Status of Software development

D. Gibin[1], T. Usher[2]

[1] Department of Physics and INFN Sez. Di Padova,
[2] SLAC National Accelerator Laboratory

ICARUS General meeting 11 September 2019

groups organization

- The ICARUS simulation and reconstruction group is responsible for the preparation and maintenance of the software tools needed for the simulation and the reconstruction of the data collected in the experiment.
- Some finer structure with subgroups operating in parallel

ICARUS simulation and reconstruction
Daniele Gibin - Tracy Usher

- **Wire signals processing , 3D reconstruction and tracking** — *Filippo Varanini*
- **Light Detection Systems Simulation & Reconstruction** — *Alessandro Menegolli*
- **CRT simulation and reconstruction** — *Chris Hilgemberg*
- **energy meas., e.m. activity ID and measurement, PID, purity meas.** — *Christian Farnese*
- **Event production** — *Maya Wospakrik*

- Weekly meetings on Monday, to share the status of all the activities, discuss problems and planning

Status of the activities - 1

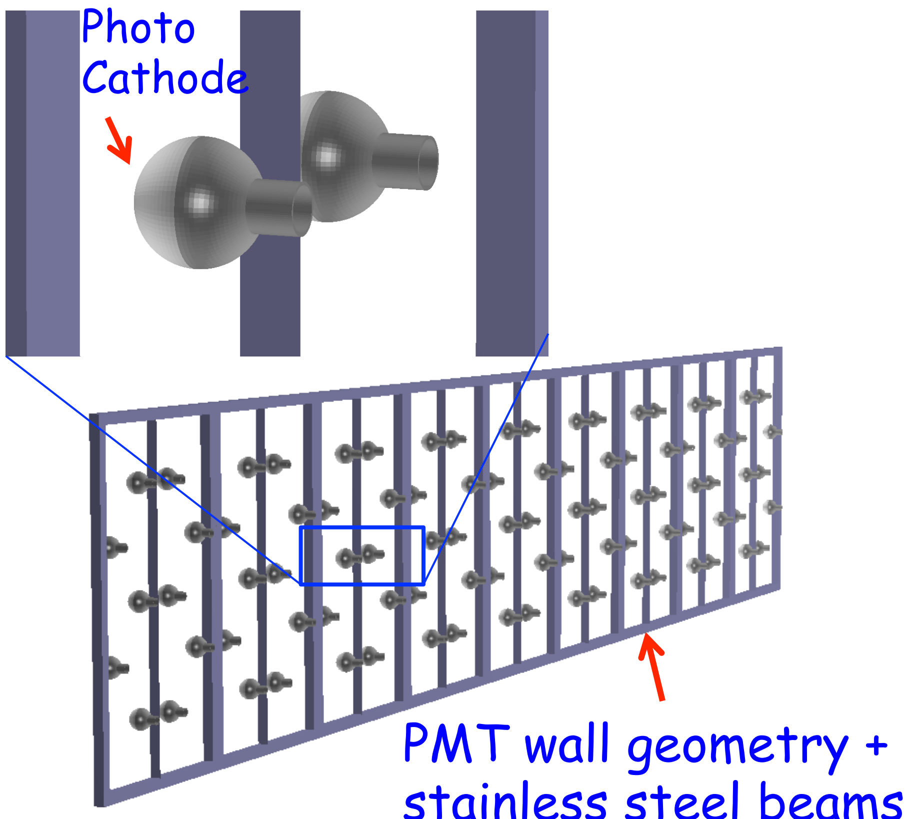
- The present activities are concentrated on the simulation of events to be ready for the arrival of the real data.
- The basic functionality for the simulation of ν interactions (BNB and NuMI beams) and cosmic rays crossing the detector is in place and operational. In particular:
  - ✓ Description of the signal provided by the new wire front end electronics. Wire noise model from past running and from testing of the new electronics;
  - ✓ PMT "library" allowing reasonable simulation of the light response in both cryostats;
  - ✓ Working CRT simulation
  - ✓ Not yet a Space Charge model (coming soon)

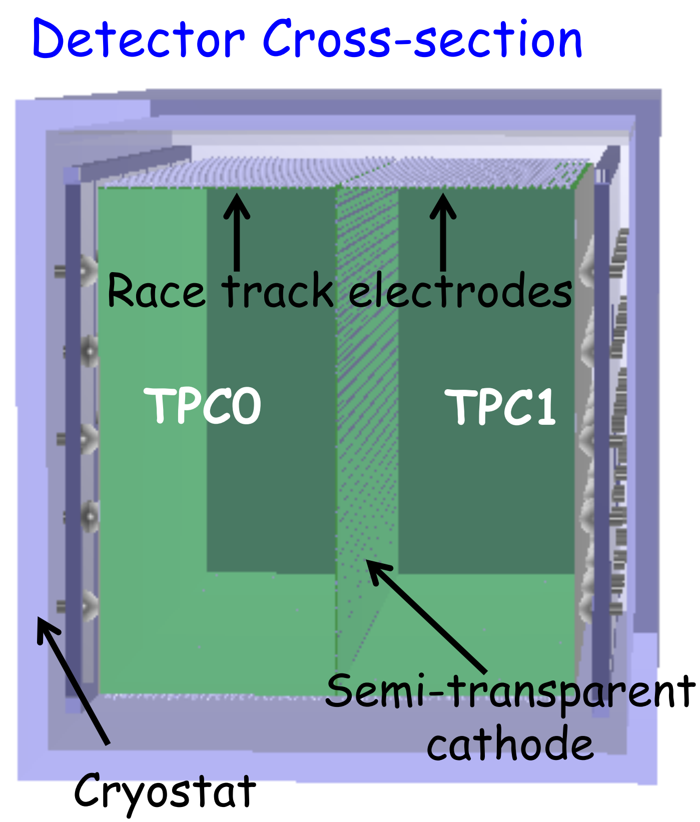
Status of the activities - 2

- Basic signal processing chain available:
  - ✓ A suite of noise filtering tools available, inspired also by the features observed in past ICARUS run;
  - ✓ Original “raw hits” path developed from the original ICARUS hit finding and fitting and treatment of the now bipolar induction signals;
  - ✓ 1D deconvolution algorithm to convert the induction signals to unipolar signals with good response and
  - ✓ Using the LarSoft hit finding algorithm
- We have optical hit and light flashes reconstruction
- We have integrated the Basic tracking and shower reconstruction
- The code to create 3D Space Points which are being fed into the deep learning effort is running
- We have some final stage code now running
  - ✓ Track fitting
  - ✓ Track momentum determination from MCS

Present detector description

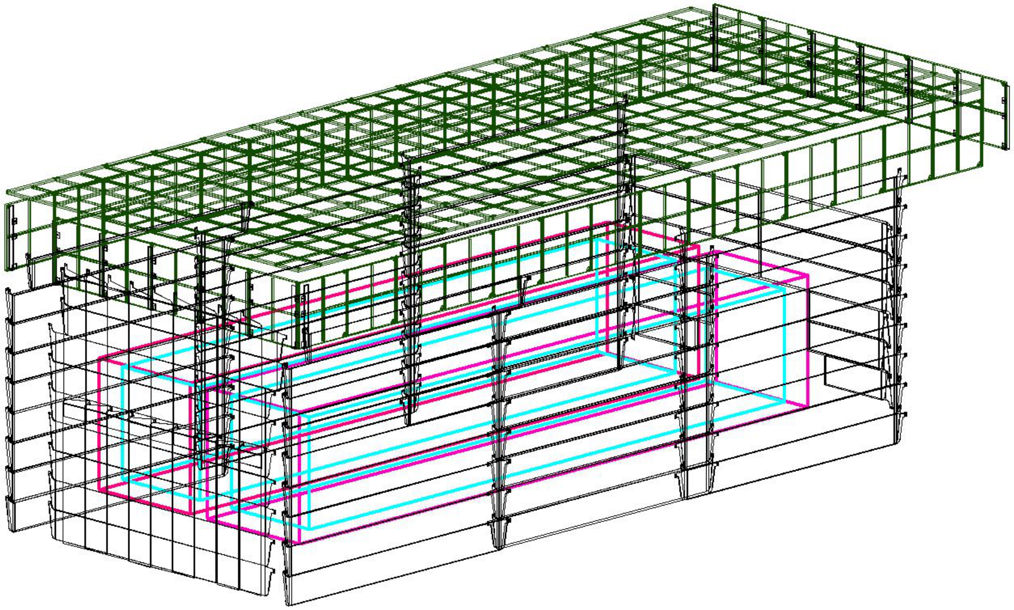
- Geometry «stable» version since 29 April 2019 includes:
  - Correct size/composition/mass of materials surrounding active LAr;
  - Realistic description of PMT;
  - Stainless steel chamber lateral mechanical structure.





CRT simulation and reconstruction

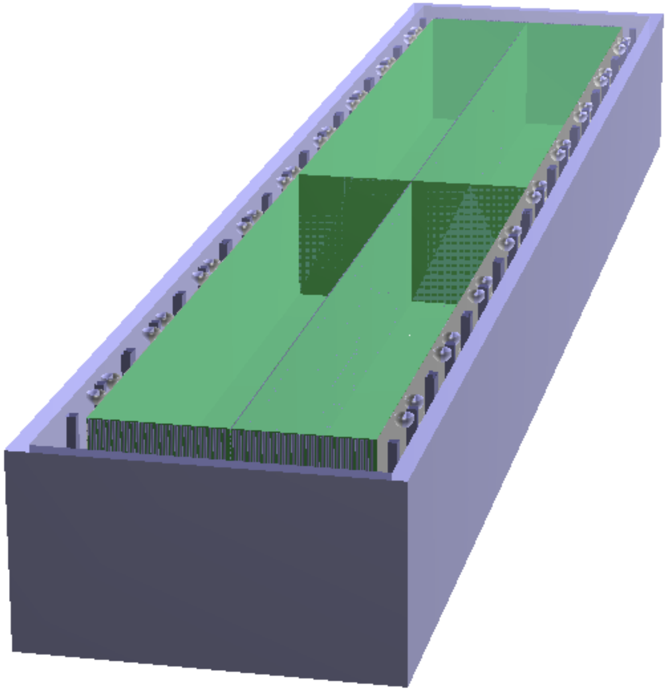
- The geometrical description of the CRT is quite realistic, including support structures and drawings of the detector

