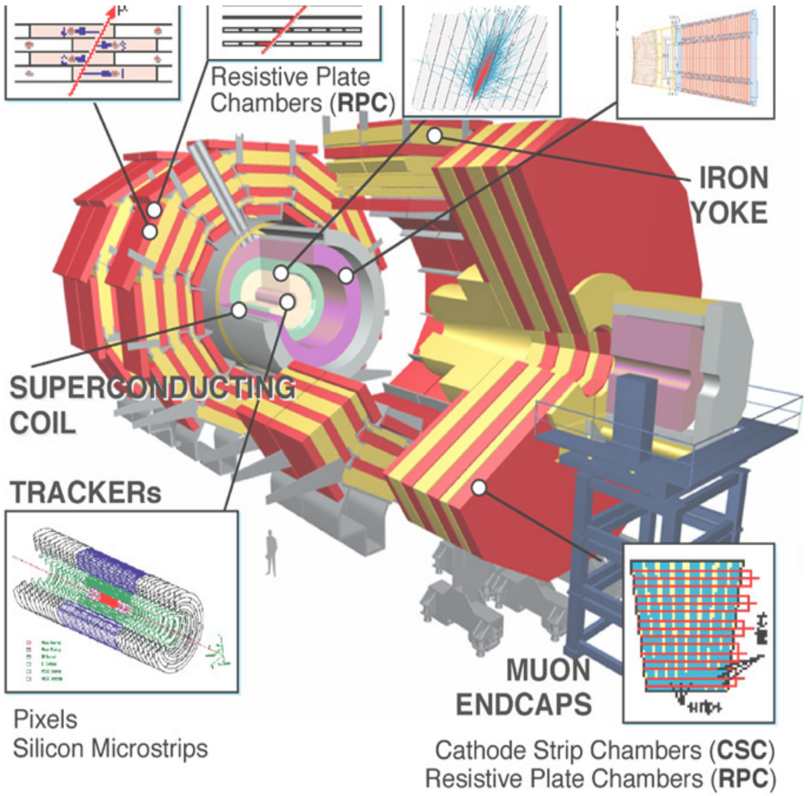
Overview of CMS Upgrades and DAQ

DAQ@LHC workshop 12-14 Mar 2013

Frans Meijers (CERN-PH-CMD)
On behalf of the CMS DAQ group

CMS design parameters and DAQ requirements

Detectors

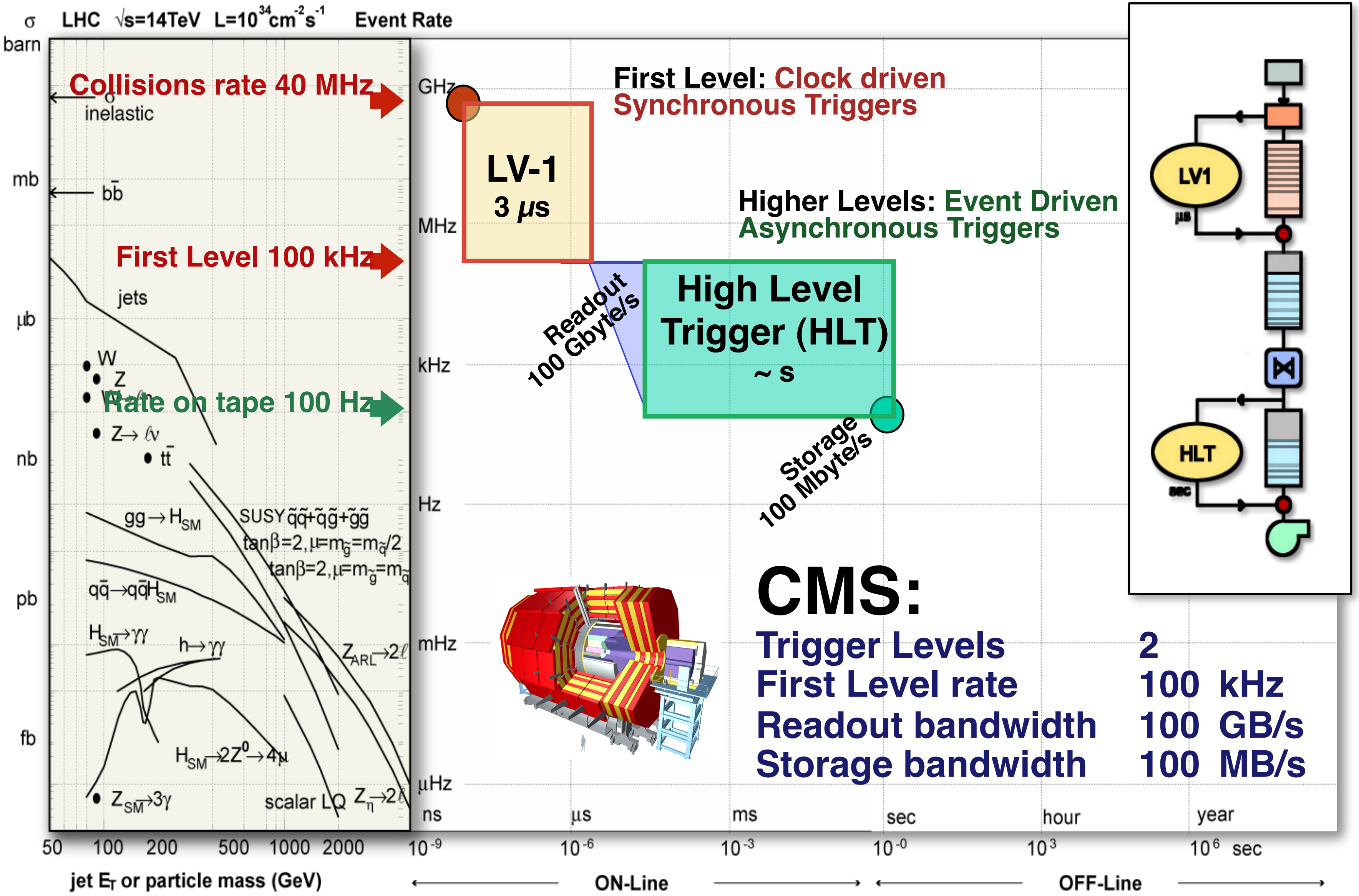


Total weight : 12,500 t
Overall diameter : 15 m
Overall length : 21.6 m
Magnetic field : 4 Tesla

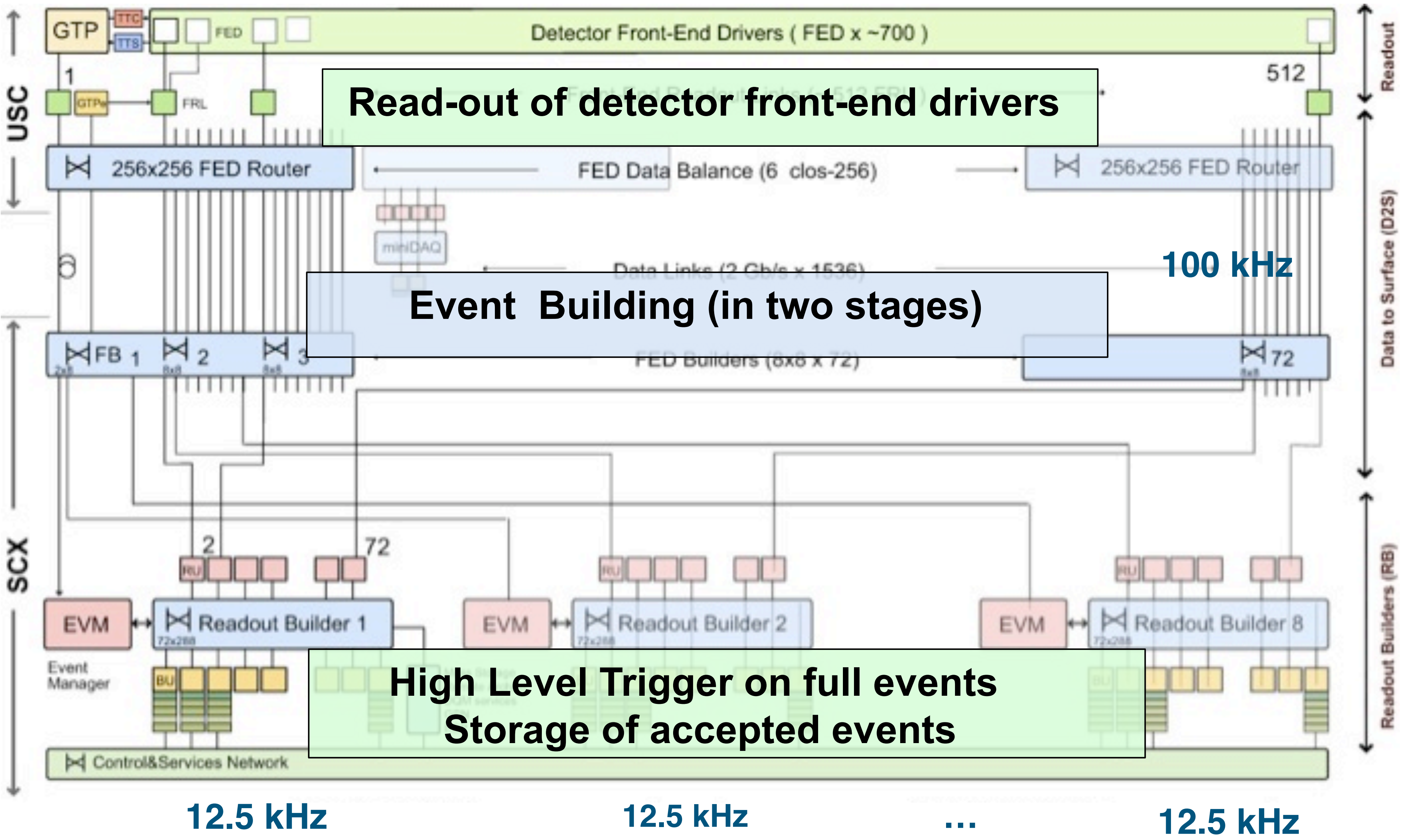
Detector	Channels	Control	Ev. Data
Pixel	60000000	1 GB	50 (kB)
Tracker	10000000	1 GB	650
Preshower	145000	10 MB	50
ECAL	85000	10 MB	100
HCAL	14000	100 kB	50
Muon DT	200000	10 MB	10
Muon RPC	200000	10 MB	5
Muon CSC	400000	10 MB	90
Trigger		1 GB	16

Average Event size	1 Mbyte
Max LV1 Trigger	100 kHz
Online rejection	99.999%
System dead time	~ %

