# The Evolution of the Trigger and Data Acquisition System in the ATLAS Experiment

Nicoletta Garelli (SLAC)

*on behalf of the* ATLAS Collaboration

CHEP2013

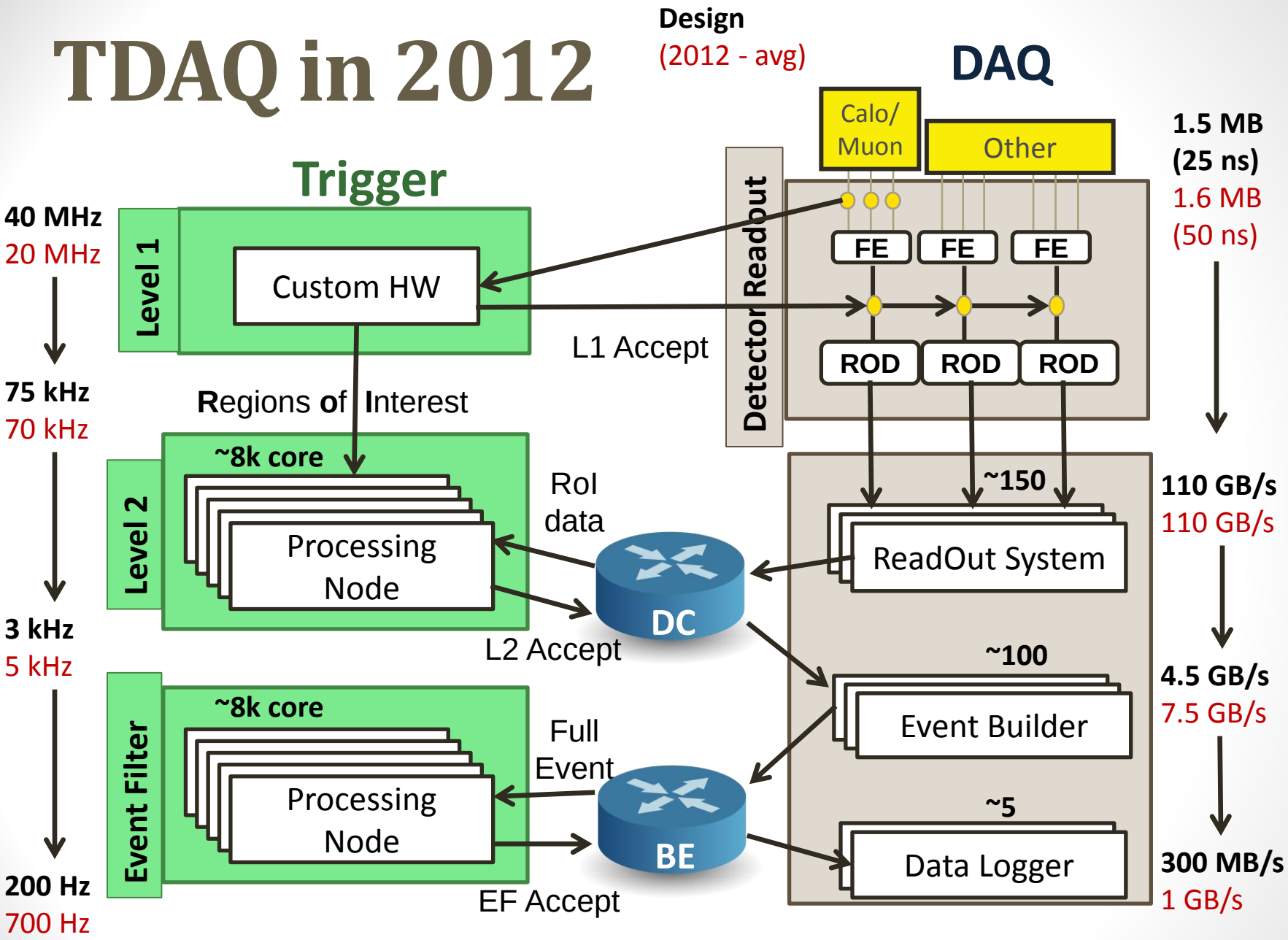
14-18 October, Amsterdam

SLAC NATIONAL ACCELERATOR LABORATORY

# Outline

- **Introduction**
  - ATLAS Trigger and Data Acquisition (TDAQ) system in 2012
  - LHC after the Long Shutdown I (LSI) → new challenges for TDAQ
- **TDAQ evolution**
  - Central Trigger Processor upgrade & new topological trigger
  - Network evolution
  - Read-Out System replacement
  - Data-flow evolution
  - High Level Trigger core software new design
- **First results with new TDAQ architecture**
- **Conclusion**

# TDAQ in 2012




*Note*: Level 2 (L2) + Event Filter (EF) = High Level Trigger (HLT)

# LHC After LS1 – TDAQ View

- LHC will improve performance after LS1 (Feb. 2013 – Nov. 2014)

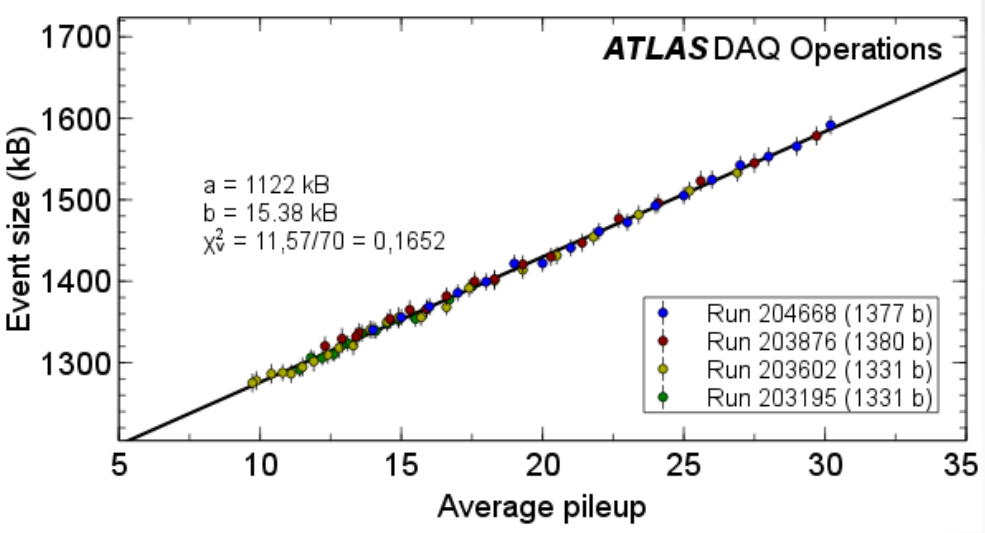
| | $E_{CM}$ [TeV] | K | Instant. Lumi [$cm^{-2}$ $s^{-1}$] | Pile-Up |
|---|---|---|---|---|
| Nominal (25 ns) | 14 | 2808 | $10^{34}$ | 23 |
| 2012 (50 ns) | 8 | 1380 | $0.76\ 10^{34}$ | 35 |
| 25 ns | 13 | 2760 | $0.85\ 10^{34}$ | 23 |
| 25 ns low emittance | 13 | 2600 | $1.5\ 10^{34}$ | **42** |
| 50 ns low emittance | 13 | 1260 | $2.0\ 10^{34}$ | **110** |

