The KM3NeT data acquisition system

Status and evolution

Tommaso Chiarusi*

Emidio Giorgio, Daniele Zito

On behalf of the KM3NeT Collaboration

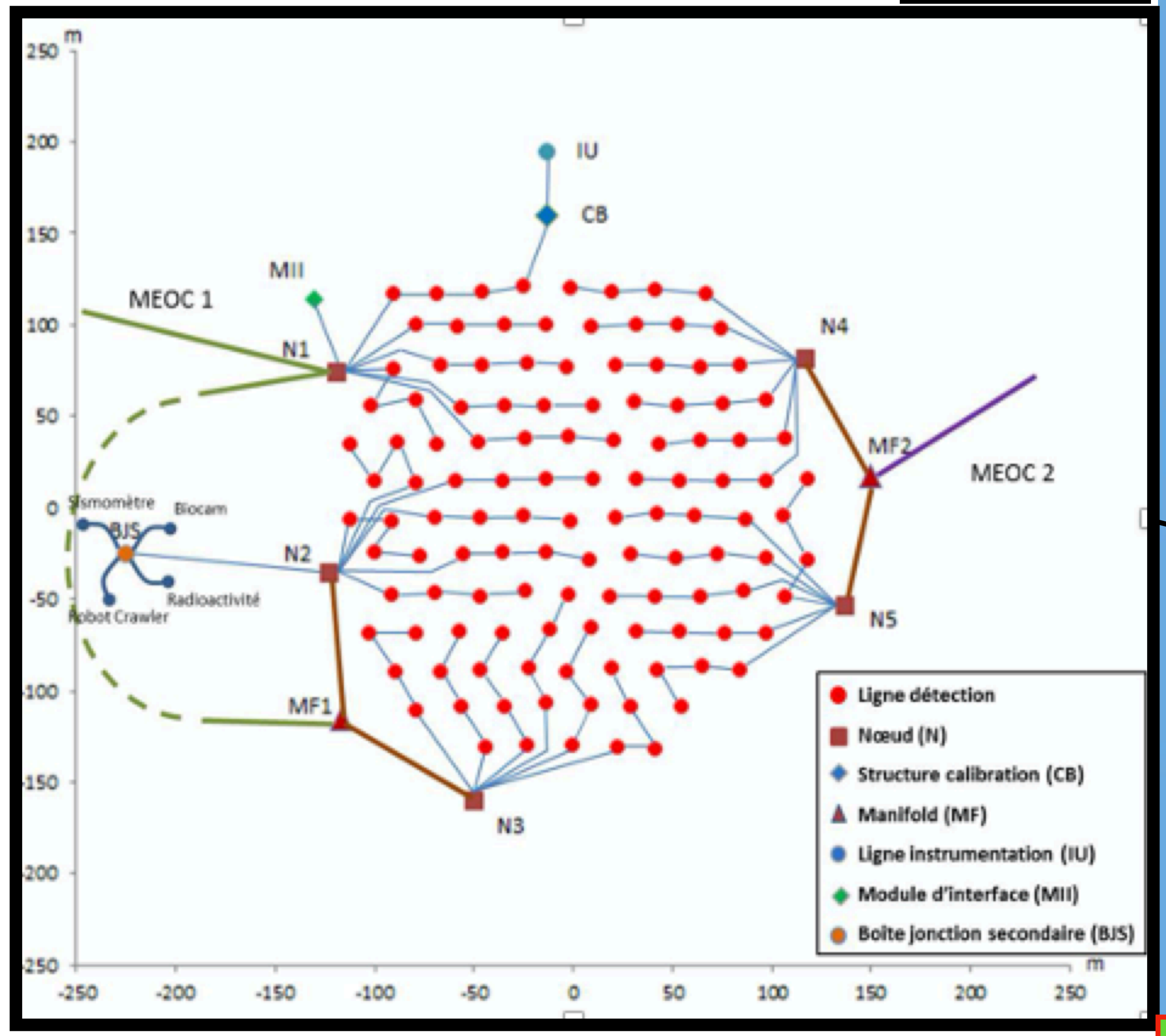
INFN

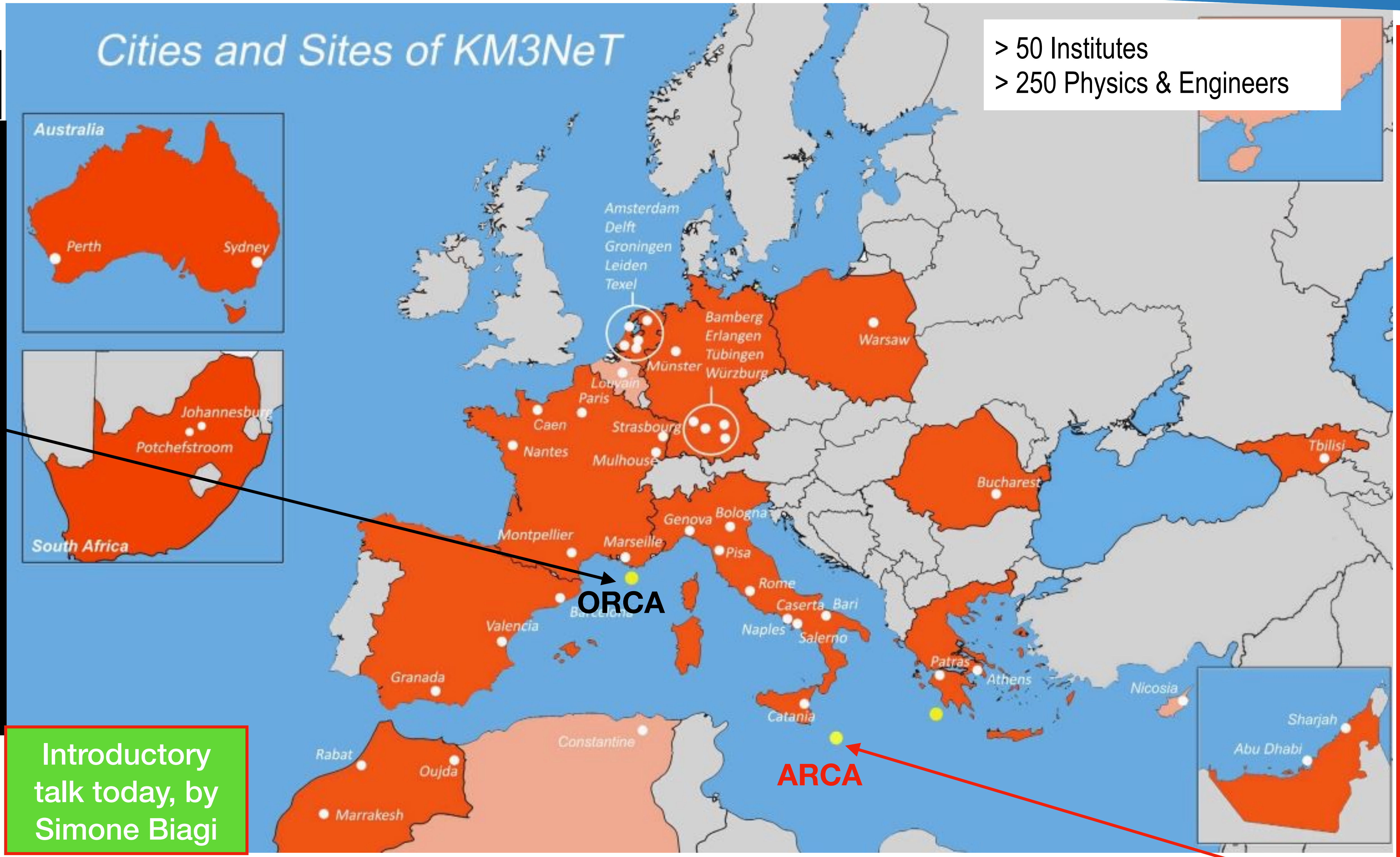
Sezione di Bologna

Laboratori Nazionali del Sud

KM3NeT

ORCA

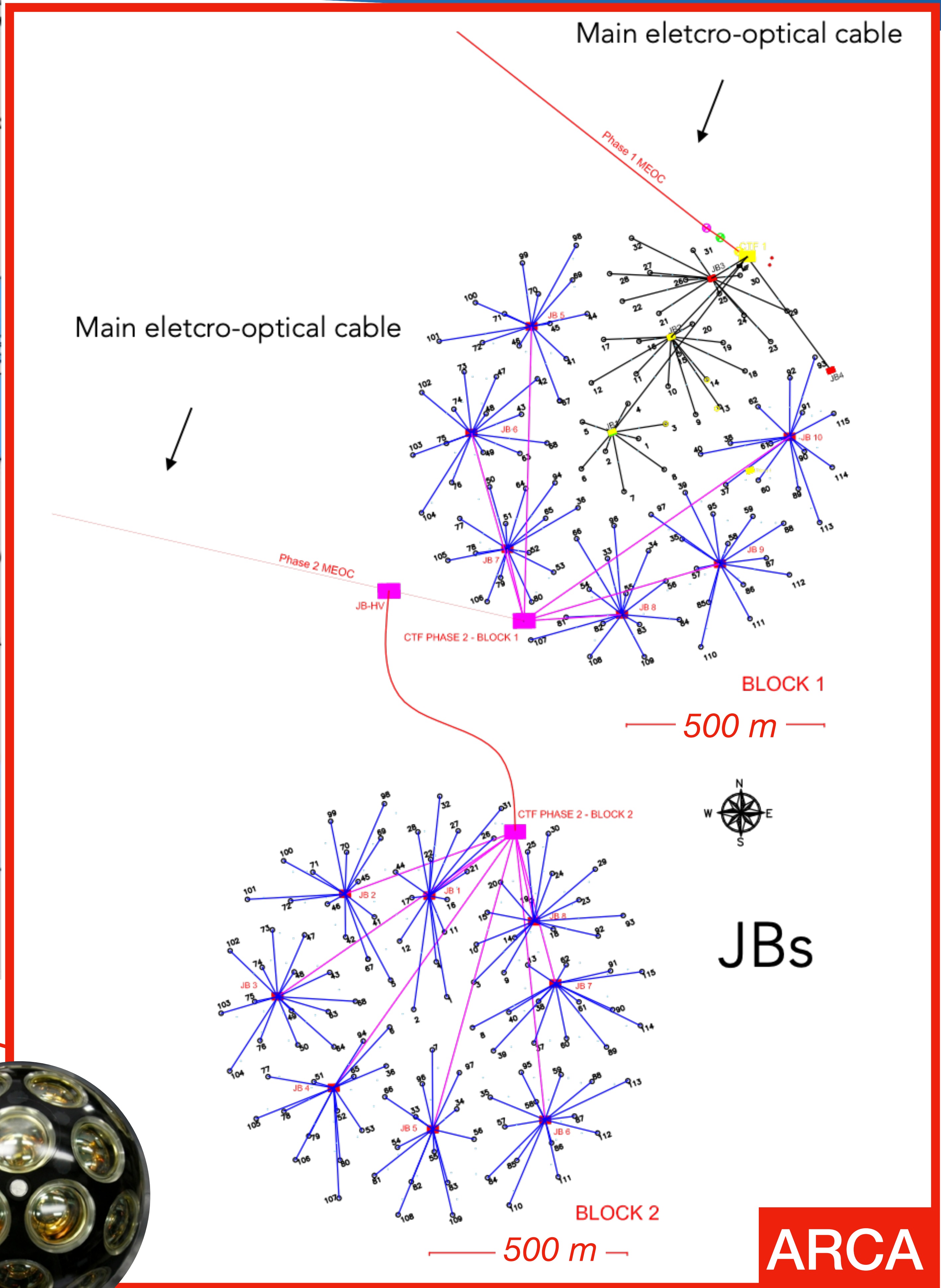




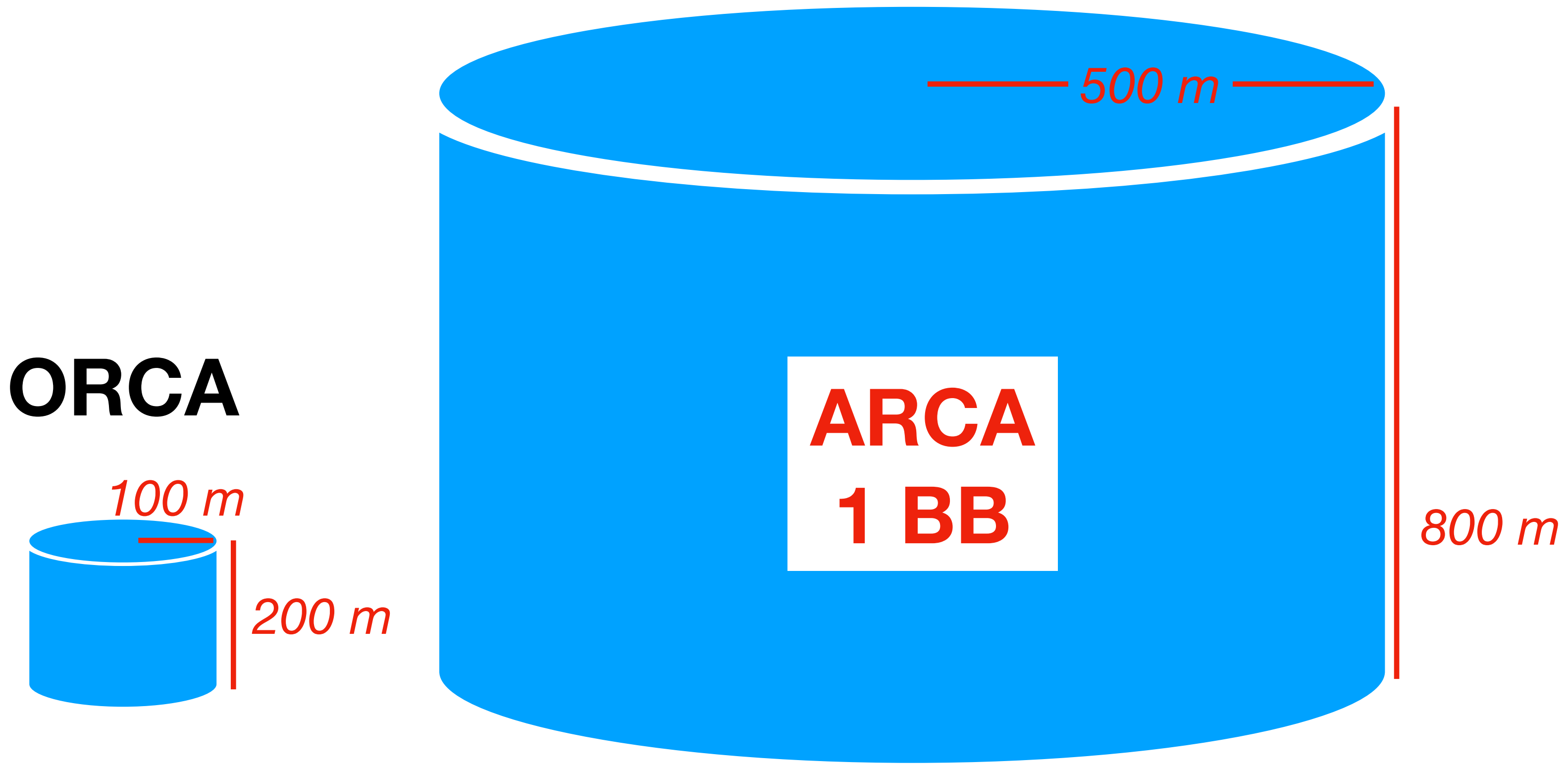
Introductory talk today, by Simone Biagi

Two detectors, same technology, different layout and physics objectives

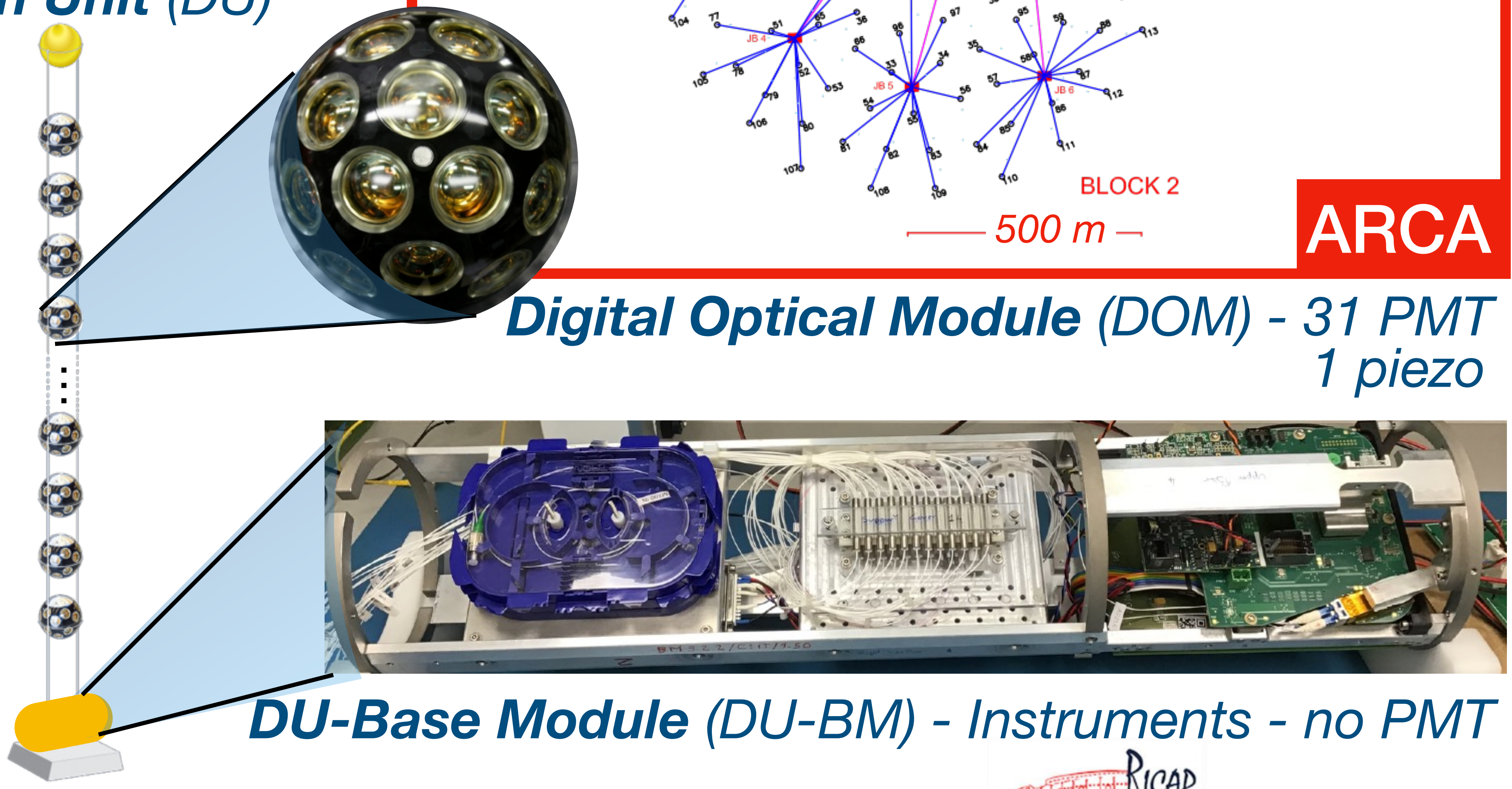
See Simone Biagi's talk



	ARCA	ORCA
Location	Italy	France
N. building blocks	2	1
N. DU per b.b.	115	115
DU distance	90 m	20 m
DOM spacing	36 m	9 m
DU height	~ 800 m	~ 200 m
Instrumented mass (Mton)	2*650	7
Depth	3500 m	2500 m



***Detection Unit** (DU)*



***Digital Optical Module** (DOM) - 31 PMT 1 piezo*

***DU-Base Module** (DU-BM) - Instruments - no PMT*

KM3NeT

- big volumes
- water optical properties (absorption & scattering of blue-green photons ~ 50-100 m)
- good angular resolution O(.1°) for sky pointing (that's neutrino ASTRONOMY)

➠ Many detection elements (N. OMs > O(1000)/km^3) deployed in bunches

➠ **SCALABLE DAQ design**

- No "beam crossing" reference such as for experiments at Colliders
- complex DAQ structures in extreme conditions (mandatory: minimal underwater complexity)

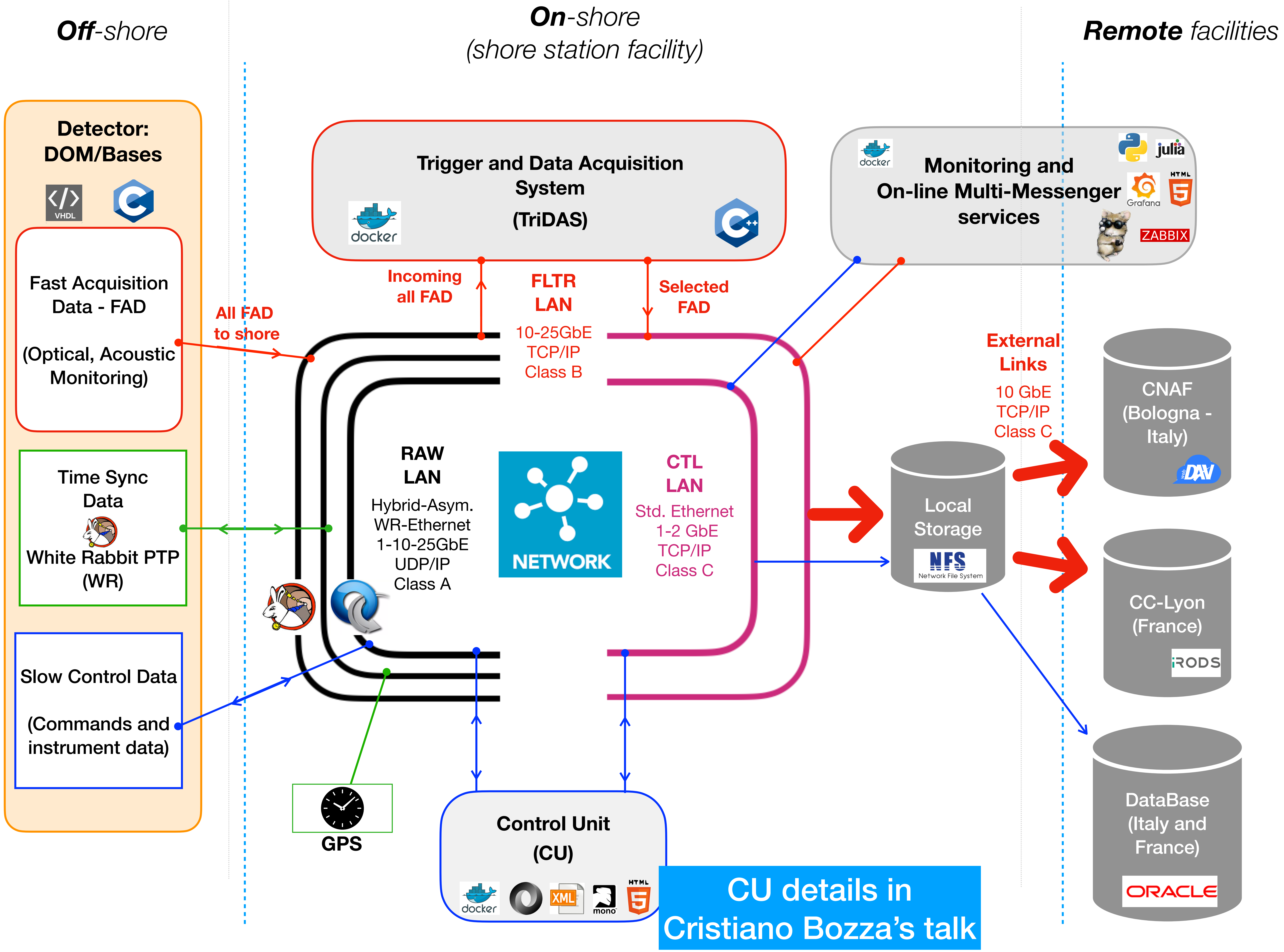
➠ **ALL DATA TO SHORE (a.k.a. *trigger-less streaming readout*) approach**

DRAWBACKS

signal-to-noise ratio extremely disfavoured :

muon rate (atmospheric dominating)	:	O(100) Hz/km^3
^{40}K decays (~constant)	:	O(10) kHz/PMT(3", 0.5 p.e. threshold)
Bioluminescence (occasional)	:	O(100) kHz/PMT(3", 0.5 p.e. threshold)

➠ High continuous throughput to shore, needed **large bandwidth switching infrastructure and a strong data reduction**