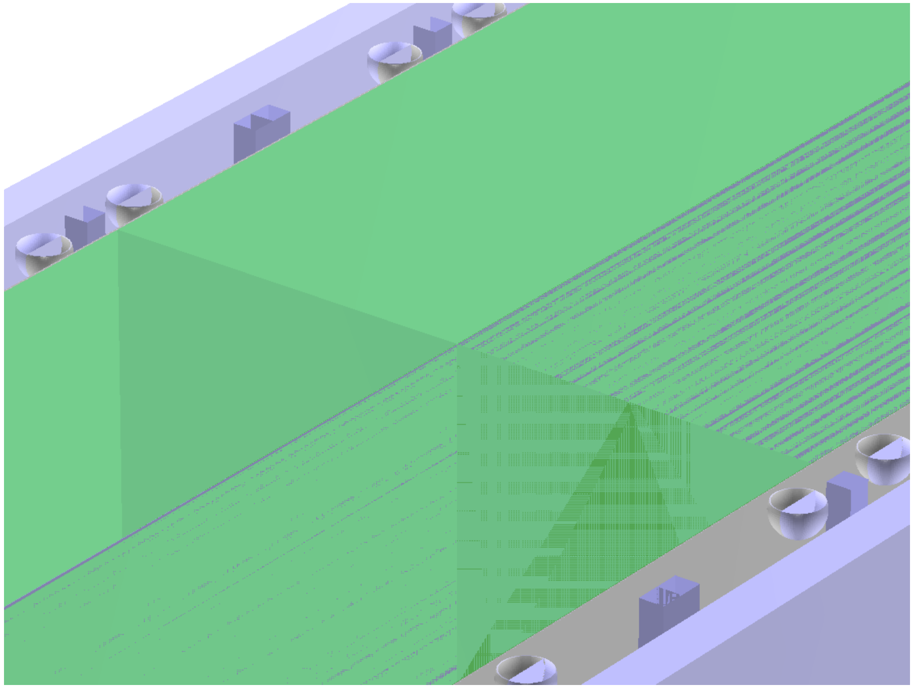


- Detector simulation has been validated with data from CRT test stand
- Reconstructed CRT hit are now available and are being analysed to evaluate detector tagging efficiency in more realistic way than exploiting truth information in MC
- See tomorrow presentation by C. Hilgenberg for updates on cosmic μ tagging efficiency and on autoveto probability for beam ν interactions.

Updated TPC geometry

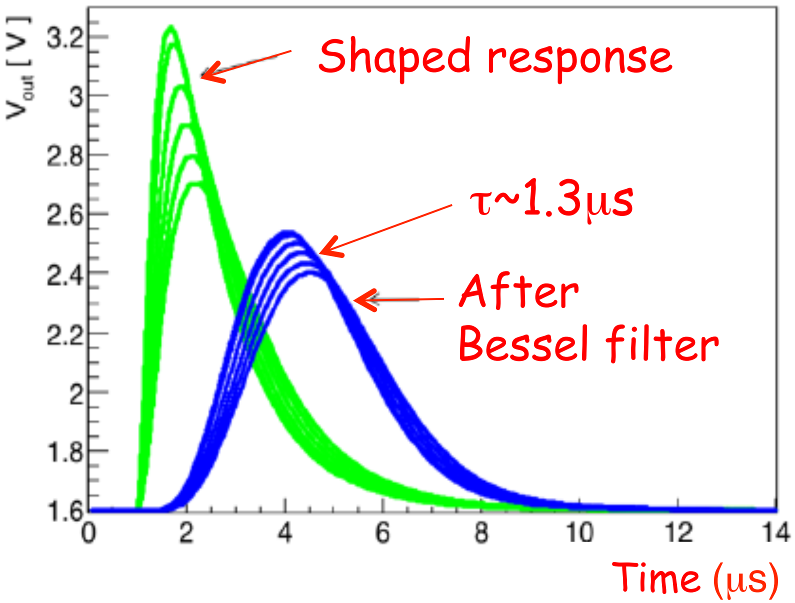
- Ongoing tests of a new version of the detector description including split wires for the Induction 1 plane (presently assuming single wires stretched along the full detector length)
  - Requiring to split also the Induction 2 and collection plane and introducing 4 TPC volumes
  - Signals from wires crossing adjacent volumes are then merged together (details in tomorrow presentation by A. Menegolli)





wire signal: from ionization to digitized waveforms

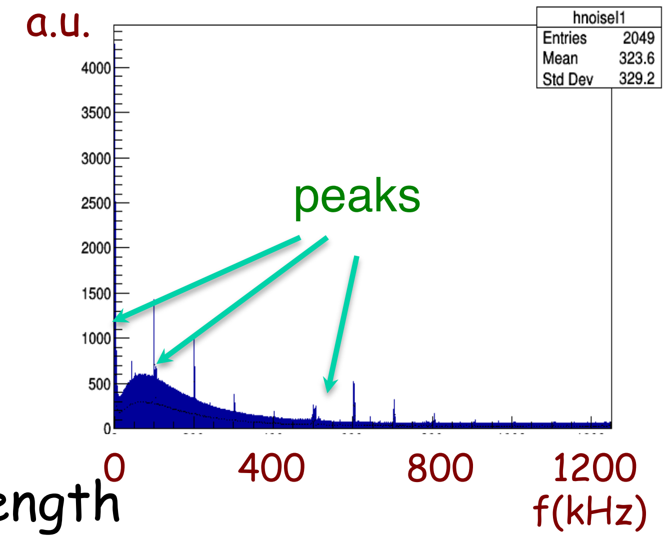
- The wire signal simulation includes:
  - distribution of ionization electrons generated by GEANT4
  - the effects of electric field in proximity of each wire-plane from an electrostatic simulation (by L. Rochester), neglecting neighboring wires effects
    - Signals on neighboring wires will be introduced soon
  - the response of the readout electronics, identical for all three planes and modeled as ***$t*\exp(-t/\tau)$***, with τ = ***1.3 μs***, from numerical approximation of shaping + Bessel filter
  - Absolute normalization taken from 50-liter TPC measurements at CERN.



For details seee the electronics paper **JINST 13 P12007(2018)**

ICARUS wire noise simulation

- Simulation includes as a first approximation the electronic noise derived from past measurements: ICARUS@LNGS + 50-liter chamber at CERN
- Frequency spectrum from FFT of LNGS noise (in Induction 1 plane) assuming:
  - Smooth component totally uncorrelated Peaks (~100 kHz harmonics) as totally correlated within a board (32 wires)



- RMS normalization (2.3 ADC#) from CERN measurements, corrected for different wire length
- New model to accommodate noise measurement directly from the data collected during the commissioning and tune the reconstruction in as realistic as possible working conditions
- More details in F. Varanini talk tomorrow.

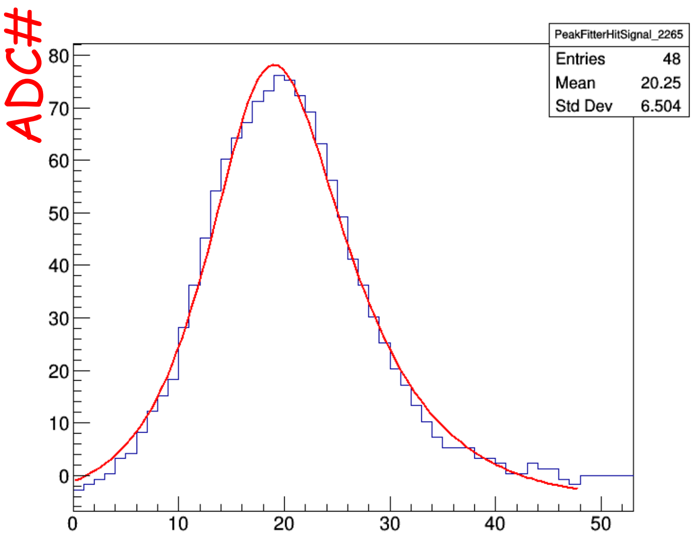
Wire signal processing

- Two different and complementary approaches are available in the code:
- Raw (used by ICARUS@LNGS): no deconvolution takes place
  - Hit-finding directly performed on raw signal from wires
  - Different hit shapes for different wire-planes
  - Offline integration/filtering of Induction-2 signal
- Deconvolution: response functions are first inverted to (ideally) recover the intrinsic drift electron signal. Then:
  - ROI identification and hit-finding performed on deconvolved signal
  - MIP hit shape is Gaussian and independent of wire-plane
  - Used in SBND/MicroBooNE: software tools can largely be shared

Raw hit finding algorithm

- Inherited from previous ICARUS algorithms, ported into Larsoft
- Based on "time-over-threshold" logic (parameters depend on wire plane)
- The bipolar signals of Induction 2 are transformed to unipolar shape by time integration followed by an appropriate baseline restoration
- Good performances for Collection and Induction-1 (>98% efficiency, <0.1 fake hits per wire) for "standard" noise conditions used in simulation
- This will have to be reviewed for higher noise (especially in Induction-1)
- Reduced performance in Induction-2 (efficiency ~90%, ~1 fake hit per wire in the whole drift region)
- Collection fitting function considers double exponential

$$f(t) = B + A \frac{e^{\frac{-(t-t_0)}{\tau_1}}}{1 + e^{\frac{-(t-t_0)}{\tau_2}}}$$



