Processing and reconstruction

10 November 2025

A. Calivà, M. Puccio, C. Ristea, R. Shahoyan, C. Zampolli

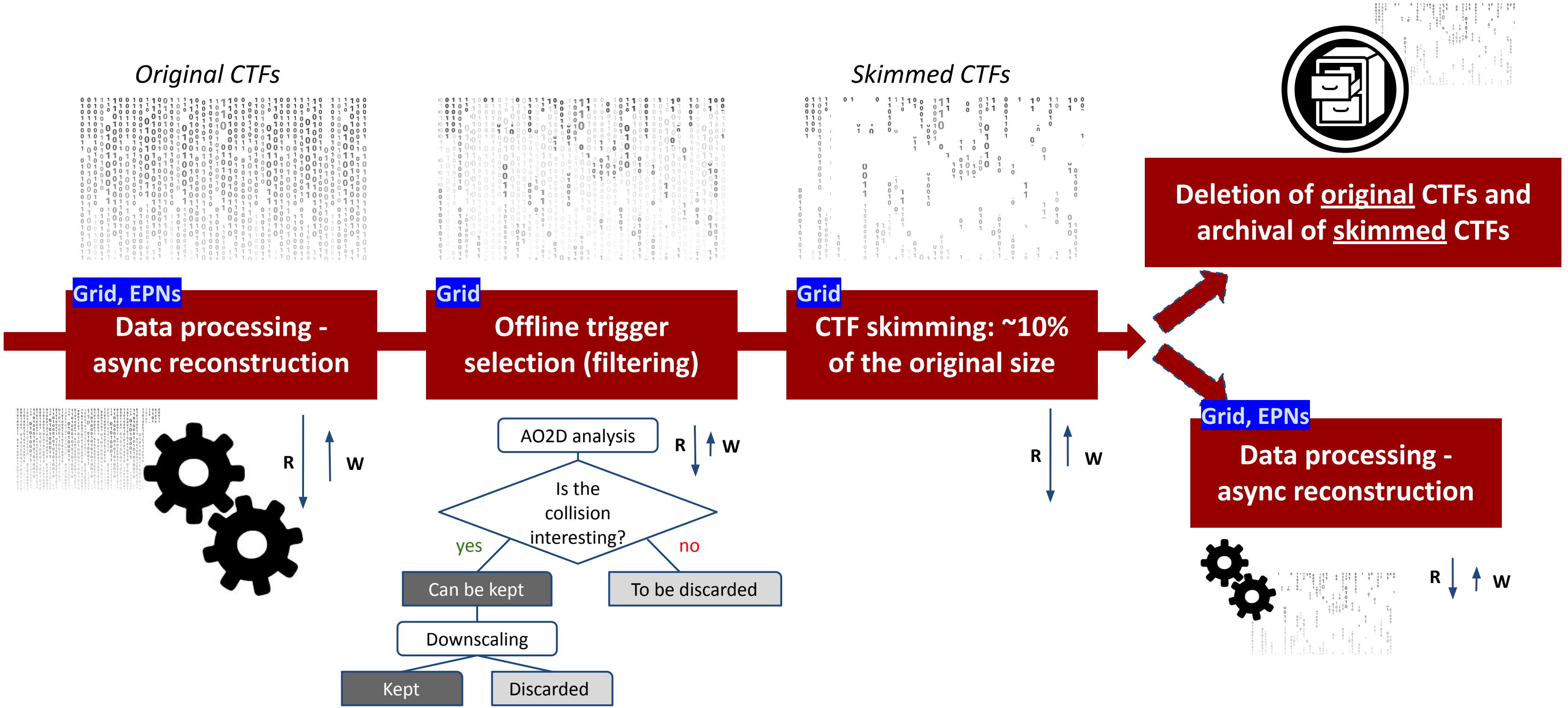
Outline

- Data processing steps: from data taking, to skimming and analysis QC
- An overview of MonALISA
- Track and vertex reconstruction

