INSTITUT FIZYKI JĄDROWEJ
IM. HENRYKA NIEWODNICZAŃSKIEGO
POLSKIEJ AKADEMII NAUK

Extreme Computing in the ALICE Experiment

Jacek Otwinowski (IFJ PAN)

PTI, AGH, 14.05.2019

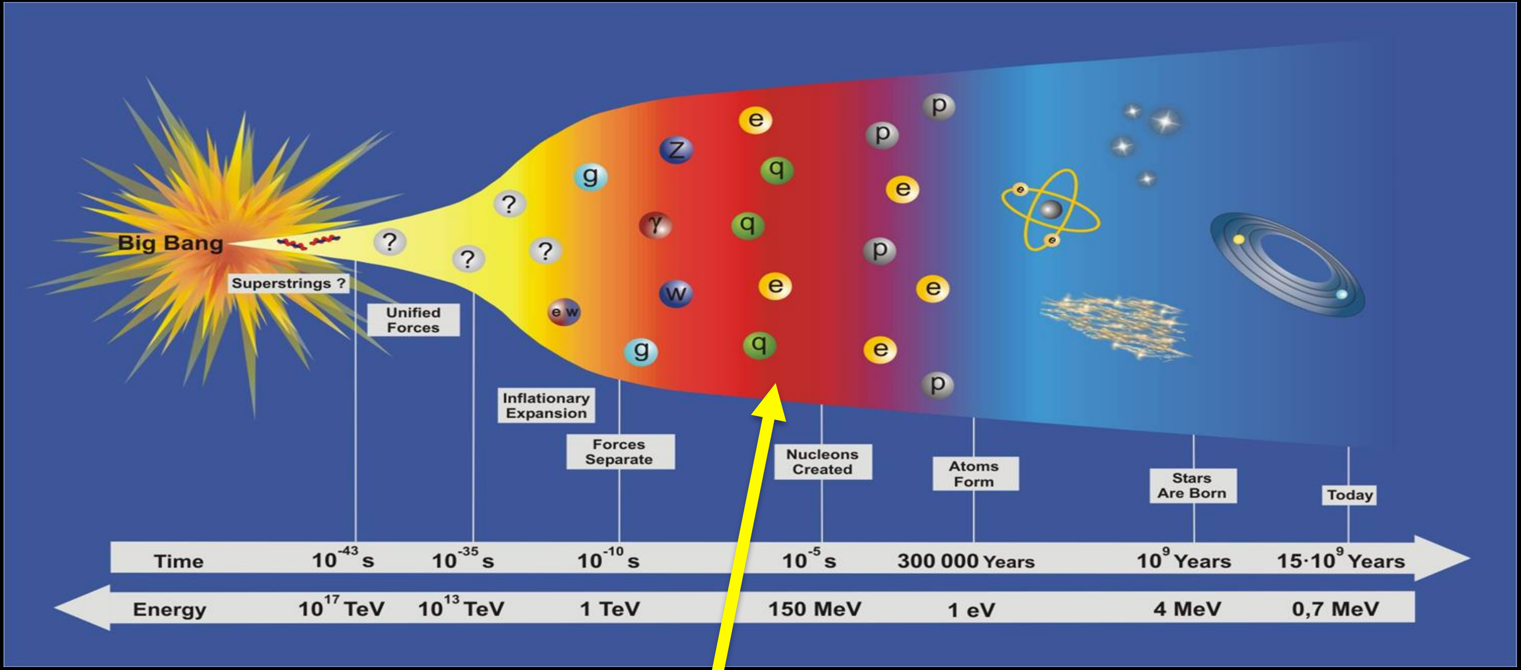
A Large Ion Collider Experiment
aliceinfo.cern.ch

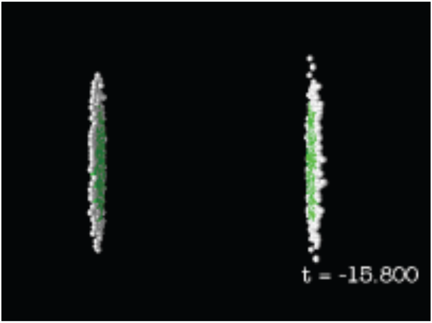
European Organization for Nuclear Research
www.cern.ch

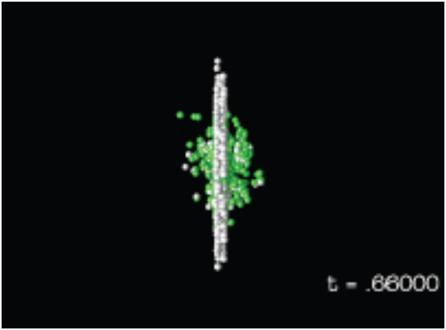
Big Bang in Laboratory

Heavy-ion collisions

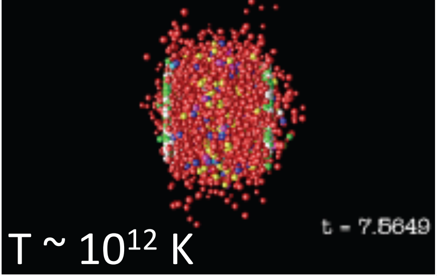
$^{208}Pb + ^{208}Pb$
(208 nucleons)



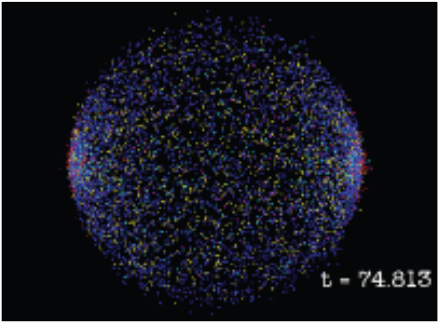




Quark-Gluon Plasma

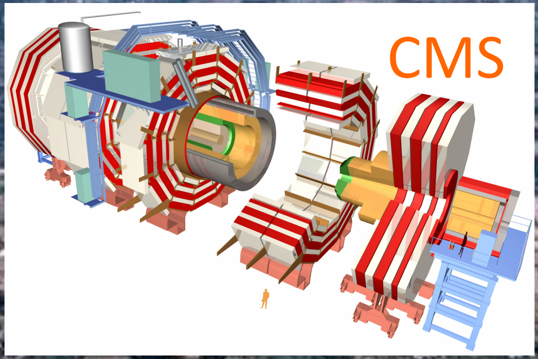


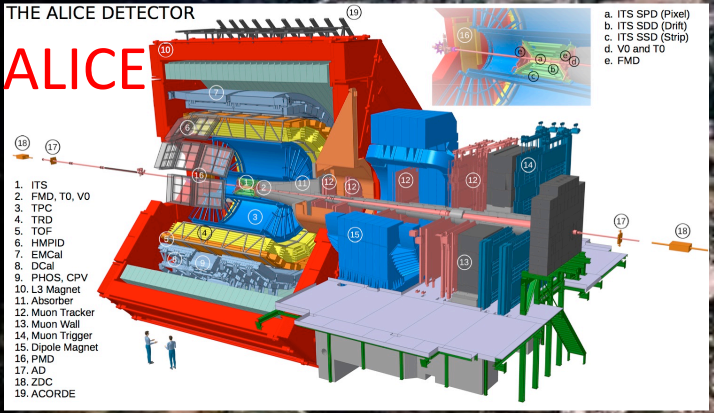
$T \sim 10^{12}$ K

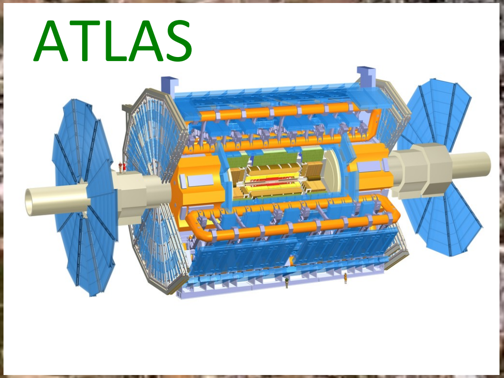


Time

Large Hadron Collider





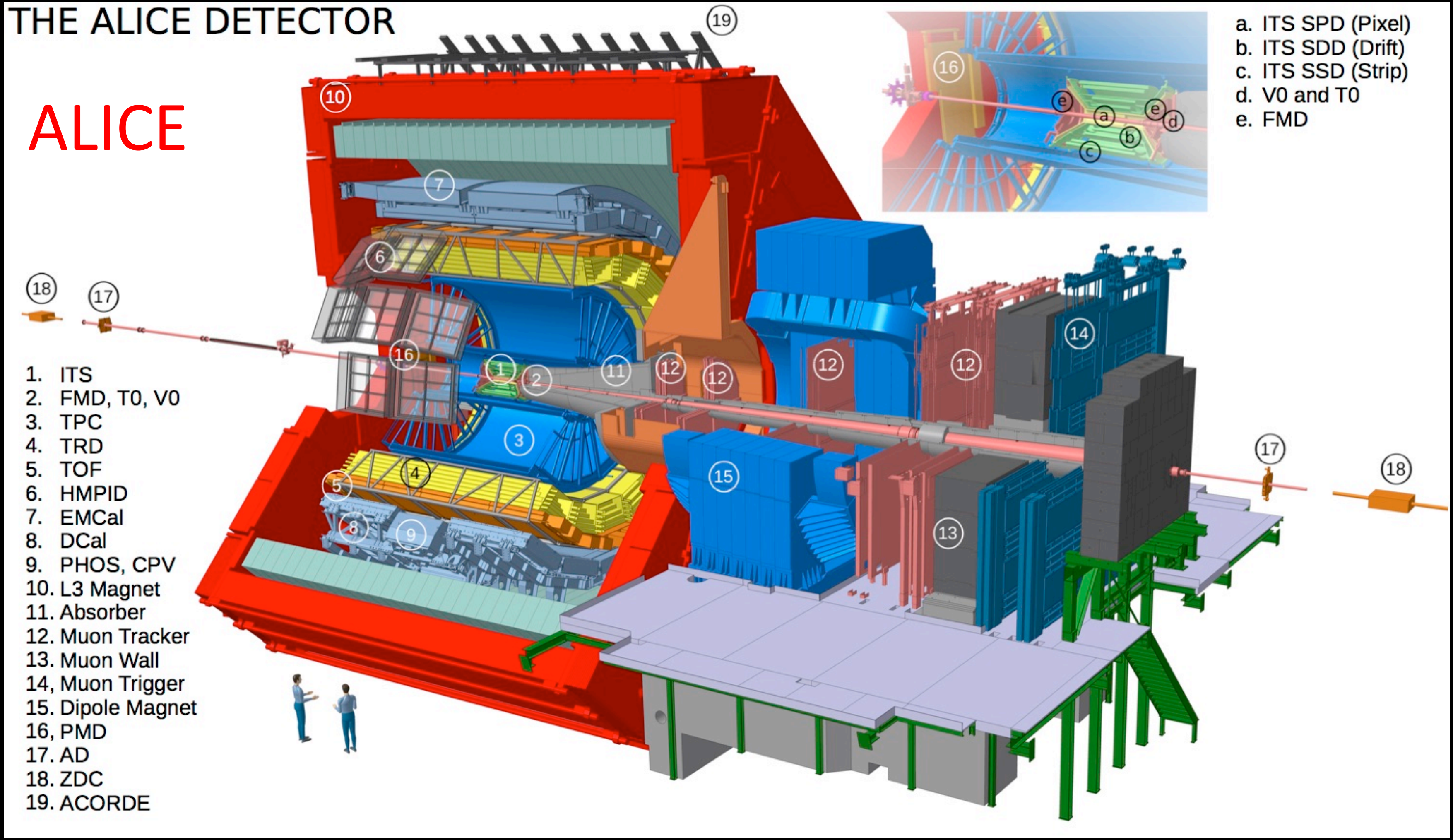


p+p at √s=13 TeV

Pb+Pb at $\sqrt{s_{NN}}$=5 TeV

(collision energy per nucleon pair)

