Streaming Orchestration Use Case: Event Identification from Streaming Data

The University of Tokyo, Quark Nuclear Science Institution
Takuya Kumaoka

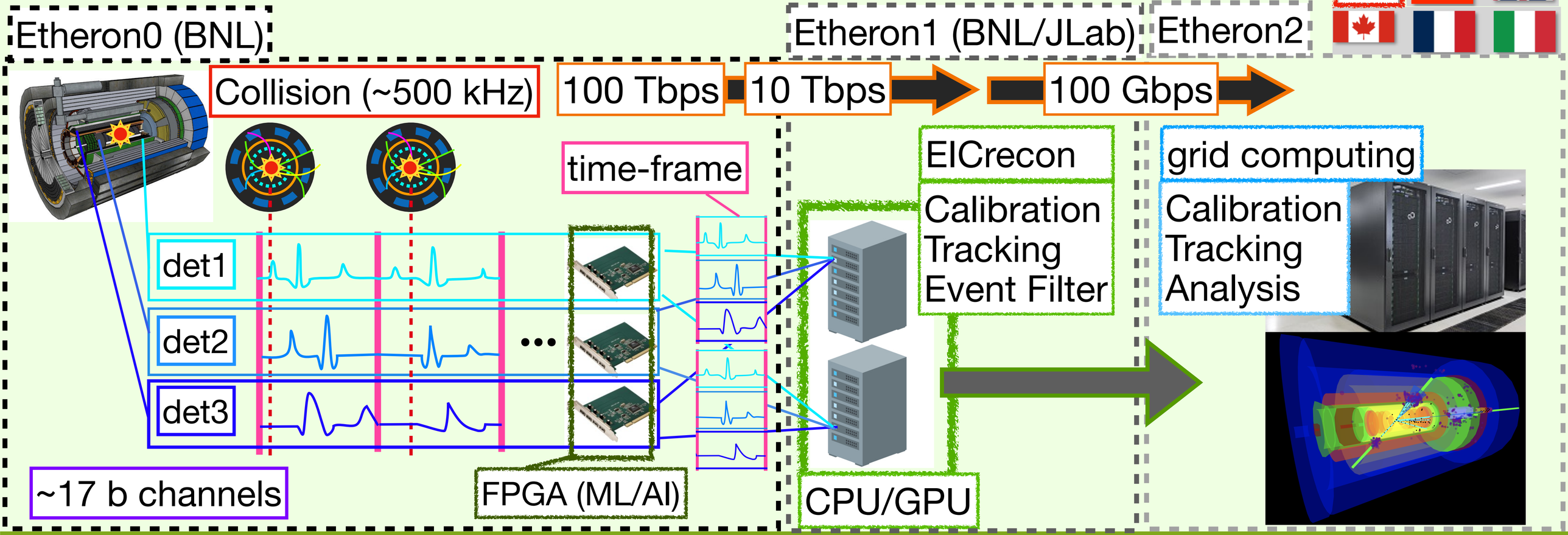
EIC-ePIC DAQ Software & Computing

Acquire zero suppression physics 500 kHz

Dominant Background (Synchrotron radiation): 14 MHz

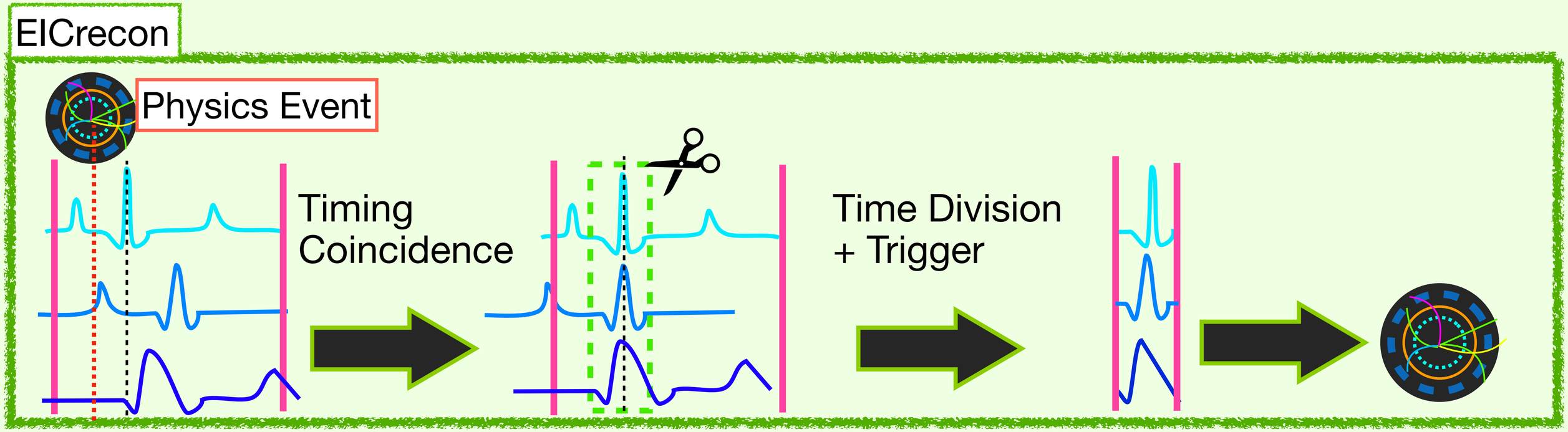
→ Data Storage: ~350 PB/year

→ **Essential to reduce data volume and high-speed processing**
(100 Tbps → 100 Gbps)



This study aim

Implement a module in EICrecon to extract only physics signals within a time frame



- There is surplus region of noises and background in a time-frame
- Extructing only physics events enalbe to reduce data volume

→ **Aim to develop a module capable of extracting all physics events,**
while meeting the requirements for data transfer rate and processing speed.

JANA2 Data Flow

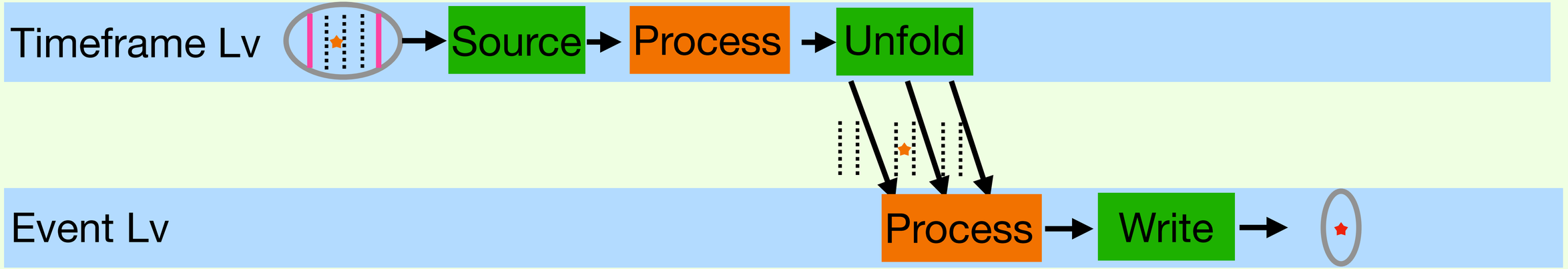
Basic Topology

sequential parallel

If the input data contains only one physics event, a single-level process is sufficient.

Timeslice Splitting Topology

※ Unfold: Split one frame into multi frames

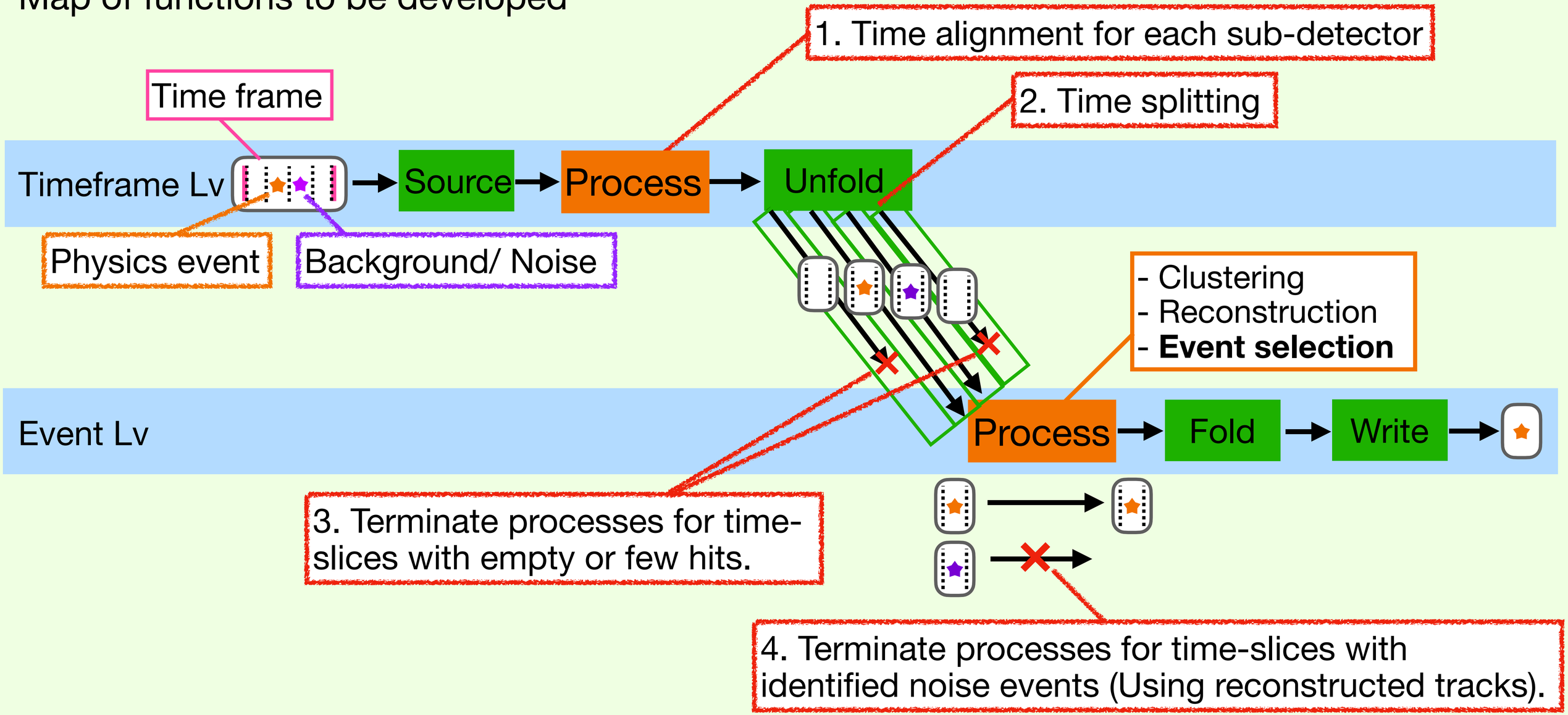


A split time frame within a time slice needs to be processed.
→ It should be handled in parallel at the event level.