Run 3: pp data challenge

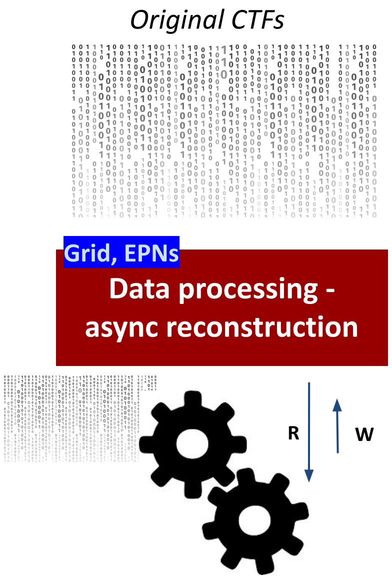
- Very large data sample collected in pp collisions
  - E.g. in 2024, ~9 PB of data taken every week
    - Data are stored in the so-called CTF (Compressed Time Frame) files
    - Organized in 10 minutes long folders in alien, under the "/$year/$period/$run" path
  - Disk buffer at P2 is ~140 PB
  - Considering ~21 weeks of pp data taking → ~190 PB from pp data taking
  - This alone exceeds the O2 buffer space, not including the fact that on the buffer we need to keep some old data
    - detector data
    - PbPb from previous year
    - other data
- Solution: reduce the size of the CTFs by applying an offline trigger selection (OTS); CTFs are then re-written in the so called "skimmed CTFs"

Steps of pp data processing



NB: sizes of symbols/images only for illustration purposes

Steps of pp data processing: C/APass

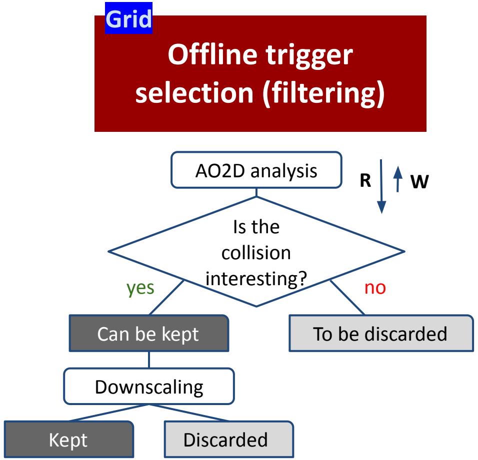


1) Calibration Pass (CPass): first reconstruction of the data, triggered automatically, runs on 10% of the CTFs, selected in a controlled-random way
2) Manual calibration to fix online calibrations, improve others… and QC to collect feedback from data quality
   a) First assessment happens already in Online; async reconstruction allows to validate and possibly reject some data (few percent in 2023 pp)
3) APass1 (Asynchronous Pass): first full reconstruction of the data, triggered by completion of manual calibration and QC
4) After APass1, TPC PID calibration for analysis is extracted
   a) Neural Network approach

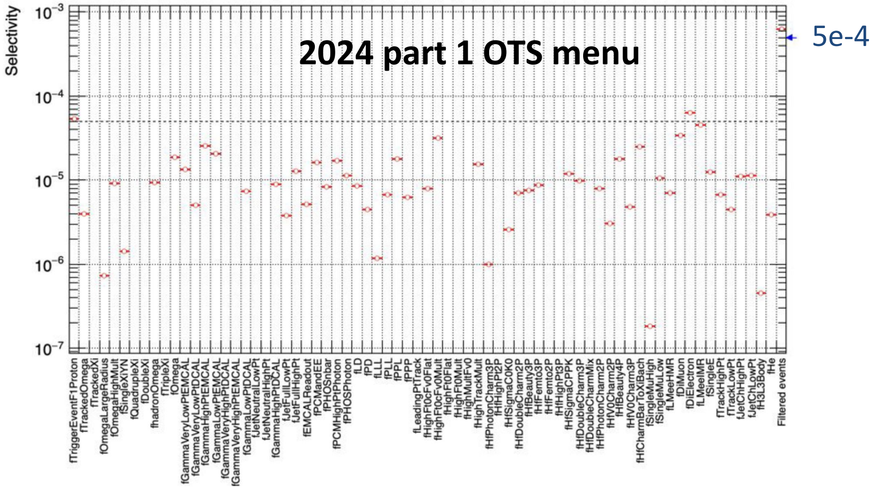
NB: sizes of symbols/images only for illustration purposes

