Real-Time System for Distribution of Clock and Control Commands with Fixed Latency

Maurício Féo – CERN

on behalf of the LHCb Online team

IEEE NPSS NUCLEAR & PLASMA SCIENCES SOCIETY

IEEE

23rd Real Time Conference

The LHCb Upgrade

When: LHC Long Shutdown 2 (by 2022)
For Runs 3 & 4 (2022 – 2029)

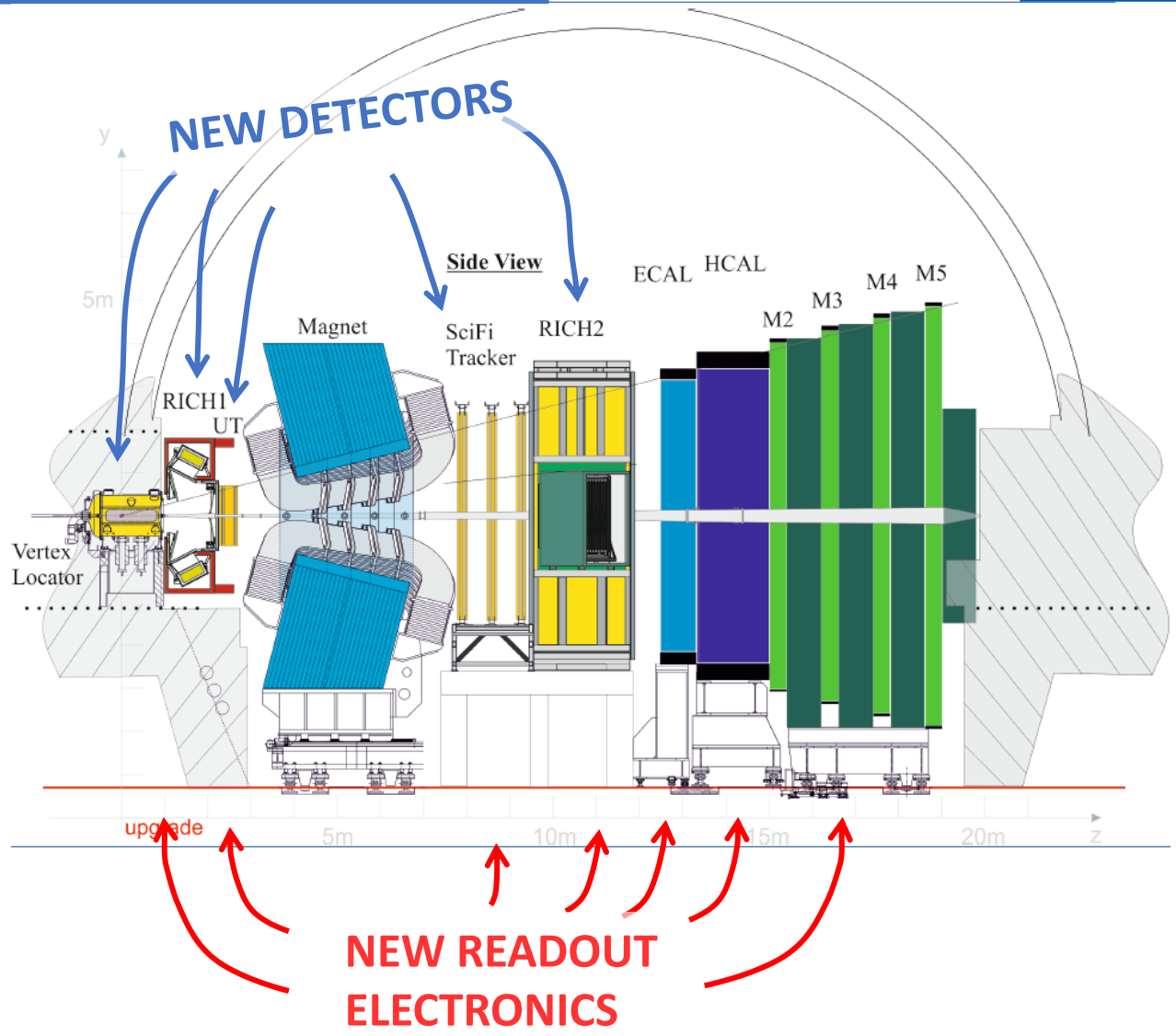
Why: To increase statistics
9 fb^{-1} (Runs 1-2) → 50 fb^{-1} (Runs 1-4)

How: Increasing instant. luminosity
5x higher → L_{inst} = $2x10^{33}$ $cm^{-2}s^{-1}$

Increasing readout rate
1 MHz → 40 MHz

This requires some **main changes:**

- Replace many sub-detectors:
  - New Tracking System
    (VeLo, UT, SciFi)
  - Partially new Particle ID System
    (RICH1 + RICH2)
- Replace of ALL the electronics:
  - No more hardware trigger
  - Event selection in software
  - **Completely new DAQ system**

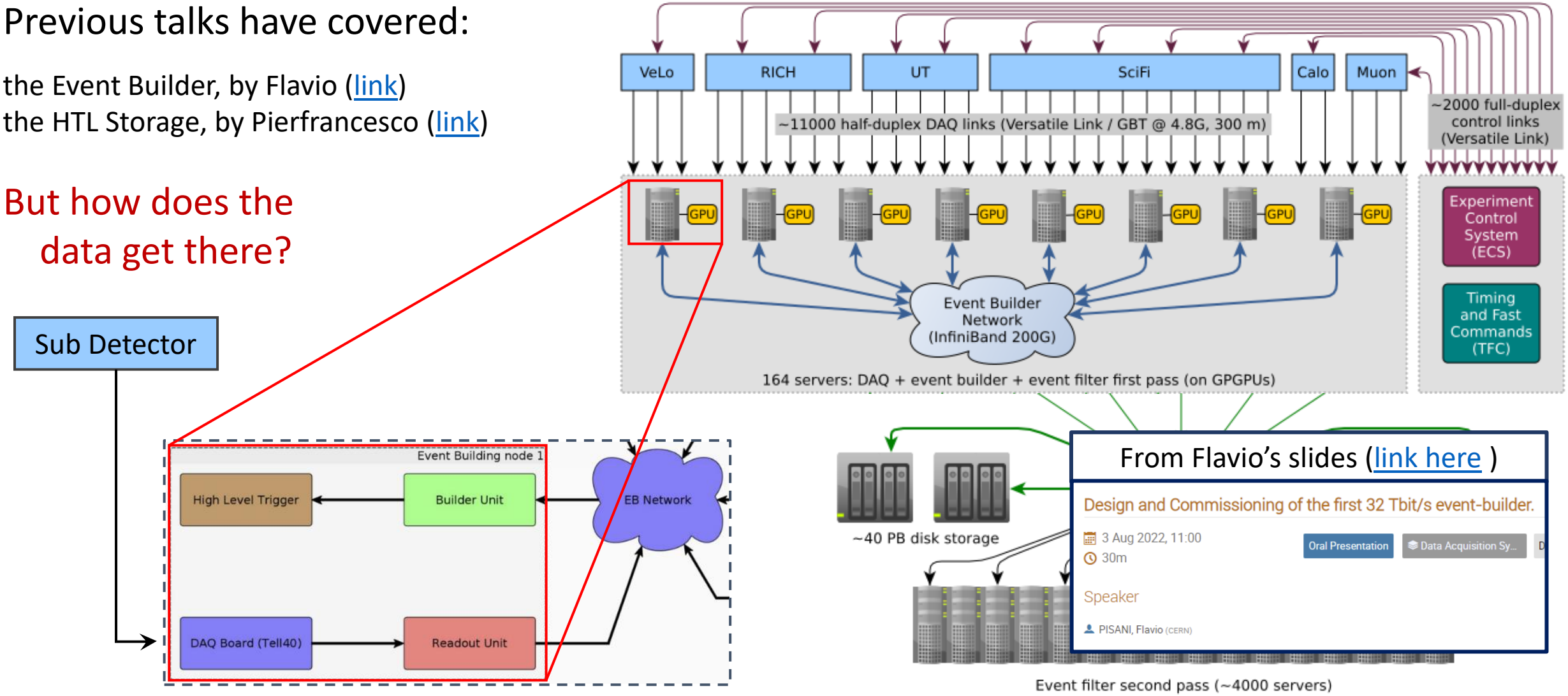


The Online DAQ System

Previous talks have covered:

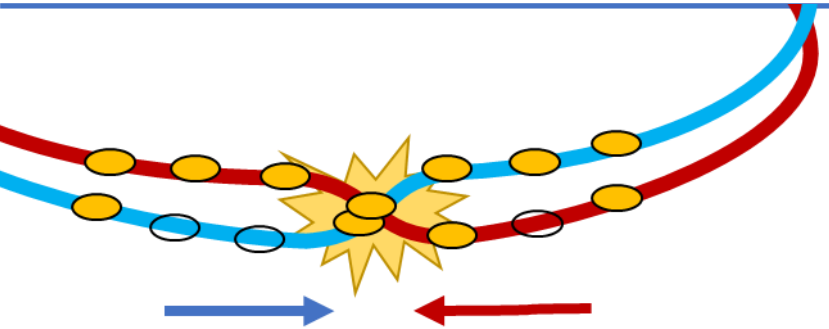
the Event Builder, by Flavio (link)
the HTL Storage, by Pierfrancesco (link)

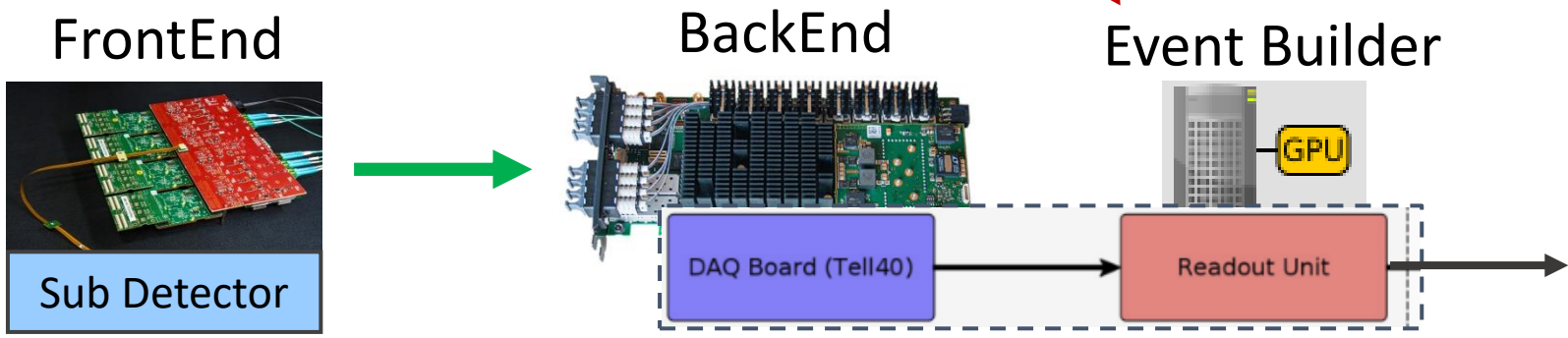
But how does the data get there?



