TDAQ mini-workshop summary

Becky Chislett

23rd September 2020

Introduction

We had a TDAQ mini workshop on 14th September providing an opportunity to hear about various different TDAQ systems and potential options for Mu2e-II

- Mu2e-II DAQ thoughts : Ryan Rivera (FNAL)
- The ATLAS TDAQ system : Catrin Bernius (SLAC)
- The ATLAS FTK experience : Lauran Tompkins (Stanford)
- The CMS trigger system : Isobel Ojalvo (Princeton)
- GPU trigger systems : Gianluca Lamanna (Pisa and INFN)

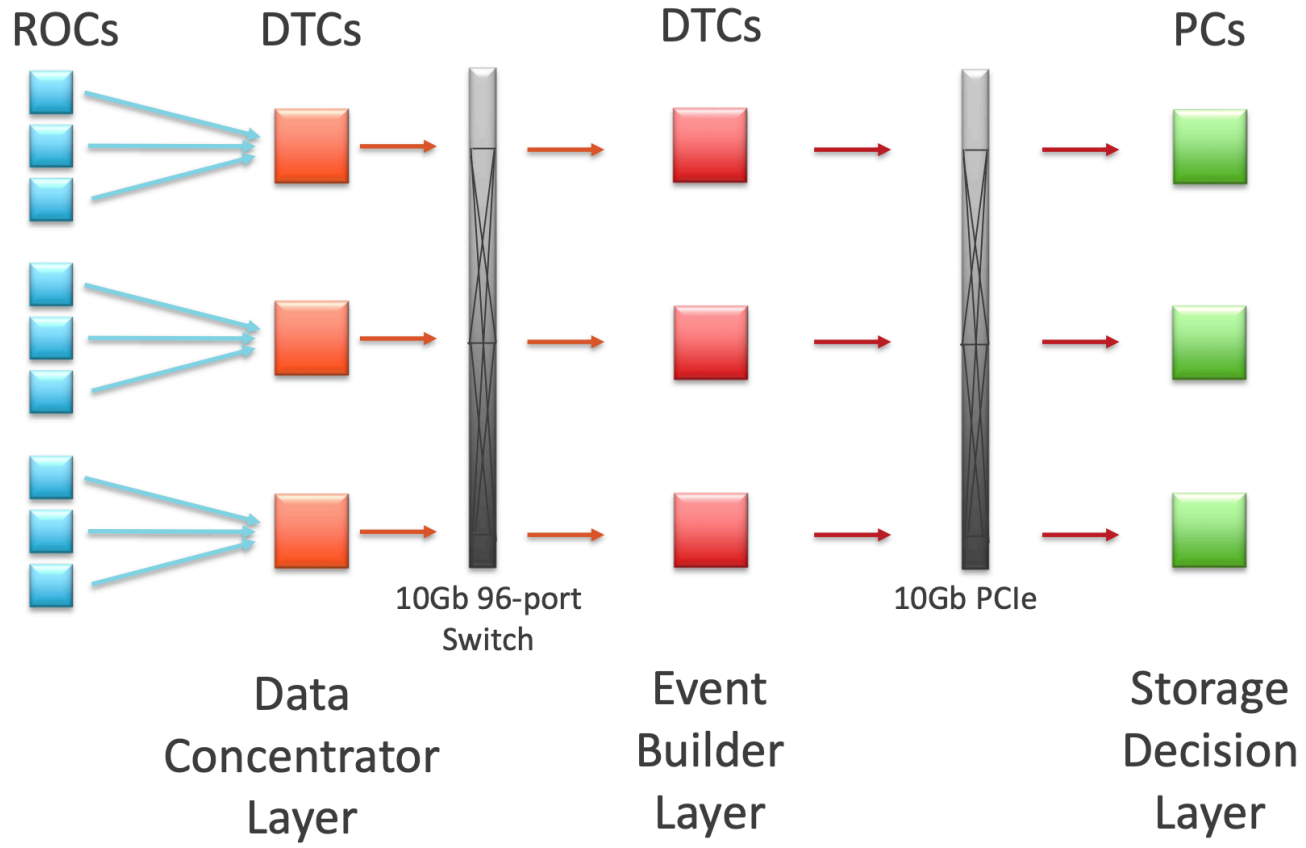
Mu2e-II TDAQ – Ryan Rivera

Introduced the requirements on the TDAQ system for Mu2e-II with some thoughts about possible routes forward

Expectations on data rates :

- ~2x detector channels, ~5x pulses on target
  ~10x data rate (40GBps for Mu2e)
- More detector channels and higher background
  ~3x event size (200KB for Mu2e)
- Assuming 2x tape storage (14PB/year) need to reduce 600KB @ 3MHz to 560MB/s
  ~3000:1 rejection

Mu2e data readout path

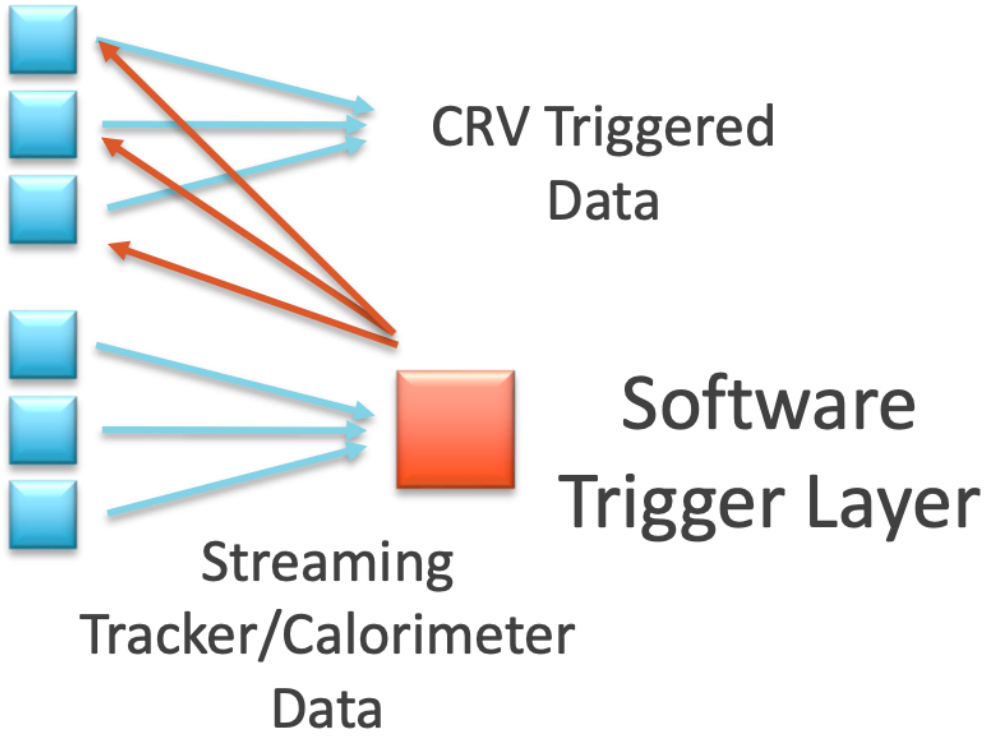


Mu2e-II TDAQ – Ryan Rivera

Other considerations :

- **Reduced or no off spill time** to readout large front end buffers (Mu2e has a second to catch up)
- **No large buffers for the CRV** leads to two options :
  - Large CRV buffers + software trigger
  - Small CRV buffers + hardware trigger
- **Streaming vs triggered data taking** :
  - Same as Mu2e : stream tracker + calo, software trigger for CRV
  - Or stream calo data + hardware trigger for tracker and CRV based on calo, software trigger for storage choice

Mu2e trigger path



Mu2e-II TDAQ – Ryan Rivera

Need to start thinking about the choice and requirements of the FPGA

- **Radiation tolerance requirements**
  - Higher for Mu2e-II : ~ calo level for CMS phase II?
  - Probably don't want to design our own rad hard links
- **Use of High Level Synthesis (HLS) :**
  - Good enough to rival VHDL/Verilog development
  - Allows physicists to develop code
- **FPGA development can start now to inform board choice / DAQ topology :**
  - Which subsystems are streaming?
  - Is a low latency level 1 hardware trigger possible? Could save money due to reduced data/buffer sizes
  - How much processing is needed for the high level trigger?

Mu2e-I DTC →

	KINTEX-7	KINTEX UltraSCALE	VIRTEX-7	VIRTEX UltraSCALE
Logic Cells (LC)	478	1,161	1,995	4,407
Block RAM (BRAM) (Mbits)	34	76	68	132
DSP-48	1,920	5,520	3,600	2,880
Peak DSP Performance (GMACs)	2,845	8,180	5,335	4,268
Transceiver Count	32	64	96	104
Peak Transceiver Line Rate (Gb/s)	12.5	16.3	28.05	30.5
Peak Transceiver Bandwidth (Gb/s)	800	2,086	2,784	5,886
PCI Express Blocks	1	6	4	6
Memory Interface Performance (Mb/s)	1,866	2,400	1,866	2,400
I/O Pins	500	832	1,200	1,456

Mu2e-II TDAQ – Ryan Rivera

TDAQ LOIs submitted for Snowmass 2021 :

1. **A 2-level TDAQ system based on FPGA pre-processing and trigger primitives**
   - ROCs (create trigger primitives, buffer event fragments)
   - L1 FPGA layer (getting trigger primitives from calo and tracker)
   - HLT layer (requests event fragments from full detector)

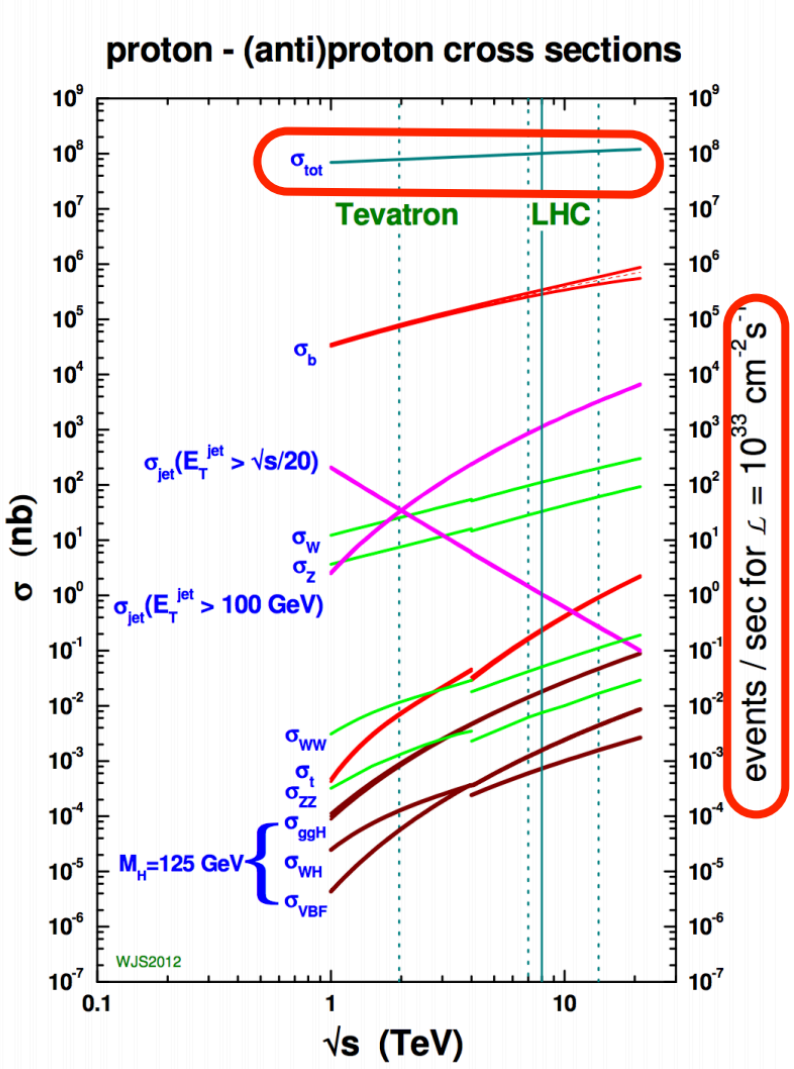
2. **A 2-level TDAQ system based on FPGA pre-filtering**
   - Leverage HLS for FPGA rejection

