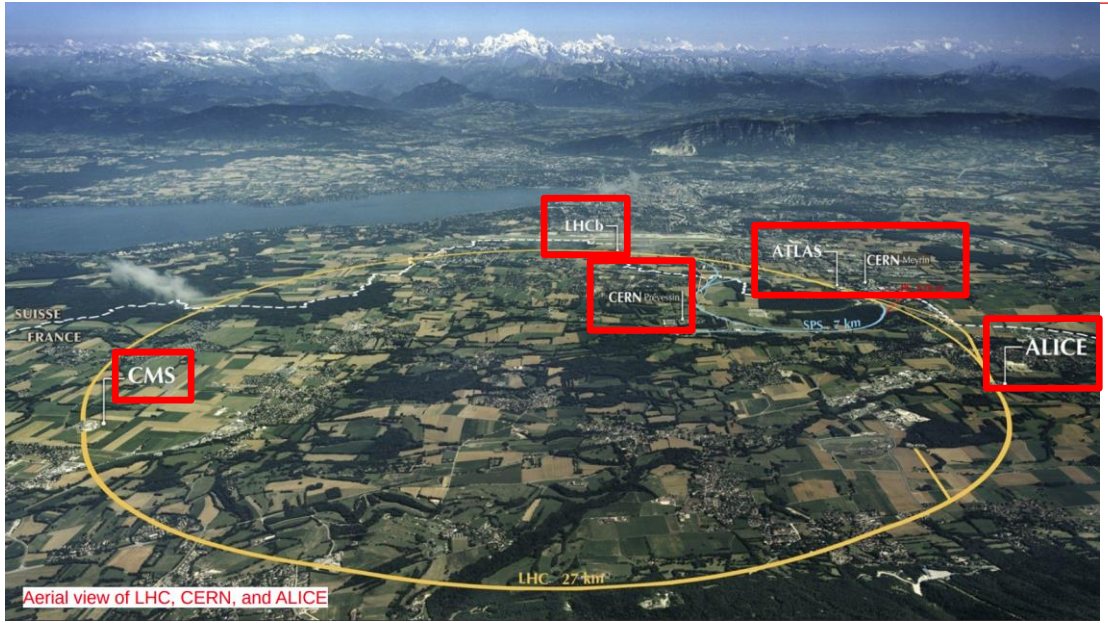
A Large Ion Collider Experiment

ALICE TRIGGER-DAQ

4th World Summit conference (EDSU2022)

Filippo Costa
for the ALICE collaboration



Aerial view of LHC, CERN, and ALICE

The LHC (Large Hadron Collider)
It consists of a 27-kilometre ring of superconducting magnets with a number of accelerating structures to boost the energy of the particles along the way.

https://www.home.cern/science/accelerators/large-hadron-collider

4 main experiments:

- ALICE
- ATLAS
- CMS
- LHCb

Filippo Costa is:

- Staff at CERN working for ALICE-DAQ
    - firmware developer for PCIe readout card (DAQ)
- Coordinator for the detector readout activities in ALICE
- Central and Detector Software Release coordinator

INTRODUCTION

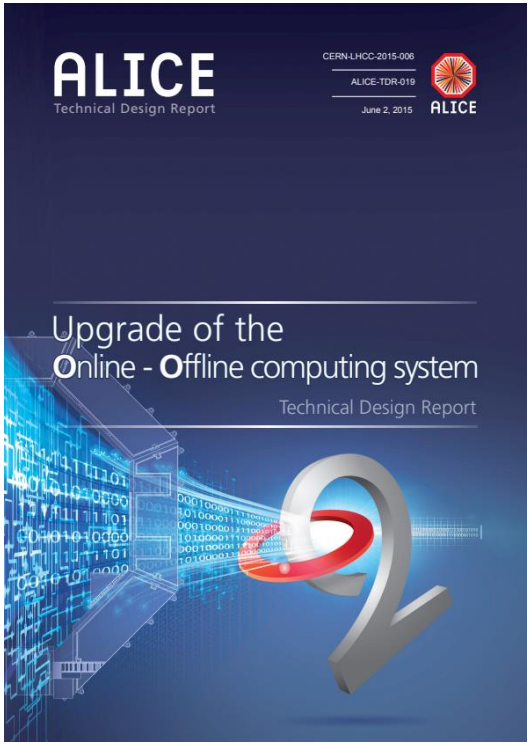
The presentation describes the DAQ and the TRIGGER systems in ALICE and the implementation of the CONTINUOUS READOUT.

The first part describes the motivation for using CONTINUOUS READOUT and the challenges that came with the upgrade.

The central part of the presentation gives details concerning the ALICE DAQ and TRIGGER systems.

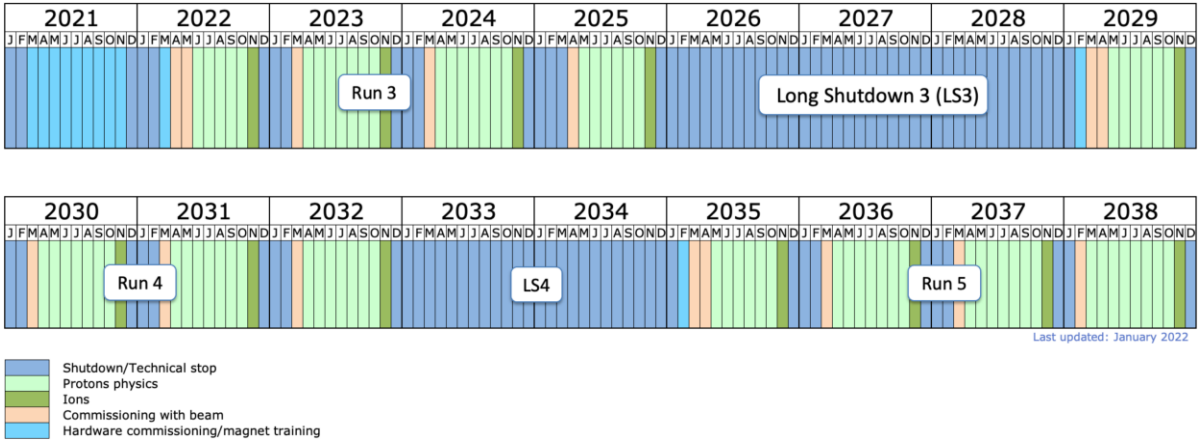
In the end results from the recent STABLE BEAM PERIOD are shown.

MOTIVATION

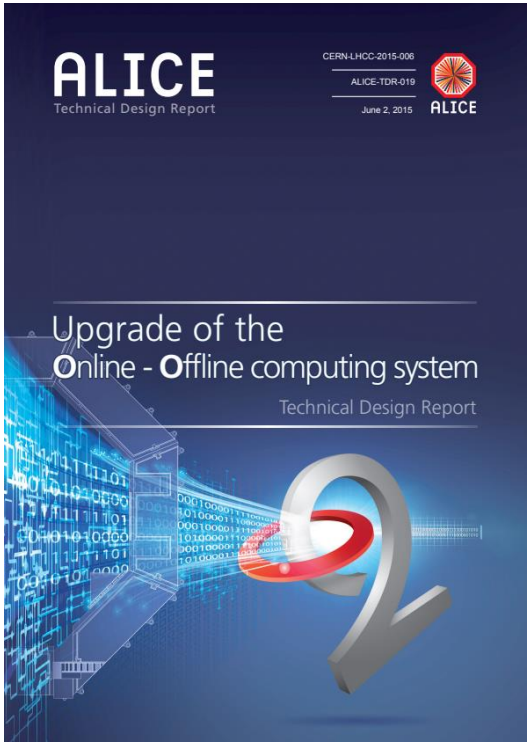


In RUN3 to keep up with the Pb-Pb 50kHz interaction rate, the TPC requires the implementation of a continuous read-out process to deal with event pile-up and avoid trigger-generated dead time.

The resulting data throughput from the detector has been estimated to be greater than **3.5TB/s for Pb–Pb** events, several orders of magnitude more than in Run 1/2.

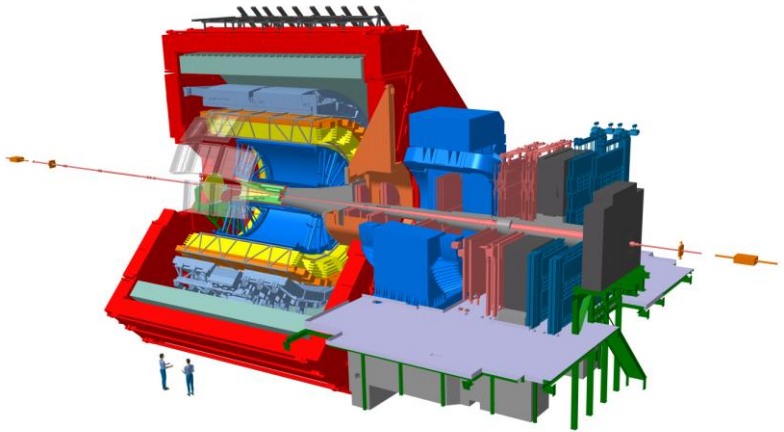


ALICE O^2 UPGRADE for the LHC RUN 3 2022

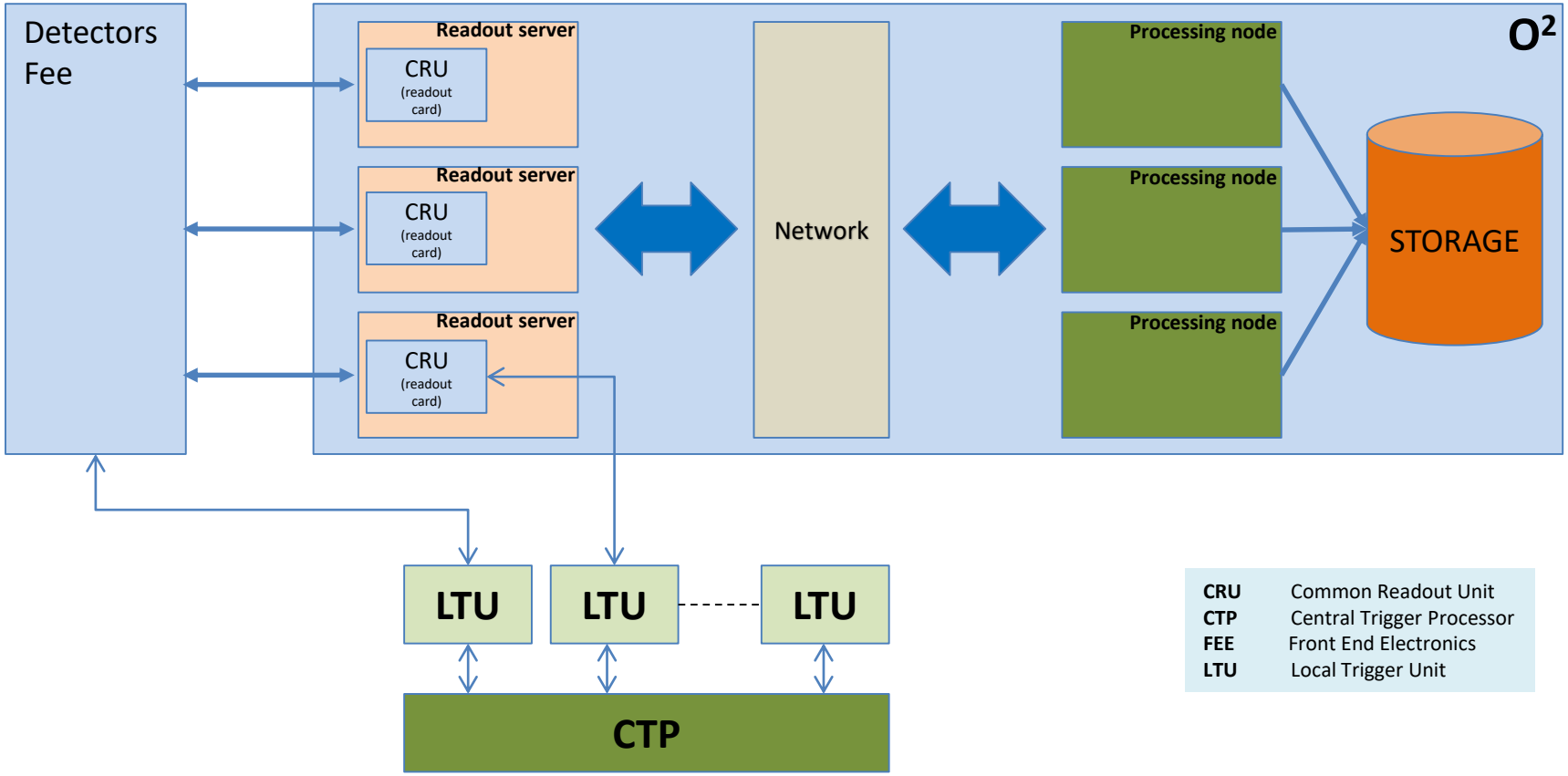


A new **ALICE Online and Offline (O^2)Computing** has been developed, the **ALICE O^2.**

The new O^2 facilities provide:
- continuous readout
- synchronous and asynchronous reconstruction during data taking (no raw data recording !!)
- two different categories of computing nodes, corresponding to the two data aggregation steps
- High-throughput system, equipped with hardware acceleration

