Towards the final TFC architecture Part II
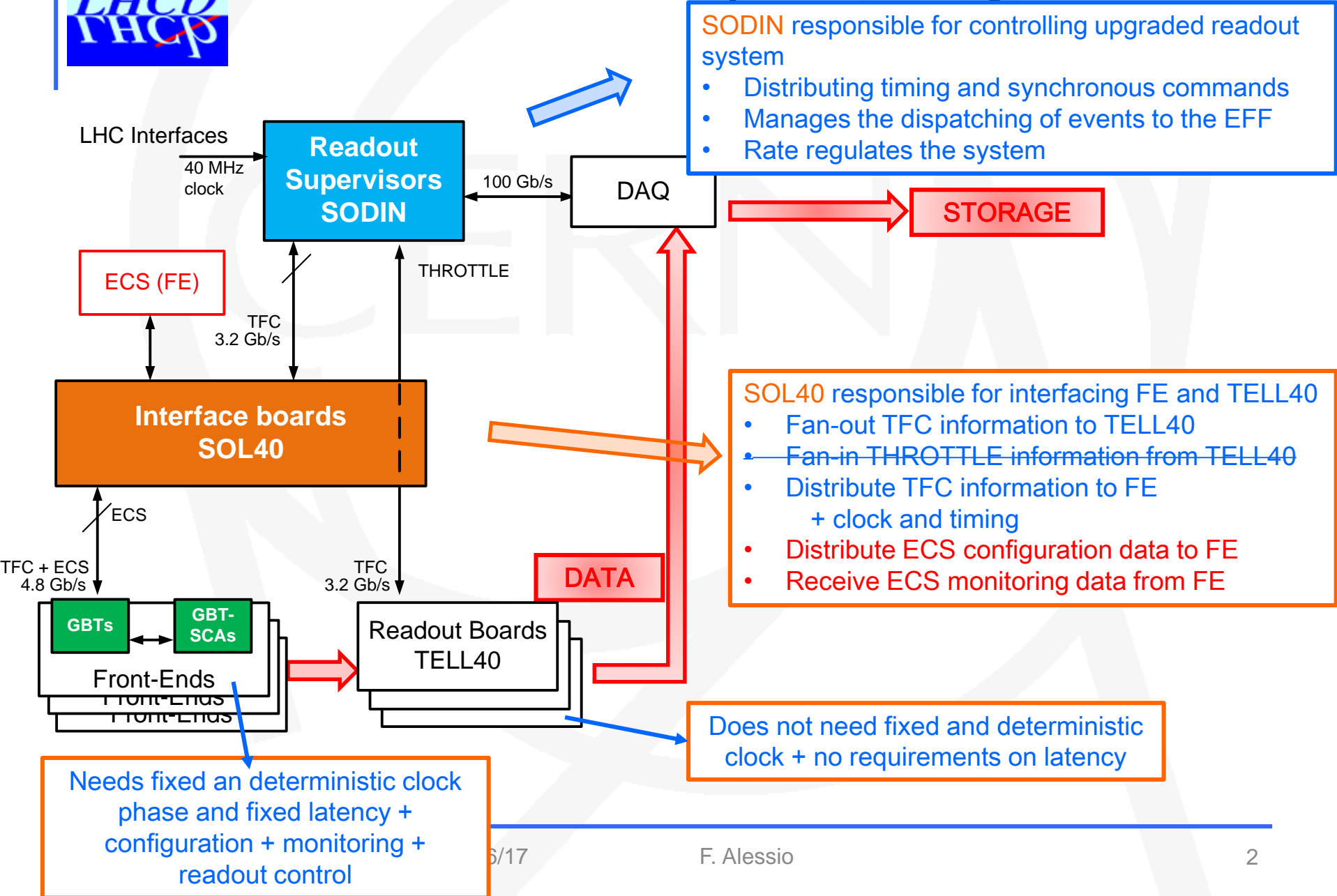
08 June 2017
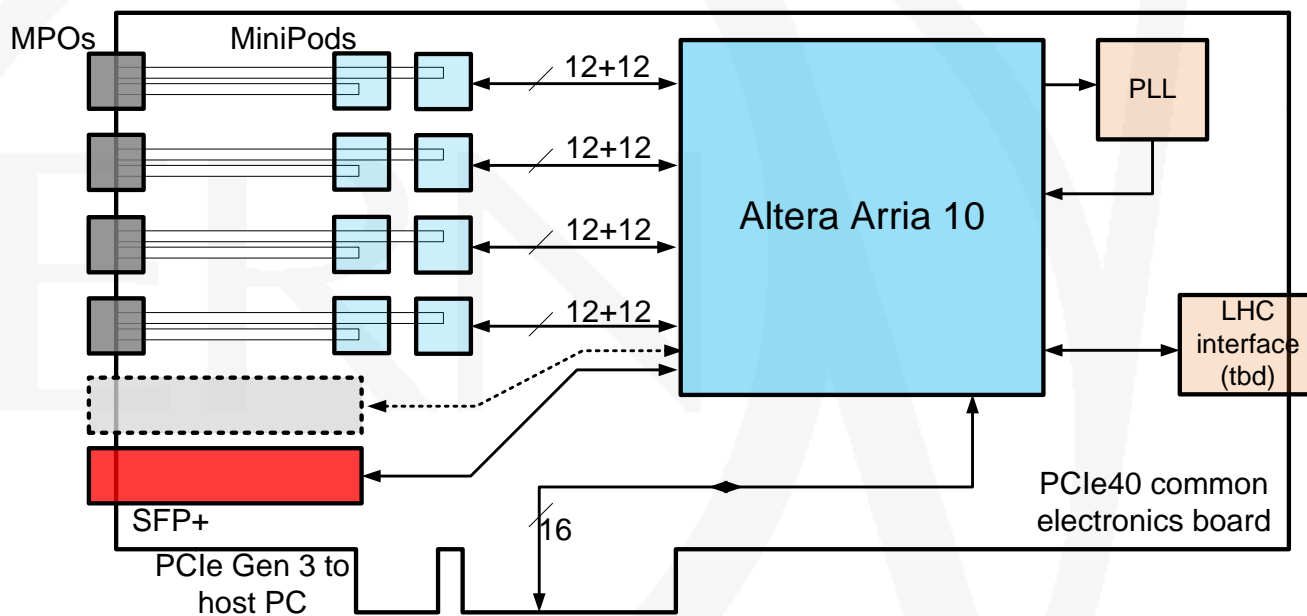
F. Alessio, CERN

