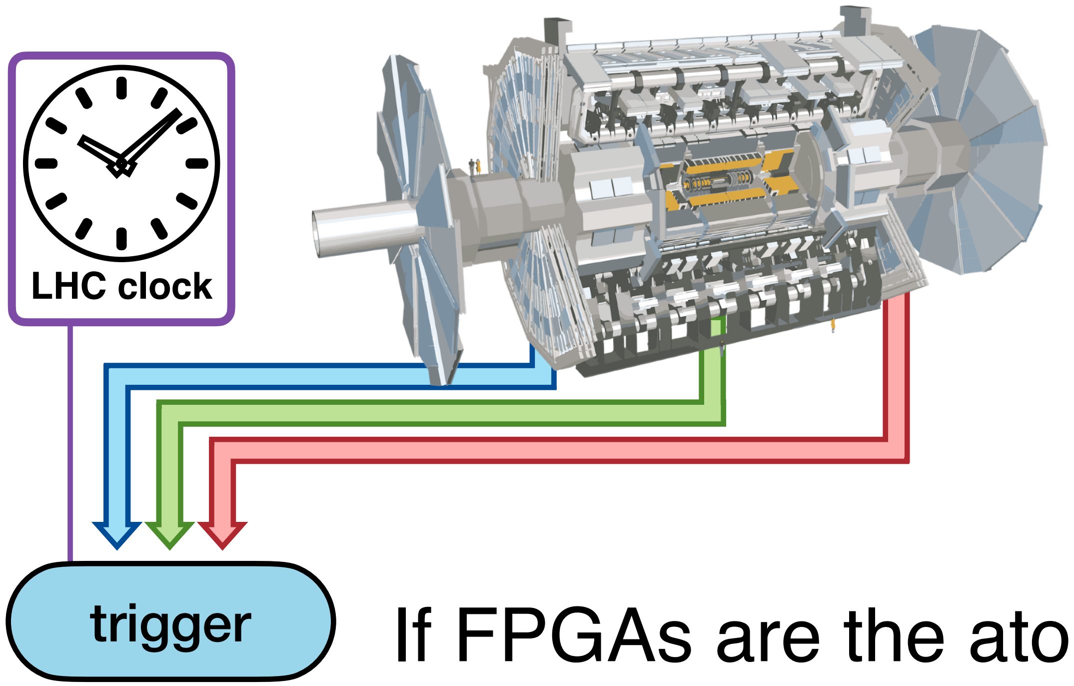
Trigger and DAQ challenges at the LHC

Christian Herwig
July 22, 2024

Plan

- **Yesterday afternoon: Concepts**
  - Introduction and physics goals of the LHC Experiments
  - The role of the trigger and DAQ systems
  - DAQ concepts: simple toy model → complex systems
- **This morning: Details and Implementations**
  - **Trigger concepts: from hardware to architecting a system**
  - The challenge of high luminosity, and the upgrades
  - ATLAS, CMS hardware triggers, and their evolution
  - Software triggers, and new analysis paradigms
  - Wrap up

Designing a rapid-reconstruction system

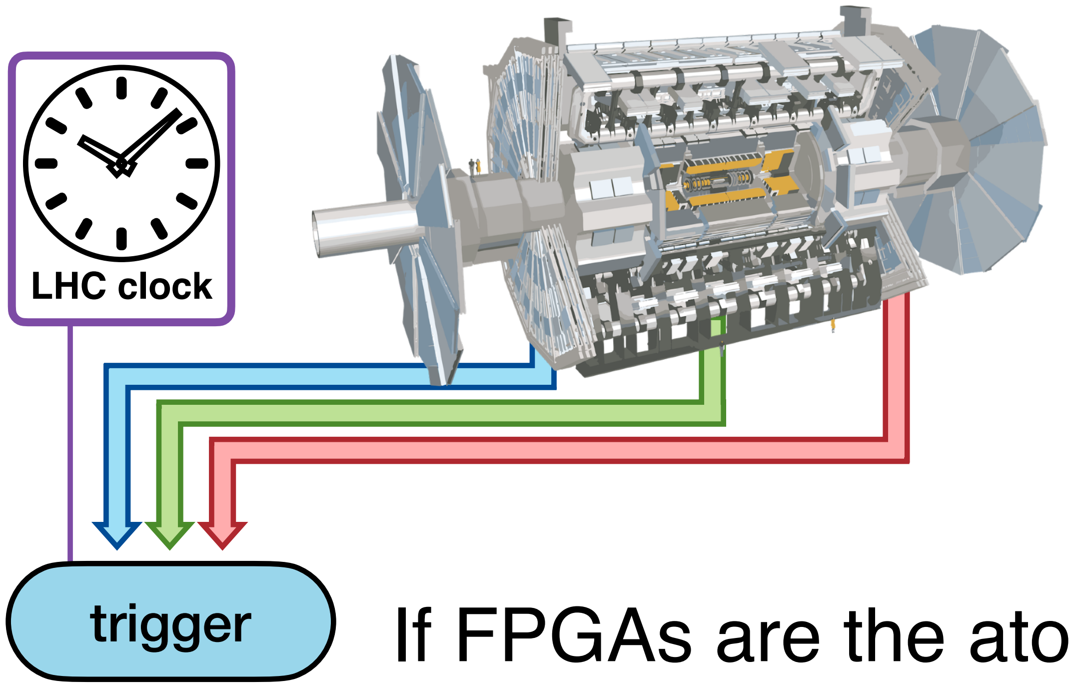


Trigger MUST process new events at the experimental frequency — quickly!

@LHC this is 40 MHz, 1/(25 nanosec)

If FPGAs are the atomic units, how to architect a complete system?

Designing a rapid-reconstruction system



Trigger MUST process new events at the experimental frequency — quickly!

@LHC this is 40 MHz, 1/(25 nanosec)

If FPGAs are the atomic units, how to architect a complete system?

Pipelined processing

Time multiplexing

Simplified algorithms

Regional reconstruction

Pipelining

Trigger systems rely on **pipelines** to perform *complicated reconstruction* tasks while handling a stream of *continuous inputs*.

- They enable systems with *long latency*, *short initiation interval*.

???

???

???

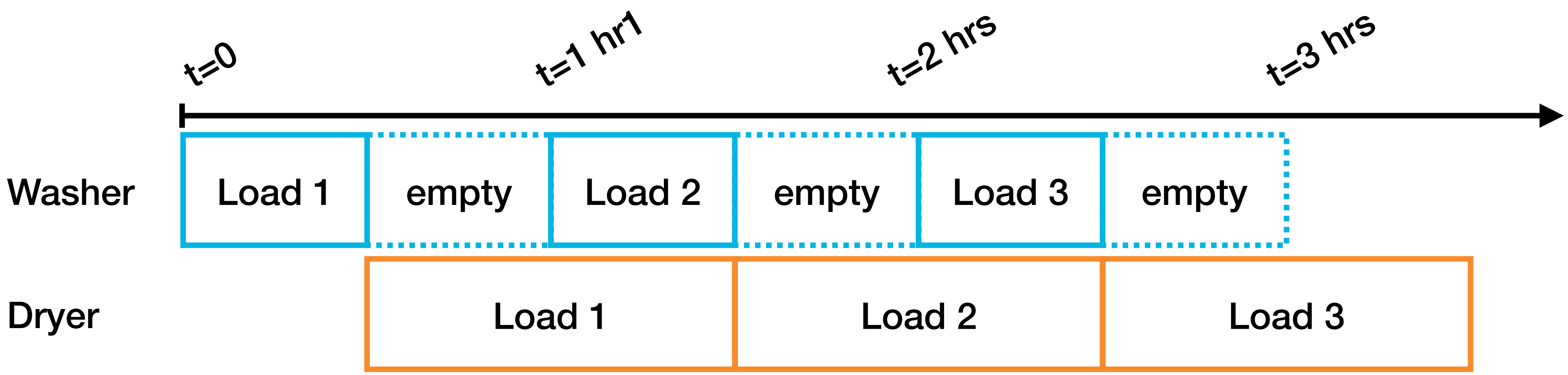
A real life example

Trigger systems rely on **pipelines** to perform *complicated reconstruction* tasks while handling a stream of *continuous inputs*.

They enable systems with *long latency*, *short initiation interval.*

Example: laundry in my apartment building:

1 washer (runs 30min) and 1 dryer (runs 1hr).



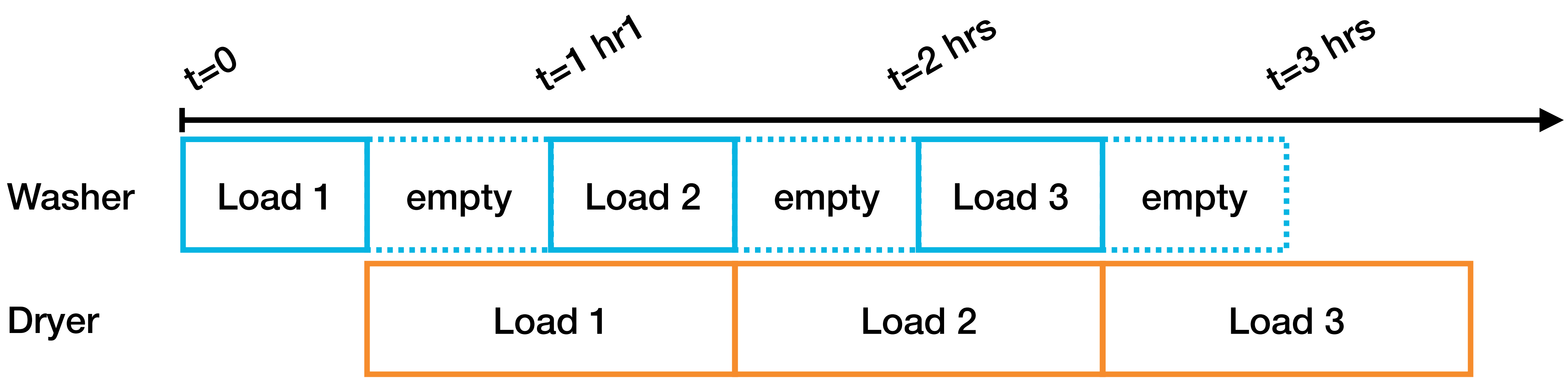
A real life example

Trigger systems rely on **pipelines** to perform *complicated reconstruction* tasks while handling a stream of *continuous inputs*.

They enable systems with *long latency*, *short initiation interval.*

Example: laundry in my apartment building:

1 washer (runs 30min) and 1 dryer (runs 1hr).



What are the ii, latency of the washer? The dryer? The entire system?

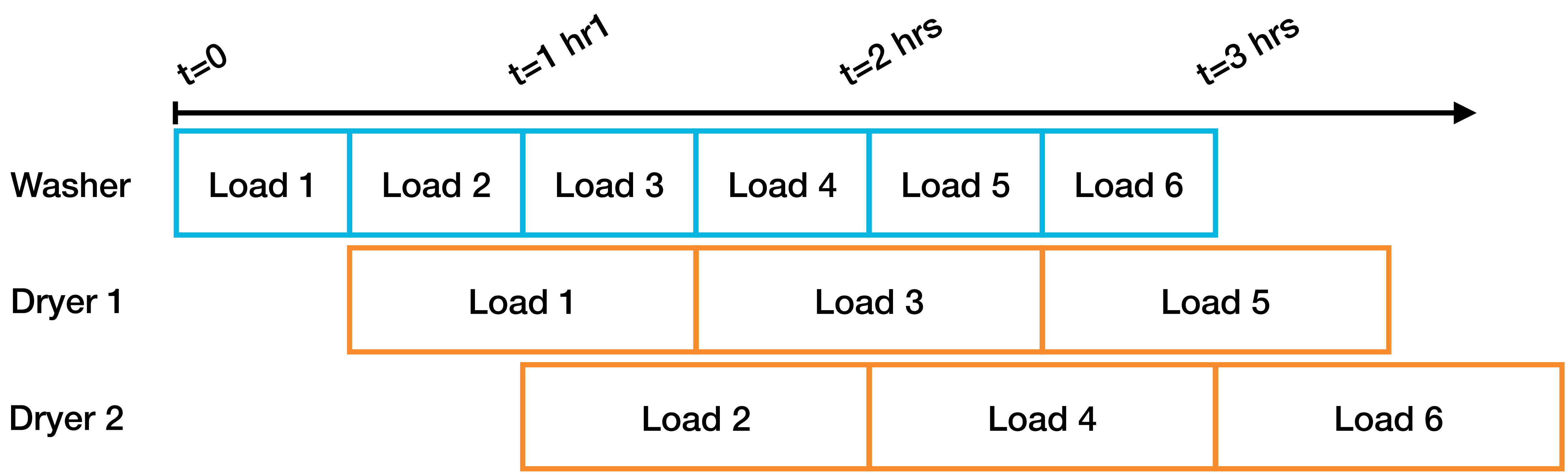
A real life example

Trigger systems rely on **pipelines** to perform *complicated reconstruction* tasks while handling a stream of *continuous inputs*.

They enable systems with *long latency*, *short initiation interval.*

Example: laundry in my apartment building:

1 washer (runs 30min) and 1 dryer (runs 1hr).

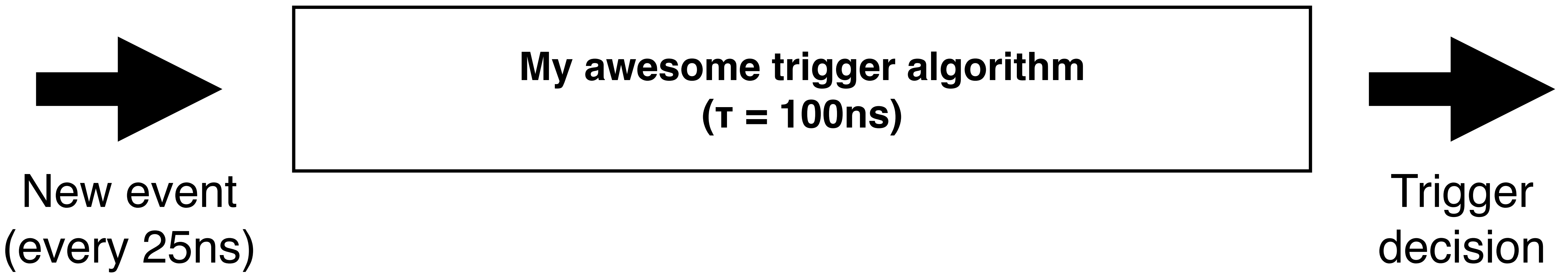


What are the ii, latency of the washer? The dryer? The entire system?

Pipelining

Trigger systems rely on **pipelines** to perform *complicated reconstruction tasks* while handling a stream of *continuous inputs*.

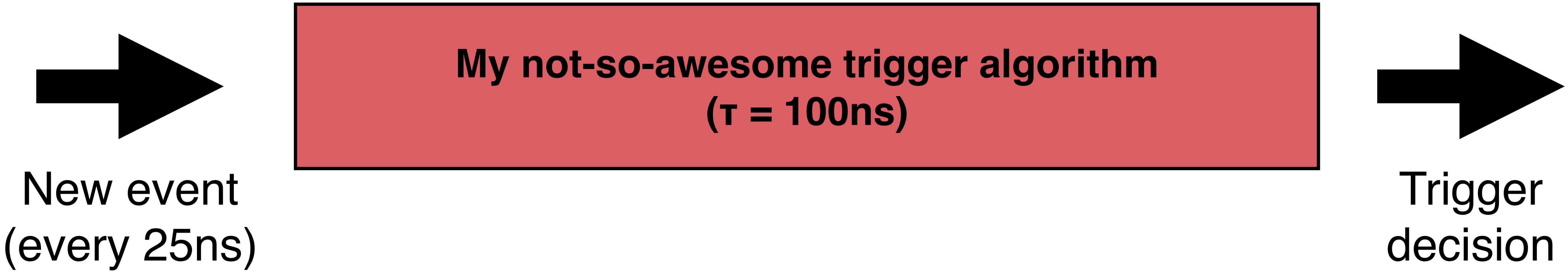
They enable systems with *long latency*, *short initiation interval.*



Pipelining

Trigger systems rely on **pipelines** to perform *complicated reconstruction* tasks while handling a stream of *continuous inputs*.

They enable systems with *long latency*, *short initiation interval.*

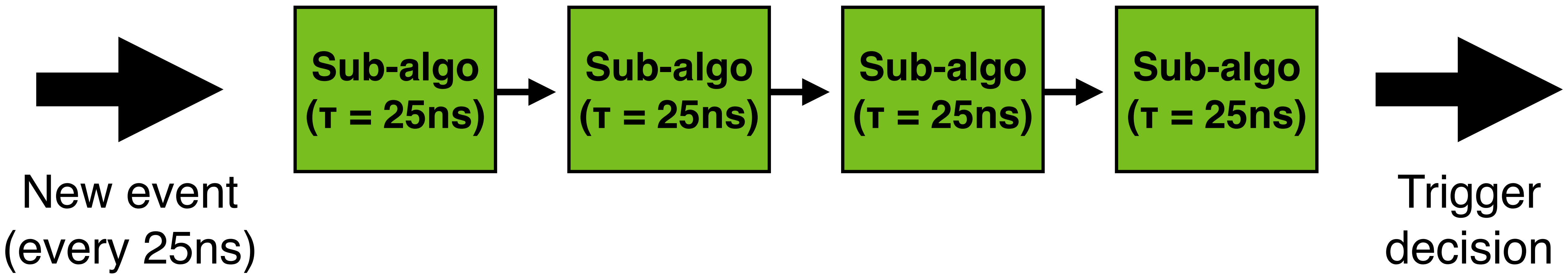


The block is still processing when a new event arrives!

Pipelining

Trigger systems rely on **pipelines** to perform *complicated reconstruction* tasks while handling a stream of *continuous inputs*.

They enable systems with *long latency*, *short initiation interval.*



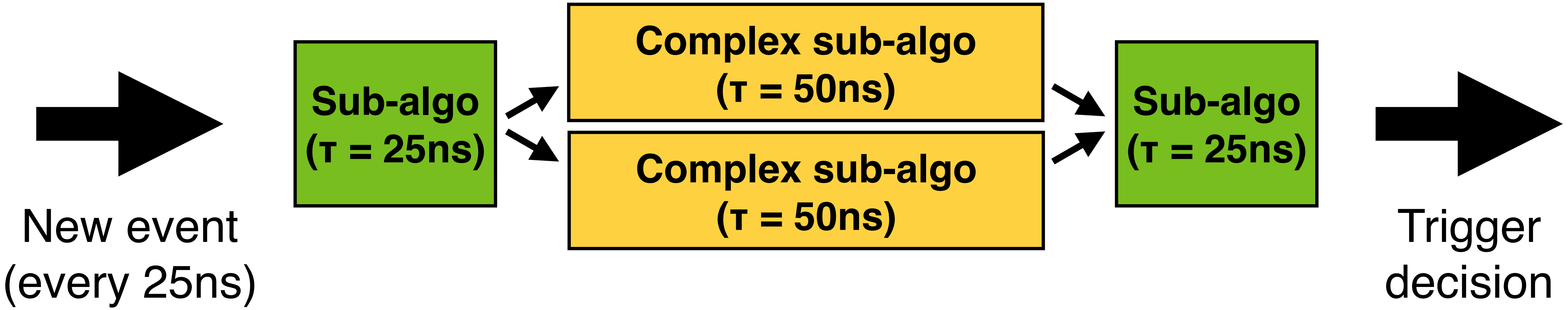
Modular logic can process many events at once

(and again, intermediate buffers ease synchronization)

Pipelining

Trigger systems rely on **pipelines** to perform *complicated reconstruction tasks* while handling a stream of *continuous inputs*.

They enable systems with *long latency*, *short initiation interval.*

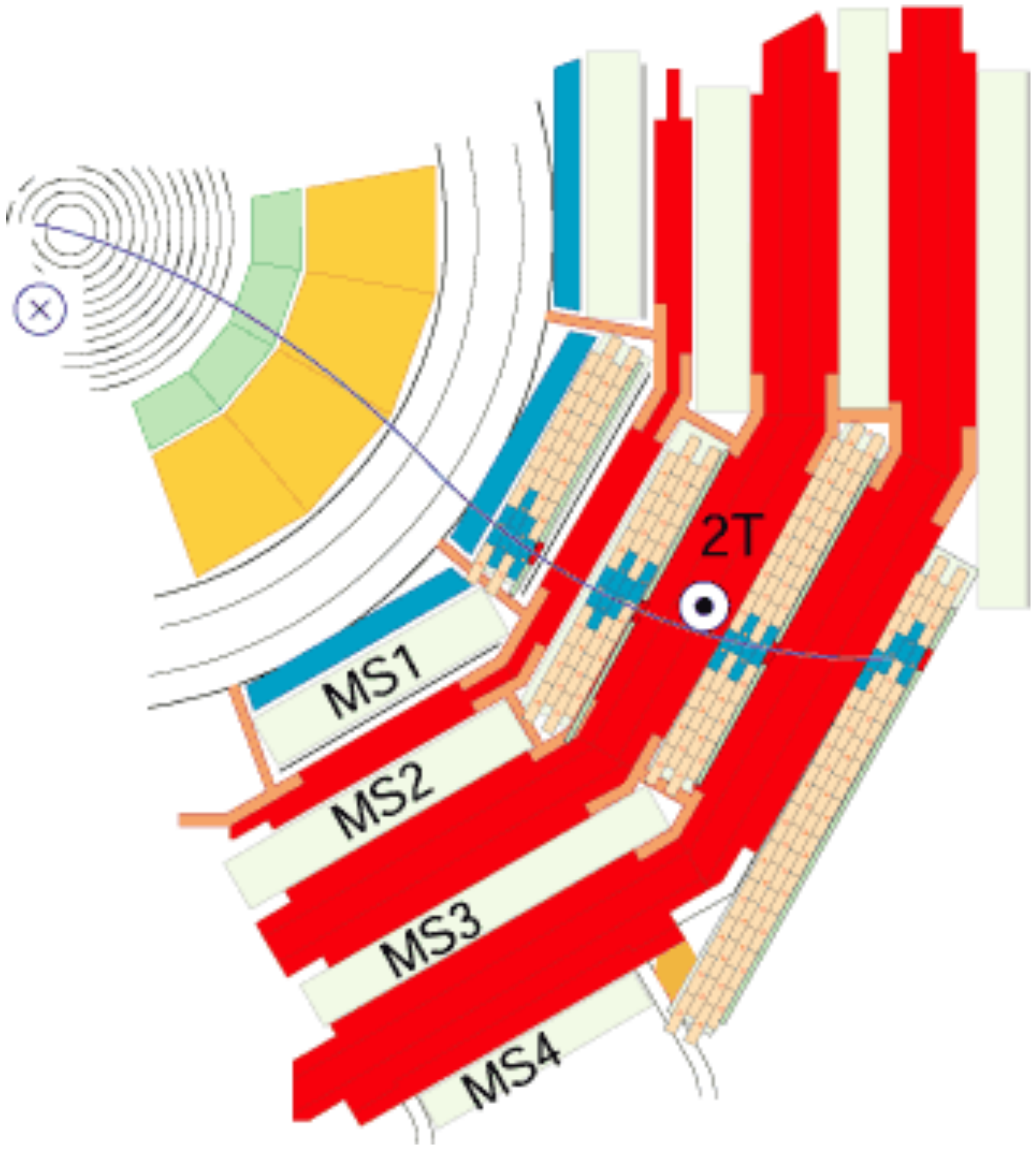


Sometimes this is inconvenient or impossible → **parallel processors**

(More complex building blocks, at the expense of more resources)

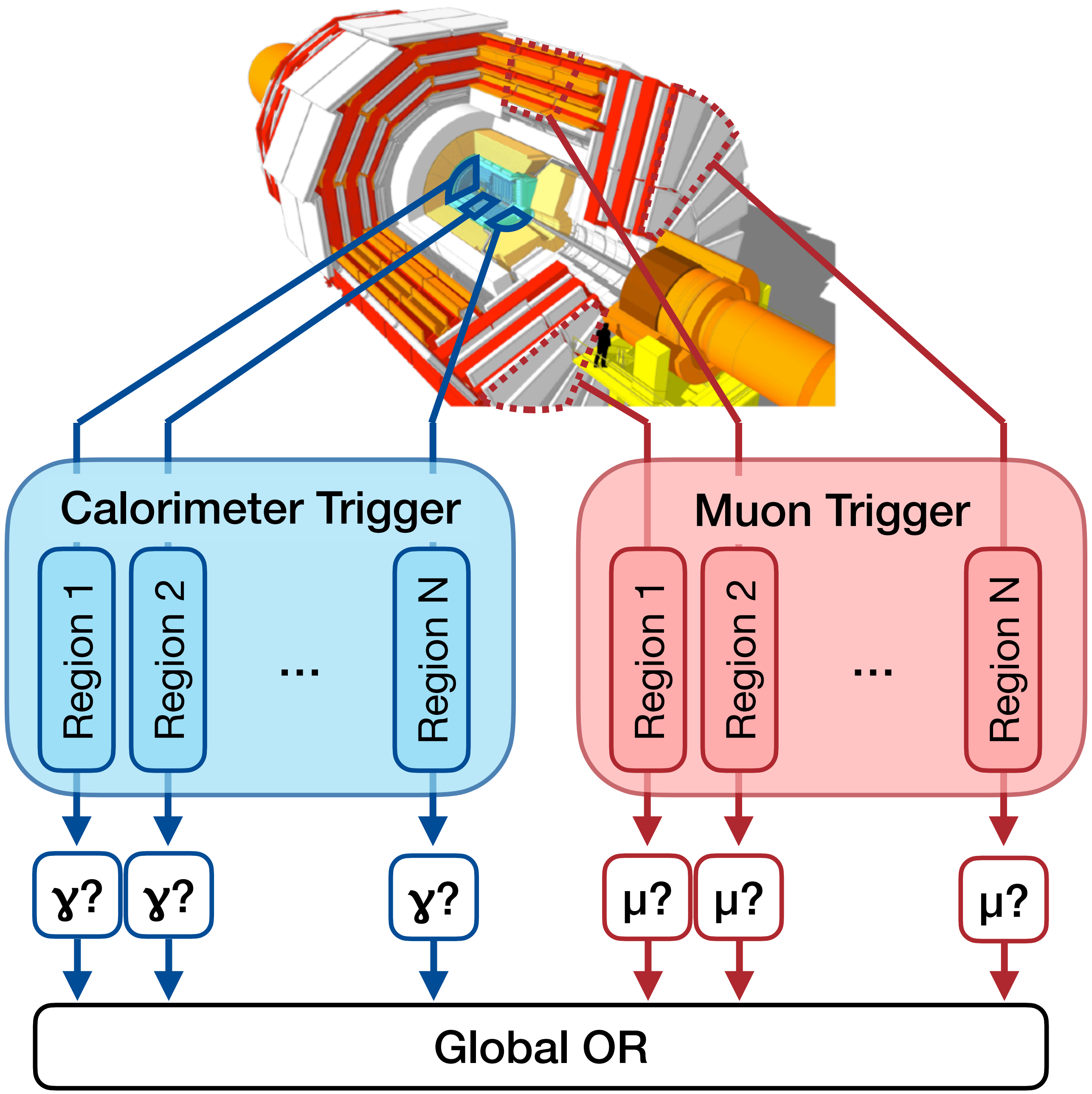
Regional processing

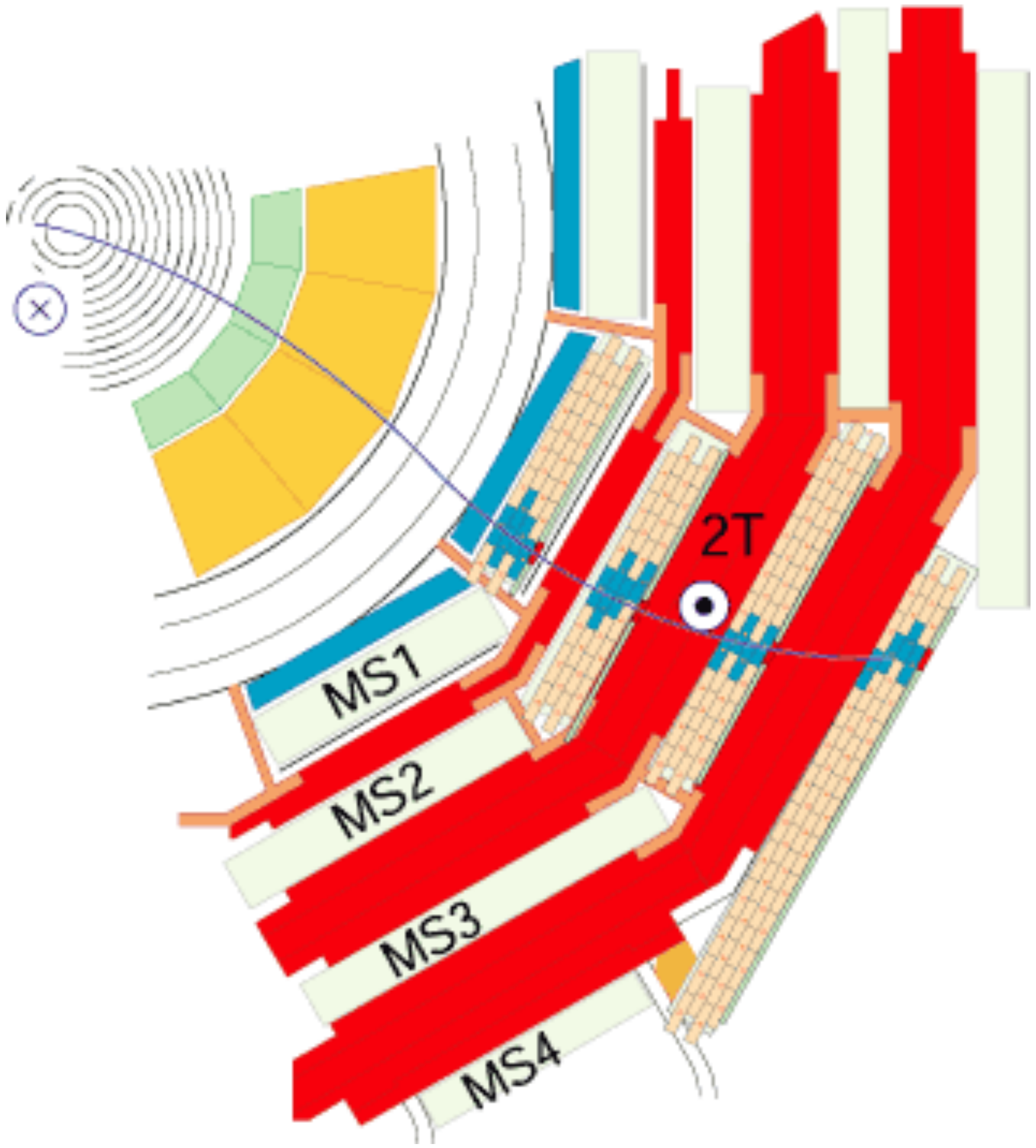
Particle reconstruction is **an** **inherently local** **task**
→ process parallel regions



Regional processing

Particle reconstruction is **an** **inherently local** **task**
→ process parallel regions

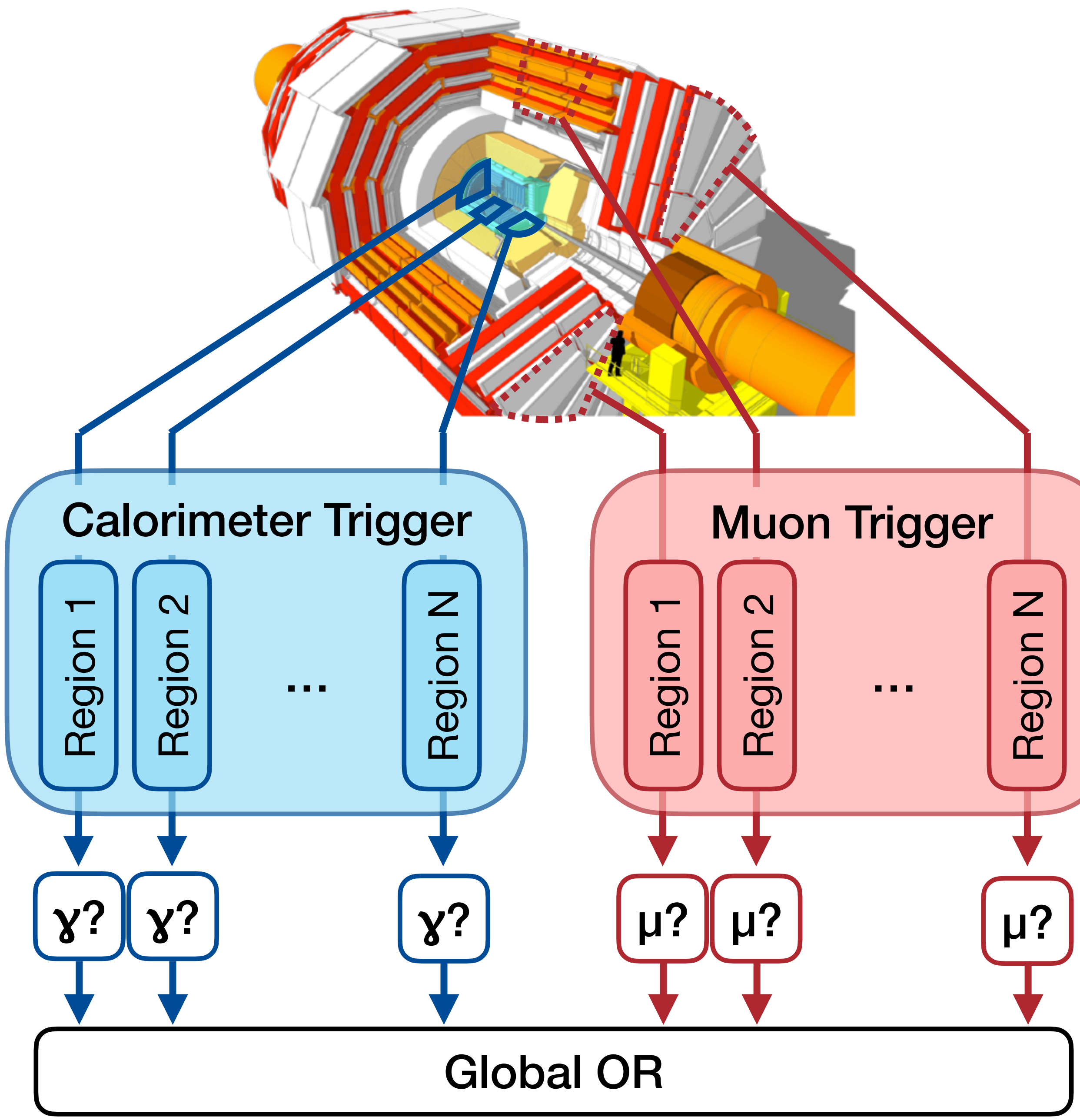


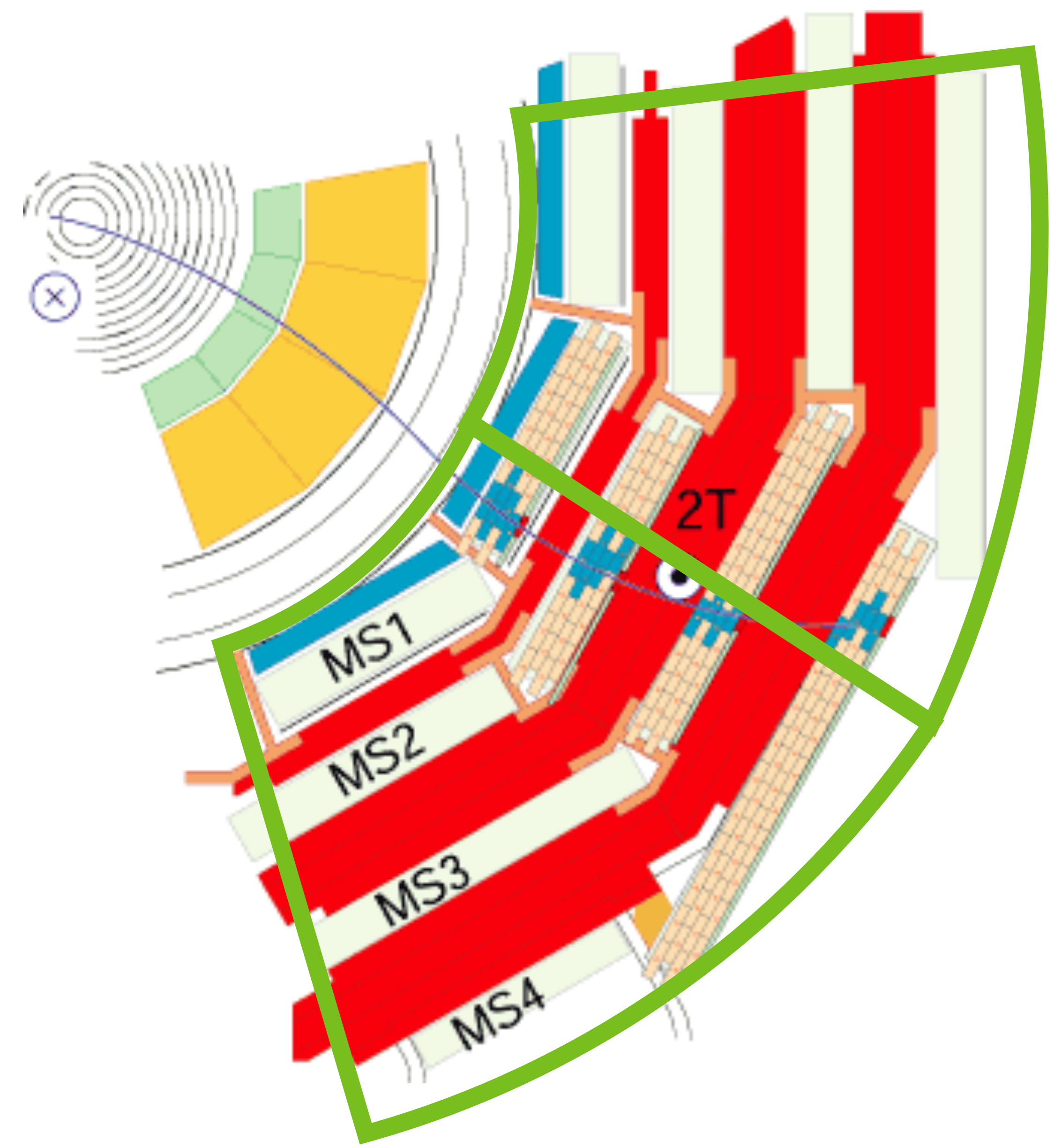


No “event building” necessary!

Regional processing

Particle reconstruction is **an inherently local task**
→ process parallel regions

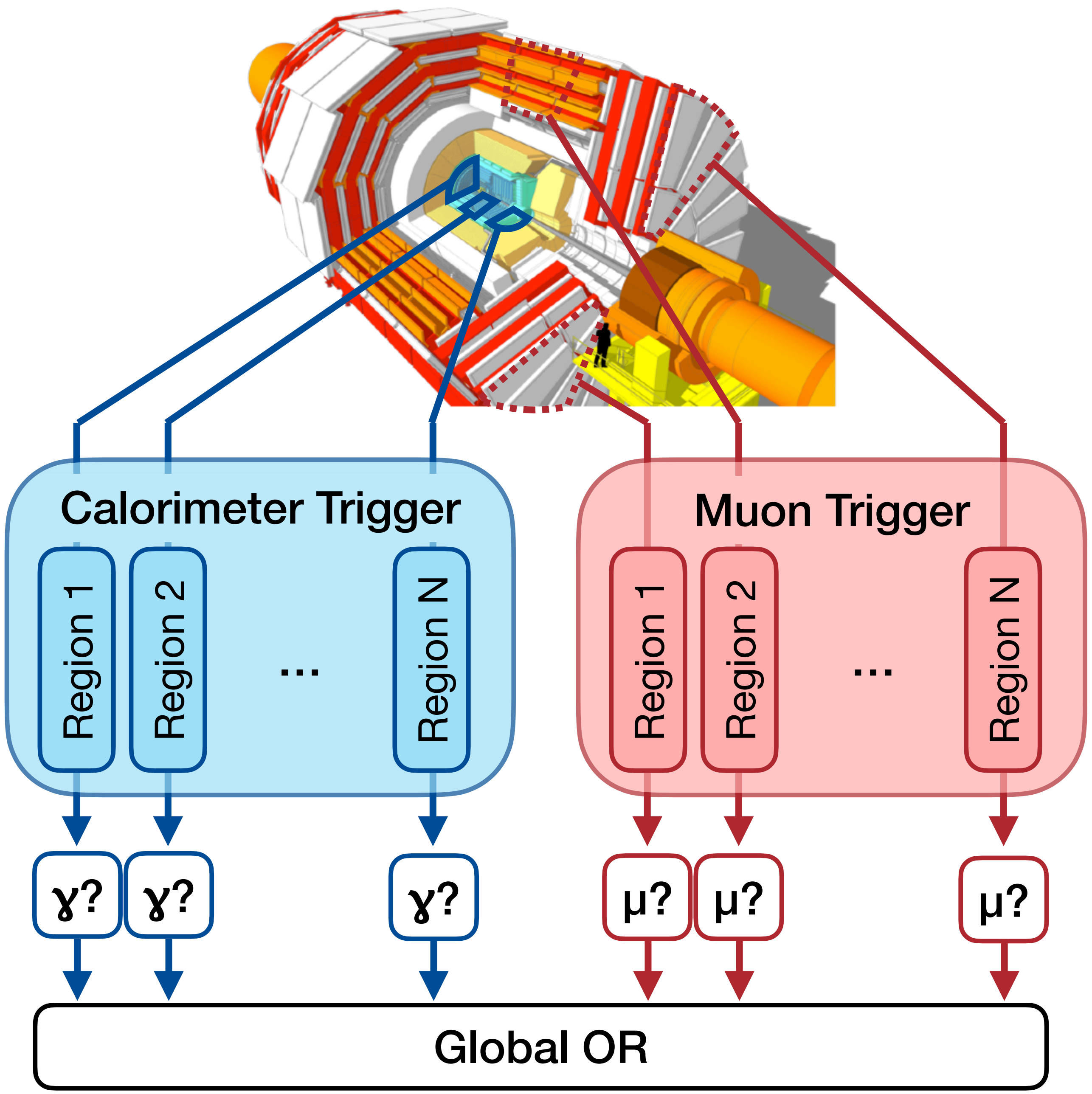


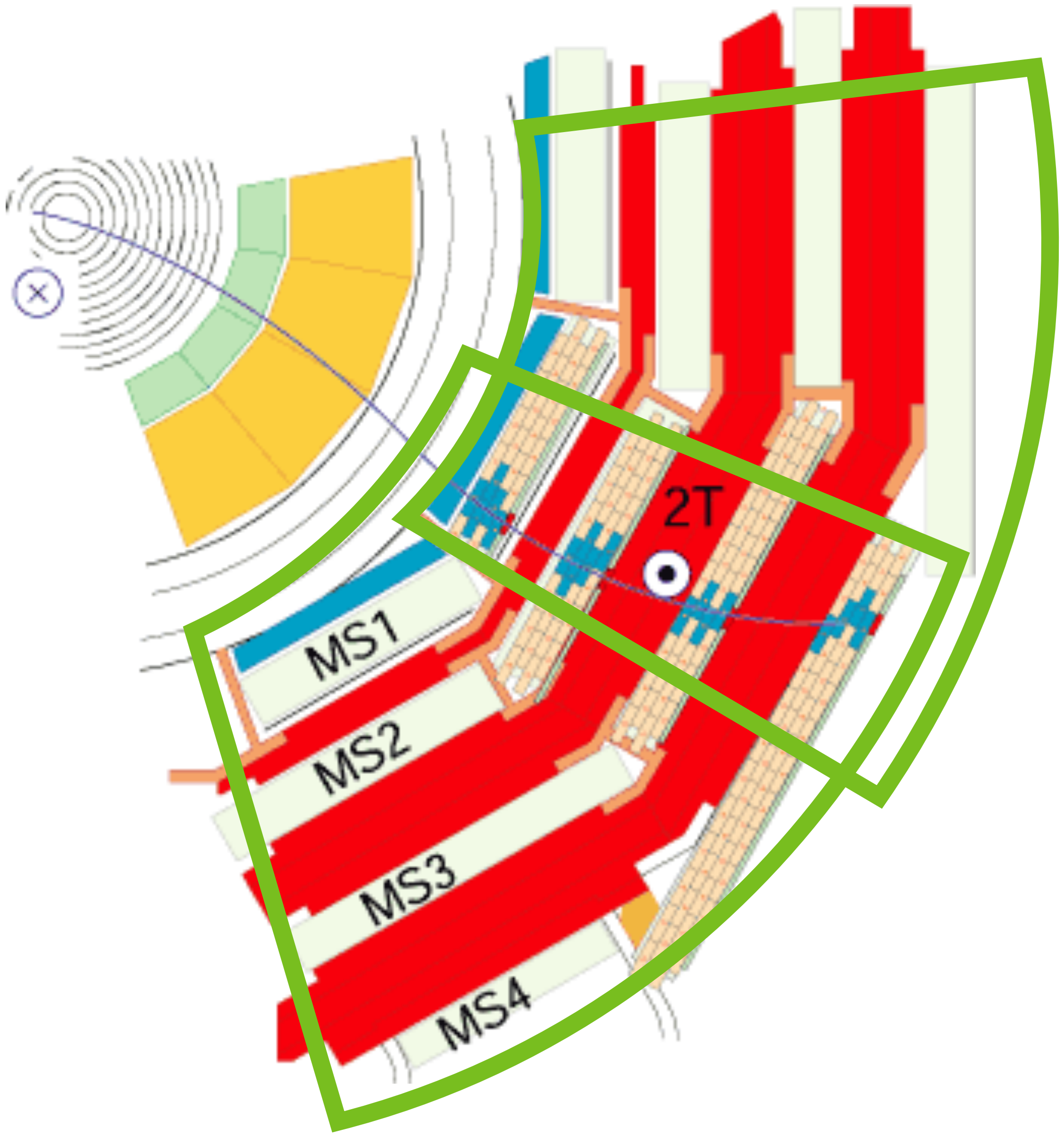


Caveat: some data-sharing always needed. How to deal with overlaps?