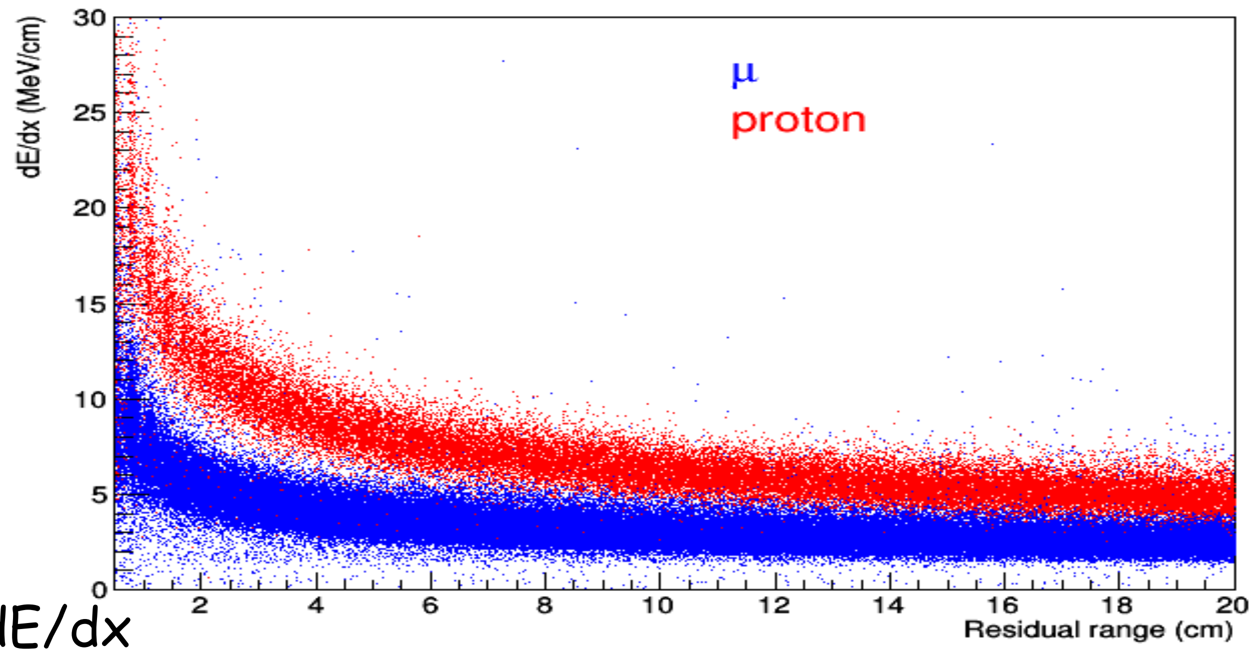
Sample-initial sample

- See tomorrow talk by F. Varanini

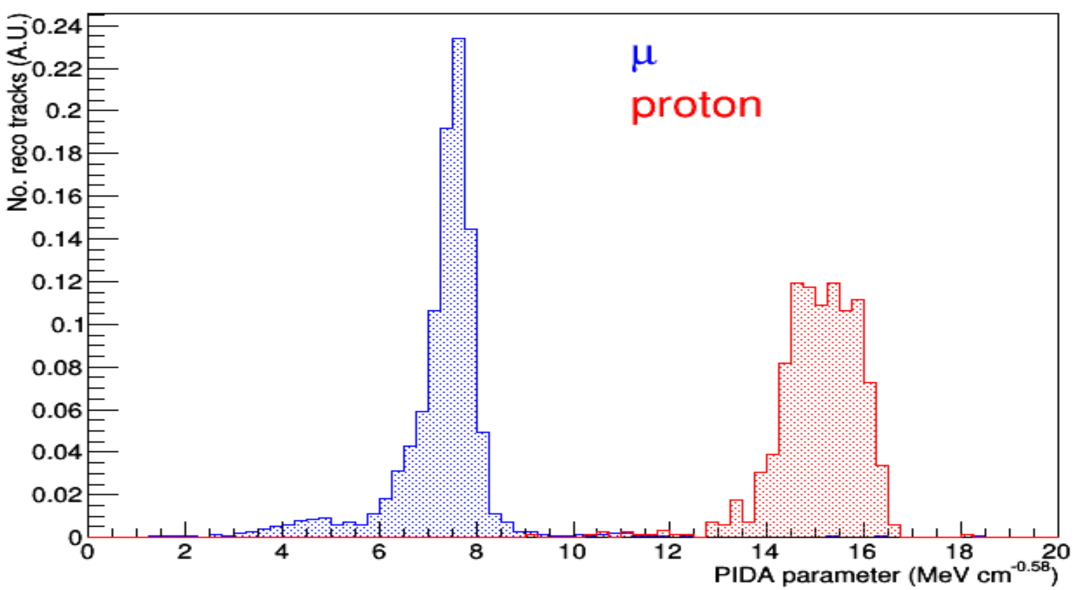
Deconvolution – 1 dimension

- Goals of deconvolution algorithm:
  - Convert raw signals into unipolar signals, nominally with a gaussian shape for all planes
  - Regularize the response across the three planes
- The method corrects for the electronics and field response components of the signals observed on the wires (the "raw" signal).
  - Diffusion, lifetime, recombination handled separately
- Deconvolved waveforms searched for candidate peaks which are fit to one or more gaussian shapes
- Current implementation treats each wire independently and does not consider possible induced effects (due to track angles) from neighboring wires.
  - Works best for tracks at relatively shallow angles to the wire planes
  - "Good enough" for commissioning and development of reconstruction algorithms
- WireCell Team are working to integrate their toolkit into the ICARUS system which will bring 2 dimensional deconvolution in the future.

Preliminary PID with dE/dx for stopping particles

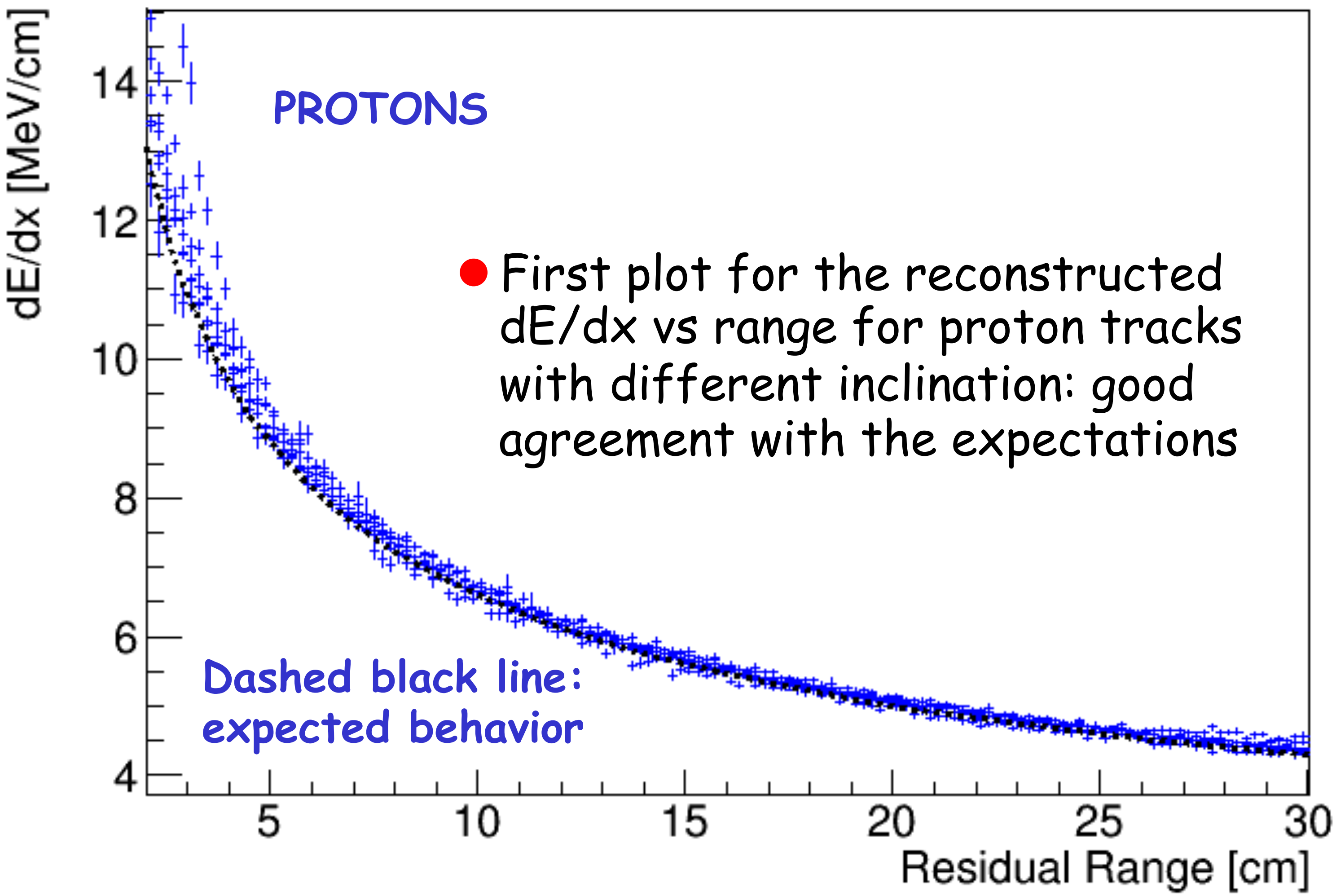


- First plot for the dE/dx vs range and for the particle identification test obtained for protons and muons.



