UKRI
Science and Technology Facilities Council

Introduction to Trigger Systems

Will Panduro Vazquez, RAL PPD

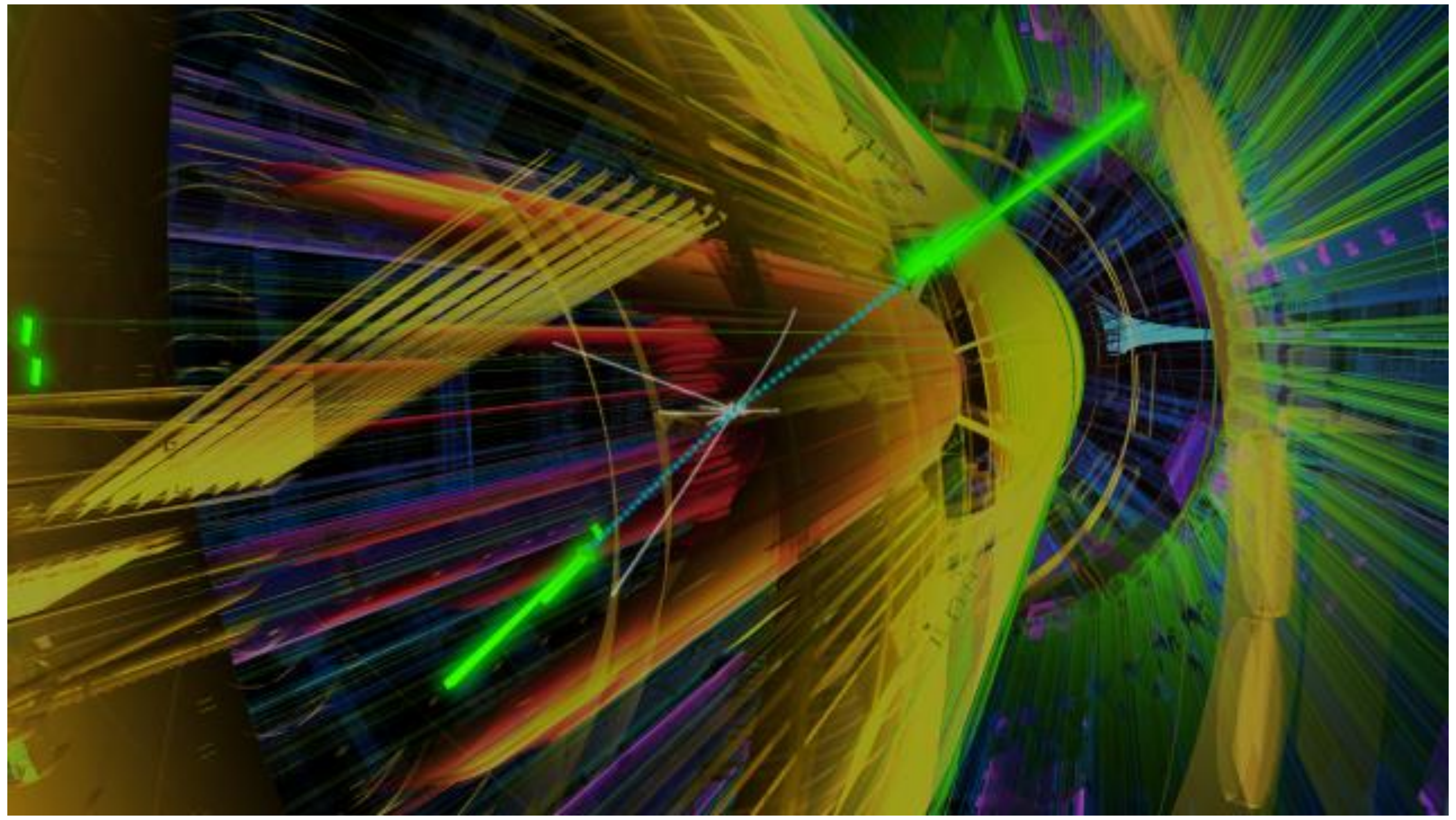
Advanced Graduate Lectures on practical Tools, Applications and Techniques in HEP
6th June 2024

Acknowledgements and Disclaimer

- Lecture and much of the material inherited from Julie Kirk (RAL PPD) – many thanks!
- I work on ATLAS – much of this talk will be collider-based, but the concepts are general

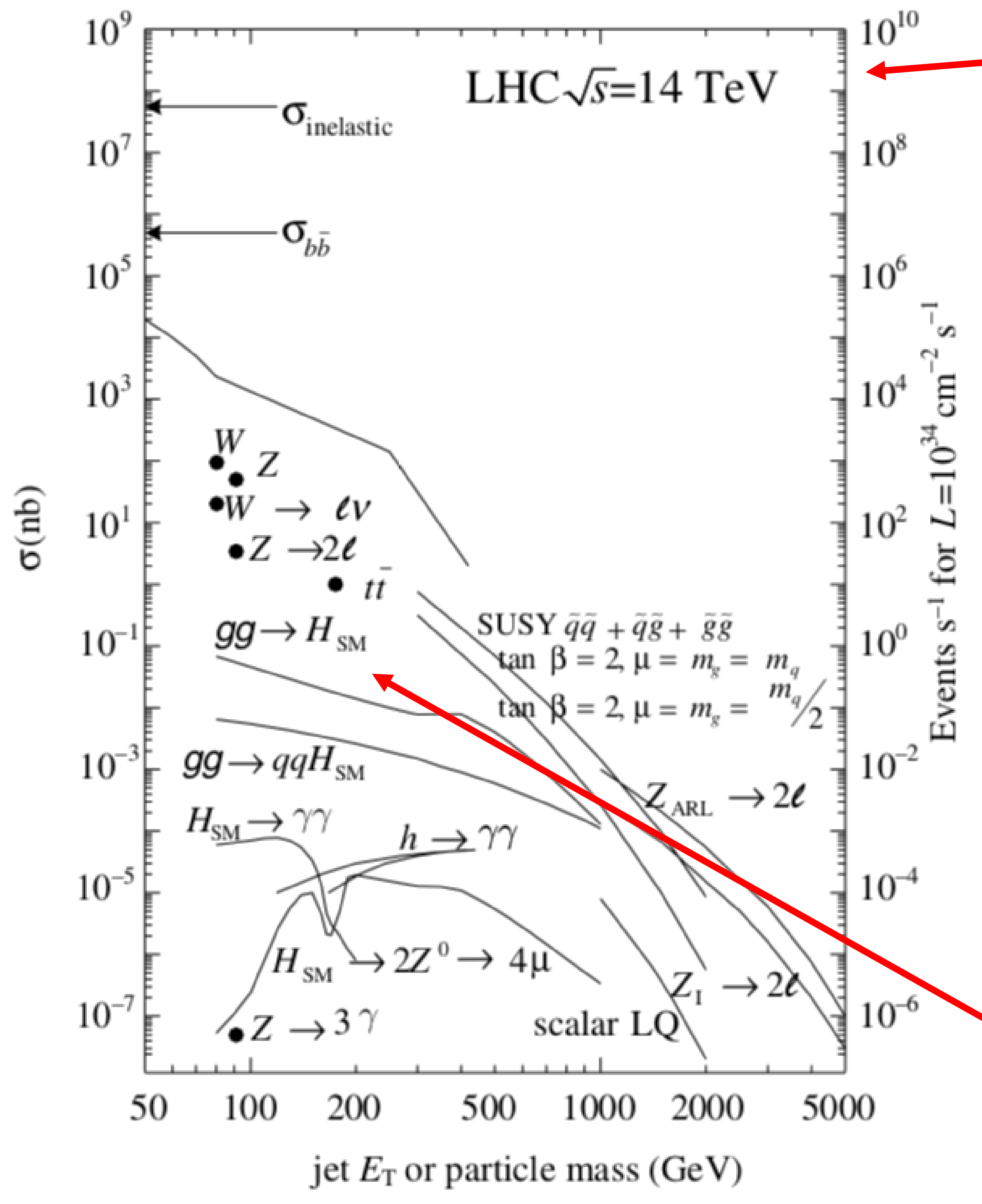
What is a trigger system, and why do we need it?

- The problem:
  - Particle physicists are typically searching for rare processes
  - Particle physics experiments, in particular at colliders, generate enormous amounts of data



What is a trigger system, and why do we need it?

- The problem:
  - Particle physicists are typically searching for incredibly rare processes
  - Particle physics experiments, in particular at colliders, generate enormous amounts of data



At the LHC: Inelastic cross-section → GHz of events

(40 MHz bunch crossing rate × ~60 p-p interactions per crossing)

Process	Cross-section (nb)	Production Rate (Hz)
Inelastic	10^8	10^9
$b - bbar$	$5x10^5$	$5x10^6$
$W \to l\nu$	15	150
$Z \to l\nu$	2	20
$t - tbar$	1	10
$H(125)SM$	0.05	0.5

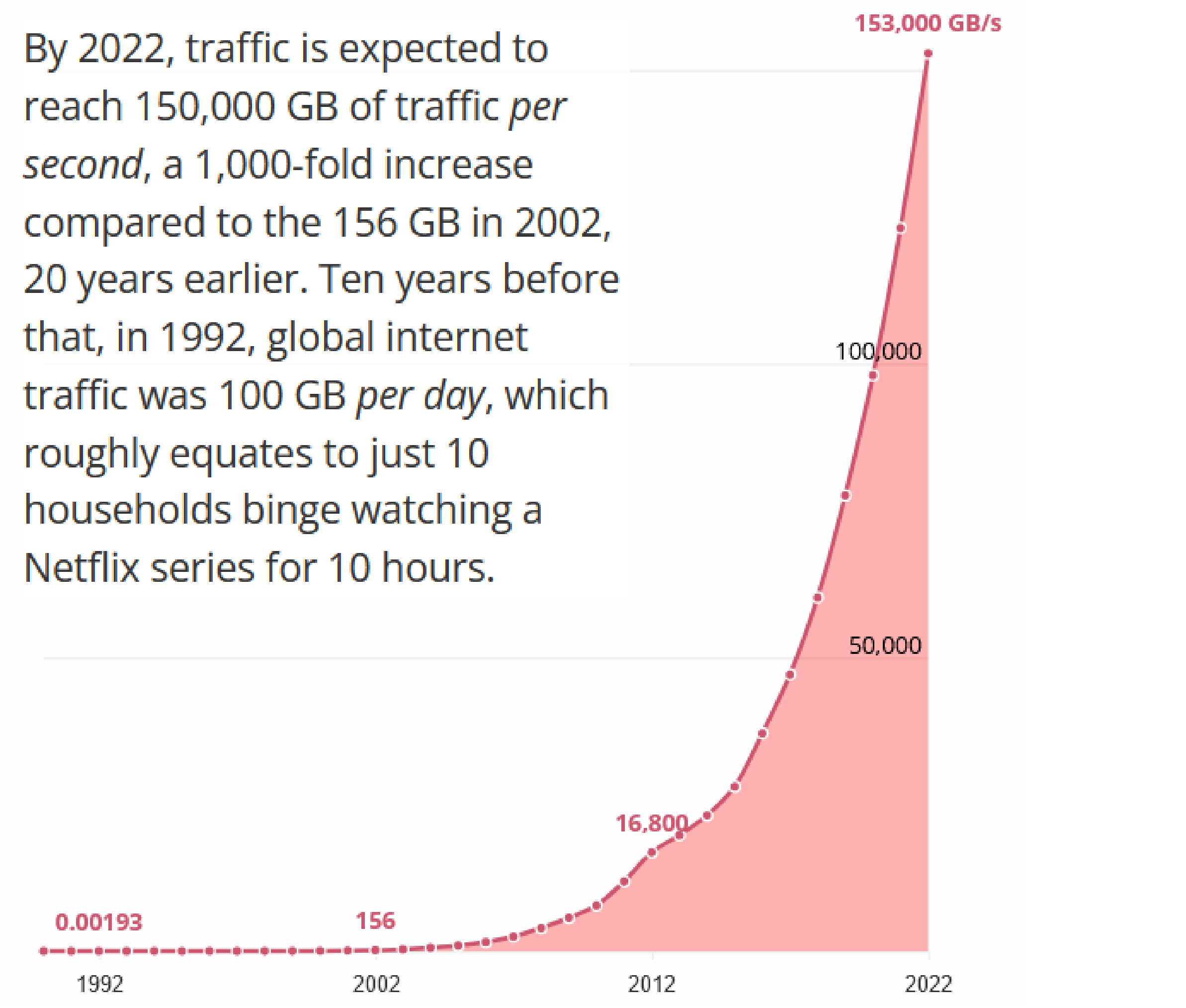
Some of the most interesting processes only happen very rarely : ~1 in 10^{11}

Science and Technology Facilities Council

Particle Physics

What is a trigger system, and why do we need it?

- The problem:
  - Particle physicists are typically searching for incredibly rare processes
  - Particle physics experiments, in particular at colliders, generate enormous amounts of data
- For example:
  - CMS and ATLAS unfiltered off-detector data rate of order 60 TB/s
  - Instantaneous global internet traffic averaged over 2022: 150 TB/s
- No storage system we could possibly afford could cope with this volume of data!



Source: WDR 2021 team calculations and Cisco Visual Networking Index: Forecast and Trends, 2017–2022.

Science and Technology Facilities Council
Particle Physics

What is a trigger system, and why do we need it?

- The problem:
  - Particle physicists are typically searching for rare processes
  - Particle physics experiments, in particular at colliders, generate enormous amounts of data
- The problem rephrased:
  - How do we maximise our ability to find these rare processes in an enormous and complex dataset, when we know that the vast majority of what we produce is (uninteresting) background?

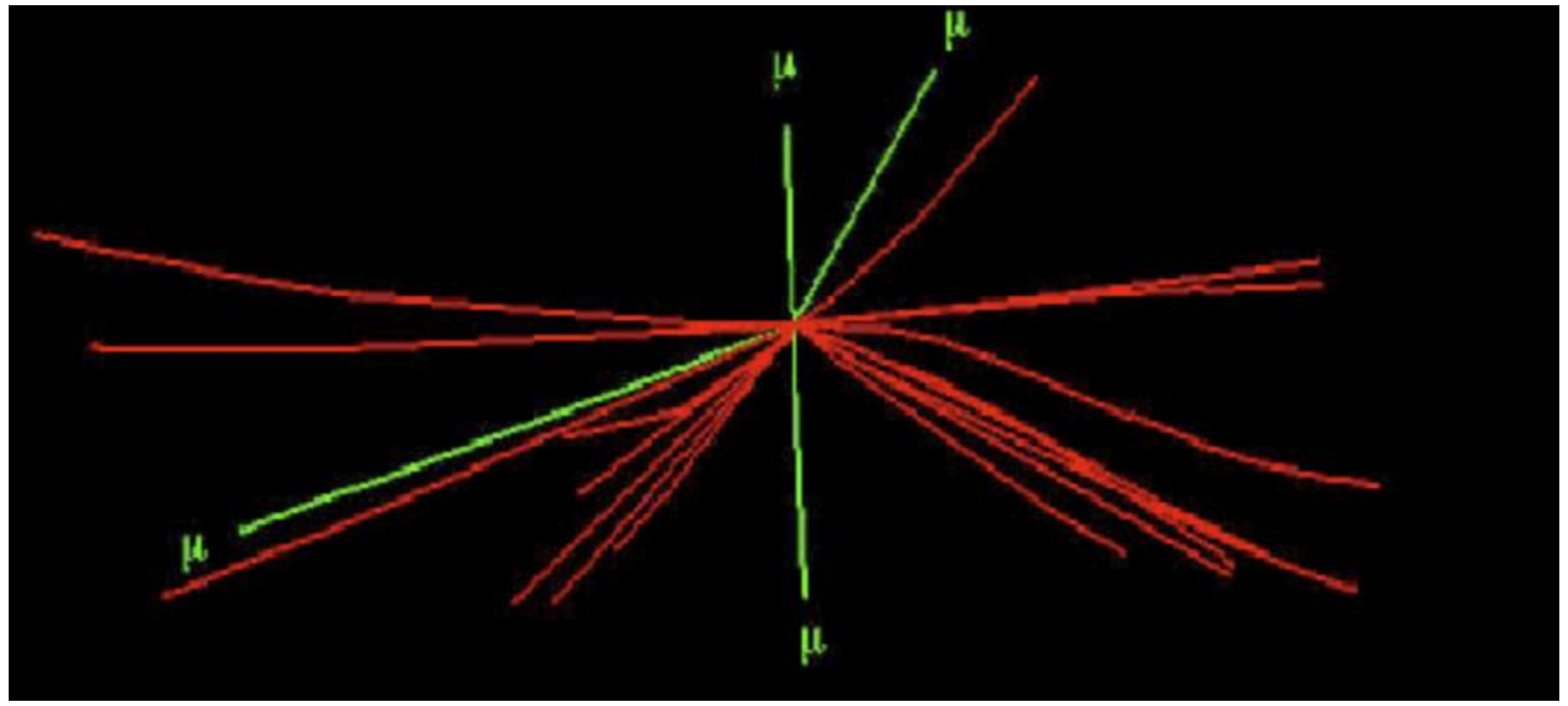
- One Higgs in every 10 billion pp interactions
- $H \rightarrow \gamma\gamma$ is even rarer, BR~10^{-3}
- 1 $H \rightarrow \gamma\gamma$ per 10 trillion interactions

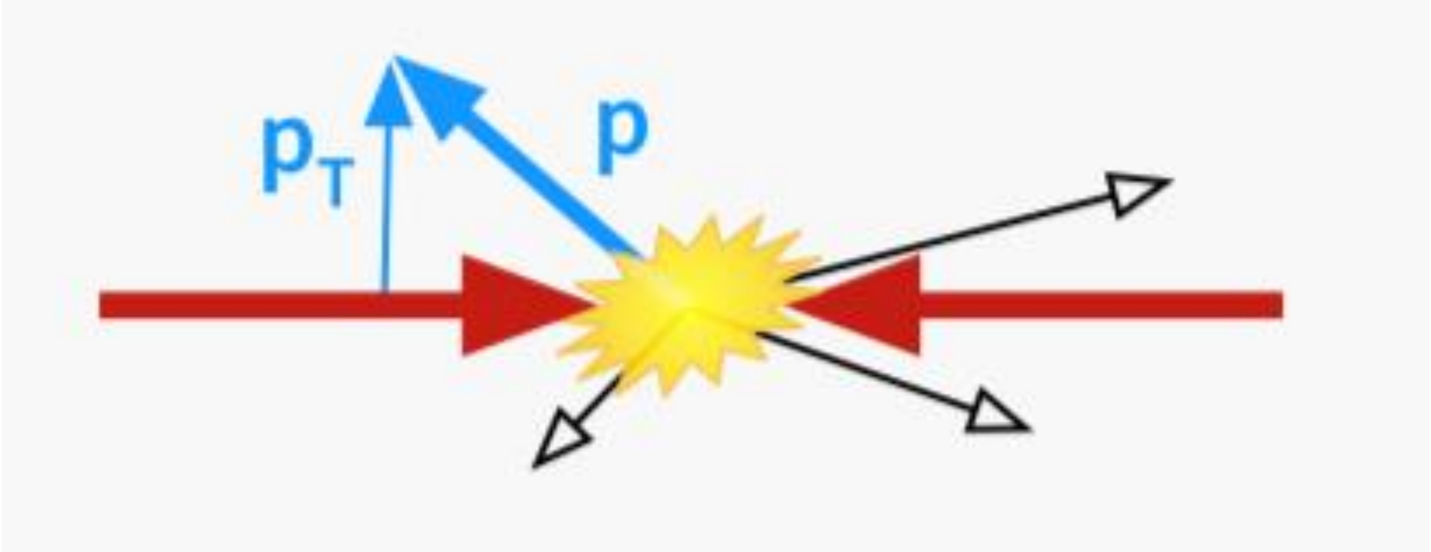
What is a trigger system, and why do we need it?

- The solution:
  - Triggers!
  - Broadly speaking:
    - *"A system that uses simple criteria to rapidly decide which events in a particle detector to keep when only a small fraction of the total can be recorded."*

What is a trigger system, and why do we need it?

- The solution:
  - Triggers!
  - Broadly speaking:
    - *"A system that uses simple criteria to rapidly decide which events in a particle detector to keep when only a small fraction of the total can be recorded."*
- Example from Colliders:
  - Interesting events usually have high-pT particles

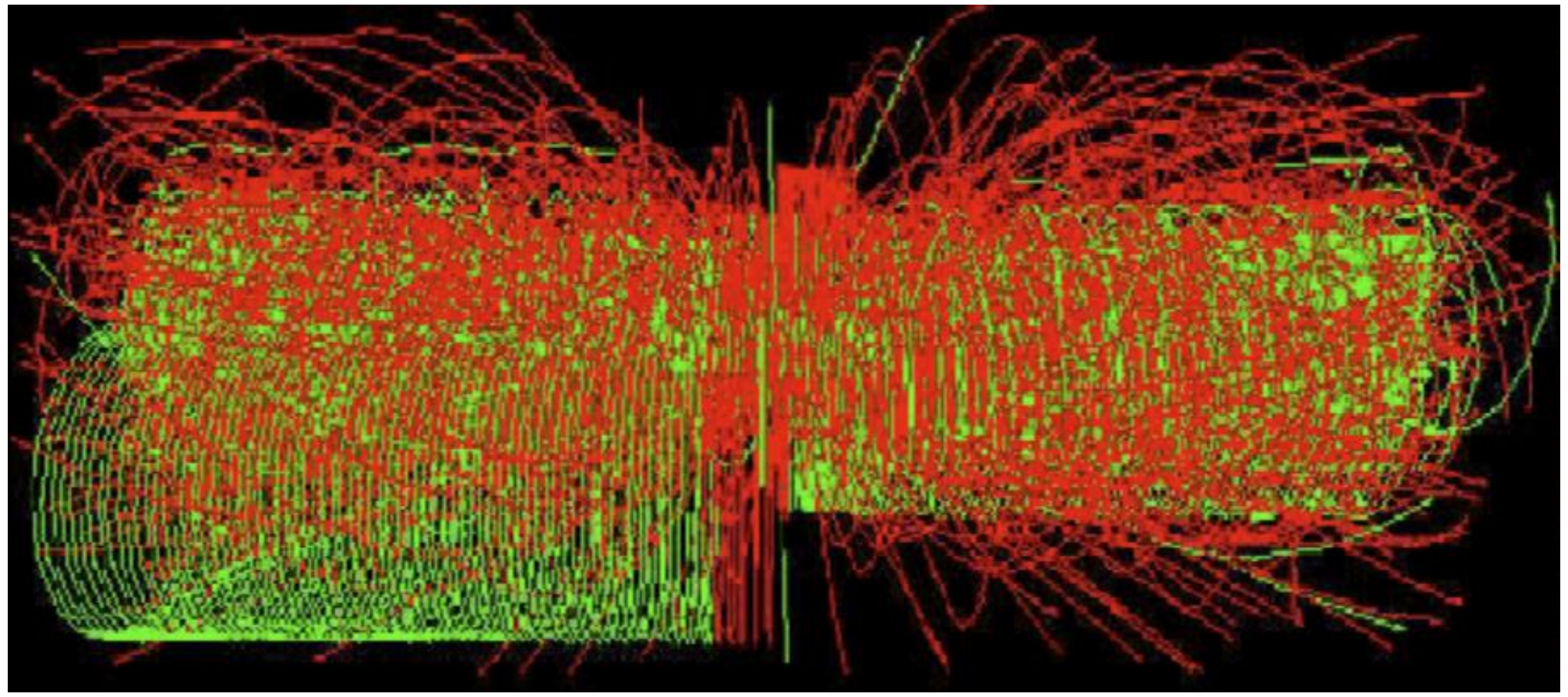




- $H \rightarrow 4\mu$, $p_T(\mu)$ ~ 30-50 GeV
- $H \rightarrow \gamma\gamma$, $p_T(\gamma)$ ~ 50-60 GeV

What is a trigger system, and why do we need it?

- The solution:
  - Triggers!
  - Broadly speaking:
    - *"A system that uses simple criteria to rapidly decide which events in a particle detector to keep when only a small fraction of the total can be recorded."*
- Example from Colliders:
  - Interesting events usually have high-pT particles
  - … hidden in a mass of low-pT background (~98% of the data!)