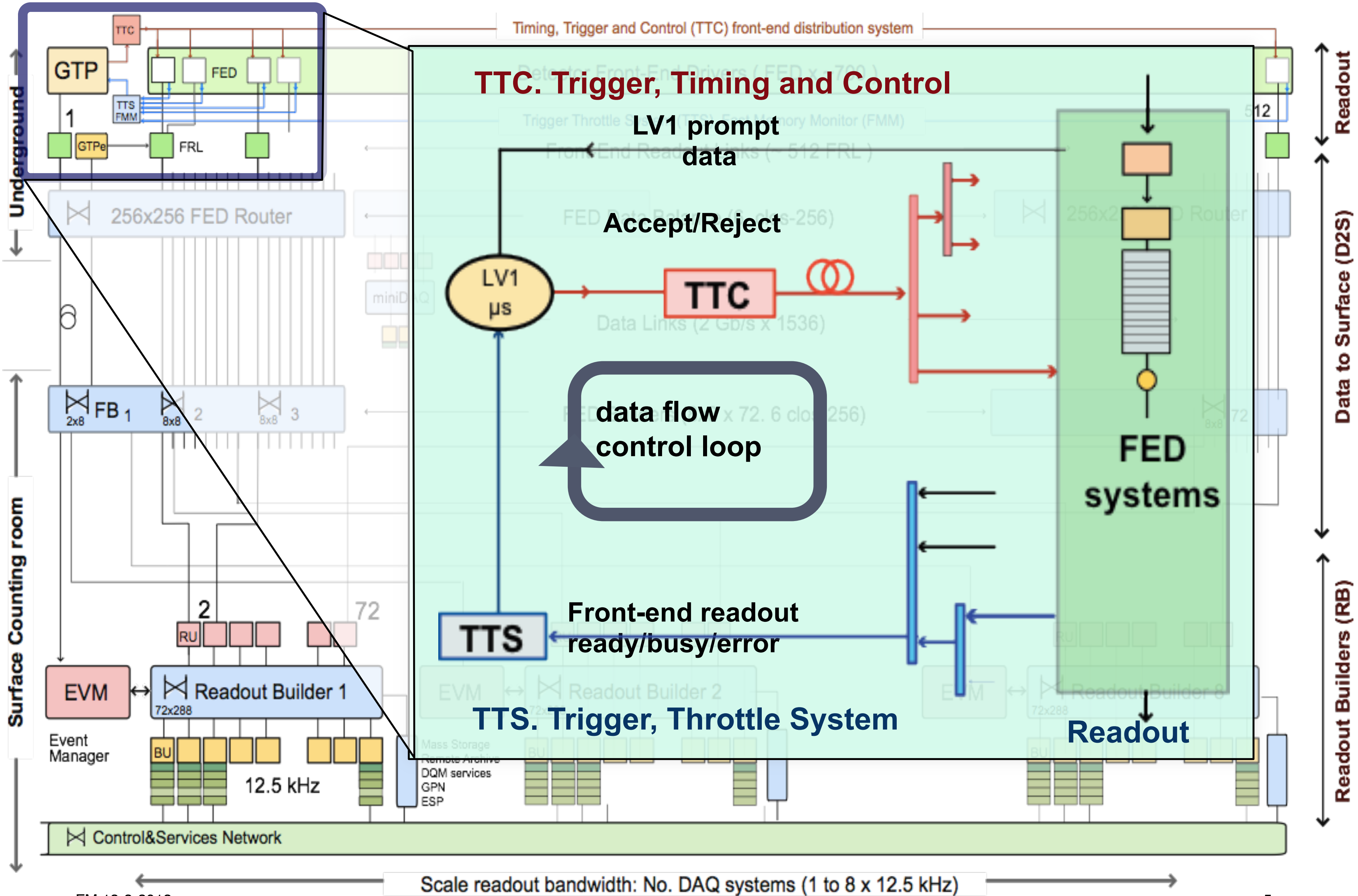
Two Trigger levels



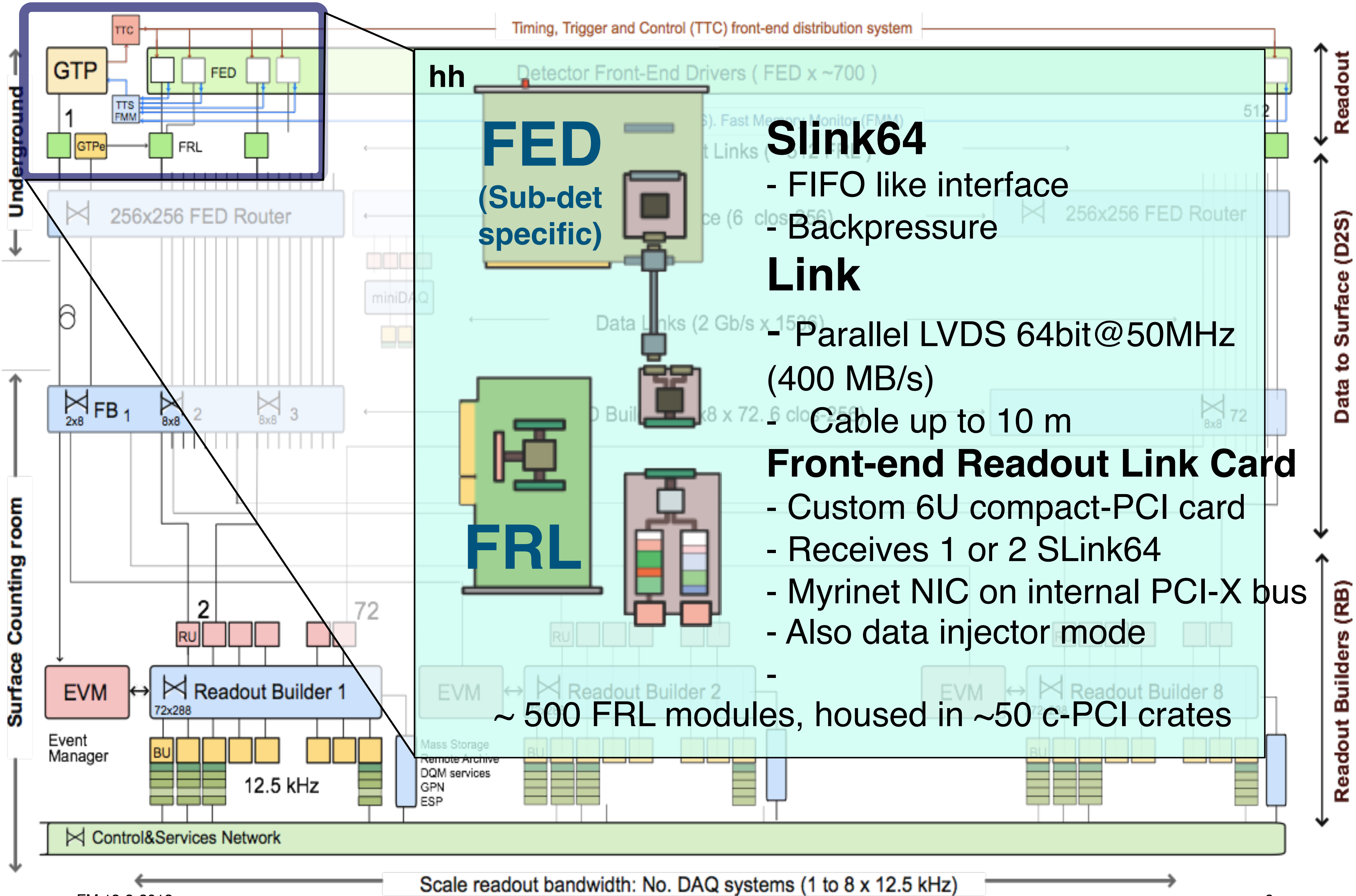
CMS DAQ1 (2008-2012)



Front-end model – lossless DAQ



Uniform interface - Readout



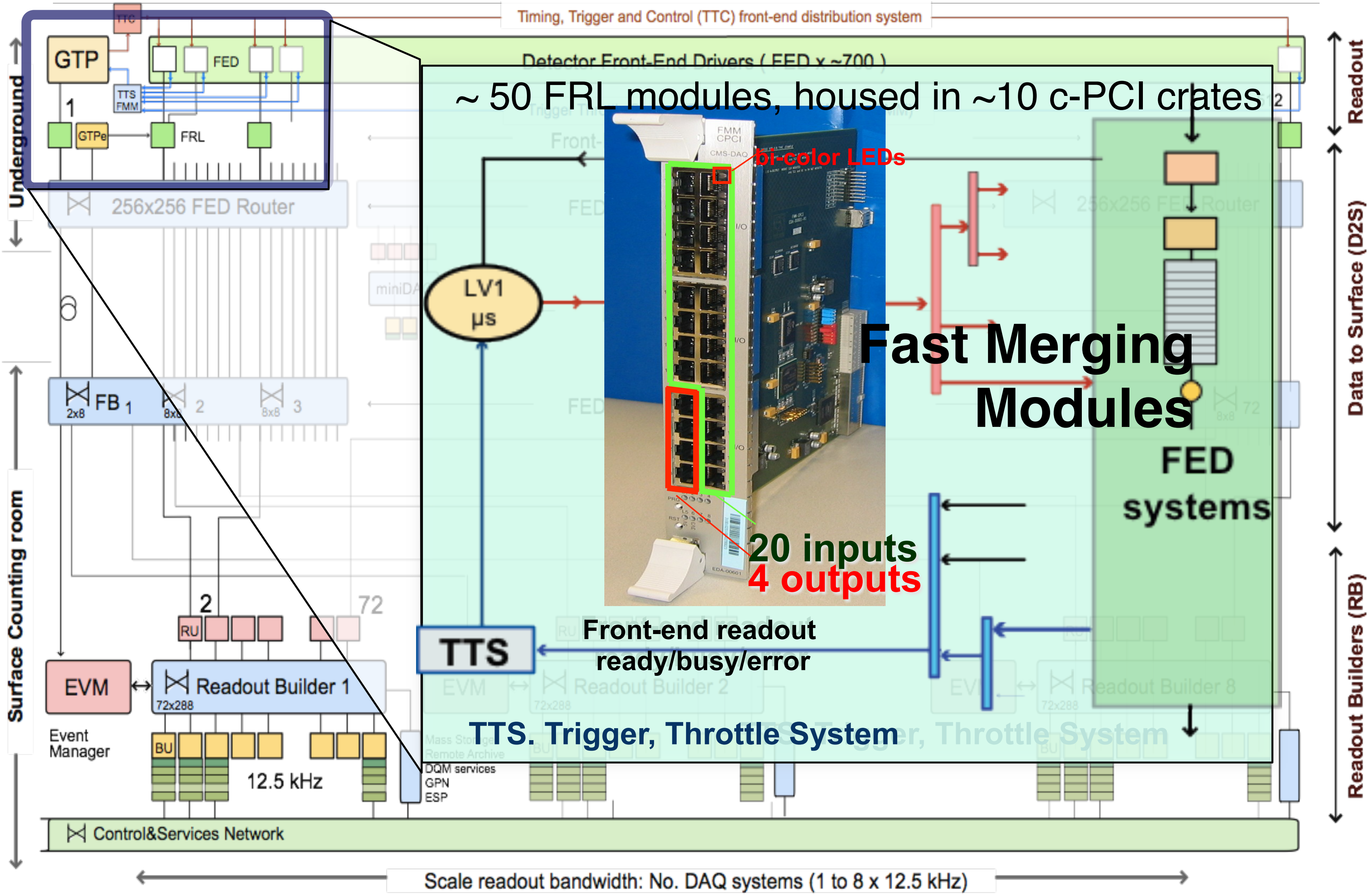
Slink64
- FIFO like interface
- Backpressure