3. **TDAQ based on GPU co-processor**
   - Using GPUs at HLT (or L0)

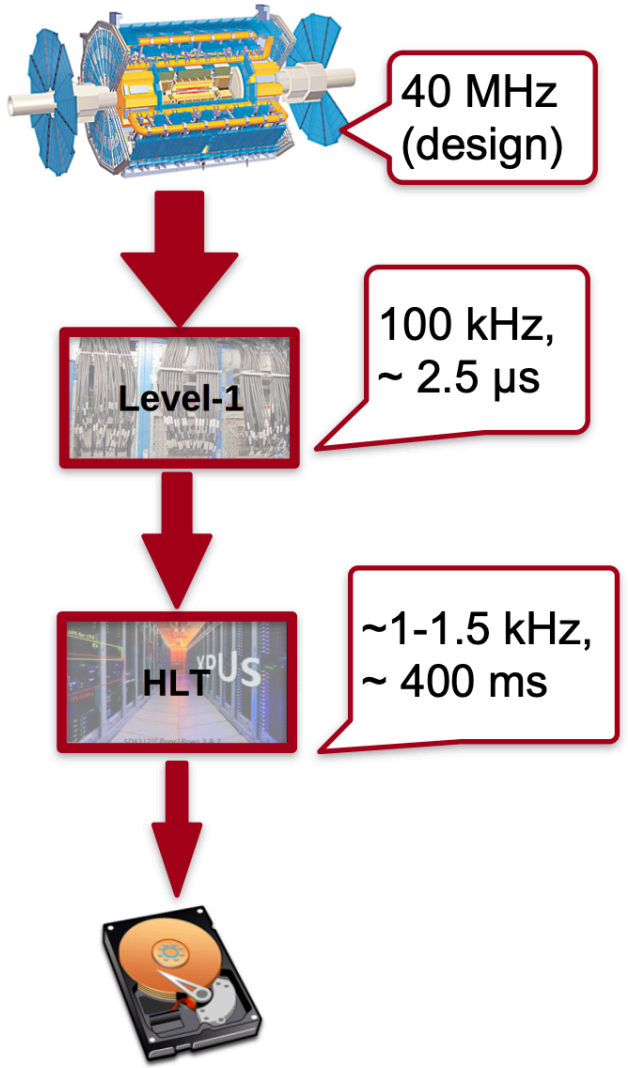
4. **A trigger-less TDAQ system based on software trigger**
   - Scale up current system.

ATLAS TDAQ – Catrin Bernius

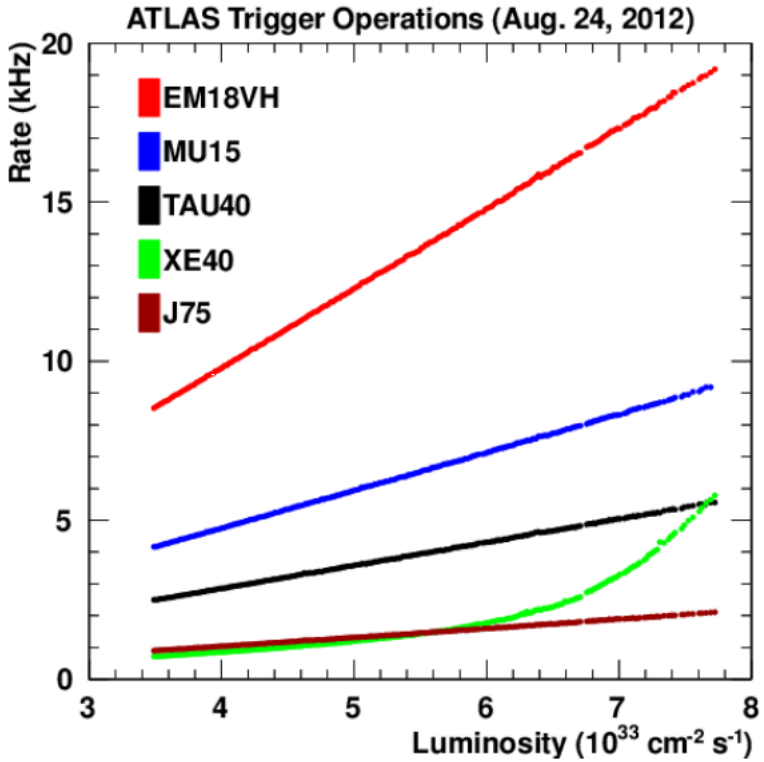
Triggering is essential in ATLAS



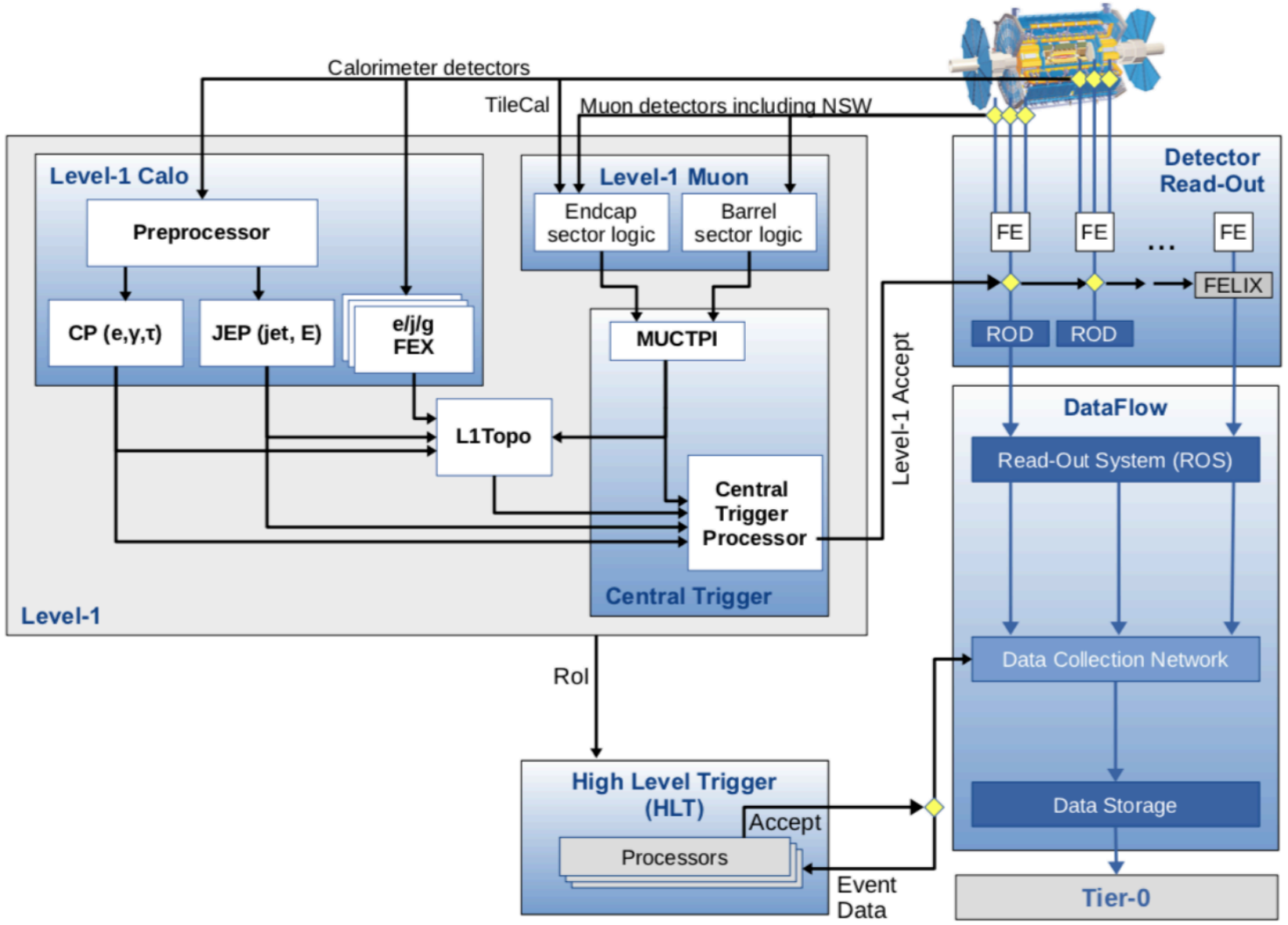
The ATLAS trigger system



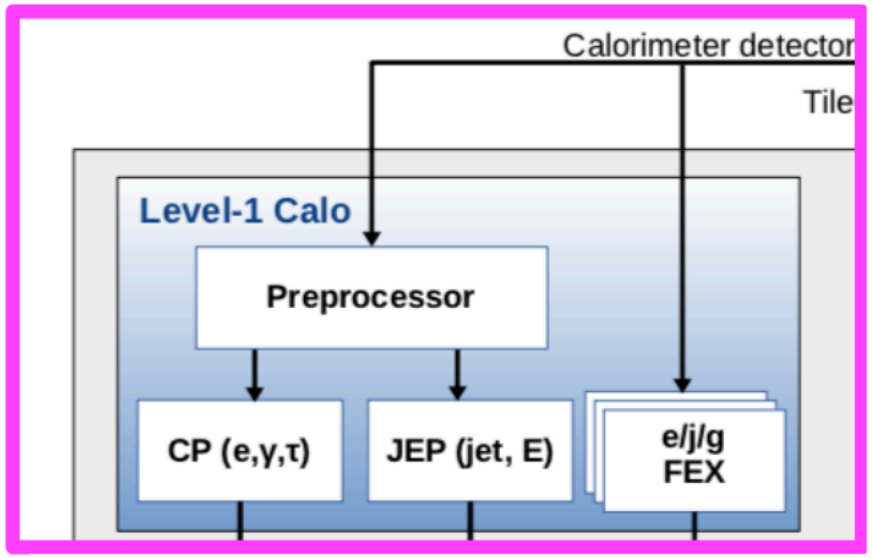
The trigger system has had to evolve with increased performance at the LHC



