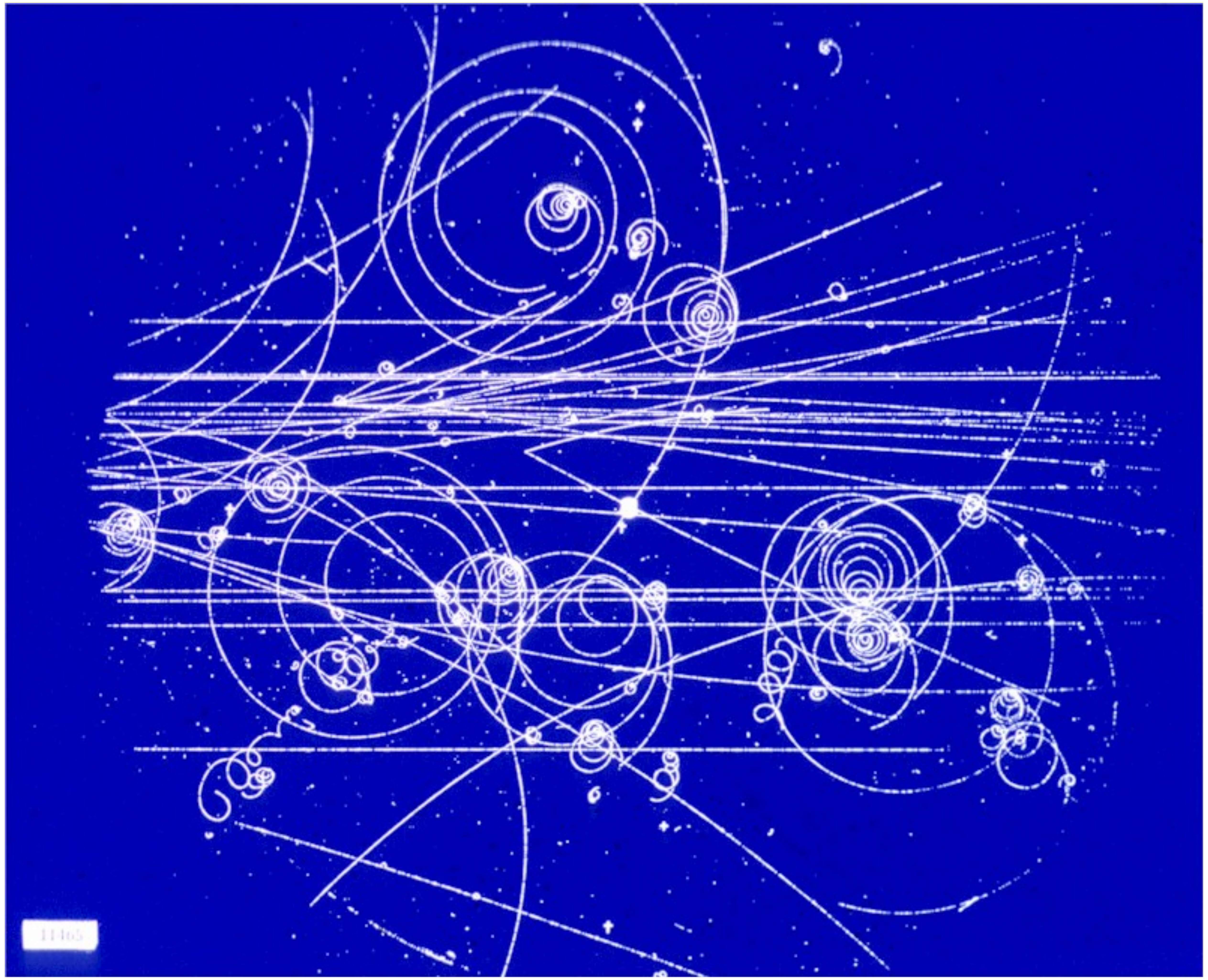
Data processing and analysis

This is where machine learning can help!

D. Bazin
National Superconducting Cyclotron Laboratory
Facility for Rare Isotope Beams
Michigan State University

Bubble chambers

- The first active target detector
  - Chamber filled with superheated liquid
  - Charged particles create local phase transitions that result in the formation of bubbles
  - Pictures of the tracks are taken with a camera
- Tedious analysis!
  - People look at each picture and identify interesting events
  - Later on automated methods were used to help, but still very slow (30 s per photo)



Bubble chambers

- The first active target detector
  - Chamber filled with superheated liquid
  - Charged particles create local phase transitions that result in the formation of bubbles
  - Pictures of the tracks are taken with a camera
- Tedious analysis!
  - People look at each picture and identify interesting events
  - Later on automated methods were used to help, but still very slow (30 s per photo)

Put your brain in a computer…

- Humans have evolved to be excellent at pattern recognition
  - Survival skill: spot the tiger hiding in the forest, the fruits hanging in the trees
  - Human brains can easily extract signal from noise and identify patterns
- Computers have never encountered this type of problem!
  - Algorithms are needed to emulate these cognitive functions in computers
  - Recent years have seen an explosion of pattern recognition methods
  - Face recognition (your cell phone), self-driving cars, etc…
  - Many of these methods can be applied to active target data analysis

Analysis phases

- Three main tasks
  - Cleaning (removing noise or spurious signals)
  - Identification (identifying the type of event)
  - Fitting (extracting physics parameters from data)
- Each phase present its own set of challenges
  - Cleaning: finding criteria that cleans effectively without removing good data
  - Identifying: anticipating classes of events (based on simulations)
  - Fitting: find robust methods and avoid local minima
- Machine learning techniques can help in all three phases…

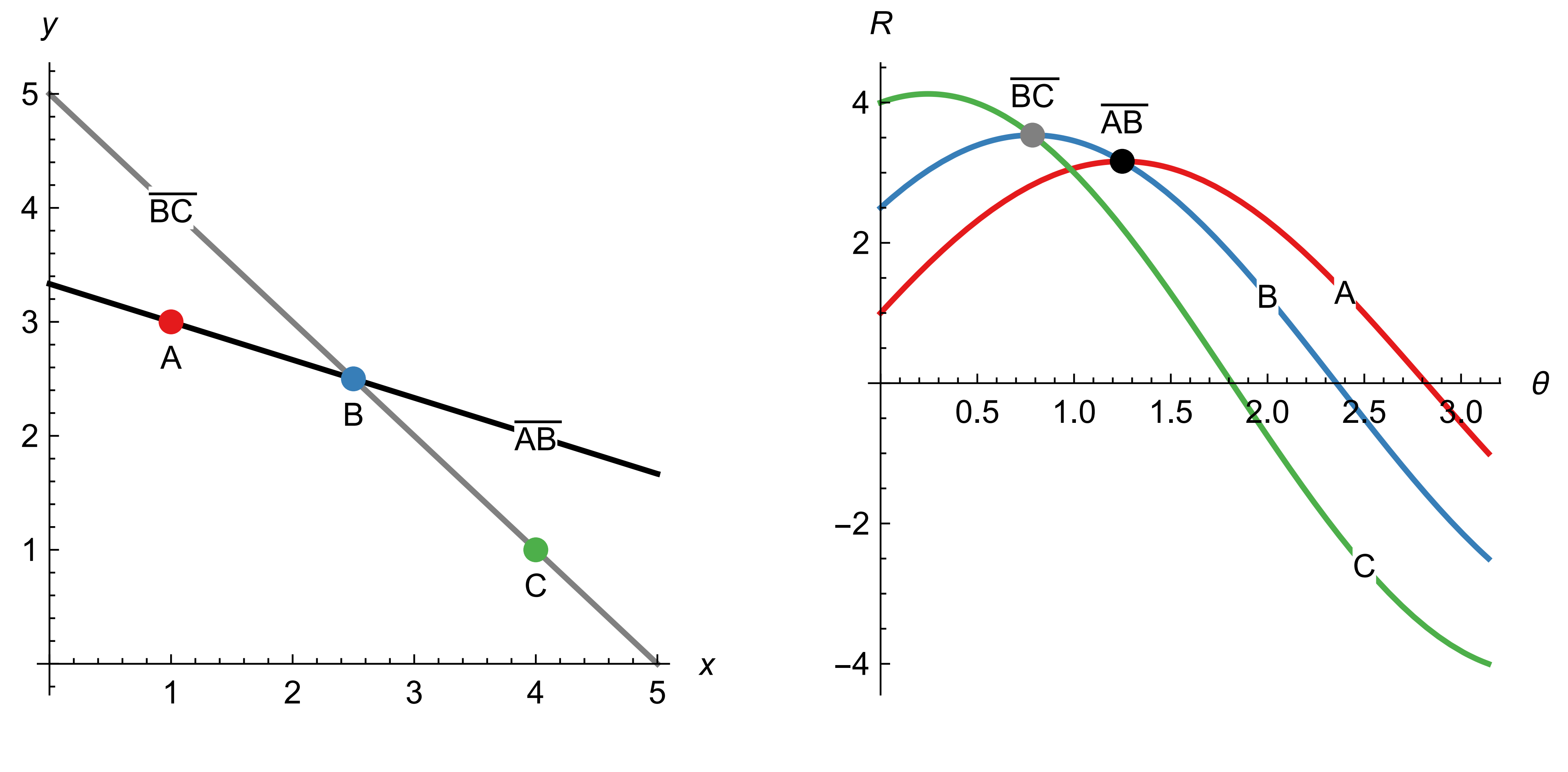
Hough transform: the eyes of a computer...

- Linear Hough transformation to find points along straight lines

$$R = x\cos(\theta) + y\sin(\theta)$$

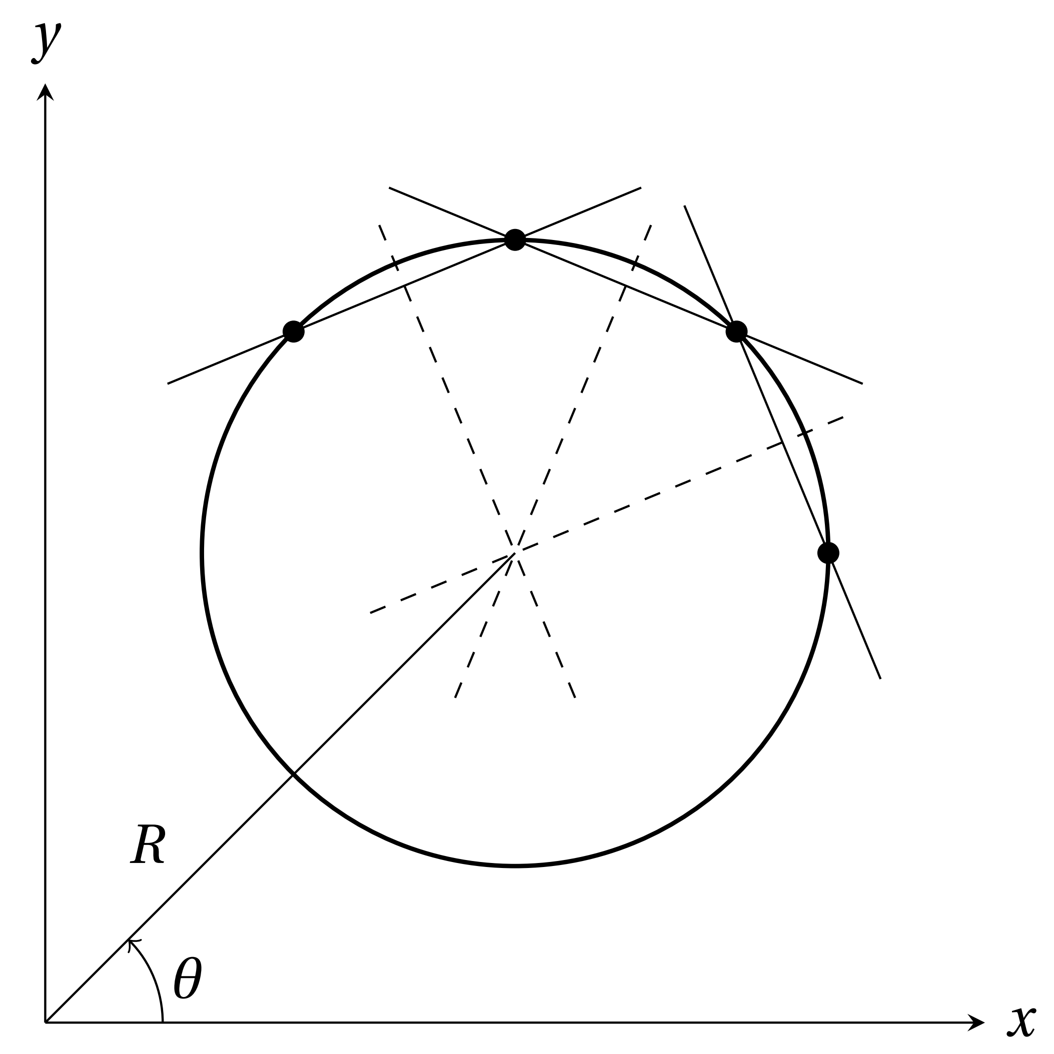
- Circular Hough transformation to find points on a circle

$$R = \frac{(x_1^2 - x_0^2) + (y_1^2 - y_0^2)}{2[(x_1 - x_0)\cos(\theta) + (y_1 - y_0)\sin(\theta)]}$$