A Large Ion Collider Experiment



1. ITS
2. FMD, T0, V0
3. TPC
4. TRD
5. TOF
6. HMPID
7. EMCal
8. DCal
9. PHOS, CPV
10. L3 Magnet
11. Absorber
12. Muon Tracker
13. Muon Wall
14, Muon Trigger
15. Dipole Magnet
16, PMD
17. AD
18. ZDC
19. ACORDE

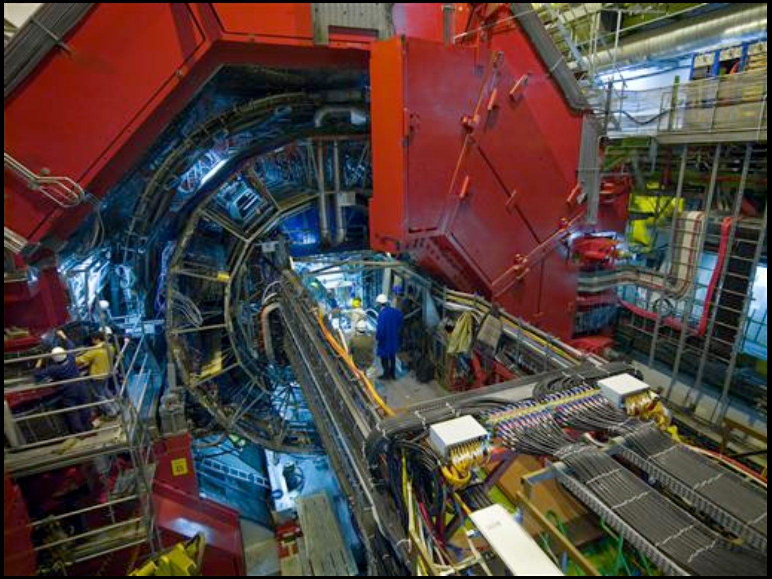
a. ITS SPD (Pixel)
b. ITS SDD (Drift)
c. ITS SSD (Strip)
d. V0 and T0
e. FMD

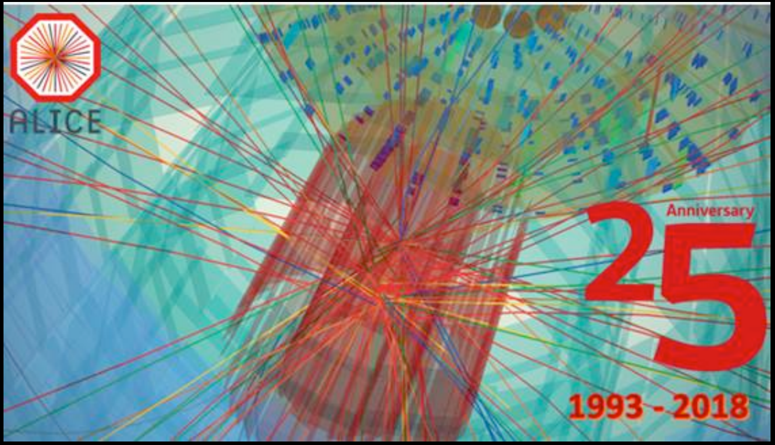
IFJ PAN (since beginning in ALICE)

- physics observables, simulations, calibration and reconstruction, data quality control

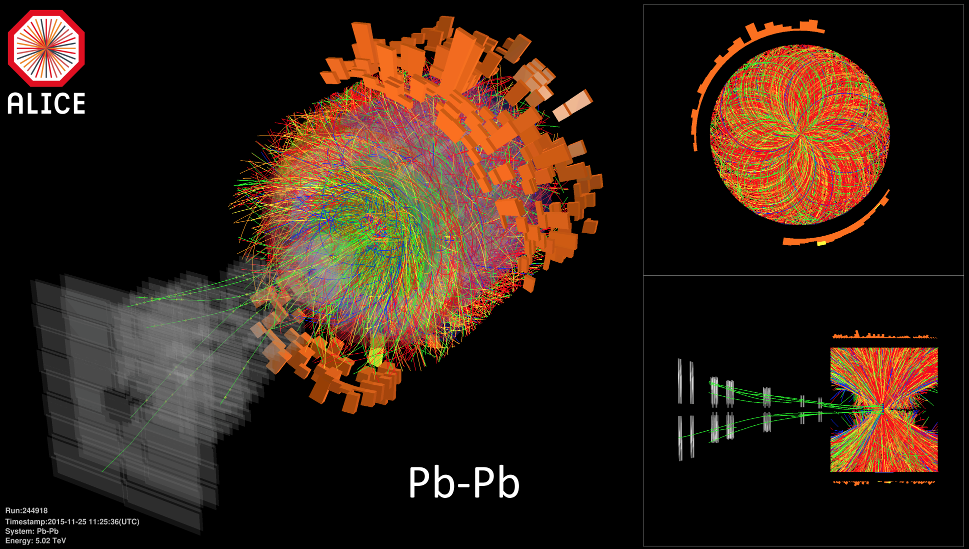
ALICE at Work since 2009

- ~ 15 years of construction work
- More than 500000 readout channels
- ~8000 charged particles in Pb-Pb collision

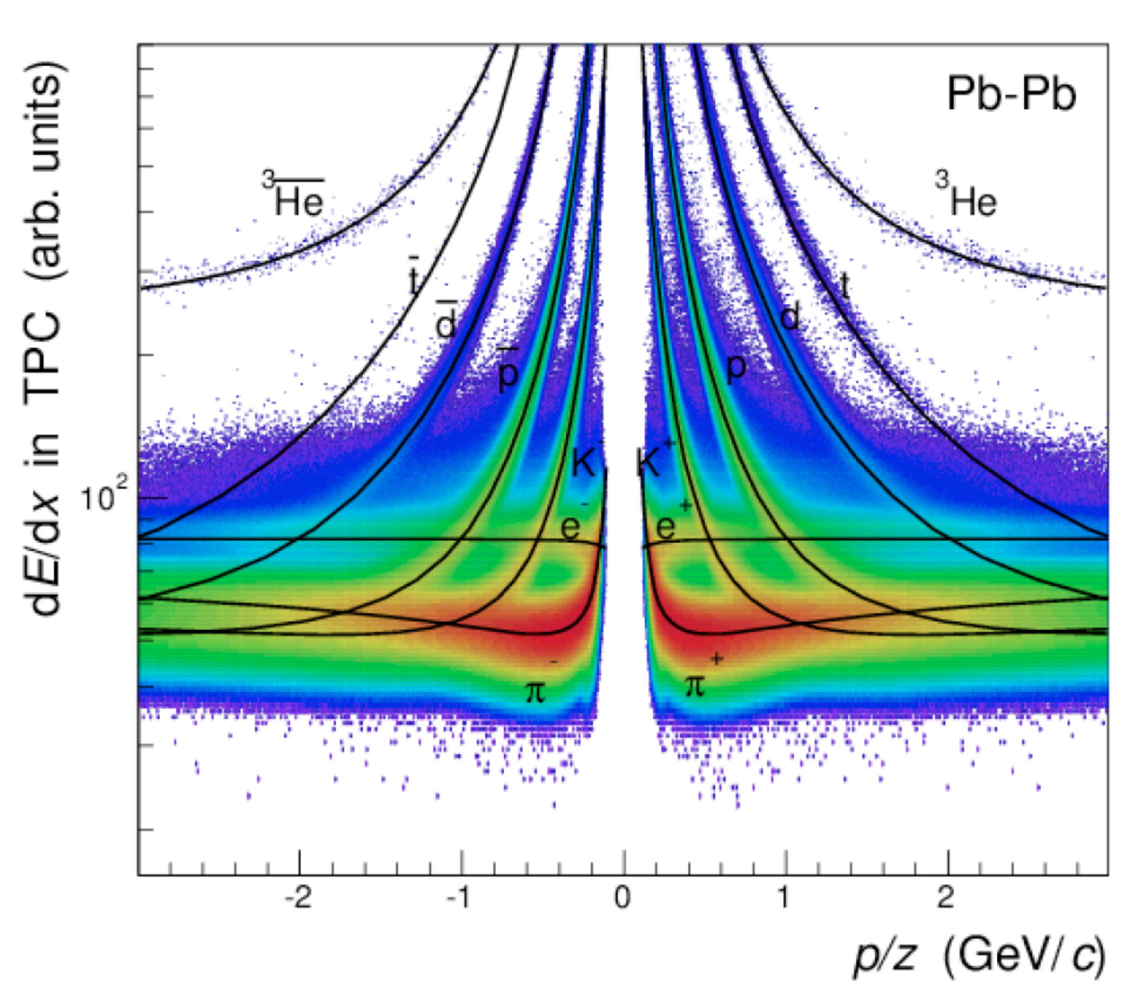




https://indico.cern.ch/event/653848

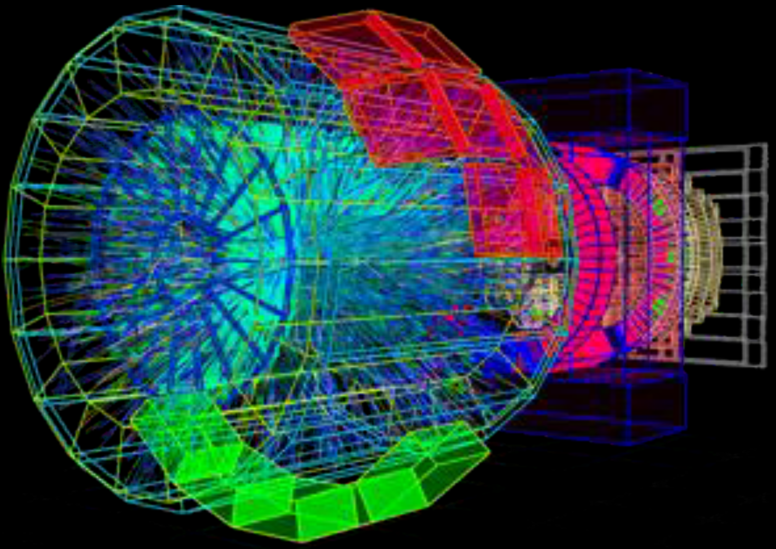


ALICE Particle Identification Capabilty



Matter and antimater is produced with the same amount at the LHC!