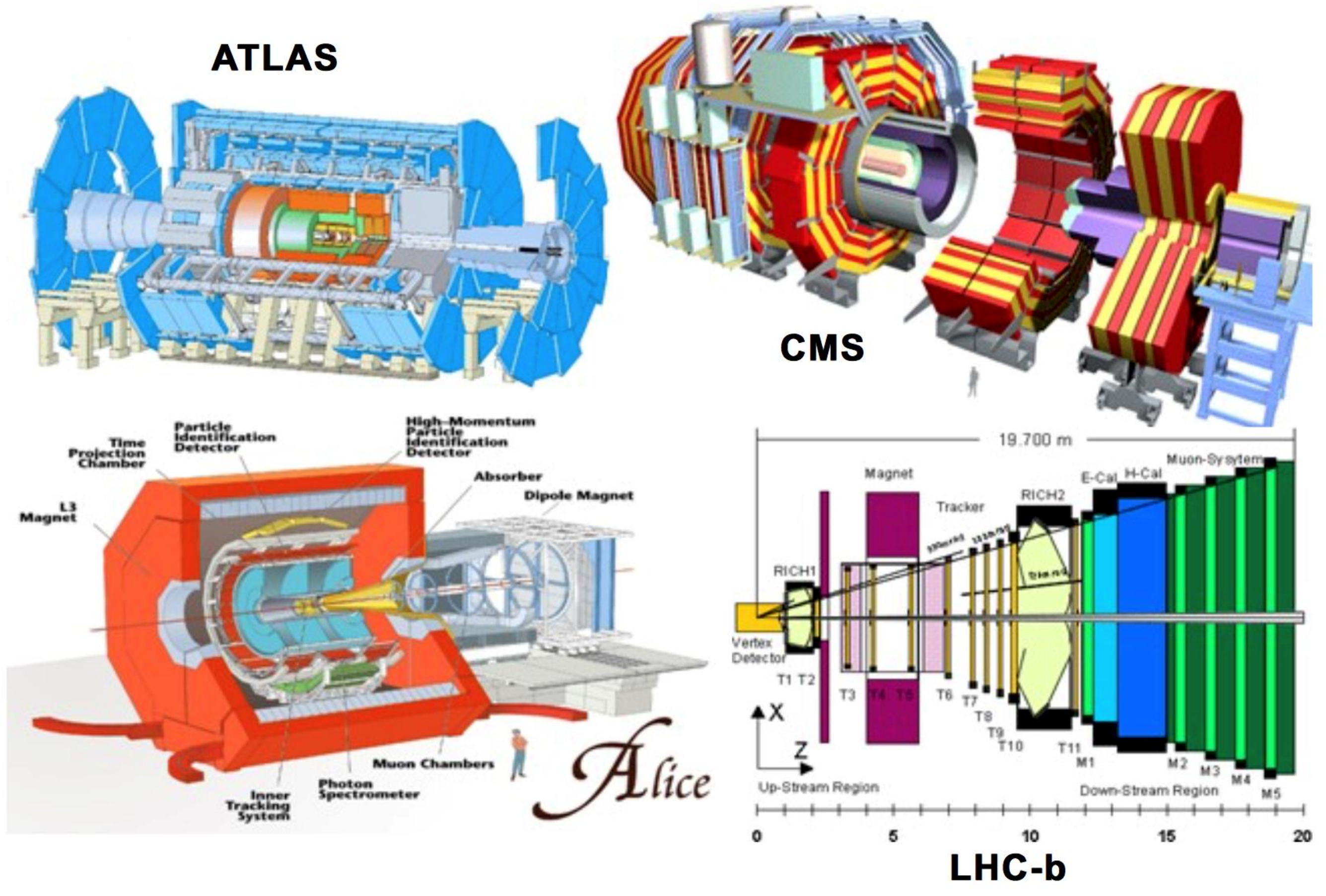


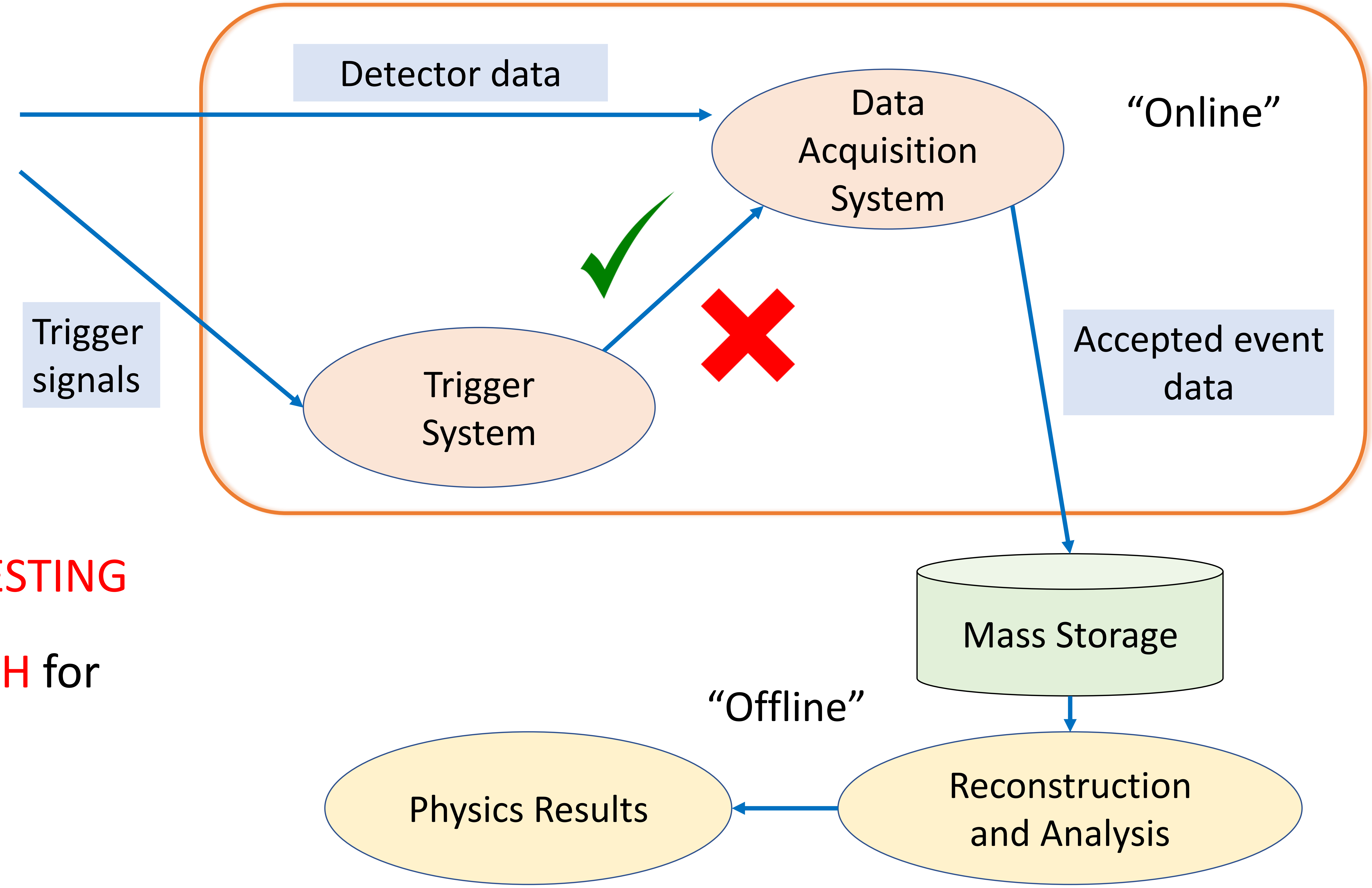
Science and
Technology
Facilities Council
Particle Physics

What is a trigger system, and why do we need it?

- The solution:
  - Triggers!
  - Broadly speaking:
    - *"A system that uses simple criteria to rapidly decide which events in a particle detector to keep when only a small fraction of the total can be recorded."*
- Example from Colliders:
  - Interesting events usually have high-pT particles
  - … hidden in a mass of low-pT background (~98% of the data!)
  - How do we design a system to preferentially detect the interesting high-pT events in real-time, storing only this small fraction for later analysis?
    - Have to get it right first time – once an event is discarded it is gone forever!

General Overview





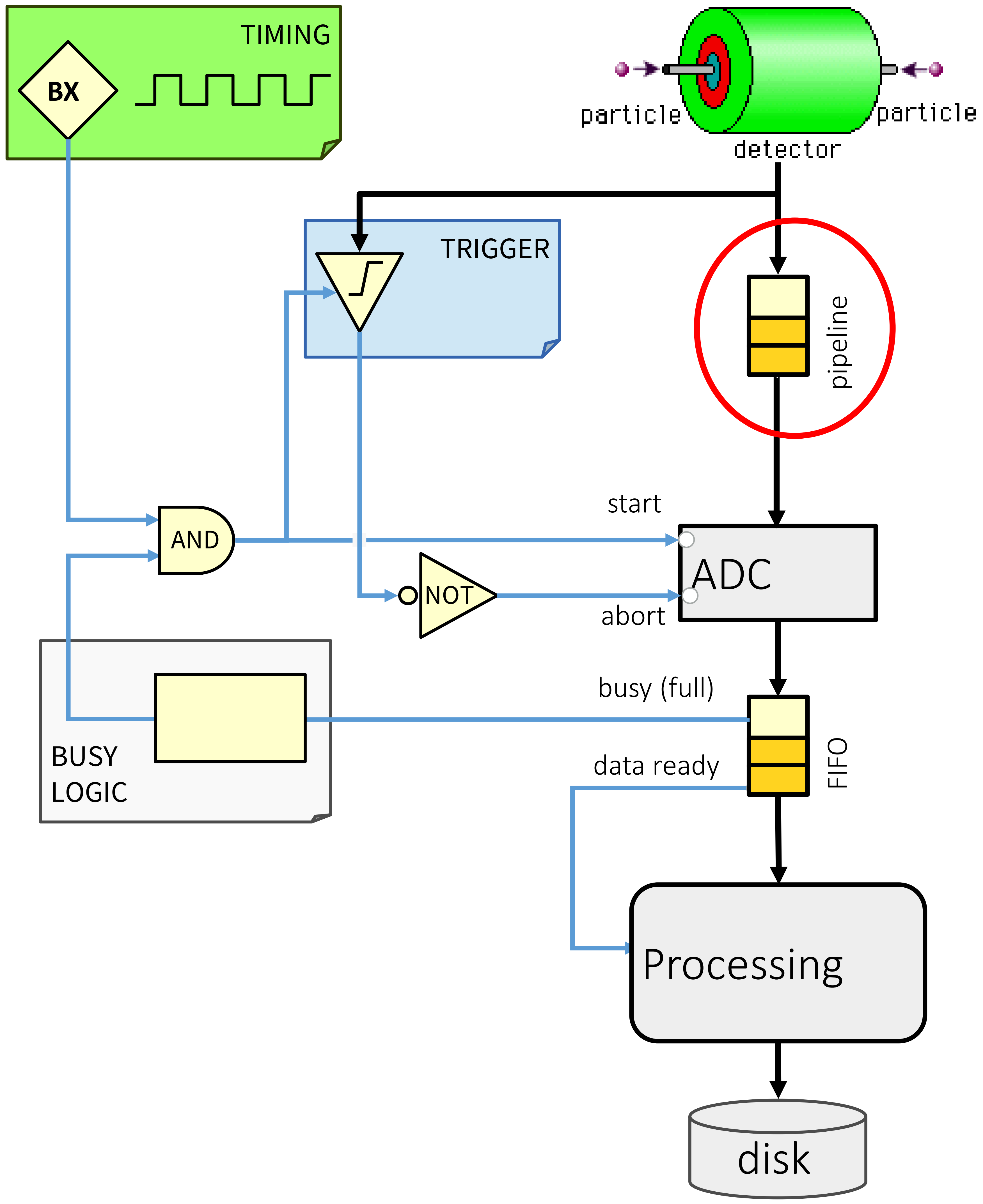
- Trigger must:
  - Make a FAST decision
  - Select the most INTERESTING events
  - Keep rate LOW ENOUGH for offline storage and reconstruction

Science and Technology Facilities Council

Particle Physics

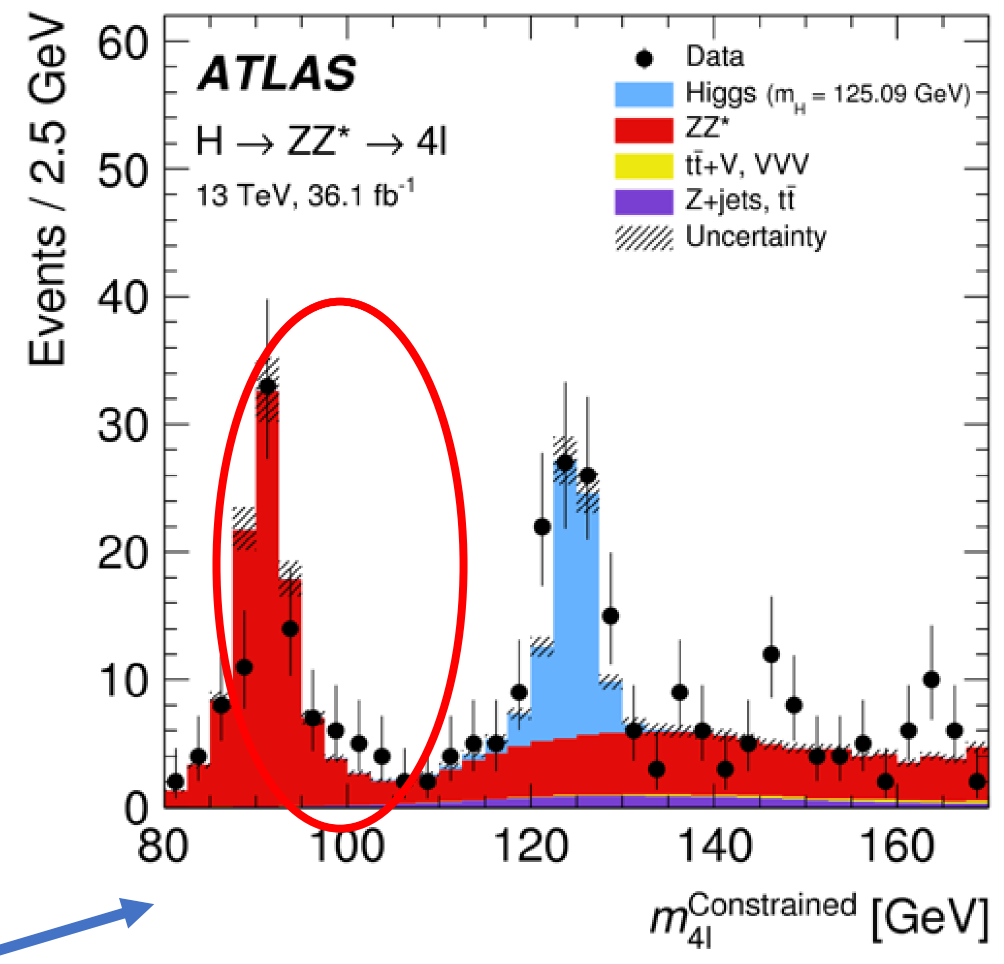
Making a Fast Decision

- At LHC, bunch crossings happen every 25 ns (40 MHz rate)
- Ideally need to decide for each bunch crossing whether to keep an event
- Huge amount of data per bunch crossing : $O(10^6\text{-}10^8)$ channels
- Time in which trigger has to operate, known as the latency, is limited by the amount of buffering available in the system

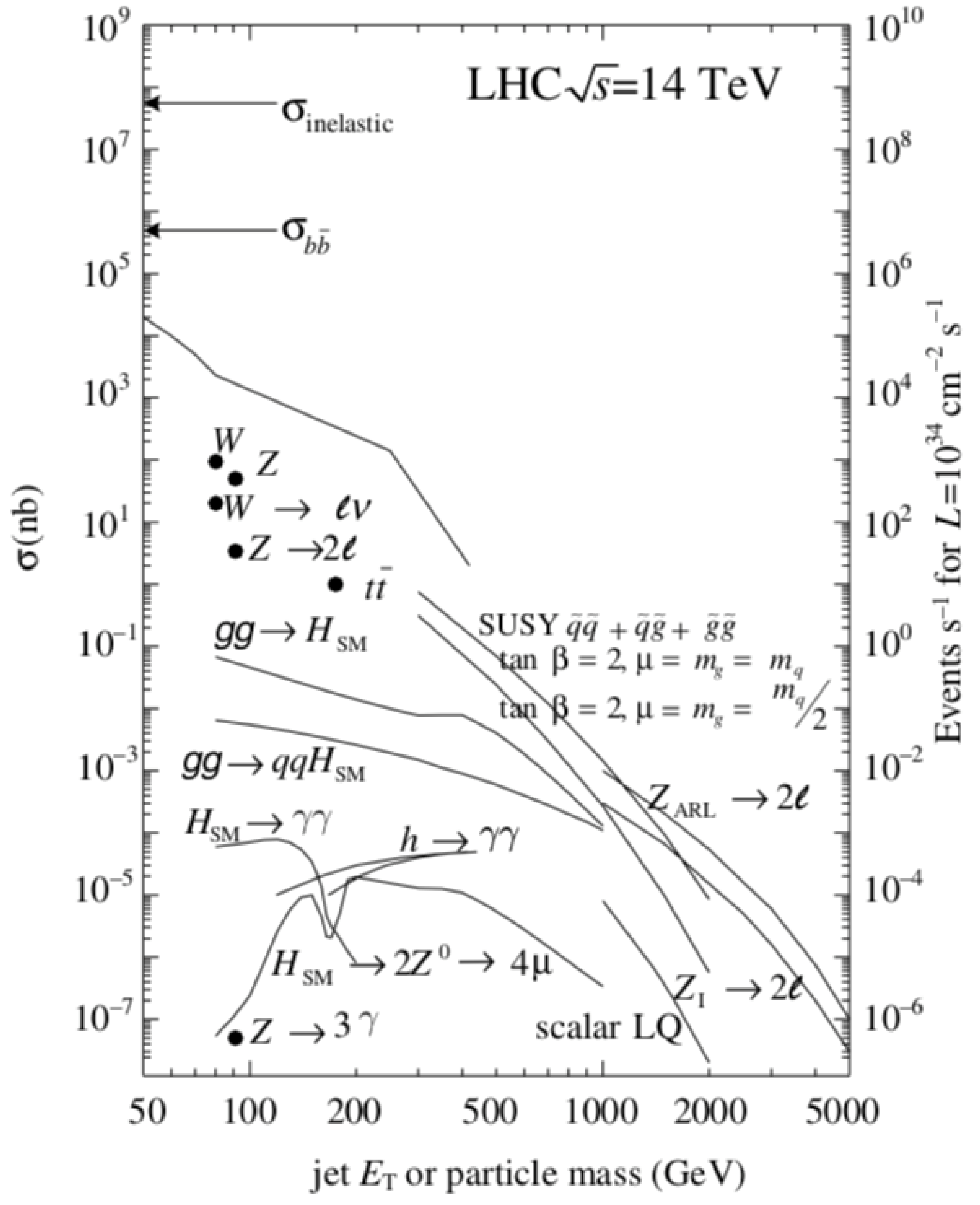


Select Interesting Events

- What is an interesting event?
  - Experiments aren't just discovery machines
  - We also need to study the Standard Model in detail to fully parameterise and understand it – in particular where it breaks down



- Physics goals for CMS/ATLAS:
  - Higgs properties
  - Search for Beyond Standard Model particles : SUSY, extra dimensions, new gauge bosons, black holes…..
  - Many interesting Standard Model studies, but these processes can occur with relatively high rate
    - e.g. $W \rightarrow l\nu$ ~150 Hz (10^{34} $cm^{-2}s^{-1}$)



Select Interesting Events

- Existing systems limit us to of the order of 1 kHz of data written to storage

- How do we prioritise sensitivity to new phenomena while also enabling us to study known landscape in more detail?
  - Typically, experiments define a Trigger "menu" (more later), which decides which events to keep of each type
  - Decision taken with input from across the experiment

Keep rate low enough

- As you heard in the previous lecture, the DAQ system collects the data from different parts of the detector, converts them to a suitable format and (ultimately) saves them to permanent storage

- DAQ bandwidth constrained by:
  - Finite storage capacity
    - Both short term (e.g. RAM) and long term (disc/tape)
  - Network connectivity (and cost)
    - Our ability to transport data from one system to another in large volumes
  - Cost of computing power
    - Both in the DAQ and trigger chains and for further processing of stored data

- Average event size determines the trigger rate, itself dictated by the number of particles in the event and the amount of data we can transport off-detector

$$BW_{DAQ} = Rate_{Trig}^{max} \, x \, Size^{Ev}$$

Keep rate low enough

- As you heard in the previous lecture, the DAQ system collects the data from different parts of the detector, converts them to a suitable format and (ultimately) saves them to permanent storage

- DAQ bandwidth constrained by:
  - Finite storage capacity
    - Both short term (e.g. RAM) and long term (disc/tape)
  - Network connectivity (and cost)
    - Our ability to transport data from one system to another in large volumes
  - Cost of computing power
    - Both in the DAQ and trigger chains and for further processing of stored data

- Average event size determines the trigger rate, itself dictated by the number of particles in the event and the amount of data we can transport off-detector

$$BW_{DAQ} = Rate_{Trig}^{max} \; x \; Size^{Ev}$$

~ 1 GB/s

~ 1 kHz

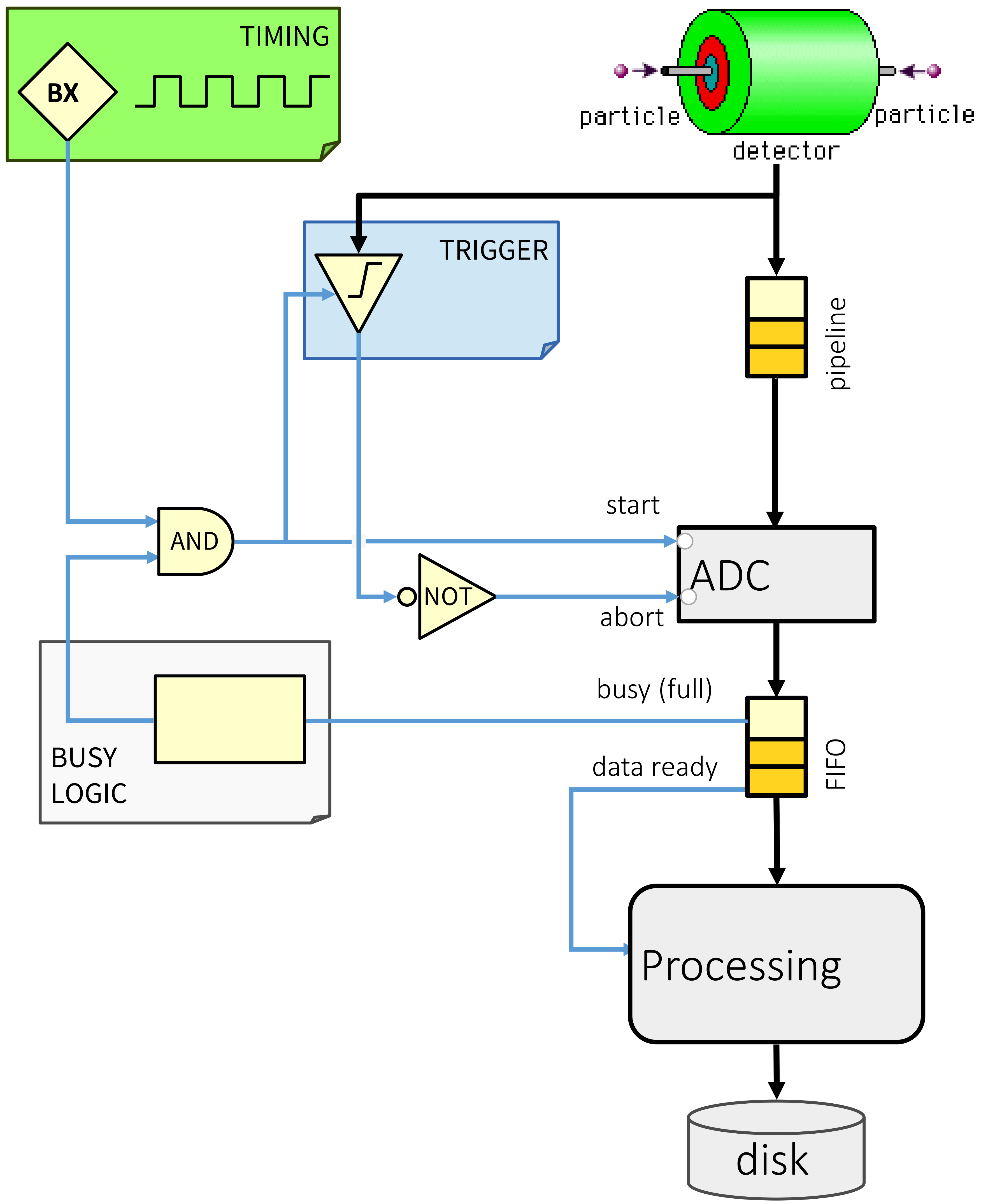
~ 1 MB

Science and Technology Facilities Council

Particle Physics

Multi-Level Trigger Systems

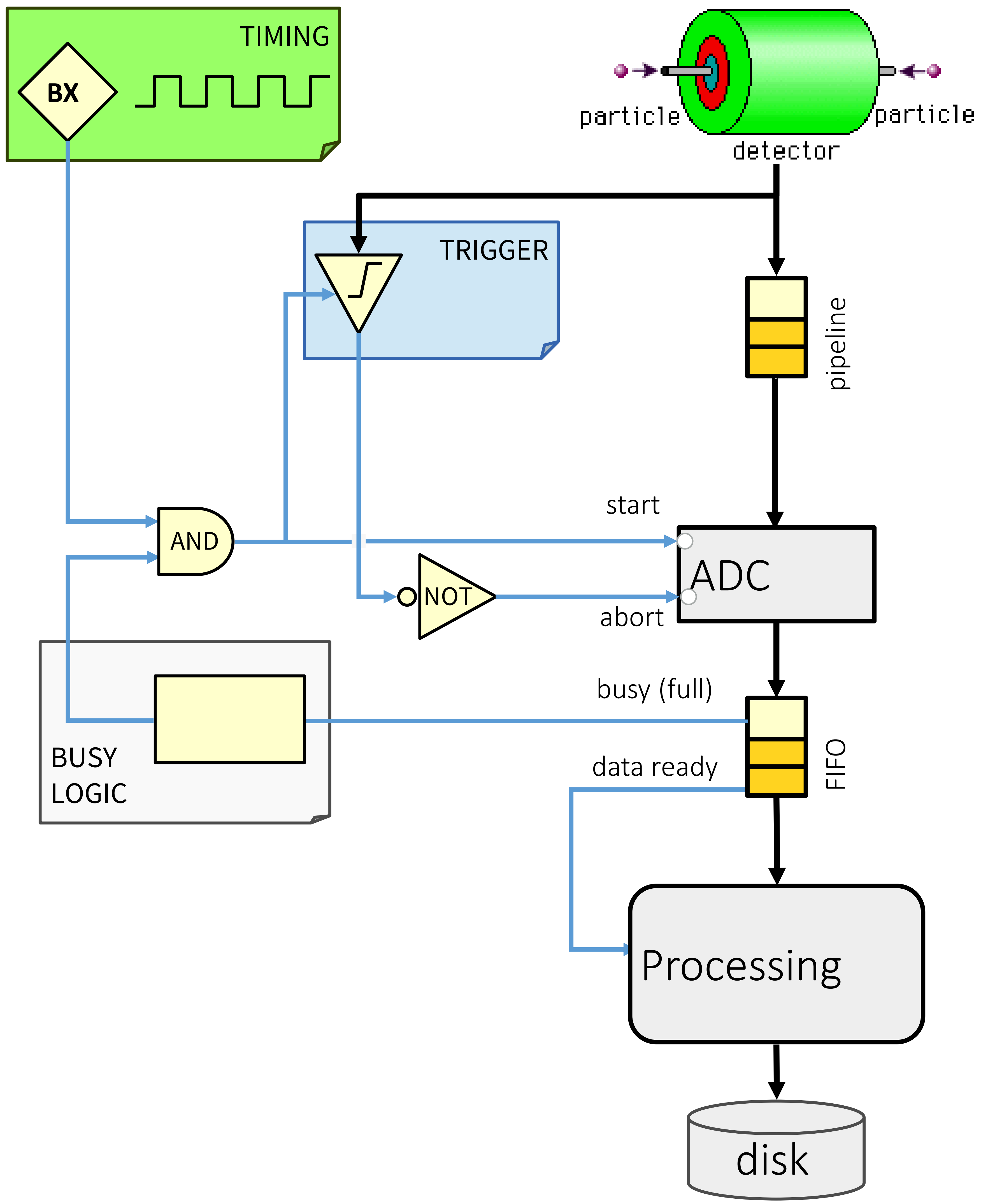
- In previous lecture, we built up a complex trigger system starting from a simple example (β decay)
- Basic idea is that of a discriminator, which generates an output signal when input is greater than some threshold
- In most collider experiments, must also integrate with timing and BUSY propagation mechanisms, as well as wider DAQ system