Regional processing

Particle reconstruction is **an inherently local task**
→ process parallel regions

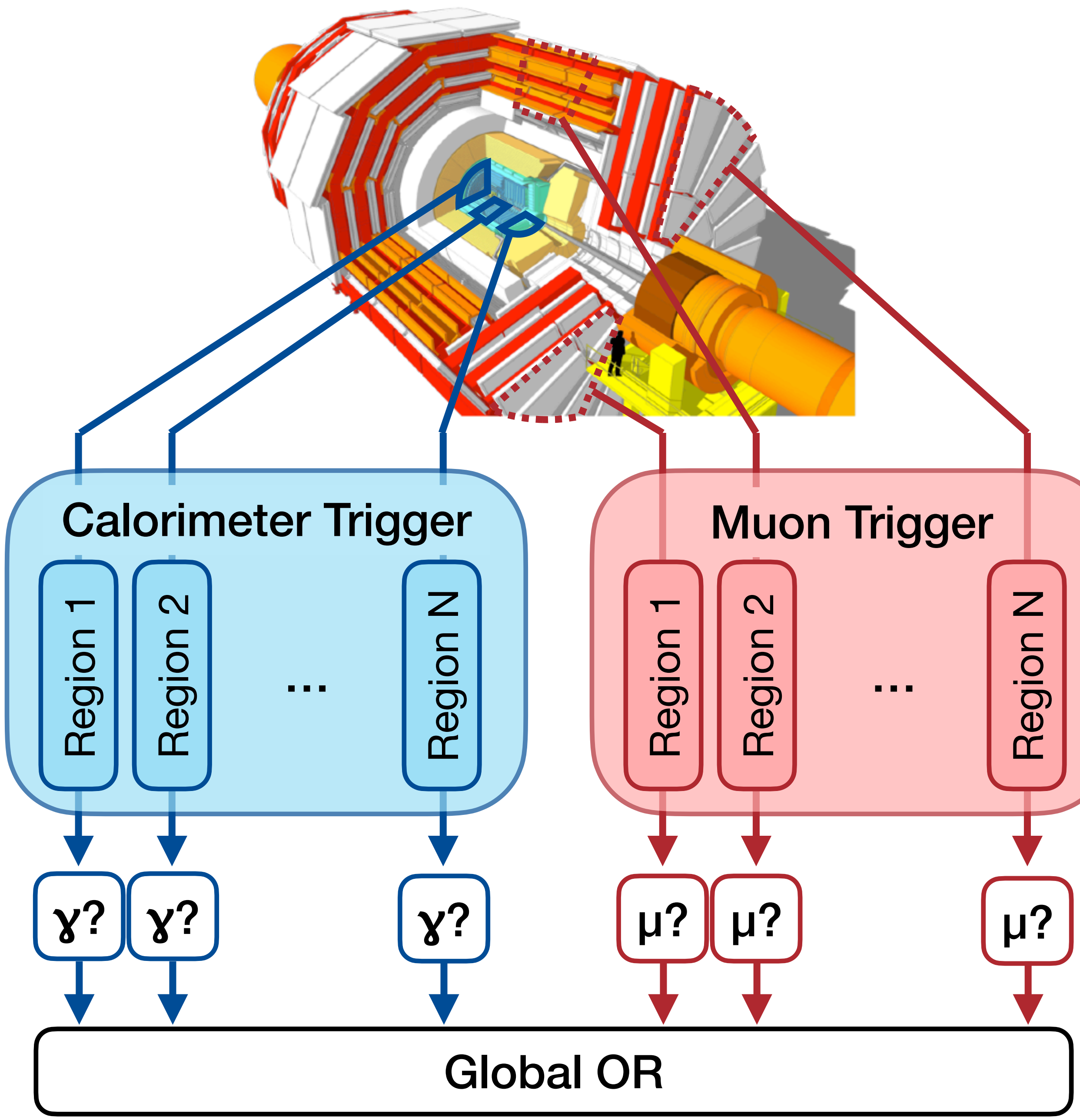


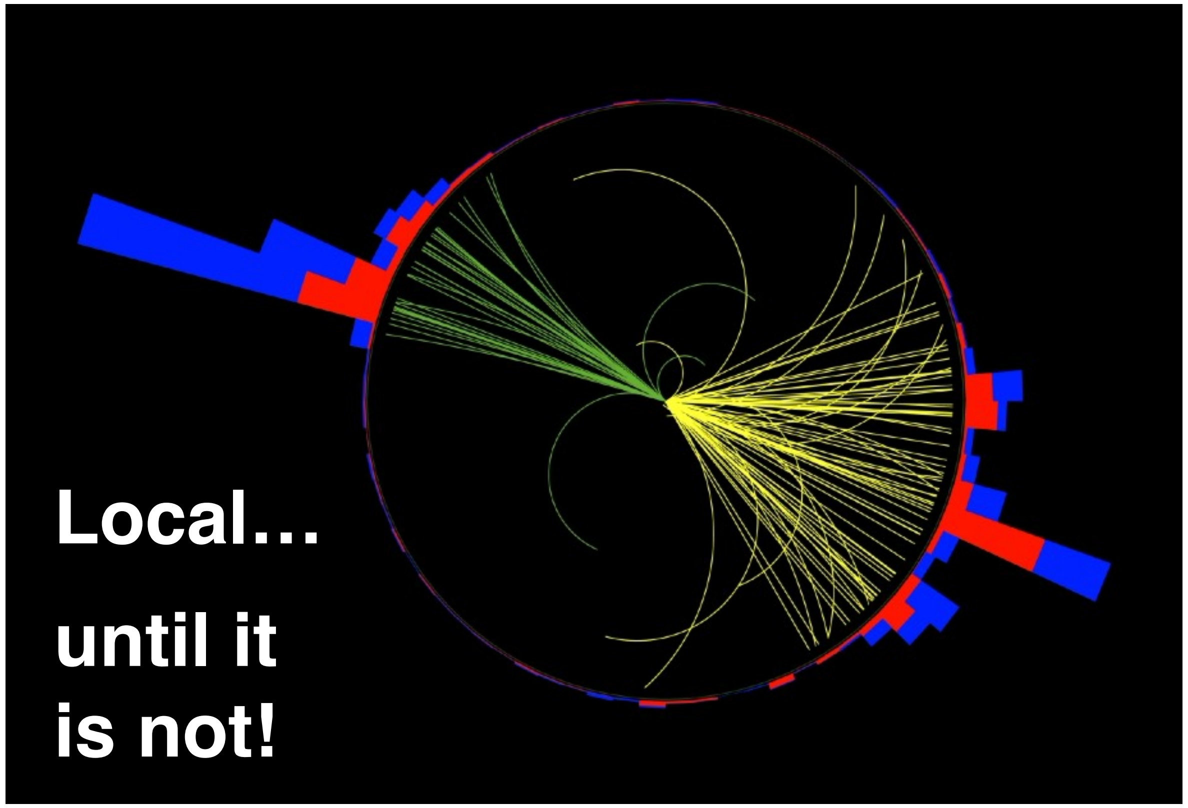


Caveat: some data-sharing always needed. How to deal with overlaps?

Regional processing

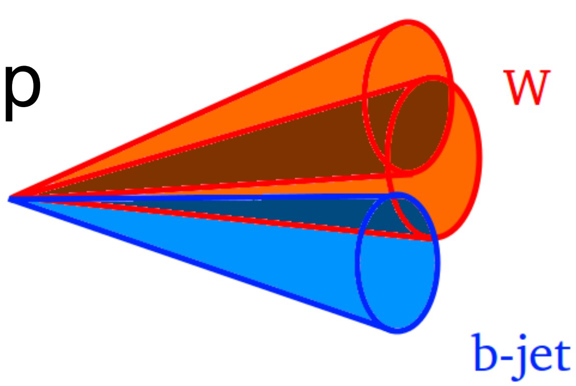
Particle reconstruction is an inherently local task
→ process parallel regions





Particle jets occupy 1/2 detector!

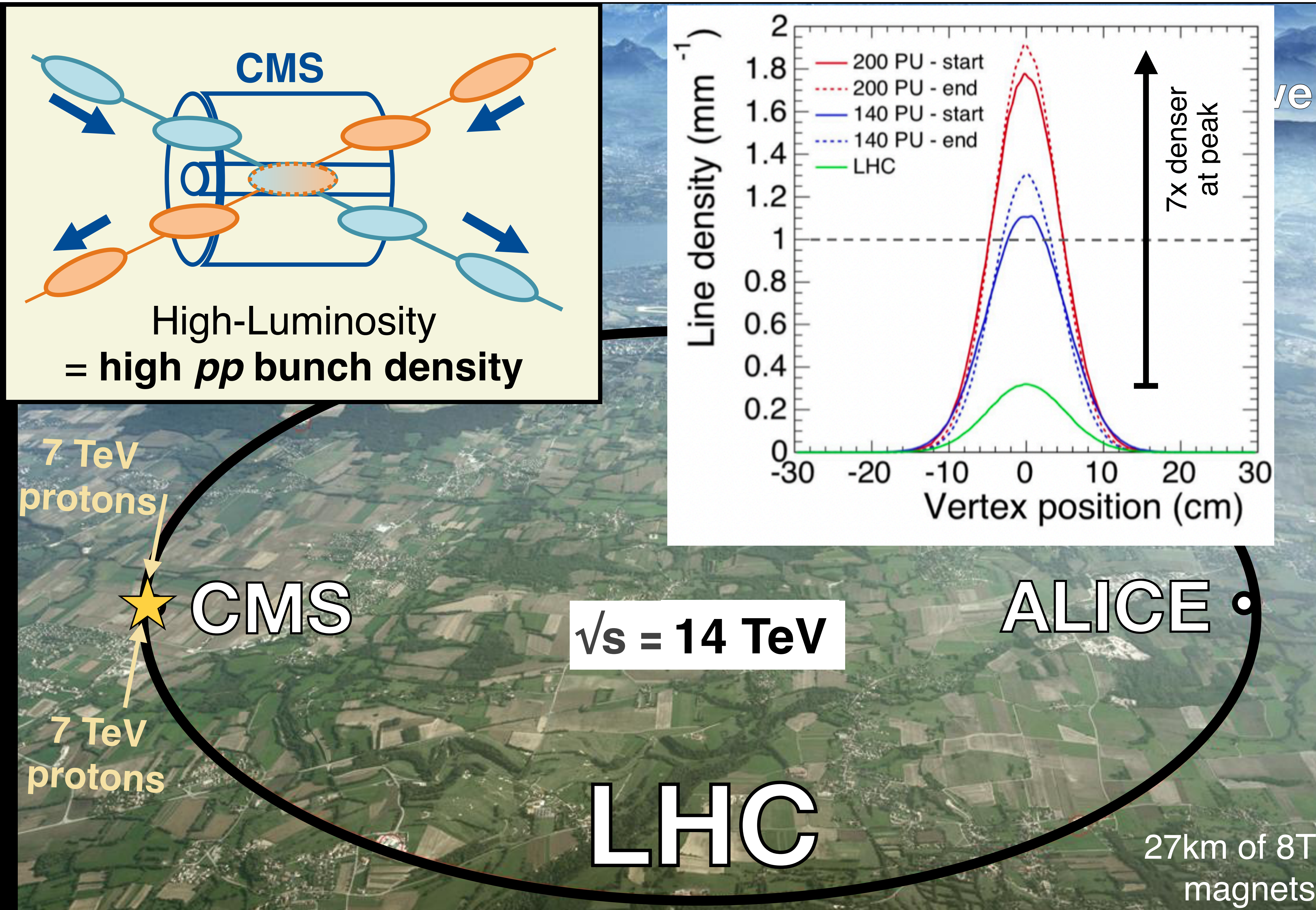
E.g. a decaying top quark with large Lorentz boost



Plan

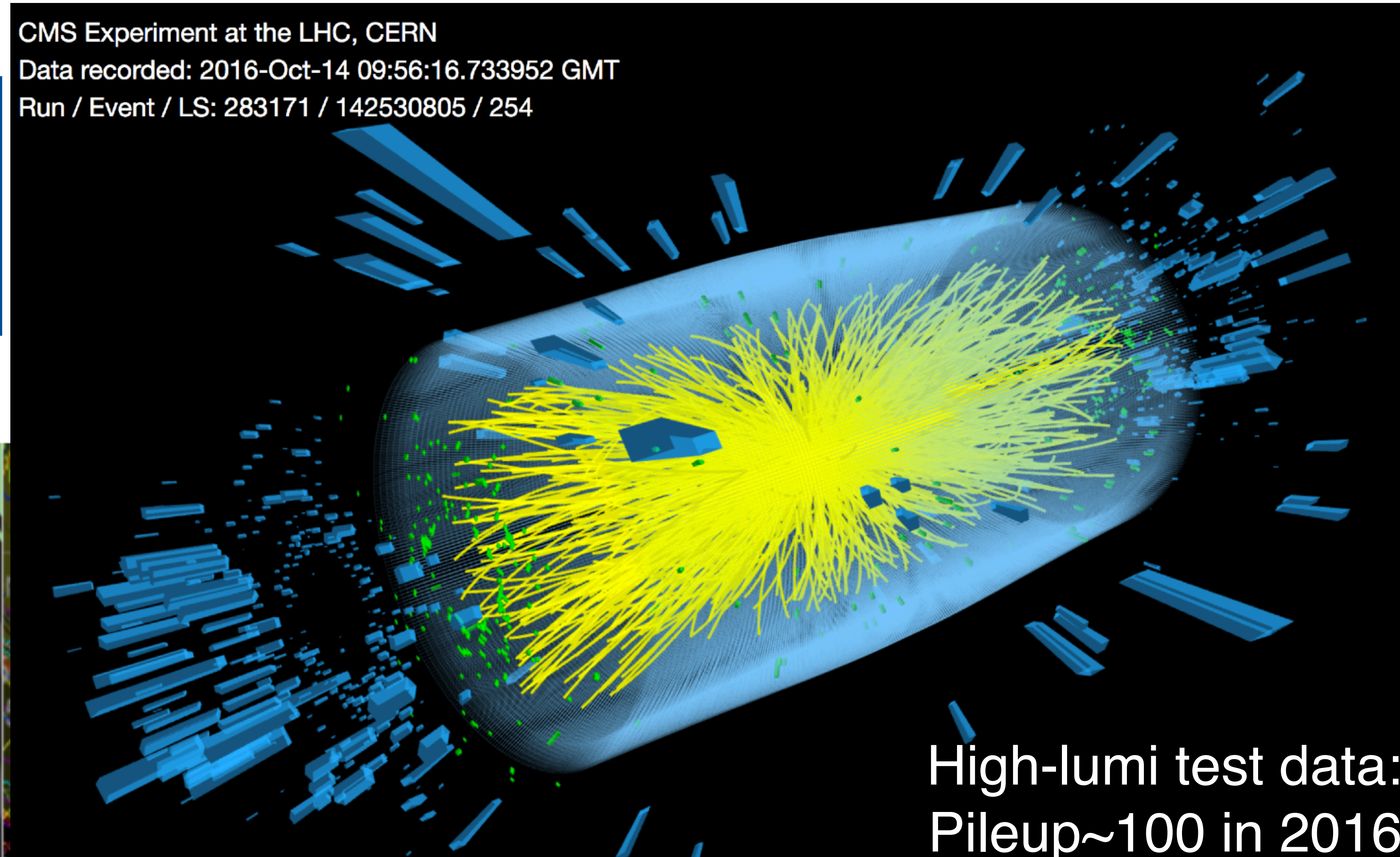
- **Yesterday afternoon: Concepts**
  - Introduction and physics goals of the LHC Experiments
  - The role of the trigger and DAQ systems
  - DAQ concepts: simple toy model → complex systems
- **This morning: Details and Implementations**
  - Trigger concepts: from hardware to architecting a system
  - **The challenge of high luminosity, and the upgrades**
  - ATLAS, CMS hardware triggers, and their evolution
  - Software triggers, and new analysis paradigms
  - Wrap up

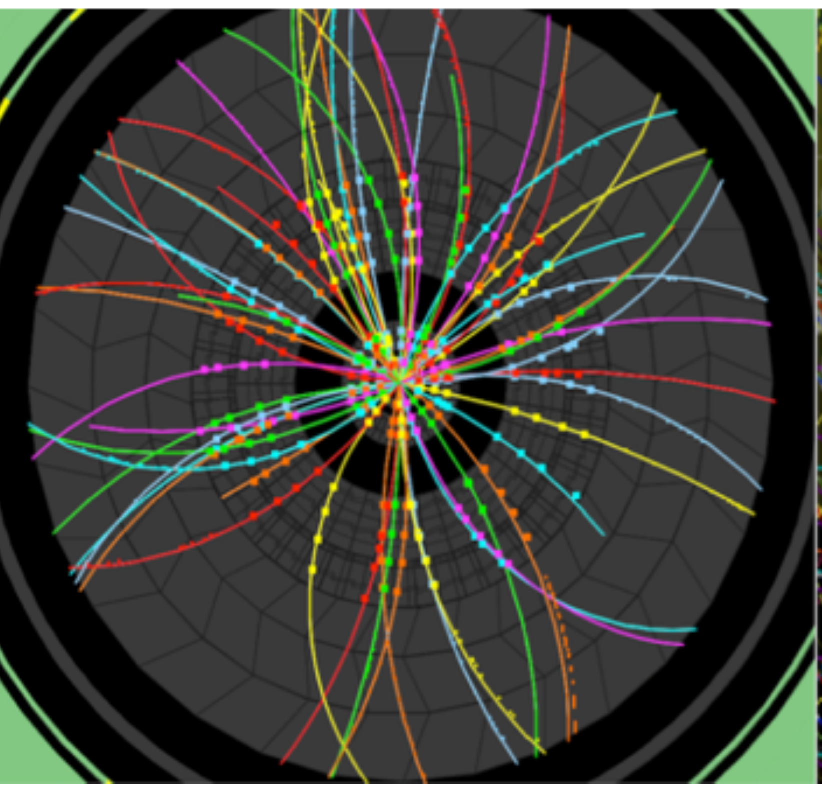
Towards the high-luminosity era



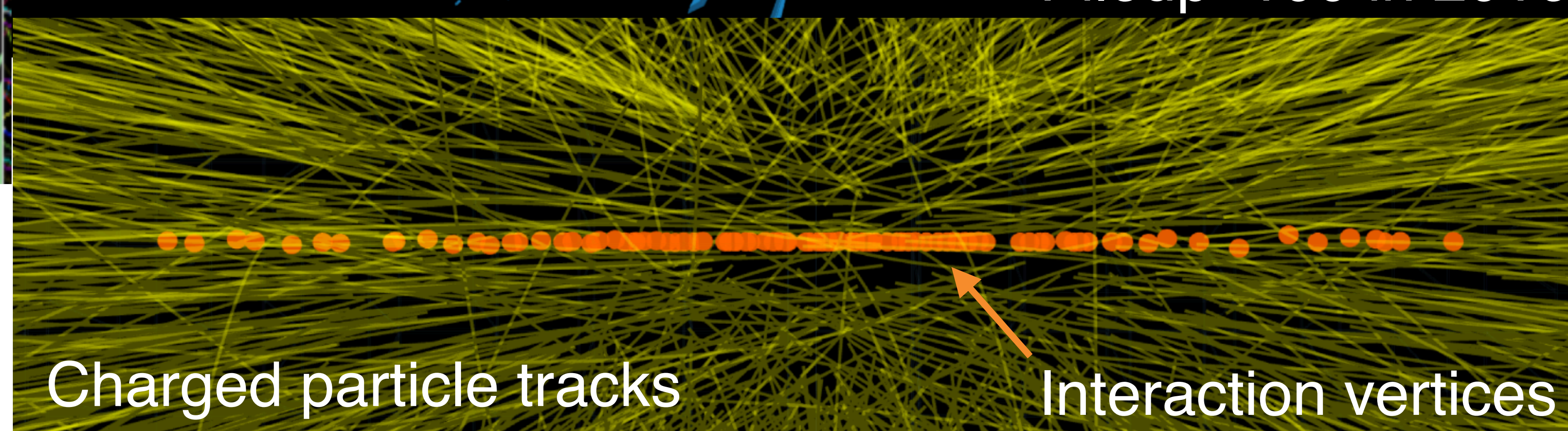
Basic challenge of High-Luminosity

Must disentangle decay products of 200 overlapping collisions.



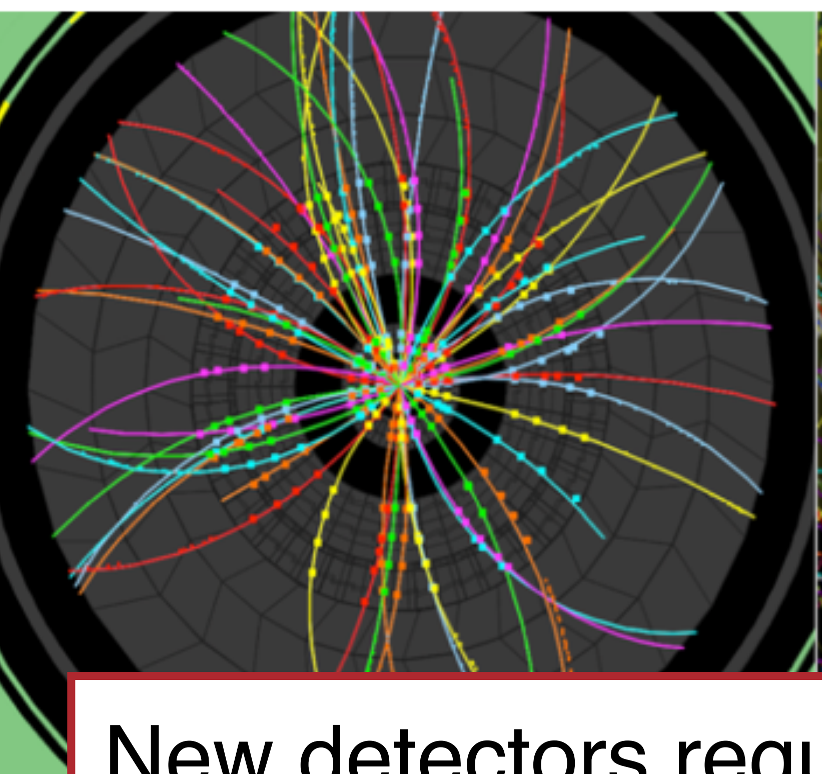


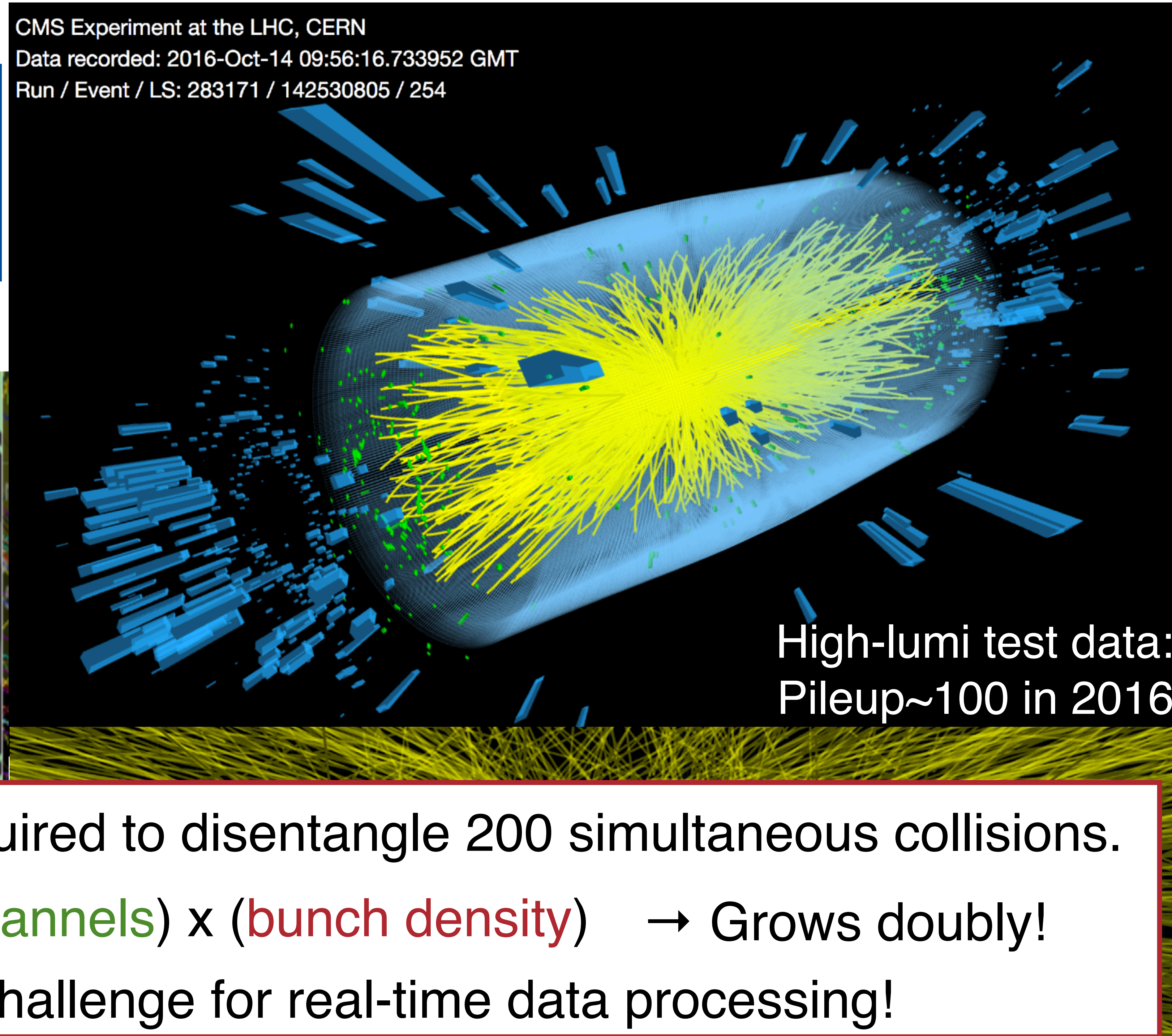
Typical collision at the LHC start (Pileup=2)



Basic challenge of High-Luminosity

Must disentangle decay products of 200 overlapping collisions.



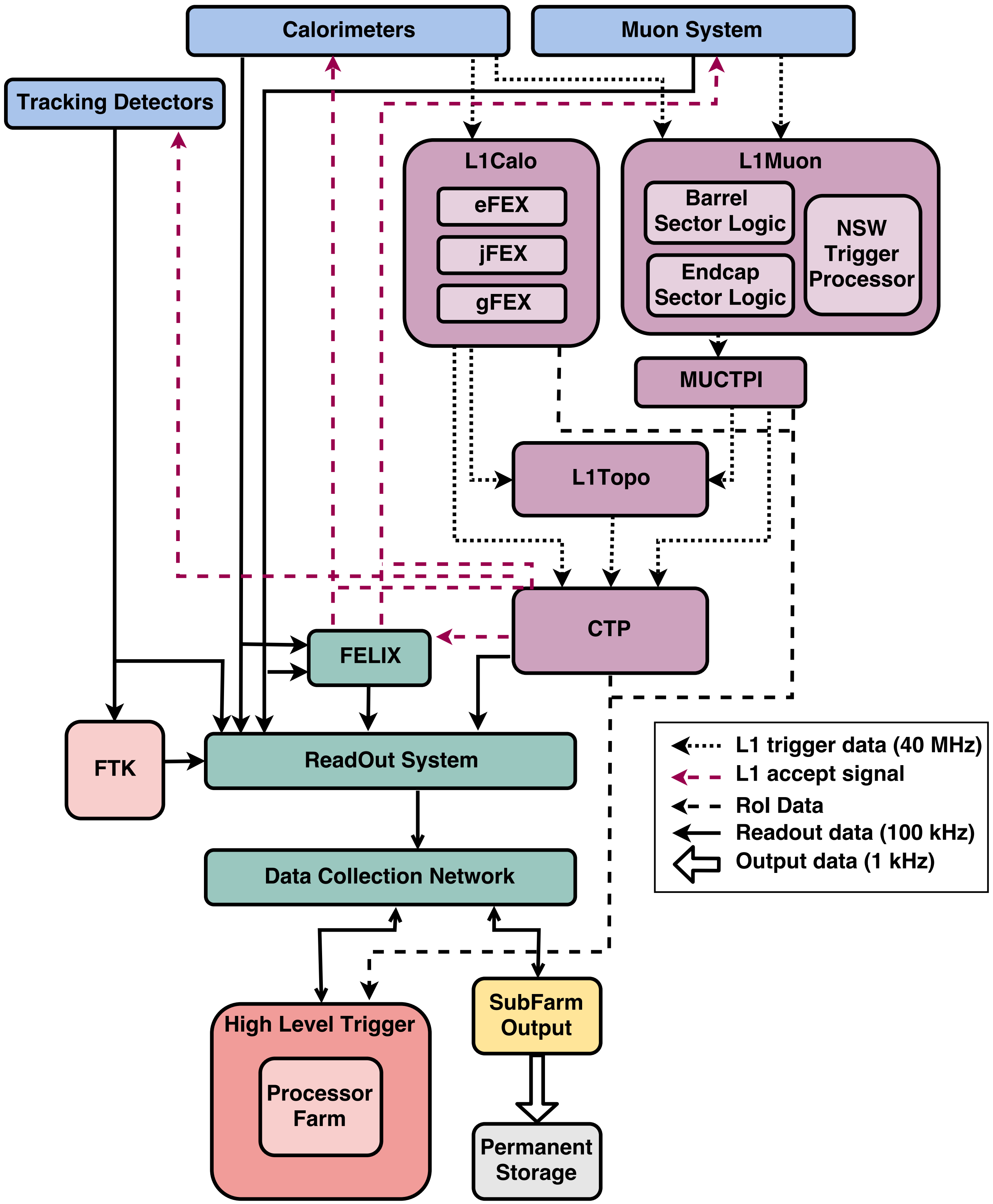


New detectors required to disentangle 200 simultaneous collisions.

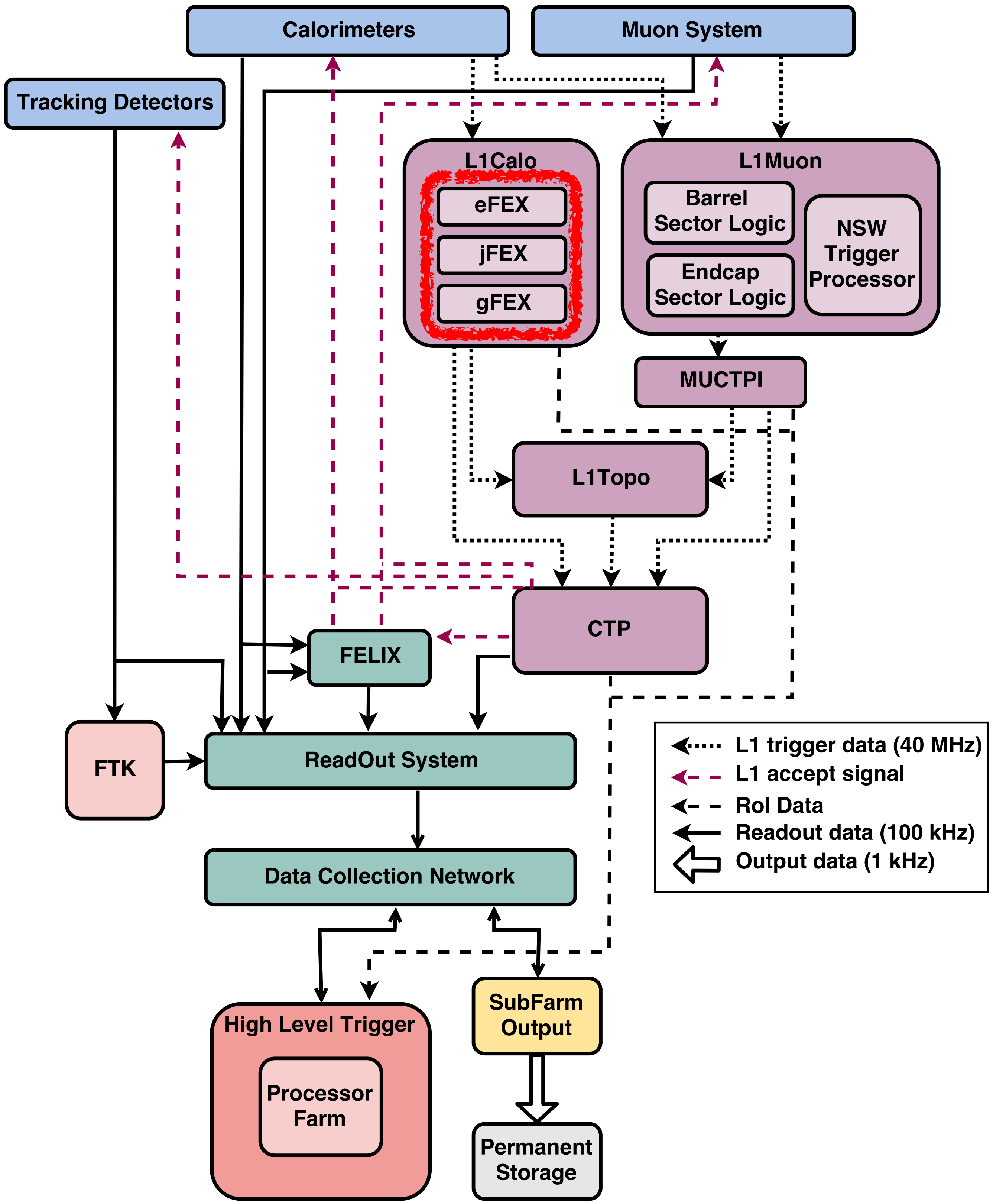
Data rate ~ (# channels) x (bunch density) → Grows doubly!

Huge challenge for real-time data processing!

The ATLAS Trigger System



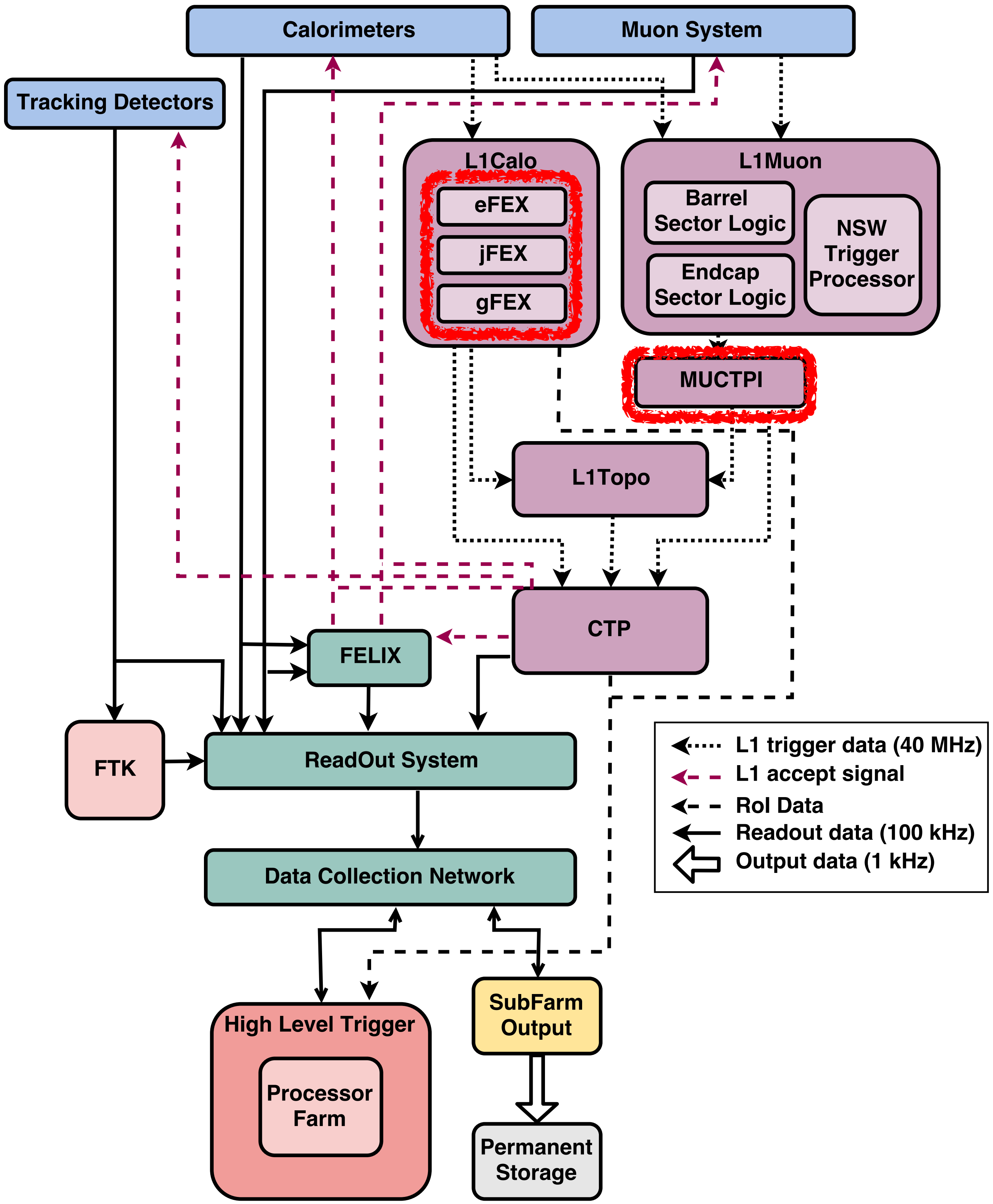
The ATLAS Trigger System



Sub-engines for

- Electrons and photons
- Hadronic jets (R=0.4, quark/gluon)
- Large-R jets (hadronic W/Z/h)

The ATLAS Trigger System

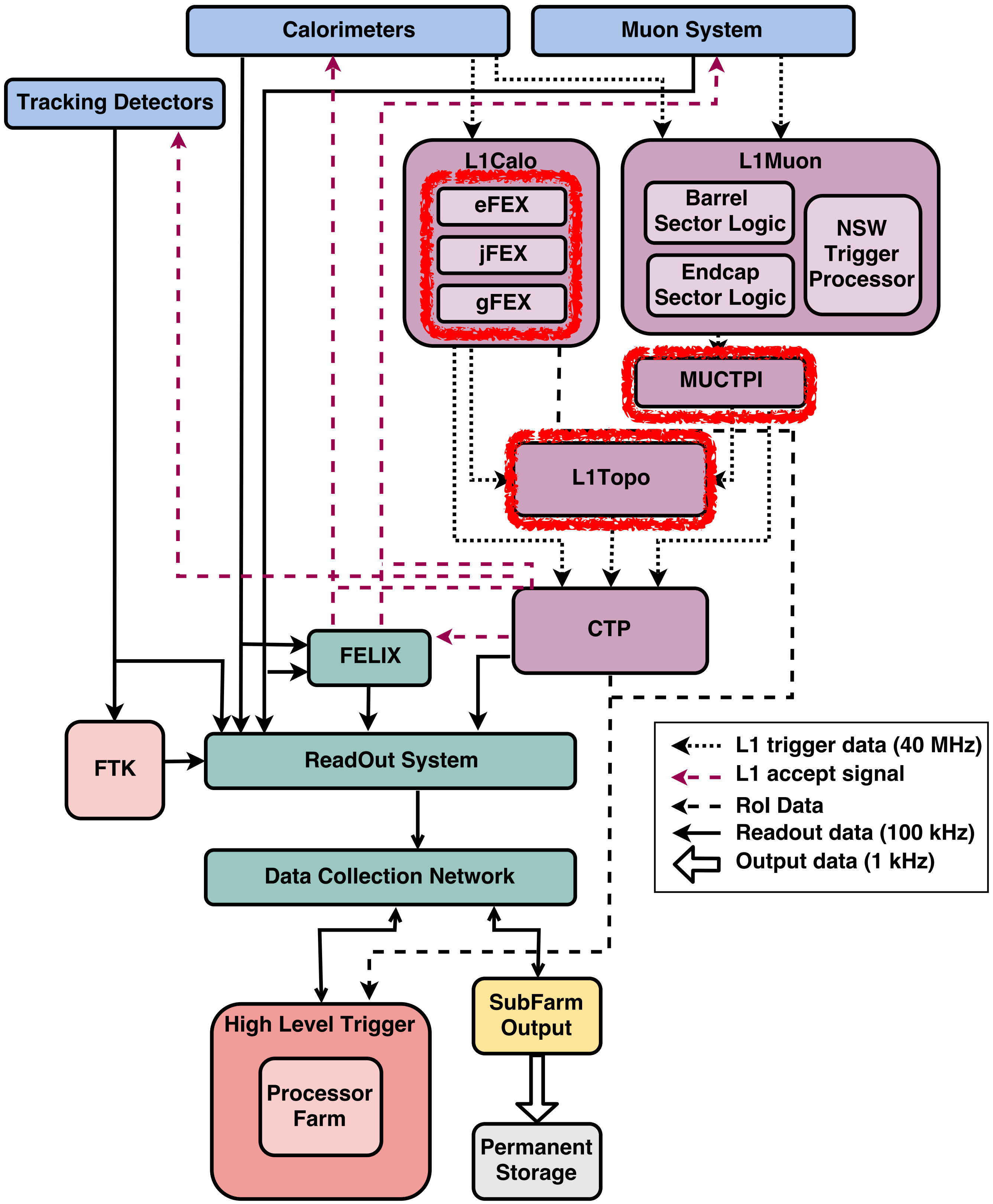


Sub-engines for

- Electrons and photons
- Hadronic jets (R=0.4, quark/gluon)
- Large-R jets (hadronic W/Z/h)

