Proton/Deuteron Separation at the LArIAT Experiment

by *Mohammed Sultan* on behalf of the LArIAT collaboration

Supervisors: Roxanne Guenette, Elena Gramellini

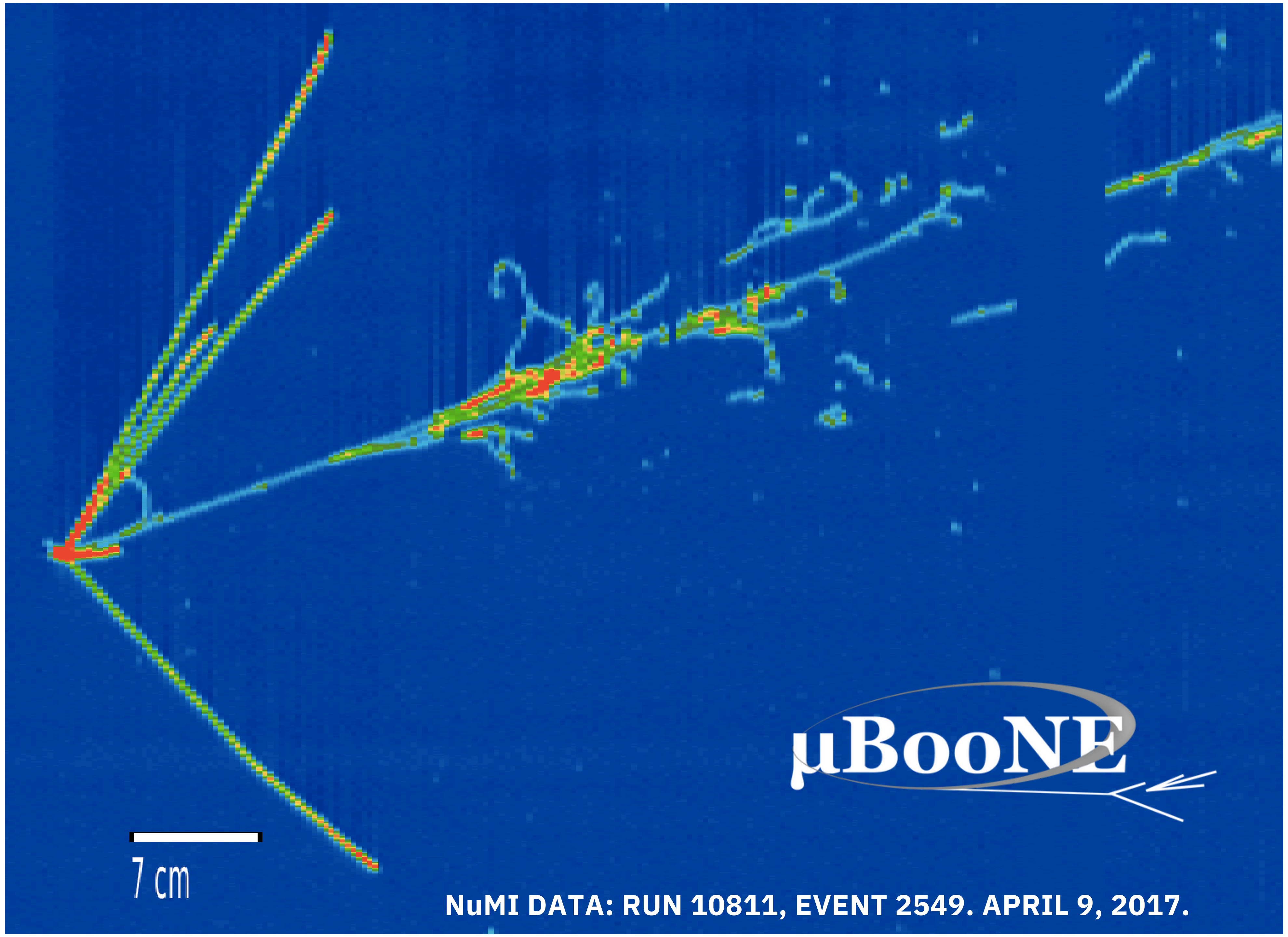
30/10/2025 - Neutrino Physics and Machine Learning (NPML) 2025

Protons and Deuterons in LArTPCs

Misidentifying deuterons as protons biases neutrino energy reconstruction. Despite their mass difference, short-range proton and deuteron tracks **exhibit similar topology and dE/dx profiles,** complicating separation.

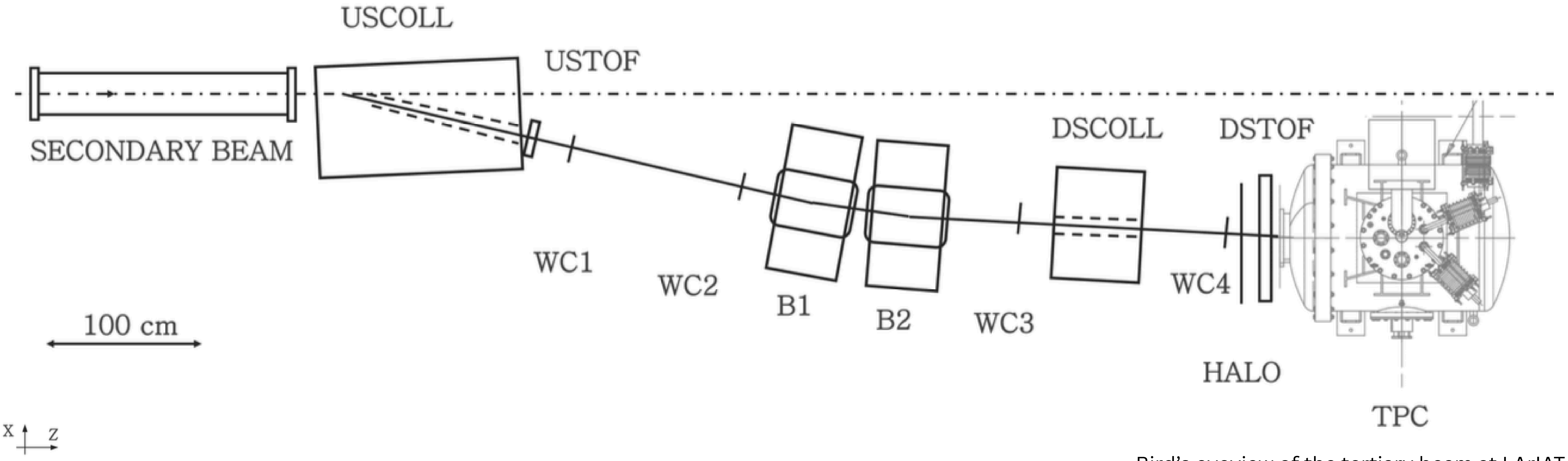
This study aimed to **search for deuterons in LArIAT data** and develop **a tagging method to reliably distinguish them from protons**.

Example LArTPC neutrino event from MicroBoone. Separating protons and deuterons is extremely difficult in these uncontrolled environments.



The LArIAT (Liquid Ar in a Testbeam) Experiment

LArIAT is a liquid argon test-beam experiment at Fermilab (2015-2017) designed **to study how charged particles: like protons, pions, kaons, and muons, interact in liquid argon**, using a precisely controlled beam where the **particle type and momentum are known**.

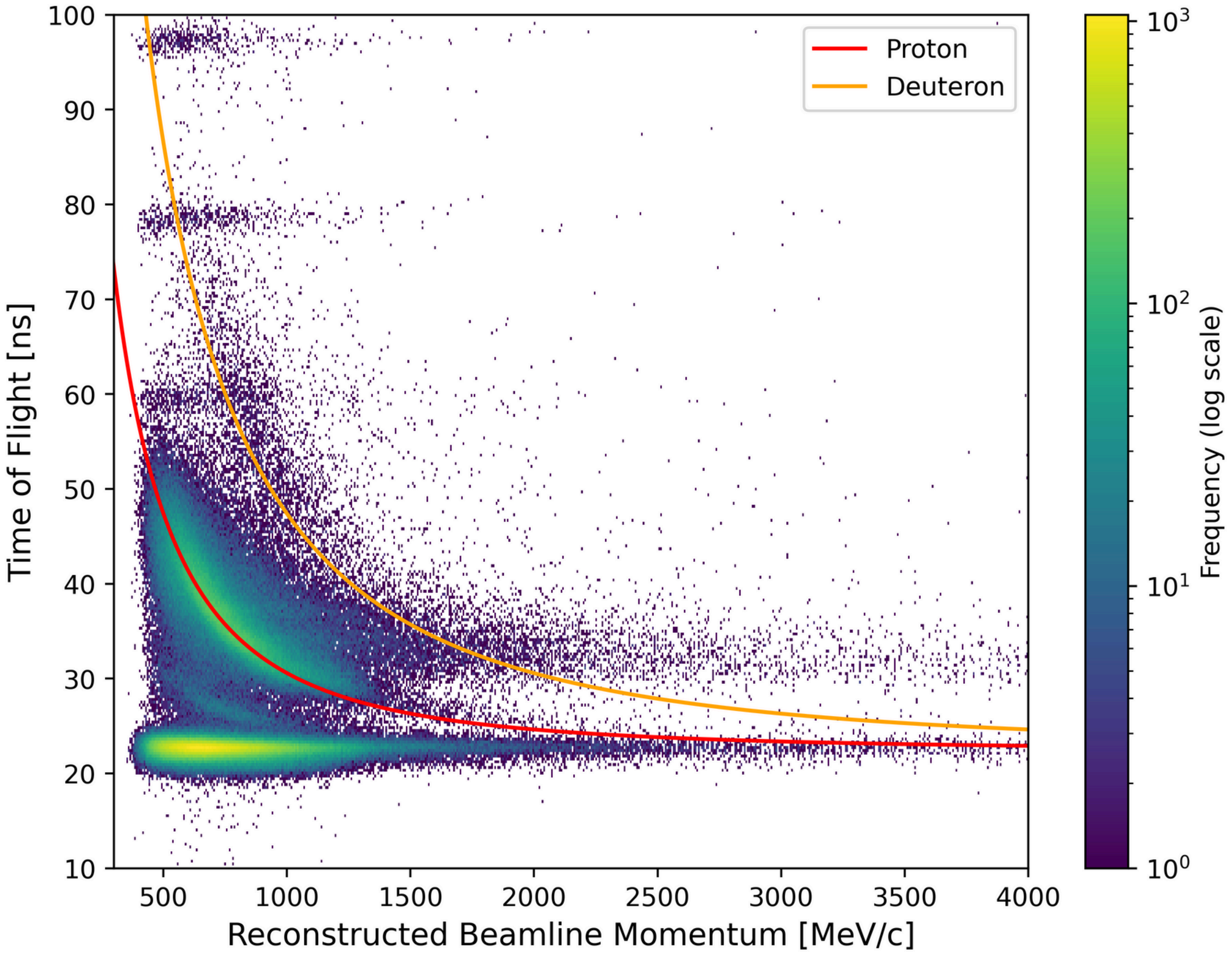


Bird's eyeview of the tertiary beam at LArIAT.
TPC dimensions are: 47 cm × 40 cm × 90 cm.

The LArIAT Experiment

Because the beam composition includes a range of hadrons, including **protons and a small but measurable population of deuterons**, LArIAT gives us a rare opportunity to study hadronic PID directly in liquid argon with ground truth labels.

Key point: we know the mass of the particle entering the TPC and we have quality cuts on the events.



Measured time-of-flight versus momentum in the beamline spectrometer system. The coloured lines indicate the expected behaviour for different particle species.

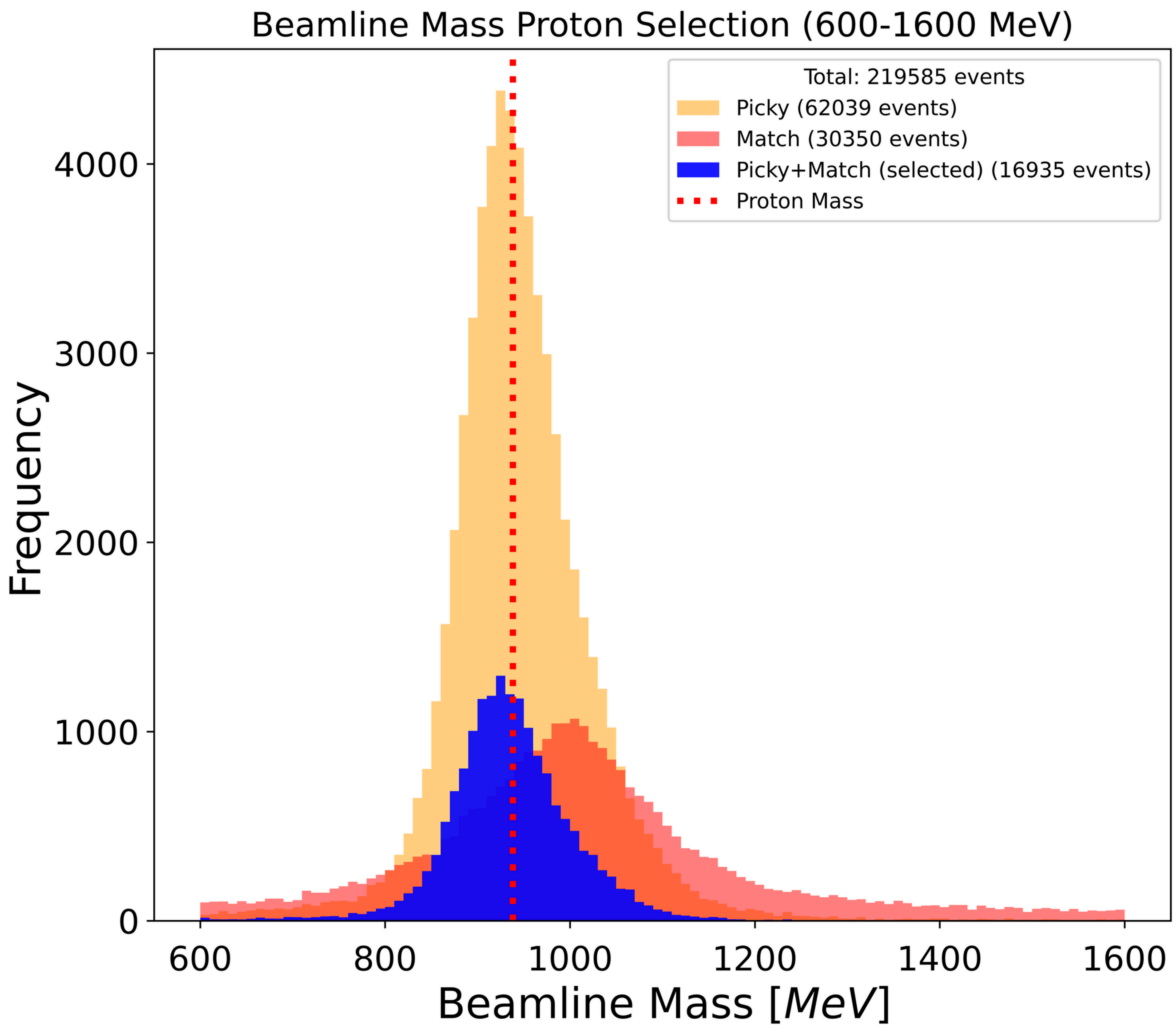
Proton Selection

Proton Mass: 938.27 MeV/c²

Selection: **600-1600 MeV**

Only the events where **both picky & match** were true were selected.

Then used reconstruction, selecting events **with only 1 reconstructed track.**

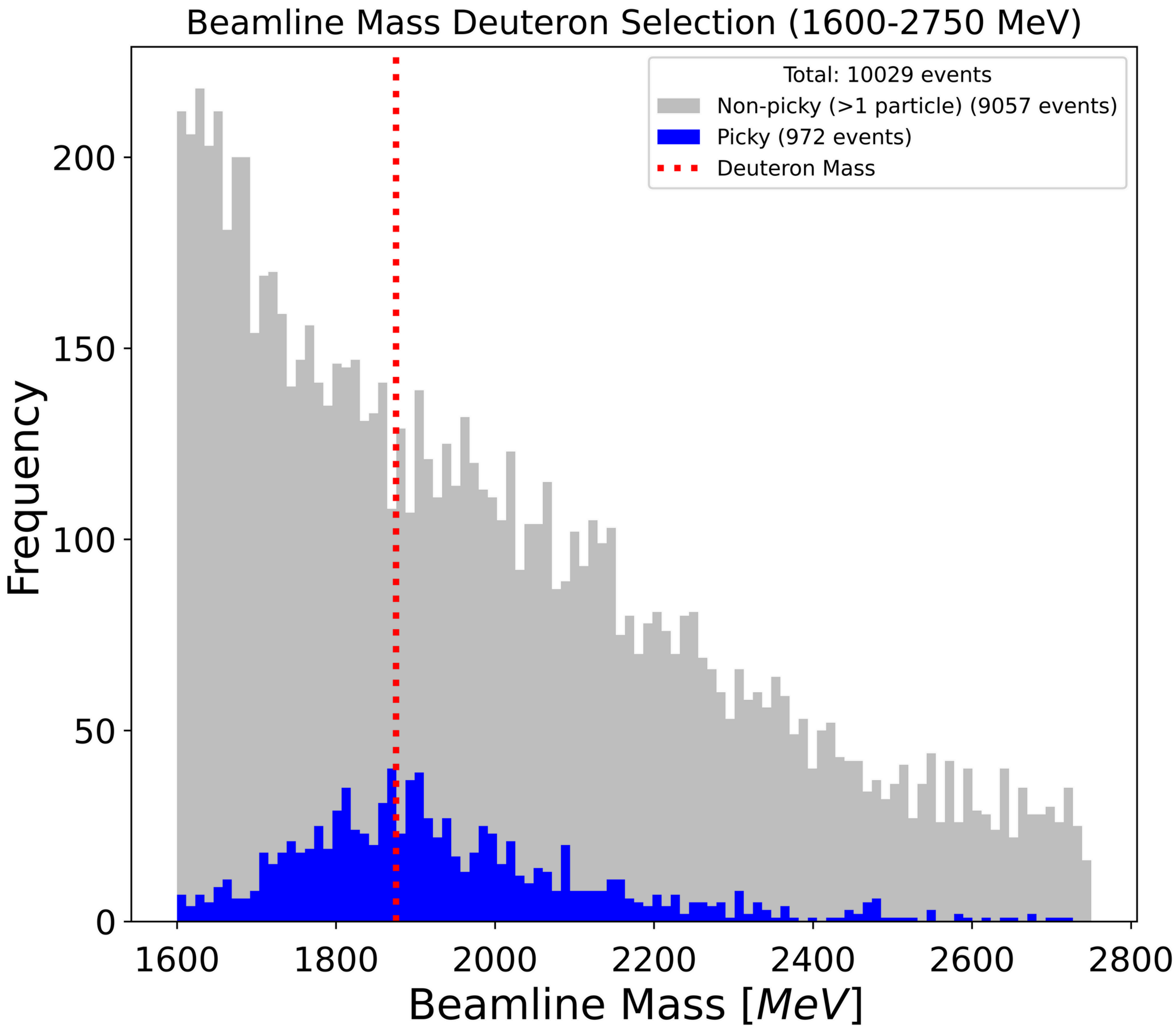


Deuteron Candidates

Deuteron Mass: 1875.6 MeV/c²

All were selected because of low stats.

1600-2750 MeV
10237 events



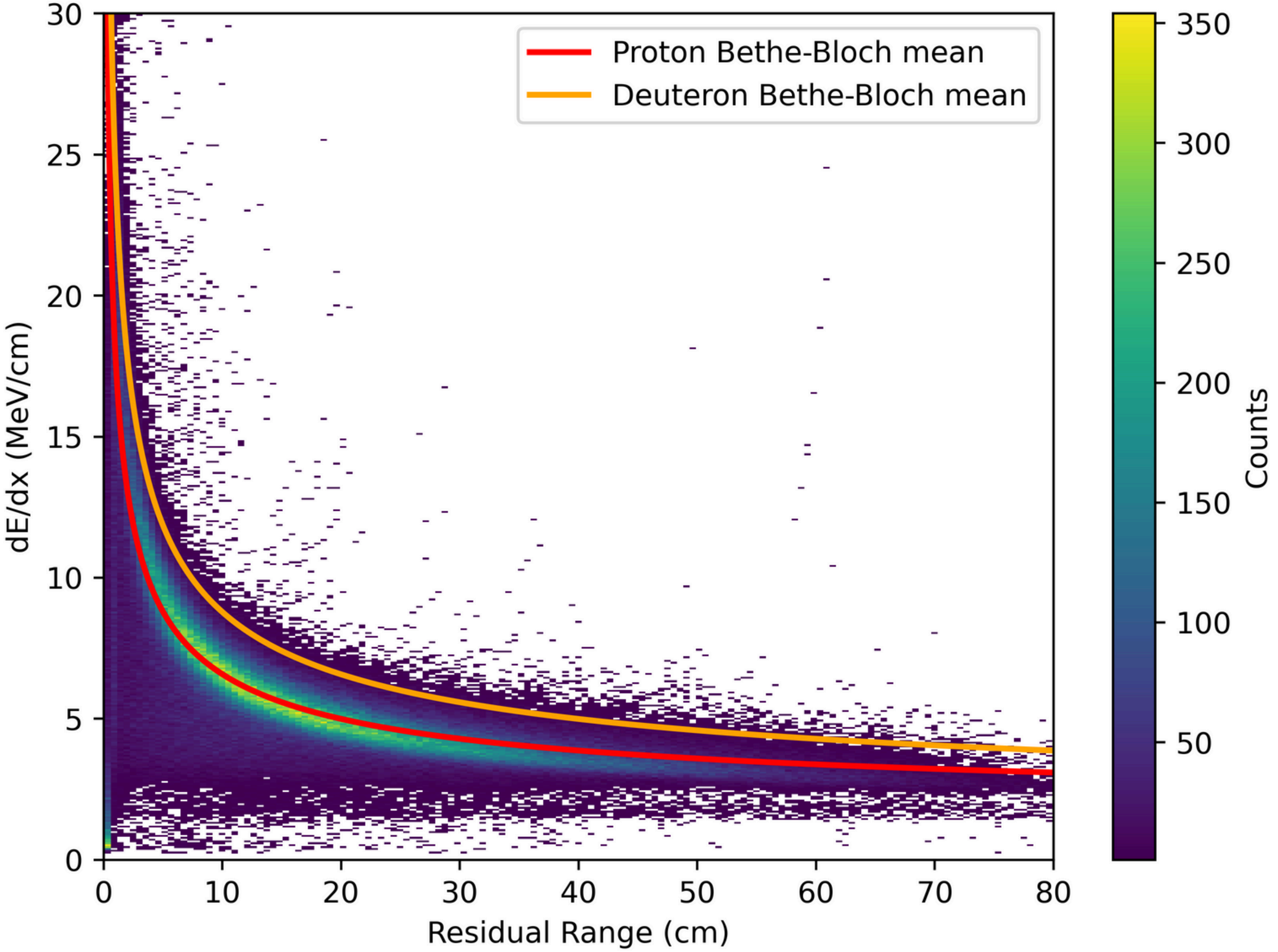
Particle Identification via *dE/dx*

The energy deposited per unit length (*dE/dx*) **depends on the particle type**. *dE/dx* vs. residual range (distance left for a particle to travel before it stops) profiles can **help identify different particles.**

Both protons and deuterons show a rising ionisation profile ending in a **Bragg peak**, but deuterons, being twice as massive as protons with the same charge, **travel more slowly at a given kinetic energy and therefore deposit denser ionisation**.

However, in small ranges, the two profiles start to overlap.

Expected dE/dx vs. residual range profiles for protons and deuterons. The heatmap is shown for actual proton data

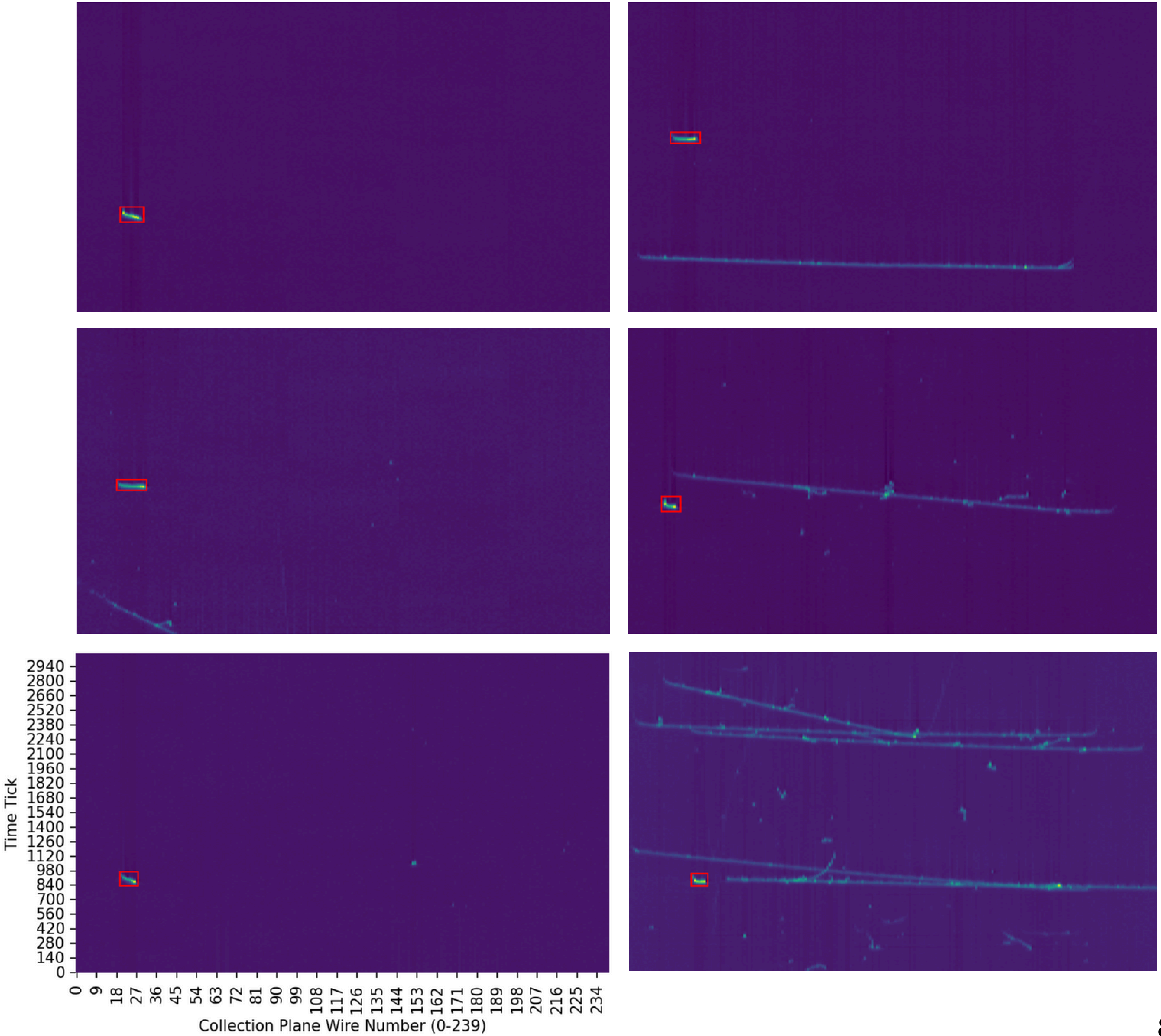


Particle Identification via *dE/dx*

As seen here, the expected **deuteron tracks range between 5-15 wires, being ~ 2 - 7.5 cm long**.