The new TFC system at a glance

LHC Interfaces

40 MHz clock → **Readout Supervisors SODIN**

100 Gb/s → DAQ

SODIN responsible for controlling upgraded readout system
- Distributing timing and synchronous commands
- Manages the dispatching of events to the EFF
- Rate regulates the system

STORAGE

ECS (FE)

THROTTLE

TFC 3.2 Gb/s

Interface boards SOL40

ECS

TFC + ECS 4.8 Gb/s

TFC 3.2 Gb/s

GBTs ↔ GBT-SCAs

Front-Ends

Readout Boards TELL40

DATA

SOL40 responsible for interfacing FE and TELL40
- Fan-out TFC information to TELL40
- ~~Fan-in THROTTLE information from TELL40~~
- Distribute TFC information to FE
     + clock and timing
- Distribute ECS configuration data to FE
- Receive ECS monitoring data from FE

Does not need fixed and deterministic clock + no requirements on latency

Needs fixed an deterministic clock phase and fixed latency + configuration + monitoring + readout control

TFC on common hardware backbone

Specs:

- Firepower of 48 bidir links
- TFC interface on SFP+ (two SFP+ in final system)
- On-board external PLL for deterministic phase resynchronization

Same exact hardware (PCIe40), firmware defines the functionality: SODIN, SOL40, TELL40

- The Host PC is the PC controlling the board (ex CCPCServer/SPECServer → GbtServer)
- In principle ~100 Gbps of ECS access and data bandwidth
- External connection to TFC/LHC peripheral boards via electrical connections (LVDS)
  - As usual, tight and proficuous collaboration with CPPM

Final TFC architecture
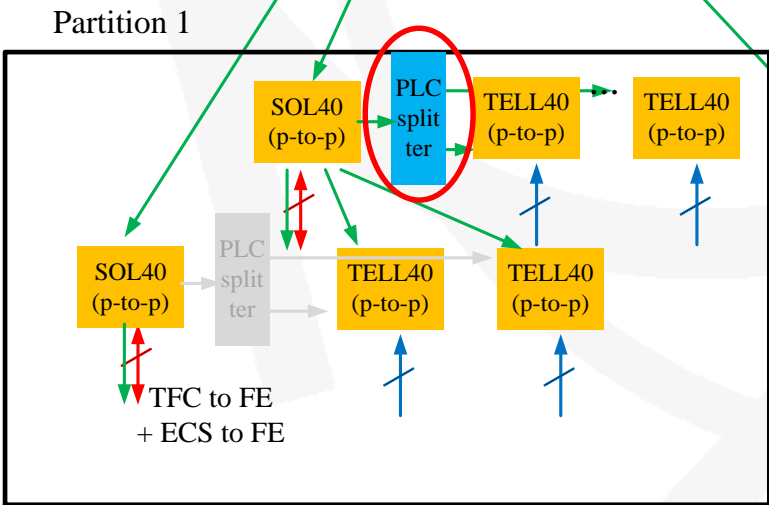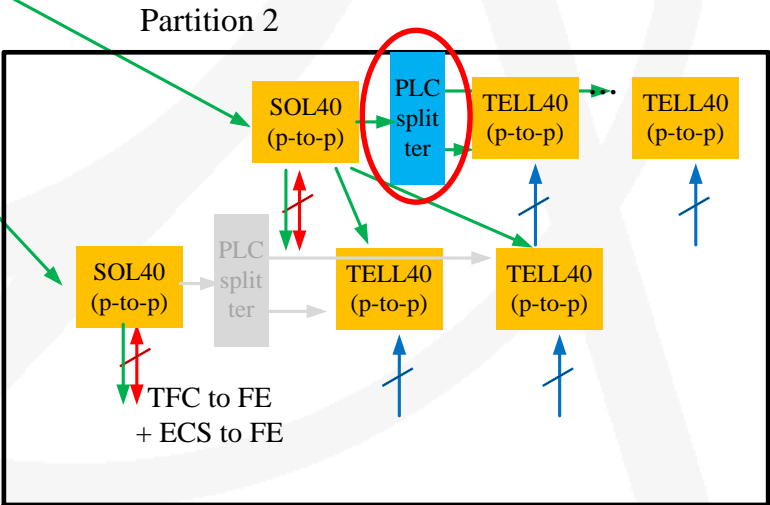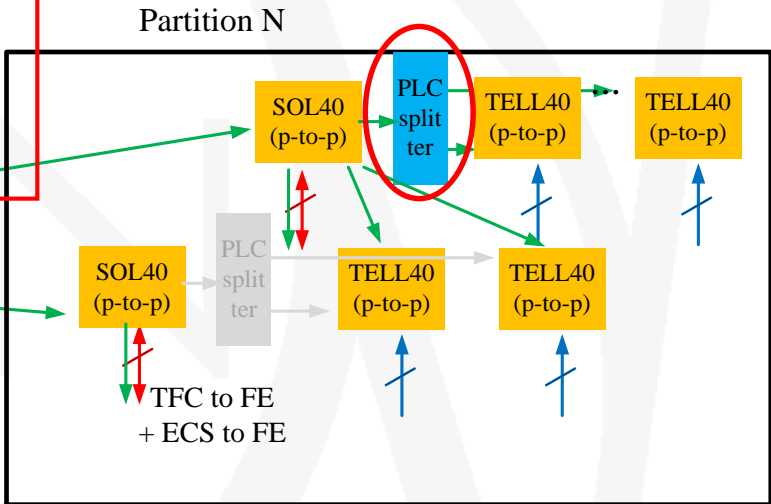
Final TFC architecture - considerations