ATLAS TDAQ – Catrin Bernius



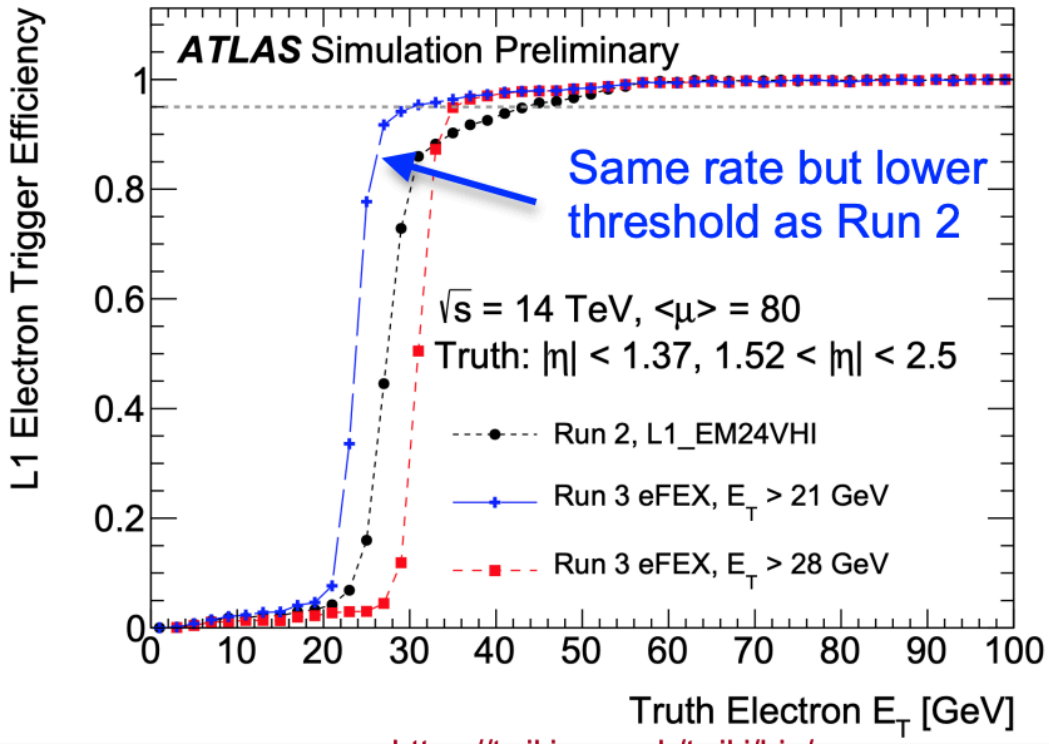
ATLAS TDAQ – Catrin Bernius



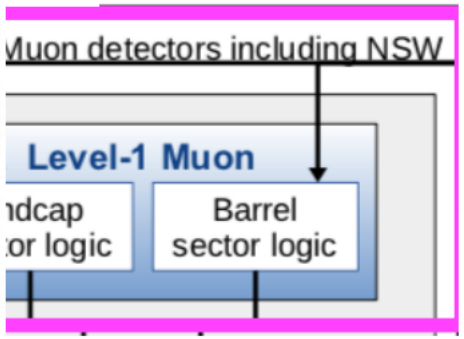
L1 calo trigger identifies electrons, photons, jets, hadronically decaying taus and global event quantities

Upgrades for run 3 :

- Better granularity Lar calo inputs to improve resolution
- ATCA-based Feature Extractors (FEX) for more sophisticated algorithms – reduced rates while keeping low thresholds

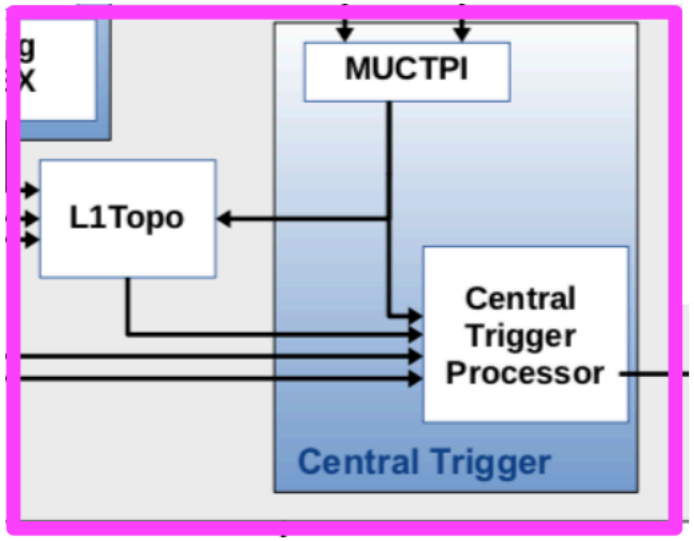


ATLAS TDAQ – Catrin Bernius

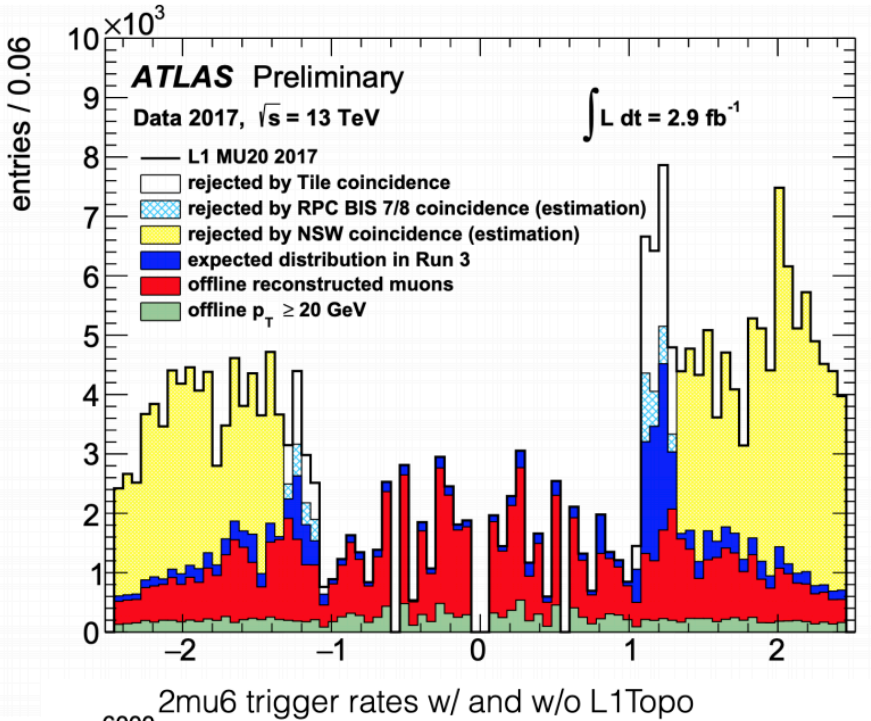


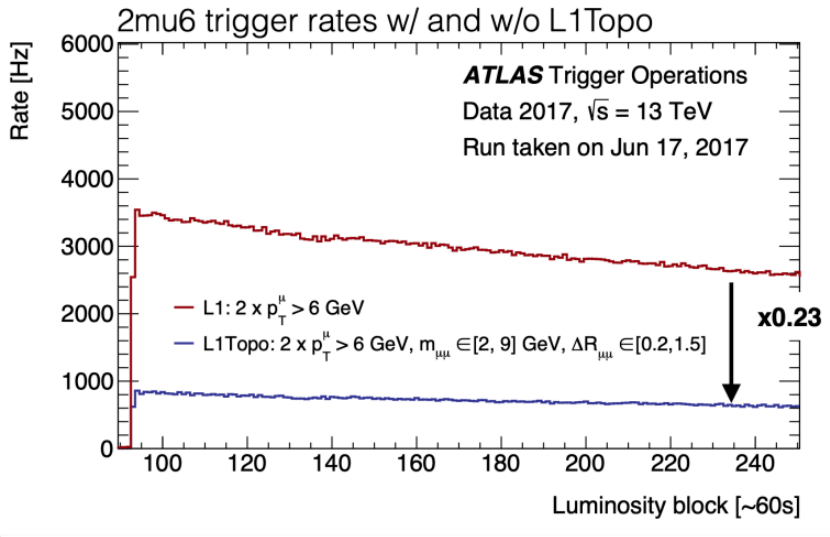
The muon trigger looks for hits in the barrel and endcap muon detectors using coincidences to reduce the rates

In Run-3 a New Small Wheel (NSW) encap should reduce the rates further without a loss in efficiency

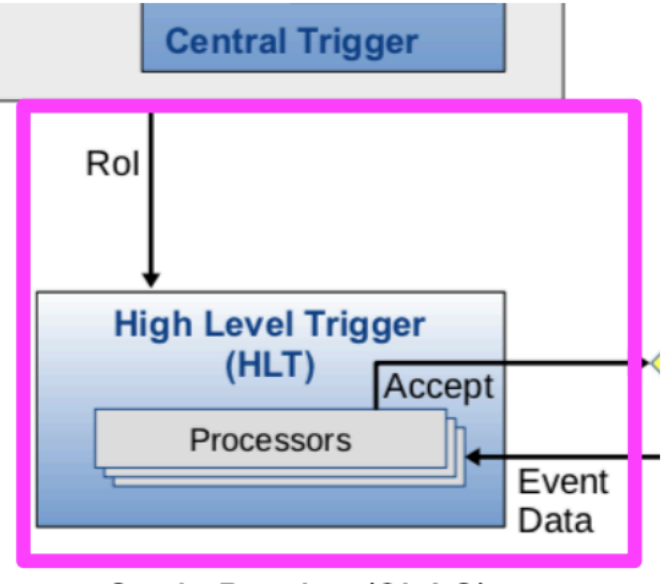


Combining information from the muon and calo triggers with other kinematic selections further reduces the rate



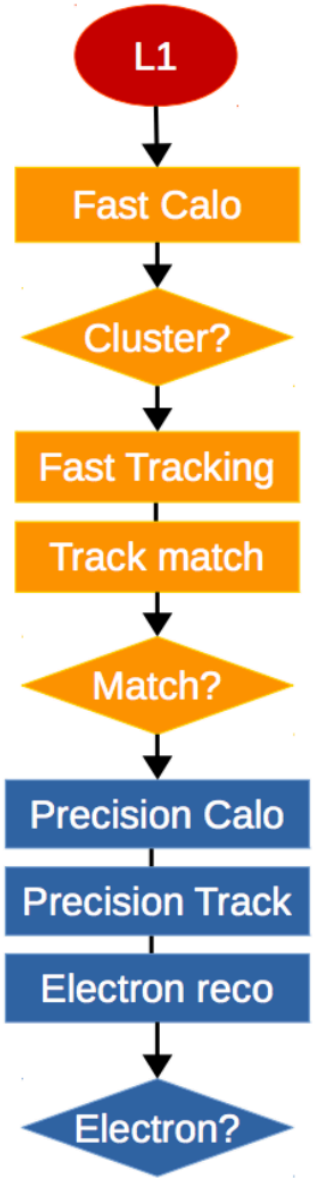


ATLAS TDAQ – Catrin Bernius



The high level trigger uses information from L1 :

