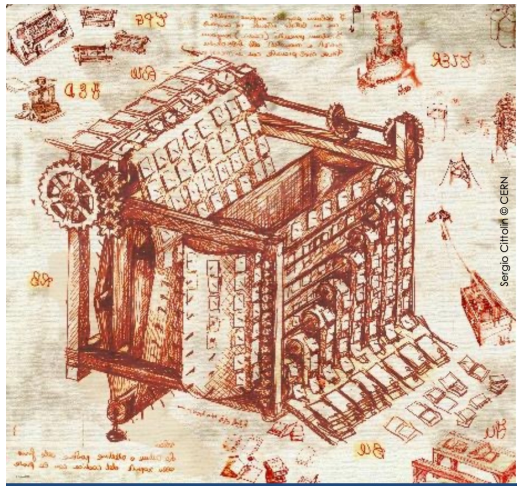
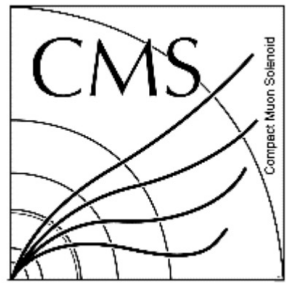
CMS detector: Run 3 status and plans for Phase 2



Srećko Morović
On behalf of the CMS Collaboration

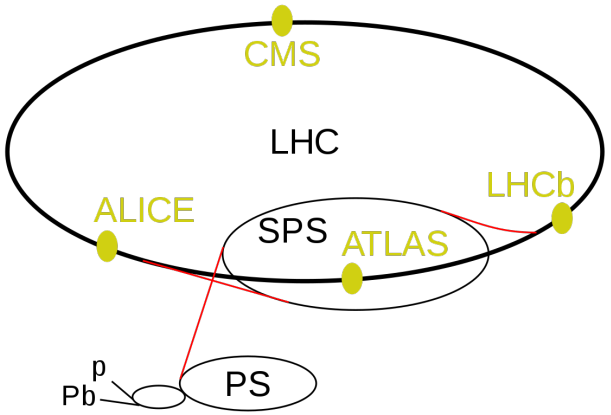
DIS 2023
29. 03. 2023.

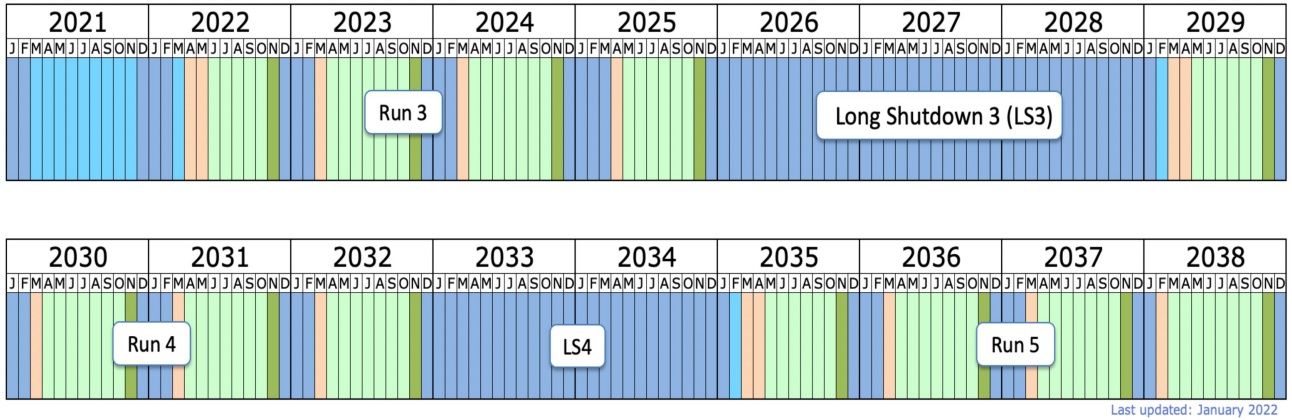


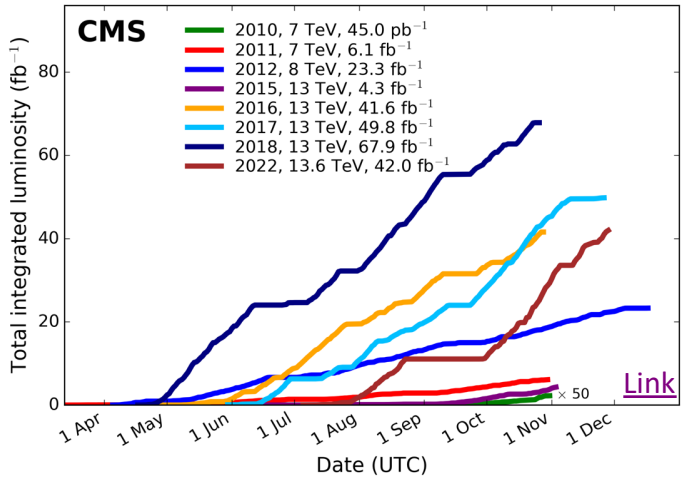
UC San Diego

Introduction

- LHC Run 2 completed in 2018 → start of Long Shutdown (LS2)
  - Upgrade of LHC and main experiments
- LHC operation restarted in 2022 with Run 3
  - Proton-proton collision energy - 13.6 TeV
  - Target: collect above the total luminosity collected previously (nearly 170 fb^{-1})
  - Increased average instantaneous luminosity (and pileup) from Run 2 (~ $2*10^{34}$ cm^2/s)
- After 2025 - LS3 and High-Luminosity LHC
  - Phase-2 CMS upgrades







Run 3 CMS detector upgrades in a nutshell

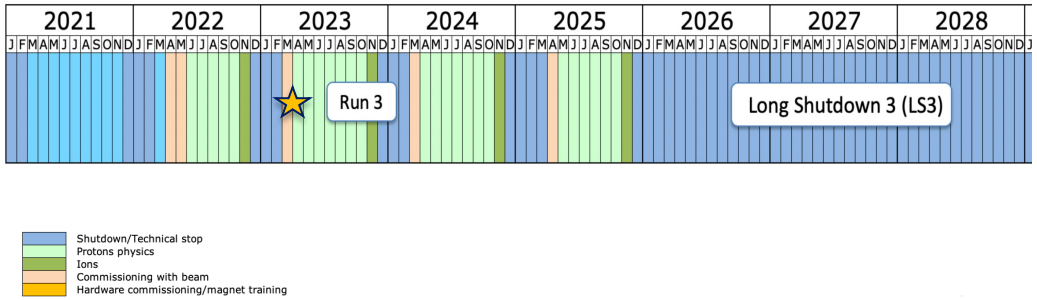
Phase-1 (LS1 → Run 2 →LS2) highlights:

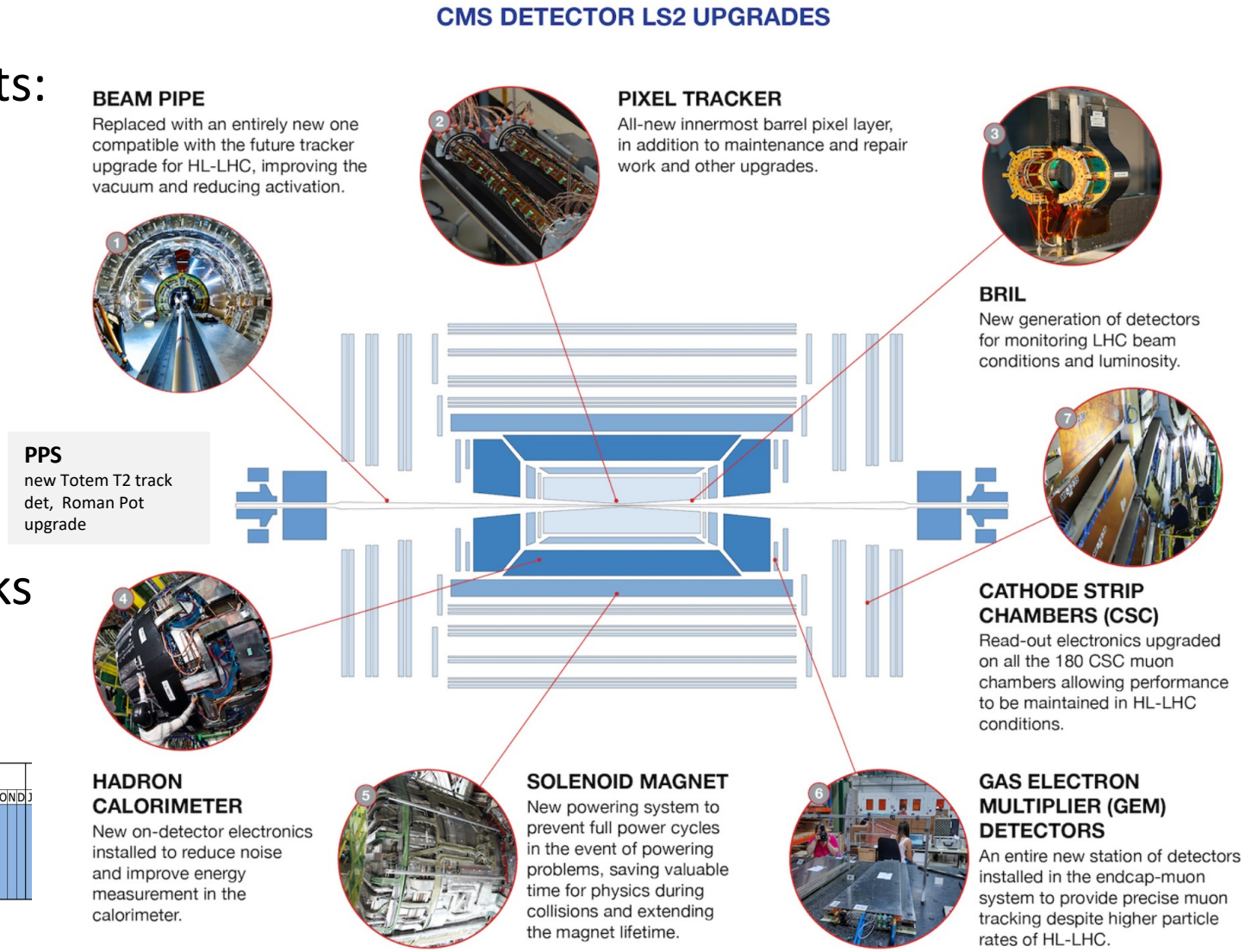
- Inner (Silicon Pixel) Tracker replacement, L1-Trigger and HCAL upgrade

Phase-2 upgrades

- Upgrade programme for the HL-LHC (discussed later…)

Many upgrade and refurbishment tasks performed in in LS2



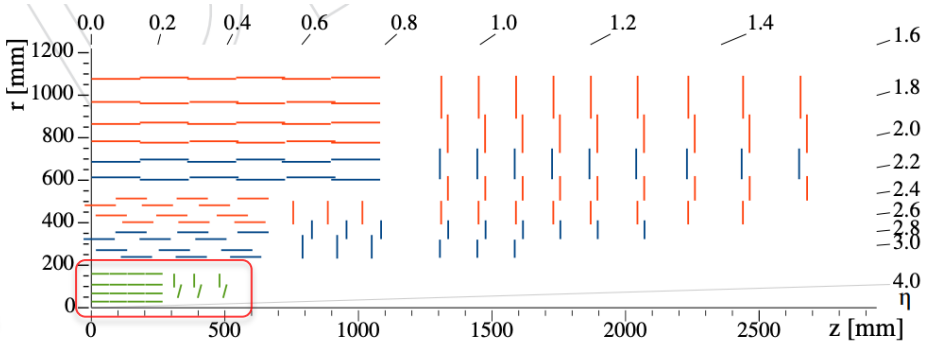


https://home.cern/press/2022/CMS-upgrades-LS2

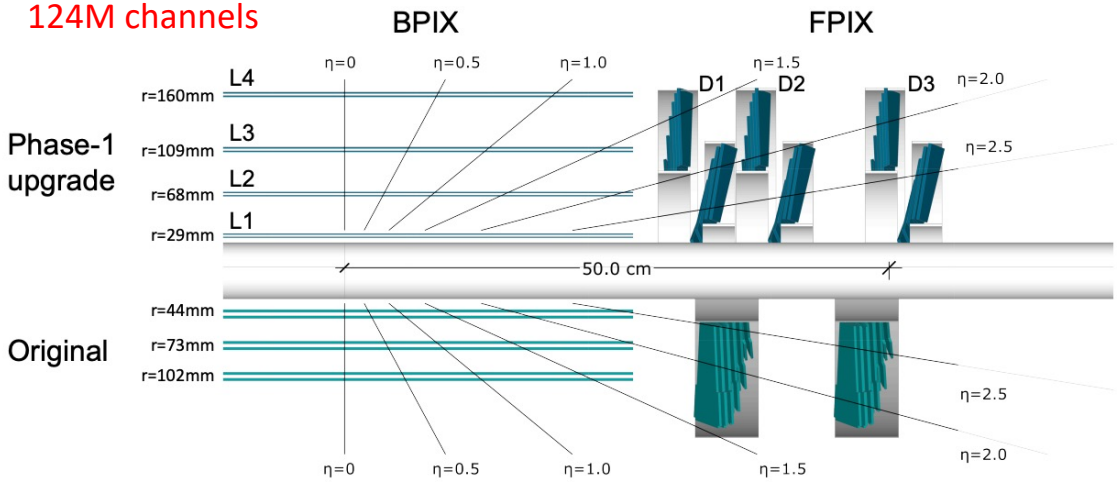
Inner Tracker Phase-1 upgrade

JINST 16 (2021) P02027

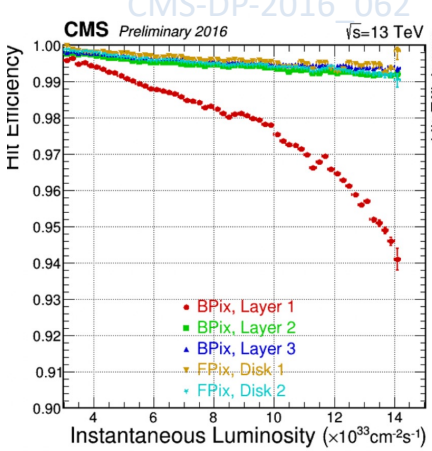
- Phase 1 Si-Pixel tracker upgrade in 2017
  - Additional layer, front-end electronics upgrade



124M channels

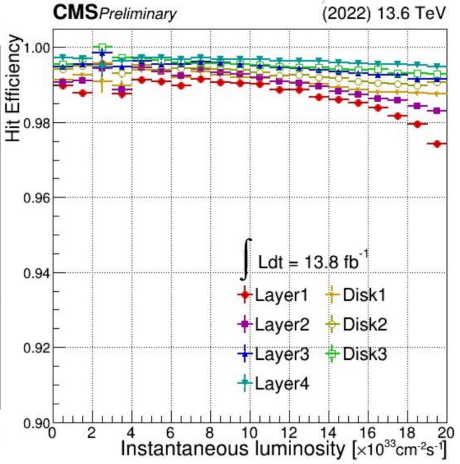


CMS-DP-2016_062

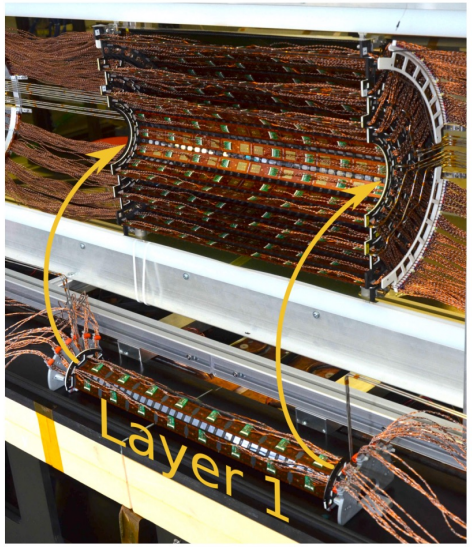


Phase-0 Pixel Tracker Hit efficiency

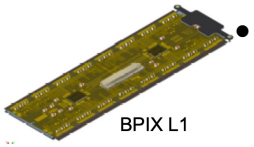
CMS DP -2022/067



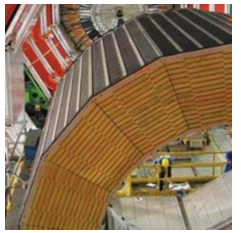
Phase-1 Pixel Tracker hit efficiency (2022)



- Refurbished in LS2:
  - Layer 1 replaced due to radiation ageing (collected 120 fb^{-1} in Run 2)
  - Upgraded on-detector electronics for the first layer:
    - Fixed issues with readout synchronization, noise shielding and radiation resistance

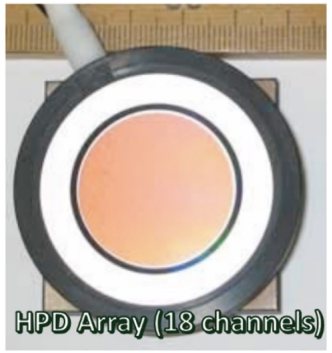


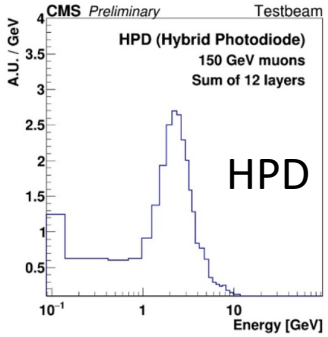
HCAL Phase-1 upgrade

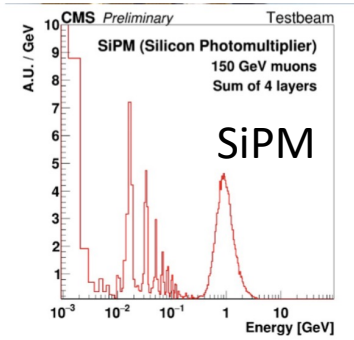


CMS-TDR-010

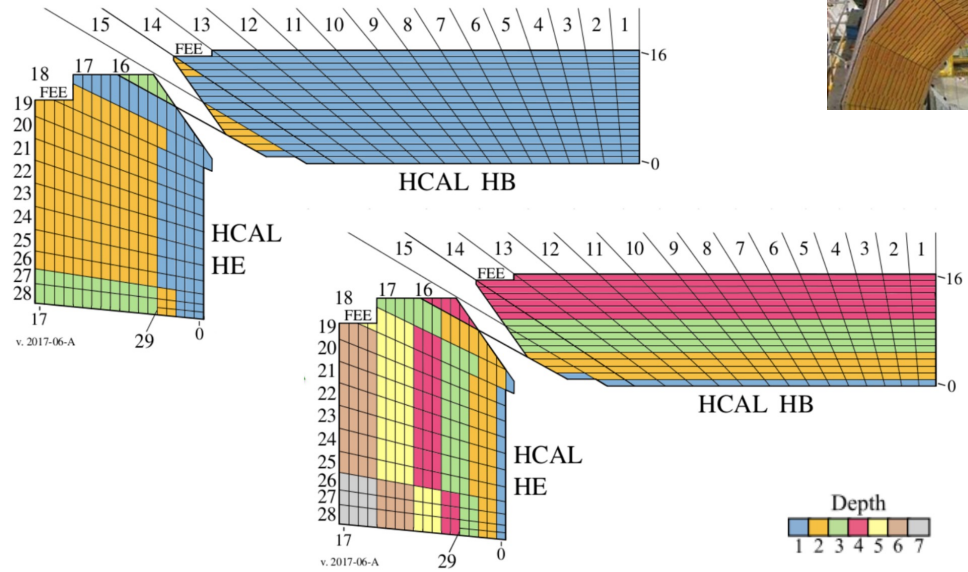
- Upgraded photodetectors and readout electronics
  - Increased longitudinal segmentation 2x to 4x
  - Completed with HCAL Barrel upgrade in LS2
    - Hybrid Photo Diodes (HPDs) → Silicon photomultipliers (SiPM) sensors
    - Improved photon sensitivity, spectrum resolution, high radiation resistance







  - Readout electronics upgrade
    - Increased readout granularity and redundanc
    - Improved data quality to Level-1 trigger



HCAL segmentation in Run 2 (top-left and Run 3 (bottom)

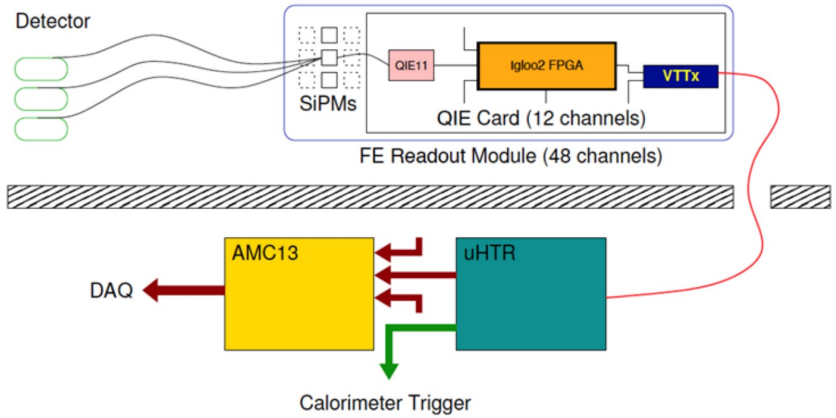
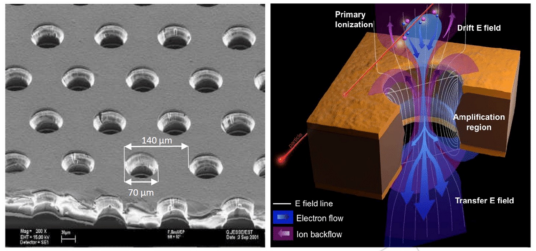
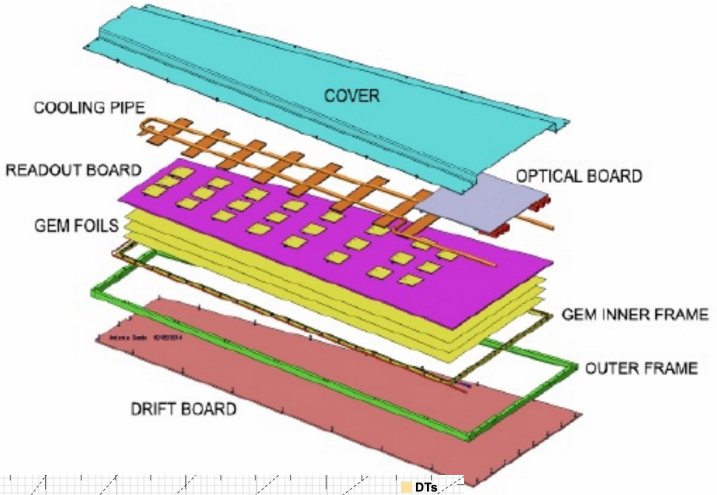


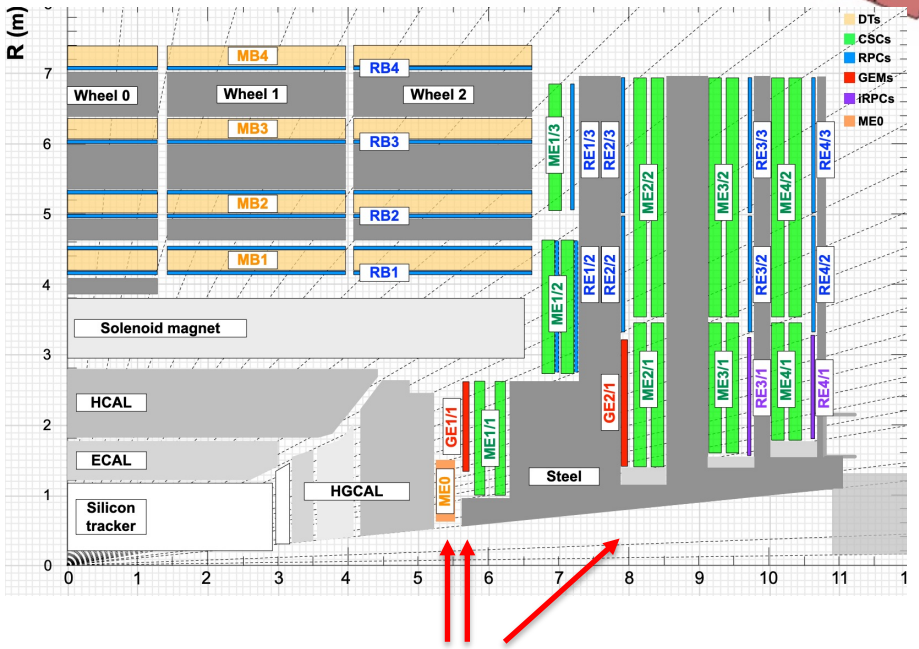
Diagram of the upgraded HB/HE readout chain

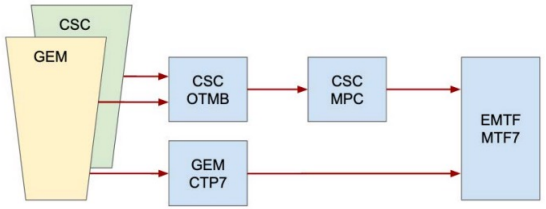
Muon system – Gas Electron Multiplier Chambers (GEM)

- GEM advantages
  - High-rate capability, adequate pattern recognition and radiation tolerance
- Staged Phase-2 upgrade
- GE/1 ring in endcap was installed in LS2
  - First of the upgrades on the road to Phase-2
    - Detector electronics ready for high-rate and radiation conditions
  - Integration with the Muon L1-Trigger



