Overlays
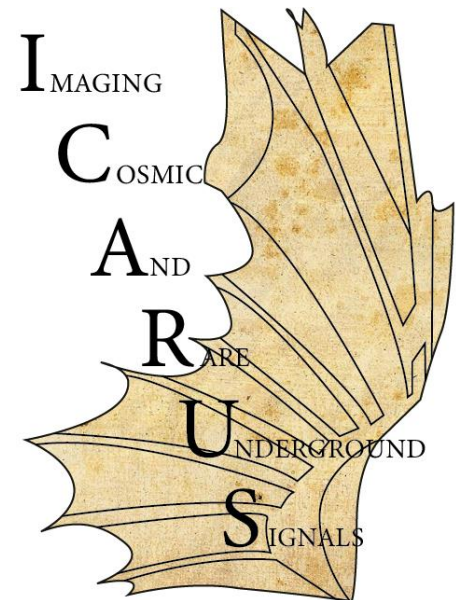
Ivan Caro Terrazas, on behalf of the Overlay Team

ICARUS Collaboration Meeting

14-Oct-2024

Overlay Overview

- Goal: **Make MC look as data-like as possible**

- The basic idea is to "overlay" simulated neutrino events onto off-beam events
  - For TPC and PMTs, we add waveforms from cosmic data and the simulated event
  - CRT hits from both data and simulations are added

- Allows for data-driven modeling of backgrounds
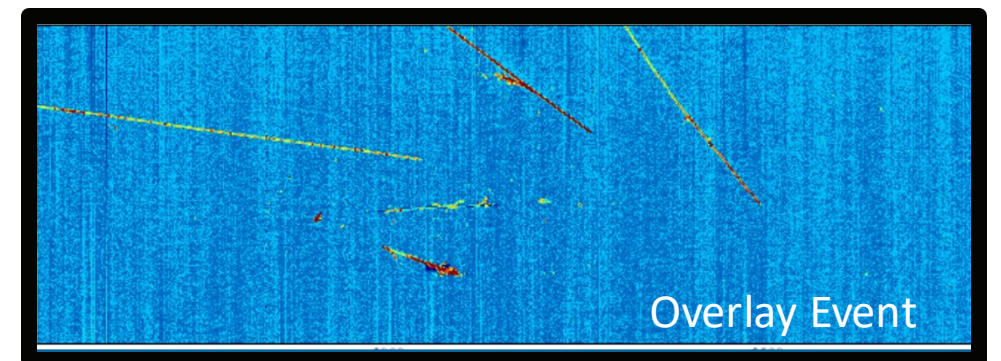 - Captures noise from TPC, PMTs, and CRT

Cosmic Data

Simulation Neutrino
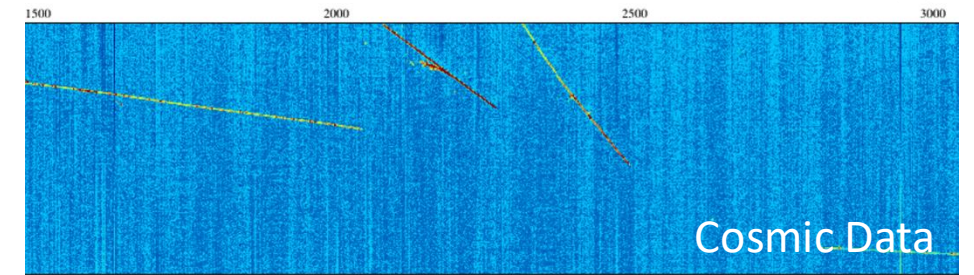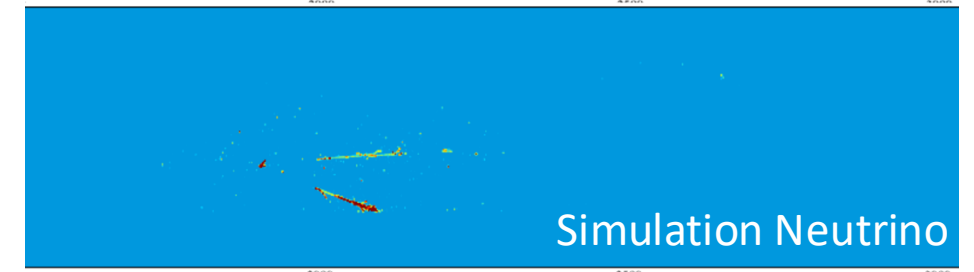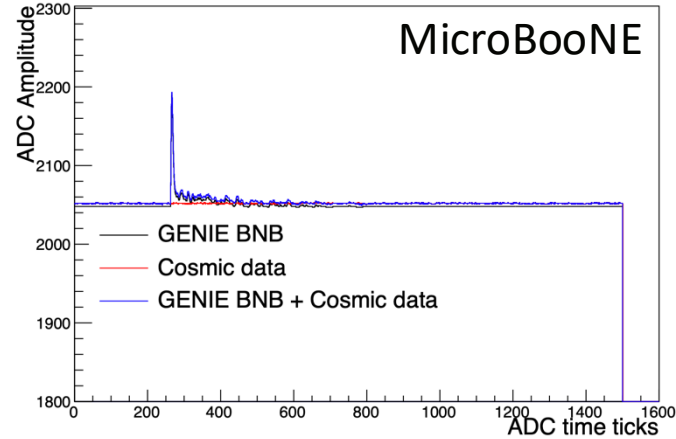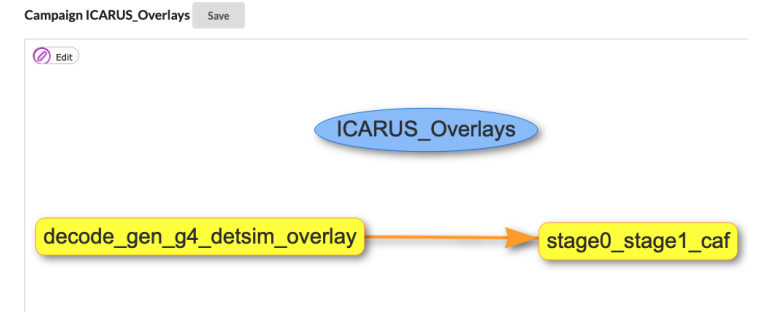
Overlay Event

