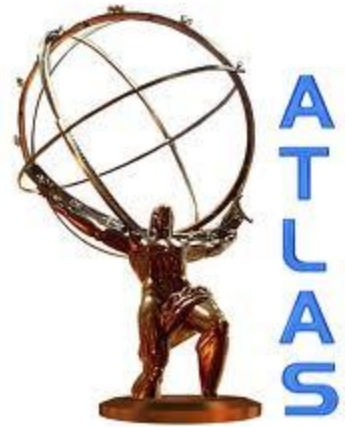




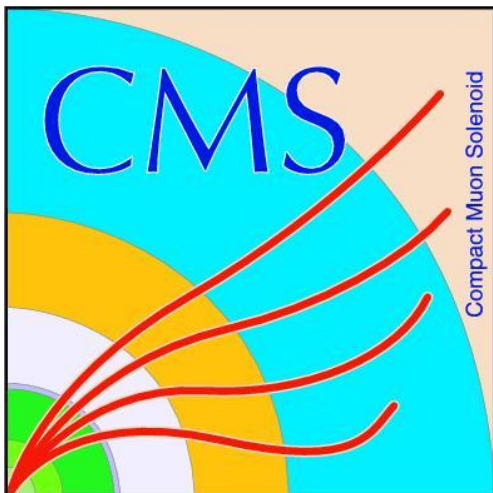
ALICE

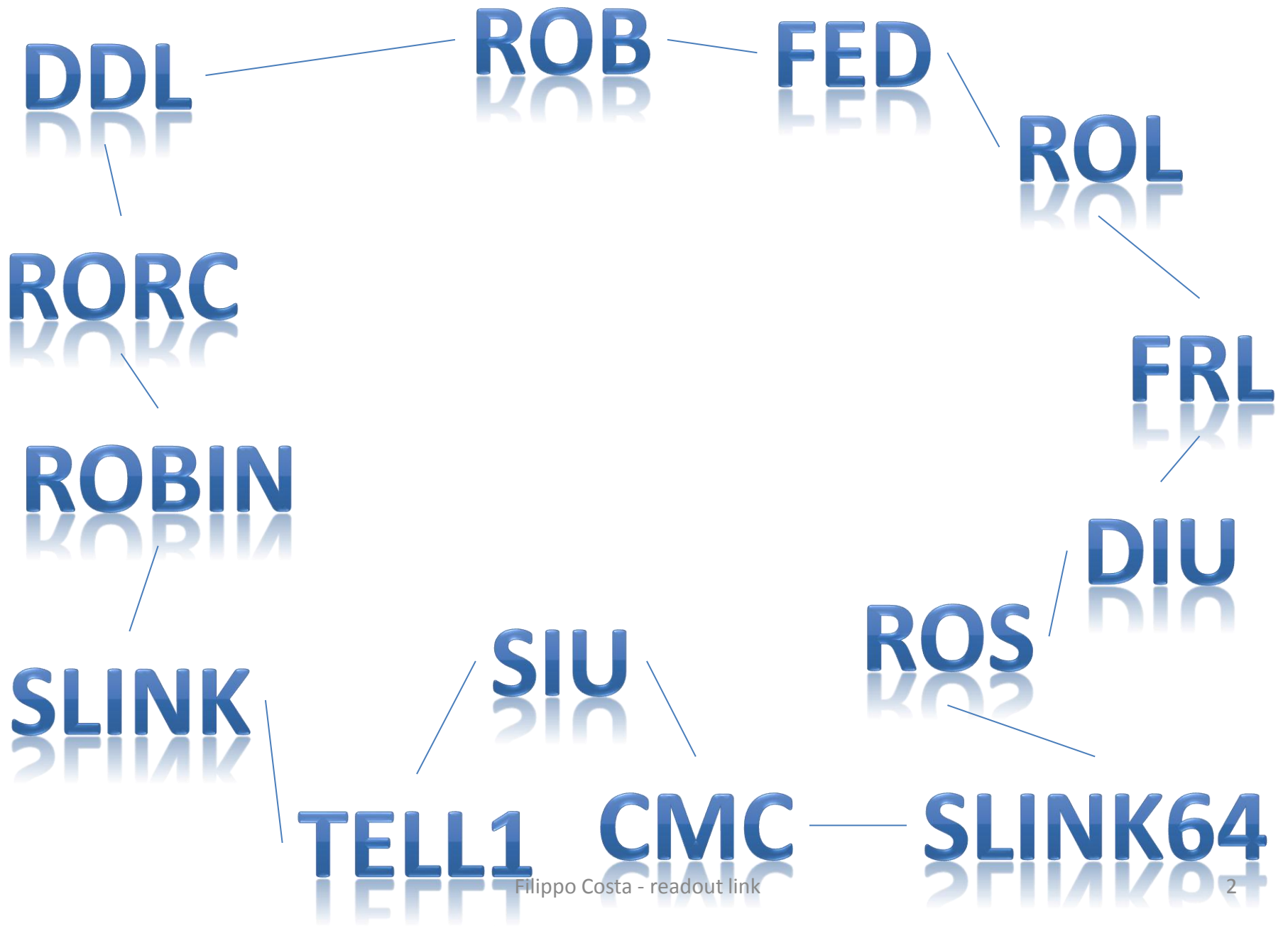
A JOURNEY OF DISCOVERY

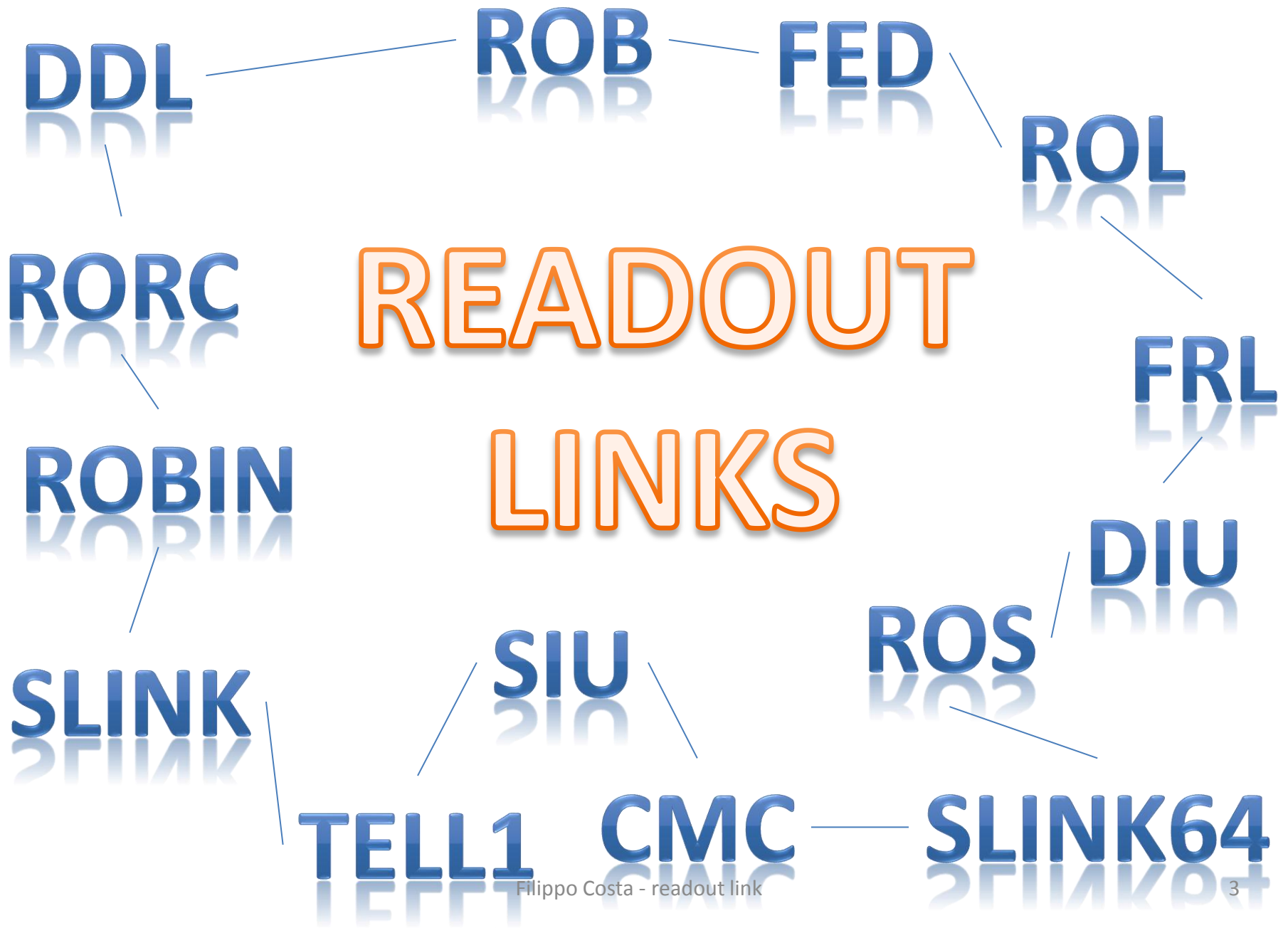


READOUT LINK

Filippo Costa



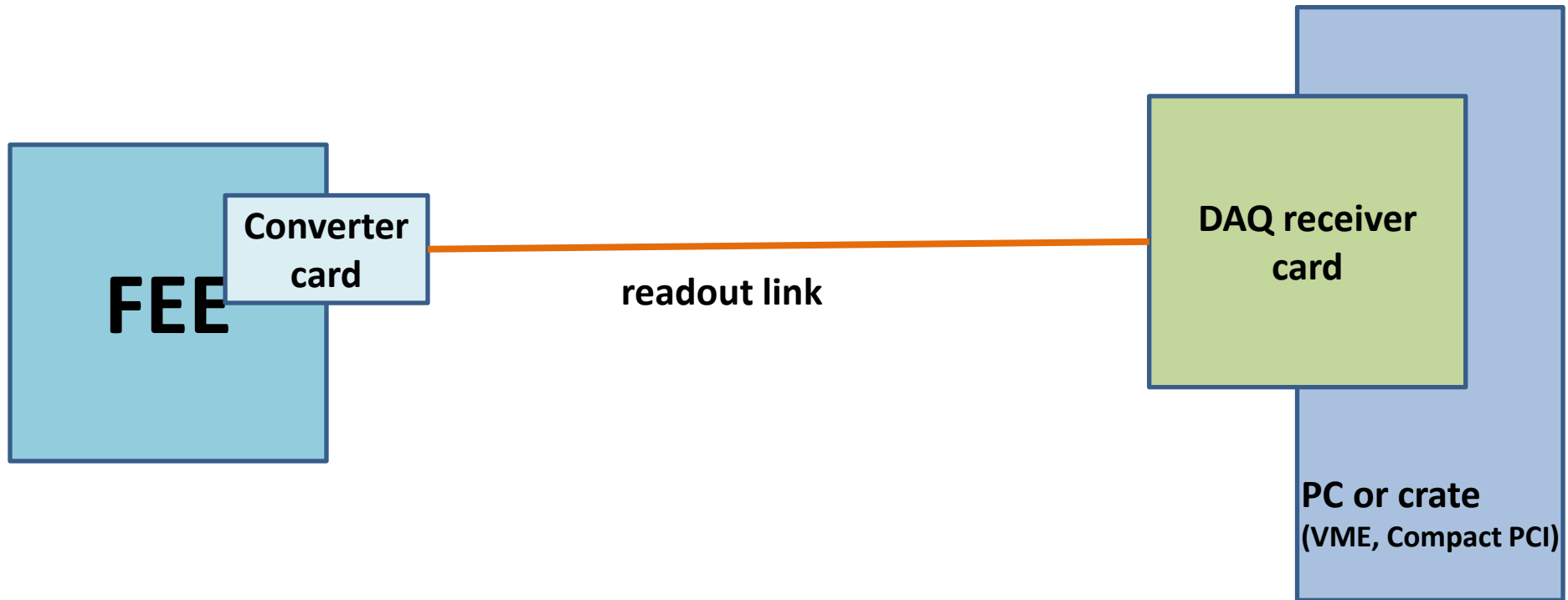


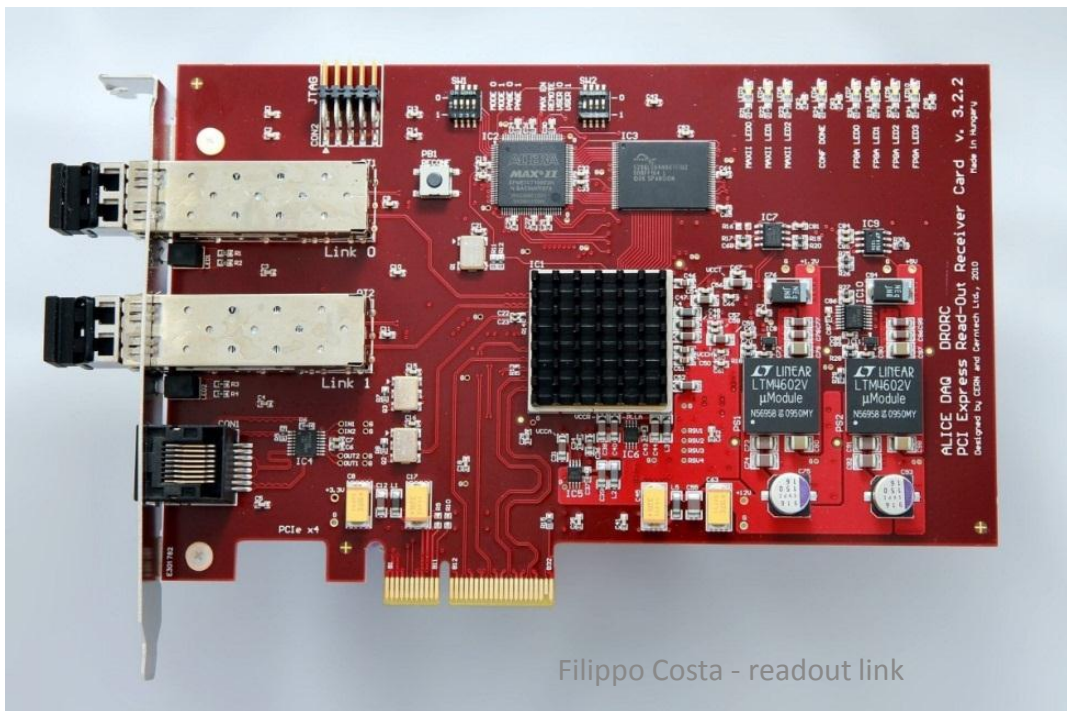


LEGEND

SIU	Source Interface Unit	CONVERTER CARD
CMC	Common Mezzanine Card	
HOLA	High-speed Optical Link for Atlas	
TELL1	Trigger ELectronics Level 1	
DDL	Detector Data Link	READOUT LINK
ROL	Readout Link	
SLINK64	Simple Link Interface	
GbE	Gigabit Ethernet	
