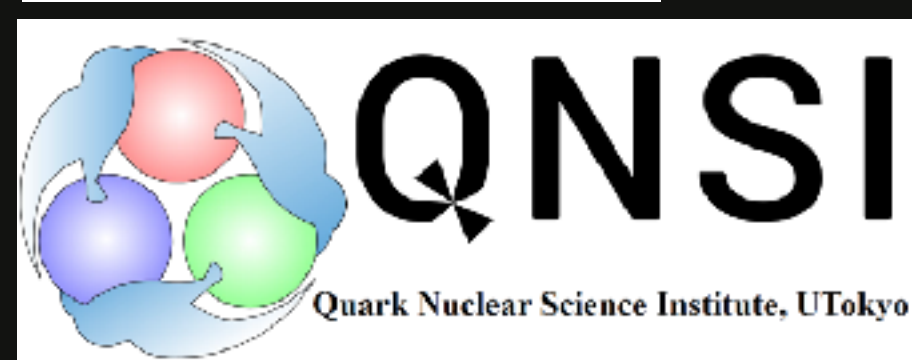
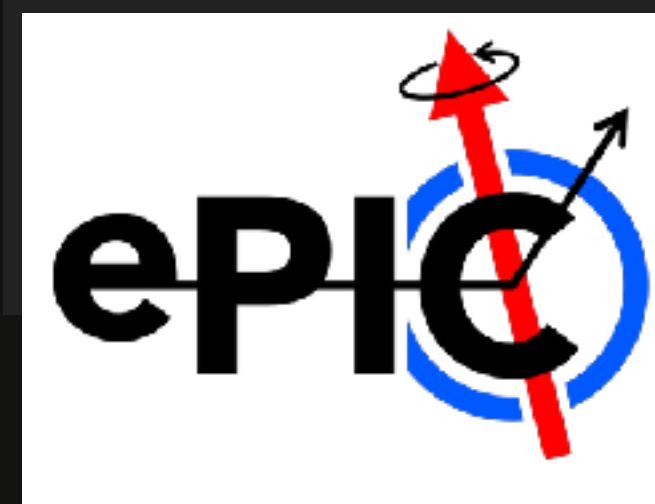


Streaming Event Identification & Classification for EIC-ePIC



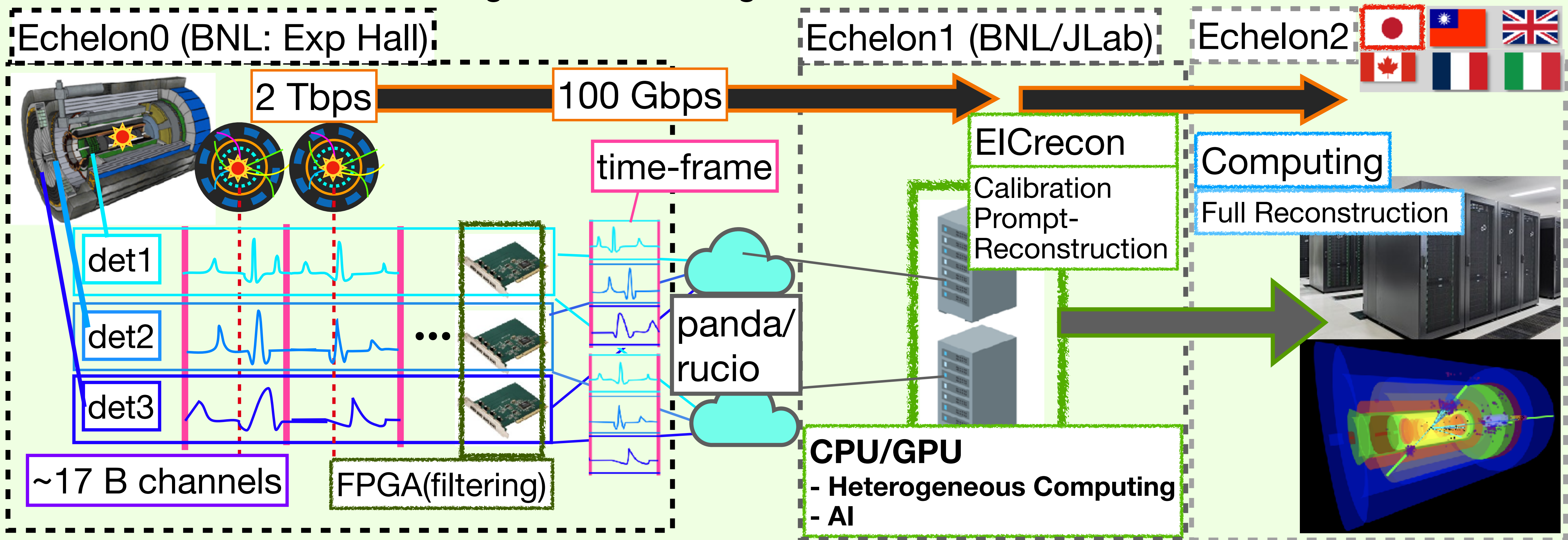
Quark Nuclear Science Institute, The University of Tokyo
Takuya Kumaoka, Taku Gunji



Tokyo City University
Marco Meyer-Conde, Oliver Partridge

EIC-ePIC DAQ Software & Computing Model

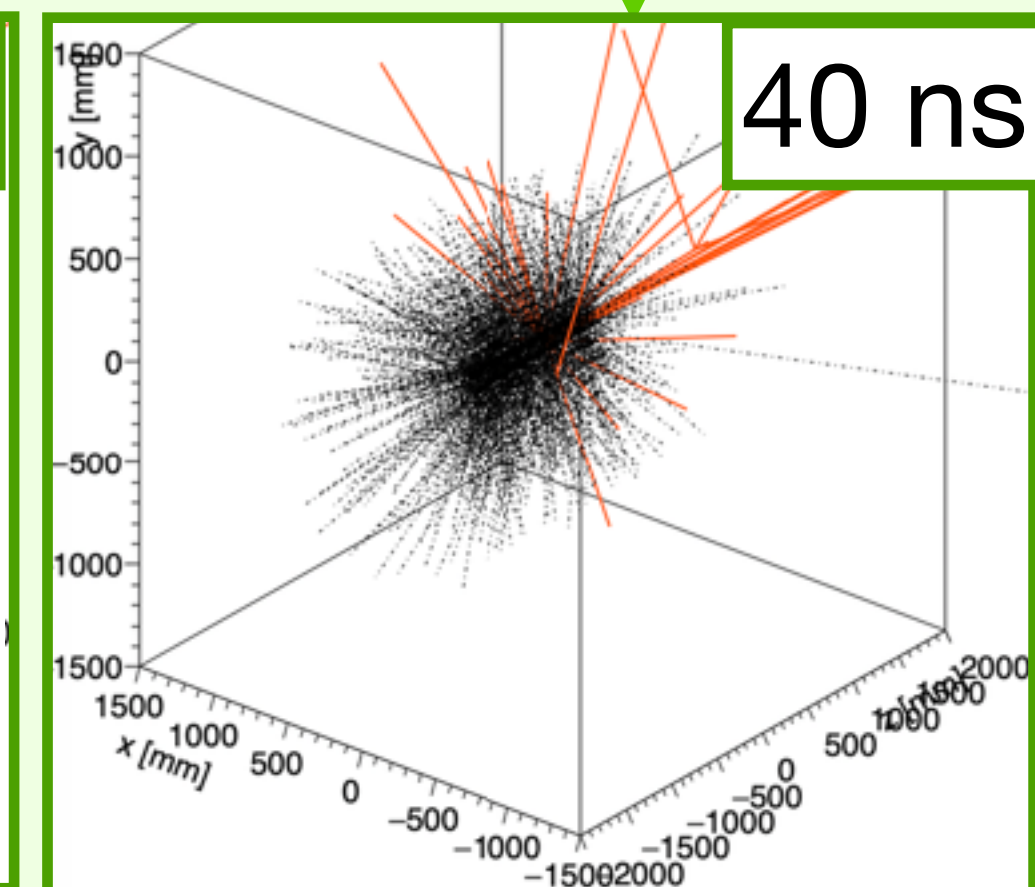
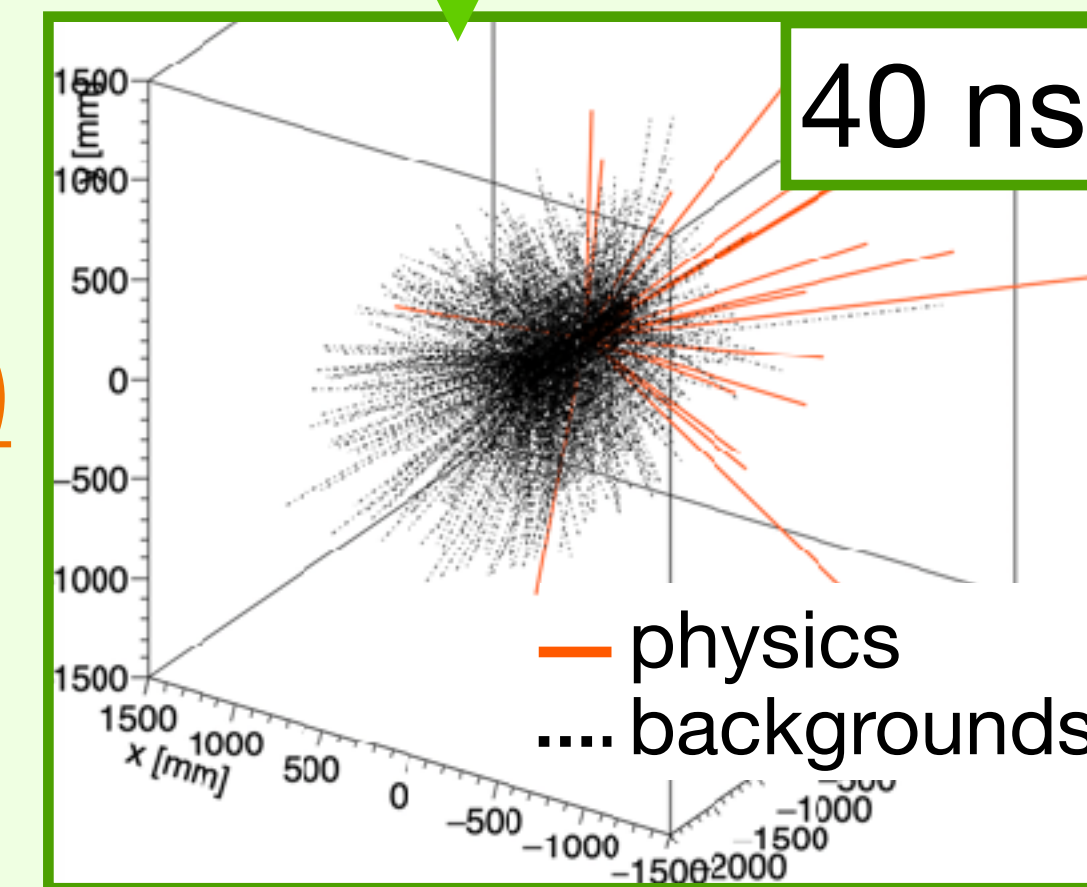
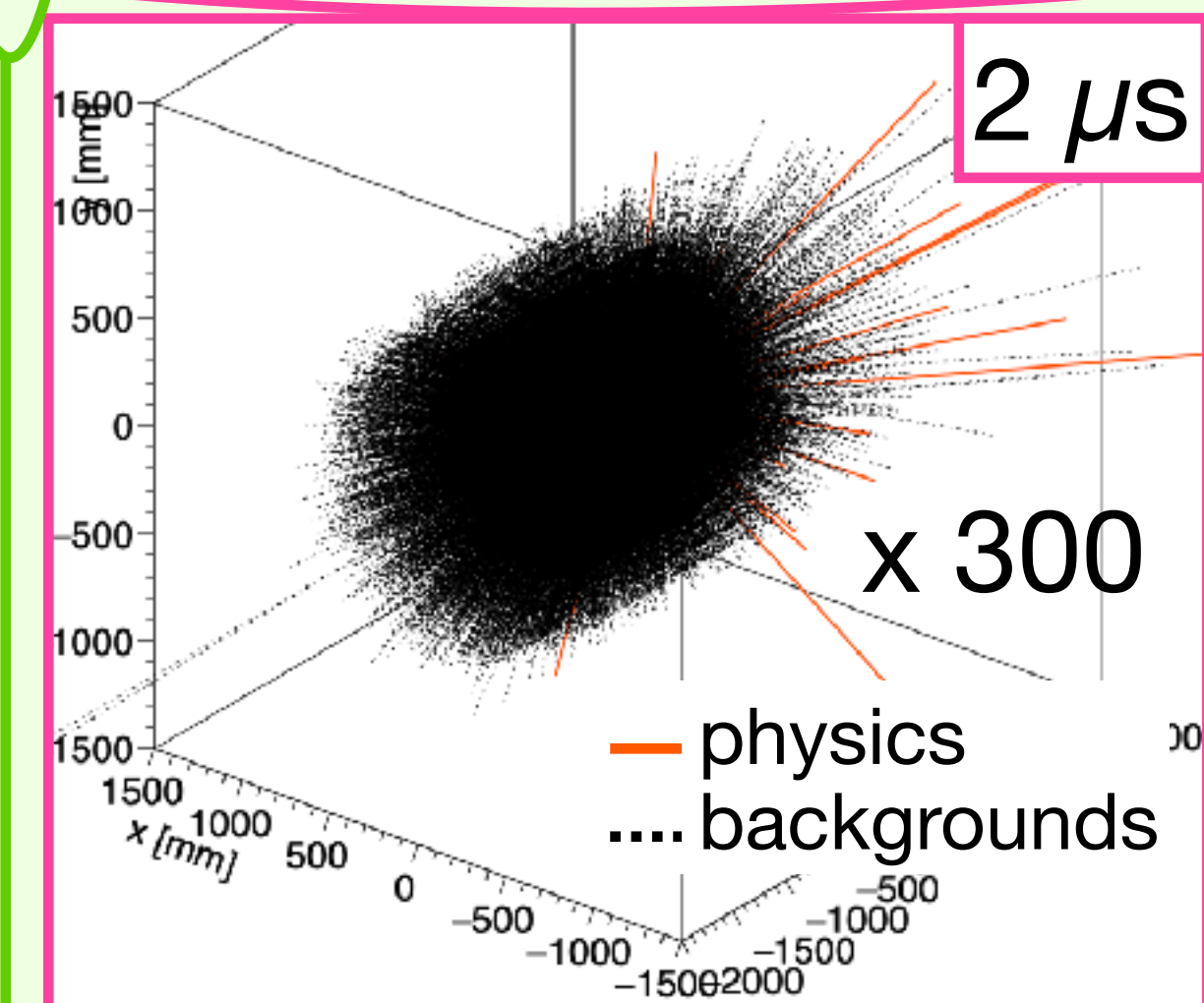
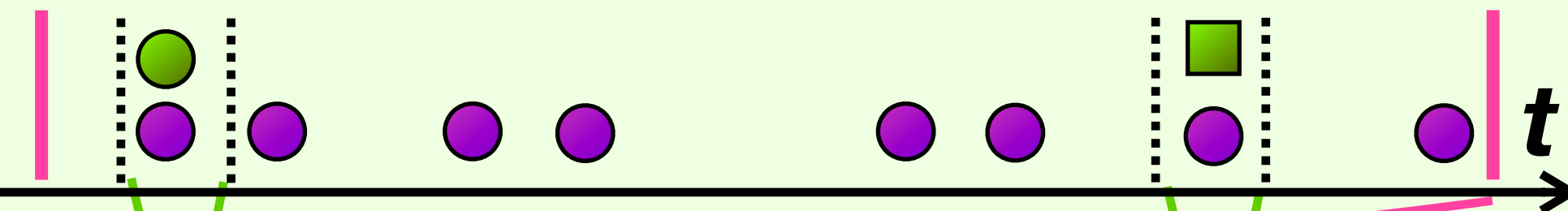
The EIC-ePIC experiment adopted a **streaming DAQ** system to record all physics events, including those that are difficult to distinguish from background events.



Purpose of this study

- Time-frame (TF) ~ 0.6 ms / super TF $\sim \times 1000$ TFs
- Acquire DIS events at 500 kHz (@ 18×275 GeV²)
- Expected background rate (dominated by Synchrotron radiation) : 3.3 GHz (@ 18×275 GeV²)
- + detector noises
- 300 DIS events/TF and background $\sim 2.0 \times 10^6$ events/TF
- **Essential to deploy streaming reconstruction to select physics event and reject background**
- + **With triggerless readout, events must be identified in software.**

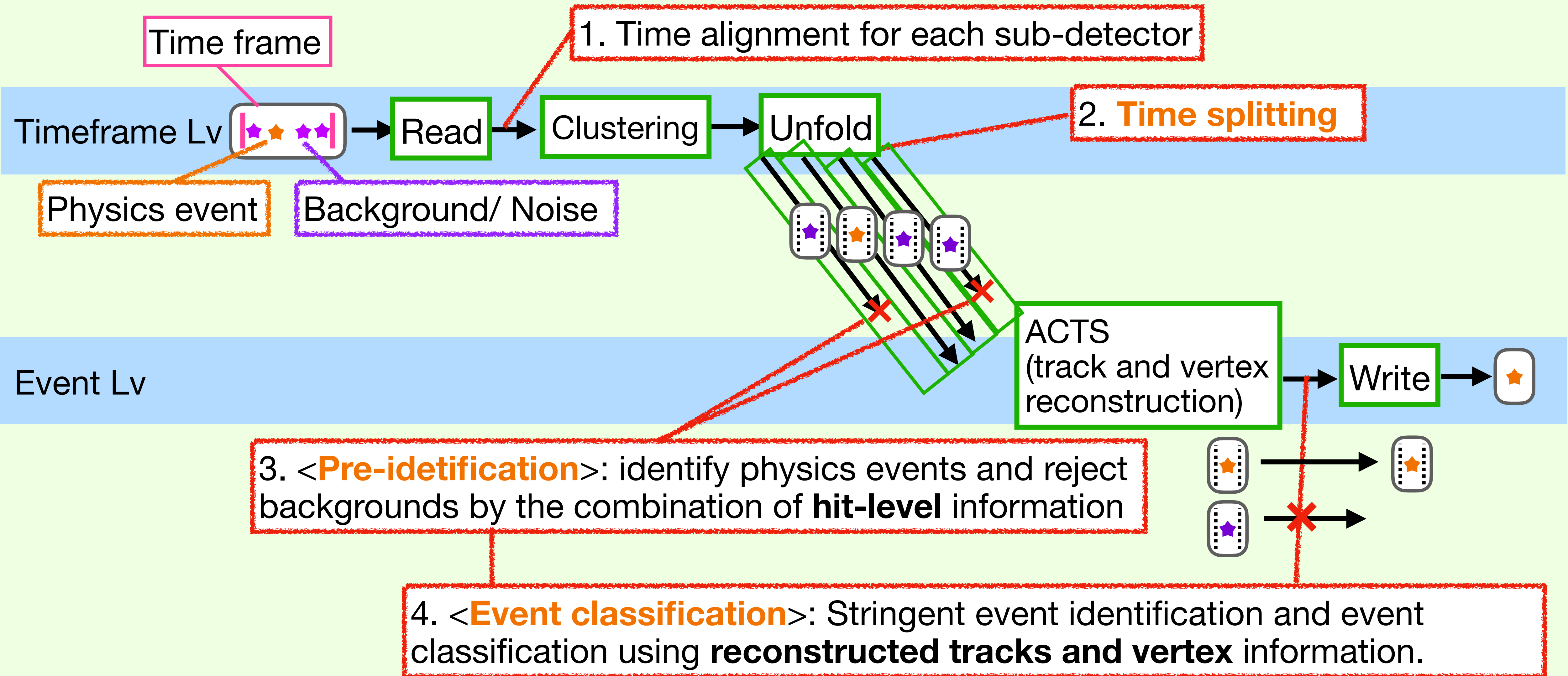
time-frame



Purpose

Development of streaming event reconstruction and algorithms for event classification (ElCrecon)

Online Event Identification System



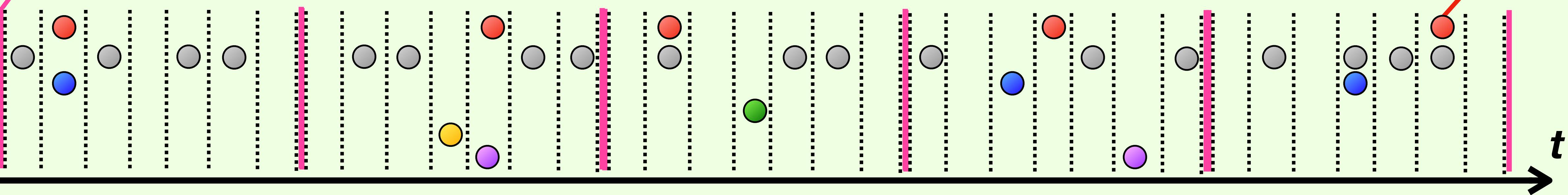
Developing Event Identification algorithms

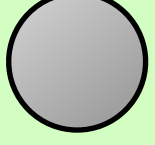
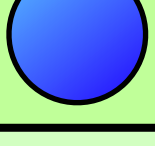
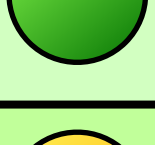
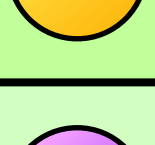
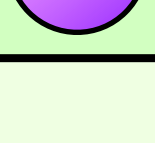
- Event pre-identification by a conventional approach (w/o ML) (T.Kumaoka)
 - Event pre-identification by a conventional approach and event classification using ML (M.Meyer, O.Partridge)
 - Event identification using ML (D.Romanov)
- } **This talk**

Simulation for Trigger Quality Evaluation

time-frame (2 μ s: 500 kHz): 1000 events)

physics events (500 kHz)



	DIS NC 10x100 $Q^2 > 1 \text{ GeV}^2$ SIDIS, DDIS, EXCLUSIVE(photo production J/Psi)	500 kHz
	Synchrontron Radiation	36.6 GHz
	Electron bremsstrahlung radiation	3.18 MHz
	Electron Touscheck scattering (intrabeam dattering)	233.5 kHz
	Electron Coulomb scattering processes	29.6 kHz
	Proton beam gas interactions	31.9 kHz

Beam energy: 10x100

Evaluate event-identification efficiency in a **blind-analysis** setting, without prior knowledge of event positions in the continuous data stream.

