CMS DAQ for Phase-II

Trigger, Online and Offline Computing Workshop

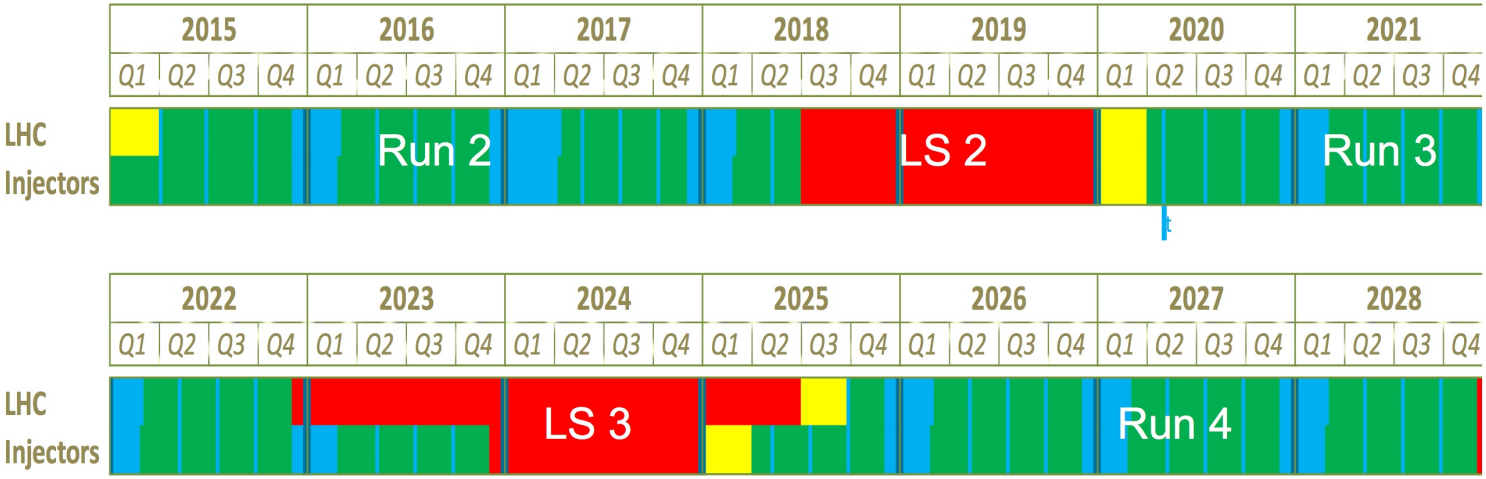
4-5 Sep 2014, CERN

Frans Meijers for CMS

Outline:

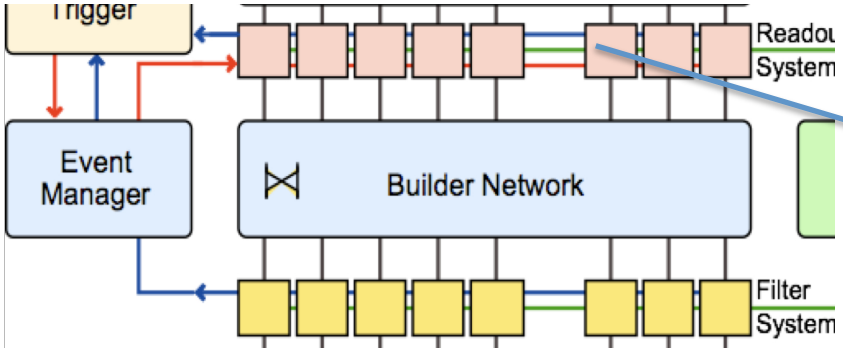
- DAQ for CMS Phase-II detector at HL-LHC
- Chapter “DAQ and Trigger Control” in the CMS Phase-II Technical Proposal