1. **# of SODINs = 1 → single Readout Supervisor**

   → total max available destinations = 48

   - Each SODIN should also transmit event data bank → interface to DAQ + PC
   - Partitioning ensured by independent cores inside SODIN FPGA
     - ✓ Routing inside FPGA done accordingly, very simple management

2. **# of SOL40s dictated by sub-detectors needs for FEs**

   - Current numerology SOL40s for FE
     - → VELO: 4, UT: 3, SCIFI: 12, RICH: 10, CALO: 2, MUON: 6 → total = 37
   - Minimum # of splitters == number of partitions (currently 13)
     - → Profiting from the second SFP+ on PCIe40 to connect SOL40 to SODIN

	Number	# DAQ interface	# PC interface	# splitter	# MPO splitter	# extra-optics	Individual total costs	Unitary cost	Cost per type of card	Total cost
SODIN	1	1	1				16	16.0	16	
SOL40 for TELL40	0		0	13	0		6	0.5	450	466 kCHF*
SOL40 for FE	37		37				444	12.0		

* Mixture of Online and sub-detector budget... ☺

Definition of partitions

Partitioning is considered essential in our architecture

- commissioning, local tests, exclude/include
- Current numerology for TELL40s
  - → VELO: 52, UT: 68, SCIFI: 144, RICH: 102, CALO: 42, MUON:28 → total = 436

Partition:

- «minimal» ensemble of detectors that can be run independently with TFC and ECS
  - o VELO = 2 (A/C side)
  - o UT = 2 (A/C)
  - o SciFi = 4 (Q1/Q2/Q3/Q4)
 - o RICH = 3 (1 for RICH1 and 2 for RICH2, A/C)
  - o CALO = 1
  - o MUON = 1

  Total = 13

  NB: partitioning has direct impact on the (cost of the) TFC architecture, optimisation!

Final TFC architecture - considerations

3. **Good flexibility to expand the system later on or to accommodate changes**
   - More available output from SODIN → ~11
     - ✓ Can even include other partitions later on

4. **No more bidirectionality with TELL40s**
   - Possible that the backpressure information will be transmitted by network to SODIN for fast feedback
   - In any case, **information must be available to ECS for monitoring**

Partition N
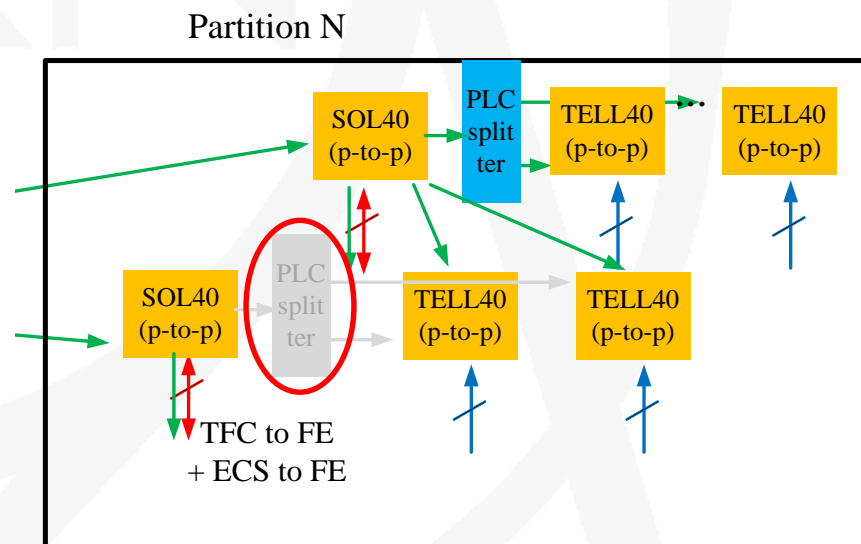
Considerations on optical splitters

Architecture supposes the use of optical splitters to distribute TFC to TELL40

→ Choose PLC* over FBT** cause not sensitive to wavelength (to be used at 1310nm) and higher degree of splitting (can go up to 1:64)

