A Large Ion Collider Experiment

ALICE Computing System for Run 3 and Run 4: the O² Project

Pierre VANDE VYVRE

Outline

- ALICE upgrade during LS2
- O^2 Project Requirements and architecture
  - Physics programme and requirements
  - Architecture
- O^2 Hardware facility
  - Technologies
  - Components
- O^2 Software
  - Software framework
  - Calibration, reconstruction and data volume reduction
  - System simulation
- O^2 in the Grid

ALICE Upgrade

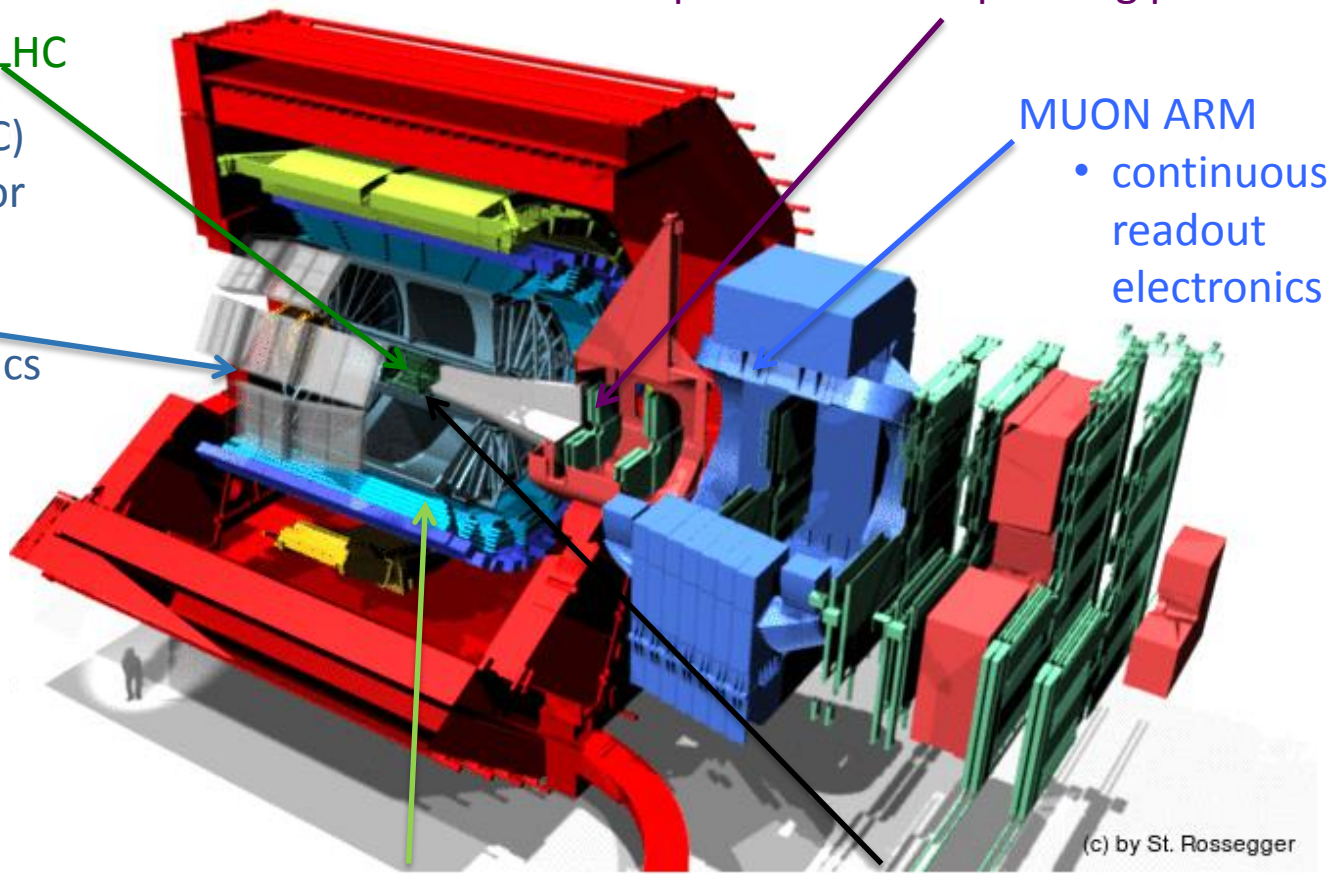
LHC after LS2: Pb–Pb collisions at up to L = 6 10^{27} $cm^{-2}s^{-1}$ ⇒ interaction rate of 50kHz

New Inner Tracking System (ITS)
- improved pointing precision
- less material -> thinnest tracker at the LHC

Muon Forward Tracker (MFT)
- new Si tracker
- Improved MUON pointing precision

Time Projection Chamber (TPC)
- new GEM technology for readout chambers
- continuous readout
- faster readout electronics

MUON ARM
- continuous readout electronics



New Central Trigger Processor

Entirely new Data Acquisition (DAQ)/ High Level Trigger (HLT)

TOF, TRD, ZDC
- Faster readout

New Trigger Detectors (FIT)

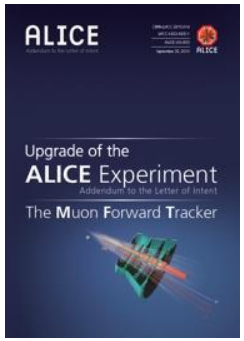
Physics programme and data taking scenarios

Year	System	$\sqrt{s_{NN}}$ (TeV)	L_{int} (pb^{-1})	L_{int} (nb^{-1})	$N_{collisions}$
2020	pp	14	0.4		$2.7 \cdot 10^{10}$
	Pb-Pb	5.5		2.85	$2.3 \cdot 10^{10}$
2021	pp	14	0.4		$2.7 \cdot 10^{10}$
	Pb-Pb	5.5		2.85	$2.3 \cdot 10^{10}$
2022	pp	14	0.4		$2.7 \cdot 10^{10}$
	pp	5.5	6		$4 \cdot 10^{11}$
2025	pp	14	0.4		$2.7 \cdot 10^{10}$
	Pb-Pb	5.5		2.85	$2.3 \cdot 10^{10}$
2026	pp	14	0.4		$2.7 \cdot 10^{10}$
	Pb-Pb	5.5		1.4	$1.1 \cdot 10^{10}$
	p-Pb	8.8		50	10^{11}
2027	pp	14	0.4		$2.7 \cdot 10^{10}$
	Pb-Pb	5.5		2.85	$2.3 \cdot 10^{10}$

ALICE Upgrade for Run 3 and 4

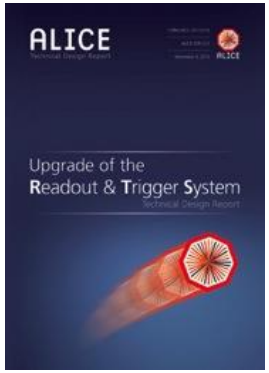
- ALICE goal:
  - Integrate a luminosity of 13 nb^{-1} for Pb–Pb collisions recorded in minimum bias mode
  - Together with dedicated p–Pb and pp reference runs
- Major detector upgrade:

Inner Tracking System (ITS)



Muon Forward Tracker (MFT)

Time Projection Chamber (TPC)



Readout and Trigger system

- New Online-Offline Computing System (O^2)

Read-out rate and data rate with Pb-Pb beams

Detector	Max read-out rate (kHz)	Data rate for Pb-Pb collisions at 50kHz (GB/s)	Fraction of the total input data rate (%)	Average data size per interaction (MB)
ACO	100	0.014		0.00028
CPV	50	0.9		0.018
CTP	200	0.02		0.0004
EMC	42	4		0.08
FIT	100	0.115		0.023
HMP	2.5	0.06		0.024
ITS	**100**	**40**	**1.2**	**0.8**
MCH	100	2.2		0.04
MFT	**100**	**10**	**0.3**	**0.2**
MID	100	0.3		0.006
PHS	42	2		0.04
TOF	200	2.5		0.05
TPC	**50**	**3276**	**97.6**	**20.7**
TRD	**90.9**	**20**	**0.6**	**0.5**
ZDC	100	0.06		0.0012
Total		**3358**		**22.5**

ALICE O^2 in a nutshell

Requirements

1. LHC min bias Pb-Pb at 50 kHz
   ~100 x more data than during Run 1
2. Physics topics addressed by ALICE upgrade
   - Rare processes
   - Very small signal over background ratio
   - Needs large statistics of reconstructed events
   - Triggering techniques very inefficient if not impossible
3. 50 kHz > TPC inherent rate (drift time ~100 µs)
   Support for continuous read-out (TPC)
   - Detector read-out triggered or continuous

New computing system

- Read-out the data of all interactions

➔ Compress these data intelligently by online reconstruction

➔ One common online-offline computing system: O^2

- Paradigm shift compared to approach for Run 1 and 2

Unmodified raw data of all interactions shipped from detector to online farm in triggerless continuous mode

HI run 3.3 TByte/s ⇩

Baseline correction and zero suppression
Data volume reduction by zero cluster finder.
No event discarded.
Average compression factor 6.6

500 GByte/s ⇩

Data volume reduction by online tracking. Only reconstructed data to data storage.
Average compression factor 5.5

90 GByte/s ⇩

Data Storage: 1 year of compressed data
- Bandwidth: Write 90 GB/s Read 90 GB/s
- Capacity: 60 PB

20 GByte/s ⇕

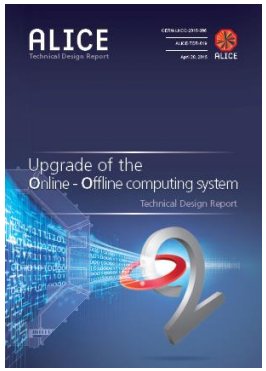
Tier 0, Tiers 1 and Analysis Facilities

⇕

Asynchronous (few hours) event reconstruction with final calibration

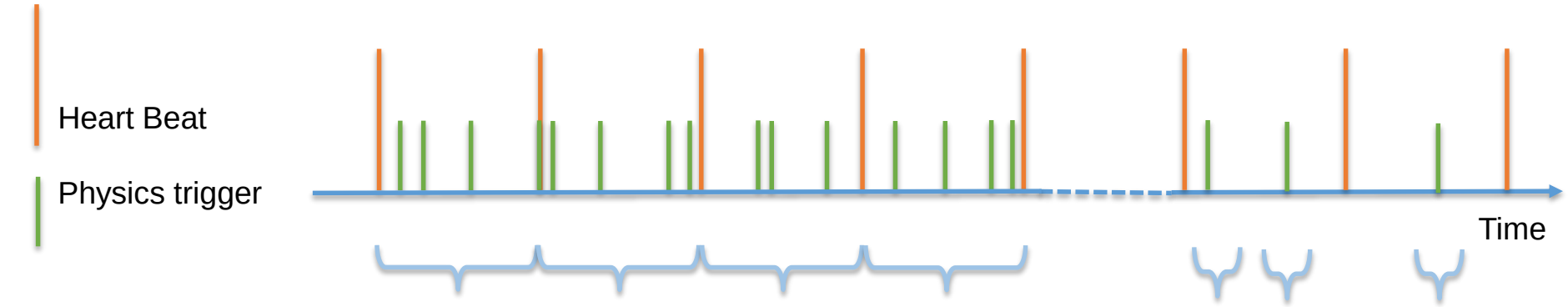
ALICE Computing Upgrade for Run 3 and 4

- New common Online-Offline (O^2) computing system
- Data volume reduction by data processing
  - HLT system had been designed to support different ways to reduce the data volume
  - RoI identification and event selection never been used
  - HLT major impact with the data volume reduction by cluster finder.
  - Since 2011, ALICE does NOT record the raw data
- O^2 system part of the data grid : export and import of workload
- Improve the efficiency on the grid
  - Analysis is the least efficient of all workloads that we run on the Grid
  - Analysis Facility: dedicated sites where AODs are collected and processed quickly



O^2 Online and Offline computing
https://cds.cern.ch/record/2011297/files/ALICE-TDR-019.pdf

Triggers, Heart Beats, Timeframe etc



Heart Beat Frames (HBF): stream of data delimited by two heart beats

Trigger data fragments

Sub-Time Frame (STF): grouping of all consecutive HBFs during a time period

Subevents

Time Frame (TF): grouping of all STFs from all FLPs for the same time period with the triggered subevents for the same time period

Compressed Time Frame (CTF): TF compressed after processing

Data flow & processing (1)

Raw data input

Local processing

Frame dispatch

Global processing

Storage

Synchronous

Detectors electronics

TPC ... ITS ... TRD

Trigger and clock

Detector data samples synchronized by heartbeat triggers

FLPs Buffering O(100)

Local aggregation

QC

Time slicing

Sub-Time Frames

Data Reduction 0

e.g. clustering

Tagging

Calibration 0

on local data, ie. partial detector

First-Level Processors (FLPs)

Partially compressed sub-Time Frames

Load balancing & dataflow regulation

EPNs

Time Frame building

O(1000)

Full Time Frames

Detector reconstruction

e.g. track finding

Data Reduction 1

Calibration 1

on full detectors

e.g. space charge distortion

Event Processing Nodes (EPNs)

