Status: hit finding and raw waveform fitting in LArSoft

Christoph Alt

June 7^{th}, 2017 / WA105 Science Board

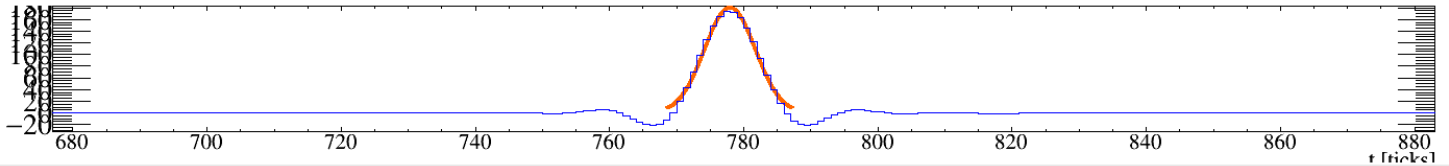
ETH zürich WA105 DUNE DEEP UNDERGROUND NEUTRINO EXPERIMENT

Motivation and overview

Current status: Apply deconvolution to raw waveforms before hit finding in all planes

→ highly motivated by bipolar induction plane signals

→ introduces distortions to the waveform



Goal: use raw waveform for hit finding and fitting in dual phase/collection planes

Content:

1. Hit shaping, finding and fitting
2. Event display in LArSoft
3. Performance of raw hit finding

1. Hit shaping, finding and fitting

Hit shaping

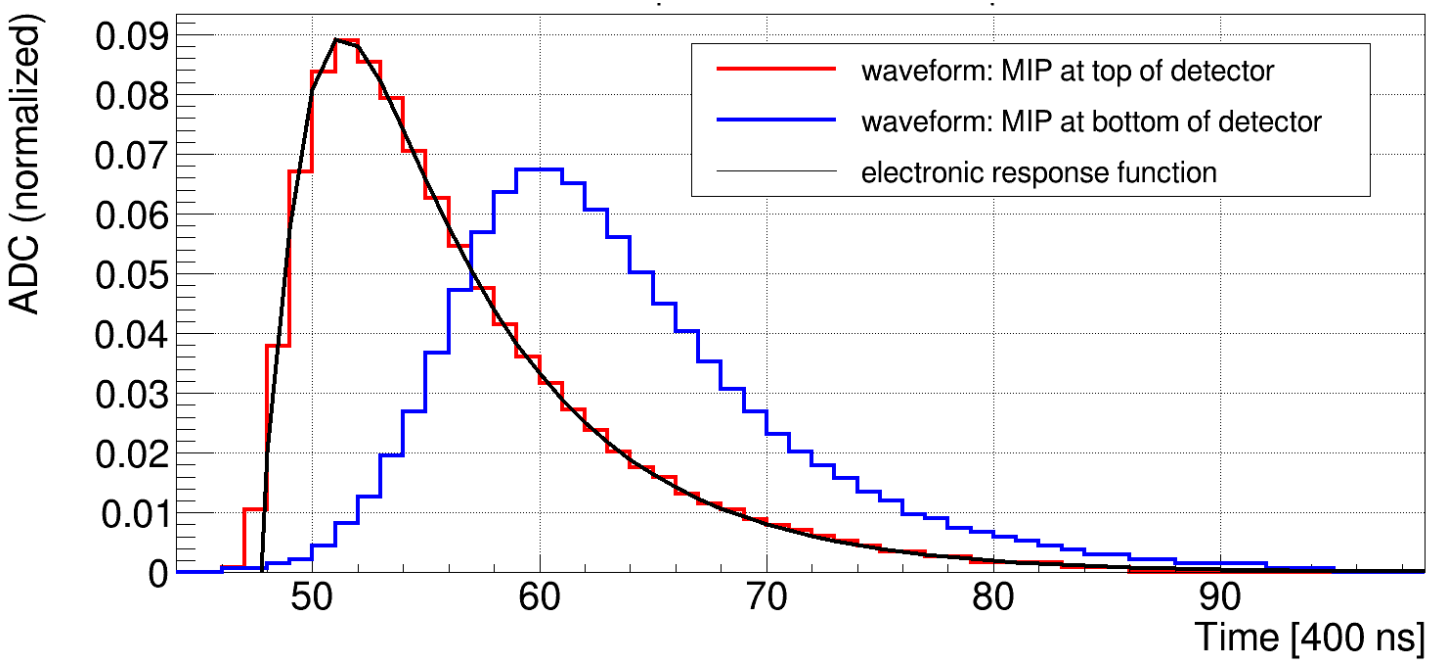
- Electronic response function for dual phase:

$$f_{FastPreAmp}(t) = \frac{\tau_D \cdot e^{\frac{-(t-t_0)}{\tau_D}} - \left(\tau_D + (t-t_0)\frac{\tau_D - \tau_I}{\tau_I}\right) \cdot e^{\frac{-(t-t_0)}{\tau_I}}}{(\tau_D - \tau_I)^2}$$

with: $\tau_D = 2.83\mu s, \quad \tau_I = 0.47\mu s$

- Waveform = electron distribution (t) $\otimes$ electronic response (t)