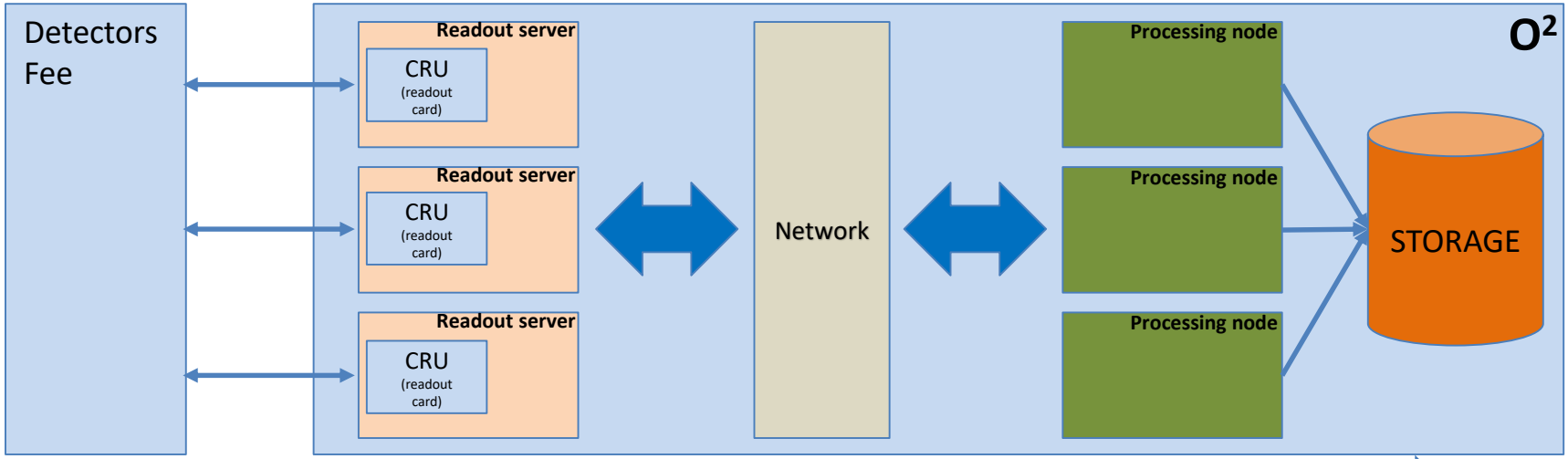


ALICE O²



CRU	Common Readout Unit
CTP	Central Trigger Processor
FEE	Front End Electronics
LTU	Local Trigger Unit

ALICE O^2



- 8000 links connect the detectors to O^2 farm
- 200 readout servers
  - 500 readout cards
- 250 processing nodes collect and store data
  - ~2000 GPU & CPU
- 1 CTP - 15 LTUs (1 per detectors)

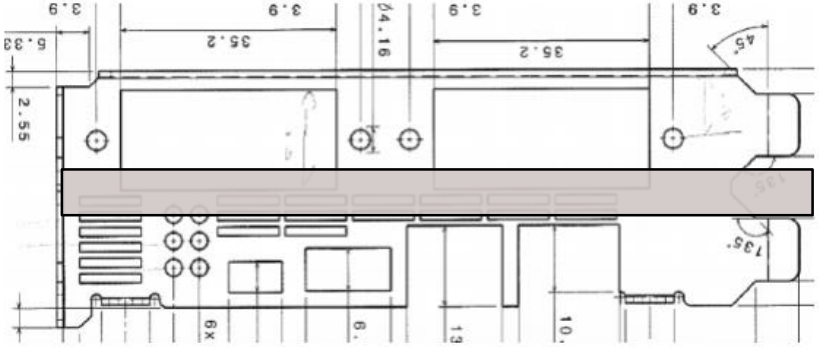
CRU	Common Readout Unit
CTP	Central Trigger Processor
FEE	Front End Electronics
LTU	Local Trigger Unit

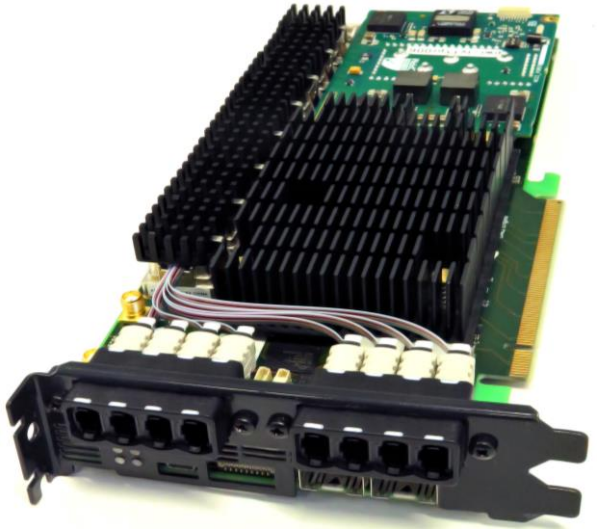

WARNING
SPOILER
ALERT

CHALLENGES:

- mechanical
- heat dissipation
- different FEE -> same data format
- high data rate
- no trigger BUSY
- heavy FPGA/GPU/CPU processing
  - data flow
  - data compression
  - data processing
- online compression

MECHANICAL INSTALLATION

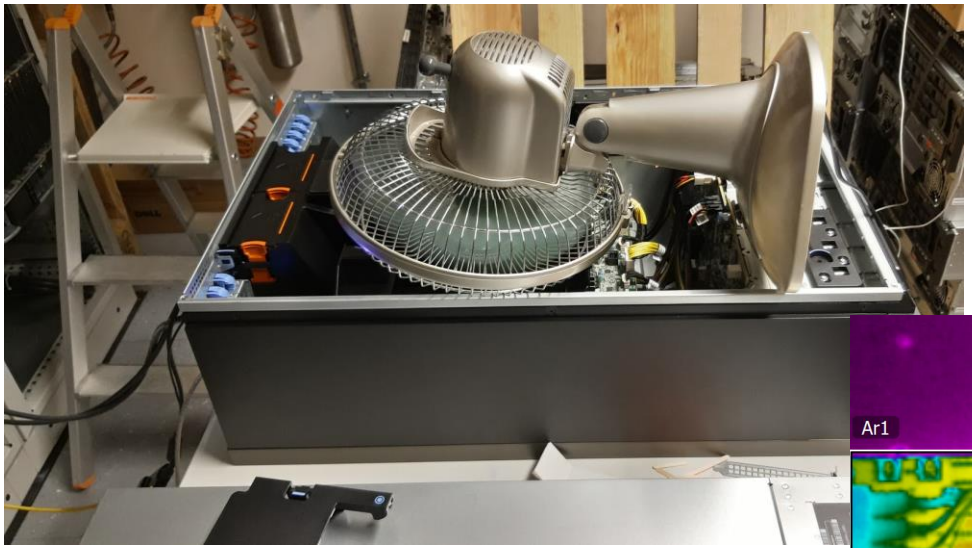




- same readout card from another CERN experiment **(LHCb's PCIe40)**
- custom firmware to address ALICE requirements:
  - *continuous readout*
  - *data processing in FPGA*
  - *distribution of the trigger messages*
  - *slow control of the FEE*

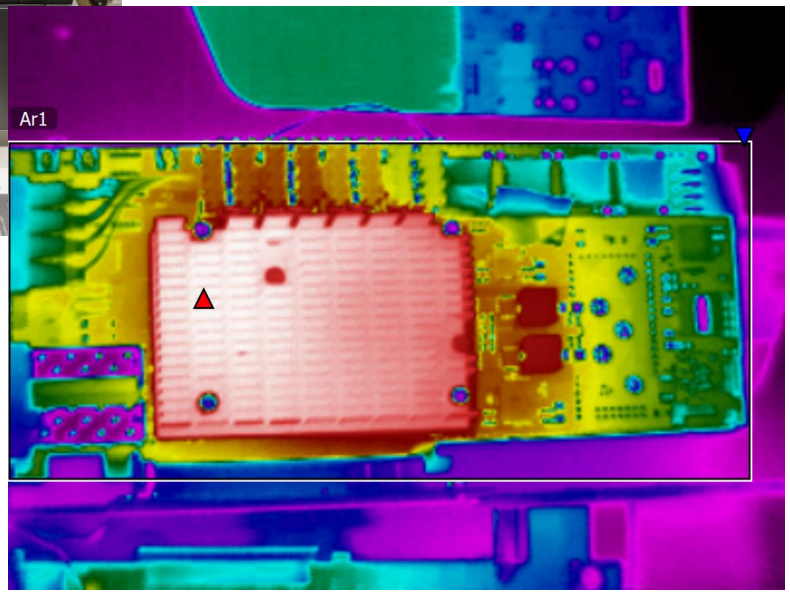
O. Bourrion et al 2021 JINST 16 P05019

HEAT DISSIPATION



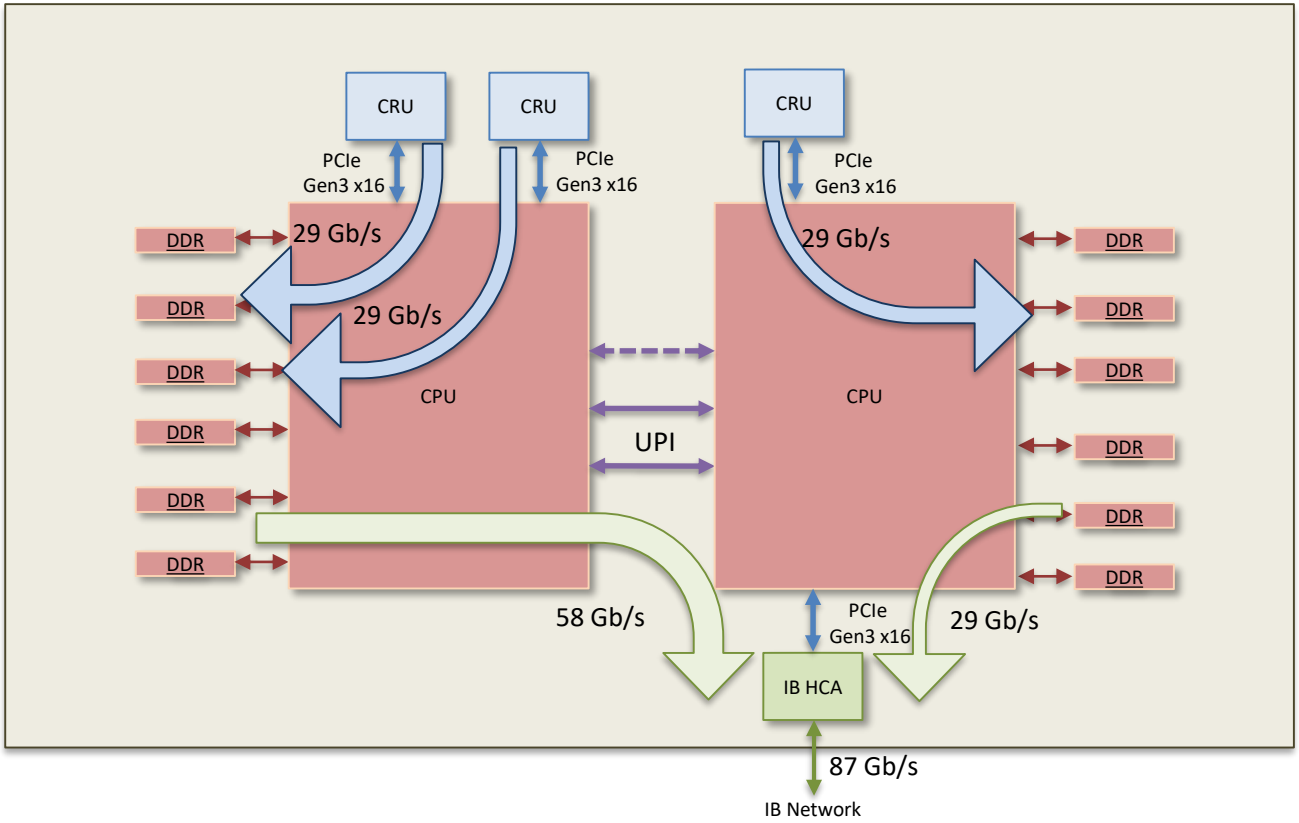
FPGA MAX temp 80 C

MAX ALLOWED TEMP
- FPGA : 60 C



CHEP 2019
https://indico.cern.ch/event/773049/contributions/3474356/

DATA FLOW in the readout server

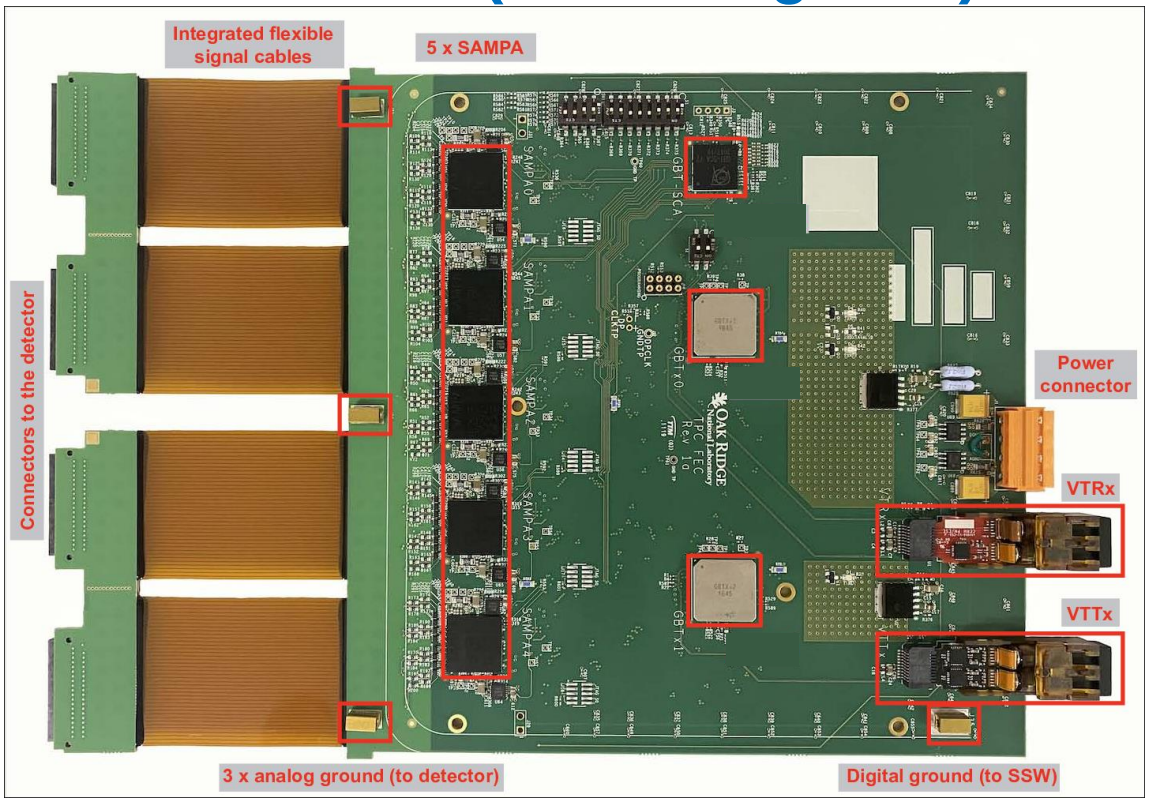


Dual CPUs system
- NUMA nodes
- BIFURCATION
- Moving data across CPUs

DIFFERENT FEE – SAME DATA FORMAT

Every detector in ALICE has a different FEE, that are readout by the same card

FEE w/o FPGA (streaming data)