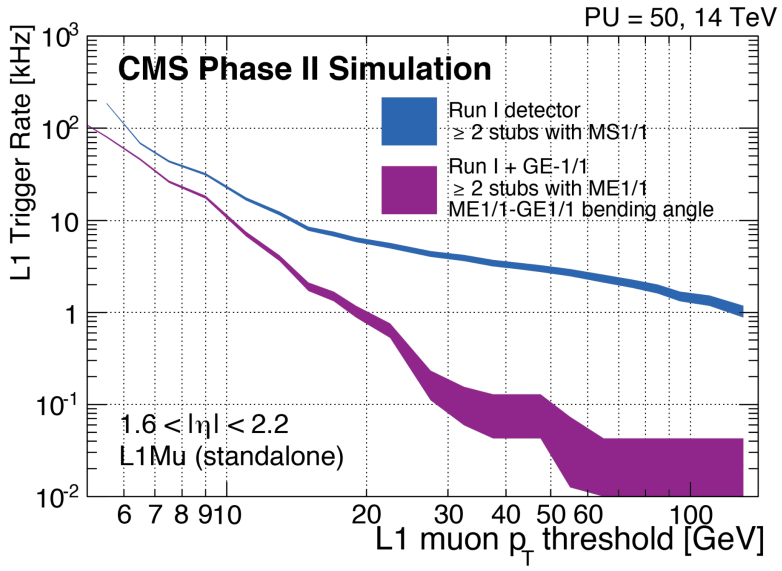






L1-Trigger Endcap muon track-finder

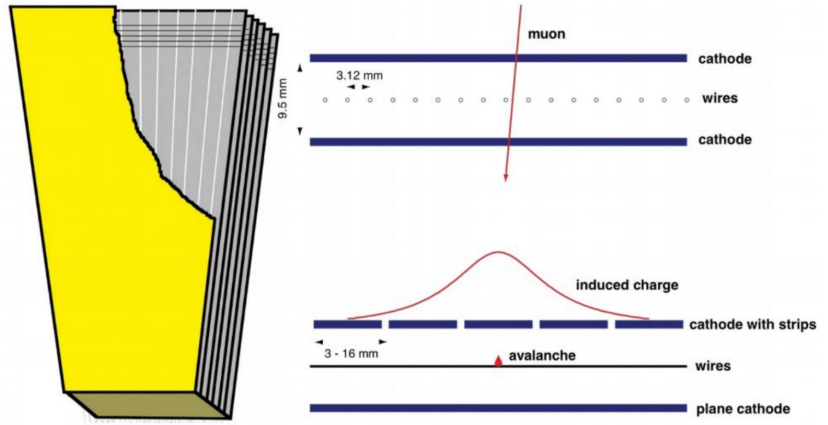
GEM Phase-2 detectors

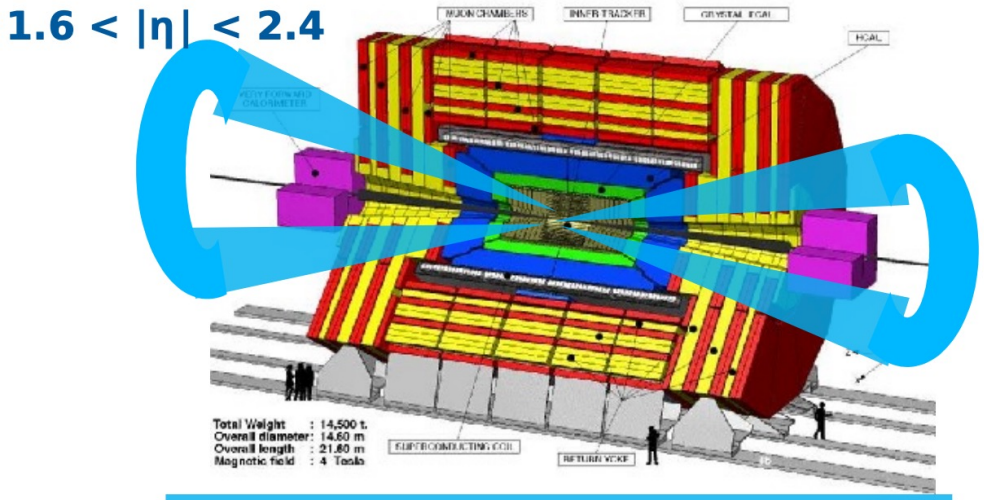


CMS-TDR-013

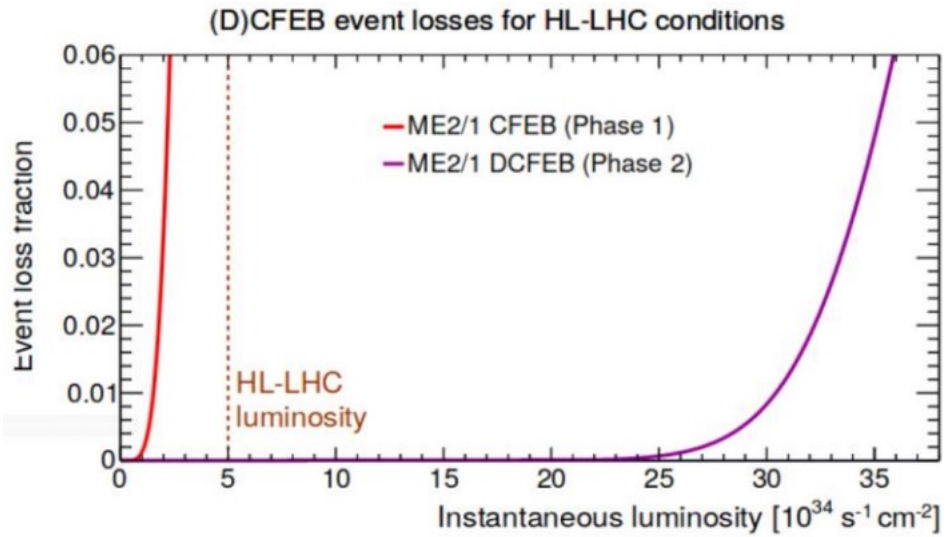
Muon system – Cathode Strip Chamber (CSC)

- Detector electronics replaced for the inner CSC rings
  - Ready for Phase-2 (opportunity in LS2)
    - Latency (12.5 μs) and L1 rate (750 kHz)
    - Improved radiation hardness / tolerance
  - Upgraded for 180 chambers



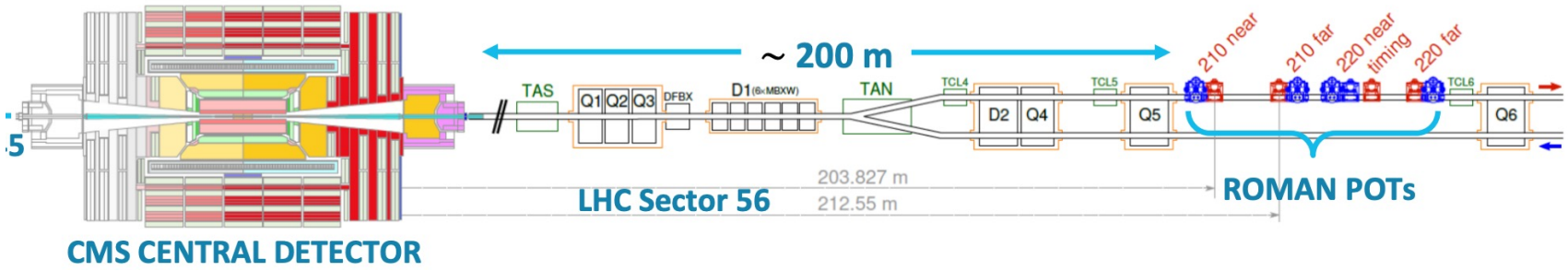


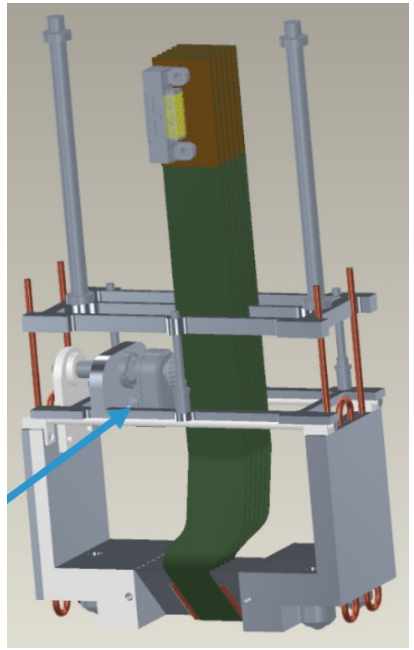
Approximate angular region of the inner rings:
ME1/1, ME2/1, ME3/1, ME4/1
= 180/540 chambers



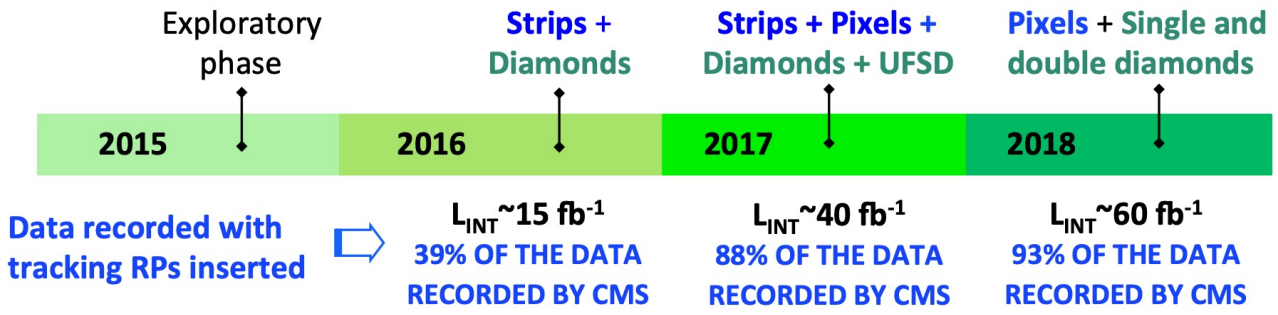
CMS-TDR-016

Precision Proton Spectrometer (PPS) in Run 3





- Designed for measuring scattered protons in forward regions of CMS in standard LHC running
  - pp -> pXp processes (X = high ET jets, gg, ZZ, W…)
  - Tracking detectors: LHC magnets used to measure proton momentum
  - Timing detectors (disentangling pileup)



- Run 3:
  - New tracker system in Run 3:
    - 3D silicon pixel sensors - movable to distribute radiation damage
  - Timing system
    - upgraded electronics and double diamond sensors (< 30ps resolution)

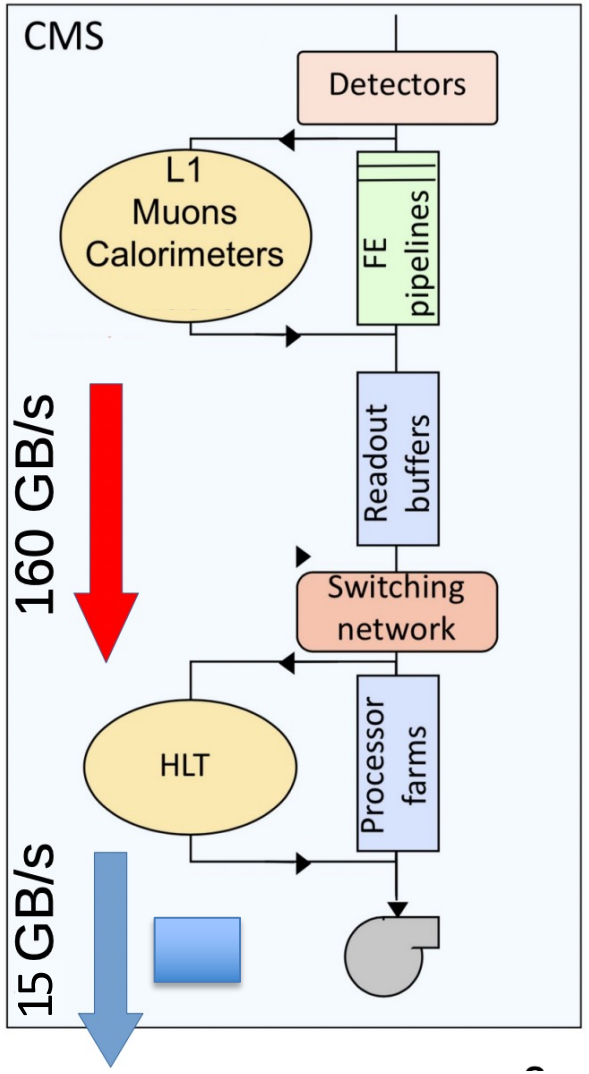
Run 3 requirements for Trigger and Data Acquisition (DAQ)

- Requirements similar to Run 2

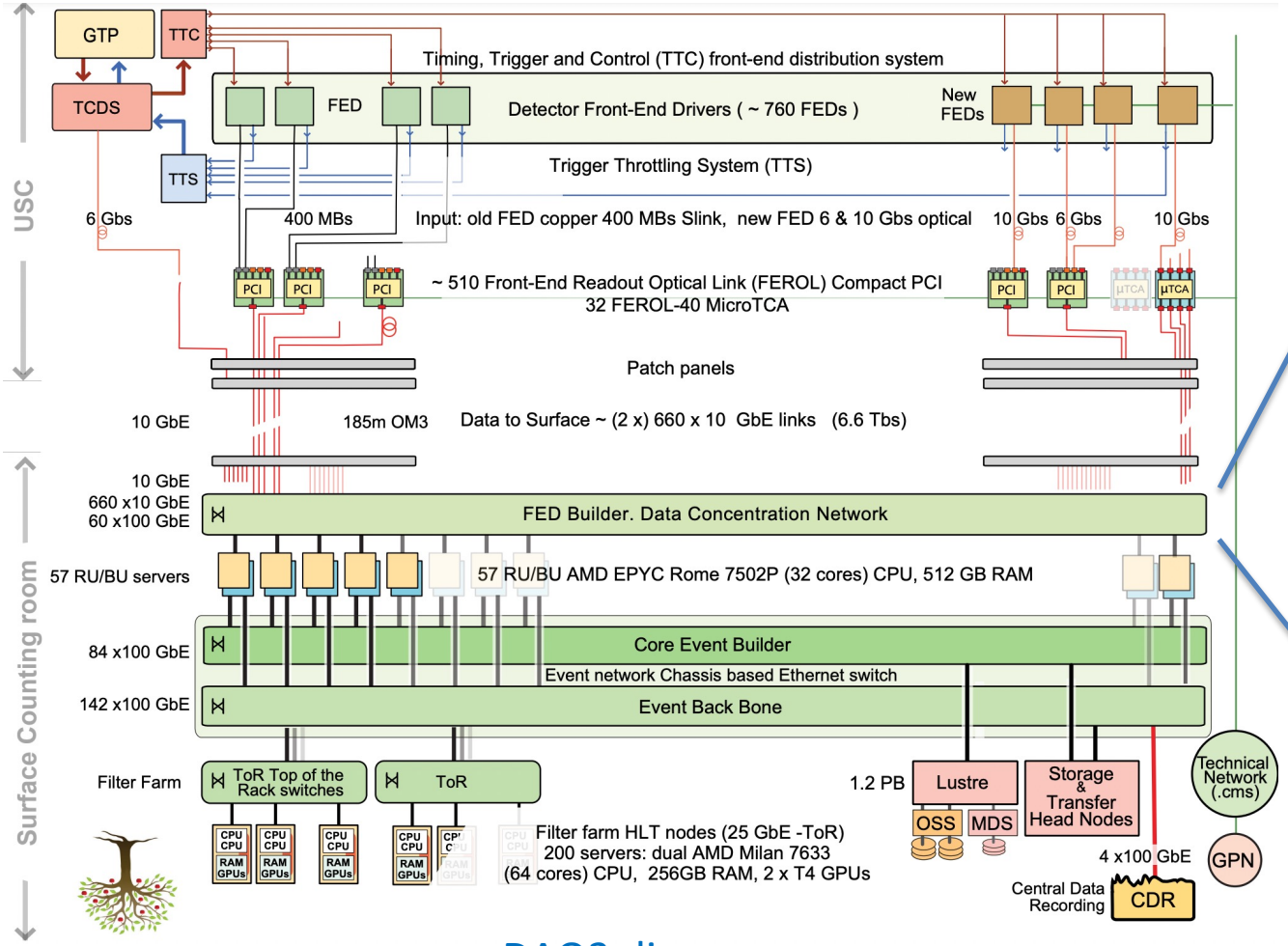
	Run 1	Run 2	Run 3
L1 accept rate	100 kHz	100 kHz	100 kHz or above
Event size *	1 MB	1.4 MB	1.6 MB **
HLT output rate (pp)	up to 1 kHz	O(1) kHz	O(5) kHz prompt and parking
HLT computing power (MHS06)	~0.21	~0.72	0.65 (+ GPUs)
HLT output bandwidth	2 GB/s	3 - 6 GB/s	8 (pp) – 15 (HI) GB/s

* raw, uncompressed size, estimate at pileup 56
**Increased mostly due to upgraded HCAL and GEM in Run 3

– Increased trigger-accept rate for heavy-ion (PbPb) run



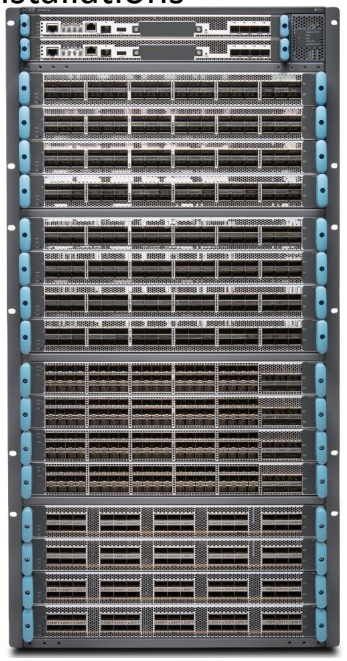
DAQ Data Concentrator in Run 3



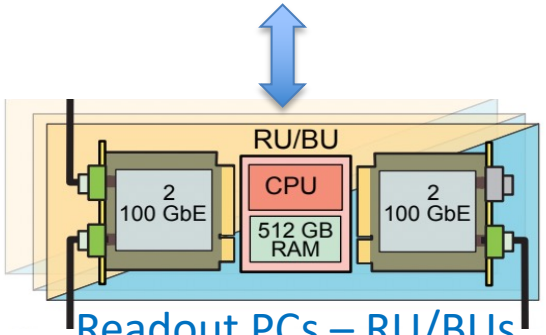
DAQ3 diagram

readout of L1-Trigger accepted (100 kHz) event *fragments* from detector electronics (*sources*)

- 10 → 100 Gbit network to transport data to the surface installations



- monolithic Juniper QFX10016 chassis switch up to **100 Gb/s Ethernet** link speed
  - Replaces 40 Gbit Ethernet (multi-switch) network fro, Rim 2

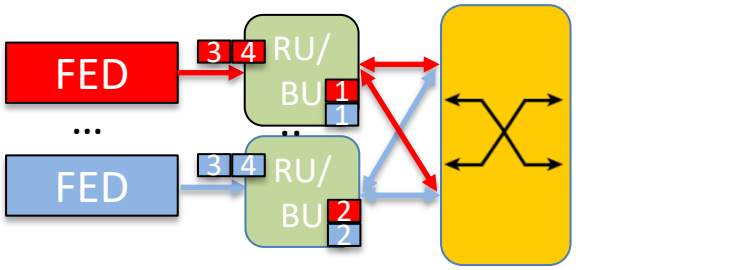


Readout PCs – RU/BUs

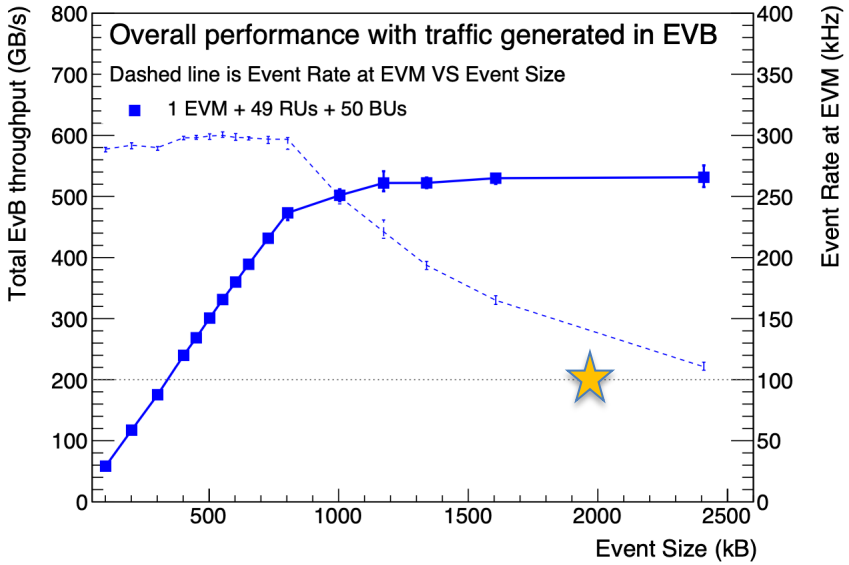
Zen2 AMD 'Rome' architecture (32 core)
3 x 100 Gb Ethernet
(2x Mellanox ConnectX-6)

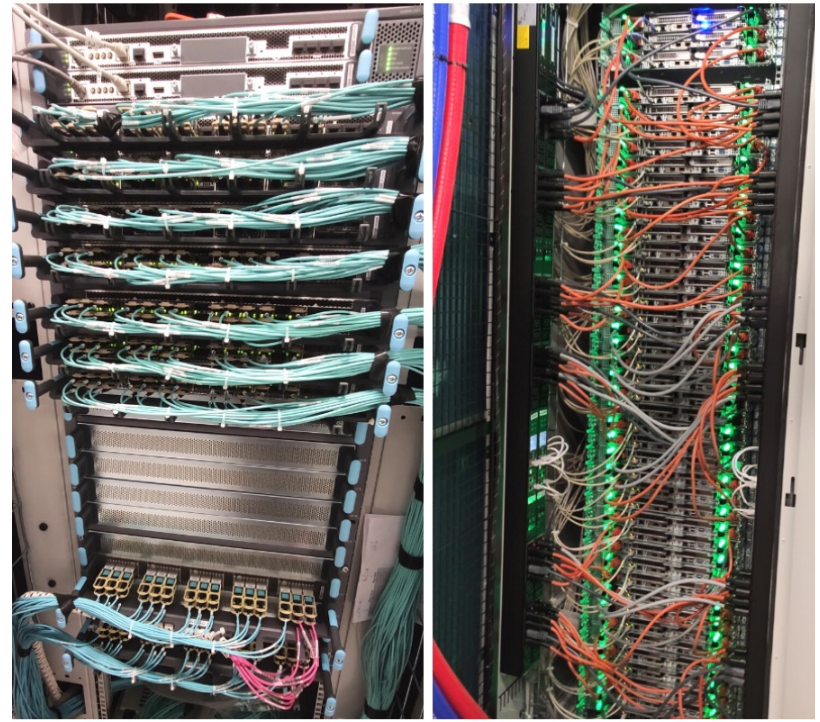
DAQ Event Building in Run 3

- Building complete events from scattered readout
- Run 3: 100 Gbit Ethernet with remote DMA (RoCE v2) – replaces Infiniband network (Run 2)
  - Using 2nd chassis based switch (Juniper) → serving also HLT and Storage and Transfer System (Event Backbone Network)
- Folded event-building:
  - Advantage - bidirectional use of network links
  - More demanding on individual PC CPU and I/O performance



Folded EVB illustration



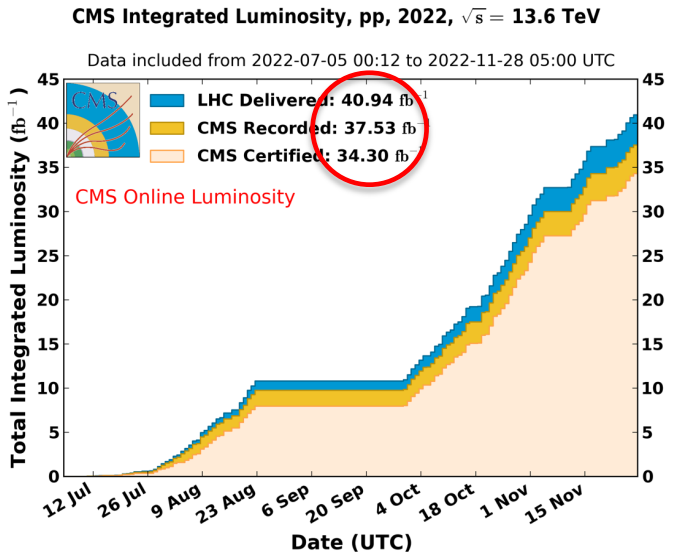


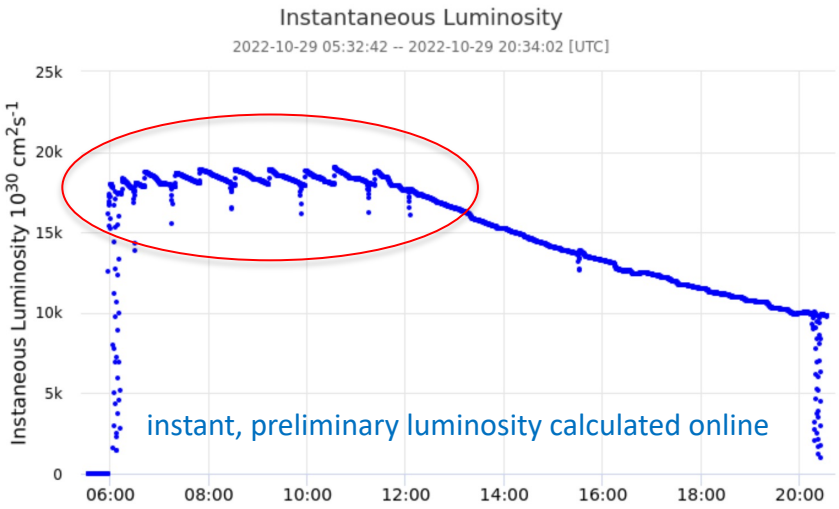
Fully cabled Run 3 EvB and Event Backbone switch

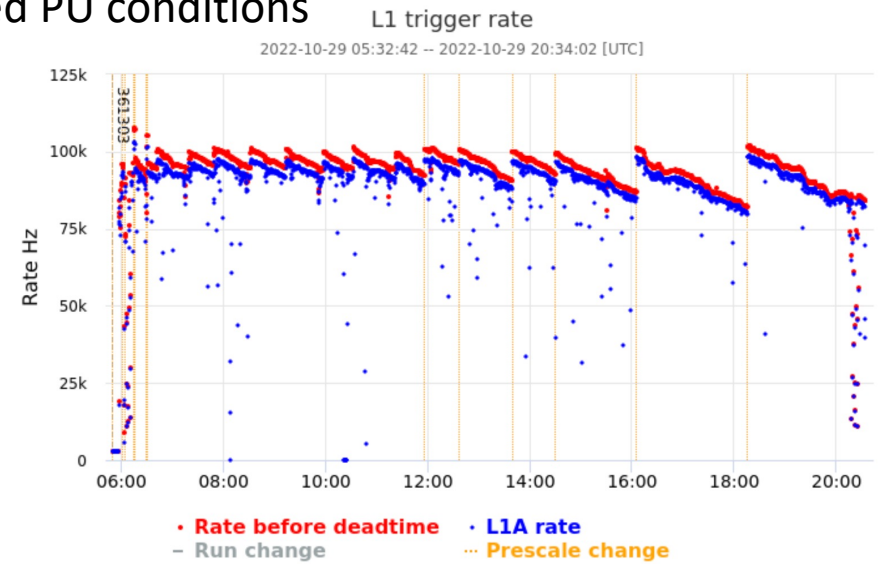
L1-Trigger and HLT → see Dener S. Lemos talk