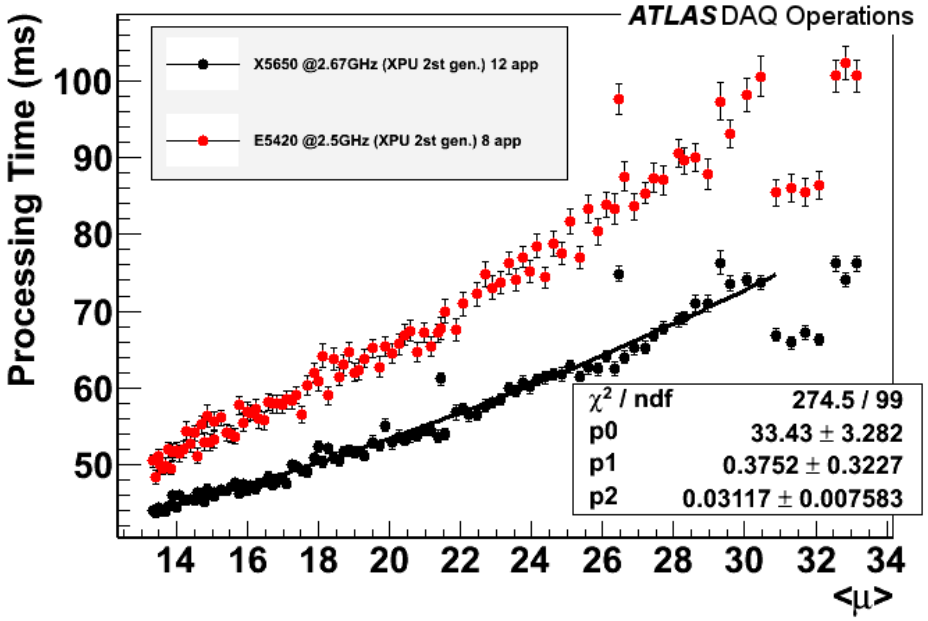
J. Wenninger, 2 July 2013

- ATLAS operating conditions in Run II:
  - More readout channels
  - Level 1 rate: 100 kHz (was 70 kHz in 2012)
  - Average storage rate: 1 kHz (was 700 Hz in 2012)
- Pile-up effects:
  - Event size
  - HLT processing time

# Pile-Up Effects in 2012

- Increase in **Event Size**
  - Nominal: 1.5 MB
  - Peak up to ~1.7 MB in 2012
  - First extrapolation: **event size ~2.3 MB at μ~80**
    - possible future deviations
- Increase in **HLT Processing Time**
  - Depends on **trigger menu**
  - Difficult to trust predictions at higher pile-up