- *NO FPGA installed on the FEE*
- *DATA coming out from the FEE is a continuous stream of information*

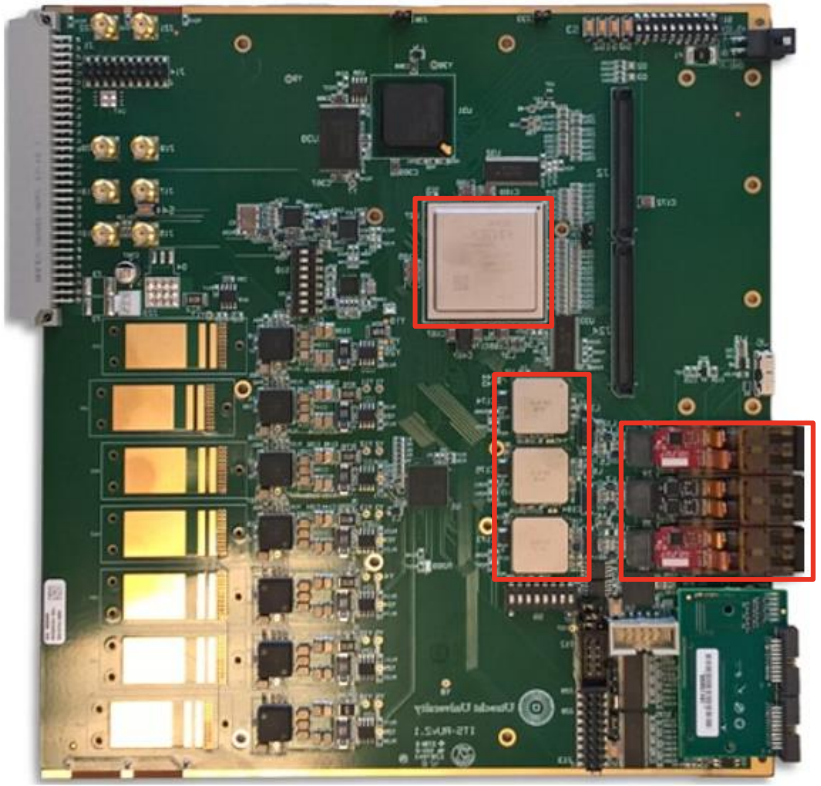
DATA

DATA

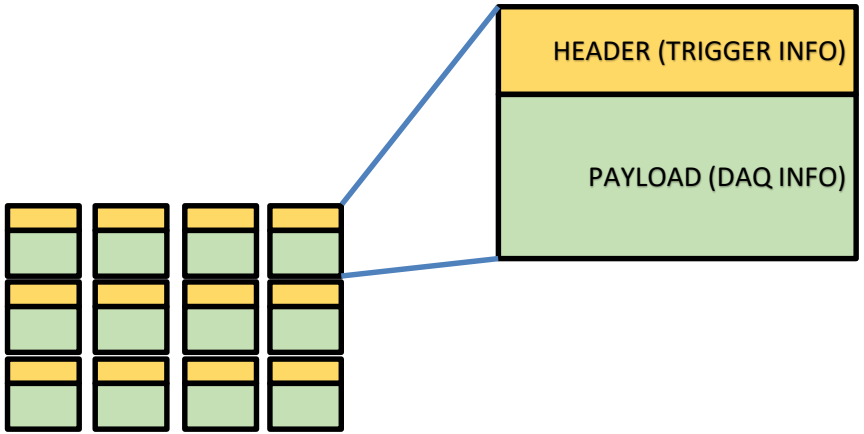
DETECTOR

DAQ

FEE w/ FPGA (packet data)



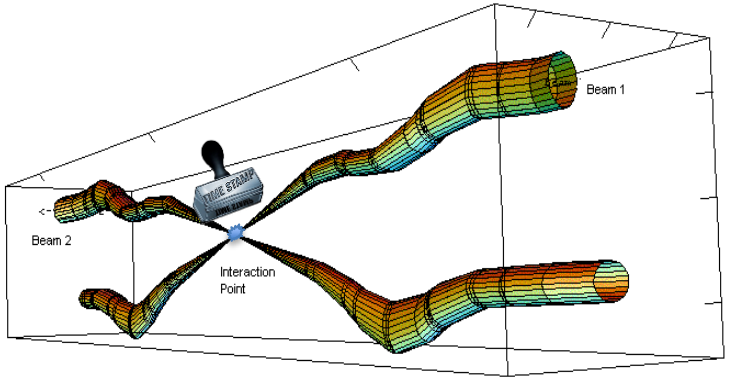
- *1 FPGA installed on the FEE (1)*
- *DATA coming out from the FEE is formatted in packets (HEADER + PAYLOAD)*

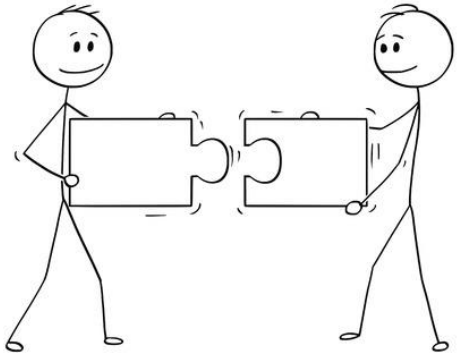


DETECTOR

DAQ

ALICE STREAMING READOUT





The software that processes data requires timing information to match streams coming from different detectors.

How do we add the trigger information in a continuous stream of data?

We invented 2 new trigger objects:

Heart Beat trigger

TIME FRAME

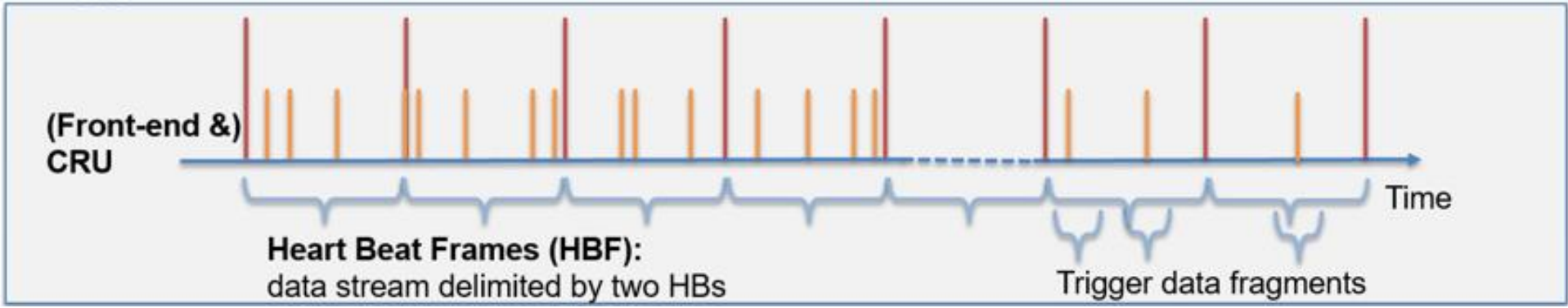
HB TRIGGER and TIME FRAME in a slide

Heart Beat (HB)
issued in continuous & triggered modes to all detectors

Physics trigger
can be sent to upgraded detectors
will be sent to non-upgraded detectors

HB rate ~10 KHz
Time Frame = 128 HB frames

Triggered read-out



CONTINUOS Vs TRIGGERED readout

"*In particle physics, a trigger is a system that uses simple criteria to rapidly decide which events in a particle detector to keep when only a small fraction of the total can be recorded.*"

In triggered mode ALICE collects data "triggered" by a specific event (***PHYSICS trigger***)

In a continuous mode ALICE collects data constantly. There is no need for a PHYSICS trigger for the FEE to generate data

Example

PHYSICS TRIGGER

Let’s take this stream of information as our ORIGINAL DATA.
The trigger is when the GOLF-CLUB hits the ball

ALICE TRIGGERED READOUT

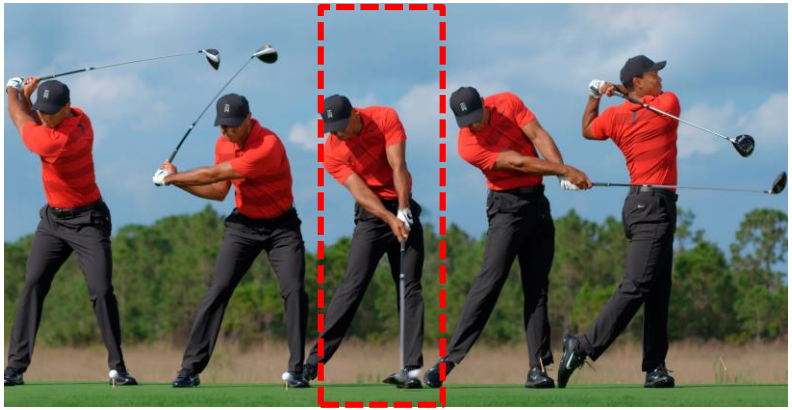
Original data

FEE data

Detector collects information upon detection of the trigger, for a given amount of time (buffer size on the electronics)

ALICE CONTINUOUS READOUT

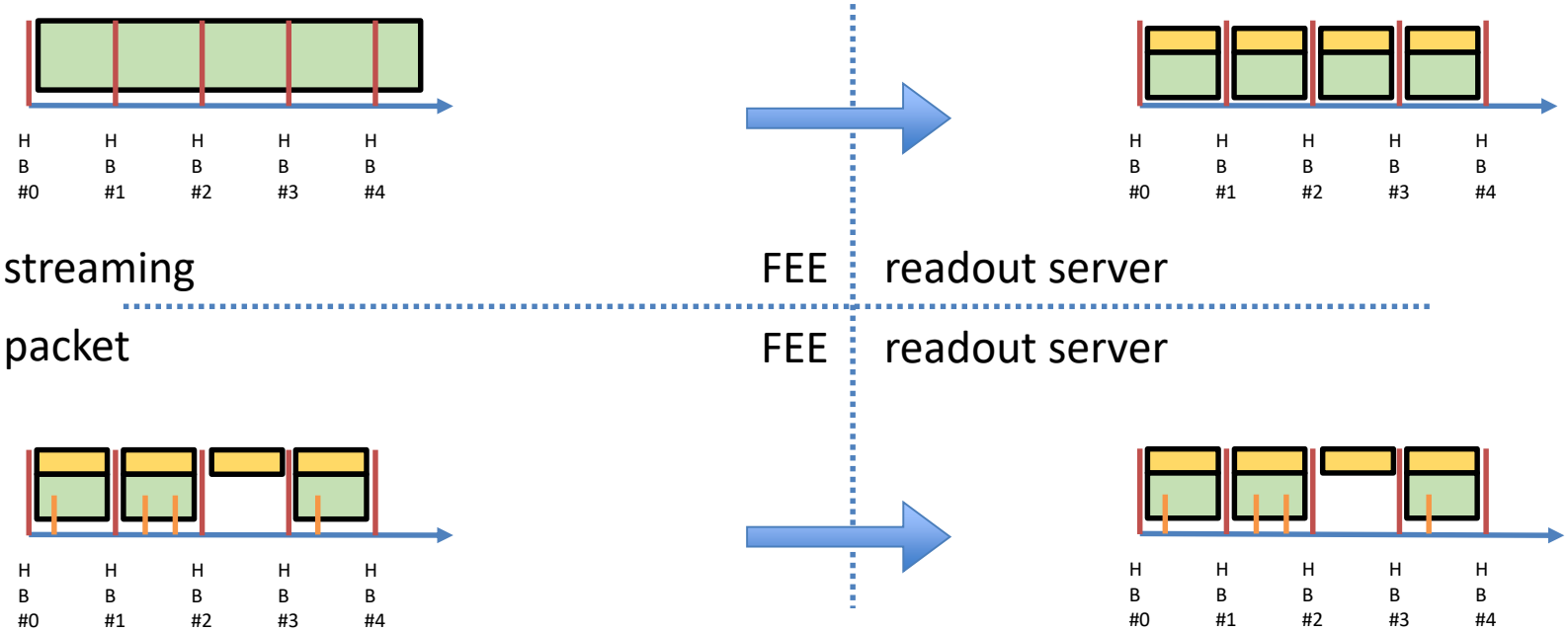
Original data



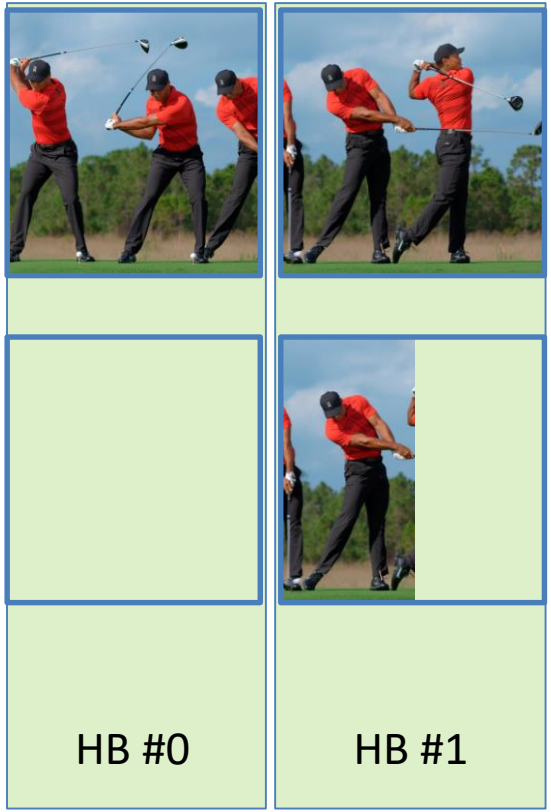
FEE data

Detector collects data constantly.
In the stream of information, it is possible to identify the PHY trigger