DPG ALICE · Data Preparation Group

Steps of pp data processing: Offline Trigger Selection



1) Offline Trigger Selection (OTS): analysis on AO2Ds from APass1 to identify collisions interesting for analysis
   a) Selected BCs are identified by different analyses
   b) Total selectivity budget is of the order of 5e-4 and it is shared among the PWGs
      i) Goal: final CTF compression of the order of 4%



Steps of pp data processing: CTF skimming

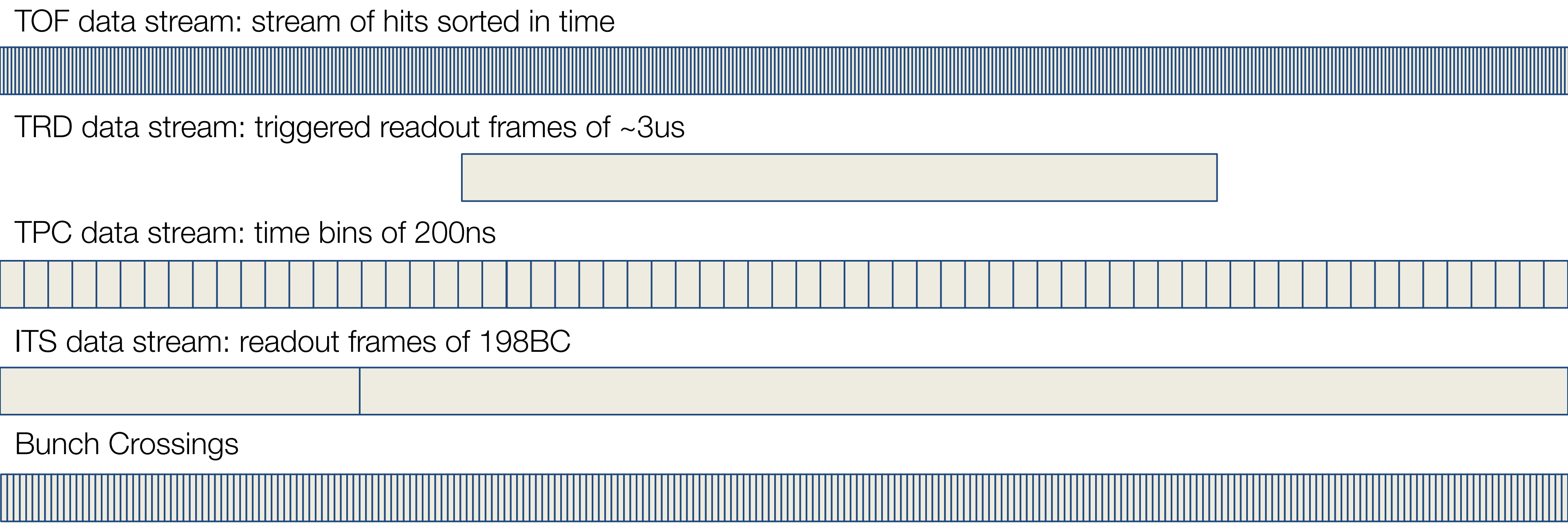
Skimmed CTFs

Grid

CTF skimming: ~4% of the original size

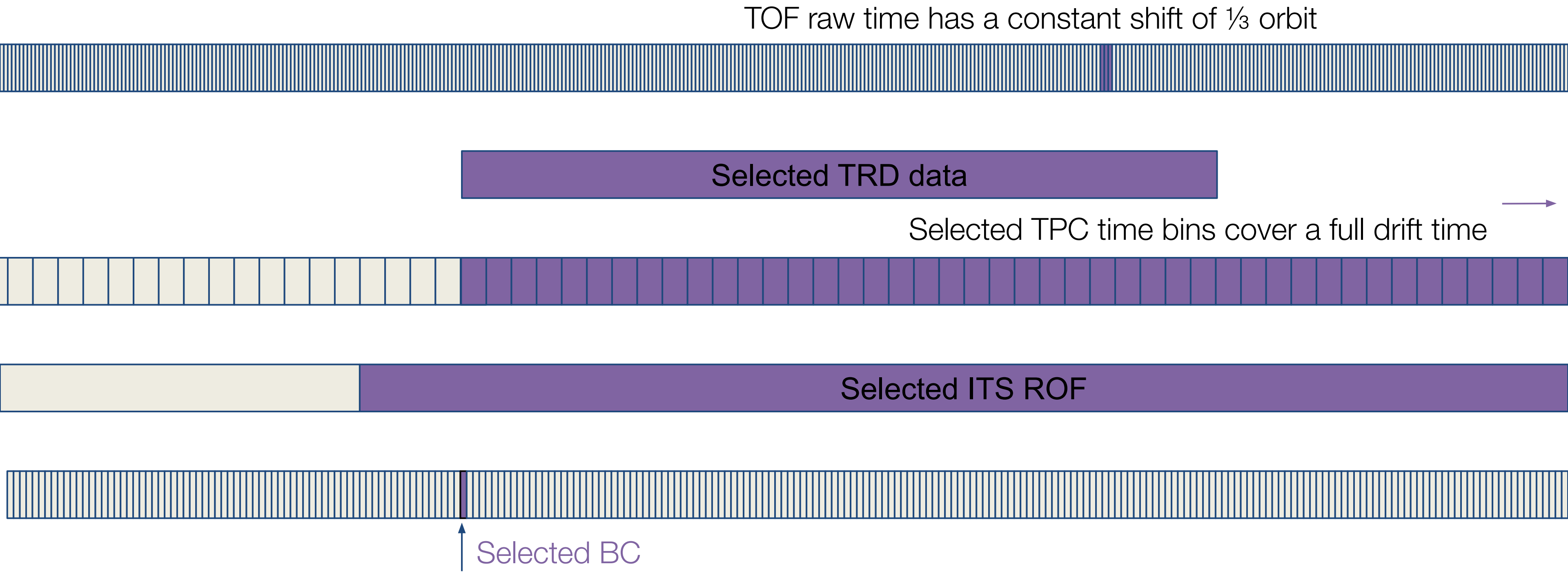
1) CTF skimming: following the decision of the OTS, CTFs are read again, and skimmed CTFs are written out selecting only the interesting BCs, with margin
   a) Margin needed in order to keep the full event information for all detectors (e.g. in TPC we keep 4000BC, a drift time)
   b) Extra margin of 1000BC added for safety
2) Original CTFs are then deleted (for good!)
   a) A 5 per mille MinBias sample is kept
      i) 2022: 1 pb^{-1} was kept (larger sample)
      ii) 2023: 0.5 pb^{-1} was kept
3) Further reconstruction passes can happen only on the skimmed CTFs (and the MinBias sample)
4) Since skimmed CTFs are only a few percent of the size of the original ones, further processing is fast
5) apass1_skimmed follows
   a) e.g. 24al:
      i) apass1 output: 5 PB
      ii) apass1_skimmed: 230 TB → <5%

CTF skimming: selecting from the datastream



Pictorial view (take it with a pinch of salt) of the data stream of 4 barrel detectors representing the situations being dealt with in the CTF skimming process

