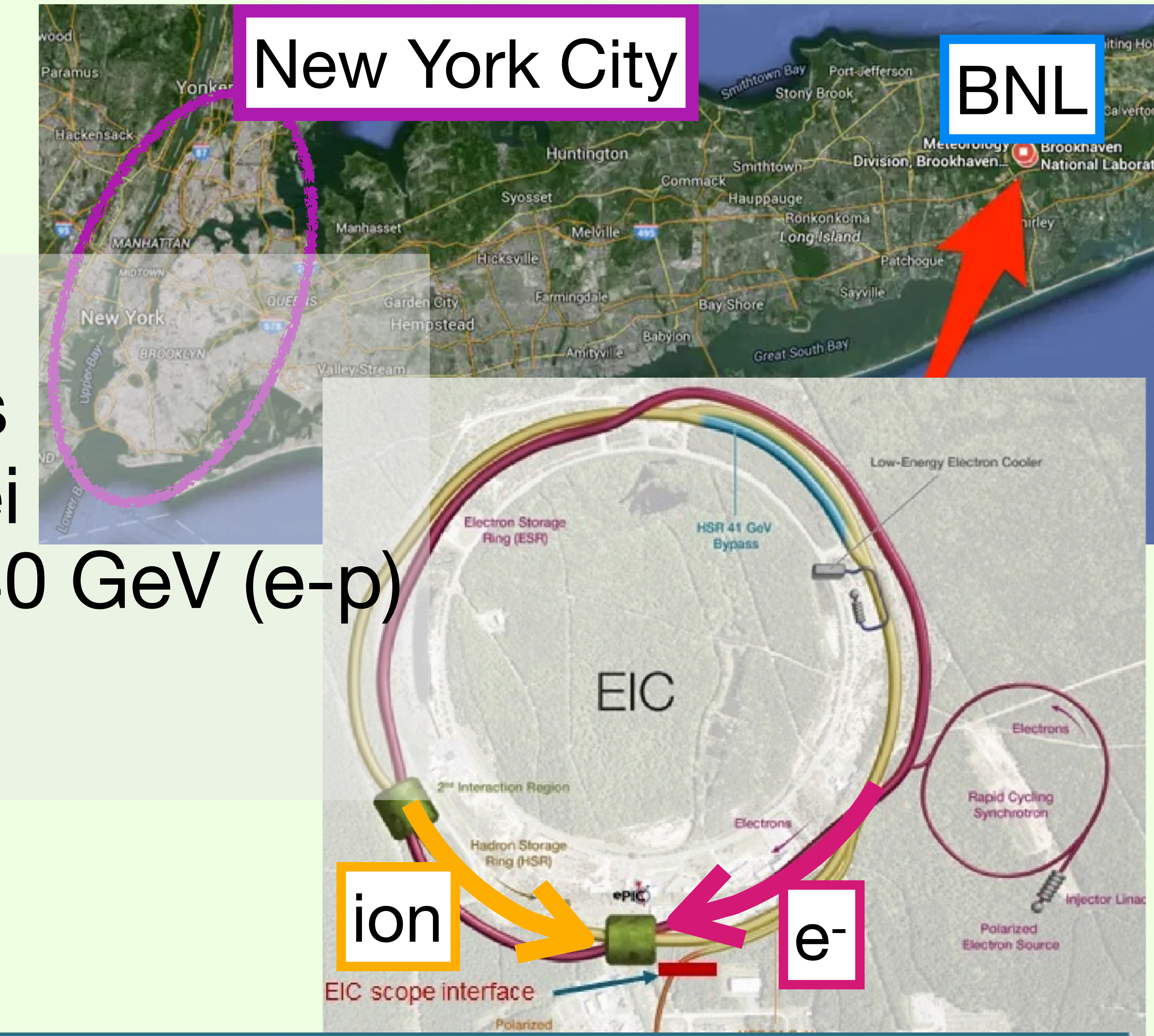
Development of streaming data reconstruction for ePIC experiment at EIC

Quark Nuclear Science Institute, The University of Tokyo
Takuya Kumaoka

Electron-Ion Collider

EIC (Electron Ion Collider) (2035-)

New accelerator of Brookhaven National Laboratory(BNL)

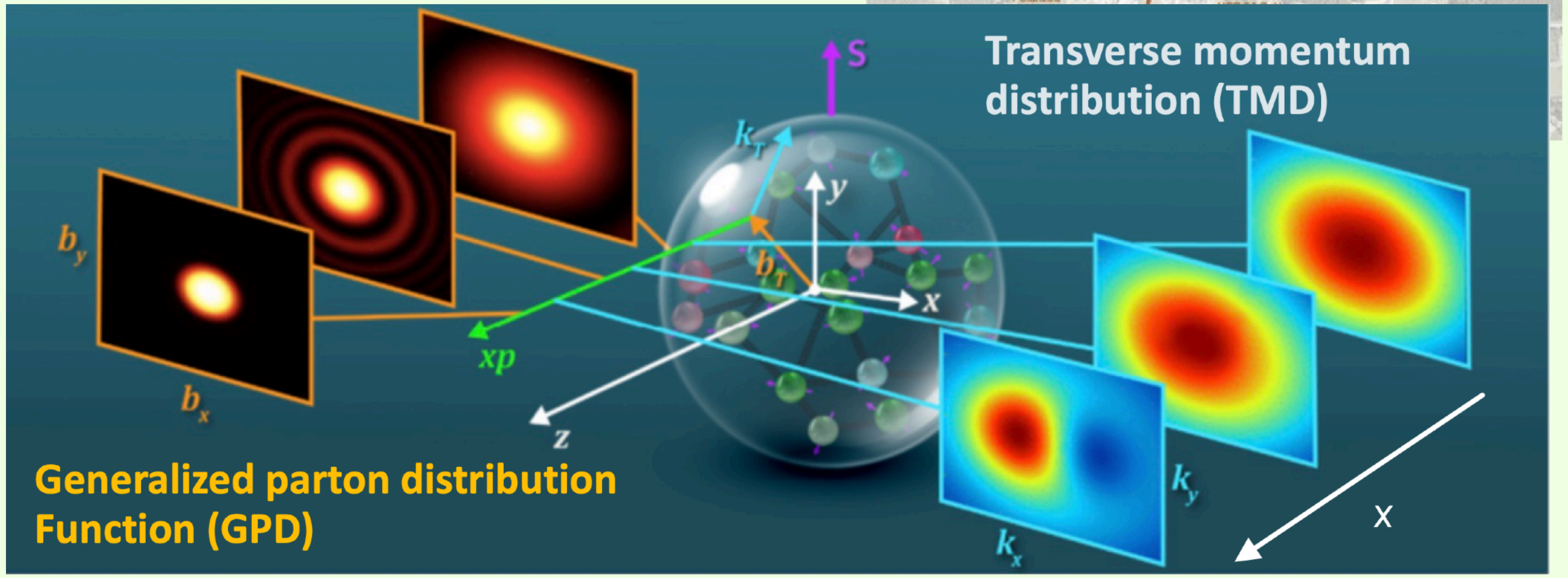


- **Characteristics**:
- A high degree of beam polarization* (~70%) for electrons and Ions
- Availability of ion beam from deuteron to the heaviest stable nulcei
- Variable center of mass energies ~20-100 GeV, upgradable to ~140 GeV (e-p)
- High collision luminosity* ~10^{33}-10^{34} $cm^{-2}s^{-1}$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

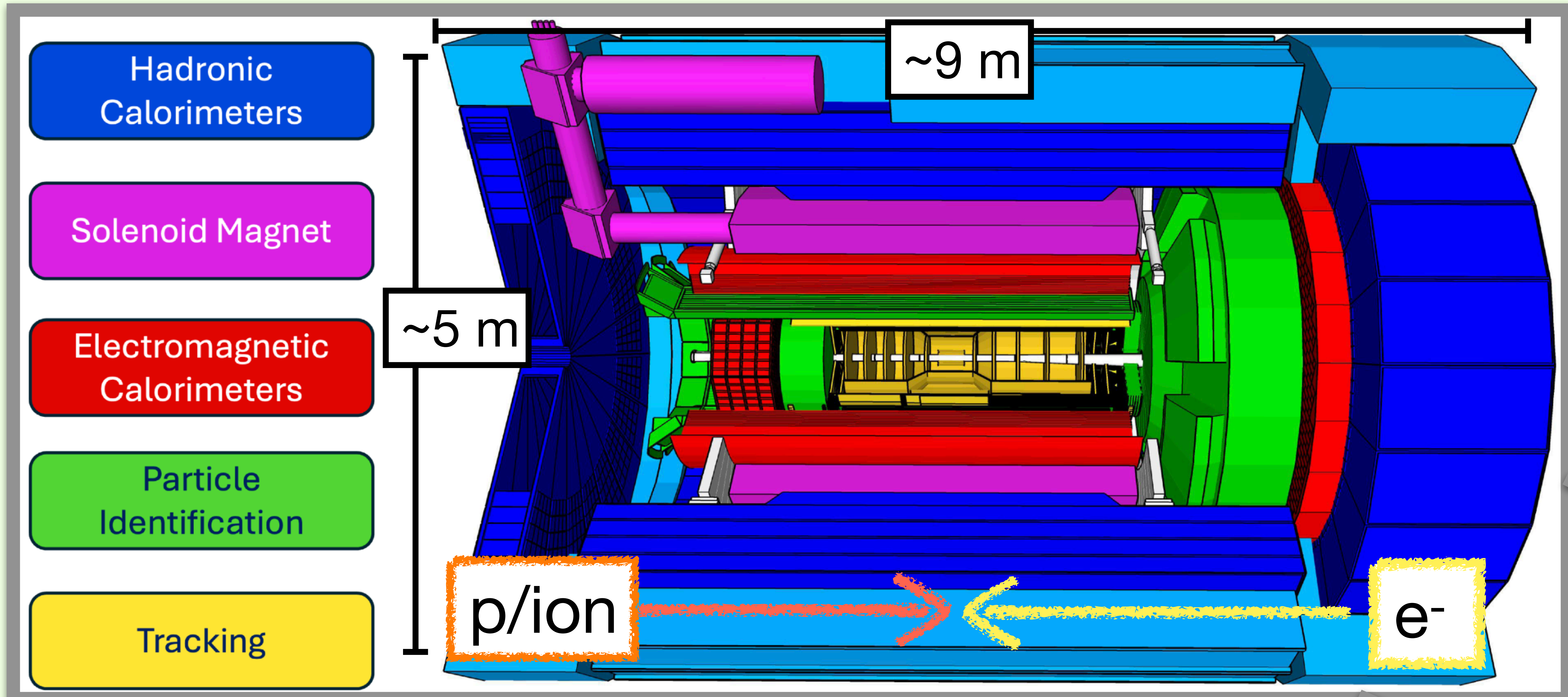
Main Physics Targets:

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?



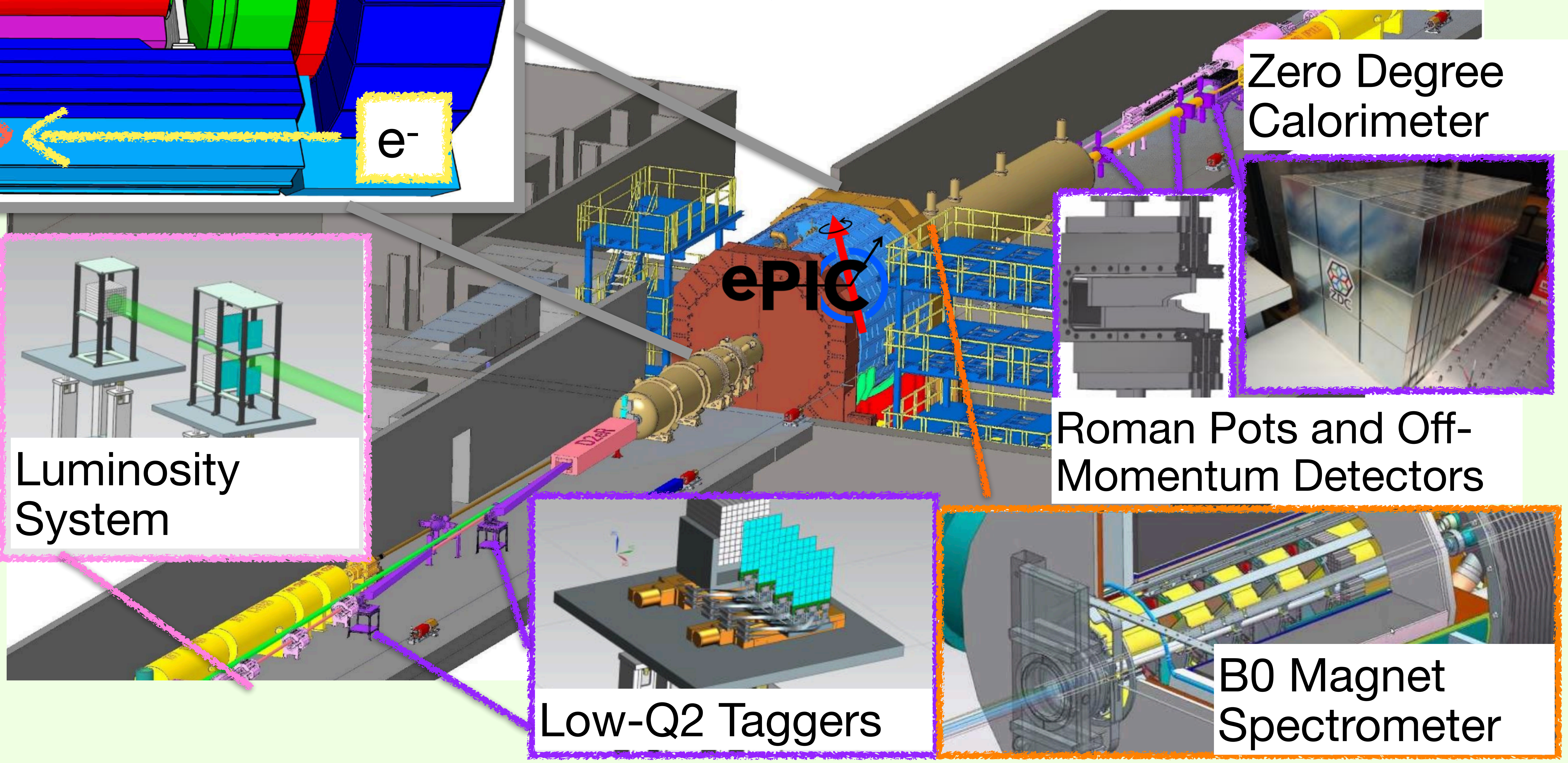
ePIC (electron-Proton/Ion Collider) detector



Calorimetry:
- 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
- MPGDs (μRWELL/μMegas)

PID:
- hpDIRC
- pfRICH
- dRICH
- AC-LGAD (~30ps TOF)



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.

