ATLAS and CMS Trigger and DAQ upgrades for the High Luminosity LHC

Imma Riu (IFAE-BIST Barcelona)
on behalf of ATLAS and CMS Collaborations

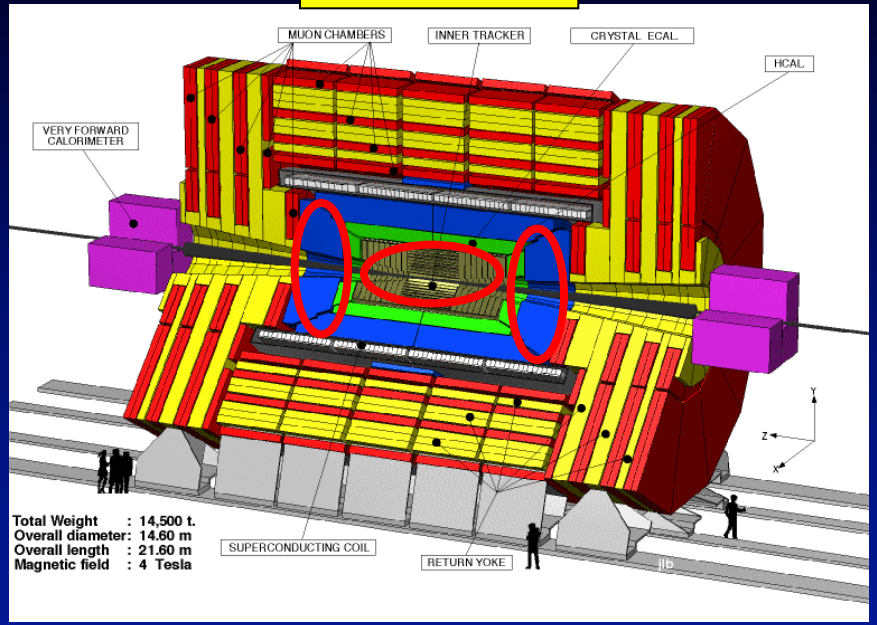
CHEP 2018, 10 July 2018
Sofia, Bulgaria

Outline

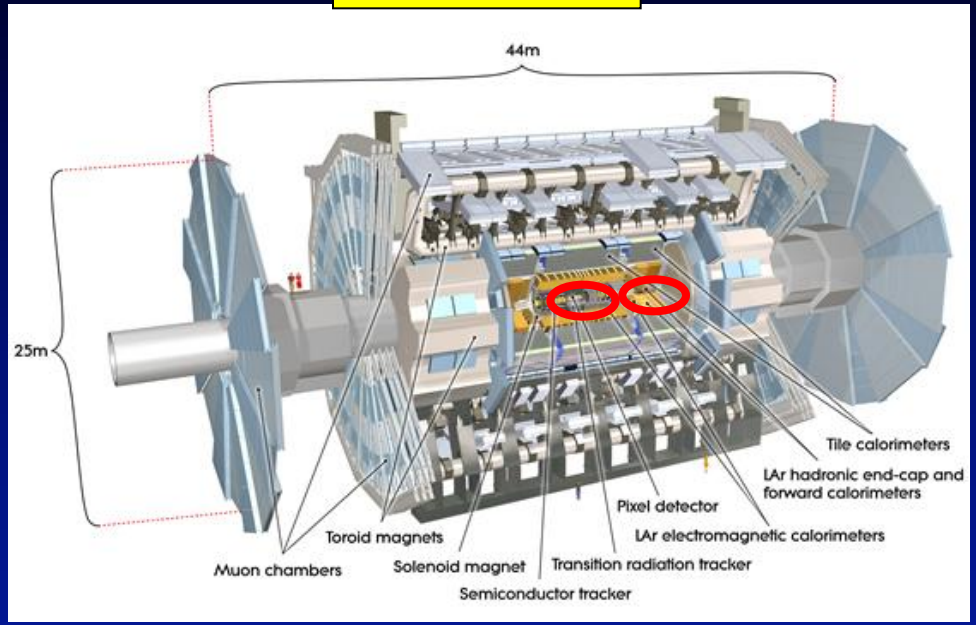
- Introduction
  - Detector upgrades and current operations
- HL-LHC challenges for TDAQ
- Upgrade physics motivation
- ATLAS and CMS trigger architecture
- CMS L1 Tracking trigger
- ATLAS Hardware Tracking for the trigger
- ATLAS and CMS DAQ description
  - Readout, dataflow, event building and storage
- ATLAS Event Filter
- Summary and outlook

Introduction: detectors and upgrade plans

CMS detector



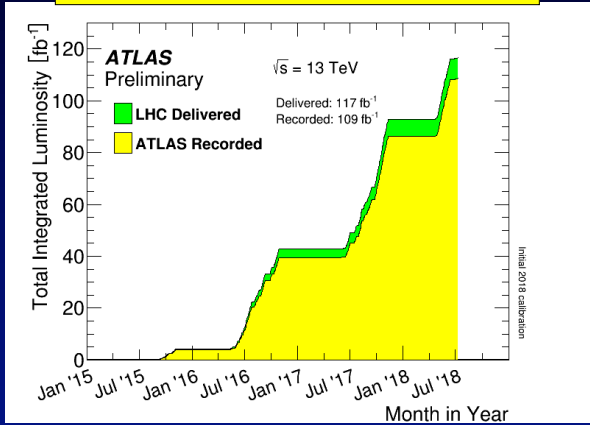
ATLAS detector



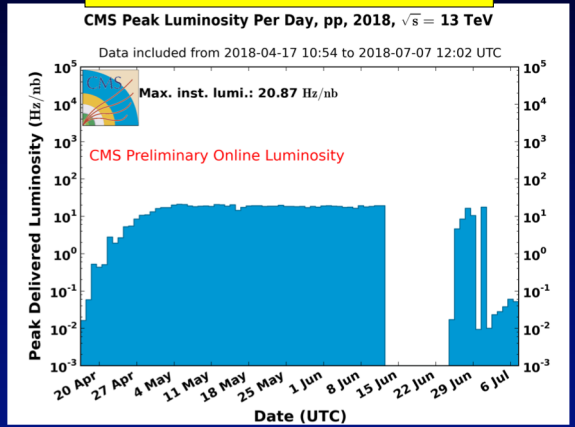
Upgrades of the Inner tracker, calorimeters and muon system planned by both detectors