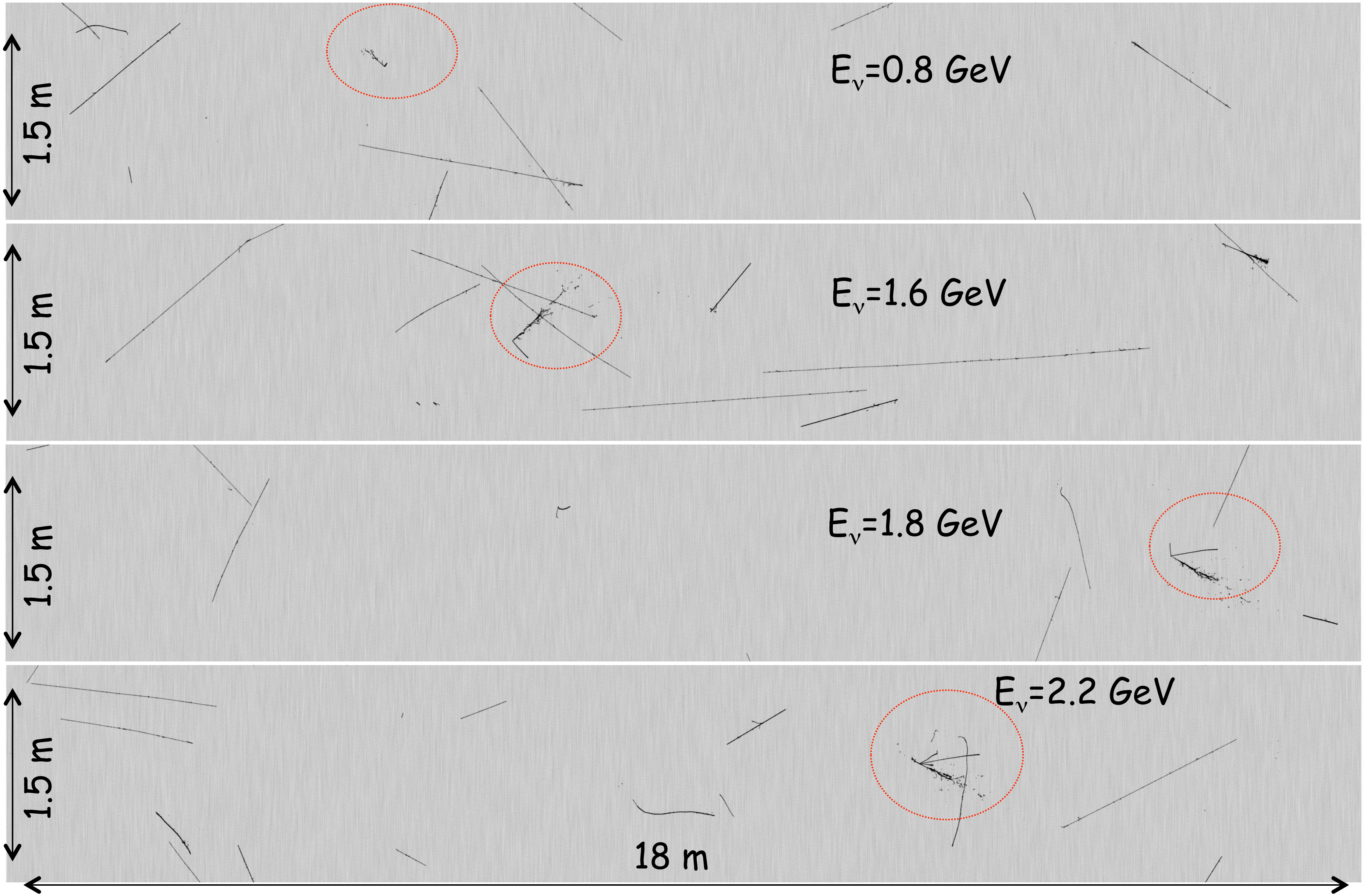
> In this analysis only the Collection view has been used for the calorimetric reconstruction

dE/dx vs range studies for protons tracks



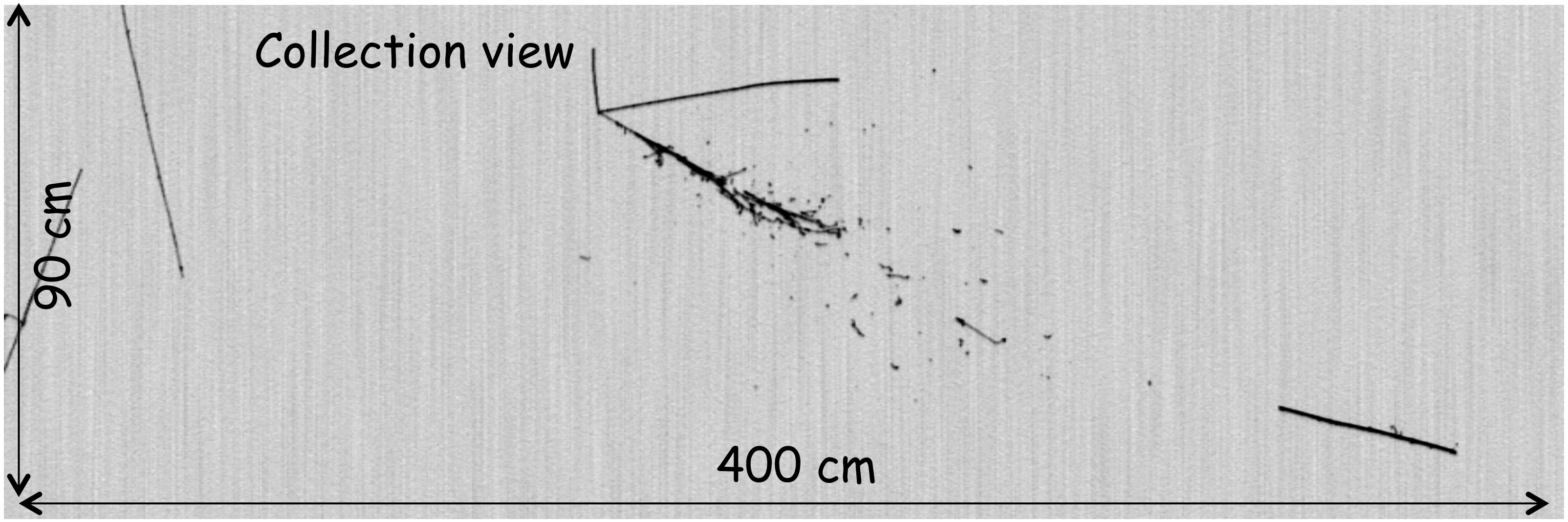
- First plot for the reconstructed dE/dx vs range for proton tracks with different inclination: good agreement with the expectations

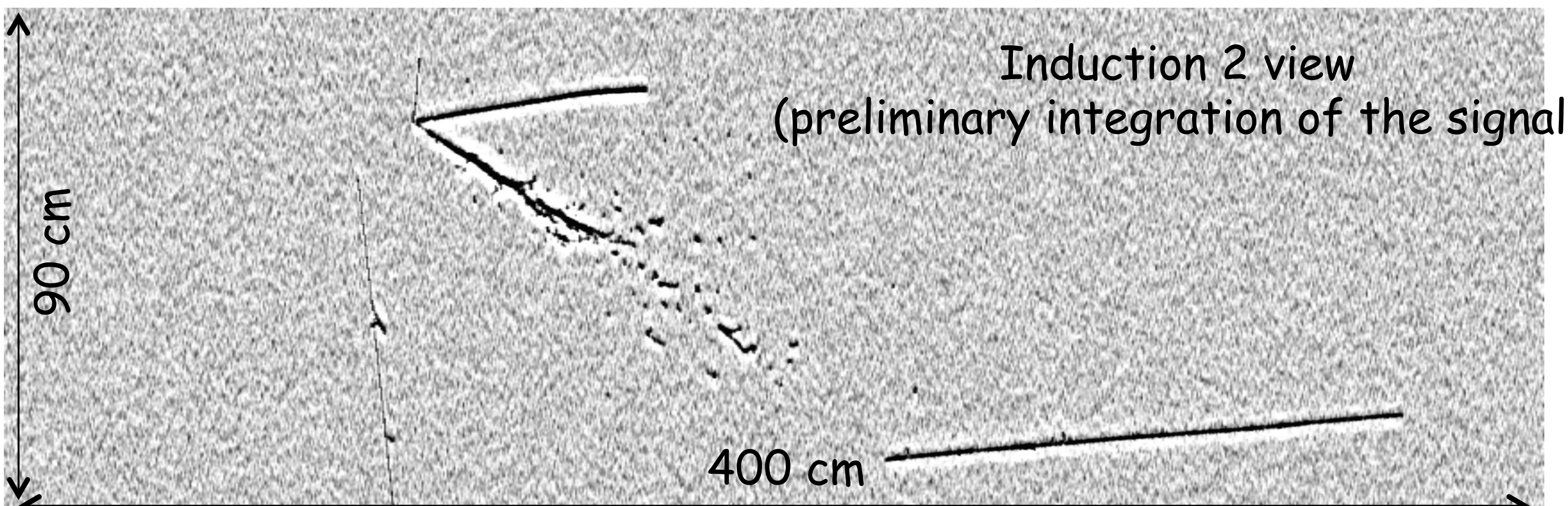
Few examples of simulated BNB events and cosmics in ICARUS



- Examples of MC νeCC events overlapping with cosmics (collection view)

