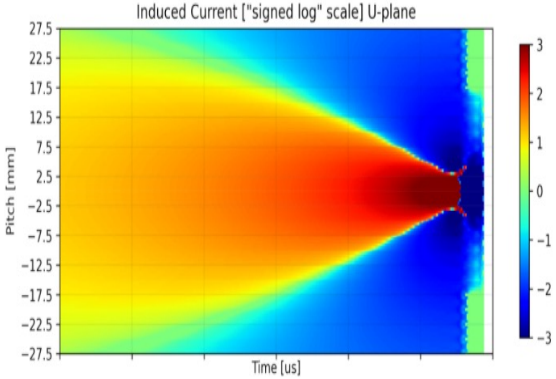
$$M(t', x') = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} R(t, t', x, x') \cdot S(t, x)\,dt\,dx + N(t', x')$$

Energy depo + diffusion
+ rasterization



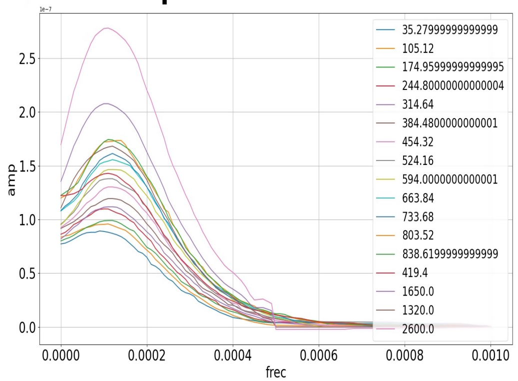
⊗

Long-range and
position-dependent field



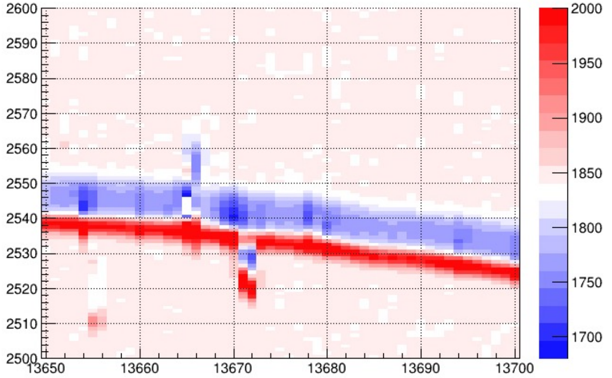
⊕

Noise
Spectrum



⟶

Final Signal



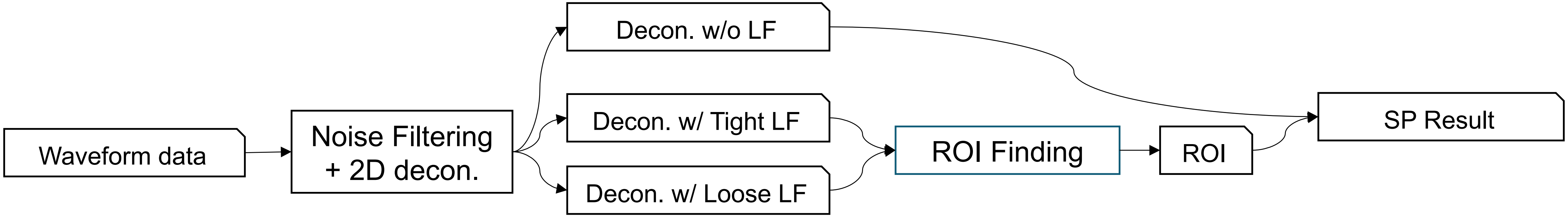
Brookhaven National Laboratory

2D-Convolution based Signal Processing

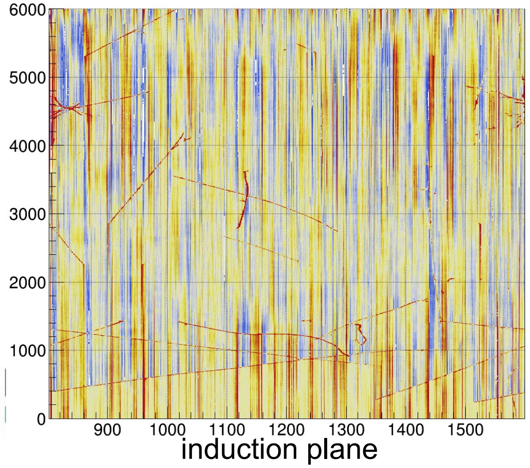
Signal Processing (SP) of LArTPC resolves charge from the original measurement:

$$S(\omega_t, \omega_x) \sim \frac{F(\omega_t, \omega_x) \cdot M(\omega_t, \omega_x)}{R(\omega_t, \omega_x)} \xrightarrow{IFT} S(t, x)$$