- **Track5 25th May: S.Rahman, Exploring Potential Pathways to Accelerate ePIC Detector Simulation**
- **Track4 26th May: S.Rahman, Scaling ePIC Simulation Production: Distributed Workflow and Data Management**
- **Track4 27th May: H.Szumila-Vance, The ePIC Streaming Computing Model**
- **Track8 28th May: S.Kay, A.Prozorov, ePIC User Learning Training and Documentation Strategies**

Purpose of this study

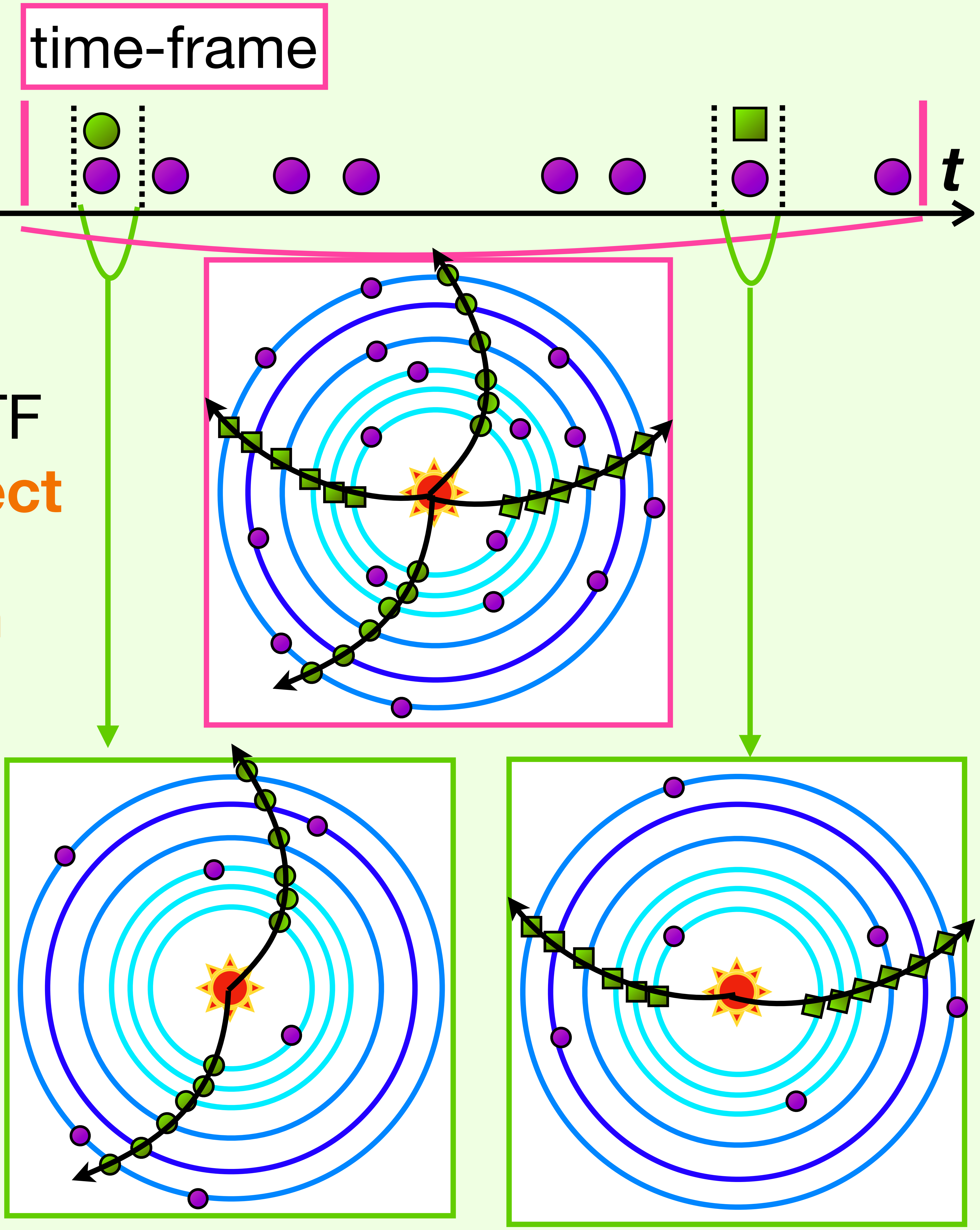
- Time-frame (TF) ~0.6 ms / super TF ~ x1000 TFs
- Acquire DIS events at 500 kHz (@ 18x275 GeV^2)
- Expected background rate (dominated by Synchrotron radiation) : 3.3 GHz(@ 18x275 GeV^2)

+ detector noises

→ 300 DIS events/TF and background ~2.0×10^6 events/TF

→ **Essential to deploy streaming reconstruction to select physics event and reject background**

+ **With triggerless readout, events must be identified in software.**



Purpose

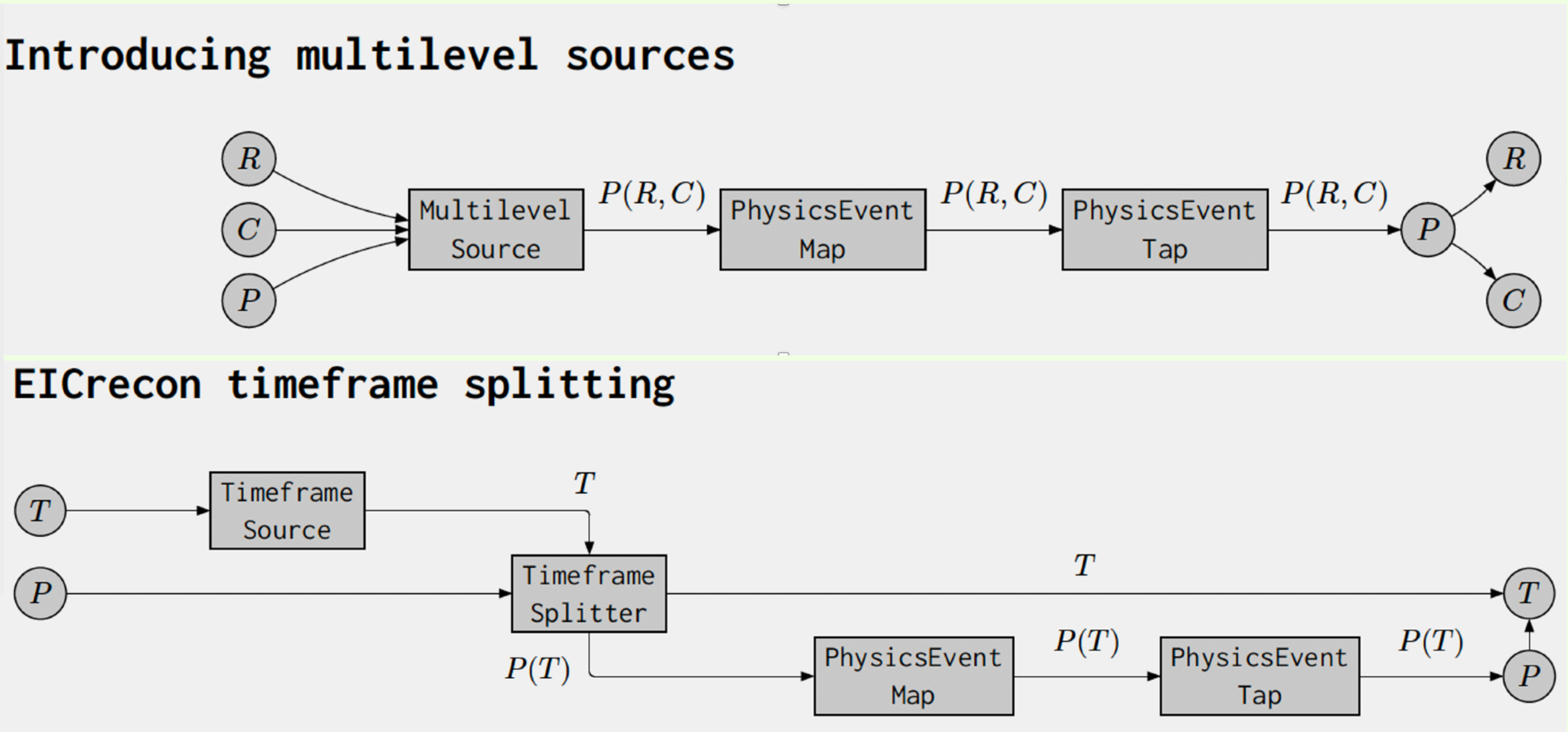
Development of streaming reconstruction and algorithms for event selection (EICrecon)

EICrecon framework

- Base framework: JANA2

General purpose software framework for future experiments developed by Jefferson Lab (JLab) （used by GlueX experiment）

Multi-threaded JANA2 framework provides a component-level hierarchical decomposition of data boundaries into Run, Timeframe, PhysicsEvent, and Subevent levels.
→Essential for streaming processing



- Data Model: PODIO (Plain Old Data Input/Output)

API data driven model developed by CERN

Based on yaml: edm4eic (git) ← edm4hep (git) ← podio (page)

- Tracking: ACTS (A Common Tracking Software)

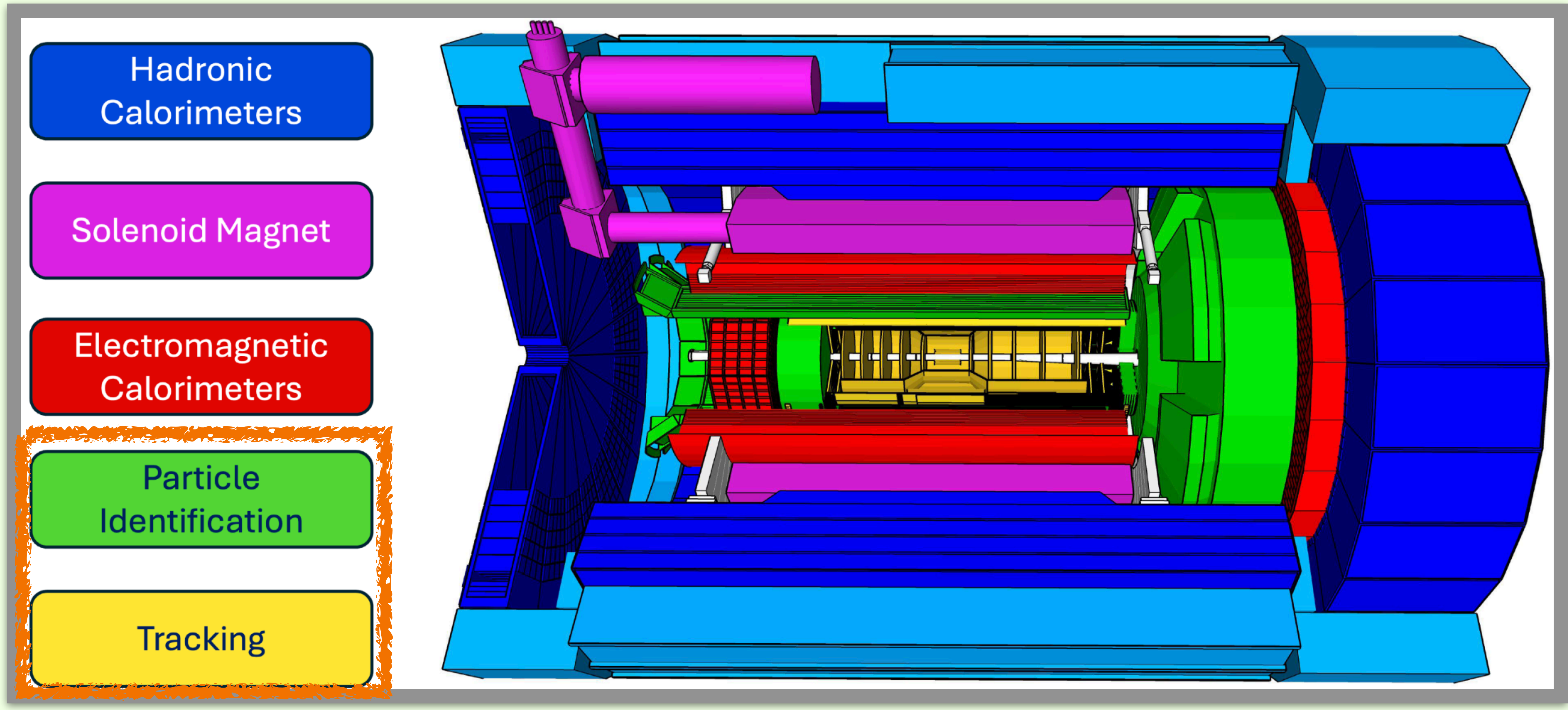
A modular, experiment-independent tracking library developed at CERN

Online Event Identification System

Sub-detectors used for the pre-identification

Need some sub-detectors's hit coincidence to extract a physics event from a time frame.

