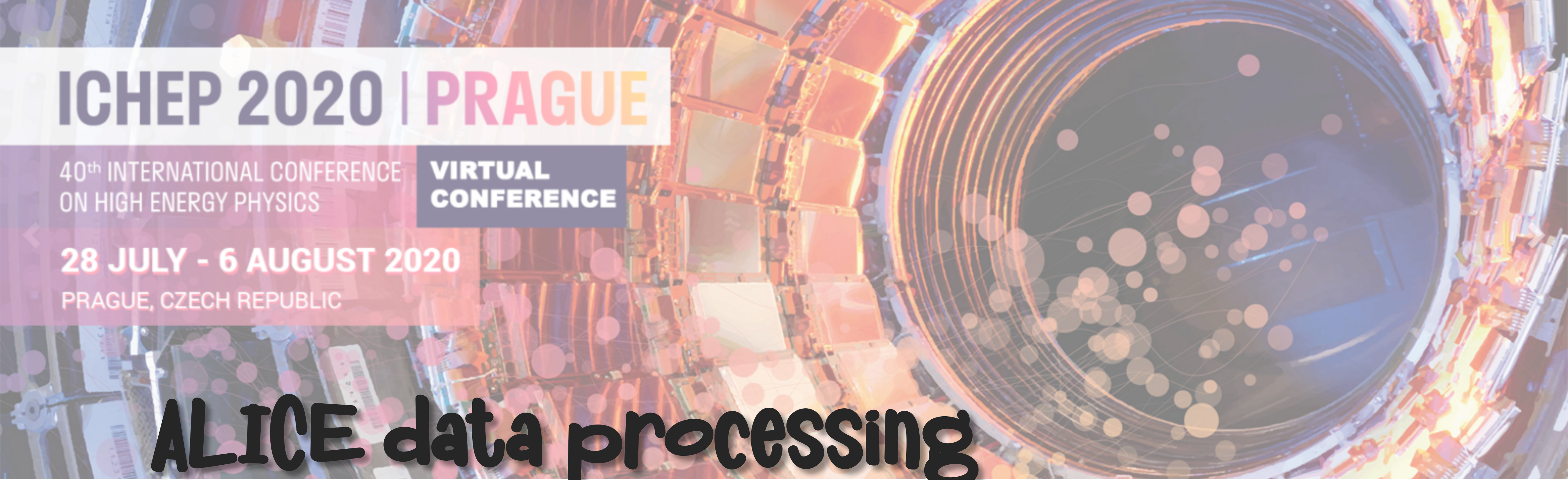


ALICE data processing for Run 3 and Run 4 at the LHC

40th INTERNATIONAL CONFERENCE ON HIGH ENERGY PHYSICS,

28 JULY – 6 AUGUST 2020, PRAGUE, CZECH REPUBLIC, VIRTUAL CONFERENCE

CHIARA ZAMPOLLI(*) for the ALICE COLLABORATION

(*) Chiara.Zampolli@cern.ch

ALICE

CERN

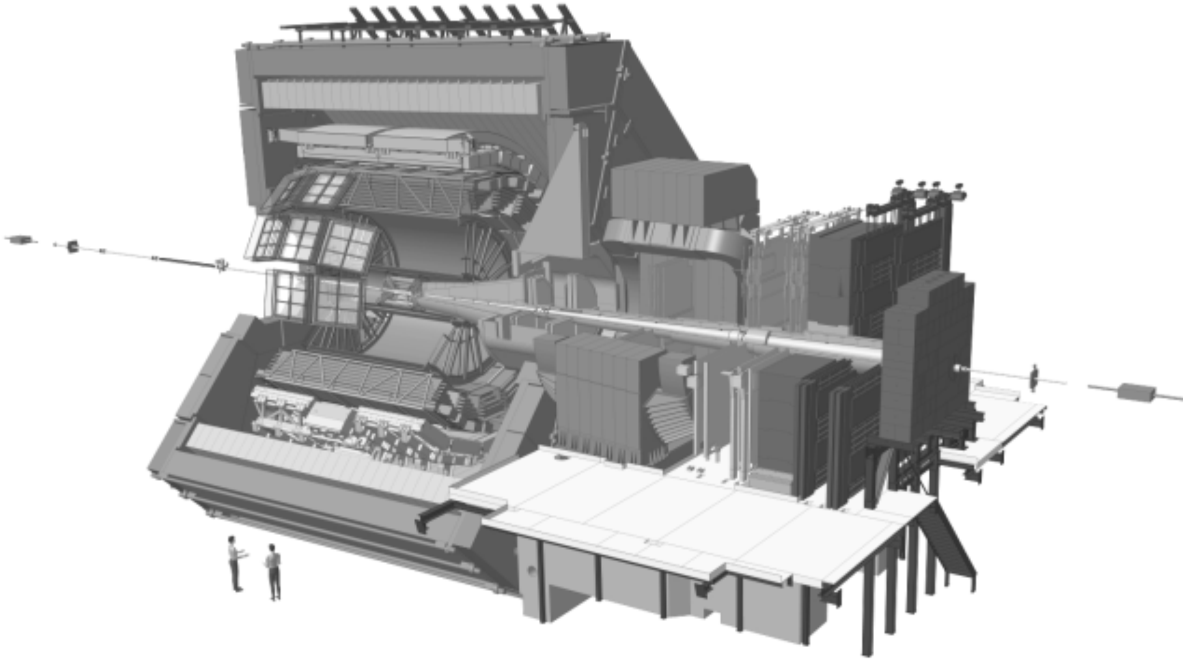
Outline – ALICE in Run 3 and Run 4

- ALICE upgrade
- ALICE data processing
- Synchronous processing
- TPC space-charge distortions
- Central barrel tracking
- O^2 processing model
- Analysis framework
- Summary and conclusions

See also:

S. Panebianco, 28th July at 19:20, session 12
M. Concas, 30th July at 8:20, this session
M. Lettrich, 30th July at 9:20, this session

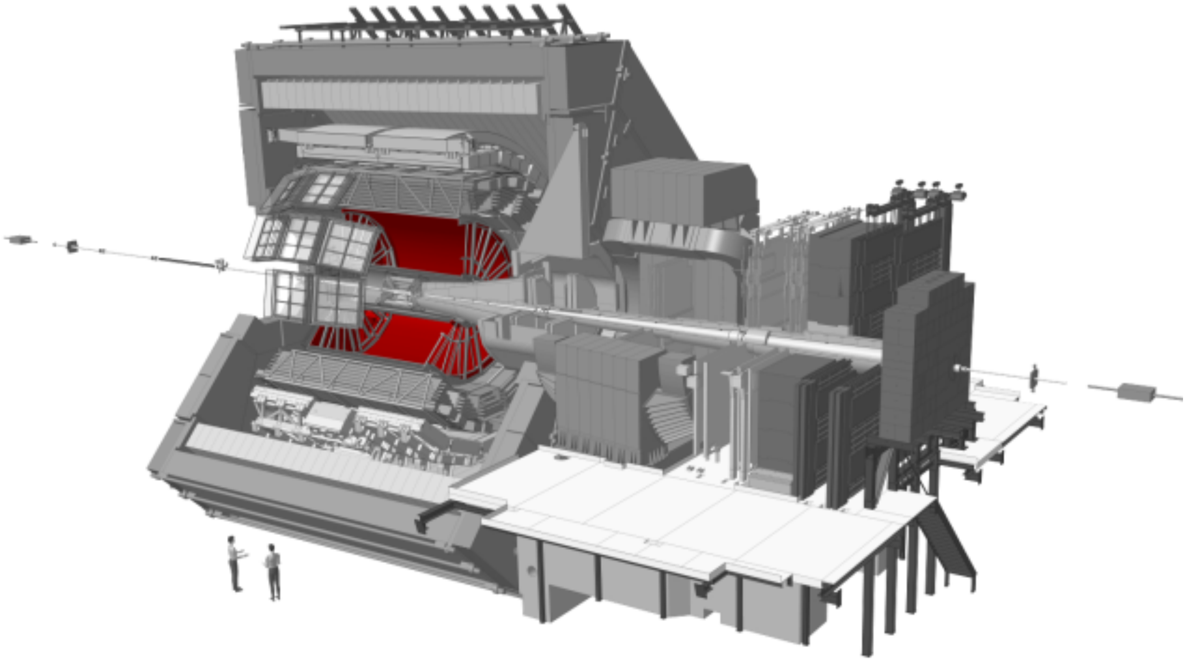
A Large Ion Collider Experiment



Dedicated heavy-ion experiment

- Barrel tracking detectors with geometrical acceptance $|\eta| < 0.9$ and full ϕ
- Precision tracking capabilities in $|\eta| < 0.9$ down to very low momenta (<100 MeV/*c*, low B-field)
- Different particle-identification detectors, some with limited geometrical acceptance
- Excellent hadron identification from low to high momenta
- Tracking detectors optimized for extremely high charged track multiplicities

A Large Ion Collider Experiment



Dedicated heavy-ion experiment

- Barrel tracking detectors with geometrical acceptance |η| < 0.9 and full ϕ
- Precision tracking capabilities in |η| < 0.9 down to very low momenta (<100 MeV/*c*, low B-field)
- Different particle-identification detectors, some with limited geometrical acceptance
- Excellent hadron identification from low to high momenta
- Tracking detectors optimized for extremely high charged track multiplicities

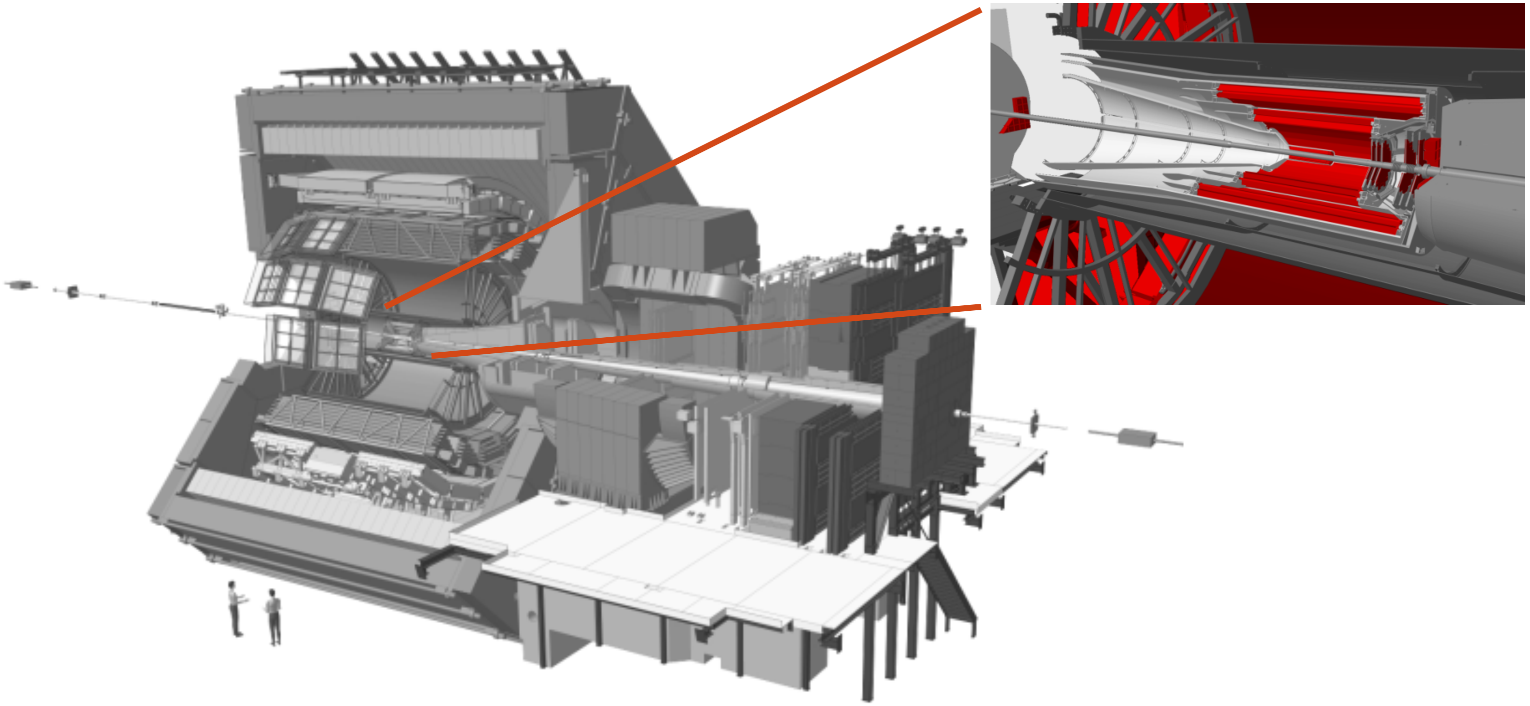
Run 2:

- Pb-Pb Interaction Rate (IR) ~7-10 kHz, trigger rate < 1 kHz limited by **TPC** readout (<3.5 kHz) and bandwidth
- Collected luminosity ($\mathcal{L}$): **~1 nb^{-1}**

Run 3 + Run 4:

- Pb-Pb IR = **50 kHz** (but also pp!), **continuous** readout
- Goal: $\mathcal{L}$ **~10 nb^{-1}** (B = 0.5 T) + **3 nb^{-1}** (B = 0.2 T)