RORC	Readout Receiver Card	DAQ RECEIVER CARD
ROBIN	Readout Buffer INput module	
FRL	Front end Readout Link	
NIC	Network Interface Card	

General picture

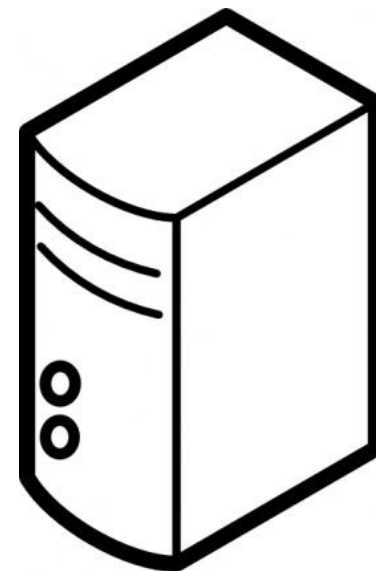


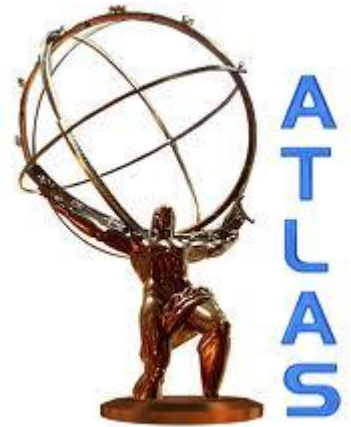


Filippo Costa - readout link

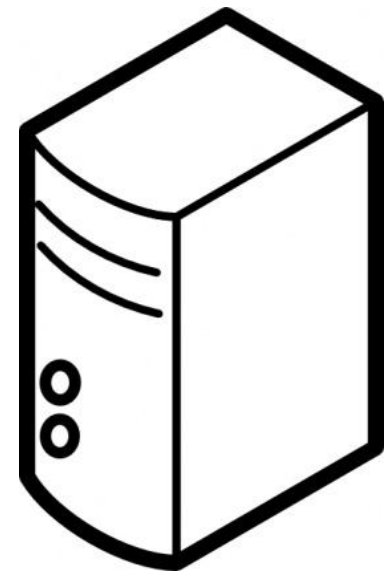


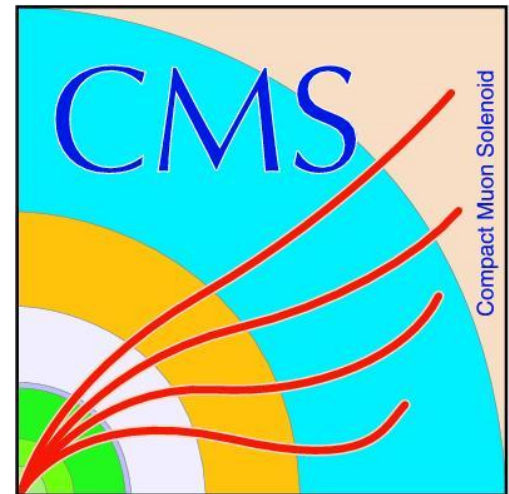
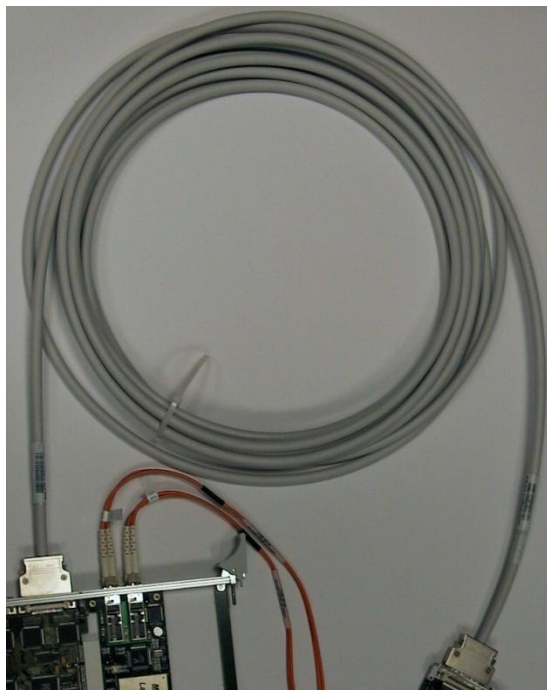
ALICE
A JOURNEY OF DISCOVERY



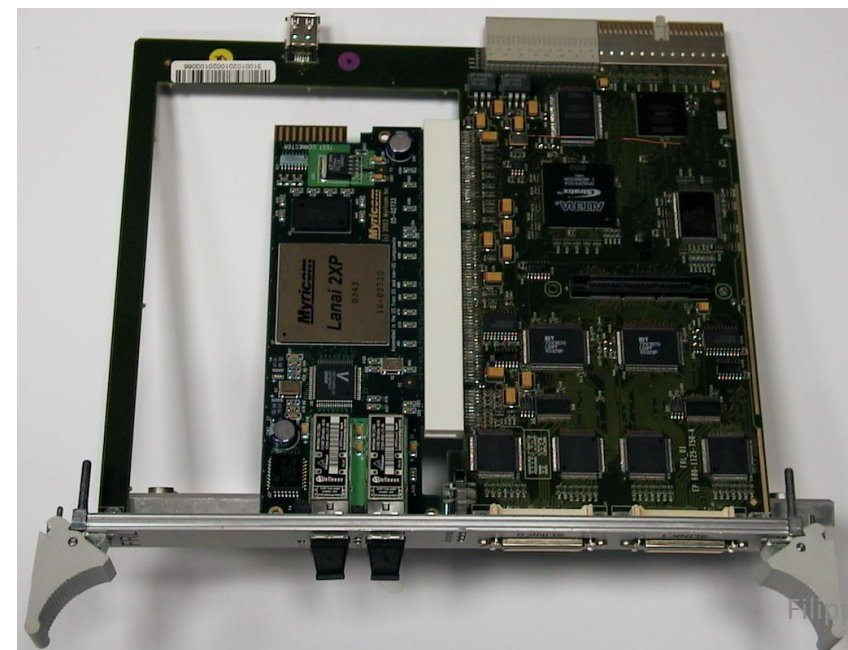


Filippo Costa - readout link



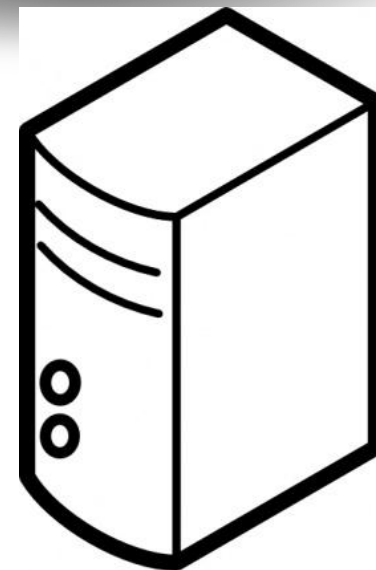
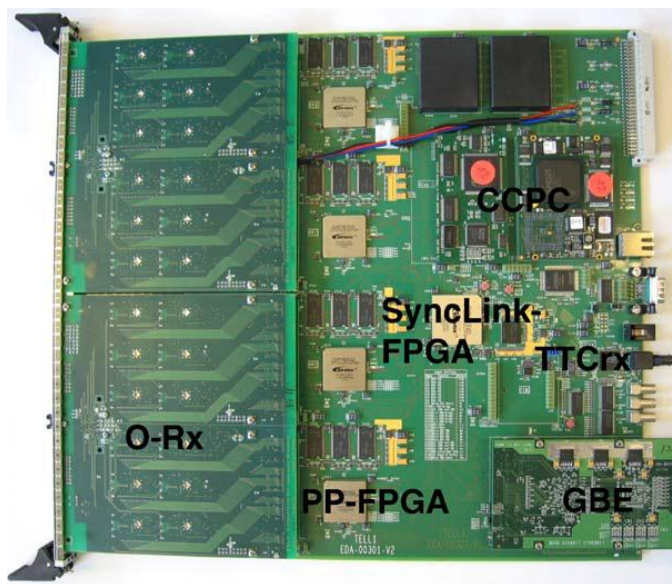


