Pandora for dual phase LArTPCs

Dom Brailsford, Maria Brigida Brunetti, Etienne Chardonnet

FD sim/reco workshop
17/09/20

Pandora for dual phase LArTPCs

- Pandora is a multi-algorithm approach to pattern recognition, deployed in many LArTPC experiments
    - Harnesses physics and detector information to maximise reconstruction potential
- Most of Pandora's algorithms are suitably detector agnostic and work 'out of the box' for dual-phase
- A few of Pandora's most powerful algorithms require three readout views (single phase LArTPCs) to operate
    - Unusable by a two readout view dual phase LArTPC
- The following slides overview a tailored end-to-end reconstruction workflow for dual phase LArTPCs in Pandora, including expansion and adaptation of key algorithms
    - The new Pandora workflow is a part of the ProtoDUNE-DP offline production for cosmic rays

2D reconstruction: using hit widths

- 2D clustering algorithm which assesses hit widths

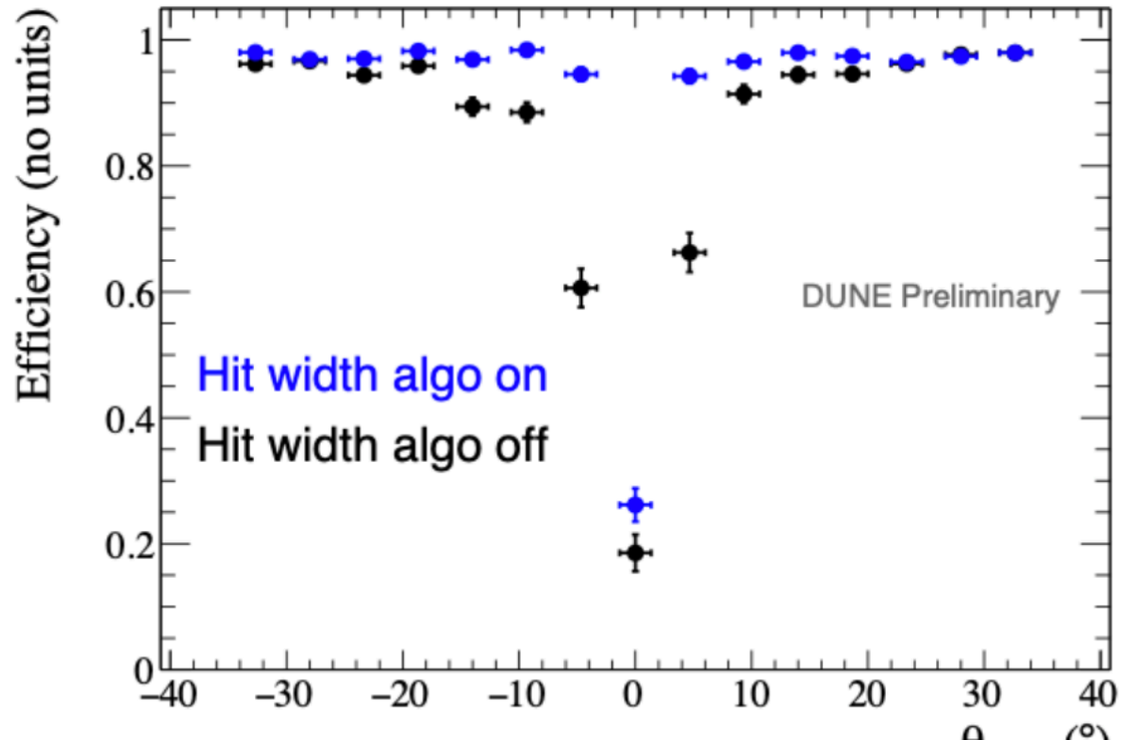
- Developed by **Isobel Mawby** for DUNE-FD

- Improves reconstruction performance for sparse tracks

  - Such tracks are typically almost parallel to the drift axis

Without hit width clustering alg

With hit width clustering alg



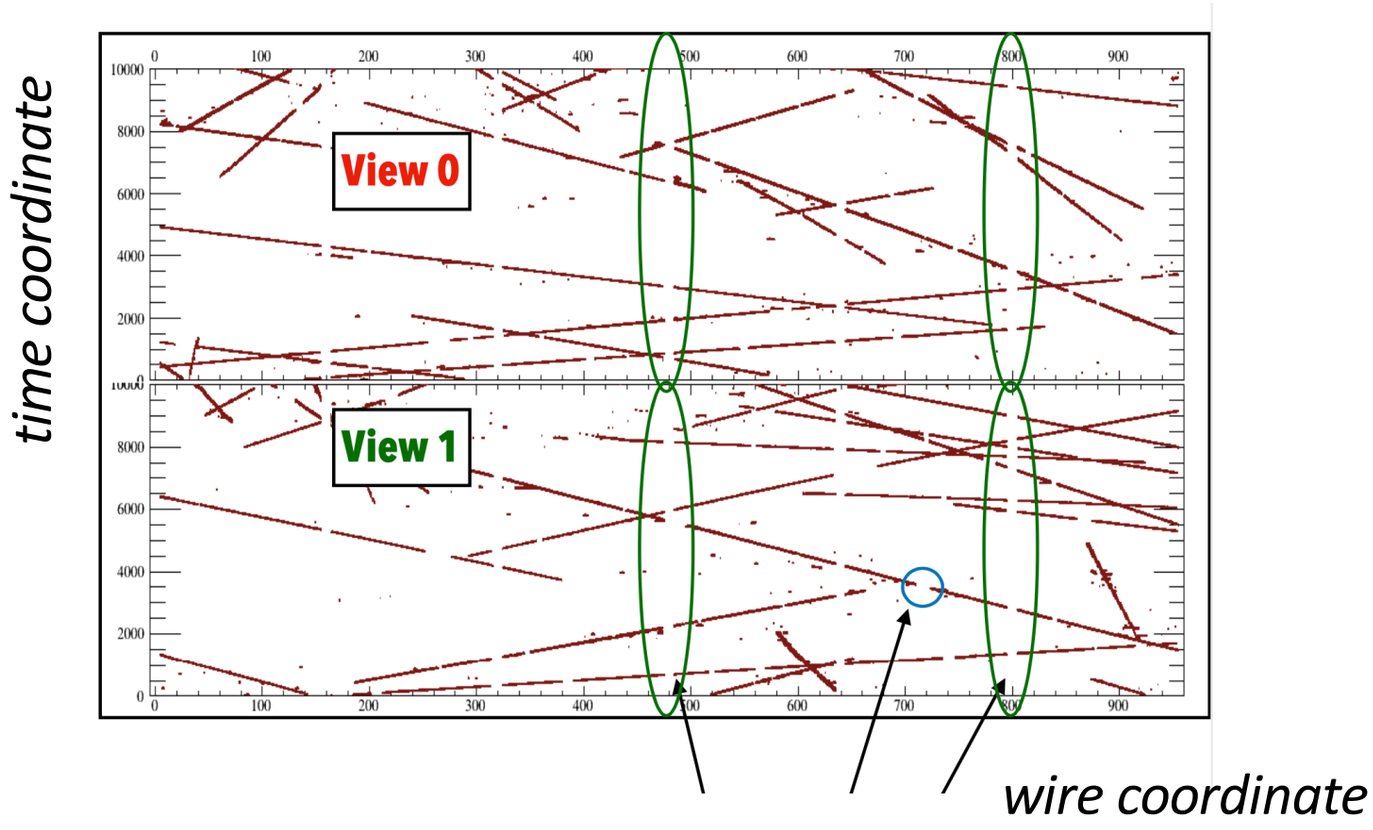
Particle gun muons in ProtoDUNE-DP

2D reconstruction: LEM and CRP gaps

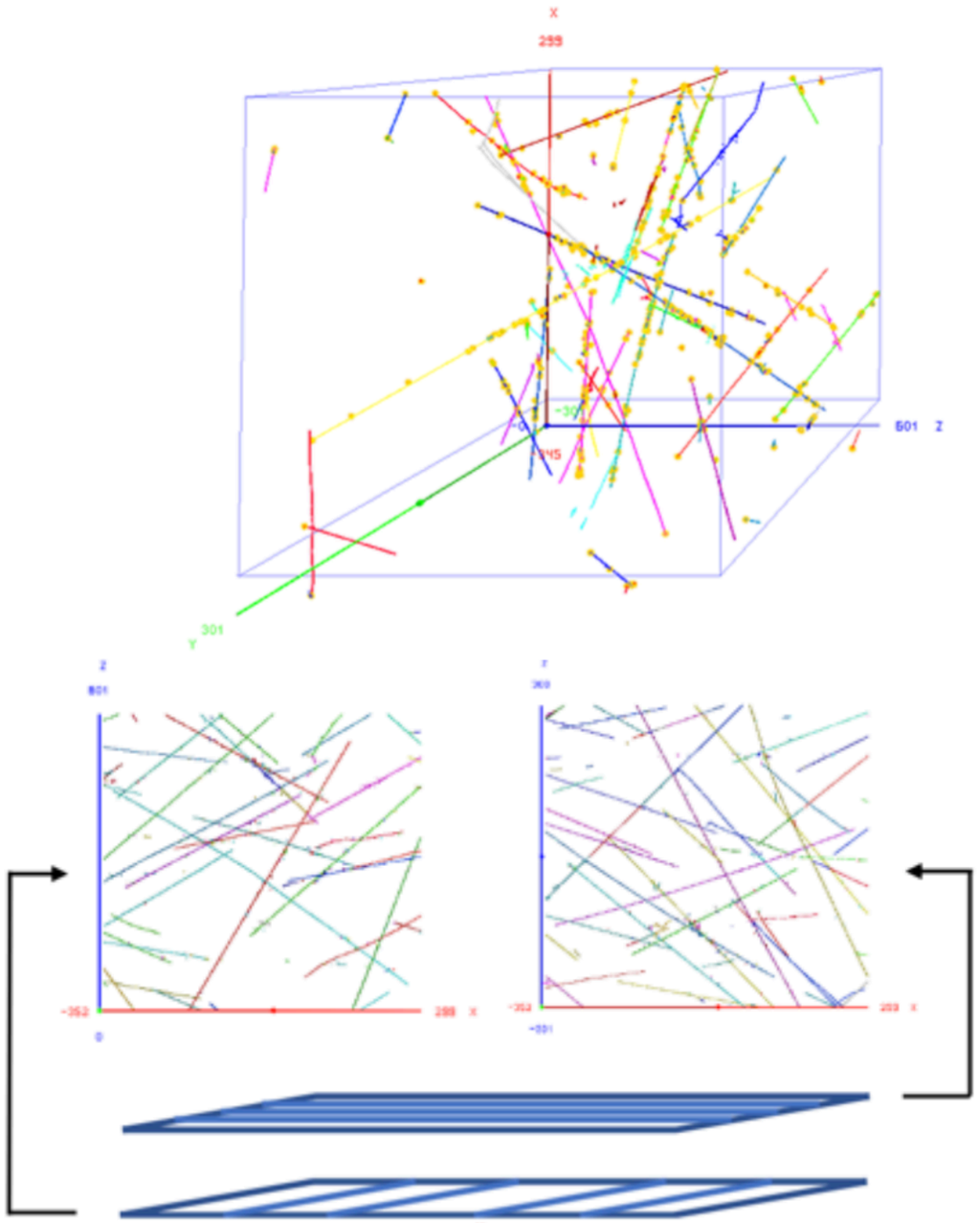
- ‘Blind’ regions of the detector between adjacent LEMS and adjacent CRPs

- Gap positions now implemented in pandora

- Knowledge of the gaps provides guidance to Pandora’s 2D clustering algorithms



2D->3D matching

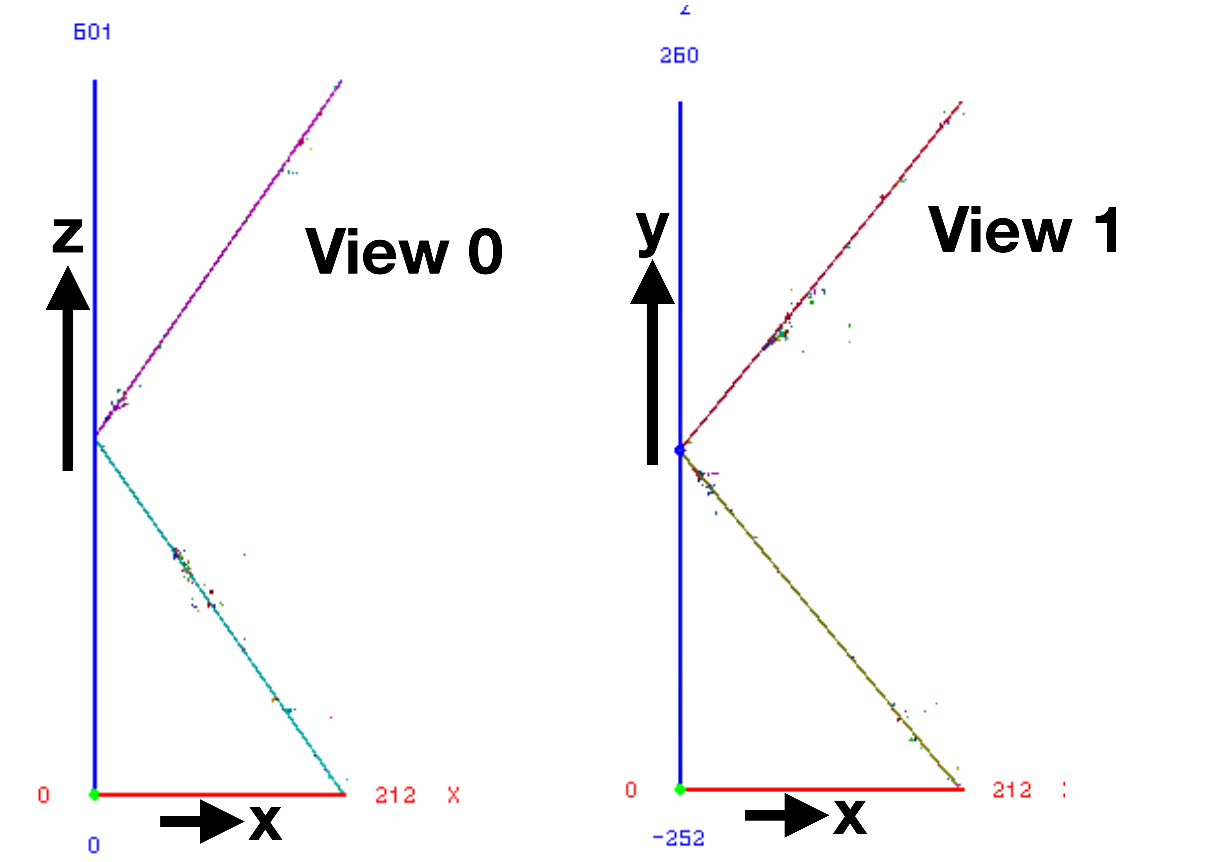


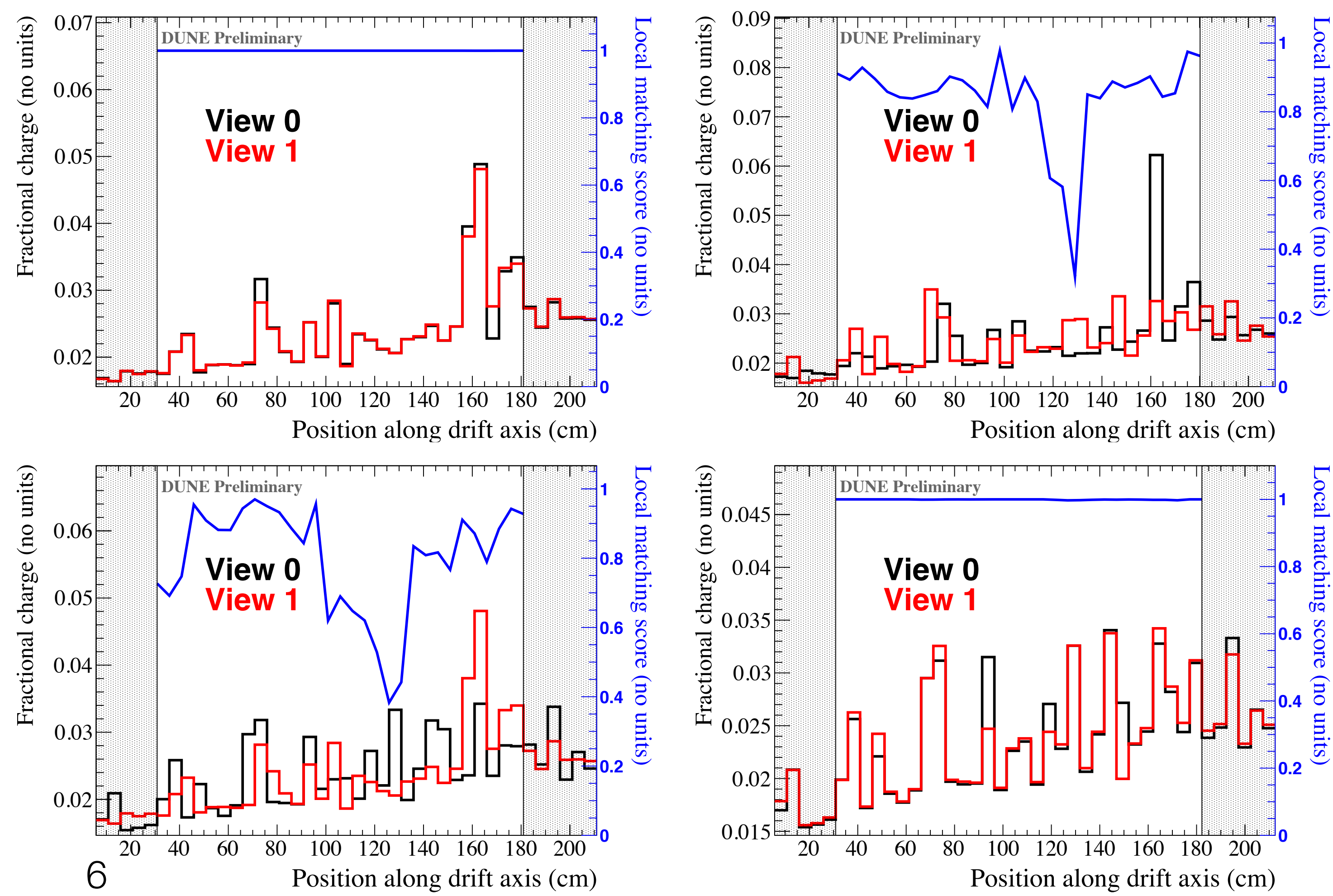
- Matches compatible 2D clusters across readout views to create 3D clusters
- Pandora for single-phase/3-view LArTPCs utilises all three views simultaneously to maximise 2D->3D matching potential
- Pandora for dual-phase/2-view LArTPCs now utilises **calorimetric information** for the first time to match clusters across views