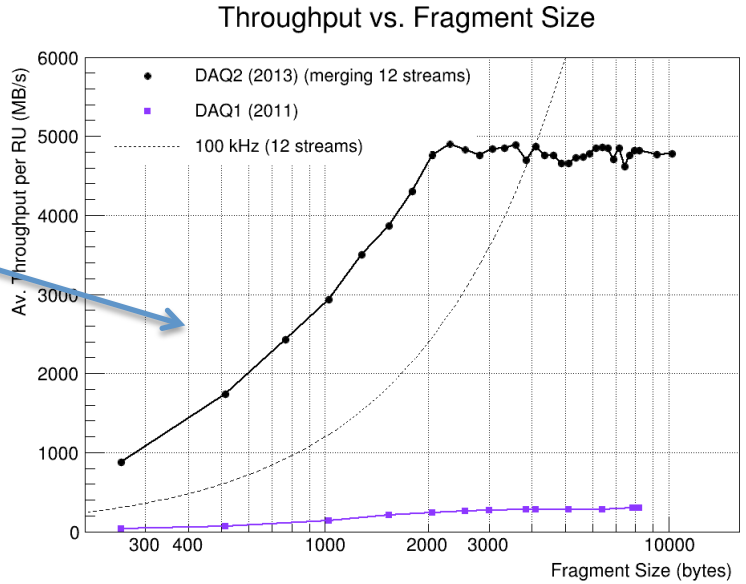
Timescale



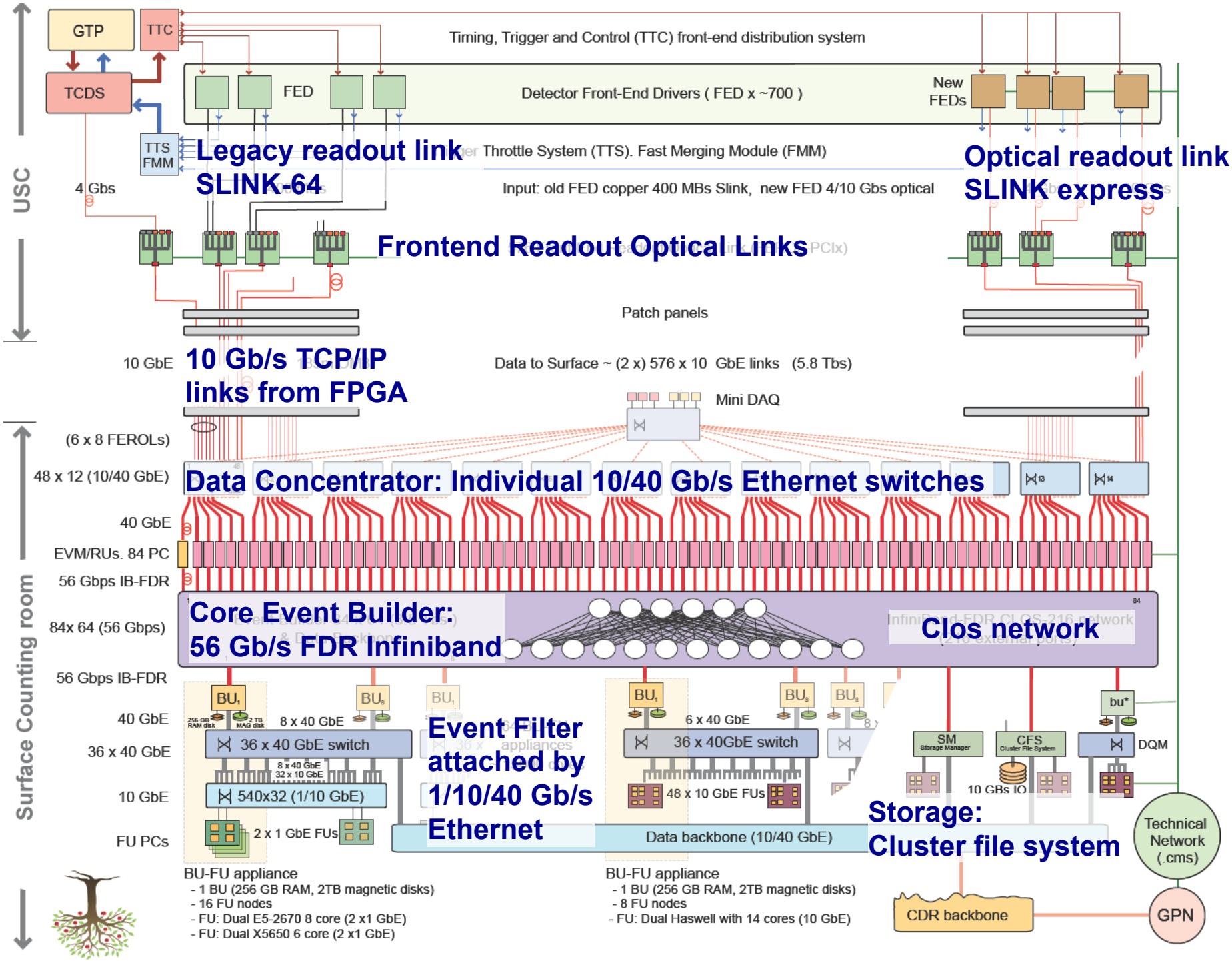
- Start of run-4 physics in 2025 is very far away
  - >10 years, two 5-year replacement cycles
    - Note: will change DAQ + HLT farm system again in LS2 (5 years old in 2019)
- Technology evolves ~exponentially (at fixed cost ?)
  - FPGA and serialisers
  - Networking
  - Computing (server PCs)
  - Storage

From DAQ1 (2007) to DAQ2 (2014)