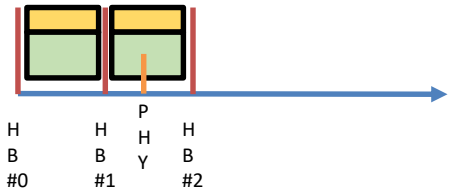
DIFFERENT FEE - SAME DATA FORMAT



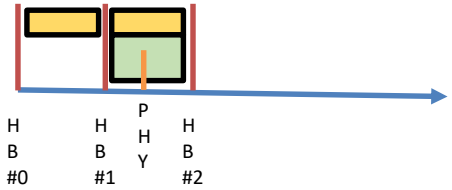
ALICE DATA FORMAT - HB FRAME



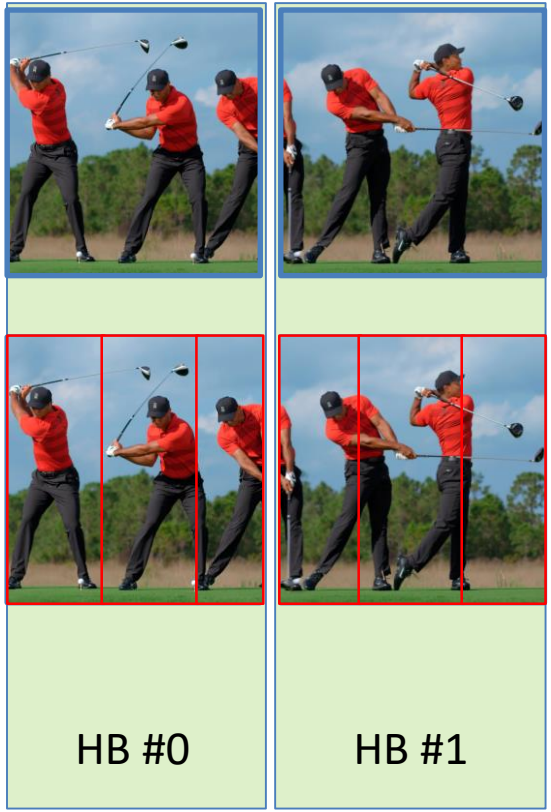
STREAMING DETECTOR



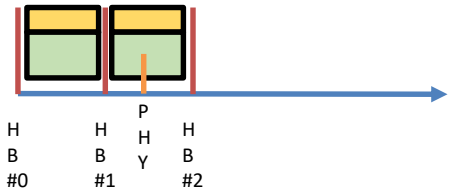
TRIGGERED DETECTOR
(1 PHY trigger)



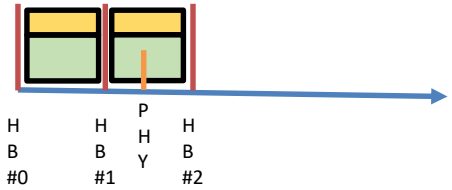
ALICE STREAMING READOUT



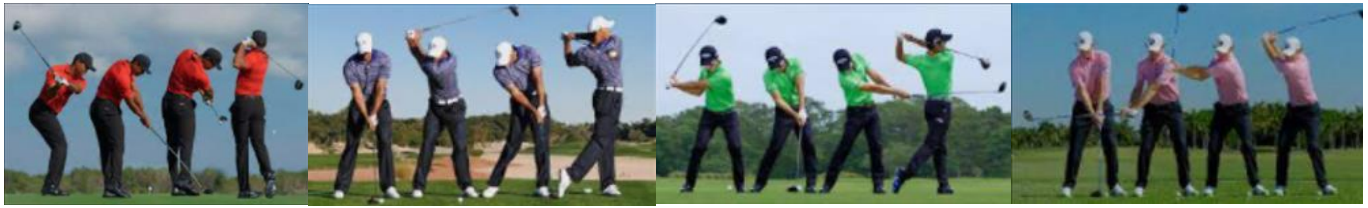
STREAMING DETECTOR

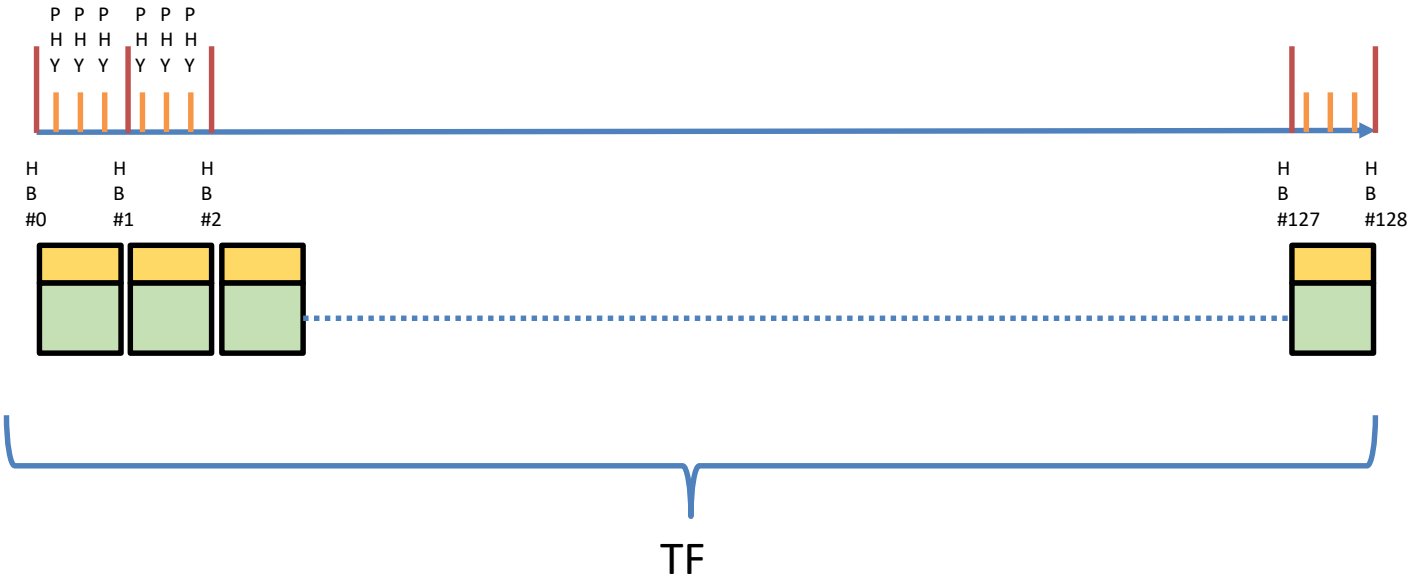


auto-TRIGGERED DETECTOR

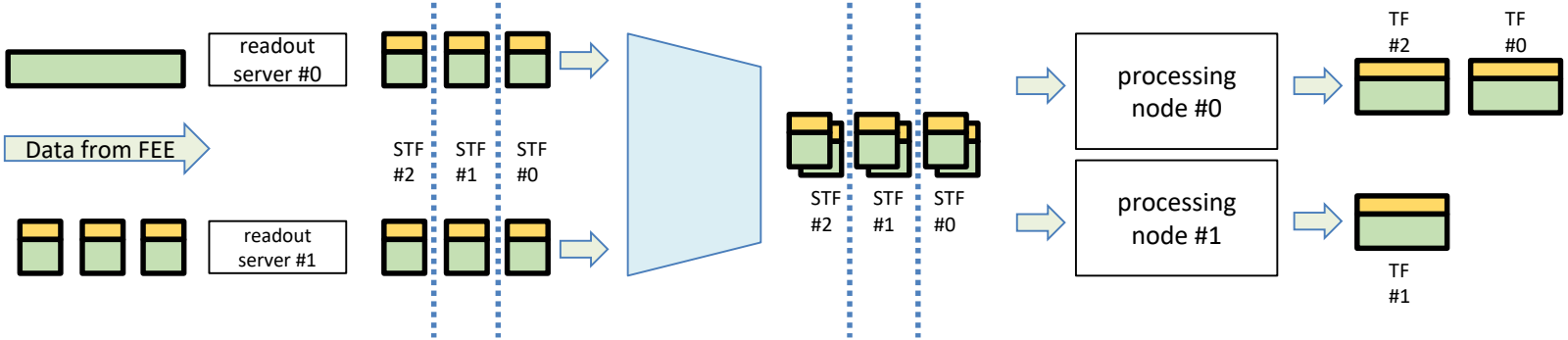


Sub - Time Frame (128 HB frame)



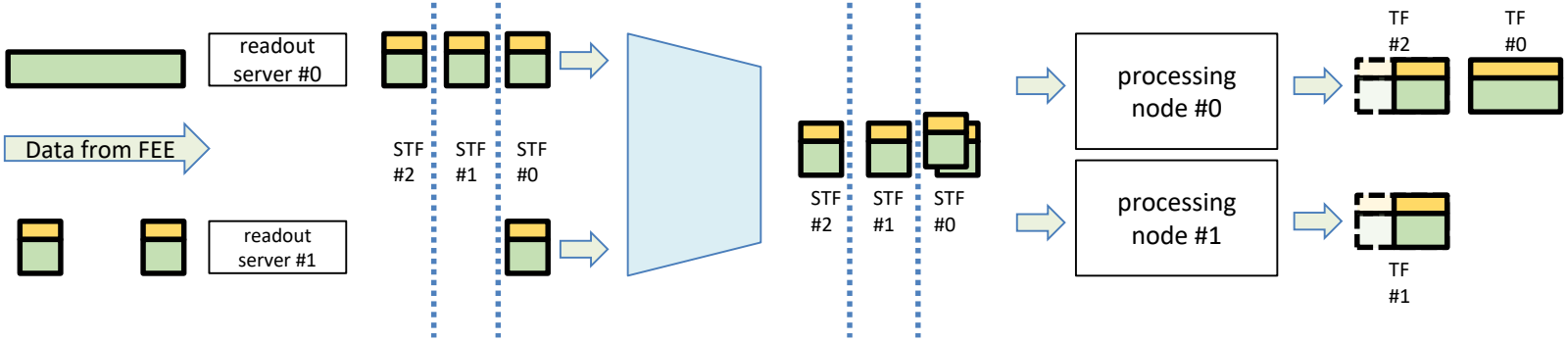


Time Frame (processing node)



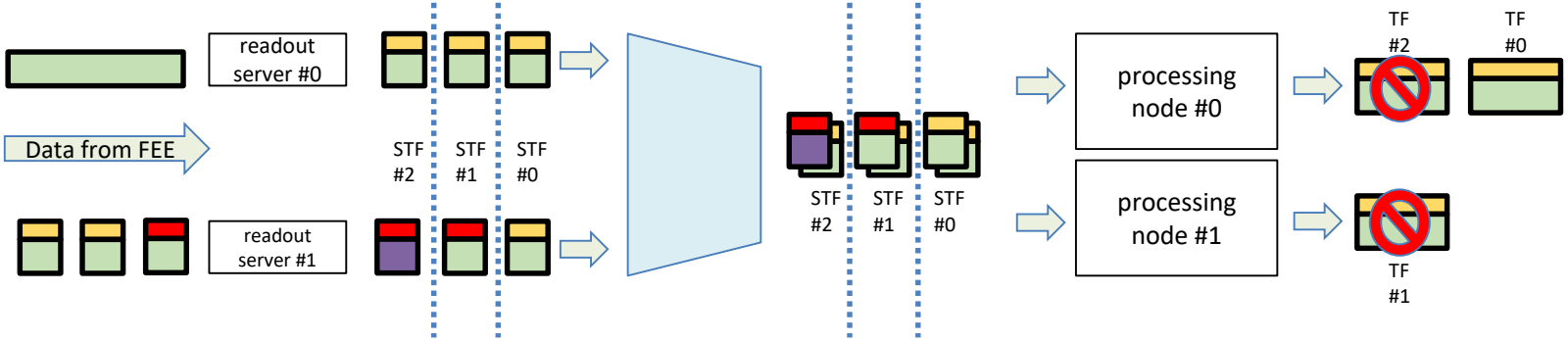
Readout server generates sub-TF.
Processing node build the complete TF collecting data from all the detectors.

Time Frame (incomplete TF)



If some readout servers can’t generate data for a specific TF, the processing node will build an incomplete TF.
That means for a specific period data from part of a detector are missing.

Time Frame (rejected TF)



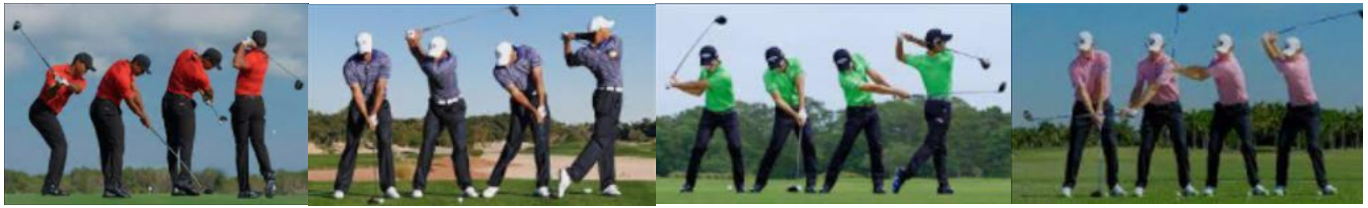
Processing node can reject the TF if the data is corrupted and can't be merged:

- RDH information not correct
- RDH information corrupted

THROTTLE DOWN DATA RATE

- HB ACCEPT/REJECT
- TF rejection
- DATA COMPRESSION

HB ACCEPT/REJECT



Original data

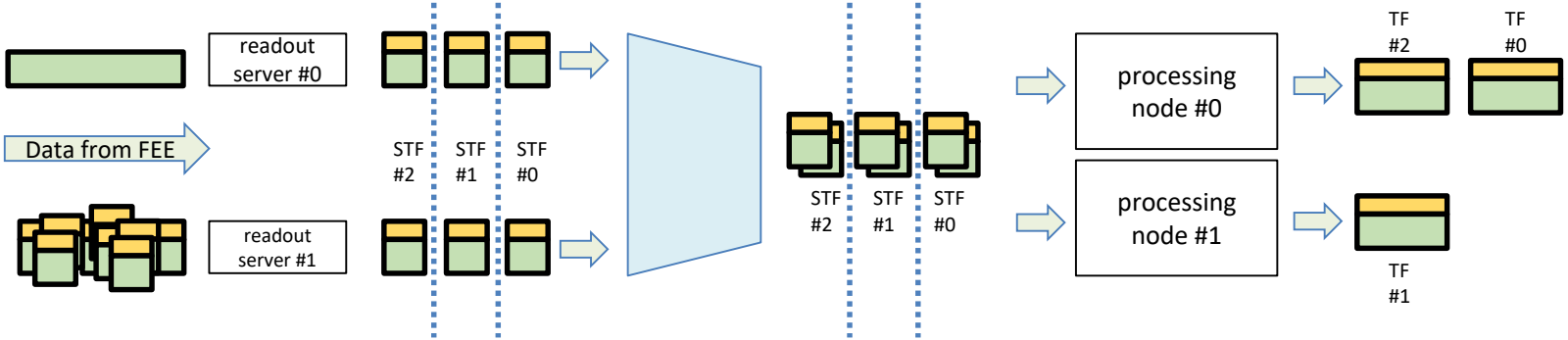
Saved data

HB #0 HB #63 HB #127 HB #128

HB A (50 %) HB R (50 %)