$\rightarrow$ Diffusion (6m drift) from red to blue in protoDUNE-DP



Added after the meeting: diffusion in LArSoft

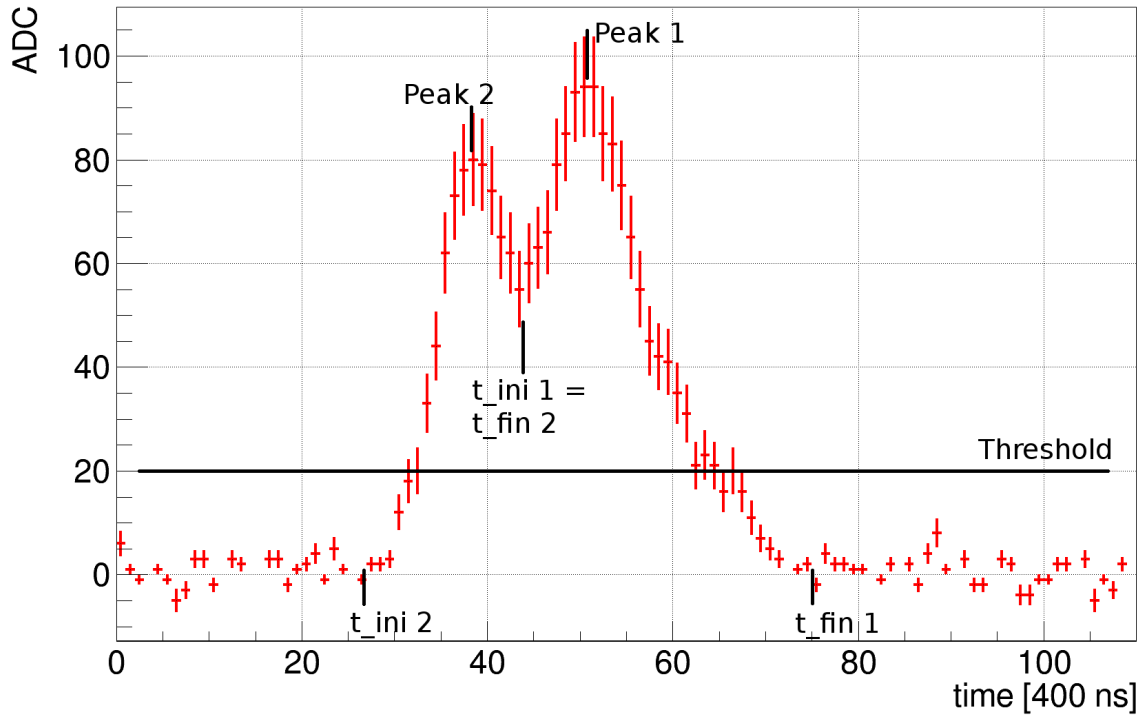
- Position of ionization charge is smeared along drift direction (longitudinal "L") and transversal to drift direction ("T") using a Gaussian smearing function with the following σ:

$$\sigma_{L/T} = \sqrt{2 \cdot D_{L/T} \cdot x_{Drift}/v_x}$$

- $D_L = 6.2 \cdot 10^{-9} cm^2/ns$
- $D_T = 16.3 \cdot 10^{-9} cm^2/ns$
- x_{Drift} = drift distance
- x_{Drift} = drift velocity (depends on electric field)

Hit finding

1. Define peaks: bin with highest ADC count above threshold
2. Define t_{ini} and t_{fin}: walk along time axis until:
   - 2.1 a bin with ADC < 0 is found, **or:**
   - 2.1 a bin followed by 3 consecutive bins with same/higher ADC count is found (criteria to be tuned, depends on gain, noise, ...)
3. Repeat step 1 and 2 for remaining region until no peak above threshold is found