Link
- Parallel LVDS 64bit@50MHz (400 MB/s)
- Cable up to 10 m

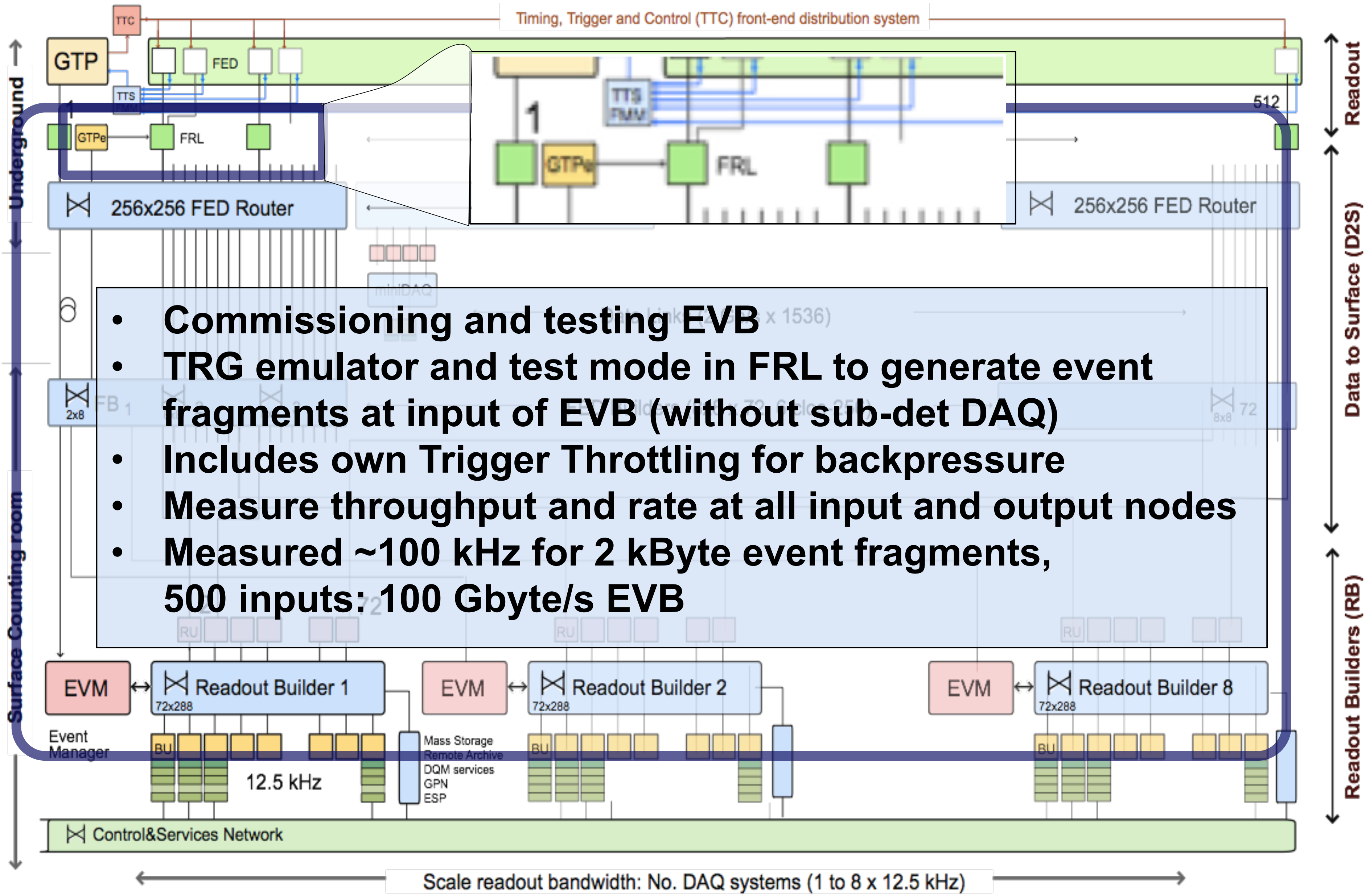
Front-end Readout Link Card
- Custom 6U compact-PCI card
- Receives 1 or 2 SLink64
- Myrinet NIC on internal PCI-X bus
- Also data injector mode
-

~ 500 FRL modules, housed in ~50 c-PCI crates

Uniform Interface – TTS

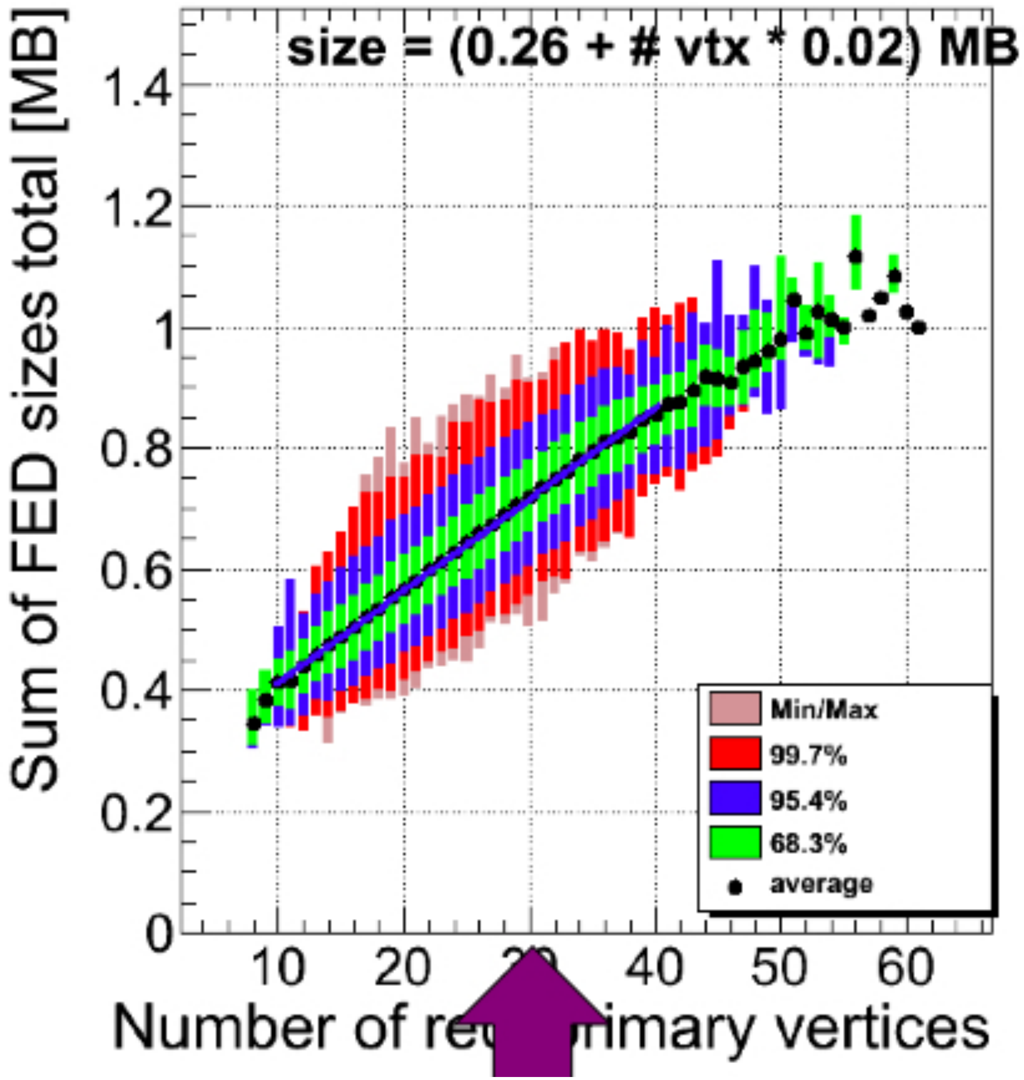


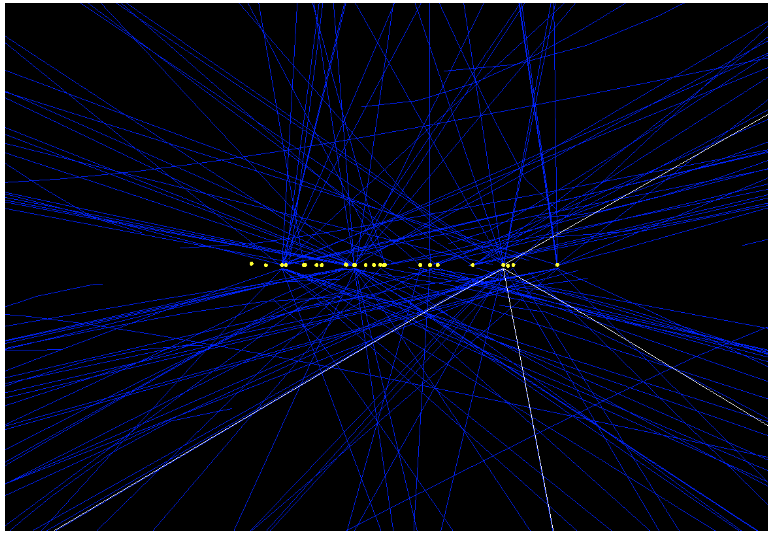
Full-EVB and emulator mode



- Commissioning and testing EVB
- TRG emulator and test mode in FRL to generate event fragments at input of EVB (without sub-det DAQ)
- Includes own Trigger Throttling for backpressure
- Measure throughput and rate at all input and output nodes
- Measured ~100 kHz for 2 kByte event fragments, 500 inputs: 100 Gbyte/s EVB

Event Size vs Pileup (50ns)





- Due to acceptance: number of reconstructed vertices = ~0.7 PileUp

DAQ @ LHC: Introduction

- Upgrade time line and terminology

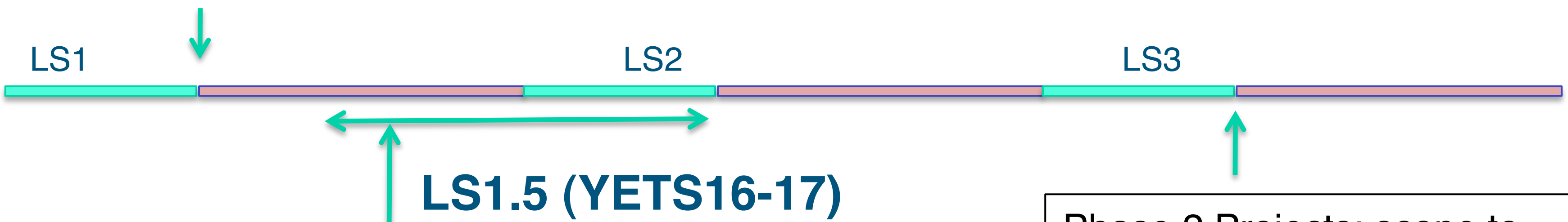
…	2010 2011 2012 2013 2014	2015 2016 2017 2018	2019 2020 2021 2022 2023	… 2030
	Phase 0 Run 1 / LS1	Run 2 / LS2	Phase I Run 3 / LS3	Phase II Run 4
	(Prepare Run 2)	(Prepare Phase I)	(Prepare Phase II)	
	Consolidation	Ultimate luminosity $\sqrt{s}$ = 13~14 TeV 25 ns bunch spacing	HL-LHC	
ATLAS / CMS	L_{inst} 1 x10^{34} $cm^{-2}s^{-1}$ $\mu \sim 27$ $\int L_{inst} \sim 50\ fb^{-1}$	L_{inst} 2-3 x10^{34} $cm^{-2}s^{-1}$ $\mu \sim$ 55–81 $\int L_{inst} > 350\ fb^{-1}$	L_{inst} 5 x10^{34} $cm^{-2}s^{-1}$ $\mu \sim 140$ [with levelling] L_{inst} 6-7 x10^{34} $cm^{-2}s^{-1}$ $\mu \sim 192$ [without levelling] $\int L_{inst} \sim 3000\ fb^{-1}$	
LHCb	L_{inst} 4-6 x10^{32} $cm^{-2}s^{-1}$ $\mu \sim 1.8$ [with levelling] $\int L_{inst} > 10\ fb^{-1}$	L_{inst} 1-2 x10^{33} $cm^{-2}s^{-1}$ $\mu \sim$ 4–6 [with levelling] $\int L_{inst} \sim 50\ fb^{-1}$		
ALICE	L_{inst} 1-2 x10^{27} $cm^{-2}s^{-1}$ [with levelling] $\int L_{inst} > 1\ nb^{-1}$	$\int L_{inst} > 10\ nb^{-1}$		