Introduction: current LHC and detectors operations

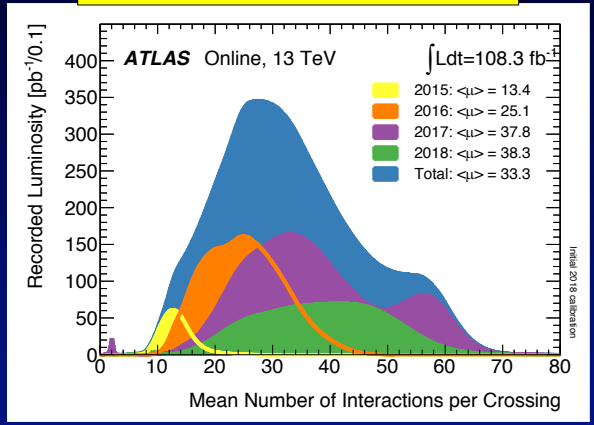
ATLAS Run 2 integrated luminosity



CMS peak luminosity in 2018

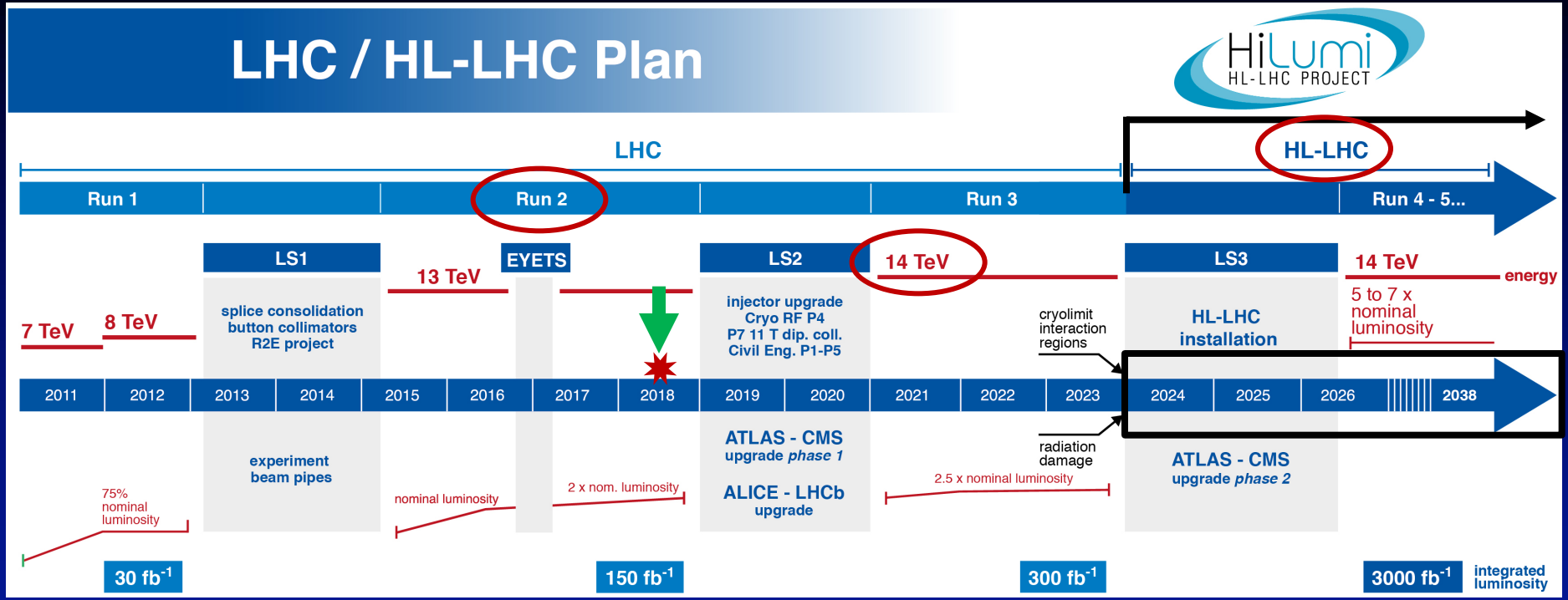


ATLAS mean # interactions / BC



- Total of Run 2 already recorded luminosity: ~117 fb^{-1} per experiment, ATLAS and CMS
  - Aimed total Run 2 luminosity: ~150 fb^{-1} per experiment
- Thanks to the excellent performance of LHC:
  - Exceeded the design luminosity of 10^{34} $cm^{-2}s^{-1}$ in 2017 and reached 2 x 10^{34} $cm^{-2}s^{-1}$ peak luminosity in May 2018
- Events with >60 interactions per bunch crossing (BC) were recorded in 2018 and some events with ~80 in 2017

HL-LHC upgrade plan overview



- Possible LHC scenarios:
  - Baseline scenario: 5.0 x 10^{34} $cm^{-2}s^{-1}$ peak luminosity and ~3 ab^{-1} integrated after ~10 years
  - Ultimate scenario: 7.5 x 10^{34} $cm^{-2}s^{-1}$ peak luminosity and ~4 ab^{-1} integrated after ~10 years

Civil engineering work just started. Ground-breaking ceremony held 15th June 2018 at CERN

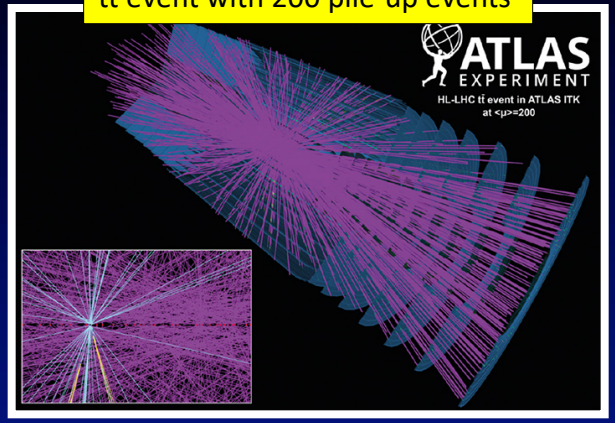
The civil engineering work for the High-Luminosity LHC gets under way. Here we see the earthmovers at work on the new 80 metre access shaft at Point 5. (Image: Julien Ordan/CERN)

Trigger and DAQ challenges in Phase-II

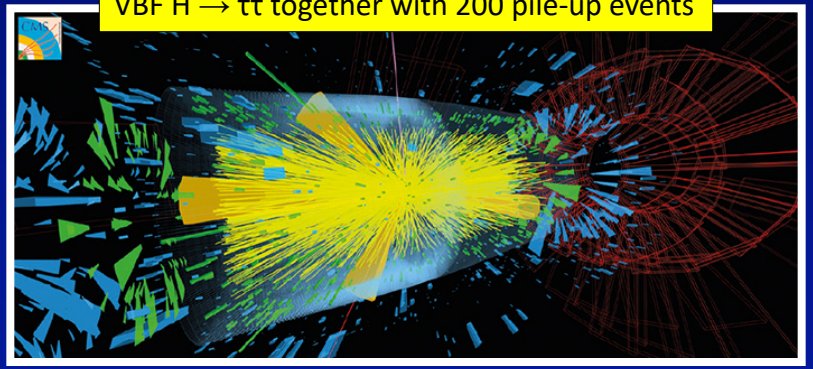
- High Luminosity consequences
  - High pile-up (up to 200 events/BC) with large number of tracks
  - Harsher radiation environment
  - High granularity detectors that need to be read out
  - Larger event size

- Requirements for Trigger and DAQ for CMS–ATLAS :
  - L1 latency increases to ~10–12.5μs (~2.5–3.2μs today)
  - Readout rate increases to 750–1000 kHz (100 kHz today)
  - Overall throughput to ~50 Tb/s (~2 Tb/s today)
  - Rate to permanent storage to ~7.5–10 kHz (~1 kHz today)

$t\bar{t}$ event with 200 pile-up events

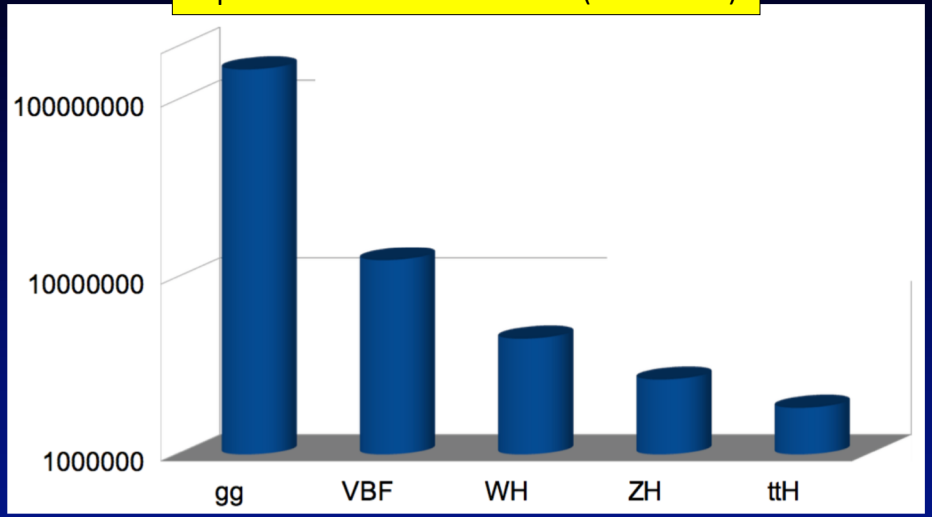


VBF H → ττ together with 200 pile-up events

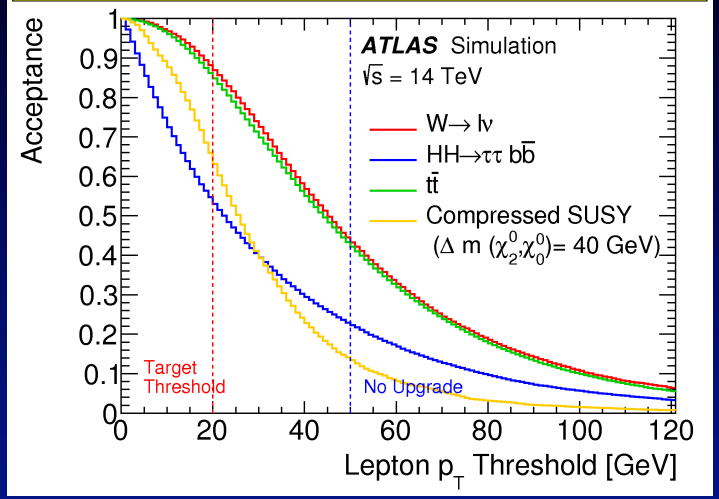


HL-LHC upgrade: physics motivation

Expected # events with 3 ab^{-1} (ECFA 2013)