QC

Sub-Time Frames
Time Frames
Compressed Time Frames
AOD

Quality Control

QC data

Compressed Time Frames

O²/T0/T1

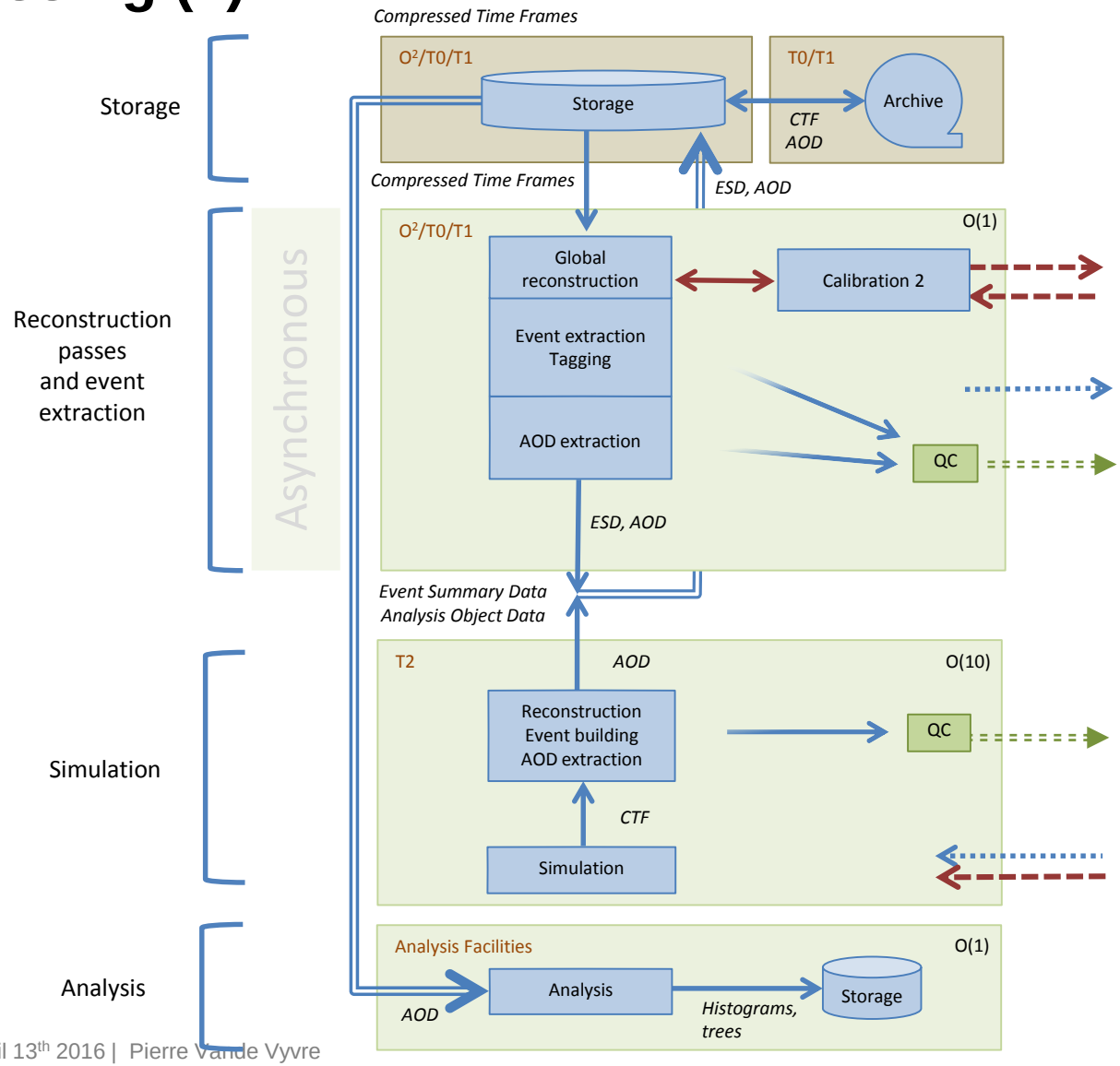
Storage

T0/T1

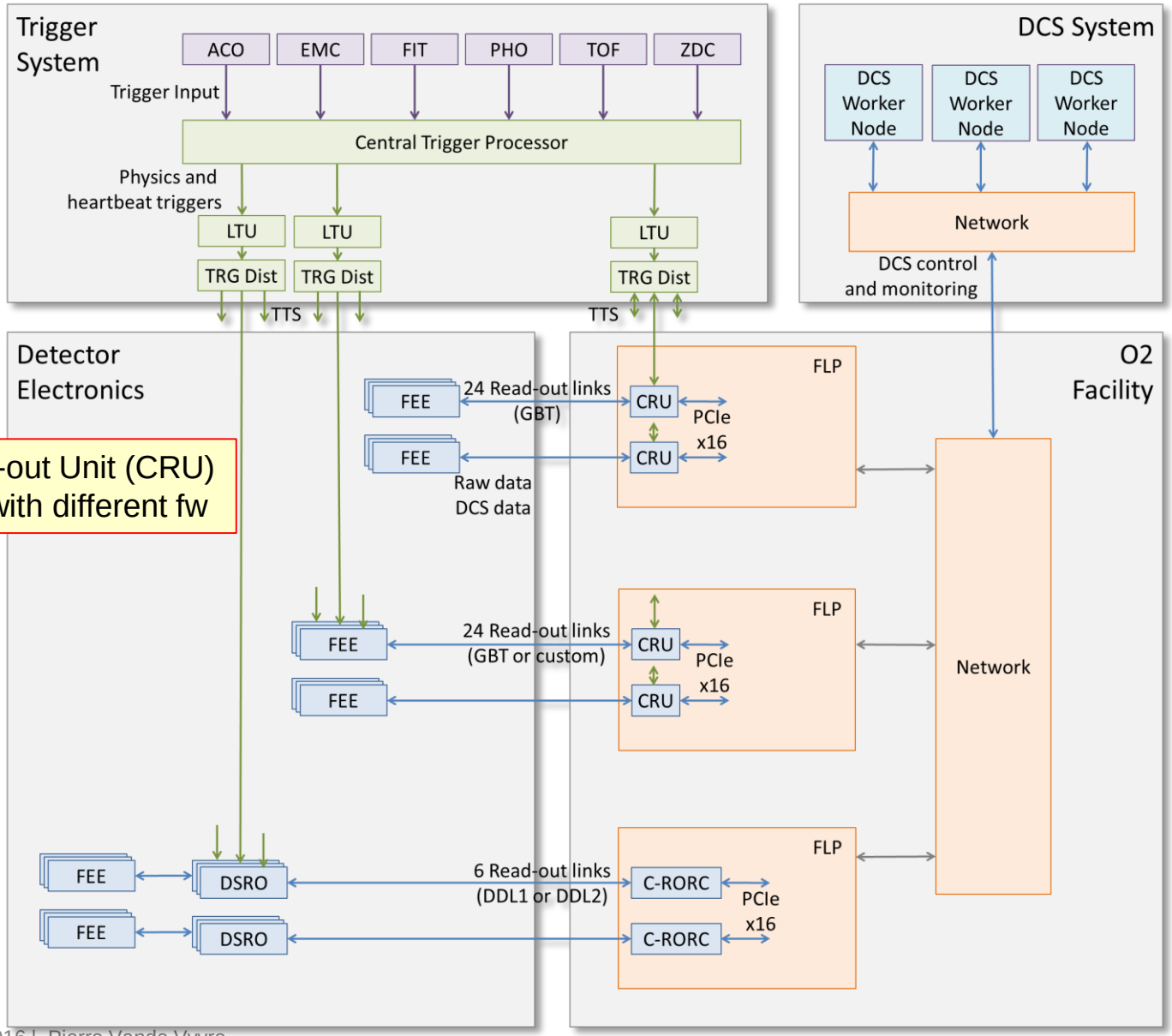
Archive

CTF
AOD

Data flow & processing (2)



Detector read-out – O² interfaces

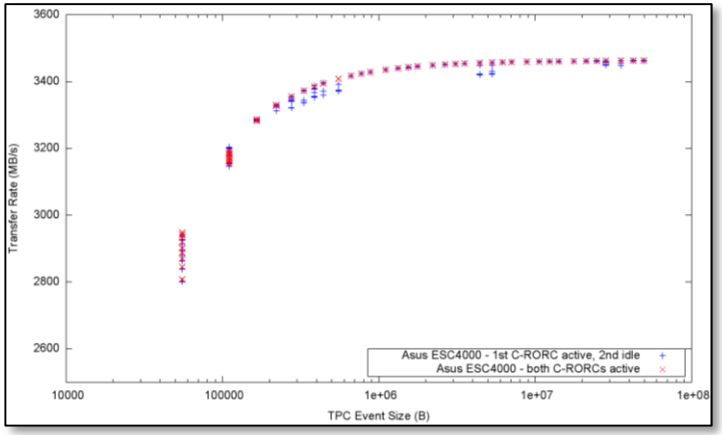


Common Read-out Unit (CRU)
LHCb PCIe40 with different fw

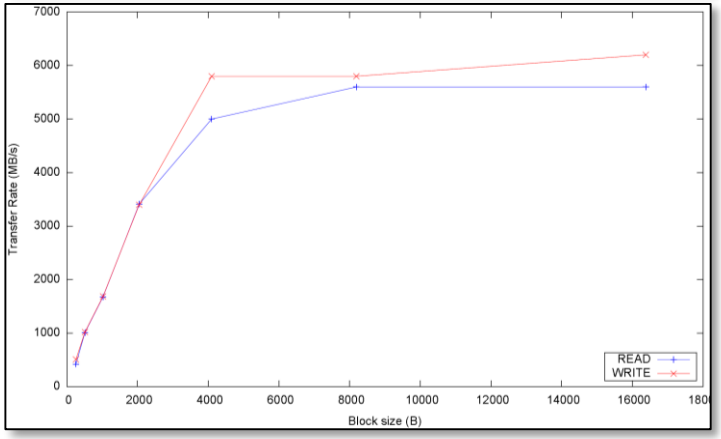
Detector links and read-out boards

Detector	Detector read-out links			Read-out boards	
	DDL1	DDL2	GBT	CRORC	CRU
ACO	1			1	
CPV	6			1	
CTP			14		1
EMC		20		4	
FIT		2		1	
HMP	14			4	
ITS			495		23
MCH			550		25
MFT			304		14
MID			32		2
PHS		16		4	
TOF			72		3
TPC			6552		360
TRD			1044		54
ZDC			1		1
Total	**21**	**38**	**9064**	**15**	**483**

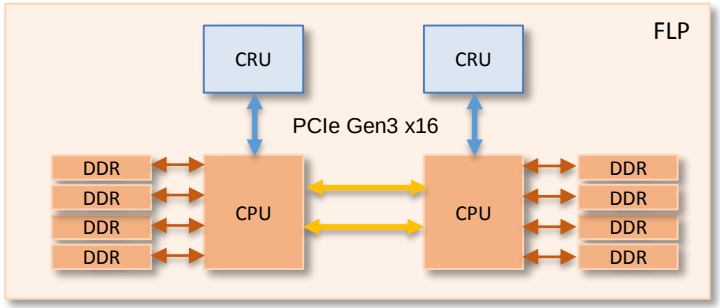
Technology: Input/Output with PCI Express



PCIe Gen2
- Up to 3.4 GB/s by device (x8) with 1 or 2 devices active slot
- Device independence