→ Use tracker detectors (**ToF**, **MPGD** and **Silicon (MAPS)**) for the event identification. (But all detectors' data is streamed)



ToF

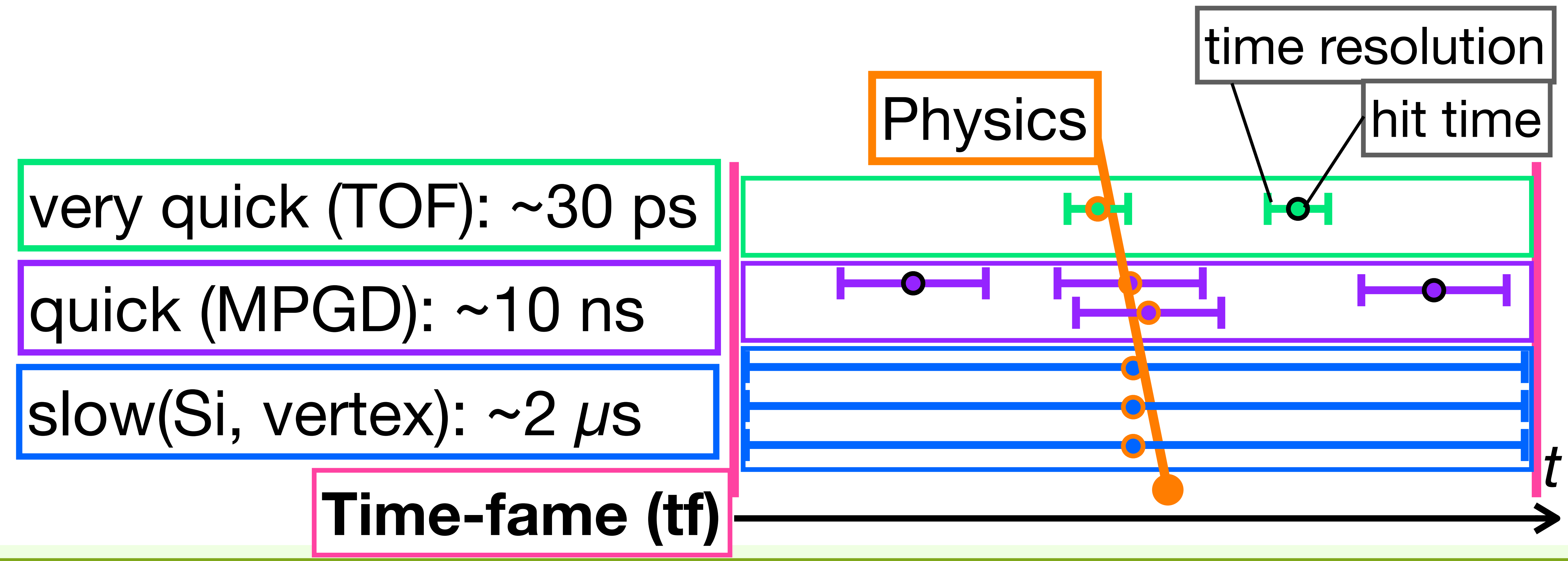
- Very high time resolution (~ 30 ps)
- One layer / There are uncovered region

MPGD

- High time resolution (~ 10 ns)
- Wide covered region
- Two layers

Silicon (MAPS)

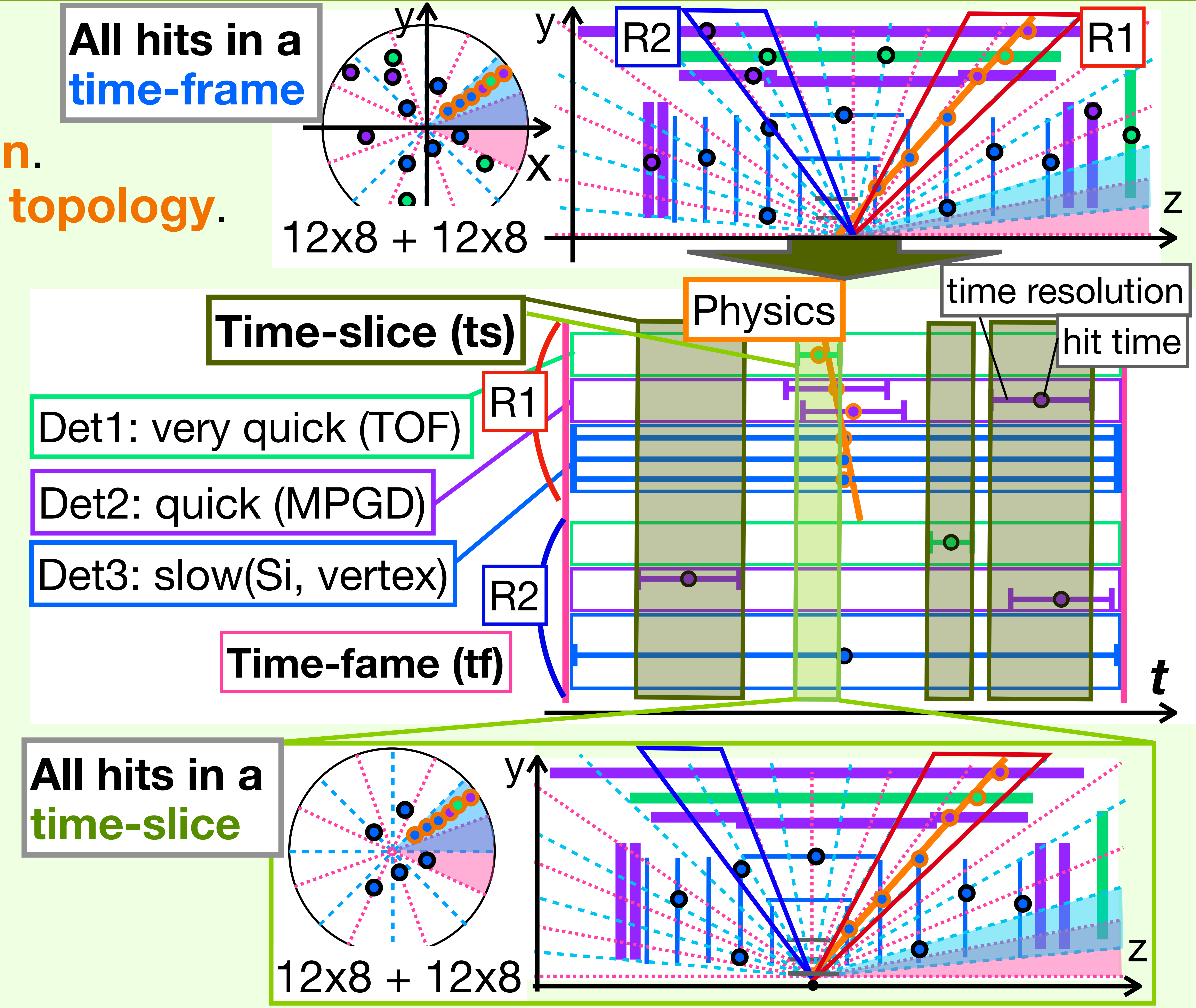
- 2 μs time window
- Wide covered region
- Multiple layers



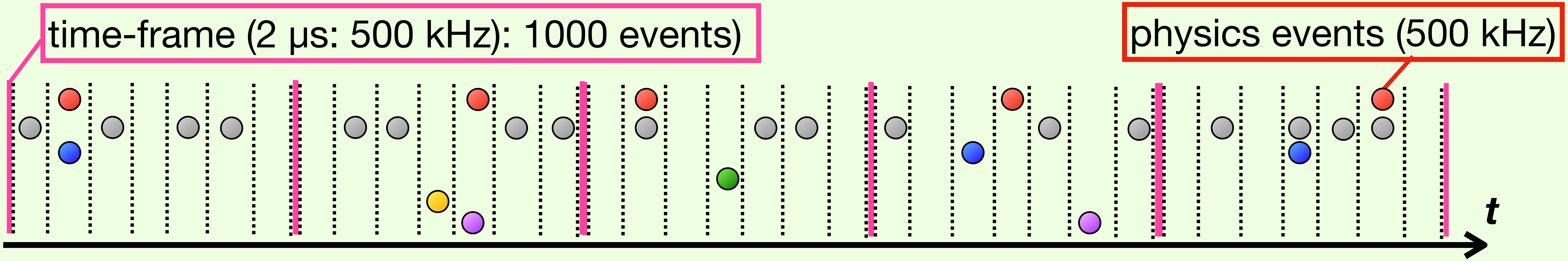
Pre-identification Algorithms

Key:
- **Hit driven** time slice
- Consider the detector's **time resolution**.
- Use coincidences in both **time** and **hit topology**.

1. Using the fastest detector (TOF) slices the time-frame.
2. Check the topology of the hits.
3. If there are more than any three hits in the same region, the time slice is accepted as event candidate.
4. Repeat steps 1–3 for all TOF hits.
5. If any unassociated MPGD hits remain, perform the same process using MPGD hits.
6. if there are no hits in TOF and MPGD in certain time slice, perform the same process based on silicon hits.



Simulation for Trigger Quality Evaluation



	DIS NC 18x275 Q^2 > 1 GeV^2 or 100 GeV^2 (Deep inelastic scattering neutral current)	500 kHz
	Synchrontron Radiation	14 MHz
	Electron bremsstrahlung radiation	317 kHz
	Electron Touscheck scattering (intrabeam dcattering)	1.3 kHz
	Electron Coulomb scattering processes	0.72 kHz
	Proton beam gas interactions	22.5 kHz

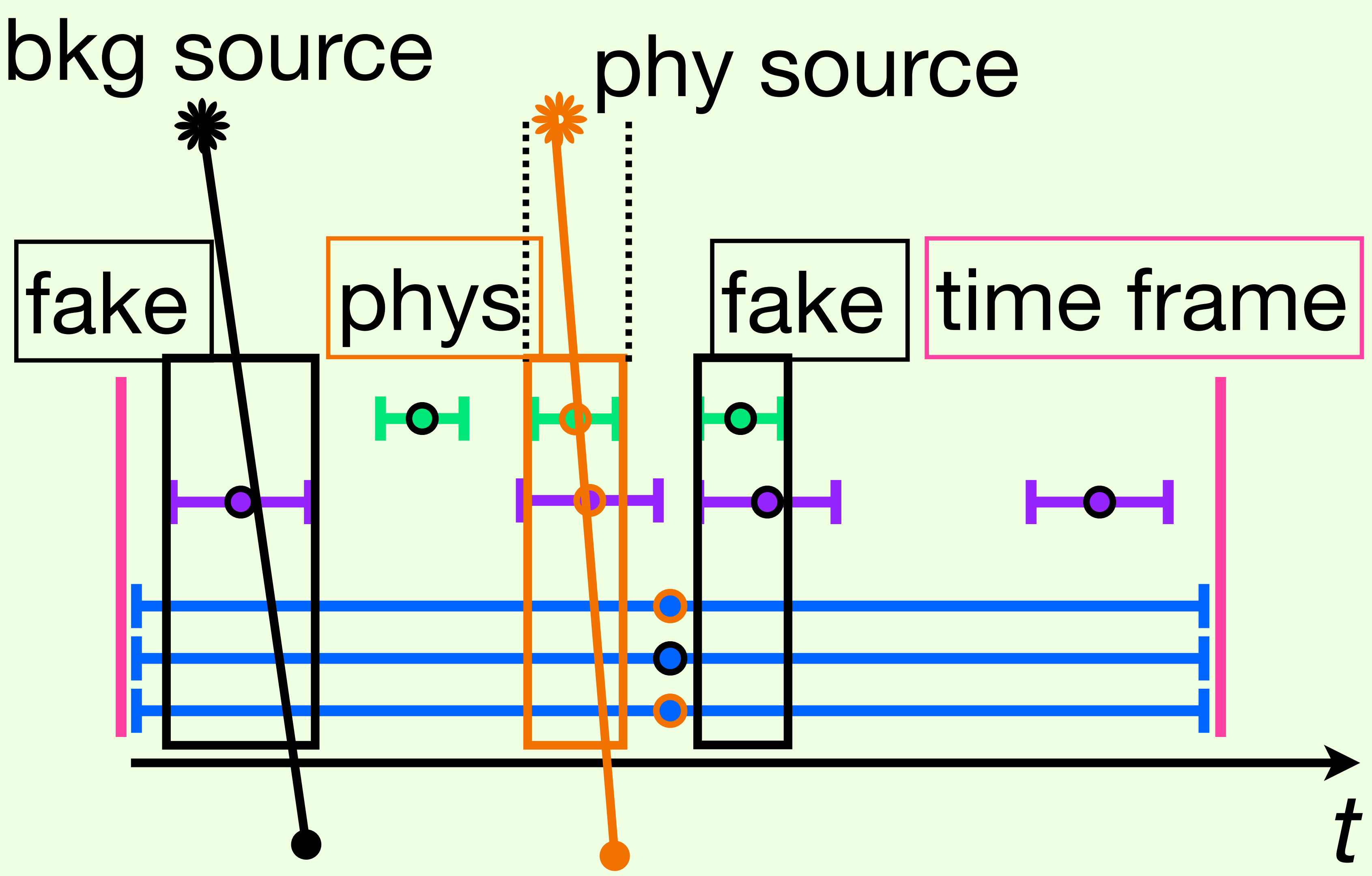
※ Very specific case to easy trigger (high Q^2).
In the real case, the high Q^2 events are lower than 500 kHz.