2D->3D matching

- Compare all pairwise cluster combinations across views
- Find the time-overlap region for each cluster pair
- Create fractional charge profiles for each cluster in said overlap region (red and black histograms)
- For each cluster comparison, calculate:
  - Global matching score*
  - Local matching scores* for regions of the overlap (blue curves)
- A high score (global or local) indicates a good match

Di-muon particle gun in ProtoDUNE-DP





***Score based on correlation coefficient p-value**

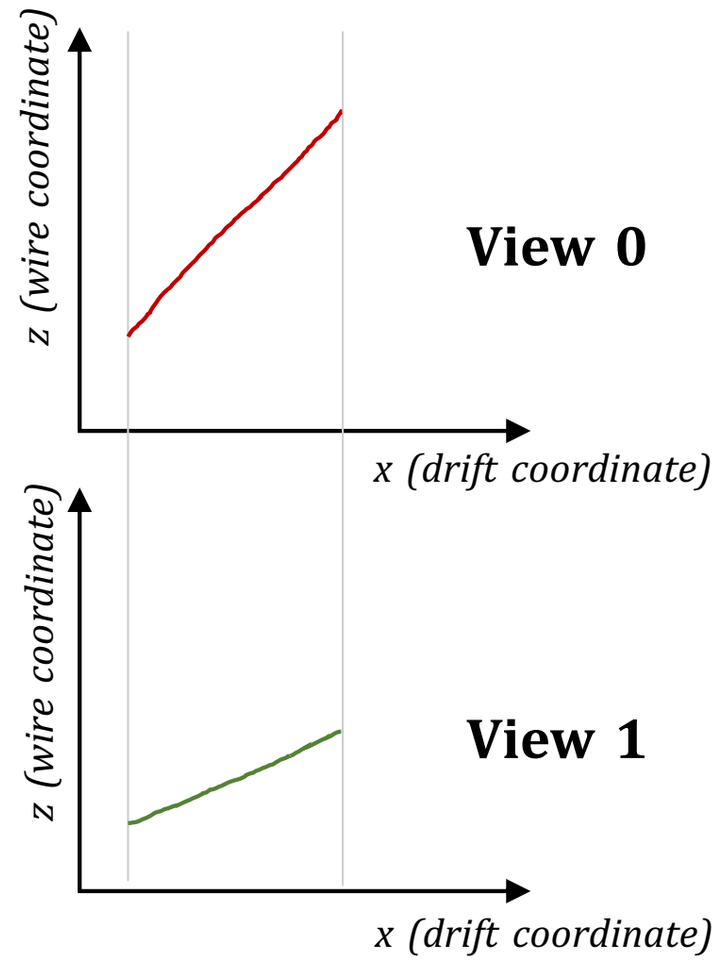
Most recent 2D->3D matching developments

- Packaged the calorimetric matching metrics into **tools** which decide if two 2D clusters match together to make a 3D particle

- Most tools target a particular topology -> harnesses geometry and calorimetry together

- Implemented tools
  - Clear tracks tool
  - Long tracks tool
  - Simple tracks tool

Clear tracks tool



- Two clusters only overlap each other
- The calorimetry matching metrics between the two clusters are suitably high

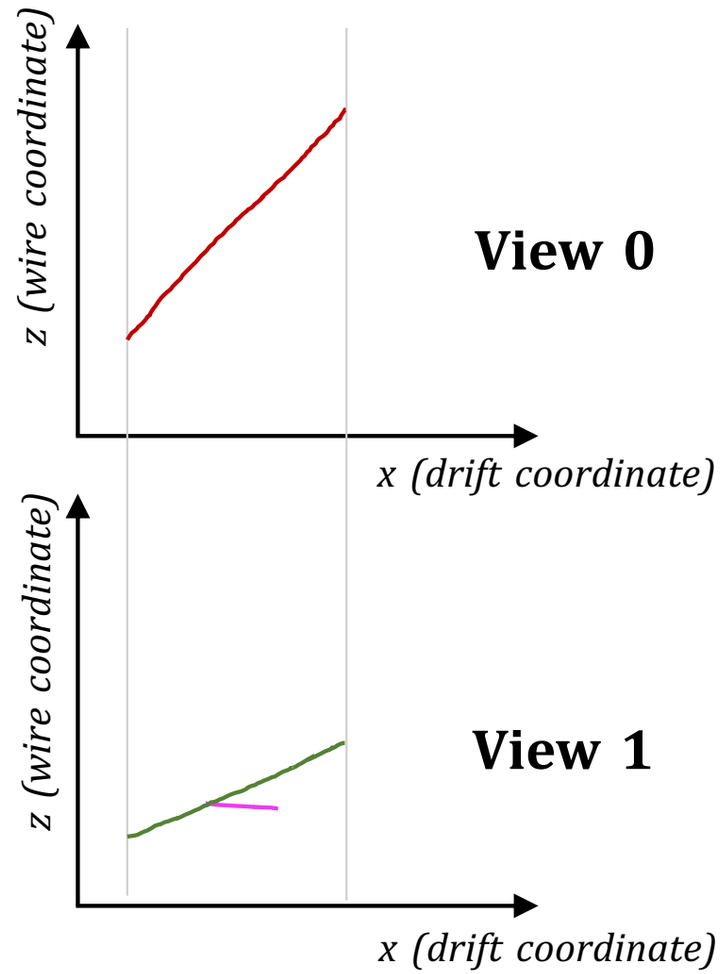
Most recent 2D->3D matching developments

- Packaged the calorimetric matching metrics into **tools** which decide if two 2D clusters match together to make a 3D particle

- Most tools target a particular topology -> harnesses geometry and calorimetry together

- Implemented tools
  - Clear tracks tool
  - Long tracks tool
  - Simple tracks tool

Long tracks tool



- Cluster(s) match multiple other clusters in the opposing view
- Ambiguity broken by picking comparisons with sufficiently long regions that calorimetrically match

Most recent 2D->3D matching developments

- Packaged the calorimetric matching metrics into **tools** which decide if two 2D clusters match together to make a 3D particle

- Most tools target a particular topology -> harnesses geometry and calorimetry together

- Implemented tools
  - Clear tracks tool
  - Long tracks tool
  - Simple tracks tool

Simple tracks tool

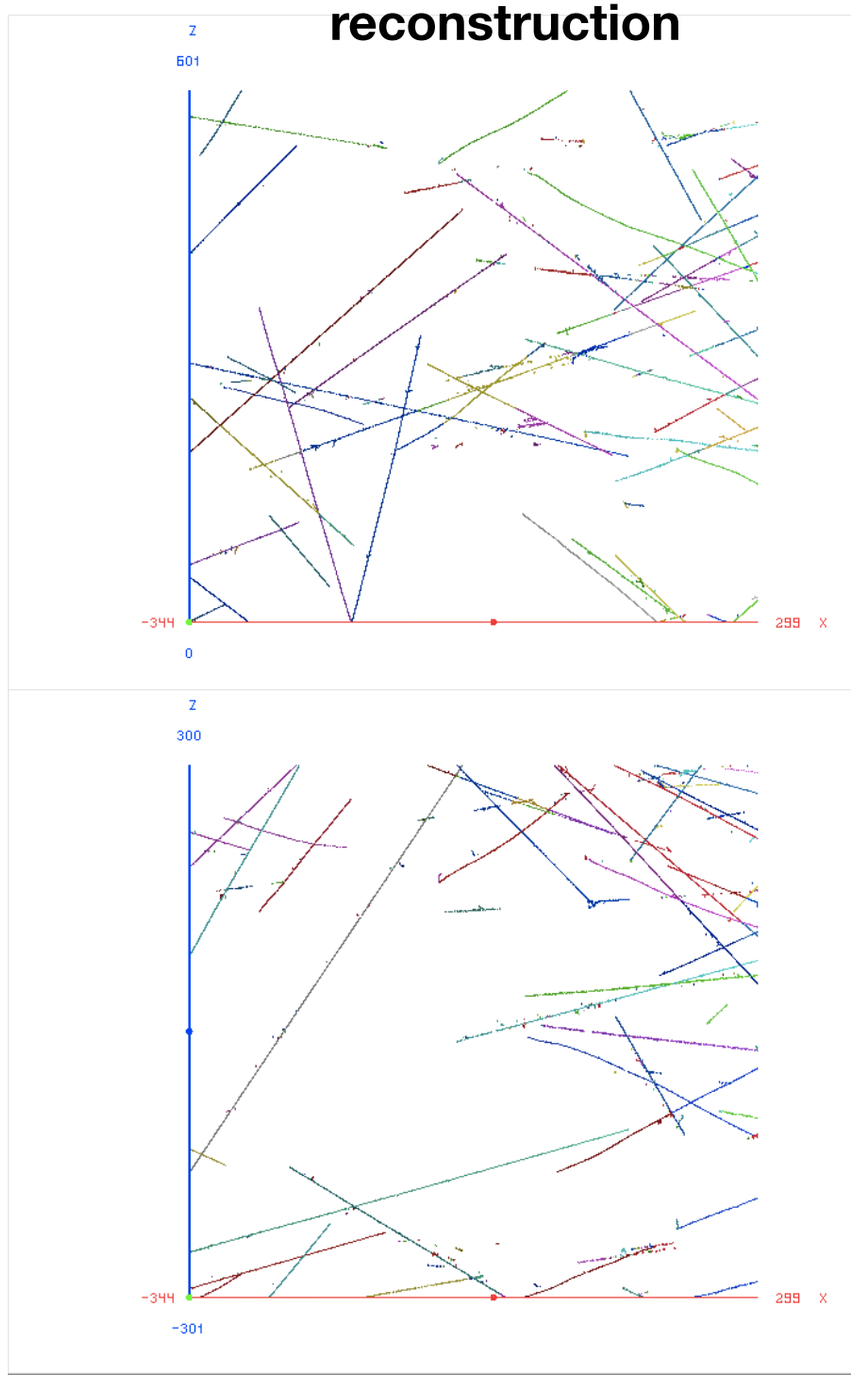
- No fancy diagram
- Ranks all remaining cluster comparisons and repeatedly picks the highest-rank comparison
- Ranking based on
  - **Locally matched fraction**
    - Fraction of local matching scores above threshold
  - **Global matching score**
  - **Number of matched points**
    - The raw number of local matching scores above threshold

Reconstruction performance

- All recent developments have been very focussed on ProtoDUNE-DP due to recent data taking preparation
- All performance assessments have revolved around simulated high multiplicity cosmic ray events

Pandora development iteration	Fraction of cosmic rays correctly reconstructed
Starting point	46%

2D projections of 3D reconstruction

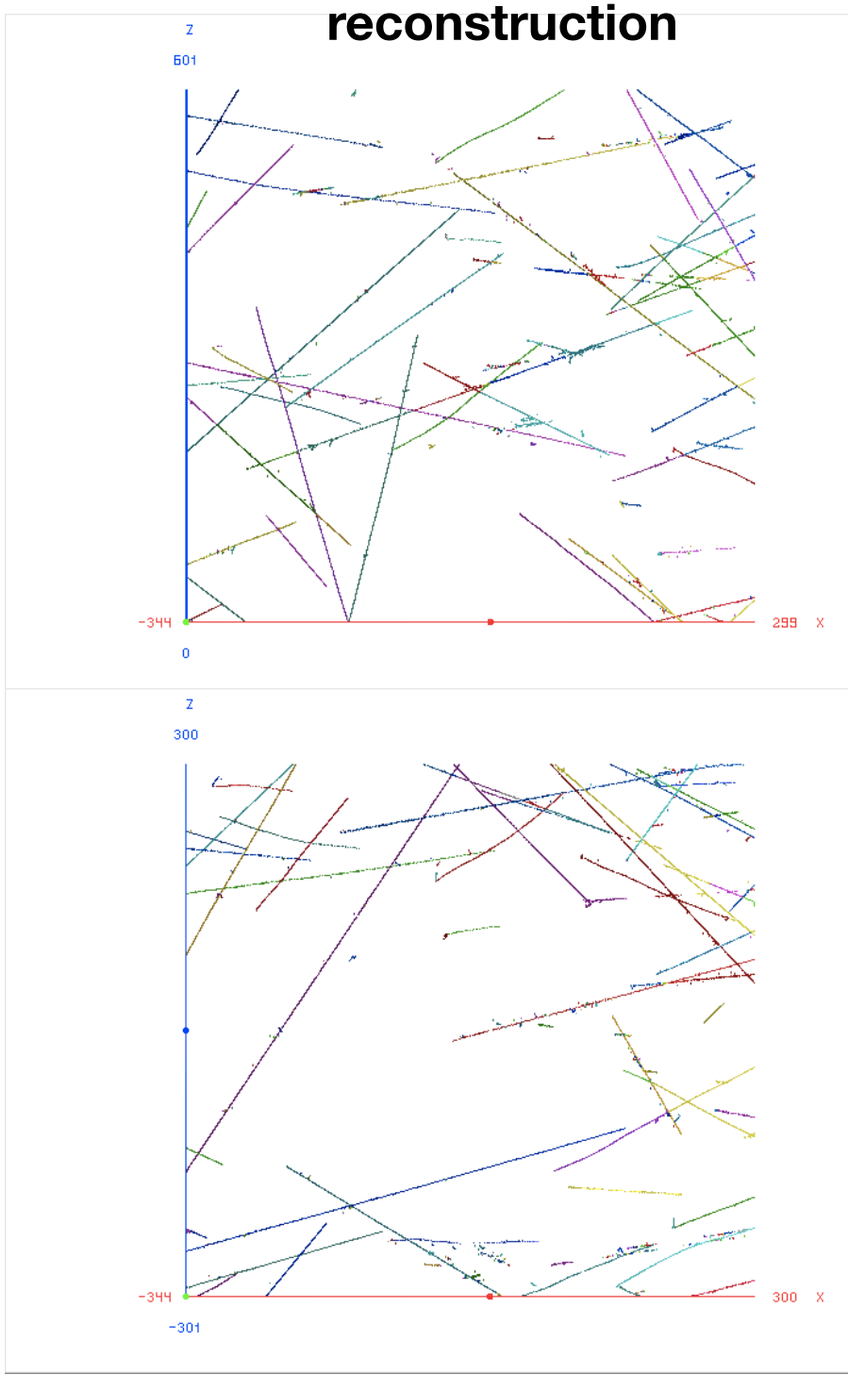


Reconstruction performance

- All recent developments have been very focussed on ProtoDUNE-DP due to recent data taking preparation
- All performance assessments have revolved around simulated high multiplicity cosmic ray events

Pandora development iteration	Fraction of cosmic rays correctly reconstructed
Starting point	46%
+ Hit width 2D clustering and two-view 2D->3D matching (clear tracks tool only)	57%

2D projections of 3D reconstruction

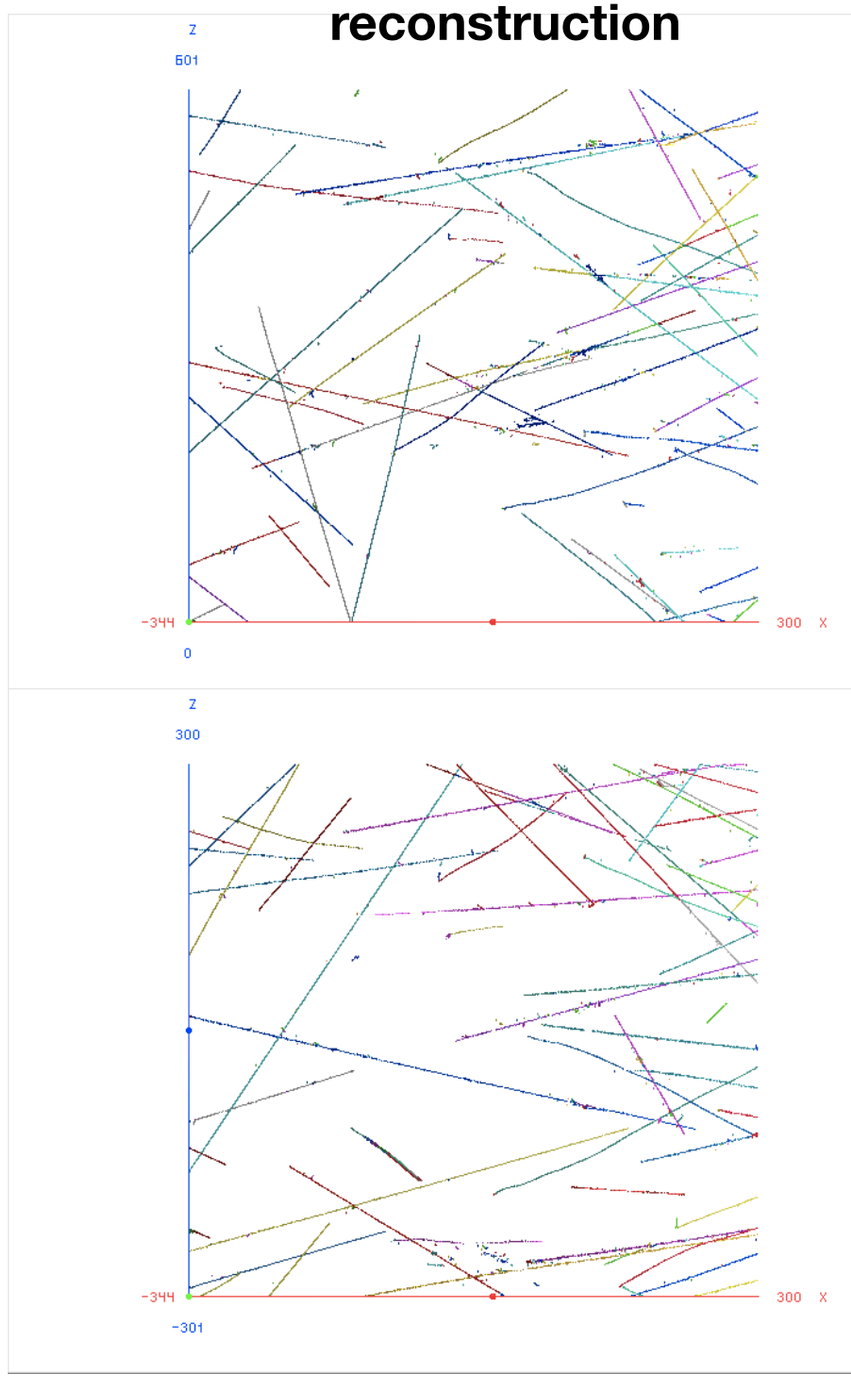


Reconstruction performance

- All recent developments have been very focussed on ProtoDUNE-DP due to recent data taking preparation
- All performance assessments have revolved around simulated high multiplicity cosmic ray events

Pandora development iteration	Fraction of cosmic rays correctly reconstructed
Starting point	46%
+ Hit width 2D clustering and two-view 2D->3D matching (clear tracks tool only)	57%
+ Long tracks and simple tracks tools	76%