A Large Ion Collider Experiment



Upgraded detectors:

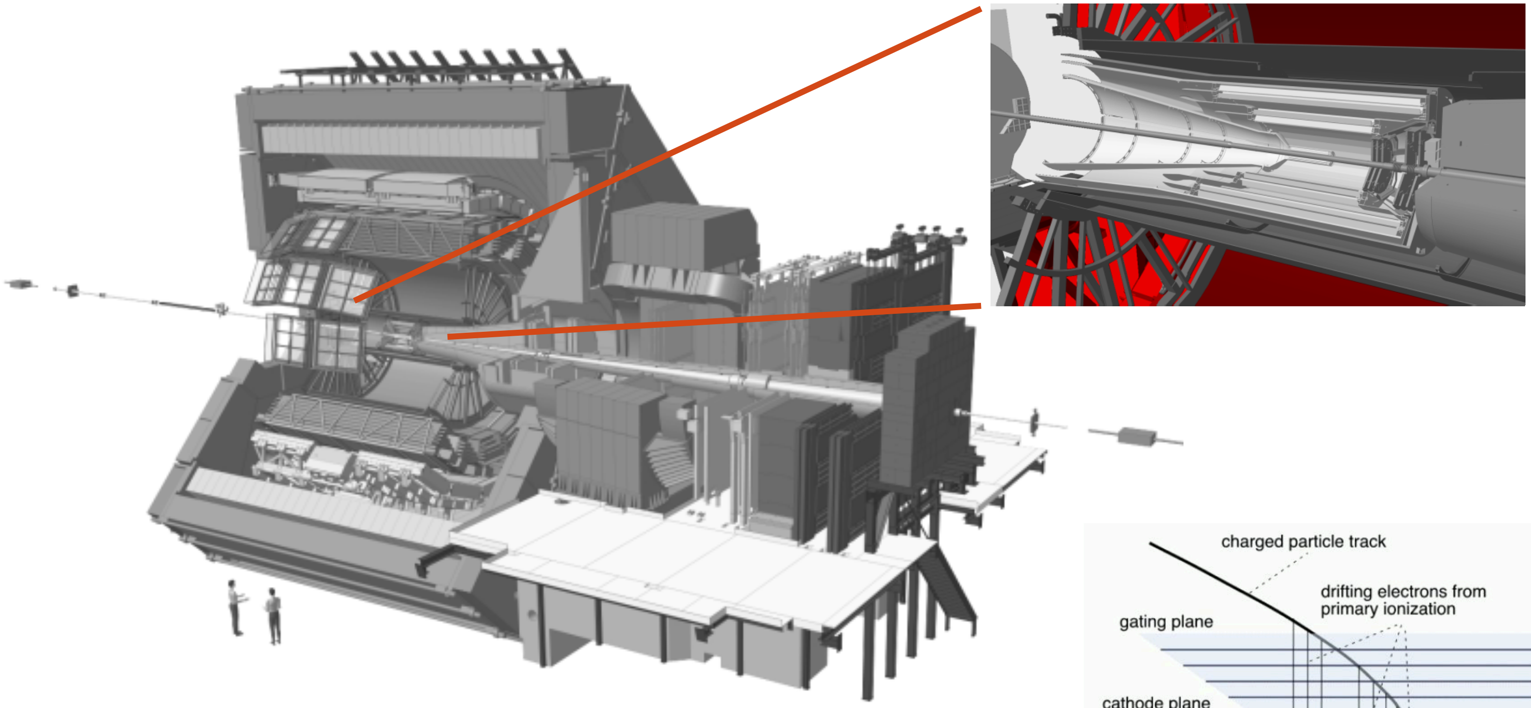
ITS, TPC, MFT, FIT

See S. Panebianco

Run 3 + Run 4:

- Pb-Pb IR = **50 kHz** (but also pp!), **continuous** readout
- Goal: $\mathcal{L}$ **~10 nb^{-1}** (B = 0.5 T) + **3 nb^{-1}** (B = 0.2 T)

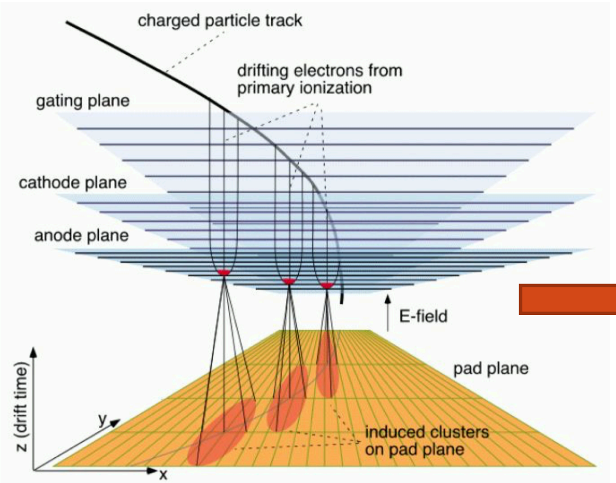
A Large Ion Collider Experiment



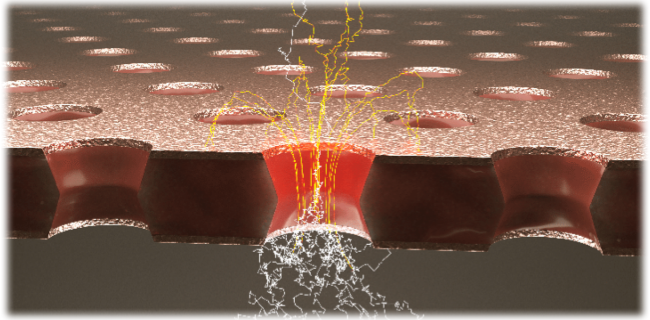
TPC: from MPWC to GEM

Intense R&D studies for detector optimization

- GEM stacks with 4 layers
- Highly optimized HV configuration



MPWC: triggered readout

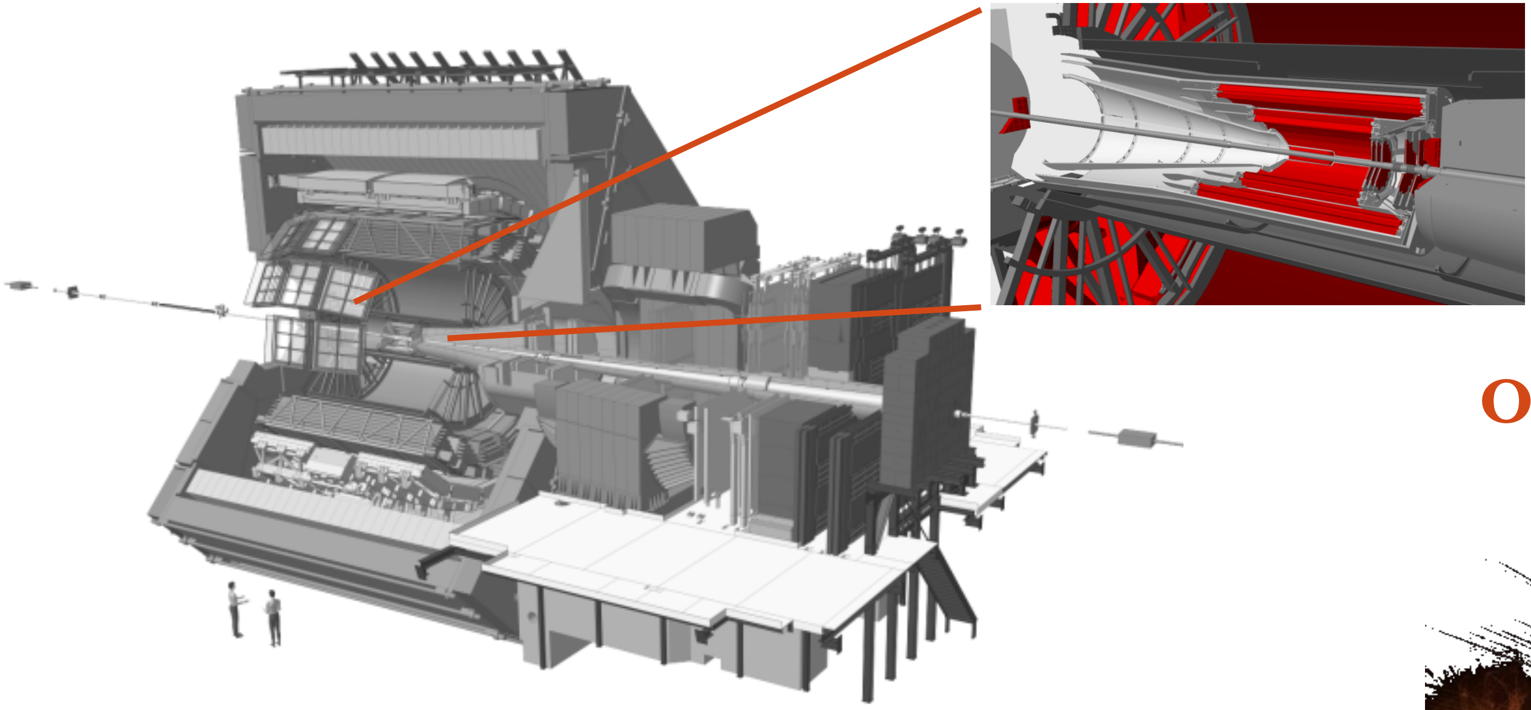


GEM: continuous readout

Run 3 + Run 4:

- Pb-Pb IR = **50 kHz** (but also pp!), **continuous**
- Goal: $\mathcal{L}$ **~10 nb^{-1}** (B = 0.5 T) + **3 nb^{-1}** (B = 0.2 T)

A Large Ion Collider Experiment



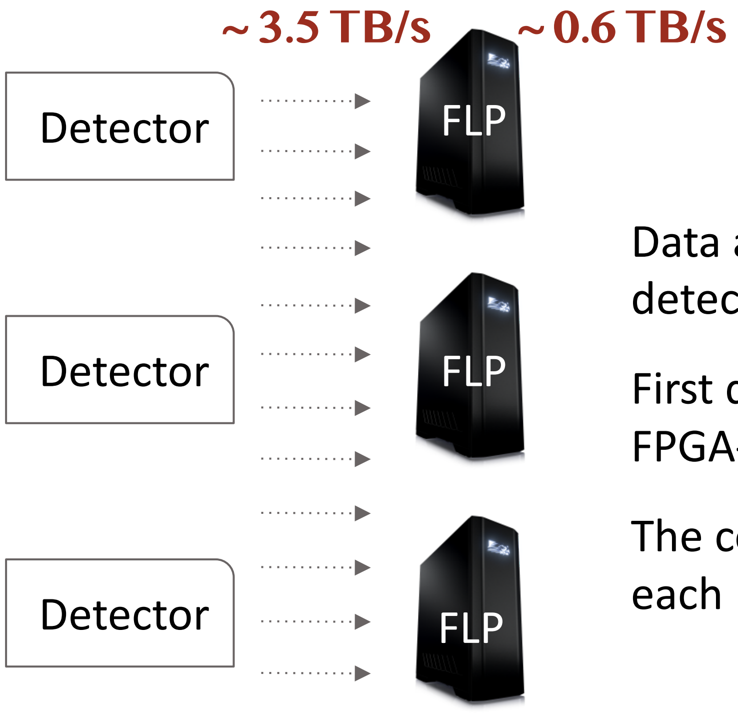
Upgraded detectors:
ITS, TPC, MFT, FIT

Upgraded data processing:
O^2 (Online-Offline processing)

Run 3 + Run 4:

- Pb-Pb IR = **50 kHz** (but also pp!), **continuous**
- Goal: $\mathcal{L}$ **~10 nb^{-1}** (B = 0.5 T) + **3 nb^{-1}** (B = 0.2 T)