“Ghost-busting” muons reconstructed in two overlapping detector segments

The ATLAS Trigger System



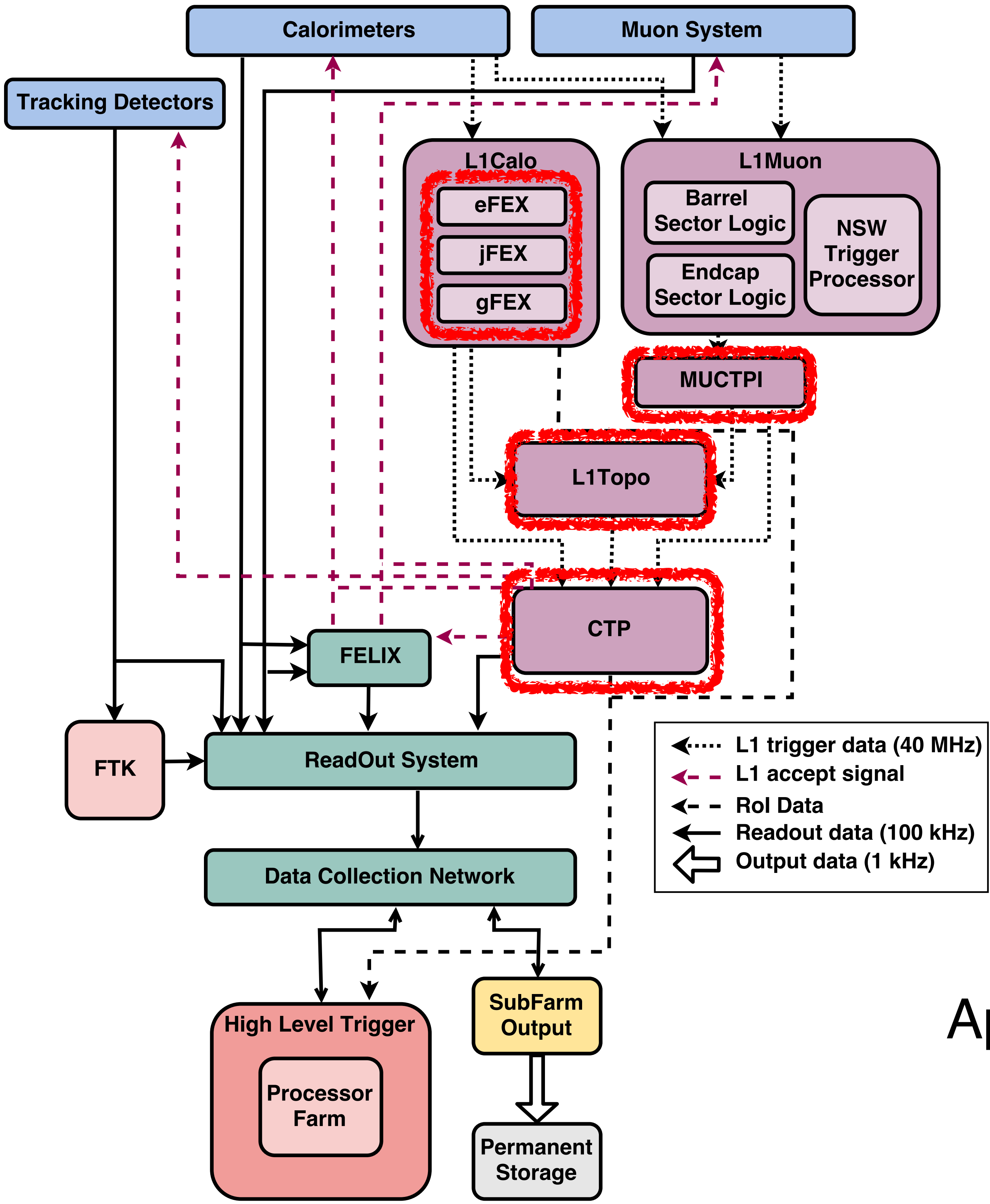
Sub-engines for

- Electrons and photons
- Hadronic jets (R=0.4, quark/gluon)
- Large-R jets (hadronic W/Z/h)

"Ghost-busting" muons reconstructed in two overlapping detector segments

Correlating multiple objects (e.g. m(μμ), dR(j,j), energy sums,…)

The ATLAS Trigger System



Sub-engines for

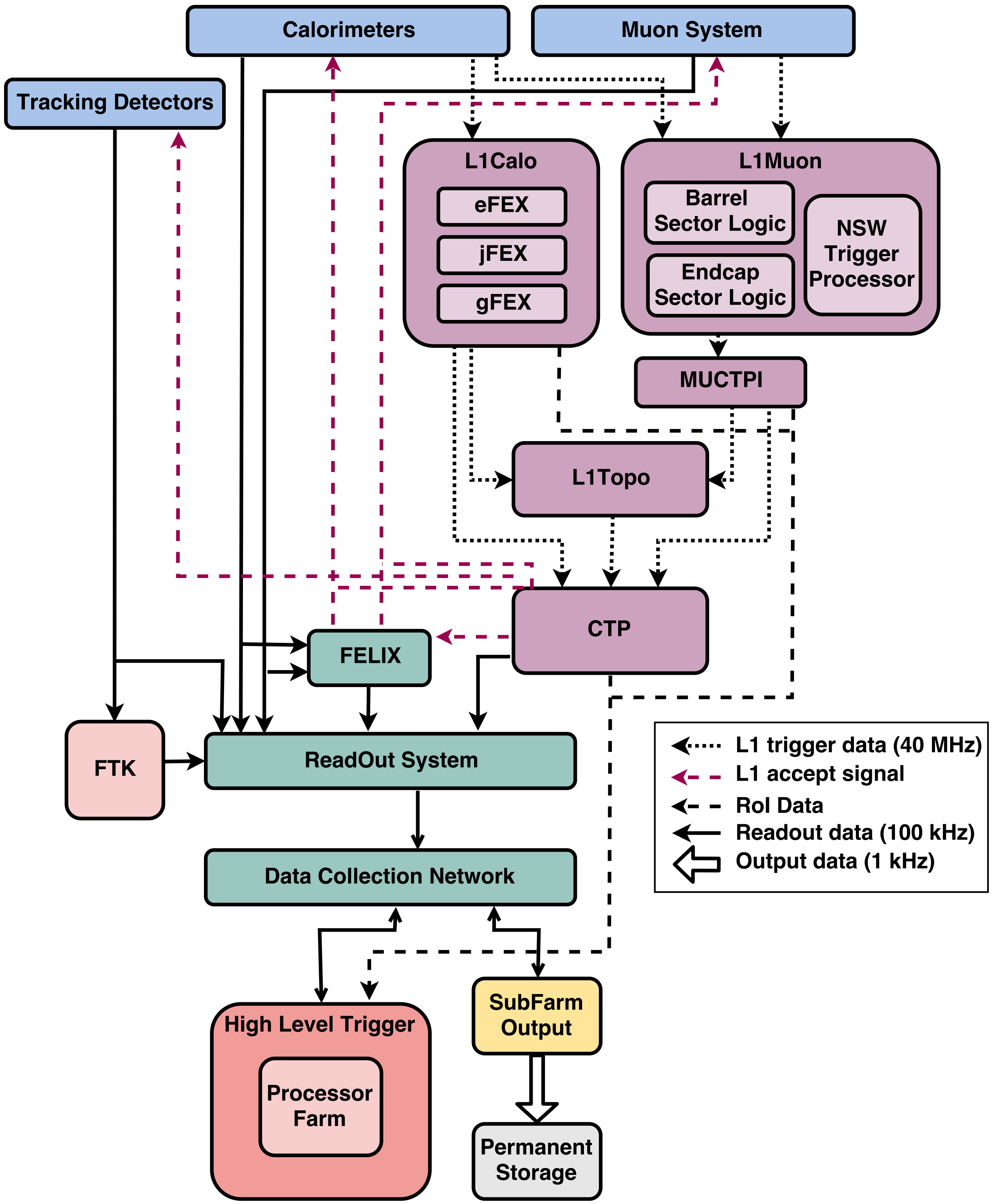
- Electrons and photons
- Hadronic jets (R=0.4, quark/gluon)
- Large-R jets (hadronic W/Z/h)

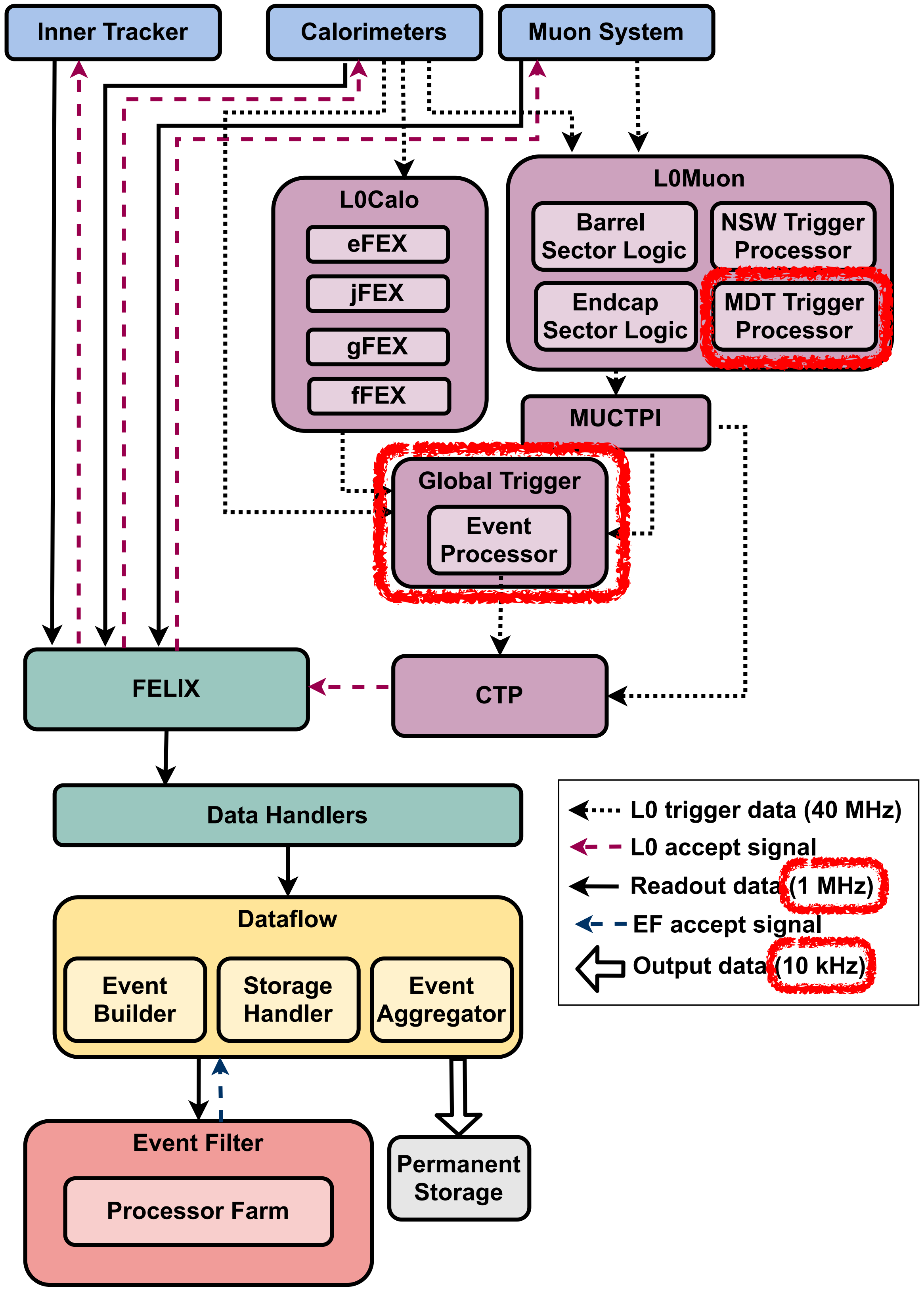
“Ghost-busting” muons reconstructed in two overlapping detector segments

Correlating multiple objects (e.g. m(μμ), dR(j,j), energy sums,…)

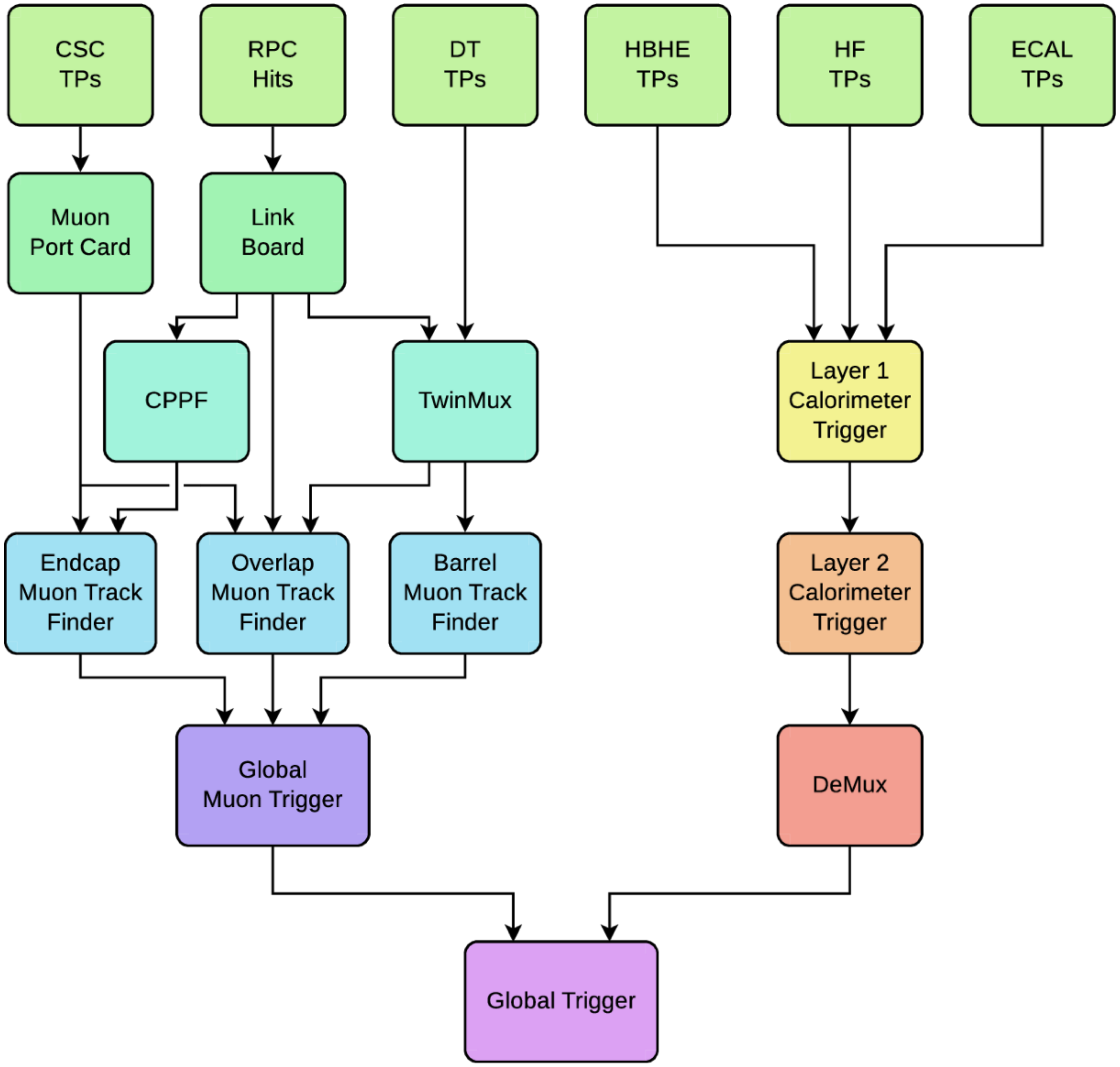
Apply the dead time rules

The ATLAS Trigger System

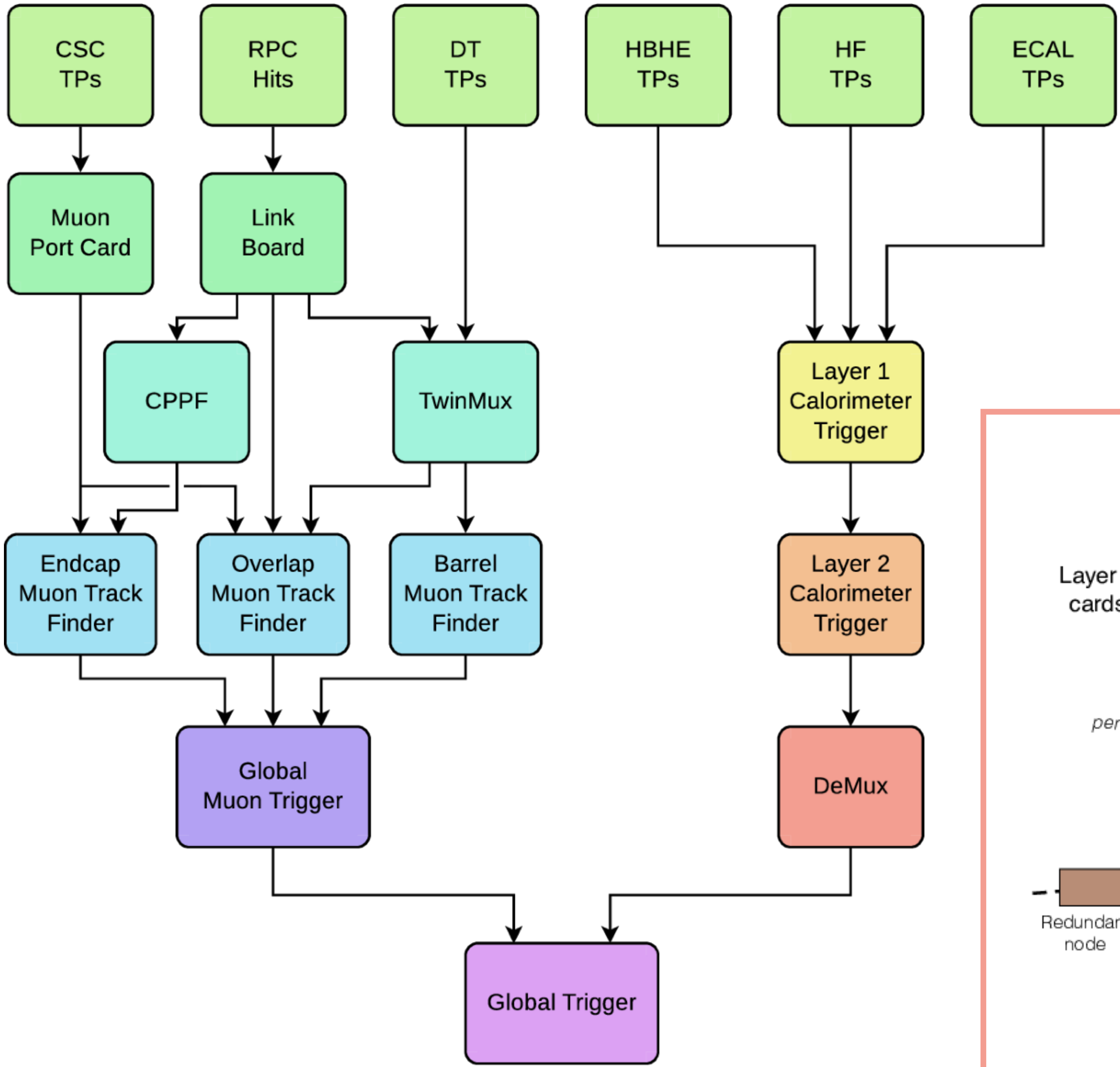




The CMS Trigger System



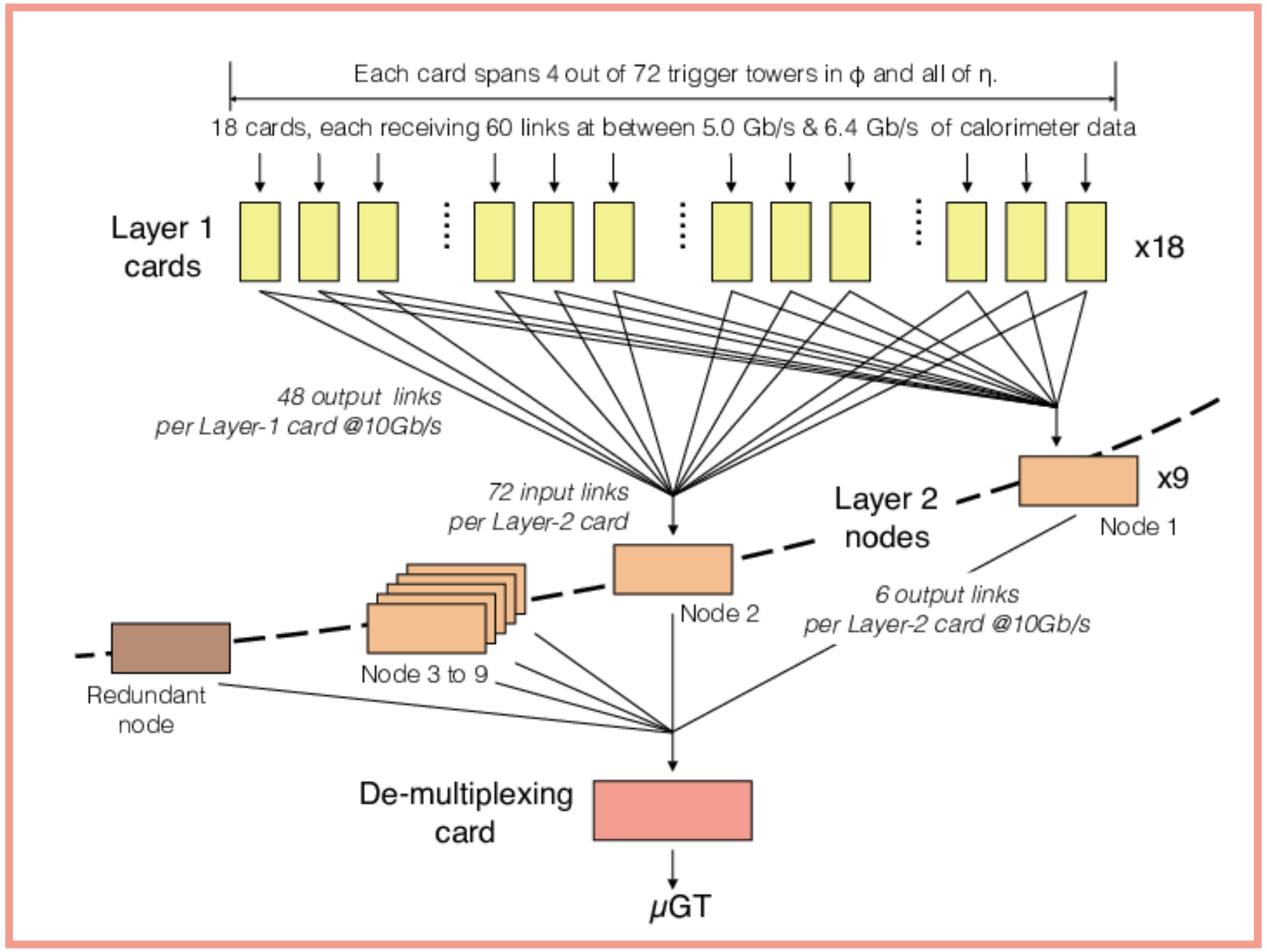
The CMS Trigger System



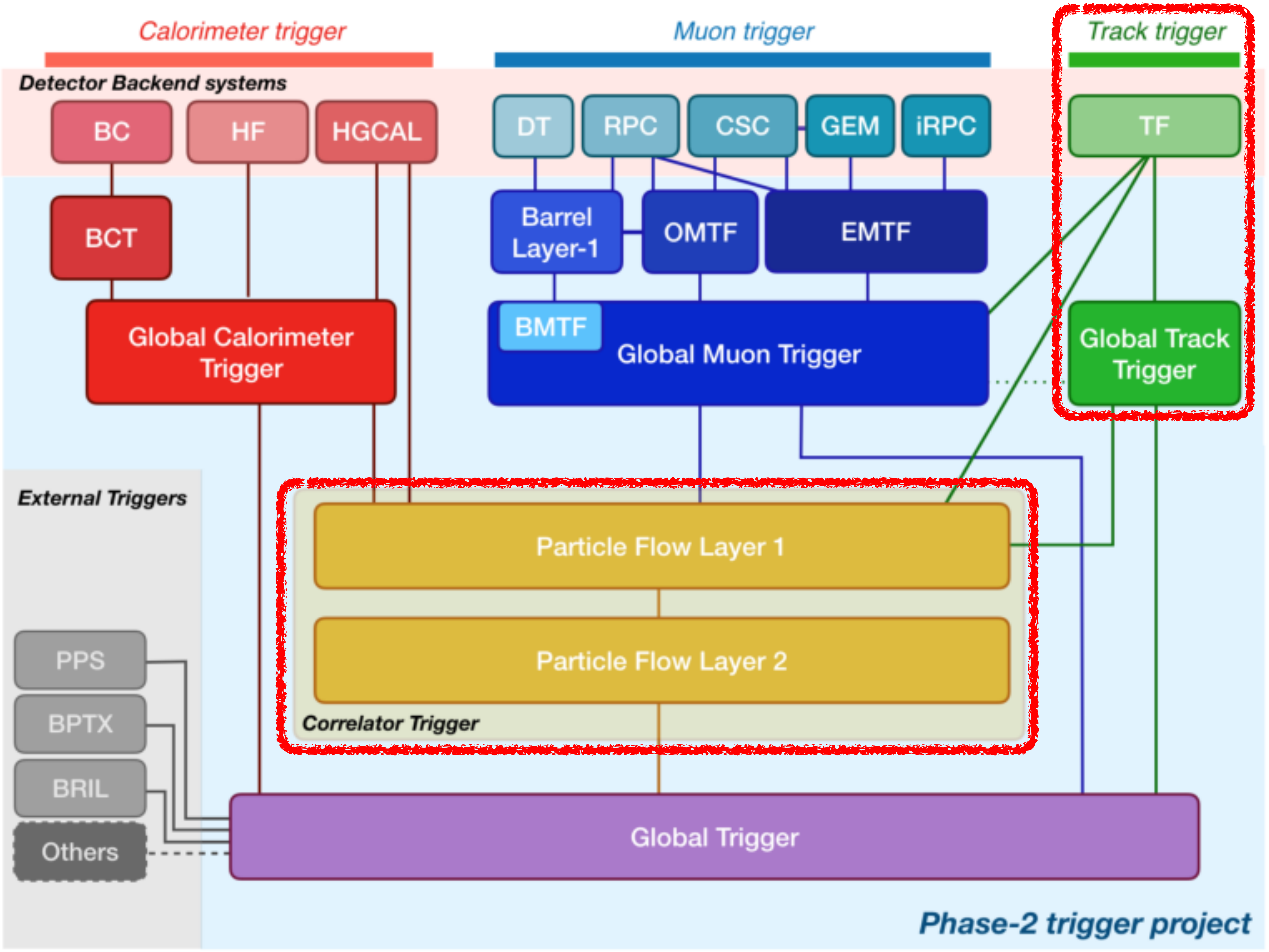
Time-multiplexing in calorimeter

Concentrate all data on 1 board!

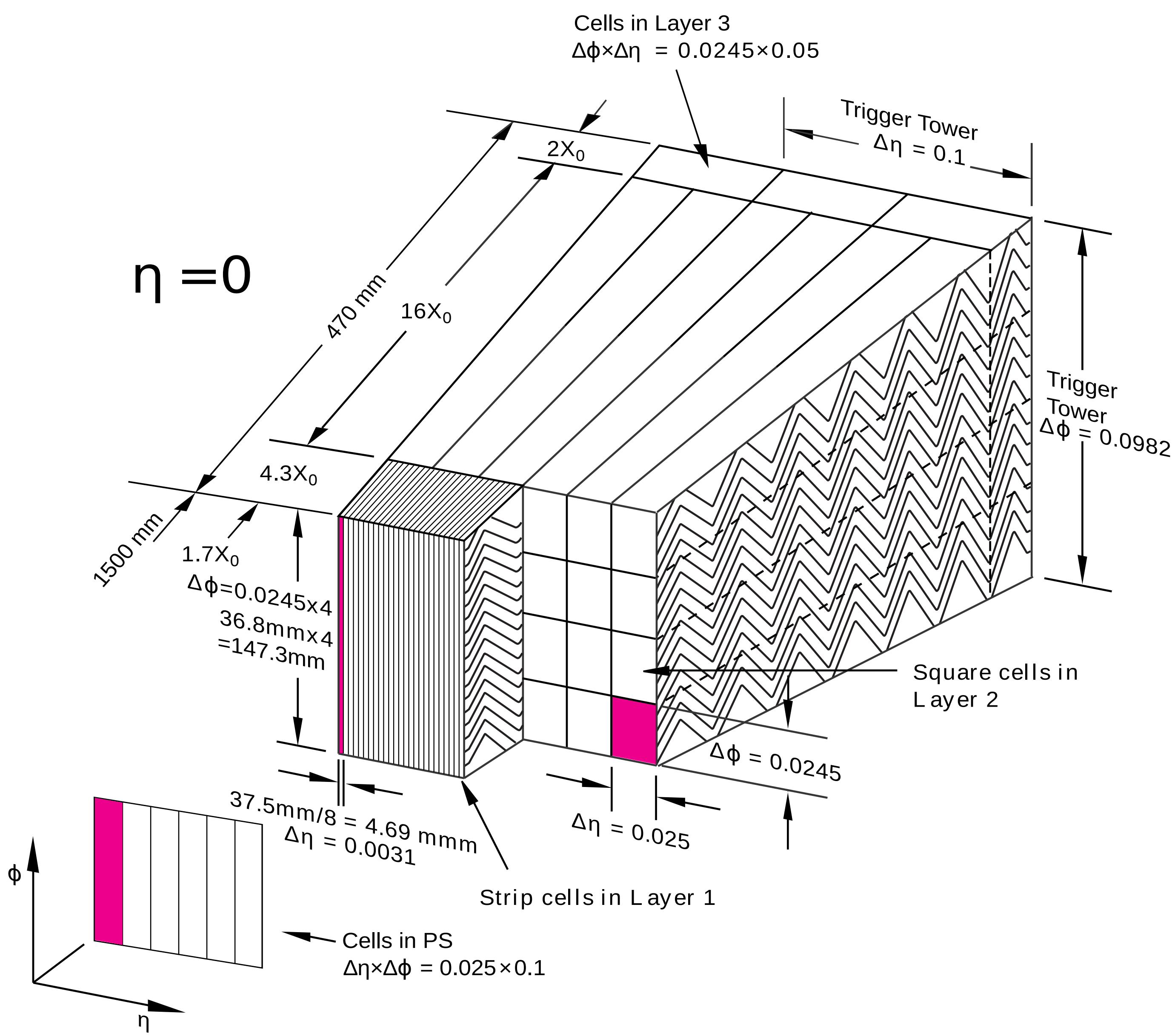
BUT, latency cost to mux and de-mux the data (~0.5μs)



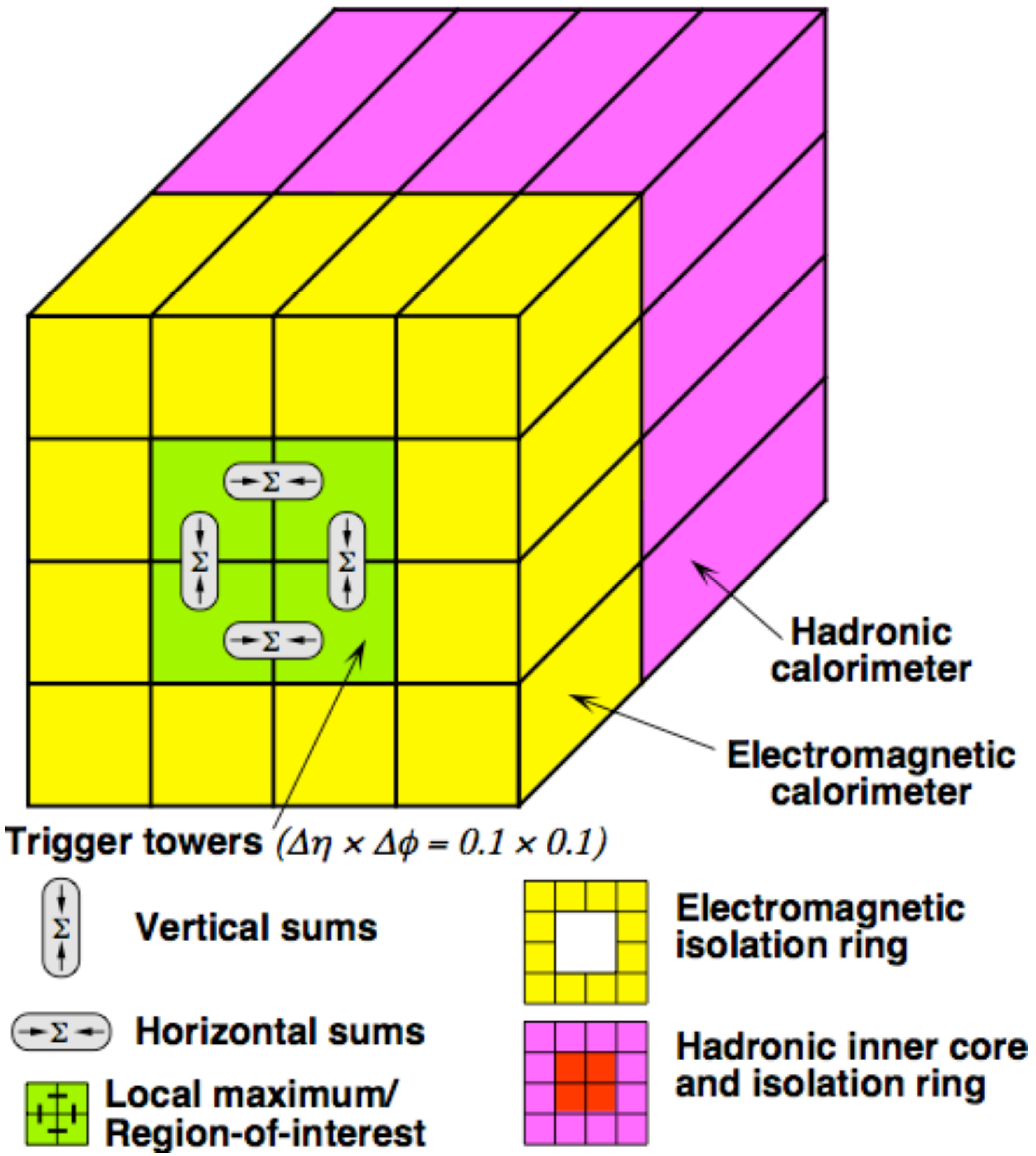
The CMS Trigger System



ATLAS HW Calorimeter trigger



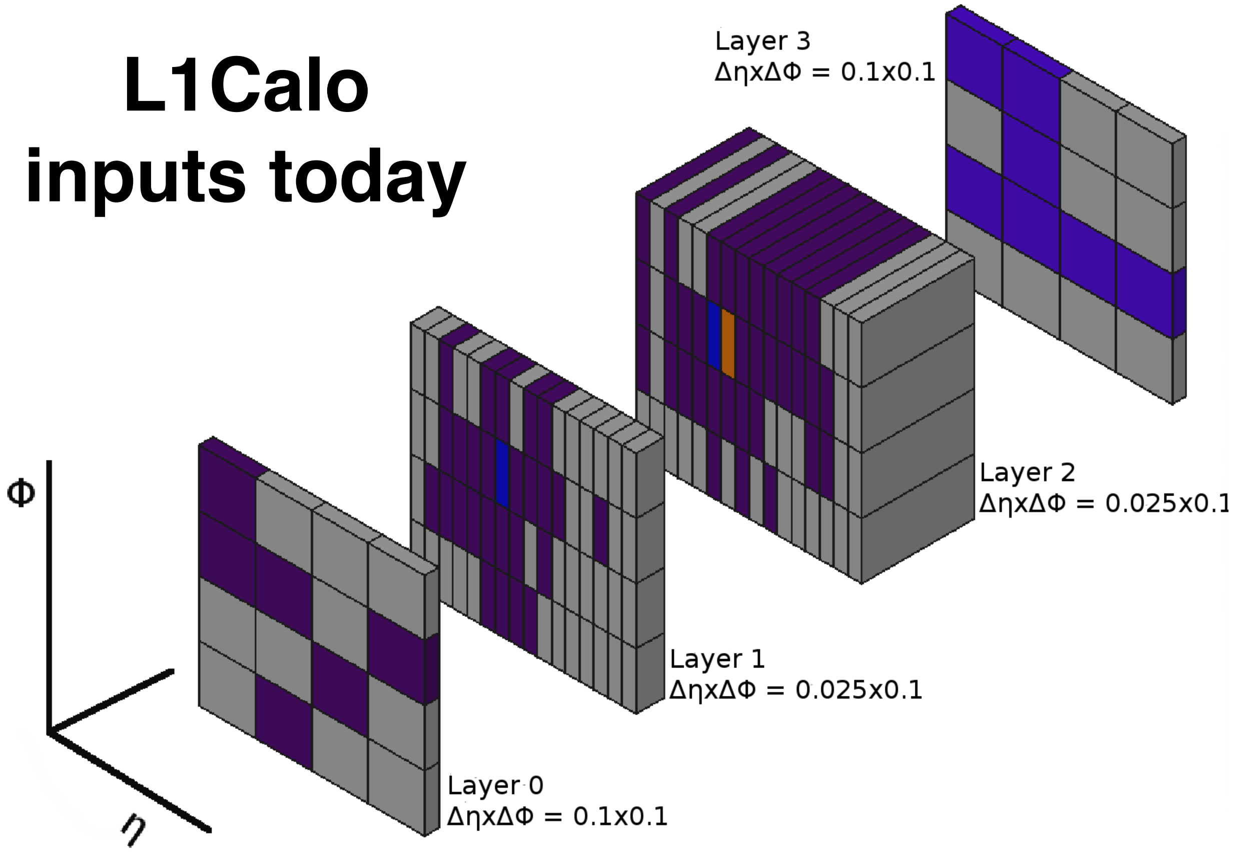
Original trigger design



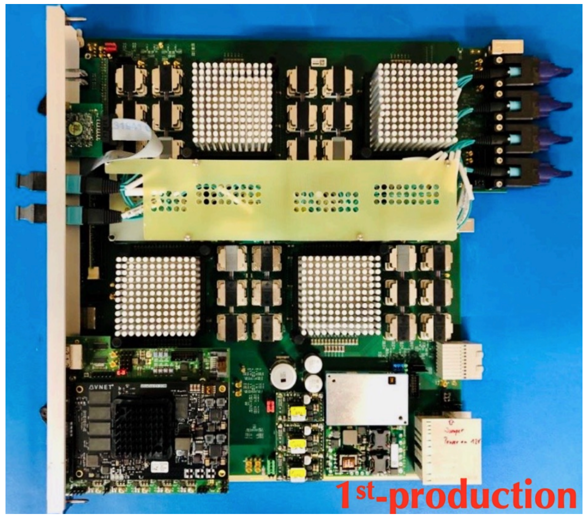
Simple algorithms (sums, local maxima), that are quick to compute in hardware.

ATLAS HW Calorimeter trigger

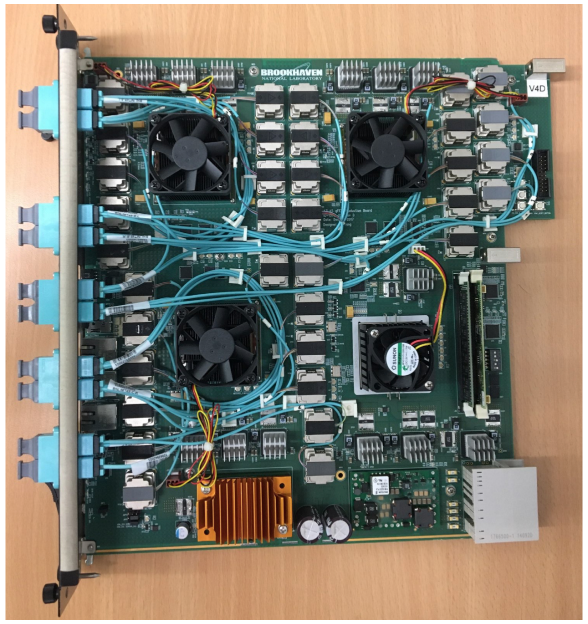
L1Calo inputs today



(a) eFEX



(b) jFEX

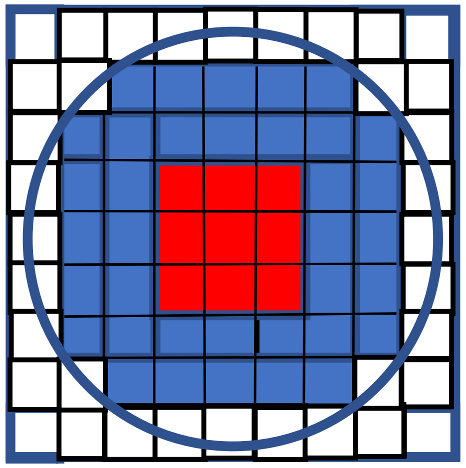


(c) gFEX

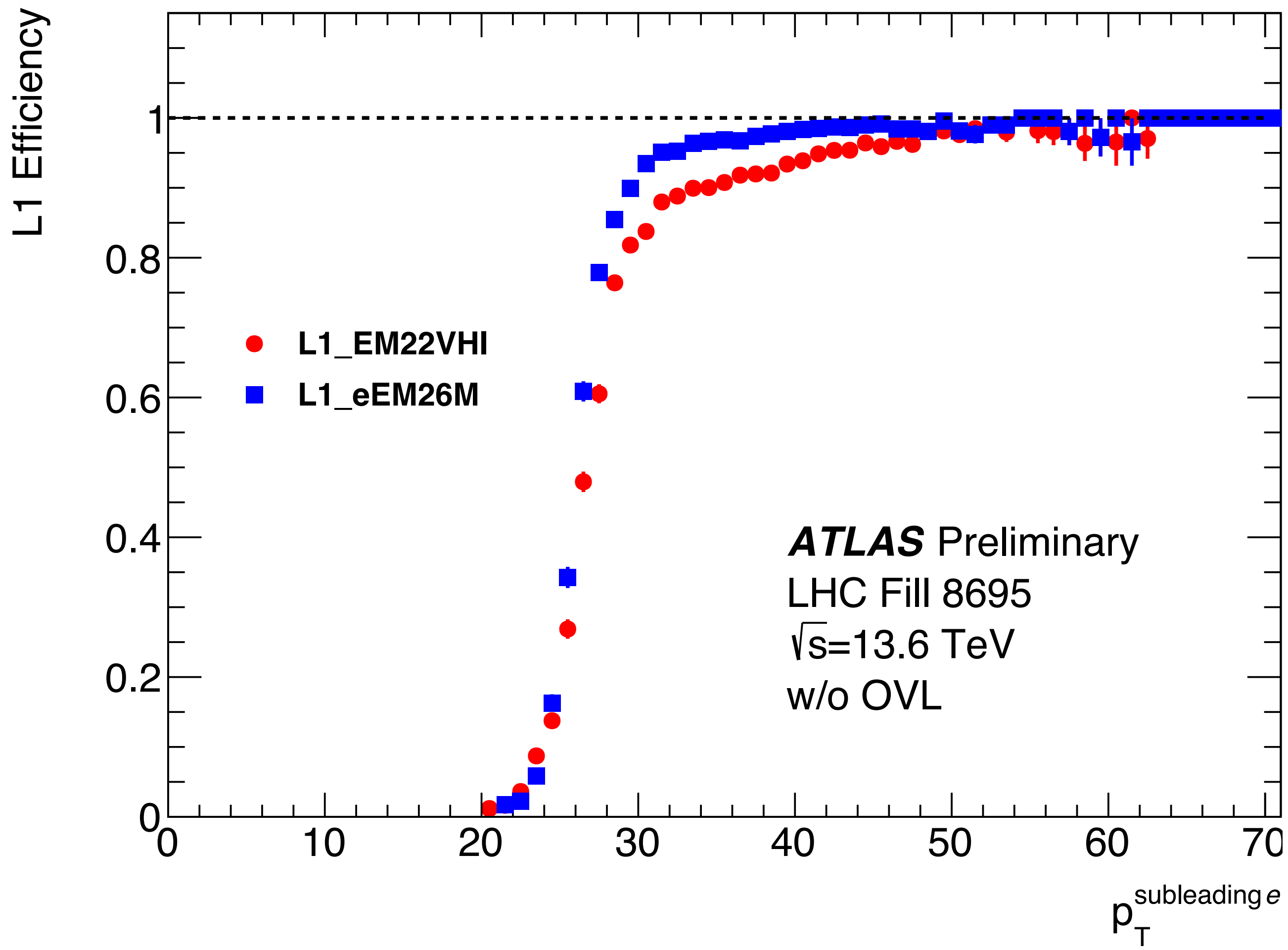
Improved granularity, thanks to new FEs, higher link speeds!