The dE/dx profiles in this range overlap strongly and thus **traditional dE/dx separation in this data was very difficult**.

Especially given the low statistics.

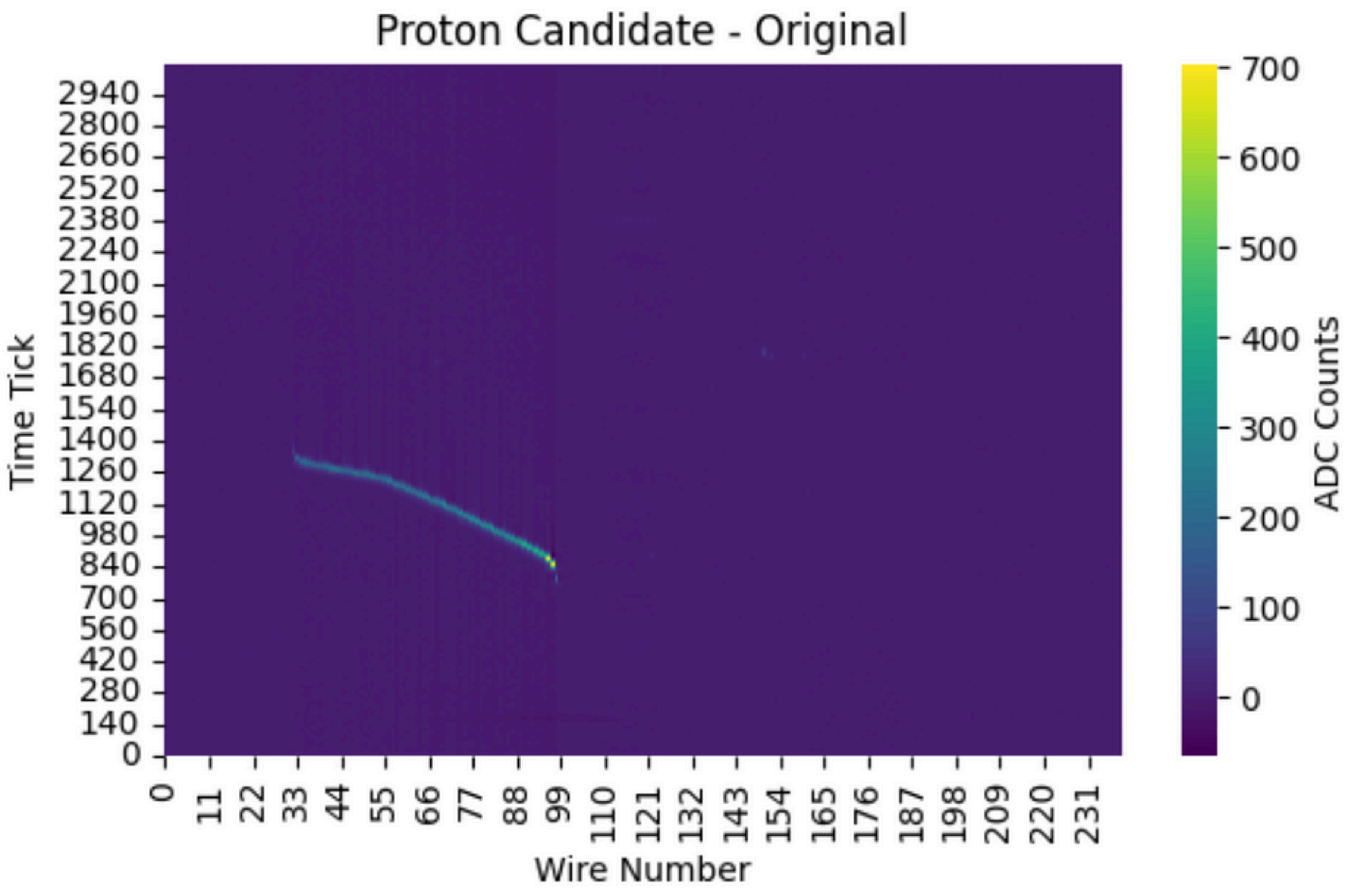


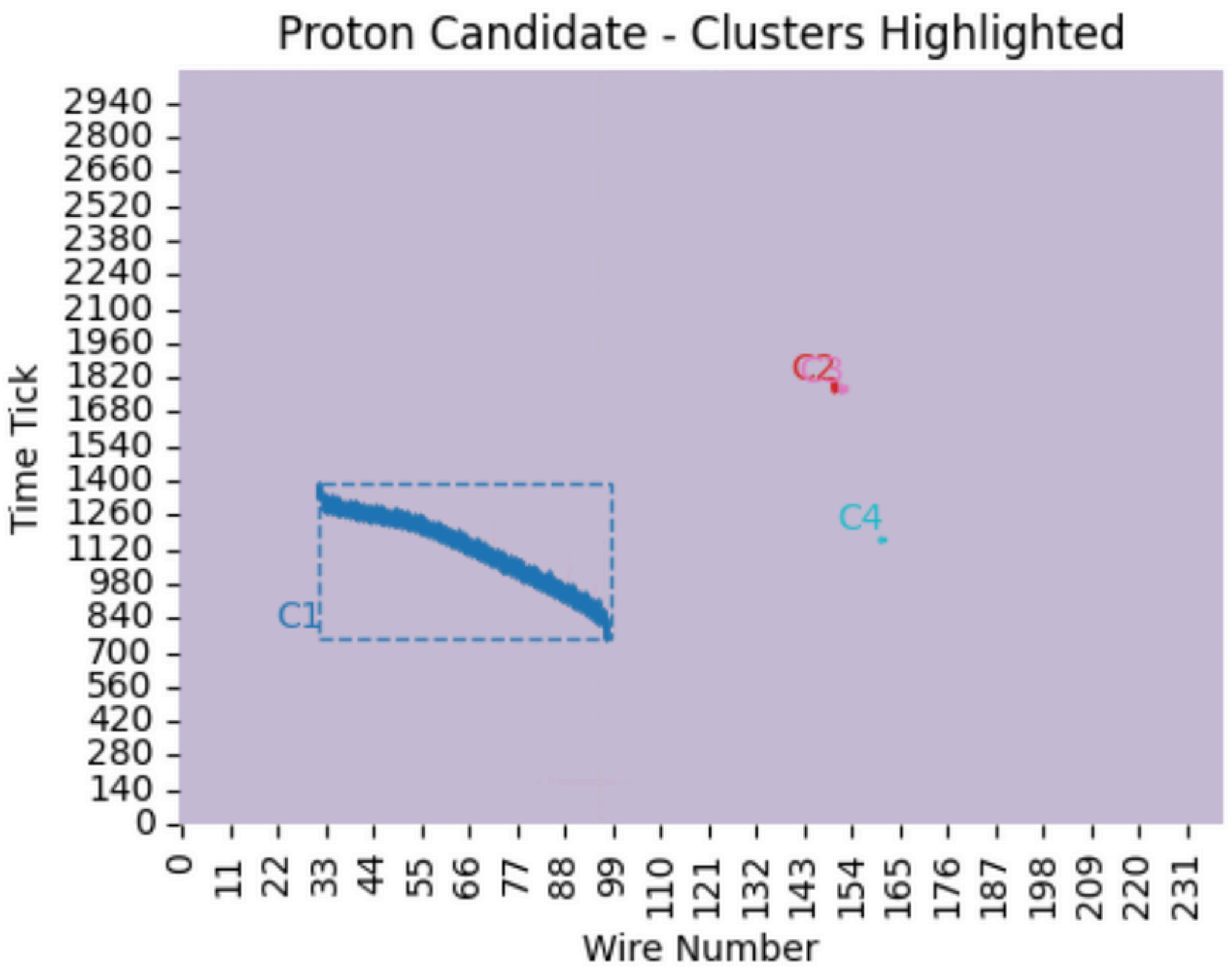
Examples of expected deuterons (visually identified by searching in the mass range, for short, highly ionising tracks)

Clustering Hits

A simple threshold on ADC was used to make a binary mask then cluster together **connected pixels (pixels next to each other in all 8 directions with ADC > threshold)**.

This was applied to all selected events, protons and deuterons. Cuts were made to discard noisy clusters and clusters were matched and paired across the two planes.





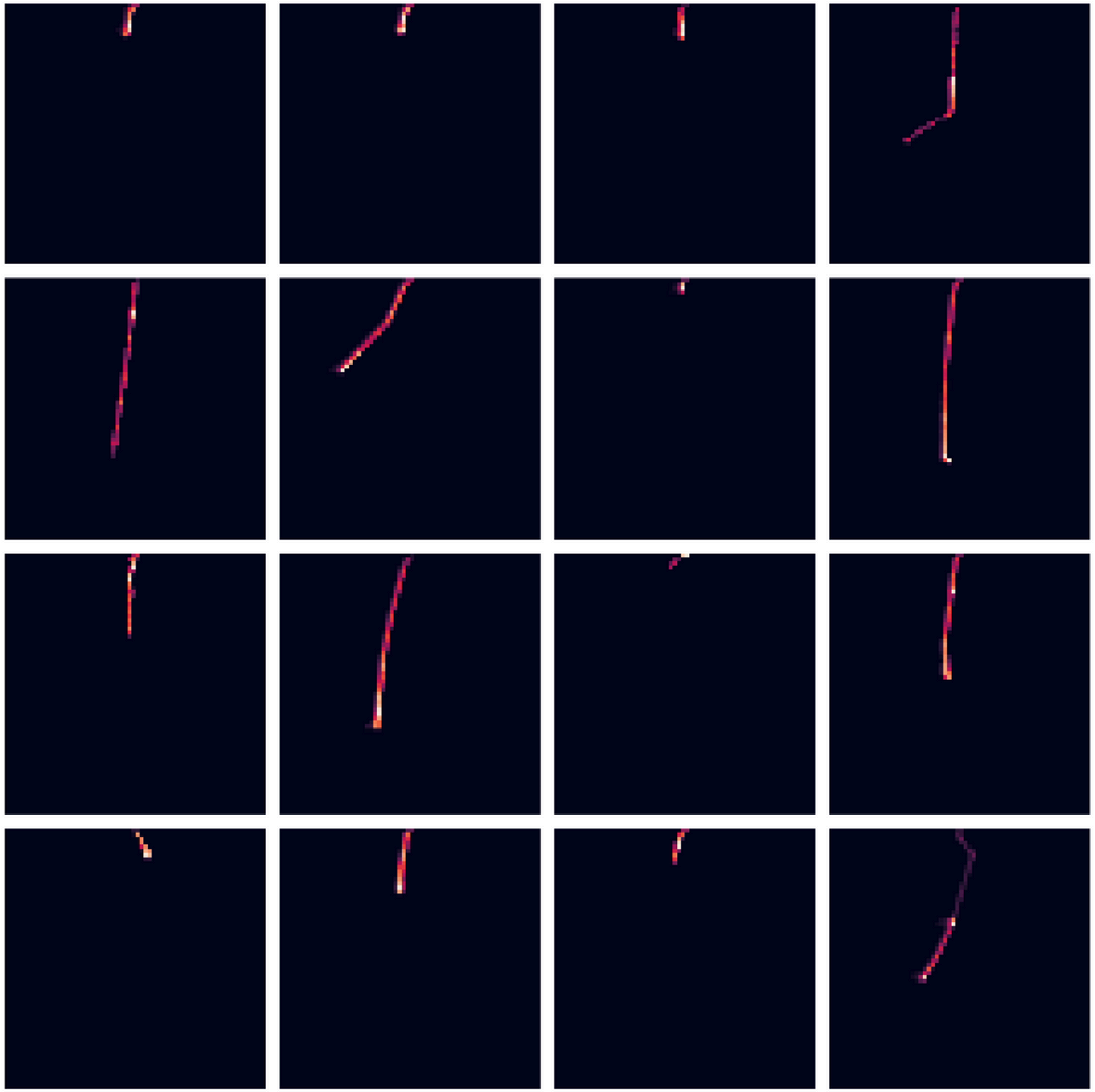
ML-based separation

Can we learn to separate protons/deuterons directly from the images?

We attempted to leverage the high statistics and reliability of the proton selection, and **treated the search for deuterons as an anomaly detection problem.**

An **autoencoder was trained on the true proton cluster images**, and then the clusters from deuteron candidate events were passed through this model, giving latent representations for each.

We expected separation either **through the image reconstruction error or in the latent space**.

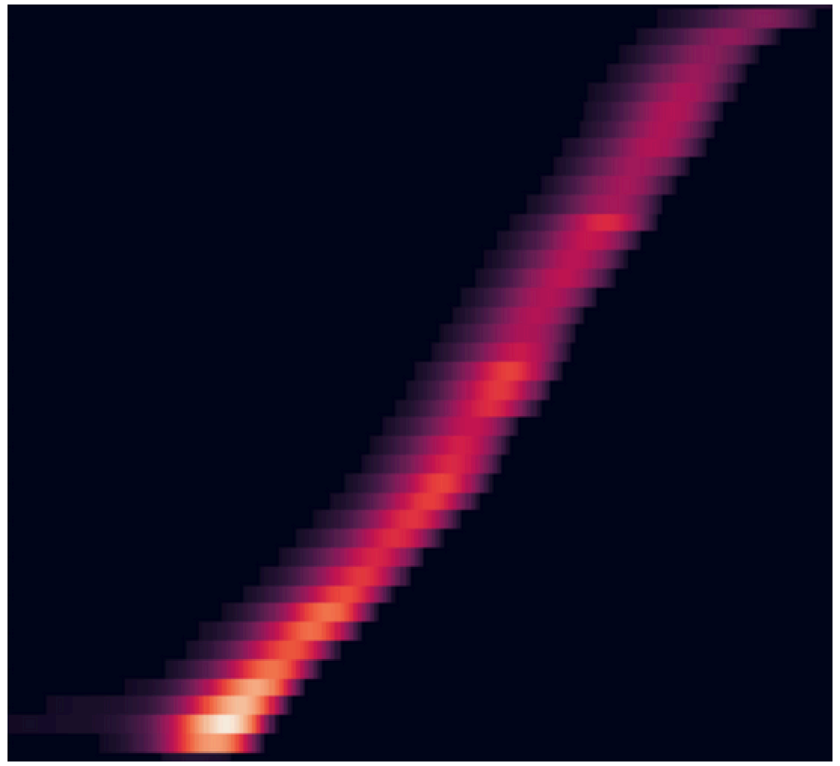


Examples of clusters from deuteron candidate events.

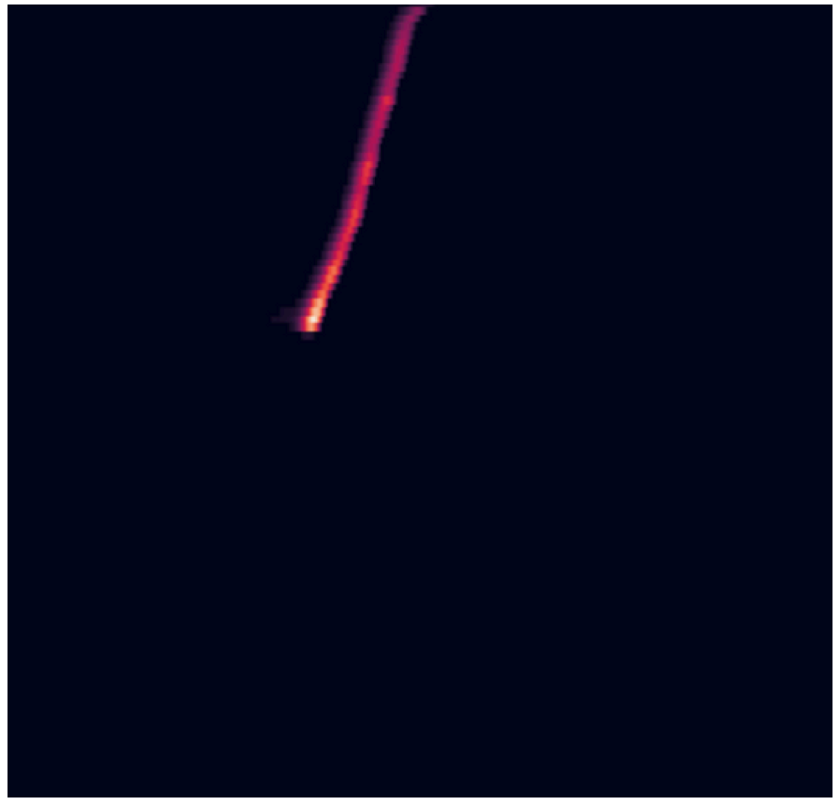
Image construction

Using clusters from each event (after cleaning):

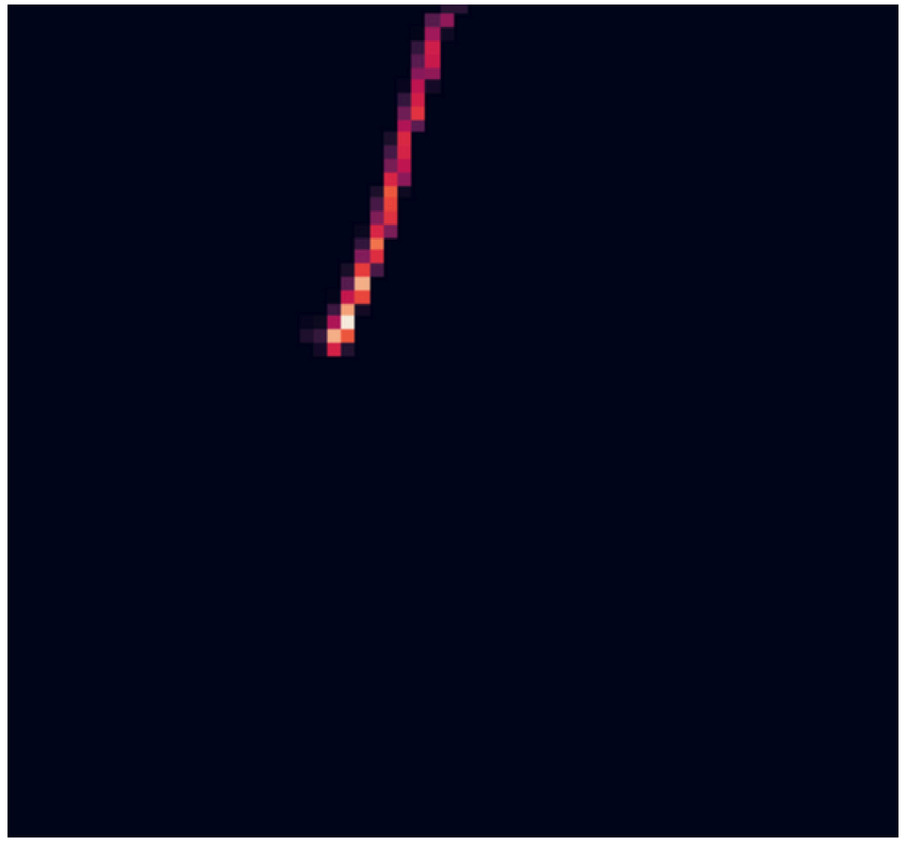
- Centring:
  - Place image on zero canvas so track lies in the centre.
  - Removes translation variability – not relevant for p/d separation.
- Downsampling:
  - Resize to 64×64 using bilinear interpolation.
  - Keeps main topology, reduces dimensionality.
- Final dataset: centred, normalised 64×64 images for training.



Cluster image from event



Padded and centred

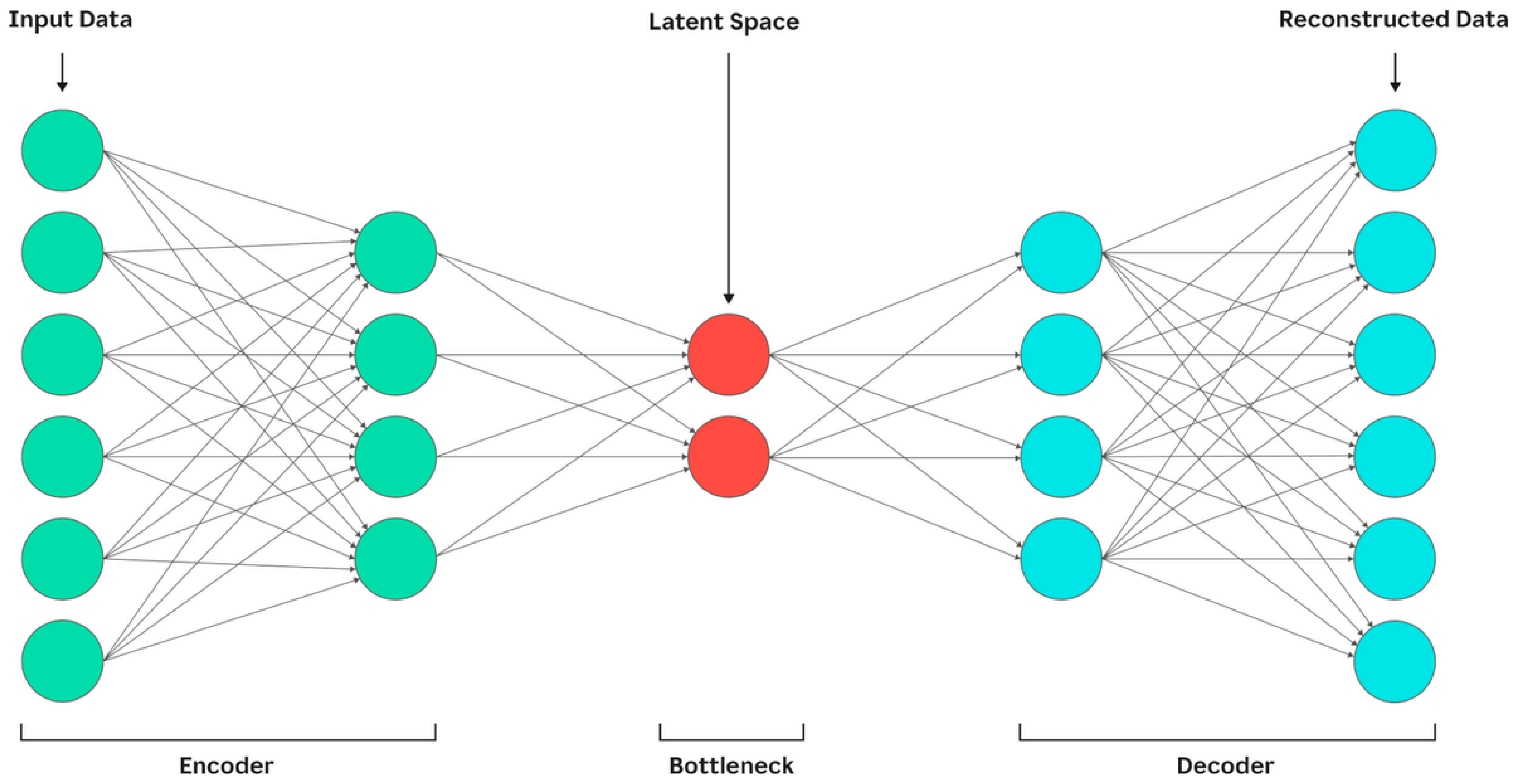


Downsampled to 64 x 64

Convolutional Autoencoder (CAE)

Goal: Learn compact latent representations of proton images → 64×64 (2-channel) charge images.

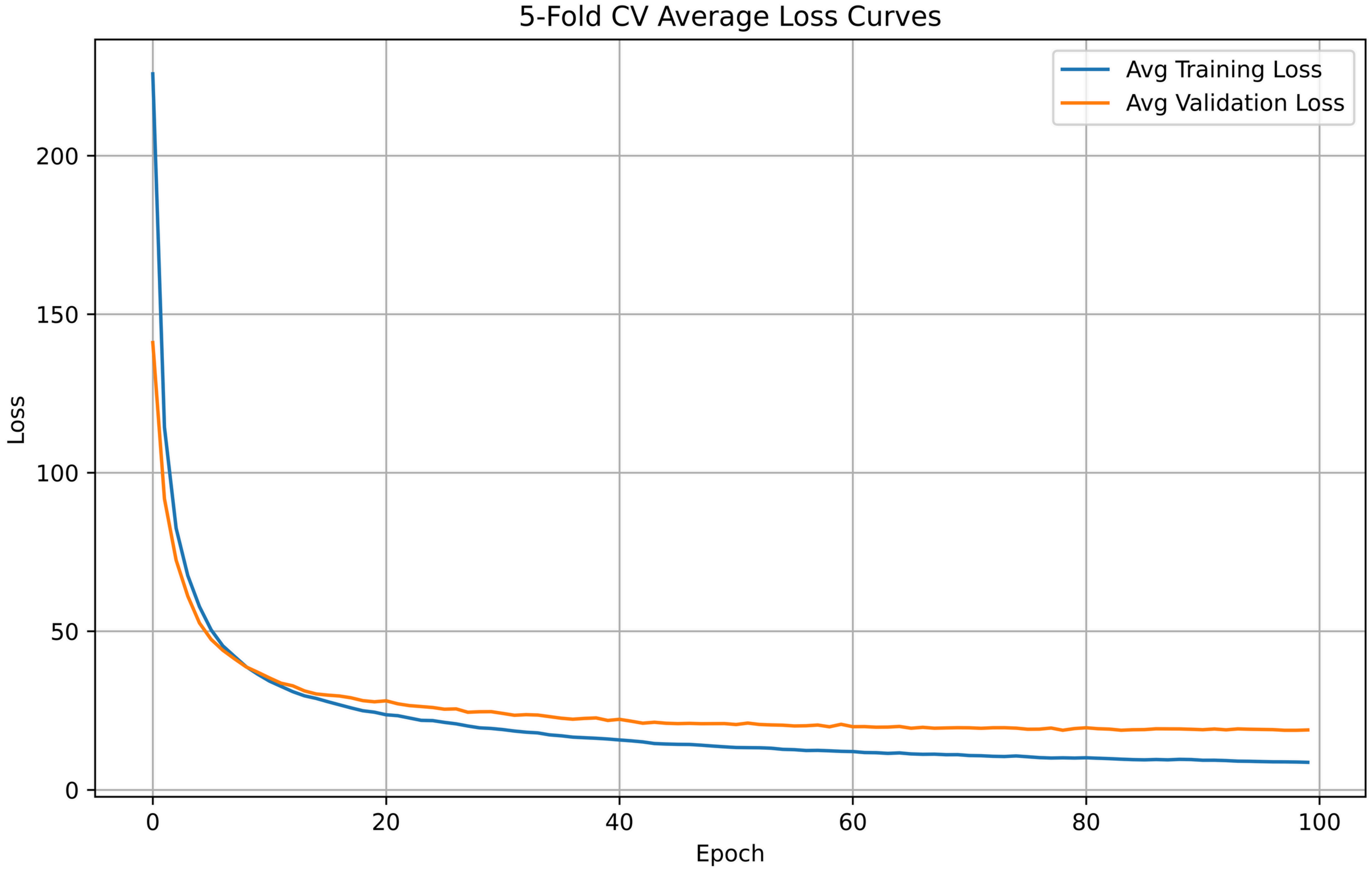
- Encoder:
  - 4 convolutional blocks (stride = 2) → progressively downsample image.
  - Channels: 1 → 32 → 64 → 128 → 256.
  - Output flattened and reduced to latent vector (dim = 16).
- Decoder:
  - Mirrors encoder using transpose convolutions.
  - Reconstructs image back to original size.
- Training objective: minimise reconstruction loss (MSE) between input and output.