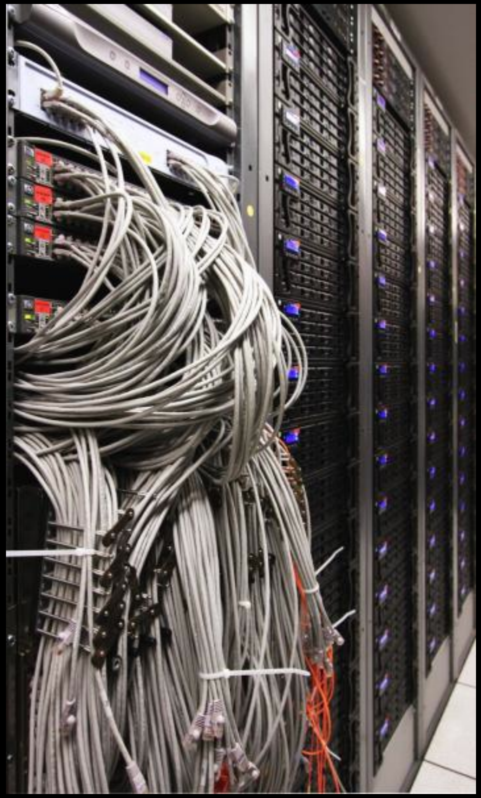
Data Processing in ALICE



~12 GB/s

DAQ and HLT (High Level Trigger)

- ~1000 CPUs and FPGAs
- Data acquisition and online reconstruction and compression



~4 GB/s

ALICE grid (AliEn)

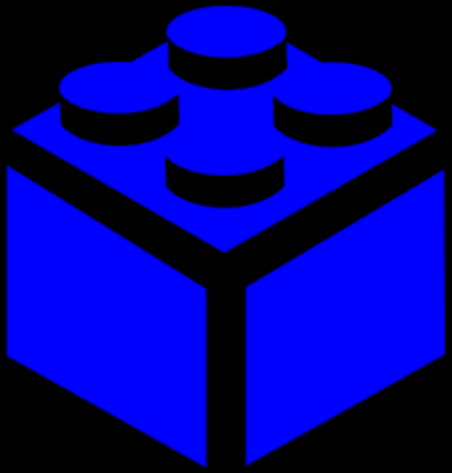
- ~ 50 PB disk storage
- ~ 60000 CPUs
- Offline data calibration, reconstruction and analysis
- Monte Carlo simulations

Current ALICE Software

aliweb.cern.ch/Offline

ROOT

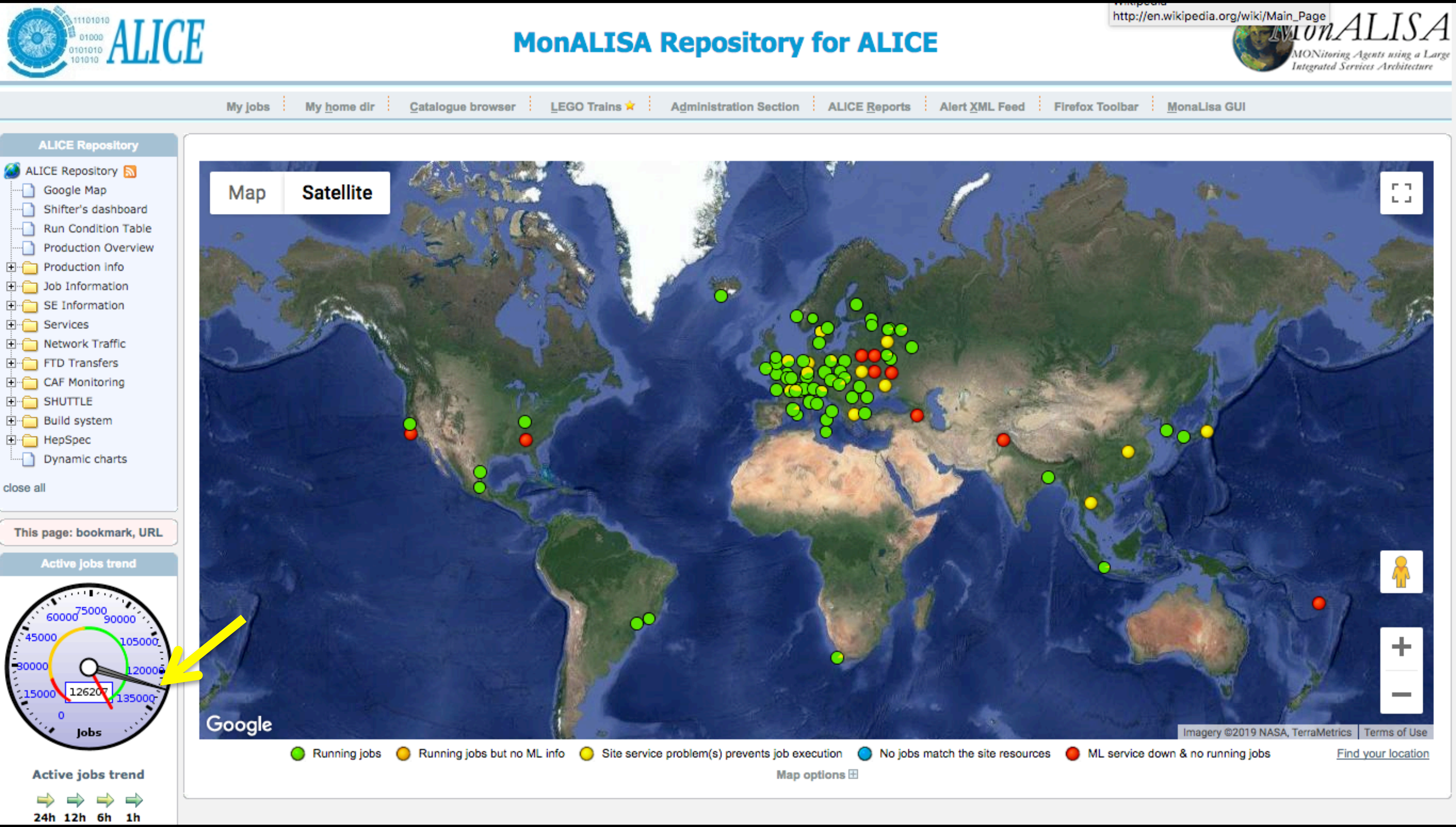
AliROOT



AliEn
+ MonALISA

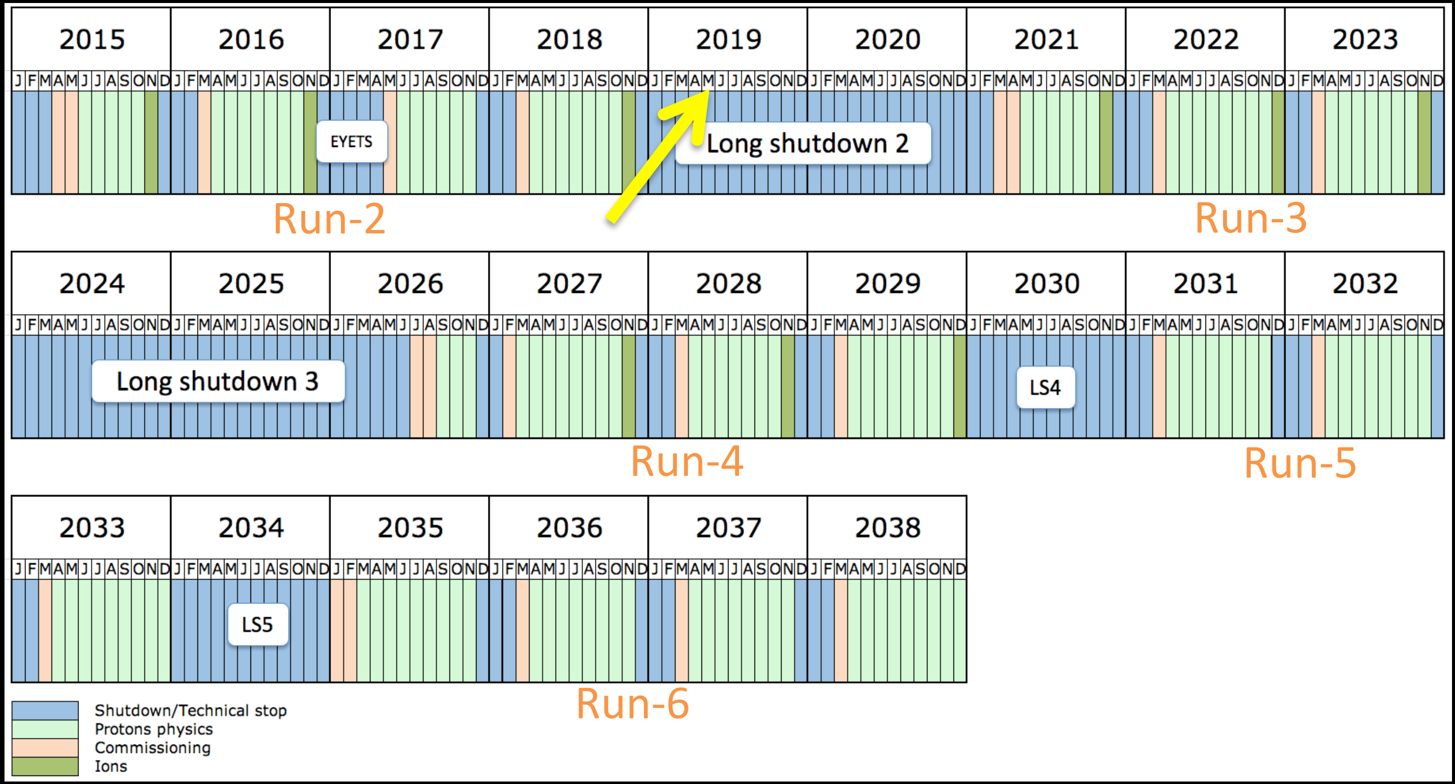
- ROOT – software framework for data analysis, visualization and storage (C++, Python, R…)
- AliROOT– ALICE software for data calibration, reconstruction and analysis based on ROOT
- AliEn + MonALISA – ALICE grid software for distributed data processing

MonALISA


ALICE
MonALISA Repository for ALICE
http://en.wikipedia.org/wiki/Main_Page
MonALISA
MONitoring Agents using a Large Integrated Services Architecture
My jobs
My home dir
Catalogue browser
LEGO Trains
Administration Section
ALICE Reports
Alert XML Feed
Firefox Toolbar
MonaLisa GUI
ALICE Repository
ALICE Repository
Google Map
Shifter's dashboard
Run Condition Table
Production Overview
Production info
Job Information
SE Information
Services
Network Traffic
FTD Transfers
CAF Monitoring
SHUTTLE
Build system
HepSpec
Dynamic charts
close all
This page: bookmark, URL
Active jobs trend
0
15000
30000
45000
60000
75000
90000
105000
120000
135000
126207
Jobs
Active jobs trend
24h
12h
6h
1h
Map
Satellite
Google
Imagery ©2019 NASA, TerraMetrics
Terms of Use
Running jobs
Running jobs but no ML info
Site service problem(s) prevents job execution
No jobs match the site resources
ML service down & no running jobs
Find your location
Map options

ALICE Future

LHC running program

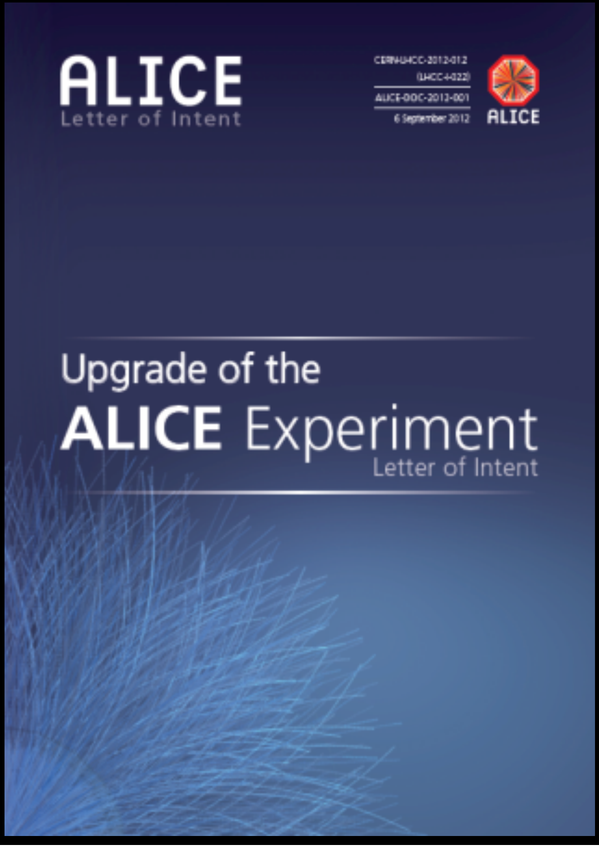


ALICE work on preparations for Run-3 and Run-4

- Detector upgrade
- Online-offline computing system upgrade
- Readout electronics and trigger upgrade