ALICE data processing in Run 3 + 4



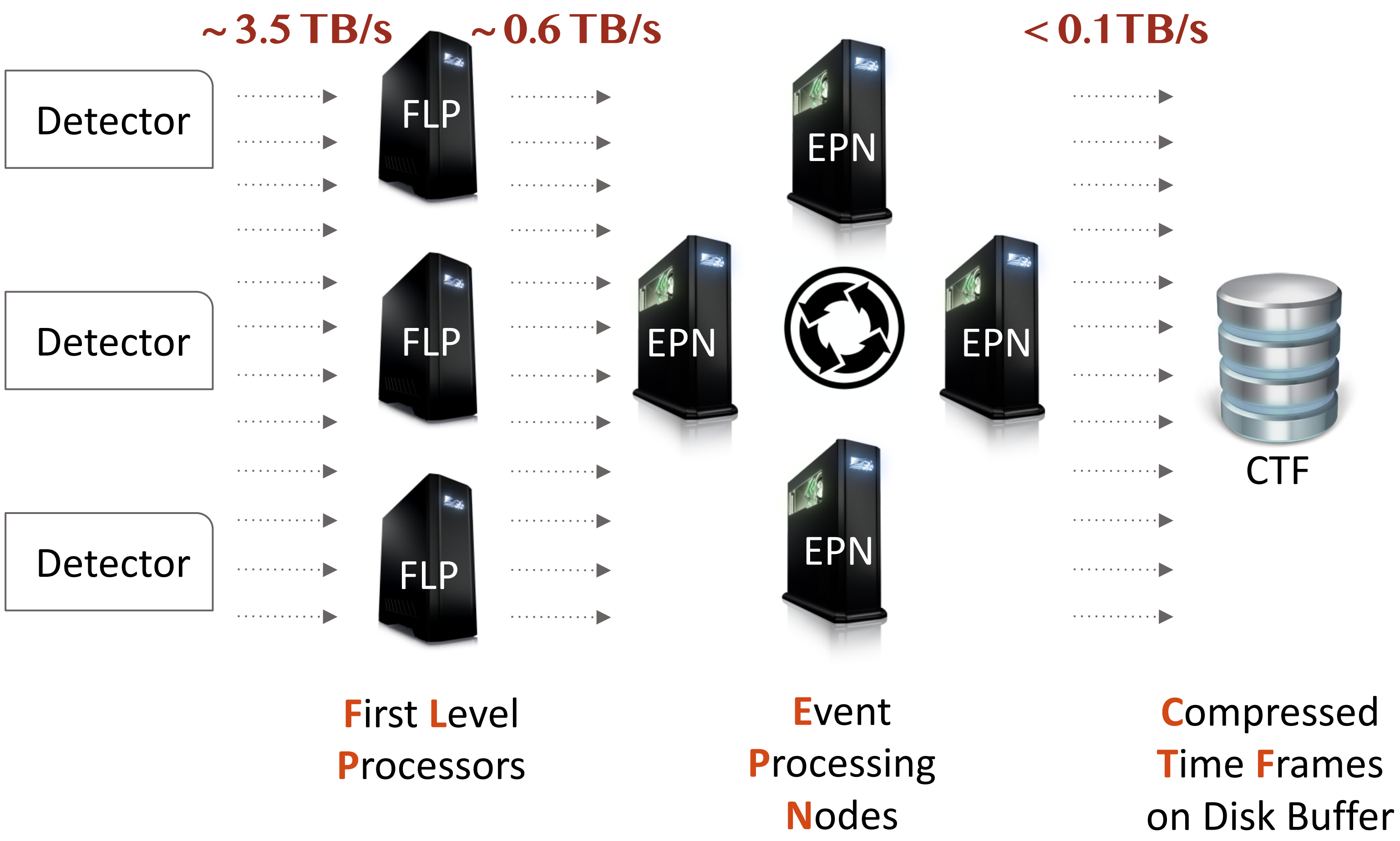
First **L**evel **P**rocessors

Data arrive at the First Level Processing nodes (FLP) from the detectors' readout links.

First data compression (zero suppression) is performed inside FPGA-based readout cards (Common Readout Unit).

The continuous data are divided into Sub-**T**ime **F**rames (TFs) on each FLP, 1 TF = 10-20 ms long (128-256 orbits) packets.

ALICE data processing in Run 3 + 4



Sub-Time Frames are merged together into complete TFs on the EPNs.

Synchronous reconstruction, calibration, data compression is performed.

Size of the farm (FLPs with CPUs + EPNs with CPUs and GPUs) such to cope with the peak rate of 50 kHz Pb-Pb data taking.

Compressed Time Frames are written to a 60 PB disk buffer, enough to keep the foreseen 1 month long Pb-Pb data sample.

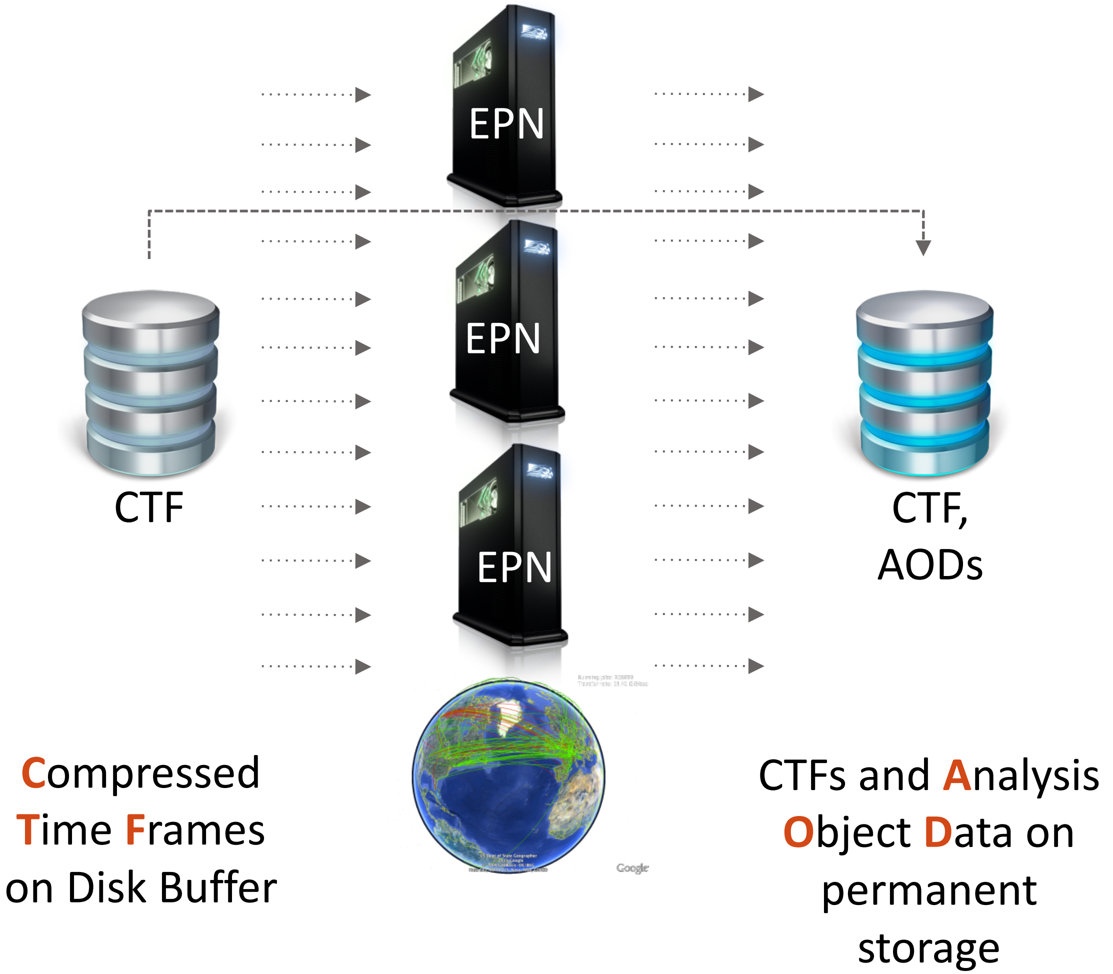
Total compression factor from detectors to disk: **35** (N.B.: cannot be compared to Run 2 due to different raw data format).

ALICE data processing in Run 3 + 4

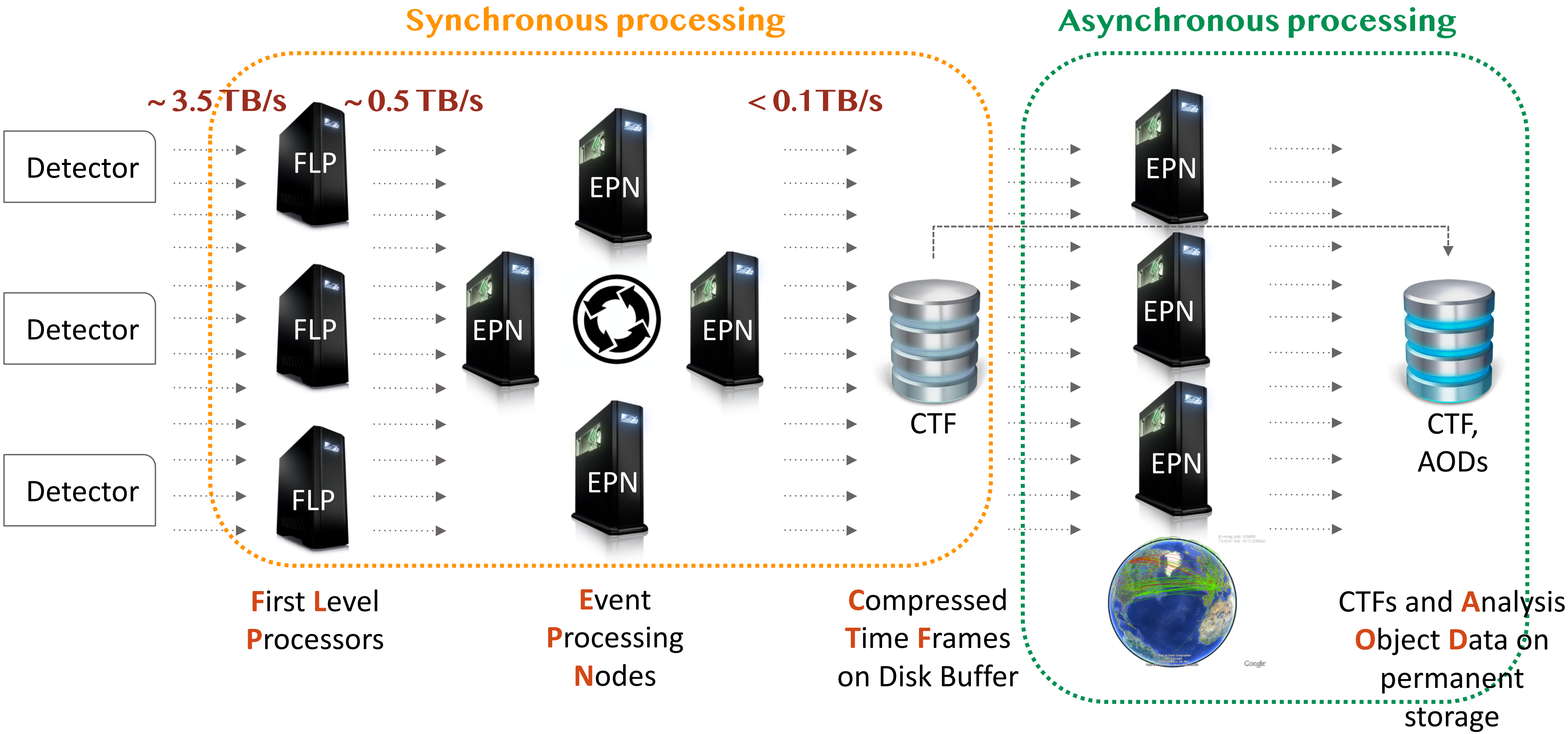
At **asynchronous** stage, a second (and possibly third) reconstruction with final calibration will be run on the O^2 EPN farm and the T0/T1s (1/3 each).

Final Analysis Object Data (AOD) will be produced and saved on permanent storage.

CTFs will be deleted from the disk buffer to make space for new data



ALICE data processing in Run 3 + 4



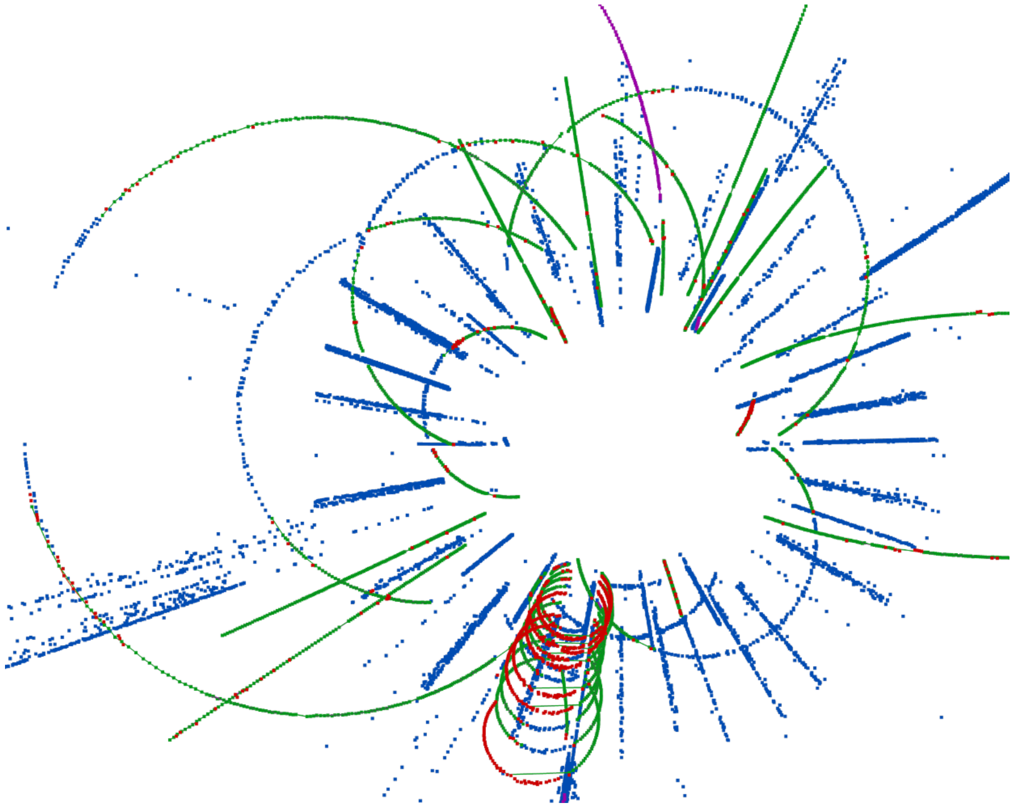
Synchronous processing

Goal of synchronous reconstruction is to reach factor 35 of compression.