White Rabbit

Openflow

ORACLE

HTML

XML

JSON

mono

VHDL

docker

python

julia

ANSIBLE

Grafana

ZABBIX

Observium

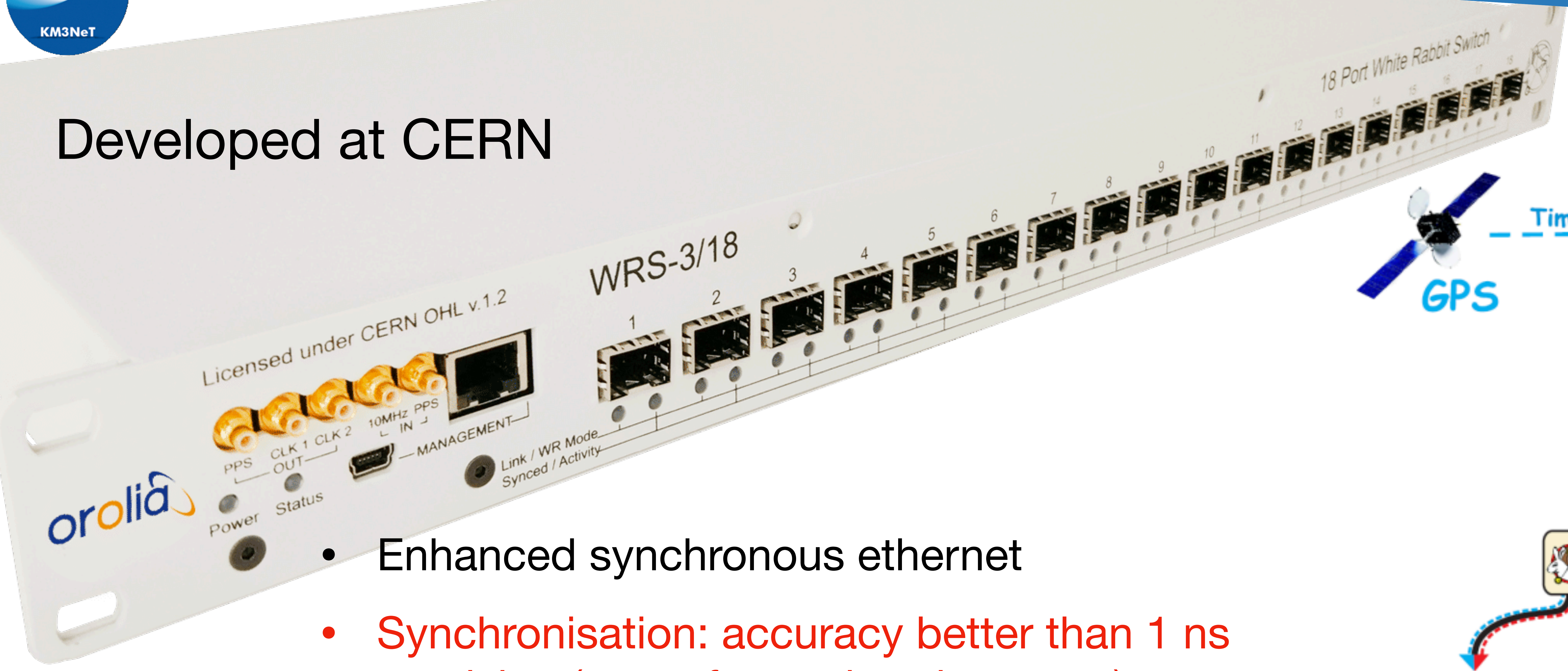
NFS
Network File Systen

CentOS

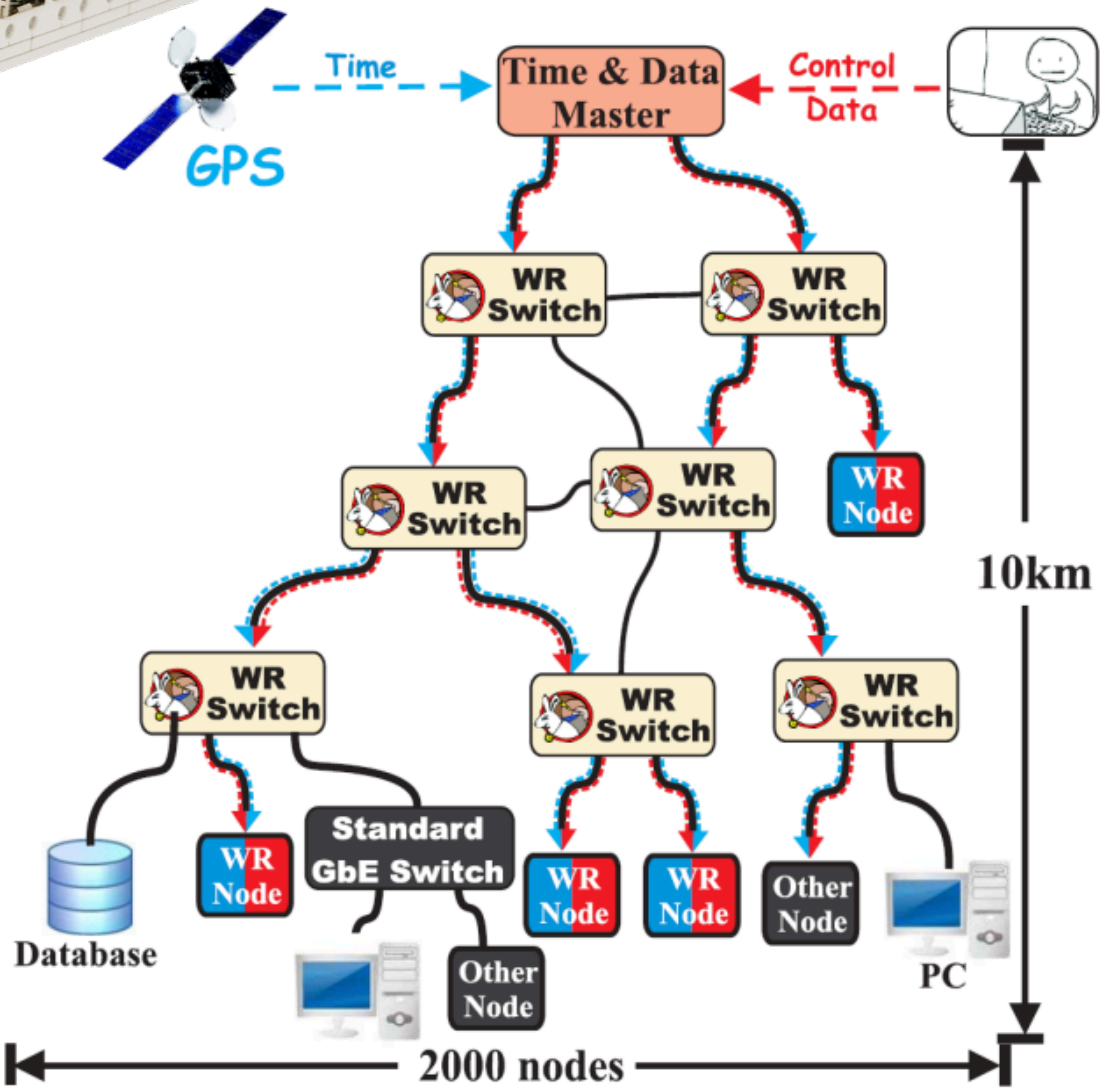
iRODS

KM3NeT

Developed at CERN

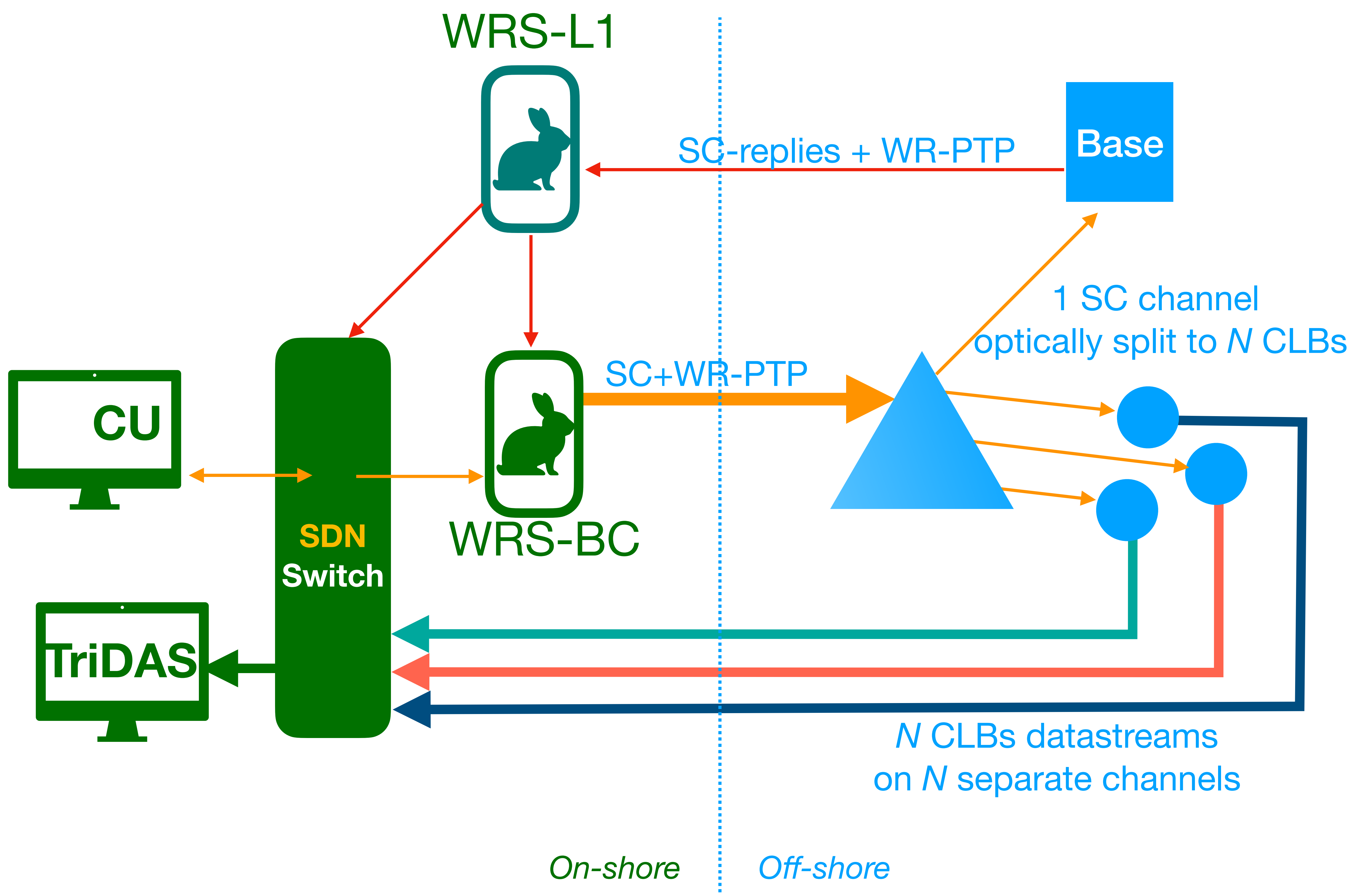


- Enhanced synchronous ethernet
- Synchronisation: accuracy better than 1 ns precision (tens of ps stdev skew max)
- Deterministic reliable and low latency control-data delivery



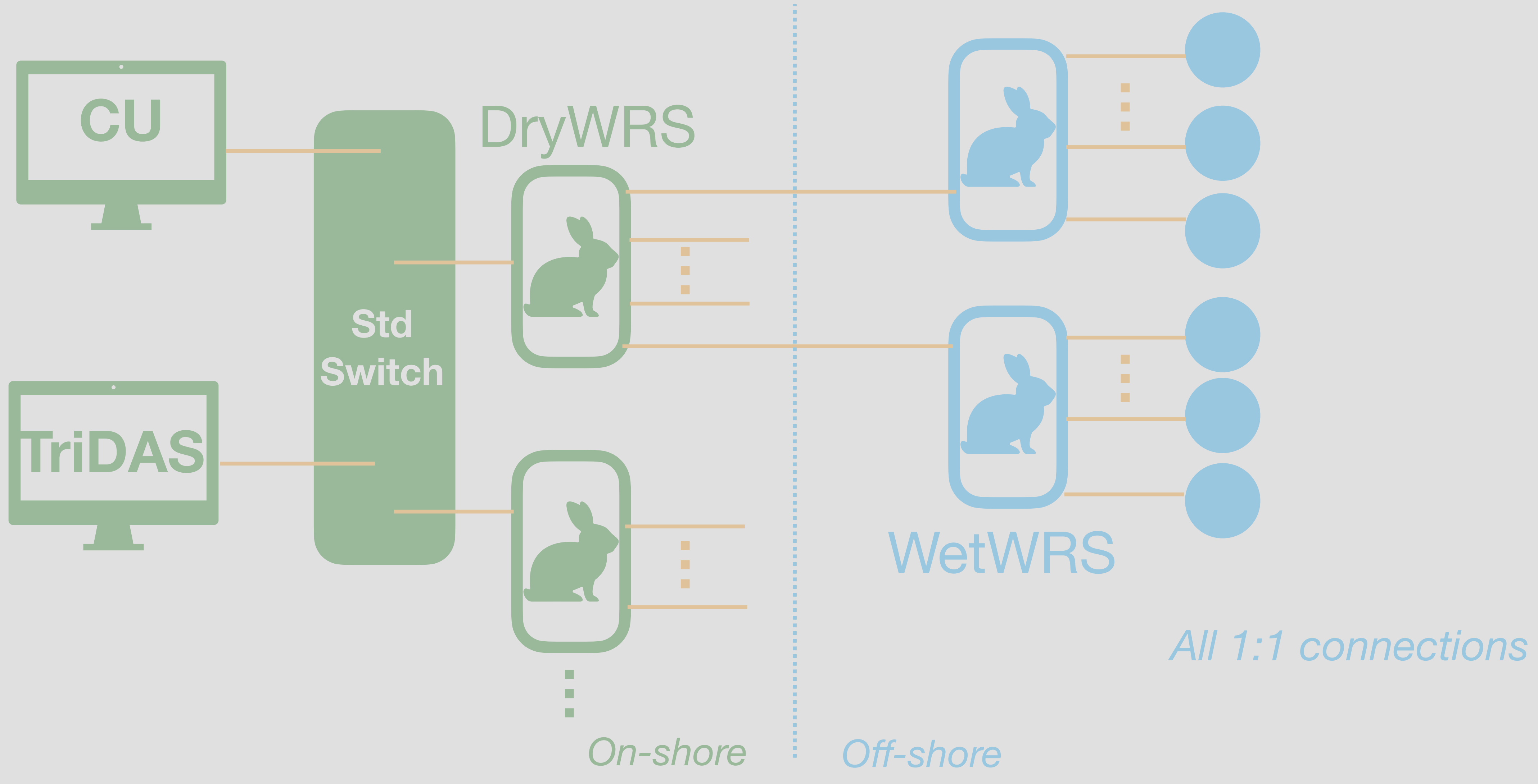
Currently used releases in KM3NeT DAQ:
Hardware: **WRS-18p-hw-v3.4**
Firmware: WR-Core **v4.2** (**customised** by Seven Solutions for KM3NeT); **v5.0.1**
Ongoing evaluation of **v6**.x

Broadcast (ARCA 32 strings; ORCA 48 strings *at least*)



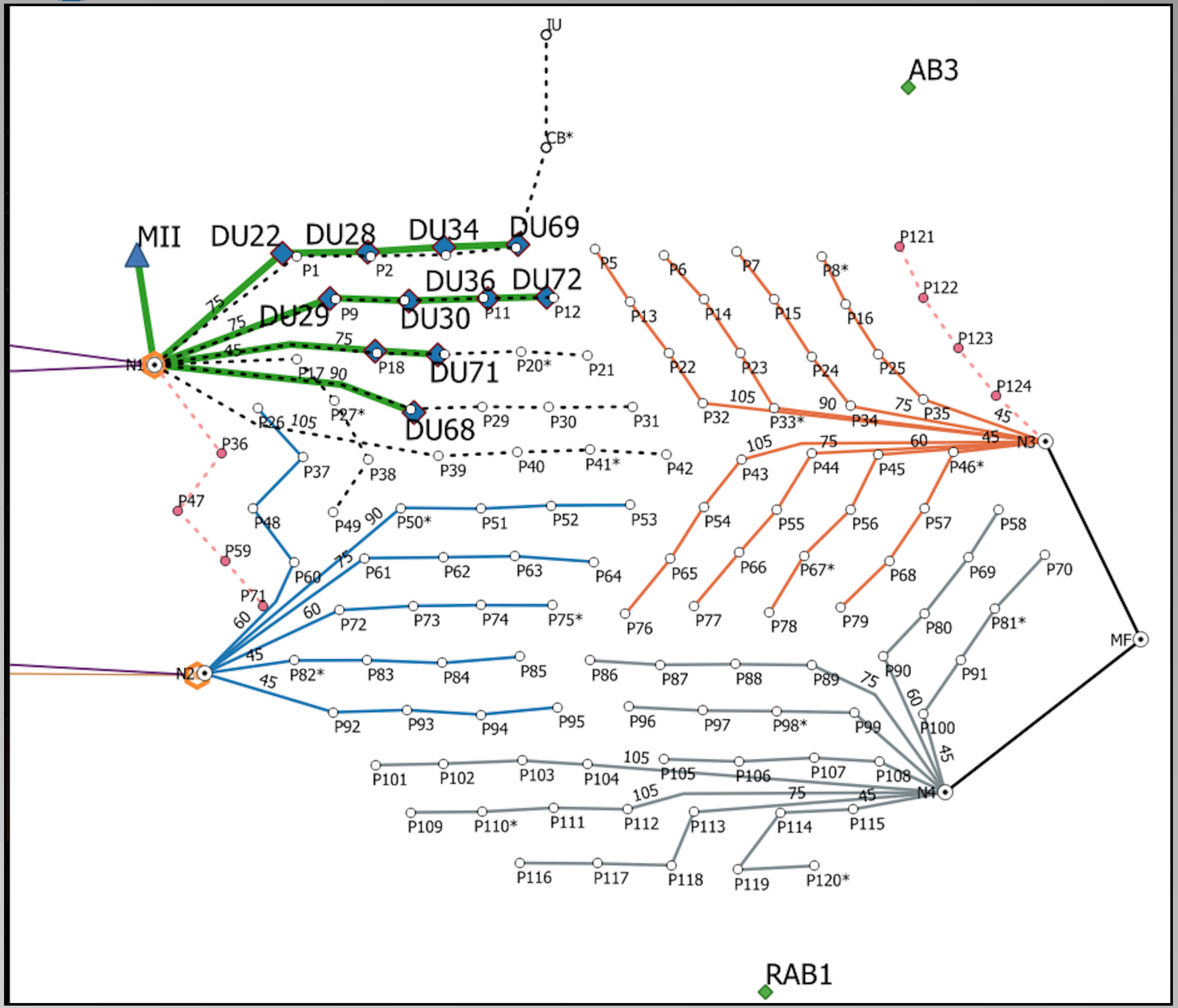
Current implementation in both ORCA/ARCA
(as well as other test-installations)

Full White Rabbit (necessary for ARCA 2 BB)

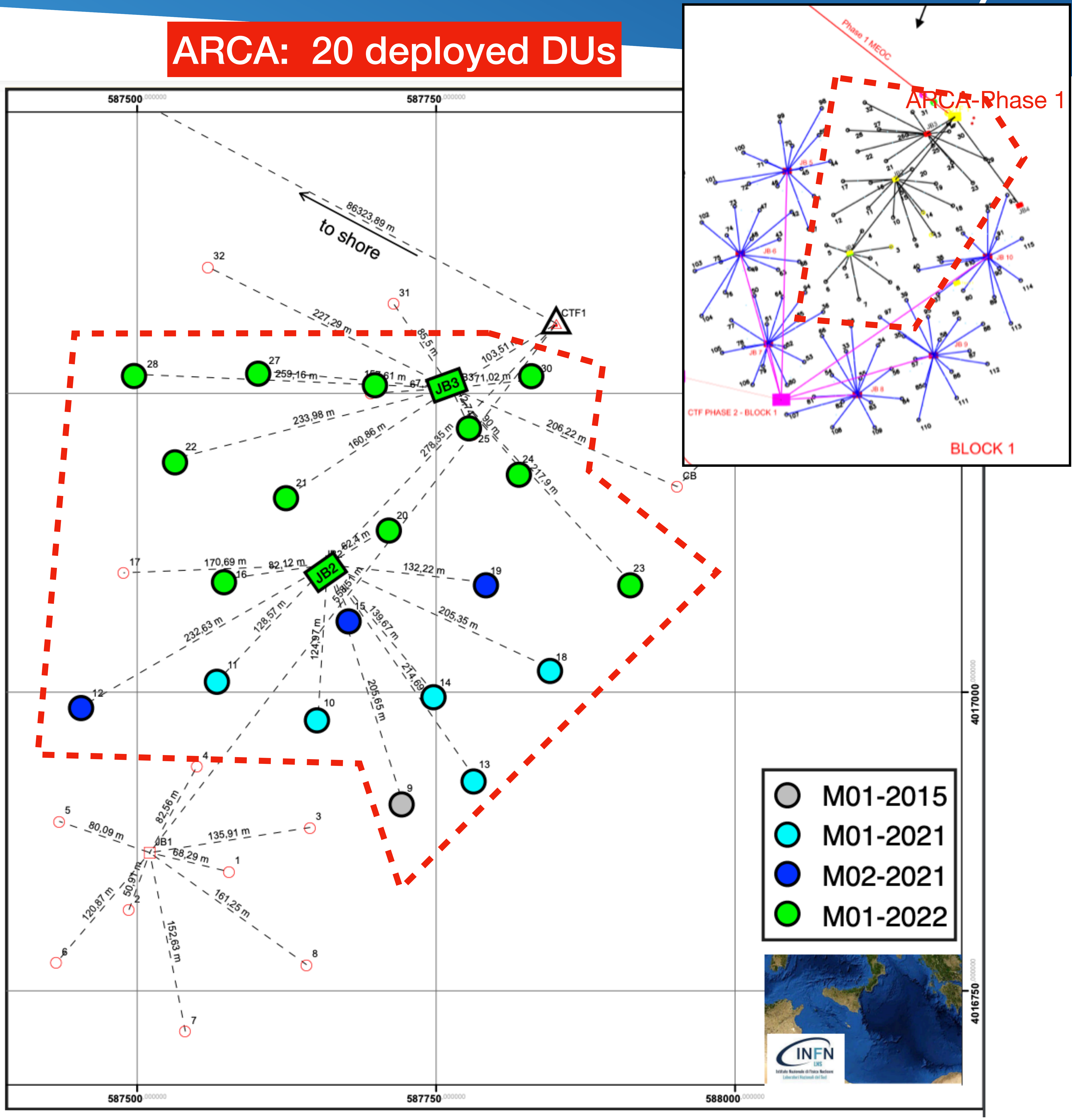


Future evolutions

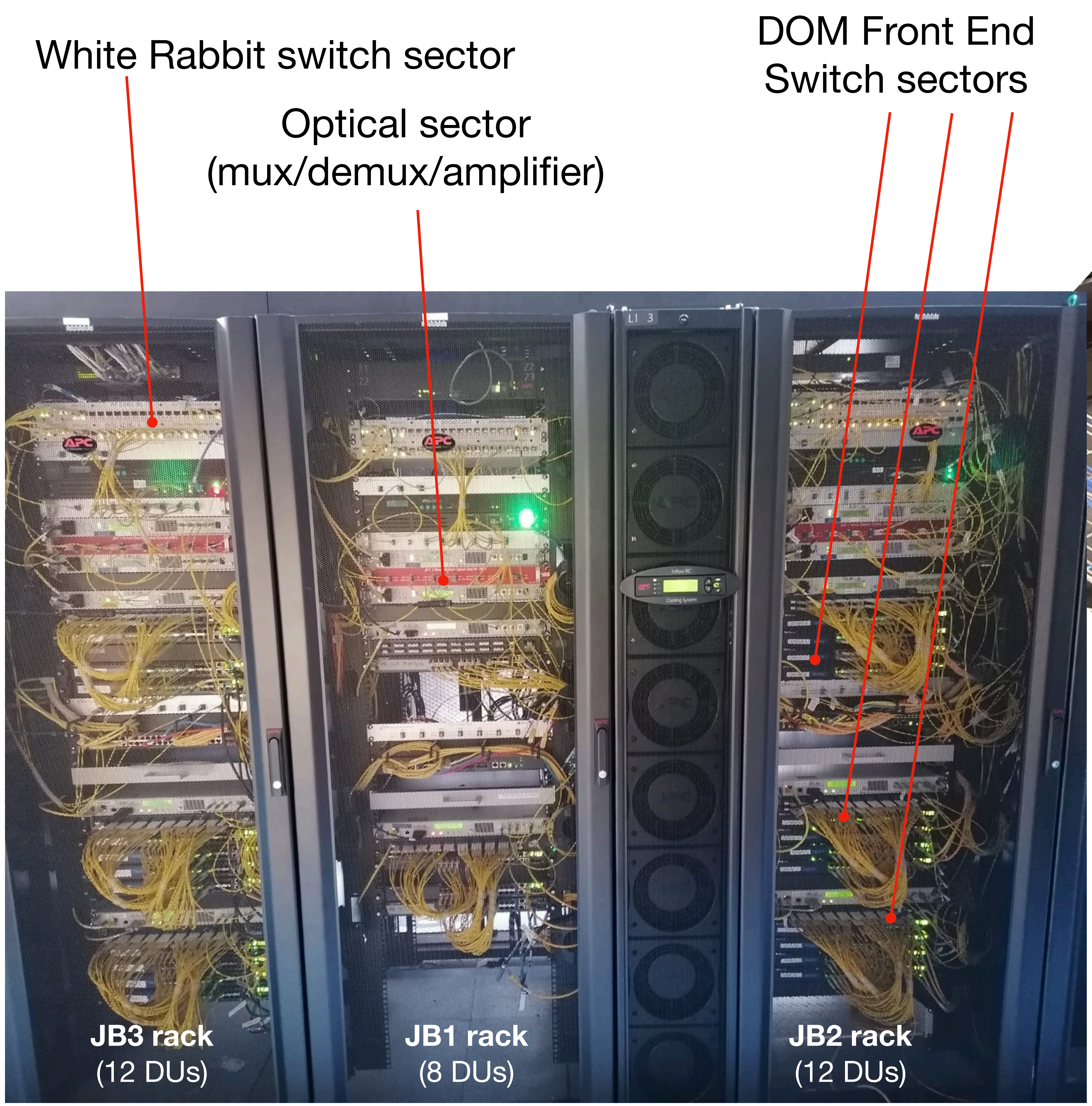
ORCA: 11 deployed DUs

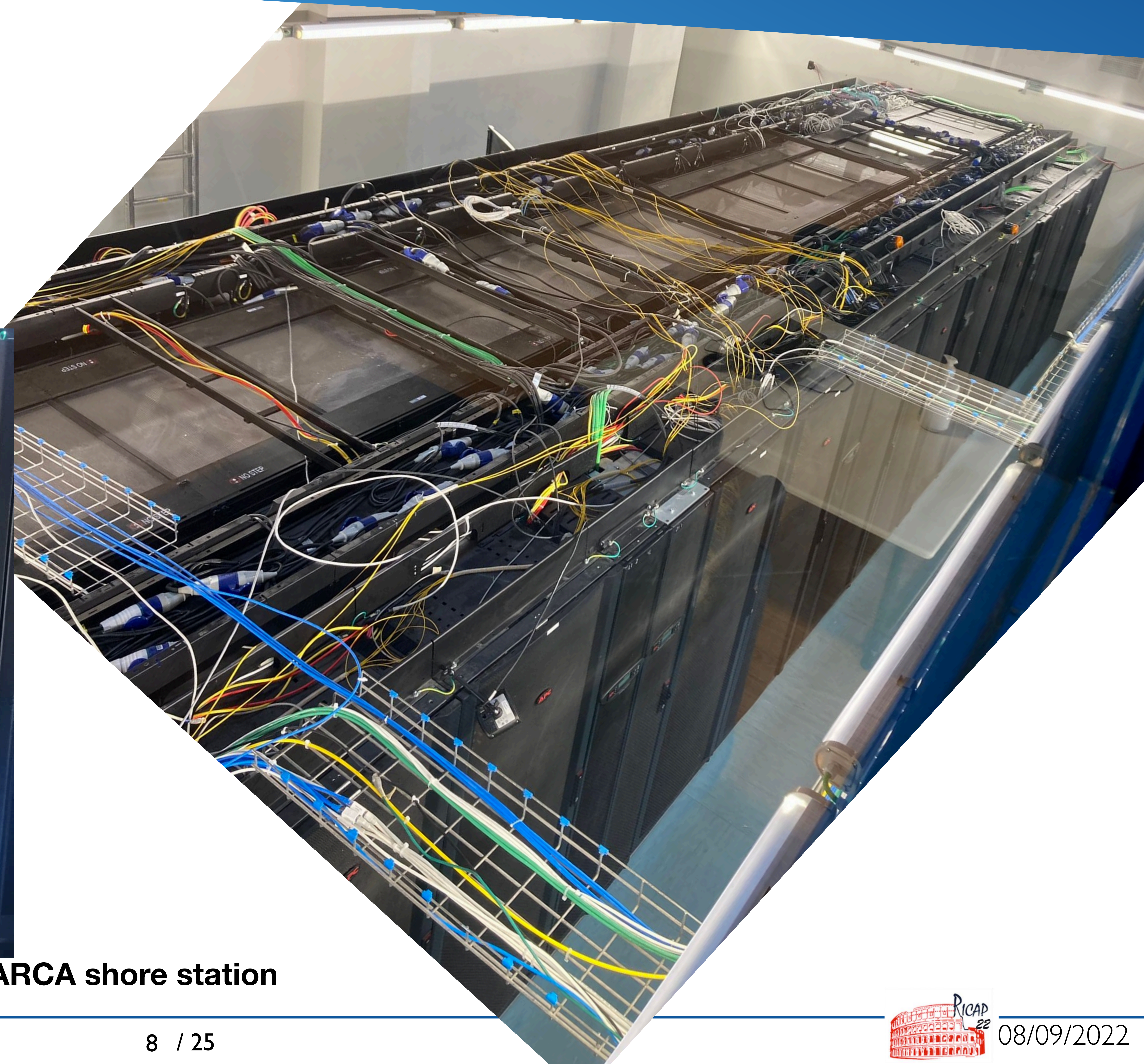


ARCA: 20 deployed DUs

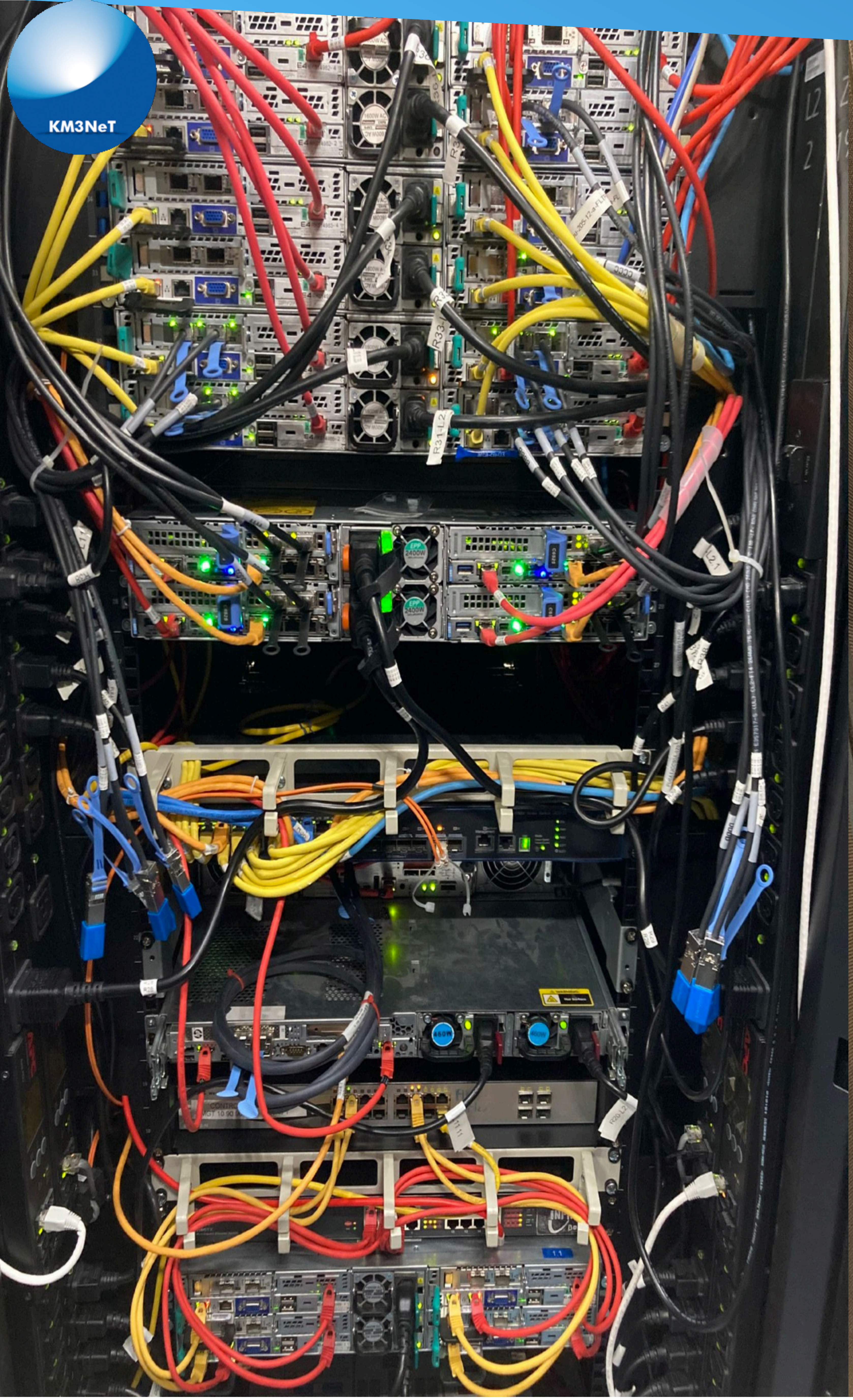


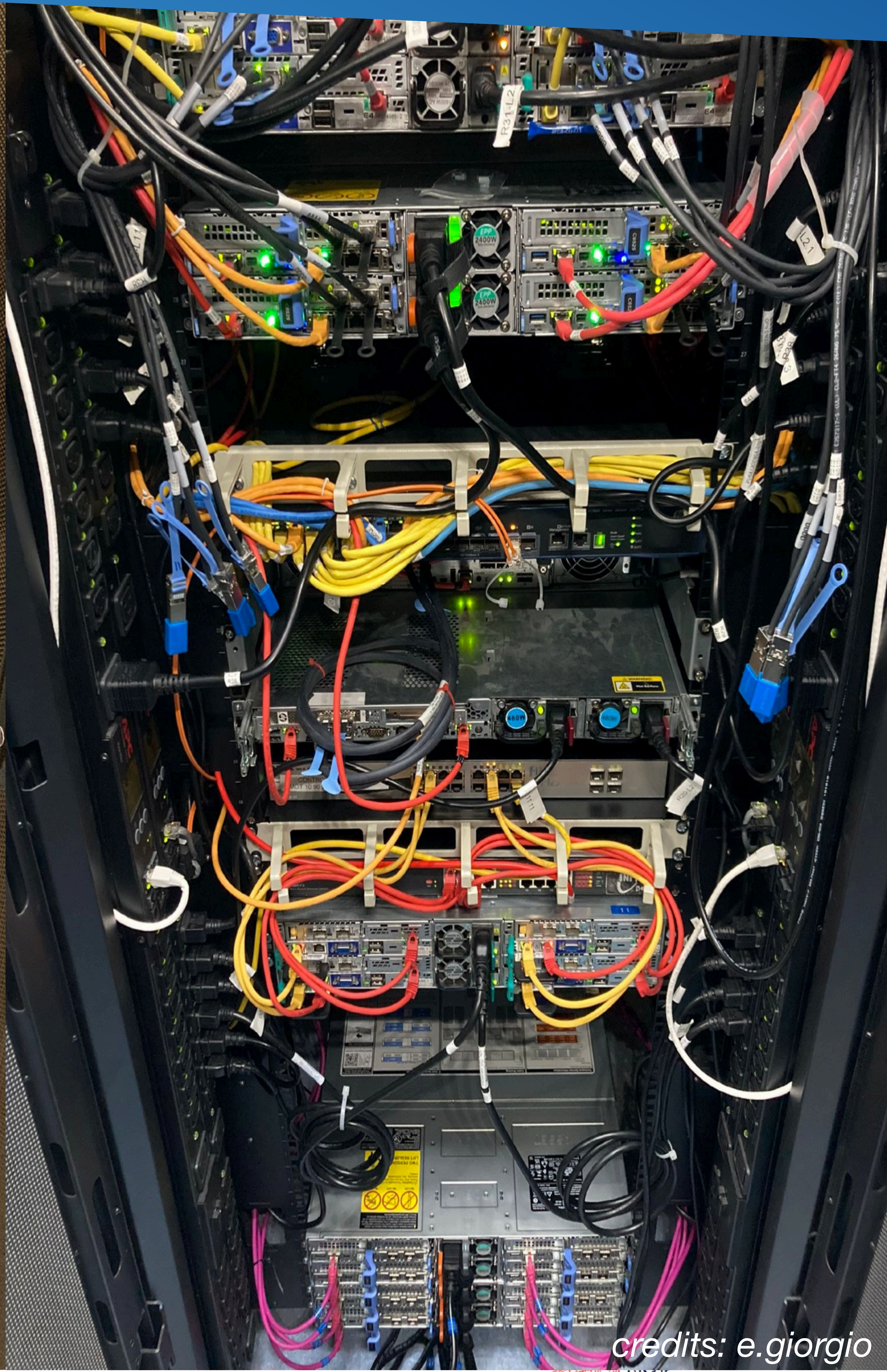
In these days of September,
new deployments for both ARCA and ORCA.
More data sources soon!