2D projections of 3D reconstruction

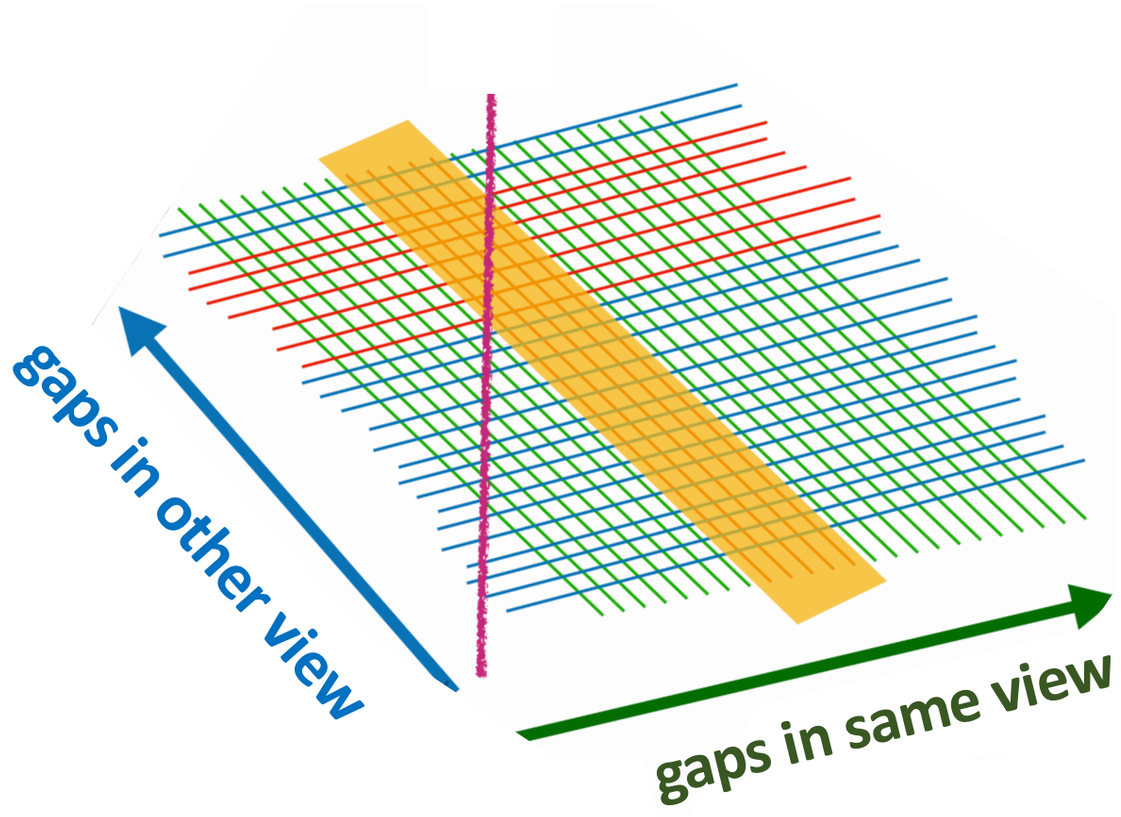


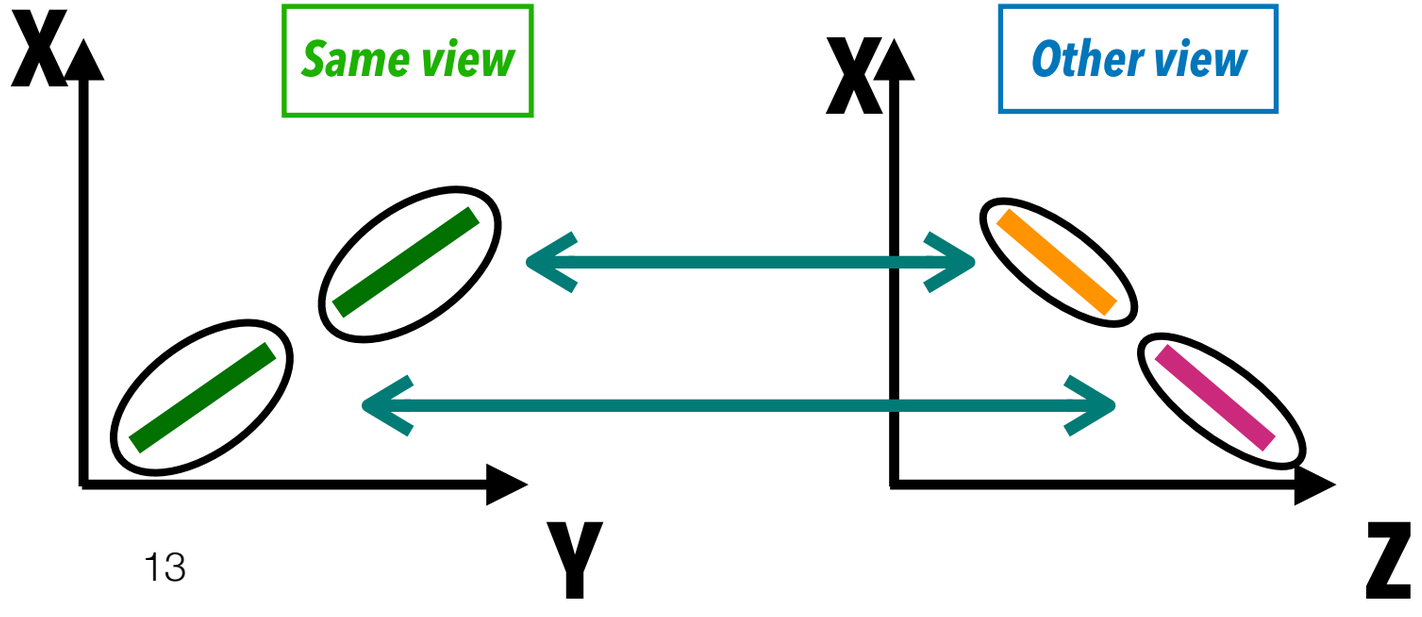
Future work

- Assess/benchmark performance to target improvements

- 2D->3D matching tool for matching across the ‘other-view’ detector gaps

  - See diagrams





Summary

- Pandora is a multi-algorithm approach to pattern recognition, currently deployed in multiple LArTPC experiments

- A few of Pandora's most powerful algorithms require three views to function and so are not appropriate for dual phase-style LArTPCs

- We have expanded/adapted some key areas of Pandora's workflow to harness features of dual-phase style LArTPCs

- There now exists a tailored complete end-to-end reconstruction workflow for dual-phase style LArTPCs in Pandora

    - The workflow already features in ProtoDUNE-DP's offline reconstruction chain

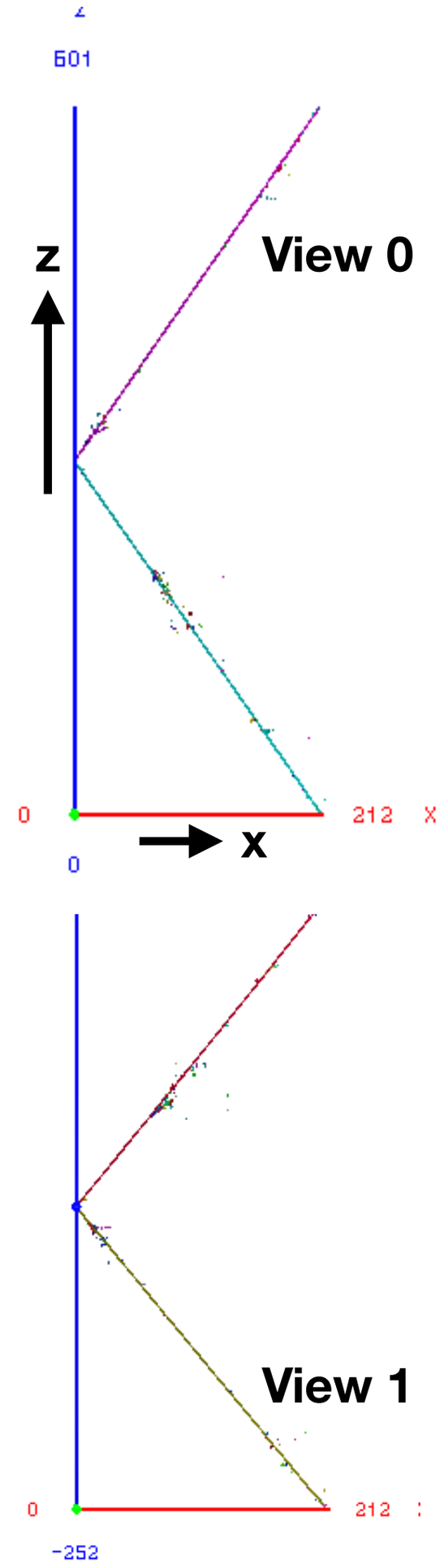
2D->3D matching

- 2D->3D matching takes 2D clusters (e.g. from each wire view) and matching them across views to make 3D objects

- Pandora's main 2D->3D matching algorithm requires a cluster in **three** distinct views to function
  - Combining positions from clusters in two of the views infers a position in the third view. A pseudo chi2 is calculated for inferred vs actual positions along the cluster

- This is problematic for any detector technology which only has two views (e.g. the CRP-based dual-phase LArTPCs)

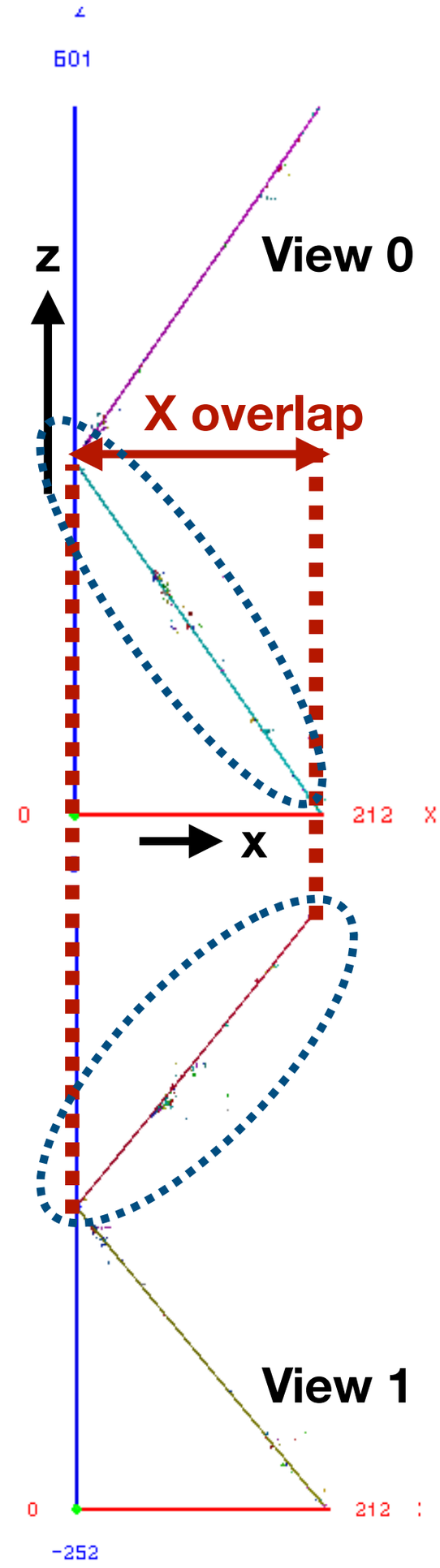
- A solution: Use the charge depositions along the 2D clusters to provide the redundant information to over constrain the matching
- The following slides outline a new pandora algorithm which harnesses the charge depositions to help inform the 2D->3D matching in a two-view LArTPC



2D->3D matching

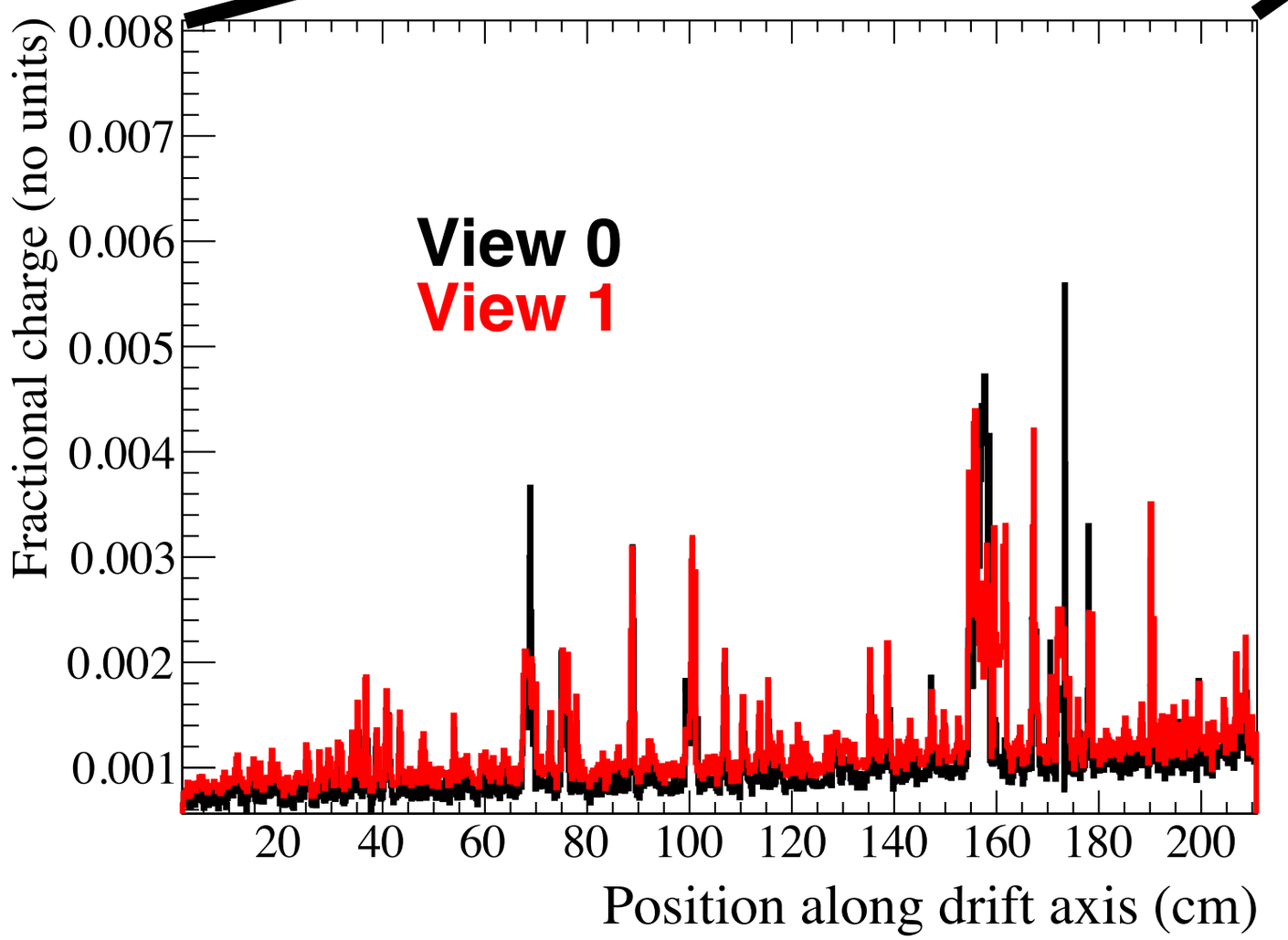
1. Compare every 2D cluster in one view with every cluster in the other view (each one of these comparisons is a **matching candidate**)

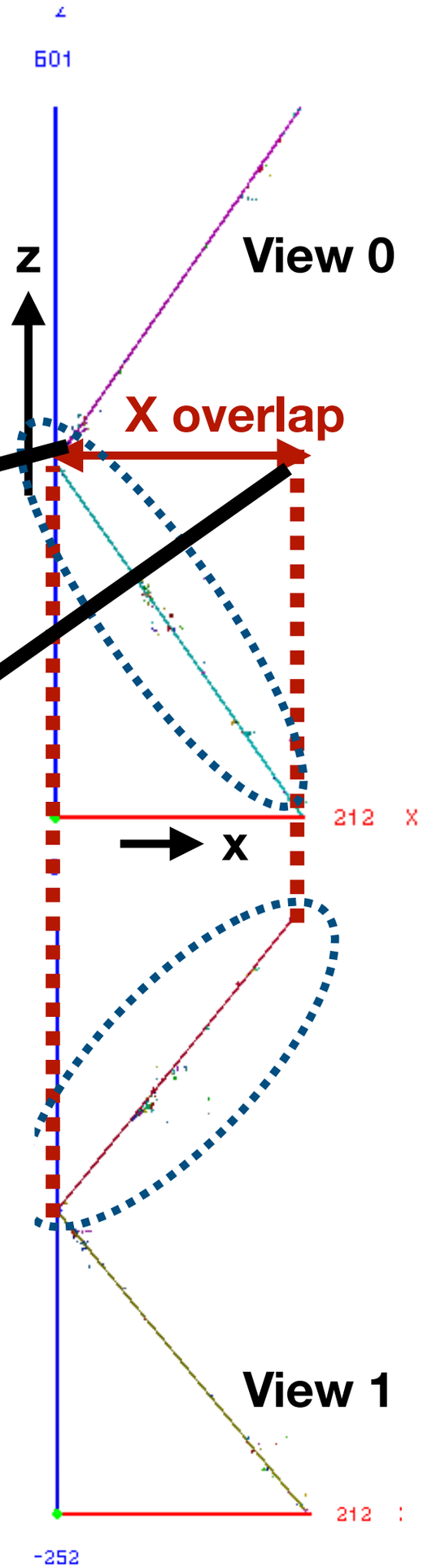
2. For each matching candidate, find the region along the drift coordinate that the two clusters share (the **X overlap**)