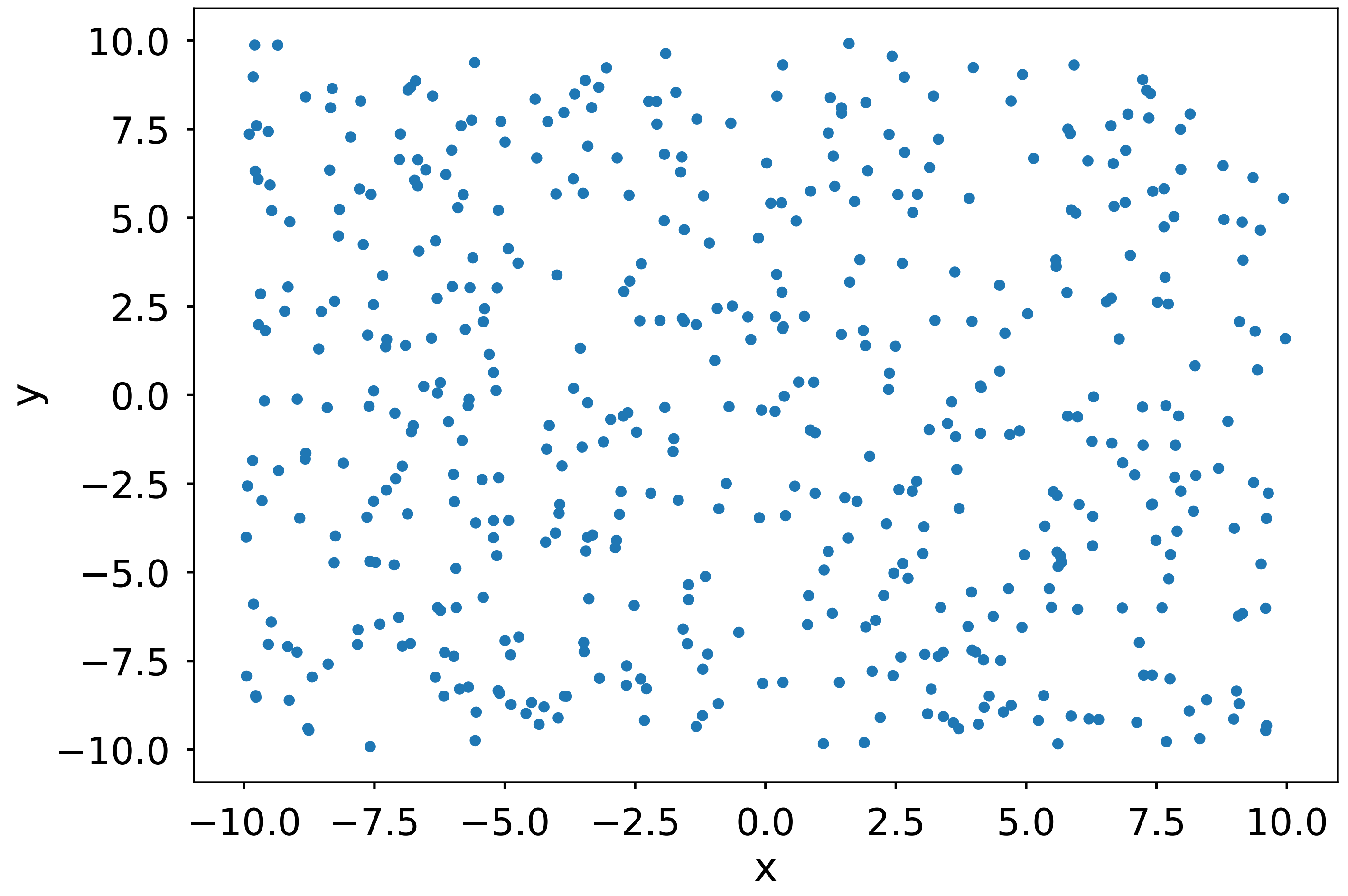


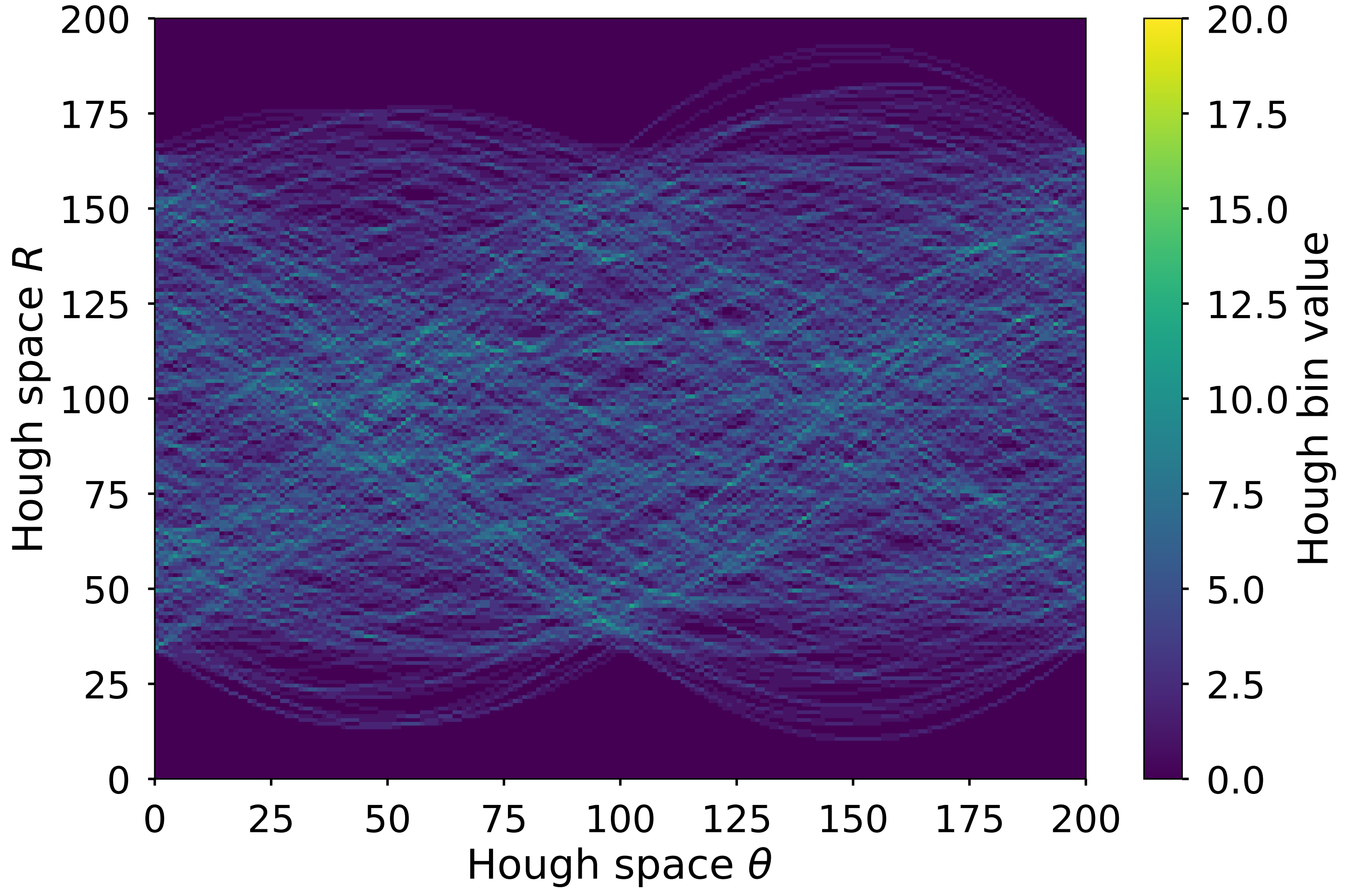
(a) Coordinate space

(b) Hough space

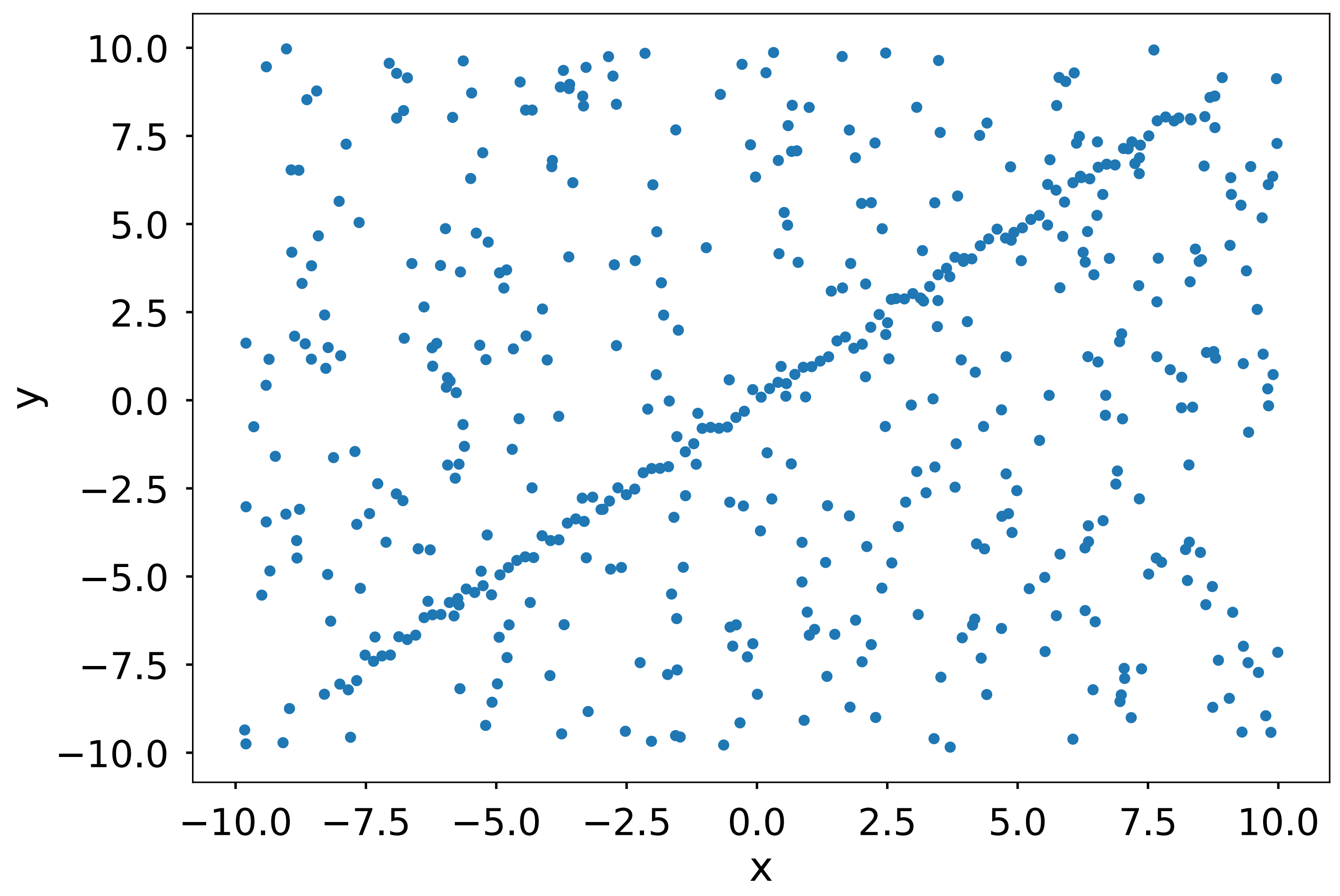


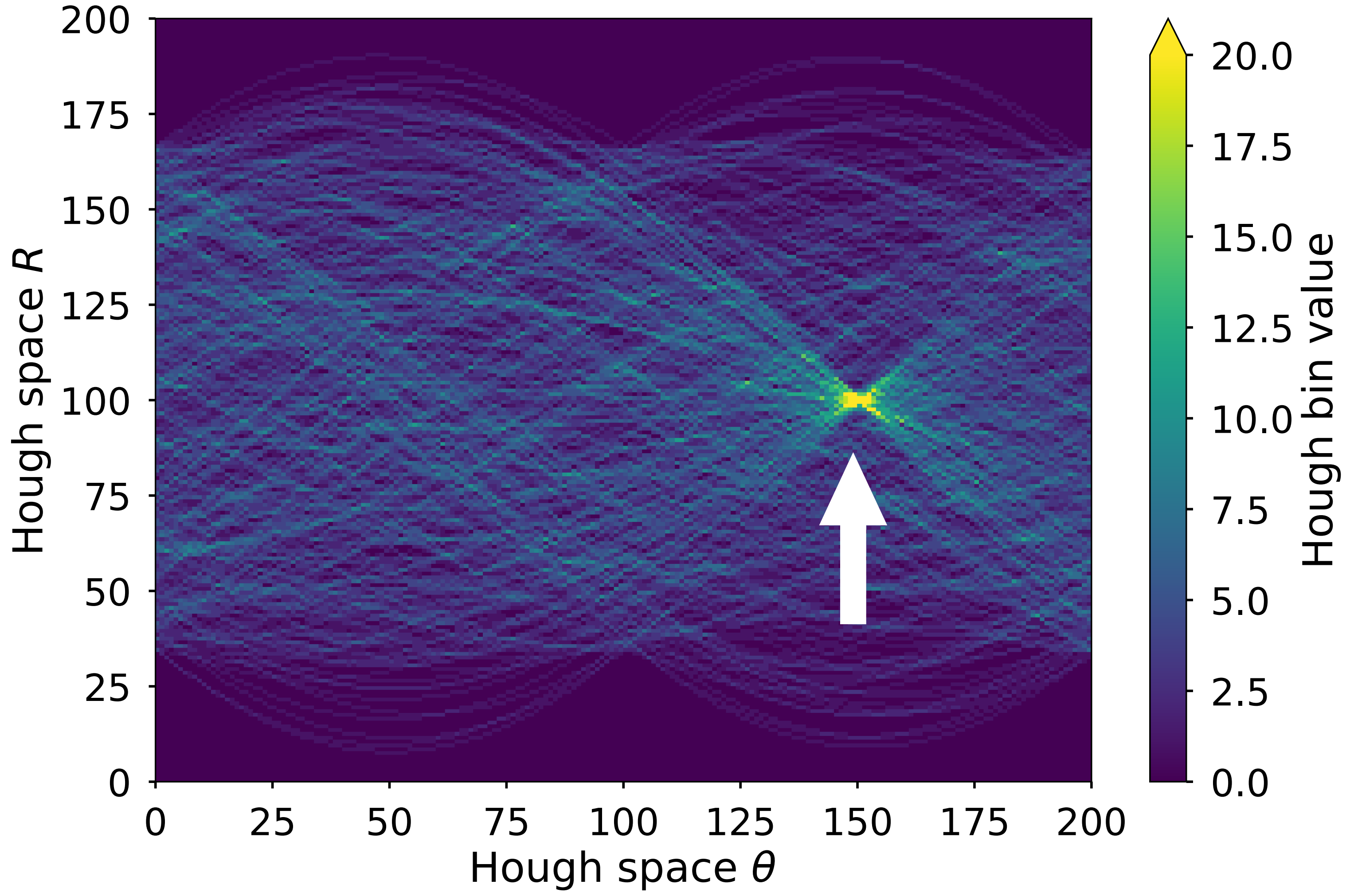
Linear Hough transform





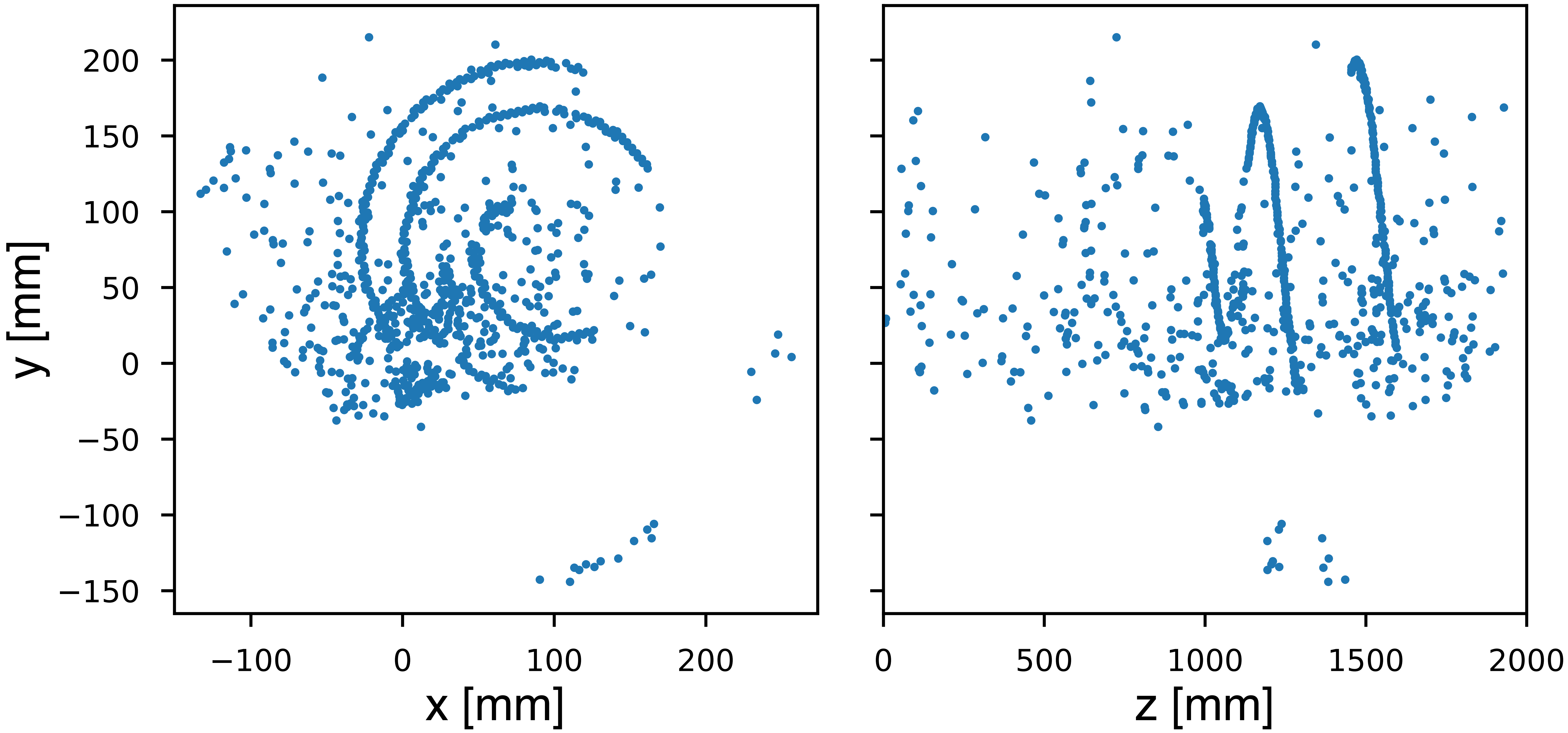