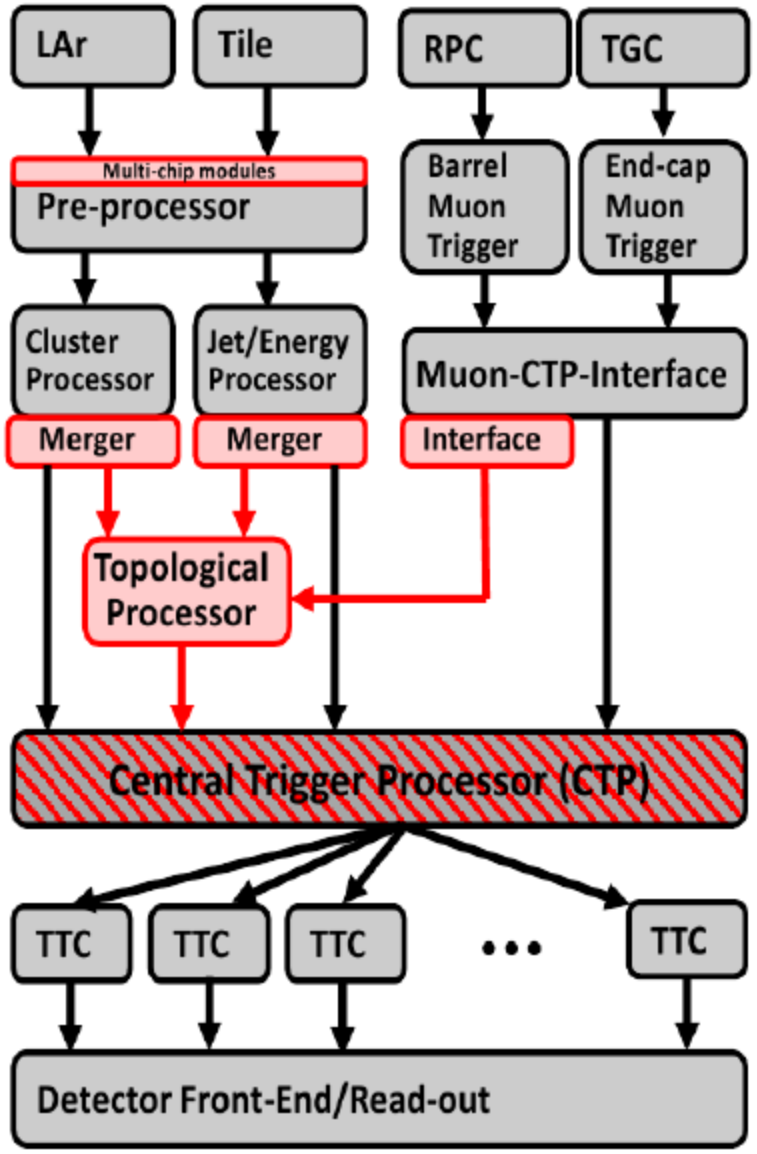



# TDAQ Requirements After LS1

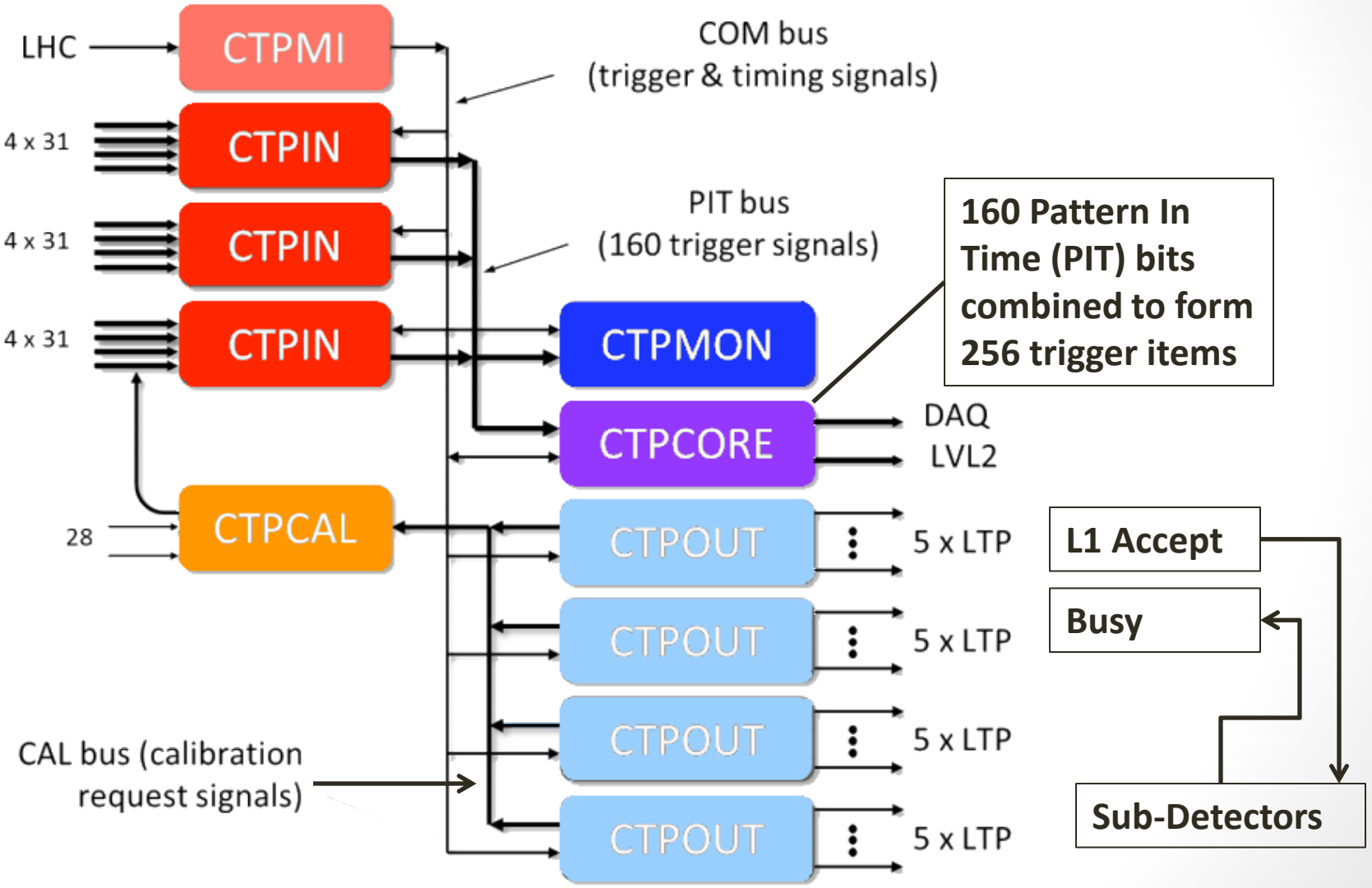
- ATLAS operating conditions after LS1 imposes
  - Higher ReadOut System (ROS) request rate
  - Increased bandwidth at HLT
- Major activities to be compliant with requirements for Run II
  1. **Upgrade Central Trigger Processor**
  2. **Replacement of core routers**
     - End of their lifetime
     - ROS connectivity not enough
  3. **Replacement of ROS**
  4. **Evolve the software** to a system easier to maintain
     - Give more flexibility to HLT
- Two new trigger systems will be introduced
  1. **Topological Processor** at Level-1
  2. **Fast Tracker** to provide full track reconstruction in input to HLT, partially available in Run II
- TDAQ has to be functional for detector recommissioning in early 2014!

# Level 1 Topological Trigger

- *Goal*: merge detailed information from trigger detectors in a single Level 1 module to determine complex observables (*e.g.* invariant mass)
- *Requirements*:
  - Receive 1 Tbps bandwidth
  - Process data within 100 ns
- **Completely new electronics: Topological Processor**
  - Input elaborated by algorithms implemented in FPGAs
  - Output sent to CTP
  - Prototypes ready based on ATCA