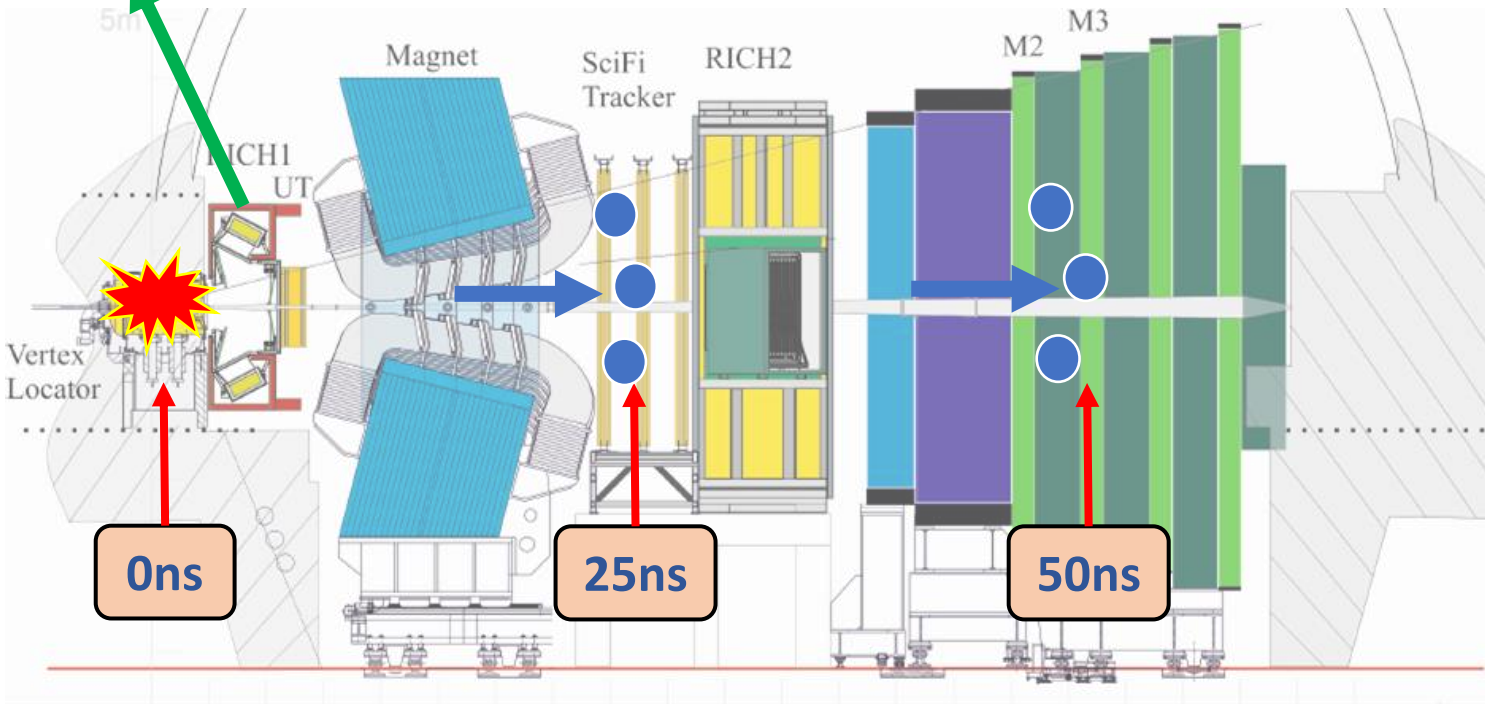
From Collision to Memory

The data path before the software





- LHC beams are divided in bunches that cross in synchrony with a clock signal
  - **~40MHz = 25ns / cross**
- Not all bunches are filled → no collision
  - **We need to select which bunch crossings to save**
- Particles arrive at different times depending on the subdetector
  - **We need to phase-align the clocks per subdetector and have fixed latency**



The TFC System: A Real-Time Architecture

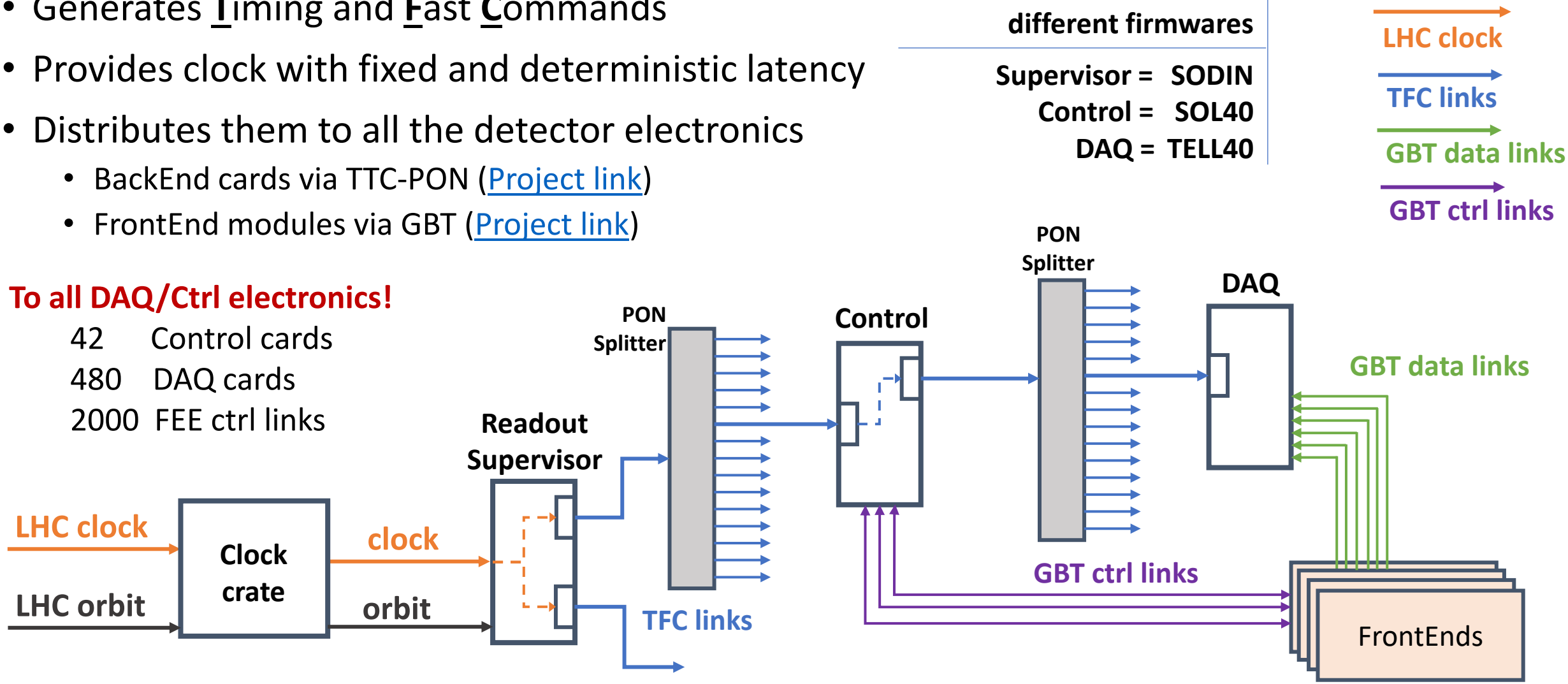
- Generates **T**iming and **F**ast **C**ommands
- Provides clock with fixed and deterministic latency
- Distributes them to all the detector electronics
  - BackEnd cards via TTC-PON (Project link)
  - FrontEnd modules via GBT (Project link)

All cards are PCIe40s with different firmwares

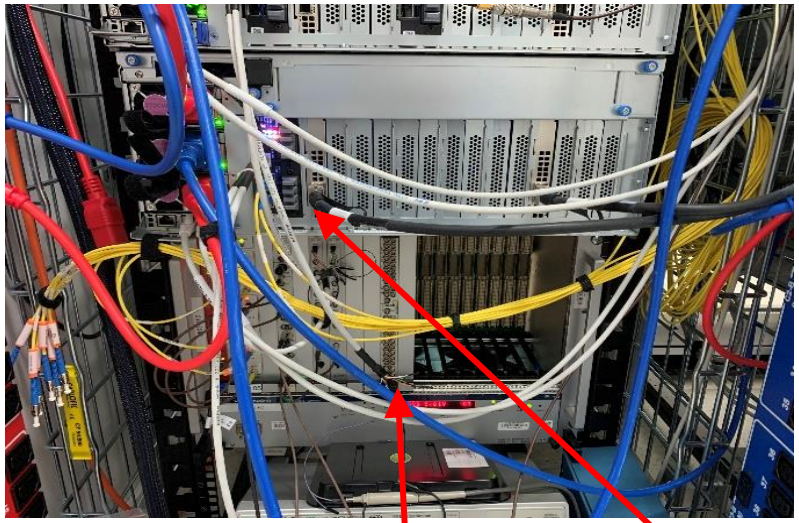
Supervisor = SODIN
Control = SOL40
DAQ = TELL40

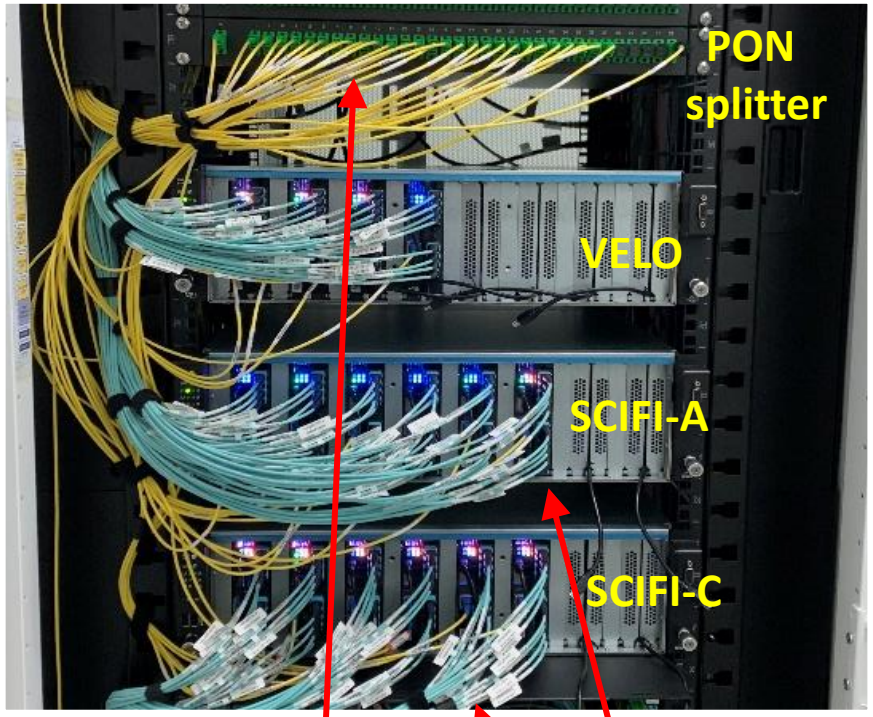
To all DAQ/Ctrl electronics!