Science and Technology Facilities Council
Particle Physics

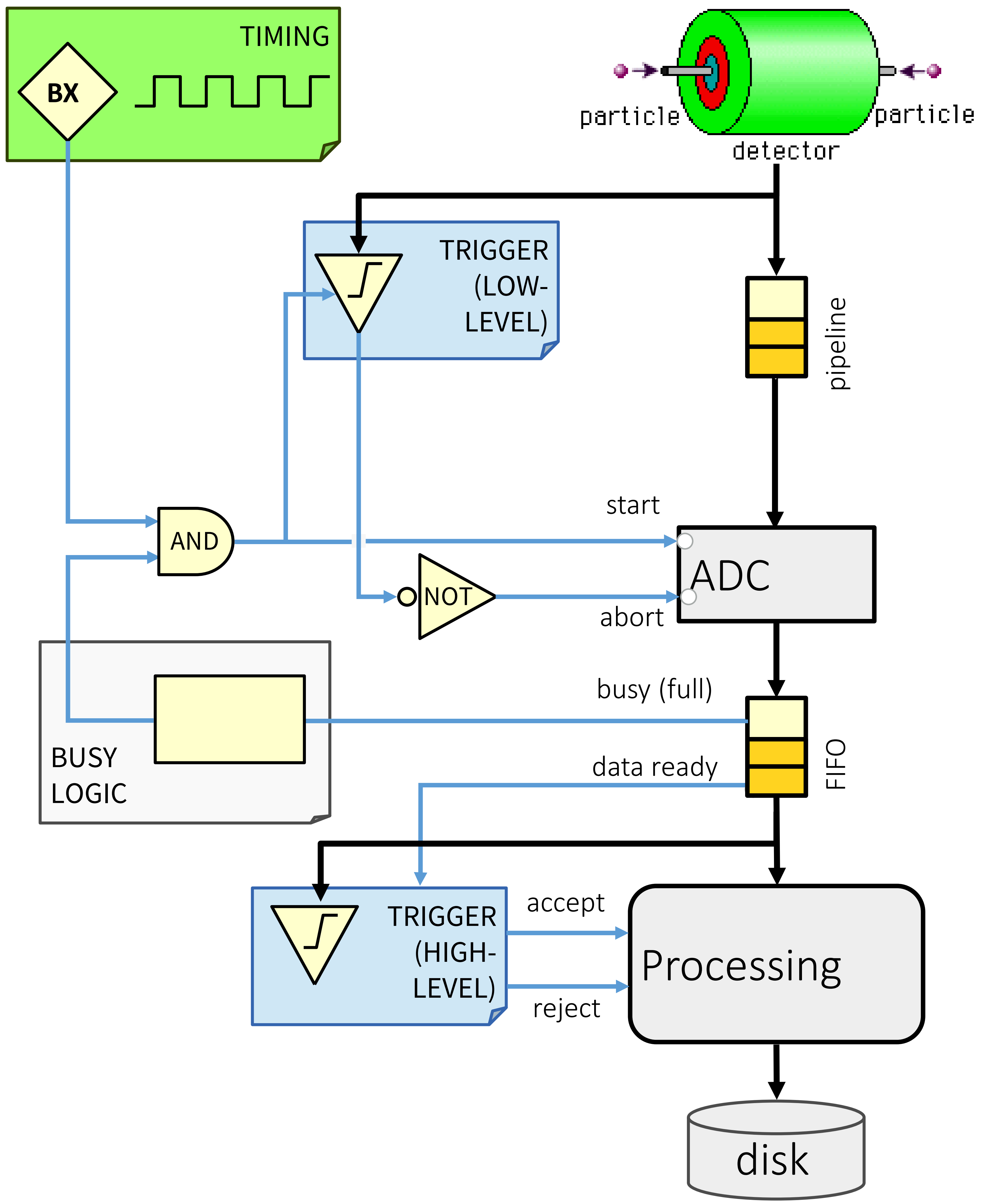
Multi-Level Trigger Systems

- Most (but not all) collider detectors implement their triggers with high-speed, typically custom, electronics in the first instance
- What happens if this hardware layer is unable to reduce the rate sufficiently?
- Typically constrained by available technology and cost



Multi-Level Trigger Systems

- Solution: add additional trigger levels to further reduce rate
- Each trigger level copes with smaller input rates, and can potentially benefit from larger buffers
- Results in increased latency, meaning more complex analysis can take place
- Experiments often opt for one or two fast hardware trigger layers followed by a slower (but still fast!) software trigger layer

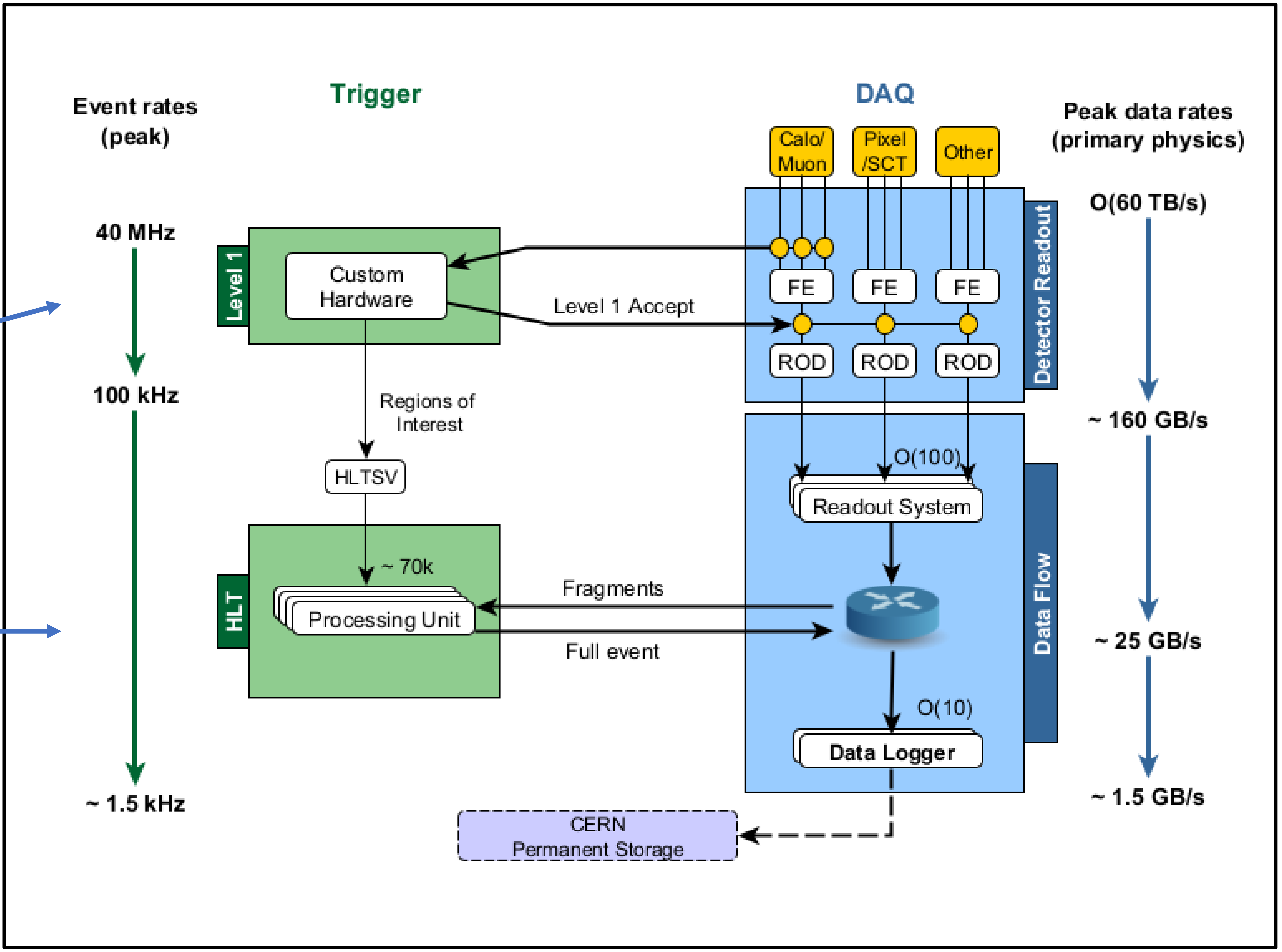


Science and Technology Facilities Council

Particle Physics

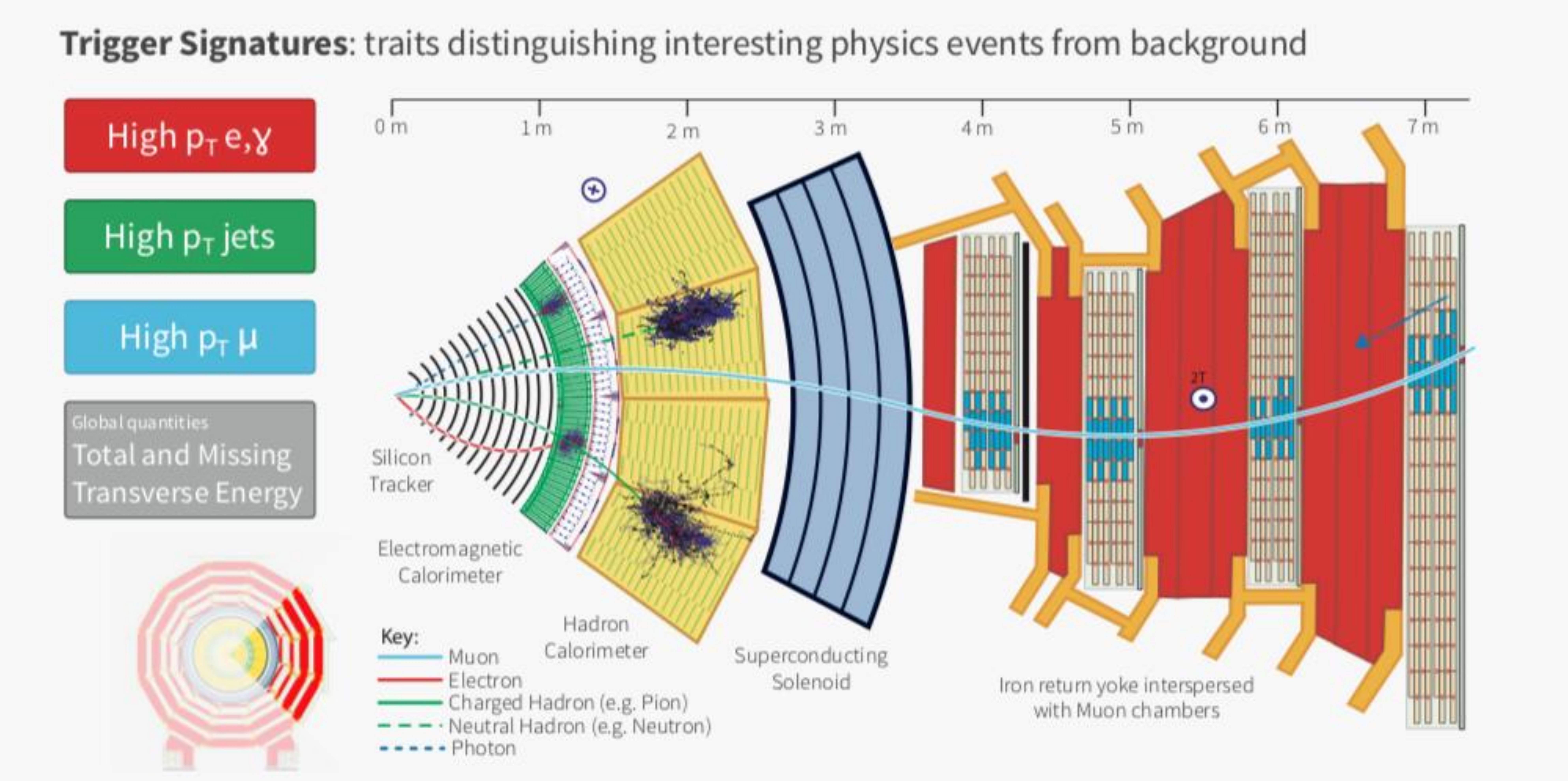
Multi-Level Trigger Systems

- Example: ATLAS TDAQ System, shown here in LHC Run 2
- Trigger decision distributed over two steps, each rejecting the vast majority of events
- L1 Trigger based on custom hardware (ASICs, FPGAs), with 2.5 μs latency
    - Keep 1 in 400 events
- HLT based on software running on commodity servers, with approaching 1 s latency
    - Keep approx. 1 in 66 events



Science and Technology Facilities Council
Particle Physics

What is the trigger looking for?



Science and Technology Facilities Council

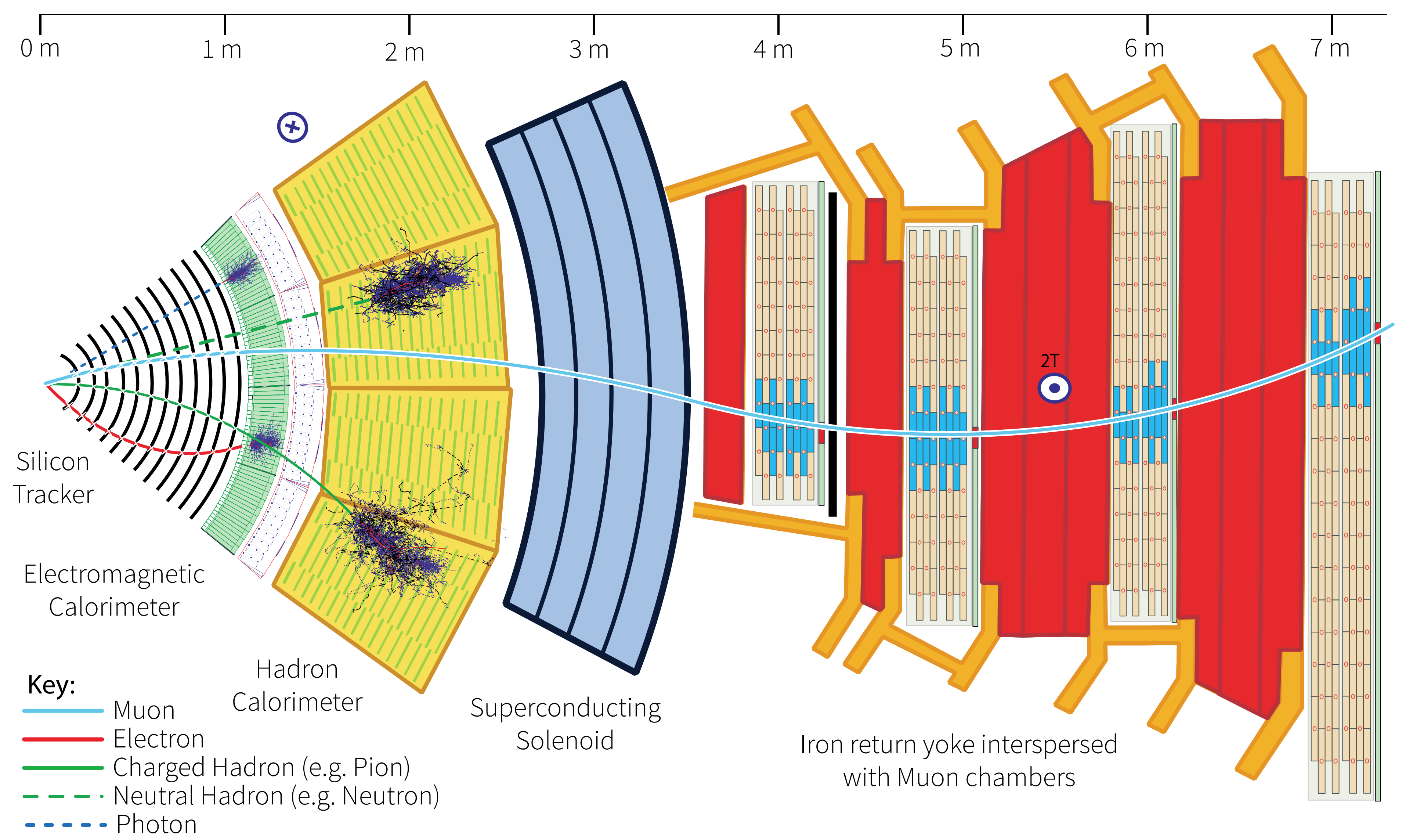
Particle Physics

Trigger Algorithms

- Trigger selection based on multiple trigger algorithms

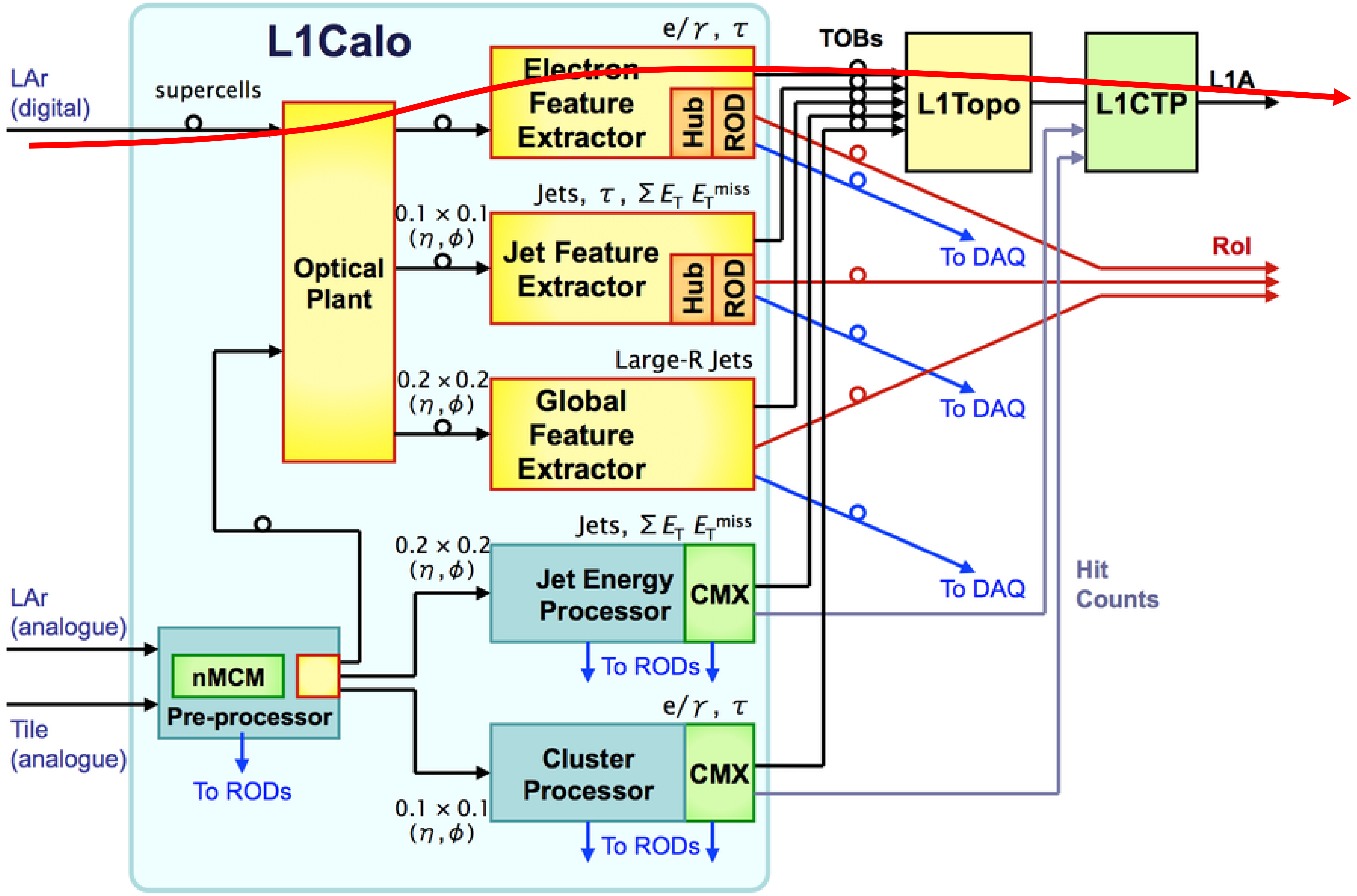
- Exploiting reduced data from detector regions to identify physics signatures
- Generally, several algorithms operate in parallel to find different signatures (trigger objects)
- e.g. calorimeter data used to find electrons + jets in parallel
- Algorithms must cover whole detector in an unbiased way
- The algorithm output is a counter or list of signatures, possibly with extra information
  - # of objects
  - pT, position, charge, 'quality', etc… for each object

- Some algorithms are 'global'
  - i.e. they encompass the whole detector
  - Examples: Missing ET, Total ET, HT, global object counts

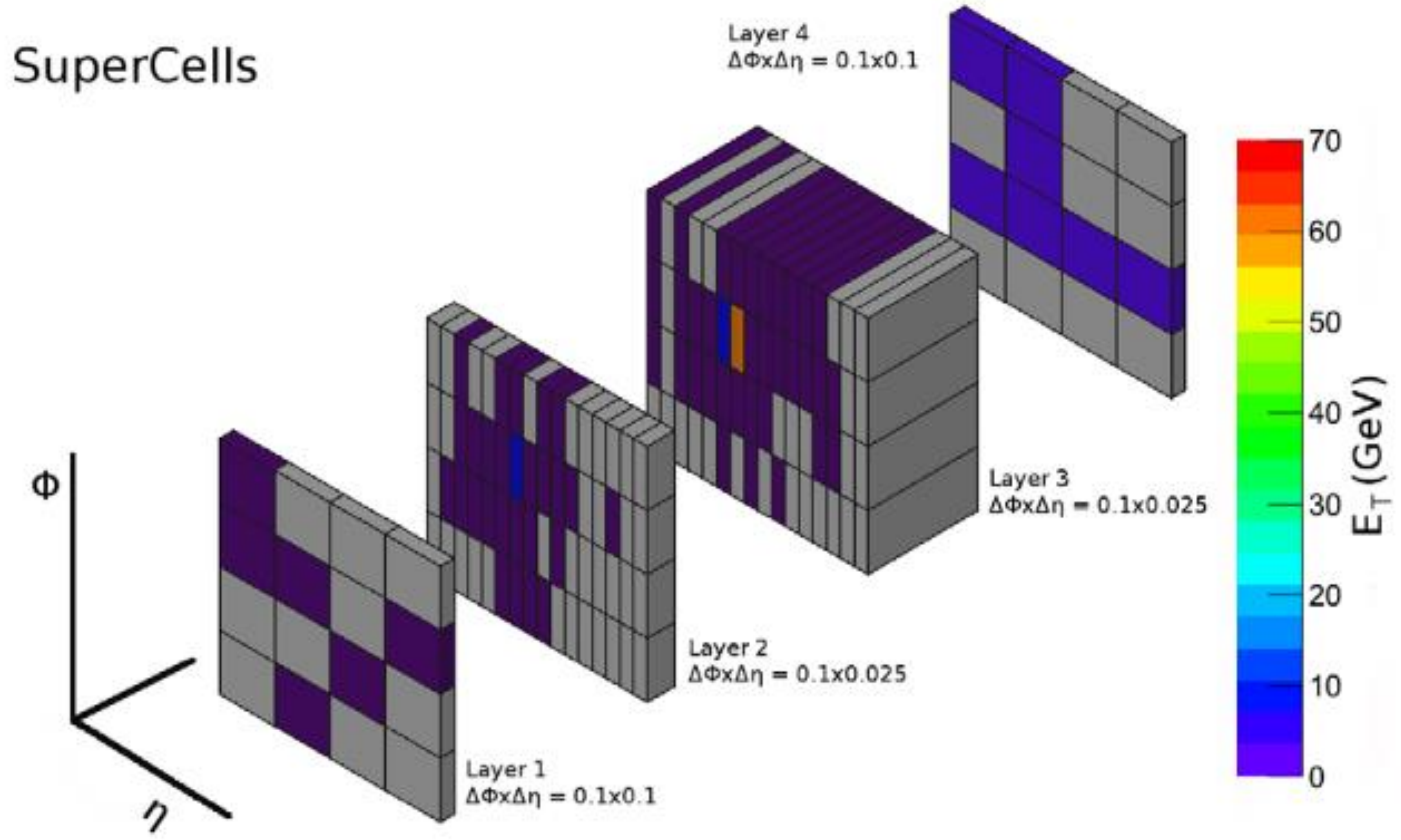


Science and Technology Facilities Council
Particle Physics

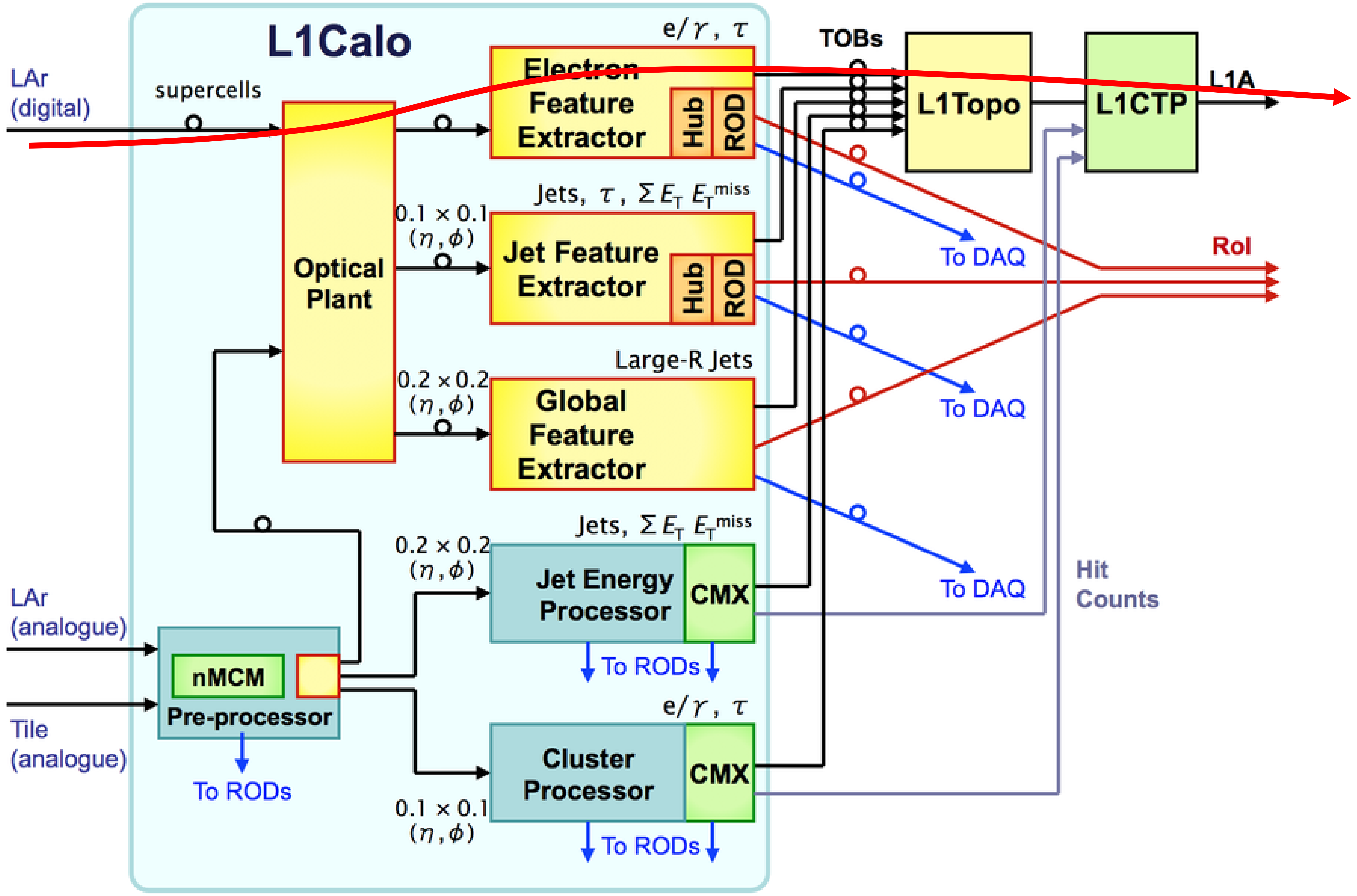
L1 Example: Electrons in the ATLAS Calorimeter Trigger in Run 3



- On-detector:
  - Signals from liquid argon calorimeter cells digitized and combined to form high granularity SuperCells
  - Supercells then formed into Trigger Towers