ALICE Upgrade Documents

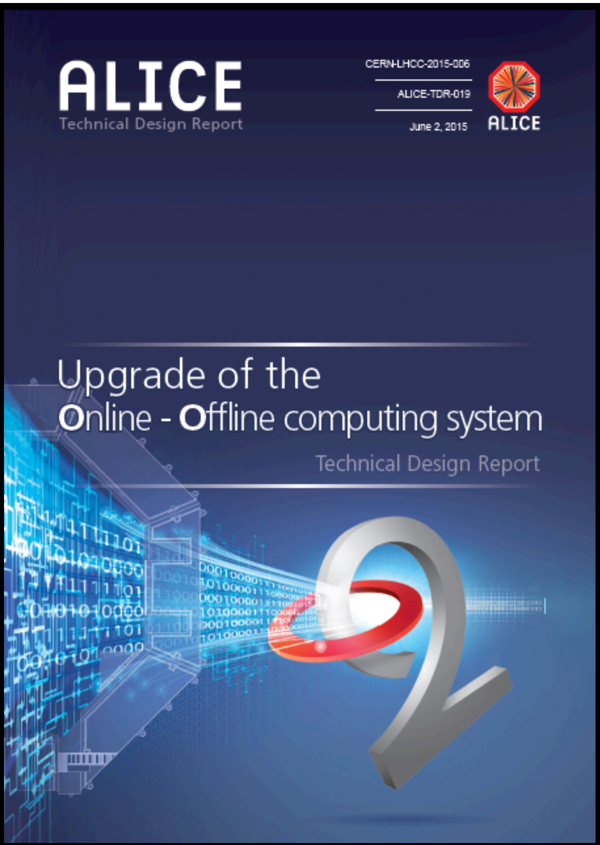
https://cds.cern.ch/record/2011297


ALICE
Letter of Intent

CERN-LHCC-2012-012
(LHCC-I-022)
ALICE-DOC-2012-001
6 September 2012

ALICE

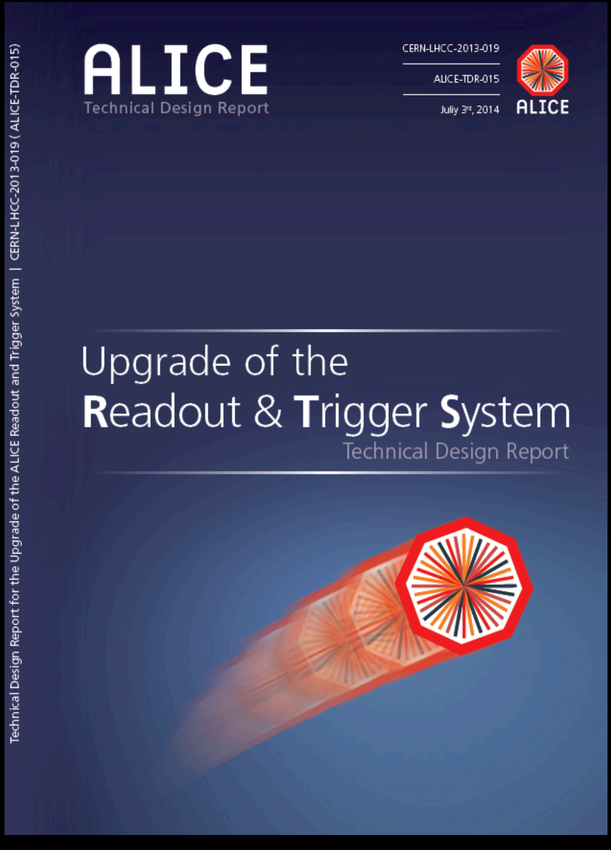
Upgrade of the
ALICE Experiment
Letter of Intent


ALICE
Technical Design Report

CERN-LHCC-2015-006
ALICE-TDR-019
June 2, 2015

ALICE

Upgrade of the
Online - **O**ffline computing system
Technical Design Report


Technical Design Report for the Upgrade of the ALICE Readout and Trigger System | CERN-LHCC-2013-019 (ALICE-TDR-015)

ALICE
Technical Design Report

CERN-LHCC-2013-019
ALICE-TDR-015
Juliy 3rd, 2014

ALICE

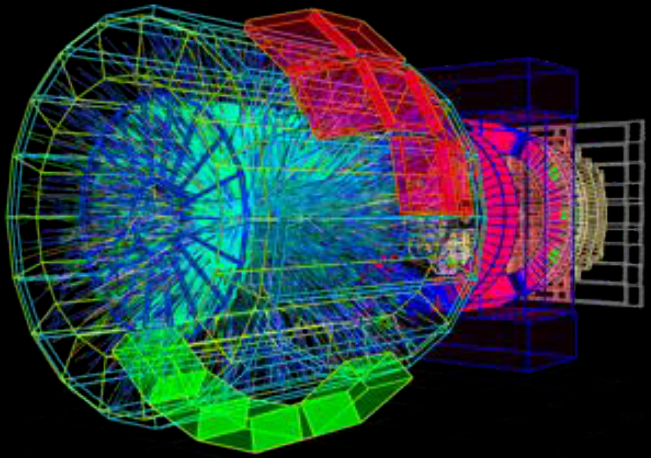
Upgrade of the
Readout & **T**rigger **S**ystem
Technical Design Report

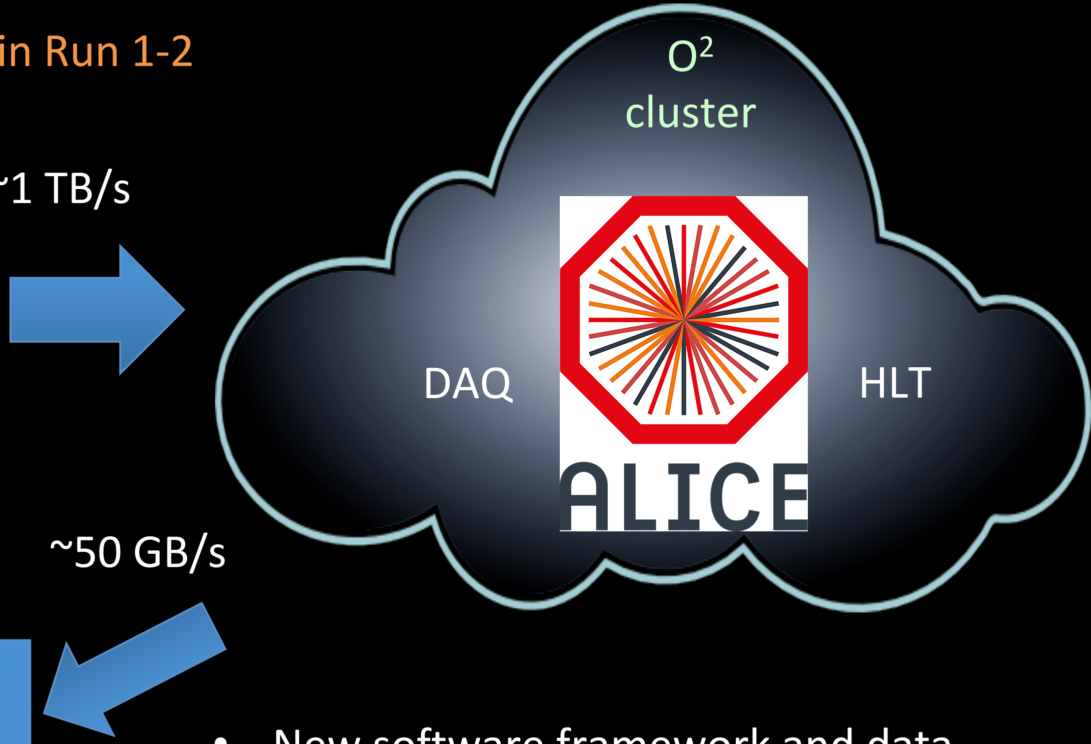
http://cds.cern.ch/record/1475243

http://cds.cern.ch/record/1603472

ALICE O^2 Online-Offline Computing

- Continuous data readout
- ~100x more data in Run 3-4 than in Run 1-2



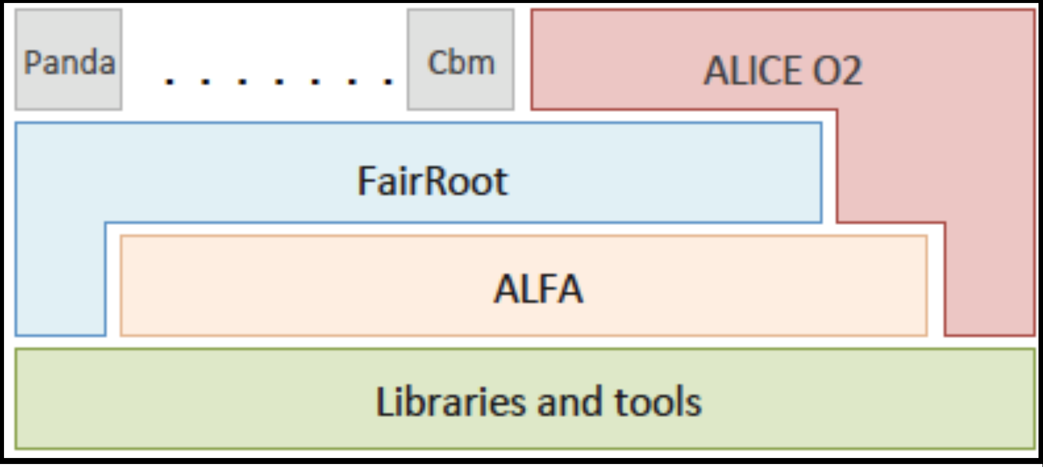


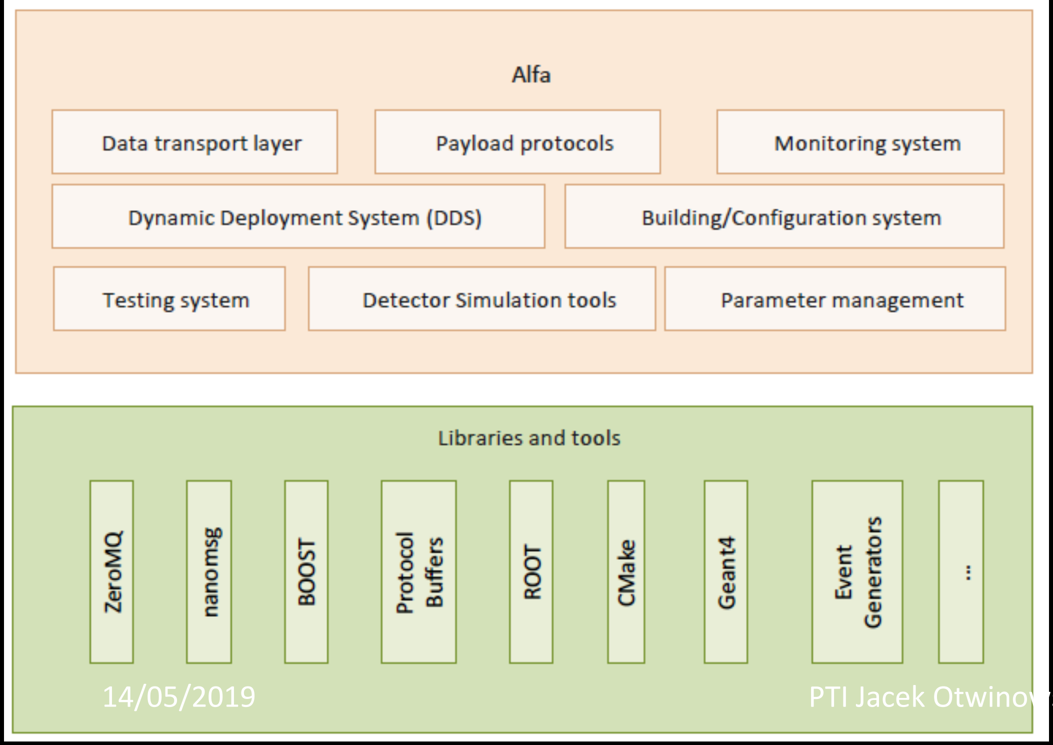
ALICE grid (AliEn 2.)