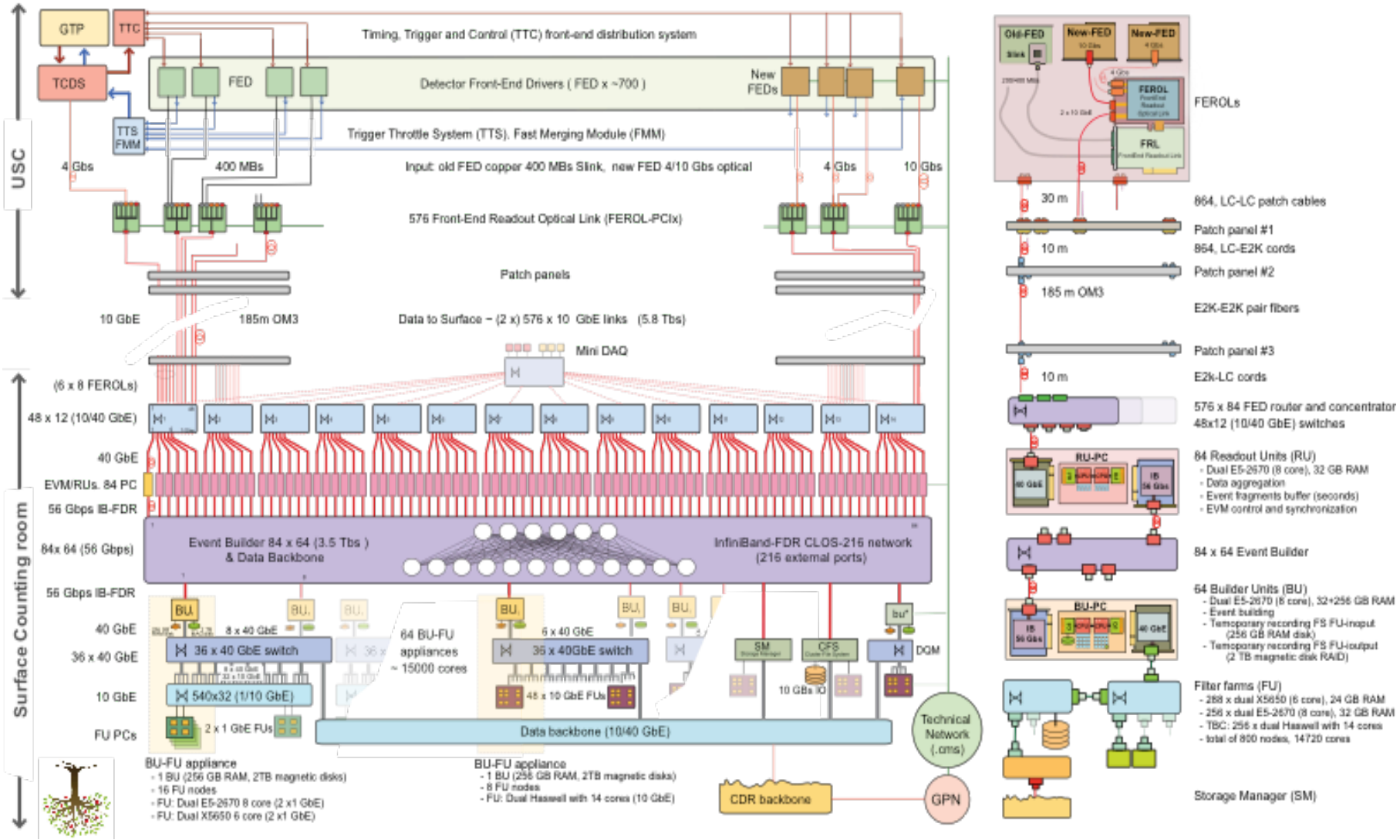




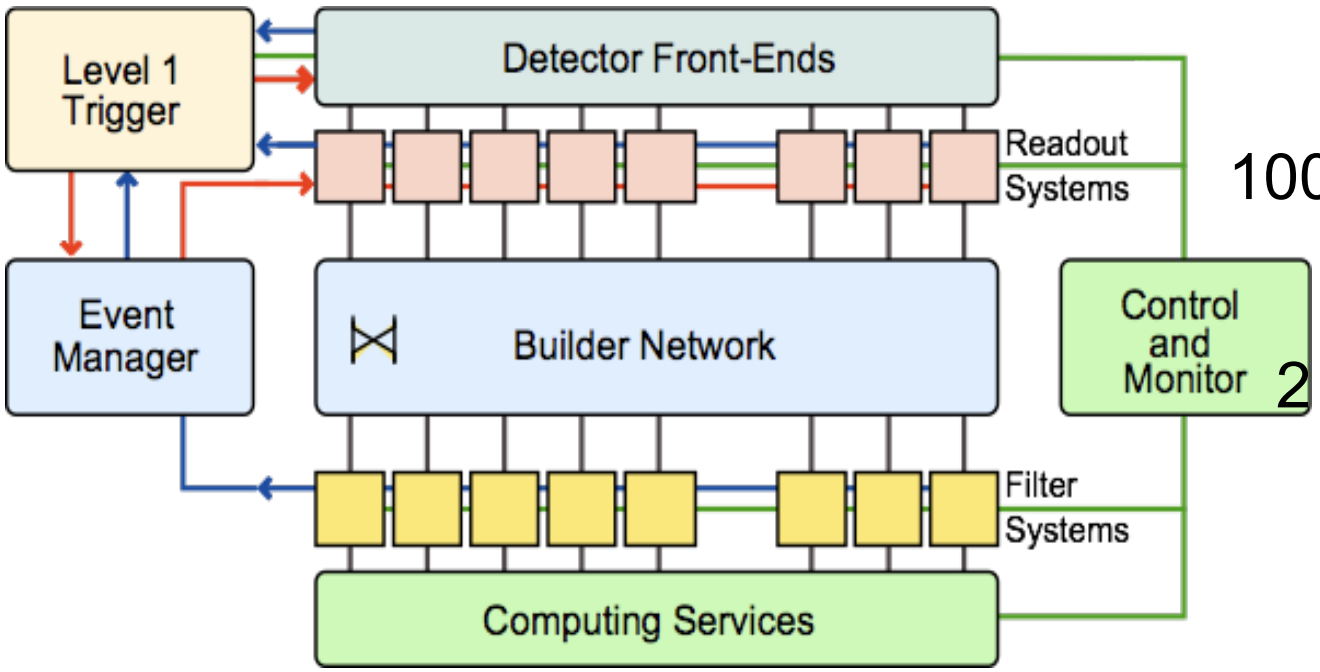
- DAQ1: Myrinet (2x2 GBps) and Ethernet (3x1 Gbps)
- DAQ2: 40 GbE and IB FDR 56 Gbps
- Factor >10x in throughput per DAQ “read-out” node


USC
Surface Counting room
GTP
TTC
TCDS
TTS FMM
Timing, Trigger and Control (TTC) front-end distribution system
FED
Detector Front-End Drivers (FED x ~700)
New FEDs
Throttle System (TTS). Fast Merging Module (FMM)
Input: old FED copper 400 MBs Slink, new FED 4/10 Gbs optical
4 Gbs
Legacy readout link
SLINK-64
Optical readout link
SLINK express
Frontend Readout Optical Links
Patch panels
10 GbE
10 Gb/s TCP/IP links from FPGA
Data to Surface ~ (2 x) 576 x 10 GbE links (5.8 Tbs)
Mini DAQ
(6 x 8 FEROLs)
48 x 12 (10/40 GbE)