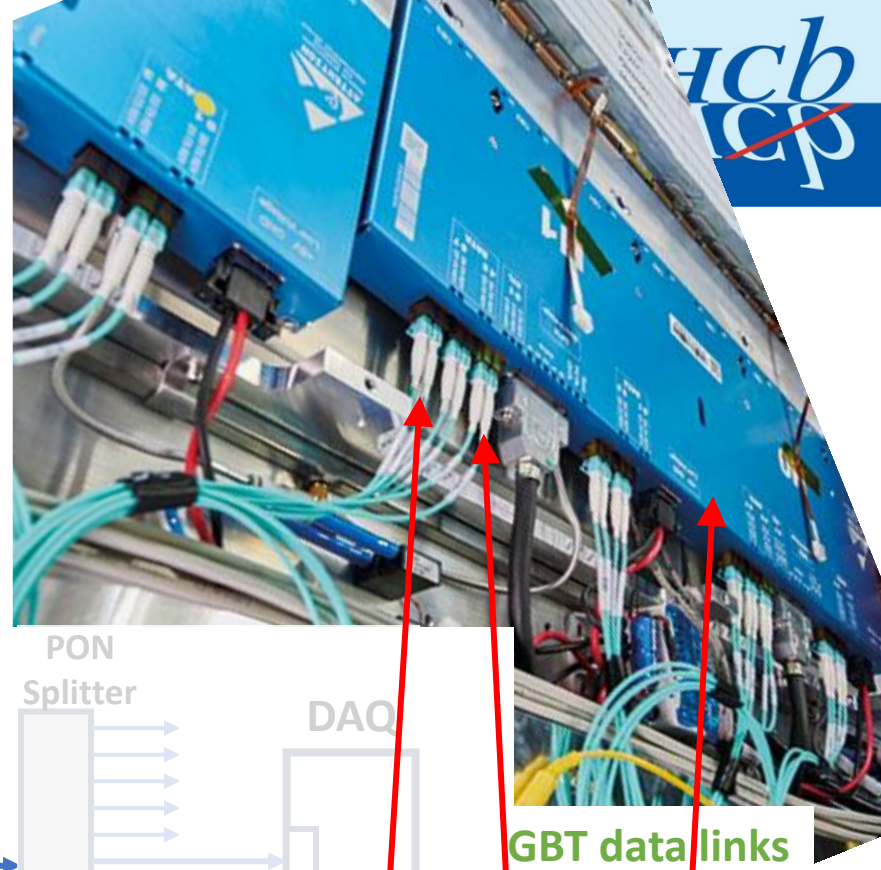
42 Control cards
480 DAQ cards
2000 FEE ctrl links



The TFC System i







PON
Splitter

Control

DAQ

GBT data links

Readout
Supervisor

LHC clock

LHC orbit

Clock
crate

clock

orbit

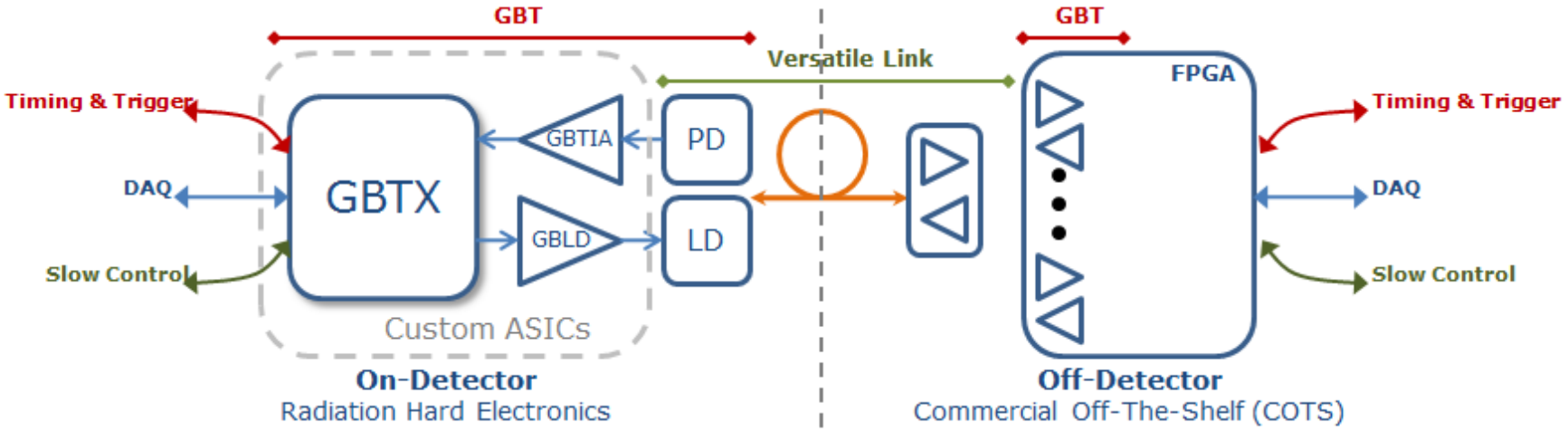
TFC links

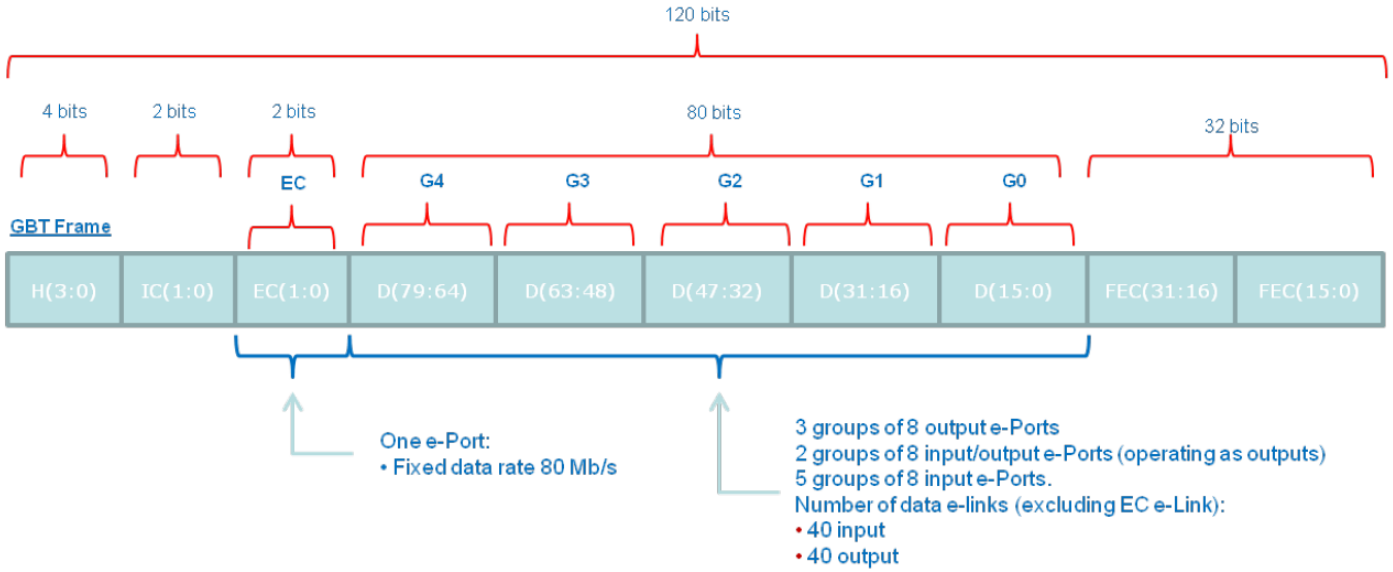
GBT ctrl links

FrontEnds

GBT Project: The GigaBit Transceiver

- The GBT Project provides a radiation hard chipset for handling control and data acquisition on frontend boards
- Designed at CERN, it is widely used on the upgrade of its experiments
- It also provides a firmware component (GBT-FPGA) that allows FPGAs to interface directly with the GBTx chip.