sequential parallel

My Study Targets

Map of functions to be developed

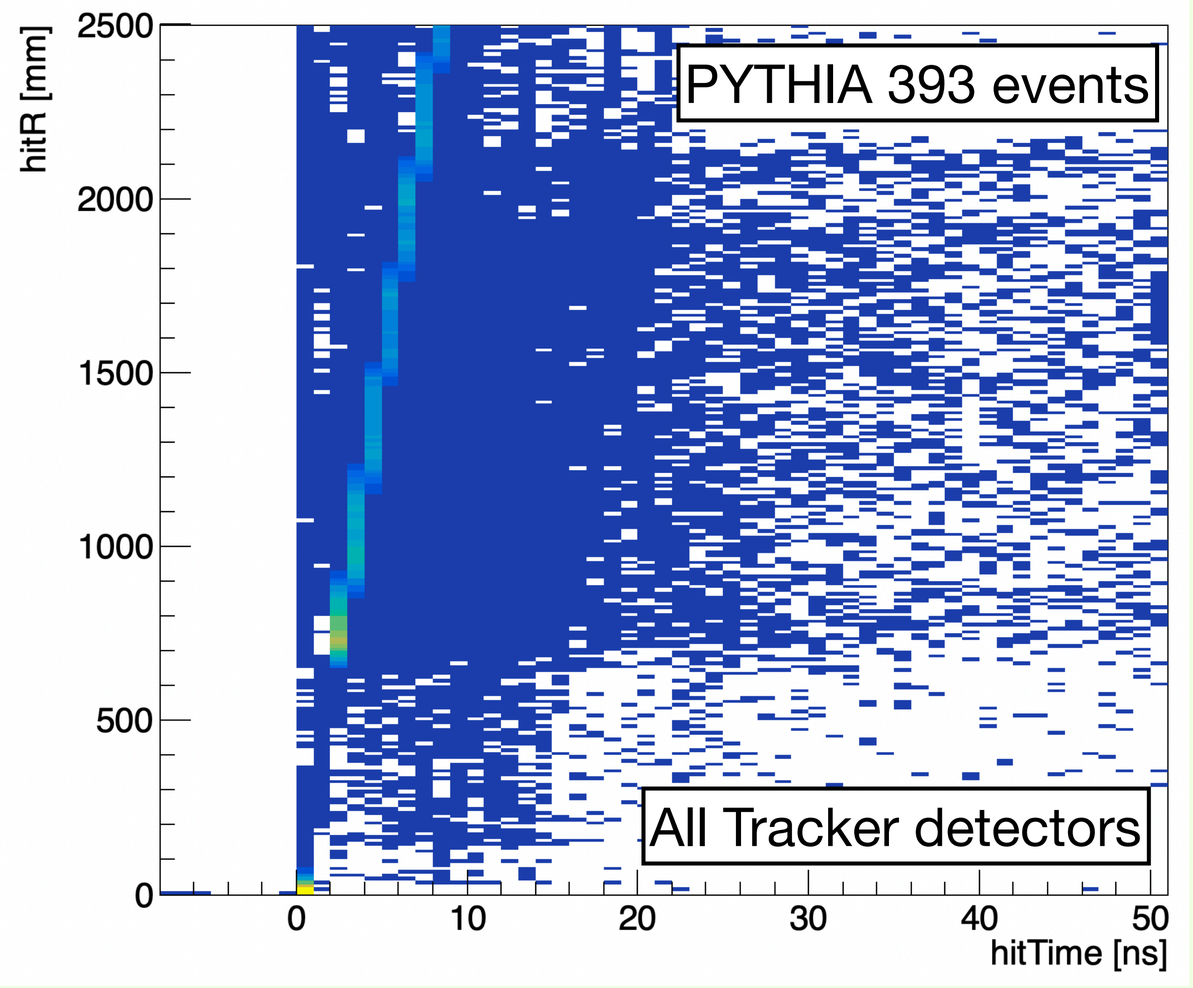


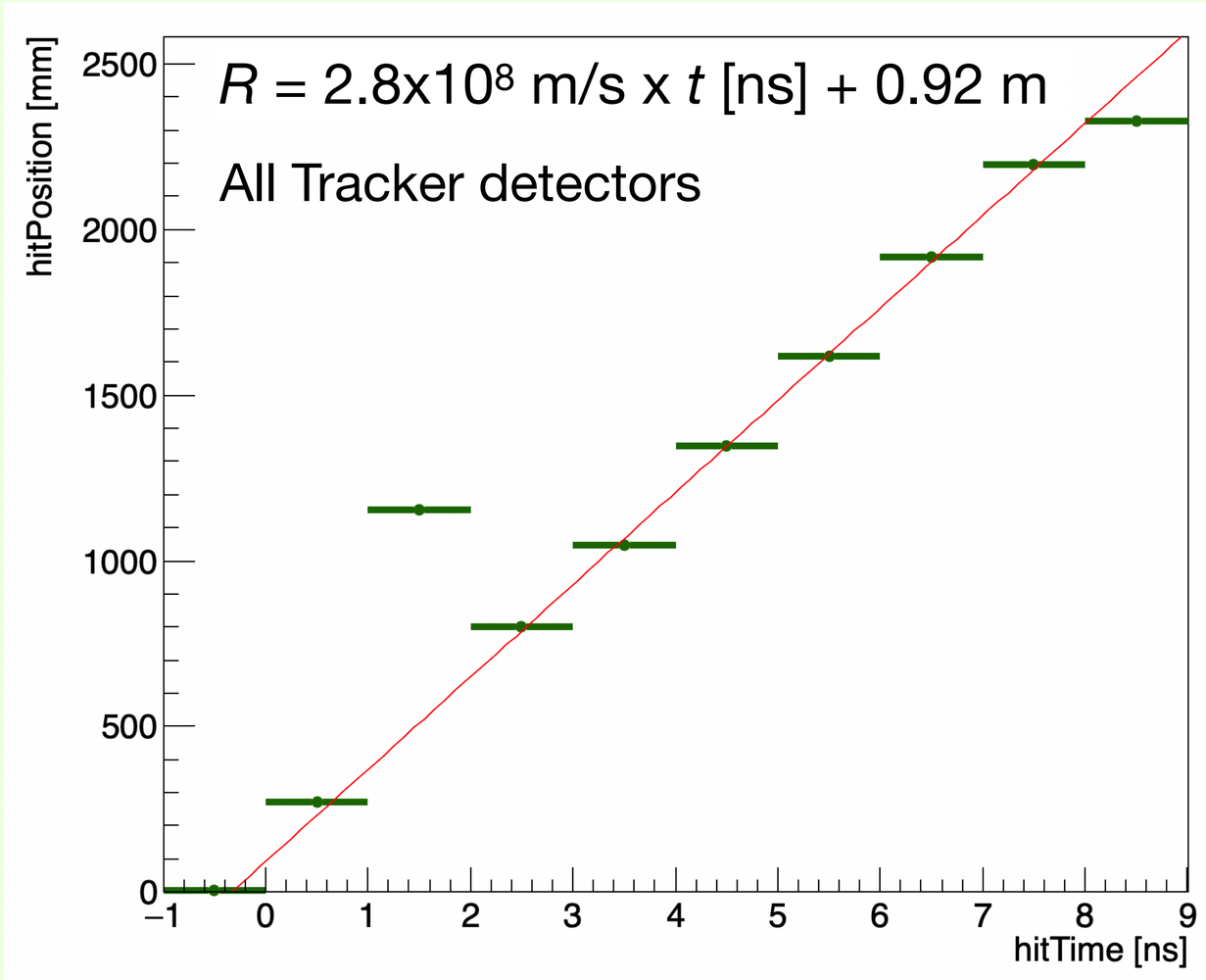
Progress of Study

1. Create a new factory to time alignment (**Done**)
2. Unfolding trivial test (**Done**)
3. Time splitting test (**Done**)
4. Timing Coincidence (**Done**)
5. Simulation with Background (**On going: - Octorber**)
6. Evaluate effects for tracking (October?)
6. Apply detector response (October-November?)
7. GPU usage (December?)

Hit time vs Hit position (Time Alignment)

These plots show the relation between hit time vs hit position to alignment hit time.





→A correlation between distance and hit time is observed.
However, since we are looking at MC hits, this result is expected (W/O detector response).

Time Alignment

Create Factory to do time alignment

Original input data

```
HitChecker: Event 188640 Hits in: 13.1356, 1.12963, 1.2371, 1.03809, 1.02581, 24.2638, 10.5798,
9.64017, 1.77883, 1.93878, 1.63005, 1.61152
```

Sort + time alignment ($t' = t$ - 0.003 [ns/mm] * R [mm])

```
HitChecker: Event 188640 Hits in: 0.863942, 0.867679, 0.889219, 0.943039, 1.06582, 1.37656,
1.41053, 1.49302, 1.5067, 1.66467, 1.68391, 1.76771
```

It seems work well.

→ **Future plan: Time response of detectors will be included.**

Create Unfold Code

JEventUnfolder (JANA2) Nathan's work:
Base of unfold code. It handles to IDs of a time frame (parent) and events (child).

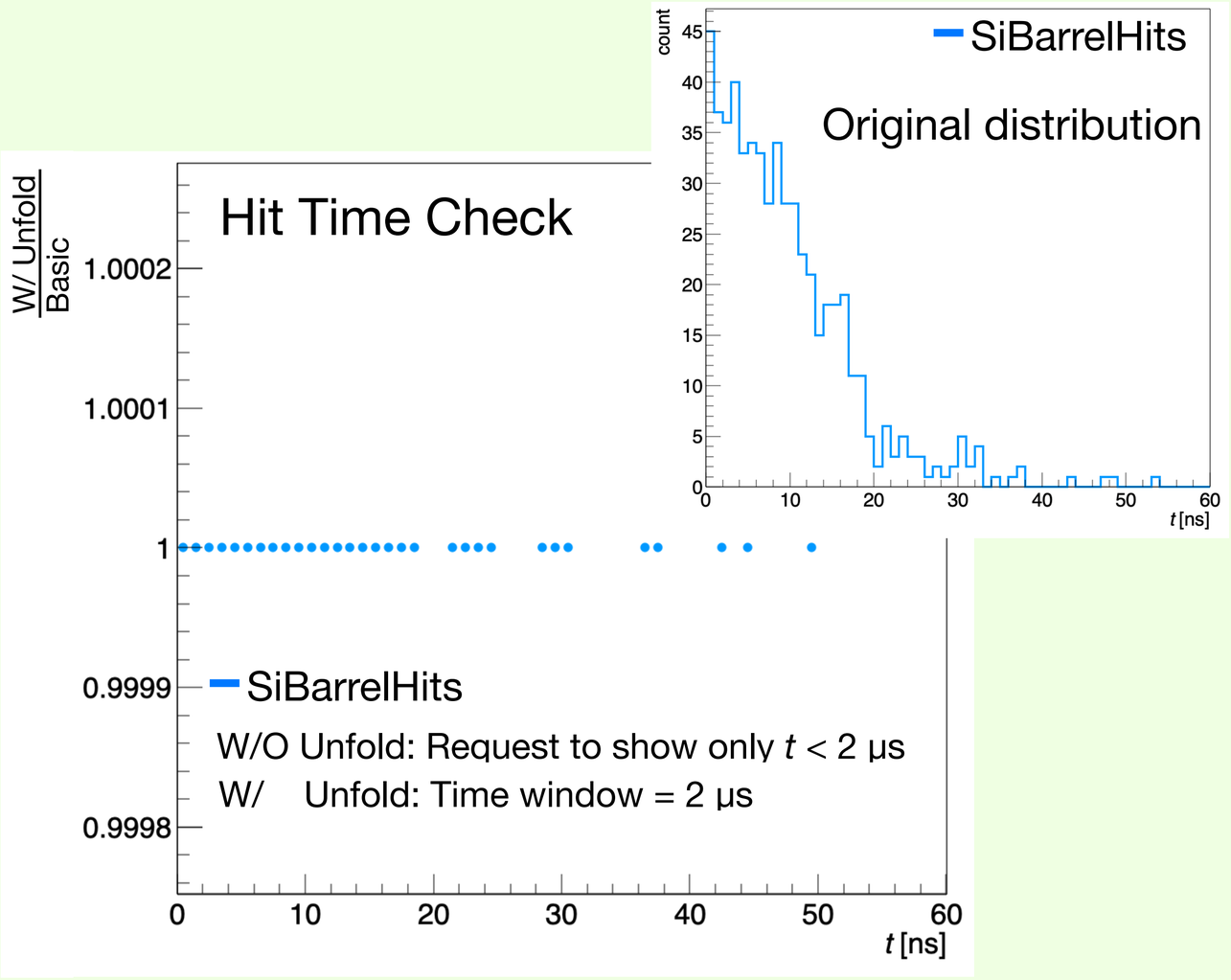
MyTimesliceSplitter (JANA2, unfolding tutorial) Nathan's work:
This code is to test JEventUnfolder.
It deparses each time region as event level.
However, it cannot be used in EICrecon directly.

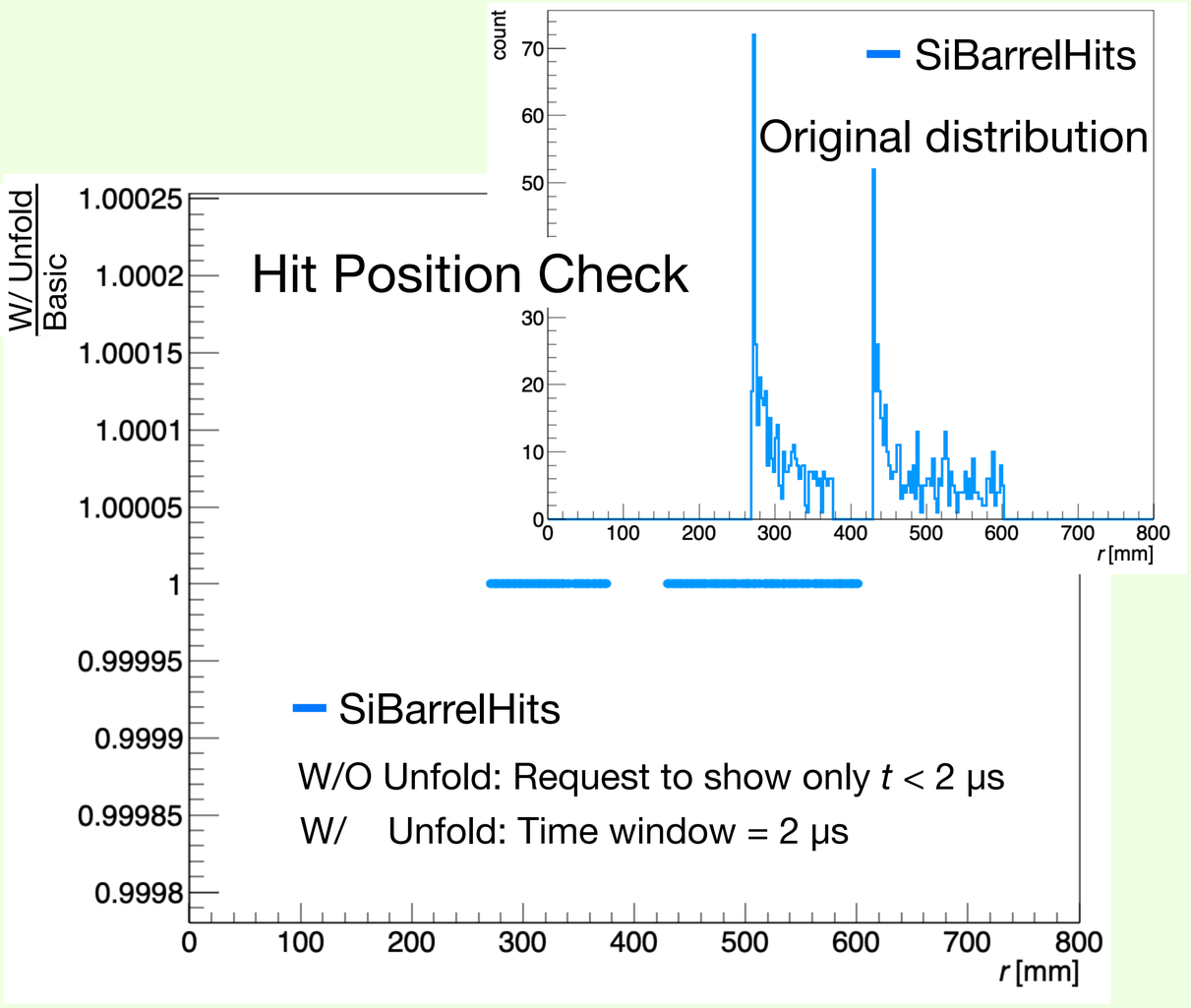
TimeframeSplitter (EICrecon):
This code performs the actual time frame splitting based on JEventUnfolder.