2D->3D matching

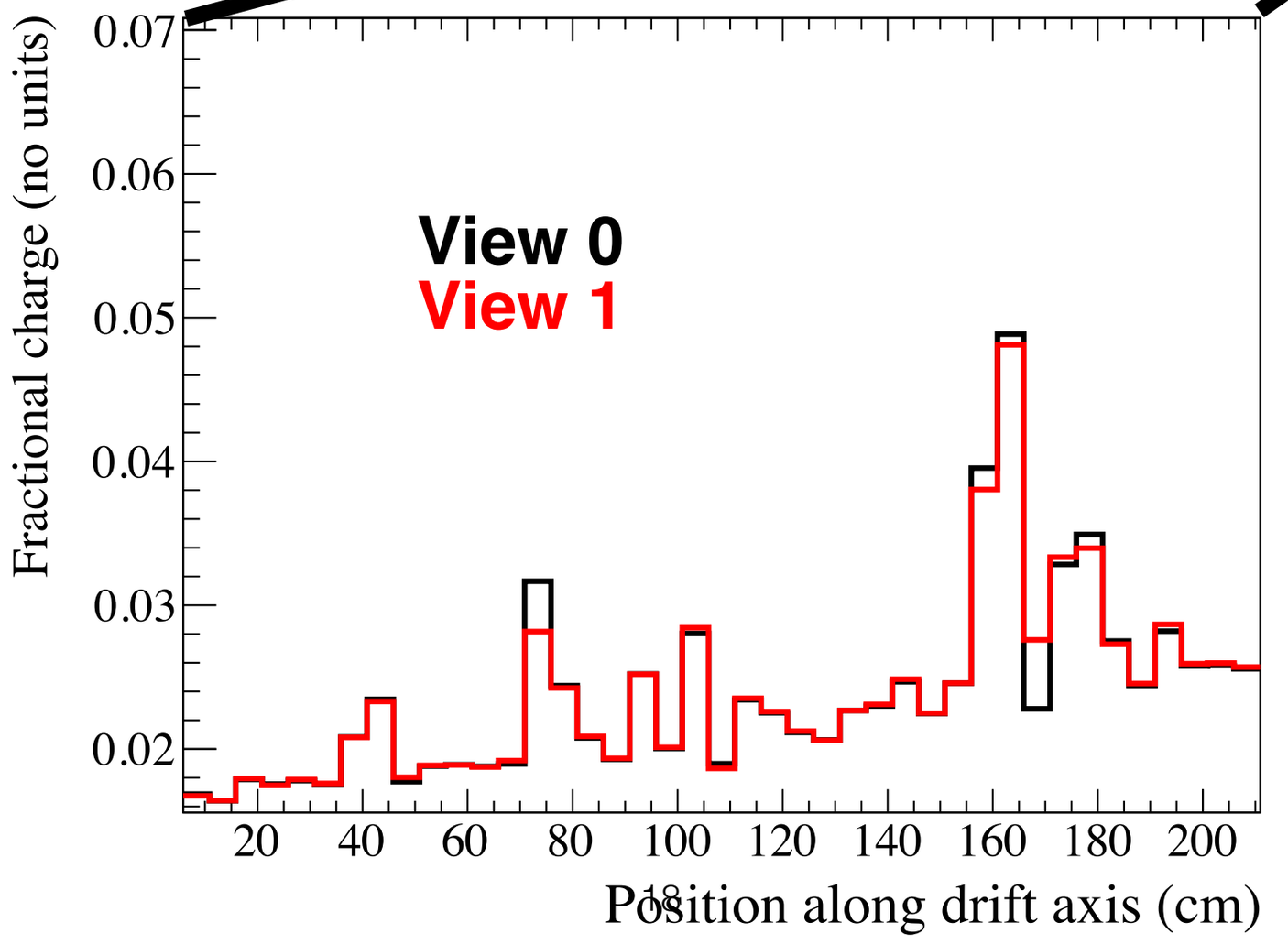
3. Extract the **fractional hit charge profiles** of the two clusters in the X overlap

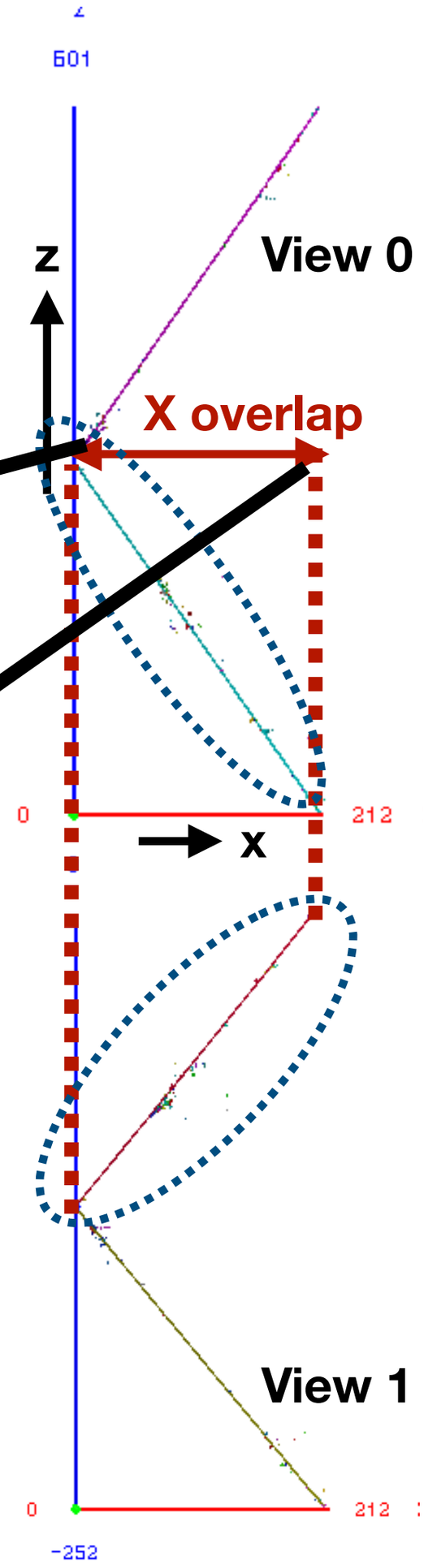




2D->3D matching

4. **Resample/downsample** the two charge profiles so that they are **equally sampled** along the x-axis

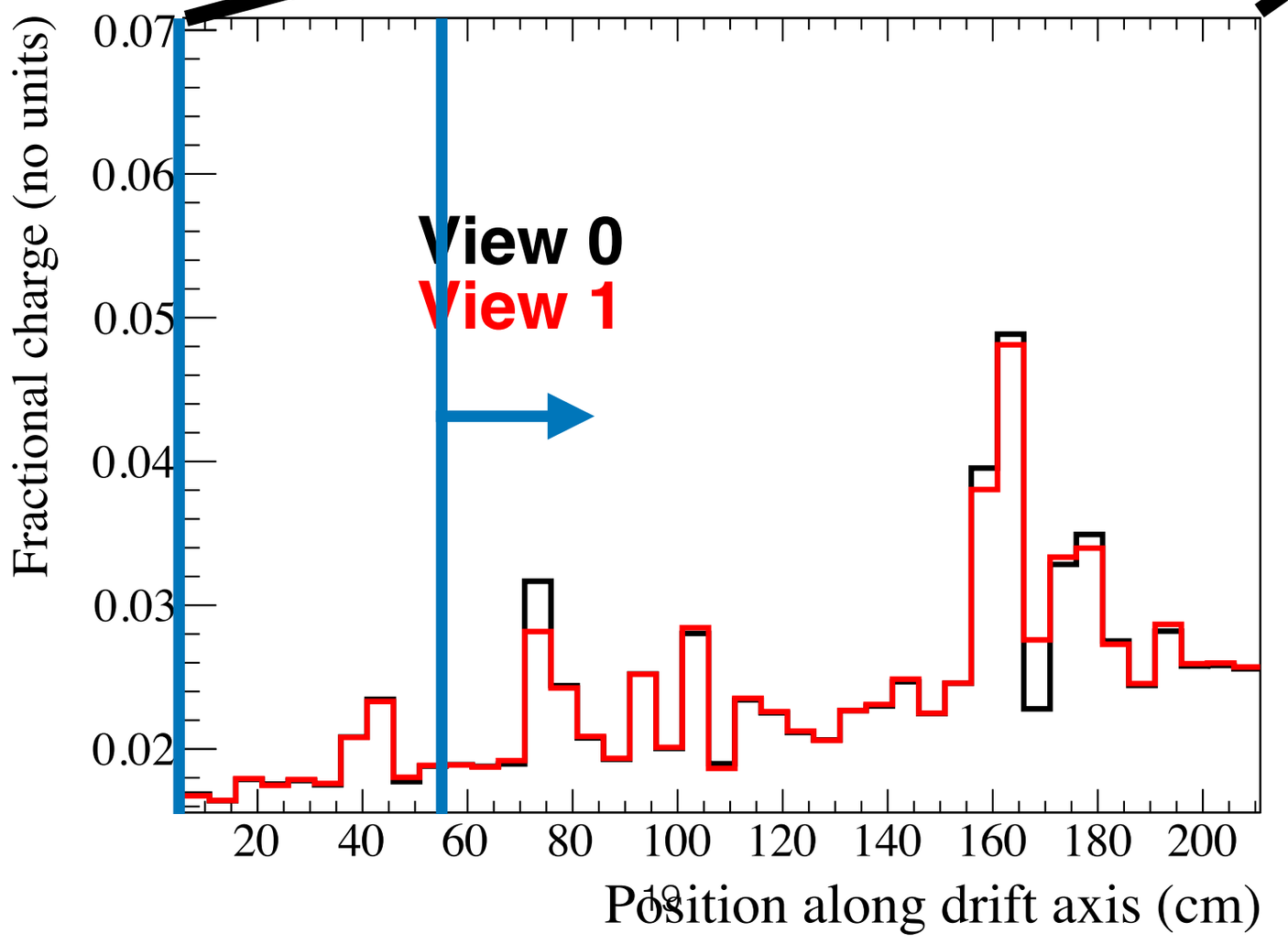


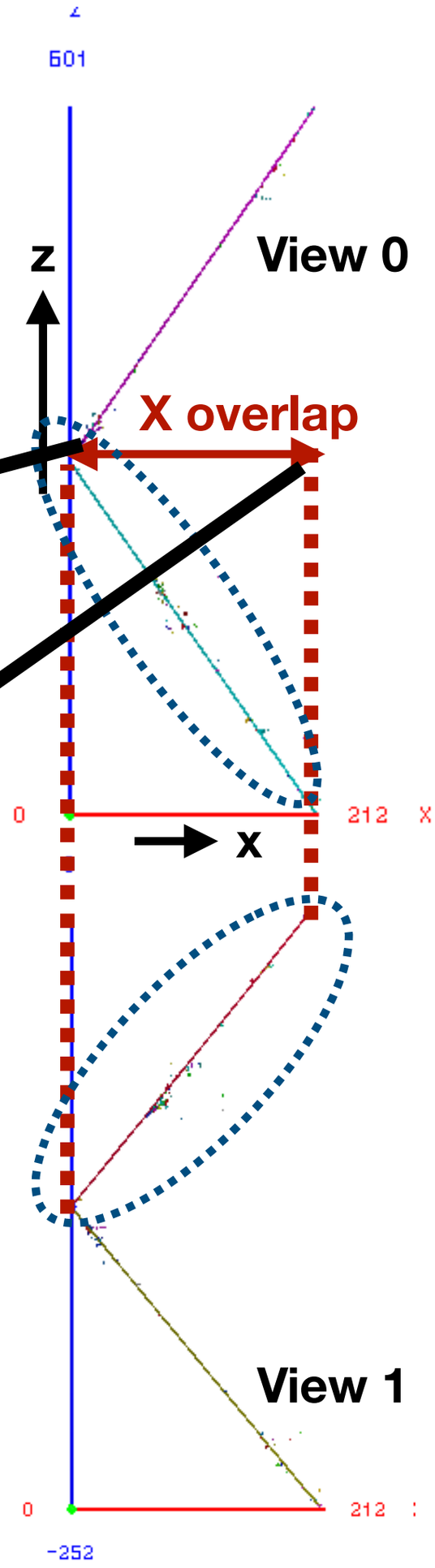


2D->3D matching

5. **Slide a window across the profiles.**
   Calculate the correlation coefficient p-value (*p*) for the points in the window

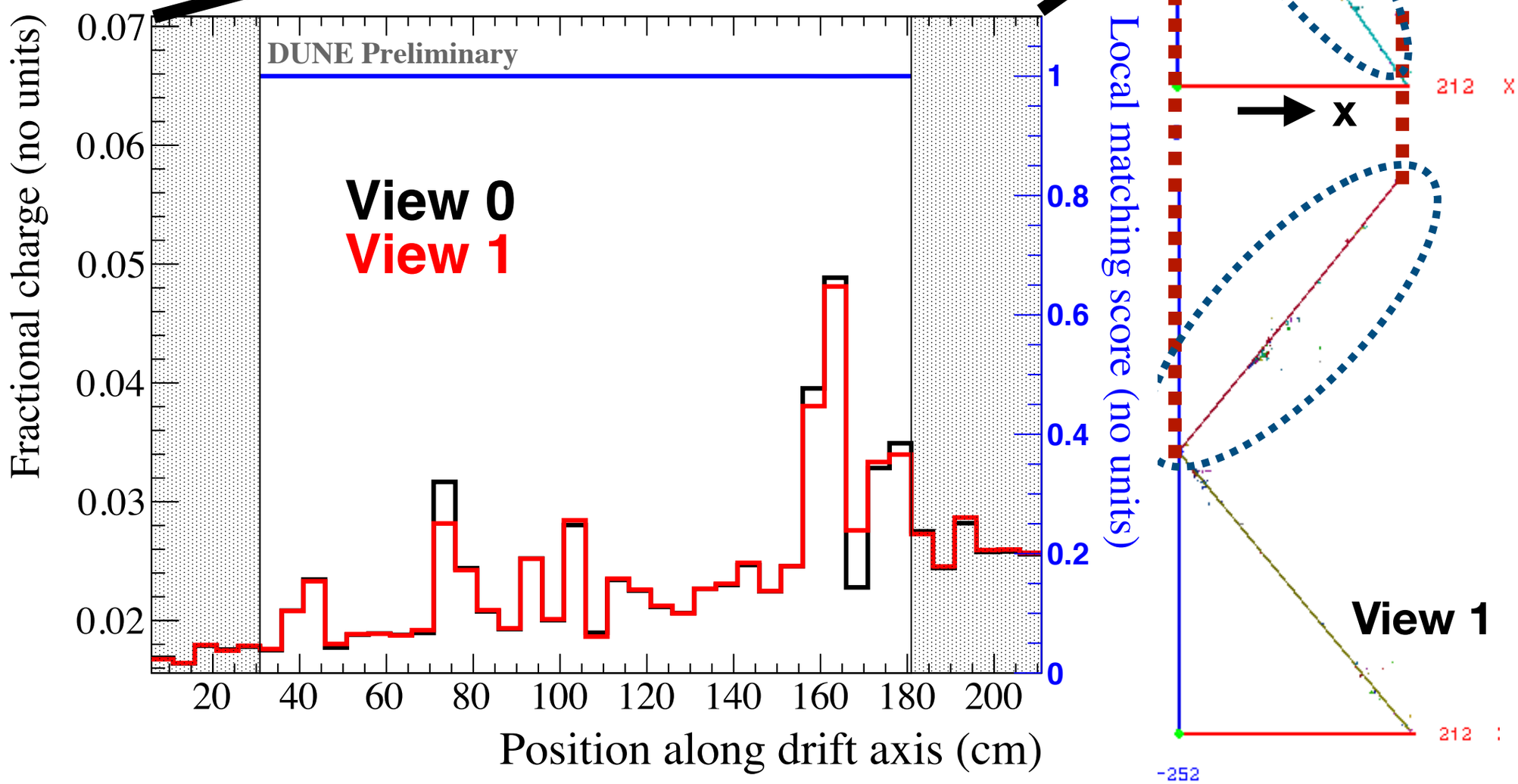
   • Define ***L*=1-*p*** for each window