CMS Upgrade program

Scope described in Technical Proposal for the Upgrade of the CMS detector through 2020
http://cdsweb.cern.ch/record/1355706 LHCC-2011-006

Three stages

LS1 Projects: in production
- Completion of muon coverage (ME4)
- Improve muon operation (ME1), DT electronics
- Replace HCAL photo-detectors in HF (new PMTs) and HO (HPD→SiPM)

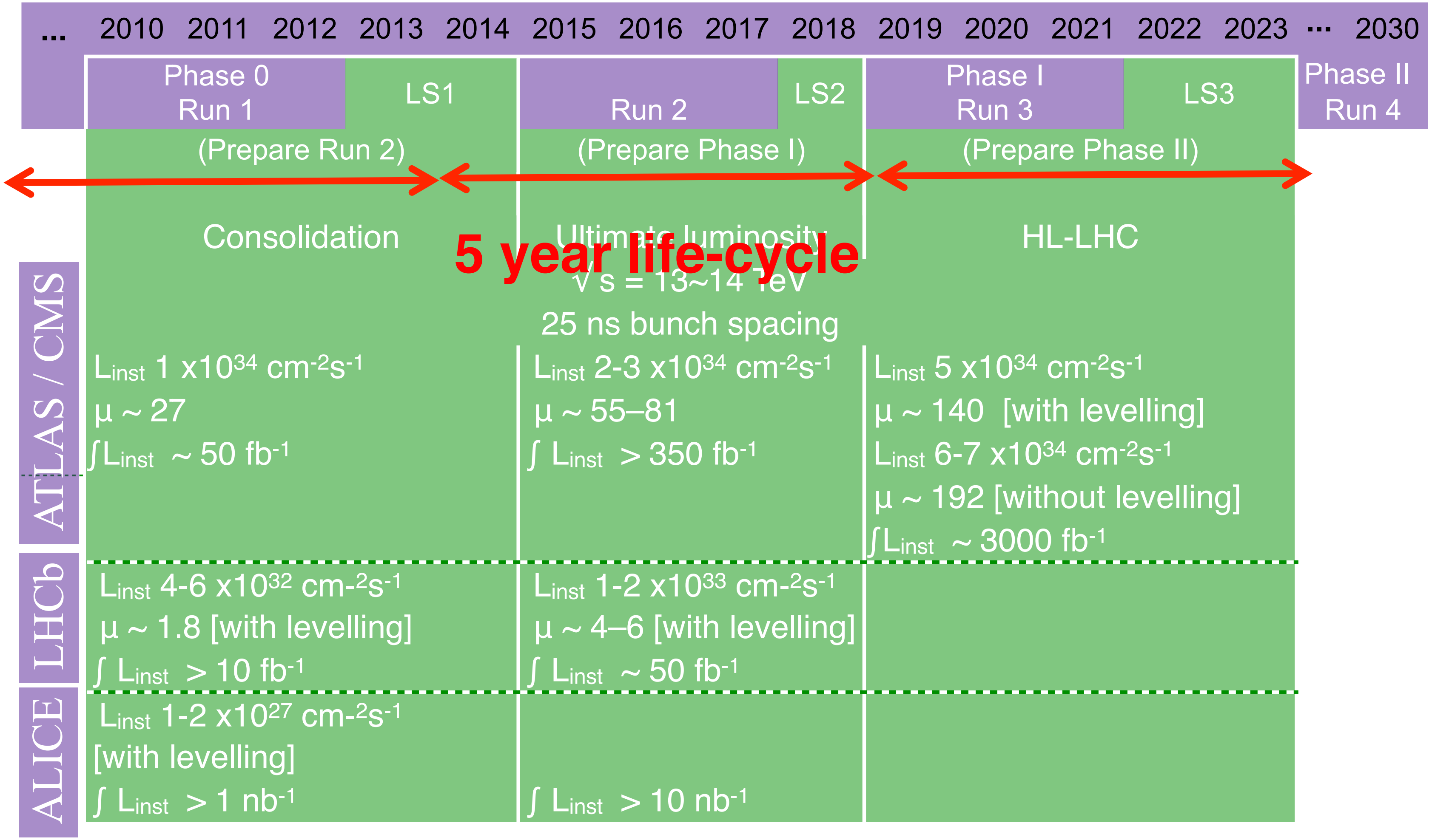


Phase 1 Upgrades:
- Pixel detector replacement (YETS16-17)
- HCAL electronics upgrade
- L1-Trigger upgrade

Phase 2 Projects: scope to be defined in Technical Proposal (2014)
- Tracker Replacement
- Track Trigger
- Forward Calorimetry and Muons?
- Further Trigger upgrade?

DAQ @ LHC: Introduction

- Upgrade time line and terminology



~Phase-I
Post – LS1
Run 2

DAQ2 for post-LS1

- **Equipment replacement cycle**
  - PC and network replaced typically each 5 years
- **New requirements**
  - Increased data sizes due to higher pile-up
  - Some sub-detectors will be replaced which lead to higher data volumes
    - Eg HCAL sensors, new Pixel
  - Some sub-det new back-end electronics in uTCA standard with serial link to cDAQ
  - Data aggregation of links with a range of 1–10 Gbps throughput
- **Keep "external" boundaries**
  - Inputs (custom electronics)
    - About 500 2-4 Gbps "Legacy" FEDs
    - About 20-100 6-10 Gbps New FEDs
  - Output to HLT farm
    - About 500 "Legacy" nodes with ~1-2 Gbps input
    - About 400 new nodes ~4 Gbps input
- **Need also to operate for Heavy-Ion conditions**

DAQ2 for post-LS1

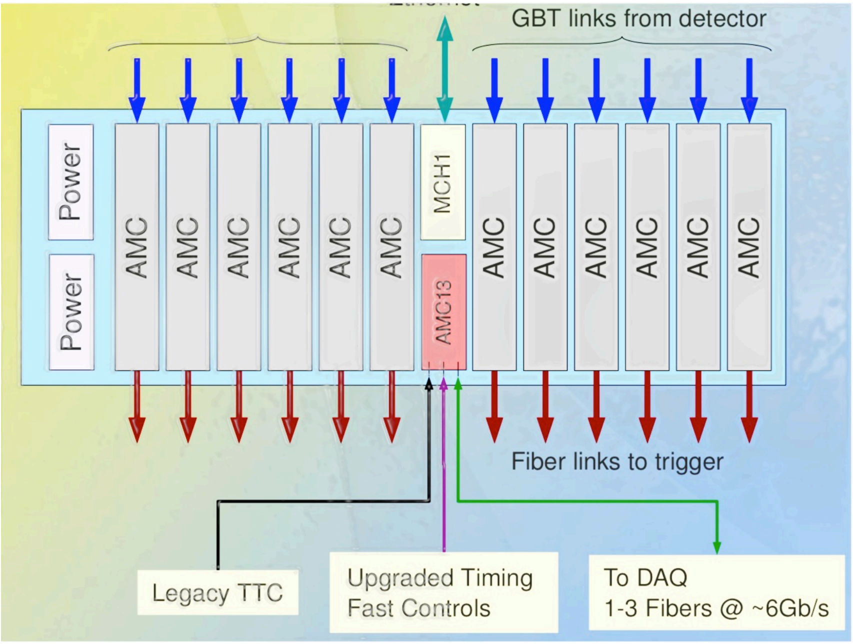
- **TCDS (Trigger Control and Distribution System)**
  - Need for more TTC partitions
  - Rationalisation of
    - TCS (Trigger Control System)
    - TTC
    - TTS (Trigger Control System)

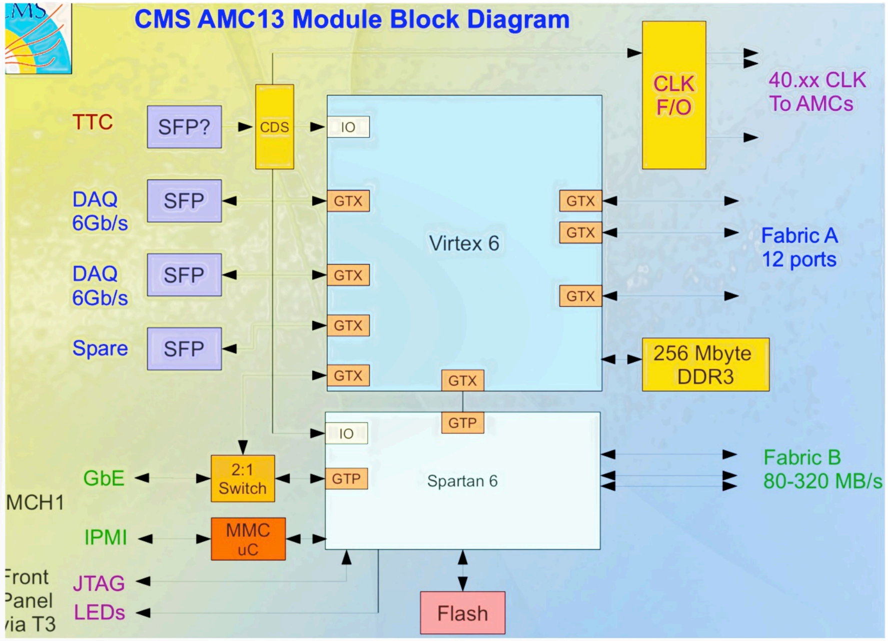
- **Re-visit implementation of lumi-section**
  - Example of a feature which was introduced as an afterthought after TDR

- **File based HLT**
  - Take advantage of advances of storage technology (in speed)
  - Write full EVB output of 100 kHz to storage (for ~1 m)
    - Absorb the HLT initialisation time
  - Full decoupling of two frameworks (XDAQ and CMSSW)

- **Possibility to use HLT farm as a cloud resource for "offline processing"**

uTCA based off-detector electronics





- Development by BU (Boston University) for HCAL
- This structure is also considered for some of the Trigger sub-systems
- AMC13 might evolve in to CMS “common platform”

- AMC13 sends data to central DAQ over multi-gbps serial link (6 Gbps in prototype)
- (P2P) Protocol for data link to central DAQ has been developed

L1 trigger upgrade (I)

Trigger Upgrade: The Plan

Upgrade the Calo, Muon and Global Triggers

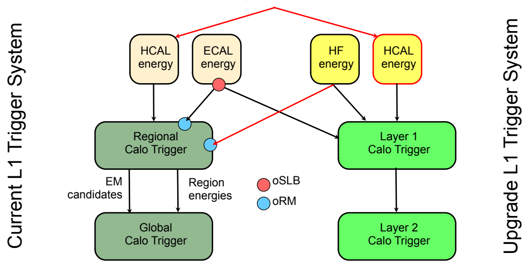
- architecture highly configurable, based mainly on 3 boards (with large FPGA, high bandwidth optics, memory for LUTs)
- parallel commissioning of new trigger while operating present trigger
- goal to provide improvements for 2015, commission full functionality for 2016

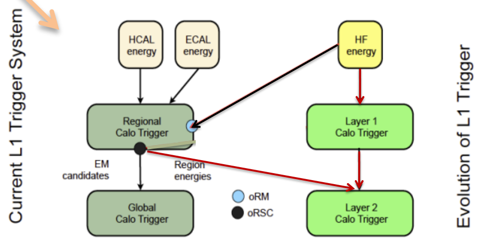
Trigger Improvements

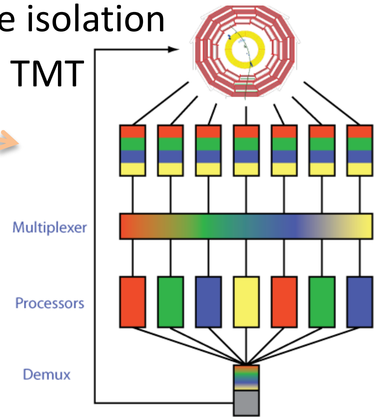
- Improved electromagnetic object isolation using calorimeter energy distributions with pile-up subtraction;
- Improved jet finding with pile-up subtraction;
- Improved hadronic tau identification with a much narrower cone;
- Improved muon p_T resolution in difficult regions;
- Isolation of muons using calorimeter energy distributions with pile-up subtraction;
- Improved global Level-1 trigger menu with a greater number of triggers and with more sophisticated relations involving the input objects.