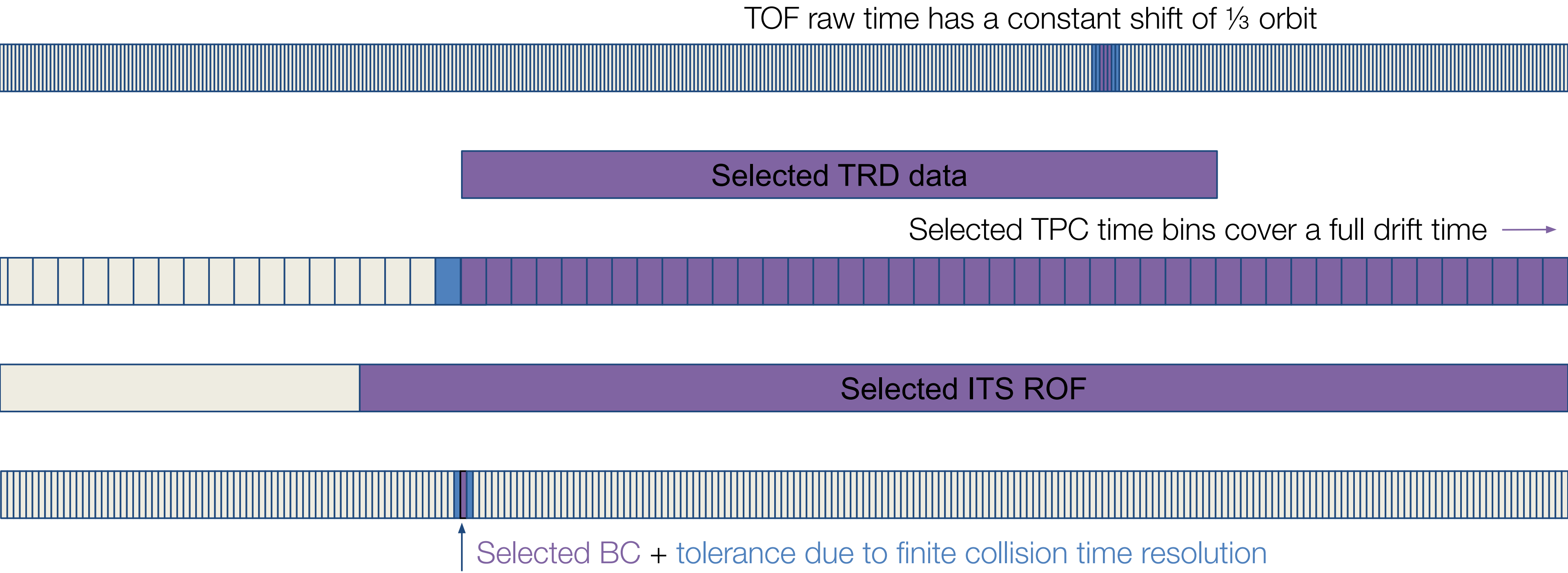
CTF skimming: selecting from the datastream



The CTF reader has been modified to read, for each detectors, only the data related to the selected BC (+margin)

- These data will contain pile-up for most detectors due to their ROF being larger than the average distance between two collisions

CTF skimming: selecting from the datastream



When selecting an event, we do not define only a selected BC, but a window of selected BCs for two reasons:

- The collision time resolution is finite, and we allow for a 4 sigma window around the measured collision time
- The found collision is associated at analysis level to the closest T0 signal, that might be few tens of BCs away

Steps of pp data processing: thinning

- Special procedure put in place to reduce the output of an apass1 while the skimming was not finalized
- Removed TPC only tracks, which are anyway not used for V0s (apart from gammas!), kinks, cascades
- Reduces by a factor 40-50% the output of apass1
  - ~3.3 PB from thinned apass4 of 2022 and 2023 (apass4 was ~6.3 PB)
  - in 2023, some periods had a larger reduction factor (~68%) due to a special table for tracking and performance studies (*trackQC* table) kept for all tracks
- Done for 2022 apass4 and 2023 apass4
- Used selectively on most of the periods in PbPb apass4 production
  - Two periods kept non-thinned for AOT tracks studies needing TPC only tracks
  - Under discussion to move the studies to use the sample of TPC only tracks kept anyway (downscaled) in the *trackQC* table

