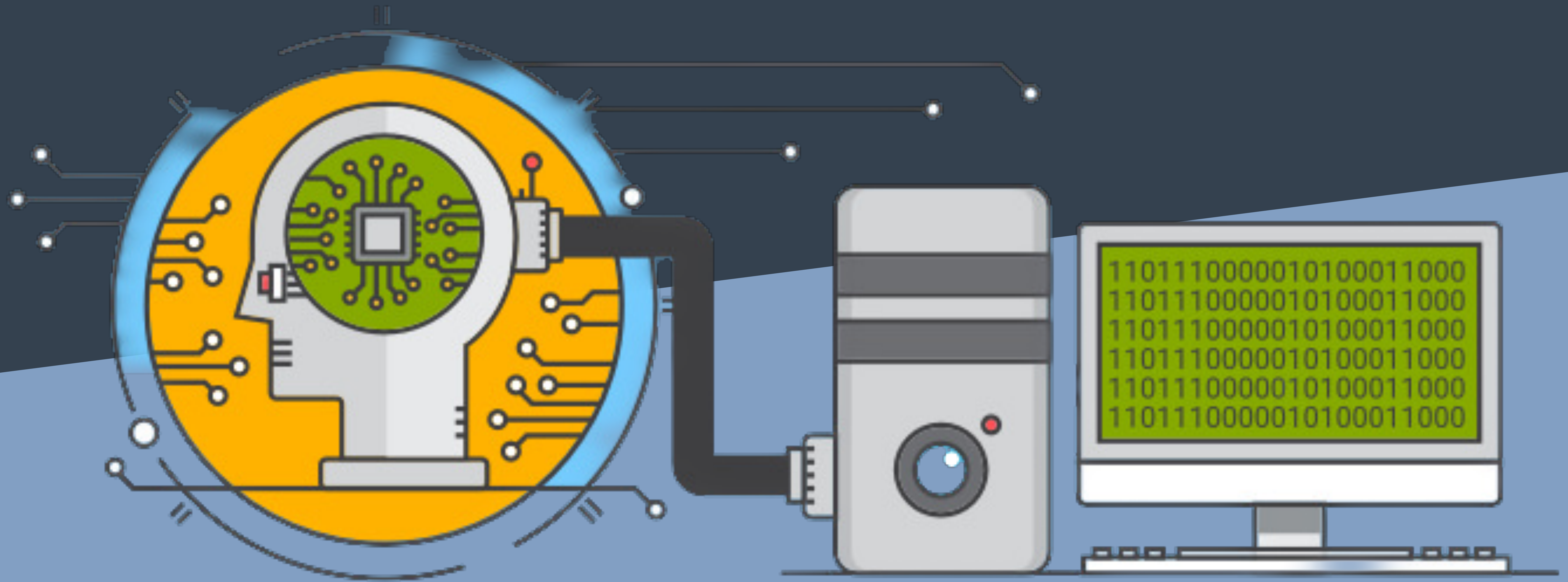
Artificial Intelligence in CLAS12

Artificial Intelligence/Machine Learning for Physics Applications

G.Gavalian (Jefferson Lab)



Richard Tyson
(University of Glasgow)

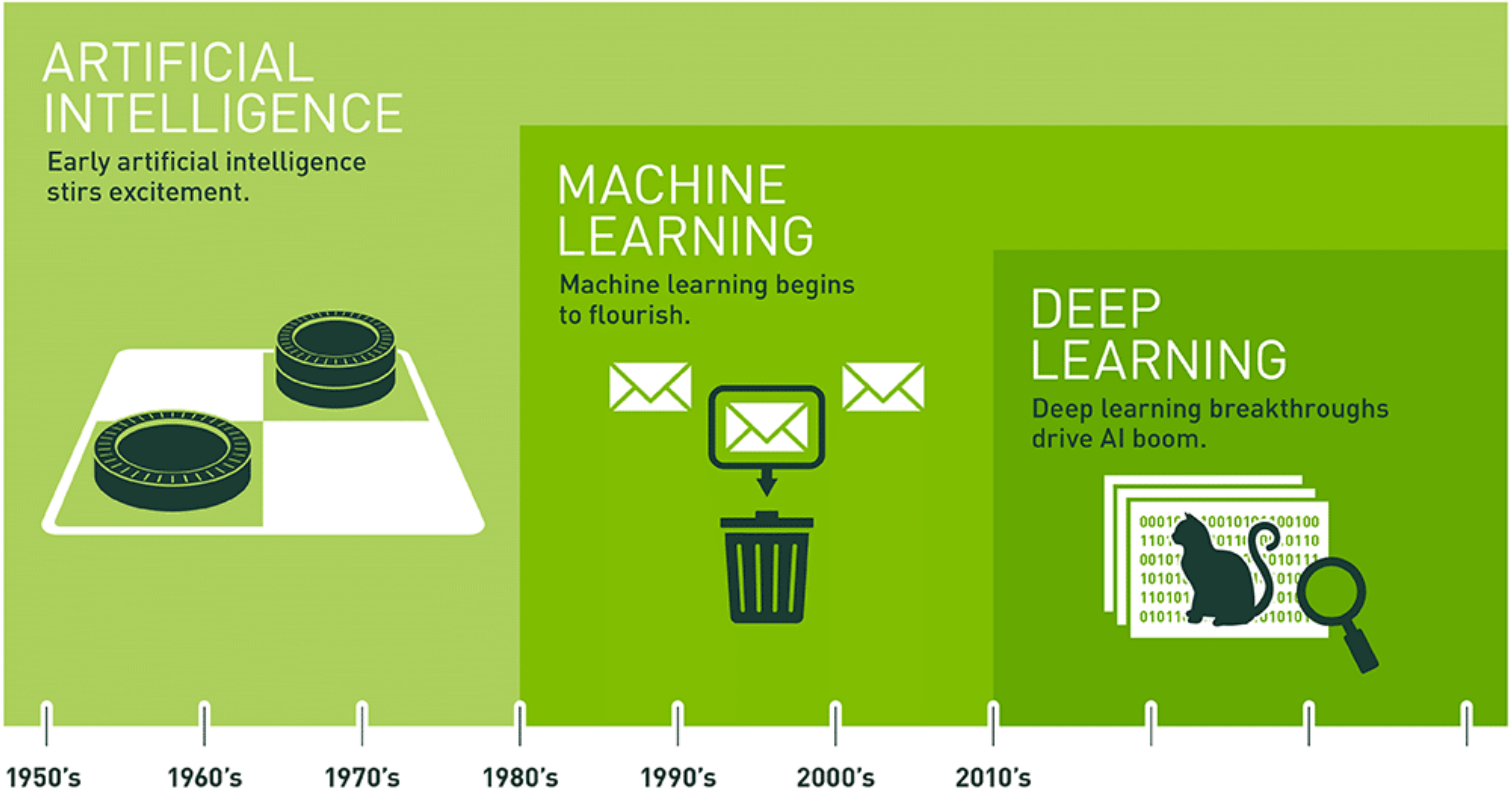
Angelos Angelopoulos (CRTC)
Polykarpos Thomadakis (CRTC),
Nikos Chrisochoides (CRTC)
Department of Computer Science,
Old Dominion University, Norfolk, VA, 23529

NSTAR (June 2024)

Outline

- Outline:
  - AI/ML advancements, direction, possibilities
  - How CLAS12 leverages AI/ML tools
  - What is the impact on physics from AI/ML
  - What is the impact on computation infrastructure
  - Looking into the future (streaming readout, other Jlab Experiments, EIC)

What is AI?



Since an early flush of optimism in the 1950s, smaller subsets of artificial intelligence – first machine learning, then deep learning, a subset of machine learning – have created ever larger disruptions.

History of AI

1950-1960

1950. Turing Test by Alan Turing
1956. Term of AI was coined

1961-1970

1961.First Industrial robot replaced humans at assembly line.
1964.Pioneering chatbot named ELIZA was developed at MIT
1966. A general purpose mobile robot developed at Stanford

1971-1990

AI Winter

1991-2000

1997. IBM's Deep Blue defeated Garry Kasparov in chess Competition
1998. An emotionally intelligent robot KISMAT was developed
1999. Sony launched pet robot dog named AIBO

2001-2010

2002. iRobot launched autonomous vacuum cleaner robot in bulk.
2009.Google built first self driving car for urban conditions

2011-2020

2011. Apple's SIRI and IBM's Watson were developed
2014.EUGENE, a chatbot passed Turing test; Amazon launched Alexa, a voice enabled intelligent virtual assistant.
2017.Google's AlphaGO beat the world's best GO player Ke Jie.

2020

- **Moxie:** A Social-Emotional Companion for kids is developed by Embodied.
- Earth's first autonomous **beehive** is developed by beewise
- **TrialJectory** is an AI enabled service to look for clinical trials.
- **BrainBox AI** is an AI system to predict a building's thermal conditions.
- Refined business process, more personalized recommendations, human like conversational skills

2021

ChatGPT

Midjourney

LLAMA

AI in Nuclear Physics

My First experience with Event Reconstruction

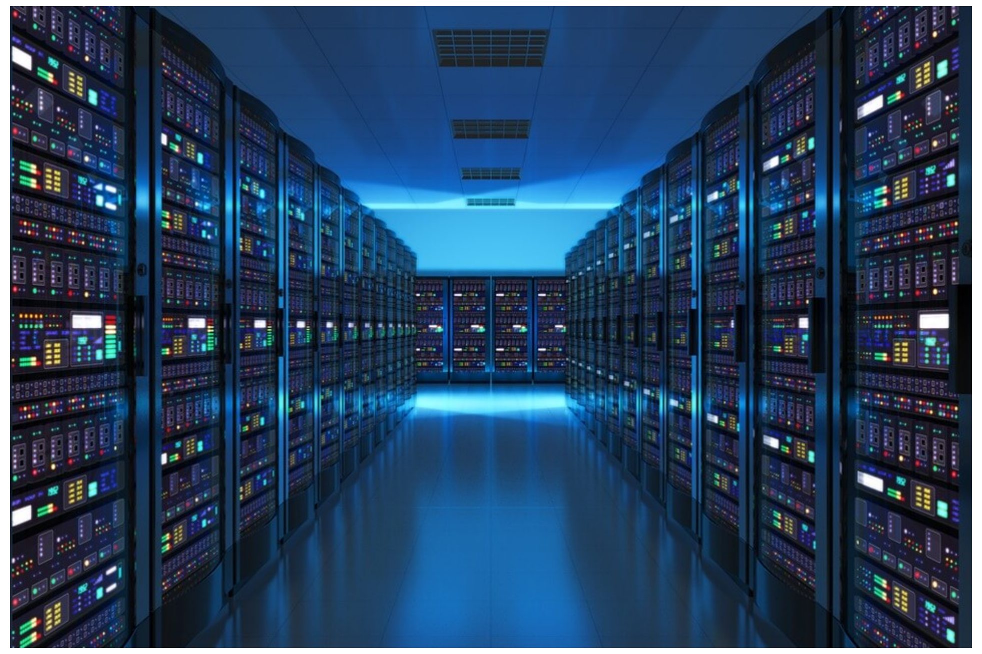
Rate: ~**0.0008** Hz (single person, assuming 20 min per event)

Earth Population: 4.767 billion (2,135 kHz assuming 56% in the age bracket 21-65)

CLAS12 event reconstruction

Rate: **2-3** Hz (single CPU) (many more channels, higher rates)

Computers now (64 Cores), 2.6 MHz

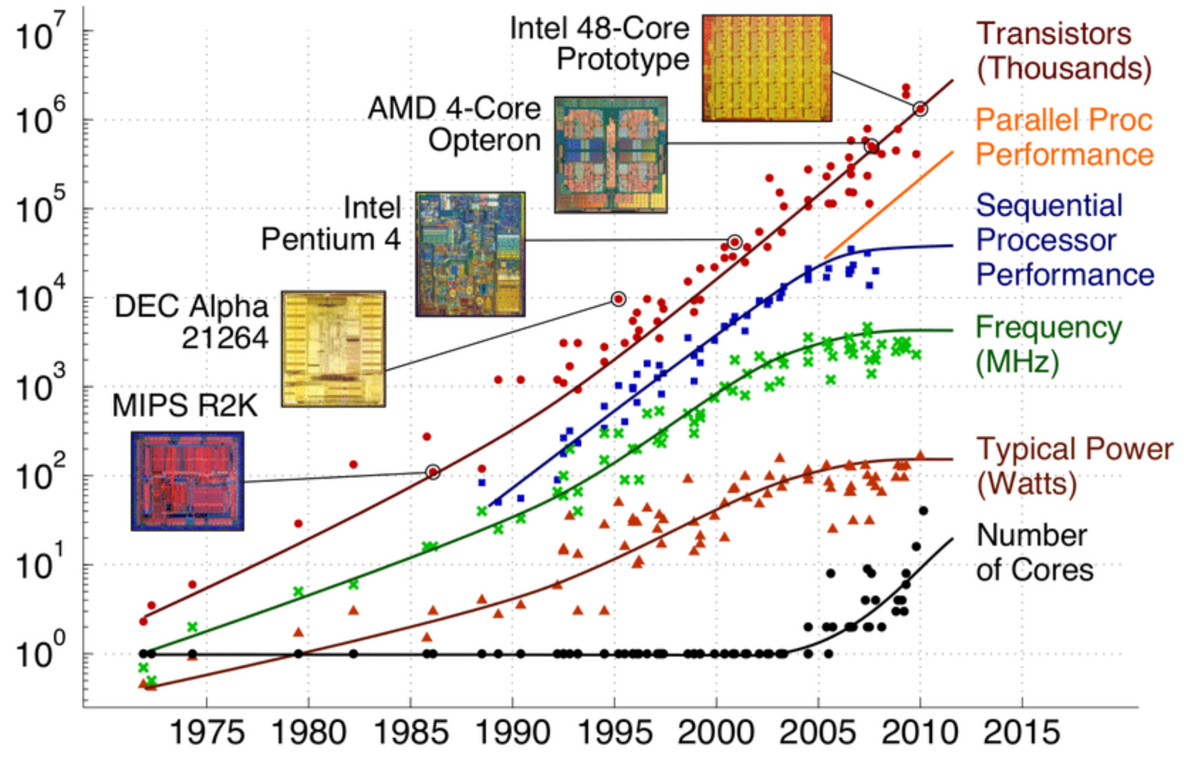


1983 — Big Jump → 1996-2001 — No change → 2023 → ?

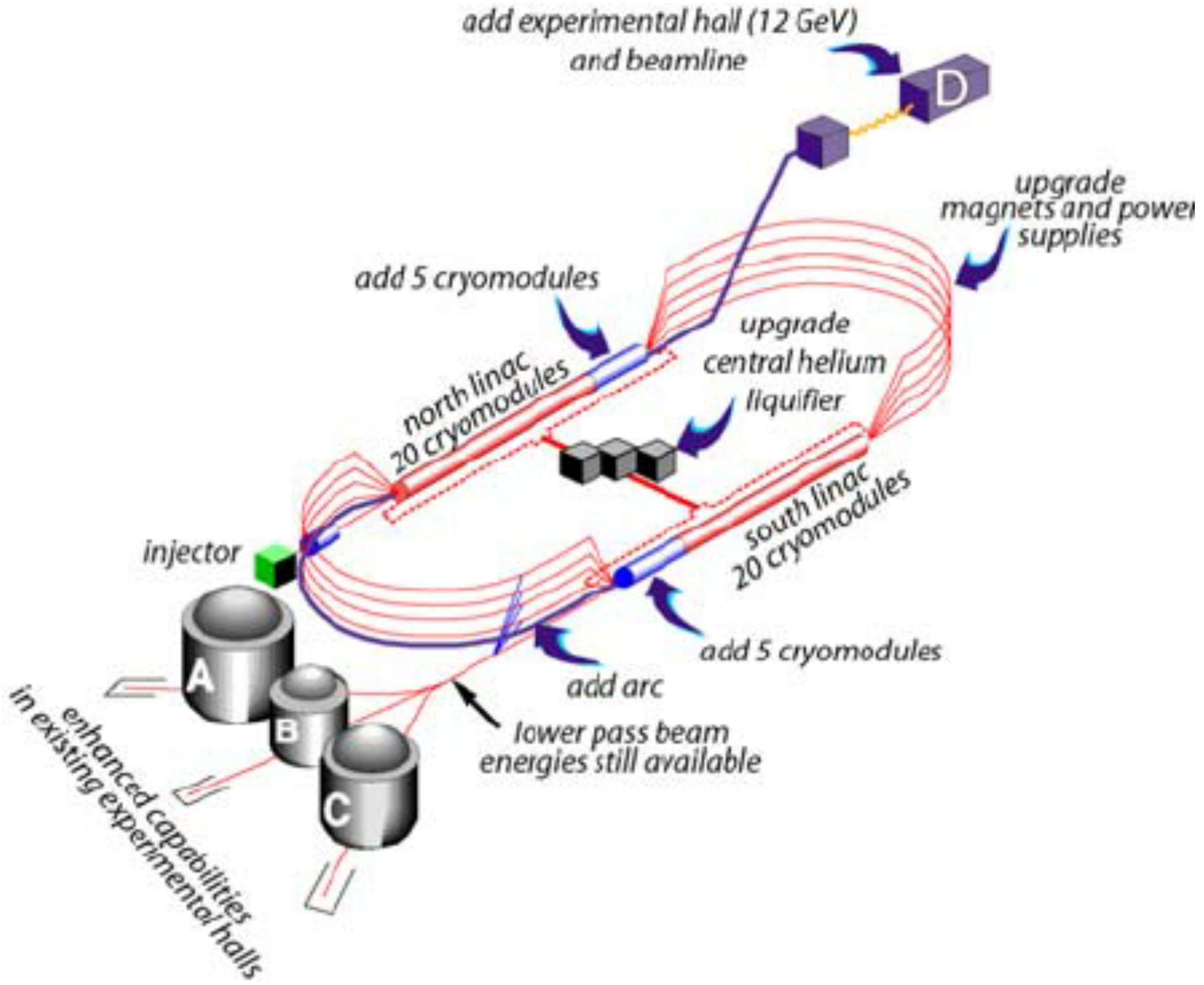
CLAS6 event reconstruction

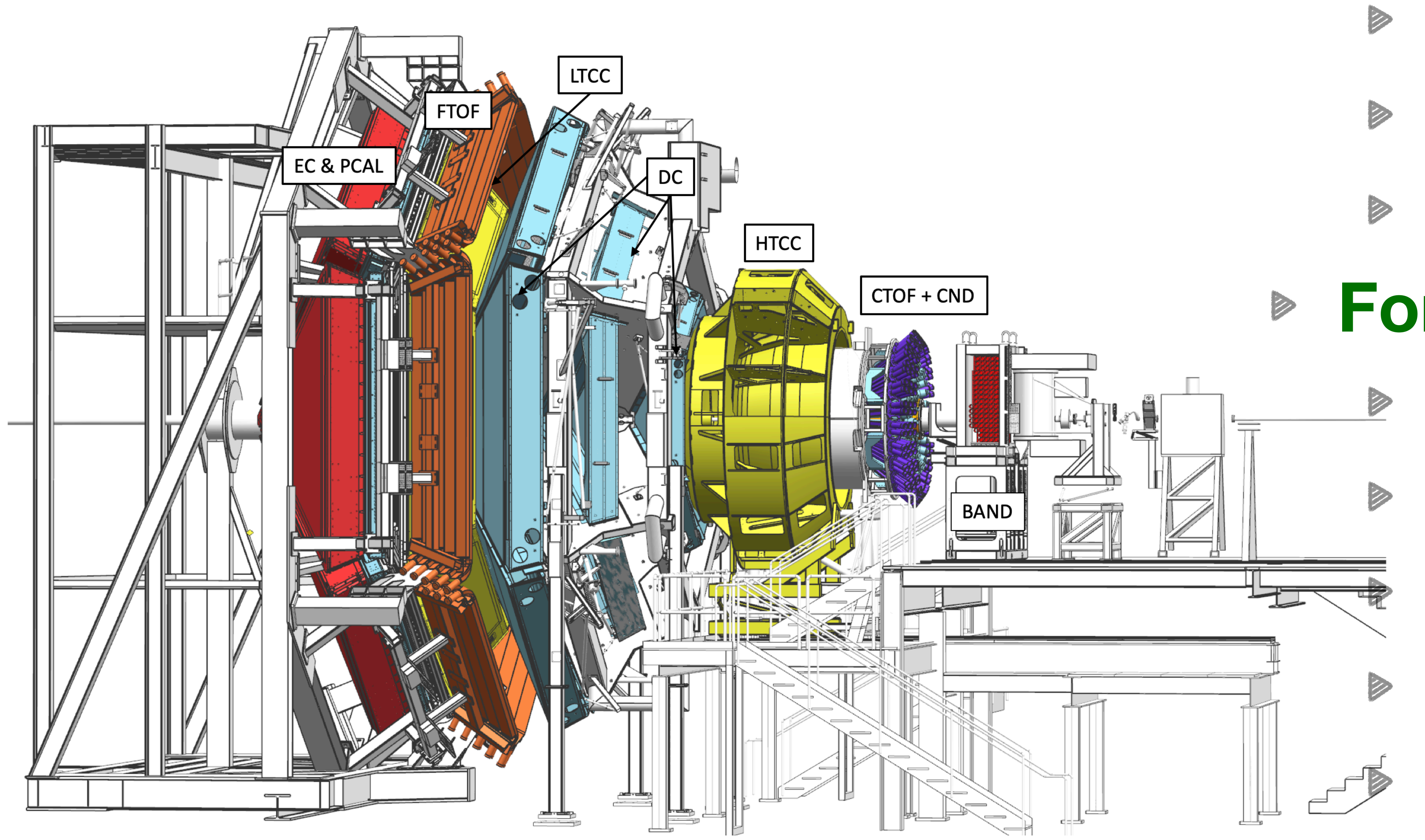
Rate: **8** Hz (single CPU)