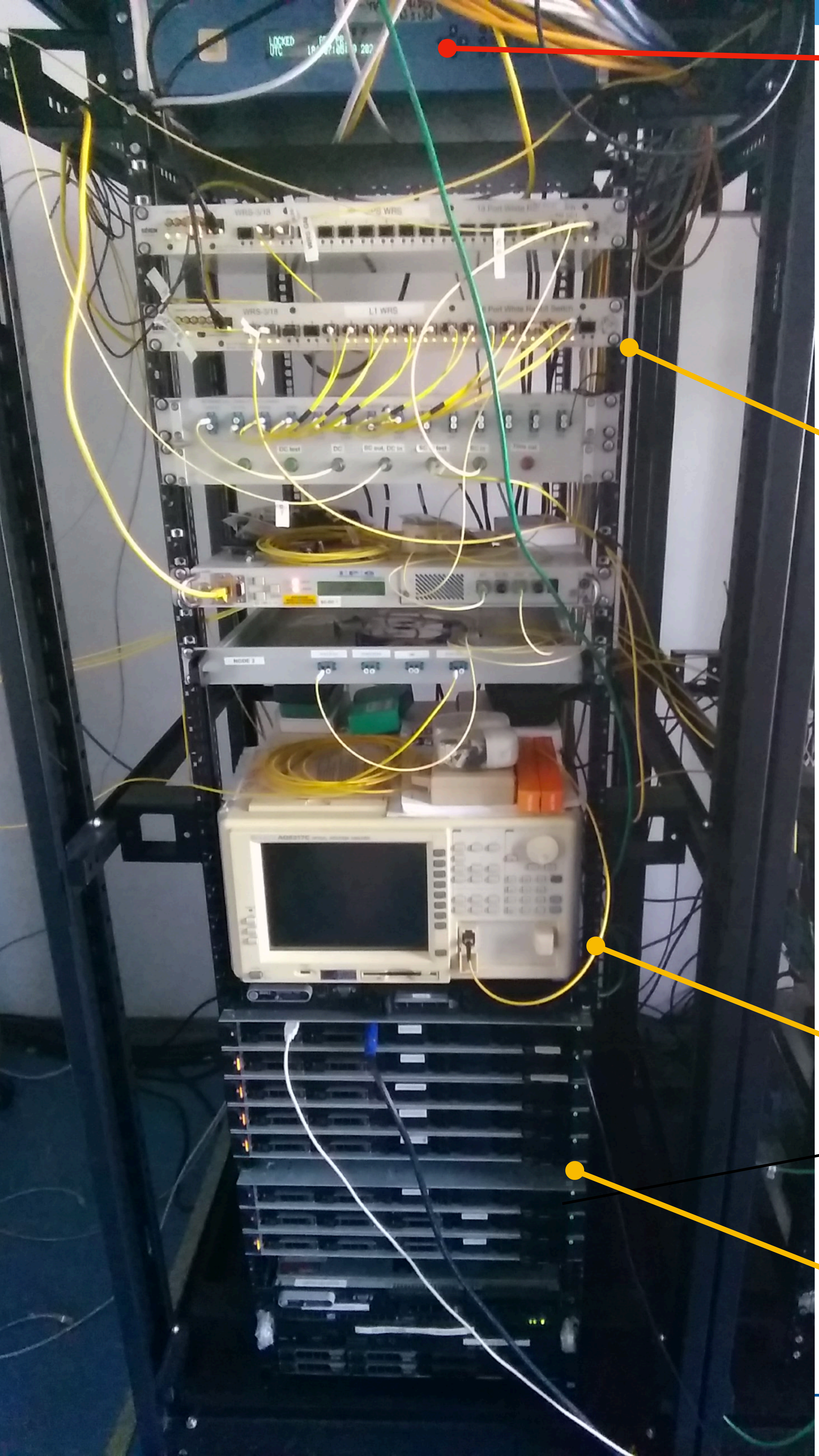


ARCA shore station





credits: e.giorgio



GPS

SCSF and SCBD
SDN switch fabric

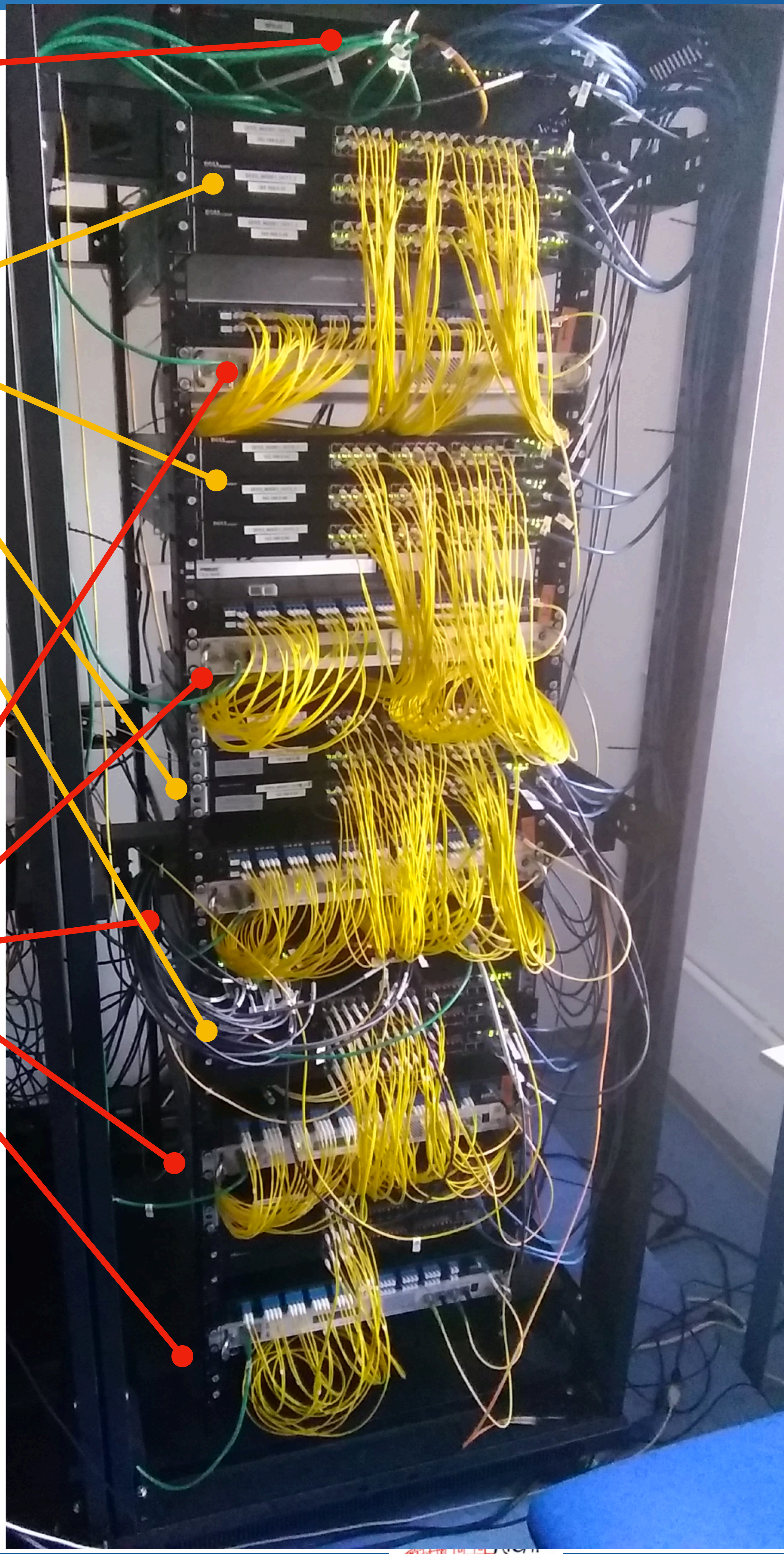
DOM Front-end Sectors

White Rabbit fabric

Demultiplexing stages for the
incoming optical signals

OSA (optical spectral analysis)

Computing servers



The ***Broadcast* WRS customisation for KM3NeT** differs from the standard point-to-point (Master-Slave) WR protocol connections.

The **Software Defined Networking** (SDN) solves the scalability issue of this asymmetric scenario with a hybrid shore-station in the Broadcast scenario.

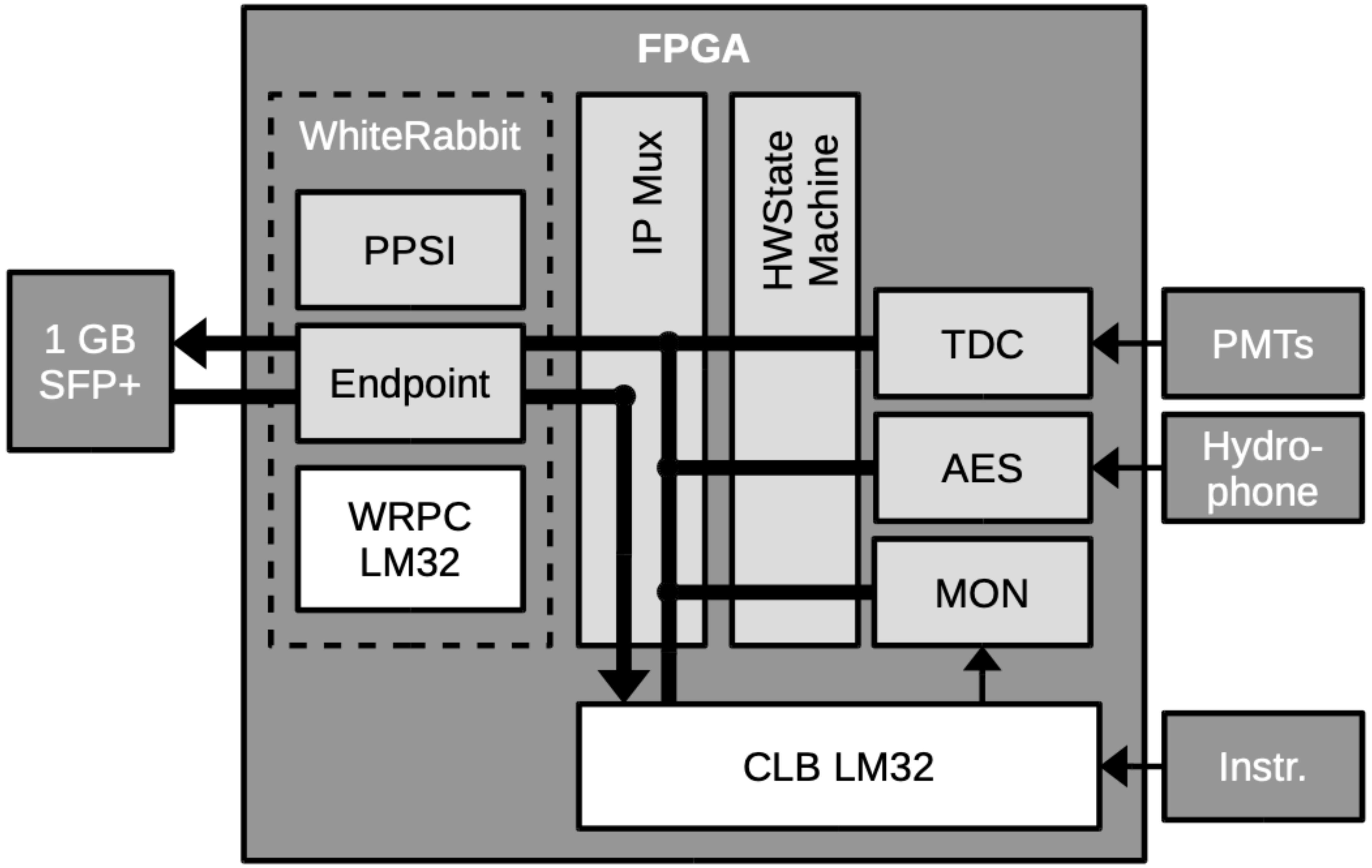
CLB FIRMWARE ARCHITECTURE

Two LM32 cores

- WhiteRabbit LM32 for timing control
- KM3NeT CLB for DAQ control / instrumentation readout

Three DAQ modules

- Time to Digital Converter (TDC) – from Photo Multiplier Tubes (PMTs)
  AES-standard receiver - from Hydrophone
- MONitoring, for performance information



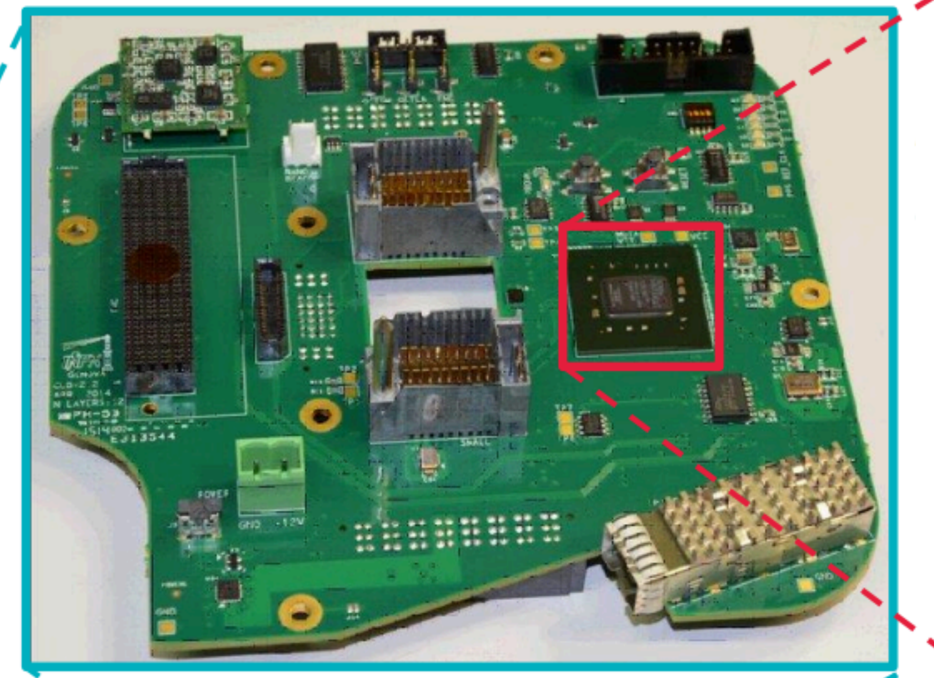
High-level diagram of CLB gateware and network data-path

Network path

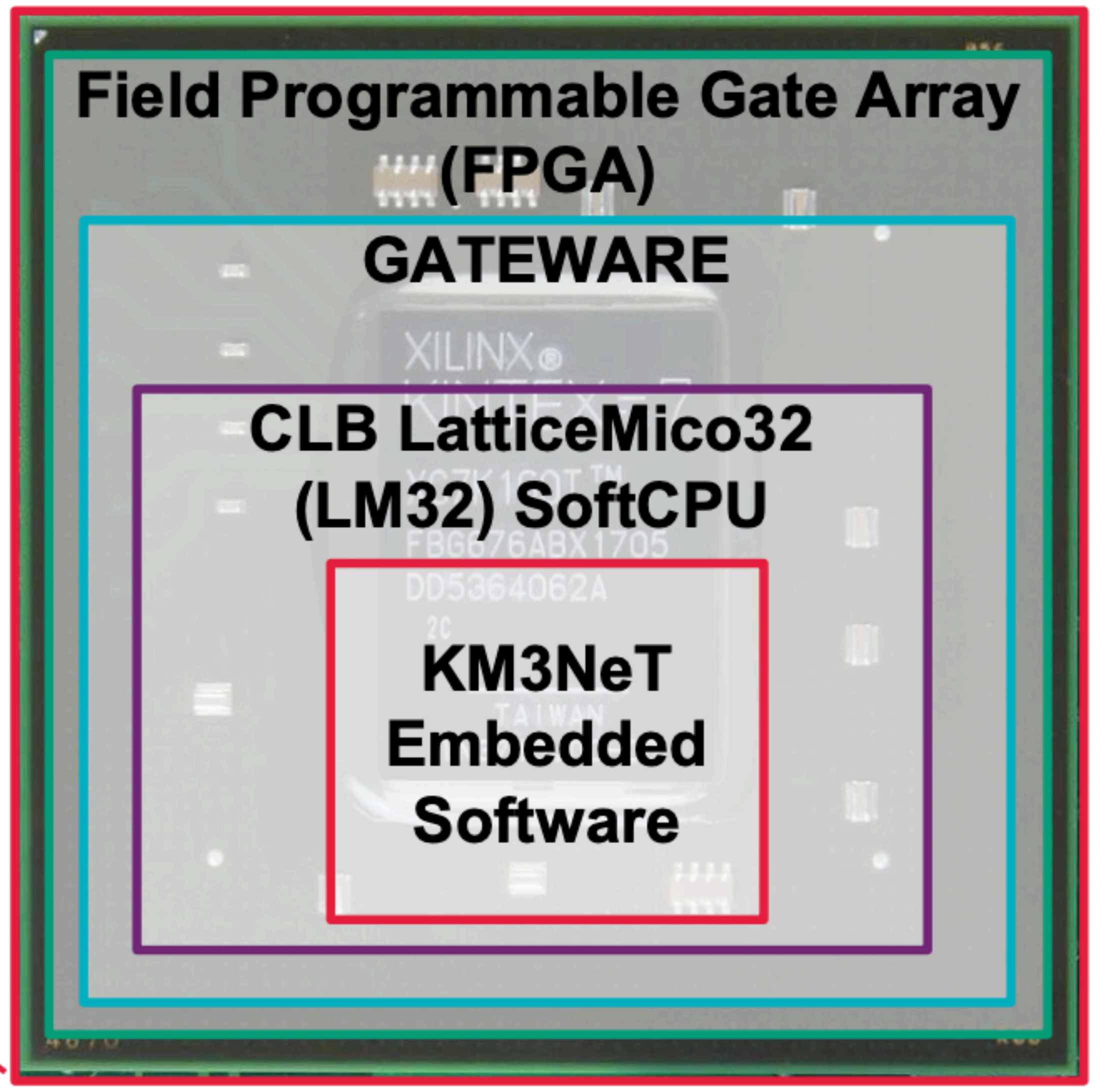
- WhiteRabbit is used for timing and intercepts and transmits timing related Ethernet packets. The remaining data is sent over IPMux to the CLB LM32
- DAQ modules generate data, subsequently annotated and framed by the HWStateMachine, wrapped as UDP packets and dispatched by the IPMux

EMBEDDED SOFTWARE IN KM3NET DETECTOR

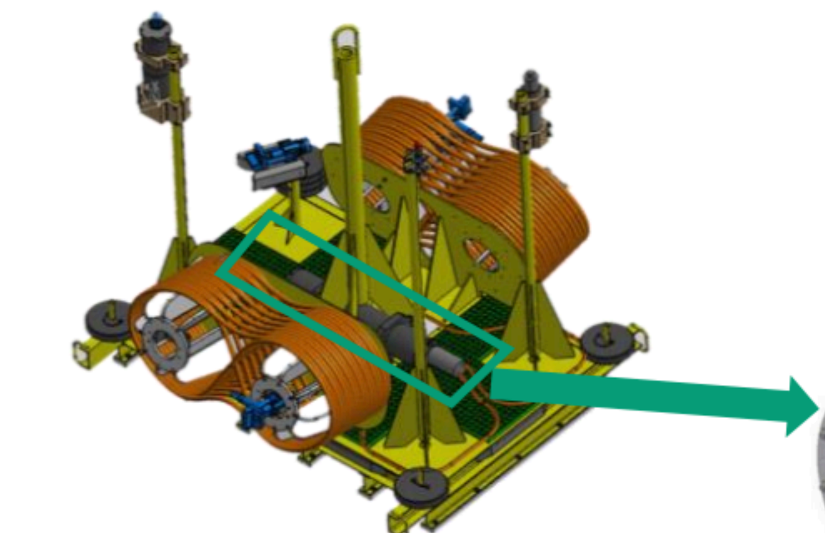
Three different detector modules feature a Central Logic Board (CLB) running the KM3NeT Embedded Software inside a SoftCPU*



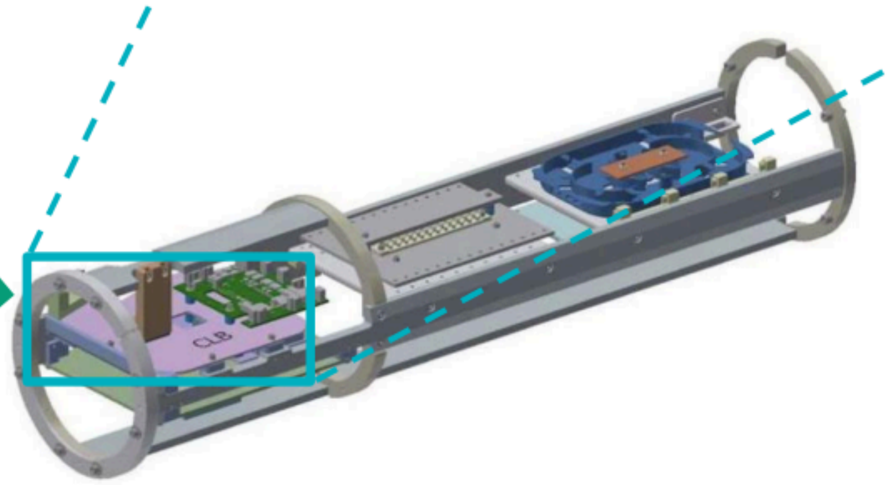
Central Logic Board (CLB) v2



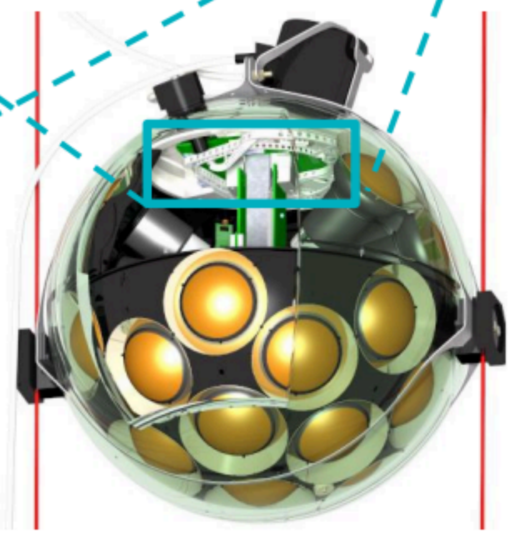
Xilinx Kintex 7 FPGA



Calibration Unit (CU) Base
(contains DU Base as controller)



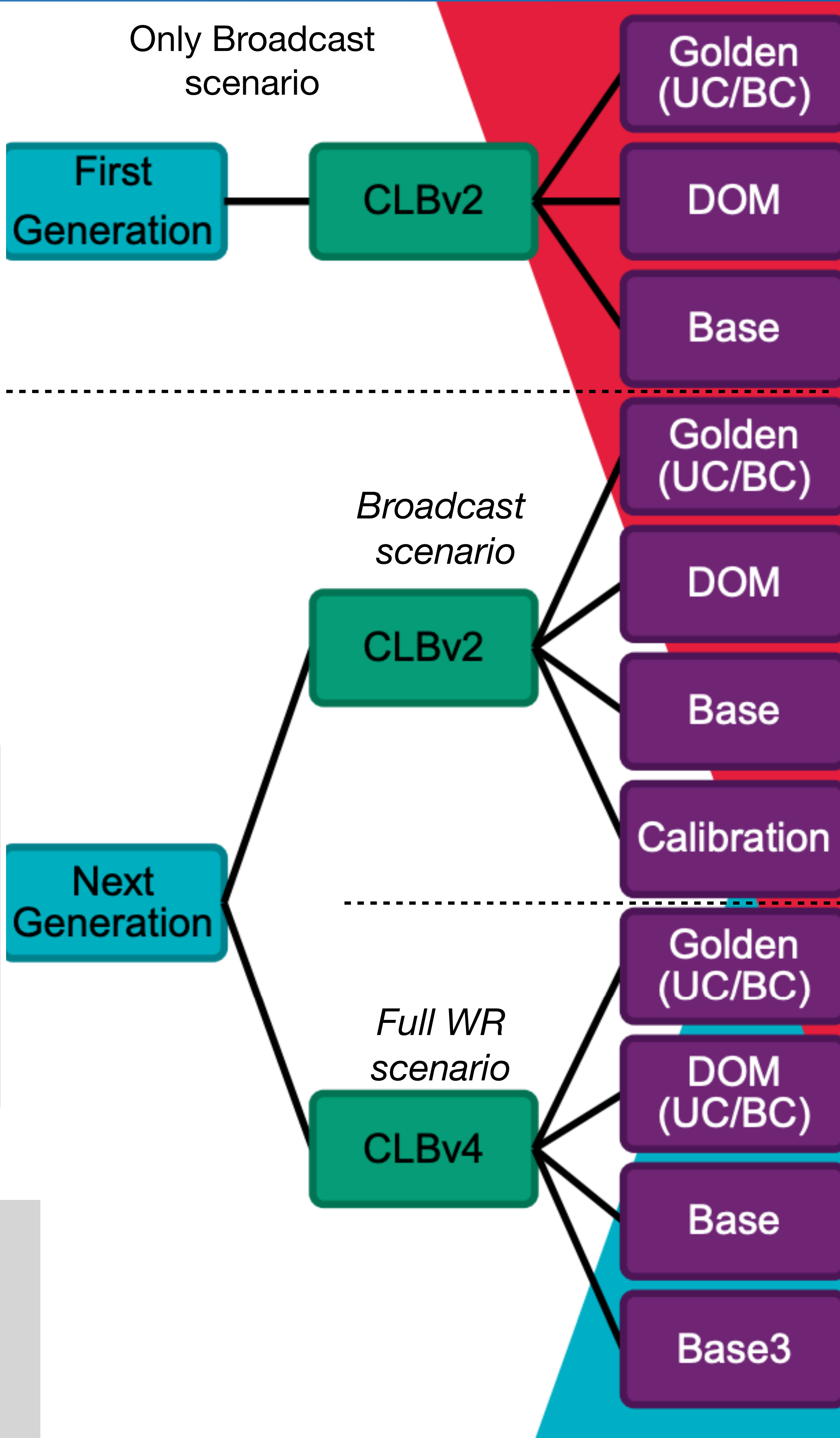
Detector Unit (DU) Base



Digital Optical Module (DOM)

* A CPU coded in a hardware language, running on FPGA programmable logic

Note on terminology used in this presentation:
- Gateware: Programmable Logic running inside an FPGA
- Embedded Software: Instructions running on an Embedded Controller (CPU)
- Firmware: The combination of Gateware and Embedded Software

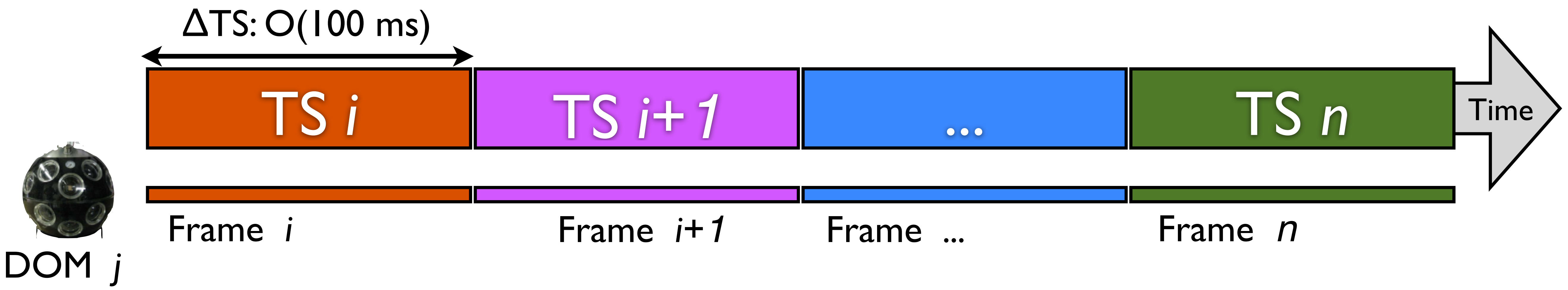


Firmware flavors for KM3NeT

FW Timeline

- 2014 First Gen prototype
- 2016 First Gen release with ARCA DU #1
- 2020 Next-Gen prototyping
- 2021 First-Gen Last release
- 2022 Next—Gen Deployment

- **Timeslice** (TS): it is the abstract subdivision of the continuity in the time-line of the experiment.
- **Frame**: it is the group of information of a certain flavour (TDC, AES, MON) occurred in a DOM during a TS.