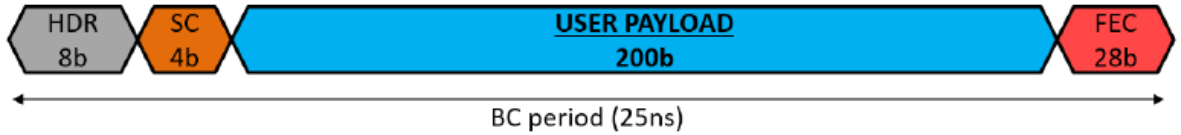
For more info:

[GBT Project link](#) (requires CERN login)

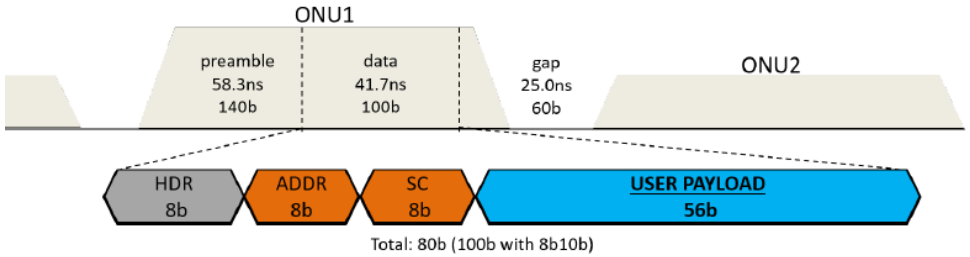
[The GBT-SerDes ASIC prototype](#) (public paper)

The TTC-PON Project

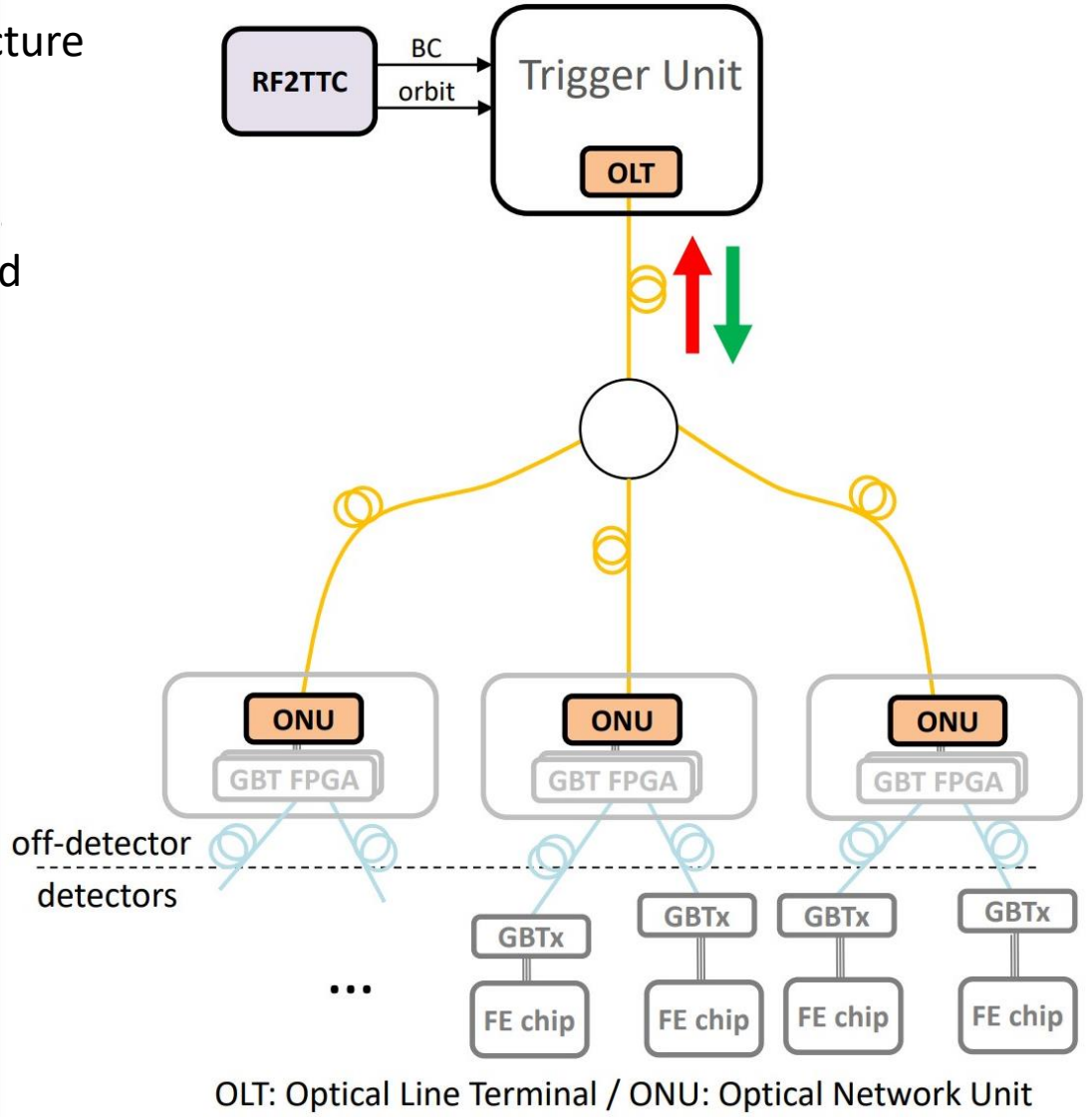
- **T**iming, **T**rigger and **C**ontrol for **P**assive **O**ptical **N**etworks
  (Link to project)
- FPGA-Based System for TFC distribution with fixed latency
- 9.6 Gbps downstream with FEC
  (8.0 for the user)



- 2.4 Gbps upstream 8b10b (Time Division Multiplexing)



- Master/ Follower architecture
  - OLT = Master
  - ONU = Follower
- Slow Control Implemented
  - Everything controllable from the master



TTC-PON: Clock Recovery

- The TTC-PON firmware manages clock and data recovery, making use of the FPGA transceivers and an external SFP+ module
- The **OLT** sets the beginning of the transmission of a new word from a strobe pulse provided by the user
- The **ONU** identifies the header position in the serial stream and generates a strobe pulse phase-aligned with the header*
- The recovered clock can be generated from the ONU Rx strobe

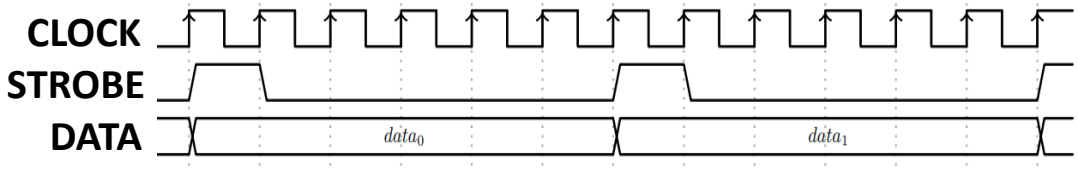
*This is achieved by using the FPGA transceivers' rxslide/bitslip functionality to shift the recovered parallel clock until it is aligned with the header. For more details:

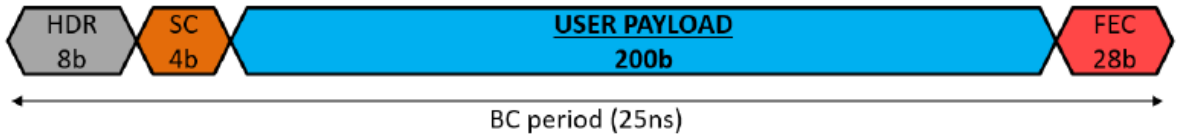
Achieving Picosecond-Level Phase Stability in Timing Distribution Systems With Xilinx Ultrascale Transceivers

Eduardo Mendes, Sophie Baron, *Member, IEEE*, Csaba Soos, Jan Troska, and Paolo Novellini

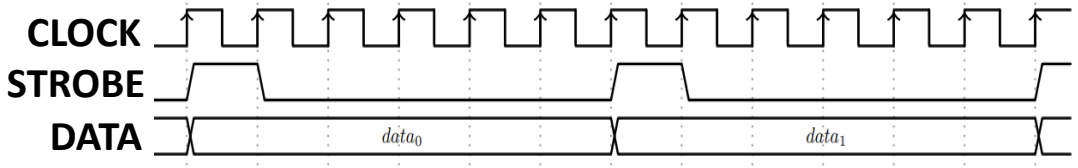
https://ieeexplore.ieee.org/document/8967127

OLT Tx





ONU Rx

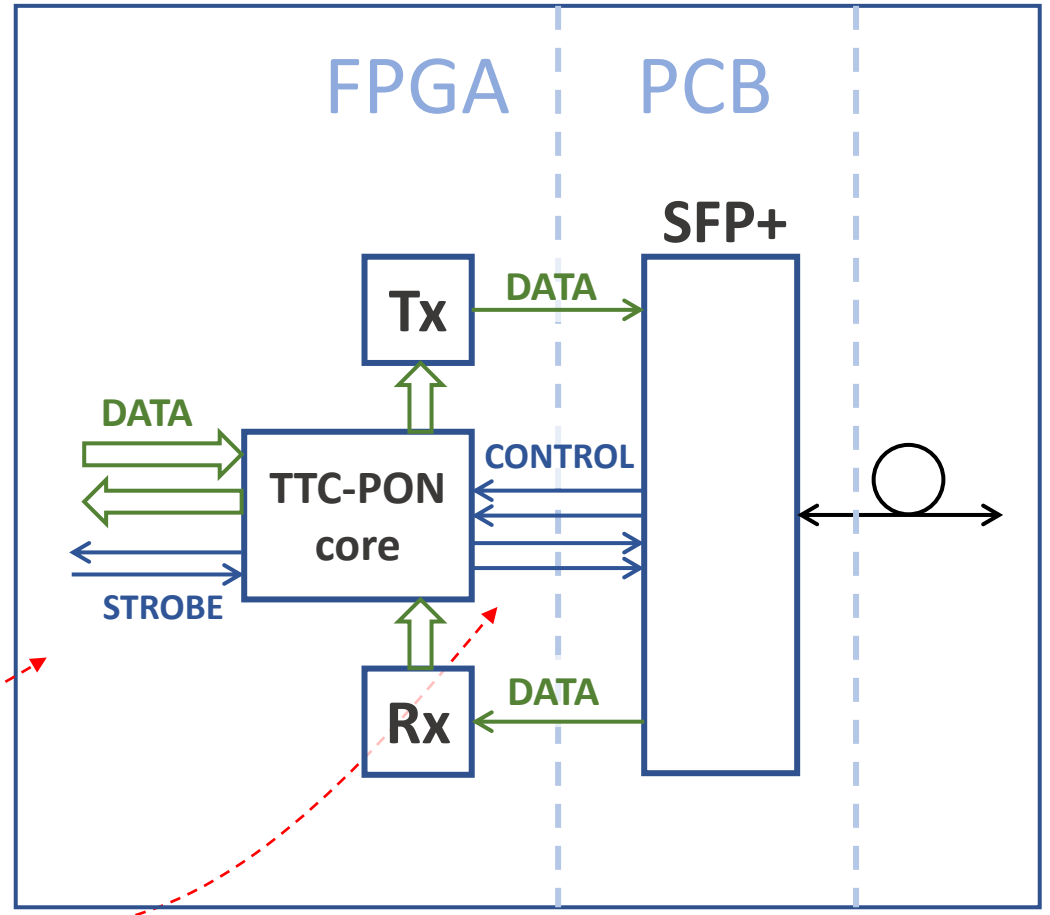


The Tx and Rx waveforms look the same, but:
On the Tx, the strobe defines the position of the data word
On the Rx, the data word defines the position of the strobe

TTC-PON: Integration into the Firmware

- The TTC-PON firmware manages clock and data recovery, making use of the FPGA transceivers and an external SFP+ module
- It provides an abstraction layer to the user
- The user still needs to handle the configuration and calibration of the transceiver, provide the interfaces of the TTC-PON core to the transceiver and the external SFP+ module and provide a way of accessing the TTC-PON registers.