Computers 4 Cores, 2.4 MHz



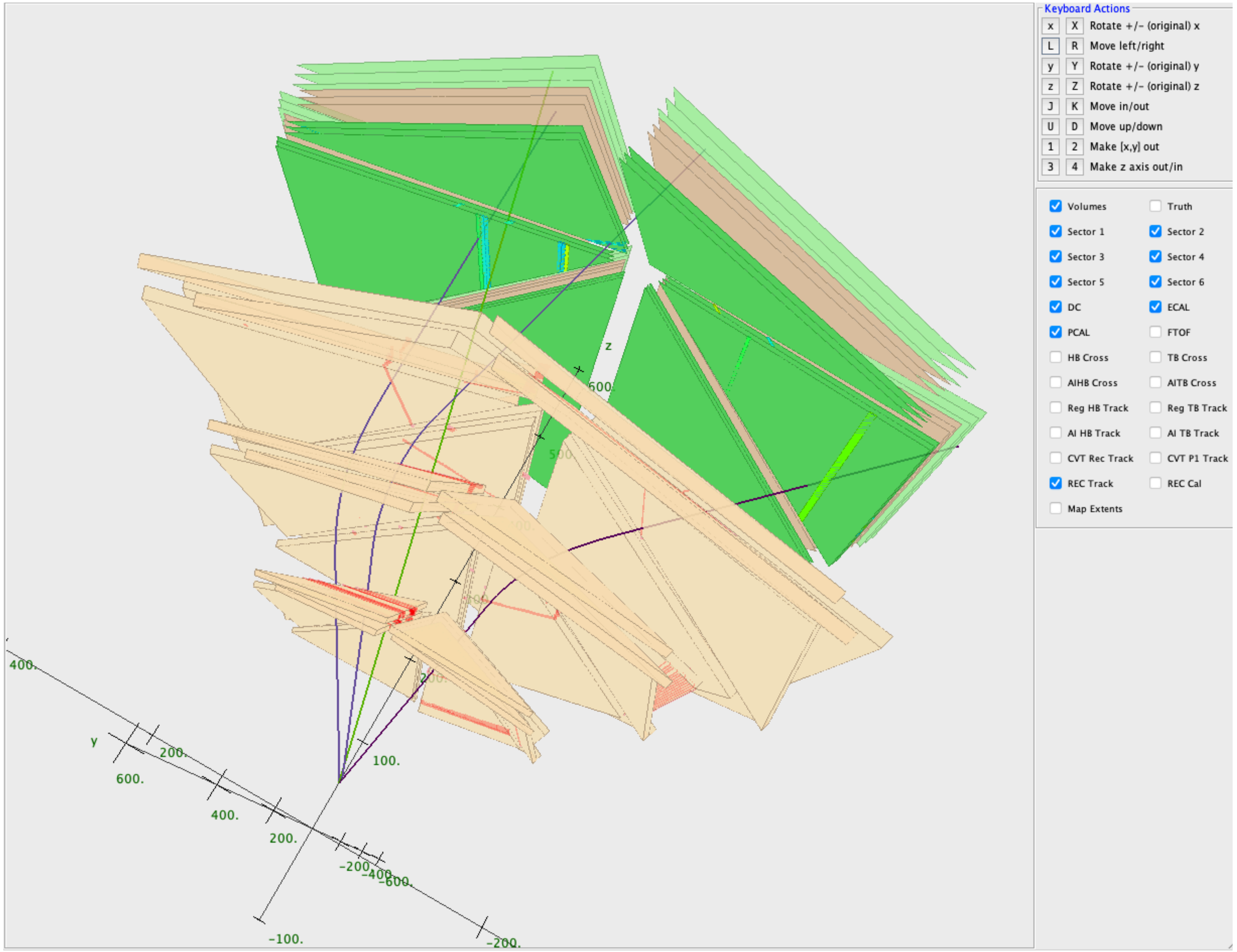
Jefferson Lab/CLAS12





- CEBAF
  - 12 GeV electron beam distributed to 4 experimental hall
  - Each experimental hall contains a detector system for specific experiments
- **Hall-B:**
  - CEBAF Large Acceptance Spectrometer (CLAS12) Located in Hall-B
  - **Central Detector:**
    - Silicon Tracker
    - Time-Of-Flight
    - Neutron Detector
  - **Forward Detector:**
    - Drift Chambers
    - Time of Flight
    - High Threshold Cherenkov Counter
    - Ring Imaging Cherenkov Counter
    - Electromagnetic Calorimeter

Drift Chambers

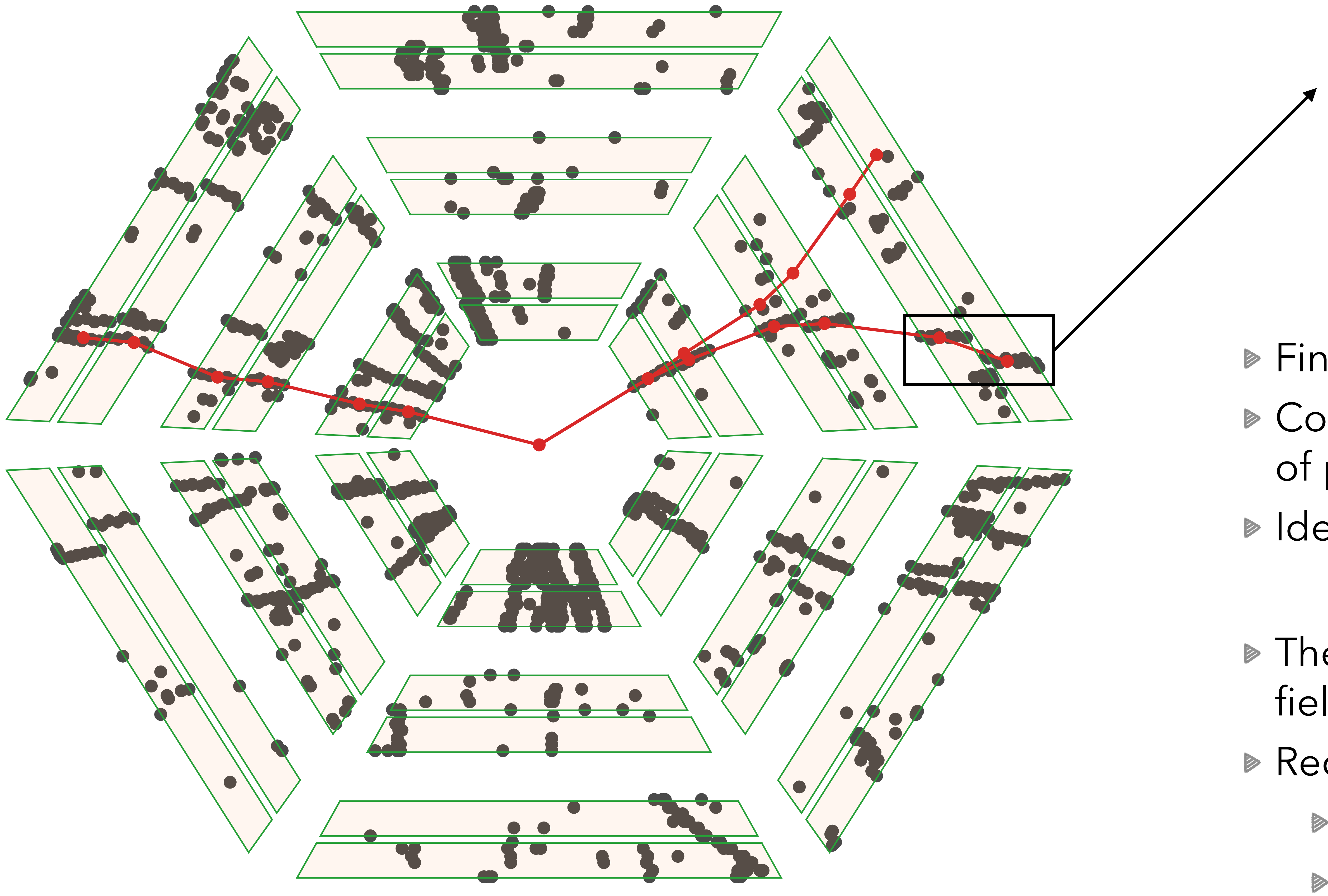


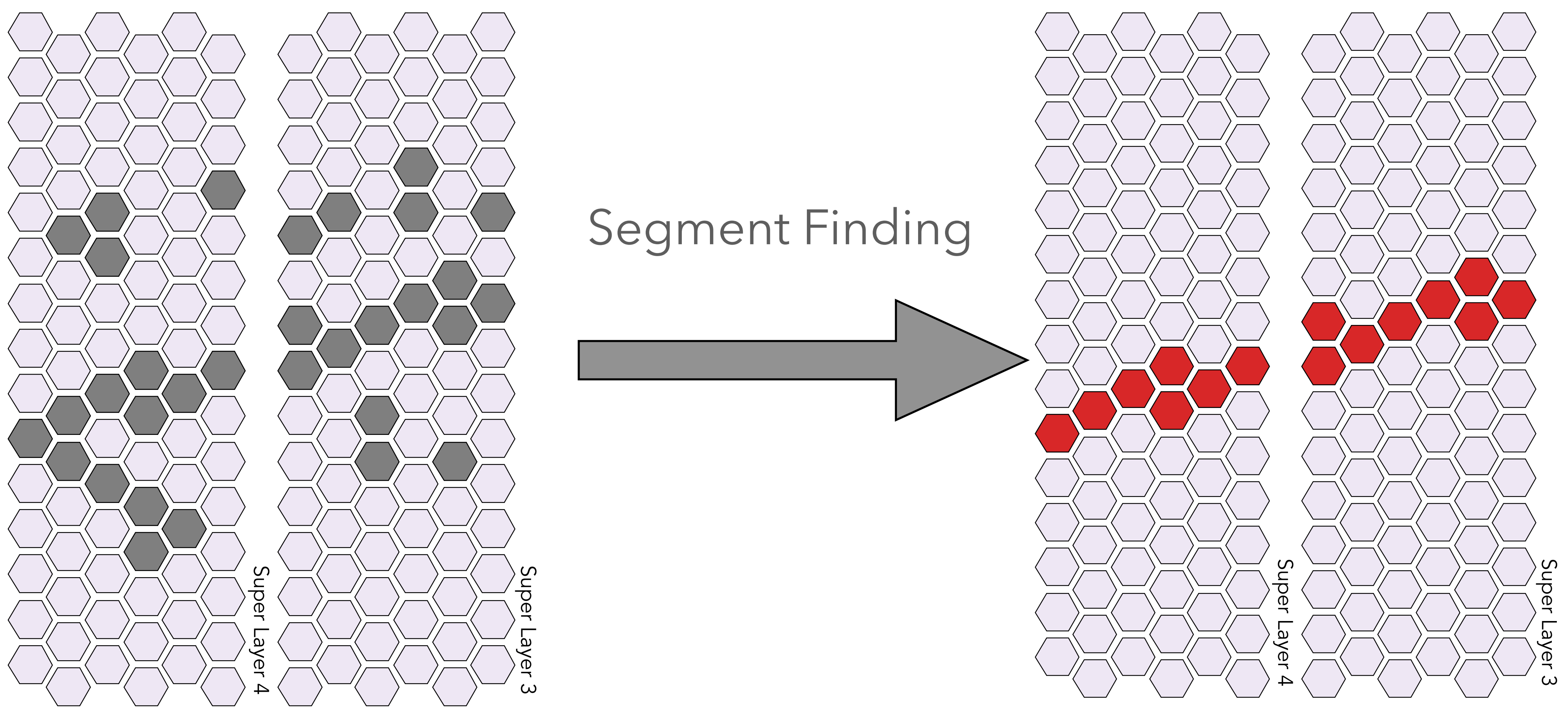
Parameter	Specification
Angular Range	5-40 degrees
Momentum Resolution	dp/p<1%
Polar resolution	1 mead
Azimuthal Resolution	1 mrad/sinT
Luminosity	10e35 cme-2se-2

Parameter	Specification
Cell Type	Hexagonal
Wire layout	6 sectors, 3 regions
Stereo	+/- 6 degree stereo
Granularity	112 wires/layer (24192)
Gas	90/12 Argon/CO2

Track Finding

- 6 sectors with 6 chambers in each sector (called super-layers)
- 6 wire planes in each super layer with 6-degree tilt relative to each other, (112 wires in each plane)

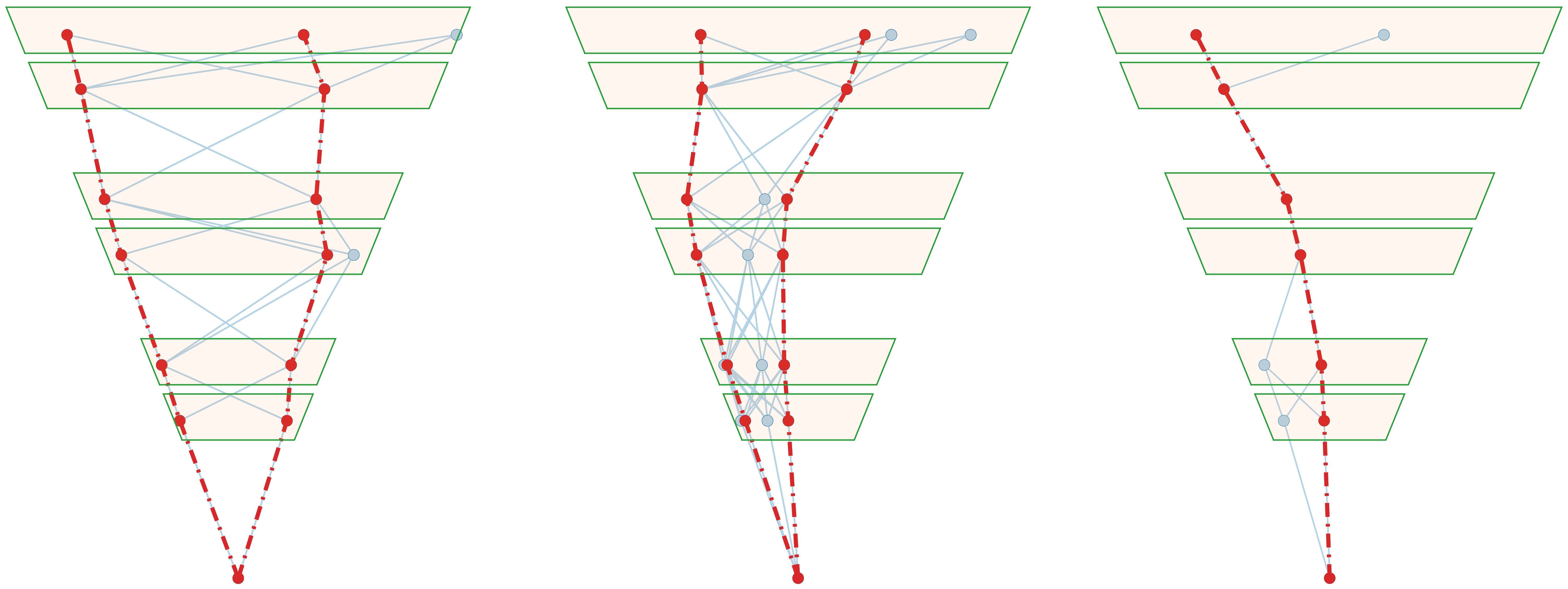


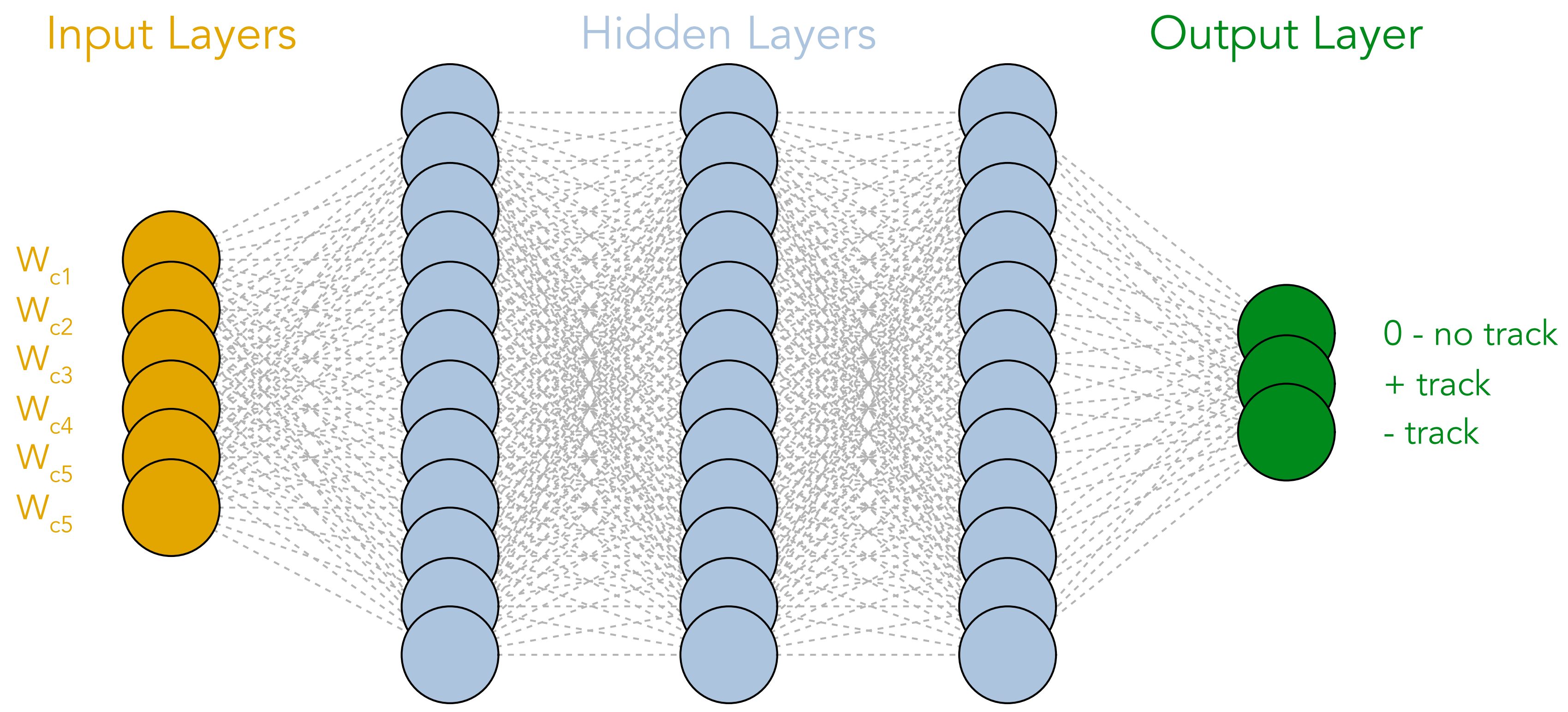


- Find segments in each super layer (remove noise)
- Combine 6 segments (one from each super layer) to make a list of possible tracks
- Identify correct combinations of segments that represent a track

- The conventional algorithm performs fit through the magnetic field to assess the goodness of the track.
- Requires:
  - Knowledge of drift chamber geometry
  - The precise value of the magnetic field in space

Track Finding



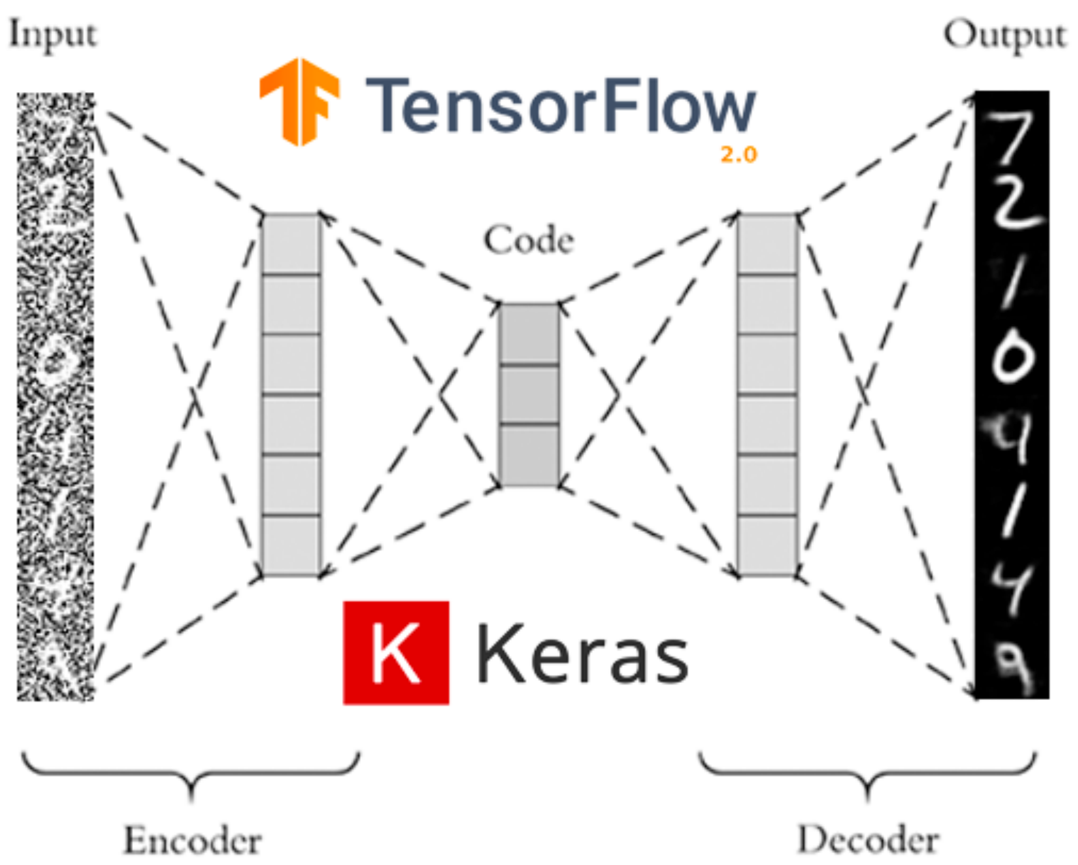


- Input: W [1..6] - average wire position of the segment
- Output: [false track, positive track, negative track]