○ Calo Trigger

- LS1: optical split (oSLB & oRM) and operate slice of upgrade in parallel
  - 2015: use prototype boards to implement improved τ, e isolation
    - 2016: grow slice to full upgrade TMT architecture baseline





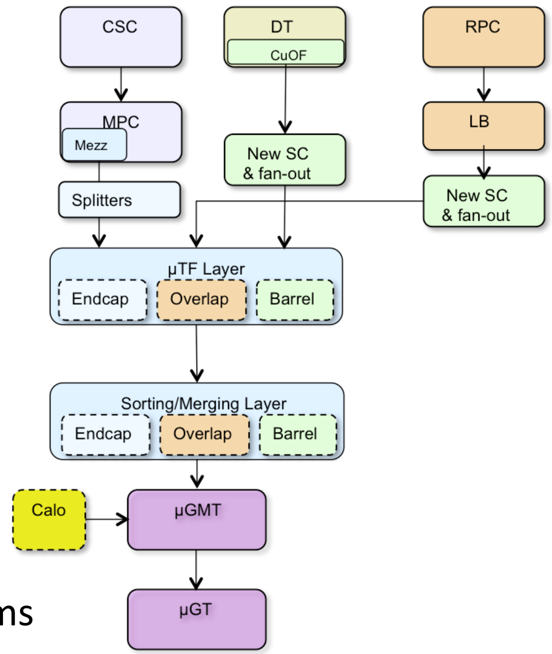


Time-Multiplexed Trigger

L1 trigger upgrade (II)

Trigger Upgrade: The Plan

- Muon
  - Upgrade/integrate Track Finders: endcap (CSCTF), barrel (DTTF) and Overlap regions
  - options for connection between Muon and Calo triggers
- Global
  - Upgrade the Trigger Control and Distribution System, separate from GT
  - Again use standard μTCA boards with large FPGAs for new algorithms

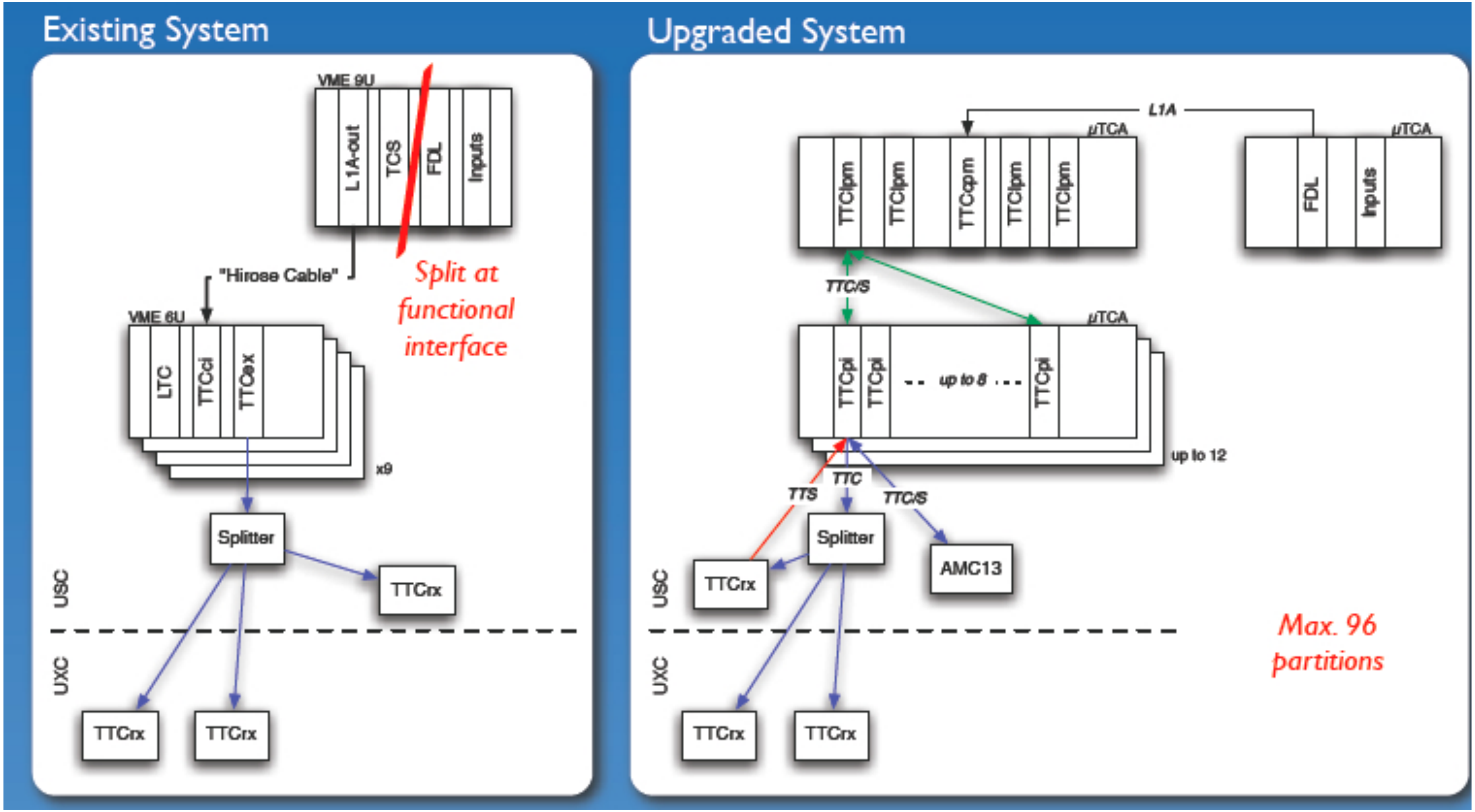


- Combine all 3 muon systems in new TF layer
  - Muon redundancy used earlier in chain
- Switch over to new system when fully produced and commissioned
  - Target: 2016
- Some options on how to connect RPC, and how TF layer factorised
- Add connection to calo trigger upgrade to provide muon isolation
  - Baseline calo regions ➔ GMT

- Cost and Schedule
  - The cost tables and schedule not yet reviewed
    - Cost scale is ~5M CHF
    - Goal to complete hardware and initial trigger firmware/software for 2016 physics

➢ Hardware is one thing – we need a physics ready trigger system (including FM, SW, trigger tables). This is a major project.

Trigger Control & Distribution System



- Operational mid 2014

DAQ2 for post-LS1

- DAQ2
  - re-implementation with up-to-date technology
  - Typically 10x less nodes with 10x more performance
    - DAQ1: 2 x 2Gbps Myrinet and 3 x 1GbE
  - Consider 10 GbE, 40 GbE, IB FDR (56 Gbps)
  - Timescale
    - Design, evaluation, order, for delivery and installation Q4 2013
    - Switchover DAQ1 to DAQ2 Apr-2014, commissioning, improvements

DAQ2 for post-LS1

FRL / FEROLs

Custom Link 10Gb/s TCP/IP

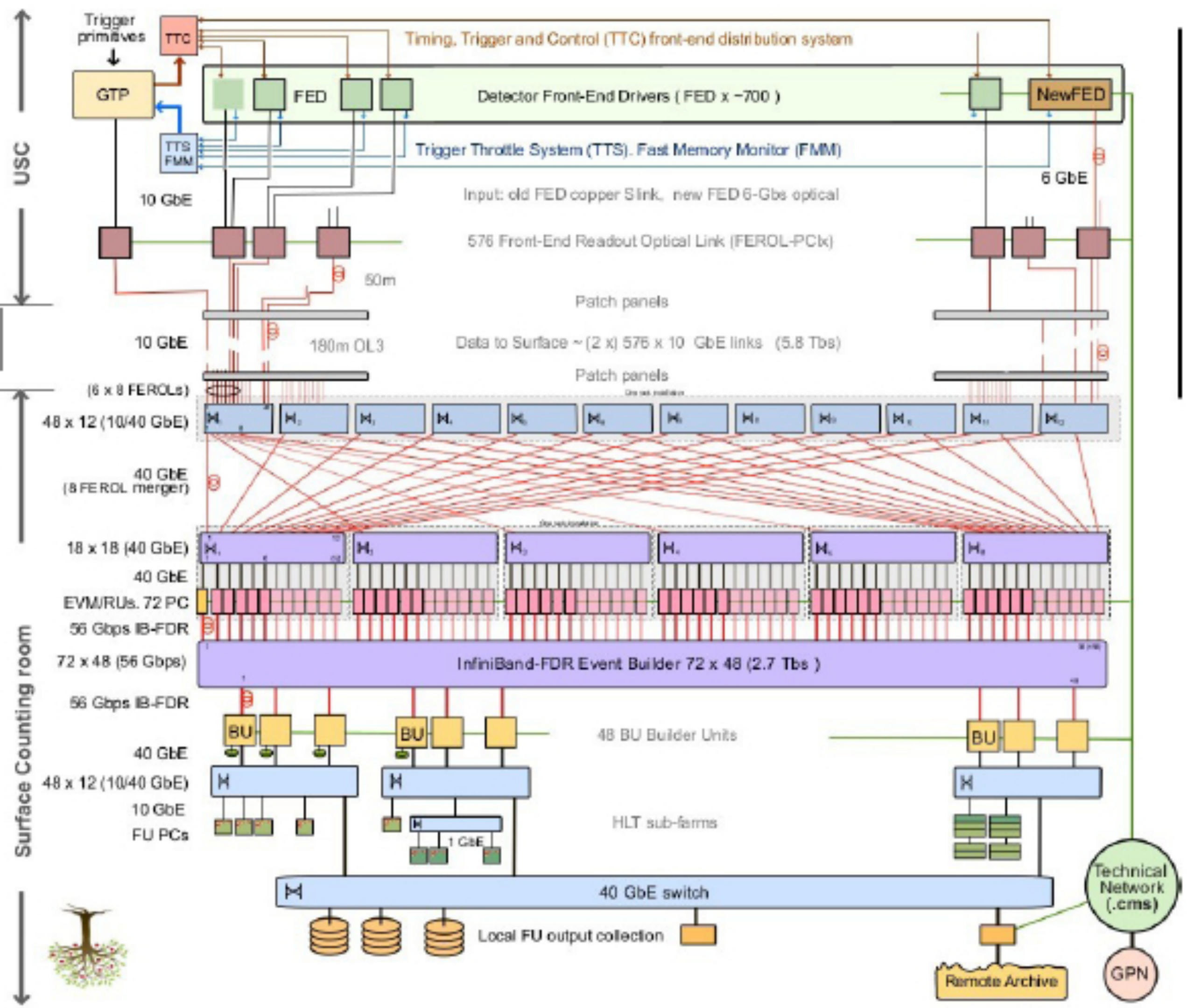
10 GB/s to 40 Gb/s translation

2 layer of switches (implements redundancy)