Compact Muon Solenoid



Filippo Costa - readout link





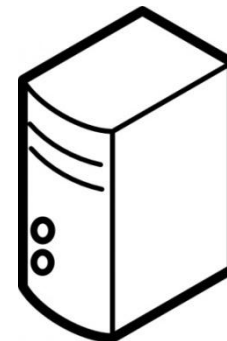
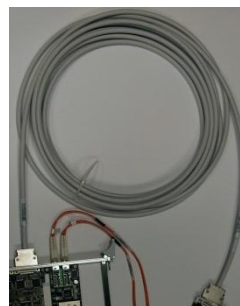
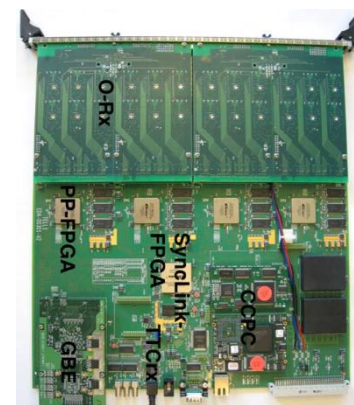
IT'S

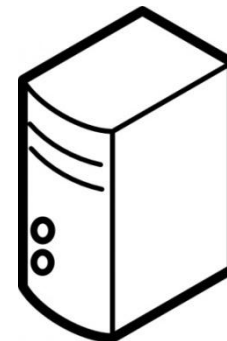
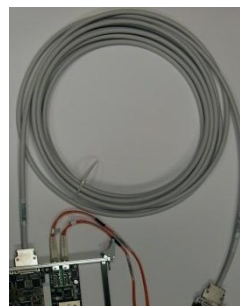
THE

SAME

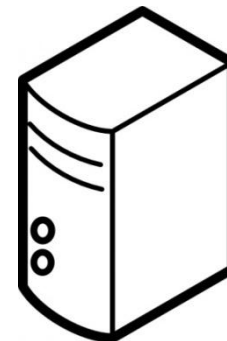
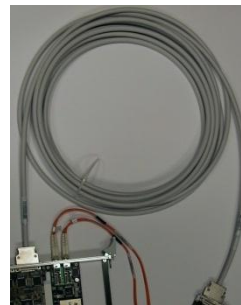
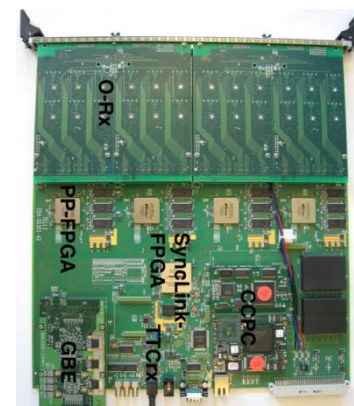
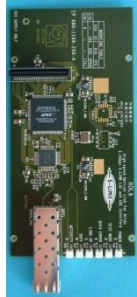
THING

(almost)



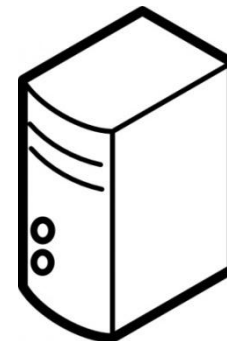
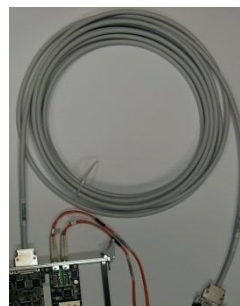
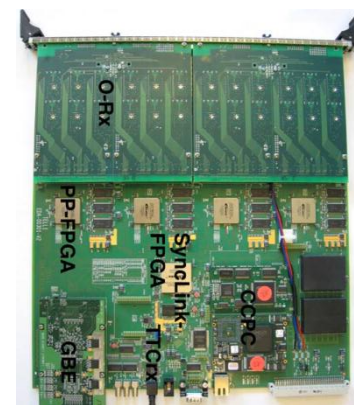
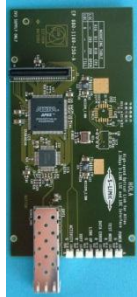


~2500
CHF
(~1300
per link)



~2500
CHF
(~1300
per link)

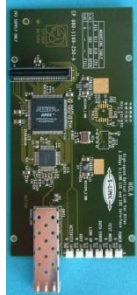
~3300
CHF
(~1100
per link)



~2500
CHF
(~1300
per link)

~3300
CHF
(~1100
per link)

1550
CHF
(~800
per link)

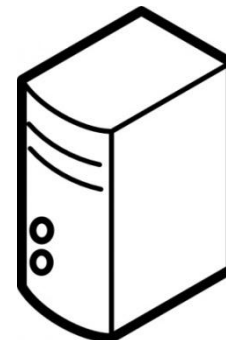
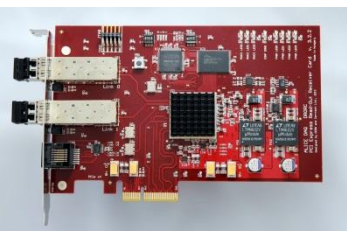


ALICE

ATLAS

CMS

LHCb



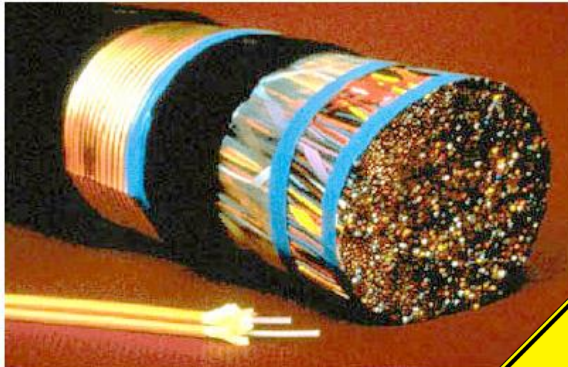
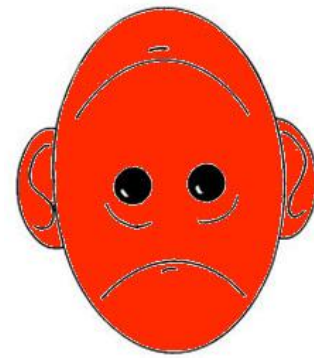
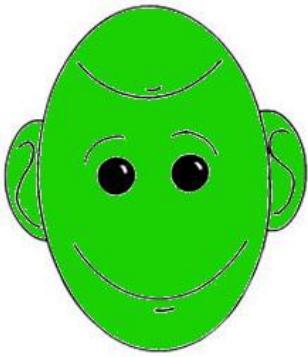
~2500
CHF
(~1300
per link)

~3300
CHF
(~1100
per link)

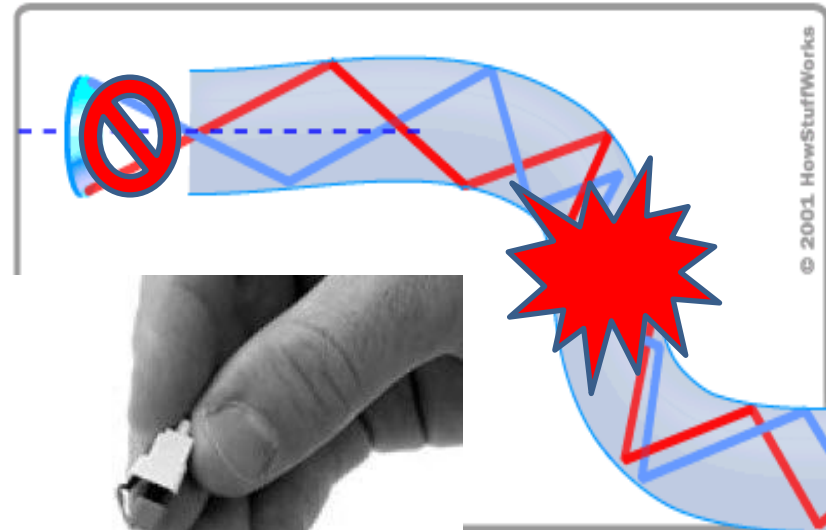
1550
CHF
(~800
per link)

> 6000
CHF
(~300
per link)

	# links	protocol	Throughput	Flow control	Main buffering
ATLAS	1600	Slink optical	160/200 MB/s	Yes	On board. 3 x 64 MB
ALICE	500	DDL optical	200 MB/s	Yes	On the PC memory 8 GB
CMS	600	Slink64 copper	400 MB/s	Yes	On board. 64 KB
LHCb	400	Ethernet	4 x 1Gb/s	NO	On board. 128 KB

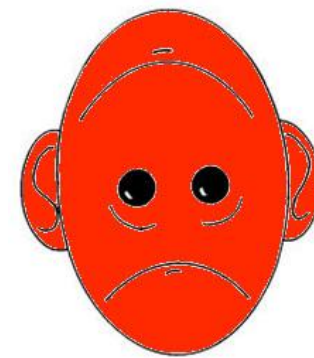
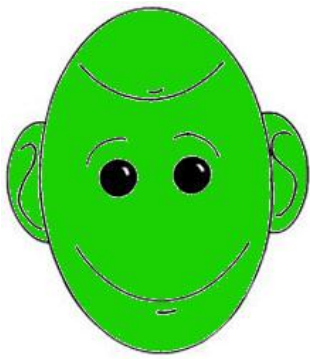


The optical fiber cable in the foreground has the equivalent capacity of the copper cable in the background.

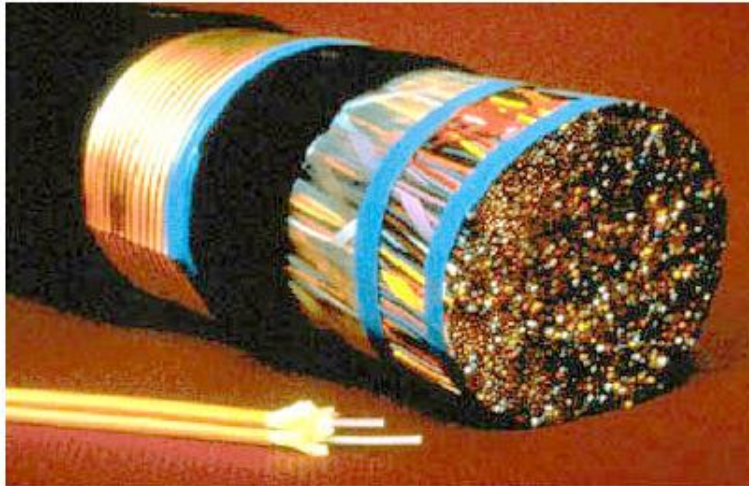


© 2001 HowStuffWorks

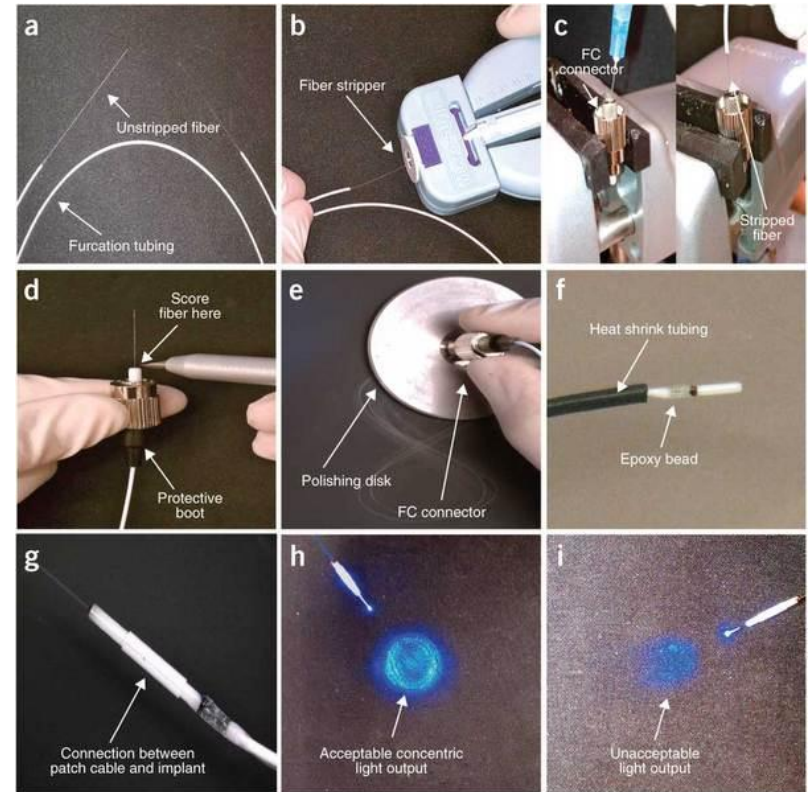
No general software tools to debug data flow.