Most relevant detector is TPC: from 3.4 TB/s to 70 GB/s

TPC data compression will consist of:

- Zero suppression
- Clusterization
- Optimized data format
- Entropy reduction
- TPC tracking, to remove clusters not associated to tracks
- Remaining clusters entropy-compressed with ANS encoding



Unassigned clusters (straight lines correspond to noisy pads)
Removed clusters
Reconstructed tracks
Failed fit

See M. Lettrich

Synchronous processing

Goal of synchronous reconstruction is to reach factor 35 of compression.

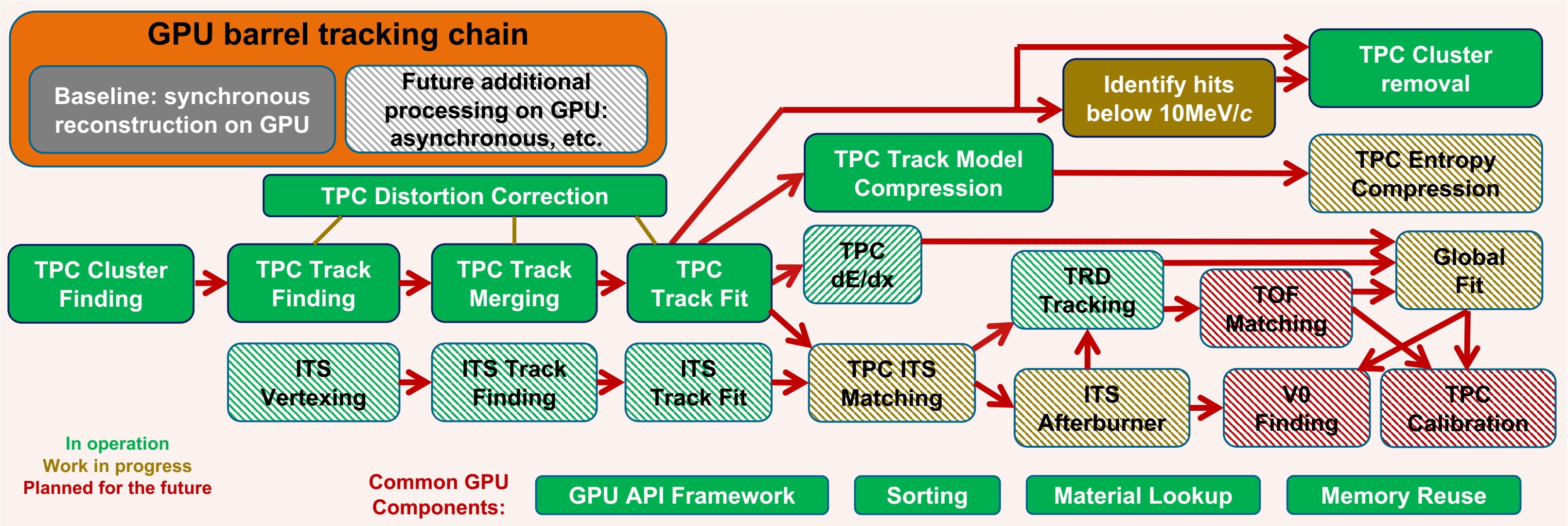
Most relevant detector is TPC: from 3.4 TB/s to 70 GB/s

TPC data compression will consist of:

- **Clusterization**
- **TPC tracking**

USE OF GPUs MANDATORY

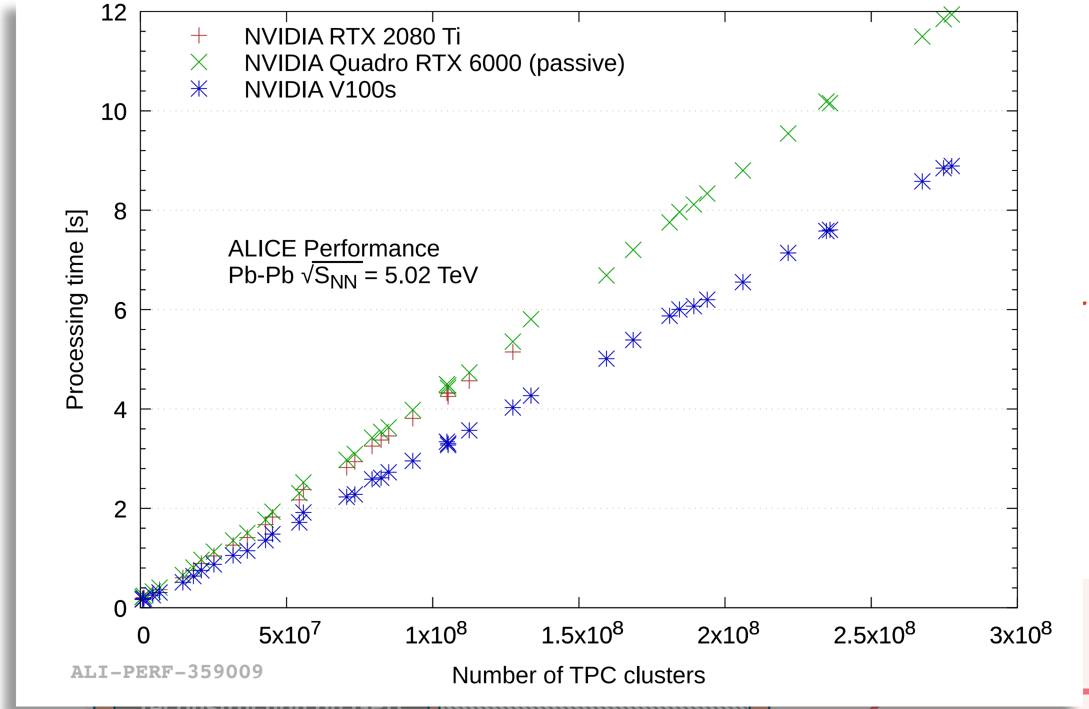
> 40x faster than CPU but only 4x more expensive



See M. Concas

Plan to exploit GPUs computing power also during asynchronous reconstruction

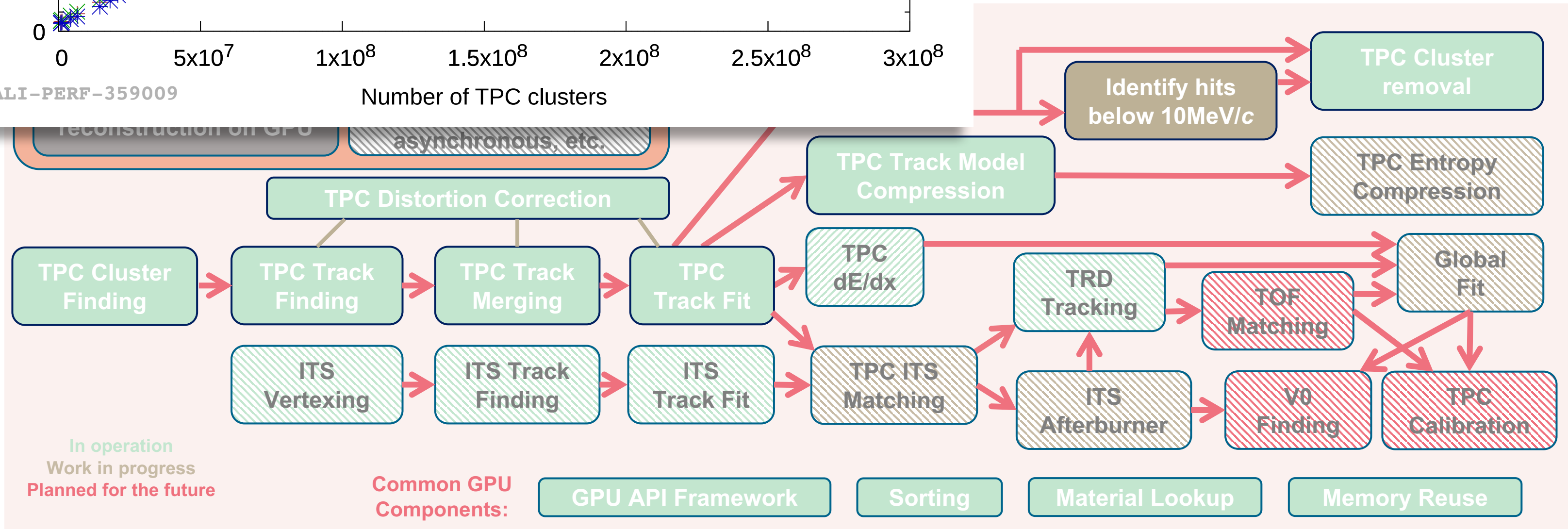
Synchronous processing



Linear dependence of GPU processing time vs number of TPC clusters tested up to 256 orbits TF → TF length does not impact on the number of needed GPUs

USE OF GPUs MANDATORY

> 40x faster than CPU but only 4x more expensive



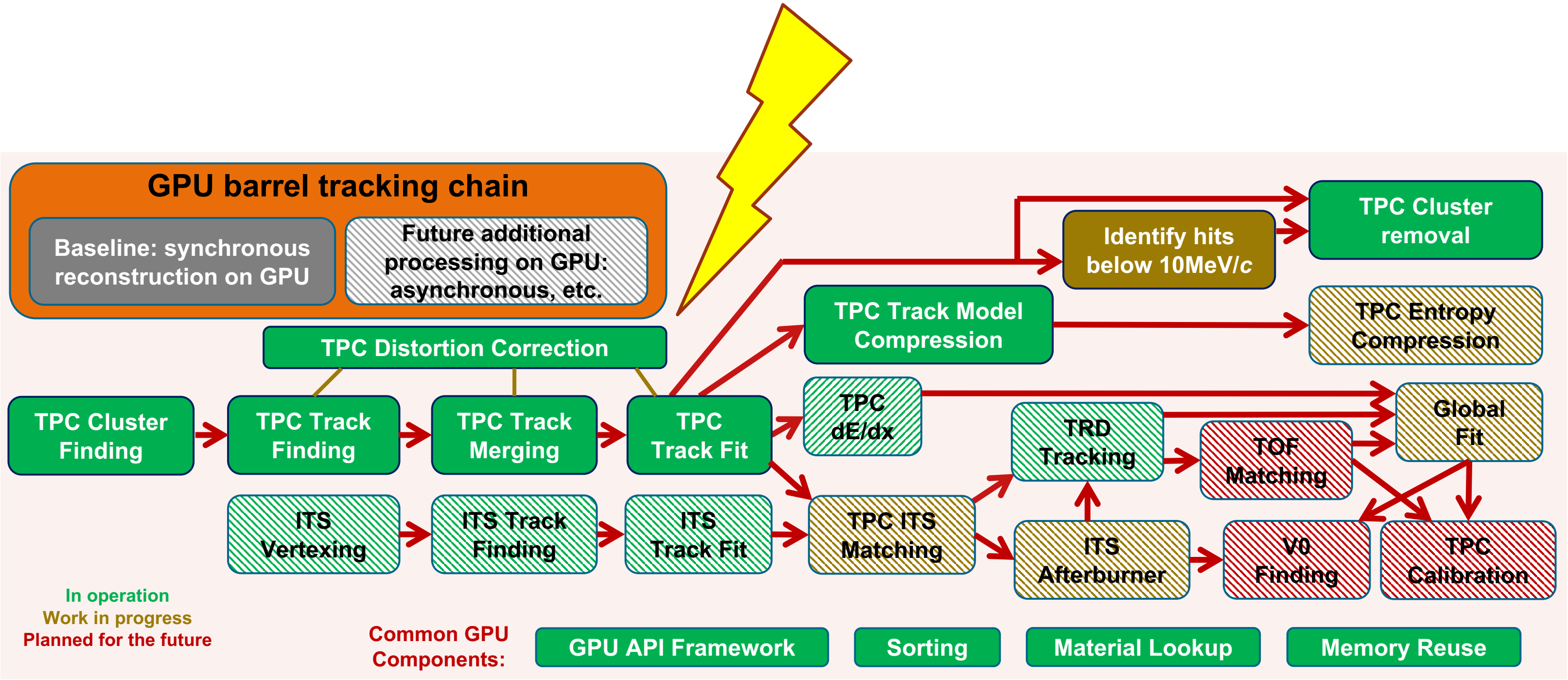
See M. Concas

Plan to exploit GPUs computing power also during asynchronous reconstruction

Synchronous processing

USE OF GPUs MANDATORY

> 40x faster than CPU but only 4x more expensive



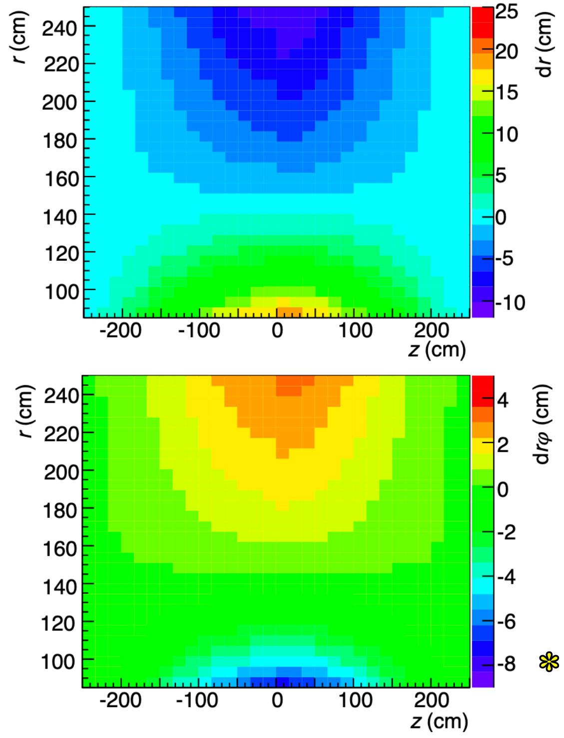
Plan to exploit GPUs computing power also during asynchronous reconstruction