Various physics signals acceptance versus lepton p_T

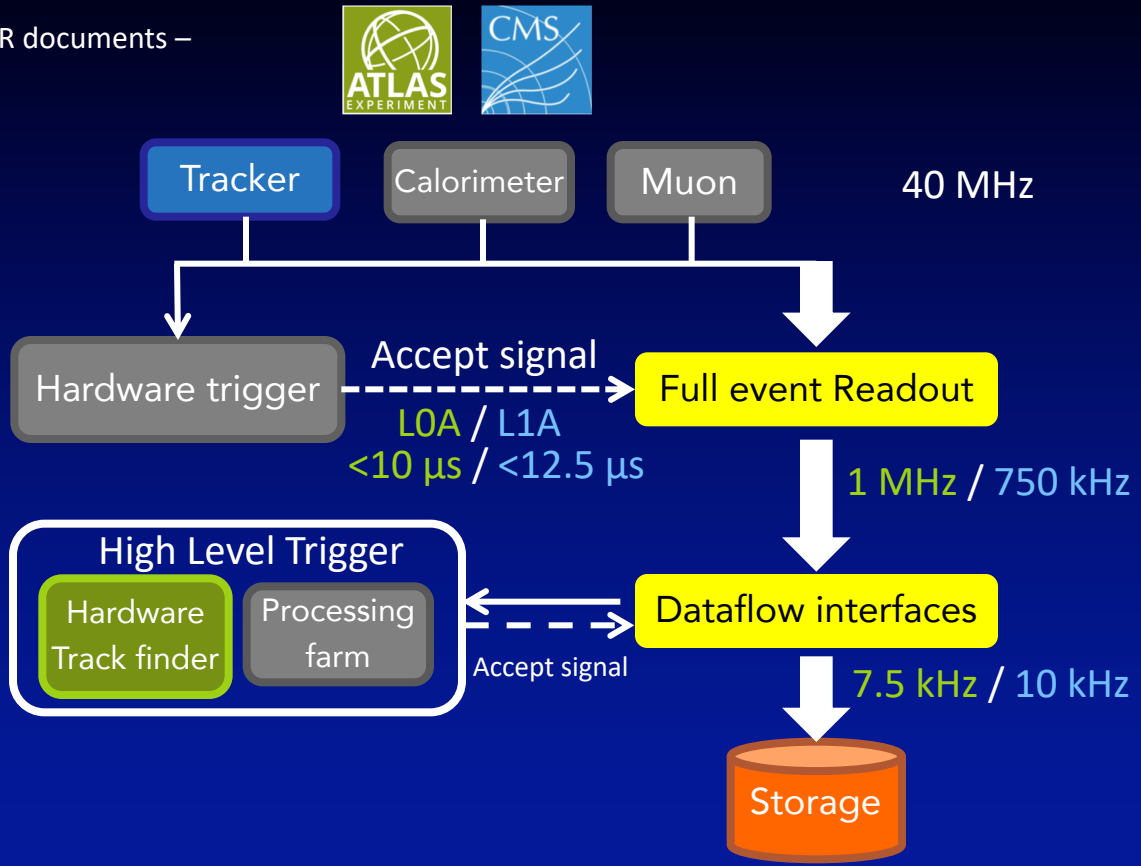


- Challenging and broad High-Luminosity LHC programme:
  - Precision measurements of the Higgs boson properties
  - Precision Standard Model measurements
  - Searches for Beyond-the-Standard-Model signatures
  - Flavour and Heavy-ion physics
- Requires keeping the p_T of the various trigger objects as low as possible
  - Electroweak scale requires low p_T leptons
  - Searches for new physics with e.g. low Δm too
  - HH measurements requires low p_T jets /b-jets

ATLAS and CMS upgrades architecture

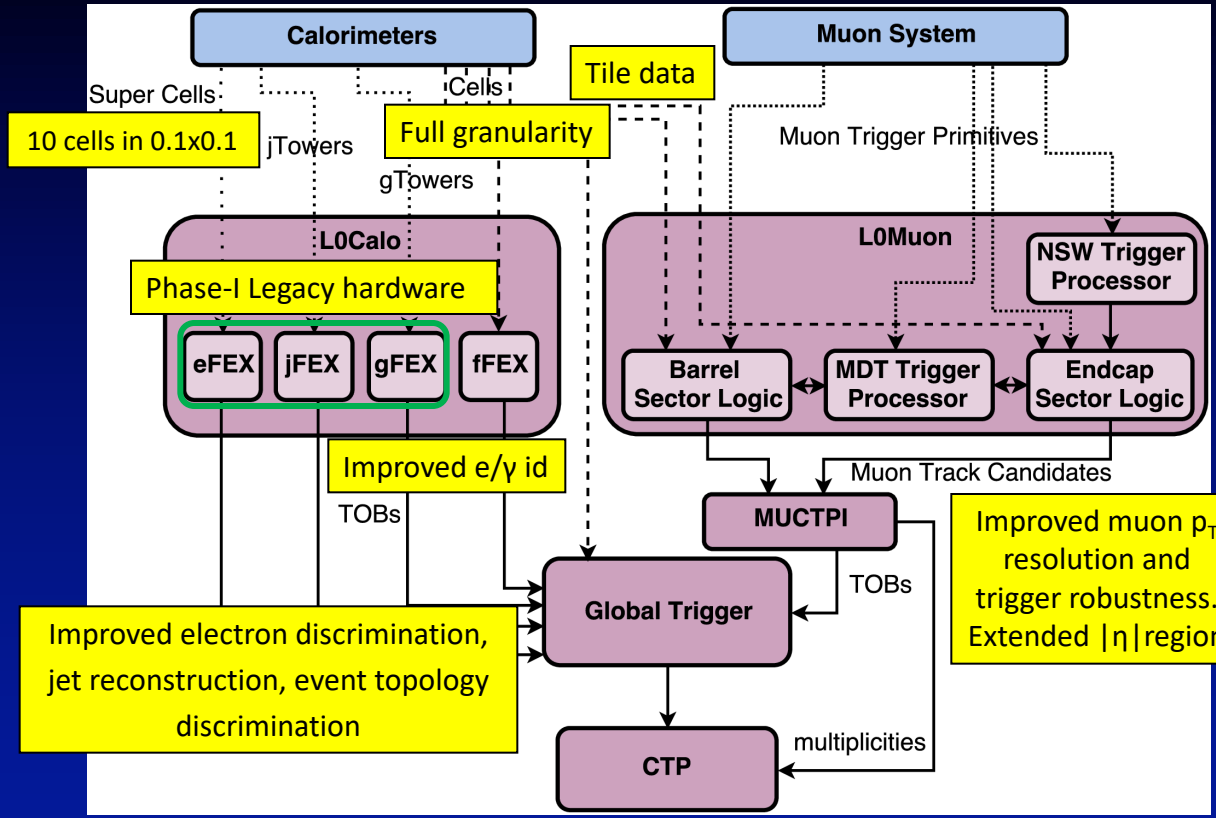
– Talk content based on 2 CMS interim and the ATLAS TDR documents –

- Two-level trigger architecture designed for both ATLAS and CMS:
  - First trigger level name differs:
    - L0 (ATLAS) and L1 (CMS)
  - Tracking data is used at L1 in CMS and at HLT in ATLAS in customised hardware
- ATLAS:
  - L0 latency/rate: <10 μs 1 MHz
  - HLT output rate: 10 kHz
  - Considers an evolution system with L0/L1 (w/ 400 kHz output rate) and HLT
- CMS:
  - L1 latency/rate: <12.5 μs <750 kHz
  - HLT output rate: 7.5 kHz



