Drift Simulation and Data Reduction

Brett Viren

Physics Department

BROOKHAVEN NATIONAL LABORATORY

DUNE Collaboration Meeting, Jan 2017

Outline

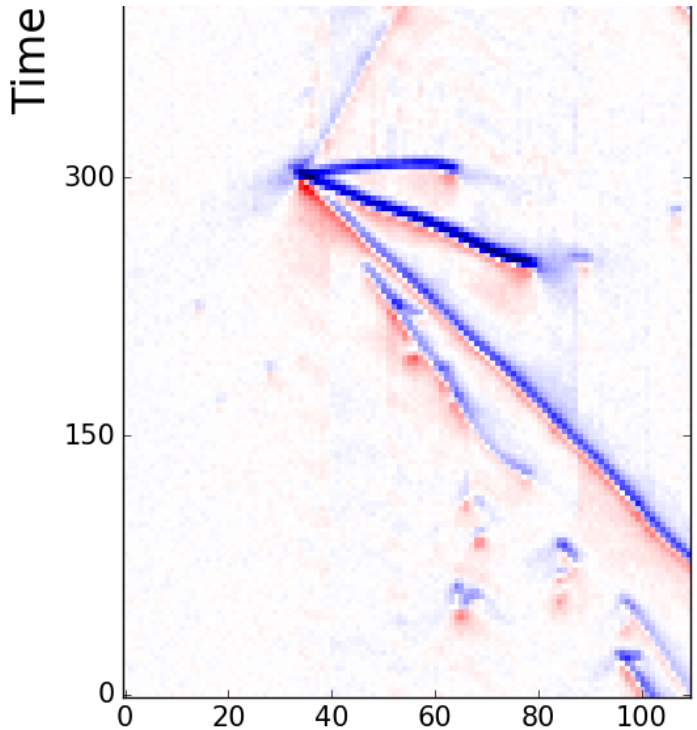
Full TPC Drift and Signal Simulation

Proposal for Initial Data Reduction Process

Motivation for a Better Simulation

- MicroBooNE observes long range induction effects and effects potentially due to intra-wire field variations.
- Current simulation (and deconvolution) does not take these into effect.
- Drift/signal sim very tied to nature of field calculations.
- RAM issues with with higher-rez fields

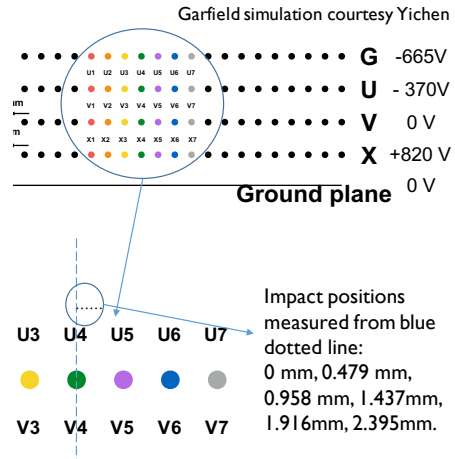
$\rightarrow$ need a fresh start

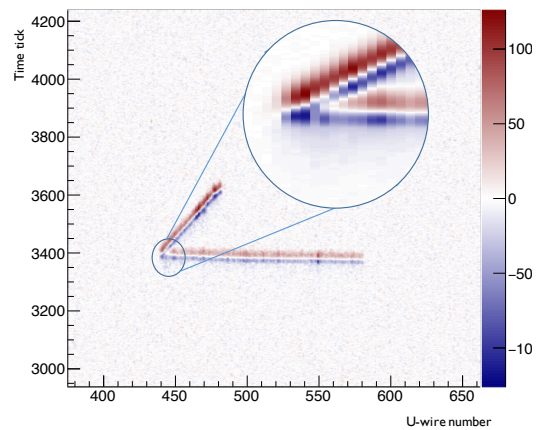


Zoom in on raw MicroBooNE event, induction plane.
(figure from Brooke and Chao.)

Initial Improvements - Xiaoyue

- Use (2D) field responses which are calculated:
  - across **7 nearest wires** (instead of 1)
  - at paths spaced by **1/10th the wire pitch** (instead of full pitch)
- Truncated 3σ Gaussian diffusion "cloud"
  - Longitudinal & transverse diffusion coefficients.
  - Charge elements sampled to weight field response drift paths.
- Some long-range induction effects start to be reproduced.





From Xiaoyue's report last collab meeting.