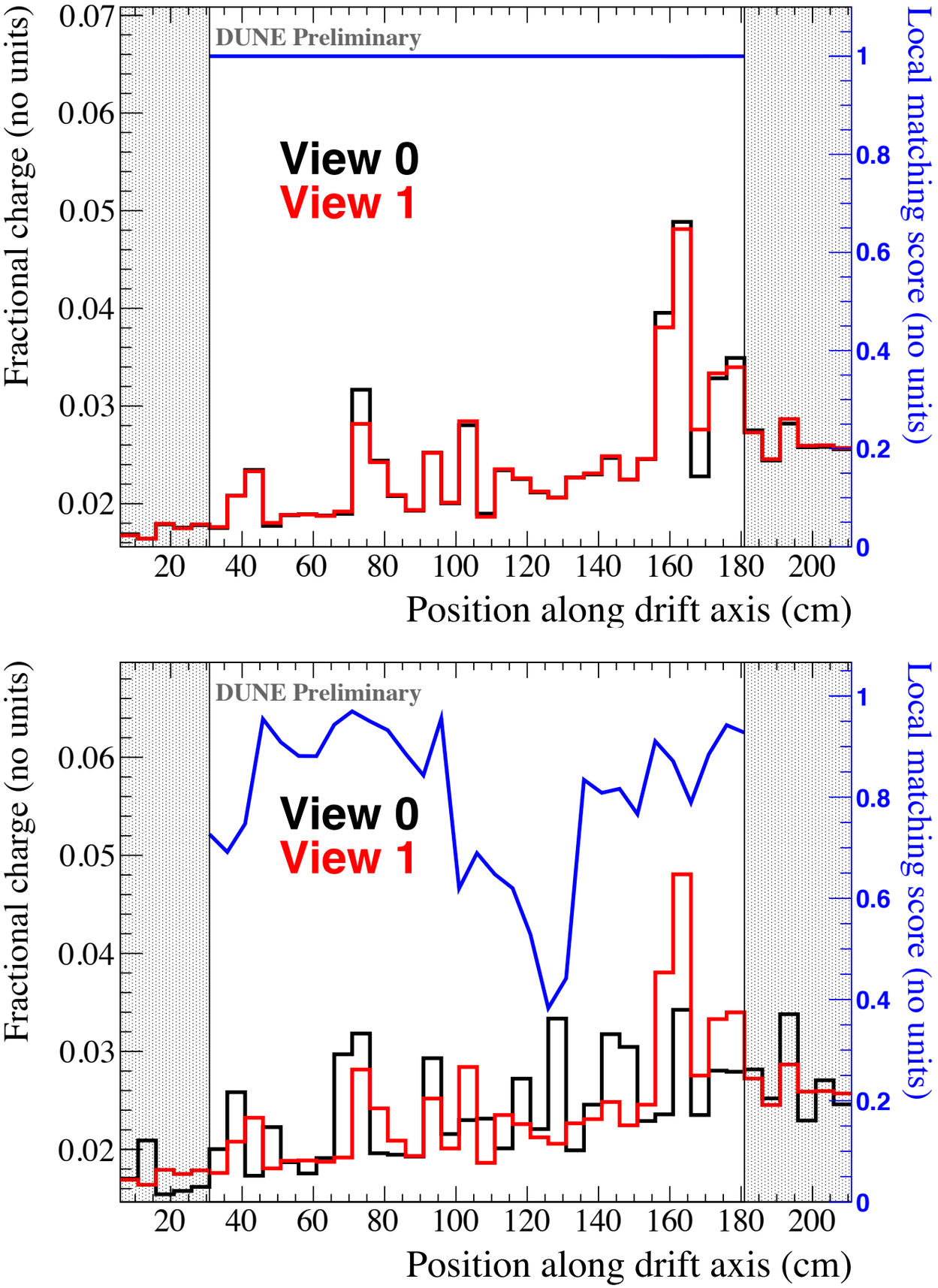


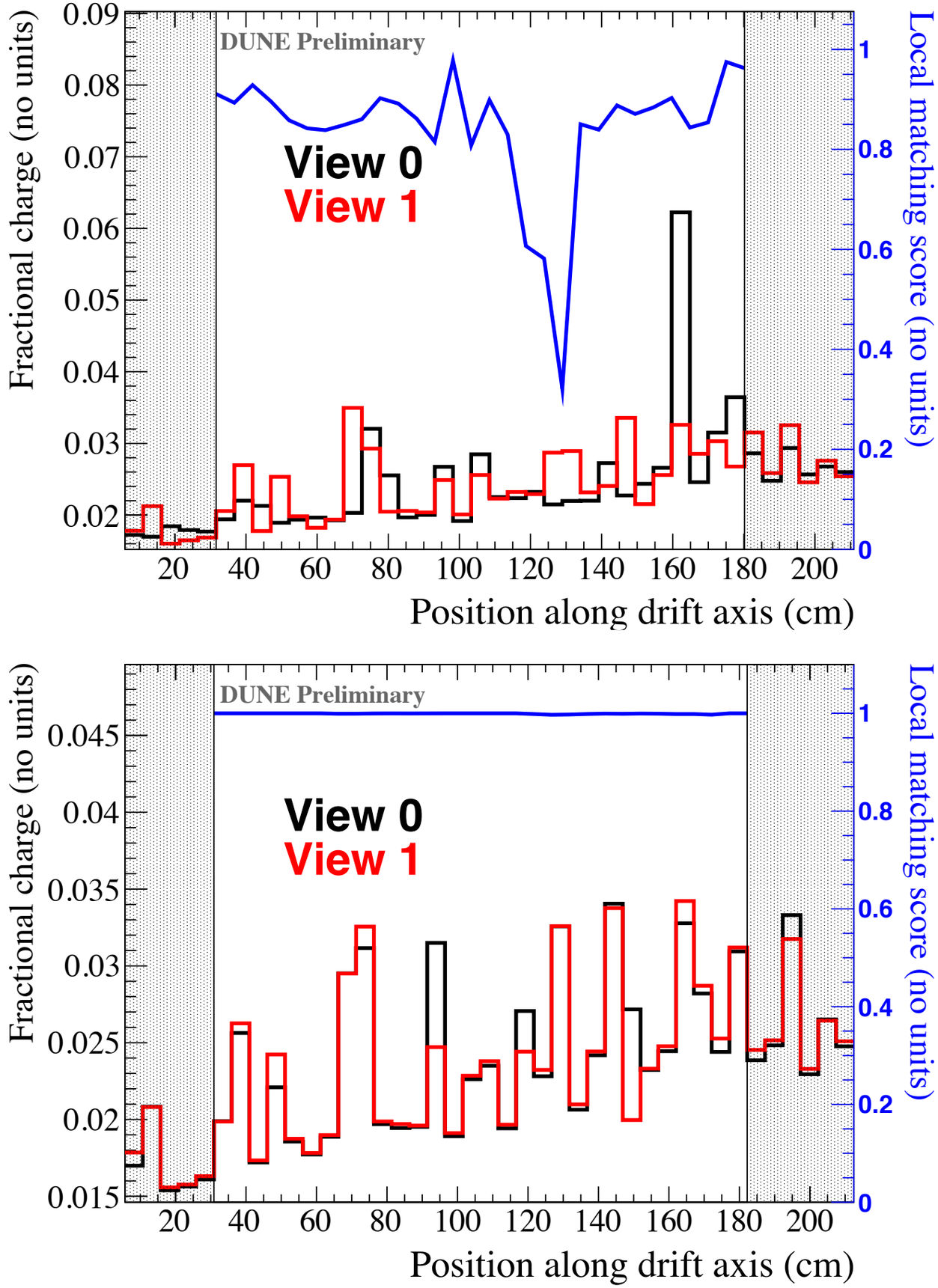
2D->3D matching

6. Fraction of windows with *L*, local matching score (blue line on plot), above threshold (currently 0.99) indicates a good match

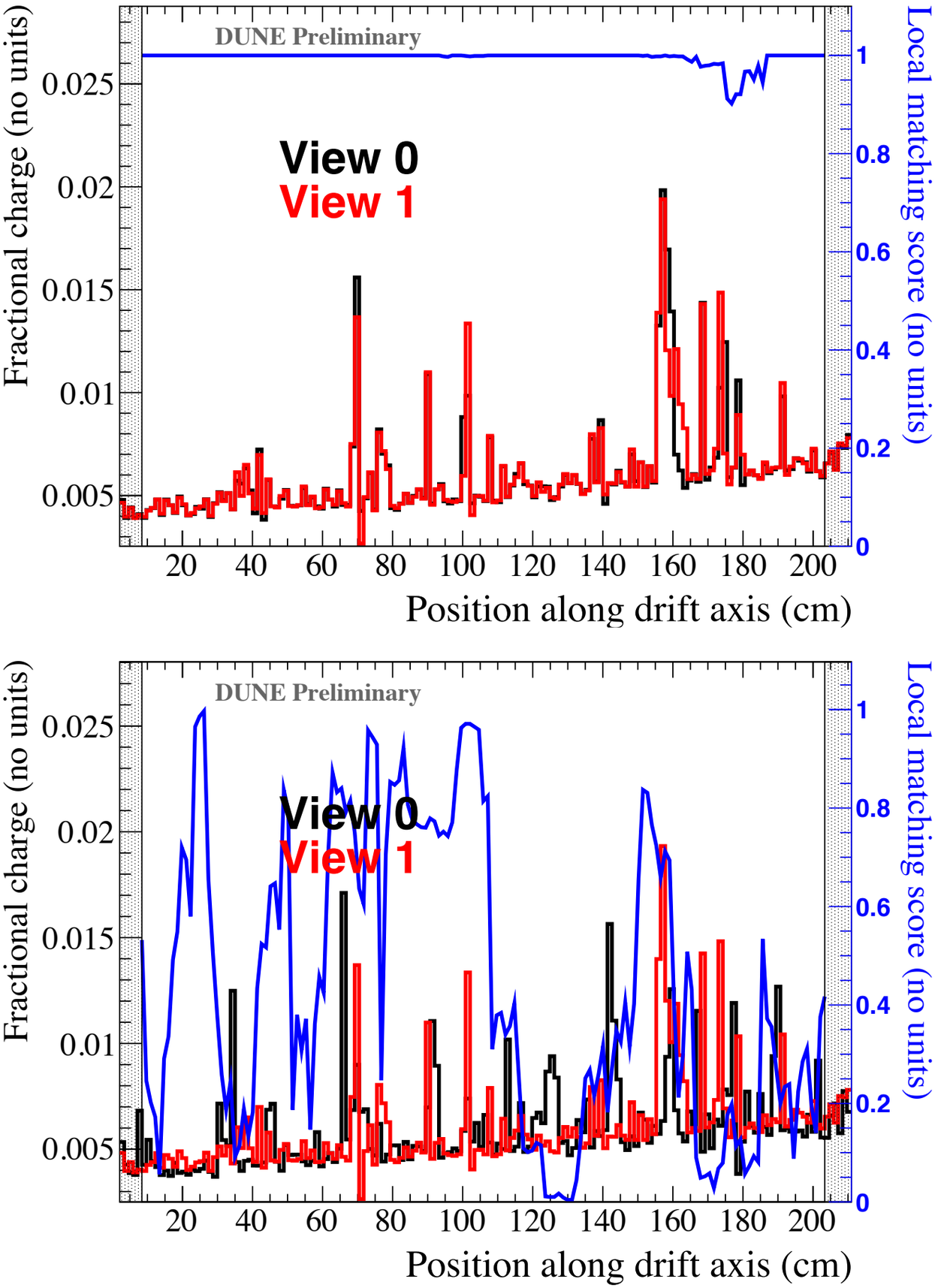


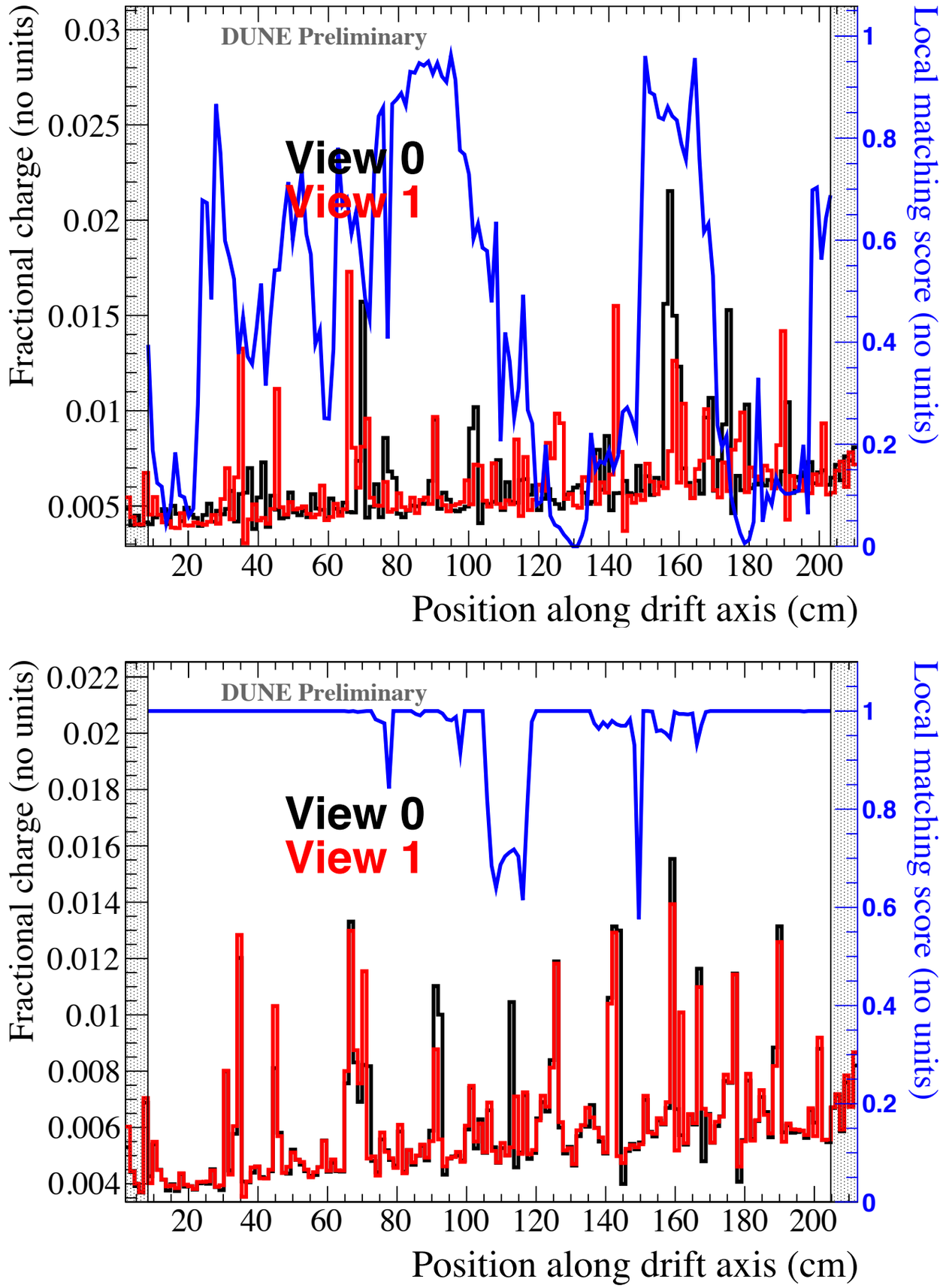
Example 2D->3D matching (MC di-muon event, Nu2020 poster, downsampling factor == 20)





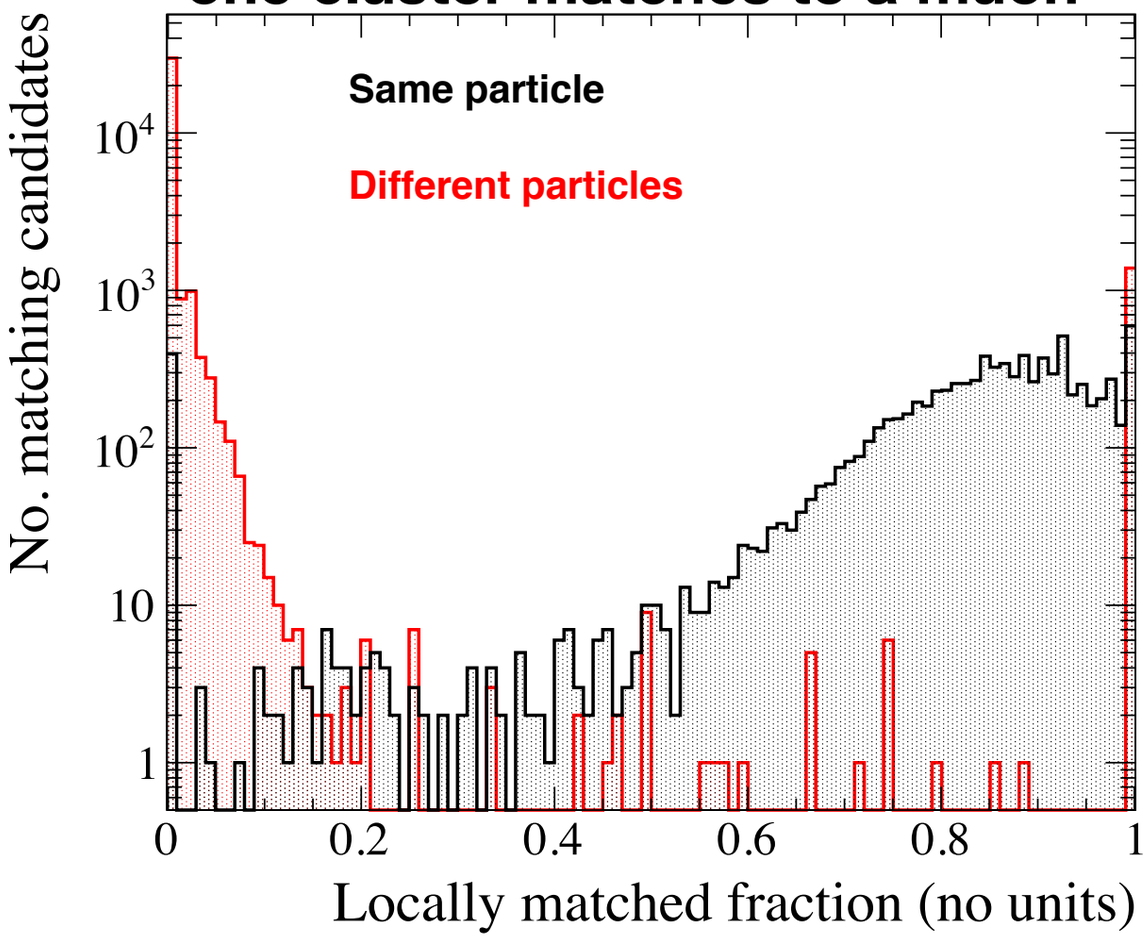
Same di-muon event, downsampling factor == 5 (factor currently used in the codebase)





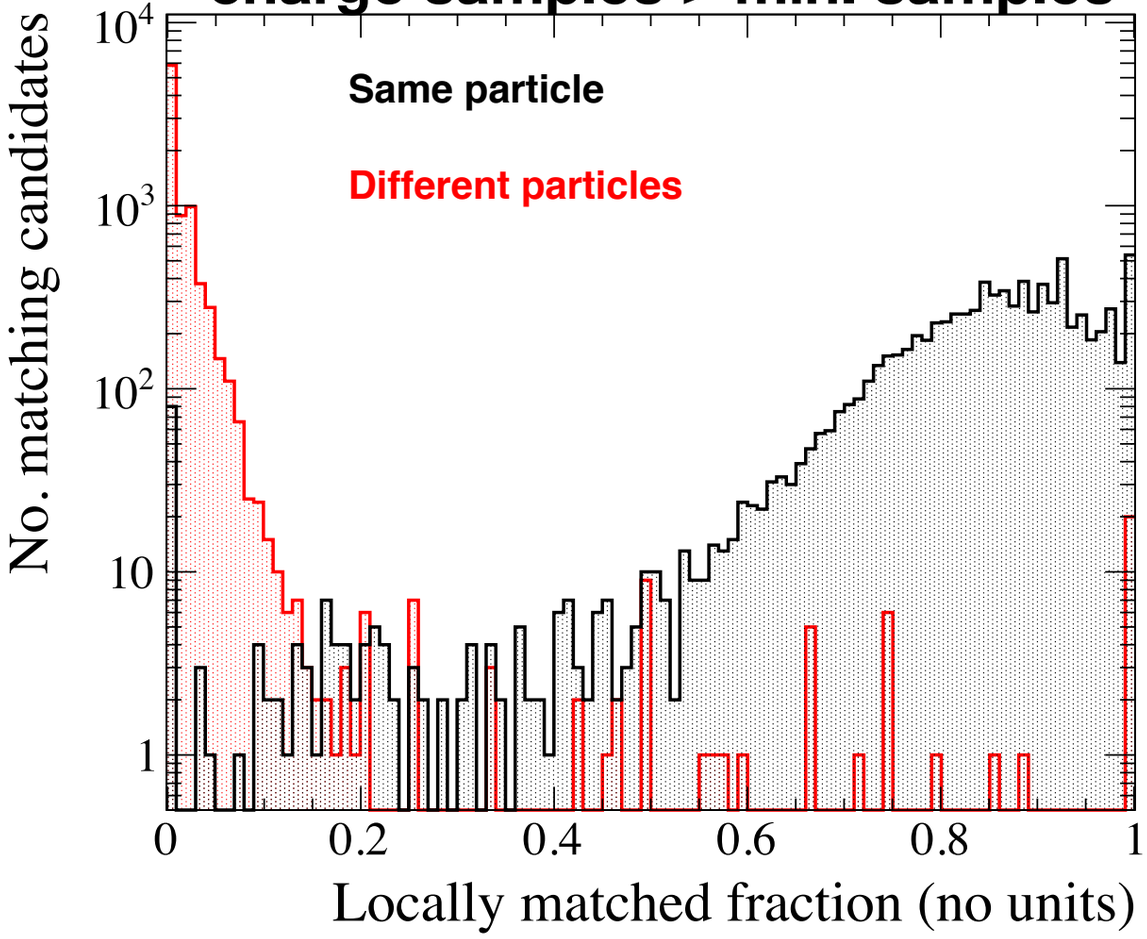
Matching performance (10 GeV di-muon, complete overlap)