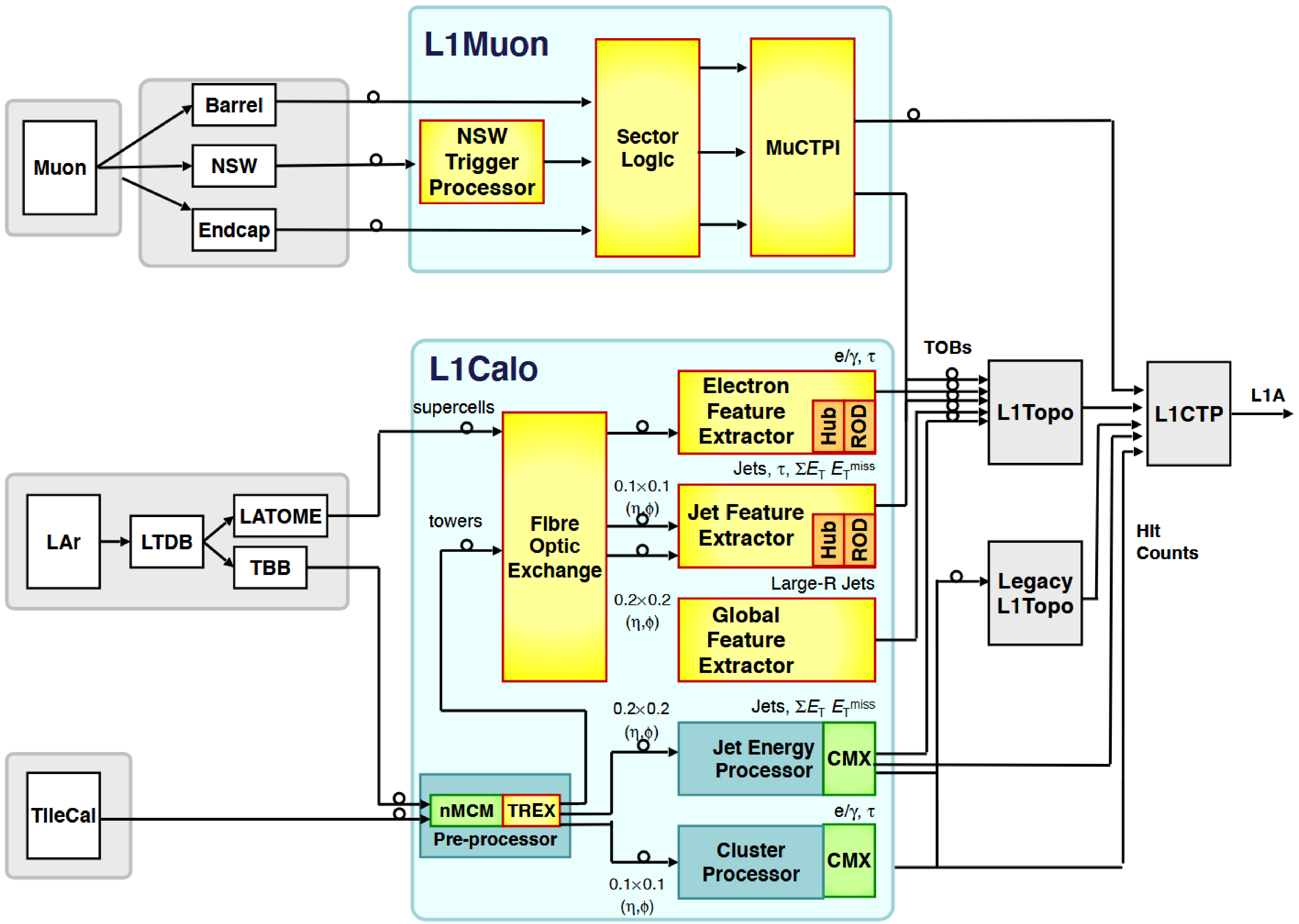
L1 Example: Electrons in the ATLAS Calorimeter Trigger in Run 3



- Off-detector - L1 Trigger
  - Electron Feature Extractor (eFEX)
    - Trigger Algorithm Performed in FPGAs
  - Forms cluster seed by finding local energy maximum in Super Cells
  - Cluster energy computed by summing adjacent Super Cells across calorimeter layers
  - Isolation cut based on cluster energy sum compared to surrounding area

L1 Example: ATLAS Central Trigger

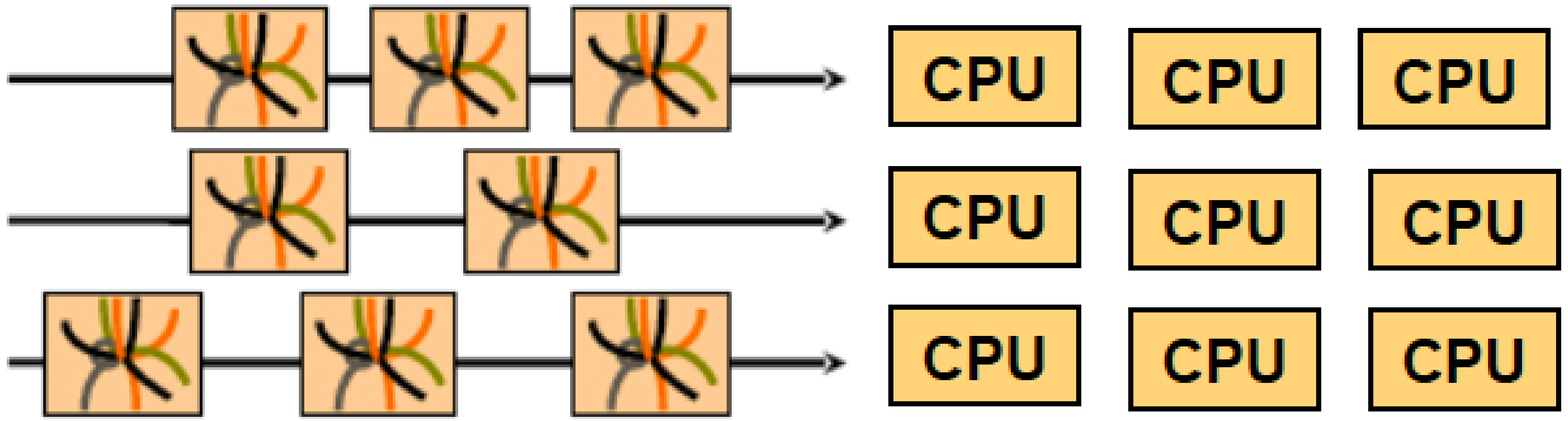
- L1 (Calo and Muon) use reduced granularity to provide fast (<2.5μs) information on particle candidates.
- May only be Muon/Calo but still a lot of info
  - Electrons, muons, taus, jets, total and missing energy
  - location, E_T, p_T threshold passed on
- Can also look at topological constraints
  - More complex checks: $\Delta\varphi, M_{JJ}, \Delta R$
- Central trigger processor (CTP) decides pass/fail
  - Electron in previous example will have to pass specific pT threshold to be accepted
- If pass, collate data from whole detector and send to **High Level Trigger** (rate ~100kHz)

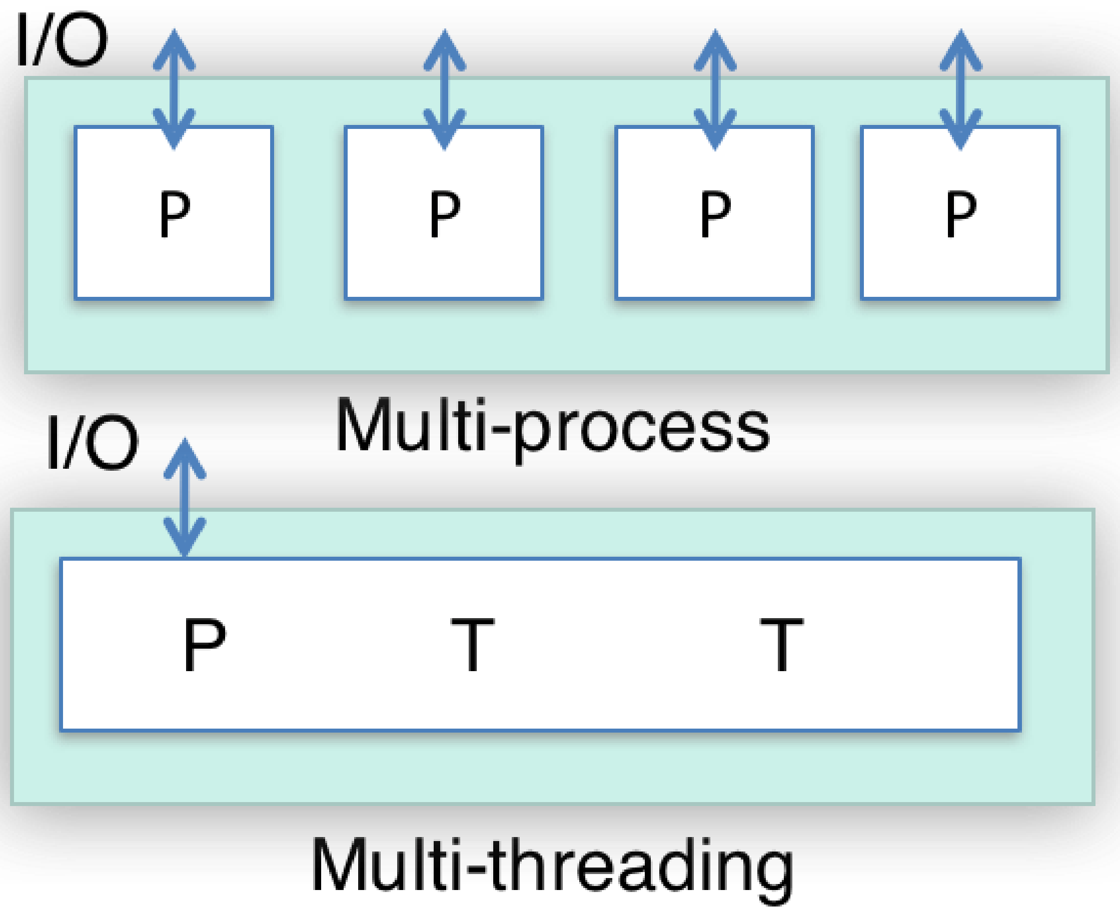


Science and Technology Facilities Council
Particle Physics

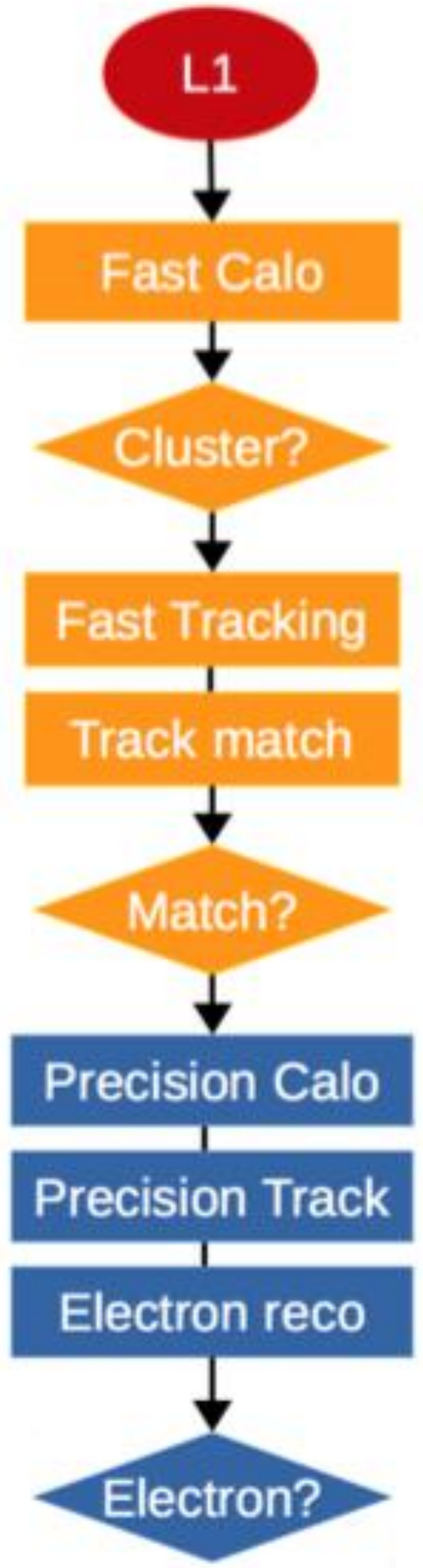
High Level Trigger: Software Trigger

- Still need to reduce rate for storage:
  - 100 kHz → 1kHz
- Networked computer farm
- Early rejection:
  - Reduce data and resources (CPU, memory….)
- Event-level parallelism:
  - Process more events in parallel
  - Multi-processing or/and multi-threading

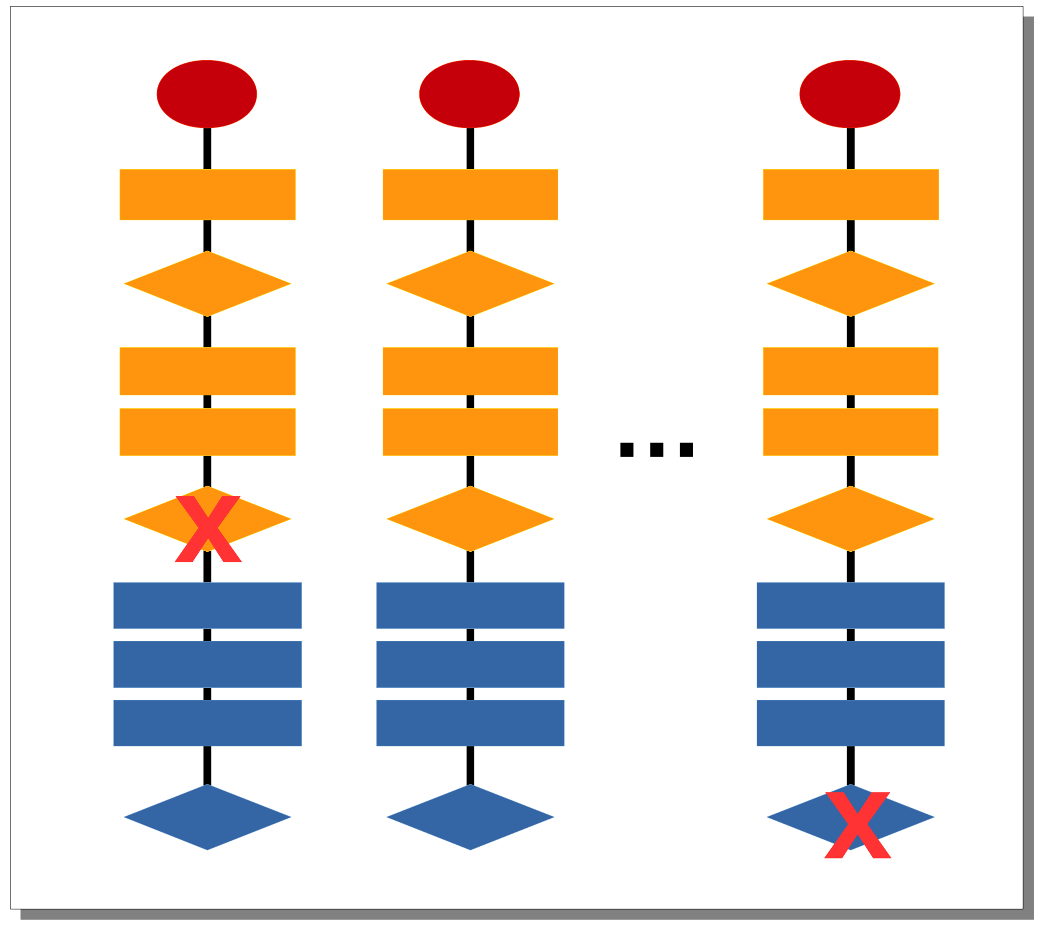




Example: HLT trigger algorithm - electrons



- Offline reconstruction too slow to be used directly
  - Takes >10s per event but HLT usually needs << 1s
- Requires step-wise processing with early rejection
- Fast reconstruction
  - Trigger-specific algorithms
  - L1-guided regional reconstruction
- Precision reconstruction
  - Offline (or very close to) algorithms
  - Full detector data available
- Stop processing as soon as one step fails



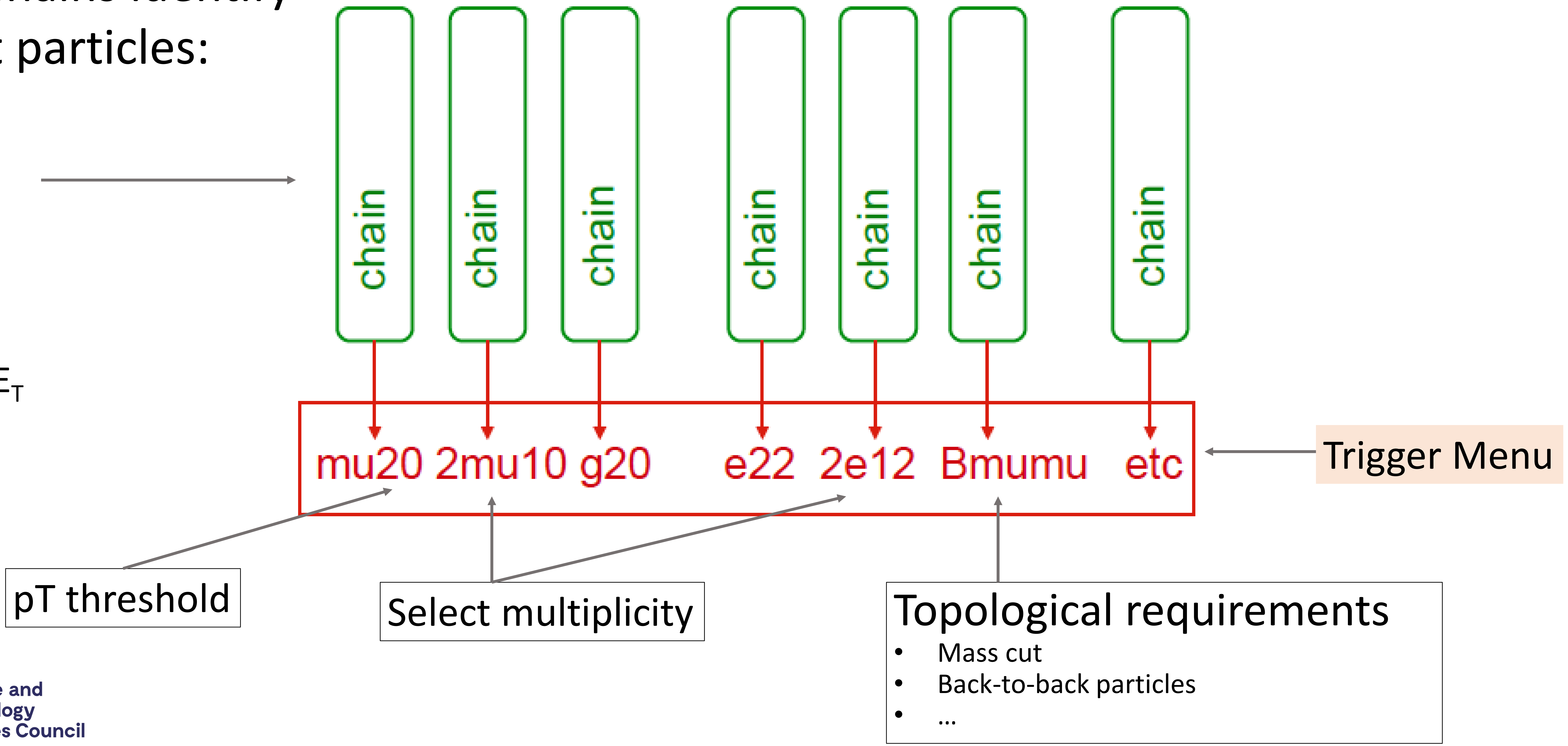
- Once again, to pass, our electron will ultimately have to satisfy a pT threshold

Trigger chain

Putting it all together: The Trigger Menu

Trigger Chains identify different particles:

- Electron
- Muon
- Photon
- Tau
- Jets
- B-jets
- Missing E_T
- …



Science and Technology Facilities Council
Particle Physics

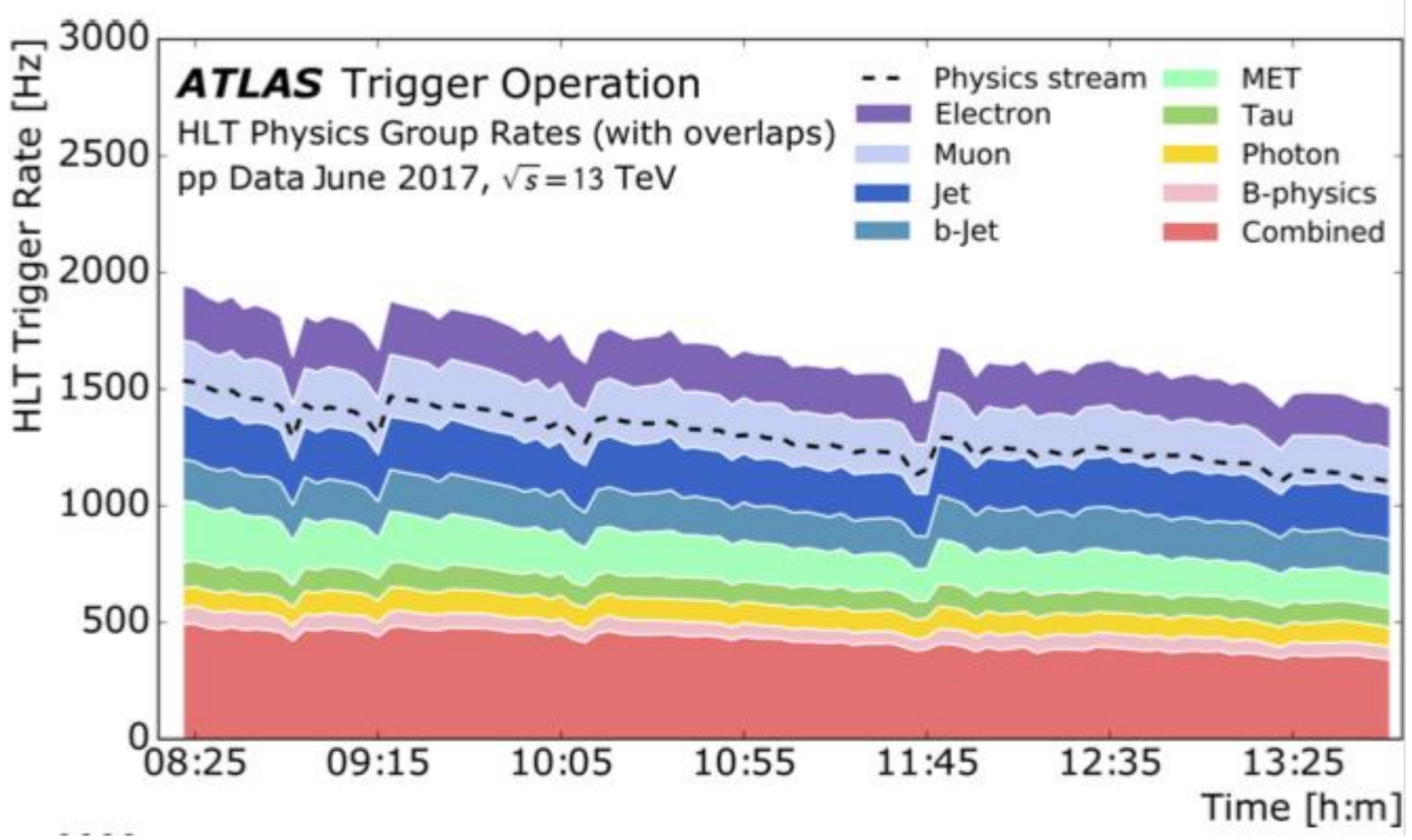
Trigger Menu – The Bigger Picture

- Goals:
  - Define the physics programme of the experiment, i.e. what is recorded
  - Each physics group defines a set of chains, corresponding to the individual triggers needed for their analysis programme

- In data taking, an event is recorded if at least one chain passes.