Event pre-Identification algorithm (Takuya's way)

1. Time slicing

Scan trigger-detector hits in each time slice (e.g. 40 ns).

2. Hit counting in a η - ϕ towers

Divide the η - ϕ space into $10 \times 10 \times 3$ towers and count the detector hits falling within each tower.

3. Tower selection

Select towers exceeding detector-specific hit thresholds.
e.g. $\text{EMCal} \geq 10$, $\text{B0} \geq 20$, $\text{ZDC} \geq 50$ hits,
 $\text{EMCal} \geq 5 \ \&\& \ \text{TOF} \ || \ \text{MPGD} > 1$

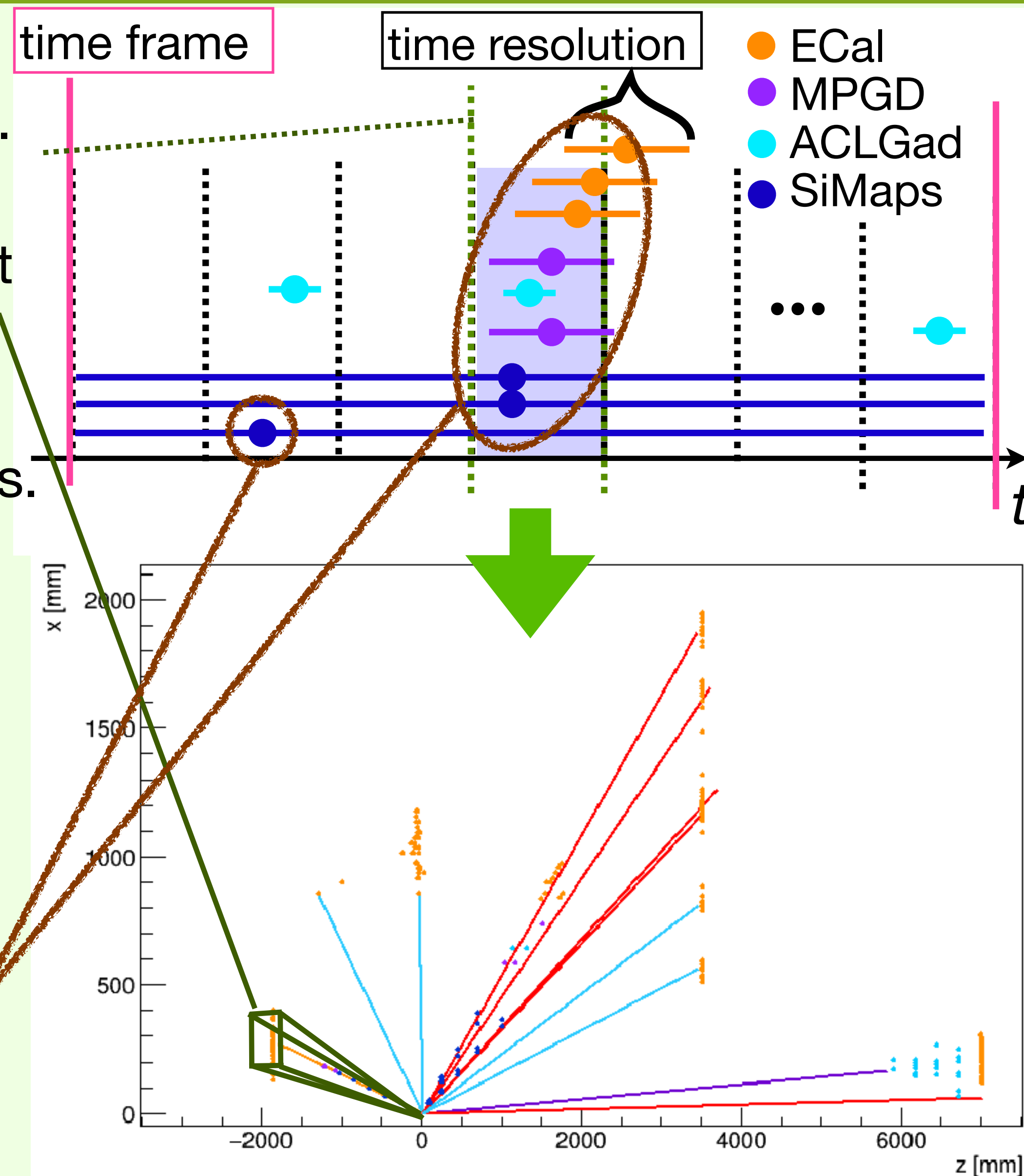
4. Trigger decision

Apply single- or multi-detector trigger tower condition.

e.g. $\text{EMCal} \times 1 + \text{B0} \times 1$, $\text{EMCal} \times 1 + \text{ZDC} \times 1$, $\text{EMCal} \times 3$

5. Event extraction

For accepted time slices, collect associated hits using detector-dependent time windows and send them to downstream reconstruction.



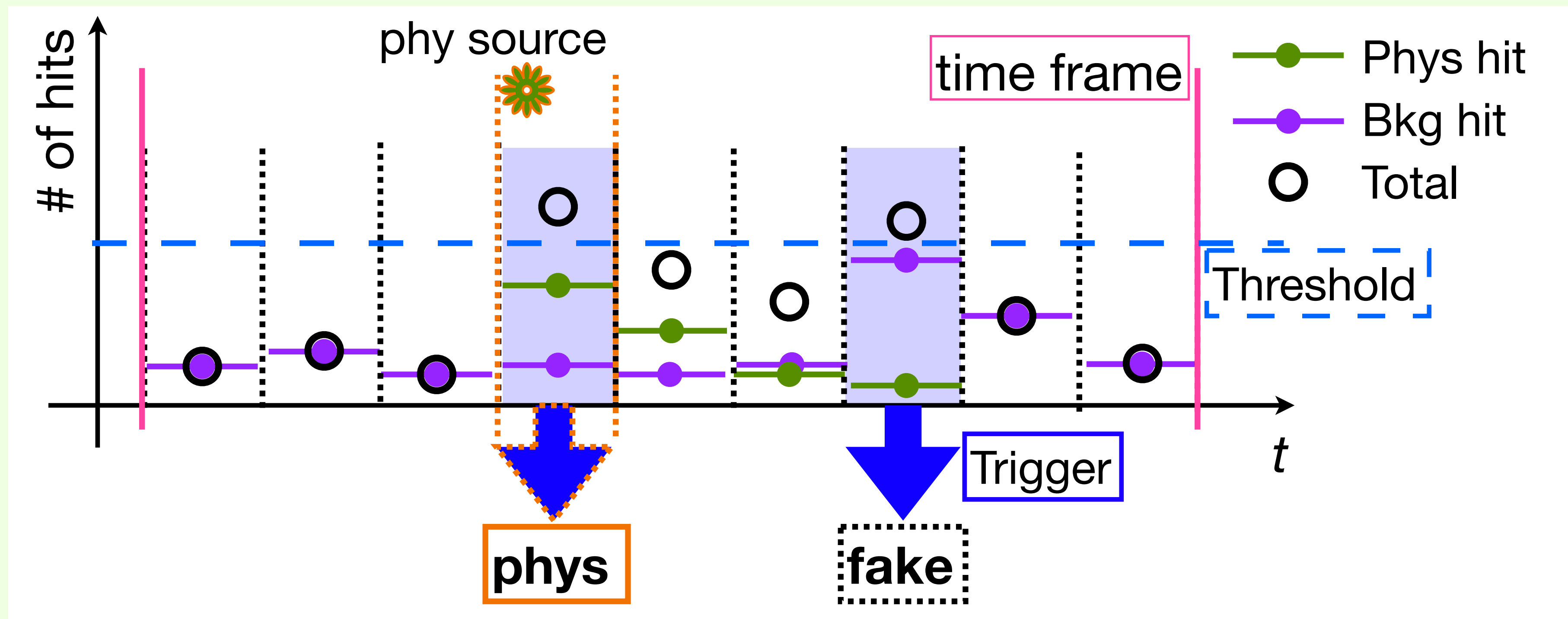
Event pre-identification Efficiency Definition

The triggered time slice containing the production time of a physics

→ **✓ physics trigger**

→ **✗ fake triggers** (Even if it includes the hits which coming from the physics events)

$$\text{Efficiency (\%)} = \frac{\text{Number of triggers of physics or fake}}{\text{Number of physics events}} \times 100$$

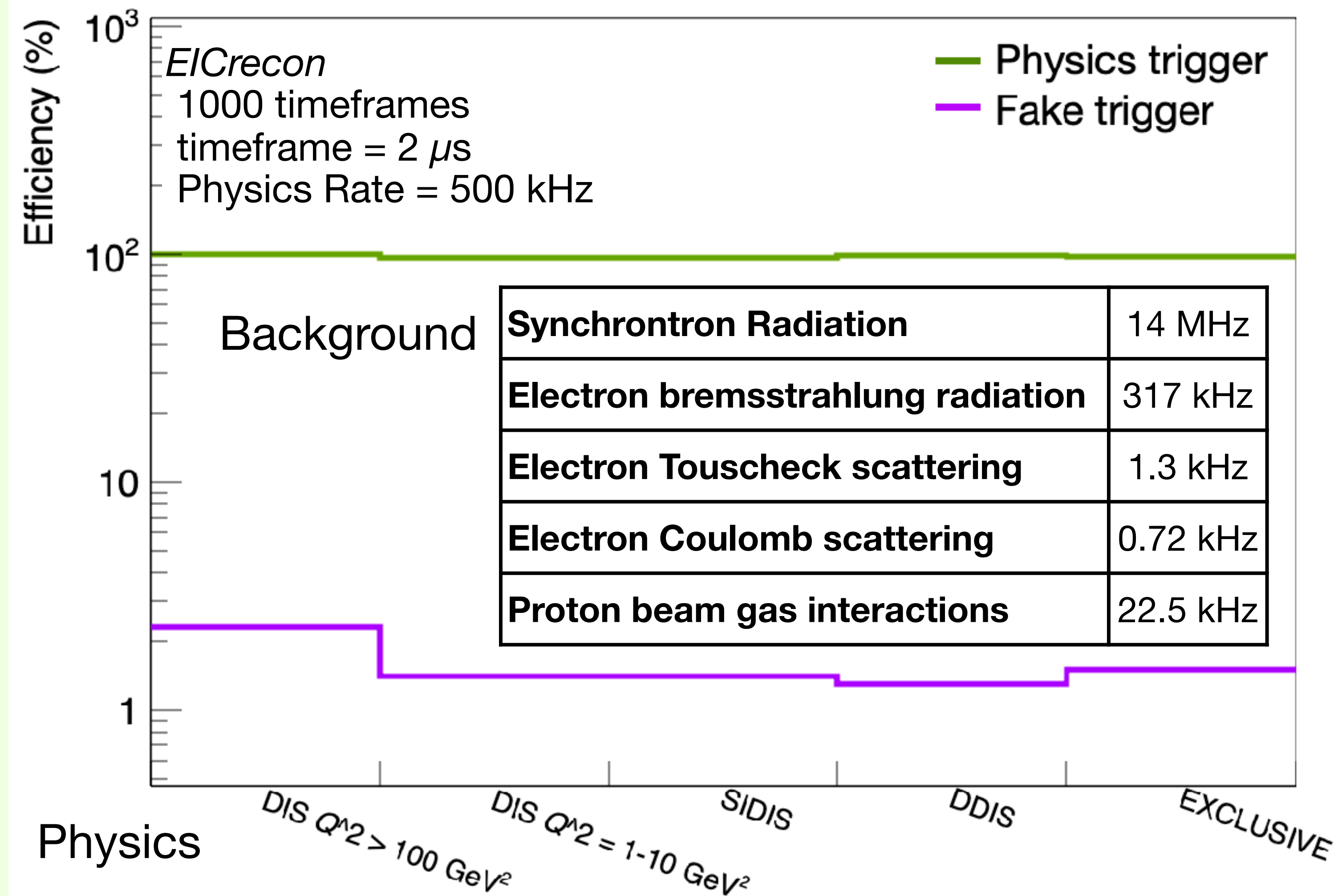


Conventional Approach to Streaming Reconstruction (Takuya)

For all **five physics event categories**, the event pre-identification efficiency is **above 95%**.
✧ This does not imply that the quality of the triggered events has been validated.

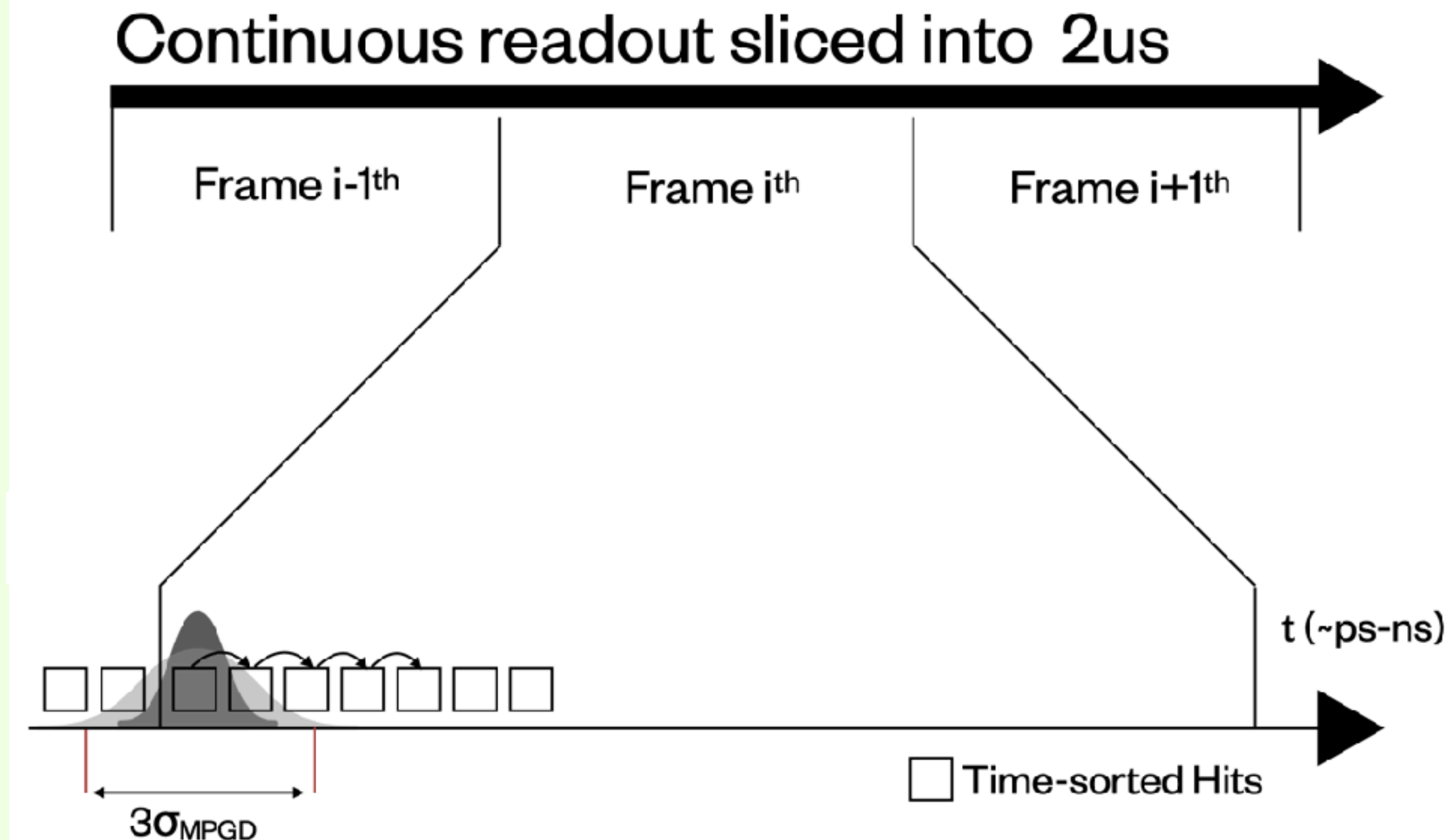
The **fake trigger rate** is less than a **few percent**.

✧ This should not be interpreted as the background reduction rate, since a single time slice may contain hits originating from multiple background events.



Event pre-identification (M.Meyer, O.Partridge)

- (1) Timing Alignment: $t_i = |t_i' - r/c|$
- (2) **Trigger Candidate:**
 - Gated Fast hits (MPGD+TOF): $3\sigma_{\text{MPGD+TOF}}$
- (3) **Timing Coincidence:**
 - Per-Detector Gating: $3\sigma_i$
(σ_i being i-th detector res.)
=> no pure time window definition)
- (4) **Tracklet Coincidence:**
 - Pointing toward IP
- (5) **Calorimeter Energy threshold**
 - Energy coincidence within sigma t_0
(attempt to introduce time error VS. E_{hit})



Trigger Candidate

$$t_0 = \frac{\sum_i w_i t_i}{\sum_i w_i}, \quad \sigma_{t_0} = \left(\sum_i \frac{1}{\sigma_i^2} \right)^{-1/2}$$

$$w_i = \frac{1}{\sigma_i^2}$$

- t_i : corrected timestamp of hit i ,
- σ_i : time resolution of detector i .

Timing Coincidence

$$|t_i - t_0| < n \sigma_i$$

where:

- t_i is the corrected hit or cluster time,
- t_0 is the candidate event-time estimate,
- σ_i is the hit or cluster time error,
- n is the configured number of standard deviations.

Tracking Coincidence

Geometrical Alignment

MPGD Inner ↔ TOF

TOF ↔ MPGD Outer

Energy Coincidence

$$|t_{\text{clu}} - t_0| < n \sigma_{t,\text{cal}}(E_{\text{clu}})$$

$$\sigma_{\text{clu}} = \frac{\sqrt{\sum_i (\sigma_{\text{hit},i} E_{\text{hit},i})^2}}{\sum_i E_{\text{hit},i}}$$

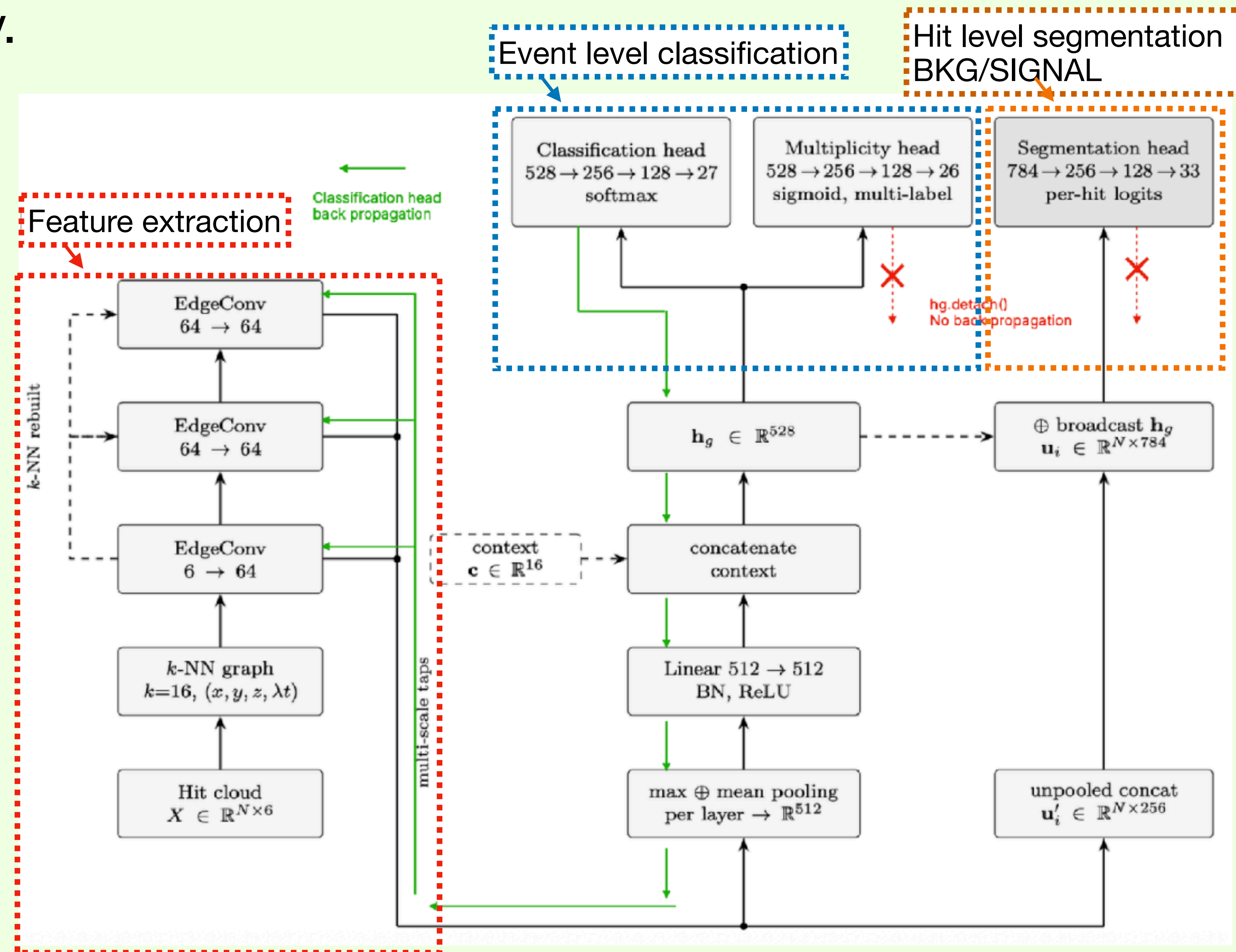
$$\sigma_{\text{hit},i} = \sqrt{\frac{a^2}{E_{\text{hit},i}} + b^2}$$

Event Classification using GNN (M.Meyer, O.Partridge)

Event Builder candidates are represented as point clouds and processed with a Dynamic Graph CNN using kNN and EdgeConv.

The GNN learns hit-to-hit correlations to classify each candidate into background or a physics-event class, while also providing hit-level segmentation.