$$R = x\cos(\theta) + y\sin(\theta)$$



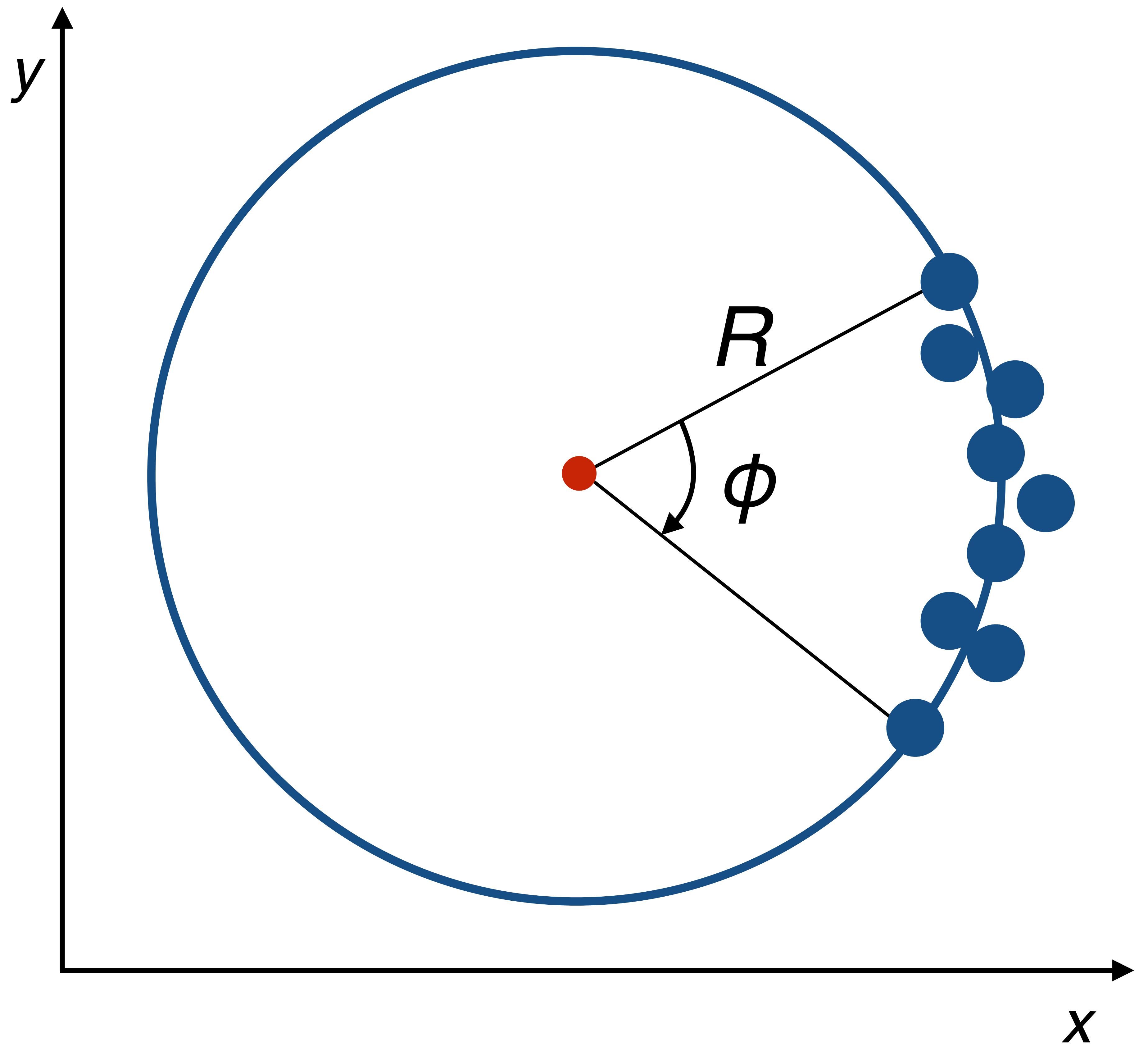


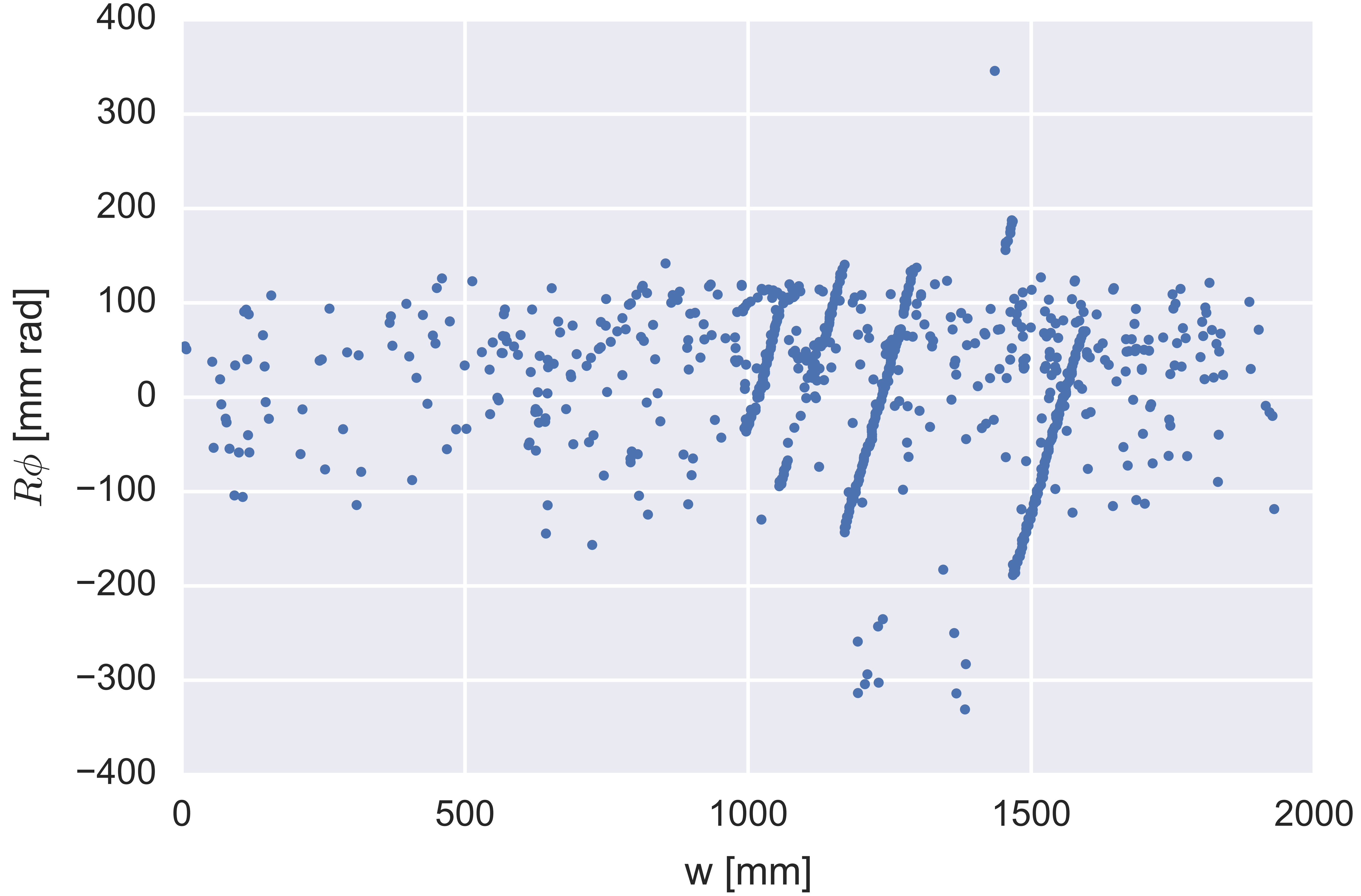
Example on AT-TPC data

- Very noisy and incomplete data!
- Use circular Hough to identify rough center of spiral
- Transform data into RΦ coordinates
- Use linear Hough to identify lines