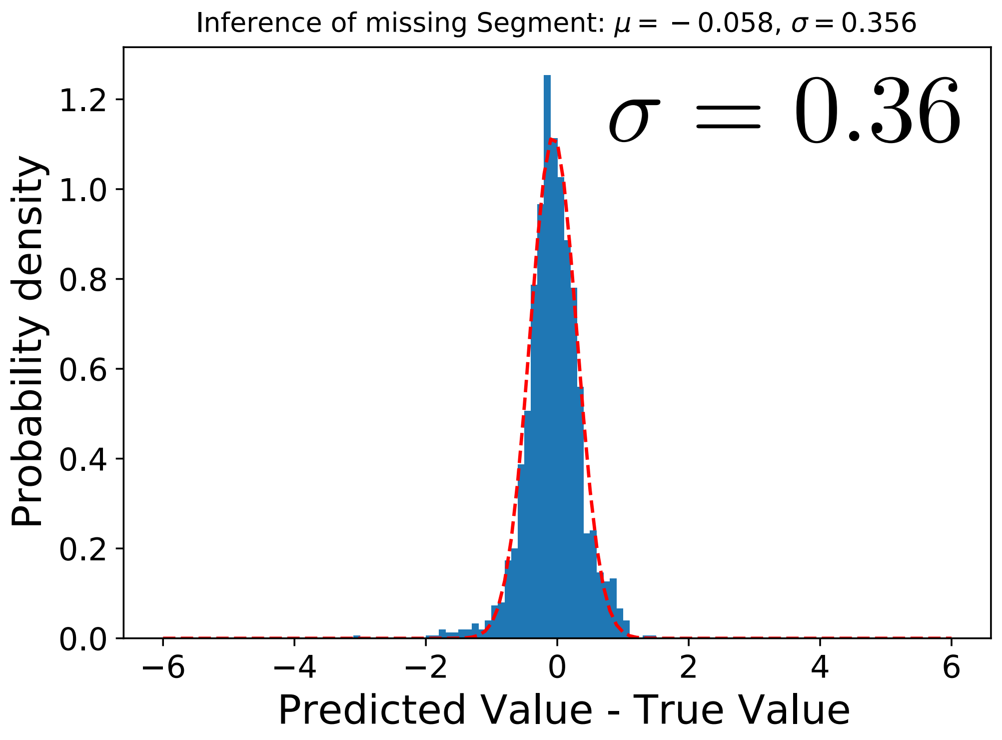
- Neural Network is trained to recognize patterns of segment combinations
- The track classifier assigns a probability of the track candidate to be positive, negative, or false track.
- The network is trained on reconstructed data where the right combinations are already found and false combinations of segments is generated by interchanging clusters from a different track

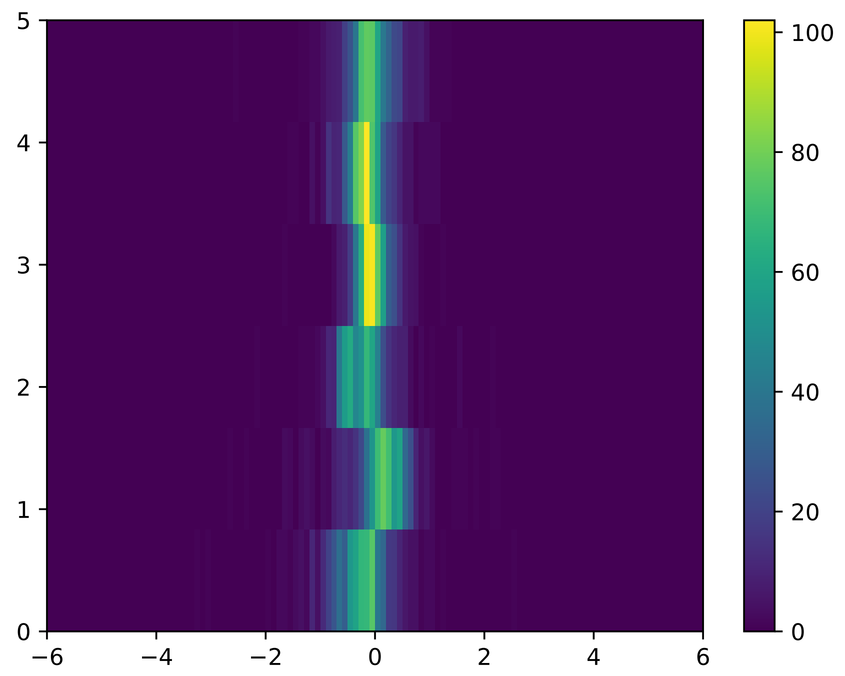
Corruption Auto-Encoder

- An auto-encoder is composed of an encoder and a decoder sub-models. The encoder compresses the input and the decoder attempts to recreate the input from the compressed version provided by the encoder.
- **Typically used for de-noising, but can be used for fixing glitches (our case).**

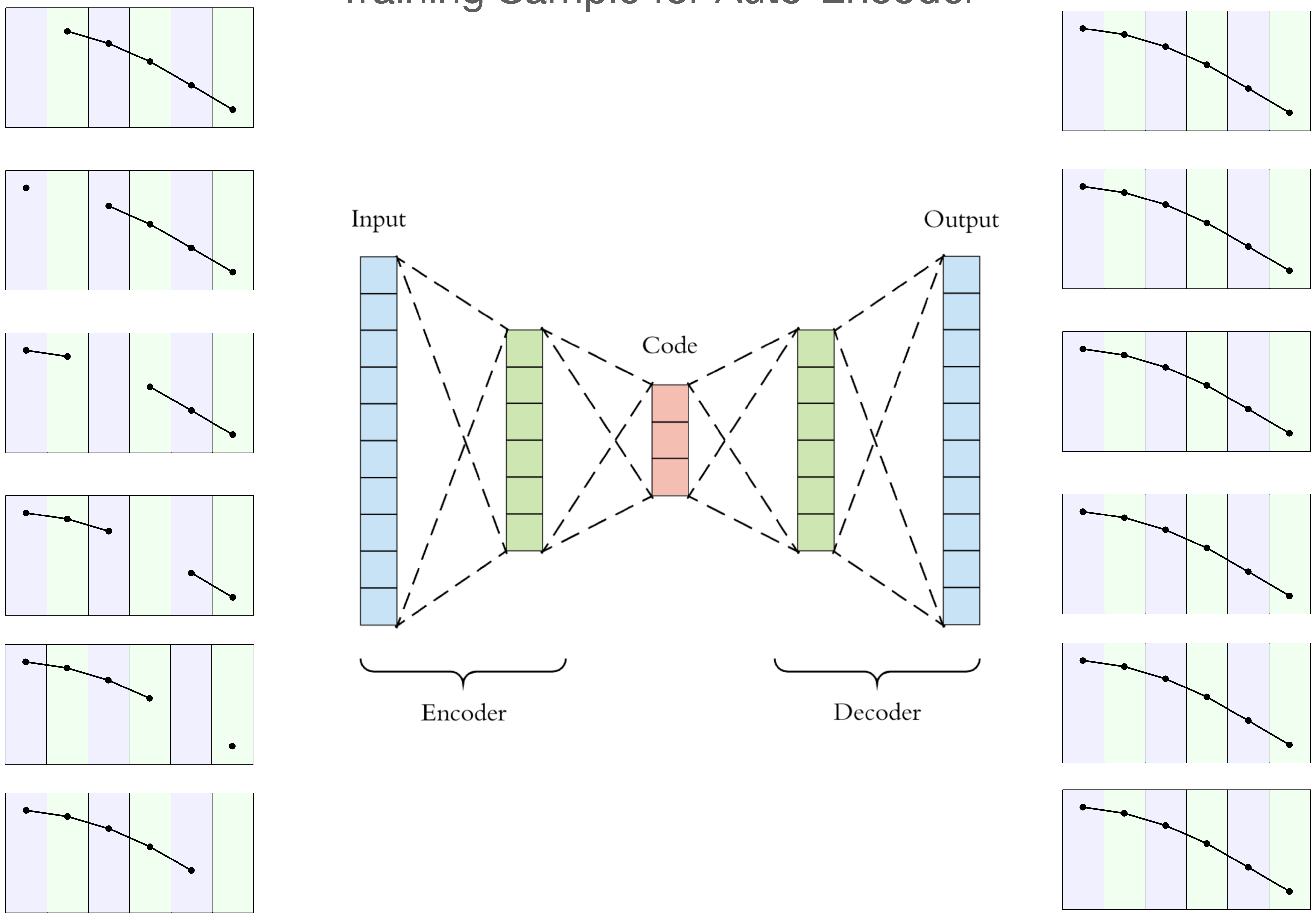


- The network Predicts the missing cluster position with a precision of 0.36 Wire



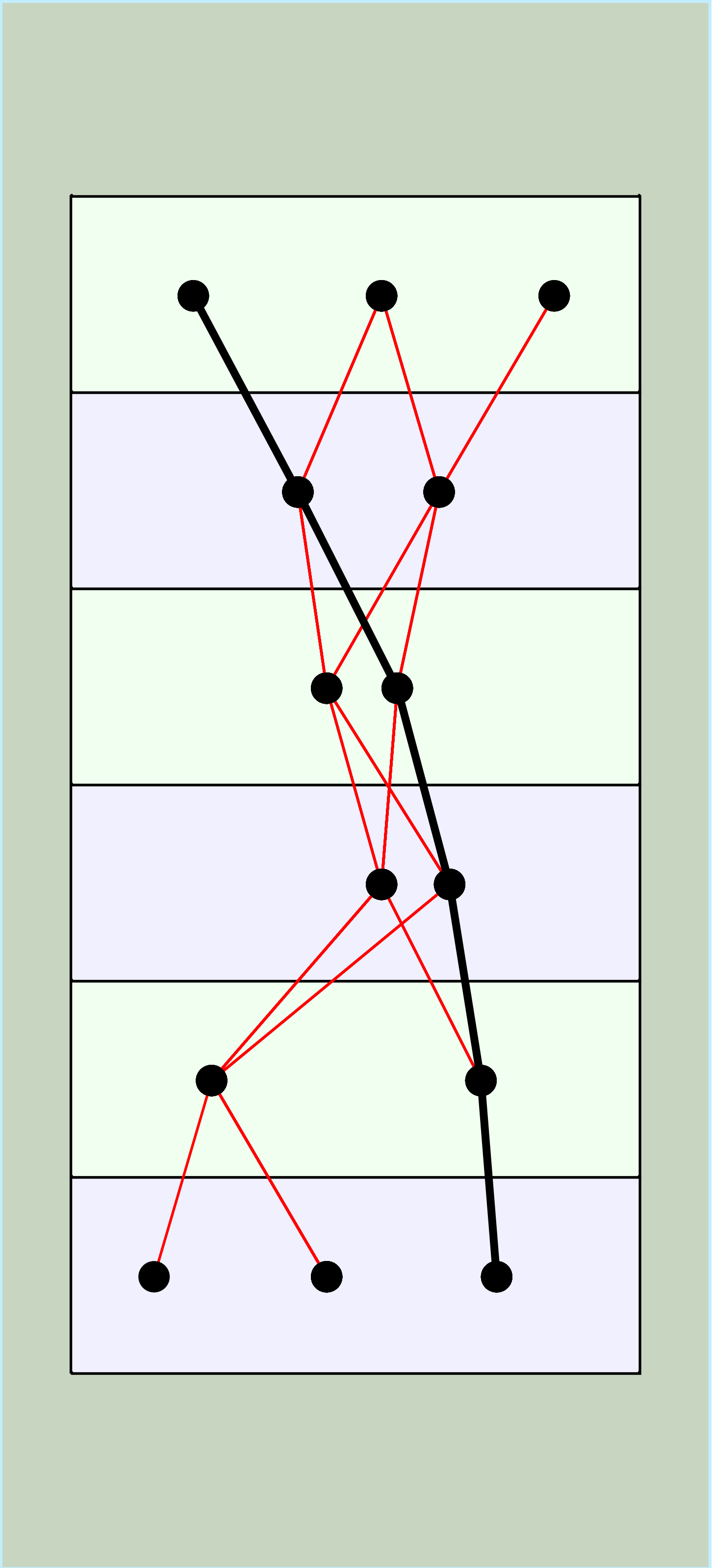


Training Sample for Auto-Encoder

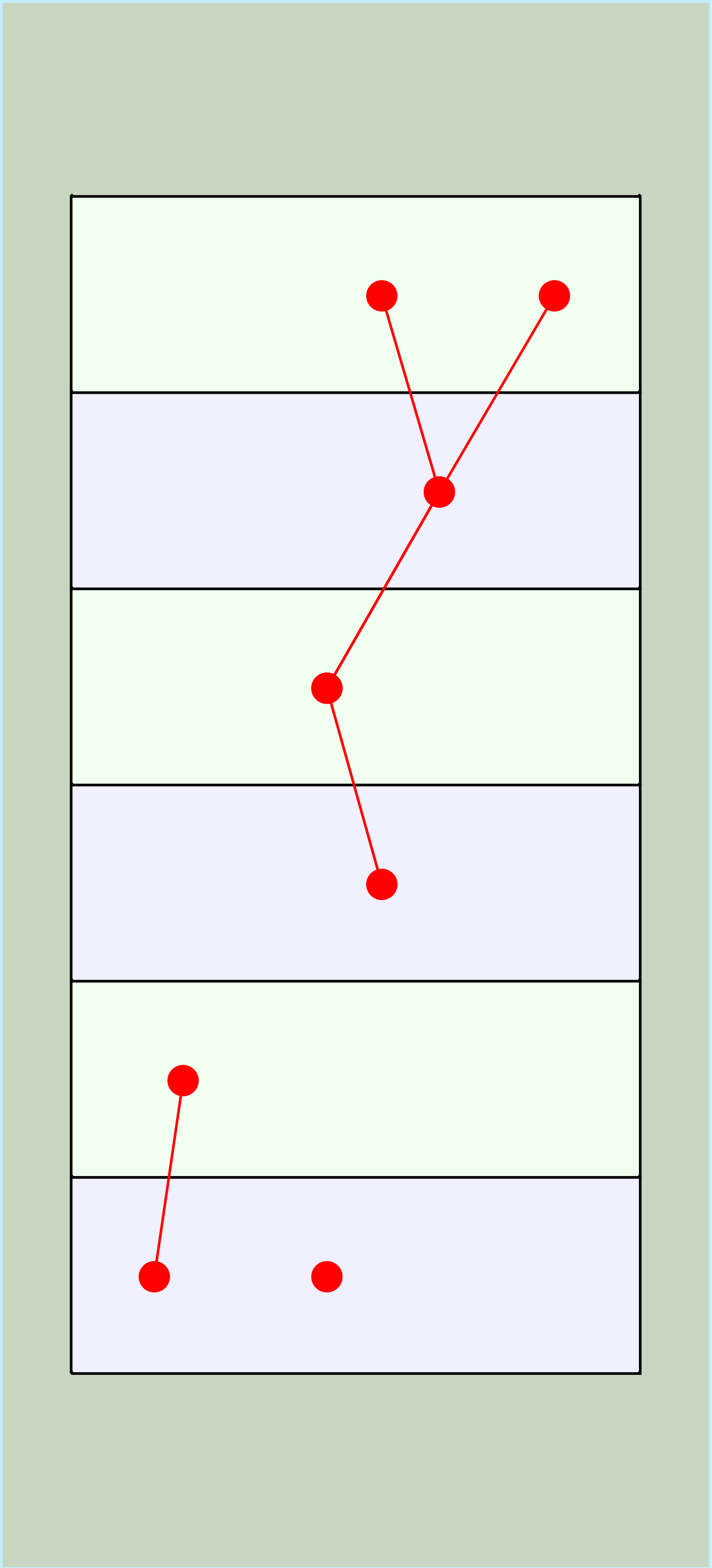


- Use Auto-Encoders to fix the missing cluster (provide a position)
- Good reconstructed tracks are used to generate training samples by removing one cluster from each super layer

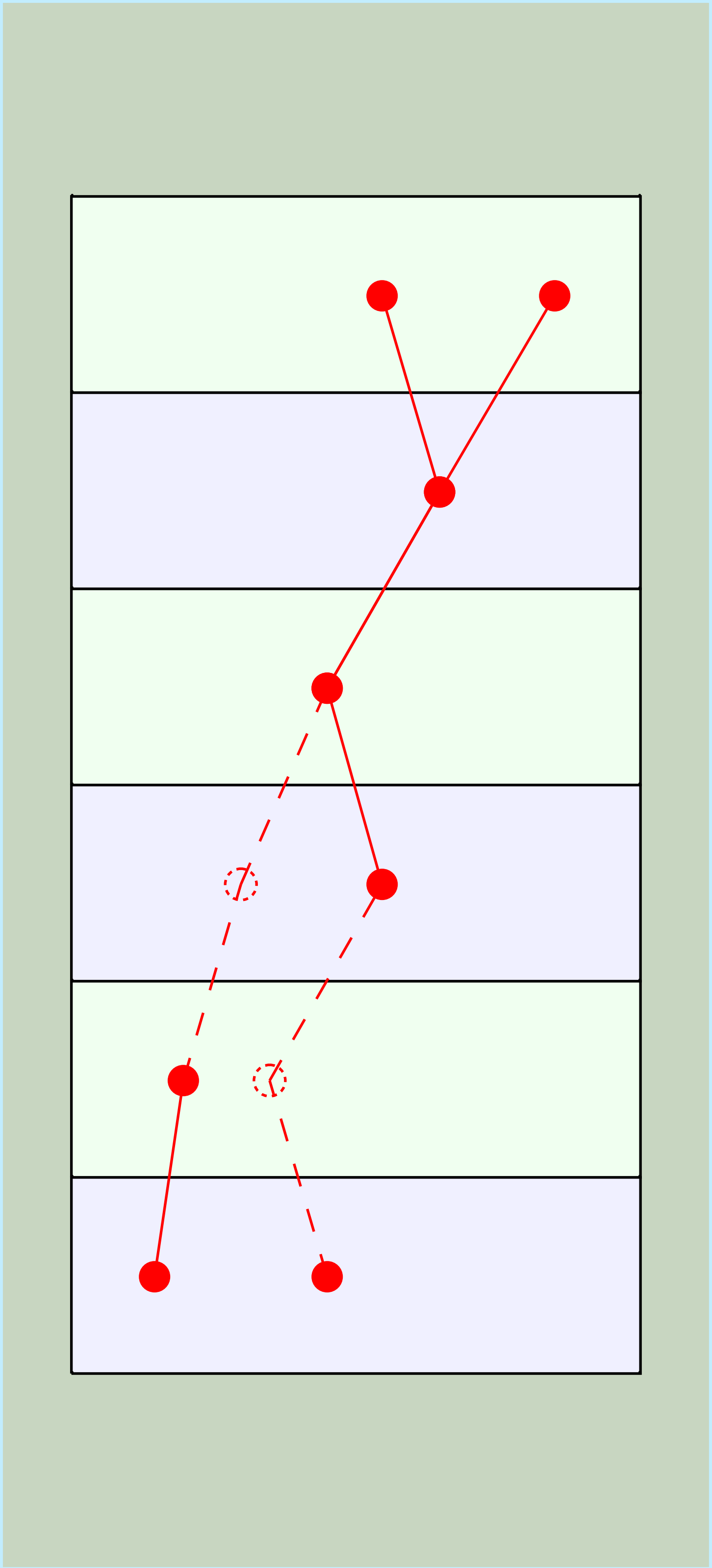
Putting all together



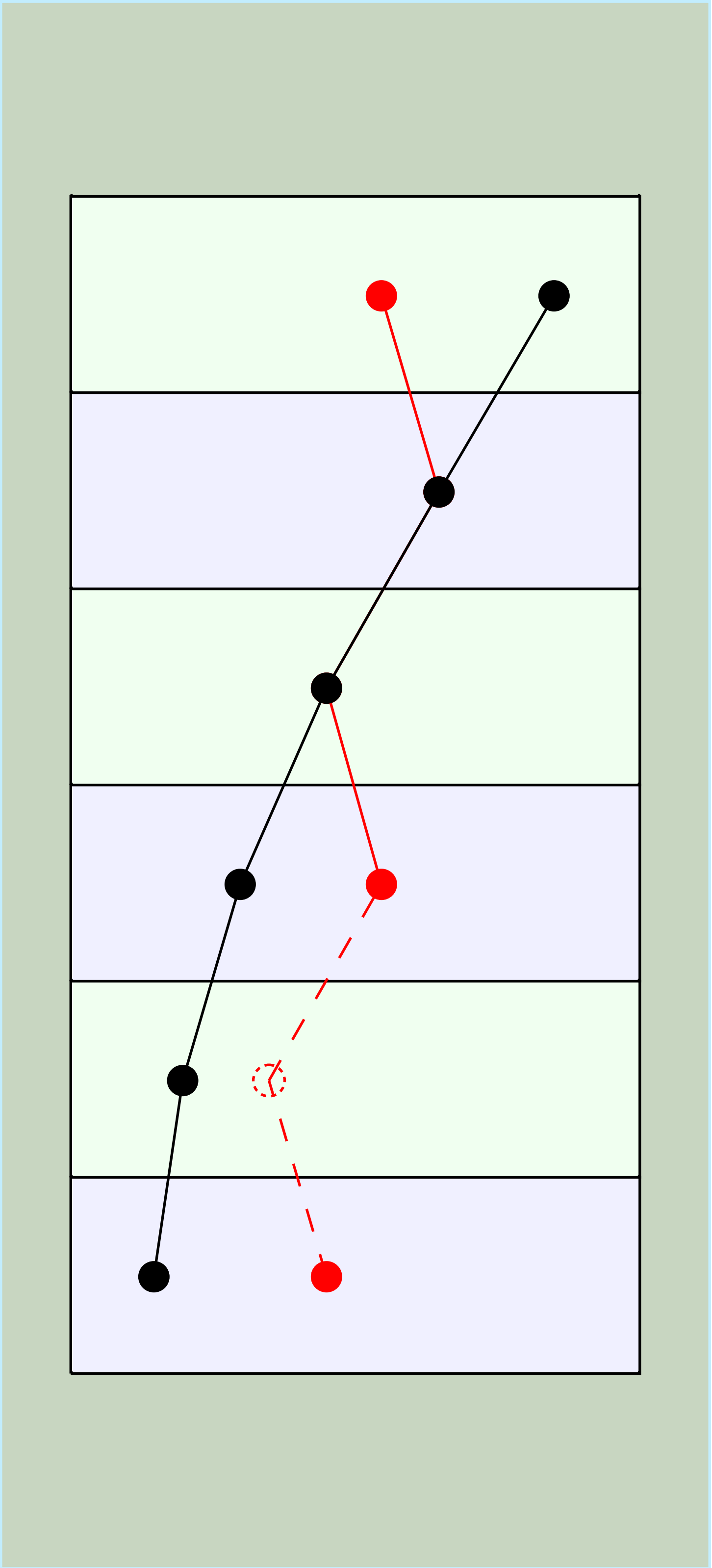
Classifier picks the correct track from 6 super-layer combinations