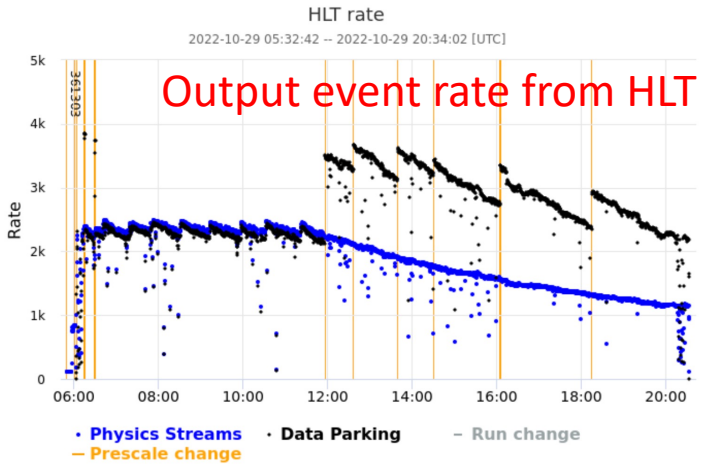
Operations in 2022

- July 5th to November 28th – Run 3 proton-proton physics run at sqrt(s)=13.6 TeV
  - 41 fb^{-1} delivered by LHC
  - Various short special runs, including high-pileup and heavy ion (PbPb) test run
- Efficient operation of CMS
  - 37.5 fb^{-1} recorded by CMS - 91.7% efficiency
  - After the initial, commissioning phase, good data quality (~ 95%)
- LHC beam parameters controlled to level luminosity at constant rate for several hours in every proton-proton fill
  - Mitigates LHC heat load limitations at higher intensity
  - Better for detectors due to controlled PU conditions



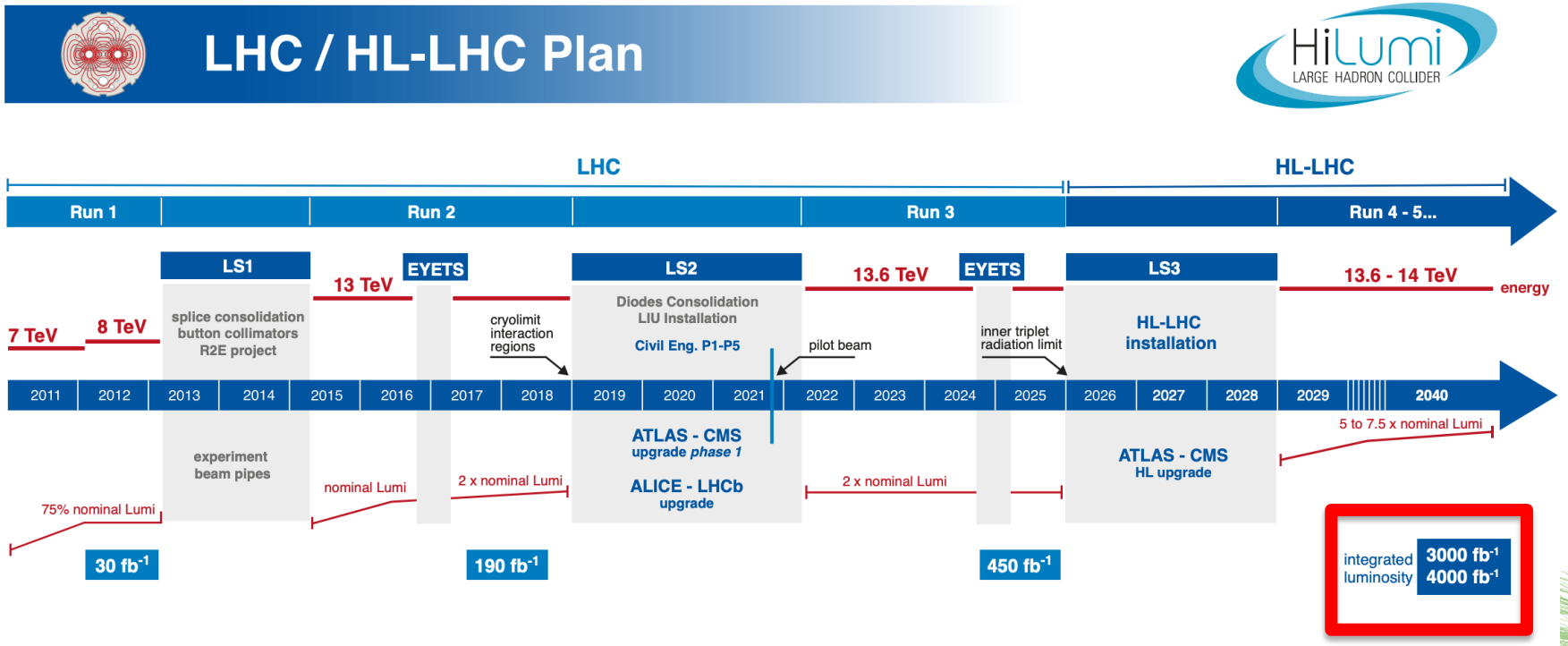


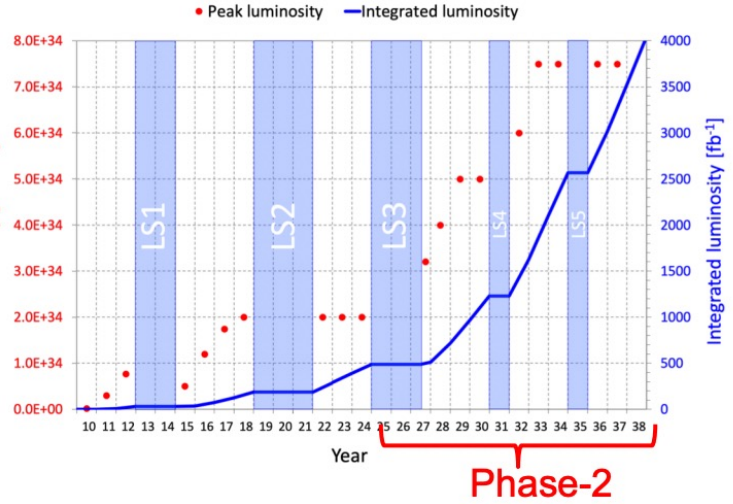




Plots from the Online Monitoring system (OMS)

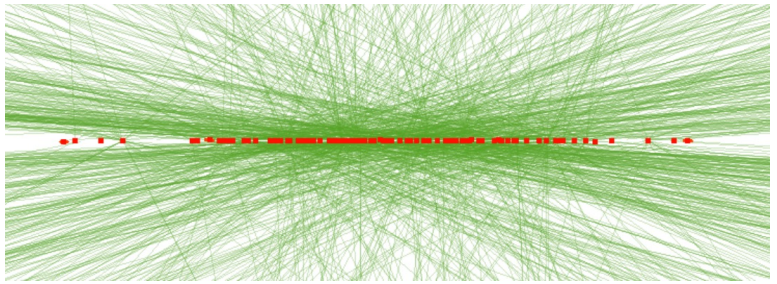
High-Luminosity LHC (Phase-2)





Aim: collection of large amount of collision data, up to 4000 fb^{-1}

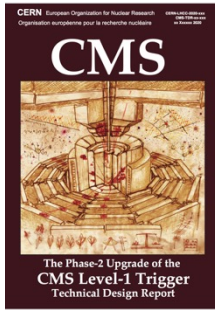
- Physics motivation: SM, Higgs coupling, SUSY, DM searches…
- New paradigms for HEP experiments to fully exploit HL-LHC luminosity
  - Upgraded as well as fully new designed detectors to cope with increased occupancy and radiation
  - detector electronics, Trigger and DAQ upgrades to handle increased event size and L1-trigger rate



140 PU event display

	$\mathcal{L}$	$\langle$ PU $\rangle$	Vertex Density	$\int \mathcal{L}$ / year
Baseline	$5 \cdot 10^{34}\,\mathrm{cm}^{-2}\,\mathrm{s}^{-1}$	140	0.8/ mm	250 fb^{-1}
Ultimate	$7.5 \cdot 10^{34}\,\mathrm{cm}^{-2}\,\mathrm{s}^{-1}$	200	1.2/ mm	> 300 fb^{-1}

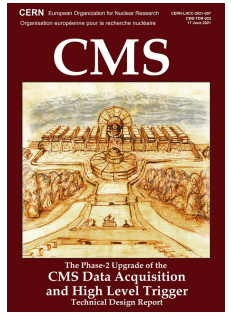
CMS HL-LHC Upgrades



L1-Trigger

https://cds.cern.ch/record/2714892

- Tracks in L1-Trigger at 40 MHz
- Particle Flow selection
- 750 kHz L1 output
- 40 MHz data scouting



DAQ & High-Level Trigger

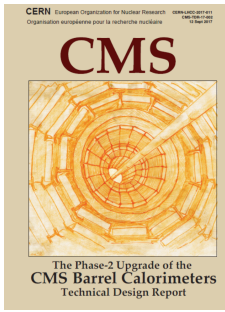
https://cds.cern.ch/record/2759072

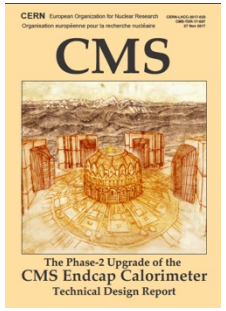
- Full optical readout
- Heterogenous architecture
- 60 TB/s event network
- 7.5 kHz HLT output

Barrel Calorimeters

https://cds.cern.ch/record/2283187

- ECAL single crystal granularity at L1 trigger with precise timing for e/γ at 30 GeV
- ECAL and HCAL new Back-End boards

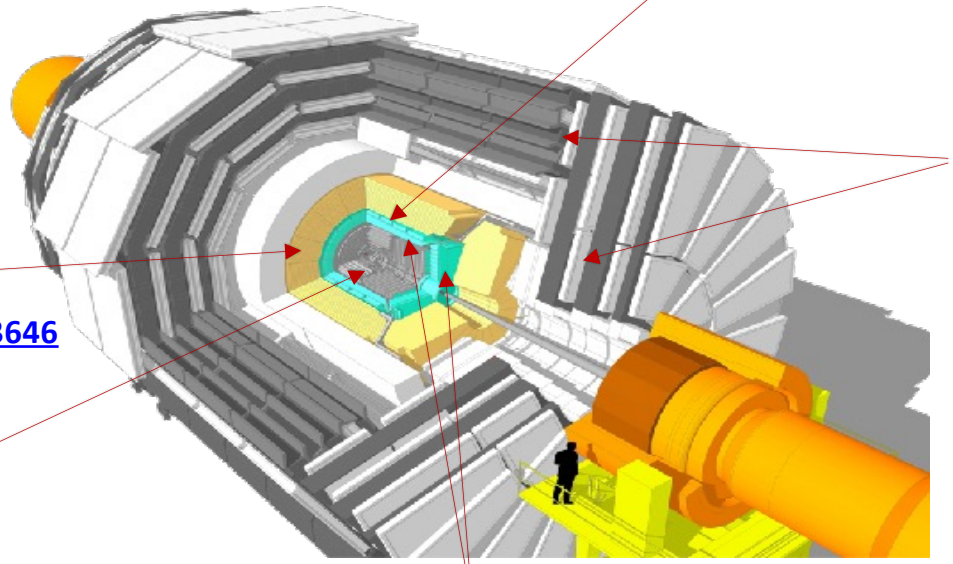




Calorimeter Endcap

https://cds.cern.ch/record/2293646

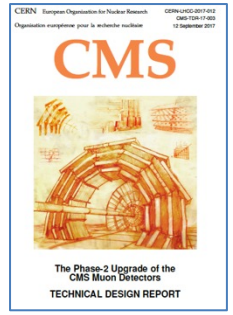
- 3D showers and precise timing
- Si, Scint+SiPM in Pb/W-SS

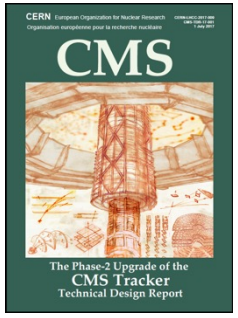


Muon systems

https://cds.cern.ch/record/2283189

- DT & CSC new FE/BE readout
- RPC back-end electronics
- New GEM/RPC $1.6 < \eta < 2.4$
- Extended coverage to $\eta \simeq 3$

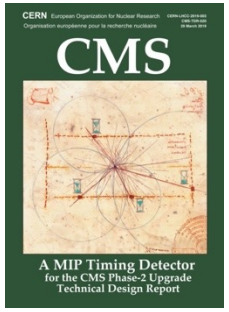




Tracker

https://cds.cern.ch/record/2272264

- Si-Strip and Pixels increased granularity
- Design for tracking in L1-Trigger
- Extended coverage to $\eta \simeq 3.8$



MIP Timing Detector

https://cds.cern.ch/record/2667167

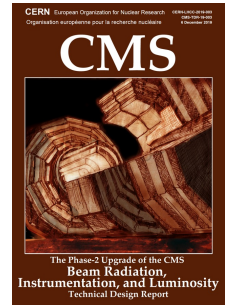
Precision timing with:

- Barrel layer: Crystals + SiPMs
- Endcap layer:
  Low Gain Avalanche Diodes

Beam Radiation Instr. and Luminosity

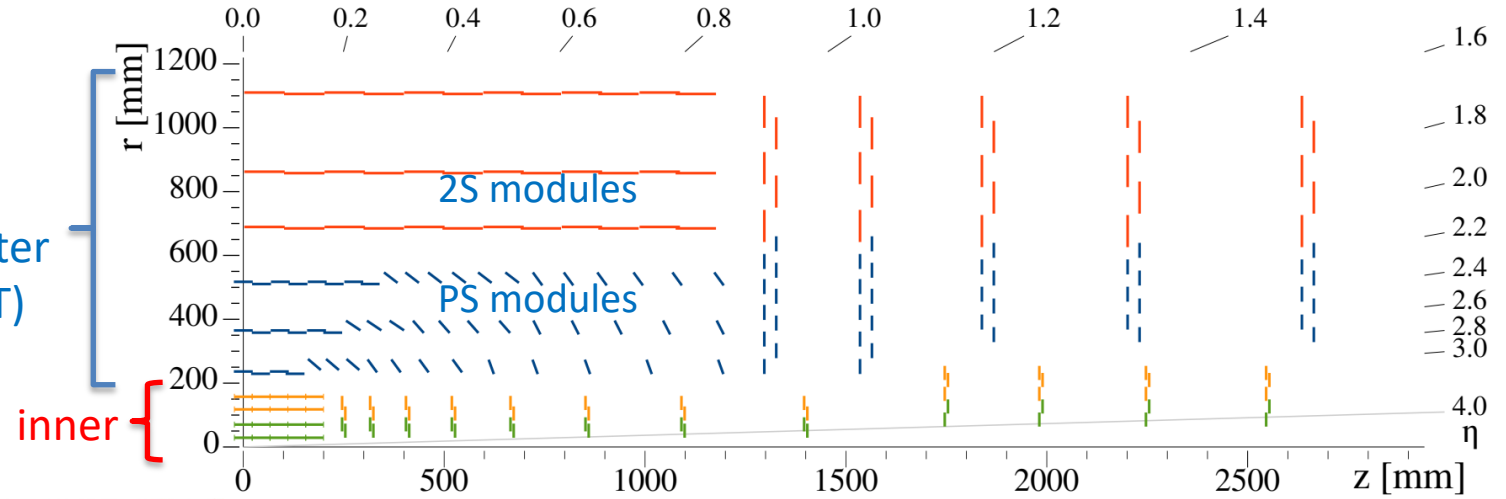
http://cds.cern.ch/record/2759074

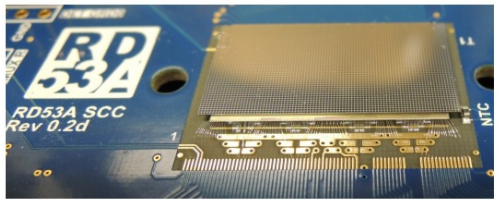
- Beam abort & timing
- Beam-induced background
- Bunch-by-bunch luminosity:
  1% offline, 2% online
- Neutron and mixed-field radiation monitors



Inner and Outer Tracker Phase-2 upgrade

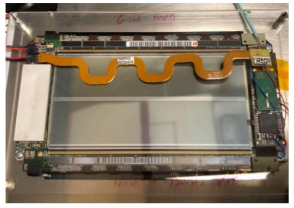
- Full replacement of inner and outer tracker detector
- Due to increased particle density with high PU → higher detector granularity
  - 20x more channels than in Phase-1
- Improved radiation hardness and reduced material budget
- Coverage increased to $\eta = 4$

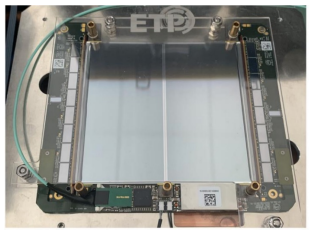




RD53A prototype Pixel chip
- CMS will use 25x100μm
(3D or planar sensors)

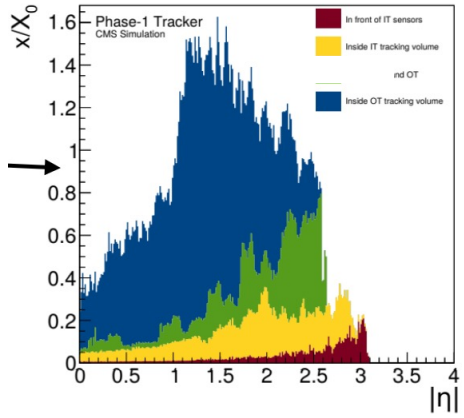
OT PS (strip-strip) module – 2.5cm x 100μm strips & 1.5mm x 100 μm MacroPixels



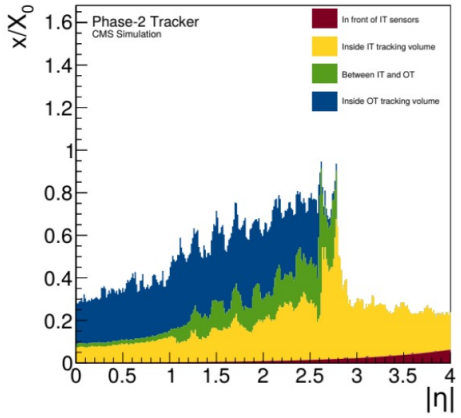


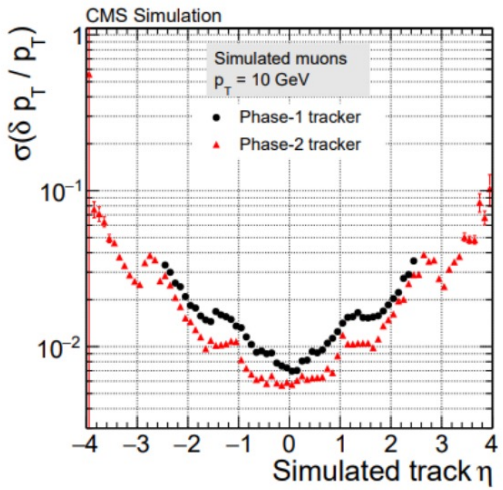
Outer Tracker 2S (strip-strip) module prototype – 5 cm x 90μm strips

Phase 1 material budget

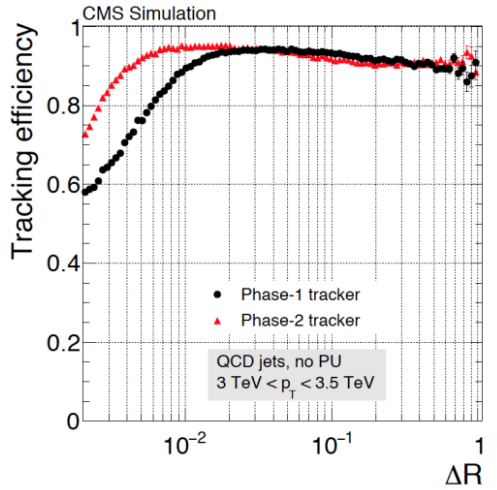


Phase-2 material budget



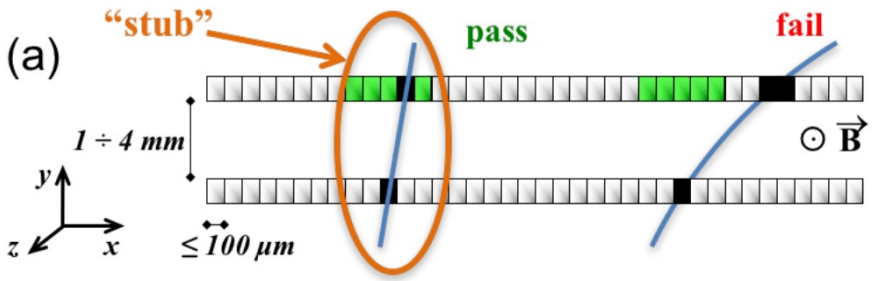


Simulated muon p_T resolution

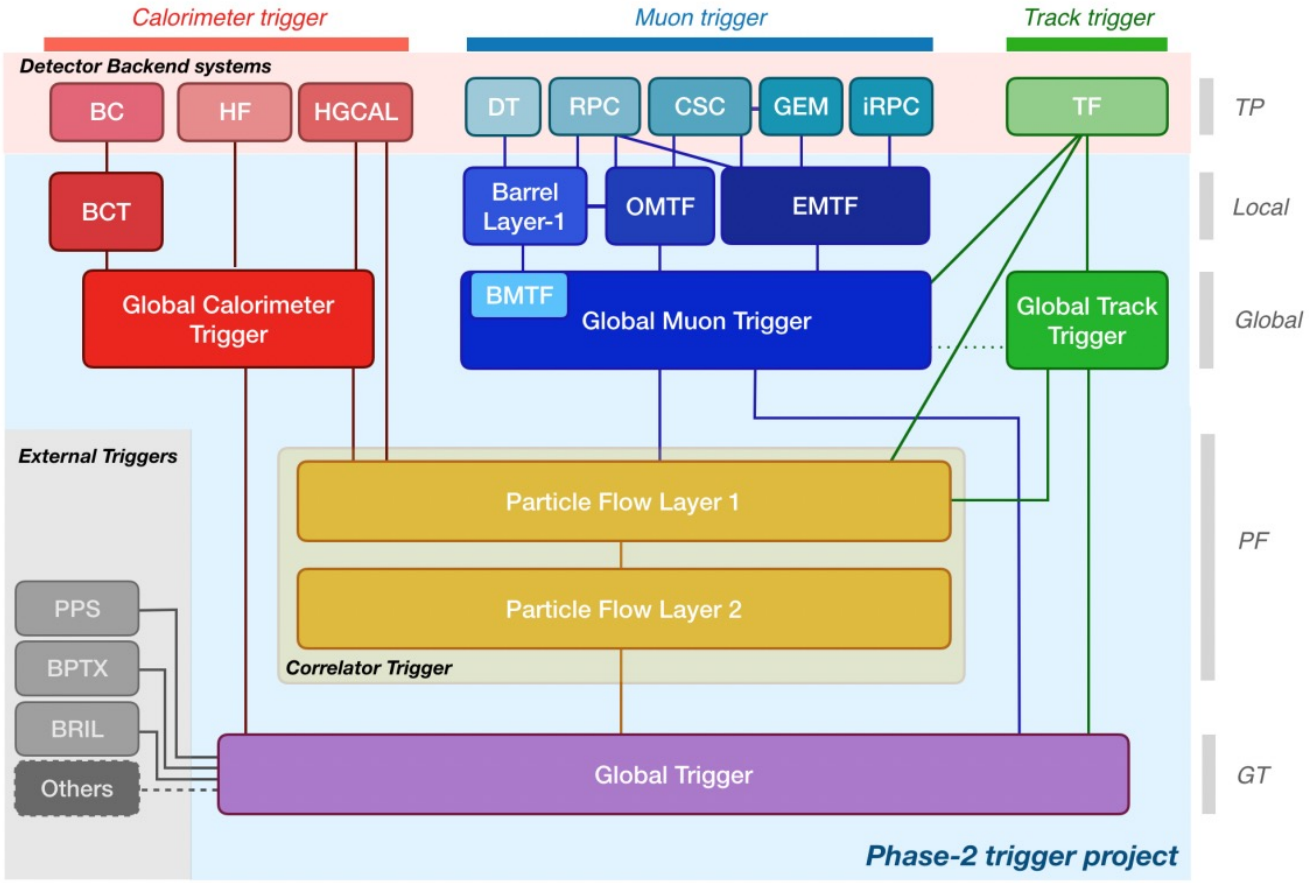


Level-1 Trigger Phase-2 upgrade

- 750 kHz L1-accept (readout) rate from the detector from the 40 MHz collision rate
- 12 μs latency budget for decision on event accept or discard
- New **Track-trigger for the Outer tracker (OT)**
  - Separation of high and low transverse momentum charged tracks, p_T > 2 GeV, in Tracker pT modules (trig. primitive filtering in PS and 2S)
  - FPGA-based parallelized track-finding: O(15k) stubs per PU 200 event



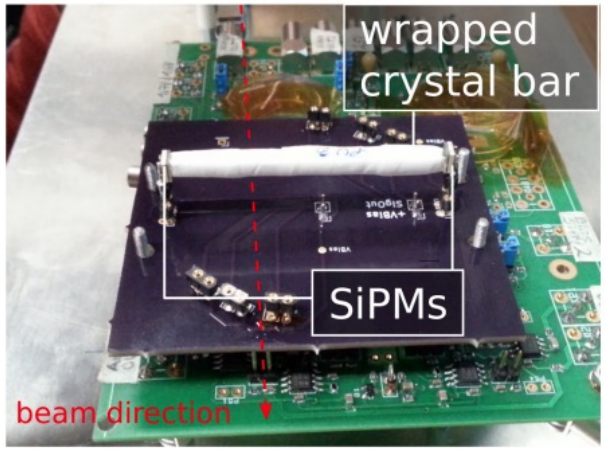
- High-precision calorimetry, vertex finding, displaced muons
- Particle flow in Correlator Trigger
  - application of an algorithm building a global event description in firmware