And a new generation of trigger hardware to match!

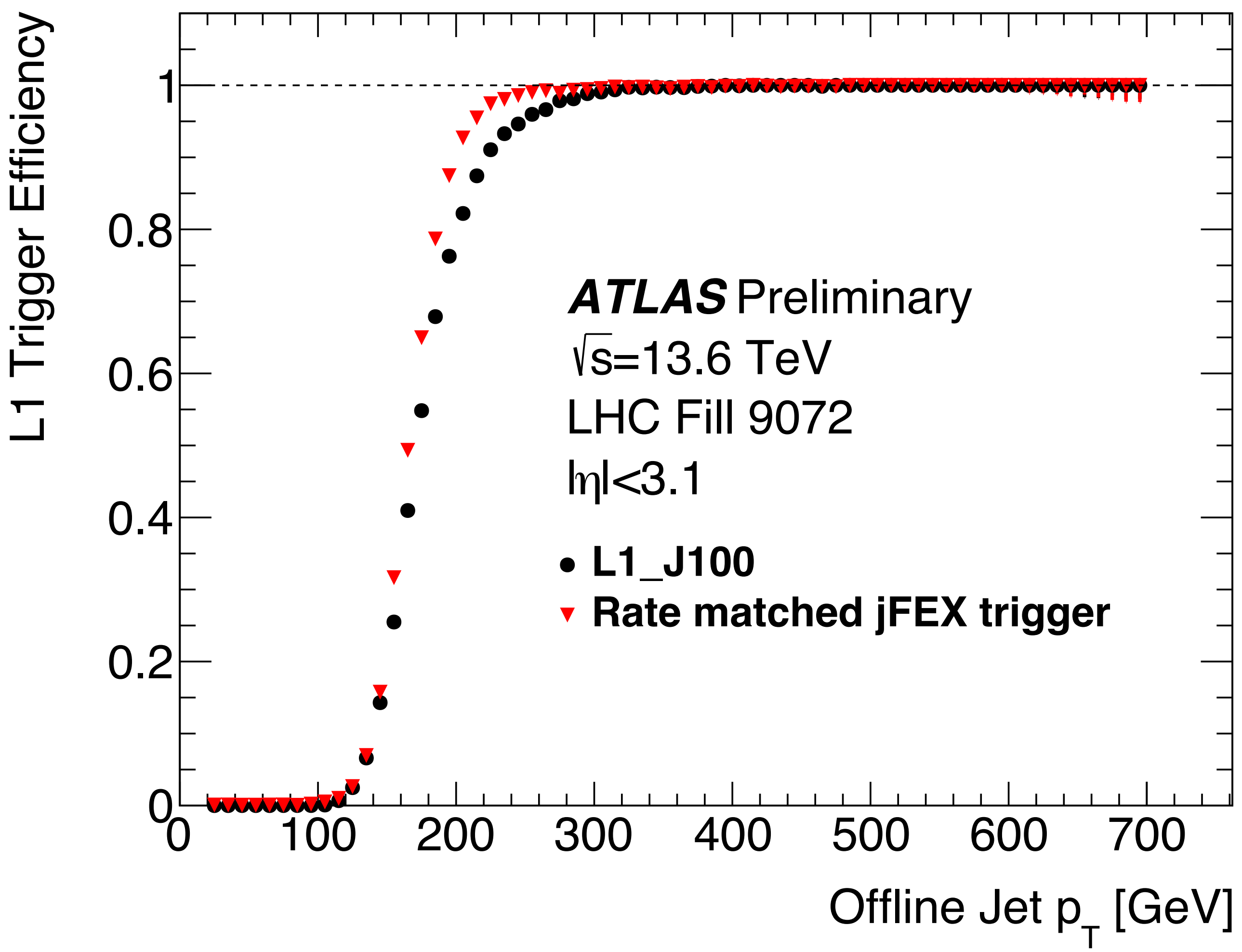
ATLAS HW Calorimeter trigger



Electrons & photons



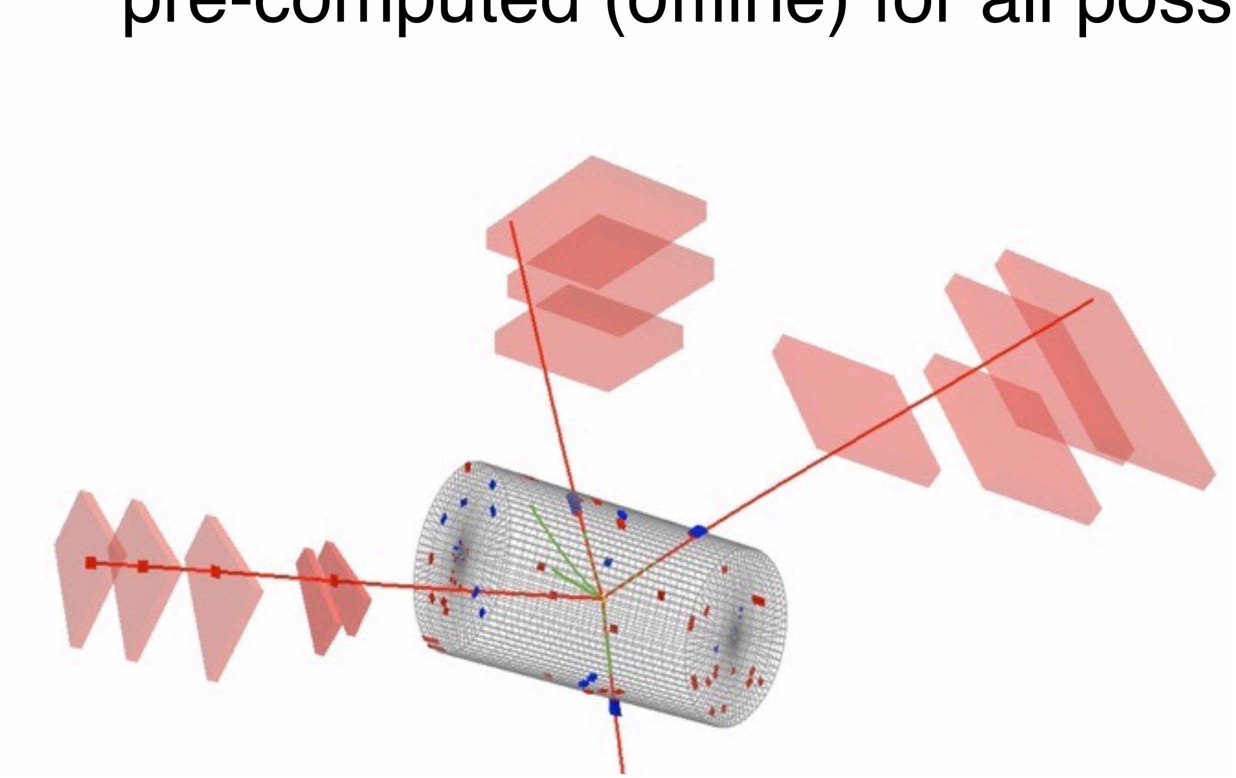
Jets

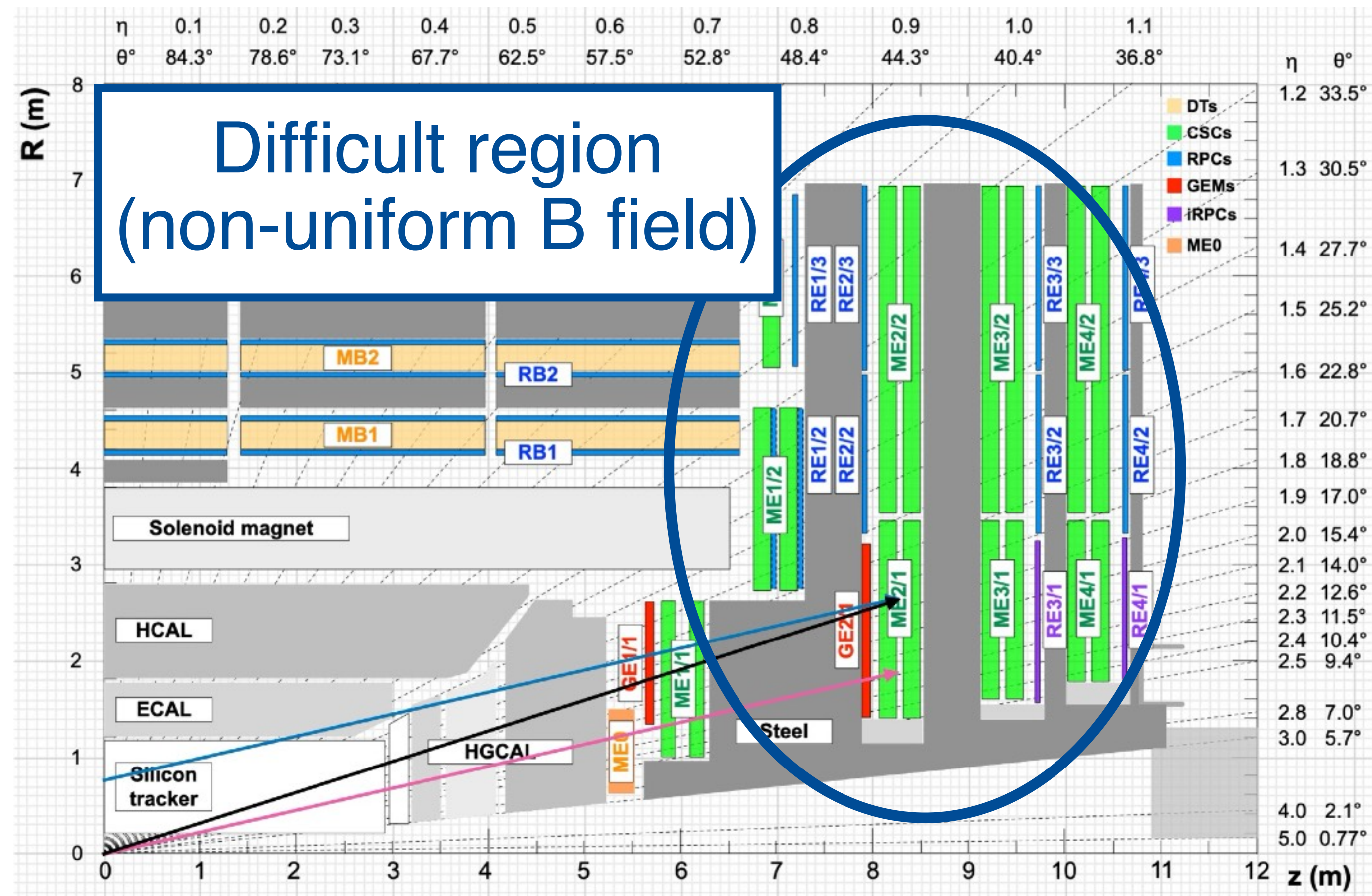


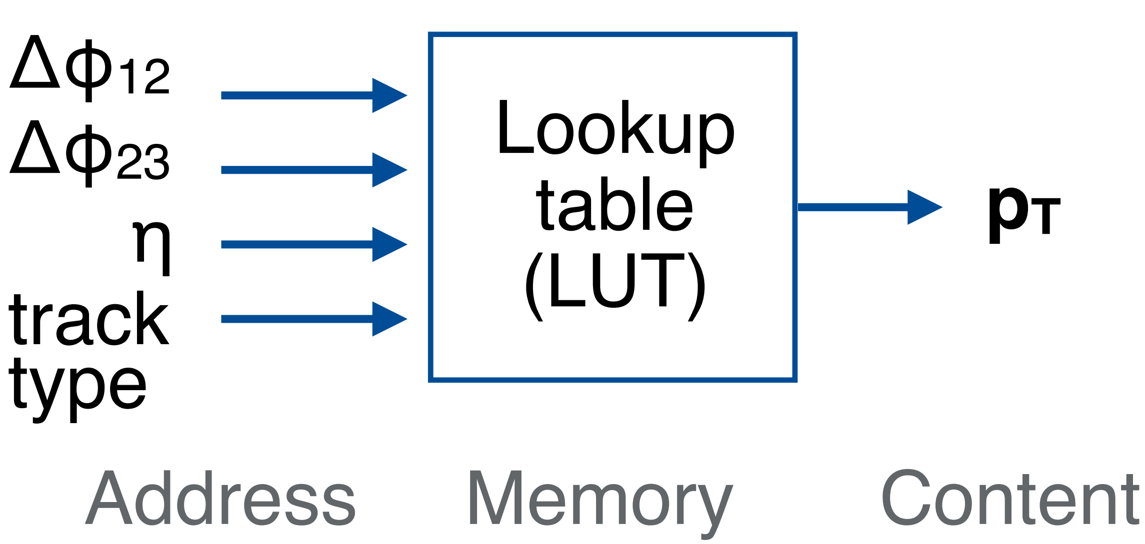
More detailed information leads to improved efficiencies (at fixed rate)

CMS HW Muon “reconstruction”

Instead of reconstructing μ p_T “on the fly”, results of an algorithm can be pre-computed (offline) for all possible input values for **fast lookup**.

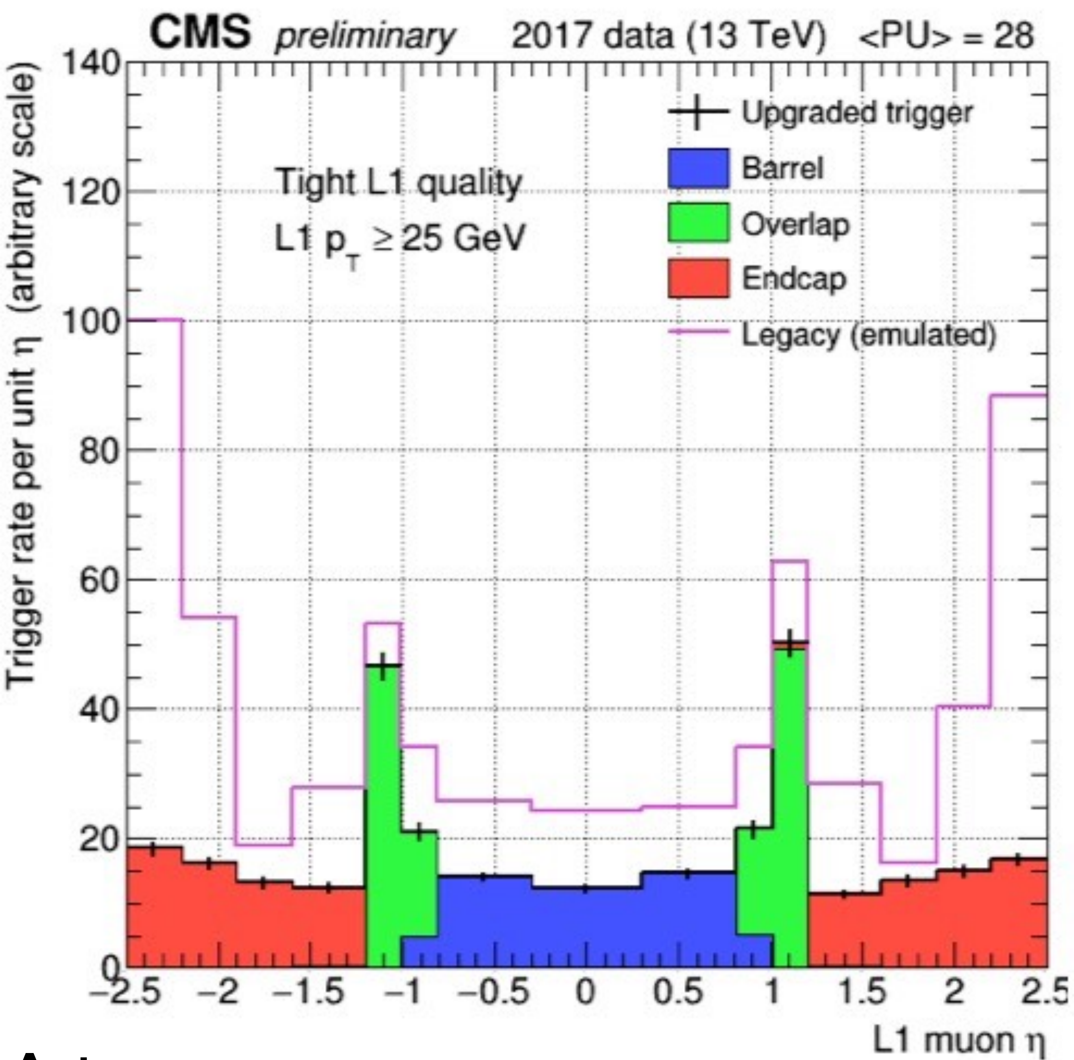


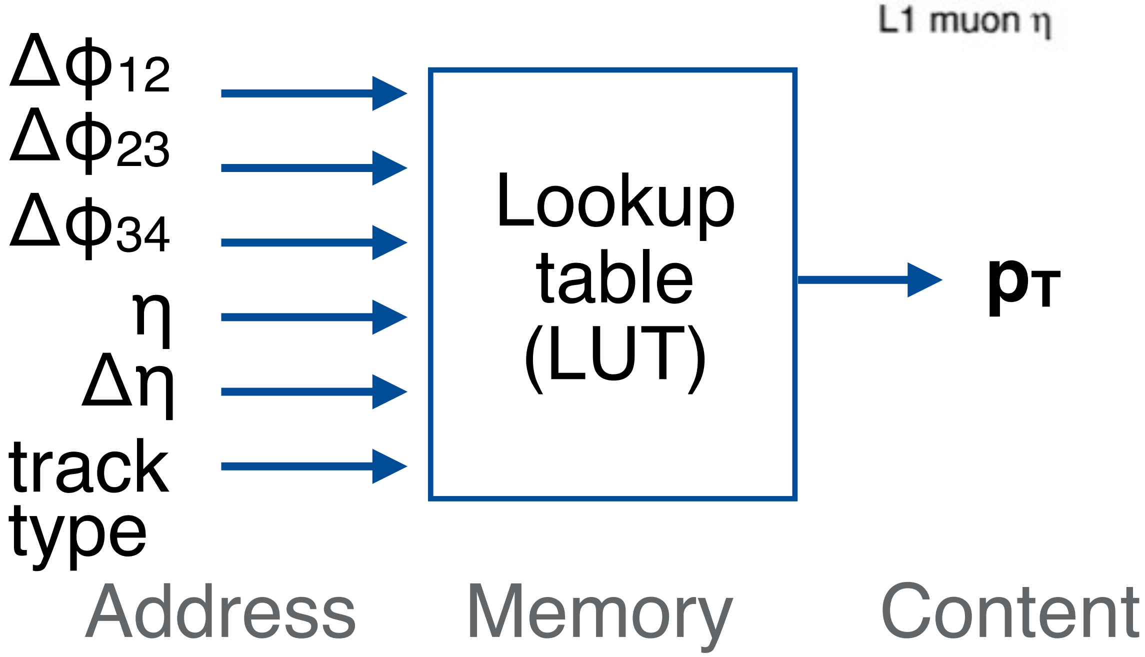




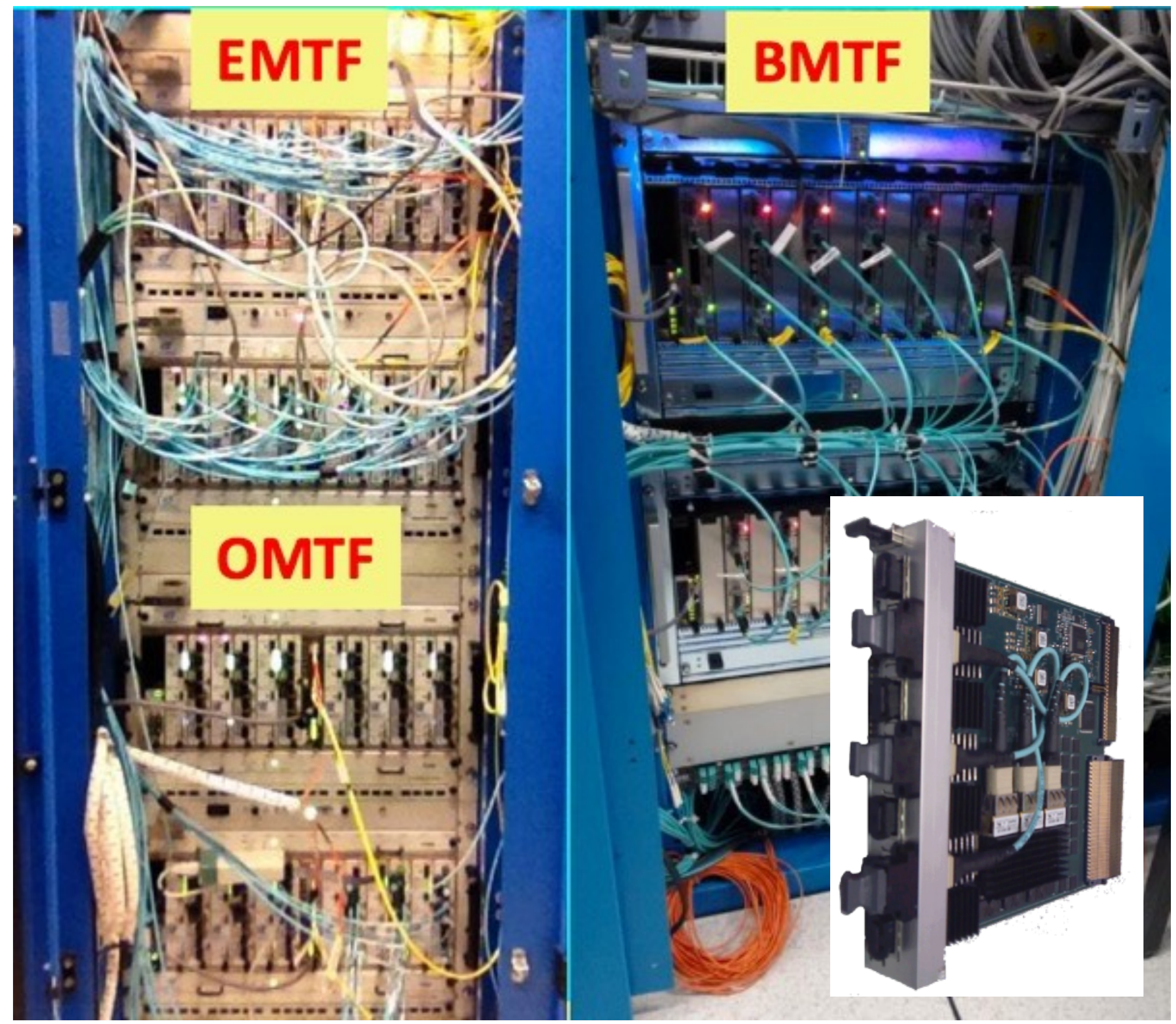
Original CMS: external RAM 22b

CMS HW Muon "reconstruction"



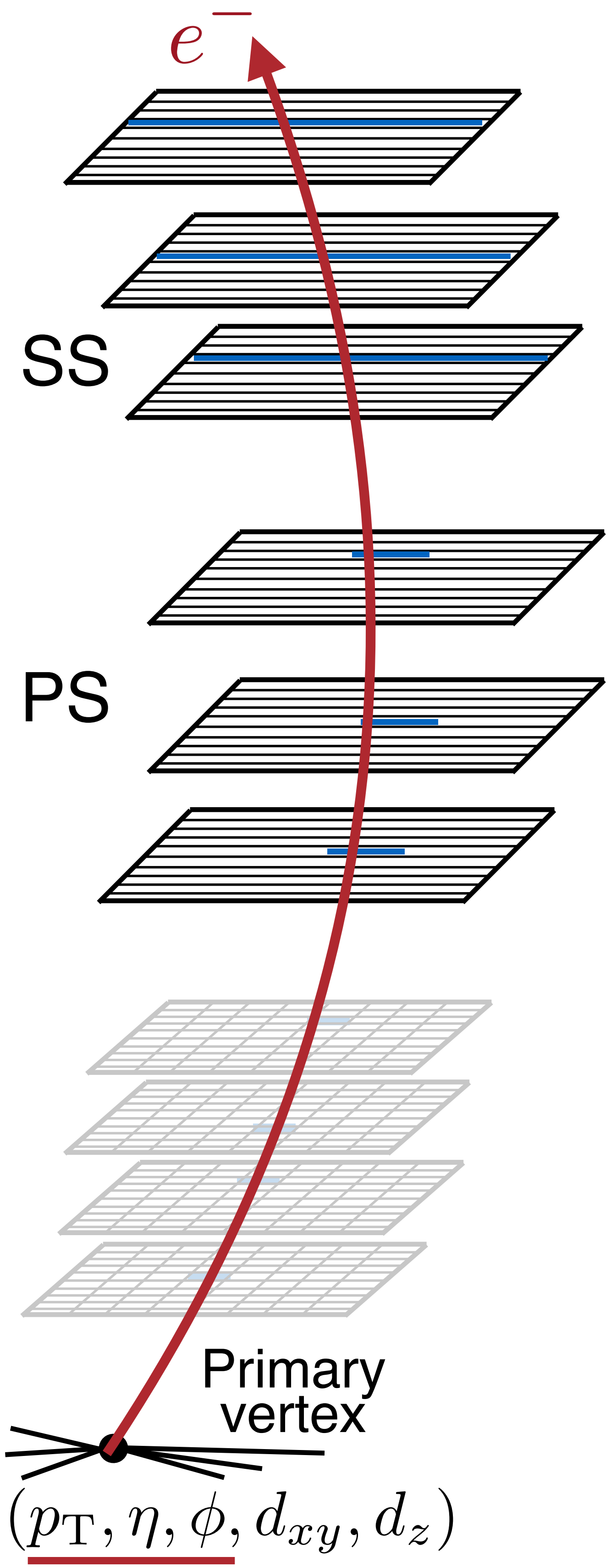


Phase-1: Larger FPGAs, internal DRAM with 30b address space



30b LUT actually encodes a Neural Net … more on that later!

Trigger-driven tracker design

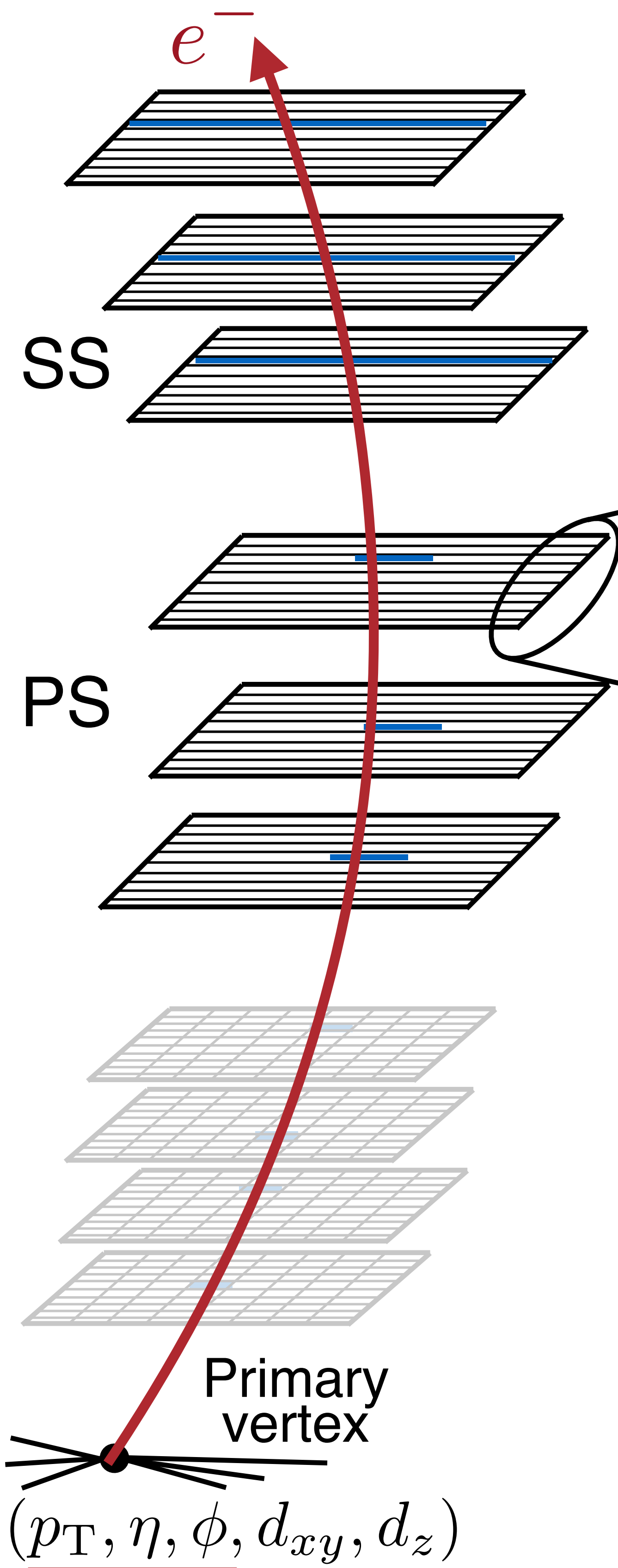


Large-radius sensors drive p_T measurement (lever arm).

Outer layers: 2 stacked sensors with 5cm strips "SS".

Inner layers: strips (2.4cm) + macro-pixel (1.5mm) "PS".

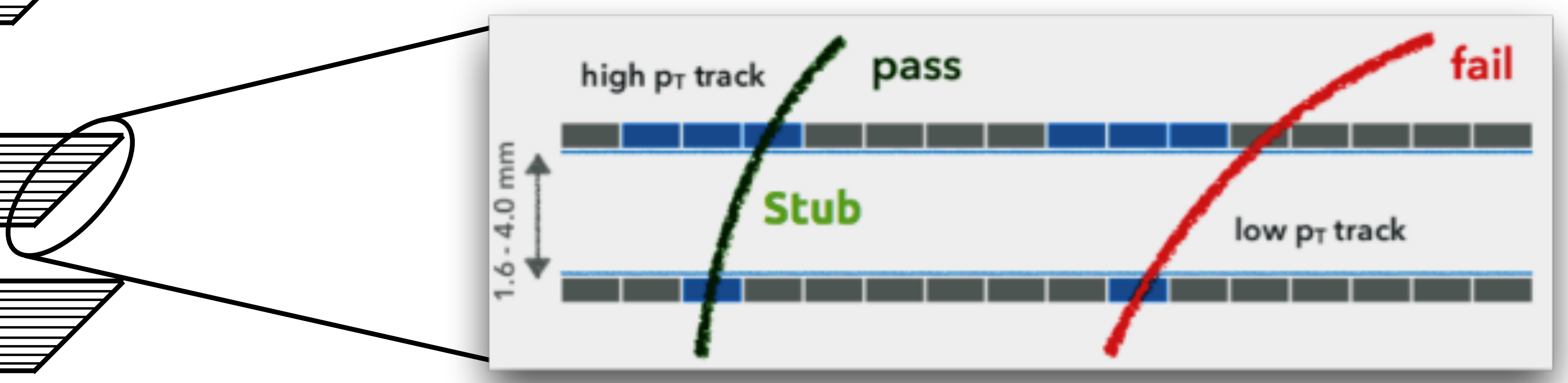
Trigger-driven tracker design



Large-radius sensors drive p_T measurement (lever arm).

Outer layers: 2 stacked sensors with 5cm strips “SS”.

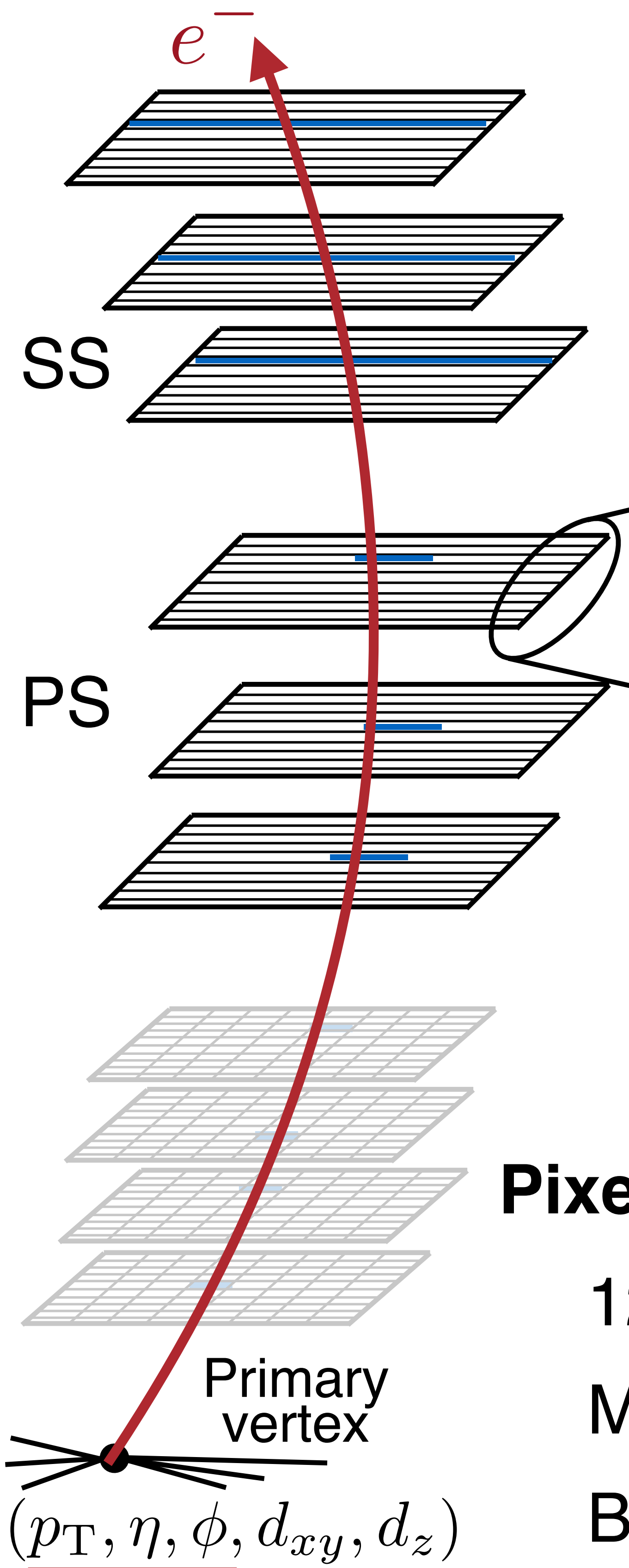
Inner layers: strips (2.4cm) + macro-pixel (1.5mm) “PS”.



Double-layer strip modules provide local p_T measurement.

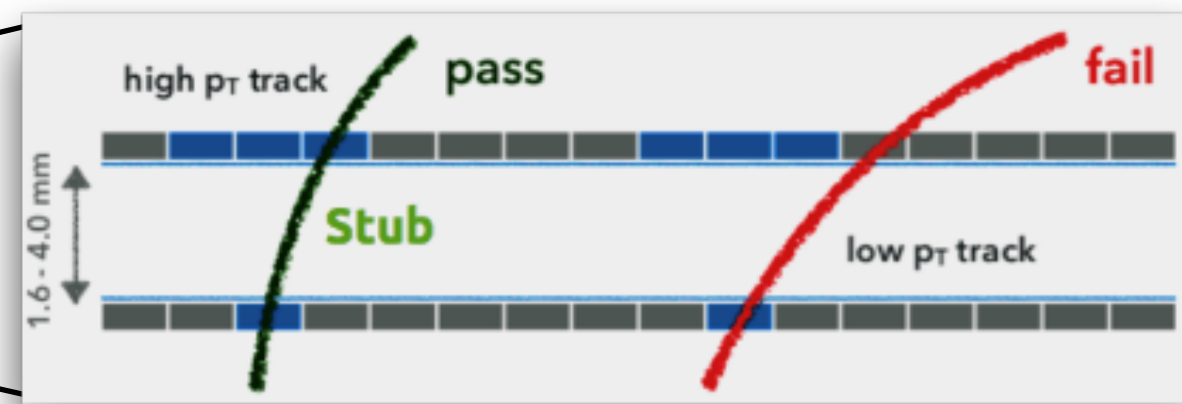
→ Intrinsic mechanism to filter hits from low-p_T tracks, allows high-p_T (2 GeV) track-finding in the trigger system!

Trigger-driven tracker design



Large-radius sensors drive p_T measurement (lever arm).

- Outer layers: 2 stacked sensors with 5cm strips "SS".
- Inner layers: strips (2.4cm) + macro-pixel (1.5mm) "PS".

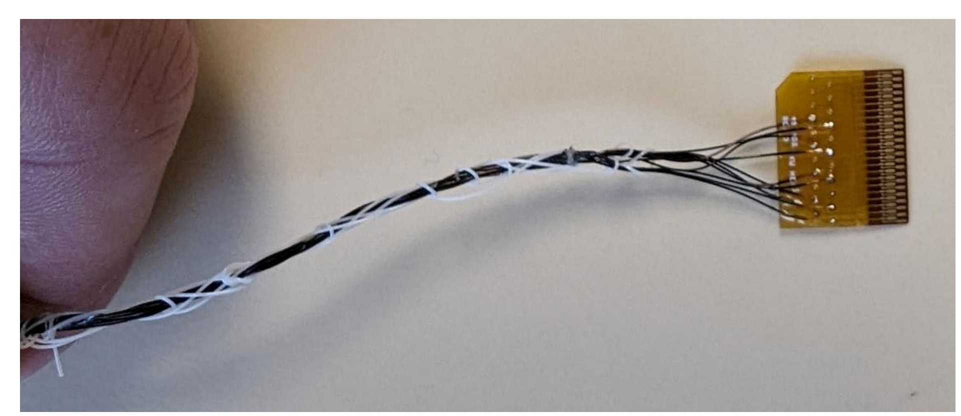


Double-layer strip modules provide local p_T measurement.

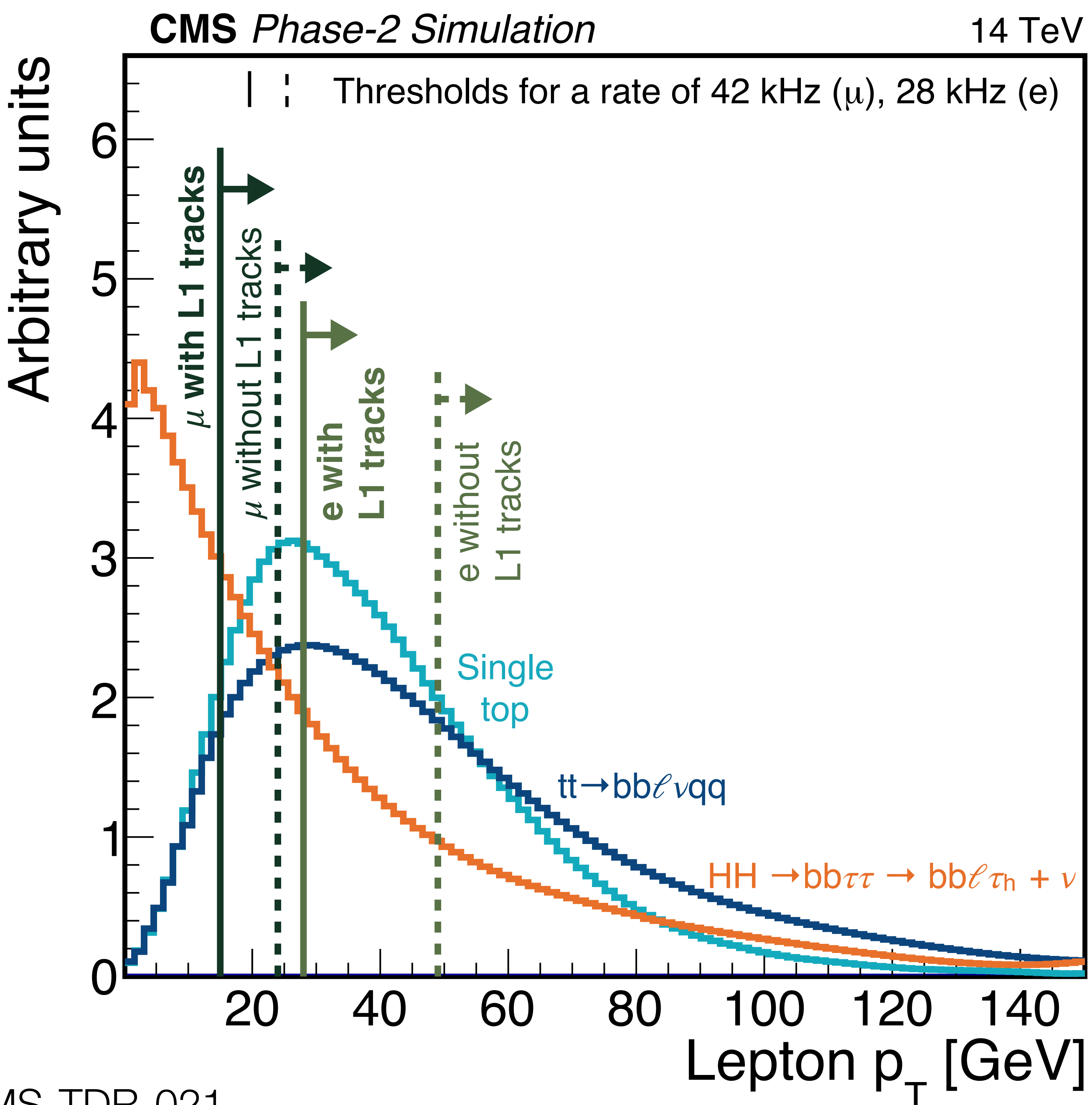
→ Intrinsic mechanism to filter hits from low-p_T tracks, allows high-p_T (2 GeV) track-finding in the trigger system!

Pixel sensors are more challenging

- 124M → 1.9B channels
- Minimize material (heat, cooling)
- But, 12.8µs buffer!