Matching cands. where at least one cluster matches to a muon



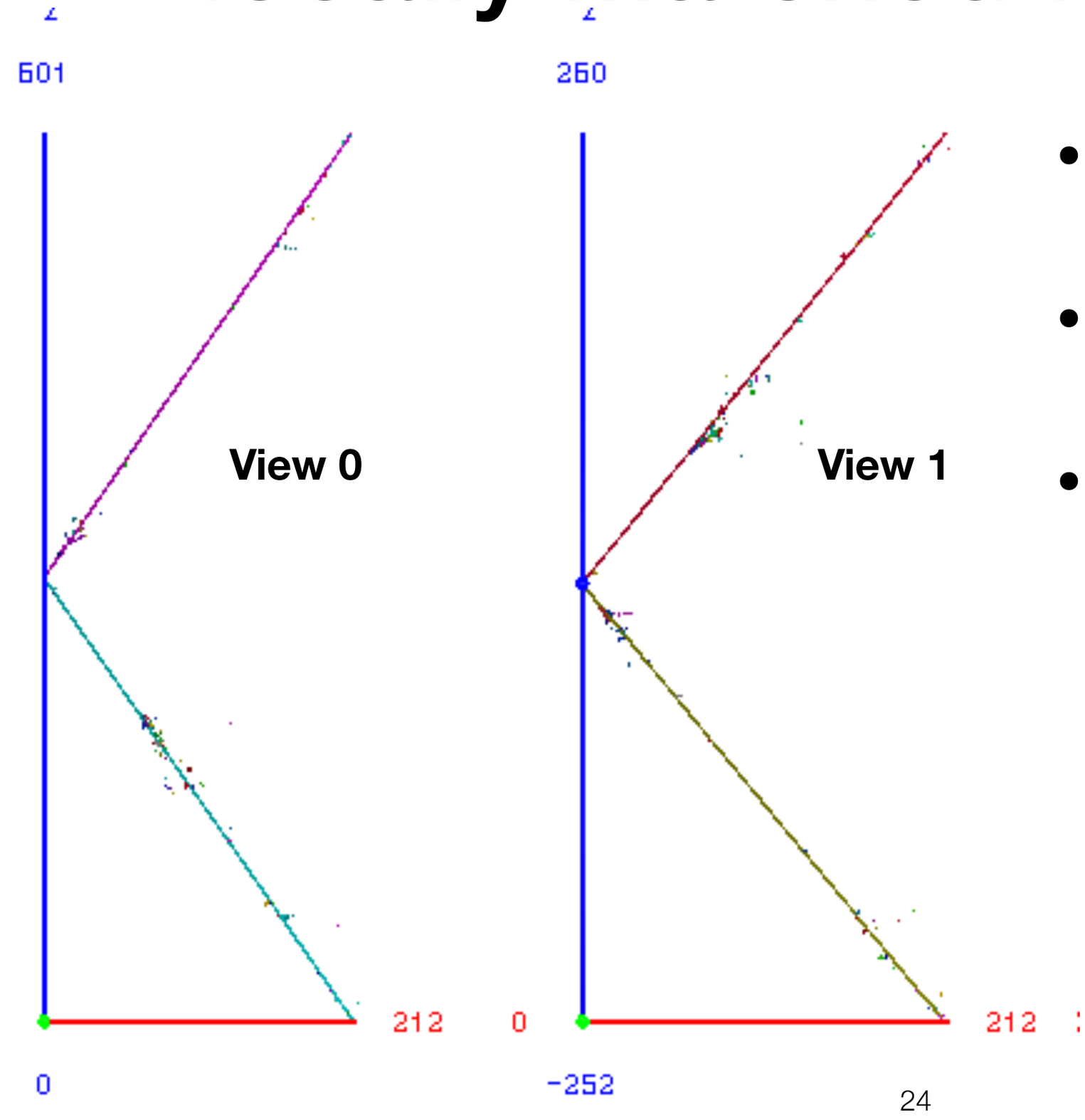
- Excess of incorrect matches at local. match. frac. == 1

Same as left and also number of charge samples > min. samples



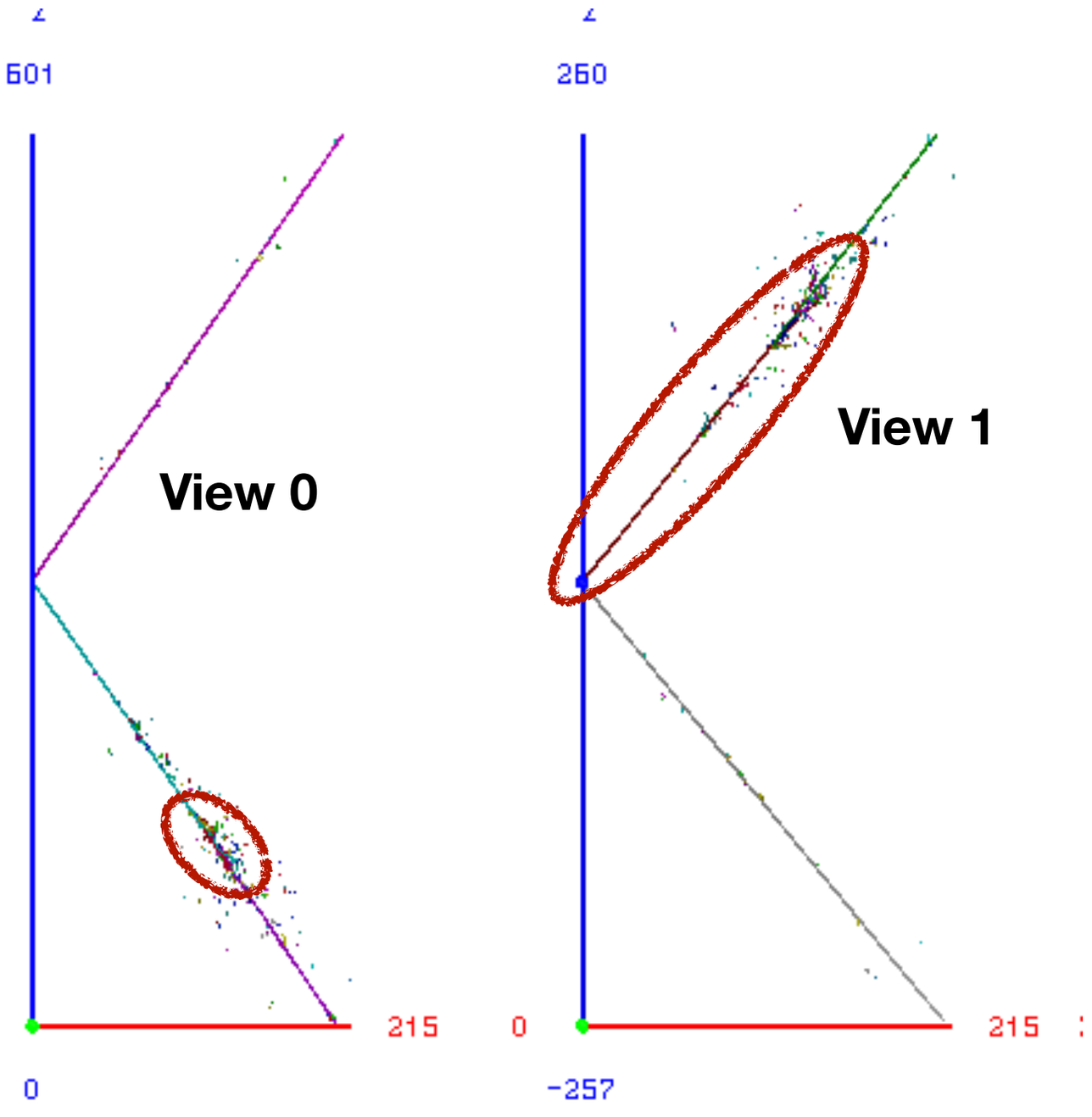
- Excess reduced when rejecting matching candidates with the minimum number of sampling points
- 4% of the correct matches are also dropped (almost all from the 0 bin)

Example incorrect match with locally matched fraction==1

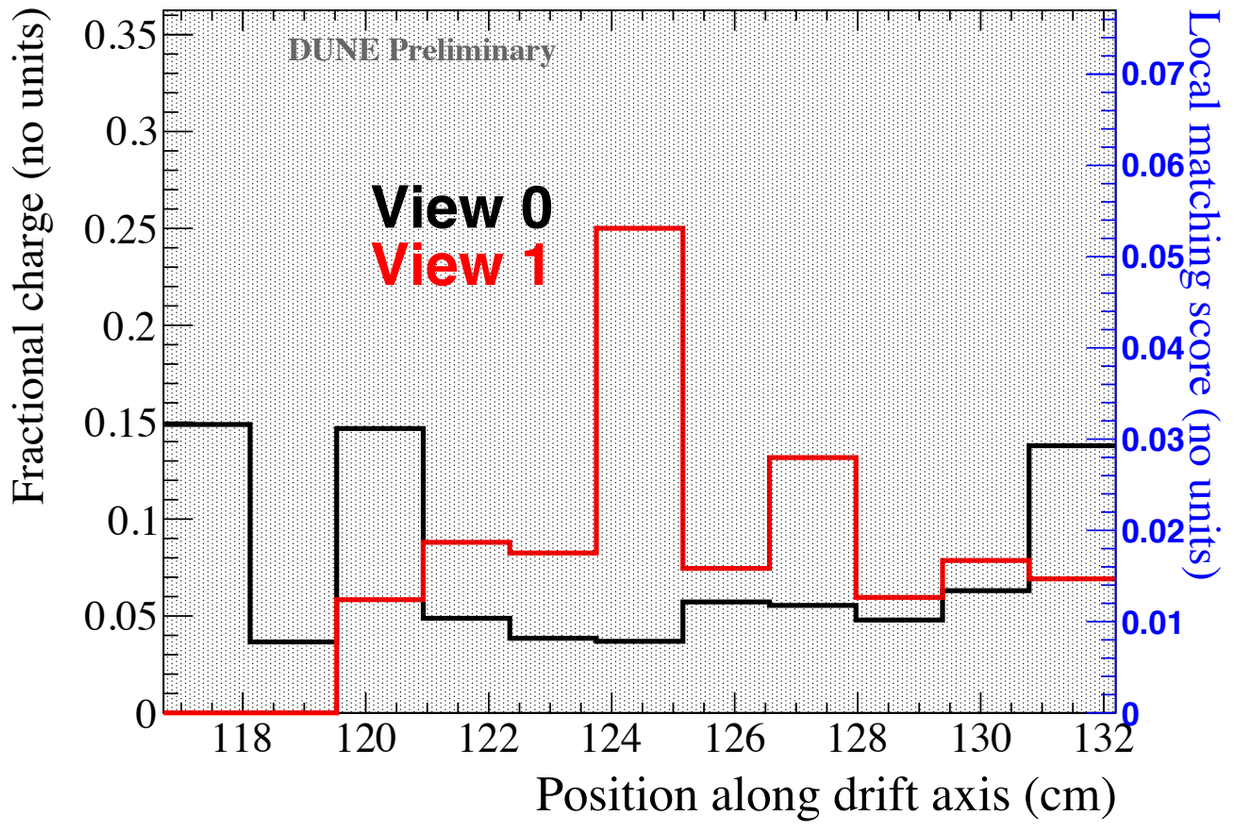


- A lot of EM activity from both muons
- The EM activity results in a lot of small 2D clusters
- The bad match consists of
  - 16 hit electron cluster in view 0
  - A very small segment of one of the muon clusters in view 1

Example ‘correct’ match where locally matched fraction==0

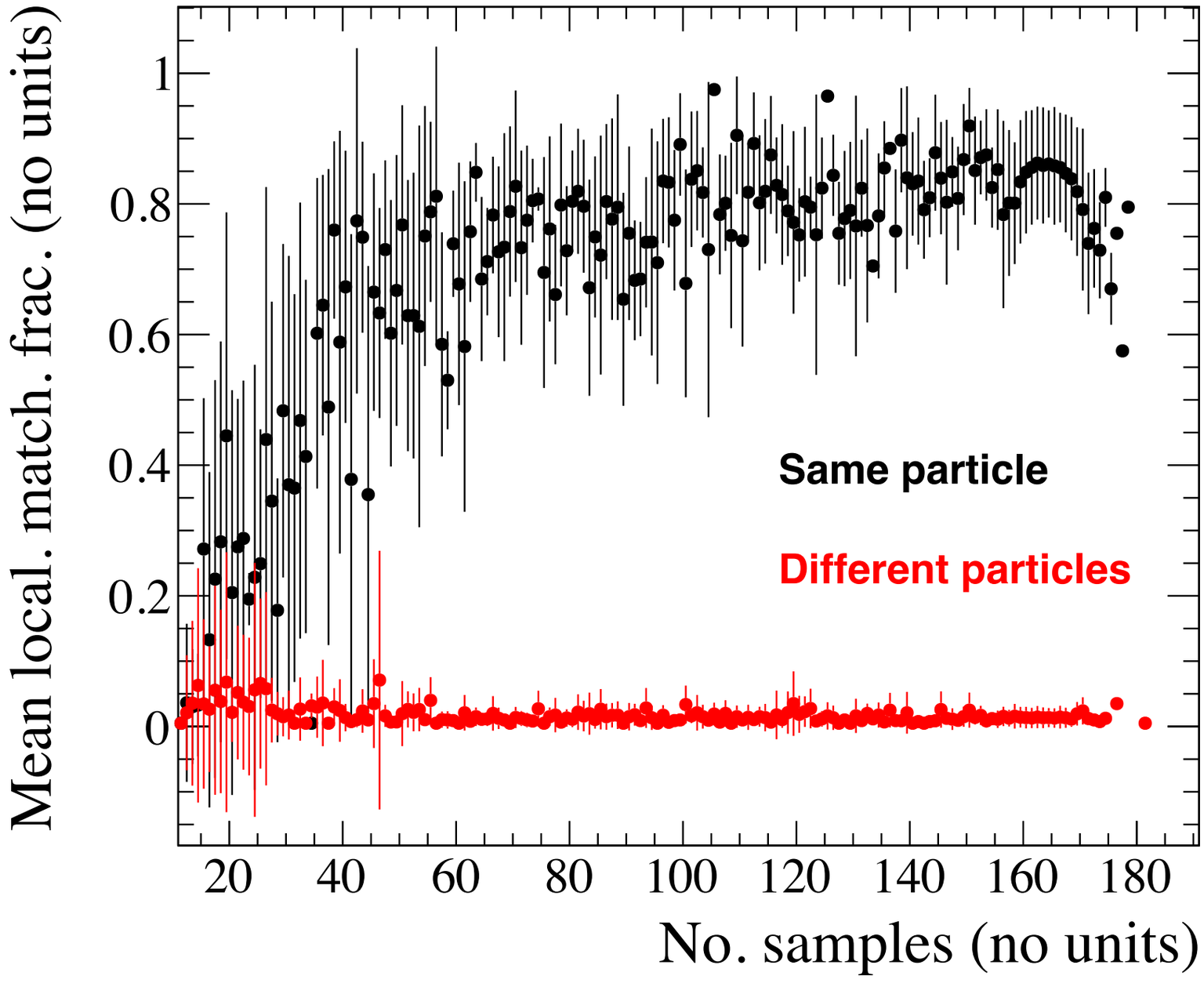


- EM activity is to blame, again
- This match consists of
  - Modestly sized view 0 cluster that truth matches to primary muon
  - A much larger primary muon cluster in view 1
  - Both clusters truth match to the same muon



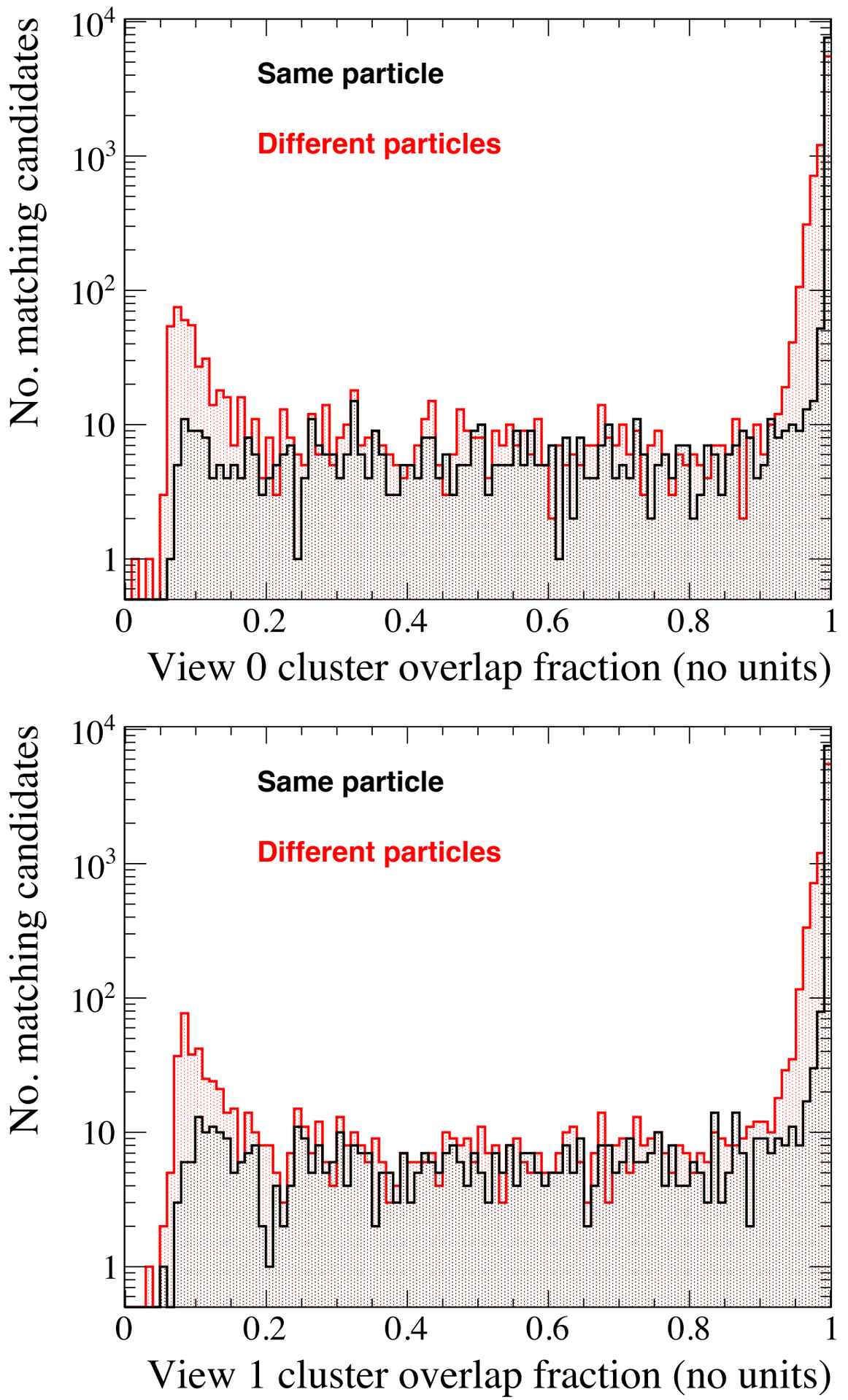
Matching performance (10 GeV di-muon, complete overlap)

- Plot shows the average local match fraction vs the number of samples (after downsampling) in each matching candidate
  - Error bars are the standard deviation (NOT standard error on the mean)
- Matching candidates with a low number of sampling points are dominated by small EM clusters