Detail of one event





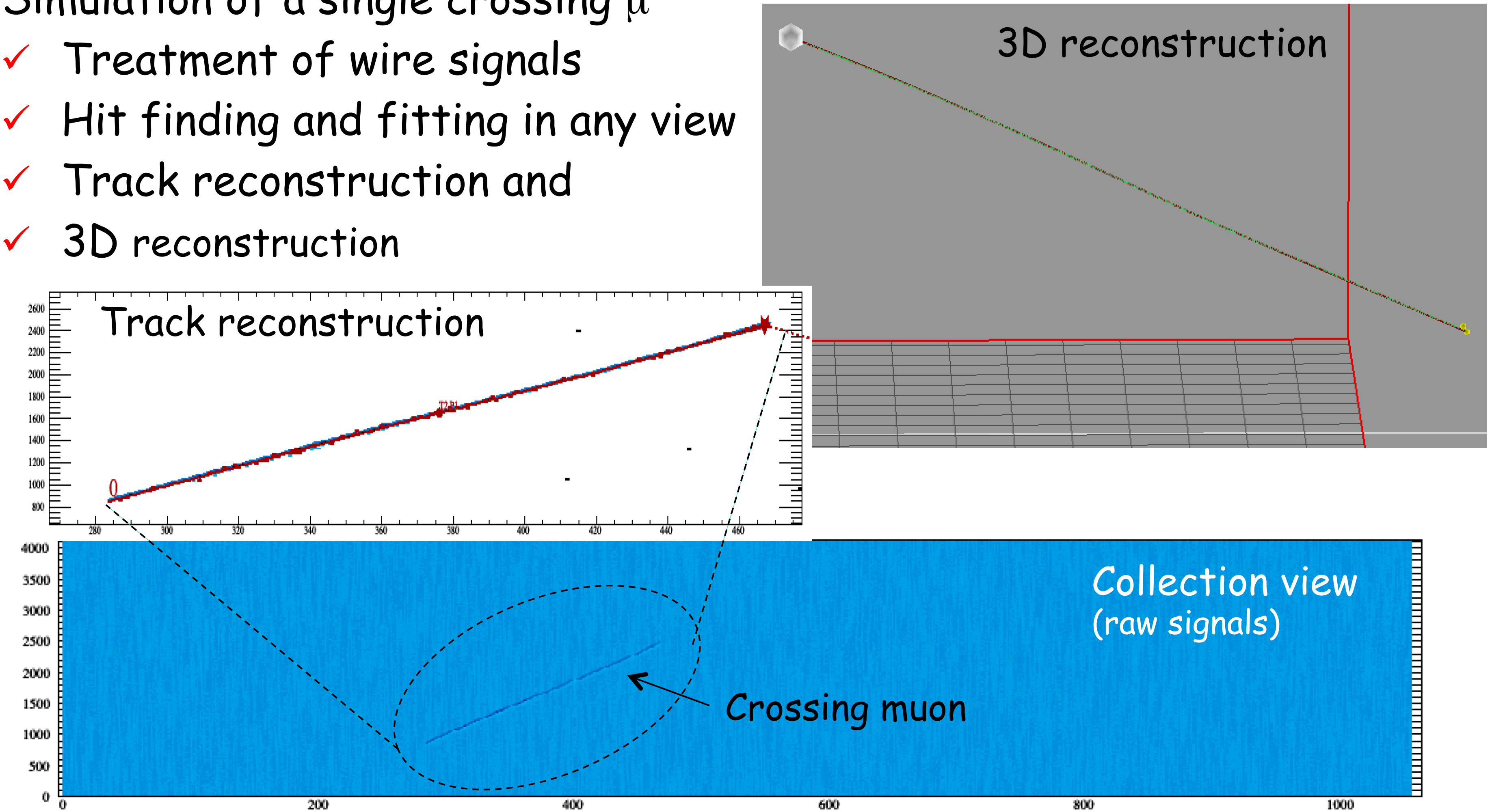
- Close view of one 1.8 GeV νeCC

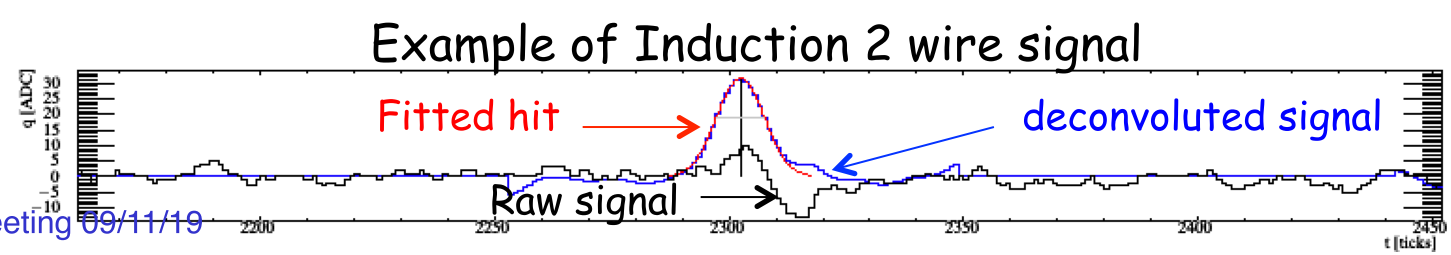
Track/Shower Reconstruction

- The hits are the final result of the signal processing
- To reduce noise hits (particularly from the middle induction layer), hits from each plane are combined to form SpacePoints
  - Fake hits from noise will not form valid SpacePoints and can be removed
- Resulting collection of hits are input to the Pandora framework
  - Three stages: overview, track finding, shower finding
  - Track and shower pattern recognition package
  - Returns candidate vertices for tracks/showers/particles
  - Creates particle hierarchy (PFParticles) for "connected" tracks and showers – describes structure of, for example, neutrino interactions
- Track candidates passed to track fitter to form fully fit track trajectories
- Shower candidates pass to SBN Shower reconstruction – *NEW!*

Example of simulation and reconstruction

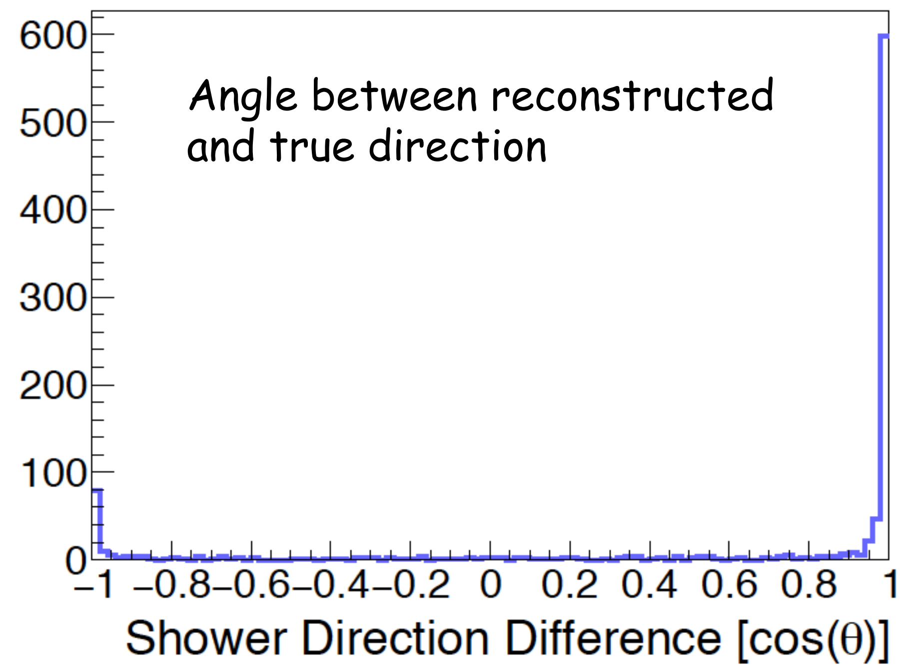
- Simulation of a single crossing μ
  - ✓ Treatment of wire signals
  - ✓ Hit finding and fitting in any view
  - ✓ Track reconstruction and
  - ✓ 3D reconstruction

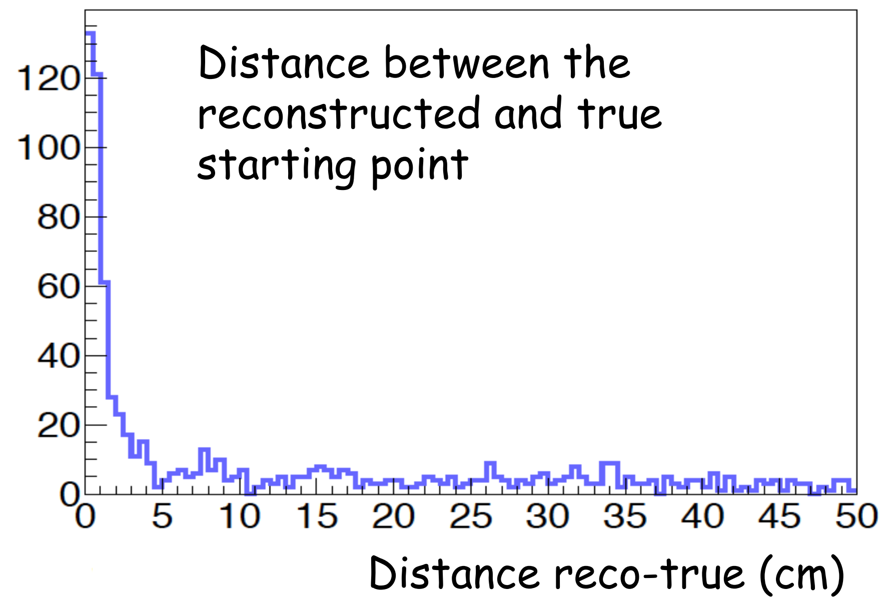




SBN Shower Reconstruction

- Successful application of the SBN shower algorithm to ICARUS simulation for electron π events
- The out of the box exercise without any special tuning looks promising even if to get performance adequate for the analysis additional customization and tuning are required
- See tomorrow talk by YT. Tsai



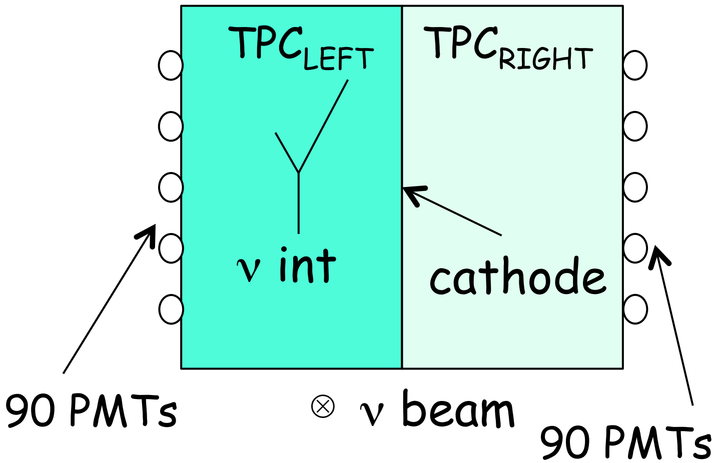


MCS momentum measurement

- Crucial for the energy reconstruction of $\nu\mu$CC interactions with not contained muons
- ICARUS algorithm (see paper ***JINST 12 P04010(2017)***) allowed measurement with average ~15% resolution over 0-4 GeV/c range
- Several relevant features w.r.t to other algorithms available in LArSoft:
  - Event-by-event estimation of drift coordinate measurement error
  - Event-by-event optimization of segment length
  - Two estimations of deflection angle (linear fit+segment barycenter)
  - χ^2-like function comparing observed/expected deflections as a function of momentum
  - Full treatment of covariance between consecutive angle deflections
- Implementation into larsoft is being finalized
- Test on fully simulated muons and characterization of resolution as a function of track length/momentum will soon follow