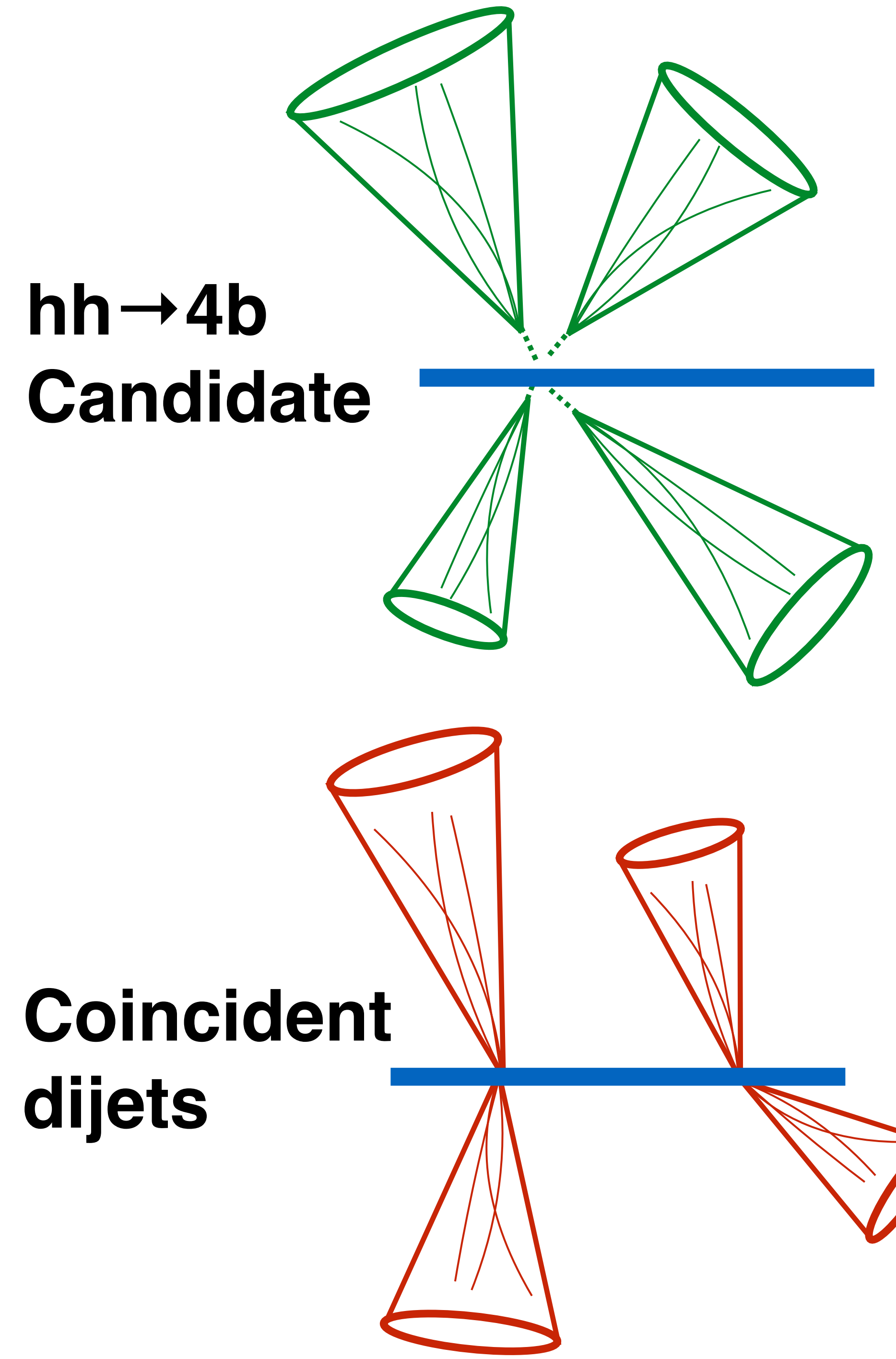


The case for tracks in the trigger



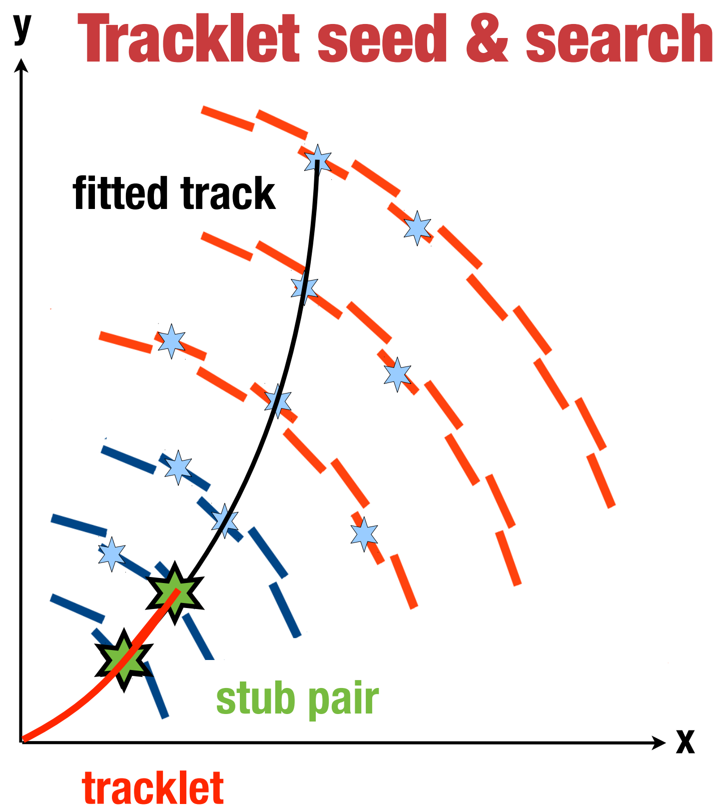
CMS-TDR-021

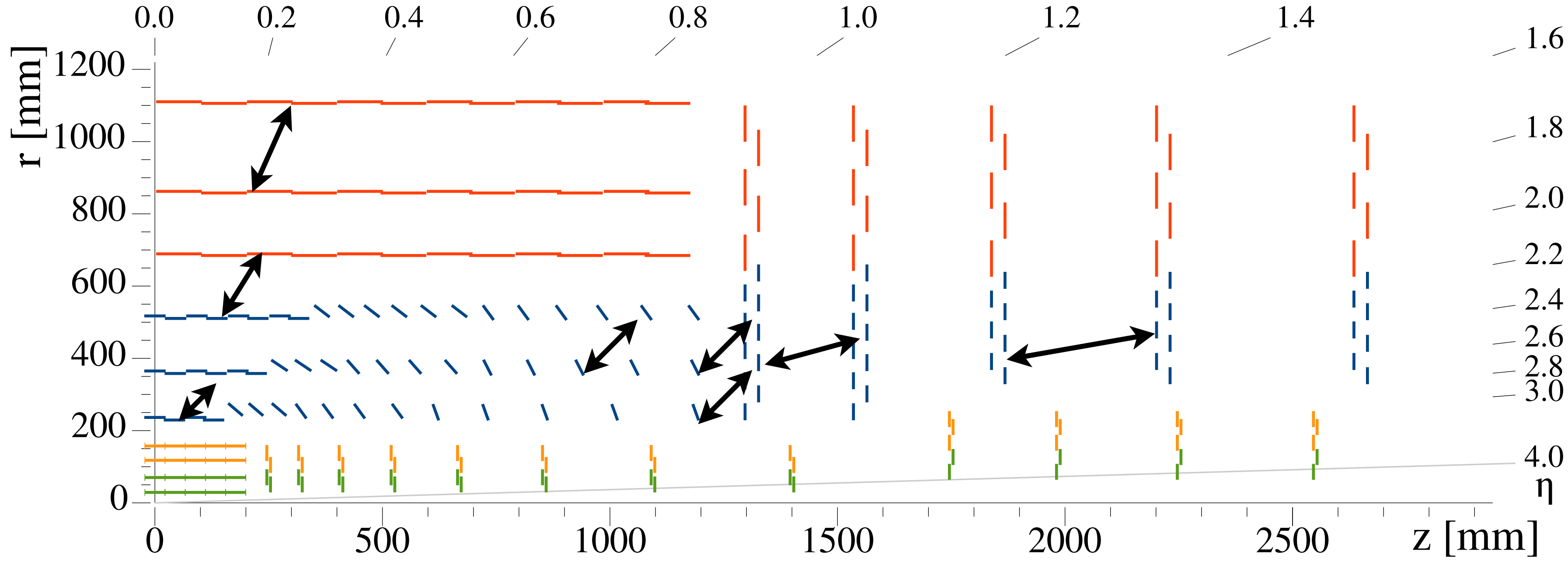
Maintain low lepton thresholds



More handles for hadronic backgrounds

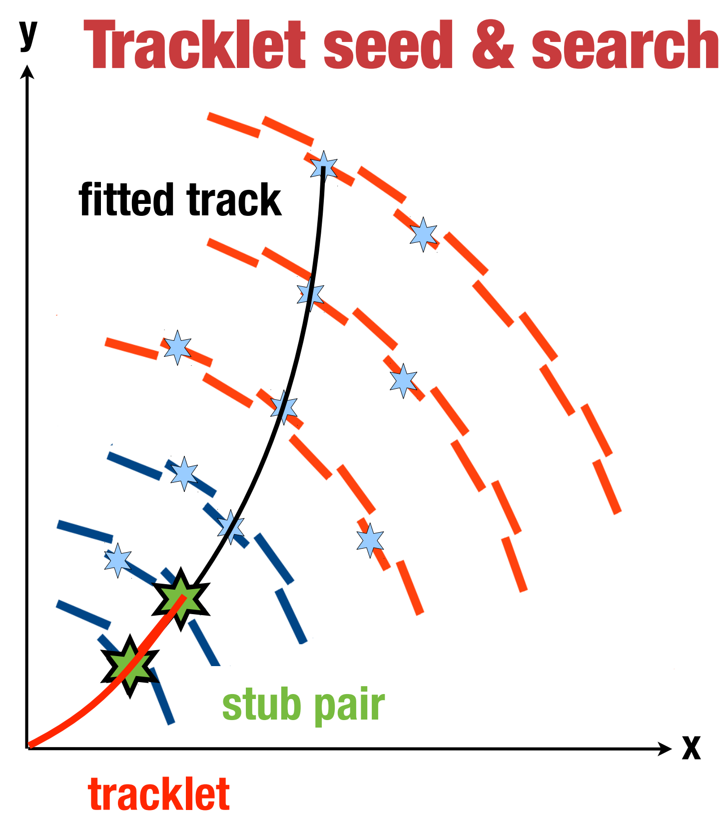
Tracking in Hardware (I)

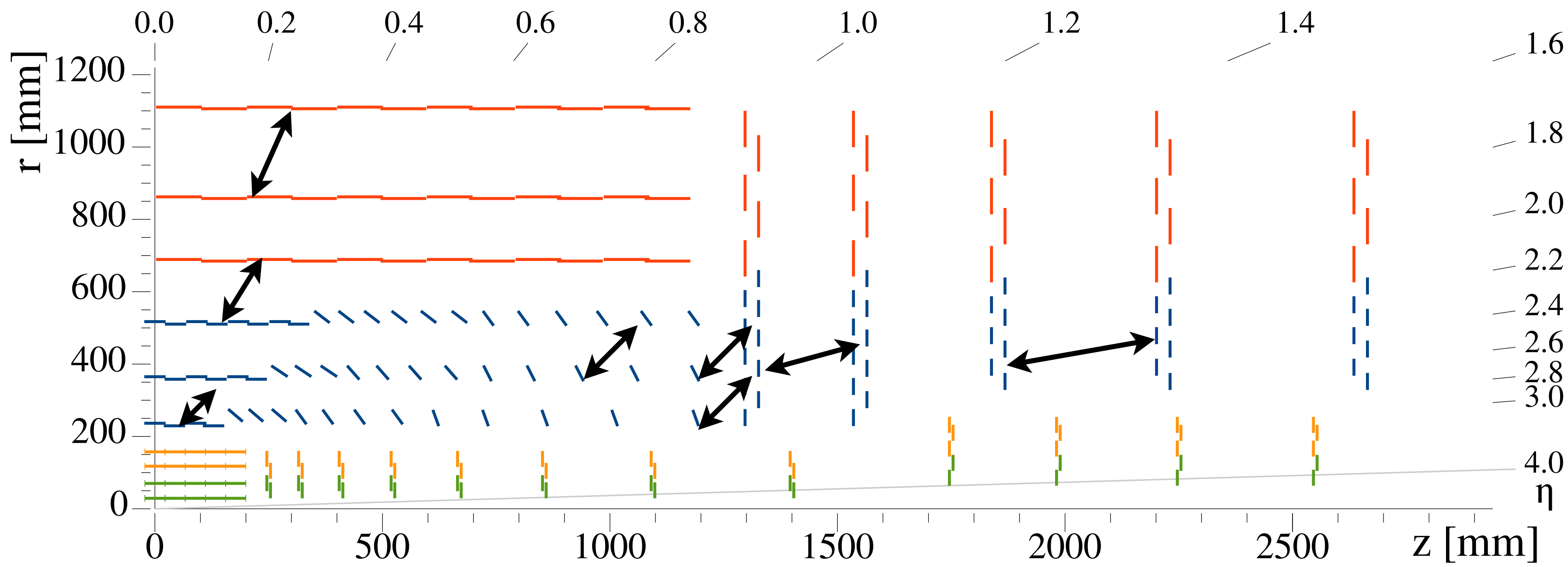




Beam spot + 2 layers form a proto-track.

Tracking in Hardware (I)

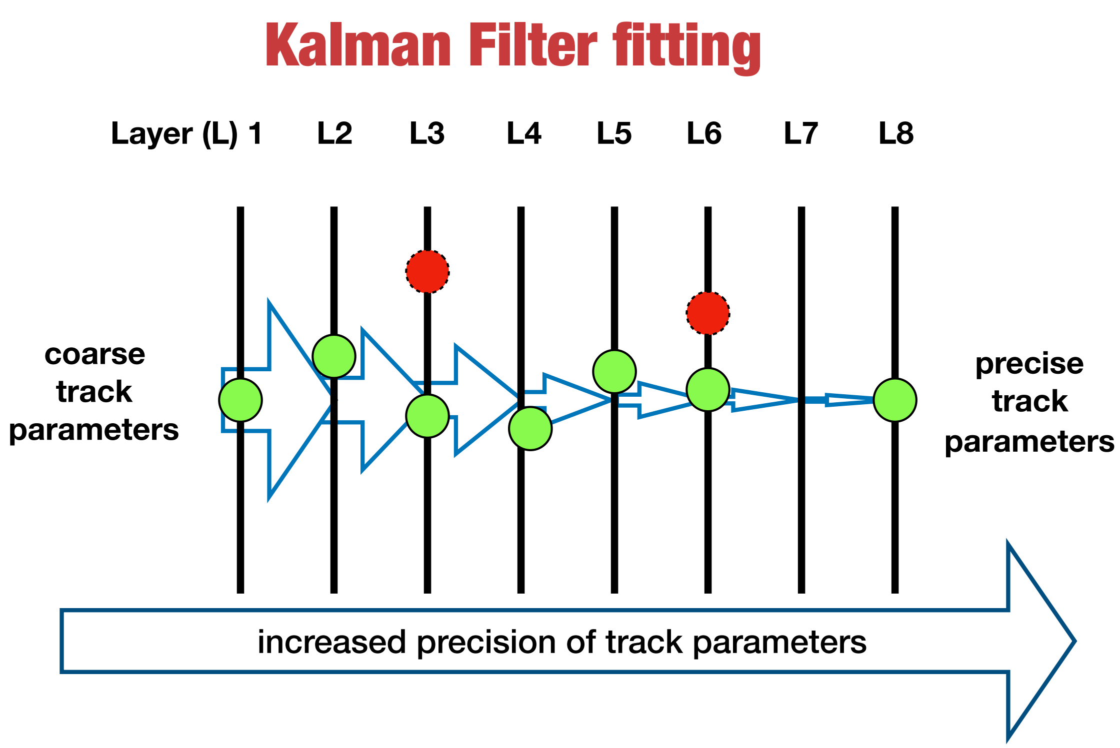




Beam spot + 2 layers form a proto-track.

Parameters are refined by extrapolating into other layers with a Kalman Filter

Baseline algo gives (p_T, η, ϕ, d_z, d_{xy})

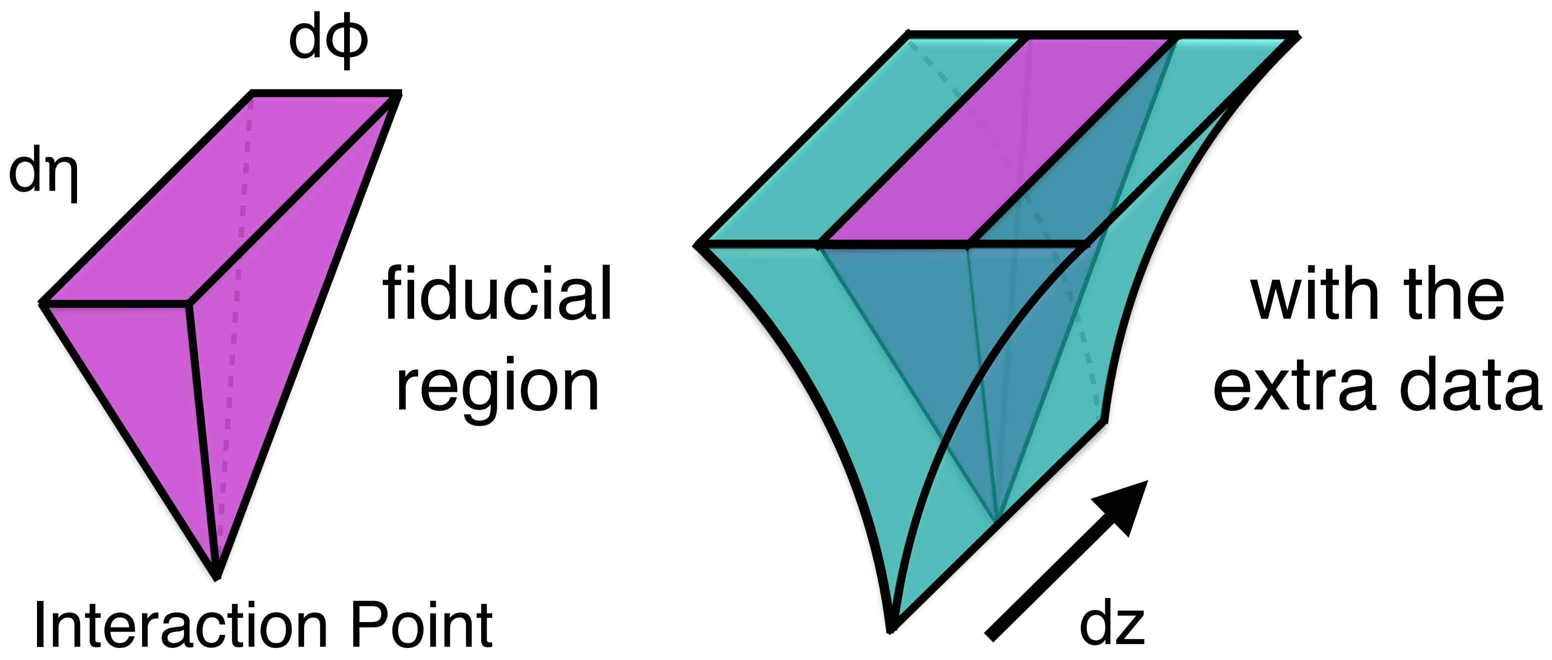


Tracking in Hardware (II)

Data-sharing is painful
for innermost detectors.

Best to use large regions

→ 9 ϕ slices

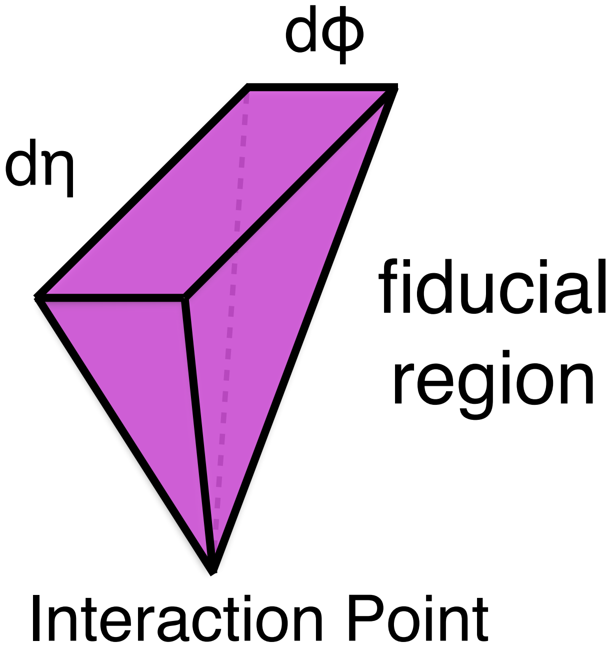


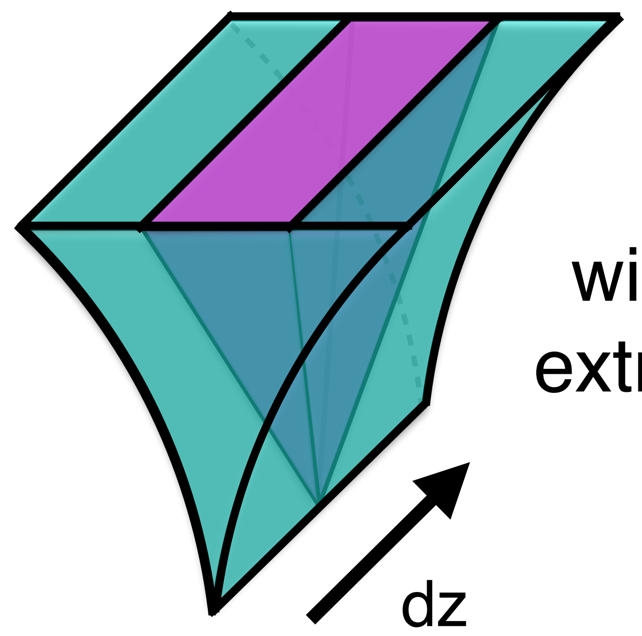
Tracking in Hardware (II)

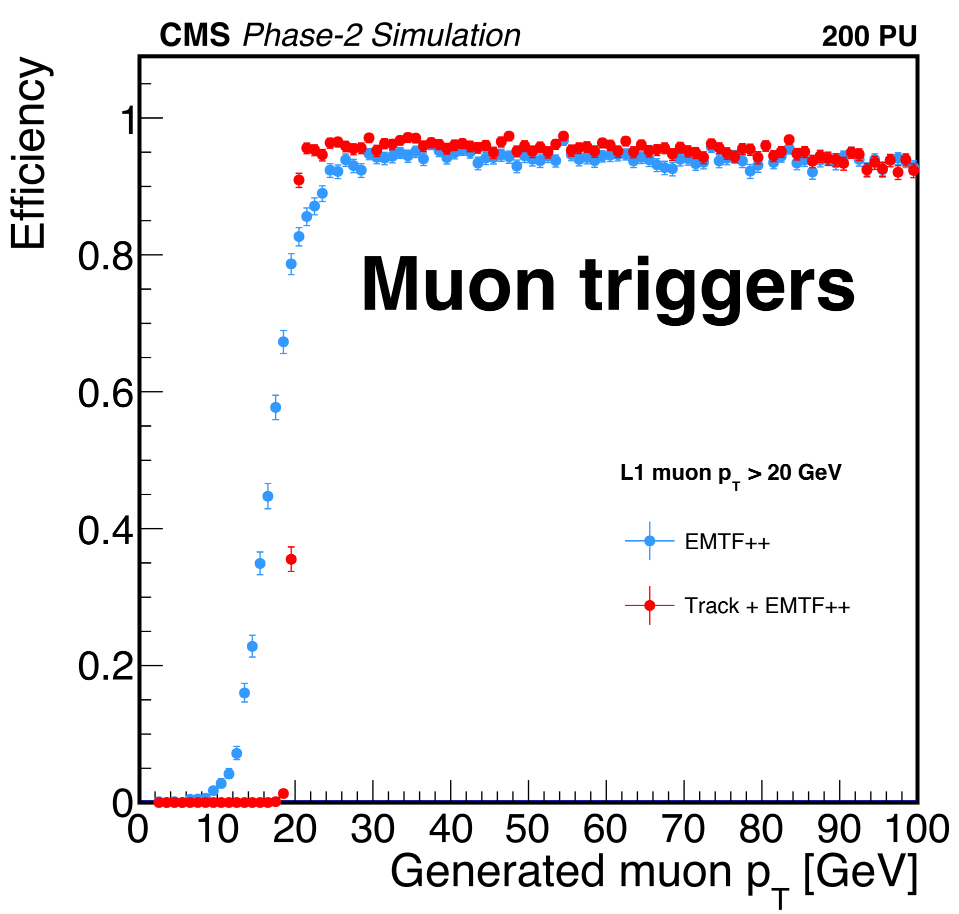
Data-sharing is painful for innermost detectors.

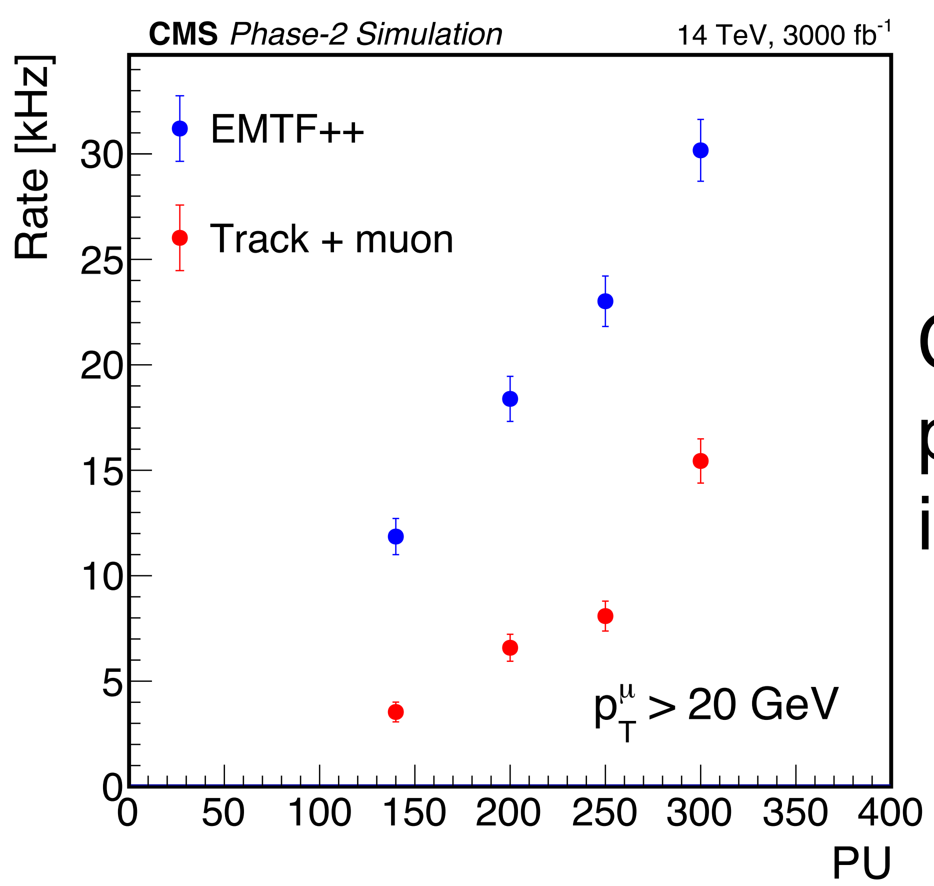
Best to use large regions

→ 9 ϕ slices





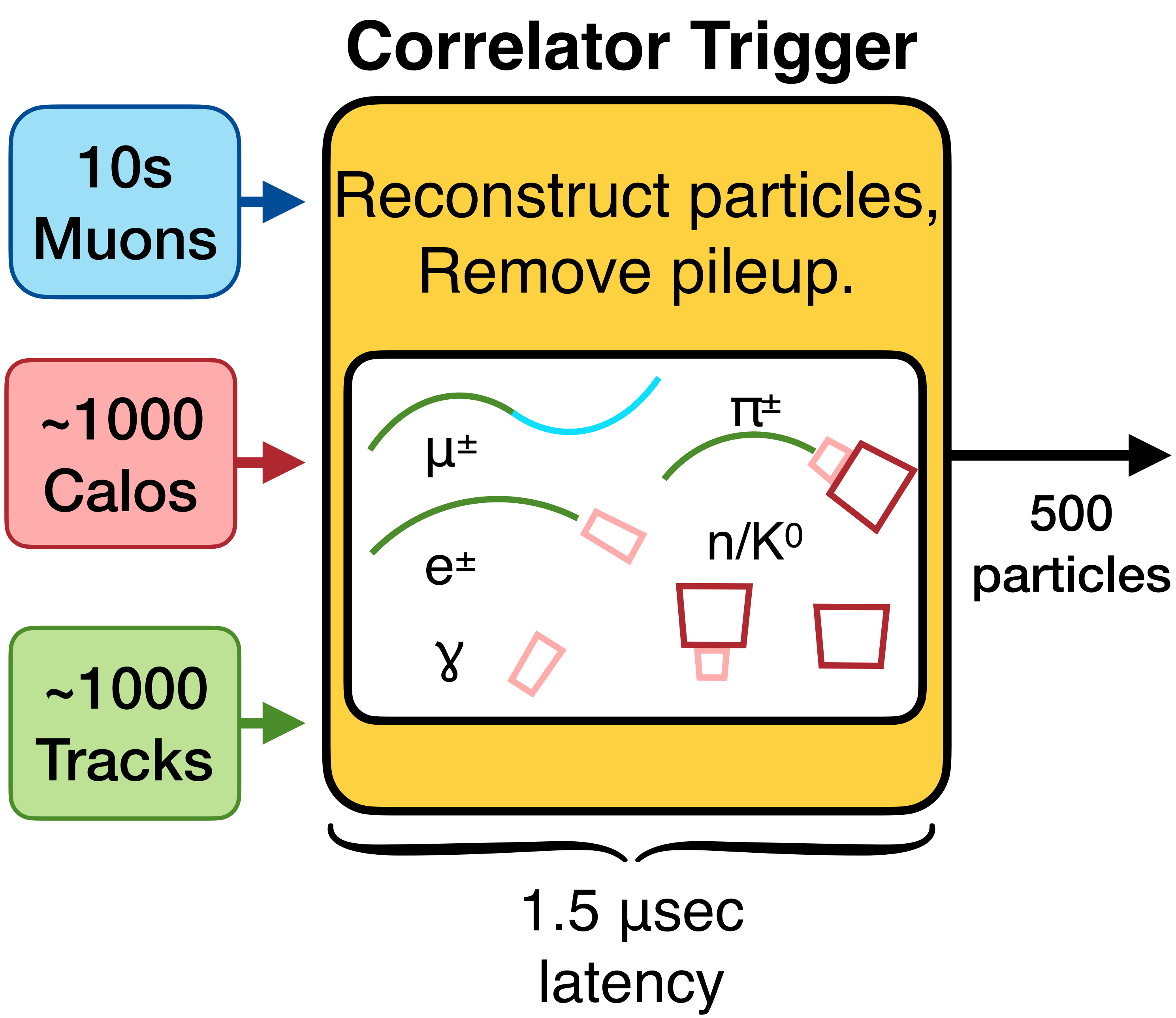


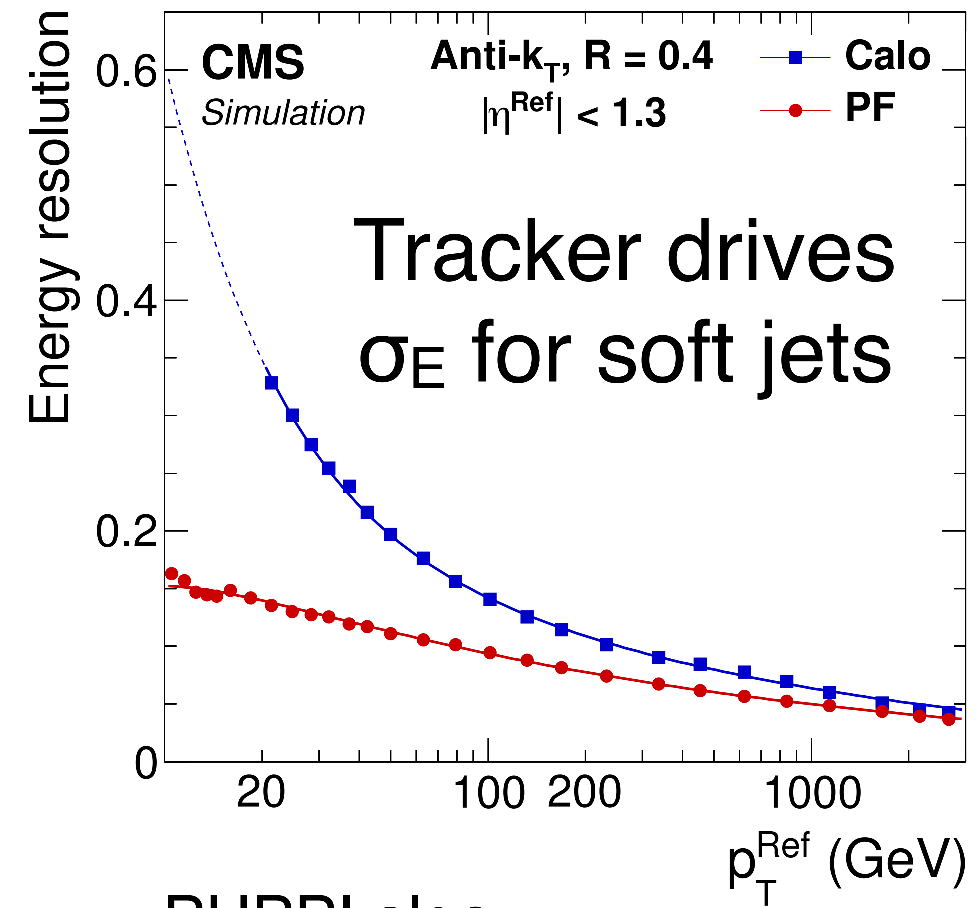


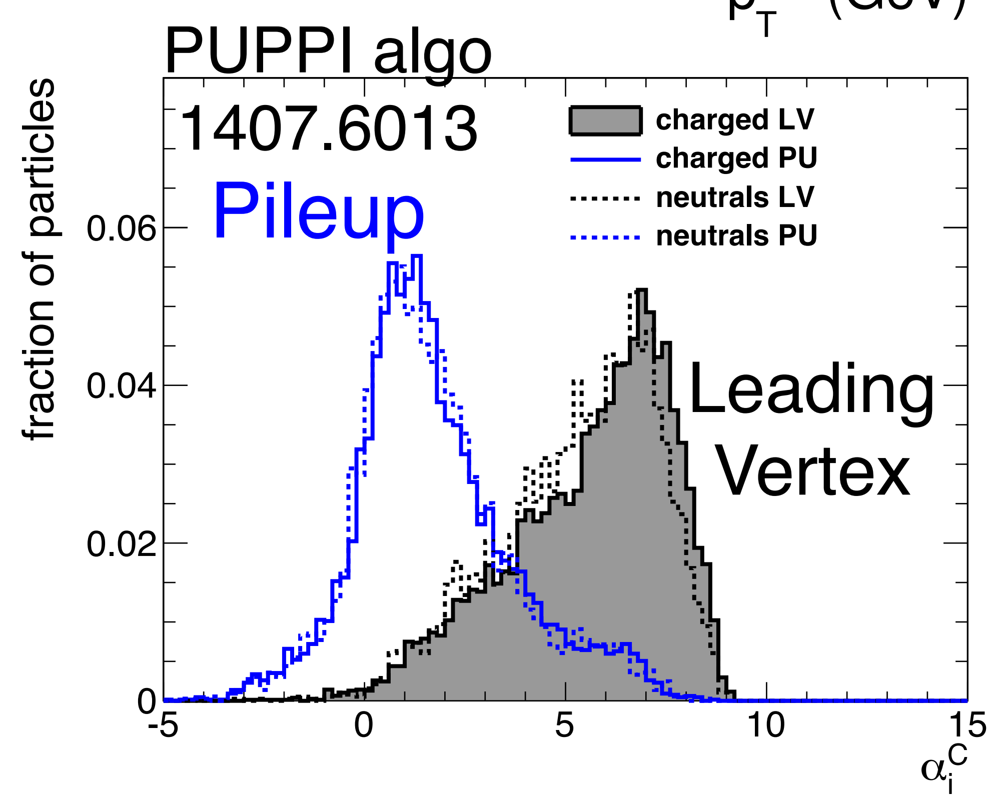
Challenging, but large potential improvements in rate & efficiency!

Fundamentally new capabilities!

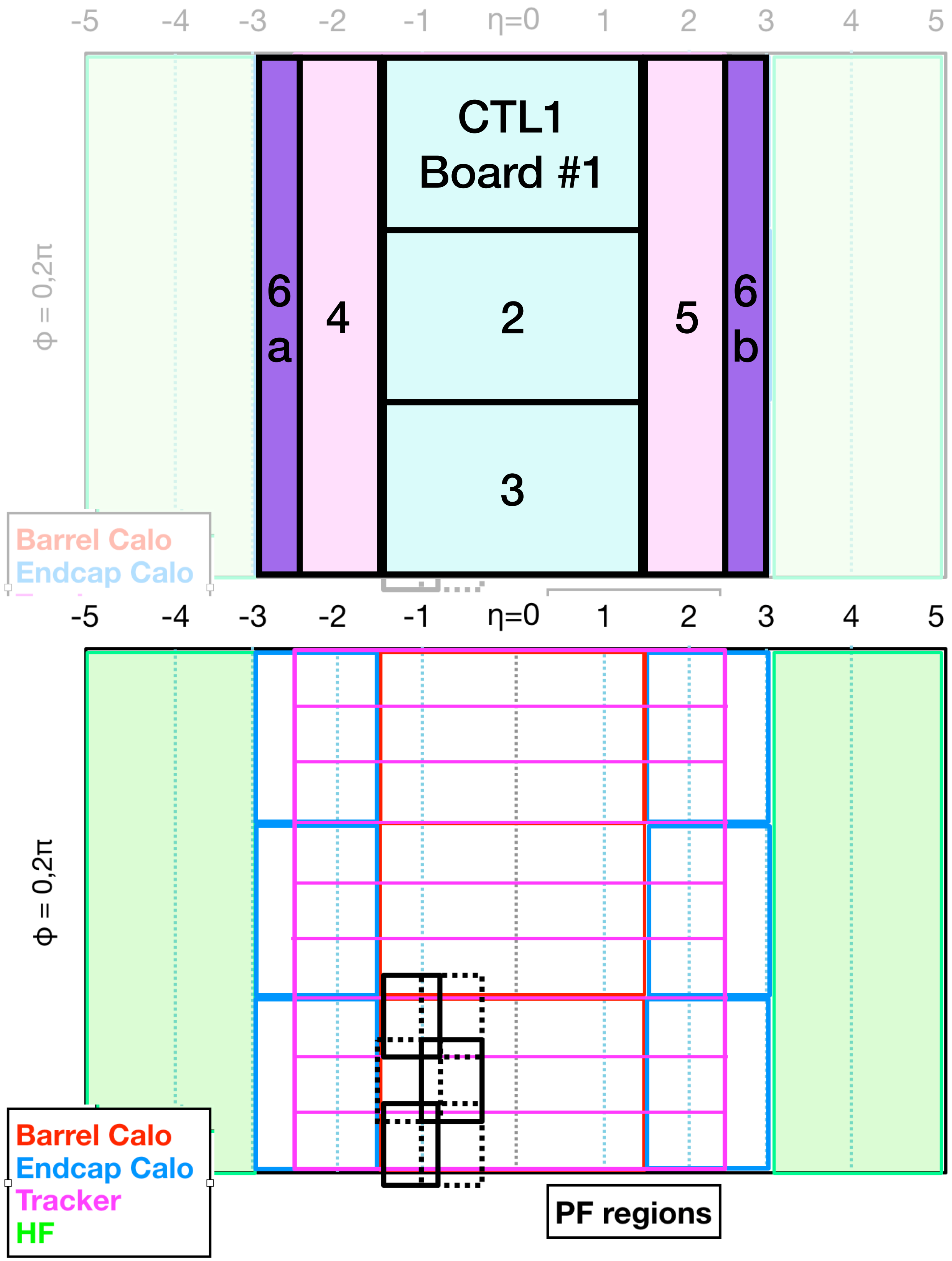
Particle flow algorithm:
the backbone of CMS reconstruction