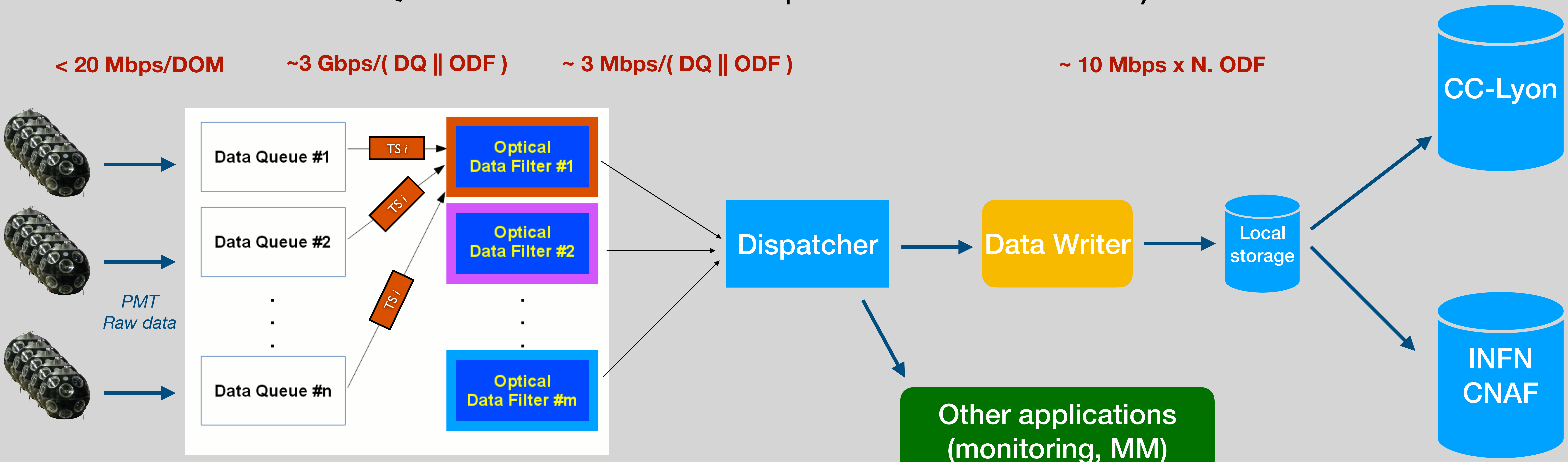


Distributing the computational load

- Each trigger algo applied to one full set of frames of one TS.
- Multiple TSs handled in parallel

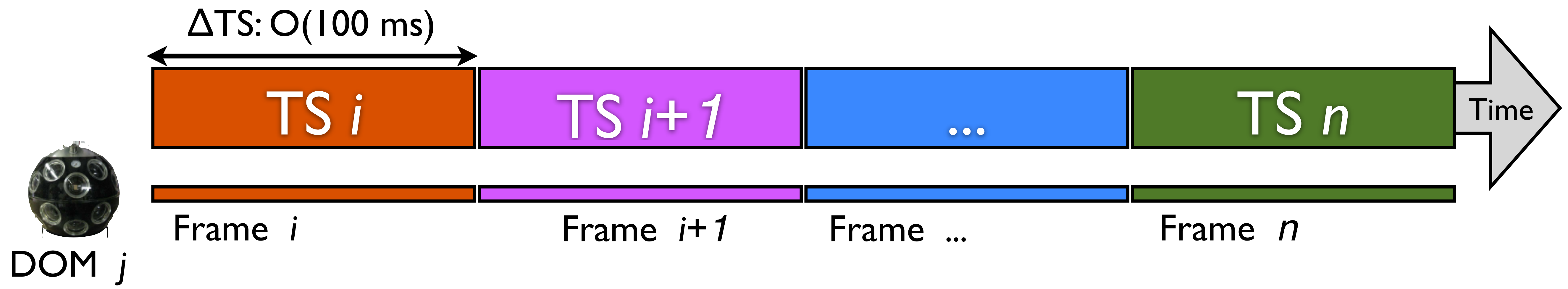
Optical World

A DQ collects data from a sector of DOMs and DU-BMs.
All DQs transfer all their data from a precise Time Slice to the very same oDF.



If needed more *Filtering Power*, just add it!

- **Timeslice** (TS): it is the abstract subdivision of the continuity in the time-line of the experiment.
- **Frame**: it is the group of information of a certain flavour (TDC, AES, MON) occurred in a DOM during a TS.

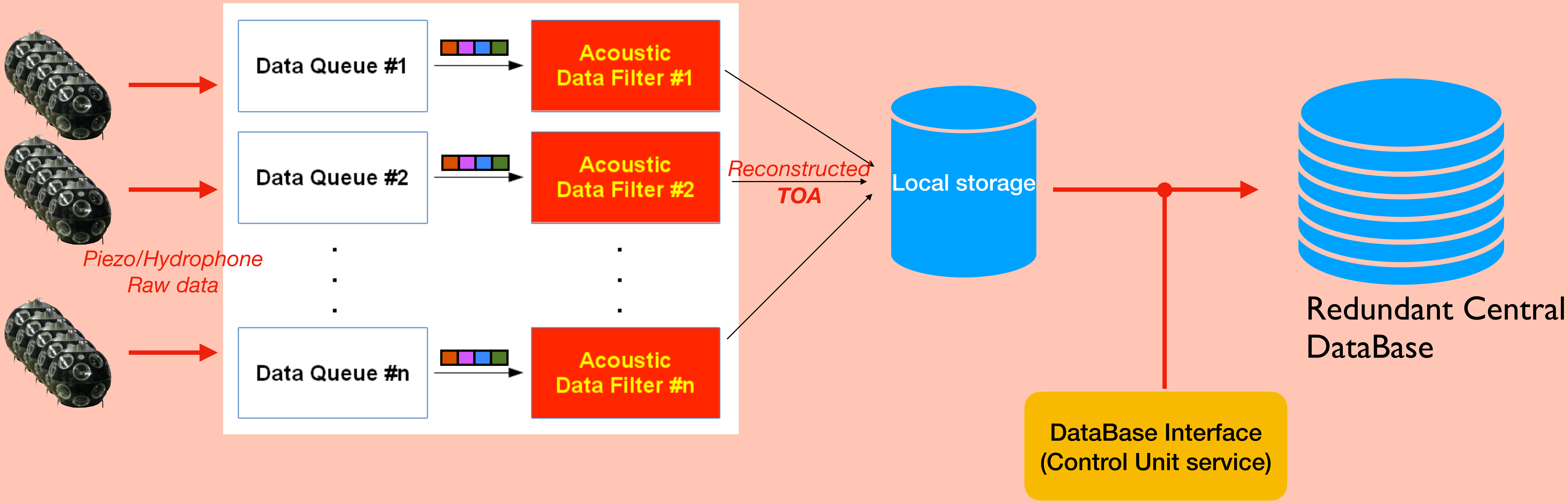


Distributing the computational load

- Each trigger algo applied to one full set of frames of one TS.
- Multiple TSs handled in parallel

Acoustic World

Acoustic data must be sent in a continuous stream, addressing all data from one DQ to a single Acoustic DF.
Independent reconstruction of the *Time Of Arrival* (**TOA**) of acoustic signals from various beacons



Basic triggers

L0: all hits over threshold (i.e. all hits sent by the CLBs)
L1: pairs of hits of the same DOM within 25(10)ns.

L2-trigger level

- **3D-Trigger** - general concept:

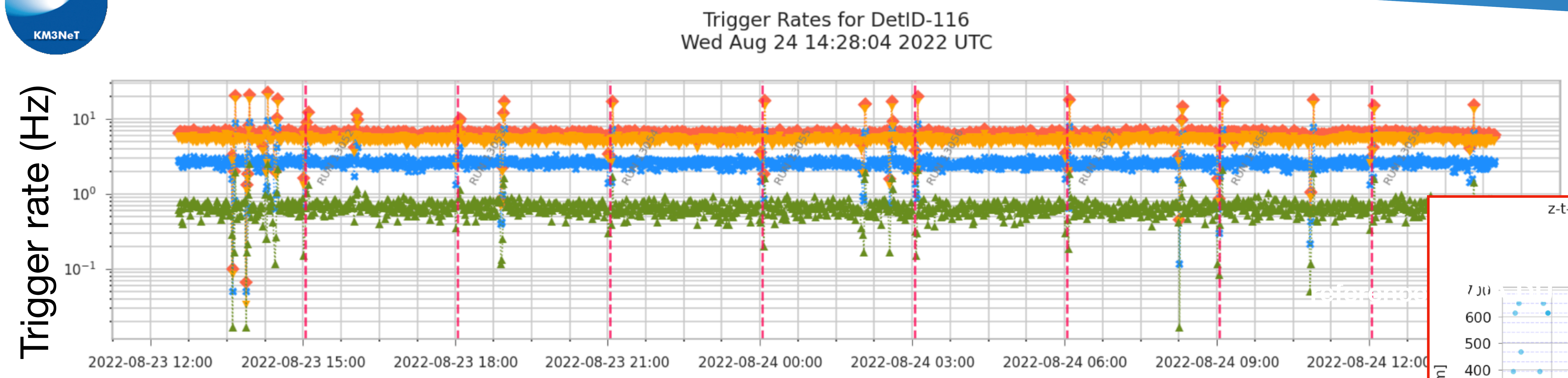
1. A minimum n. of ***consecutive*** L1s ≥ N_{th} within a ΔT (at least $n_{DOM} \geq 2$ or 5)
2. *3D*-causality filter : $|t_i - t_j| \leq |\vec{x}_i - \vec{x}_j| \frac{n}{c} + T_{MaxExtra}$
3. The trigger is set if the n. of satisfying hits is ≥ N'_{th}

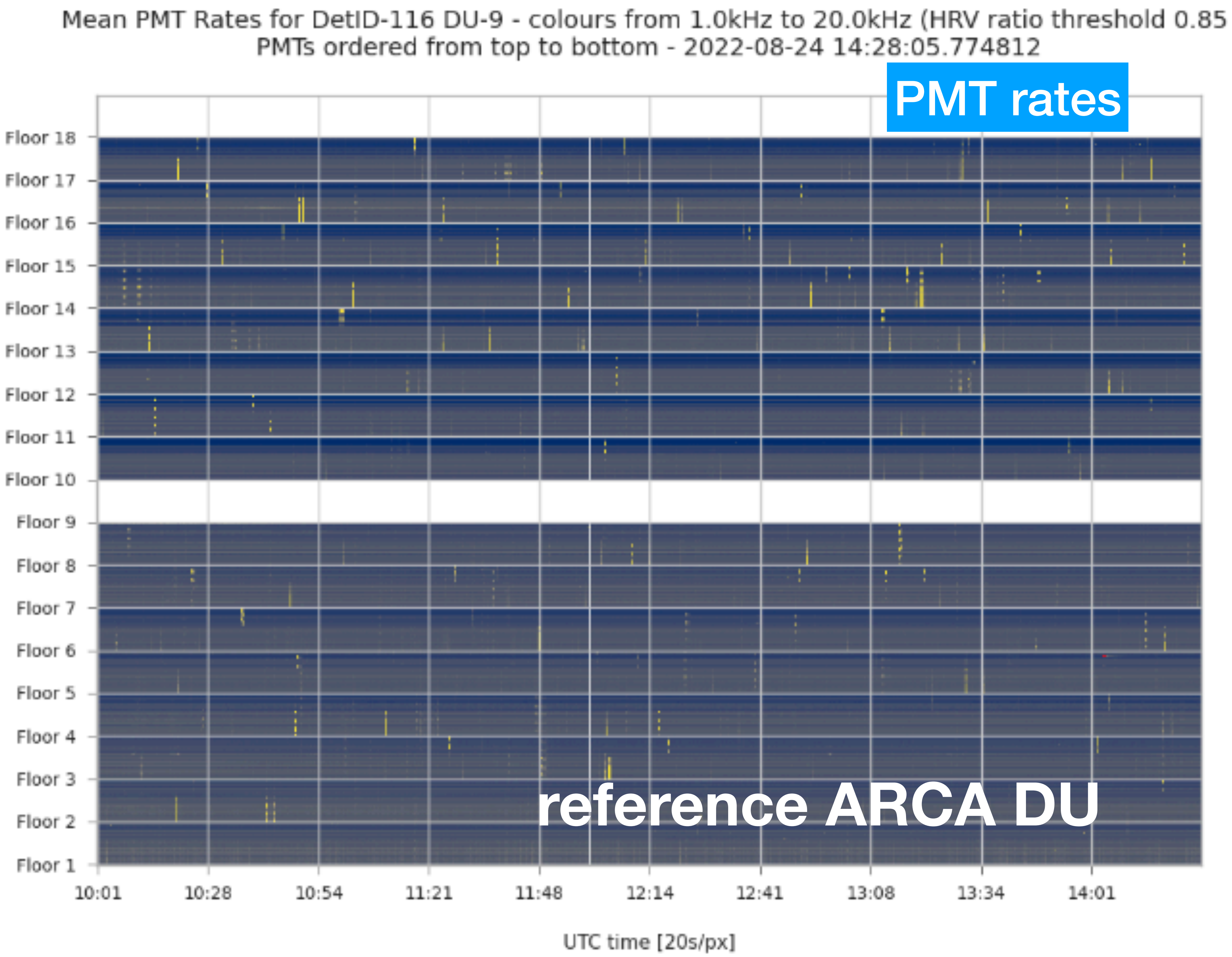
- **3D-Muon/Shower**
  Assumes an extended track-like / short pulse shape for the event topology

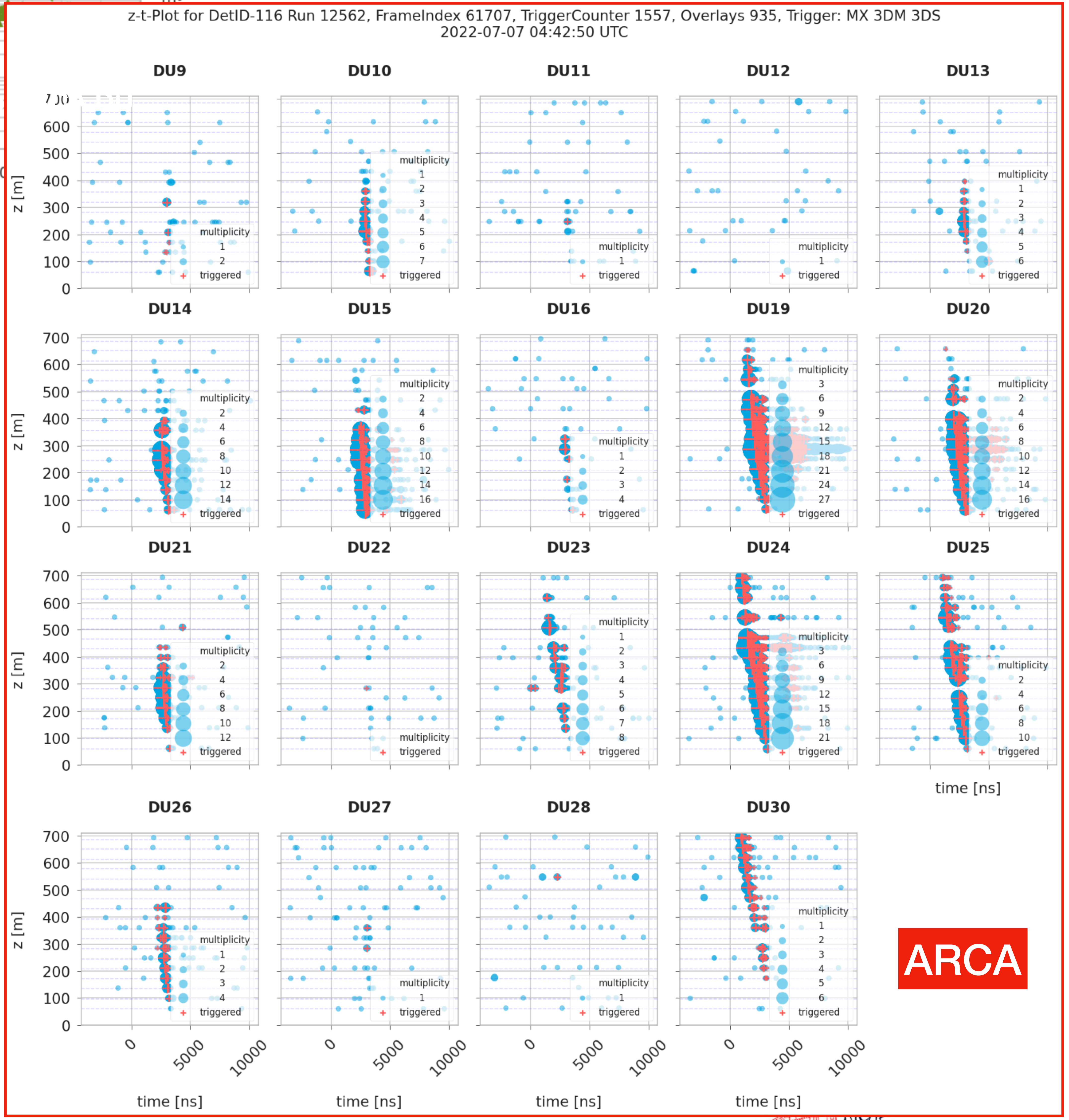
- **MX-Shower**
  Combines L0s and L1s within a limited space in the detector.

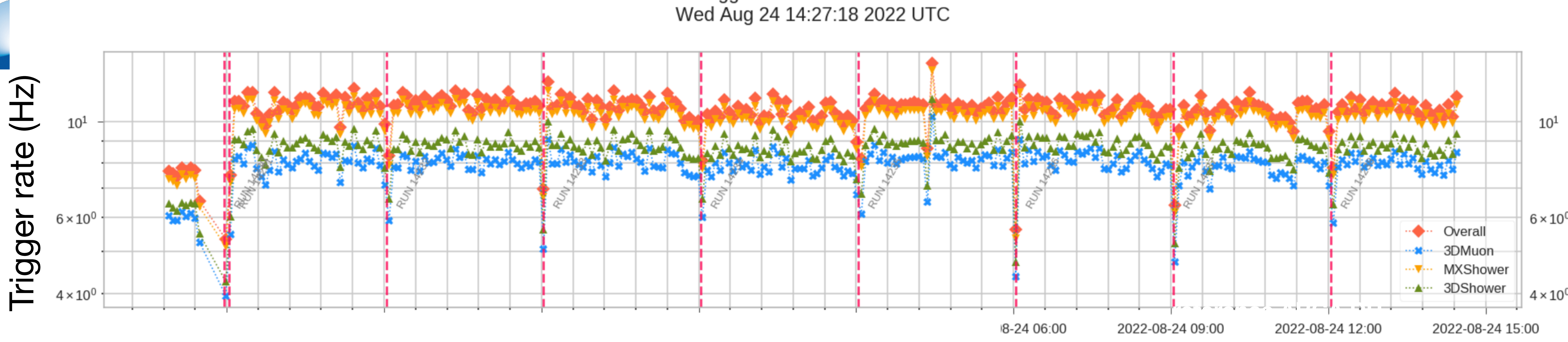
Trigger settings passed to the Data Filters via the run setups by the Control Unit

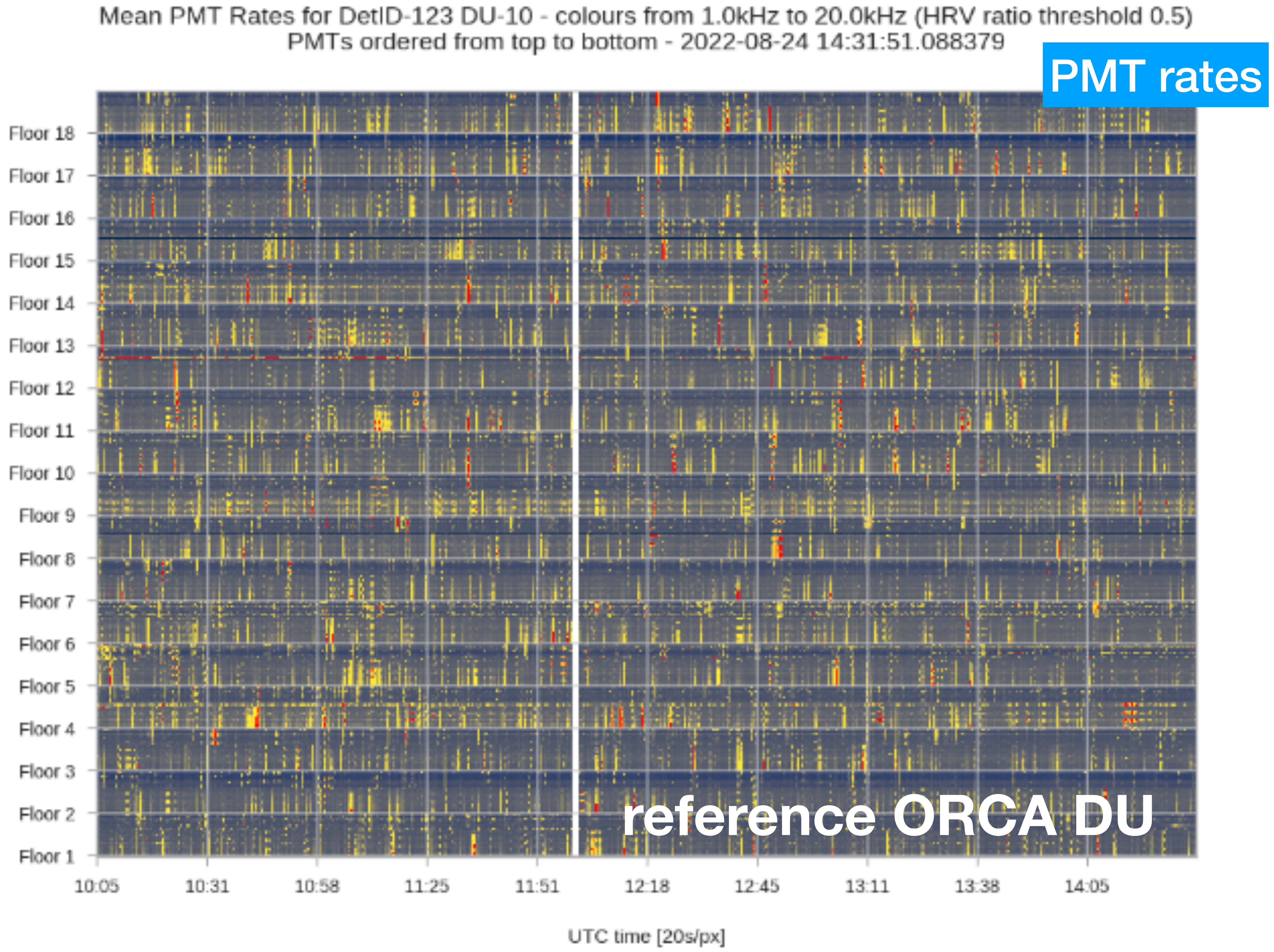
Trigger algorithms developed within a large C++ software framework, ***Jpp.*** The same codes are used for the on-line DAQ as well as off-line analysis.

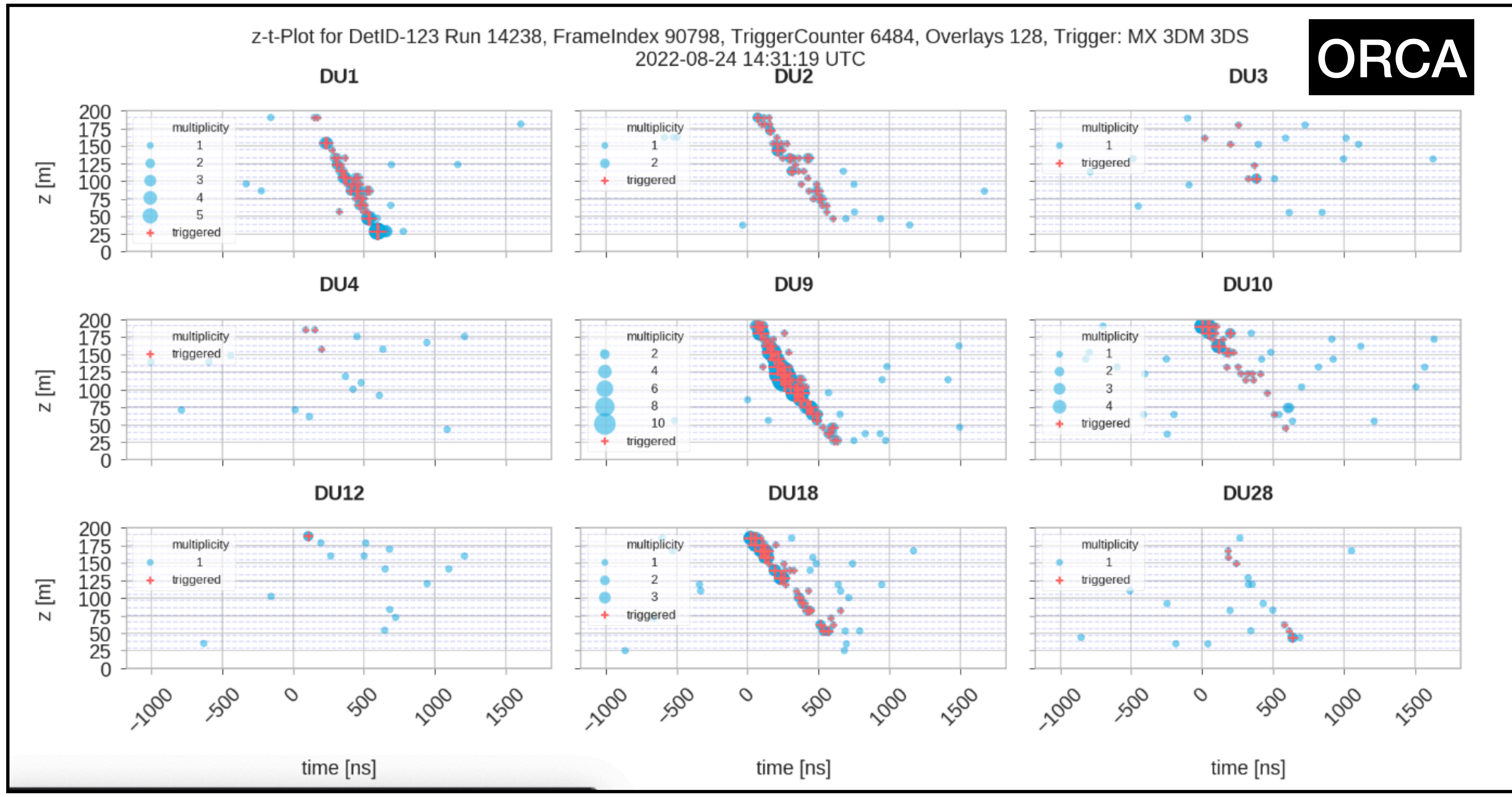

Trigger Rates for DetID-116
Wed Aug 24 14:28:04 2022 UTC
Trigger rate (Hz)
2022-08-23 12:00
2022-08-23 15:00
2022-08-23 18:00
2022-08-23 21:00
2022-08-24 00:00
2022-08-24 03:00
2022-08-24 06:00
2022-08-24 09:00
2022-08-24 12:00