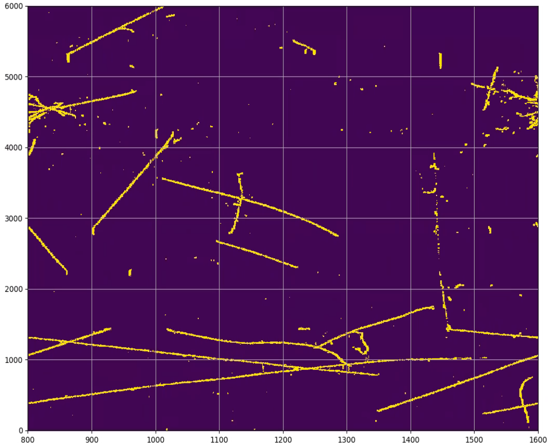
- *"2D deconvolution": assuming translational symmetry in the third dimension*
- *Utilize the signal/noise separation in both frequency and time domain*



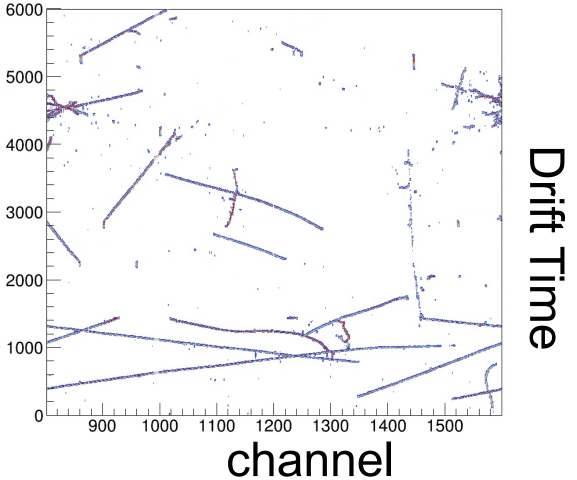
Decon. w/o LF (low frequency) filter
Waveform ⟶ charge, dense