DPG ALICE · Data Preparation Group

Steps of pp data processing: straining

- New step in the processing of pp data:
- Goal: extract luminosity information from apass1 (not skimmed!)
  - e.g. from 24am:
    - apass1 output = 1.8 PB
    - apass1_strained = 8.66 TB → <5 per mille
- After apass1_skimmed and straining are done, apass1 can be deleted

DPG ALICE • Data Preparation Group

Analysis QC

- After reconstruction, some basic analysis are run centrally in a child job
- They are not considered as a critical step: if something fails, they're not re-run
- They use the same sw as the reconstruction
- Very useful for QC
- Every PWG can add its own analysis and report at the aQC meeting (Tue 3 pm)

DPG ALICE · Data Preparation Group

https://alimonitor.cern.ch/production/raw.jsp

Data period, JIRA-ID

RAW Production Cycles

Processing requests »

							Raw data		Reconstructed						Timing		Software versions	»
24am, !relval											ESDs			Output				
Production	**Description**		**Col.**	**Status**	**Run Range**	**Runs**	**Chunks**	**Size**	**Chunks**	**%**	**Size**	**%**	**Events**	**Size**	**Running**	**Saving**		**Err**
LHC24am_apass1_debug	LHC24am - apass1 of LHC24am debug - 13.6 TeV, CPU, O2-5308		pp	Running	555967 - 555967	36	2,387,592	21.22 PB	0	0%	0 B	-	0	28.24 GB	8d 22:51	1d 21:13	O2PDPSuite::async-async-v1-01-12a-slc9-alidist-async-v1-01-02-1	
LHC24am_apass1_skimmed	LHC24am - apass1_skimmed of LHC24am - 13.6 TeV, CPU, O2-5025		pp	Running	555344 - 555976	1,132	93,234,904	830.1 PB	0	0%	0 B	-	0	74.81 TB	74y 69d	12y 320d	O2PDPSuite::async-async-v1-01-12a-slc9-alidist-async-v1-01-02-1	
LHC24am_skimming	LHC24am - CTF skimming and re-writing of LHC24am - 13.6 TeV, O2-5111		pp	Running	555344 - 555976	1,132	93,234,904	830.1 PB	0	0%	0 B	-	0	846.1 TB	83y 206d	48y 117d	O2PDPSuite::async-async-v1-01-11-slc9-alidist-async-v1-01-01-1	
LHC24am_apass1	LHC24am - apass1 of LHC24am - 13.6 TeV, CPU, O2-5308		pp	Running	555344 - 555976	1,132	93,234,904	830.1 PB	0	0%	0 B	-	0	1.575 PB	1493y 336d	66y 362d	O2PDPSuite::async-async-v1-01-12a-slc9-alidist-async-v1-01-02-1	
LHC24am_cpass0	LHC24am - cpass0 of LHC24am - 13.6 TeV, EPN, O2-4786		pp	Running	555344 - 555976	1,131	93,231,081	830 PB	0	0%	0 B	-	0	162.1 TB	71y 335d	1y 292d	O2PDPSuite::async-async-v1-01-12-slc8-alidist-async-v1-01-01-1	
5 productions							**375,323,385**	**3.263 XB**	**0**		**0 B**		**0**	**2.633 PB**	**1723y 226d**	**130y**		