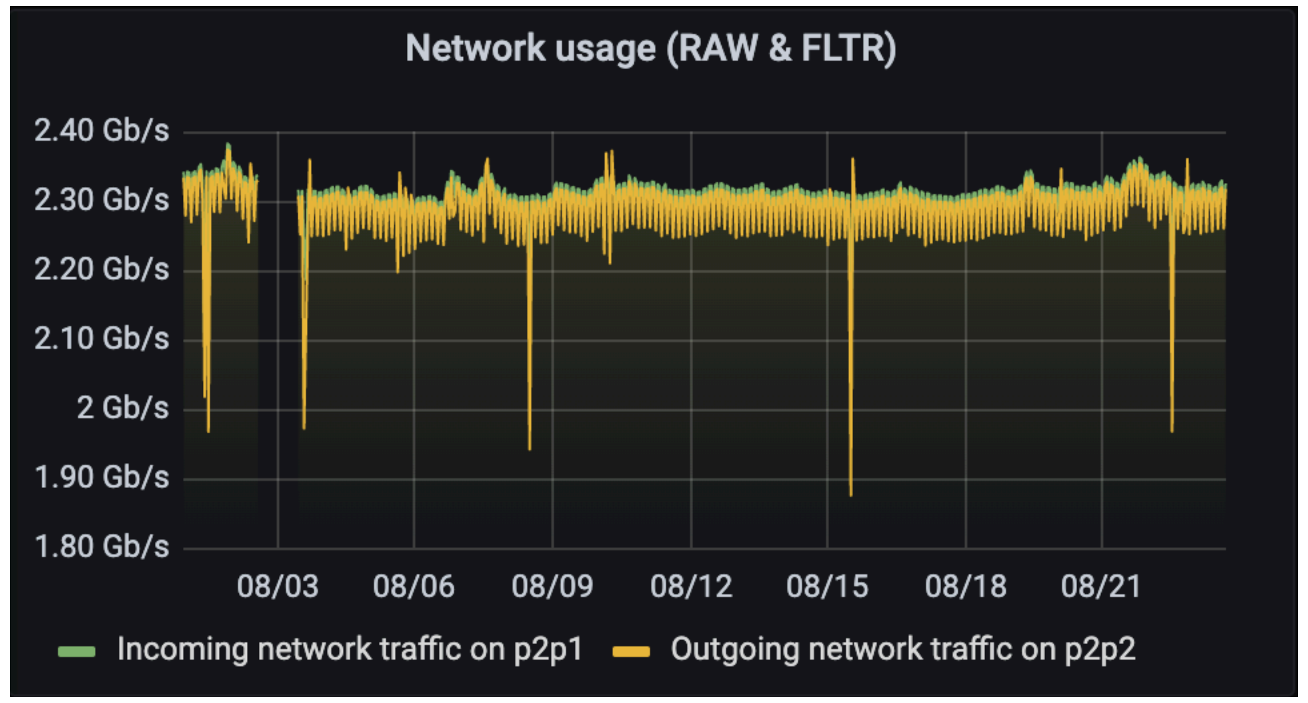

Mean PMT Rates for DetID-116 DU-9 - colours from 1.0kHz to 20.0kHz (HRV ratio threshold 0.85
PMTs ordered from top to bottom - 2022-08-24 14:28:05.774812
PMT rates
Floor 18
Floor 17
Floor 16
Floor 15
Floor 14
Floor 13
Floor 12
Floor 11
Floor 10
Floor 9
Floor 8
Floor 7
Floor 6
Floor 5
Floor 4
Floor 3
Floor 2
Floor 1
reference ARCA DU
10:01
10:28
10:54
11:21
11:48
12:14
12:41
13:08
13:34
14:01
UTC time [20s/px]

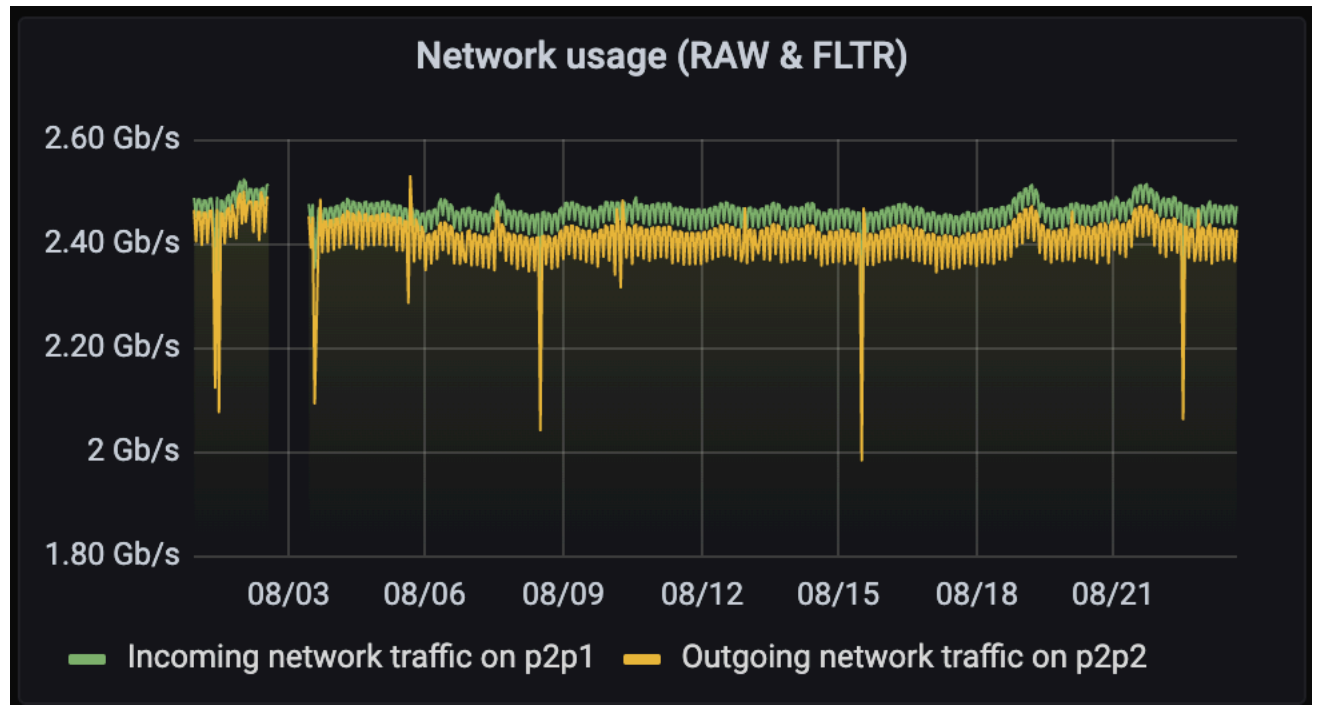

z-t-Plot for DetID-116 Run 12562, FrameIndex 61707, TriggerCounter 1557, Overlays 935, Trigger: MX 3DM 3DS
2022-07-07 04:42:50 UTC
DU9
DU10
DU11
DU12
DU13
DU14
DU15
DU16
DU19
DU20
DU21
DU22
DU23
DU24
DU25
DU26
DU27
DU28
DU30
z [m]
time [ns]
multiplicity
triggered
ARCA


Wed Aug 24 14:27:18 2022 UTC
Trigger rate (Hz)
Overall
3DMuon
MXShower
3DShower
2022-08-24 09:00
2022-08-24 12:00
2022-08-24 15:00


Mean PMT Rates for DetID-123 DU-10 - colours from 1.0kHz to 20.0kHz (HRV ratio threshold 0.5)
PMTs ordered from top to bottom - 2022-08-24 14:31:51.088379
PMT rates
reference ORCA DU
UTC time [20s/px]

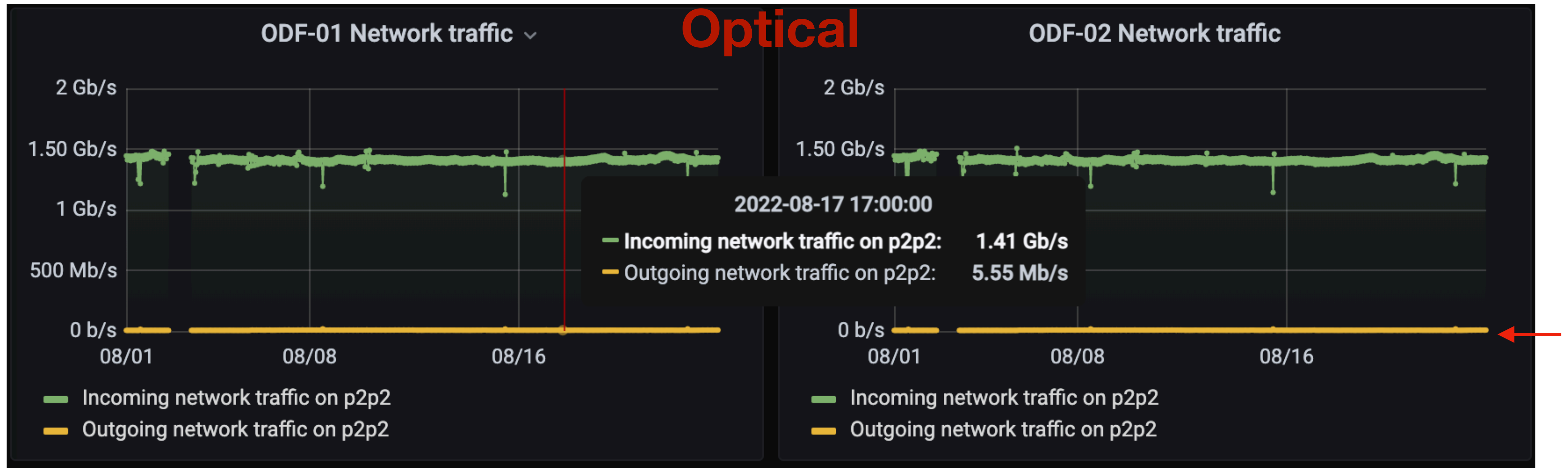

z-t-Plot for DetID-123 Run 14238, FrameIndex 90798, TriggerCounter 6484, Overlays 128, Trigger: MX 3DM 3DS
2022-08-24 14:31:19 UTC
ORCA
DU1
DU2
DU3
DU4
DU9
DU10
DU12
DU18
DU28
z [m]
time [ns]
multiplicity
triggered



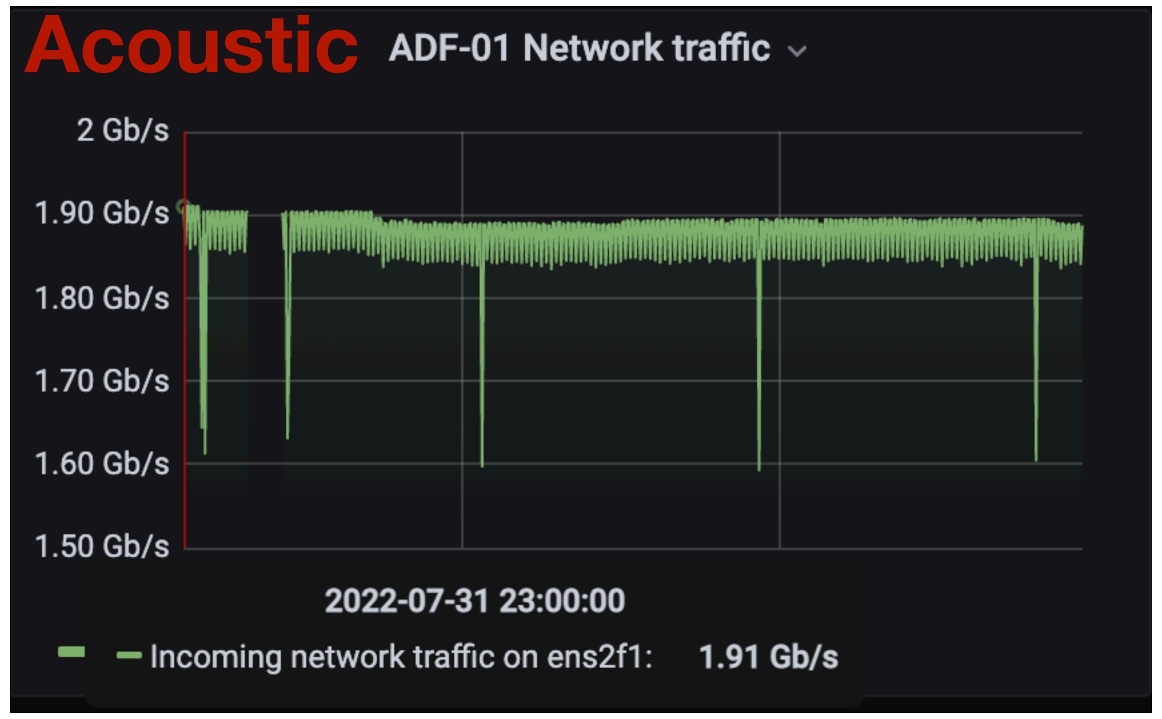


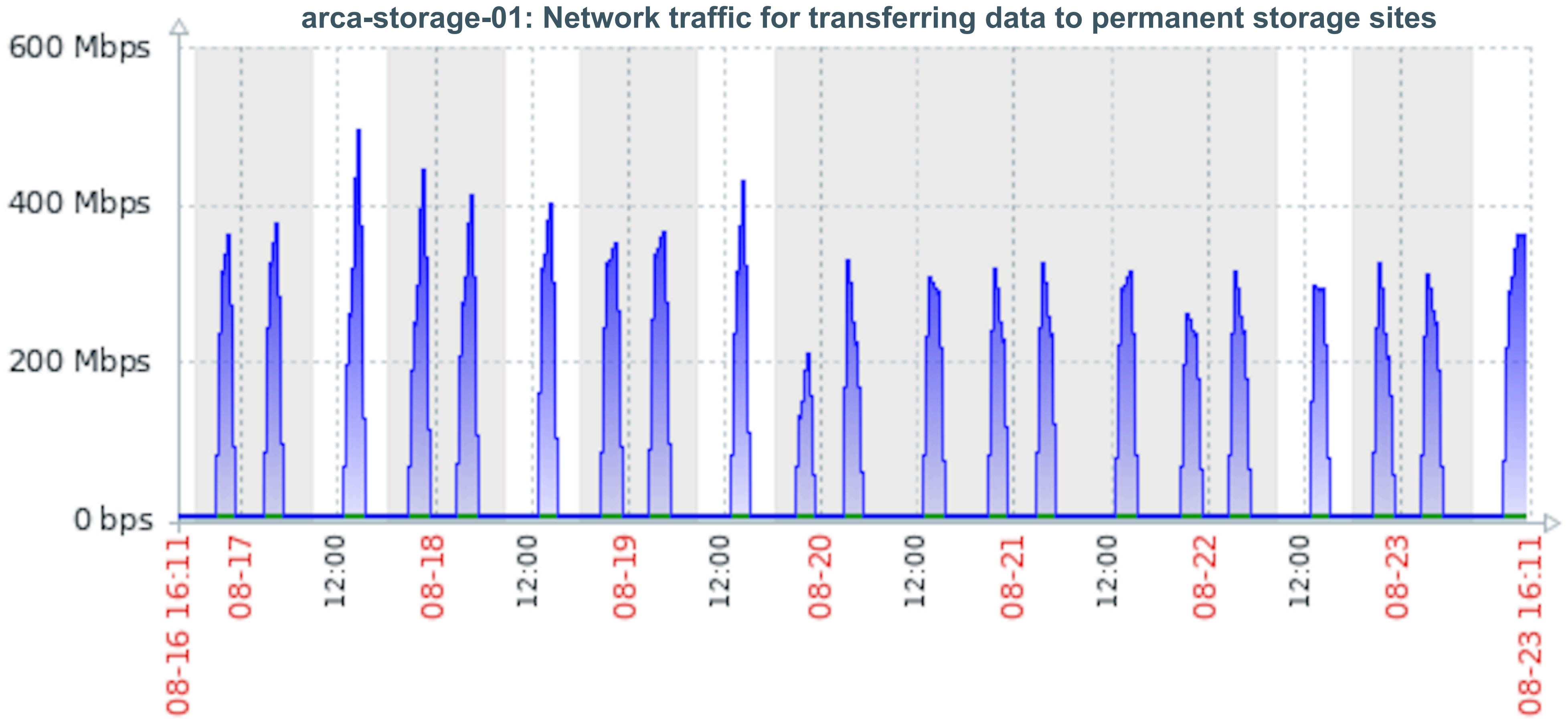
DataQueue level:

- receive and route to Data Filters (O+A)



data to Data Writer (i.e. data filtered by ~3 order of magnitued)





Periodic data transfer to permanent storage @CC-Lyon and @CNAF

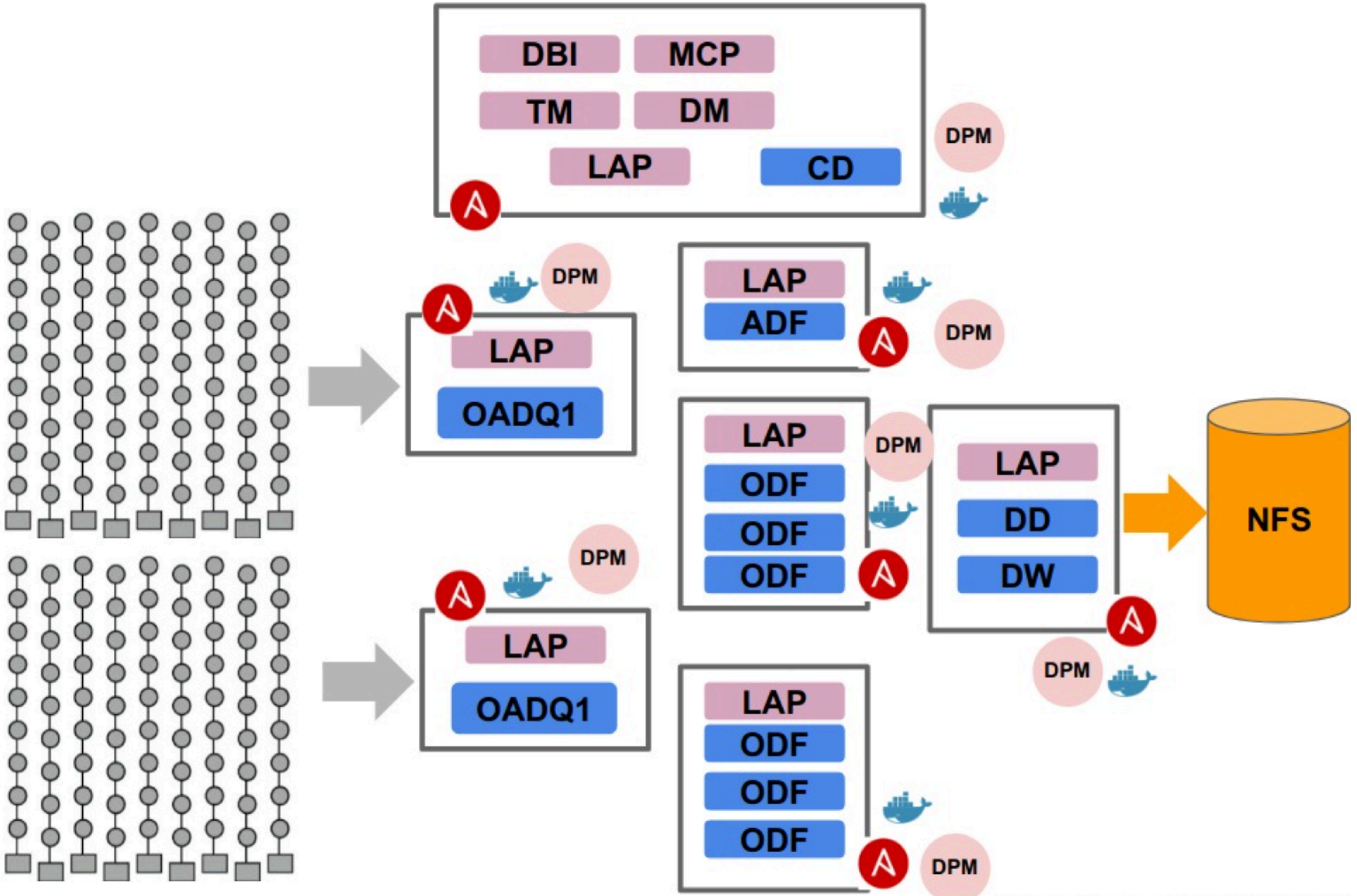
AIACE is an ANSIBLE-based collection of *playbook* for installing and configuring the computing resources and network devices.

DOCKER images -> independent container for each DAQ process, their deployment is handled via AIACE

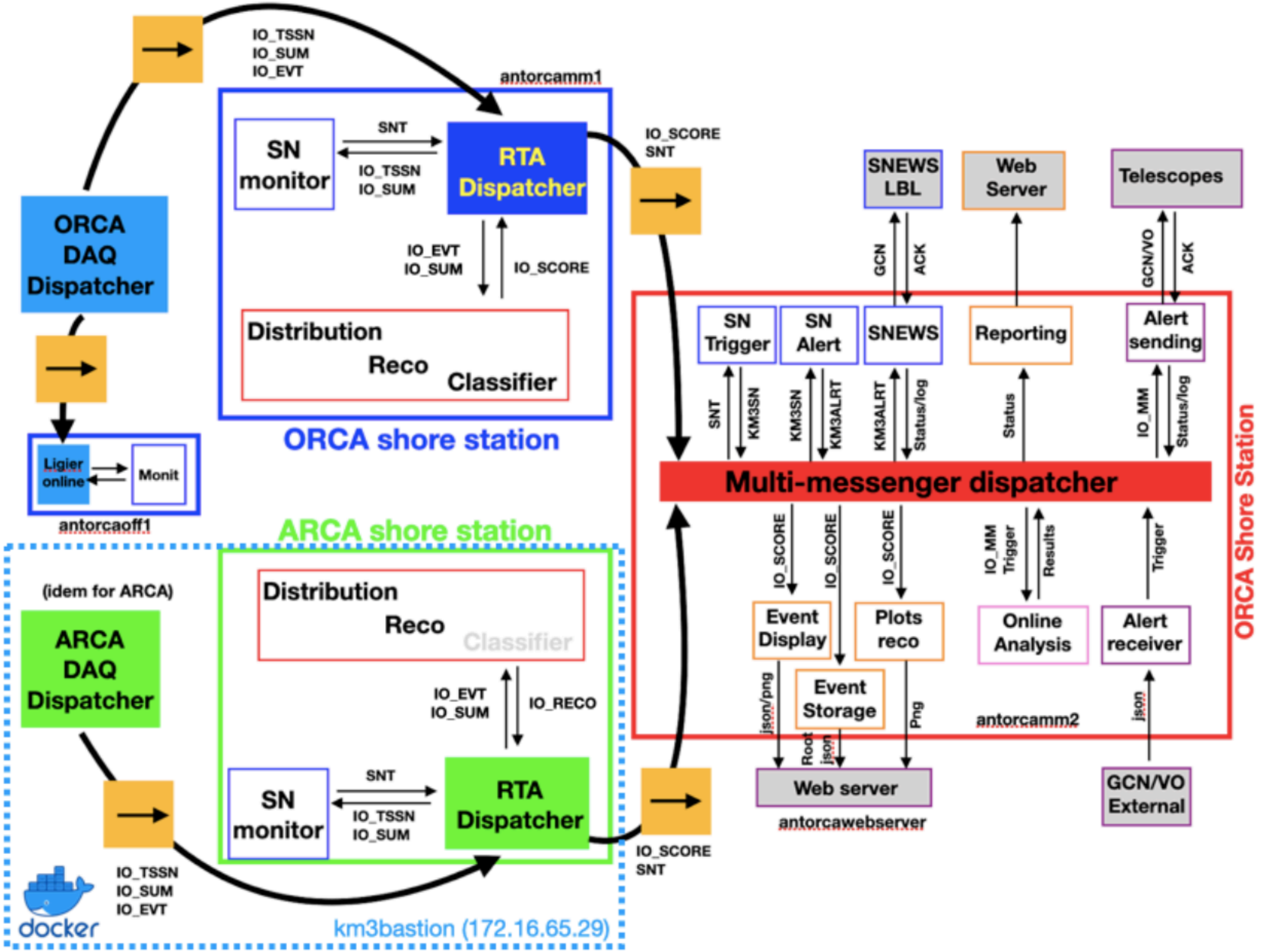
Dynamic Provisioning Manager (DPM) system for "keeping-alive" the DAQ processes and role manager

At present, both for **ARCA19** and **ORCA11**:

- 2x DataQueue processes (on 2 independent server)
- 30x Optical Data Filter processes (on 2 independent server)
- 1x Acoustic Data Filter (on 1 independent server)
- 1x DataWriter together with 1 DataDispatcher (on 1 independent server)

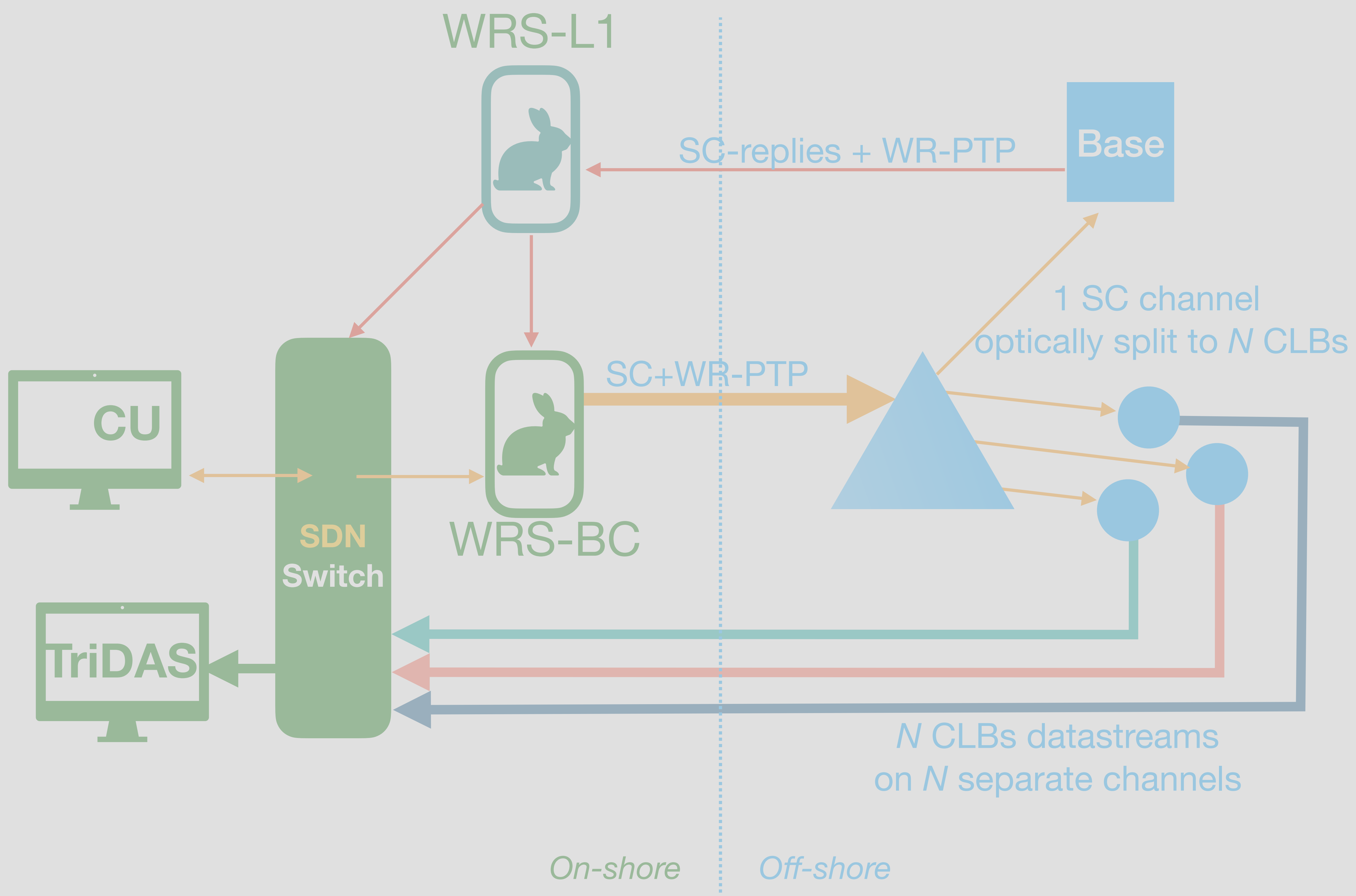


- Event processing done separately for ARCA and ORCA at each shore station
- Same **processing structure** but different software organisation (in ARCA the docker approach is adopted).
- The output of the reconstructed events by ARCA and ORCA at the end of each run (.json files) is stored in a common dispatcher (MM dispatcher)



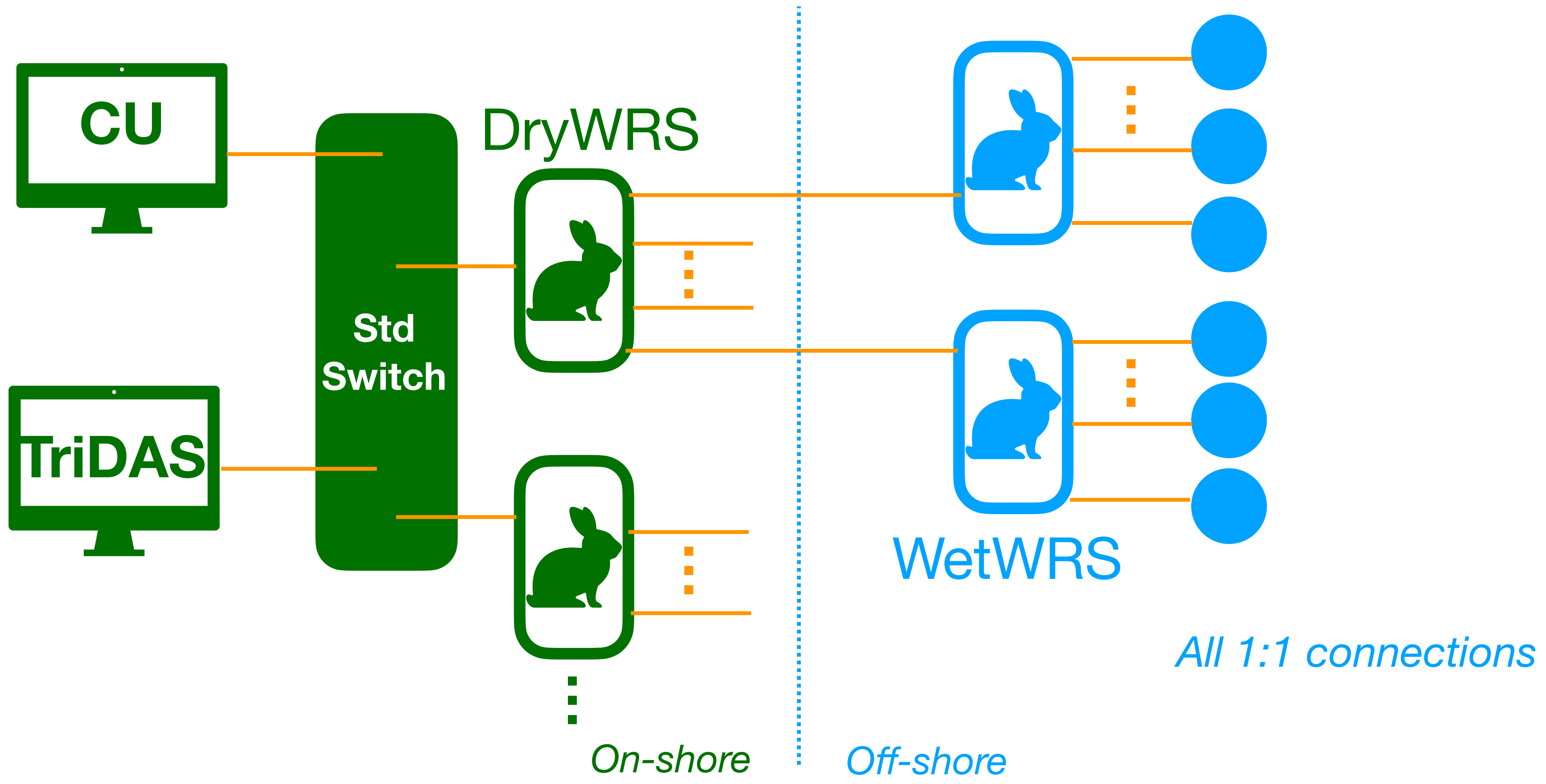
KM3NeT

Broadcast (ARCA 32 strings; ORCA 48 strings *at least*)