Autoencoder architecture schematic.

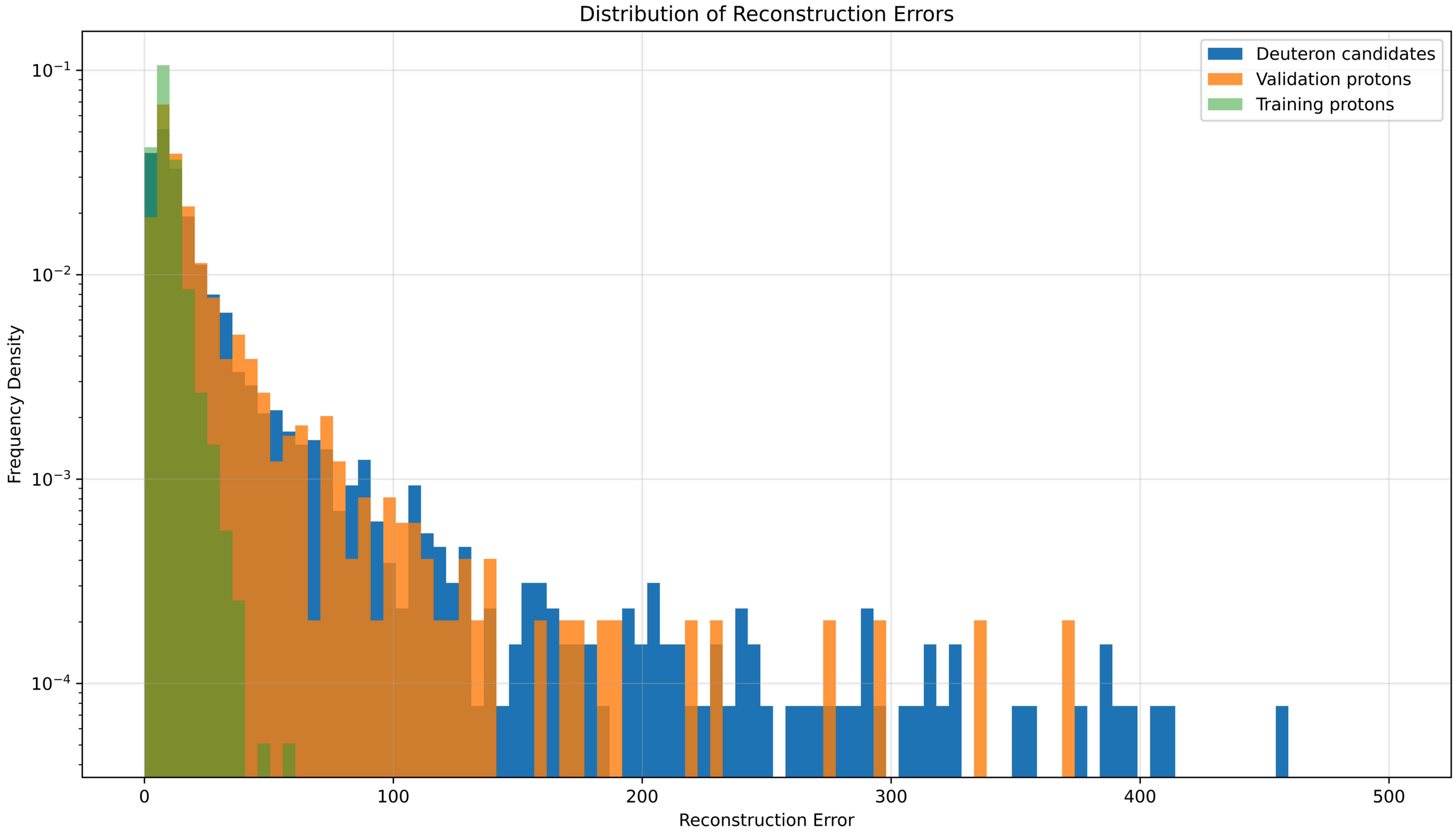
Training

- Training = 5033 images
- Validation = 1258 images
- Both curves decrease smoothly and converge, indicating stable training and no overfitting.
- The small gap between training and validation losses suggests good generalisation, the model captures relevant structure without memorising noise.
- Loss plateaus after ~60 epochs → sufficient training duration.



Training

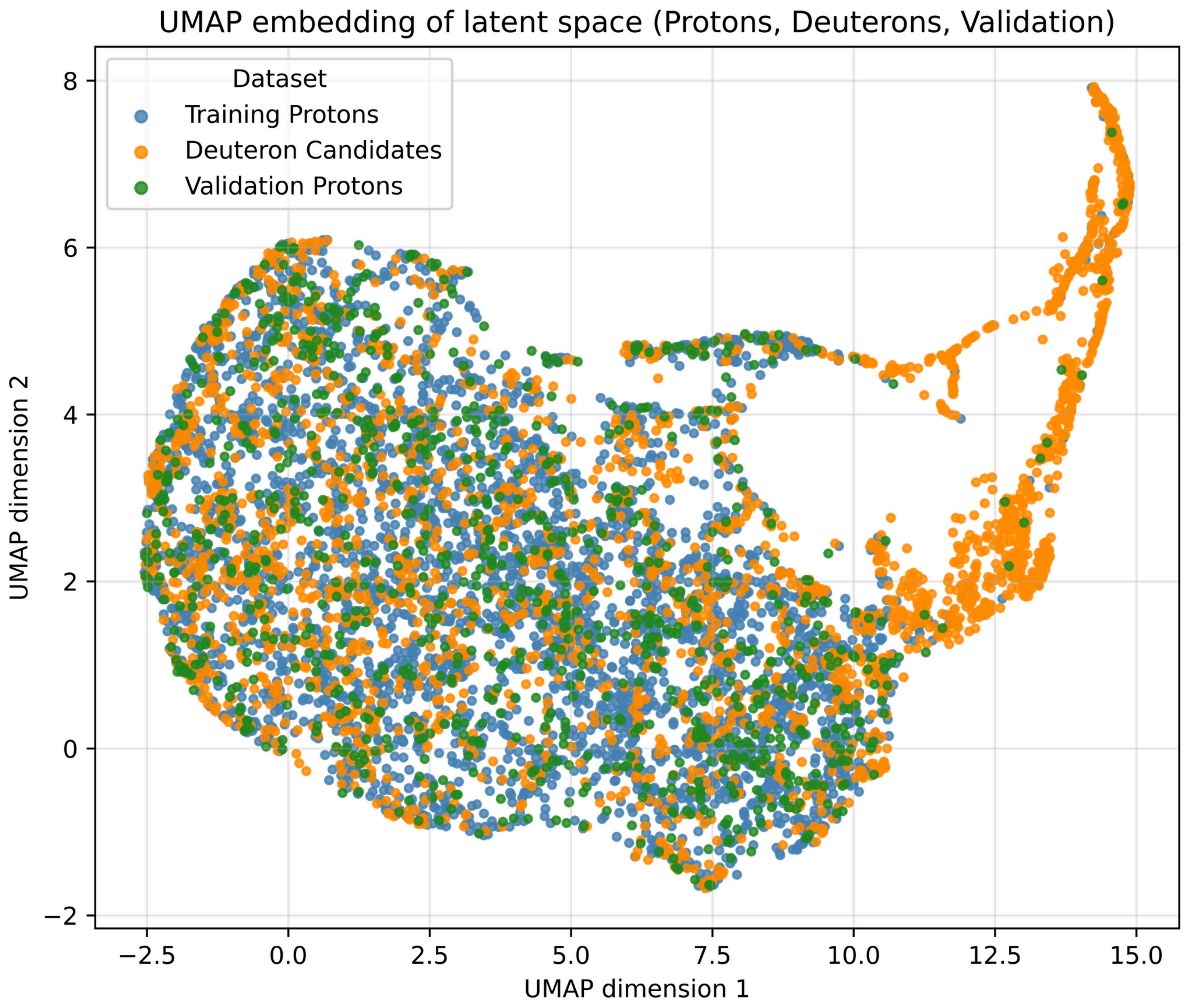
- Model trained only on protons → error acts as an anomaly score.
- Protons (train & val) cluster at low errors, showing good generalisation.
- Deuterons candidates show a broader, higher-error tail, indicating the model recognises them as distinct from protons. (5115 images)

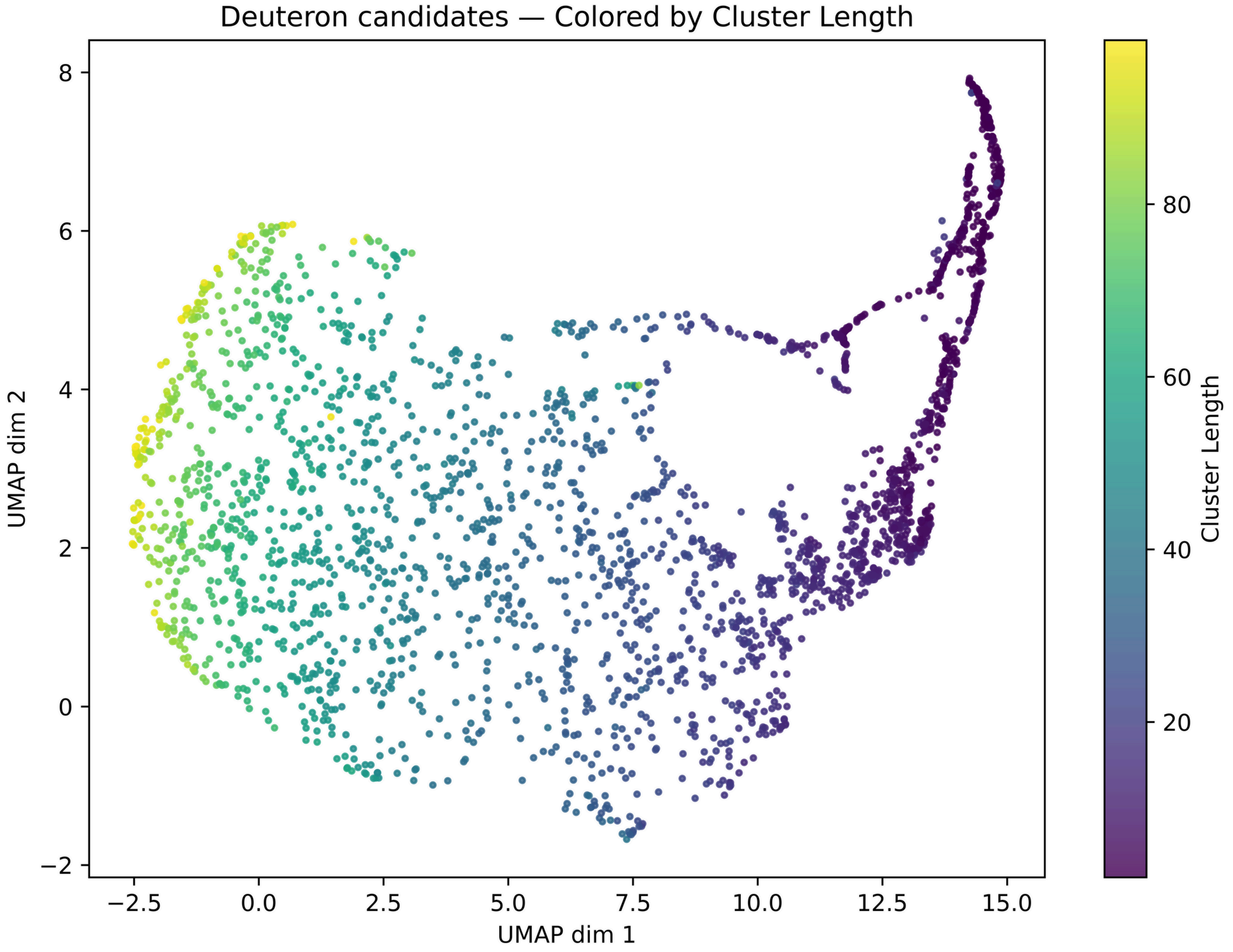


The Latent Space

UMAP helps reduce dimensions for visualisation of the latent space.

- Many tracks in the deuteron candidates can be **seen overlapping the proton regions** (suggesting protons present)
- **Many cluster separating away from the proton region** (suggesting non-protons)
- These could be noise or deuterons.



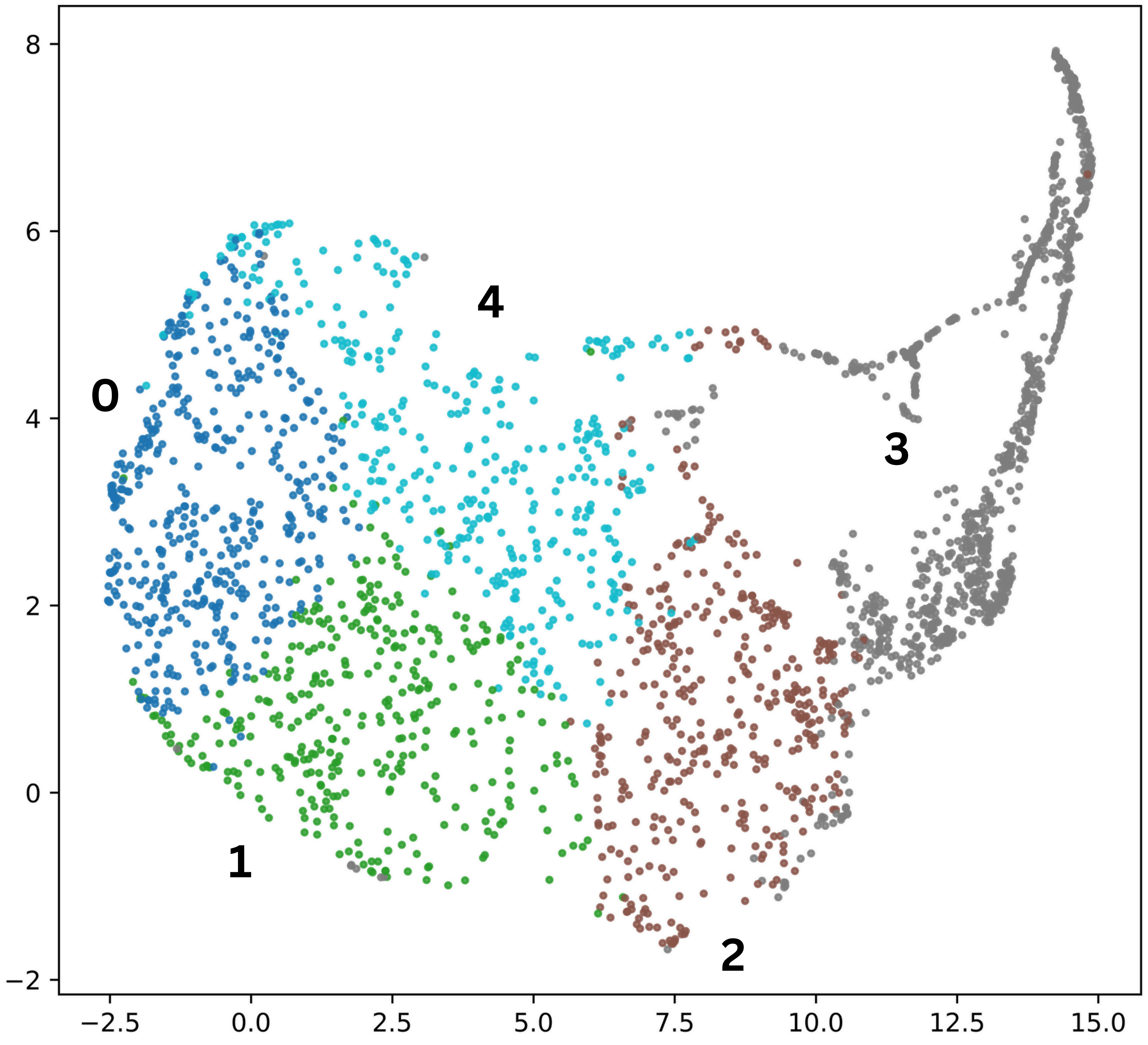


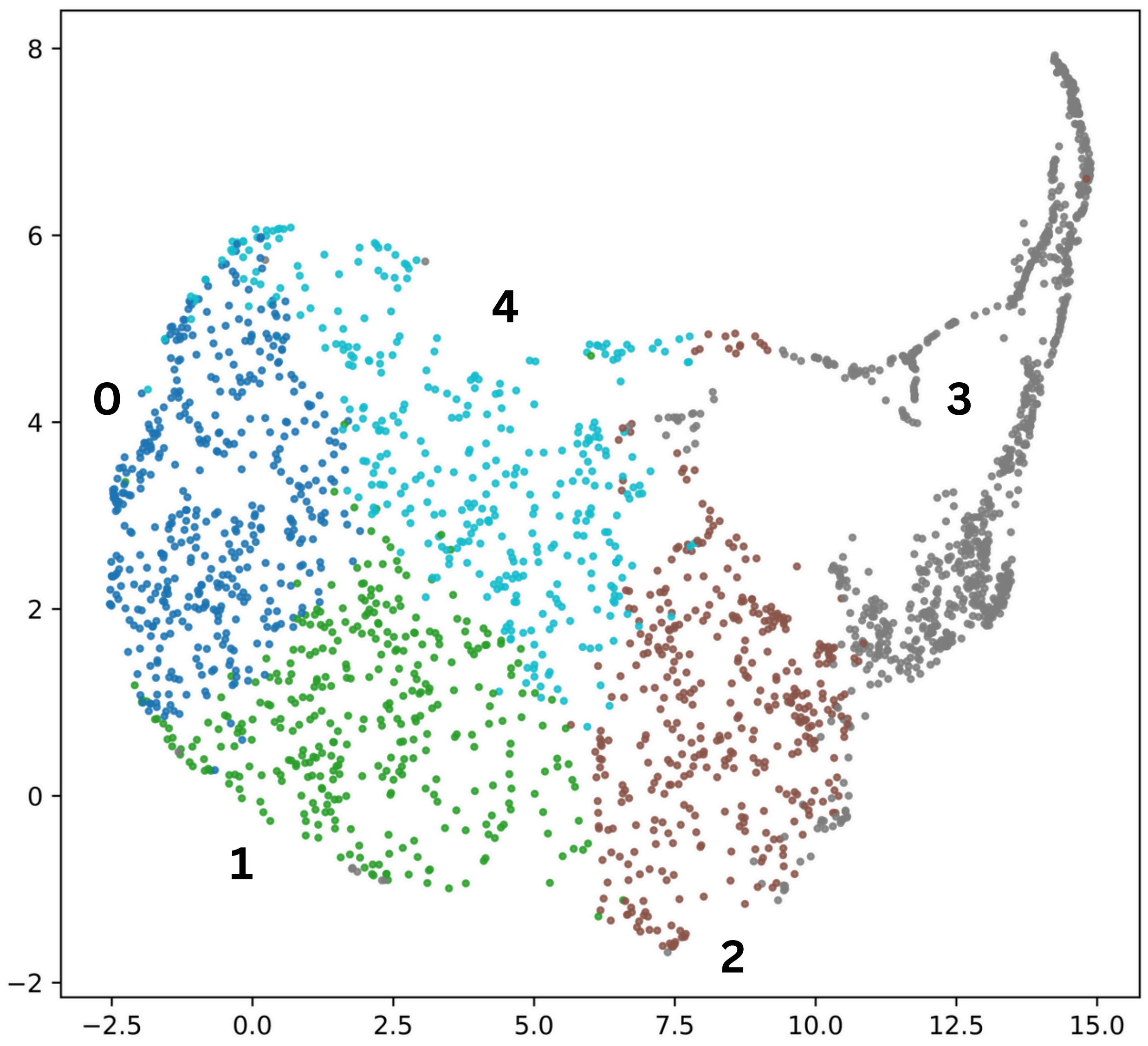
The latent space seems to be encoding the length of the unseen particles.

Running K-Means Clustering

In order to probe the different regions of the latent space we can look at cluster displays in different regions.

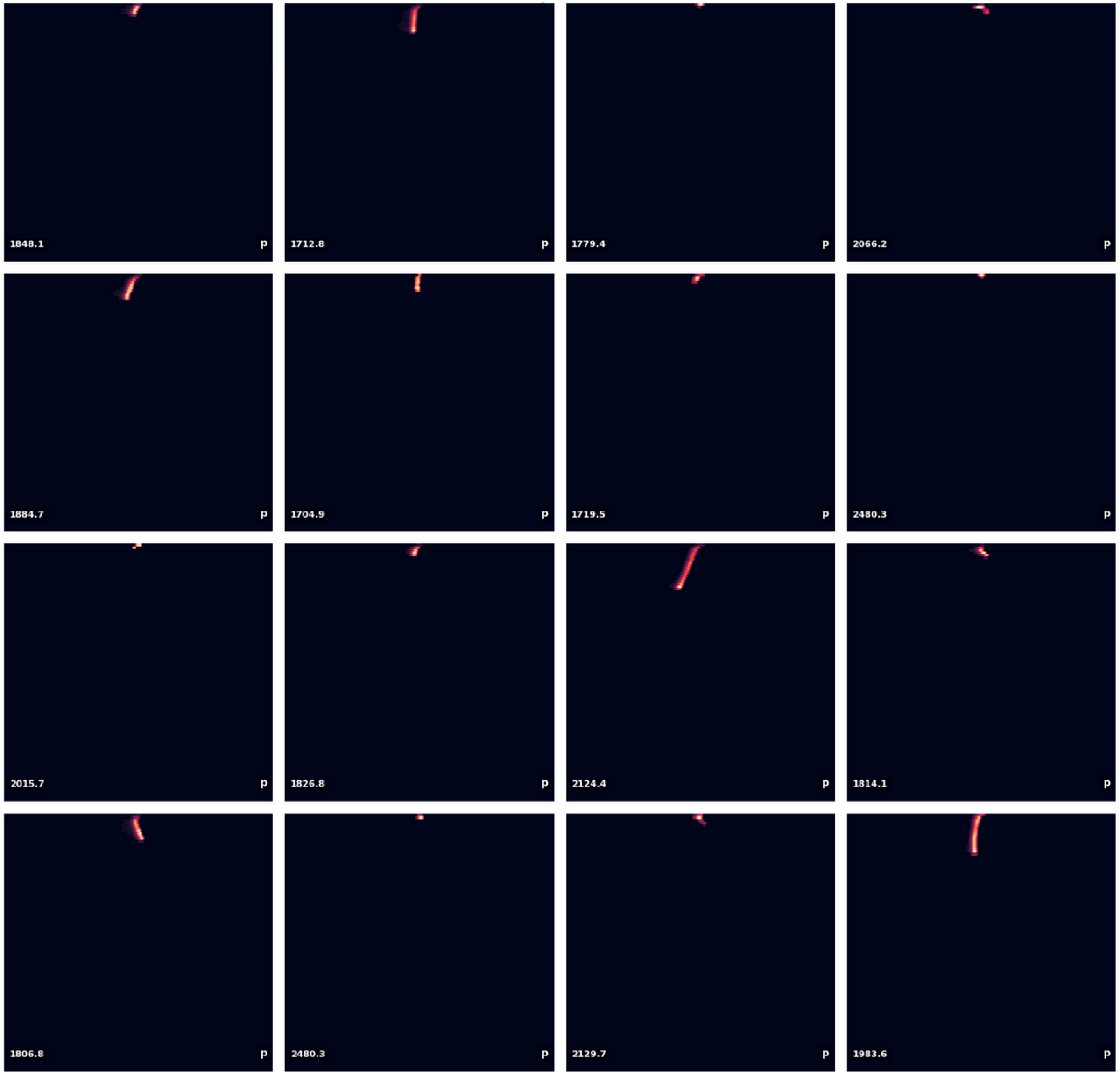
Used k = 5 on the data in 16D

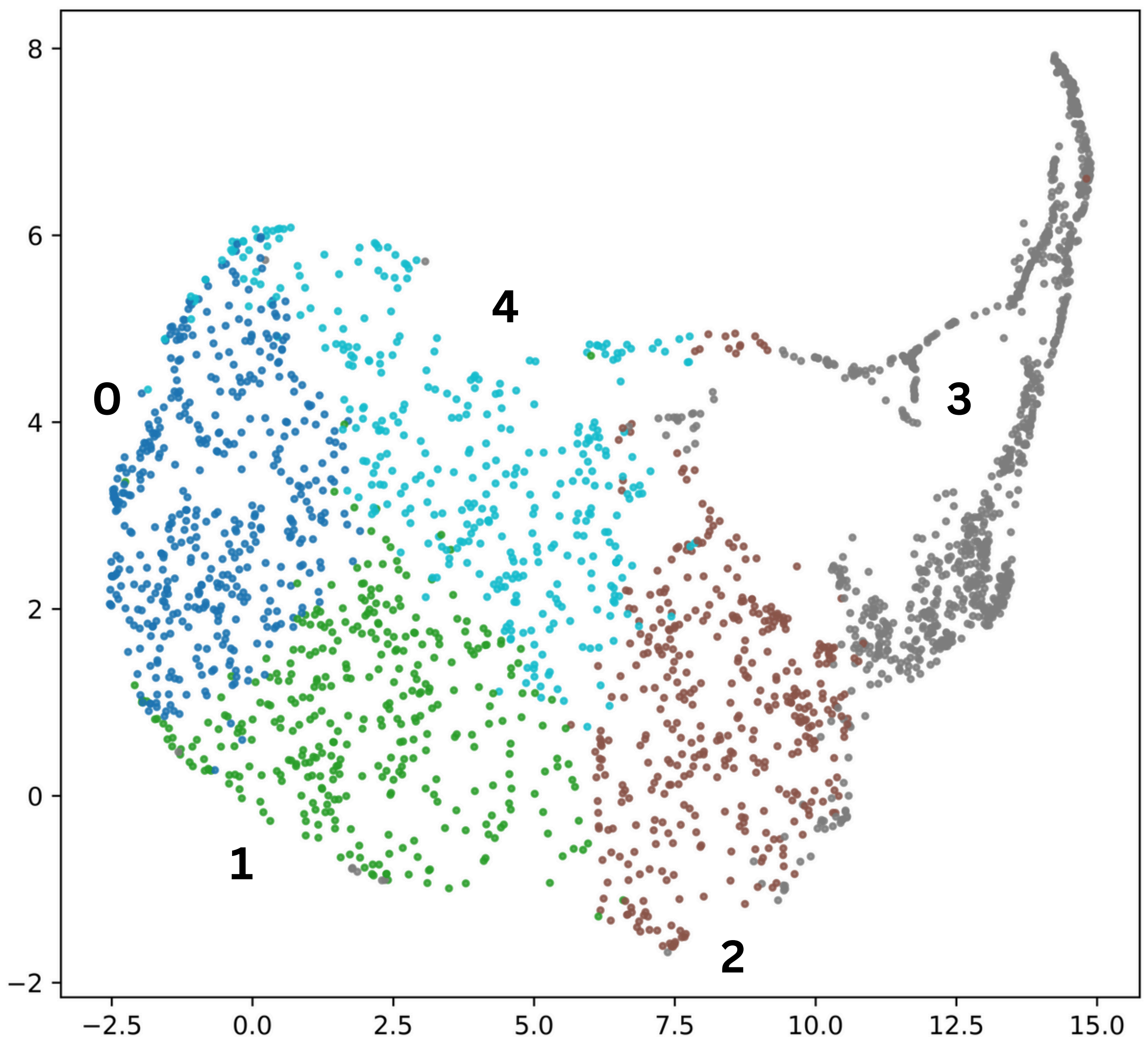




This region does not overlap the protons in the 2D UMAP. Visually match expectations for deuterons.