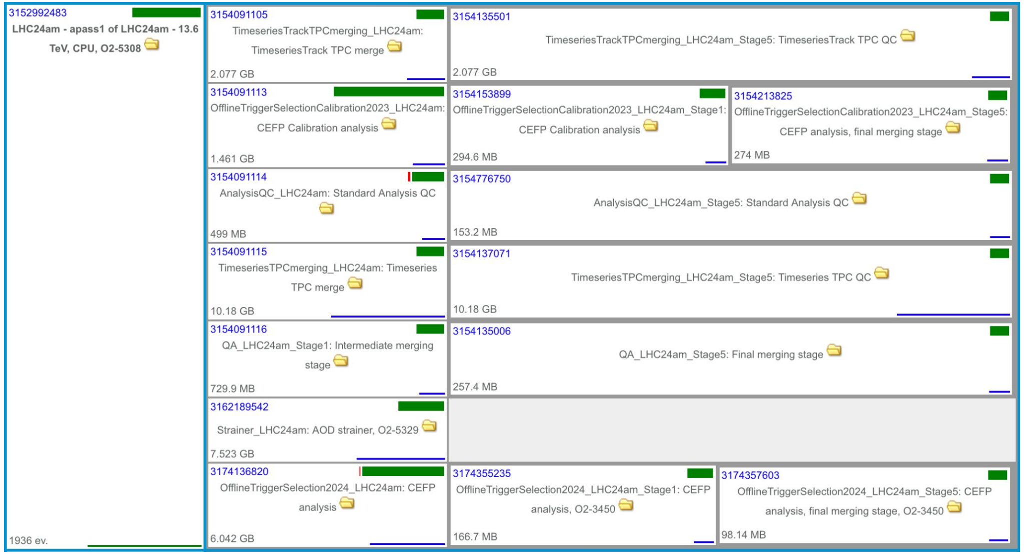
Main reconstructions

- Transient: cpass0 (10% of data), apass1
- Permanent: apass1_skimmed (NB ~6% x apass1), skimming(→TAPE), debug (1 CTF file every 10’, w/ full output stored)

DPG ALICE • Data Preparation Group

Production chains in MonALISA

apass1



https://alimonitor.cern.ch/prod/

Production type: - All -												
Production info							Jobs status					
	24am											
ID	Tag	Status	Done%	Cfg	Out	Links	Total	Done	Active	Waiting	Runs	In ev
30850	Strainer_LHC24am	Completed	99%				49082	49076	0	0	53 (555344 - 555976)	
30788	OfflineTriggerSelection2024_LHC24am	Running	99%				235047	234587	0	1	53 (555344 - 555976)	
30905	OfflineTriggerSelection2024_LHC24am_Stage1	Completed	100%				4981	4981			53 (555344 - 555976)	
30902	OfflineTriggerSelection2024_LHC24am_Stage5	Completed	100%				2264	2264			53 (555344 - 555976)	
30783	QA_LHC24am_Stage1	Running	99%				7790	7782	5	0	53 (555344 - 555976)	
30678	QA_LHC24am_Stage5	Completed	99%				3488	3485	0	0	53 (555344 - 555976)	
30782	OfflineTriggerSelectionCalibration2023_LHC24am	Running	99%				446284	442520	778	0	53 (555344 - 555976)	
30784	OfflineTriggerSelectionCalibration2023_LHC24am_Stage1	Running	99%				4878	4870	2	0	53 (555344 - 555976)	
30785	OfflineTriggerSelectionCalibration2023_LHC24am_Stage5	Running	99%				1219	1218	0	0	53 (555344 - 555976)	
30680	TimeseriesTPCmerging_LHC24am	Completed	99%				9924	9895	0	0	53 (555344 - 555976)	
30682	TimeseriesTPCmerging_LHC24am_Stage5	Completed	99%				3458	3443	0	0	53 (555344 - 555976)	
30679	AnalysisQC_LHC24am	Running	93%				25093	23544	314	0	53 (555344 - 555976)	
30683	AnalysisQC_LHC24am_Stage5	Completed	99%				3116	3114			54 (-1 - 555976)	
30677	TimeseriesTrackTPCmerging_LHC24am	Completed	99%				9924	9893	0	0	53 (555344 - 555976)	
30681	TimeseriesTrackTPCmerging_LHC24am_Stage5	Completed	99%				3456	3445	0	0	53 (555344 - 555976)	
7 productions			99%				810004	804117	1099	1		