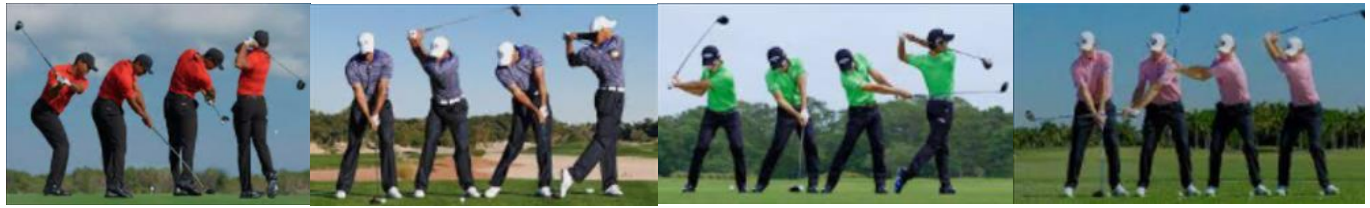
TF

Time Frame (incomplete TF)

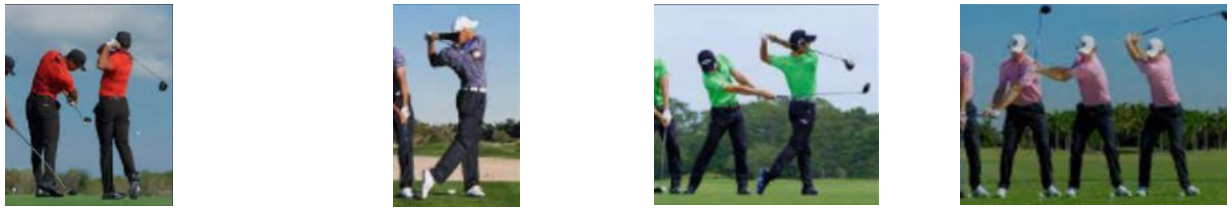


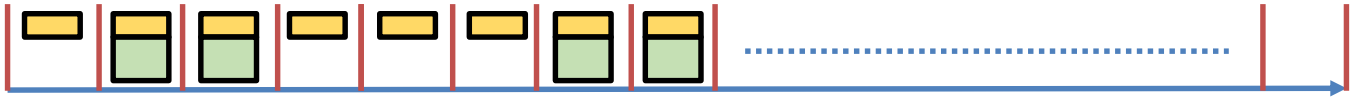
The detector is generating more data than what the hardware/software can handle, some of these information will be rejected.
We still build the TF, from the trigger information it is complete (it contains all the 128 ORBIT information), but part of the payload is rejected.

DATA COMPRESSION/PROCESSING



Original data



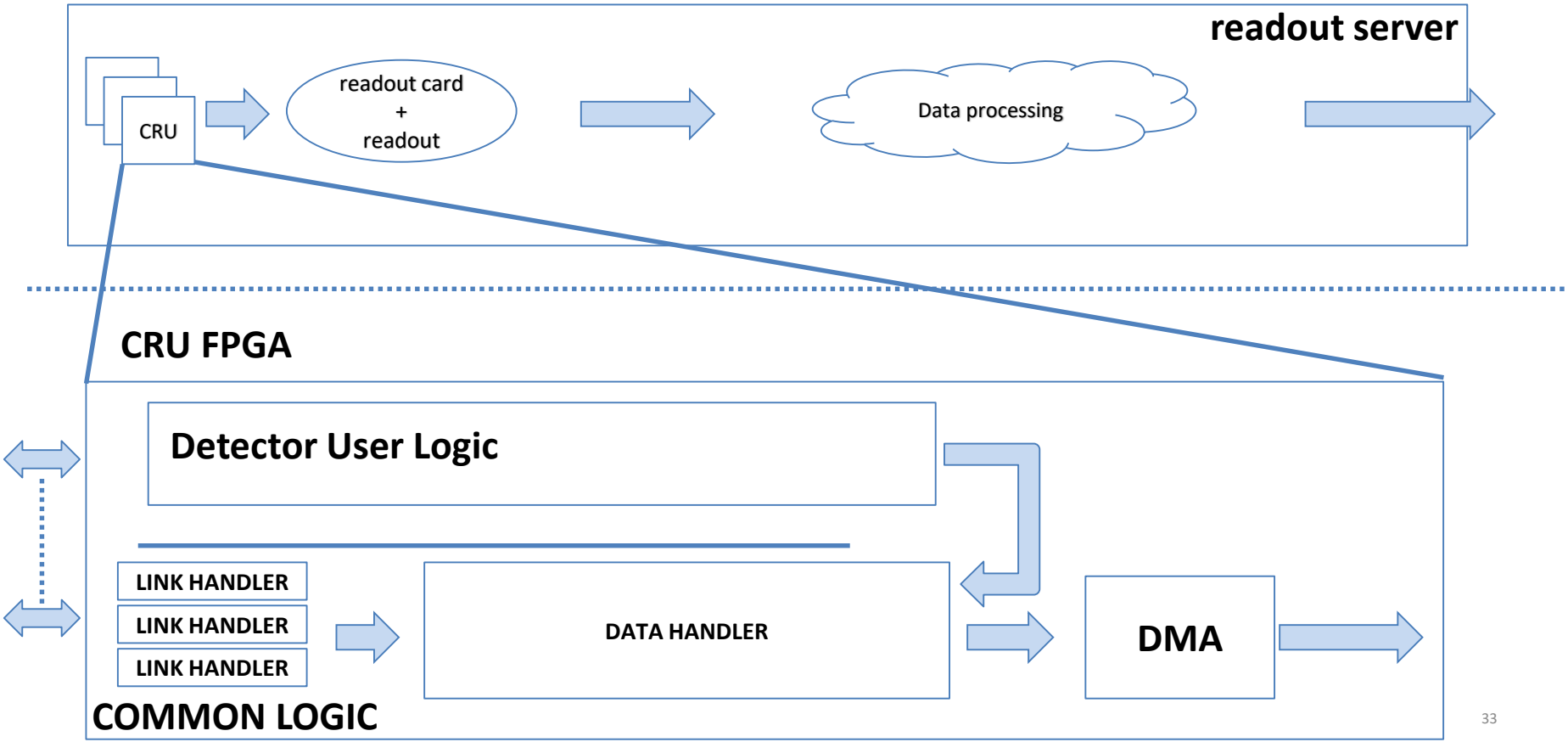


HB #0 HB #1 HB #2 HB #3 HB #4 HB #5 HB #6 HB #7 HB #8 HB #127 HB #128

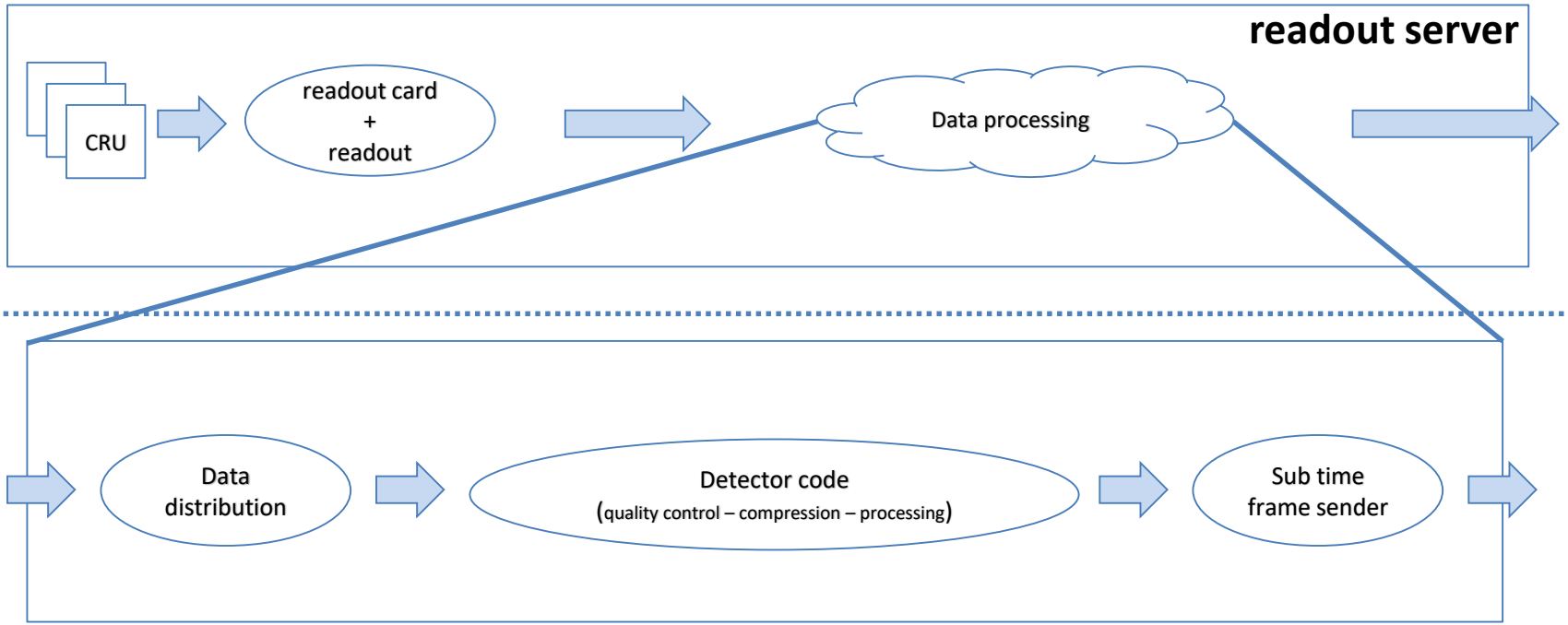
Saved data

TF

DATA COMPRESSION /PROCESSING

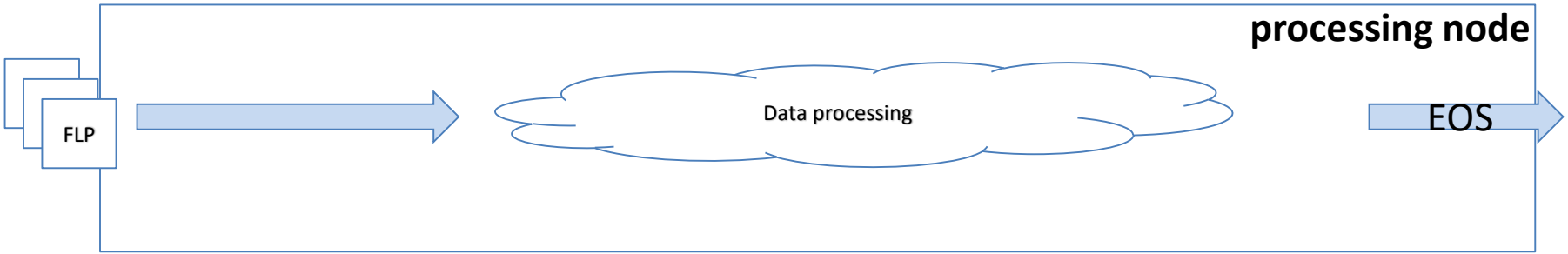


DATA COMPRESSION/PROCESSING (CPU)



DATA can be compressed by detector software running on the FLP before reaching the EPNs.

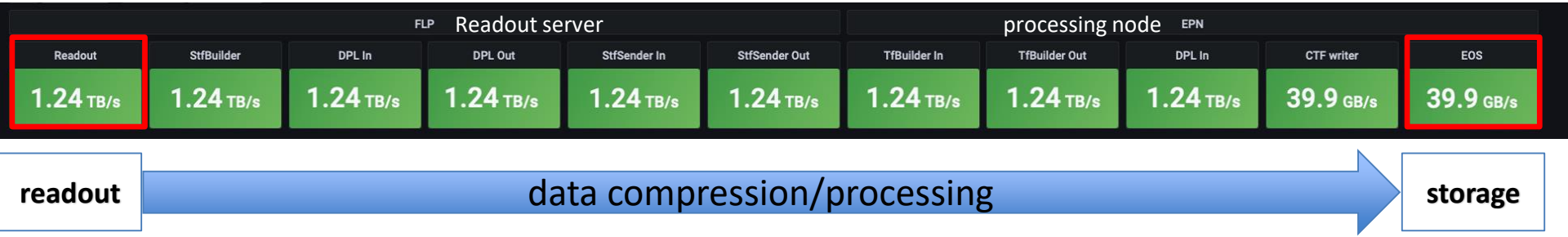
DATA PROCESSING (CPU - GPU)



https://indico.phy.ornl.gov/event/112/contributions/469/attachments/496/1348/2021-12-09_Streaming_Readout_Workshop_O2.pdf

CONCLUSIONS …

- The ALICE O^2 system:
  - Hardware is installed
  - Software is in constant evolution
- ALICE is collecting data with beam since beginning of 2022:
  - High data rate tests at different IR (2, 3, 4 MHz)
  - Data rate is larger than what we expected (almost a factor 2), but we are working to mitigate the effect
- **HI test-run happening next week**