Need to implement time-splitting function without losing physics events.

Compare Original or W/ Unfolding for Position and Energy

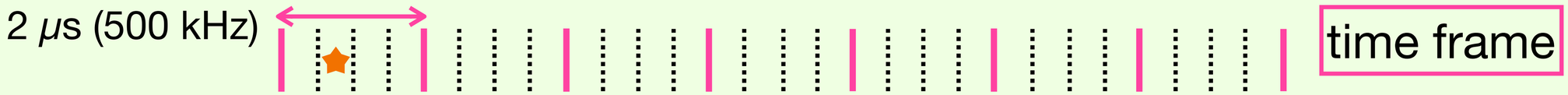
Trivial test: Compare the outputs between the basic topology and the unfold topology





The results show a complete match between cases W/O and W/ unfolding.
These results are reasonable.

Time Splitting results




```
23:44:23.426  [info] Timeslice
HitChecker: Event 188640 Hits in: , 1.02581, 1.03809, 1.12963, 1.2371, 1.61152, 1.63005, 1.77883, 1.93878, 9.64017, 10.5798, 13.1356, 24.2638

child_idx = 0:: TimeframeSplitter: timeslice 188640 iTimeSlice 0 eTimeSlice 4
23:44:23.426  [info] PhysicsEvent
HitChecker: Event 188640 Hits in: , 1.02581, 1.03809, 1.12963, 1.2371, 1.61152, 1.63005, 1.77883, 1.93878

child_idx = 1:: TimeframeSplitter: timeslice 188640 iTimeSlice 4 eTimeSlice 8
23:44:23.925  [info] PhysicsEvent
HitChecker: Event 188640 Hits in:

child_idx = 2:: TimeframeSplitter: timeslice 188640 iTimeSlice 8 eTimeSlice 12
23:44:24.267  [info] PhysicsEvent
HitChecker: Event 188640 Hits in: , 9.64017, 10.5798

child_idx = 3:: TimeframeSplitter: timeslice 188640 iTimeSlice 12 eTimeSlice 16
23:44:24.590  [info] PhysicsEvent
HitChecker: Event 188640 Hits in: , 13.1356

child_idx = 4:: TimeframeSplitter: timeslice 188640 iTimeSlice 16 eTimeSlice 20
23:44:24.889  [info] PhysicsEvent
HitChecker: Event 188640 Hits in:
```

Input data in one time frame

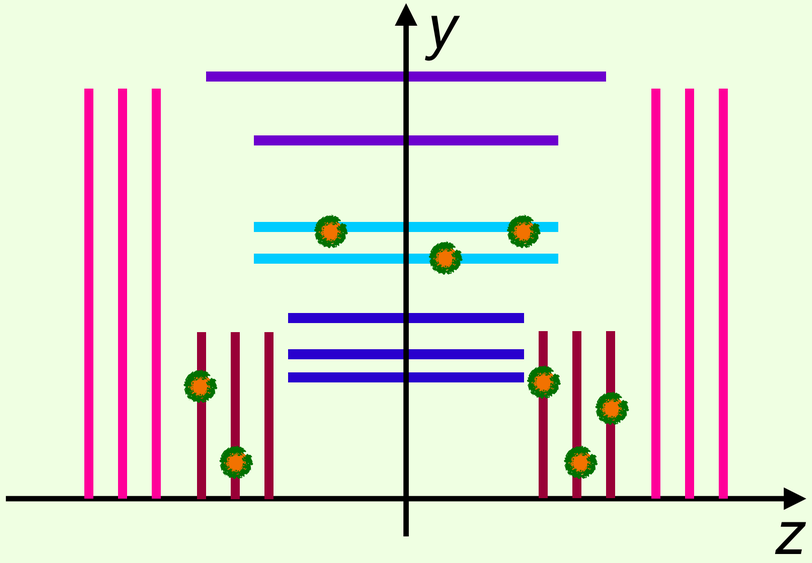
separate into 4 ns
→This value itself has no meaning now. It should be determined by hit time distributions.

As the time split test, it works well.

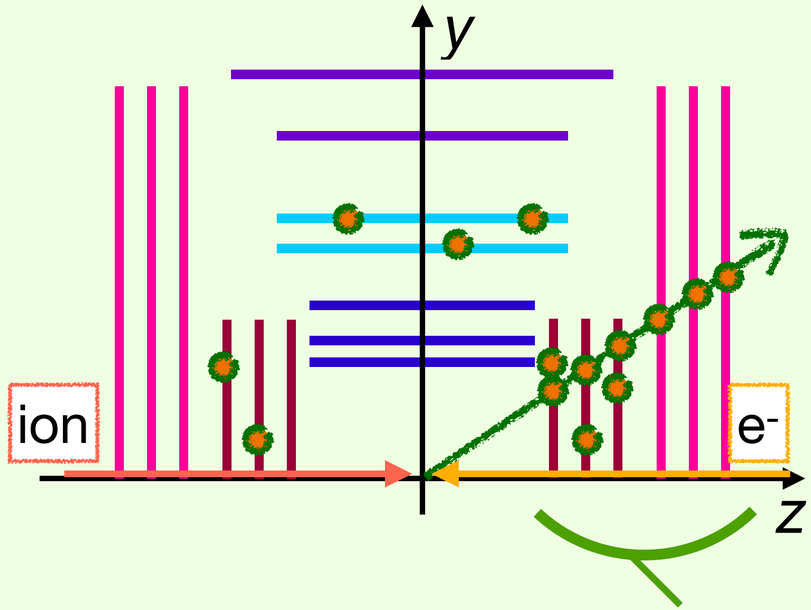
Next: Develop algorithms to decline the intervals of empty or small hits.

Timing-Coincidence

Noisy detectors

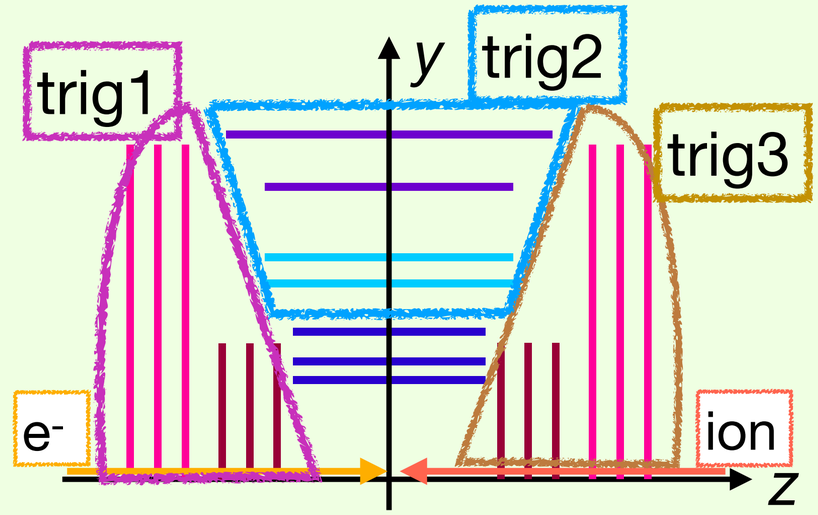


Ideal Track hits



timing coincidence

Simple Triggers



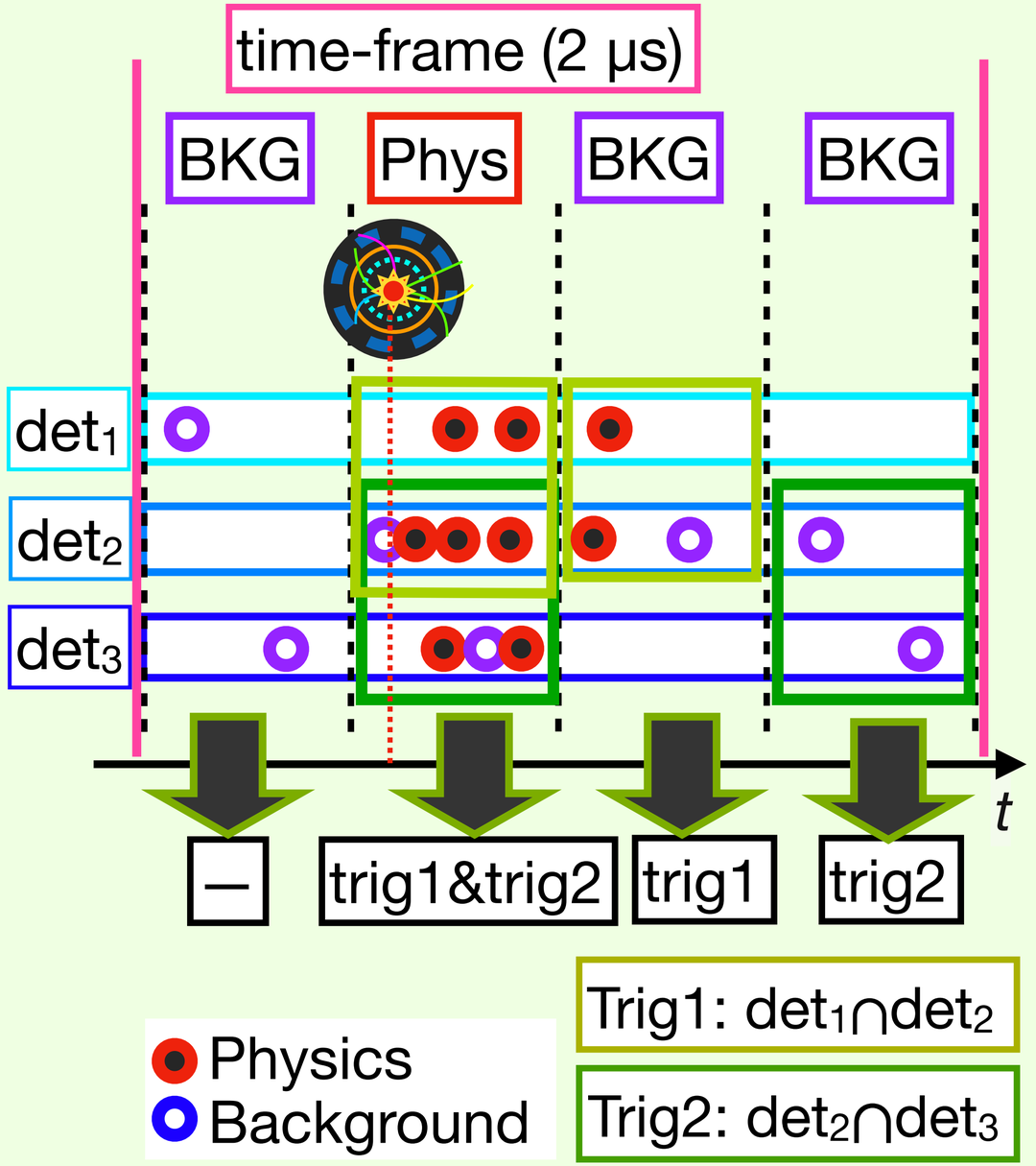
Noisy detectors fire a detector has his high-frequency.
→ The time-coincident using multiple sub-detectors reduce the noise hit time region.

Evaluate Trigger Performance

In physics simulations, where the event properties (time, number of hits, etc.) are known

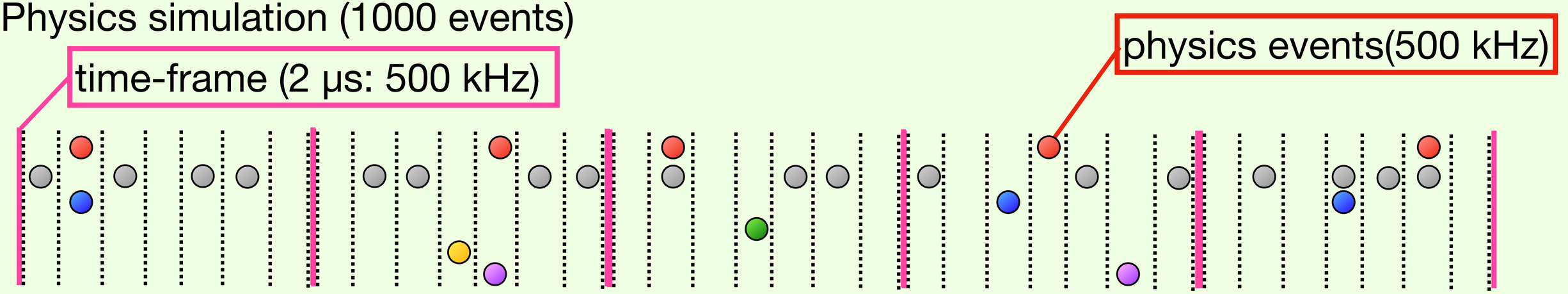
→ It is possible to define physics and noise regions within each time-frame.

- Separate a time-frame by arbitrary-width time slices (e.g., 20 ns).
- Regard the time-slice containing the physics event as a “Physics region”.
  - Other every time-slices are regarded as “Background region”, even if they have hits from physics event.
- Triggers are constructed from hit combinations across multiple sub-detectors.
  (In this study, the number of hits, positions, and timing have not yet been considered.)