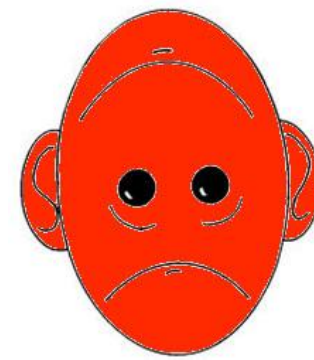
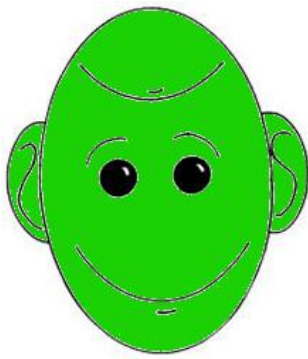


- S-Link is a that can be implemented in an FPGA (not resources demanding)



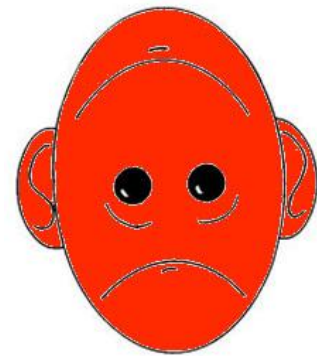
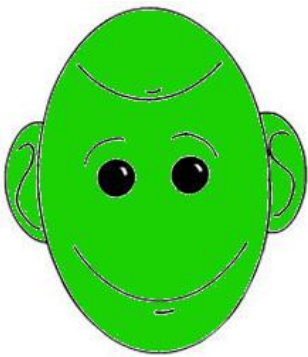
The optical fiber cable in the foreground has the equivalent capacity of the copper cable in the background.





- Very robust link.
- Cost is pretty low.





(Untitled) - Wireshark

File Edit View Go Capture Analyze Statistics Help

Filter: Expression... Effacer Appliquer

No.	Time	Source	Destination	Protocol	Info
5	12.117859	192.168.254.128	192.168.254.2	DNS	Standard query A kernel.org
6	12.176017	192.168.254.2	192.168.254.128	DNS	Standard query response A 204.152.191.5 A 204.152.191.37
7	12.177666	192.168.254.128	kernel.org	TCP	proofd > wwv [SYN] Seq=0 Len=0 MSS=1460
8	12.347220	kernel.org	192.168.254.128	TCP	wwv > proofd [SYN, ACK] Seq=0 Ack=1 Win=64240 Len=0 MSS=1460
9	12.347469	192.168.254.128	kernel.org	TCP	proofd > wwv [ACK] Seq=1 Ack=1 Win=64240 Len=0
10	12.347993	192.168.254.128	kernel.org	HTTP	GET / HTTP/1.1\r\n
11	12.348111	kernel.org	192.168.254.128	TCP	wwv > proofd [ACK] Seq=1 Ack=415 Win=64240 Len=0
12	12.518267	kernel.org	192.168.254.128	TCP	[TCP segment of a reassembled PDU]
13	12.519468	kernel.org	192.168.254.128	TCP	[TCP segment of a reassembled PDU]
14	12.519622	192.168.254.128	kernel.org	TCP	proofd > wwv [ACK] Seq=415 Ack=2897 Win=64240 Len=0
15	12.618072	192.168.254.128	kernel.org	TCP	proofd > wwv [SYN] Seq=0 Len=0 MSS=1460
16	12.693065	kernel.org	192.168.254.128	TCP	[TCP segment of a reassembled PDU]
17	12.697060	kernel.org	192.168.254.128	TCP	[TCP segment of a reassembled PDU]
18	12.697983	kernel.org	192.168.254.128	TCP	[TCP segment of a reassembled PDU]
19	12.697101	kernel.org	192.168.254.128	TCP	[TCP segment of a reassembled PDU]
20	12.697336	192.168.254.128	kernel.org	TCP	proofd > wwv [ACK] Seq=415 Ack=8689 Win=64240 Len=0

Frame 10 (468 bytes on wire, 468 bytes captured)

Ethernet II, Src: Vmware_f8:3a:35 (00:0c:29:f8:3a:35), Dst: 192.168.254.2 (00:50:56:f6:1a:eb)

Internet Protocol, Src: 192.168.254.128 (192.168.254.128), Dst: kernel.org (204.152.191.5)

Transmission Control Protocol, Src Port: proofd (1093), Dst Port: wwv (80), Seq: 1, Ack: 1, Len: 414

Source port: proofd (1093)

Destination port: wwv (80)

Sequence number: 1 (relative sequence number)

Next sequence number: 415 (relative sequence number)

Acknowledgement number: 1 (relative ack number)

Header length: 20 bytes

Flags: 0x18 (PSH, ACK)

Window size: 64240

Checksum: 0x1826 [correct]

Hypertext Transfer Protocol

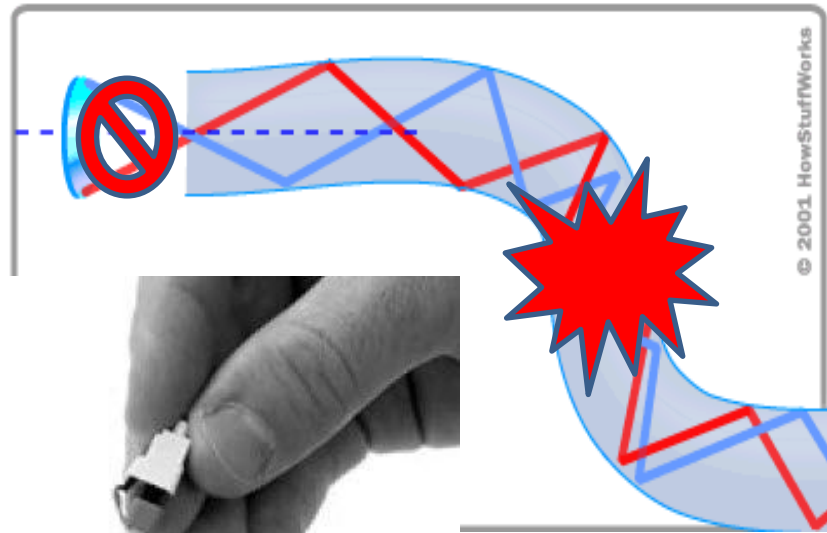
GET / HTTP/1.1\r\n

0020 bf 05 04 45 00 50 13 12 3e 91 02 e2 cd 6c 50 18 ...E.F. >...LP.

0030 fa f0 18 26 00 00 47 45 54 20 2f 20 48 54 54 50 ...&.GE T / HTTP

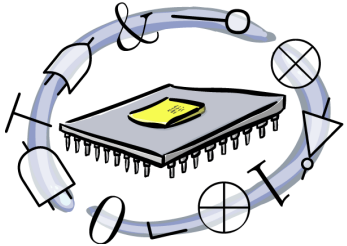
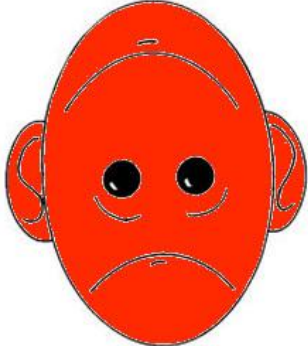
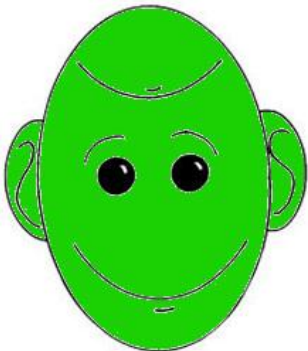
0040 2f 31 2e 31 0d 0a 48 6f 73 74 3a 20 6b 65 72 6e /1.1.Ho st: kern

Sequence number (tcp.seq), 4 bytes | P: 107 D: 107 M: 0 Drops: 0



GOL

IN GENERAL





The Future

RUN 2 ↗



We are busy updating the experiment for you and will be back shortly

ALICE

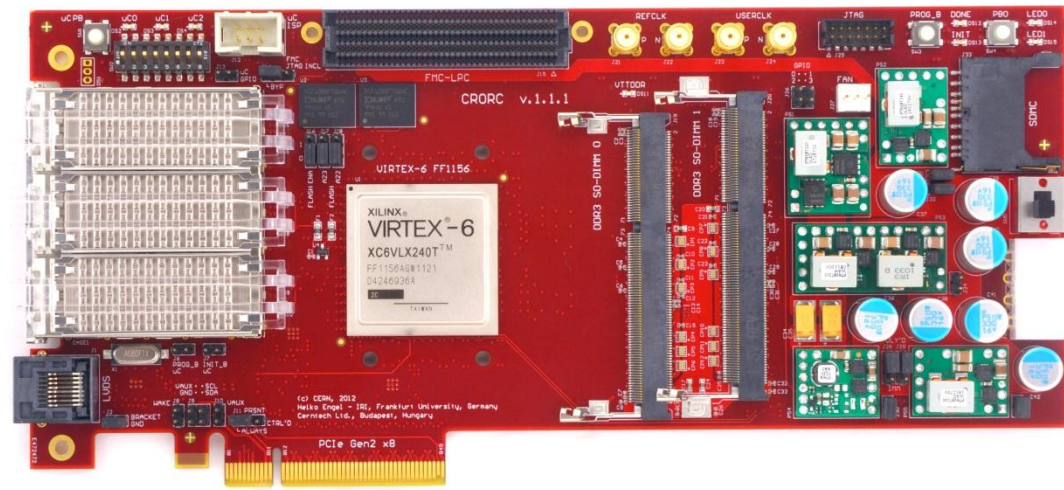
A JOURNEY OF DISCOVERY

- Faster readout and higher data throughput for TPC and TRD.
- Higher data throughput between HLT to DAQ.