ROI (region of interest):
Hit finding, sparsify



SP (signal processing) result:
Sparse, ionization charge



Wire-Cell 3D Imaging Principle

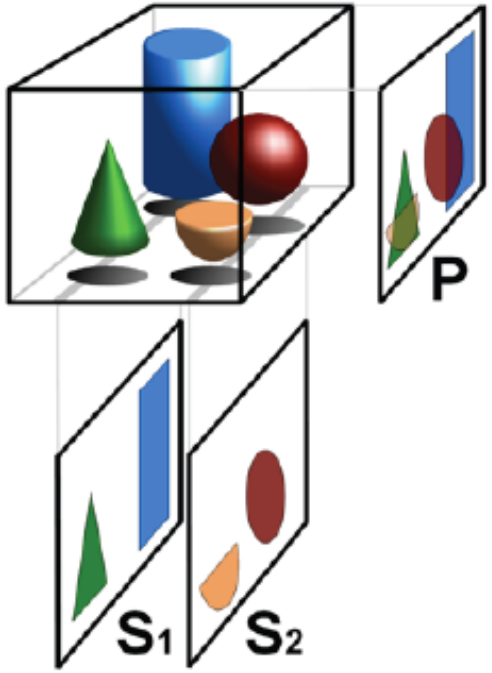
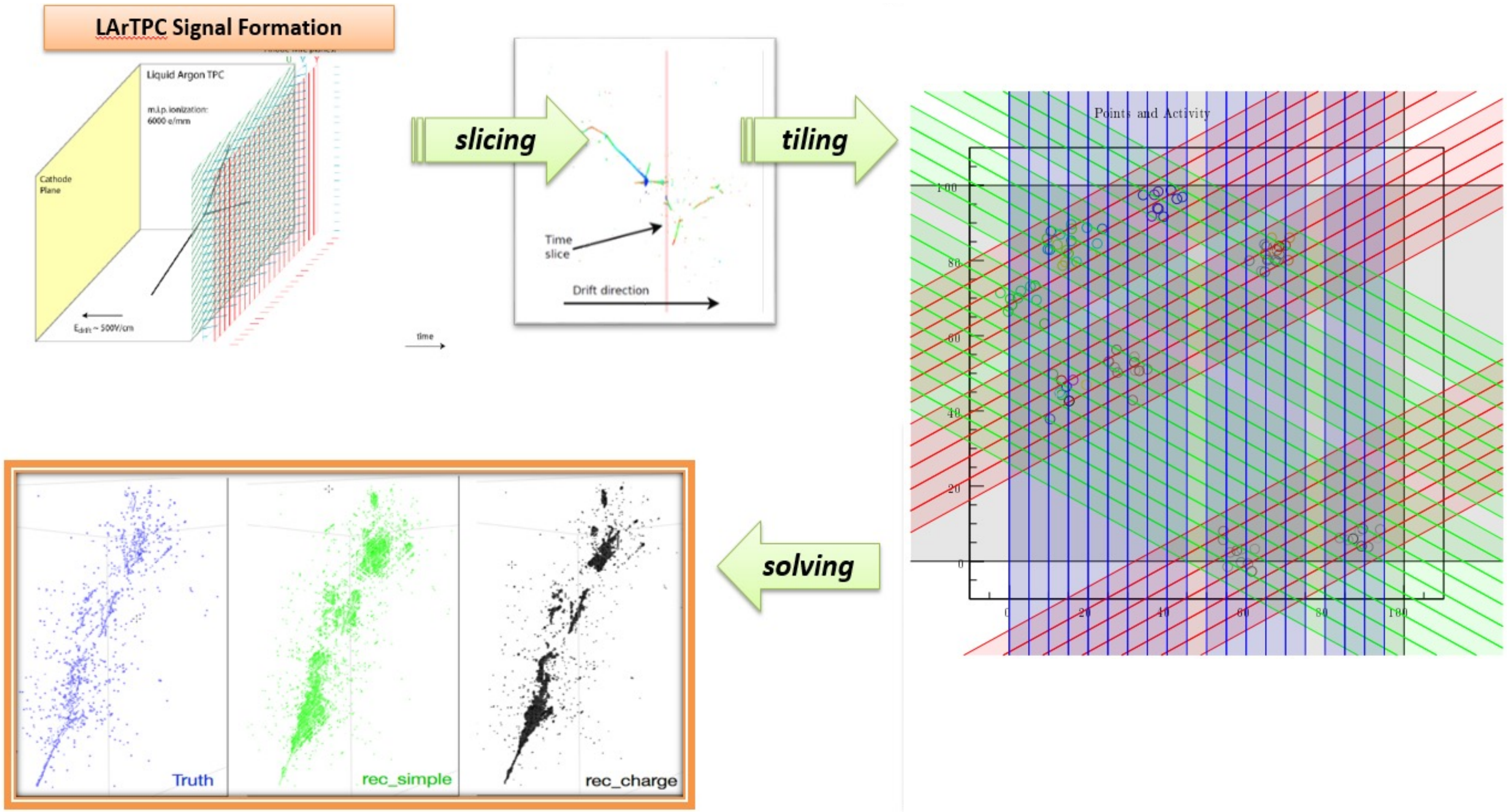


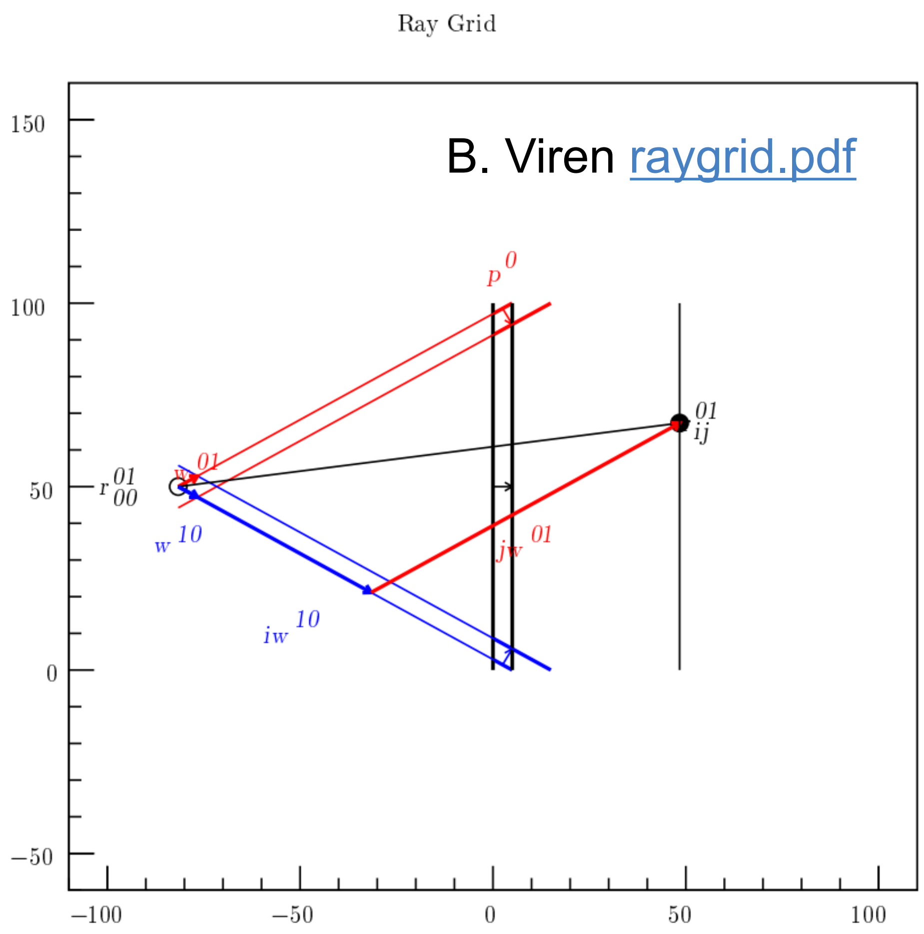
Fig.1:Basic principle of tomography: superposition free tomographic cross sections S1 and S2 compared with the projected image P