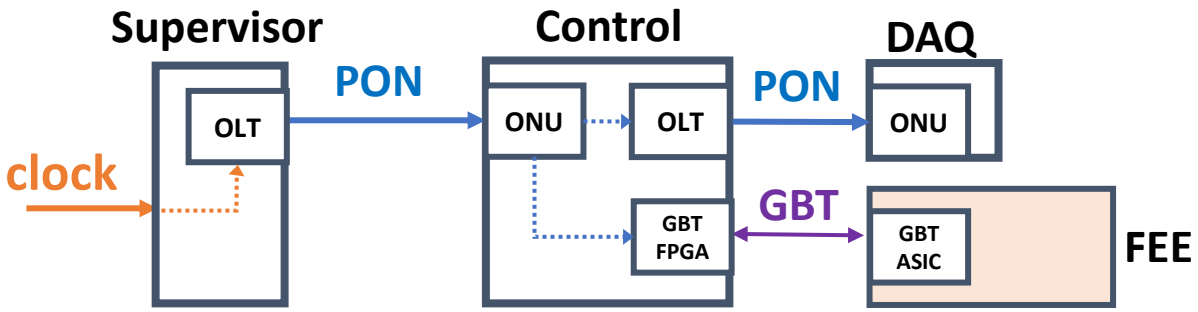
  - The strobe pulse is an input to the OLT and output from the ONU
  - The SFP+ control signals differs from the ONU and OLT

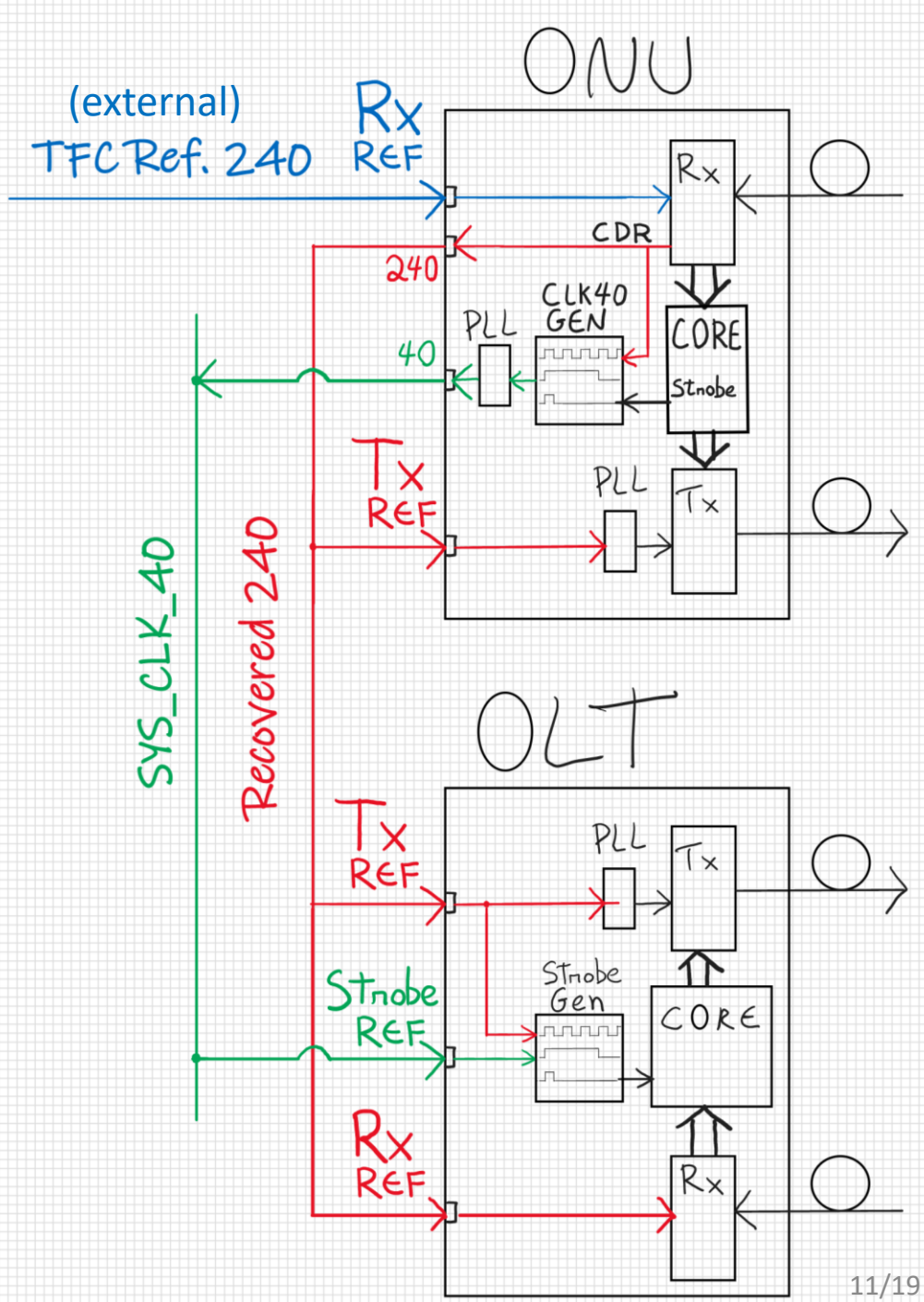


Simplified diagram of the TTC-PON core integrated into the firmware

TFC System: Clock Distribution

- The TFC System uses the TTC-PON firmware component to distribute the clock between the Supervisor and Control cards, and between the Control and DAQ cards
- In the Control card, the strobe pulse used by the OLT is generated from the 40MHz clock recovered from the ONU. This way the clock propagated downstream to the DAQ cards is phase-locked with the one received from the Supervisor card
- The clock to the FrontEnd electronics uses instead the GBT protocol
  - The BackEnd firmware uses instances of the GBT-FPGA core
  - The FrontEnd boards use the GBTx ASIC



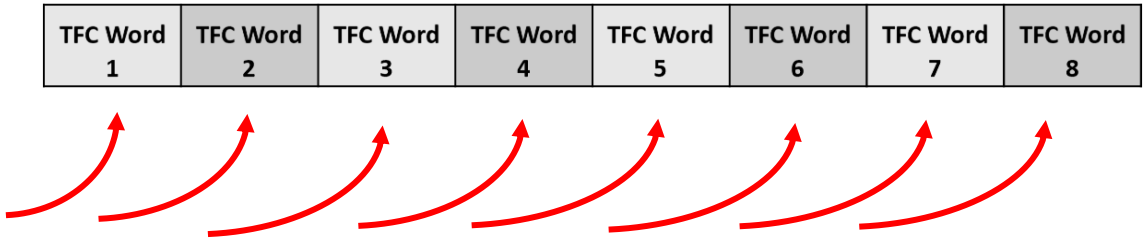


TFC System: Control Commands

The Readout Supervisor generates a TFC word (25b) for each partition* according to the run settings

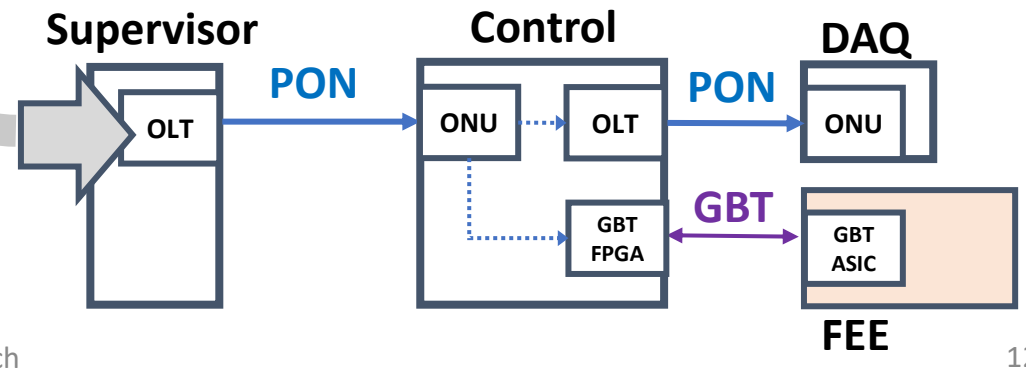
TFC Word
BXID Reset
FE Reset
Header Only
NZS Mode
Bx Veto
Trigger
Snapshot
Synch
Calib Type
Trigger Type
Etc.

The TFC words are then packed into 2 streams of 200b



And then distributed to the DAQ cards and FEE

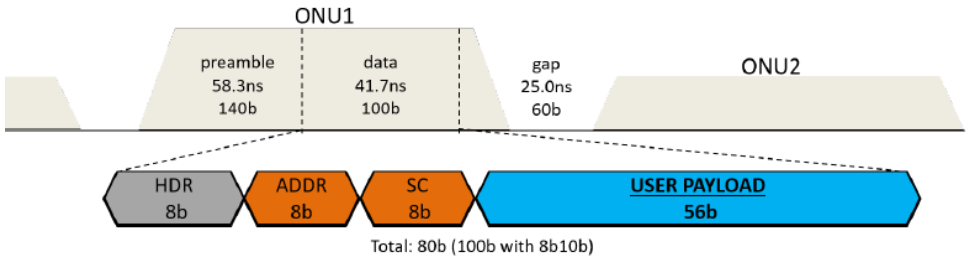
It uses the same stream of data from where the clock is recovered



* Our detector can be divided into partitions: It allows us to simultaneously send different commands and run different slices of the detector independently.
The control cards can be configured by software to "listen" to any partition.