ALICE

A JOURNEY OF DISCOVERY



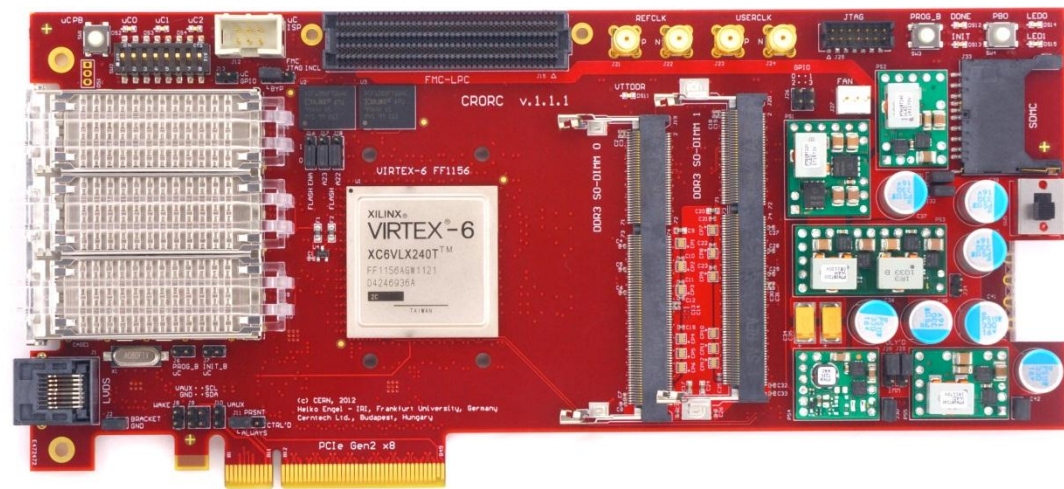
- **PROTOCOL UPGRADE:**
 - faster DDL protocol from 2 Gb/s to 6 Gb/s.
- **HARDWARE UPGRADE:**
 - **INPUT:** from 2 ch @ 2Gb/s to 12 ch @ 6 Gb/s.
 - **OUTPUT:** from PCIe gen.1x4 to PCIe gen.2x8.
- Current RORC firmware implemented on new hardware.
- Same software framework, only driver upgrade to handle the new hardware.



We are busy updating the experiment for you and will be back shortly



- 64-bits PCI technology , obsolete and causes problems for upgrading DAQ PCs.
- Number of readout links will increase.
- Reduce amount of RACK space.
- Increase data throughput from ROS to HLT after LS1.



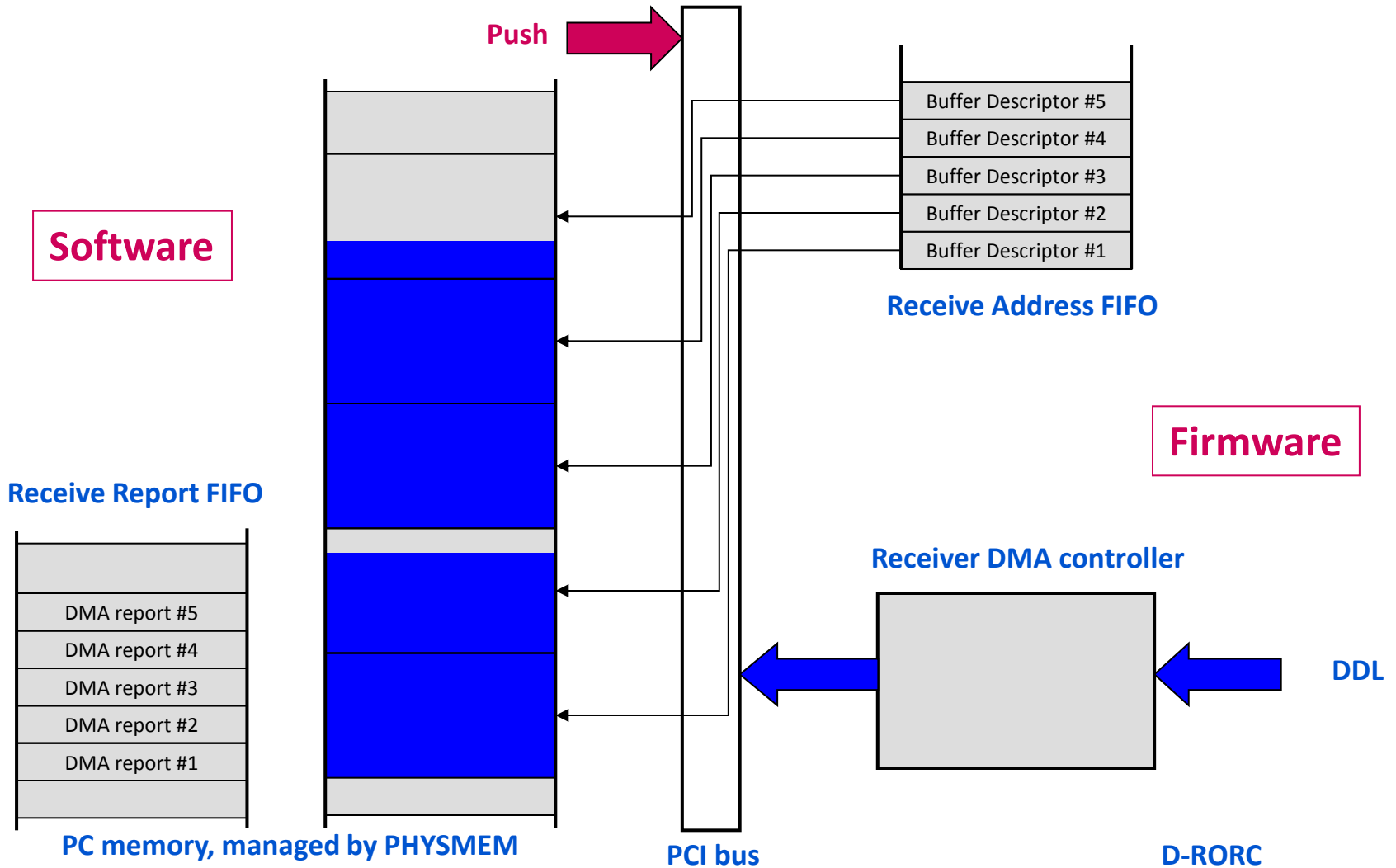
- Use of the ALICE CRORC, but is called **ROBINNP**.
- **HARDWARE UPGRADE:**
 - **INPUT:** from 3 ch @200 MB/s to 12 ch @ 200 MB/s (6Gb/s)
 - **OUTPUT:** from PCI 64 bit 66MHz to PCIe gen.2x8.
 - Bigger buffering capacity (up to 4 GB of DDR2/3).
- New firmware (the PPC present in the ROBIN will be moved out and implemented in software).

Same hardware – different features

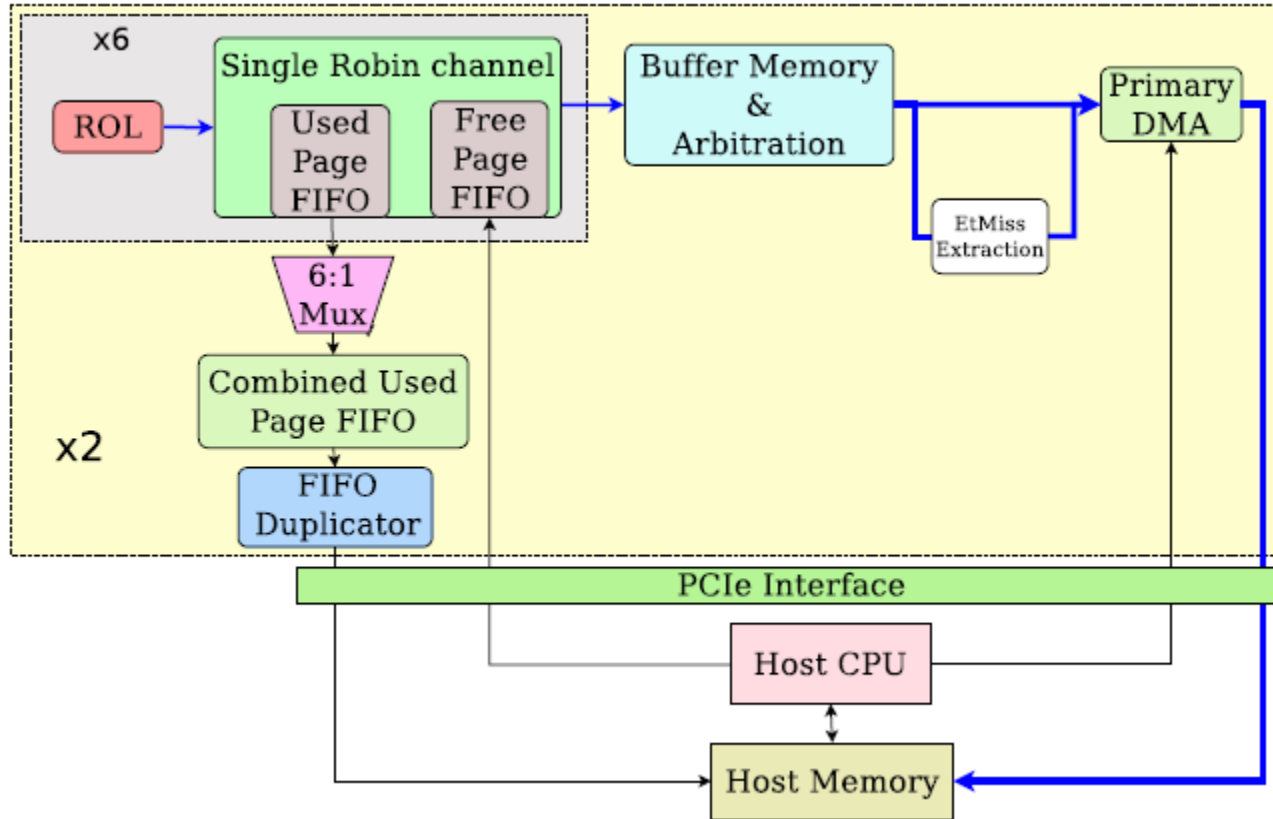
ALICE	ATLAS
12 ch @ 6 Gb/s	12 ch @ 2 Gb/s (in average, if needed 6Gb/s)
No use of internal memory	Up to 4 GB of DDR2/3 used to buffer events in the board



ALICE CRORC FIRMWARE



ATLAS ROBINNP firmware





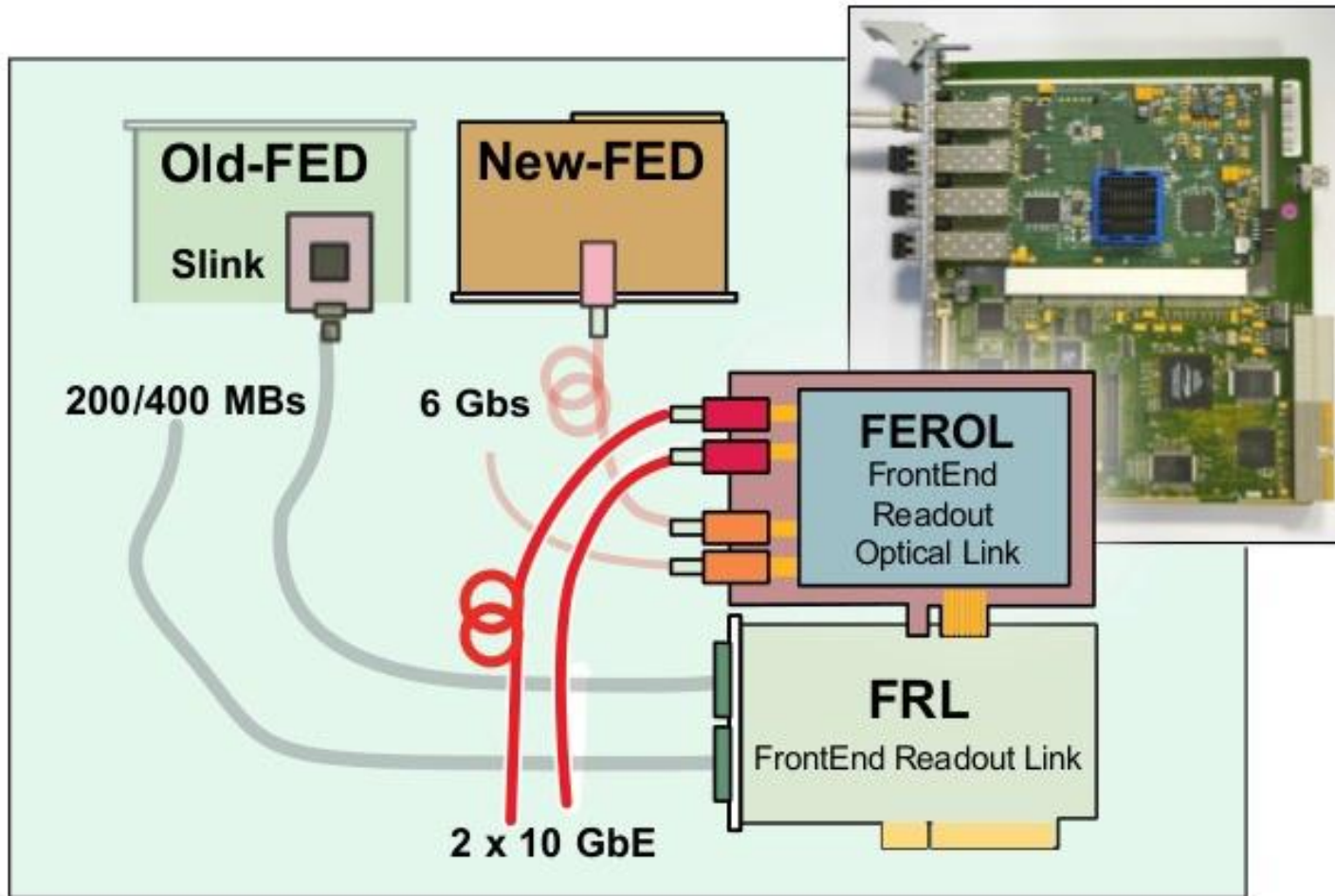
We'll be
back
soon.