Data Concentrator: Individual 10/40 Gb/s Ethernet switches
40 GbE
EVM/RUs. 84 PC
56 Gbps IB-FDR
84x 64 (56 Gbps)
Core Event Builder:
56 Gb/s FDR Infiniband
Clos network
56 Gbps IB-FDR
40 GbE
8 x 40 GbE
36 x 40 GbE
36 x 40 GbE switch
8 x 40 GbE
32 x 10 GbE
10 GbE
540x32 (1/10 GbE)
FU PCs
2 x 1 GbE FUs
Event Filter attached by 1/10/40 Gb/s Ethernet
6 x 40 GbE
36 x 40GbE switch
48 x 10 GbE FUs
Data backbone (10/40 GbE)
SM
Storage Manager
CFS
Cluster File System
10 GBs IO
bu*
DQM
Storage:
Cluster file system
Technical Network (.cms)
GPN
CDR backbone
BU-FU appliance
- 1 BU (256 GB RAM, 2TB magnetic disks)
- 16 FU nodes
- FU: Dual E5-2670 8 core (2 x1 GbE)
- FU: Dual X5650 6 core (2 x1 GbE)
BU-FU appliance
- 1 BU (256 GB RAM, 2TB magnetic disks)
- 8 FU nodes
- FU: Dual Haswell with 14 cores (10 GbE)