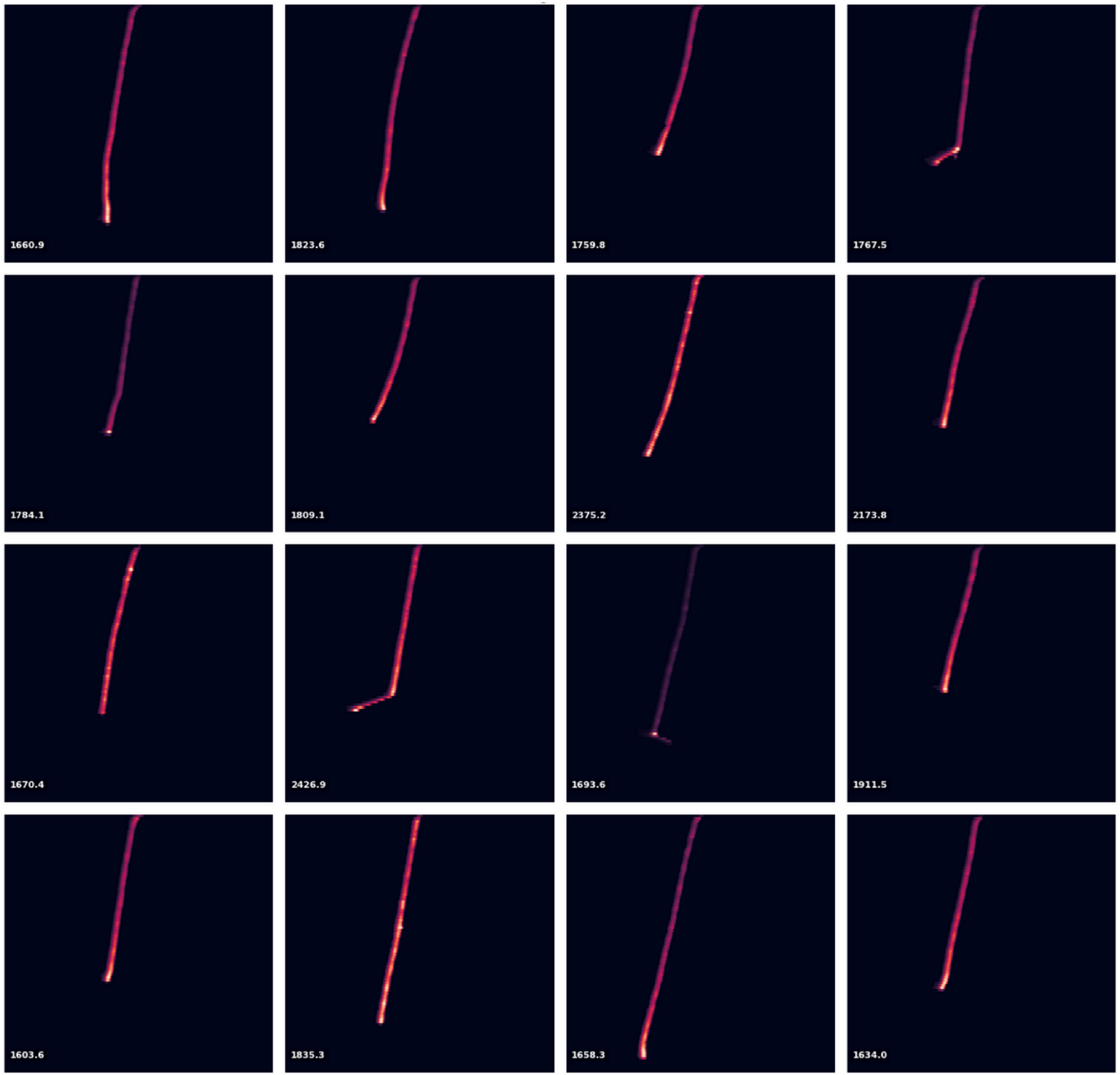
Region 3

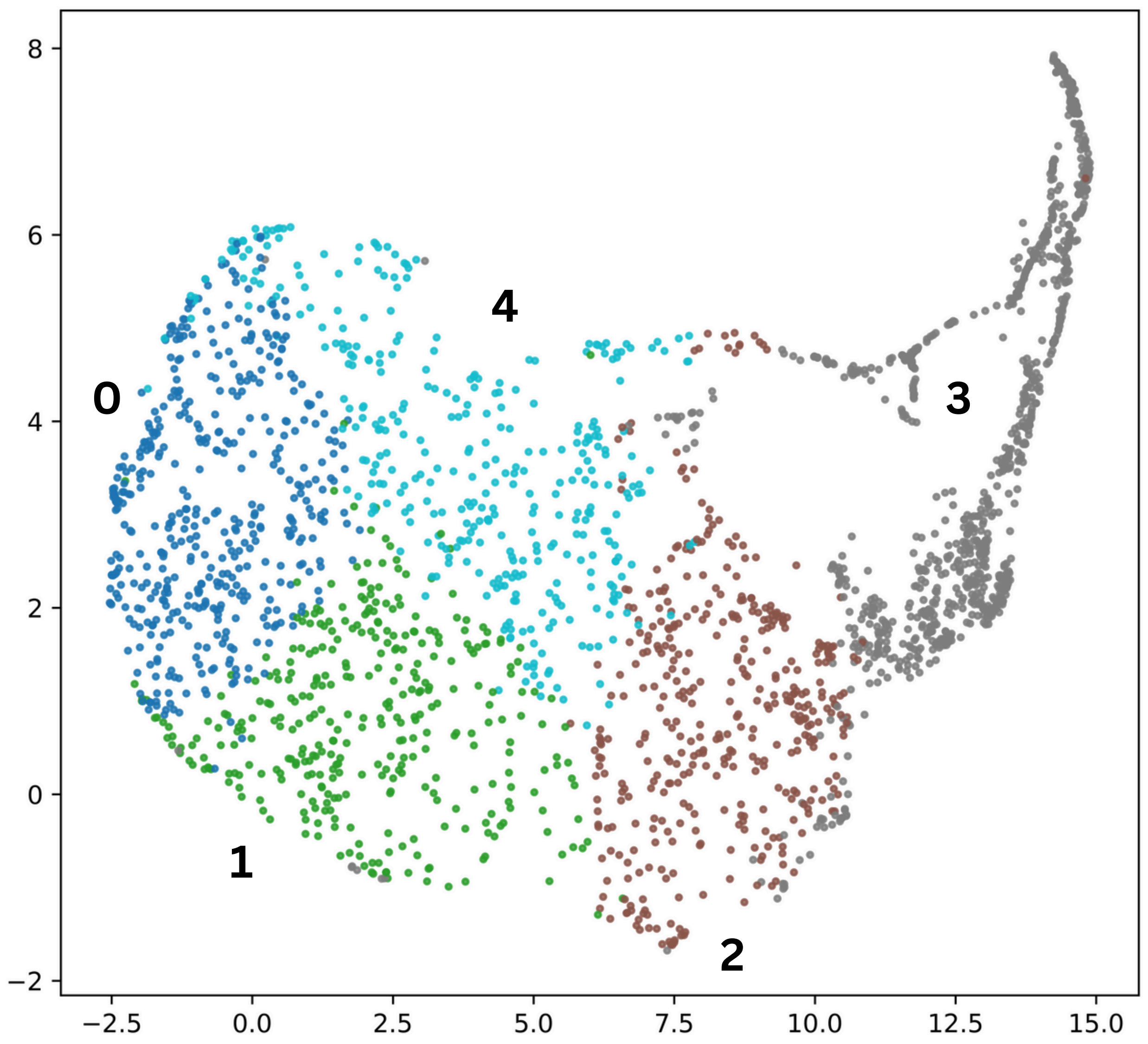




This region does not overlap the protons in the 2D UMAP.

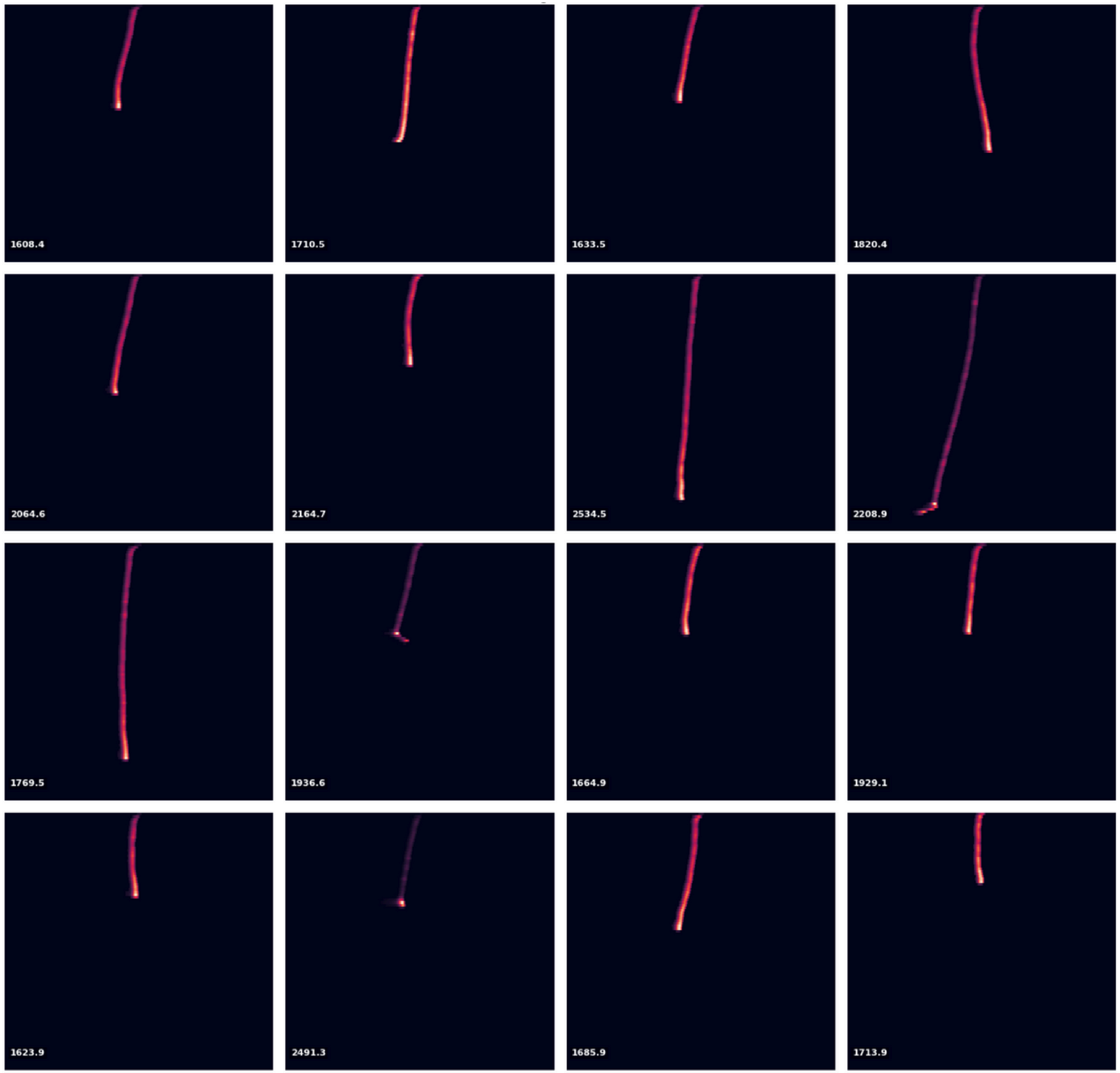
more from Region 3 (only 1 particle)

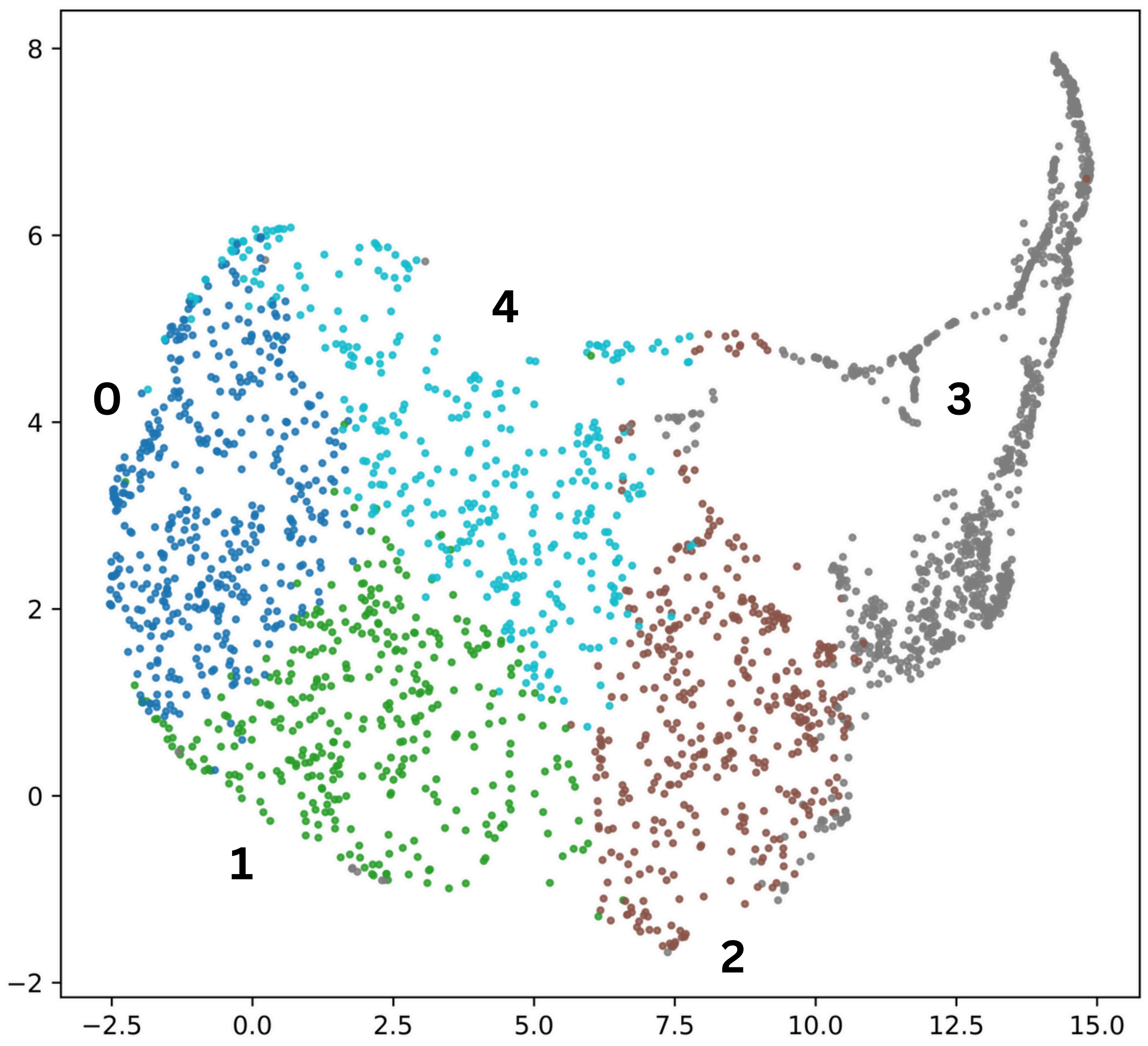




This region overlaps the protons in the 2D UMAP.

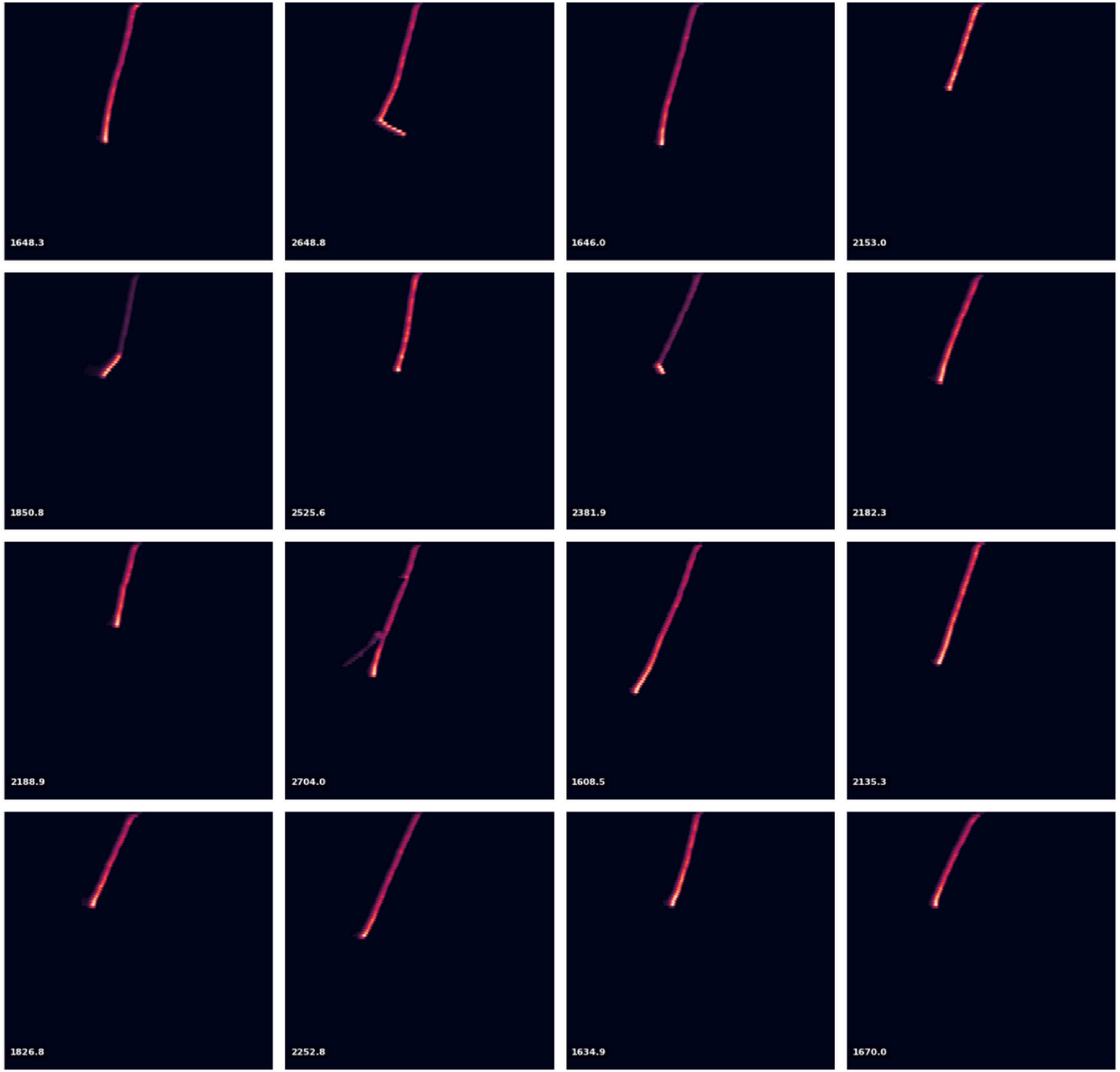
Region 0

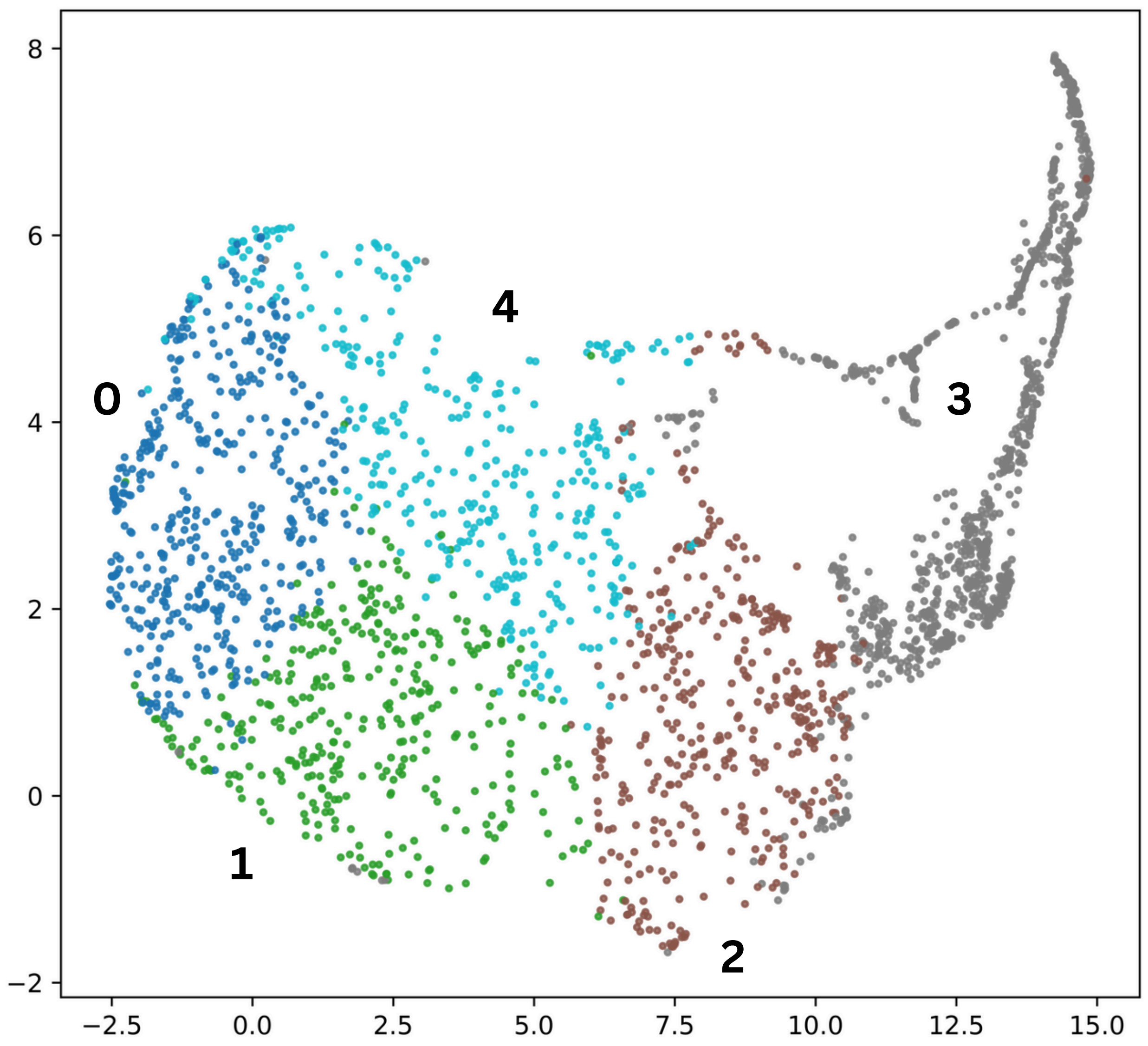




This region does overlap the protons in the 2D UMAP.