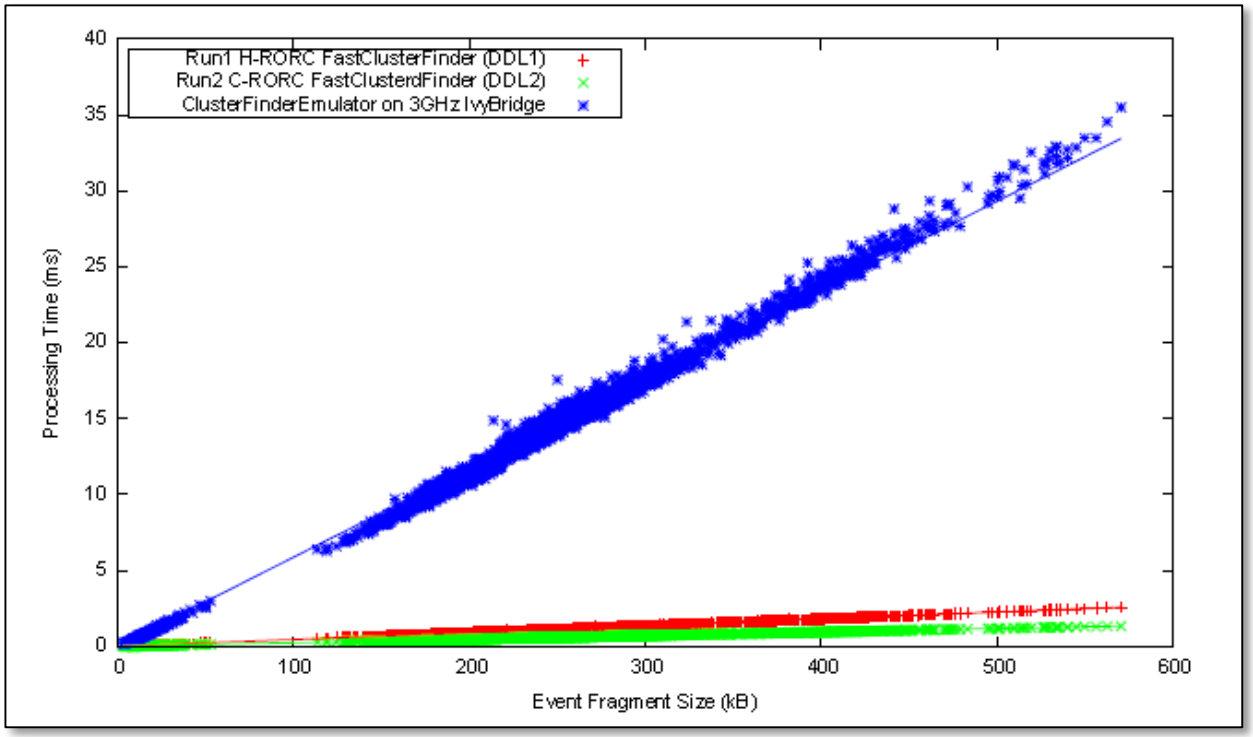


PCIe Gen3
- Up to 6 GB/s or 48 Gb/s for one x8 slot
- 1 slot PCIe Gen3 x16 delivers up to 96 Gb/s
- 1 CRU needs 20 GBTs x 4.0 Gb/s = 80 Gb/s of I/O bandwidth
- Gen3 is adequate for O^2 (x16)
- PCIe backward compatibility will allow to use Gen4 platforms or devices if desired



FLP Input-Output
- FLP with two CPUs:
  Each CPU : 16 PCIe Gen3 lanes (out of 40) used for the CRU interface

Technology: Hardware acceleration with FPGA



FPGA

- Acceleration for TPC cluster finder versus a standard CPU
- 25 times faster than the software implementation
- Use of the CRU FPGA for the TPC cluster finder

Facility Design: Read-Out boards and FLP

- Read-Out boards
  - C-RORC (# 15)
    - Input : up to 6 DDL1s and DDL2s
    - Output: PCIe Gen 2 x8
  - CRU (#483)
    - Input : up to 24 GBTs
    - Output: PCIe Gen 3 x16
    - Processing: TPC cluster finder (#360)
- FLP (#268)
  - Input: 1 or 2 read-out boards
    Network input for the DCS FLP
  - Output: up to 18 Gb/s for
    the TPC inner chamber FLP
  - Processing: ITS cluster finder (# 23)

Table 10.1: Number of read-out boards and FLPs per detector to O^2 system.

Detector	Number of read-out boards	Read-out board type	Number of FLPs
ACO	1	C-RORC	1
CPV	1	C-RORC	1
CTP	1	CRU	1
DCS	1	Network	1
EMC	4	C-RORC	2
FIT	1	C-RORC	1
HMP	4	C-RORC	2
ITS	23	CRU	23
MCH	30	CRU	15
MFT	14	CRU	7
MID	1	CRU	1
PHS	4	C-RORC	2
TOF	3	CRU	3
TPC	360	CRU	180
TRD	54	CRU	27
ZDC	1	CRU	1
Total			268

Technology: Hardware acceleration with GPU

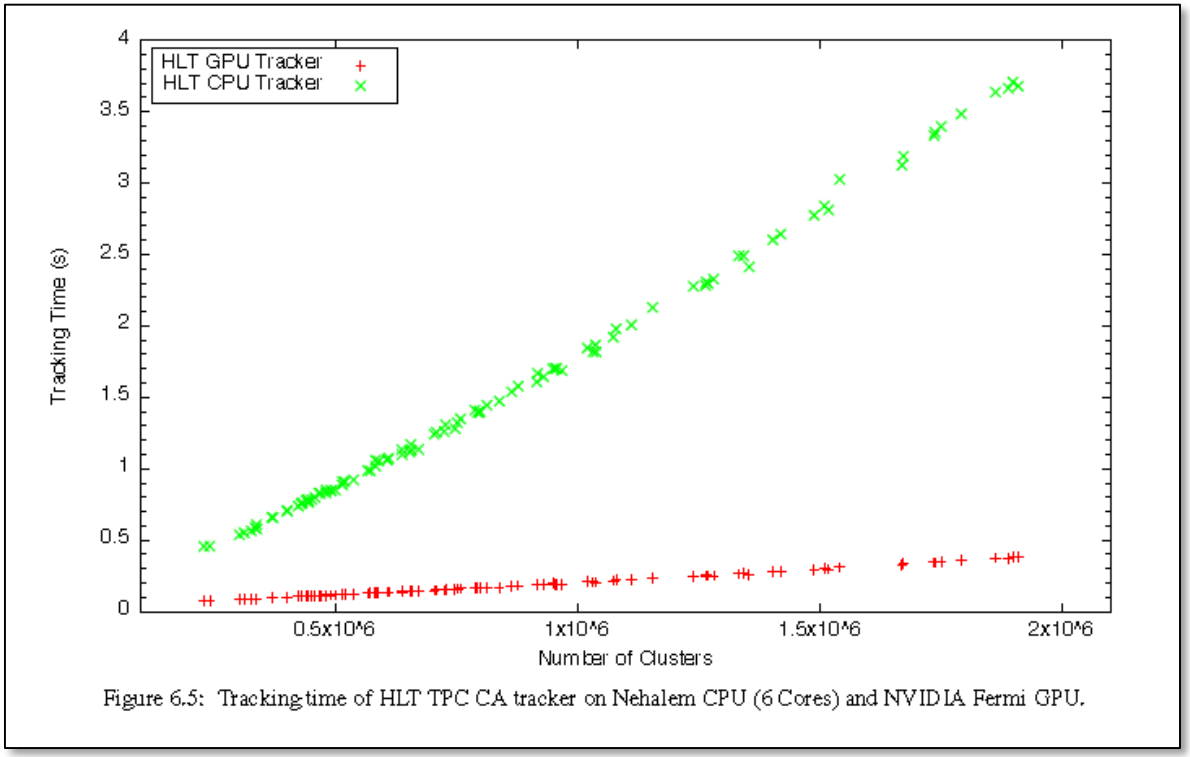


Figure 6.5: Tracking-time of HLT TPC CA tracker on Nehalem CPU (6 Cores) and NVIDIA Fermi GPU.

GPU

- TPC Track Finder based on the Cellular Automaton principle to construct track seeds.
- It is implemented for OpenMP (CPU), CUDA (Nvidia GPU), and OpenCL (AMD GPU).
- 1 GPU replaces 30 CPU cores and uses 3 for I/O
- Use of GPUs for the TPC track finder in the EPNs

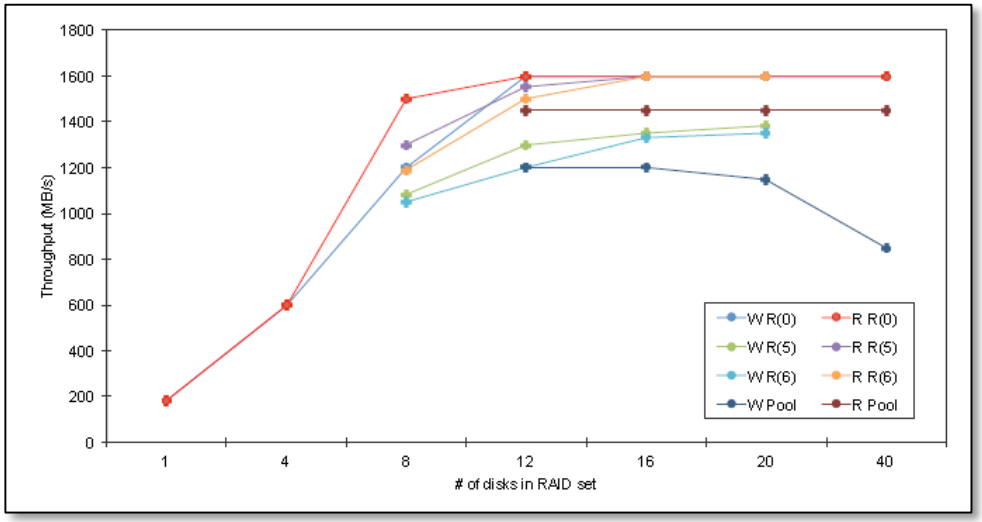
Facility Design: EPN and GPUs

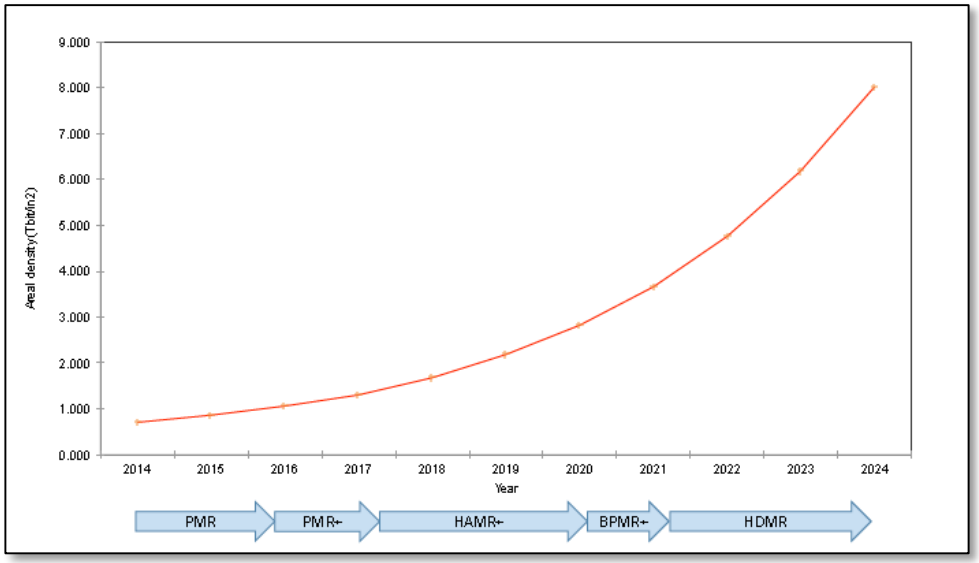
Table 8.8: CPU requirements for processing the data from Pb–Pb interaction at 50 kHz in the synchronous mode.

Detector	Process	Processing requirement [CPU cores or GPUs.]	Processing Platform	System reference
TPC	Calibration	1000	CPU	Intel I7-4600U 2.70 GHz
TPC	Track seeding, following	5000	GPU	AMD S9000
TPC	Track merging, fitting	15000	CPU	Intel I7-980X 3.60 GHz
ITS	Tracking	75000	CPU	Intel I7-2720QM 2.20 GHz
MCH	Preclustering	200	CPU	Intel I7 2.30 GHz
MCH	Clustering	5000	CPU	Intel I7 2.20 GHz

- EPN
  - 2 CPUs with at least 32 cores
    By then the most powerful CPUs will include 40 cores
  - 2 GPU cards each with 2 GPUs
  - 1500 nodes needed

Technology: data storage





- Hard disks enclosed into storage arrays of up to 80 hard disks in 4 U
- Possible technologies for storage arrays attachments
  - Fibre Channel: 8, 16 GB/s
  - ISCSI: depending on network speed
  - Infiniband: 40, 56 Gb/s
  - SAS: 12 Gb/s (through servers)
  - O^2 : data servers + storage arrays
- Bandwidth with a set of 12 hard disks
  - 1.4 GB/s write
  - 1.6 GB/s read
- Capacity
  - In 2015: 1.0 Tb/in^2, 6/8 TB disks for enterprise/consumer
  - In 2019: 2.2 Tb/in^2, 12/20 TB disks should be available

Facility Design: Data Storage

Table 4.4: Number of reconstructed collisions and storage requirements for different systems and scenarios.