→ Better stability across all ratios

Not enough power budget from PCIe40 through standard SFP+

Specification

| Parameter		Specifications											
Operating Wavelength(nm)		1260~1650											
Type		1×2	1×4	1×8	1×16	1×32	1×64	2×2	2×4	2×8	2×16	2×32	2×64
Insertion Loss(db)	typical	3.6	6.8	10	13	16	19.5	4	7	10.5	13.5	16.5	20.5
Uniformity(db)	typical	0.4	0.5	0.5	1	1	2.5	0.6	0.8	1	1.5	2	2
PDL(db)Max	typical	0.1	0.1	0.1	0.2	0.2	0.2	0.1	0.1	0.1	0.2	0.2	0.2
Wavelength department loss(db)	typical	0.1	0.1	0.1	0.1	0.2	0.2	0.1	0.1	0.1	0.1	0.2	0.2
Temperature department loss(db)	typical	0.3	0.3	0.3	0.4	0.4	0.4	0.3	0.3	0.3	0.4	0.4	0.4
Directivity(db)Min		55											
Return Loss(db)Min		55(50)											
Operating temperature		-40~+85											
Strorage temperature		-40~+85											

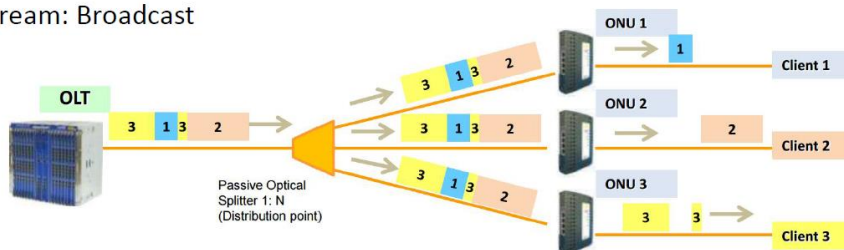*Planar Lightwave Circuit
**Fused Biconical Taper

Optics from PON?

EP-ESE group has been looking into PON for the future TTC systems since many years and they have found some qualified optics for this
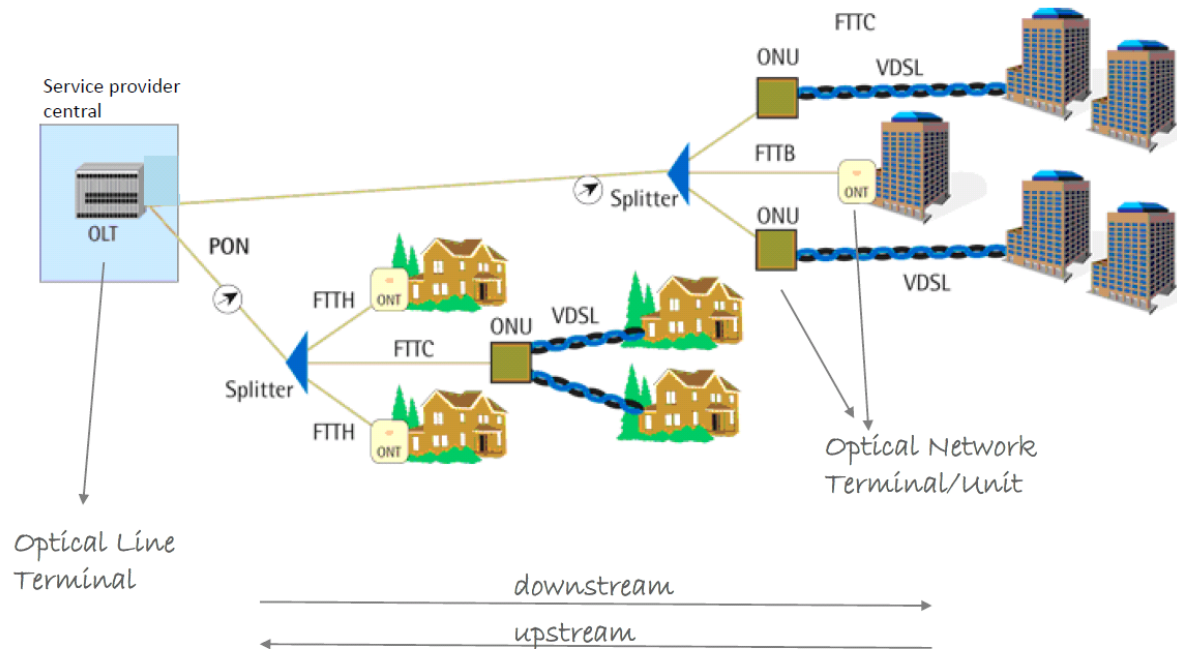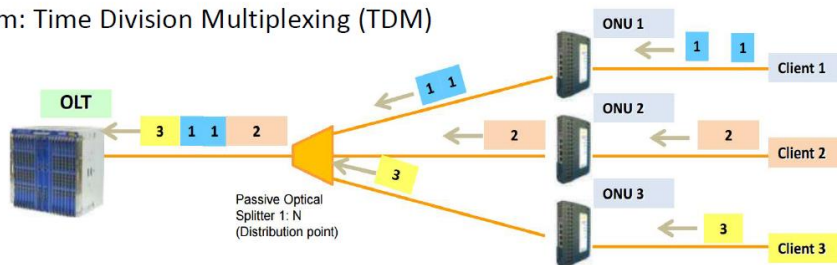
Optics from PON?