- Menu design is driven by:
  - Physics programme
  - Rate limitations at L1/HLT
  - Online resources (CPU/bandwidth)

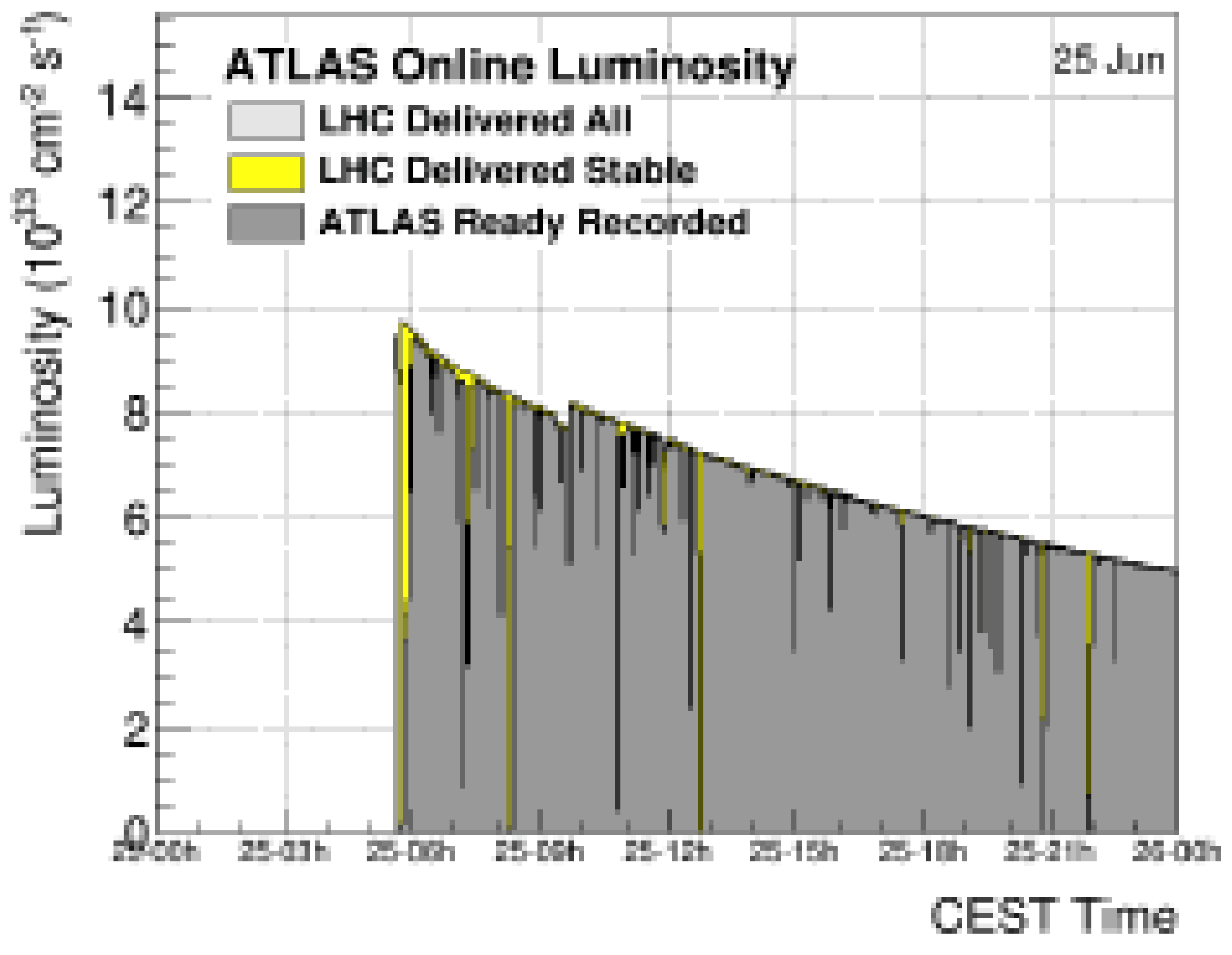
- Menu consists of:
  - Primary physics triggers
  - Support triggers – measure efficiencies and backgrounds
  - Calibration triggers – needed by detector groups (e.g. L1 only triggers)
  - Monitoring triggers – check everything is working (e.g. $Z \rightarrow ll$)

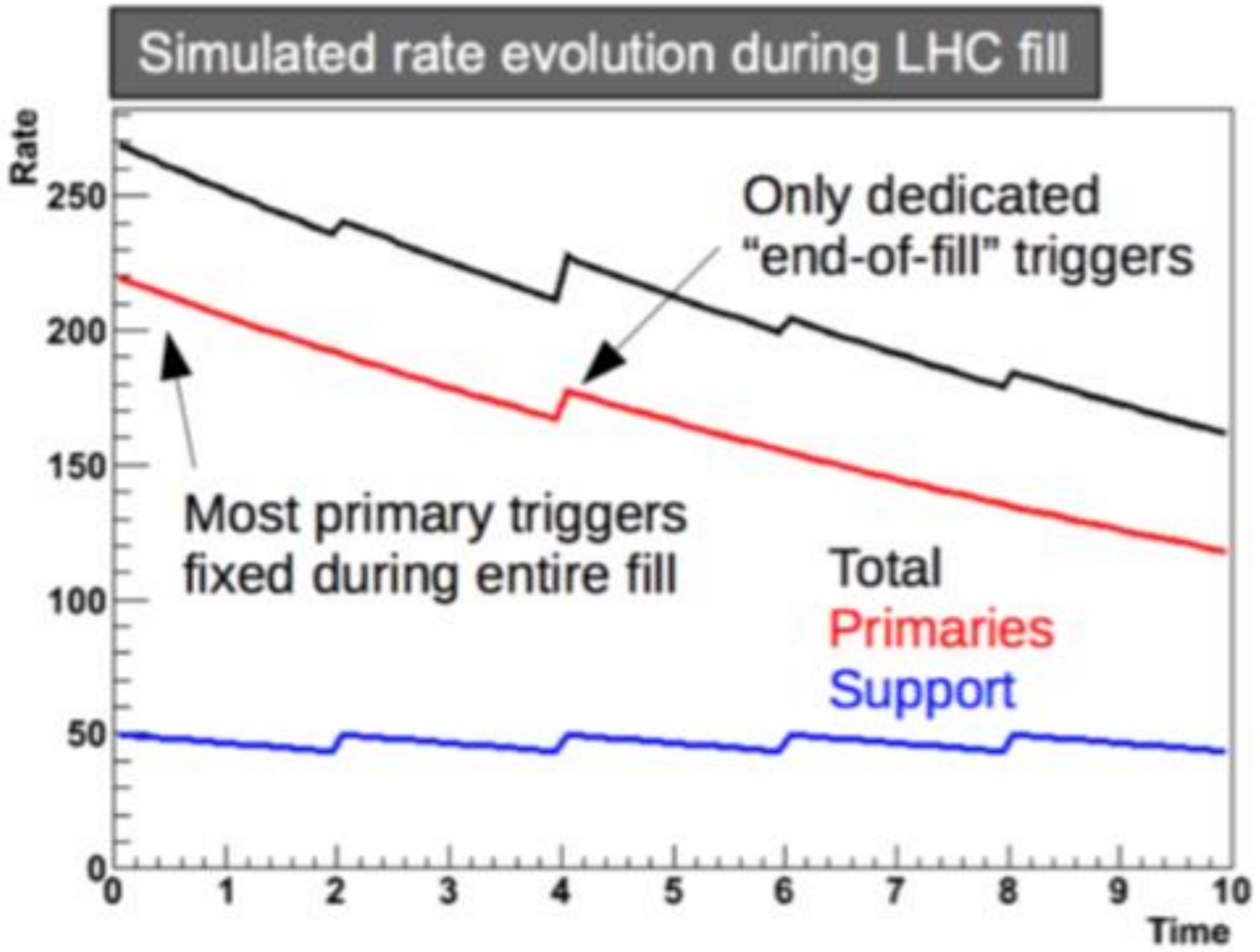


More details about menus, and how you go about measuring the performance of your triggers, in the extra slides!

Trigger Prescales

- Menu varies with luminosity, time and running conditions
- Not all triggers run at full rate:
  - Rate might be too high
  - Sub-sample may be enough to fulfil needs (support triggers)
- Trigger prescale reduces rate:
  - Prescale N means accept '1 in N' events passing this trigger
  - Prescale can be fractional
    - Fractional presales implemented through random number generation rather than simple periodic selection
    - Gives more flexibility to optimise rates and throughput
  - Apply L1 or HLT prescales
  - Can change during run – lower prescales as luminosity drops, add in 'end-of'fill' triggers to optimise use of resources





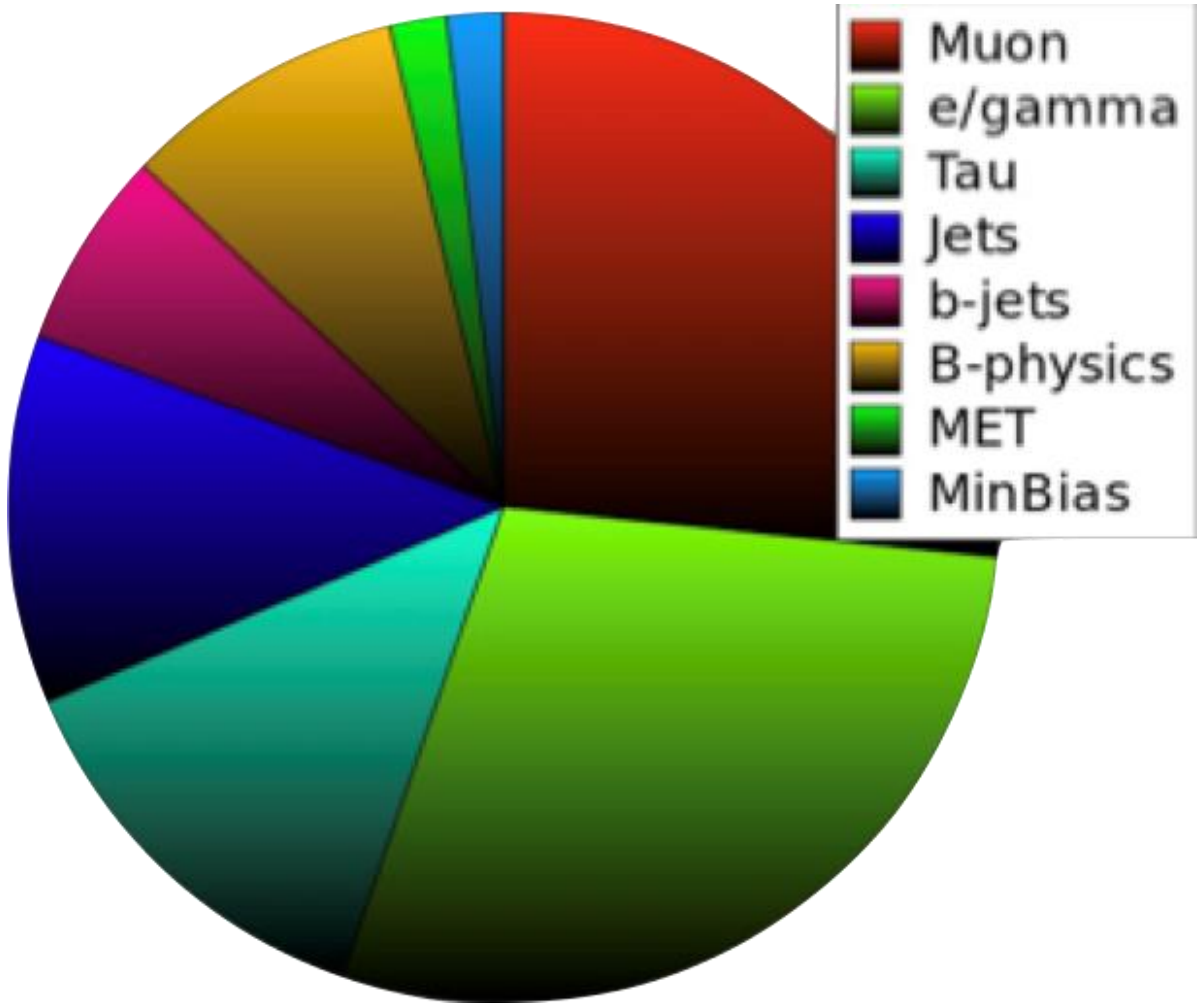
Science and Technology Facilities Council

Particle Physics

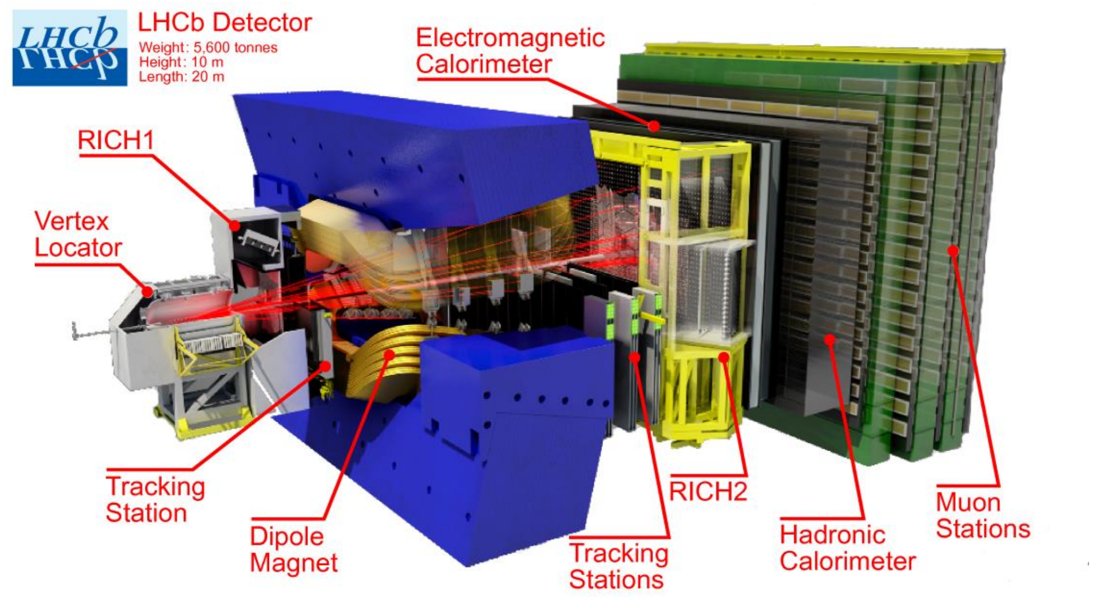
Rate Allocation: “Physics versus Bandwidth”

Lower thresholds always desirable, but the physics coverage must be balanced against online and offline computing cost

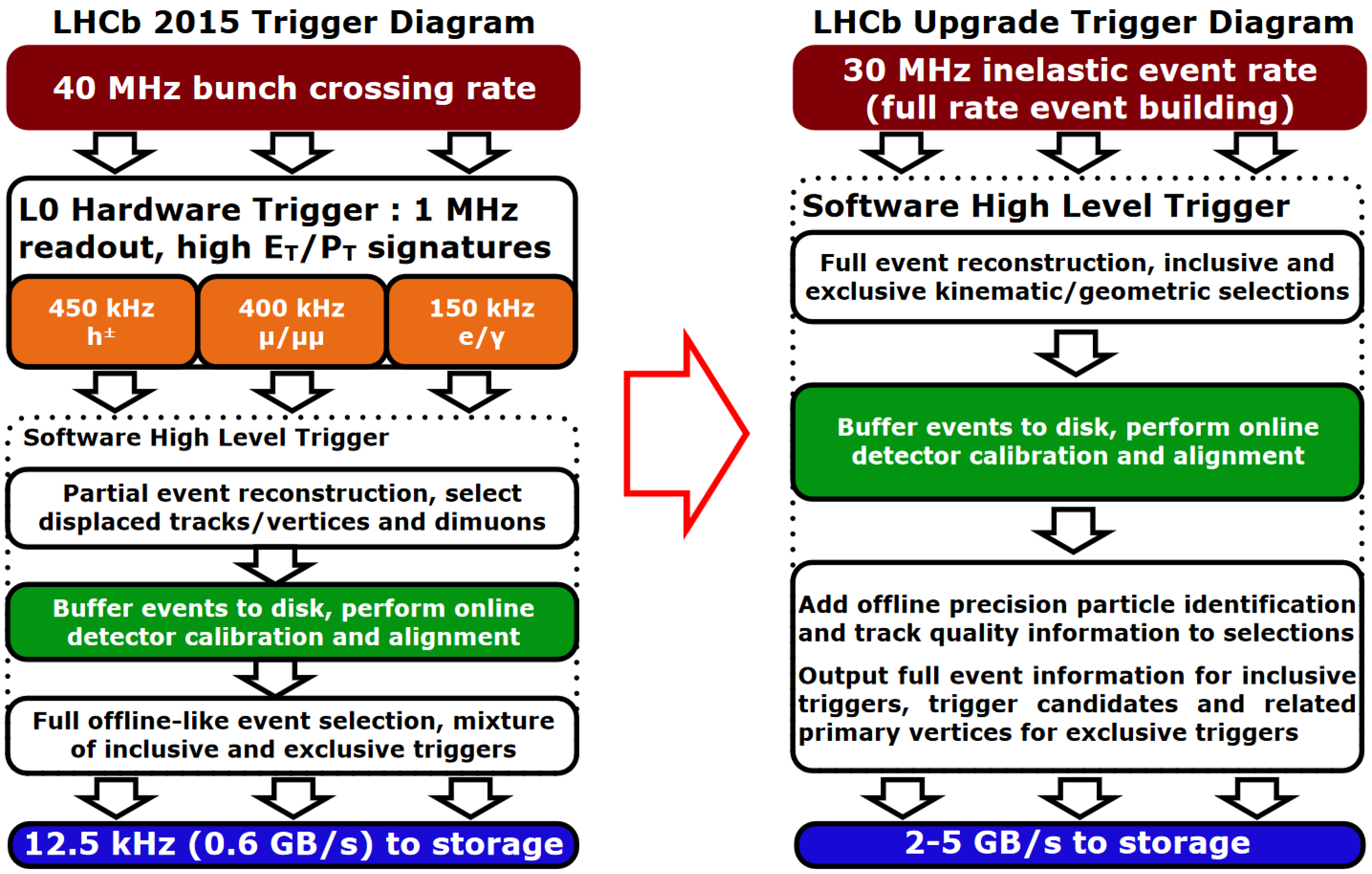
- Target : the final available DAQ bandwidth
- The rate allocation to each trigger signature
  - Physics goals (plus calibration, monitoring samples)
  - Required efficiency and background rejection
  - Bandwidth consumption
- When designing the menu: check predicted rates using previously recorded unbiased data



LHCb Trigger Upgrade



- From Run 3 (2022), LHCb has essentially done away with its hardware trigger
- Instead, all detector data are read out directly into High Level Trigger (software), which features GPU-based acceleration
- Possible due to (relatively) small event size (approx. 10% of CMS/ATLAS), but involves very complex reconstruction at high rates in HLT



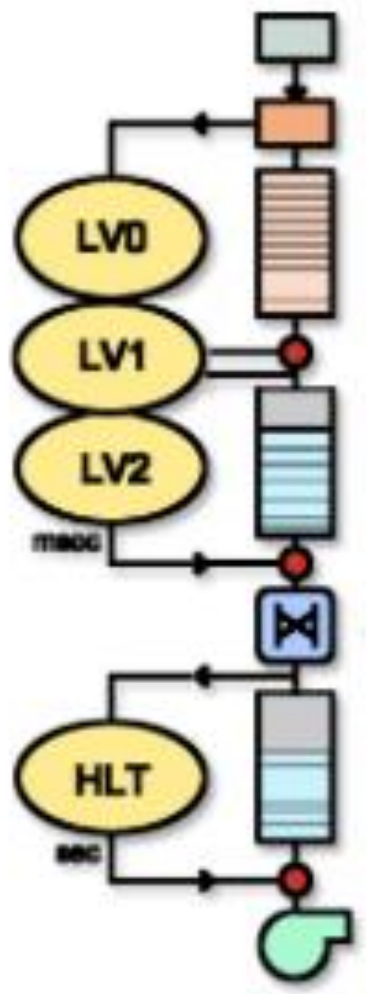
Science and Technology Facilities Council
Particle Physics

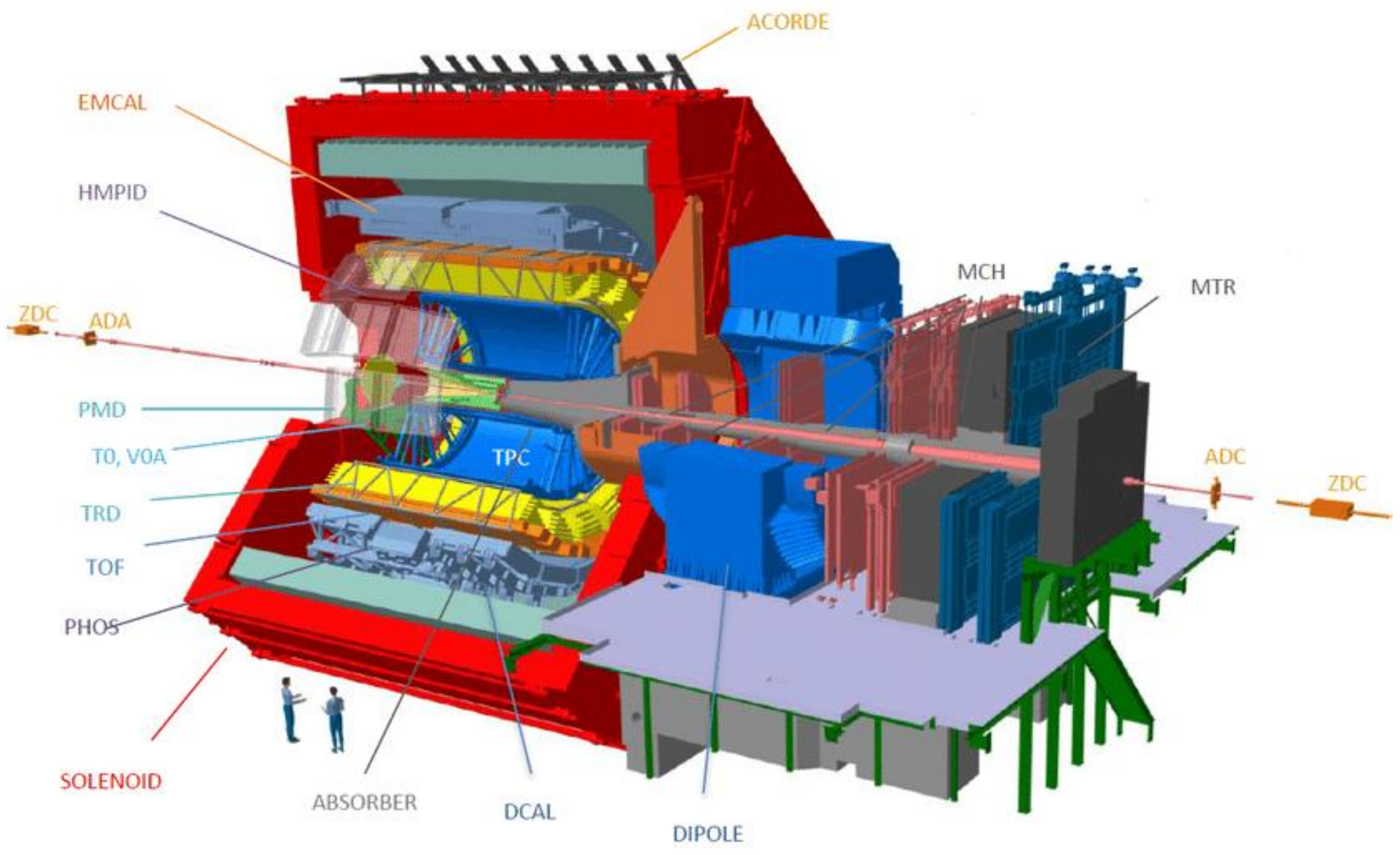
ALICE Trigger

- Run 1 and 2 Trigger
  - 3 hardware levels
  - 1 software

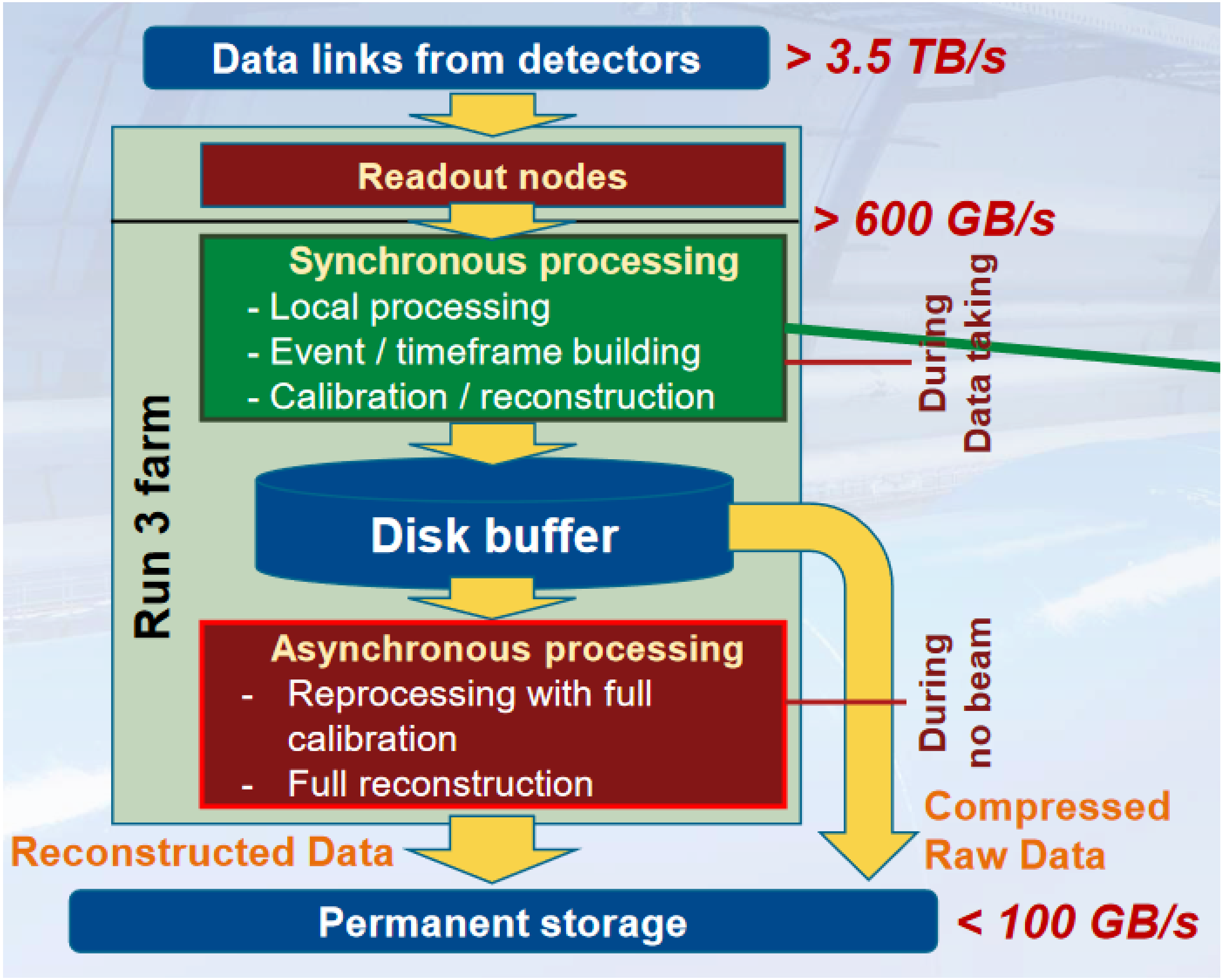
Latency
1.2μs
6.5μs
100μs

Software



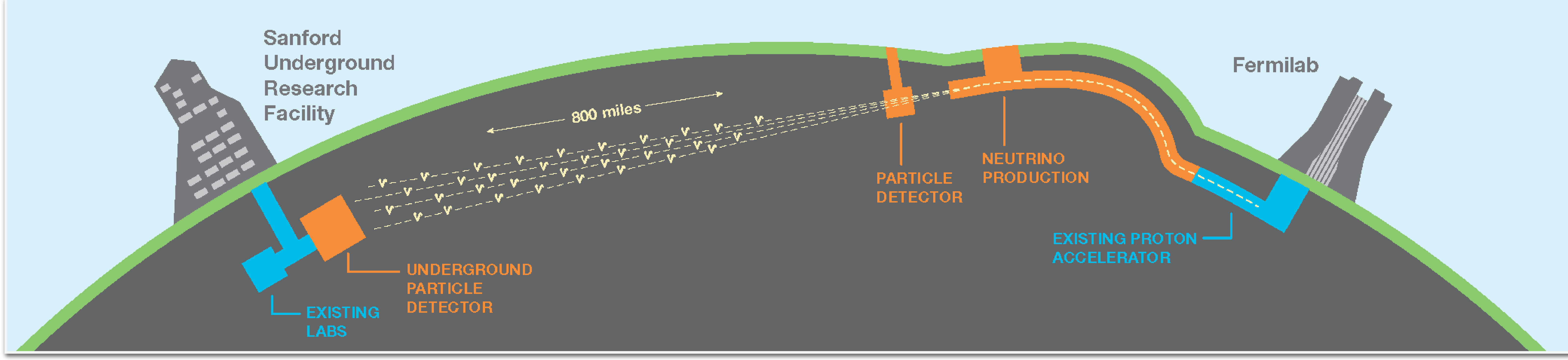


- In Run 3, moved to integrated online-offline ‘O^2’ system with continuous readout
- Introduction of new Fast Interaction Trigger (FIT) minimum bias trigger with high precision timing and luminometry
- Making use of GPU-accelerated reconstruction