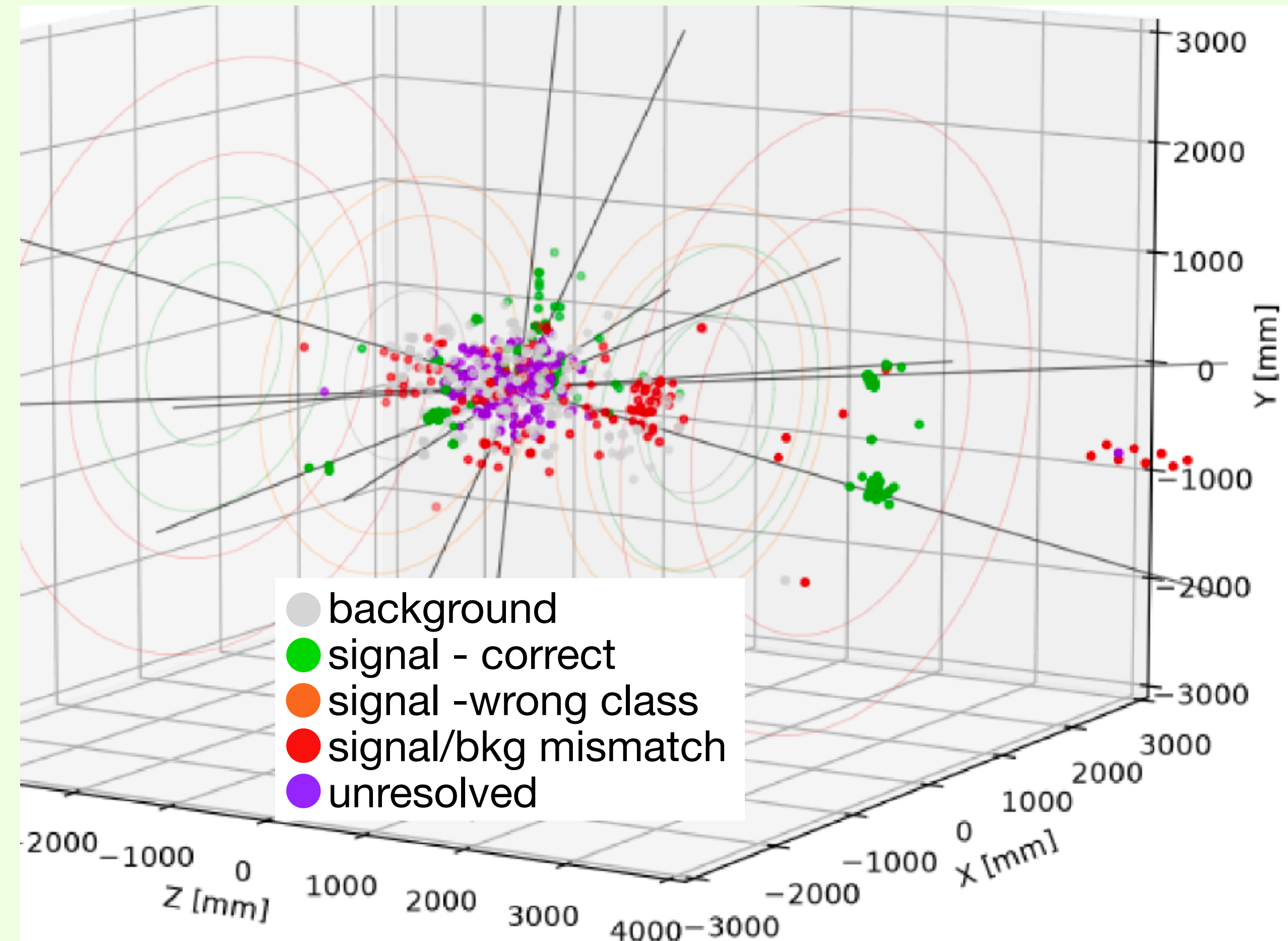
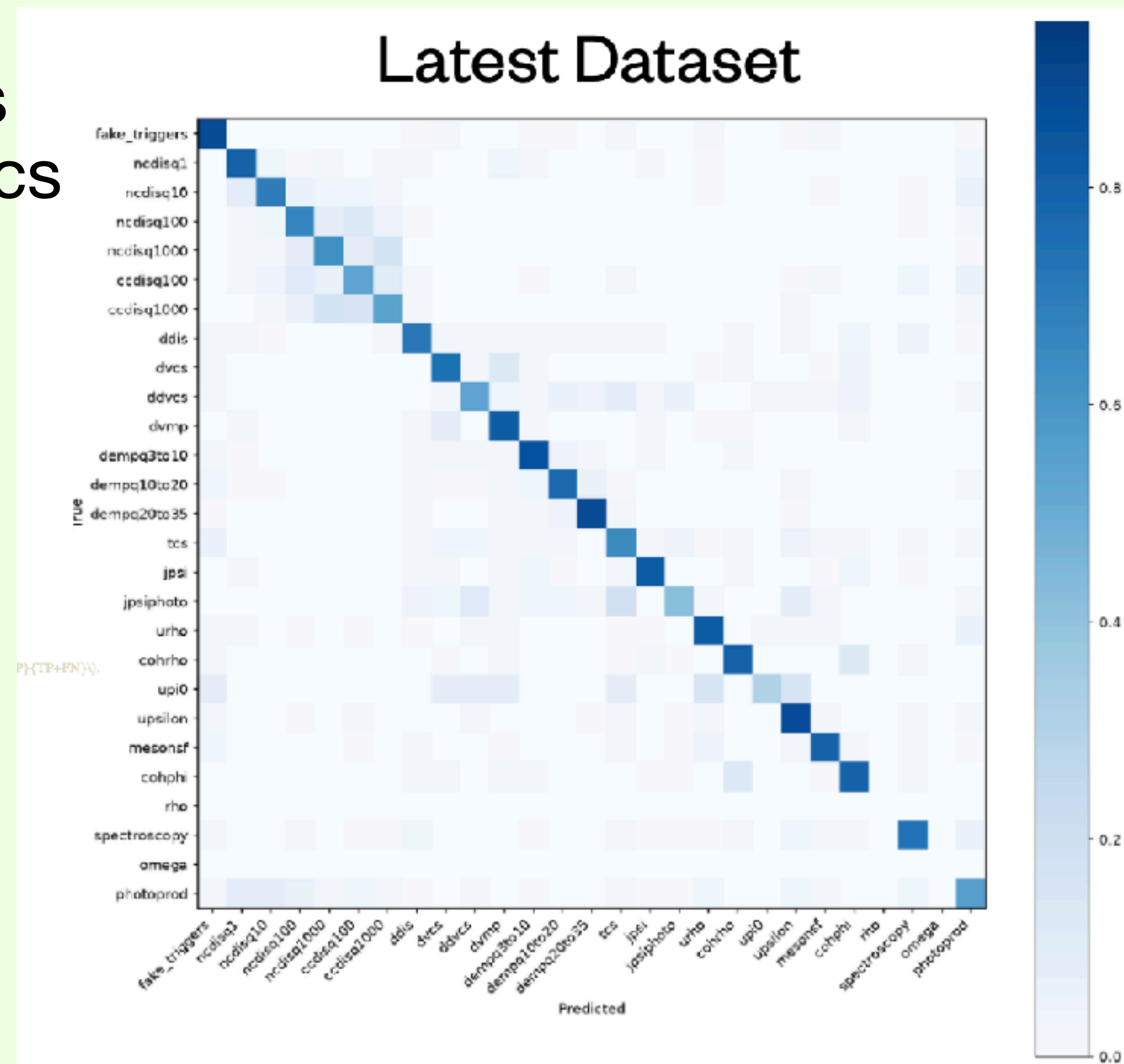
The trained model is deployed in EICrecon through ONNX for ML-based pre-filtering before the computationally expensive full reconstruction.



Event Classification using GNN (M.Meyer, O.Partridge)

- Preserves approximately 99% of signal hits while reducing tracking combinatorics and accelerating tracking by up to 6.5×
- The segmentation of hits between physics and background also works well.
There are unresolved hits in simulation, if it is solved the classification level will be better.

27 kinds
of physics
events



Summary and Outlook

Summary

We are developing methods for event identification and classification within a timeframe. We have already prepared benchmark algorithms and implemented them in ElCrecon. Using simulated data, we have evaluated their performance and confirmed that they achieve high efficiencies.

Outlook

- Event identification
 - ◆ Validate the quality of triggered events (tracks and calorimeter clusters).
 - ◆ Test with more realistic simulation (timeframe size, detector noise, collision rate).
 - ◆ Modularize the algorithm to enable comparisons with other approaches.
 - ◆ Implement multithreading and explore GPU acceleration.
 - ◆ Implement alignment and calibration line.

Backup Slides

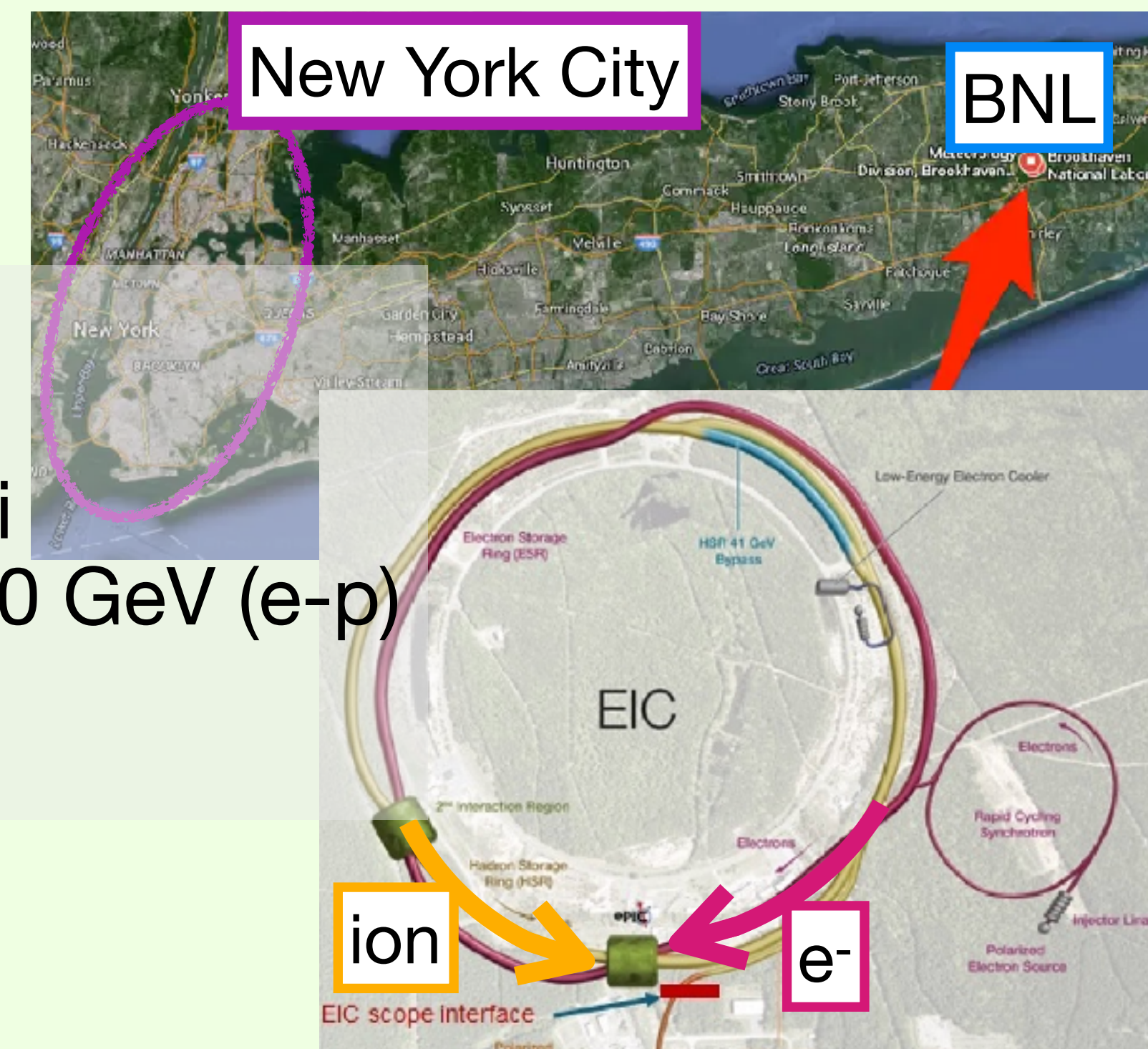
Electron-Ion Collider

EIC (Electron Ion Collider) (2035-)

New accelerator of Brookhaven National Laboratory(BNL)

- Characteristics:

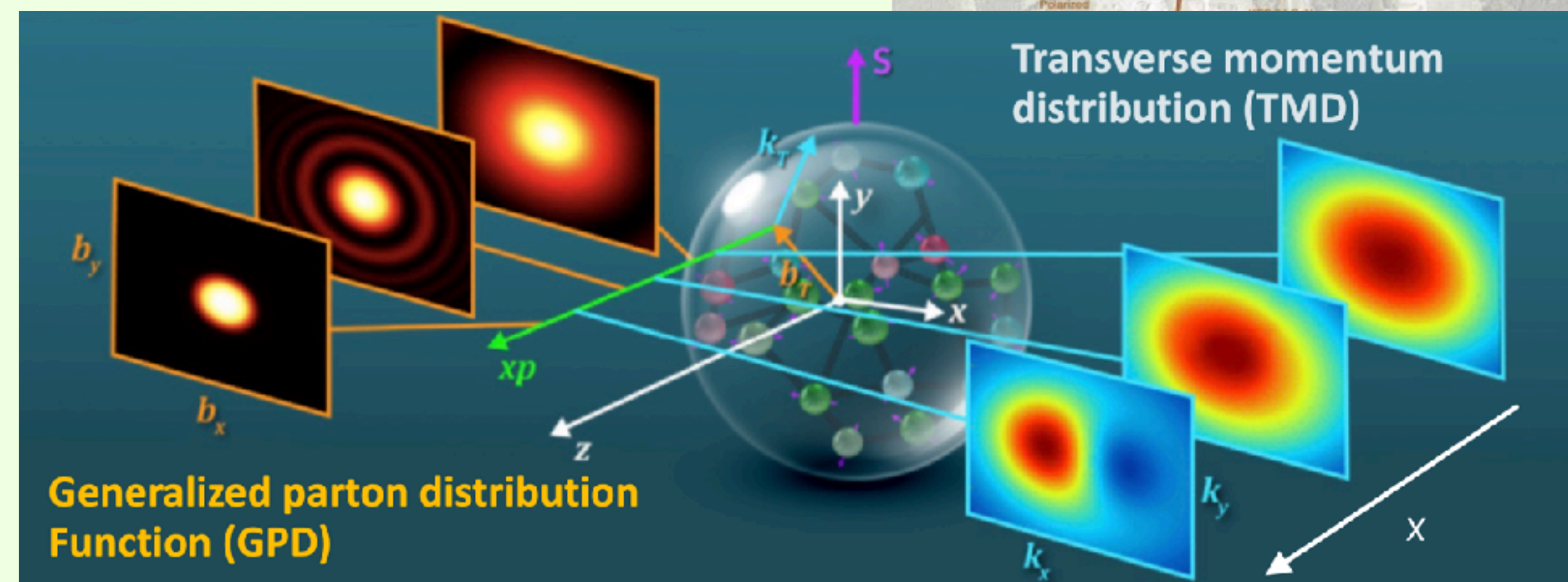
- A high degree of beam polarization* (~70%) for electrons and ions
- Availability of ion beam from deuteron to the heaviest stable nuclei
- Variable center of mass energies ~20-100 GeV, upgradable to ~140 GeV (e-p)
- High collision luminosity* $\sim 10^{33}$ - 10^{34} cm⁻²s⁻¹
- Possibly more than one interaction region



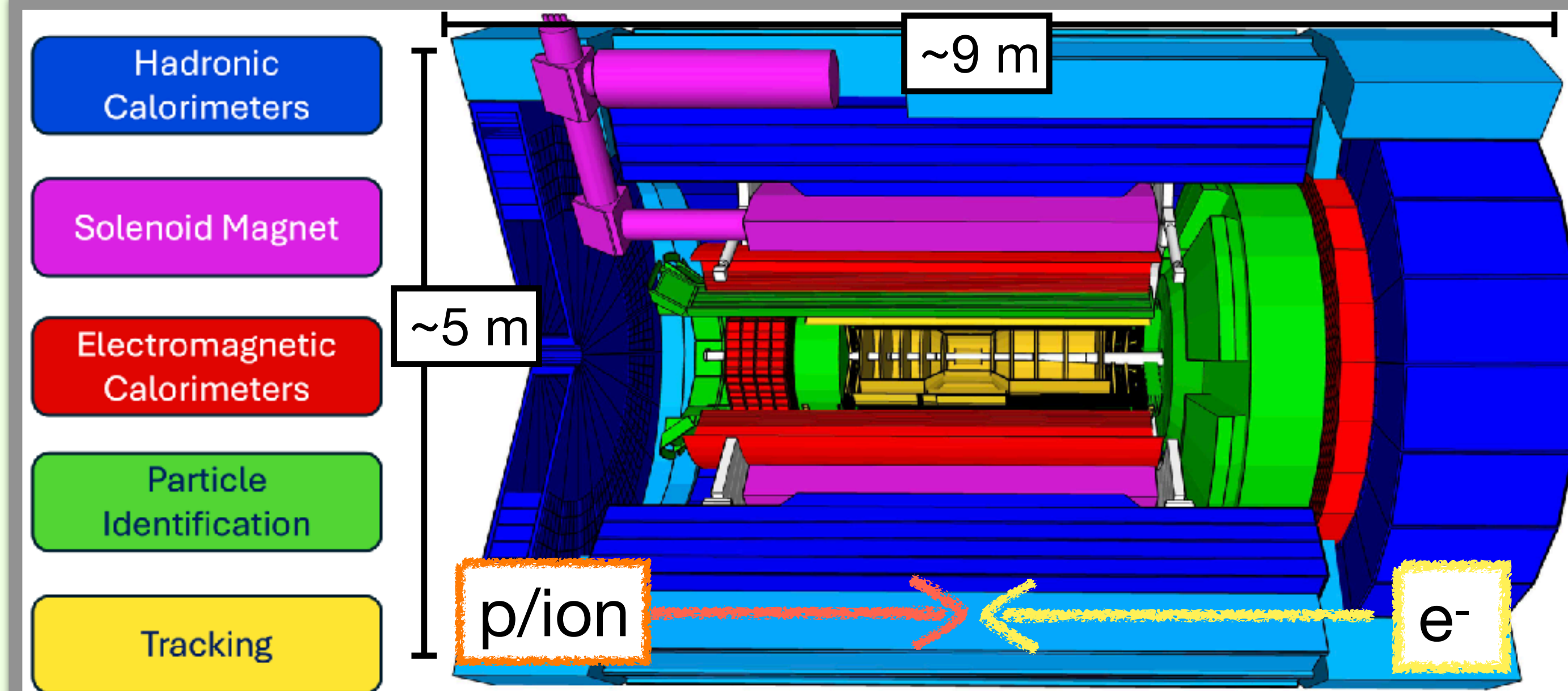
Main Physics Targets:

Nucleon and Nucleus structure using polarized e+p and e+A collisions over the wide beam energy

- How does the **mass** of the nucleon arise?
- How does the **spin** of the nucleon arise?
- What are the emergent properties of dense systems of gluons?



ePIC (electron-Proton/Ion Collider) detector / Time resolution



Calorimetry: (ECal: ~ 20 ns)

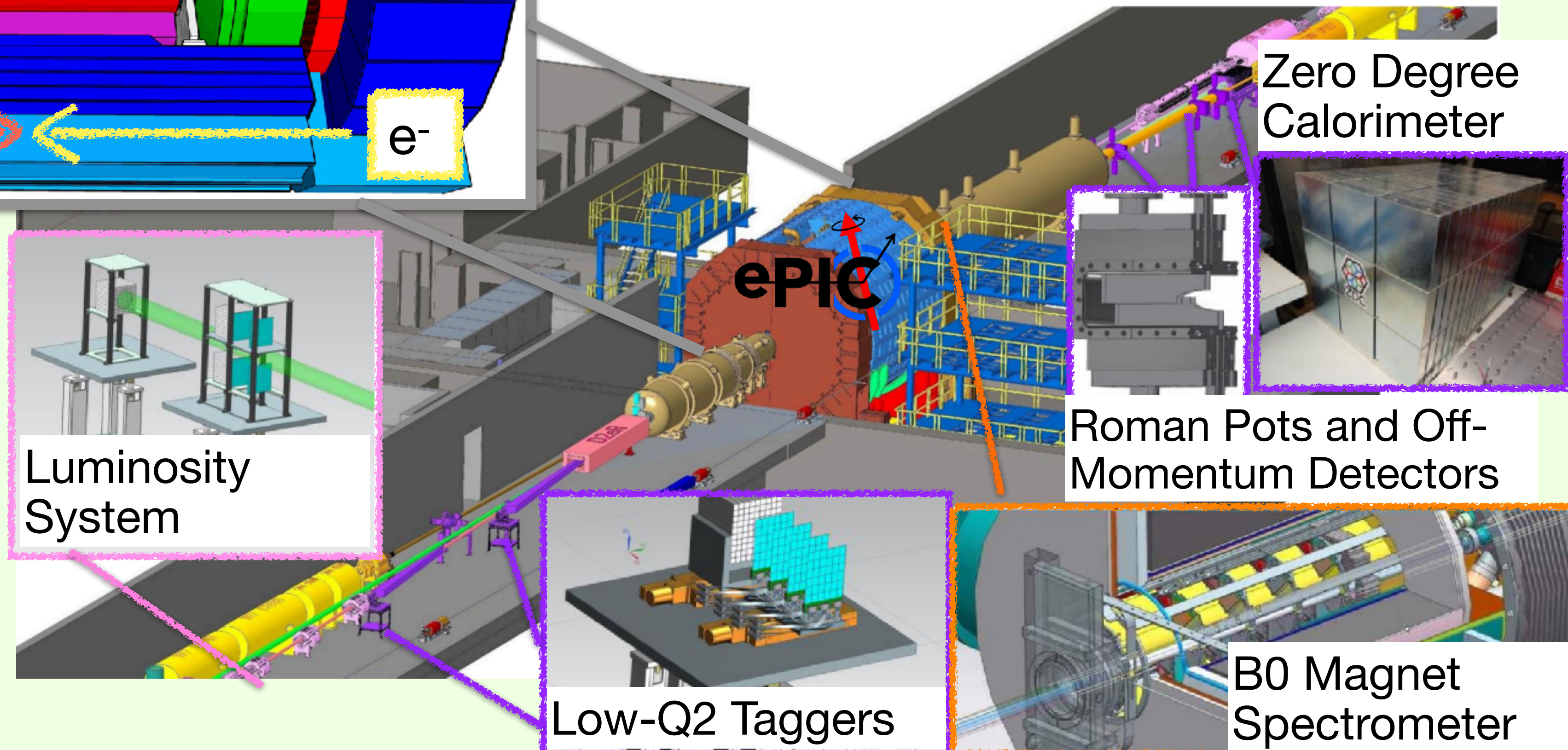
- Imaging Barrel EMCal
- PbWO₄ EMCal in backward direction
- Finely segment EMCal+HCal in forward direction
- Outer HCal (sPHENIX re-use)
- Backwards HCal (tail-catcher)