EP-ESE group has been looking into PON for the future TTC systems since many years and they have found some qualified optics for this

Latest results from optics

Test setup in lab

SFP+ → GoFoton/SuperXon optical modules (N2a class)

- OLT: SOGX6292-PSGB
- ONU: SOGX2629-PSIGA
- Same size as the SFP+, PCIe40 already designed to accept such modules

Splitters → CubeOptics PON

- 1:64 split ratio (composed of 1:32 and 1:2)

Single Mode Fibers, 200m (NB: each km adds -0.5 dB of loss)

S. Baron et al.

Label	Contribution	Power
A	Downstream Power Budget (with FEC)	29.68 dB
B	Attenuation (1:64)	22.8 dB
C = A- B	**Downstream Optical Power Margin (with FEC)**	**6.88 dB**
D	Downstream Power Budget (with no FEC)	26.68 dB
E= D-B	**Downstream Optical Power Margin (with FEC)**	**3.88 dB**

Latest results from optics

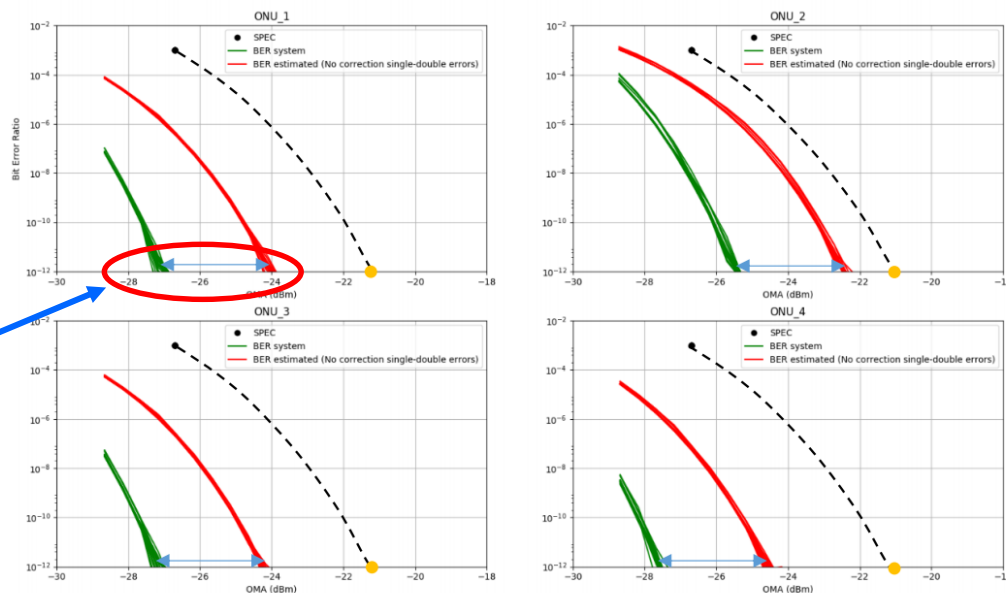**Already enough power budget with no FEC**

→ **FEC adds 3 dB of power to the margin**

S. Baron et al.

- FEC used is BCH(120,106) with single- and double-error correction
- Running at XGPON TTC adapted bandwidth (9.6 Gbps)

Decided to tag along with ESE group and ALICE to buy the optics for OLTs and ONUs to distribute TFC to the TELL40

Gain due to FEC and PON

--------	BER plots using the FEC (50 plots with confidence level of 95%)
--------	BER plots extrapolated from green plot for an scrambled but uncorrected stream
●	Maximum Rx Sensitivity specified on ONU datasheet with PRBS 2^{31} -1test pattern @9.953Gbit/s, BER ≤1×10^{-3} (converted in OMA)
--------	Extrapolated specified BER plot from datasheet value + measured BER plot
●	Extrapolated maximum receiver sensitivity (EOL=End Of Life) with PRBS 2^{31} -1test pattern @9.953Gbit/s, BER ≤1×10^{-12}
←→	FEC coding gain

Final TFC architecture

Protocols:

SODIN ←→ SOL40

8b10b classical with internal FPGA CDRs, this link to recover and regenerate clock

SOL40←→TELL40

Plan A: TTC-PON protocol (no FEC)
Plan B: 8b10b classical with internal FPGA CDRs

Possibility to transmit low-bandwidth information back to SOL40 for ECS monitoring (upstream)

SOL40 ←→ FE

GBT protocol using GBT-FPGA core

Final TFC architecture – extra costs

Equip each SOL40 used to control TELL40s in a partition with OLT specific optics

- Minimum number 13, but allow for each partitions to have at least two SOL40-OLTs → 26

- Cost enquiry is around +500 USD per optics wrt classical SFP+

Equip all TELL40s with ONUs specific optics

- No extra-cost wrt to SFP+, cost enquiry shows comparable values (100USD)

Decided to add 35% spares to make sure we don't risk any technology obliteration

- 35 OLTs
- 564 ONUs
- 35 splitters

} Cost of 91.4 kCHF (wrt 73.9 kCHF w/o PON optics)

Final TFC architecture – extra considerations on location

TTC-PON optics are qualified for >> km distance and split ratio 1:64

→ Leave freedom to where to physically put TFC system

(reminder: data center decided to be at pt8 in surface)

Case 1: put ALL TFC system undeground (D2/D3)

Pros:

- Short distances between SODIN-SOL40 and SOL40-FE
- Extra free space in container
- SODIN located where LHC interfaces are

Cons:

- Need to pull ~3000 duplex fibers across shielding wall (TFC to FE)
- Need to pull ~50 single mode fibers upstairs (splitters being upstairs)
- Need to bring SODIN data bank upstairs

Final TFC architecture – extra considerations

TTC-PON optics are qualified for > km distance and split ration 1:64

→ Leave freedom to where to physically put TFC system

(reminder: data center in surface)

Case 2: put ALL TFC system in surface

> Pros:
>
> - Everything in the same place, no need to pull extra-fibers
> - Short single-mode fibers between SOL40→splitter→TELL40