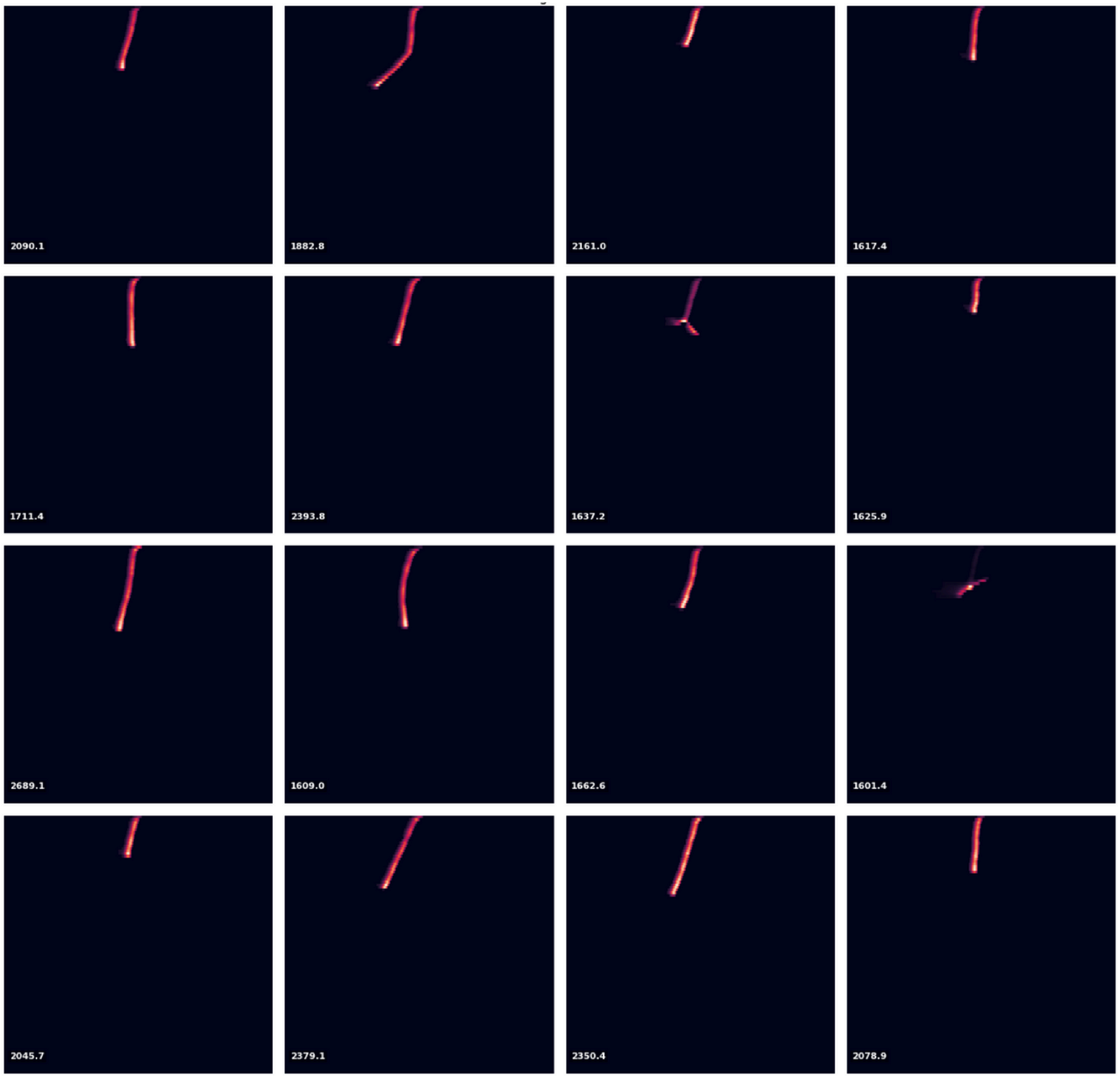
Region 4

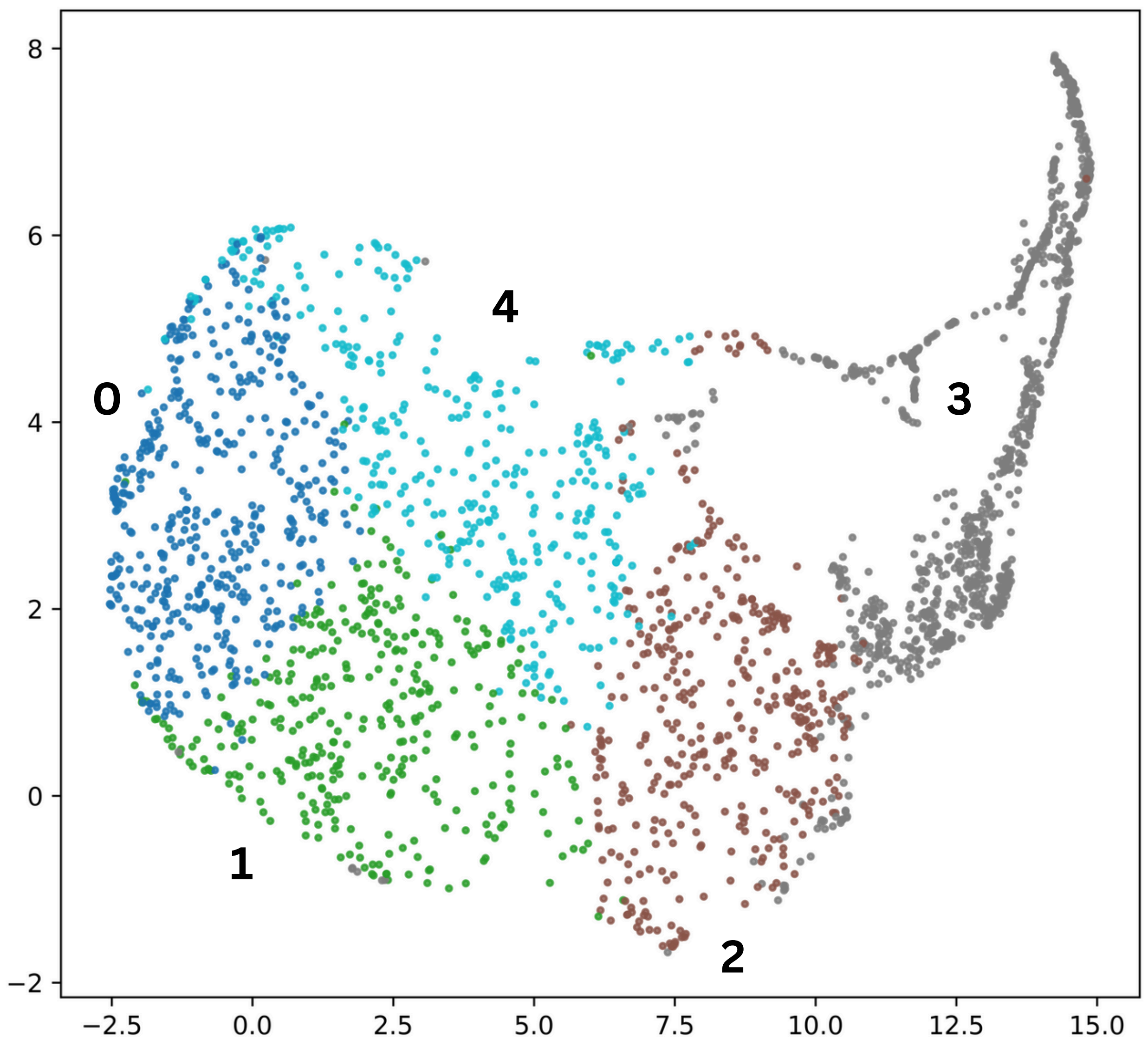




This region does overlap the protons in the 2D UMAP.

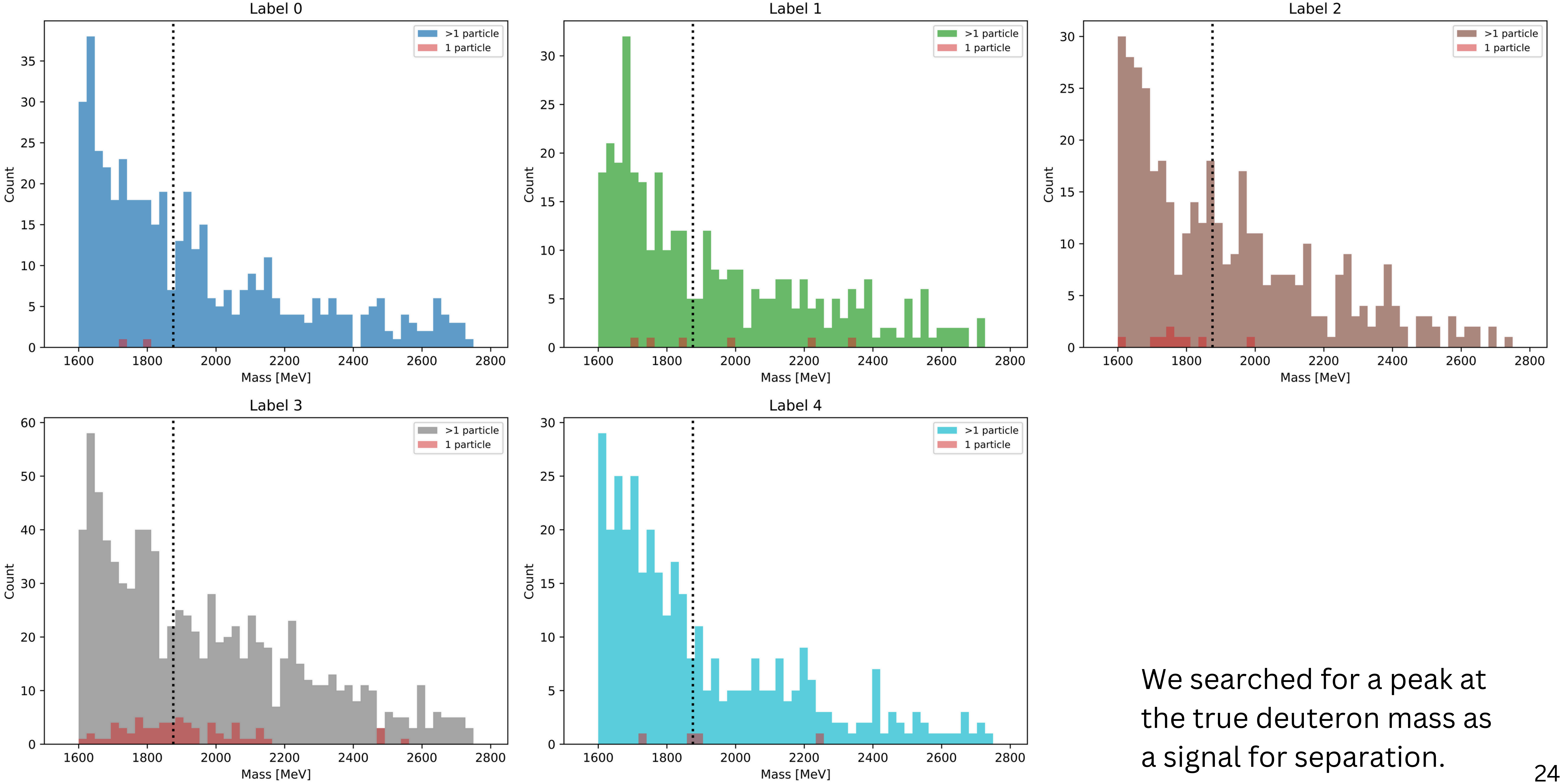
Region 1



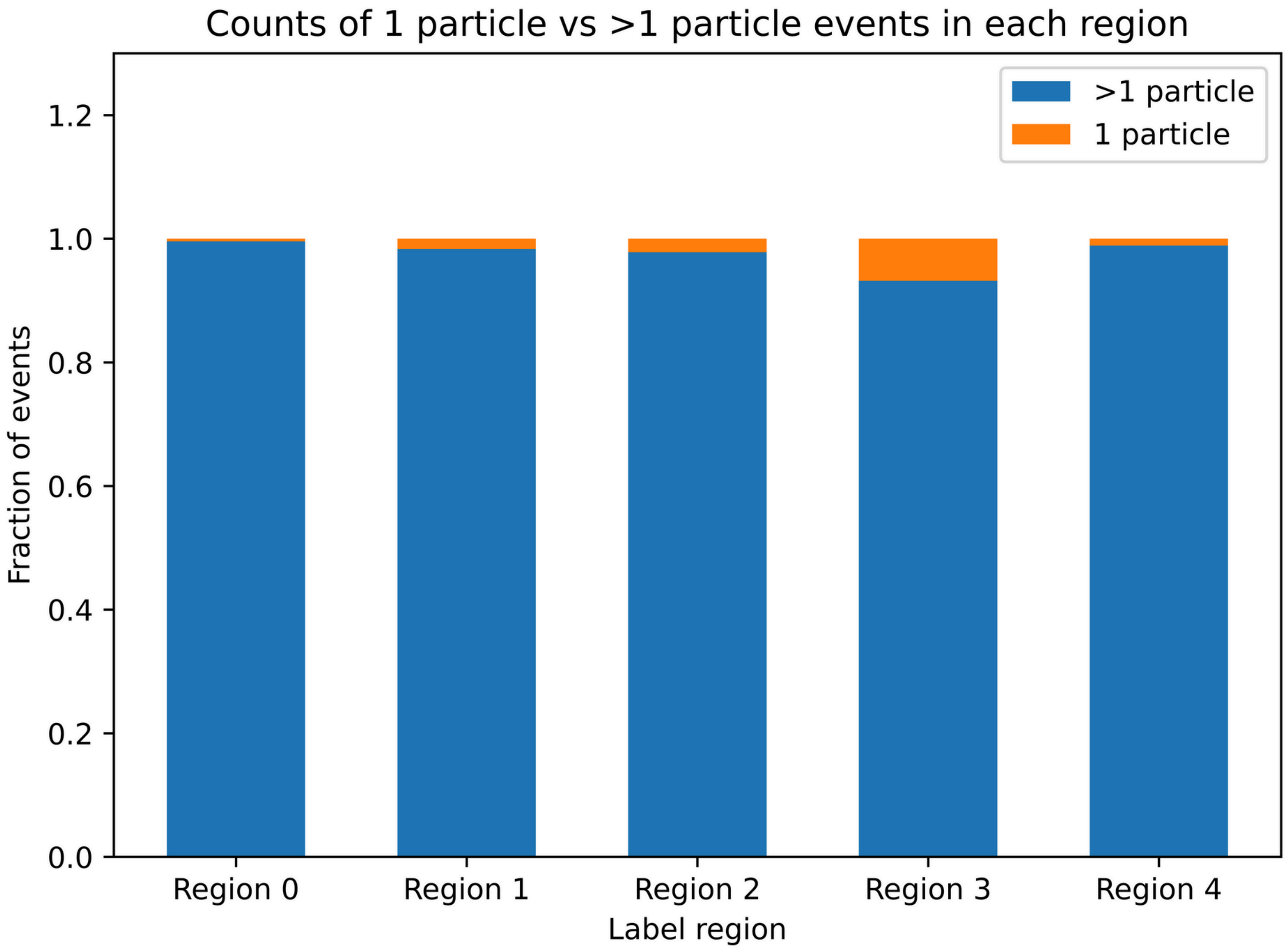


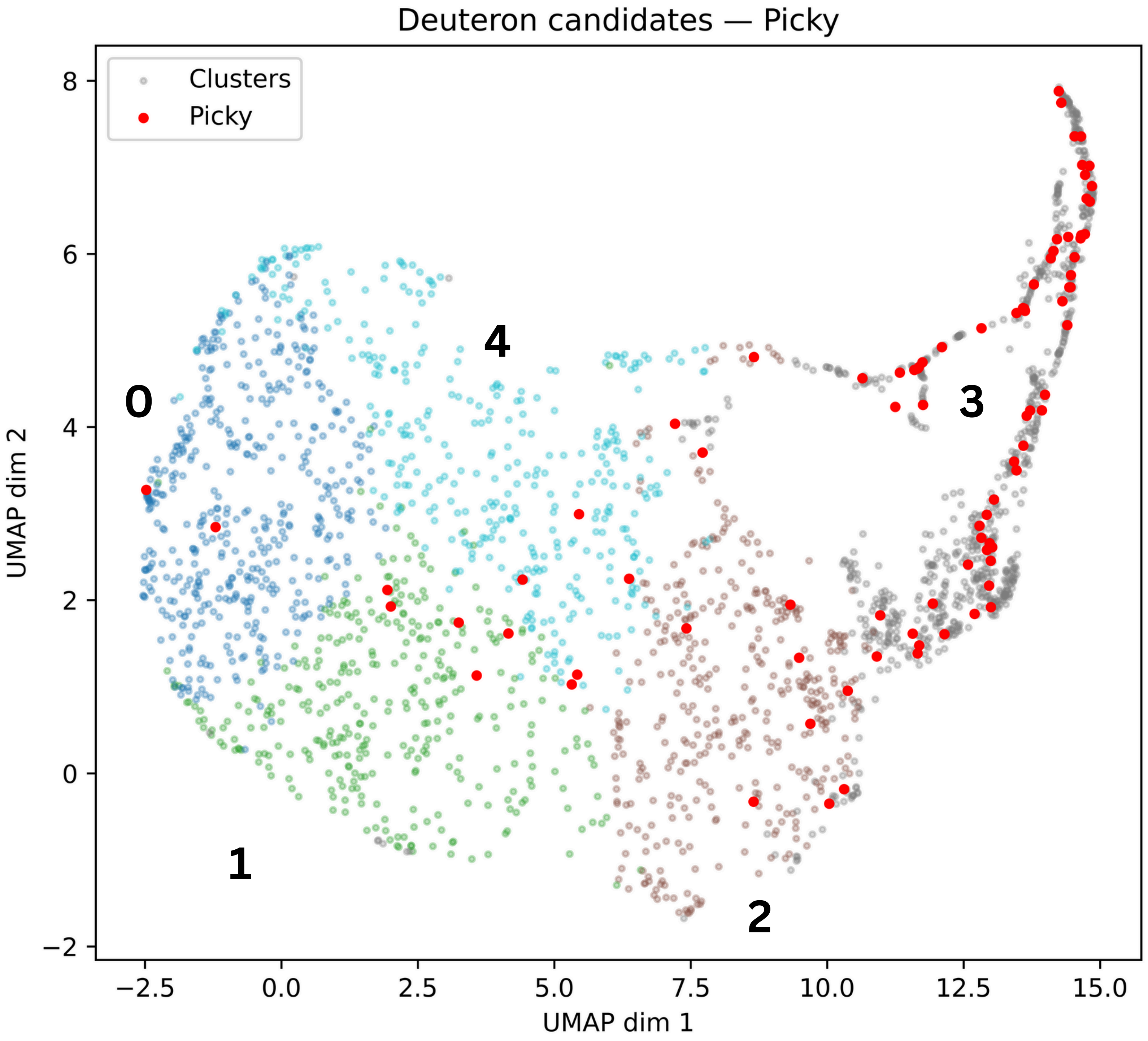
This region does overlap the protons in the 2D UMAP.

Region 2



We searched for a peak at the true deuteron mass as a signal for separation.



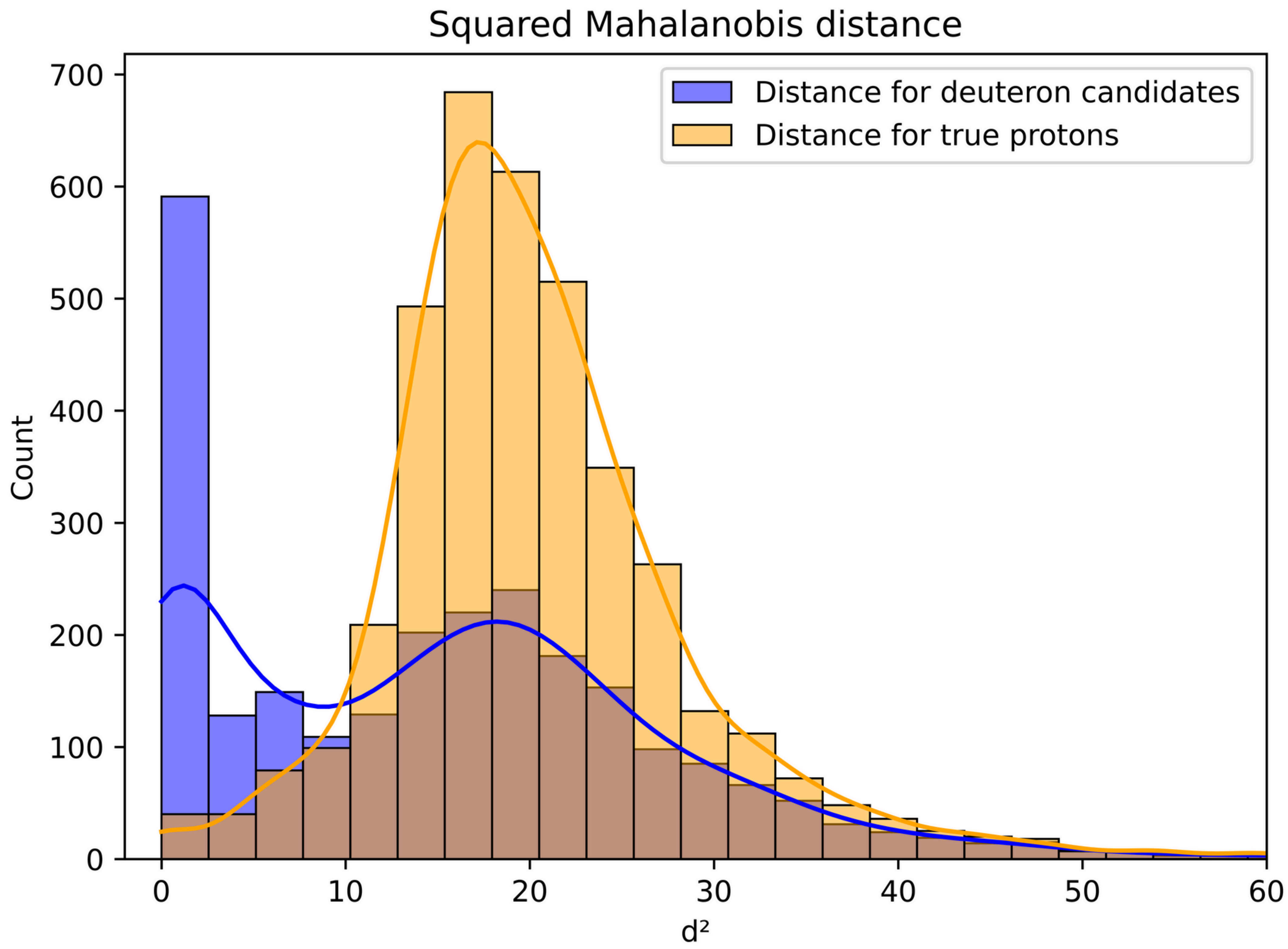


We searched for over-representation of 1 particle events in a cluster as a signal for separation.

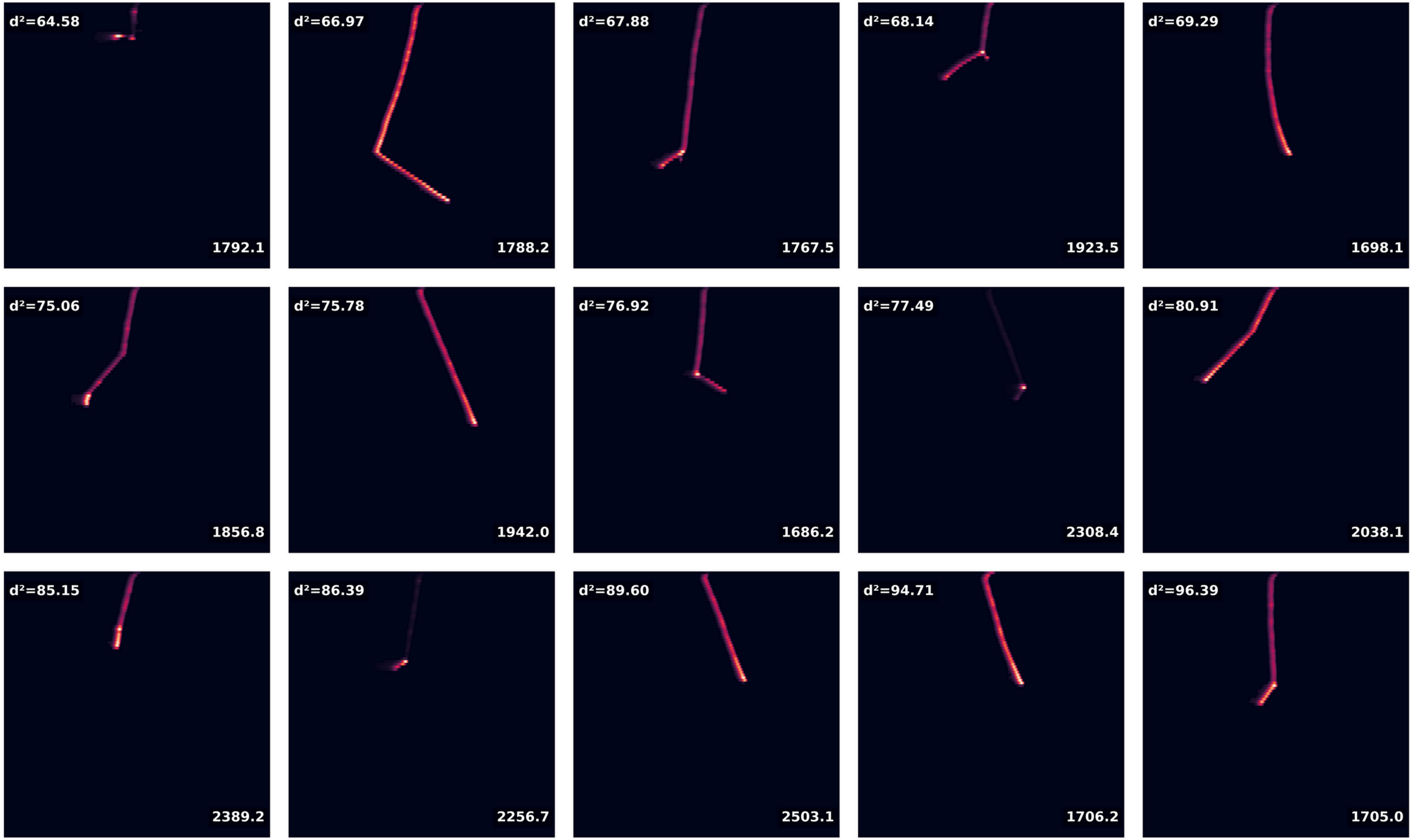
$$D_M^2(x,y) = (x-y)^T \Sigma^{-1} (x-y)$$

Using the **1 particle deuteron events** from region 3 in the latent space as reference points, **we used the Mahalanobis distance as a measure of being "deuteron-like".**

Mahalanobis takes the covariance between the different dimensions in the latent space into account when calculating the distance. When covariance = identity matrix, it reduces to Euclidean distance.