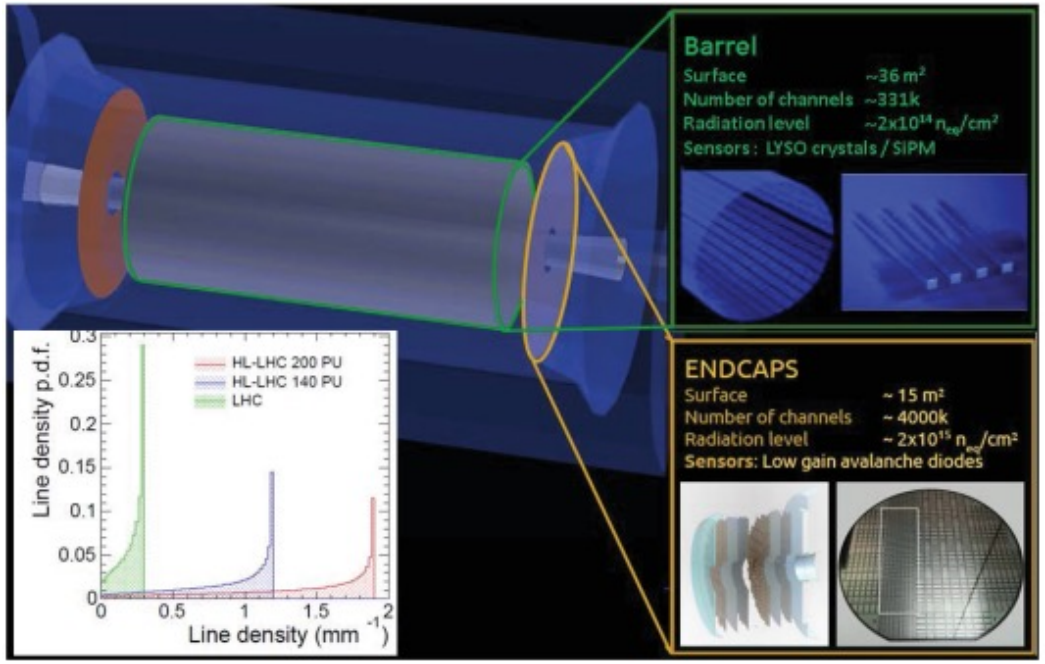


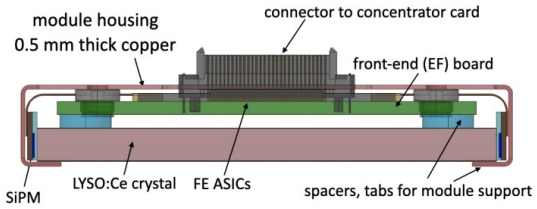
- 40 MHz L1-scouting system
  - Readout of trigger-primitives and trigger data, storing for analysis
  - Demonstration system in Run 3

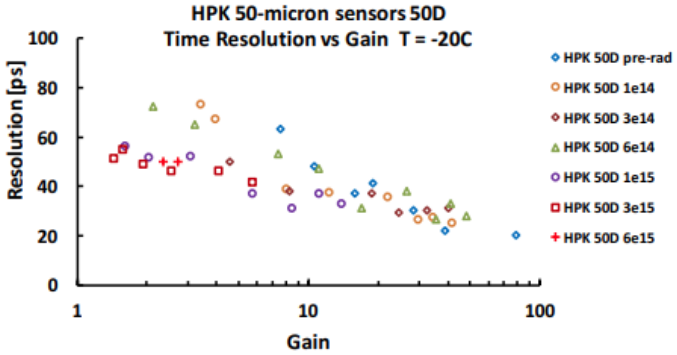
MIP Timing Detector - MTD

- Timing detector located between tracker and calorimeters

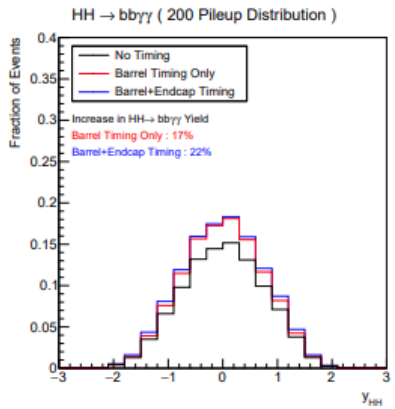


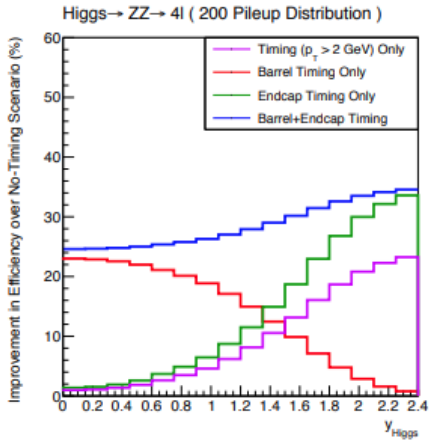






Timing resolution ~ 30 ps

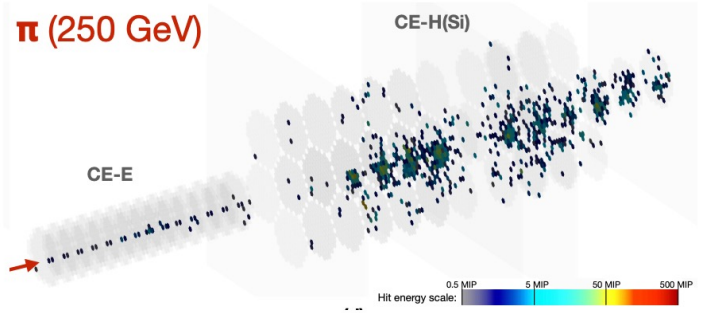




- Completely new capability to CMS
  - precise measurement of production time of MIP particles →disentangle PU interactions (beam-spot spread ~ 180 ps)
  - Improved performance of b-jet indentification and missing p_T measurement
  - Enhanced capability of searches for long-lived particles

High Granularity Calorimeter - HGCAL

- Radiation-tolerant imaging **Endcap** calorimeter
  - Replaces existing Pre-shower, ECAL and HCAL (due to end-of-life radiation, trigger latency…)
  - *5D* (spatial, energy, precision-timing) measurement of shower shape
  - participates in triggering
  - 6.5 million channels : silicon cells (0.5-1cm2) and tiles (4-32 cm) – readout by SiPMs



8'' Si module tile

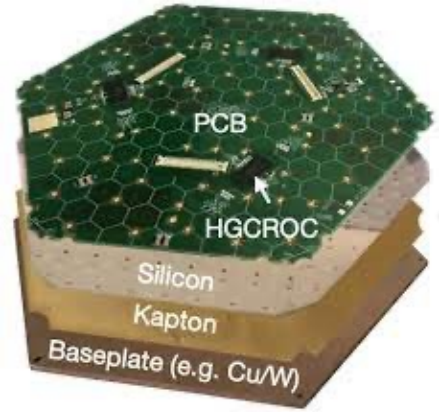
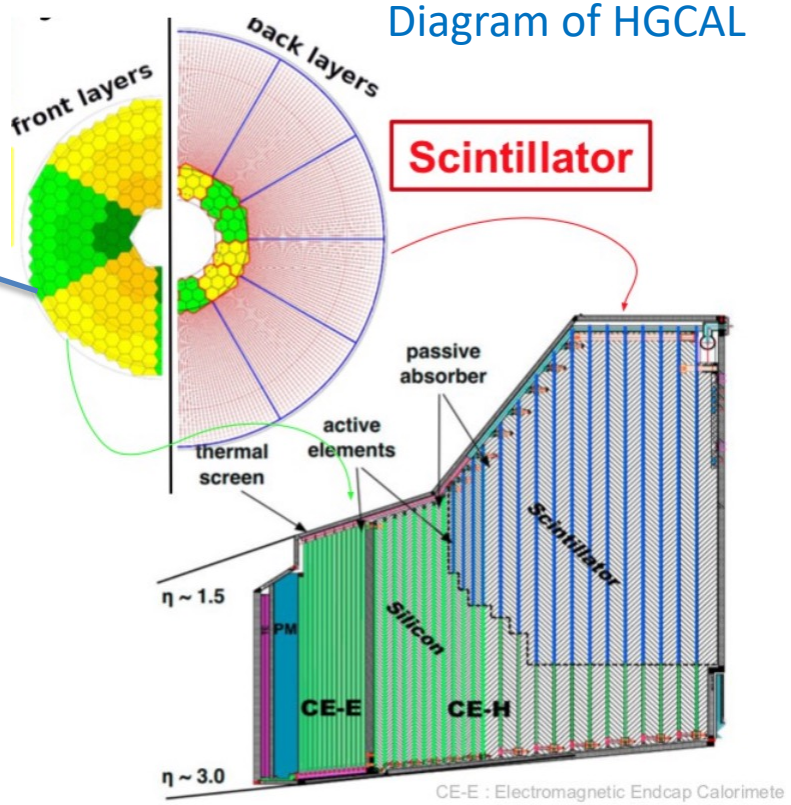
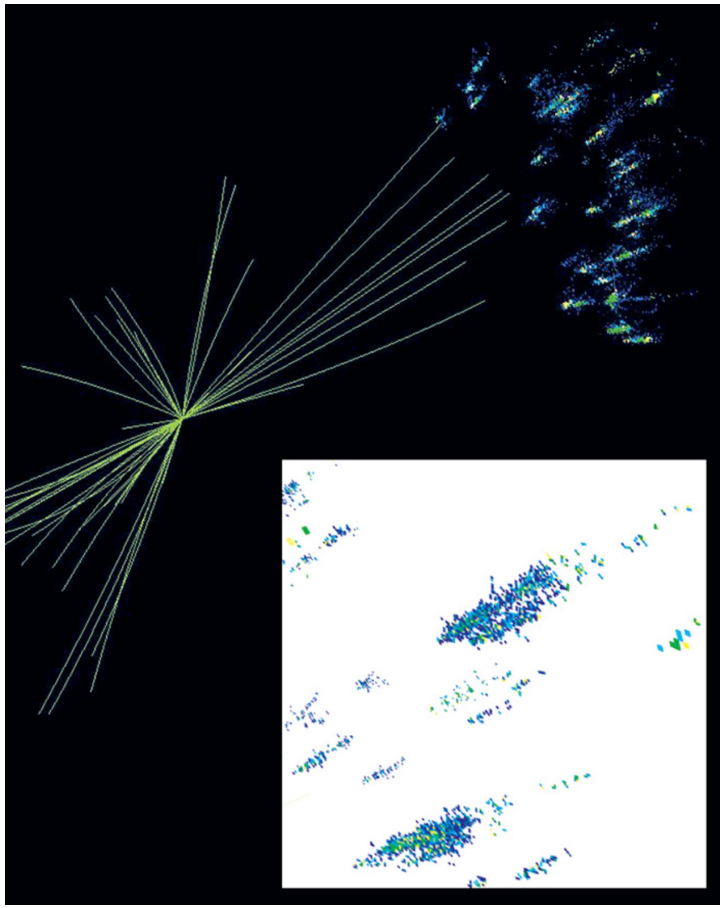


Diagram of HGCAL

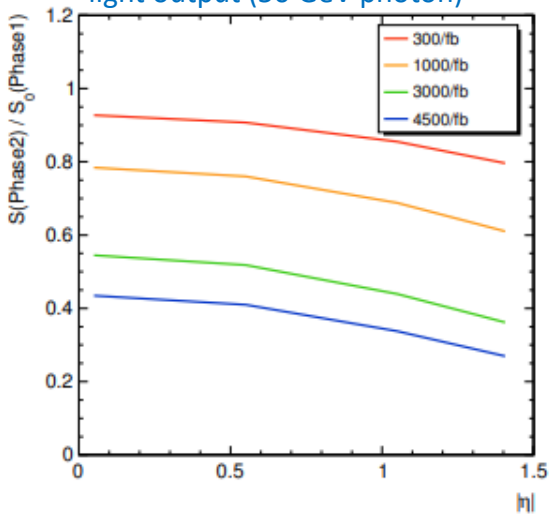




ECAL Barrel Phase-2 upgrade

- Refurbishment of existing barrel supermodules
  - PbWO3 crystals – long lifetime in radiation conditions
- New on-detector electronics and readout
  - Single-crystal trigger-primitive generation
  - Increased buffers for larger 12μs L1-accept latency and 750 kHz readout rate
  - improved timing and noise filtering, and re-optimized APD pulse shaping

Expected fraction of initial crystal light output (50 GeV photon)



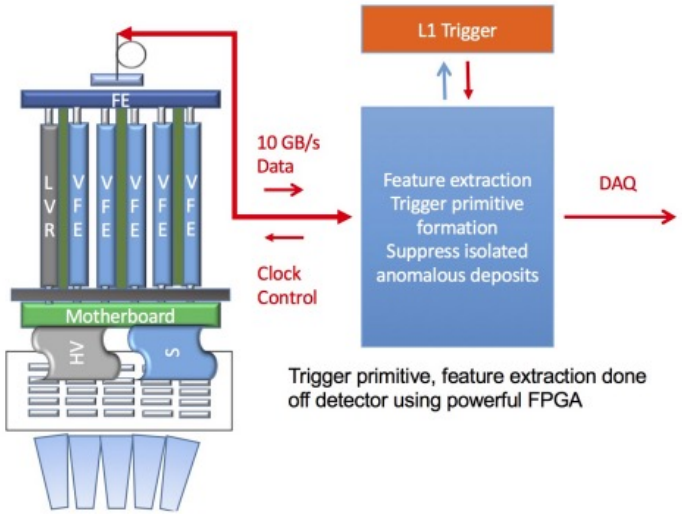


Figure 1.4: Block diagram describing the upgrade EB electronics architecture.

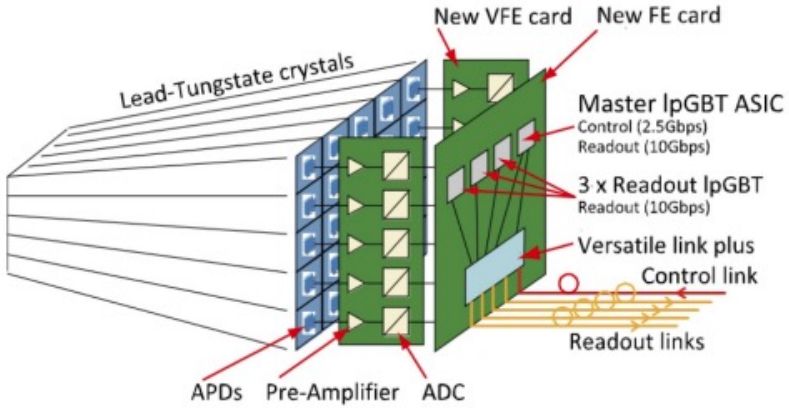
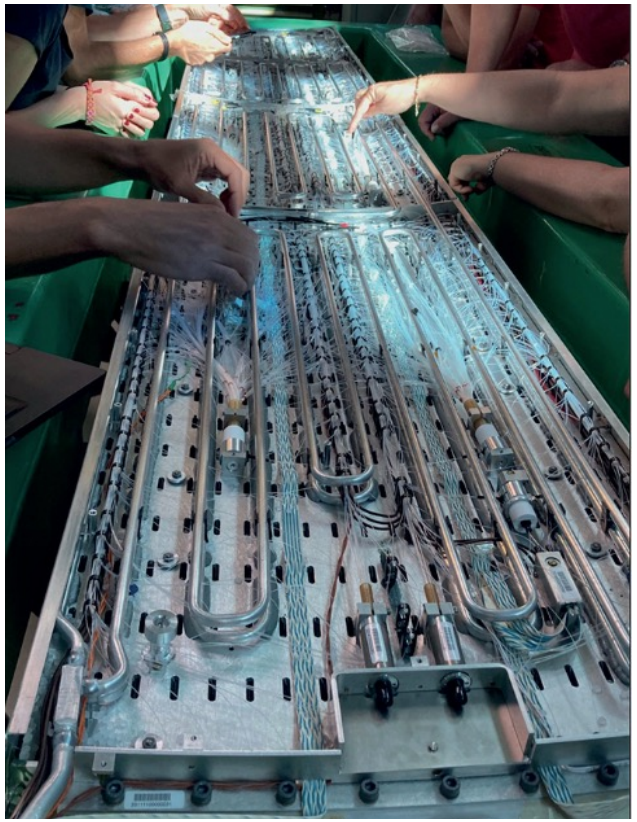


Figure 1.5: Schematic of the upgrade EB electronics architecture.

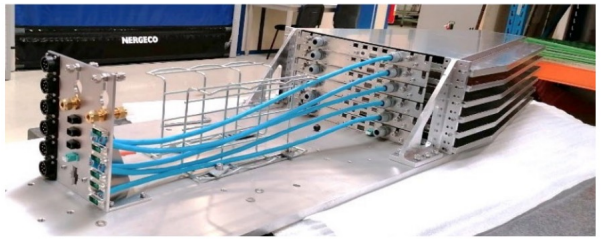
Refurbished spare ECAL barrel supermodule



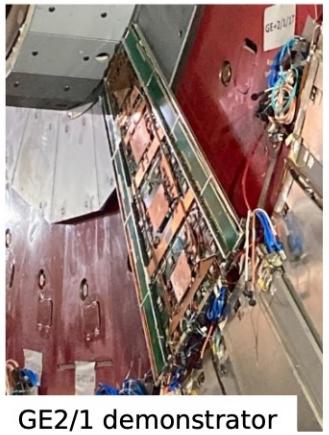
November 2022 electron test beam
Pedestal mean, RMS and pulse shape as expected and consistent over all channels

Muon system Phase-2 upgrades

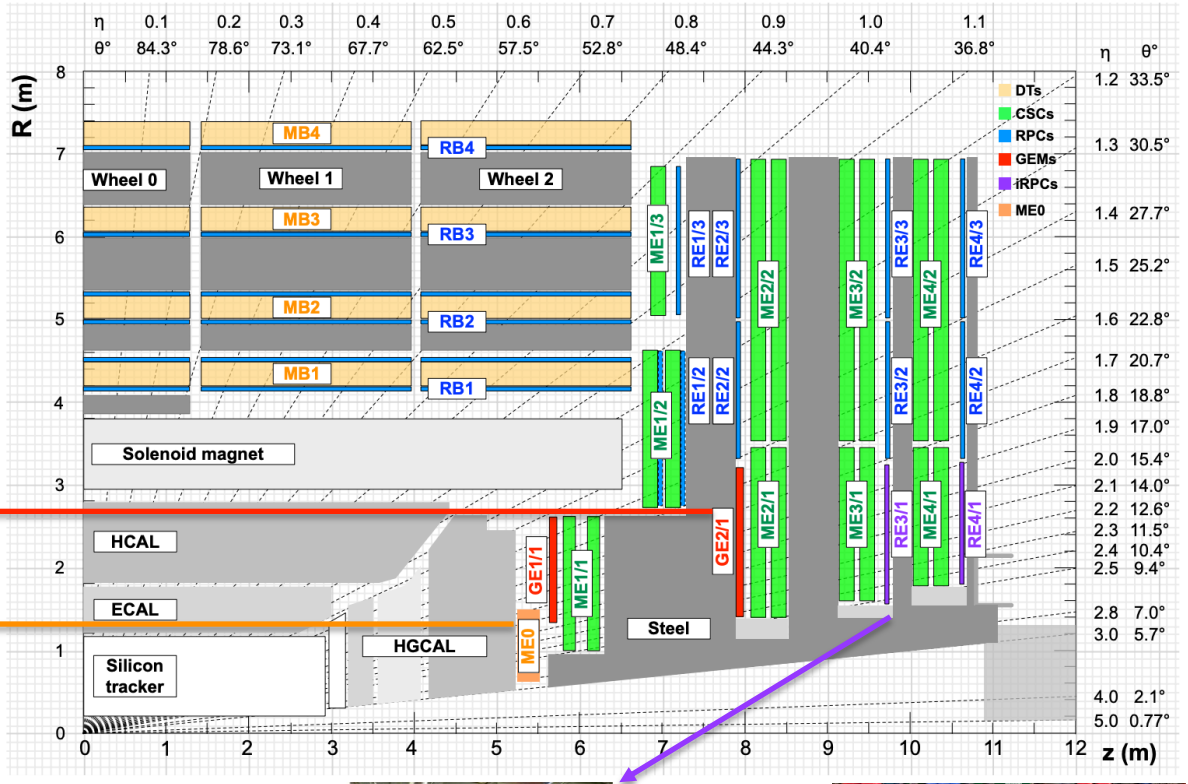
- Existing DT, CSC, RPC detector upgrade electronics
  - For HL-LHC high-rate readout, latency requirements and improved performance
  - the DT readout electronics upgrade (ongoing slice test) – improved time resolution
  - the legacy RPC system readout electronics upgrade - adding better time estimation, providing a sub-bunch-crossing time window of 1.56 ns
- New detectors in the forward region
  - Improved geometrical acceptance for muons
- New GEM detectors
  - ME0 – 6-layer triple-GEM detectos at $|\eta| = 2.4 - 2.8$
    Capable of high hit rate



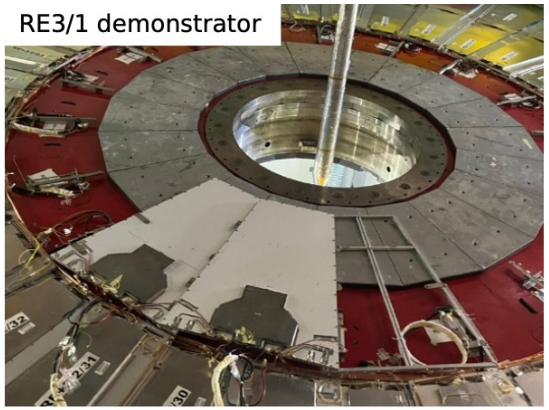
  - GE 1/1 – 2 layers installed in LS2
  - GE 2/2 - demonstrator installed in LS2

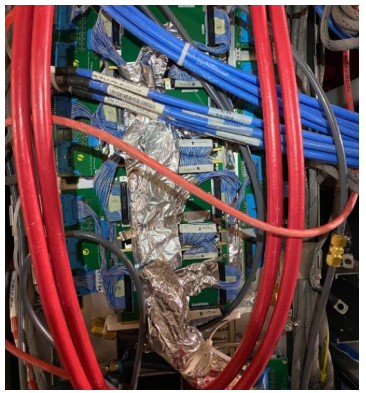


GE2/1 demonstrator



RE3/1 demonstrator



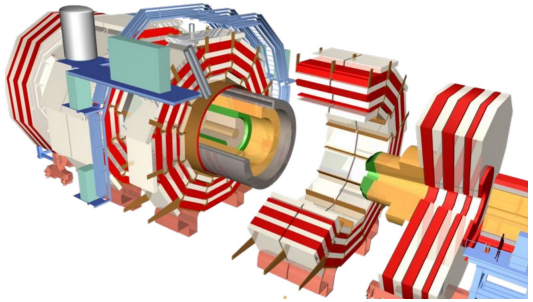


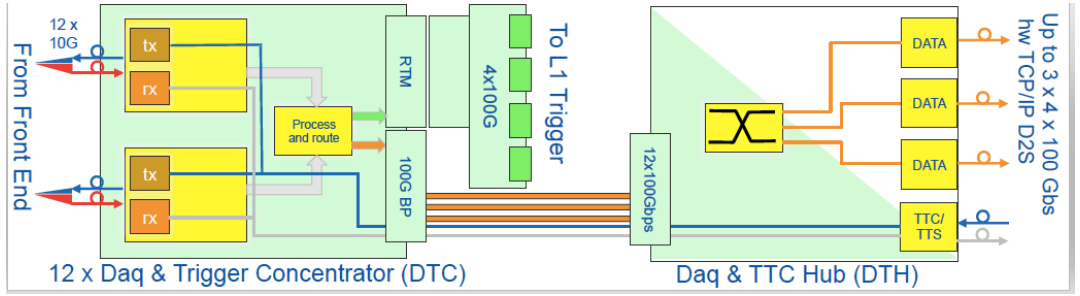
DT slice test electronics (Theta splitter)

- Improved Resistive Plate Chambers (iRPC)
  - Redesigned for high background in HL-LHC conditions at high-η
  - Improved hit rate and time and spatial resolution
  - Demonstrator installed in LS2

Phase 2 Readout and DAQ interface

Front end radiation-hard electronics





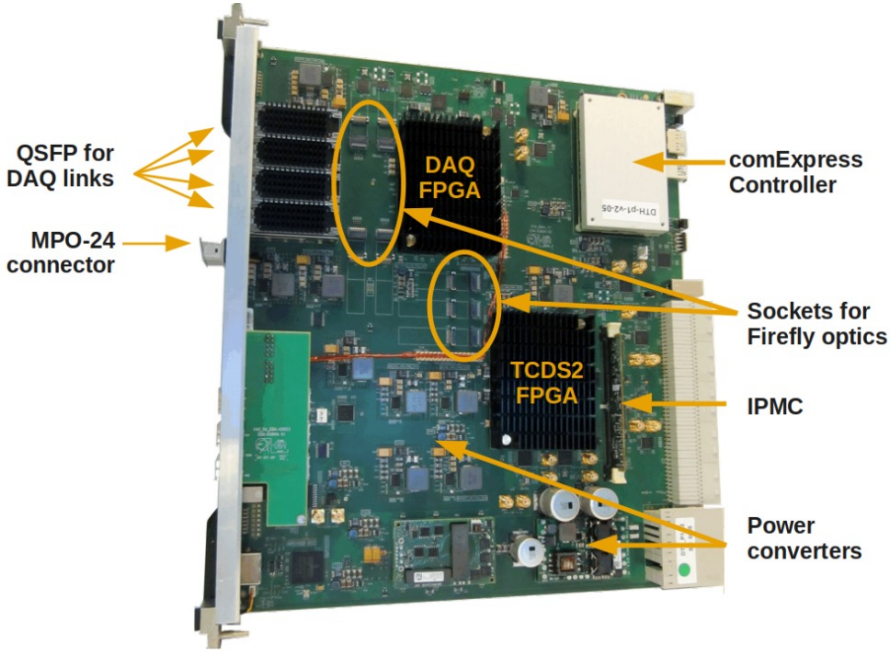
40 MHz trigger primitive
750 kHz full-event readout
(~ 8 MB/ event)