Ethernet / Infiniband translation

Infiniband Event builder

HLT farm



~Phase-I
Post – LS2
Run 3

DAQ for post-LS2

- Adiabatic changes for CMS
  - Increased data sizes due to higher pile-up
  - Some sub-detectors will be replaced which lead to higher data volumes
    - Eg HCAL sensors
  - More sub-det new back-end electronics in uTCA standard with serial link to cDAQ

- Equipment replacement cycle
  - PC and network replaced typically each 5 years

- Two scenarios
  - “box to box” replacement
  - Re-implement with up-to-date technology (like DAQ2)

Phase-II
Post – LS3
Run 4

CMS Phase-II

- HL-LHC
  - IL of 5 x 10**34, pileup 100-200
- Detector
  - New Tracker
  - Forward Calo?, Muons?
- DAQ and trigger
  - Track trigger
  - All sub-detectors will have new off-detector electronics
  - Entirely new central-DAQ system

15

Trigger Performance and Strategy – Interim Report

- Key goal: maintain the physics acceptances of leptonic, photonic, and hadronic trigger objects similar to 2012 (especially for low-mass processes like Higgs)
- Two key components under consideration for Phase 2:
  1. **L1 tracking trigger**
  2. **a significant increase of L1 rate, L1 latency and HLT output rate**
- Tracking at L1 will *help* maintain rates for muons, electrons & possibly taus. Only limited improvement expected for photons & hadronic objects
- For these, it may be important to increase L1 rate substantially. An increase in rate requires significant changes to frontend electronics, so also consider
  3. **Increasing L1 latency from present 4 μs (Tracker) or 6.4 μs (ECAL) limit**

  Allows more time for more sophisticated algorithms in new FPGAs and architecture

➢ "Target parameters" to focus the discussion
  – **1 MHz rate and 20 μs latency**

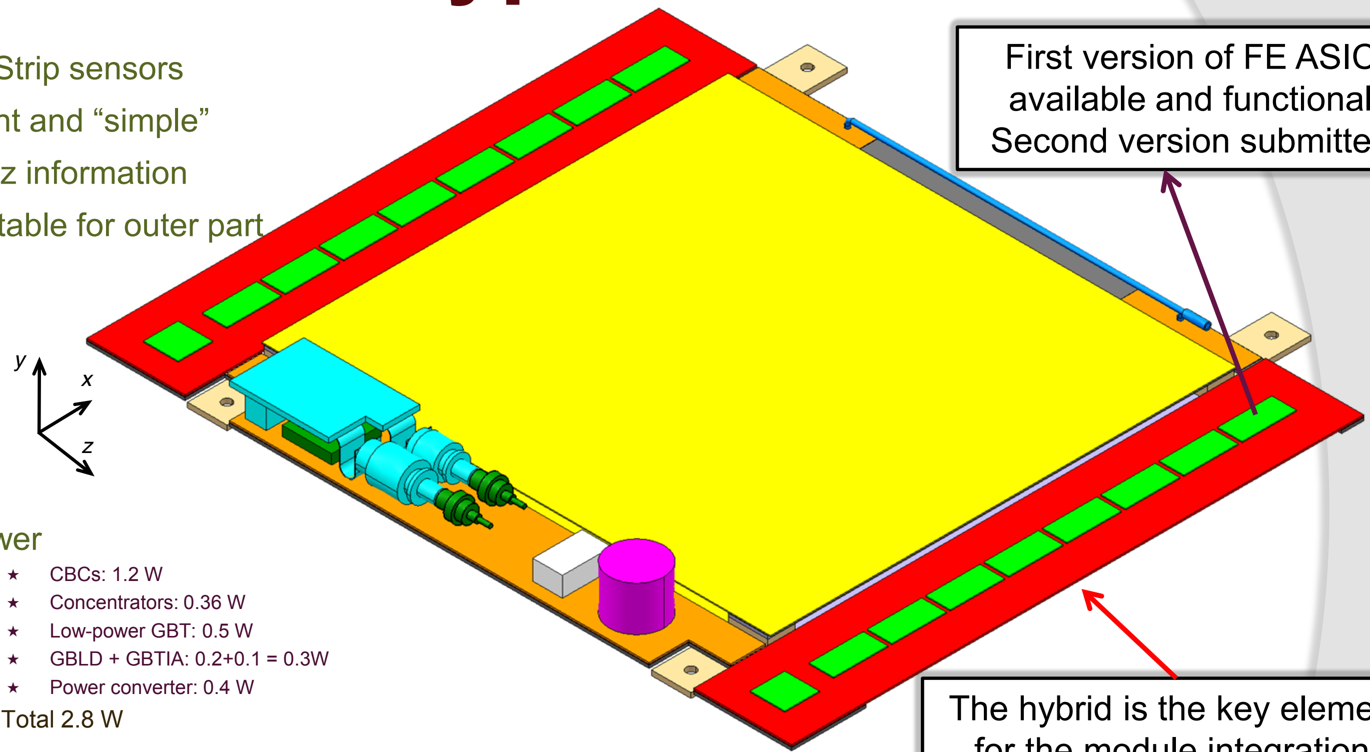
- CDAQ/HLT initial look: trends for networking/switching and multi-core computing circa 2023
  - "1 MHz input looks feasible" → output rate would be up to 10 kHz

2. & 3. Basic Parameter Scenarios

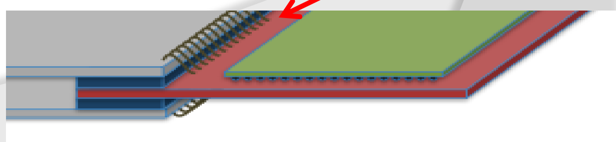
- Surveyed subsystems, DAQ, Computing led to consideration of following basic parameter scenarios (so far):
  - Scenario 1: L1 rate = 100 kHz, L1 Latency = 6.4 μs (present = 4 μs)
    - Used up to now to guide Phase 2 Tracker
  - Scenario 2 ("non-invasive"): L1 rate = 150 kHz, L1 Latency = 6.4 μs
    - Survey among sub-systems, (e.g. ECAL), suggests that L1 trigger rate can go up to 150 kHz without change of front-end electronics (to be further confirmed).
  - Scenario 3: L1 rate = up to 1MHz, L1 Latency: up to 20 μs
    - Survey suggests feasible *IF significant upgrades are carried out*
    - To set the scale: Task Force on EB FEE replacement → ~10M CHF and 26 months of shutdown
- **Clearly any such change requires good physics justification, and estimates of work/cost for each subsystem**
- **Aim for final decision on this by early 2014**
- In the interim, propose that ongoing work for Phase 2 be compatible with all scenarios
    - Implies design changes for upgrade electronics (*e.g.* Tracker)

p_T modules types: "2S Module"

- 2x Strip sensors
- Light and "simple"
- No z information
- Suitable for outer part



- Power
  - CBCs: 1.2 W
  - Concentrators: 0.36 W
  - Low-power GBT: 0.5 W
  - GBLD + GBTIA: 0.2+0.1 = 0.3W
  - Power converter: 0.4 W
  - Total 2.8 W
- ≈ 5 cm long strips, ≈ 90 μm pitch, ≈ 10x10 cm^2 overall sensor size
- Wirebonds from the sensors to the hybrid **on the two sides**
  - 2048 channels on each hybrid
- Chips bump-bonded onto the hybrid



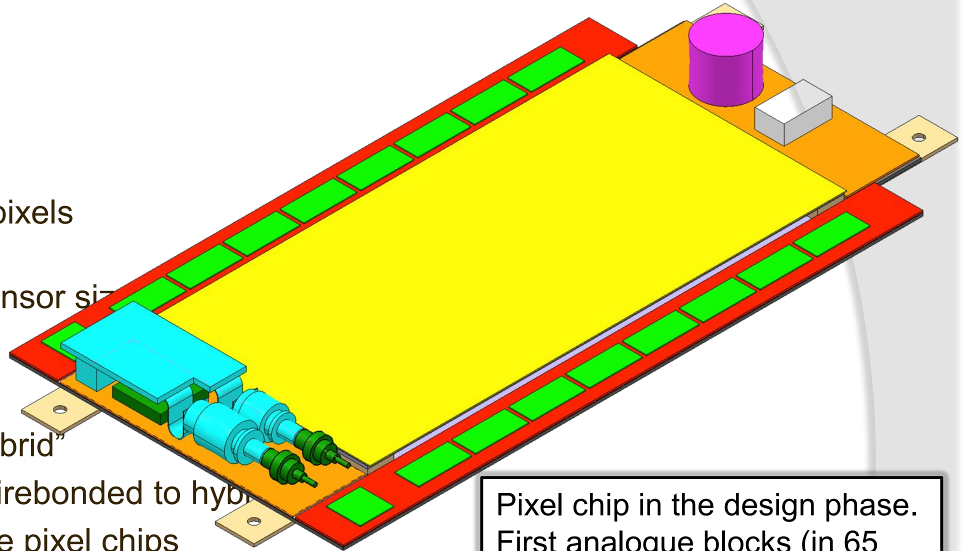
p_T modules types: “PS Module”

➢ Sensors:
  - Top sensor: strips
    - ★ 2×25 mm, 100 µm pitch
  - Bottom sensor: long pixels
    - ★ 100 µm × 1500 µm
  - ≈ 5x10 cm^2 overall sensor si

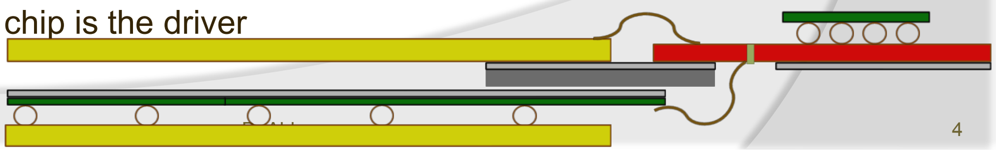
➢ Readout:
  - Top: wirebonds to “hybrid”
  - Bottom: pixel chips wirebonded to hyb
  - Correlation logic in the pixel chips

➢ No interposer, sensors spacing tunable

➢ Power estimates
    - ★ Pixels + Strips + Logic ~ 2.62 + 0.51 + 0.38 W = 3.51 W
    - ★ Low-power GBT + GBLD + GBTIA ~ 0.5 + 0.2 + 0.1 = 0.8 W
    - ★ Power converter ~0.75 W
  - Total ~ 5.1 W, pixel chip is the driver



Pixel chip in the design phase. First analogue blocks (in 65 nm) to be submitted in 2013.