> Cons:
>
> - SODIN not in same place as LHC interfaces, re-route LHC clock
>   NB: LHC clock already in surface, so minor intervention
> - Long distance SOL40→FE for timing distribution
> - No extra free space in container (TFC is ~10% of the system)

Discussion already started on where to put everything,

Outlook

MiniDAQ2 are now available at CERN. Next steps:

- Create a small cluster: SODIN->SOL40->TELL40(and generic FE)
- Re-implement 8b10b TFC protocol in Arria X with specific SFP+ link
  - Work in collaboration with Marseille, as usual ☺
  - Perform tests with latency, clock phase and timing/readout control distribution
- Obtain few samples of new optics and splitter and re-do tests
- Include new TTC-PON firmware core in firmware and re-do tests
- The dream is to run few MiniDAQ2 in parallel to the current experiment... ☺

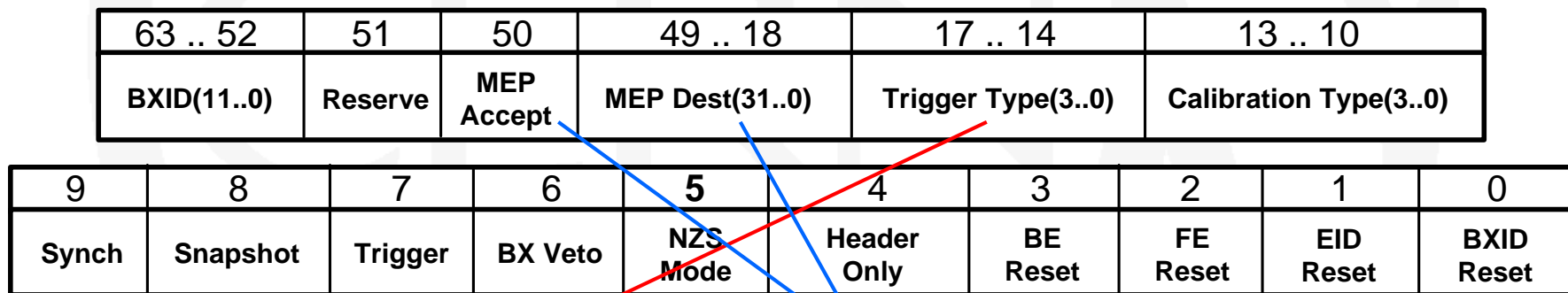NB: numbers for SOL40+TELL40 from sub-detectors have an impact on TFC architecture.

Backup

Costs per item

Costs per item	kCHF
DAQ interface	4
PC interface	4
PCIe40	8
OLT transmitter	0.5
ONU receiver	0.1
PON splitter	0.5
MPO splitter	0.15
MiniPODs (pair of tx/rx)	0.9

TFC commands and data

TFC information will remain the same: 64 bits @ 40 MHz encoded in 8b10b

→ 2.56 Gbps user data, 3.2 Gbps with encoding (no extra logic needed)

63 .. 52	51	50	49 .. 18	17 .. 14	13 .. 10
BXID(11..0)	Reserve	MEP Accept	MEP Dest(31..0)	Trigger Type(3..0)	Calibration Type(3..0)

9	8	7	6	**5**	4	3	2	1	0
Synch	Snapshot	Trigger	BX Veto	NZS Mode	Header Only	BE Reset	FE Reset	EID Reset	BXID Reset

MEP accept command when MEP ready:
- Take MEP address and send all fragments to that address
- Dynamic mechanisms based on well-oiled mechanism used today
- Investigations ongoing to see if a different mechanism is more suited for upgraded system
- This may possible be dropped at the TFC level

Trigger command to accept events for that BXID:
→ Trigger acting as a VETO to rate regulate the system
- Not based on physics decision (NO LLT)
- Unbiased (random) rate regulation of the system by reducing the rate at which trigger is set → based on «gears»

A subset of these commands is relayed to the FEs

Rate regulation to reduce output bandwith

→ Rate regulation in SODIN is considered as the safest mechanism to reduce output bandwidth or relieve backpressure

→ Out of the full 40 MHz, only ~28 MHz contains beam-beam, ~4 MHz contains beam-empty/empty-beam, ~8 MHz contains emtpy-empty
   - ✓ SODIN can have a programmable mask that defines the rate of crossing types to be kept
   - ✓ 100% for bb, 25% for eb/be and 10% for ee
   - ✓ Effectively the full input rate is 30 MHz.

	ee	$eb+be$	bb
Number of crossing	762	180+180	2442
Rate	8.6 MHz	4.06 MHz	27.5 MHz
Accepted fraction	10%	25%	100%