~50000
FE optical links

lpGBT

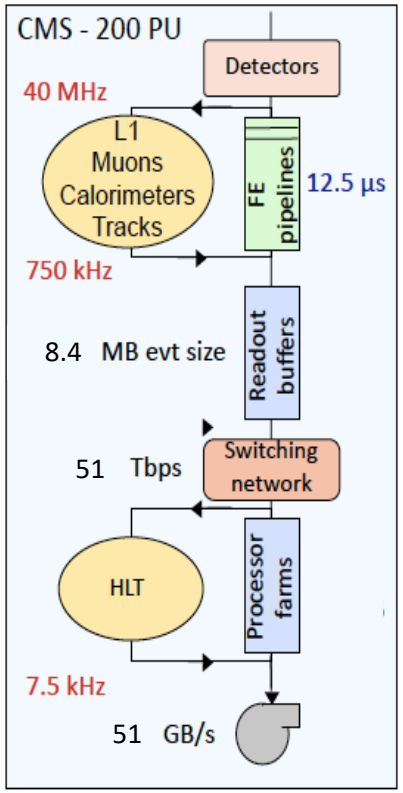
ATCA modular electronics

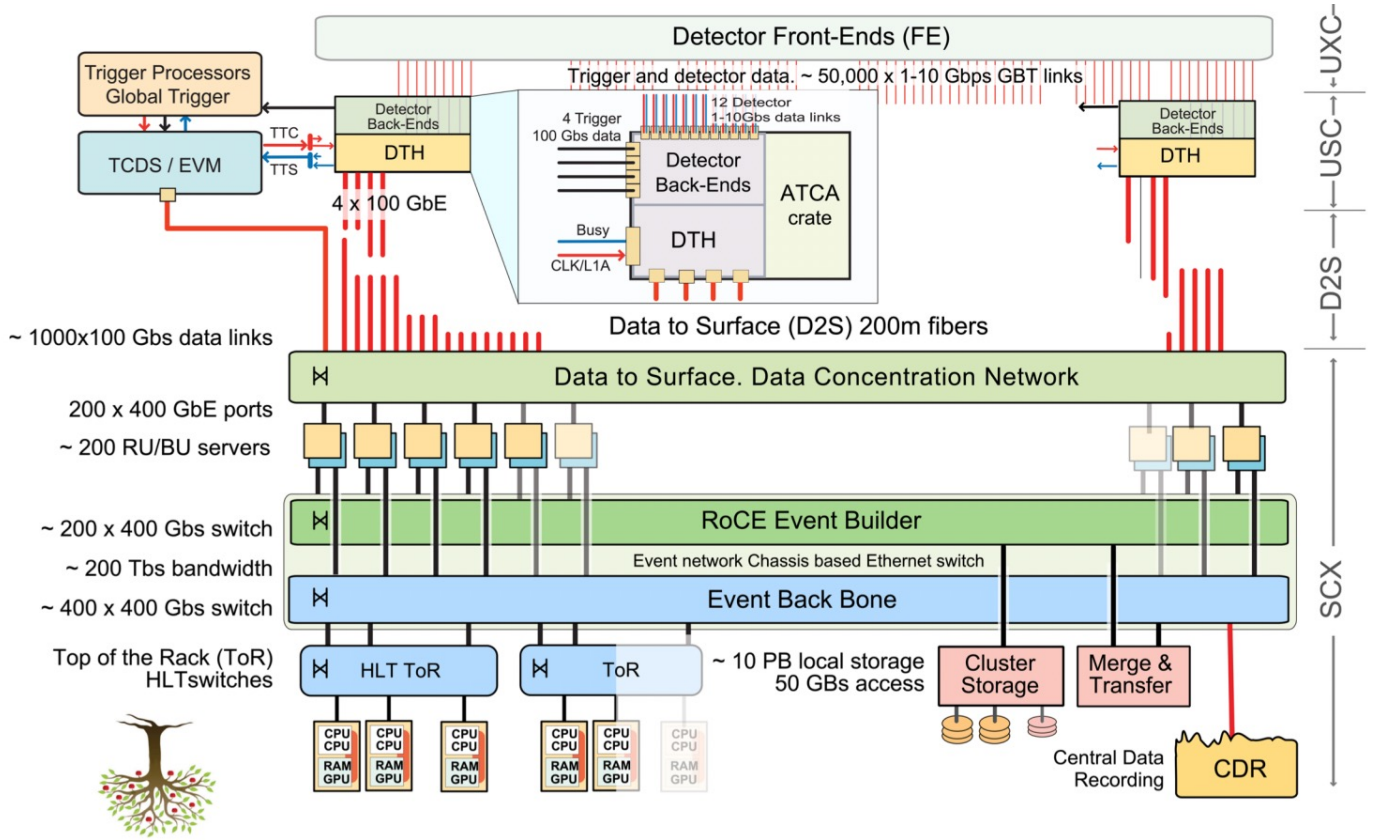
- DAQ and Trigger Hub (DTH)
  - Custom ATCA board for high-bandwidth readout
    - 2x Xilinx FPGA and HBM memory
  - Input: Firefly 100 Gbps inputs (4x or 8x)
  - Output: 100 Gb/s Ethernet (4x or 8x) to DAQ data concentrator - data-to-surface
  - Precise clock distribution
  - installed in ATCA crate with detector-backend electronics



v2 prototype of the DTH-400 board

Phase-2 evolution of CMS DAQ





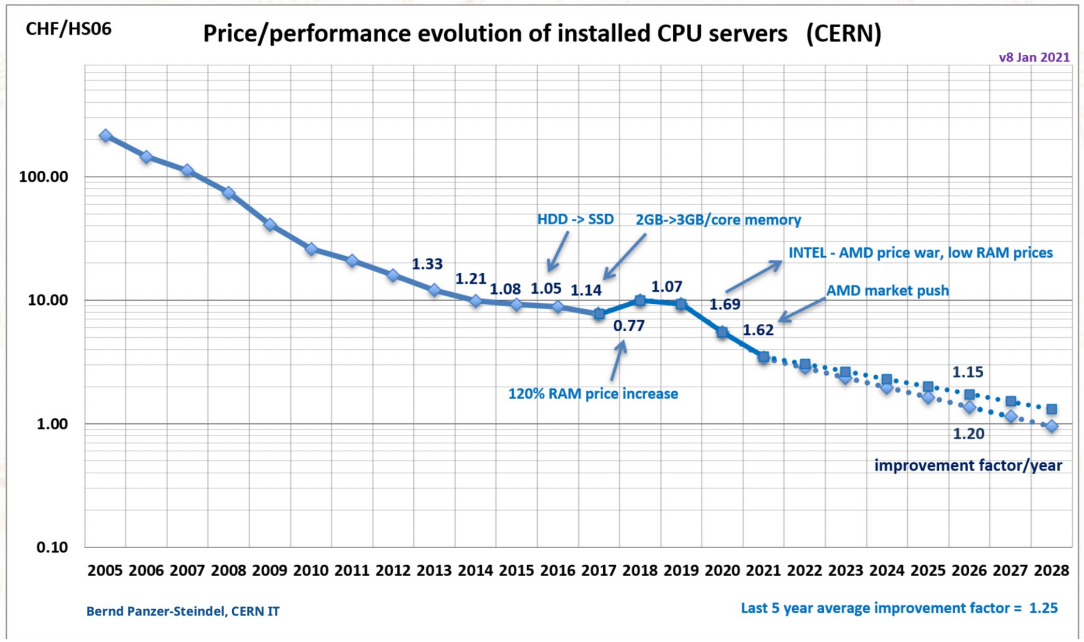
- DAQ Phase-2 reaquirements compared to Run 3
  - 7.5x L1 rate, ~4x event size → ~ 30x **readout bandwidth**
- Aiming to achieve by scaling up current DAQ architecture, taking advantage of hardware evolution
  - 400 Gbps Ethernet (or equivalent) and new gen. computers for I/O

Orbit building

- In place of the event-building
- 11 kHz orbit rate in DAQ
- Events from an orbit will be built in HLT nodes

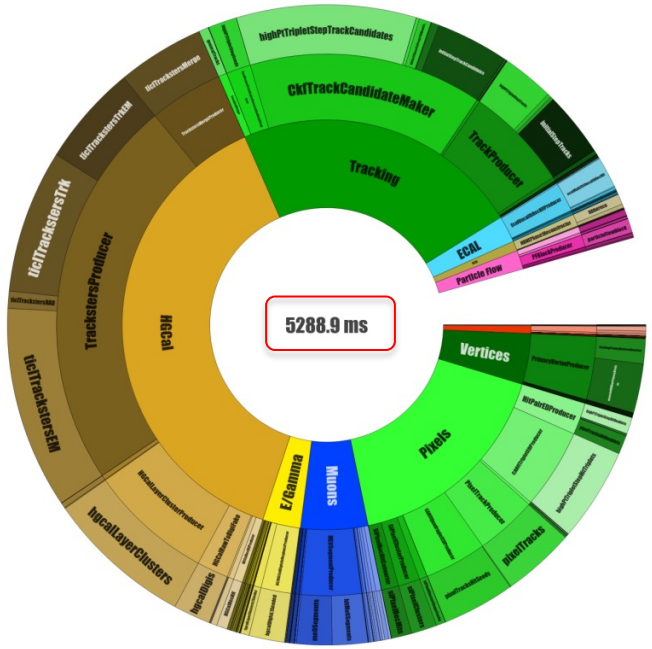
Phase 2 HLT requirements

- Large challenge
  - Estimated 50x computing power increase (for Run-5) with current approach
    - Due to high pileup, L1-rate and increased requirements reconstructing upgraded and new detectors (mainly Tracker and HGCAL)



Strategy:

- Technology evolution of CPUs (cost reduction)
- Coprocessors (GPUs)
  - Assuming similar price/performance evolution
  - Goal: adapt up to 80% HLT code (processing time) to GPU algorithms
- Algorithmic improvements, optimizations and quality improvements in development



Measurement of Phase-2 event reconstruction time on MC-simulated events

Conclusion

- Extensive upgrade program in LS2 was completed
  - Upgraded and refurbished systems were commissioned in first half of 2022
  - Successful first year of Run 3 with 37.5 fb^{-1} recorded by CMS
  - 34 fb^{-1} certified for the physics analysis
  - Overall efficient operation of LHC and CMS in this period

- HL-LHC upgrade program
  - Large R&D effort in progress withing the CMS Collaboration on the upgrade
  - Advancing from final engineering and prototyping into production phase
  - Software reconstruction development in preparation for running with new and upgraded detectors

BACKUP

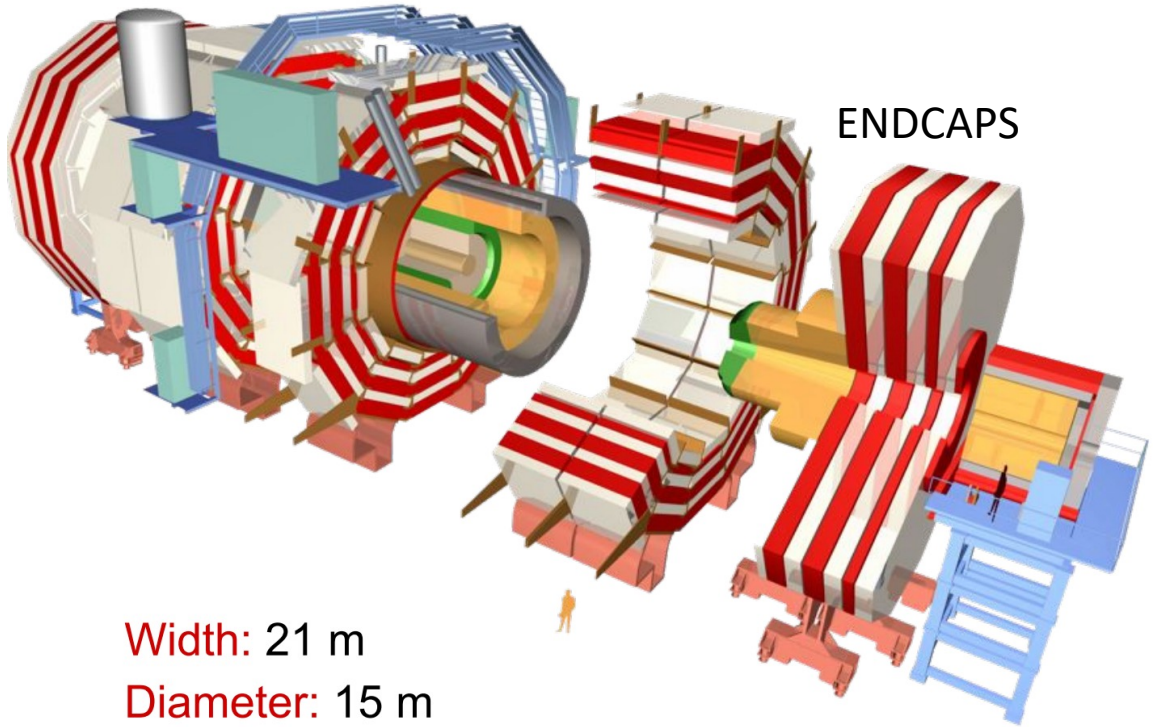
Compact Muon Solenoid (CMS)

- ~ 30 MHz collision rate for proton bunches (25 ns spacing → 40 MHz)
- 2-stage filtering of collision records: Level 1 (hardware) and Level 2 software

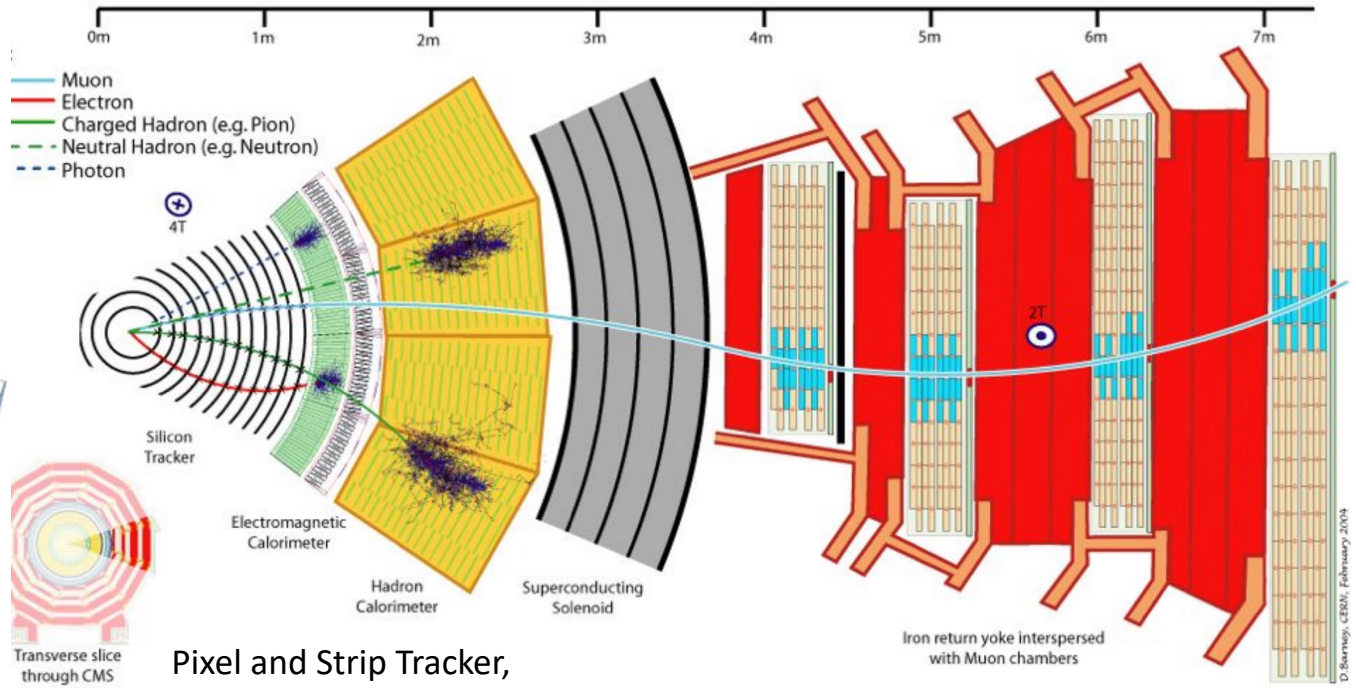
Compact Muon Solenoid

BARREL

ENDCAPS



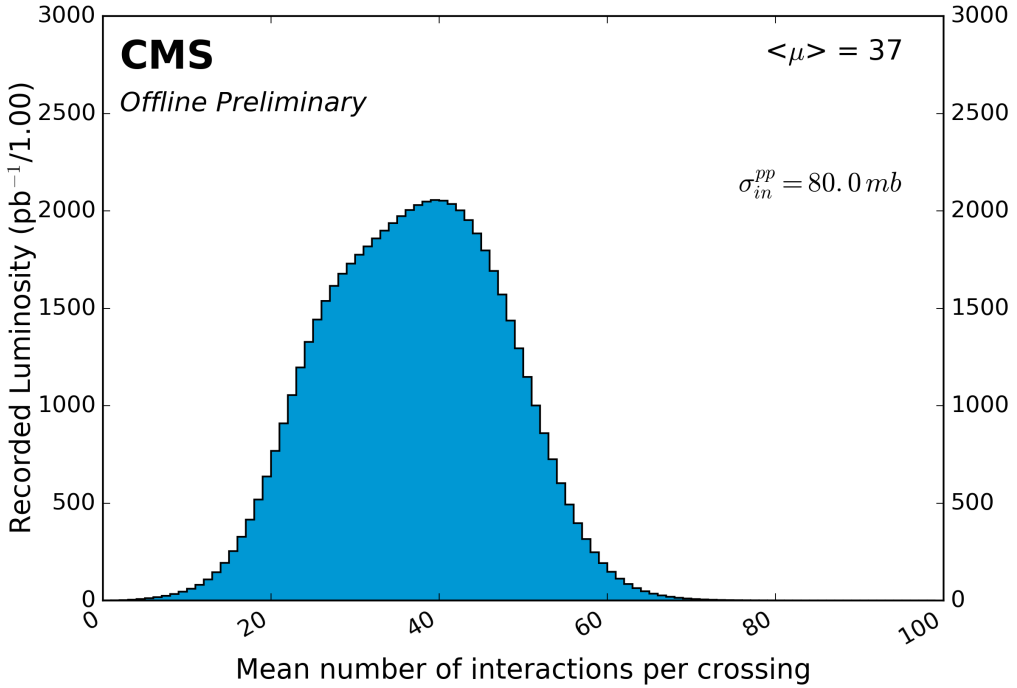
Width: 21 m
Diameter: 15 m
Weight: 15000 t
Magnetic field: 3.8 T (solenoid)



Pixel and Strip Tracker, ECAL, HCAL

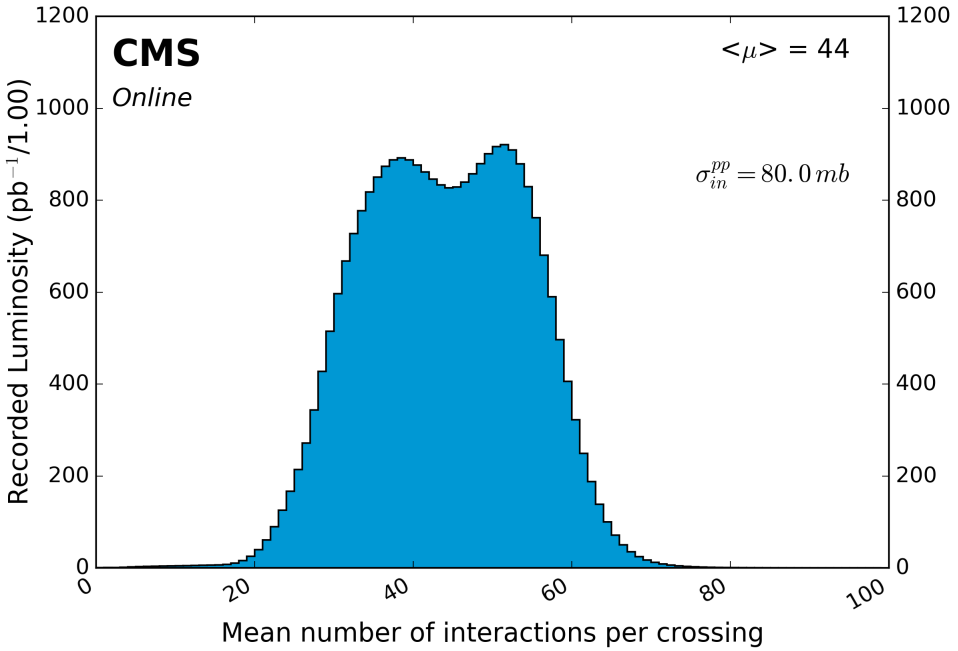
Muon detectors: DT, CSC, RPC, GEM

2018 (Run 2) pileup distribution



https://cern.ch/cmslumi/publicplots/2018/pileup_pp_2018_80000_Normtag.png

2022 (Run 3) pileup distribution

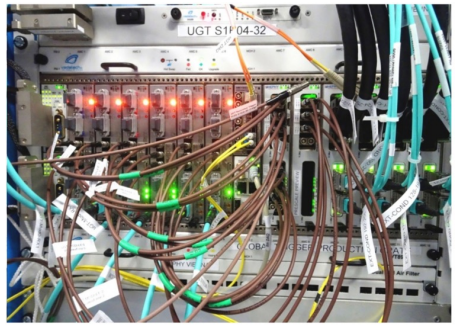


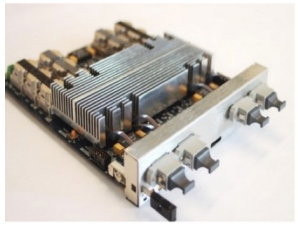
https://cmslumi.web.cern.ch/publicplots/2022/pileup_pp_2022_80000_Normtag.png

Level-1 Trigger in Run 3

N. Dev et al 2017 JINST 12 C02014

- Phase-1 Trigger upgrade – in use since Run 2 (2016)
  - New electronics (μTCA standard)
  - Improved performance at increased pileup and luminosity
  - Full-granularity of calorimeter exploited
  - Combining muon detector information early in the chain (improving efficiency and rate reduction)
  - Supporting more sophisticated algorithms





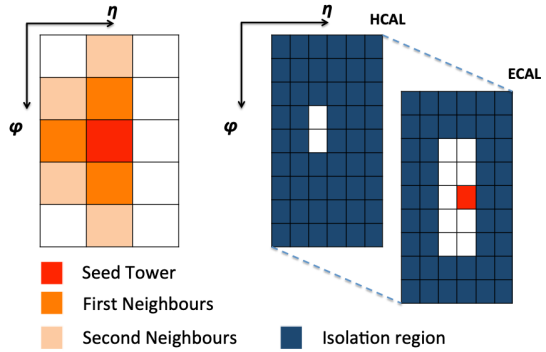
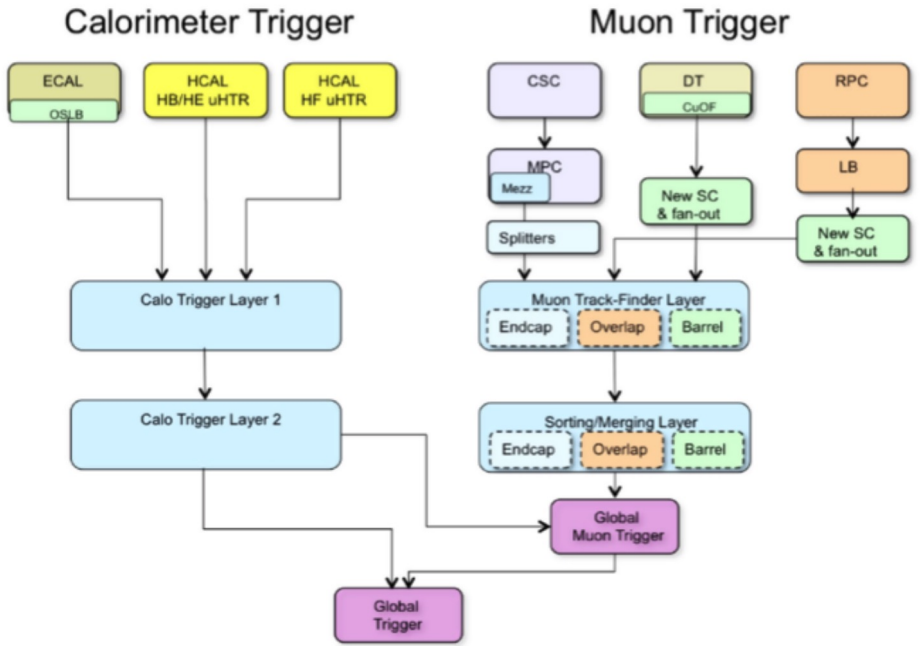
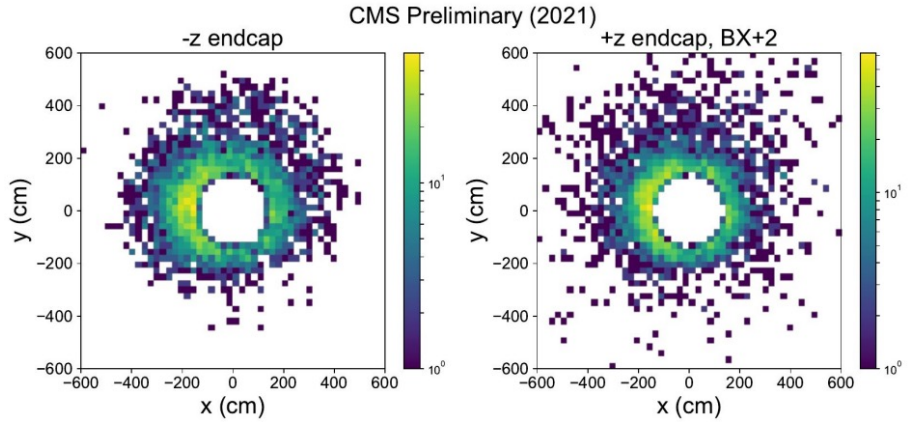


Illustration of L1 calorimeter clustering and hadronic isolation region



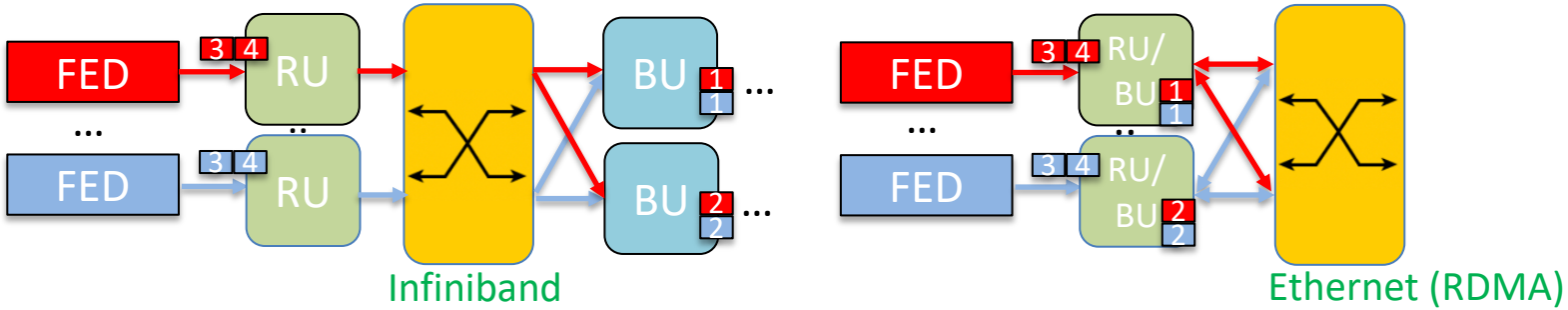
- L1-scouting
  - Reading L1-trigger data at the full bunch-crossing rate (passively) from the L1 system
  - Output saved to storage for analysis
  - Physics at the bunch-crossing rate → search for rare processes & difficult to trigger signatures at L1-Trigger resolution
  - A demonstration system in Run 3 for the Phase-2 upgrade