Year	System	Collisions	Storage CTF (PB)	Storage Calibration (TB)	Storage ESD/AOD (PB)	Required CPU seconds (single CPU core)
2020	pp	$2.7 \cdot 10^{10}$	1.5	5	0.6	$1.7 \cdot 10^{10}$
	Pb–Pb	$2.3 \cdot 10^{10}$	37	23	15	$2.8 \cdot 10^{11}$
2021	pp	$2.7 \cdot 10^{10}$	1.5	5	0.6	$1.7 \cdot 10^{10}$
	Pb–Pb	$2.3 \cdot 10^{10}$	37	23	15	$2.8 \cdot 10^{11}$
2022	pp	$4.3 \cdot 10^{11}$	23	76	9.2	$2.7 \cdot 10^{11}$
2025	pp	$2.7 \cdot 10^{10}$	1.5	5	0.6	$1.7 \cdot 10^{10}$
	Pb–Pb	$2.3 \cdot 10^{10}$	37	23	15	$2.8 \cdot 10^{11}$
2026	pp	$2.7 \cdot 10^{10}$	1.5	5	0.6	$1.7 \cdot 10^{10}$
	Pb–Pb	$1.1 \cdot 10^{10}$	18	11	7.2	$1.3 \cdot 10^{11}$
	p–Pb	$1.0 \cdot 10^{11}$	10	20	4.0	$7.2 \cdot 10^{10}$
2027	pp	$2.7 \cdot 10^{10}$	1.5	5	0.6	$1.7 \cdot 10^{10}$
	Pb–Pb	$2.3 \cdot 10^{10}$	37	23	15	$2.8 \cdot 10^{11}$

	Capacity (PB)
CTF	38.5
ESD (1x)	5.8
AOD (2x)	7.7
Redundancy	8.3
Total	60.3

See talk of U. Fuchs today 14:35 “ALICE and LHCb storage system”

- Data Storage needs
  - Bandwidth: Write: 90 GB/s
    Read: 90 GB/s
  - Capacity: 60 PB
- 1 storage array in 2019:
  - 1.2 PB of raw capacity, 1.0 PB with redundancy
  - 7 GB/s of bandwidth
  - 68 units needed with 34 data servers

Technology: FLP-EPN networking

- 3 possible technologies:
  - Ethernet: 10, 40 and 100 Gb/s
    - 32 ports 100 Gb/s
  - Infiniband: 10, 20, 40, 56, 100 Gb/s
    - 36 ports EDR 100 Gb/s
  - Omni-Path: 100 Gb/s
- Available bandwidth on one node adequate for the O^2 needs (FLP TPC out: up to 20 Gb/s)
  - 270 ports at 40 Gb/s or more
  - 1500 ports at 10 Gb/s
- Technology choice: cost/performance
- Software framework: network agnostic

	Protocol	Maximum (GB/s)	Measured (GB/s)
IB QDR 40 Gb/s	Native IB verbs	4.0	3.9
IB FDR 56 Gb/s	Native IB verbs	6.8	5.6
IB FDR 56 Gb/s	IPoIB TCP	6.8	2.5
Eth 40 Gb/s	TCP	5.0	4.9

Hardware Facility

Detectors

270 FLPs
First Level
Processors
(FLPs)

1500 EPNs
Event Processing
Nodes
(EPNs)

34 Storage
Servers

68 Storage
Arrays

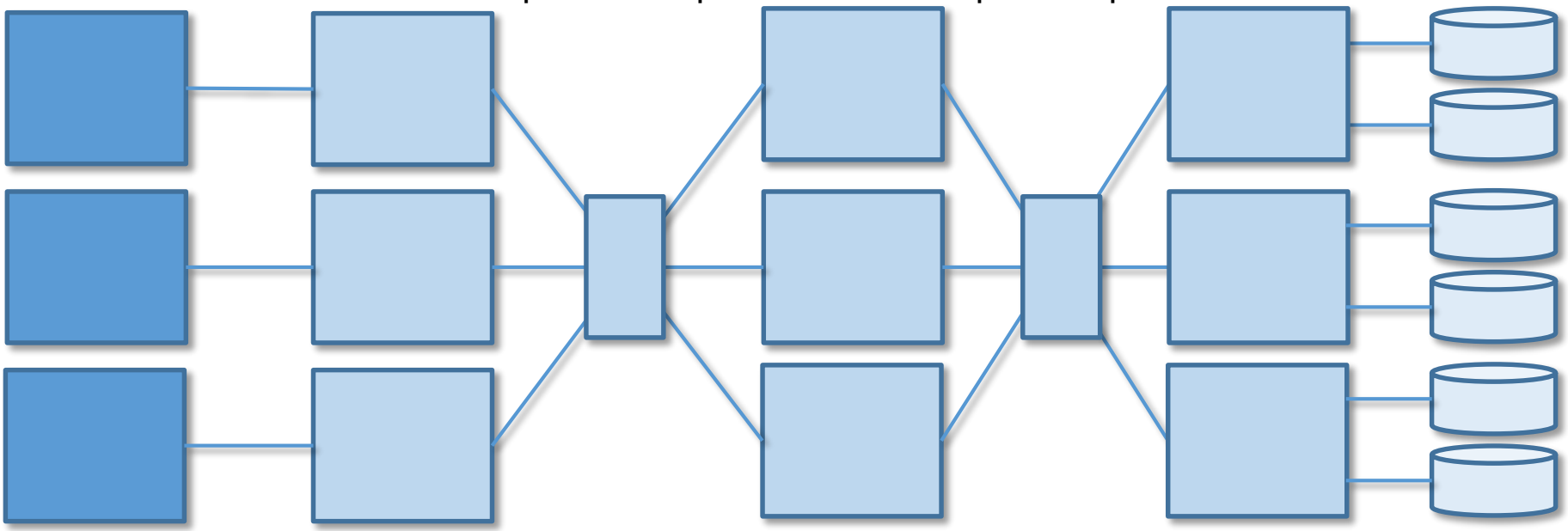
Switching
Network

Storage
Network

9000 Read-out
Links

Input: 270 ports
Output : 1500 ports

Input: 1500 ports
Output : 34 ports



3.3 TB/s

500 GB/s

90 GB/s

Facility Design: infrastructure

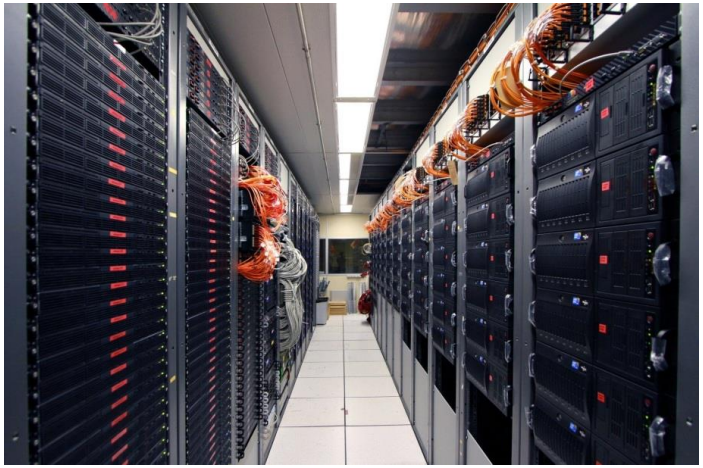
Table 10.5: Location, rack space, power and cooling needs of the O^2 facility.

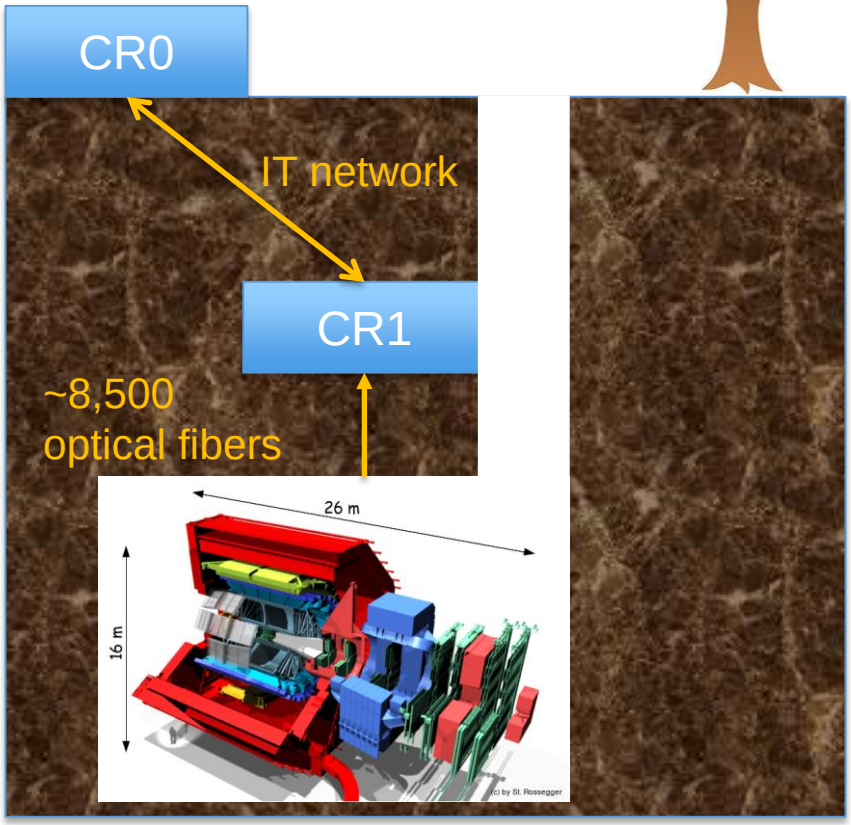
Type	Number of items	Item height (U)	Location-Total height (U)	Number of Racks	Power per rack (kVA)	Total power (kVA)	Cooling per rack (kW)	Total cooling (kW)
FLP	250	2	CR1 - 35	18	12	216	14	252
EPN	1500	1	CR0 - 54	34	50	1700	50	1700
SEPN	50	1	CR0 - 54	1	12	12	12	12
Storage	34	9	CR0 - 54	7	50	350	50	350
Network								
Dataflow	10	3	CR1 - 40	2	12	24	12	24
Dataflow	15	3	CR0 - 54	2	12	24	12	24
Control	4	1	CR1 - 40	1	12	12	12	12
Control	14	1	CR0 - 54	1	12	12	12	12
Services	110	1	CR1 - 40	4	12	48	12	48
Total			CR1	25		300		336
			CR0	45		2086		2086
Grand total				70		2398		2434

- O^2 facility located at the LHC P2
- Two Counting Rooms (CR) available each with 40 racks
- One room (CR1) renovated during LS1 will be reused
- New additional computing room (CR0) needed on the surface
- Existing power distribution at P2: UPS 500 kVA and 270 kW of normal power in each CR. Major upgrade needed.
- Installed cooling power in the CRs ~ 200kW. Major upgrade needed.

Computing Rooms

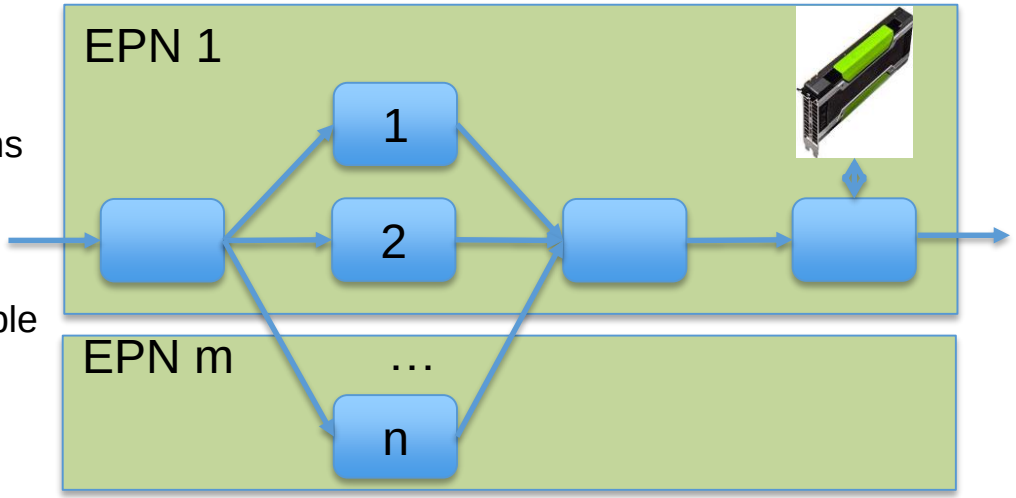
- CR0: new room and infrastructure needed on the surface
  - Market Survey MS-4108/EP – Survey of companies for the supply of Container Data Centers (24 companies contacted so far)
  - Water or free air cooled
  - Will be followed by the issue of an invitation to tender to qualified and selected firms in Q3/2016
  - Done with LHCb
- CR1:
  - Reuse existing room
  - Adequate power and cooling for the detector read-out