Some low distance clusters



Some high distance clusters

Conclusion

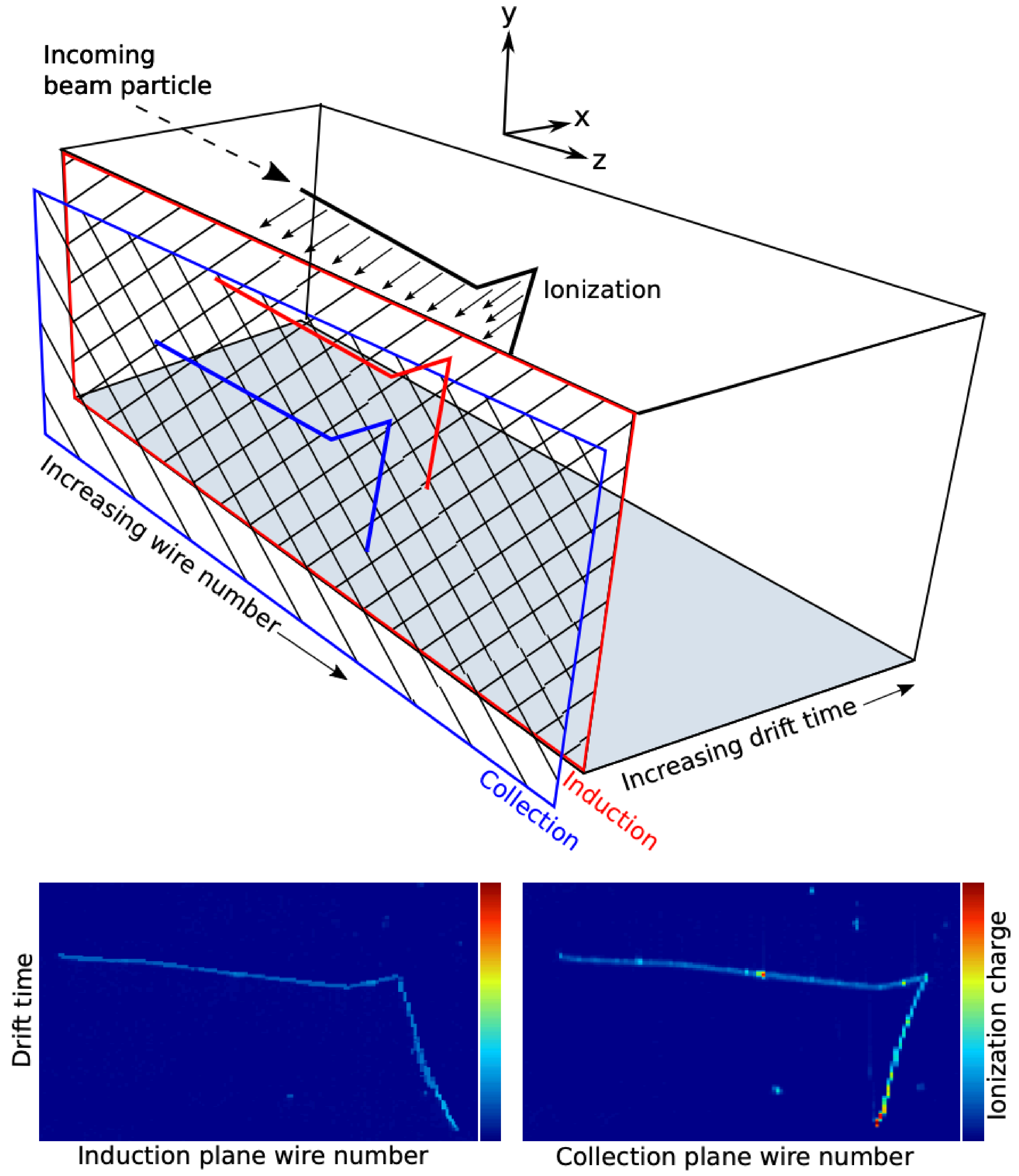
- **Proton–deuteron separation is critical** for accurate neutrino measurements (e.g. energy reconstruction), yet extremely challenging due to similar short-track topologies.
- **LArIAT provides the ideal environment** to study and disentangle these particles, offering controlled beamline data with particle-type priors, but reconstruction inefficiencies and noise complicate identification.
- **Our convolutional autoencoder approach** enables separation directly from raw images, revealing latent representations that capture features like length, allowing us to localise “deuteron-like” regions, recover missed deuterons from noisy data and separate them from protons.
- **The selection of reference points** for the Mahalanobis distance can be further refined to make sure noisy clusters don’t make through.

Questions?

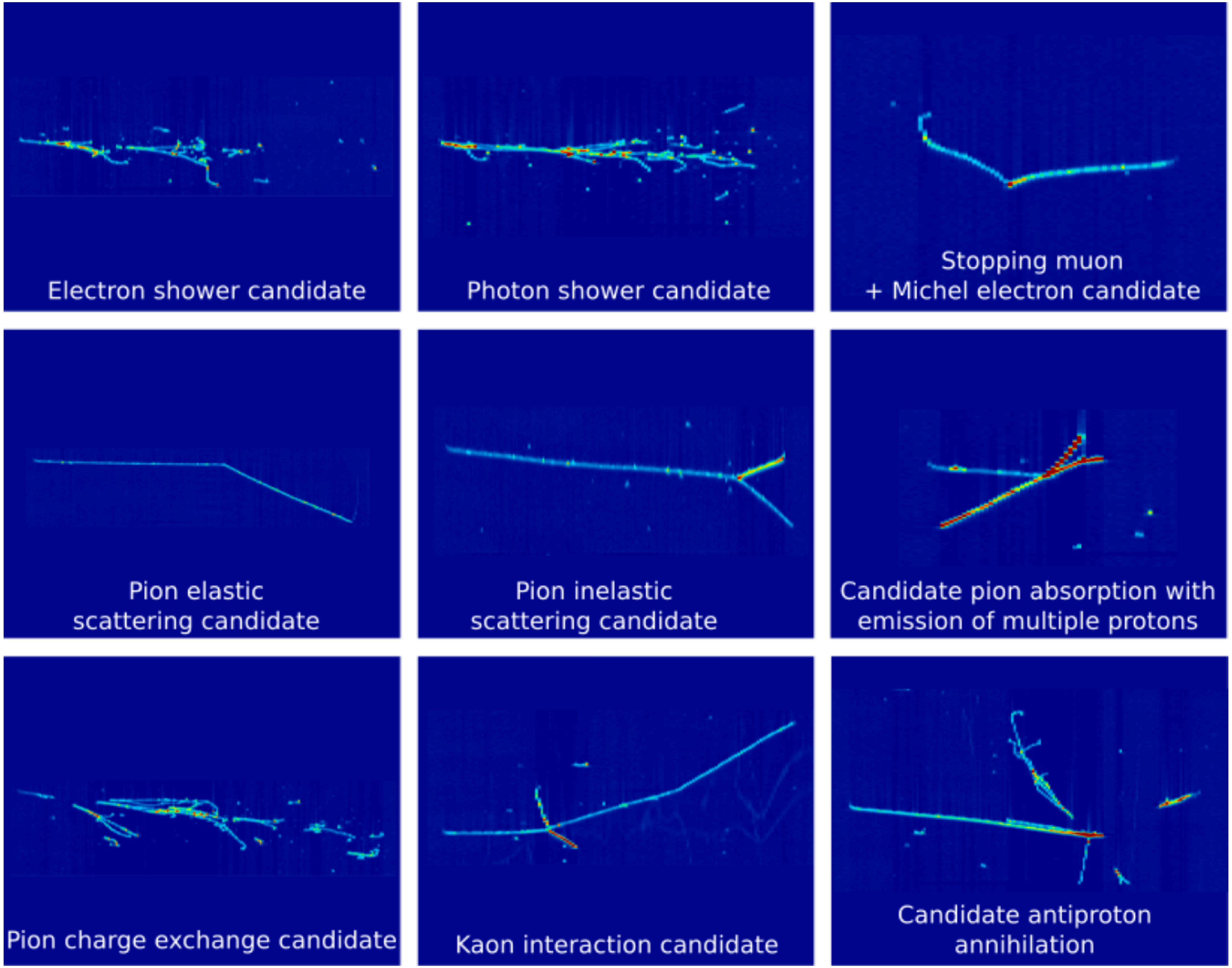
Neutrino Physics and LArTPCs

- The universe **should contain equal amounts** of matter and antimatter, yet **we observe a clear matter excess**.
- This imbalance is **not explained** by the Standard Model.
- A key ingredient is Charge–Parity (CP) violation: break in symmetry between matter and antimatter.
- Neutrino oscillations provide a probe by comparing the behaviour of neutrinos and antineutrinos.

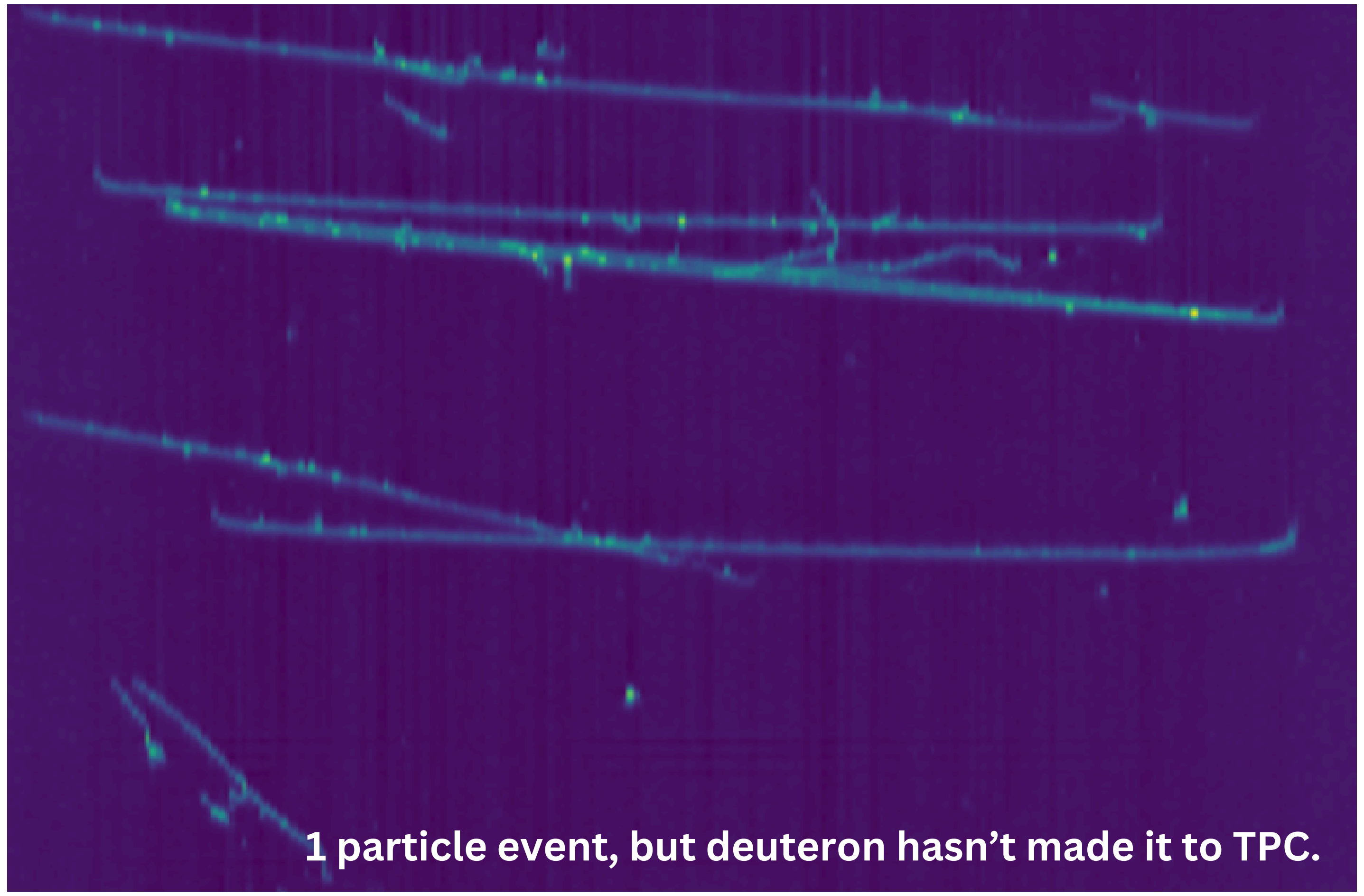
A **Liquid Argon Time Projection Chamber** (LArTPC) is a kind of neutrino detector that **can reconstruct particle paths from neutrino interactions with high precision**. LArTPCs have emerged as a central tool in neutrino physics.

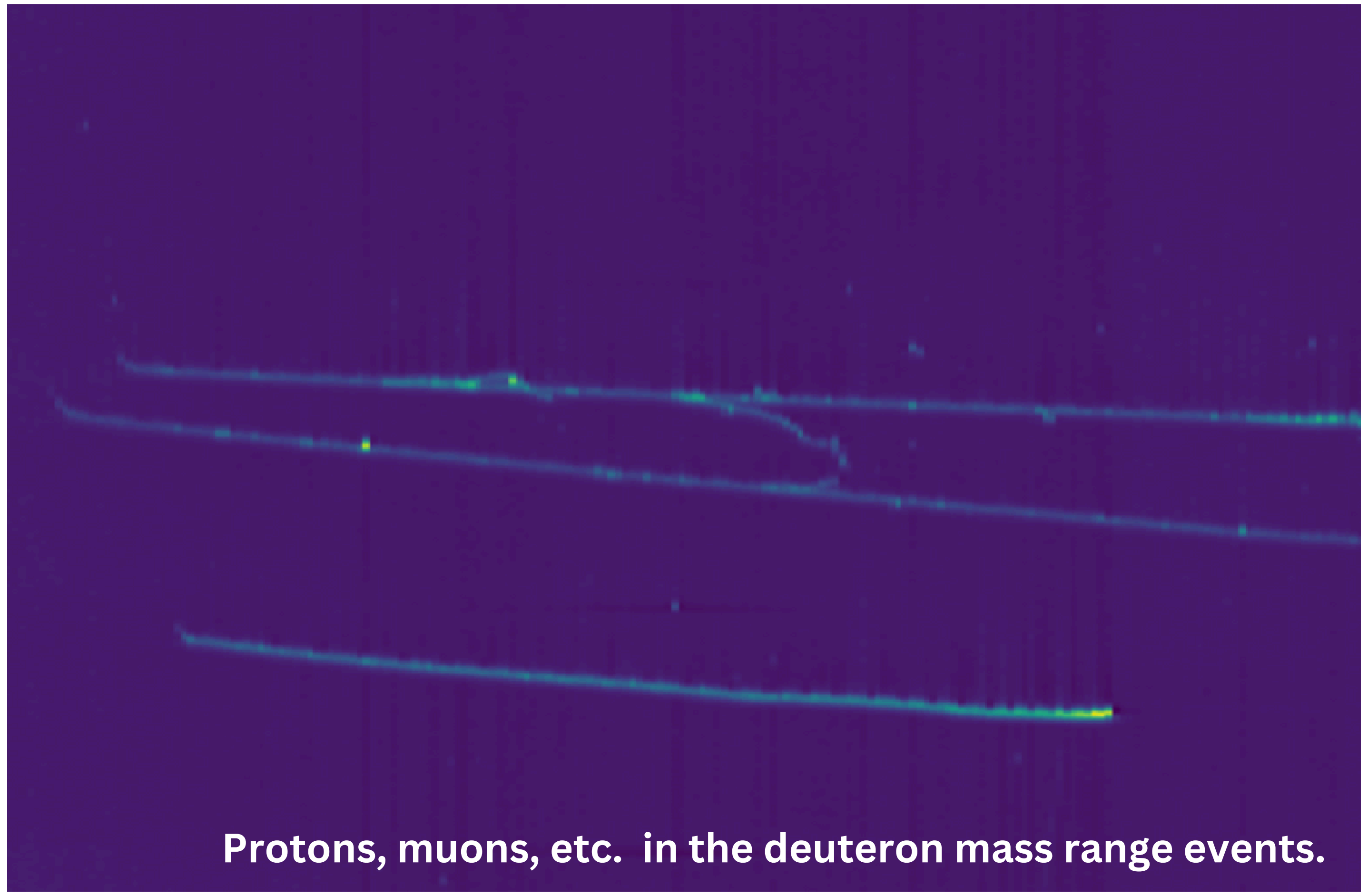


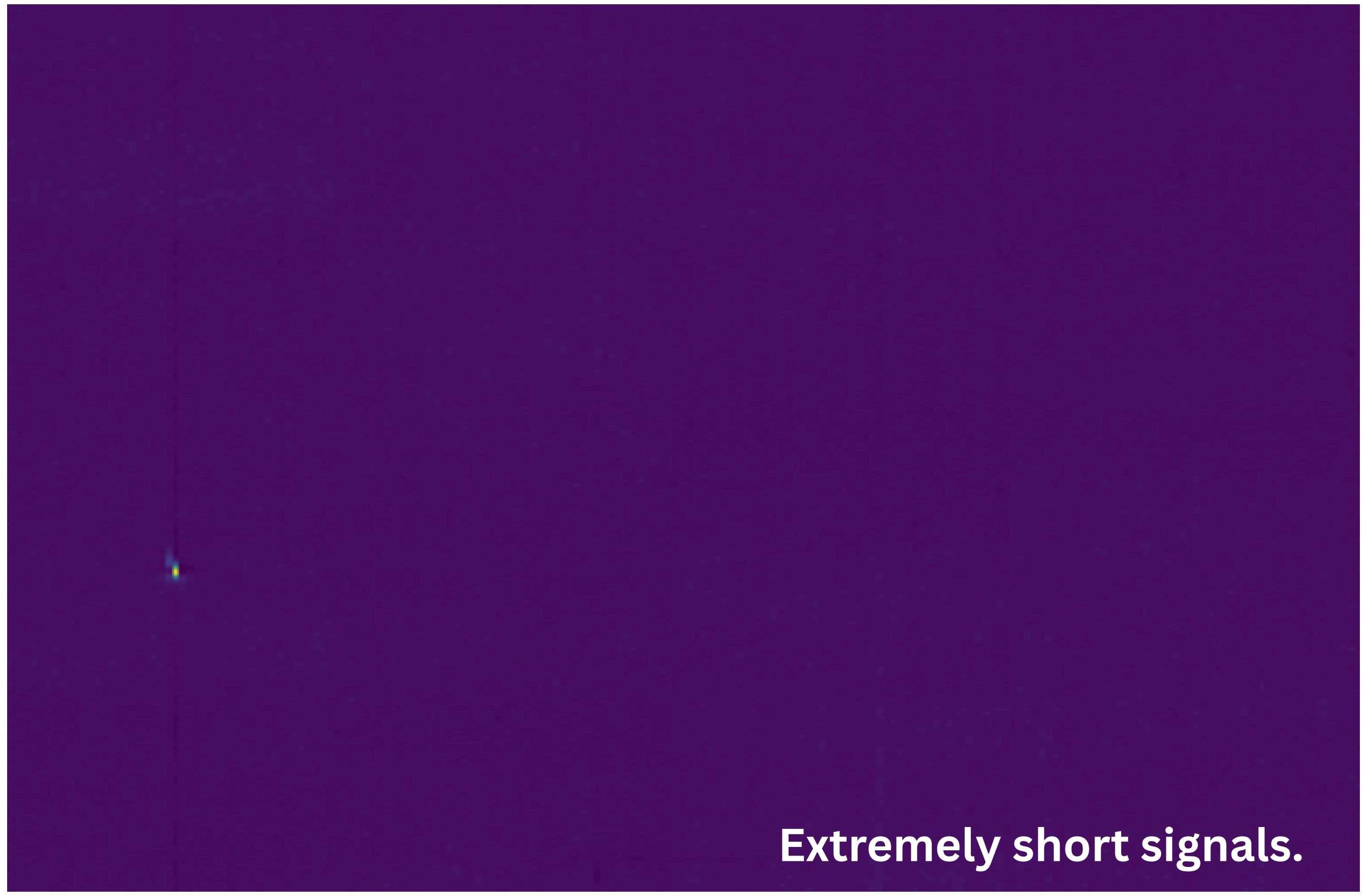
Schematic of the LArTPC operation. An incoming charged particle ionises the liquid argon, producing electrons that drift under an electric field towards the wire planes. Induction and collection planes record complementary 2D projections of the track, which can be combined to reconstruct the full 3D trajectory.



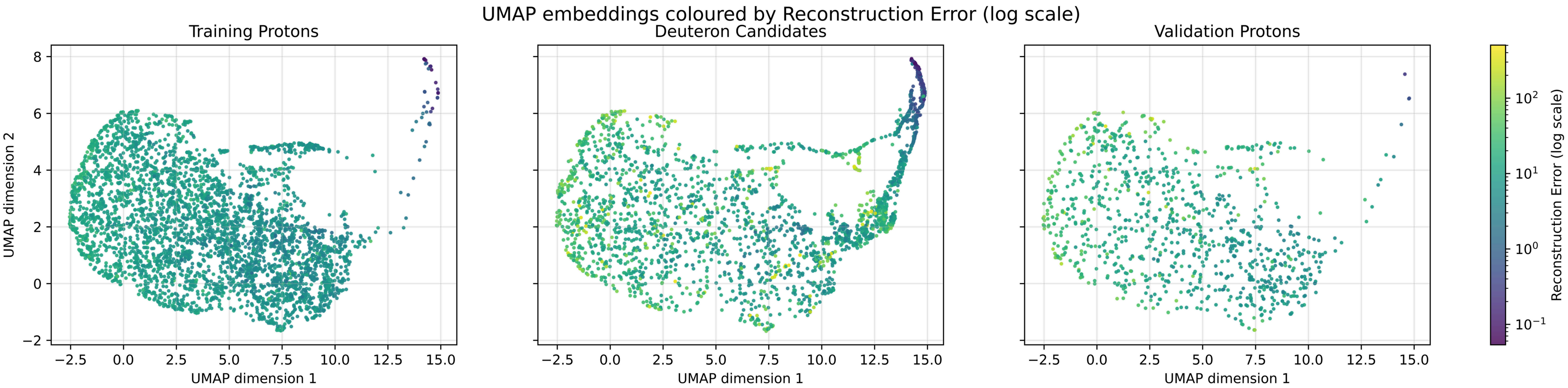
Example LArIAT events. Tracks enter from the left. Collection plane only. The colour indicates density of ionisation charge. Different scales in each graph.


1 particle event, but deuteron hasn’t made it to TPC.


Protons, muons, etc. in the deuteron mass range events.

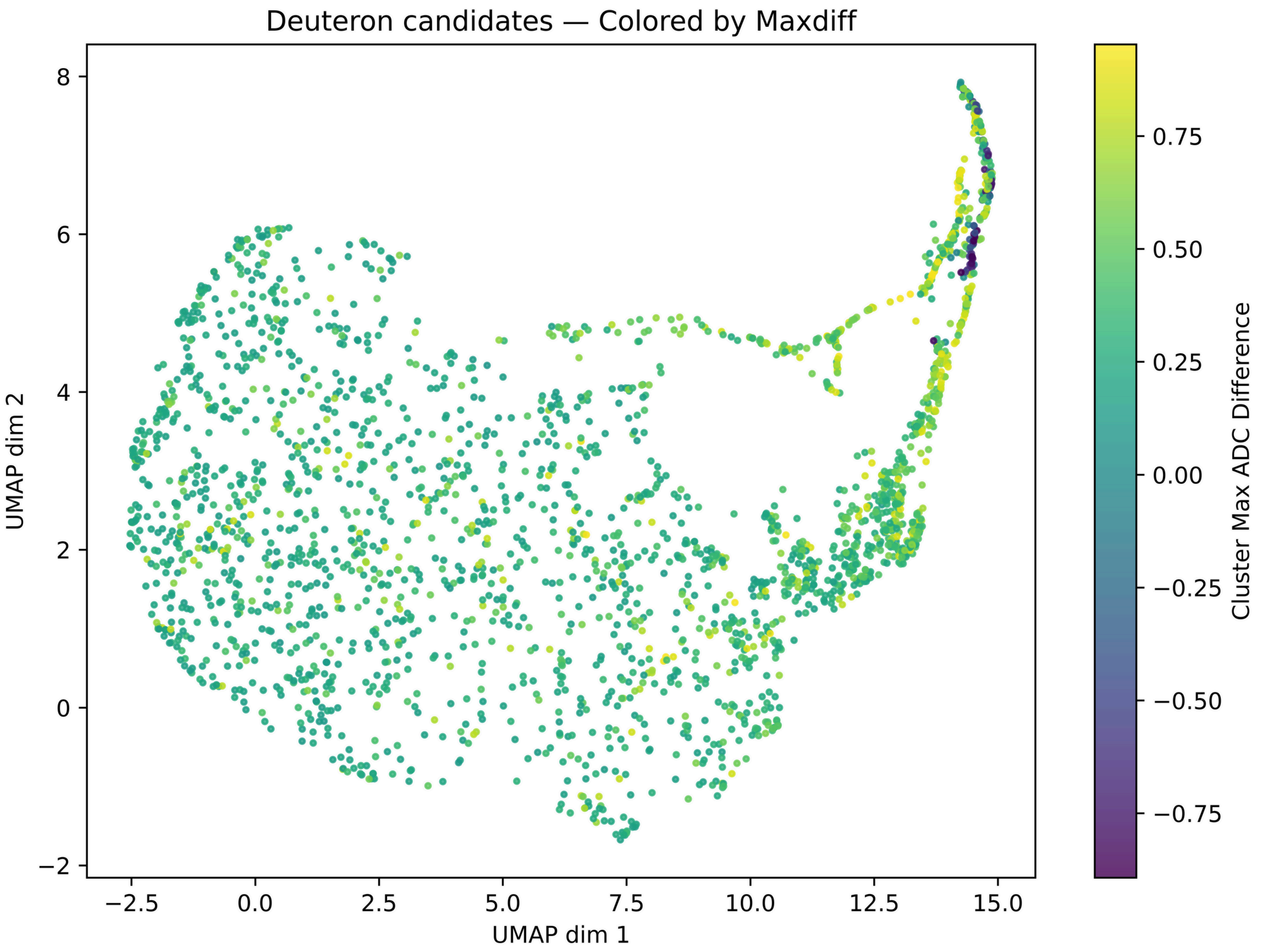

Extremely short signals.