CMS DAQ for LHC run 2



Two level CMS DAQ system



40 MHz

100 kHz -> 750 kHz

2 Tbps -> ~50 Tbps

1 kHz -> ~7.5 kHz

CERN-LHCC-2014-XXX
CMS-XXX-XXX
XX September 2014

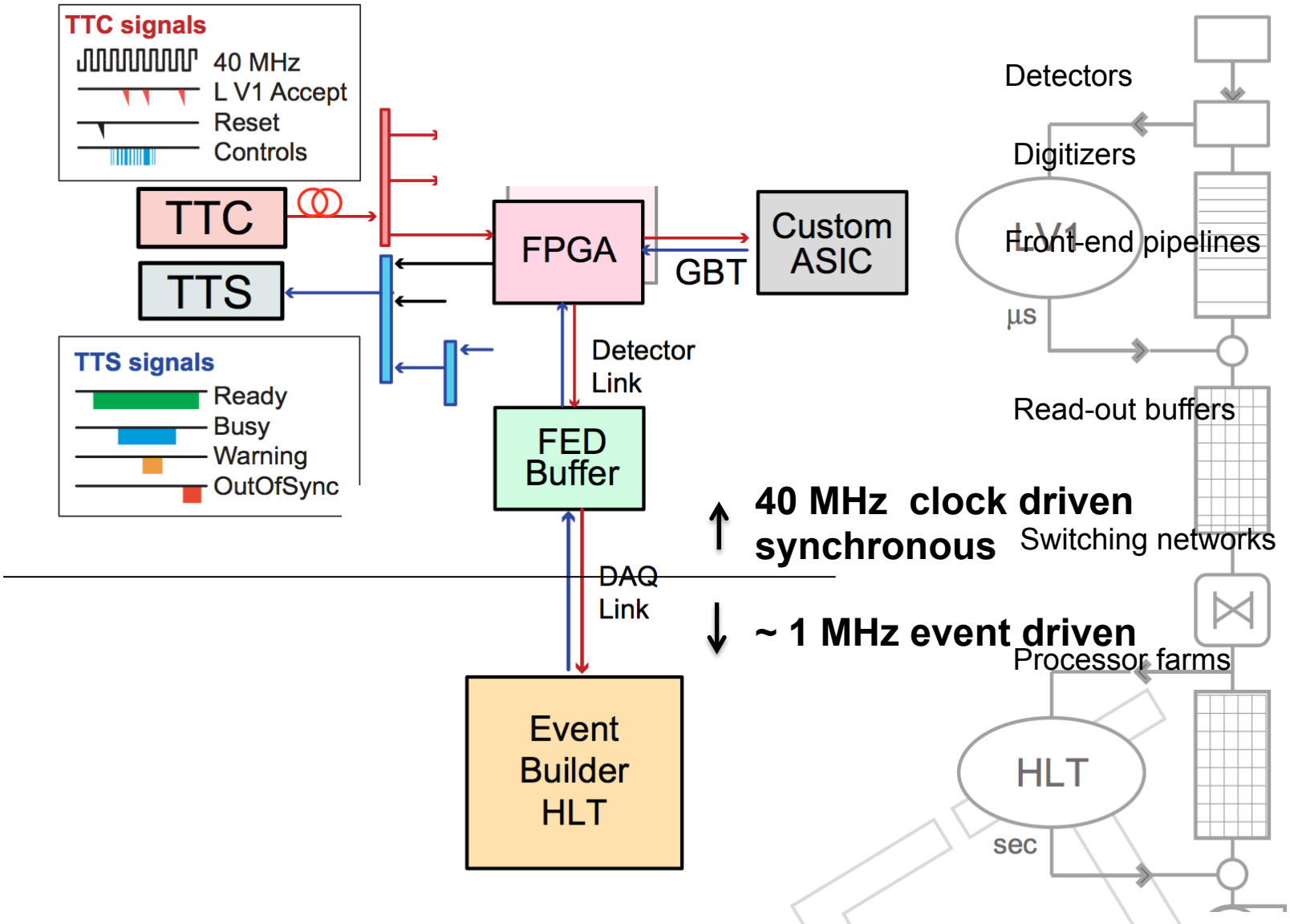
TECHNICAL PROPOSAL FOR THE PHASE-II UPGRADE OF THE COMPACT MUON SOLENOID

This Technical Proposal presents the upgrades foreseen to prepare the CMS experiment for the High Luminosity LHC. In this second phase of the LHC physics program, the accelerator will provide to CMS an additional integrated luminosity of about 2500 fb^{-1} over 10 years of operation, starting in 2025. This will substantially enlarge the mass reach in the search for new particles and will also greatly extend the potential to study the properties of the Higgs boson discovered at the LHC in 2012. In order to meet the experimental challenges of unprecedented p-p luminosity, the CMS collaboration will need to address the aging of the present detector and to improve the ability of the apparatus to isolate and precisely measure the products of the most interesting collisions. This document describes the conceptual designs and the expected performance of the upgrades, along with the plans to develop the appropriate experimental techniques. The infrastructure upgrades and the logistics of the installation in the experimental area are also discussed. Finally, the initial managerial organization and cost estimates of the upgrades are presented.

DAQ/HLT parameters

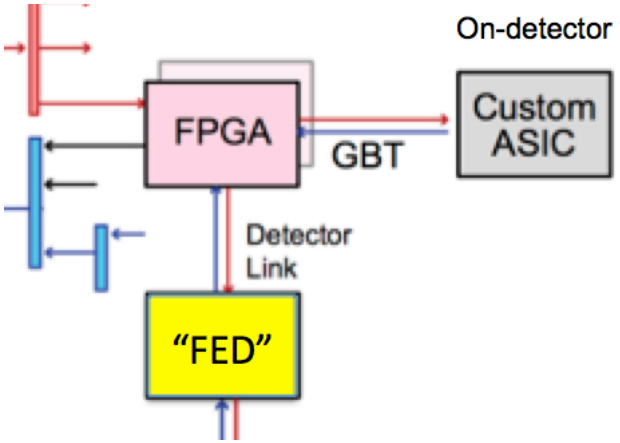
	LHC Run-I	LHC Phase-I upgr.	HL-LHC Phase-II upgr.	
Energy	7-8 TeV	13 TeV	13 TeV	
Peak Pile Up (Av./crossing)	35	50	140	200
Level-1 accept rate (maximum)	100 kHz	100 kHz	500 kHz	750 kHz
Event size (design value)	1 MB	1.5 MB	4.5 MB	5.0 MB
HLT accept rate	1 kHz	1 kHz	5 kHz	7.5 kHz
HLT computing power	0.2 MHS06	0.4 MHS06	6 MHS06	13 MHS06
Storage throughput (design value)	2 GB/s	3 GB/s	27 GB/s	42 GB/s