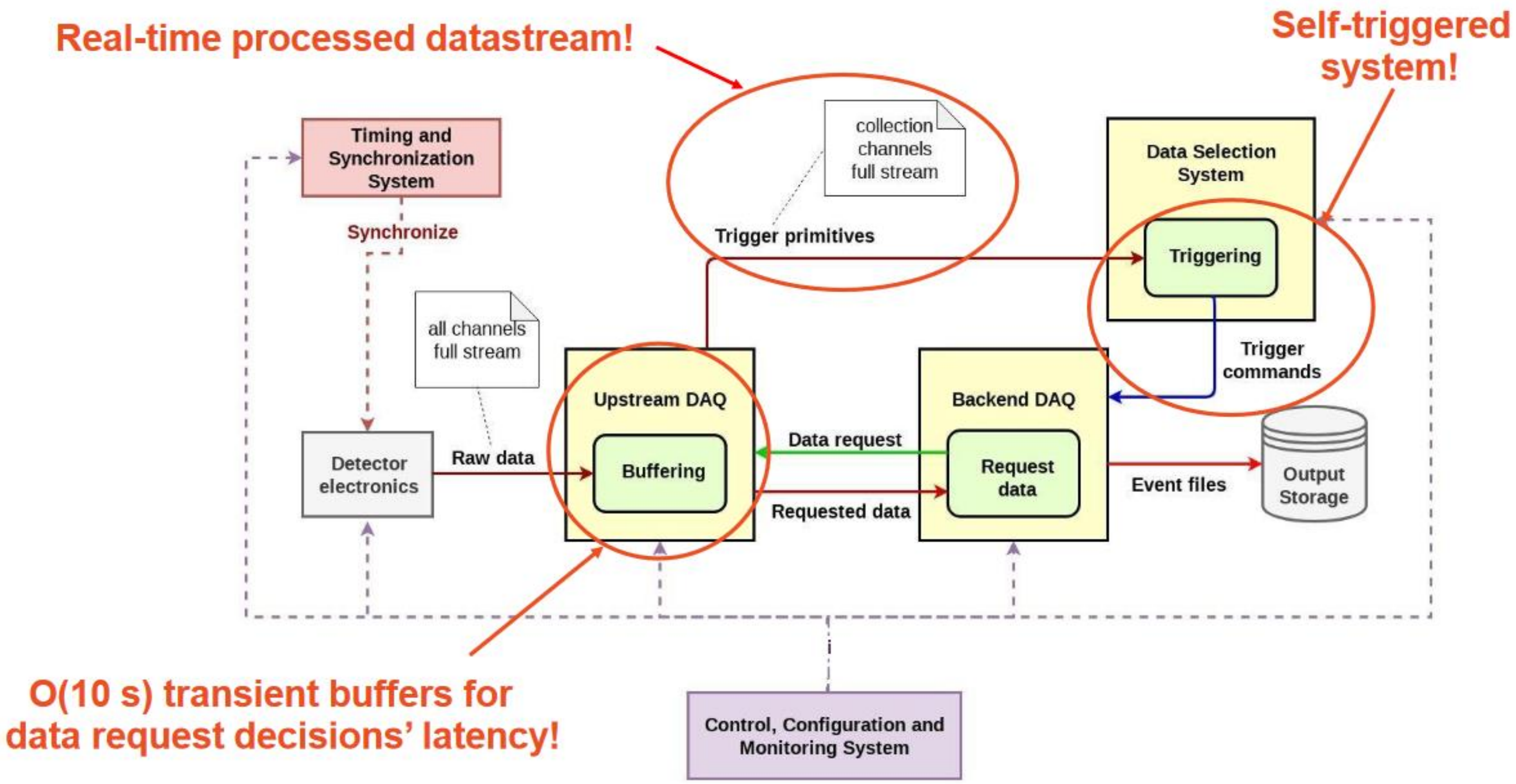


Science and Technology Facilities Council
Particle Physics

DUNE Trigger System

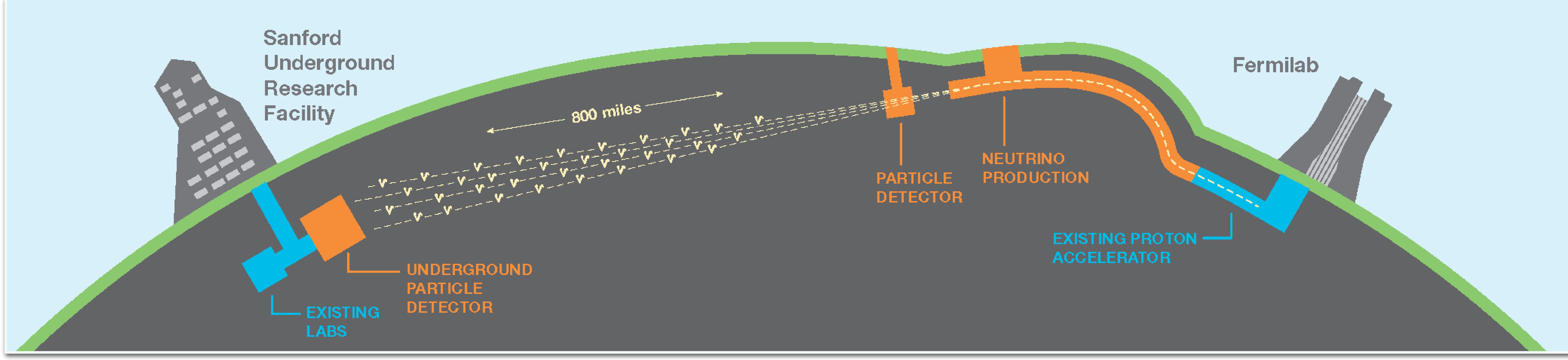


- Single trigger level with two modes of operation:
  - Interaction triggers (interesting localized activity somewhere in the detector)
    - Examples: beam triggers, cosmic rays and photon detection.
  - Supernova Neutrino Burst (SNB) triggers (sufficient activity in the detector to suggest a SNB)
    - All data are stored for 100 sec window including O(10 s) before the trigger signal.
    - External Trigger Interface (ETI) to pass messages to Supernova Early Warning System (SNEWS)

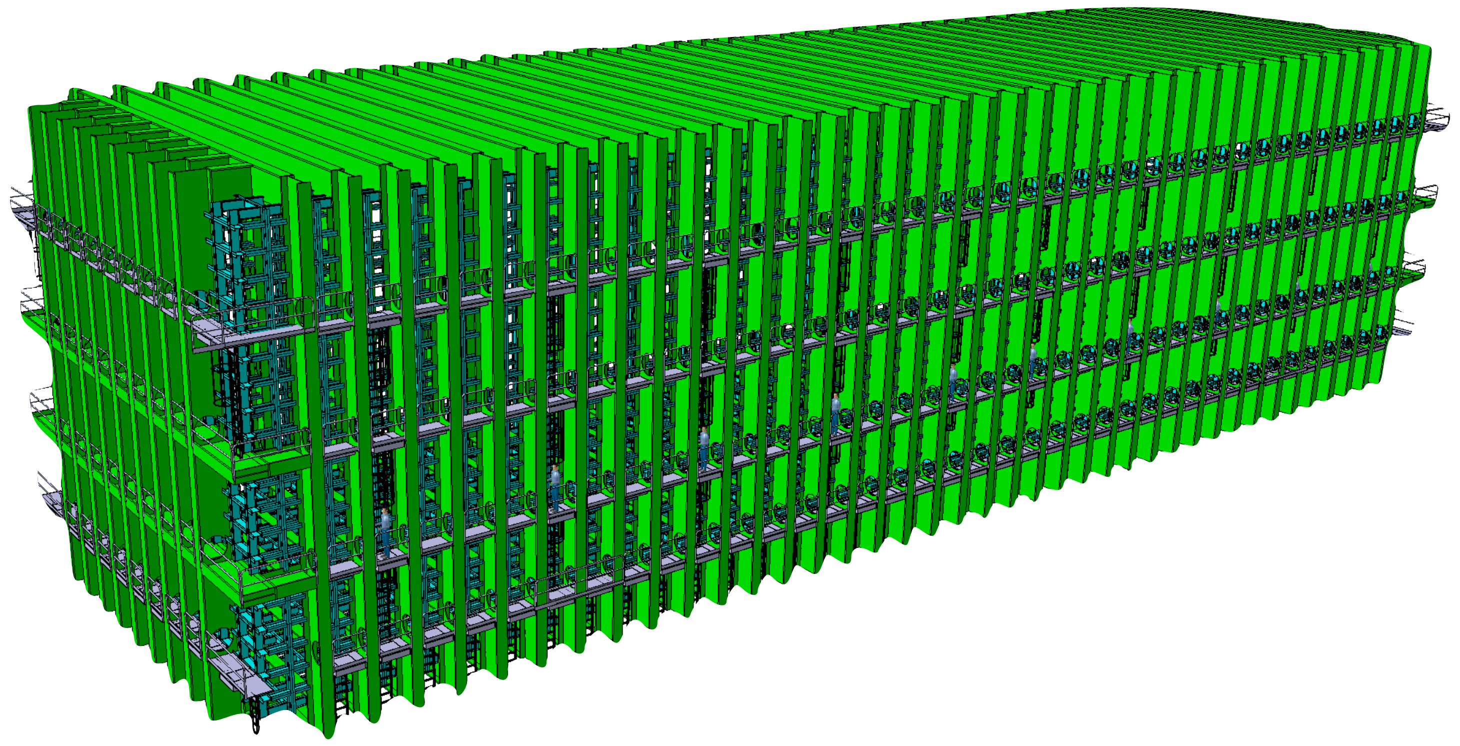


Science and Technology Facilities Council
Particle Physics

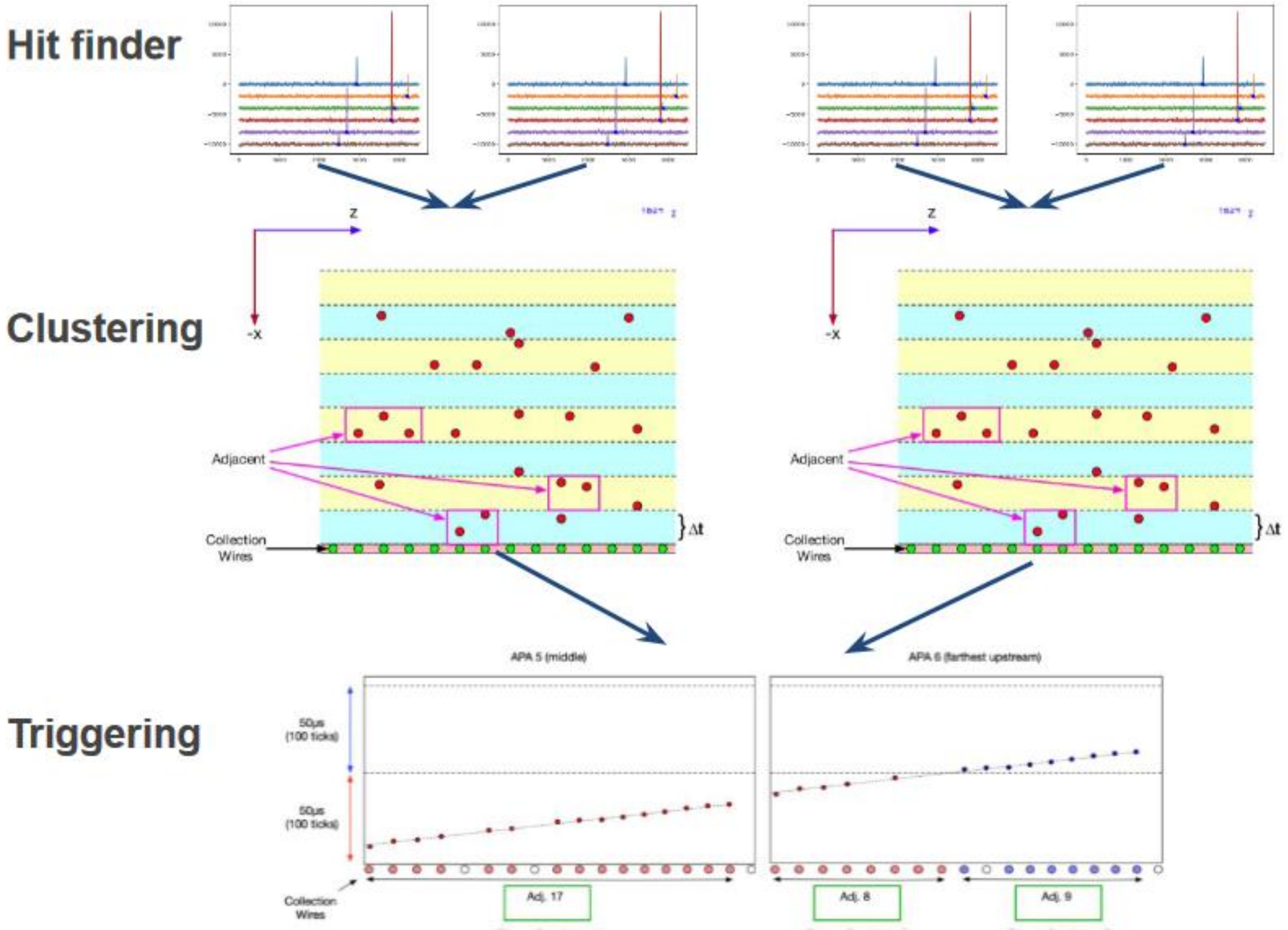
DUNE Trigger System

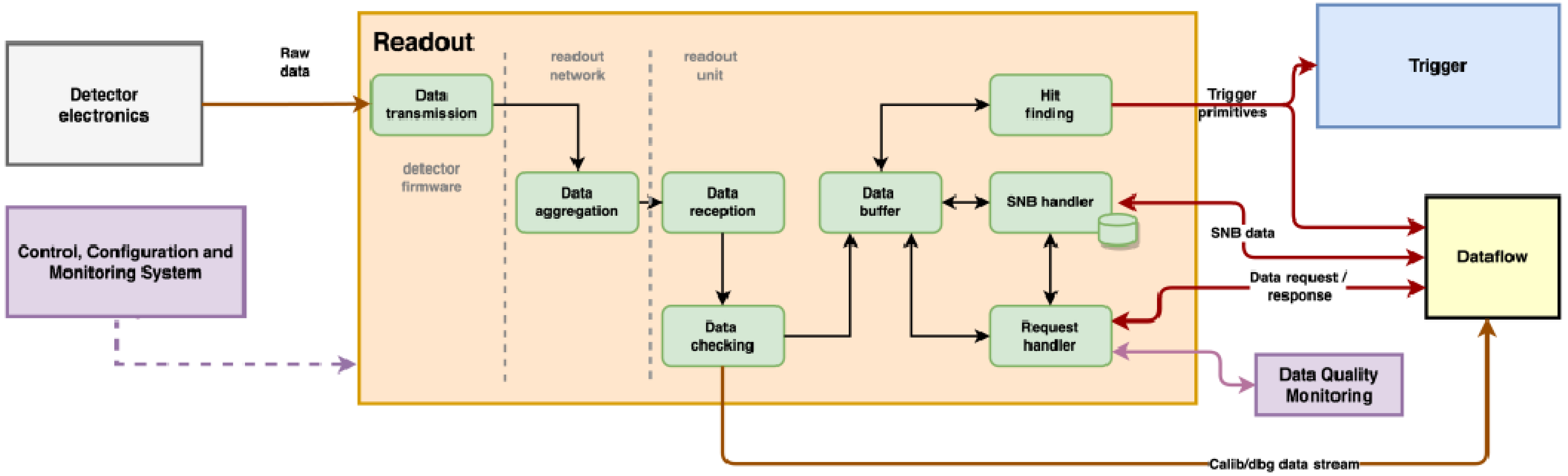


- Software-based (SIMD) hit finding to produce Trigger Primitives, followed by clustering and selection



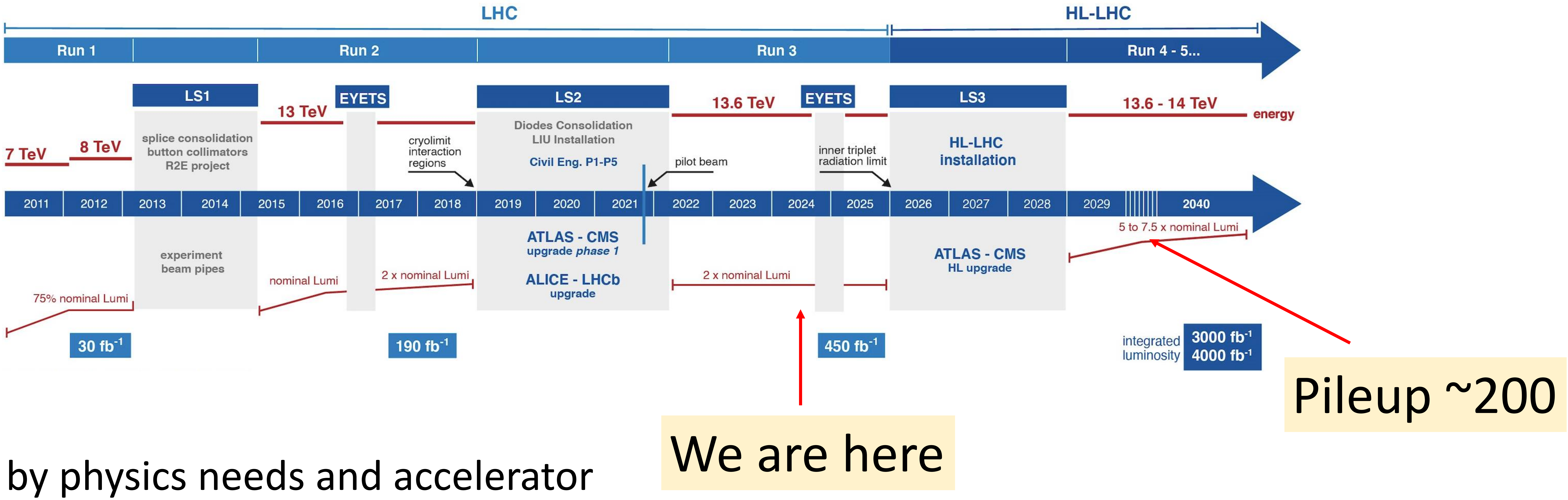
DUNE Far Detector comprised of four 17 kT Liquid Argon Time Projection Chamber (LArTPC) modules





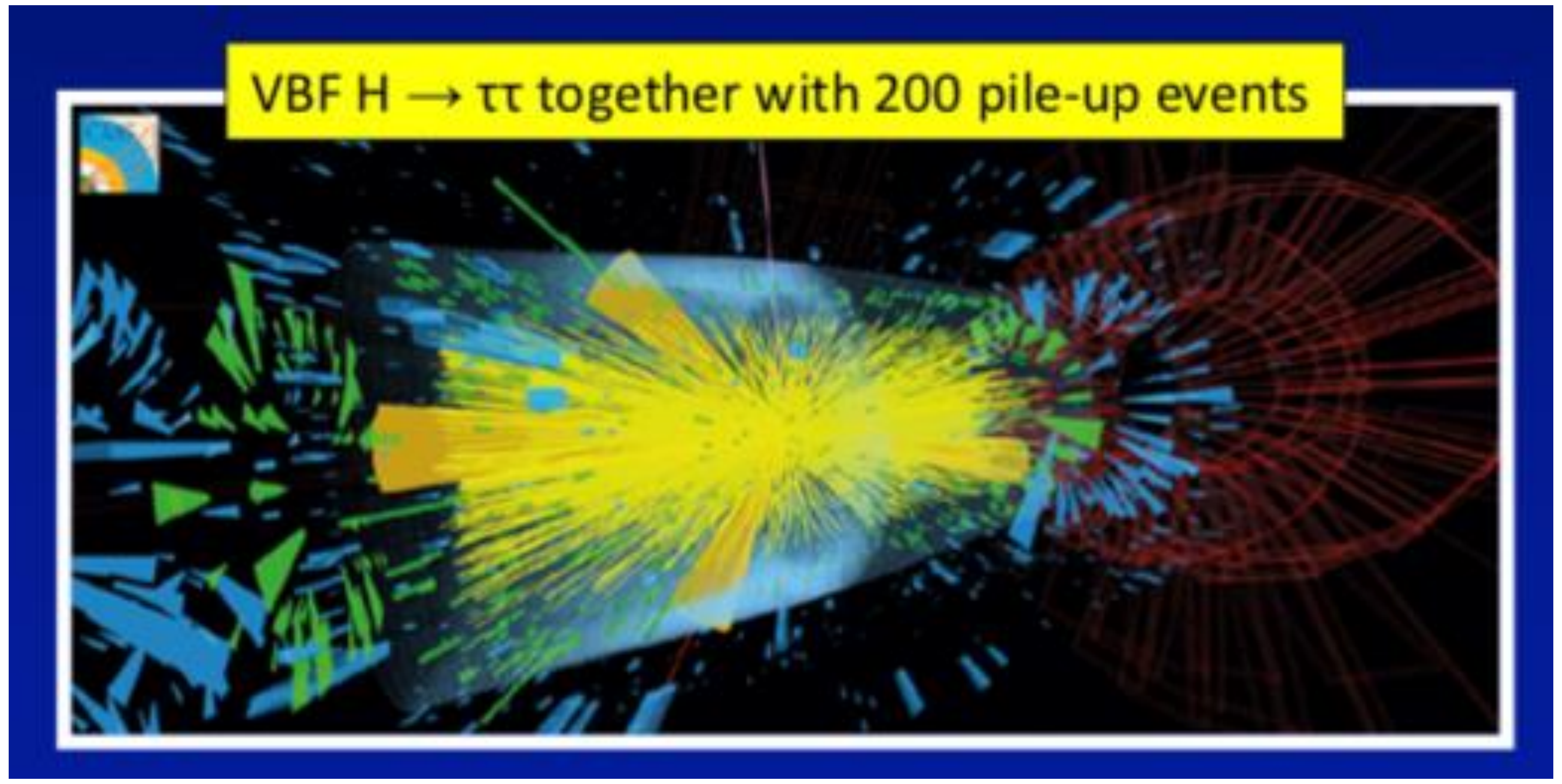
Science and Technology Facilities Council

High Luminosity LHC



- Trigger driven by physics needs and accelerator environment

- Future HL-LHC:
  - ~200 interactions per bunch crossing (pileup) – c.f. ~ 65 in Run 3
  - Higher granularity detectors (now billions of channels)
  - All implies much larger event size



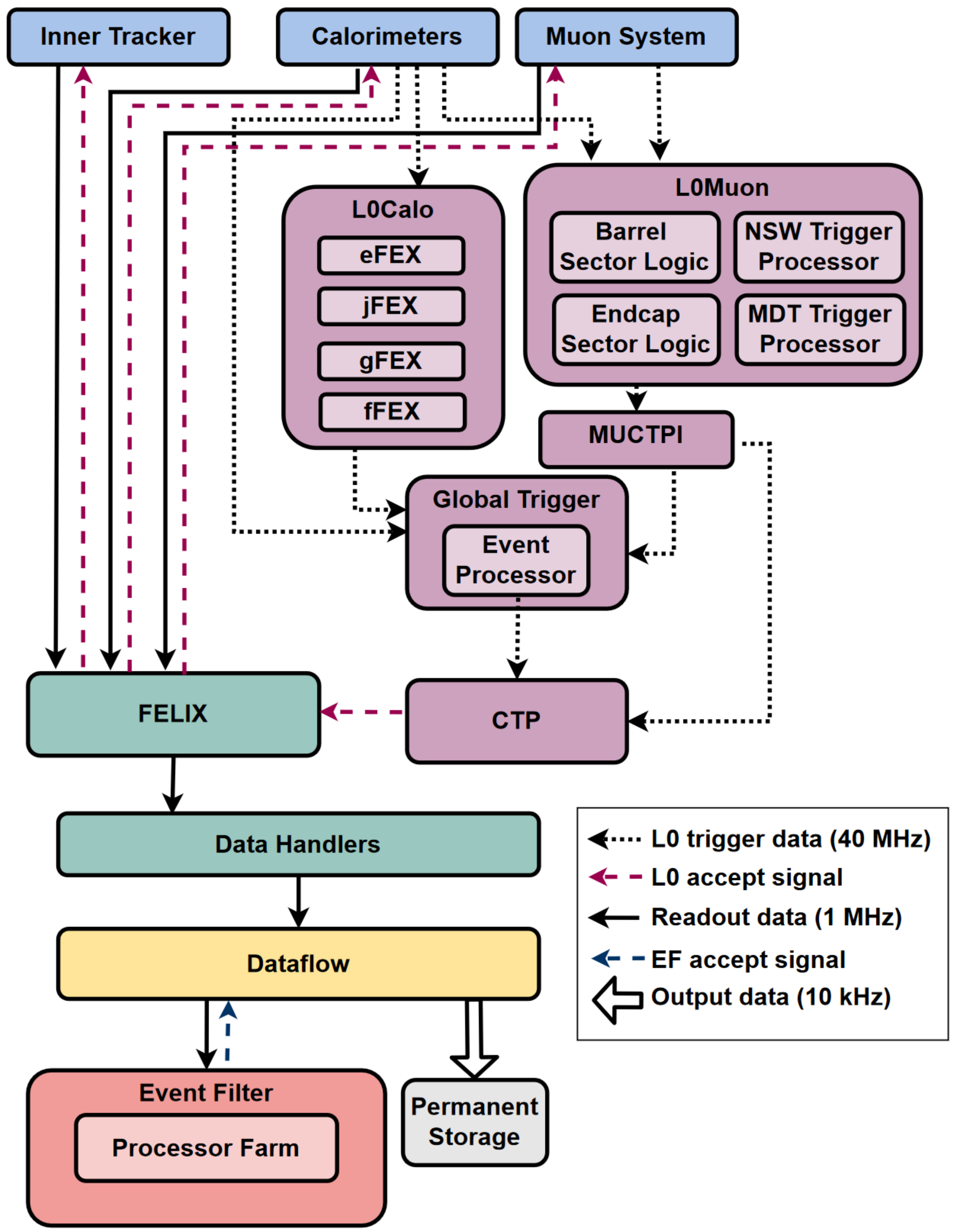
Science and Technology Facilities Council

Particle Physics

High Luminosity LHC

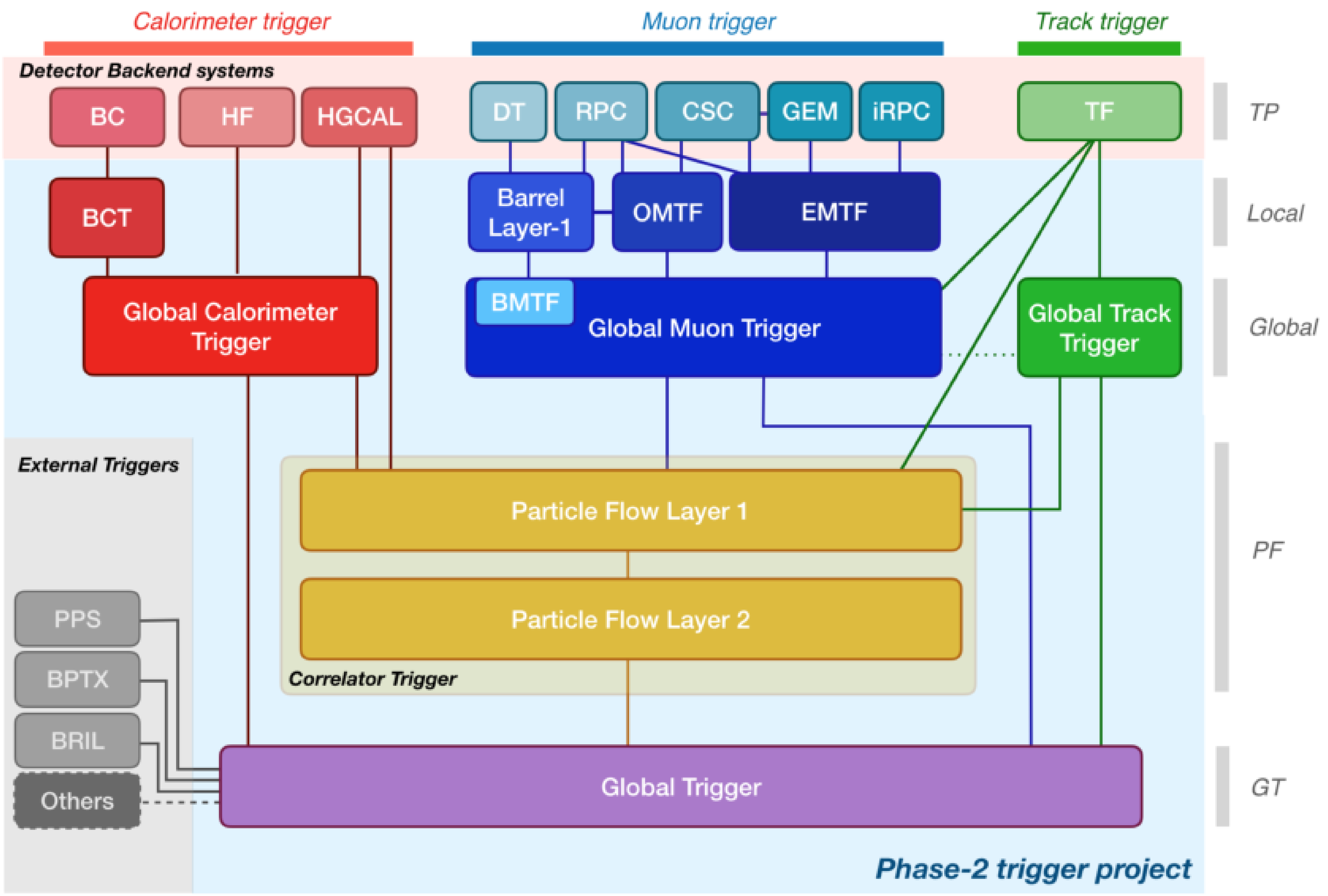
- BUT physics need to keep trigger thresholds close to today's values
- ATLAS:
  - L1 latency increases to 10μs (2.5μs today)
  - Event size increase to 4.6 MB (1-2 MB today)
  - L1 trigger rate increases 1 MHz (100 kHz today)
  - Rate to storage ~10 kHz (~1 kHz today)