https://en.wikipedia.org/wiki/Tomography



"Three-dimensional Imaging for Large LArTPCs", JINST 13, P05032 (2018)

Ray grid



Convenience way to calculate wire crossing projections
Multiple non-orthogonal 2D coordination system

- one for each wire plane pair

2D crossing coord:

$$r_{ij}^{lm} = r_{00}^{lm} + jw^{lm} + iw^{ml}$$

Projection to the pitch direction of the target plane:

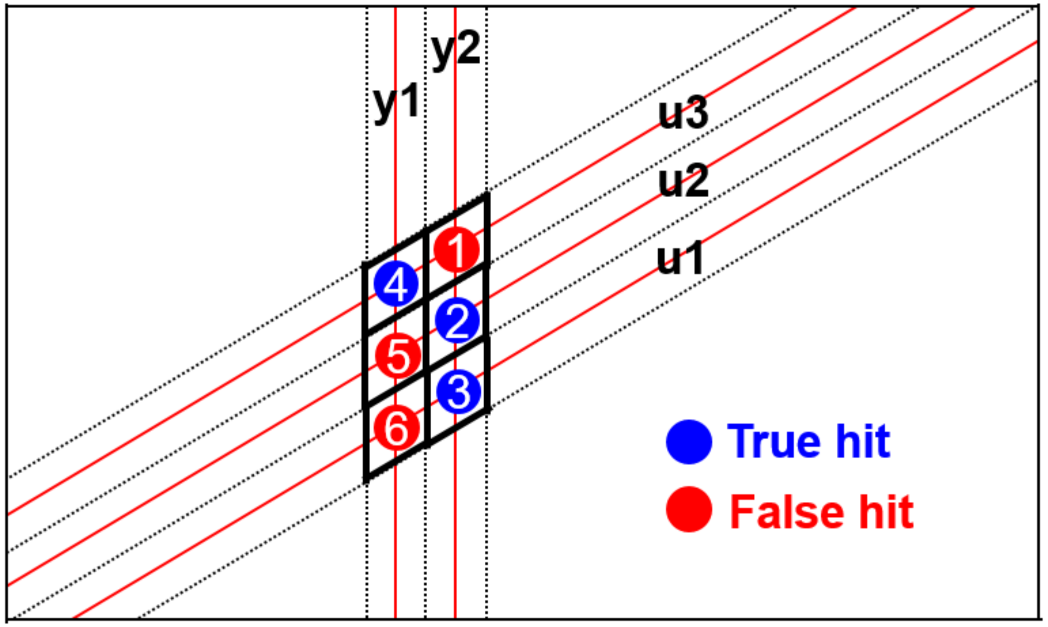
$$p_{ij}^{lmn} = (r_{ij}^{lm} - c^n) \cdot \hat{p}^n$$

Expanded:

$$P_{ij}^{lmn} = r_{00}^{lm} \cdot \hat{p}^n + jw^{lm} \cdot \hat{p}^n + iw^{ml} \cdot \hat{p}^n - c^n \cdot \hat{p}^n$$

w^{lm}: displacement vector along layer-m spaced by layer-l

Solving: usage of Charge, Sparsity, Positivity, Proximity



measured charges on Wires

$$y = A \cdot X$$

true charge to be resolved

$$\begin{pmatrix} y1 \\ y2 \\ u1 \\ u2 \\ u3 \end{pmatrix} = \begin{pmatrix} 0 & 0 & 0 & a & a & a \\ a & a & a & 0 & 0 & 0 \\ 0 & 0 & a & 0 & 0 & a \\ 0 & a & 0 & 0 & a & 0 \\ a & 0 & 0 & a & 0 & 0 \end{pmatrix} \begin{pmatrix} H1 \\ H2 \\ H3 \\ H4 \\ H5 \\ H6 \end{pmatrix}$$