Raw Data

Merging Signals

Production

Raw Data: Data Collected

- ICARUS uses minimally biased off-beam data to make overlays: offbeam[bnb/numi]minbias

- We collect these types of events in "nominal" data-taking and in special runs dedicated to overlays

- We aimed to take 10x the neutrino triggers
  - Nu trigger = ([Beam Triggered Rate] – [Off-Beam Triggered Rate])

Run 2 Data (Dec-2022->Jul-2023)

Data Stream	Config Name	# of events(triggers)
bnbmajority	Physics	1745833
numimajority	Physics	1036576
offbeambnbmajority	Physics	1358848
offbeamnumimajority	Physics	604722
offbeamnumiminbias	Physics	296344
offbeambnbminbias	Physics	2791336

BNB: 7.98x and NuMI: 7.15x

Run 3 Data (Mar-2024->**Today**)

Data Stream	Config Name	# of events(triggers)
bnbmajority	Physics	1256371
numimajority	Physics	1132502
offbeambnbmajority	Calibration_MAJORITY	674042
offbeambnbmajority	Physics	982079
offbeamnumimajority	Calibration_MAJORITY	2794619
offbeamnumimajority	Physics	737720
offbeamnumiminbias	Physics	298886
offbeambnbminbias	Physics	1854200
offbeambnbminbias	Overlay	1242824
offbeambnbminbias	Overlay AL9	2259792

BNB: 20.61 and NuMI 14.32x

- Data from newer runs is incompatible with Run2 data
  - Trigger (ex. adders) and noise are different

- We added around ~3M events of overlay data over the shutdown period
  - This boosts our available datasets **to above 10x** the Run3 neutrino triggers for both BNB and NuMI

- Note: Overlay AL9 config has bug that mislabels data stream metadata to "unknown"

Raw Data: FTS Infrastructure

- ICARUS introduced a new DAQ configuration for generating overlay data:
  - `Overlays_MINBIAS_Standard_4Hz_WithTPCCompression_Run3_00001`
  - Effectively the same as the Calibration config. + 4Hz configuration
- Files with Physics (from Nominal data taking) or Overlays in metadata configuration name are sent to the same tape file family
  - "Physics_" or "Overlay_" ==> FTS will move files to File Family: data_artroot_raw_offbeambnbminbias
    - Everything else (For example Calibration config) ==> File Family: raw
  - Only applies to offbeambnbminbias data stream
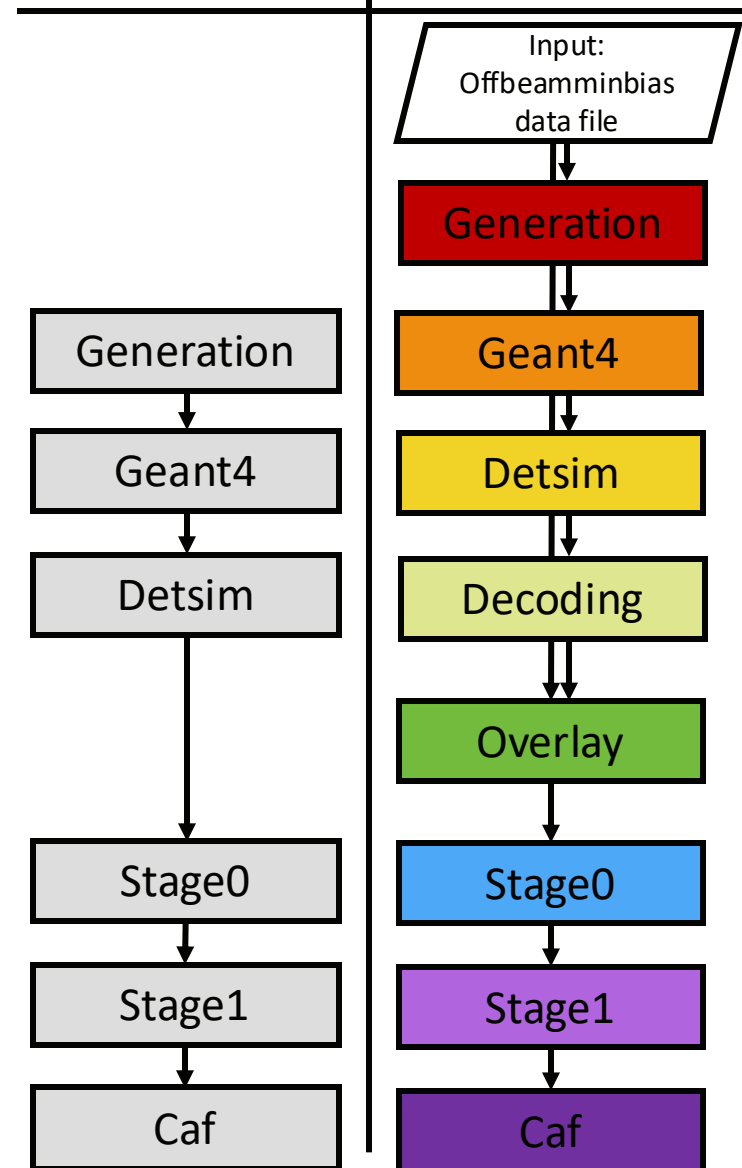- **Faster readback from tape when we want to generate new MC in the future**