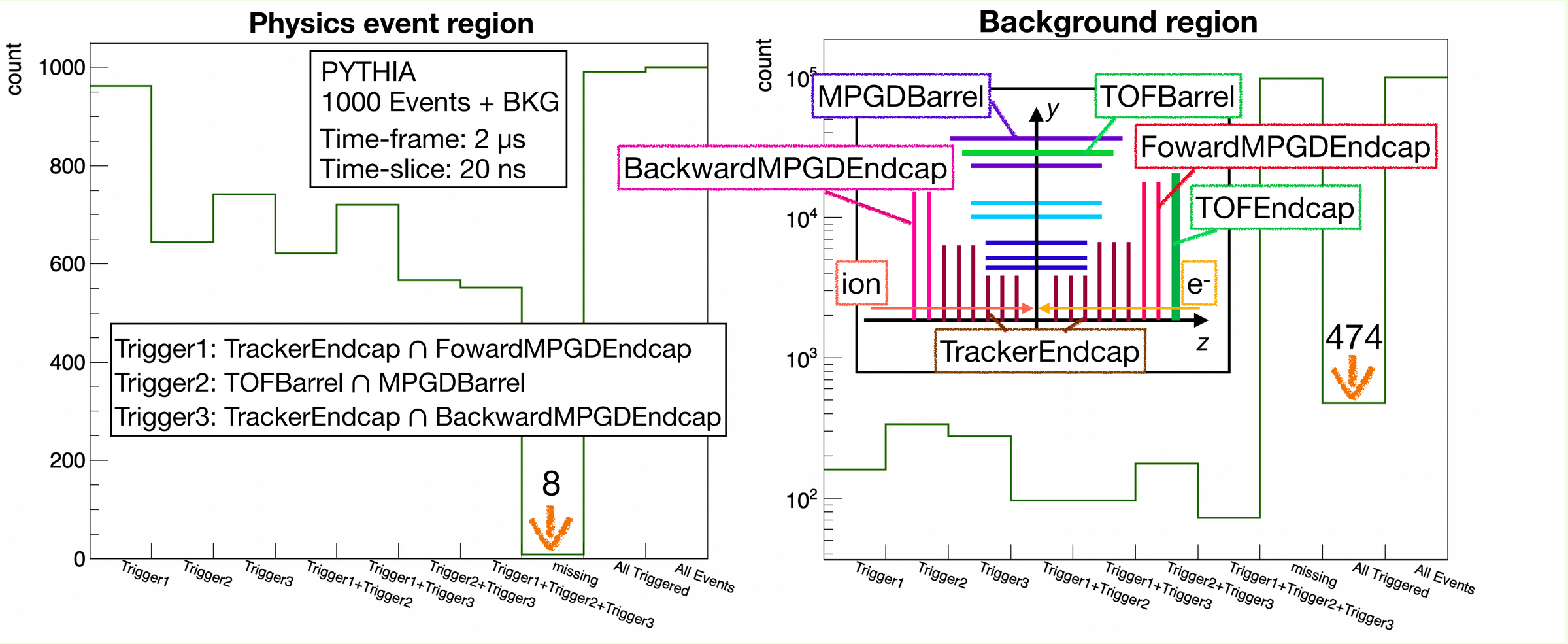
Simulation Setup

Physics simulation (1000 events)



	DIS NC 18x275 $Q^2 > 1$ (Deep inelastic scattering neutral current)	500 kHz
	Synchrontron Radiation	14000 kHz
	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

Trigger Efficiency Performance



※ Background region is not background physics

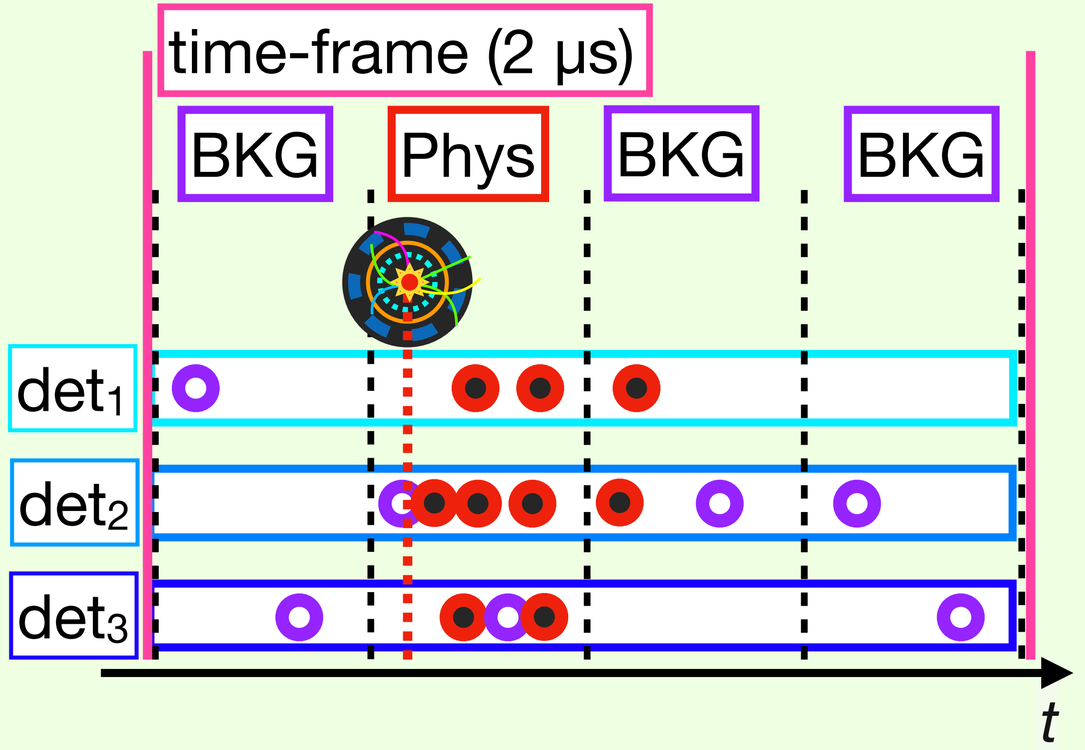
Physics region: Efficiency 99% (992/1000)

Background region: Reduction Ratio 99% (411/10000)

Trigger Efficiency for each Time-slice Width.

- All time slice width could capture over 98% physics events.
→ Wider time-window can recover edge events

- The background reduction rates for all setting are not significantly changed.
Dominant background event is x28 # of physics.
→ **117/28000 → 0.4% (99.6% rejection)**



trigger/time slice width [ns]	5	10	20	200
Physics trigger rate (trigger/all events)	986/1000 = 98.6%	989/1000 = 98.9%	992/1000 = 99.2%	997/1000 = 99.7%
BKG trigger rate (trigger/all events)	1262/400000 = 0.3%	784/200000 = 0.4%	411/100000 = 0.4%	117/10000 = 1.2%

Progress of Study

1. Create a new factory to time alignment (**Done**)
2. Unfolding trivial test (**Done**)
3. Time splitting test (**Done**)
4. Timing Coincidence (**Done**)
5. Simulation with Background (**On going: - Octorber**)
6. Evaluate effects for tracking (October?)
7. Apply detector response (digitization/noise) (October-November?)
8. GPU usage (December?)

Time-splitter document: [link]

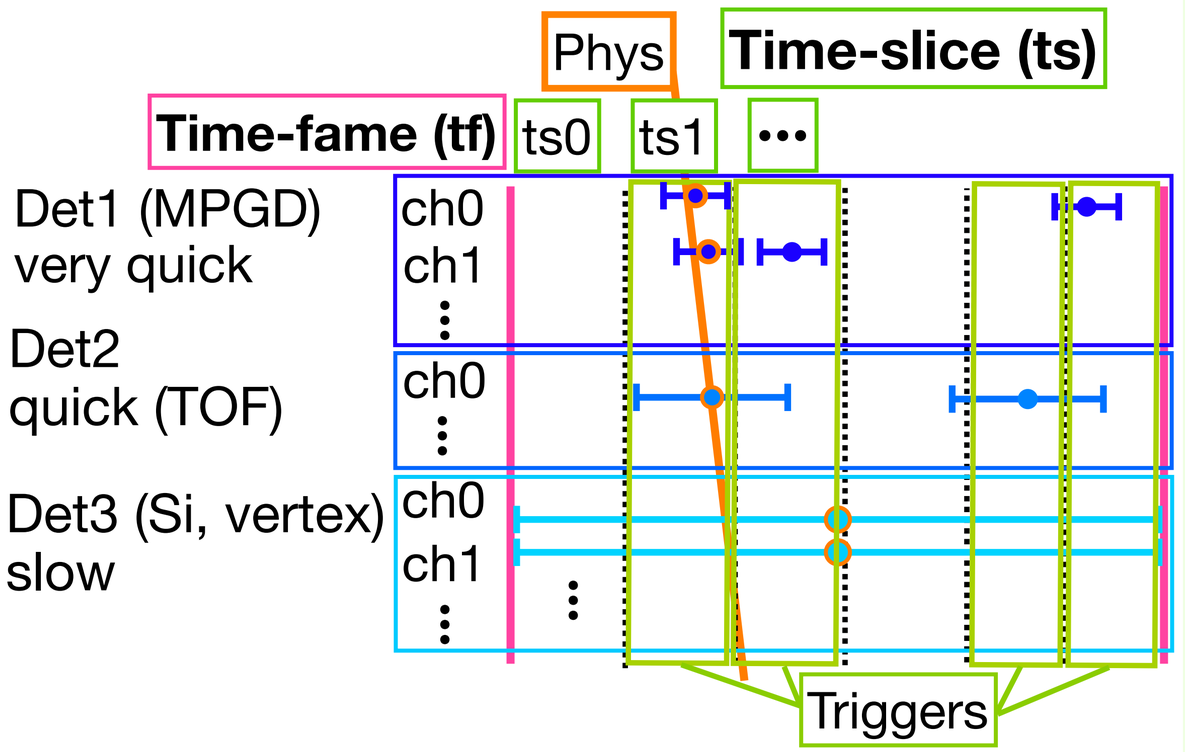
New event build algorithm (considering time resolution of detector)

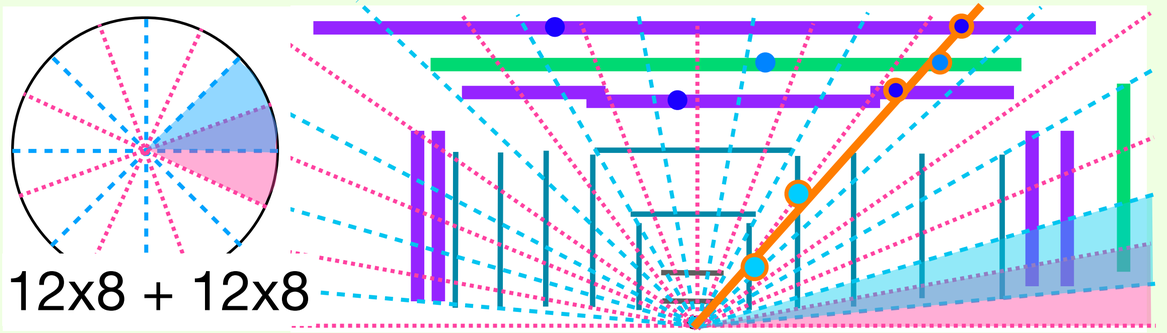
- There is a possibility overwrap time-frame edge by considering hit time resolution.

- Each sub-detector has different time resolution.

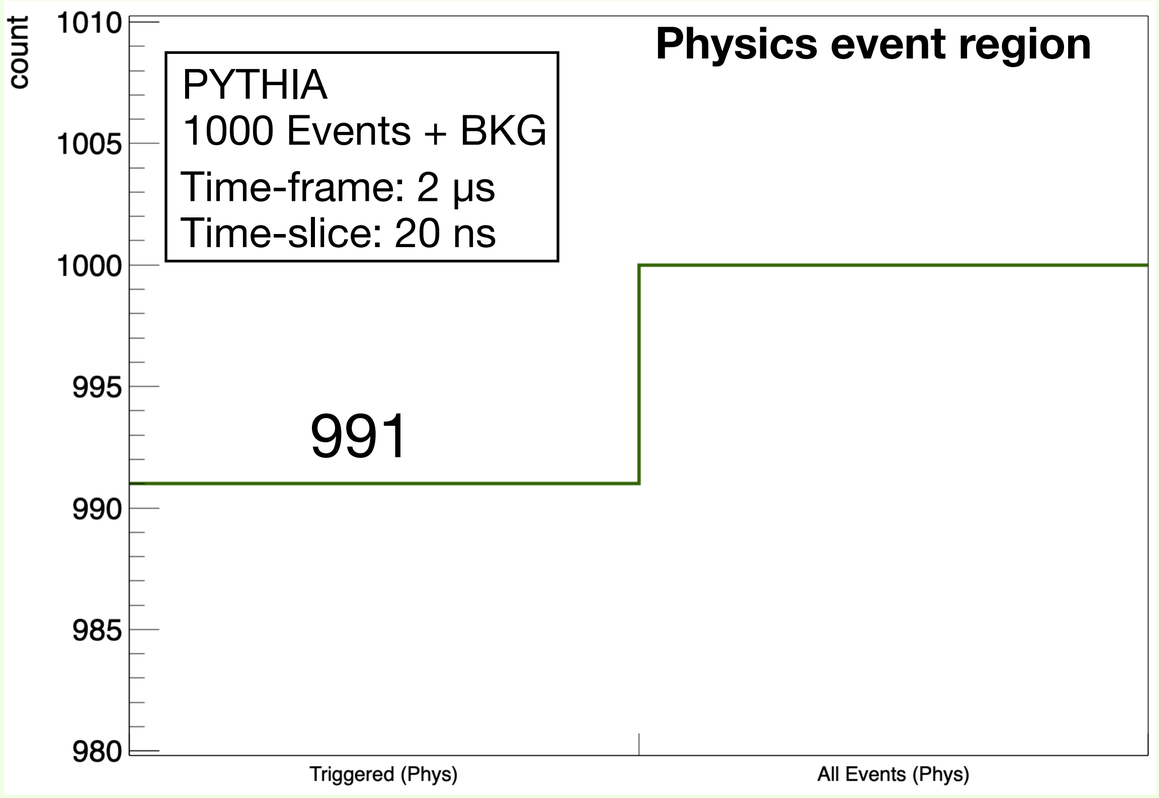
- Silicon detector has large time-resolution (~2 μs).
→ To reduce background, we need to use quick detector like MPGD(~10 ns)/TOF (30 ps).