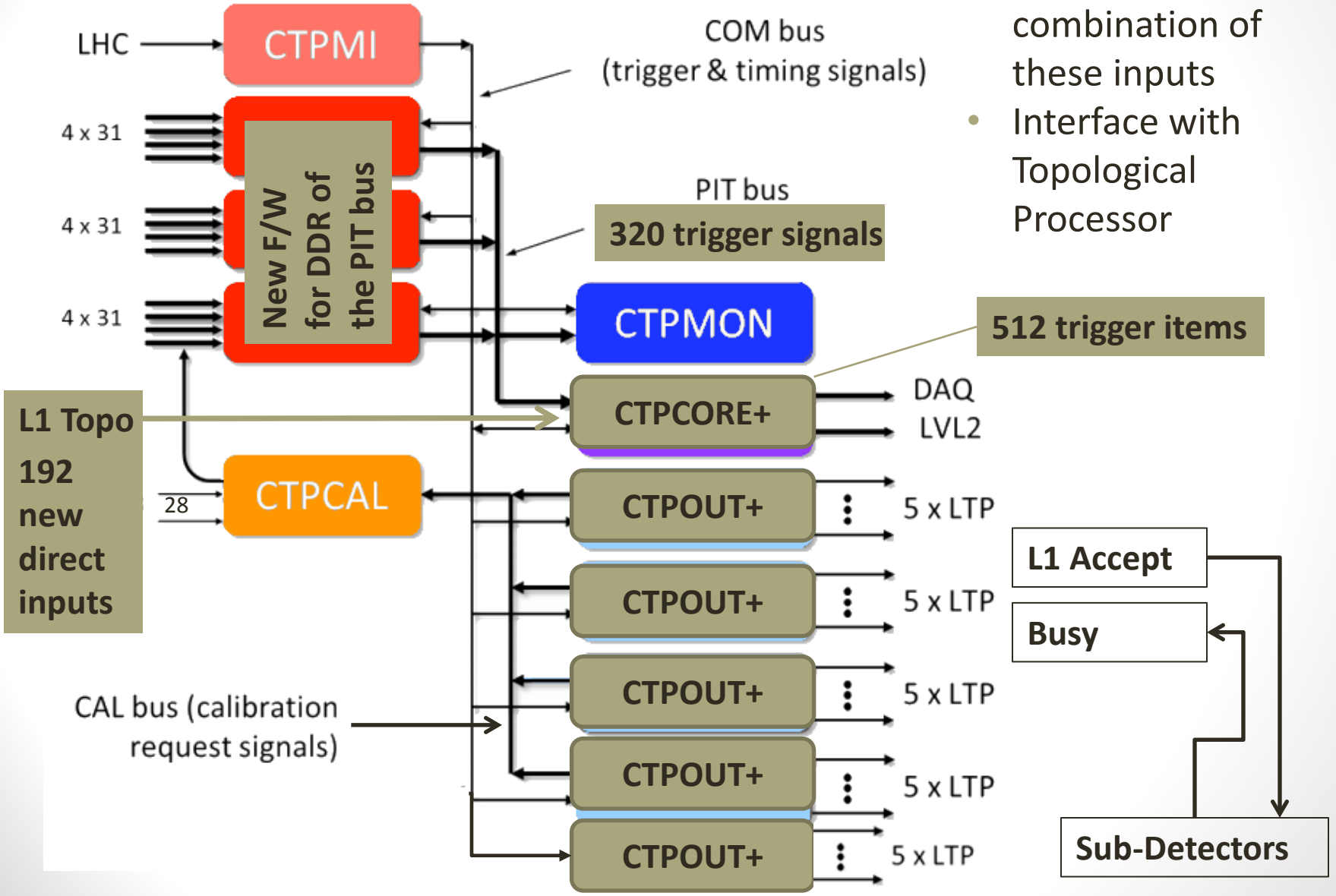


# Central Trigger Processor (CTP) in 2012

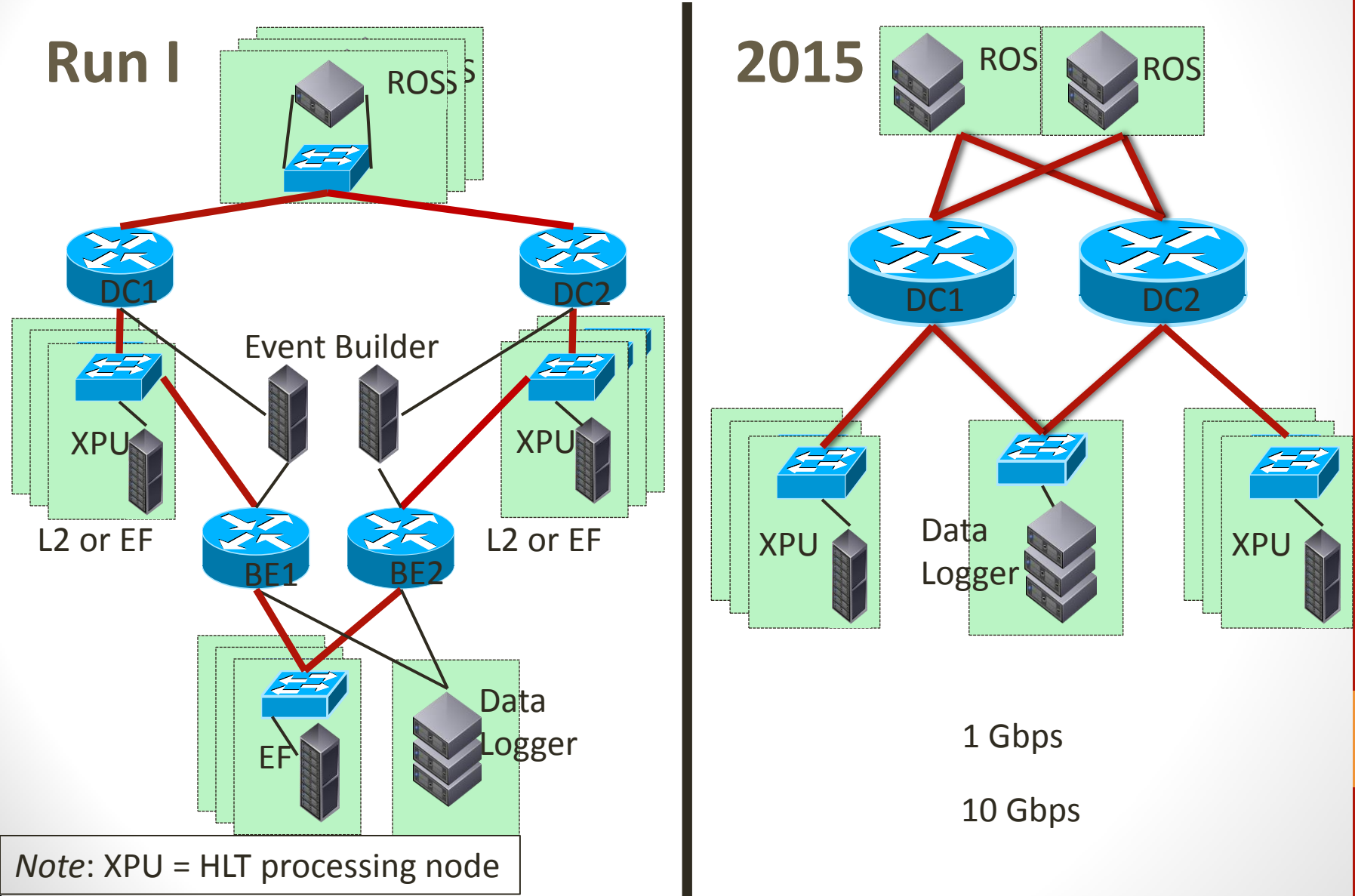


# CTP Upgrade

- Process double the number of trigger inputs and logical combination of these inputs
- Interface with Topological Processor