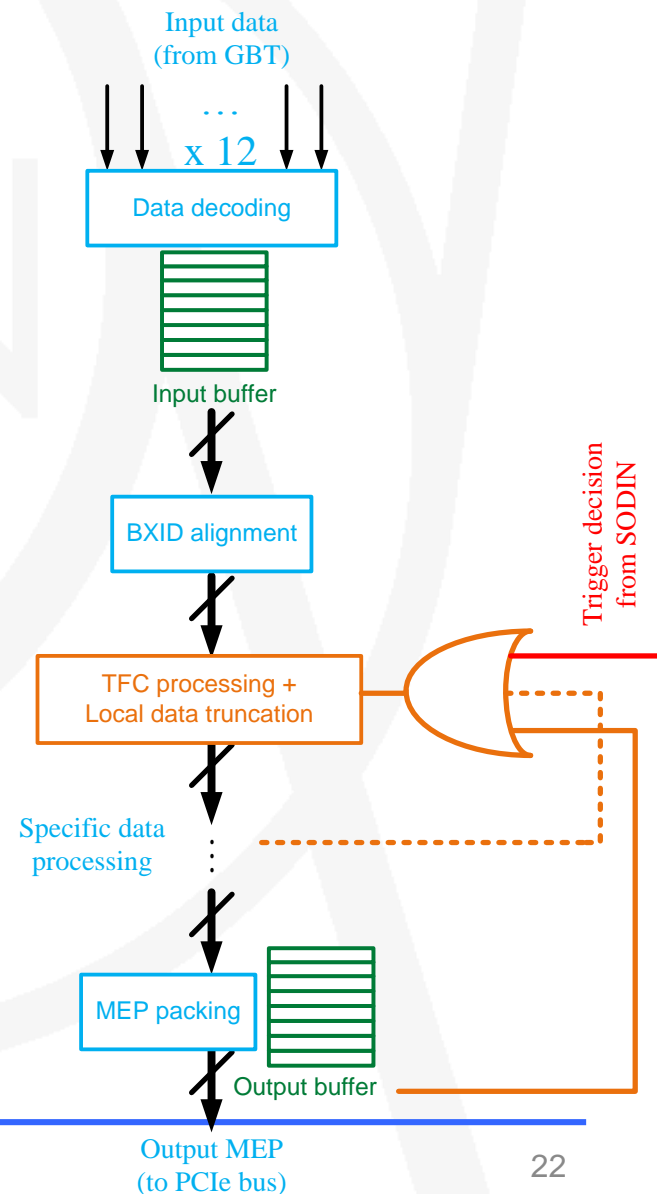→ This was described in EDMS 1606939 as Renaud mentioned a while ago.

Drop SODIN-TELL40 bidir

Upstream communication seems the most complicated issue in the TTC-PON architecture and probably we don't even need it

1. TELL40 will throttle locally
2. Roundtrip between SODIN-TELL40 for a specific BXID would be complicated and adds logic in the TELL40
3. Need to fix a maximum latency (SODIN-TELL40 for trigger)
   - Complicated in a fully asynchronous system...

Information on backpressure rather sent through network : will be investigated

➔ SODIN could apply «gears» to its trigger rate by reducing the rate by the same proportion

Input data (from GBT)

x 12

Data decoding

Input buffer

BXID alignment

TFC processing + Local data truncation

Specific data processing

MEP packing

Output buffer

Trigger decision from SODIN

Output MEP (to PCIe bus)

TFC output bandwidth

SODIN will also transmit a data bank to the DAQ for each «accepted event»

- Also defines the MEP size (can be dynamic and also very big)
- Currently is 44 bytes @ 1 MHz → 352 Mbps / 44 bytes @ 40 MHz → ~14 Gbps
  - ✓ Can be expanded to include more informations as we have ~100 Gbps for central data taking
  - ✓ Can be reduced to allow **max 10 partitions to run in parallel** (~10 Gbps → 32 bytes/8 words)

| Word	31 30 29 28 27 26 25 24 23 22 21 20 19 18 17 16	15 14 13 12 11 10 9 8 7 6 5 4 3 2 1 0
#H0	Size in bytes = 44	Magic Patter = 0xCBCB
#H1	Source ID = "ODIN S/N"	Version = 2	Type = 0x10
#D0	Run Number (31 .. 0)	
#D1	Event Type (31 .. 0)	
#D2	Orbit ID (31 .. 0)	
#D3	Event ID (63 .. 32)	
#D4	Event ID (31 .. 0)	
#D5	GPS Time (63 .. 32)	
#D6	GPS Time (31 .. 0)	
#D7	Error Flags	Detector Status (23 .. 0)
#D8	LHC info	BX Type	R	R	Trigger Type	R	R	R	R	Bunch ID

Last topic: SOL40 firmware
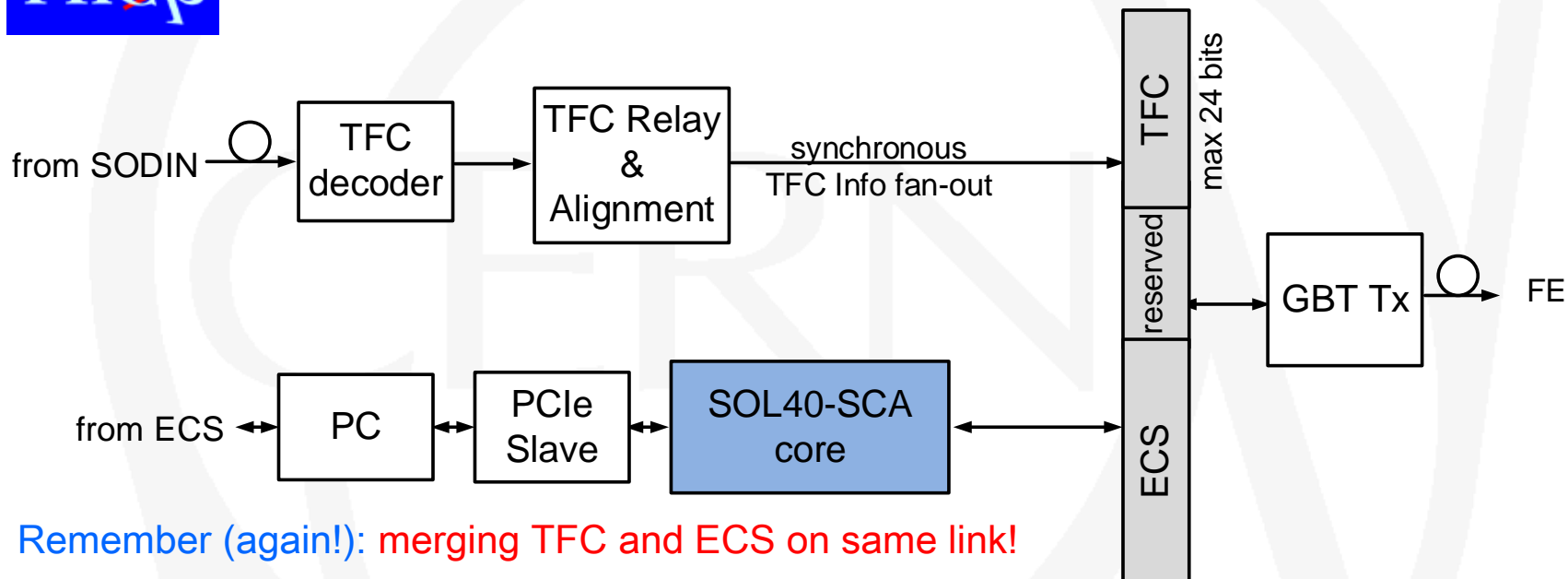
Remember (again!): merging TFC and ECS on same link!