# Production: Workflow

Nominal MC Workflow | **Overlay Workflow**

Overlay Workflow:
- Input: Offbeamminbias data file
- Generation
- Geant4
- Detsim
- Decoding
- Overlay
- Stage0
- Stage1
- Caf

Nominal MC Workflow:
- Generation
- Geant4
- Detsim
- Stage0
- Stage1
- Caf

- Added two steps to the neutrino simulation stages
  - Decoding stage: Decode artdaq fragments into data products
  - Overlay stage: Adding PMT and TPC wire waveforms to the corresponding one from data
- Changes in Detsim fcl file to not include noise simulation. Changes/validations needed:
  - Identify "bad" channels and remove signals from these regions
  - Adjust signals based on electron lifetime from data event (more on this in later slide)

Input: Offbeamminbias data file

1
- Generation
- Geant4
- Detsim
- Decoding

2
- Overlay
- Stage0
- Stage1
- Caf
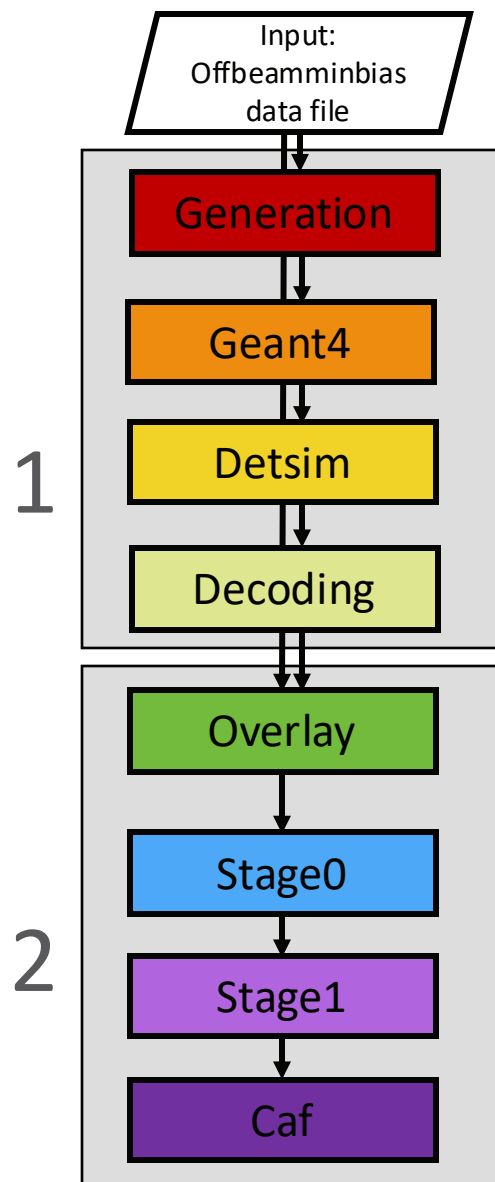
- Updated workflow order for running overlay-making steps:
  1. Generation+Geant4+Detsim+Decoding
     - The decoding of the artdaq fragments is done as the end step, making the job less disk intensive. Now requires ~50 GB of disk
  2. Overlay+Stage0+Stage1+Caf