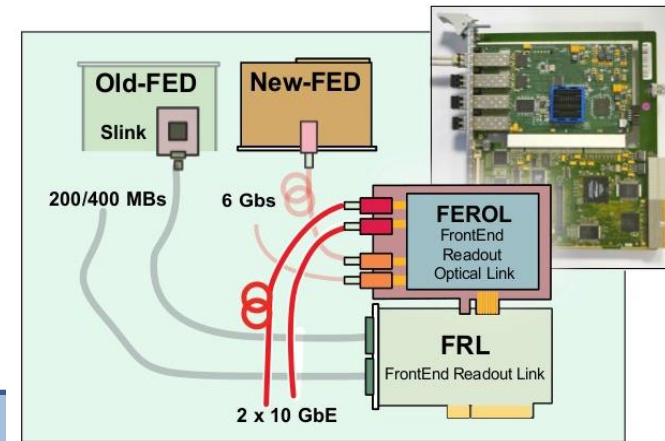
We are busy updating the experiment for you and will be back shortly

- Hardware is aging (Myrinet) ... needs for new components and replacements.
- New components are not compatible with the current hardware.
- Faster readout and higher data throughput, needed by trigger and HCAL.

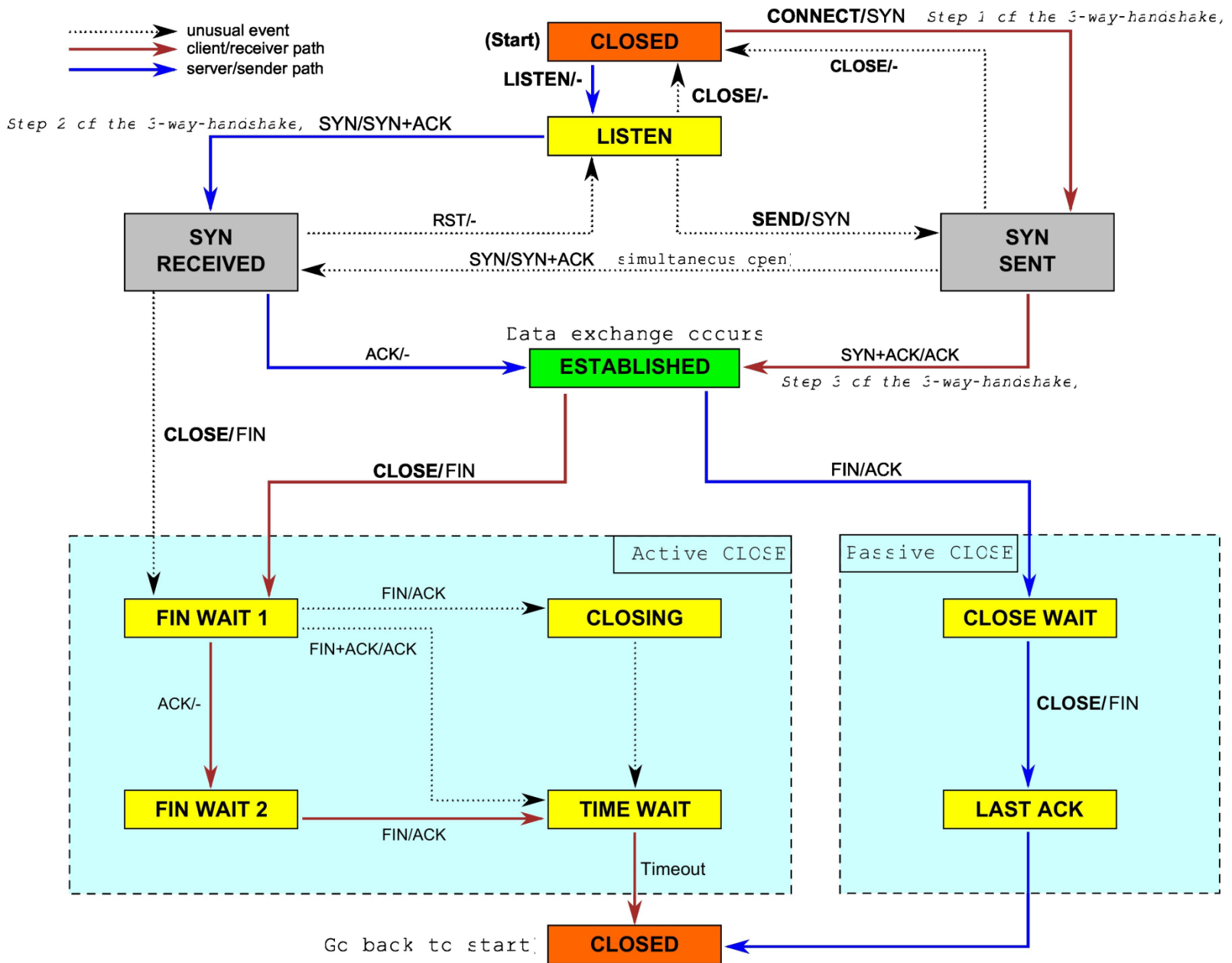
CMS

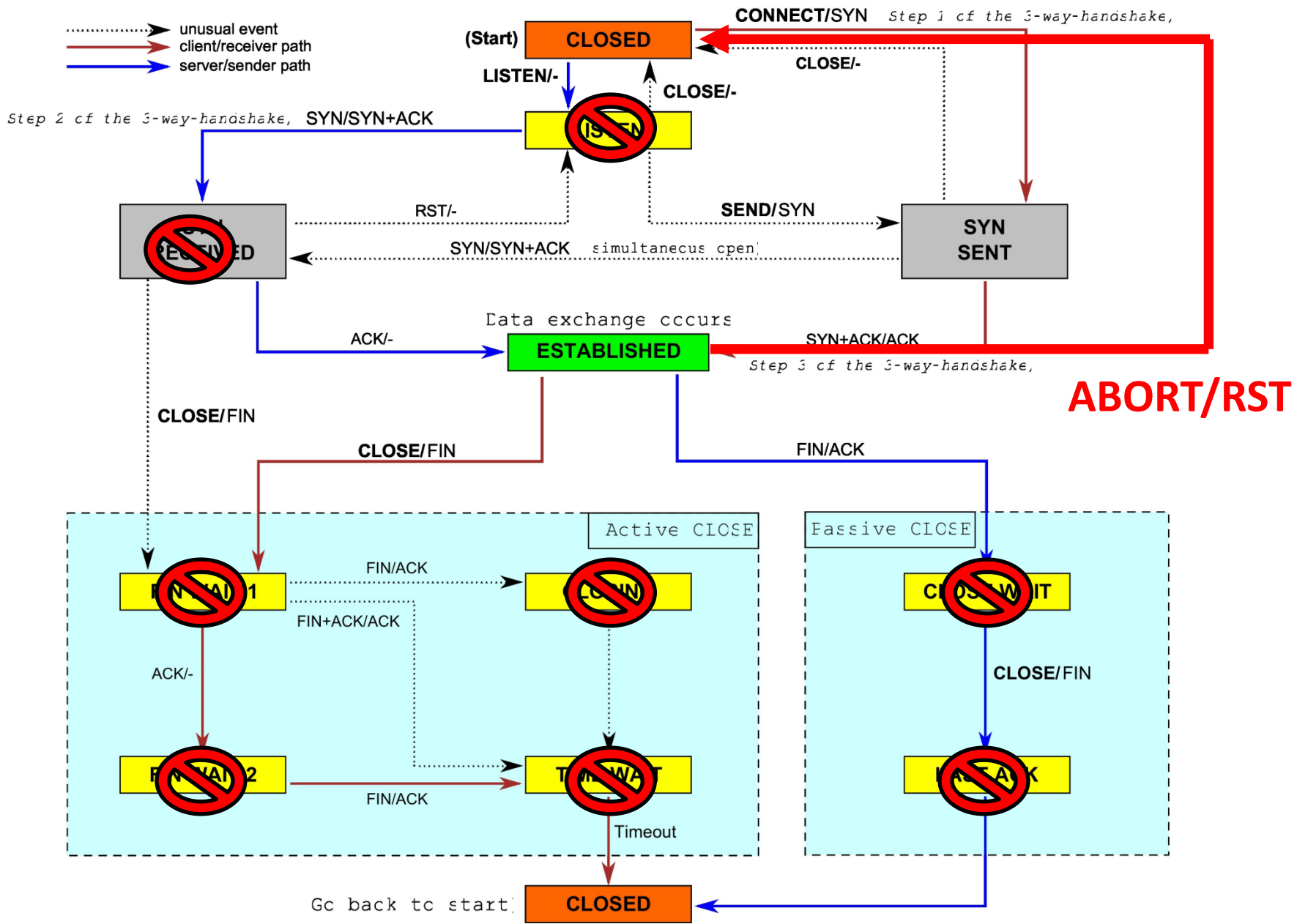
Compact Muon Solenoid



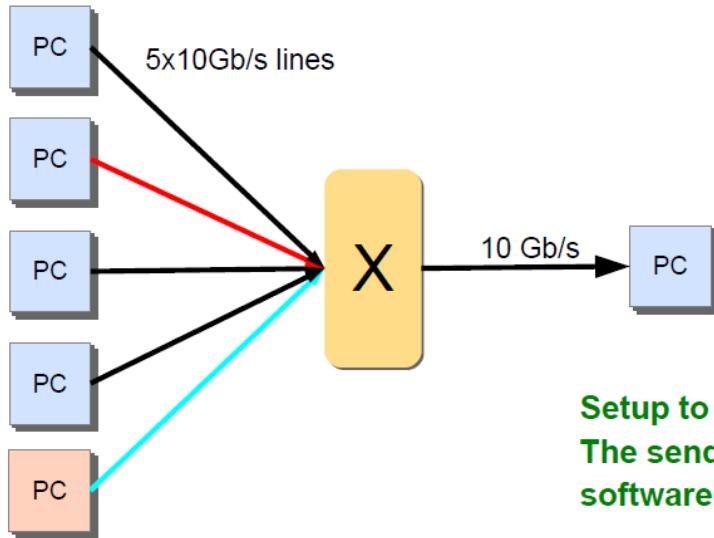


- The existing DAQ receiver card (FRL) stays in place to support legacy FEDs.
- New FEROL card has 2 x 10 Gb/s and 2 x 6 Gb/s channels.
- No CMC , but IPCORE, same interface to the FEE.
- The new 6 Gb/s link has re-transmit feature.
- 500 MB internal buffer.
- 2 input links are treated independently and can share the same 10 Gb/s.
- The second link 10 Gb/s will be used to receive 10 Gb/s in future (1 Input 1 Output)
- 10 Gb/s reduced TCP/IP implemented in FPGA.