→ SOL40 firmware will take care of doing all the complicated bit manipulations needed to control GBTs, its SCAs and your FE chipsets

- Single firmware for everybody with all necessary features and requirements
    - → minimize unconformities, highly programmable and customizable
    - → modular (can get a core and put it somewhere else in an eval board)
- Centrally provided with ECS support

→ In addition it will distribute the TFC commands from SODIN to FE with fixed latency.

SOL40-SCA core
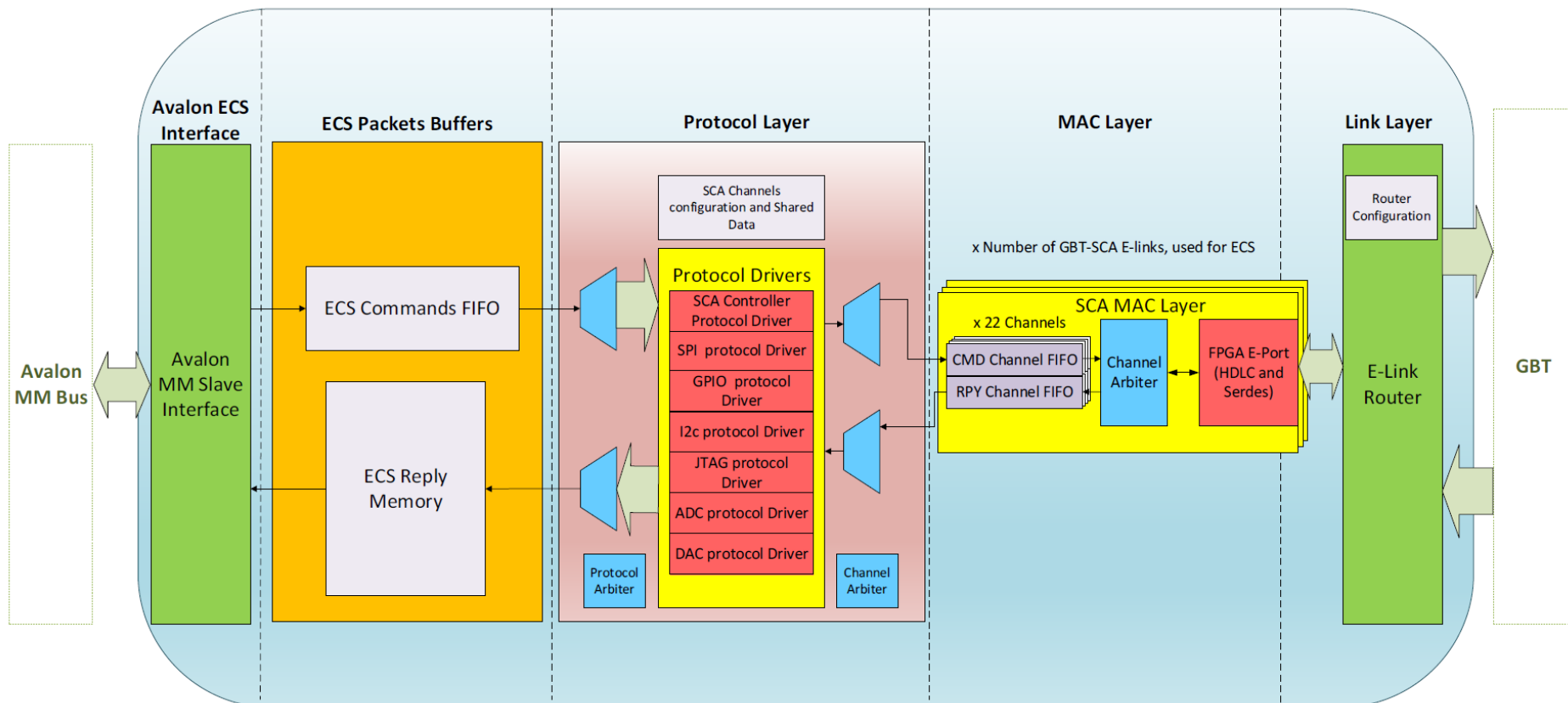
1 core per GBT, scalable support for SCAs, all SCA buses, *vendor/FPGA independent*, modular control, error detection capabilities
This is a CERN-wide development: i.e. the other experiments are going to use it.