- Previous workflow:
  - Decoding+Generation+Geant4+Detsim →Stage0+Stage1+Caf
  - More disk intensive on the workernode

- Issues to fix for more efficient use of resources
  - Reduce file size after the Overlay stage
    - File size: ~500 MB/event

- Currently, setup to add data/simulation:
  - TPC: raw::RawDigit (data + simulation)
  - PMT: daqPMT (data) and opdaq (simulation)
  - CRT: crthit::stage0

- Updated Overlay code to run on Run3 data
  - Updated to work with icaruscode v09_90_00
  - Now lives in a [feature branch](#)
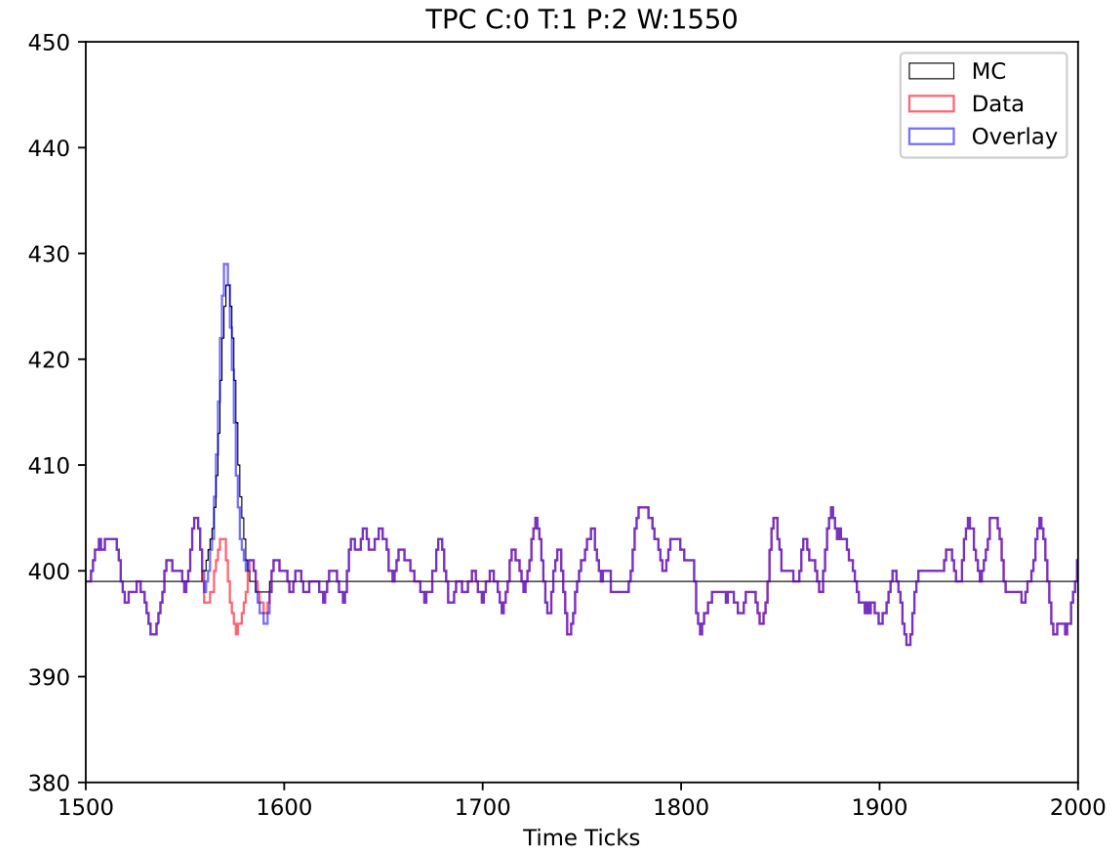
Merging Signals: Electron Lifetime

- The simulated electron lifetime needs to match that of the data in overlay MC

- We have an existing electron lifetime saved per-run in an SQLite database file (from calibration work)

- Gray has integrated a lookup of this electron lifetime into overlay MC (on a feature branch)

- Uses code infrastructure put in place to solve this problem on MicroBooNE

- The code is here:
https://github.com/SBNSoftware/icaruscode/blob/feature/gputnam-overlays/icaruscode/Overlays/ICARUSDrifter.cxx