Tracking:

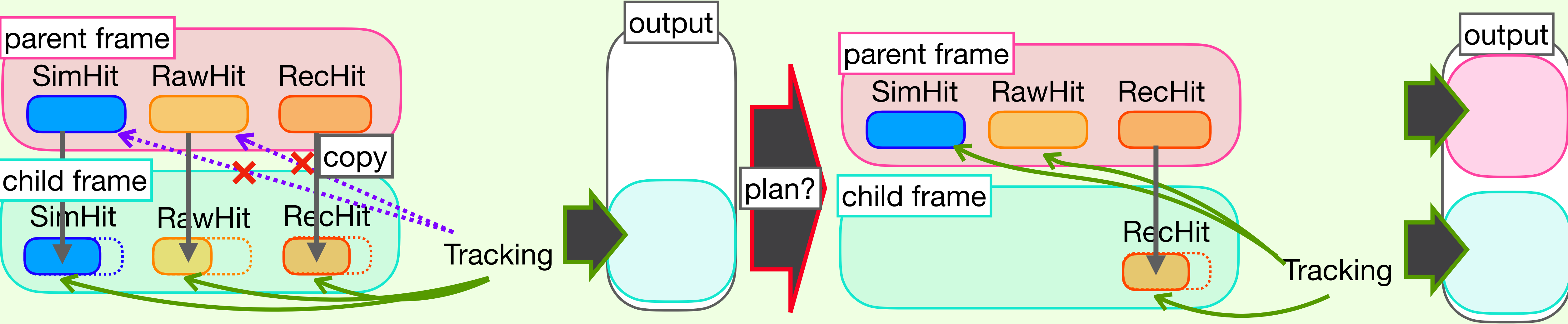
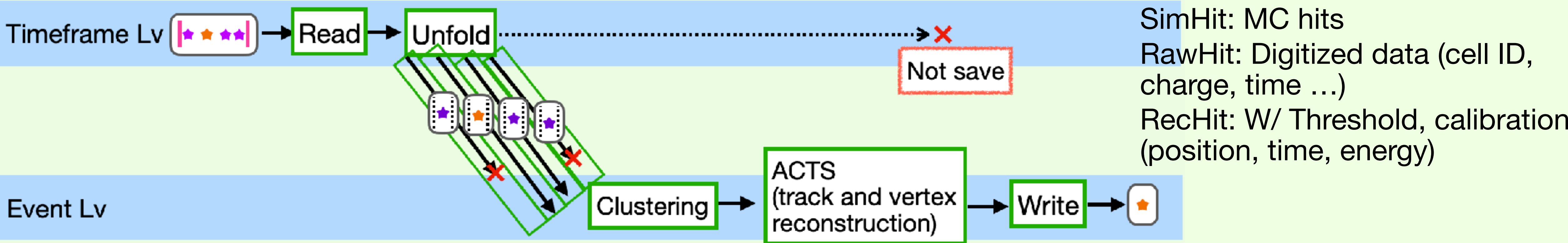
- New 1.7 T solenoid
- Si MAPS tracker (~ 2 μ s)
- MPGDs (~ 10 ns)

PID:

- hpDIRC
- pfRICH
- dRICH
- AC-LGAD (~ 30 ps TOF, B0)



1. Which information is important for the reconstruction group? / How should we need to make a output data structure?



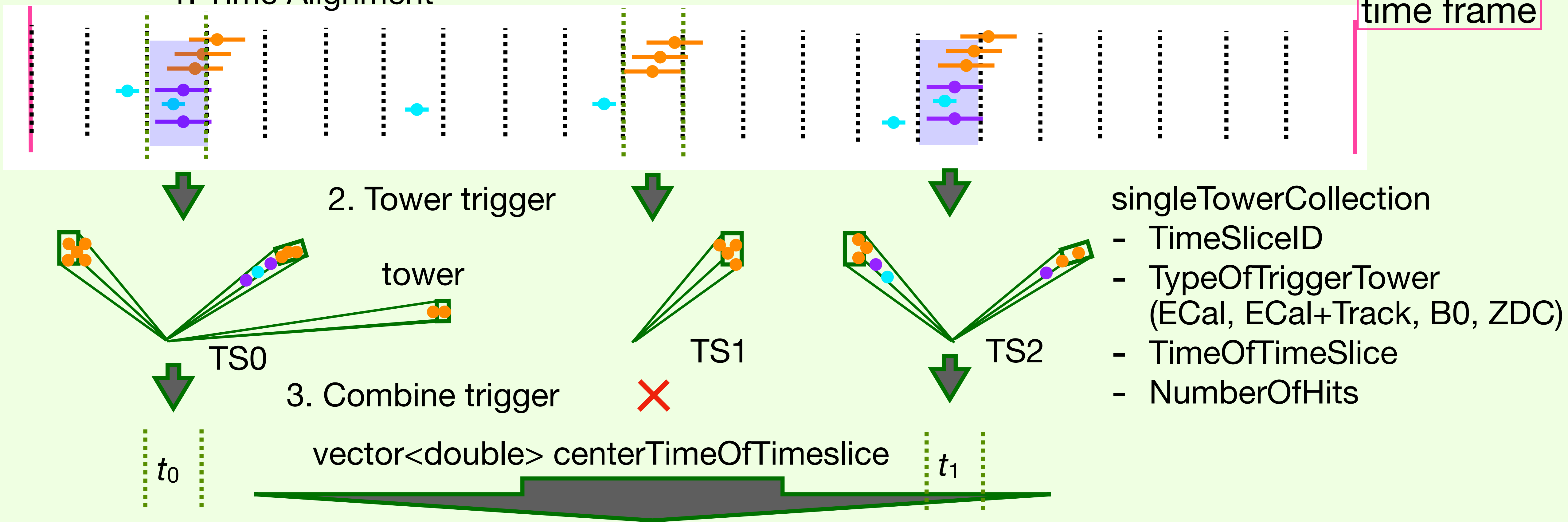
In the current architecture, a factory cannot read the collection in a different level. Therefore, the unfold copies every collection which the down stream factory now.

It is possible to avoid unnecessary copies, but the output file format will change with the single level way. It requires the offline analysis code modification.

Modularize the current pre-event identification

1. Time Alignment ➡
2. Scan Time slice and make time&geometrical coincidence trigger towers ➡
3. Combine trigger using the towers and send time slice time vector to the unfold ➡
4. Unfolder

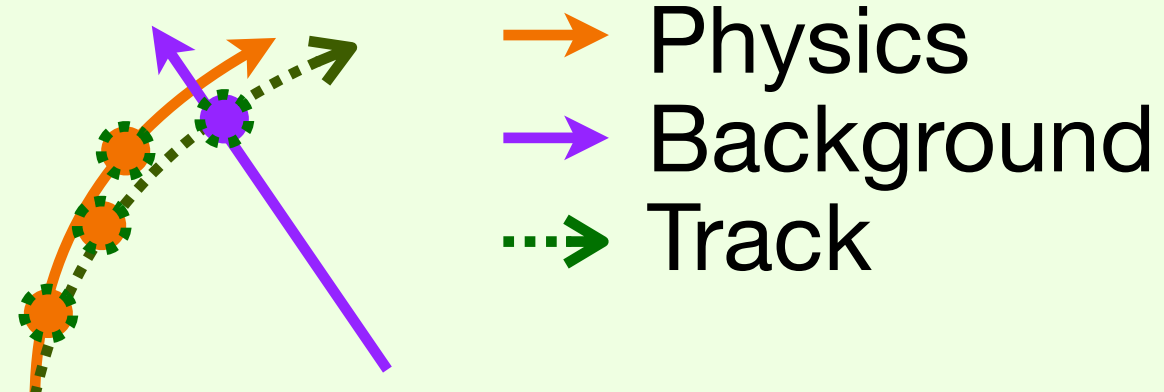
1. Time Alignment



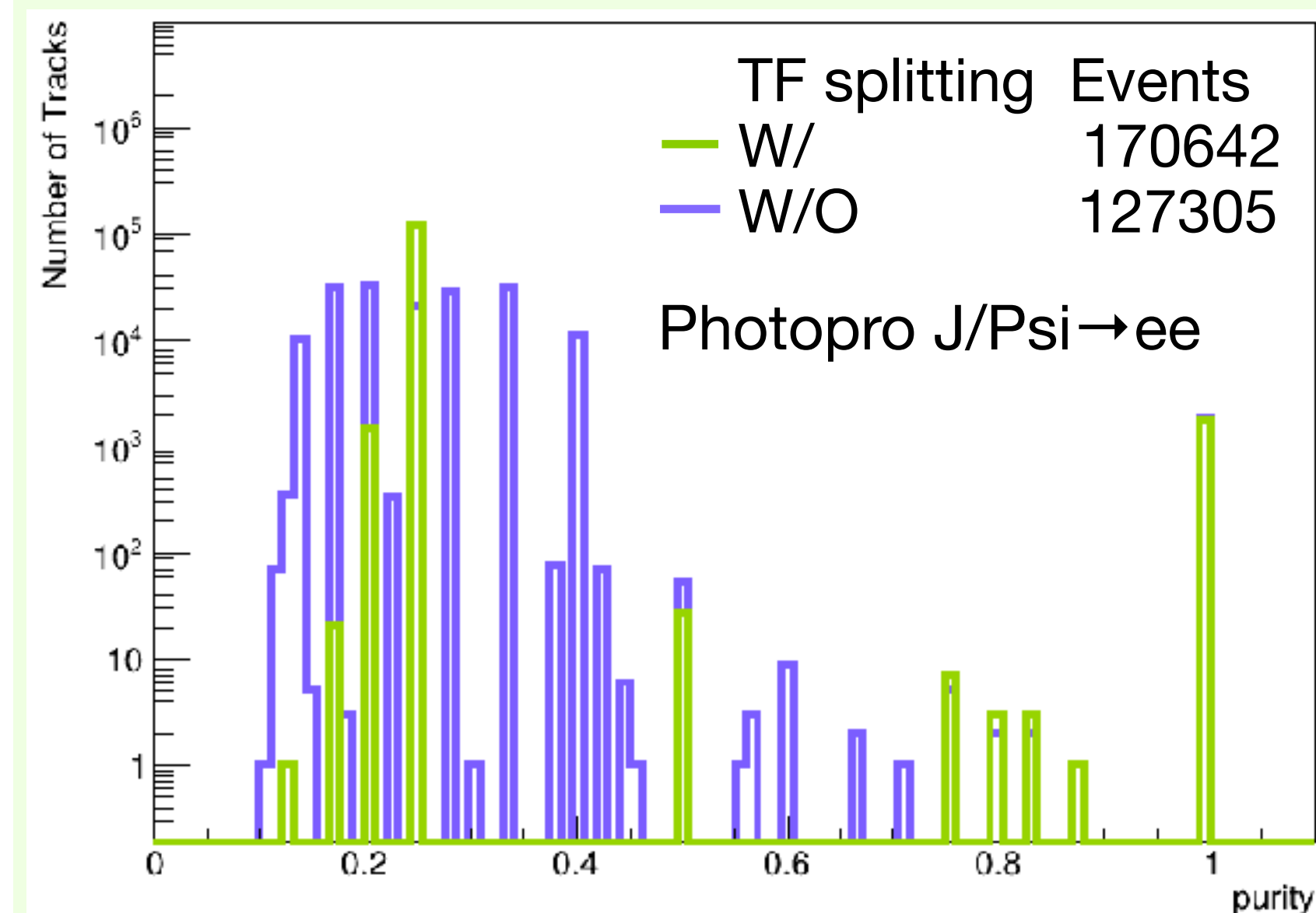
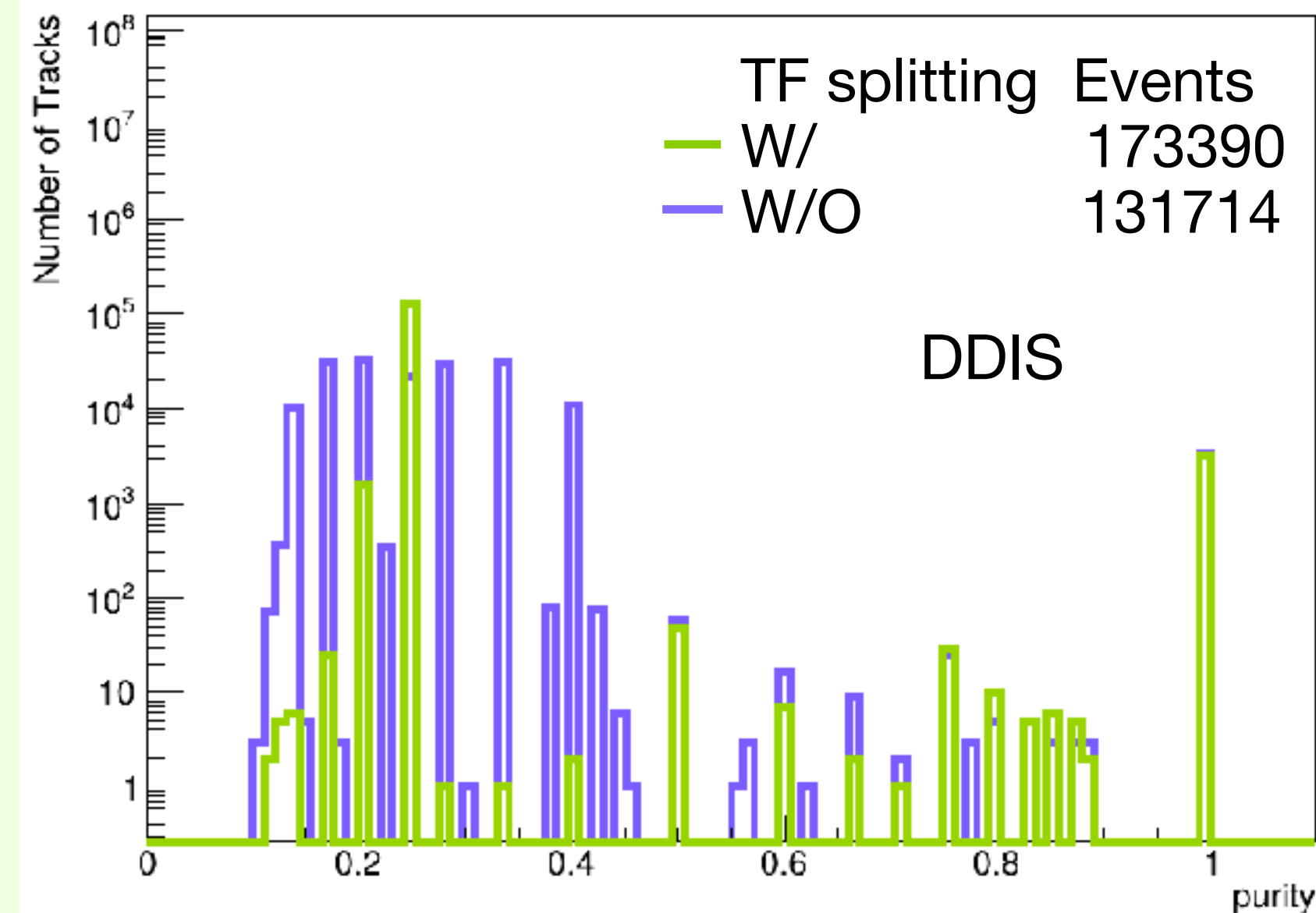
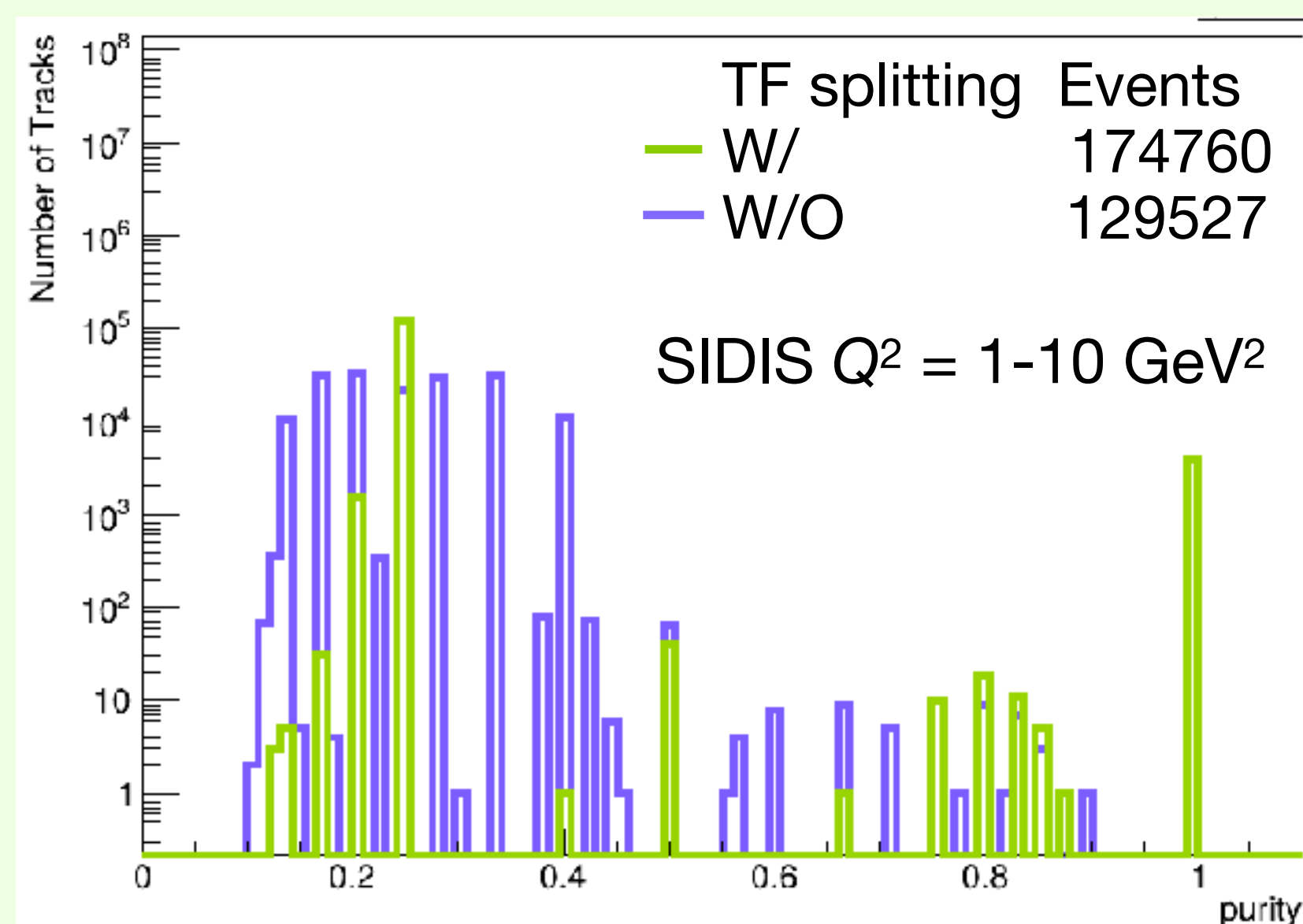
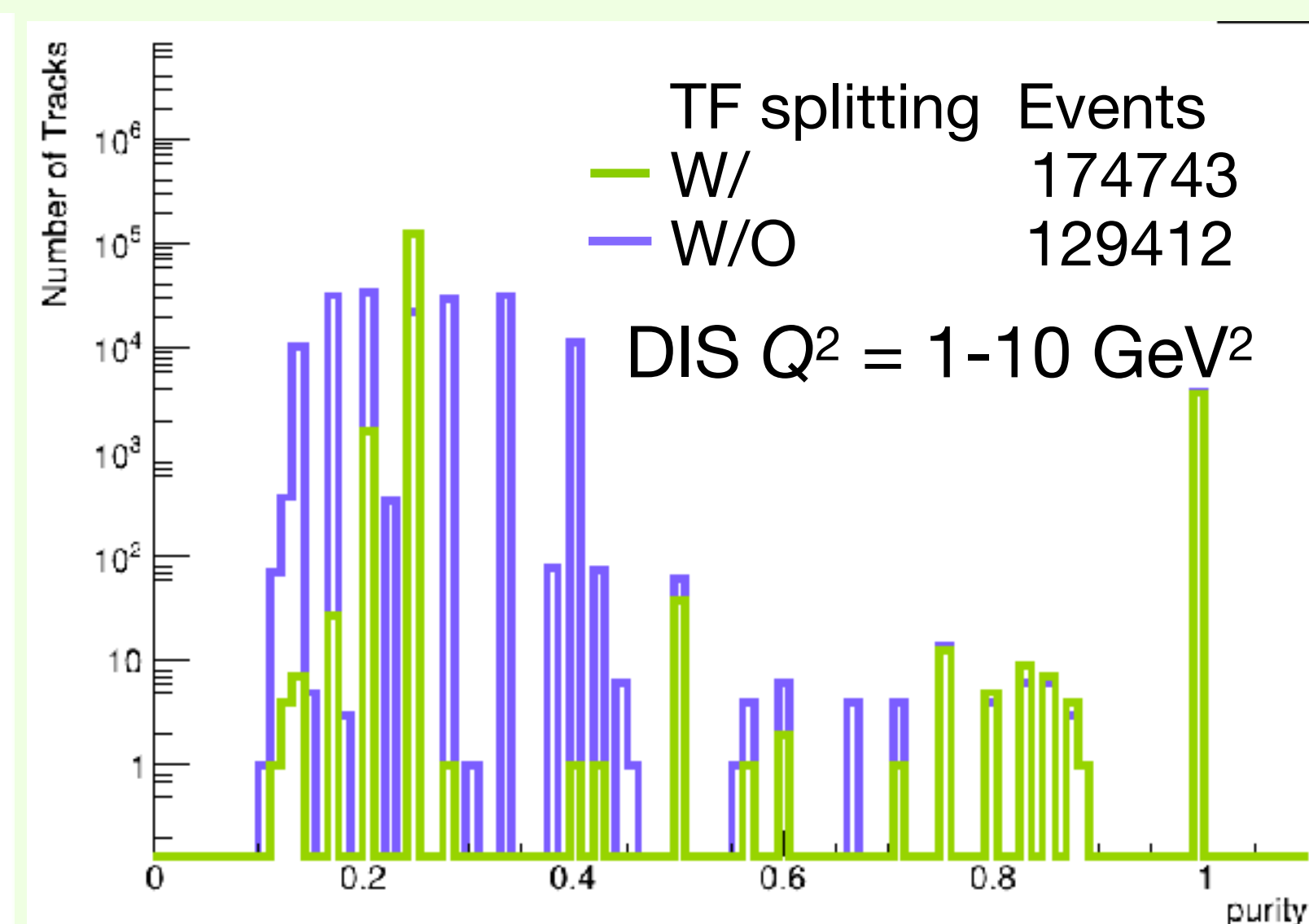
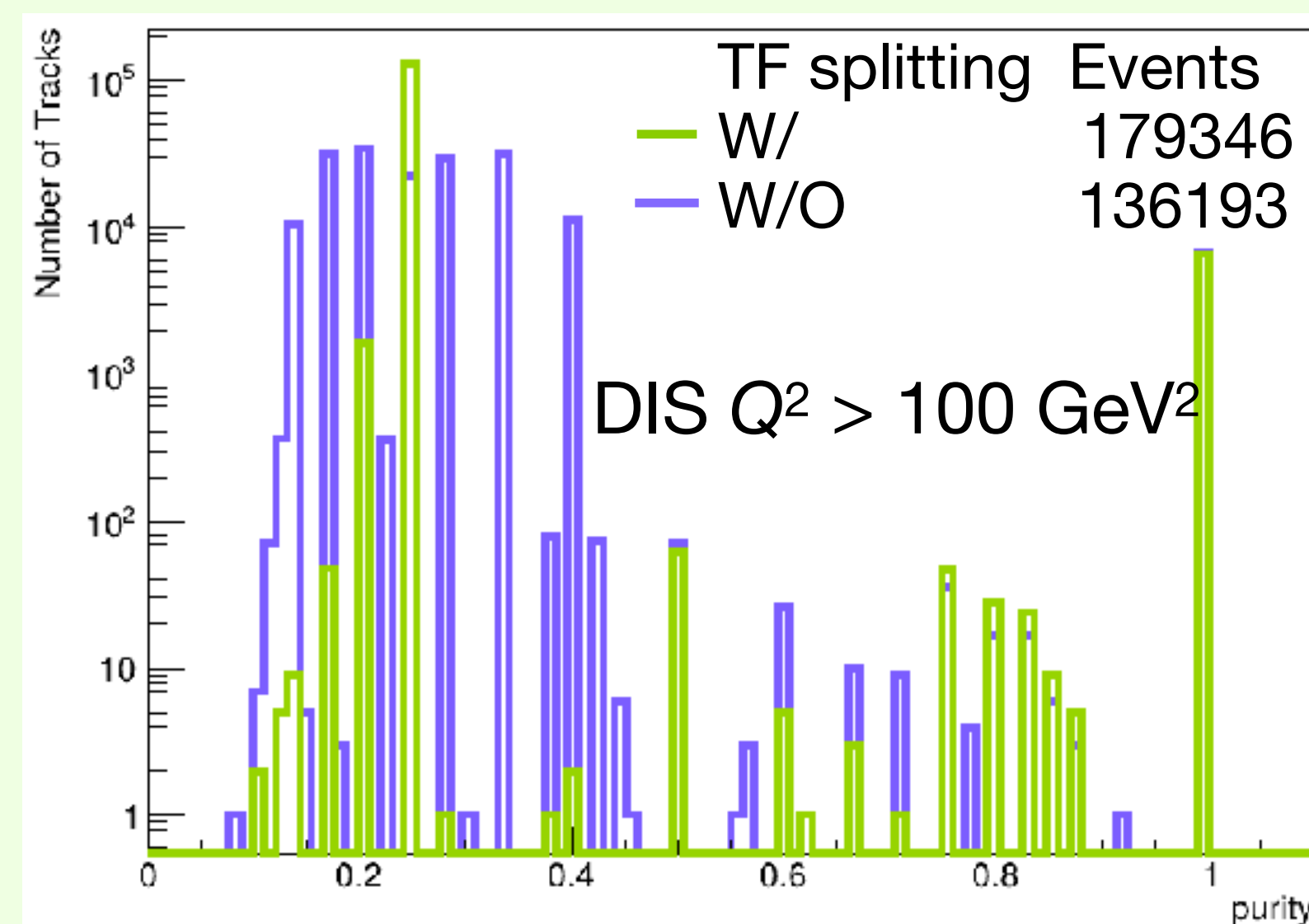
4. Unfold: Create time slice frames and register all detector's hits into the frames, respectively.