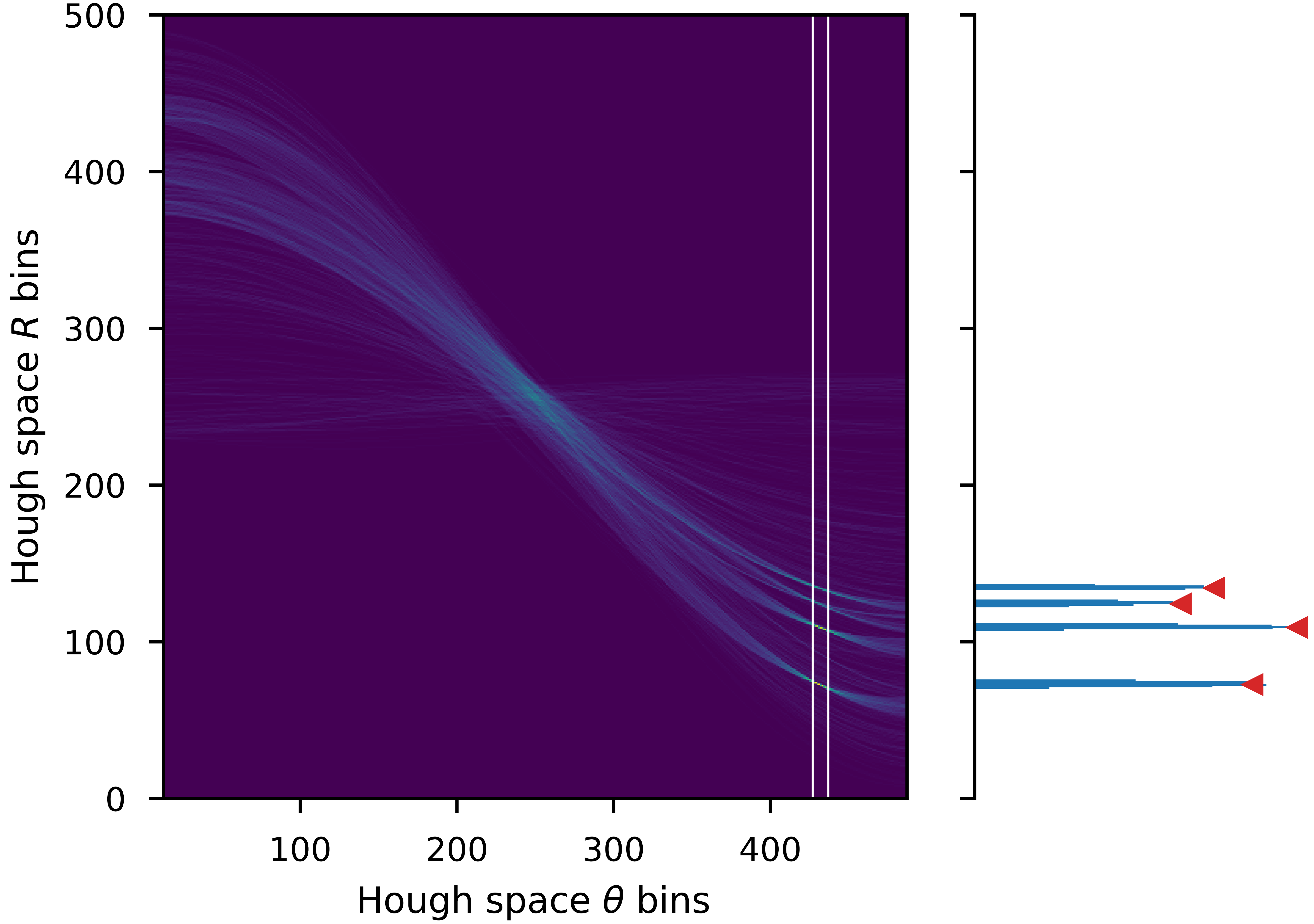
Cleaning AT-TPC data

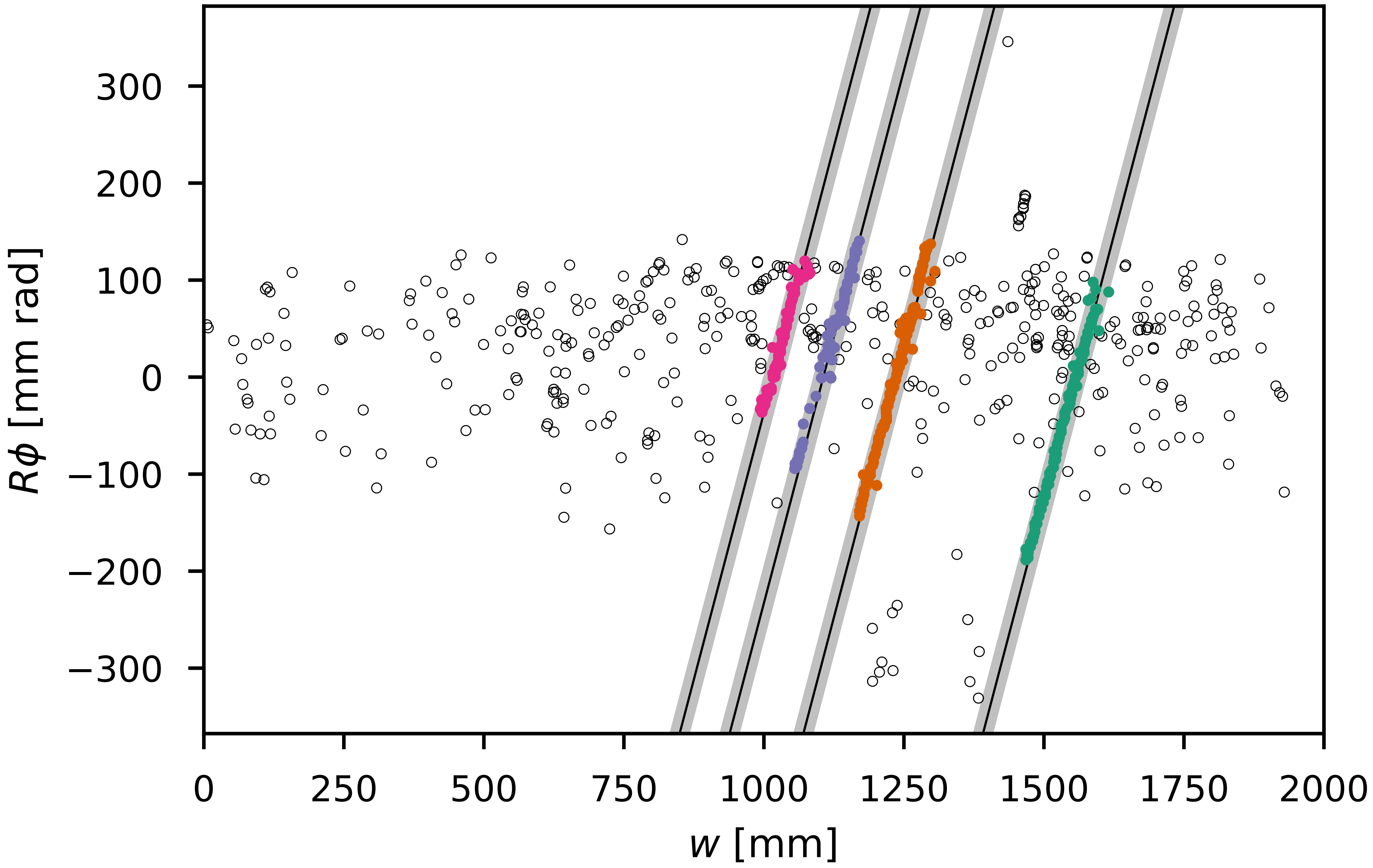




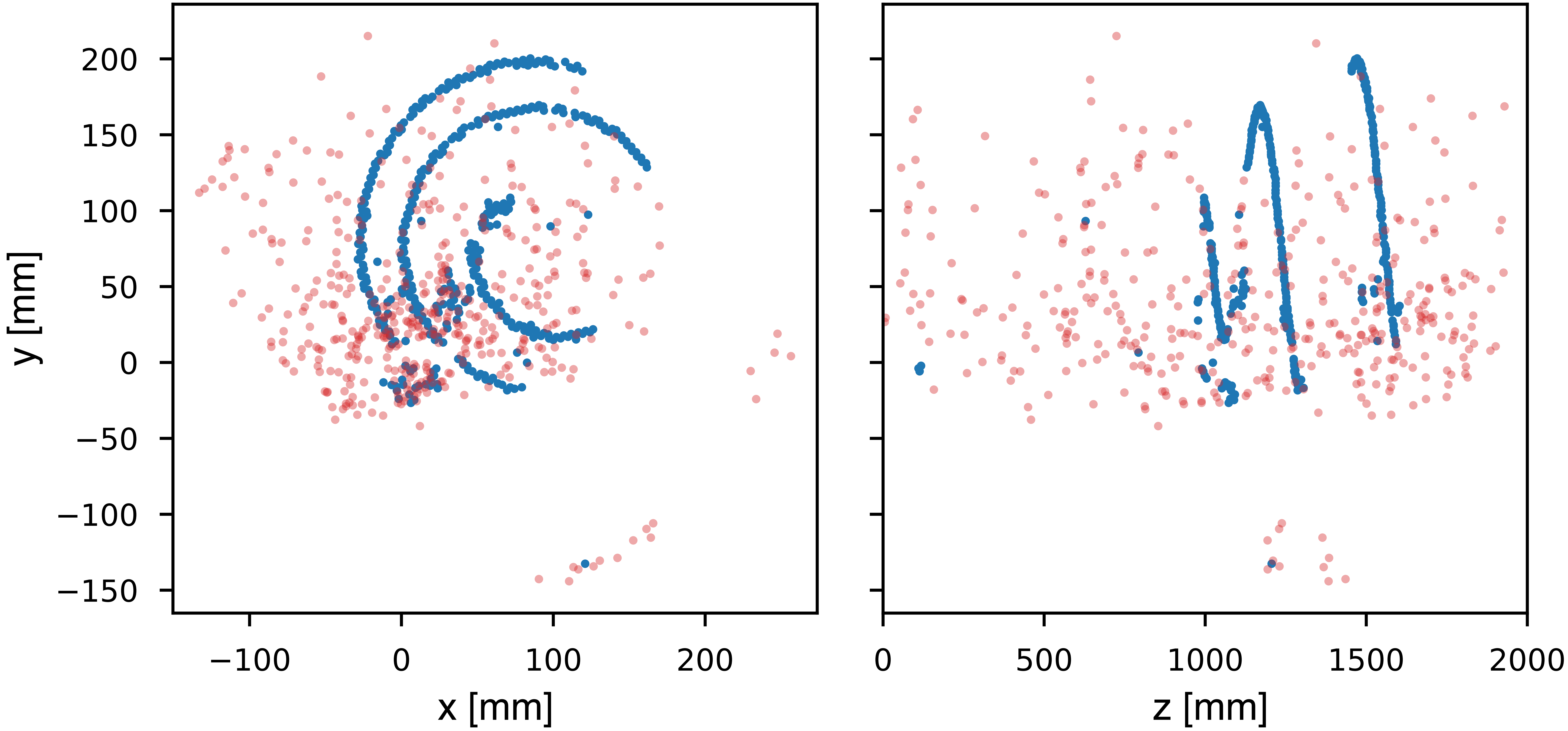
$$w = R\phi$$

Cleaning AT-TPC data





Cleaning AT-TPC data

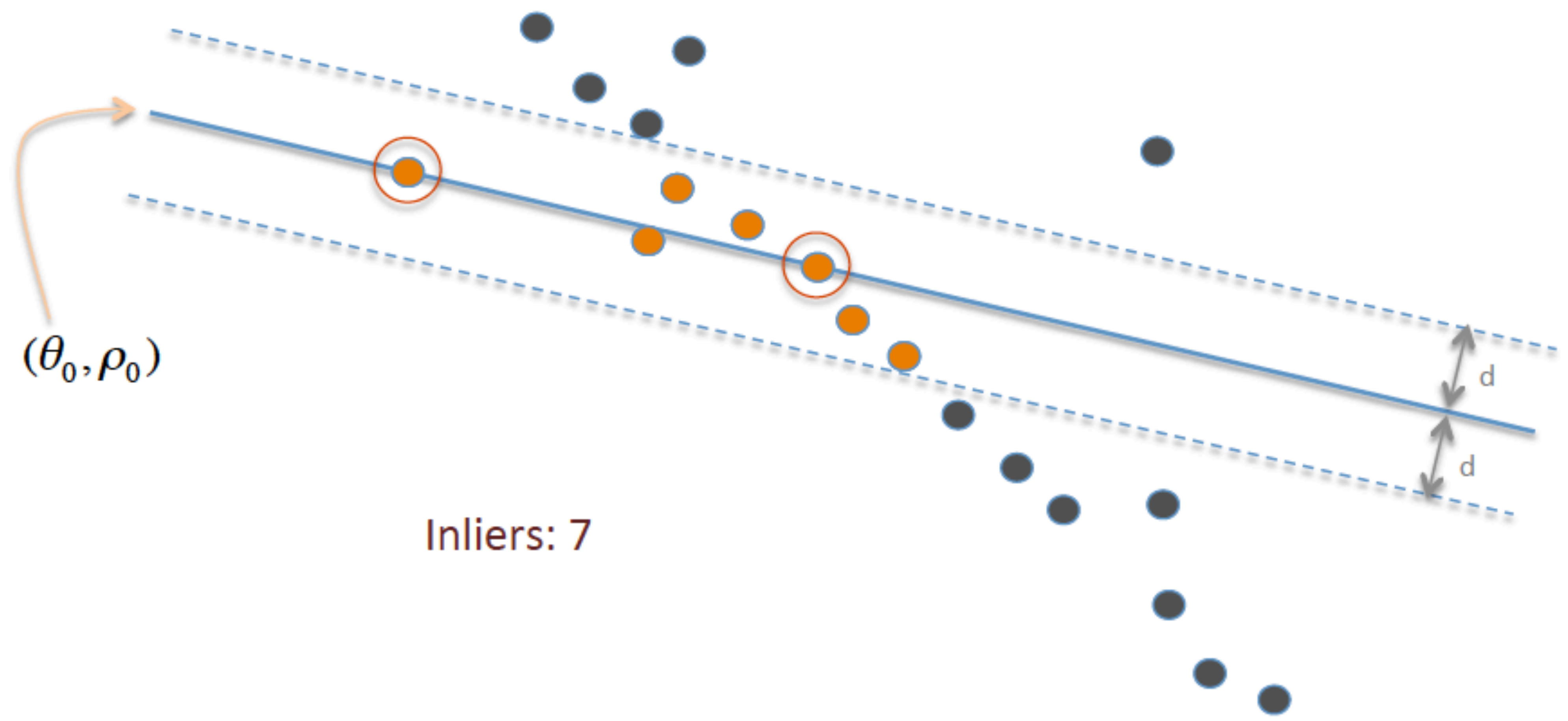


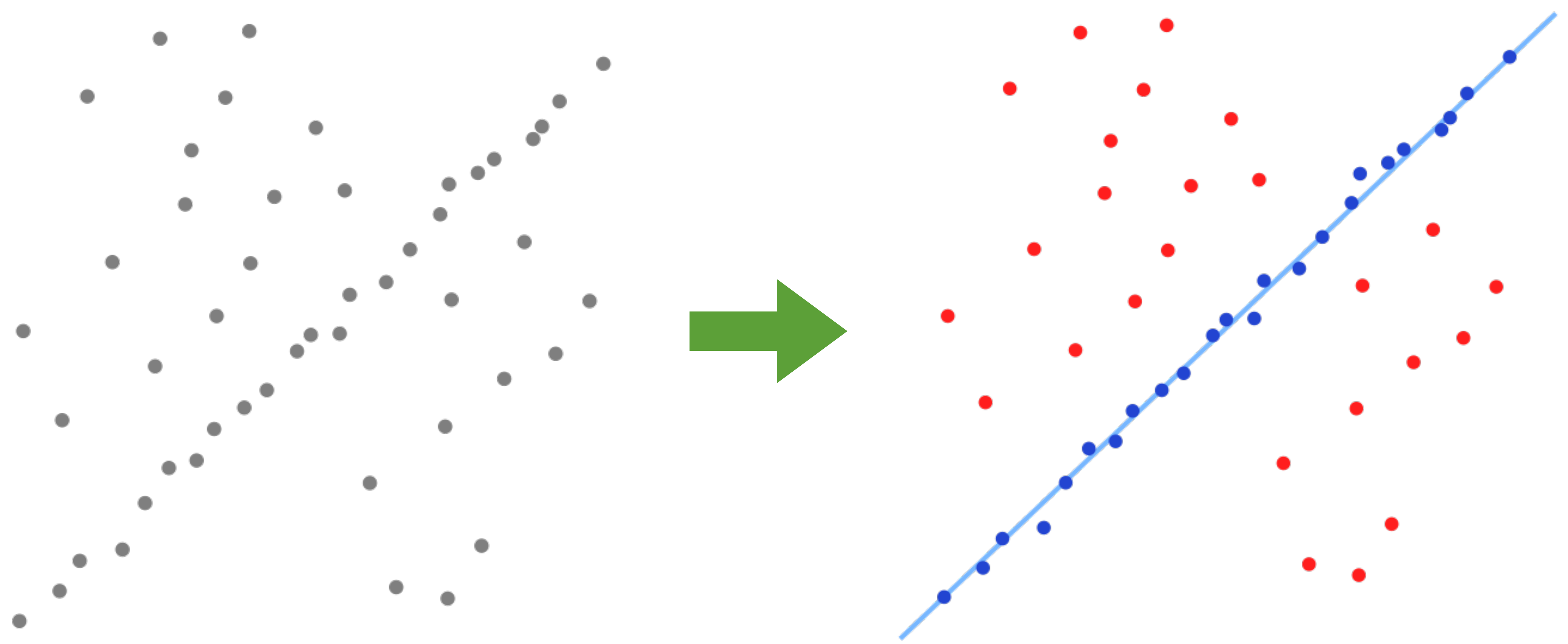
J. Bradt, Ph. D. Dissertation, 2017

https://publications.nscl.msu.edu/thesis/%20Brandt_2017_5279.pdf

RANSAC method

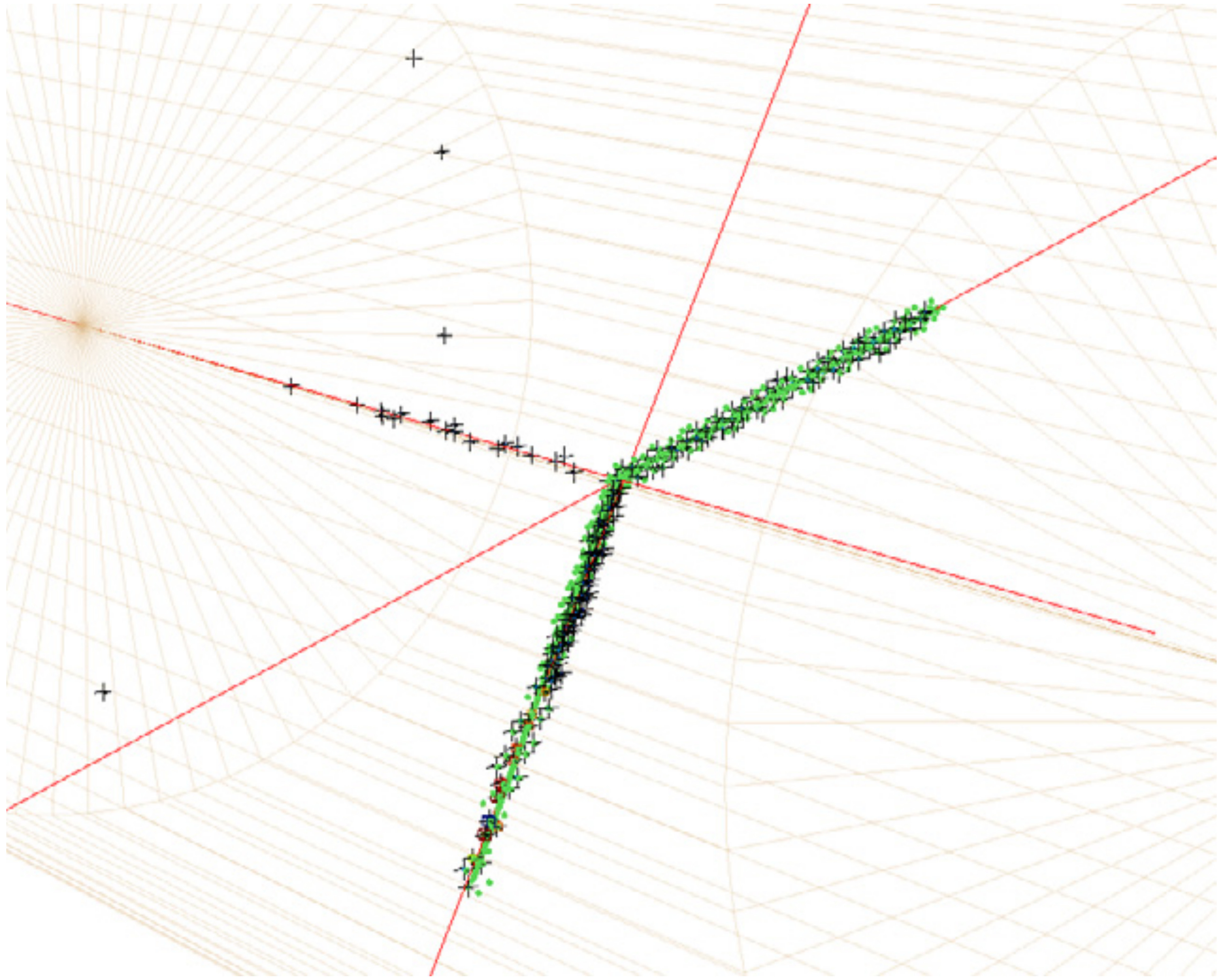
- Random Sample Consensus
  - Select random sample from data
  - Fit model to this subset
  - Test fit on other data than subset
  - Data that fits well is part of consensus
  - Repeat and keep best fit
- Parameters
  - Maximum distance from fit for inliers
  - Minimum number of points for good fit





RANSAC method

- Advantages
  - Can find tracks in 3D