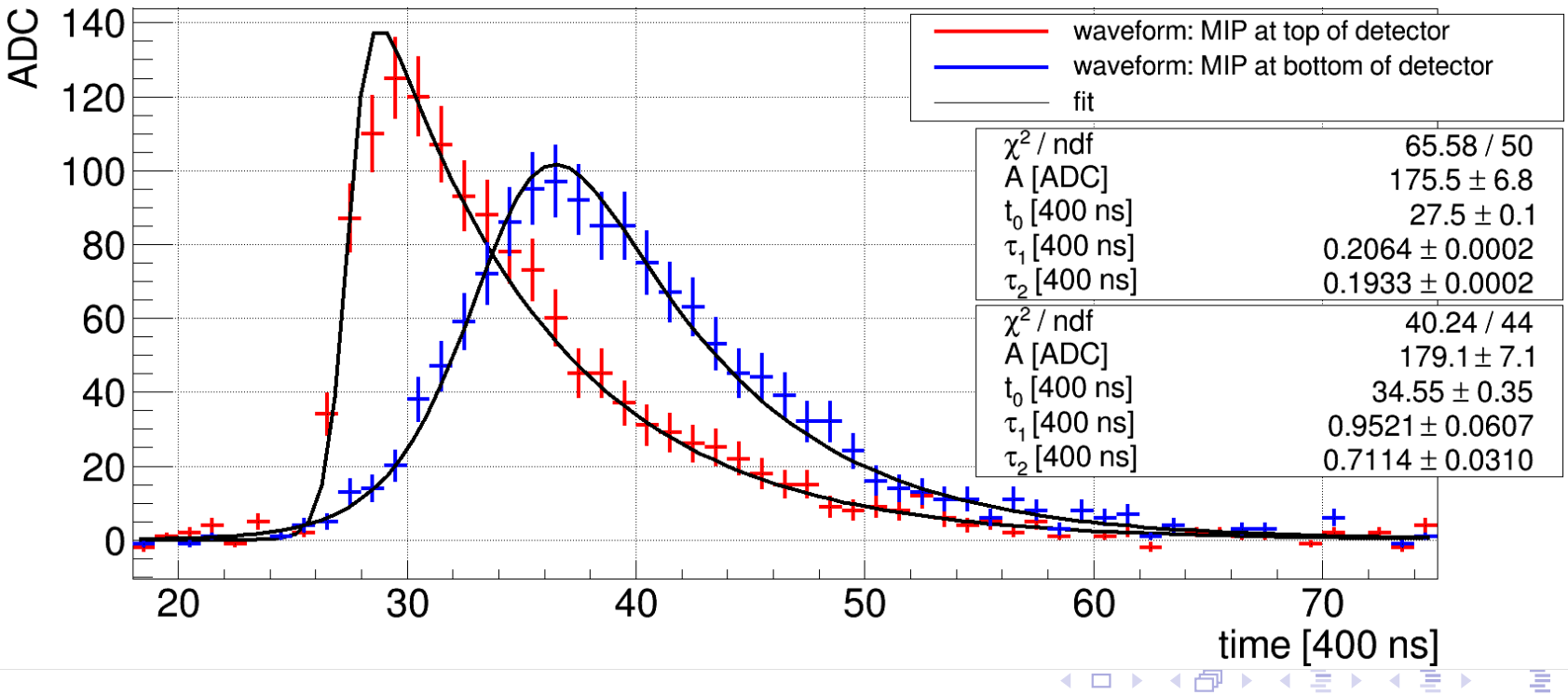


Hit fitting: single hits

- Fit every identified peak from t_{ini} to t_{fin} :

$$\text{fit function } f(t) = A \cdot \frac{e^{\frac{t-t_0}{\tau_1}}}{1+e^{\frac{t-t_0}{\tau_2}}}$$

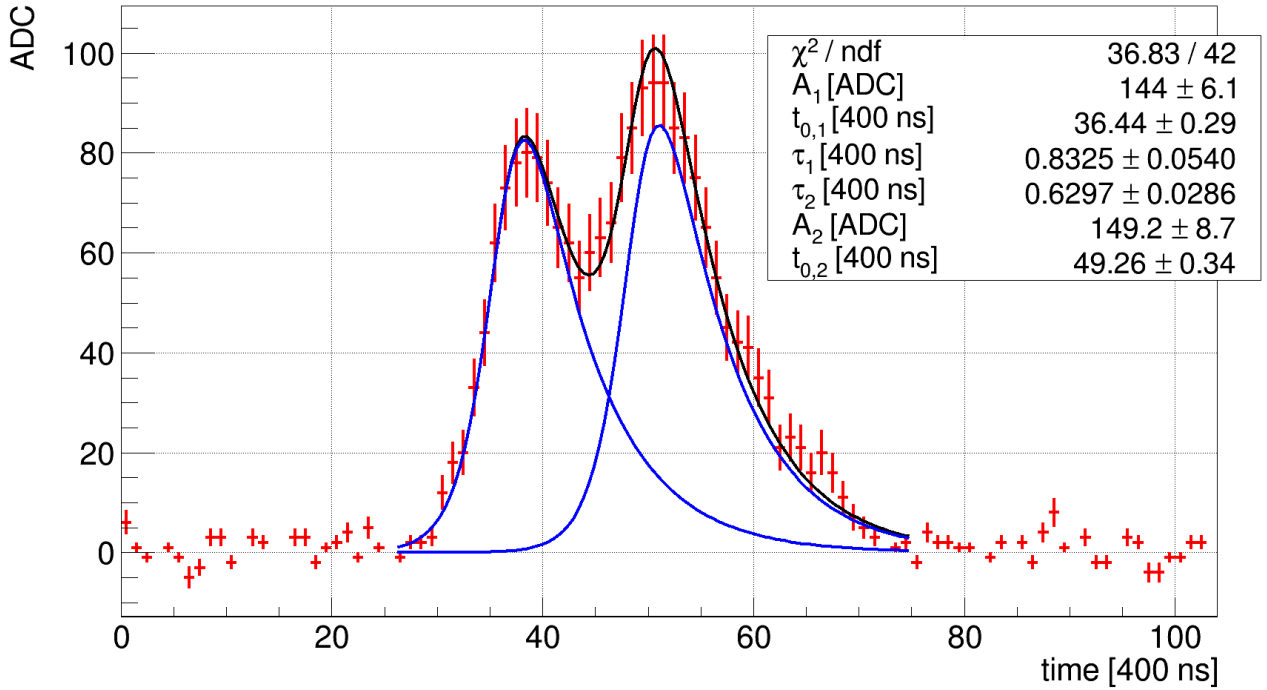
- Seeds: t_0 = peak position-3, $\tau_1 = \tau_2 = 0.5$, A = 1.65*ADC(peak)