Space-Charge Distortions in the TPC

TPC GEM configuration designed to reduce to the minimum the ion backflow (< 1%)

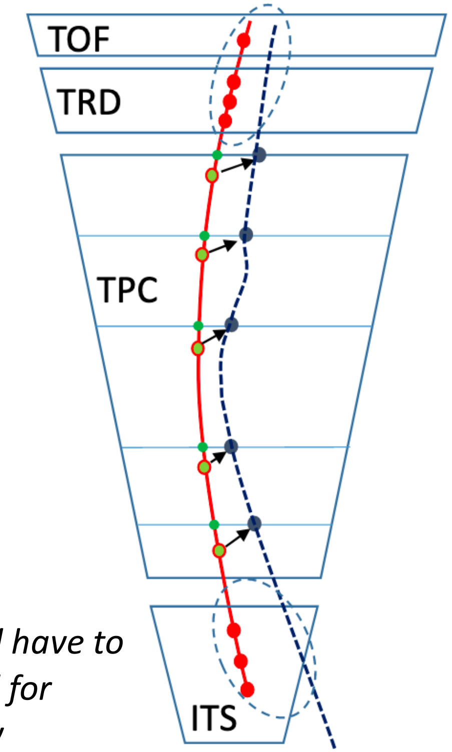
- Still, positive charge accumulating and moving in the TPC → modified E-field → distortions in the TPC

Expectations for Pb-Pb @ 50 kHz:



http://cds.cern.ch/record/1622286/

Run 2 strategy for average (*) distortions:



(*) Fluctuations will have to be accounted for separately

→ Interpolation of refitted ITS, TRD and TOF **track segments** to the TPC as **reference points** for the **true track position** — sync

→ Collect ΔY, ΔZ between **distorted clusters** and **references** in TPC sub-volumes (voxels) — sync

→ Extract 3D distortion vector in every voxel — offline

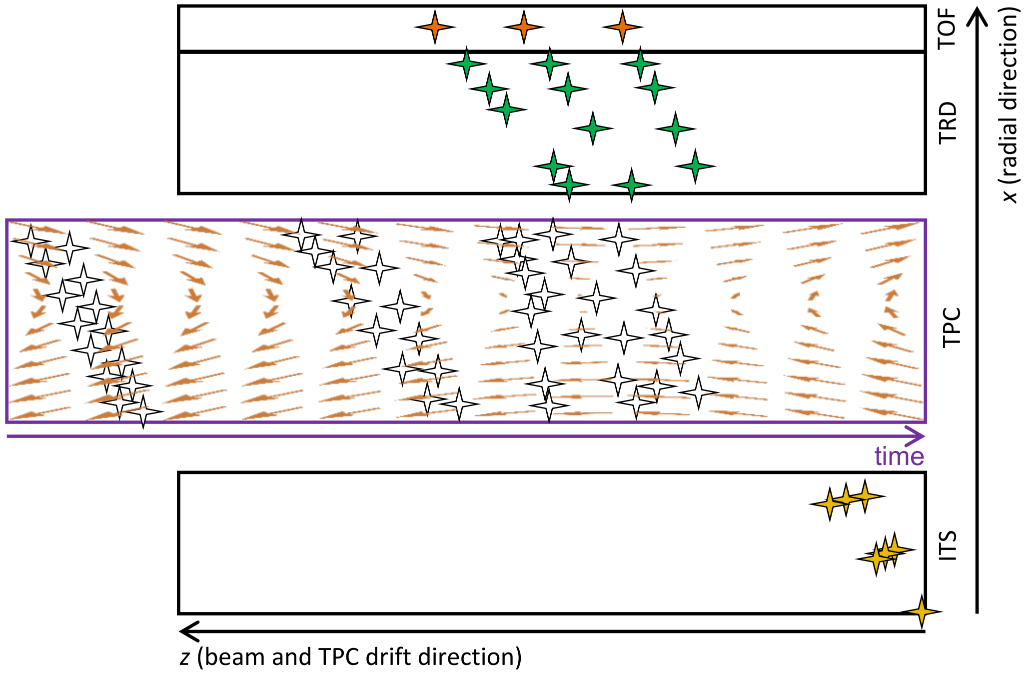
→ Use during asynchronous reconstruction — async

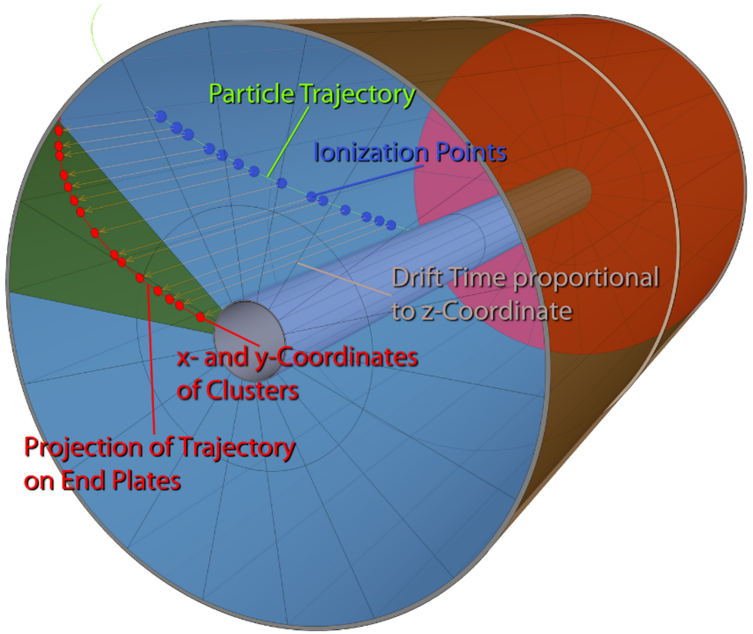
Full ITS-TPC-TRD-TOF reconstruction of a fraction (~0.6%) of the tracks at synchronous stage

Tracking in the central barrel

Challenge in Run 3 + 4!

- **overlap** of multiple collisions (5 collisions in the TPC drift time @50 kHz Pb-Pb)
- with TPC clusters without a well-defined *z* coordinate, but **just a time** (*t*)
- presence of **distortion** corrections that are position dependent



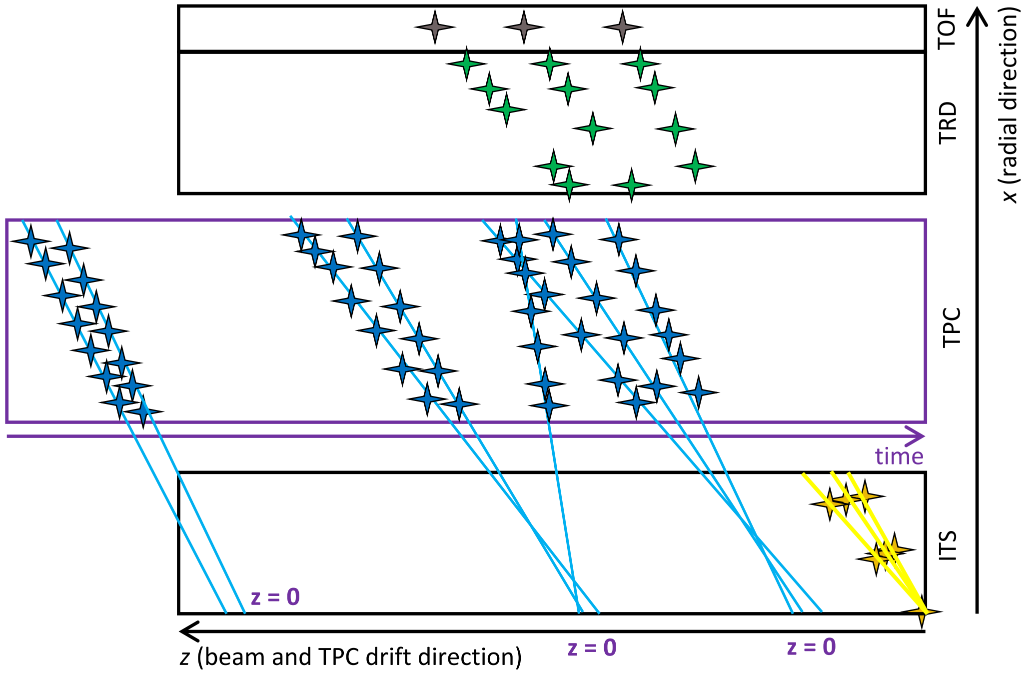


$$z = (t - t_{vertex}) * v_{drift}$$

Tracking in the central barrel

Challenge in Run 3 + 4!

- **overlap** of multiple collisions (5 collisions in the TPC drift time @50 kHz Pb-Pb)
- with TPC clusters without a well-defined *z* coordinate, but **just a time** (*t*)
- presence of **distortion** corrections that are position dependent

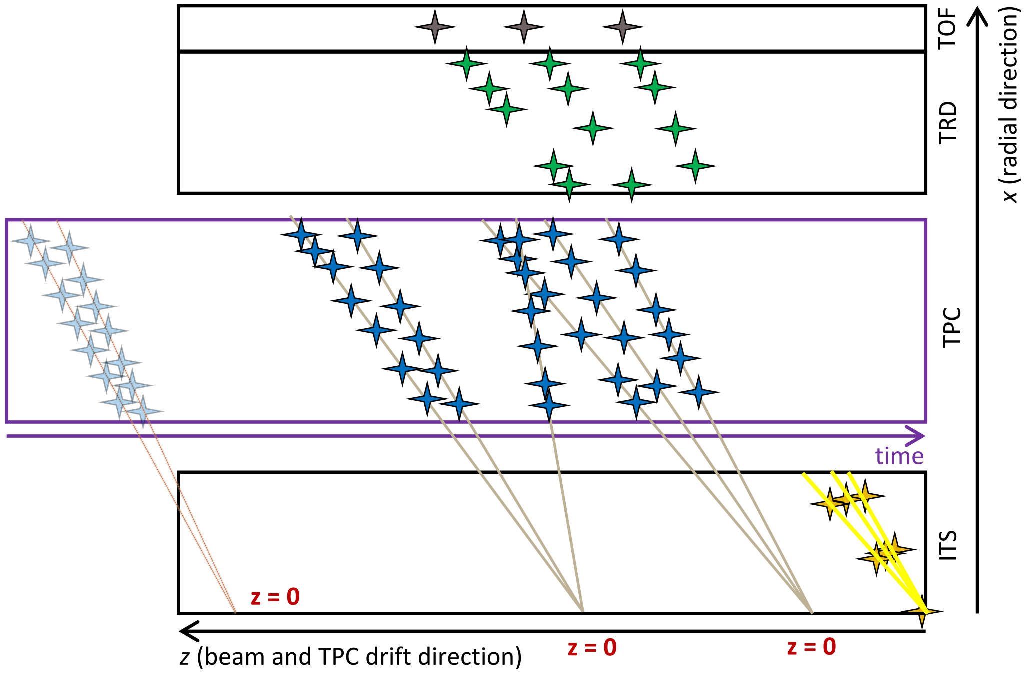


- Standalone ITS tracking
- Standalone TPC tracking, scaling *t* linearly to an arbitrary *z*.
- Extrapolate to *x* = 0, define *z* = 0 as if the track was primary → good enough at this stage (sync!)
- Track following to find missing clusters

Tracking in the central barrel

Challenge in Run 3 + 4!