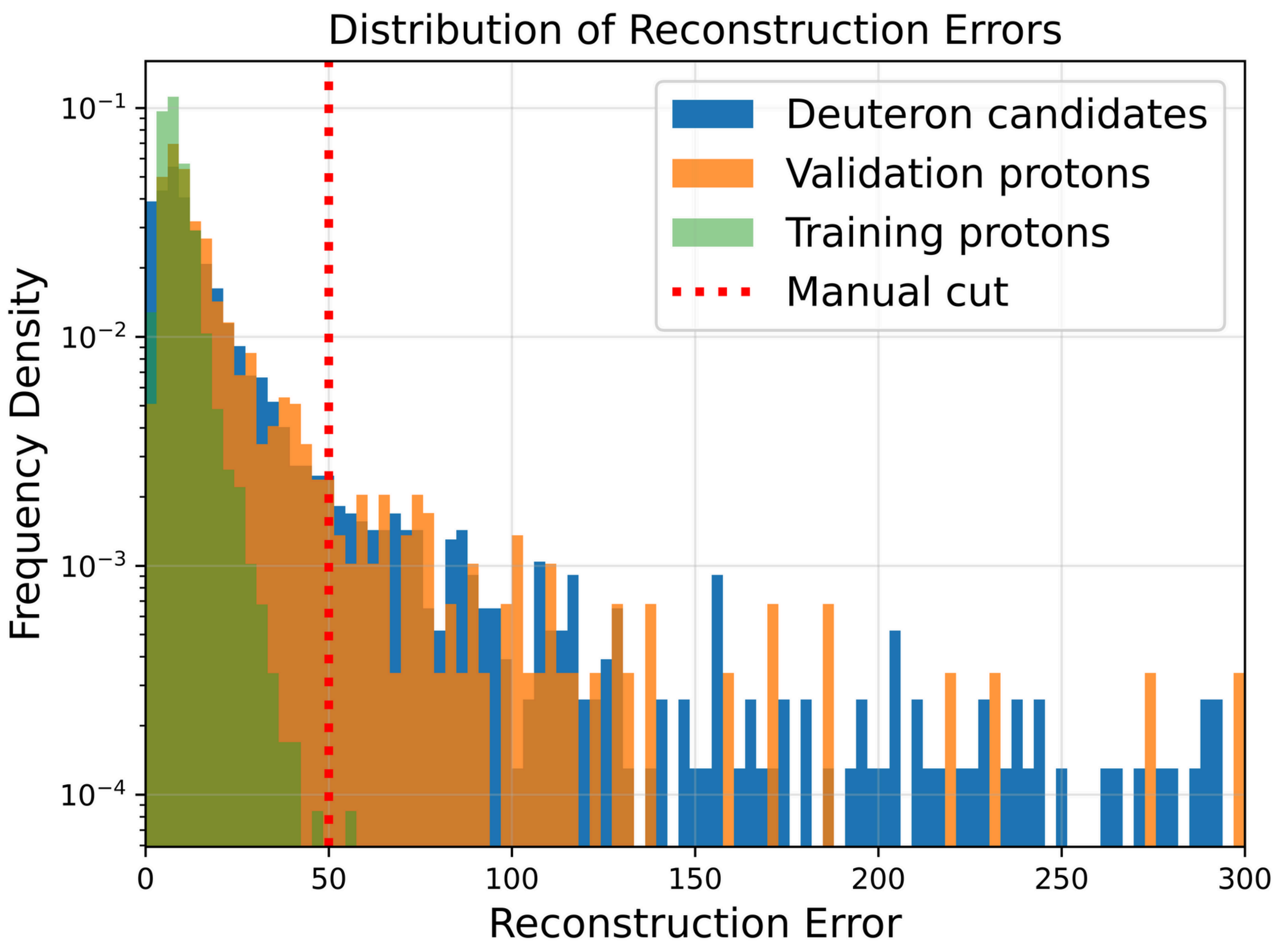
Noise Modes

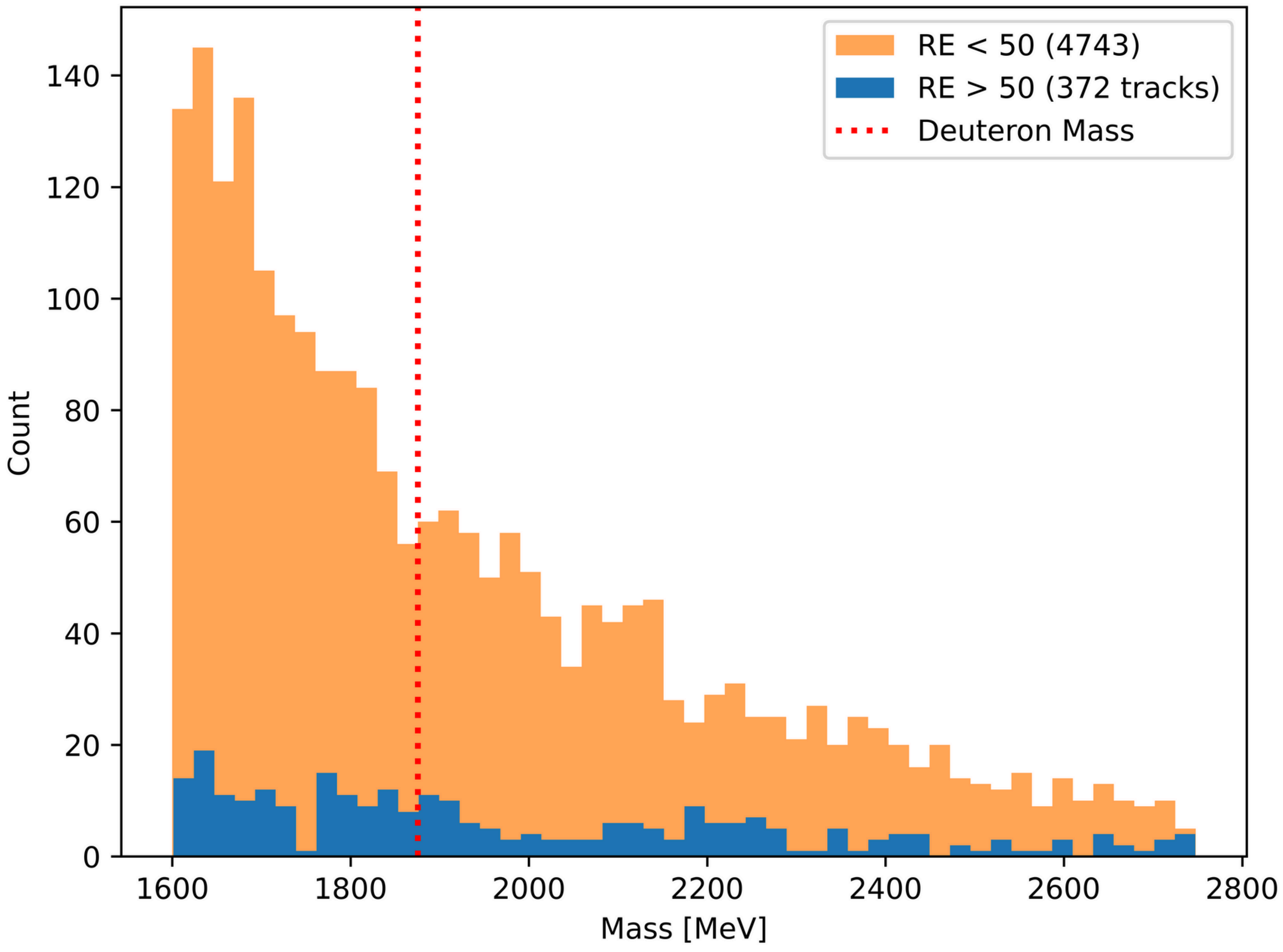


Separation using reconstruction error only was inviable.

The latent space seems to be encoding the energy profiles of the particles as well.



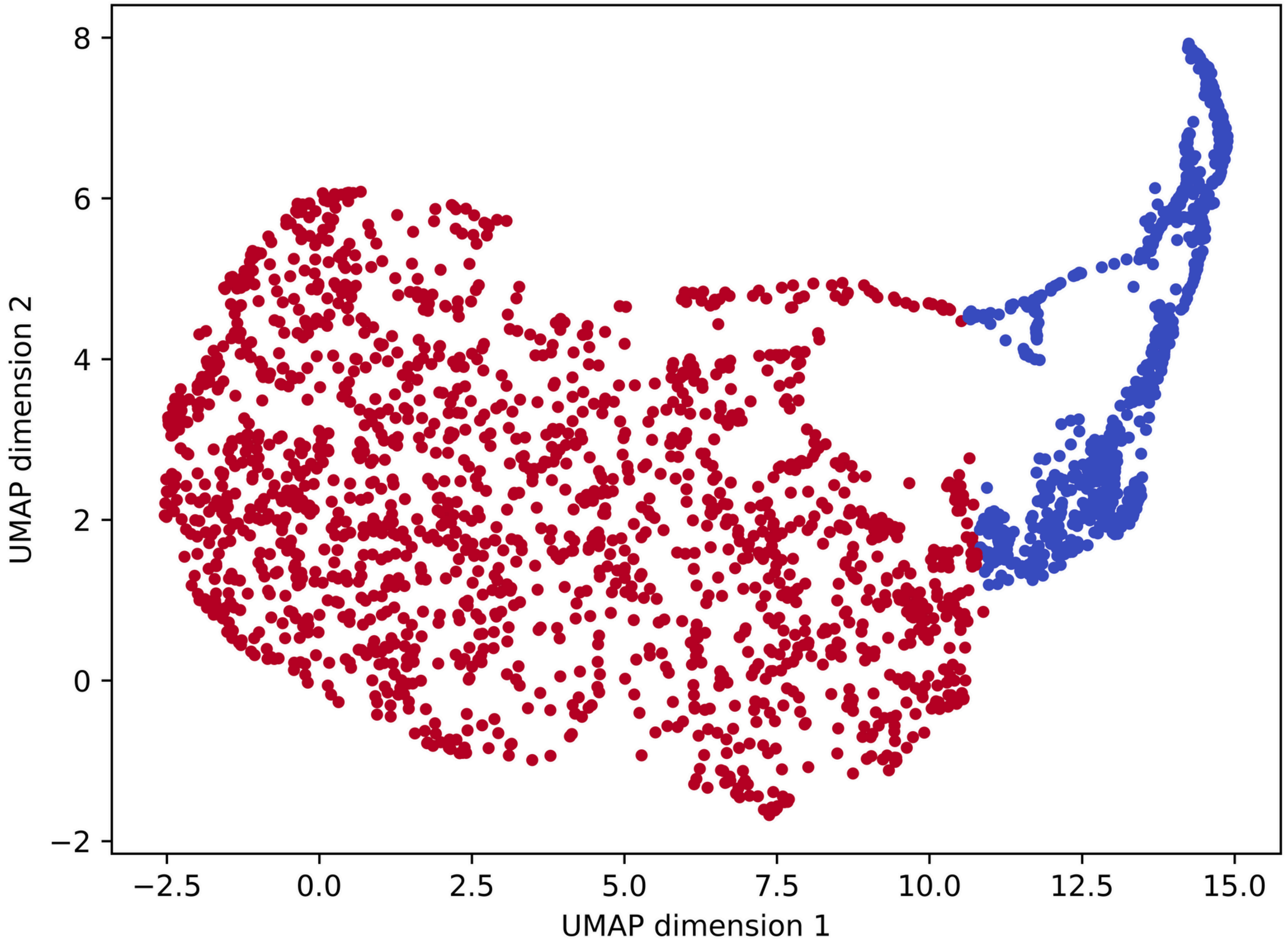


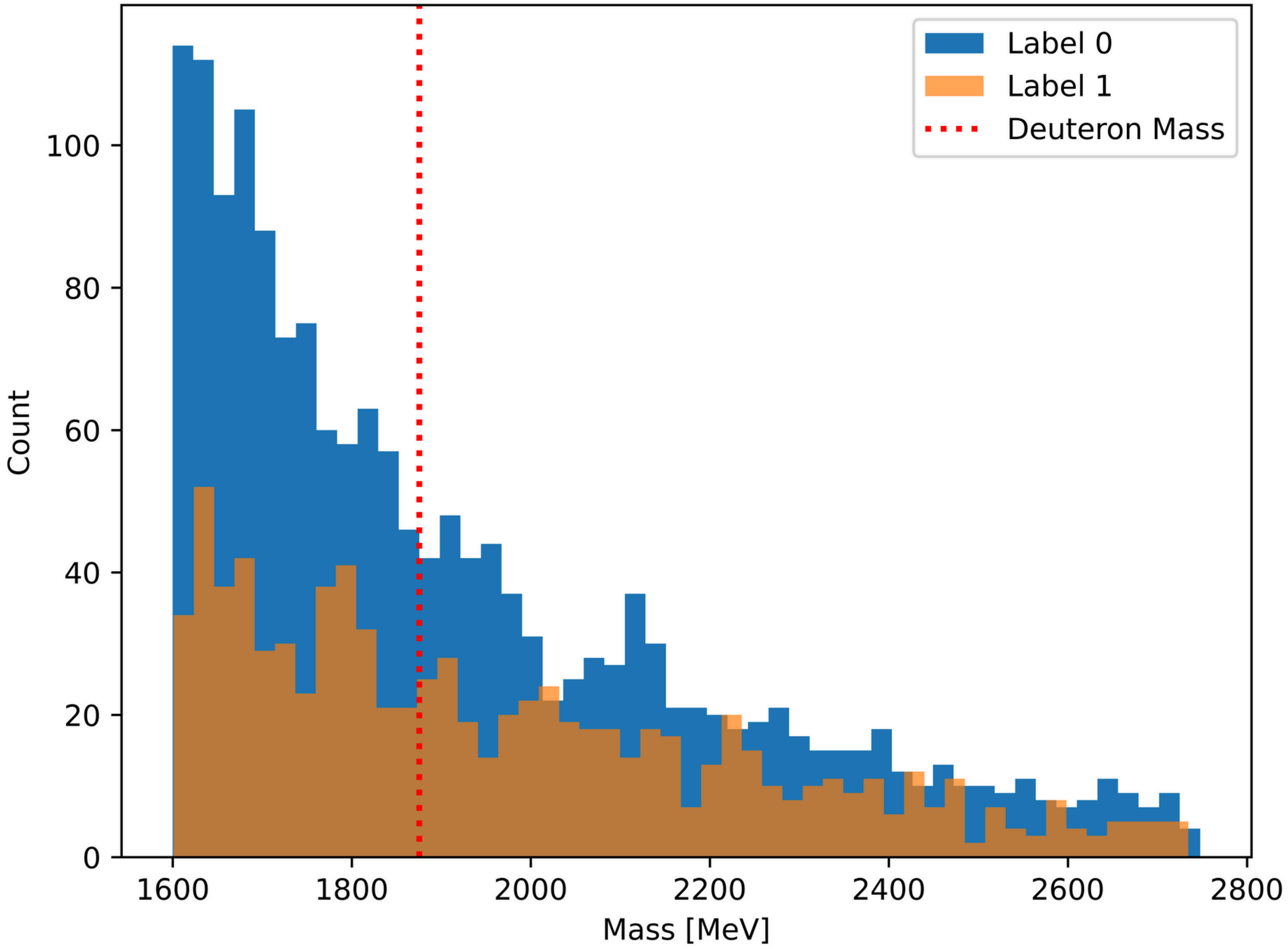


A manual cut on images with high reconstruction error (> 50), presumed to be non-protons was made. Separation was hard to determine.

Linear Cut

We began with a linear cut on the 2D UMAP space to separate proton-like and non-proton candidates. No separation was clear.



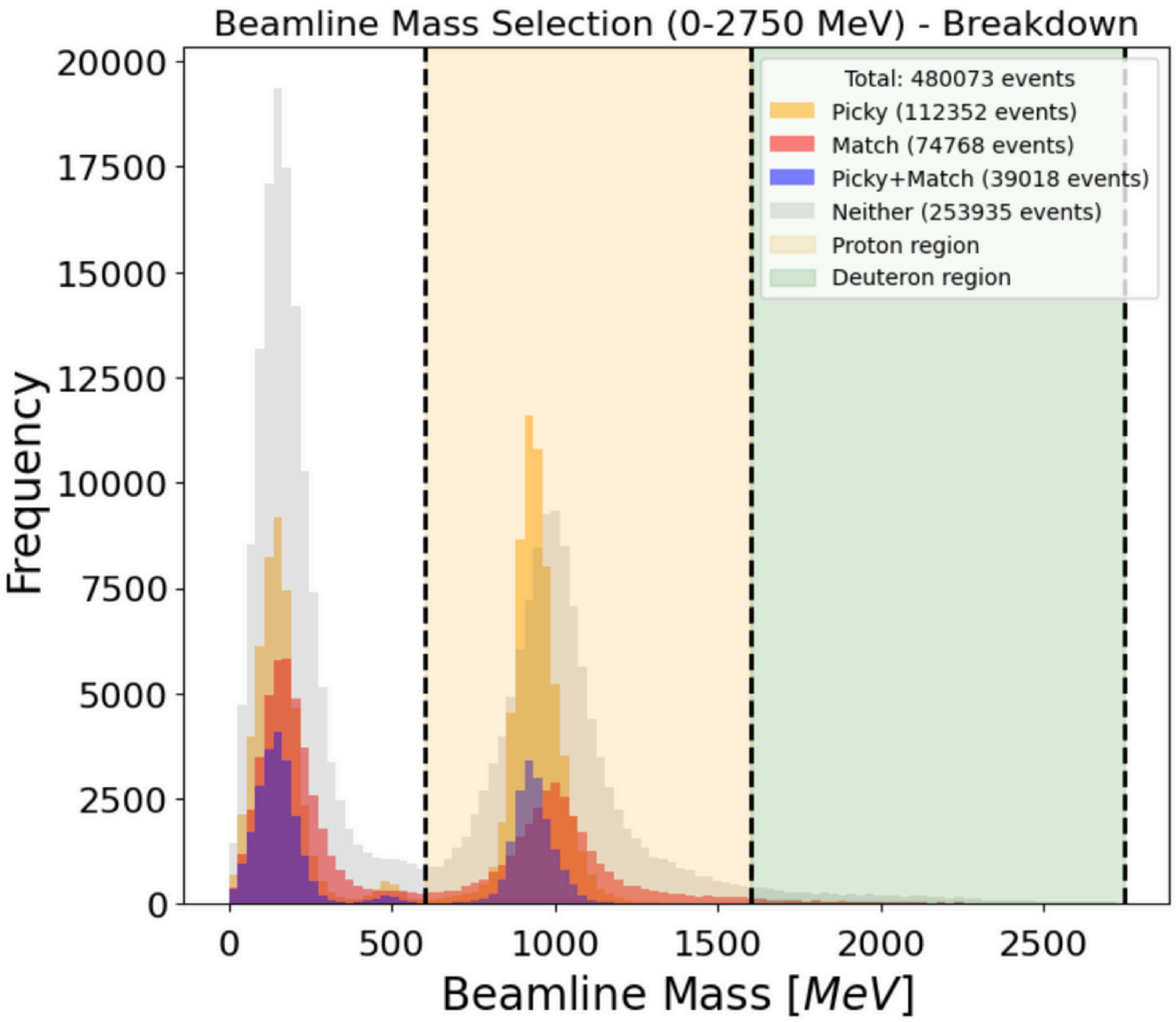


Identifying protons & deuterons

Two variables capture event quality in the reconstruction data:

Picky: where **only one particle goes through the wire chambers** before the LArTPC

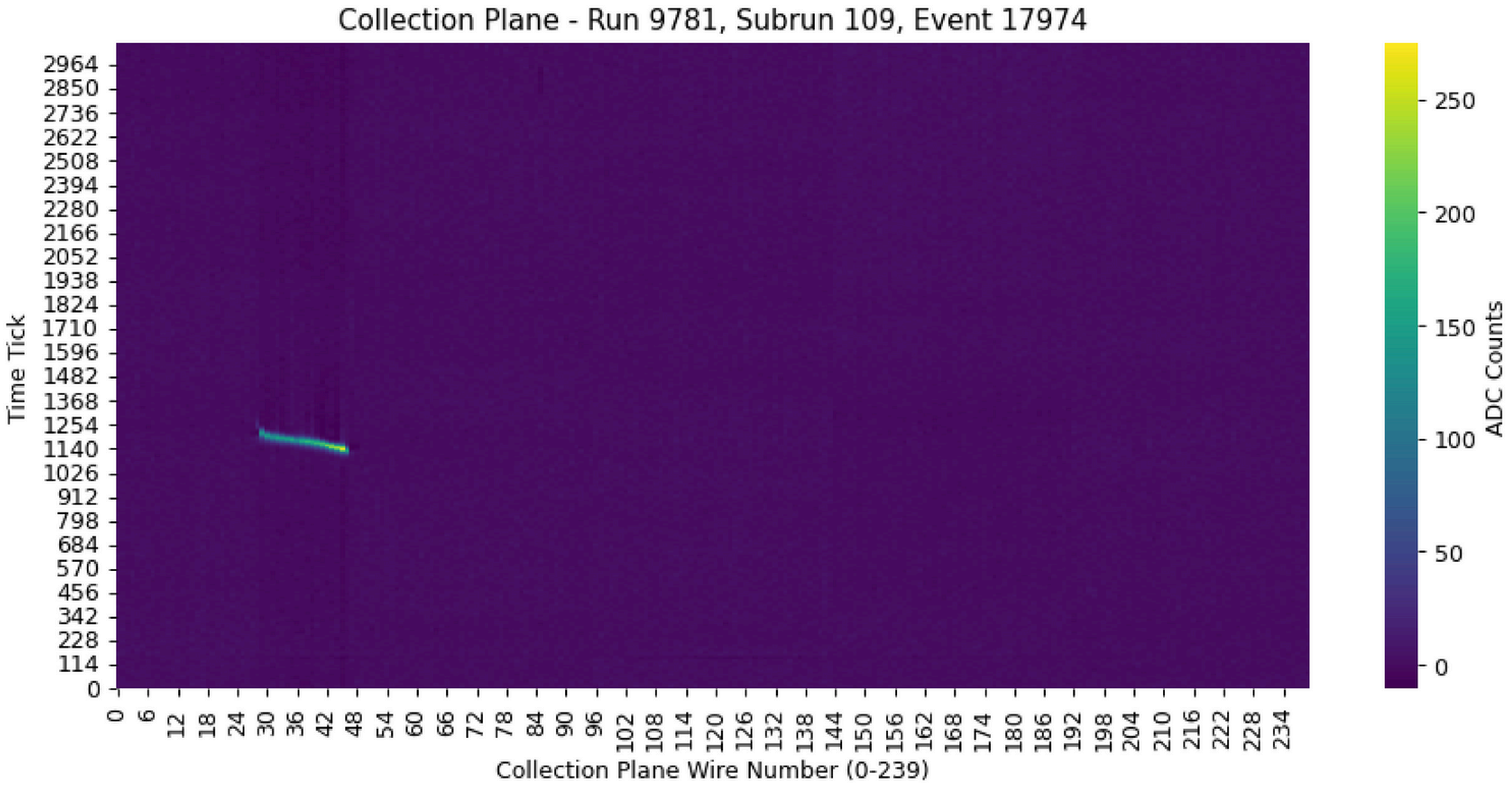
Match: where the particle position at the **wire chambers matches the position projected on the face of the LArTPC**



Cleaning Proton Selection

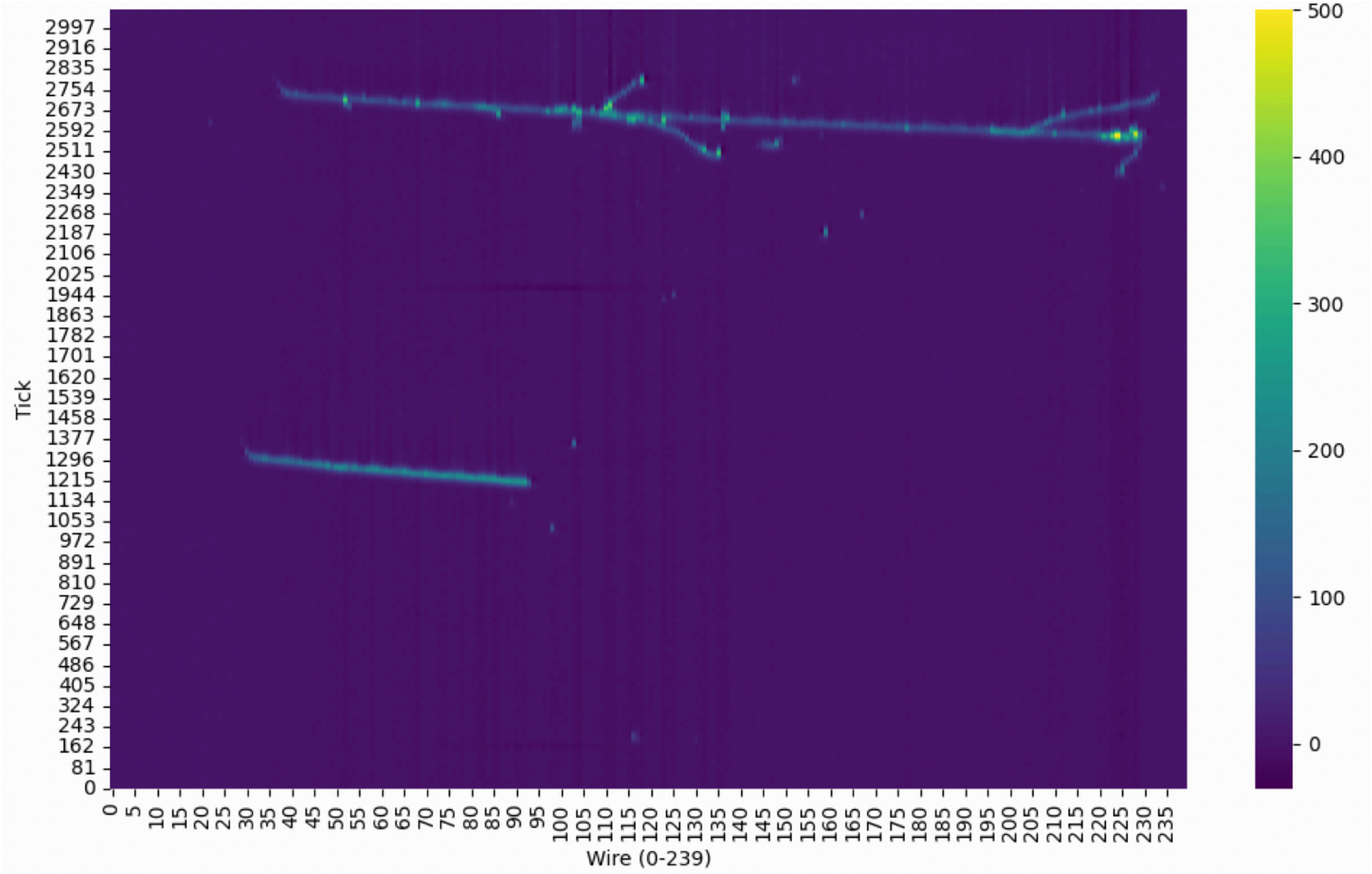
Using reconstruction data, chose proton candidates with **one track only**, keeping events like these.