Hit fitting: multi hits

- If $t_{fin,i} = t_{ini,i+1}$, perform single fit over whole range (black)
- Each summand of the fit function corresponds to a single hit (blue)

$$\text{fit function } f(t) = \sum_i A_i \cdot \frac{e^{\frac{t-t_{0,i}}{\tau_1}}}{1+e^{\frac{t-t_{0,i}}{\tau_2}}}$$



Hit fitting: calorimetry

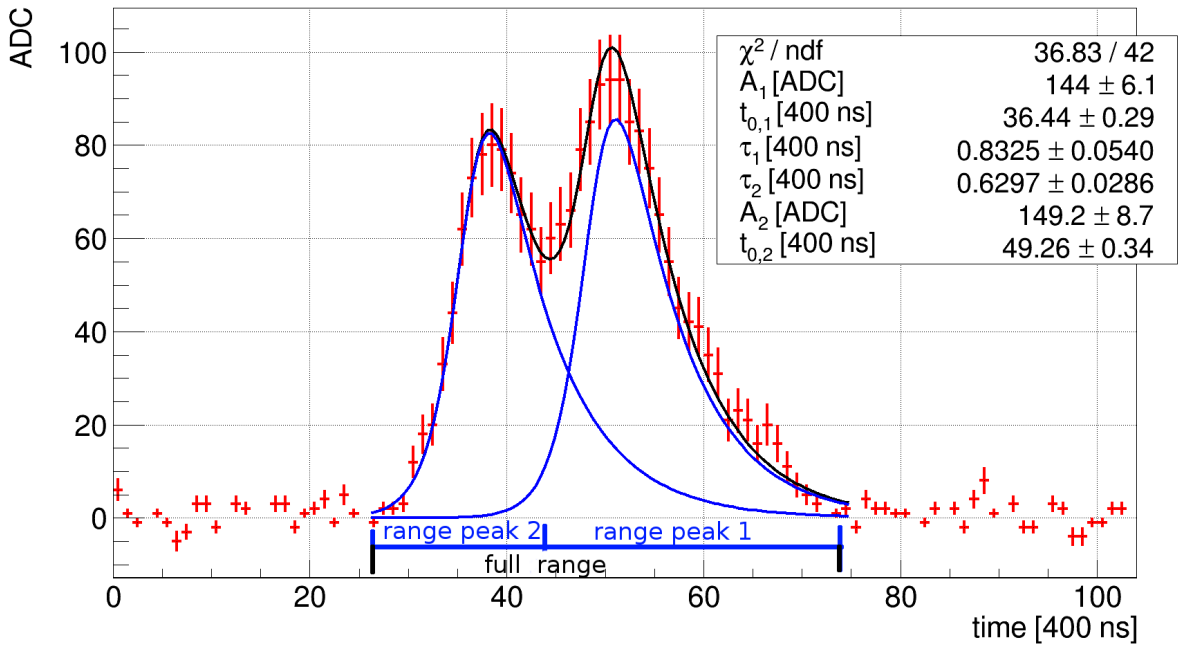
Single hits:

1. Integral of the fit and sum of ADC counts over fit range deliver same result

Multi hits:

1. Integral of each summand of the fit function over full range and sum of ADC counts within peak range deliver different results

→ go for the integral in both cases



Hit fitting: missed hits (to be improved)

Problem:

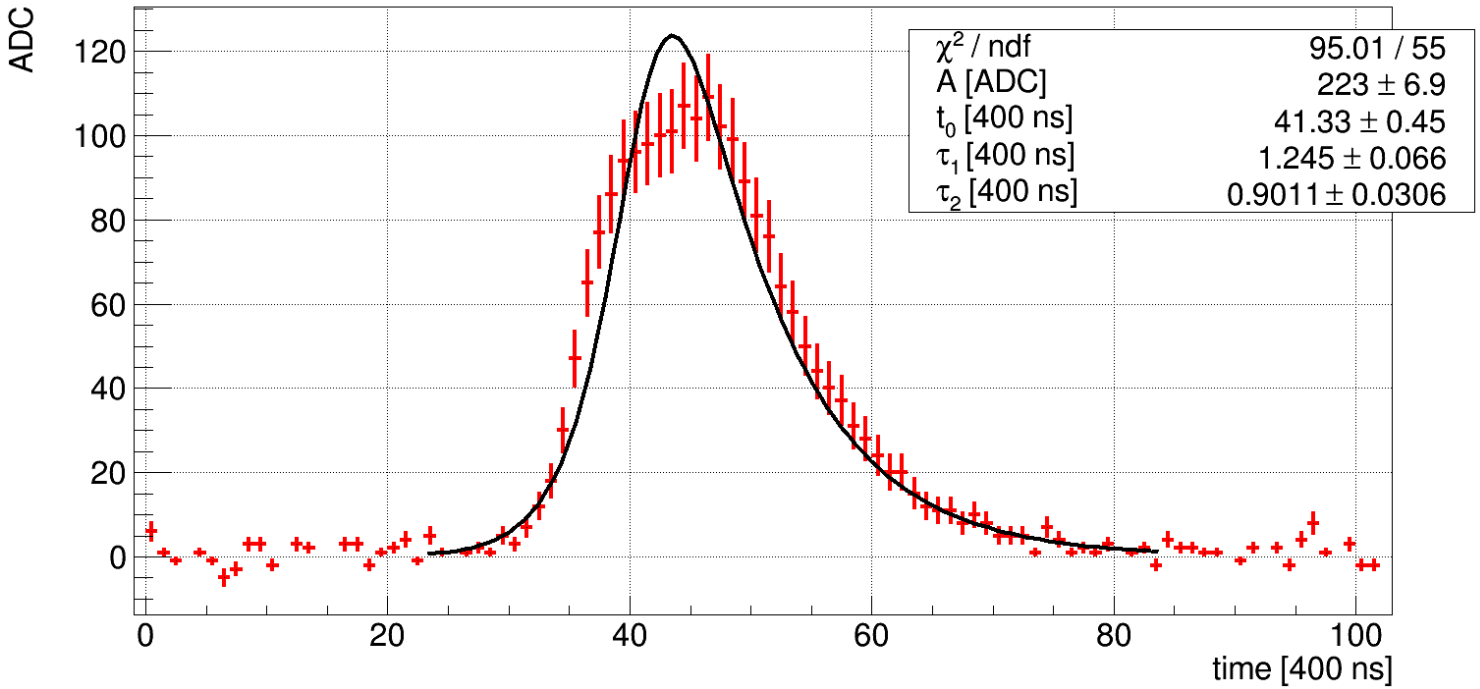
- n+1 hits misidentified as n hits (see example in plot (n=1))

→ results in a bad fit

Possible solution:

- If $\chi^2 > \chi^2_{Retry}$ for initial fit, try to fit n+1 hits

→ fitter needs good seeds/limits for additional hit (work ongoing)

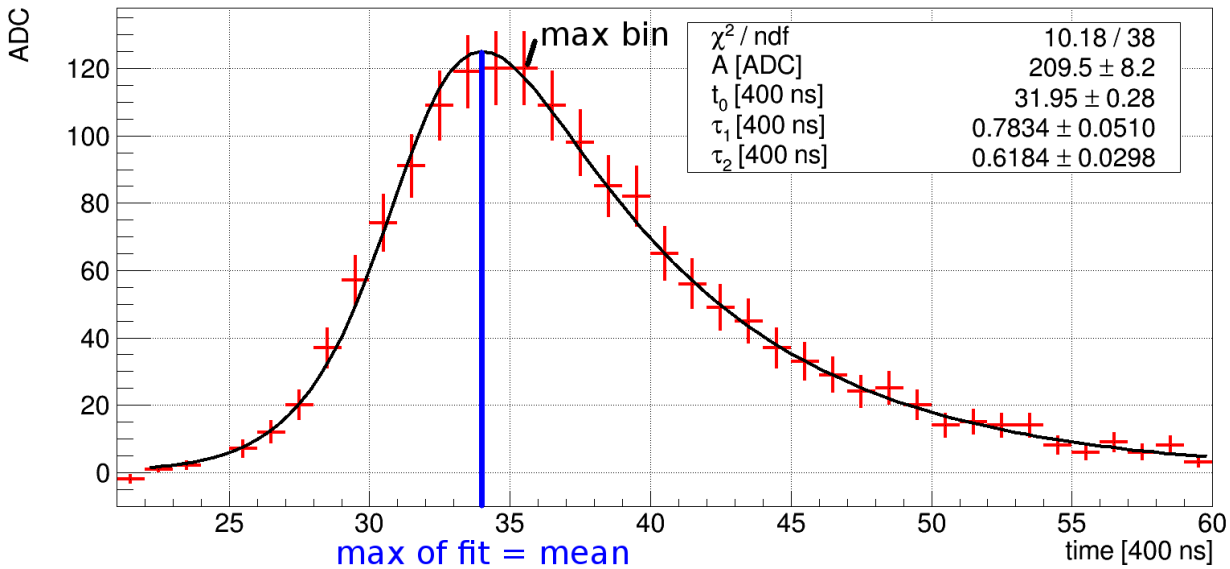


Hit fitting: saving fit parameters in recob::Hit

- Save fit parameters for each hit in a recob::Hit object
- In LArSoft, recob::Hit expects Gaussian fits of deconvoluted waveforms and is used as input for clustering and tracking