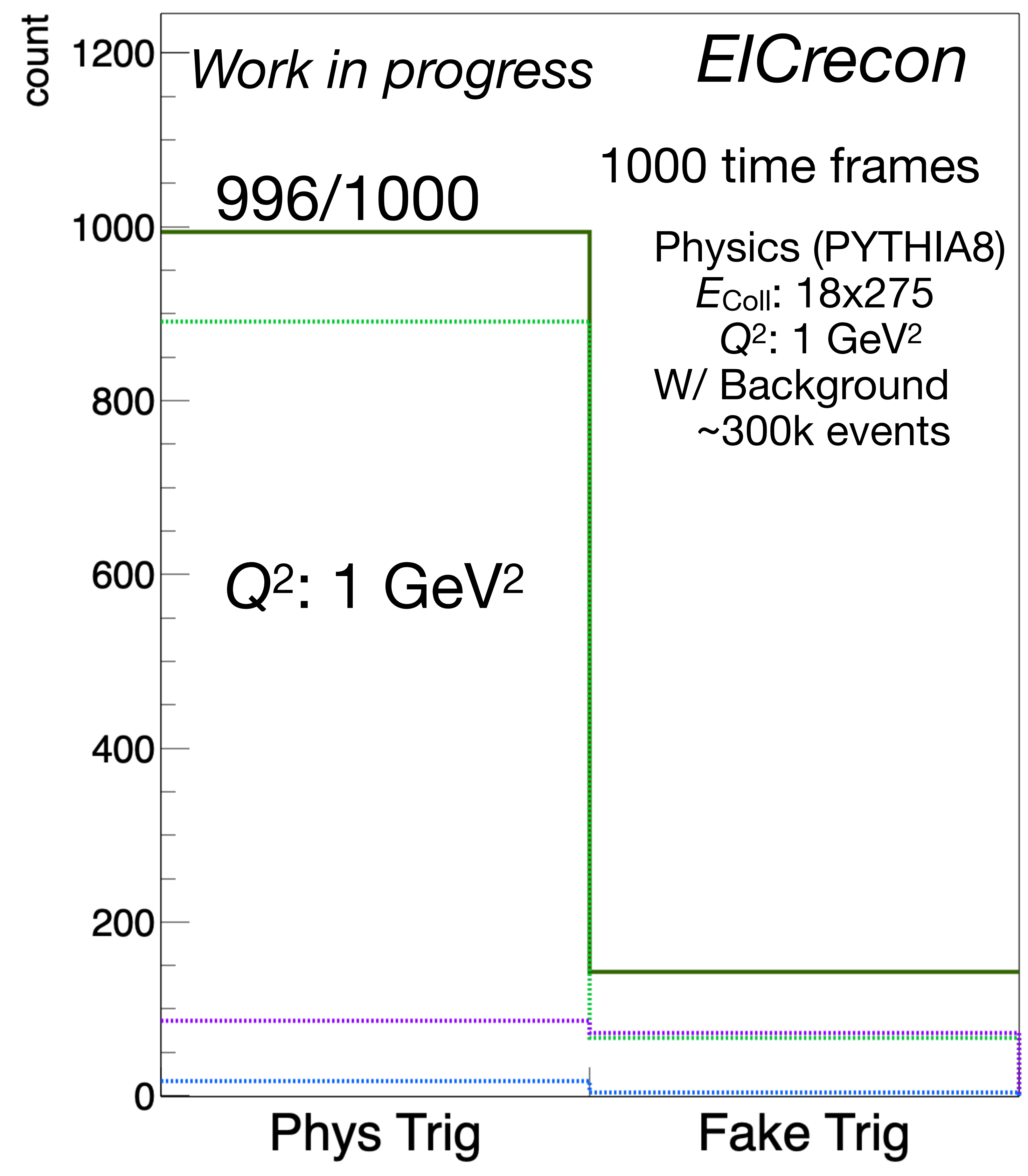
※ Recently we found the synchrotron radiation should be **14 MHz → 3.3 GHz**.

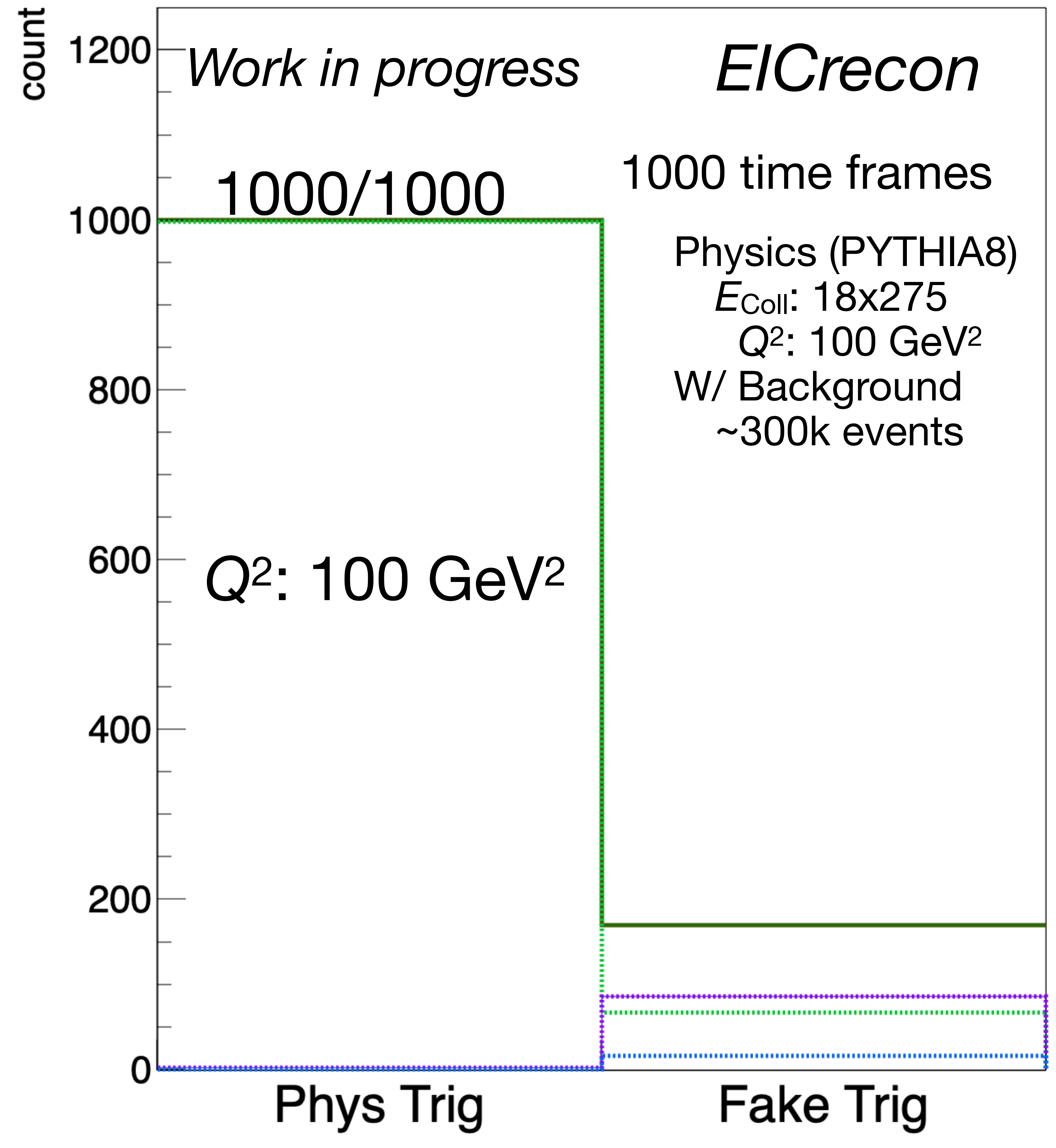
Evaluate event-identification efficiency in a **blind-analysis** setting, without prior knowledge of event positions in the continuous data stream.

Pre-identification Efficiency

Time slice containing the production time of a physics
→ ✔ **physics trigger**
→ ❌ **fake triggers**





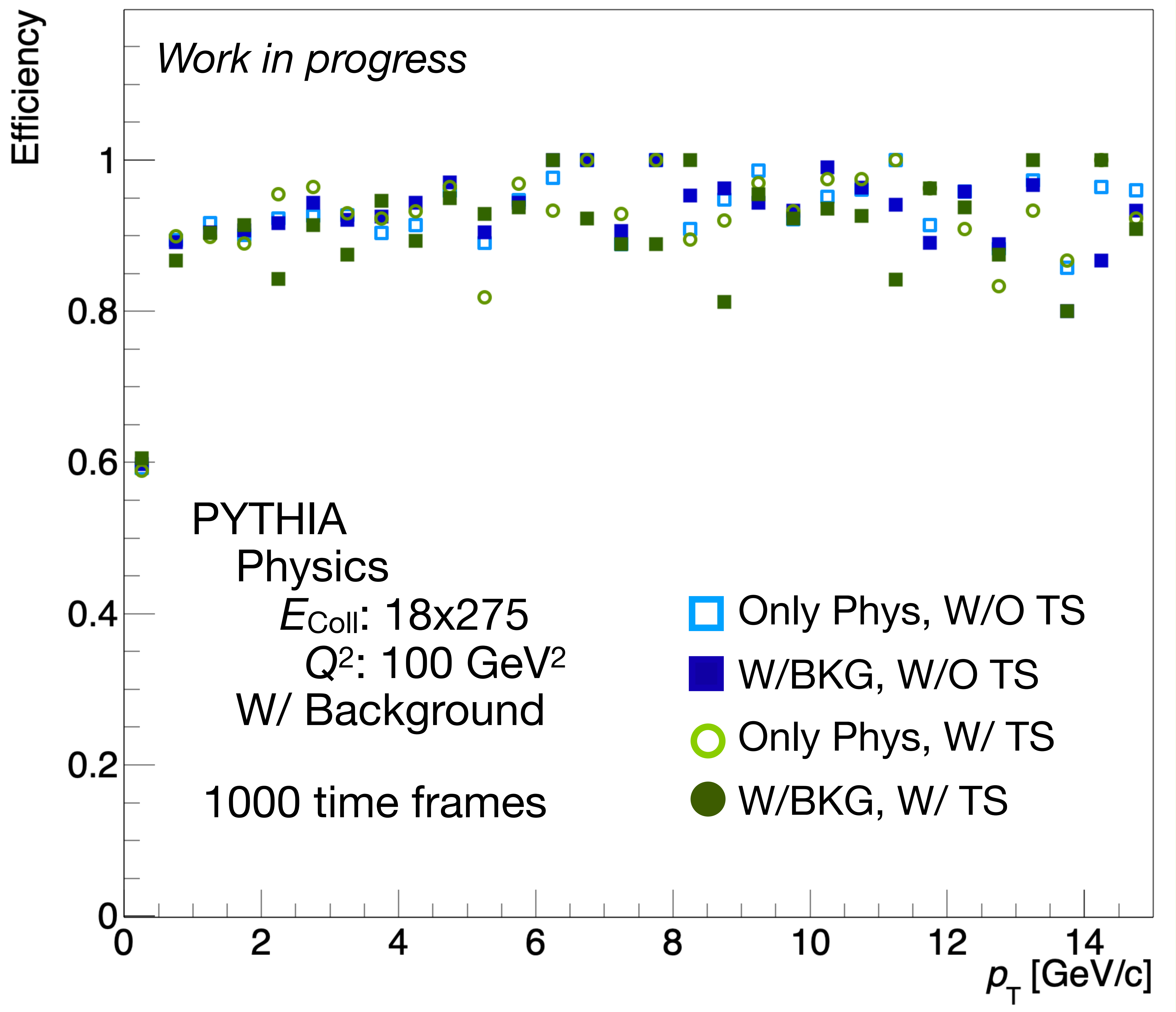


For DIS evens, the pre-identification efficiency achieved **over 99% for higher Q^2 > 1 GeV^2**

The background simulation is still old one, so it is not enough.
→ Need to test this with new background simulation.