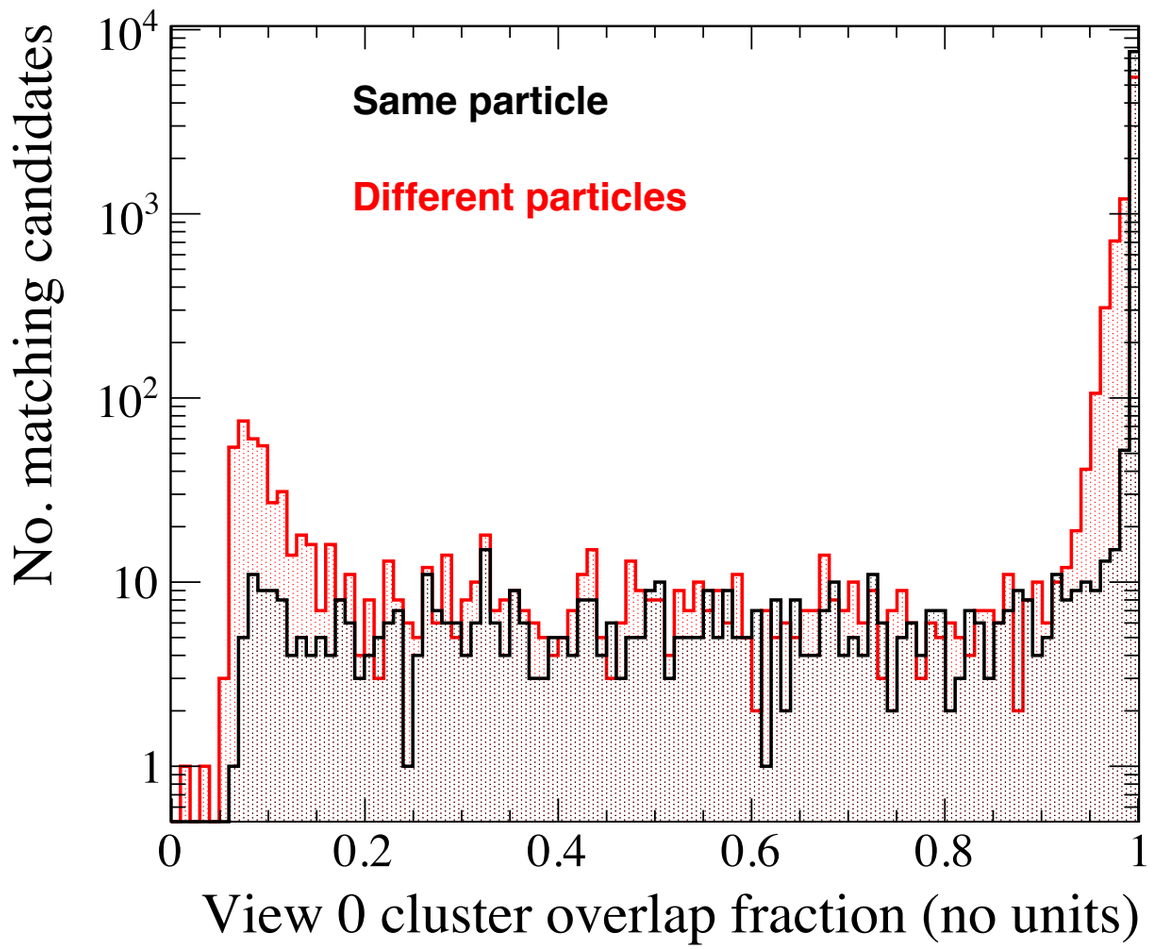
But does the calorimetry-based metric provide any more information than checking the the cluster X geometry?

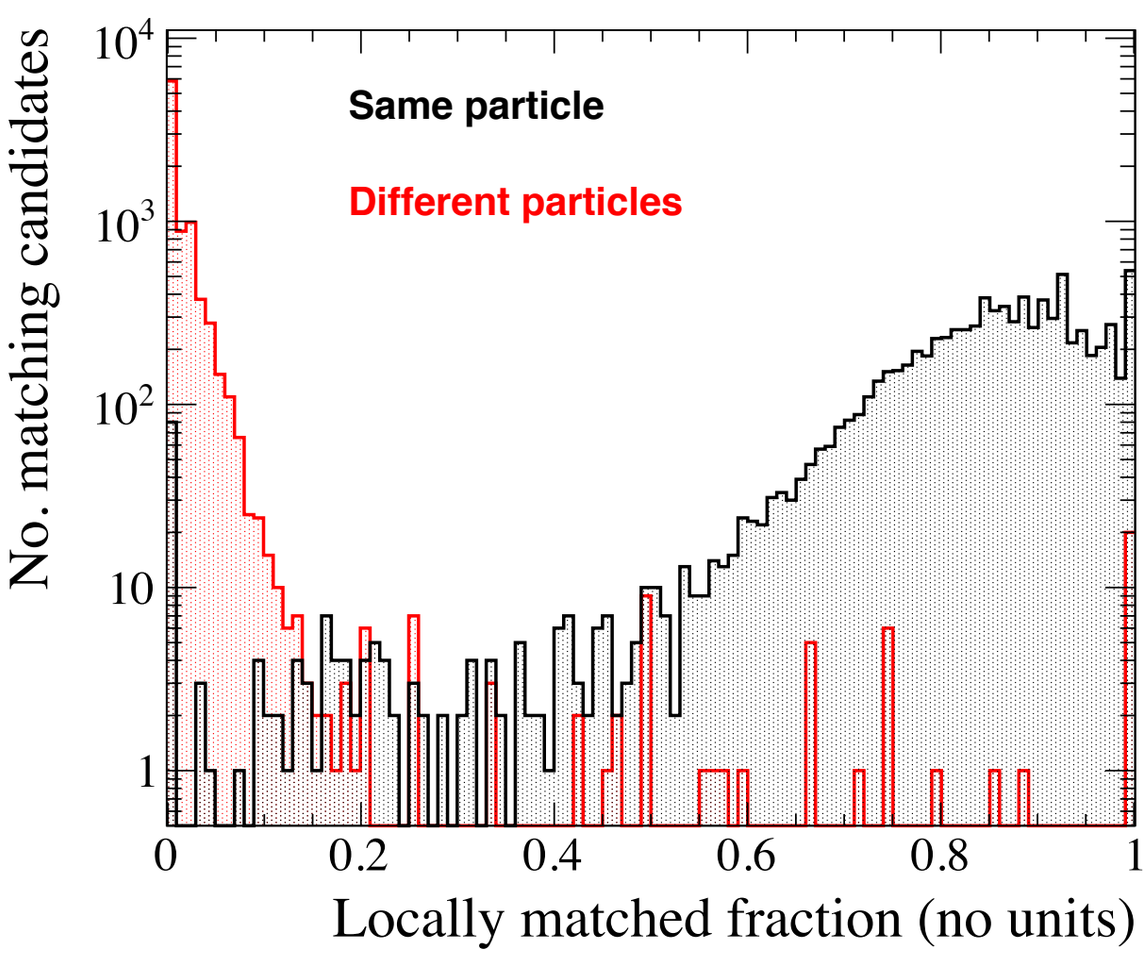
- The two pairs of distributions show the fraction of the cluster span contained in the X overlap
  - For the 10 GeV di-muon sample
- The geometry-based separation is very minimal
- **Disclaimer**: This sample was specifically chosen so that there was maximal ambiguity in the cluster X geometry

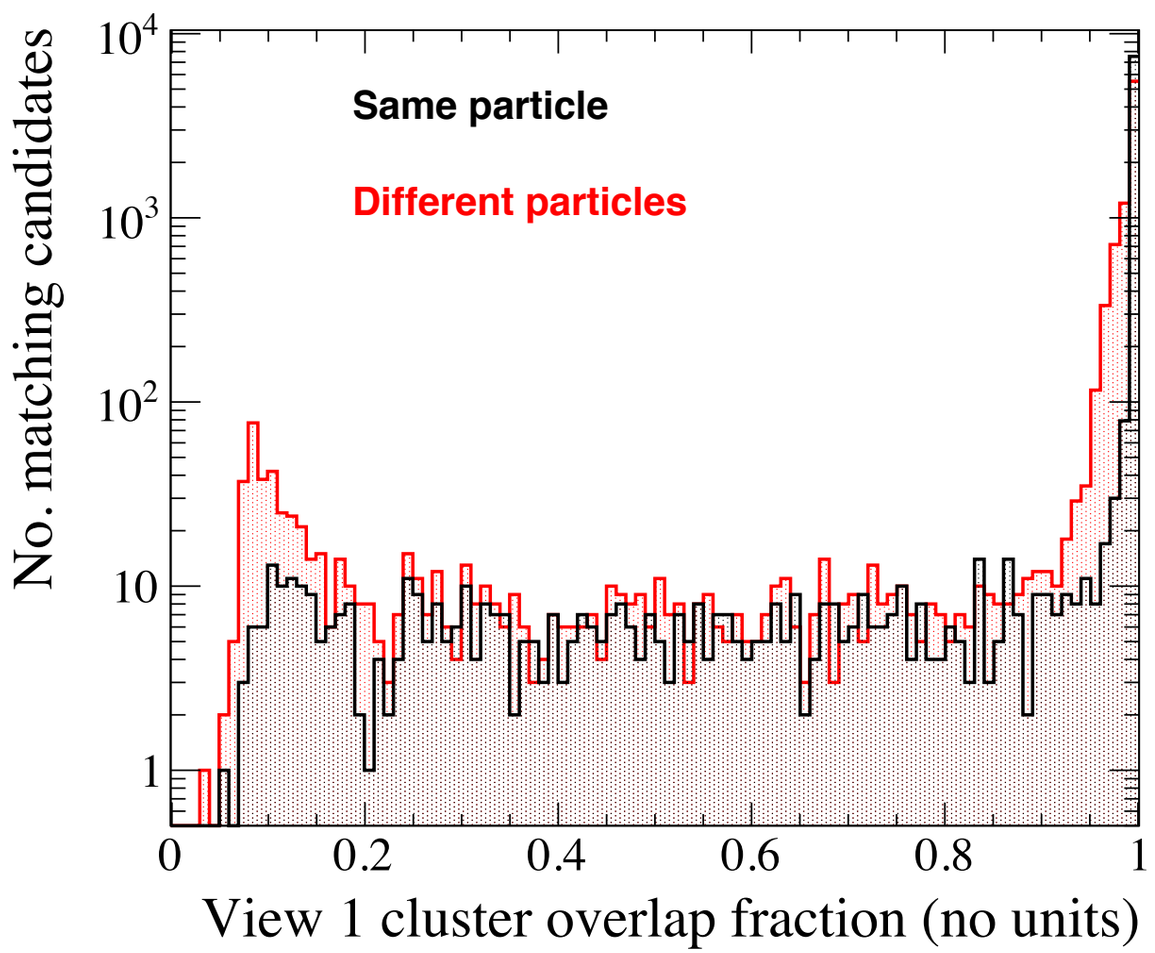


But does the calorimetry-based metric provide more information than checking the the cluster X geometry?

- Compare the geometry-based metrics (RHS plots) with the calorimetry-based metric (bottom)







Correlation coefficient's p-value

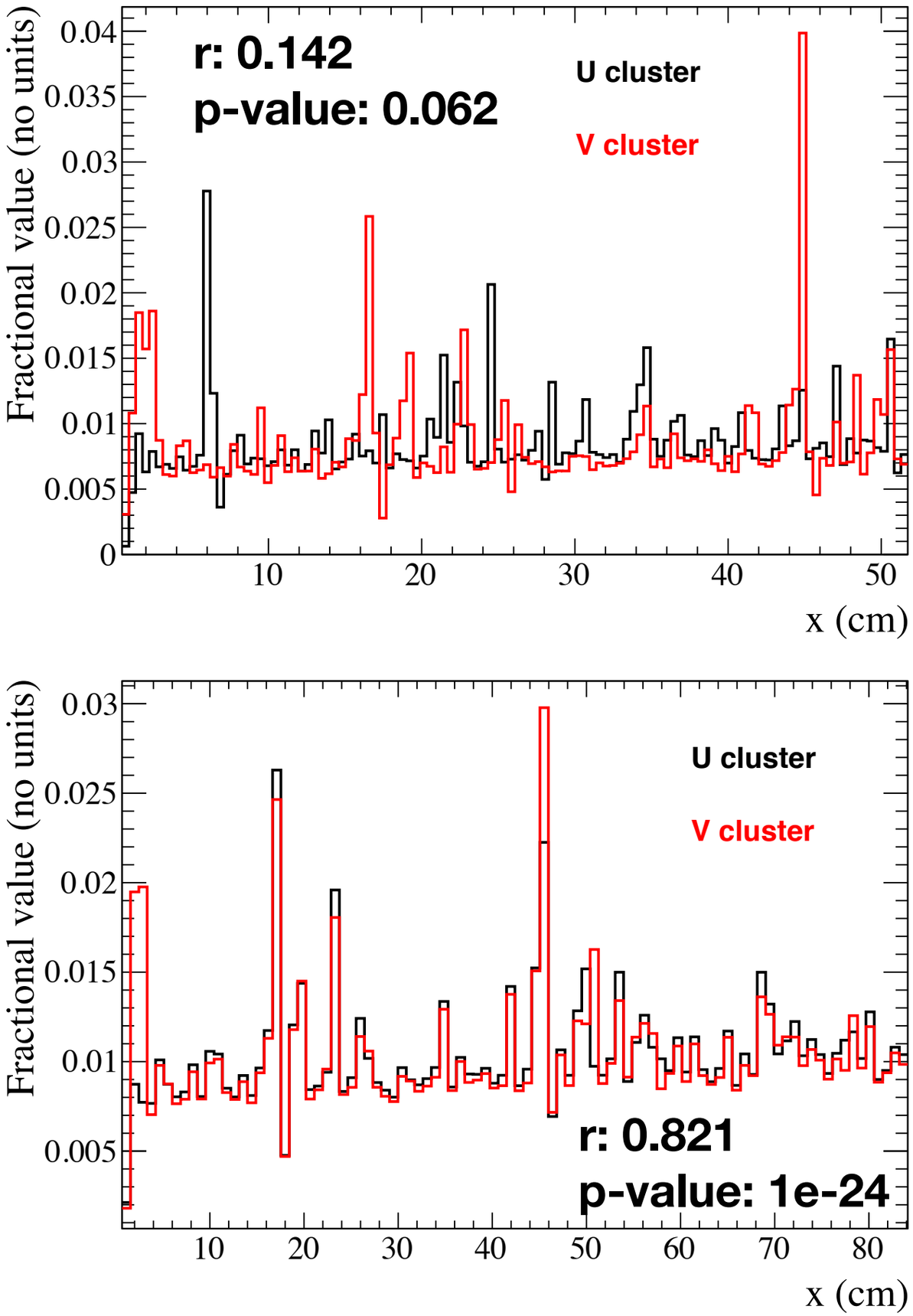
- For uncorrelated bivariate normal distribution pairs, the correlation coefficient follows a Student t-distribution with n-2 degrees of freedom

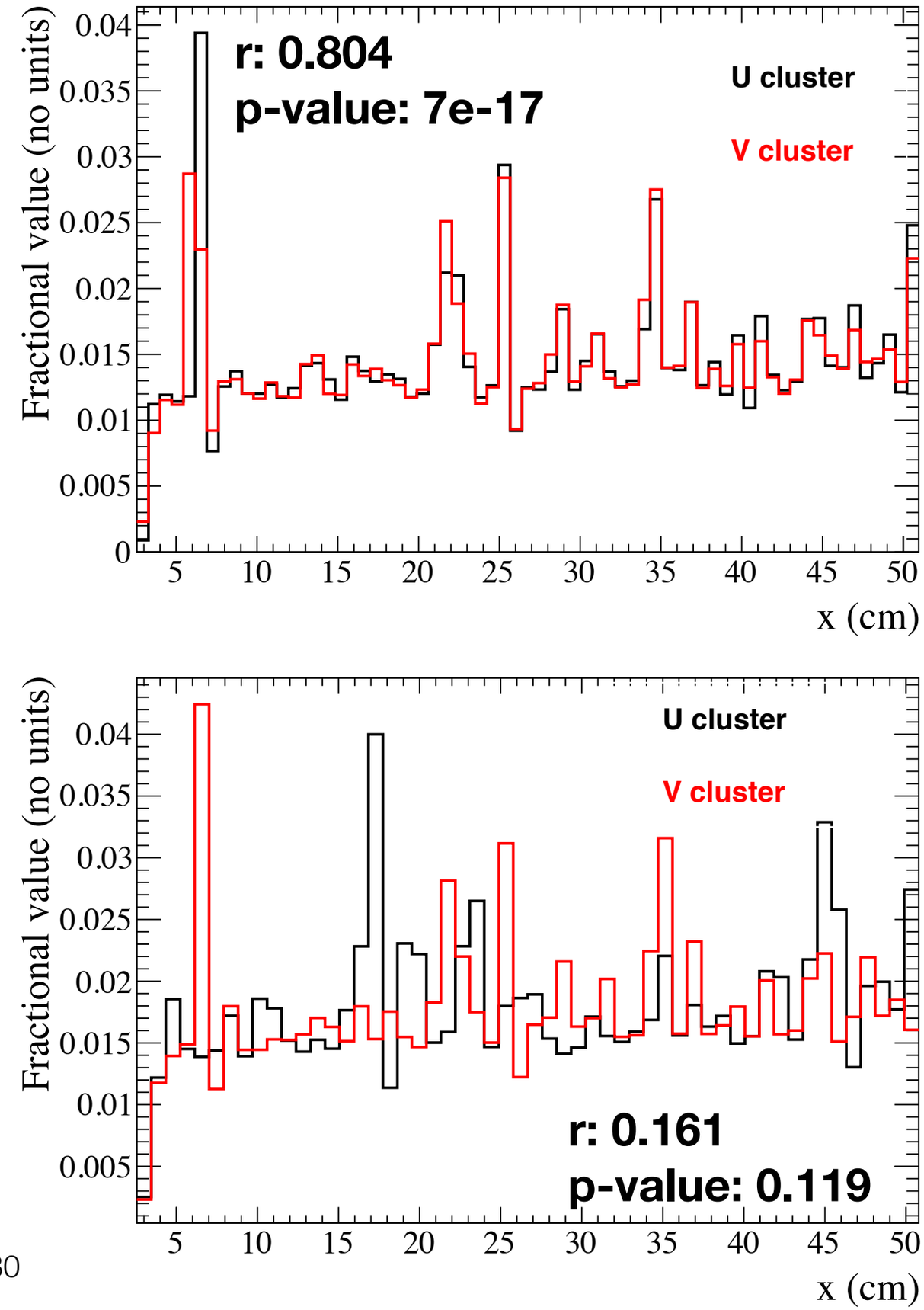
- The t-value is

$$t = r\sqrt{\frac{n-2}{1-r^2}}$$

- P-value is calculated by integrating the t-distribution above the calculated t value (a one tailed test)
  - H0: r==0
  - H1: r>0
- The t-distribution supposedly approximately holds for non-gaussian variables, provided the sample sizes are large enough. I'll revisit this in a few slides

Resampled fractional charge profiles (di-muon sample)