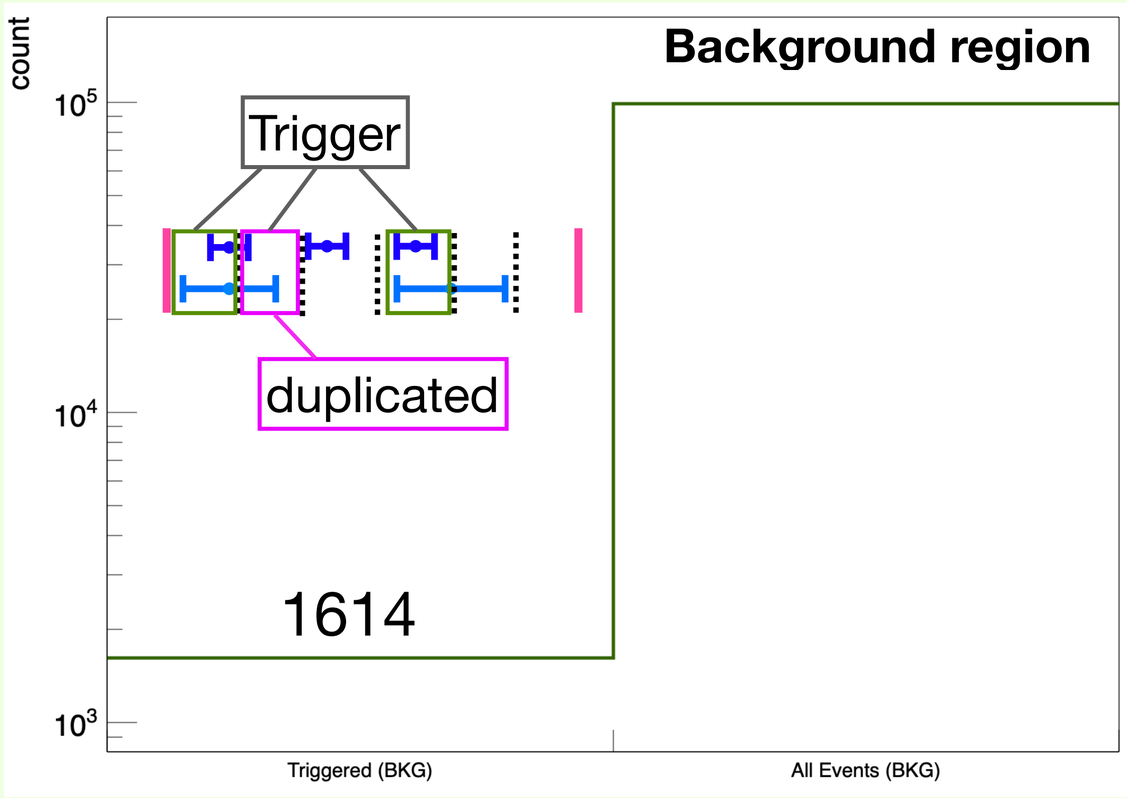
- Request over two hits for one region divided into ~200 (ηxϕ = 2x (12x8))





New Trigger Algorithm of Efficiency





- The efficiency could be kept 99.1%
- Wrong trigger increase.

→ There are duplicated data by considering time resolution

→Need to consider delete such duplicated triggers and more strong selection after reconstruction.

Backup Slides

Simulation Samples

rates in kHz	5x41 GeV	5x100 GeV	10x100 GeV	10x275 GeV	18x275 GeV	Vacuum
Total ep	12.5 kHz	129 kHz	184 kHz	500 kHz	83 kHz	
hadron beam gas	12.2kHz	22.0kHz	31.9kHz	32.6kHz	22.5kHz	10000Ahr
	131.1kHz	236.4kHz	342.8kHz	350.3kHz	241.8kHz	100Ahr
electron beam gas (Bremsstrahlung scatterings)	2181.97 kHz	2826.38 kHz	3177.25 kHz	3177.25 kHz	316.94 kHz	10000Ahr
electron beam gas (Coulomb losses - w/ collimators)		116.57 kHz		29.56 kHz	0.86 kHz	10000Ahr
electron intrabeam (Touschek losses - w/ collimators)		1112.3 kHz		233.5 kHz	0.55 kHz	
DIS eA	kHz	kHz	kHz	/	/	
hadron beam (Au) gas	7.36kHz	10.3kHz	10.3kHz	/	/	10000Ahr
	79.1kHz	110.7kHz	110.7kHz	/	/	100Ahr
electron SR					3324 MHz	

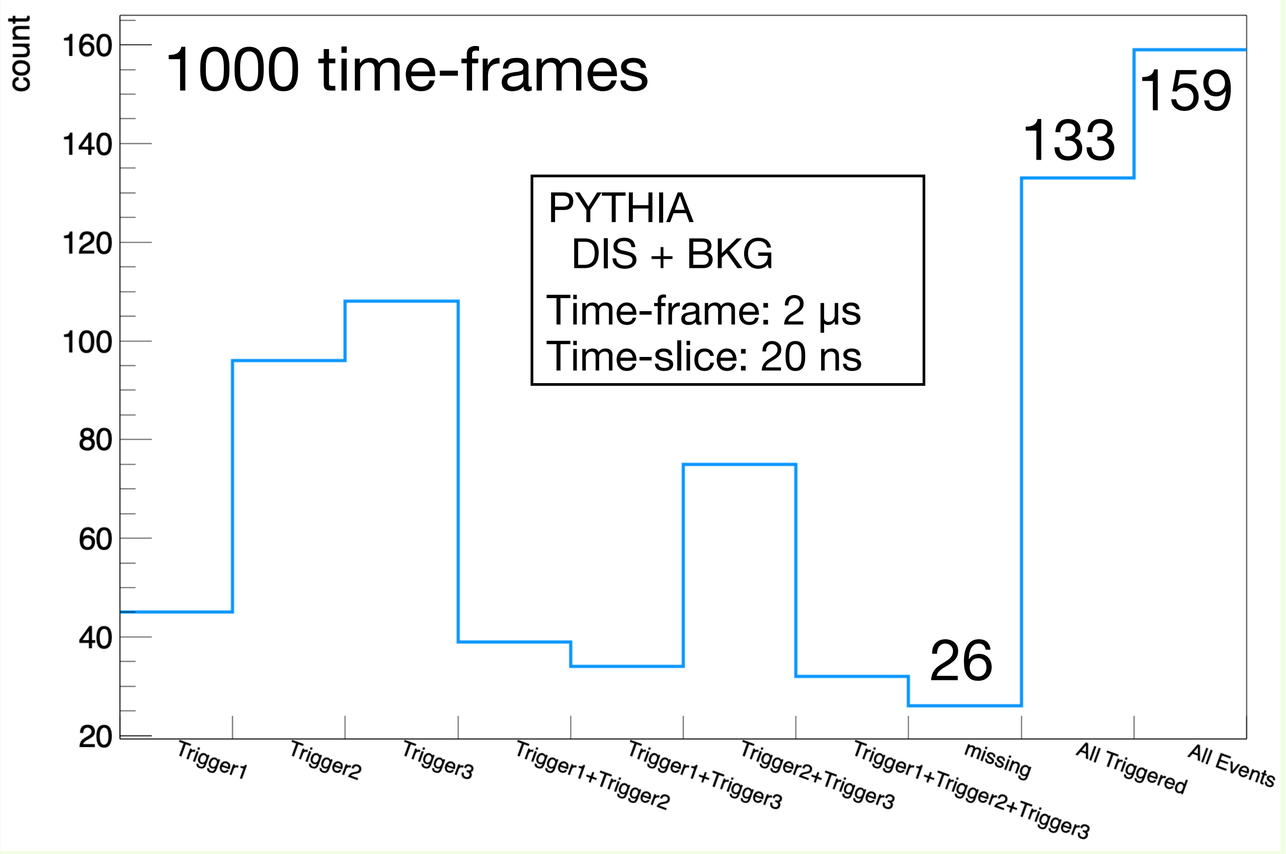
https://wiki.bnl.gov/EPIC/index.php?title=Background

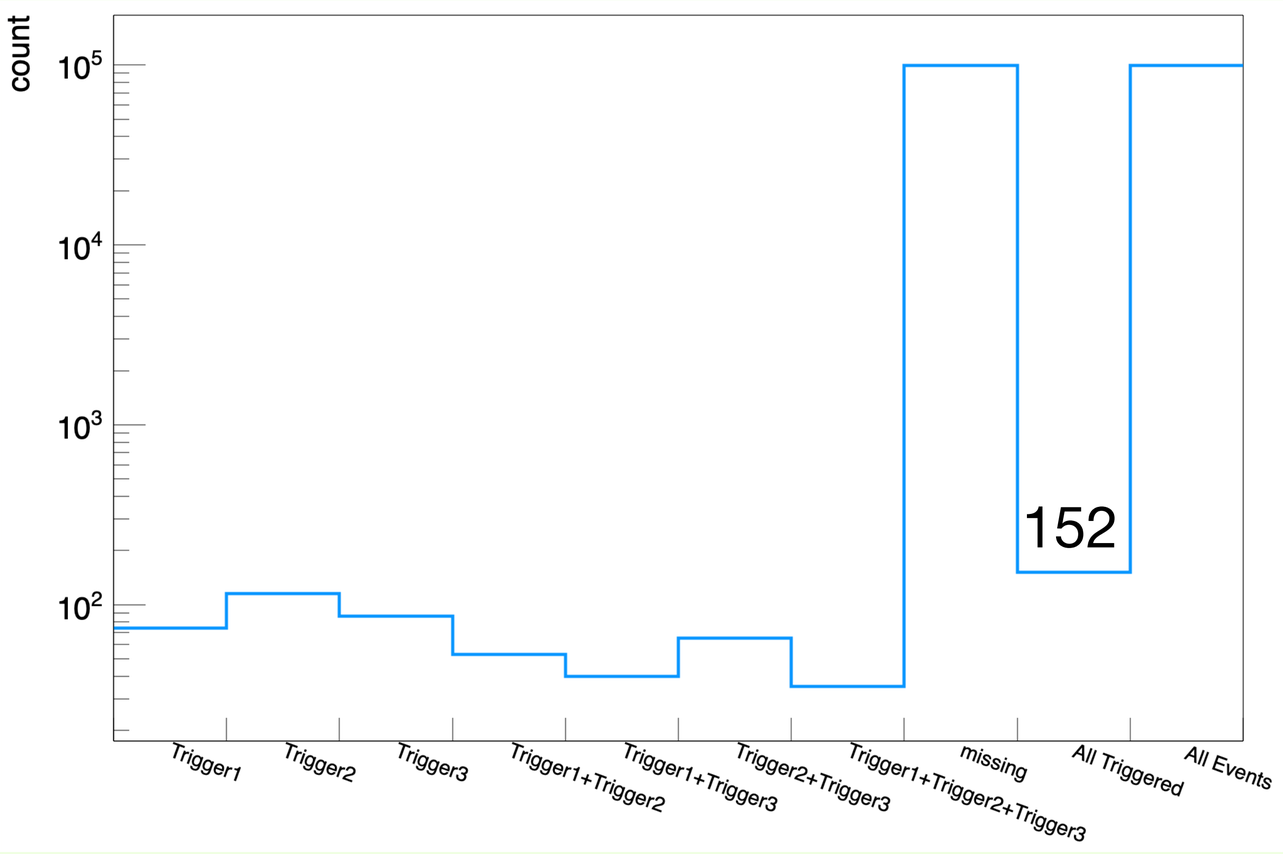
This list says the collision energy is different, but other page says no…
→ I am asking Sakib but I have not yet gotten.
If this list is collect I want to test 18x275 GeV with 500 kHz.

Case2 Simulation (83 kHz)

83 kHz Physics events (Case2)

https://github.com/eic/eic.github.io/blob/master/_resources/background_mixed_samples.md





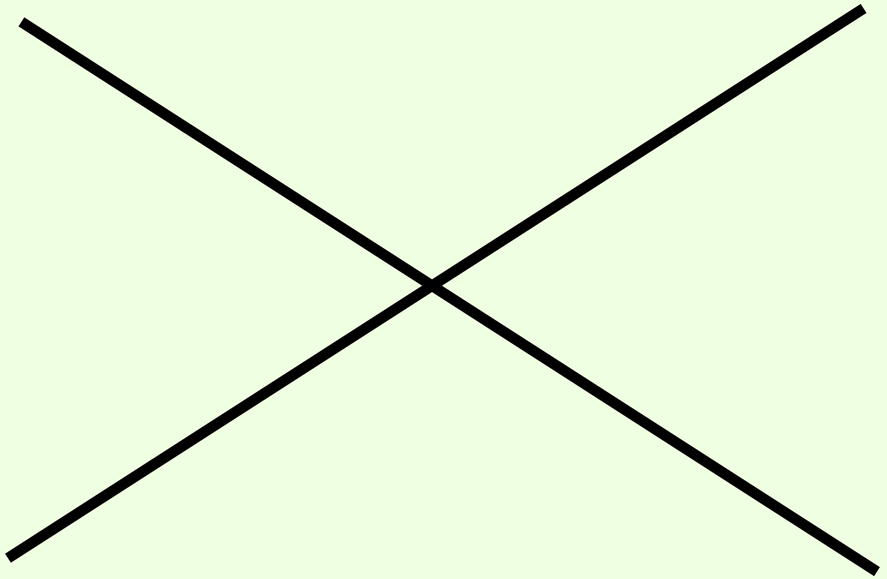
In the case2, the physics event is same with the case1.
→The trigger rate should be same, however very lower than the result of case1.
→ Searching for the reason.