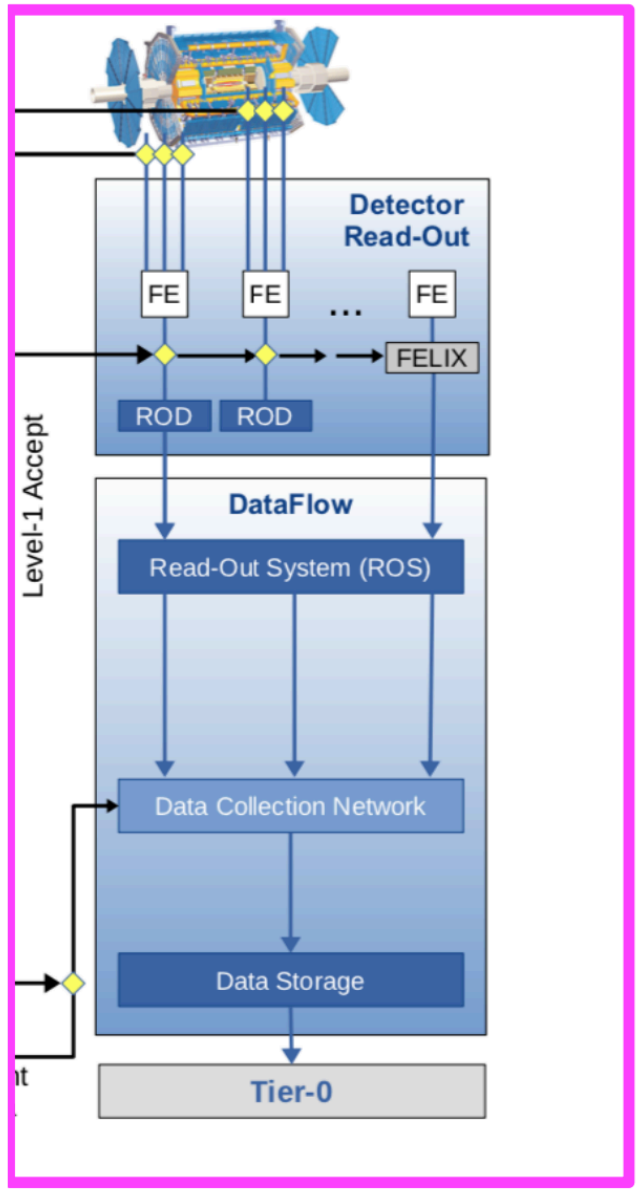
- Fast reconstruction based on RoIs
- Precision reconstruction using full detector
- Rejection is any step fails



The data flow is :

- L1 accept sends data to the Read Out Drivers (RODs)
- The data is sent via fibre to the Read Out System
- The data collection network handles communication to the HLT
- Accepted events are sent to storage

The trigger menu determines what data is recorded with appropriate prescales determined by the physics



ATLAS TDAQ – Catrin Bernius

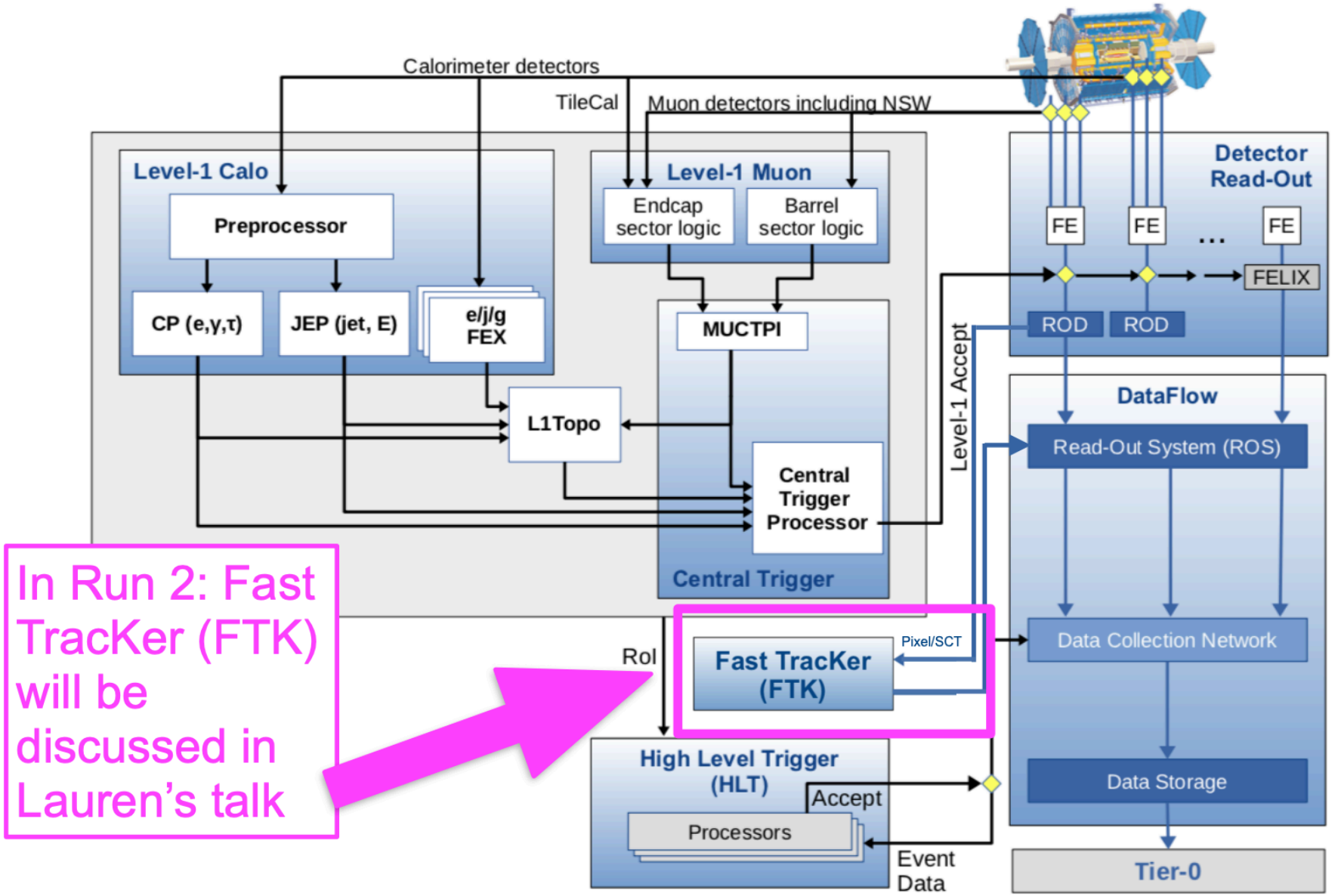
Experience from Run 1 and Run 2 has shown that the ATLAS TDAQ system is able to efficiently record data while dealing with various constraints and challenging conditions

- **Evolution of the TDAQ system** in terms of hardware and software important to maintain physics acceptance and efficiency

- **Improvements for Run 3** are focusing on new L1 hardware and improved HLT algorithms
  - Upgraded L1Calo and L1Muon system
  - Moving closer to offline reconstruction through AthenaMT

- **Versatile trigger menu** to record data for a wide range of physics analyses
  - Aim at exploiting the total bandwidth for physics even better and to extend the phase space for physics discovery

ATLAS FTK – Lauren Tompkins

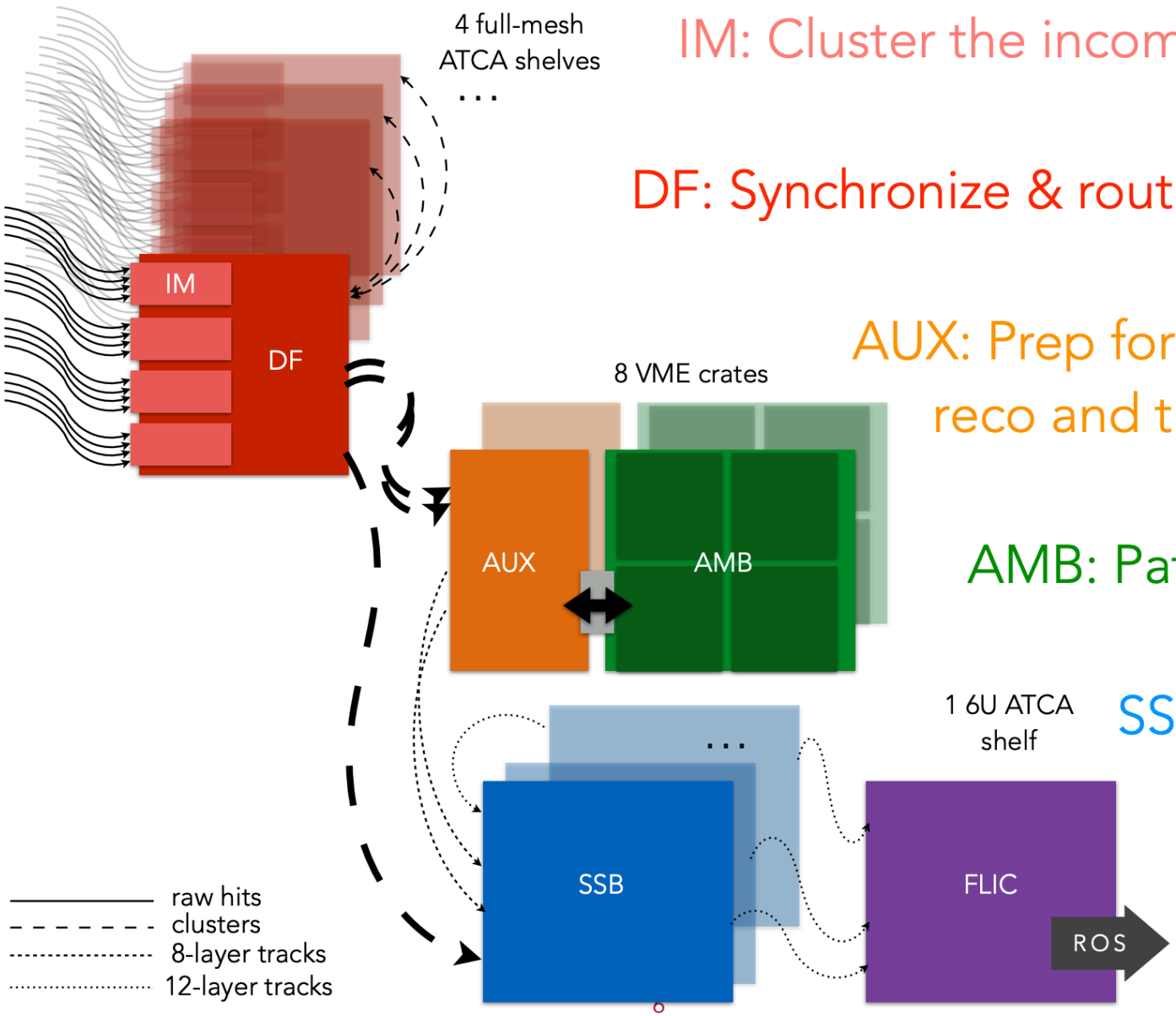


In Run 2: Fast TracKer (FTK) will be discussed in Lauren's talk

The Fast Tracker (FTK) is designed to run in parallel with the HLT and provide full silicon tracking for each event by :

- Parallelisation
- Reduction in data volume (clusters become coarse hits)
- Use pre stored patterns
- Simplified algorithms
- Hardware (FPGAs/ASICs)

ATLAS FTK – Lauren Tompkins



IM: Cluster the incoming data (128)

DF: Synchronize & route to towers (32)