- Use higher granularity, more complex algorithms throughout
  - New hardware featuring more advanced FPGAs
  - New Global Trigger, to perform offline-like algorithms on calorimeter and muon data in hardware
  - New FEX for forward region (fFEX)
- Investigate FPGA and GPU accelerators for use in HLT (Event Filter)
- Deployment of Machine learning-based algorithms throughout
- Increased used of Trigger Level Analysis

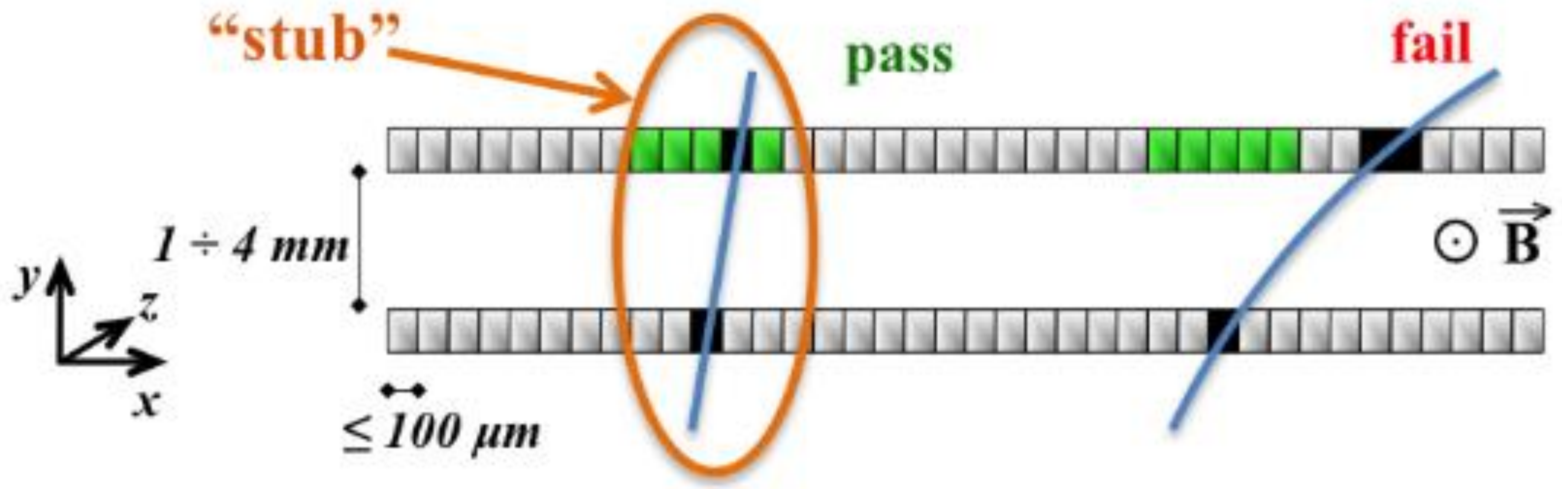


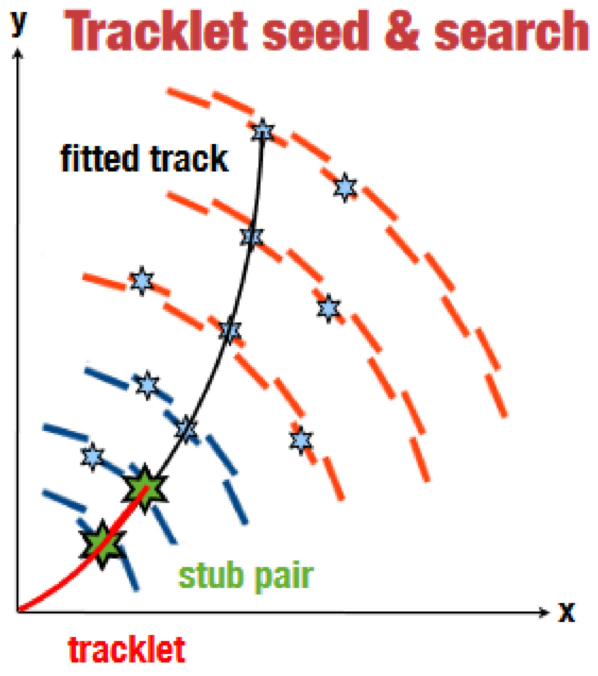
High Luminosity LHC

- CMS:
  - L1 latency increases to 12.5 μs (4 μs today)
  - Event size increase to 7.4 MB
  - L1 trigger rate increases 750 kHz (100 kHz today)
  - Rate to storage ~7.5 kHz

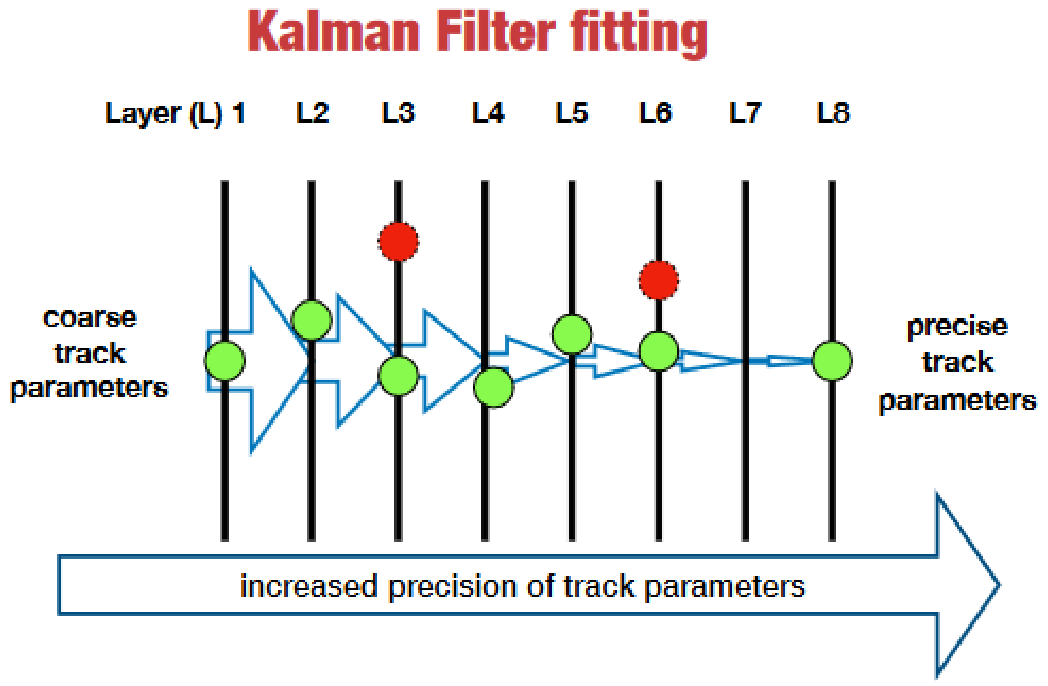


- Introduce tracking trigger in hardware at L1
- Investigate FPGA and GPU accelerators for use in HLT (Event Filter)
- Deployment of Machine learning-based algorithms throughout
- Increased used of Trigger Level Analysis





+



Summary

- *"A system that uses simple criteria to rapidly decide which events in a particle detector to keep only a small fraction of the total can be recorded."*

- Trigger strategy is a trade-off between physics requirements and affordable systems and techn

- Main trigger requirements:
  - High efficiency but with control of rates
  - Knowledge of effect of trigger selection on signal and background events
  - Flexibility and redundancy

- Most triggers are underpinned by some fundamental concepts, but there is a lot of innovation happening now right across the field as technology continues to evolve
  - This will be crucial as we face the challenges of HL-LHC and further afield.

- Trigger is vital - if don't trigger an event it is lost forever!

Science and Technology Facilities Council

Particle Physics

Questions?

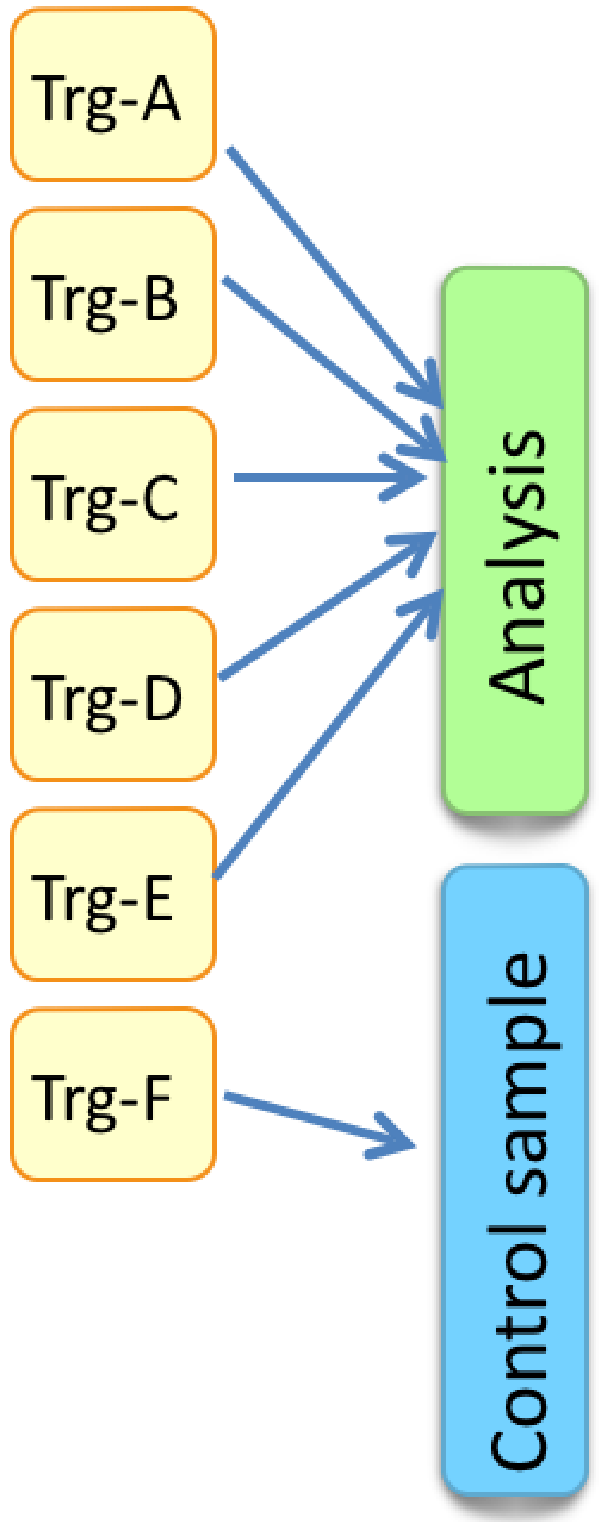
Extra Slides

Designing a Menu

- Defines the physics programme/reach of the experiment
- Collection of physics trigger, associated back-ups, triggers for calibration and monitoring

- It must be
  - Redundant - to ensure the efficiency measurement
  - Flexible - to adapt to changes of the environment and the physics goals, e.g. detectors, machine luminosity,...

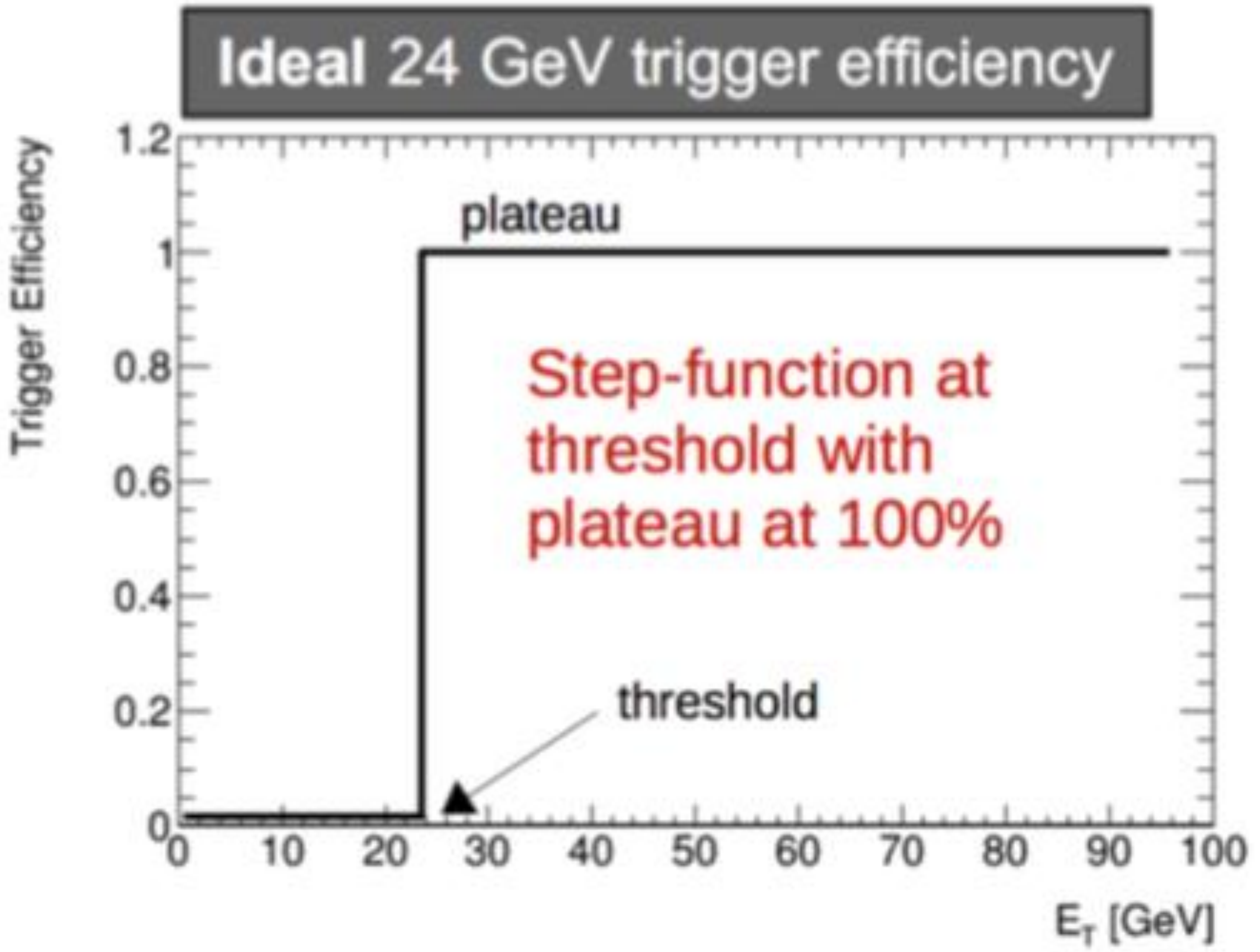
- Central to the physics programme
  - Each analysis served by multiple triggers and different samples from the most inclusive to the most exclusive
  - Ideally, it will collect events (some, at least) from all relevant processes and provide physics breadth and control samples



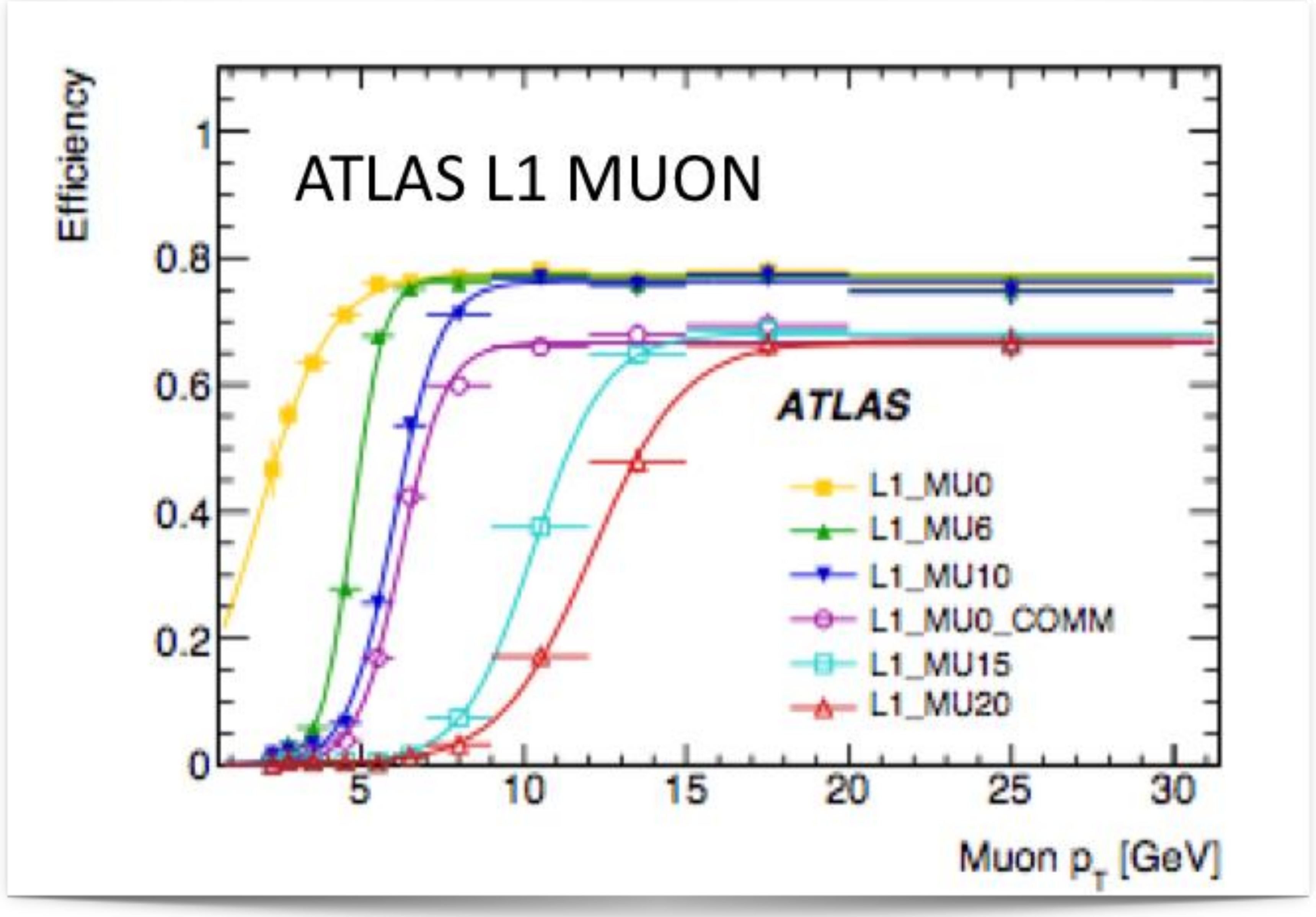
Physics analysis perspective

Physics analyser wants to know:

- Where is the trigger **turn-on**?(maximal efficiency wrt offline objects)
- What is the peak **efficiency**? (is it 100%, or do I need a scale factor)
- Is it **prescaled**? (Do I need a correction?)



- Resolutions
- Inefficiencies
- Online/offline differences



Measuring efficiency

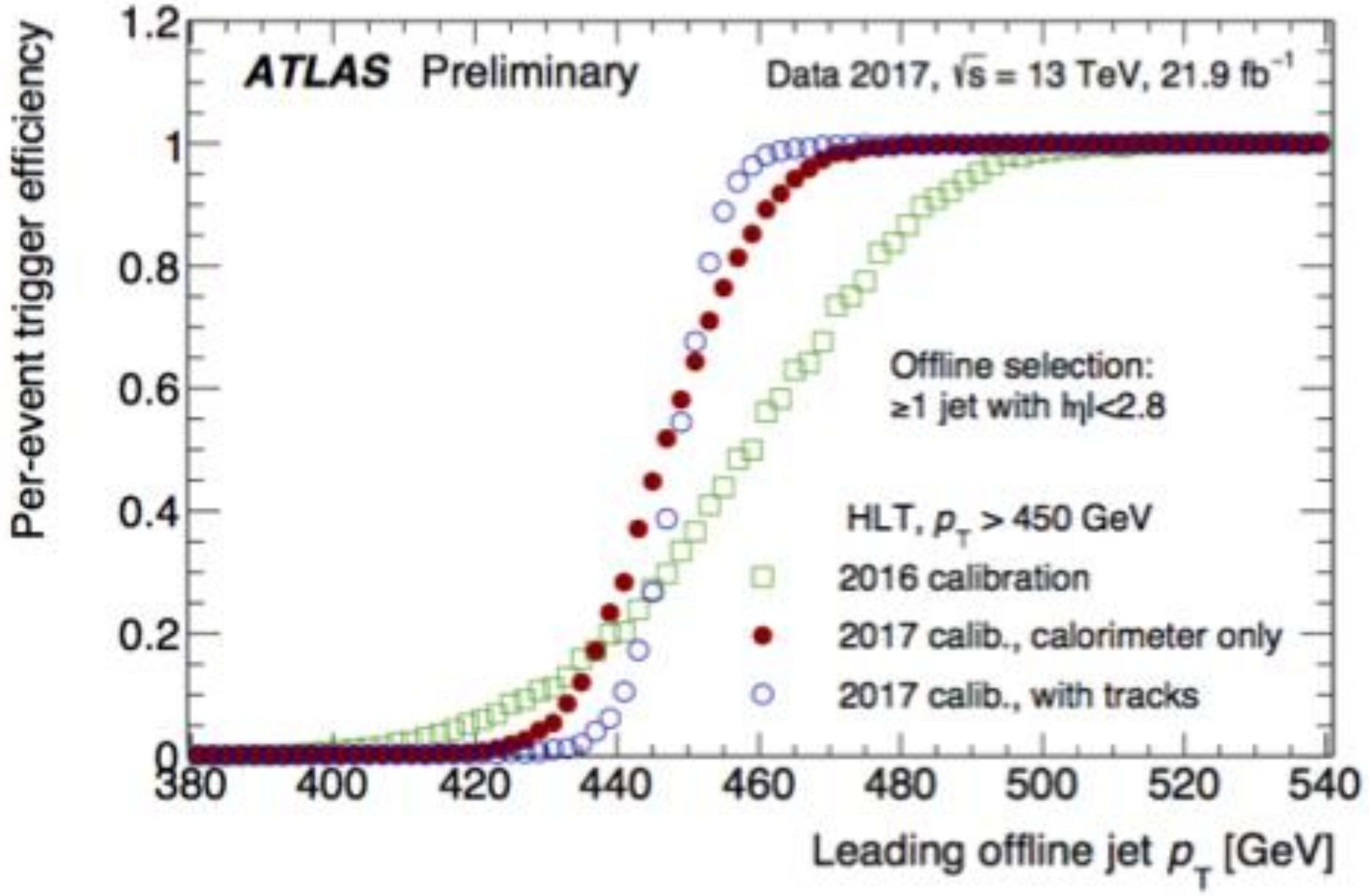
How to measure the efficiency of your trigger?