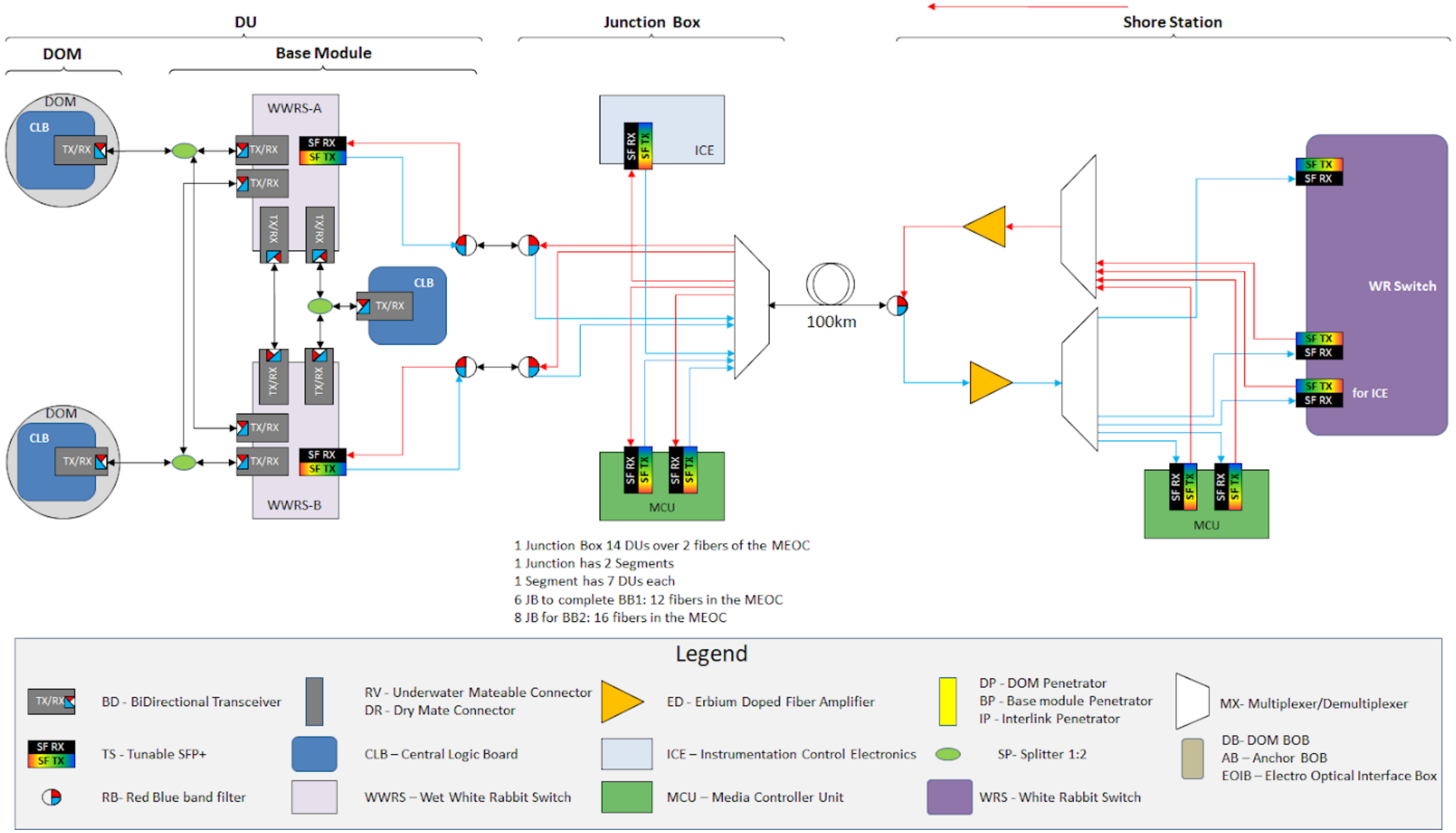
Current implementation in both ORCA/ARCA
(as well as other test-installations)

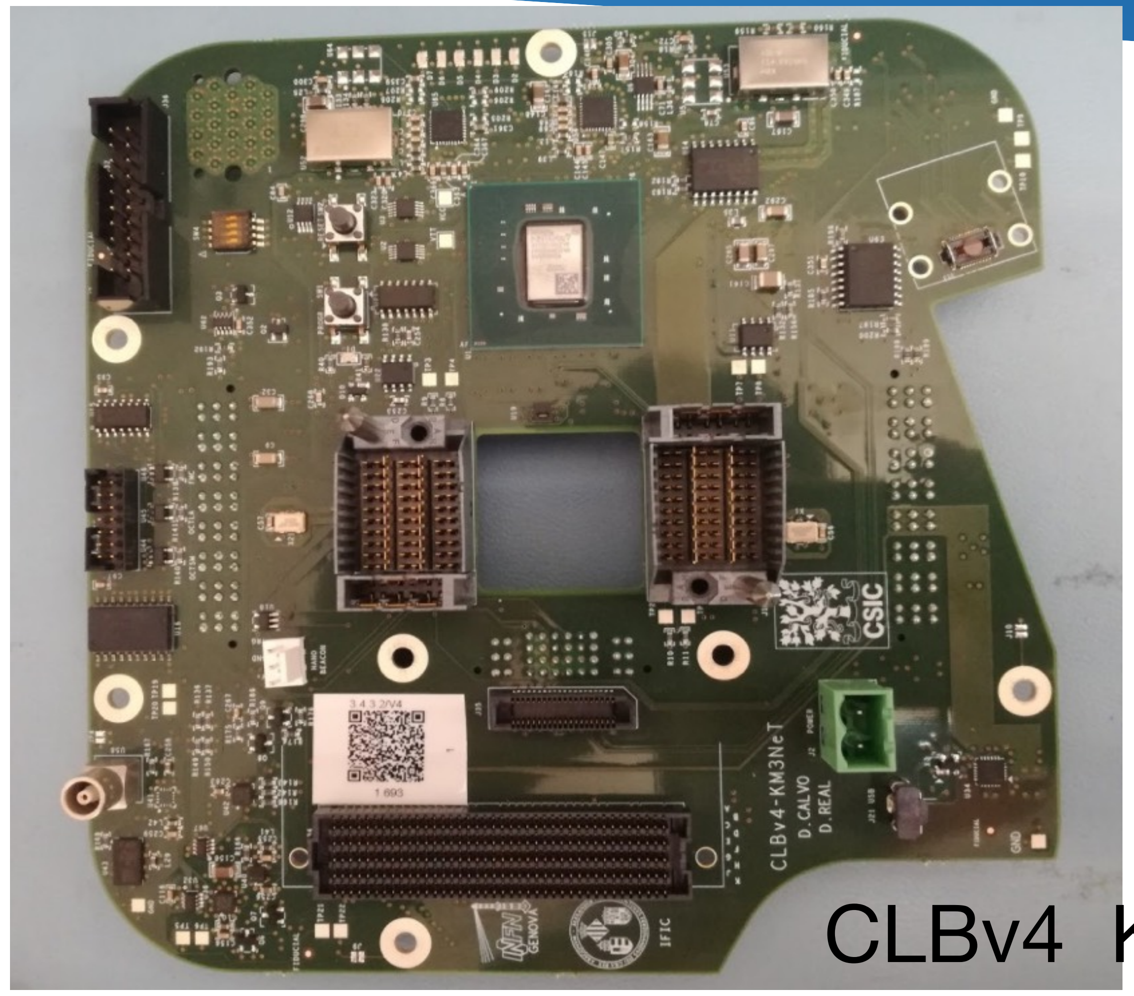
Full White Rabbit (necessary for ARCA 2 BB)



Future evolutions

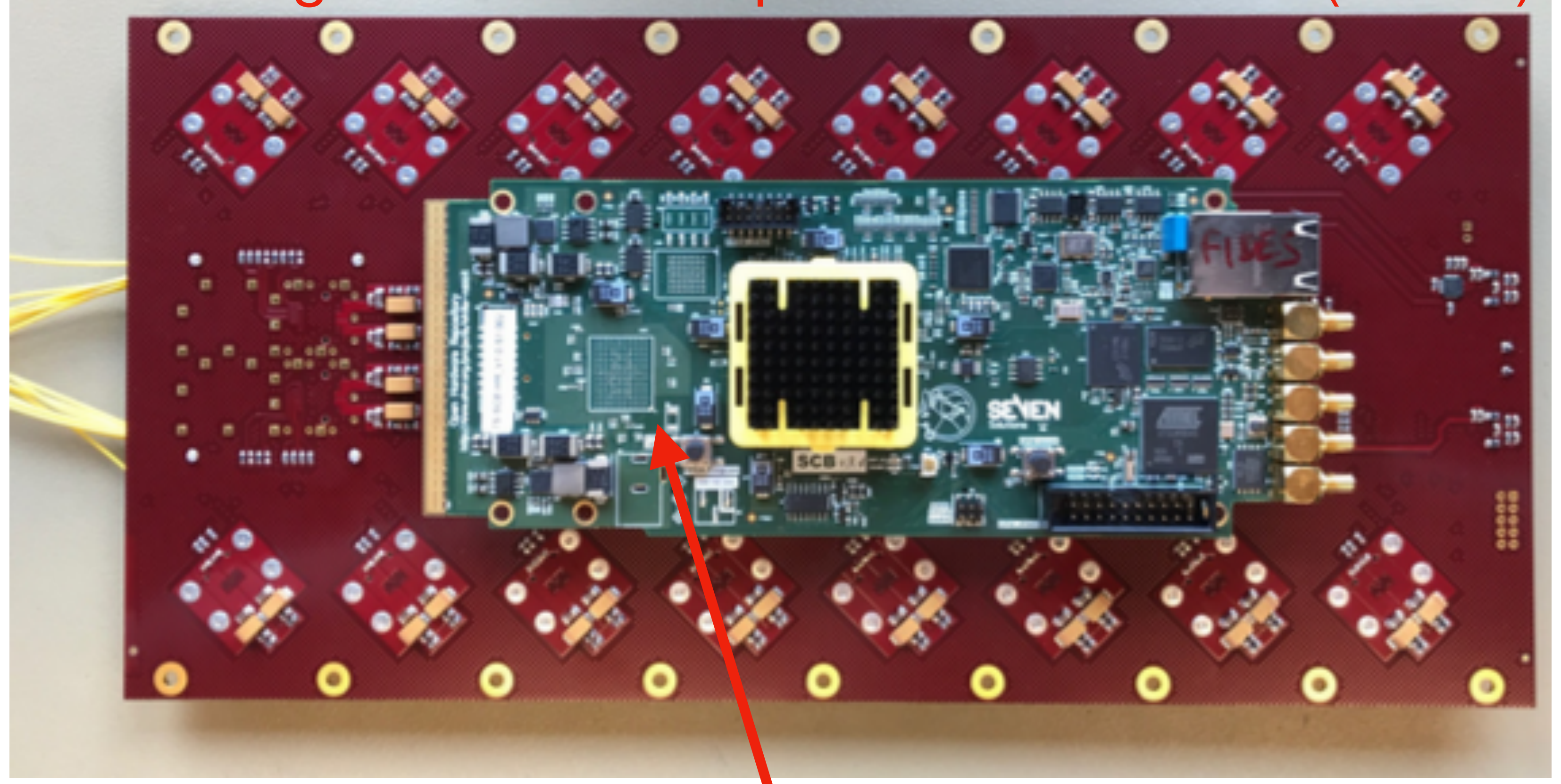
KM3NeT Phase 2.0 - ARCA Optical System - Overview



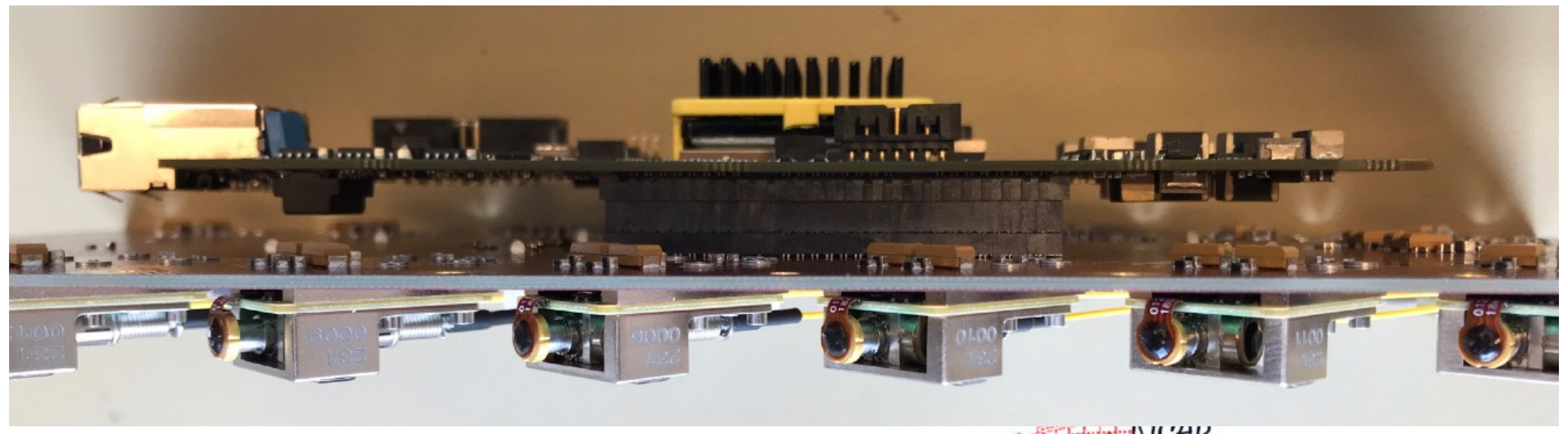


CLBv4 KM3NeT dev (IFIC)

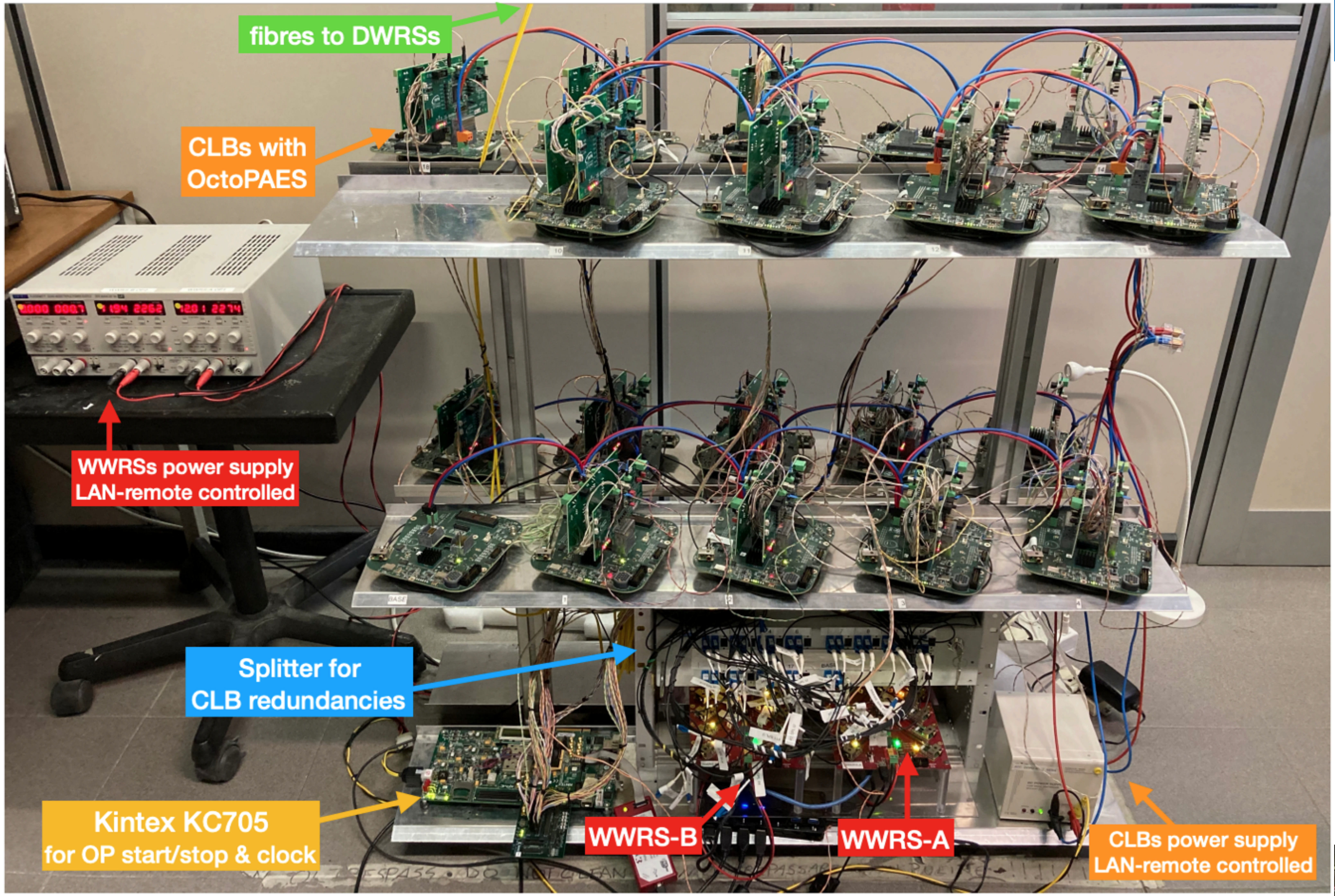
Switching "Glenair" backplane KM3NeT dev. (Nikhef)



SCB by SevenSolutions



- 2 tunable SFP+DWDM long range transceivers for connecting with the on-shore station
- 2 *Wet*WRS per DU: 9 DOMs each
- 1 BM CLB connected to both of the two WRSs (cold redundancy applied)
- 23 bidirectional short range transceivers (**high reliability**) for DOM connections (9x2), CLB connection (3), inter-WRS connection



One full DU with CLBv4 boards and GA transceivers/connections

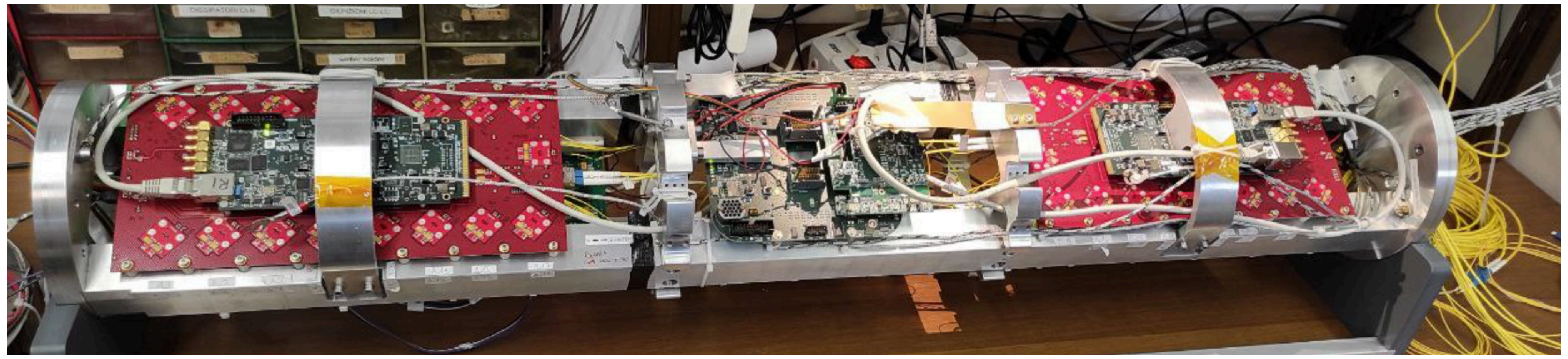
One integrated DU-BM with the WetWRS

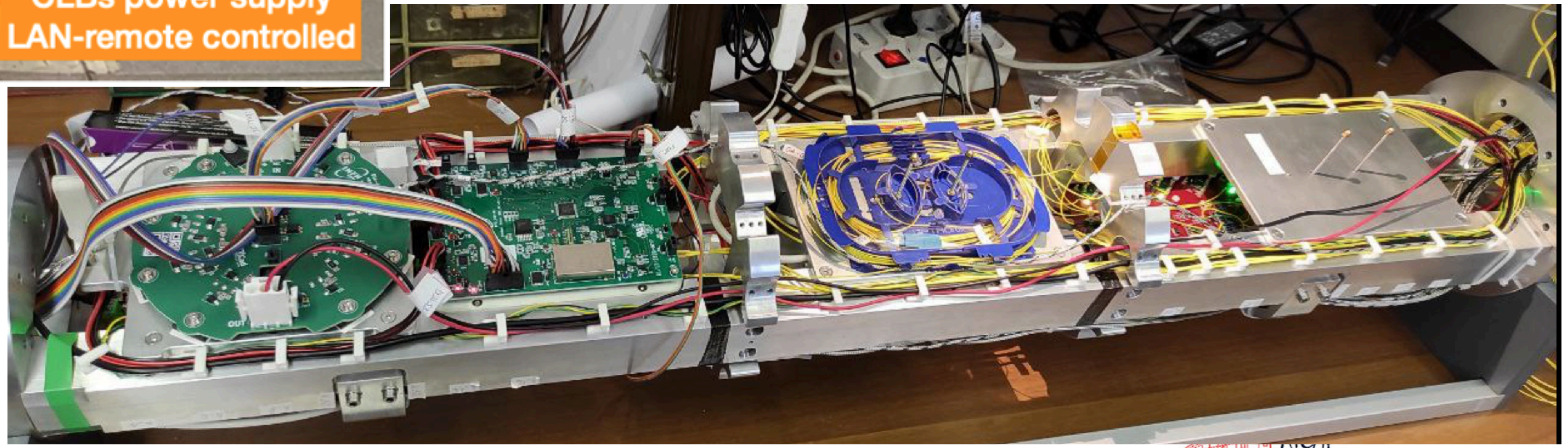
Custom electronic boards by INFN-Bo, the *OctoPAES* (emulation PMT and piezo/hydro), => tested runtime conditions of the DU:

- throughputs of various channels (PMT/ACU/MON)
- effectiveness of NG-Firmware for CLBv4
- control of DU and BM CLBv4 boards
- temperatures and power consumptions

Main boards (WWRS, power borards) subjected to HALT test

The full design is going under *Product Readiness Review* (end of 2022).

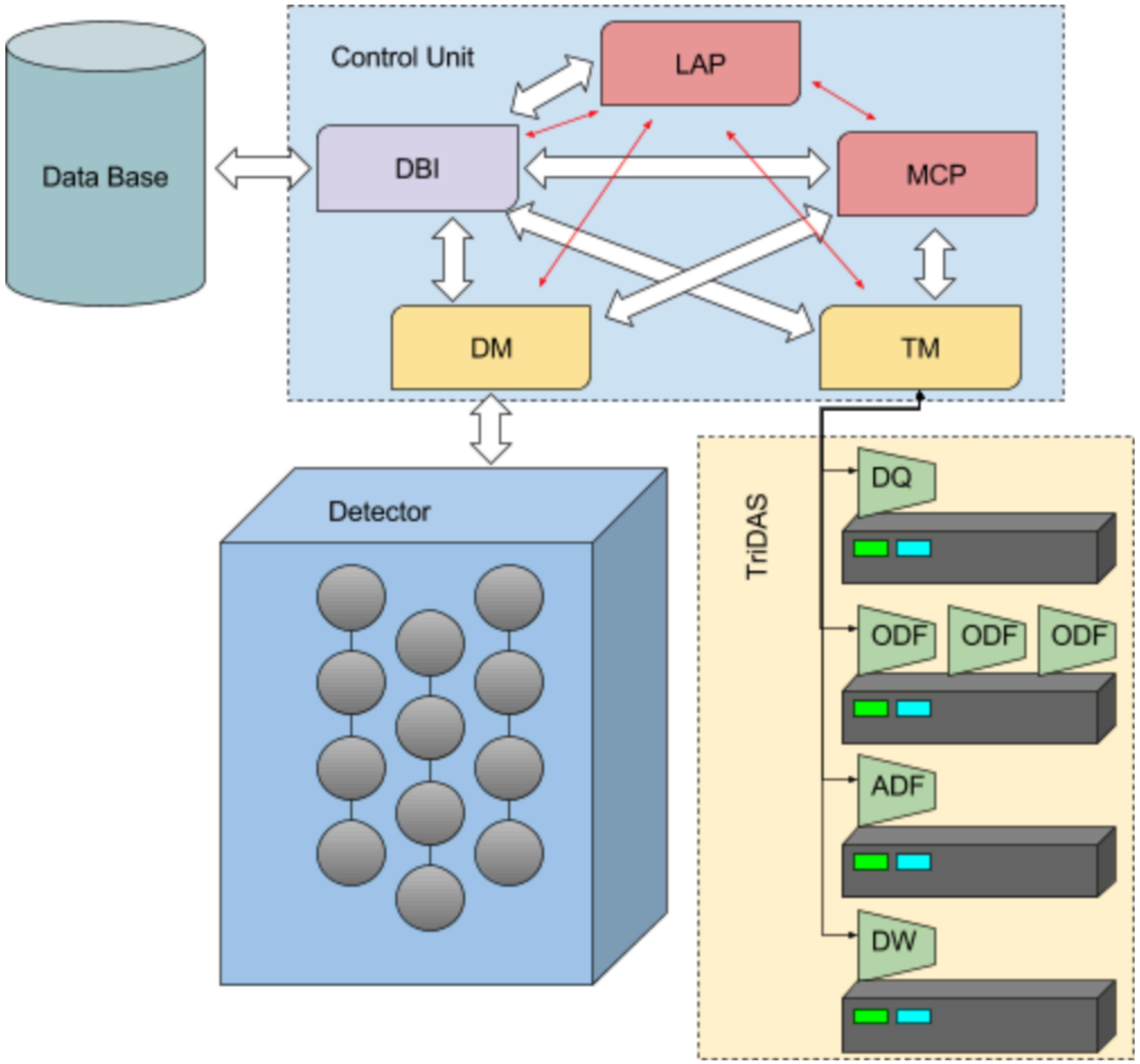




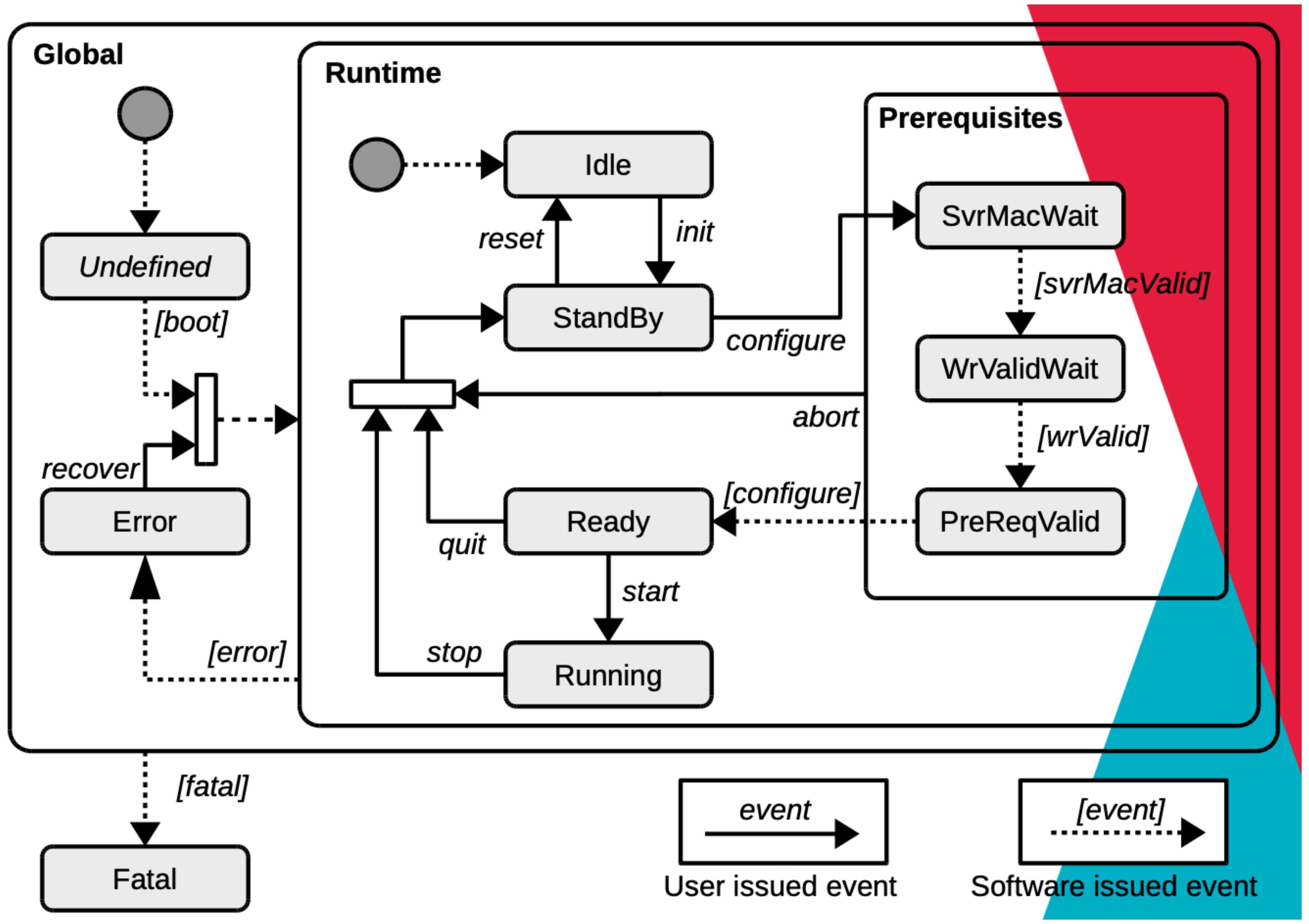
- **Scalable and modular** DAQ model
- **Dimensioned for large scale undersea neutrino detectors**, at least 2 building blocks with O(5000) endpoints
- **High throughput network** extending from on-shore to off-shore for O(100) km in a wide O(km3) volume
- Frontier technologies such as
  - **White Rabbit** for sub nanosecond precision distribution over ethernet
  - **Software Defined Networks**, for dealing with a highly asymmetric network technologies
- **Most reliable and modern technologies** used for handling the data taking control, the processing of the streaming readout and the monitoring
- Served software deployment and configuration within a **docker containerised computing infrastructure**.
- Big innovations with WRS infrastructure:
  1. Detection Units "(r)evolution" from the **"Broadcast" to the "Full-WR"** scenario (integration of the 2 scenarios)
  2. **Next Generation firmware** for the Central Logic Boards.
- NG-fw deployment foreseen i**n these weeks**.
- Currently active with ARCA 19 and ORCA 11… **soon more DUs**

Thanks for your attention !