12/14/2012 4

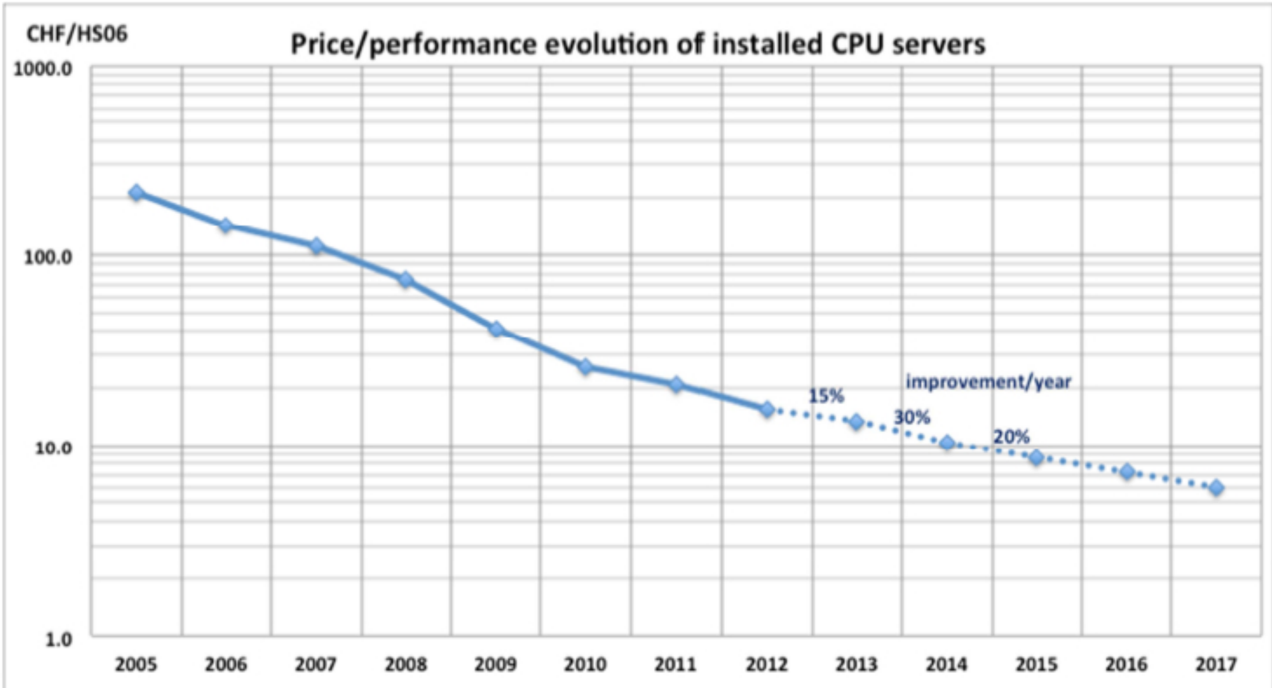
Latency and trigger rate

- Latency
  - Long pixel chip design already compliant
    - 1024 cell pipeline, 25.6 μs
  - CBC requires one design iteration
- L1 accept rate
  - Requires data reduction in the readout path for the CBC
  - Could be implemented in the CBC or in the concentrator
    - Advantages and disadvantages under discussion
  - Not a big margin left for 1 MHz frequency
    - Probably OK?
- Bottom line:
  - New specs can be implemented, with significant effort
  - We need to decide now

12/14/2012 D. Abbaneo 5

HLT with “normal” PCs

Server PCs in CERN/IT data center



Dual CPU servers, cost normalised to 2GB memory/core
Forecast 20% improvement per year
Gives 0.8^10=0.10 in 10 years, so gain factor ~10

FM 29-06-2012 Trigger Strategy - DAQ 17

DAQ

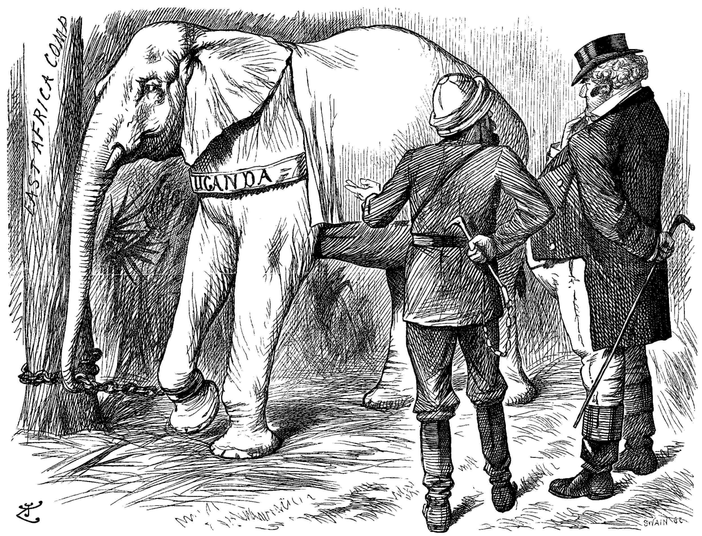
- Post-LS3 assumptions
  - Replaced BE electronics
  - 2 level trigger-daq (as now), full events at HLT (as now)
  - Assume 10 MB events size (1 MB now)
- For 1 MHz L1A, 10 MB event size
  - Assume 100 Gbps DAQ link (between BE and cDAQ)
  - Canonical system
    - 1000 FEDs with 100 Gbps DAQ link
    - Switch throughput 100 Tbps
- 40 MHz L1 appears impossible
  - Due to on-detector electronics

OTHER

HLT farm

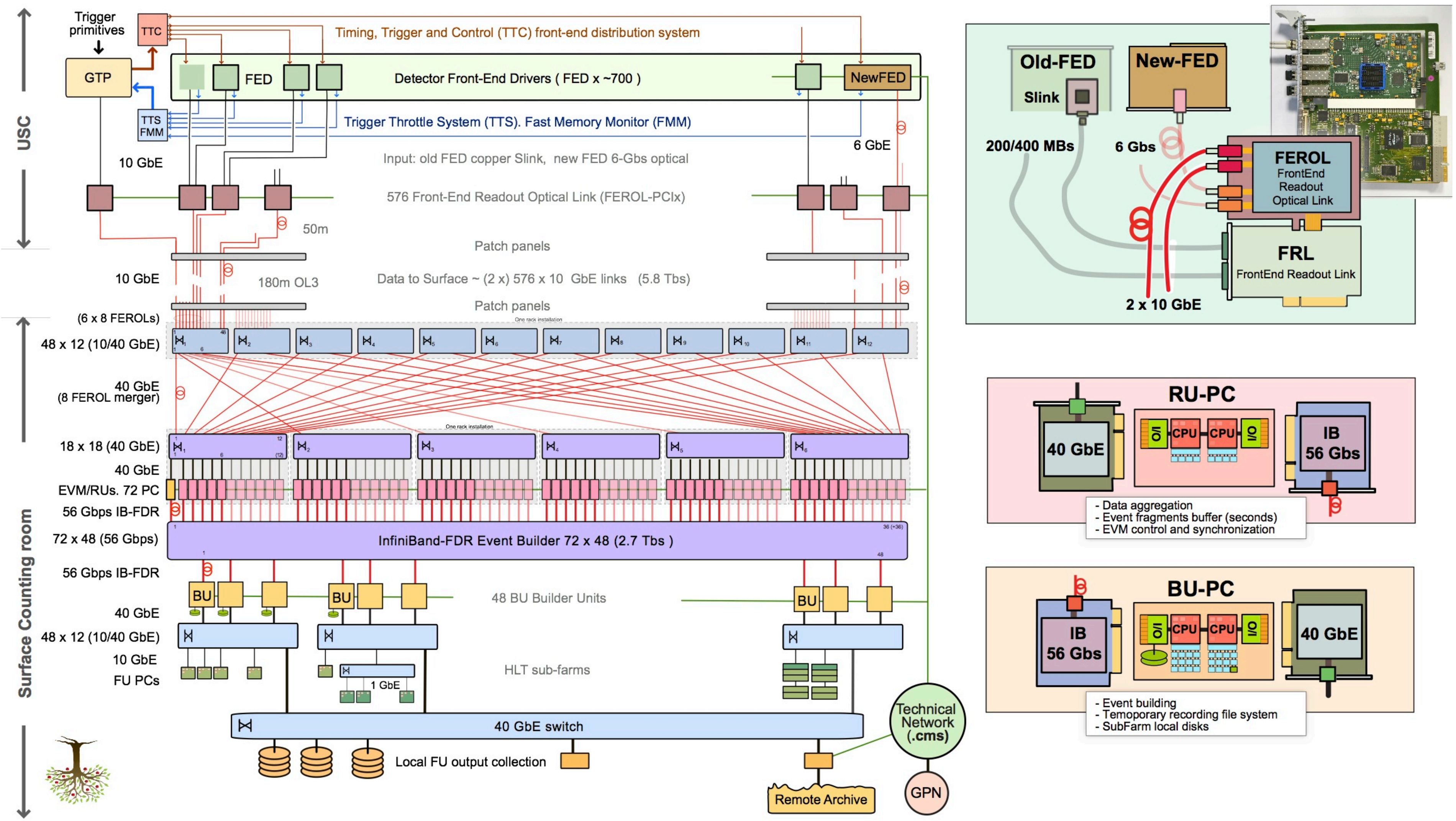
- So far, using dual-CPU (x86) server PCs
- Strategy to deploy “offline” framework and processors (CMSSW)
- Work of fully-built events
- Actually, after LS1
  - intend to run in offline mode (file to file)
  - rely on efficient multi-thread version

- No specific HLT work done on other platforms
  - GPU
  - Large number of cores a la, Xeon-Phi