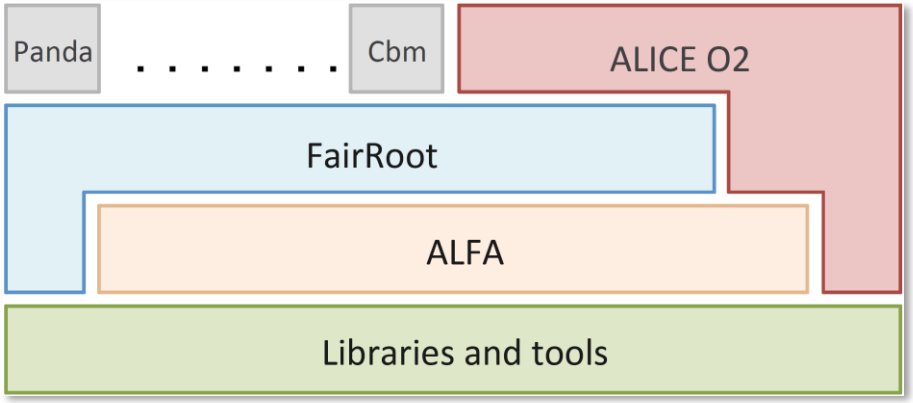




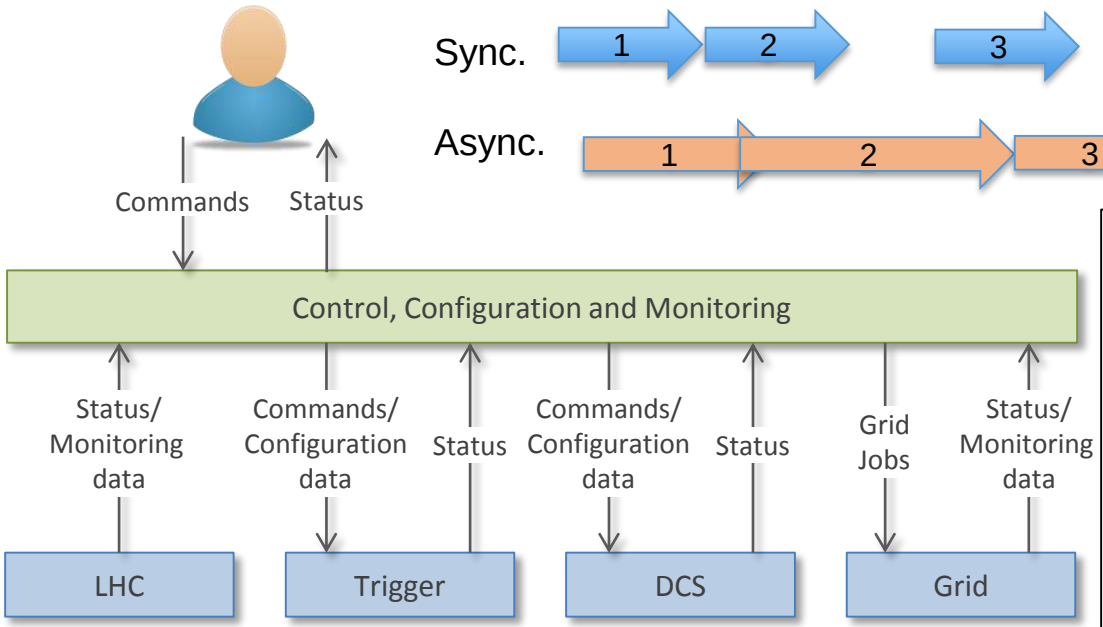
Software architecture

- Message-based multi-processing
  - Use of multi-core by splitting large programs into smaller size processes
    - Ease of development
    - Ease to scale horizontally
  - Multi-threading within processes still possible
  - Processes communicating by messages
  - Support for hardware acceleration
- ALFA
  - Prototype developed in common by experiments at FAIR and ALICE
  - Based on the ZeroMQ package
  - Data transport
  - Dynamic Deployment System





Control, configuration and monitoring



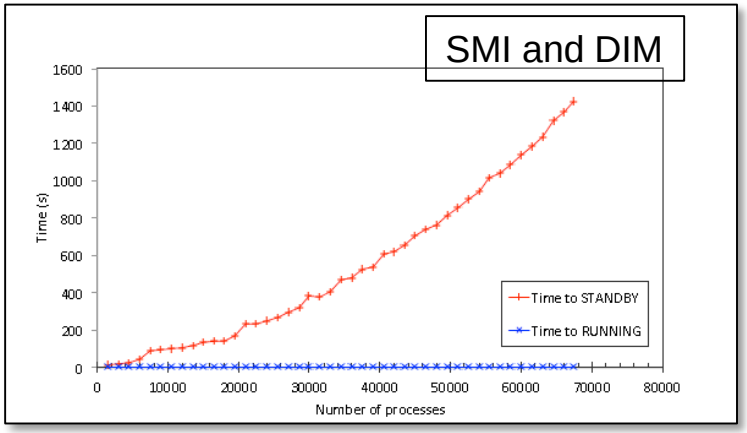
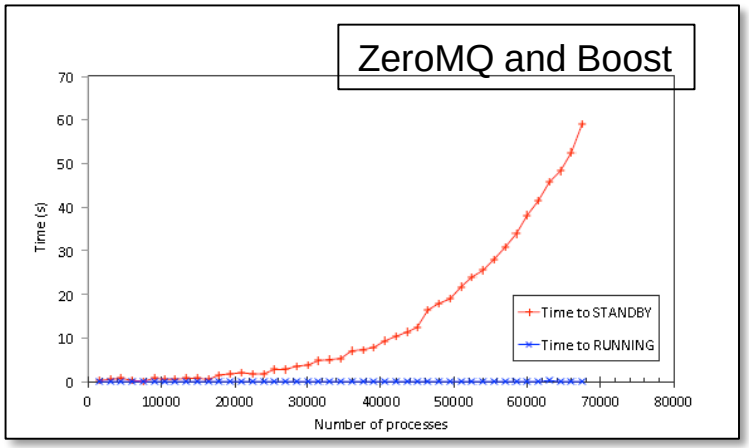
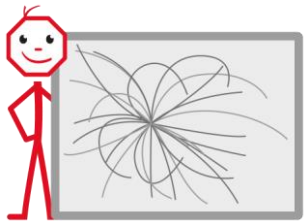


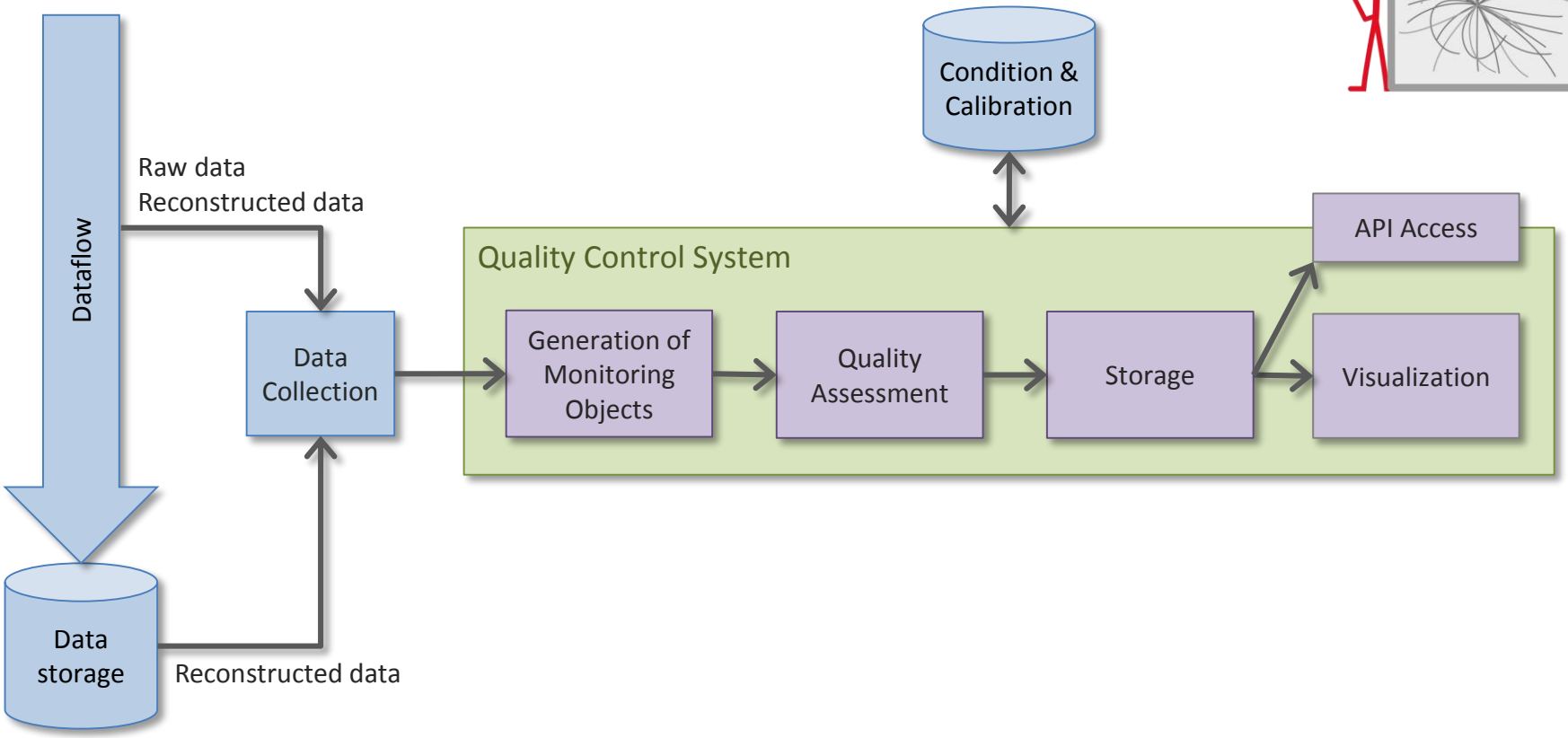
Table 6.7: Tools to implement CCM functions.

Module	Function	Tools
All	Inter Process Communication	DIM, ZeroMQ
Control	Start/stop processes	DDS
Control	Send commands to processes	SMI, ZeroMQ
Control	Task Management	SMI
Control	State Machine	SMI, Boost Meta State Machine
Control	Automation	SMI
Configuration	System Configuration Management	Puppet, Chef
Configuration	Configuration Distribution	ZooKeeper
Configuration	Dynamic Process Configuration	ZooKeeper
Monitoring	Data Collection and Archival	MonALISA, Zabbix
Monitoring	Alarms and Action Triggering	MonALISA, Zabbix