System Overview R/O electronics, DAQ and HLT



Sub-detector DAQ

- There is a lot of communality ..



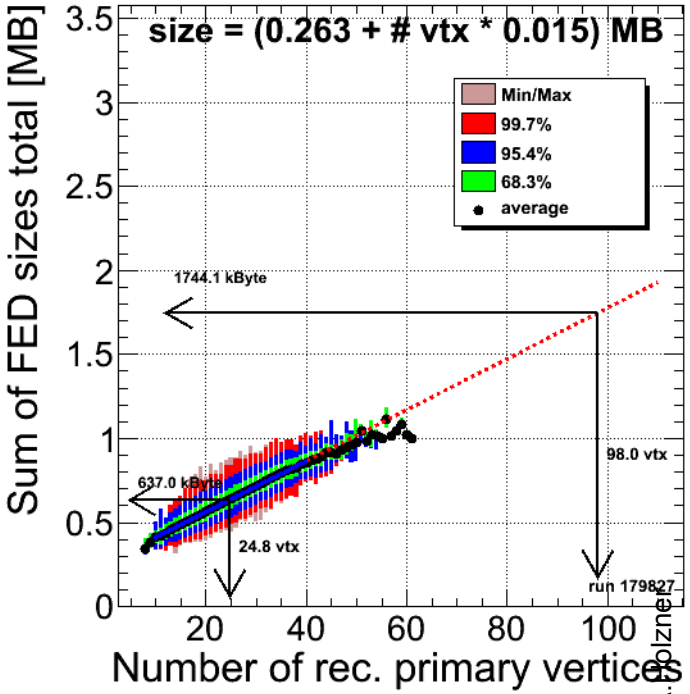
- Differences
  - Type of links between on- and off-detector electronics
    - Many will use GBT/VL at 5/10 Gbps
  - Splitting of Trigger and DAQ data at
    - on-detector (hence different links), or
    - off-detector
  - Sending full data at 40 MHz from on-detector electronics or sparsified
  - Yes/No data reduction in off-detector electronics
  - …

Sub-detector parameters

sub-detector	#links on- to off-detector	type	use	data reduction	estimated event size	# DAQ links (100 Gbps)
TK-outer	15 k	GBT (3.2 Gbps)	DAQ + Trig 20%+80%	on-det.	0.5-0.6 MB	100
TK-pixel	2 k	GBT (6.4 Gbps)	DAQ only	on-det.	0.7-1.0 MB	200
ECAL-barrel	12 k	GBT (3.2 Gbps)	streaming	off-det.	1.2 MB	200
HCAL			streaming	off-det.	0.2 MB	40
HB	1296	GBT (3.2 Gbps)				
HF	864	GBT (3.2 Gbps)				
HO	768	GOL (1.0 Gbps)				
Forward-calo						
option Shaslik	12 k	GBT (3.2 Gbps)	streaming	off-det	1.2 MB	200
option HGCal	6 k	10 Gbps	DAQ only	on-det ?	1 MB	200
Muons DT	5760	GBT (3.2 Gbps)	streaming	off-det.	0.1 MB	20
Muons CSC	1080	GBT (3.2 Gbps)	DAQ + Trig 50%+50%	off-det.	0.1 MB	20
Trigger	N/A	N/A	N/A	N/A	0.1 MB	20
Event Builder					≈4.1-4.5 MB	≈ 800