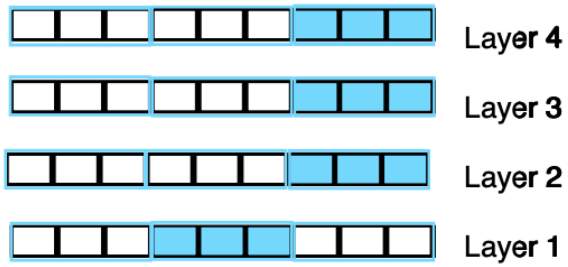
AUX: Prep for 8-layer pattern reco and track fit (128)

AMB: Pattern reco (128)

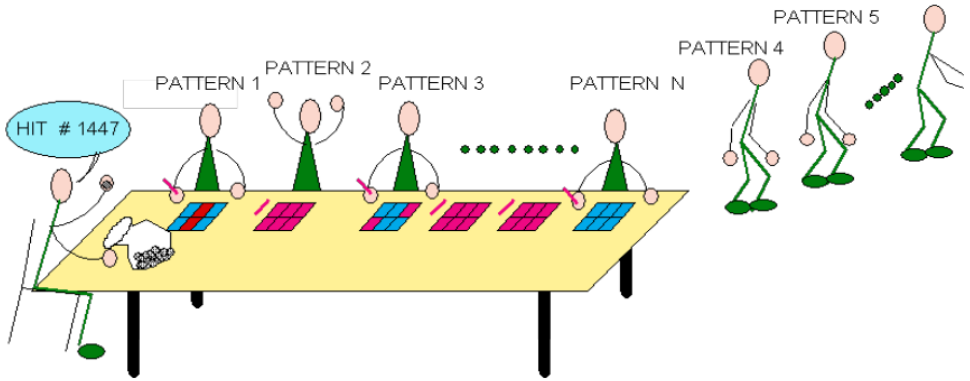
SSB: 12-layer fit (32)

FLIC: Format for HLT (2)

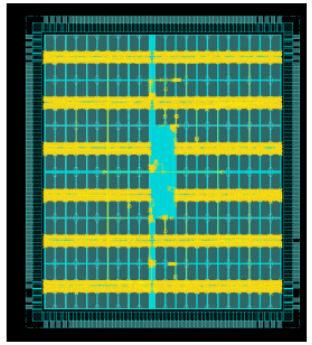
ATLAS FTK – Lauren Tompkins



- Hits are ganged together into coarse resolution hits



- All possible patterns of coarse resolution hits determined from simulation

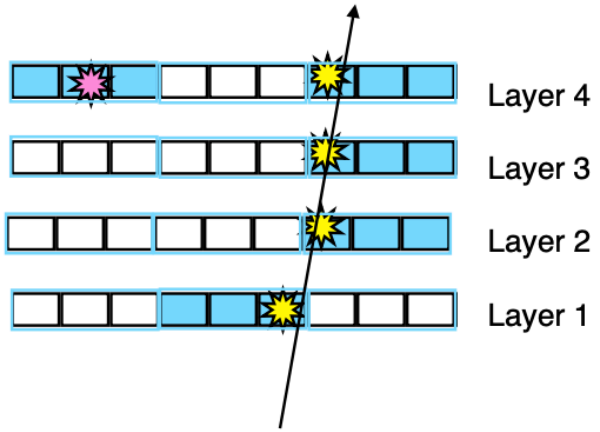


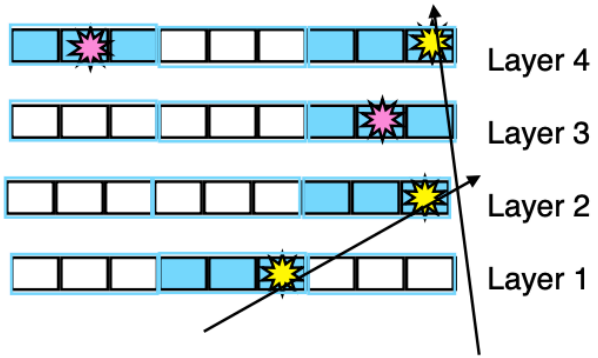
AMChip06

- Custom associative memory chips are used to **compare hits** to O(10^9) patterns **simultaneously** (bingo cards)

ATLAS FTK – Lauren Tompkins

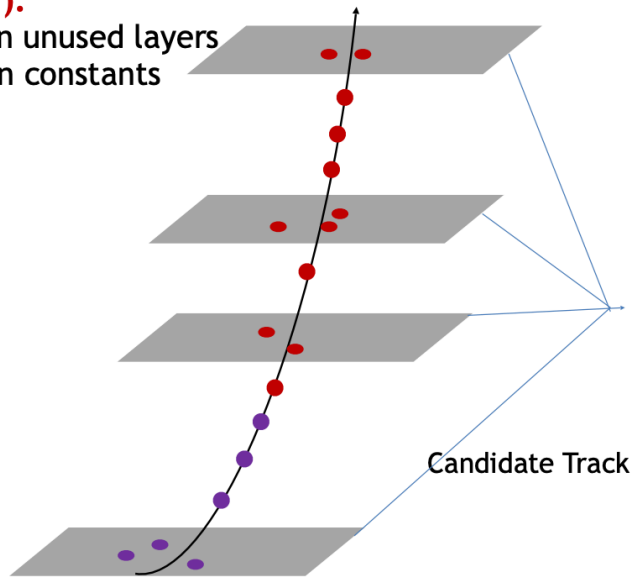
But most matched patterns are just from a random selection of hits…



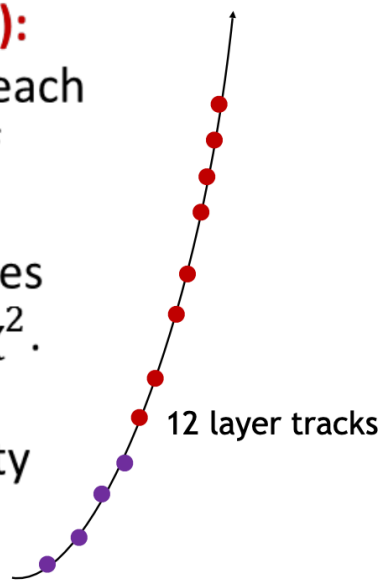


Need final track production

Extrapolator (EXT):
Finds candidates in unused layers using extrapolation constants



Track Fitter (TF):
Fits track with each combination of possible candidates, takes one with best χ^2. (Nominal and Pix/SCT Majority Fitters)



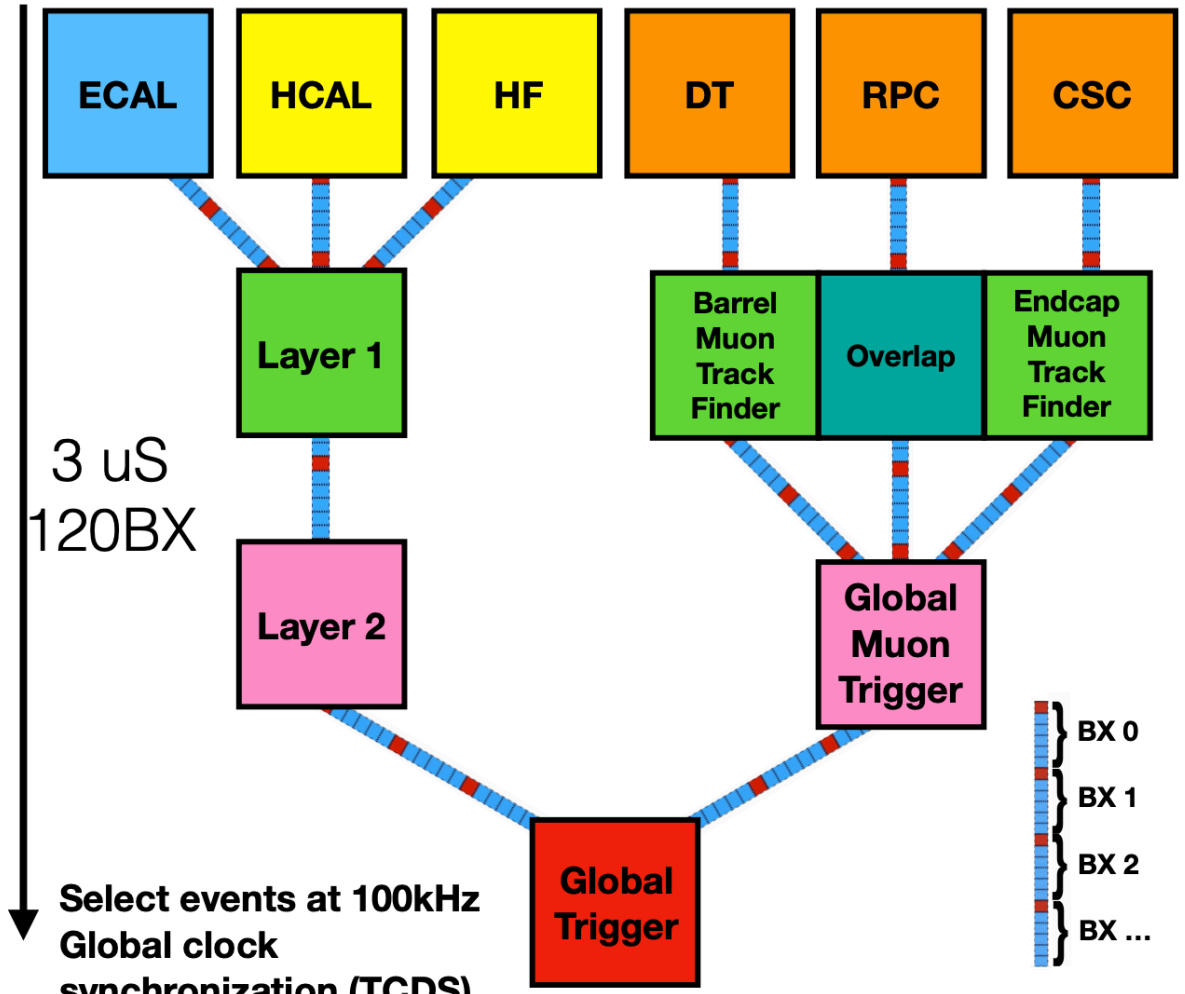
ATLAS FTK – Lauren Tompkins

UCL

- Heterogeneous systems are difficult to integrate and debug:
  - Mixed ATCA/VME ; mixed Xilinx/Altera — good reasons at the time but made life more challenging
- Resource usage estimates are hard without firmware in hand — nearly every FPGA was full
- Data push architecture with no external synchronization source is difficult in real (buggy) conditions
- Need significant engineering involvement of firmware writing
  - HLS was not mature when much of FTK development was happening (FW development started in ~2012)
- Early emphasis on interfaces/unit tests/CI&CD/simulation test benches would've streamlined integration and development
- Monitoring FW and online software is critical for debugging in situ
- Dedicated person-power is critical!