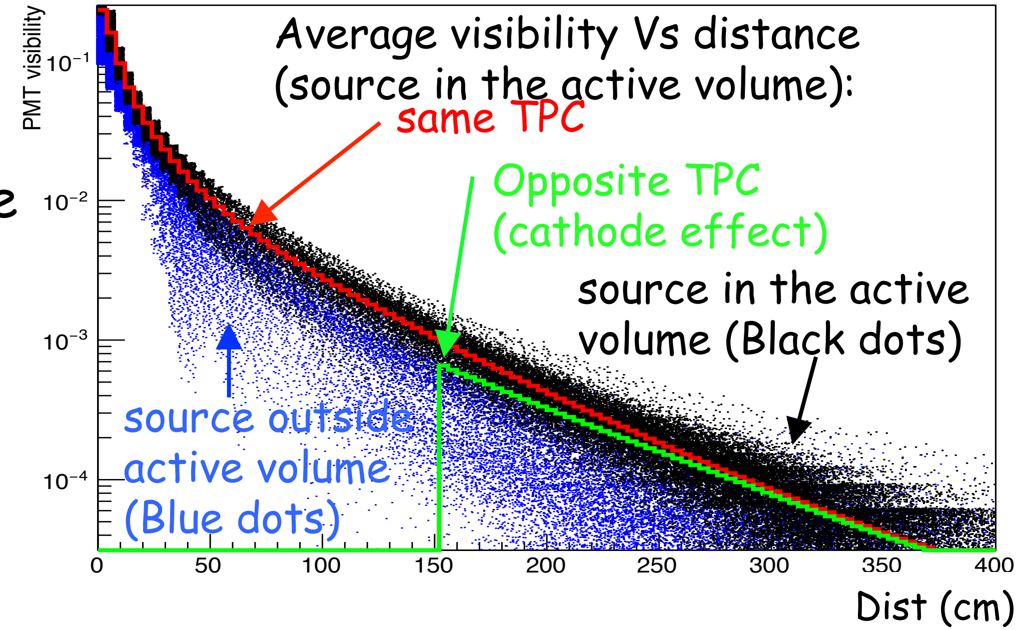
LAr scintillation light simulation

- A computer-intensive simulation is first performed: 10^5 scintillation photons are generated in a grid (5x5x5 cm^3 cells) over full LAr volume and traced accounting for the detector geometry, surface reflectivity (0 at present), absorption, Raileigh scatt., cathode transparency etc.
- The probability (*visibility*) that a photon generated in a given point of the LAr hits any PMT is computed as a 3D map called photon library
- photo-electrons generated by ionizing tracks on each PMT are parameterized as the visibility of their production point (stored into the photon library) × the PMT quantum efficiency (7% at the moment).
- The arrival time distribution is then parameterized as a function of the distance from the light production point and individual PMT
- Optimization of the access to the photon library and of the required memory (using the symmetry between the two modules) allows to exploit this approach also in case of the crowded cosmic ray events (see talk by G. Petrillo tomorrow)

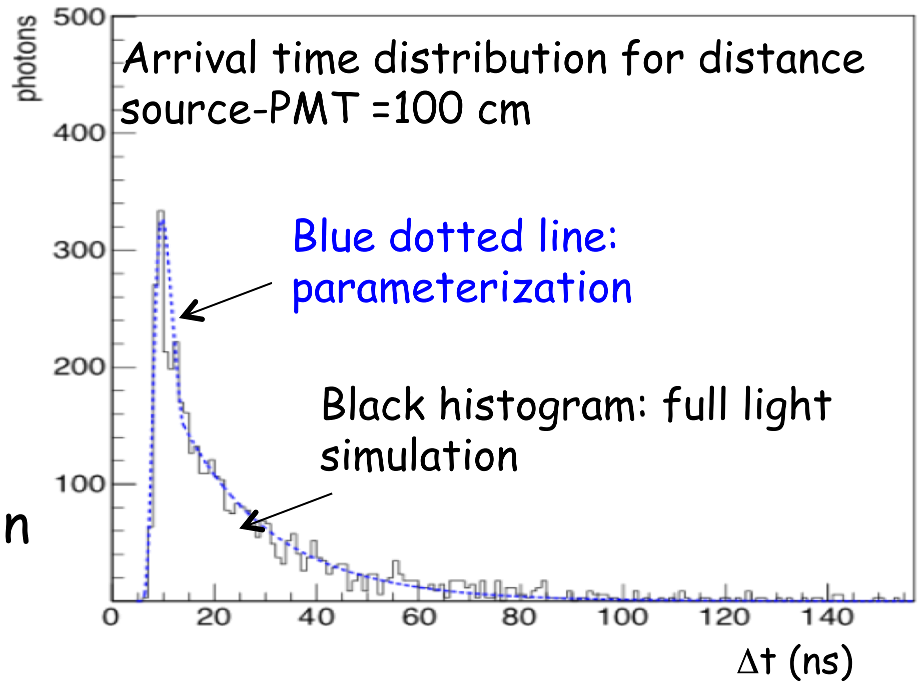


LAr scintillation light simulation

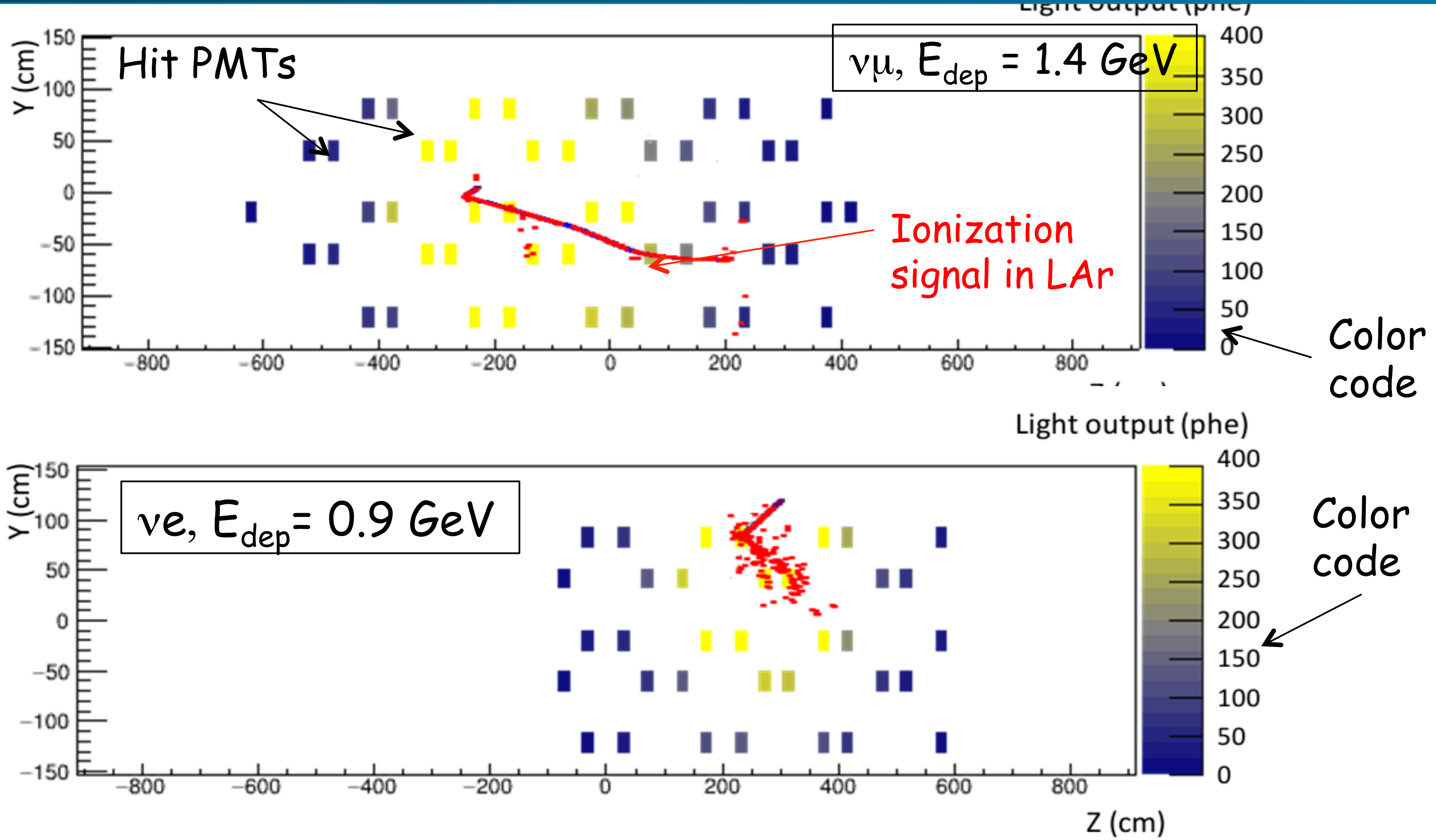
- Visibility mostly driven by distance light source-PMT, with large fluctuations reflecting the detailed geometry
- 58% cathode transparency
- Passive LAr volume screened by the race tracks



- The time distribution of photons on each PMT parameterized by members of the SBND coll. as a function of the distance between the production point and the PMT
- This parameterization appears to be working also for the ICARUS simulation as shown by a comparison with full light simulation

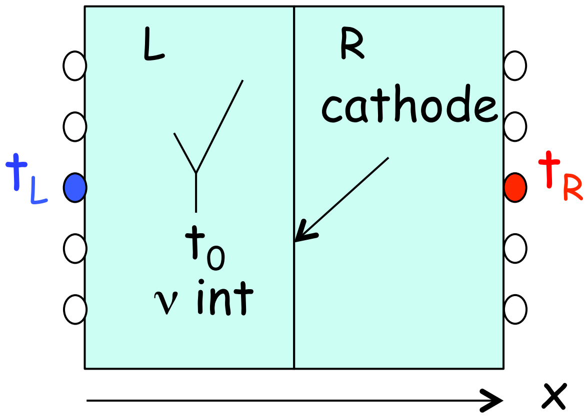


MC event simulation of TPC charge tracks and PMT hits

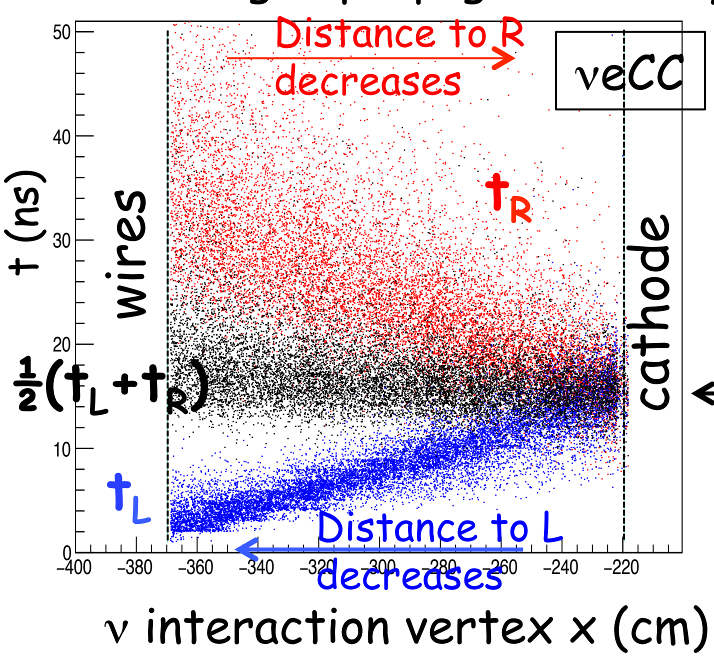


- Projection of the simulated events on the TPC wire plane:
  - The ionization signals in the LAr are shown in red;
  - The hit PMTs (> 10 phe) are represented as boxes with a color code for their signal amplitude.