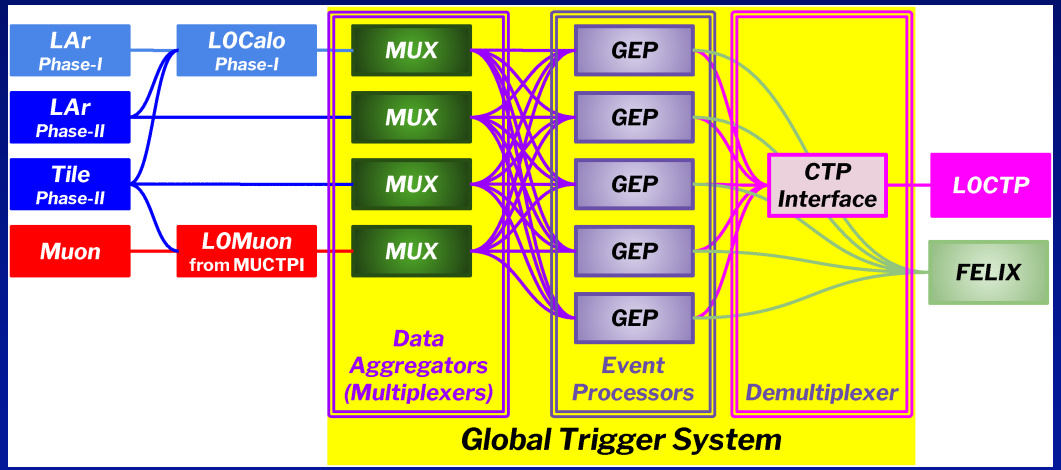
ATLAS Level-0 architecture

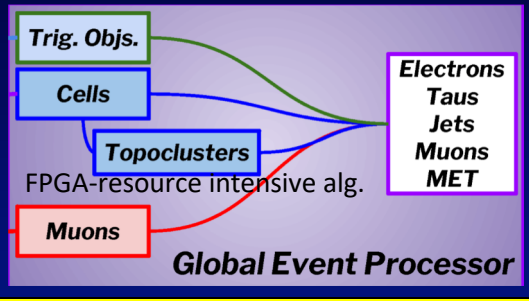
- L0Calo (5 ATCA shelves):
  - New in Phase-I and baseline for Phase-II
  - New forward-FEX (fFEX) in Phase-II for fwd EM and jets
- L0Muon (15 ATCA shelves):
  - New electronics and readout chain, new firmware
  - Use data from MDT, New Small Wheel (extend $|\eta|<2.6$)
- Central Trigger (1+1 shelves):
  - New Central Trigger Processor (CTP)
  - New MUon-to-CTP-Interface (MUCTPI)
- L1 latency (<10 μs):
  - 6.9 μs best estimate



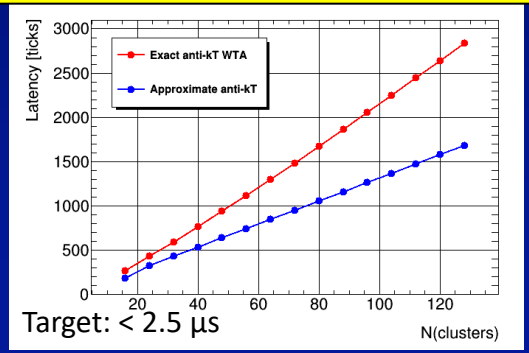
ATLAS Global Trigger description

- Time-multiplexed system concentrating data of full event into a single processor at 40 MHz
  - Input: >2300 optical fibres with different link speeds/protocols
  - Three steps: data aggregation layer (MUX) time-multiplexing, event processor layer (GEP), final demultiplexing Global to CTP interface
- ATCA-based Global Common Module
  - Common Module for MUX and GEP that differ in firmware



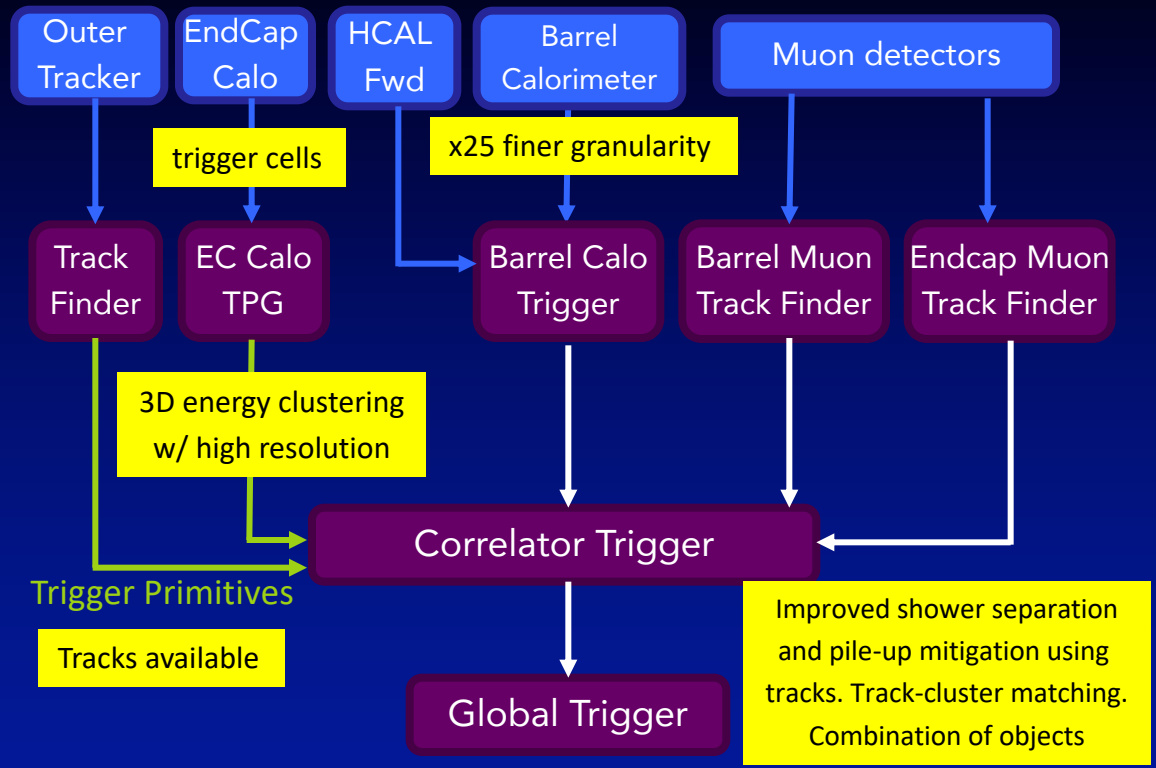


Latency of anti-k_t algo. variants in FPGA clock-cycles (320 MHz)



CMS Level-1 architecture

- Muon track finder
  - Use of additional chambers
  - Potential use of Kalman Filter
- High Granularity Calorimeter Trigger Primitive Generator
  - Use ~900k channels in the trigger (out of 6M)
- Track Finder Trigger Primitive Generator (TPG)
  - Use the outer layers
- Correlator trigger
  - Combines objects to build high-level trigger objects and particle-flow identification
- Latency target for front-ends to receive the L1A signal: 9.5 μs



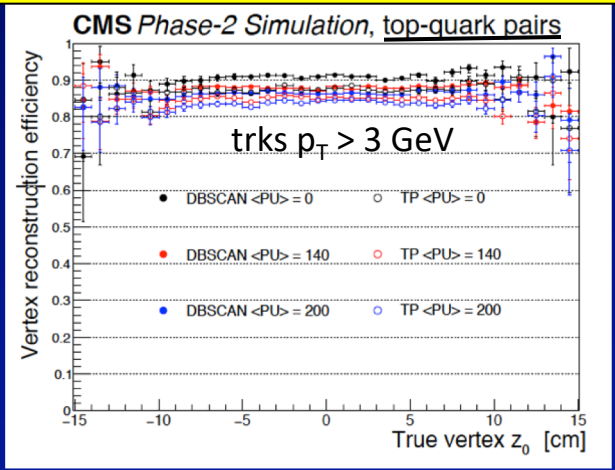
"Particle Flow reconstruction in the Level-1", G. Petrucciani