CMS TDAQ - Isobel Ojalvo

Digitization of Detector Signals
+
Trigger Primitive Generation
↓
Regional Triggers
↓
Sub-detector Trigger
↓
Global Trigger



L1 trigger:

- Reduces 1 GHz to 100 KHz
- Each event is held for ~120 bunch crossings while the decision is made

HLT :

- Reduces 100 kHz to 100s Hz
- Similar to ATLAS – processing stops when an event fails

Pipeline System
Frequency 40MHz
Feed-Forward Algorithms
- (no backwards loops)
Highly Distributed
Full Event Processed at the GT

Designed on custom built electronics employing high speed links (I/O) and (ASICs +) FPGAs

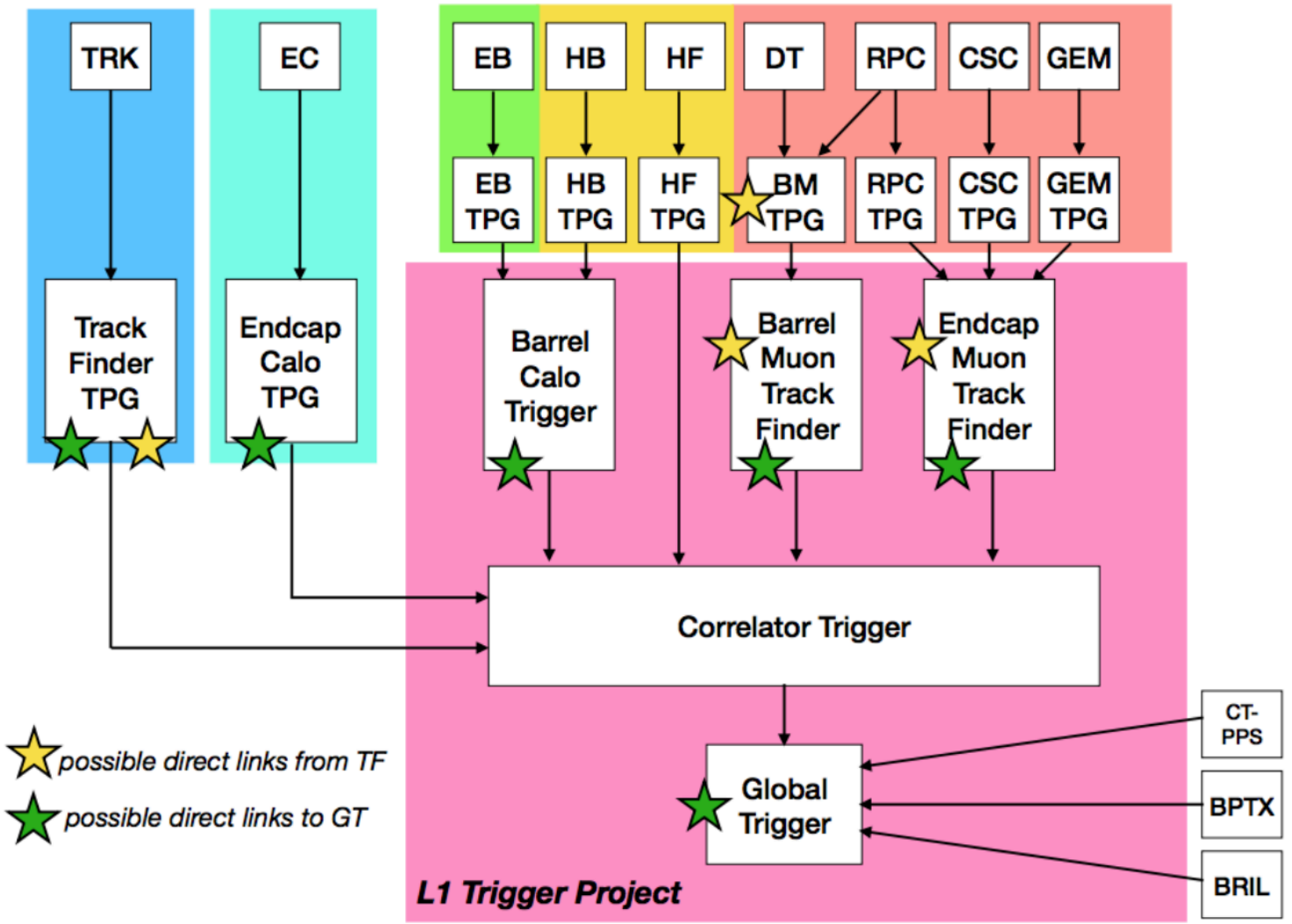
CMS TDAQ - Isobel Ojalvo

The system is being upgraded for HL-LHC

- Higher granularity calo inputs
- New track trigger
- Trigger within 12.5us with max rate 750 kHz

5 uS
200BX

2.5 uS
100BX



CMS TDAQ - Isobel Ojalvo

The Phase II upgrade plans to bring offline reconstruction to L1 and increase flexibility using

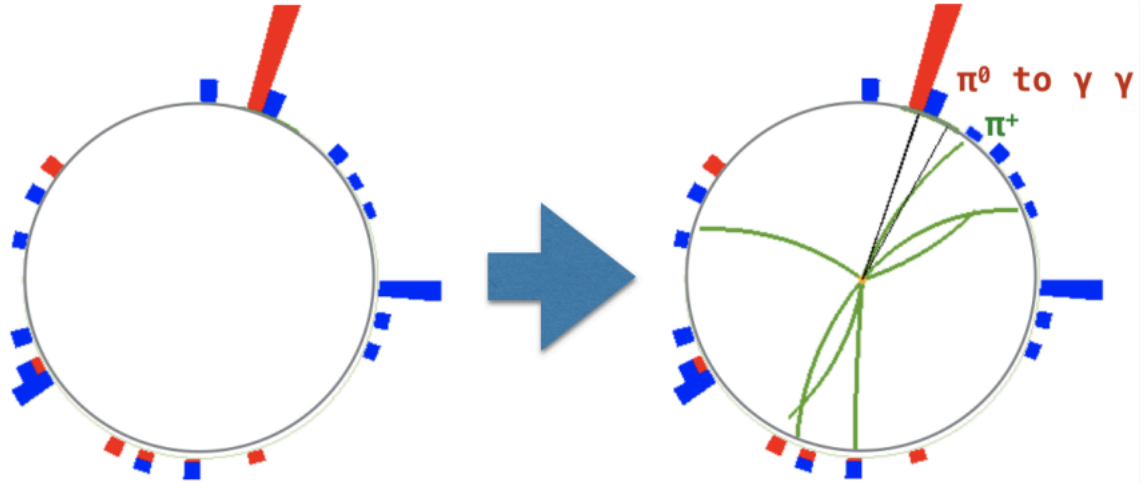
- Large multi-purpose Ultrascale+ class FPGAs
- Multi-Gigabit transceiver links (16-28 Gb/s)

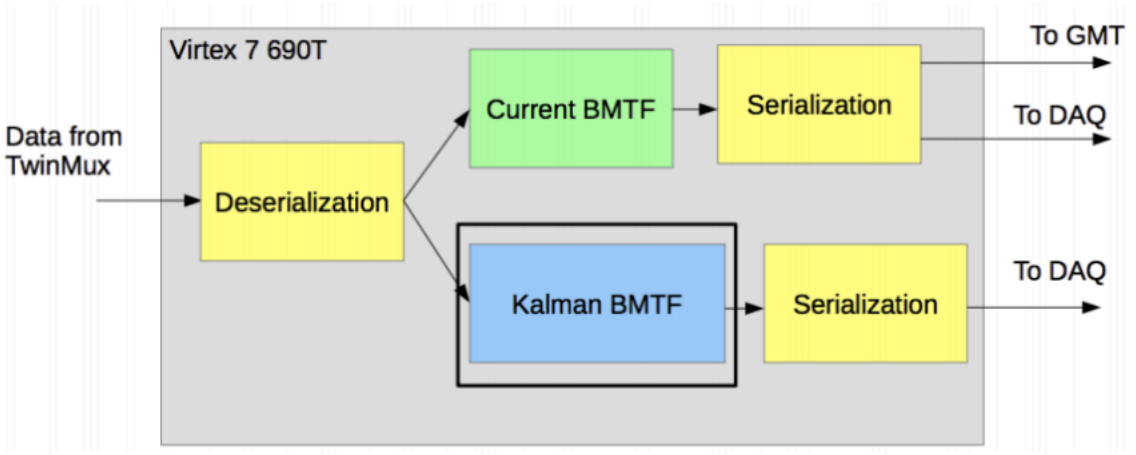
A new track trigger is being implemented :

- Create stubs by matching layers in the outer tracker
- P_T measurement by matching to inner layers

Big focus on using high level synthesis for algorithm development :

- Kalman filter muon track finder – uses DSP cores to reduce FPGA resources due to the large amount of matrix maths
- NN for muon p_T assignment using the HLS4ML tool kit





CMS TDAQ - Isobel Ojalvo

UCL

Level 1 Trigger (Run-1, Run-2 and Run-3)

- Select 100 kHz interactions from 1 GHz
- Processing is synchronous & pipelined
- Decision latency is 3 μs
- Algorithms run on local, coarse data (Calo + Muon)
- Hardware is modular and "generic" possibility to modify algorithms

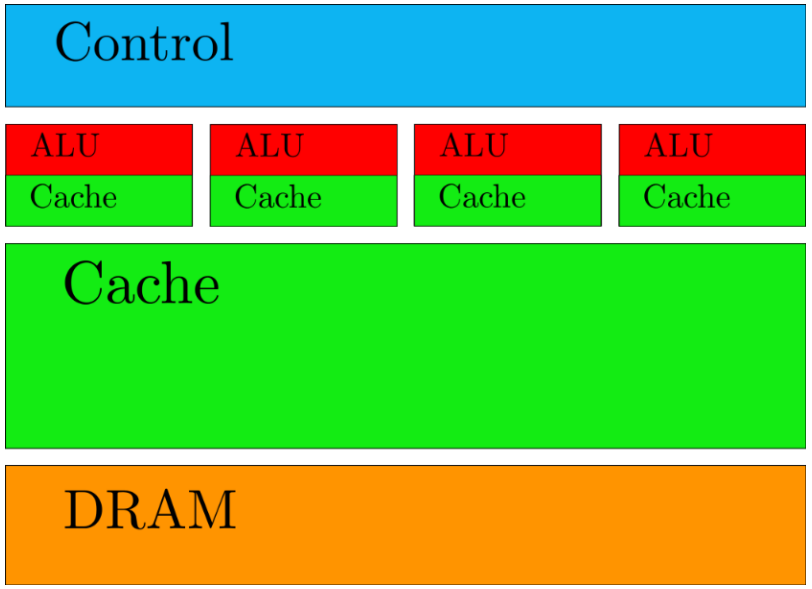
Higher Level Triggers

- Depending on experiment, done in one or two steps
  - If two steps, first is hardware region of interest
- Then run software/algorithms as close to offline as possible on dedicated farm of PCs

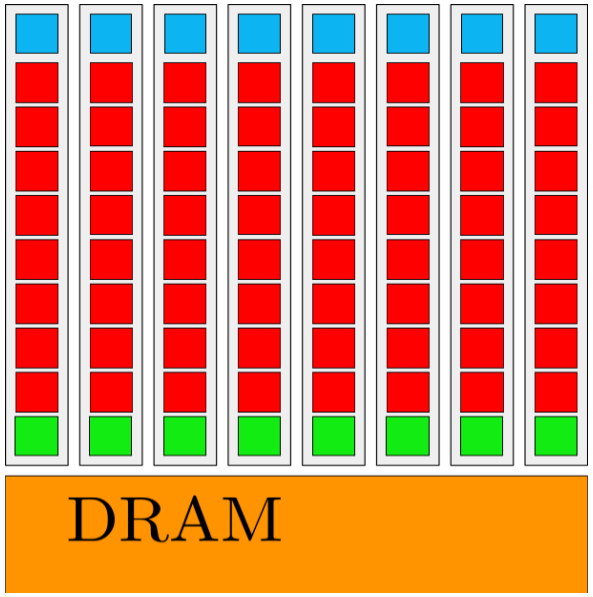
Phase-2 Level 1 Trigger

- Bring tracking to L1T (new triggerable tracker is the key)
- Migrate algorithms (not constraints) from HLT L1
  - Keep the processing parallel
- Keep the pipeline, increased latency for track building