- ~ Exabyte of data
- Offline data calibration, reconstruction and analysis
- Monte Carlo simulations

- New software framework and data model for parallel data processing
- Heterogeneous system FPGAs, GPUs, CPUs
- Optimized I/O
- Full virtualization (CernVM)

ALICE O^2 Software Ecosystem



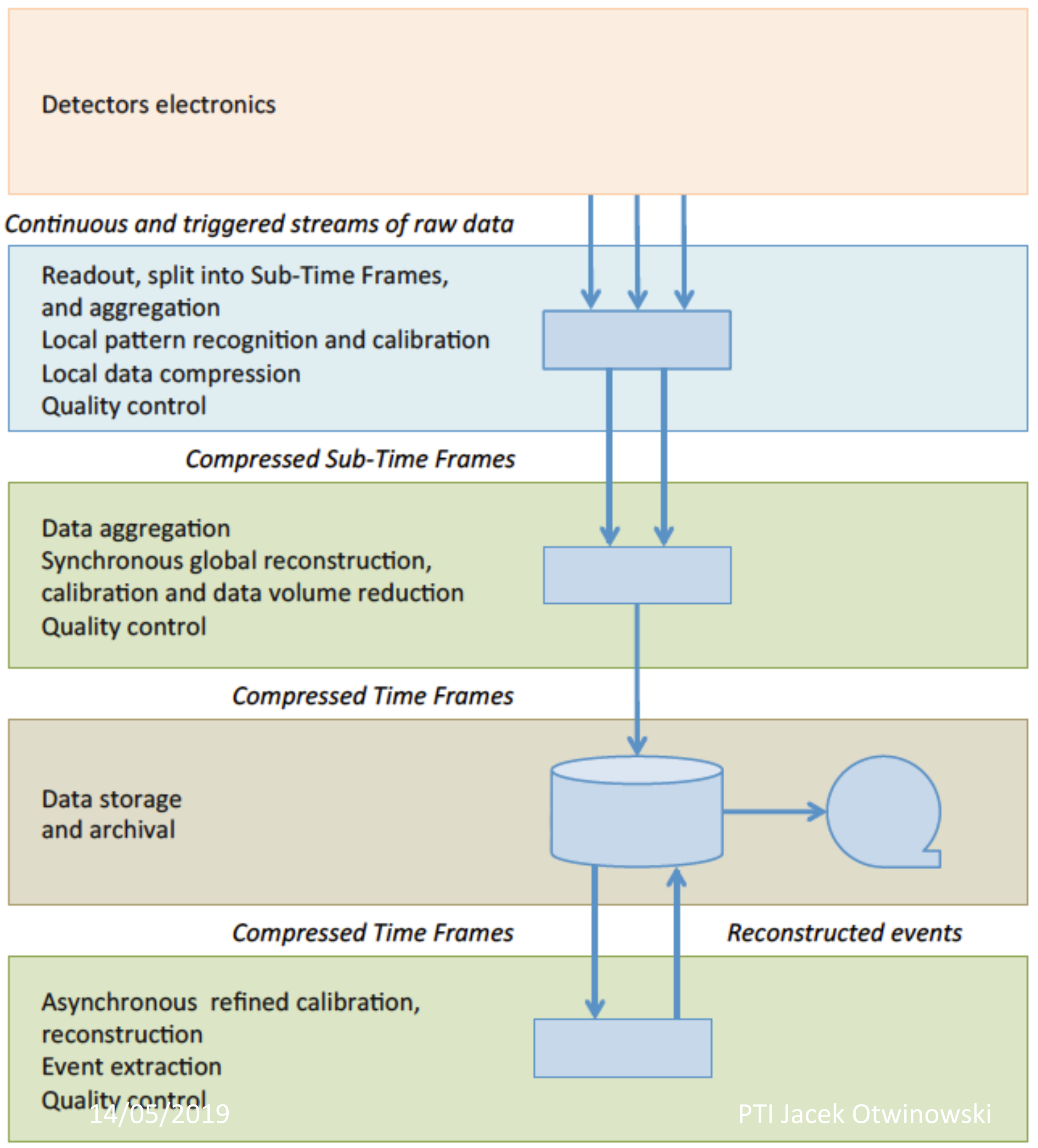


ALFA (ALICE-FAIR) concurrency framework for efficient parallel data processing on heterogeneous systems

- Data transport layer based on ZeroMQ/nanomsg
- Several data serialization/ deserialization standards
  - BOOST serialization
  - ROOT streamers
  - …
- Dynamic Deployment System (DDS)

https://cds.cern.ch/record/2011297

ALICE O^2 Functional Flow



Detectors electronics

Continuous and triggered streams of raw data

Readout, split into Sub-Time Frames,
and aggregation
Local pattern recognition and calibration
Local data compression
Quality control

Compressed Sub-Time Frames

Data aggregation
Synchronous global reconstruction,
calibration and data volume reduction
Quality control

Compressed Time Frames

Data storage
and archival

Compressed Time Frames

Reconstructed events

Asynchronous refined calibration,
reconstruction
Event extraction
Quality control

https://cds.cern.ch/record/2011297

- Raw data in time frames
- O^2 cluster (FLPs, EPNs)

First Level Processors (FLPs)
O(250)

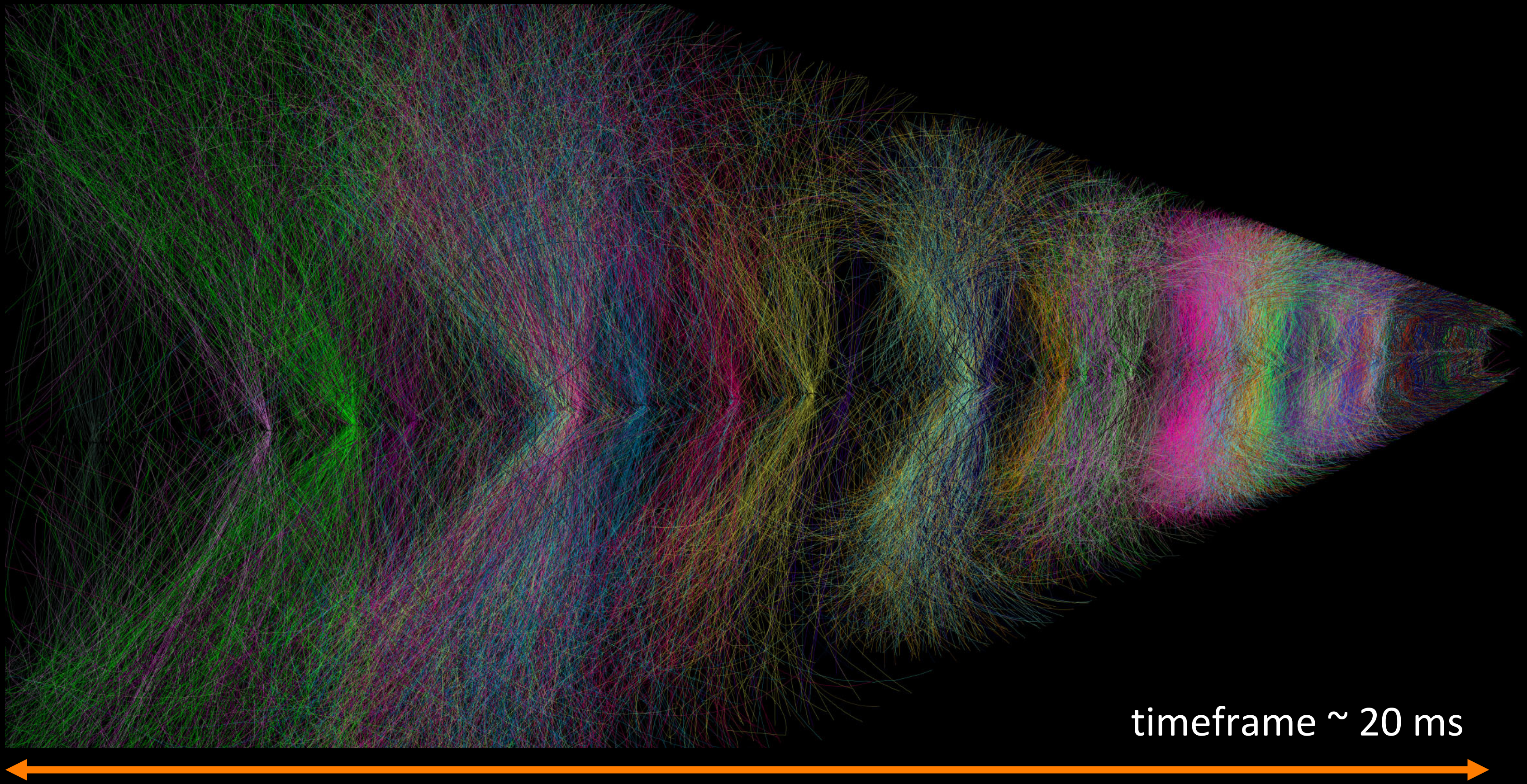
Event Processing Nodes (EPNs)
O(1000)

Data movers (O^2 cluster -> grid)
ALICE grid storage servers
Data Base servers

ALICE grid nodes and EPNs

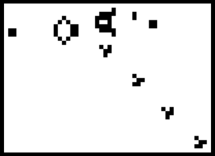
ALICE Tracking in Run-3 and Run-4

Continues data readout



- Several collision events in one timeframe
- Tracking for continues data readout in timeframes will be done on GPUs