"Overview of the CMS L1-trigger for HL-LHC", R. Cavanaugh

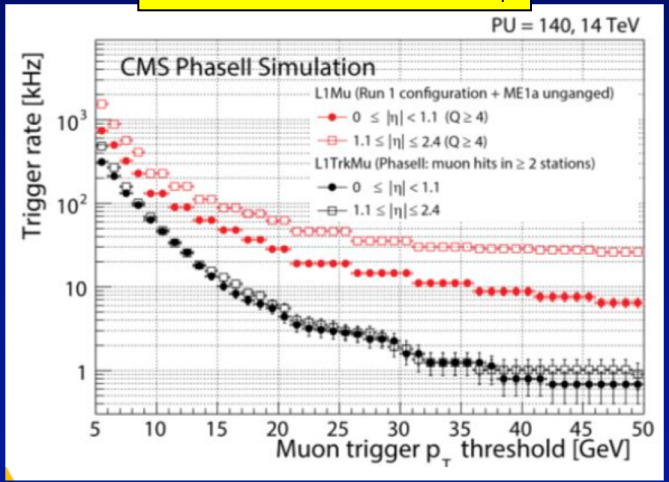
CMS tracker data in L1: motivation

- Tracker data is used at L1 in CMS
  - Essential as the rate without tracking is ~4 MHz with 200 pile-up events
  - Allows vertexing at L1 (good performance observed of algorithm in FPGA)
  - Object p_T thresholds can be maintained similar to thresholds used in Run 1

Hard interaction primary vertex reco. eff. within 1.5 mm



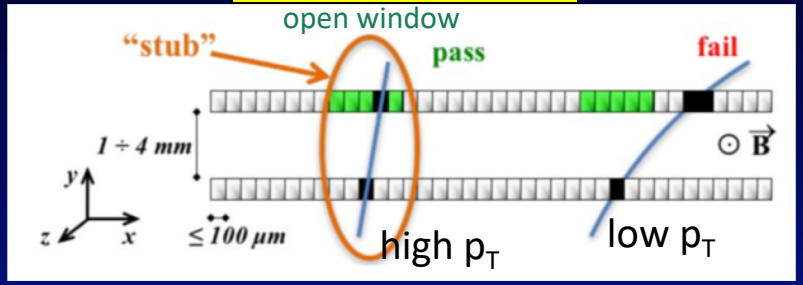
Trigger rate versus muon p_T



CMS L1 tracking description

- Tracker provides limited hit data at full 40 MHz rate
  - No use of inner pixel tracker hits (Inner tracker)
  - Only hits from outer 6 tracking layers used (Outer tracker)
- p_T calculation provided by using special p_T modules
  - Inspection of pairs of closely spaced silicon sensors, separated O(1.6-4) mm, to see if signals are consistent with the passage of a high p_T particle
  - Only for compatible hit pairs with p_T>2-3 GeV the so-called "stubs" are generated
  - Factor ~10 data reduction achieved
- Significant data rate:
  - Average 15000 stubs/25 ns sent from detector to Track Finder (trk. reco. & fit) located underground
  - Total bandwidth O(20) TB/s
- Track finding from stubs must be performed in ~4 μs

CMS "stub" concept



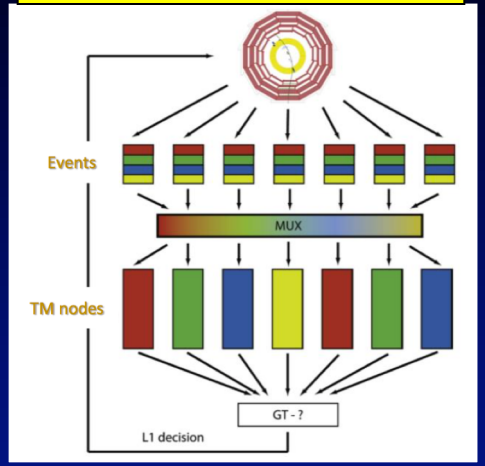
CMS L1 latencies

Transmission of stubs to BE electronics	1 μs
Correlation of trigger primitives (inc. tracks)	3.5 μs
Broadcast of L1 accept to FE buffers	1 μs
Safety Margin	3 μs

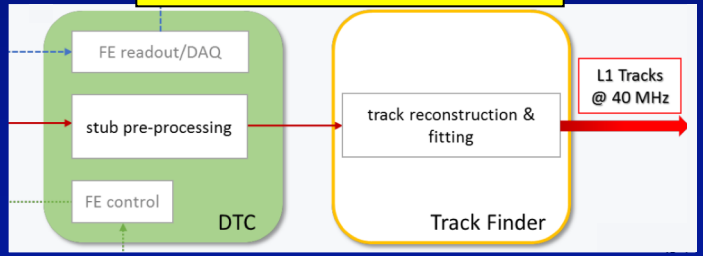
CMS L1 track finder architecture

- Outer Tracker cabled into nine regions
  - Use of time-multiplexing to parallelise data processing
  - Time-multiplexing directs data from multiple sources to a single processing node
- Data processing in two stages
  - DAQ, Trigger and Control layer (DTC)
    - Stub pre-processing
    - Time-multiplexing into a single processing node
    - Size: 216 DTC boards (~600 Gb/s) in 18 ATCA shelves
  - Track Finding Processing layer (TFP)
    - Operates on an independent **η,φ** region
    - Track finding from stubs, fitting and transmission of tracks to the L1 Correlator Trigger
    - Size: 144 TF boards (~1 Tb/s), 12-18 ATCA shelves
- System implemented in ATCA-based boards with FPGAs

Time-multiplexing concept



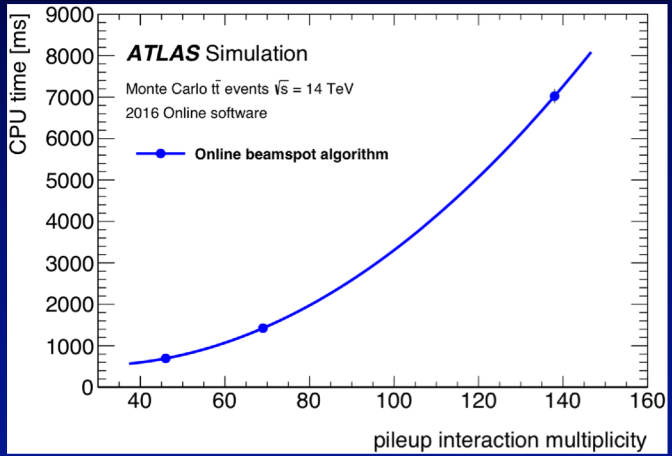
Two step data processing



ATLAS tracker data in EF: motivation

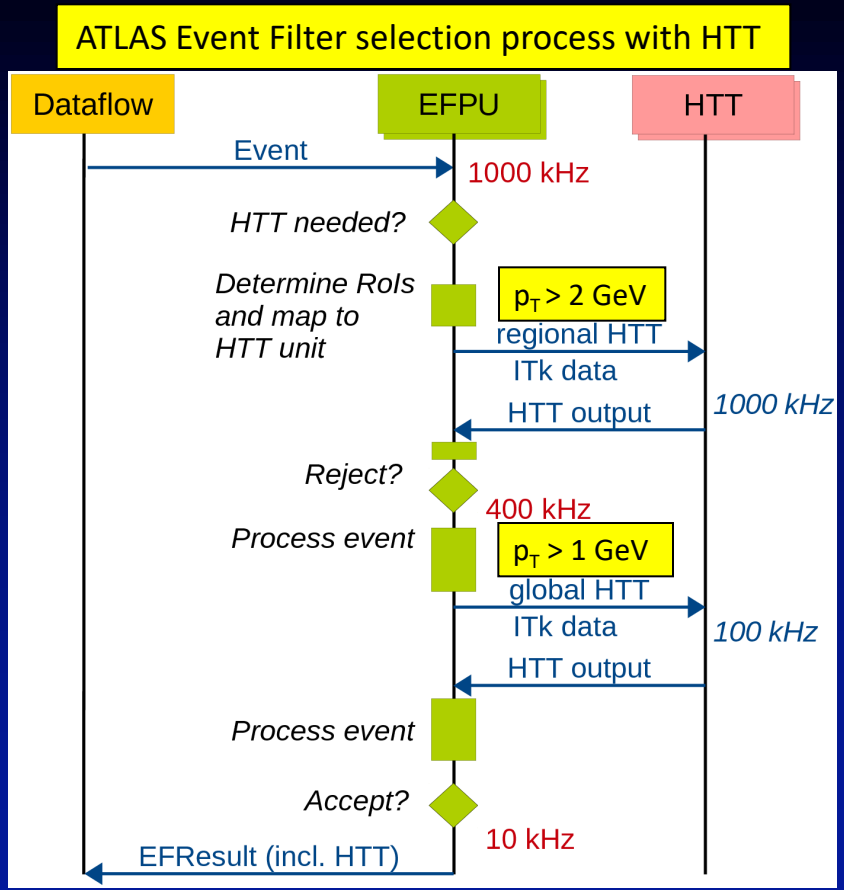
- Motivation for the ATLAS Tracker usage in the EF
  - The trigger objects p_T thresholds need to be low
  - A way to reduce the data rate a factor of 10 from 1 MHz
- Unrealistic solutions
  - CPU time for software-based track reconstruction scales geometrically with pile-up (~30k dual-socket servers estimate)
  - Using the Run 3 Fast TracKer (FTK) hardware does not scale at HL-LHC
    - A factor of 20 input bandwidth increase with respect to FTK is required
- Current plan
  - Designed a new custom hardware-based track finding co-processor known as the Hardware Tracking for Trigger (HTT)
    - Allows p_T threshold to be reduced from 10 to 2 GeV
    - Factor 10 reduction in CPU power required in the HLT farm (from 30k to 3k dual-socket servers)