TFC System Upstream: Real-Time Monitoring

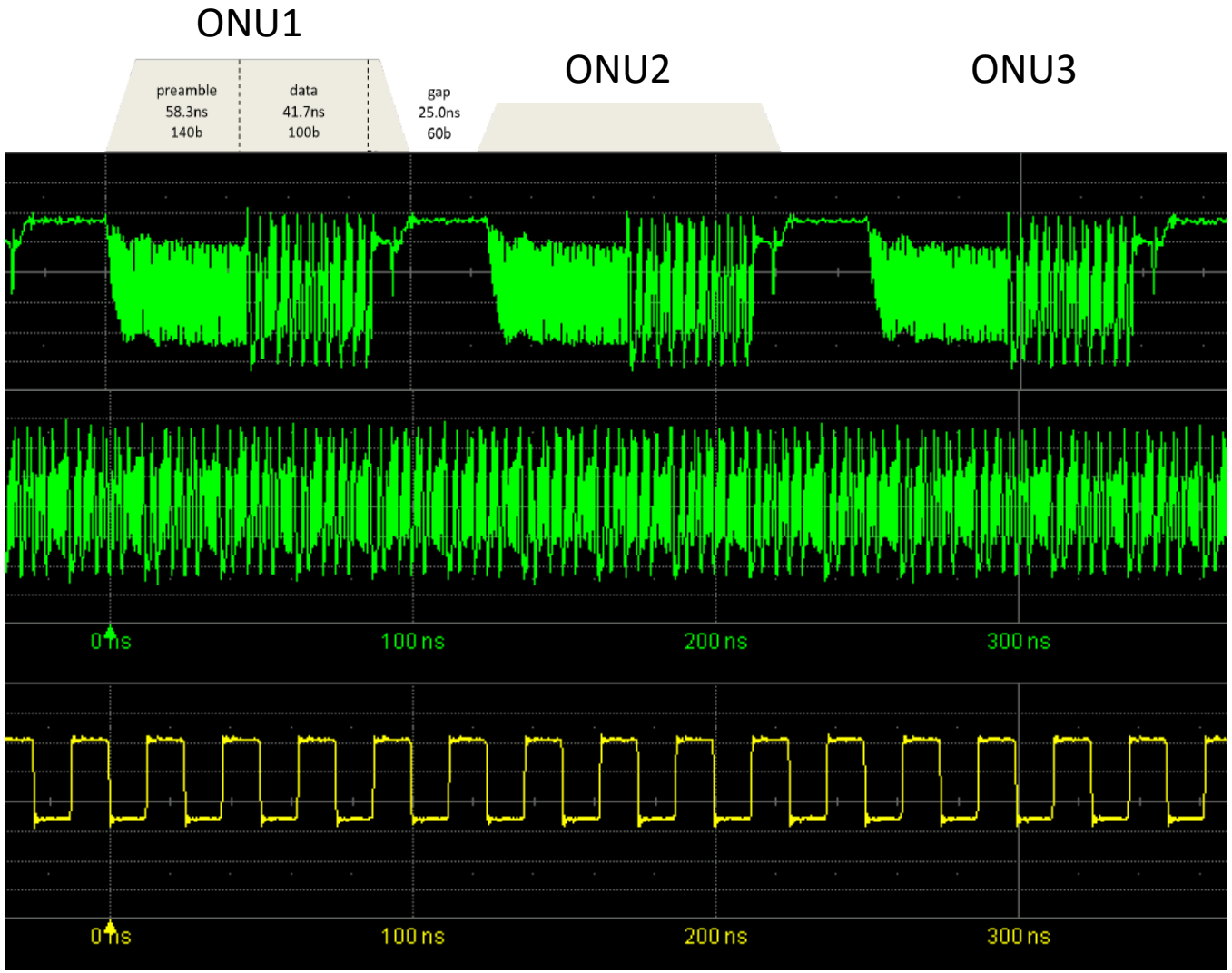
- The cards of the TFC System can also send data upstream using the TTC-PON



- The readout period is limited by the duration of a round-robin, which is proportional to the amount of ONUs in the network

- For the maximum amount of 32 ONUs in 2 levels of hierarchy, the readout period would be 32 x 2 x 5 = **320 clocks** or **8us** (**125kHz**)

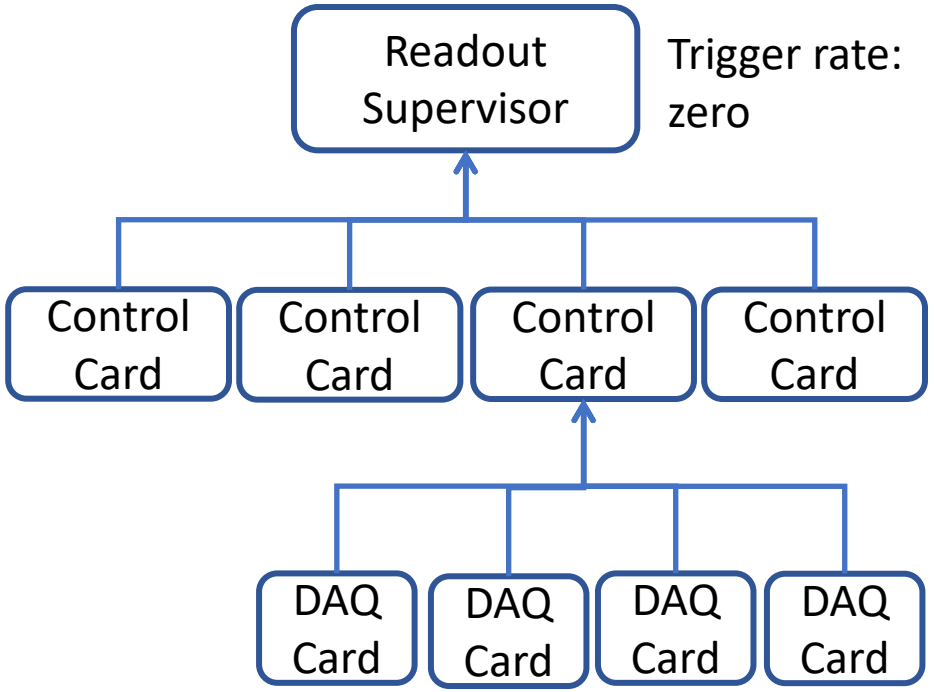
- This is the worst-case scenario when reading bits from ALL cards in the system



TFC System Upstream: Real-Time Monitoring

Application example: **Throttle**

- Throttle is a “bit” in the DAQ card that indicates when it can’t handle many more events (could be caused by any bottleneck down the data path)
- The throttle bit is sent upstream
- The Readout Supervisor automatically reacts to it by reducing the trigger rate
- This happens orders of magnitude faster than what a SCADA system would allow, preventing the whole DAQ system from going into error