  - Once best fit is found, eliminate inliers from data set and run again
  - Can find multiple tracks
  - Fast algorithm (at least for straight tracks)
- Drawbacks
  - Requires parameter adjustment
  - No convergence criteria

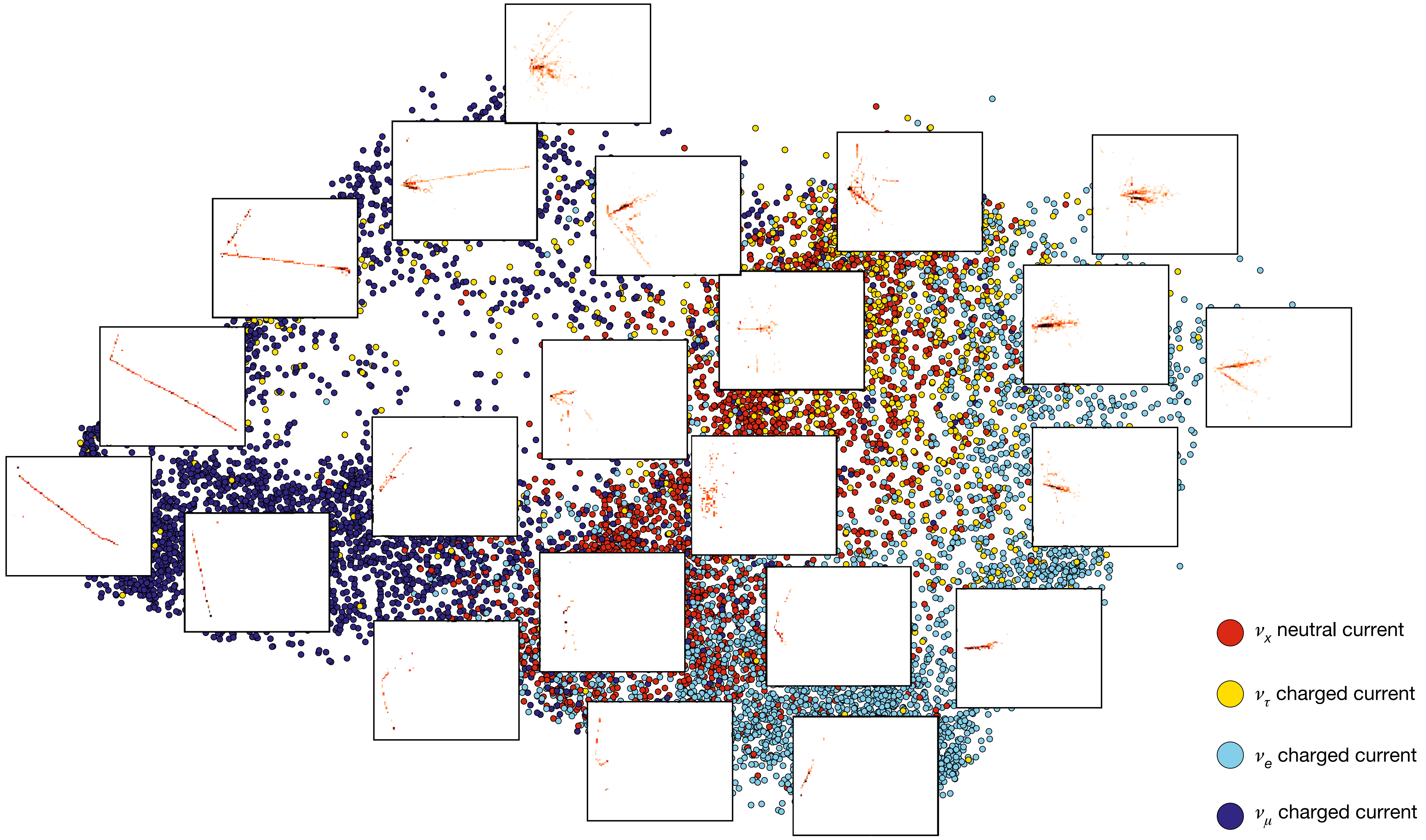


^{4}He + ^{4}He elastic scattering in AT-TPC

Y. Ayyad et al., NIM A880, 166 (2018)

Event identification

- Determine classes of events
- Not all events are interesting
- Categorize events based on their data
- Unsupervised learning: discover new classes of events
- Ultimately: make trigger decision during data taking
- This is already being done in high energy physics!

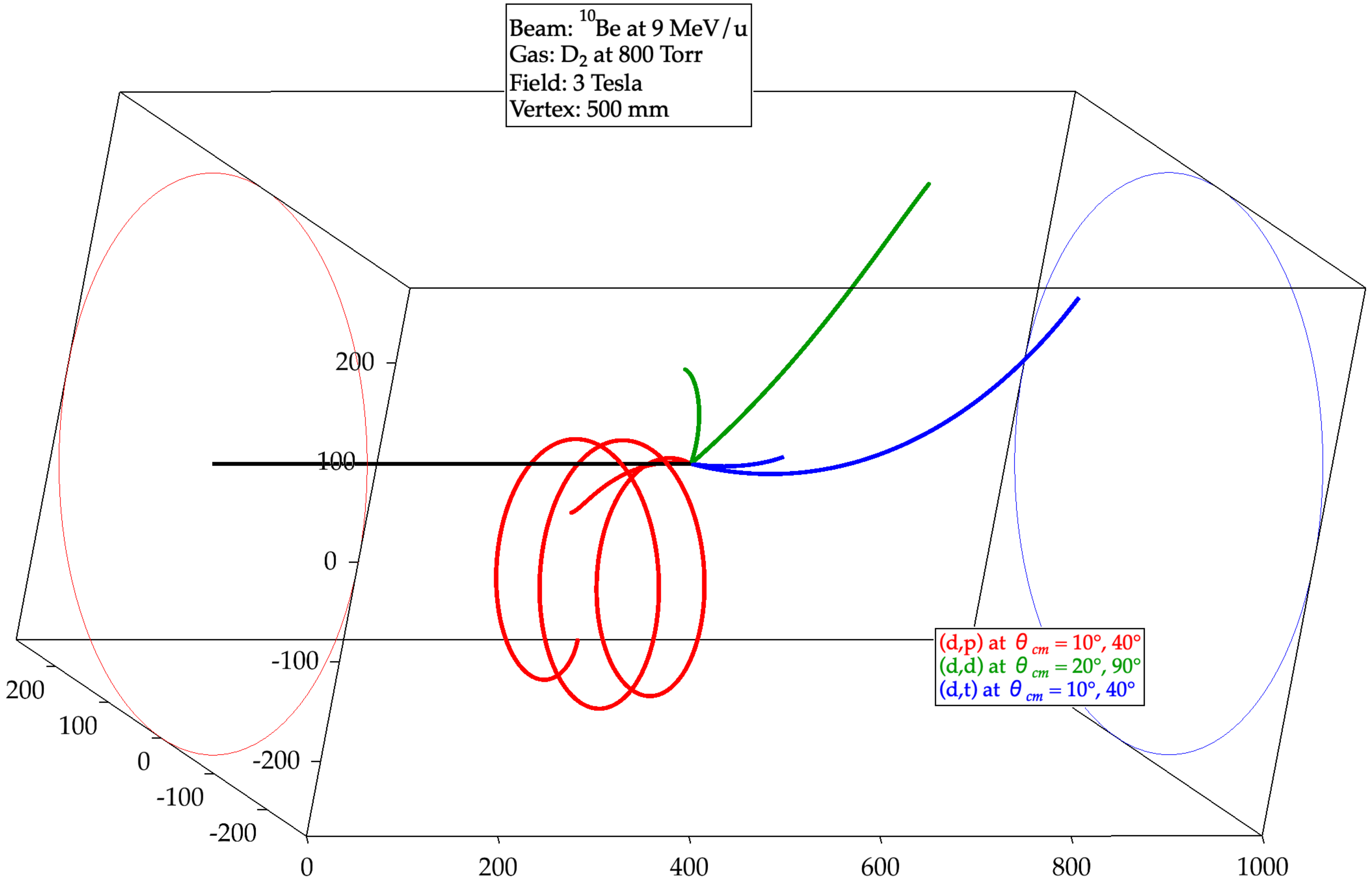


Training data for CNN analysis of NOvA

A. Radovic et al., Nature 560, 41–48 (2018)

A simple approach...