General / Dataflow
run 528987
thresholds PHYSICS
now-5m to now-5s
FLP
Readout 117 GB/s
StfBuilder 117 GB/s
DPL In 28.6 GB/s
DPL Out 27.5 GB/s
StfSender In 116 GB/s
StfSender Out 116 GB/s
EPN
TfBuilder In 161 GB/s
TfBuilder Out 161 GB/s
DPL In 116 GB/s
CTF writer 1.57 GB/s
EOS 276 MB/s
Public / FLP Big Screen display
Run 529006
529006
Start of run 2022-11-10 08:13:42
Env ID 2beXxbXwy8V
Run number 529006
Detectors CPV EMC FDD FT0 FV0 HMP ITS MCH MFT MID PHS TOF TPC TRD ZDC
State RUNNING
Run type PHYSICS
FLP
Readout 113 GB
StfBuilder 113 GB
StfSender 112 GB
EPN
TFBuilder 108 GB/s
DPL in 112 GB/s
CTF Writer 988 MB/s
ALICE DCS monitoring
Detector Control System
08:41:39 Thu, 10/11/2022
Magnets
Dipole on positive 6000 A 682 mT
Solenoid on positive 30000 A 452 mT
ALICE Permit
ALICE injection safe
Beam permit
Injection permit 1
Injection permit 2
Dipole beam permit
Detectors
CPV READY
EMC READY ss
FDD READY
FT0 READY
FV0 READY
HMP READY
ITS READY
MCH READY
MFT READY
MID READY
PHS READY ss
TOF READY
TPC READY
TRD READY
ZDC READY
Alarms
DSS Ok
CSAM Ok
LHC status
STABLE BEAMS
no handshake active
DCS on Thu 10/11/2022, 01:15
ALICE is taking physics.
LHC on Thu 10/11/2022, 07:59
fill for VdM scans
LHCb
following in this fill:
ALICE

What did I learn so far

What did I learn so far

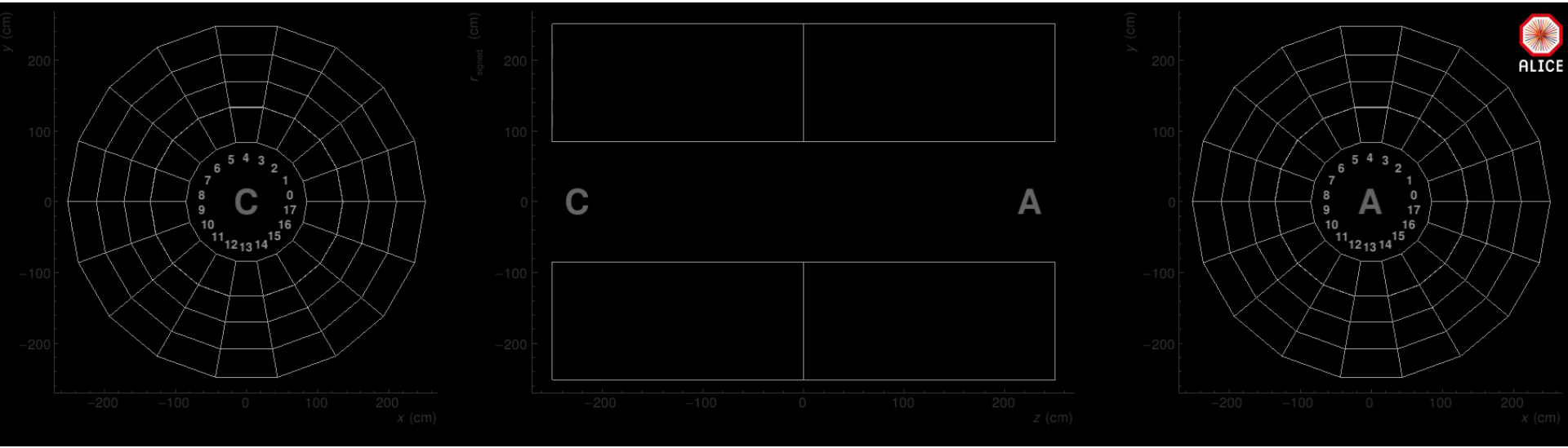
HARDWARE

- Maintenance has a big cost
- Use as much as possible the same platform (FEE – READOUT CARD – DAQ protocol)
- Complex readout card (interface DCS, TRG, DAQ) has a lot of benefits, but it becomes a single point of failure

SOFTWARE

- Build proper test system and monitoring tools. During operation is difficult to debug the software in production
- Find a way to install small components and fast. During stable beam period it is difficult to allocate slot to install large sw components that can take hours to deploy and test
- Break it soon, fix it earlier:
    - Do not wait to have the perfect solution before release a feature. Start small and simple and grow from there
    - Do not be afraid to find bugs

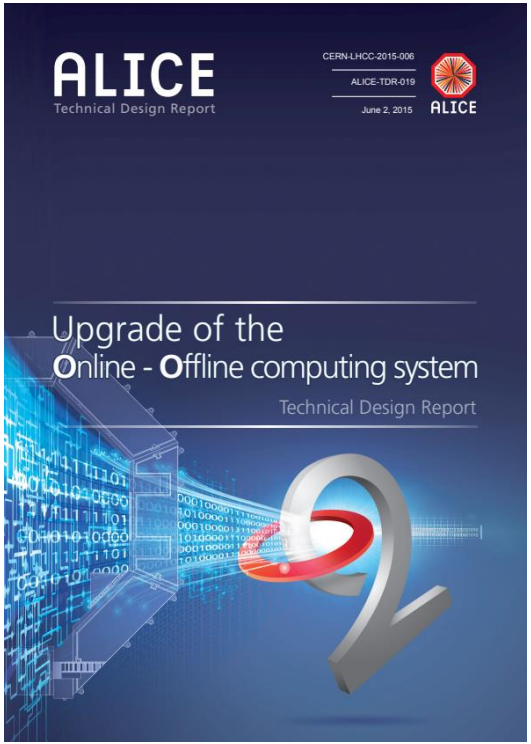
TPC CONTINUOUS READOUT



Thank you for your attention

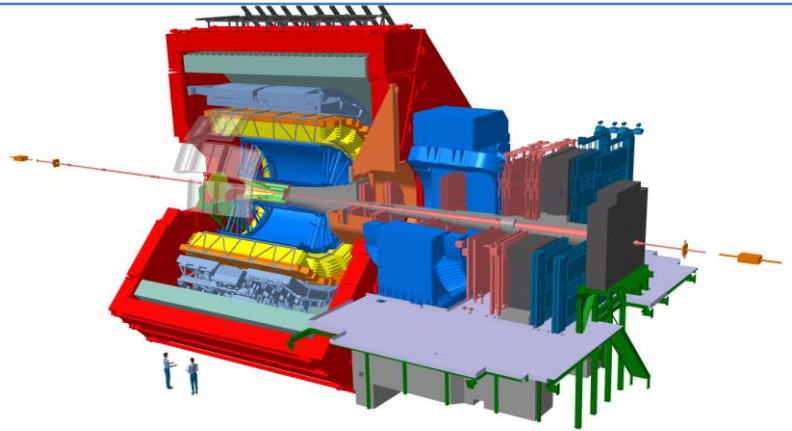
BACKUP SLIDE

ALICE NEW DETECTORs for the LHC RUN 3 2022



The **ALICE detector upgrade** includes:

- A new, high-resolution, low material Inner Tracking System (**ITS**).
- An upgrade of the Time Projection Chamber (**TPC**) consisting of the replacement of the wire chambers with Gas Electron Multiplier (**GEM**) detectors and new continuous read-out electronics.
- The addition of a Muon Forward Tracker (**MFT**).
- A new Fast Interaction Trigger (**FIT**) detector.
- An upgrade of the read-out electronics of several detectors:
    - Muon CHamber System (**MCH**),
    - Muon IDentifier (**MID**),
    - Transition Radiation Detector (**TRD**) ,
    - Time-Of-Flight detector (**TOF**),
- A new Central Trigger Processor (**CTP**).



CRU TRIGGER MODE (PHY TRIGGER IN HB FRAME)

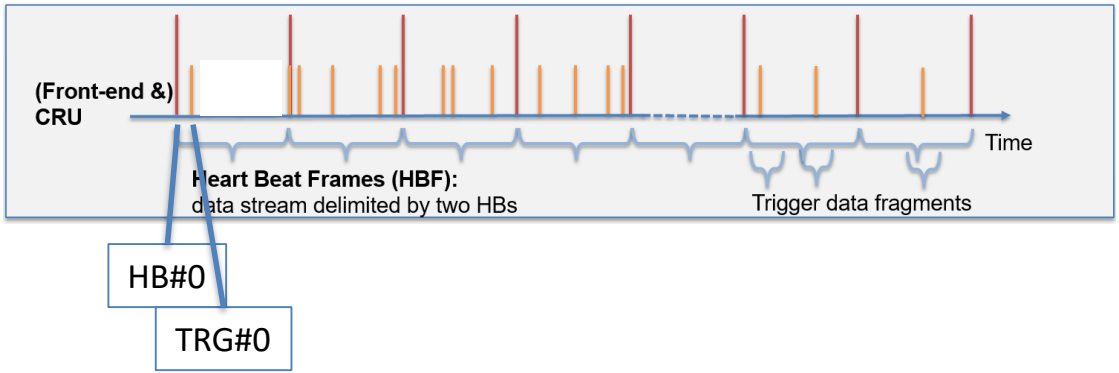
HB#0 START

PAYLOAD OF TRG#0

HB#0 END

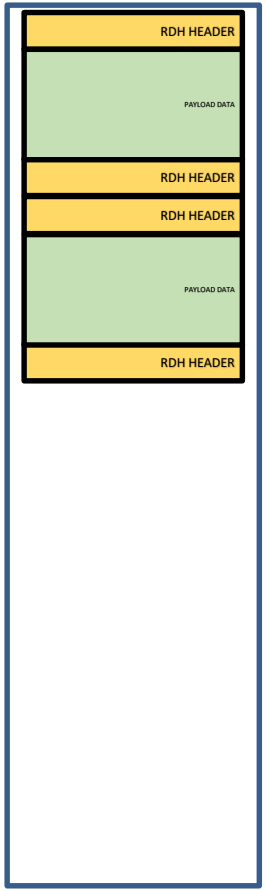
HB#1 START

HB FRAME
1 TRG



TRG information (ORBIT+BC) are embedded in the payload if needed.
The RDH contains HB trigger information (ORBIT + BC)

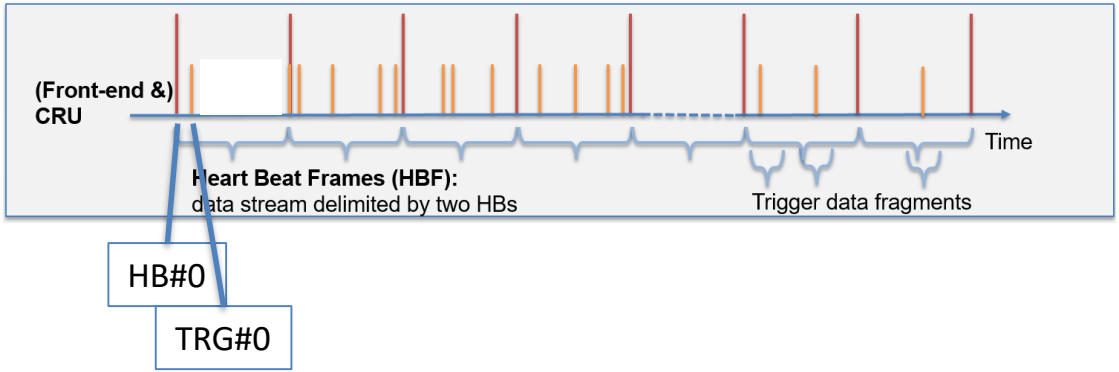
CRU CONT MODE (STREAMING DETECTOR)



HB#0 START

HB#0 END
HB#1 START

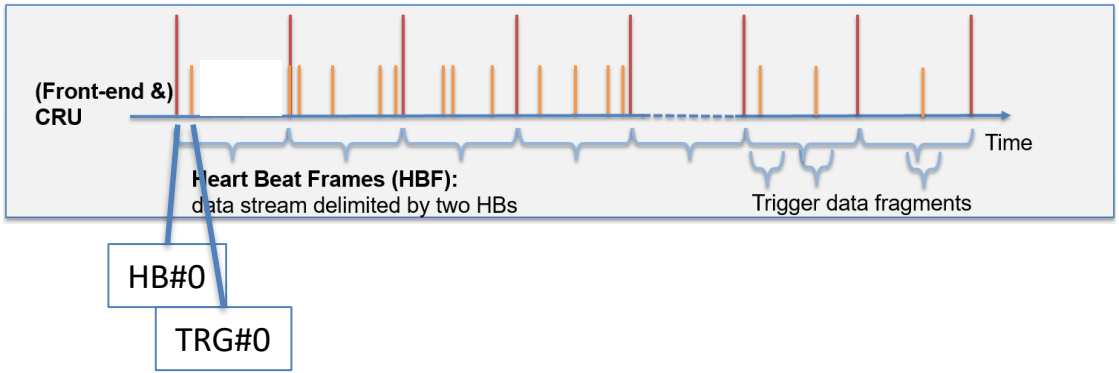
HB FRAME



CRU chops the data and generates HB frame. Every HB frame consists of several data block (RDH + PAYLOAD). The content of the payload is RAW data from the GBT.
CRU generates new HB frames for every HB trigger

Figures are there to help the description of the concept.
They do not represent a realistic data format.
In this case multiple data blocks are represented as a single block

CRU + UL CONT MODE (STREAMING DETECTOR)



The UL generates the HB frames.
The content of the payload depends from the result of the UL.
The HB frame can be empty.