Tracking performance in the triggered events

- The **efficiency** of W/ time slice were mostly consistent with W/O ones.
- The **purities** of W/ time slice W/O ones.
- **Track/vertex χ^2** and track p_T become a little better.
- **Process time** become faster using the time splitting (**-10%** / 1000 time frames (for large number of hits events: **-50%**))



(※ **This results are only for Q^2=100 GeV^2, and background is previous version (less)**)

Outlook

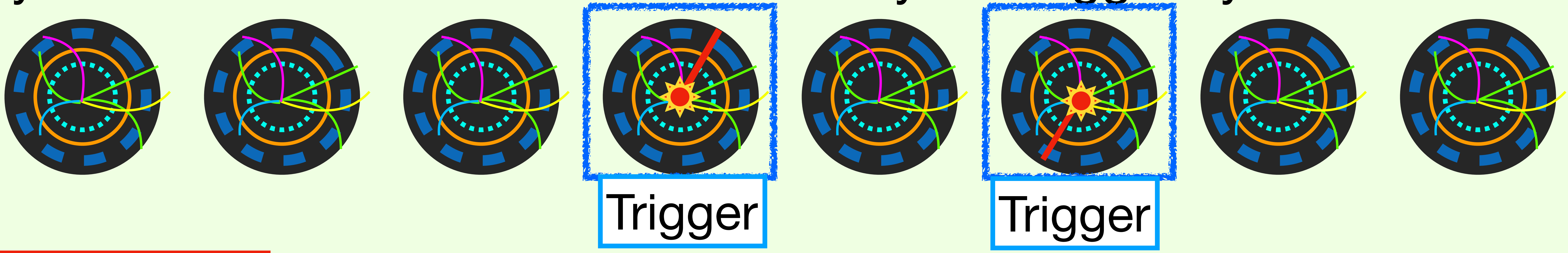
- Evaluate the pre-identification algorithm under higher background occupancy.
- Estimate the pre-identification efficiency for various physics processes.
- Integrate all detectors, including far-forward and far-backward systems.
- Apply multi-threading.
- Explore GPU acceleration.
- Develop a machine learning approach for event identification.
- Enhance detector digitization and include detector noise.

Backup Slides

Streaming Readout（SRO）

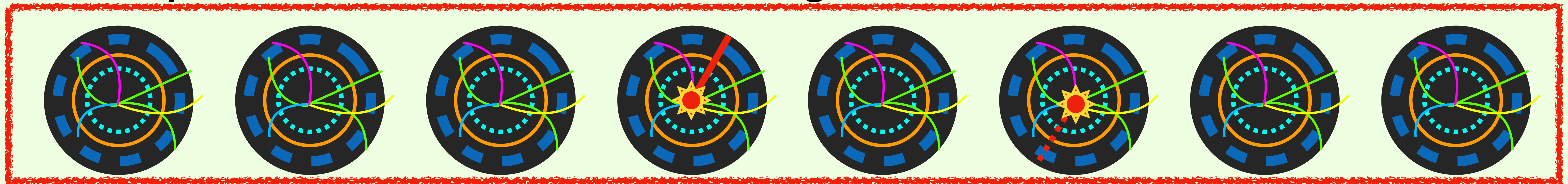
Trigger Readout

In high-energy physics experiments, the collision rate is extremely high, whereas the high-energy interactions of interest occur only rarely.
→ only characteristic events are selected by the trigger system and stored for analysis.



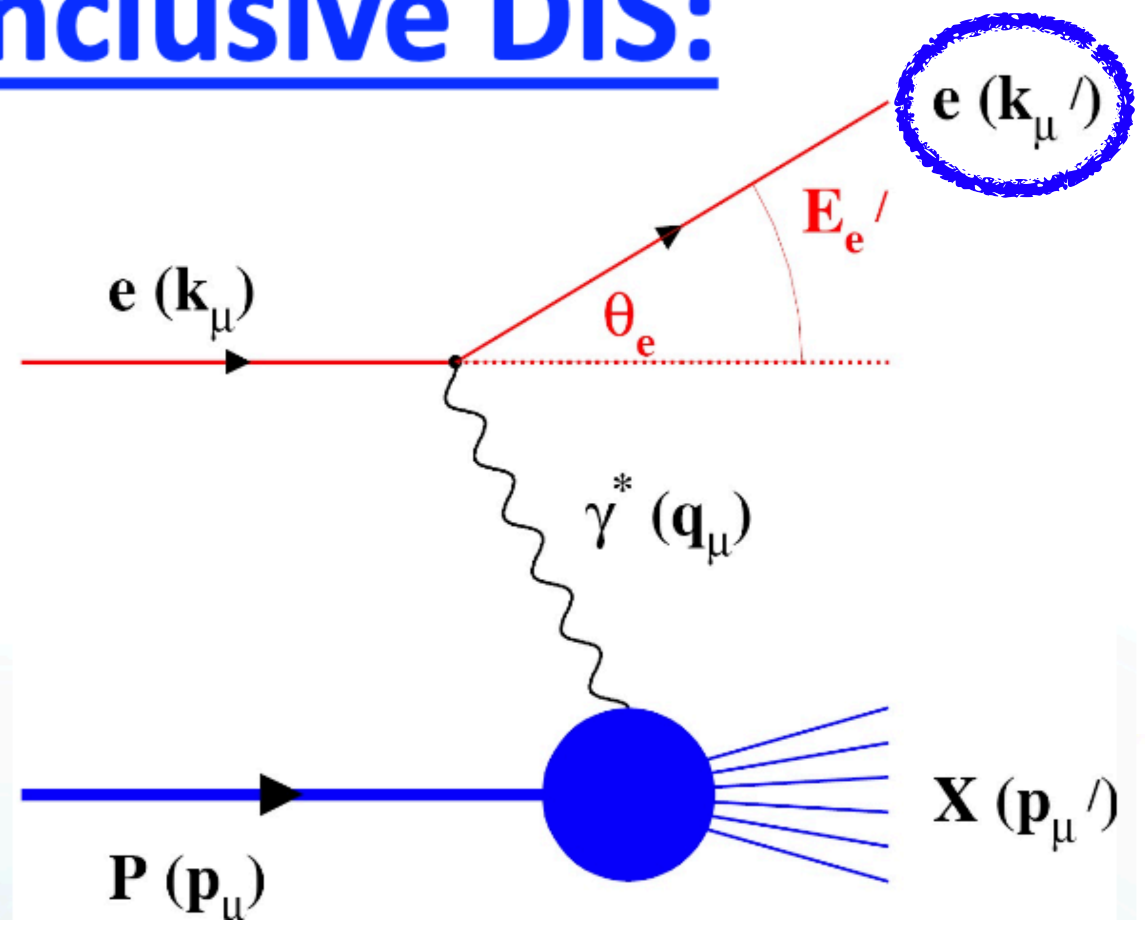
Streaming Readout

The target physics events of the EIC-ePIC experiment are difficult to distinguish from background.
→ Modern hardware and network technologies have enabled the establishment of the SRO.
(Already implemented in ALICE Run 3 and partially in RHIC-sPHENIX)
→ Data compression allows the recording of all events.



ePIC Detector Concept

inclusive DIS:

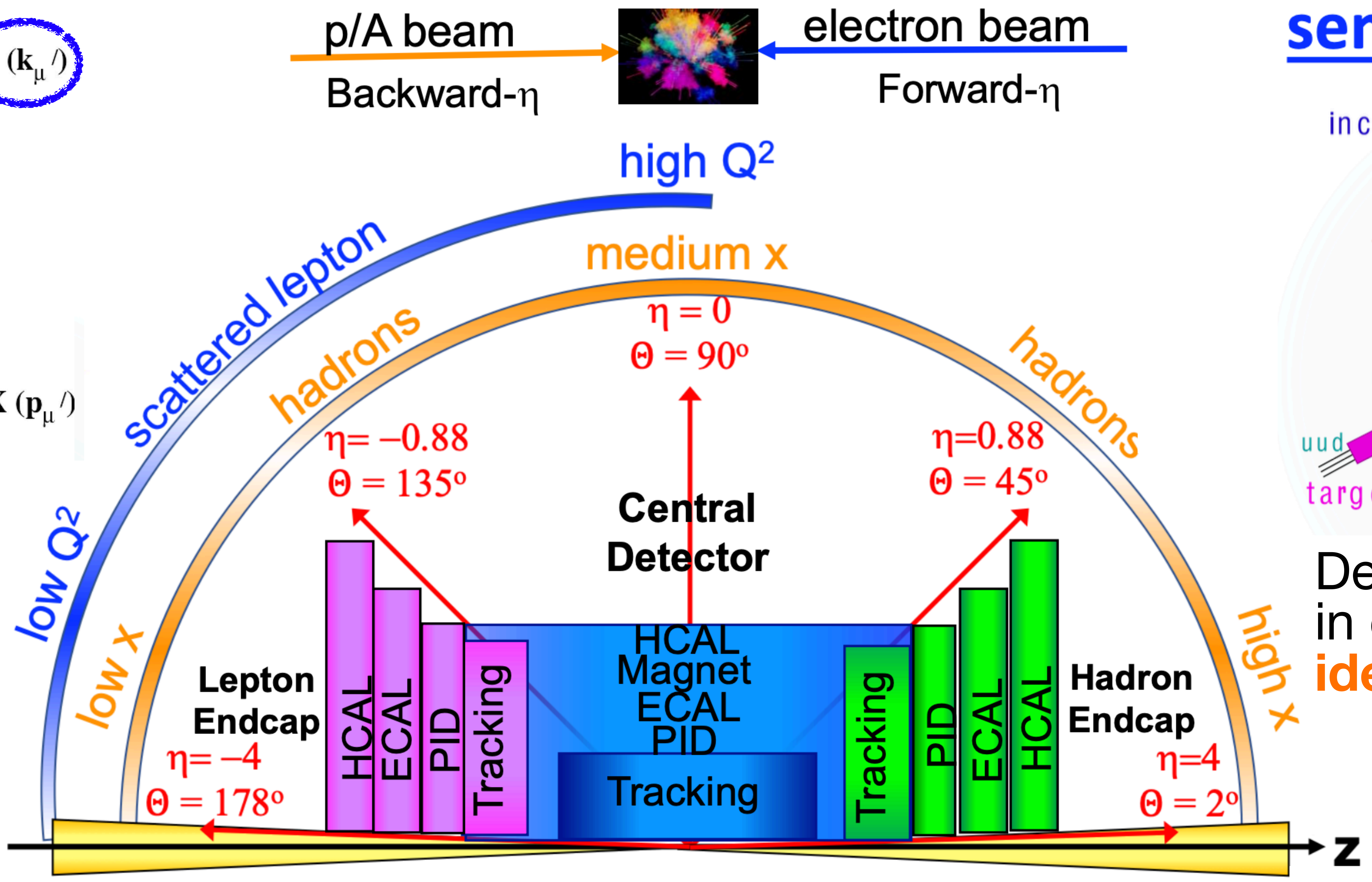


Detect only the **scattered lepton** in the detector

Very low Q^2 scattered lepton Bethe-Heitler photons for luminosity

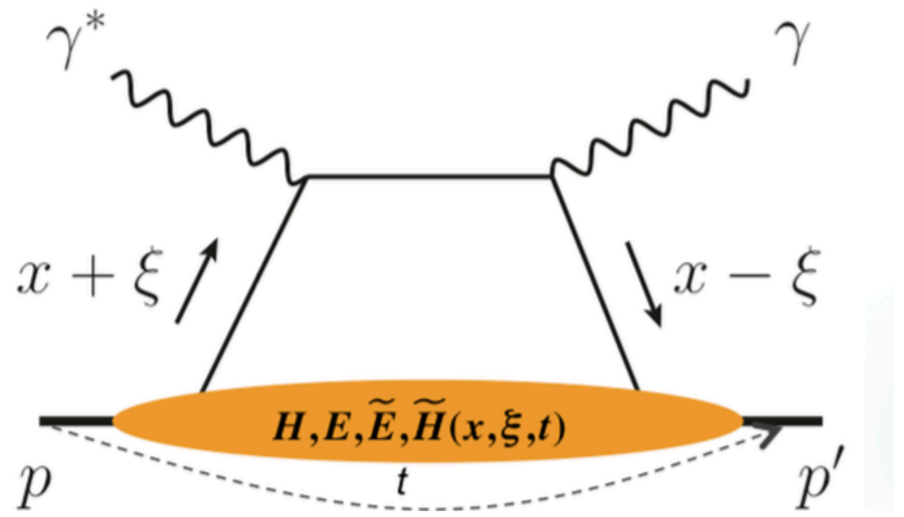
Luminosity Detector

Low Q^2-Tagger

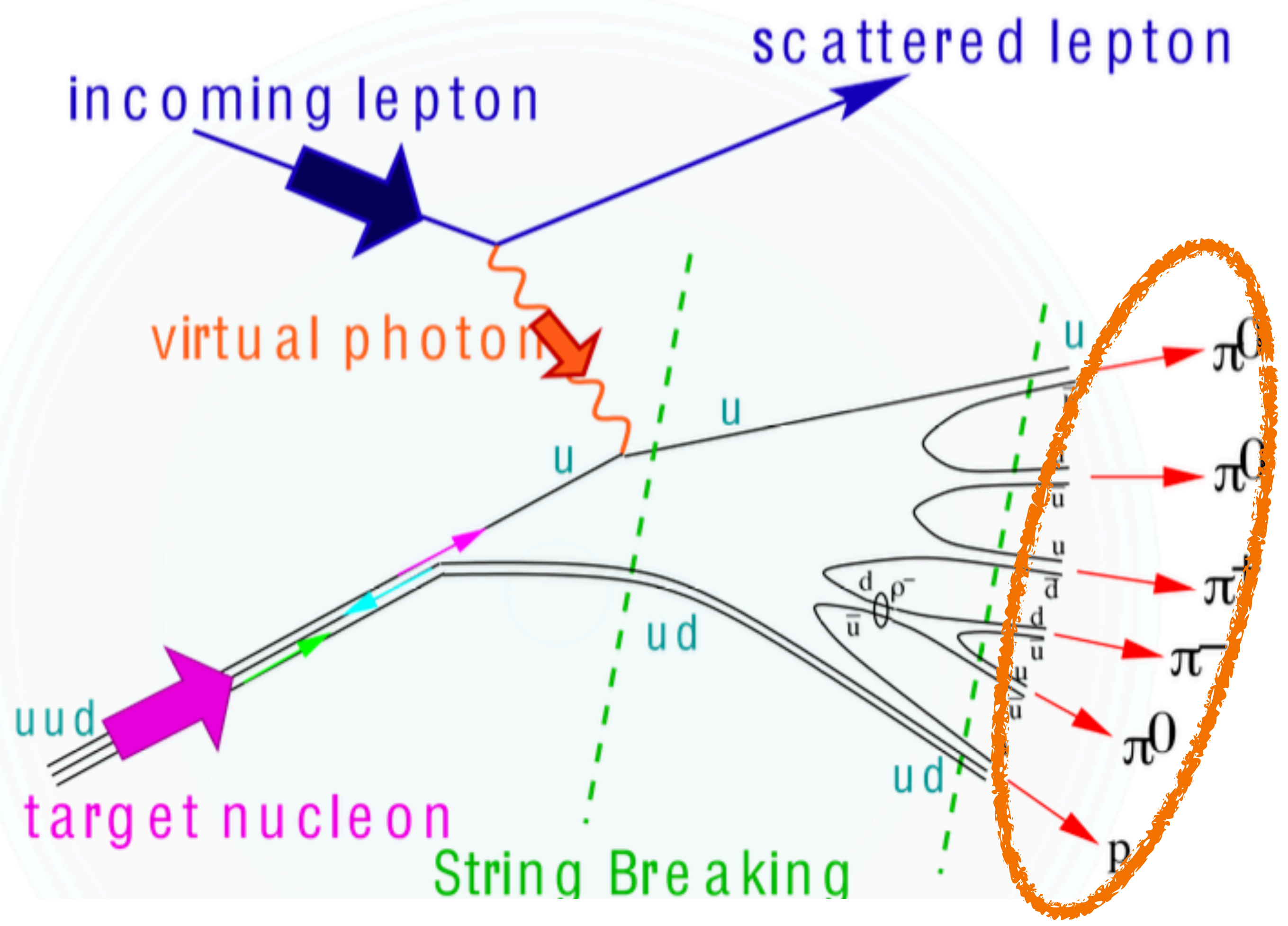


exclusive DIS

Detect the scattered lepton, identify produced hadron/ jets and **target remnants**



semi-inclusive DIS



Detect the scattered lepton in coincidence with **identified hadrons/jets**

Particles from nuclear breakup and from diffractive reactions

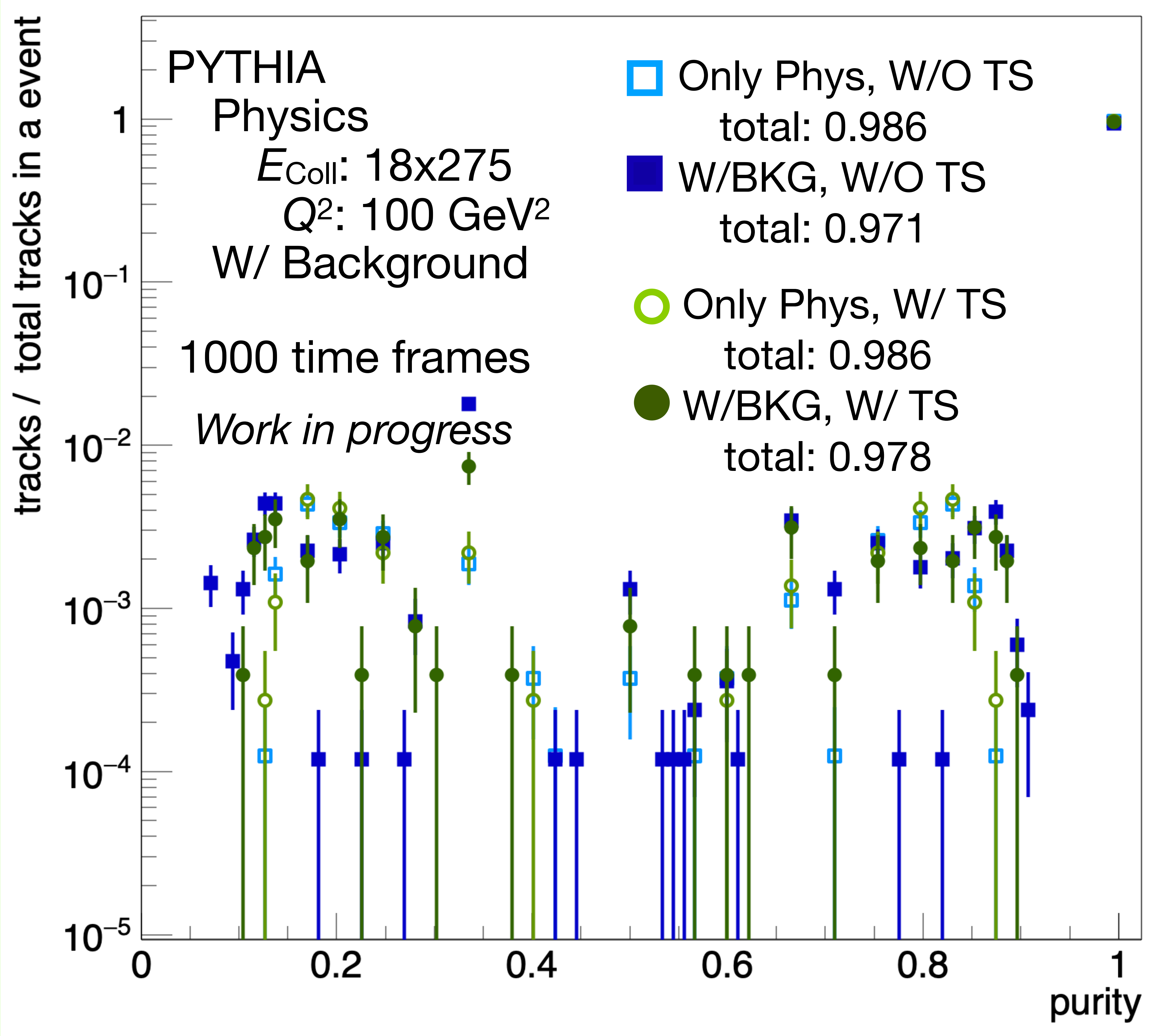
ZDC

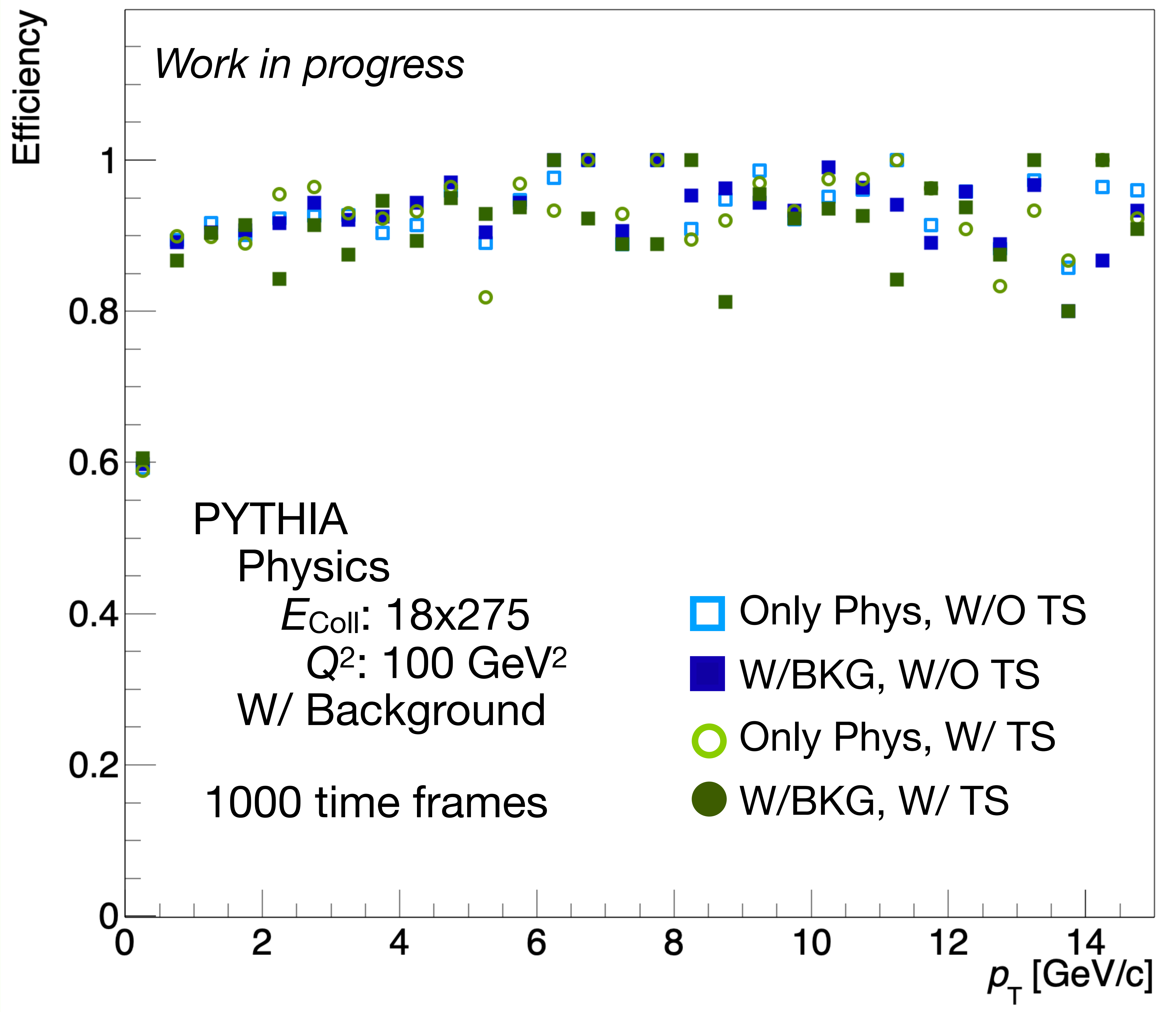
Forward Tracking

Time alignment of DIS events

Fit the mean hit time for each region.