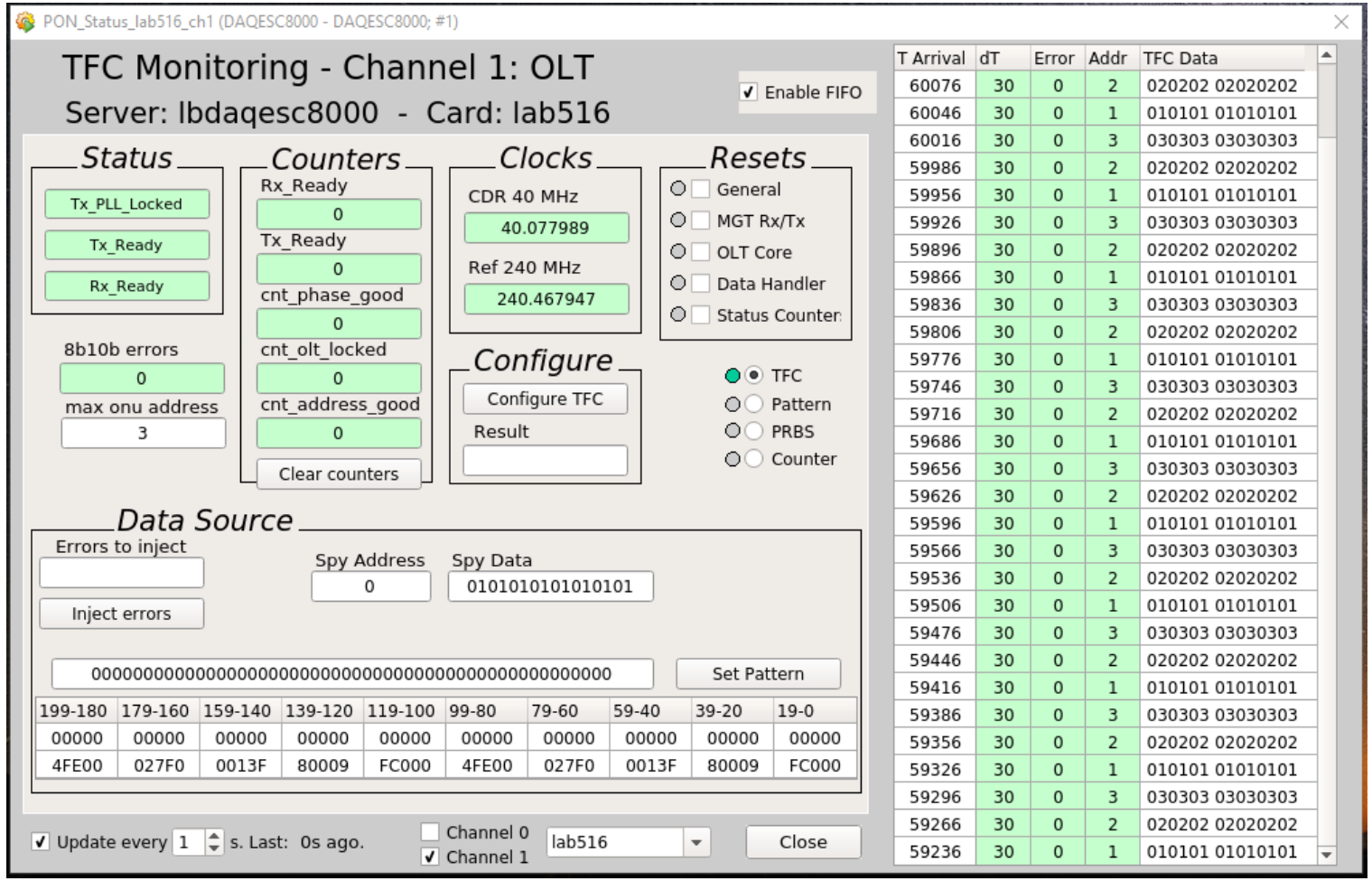


TFC System: Link Quality & Slow Monitoring

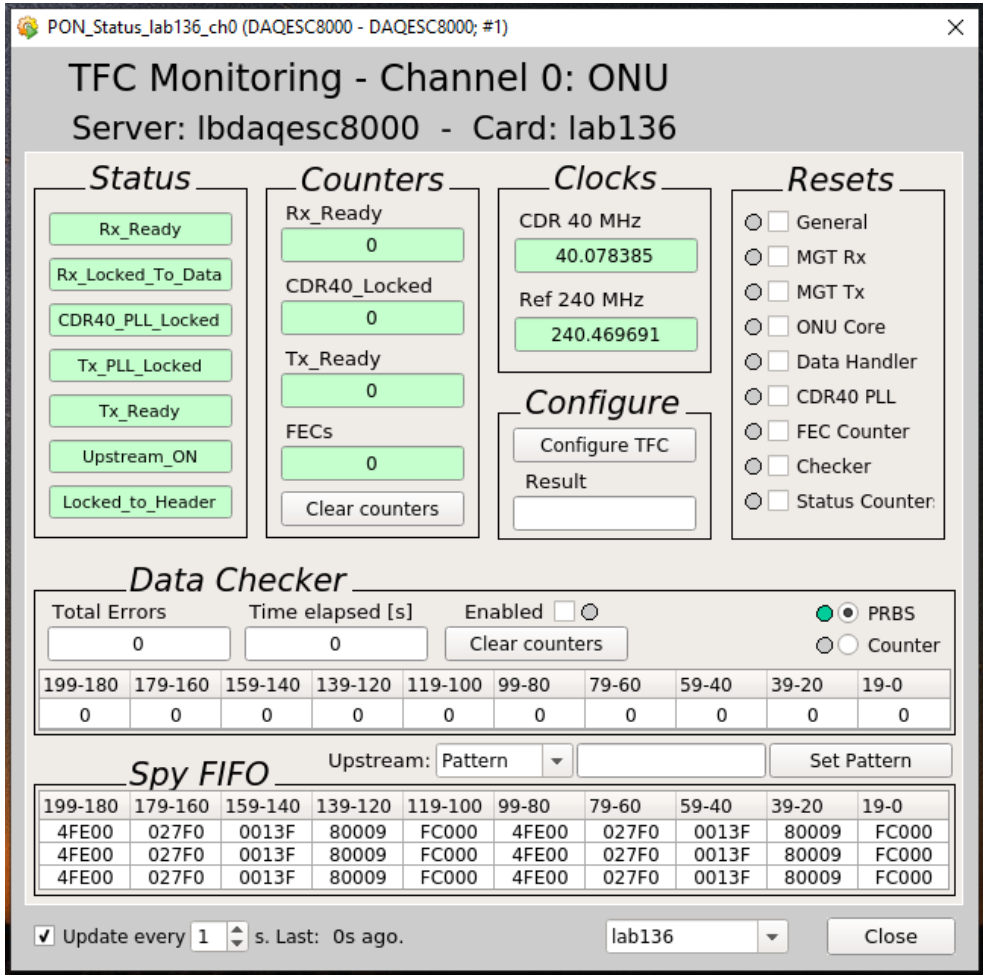
- Custom data generator + data checker
  - Pattern / Counter / PRBS
- Firmware implemented counters
  - PLLs Locked
  - Transceivers Locked / Ready
  - Header Aligned / Locked to Burst
  - FECs (downstream) / 8b10b errors (upstream)
- SCADA software for monitoring and control

TFC System: Link Quality & Slow Monitoring

OLT Monitoring Panel

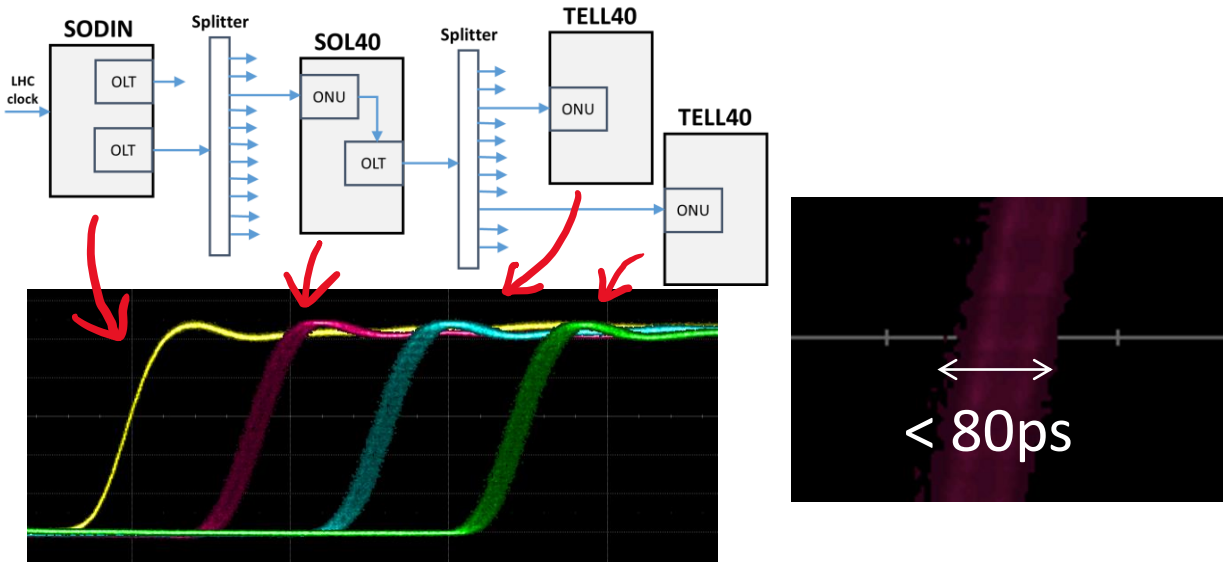


ONU Monitoring Panel



TFC System: Link Quality & Slow Monitoring

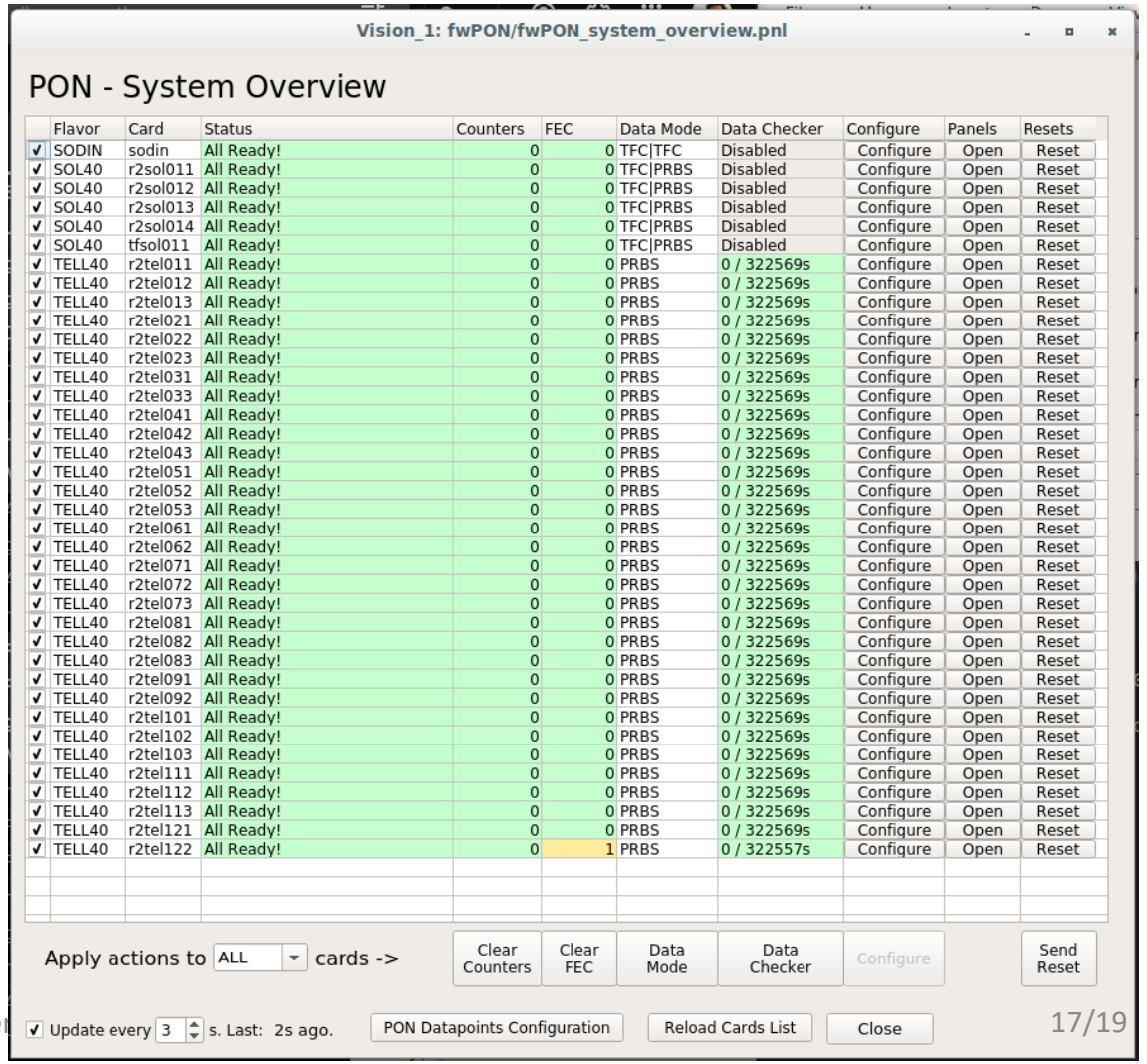
- Testing recovered clock phase
  - 36.000 resets of the Supervisor (SODIN)
  - Clocks recovered with Std. Dev. of **80ps**



- Longest PRBS Run: 32 DAQ cards for almost 4 days
  - 0 errors over 2.6×10^{15} bits transmitted
  - 1 single Forward Error Correction
  - Equivalent of 95% CL for having < 1 error per 2h for 500 cards

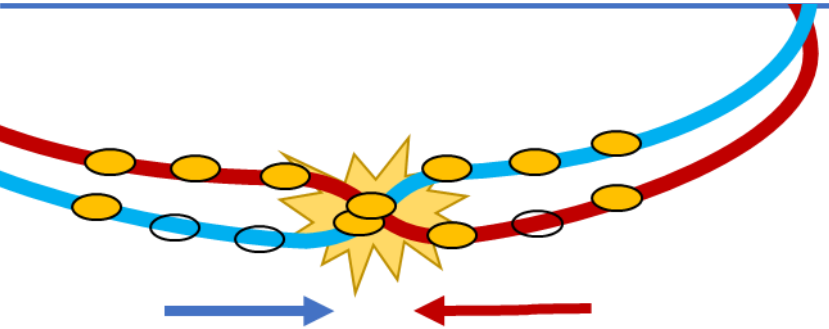
(Not all bits would actually cause an error but more statistics is still desirable)