CPU time versus pile-up interactons



Two-step ATLAS tracking in HTT

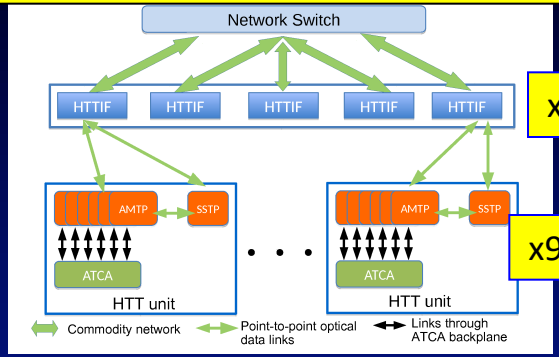
- Regional HTT
  - Level-0 information provides trigger path and Regions of Interest
  - Partial tracker event data pulled from storage (~10% on average)
    - Used hits from 8 outermost trk layers only
    - Tracks of $p_T > 2$ GeV sent back to EF
  - EF Processing Units (EFPU) makes use of rHTT tracks and helps to reduce rate to 400 kHz
- Global HTT
  - If required, full hit data (strip and pixel) pulled from storage
  - Full tracking ($p_T > 1$ GeV) performed by HTT
  - EFPU uses gHTT tracks, in combination with offline-like analysis
    - Helps reduce rate to the required 10 kHz



HTT hardware description

- ATCA-based system w/ 672 tracking processors (TP), each handling a small**η-φ** region, in 56 ATCA shelves
- Single base card differentiated by mezzanines implement the 1st and 2nd stage processors (~300 W)
  - regional HTT– AMTP boards (6 AMTP / HTT unit)
    - Uses Associative Memories (AMs) to match roads from clusters in up to 8 layers
    - Hit clustering and data organization
    - Track Fitting on roads performed in an FPGA
    - Pattern Recognition Mezzanines (PRM)
  - global HTT – SSTP board (1 SSTP/ HTT unit)
    - Second Stage (SS) TP receives 1st stage tracks/clusters from AMTPs, plus full event data from HTTIF
    - Hit clustering in new layers
    - Track extrapolation and re-fitting performed in FPGAs
    - Track Fitting Mezzanines (TFM)

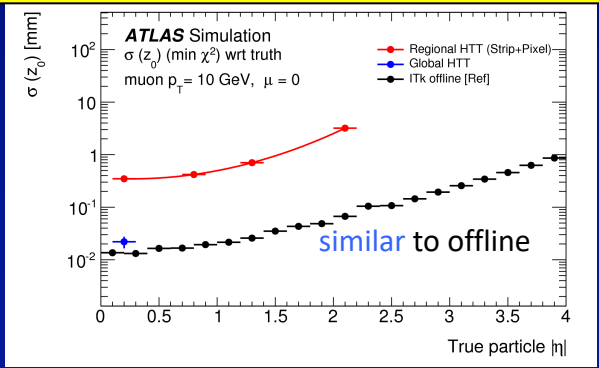
Hardware Tracking for the Trigger system overview