First MC computation of the event time from inner PMTs



Scint. light propagation delay



- The light propagation from the tracks to PMTs introduces a delay between the time t_0 of ν interaction and the first PMT light signal detected in the left and right chambers, t_L and t_R.
- This delay increases with the track distance from PMTs but the sum $(t_L+t_R)/2$ is approximately independent from the position of ν interaction along the drift direction.
- $\frac{1}{2}(t_L+t_R) \sim t_0+(16\pm2)$ ns ➔ can provide a first evaluation of ν interaction time t_0, almost independent from the position of the ν vertex and with a resolution of ≈2 ns.

$\frac{1}{2}(t_L + t_R) \sim t_0+16$ ns

(see talks by A. Menegolli and G. Petrillo tomorrow)

Further improvements on Light simulation and reconstruction

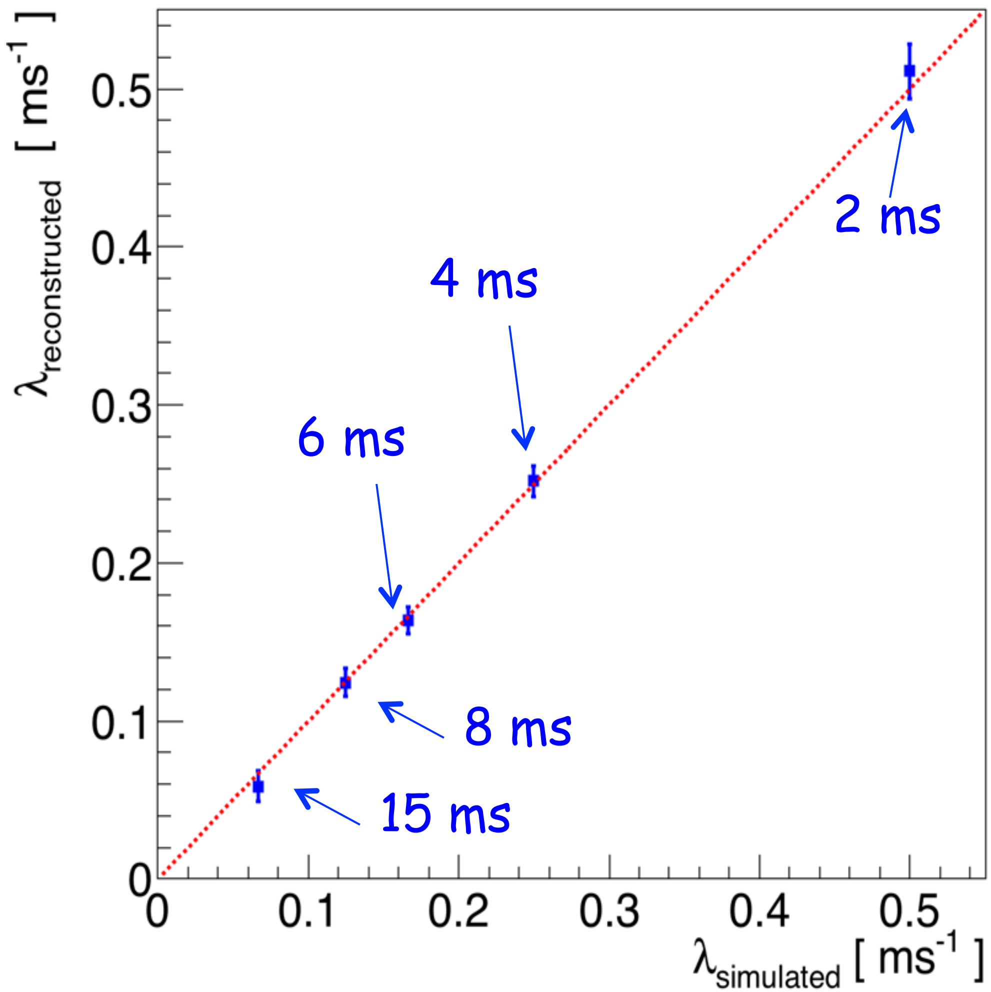
- PMT waveforms are generated including PMT noise, fluctuations in PMT gain (see G. Petrillo talk tomorrow for details)
- Mitigation of the memory/processing time (500 MHz sampling rate) by simulating behavior of the readout logic, with options to readout smaller portions of the waveform based on signals going above threshold or the presence of a beam signal
- Simulation of the signals generated by the PMT front end electronics and exploited for the trigger (see G. Petrillo talk)
- At the moment the full chain from digitized signals to global objects:
  - Hit PMT ("OpHit") identification;
  - Grouping of hit PMTs into clusters ("OpFlash") based on their time coincidence
- The full chain is committed in the ICARUS code. Needed their debug over large event statistics, together with fine tuning and and optimization of the clustering algorithm
  - Reconstruction needs of course validation and tuning with the real data during commissioning

Tools for the commissioning and first analysis

- In principle the basic tools for the commissioning are in place
- A first exercise to measure the LAr purity has been validated with simulated CR
- After initial
  - understanding of the actual of wire noise and how to treat it
  - verification of the PMT system and light signal reconstruction
- It should be possible
  - start looking at the tracking and the optical reconstruction
  - *Select clean cosmic tracks for the calibration and monitoring of the detector*
  - Move on to track/flash matching and
  - Start some basic analyses, e.g. stopping muons and Michel electrons
- …Select and measure neutrino interactions…

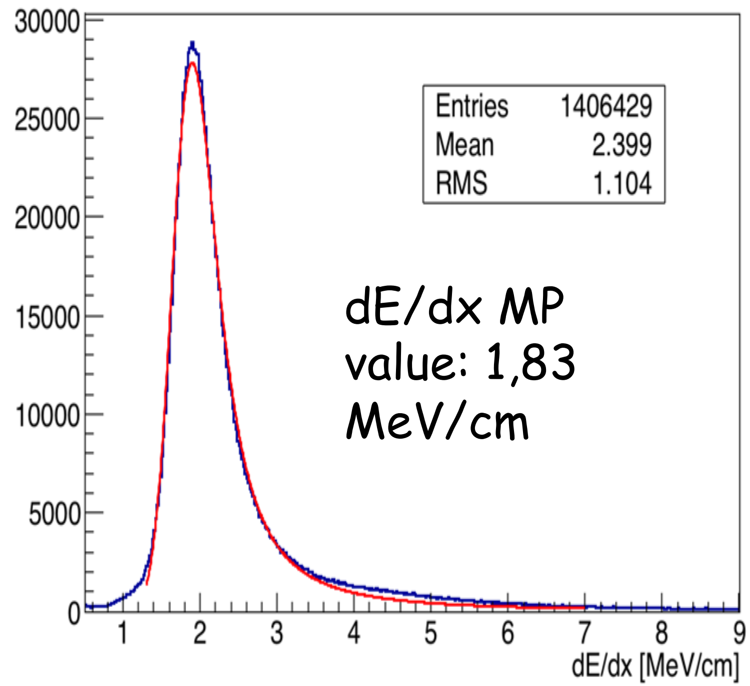
Purity measurement at FNAL: results from the offline method

- Method derived from the tool developed for the LNGS underground operation and applied to the events reconstructed by the standard reco tools (based on 200 events statistics):
- Small residual biases on $\lambda = 1/\tau$:
  - Slight underestimation of $\lambda = 1/\tau$ for $\tau > 8$ms
    ➔ maximum 0.8% underestimation of the deposited energy at high purity level;
  - overestimation of $\lambda = 1/\tau$ for $\tau < 4$ms,
    ➔ maximum 1.1% overestimation of the deposited energy for low purity level;



calibration of TPC wires

- During the commissioning an important task will be validate the reconstruction algorithms and calibrate/equalize the detector response with physical signals, exploiting the available tools
- A first possibility is represented by the study the dE/dx along cosmic μ tracks selected amongst the large flux of passing cosmic rays:
  - ✓ *The muons used for the purity measurement can be a starting point*
  - ✓ *First preliminary test performed on 5200 MC cosmic events. On average:*
    - ❖ *2 selected μ tracks per event*
    - ❖ *each Collection wire crossed by a selected ~ 250 tracks*
  - ✓ *Fitting the MP value of the histogram of dE/dx in each wire, it is possible to obtain the calibration constant of each wire with good accuracy*



The knowledge of the t_0 time of each used track is required!

Machine Learning approach to event reconstruction -1

- Plan for physics reconstruction based on ML technologies will be presented tomorrow by F. Drielsma

Computer Vision & Machine Learning for ICARUS Physics Reconstruction

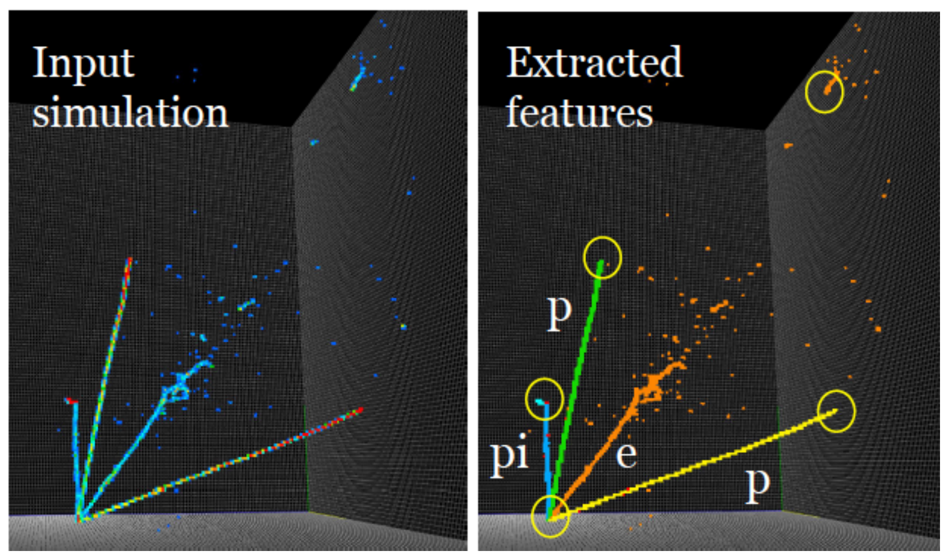
SLAC

Computer Vision: maximize physics output

- LArTPC = visually intuitive physics in hi-resolution image
- Image analysis is hard...
- Machine learning techniques recently made a breakthrough, became the go-to solution for algorithmic image analysis

Reconstruction

Machine learning is used to extract physics feature information efficiently, just like reconstruction (vertex finding, particle ID ...)