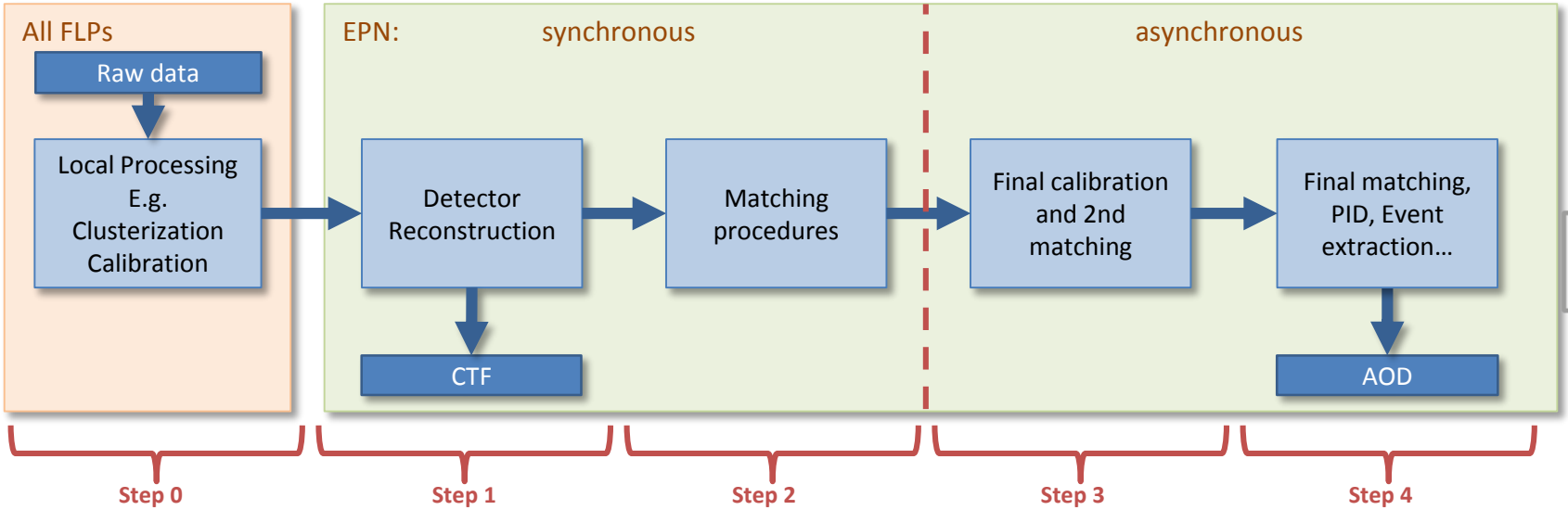


Quality control and visualization



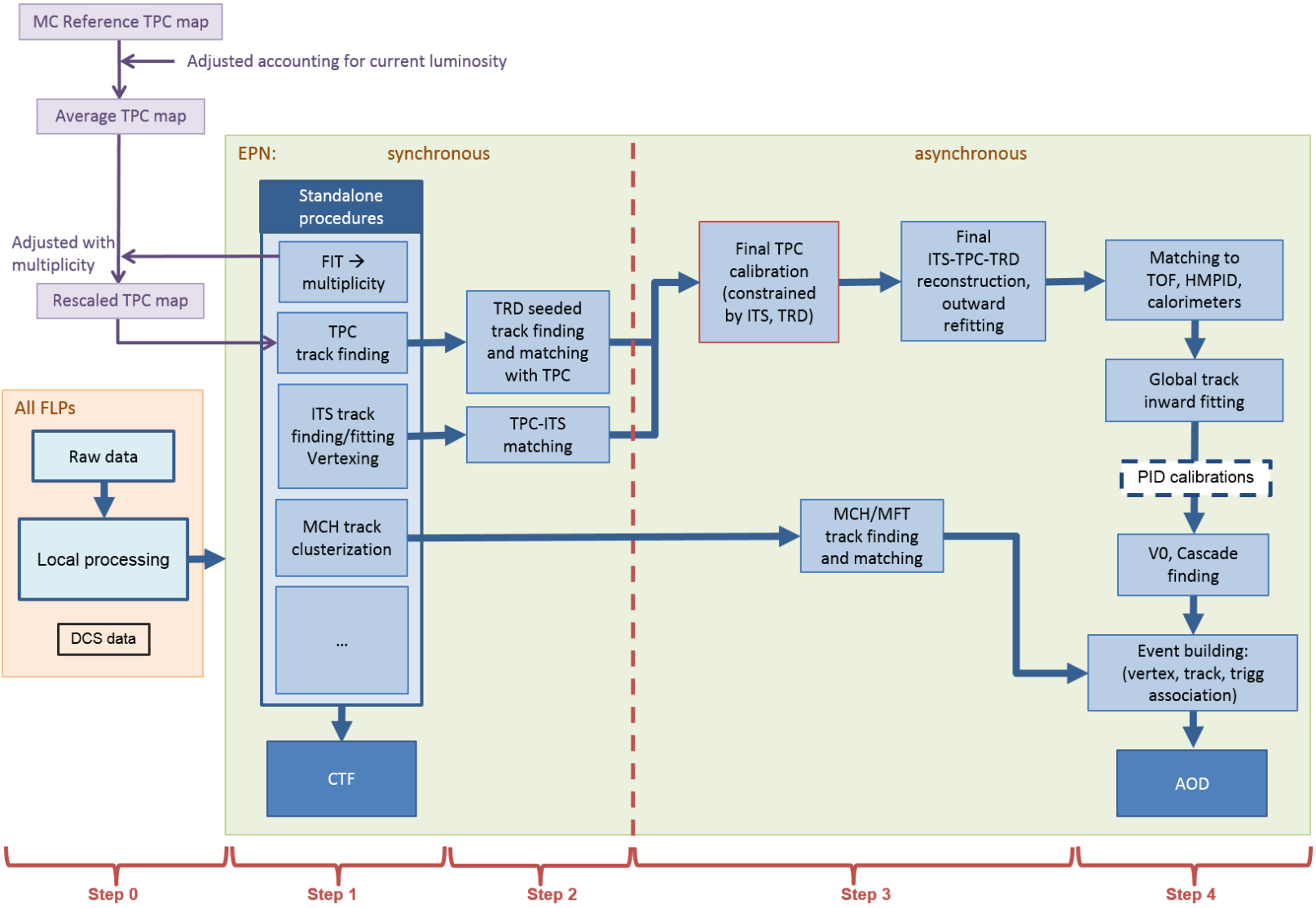


Calibration and reconstruction flow



	Compression Factor		Bandwidth (GB/s)		
	CRU/ FLP	EPN	FLP Input	FLP Output	EPN Output
TPC	2.5	8.0	1000	400	50
Other detec.	1.0	2.5	100	100	40
Total			1100	500	90

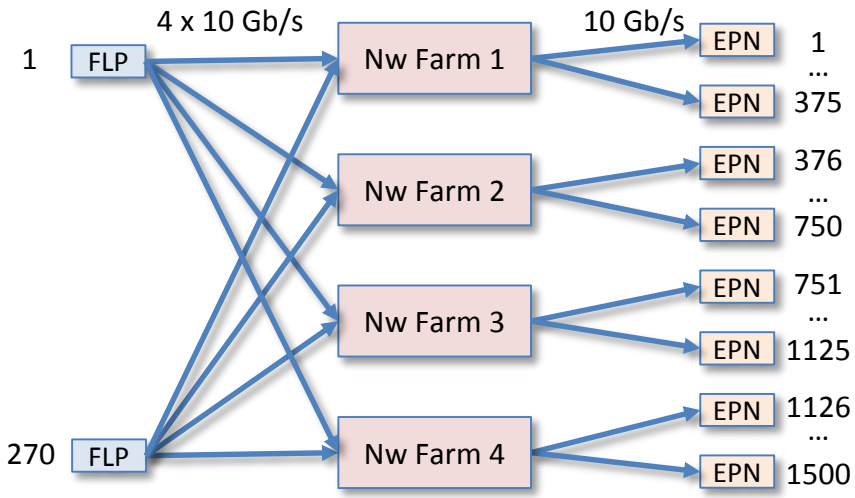
Calibration and reconstruction flow details

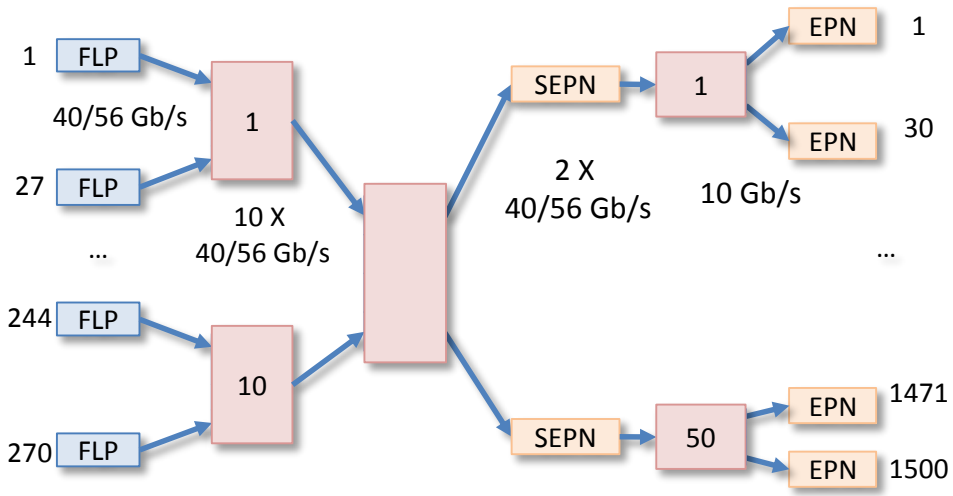


Facility Simulation: FLP-EPN network

See talk of I. Legrand & and M. Astigarraga today at 12:00 “Run 3 data flow systems simulation”

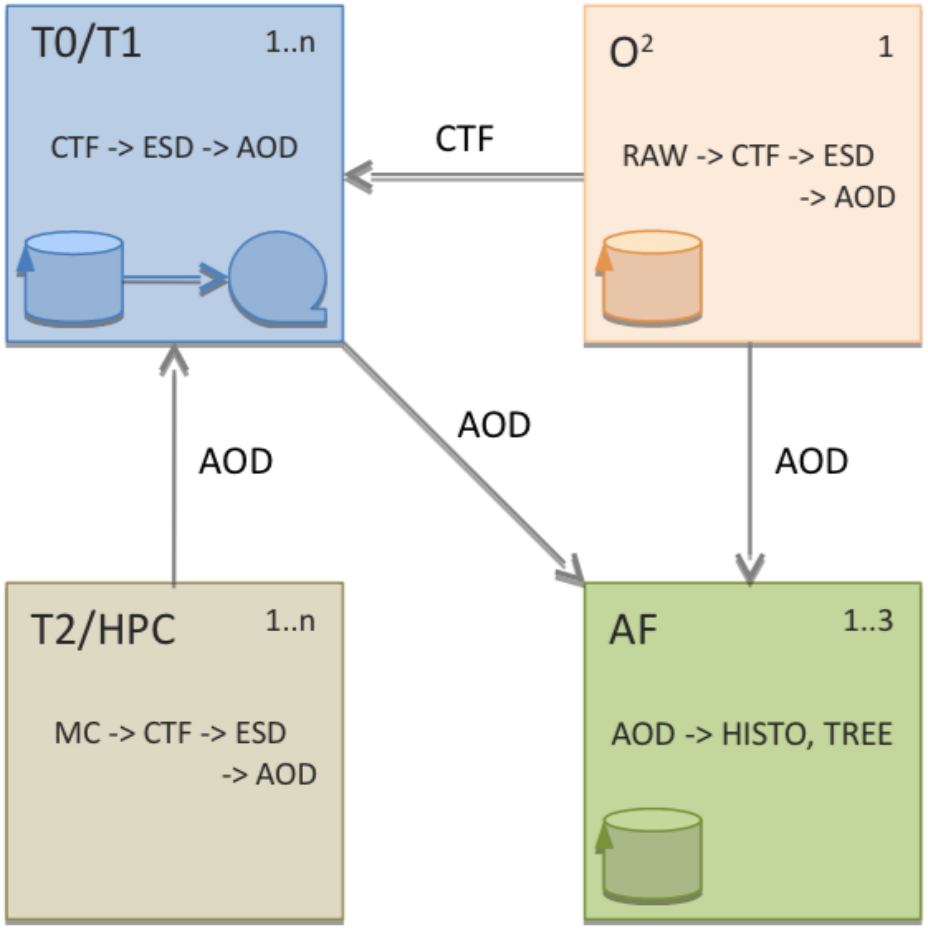
- 3 technologies:
  - Eth 10, 40, 100 Gb/s,
  - IB-FDR 56 Gb/s, IB-EDR 100 Gb/s
  - Intel Omniscale 100 Gb/s
- 270 FLP ports at 40/56/100 Gb/s + 1500 EPN ports at 10/40/56 Gb/s
- 3 network layout considered
  1. All ports at 40/56 Gb/s
  2. Limit the total throughput by splitting the system into 4 identical subsystems
  3. Limit the number of ports on a high-speed network by the use of Super-EPN





O2 System and the Data Grid

Reconstruction
Calibration
Archiving



Reconstruction
Calibration

Simulation

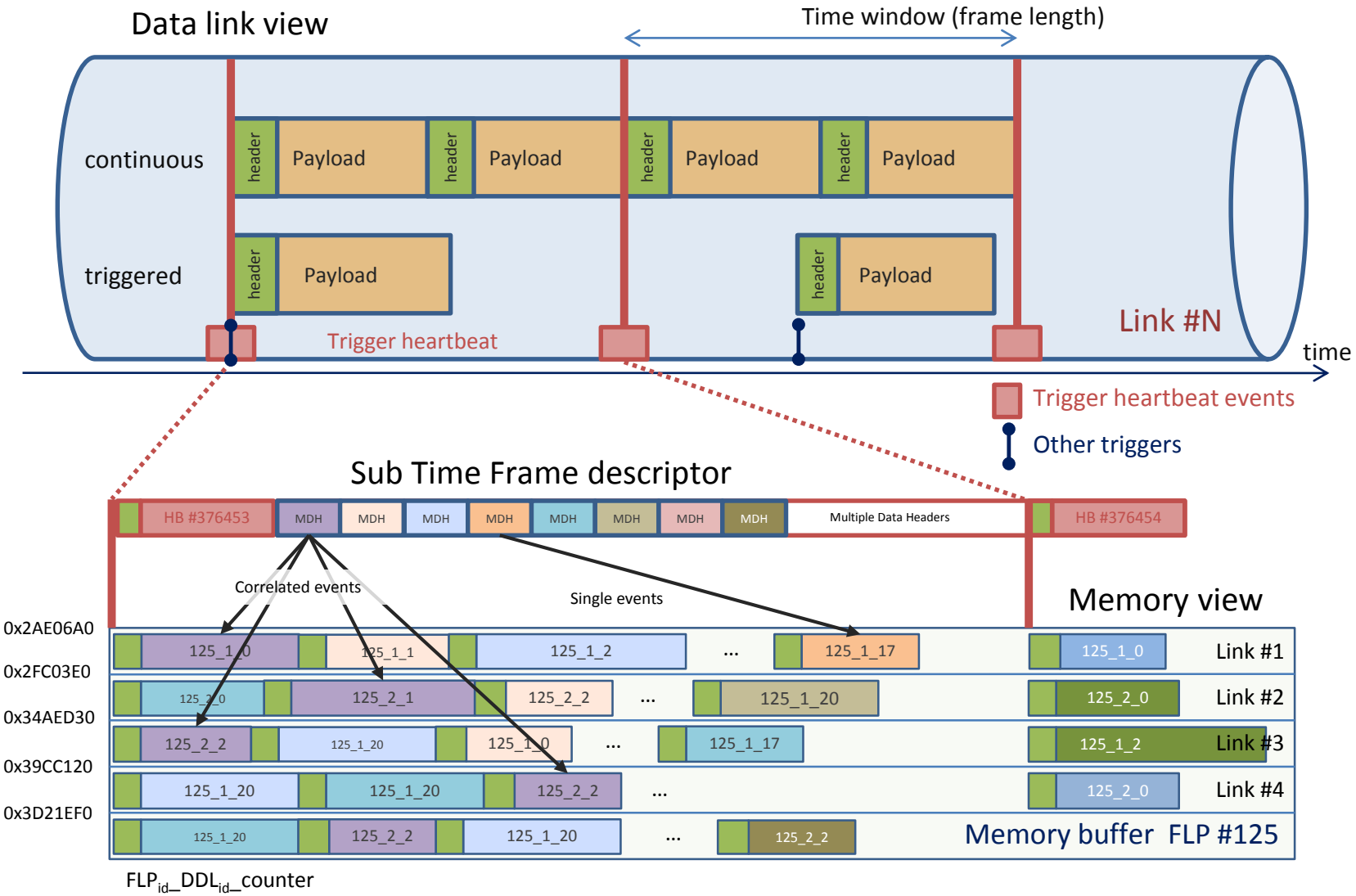
Analysis

ALICE O² Technical Design Report

https://cds.cern.ch/record/2011297/files/ALICE-TDR-019.pdf

Backup slides

Software design: data model



TPC data reduction

- TPC unmodified raw data rate: ~3.3 TB/s
- TPC zero-suppressed data rate: ~1TB/s
- Cluster finding (factor 2.5)
  - Compression ~1.25 in FPGA cluster finding
  - Data format optimisation ~factor 2
- Cluster compression (factor 3)
  - Currently: entropy coding (Huffman) - factor ~1.6
  - O2: compression applied to more cluster parameters
  - O2: additional compression using tracking information
- Background identification: factor ~2
- Charge transformations for clusters associated to tracks: factor ~1.35
- Total data compression by a factor ~20