Case3 Simulation (Only background)

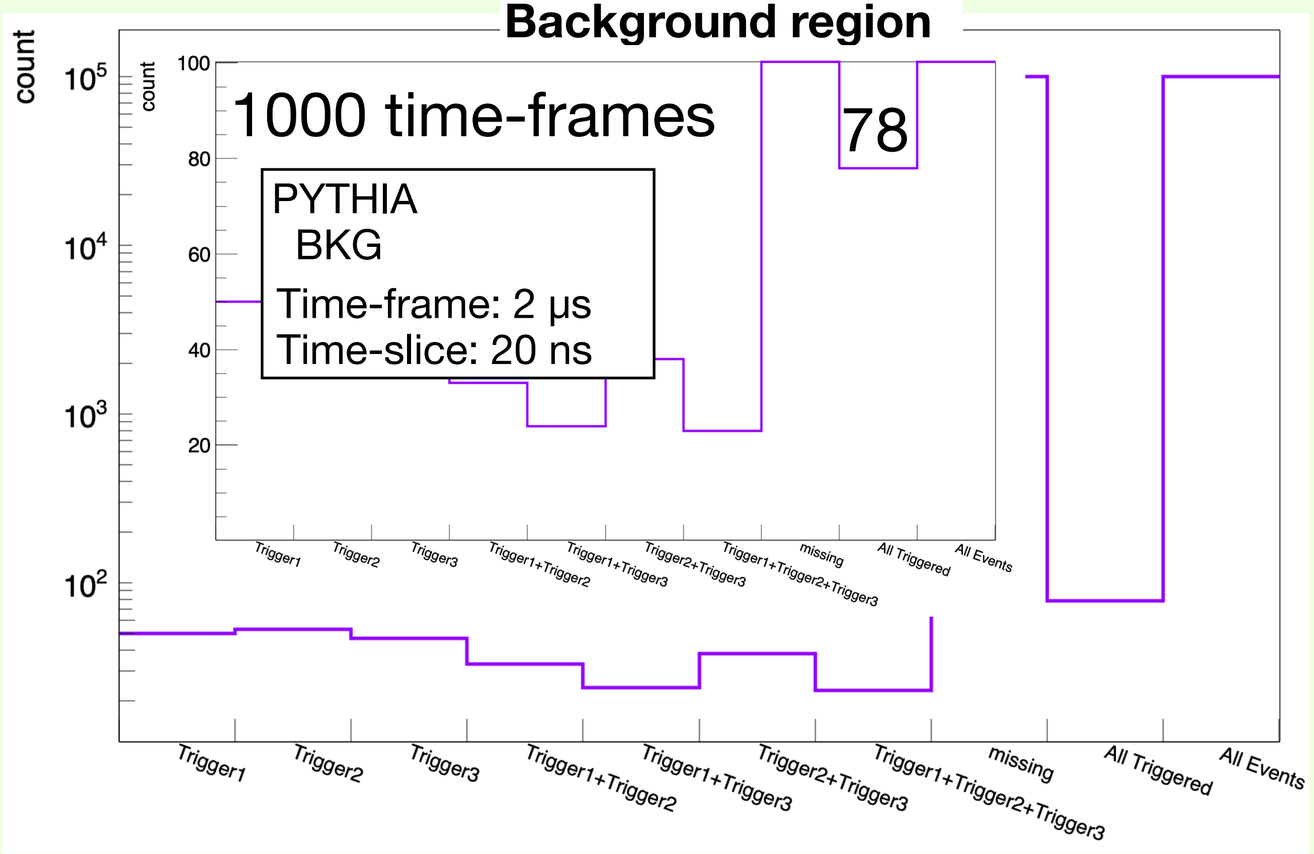
Only background (Case3)

https://github.com/eic/eic.github.io/blob/master/_resources/background_mixed_samples.md

Physics event region



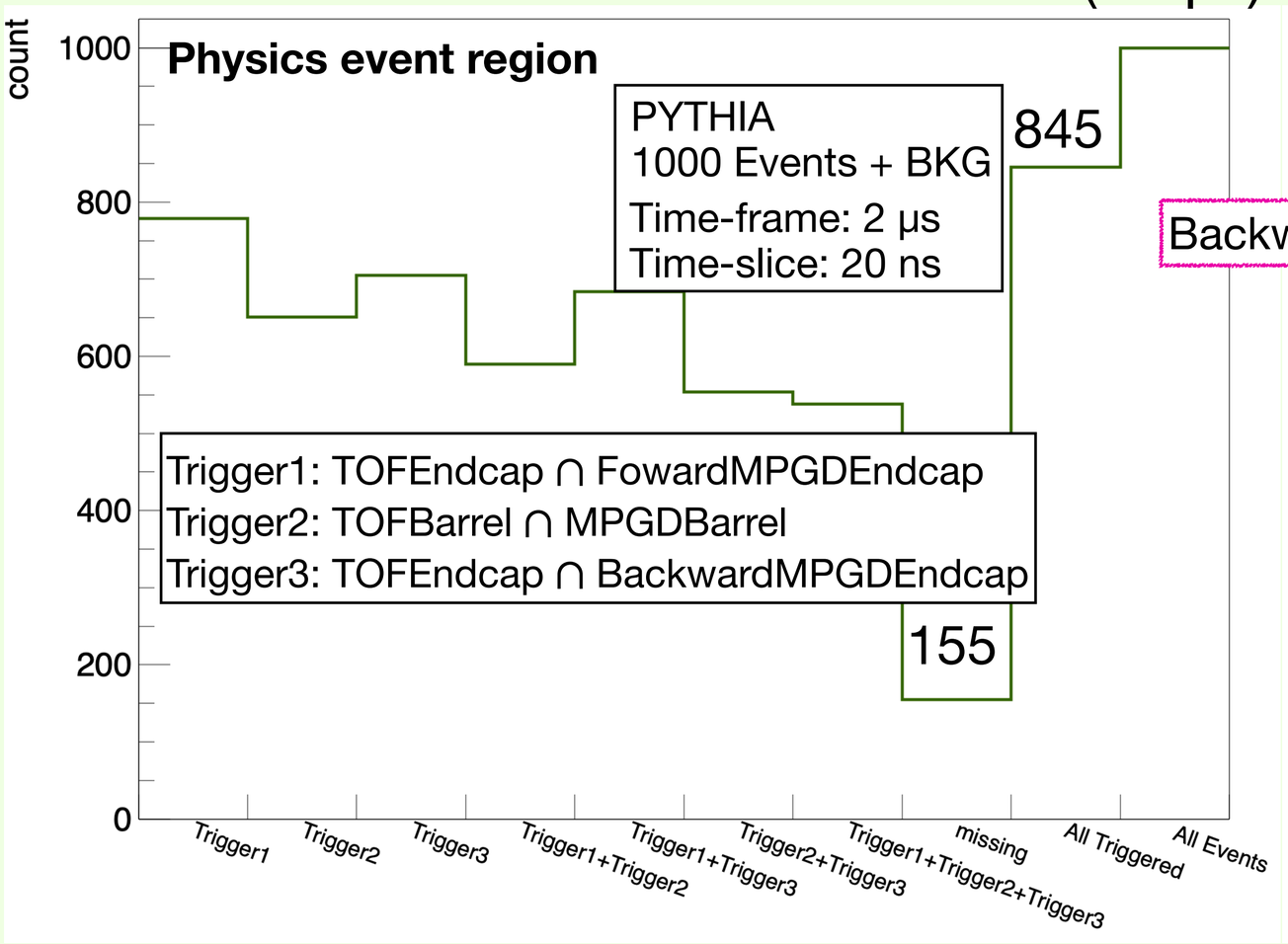
Background region

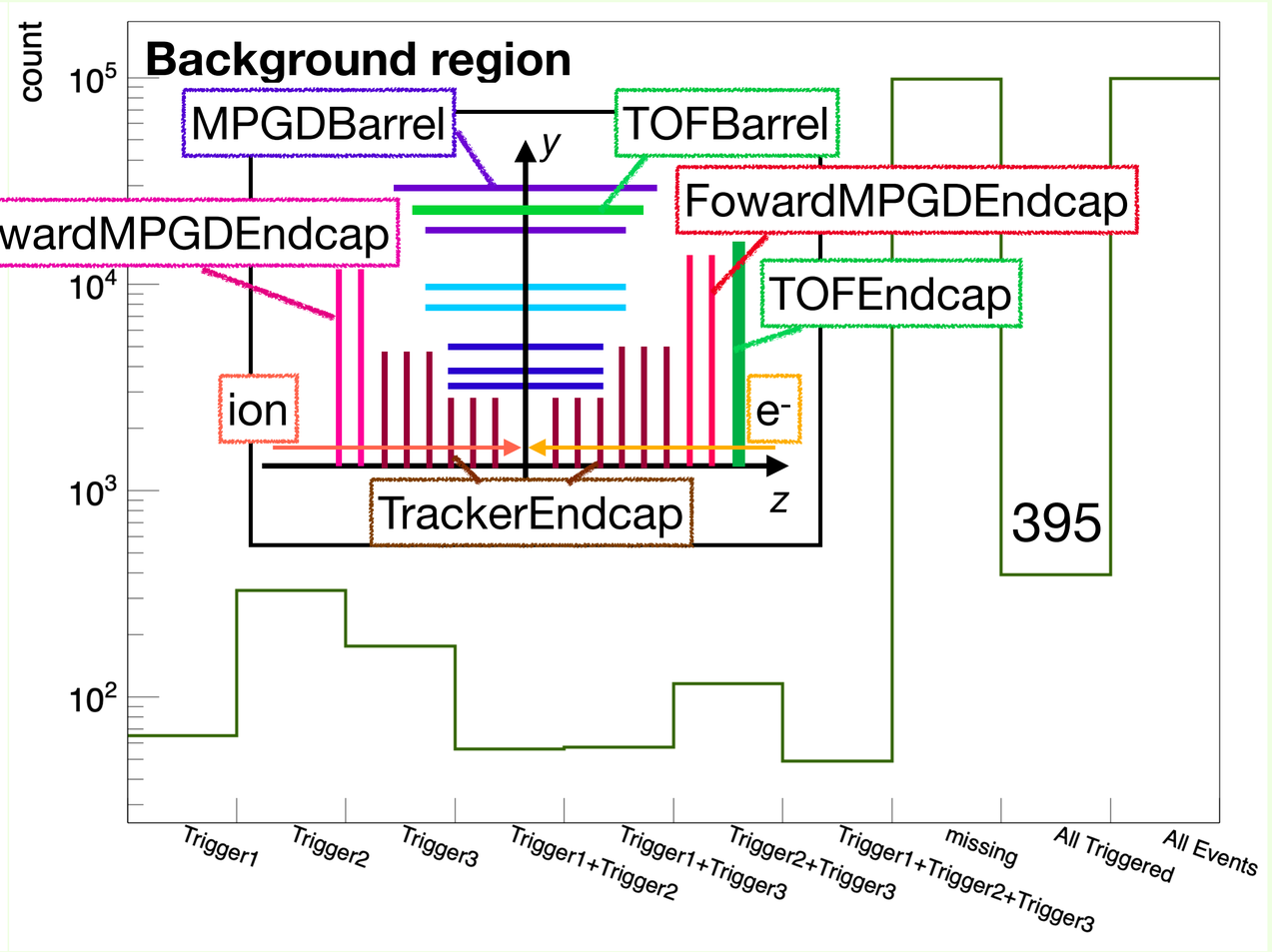


In the case3, is no background.
And background rejection ratio is consistent with other cases.

Remove Silicon Detector from Trigger Setup

Silicon detector is a little slow detector (~2 μs)