CONGESTION CONTROL



Senders:
 125 kHz, 2048 bytes
 2.048 Gb/s

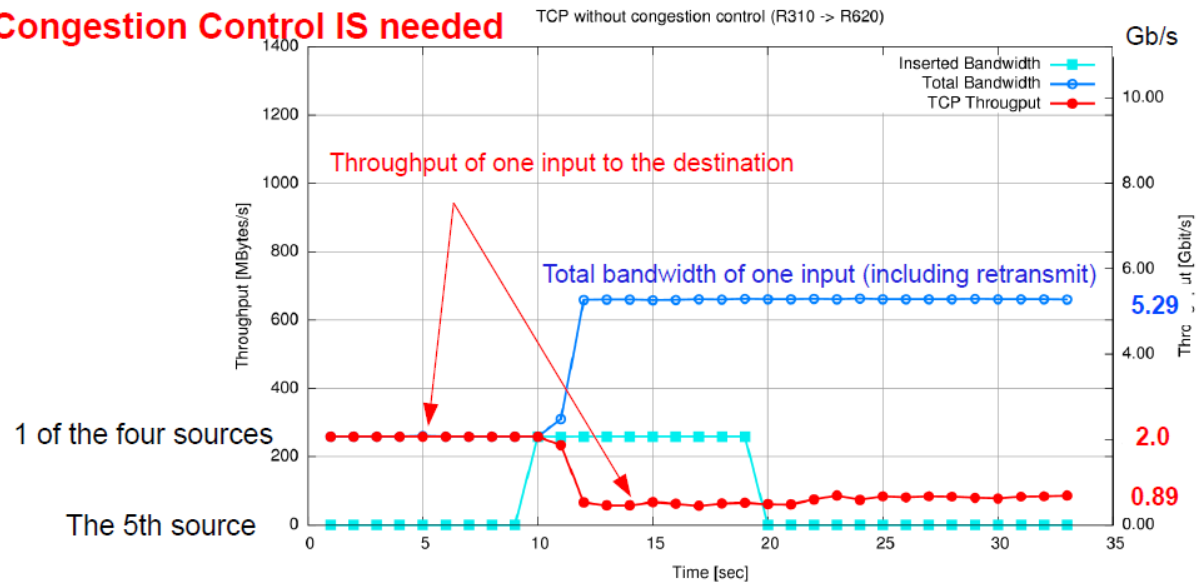
$5 \times 2,058 = 10.24 \text{ Gb/s}$

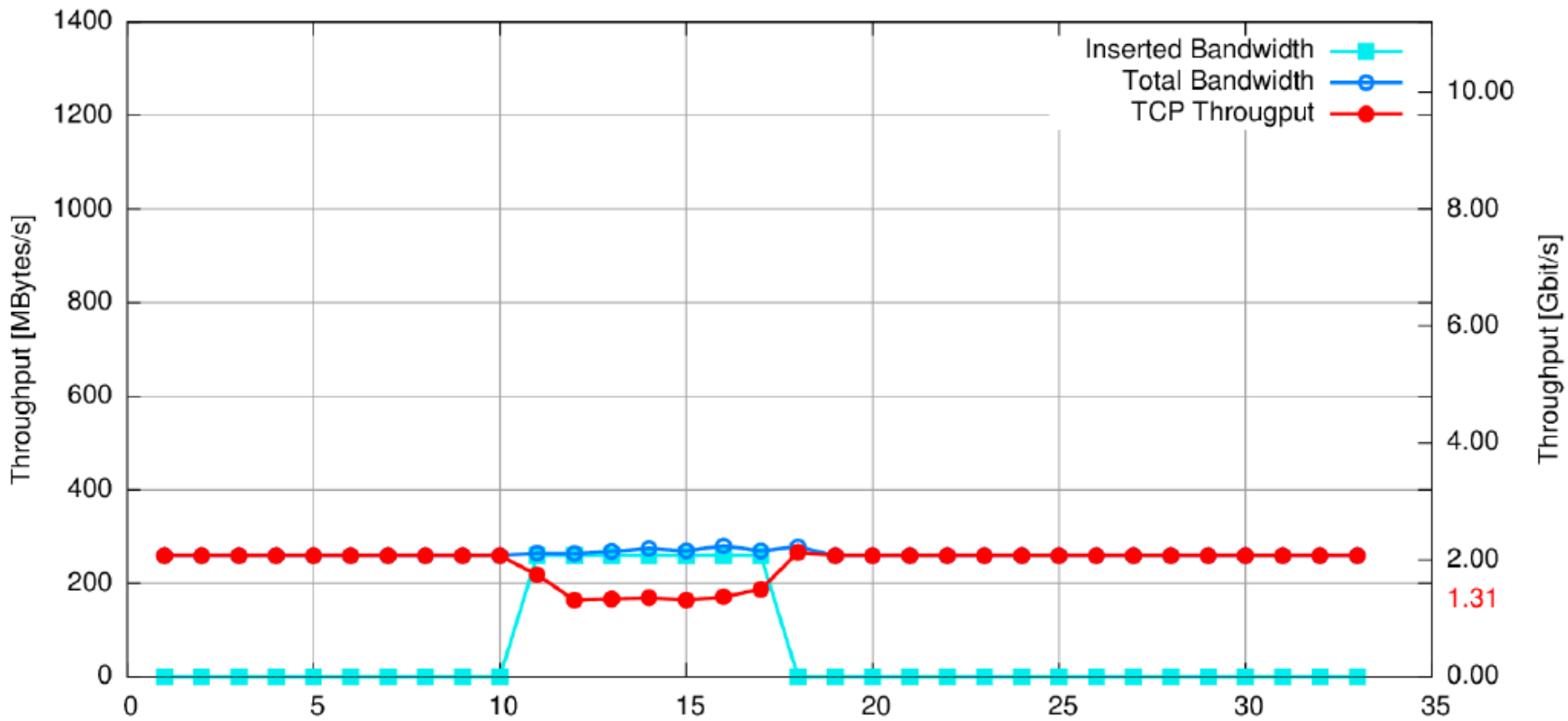
A little bit of congestion

Setup to generate congestions
The senders emulate the protocol with software

- **The system does not recover**
 - The available bandwidth is eaten up by a lot of re-transmits. Due to the re-transmits congestion becomes worse and worse and throughput becomes very small.

Congestion Control IS needed







The don't have special plan for the upgrade during LS1 ...
however ...



Look (again) what LHCb has done



They have achieved goals that other experiments are planning in the future:

- Common readout unit (TELL1) used by all the detector groups.
- Use of Ethernet protocol.



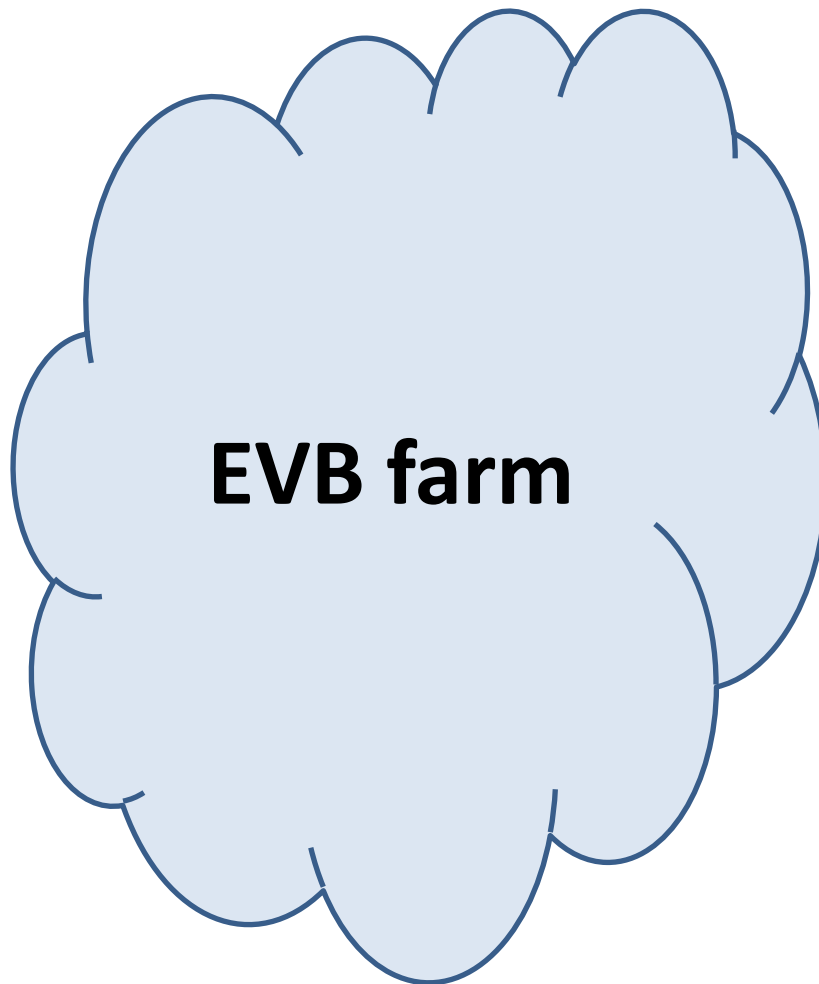
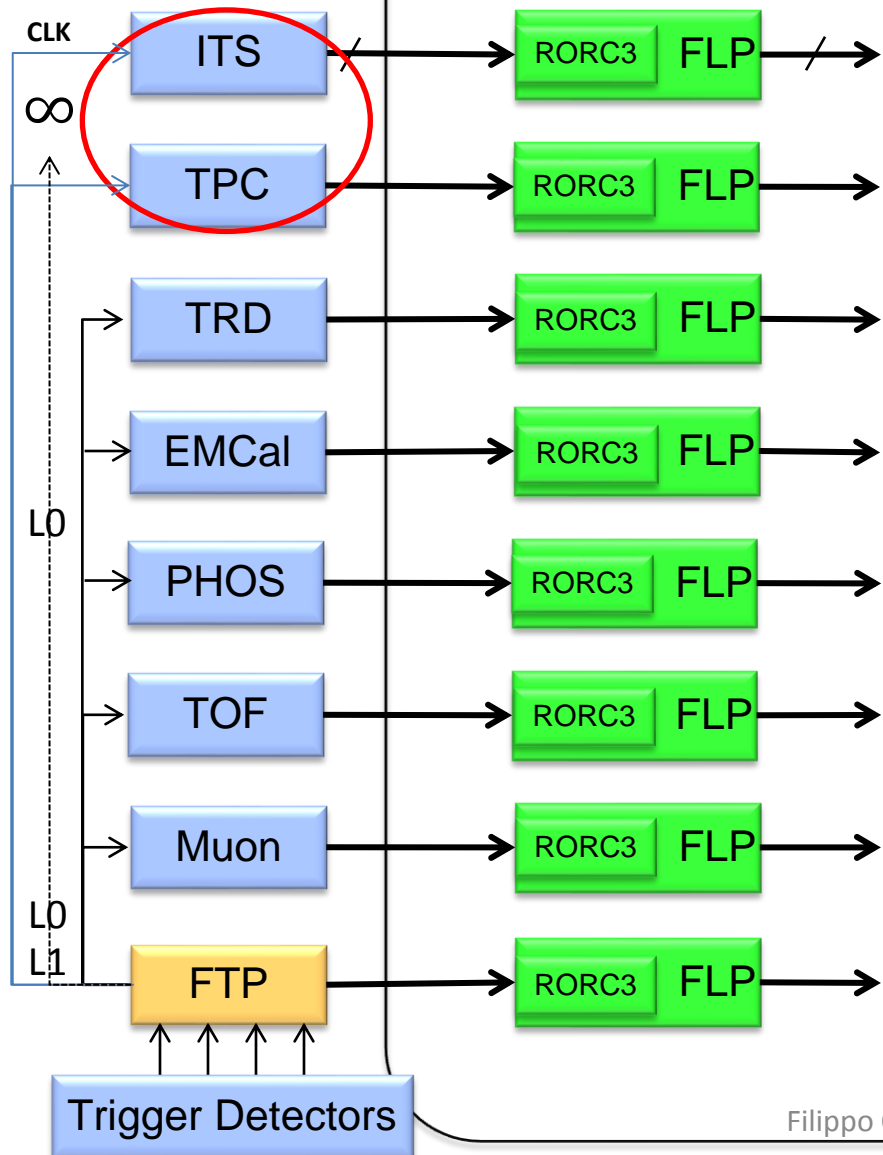
The Future