→ from Gaussian fit: width, mean and amplitude

- Need to define width and mean from double-exponential fit:
- width: $\tau_1 + \tau_2$ (too small, needs to be tuned!)
- amplitude: A, mean: see plot



2. Event display in LArSoft

Event display: raw fits (many thanks to Robert)

Goal:

- Plot raw fits in event display

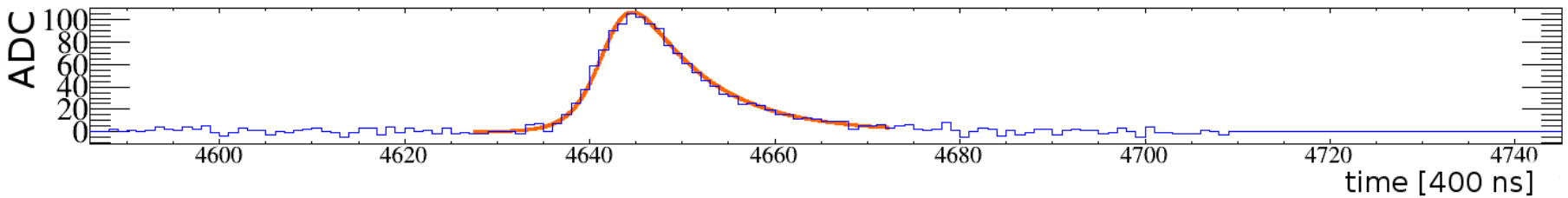
Problem:

- recob::Hit doesn't provide space to save the raw fit parameters (raw fit: 4 parameters, Gaussian fit: 3 parameters)

Solution:

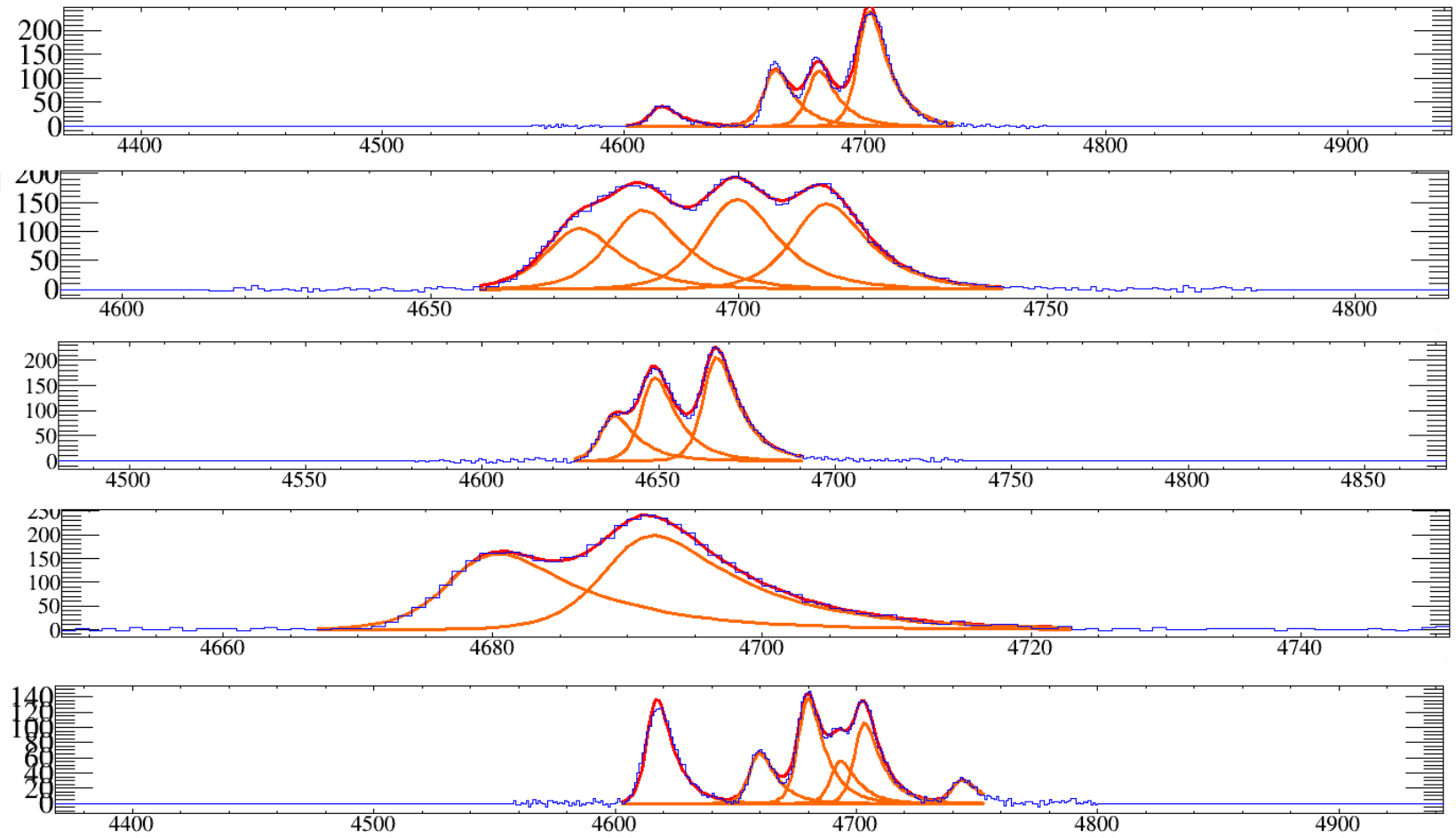
- Save raw fit parameters in a separate vector and assign this vector to the recob::Hit object