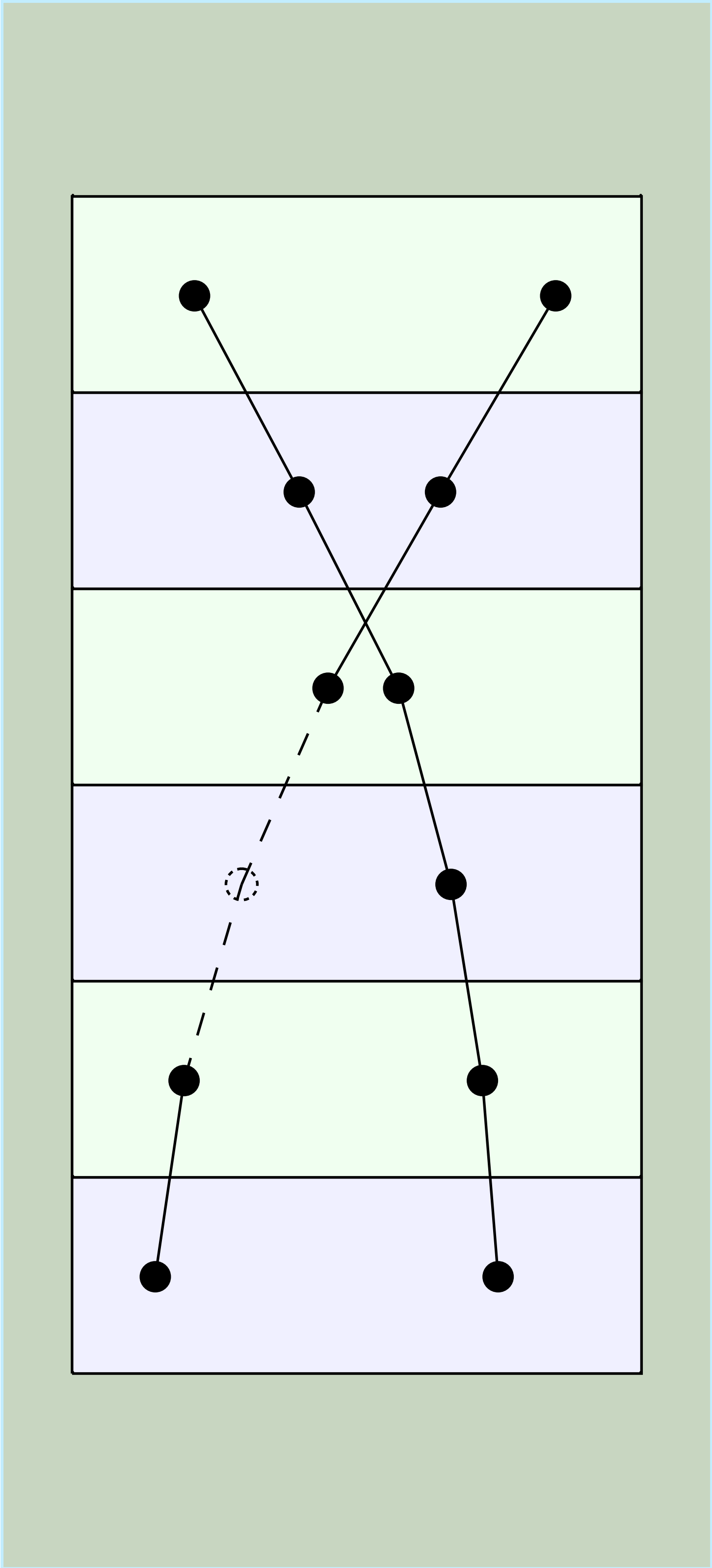
Remove all clusters belonging to identified track



Construct pseudo-clusters for all 5 super layer combinations using Corruption Auto-Encoder

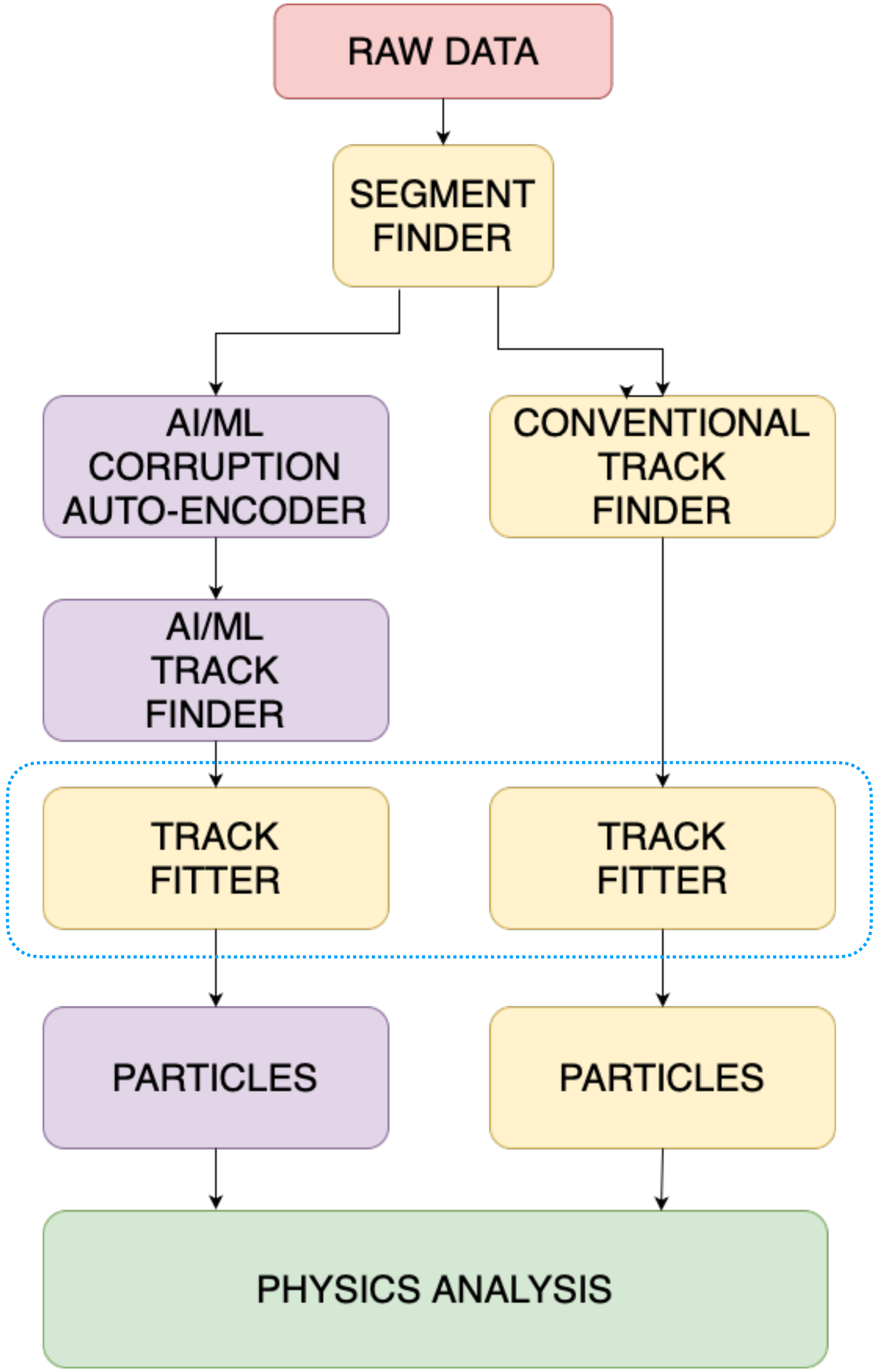


Identify tracks using 6 super layer candidates with pseudo-clusters



Voila!

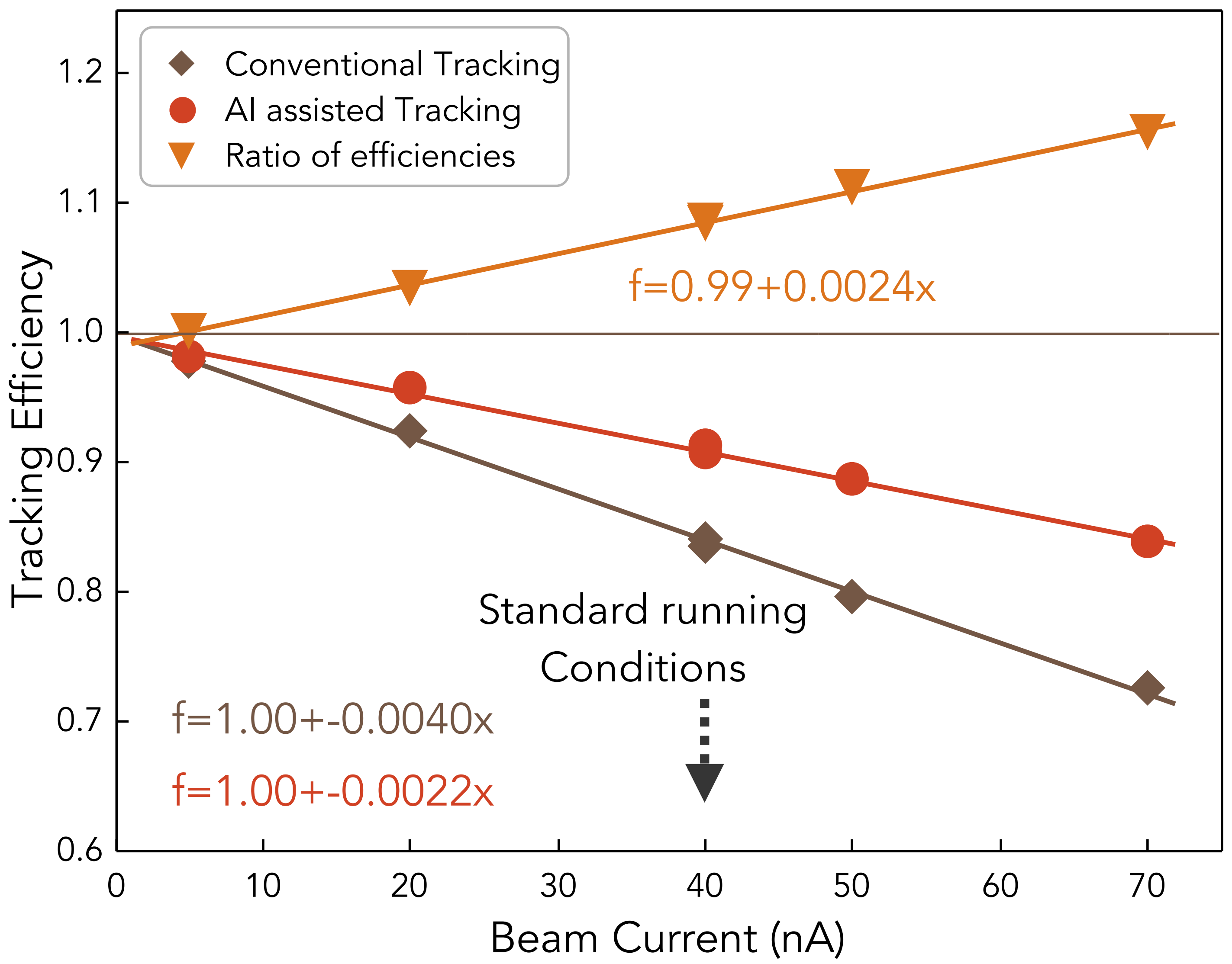
Reconstruction Workflow



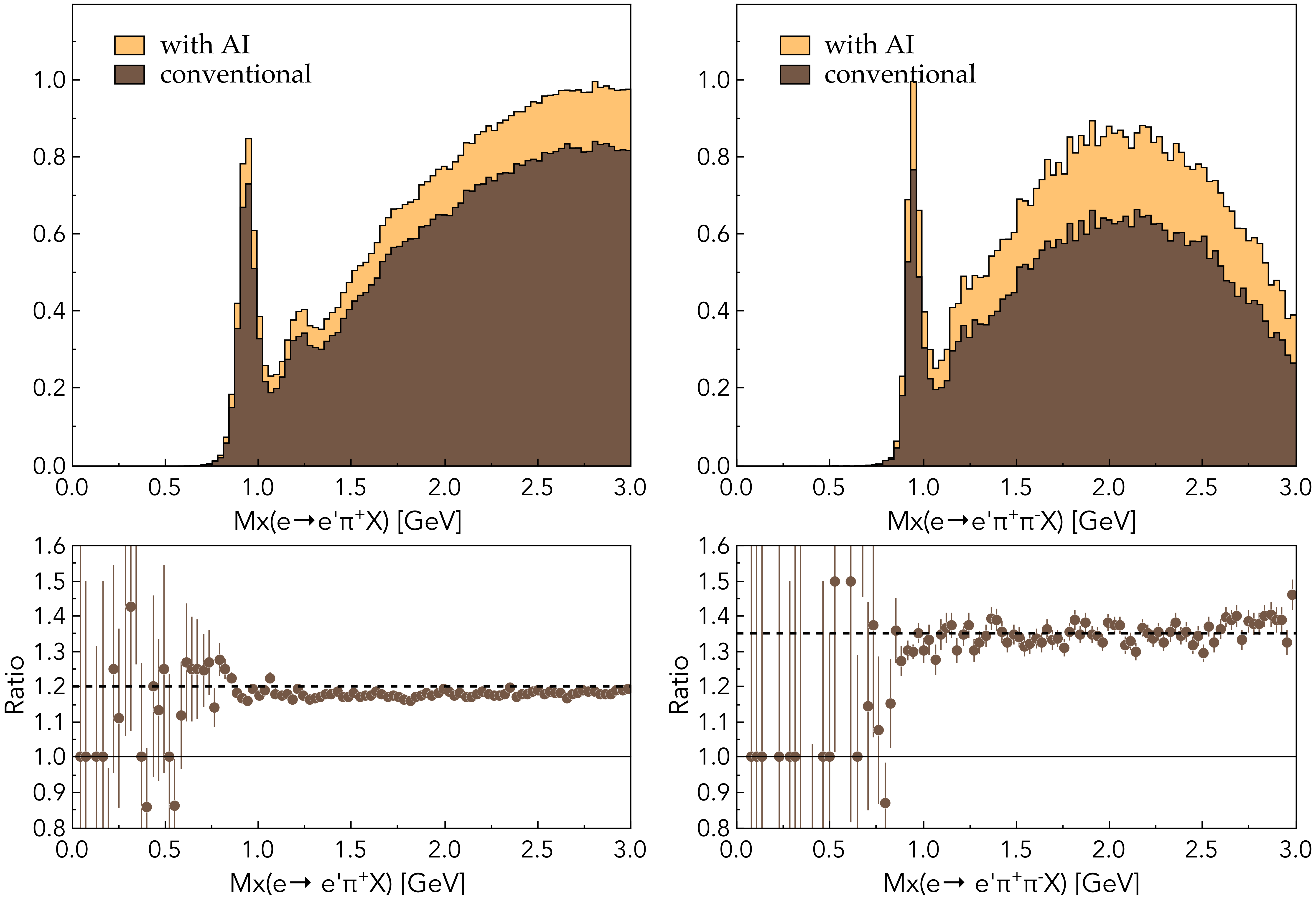
- The reconstruction workflow implements AI/ML track-finding tools that work in parallel with the conventional algorithm.
- AI-assisted tracking uses MLP to identify track candidates from the segments found by conventional segment-finding algorithms.
- The same track fitter (using Kalman-Filter) is used by both workflows.
- Two different outputs are produced from the identified particles from each tracking workflow.
- Physics analyses are performed to assess the efficiency and analyze different event topologies.

Physics Results

- Single particle efficiency increases by ~10% in standard running conditions.
- The impact on physics for a multi-particle final state is dramatic (20% for the two-particle final state and ~35% for the three-particle final state)
- The tracking code speedup is ~30%.



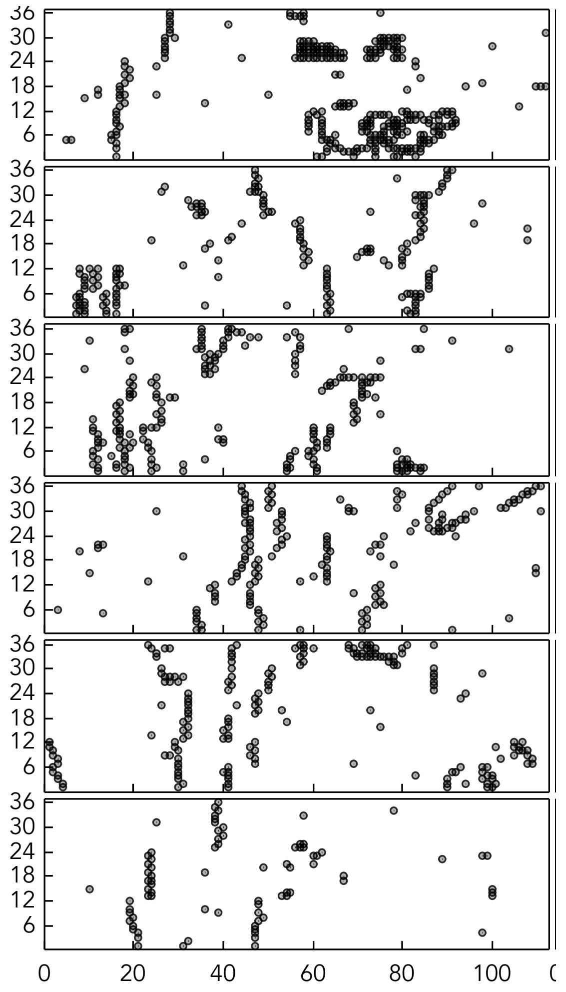
$ep \to e'\pi^+(X)$ $\qquad$ $ep \to e'\pi^+\pi^-(X)$

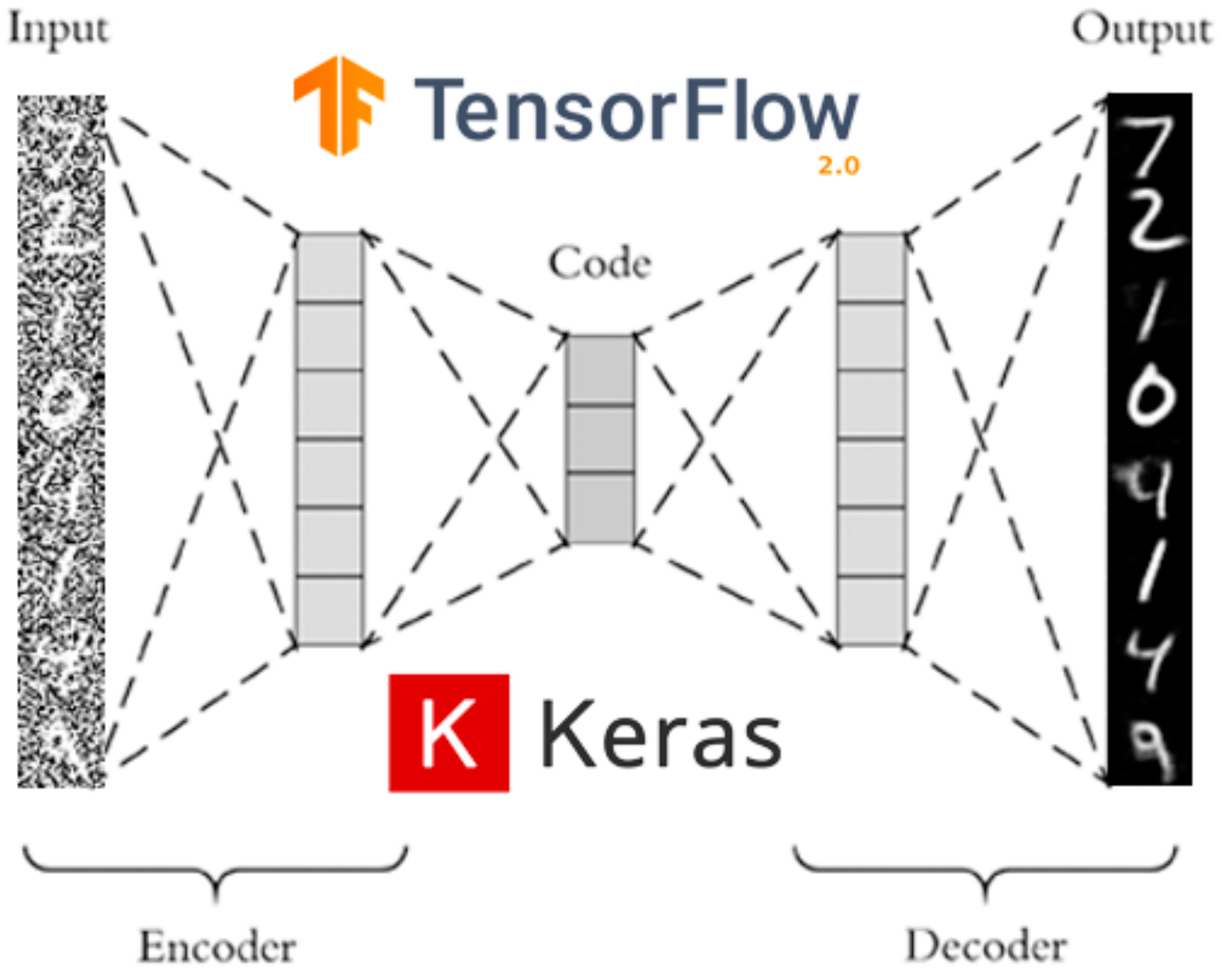


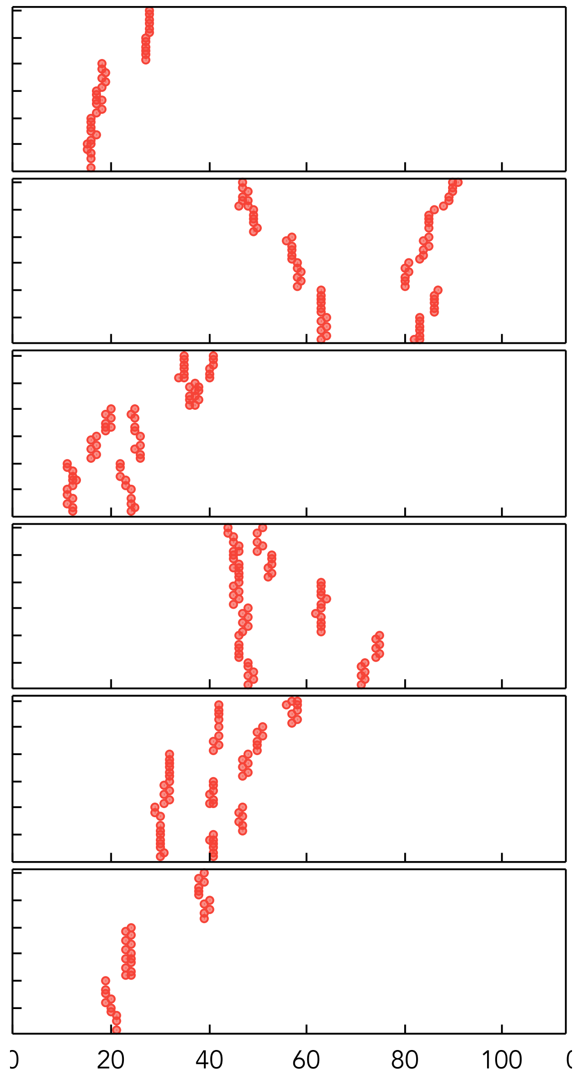
- With increased background (increased combinatorics) the efficiency of conventional tracking drops.
- AI-assisted tracking provides slower efficiency degradation with increased luminosity