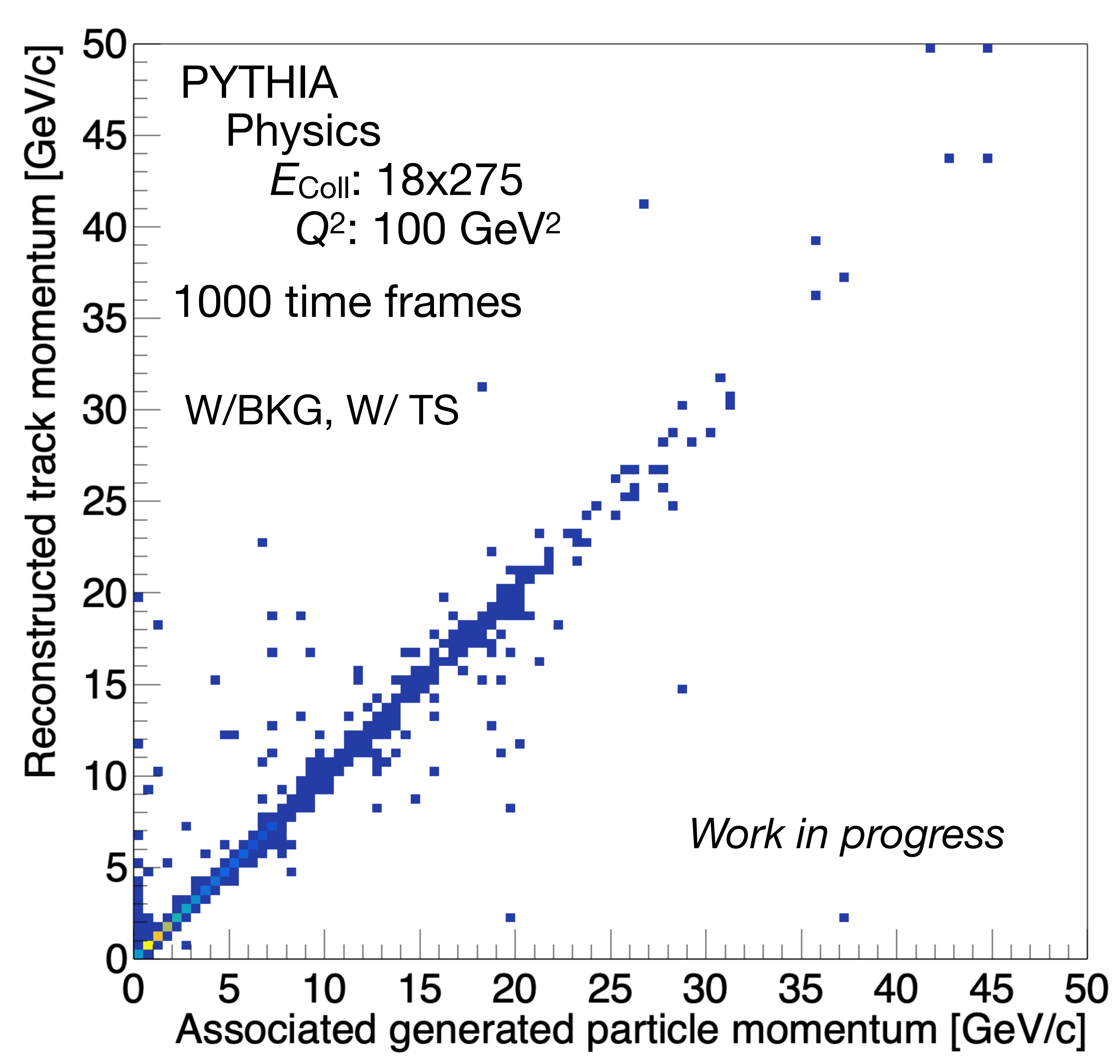
Track Purity and Efficiency Evaluation

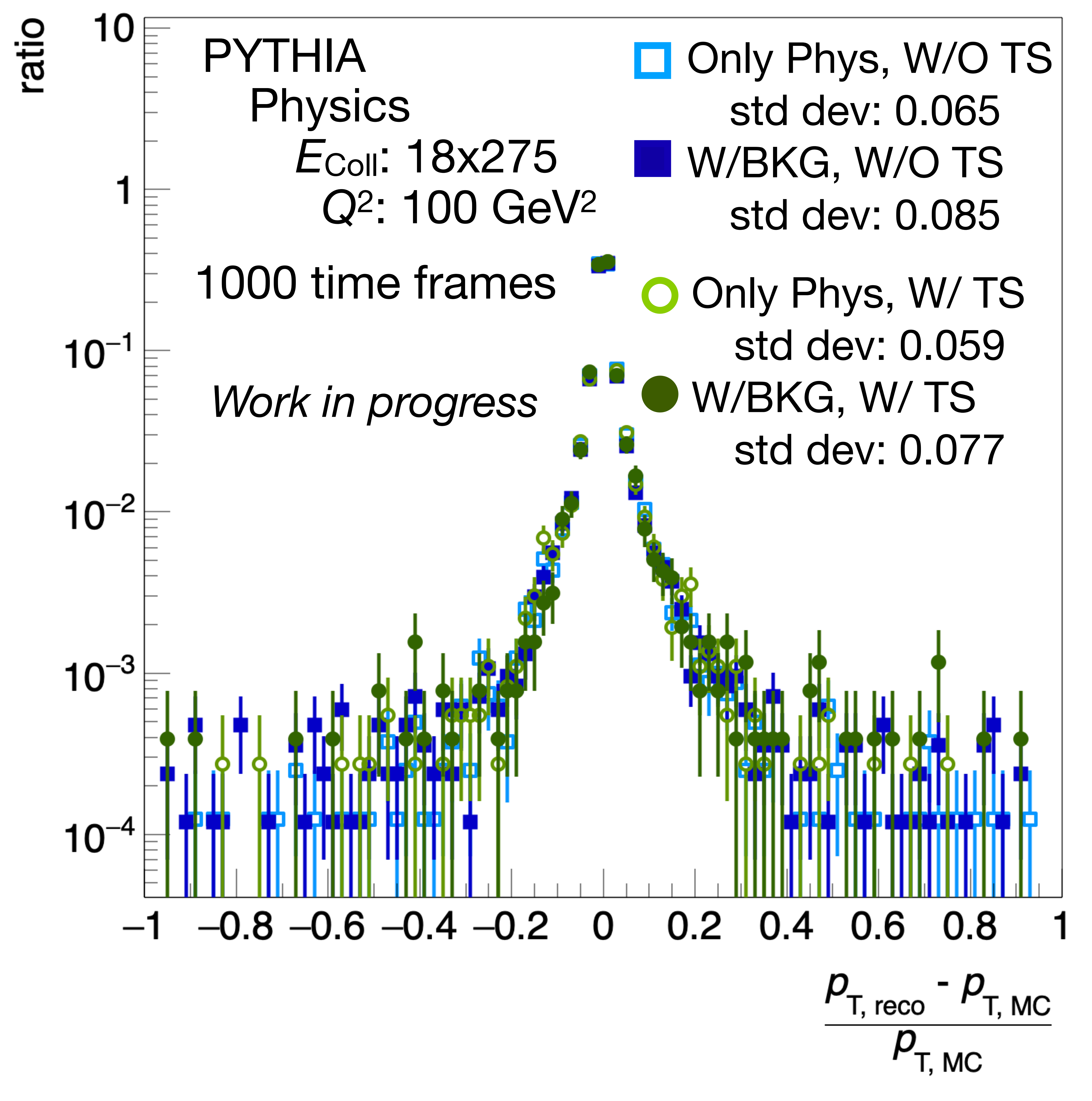




The purities of W/ time slice were a **little better** than W/O ones.
The efficiency of W/ time slice were mostly consistent with W/O ones, but a **little worse**.

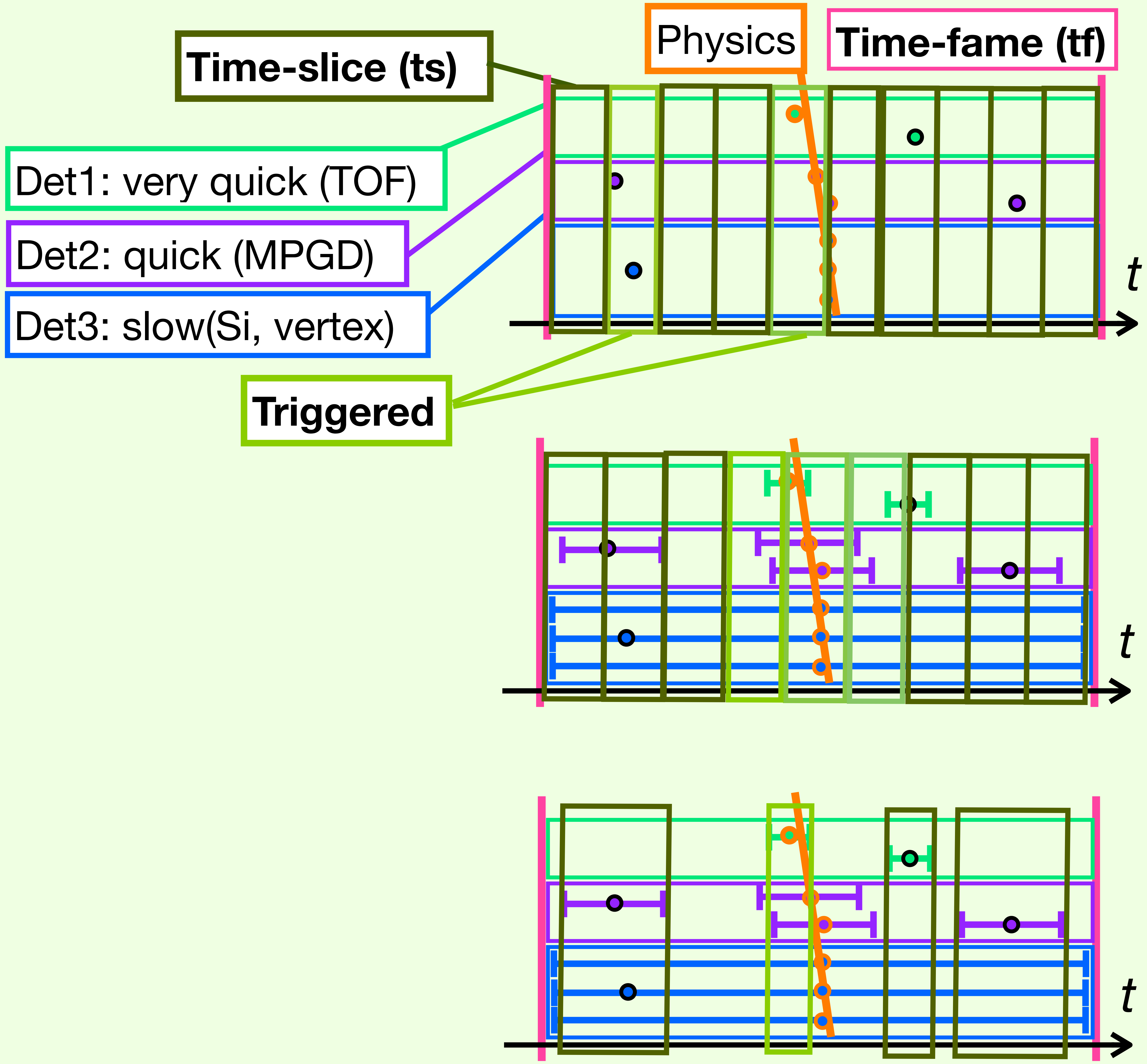
Track p_T Resolution





Good linearity between MC and reconstructed-track p_T.
The p_T resolution of the time splitting is a little better than W/O time splitting.

Time slice algorithms



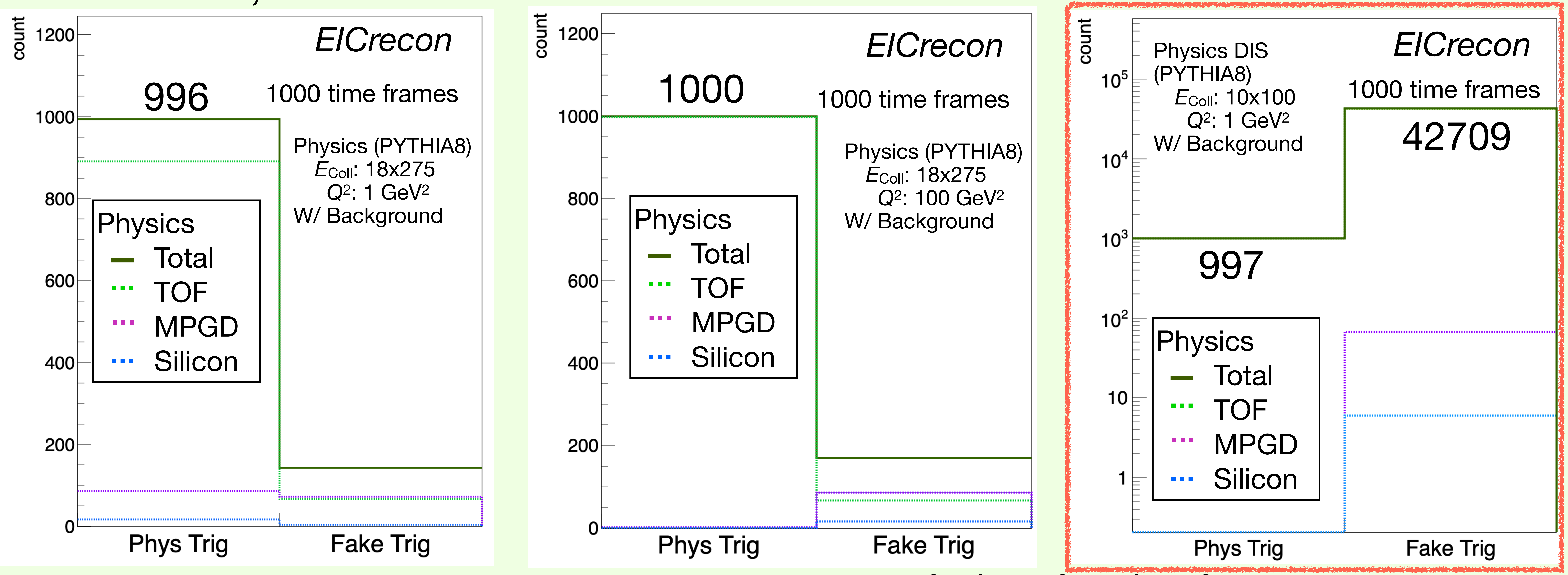
ver1: Slice a time frame by **constant time** slices
pros: very simple
cons: does not consider the detectors' time resolution.
does not considering hits' topology

ver2: Slice a time frame by **constant time** slices
pros: very simple
considering hits' topology
cons: there are many duplicated time slices including same hits and there are many fake times slices

ver3: Slice a time frame by **detector resolution base**
pros: realistic and sophisticated time slice.
considering hits' topology.
cons: a little complicated.

Pre-identification Efficiency

- There were problem to count the triggered physics event.
→ Fixed them, but there are still some concerns.