matrix determined by geometry, a=1

- The goal is to differentiate the true hits from fake ones by using the charge information
  - ~ large charge → true hits
  - ~ zero charge → fake hits
  - **correct SigProc is important,** [J. Jo's talk.]
- Sparsity, positivity, and proximity information are added through compressed sensing (L1 regularization)

$$\text{L1 reg.} \quad O(N!) \rightarrow O(m \times N)$$

$$\chi^2 = (y - A \cdot x)^2 + \lambda \cdot \sum_i |x_i|$$

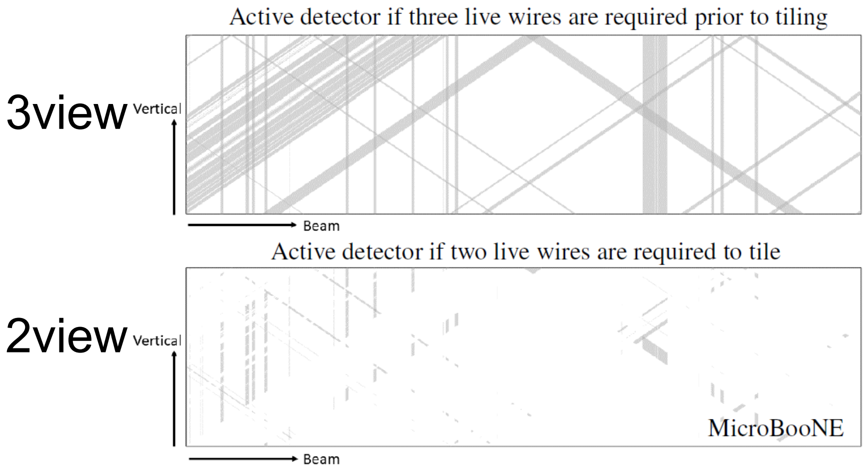
E. Candes, J. Romberg, T. Tao arXiv-math/0503066