- **overlap** of multiple collisions (5 collisions in the TPC drift time @50 kHz Pb-Pb)
- with TPC clusters without a well-defined *z* coordinate, but **just a time** (*t*)
- presence of **distortion** corrections that are position dependent

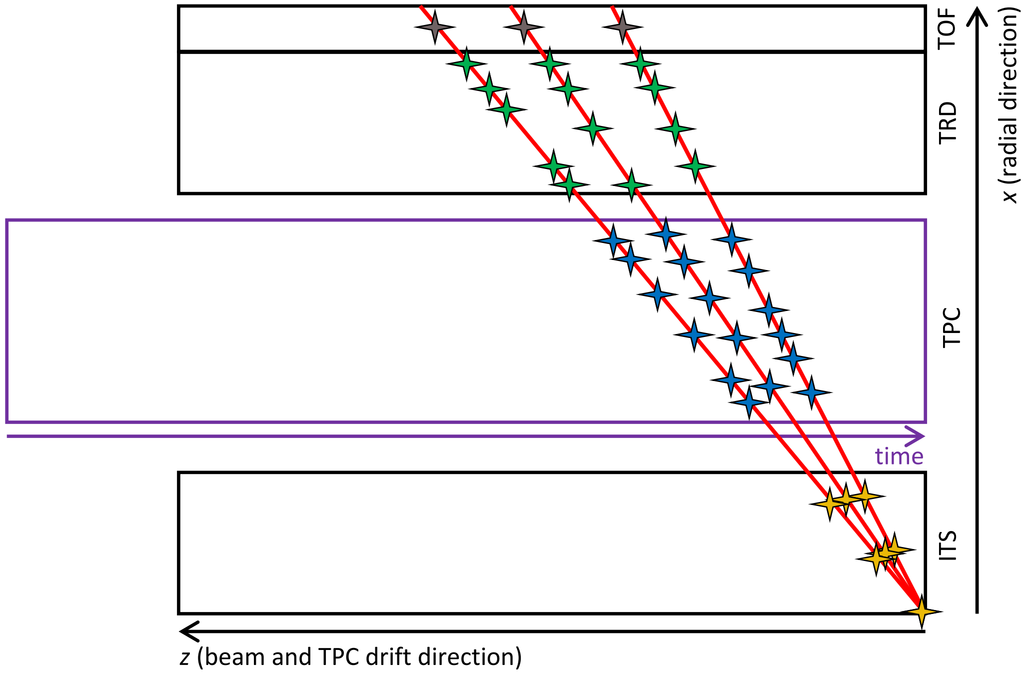


- Standalone ITS tracking
- Standalone TPC tracking, scaling *t* linearly to an arbitrary *z*.
- Extrapolate to *x* = 0, define *z* = 0 as if the track was primary → good enough at this stage (sync!)
- Track following to find missing clusters
- Refine *z* = 0 estimate, refit track with best precision
- Find ITS-TPC track compatibility using times

Tracking in the central barrel

Challenge in Run 3 + 4!

- **overlap** of multiple collisions (5 collisions in the TPC drift time @50 kHz Pb-Pb)
- with TPC clusters without a well-defined *z* coordinate, but **just a time** (*t*)
- presence of **distortion** corrections that are position dependent



- Standalone ITS tracking
- Standalone TPC tracking, scaling *t* linearly to an arbitrary *z*.
- Extrapolate to *x* = 0, define *z* = 0 as if the track was primary → good enough at this stage (sync!)
- Track following to find missing clusters
- Refine *z* = 0 estimate, refit track with best precision
- Find ITS-TPC track compatibility using times
- Match TPC track to ITS track, fixing *z*-position and *t* of the TPC track
- Refit ITS + TPC track outwards and inwards
- Prolong into TRD / TOF

More on TPC Space-Charge Distortions

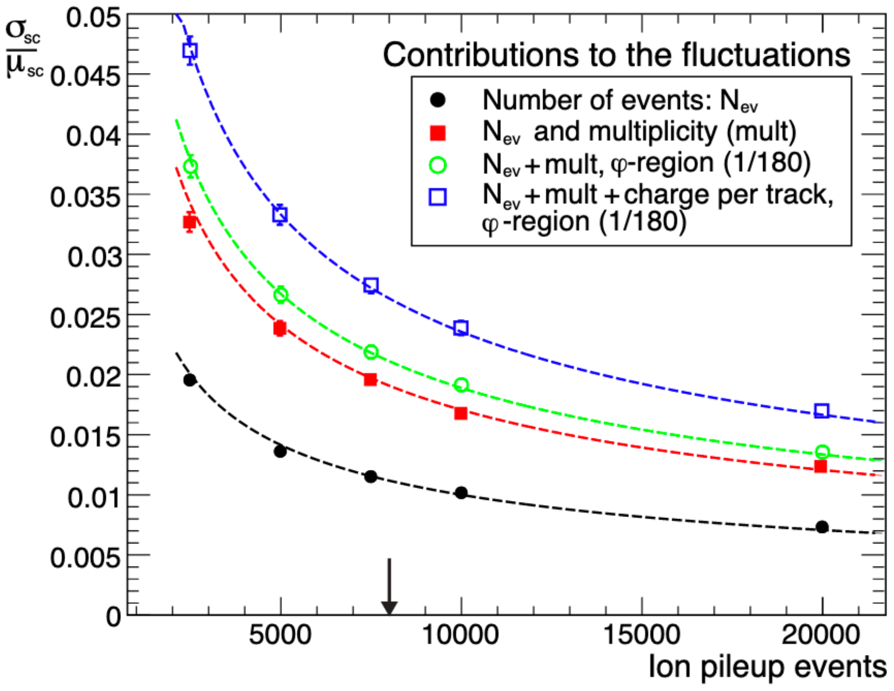
Ion Back Flow

$t_{drift, ion}$ = 160 – 200 ms

IR = 50 kHz

→ Ions from 8000 – 10000 events in the TPC drift volume

→ Space-charge fluctuations ~2 – 3% (5 – 7 mm >> intrinsic resolution, 200 μm)



http://cds.cern.ch/record/1622286/

Synchronous processing

Required resolution: O(mm)

Will use → Average distortion map, scaled to occupancy. Fluctuations corrected in 1D.

Will produce → History of digital currents (charge at the readout plane) integrated over 1 ms.

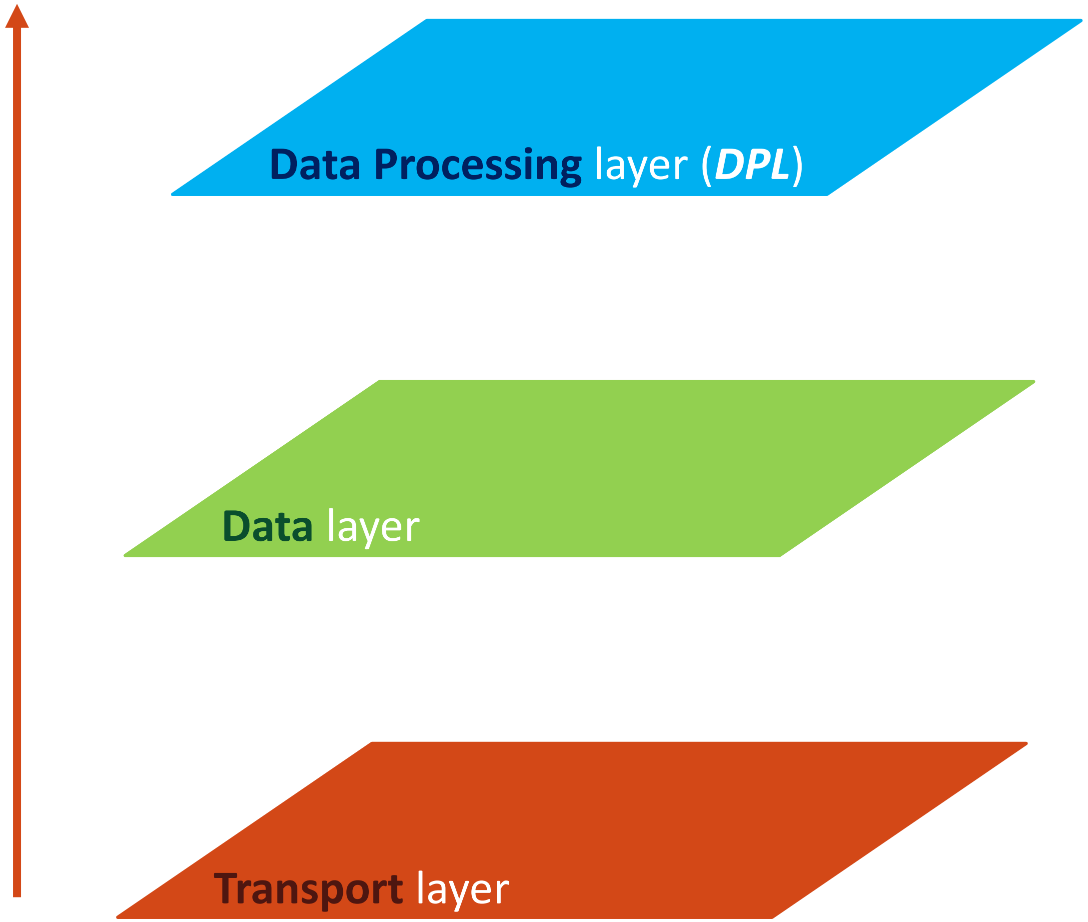
Asynchronous processing

Required resolution: O(100 μm)

Will use → Distortion map from ITS-TPC-TRD-TOF interpolation. Fluctuations corrected in 3D every 5 ms with the digital currents' history from previous 160 ms.

O^2 processing model

ALICE Run 3 + 4 processing software (reconstruction) / framework used at all stages: synchronous phase, asynchronous phase, simulation…



“Translator” of the user’s computational problem in a low-level topology of devices exchanging messages

Declarative.

Reactive-like design (push data, don't pull).

Integration with the rest of the production system.

ALICE-specific description of the messages

Header (extensible) + payload.

Computer language agnostic. Extensible. Suitable for GPU.

Multiple data formats and serialization methods: custom data structure GPU oriented; ROOT; Apache Arrow.

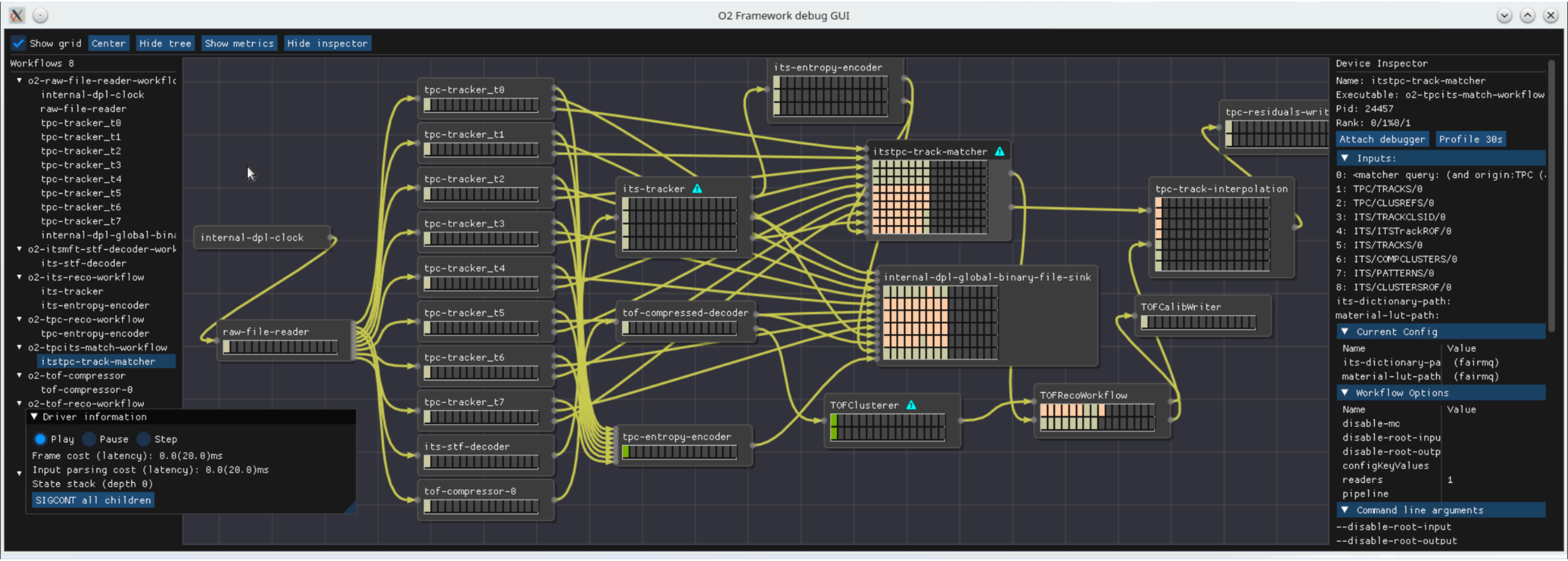
FairMQ message passing architecture

Standalone general processes (devices).

Shared memory backend.