ALICE Tracking Algorithms



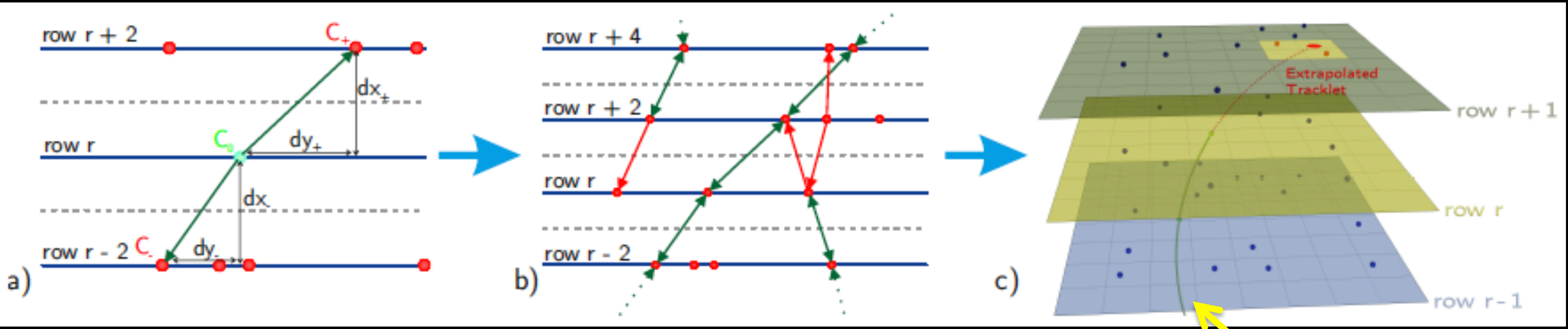
Cellular Automaton

Kalman filter

track forming

track concatenating

track following

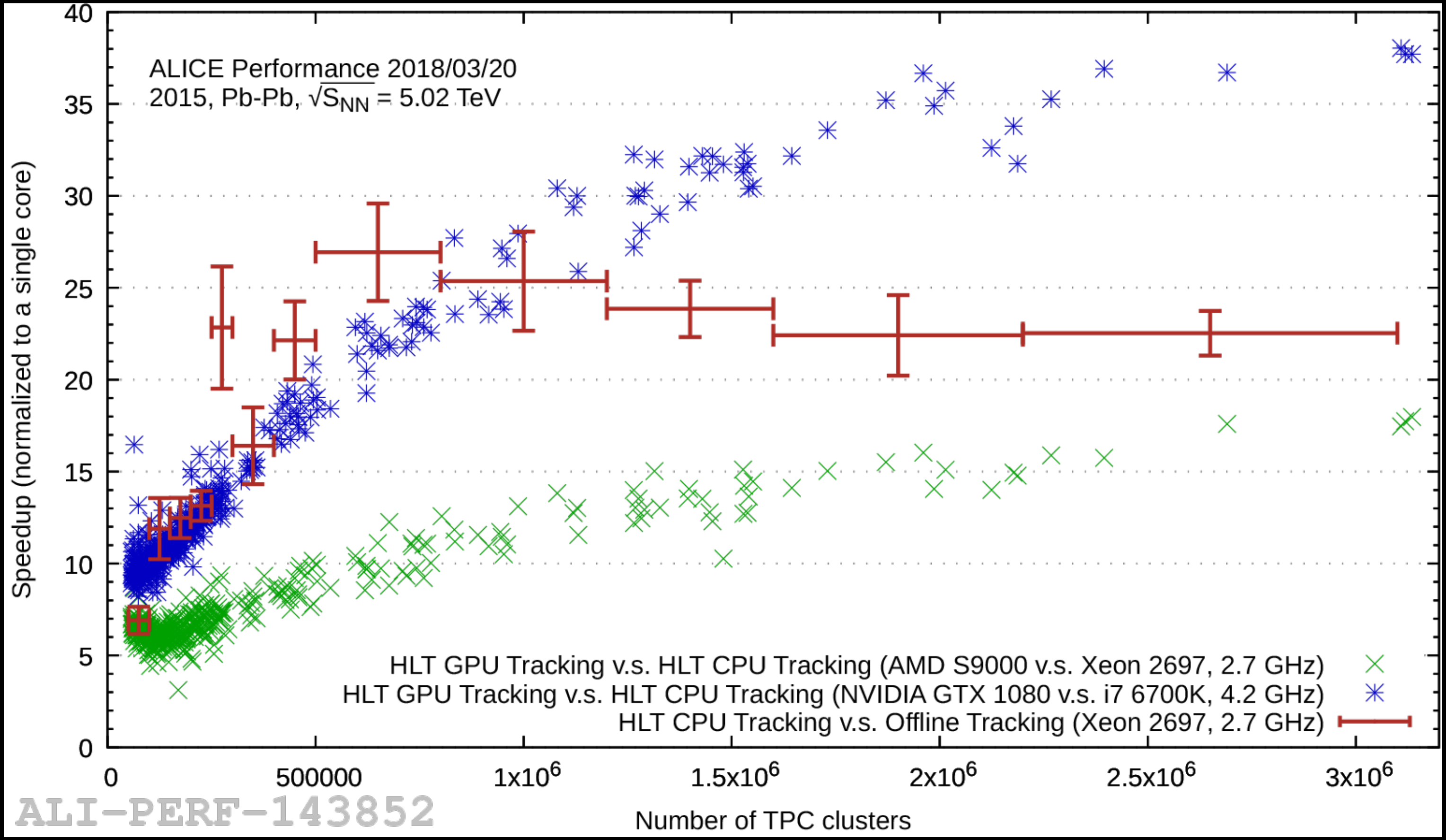


Helix (track model)

- Cellular Automaton for finding short track candidates (track forming and concatenating)
- Kalman filter for track fitting and extrapolation (track following)

https://arxiv.org/abs/1709.00618

ALICE Tracking Performance on GPUs

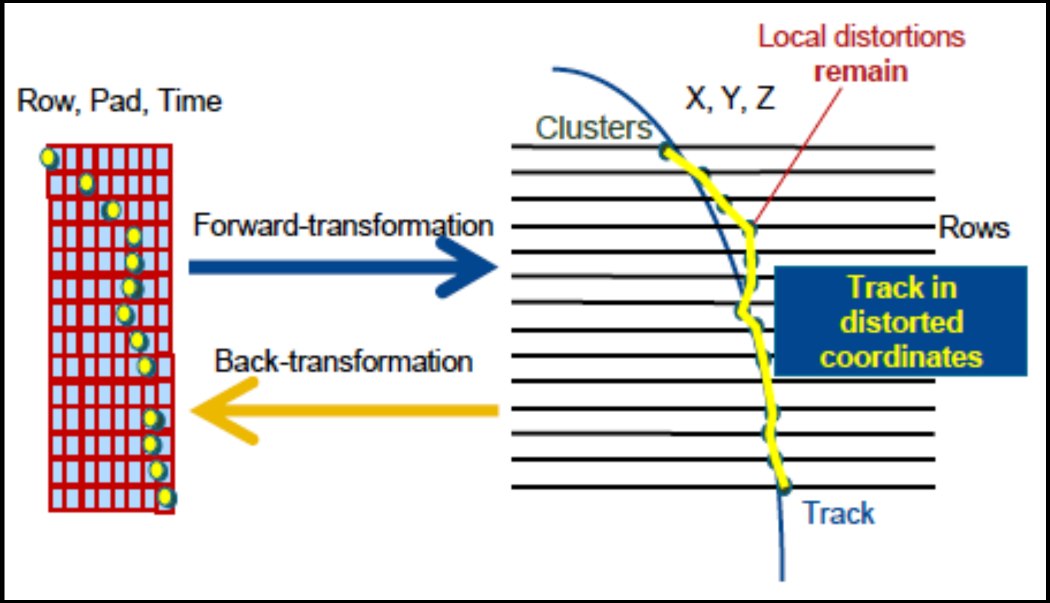


- Modern GPU replaces 40 CPU cores (4.2 GHz)
- 20 ms timeframe tracking needs ~20 s on GPU

  → ~1500 GPUs for synchronous ALICE tracking

ALICE Data Compression

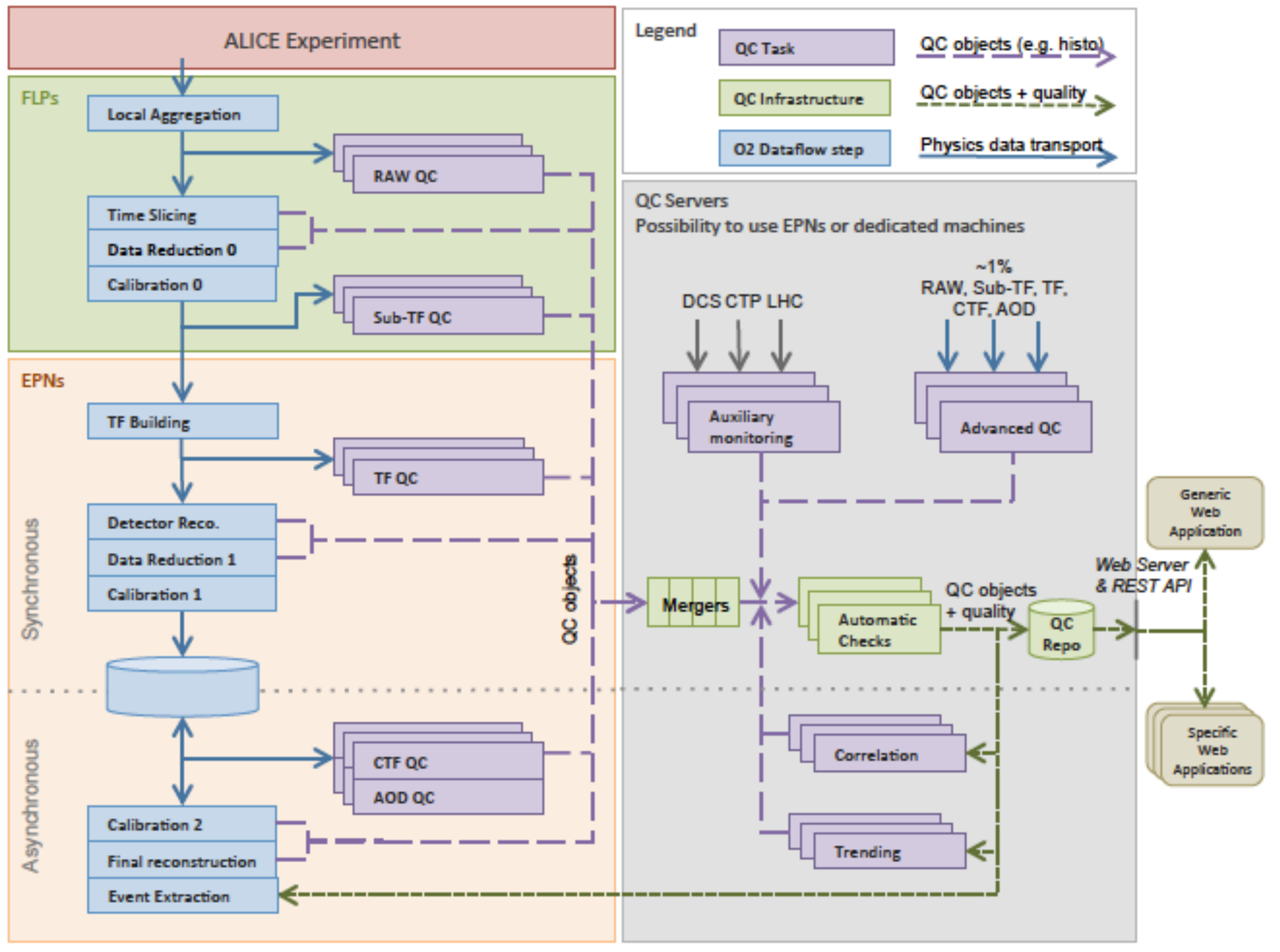
Required online data compression by factor ~20 : 1 TB/s → 50 GB/s



- Non-lossless compression
  - Clusters finding with FPGAs
  - Removal of clusters of low momentum tracks
- Lossless compression
  - Huffman or arithmetic entropy encoding
  - Storing only residuals to the clusters but not all cluster coordinates