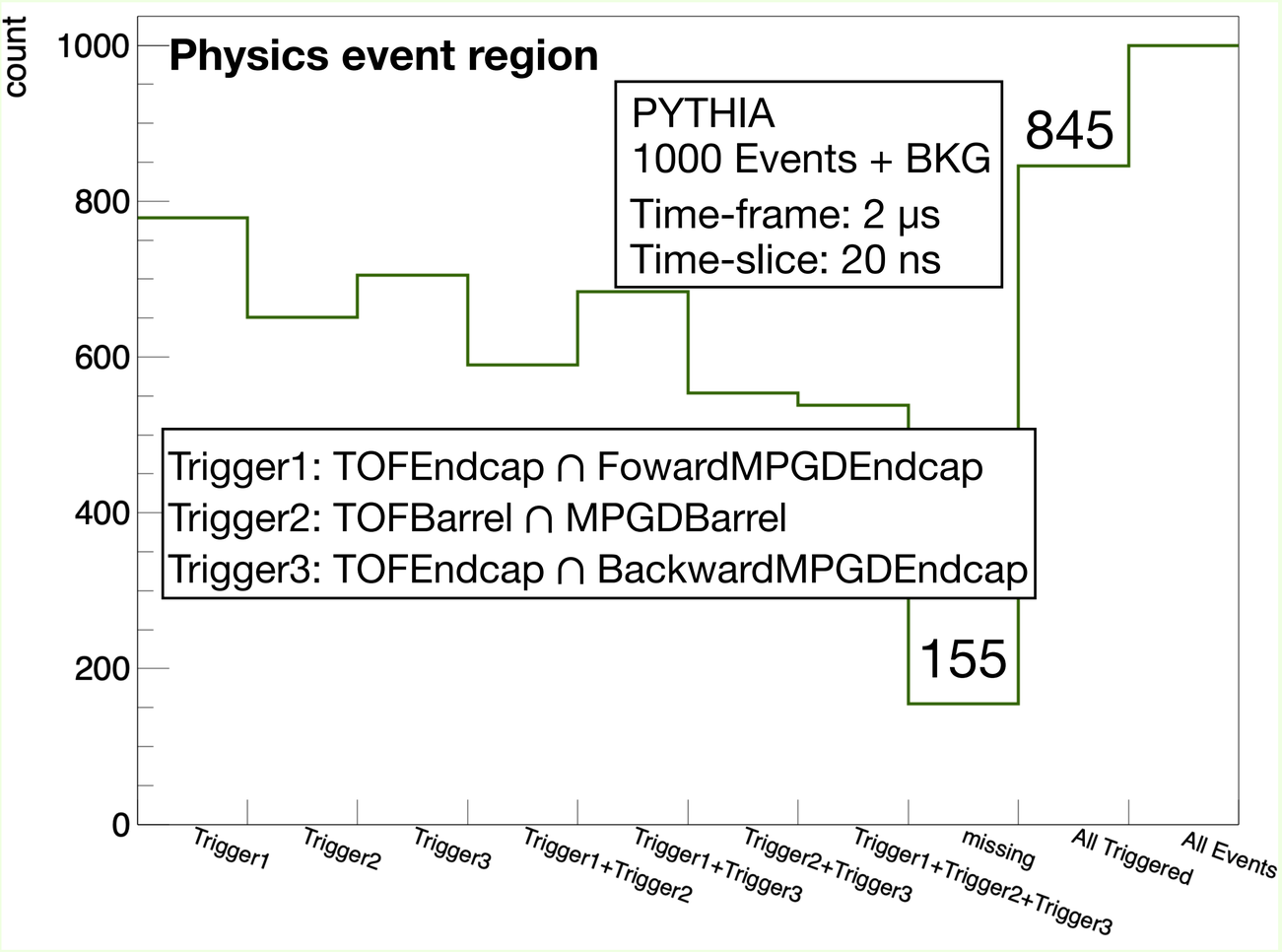


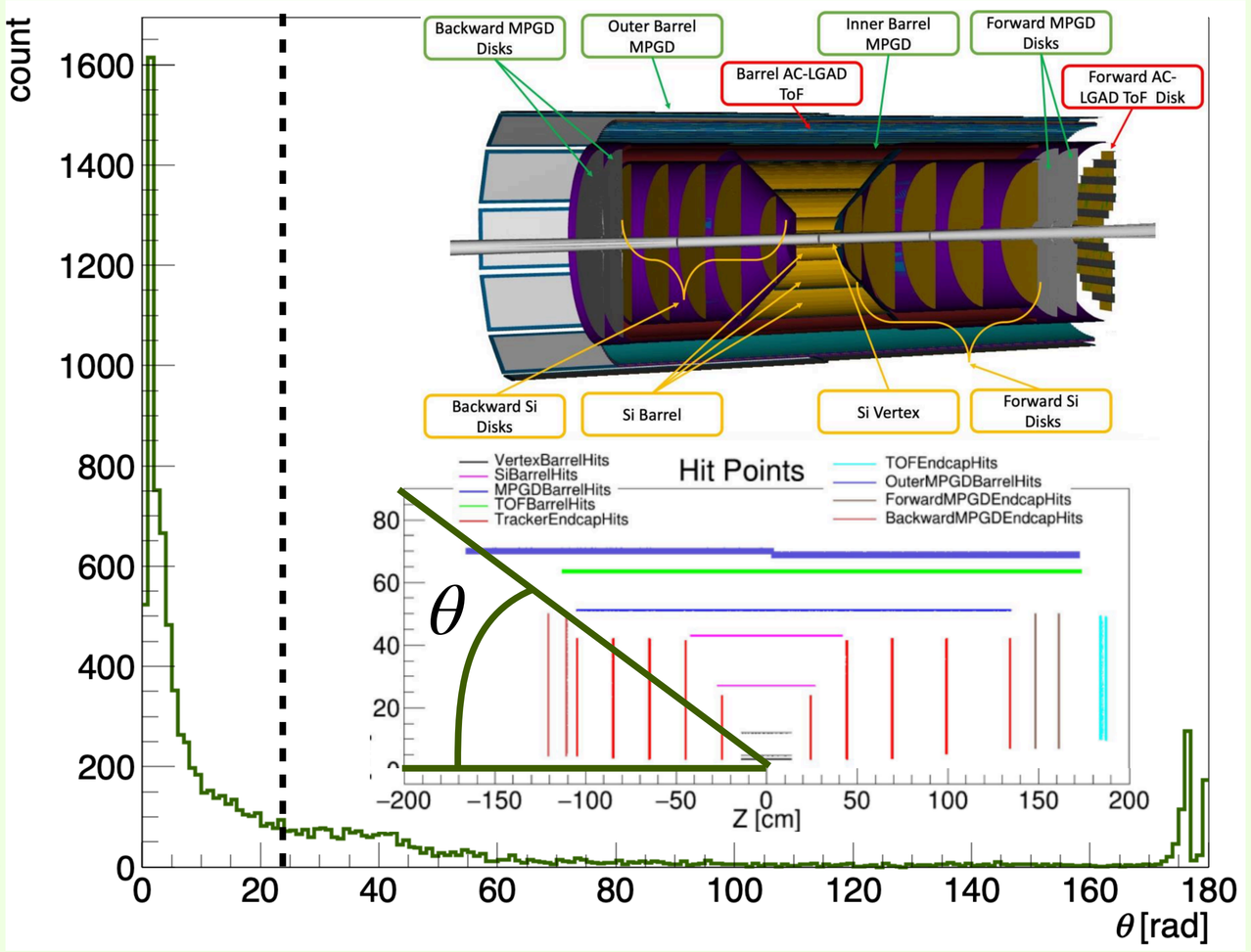
I remember there is no TOF detector at backward region.

→ The missing evens have only backward region.

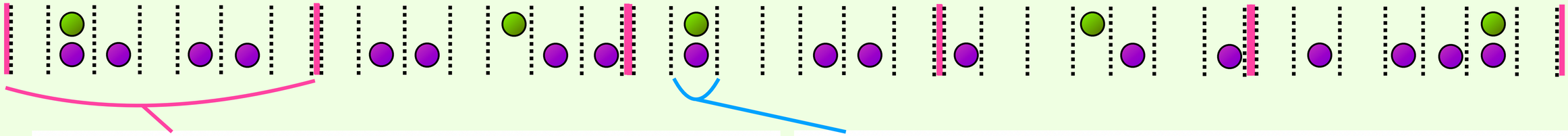
→ I have to use the silicon endcap or loose trigger like only use MPGD.