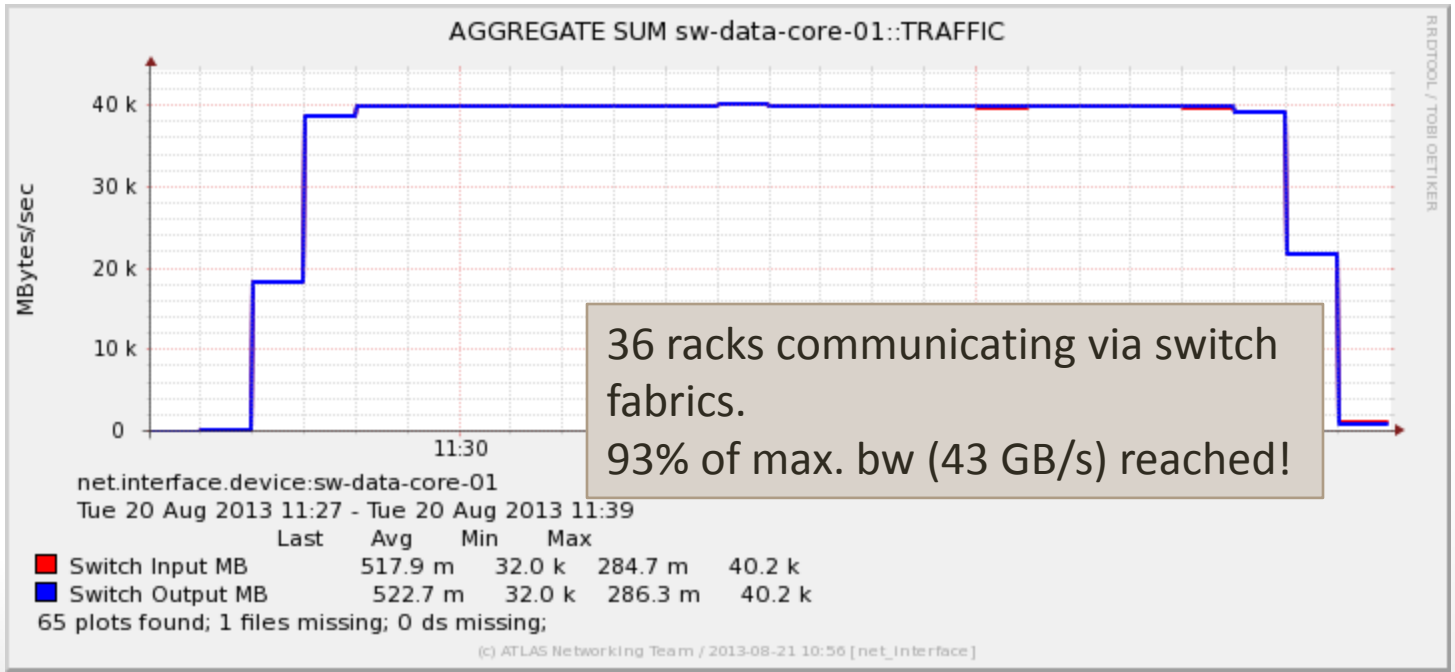
# TDAQ Network Evolution




*Note*: XPU = HLT processing node

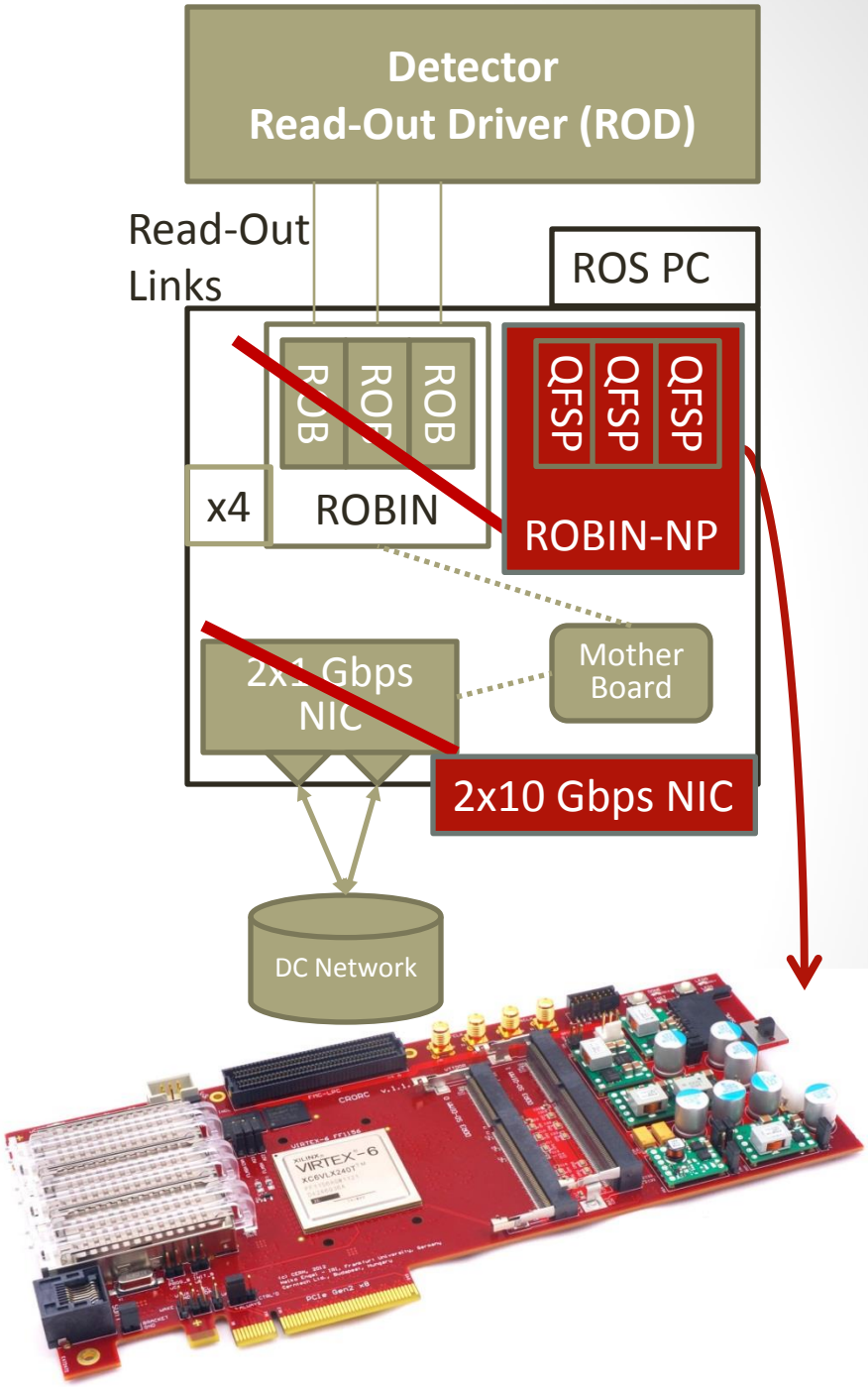
# Network Evolution First Results

- All connectivity in place, but switches and links to ReadOut System
- **New network architecture validated** during new data flow software tests (summer 2013)
- Core chassis equipped with new **Brocade 24-port 10 Gbps blade**
  - performance tests showed that they meet specifications
    - 18 fully non-blocking full duplex ports