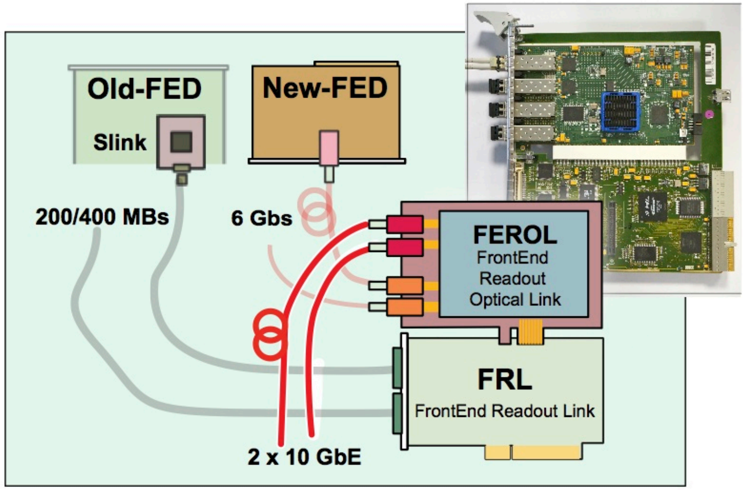


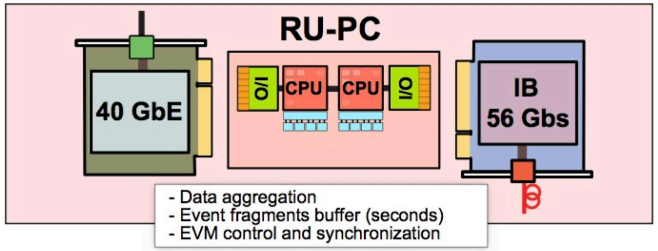
BACKUP MATERIAL

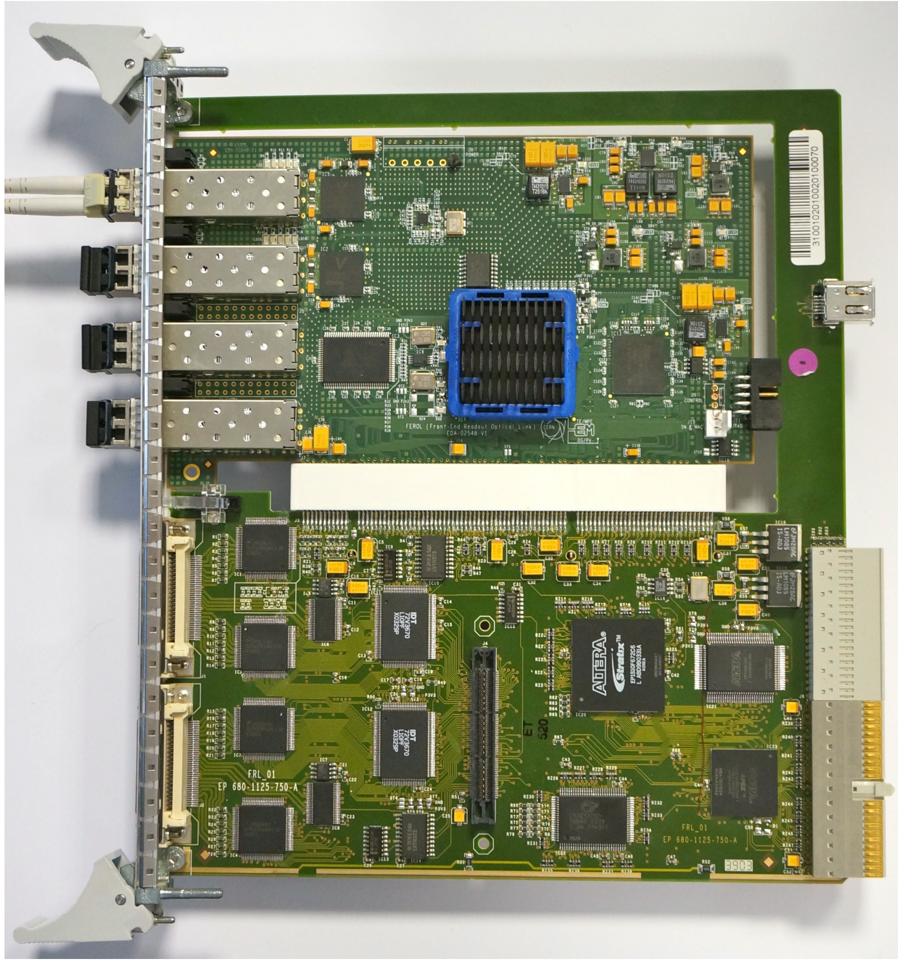
CMS DAQ upgrade for post LS1 (DAQ2)



FEROL

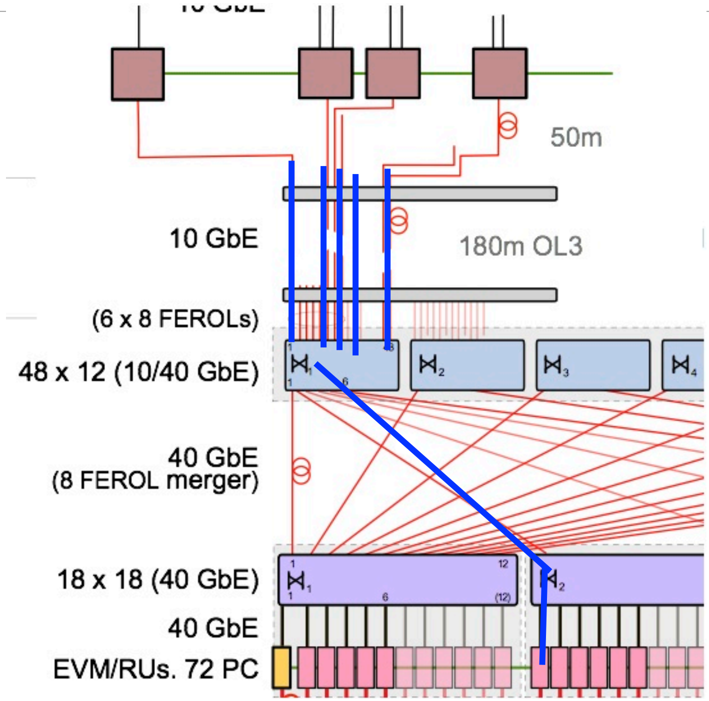






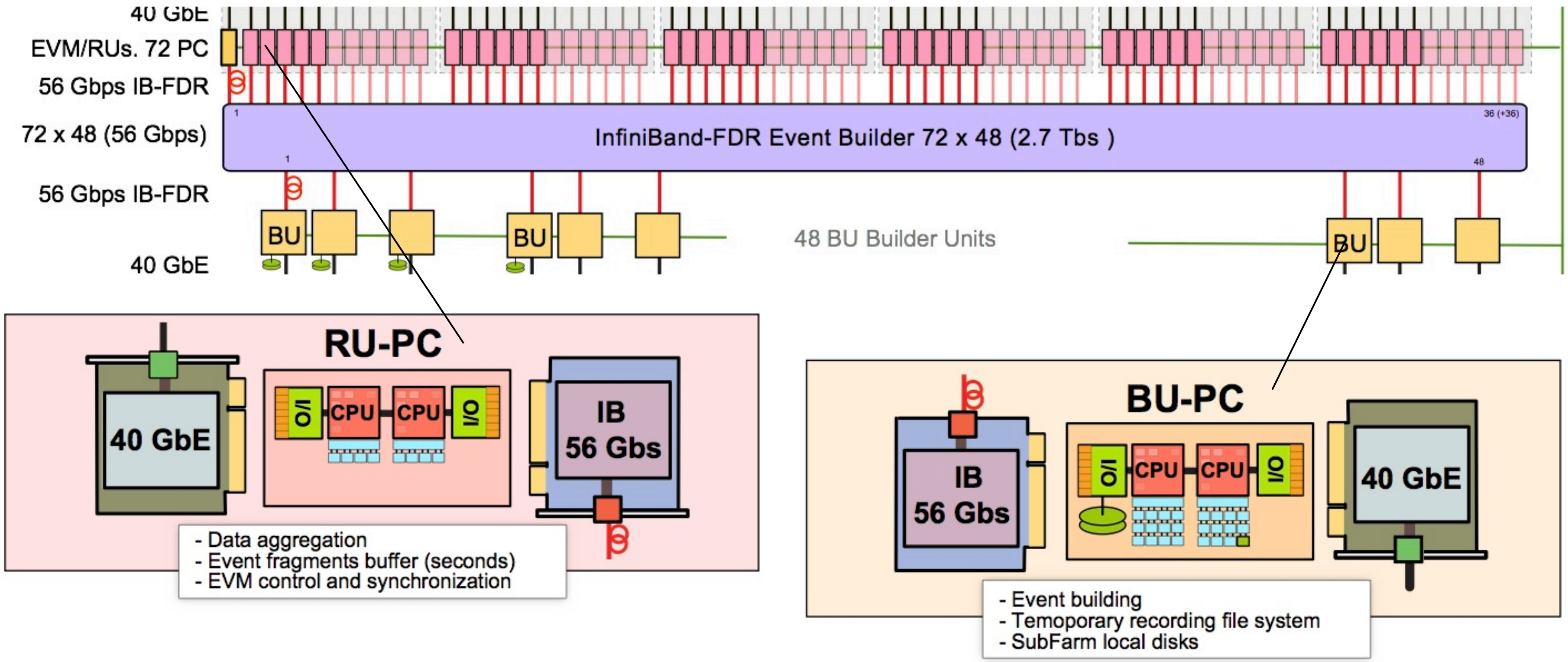
- FEROL
  - Input: custom protocol
  - Output: 10 GbE serial, reduced TCP/IP sender in FPGA
- Receiver with NIC in PC with standard driver and TCP/IP stack

FEROL aggregation



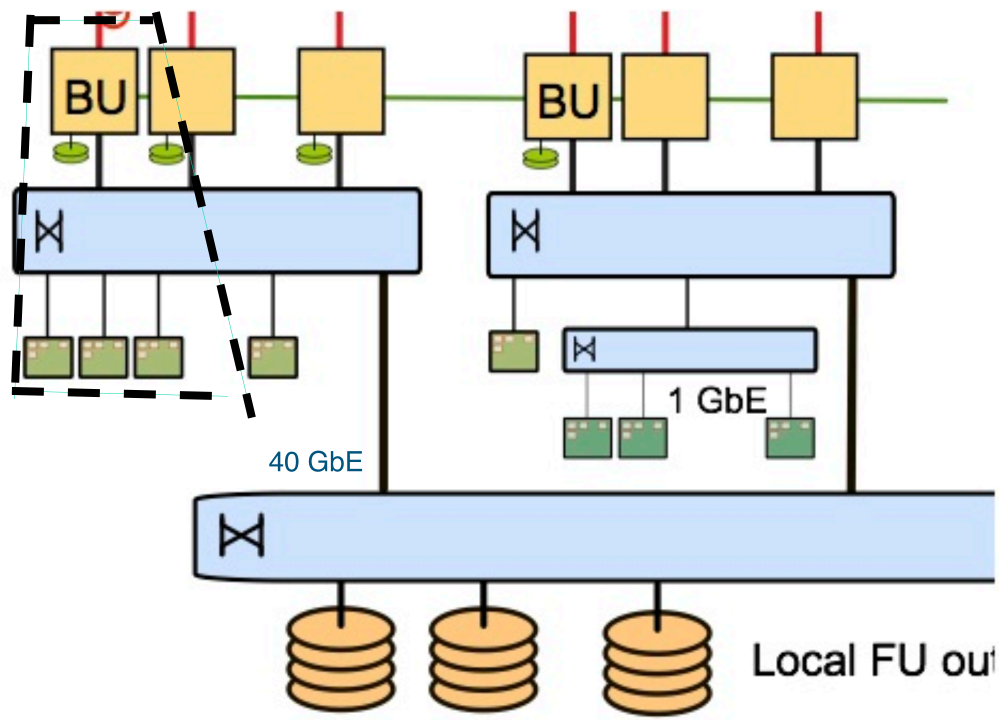
- Aggregation n-to-1, example
  - 16 FEROLs each sending 2 Gbps over 10 GbE link
  - Concentrated in one 40 GbE NIC into PC
  - Reliability and Congestion handled by TCP/IP
- USC – SCX 180m,
  - with OM3 fibres up to 200 m
  - 40 GbE (with 4 lanes 10 Gbps) max. is 150 m – NOT feasible
- Network useful to re-configure when fault with optic, PC, etc

IB EVB



- Performance Scaling with multi-layer switch network?
  - 3 layer Clos
- Implement with “Director” switch or 36-port units?

HLT subfarm



- HLT divided in 48 sub-farms, each with 1 BU and typically 24 HLT nodes
- BU writes to filesystem on ramdisk (~256 GB) with 2-4 GB/s
- HLT nodes (~24) in sub-farm cross mount filesystem
- HLT sub-farm output (1 in 100 events) collected on BU onto normal disk
- HLT output collected from all sub-farms to NAS