Tracking purity

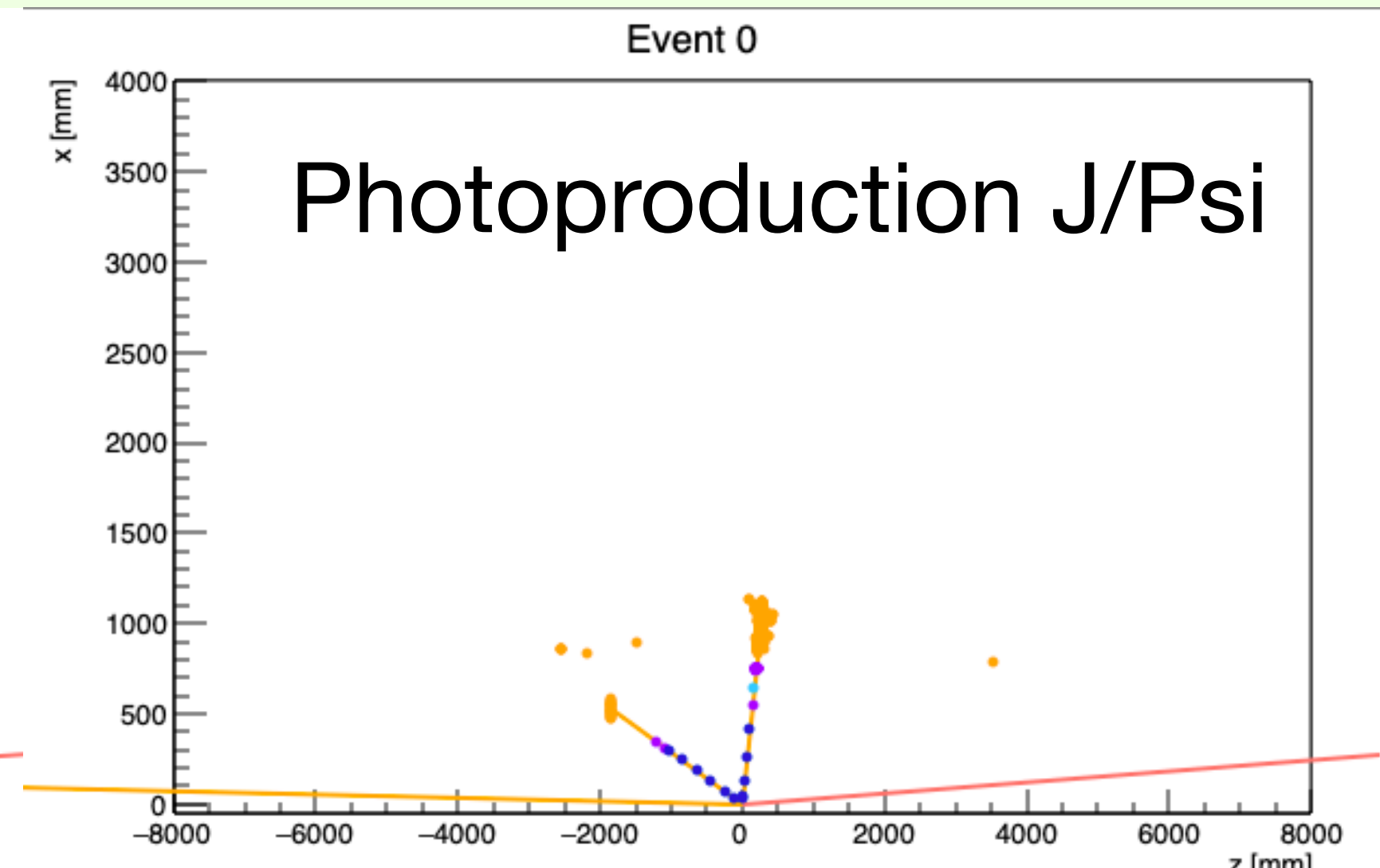
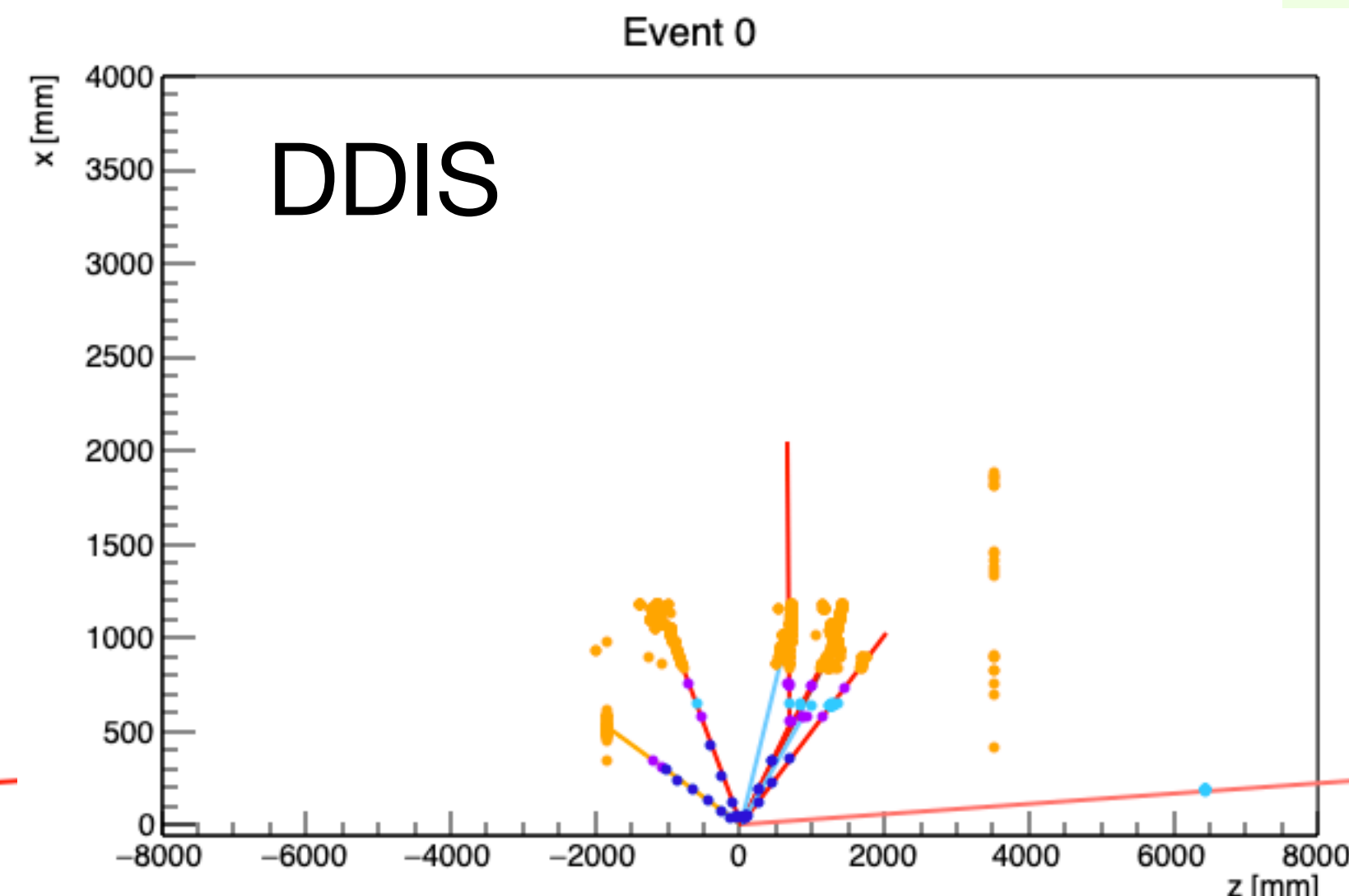
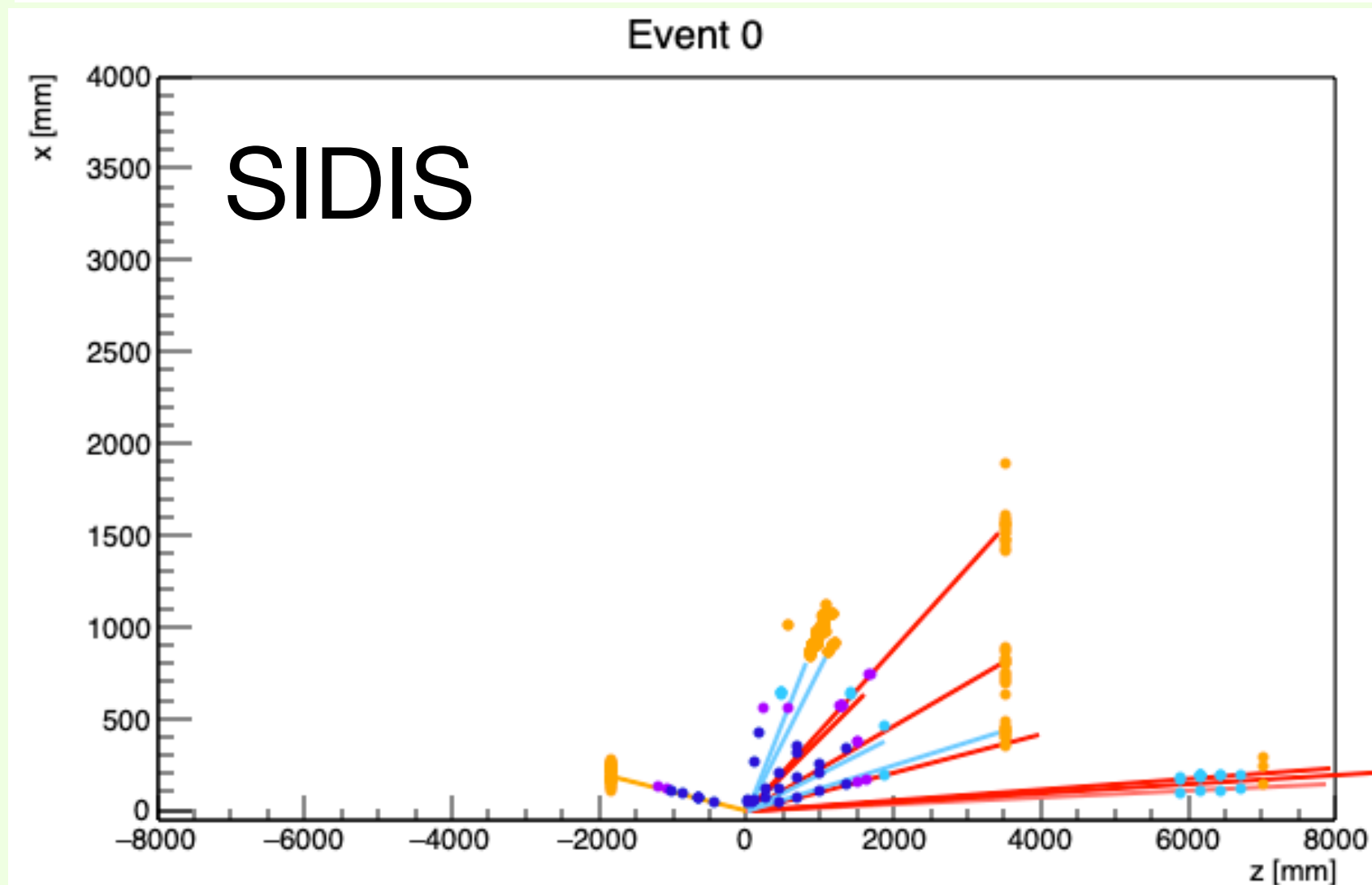
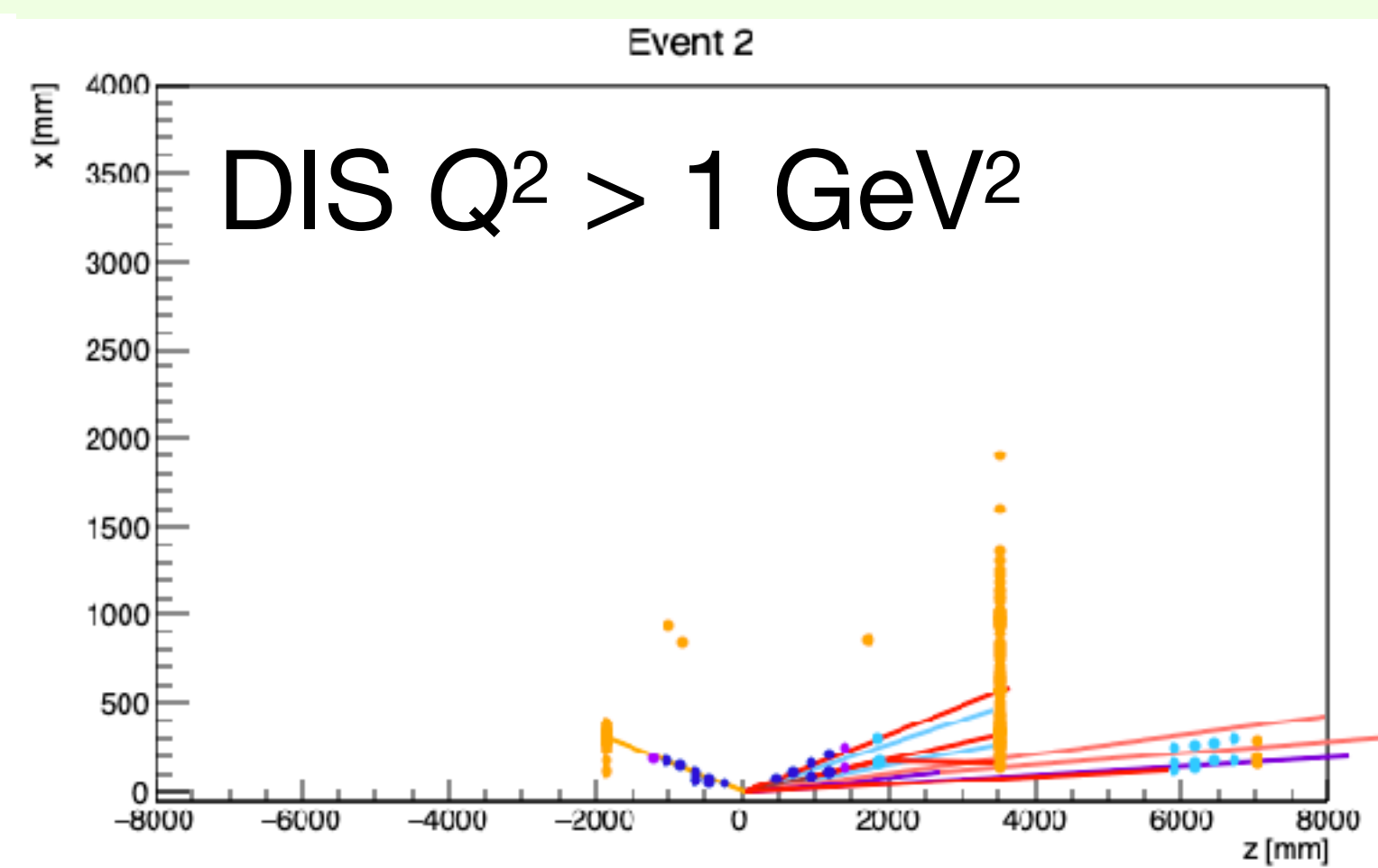
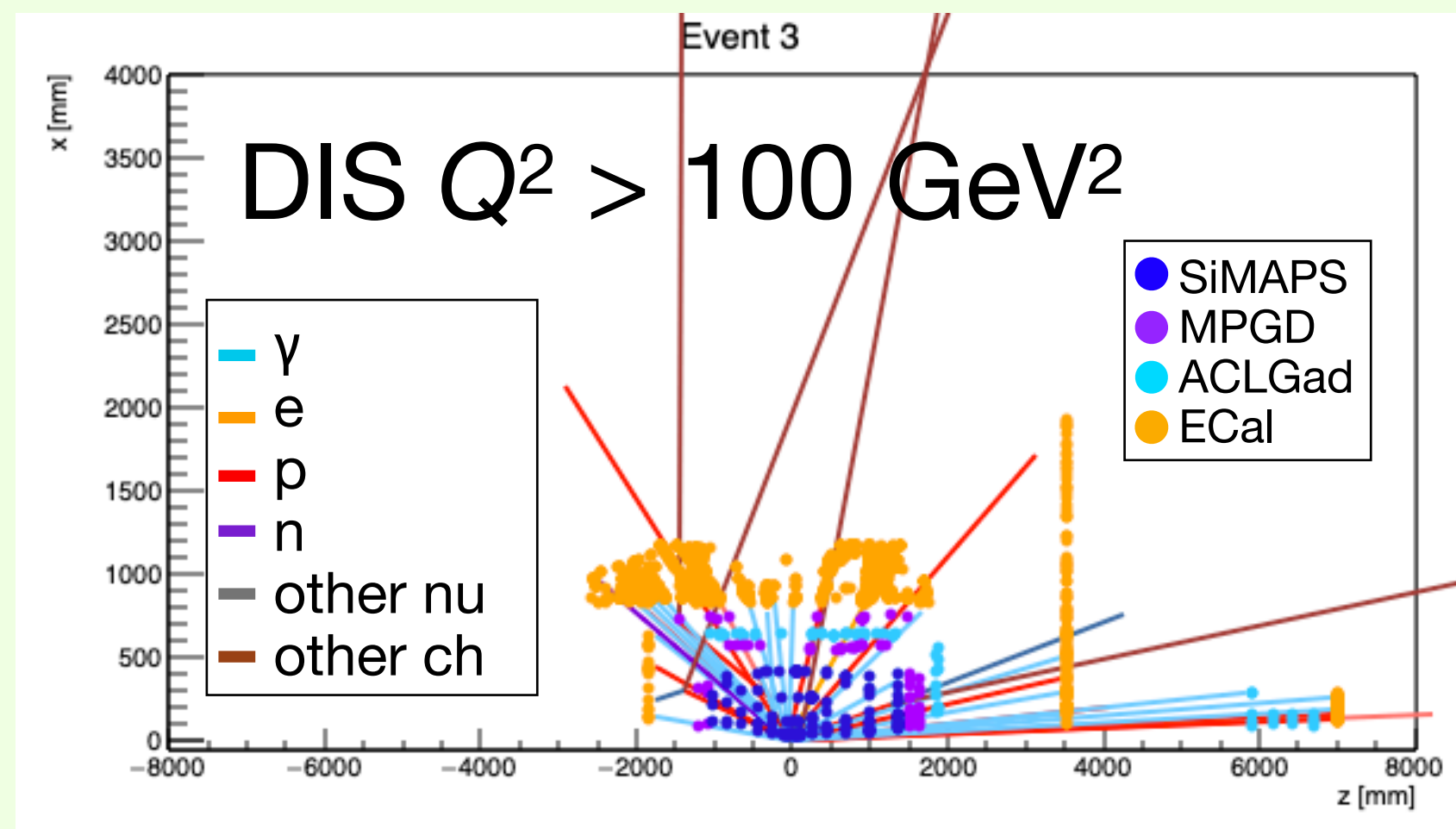
Purity: Fraction of track measurements associated with the dominant MC particle.