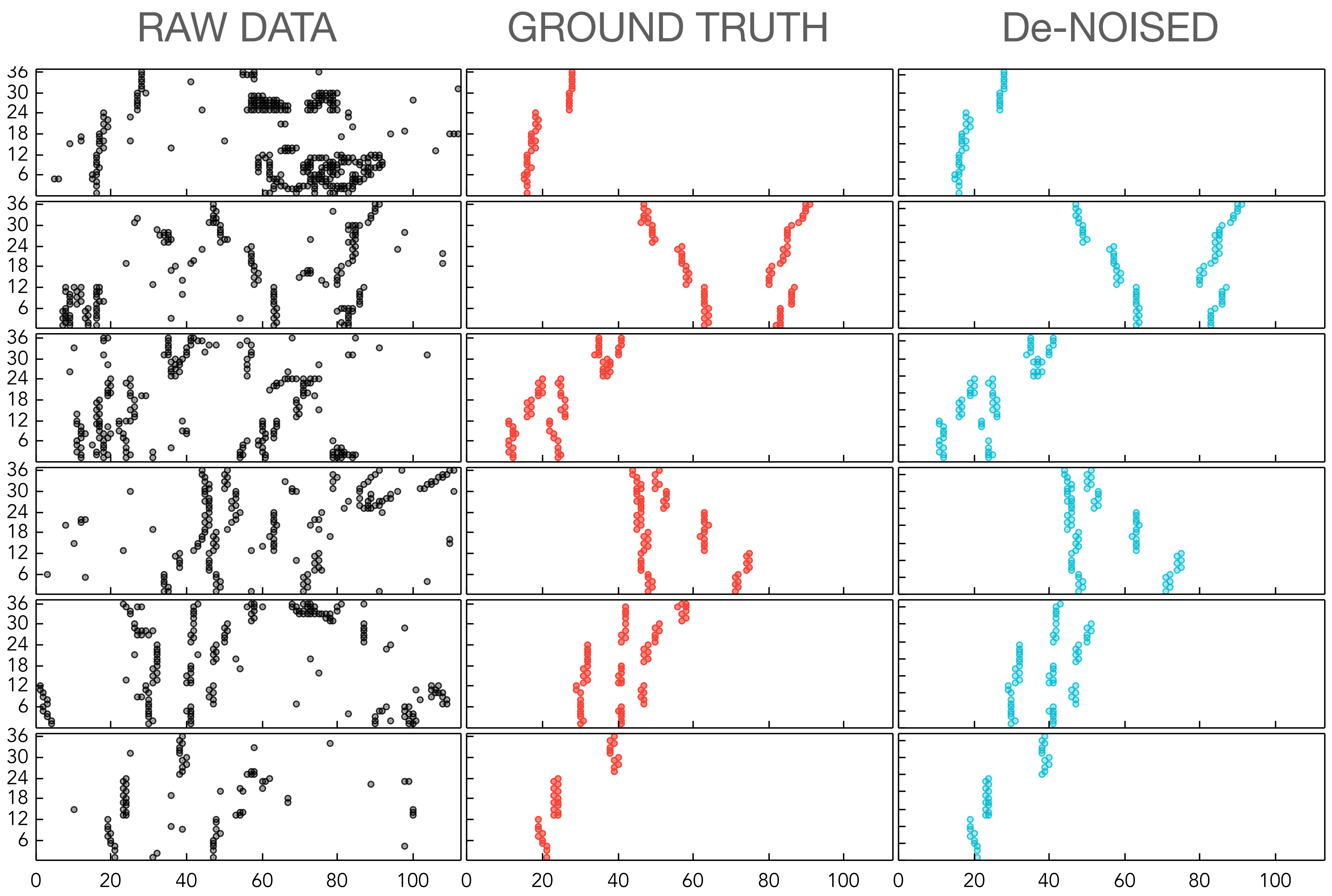
Denoising

- Convolutional Auto-Encoder is used to de-noise raw data from drift chambers.
- The network is trained on reconstructed data with track hits isolated from raw DC hits.
- The network is able to isolate hits that potentially belong to a valid track through drift chambers

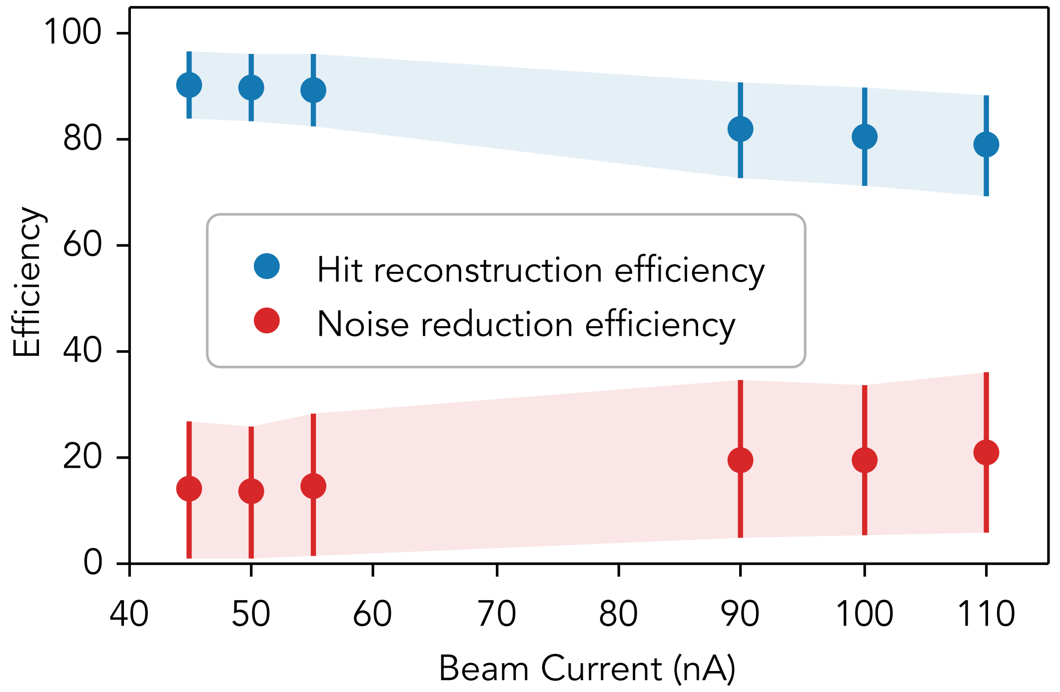


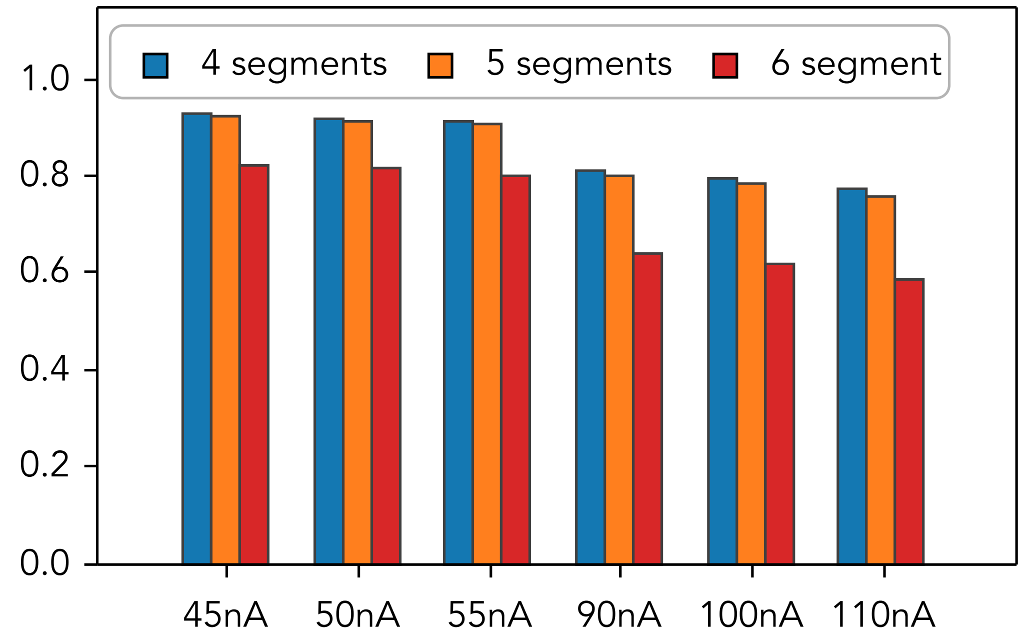




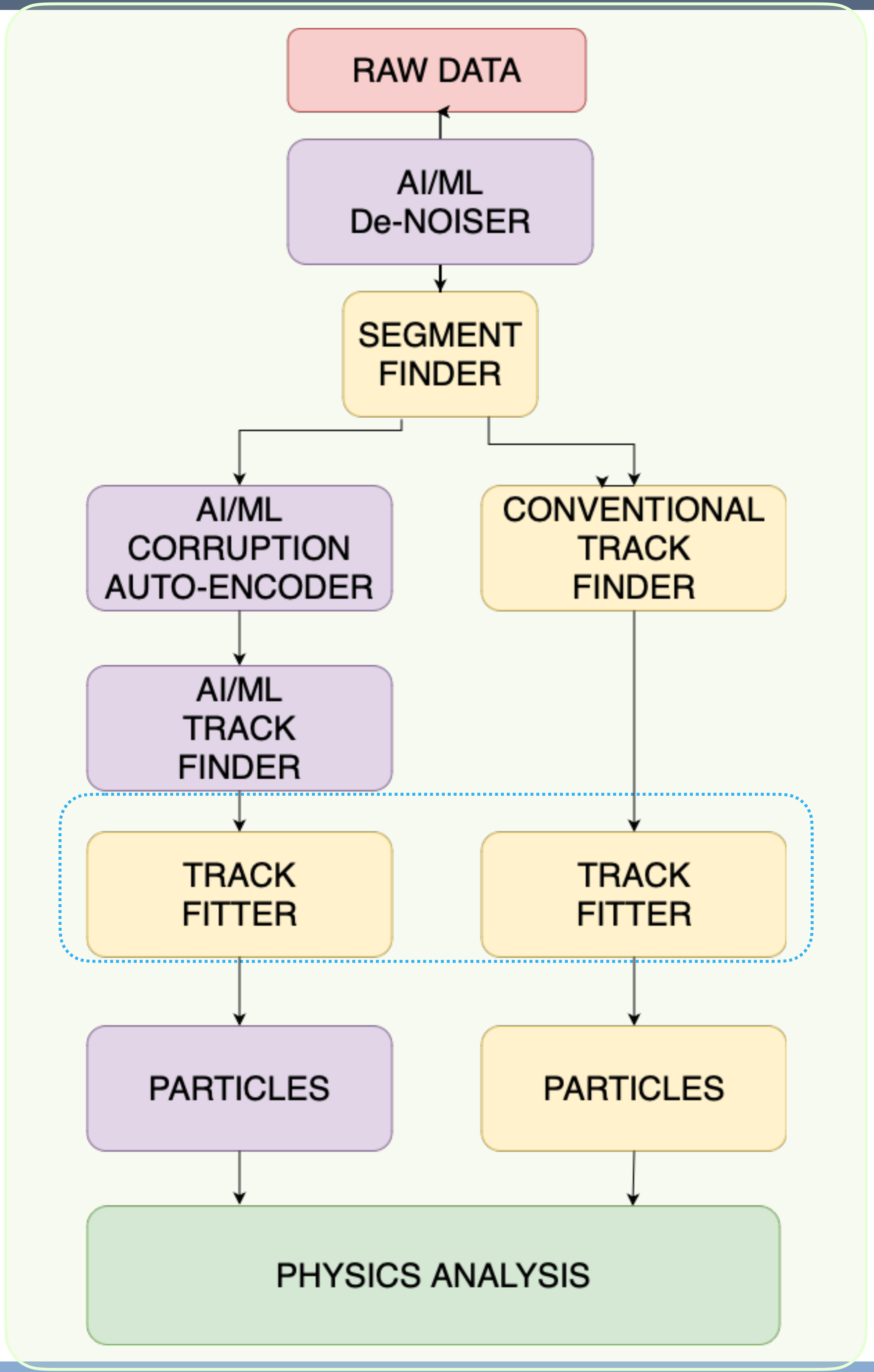


Network Performance Summary

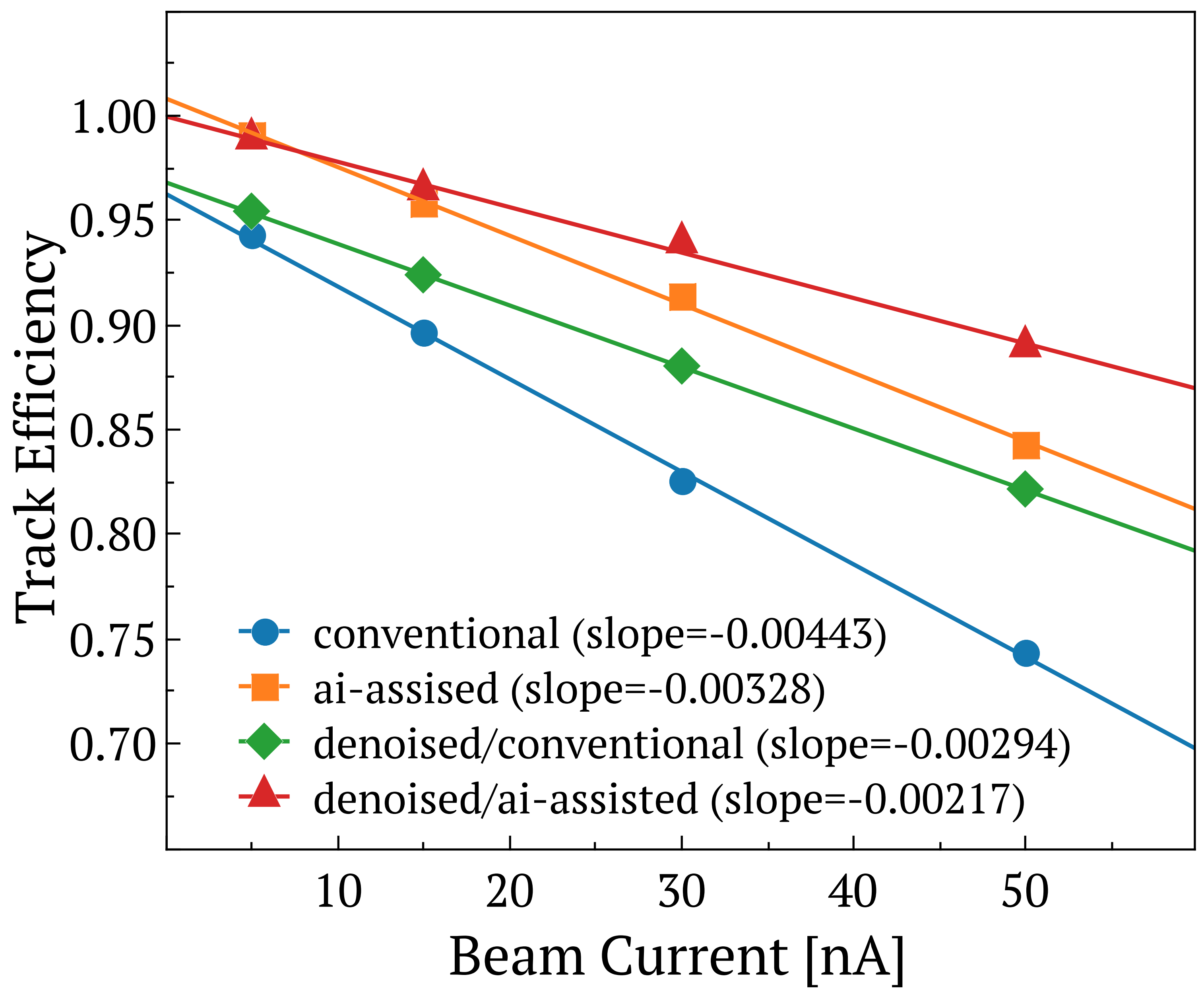




Reconstruction Workflow



- The reconstruction workflow implements AI/ML track-finding tools that work in parallel with the conventional algorithm.
- **First, the data (TDC values from Drift Chambers) is passed through de-noiser**
- Then the conventional algorithm finds segments.
- AI-assisted tracking uses MLP to identify track candidates from the segments found by conventional segment-finding algorithms.
- The same track fitter (using Kalman-Filter) is used by both workflows.
- Two different outputs are produced from the identified particles from each tracking workflow.
- Physics analyses are performed to assess the efficiency and analyze different event topologies.



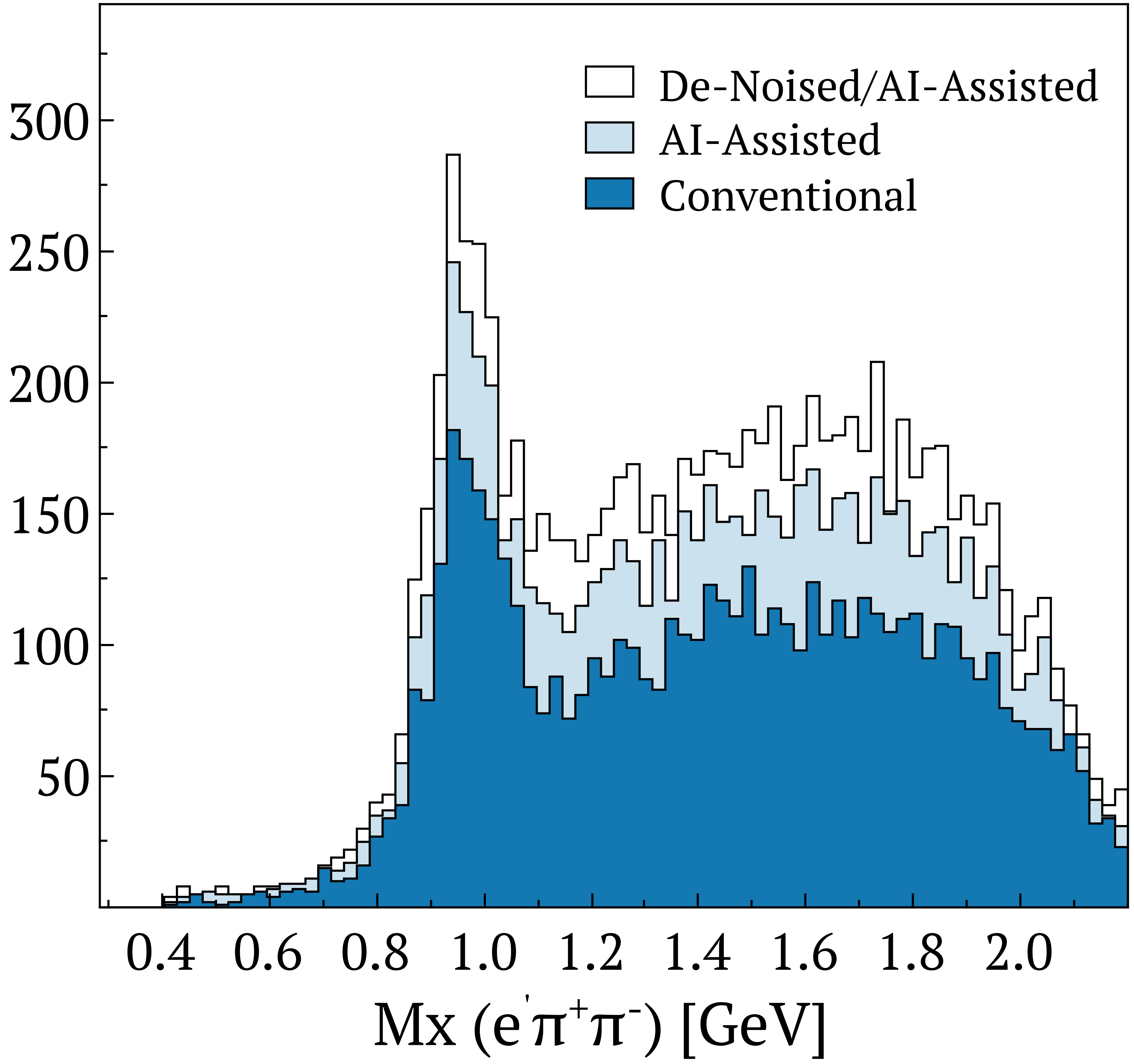
- Single track efficiency with beam current
- De-noising improves the slope with conventional tracking.
- The combination of denoised/ai-assisted yields the best track efficiency, ~18% higher than conventional.
- What is the physics impact?