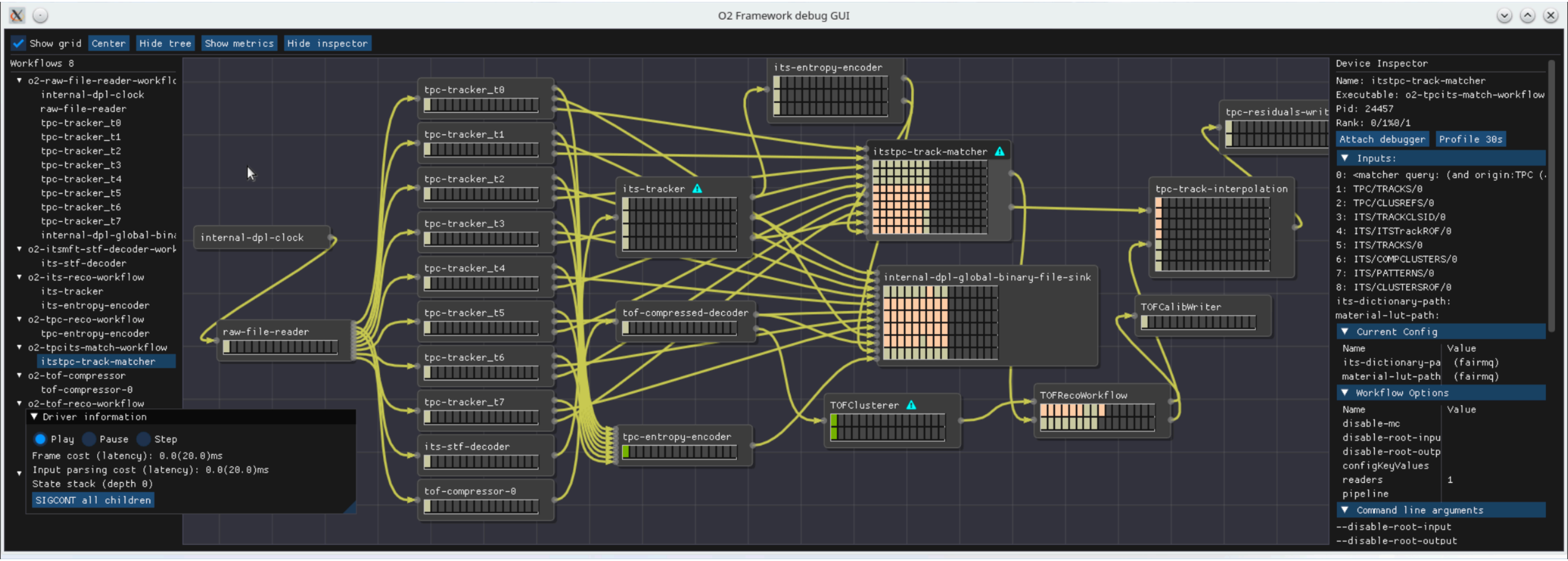
O² processing model

ALICE Run 3 + 4 processing software (reconstruction) / framework used at all stages: synchronous phase, asynchronous phase, simulation…

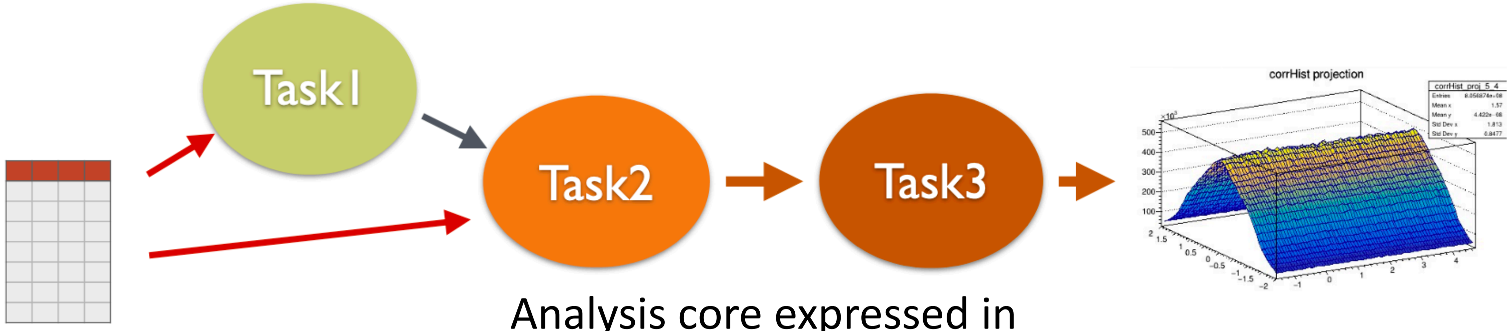


O² processing model

ALICE Run 3 + 4 processing software (reconstruction) / framework used at all stages: synchronous phase, asynchronous phase, simulation… **analysis**



Analysis framework



Analysis core expressed in the form of a **task**

- legacy from Run 1 + 2
- filters and selections
- merging, concatenation of tables

ROOT serialized output

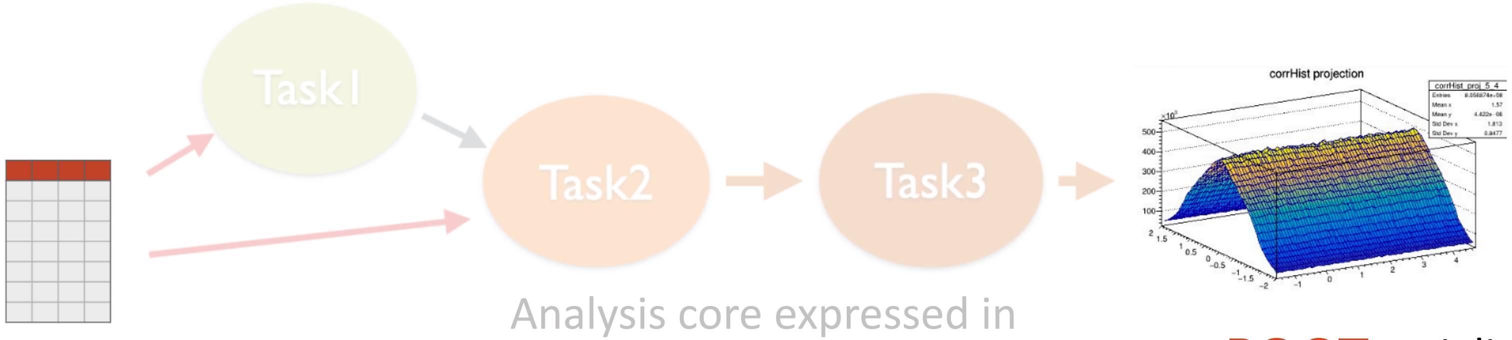
Data model for analysis based on **flat tables** arranged in a relational-database-like manner:

- minimise I/O cost
- improve vectorisation / parallelism

↓

Apache Arrow hidden behind a classic C++ API

Analysis framework



Analysis core expressed in the form of a **task**
- legacy from Run 1 + 2
- filters and selections
- merging, concatenation of tables

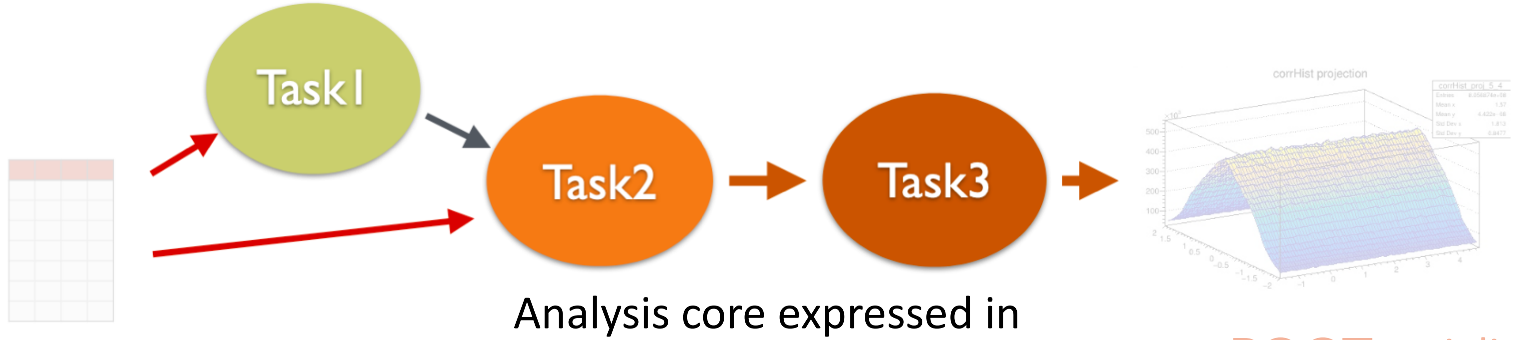
ROOT serialized output

Data model for analysis based on **flat tables** arranged in a relational-database-like manner:
- minimise I/O cost
- improve vectorisation / parallelism

Apache Arrow hidden behind a classic C++ API

O^2 Analysis model is **DECLARATIVE**: the user will specify inputs and outputs

Analysis framework



Analysis core expressed in the form of a **task**

- legacy from Run 1 + 2
- filters and selections
- merging, concatenation of tables

ROOT serialized output

Data model for analysis based on **flat tables** arranged in a relational-database-like manner:

- minimise I/O cost
- improve vectorisation / parallelism

↓

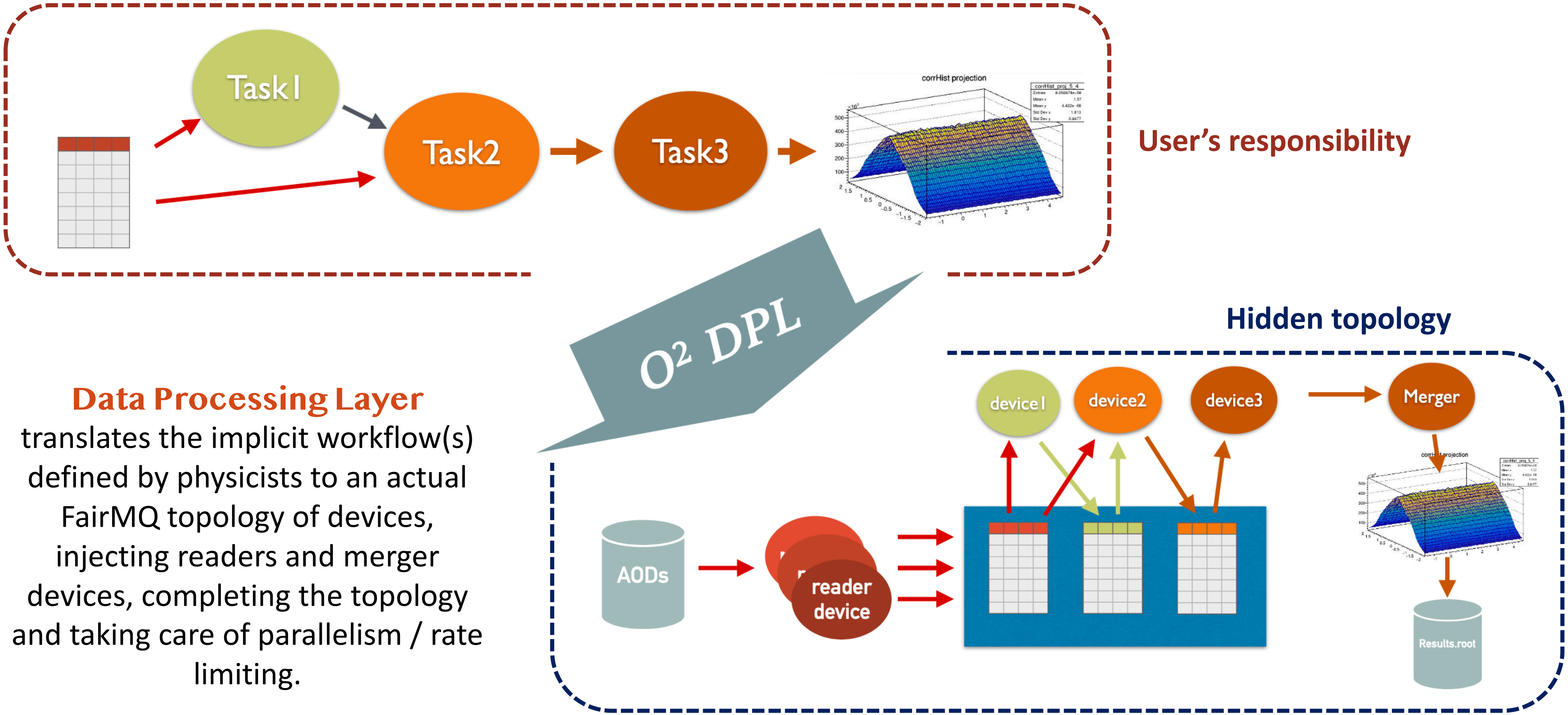
Apache Arrow hidden behind a classic C++ API

O^2 Analysis model is

DECLARATIVE: the user will specify inputs and outputs

IMPERATIVE: the user will specify the processing algorithm

Analysis framework



Data Processing Layer
translates the implicit workflow(s) defined by physicists to an actual FairMQ topology of devices, injecting readers and merger devices, completing the topology and taking care of parallelism / rate limiting.

Analysis and computing model

100x more collisions to analyze with respect to Run 1 + 2, ~x30 increase in AOD total size.