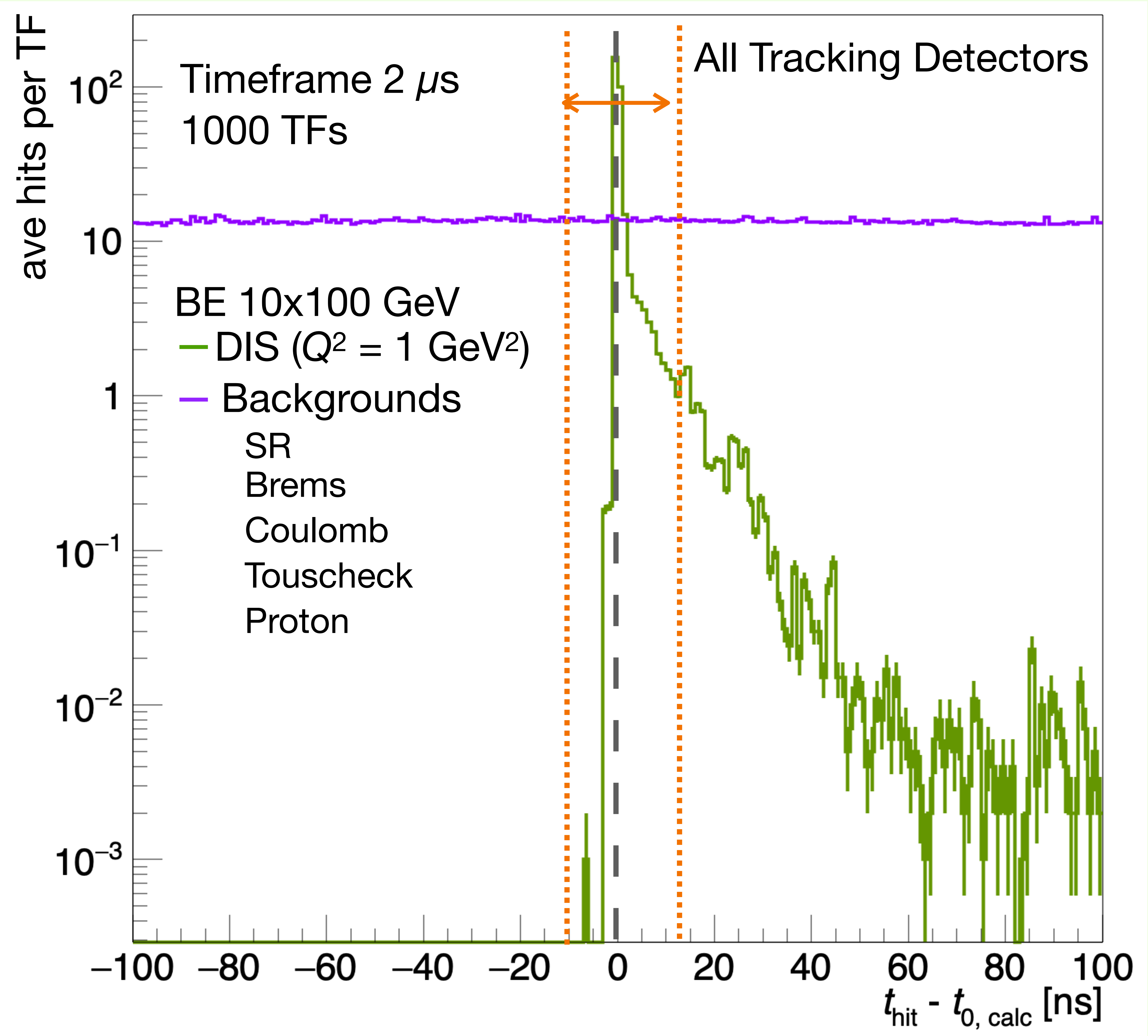


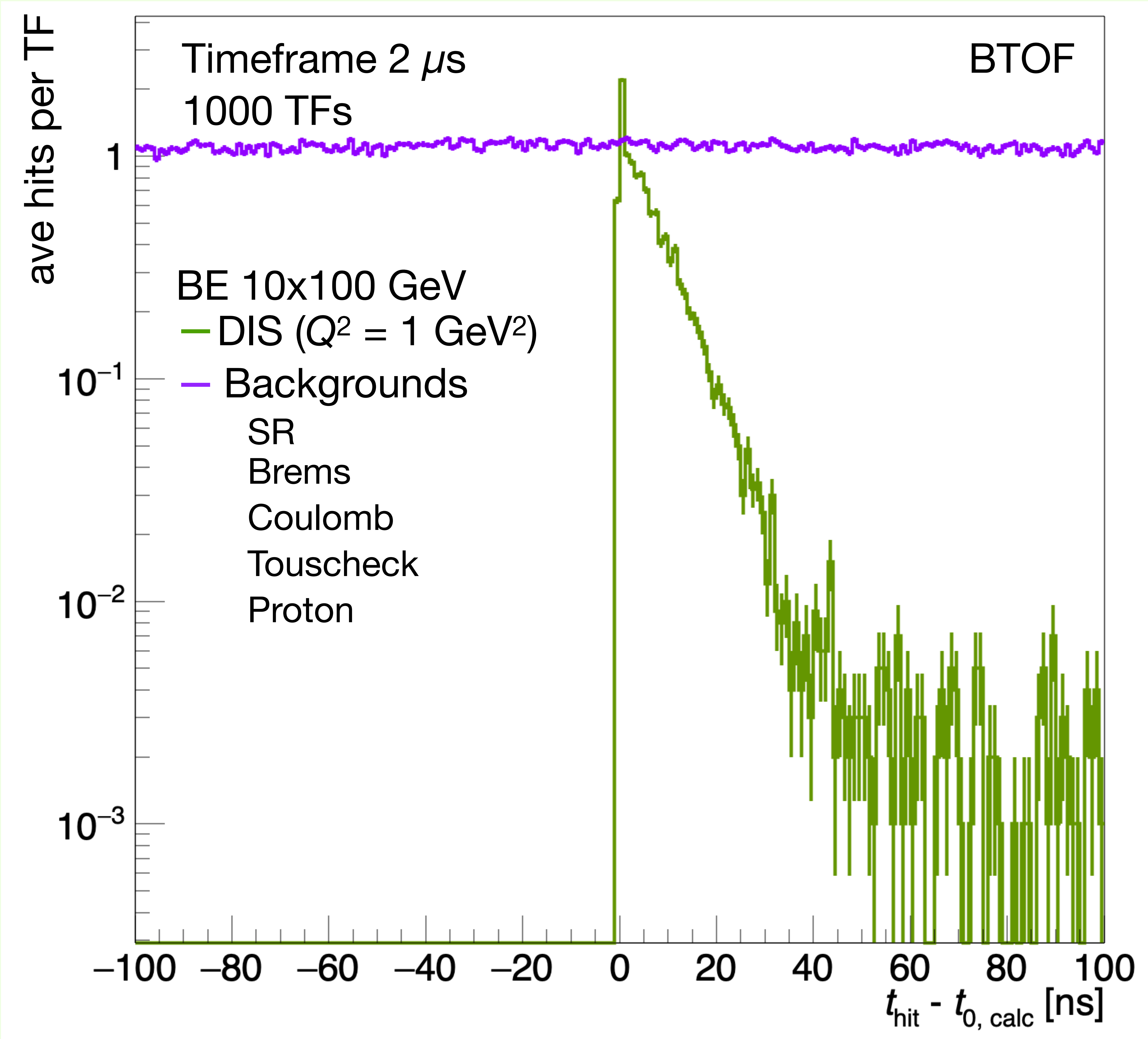
- Found the pre-identification can triggered most low Q^2 (> 1 GeV^2) DIS events.
- In the new bkg simulation, large number of fake triggers were created.
→ I am checking the details now.

Outlook map

5. Test online chain (E0 - E1 orchestraton::connect detector/rucio) (- Sep?)
1. Test for various physics events (low Q^2 / 10x100 beam energy) (- April)
Time frame
Timeframe Lv
Read
Digitization
Unfold
4. Apply multi-threads (or GPU) (- Aug)
Physics event
Background/ Noise
6. Apply detector response (???)
Clustering
ACTS
(track and vertex reconstruction)
Event Lv
Write
2. Up date the pre-identification using all sub-detectors (- April)
3. Implement "Event identification": Stringent event identification and event classificaiton using reconstructed tracks and vertex information. (- Jun)
7. Add calibration stream line (???)

Number of Hits in a TF

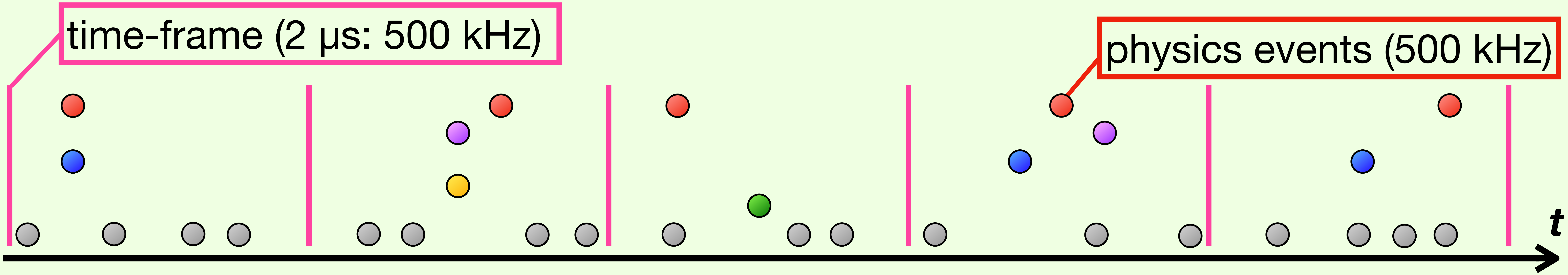




95% DIS hits: 10 ns, # of DIS hits 298, # of BKG hits 297

99% DIS hits: 30 ns, # of DIS hits 310, # of BKG hits 310

Update of new simulation



rates in kHz	Old	New	Vacuum
Total ep	500 kHz	500 kHz	
hadron beam gas	22.5 kHz	31.9 kHz	10000Ahr
		342.8 kHz	100Ahr
e beam gas (Bremsstrahlung scattering)	317 kHz	3177.3 kHz	10000Ahr
e beam gas (Coulomb losses - w/ collimators)	0.72 kHz	29.56 kHz	10000Ahr
e intrabeam (Touschek losses - w/ collimators)	1.3 kHz	233.5 kHz	
electron SR	14 MHz	36.6 GHz	

Time resolution of each sub-detector

Tracking detector

- Silicon MAPS: ~2 μs[ref, p11]
- MPGD: ~10 ns [ref, p11]
- TOF: ~30 ps[ref, p11]

PID

- hpDIRC: ~50 ps [ref, sec2]
- dRICH: ~20 ps [ref, p2]
- pfRICH / HRPPD / LAPPD: ~50 ps [ref, sec2]

Calorimeter

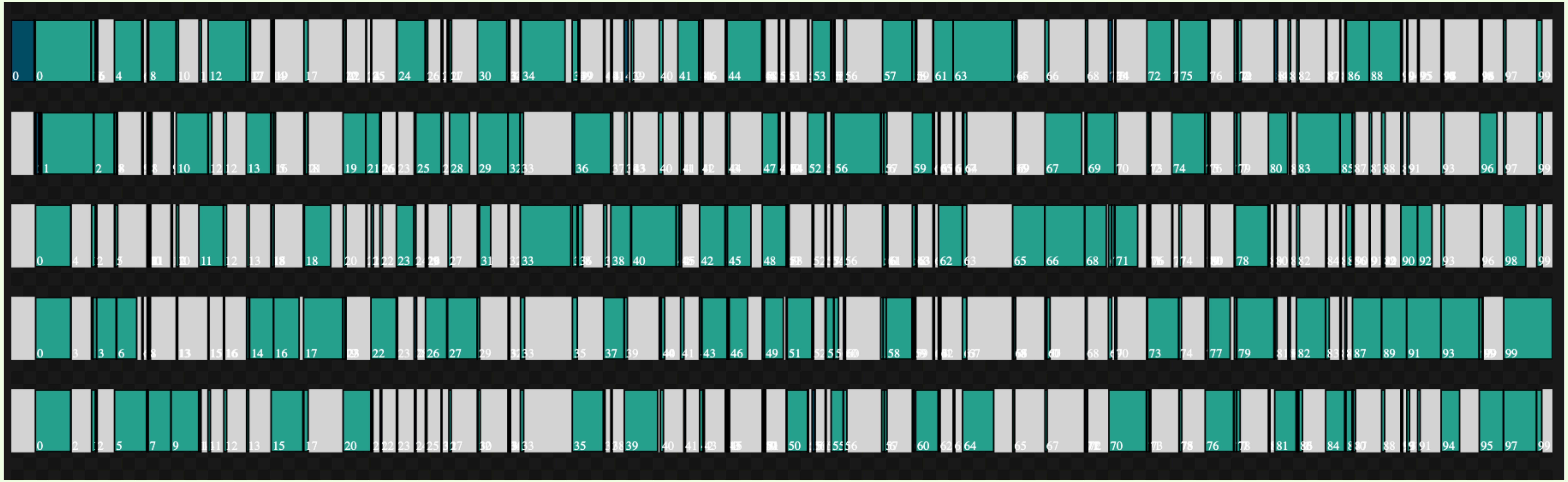
- Barrel Imaging Calorimeter: ~ 3.25 ns [ref]
- Backward/Forward EMCal、Barrel/Forward/Backward HCal: ?? s
- Zero Degree Calorimeter: ~30ps [ref]

Far-forward（Roman Pots / Off-Momentum Detectors / B0)30 30 ps [ref]

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.