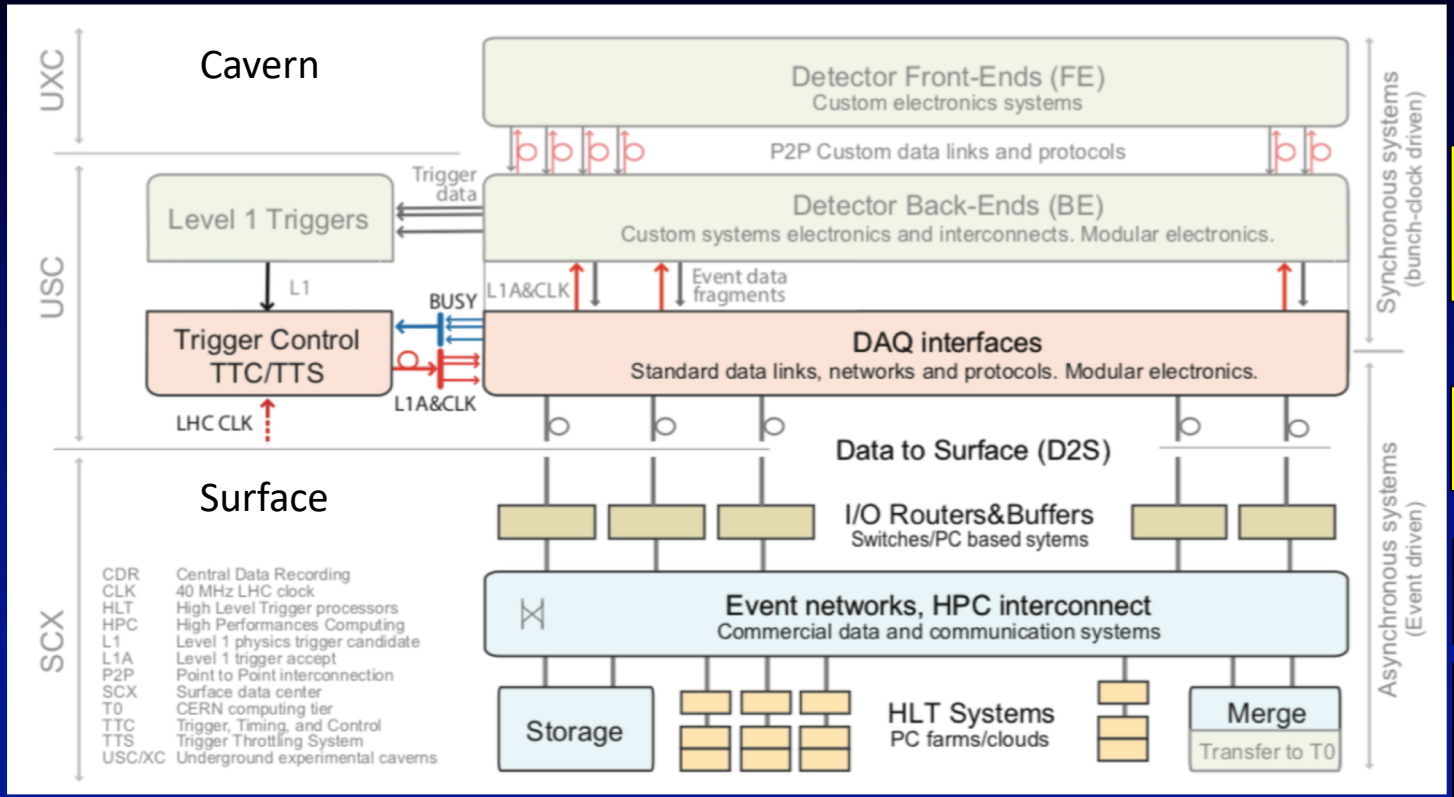
x24 HTTIF

x96 HTT units

Comparison of z_0 resolution from 1st & 2nd stage fitting



CMS DAQ system architecture



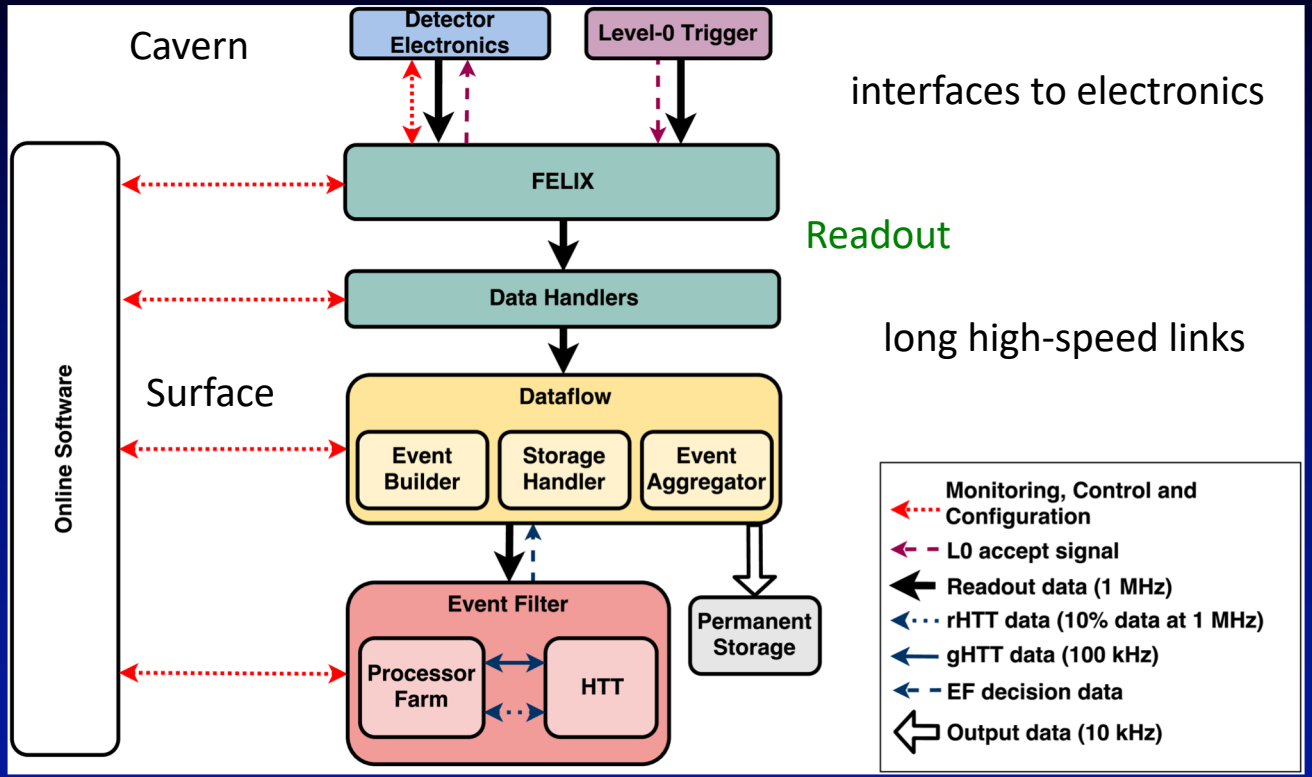
synchronous timing, trigger accept and fast control signals

high speed lossless protocol links

high performance switch network for EB

Accepted events are stored locally before long-term storage

ATLAS DAQ system overview



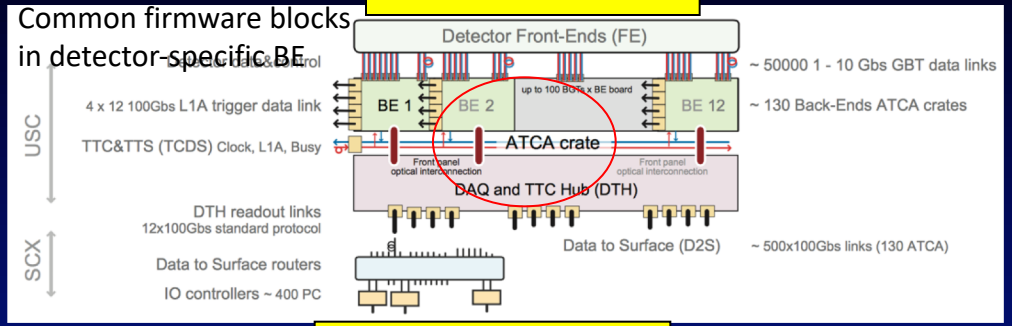
CMS and ATLAS detector interface and readout

- CMS
  - ATCA-based with detector-specific Back-End boards embedding DAQ firmware for data transmission via front panel
    - Deal with DAQ and timing hub
  - Data sent to surface over reliable protocol (e.g TCPIP) to I/O processors

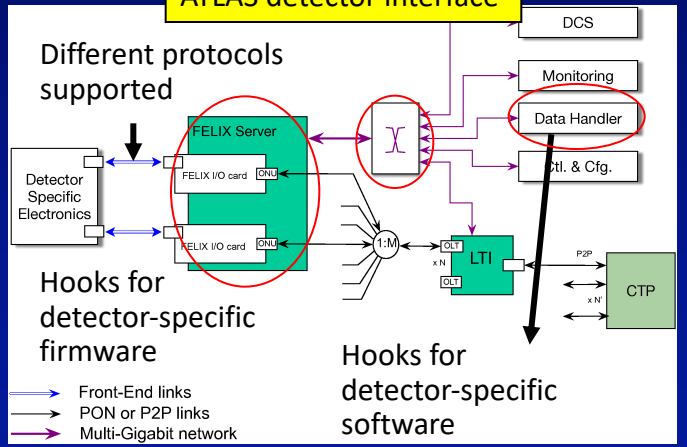
- ATLAS
  - PCIe-based FELIX cards hosted in servers:
    - Custom Front-End LInk eXchange (FELIX), the new interface to detector-specific electronics including limited detector-specific firmware
  - Data Handler servers running a software application that deals with event fragments and embeds detector-specific software

"FELIX: the new detector interface for the ATLAS experiment", S. Kolos

CMS detector interface



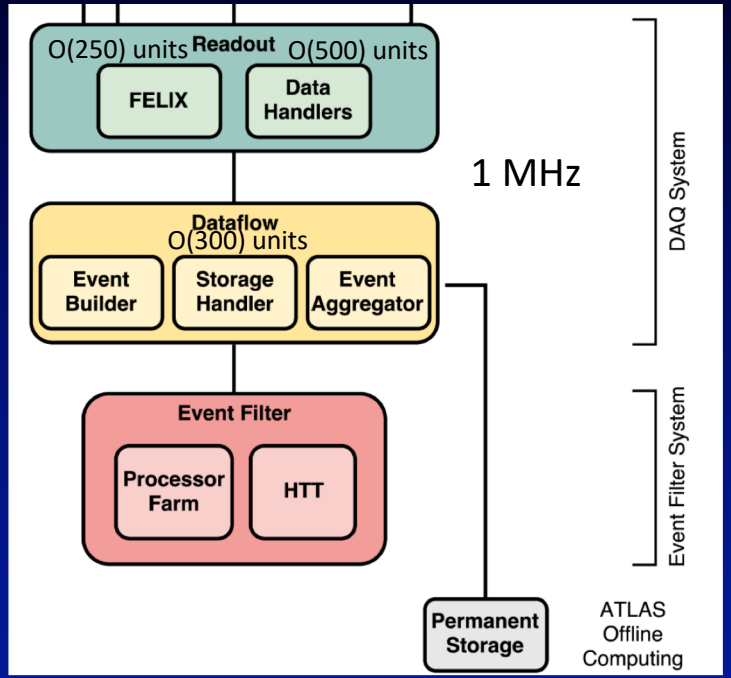
ATLAS detector interface



ATLAS Dataflow system

- Functionality description
  - Buffers data before, during and after the Event Filter decision
  - Provides partial and full event access as needed and transfers data to permanent storage
  - Managed by commodity software