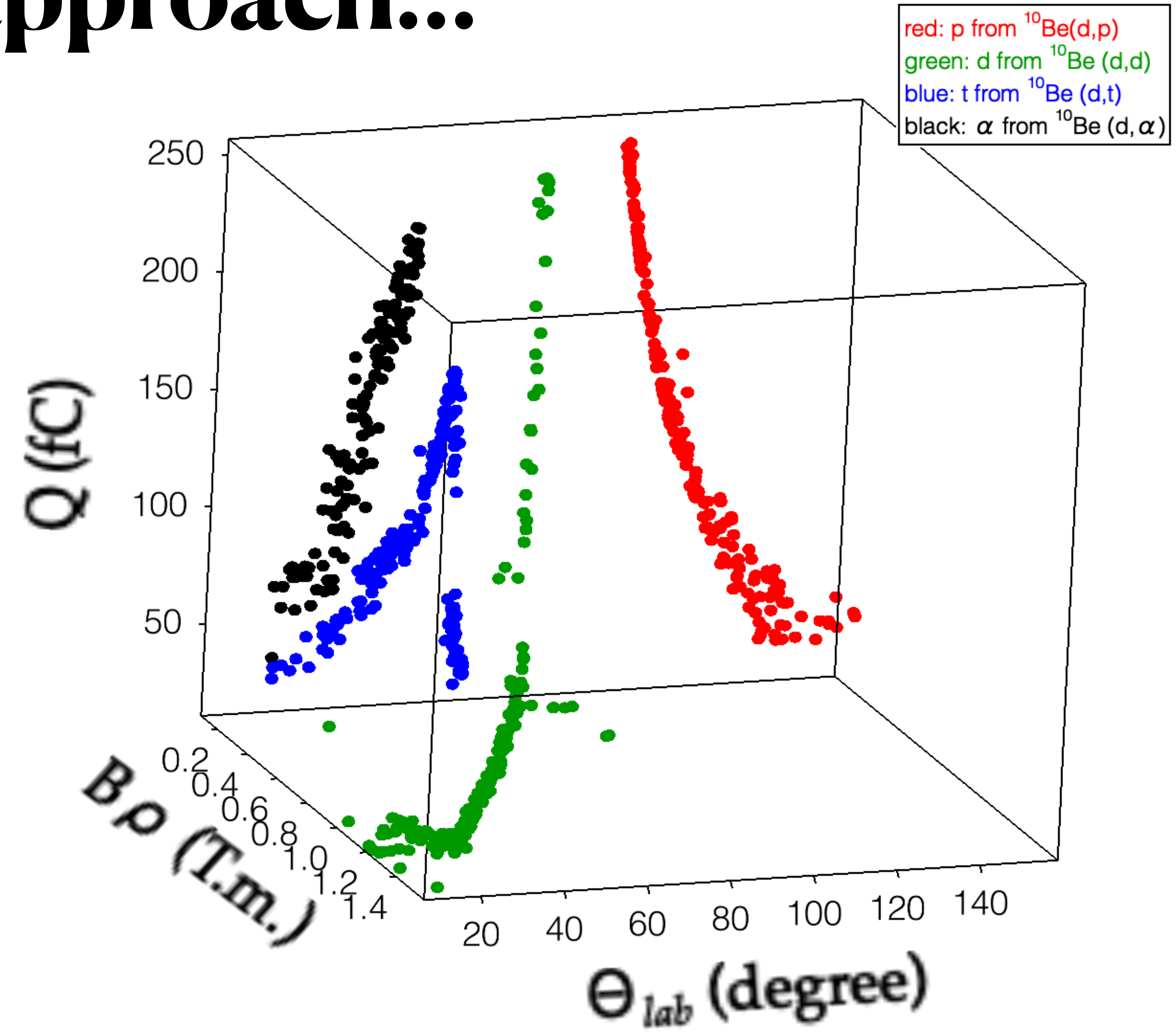
- Different reactions
  - Different recoil particles
  - Different kinematics
- Consider ^{10}Be beam on deuterium target
  - ^{10}Be(d,d'): recoiling d
  - ^{10}Be(d,p): recoiling p
  - ^{10}Be(d,t): recoiling t
  - ^{10}Be(d,α): recoiling α
- Event identification based on 3 parameters extracted from tracks



Simulation

A simple approach...

- Different reactions
  - Different recoil particles
  - Different kinematics
- Consider ^{10}Be beam on deuterium target
  - ^{10}Be(d,d'): recoiling d
  - ^{10}Be(d,p): recoiling p
  - ^{10}Be(d,t): recoiling t
  - ^{10}Be(d,α): recoiling α
- Event identification based on 3 parameters extracted from tracks

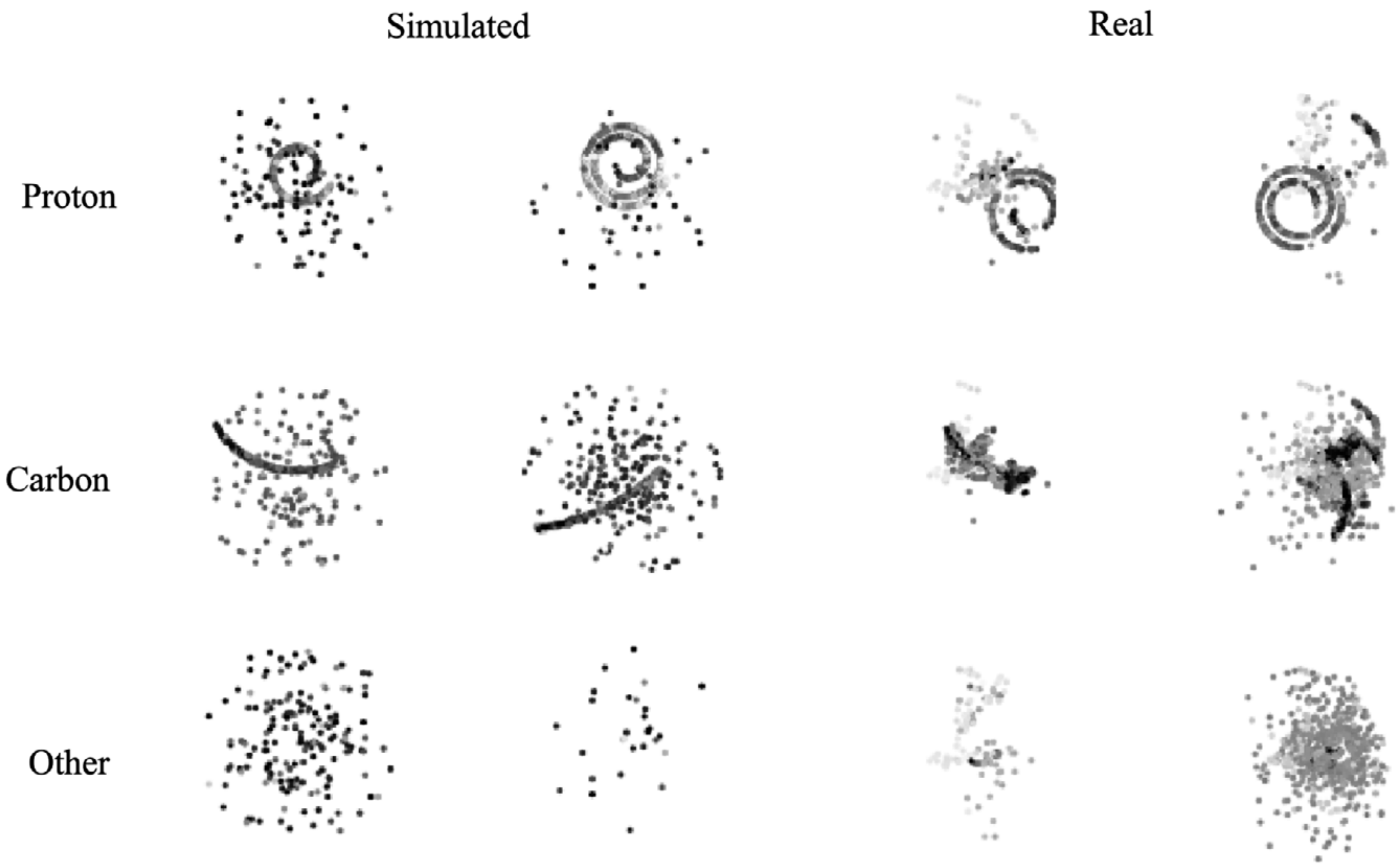


Simulation

A machine learning approach…

Tested on AT-TPC data

- ^{46}Ar scattering on C_4H_{10}
- Recoil can either be p of ^{12}C
- Both types of events are observed in the AT-TPC
- Neural networks are trained on simulated data only
- Downsampling necessary to reduce number of inputs



M. P. Kuchera et al., NIM A940, 156 (2019)

Encouraging results

- Different ML algorithms
  - LR: Logistic Regression
  - FCNN: Fully-connected Neural Network
  - CNN: Convolution Neural Network
- Different training and testing sets
  - Data labeling is a painful task (think about bubble chamber times!)
  - CNN based on VGG16 seems to perform the best

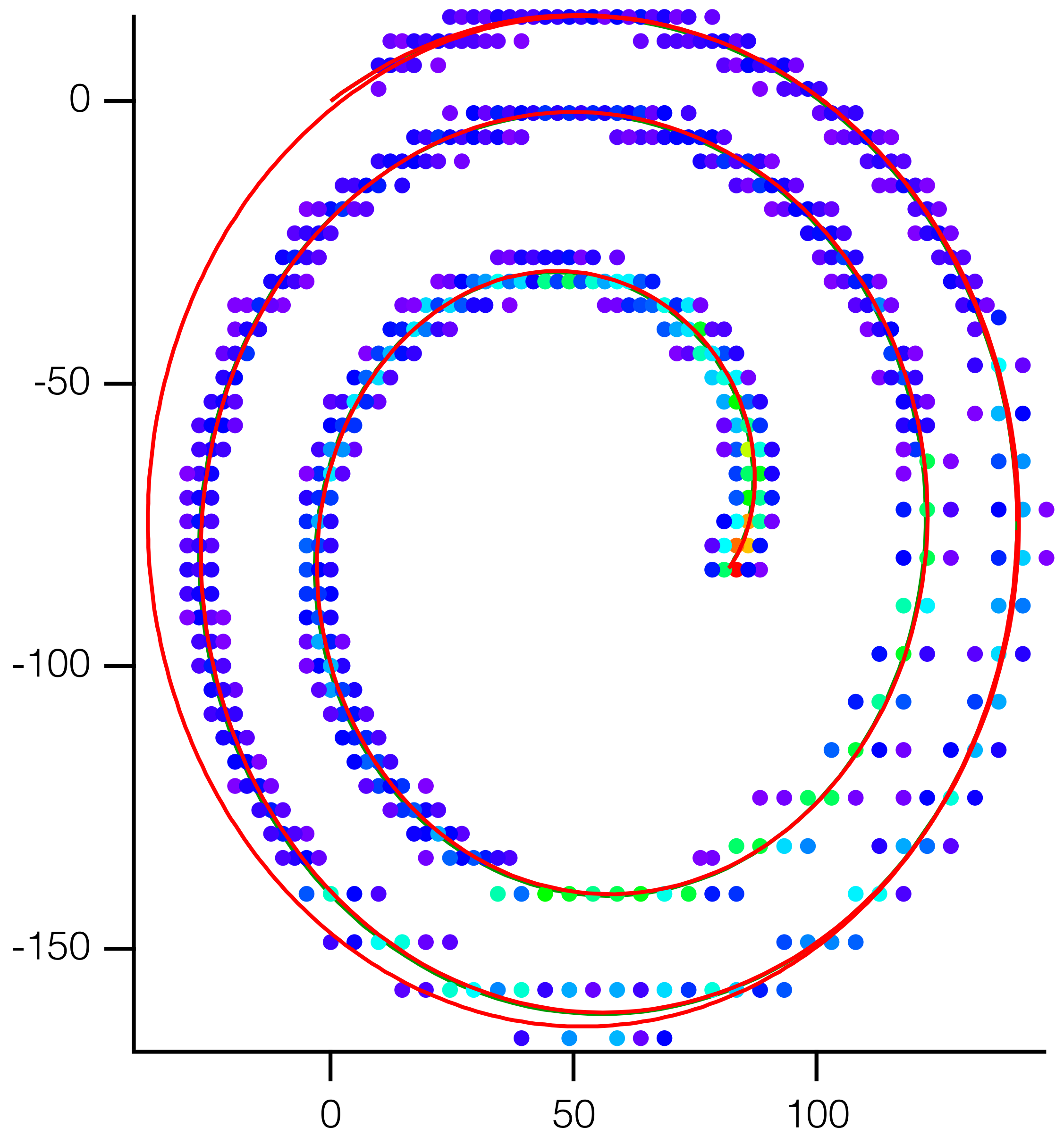
Table 4
Results in the exp → exp regime: these results were obtained by training and testing on hand-labeled data. The train and test sets came from a small dataset of fewer than 3000 events as reported in Table 1.