ex)
 $\text{purity1} = (\text{orange} \times 3) / (\text{orange} \times 3 + \text{purple} \times 1) = 75\%$
 $\text{purity2} = (\text{purple} \times 1) / (\text{orange} \times 3 + \text{purple} \times 1) = 25\%$



Event Displays



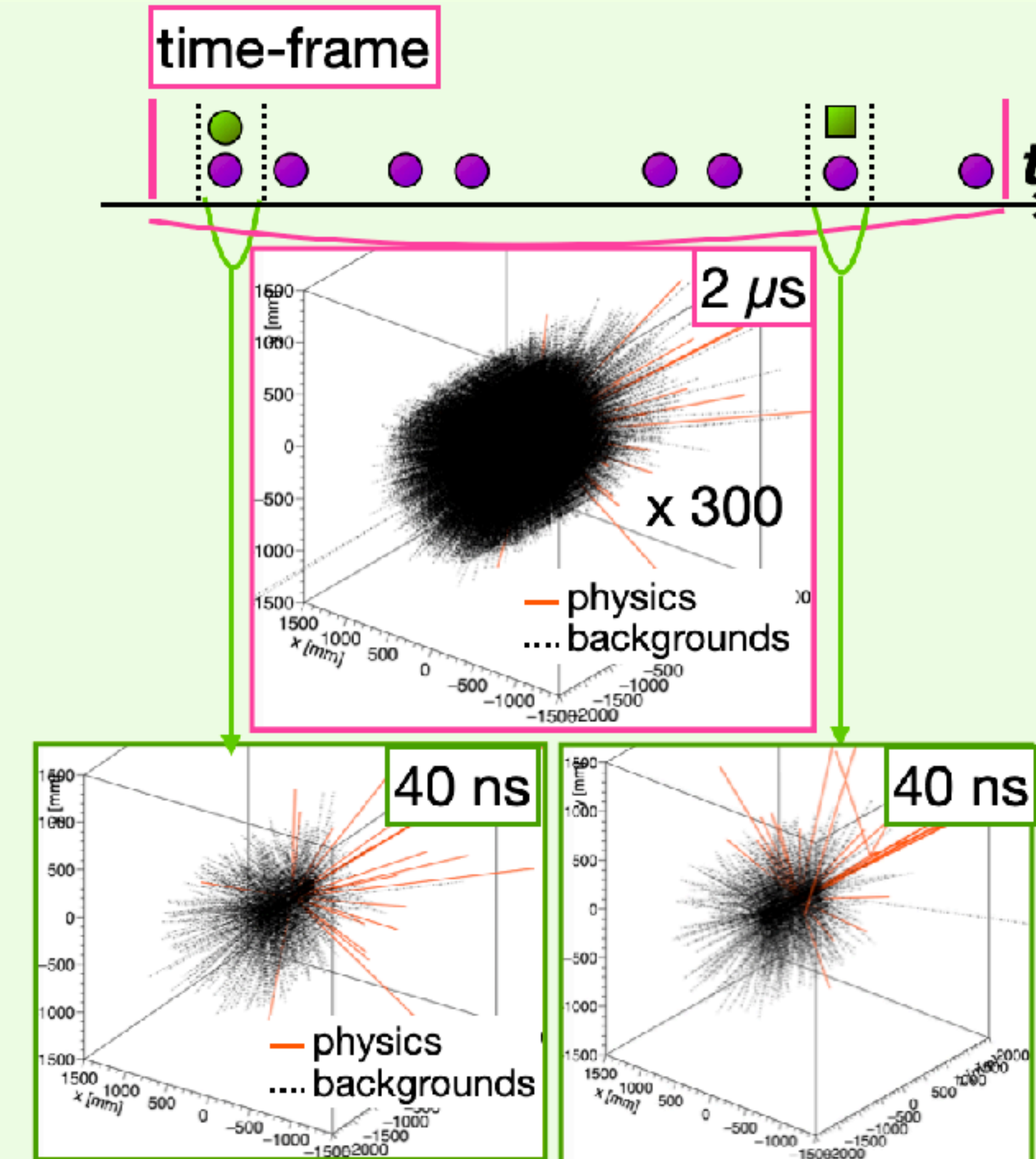
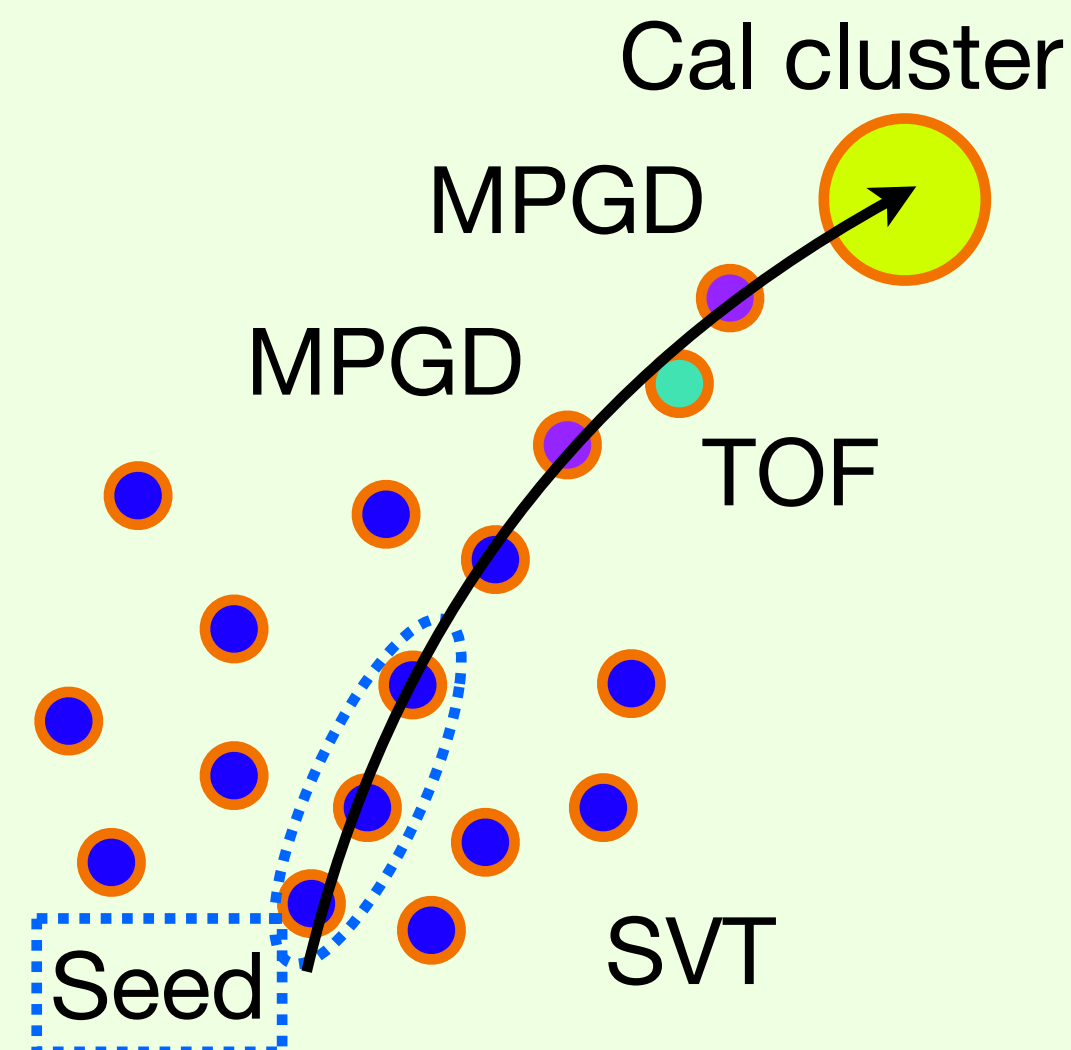
5. About interpolate tracking / tracking W/ timing information

Current ACTS tracking

- The current ACTS tracking starts with **seeds formed from the inner tracking layers** and propagates them outward. However, the inner SVT is expected to have: high hit **occupancy** from synchrotron-radiation backgrounds, high hit density due to its small radius, and relatively **poor timing resolution**.
- Would it be beneficial to instead start from hits in the outer layers and propagate inward?

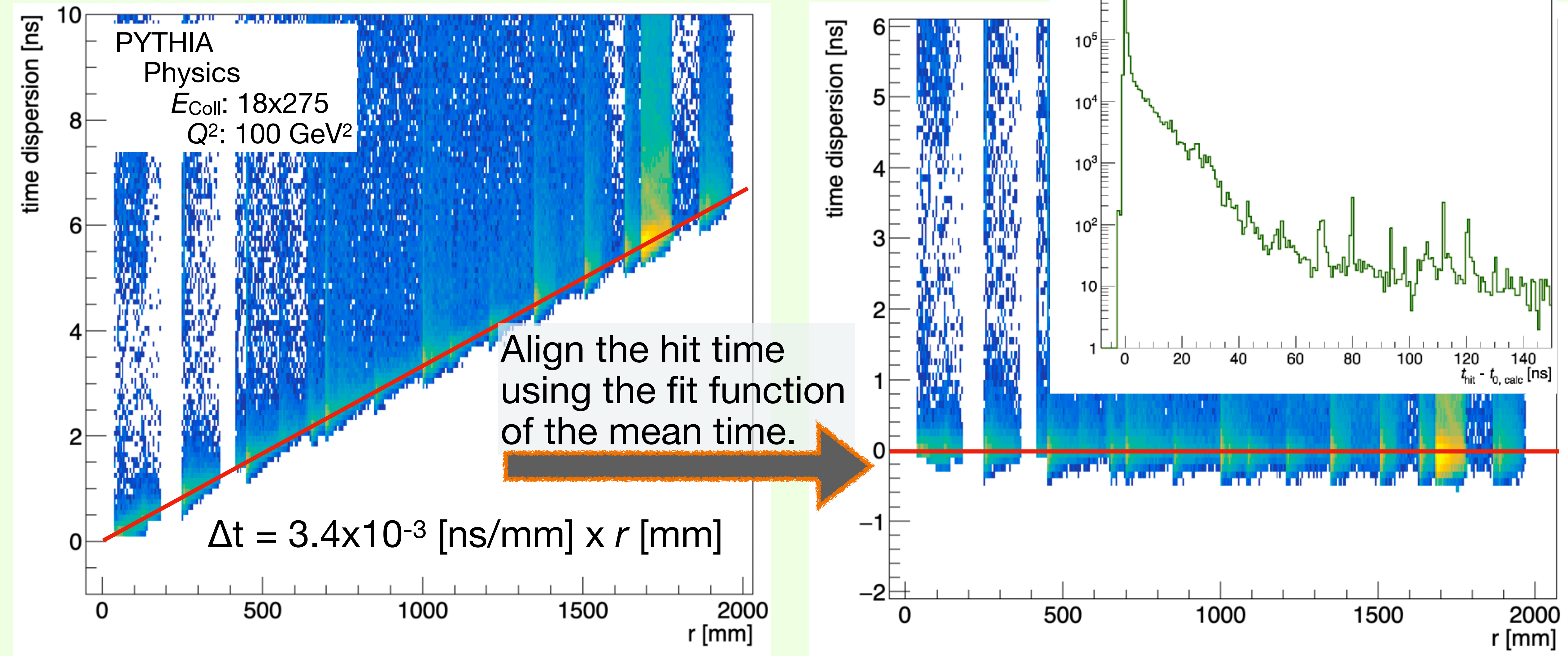
Points for discussion

- Is an **outside-in** tracking approach reasonable for this environment?
- What is the **priority or feasibility** of developing such an approach?
- Is there anything the **event-building stage** could provide to help the tracking?



Hit time dispersion of DIS events

Time alignment of DIS events



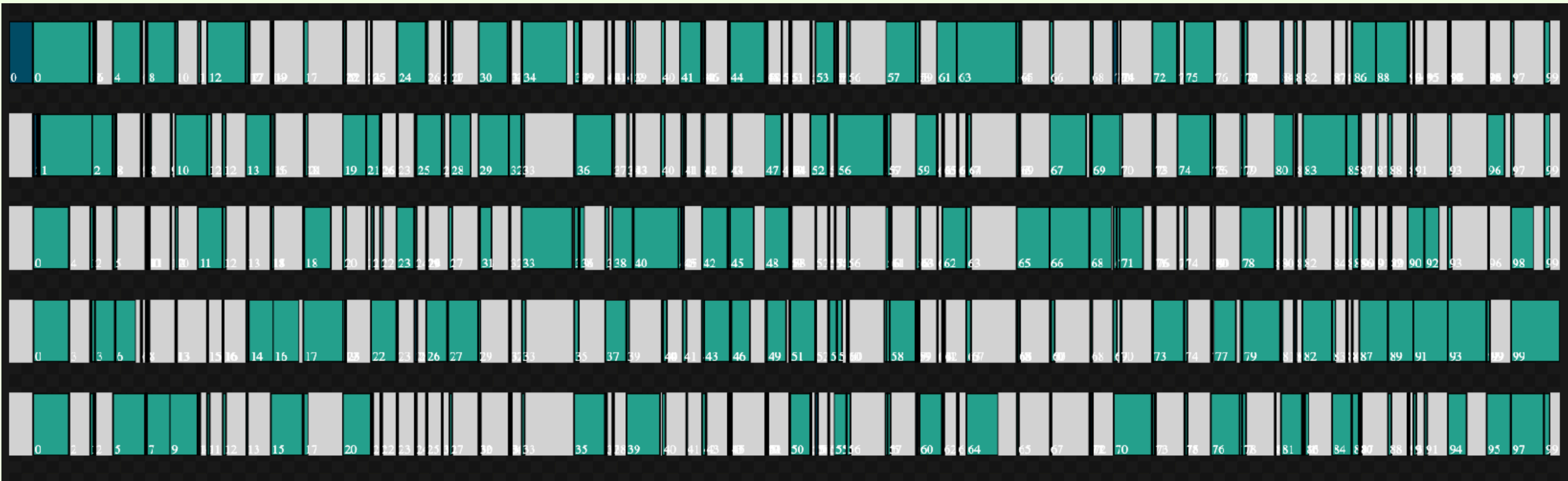
Fit the mean hit time for each region.

nthread test

nthread=1: 456.223 s / 100 time frame



nthread=5: 541.52 s / 100 time frame



Despite increasing the number of threads correctly, the processing time increased.
→ Need to understand the Nathan's evaluator and the reason.

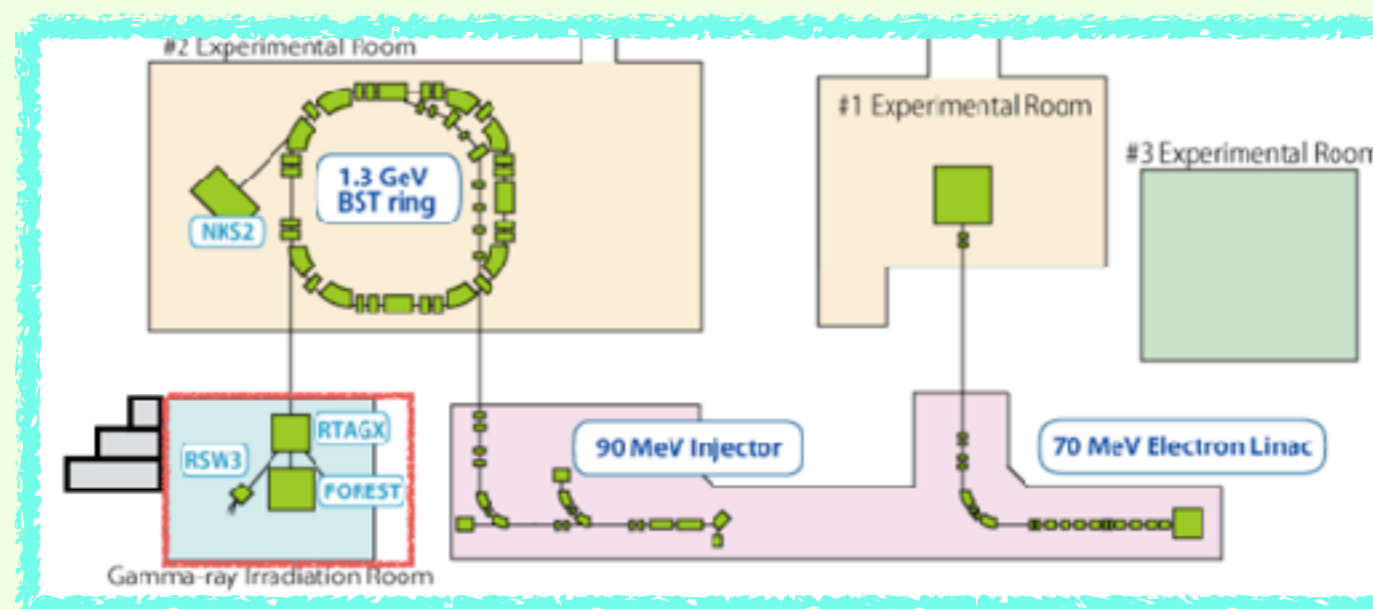
ePIC Streaming DAQ Test with a Local Beam Test

Japan barrel TOF R&D team held a beam test at [RARiS](#) (Tohoku, Japan), 3rd-7th March 2026

GeV- γ beam line:

positron: $< 1 \text{ GeV}/c$, $< 800 \text{ MeV}/c$ ($\sim 30 \text{ Hz}$)

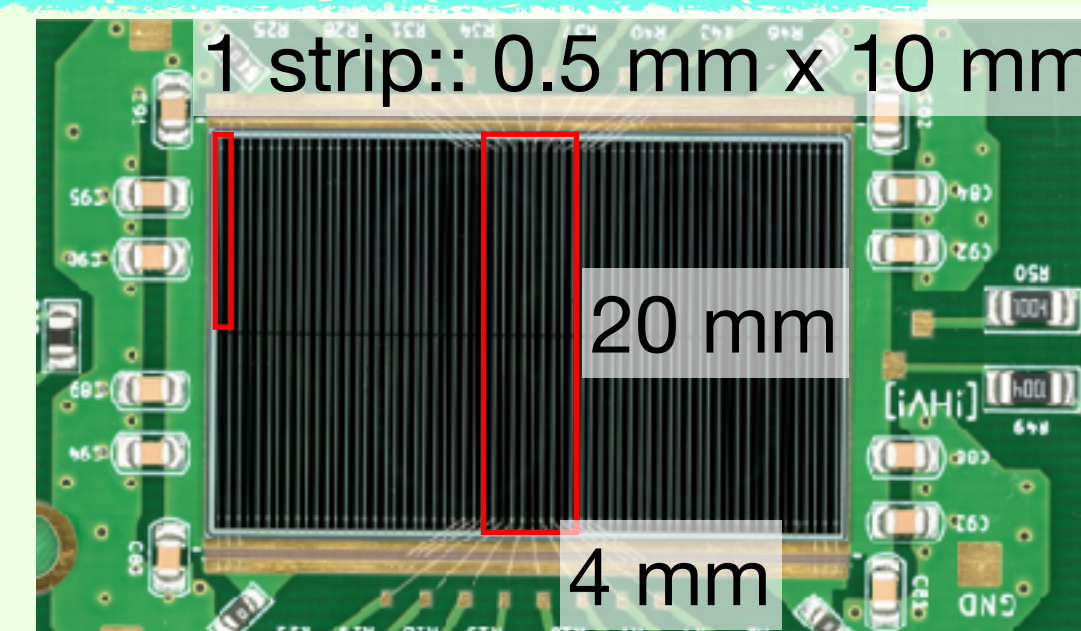
electron: $< 1 \text{ GeV}/c$ ($\sim 30 \text{ Hz}$)



Detector:

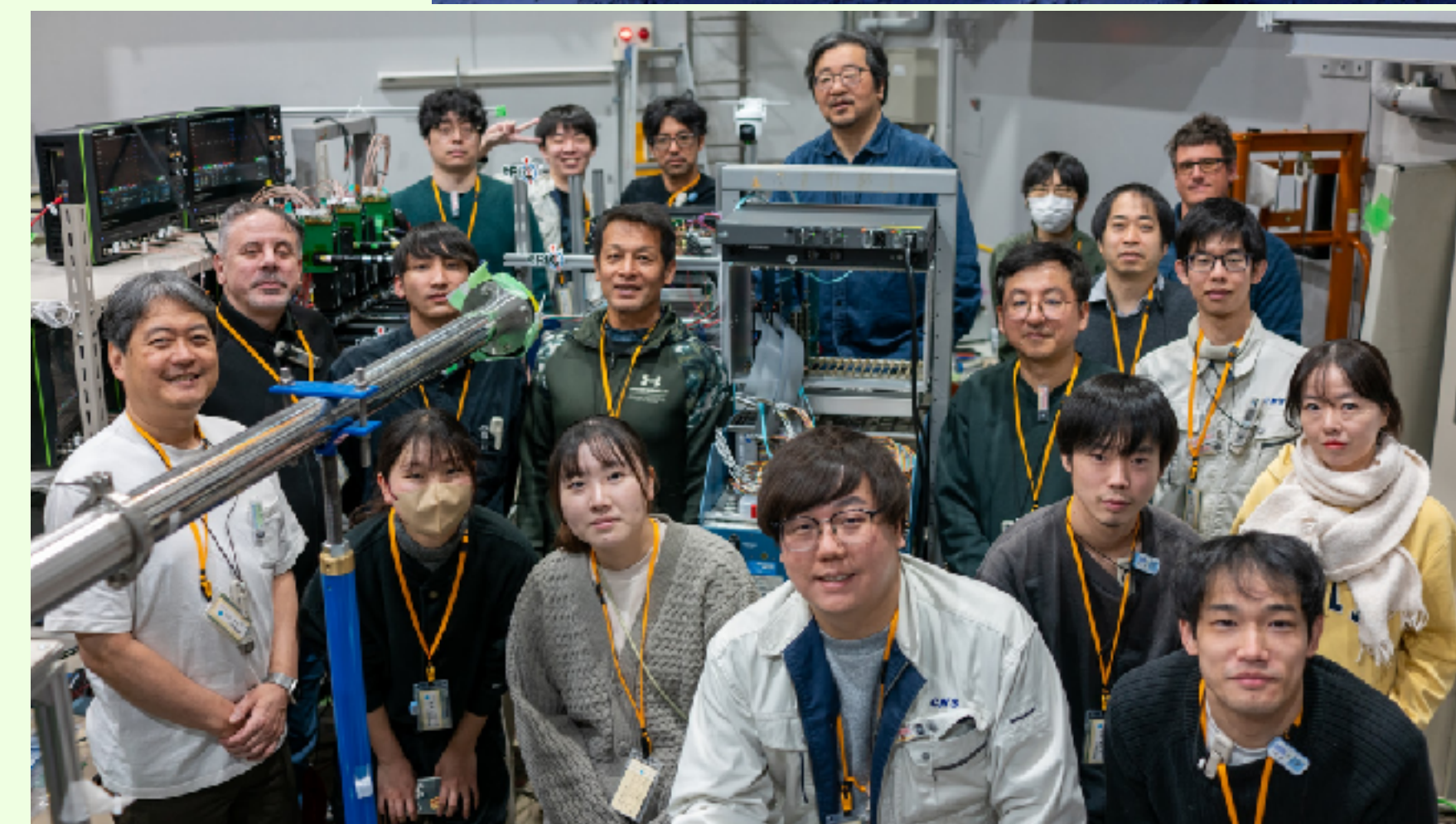
AC-LGAD (AC-coupled Low-Gain Avalanche Diode):

- High time resolution $\sim 30 \text{ ps}$
- High granularity



Beam test

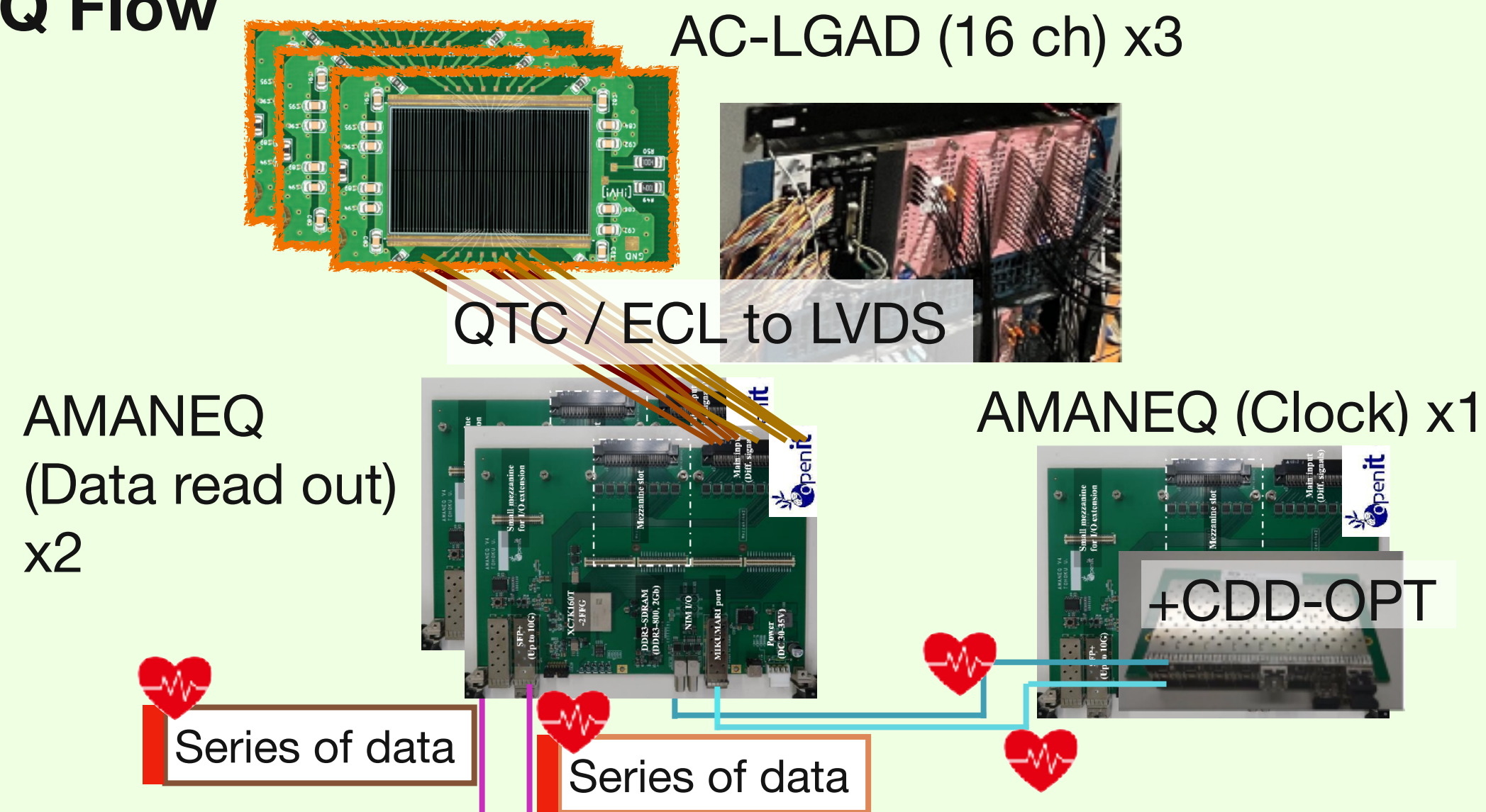
- This beam test was conducted by 23 people.
- We had three parallel lines.
- **line1: mini ePIC streaming test with SPADI using AC-LGAD**
 - N.Kobayashi (RCNP, Osaka Univ), T.Takahashi (RCNP, , Osaka Univ), S. Y. Ryu (RCNP, Osaka Univ), T.Gunji (QNSI, Univ Tokyo), M.Kaneta (Tohoku Univ), K.Ono (Shinshu Univ)
- line2: AC-LGAD performance test using Waveform (Ochiroscope)
- line3: RFSoc DAQ test for AC-LGAD performance estimation



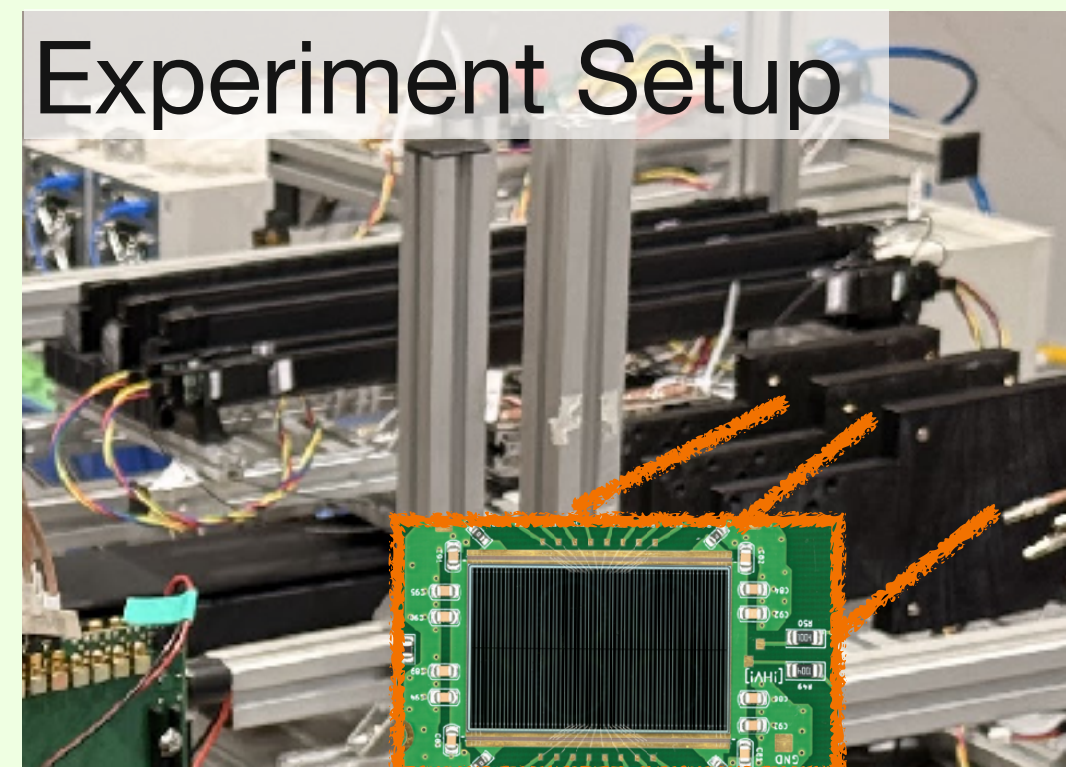
Developed Two Streaming DAQ Lines

Tried mini ePIC streaming orchestration (Echelon0-Echelon2)

DAQ Flow



Experiment Setup



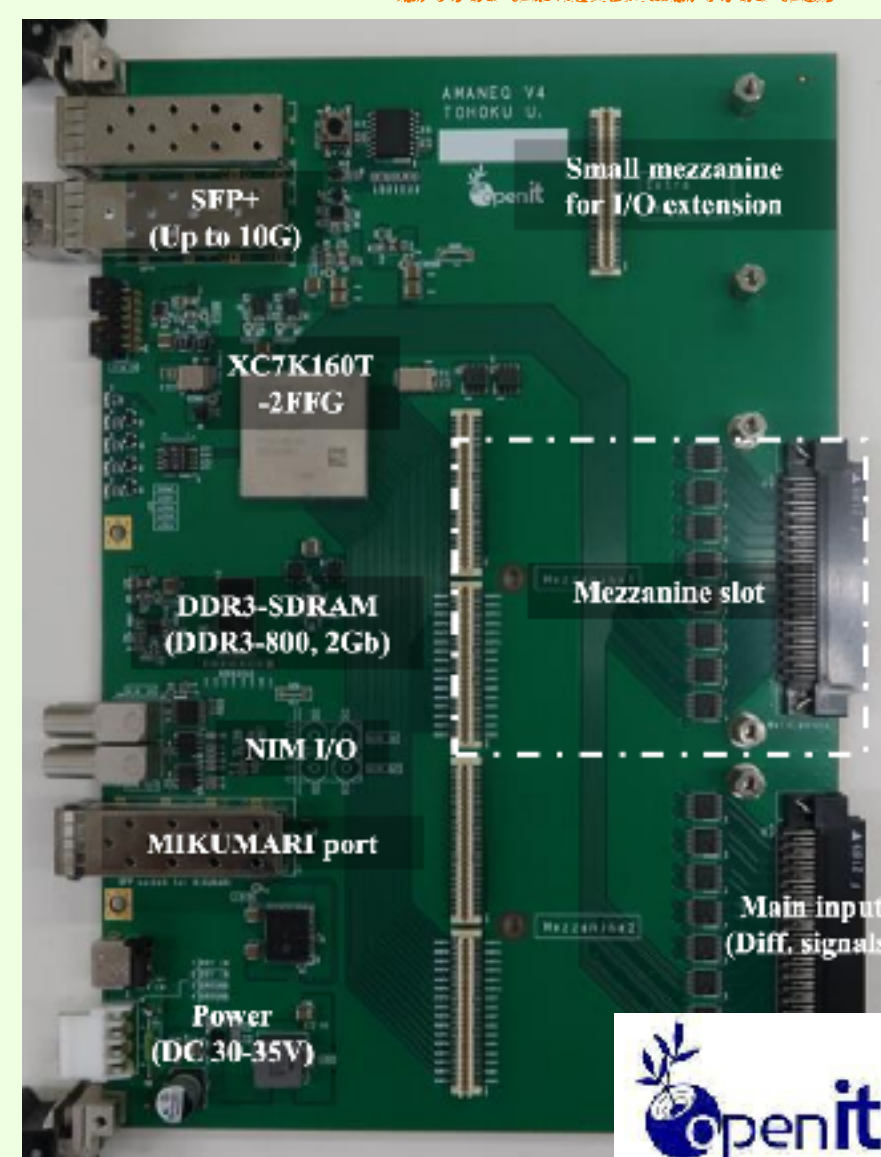
AMANEQ/MIKUMARI Setup



AMANEQ [ref]

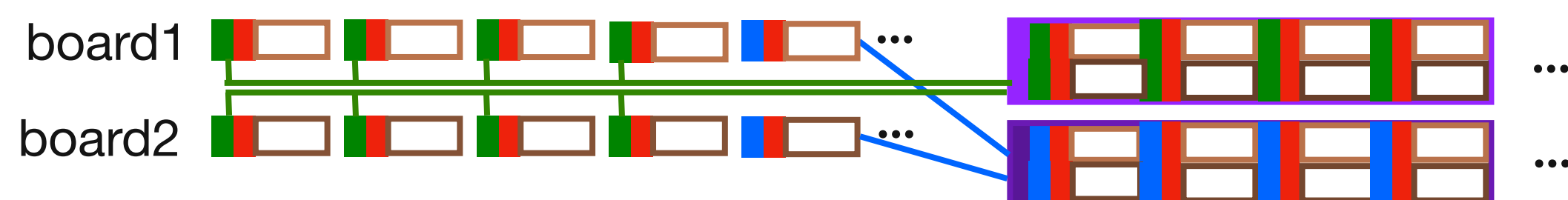
(A main electronics for network oriented trigger-less data acquisition system)

- Kintex7 with speed grade -2
 - Transceiver bandwidth up to 10 Gbps (10 GbE)
- Two mezzanine slot for functionality extension
 - HR-TDC
 - Clock distribution
 - Input-channel extension
 - etc...
- A jitter cleaner (CDCE62002)
- DDR3-SDRAM as a de-randomizer



Desktop PC (NestDAQ) 524 μs

- Sampler
- SubTimeFrameBuilder - TimeFrameBuilder



- Output

1. File sink (EDM4eic) → XRootD → Univ Osaka
2. Send a data as message queue

(3. →GPU logic filter→TimeSlicer→FileSink)

AMANEQ→EDM4eic→ElCrecon

We had not use ElCrecon for actual data not a simulation file produced by npsim.

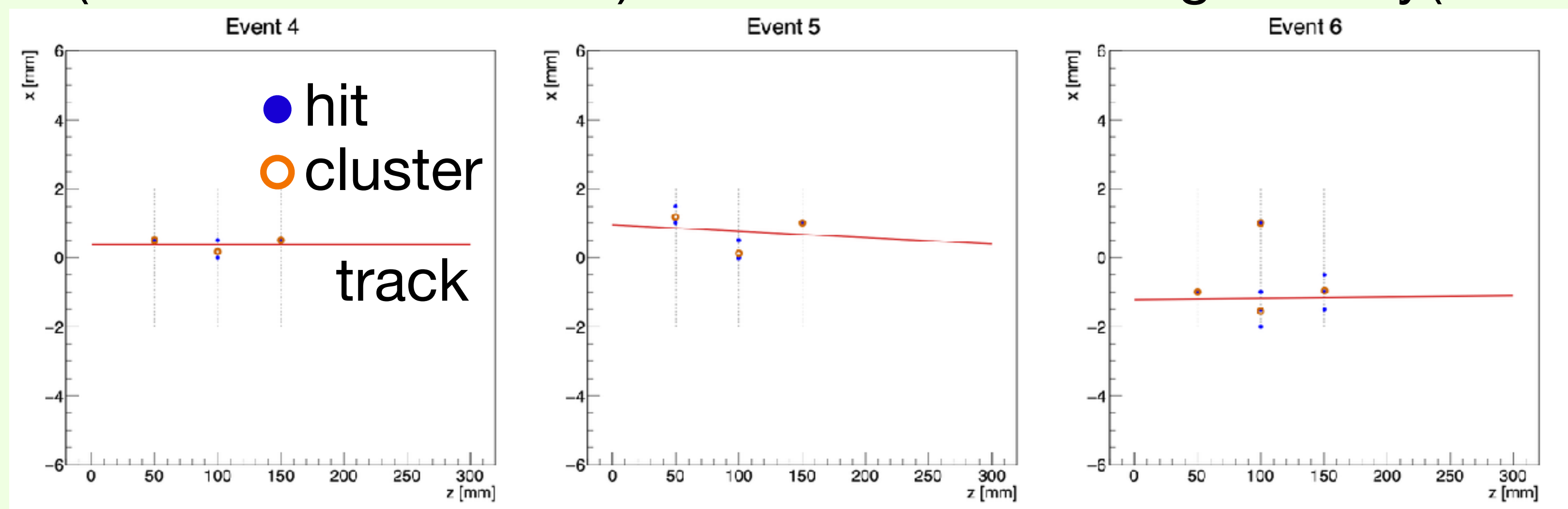
→ Establish to use ElCrecon for a beam test and actual data with SPADI-A DAQ system.

- AMANEQ (TDC) → Sink a file written as edm4eic::RawTrackerHit (N.Kobayashi: RCNP Univ of Osaka) [[git:EDM4eicSink.cxx](https://github.com/EDM4eicSink.cxx)]

- EDM4eic file → ElCrecon (K.Ono: Shinshu Univ, T.Kumaoka: QNSI Univ of Tokyo)

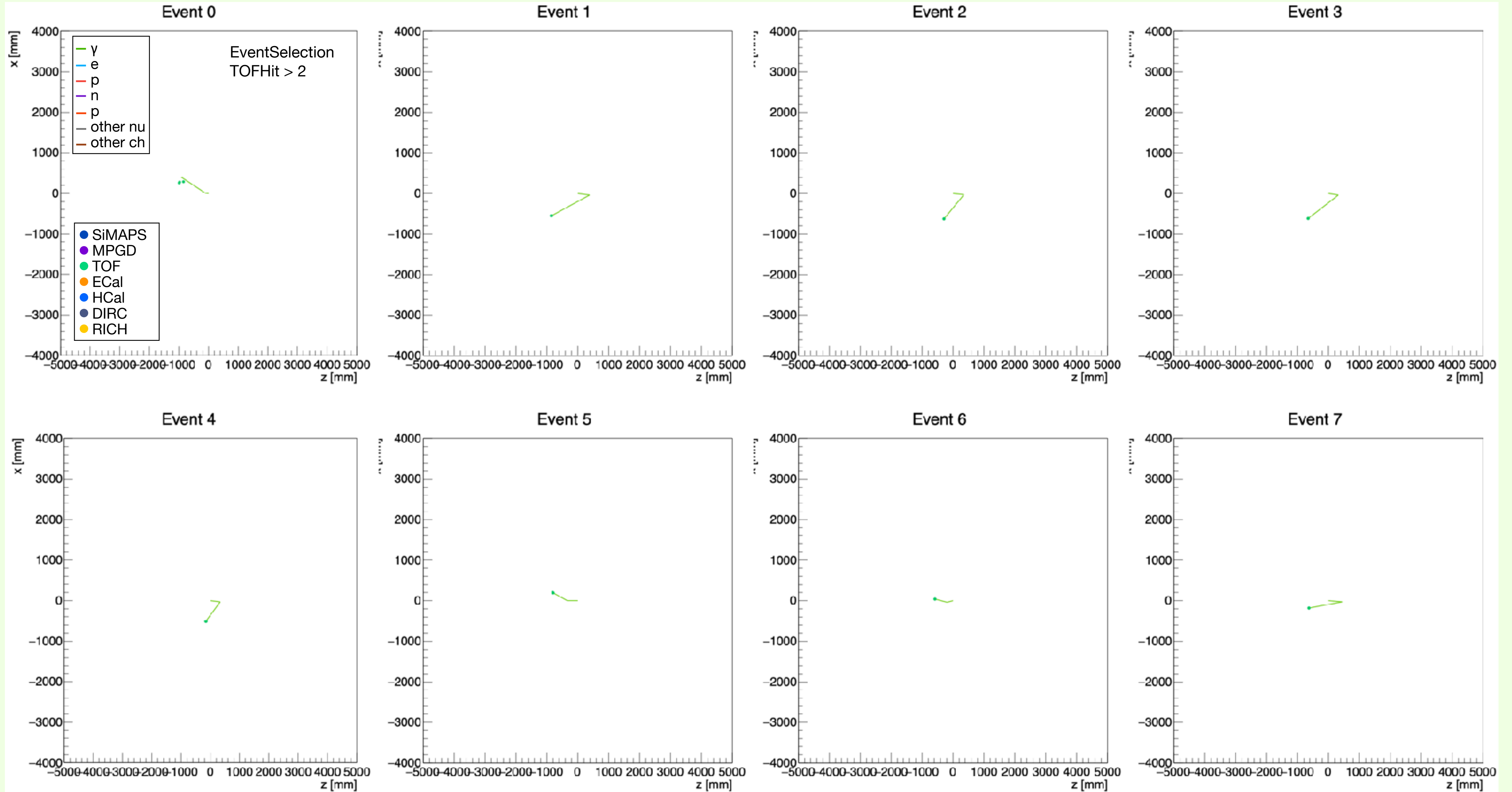
Run the ElCrecon (based on [BTOF Plugin](#)):

Original EventSource (edm4eic::RawTrackerHit)→ LGADHitCalibration_factory (edm4eic::TrackerHit)→ LGADHitClustering_factory(edm4eic::Measurement2D)

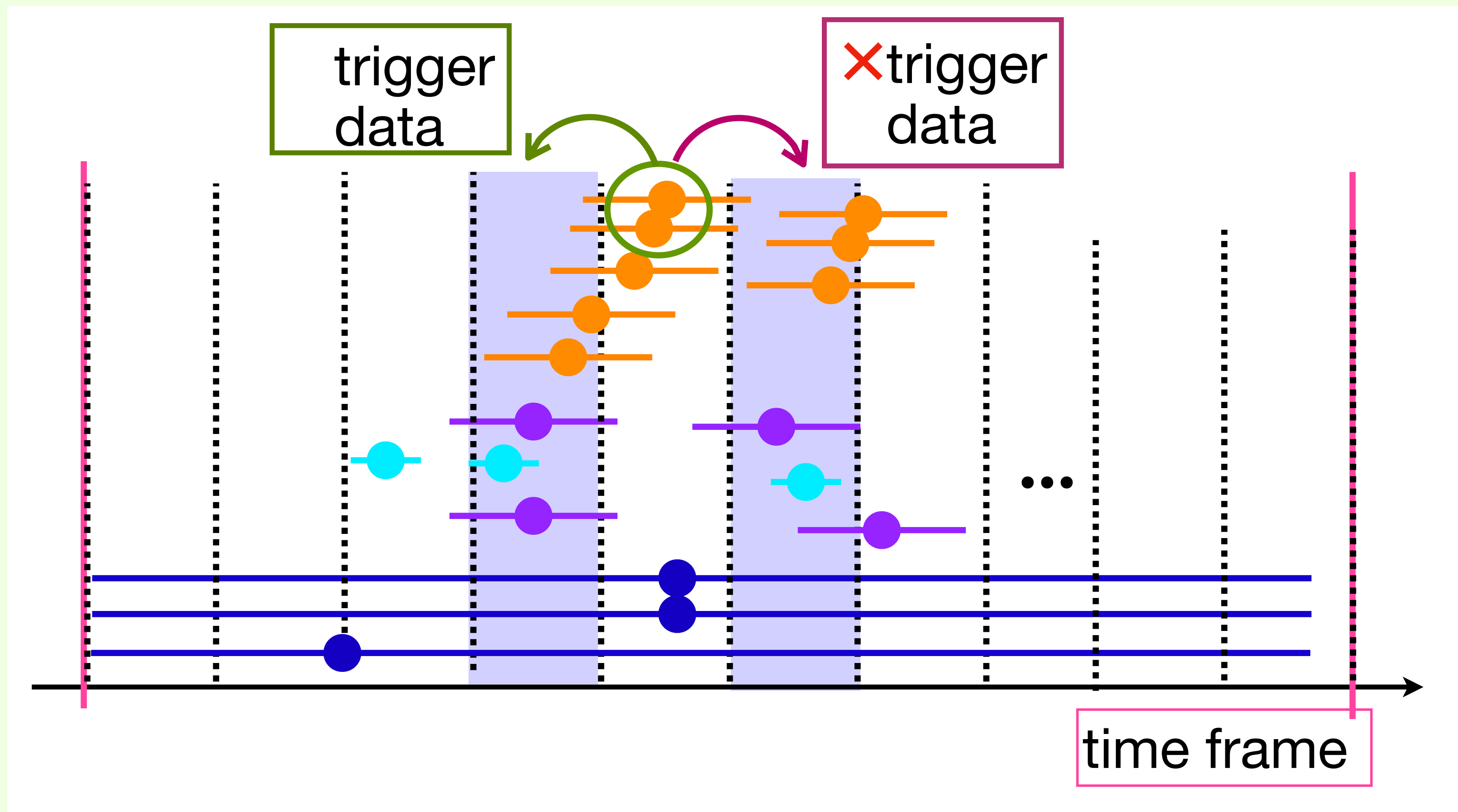


Outlook - Implement the timeframe splitter in this stream.