Backup Slides



The Control Unit components and their relationships. White and black arrows represent flows of information and/or control signals. Red arrows show the flow of authentication information. The flow of data from the TriDAS to the final storage is not shown.



The CU is a collection of (web) services which, via a state machine, drive

- the Detector
- the computing processes
- the interactions with DB for
  - runsetups, calibrations
  - Instruments data logging

Details in Cristiano Bozza's talk

CLB Optical Format Structure

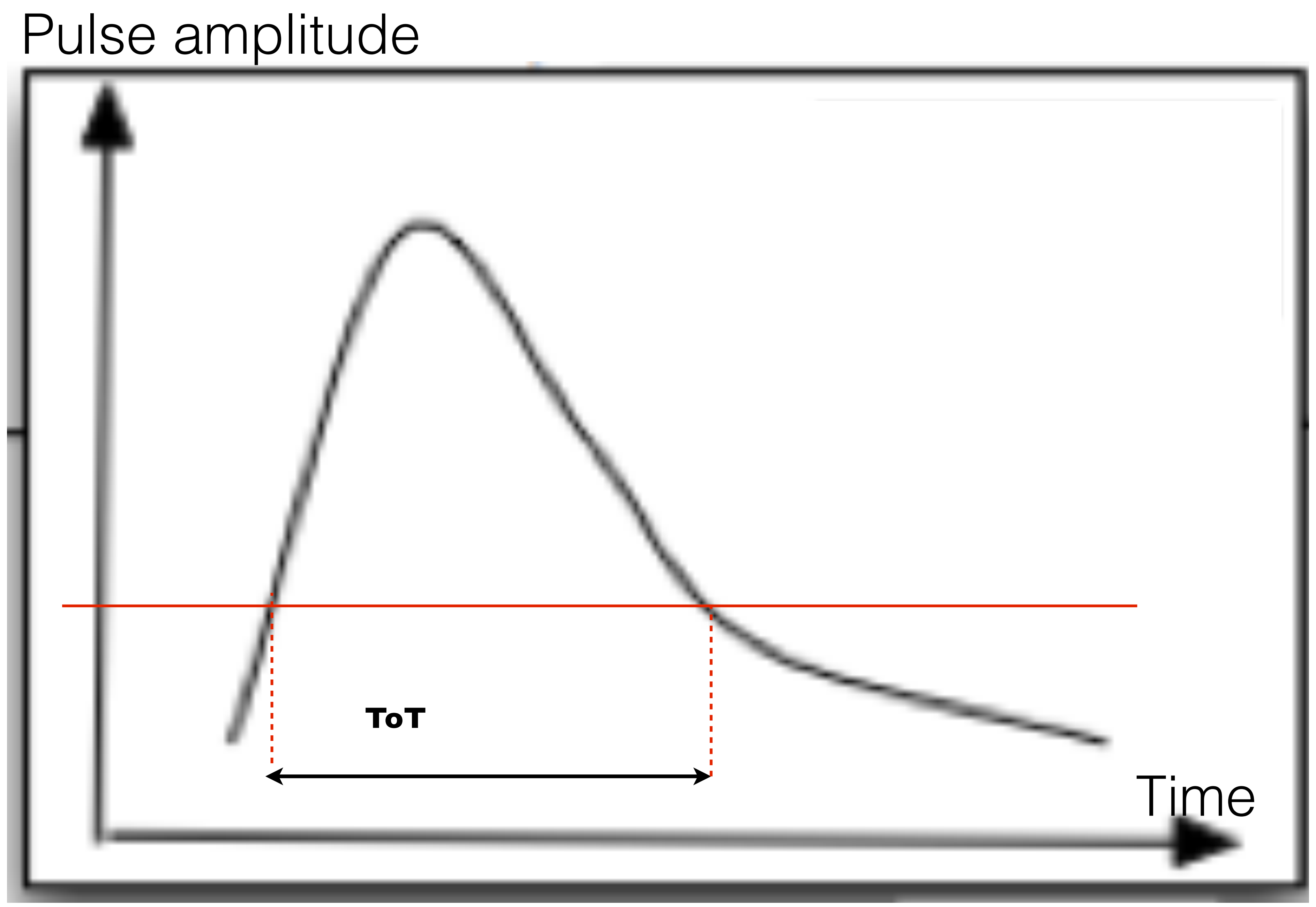
Size (bit)	Description
448	DAQ Common Header
8	TDC channel
32	Time Stamp
8	Pulse Width
8	TDC channel
32	Time Stamp
8	Pulse Width
...	...
8	TDC channel
32	Time Stamp
8	Pulse Width

One hit (6B)

- TDC (PMT) channel: 0 to 31
- Timing: counter of ns ∊ [0,1e8]
- Pulse width: Time over Threshold in ns ∊ [1,256]

Timing

Absolute time of a hit, with the precision of **1 ns.**



Time over Threshold

ToT <=> pulse amplitude.

Optimal gain (~10^6) => **ToT of 26,4 ns for single photo-electron**

Possibility to activate the ***Multi-Hit* feature** for longer pulses

CLB Optical Format Structure

Size (bit)	Description
448	DAQ Common Header
8	TDC channel
32	Time Stamp
8	Pulse Width
8	TDC channel
32	Time Stamp
8	Pulse Width
...	...
8	TDC channel
32	Time Stamp
8	Pulse Width

One hit (6B)

- TDC (PMT) channel: 0 to 31
- Timing: counter of ns ∊ [0,1e8]
- Pulse width: Time over Threshold in ns ∊ [1,256]

Timing

The **DOM is providing the absolute time of a hit**, with the precision of **1 ns.**
Hit's timestamp is then to be composed of:

- **Coarse timing** [s] + **Quasi fine timing** [16ns], from the CLB Common Header, gives the *absolute time of the Timeslice*.
- **Fine timing** [ns] from the TDC counts the *ns since the beginning of the Timeslice*

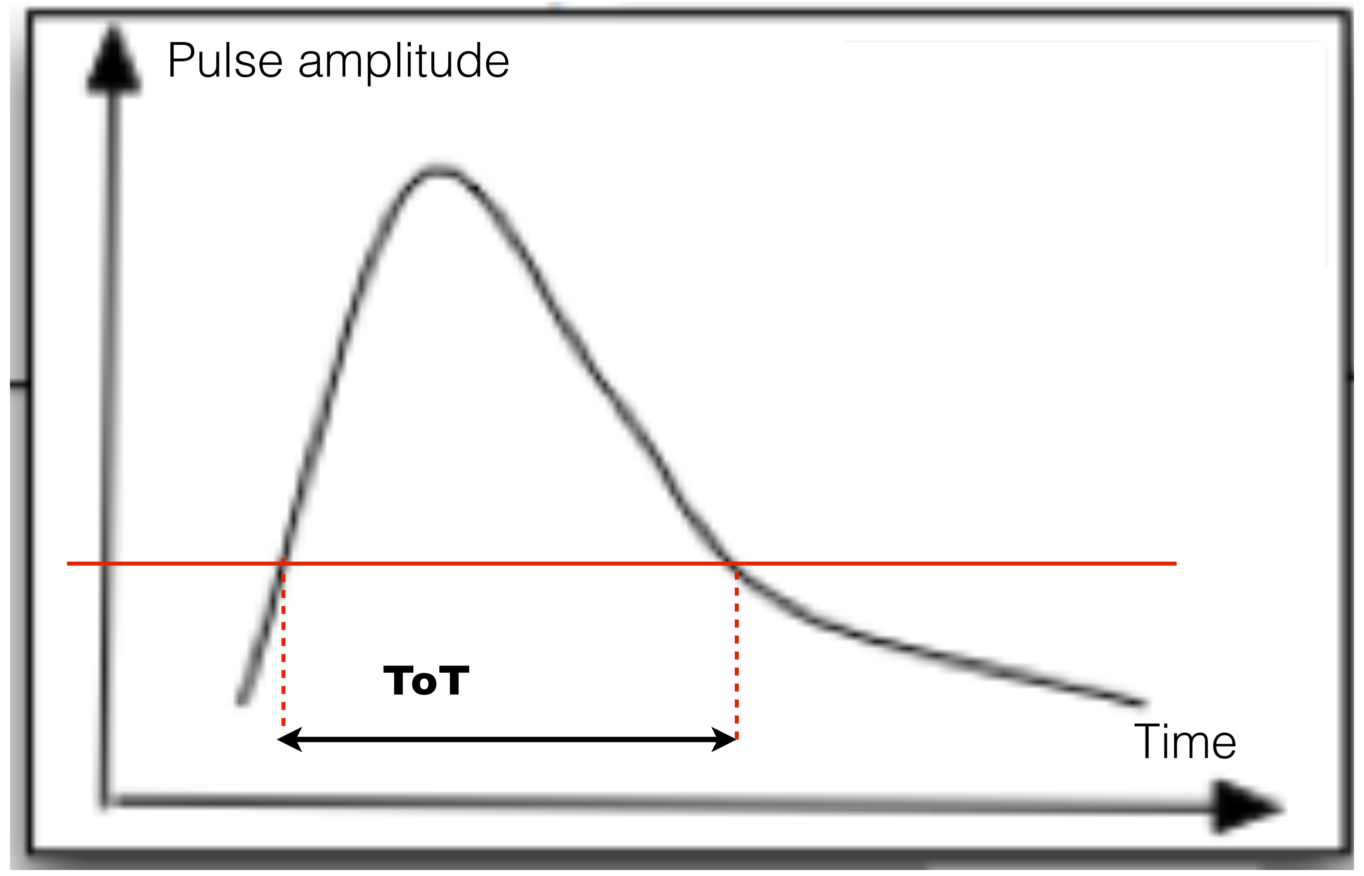
Time over Threshold

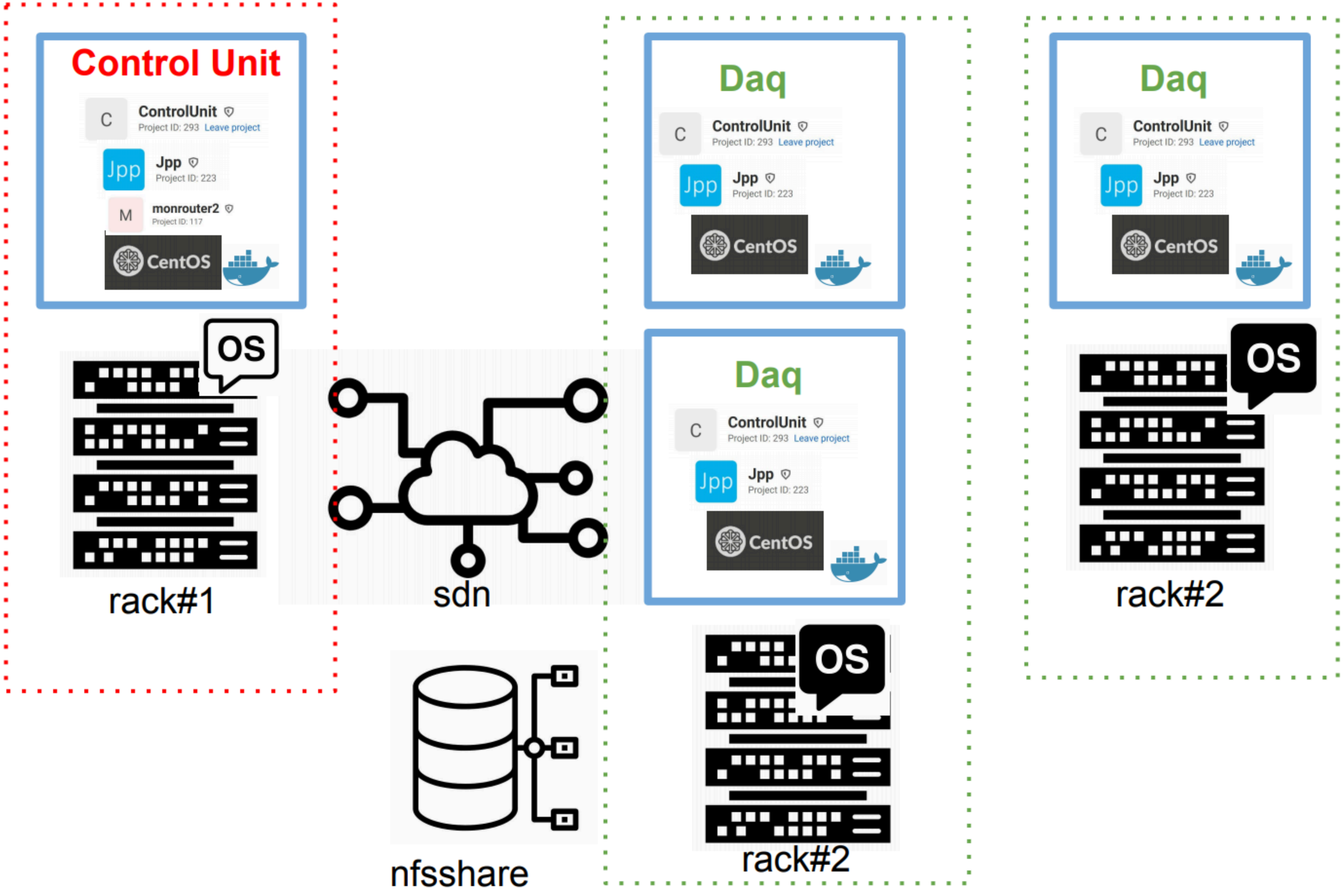
The **ToT is directly related to the pulse amplitude**. So it is a proxy for the number of photoelectrons.

The working point of the PMTs, with the optimal gain (~10^6), implies a **ToT of 26,4 ns for single photo-electron**

When long pulses exceed 256ns, if the ***Multi-Hit* feature** is active, the original hit is fragmented in subsequent pieces.

In TDC channel: apparently more hits than the actually occurred!
In MON channel: the correct rate is reported.




Control Unit
ControlUnit
Project ID: 293 Leave project
Jpp
Project ID: 223
monrouter2
Project ID: 117
CentOS
OS
rack#1
sdn
nfsshare
Daq
ControlUnit
Project ID: 293 Leave project
Jpp
Project ID: 223
CentOS
Daq
ControlUnit
Project ID: 293 Leave project
Jpp
Project ID: 223
CentOS
OS
rack#2
Daq
ControlUnit
Project ID: 293 Leave project
Jpp
Project ID: 223
CentOS
OS
rack#2

Optical data for Physics

Case	n_{DU}	n_{DOMs}	$n_{pmt/DOM}$	ν_{single}/PMT (kHz)	hit size (bit)	$\nu_{trigger}$ (Hz)	Event window (μs)
KM3NeT–Ph1, It	24	18	31	15	50	40	6
KM3NeT–Ph1, *Fr*	7	18	31	15	50	13	6
KM3NeT–1 Block (Ph2, *Fr*)	115	18	31	15	50	220	6
KM3NeT–2 Blocks (Ph2, *It*)	230	18	31	15	50	440	6

Case	DOM thp (Mb/s)	DU thp (Gb/s)	Det thp (Gb/s)	Sel thp (MB/s)	Sel thp (TB/day)	Stored (TB/y)	event size(kB)
KM3NeT–Ph1, It	23.0	0.4	10.0	1.6	0.13	49.0	7.5
KM3NeT–Ph1, *Fr*	23.0	0.4	2.9	0.4	0.03	12.0	2.2
KM3NeT–1 Block (Ph2, *Fr*)	23.0	0.4	48.0	14.0	1.20	440.0	36.0
KM3NeT–2 Blocks (Ph2, *It*)	23.0	0.4	96.0	44.0	3.80	1400.0	72.0

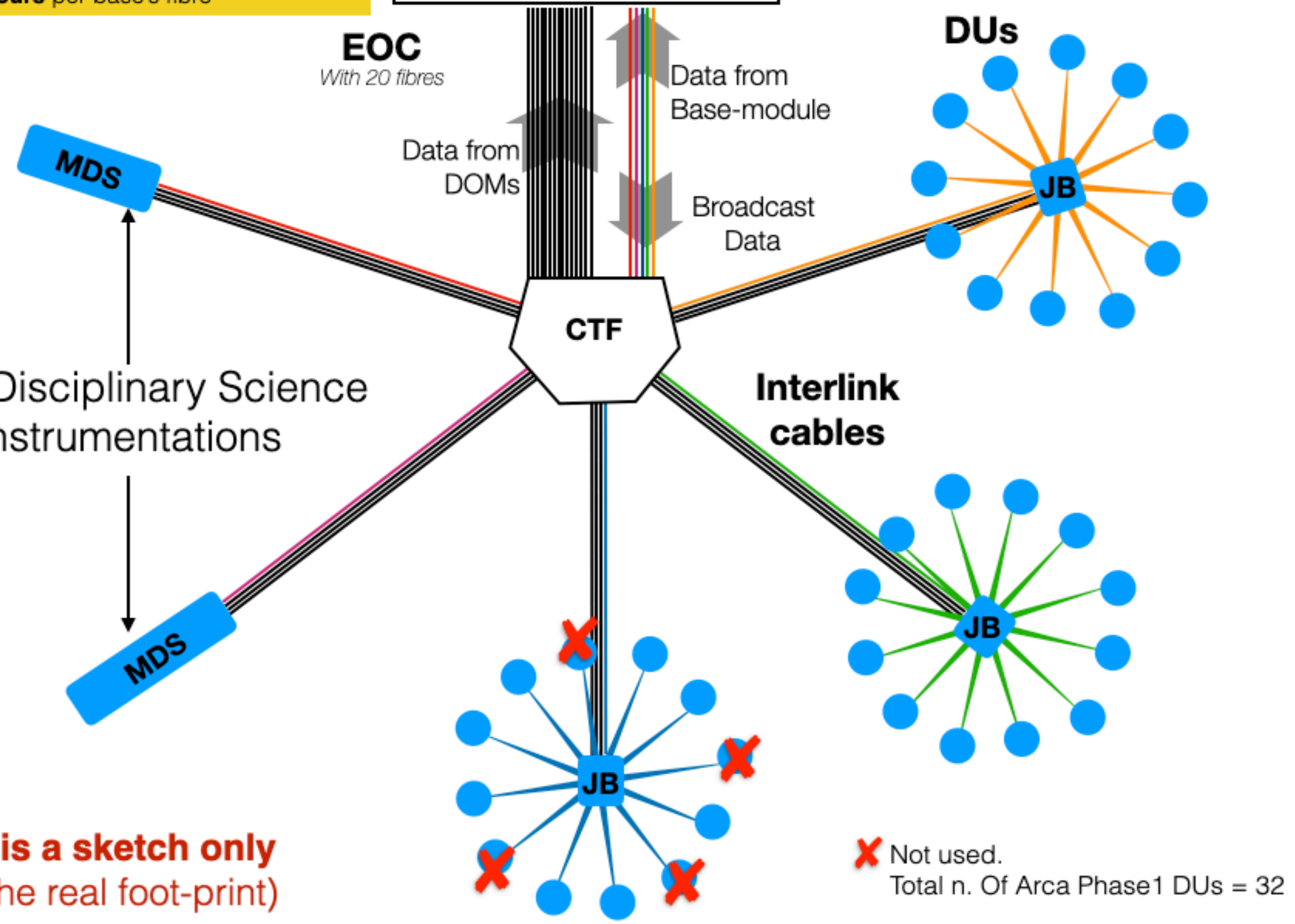
Acoustic data for positioning

Case	Raw Thp/Sensor (Mb/s)	Raw Thp/DU (Mb/s)	Raw Thp/Detector (Gb/s)	TOA (Mb/s)	Positions (Mb/s)	Storage (TB/y)
Phase 1–*It*	13.0	240.0	5.7	0.20	0.08	1.10
Phase 1–*Fr*	13.0	240.0	1.7	0.06	0.02	0.32
1 Block, Ph2 *Fr*	13.0	240.0	27.0	0.94	0.38	5.20
2 Blocks, Ph2 It	13.0	240.0	55.0	1.90	0.75	10.00

```
SamplingRateHz = 195.3 × 10^3;
ResolutionBit = 24;
NChannels = 2;
```

DAQ Overview: ARCA - Phase I sea infrastructure (asymmetric network)

up to 80 colours / fibre (spaced bye 50 MHz) used:
- **72 colours** per DOM's fibre
- **12 colours** per base's fibre



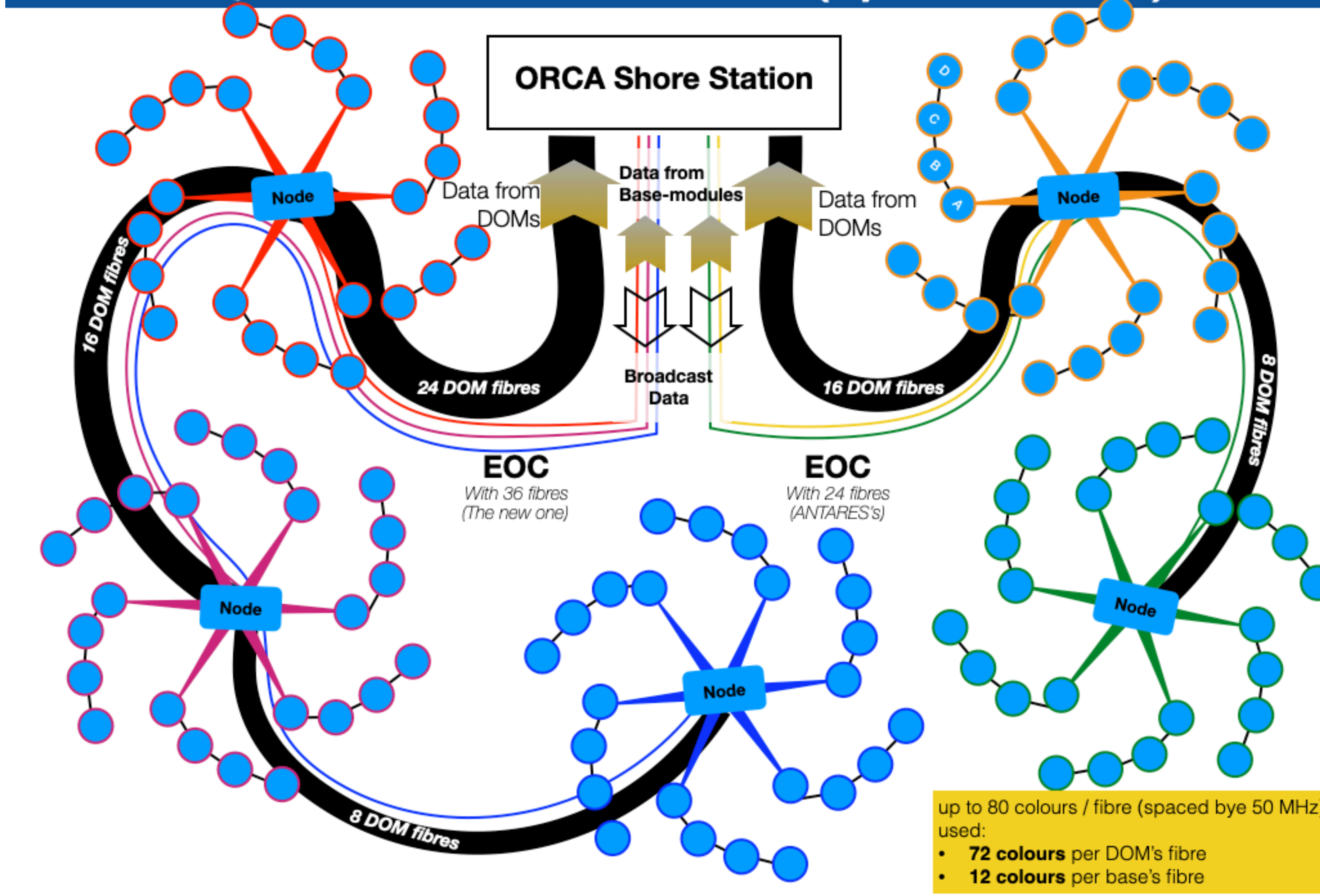
This is a sketch only
(not the real foot-print)

32 ARCA DUs only
(ARCA-Phase 1)

. Optical Connection in the "Broadcast Scenario"