Event Size at EVB level

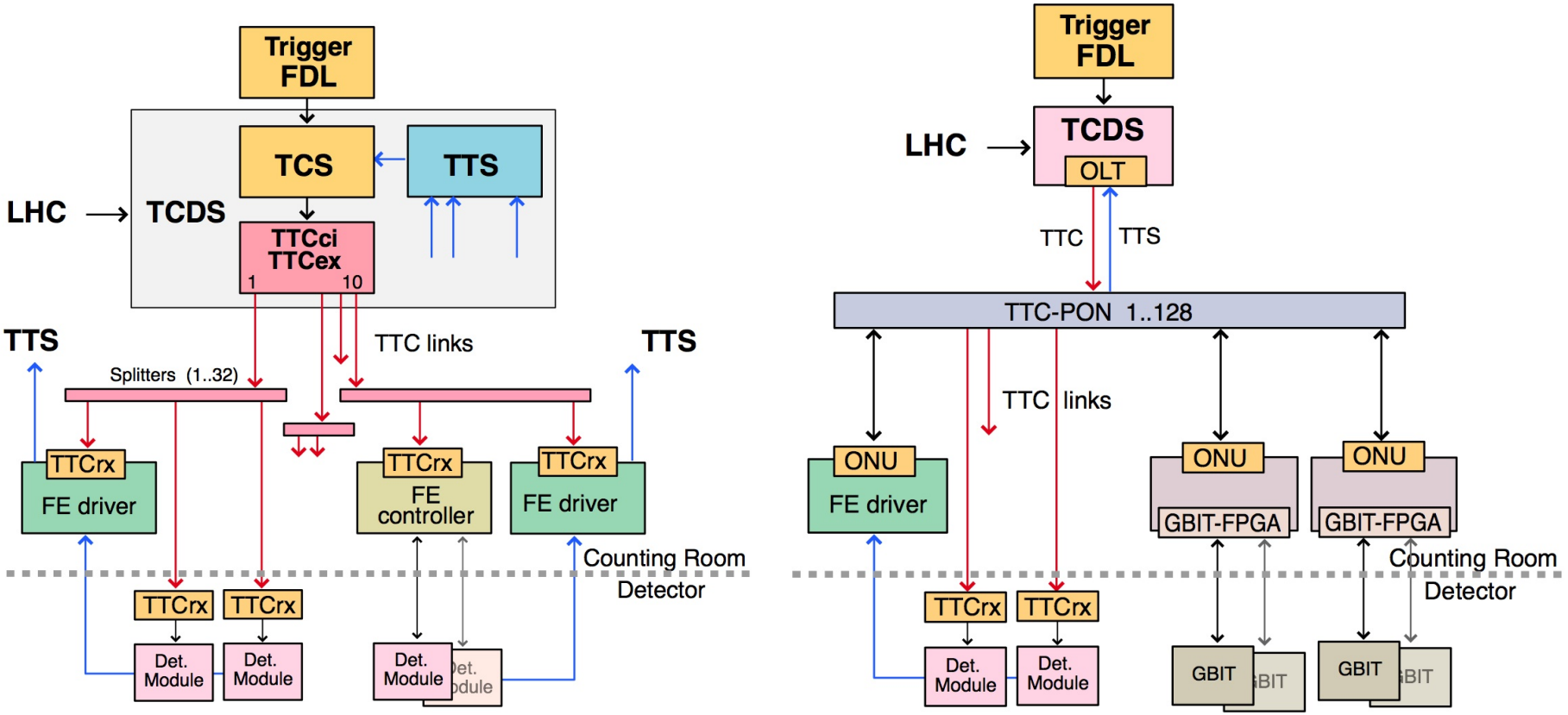
- Run1 detector
  - Linear with PU
  - Dominated by TK



Courtesy A. Holzner

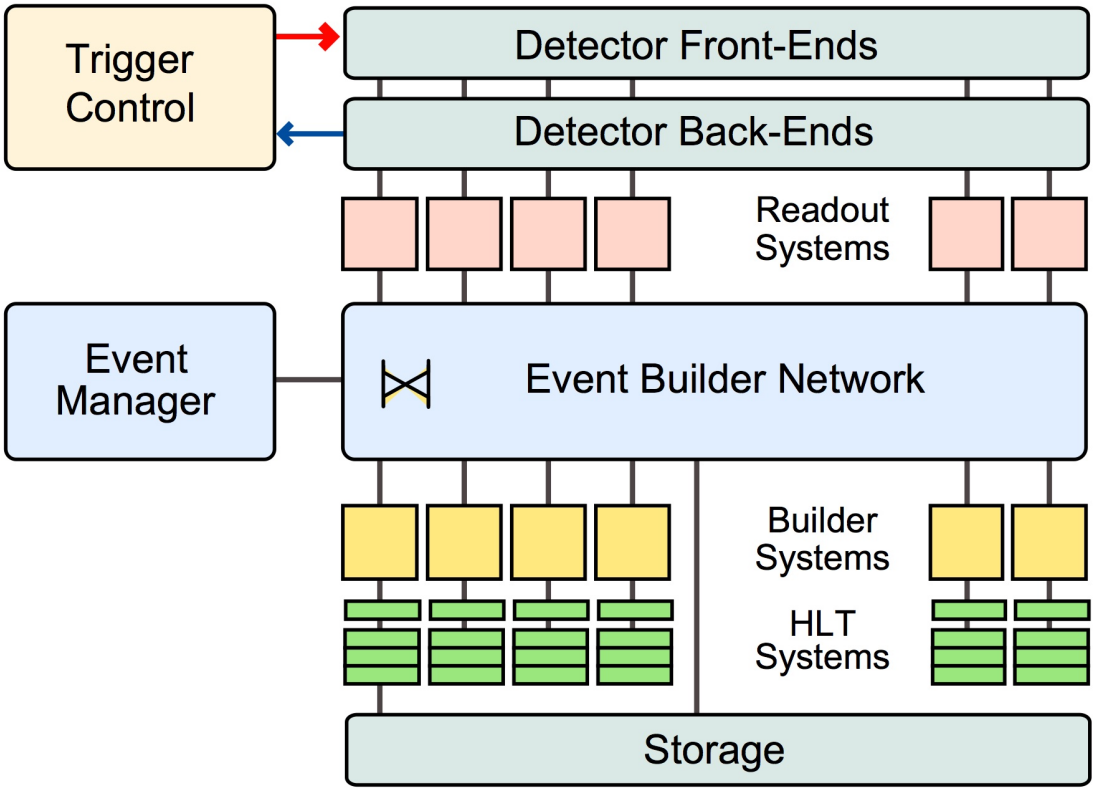
- Phase-II detector
  - Mostly new detector and R/O electronics and/or strategy
  - Dominated by calorimeters
  - Calorimeters: ~constant with PU
    - To be decided on “selective readout”
  - Trackers: PU = 140 – 200 outer: 0.5 – 0.6 MB, pixel: 0.7 – 1.0 MB