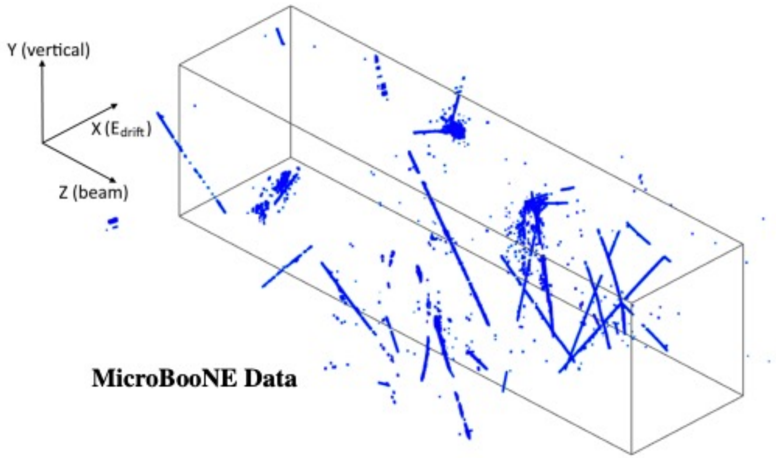
Brookhaven National Laboratory

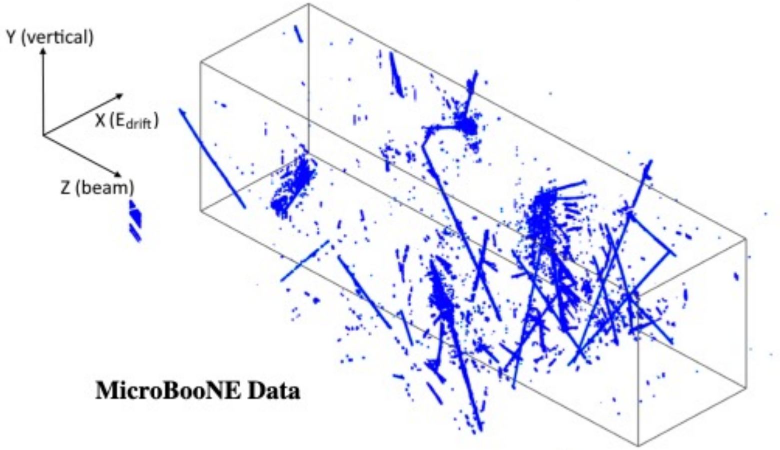
de-ghosting

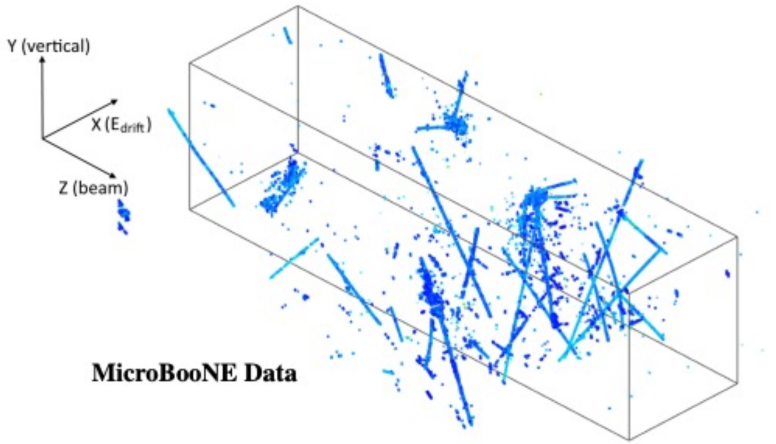
- Solving alone cannot eliminate all ghosts
- In MicroBooNE, the situation is worse when 2-view blobs are allowed
  - 10% dead channels → 3view only is not acceptable
  - **2view tiling is needed → more ghosts**
  - https://arxiv.org/abs/2011.01375
- de-ghosting: larger, connected blobs tends to be true
  - future AI/ML opportunity

dead regions

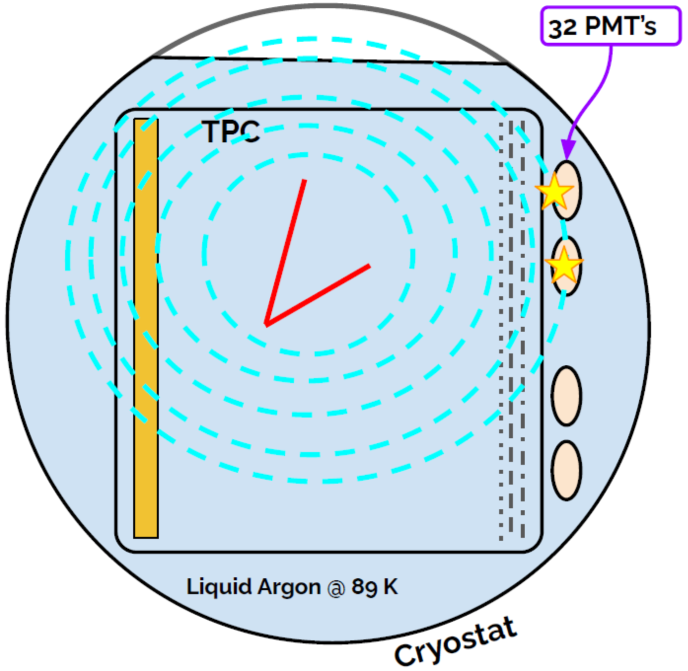




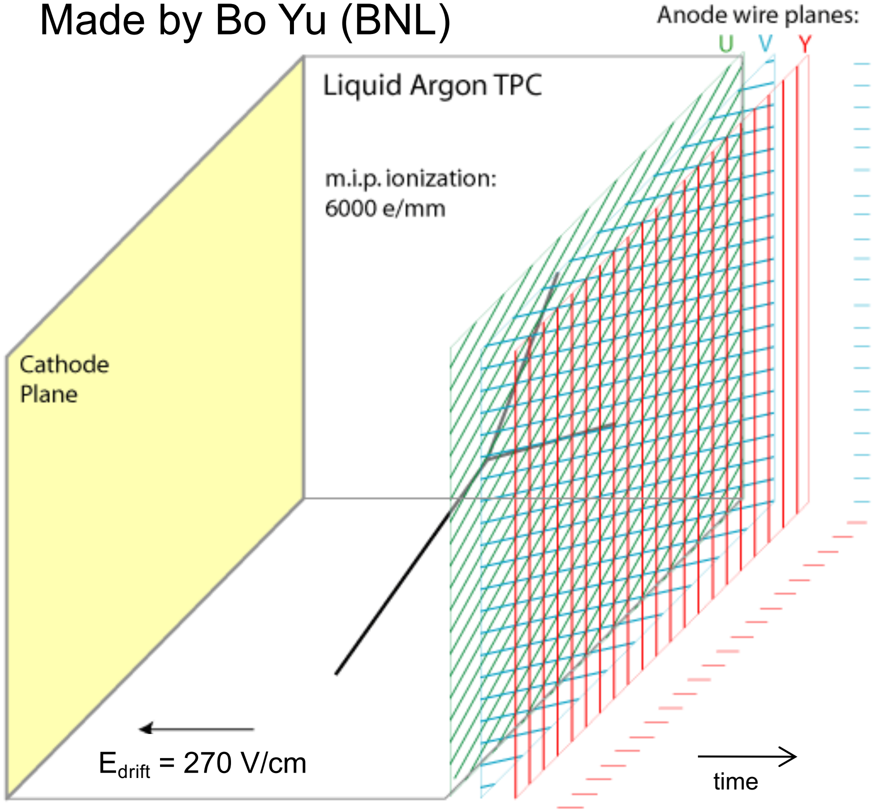




Cluster-flash (light) Matching



Made by Bo Yu (BNL)