https://arxiv.org/abs/1709.00618

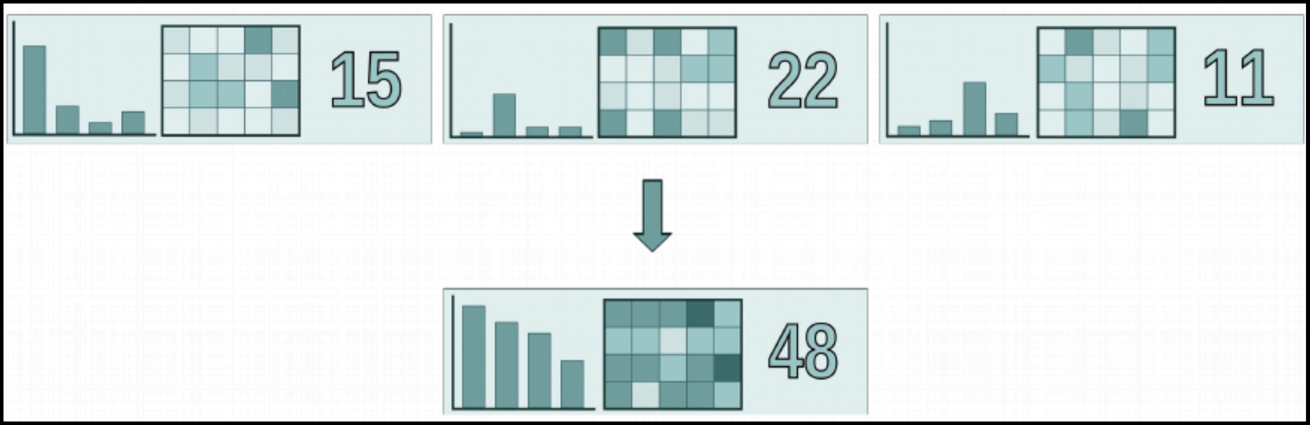
ALICE O^2 Data Quality Control

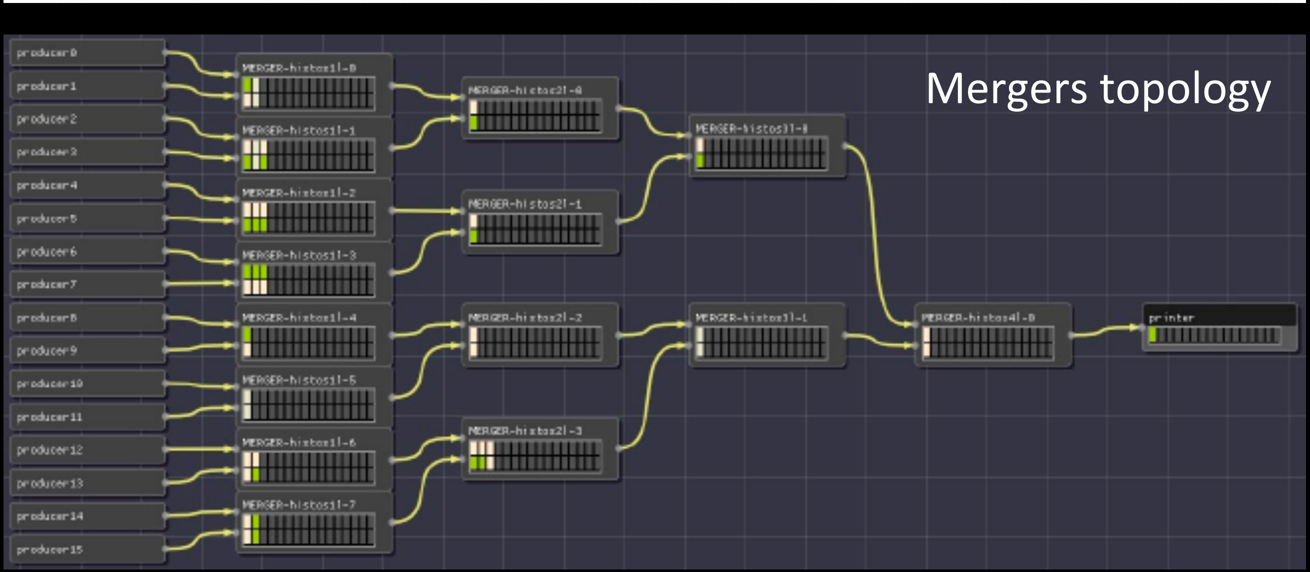


https://cds.cern.ch/record/2011297

Development in collaboration with AGH (Department of Automatic Control and Robotics EAIiIB, Department of Computer Science IET)

Data Mergers

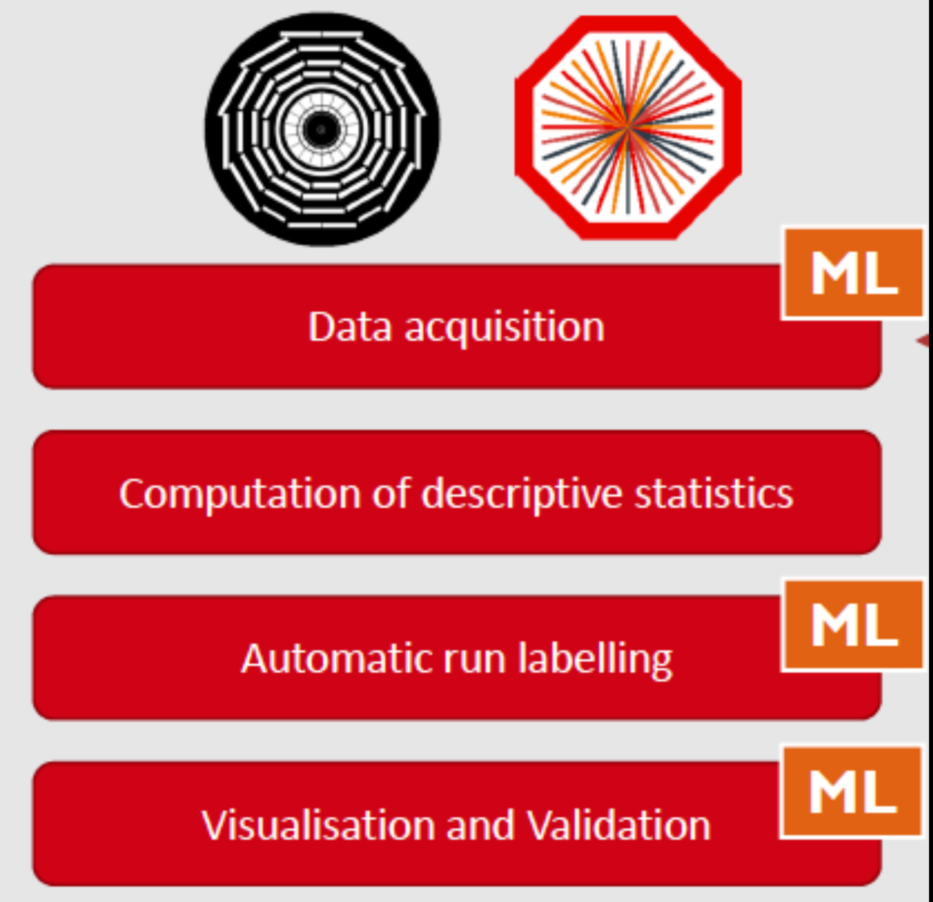




Mergers topology

Author: Piotr Konopka (PhD student AGH/CERN)
Tests and benchmarks on Prometheus cluster: Paweł Palimąka (Master student AGH)

Data Quality Control and Machine Learning



ML

Data acquisition

Computation of descriptive statistics

ML

Automatic run labelling

ML

Visualisation and Validation

Classification of anomalies (needed: labeled dataset)

Regression of one value which may indicate anomalies (needed: dataset with known values)

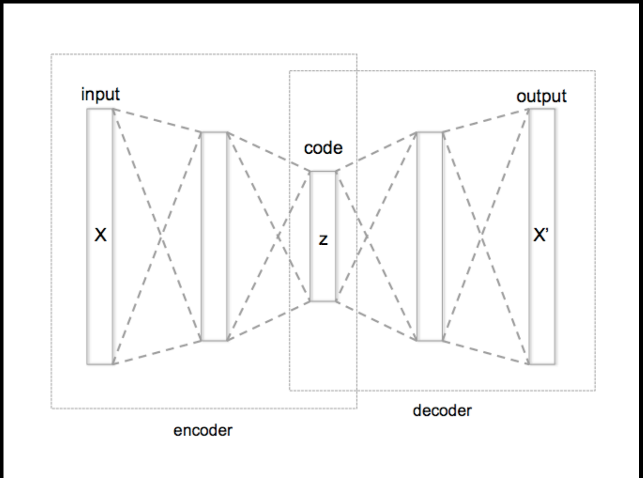
Clustering of unknown data and searching for outliers (needed: noisy data)

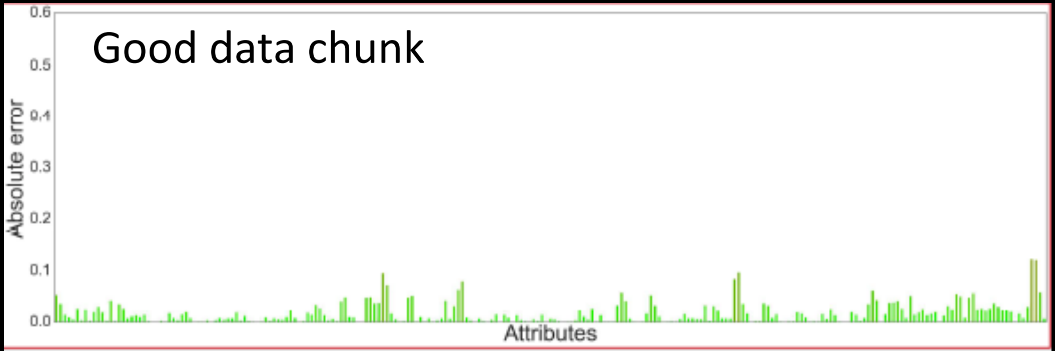
Dimensionality reduction for sparse data representation and searching of outliers (needed: high dimensional data)

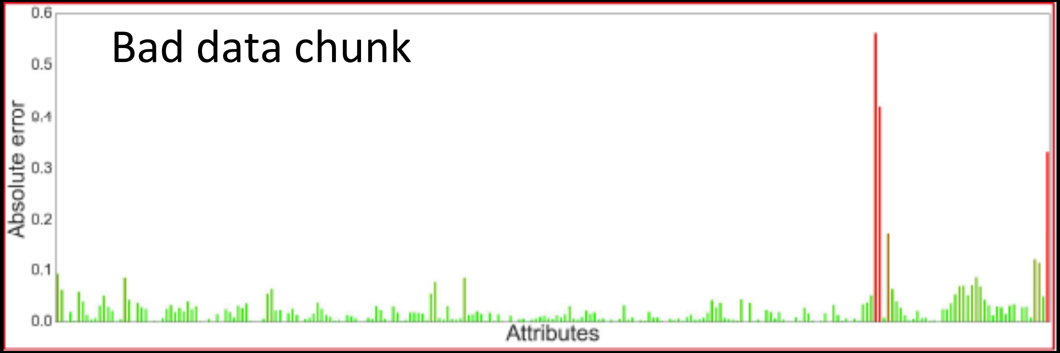
Unsupervised Learning with Autoencoders

- 2508 data chunks (91 warnings, 71 outliers)
- One data chunk ~ 15 min. time interval
- 242 attributes

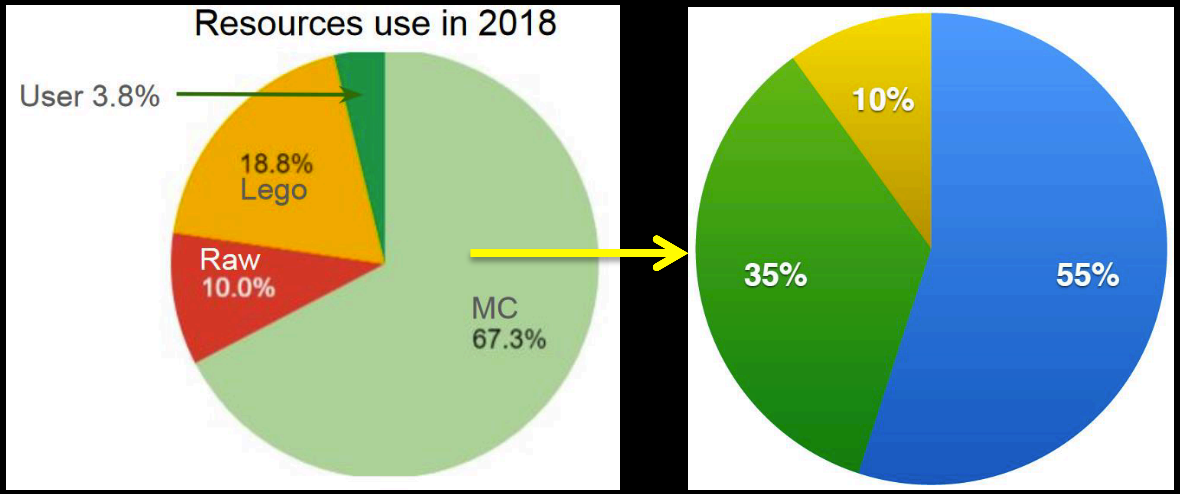
- Deep bottleneck autoencoder with 5 fully connected layers