System Monitoring Panel

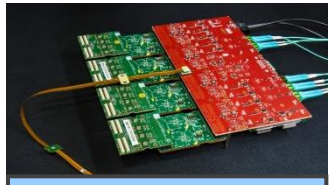


	Flavor	Card	Status	Counters	FEC	Data Mode	Data Checker	Configure	Panels	Resets
✓	SODIN	sodin	All Ready!	0	0	TFC|TFC	Disabled	Configure	Open	Reset
✓	SOL40	r2sol011	All Ready!	0	0	TFC|PRBS	Disabled	Configure	Open	Reset
✓	SOL40	r2sol012	All Ready!	0	0	TFC|PRBS	Disabled	Configure	Open	Reset
✓	SOL40	r2sol013	All Ready!	0	0	TFC|PRBS	Disabled	Configure	Open	Reset
✓	SOL40	r2sol014	All Ready!	0	0	TFC|PRBS	Disabled	Configure	Open	Reset
✓	SOL40	tfsol011	All Ready!	0	0	TFC|PRBS	Disabled	Configure	Open	Reset
✓	TELL40	r2tel011	All Ready!	0	0	PRBS	0 / 322569s	Configure	Open	Reset
✓	TELL40	r2tel012	All Ready!	0	0	PRBS	0 / 322569s	Configure	Open	Reset
✓	TELL40	r2tel013	All Ready!	0	0	PRBS	0 / 322569s	Configure	Open	Reset
✓	TELL40	r2tel021	All Ready!	0	0	PRBS	0 / 322569s	Configure	Open	Reset
✓	TELL40	r2tel022	All Ready!	0	0	PRBS	0 / 322569s	Configure	Open	Reset
✓	TELL40	r2tel023	All Ready!	0	0	PRBS	0 / 322569s	Configure	Open	Reset
✓	TELL40	r2tel031	All Ready!	0	0	PRBS	0 / 322569s	Configure	Open	Reset
✓	TELL40	r2tel033	All Ready!	0	0	PRBS	0 / 322569s	Configure	Open	Reset
✓	TELL40	r2tel041	All Ready!	0	0	PRBS	0 / 322569s	Configure	Open	Reset
✓	TELL40	r2tel042	All Ready!	0	0	PRBS	0 / 322569s	Configure	Open	Reset
✓	TELL40	r2tel043	All Ready!	0	0	PRBS	0 / 322569s	Configure	Open	Reset
✓	TELL40	r2tel051	All Ready!	0	0	PRBS	0 / 322569s	Configure	Open	Reset
✓	TELL40	r2tel052	All Ready!	0	0	PRBS	0 / 322569s	Configure	Open	Reset
✓	TELL40	r2tel053	All Ready!	0	0	PRBS	0 / 322569s	Configure	Open	Reset
✓	TELL40	r2tel061	All Ready!	0	0	PRBS	0 / 322569s	Configure	Open	Reset
✓	TELL40	r2tel062	All Ready!	0	0	PRBS	0 / 322569s	Configure	Open	Reset
✓	TELL40	r2tel071	All Ready!	0	0	PRBS	0 / 322569s	Configure	Open	Reset
✓	TELL40	r2tel072	All Ready!	0	0	PRBS	0 / 322569s	Configure	Open	Reset
✓	TELL40	r2tel073	All Ready!	0	0	PRBS	0 / 322569s	Configure	Open	Reset
✓	TELL40	r2tel081	All Ready!	0	0	PRBS	0 / 322569s	Configure	Open	Reset
✓	TELL40	r2tel082	All Ready!	0	0	PRBS	0 / 322569s	Configure	Open	Reset
✓	TELL40	r2tel083	All Ready!	0	0	PRBS	0 / 322569s	Configure	Open	Reset
✓	TELL40	r2tel091	All Ready!	0	0	PRBS	0 / 322569s	Configure	Open	Reset
✓	TELL40	r2tel092	All Ready!	0	0	PRBS	0 / 322569s	Configure	Open	Reset
✓	TELL40	r2tel101	All Ready!	0	0	PRBS	0 / 322569s	Configure	Open	Reset
✓	TELL40	r2tel102	All Ready!	0	0	PRBS	0 / 322569s	Configure	Open	Reset
✓	TELL40	r2tel103	All Ready!	0	0	PRBS	0 / 322569s	Configure	Open	Reset
✓	TELL40	r2tel111	All Ready!	0	0	PRBS	0 / 322569s	Configure	Open	Reset
✓	TELL40	r2tel112	All Ready!	0	0	PRBS	0 / 322569s	Configure	Open	Reset
✓	TELL40	r2tel113	All Ready!	0	0	PRBS	0 / 322569s	Configure	Open	Reset
✓	TELL40	r2tel121	All Ready!	0	0	PRBS	0 / 322569s	Configure	Open	Reset
✓	TELL40	r2tel122	All Ready!	0	1	PRBS	0 / 322557s	Configure	Open	Reset

Summary



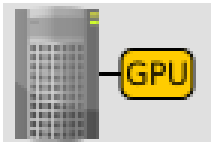
FrontEnd

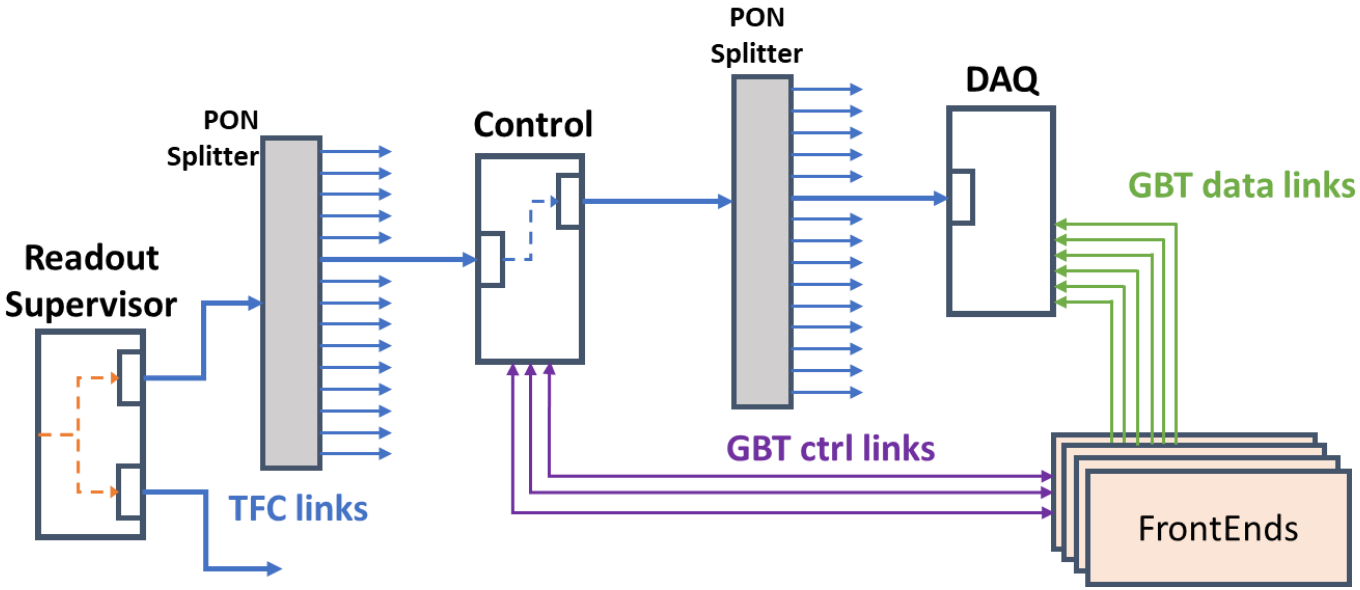


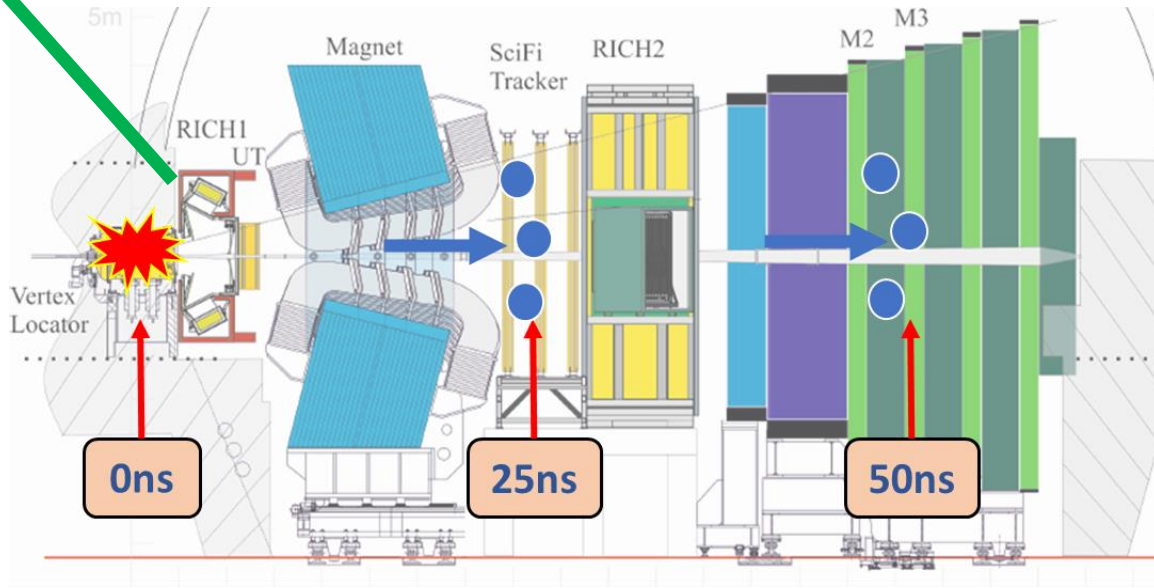
Sub Detector

DAQ Card

Event Builder







Conclusion

- Downstream system fully functional and it is currently in use
  - Since October/2021 (for commissioning of the subdetectors)
  - Since 30 days ago for data taking of LHC Run3

- Upstream system tested and being commissioned
  - Tested in a single sub-detector, expected deployment to the whole experiment by Sept/2022
  - Full characterization still pending
  - Results are, however, very promising :)

Obrigado!

Questions are welcome :)