GPU trigger systems – Gianluca Lamanna



- Large main memory
- Fast clock rate
- Large caches
- Branch prediction
- Powerful ALU
- Relatively low memory bandwidth
- Cache misses costly
- Low performance per watt

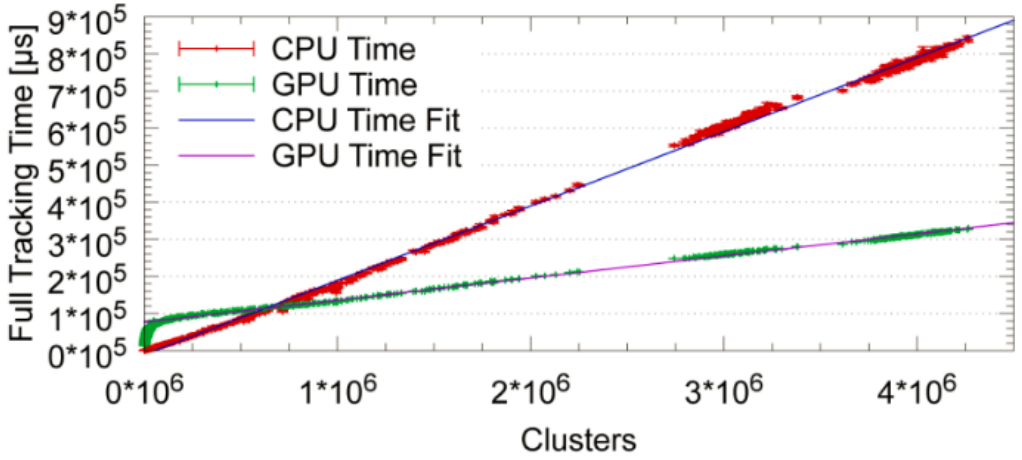


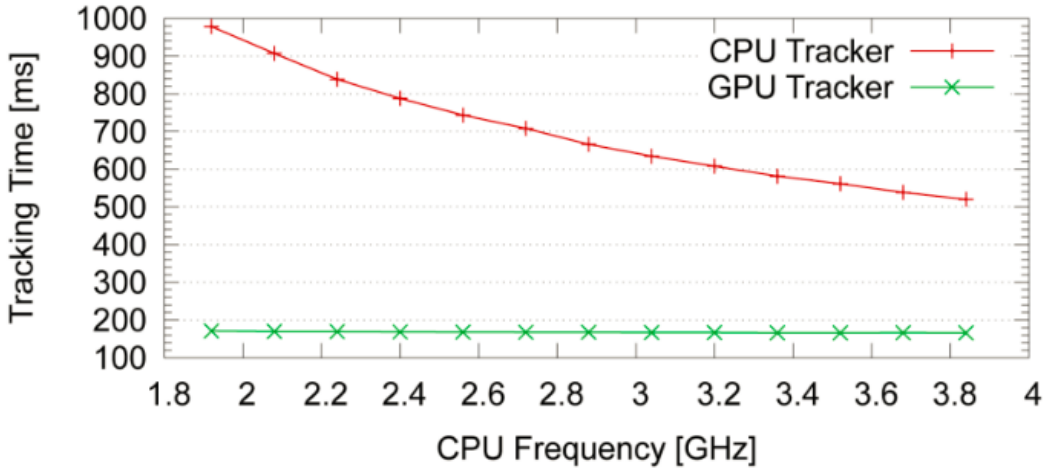
- High bandwidth main memory
- Latency tolerant (parallelism)
- More compute resources
- High performance per watt
- Limited memory capacity
- Low per-thread performance
- Extension card

GPU trigger systems – Gianluca Lamanna

ALICE TPC online tracking : 20000 tracks/event

Use of GPUs halves the number of computer nodes



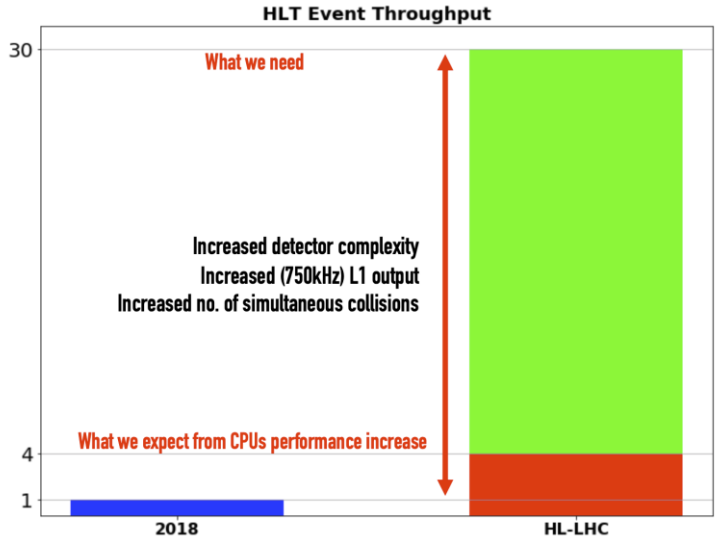


The reconstruction is dominated by the TPC

20-25 times speed up using GPUs

Task name	CPU Time [s]	GPU Time [s]
TPC sector track finding	706	11
TPC track merging	40	2
TPC track fit	300	6
TPC looping track following	150	6
TPC data track-based compression	100	2
Sum	1296	27
ITS clustering	10	
TPC-ITS track matching	1	
Global track matching to TRD	1	
Global track matching to TOF	1	
ITS tracking	10	
ITS tracklet vertexer (seeding)	1	
ITS (MFT) data compression	3	
TPC data entropy compression	35	
TPC gain calibration	10	
TPC distortions calibration with residuals	20	
Sum	92	
Total	1388	

GPU trigger systems – Gianluca Lamanna

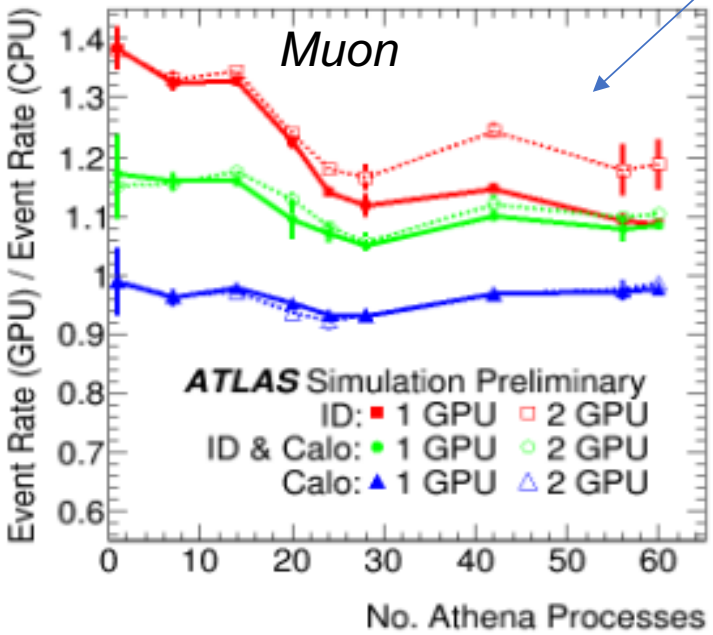


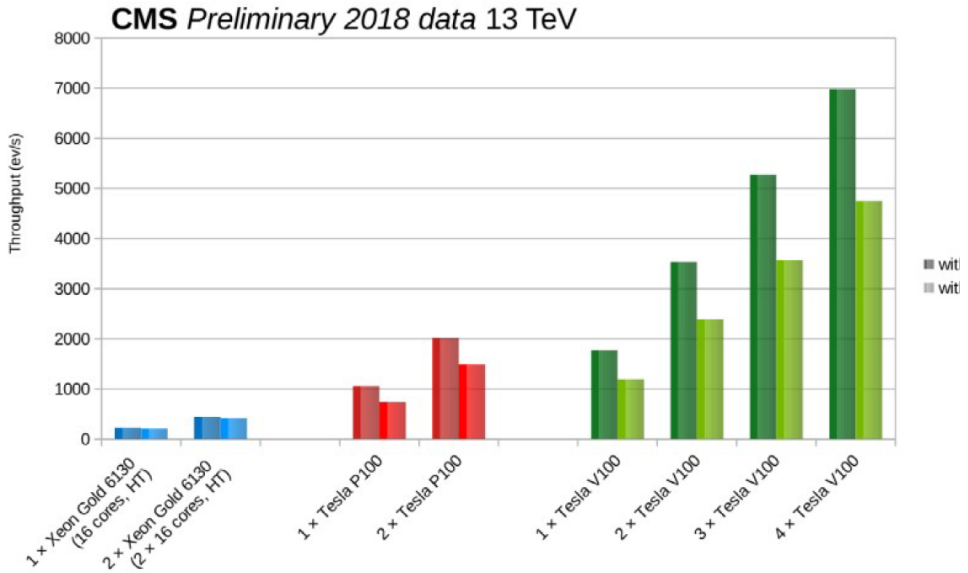
In the HL-LHC era the HLT computing load will increase significantly (~30x)

CMS have GPU based tracking ready for run3 for some reconstruction steps

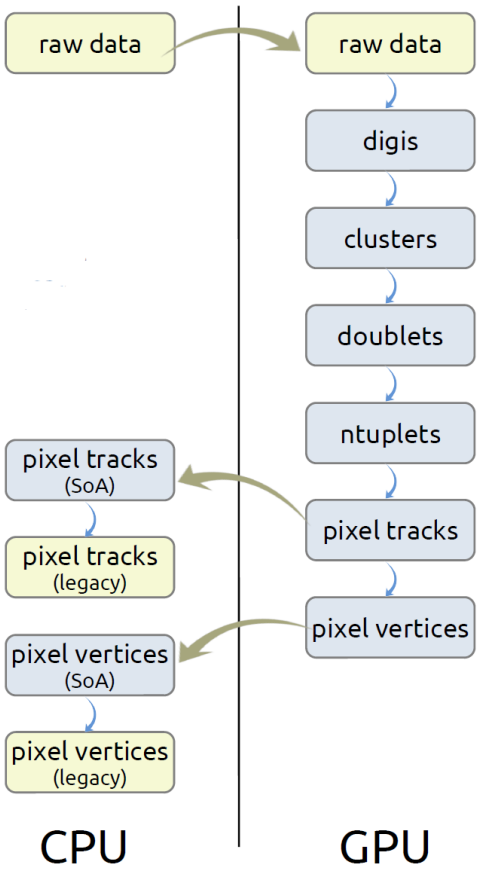
→ Possible 80% reduction in the farm

ATLAS tested GPUs but the gains were marginal due to athena not supporting concurrency and multithreading