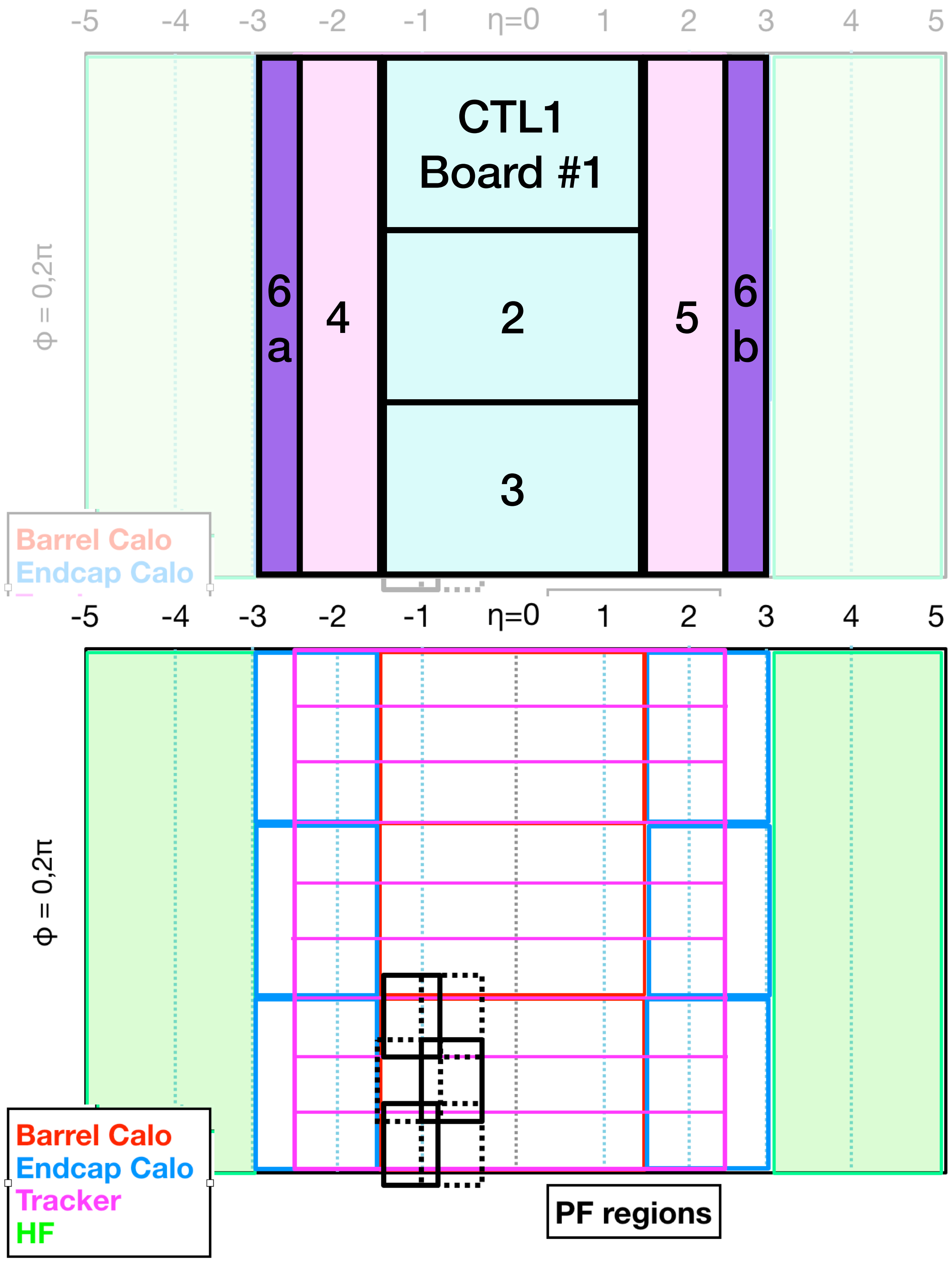


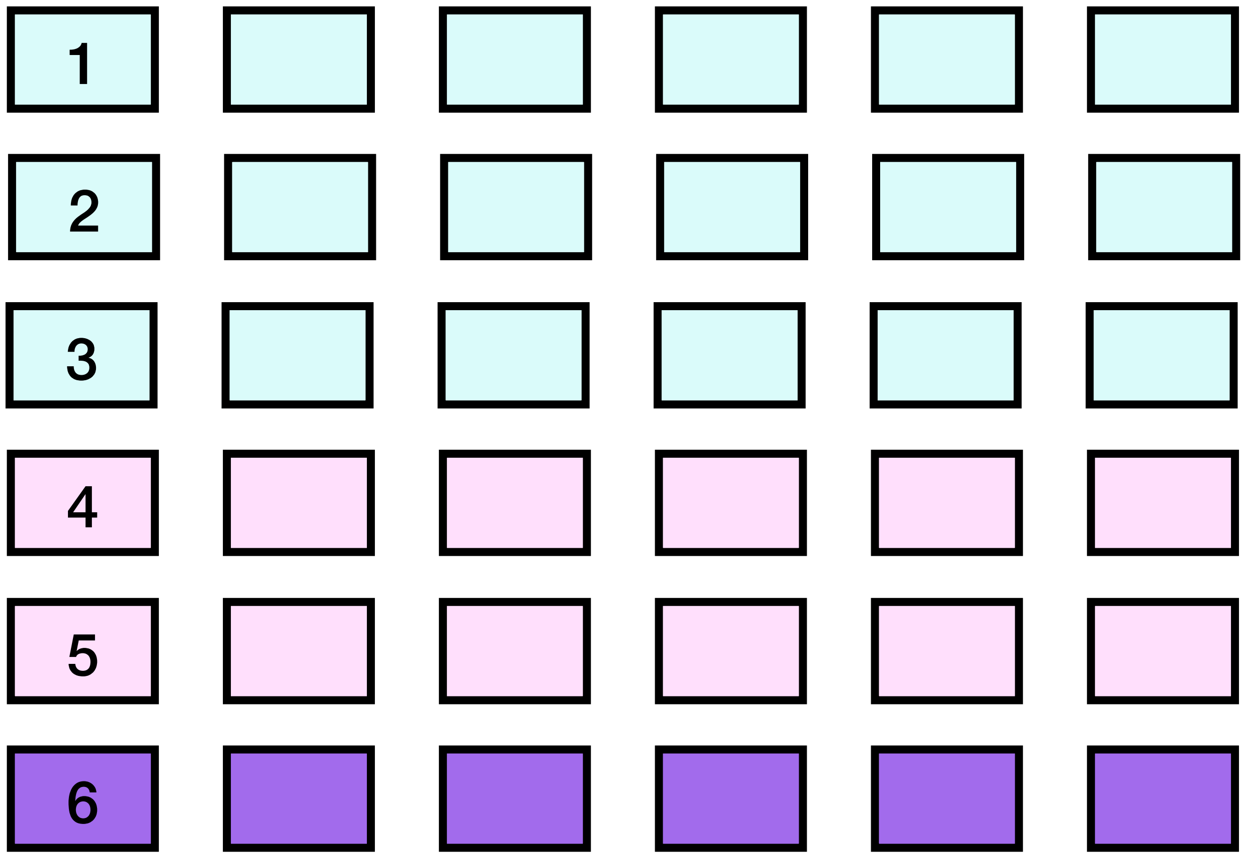


A highly parallel system

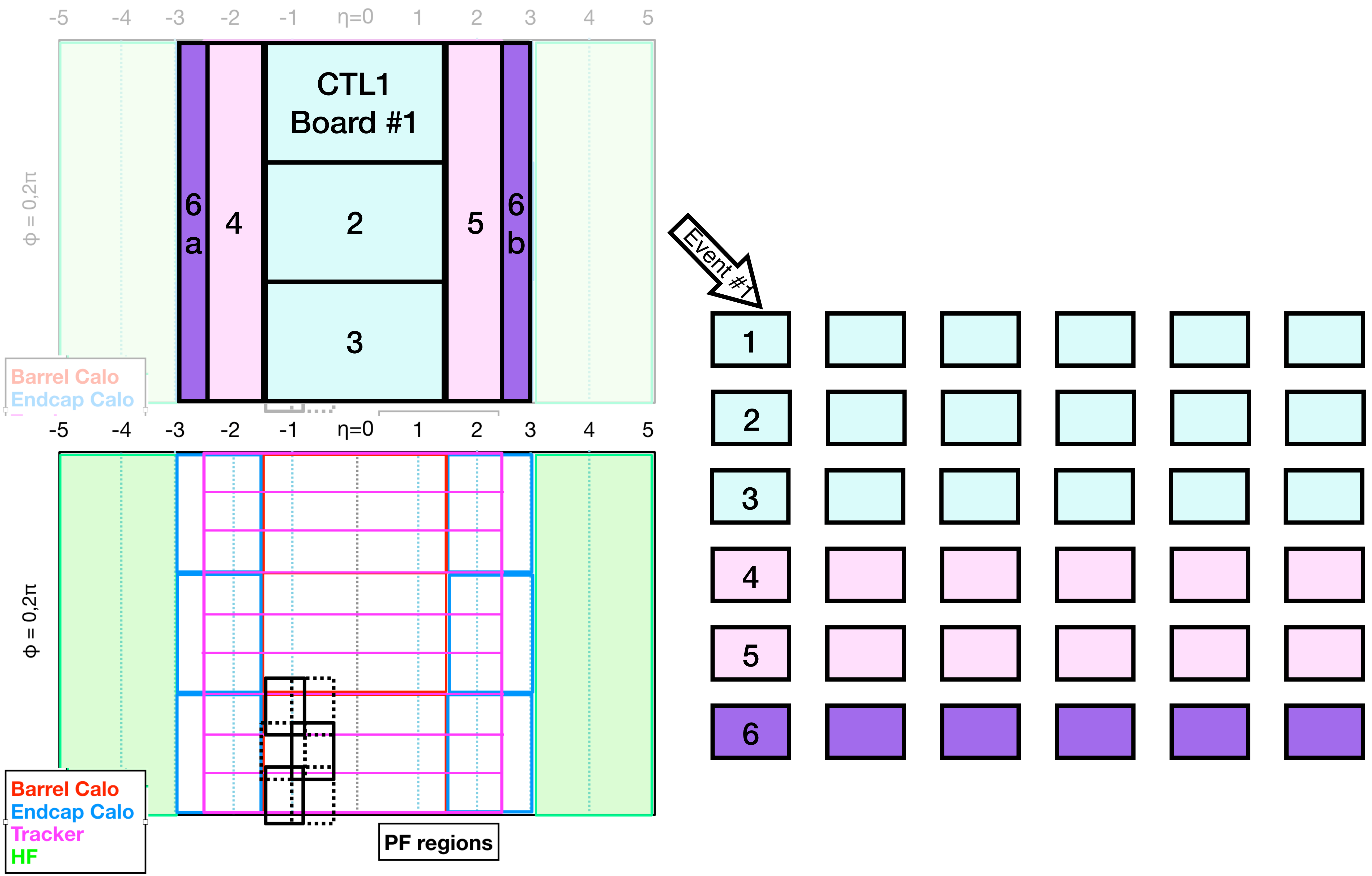


A highly parallel system

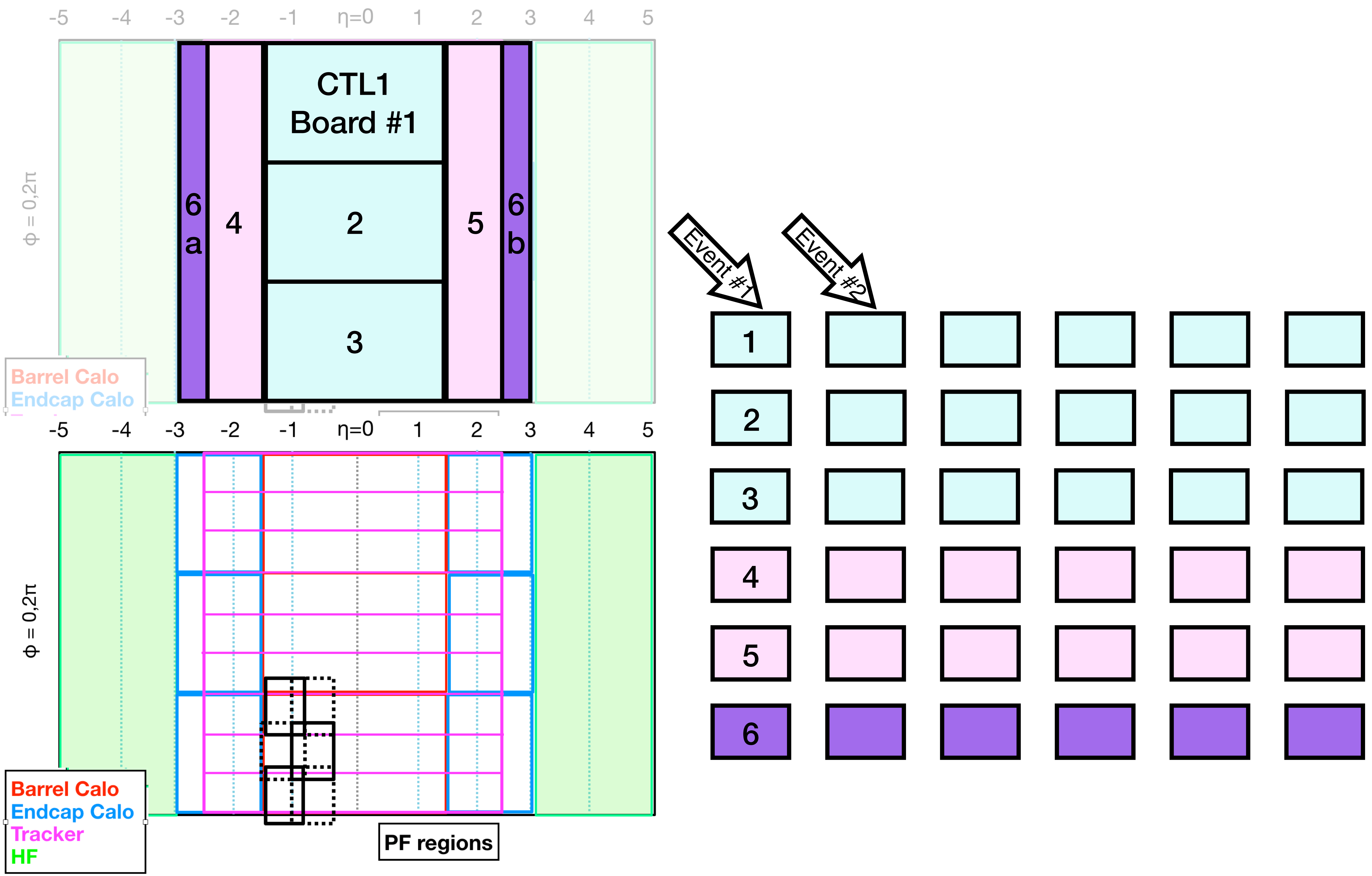




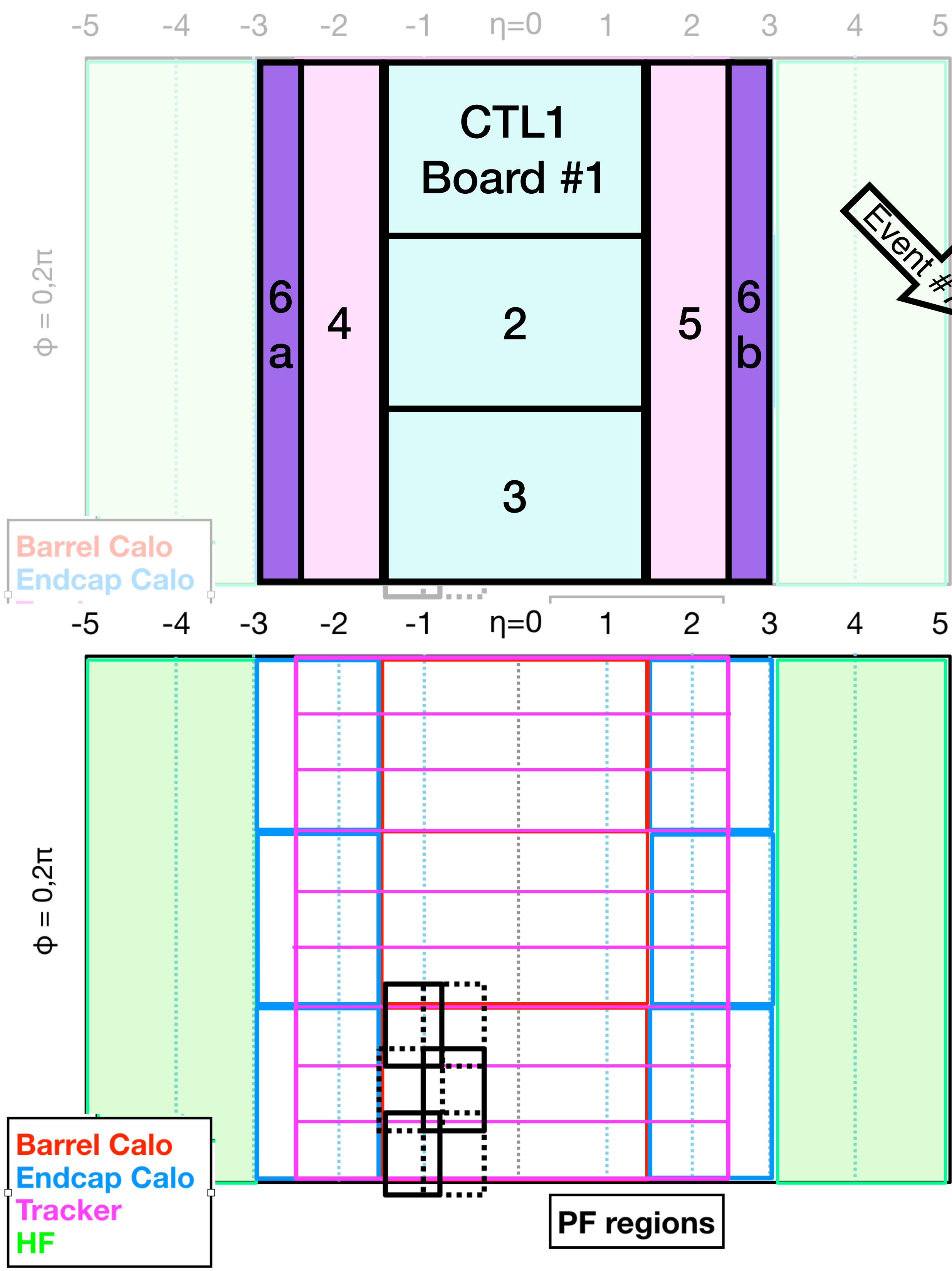
A highly parallel system

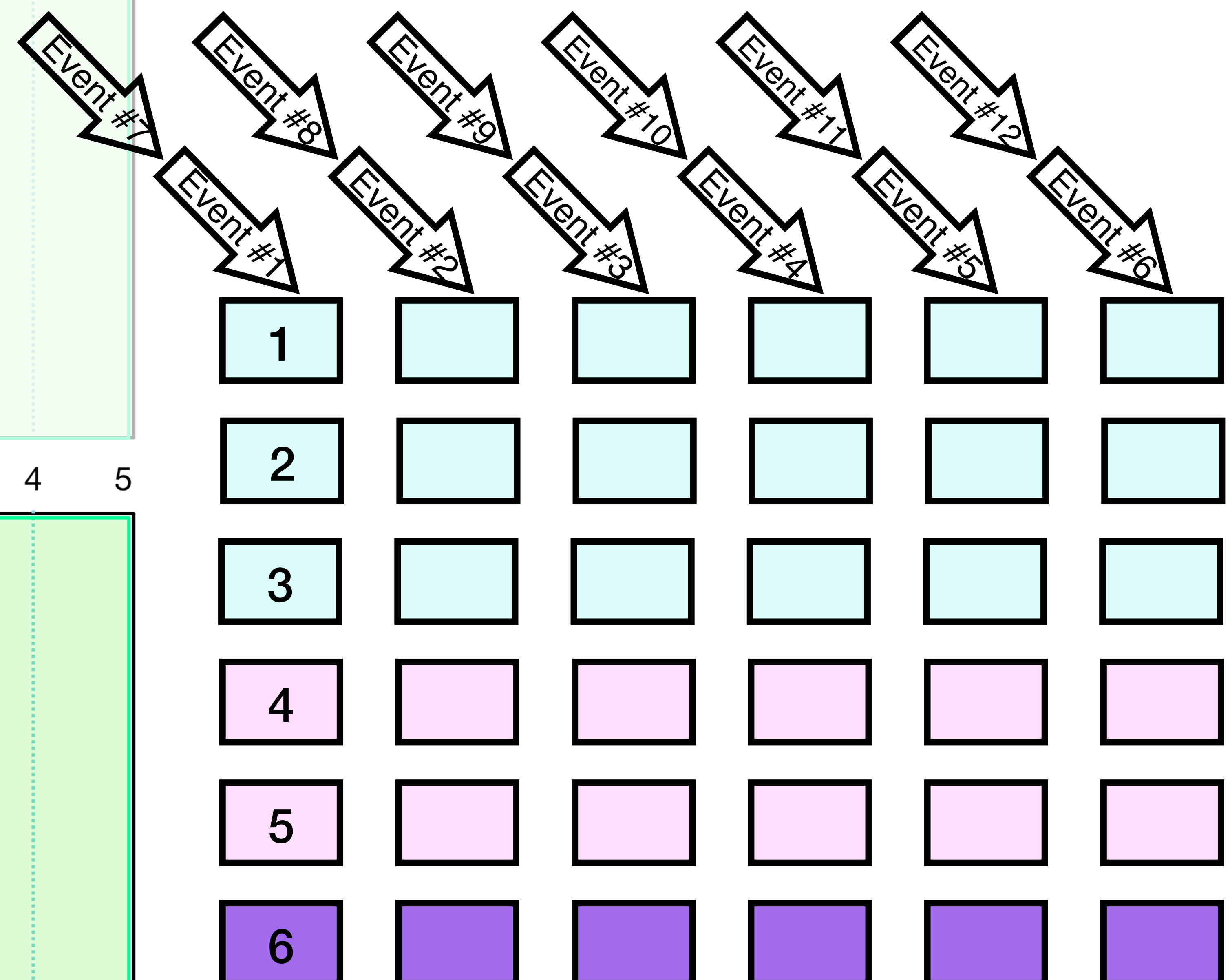


A highly parallel system



A highly parallel system



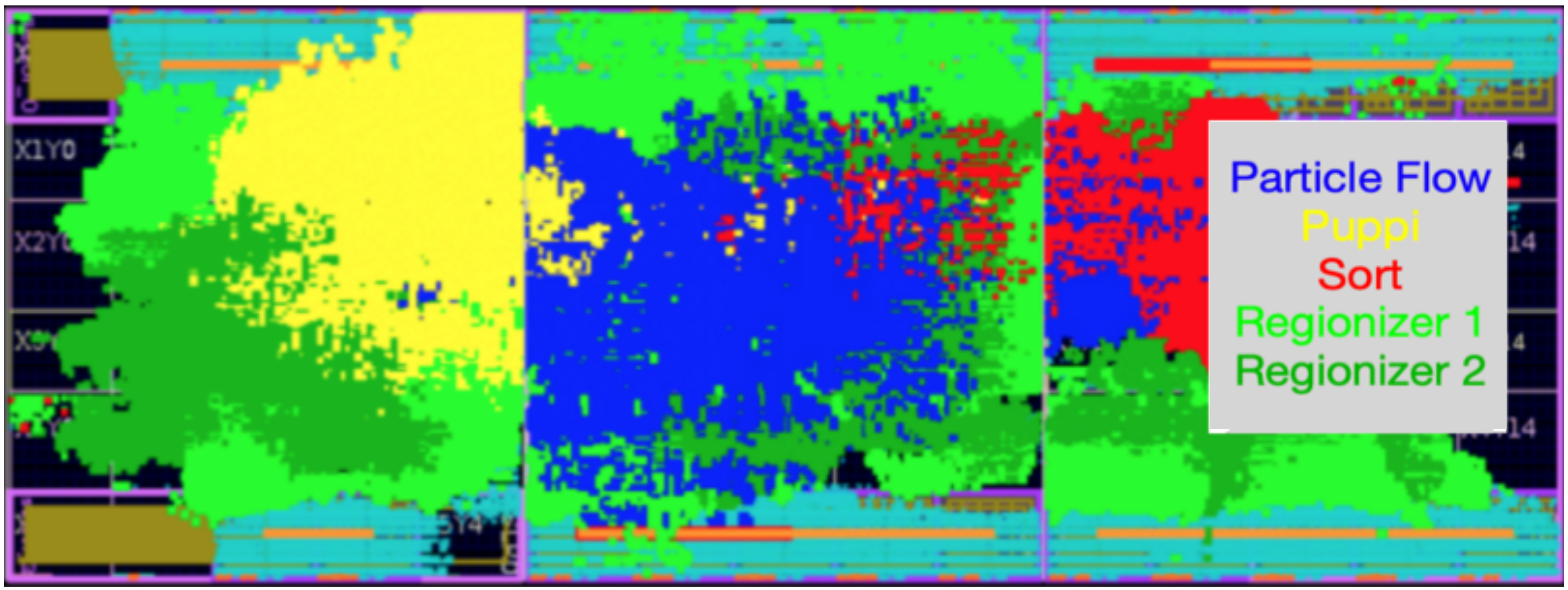


Programming an FPGA

"Firmware" specifies how the logic gates should be configured.

Custom language (vhdl/verilog) for concurrent signals.

Abstract logic → components → "place and route"



Since recently, can program in C with high level synthesis (HLS), significantly reducing barriers of time and expertise.

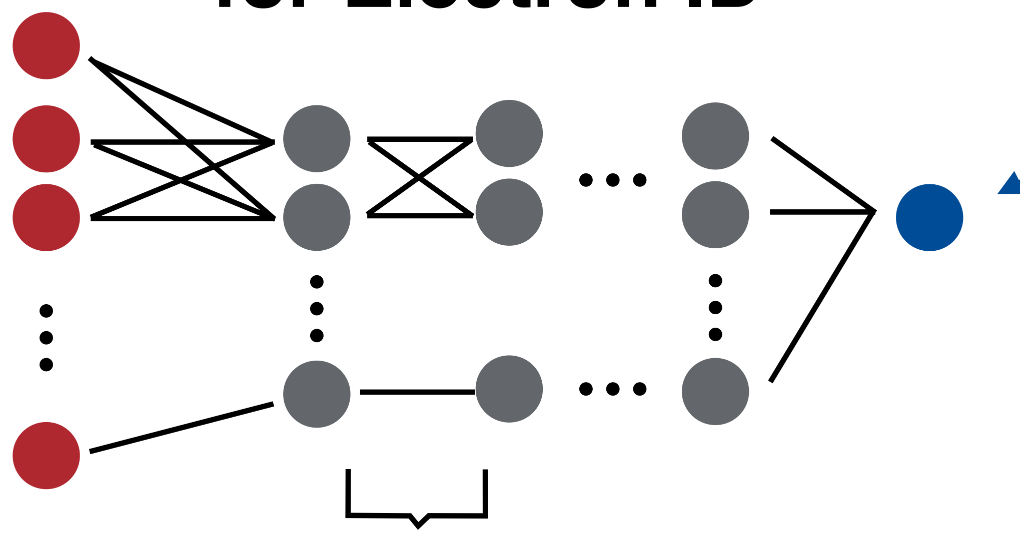
Aside: Machine learning on FPGAs

Machine Learning methods can unlock state-of-the-art performance.

From particle identification to full event selection

A Deep Neutral Network for Electron ID

Inputs:
Kinematics
Track quality
Shower shape
⋮
And so on…



Output: "Is Electron?"

Core of each NN "layer":
an N→M matrix multiplication

$$y_i = \sigma(w_{ij} x_i + b_i)$$

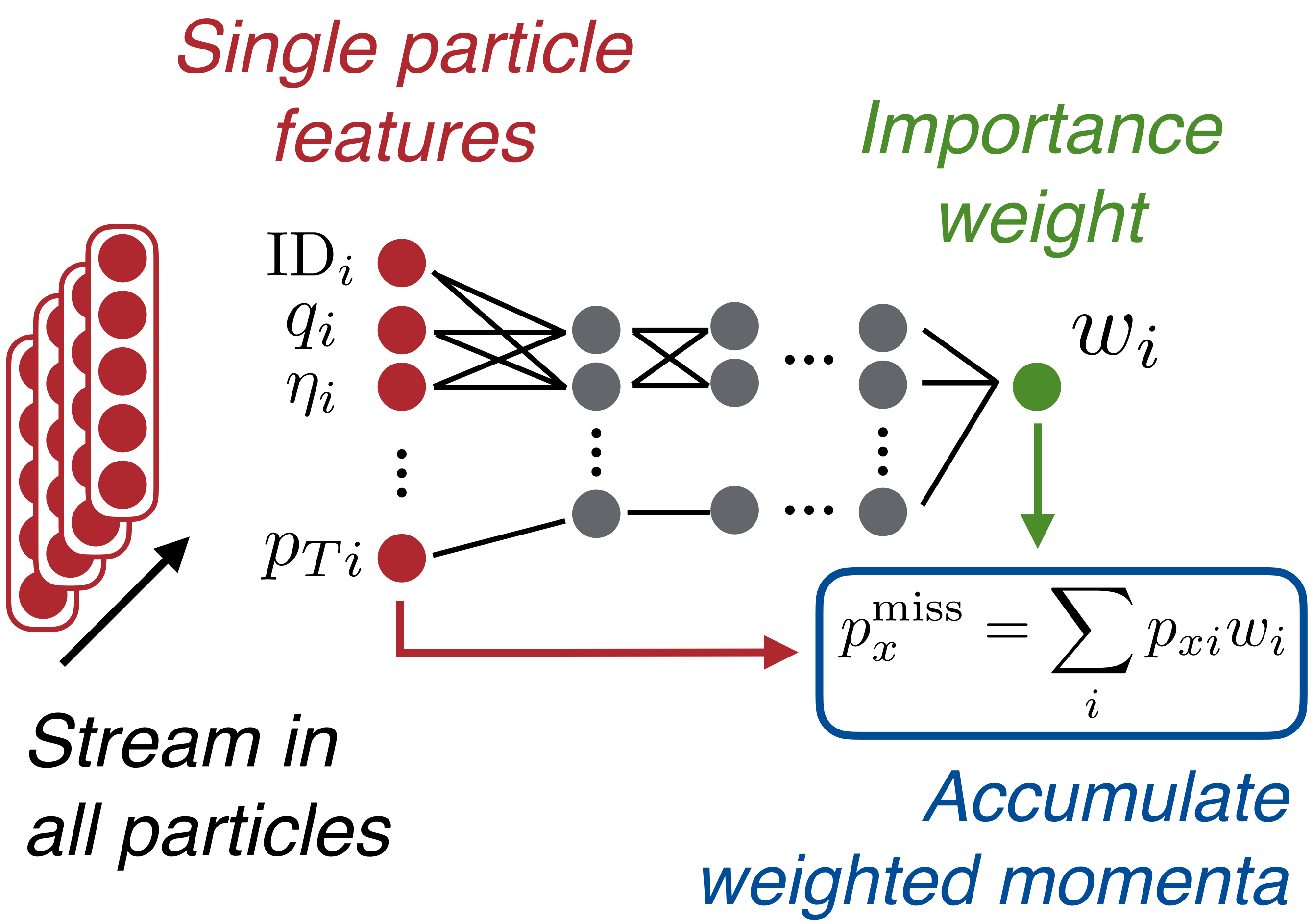
Non-linearity, e.g. $\sigma(x_i)=\max(x_i,0)$

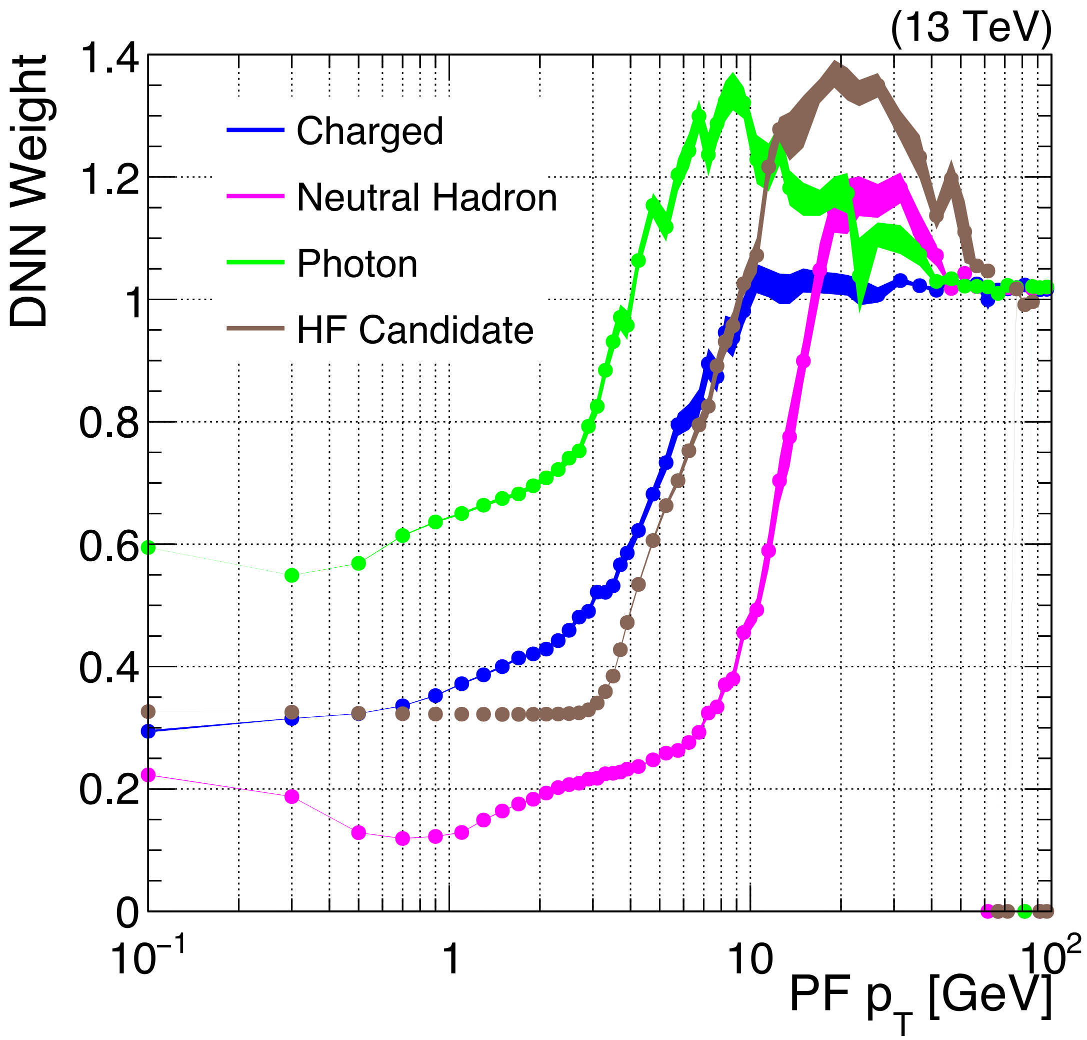
Matrix multiplication

hls4ml, for global event interpretation

After reconstructing all ~100 particles, how can we optimally use them?

- ML algorithms (like neural networks) offer one solution





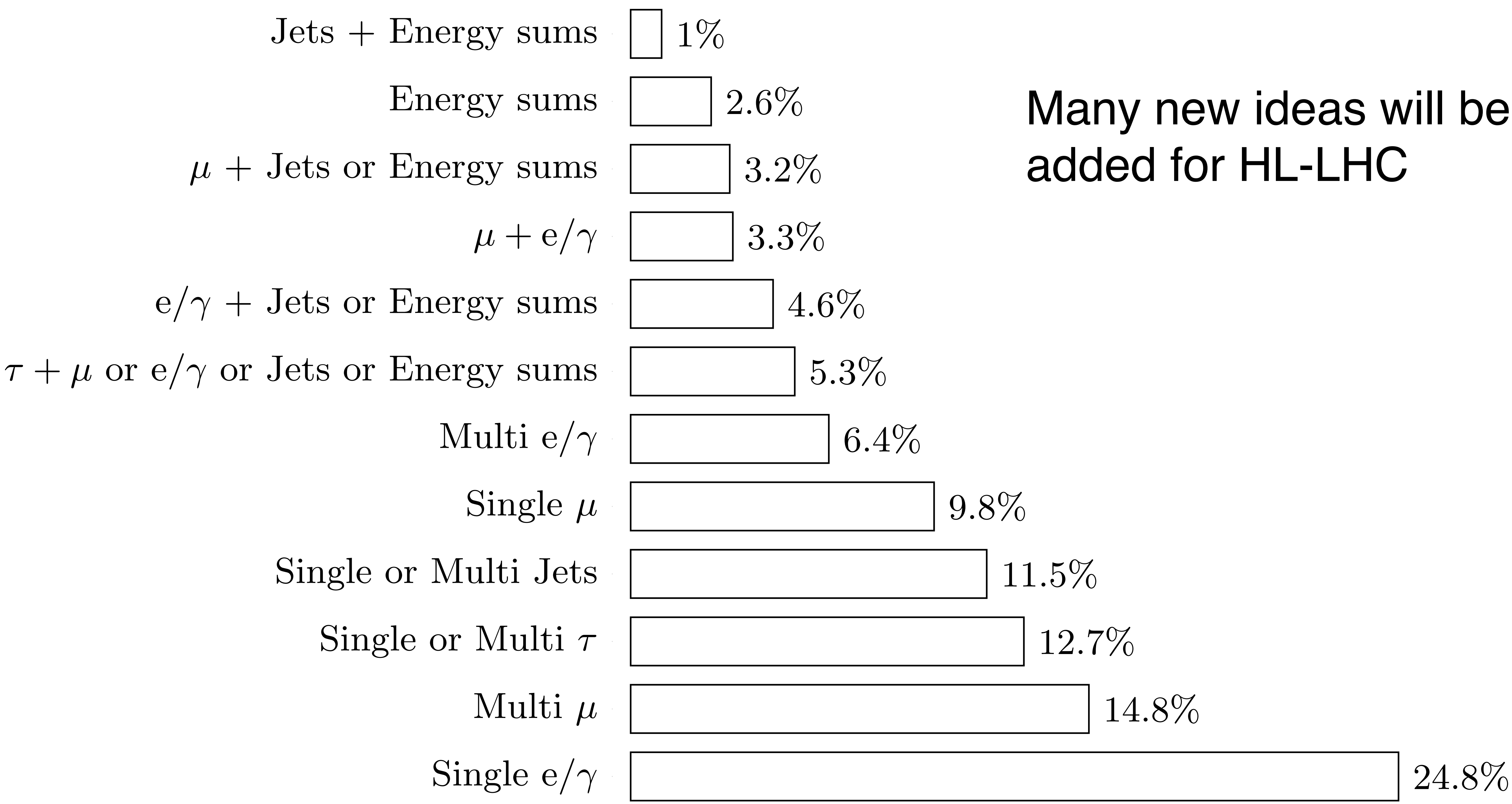
Offline model customized for trigger application (e.g. model size, precision).

- Challenge: repeating calculation for all 100 particles / event!

hls4ml

Level-1 Trigger Menu

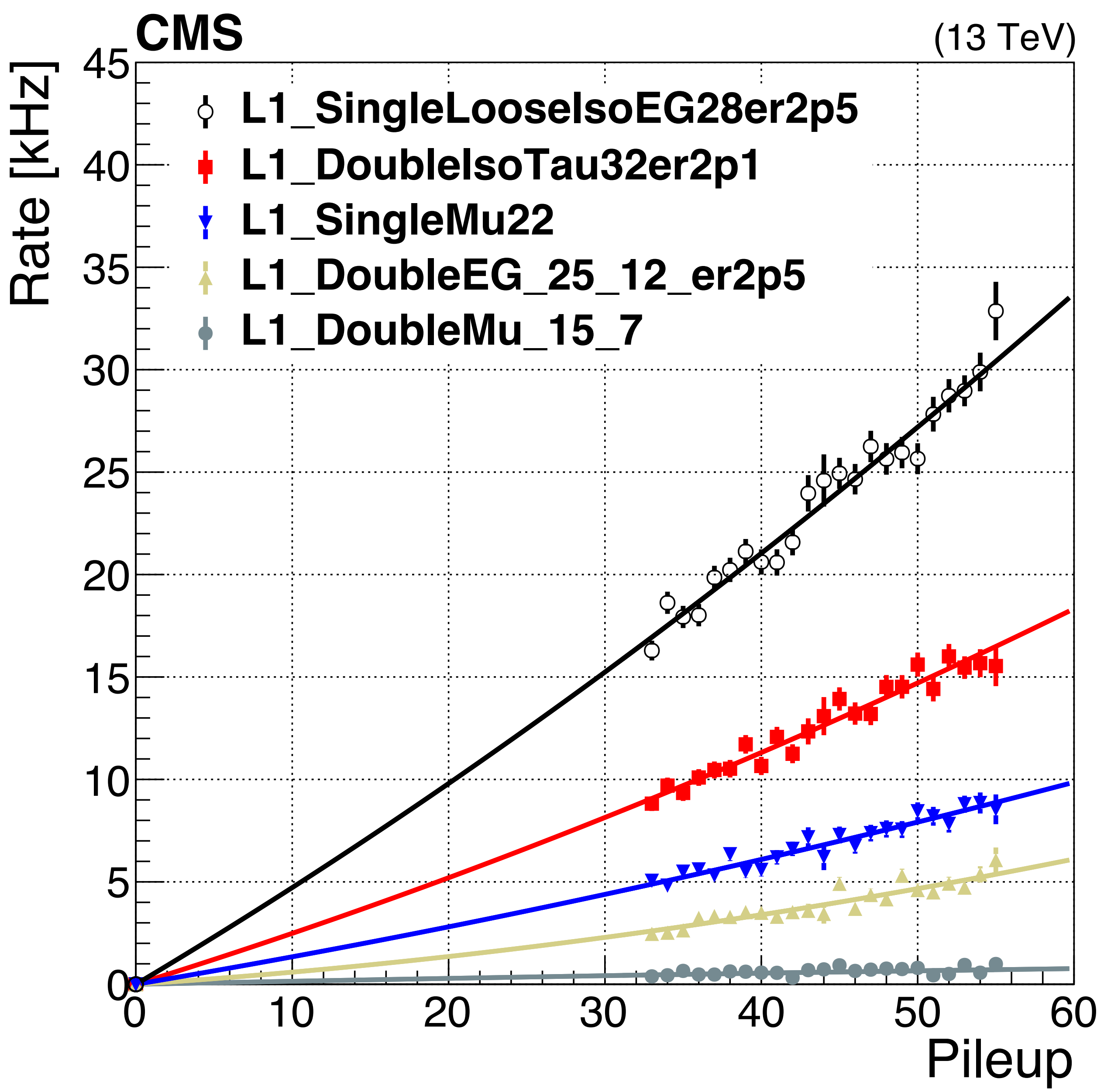
Categories of trigger path occupying the largest rate for Run 2 CMS.

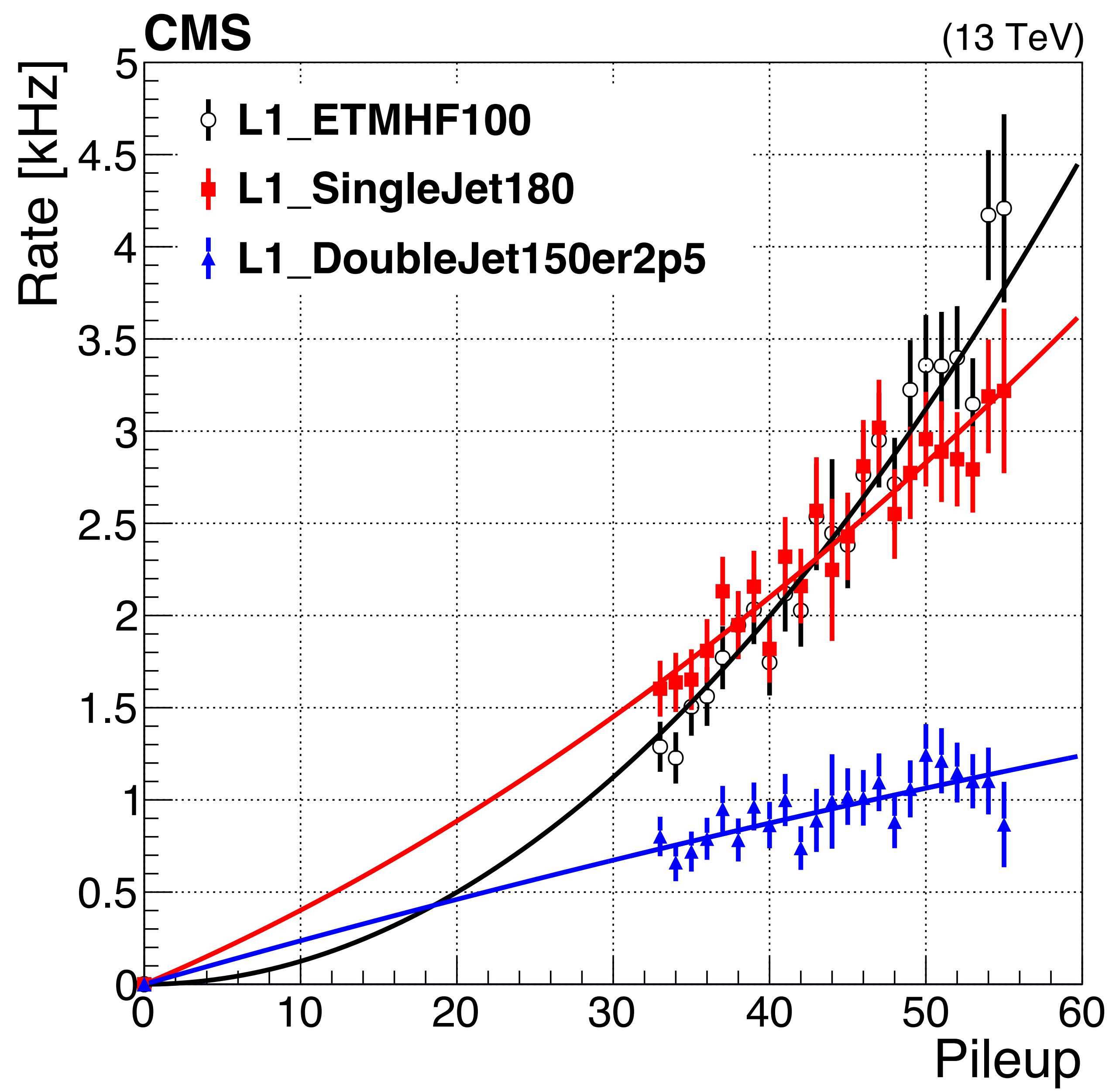


Many new ideas will be added for HL-LHC

Level-1 trigger rates

Menus with different energy thresholds target different inst. luminosity.
Ideally the rate of accepted events scales linearly with pileup.