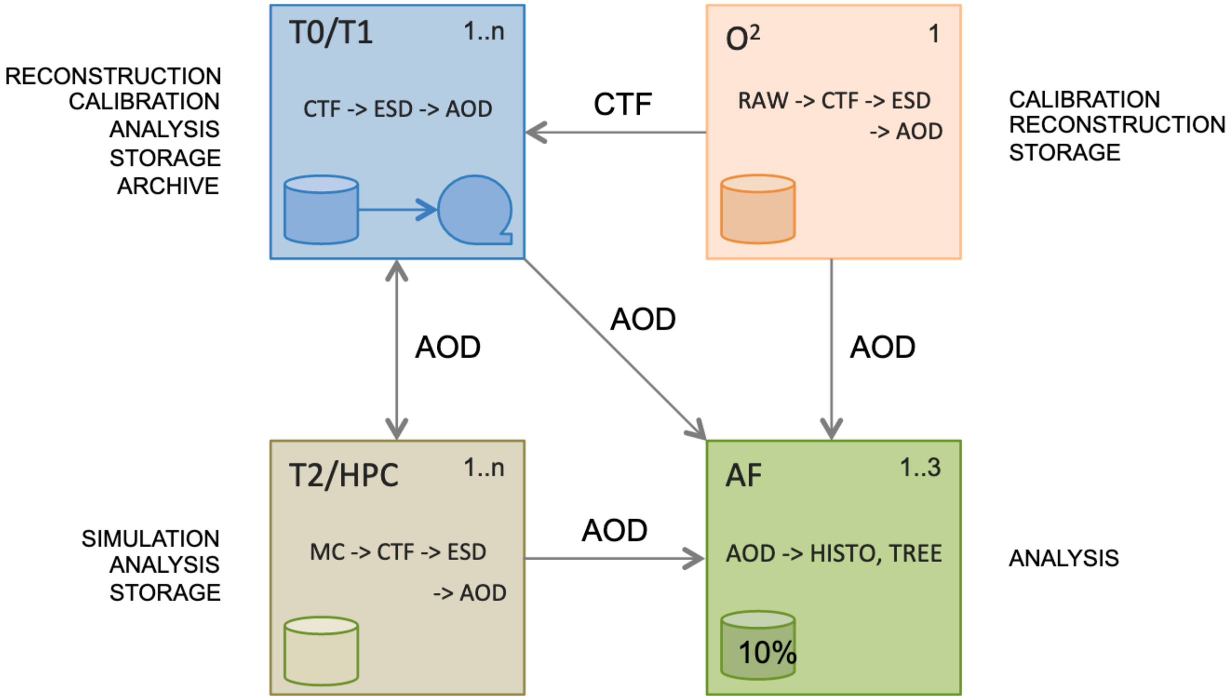
AOD stored only (size) in a **flat** data structure (performance).

Analysis oriented **skimmed** (event and track selection, information reduction) ntuples for further optimization.

10% of the reconstructed and simulated data copied to the O^2 **Analysis Facility** for fast turnaround cycles and analysis validations → **not exclusively distributed analysis model**.

Full samples analyzed on the **Grid**.

Organized analysis to minimize data access.



Summary and conclusions

ALICE is undergoing a **major upgrade** in view of the upcoming Run 3 + 4.

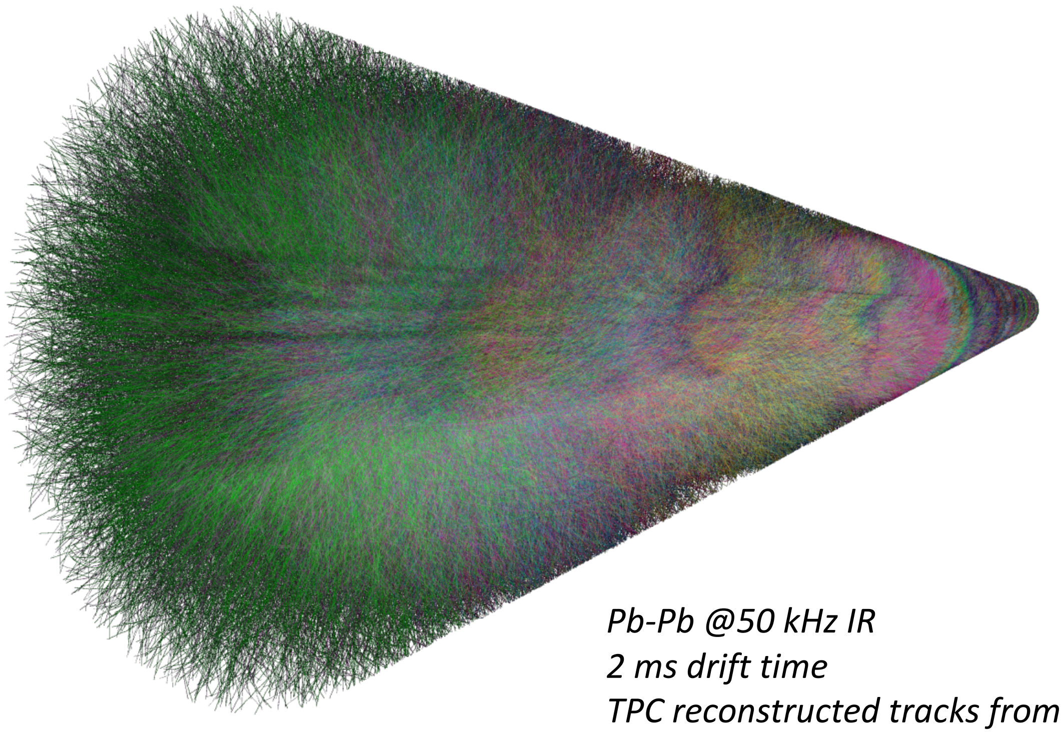
- Detectors, processing, computing (aka O^2), and analysis framework

The **synchronous stage** will allow to achieve a factor of 35 in data reduction

- Relying on full TPC reconstruction, and partial/full reconstruction of the other detectors (also for calibration)

The **O^2 processing framework** will have its foundations on three main layers, for data transport, data model, and data processing

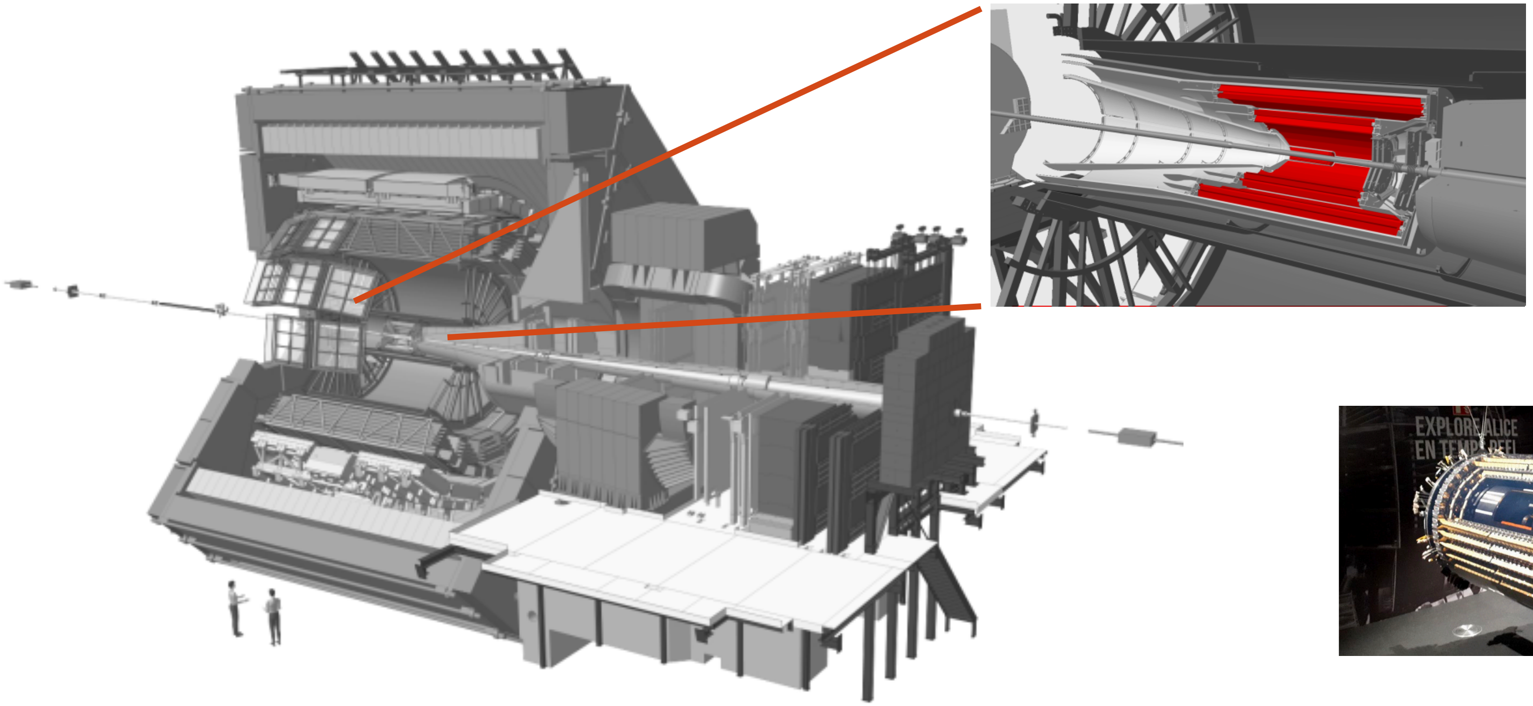
- They will ensure a consistent and homogeneous model for processing over all stages and activities – synchronous/asynchronous reconstruction, simulation, quality control, calibration



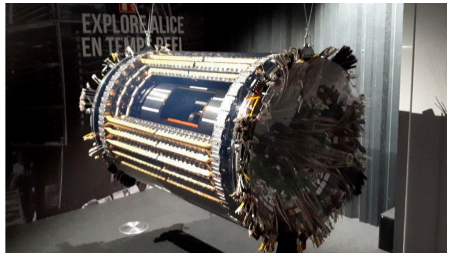
Pb-Pb @50 kHz IR
2 ms drift time
TPC reconstructed tracks from different colour-coded events

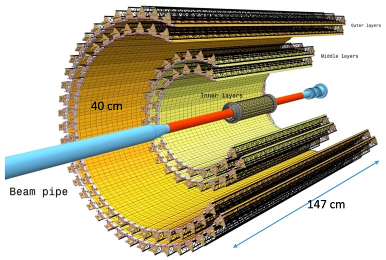
Backup

A Large Ion Collider Experiment



ITS: from Si pixel, drift, strip detectors to MAPS (CMOS Monolithic Active Pixel Sensors)





	ITS	ITS2
Rate	1 kHz	100 kHz
Thickness	~1.14% X_0	0.3% X_0, 0.8% X_0
Pixel size	425 (xy) µm x 50 µm (z)	29 µm (xy) x 27 µm (z)
n. layers	6 (only 2 with pixels)	7

Run 3 + Run 4:

- Pb-Pb IR = **50 kHz** (but also pp!), **continuous**
- Goal: $\mathcal{L}$ **~10 nb^{-1}** (B = 0.5 T) + **3 nb^{-1}** (B = 0.2 T)

Synchronous processing

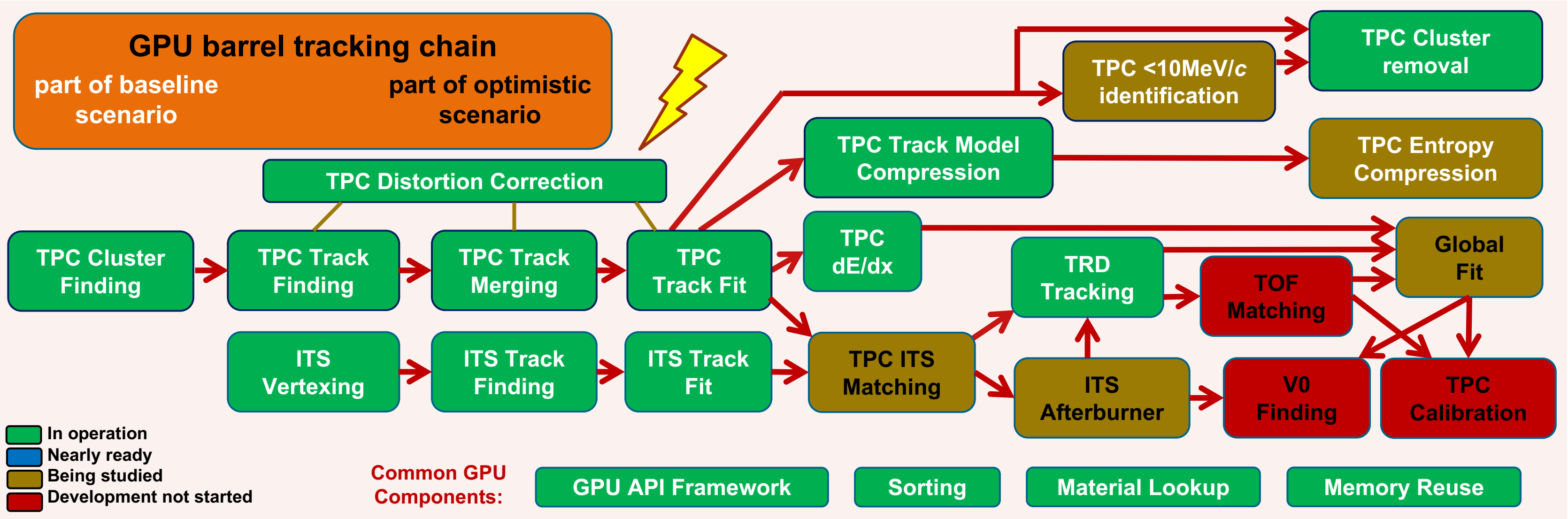
Goal of synchronous reconstruction is to reach factor 35 of compression.

Most relevant detector is TPC: from 3.4 TB/s to 70 GB/s

TPC data compression will consist of:

- **Clusterization**
- Optimized data format
- Entropy reduction
- **TPC tracking**, to remove clusters not associated to tracks

USE OF GPUs MANDATORY



Plan to exploit GPUs computing power also during asynchronous reconstruction