Machine Learning application to event reconstruction -2

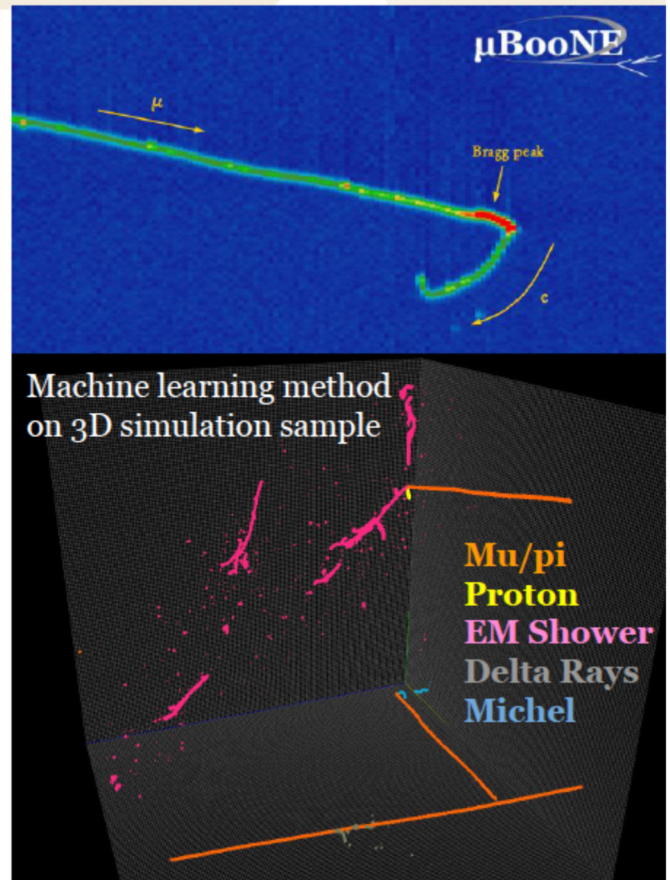
- Plan for application of ML technologies to Michel electron reconstruction will be presented tomorrow by L. Domine

Toward Detector Commissioning:
Machine Learning based Michel Reconstruction

SLAC

Michel Electron Reconstruction

- One commissioning analysis goal, also useful for **detector calibration**
- **Algorithmic method possible**: MicroBooNE published with efficiency ~2% with purity ~80 to 90%.
- Recent demonstration of >95% purity and efficiency on 3D simulation sample (for DUNE-ND pixel study)
- Apply this method for 2D projection images from ICARUS detector with extensions beyond publications
  - Mitigation of data vs. simulation domain discrepancy
  - Clustering of radiated gamma rays for better energy estimation



Notes on Simulation/Reconstruction Flow

- Simulation and Reconstruction running in "multi-TPC" mode
  - Detector simulation outputs "RawDigits" per TPC
  - Signal Processing runs separately on each RawDigit collection
  - SpacePoint formation merges output hit collection by Cryostat
  - Pandora pattern recognition runs separately on each Cryostat hit collection
  - Goal is to help enable future parallelization of the modules to decrease wall clock time for running jobs
  - This also does help reduce overall memory usage

- Reconstruction runs in two job steps
  - First step performs signal processing, gauss hit production, SpacePoint formation, pandora pattern recognition on gauss hits and PMT hit and flash reconstruction
  - Second step uses noise filtered waveforms from first step and runs the "raw hit" reconstruction, then pandora pattern recognition on the raw hits and then the CRT reconstruction

Production

- Production on the GRID are now smoothly run through the POMS system, with a nice web interface to control and monitor productions

- Centralized source of information at the wiki page https://cdcvs.fnal.gov/redmine/projects/icarus-production/wiki/ :
  - general information, documentation, tutorials
  - Instruction to place request to the production
  - Links to past/ongoing productions

Production – 2

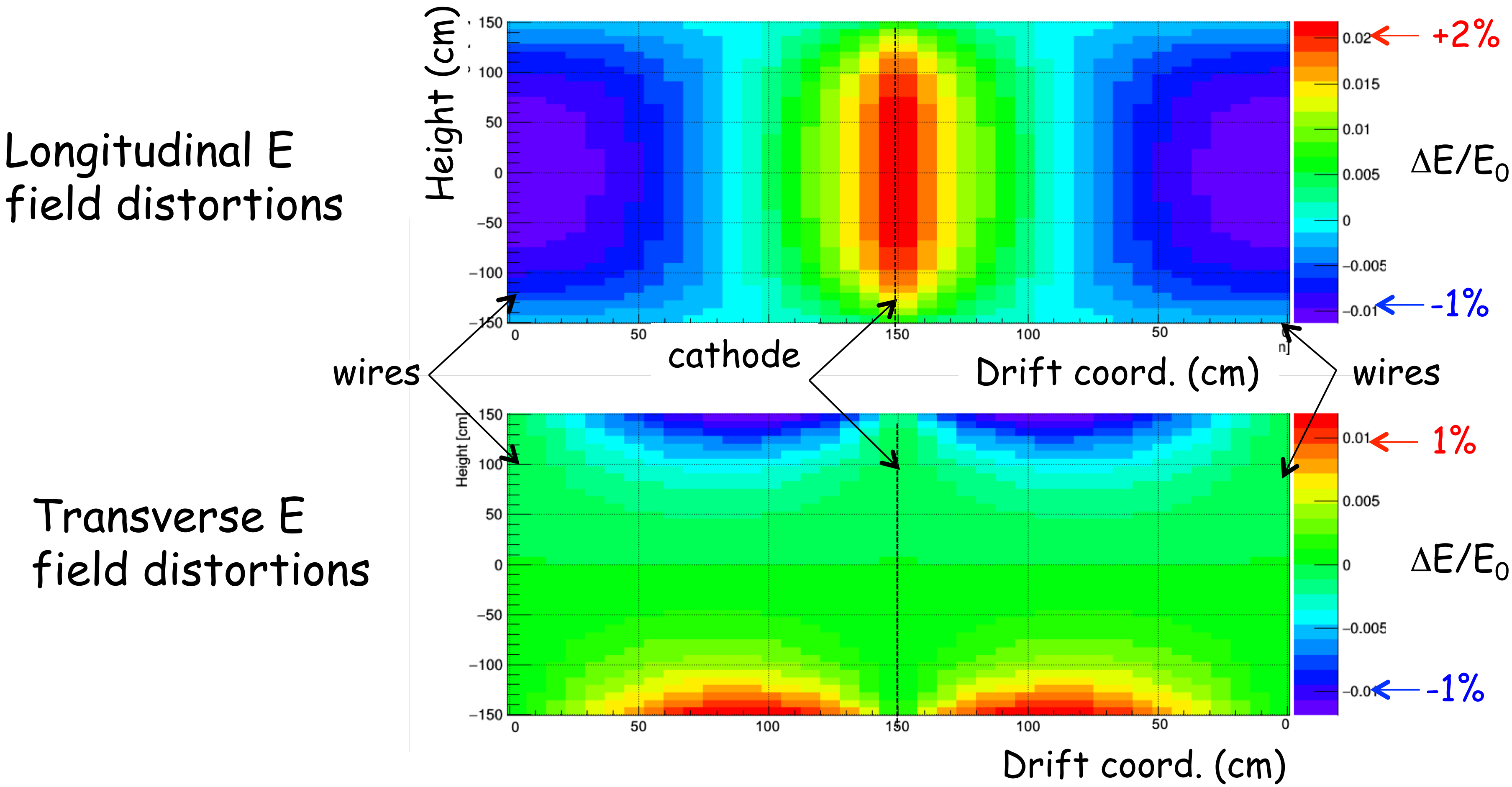
- ~400k events of different kind (including BNB and NuMI neutrino interactions, cosmic rays and single particles) have been simulated and reconstructed by the production system and are available for the analysis.
- Significant improvements in memory size and CPU requests have been obtained thanks to detailed optimization of the code.
- Presently the job success rate is ~95%. Failure for large memory/large CPU events exceeding the standard allocated memory/CPU.
- Some resources have been allocated to process and store the amount of data expected during the commissioning phase.

- More details in tomorrow presentation by M. Wospakrik.

To be done for commissioning…

- Lots of activities connected with the commissioning will start as the first real data will become available
- Accurate calibration with cosmics
- Validate reconstruction with real data:
  - Handle the real noise conditions
  - study hit reconstruction efficiency in real conditions
  - track/shower reconstruction
  - Implement space charge effects description and check with data Complete the
  - Validate light flash reconstruction and its association with TPC events
  - Complete the CRT reconstruction and track CRT association
- The present group can provide and maintain the basic functionalities, an enlarged participation would be necessary to launch first analyses including first selections and reconstruction of ν candidates: **contributions by new people is really welcome!**

Further improvement: inclusion of space charge effects



- few mm longitudinal distortions observed in Pavia surface run
- First calculation of Electric field distortion with Comsol® code
- Needs to be implemented into the simulation/reconstruction code...

Higher level analysis to study neutrinos and reject cosmics

- A very effective suppression of cosmic events can be operated by the CRT detector with small reduction of genuine ν interaction
- We should exploit to the best the light and TPC information to compensate the initial incompleteness of the CRT coverage
  - Accurate reconstructing of the PMT signals, determining the timing at ns level and their pulse height.
  - The pattern of hit PMTs and of the detailed (ns scale) time evolution of the light flash on PMTs would provide powerful discrimination between crossing μ and ν interactions, as indicated from first neural networks studies (and from ML)
  - Identify the interaction corresponding to the triggered PMTs
  - Exploit the differences between passing cosmics and ν interaction in the reconstructed event using the TPC information.
  - Optimize the shower identification and electron tagging from initial dE/dx