→ use MVAWriter (lardata/ArtDataHelper/MVAWriter.*)



Event display: raw fits

- For multi hits, plot both the fit function (red) as well as each summand of the fit function, describing single hits (orange):

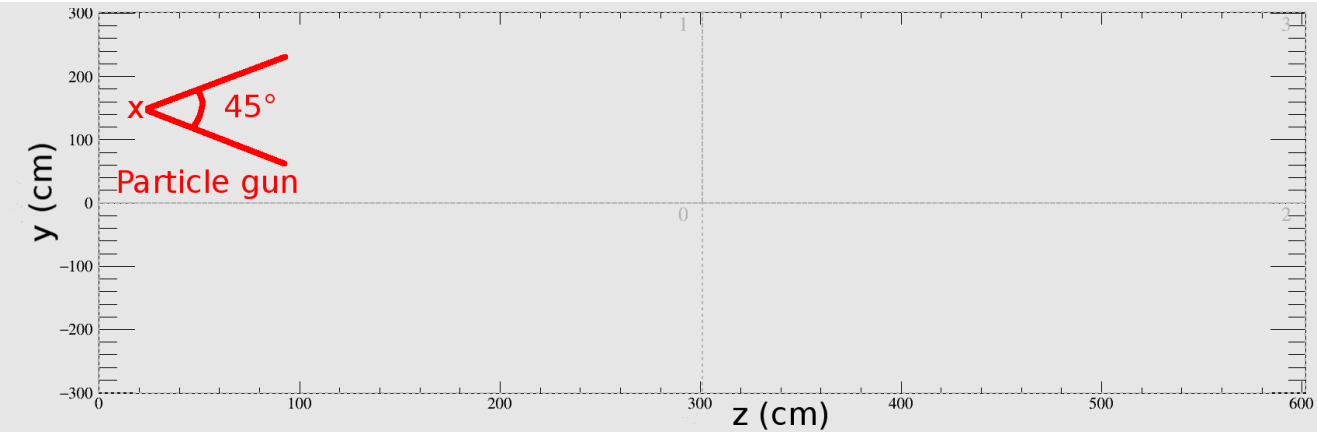


3. Performance of the new algorithm

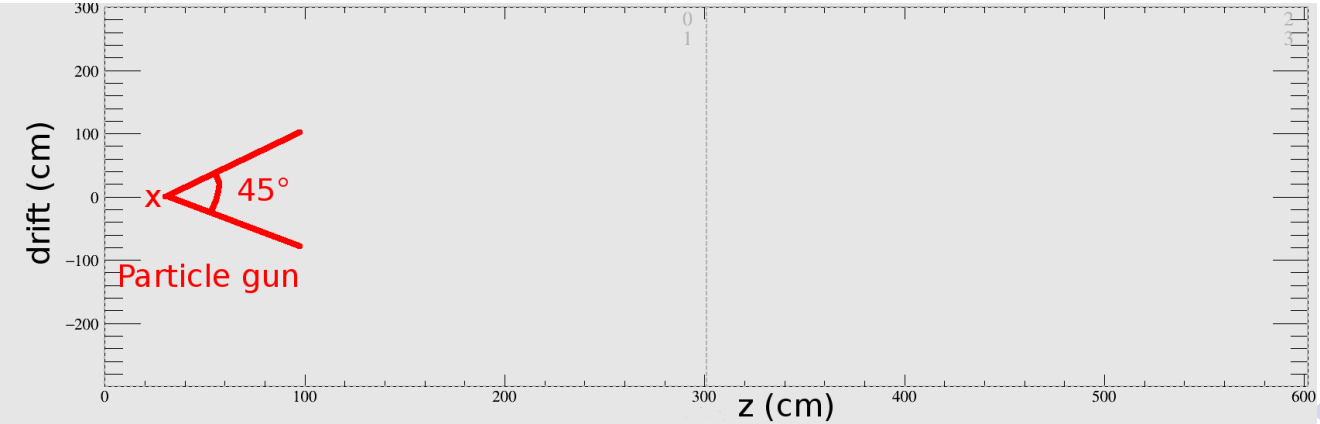
Performance of raw hit finding: data set