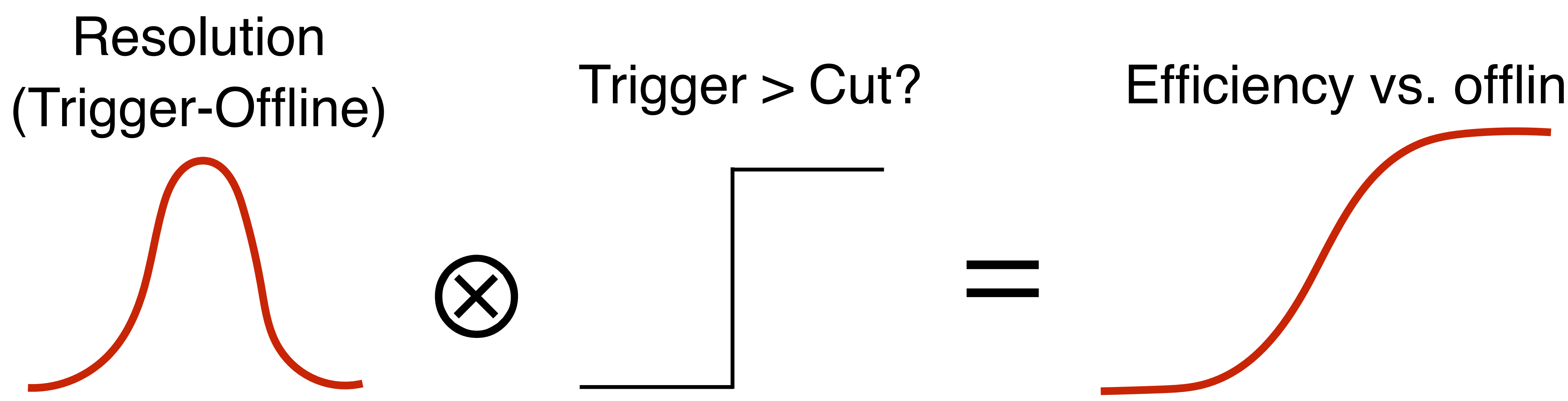




Software Triggers

Aim for reconstruction as close as possible to offline

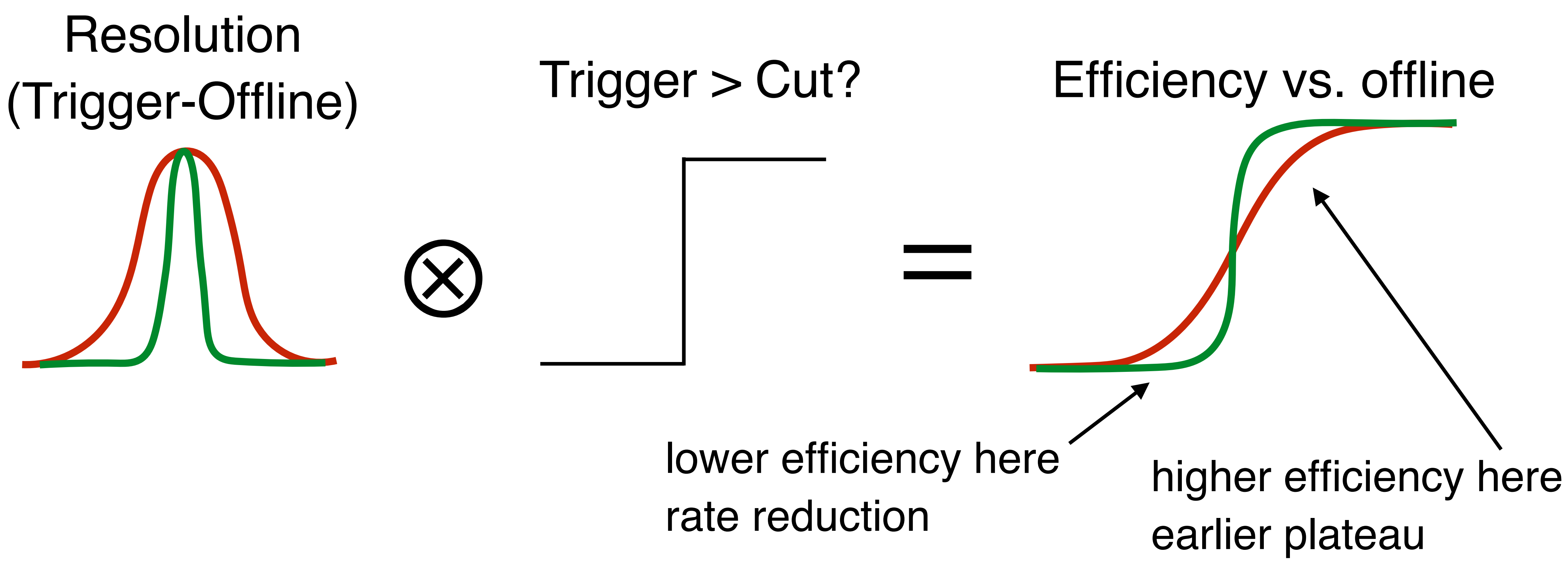
Lower thresholds and reduces systematic biases for analysis



Software Triggers

Aim for reconstruction as close as possible to offline

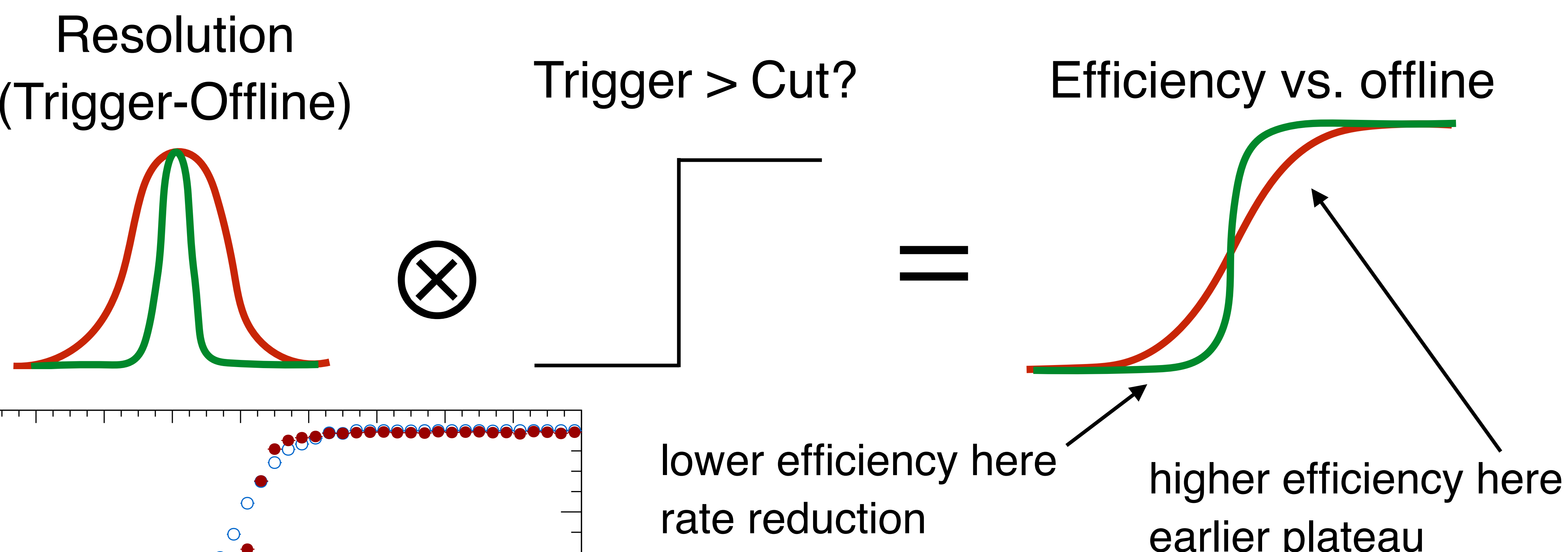
Lower thresholds and reduces systematic biases for analysis

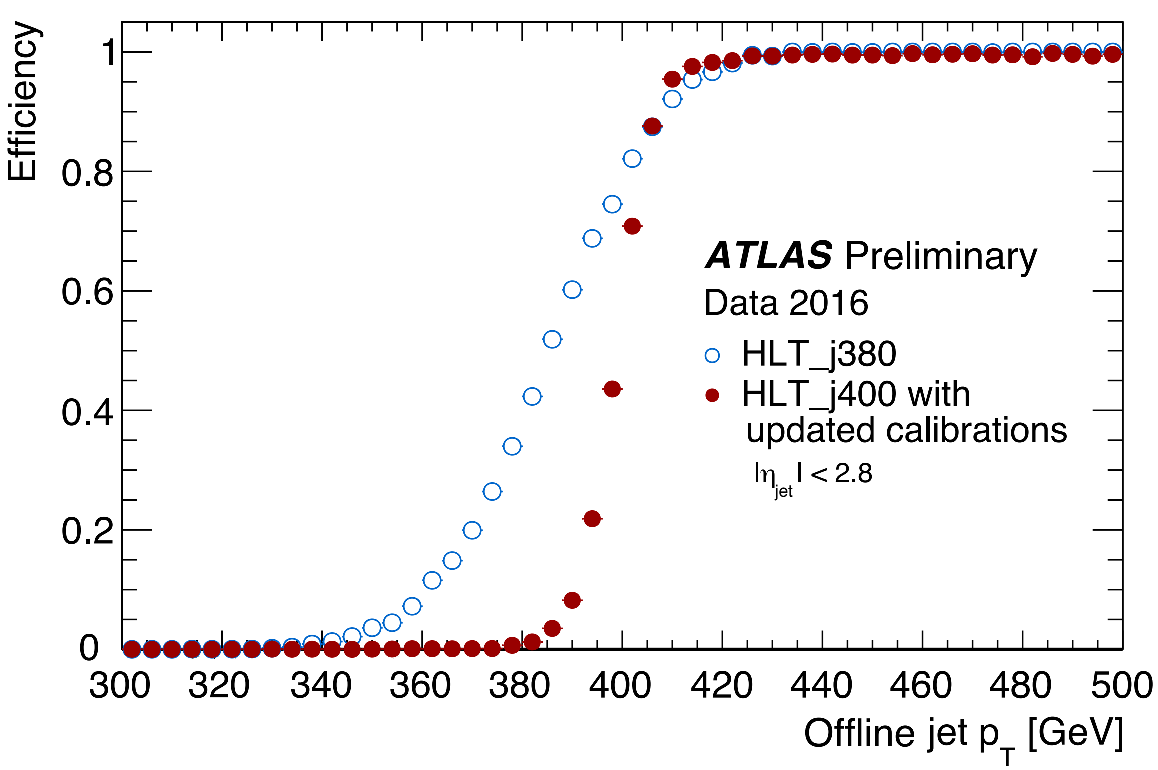


Software Triggers

Aim for reconstruction as close as possible to offline

Lower thresholds and reduces systematic biases for analysis

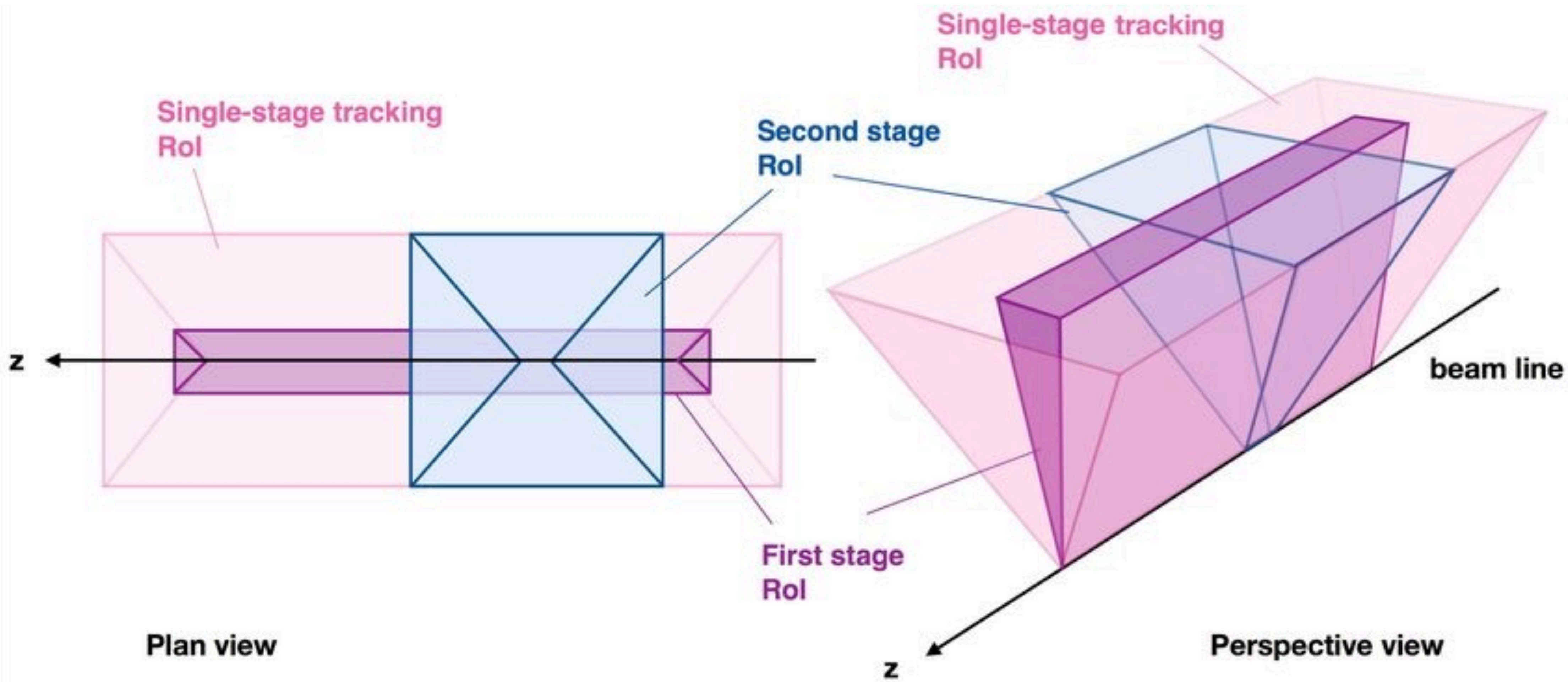




E.g. jets performance before/after the full calibration was ported to HLT

SW trigger tracking at ATLAS

At ATLAS/CMS today, software trigger relies heavily on tracking.
Offline tracking is computationally expensive, how to reduce?

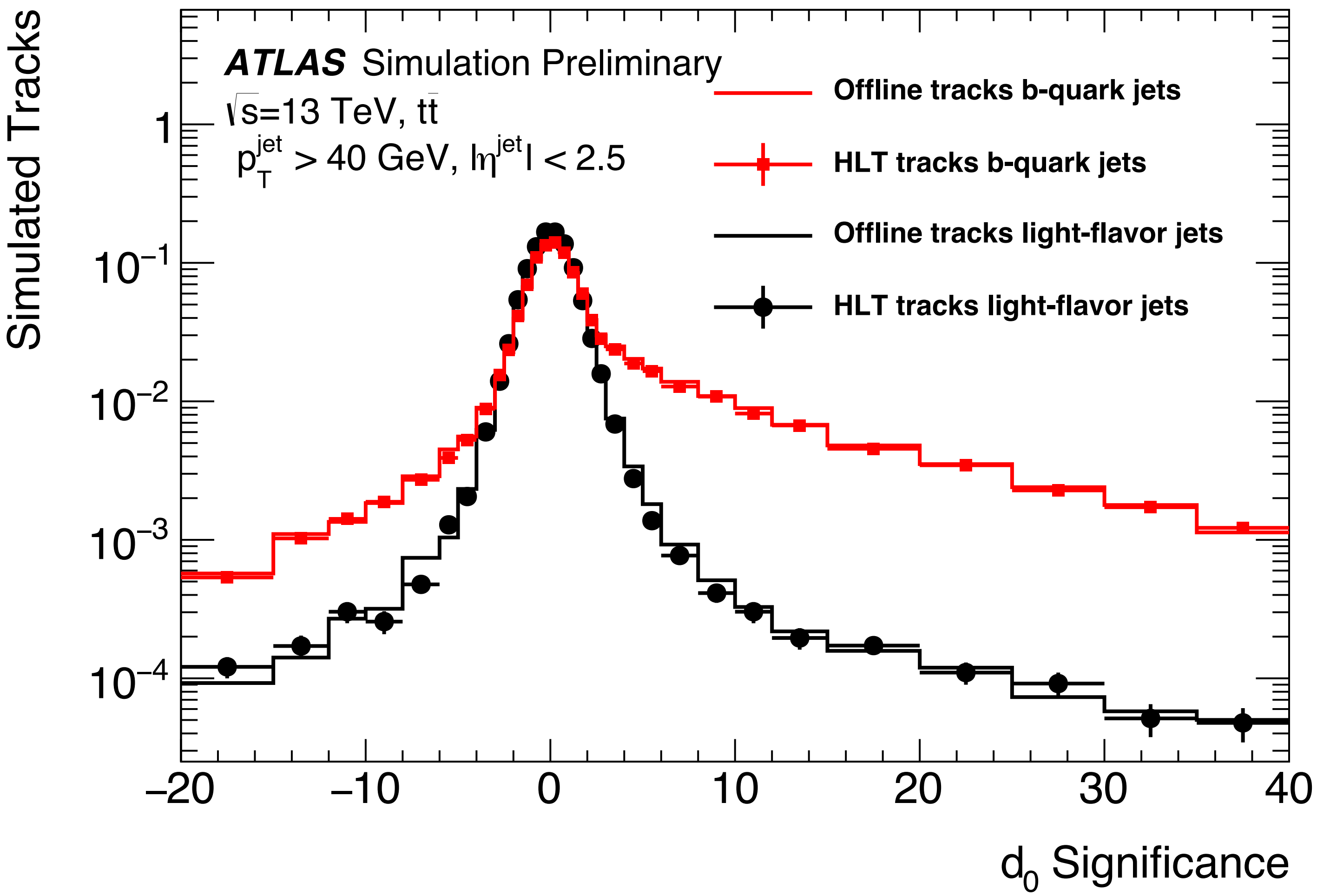


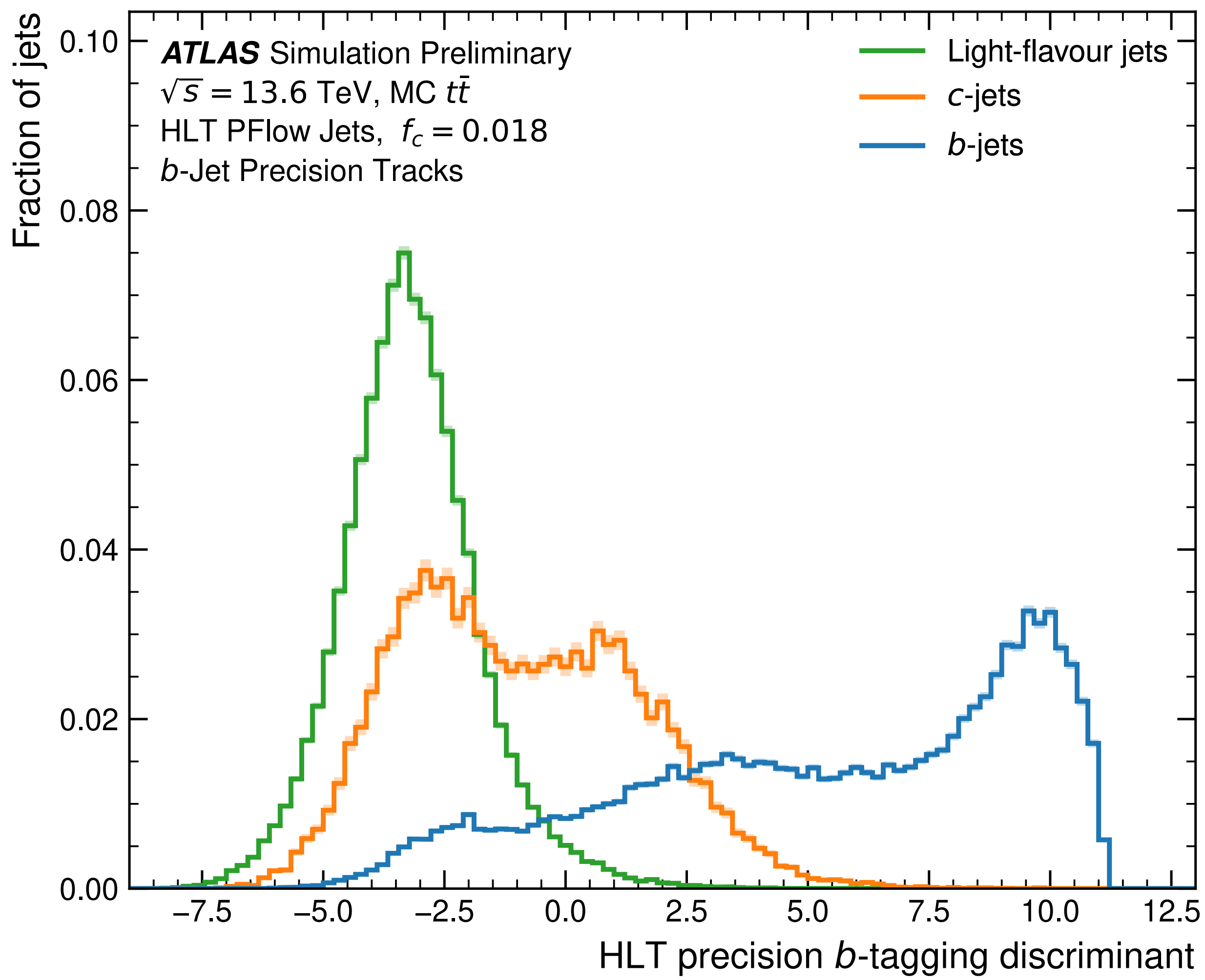
ATLAS breaks task into “speed” and “precision” stages, within ROIs.

SW trigger tracking at ATLAS

At ATLAS/CMS today, software trigger relies heavily on tracking.

Offline tracking is computationally expensive, how to reduce?



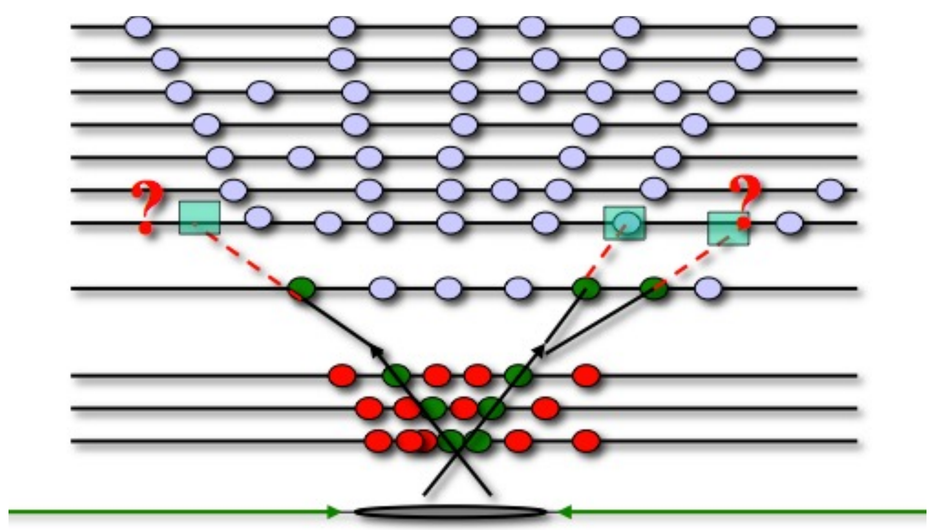


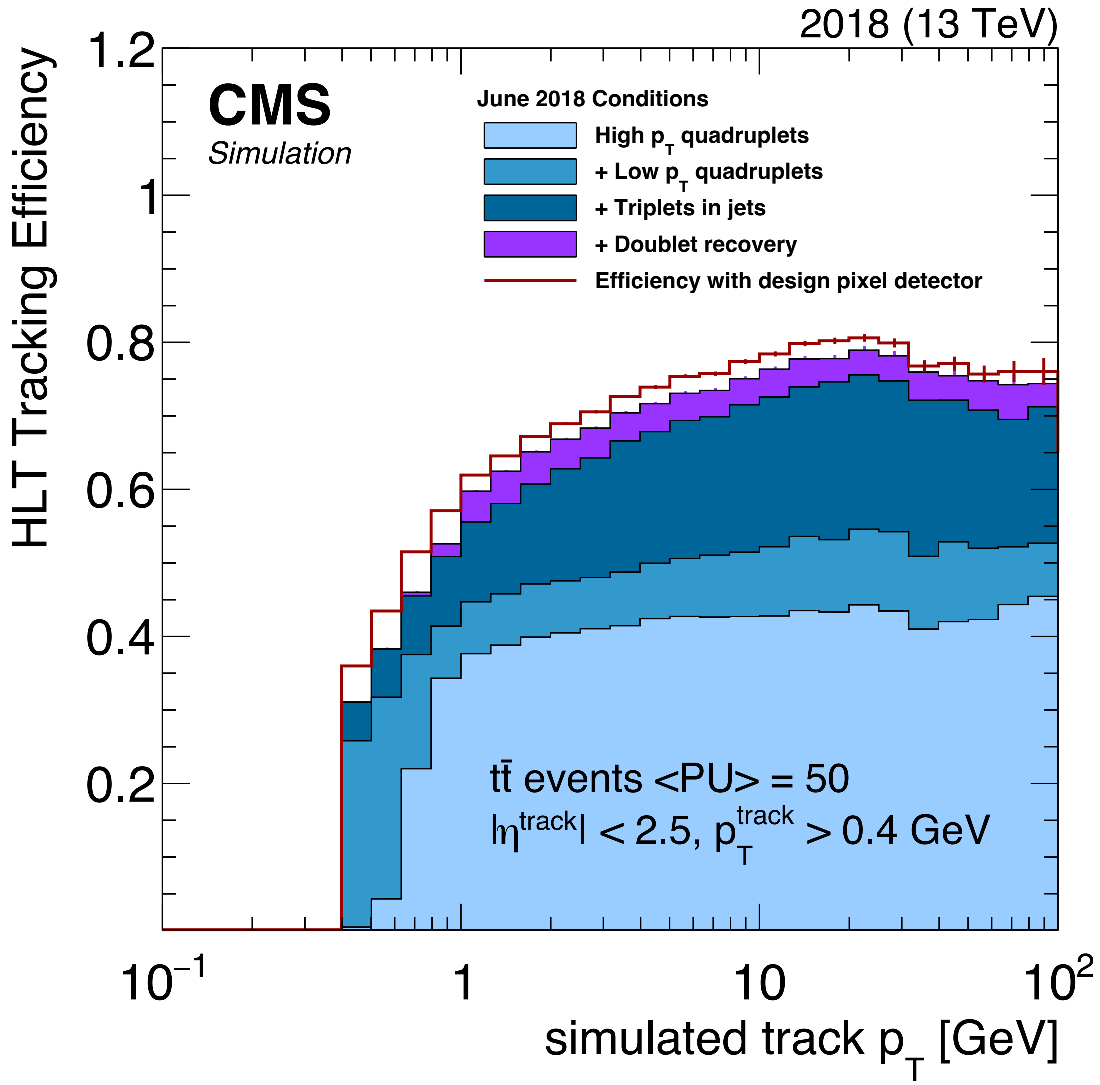
HLT accurately measures impact parameters, allows online b-tagging

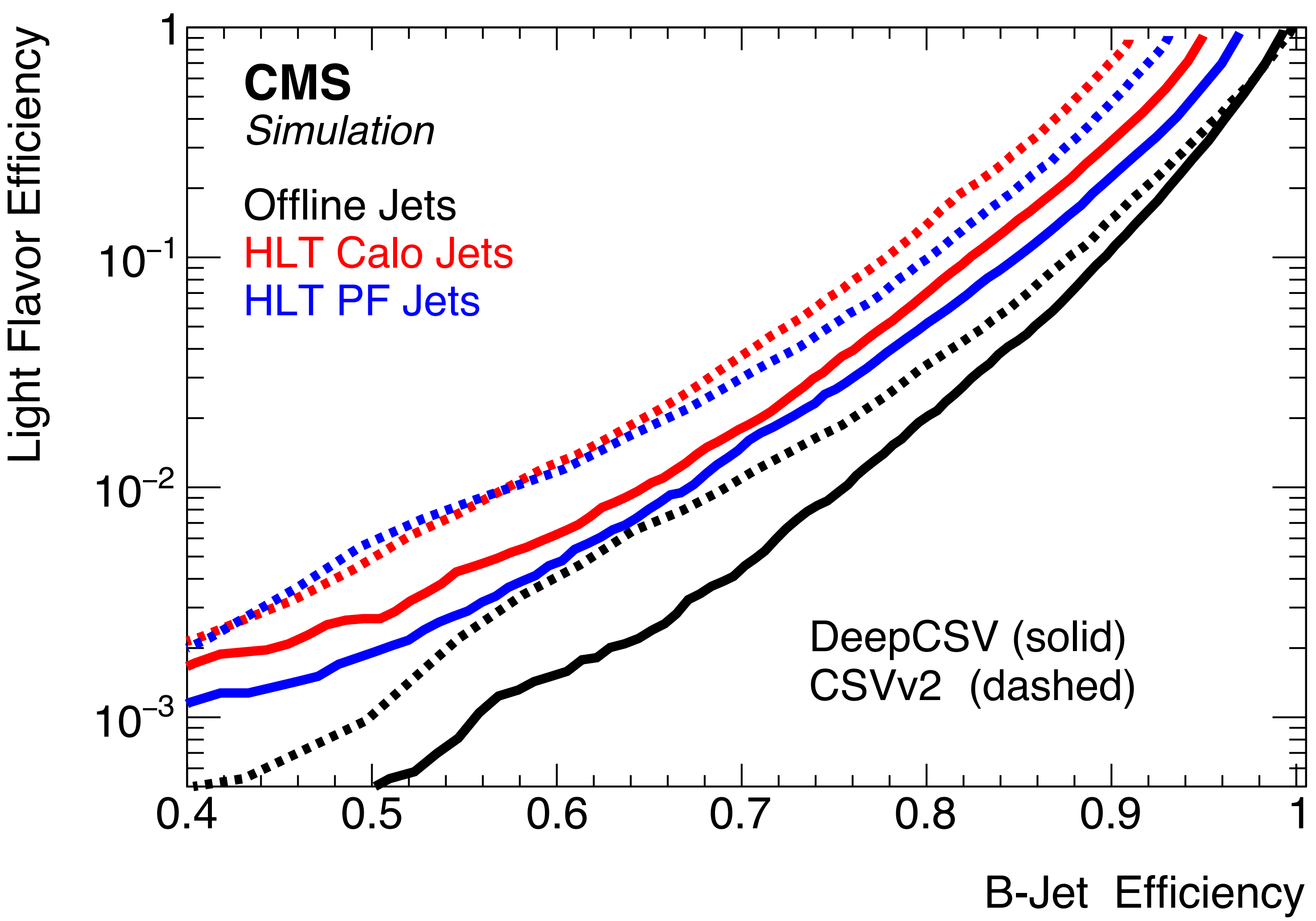
SW trigger tracking at CMS

In Run 2, a multi-stage approach was used.

Find "easy" tracks first, remove hits, and loosen validity window for the next stage.

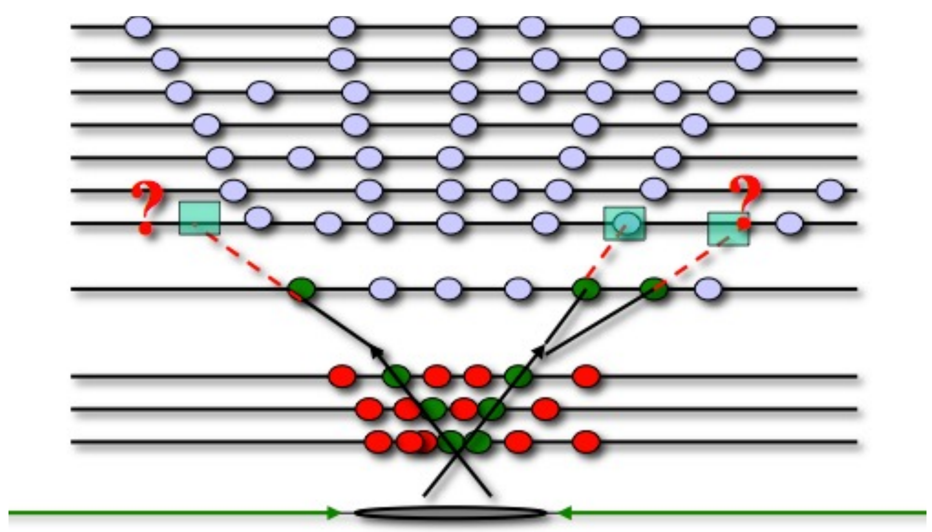


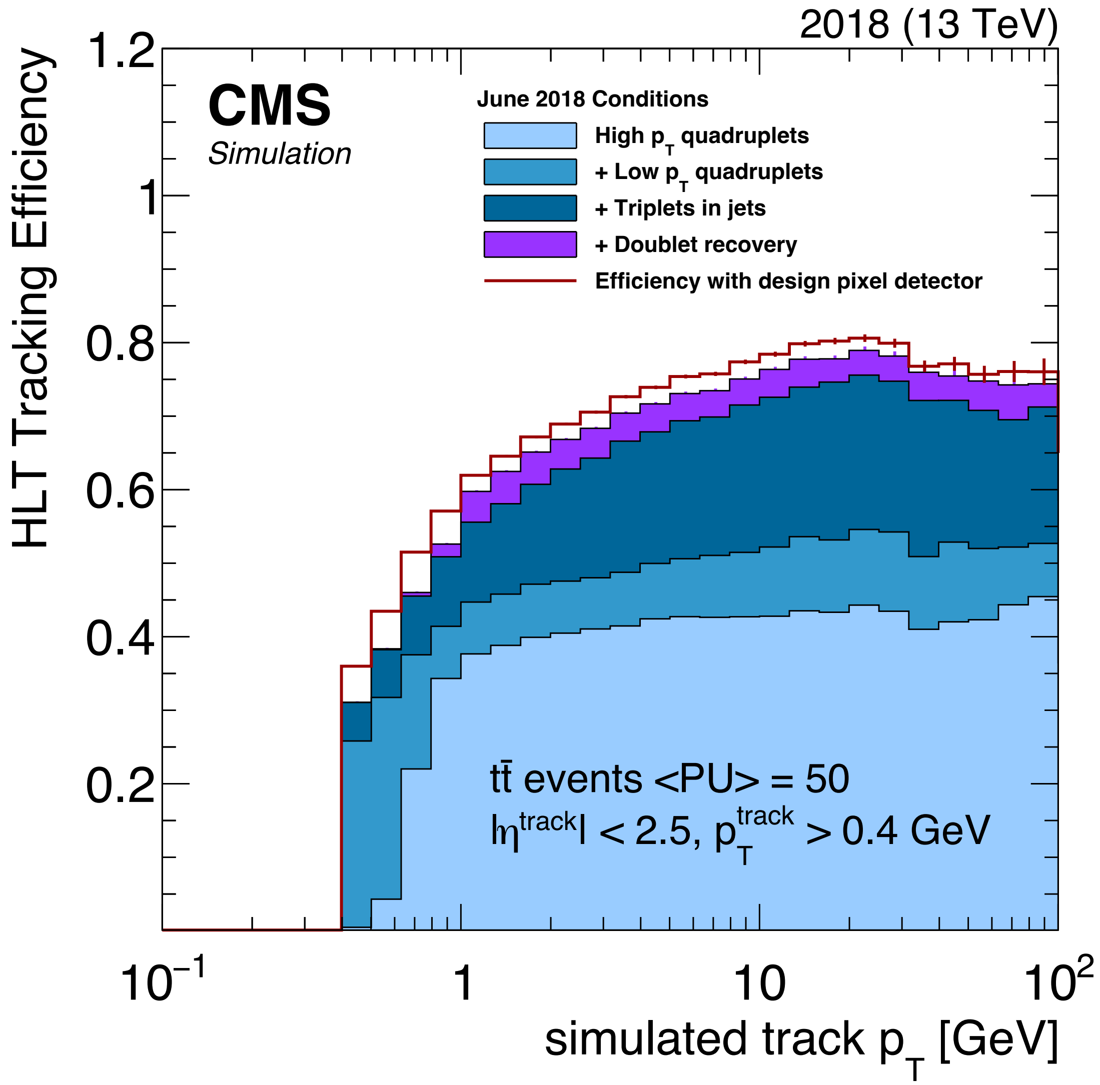


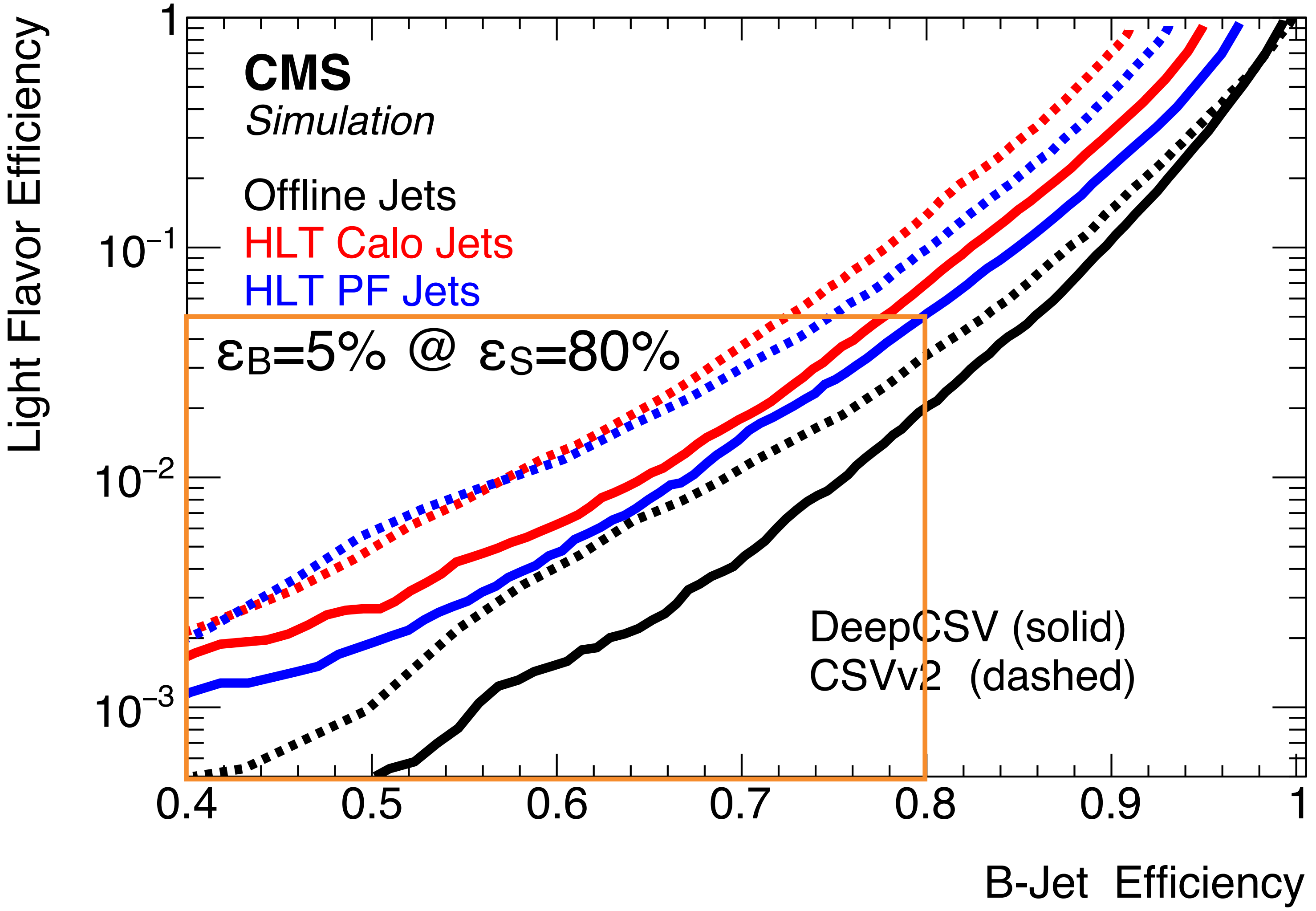


SW trigger tracking at CMS

In Run 2, a multi-stage approach was used.
- Find “easy” tracks first, remove hits, and loosen validity window for the next stage.