Project Organisation

Management

Figure 2.1: The O^2 project organisation. (UCG)

O^2 TDR authors

11.11 O^2 TDR authors

The following people have contributed to the work presented in this TDR (The arabic institute numbers refer to the list from Tab. 11.2 and the roman institute numbers refer to footnotes):

T. Achalakul[22], D. Adamova[4], N. Agrawal[15], K. Akkarajitsakul[22], A. Alarcon Do Passo Suaide[1], T. Alt[11], M. Al-Turany[10], C. Alves Garcia Prado[1], Ananya[15], L. Aphecetche[7], A. Avasthi[15], M. Bach[11], S. Bae[17], R. Bala[14], G.G. Barnaföldi[13], A. Bhasin[14], J. Belikov[9], F. Bellini[i], L. Betev[2], L. Bianchi[27], T. Breitner[11, 12], P. Buncic[2], F. Carena[2], W. Carena[2], B. Changaival[22], S. Chapeland[2], M. Cherney[30], V. Chibante Barroso[2], H. Cho[18], P. Chochula[2], F. Cliff[ii], F. Costa[2], P. Crochet[5], L. Cunqueiro Mendez[2], S. Dash[15], E. David[13], C. Delort[2], E. Dénes[13], T. Dietel[21], R. Divià[2], B. Doenigus[10], H. Engel[12], D. Eschweiler[11], U. Fuchs[2], A. Gheata[2], M. Gheata[20], A. Gomez Ramirez[12], S. Gotovac[3], S. Gorbunov[11], L. Graczykowski[19], A. Grigoras[2], C. Grigoras[2], A. Grigore[2], R. Grosso[27], R. Guernane[6], A. Gupta[14], N. Hadi Lestriandoko[16], F. Hensan Muttaqien[16], I. Hřivnáčová[8], P. Hristov[2], C. Ionita[2], M. Ivanov[10], M. Janik[19], H. Jang[17], P. Jenviriyakul [22], S. Kalcher[11], N. Kassalias[1], U. Kebschull[12], R. Khandelwal[15], S. Kushpil[4], I. Kisel[11], G. Kiss[13], T. Kiss[13, iii], T. Kollegger[10], M. Kowalski[iv], M. Kretz[11], M. Krzewicki[11], I. Kulakov[12], Laosooksathit[29], C. Lara[12], A. Lebedev[10], I. Legrand[31], V. Lindenstruth[11], A. Maevskaya[v], P. Malzacher[10], A. Manafov[10], A. Morsch[2], E. Mudnic[3], S. Na Ranong [22], B. Nandi[15], M. Niculescu[20], J. Niedziela[19], J. Otwinowski[vi], J. Ozegovic[3], V. Papi[3], S. Park[17], P. Phunchongharn[22], P. Pillot[7], M. Planinic[vii], J. Pluta[19], N. Poljak [vii] S. Prom-on[22], K. Pugdeethosapol [22], S. Pumma[22], A. Rabalchenko[10], S. Rajput[14], K. Read[28], A. Ribon[2], M. Richter[viii], D. Rohr[11], D. Rosiyadi[16], G. Rubin[13], R. Sadikin[16], R. Shahoyan[2], A. Sharma[14], G. Simonetti[2], O. Smorholm [ii] C. Soós[2], S. Sumowidagdo[16], J. Sun[18], M. Szymanski[19], A. Telesca[2], J. Thäder[25], A. Timmins[27], A. Udupa[15], P. Vande Vyvre[2], F. Vennedey[11, 12], L. Vickovic[3], B. von Haller[2], S. Wenzel[2], N. Winckler[10], T. Wirahman[16], K. Yamnual [22], C. Zampolli [i], and M. Zyzak[12].

[i] University and Sezione INFN, Bologna, Italy
[ii] School of Physics and Astronomy, University of Birmingham, Birmingham, United Kingdom
[iii] Cerntech Ltd., Budapest, Hungary
[iv] The Henryk Niewodniczanski Institute of Nuclear Physics, Polish Academy of Sciences, Cracow, Poland
[v] Institute for Nuclear Research, Academy of Sciences, Moscow, Russia
[vi] Polish Academy of Sciences, Cracow, Poland
[vii] Institute Rudjer Boskovic, Zagreb, Croatia
[viii] Department of Physics, University of Oslo, Oslo, Norway

Project Organisation

Institutes

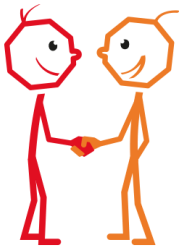


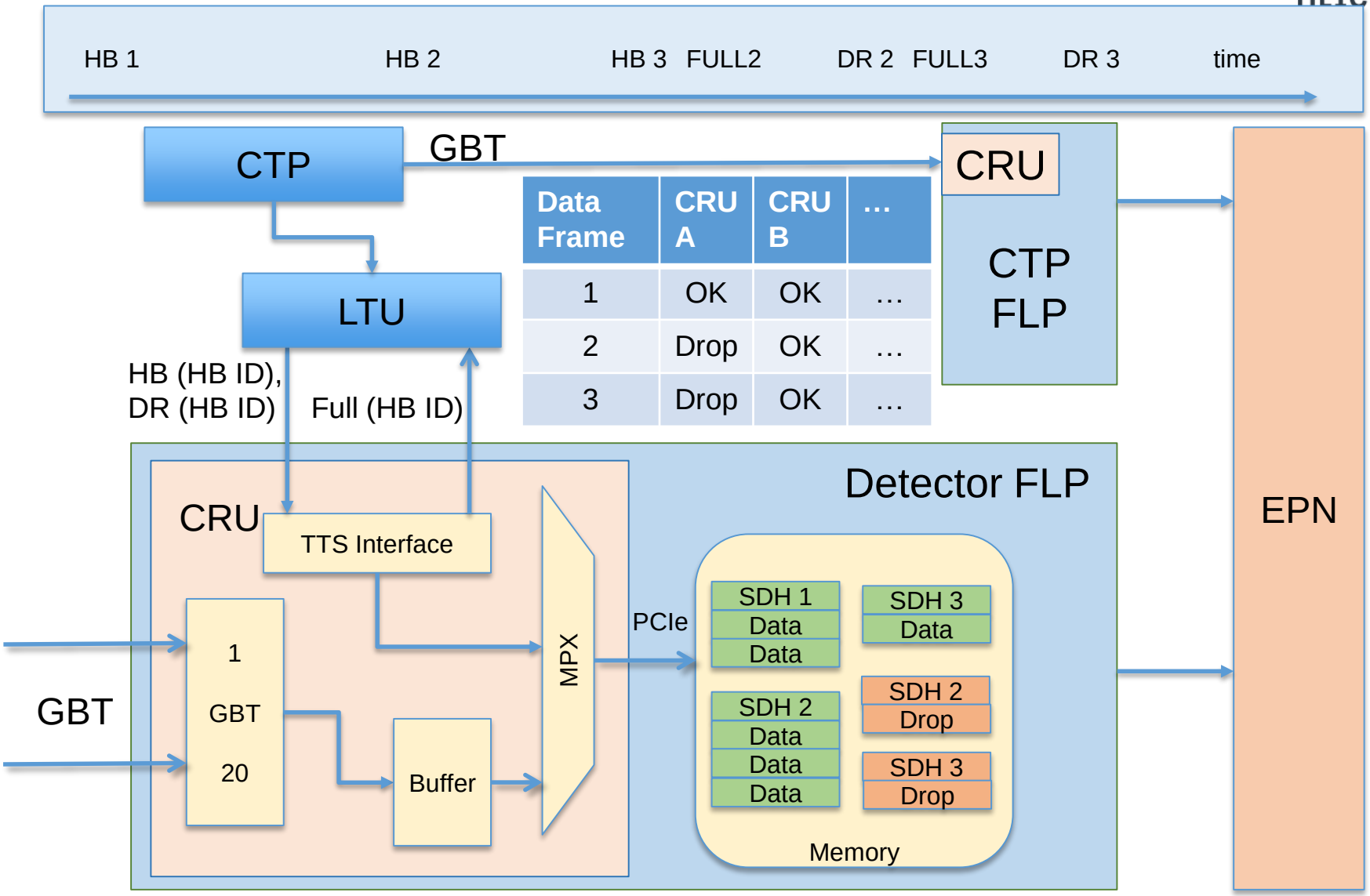
Table 11.2: Institutes participating in the O^2 Project.

	Country	City	Institute	Acronym
1	Brasil	São Paulo	University of São Paulo	USP
2	CERN	Geneva	European Organization for Nuclear Research	CERN
3	Croatia	Split	Technical University of Split	FESB
4	Czech Republic	Rez u Prahy	Nuclear Physics Institute, Academy of Sciences of the Czech Republic	ASCR
5	France	Clermont -Ferrand	Laboratoire de Physique Corpusculaire (LPC), Université Blaise Pascal Clermont-Ferrand, CNRS-IN2P3	LPC
6	France	Grenoble	Laboratoire de Physique Subatomique et de Cosmologie (LPSC), Université Grenoble-Alpes, CNRS-IN2P3	LPSC
7	France	Nantes	SUBATECH, Ecole des Mines de Nantes, Université de Nantes, CNRS-IN2P3	SUBATECH
8	France	Orsay	Institut de Physique Nucléaire (IPNO), Université Paris-Sud, CNRS-IN2P3	IPNO
9	France	Strasbourg	Institut Pluridisciplinaire Hubert Curien	IPHC
10	Germany	Darmstadt	Research Division and ExtreMe Matter Institute EMMI, GSI Helmholtzzentrum für Schwerionenforschung	GSI
11	Germany	Frankfurt	Frankfurt Institute for Advanced Studies, Johann Wolfgang Goethe-Universität	FIAS
12	Germany	Frankfurt	Institut für Informatik, Johann Wolfgang Goethe-Universität Frankfurt	IRI
13	Hungary	Budapest	Wigner RCP Hungarian Academy of Sciences	WRCP
14	India	Jammu	University of Jammu	JU
15	India	Mumbai	Indian Institute of Technology	IIT
16	Indonesia	Bandung	Indonesian Institute of Sciences	LIPI
17	Korea	Daejeon	Korea Institute of Science and Technology Information	KISTI
18	Korea	Sejong City	Korea University	KU
19	Poland	Warsaw	Warsaw University of Technology	WUT
20	Romania	Bucharest	Institute of Space Science	ISS
21	South Africa	Cape Town	University of Cape Town	UCT
22	Thailand	Bangkok	King Mongkut's University of Technology Thonburi	KMUTT
23	Thailand	Bangkok	Thammasat University	TU
24	Turkey	Konya	KTO Karatay University	KTO
25	United States	Berkeley, CA	Lawrence Berkeley National Laboratory	LBNL
26	United States	Detroit, MI	Wayne State University	WSU
27	United States	Houston, TX	University of Houston	UH
28	United States	Knoxville, TN	University of Tennessee	UTK
29	United States	Oak Ridge, TN	Oak Ridge National Laboratory	ORNL
30	United States	Omaha, NE	Creighton University	CU
31	United States	Pasadena, CA	California Institute of Technology	CALTECH