De-Noising/AI-Assisted Tracking

AI-Assisted Tracking
Increased statistics by ~35%
Compared to Conventional

De-Noised/AI-Assisted Tracking
Increased statistics by ~56%
Compared to conventional

Three detected particles in the event



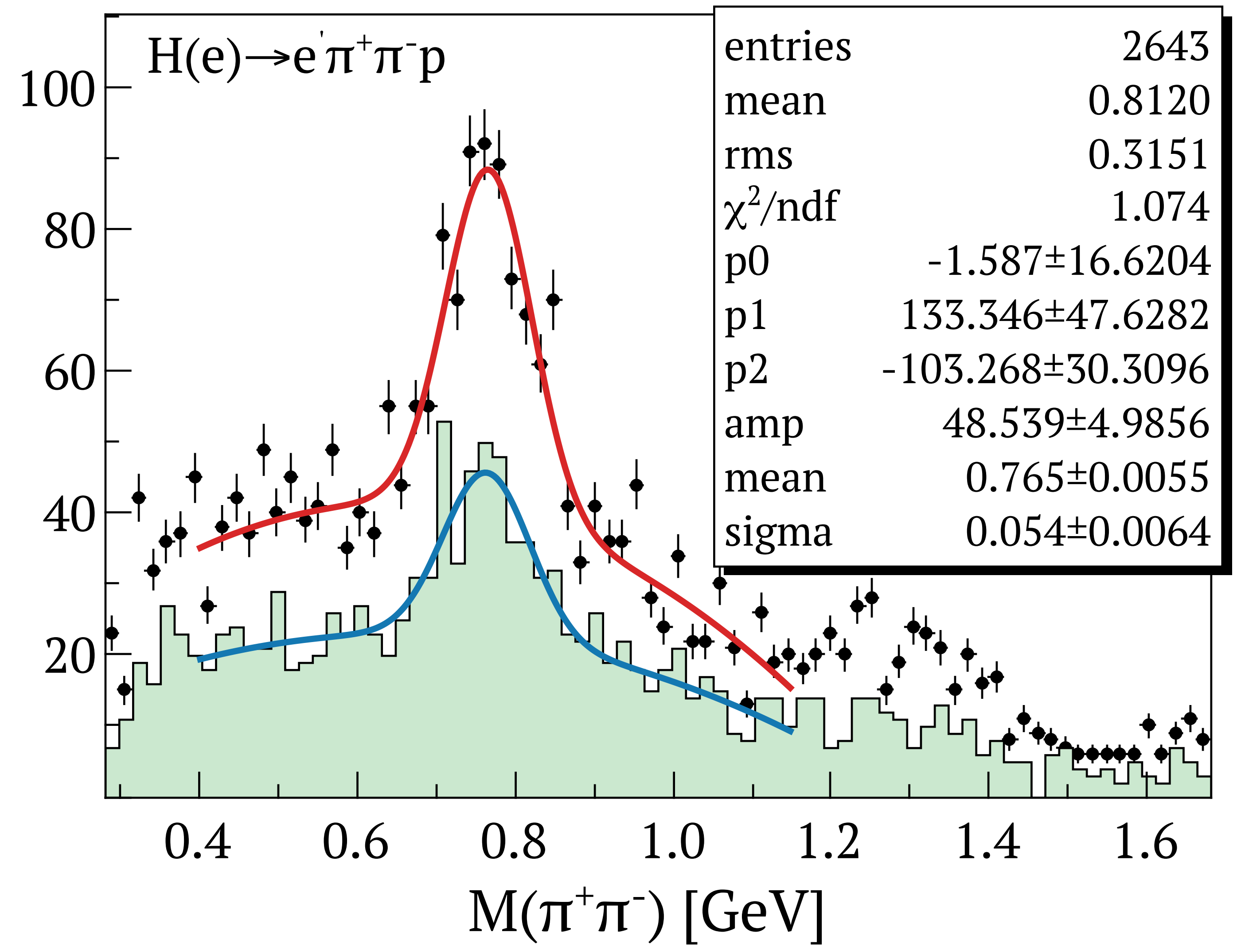
De-Noising/AI-Assisted Tracking

AI-Assisted Tracking
Increased statistics by ~35%
Compared to Conventional

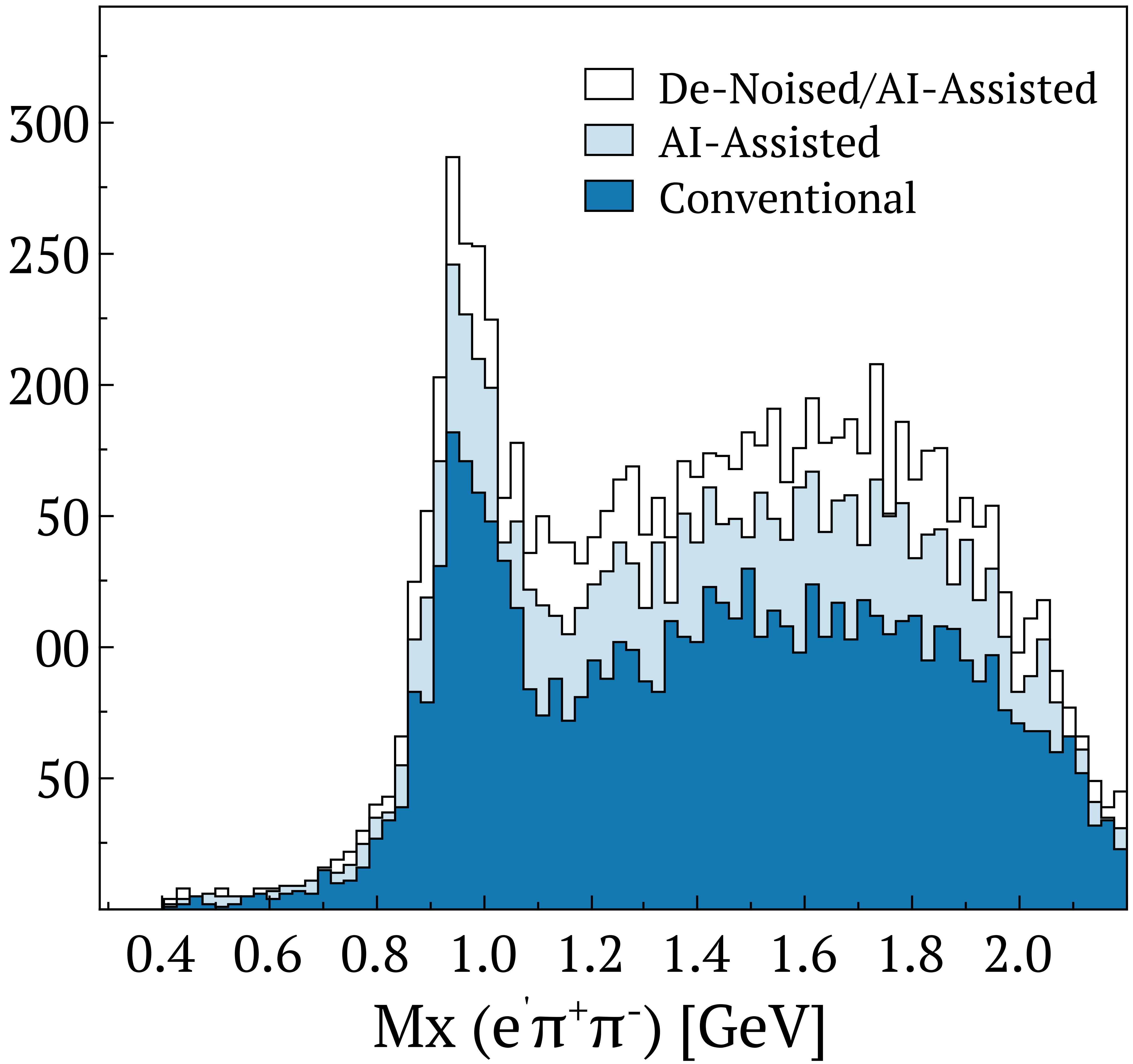
De-Noised/AI-Assisted Tracking
Increased statistics by ~56%
Compared to conventional

4-particle final state with
De-Noised/AI-Assisted
Increased statistics by ~134%
Compared to conventional

Relative Gain 2.34

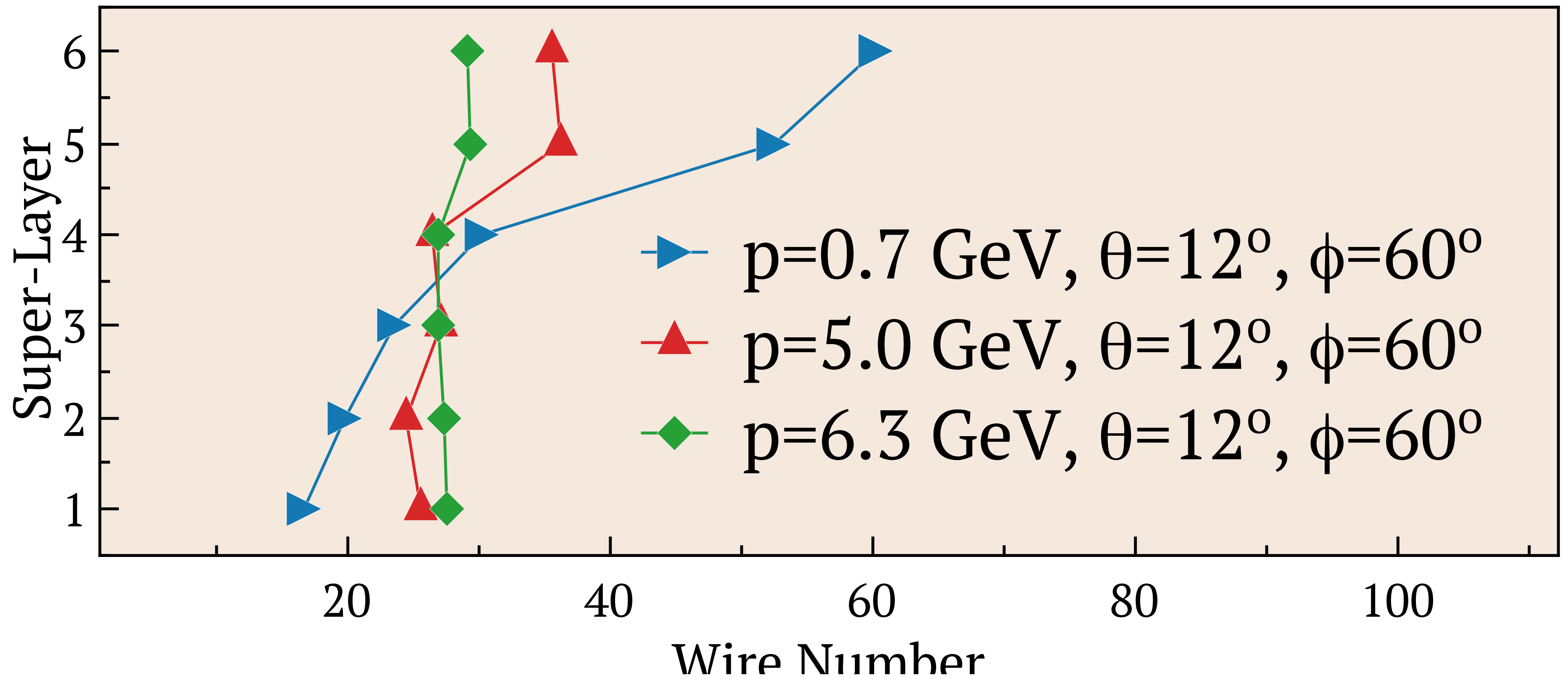


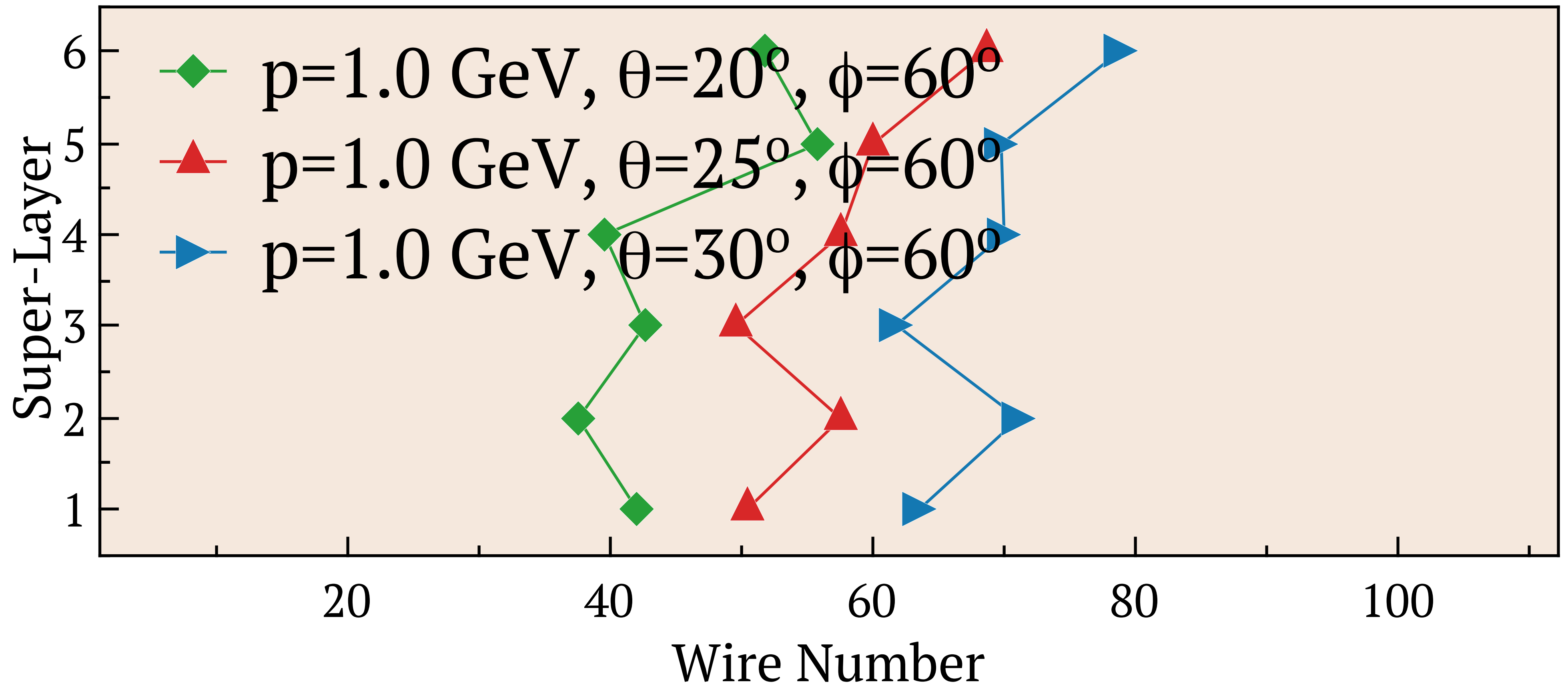
Three detected particles in the event

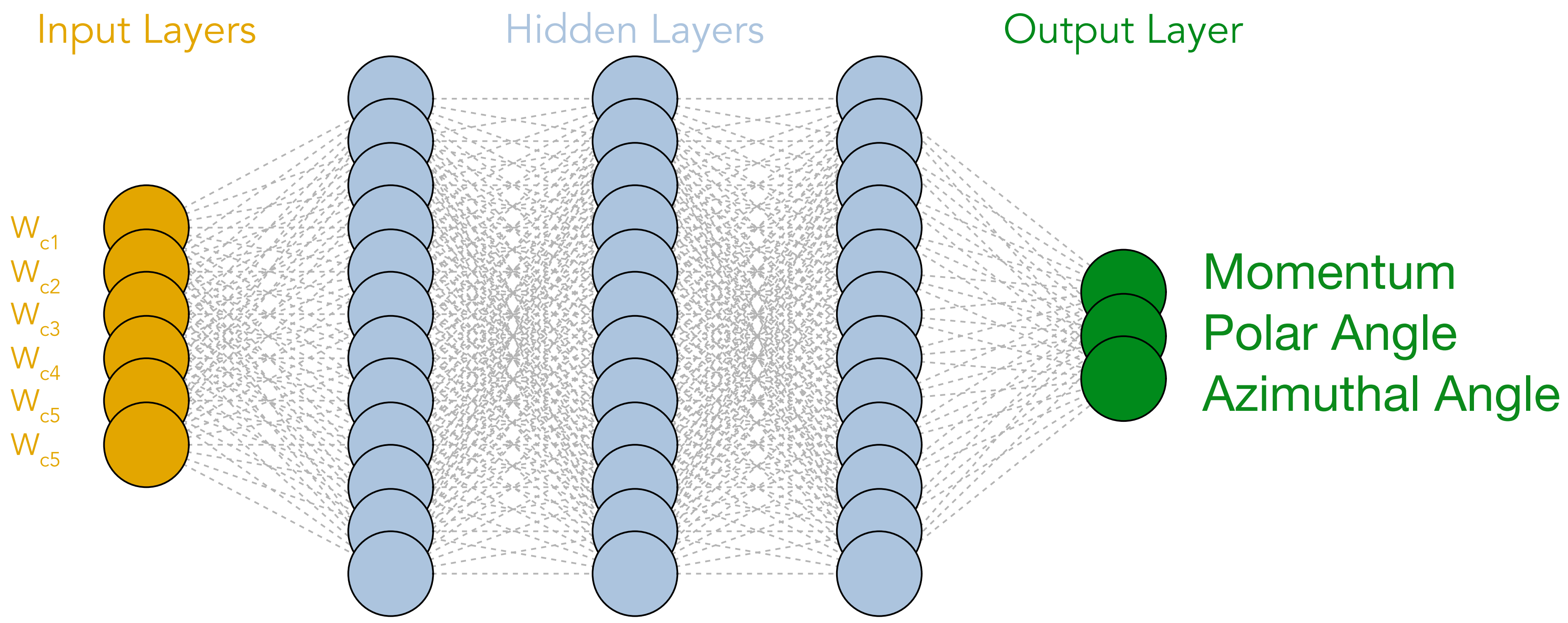


Regression Network

- Each track has unique segment combinations that correspond to particle momentum and direction.