Early results with beams: beam halo muons



Core Event Building

- Scattered detector readout → complete-event payloads
- Run 3 DAQ evolution – folded event-building (EvB)
  - Advantage: bidirectional use of network links
  - But more demanding on individual PC CPU and I/O performance



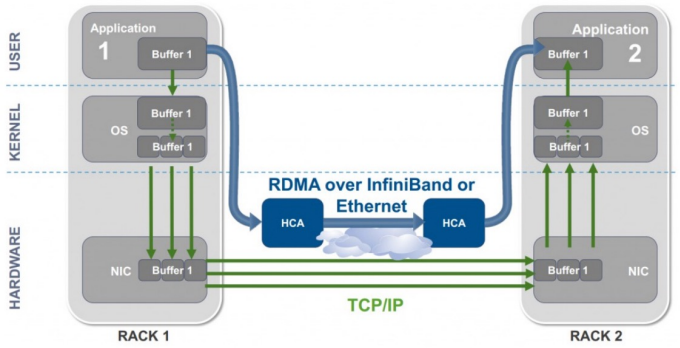
Linear EVB (Run 2)

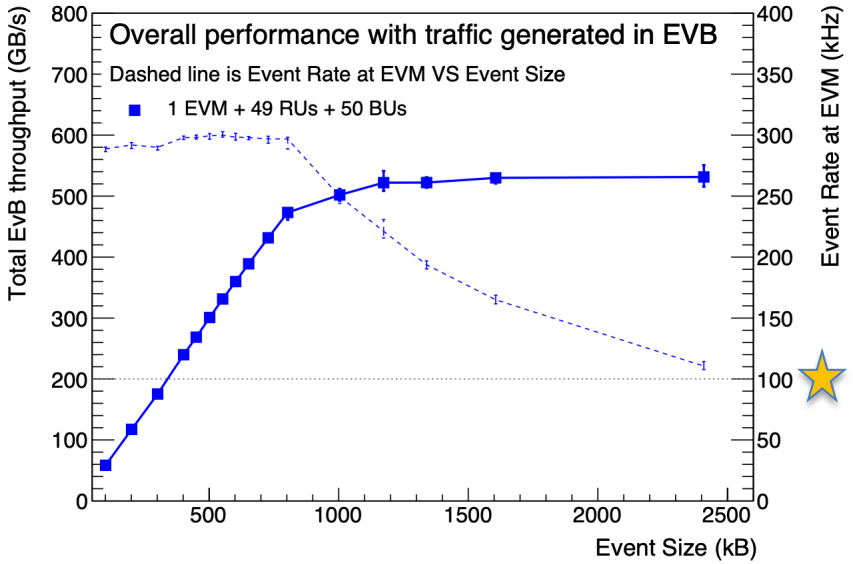
Folded EVB (Run 3)

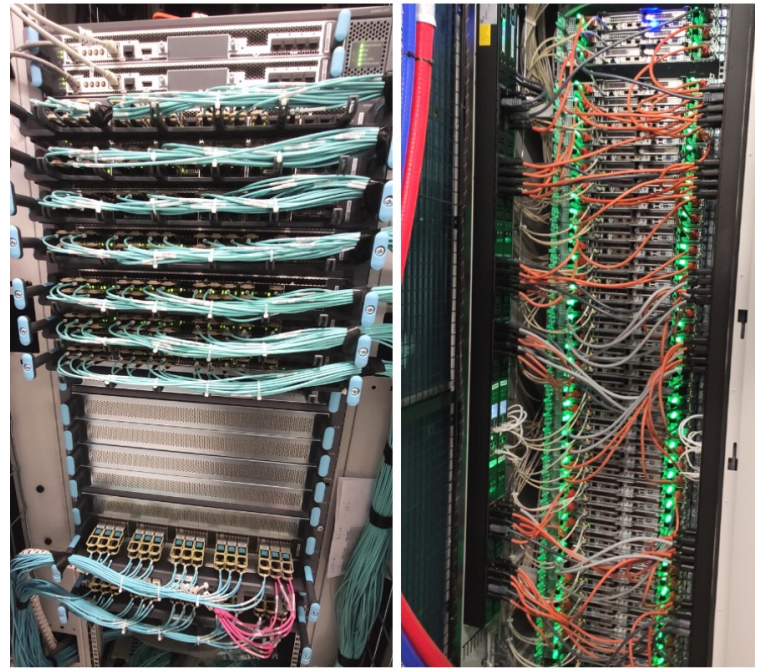
Chassis-based Ethernet switch (Juniper)

- Single switch for readout (672 x 10 Gb/s inputs) to RU/BU PCs
- Event Building network – 100 Gb/s Ethernet links
  - O(50) links to RU/BU PCs
- Event Backbone network
  - 100 Gb/s EVB ← → 25 GB/s Ethernet links
  - Serving HLT, Storage filesystem and data transfer from P5 to CERN computing center (Tier-0)

New in Run 3: 100 Gb/s Ethernet with remote DMA (RoCE v2)



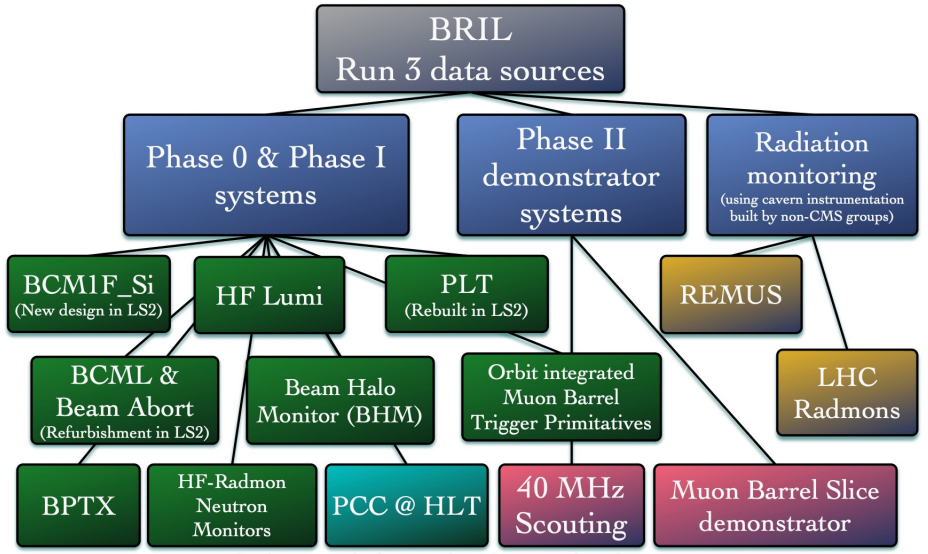




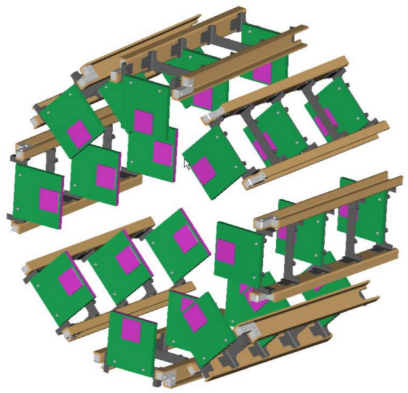
Fully cabled Run 3 EvB switch

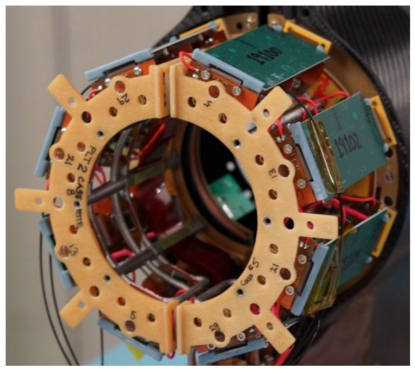
Beam Radiation Instrumentation and Luminosity (BRIL)

- Beam and radiation monitoring and protection
- Performs luminosity measurement
  - real-time for immediate feedback to LHC and detector operations, and for 'offline' analysis
  - Using detectors - Pixel tracker, HF, Muon system
  - Using deficated luminometers - PLT and BCM1F

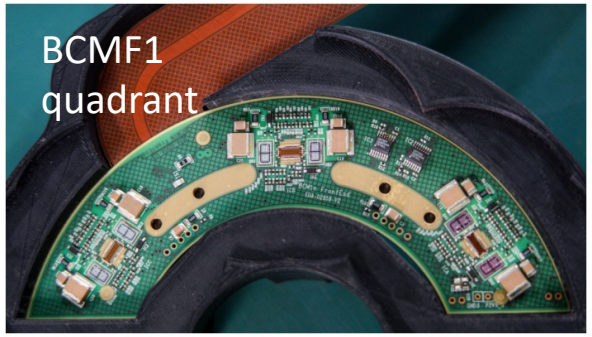


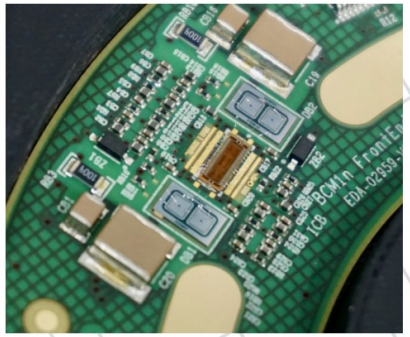
- Pixel Luminosity Telescope (PLT) –
  - 8 silicon pixel detectors located at each end of CMS
  - refurbished for Run 3 due to radiation damage



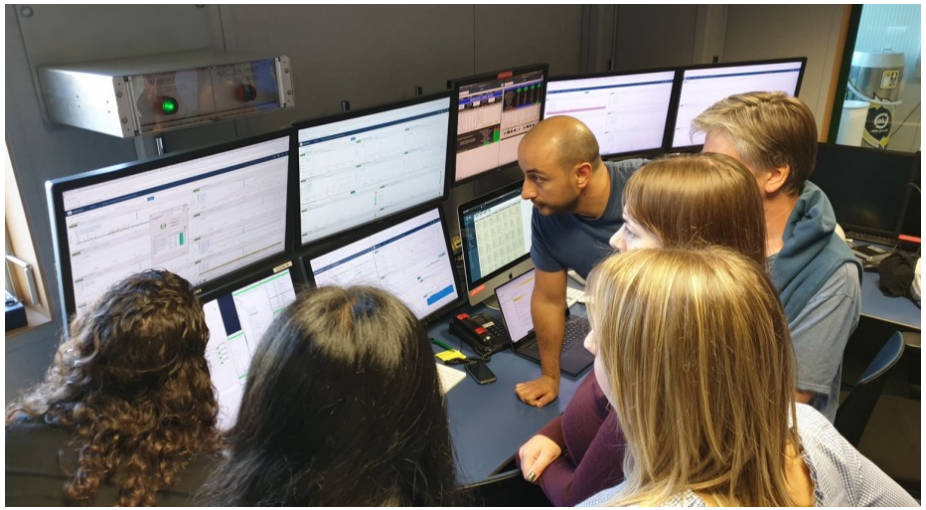


- Fast Beam Conditions Monitor (BCM1F)
  - Fast particle counter
  - Redesigned in Run3 : all silicon detector
    - Improved stability, radiation tolerance and linear response
  - Located 1.8m from IP at radius of ~ 6 cm



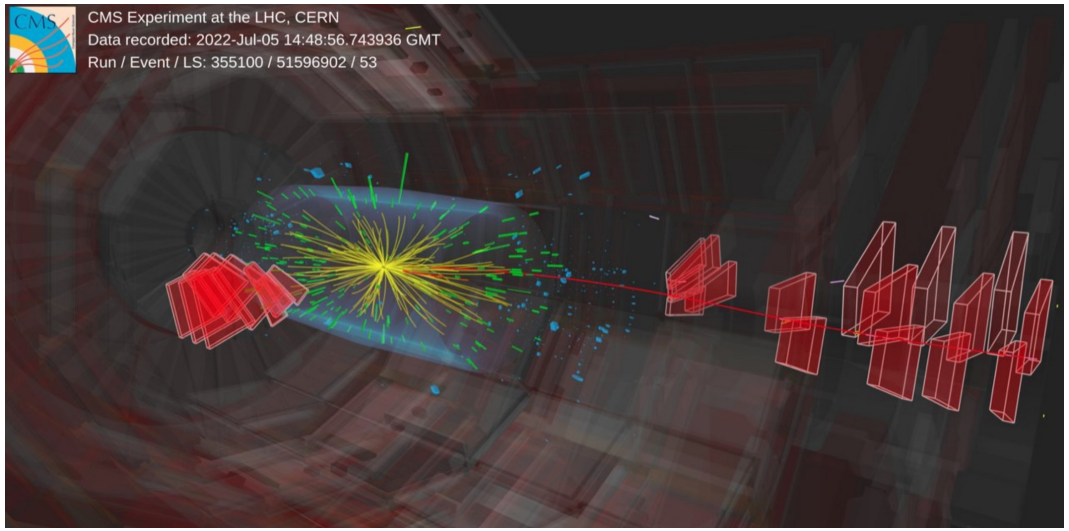


Start of the Run 3

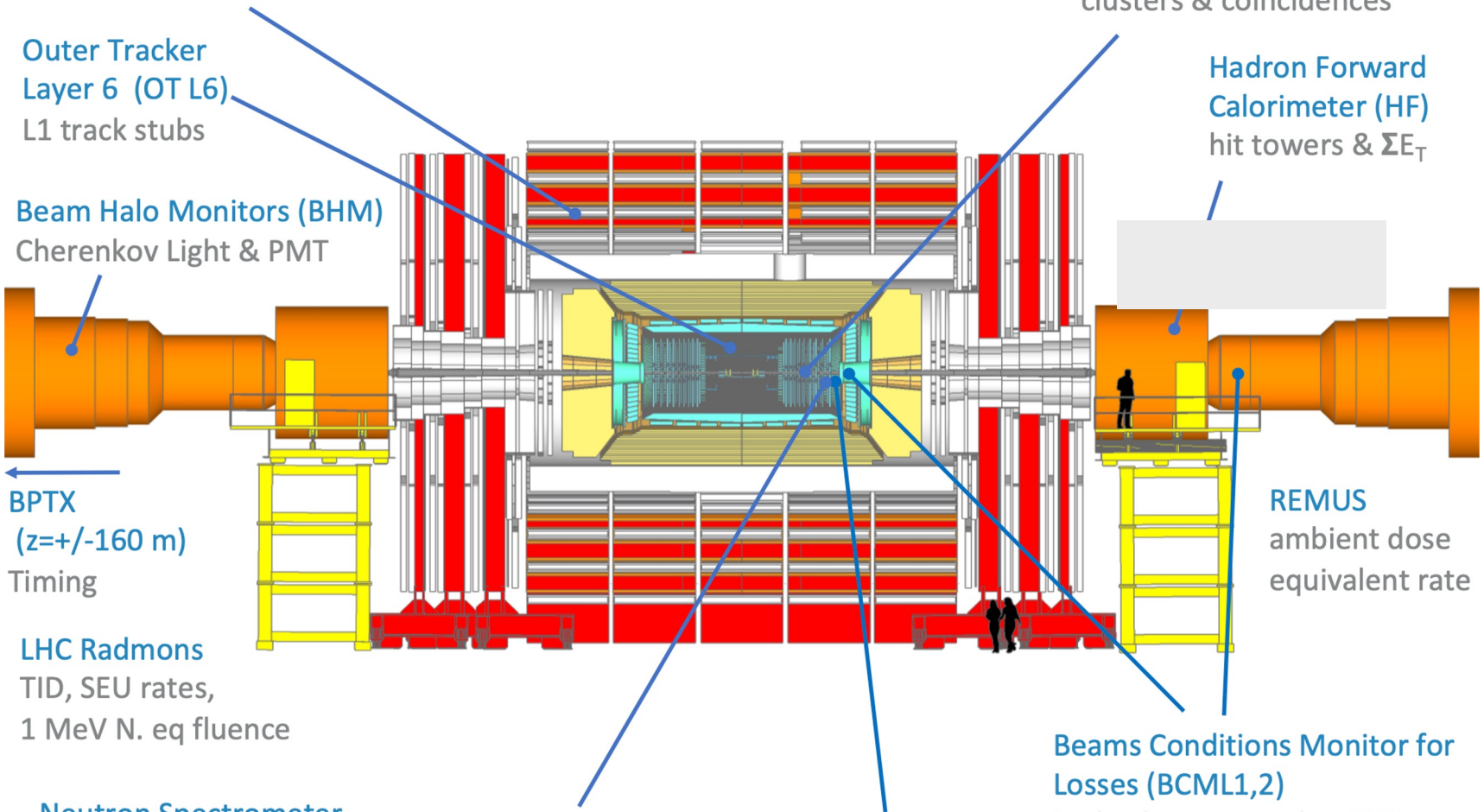


- Intense period of commissioning with cosmics and first beams circulating in the LHC
- First sqrt(s) 900 GeV collisions – May 27 2022.

- July 5th 2022. – start of 13.6 TeV Run 3 proton-proton physics run (media day)

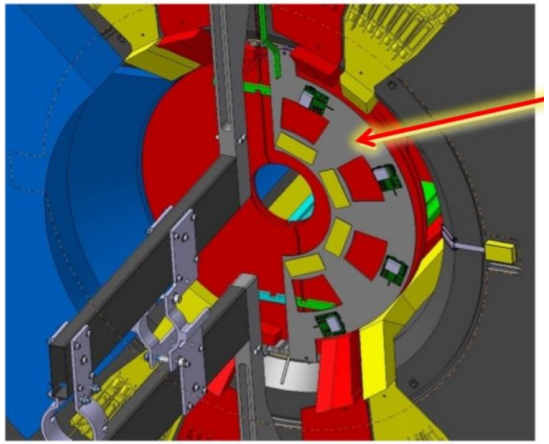


BRIL: 15 technical systems for HL-LHC



14 BRIL technical systems for Phase-2

- Bunch-by-bunch luminosity
  - Target: 1% precision for analysis, (2% online)
- FBCM
  - Dedicated luminometer
  - Si pad diodes
  - sub-BX timing and radiation resistance
  - Adapting Phase-2 inner tracker electronics



High Level Trigger refresh and performance

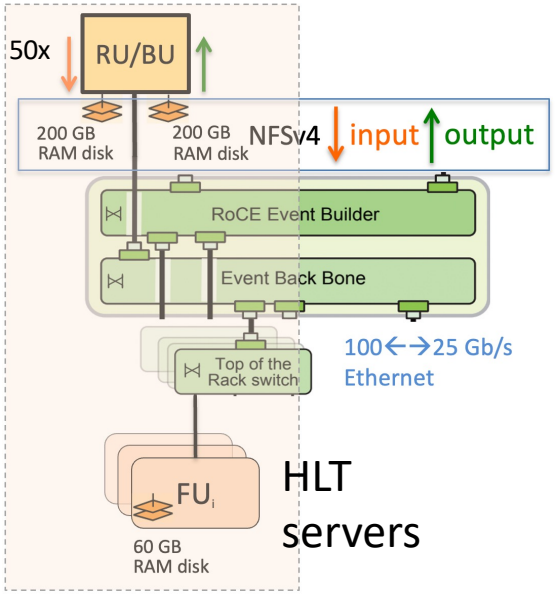
- Full-event reconstruction and filtering in the HLT
  - New generation of Server CPUs replacing end-of-life Run 2 nodes
  - Increased level of multi-threading in Run 3 → improved memory usage
  - Various algorithmic improvements and taking advantage of detector upgrades
  - Heterogeneous computing: GPU coprocessors (pilot for Phase-2)
  - Output to a distributed LustreFS filesystem and transfer to Tier-0

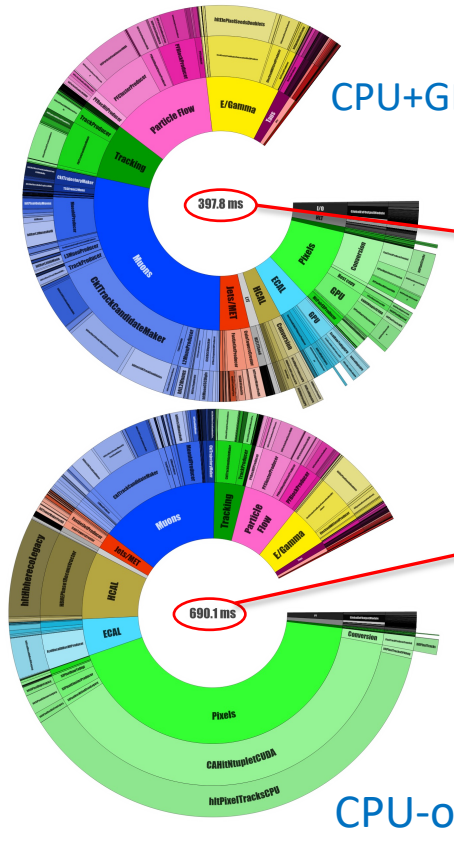
Nvidia Tesla T4

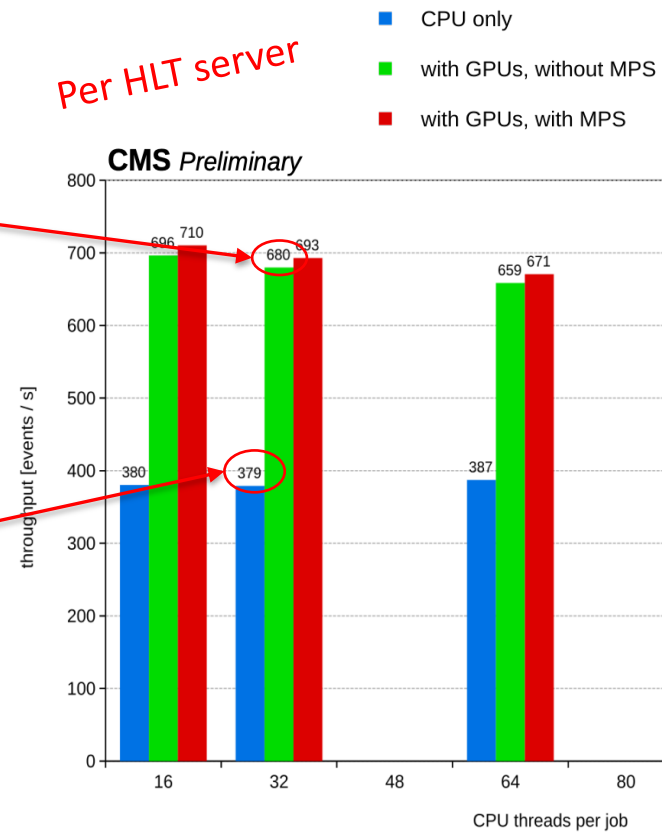
- CPU → GPU offloading for HLT reconstruction
  - Initially implemented for: Pixel, HCAL, ECAL
- Achieved ~= 40%

2022 HLT Farm

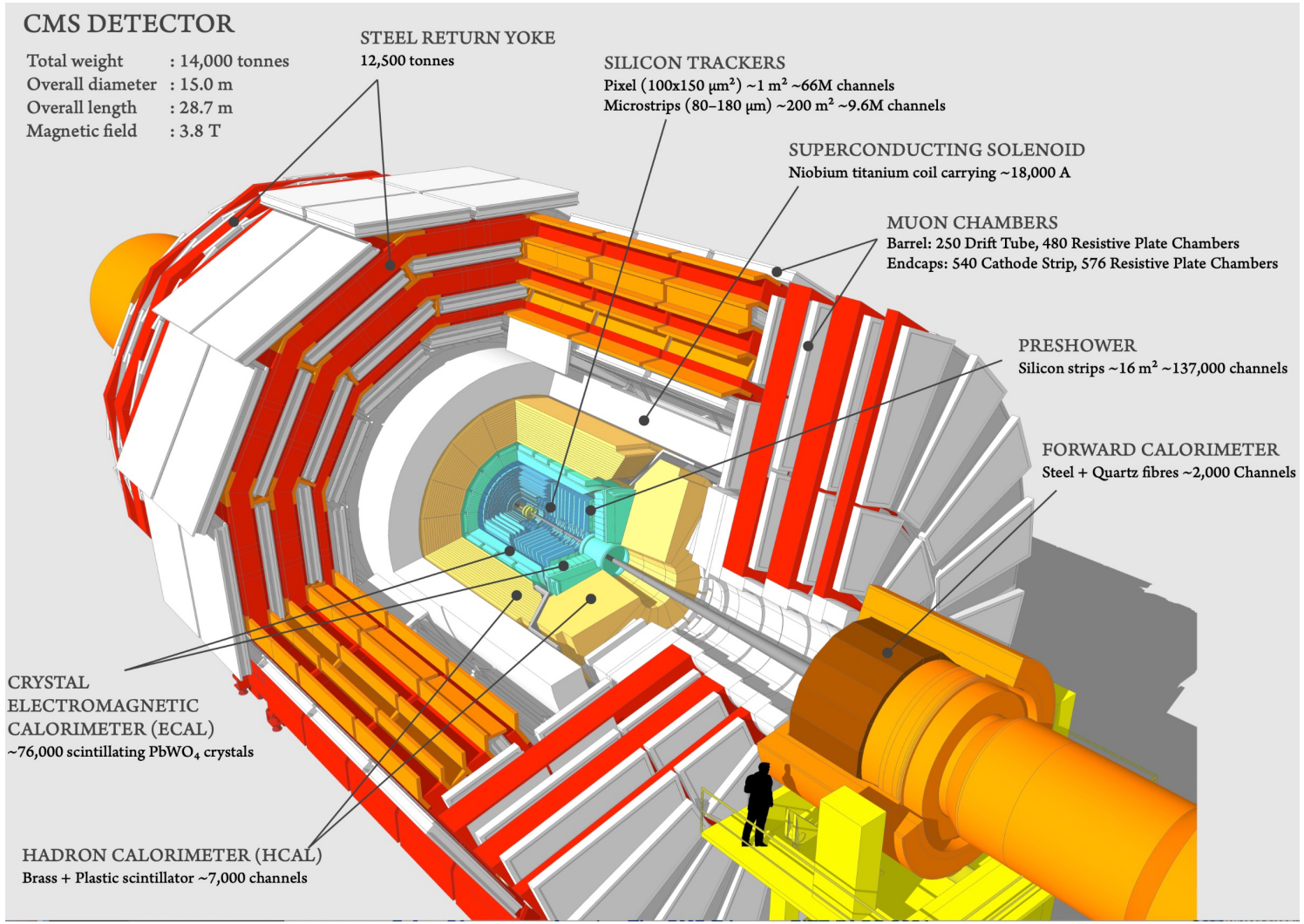
CPU and memory	2 x AMD Milan 7763 (64 cores @2.45 GHz) 256 GB RAM
GPU	2x Nvidia Tesla T4
# servers	200
# cores	25600
kHS2006 benchmark	645