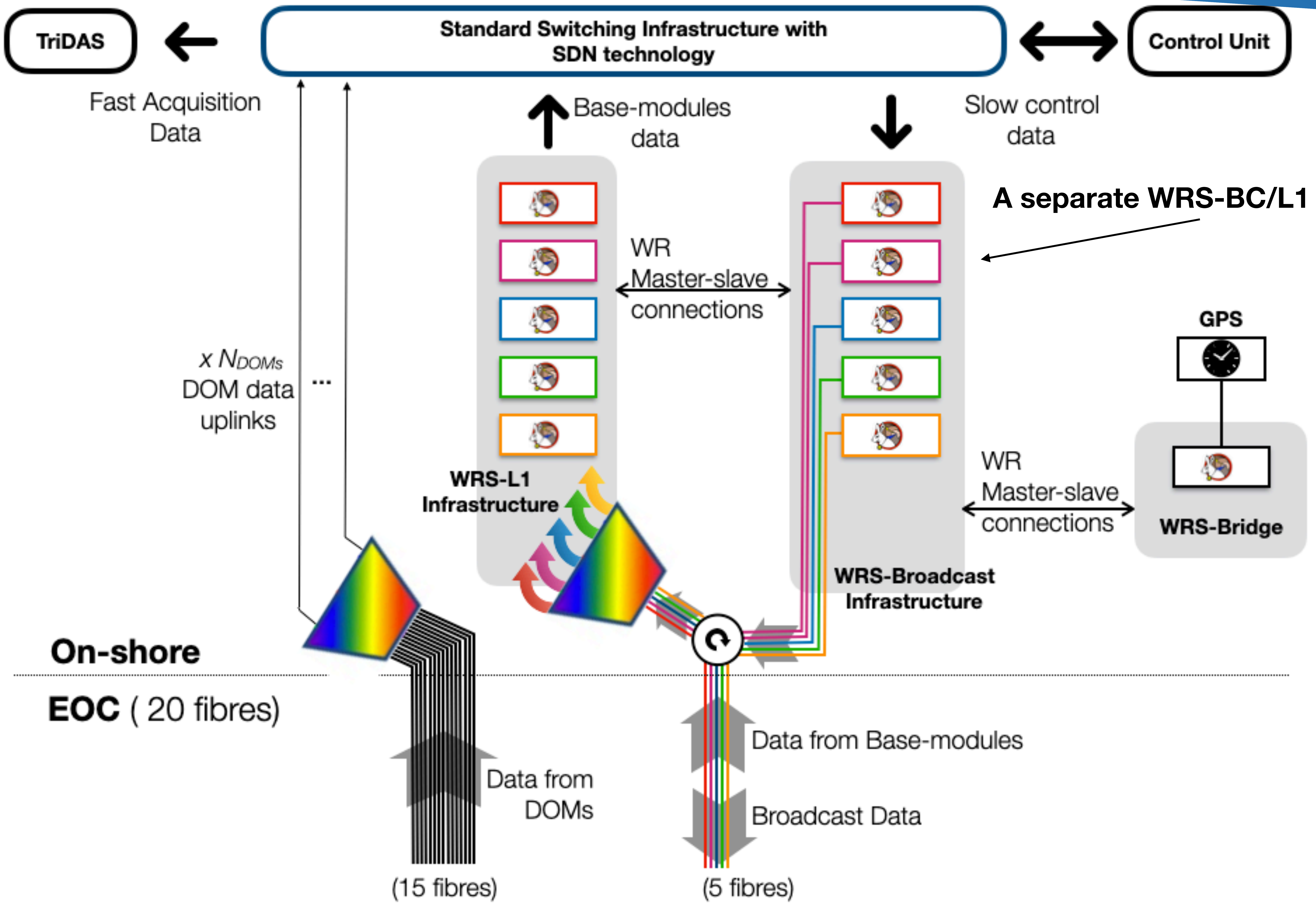
115 ORCA DUs

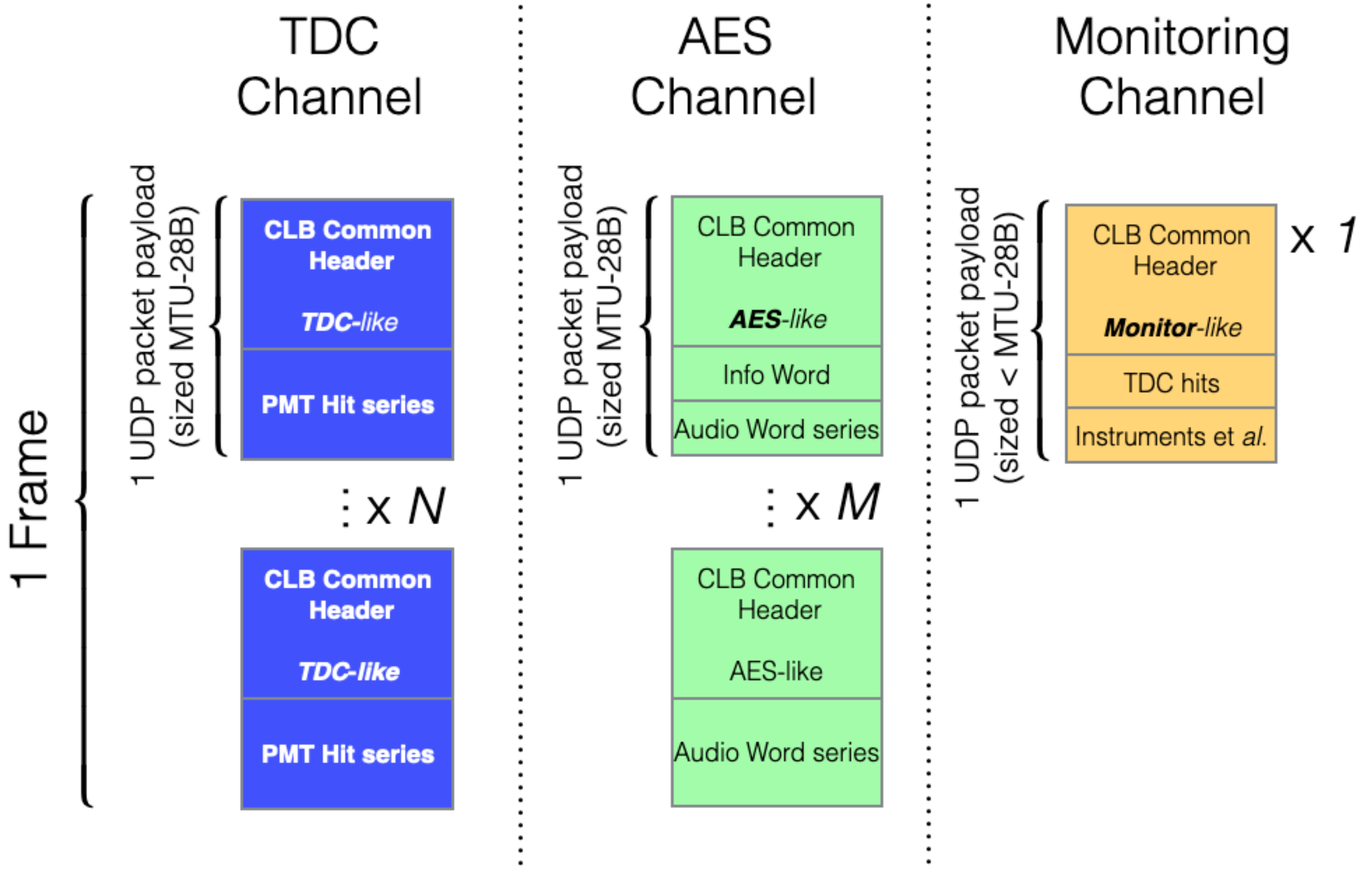
DAQ Overview: ORCA sea infrastructure (asymmetric network)

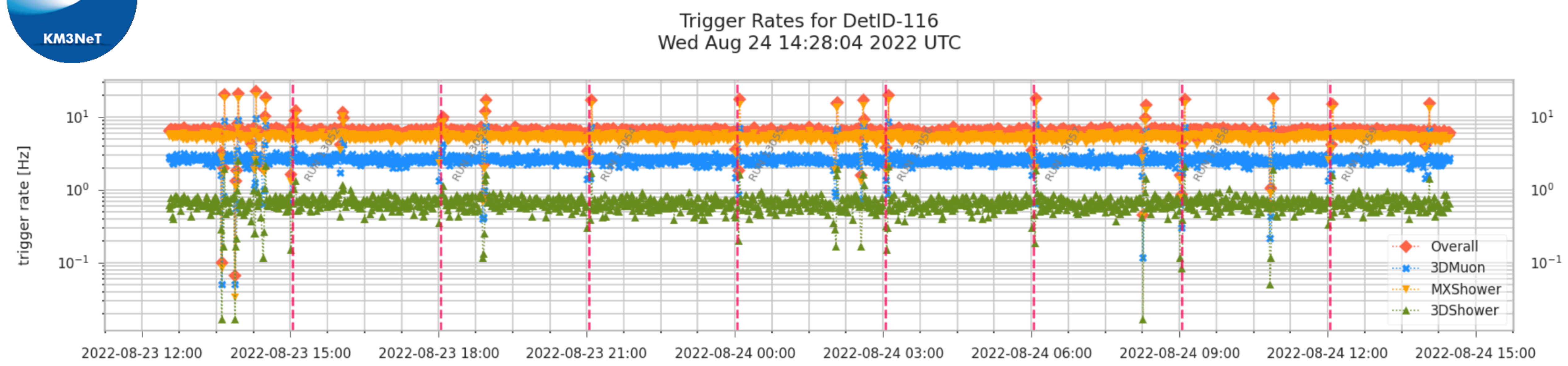


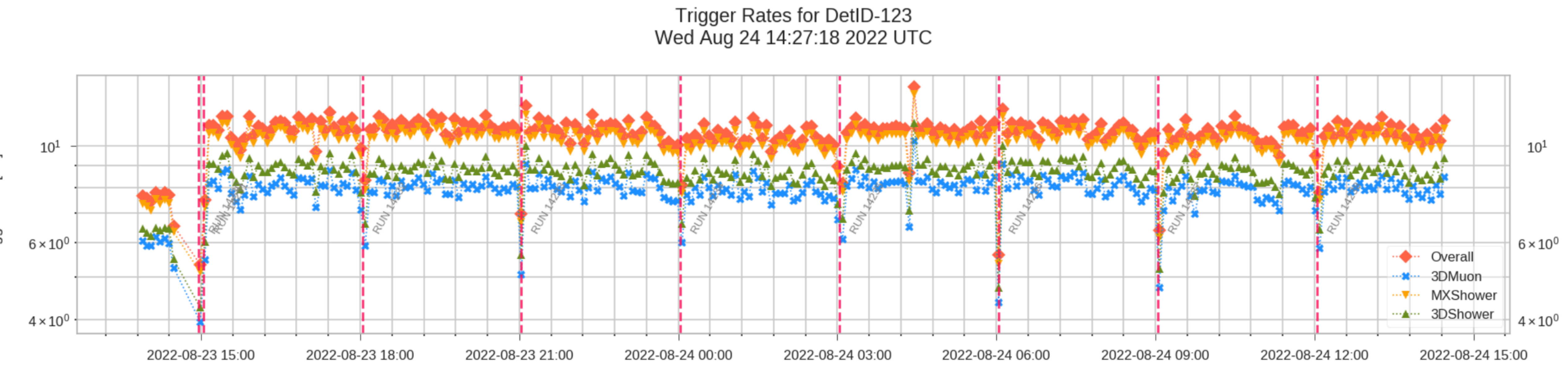
up to 80 colours / fibre (spaced bye 50 MHz) used:
- **72 colours** per DOM's fibre
- **12 colours** per base's fibre

A separate WRS-BC/L1 branch for each JB

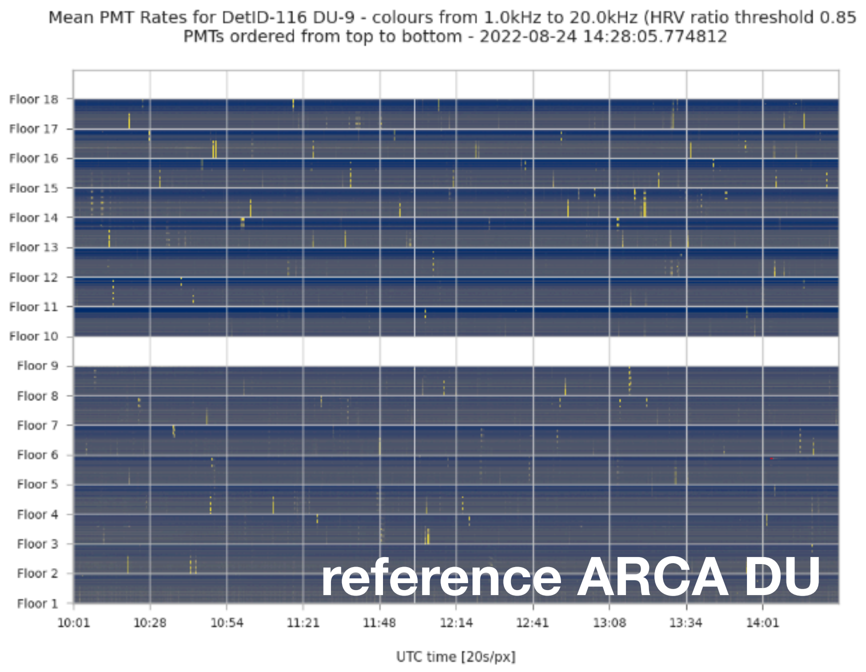


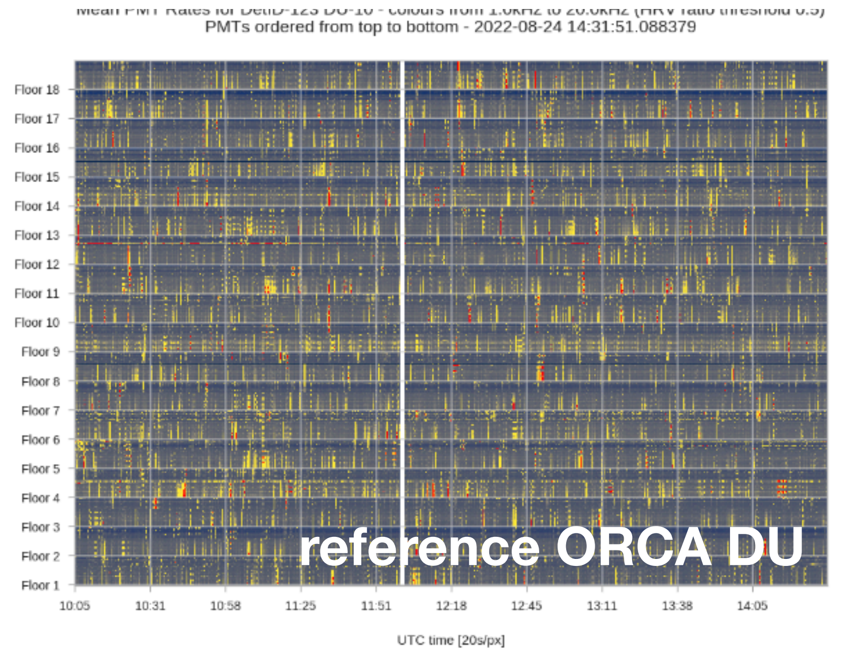

TDC Channel
AES Channel
Monitoring Channel
1 Frame
1 UDP packet payload (sized MTU-28B)
CLB Common Header
TDC-like
PMT Hit series
x N
CLB Common Header
TDC-like
PMT Hit series
1 UDP packet payload (sized MTU-28B)
CLB Common Header
AES-like
Info Word
Audio Word series
x M
CLB Common Header
AES-like
Audio Word series
1 UDP packet payload (sized < MTU-28B)
CLB Common Header
Monitor-like
TDC hits
Instruments et al.
x 1





PMT rates



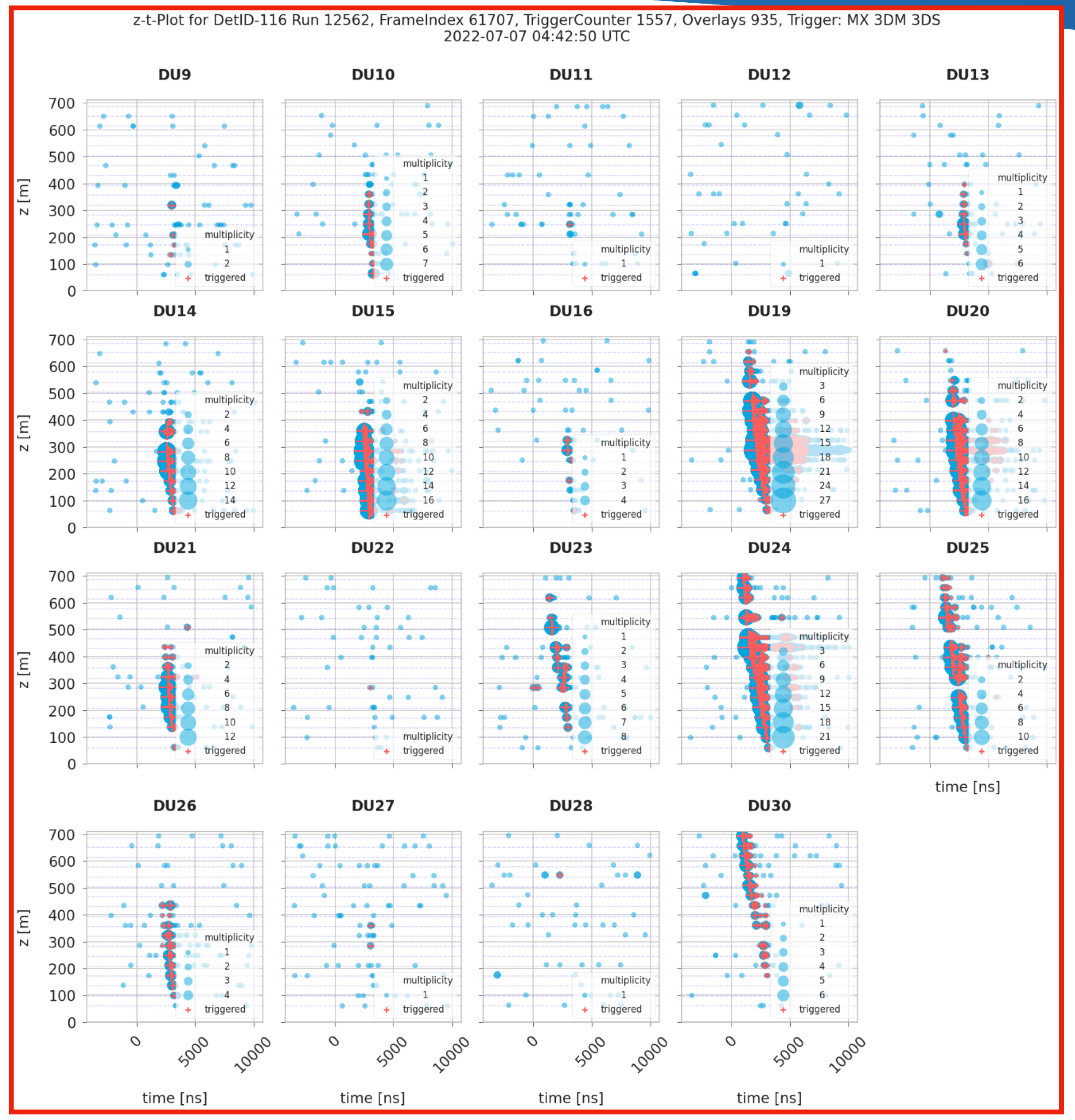


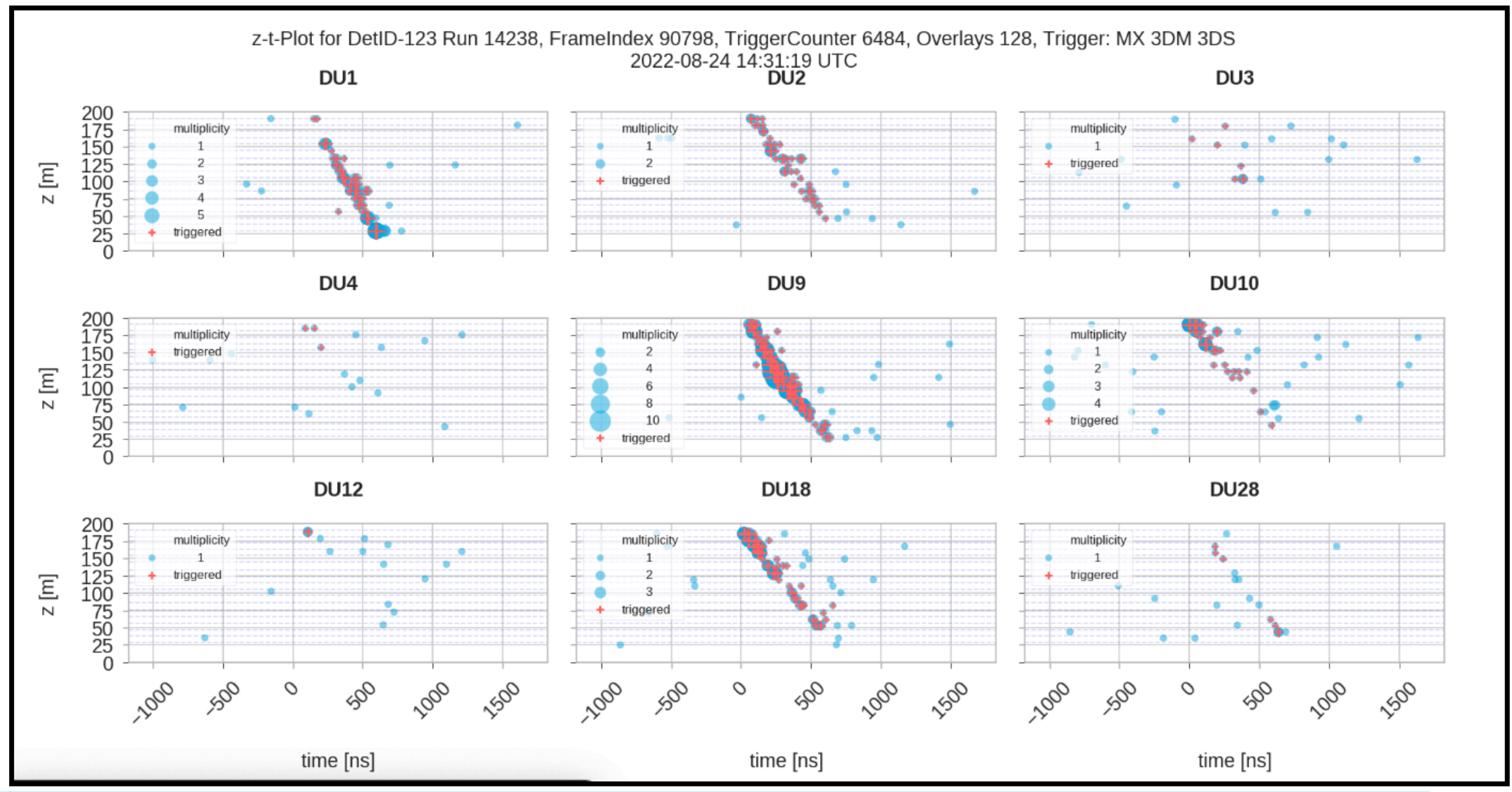
ARCA

@-3500m

ORCA

@-2500m





L0: all hits over threshold (i.e. all hits sent by the CLBs)

L1: pairs of hits of the same DOM within 25(10)ns.

L2-trigger level

http://svn.km3net.de/Jpp/trunk/documentation/JTrigger/

- **3D-Trigger** - general concept:

1. A minimum n. of ***consecutive*** L1s ≥ N_{th} within a ΔT (at least $n_{DOM} \geq 2$ or 5)
2. *3D*-causality filter : $$|t_i - t_j| \leq |\vec{x}_i - \vec{x}_j|\frac{n}{c} + T_{MaxExtra}$$
3. The trigger is set if the n. of satisfying hits is ≥ N'_{th}

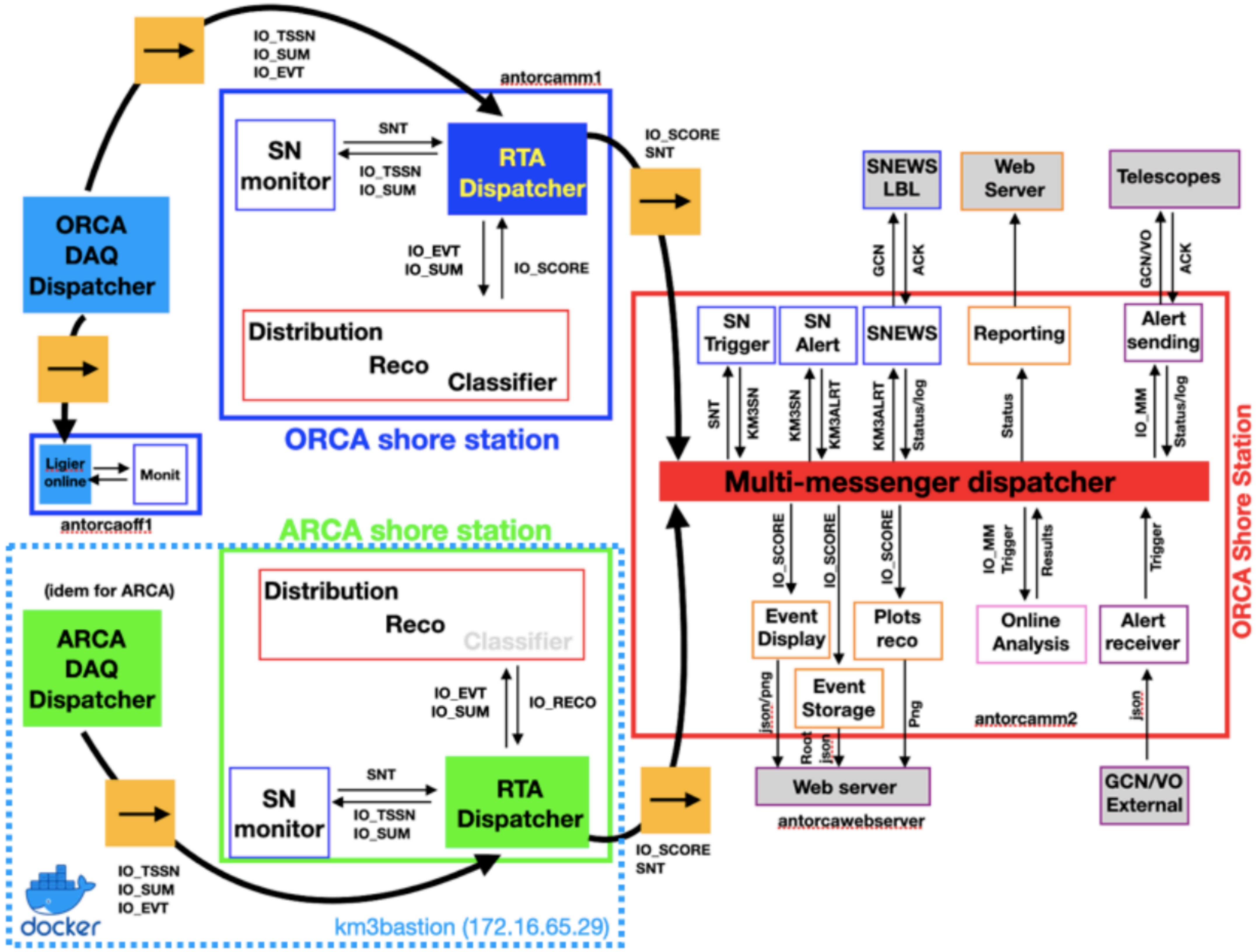
- **3D-Muon/Shower**
  Assumes an extended track-like / short pulse shape for the event topology

- **MX-Shower**
  Combines L0s and L1s within a limited space in the detector.

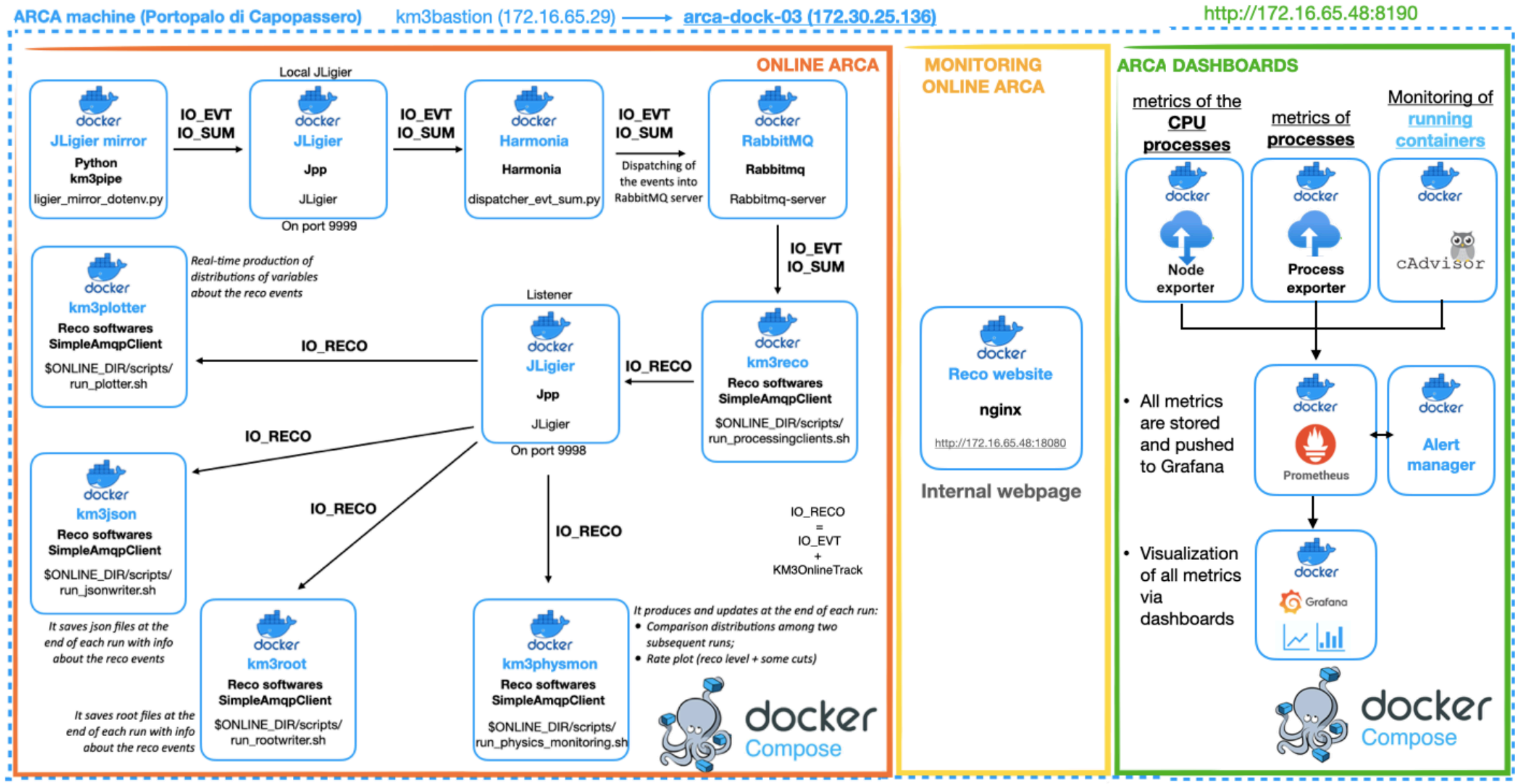
ctMin (minimal cosine of the angle between PMTs axis) used as alignment parameter

Trigger settings passed to the Data Filters via the run setups by the Control Unit

Trigger algorithms developed within a large C++ software framework, ***Jpp***. The same codes are used for the on-line DAQ as well as off-line analysis.



- Event processing done separately for ARCA and ORCA at each shore station
- Same **processing structure** but different software organisation (in ARCA the docker approach is adopted).
- The output of the reconstructed events by ARCA and ORCA at the end of each run (.json files) is stored in a common dispatcher (MM dispatcher)



AIACE is a collection of *playbook* for the installation and configuration of the computing resources and network devices. It makes use of *ANSIBLE* for building and multiplexing the command to be executed on the various servers. This is used to coordinate both the OS and software deployment as well as their configuration.

DOCKER images, one for each DAQ process, are compound of the dedicated libraries and service software.
The deployment of DAQ processes is handled via AIACE, and consists of the creation of many independent Docker containers on the due needed servers.

Docker images includes also ControlUnit services, like the LAP to apply the **Dynamic Provisioning Manager** (DPM) which "keep-alive" the DAQ processes and organises the roles of the various servers

At present, both for **ARCA19** and **ORCA11**:

- 2x DataQueue processes (on 2 independent server)
- 30x Optical Data Filter processes (on 2 independent server)
- 1x Acoustic Data Filter (on 1 independent server)
- 1x DataWriter together with 1 DataDispatcher (on 1 independent server)