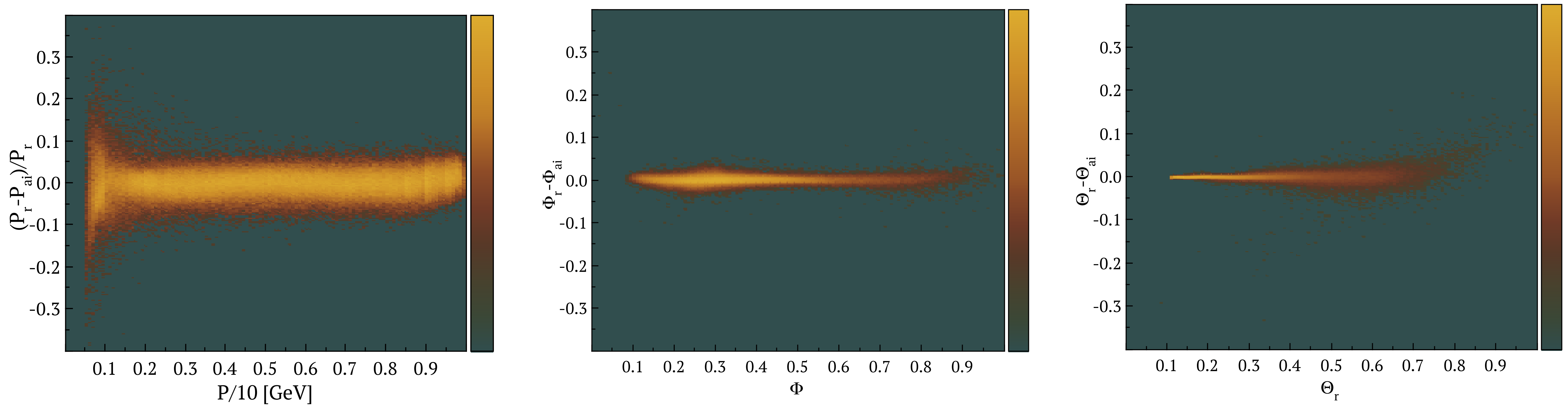




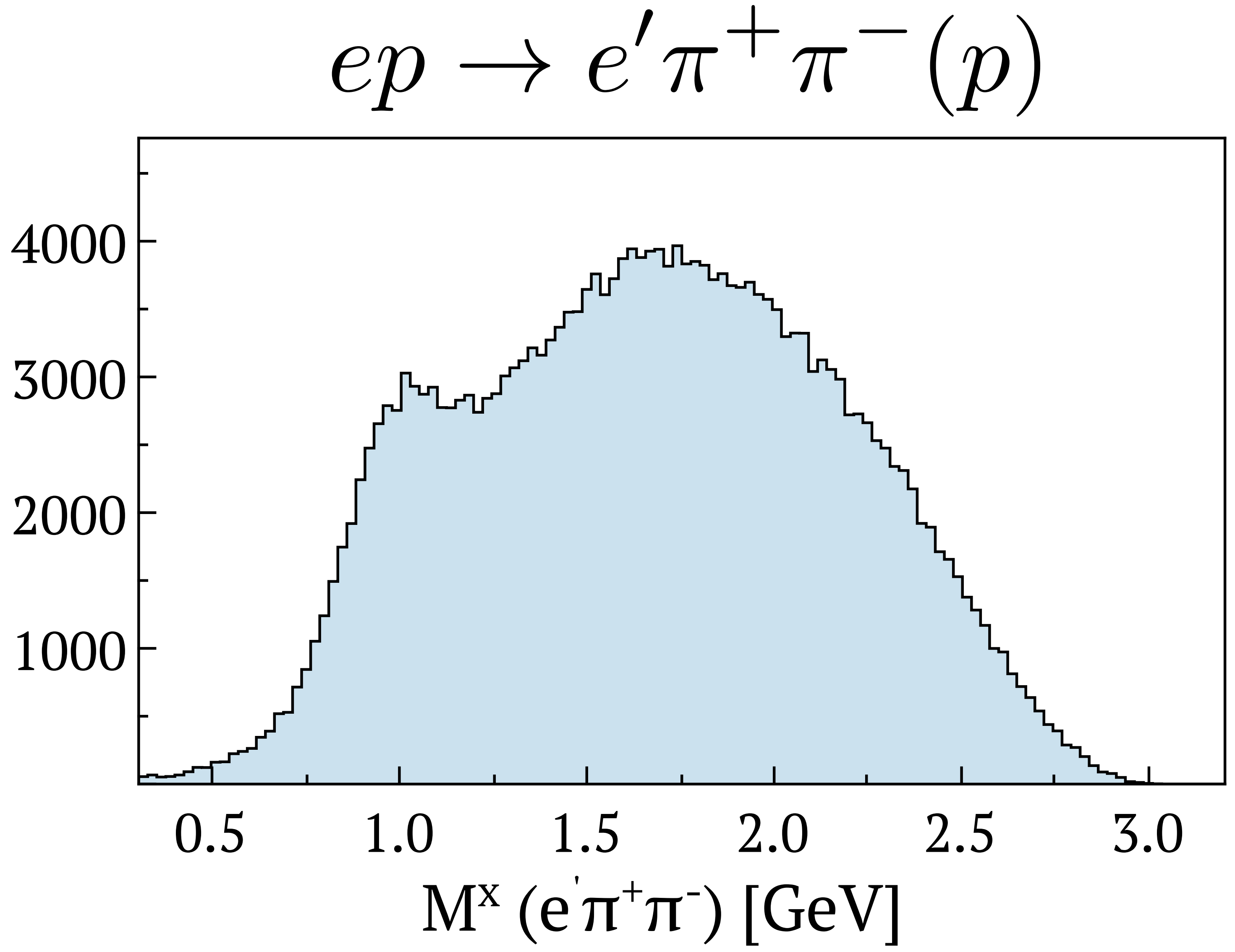


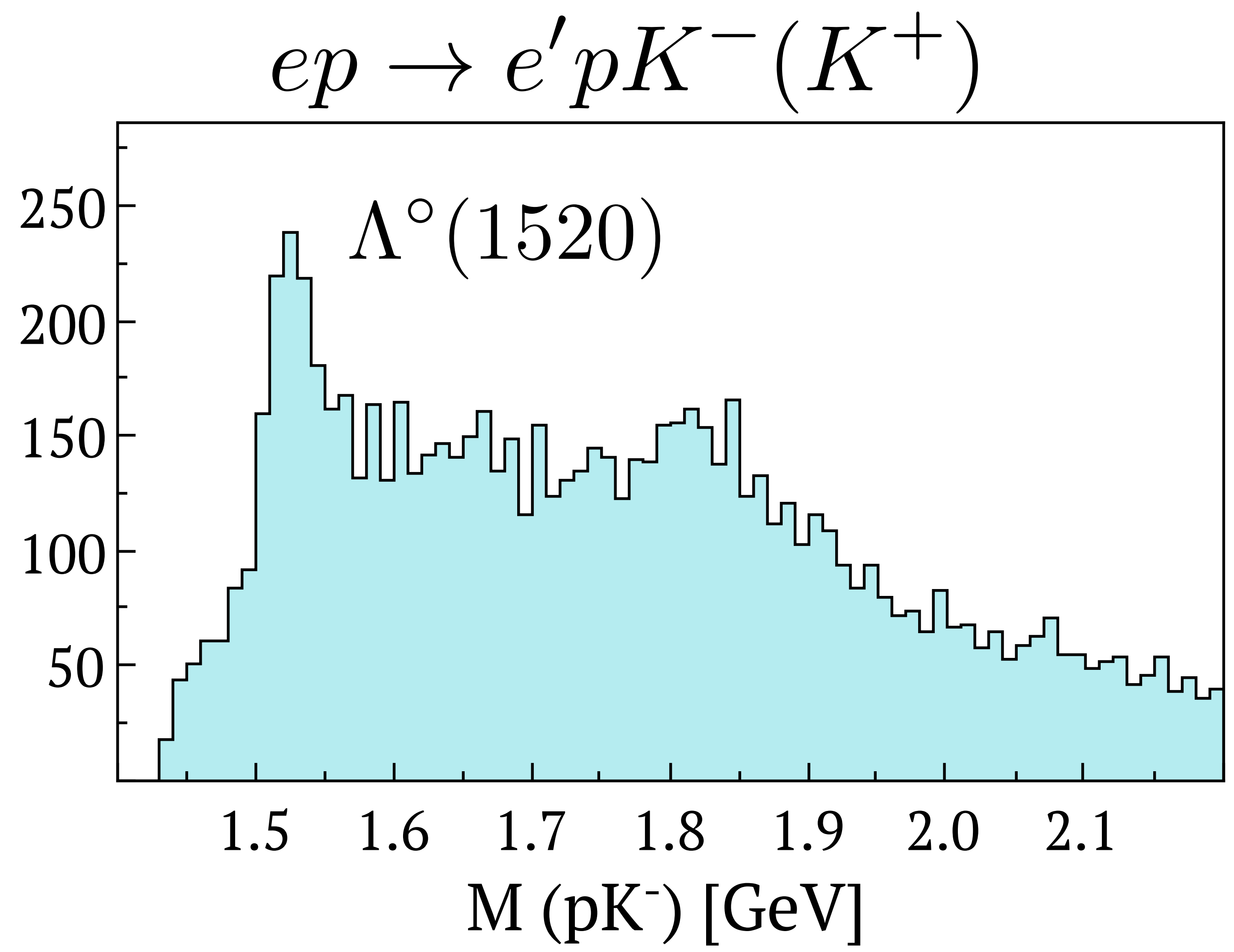
- The network is trained on 6 input parameters, corresponding to to average wire position of segments in each super-layer.
- The output is the momentum of the particle and azimuthal and polar angles.

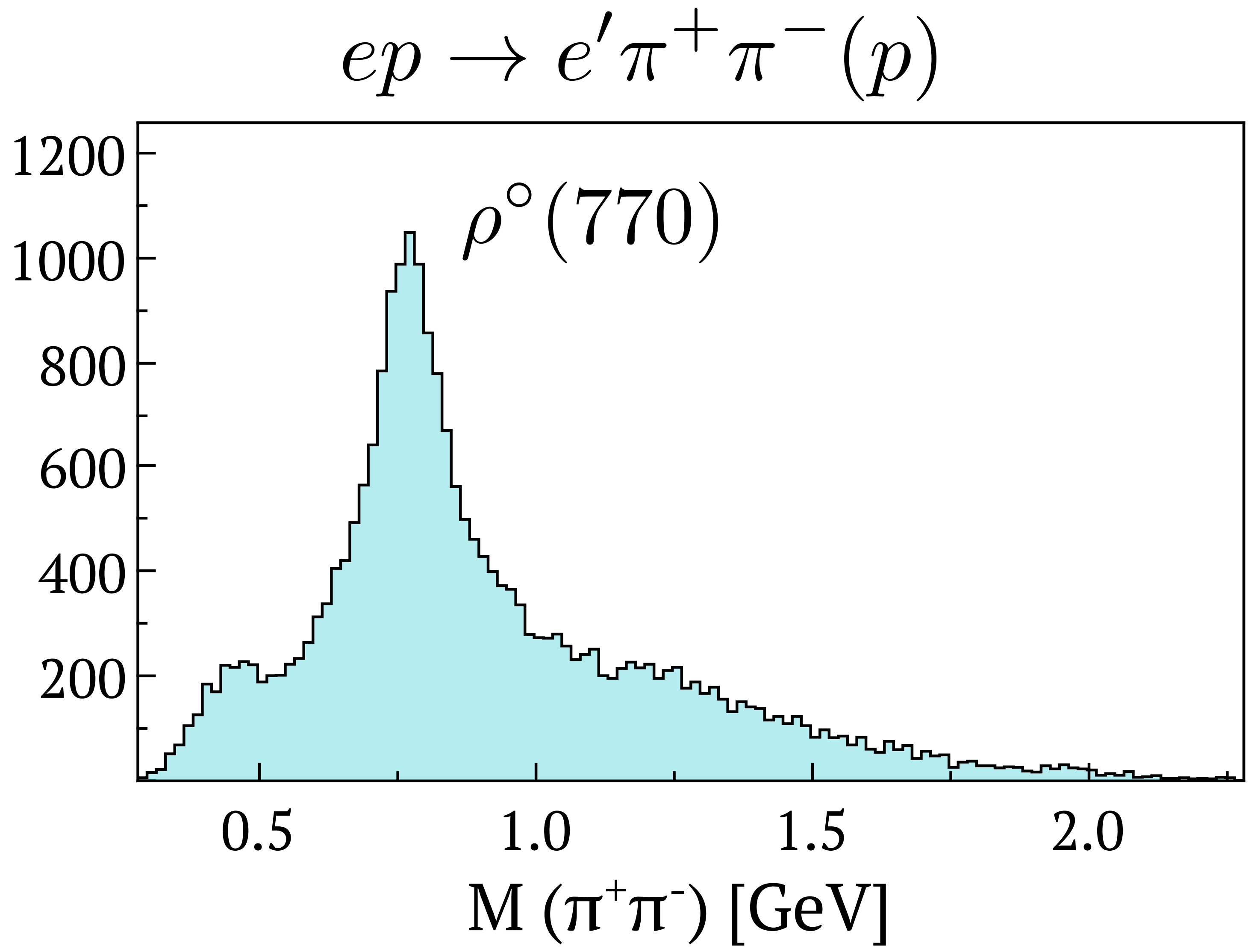
Accuracy of particle parameter predictions:



Physics Analysis



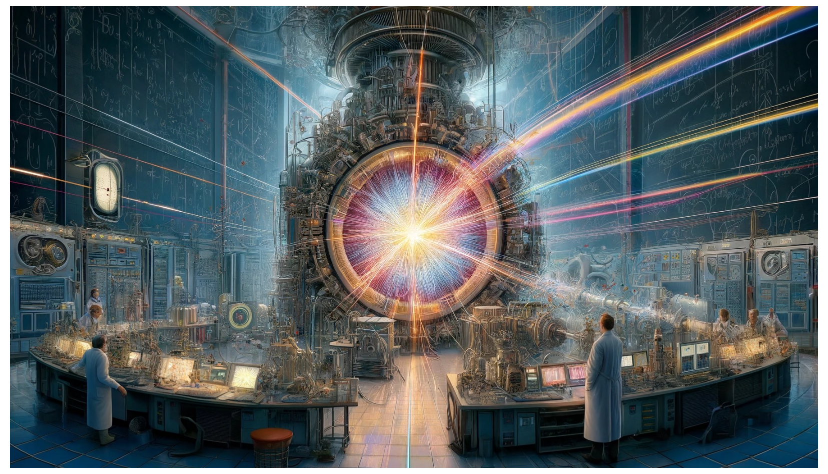




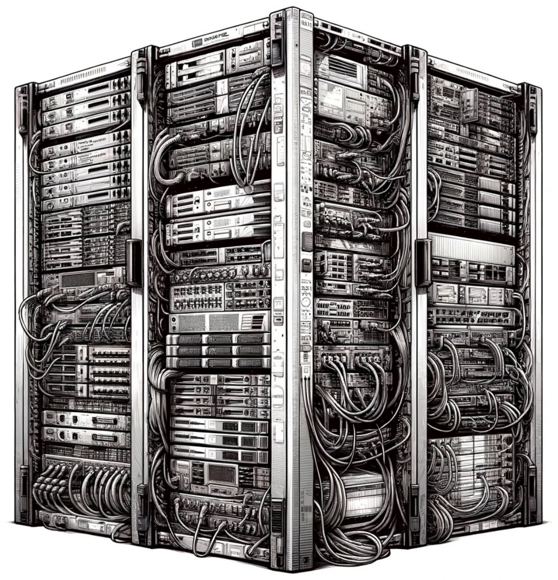
- Physics event reconstruction is based only on TDC hits in Drift Chambers.
- No calibration databases are used
- No Timing information from Time-Of-Flight Counters

- Does not provide Particle ID (feature is coming soon)
- The highest energy negative particle is assumed to be an electron
- Positive particles are assigned pion ID, for other analysis the mass of the desired positive particle is used (proton for example)
- Or kaon mass for the second negative particle (lowest momentum) for lambda analysis

InstaRec performance



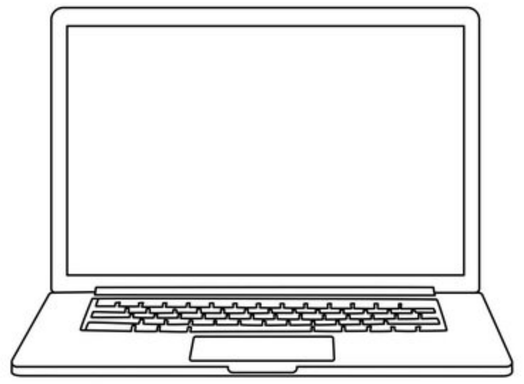
Data Collected at rate:
12,000 interactions per second
100 M events in ~4 hours



Data Processing
768 cores used
10 hours to reconstruct particles

Data Trains
Sort data by interactions
Each output is a specific physics channel
2 hours for sorting

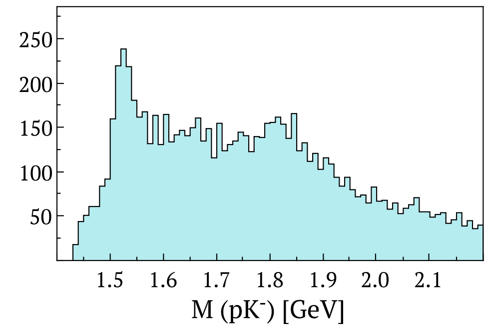
Experiments are conducted for 1-2 month
Processing data from one experiment takes ~3 month

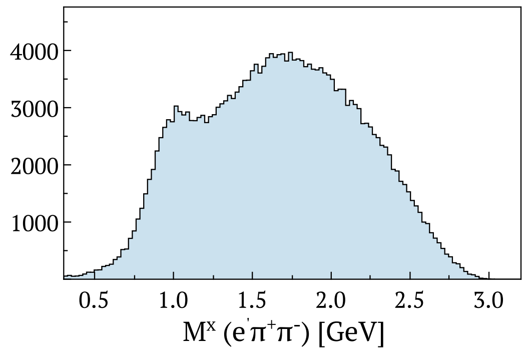


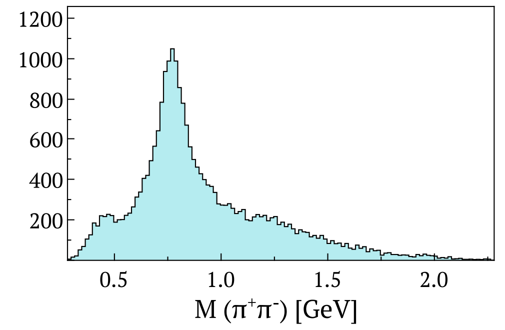
MacBook Pro M3
8 cores

INSTAREC
The track reconstruction running on a laptop
Reconstructs physics final states and sorts them

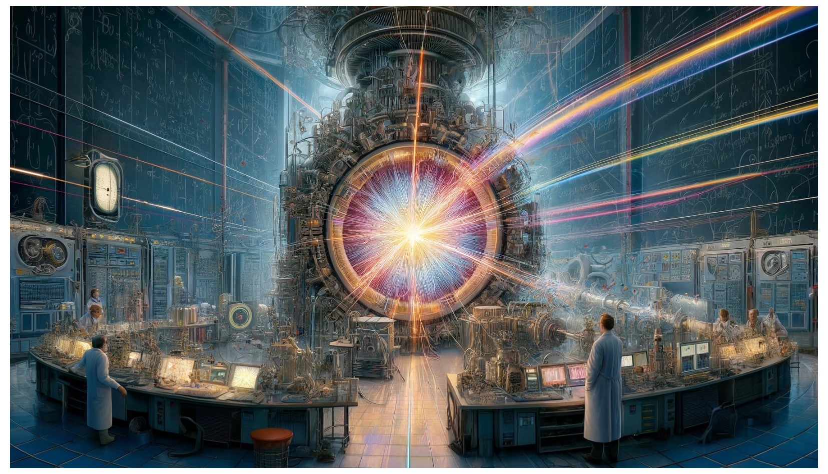
? Hours



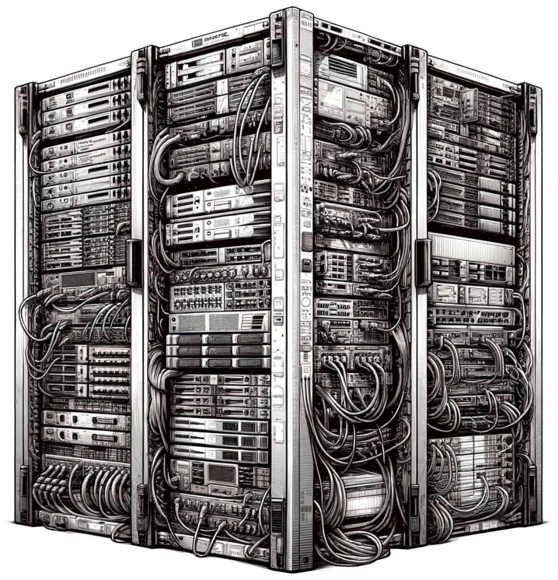




InstaRec performance



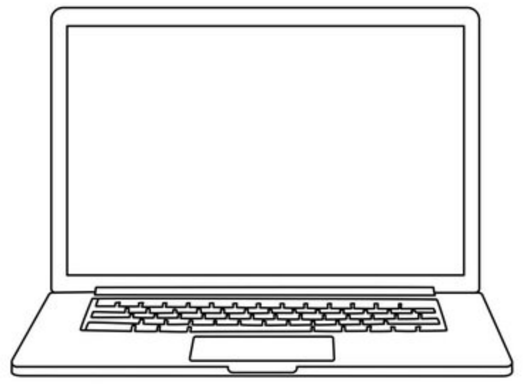
Data Collected at rate:
12,000 interactions per second
100 M events in ~4 hours



Data Processing
768 cores used
10 hours to reconstruct particles

Data Trains
Sort data by interactions
Each output is a specific physics channel
2 hours for sorting

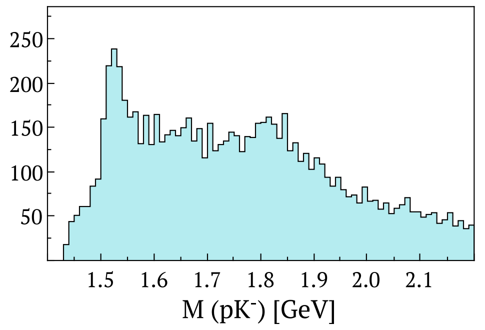
Experiments are conducted for 1-2 month
Processing data from one experiment takes ~3 month

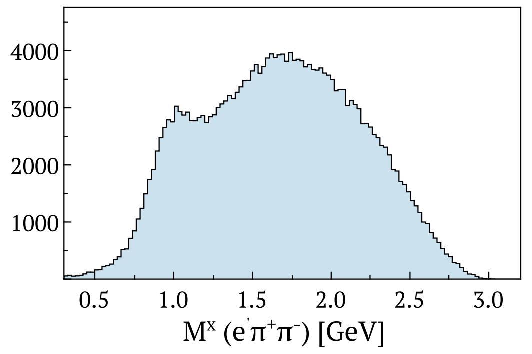


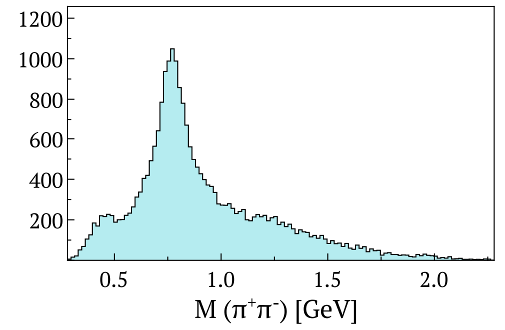
MacBook Pro M3
8 cores

INSTAREC
The track reconstruction running on a laptop
Reconstructs physics final states and sorts them

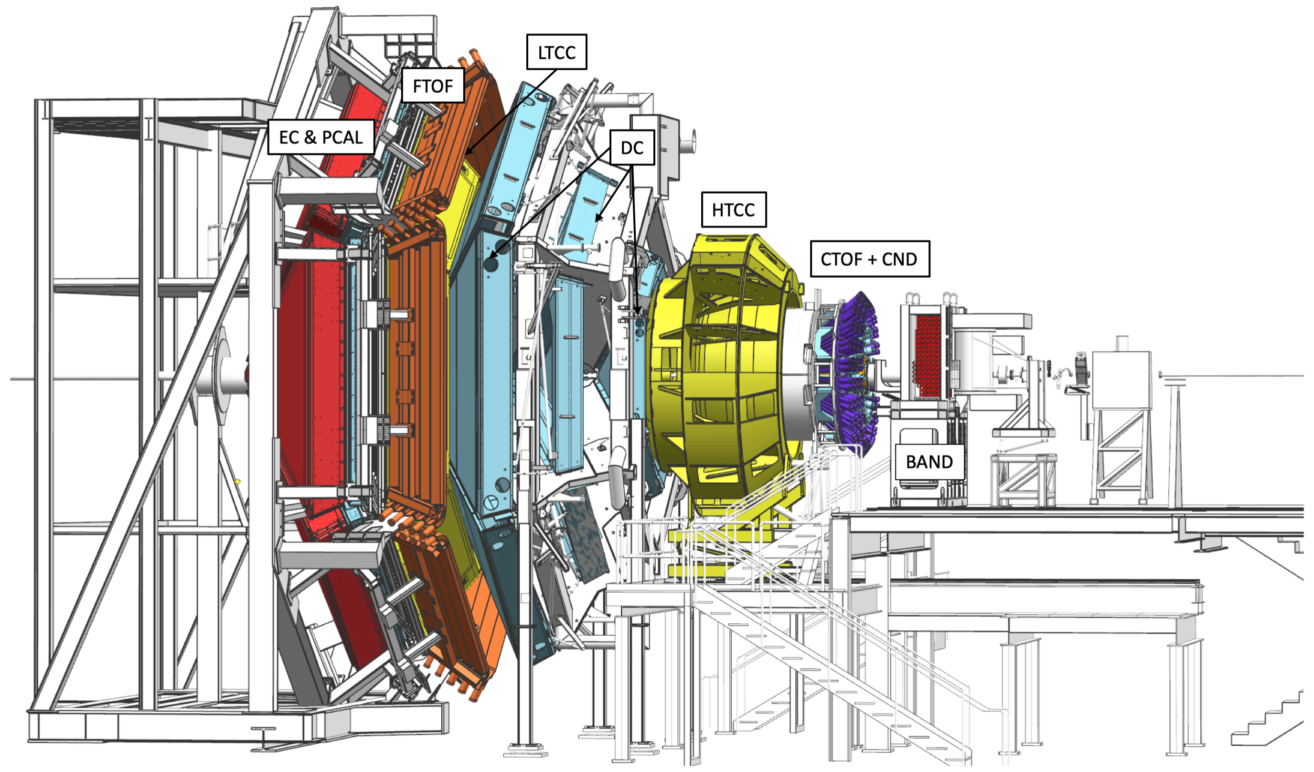
25 Minutes



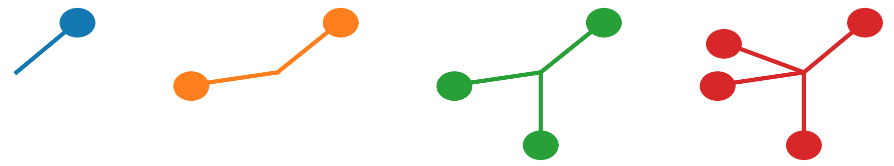


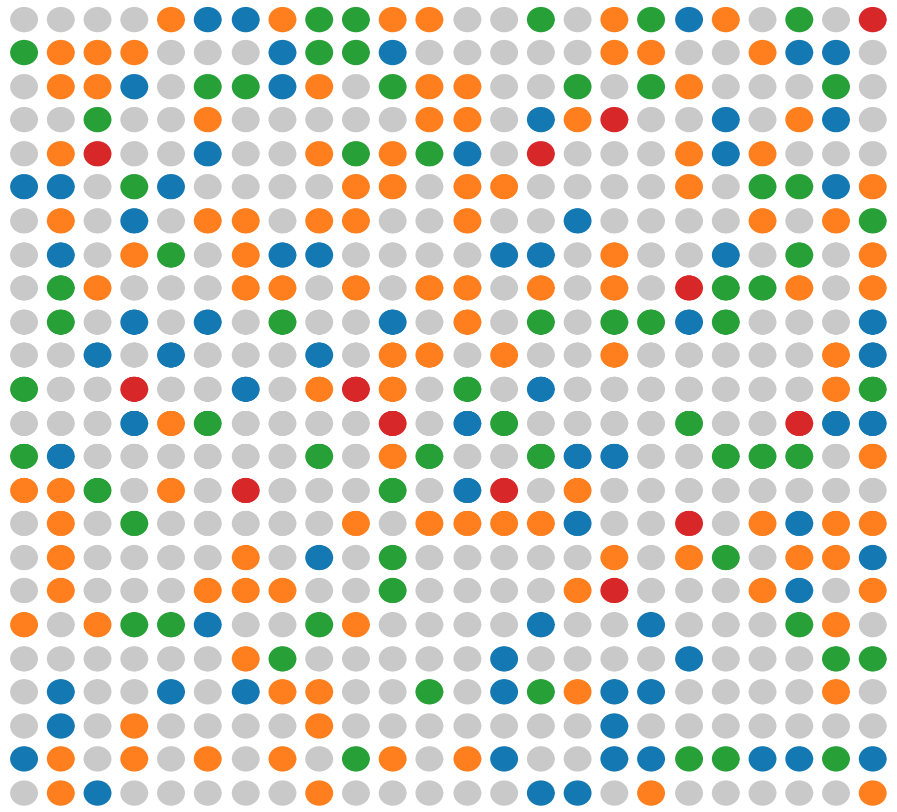


Online Workflow



- DAQ rate 12,000 Hz
- Events recorded in chronological sequence
- Every event has to be reconstructed and then separated by event topology for each analysis group in the collaboration