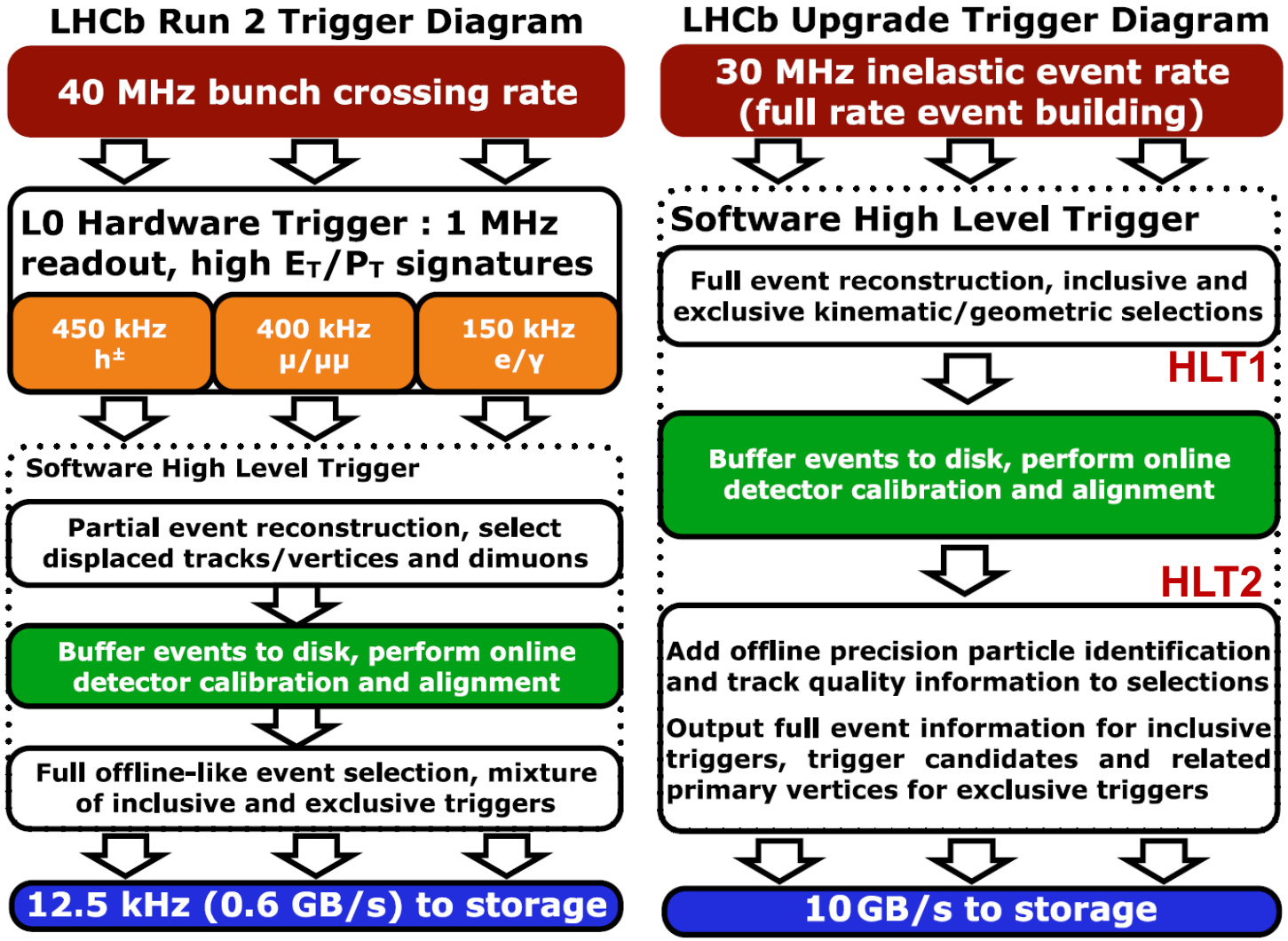




Tested during Run2 data taking



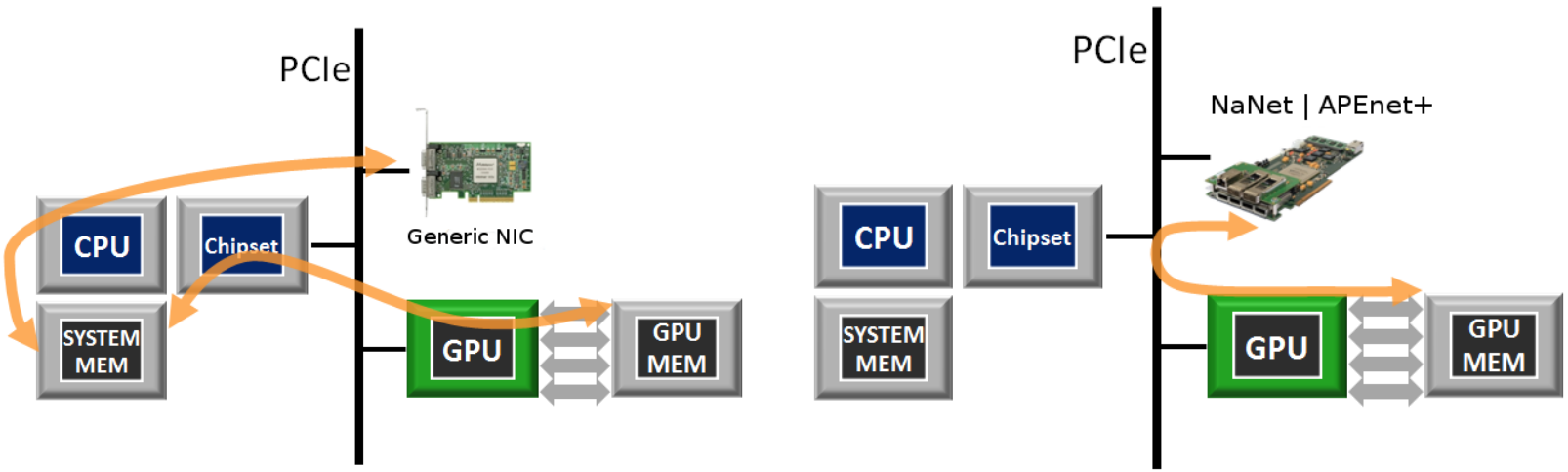
GPU trigger systems – Gianluca Lamanna

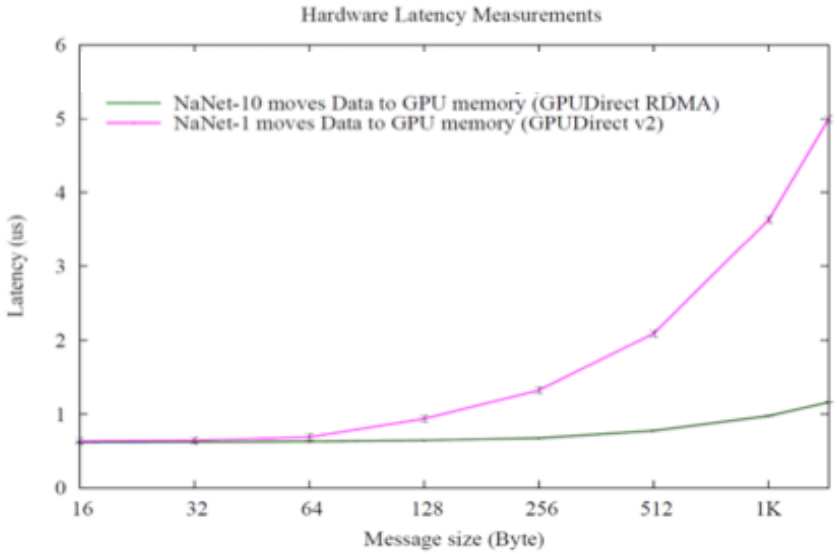


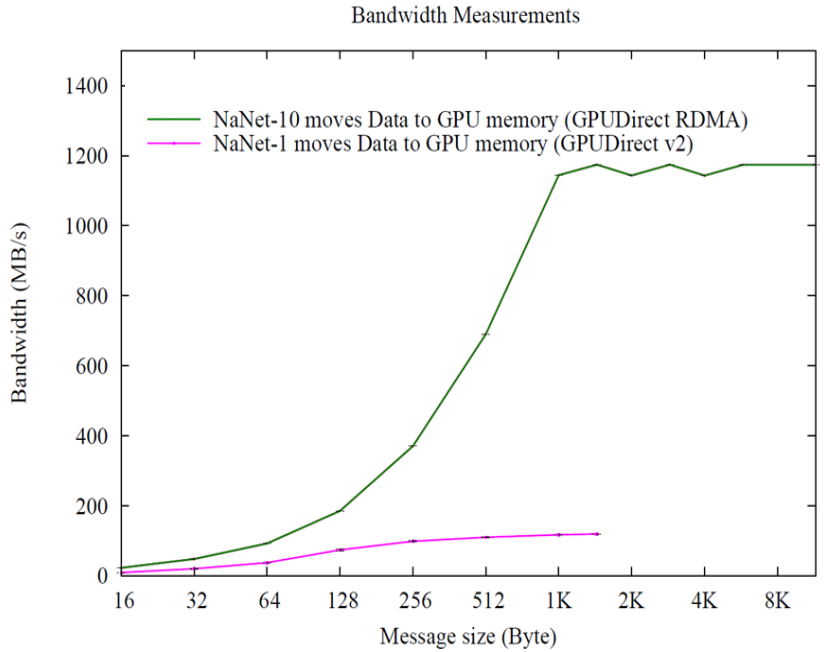
LHCb DAQ for run 3 will remove the L0 hardware trigger :

- Use of GPUs for HLT1 for full charged track reconstruction at 30MHz
  - Event rate reduced to 1MHz
  - Data rate 40 Tb/s to 1-2 Tb/s
- Next step to use GPUs for HLT2

GPU trigger systems – Gianluca Lamanna







The main problem with GPU computing is latency which is dominated by the double copy in the host RAM – hard to use GPUs for low level triggering

NaNet aims to solve this by optimising the data transfers with the GPU – will be used for the NA62 trigger system

GPU trigger systems – Gianluca Lamanna

- **The GPUs can be very useful to decrease the size of online computing farm and the dimension of the network**
- **Apart from a matter of saving money, some time this means have the possibility to have more Physics**
  - Better use of trigger bandwidth
  - Increase of trigger efficiency and purity
- **A hybrid system using both FPGA (for reliable and low latency operations) and GPU (for high computing throughput)** is probably a good solution for next generation experiments with demanding TDAQ requirements

Conclusions

We had a very interesting and informative Mu2e-II TDAQ workshop learning about the systems in use in other experiments, the upgrades and the path going forward which can feed into the plan for Mu2e-II going forward

- The evolution of the ATLAS TDAQ system with changing conditions of the LHC as well as the current updates for run 3 both in terms of hardware and algorithm development
- The lessons learnt through the development of the ATLAS FTK for online tracking
- The plans for CMS in the difficult environment of the HL-LHC with an emphasis on HLS and including a track trigger
- The ways that GPUs can be used in a variety of different ways in a range of experiments in order to reduce the size of computer farms as well as in triggering systems

A huge thanks to all the people who gave talks and participated in the workshop