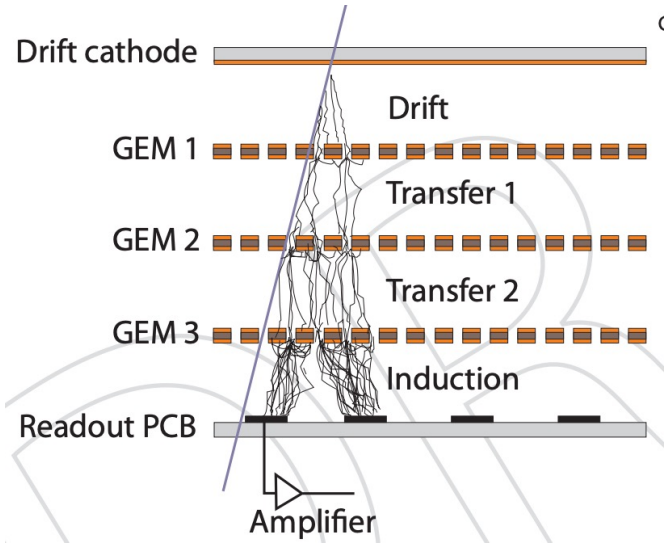


Compact Muon Solenoid

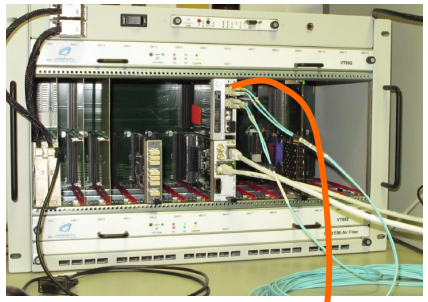


Muon system – Gas Electron Multiplier Chambers (GEM)

- GEM detectors
  - High-rate capability, adequate pattern recognition → more efficient muon reconstruction
  - radiation resistance
  - Important for phase-2 (5 kHz /cm^2 hit rate)



BACKUP: CMS Readout in Run 2 and Run 3

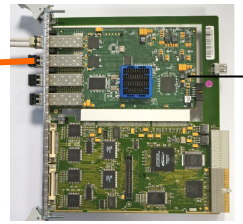


µTCA electronics

FED Fragments 2..8 kB

Optical SLINK-express
6 Gb/s - 10 Gb/s

100 kHz readout



Frontend-Readout Optical Link

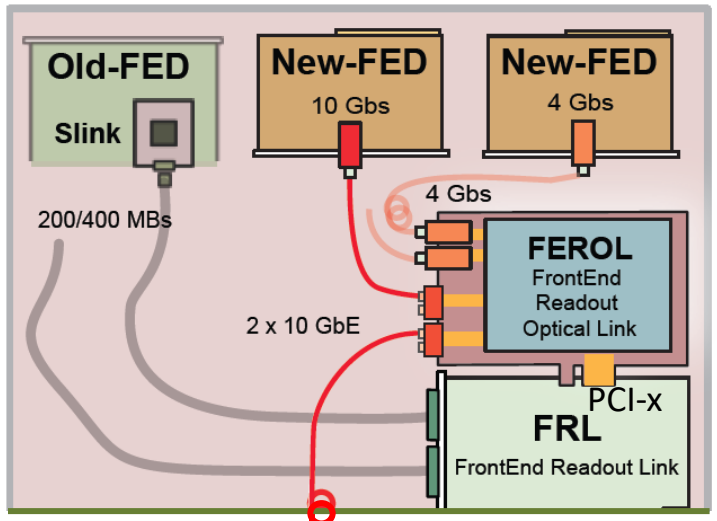
FEROL

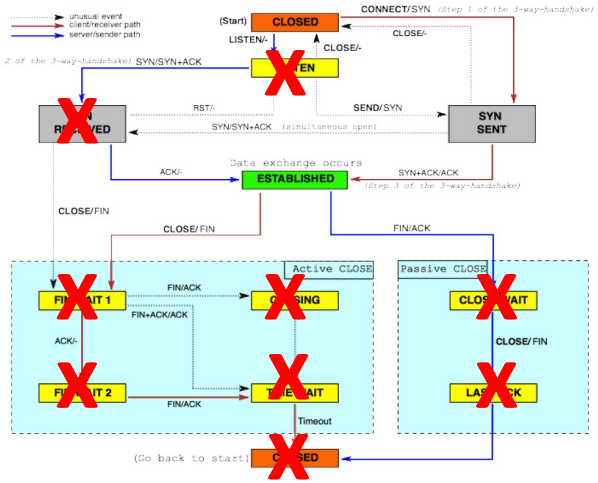
- 6 - 10 Gb/s SLinkExpress inputs
- 10 Gb/s simplified TCP/IP output (Ethernet) implemented in FPGA
- High-throughput FEROL40 µTCA card – used with Pixel upgrade

- Phase-1 (Run 2) upgrade readout system remains in use in Run 3
  - Run-1 era detector hardware: still in use interfaced through the old FRL (PCI-X) carts
- TCDS – timing and clock distribution system
  - Introduced in Run 2 and fundamentally unchanged in Run 3

- Run 3:
  - 759 FED fragments → ~ 510 readout cards (FEROLs)
  - added GEM detector readout and 8 new HCAL channels
  - Upgrade of VME-PCs - control subdetector crates

Run 2 and Run 3 - FEROL TCP/IP





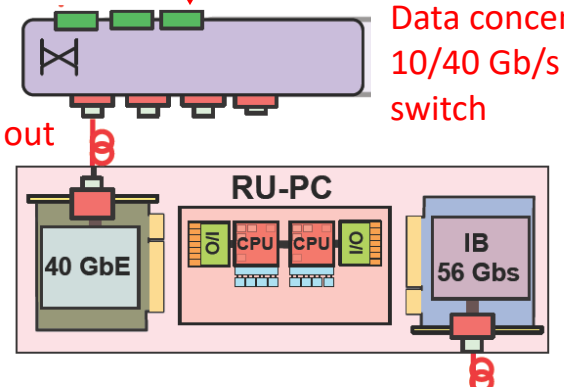
Simplified unidirectional TCP/IP implemented in FPGA

10 Gb/s simplified TCP/IP from an FPGA

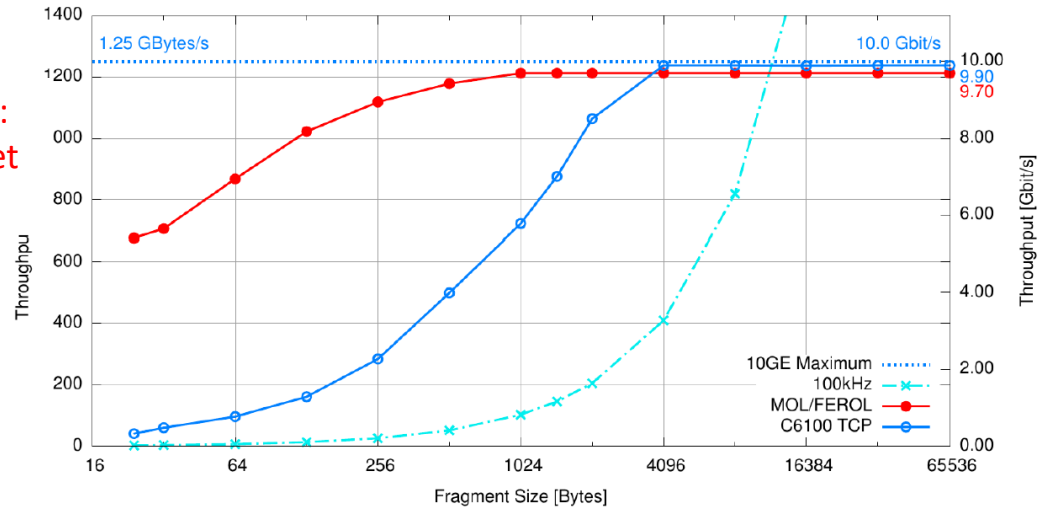
48 x 10 Gb/s in

6 x 40 Gb/s out

Data concentration: 10/40 Gb/s Ethernet switch



Point to point link performance: 9.7 Gb/s for fragments > 1 kB



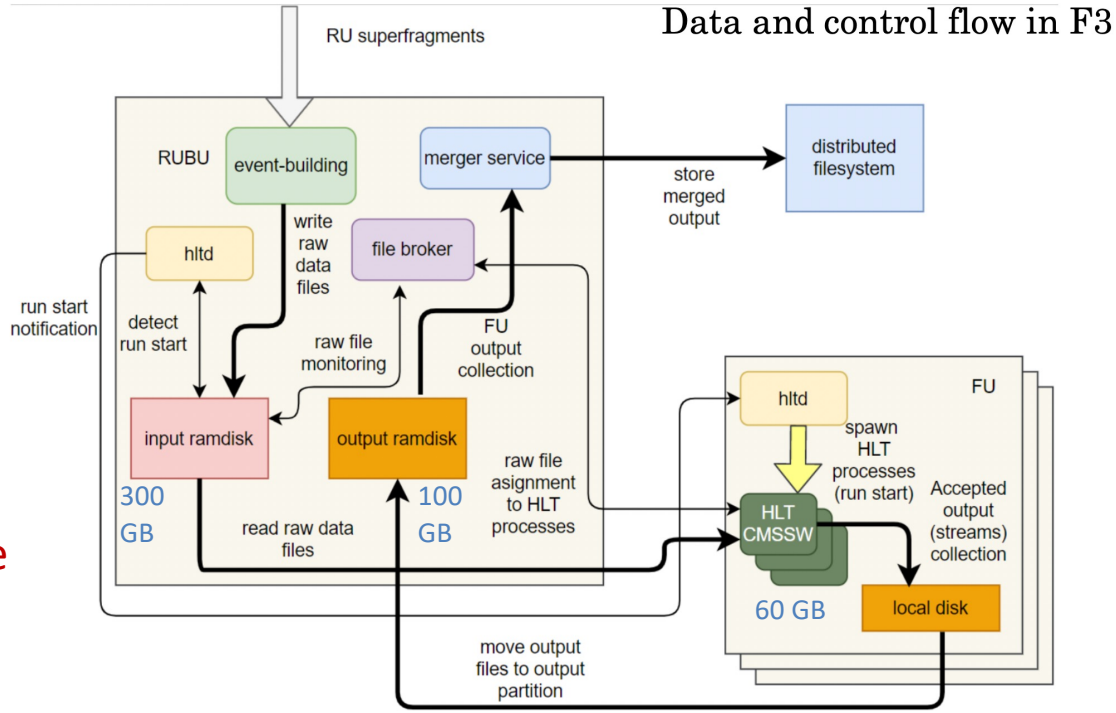
File-based Filter Farm

- Asynchronous service-based approach
  - HLT daemon service (Python 3) based on filesystem events
    - Spaws and controls CMSSW processes running HLT reconstruction and selection
  - Merging
    - 1st level: aggregation of HLT output on FU shipping to an output RAM disk buffer on RU/BU (overall ~ kHz)
    - In total, 3 merging levels (per-FU, per RU/BU (appliance) and global) to get the data fully aggregated in single location
  - Dedicated elasticsearch based monitoring system

- New functionality in Run 3: Dynamic assignment of **FU Groups** to RU/BU nodes
  - Flexibility in case of HW issues and (e.g. replacing RU/BU with hot-spares), prevents

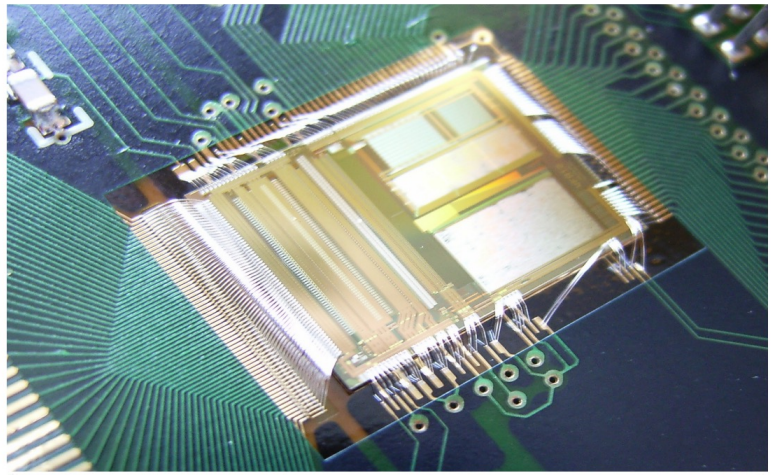
Cconsidered a suitable design for the Phase-2 upgrade

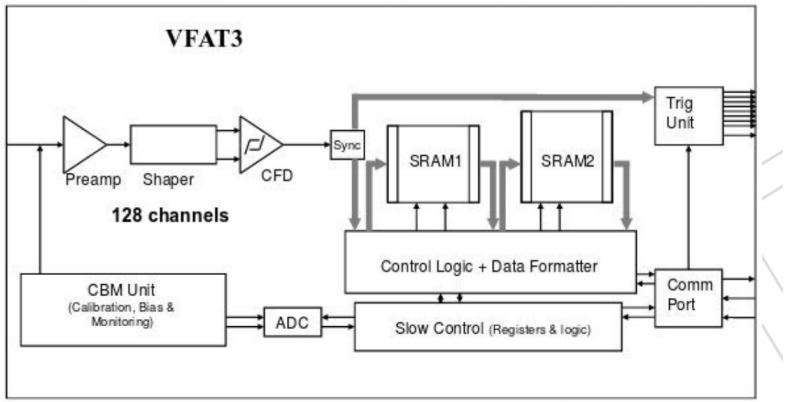
Data and control flow in F3

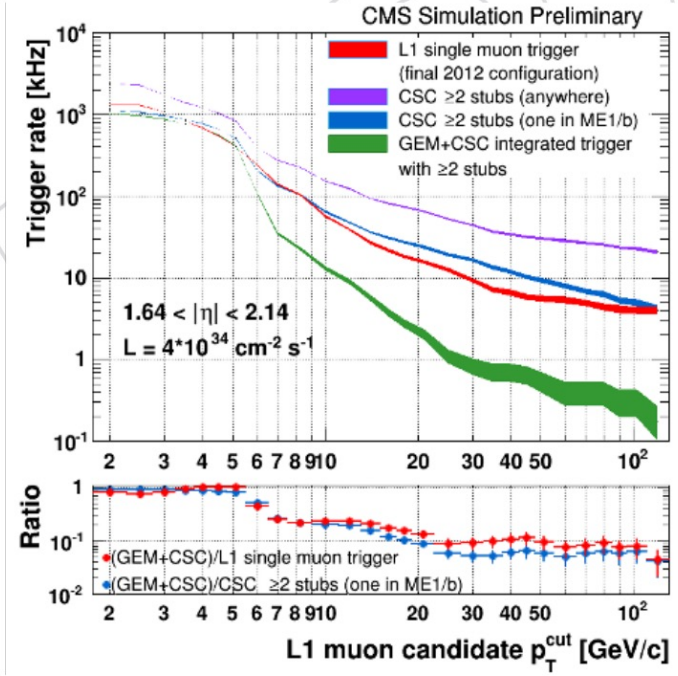


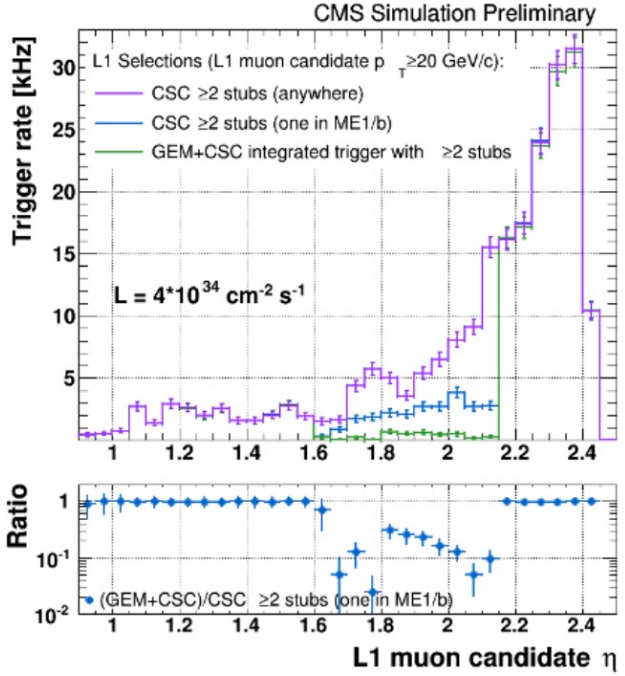
Muon system – Gas Electron Multiplied Chambers (GEM)

- New on-detector/readout electronics



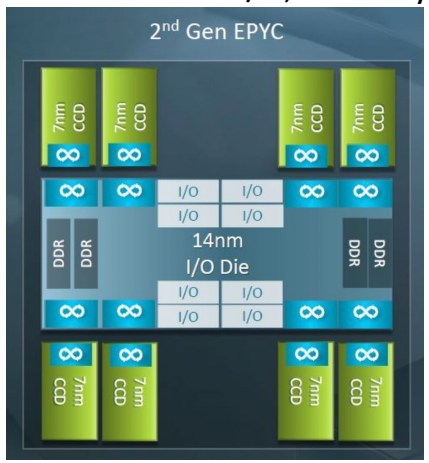


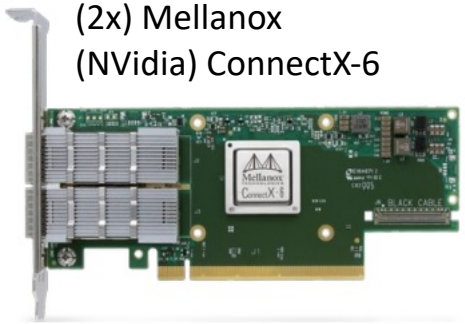




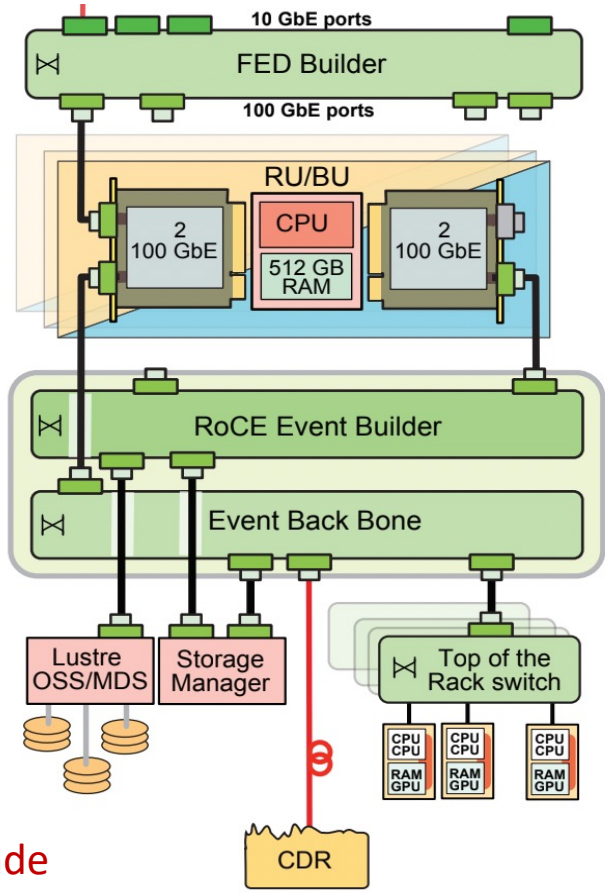
I/O nodes – “RUBU”s

Zen2 AMD architecture
8 x 4-core CCDs chiplets (16 MB L3 cache)
Interconnect: I/O, memory





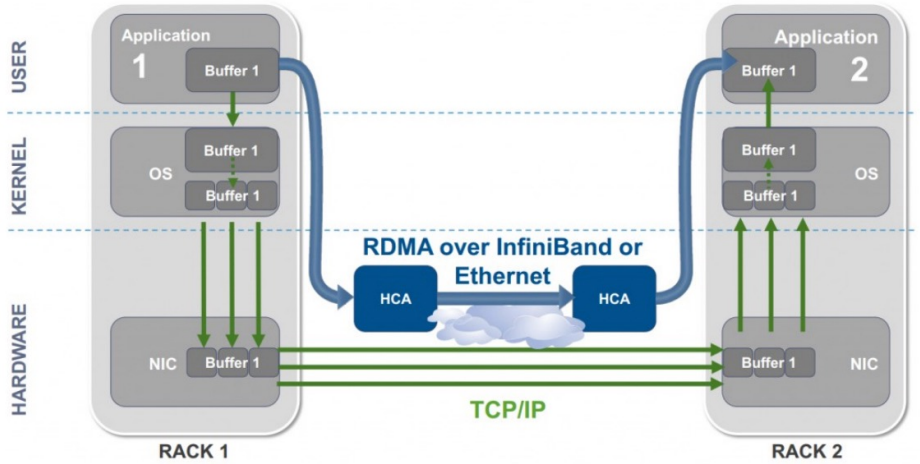
(2x) Mellanox
(NVidia) ConnectX-6



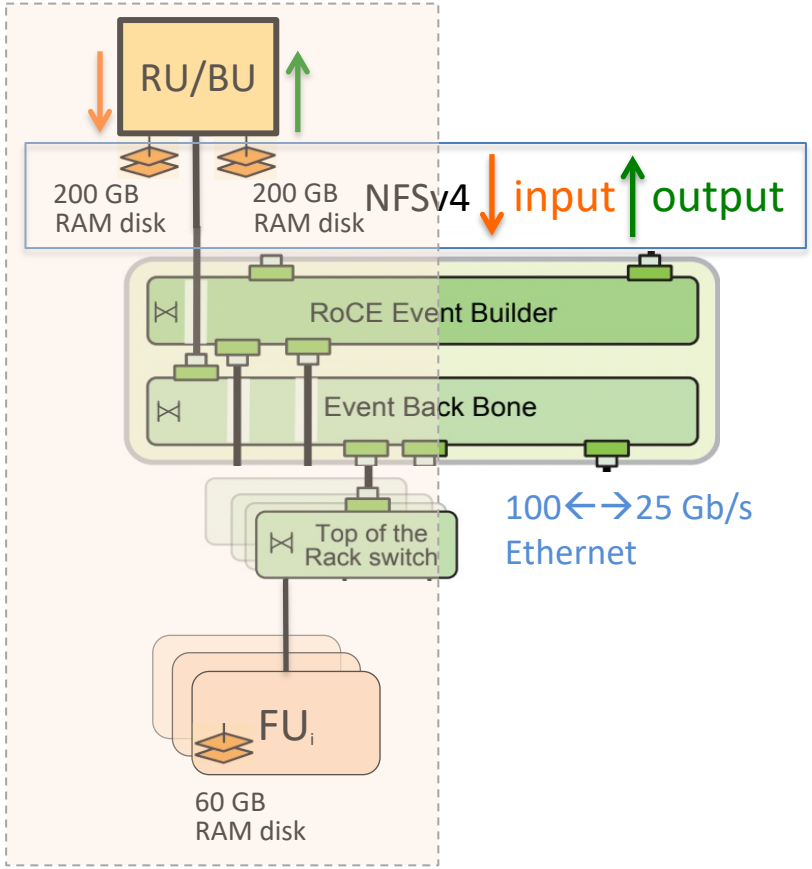
- Dell R7515 nodes (up to 57 in production):
  - 32 Core AMD Epyc 7502P processor (single-socket)
    - Evolution over 8-core / 16-core RU, BU Intel servers in Run 2
  - 512 GB RAM DDR4 - 8 channels
    - application and data buffering
  - Experimenting with single-socket systems for phase-2 upgrade
  - 2 x Mellanox ConnectX-6 Dual-100 GB/s NICs (PCIe Gen4)
    - Utilising 3 x 100 GB/s Ethernet links: for data-concentrator, EVB, Event-Backbone networks
    - New for Run 3 EVB:RoCE v2 (RDMA over Converged Ethernet)
      - Possible with a monolythic switch and more cost effective than Infiniband – interest in technology for Phase-2
    - Gen5 PCIe cards feasible for Phase-2

New technology - RoCE

- RDMA
  - Direct memory access between servers (userspace applications) without involving host CPUs – acceleration provided by host NICs
  - Part of Infiniband standard
  - Used in Run 2 DAQ:
    - Via Infiniband Verbs software implementation
- RoCE – RDMA over Converged Ethernet
  - Encapsulation of IB (RDMA) over the Ethernet
  - Same software API
  - Requirement: **loss-less non-blocking Ethernet**
    - Possible with a chassis-based switch
- Adopted for Run 3 DAQ – RoCE v2
  - Cost effective and easier to manage than IB
  - **reused** Run 2 IB software implementation with a minor adaptation



High Level Trigger



HLT farm connected via a high-performance network

- Run 3: Top-of-Rack switches (5 racks * approx. 40 HLT nodes) – total 200 nodes
  - Per rack: 8 x 100 Gb/s uplinks to the Event Backbone network
  - In total 200 nodes HLT nodes with 25 Gb/s Ethernet connectivity

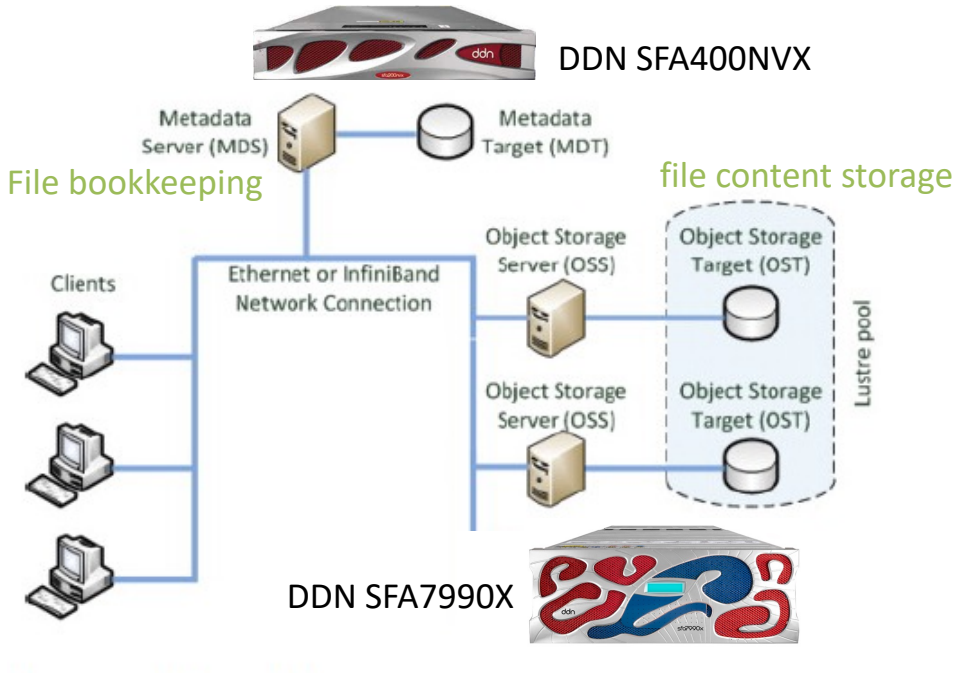
File Based Filter Farm (F^3)