Trigger Control and Distribution



- PON technology (commercial components):
  - Bi-directional
  - Downstream 100x TTC BW (could send 216 bit per BX)
    - Control readout and/or event building

EVB, HLT facility and Storage



- Same design as current system, nodes and networks with better performance
  - HLT on fully built events
- EVB: e.g. 800 links with 100 Gbps, 33% efficiency, 80 Tbps switch

Areas of R&D

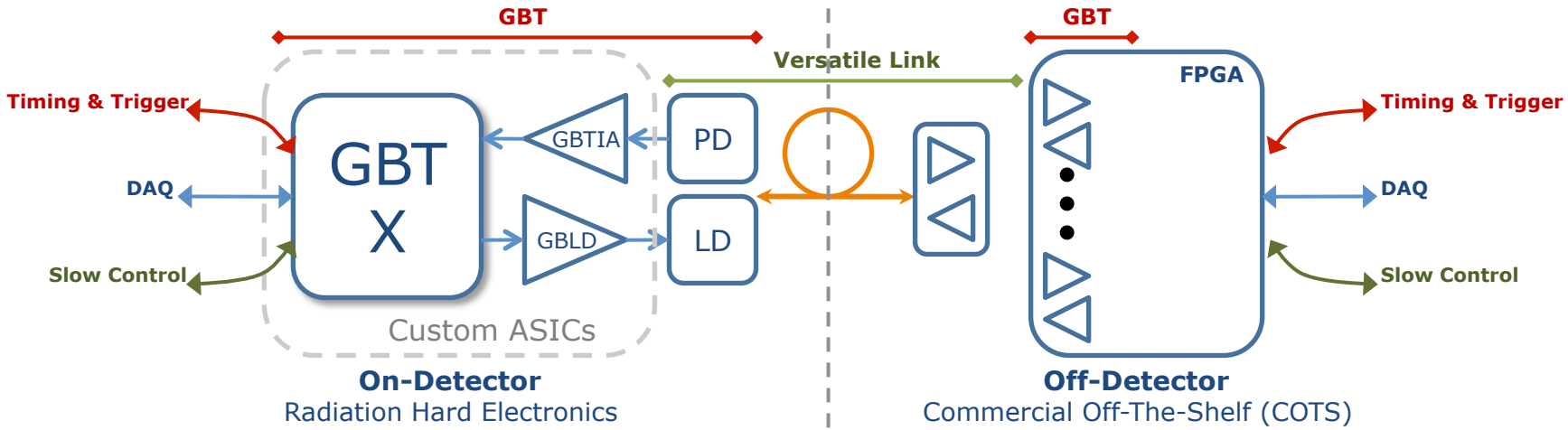
- Areas
  - FPGA, links and serialisers
  - Networking for event building and event distribution
  - Processor nodes for DAQ
  - Processor nodes for HLT
  - Storage
  - Software

Conclusion

- DAQ/HLT in the context of overall CMS
  - DAQ/HLT is not the “limiting” system; limit comes from
    - Sub-detector front-end electronics
    - Off-line processing capacity
  - Timescales
    - DAQ/HLT equipment replaced each 4-5 years
    - LS3 is a period for such replacement, new equipment will accommodate additional resources for Phase-II subdet and L1 upgrades
    - TDR 5 years before installation in LS4: ~2019, provided ongoing R&D program
    - Required performance re-iterated and trade-off HLT / offline can be made then

EXTRA

GBT / VL



- Upstream 5 / 3.2 Gbps (low power) or 10 / 6.4 Gbps signalling / effective
- Downstream 5 Gbps