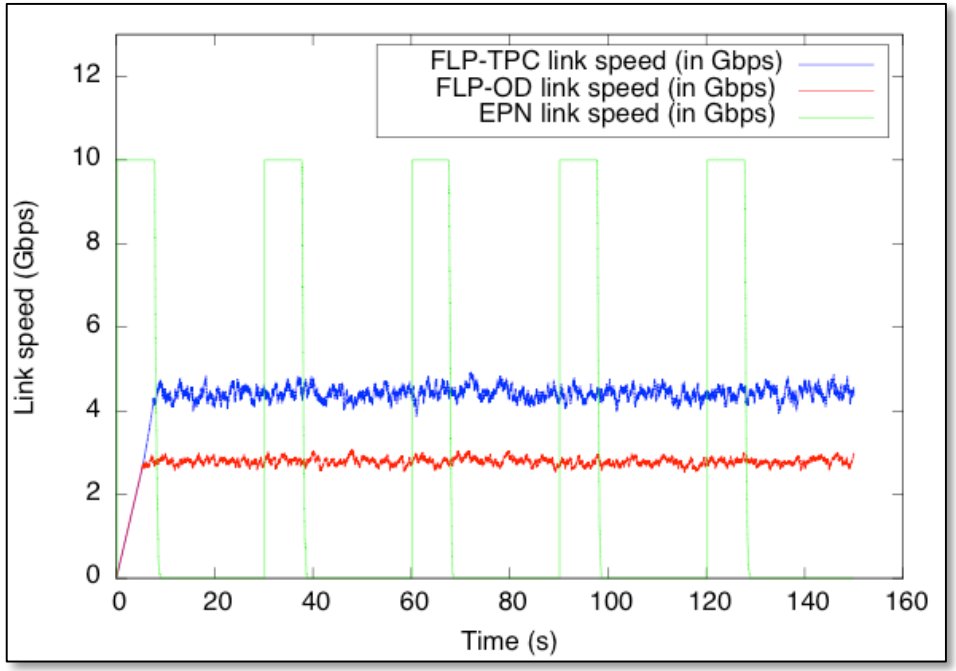
Technology: Future-bets

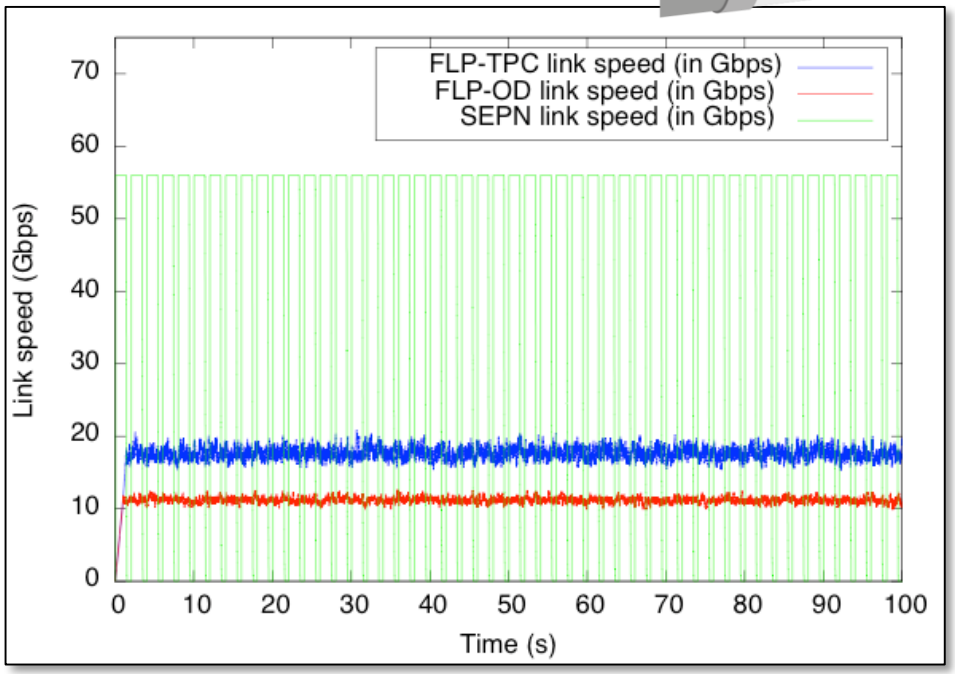
Technology	O^2 need	Comment
I/O	~77 Gb/s on 1 slot	Available now: PCIe Gen3. Gen4 backward compatibility.
FPGA	24 GBT receivers + 24 cluster finders	Available now as engineering samples: Altera Arria 10 GX
CPU (# Cores)	2 x 32 cores	14 cores (Xeon 4850) 18 cores (Xeon 8880) Bet: 32 in the affordable mid-range
GPU	TPC track seeding and following 0.1 s	Available now : AMD S9000. Bet: cost decrease slope.
Network	≥ 40 Gb/s	Available now in 2 technologies. Probably 3 technologies at the time of purchase.
Data storage	20 TB hard disk	Now: 8 TB Bet: a density yearly increase of 20%

Continuous read-out: data flows



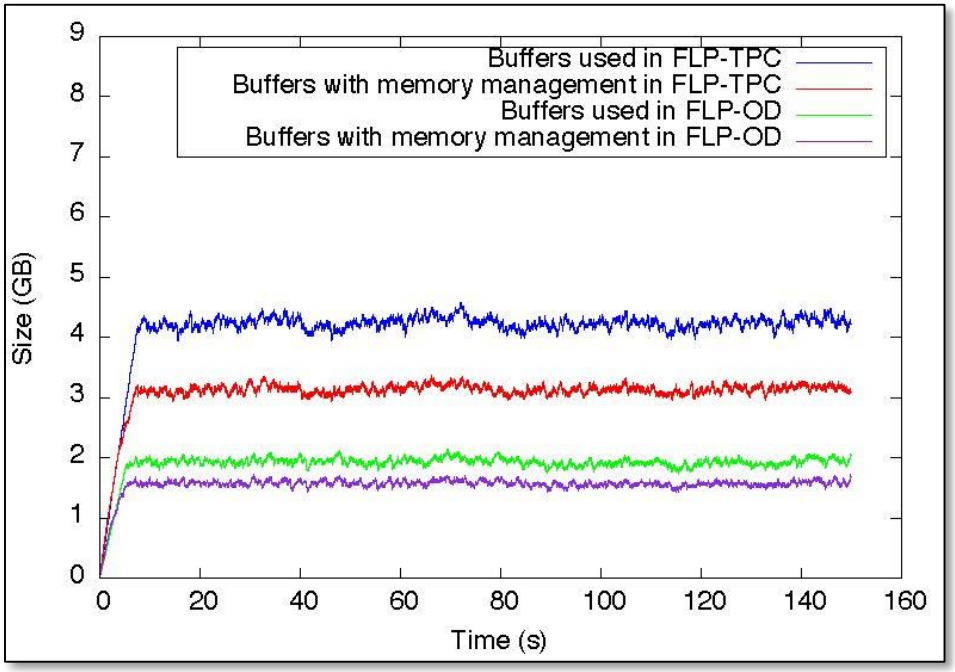
Facility Simulation: Link speed

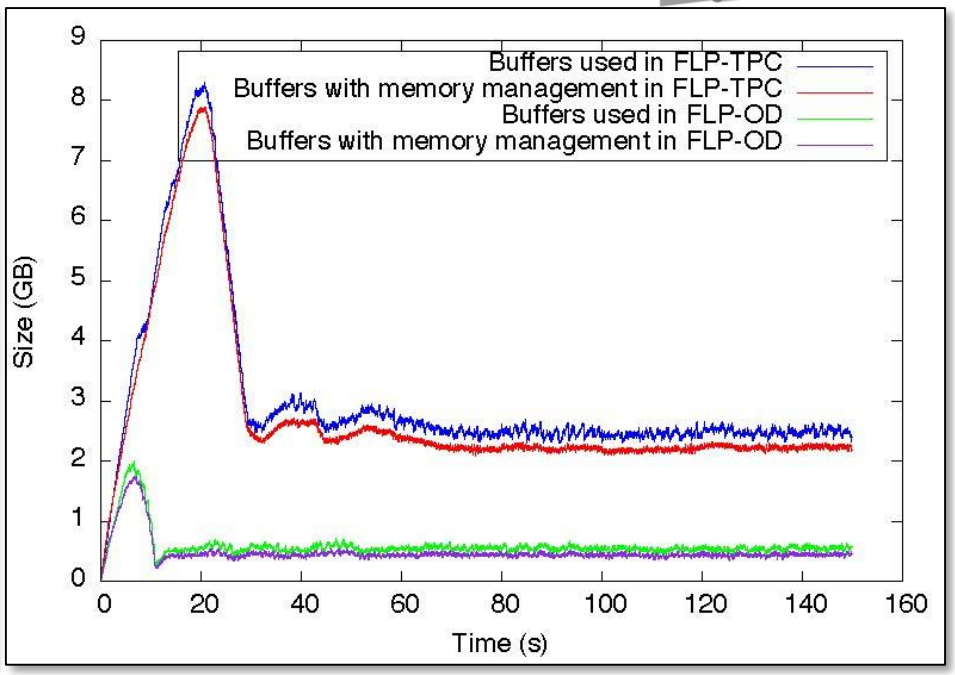




- Left Layout 2 – Right Layout 3
- Distinguish between TPC and Other Detector (OD) FLPs

Facility Simulation: FLP buffering need





- Left Layout 2 – Right Layout 3
- Distinguish between TPC and Other Detector (OD) FLPs
- With and without buffer management releasing data blocks before a full Time-Frame is transferred

Facility Simulation: Scalability

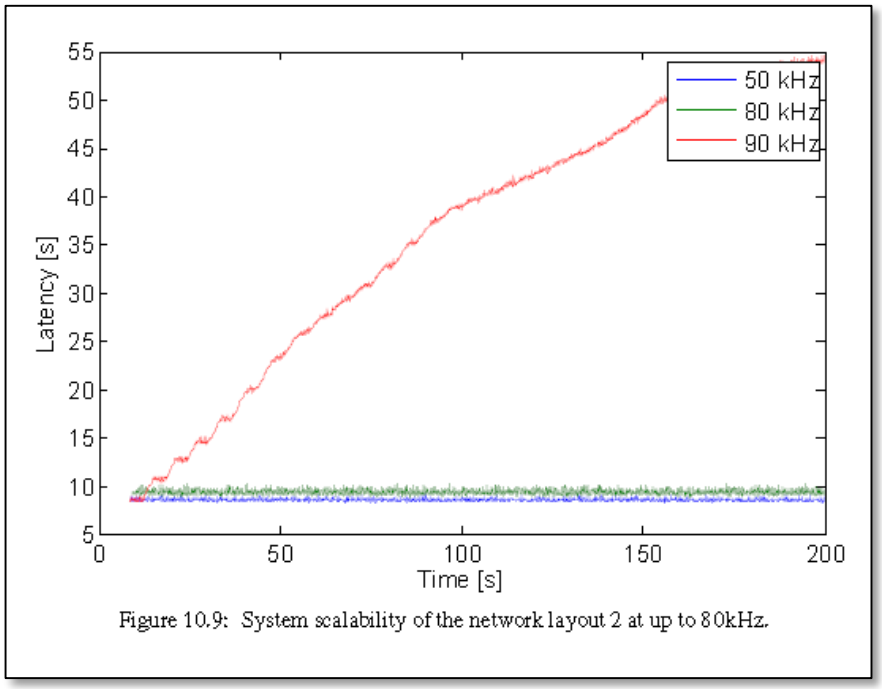


Figure 10.9: System scalability of the network layout 2 at up to 80kHz.

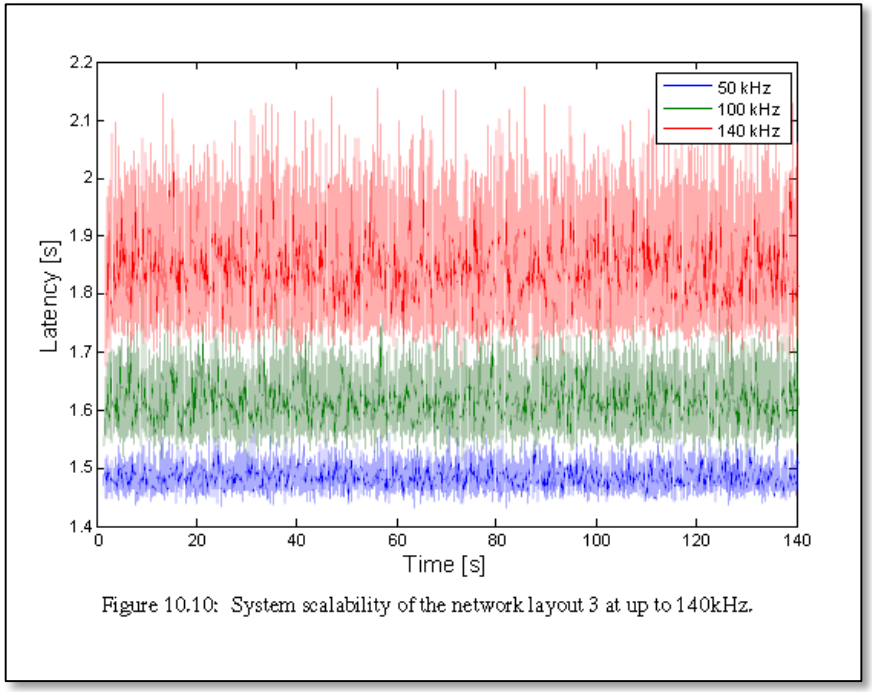


Figure 10.10: System scalability of the network layout 3 at up to 140kHz.

- Left Layout 2 (4x10 Gb/s) – Right Layout 3 (56 Gb/s)
- Scalability measured via the latency of a Time-Frame in the system