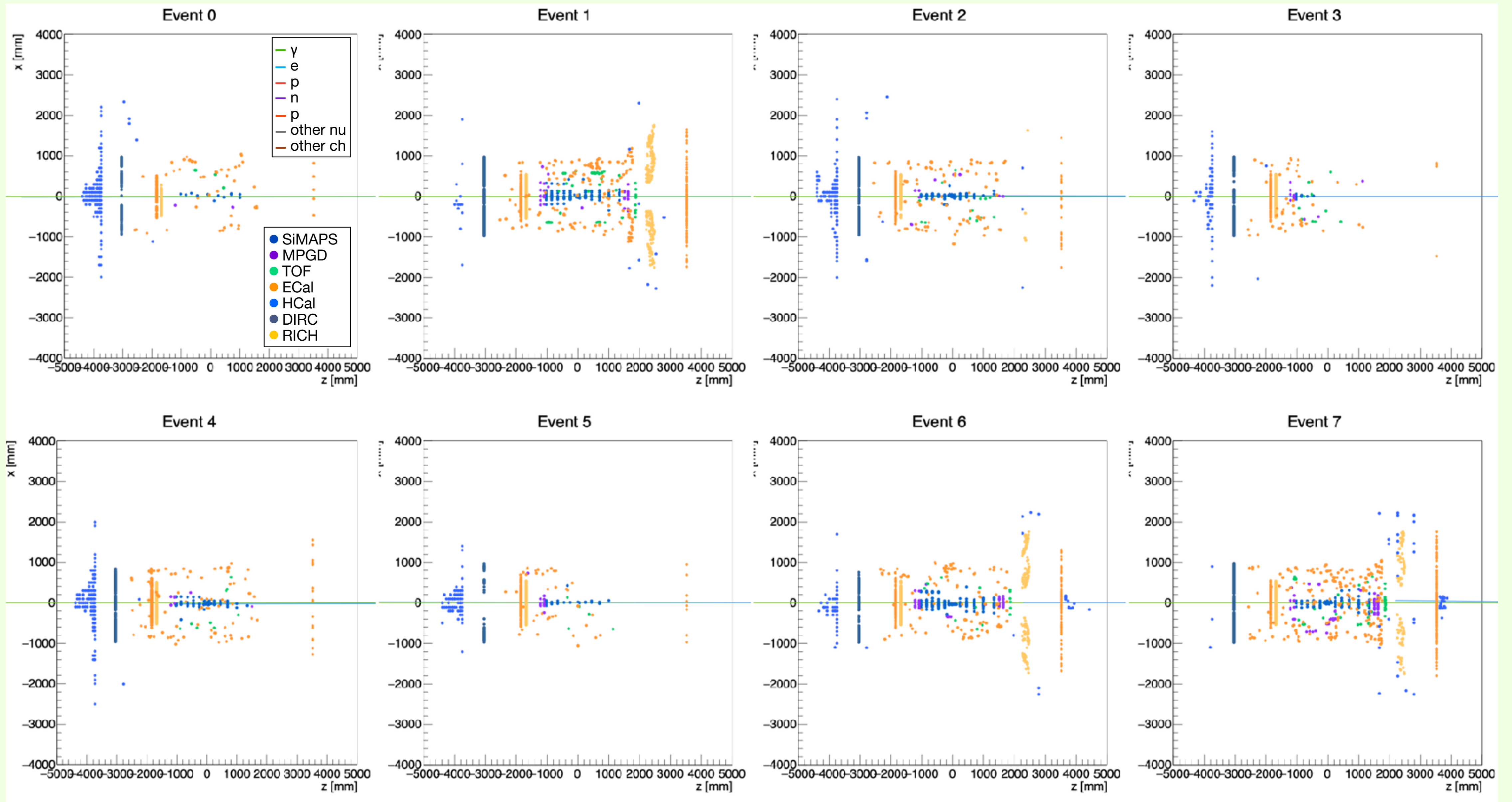
Synchrotron radiation



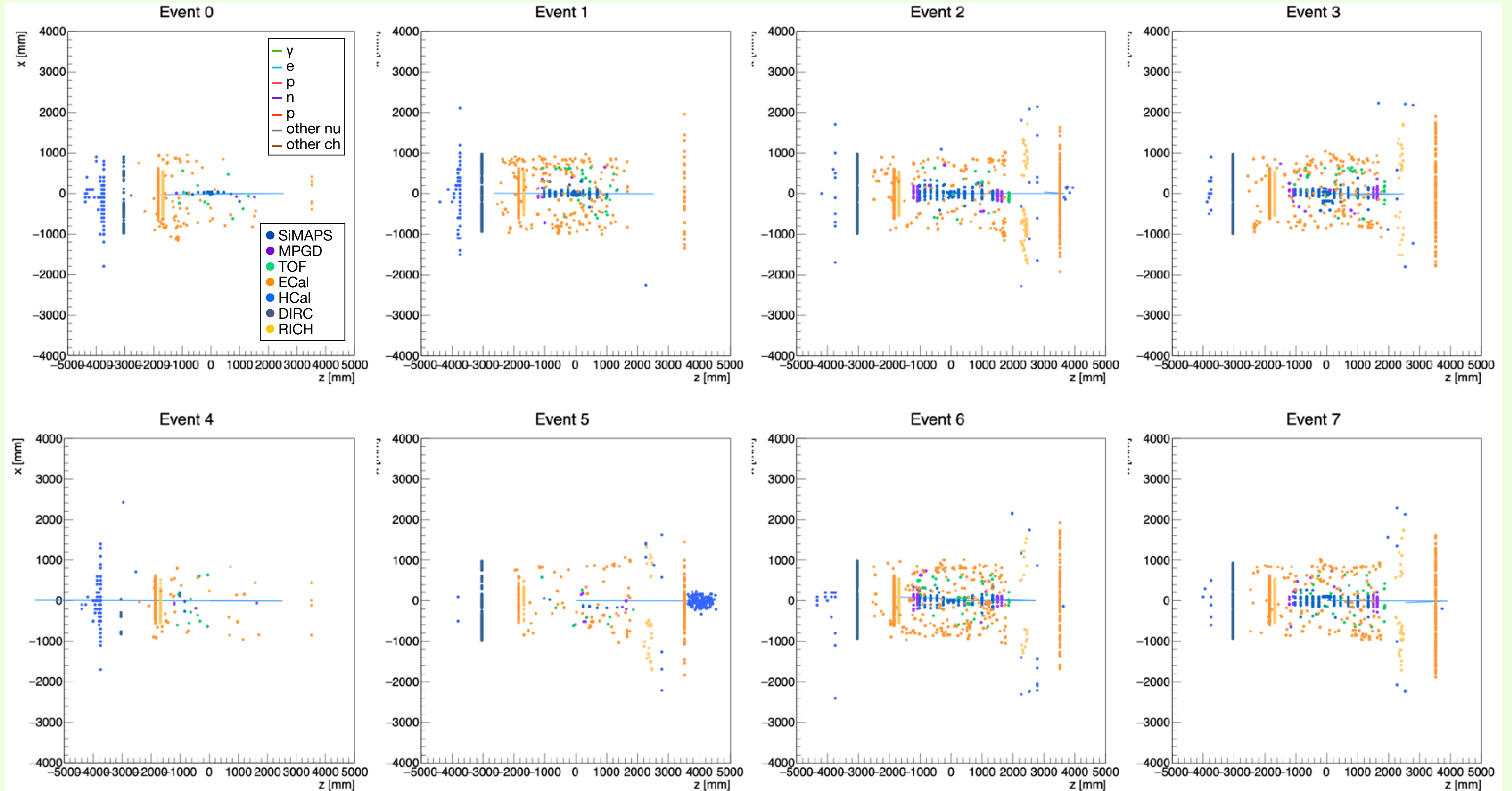
Falling within the

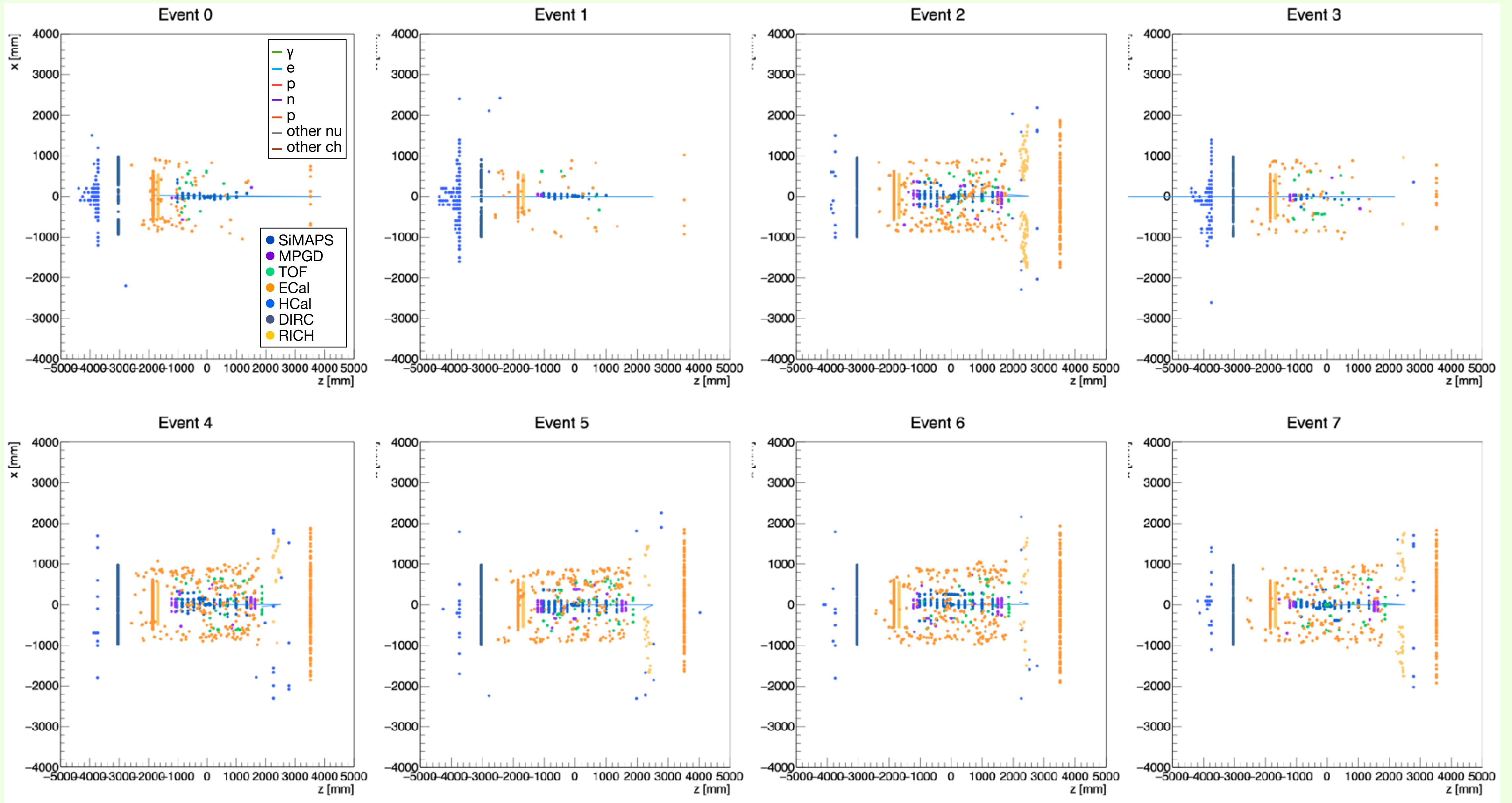


e Bremsstrahlung

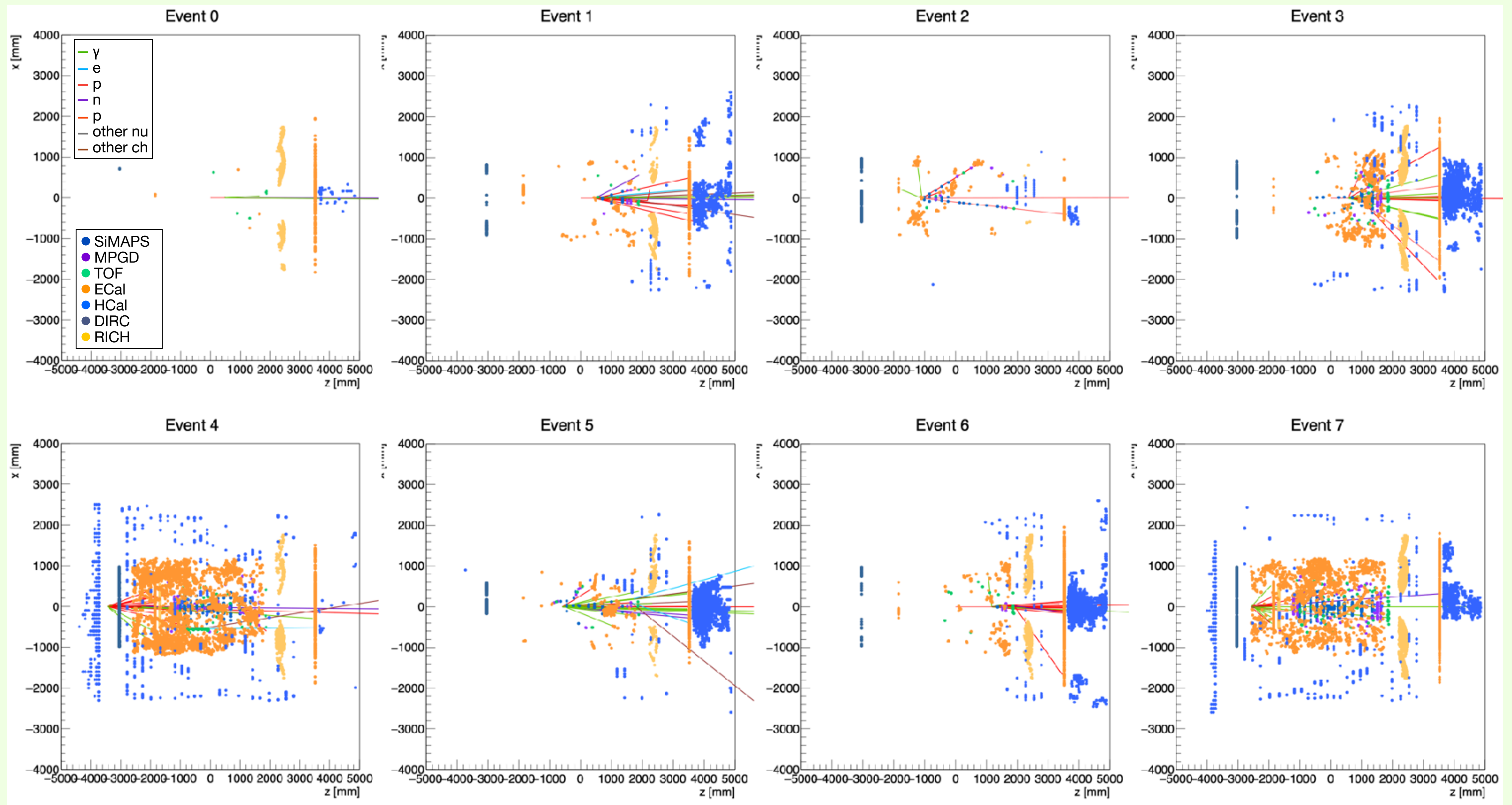


Coulomb losses





Proton beam gas



How to validate the data quality?

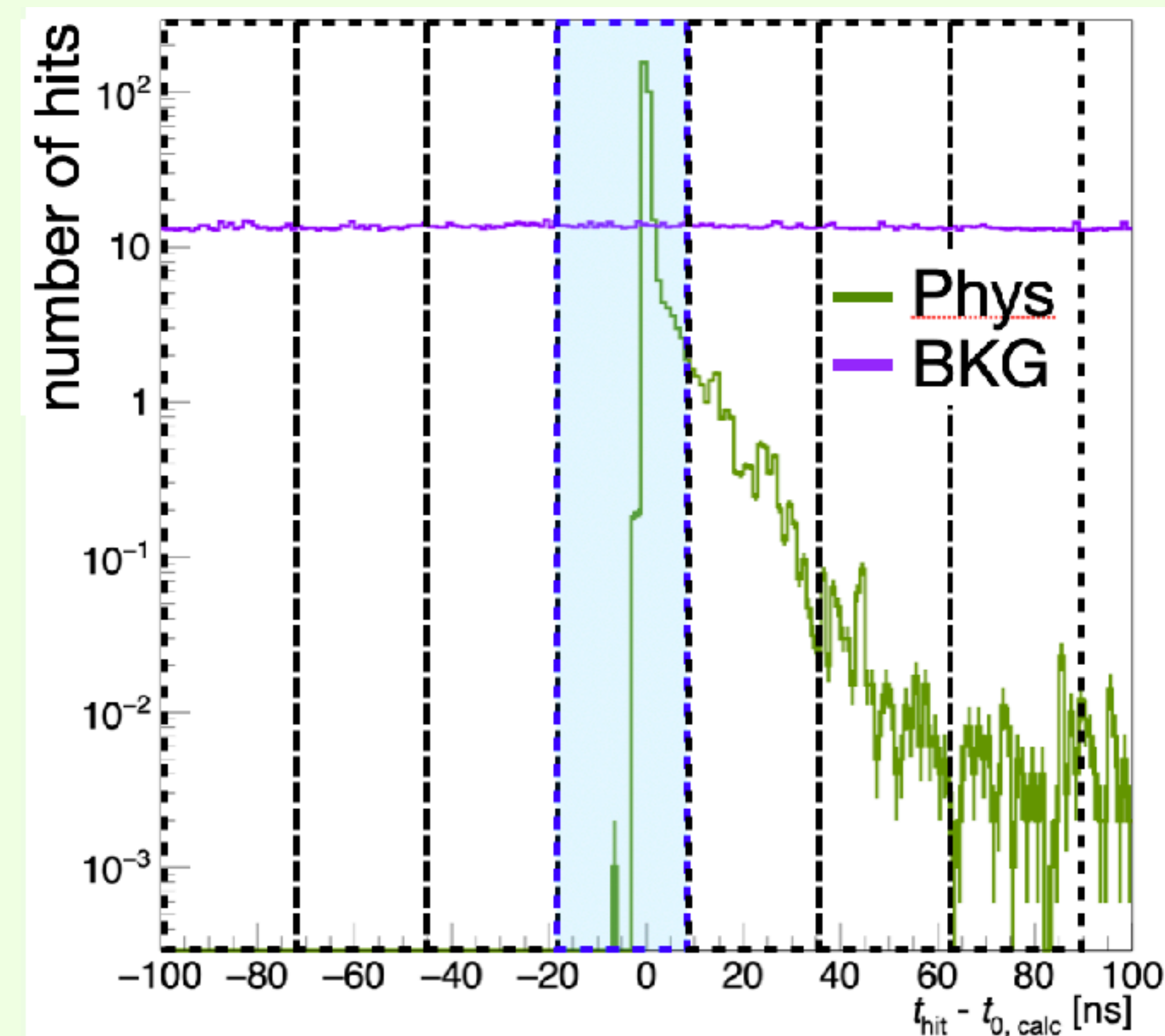
The current pre-event identification does not guarantee the quality of the physics event. It only indicates that a physics collision occurred within the triggered short time window.

→ Some hits may fall outside the selected time window, particularly those from slow, low-momentum particles or from the late stages of particle showers.

→ However, this may not be a problem as long as the physics events can still be reconstructed with acceptable quality.



How can we validate the triggered events?



Tracking Efficiency

$$\text{efficiency} = \frac{N(\text{reconstructed ch particle})}{N(\text{MC ch particle } |\eta| < 3.5)}$$