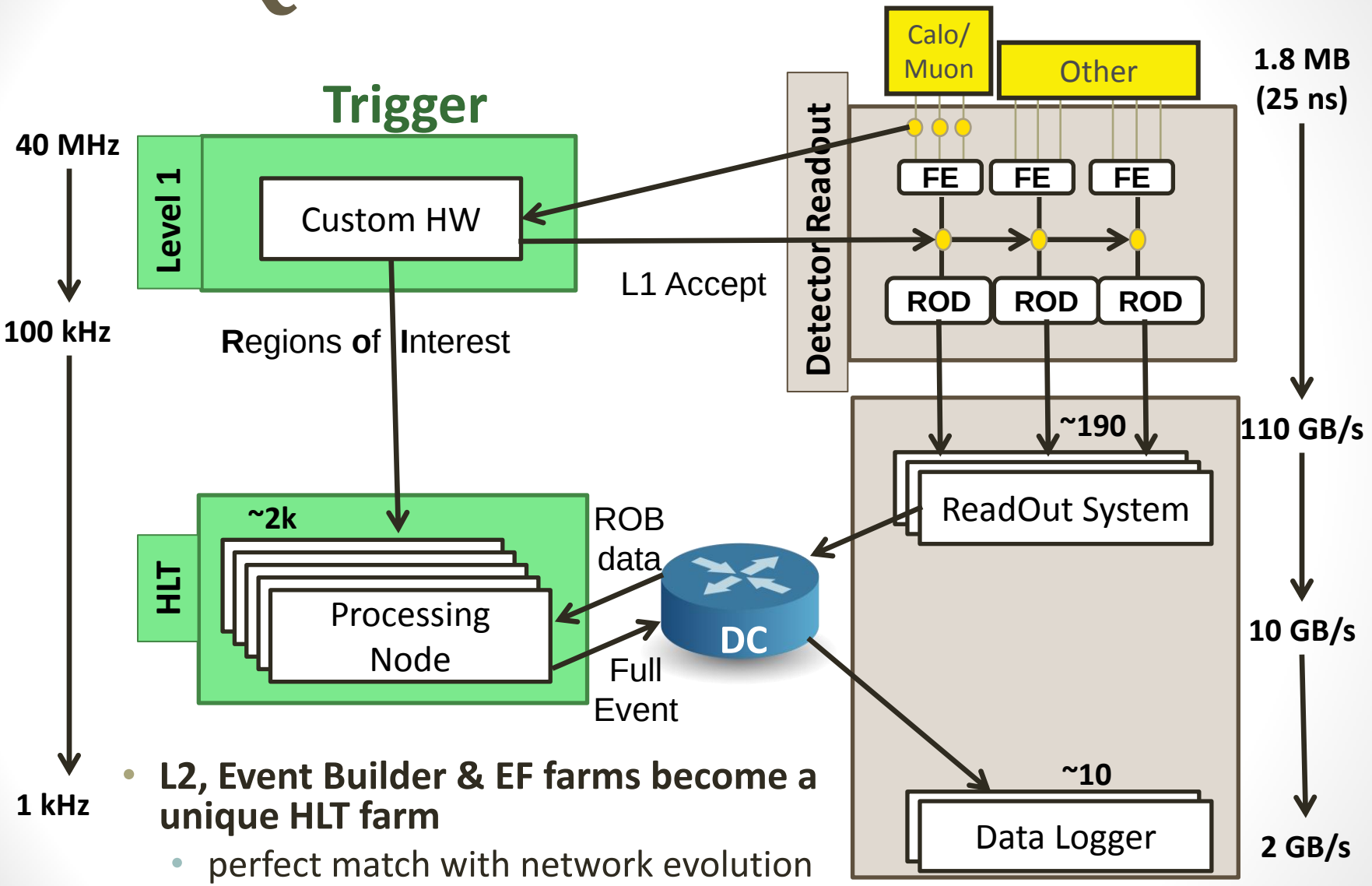


# ROS Evolution

- ReadOut System (ROS) in Run II
  - Sustain 100 kHz Level 1 rate
  - More readout links, but limited rack space: from 150 to **~190 PCs**, but with one PCI board (was 4)
  - More bandwidth: from 2x 1 Gbps to **2x 10 Gbps NIC**
- New PCI board: **ROBIN No Processor (NP)**
  - ALICE C-RORC with custom F/W
  - PCI-X replaced with **PCI-Express**
  - CPU of host PC will do job of on-board processor of current board
  - **3 QSFP transceivers** to readout up to 12 links
- New ROS: challenging project