- 100 single μ^- with $P_{\mu^-} = 500$ MeV in protoDUNE dp geo
- isotropic distribution within 45° in both planes

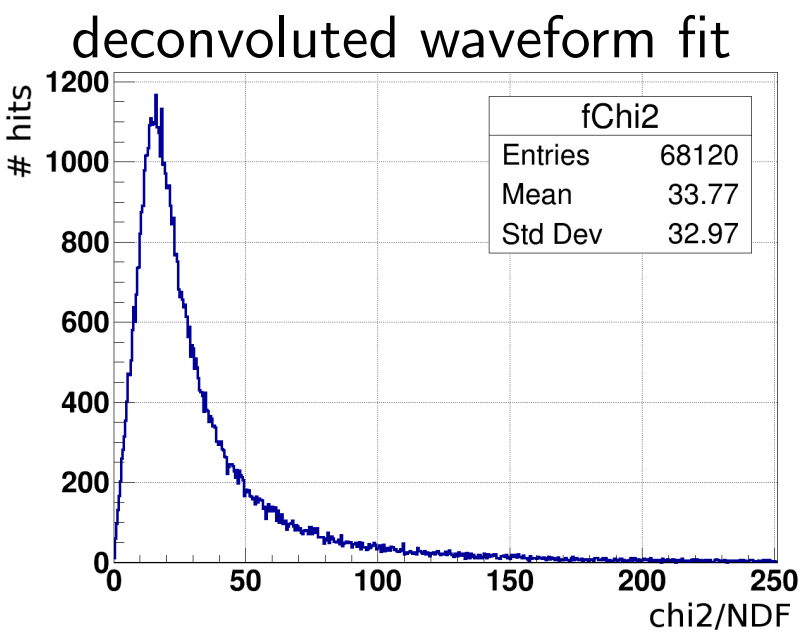
Top view:

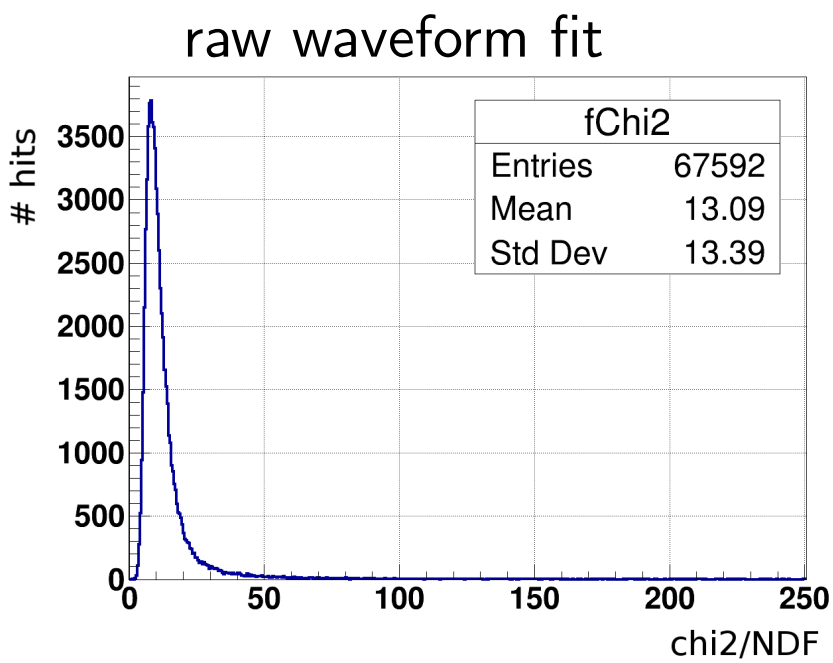


Side view:



Performance of raw hit finding: fitting





- Average χ^2/NDF: 34 (deconvoluted) vs. 13 (raw)
- Number of hits: $\frac{\text{raw}}{\text{deconvoluted}} = 99.22\,\%$

$\rightarrow$ will further improve with the 'n+1 fit' for missed hits

Performance of raw hit finding: tracking efficiency

Efficiency definition for track reconstruction:

1. Completeness $\geqslant 50\,\%$
2. Purity $\geqslant 50\,\%$
3. $75\% \leqslant \frac{L_{reco}}{L_{truth}} \leqslant 125\%$

more details: `https://indico.fnal.gov/getFile.py/access?contribId=57&sessionId=16&resId=0&materialId=slides&confId=10641`

Results:

- Deconvoluted: 97 %
- Raw: 94 %

$\rightarrow$ 6 events failed $\frac{L_{reco}}{L_{truth}}$ criteria because of split tracks

- Clustering can move single hits within their widths when assigning them to a cluster. Clustering tends to split tracks more often for small widths. This is what happened here.

Summary and outlook

Summary

- Raw hit finding and fitting for dual phase is working stable
- Delivers good results for hit finding and tracking efficiency already

Outlook

- Tune input for clustering and tracking (width)
- Improve hit finding and fitting for missed hits
- Check calorimetry (ADC counts to charge conversion)