CRU TRIGGER MODE (PHY TRIGGER IN HB FRAME)

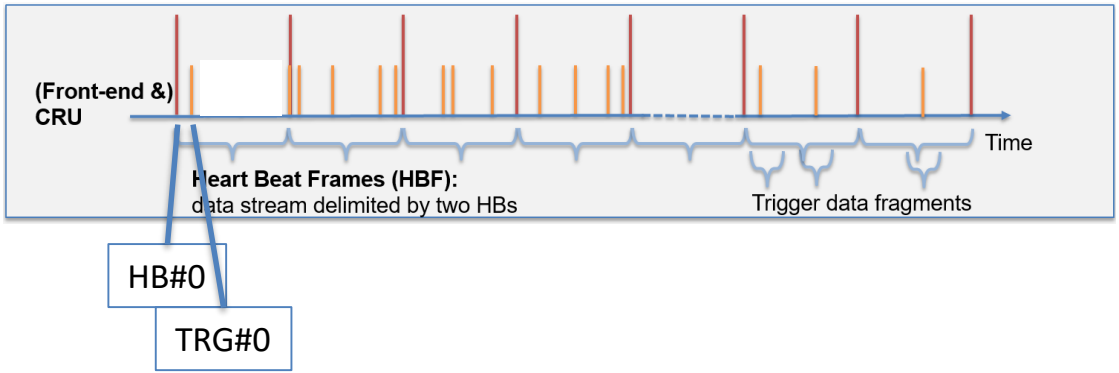
HB#0 START

PAYLOAD OF TRG#0

HB#0 END

HB#1 START

HB FRAME
1 TRG



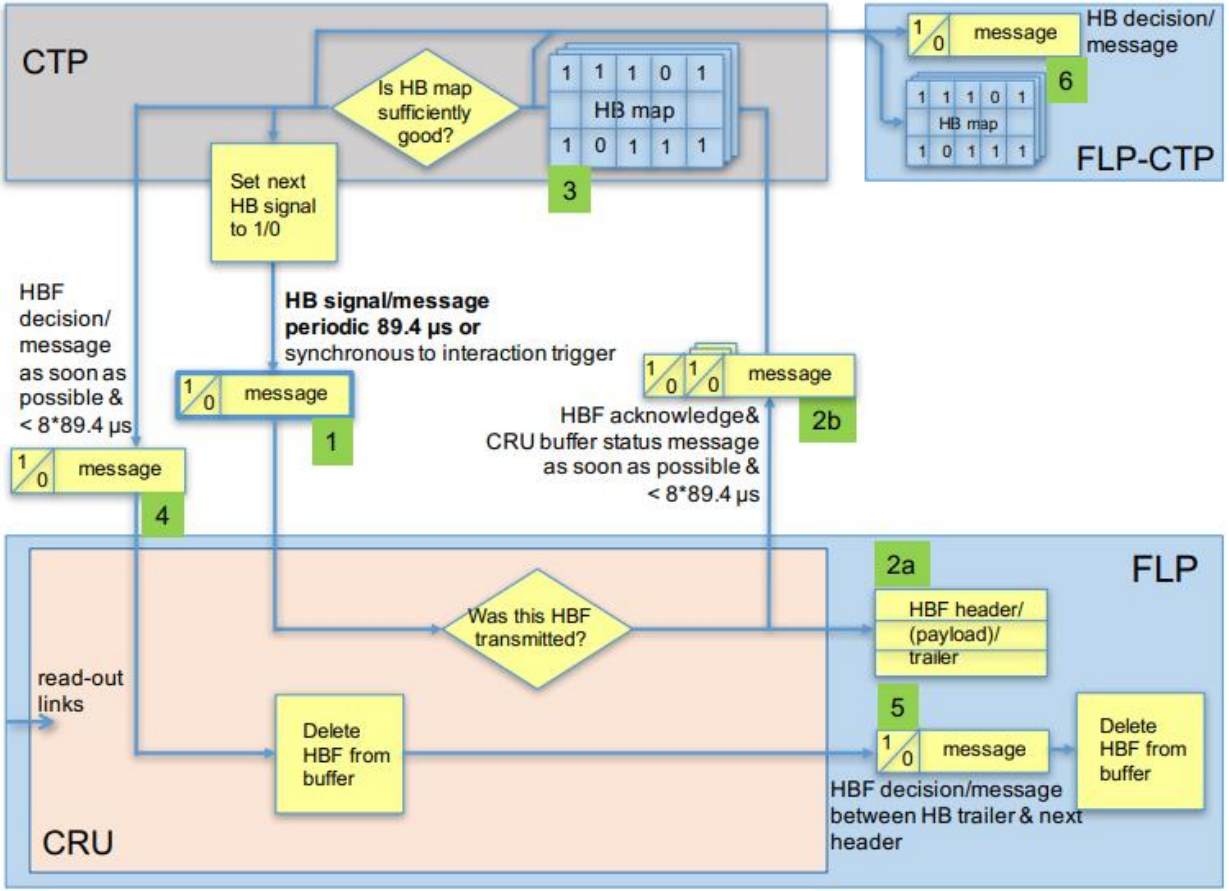
TRG information (ORBIT+BC) are embedded in the payload if needed.
The RDH contains HB trigger information (ORBIT + BC)

RDH

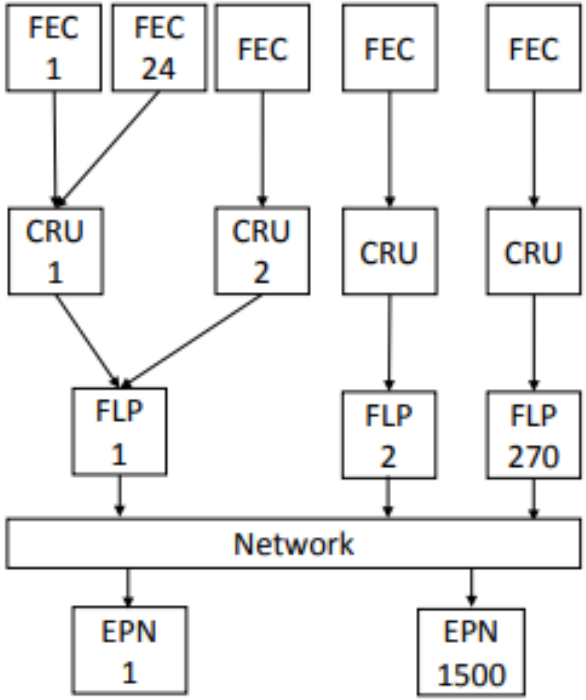
[31-0]			
FEE ID [31-16]		HEADER SIZE [15-8]	HEADER VERSION [7-0]
RESERVED [31-16]		SOURCE ID [15-8]	PRIORITY BIT [7-0]
MEMORY SIZE [31-16]		OFFSET NEW PACKET [15-0]	
DW [31-28]	CRU ID [27-16]	PACKET COUNTER [15-8]	LINK ID [7-0]
RESERVED [31-12]		BC [11-0]	
ORBIT [31-0]			
RESERVED [31-0]			
RESERVED [31-0]			
TRG TYPE [31-0]			
RESERVED [31-24]	STOP BIT [23-16]	PAGES COUNTER [15-0]	
RESERVED [31-0]			
RESERVED [31-0]			
DETECTOR FIELD [31-0]			
RESERVED [31-16]		PAR BIT [15-0]	
RESERVED [31-0]			
RESERVED [31-0]			

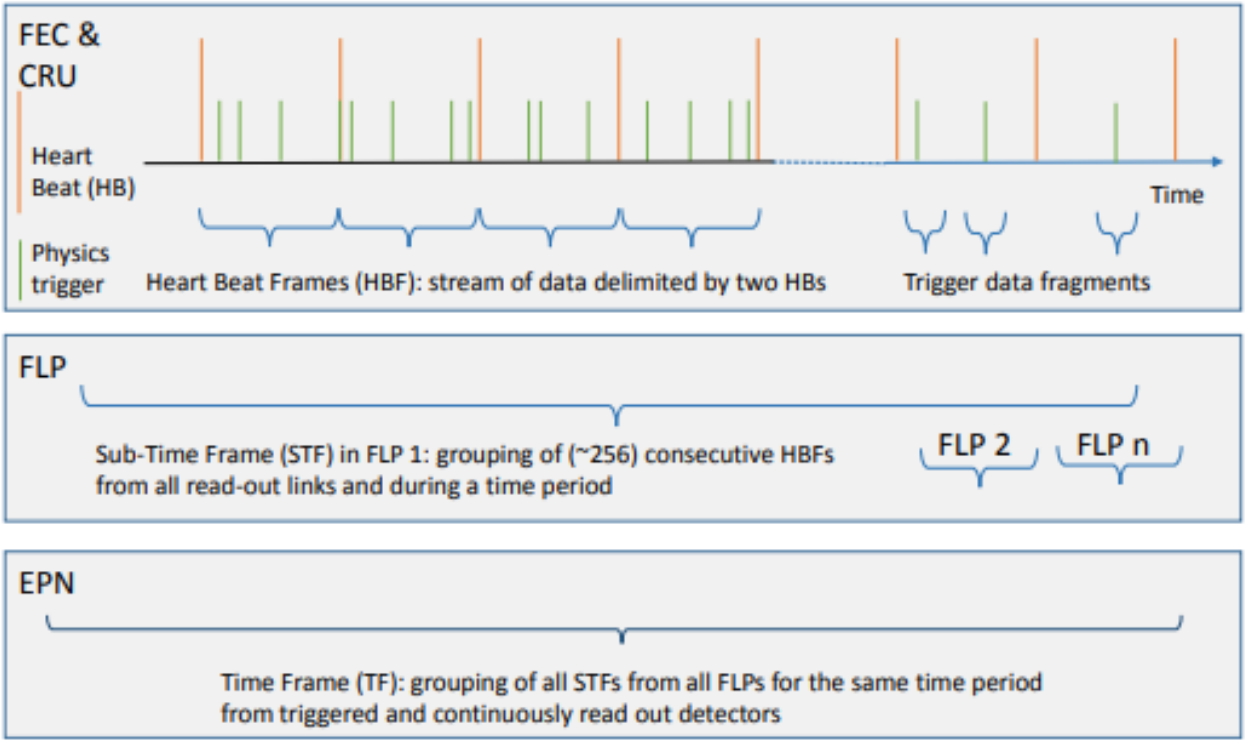
0
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15

HB MAP


CTP
Is HB map sufficiently good?
1 1 1 0 1
HB map
1 0 1 1 1
3
Set next HB signal to 1/0
HB signal/message
periodic 89.4 µs or
synchronous to interaction trigger
HBF decision/ message as soon as possible & < 8*89.4 µs
1/0 message
1
1/0 message
4
1/0 1/0 message
2b
HBF acknowledge& CRU buffer status message as soon as possible & < 8*89.4 µs
1/0 message
HB decision/ message
6
1 1 1 0 1
HB map
1 0 1 1 1
FLP-CTP
FLP
2a
HBF header/ (payload)/ trailer
Was this HBF transmitted?
read-out links
Delete HBF from buffer
5
1/0 message
HBF decision/message between HB trailer & next header
Delete HBF from buffer
CRU

Sub Time Frame – TIME FRAME ID





TRIGGER BITS

Bit	Name	Comment
0	ORBIT	ORBIT
1	HB	Heart Beat flag
2	HBr	Heart Beat reject flag
3	HC	Health Check
4	PhT	Physics Trigger
5	PP	Pre Pulse for calibration
6	Cal	Calibration trigger
7	SOT	Start of Triggered Data
8	EOT	End of Triggered Data
9	SOC	Start of Continuous Data
10	EOC	End of Continuous Data
11	TF	Time Frame delimiter
12	FErst	Front End reset
13	RT	Run Type; 1=Cont, 0=Trig
14	RS	Running State; 1=Running
...	...	Spare
27	LHCgap1	LHC abort gap 1
28	LHCgap2	LHC abort gap 2
29	TPCsync	TPC synchronisation/ITSrst
30	TPCrst	On request reset
31	TOF	TOF special trigger

GBT

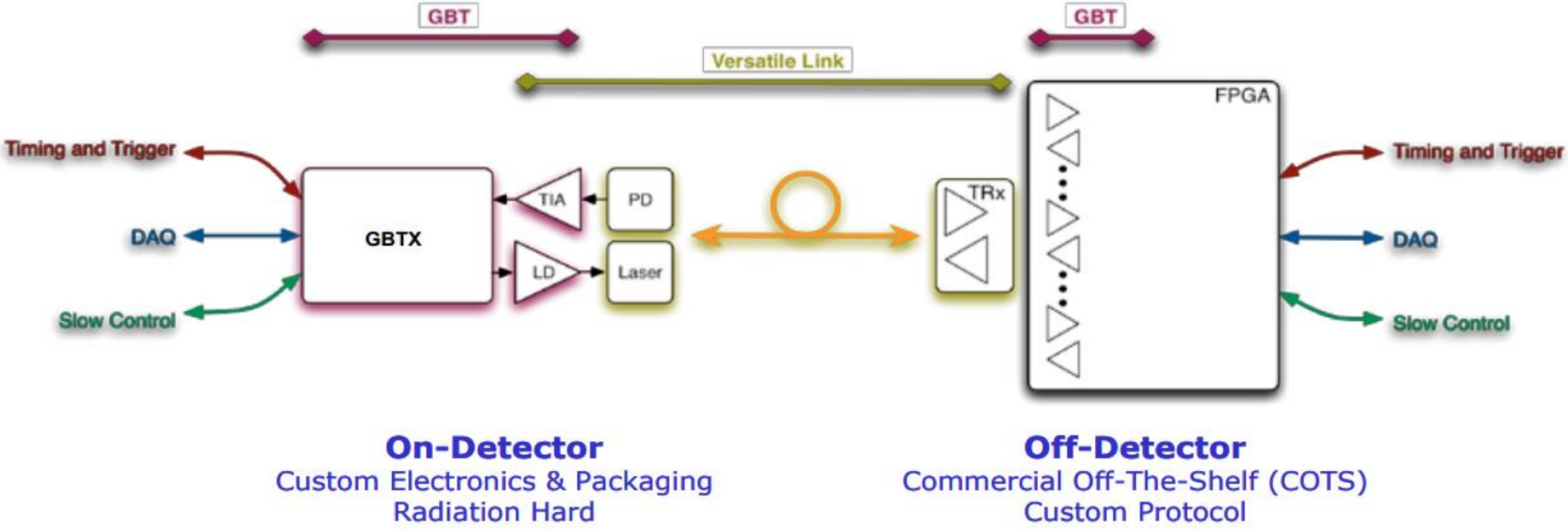


Figure 1: Link architecture with the GBT chip set and the Versatile Link opto-components.