- Purpose: check that data and simulation are on the same footing
- Plan: make a Corsika in-time simulation + off-beam unbiased data
  - Use CRT-tagged tracks in data and simulation for comparisons
  - Use calibration-tuples for studies
- Updated production workflow to include the use of filter-generated samples
  - For these cases, we want filter-based outputs from simulation (e.g. in-time cosmics, exclusive final states, etc), and simply rejecting events wastes data I/O. ICARUS inherited code from MicroBooNE (Thanks Wes and Joseph) to handle this
    - Generate a lot of events on the worker node, filter them, and write them to a file on the worker node
    - Read in a data file, and using a specialized gallery-based file reader module, copy in data from the simulation file
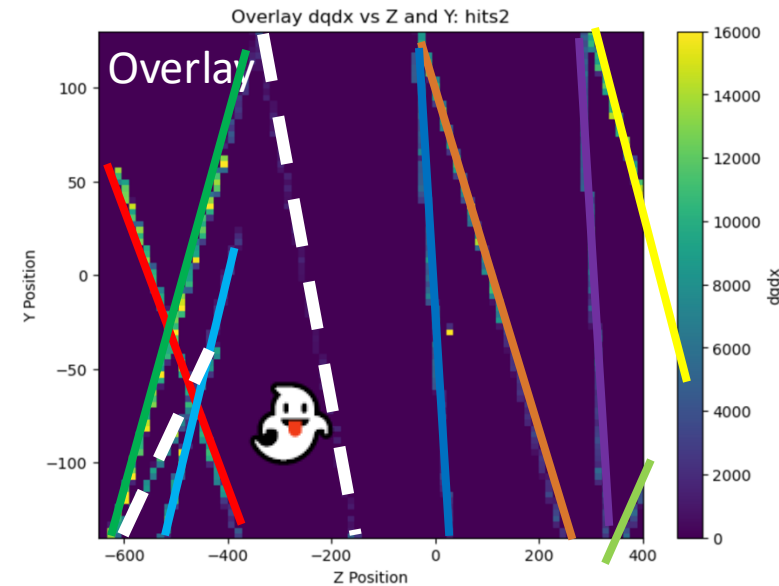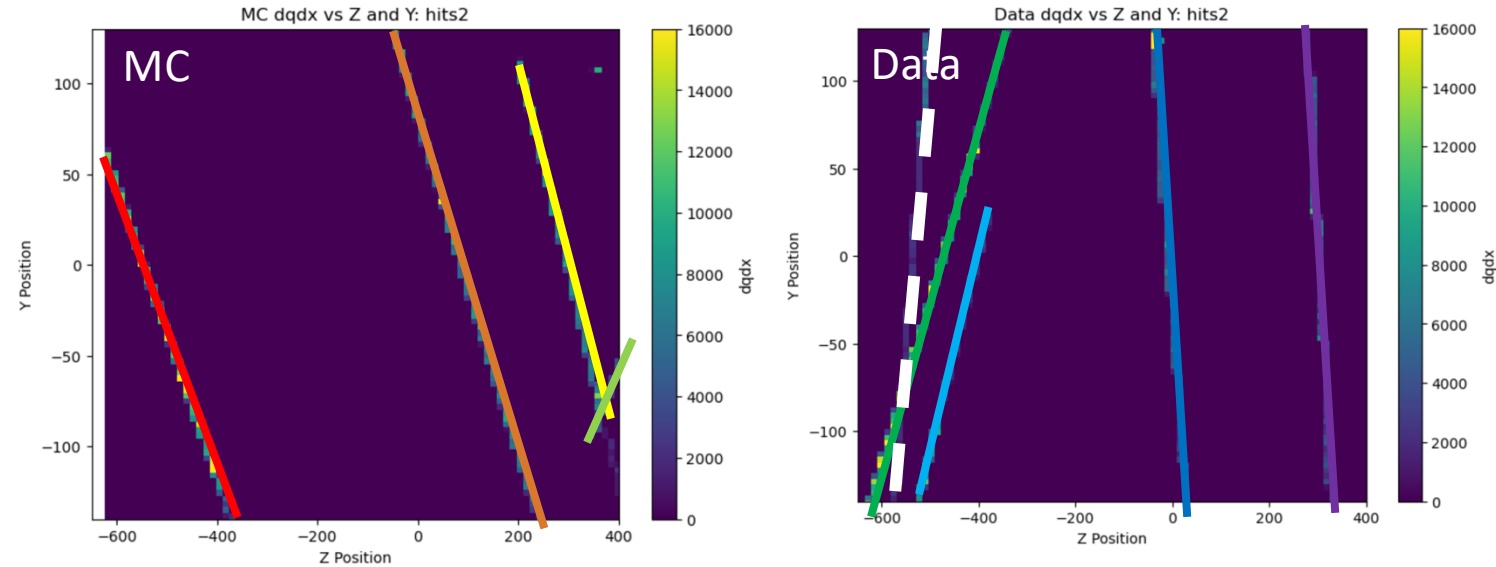