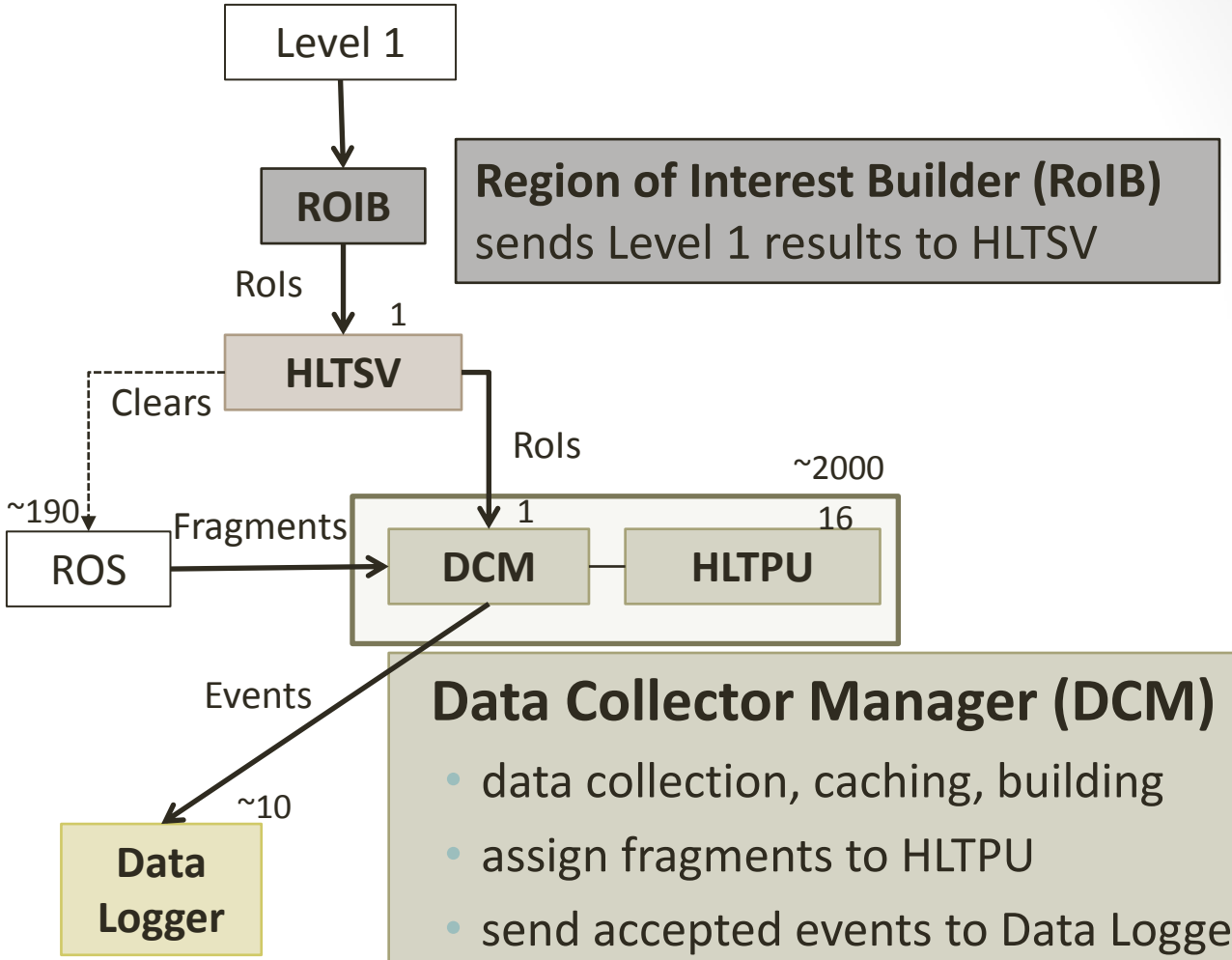
# TDAQ After LS1




- **L2, Event Builder & EF farms become a unique HLT farm**
  - perfect match with network evolution
  - automatically balanced distribution of computing resources

# Data Flow New Architecture

**HLT SuperVisor (HLTSV)** dispatches Regions of Interest to available HLT nodes [Run I: Level 2 supervisors farm]




**Region of Interest Builder (RoIB)** sends Level 1 results to HLTSV

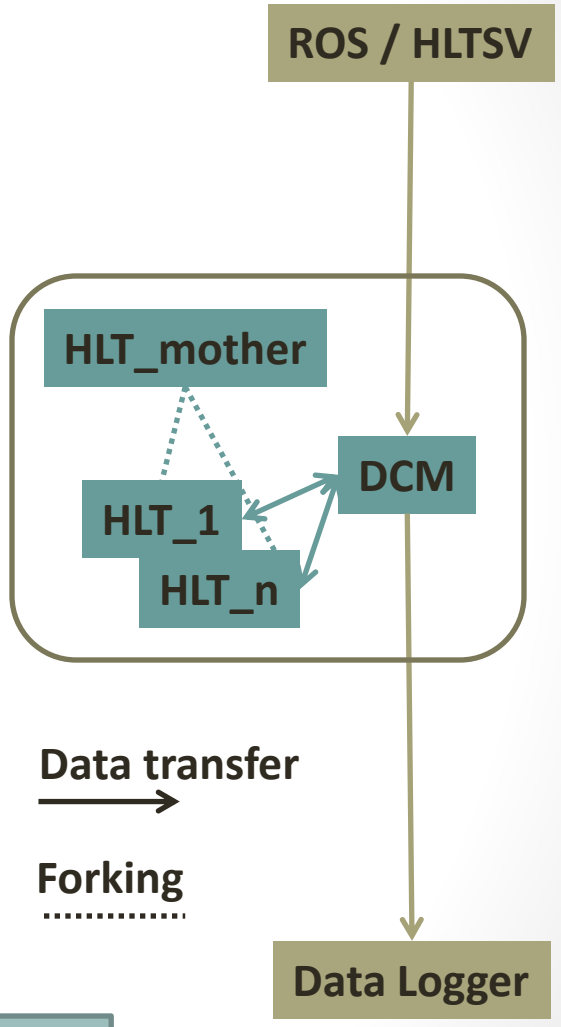
- **Common data communication library**
  - **asynchronous I/O design**, based on boost::asio
  - improve maintainability

## Data Collector Manager (DCM)

- data collection, caching, building
- assign fragments to HLTPU
- send accepted events to Data Logger
- ***n* HLT Processing Units (HLTPU)**
  - execute HLT algorithms

# HLT New Design

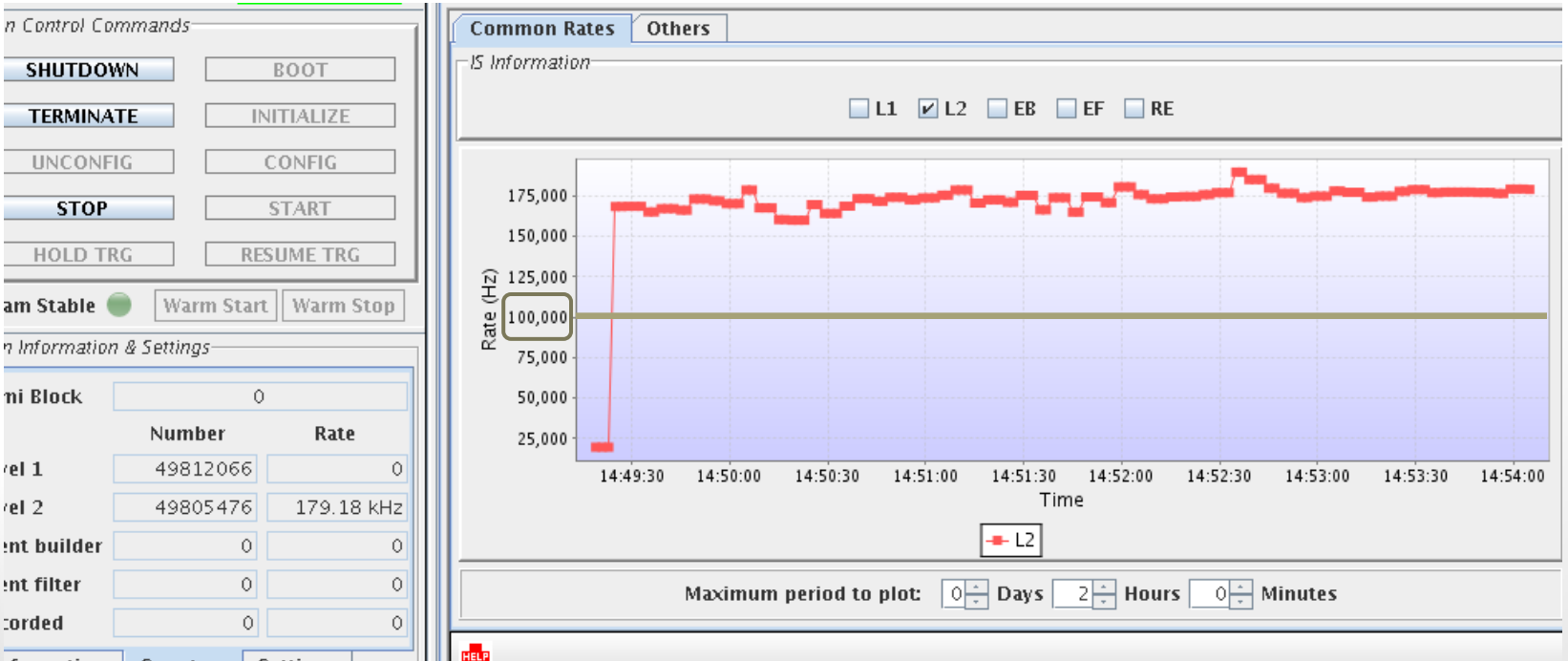
- **Former Level 2 and Event Filter algorithms run in one process**
  - Limited network transfer of data from Level 2 to Event Filter algorithms replaced by unlimited in-memory transfer
  - Data Collection optimization
  - Flexible event building
- **Mother Process** forks to exploit kernel's Copy on Write feature (CoW)
  - HLT workers (multi-process HLT) created at start of data taking session
  - Substantial amount of memory saving