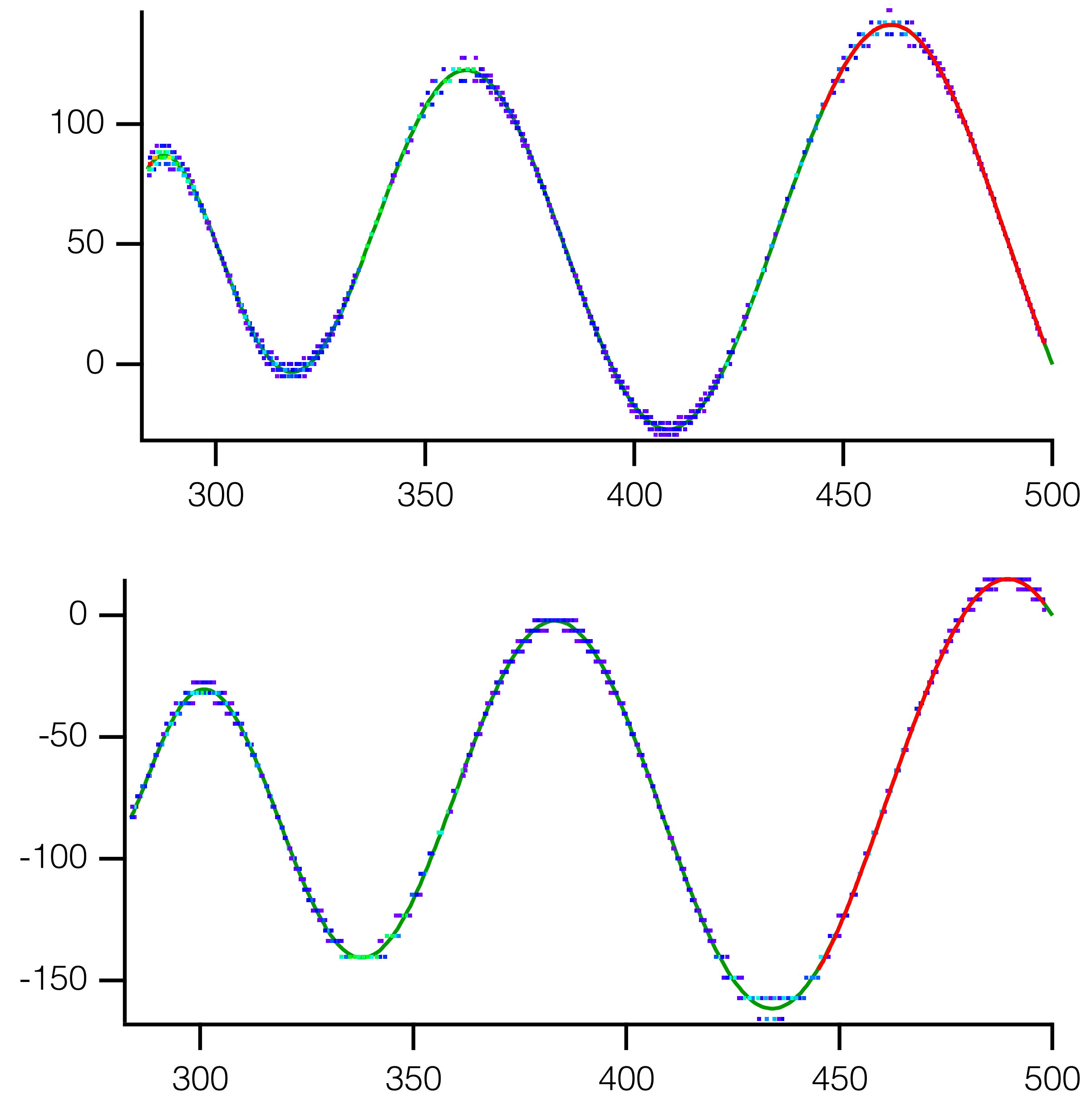
Algorithm	Learning task	Precision	Recall	F1
LR	Binary	0.78	0.58	0.66
	Multiclass	0.77	0.67	0.72
FCNN	Binary	0.85	0.54	0.66
	Multiclass	0.83	0.62	0.71
CNN	Binary	0.98	0.84	0.90
	Multiclass	**0.96**	**0.90**	**0.93**

M. P. Kuchera et al., NIM A940, 156 (2019)

Fitting

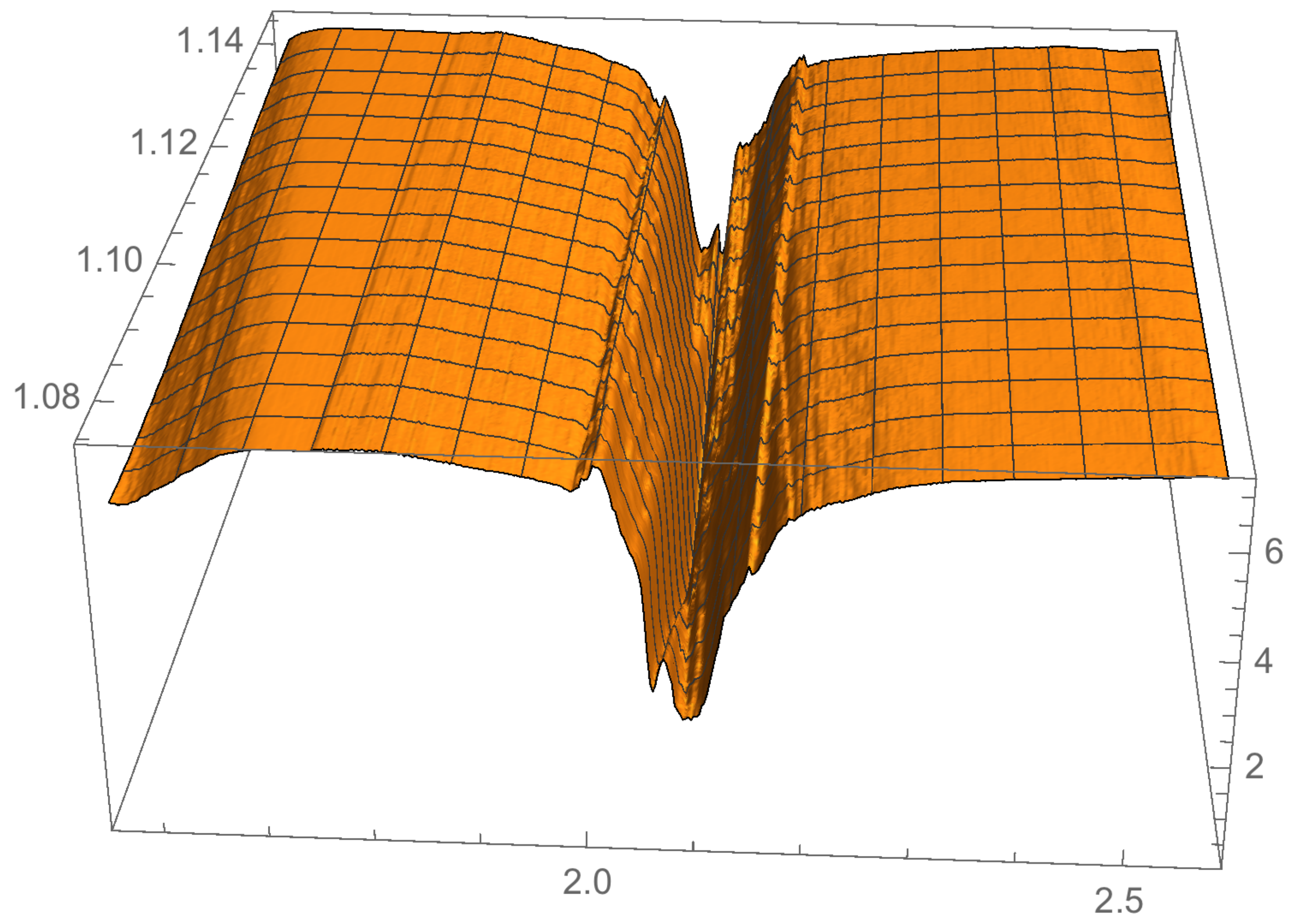
- Extract physical parameters of particles from their tracks and detector parameters
- Complex relationship between physical parameters and track data
- Can only be simulated numerically
- High accuracy needed because of demanding inverse kinematics

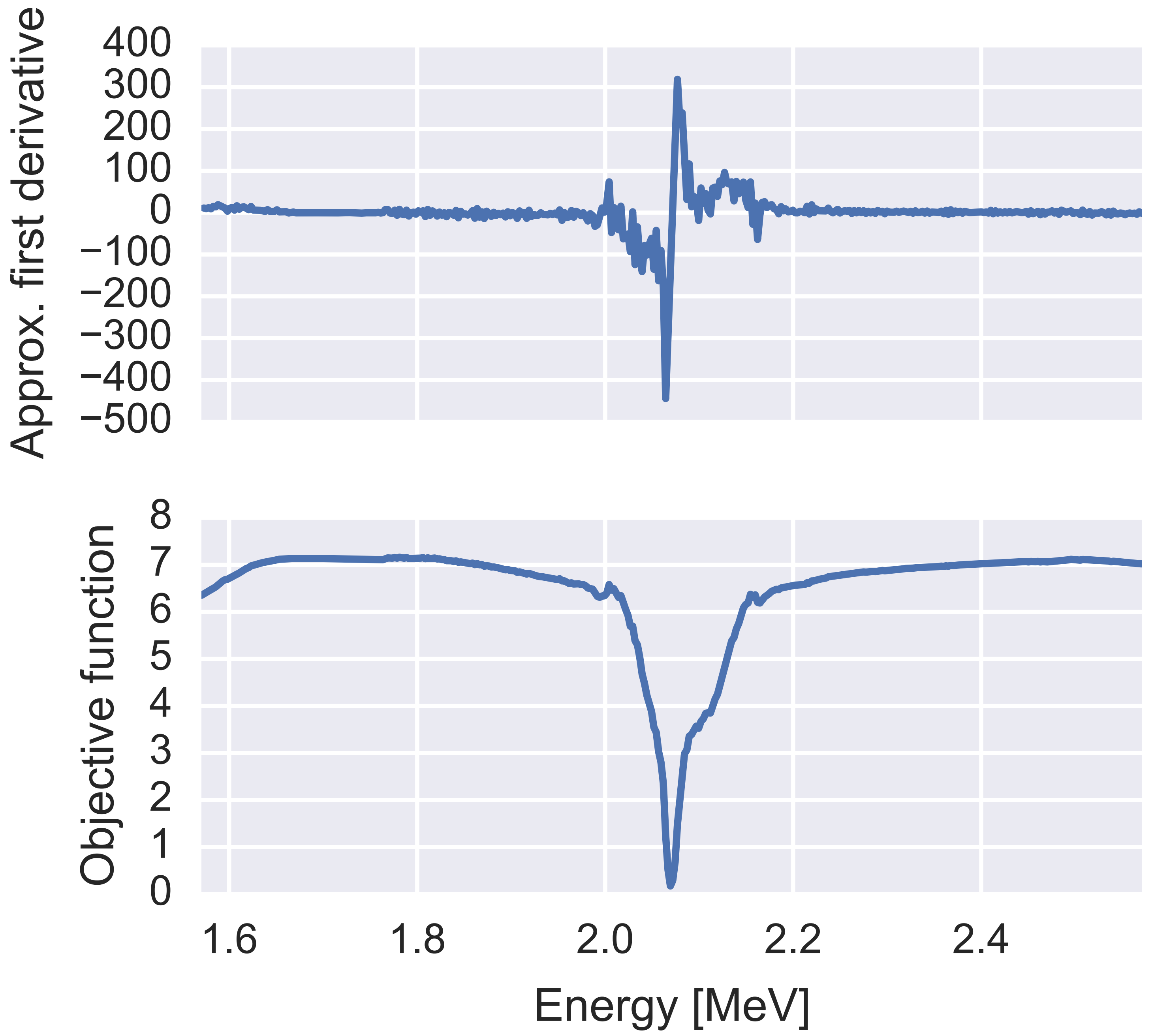




Gradient methods

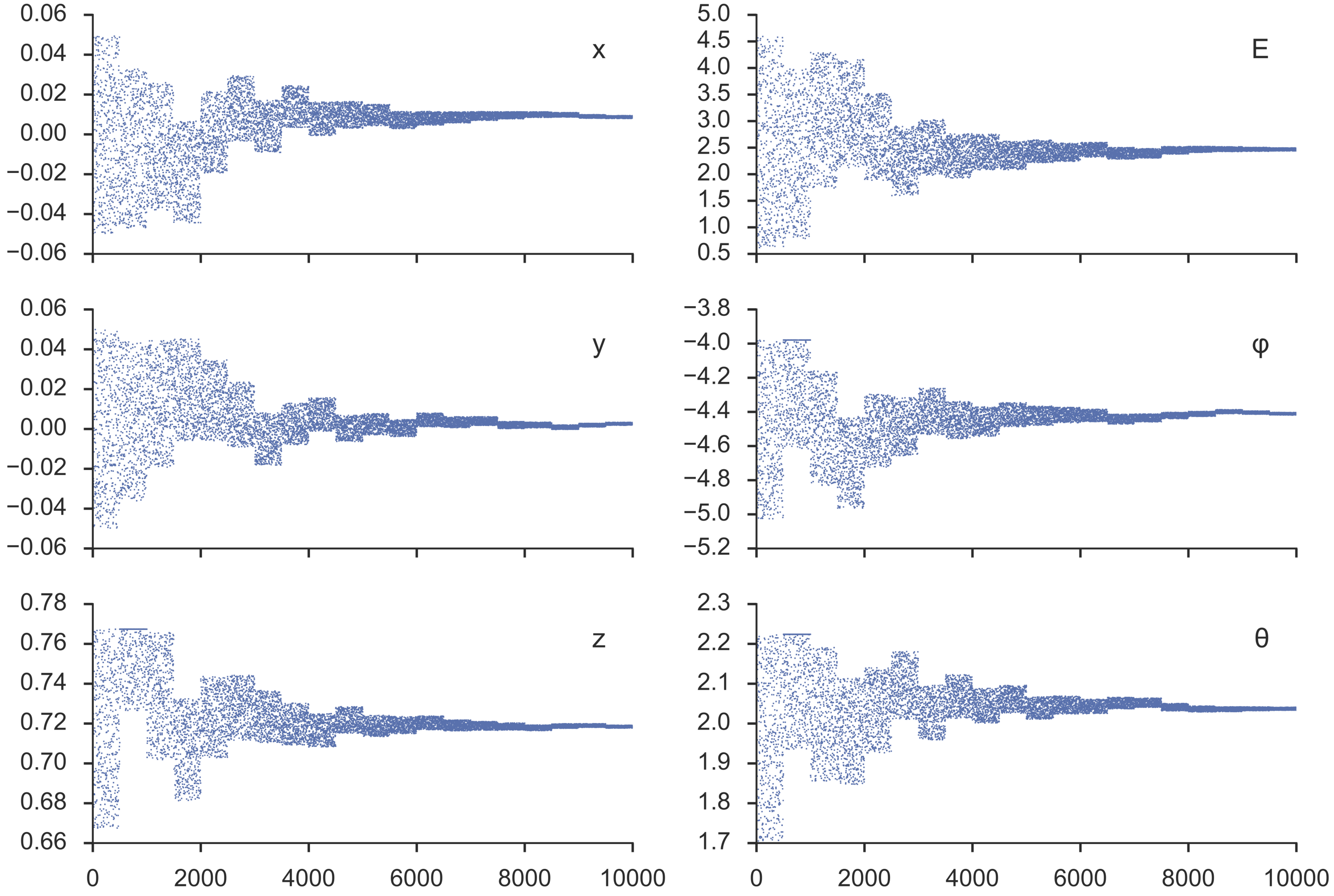
- Rely on good derivatives of the objective function
- Signal variation on pads highly non-linear
- Example shows energy χ^2 as a function of energy and scattering angle
- Algorithm most likely to get trapped in local minimum