- **One event study** comparing Data, MC, and Overlay YZ hits vs dQdx
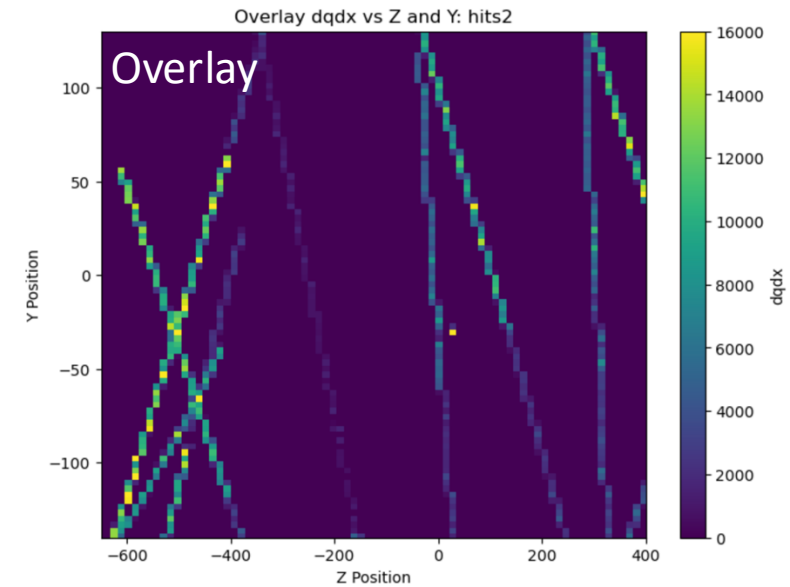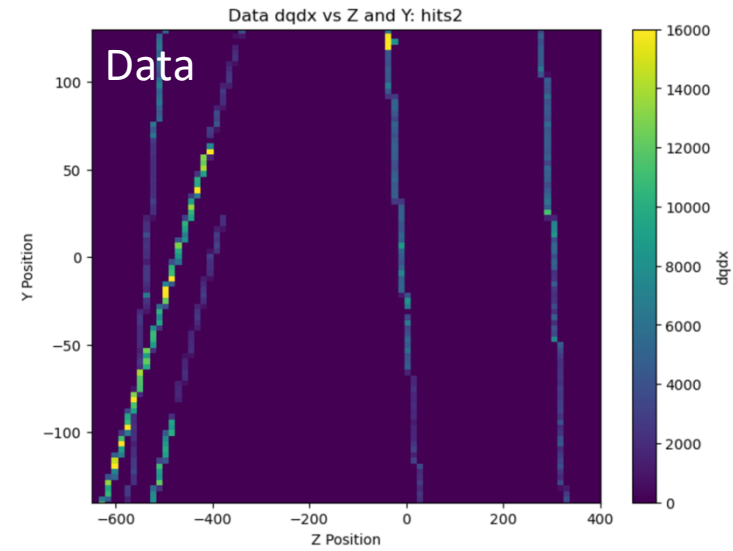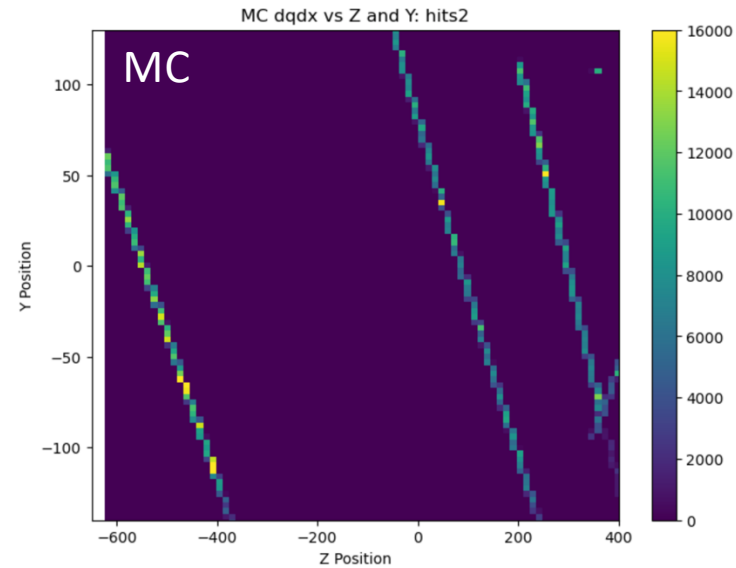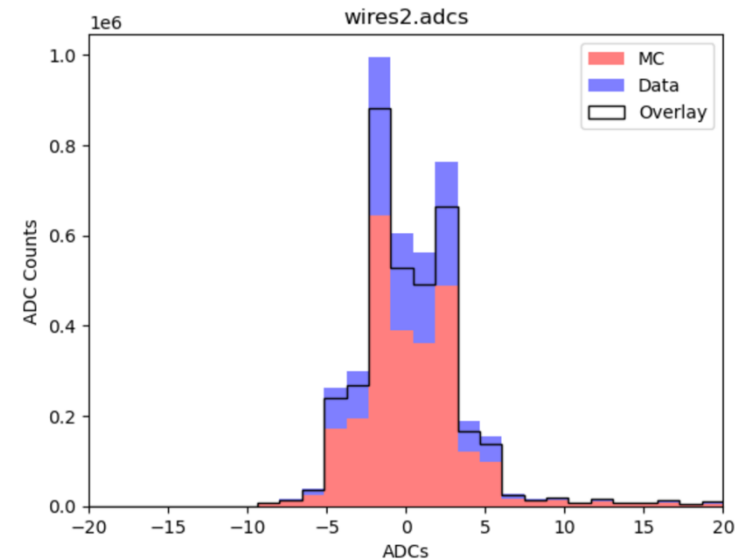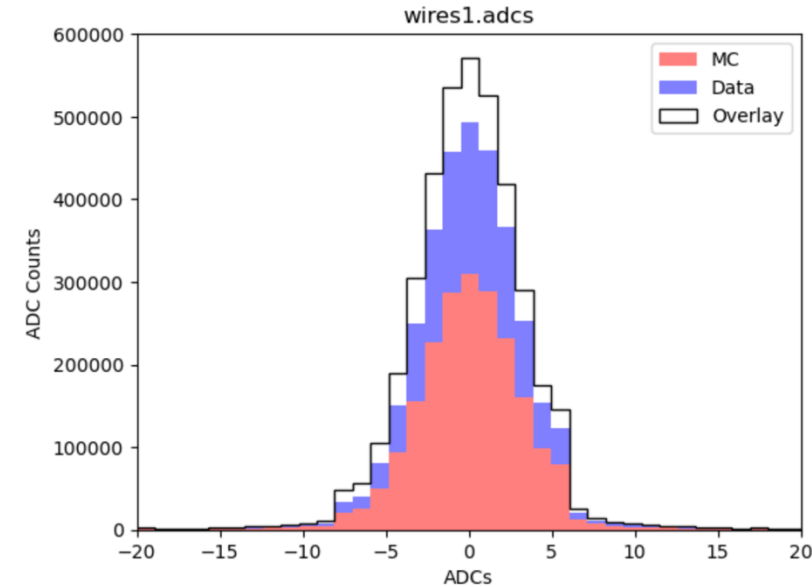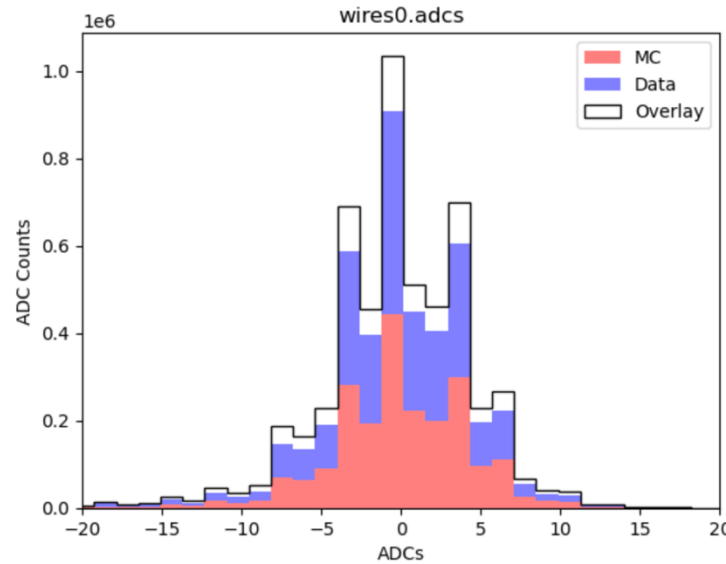
- **One event study** comparing Data, MC, and Overlay YZ hits vs dQdx