Next gen signal simulation overview

- Retain 1/10th pitch path spacing
- Support input fields responses over arbitrary range
  - Currently using 21 nearest wires
- New convolution algorithm to minimize RAM usage.
  - Same optimization strategy as Xiaoyue.
- Implemented in Wire Cell Toolkit (instead of prototype) to facilitate integration into LArSoft.

Initial Results

Remaining Work

- Various validation and performance tuning.
- Low level functions are done, need proper Wire Cell Interface layer packaging.
- LArSoft integration via Art Service Interface

Motivation for Initial Data Reduction

- protoDUNE/SP will produce 2.5 PB in two months run
  - ATLAS Run 1 level, uBoone is 1 PB/year.
- Want to re-run reconstruction many times on the data.
- People need tidy samples for variety of studies.
- No matter what, first processing steps are:
  1. Software noise filtering (hopefully minimal!)
  2. Signal processing / response deconvolution
- After this processing any usable signal is above noise (by definition) so it is safe to save only **signal regions-of-interest** (ROIs).
  - → let's say "signal ROI" instead of "zero suppression" which is what is used on raw data.

Reduction Proposal

- Run initial data reduction process ASAP after raw data is written ("keep up" mode).
- **Still save full raw data to tape!**
  - Eg, for understanding detector during commissioning.
- Use reduced data set for:
  - production "keep up" and re-processing jobs.
  - easier sharing of fresh data to look for issues.
  - facilitate tidy samples for reco development.
  - allow every collaborator to have their own full copy!
- If ever we improve reduction, go back to full raw data.

Reduction Methods, Options and Evaluation

sample	use a few uBoone raw events (`ADC` data tier)
scaling	protoDUNE has $\sim 2\times$ channels of uBoone
noise filter	reduces entropy, helps compression (`SUB`)
	$\rightarrow$ assuming all goes well represents expectation of protoDUNE data
ROI selection	throw away noise waveform pieces (`ROI`)
resample	sum N neighboring bins to **remove oversampling** (`RBn`)
	$\rightarrow$ but do not violate Nyquist theorem!
truncation	from 32 bit float working samples to 16 bit int
packing	12, 16 or 32 bit packing into 32 bit output arrays.
compression	apply ROOT compression (0-9)

Look at **reduction factor** and **compression time** as a function of all meaningful combinations of above options.

Reduction Options and Efficacy

tier	bits	Z Lvl	Z time	factor	2.5 PB
ADC	12	5	6.8 sec	**4.2**	0.6 PB
ADC	12	9	**4.3 min**	5.5	0.5 PB
SUB	12	5	5.6 sec	6.0	0.4 PB
SUB	12	9	**3.6 min**	8.0	0.3 PB
ROI	16	9	0.5 sec	150	17 TB
RB2	16	9	0.3 sec	260	9.6 TB
RB3	16	9	0.2 sec	415	**6.0 TB**
RB4	16	9	0.1 sec	540	4.6 TB

- Noise filtering and signal processing takes 4.5 min/trig on protoDUNE
  - → Z time above is scaled to 6 APA protoDUNE channel count.
- 12 bits means fully packing 8 samples across 3 words (32 bit int).
- Can achieve expected factor of 4 raw data compression at uBoone noise level.
  - → expect factor 6-8 raw data compression if low-noise, but Z time grows fast!
- Z time is insignificant once any data reduction is done.
- Only counts the TPC data, other data sizes may become significant!
- RB4 rebins to exactly the Nyquist limit, **RB3 is the safe recomendation**.

Implications for Reco

Validate Gaussian Hit Finder is not impacted by rebinning.

- Must be okay or Nyquist is wrong!
- We do better than Nyquist by summing instead of resampling.
- Simple check by Xin, $\mu = 10, \sigma = 2 \longrightarrow$
  - We recommend RB3.
- Reduction is otherwise "invisible" after Gaussian hit finding.

RBN	const	μ	σ
1	10.0	10.0	2.00
2	19.8	10.0	2.02
4	38.5	10.0	2.08
8	68.9	10.0	2.31

Need development for:

- Art input module to read in reduced data
  - $\rightarrow$ needed for raw data in any case, reduced data will be similarly packed
- Art Service interface to the reduction code will allow running it in LArSoft.
  - $\rightarrow$ will run in "keep up" in plain-art or stand-alone to conserve RAM

Summary

simulation Technical progress on implementation for an improved simulation which allows for long-range induction effects and intra-wire field variations has been made

reduction A data reduction stage can give us a much smaller sample keeping all useful signal for use by remaining processing chain and enables all collaborators to easily obtain a copy.