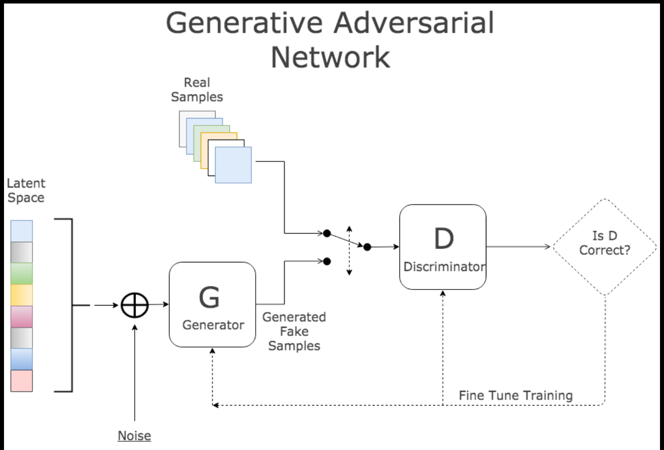
Monte-Carlo Simulations in ALICE



- particle propagation through detector
- digitization (detector response)
- rest (rec., QA/QC,…)

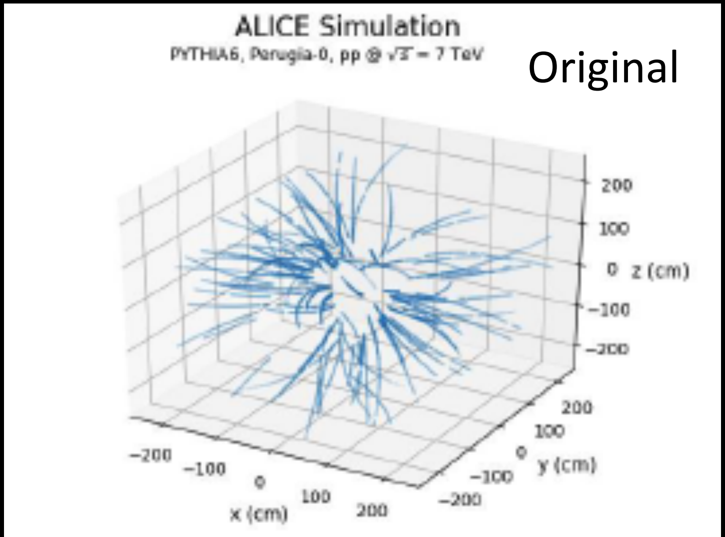
- More than 2/3 resources spend on MC simulations (Geant3/Geant4) in Run-2
- Expected 100 times more data in Run-3 and Run-4
  → Cannot be covered with the current simulation software
- Possible directions: fast simulations, embedding, optimizing current software…

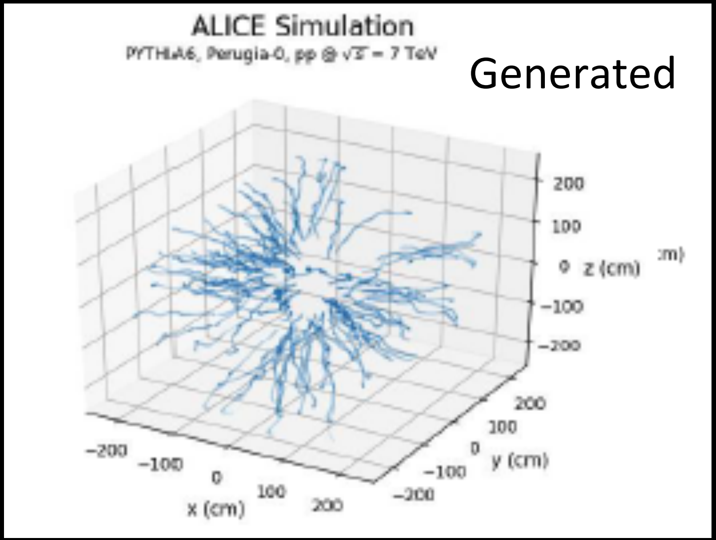
Cluster Simulations with Generative Adversarial Networks (GANs)



- Training on original reconstructed data
- Conditional Deep Convolutional GAN

- Simulation speed-up ~25 (CPU), ~250 (GPU)
- But we are not there yet…





ALICE Collaboration

- 41 countries, ~176 institutions, ~1800 scientists
- Opportunities for master and PhD students in ALICE
  - Development of novel O^2 computing system under good supervision
  - CERN student programmes (paid by CERN)
    - Summer student programme
    - Technical student programme
    - CERN Openlab summer student programme
    - Short-term internship programme
  - CERN doctoral programme (paid by CERN)
  - Short-term internships at CERN sponsored by ALICE

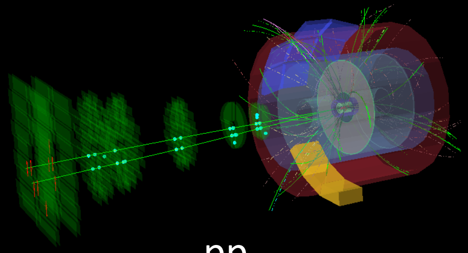
https://jobs.web.cern.ch/join-us/students

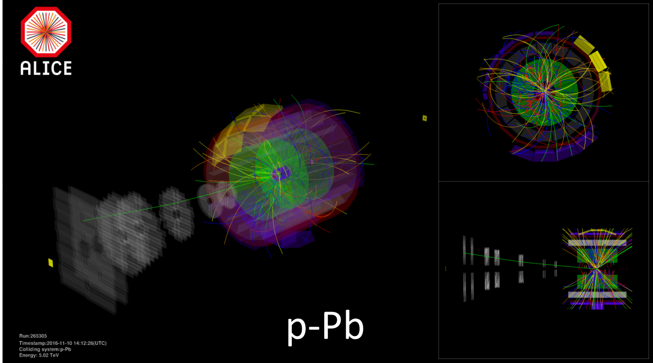
You are welcome to join and participate in developments!
Contact: jacek.otwinowski@ifj.edu.pl

Backup

ALICE at work since 2009

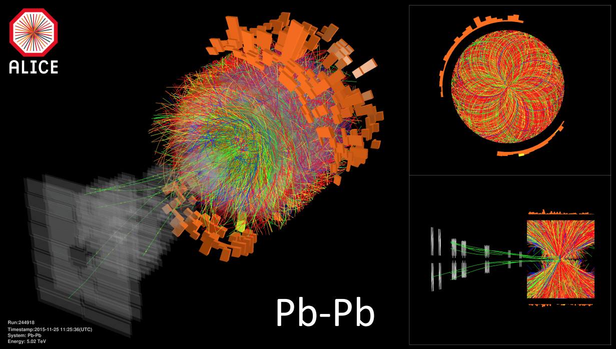
System	Year	$\sqrt{s_{NN}}$ (TeV)	L_{int}
Pb-Pb	2010-2011 2015 2018	2.76 5.02 5.02	~75 μb^{-1} ~250 μb^{-1} ~0.9 nb^{-1}
Xe-Xe	2017	5.44	~0.3 μb^{-1}
p-Pb	2013 2016	5.02 5.02, 8.16	~15 nb^{-1} ~3 nb^{-1}, ~25 nb^{-1}
pp	2009-2013 2015-2018	0.9, 2.76, 7, 8 5.02, 13	~200 μb^{-1}, ~100 μb^{-1}, ~1.5 pb^{-1}, ~2.5 pb^{-1} ~1.3 pb^{-1} , ~59 pb^{-1}



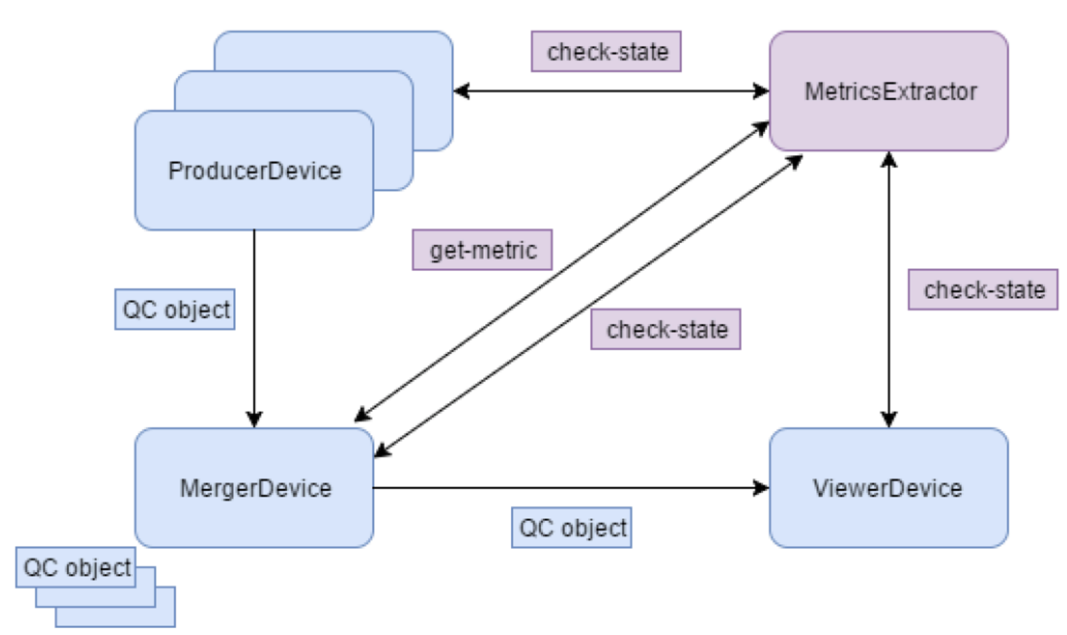


- Energy and system dependence studies of particle production are possible
- Large statistics of pp, p-Pb and Pb-Pb collisions at the same $\sqrt{s_{NN}}$

→ precise comparison studies



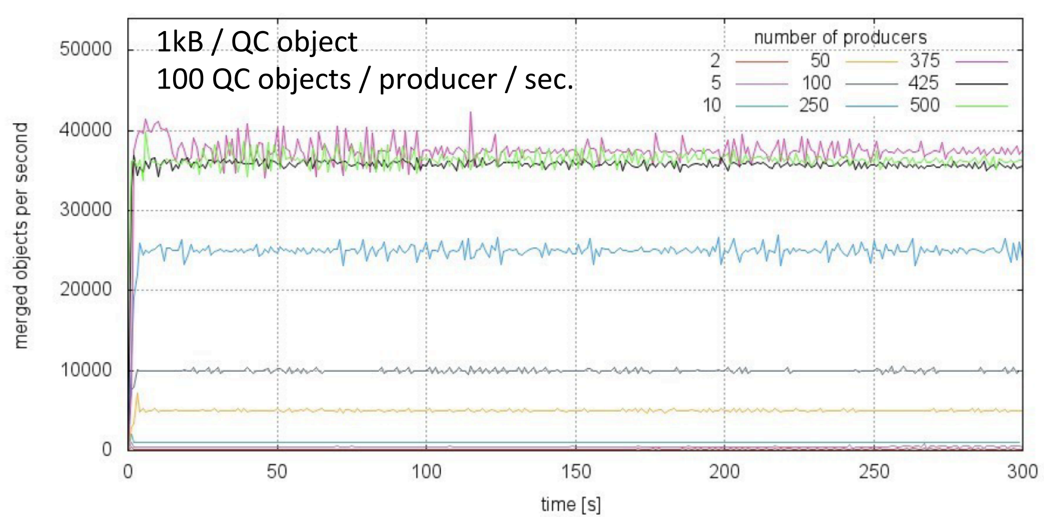
First prototype of merger device



Parameters: buffer size, QC object size, QC object type, number of producers per merger…

Metrics: CPU usage, RAM usage, average merging time, merged objects per second …

Execution on PL-Grid (Prometheus)



Patryk Lesiak,
Master Thesis 2016, Faculty of Physics and Applied Computer Science AGH.