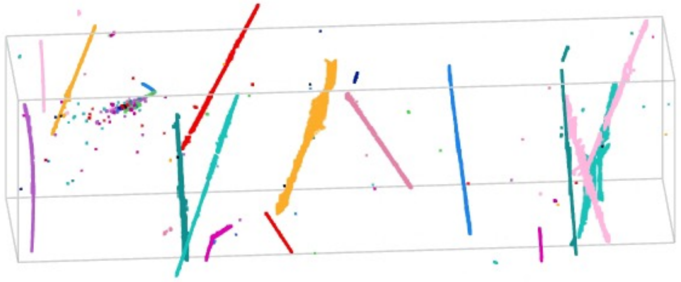
PMTs detect the scintillation light, time ~ns

Drift velocity 1.1 mm/μs → several ms drift time

- In LArTPC, the light (PMT) readout and charge (TPC) readout systems are decoupled
- The identification of neutrino interaction candidate requires matching the charge signal with the light signal in order to obtain the event time

Matching Principle

JINST 16 P06043 (2021)



Hypotheses Selection

- **Light signal proportional to (reconstructed 3D) charge**
- Known light acceptance given position
- Predicted vs. Measured light pattern with Compressed Sensing

Core Charge-Light Matching Algorithm

$$\chi^2 = \sum_i \sum_j \chi_{ij}^2 + \chi_{p1}^2 + \chi_{p2}^2 + \chi_{p3}^2$$
Overall test statistics to be minimized

$$\chi_{ij}^2 = \frac{(M_{ij} - \sum_k a_{ik} \cdot P_{ikj} - b_i \cdot M_{ij})^2}{\delta M_{ij}^2}$$
Comparison of the measured and predicted light pattern

Rule 1st

$$\chi_{p1}^2 = \sum_i \frac{(\sum_k a_{ik} - 1)^2}{c_1^2}$$
Each charge cluster can only be used once

Rule 2nd

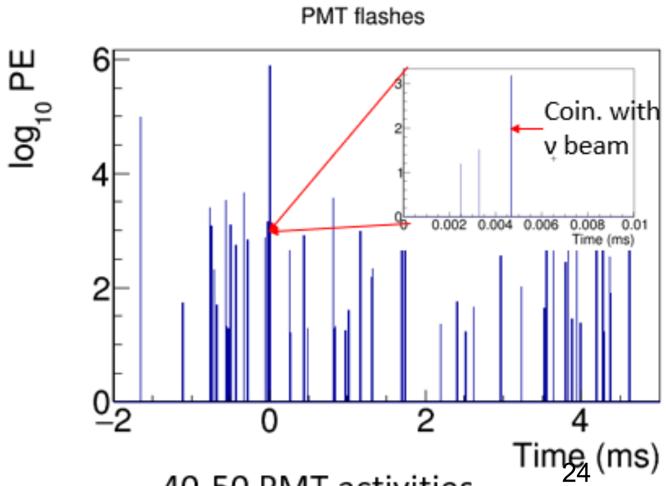
$$\chi_{p2}^2 = \sum_i \frac{b_i^2}{c_2^2}$$
Observed light flash may not correspond to any charge cluster

$$\chi_{p3}^2 = \lambda \cdot \sum_i \sum_k a_{ik}$$
Compressed sensing to select the best pairs

M: Measured Light Pattern
P: Predicted Light Pattern
δ : Uncertainty
i : ith Light Flash
j : jth PMT
k : kth Charge Cluster

Aggressively pursue charge-light matching
Additional cuts to examine the "light mismatch" events

All possible hypotheses
- One cluster → <u>at most one flash</u> (inefficiency in the light system)
- One flash → <u>many or zero</u> TPC clusters within corresponding active volume (activities in inactive volume)