GBT ASIC

- rad-hard chip (so it can sit on the FEE of a detector in the rad-zone)
- It provides different interfaces (e-link) to communicate with other chips installed in the FEE
- It provides dedicated communication to the SCA chip
- It recovers the clock and on the same link it transmits
  - Clock
  - DATA
  - SC

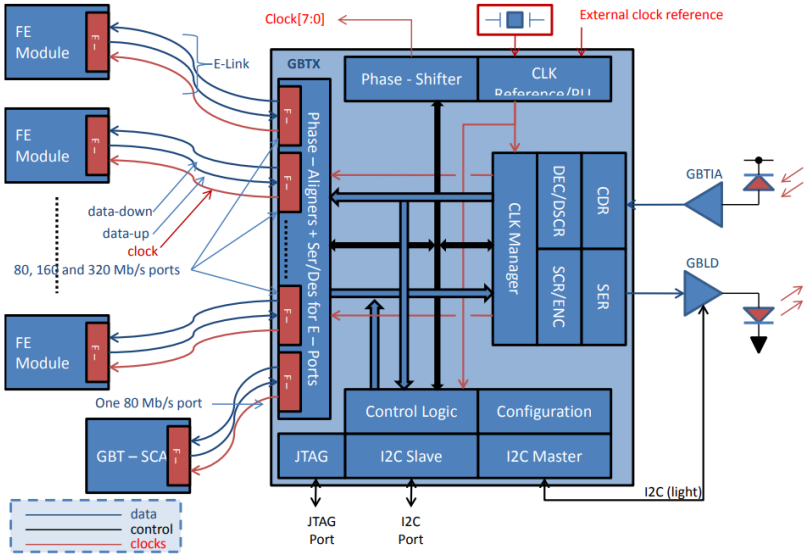


Figure 2 GBTX architecture and interfaces.

LINK RATE

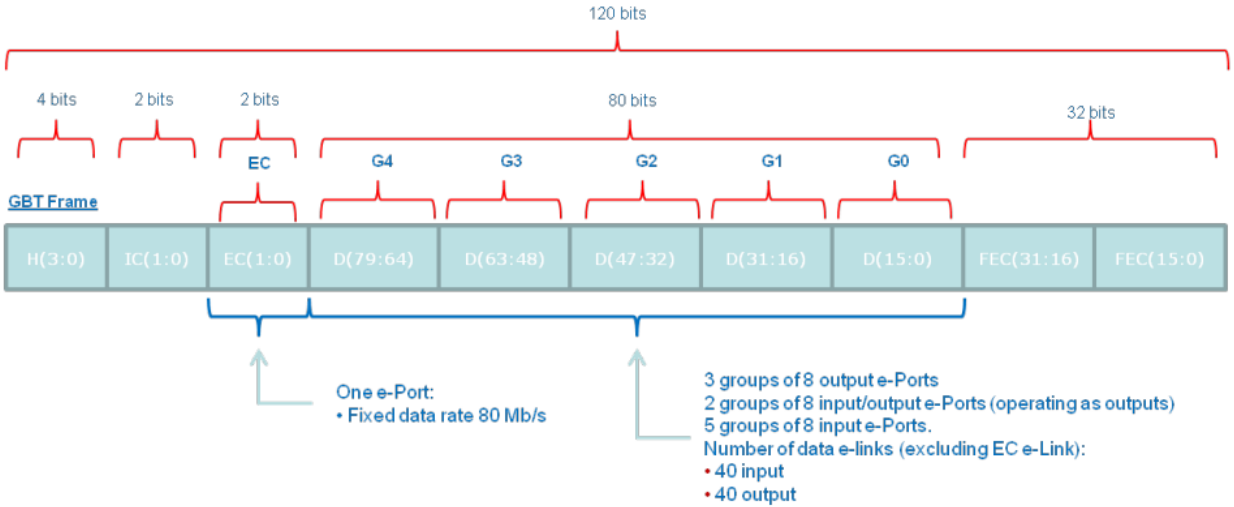


Figure 6 GBT frame structure

- The 120-bit “GBT frame format”, sketched in Figure 6, is transmitted during a single LHC bunch crossing interval (25 ns), resulting in a line rate of 4.8 Gb/s (120 bit * 40 MHz)
- The DATA field is 80 bit => 3.2 Gb/s
- The EXTENDED DATA field is 112 bit => 4.48 Gb/s
  - These 2 are the data rate of the GBT we use in ALICE

USABLE DATA RATE

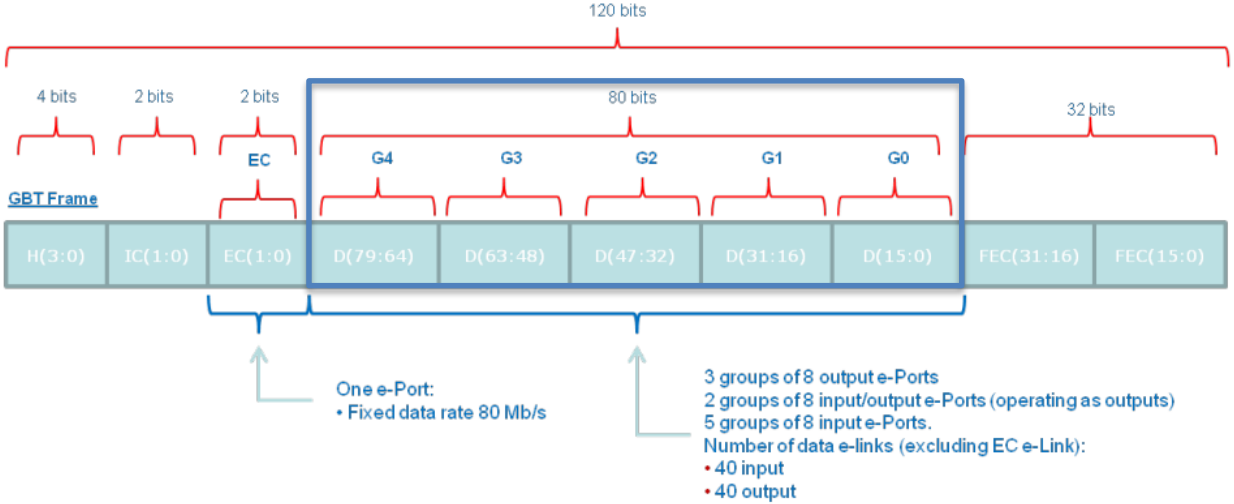


Figure 6 GBT frame structure

- The 120-bit “GBT frame format”, sketched in Figure 6, is transmitted during a single LHC bunch crossing interval (25 ns), resulting in a line rate of 4.8 Gb/s (120 bit * 40 MHz)
- **The DATA field is 80 bit => 3.2 Gb/s**
- The EXTENDED DATA field is 112 bit => 4.48 Gb/s
  - These 2 are the data rate of the GBT we use in ALICE

USABLE DATA RATE

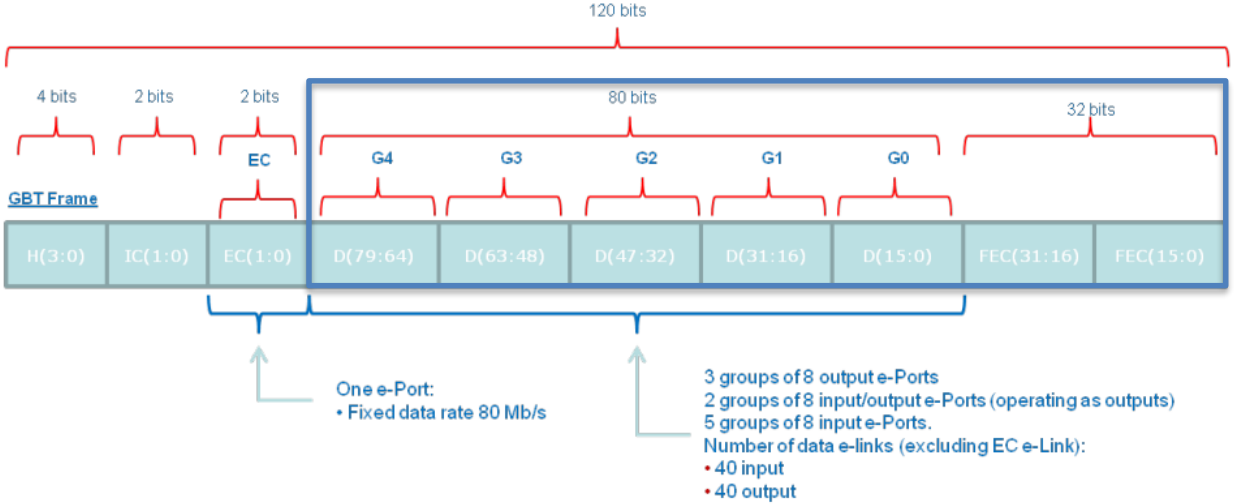
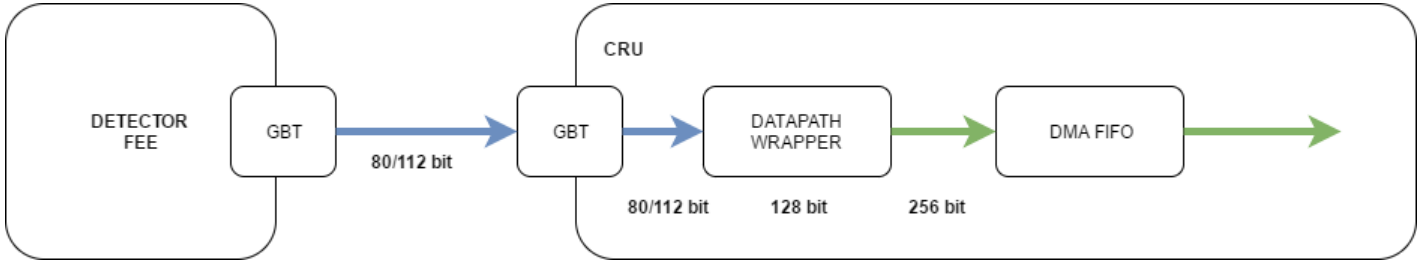


Figure 6 GBT frame structure

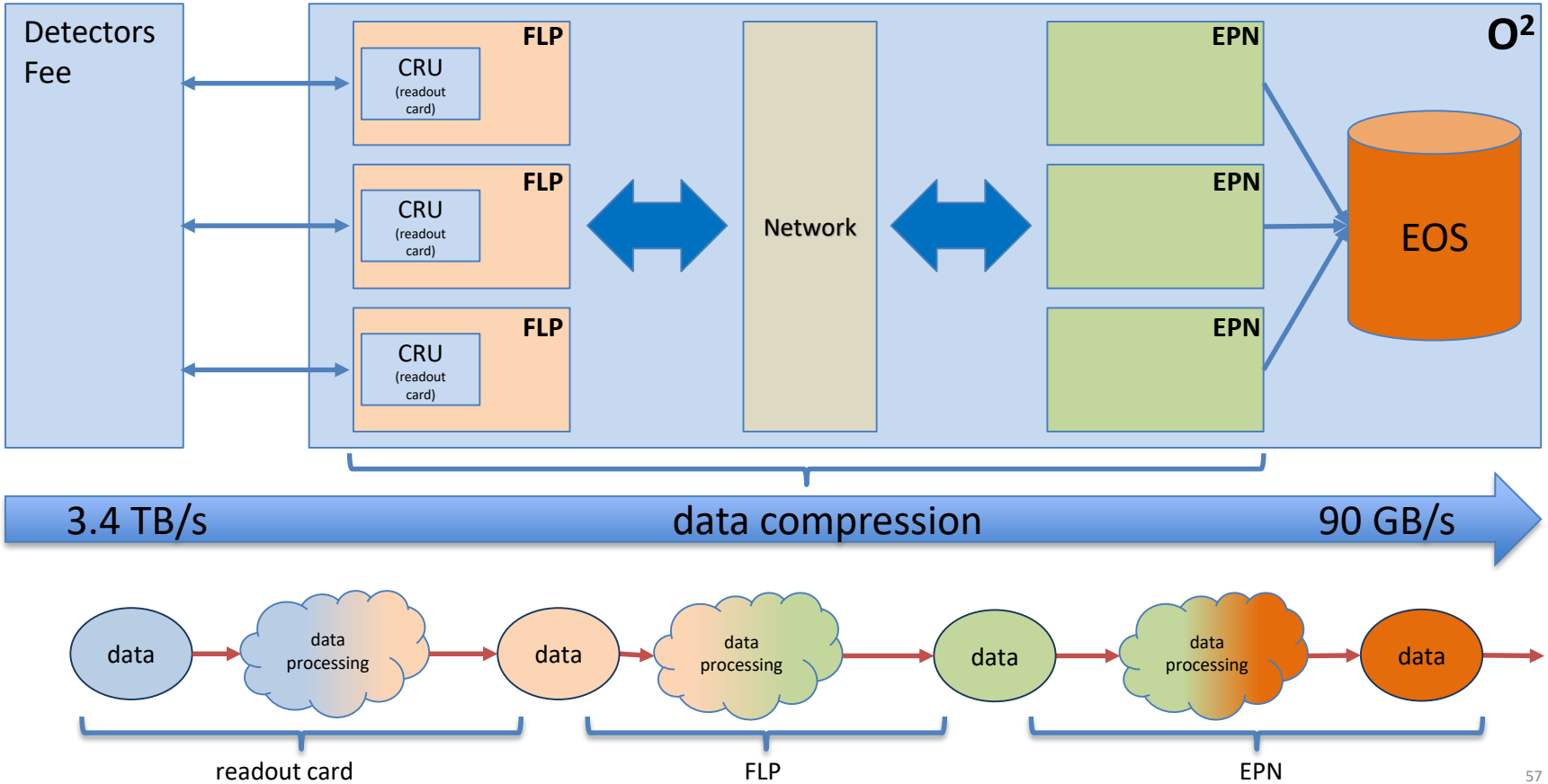
- The 120-bit “GBT frame format”, sketched in Figure 6, is transmitted during a single LHC bunch crossing interval (25 ns), resulting in a line rate of 4.8 Gb/s (120 bit * 40 MHz)
- The DATA field is 80 bit => 3.2 Gb/s
- **The EXTENDED DATA field is 112 bit => 4.48 Gb/s**
  - These 2 are the data rate of the GBT we use in ALICE

DATA RATE inflation (from 80 to 128 bit)



- To work with 80 or 112 bit is not ideal when moving words that should be a multiple of 32 bit.
- DATA word is inflated to 128 bit when the information enters in the CRU for better handling the data format of all the detectors.

DATA FLOW in O²



HB TRIGGER and TIME FRAME

Heart Beat (HB)

issued in continuous &
triggered modes
to all detectors

LHC clock 40 MHz
3564 Bunch Crossing in 1 ORBIT
ORBIT rate ~10 KHz
Time Frame = 128 Orbits

Continuous read-out