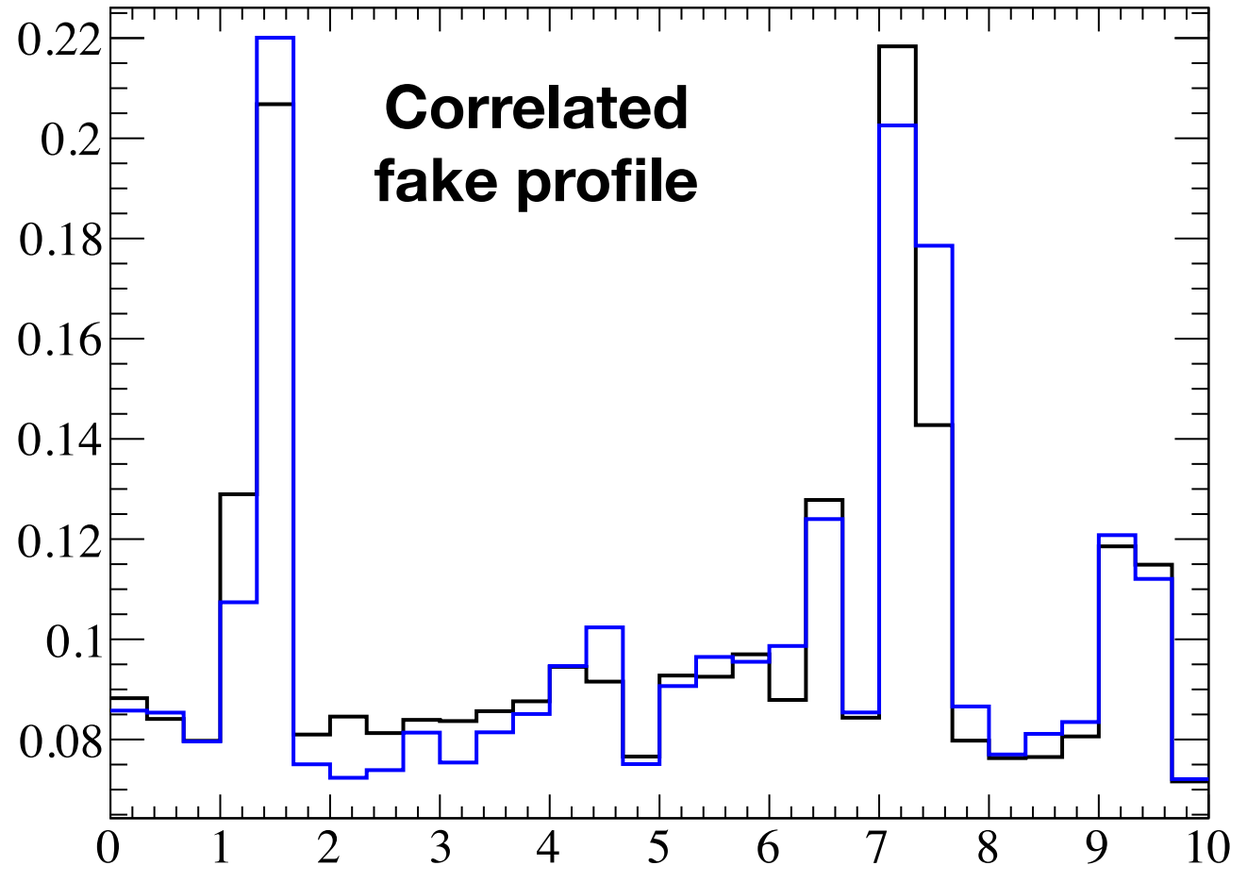


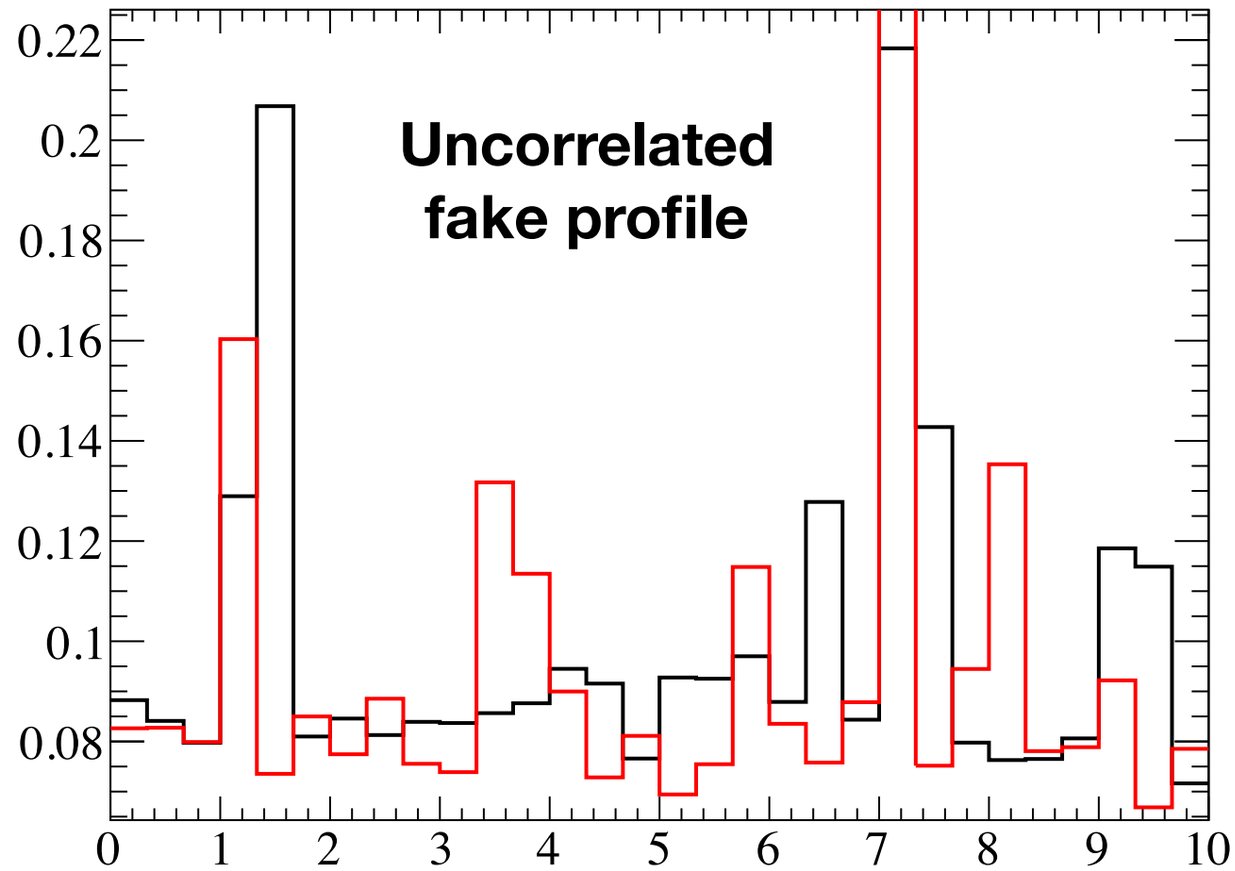
Toy study

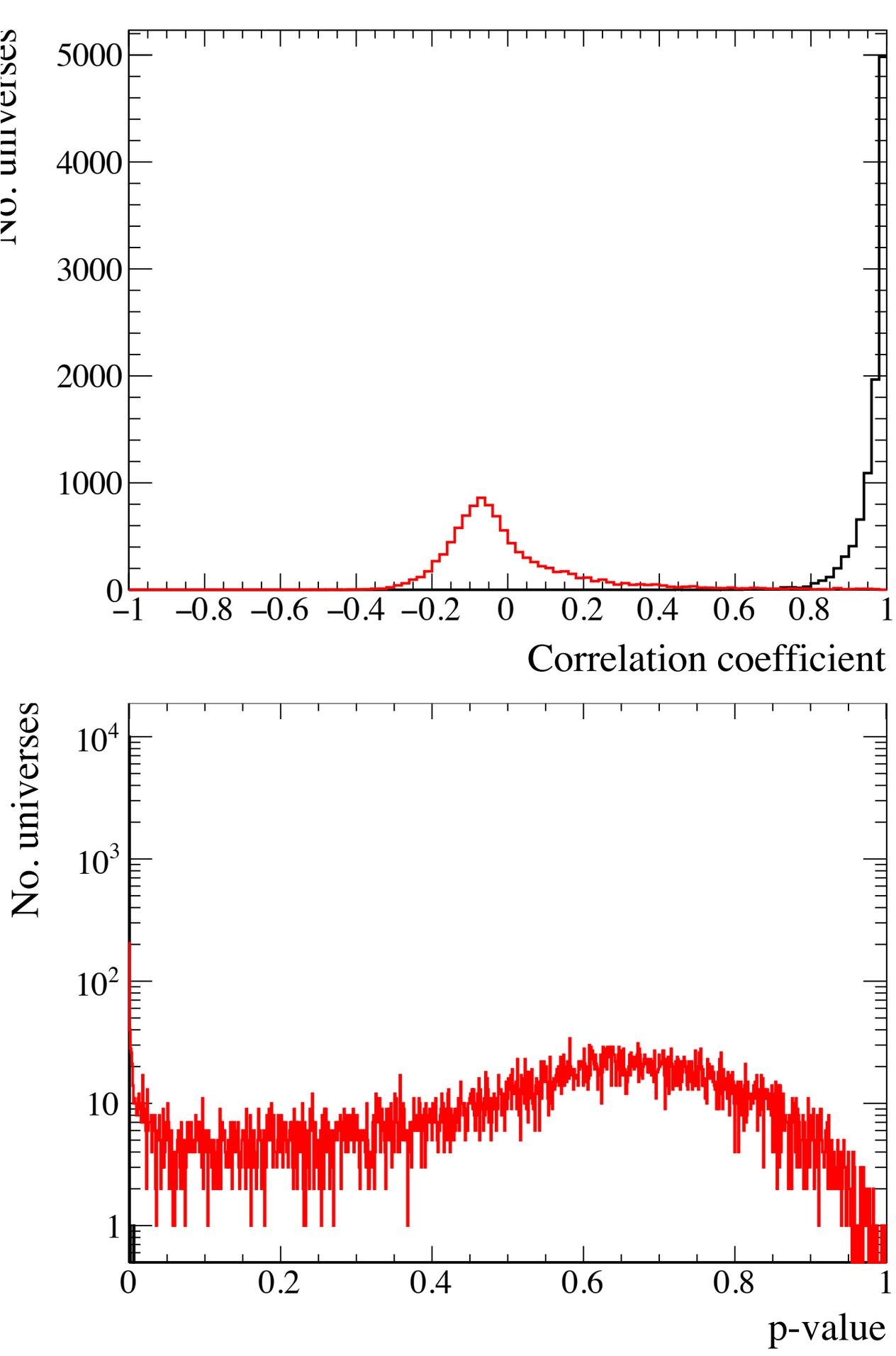
- Revisiting the student t-distribution assumption

- Produce 10000 fake fractional charge profiles
  - Fill 3 histograms with landau throws, smeared with a gaussian
  - Two hists. are filled with the same landau values but smeared separately
  - Third hist filled with separate landau values
  - Each bin is filled N times with distinct throws to mimic the downsampling
- Calculate correlation coefficient and p-value

- Landau (315, 13)
- Gaus (1,0.1)
- N hist bins == 30
- N samples per bin == 5



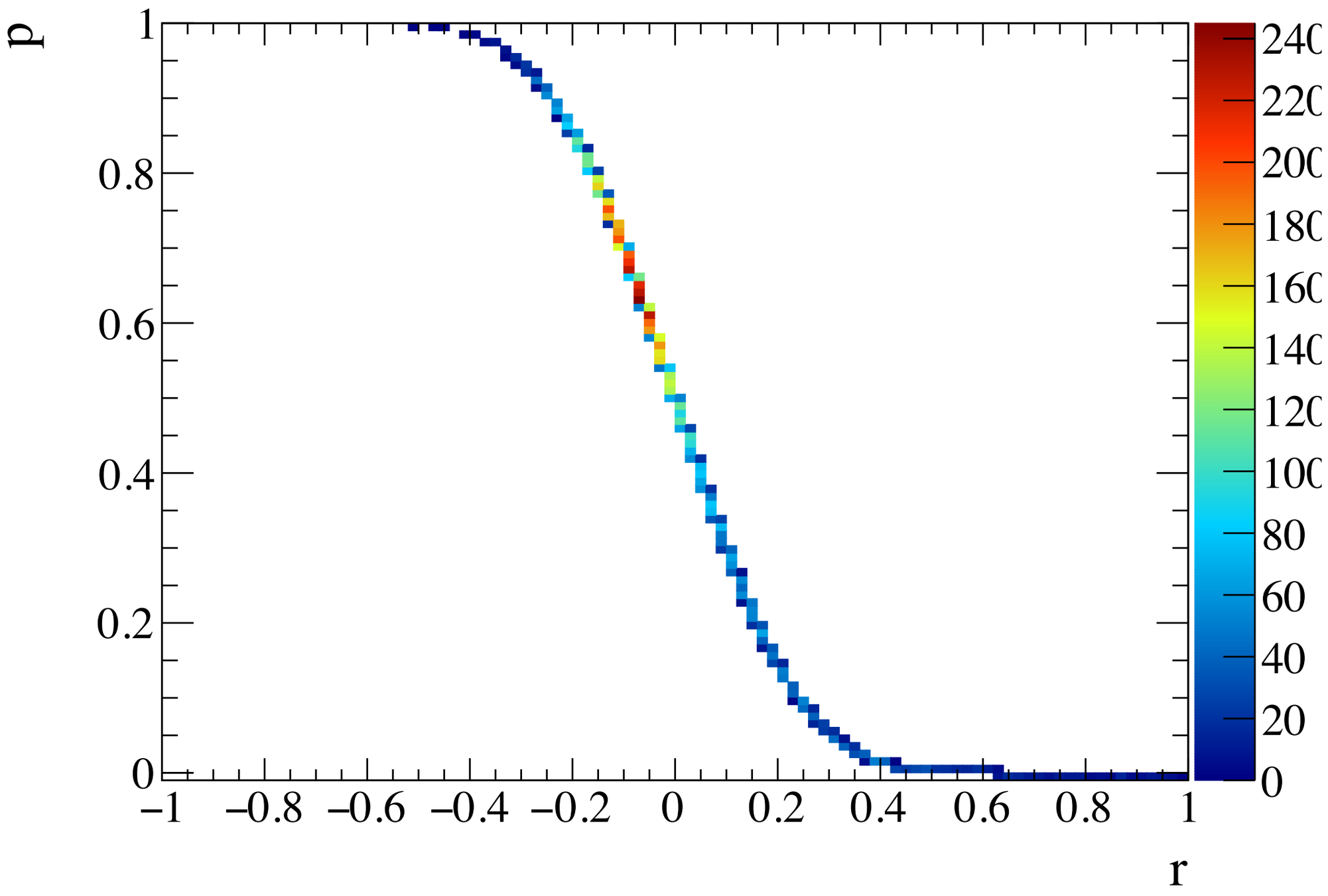


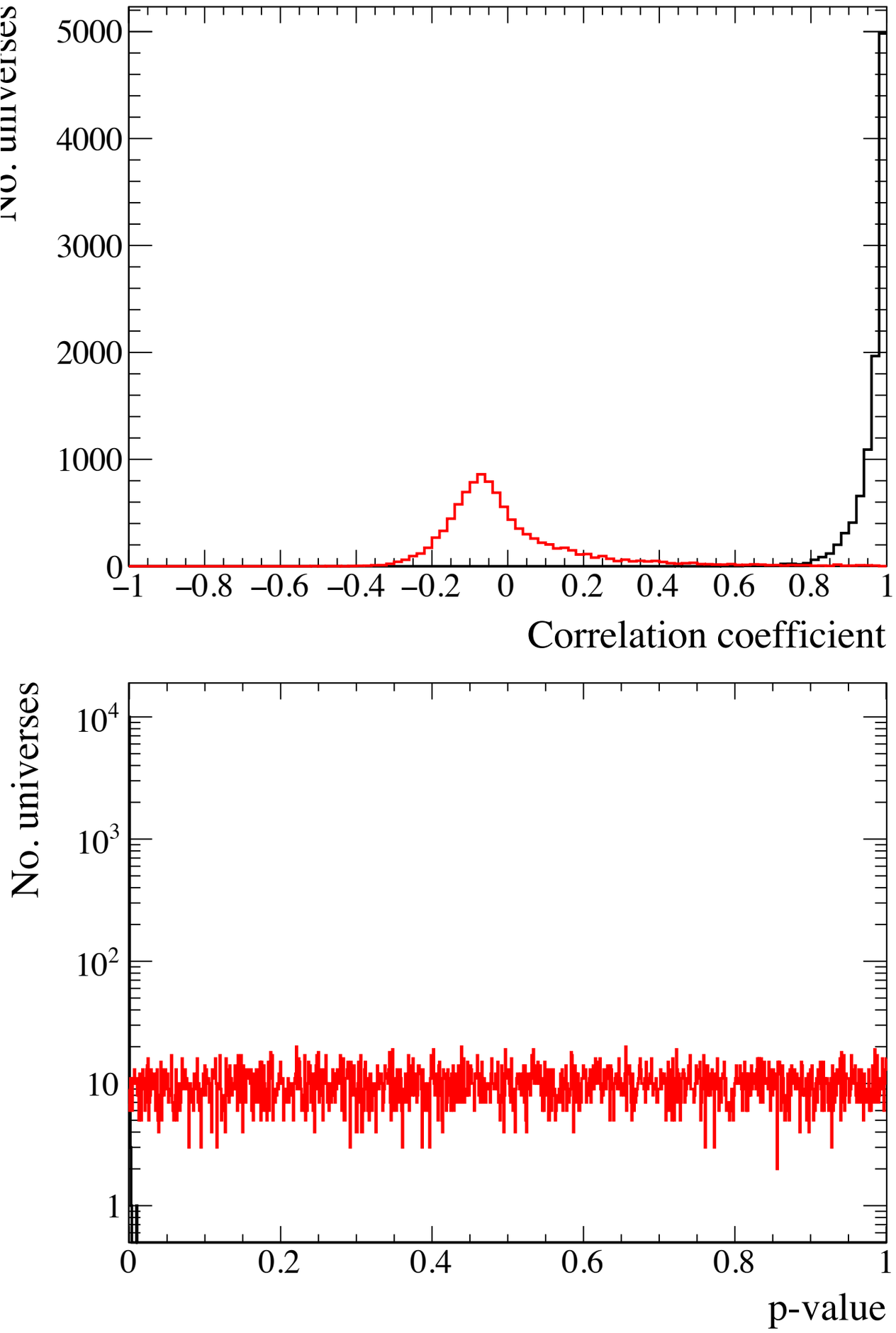


Toy study

- Top plot shows correlation coefficient for the 10,000 universes
  - Black: correlated distributions
  - Red: uncorrelated distributions
- Bottom plot shows corresponding p-values
  - The red distribution should be flat, but it is not

P-value vs r (t-distribution)





Toy study

- Instead, calculate the p-value using permutation tests
  - Randomly shuffle the bins for one distribution in a comparison and recalculate r
  - P-value == fraction of times you measure an r that is more extreme than your original r measurement
- Top plot shows correlation coefficient (same as previous slide)
- Bottom plot shows corresponding p-value

p-value vs r (permutation test)