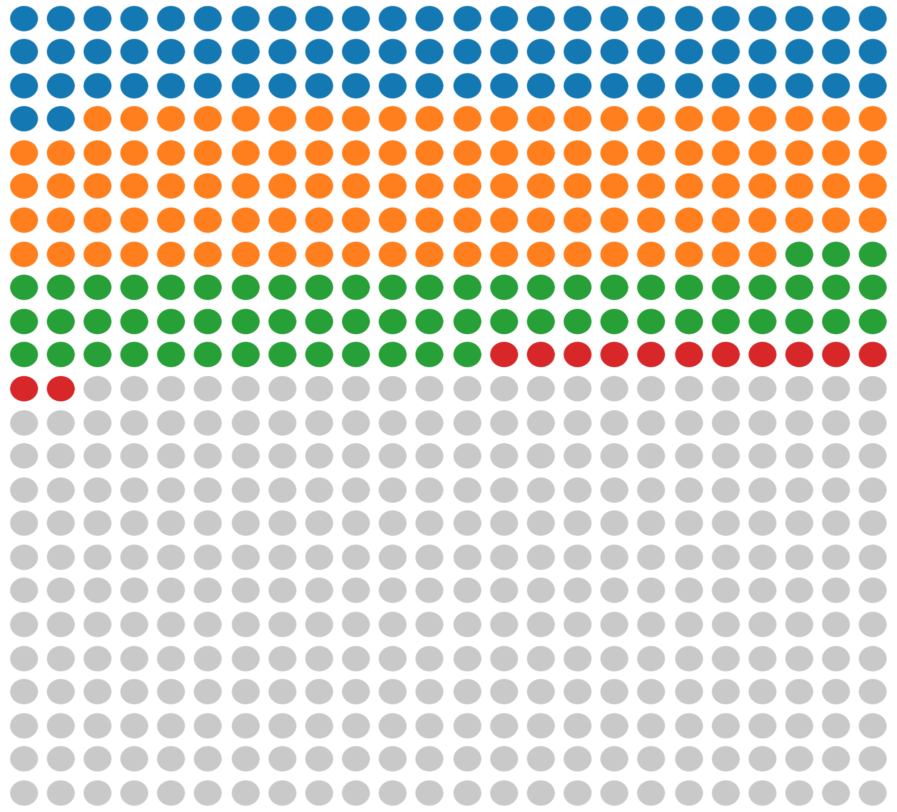




Event File

InstaRec

Rate: 96,000 Hz on a Laptop

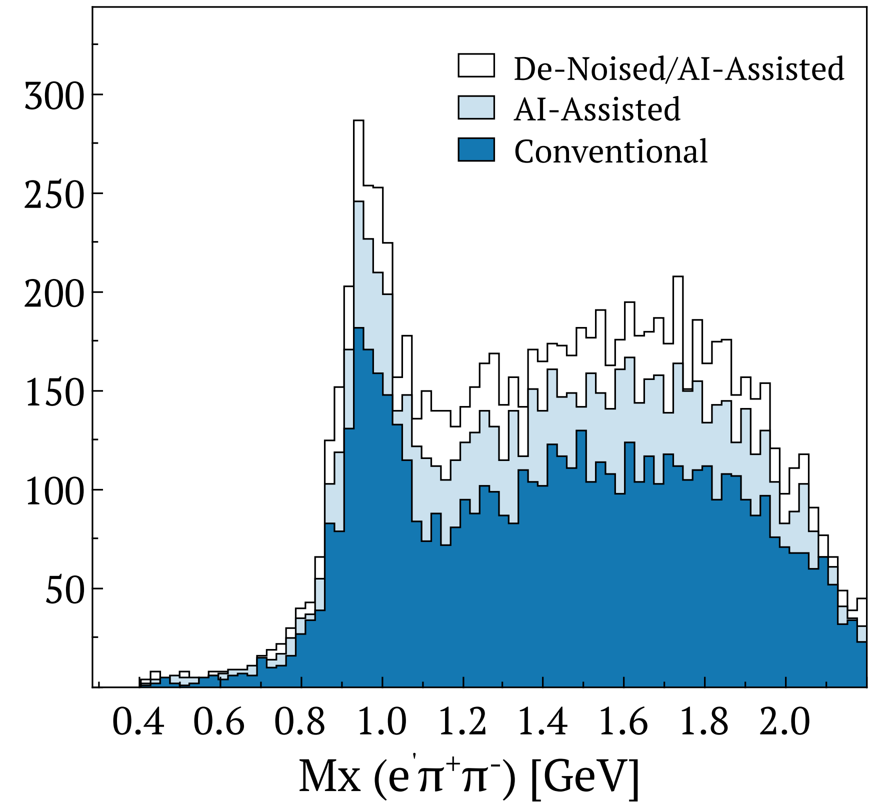


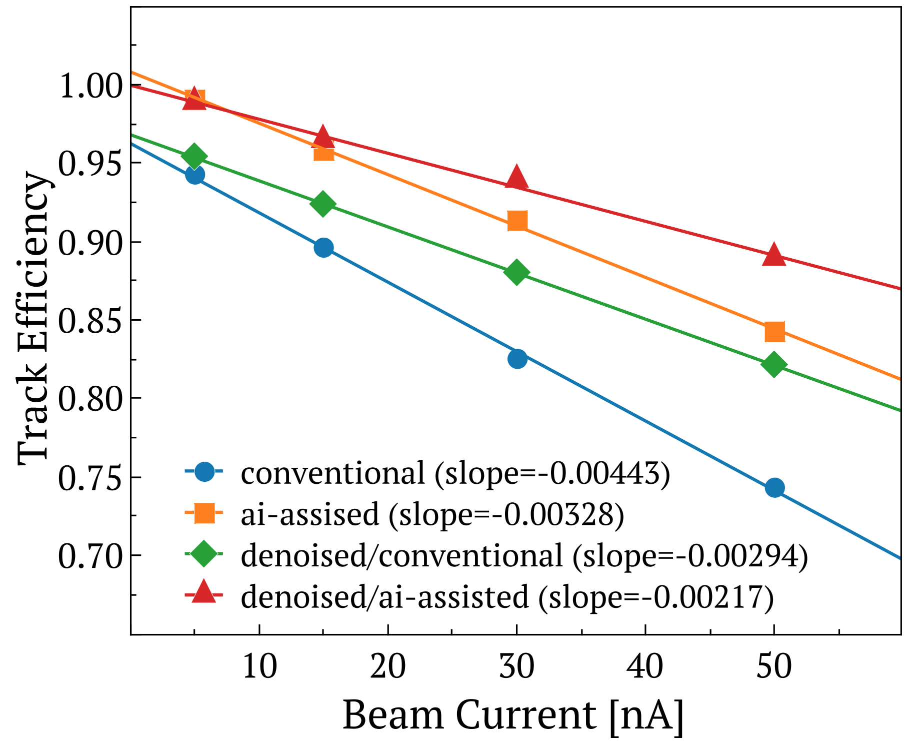
- Events in the output are sorted by topology
- Reconstructed events don’t have to be post-processed for each analysis group.
- Trigger impurities are removed, significant speed-up of data processing
- Data monitoring and calibration become possible in real-time.

Conclusions

- AI-Assisted Tracking/De-Noising
  - Already implemented in the standard CLAS12 workflow
  - Increase in single particle efficiency
  - Improvement in luminosity dependence of tracking efficiency
  - Yields to increased physics statistics
- AI-based fast reconstruction is being developed (**InstaRec**):
  - Will be integrated with the online reconstruction for data quality monitoring
  - Level-3 trigger will use this to tag events for reconstruction algorithms
  - Particle Identification is being developed
- Future:
  - Experience in CLAS12 can be applied to upcoming experiments at Jlab
  - This is the future of streaming readout, where event identification has to be done in real-time

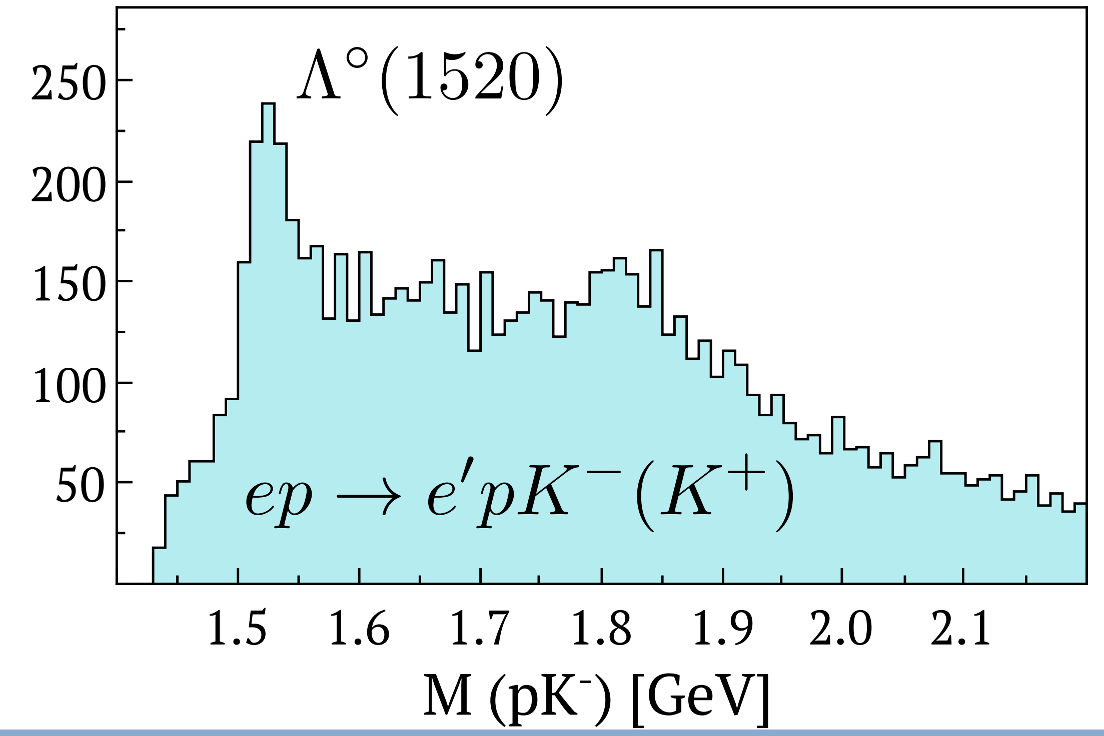
The work presented here is done In Java





We are not in AI prototyping stage, we are in the age of AI

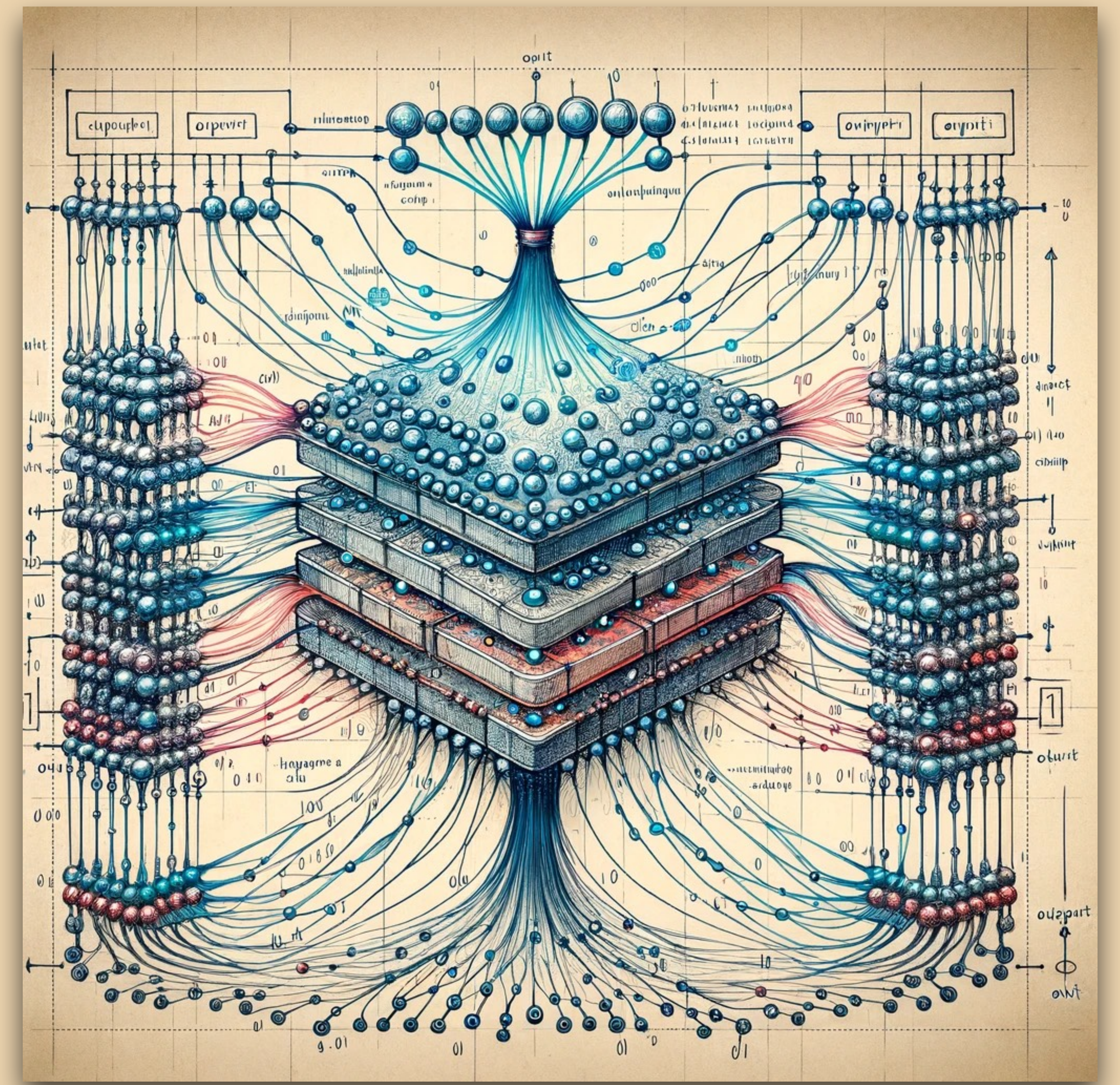
And we will be glad to share our experience with other Halls and Experiments



BACKUP SLIDES

01000111
01000001
01000111
01001111

DALL.E 3



What is Deep Learning?

Solving problems that do not have an analytical solution

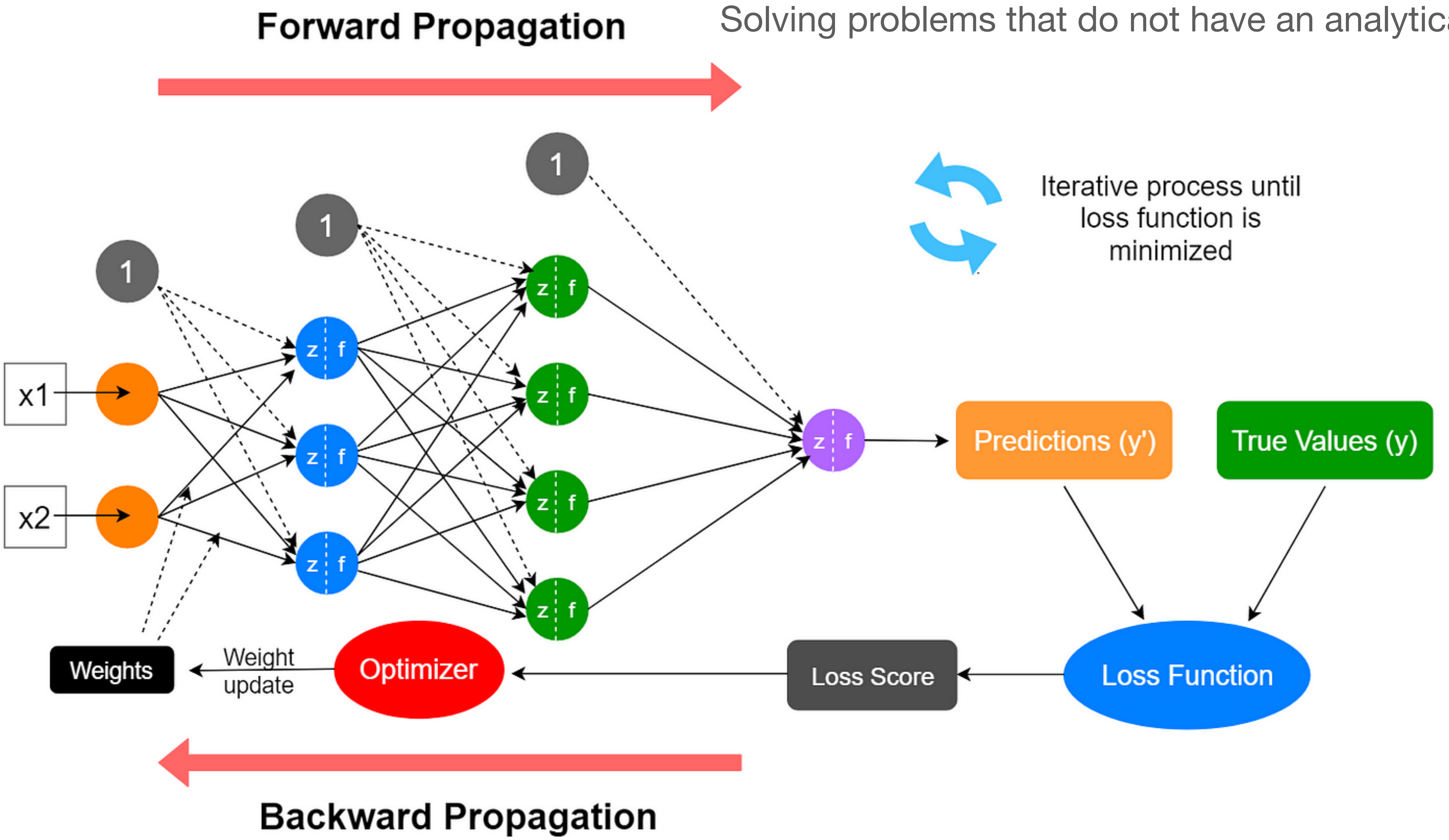


Image Generation DALL-E

Here are the images showing a footballer, a goalkeeper, and a defender playing football on Mars.

Here are the images depicting four people playing ice hockey on the moon, each wearing Nike brand skates. Earth and Saturn are visible in the background.

InstaRec performance

2,135 kHz event reconstruction for the whole world.

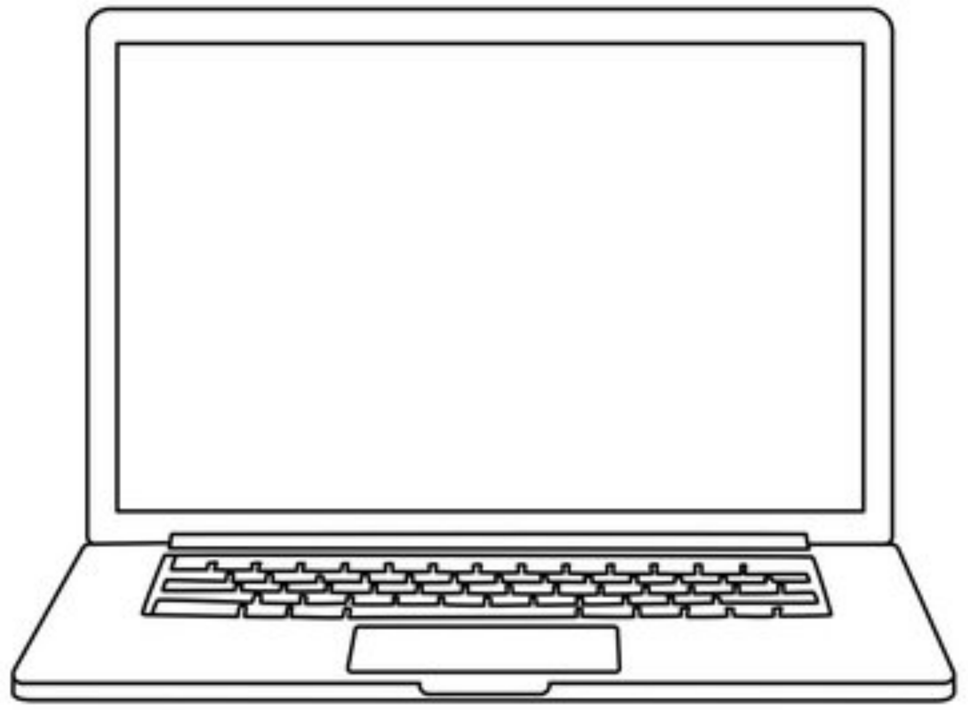
Earth's population almost doubled since then so ~4,000 kHz

~4000 kHz/96 kHz = **42** Laptops

Coincidence?
Or the answer to everything.

Running InstaRec on Laptop

12 threads



Reconstruction Rate 8 kHz (M3) per core (96 kHz multithreaded)

Benchmark

- M1 ARM processors are more performant compared to x86 counterparts
- Simple matrix multiplication code (C++) tested single-treaded NxN matrix multiplications
- M1 outperforms AMD (IFARM1901) by a significant margin.
- Maybe moving to ARM machines in the future will provide better performance for AI applications?