RUN 3 ↗

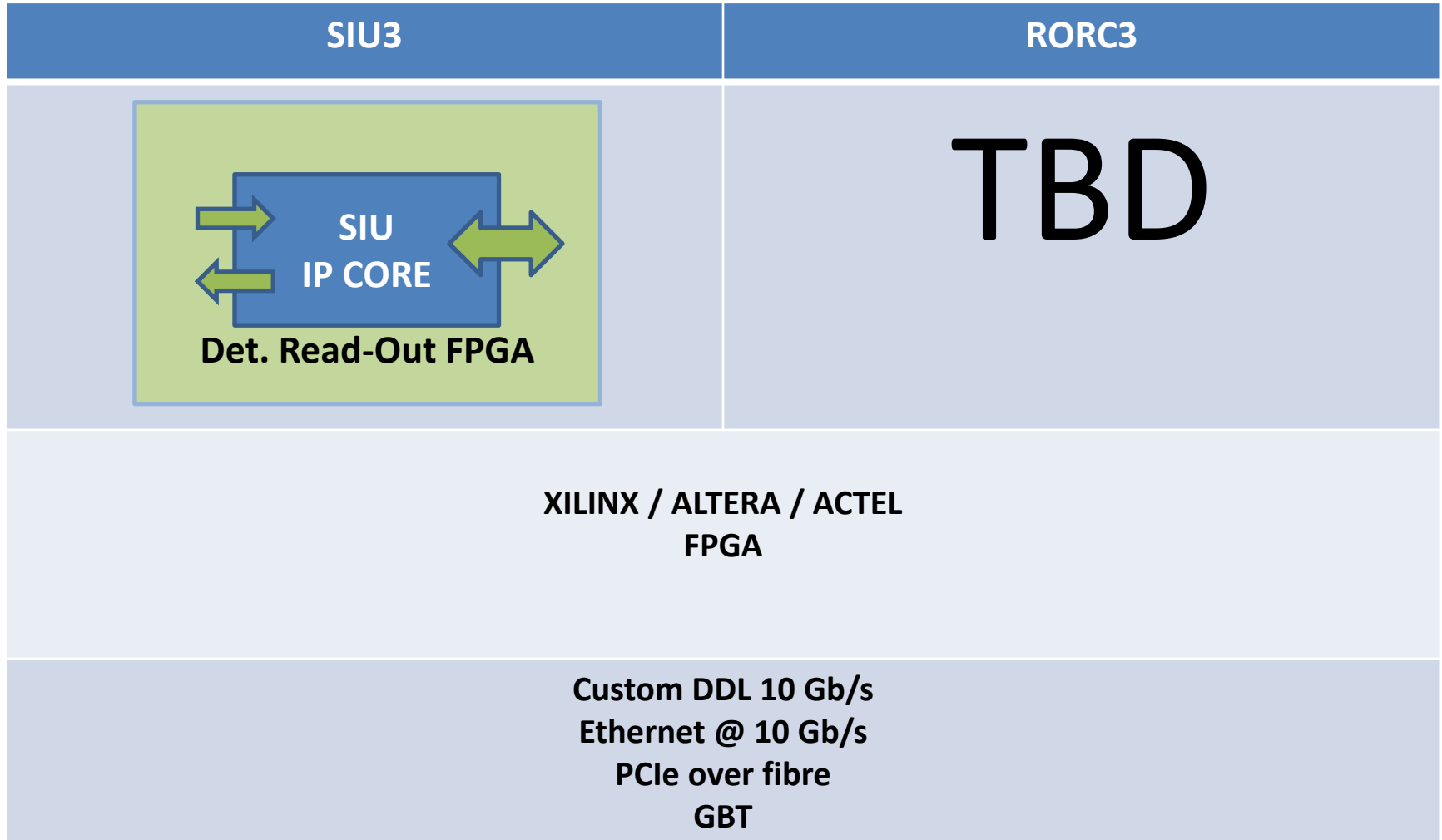
LS2 Online Upgrade

DAQ and HLT

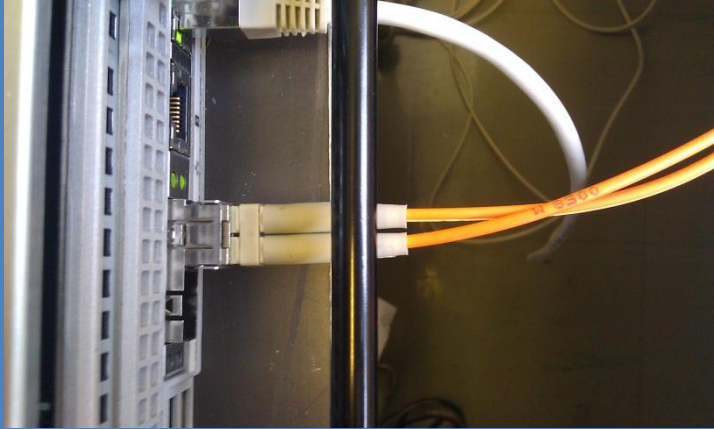
~ 2500 DDL3s 10 Gb/s (~1 TB/s in input)



ALICE readout link



DDL 3 (UDP Ethernet)



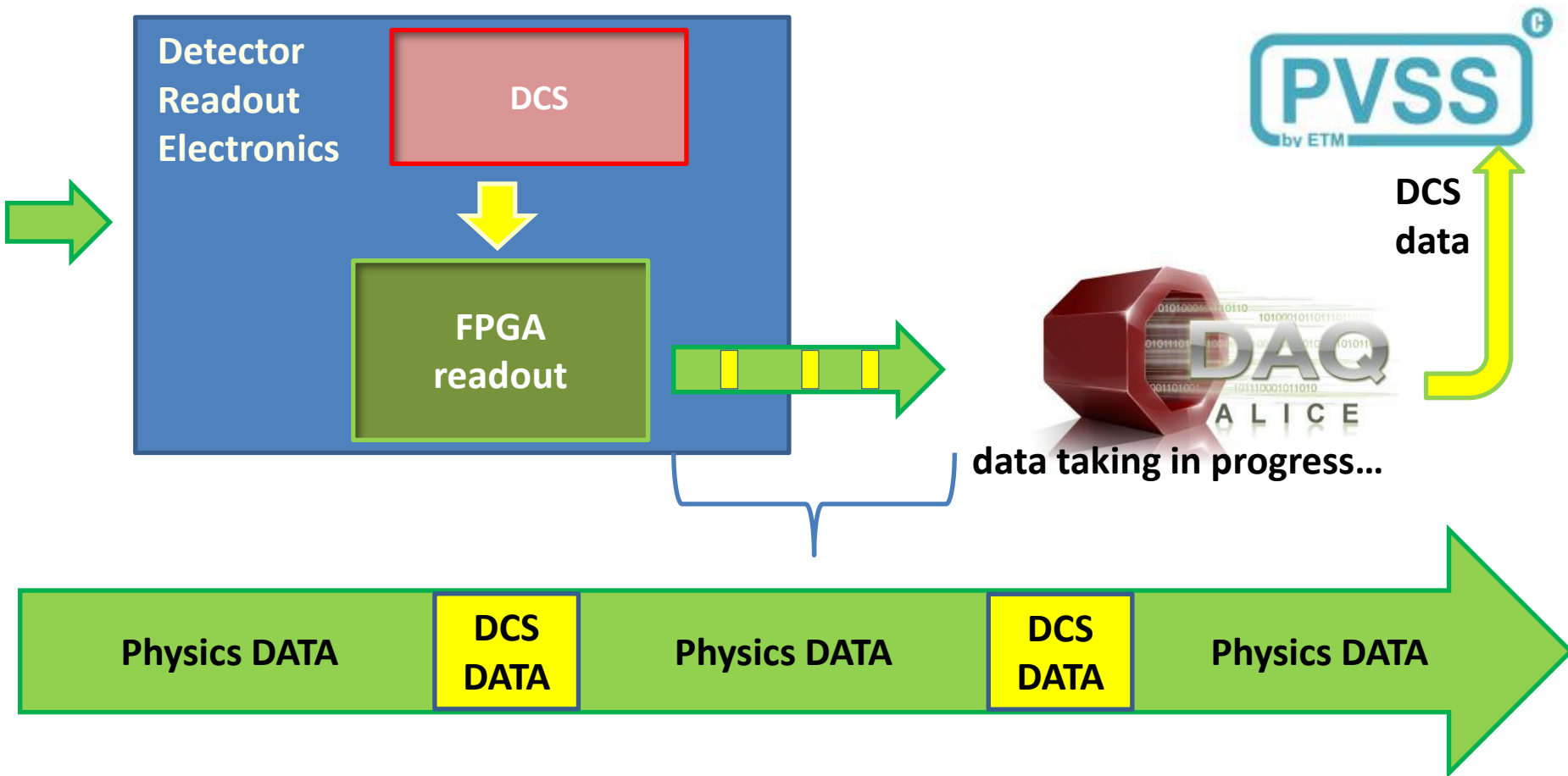
10 Gb/s embedded in the motherboard

SFP+ 10 Gb/s

DDL optical fibre

XILINX VIRTEX 6

DAQ+DCS proposal of sharing DDL3



Our proposal is to share the same physical link between the 2 data traffics:
99% PHYSICS data
1% or less, DCS monitoring data



Detector link	Readout electronic
GBT	TELL40
10 GbE (production and early tests) 40 GbE, Infiniband or PCIe Reduced TCP/IP or advanced UDP	

CONGRATULATIONS TO ALL OF US !!!!



WHO'S AWESOME?

You're awesome.

- ALICE, ATLAS, CMS and LHCb have different physics goals, but from the readout link point of view they try to achieve the same thing, move data from A to B. Why things are so different? Requirements and data throughput are not the same, but features could be adapted, instead writing them from scratch.
- Common effort to standardize the link and the electronics, ultimate goal:
 - common readout unit for detector groups,
 - use of commercial hardware and industrial-standard protocols.
- How to increase the knowledge sharing between all the experiments groups? It is always a joy to share good experience with a protocol or a link, but maybe we could learn more from things that went wrong (and usually are not showed at conferences)



Philippe Farthouat

Dominique Gigi

Benedetto Gorini

Markus Joos

Niko Neufeld

Christoph Schwick

Petr Zejdl