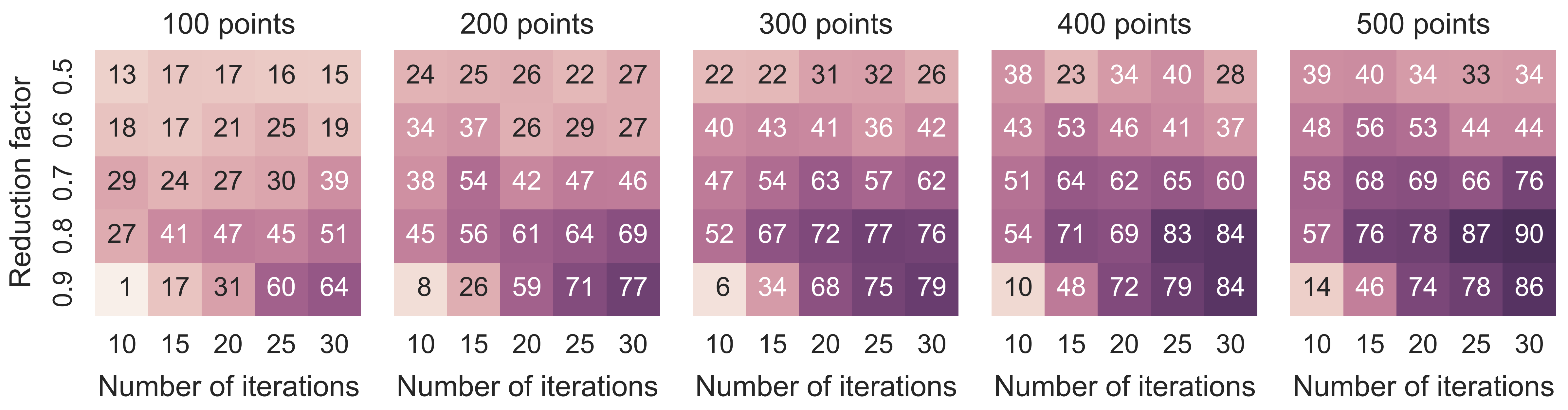
Monte-Carlo fitting

- Iterative method based on simulation
- Generate set of randomly simulated events within parameter space
- Select track with the lowest value of the objective function
- Recenter the parameter space around best value and shrink by scale factor
- Repeat process for a fixed number of iteration or until achieved precision is satisfactory

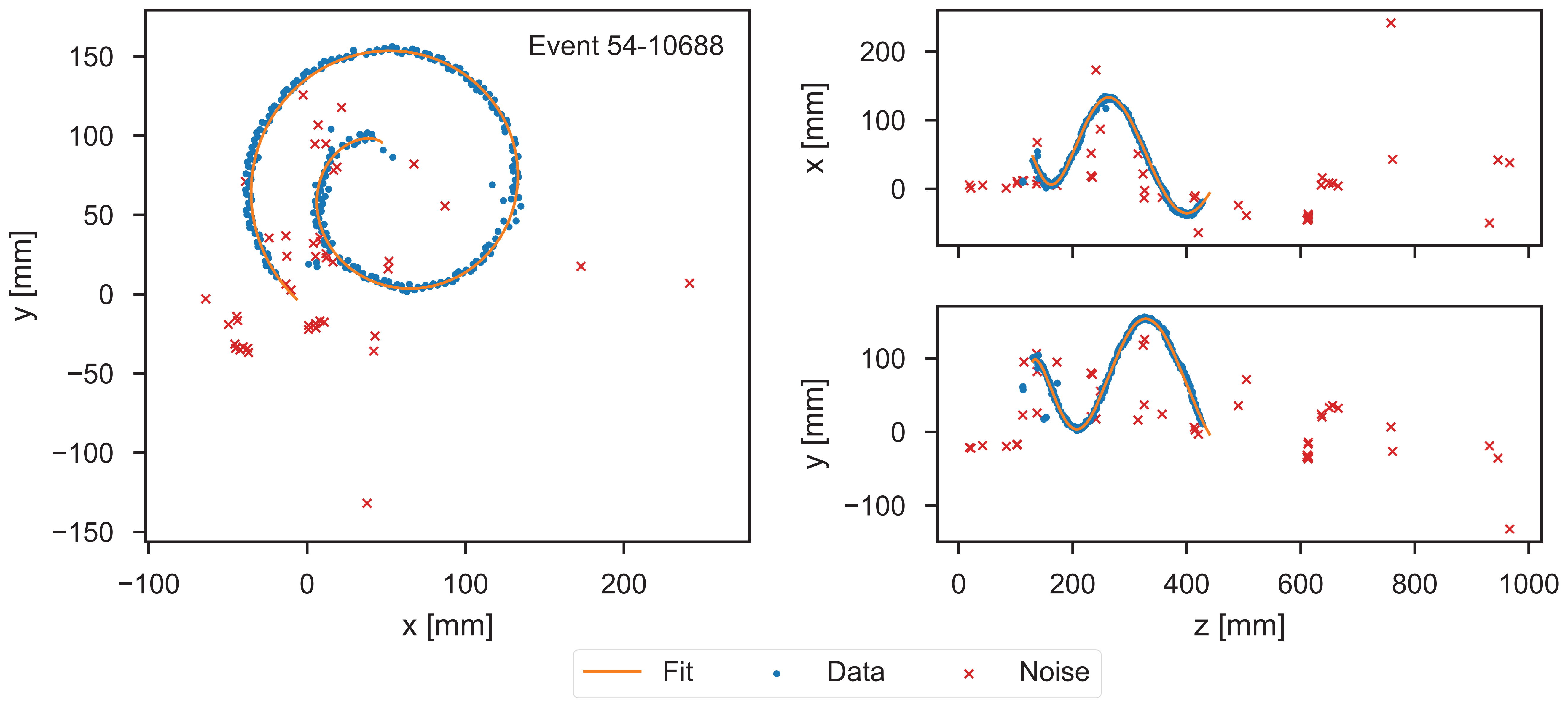


Monte-Carlo fitting

- Advantages
  - Very robust against non-linearities
  - Unlikely to fall into local minima
  - So far has given the best results
- Drawbacks
  - Relies on simulations
  - Requires numerous trials (1000's per event)
  - Needs parallel programming and processing



Performance study

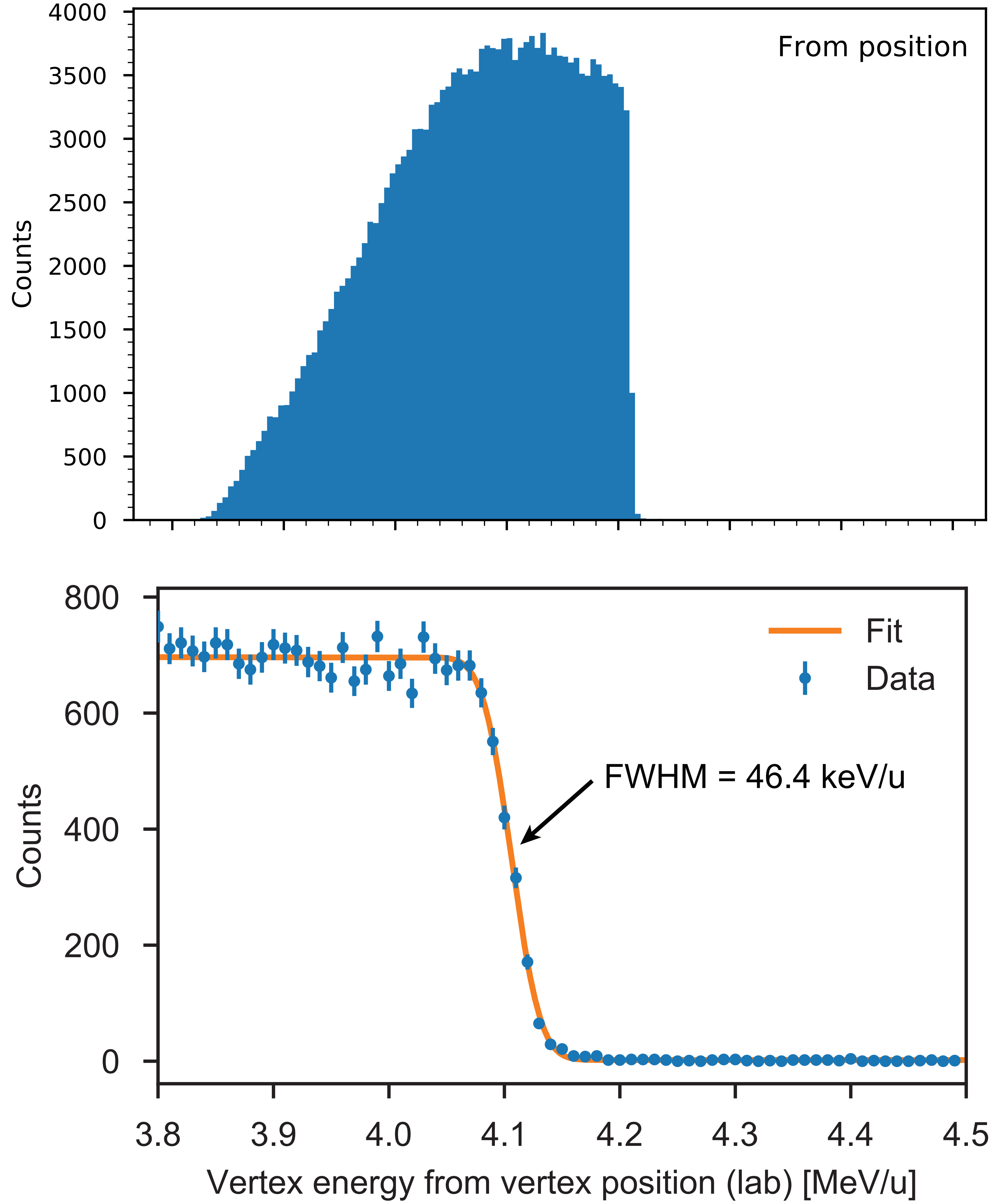


Example from AT-TPC data

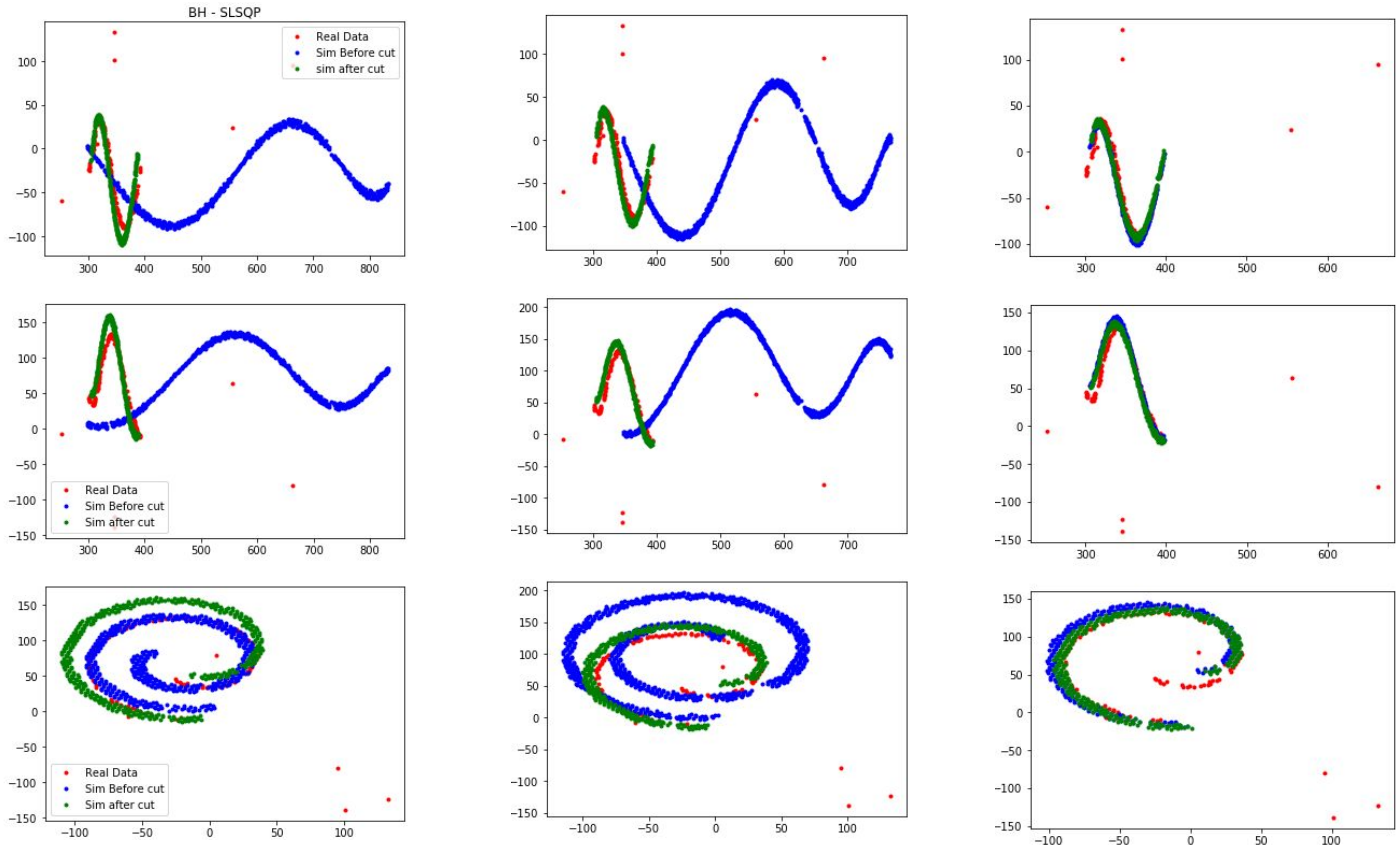
Example of achieved resolution

- Vertex energy determined from position in AT-TPC
- Proton elastic scattering from 46Ar
- Sharp falloff of vertex distribution corresponds to energy after entrance window of AT-TPC
- Resolution of vertex energy is FHWM of 46 keV/u

J. Bradt et al., PLB 778, 155 (2018)



Other methods…



Basinhopping Monte Carlo Diff Evolution

- Red: data
- Blue: fit with noise
- Green: fit w/o noise

Courtesy of M. Kuchera and C. Chen (Davidson college)

Concluding remarks

- Analysis of Active Target data is challenging
  - Very different from High Energy Physics TPC data
  - Techniques used in HEP require adaptation at the very least
  - AT community is in the early learning stages
- Nowadays fields of AI and ML are exploding
  - Many new algorithms and techniques are being developed
  - Computing power is becoming less of a limitation
  - Physicists need to open their eyes and ears to many other domains to take advantage of new techniques and algorithms