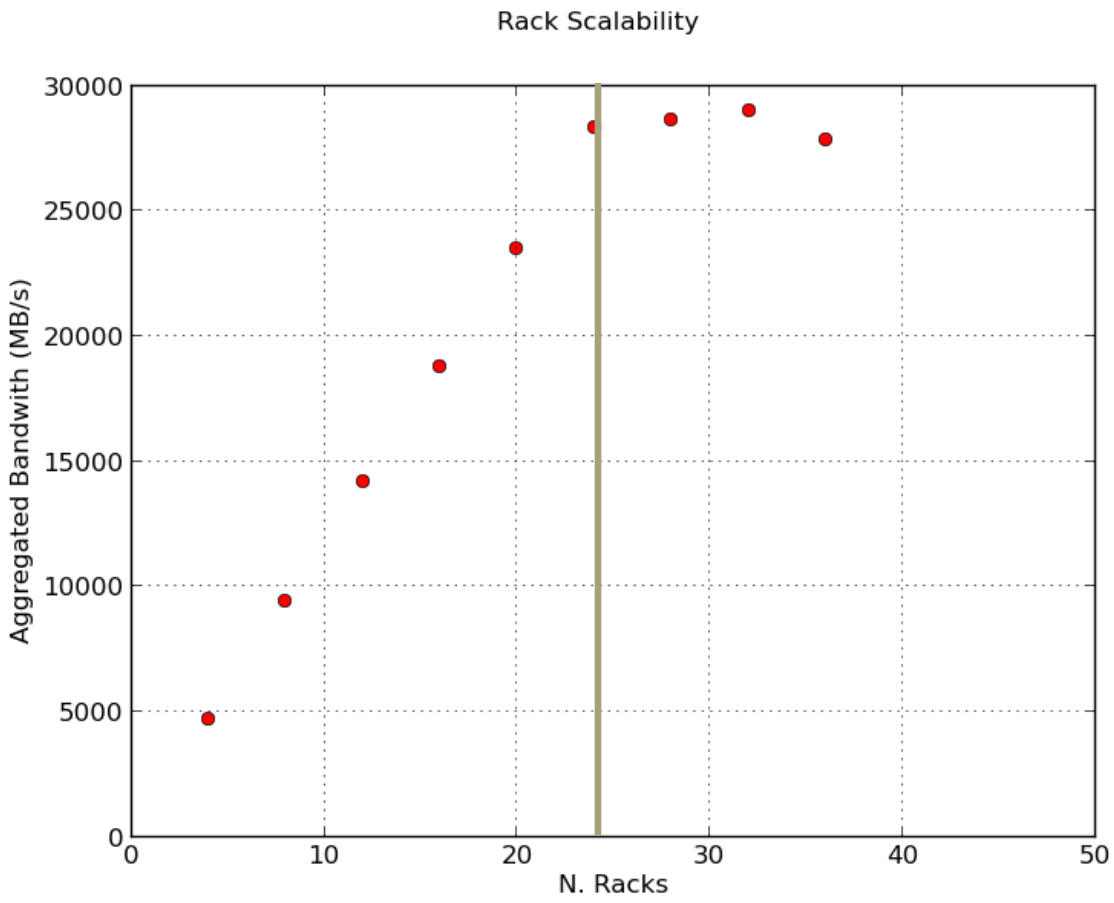

More in K. Nagano's talk (Mon 14, afternoon)
More in T. Bold's talk (Tue 15, afternoon)

# HLT SuperVisor (HLTSV) Rate

- First test: **HLTSV application can sustain 100 kHz Level 1 rate**
- Test conditions
  - 2x 1 Gbps (final design will be 2x 10 Gbps)
  - No RoI Builder, reading S-links will consume CPU
  - L1 rate ~175 kHz (80% of total bandwidth)
- **Scalability of HLTSV proved**



# HLT Racks Scalability




- *Goal*: measure network scalability increasing number of HLT racks (aka HLT bandwidth)
- **Max aggregated bandwith: 252 Gbps** (~31 GB/s)
  - Now limited by ROS installed bandwidth
  - Once ROS 2x 10 Gbps installed → limited by HLT input
- Saturation expected at ~24 HLT racks
  - 10 Gbps per rack
- Flat bandwidth after ROS saturation

# Conclusion

- **TDAQ system is evolving to satisfy Run II requirements**
- Level 1 topological trigger and CTP upgrade under development
- Unified data network has been installed and validated
  - ReadOut System performance limitation will be removed
- New Data-Flow simplified architecture has been succesfully tested
  - Ready to sustain 100 kHz Level 1 rate
  - Homogeneous and flexible HLT system
  - Automatically balanced distribution of HLT computing resources
- New exploited technologies satisfactory
- **LS1 schedule very dense, but TDAQ activities well under way**
  - No show stoppers so far
  - First performance tests promising
  - **Ready to recommission ATLAS in early 2014** (with old ROS PCs)