- Main components
  - Storage Handler
    - High throughput (7.8 TB/s) large storage system (up to 36 PB) buffering event data before and during EF processing
  - Event Builder
    - Logical interface of the dataflow to Data Handlers and Event Filter
    - Manages consistency of distributed data and data access
  - Event Aggregator
    - Receives full events at 10 kHz, buffers them for up to 48h (10 TB) and throughputs to permanent storage at 60 GB/s



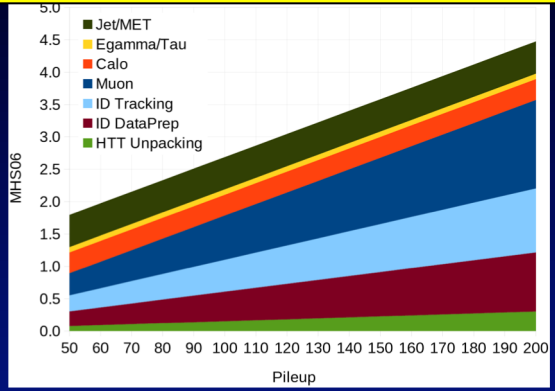
ATLAS and CMS event building and storage

- ATLAS
  - 100 Gb/s interconnects (e.g Ethernet) as baseline
  - Data injected directly from the readout into storage elements
  - Introduction of logical event building, i.e. data fragments temporarily stored in large distributed high performance storage (~1h)
    - The Event Builder is the book-keeper of the locations of event fragments in the storage system
- CMS
  - 200 Gb/s interconnects (e.g. Infiniband HDR) as baseline
  - I/O servers (~500) receiving data from readout (TCP) acting both as event building sources and destinations
  - Complete events temporarily stored (~5 PB) in high-speed storage to serve HLT (~1 min)
- Permanent storage: ATLAS and CMS have modular, scalable systems based on Hard Disk drives to aggregate accepted events into files and transfer them to Tier0
  - ATLAS foresees to use the same distributed storage system used for temporary buffering before HLT while CMS will decide the implementation later

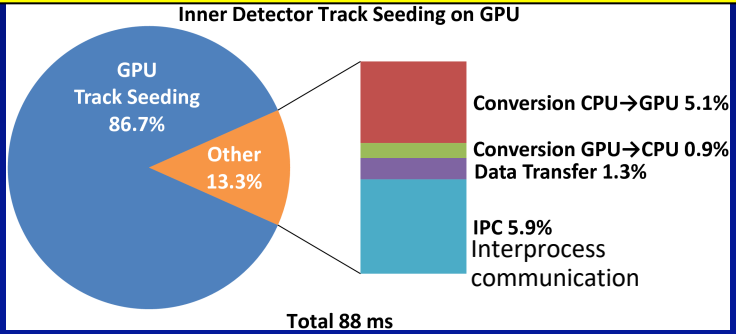
Event Filter farm in ATLAS

- The Event Filter (EF) farm
  - Responsible to perform the HLT selection in a massively parallelised way
  - Both hardware and software need upgrading
    - Most cost-effective are commodity PC servers
- Current farm size estimate
  - 4.5 MHS06 (+HTT) to handle a L0 rate of 1 MHz
    - CMS estimate is 9.2 MHS06
  - 3000 dual-socket servers in 38 racks
- Test of GPGPU
  - Tested two scenarios: up to four NVIDIA GK210GL Kepler archi. and a GTX1080 PCIe card w/ Pascal archi.
  - Cost to increase the farm throughput would be similar when adding GPGPU or CPU
  - With GPUs found very important to implement suitable event data format to avoid time-consuming conversions

ATLAS EF CPU extrapolation versus event pile-up



Breakdown of task time/event in a Kepler system



Summary

- Plans for the Trigger and Data Acquisition systems of ATLAS and CMS for the High-Luminosity Upgrades are detailed in a Technical Design Report or interim documents
  - They will support an exciting and diverse physics programme
- Both trigger systems are designed in two levels: L0 and EF in ATLAS, L1 and HLT in CMS
  - Flexible and scalable architecture to accommodate LHC conditions and physics needs
- The use of tracking information in the trigger is planned in both experiments to reduce rate
  - CMS plans to do tracking using limited hit data at L1 at 40 MHz while ATLAS plans to do it within EF in a co-processor at 1 MHz
- The planned DAQ systems rely on the experience accumulated through Runs 1 and 2
  - ATLAS and CMS DAQ systems mainly differ on the implementation of detector interfaces
  - Most components need to be re-implemented to sustain increased event size and trigger rates
- Common R&D projects planned or ongoing to design the final system

Outlook: examples of common aspects

- Extensive use of big FPGAs
  - Power delivery and thermal management and optics are key challenges
  - Appropriate manpower expertise needed
- Use of Multi-Gigabit Transceiver links designed beyond 10 Gb/s (16 and 25 Gb/s)
  - Required given the expected detector data rates
- Extensive use of ATCA crates
  - Cooling is a key challenge
- Potential use of GPUs
  - Careful programming for an efficient performance
- Increased complexity of system-level integration, maintenance and operations
  - Use of identical blades with different firmware foreseen in several places
  - Firmware management needed

THANK YOU !

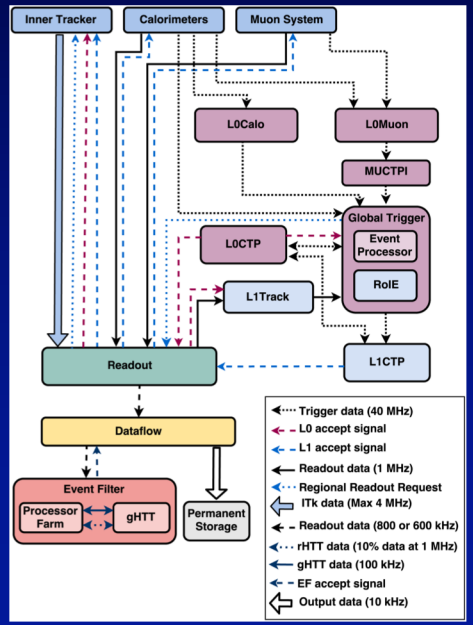
References

- Technical Design Report for the ATLAS TDAQ system Phase-II Upgrade
  - https://cds.cern.ch/record/2285584
- Interim Design Report of the CMS DAQ Phase 2 Upgrade
  - https://cds.cern.ch/record/2283193
- Interim Design Report of the CMS L1 Trigger Phase 2 Upgrade
  - https://cds.cern.ch/record/2283192

BACKUP

ATLAS evolved system

- Two-level trigger before event readout
  - L0 and L1
- Based on regions of interest
- Regional tracking done at 4 MHz or 2 MHz (instead of 1 MHz)



System Component	Hooks for Evolved System	Additional Hardware/Firmware Needed for Evolved System
Level-0 Calo Level-0 Muon	sufficient FPGA resources	minor firmware changes
Global Trigger	extra transceivers	additional MUX modules to receive information from L1Track; extra GCM modules configured as RoIE; extra GCM module for L1CTP interface; additional GEP firmware
Central Trigger	extra optical connectivity	add L1CTP plus additional patch panels and fibres
Readout	none	additional FELIX I/O Cards, servers, and Data Handlers; new FELIX/Data Handler firmware and software; low-latency links to L1Track;
Dataflow	none	larger bandwidth requirements
Event Filter	none	significant increase in computing power
HTT	rHTT hardware and firmware must meet L1Track latency requirement	separation of regional and global functionality; additional AMTP; new firmware

Trigger	Latency requirement	Level-0 rate [MHz]	Trigger threshold [GeV]
rHTT	No	1	2
L1Track	6.0 μs	2–4	4

Decision making to evolve

ATLAS risk mitigation: two-level trigger system

- The baseline L0 system can evolve to a L0/L1 system
  - L1Track is introduced at L1
    - Reuses HTT hardware from the baseline system
  - Additional event rejection performed at L1
  - L1 output trigger rate from 1 MHz to 600-800 kHz
- Criteria for evolution
  - Hadronic trigger rates higher than expected
    - Large uncertainties in the hadronic trigger rates
  - Inner Itk pixel occupancy higher than expected
    - "Fast-clear" after L0 reduces readout bandwidth per FE chip