- Concept introduced in Run 2, kept for Run 3
- Input, output (event data and metadata), monitoring and logging flow using **files (filesystem)**
  - **Network filesystem (NFS) v4** over standard TCP/IP as transport protocol
- Reduced coupling between DAQ and HLT software platforms
  - HLT uses standard CMS offline software (DAQ-specific code implemented as modules)

RAM disks (Linux **tmpfs**) used as data buffers

- Buffer on RU/BUs storing completed events from EVB and HLT output
  - 300 GB + 100 GB per RUBU, 60 GB per FU – several minutes of data buffer
  - Each RU/BU ramdisk NFS-mounted by 3-4 FU nodes (**HLT Appliance**)
- From run 3 – using 2 MB huge pages (Linux kernel feature)
  - Necessary for memory I/O performance matching 100 Gb/s link speed

CMS storage and transfer



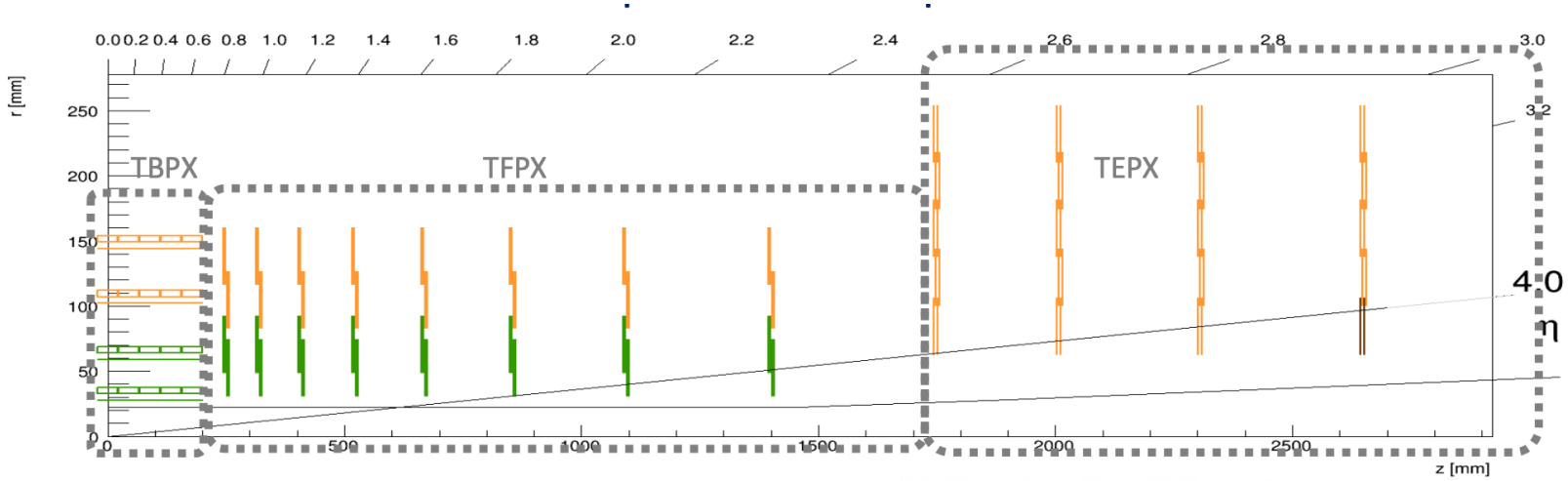
Lustre global filesystem

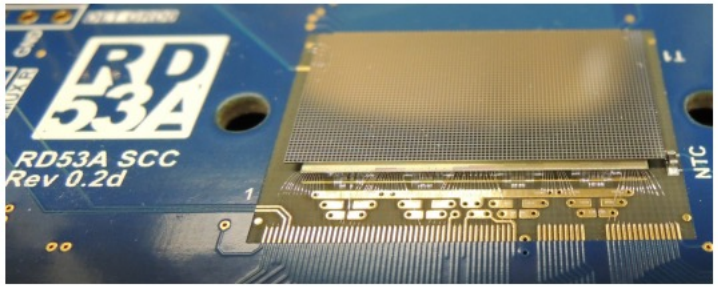
- Write buffer for HLT output aggregated from FUs and RU/BUs
- I/O requirements driven by the Heavy-Ion programme (over 15 GB/s to storage)
- New system installed for Run 3
  - 1.2 PB usable disk space (several days of data taking) - HDDs
  - Network: Event-Backbone
  - 4 x 100 Gb/s links for **transfer to Tier-0**
- **15 GB/s** – stable performance

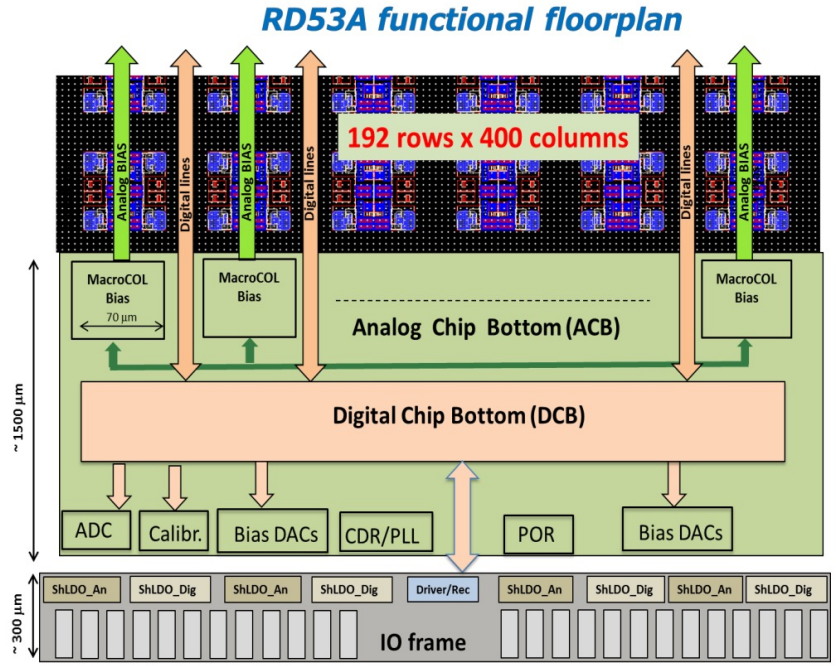
  Scaled up system envisioned for the Phase-2 upgrade

Inner Tracker Phase-2 upgrade

- Expected exposure to radiation levels 1.2 Grad and 2.3 x 10^16 n_{eq}/cm^2
- hit rate 3.2 GHz/cm^2
- 65nm CMOS RD53 (B variant) chip (developed for ATLAS and CMS)
  - Pixel size 25 x 100 µm (3D or planar)
  - Active components integrated into the chip

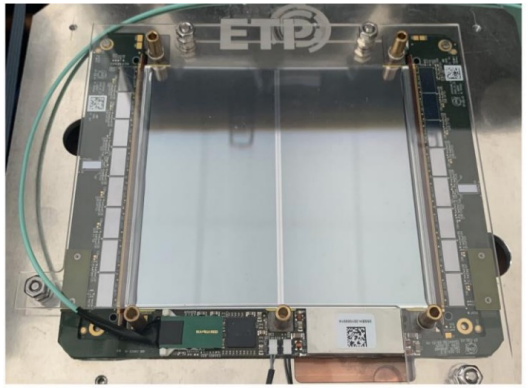


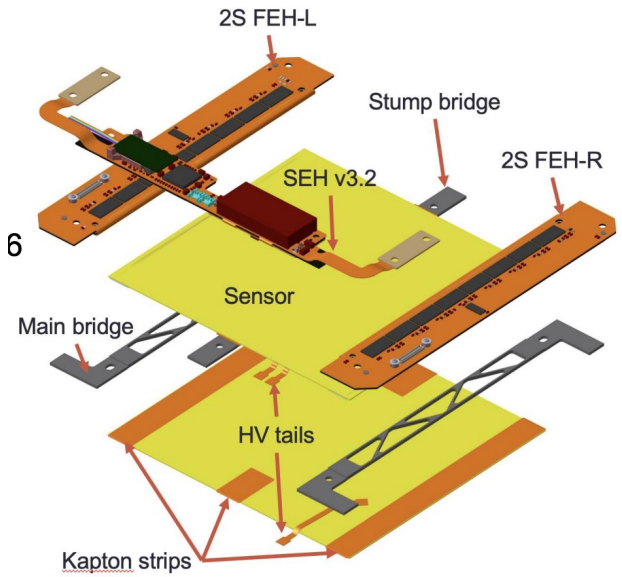


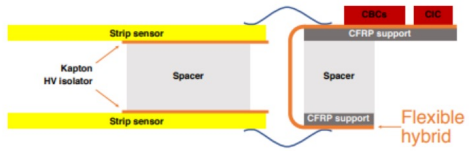


Outer Tracker Phase-2 upgrade

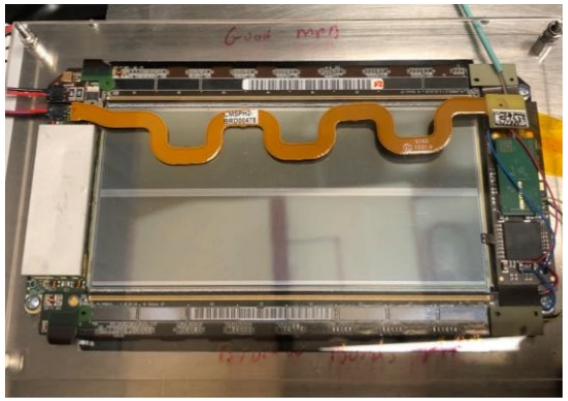
- 2S module (outer layers)
  - 10 x 10 cm
  - 90μm pitch, 5 cm strip length (z)
  - 4x 1016 strips (2x per layer)
  - Coarse z coordinate

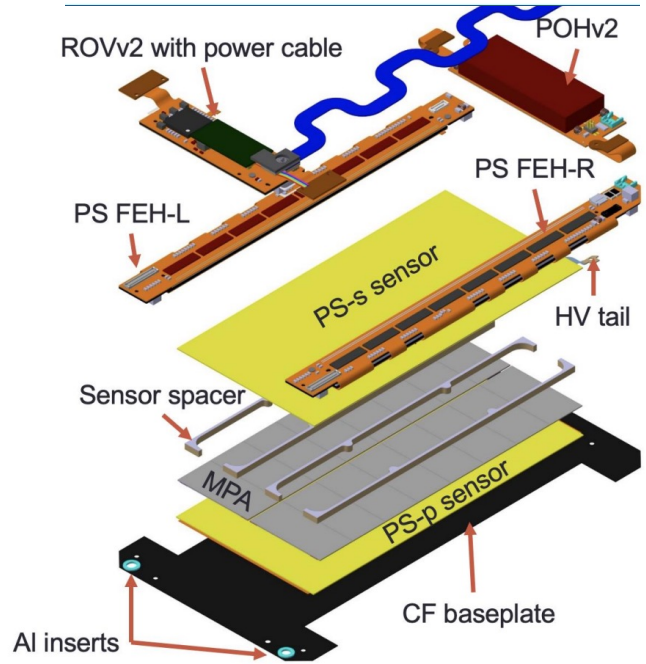


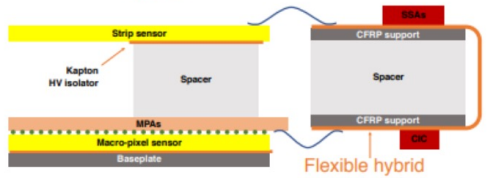




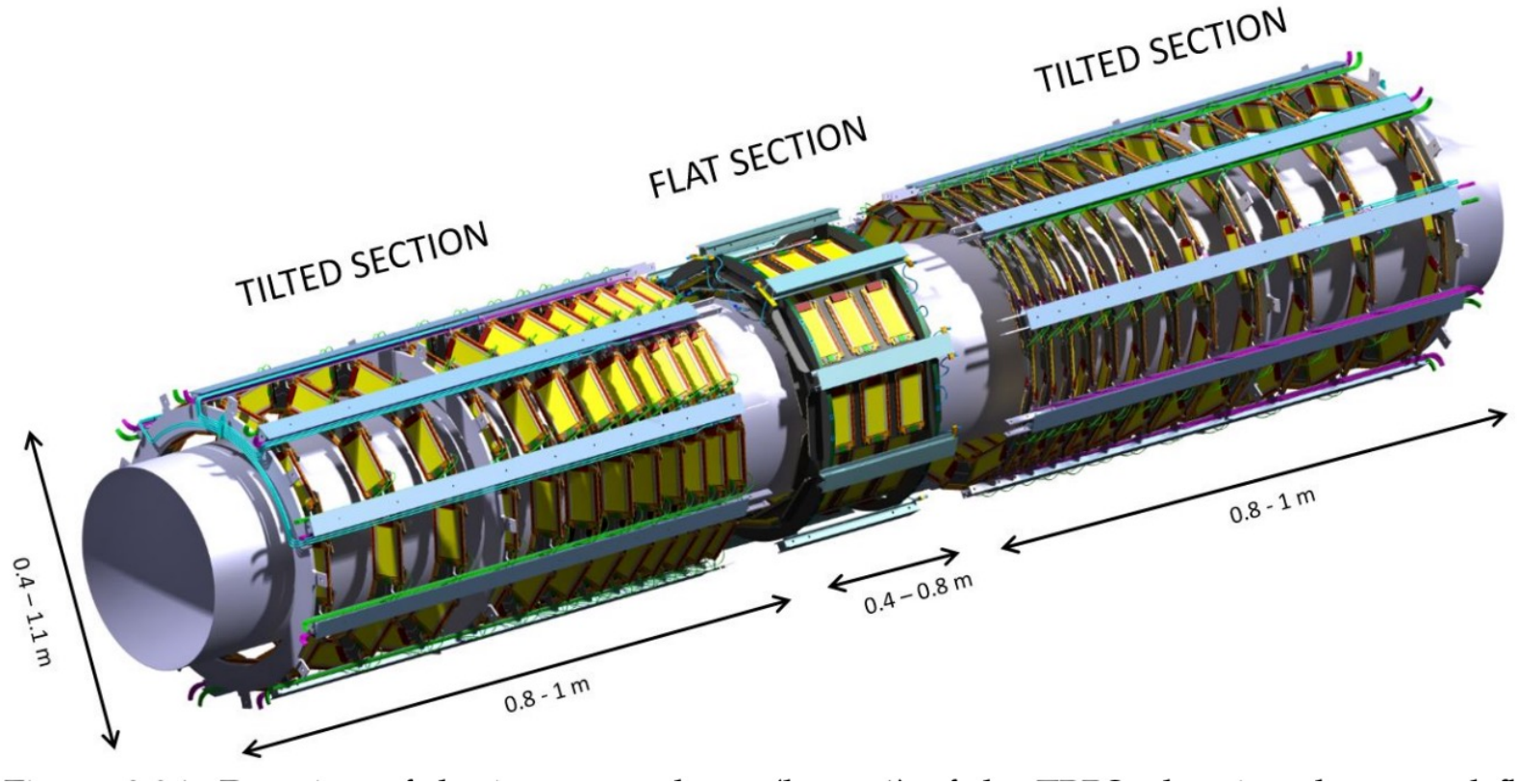
- PS module (inner layers)
  - 5x10 cm
  - PS-p (pixel) layer
    - MacroPixel 1.5mm x 100 μm
    - 32x960 pixels
  - PS-s (strip) layer
    - Length 2.5 cm (z)
    - 100 μm pitch
    - 2x960 strips
  - Accurate z-coordinate



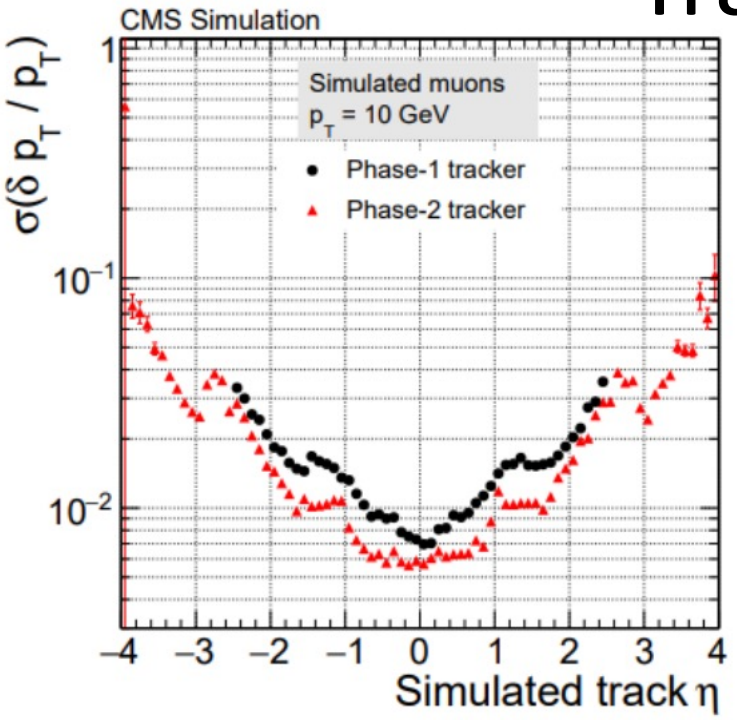


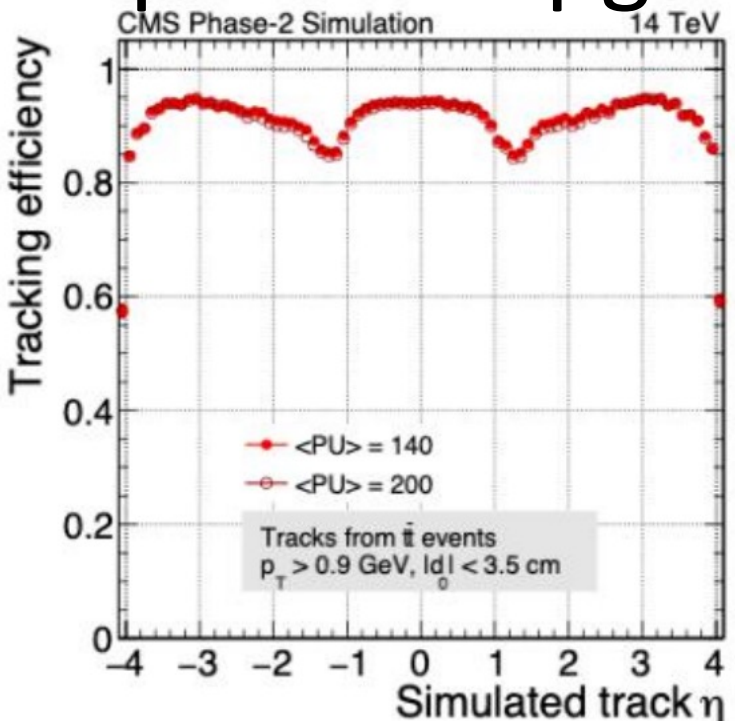


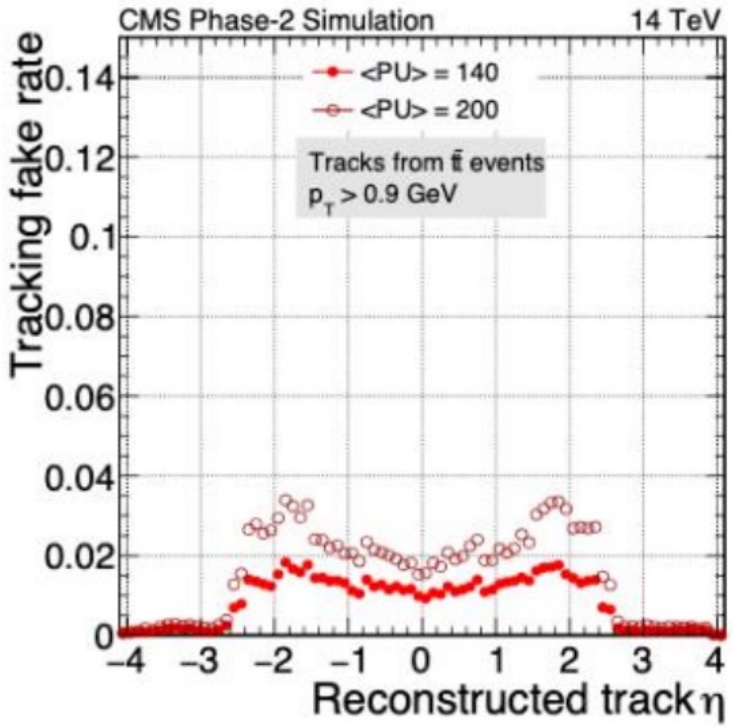
- Inner layer of the Phase-2 Outer Barrel Tracker



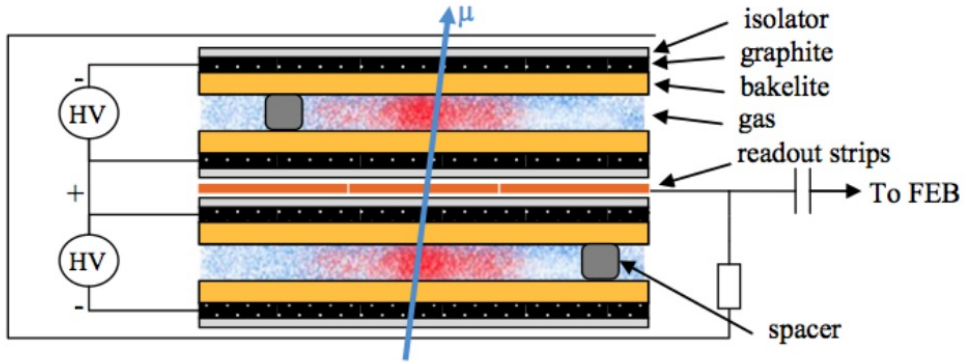
Tracker phase 2 upgrade







	CMS Outer Tracker Phase 1	CMS Outer Tracker Phase 2
Total surface	~ 206 m2	~ 191 m2
# Modules	15148	13296
# Strips	9.3 M	42 M
# Macro-pixels	0	172.5 M
L1 trigger rate	100 kHz	750 kHz
Latency	3.2 μs	12.5 μs
Powering	direct	DC –DC converters
Cooling	C_6F_{14}	CO_2



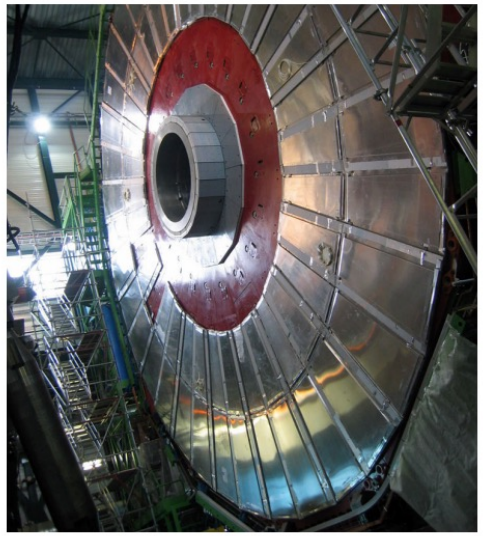
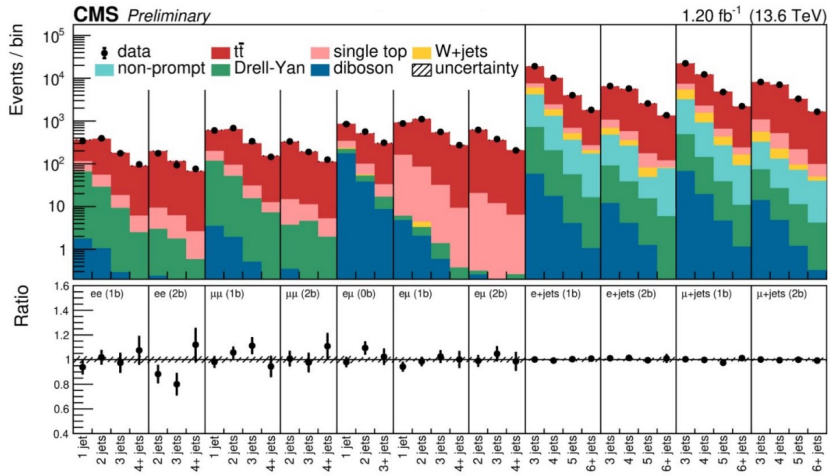
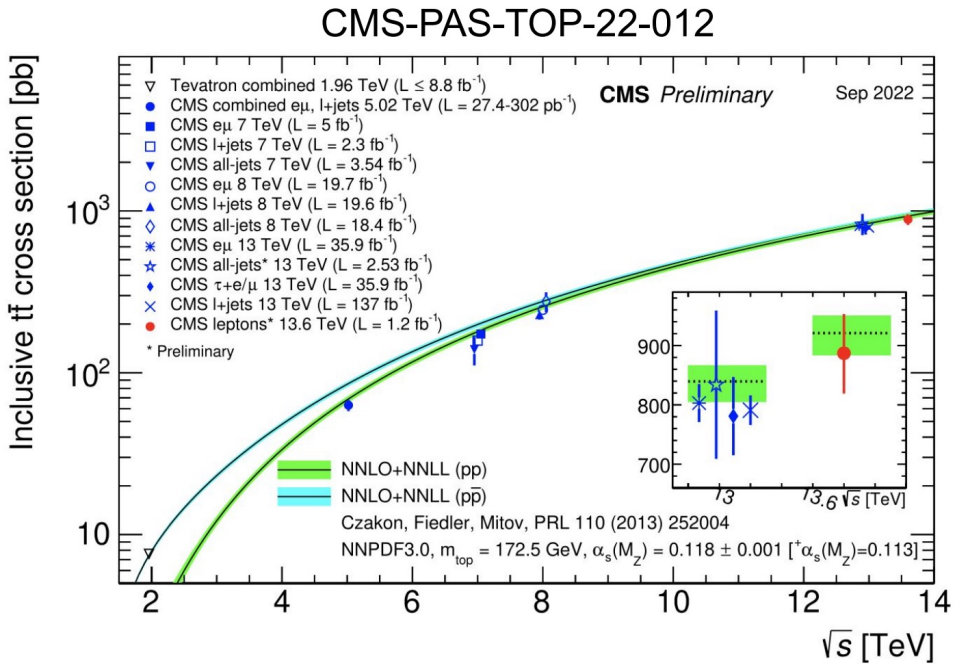


Figure 1.7: Left: Working principle of the double gap RPCs in CMS [23]. Right: RPC endcap chambers RE1/2 and RE1/3 after installation. The iron yoke is shown in red.

Early results at 13.6 TeV





- Result: $\sigma = 887^{+38}_{-42}(\text{stat+sys}) \pm 55(\text{lumi})$ pb
- Consistent with SM prediction: 921^{+29}_{-37} pb