40-50 PMT activities

Brookhaven National Laboratory

PROCESS NAME	MODULE LABEL	PRODUCT INSTANCE NAME	DATA PRODUCT TYPE	SIZE
SinglesGen	generator		std::vector<simb::MCTruth>	1
SinglesGen	rns		std::vector<art::RNGsnapshot>	1
SinglesGen	TriggerResults		art::TriggerResults	1
G4	IonAndScintExternal		std::vector<sim::SimEnergyDeposit>	202
G4	elecDrift		std::vector<sim::SimChannel>	1735
G4	PDFastSimAr		std::vector<sim::SimPhotonsLite>	88
G4	rns		std::vector<art::RNGsnapshot>	8
G4	IonAndScint		std::vector<sim::SimEnergyDeposit>	38567
G4	TriggerResults		art::TriggerResults	1
G4	PDFastSimAr		std::vector<sim::OpDetBacktrackerRecord>	88
G4	largeant		std::vector<simb::MCParticle>	16
G4	largeant	LArG4DetectorServicevolTPCActive	std::vector<sim::SimEnergyDeposit>	38567
G4	largeant	LArG4DetectorServicevolCryostat	std::vector<sim::SimEnergyDeposit>	202
G4	PDFastSimXe		std::vector<sim::OpDetBacktrackerRecord>	88
G4	PDFastSimXe		std::vector<sim::SimPhotonsLite>	88
G4	largeant		sim::ParticleAncestryMap	-
G4	largeant		art::Assns<simb::MCTruth,simb::MCParticle,sim::GeneratedParticleInfo>	16

76k

IonAndScint

genie

38k

MARLEYGen	K40GenInCathode		std::vector<simb::MCTruth>	1
MARLEYGen	Rn220ChainPb212GenInLAr		std::vector<simb::MCTruth>	1
MARLEYGen	Rn222ChainRn222GenInLAr		std::vector<simb::MCTruth>	1
MARLEYGen	K40GenInAnode		std::vector<simb::MCTruth>	1
MARLEYGen	U238ChainGenInCathode		std::vector<simb::MCTruth>	1
MARLEYGen	CavernNGammasAtLAr1x8x14		std::vector<simb::MCTruth>	1
MARLEYGen	Th232ChainGenInCathode		std::vector<simb::MCTruth>	1
MARLEYGen	K42From42ArGenInUpperMesh1x8x14		std::vector<simb::MCTruth>	1
MARLEYGen	Rn222ChainGenInPDS		std::vector<simb::MCTruth>	1
MARLEYGen	K42From42ArGenInLAr		std::vector<simb::MCTruth>	1
MARLEYGen	Rn222ChainFromPb214GenInUpperMesh1x8x14		std::vector<simb::MCTruth>	1
MARLEYGen	foamGammasAtLAr1x8x14		std::vector<simb::MCTruth>	1
SupernovaG4Stage1	largeant	LArG4DetectorServicevolCryostat	std::vector<sim::SimEnergyDeposit>	1972287
SupernovaG4Stage1	largeant		sim::ParticleAncestryMap	-
SupernovaG4Stage1	largeant	LArG4DetectorServicevolTPCActive	std::vector<sim::SimEnergyDeposit>	776790
SupernovaG4Stage1	largeant		std::vector<simb::MCParticle>	131607
SupernovaG4Stage1	rns		std::vector<art::RNGsnapshot>	8
SupernovaG4Stage1	largeant		art::Assns<simb::MCTruth,simb::MCParticle,sim::GeneratedParticleInfo>	131607
SupernovaG4Stage1	elecDrift		std::vector<sim::SimChannel>	87248
SupernovaG4Stage1	PDFastSimArExternal		std::vector<sim::OpDetBacktrackerRecord>	184
SupernovaG4Stage1	PDFastSimAr		std::vector<sim::OpDetBacktrackerRecord>	184
SupernovaG4Stage1	PDFastSimAr		std::vector<sim::SimPhotonsLite>	184
SupernovaG4Stage1	PDFastSimArExternal		std::vector<sim::SimPhotonsLite>	184
SupernovaG4Stage1	IonAndScintExternal		std::vector<sim::SimEnergyDeposit>	1972287
SupernovaG4Stage1	TriggerResults		art::TriggerResults	1
SupernovaG4Stage1	IonAndScint		std::vector<sim::SimEnergyDeposit>	776790
G4Stage2	PDFastSimXeExternal		std::vector<sim::SimPhotonsLite>	184
G4Stage2	PDFastSimXeExternal		std::vector<sim::OpDetBacktrackerRecord>	184
G4Stage2	PDFastSimXe		std::vector<sim::SimPhotonsLite>	184
G4Stage2	TriggerResults		art::TriggerResults	1
G4Stage2	PDFastSimXe		std::vector<sim::OpDetBacktrackerRecord>	184
G4Stage2	rns		std::vector<art::RNGsnapshot>	4

Supernova?

777k

(1972+776)*2 = 5496k