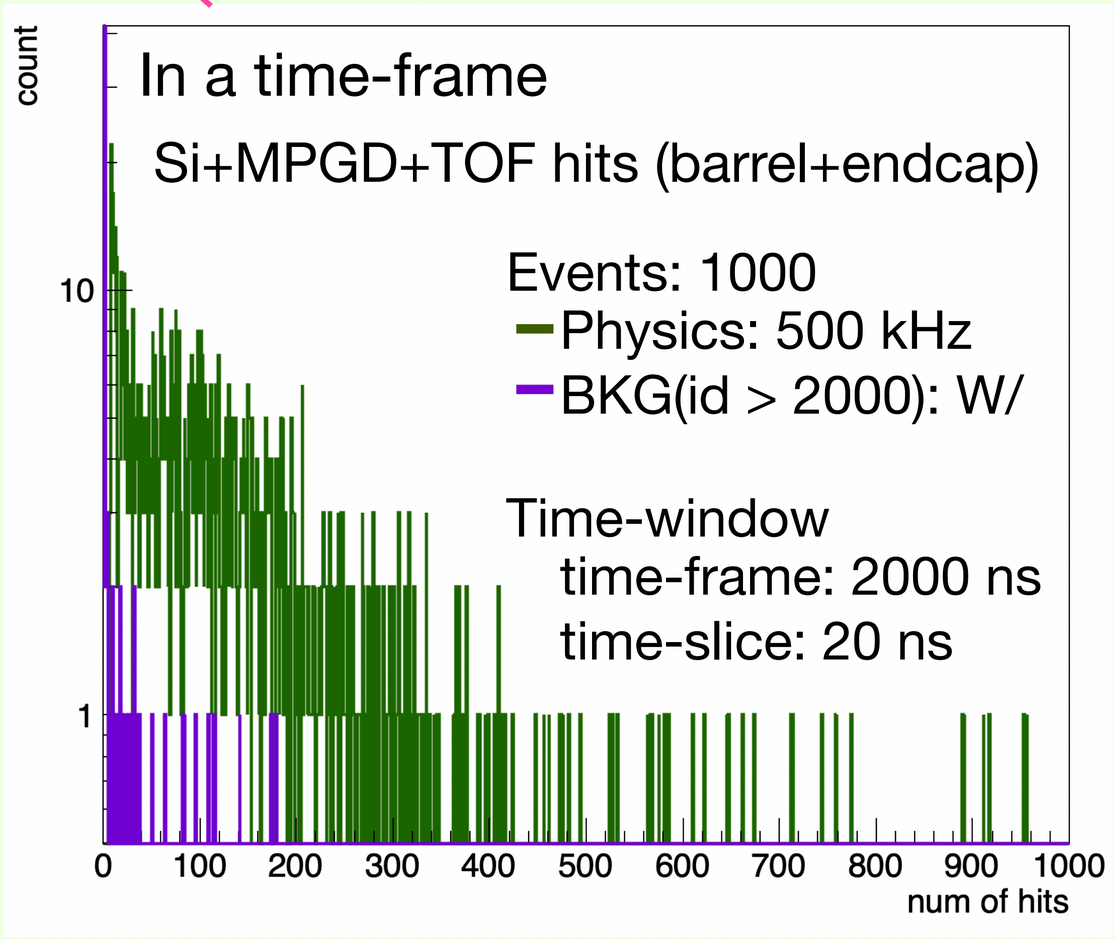
Missing Event Characteristics

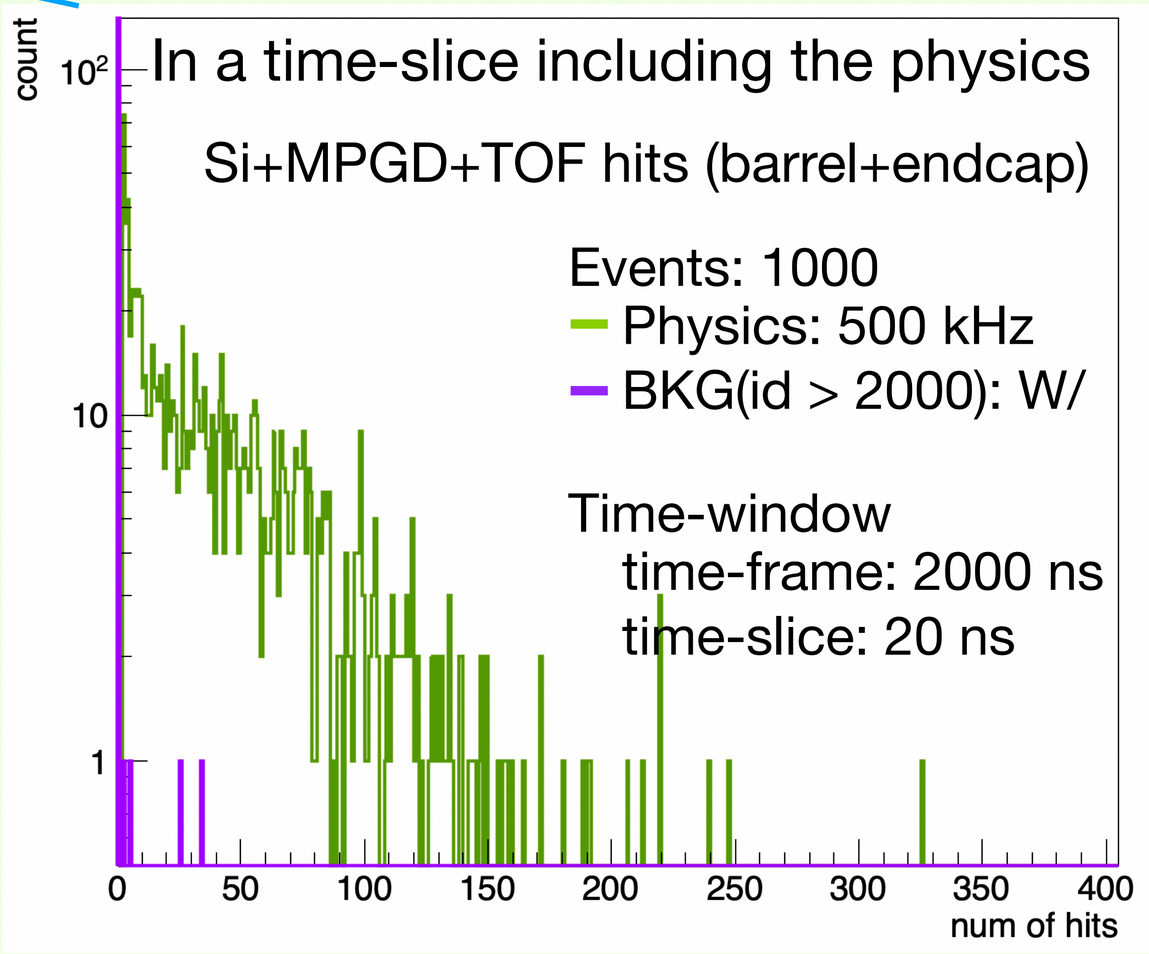




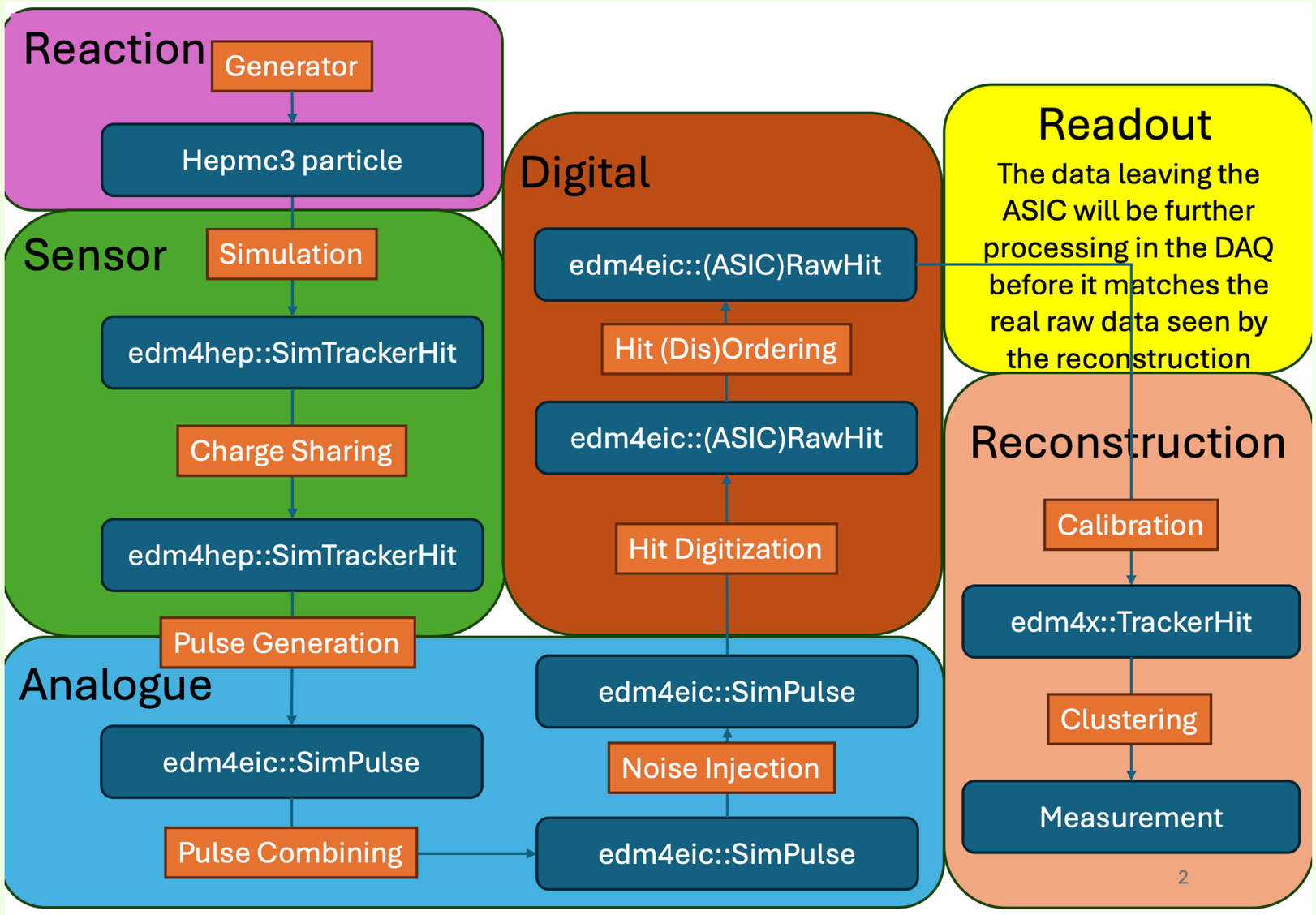
Number of Physics/Background Hits in Time-Window







EICrecon Workflow



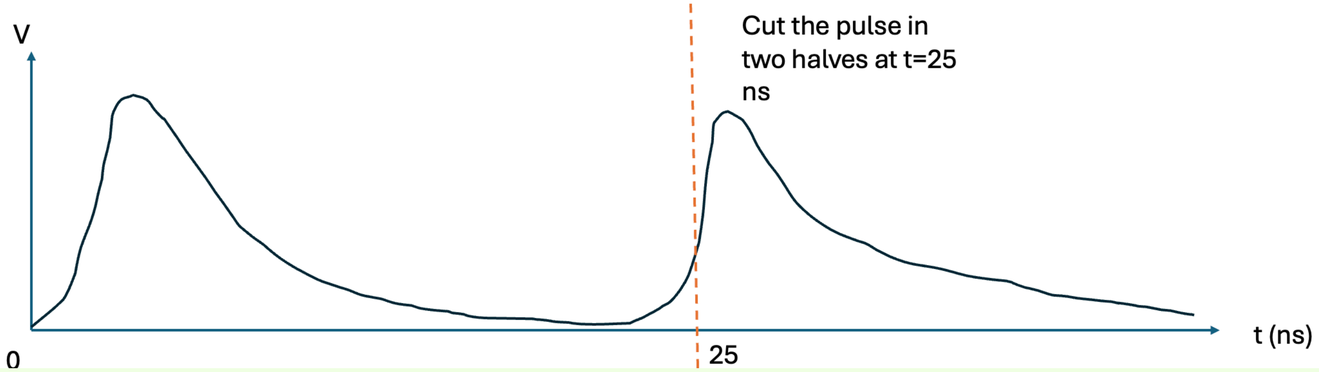
Simon Gardner

https://indico.bnl.gov/event/27734/contributions/106318/attachments/61235/105199/EICrecon%20Digitization%20Developments.pdf

Slides of the TOF Digitization Report 1

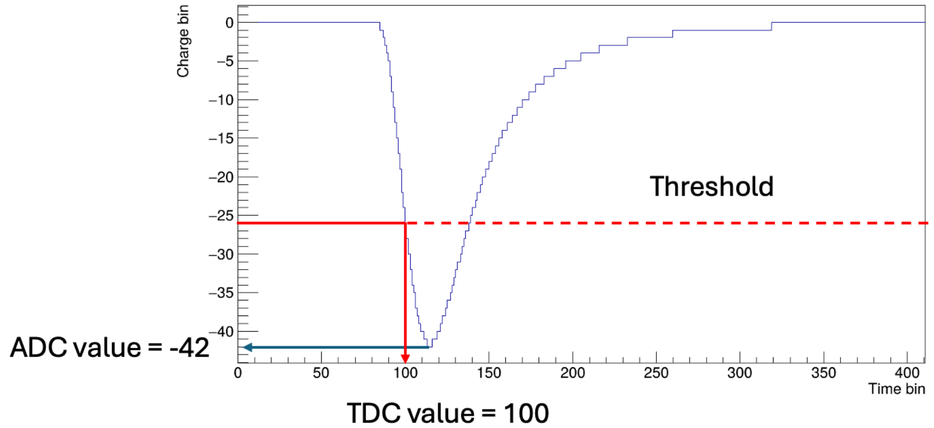
Pulse Discretization: SiliconPulseDiscretization

- Pulse needs to be cut into 25 ns segments.
- If peaks happens near the 25 ns boundary, it will be registered as two different hit.



Pulse digitization: EICROCDigitization

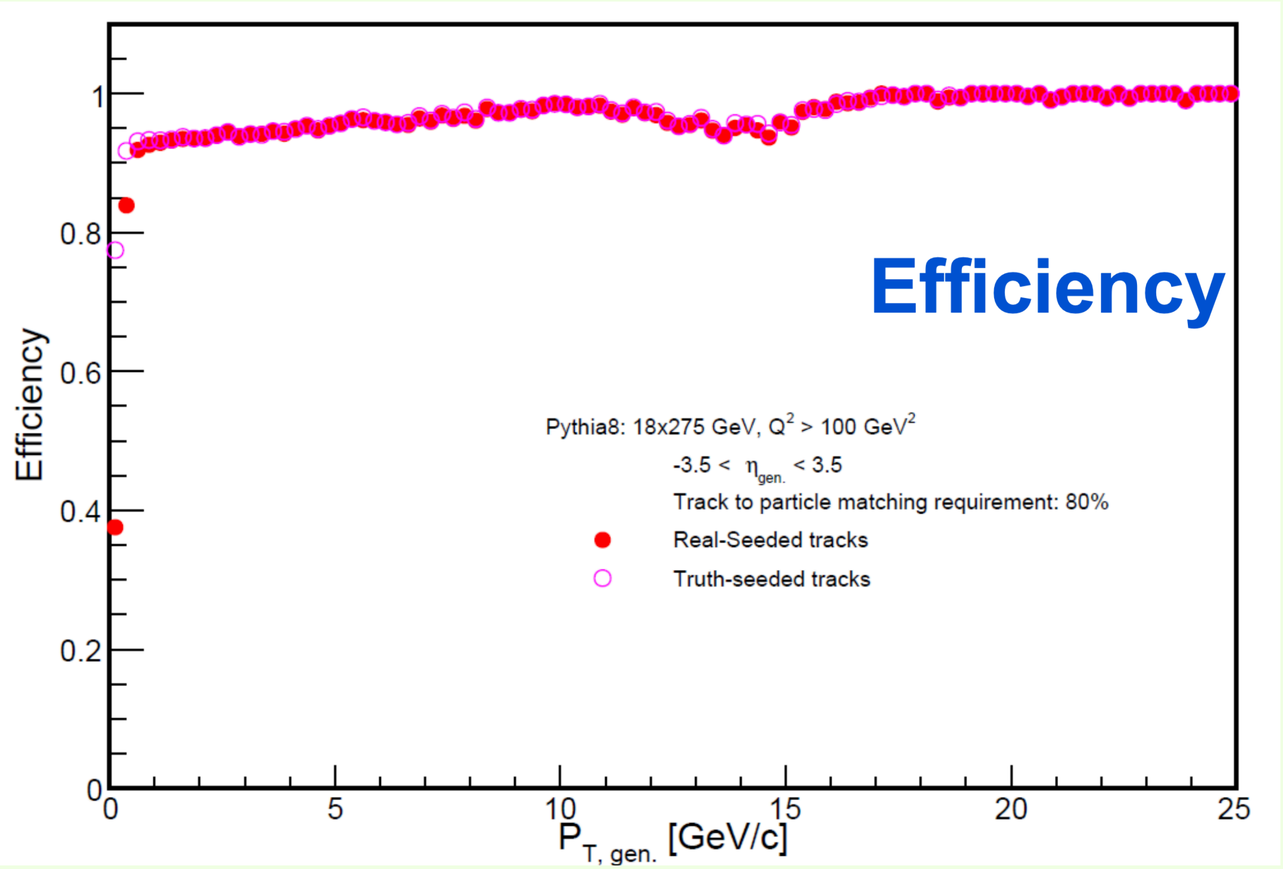
- Time when Charge value crosses a threshold -> TDC value.
- Max(Min) pulse height -> ADC value

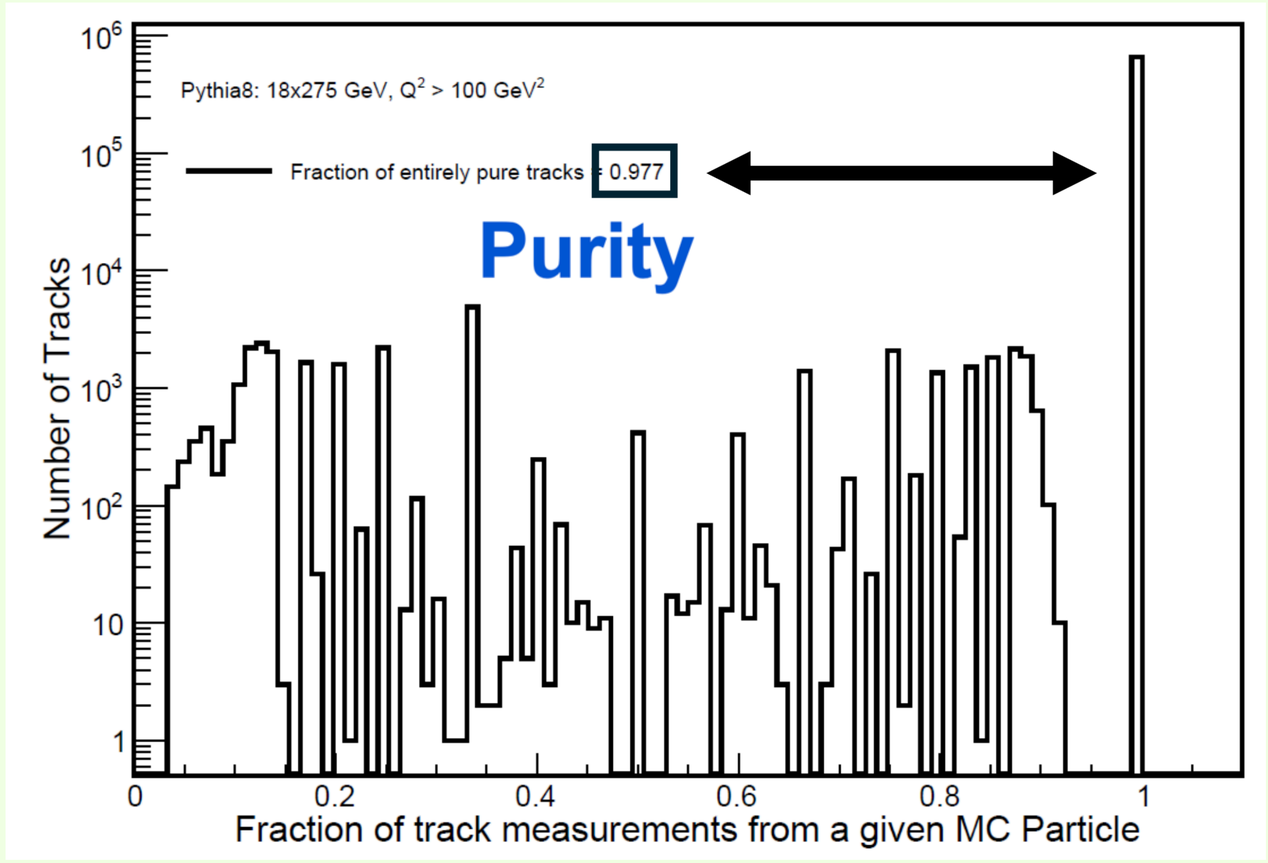


https://indico.bnl.gov/event/28217/

https://github.com/eic/EICrecon/pull/1779/files

Tracking Quality Evaluation





https://indico.jlab.org/event/934/contributions/17080/attachments/13196/21252/tracking_summary_071825.pdf

https://github.com/eic/snippets/tree/main/Tracking/DISAnalysis