- We see the tracks from MC and Data are overall present in overlay

  - Unaccounted (Ghosts?) tracks in overlay event around Z = -600 cm and Z = -200 cm

- Further investigation needed

- **One event study** comparing Data, MC, and Overlay ADC distributions for the 3 wire planes

- The data and MC are stacked histograms with overlayed overlay histogram

- The overlay distribution shape looks like underlying data+MC with some underestimation from data+MC in the first two induction planes

- **One file (50 events) study** comparing Data, MC, and Overlay collection plane hits distribution

- Distribution discrepancies need to be investigated

- Larger statistic samples needed
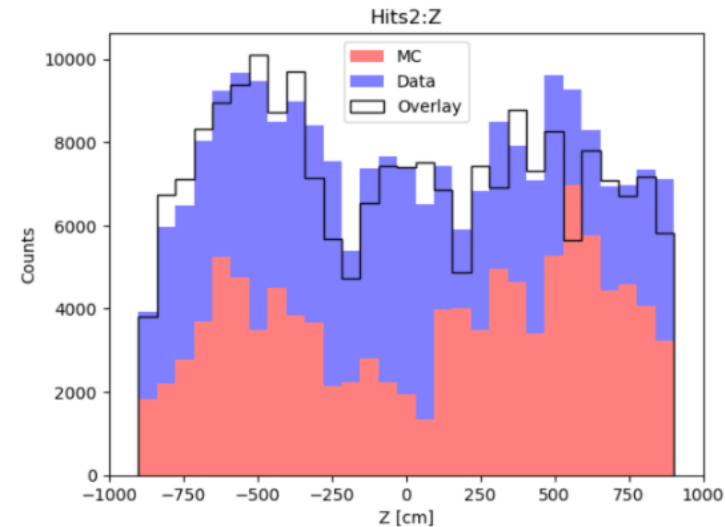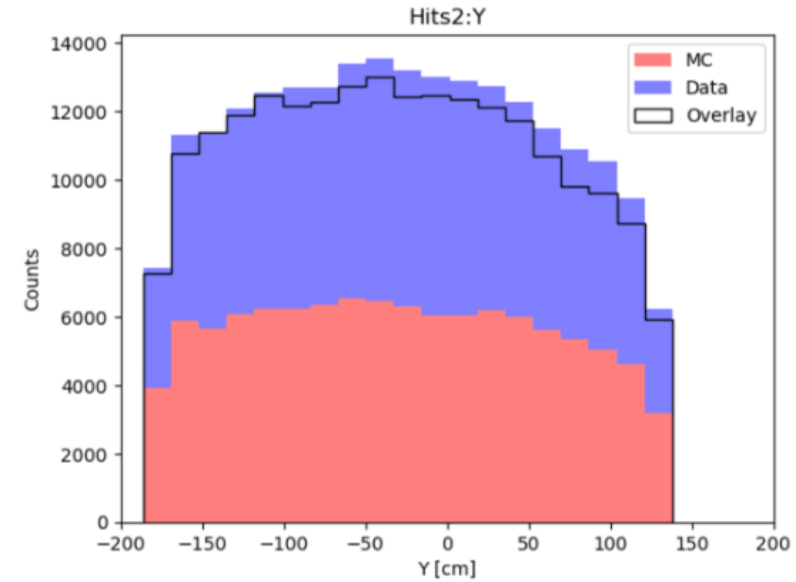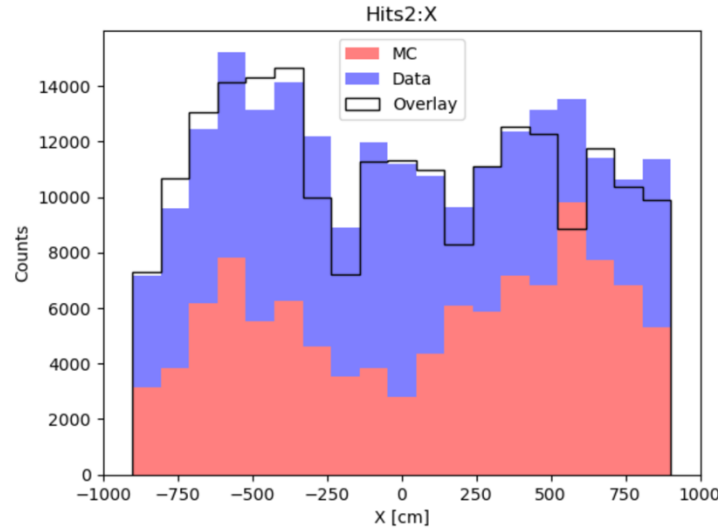
- Make a larger sample of ~1000 events
- List of validation studies (Comparisons between data and simulation):
  - dQ/dx distributions as a function of:
    - Drift location
    - YZ plane
  - Hit efficiencies as a function of track angle
  - Track length as a function of track angle
  - Dead channel checks
    - We need to be careful not to overlay signals over dead channels
- PMT signals need work for validation. Should we use overlays with only TPC and CRT signals validated for our first rounds of analysis?
  - We can end up reprocessing the PMT reconstruction with the updated calibration
  - What data products do we need to enable reprocessing of only the PMT simulation?
- Study the performance of the truth-matching
  - We want to match reconstructed objects to the truth from the simulation

- Validation could take about 2-3 months of effort
  - Most production infrastructure is in place
  - The overlay code is working

- Making overlays using Run2 and Run3 data could take less than 6 months
  - Takes ~6 months to reprocess ALL Run 2 and 3 data (From yesterday's discussion)
  - The data needed for making overlays is a subset of that: off-beam minbias data stream
  - Run2 data lives on disk -> No prestaging needed
  - Run3 data needs to be prestaged

- We would also need the YZ simulation for both validations and production of overlays

- ICARUS has taken enough off-beam unbiased data to make overlays to cover for Run3: BNB: 20.61x and NuMI 14.32x Run3 neutrino triggers
- Updated FTS to store overlay input data in dedicated file family
- Updated overlay code to allow data and filtered generated events to be on the same artroot event level
  - Use code infrastructure from MicroBooNE
- Validation studies started. Expect more updates in future meetings
- To-Dos:
  - Make a larger validation sample (~1000 events)
  - Integrate Gray's electron lifetime lookup code to the main overlay code
  - Incorporate the YZ simulation when available
  - Produce validation studies with larger statistics
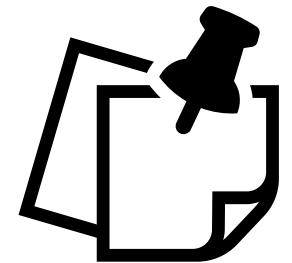
Anyone interested in joining is welcome to talk to me or anyone on the team.
Slack Channel: #icarus-overlay

Backup