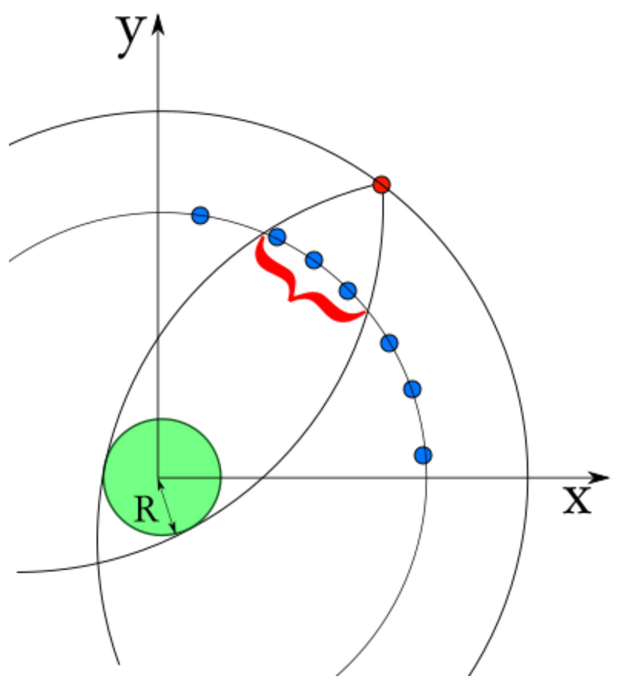


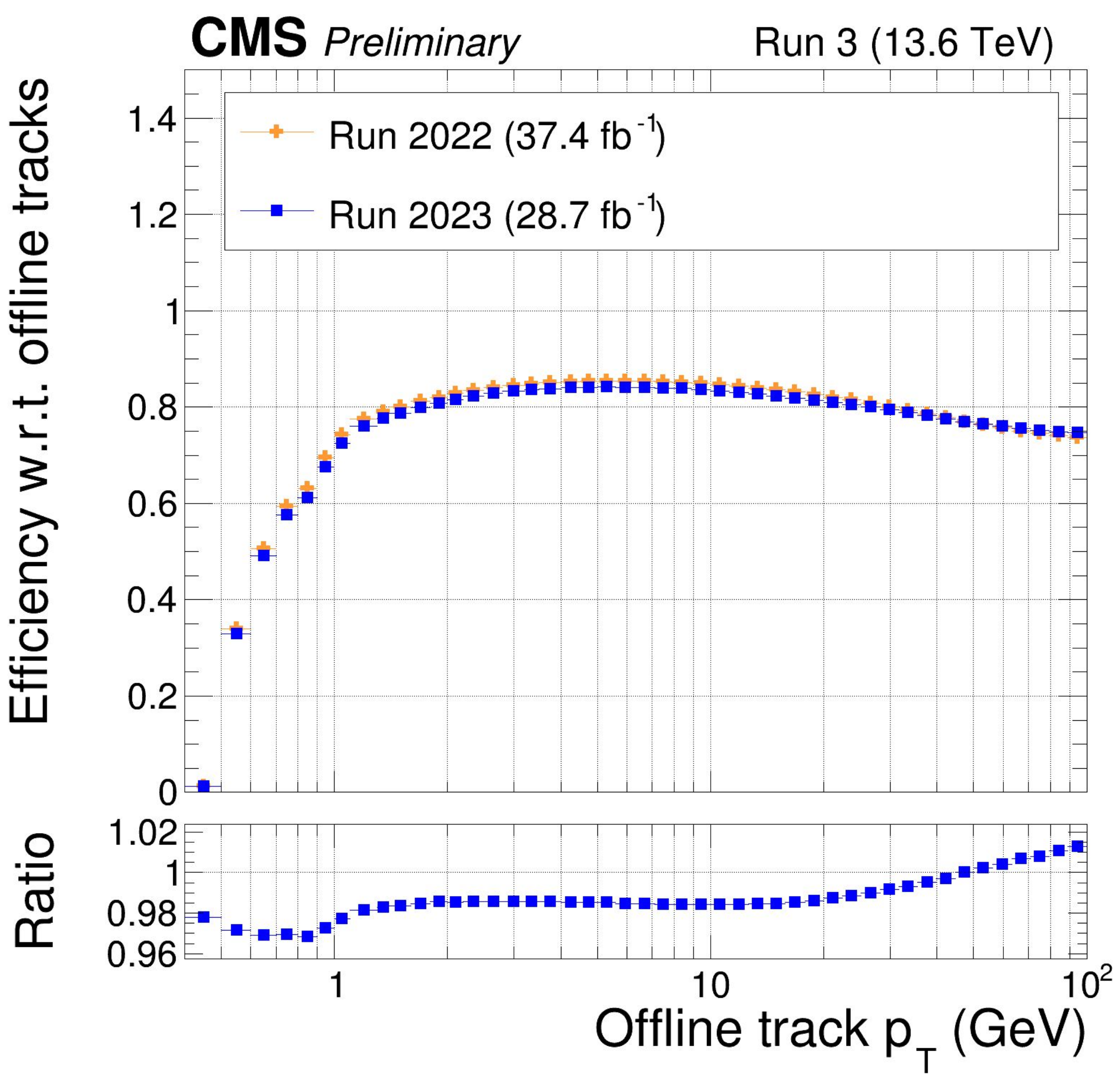
SW trigger tracking at CMS

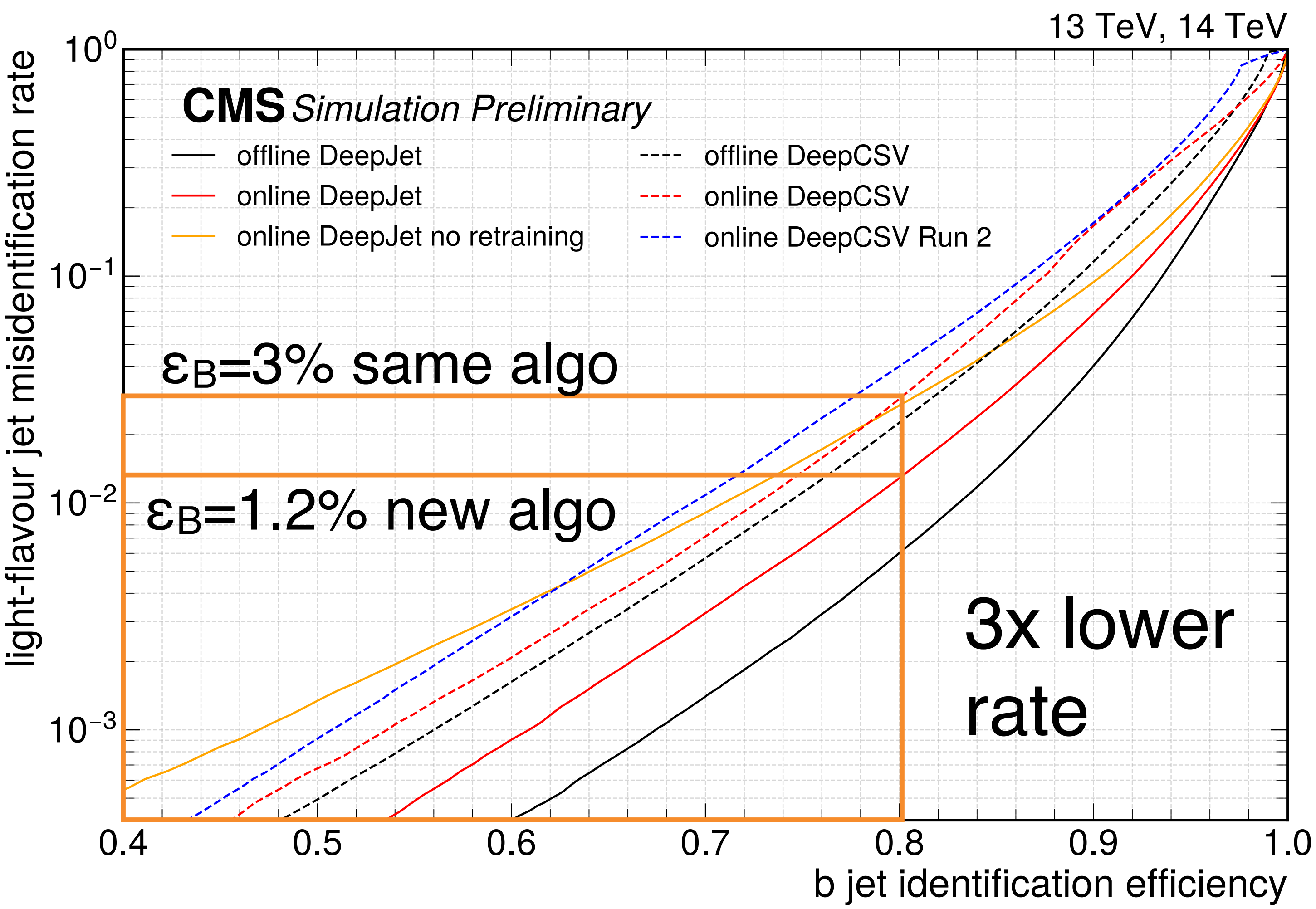
For Run 3, much of the task is offloaded to GPUs.

Hit unpacking, clustering, and "pixel track" formation

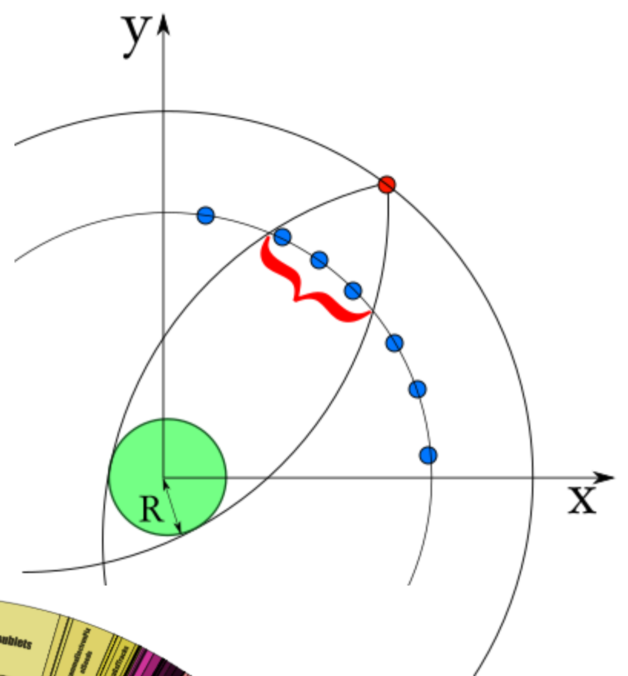
Pixel tracks (3+ hits) seed a single-stage approach.





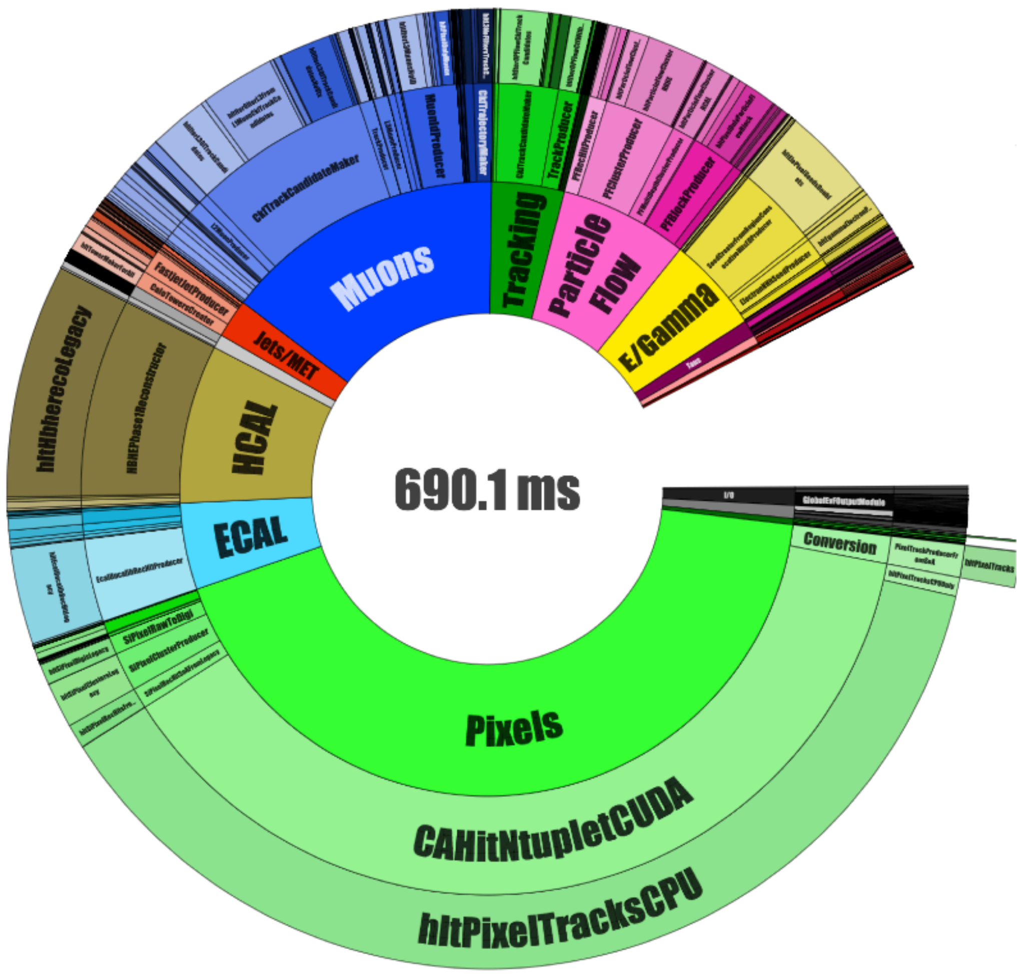


SW trigger tracking at CMS

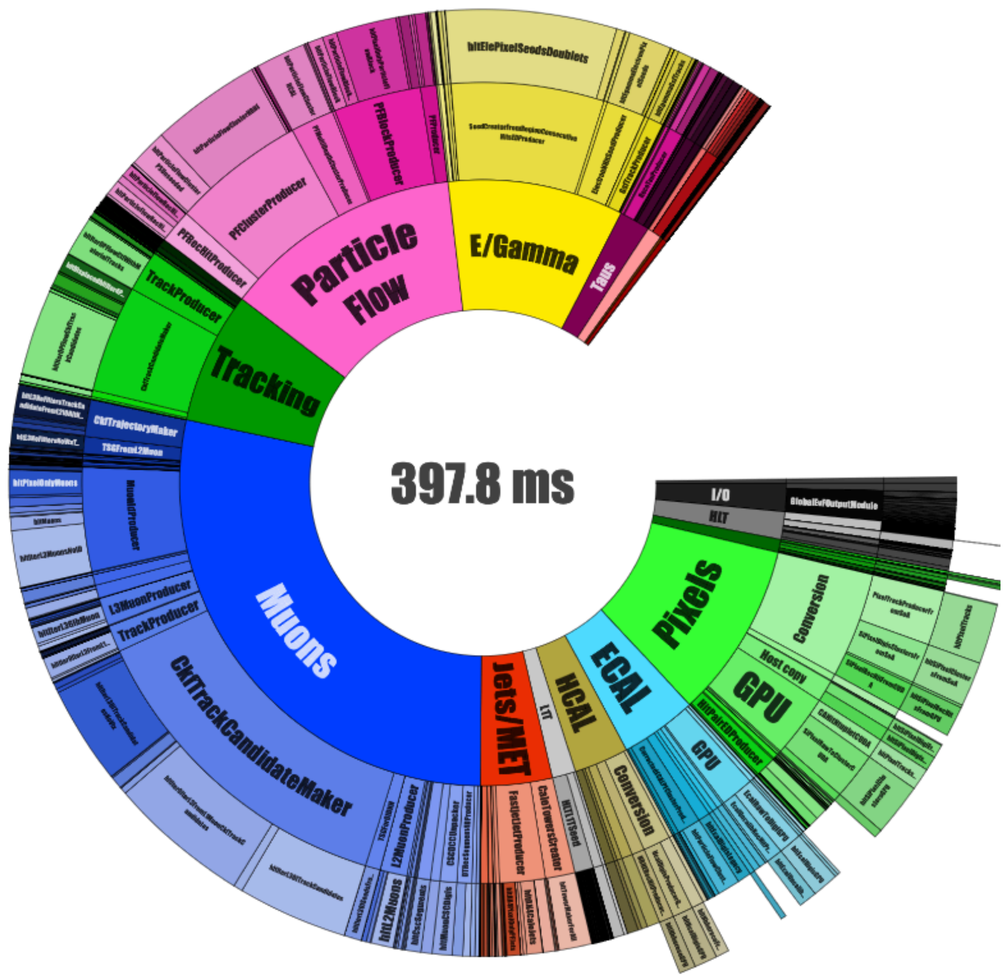


For Run 3, much of the task is offloaded to GPUs.
  Hit unpacking, clustering, and “pixel track” formation

Pixel tracks (3+ hits) seed a single-stage approach.



CPU only



CPU+GPU only

High-Level Trigger Menu

Example from CMS Run 2

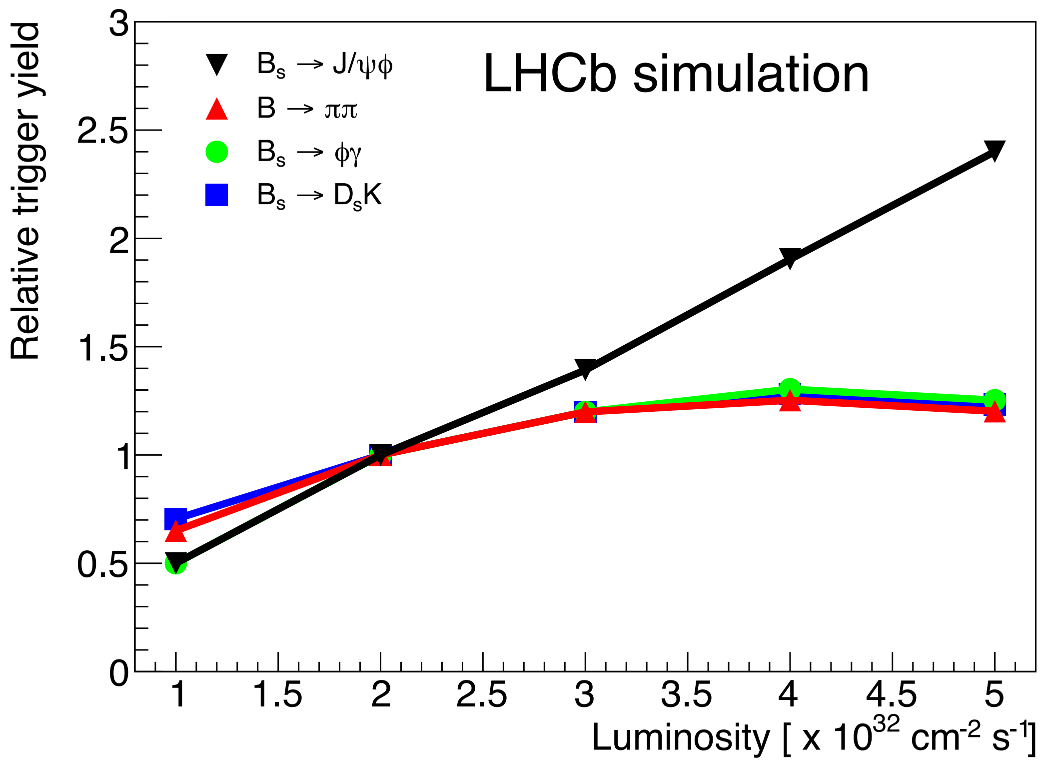
HLT path	L1 thresholds [GeV]	HLT thresholds [GeV]	Rate [Hz]
Single muon	22	50	49
Single muon (isolated)	22	24	230
Double muons	–	37, 27	16
Double muons (isolated)	–	17, 8	32
Single electron (isolated)	30	32	180
Double electrons	25, 12	25, 25	16
Double electrons (isolated)	25, 12	23, 12	32
Single photon	30	200	16
Single photon (isolated), barrel only	30	110	16
Double photons	25, 12	30, 18	32
Single tau	120	180	16
Double taus	32	35, 35	49
Single jet	180	500	16
Single jet with substructure	180	400	32
Multijets with b-tagging	$H_T > 320$ jets $> 70, 55, 40, 40$	$H_T > 330$ jets $> 75, 60, 45, 40$	16
Total transverse momentum	360	1 050	16
Missing transverse momentum	100	120	49

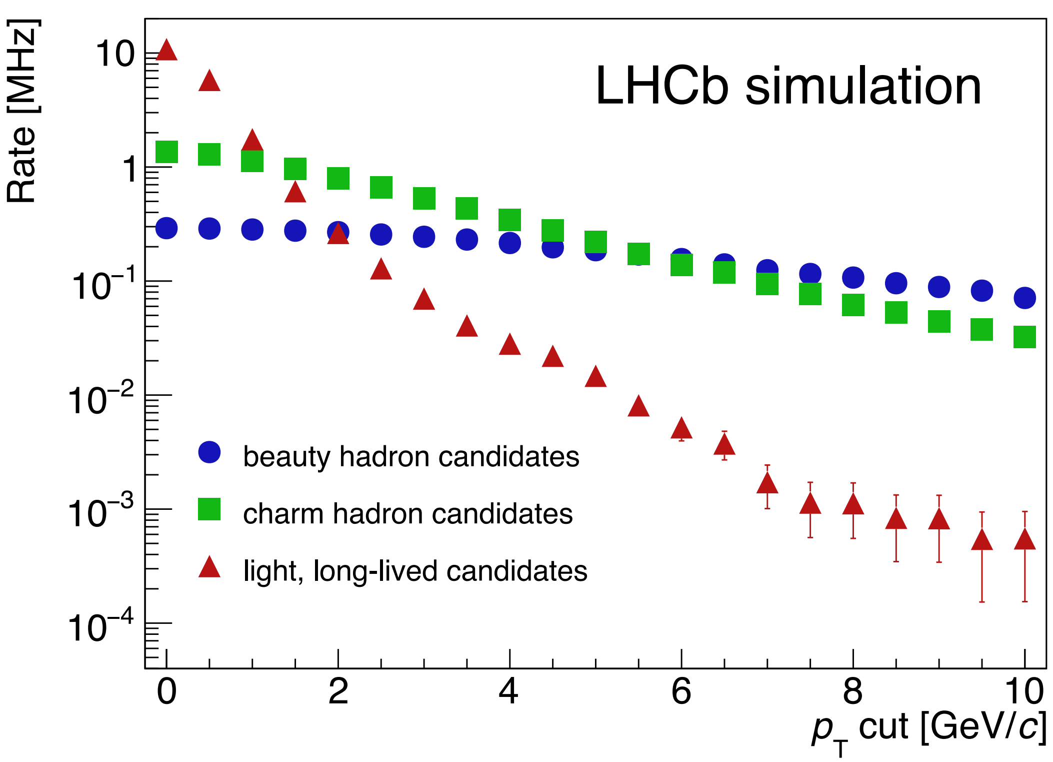
The LHCb Trigger System

In Run 3, LHCb is collecting events with 5x higher collision density.

Legacy hardware trigger rate places limitations for key channels.

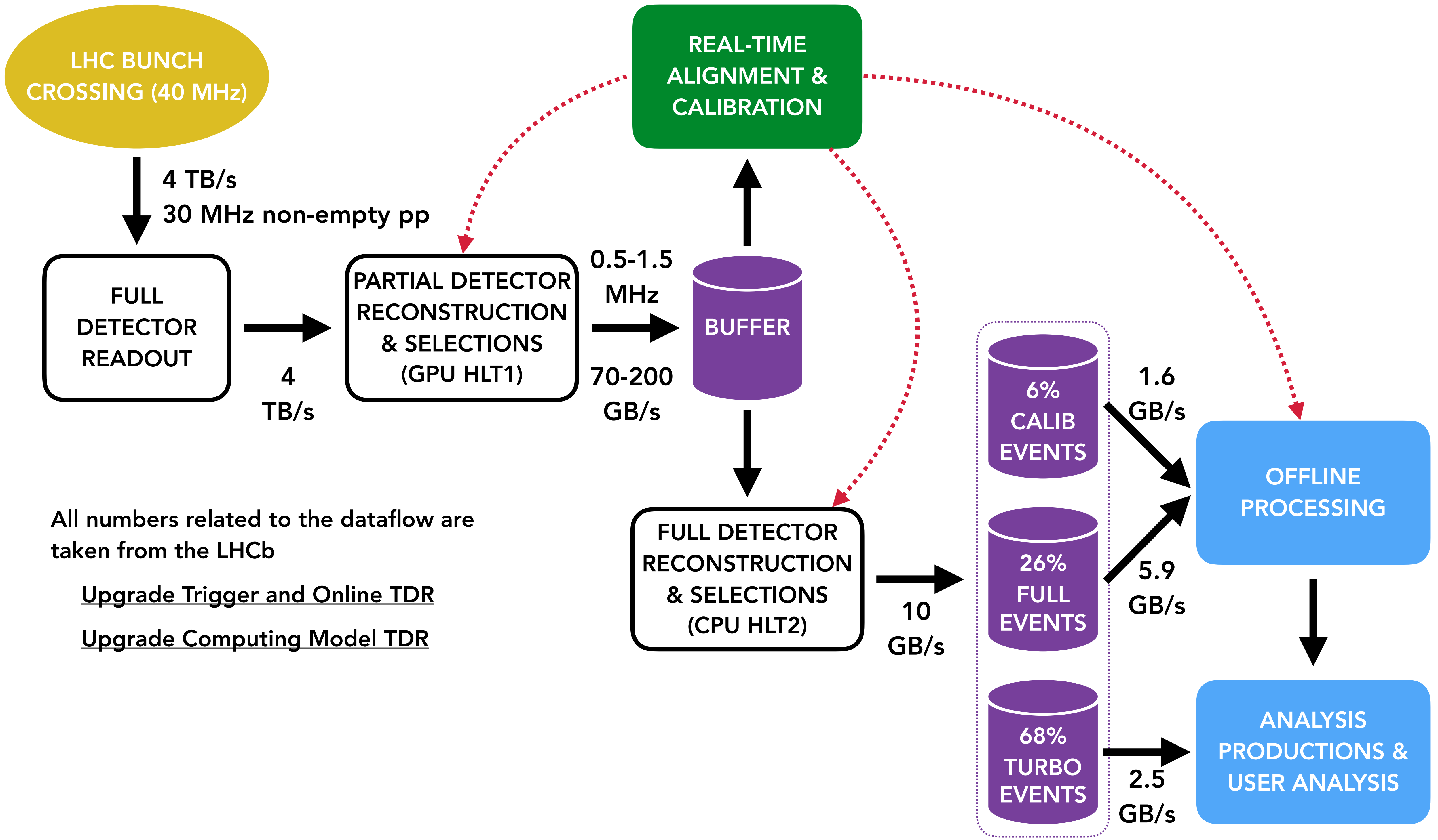
Requirements on p_T and displacement are not enough!





The LHCb Trigger System

For Run 3, transitioned to completely software-based trigger.



All numbers related to the dataflow are taken from the LHCb

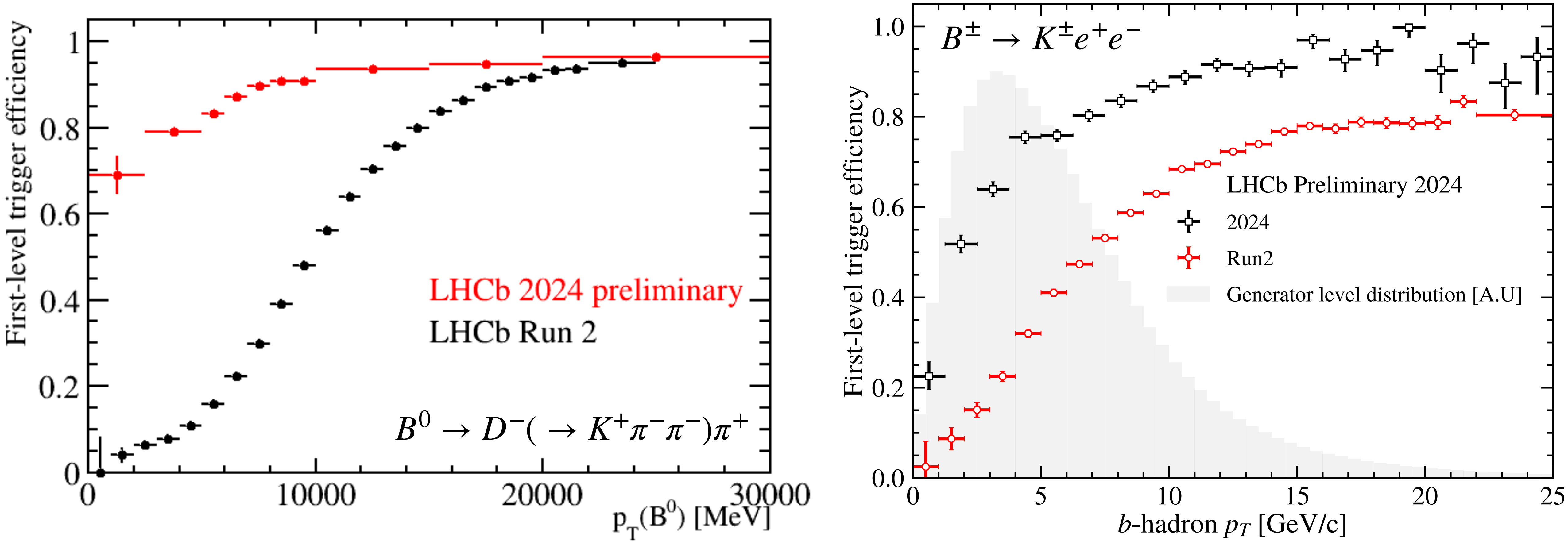
Upgrade Trigger and Online TDR

Upgrade Computing Model TDR

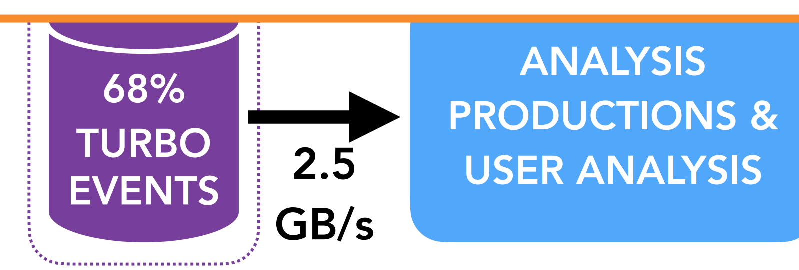
The LHCb Trigger System

For Run 3, transitioned to completely software-based trigger.

REAL-TIME

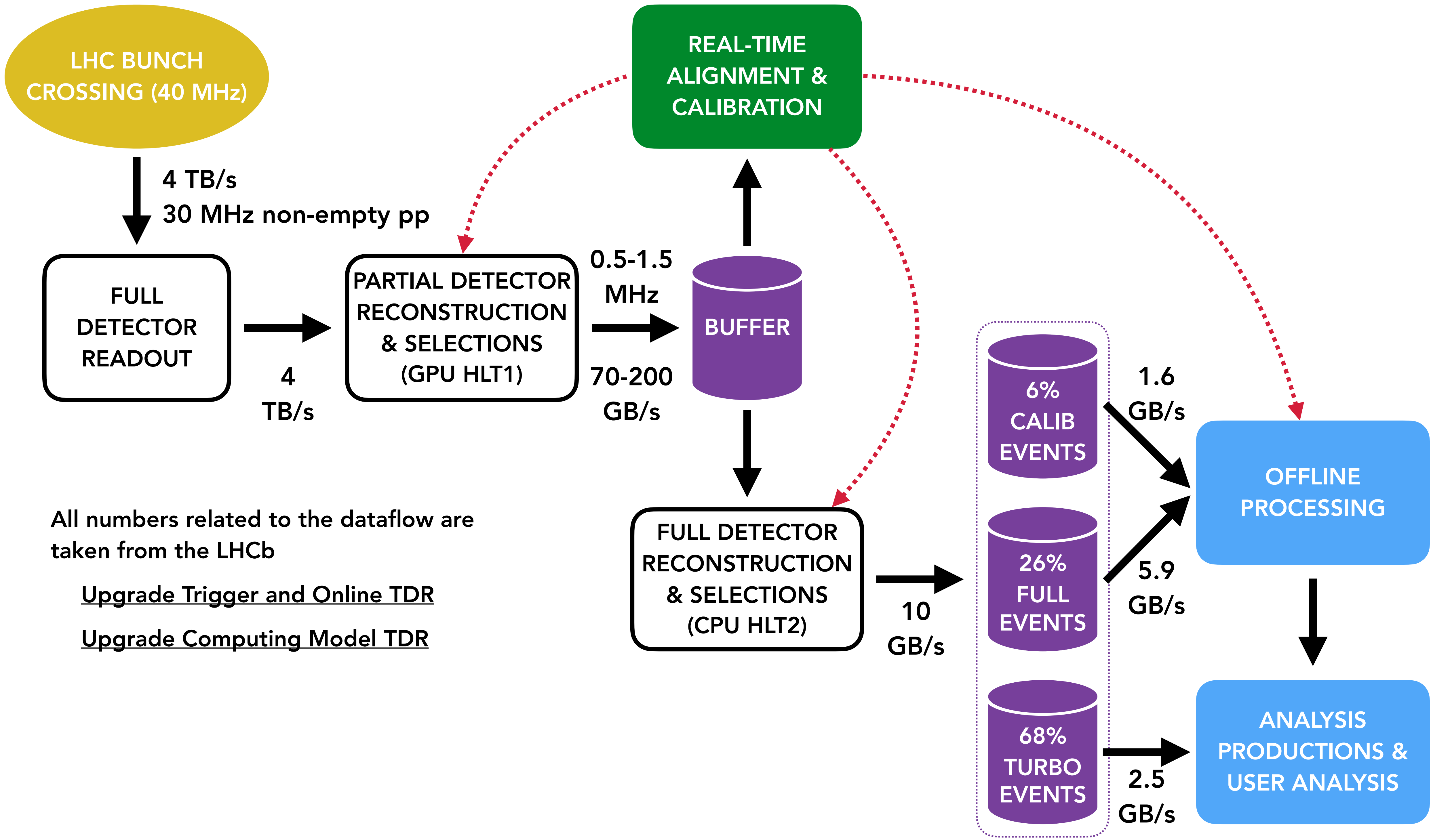


Largest improvements for charm physics and electron channels



The LHCb Trigger System

For Run 3, transitioned to completely software-based trigger.

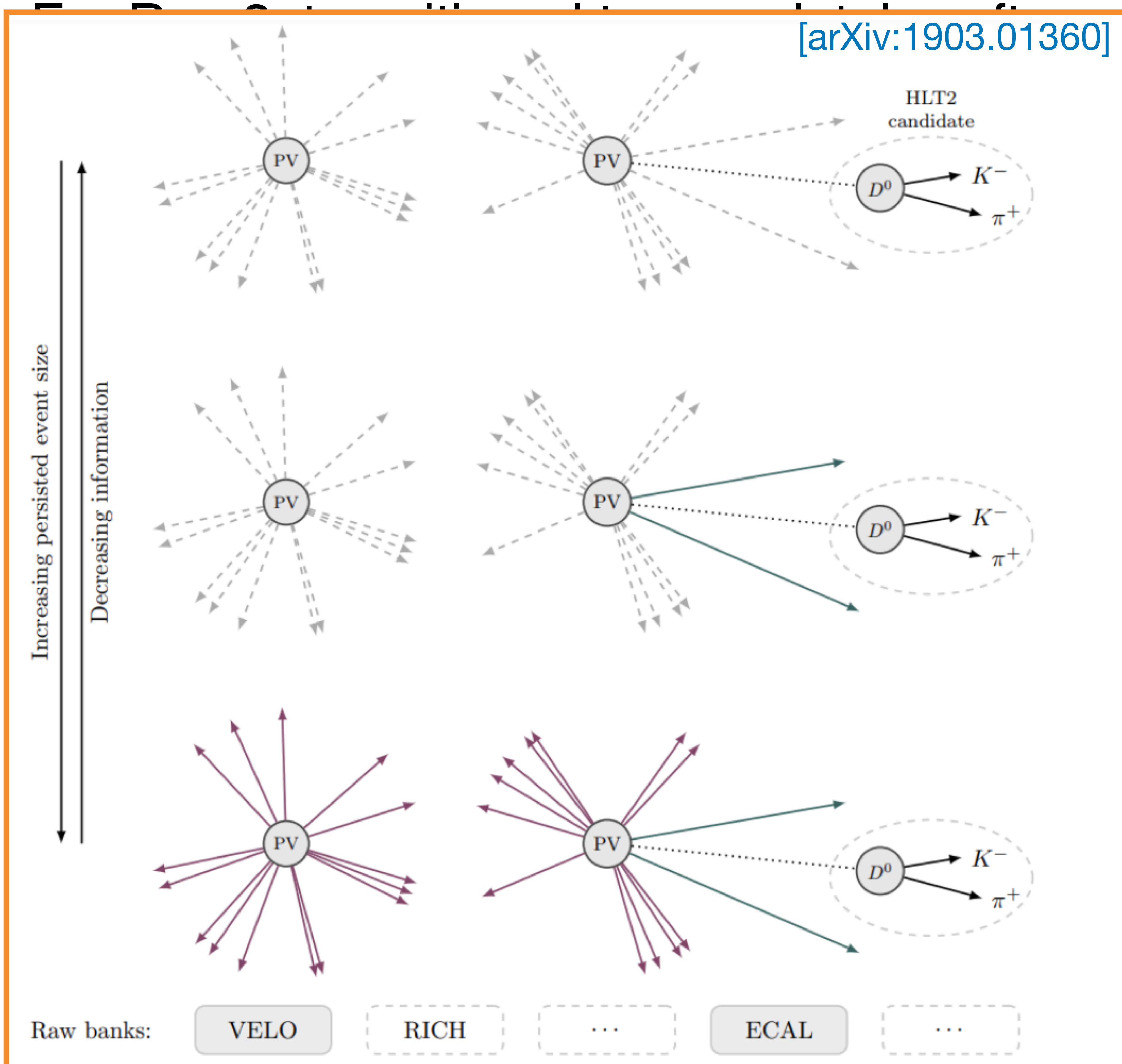


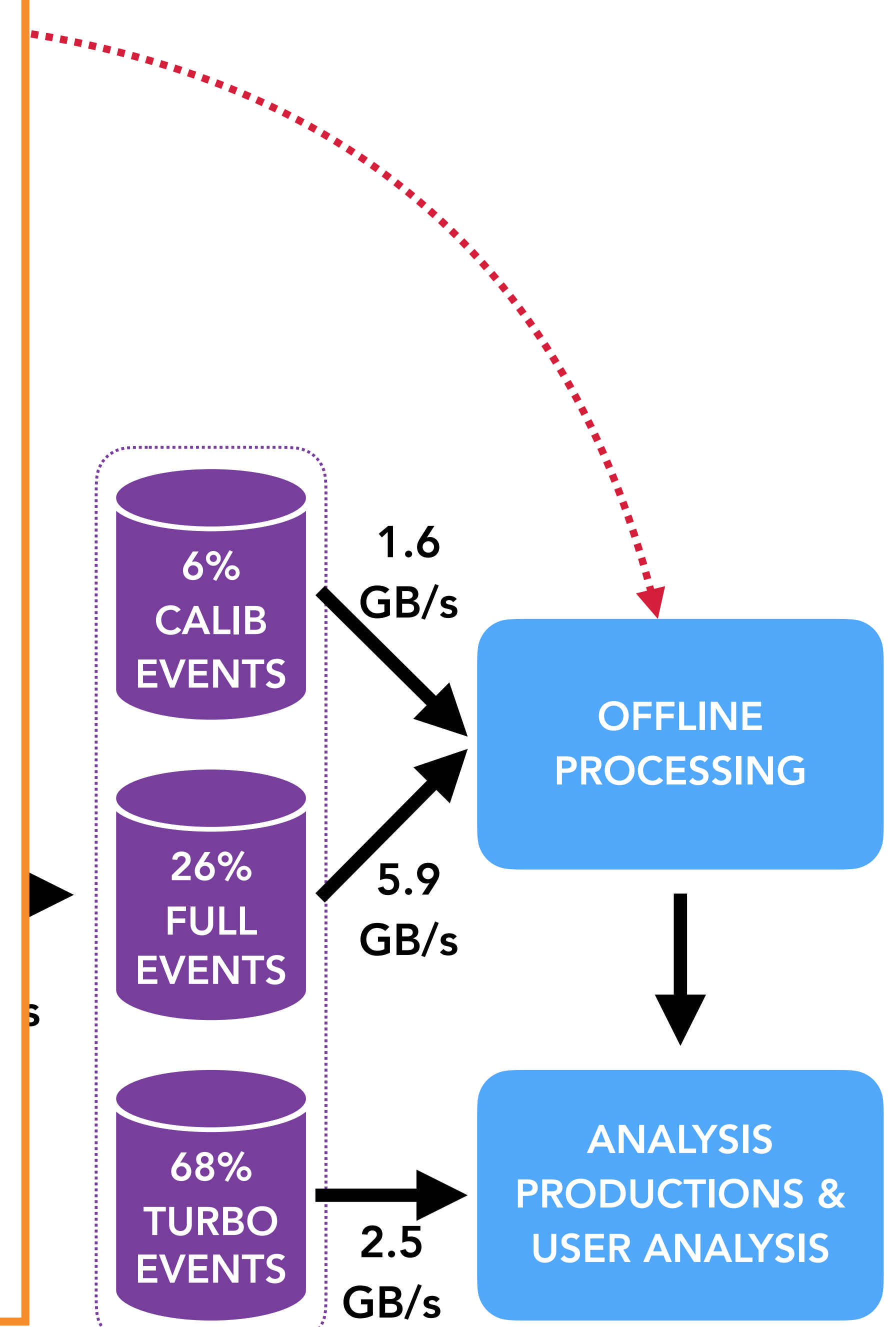
All numbers related to the dataflow are taken from the LHCb

- Upgrade Trigger and Online TDR
- Upgrade Computing Model TDR

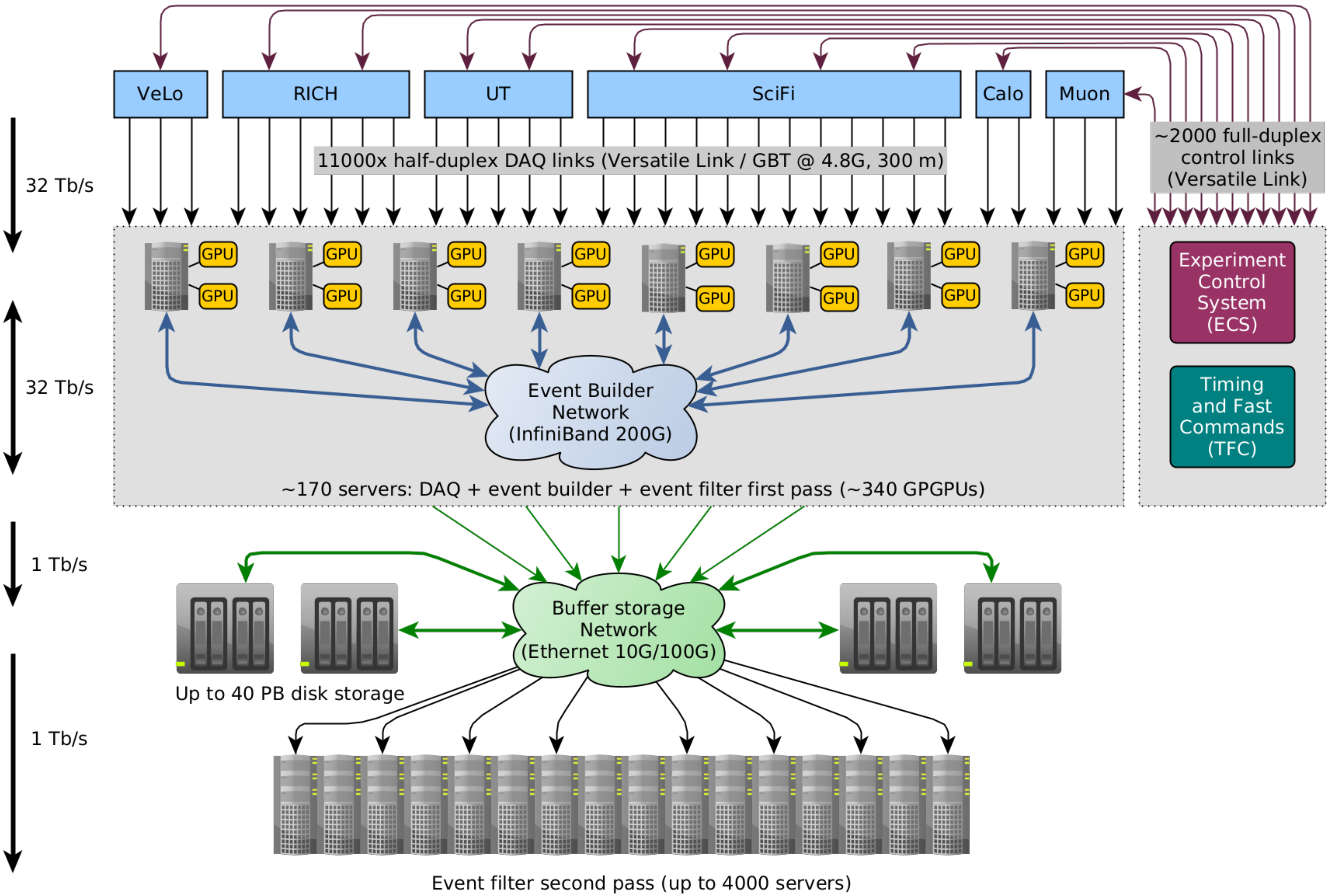
The LHCb Trigger System

e-based trigger.





The LHCb Trigger System



Wrapping up

- LHC trigger systems have continuously evolved to accommodate the changing conditions, detectors, and physics goals of the experiments
  - Rapidly improving technology plays a huge role
    - Faster links, larger chips, offering more compute with less power
- ATLAS, CMS, and LHCb have explored different strategies, sharing/borrowing ideas at times to great effect!
- High luminosity presents a new challenge for all experiments to face.
  - Detectors are being built “around the trigger”, to great effect!
- Thanks for your attention!