Define efficiency w.r.t the offline reconstructed objects

$$\epsilon_{trigger} = \frac{N_{trigger}}{N_{offline}}$$

Various methods:

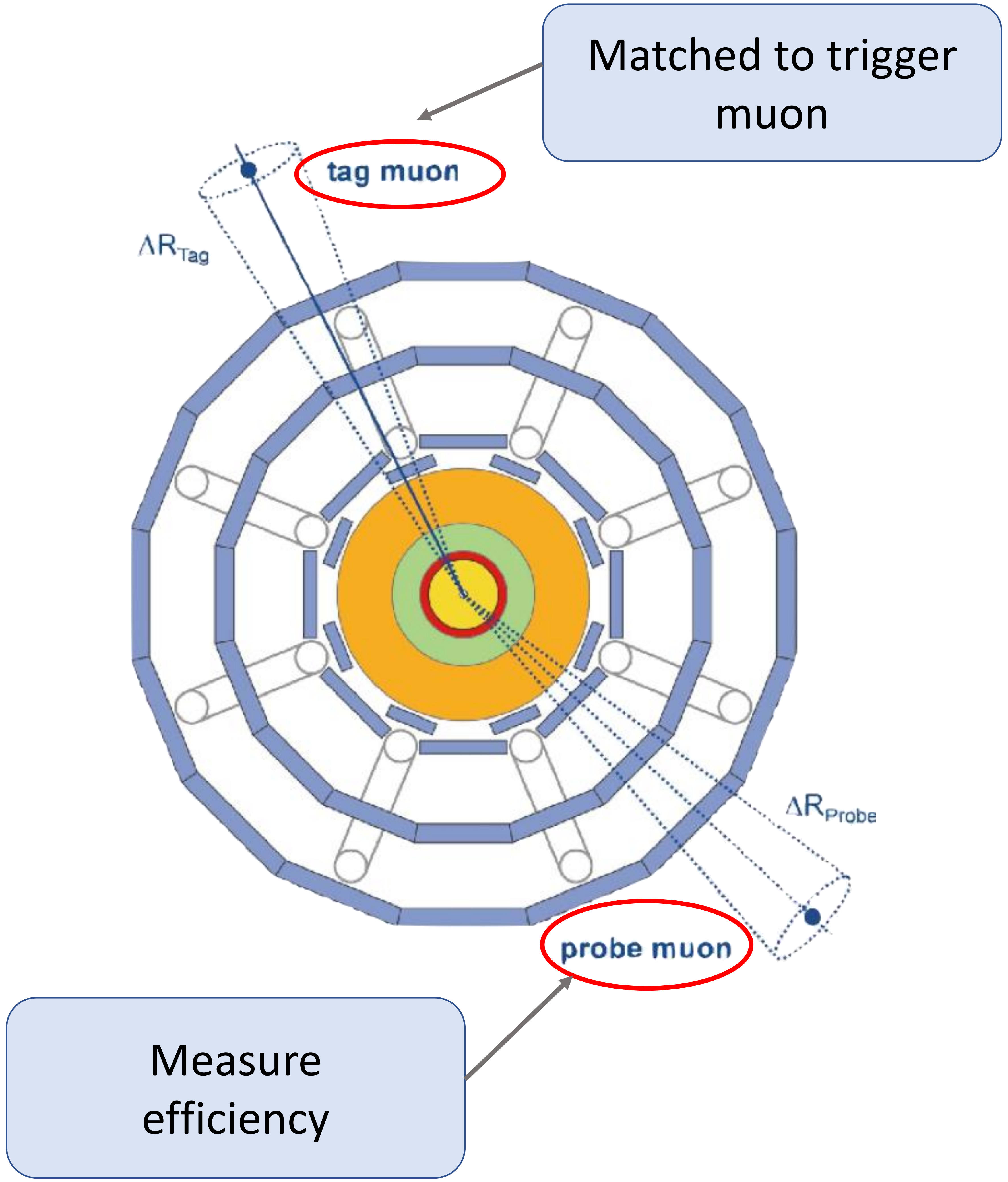
- ➢ Tag-and-probe
  - ➢ Trigger on one particle (tag), measure how often another (probe) passes trigger, e.g. $Z \rightarrow ll, \quad J/\psi \rightarrow ll$
- ➢ Boot-strap
  - ➢ Use sample triggered by looser (prescaled) trigger to measure efficiency of higher threshold trigger
- ➢ Orthogonal trigger
  - ➢ Use sample triggered by one trigger (e.g. muon trigger) to measure efficiency of a different, independent trigger (e.g. jet trigger)
- ➢ Simulation/emulation : MC



Science and Technology Facilities Council

Particle Physics

Measuring efficiency : tag-and-probe



Exploit a well know physics process (e.g $Z \rightarrow ll,\ \ J/\psi \rightarrow ll$) to select a very clean sample

Applicable on specific signatures (typically leptons)
Requires careful fake control

How?
Online:
- Trigger on independent signature (e.g. single muon)

Offline:
- Reconstruct the event, e.g. 2 muons in Z mass window (use tight selection for high purity)
- Match offline muon to trigger muon (Tag)
- Measure trigger efficiency for other (Probe) muon

Efficiency as a function of:

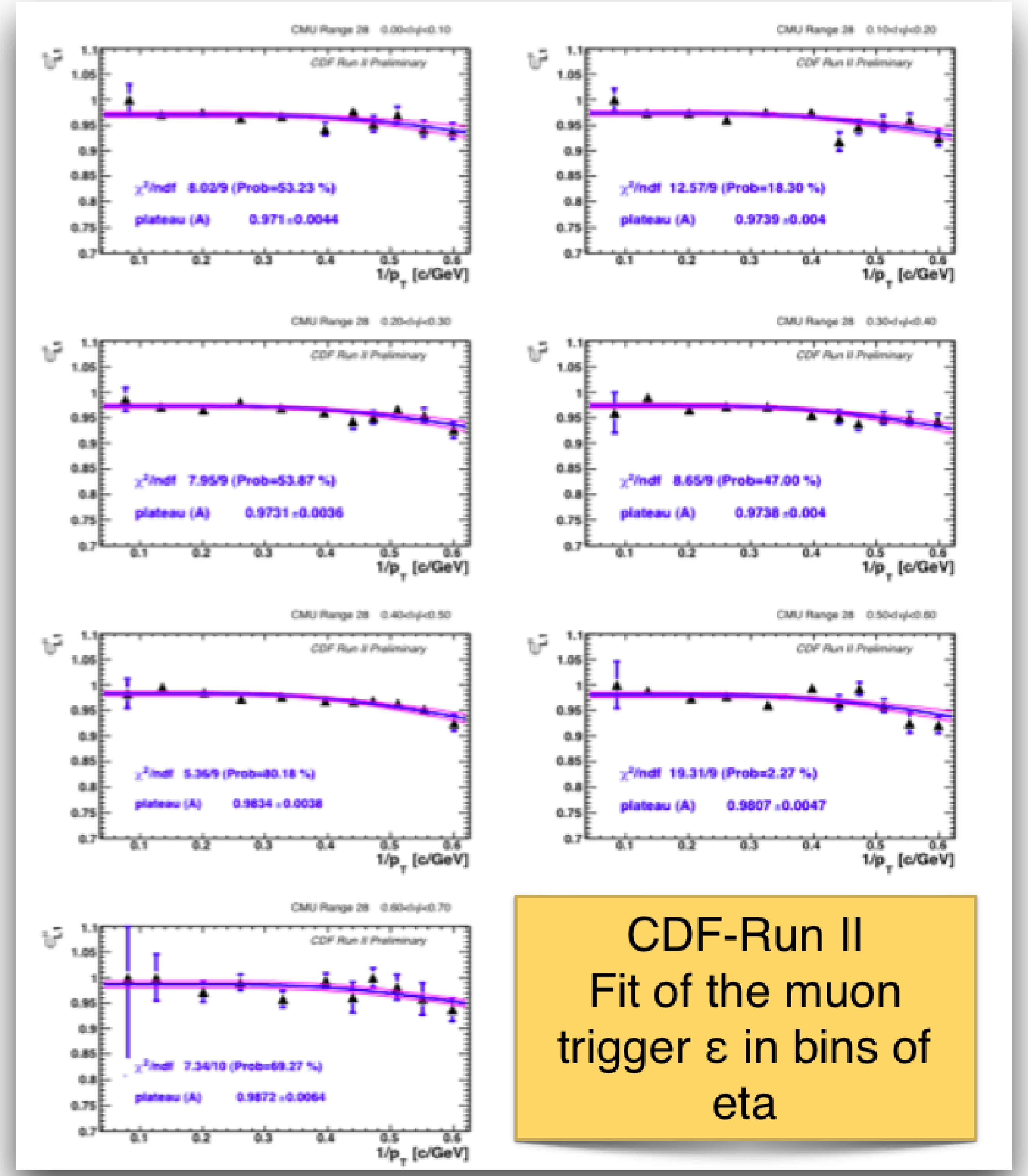
Trigger efficiency can vary rapidly due to changes in

- ➢ Detector geometry (η, φ)
- ➢ Trigger hardware (ageing, dead channels, etc…)
- ➢ Trigger definition
- ➢ Trigger algorithms

Analyses must track all of these changes

Multi-dimensional study of the efficiency:
$\varepsilon(p_T, \eta, \varphi, run\#)$

Fit the turn-on curves for different bins of η, φ, p_T



Science and
Technology
Facilities Council

Particle Physics

Monte Carlo and Scale Factors

Triggers have to be emulated in simulated data (Monte Carlo)

BUT… MC samples are produced **before** the data are recorded

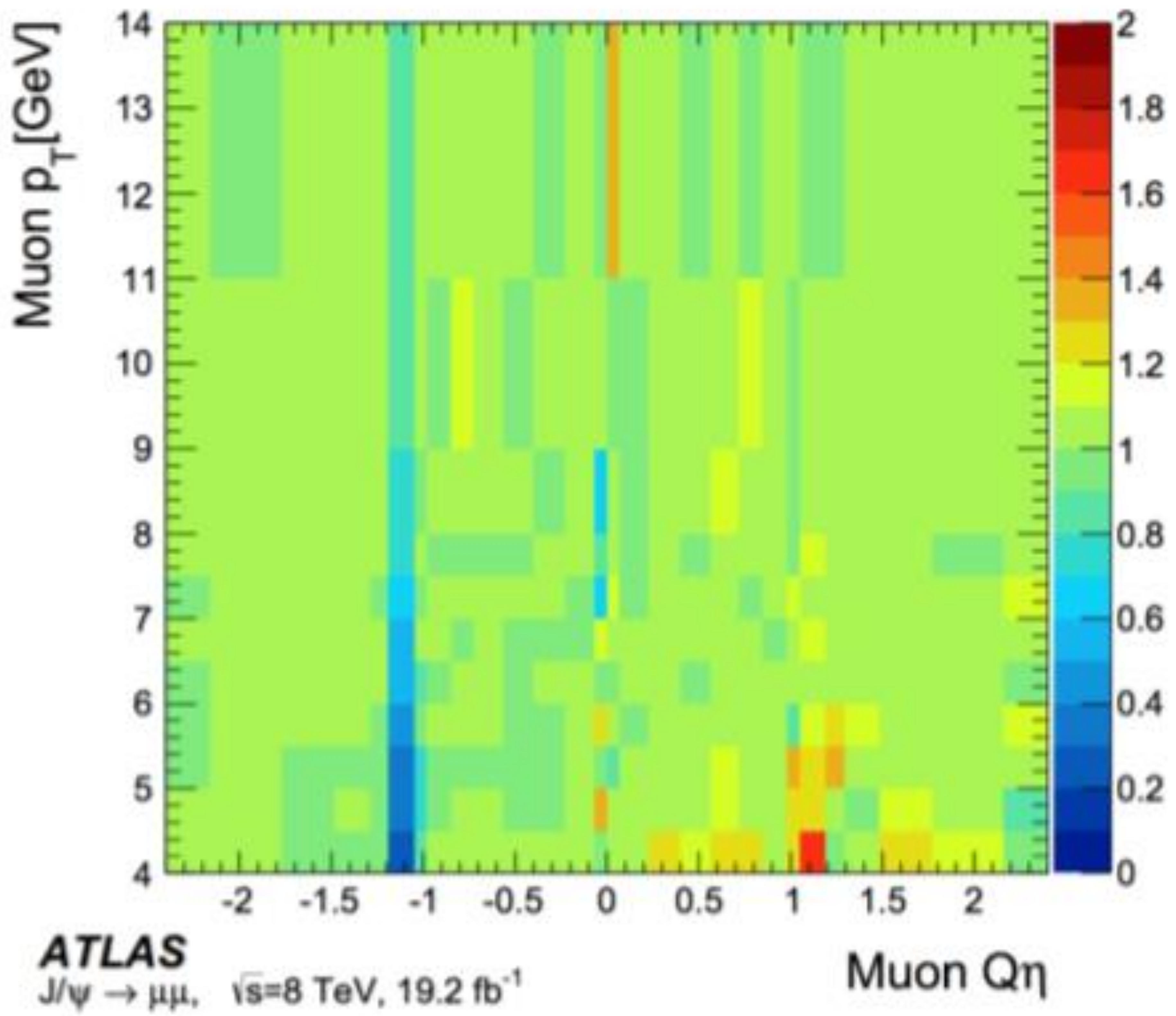
MC contains best-guess trigger menu (plus backups for possible future triggers). **Never** think of everything

Differences between data/MC occur due to:

- Different running conditions: pileup, luminosity profile
- Trigger menu changes
- Improvements/ bug fixes

Scale factors used:

- Correct MC to match observed data
- Parameterise in terms of p_T, η, ϕ, etc.



Science and Technology Facilities Council

Particle Physics

How to design a trigger

Understand your physics:
- What particles are in final state, how high is pT?

Understand the existing trigger menu
- Is there a trigger in place that will accept your events? → It's covered – job done!!

If not think up a new trigger:
- Can you combine particles, e.g. muon + 2 b-jets
- Can you use topology of event, .e.g. invariant mass, back-to-back topologies
- Remember trigger selection should be looser than offline

Keep it simple
- Less bias
- Less need for supporting triggers

Would other analyses profit from your trigger?
- More analyses there are the more likely your trigger will be accepted to run online

If possible, base it on an already existing trigger
- Already validated

Check the rate
- Will this new trigger fit into the trigger menu or do you need a prescale?

How to design a trigger

Simple

- Easy to commission, debug and understand

Inclusive

- One trigger for many analyses
- Discover the unexpected!

Robust

- Trigger runs millions of times per second – strange conditions will occur, be prepared
- Be immune to detector problems
- Be prepared for changes in beam conditions

Redundant

- Signal selected by more than one trigger
- Help understand biases and efficiencies
- Safety – backup in case of high rates or problems

Don't forget supporting triggers for efficiency measurement and background studies

Trigger Level Analysis (TLA)

Search analyses don't tend to like prescaled triggers:

- Immediate efficiency loss at trigger level
- Signal events could be lost

Prescales used to keep rates under control

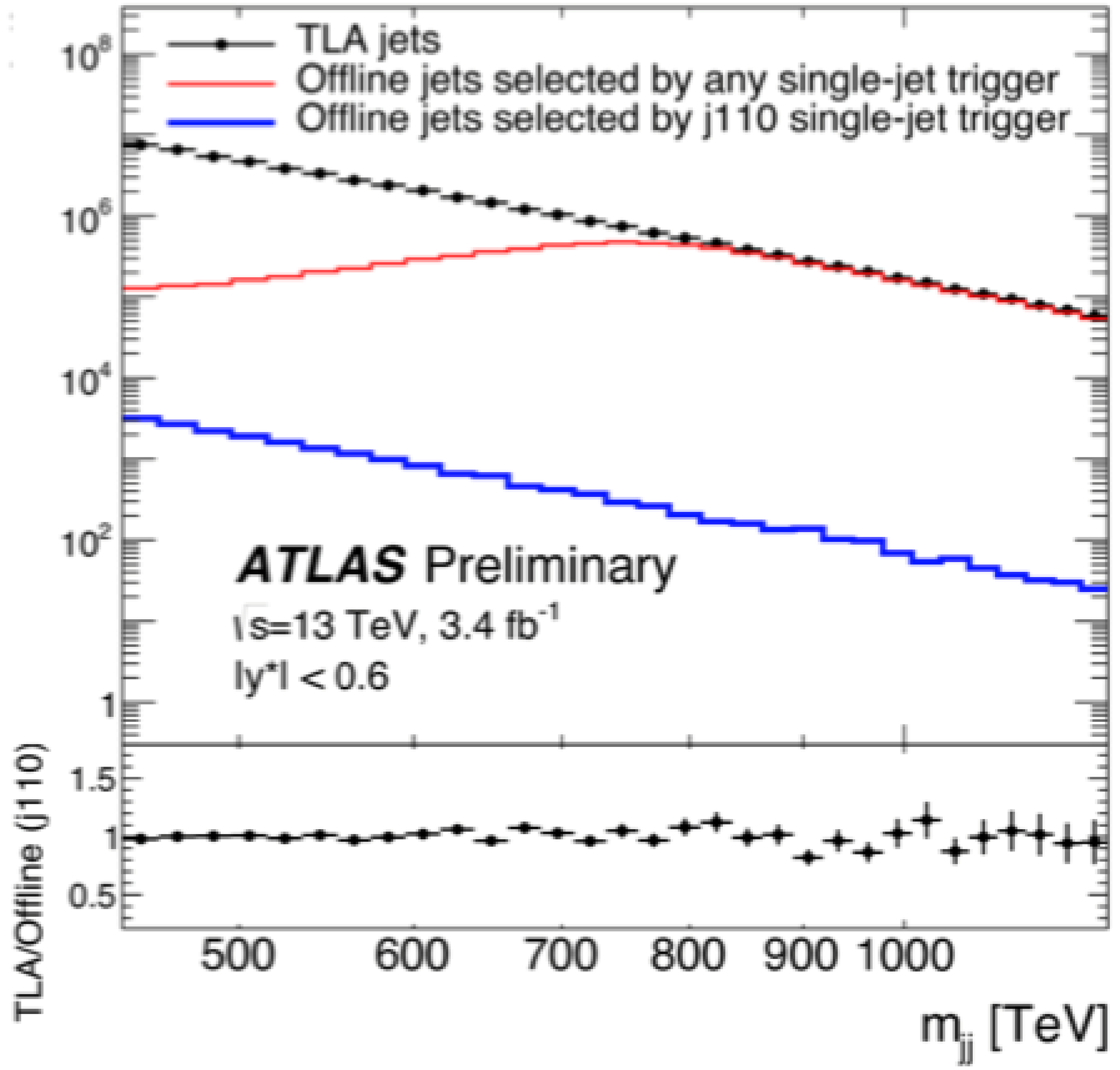
Have another dial to turn: event size

Reduce the size of your event by only saving objects you need

- Run unprescaled again!!

TLA jets:

- Only save the leading few HLT trigger jets with selective variables
- Form di-jet invariant mass and push below threshold allowed for normal jet triggers



Particle Physics

Science and Technology Facilities Council

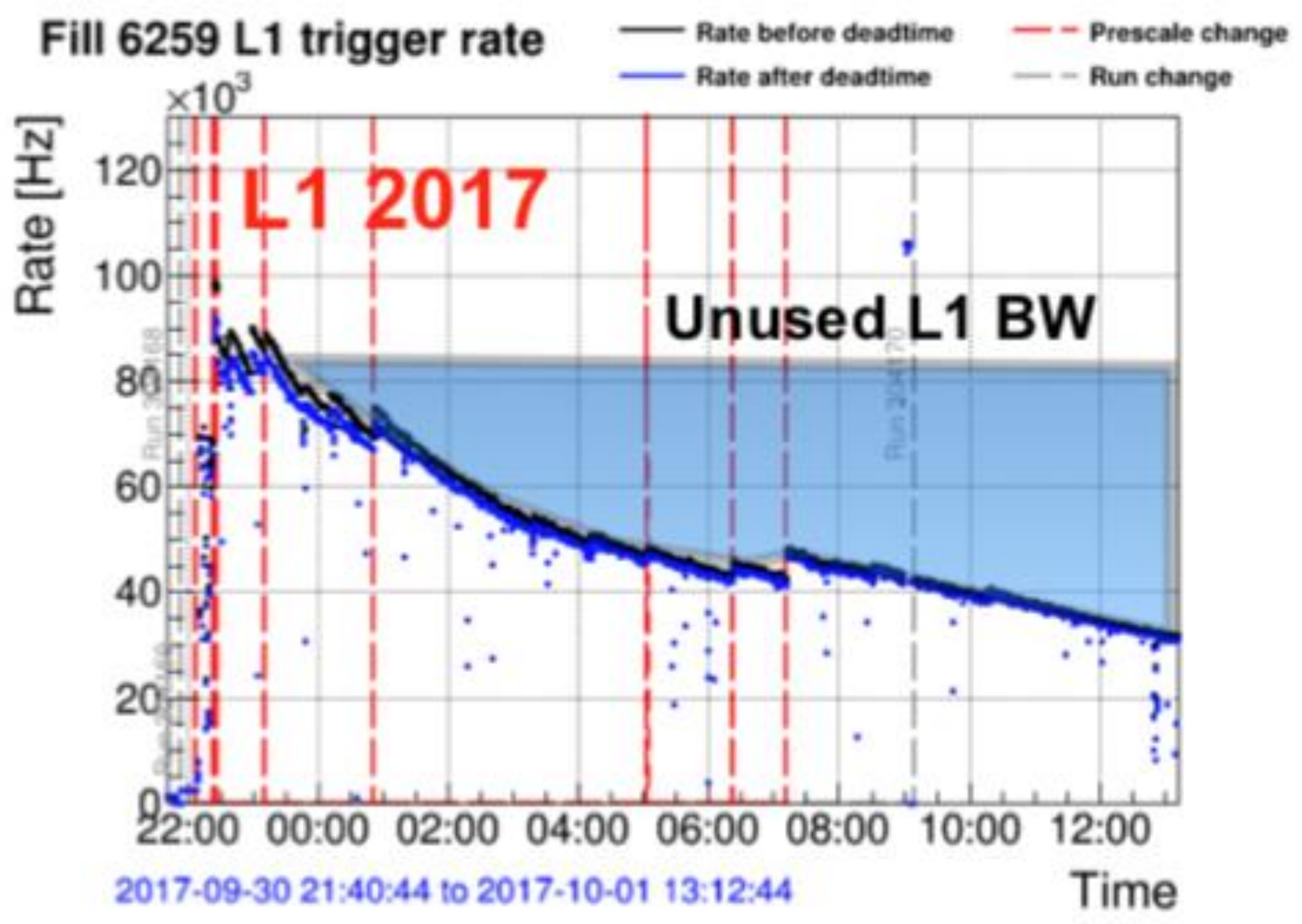
Parked / delayed data streams

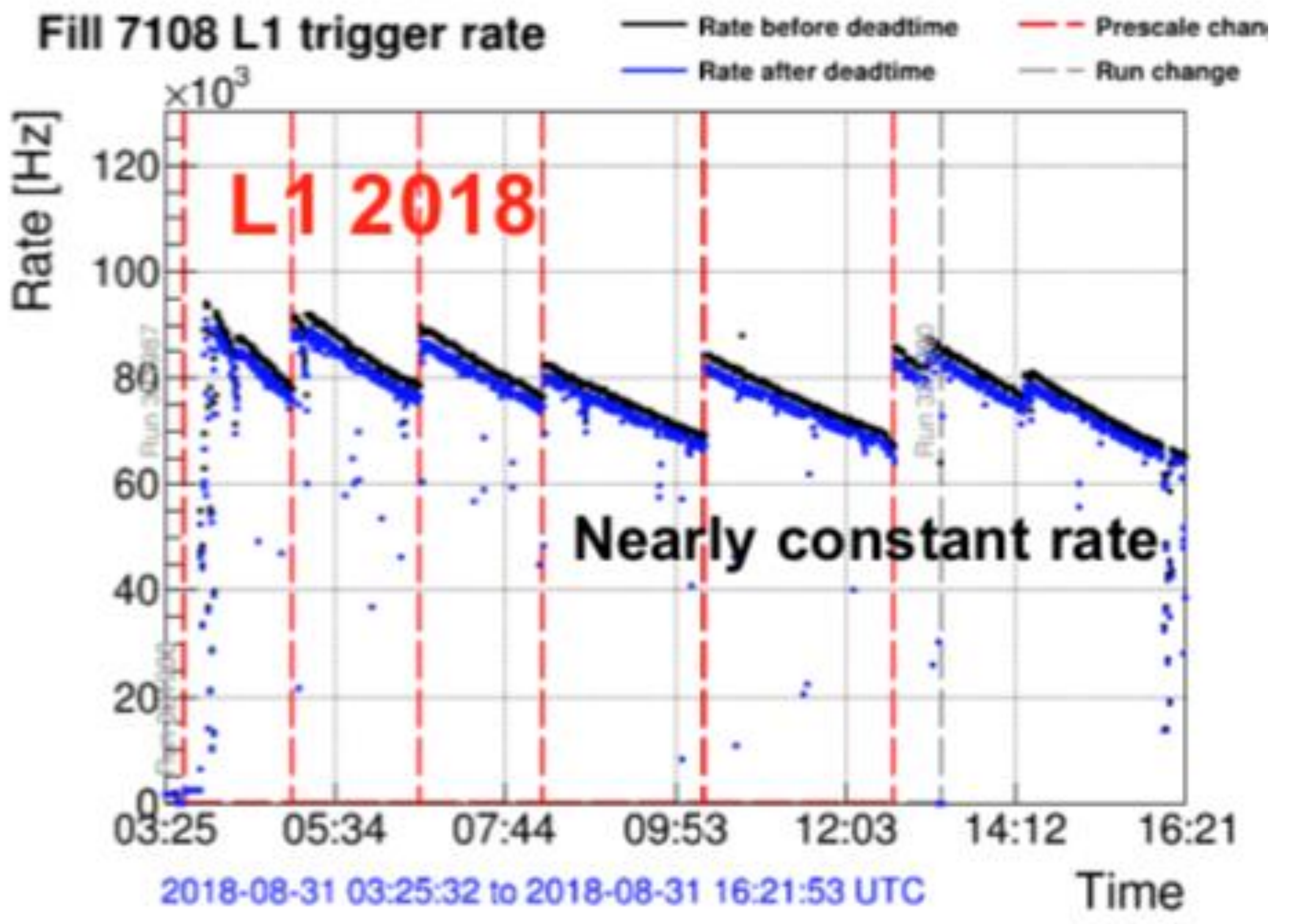
What if we don't mind waiting for the data?

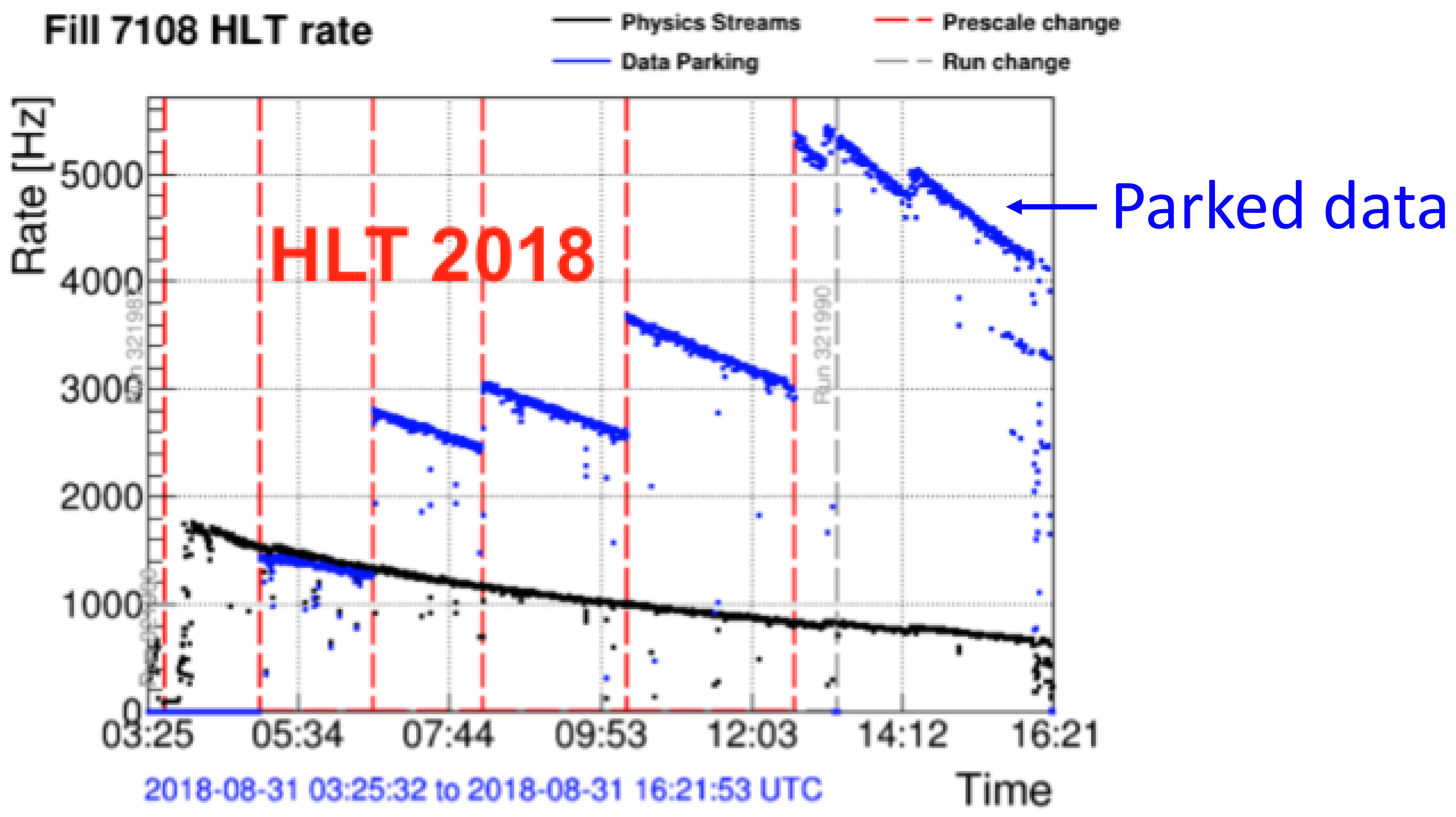
B-physics analyses at CMS/ATLAS

- Often take a while for analyses to be completed (personpower, complicated analysis)
- Can afford to wait O(6 months) after da taking for the data
- During the run "park" the data – reconstruct when computer power is available (between fills or end of year)

CMS recorded 10^{10} unbiased B hadron decays in 2018







Science and Technology Facilities Council

Particle Physics

Thank you

Search: STFC Particle Physics

@PPD_STFC

@STFC_Matters

Science and Technology Facilities Council