This brought down the number of **proton candidates to 7636.**



Cleaning Proton Selection

This cut removed events like these.



Cuts

Selection	Events
No cuts	354,879
+ Keep only clusters > 3 wires long	87,712
+ constant max ADC difference	87,703
+ remove clusters with only negative ADC change	77,721
+ remove clusters > 180 wires	73,878
+ ADC threshold (100 for collection, 50 for induction)	61,080
+ only clusters within bounding box	11,889

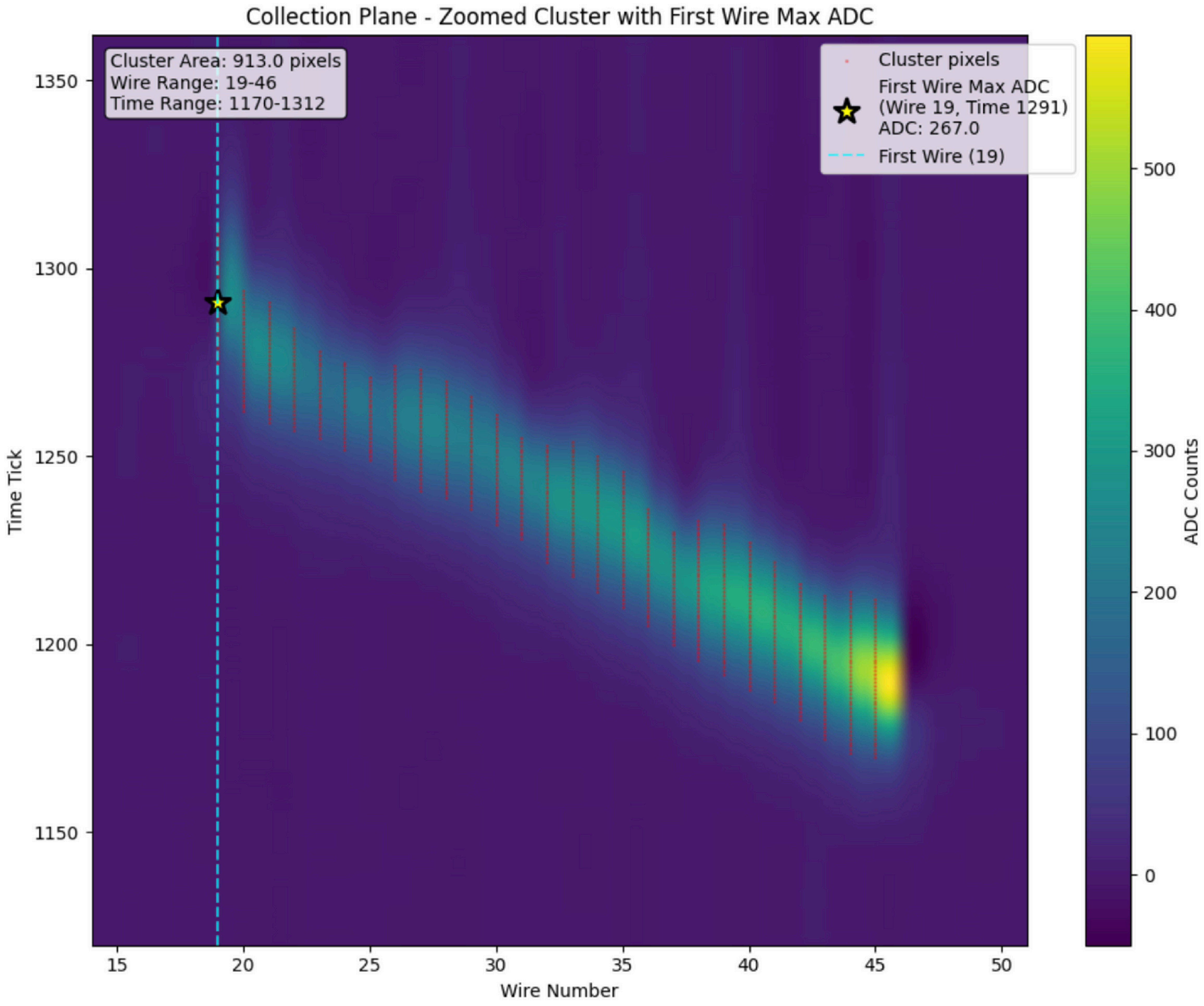
Cluster counts (induction + collection) for deuteron candidates after successive cuts.

Finding track vertices

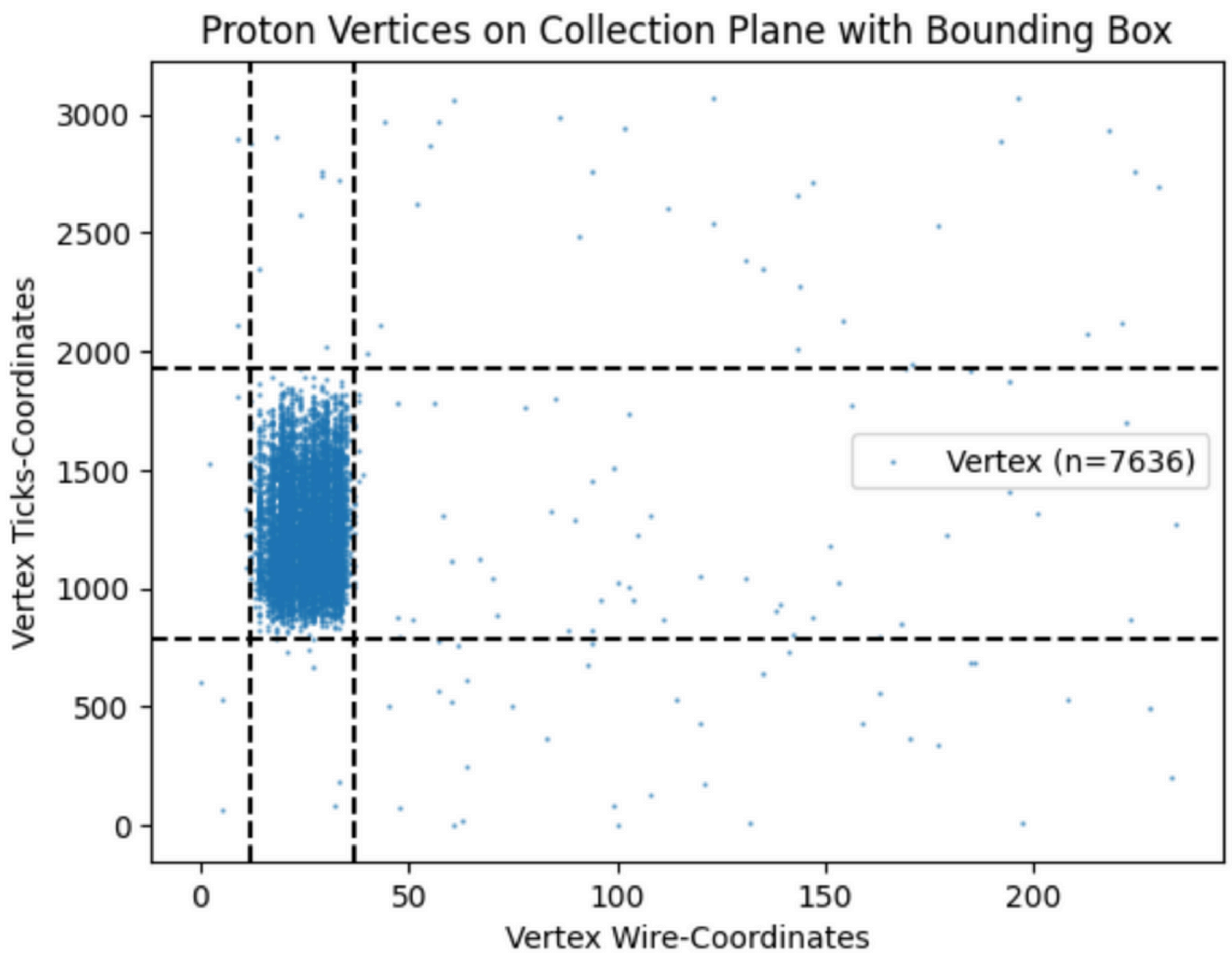
Developed a heuristic to **locate cluster vertices via max ADC on the first wire**.

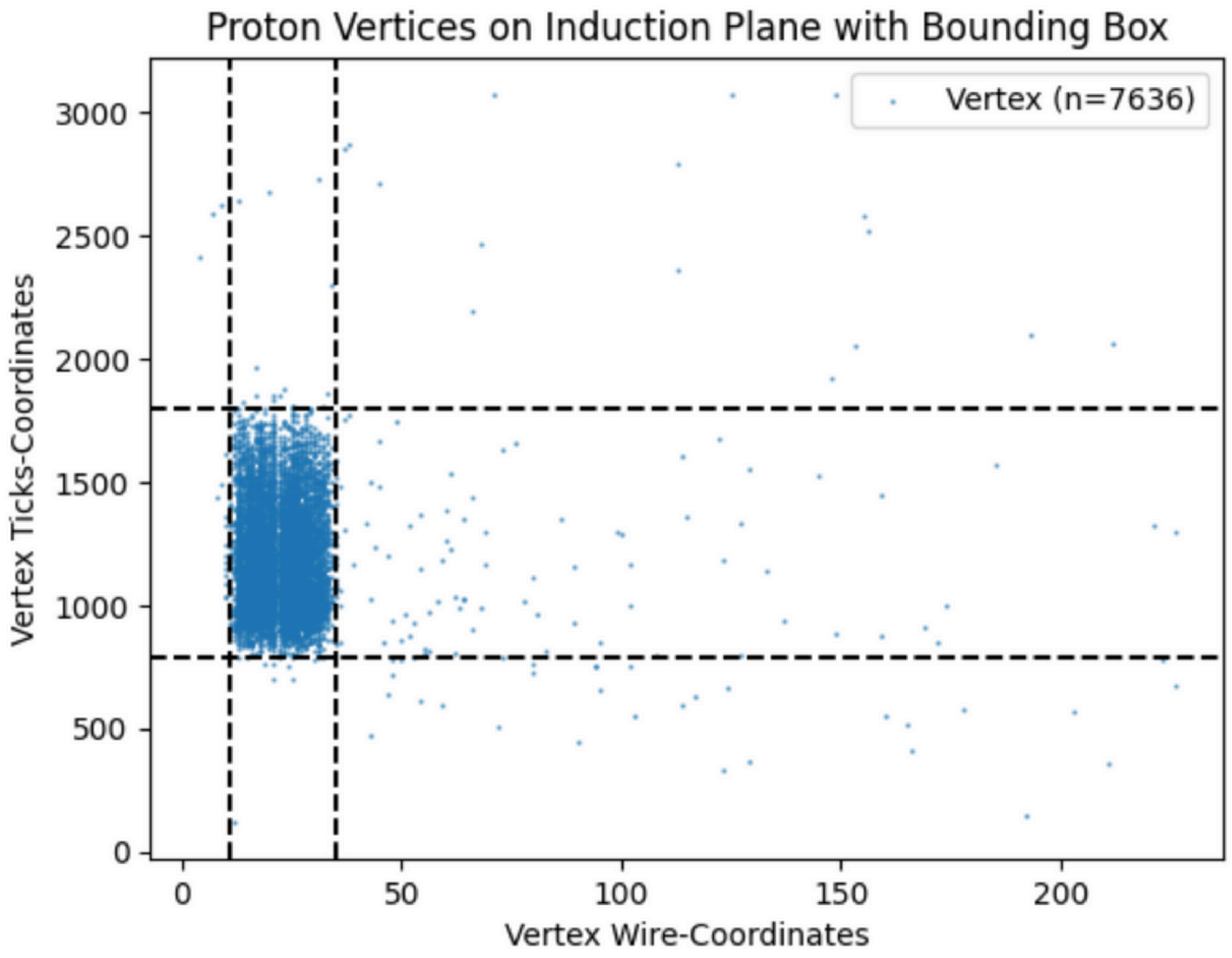
Applied to all proton candidates.

Aimed to identify typical detector regions for clean proton signals to inform cuts on deuteron candidates.



Cleaning Deuteron Selection

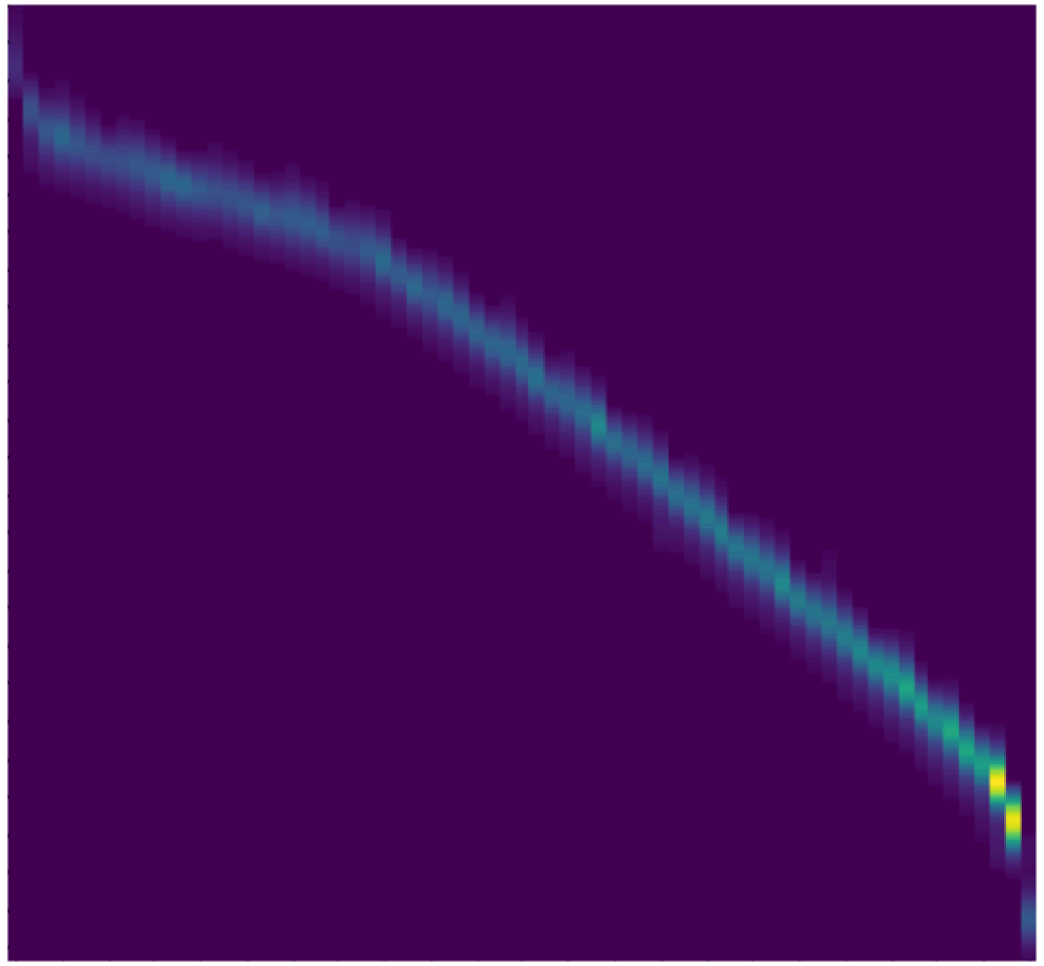




Used this to **cut all deuteron candidates that had no activity within these bounding boxes** in collection or induction (ADC threshold = 100). **Leaving 6961 deuteron candidate events.**

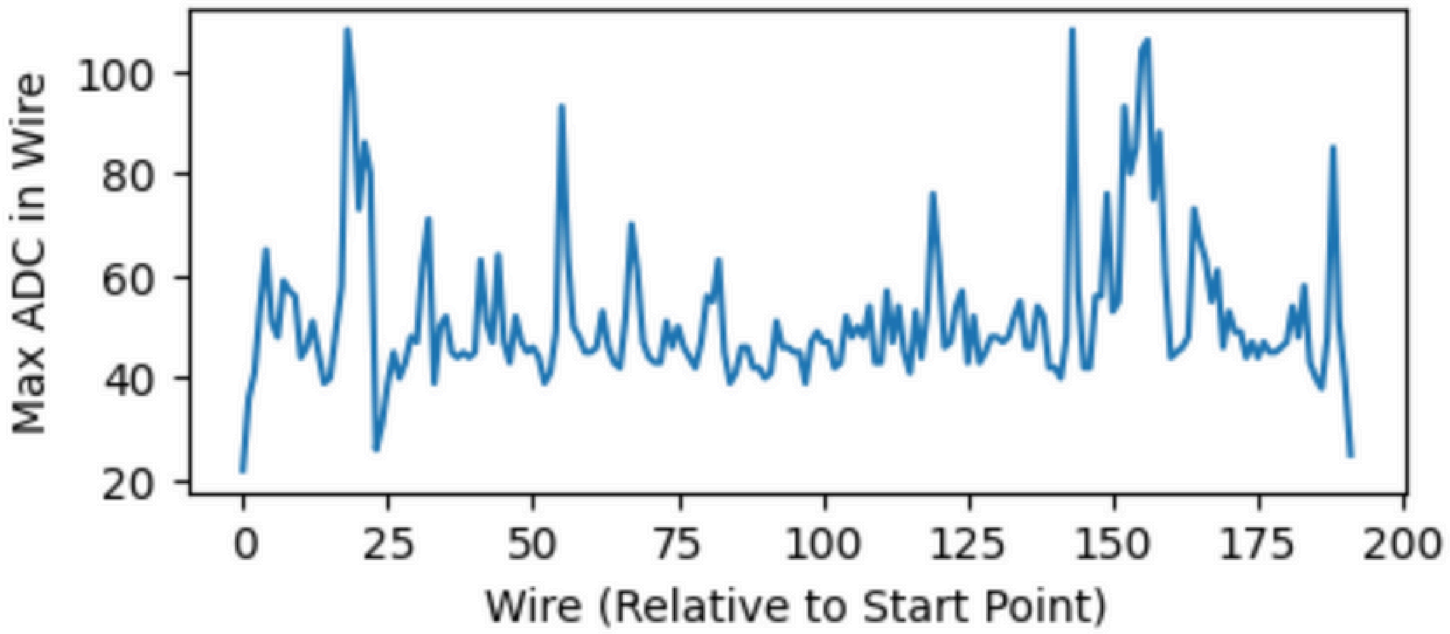
Finding events with signal (p/d).

1. **Ran clustering on all events** (protons + deuterons), and recorded each cluster as a separate matrix.

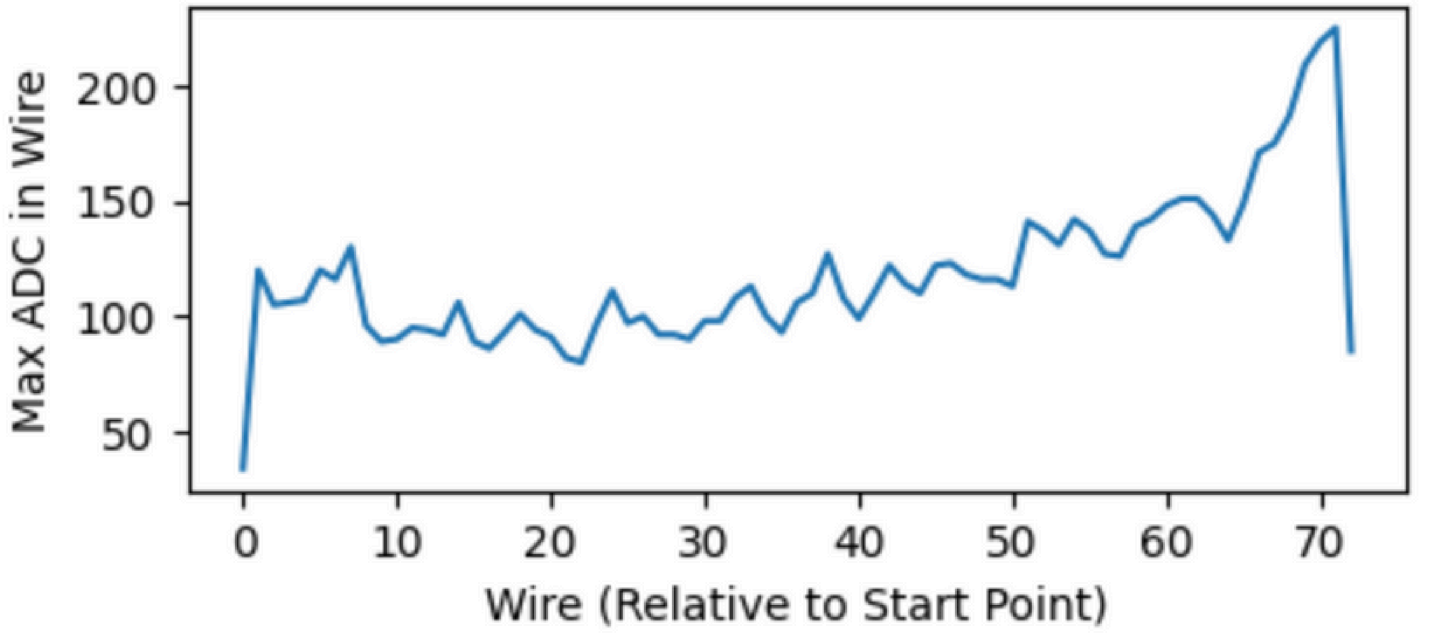


2. For each cluster, **recorded max ADC on each wire to capture trend of energy deposition**.

Potential Muon (fluctuations due to delta rays + low baseline)



Potential Proton (consistent growth + peak at end)



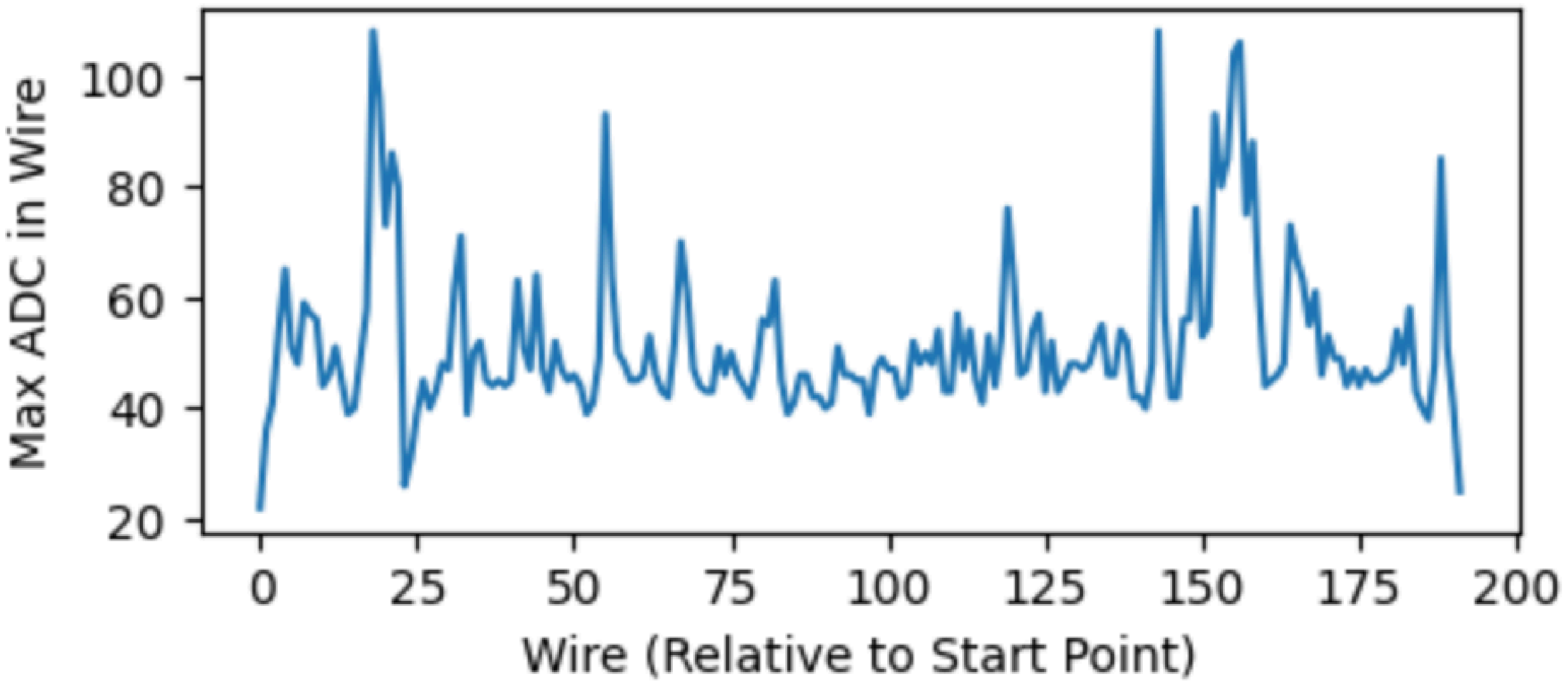
Finding clean deuterons

For each cluster, **computed ADC differences between consecutive "wire max ADC" signals**.

Extracted **the maximum positive difference as a feature**.

Used to distinguish signal clusters (protons/deuterons) from noise (e.g. muons). **Expecting large numbers for noise, and small for signal.**

Potential Muon (fluctuations due to delta rays + low baseline)



Potential Proton (consistent growth + peak at end)