Output merging

run 560123

reco 758 / 759

grid apass3

eff 99.9%

CPU time 19y 44d

disk size 16.59 TB

/0730	/0730/Stage_1	/0730/Stage_2	/0730/QC
123 / 123	13 / 13	2 / 2	1 / 1
100%	100%	100%	100%
2y 115d	31m 51s	17m 14s	44s
2.108 TB	958.3 MB	184.6 MB	98.75 MB

/0720	/0720/Stage_1	/0720/Stage_2	/0720/QC
208 / 209	21 / 21	3 / 3	1 / 1
99.50%	100%	100%	100%
5y 273d	1:21	5m 22s	**18m 43s**
5.182 TB	1.661 GB	292.7 MB	**235.6 MB**

/0710	/0710/Stage_1	/0710/Stage_2	/0710/QC
213 / 213	22 / 22	3 / 3	1 / 1
100%	100%	100%	100%
6y 75d	1:34	3m 21s	41s
5.33 TB	1.737 GB	292.6 MB	109.4 MB

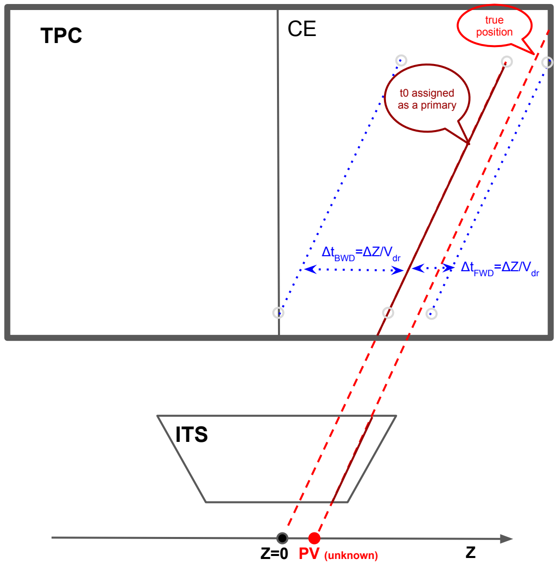
/0700	/0700/Stage_1	/0700/Stage_2	/0700/QC
214 / 214	22 / 22	3 / 3	1 / 1
100%	100%	100%	100%
4y 310d	55m 29s	12m 44s	1m 12s
3.972 TB	1.543 GB	283.6 MB	106 MB

Output single file/run: QC_fullrun.root, AnalysisResults_MergedAnalysis_fullrun.root, TPC_timeseries/track.root, OTS selection

Details of track and vertex timestamps assignment and their uncertainties in Run 3 reconstruction

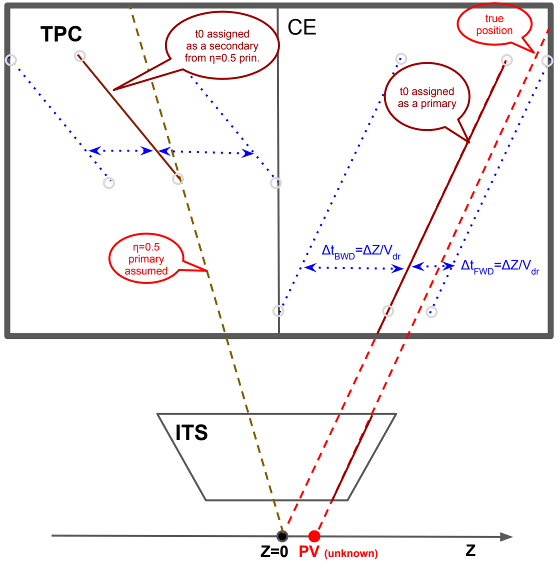
ITS, TPC tracks time assignment

- Independent standalone tracking in ITS and TPC
  - ITS tracks have precise positions. Time uncertainty corresponding to ITS readout frame (ROF) length (~5μs in pp, ~15μs in PbPb) is assigned to timestamp defined as a middle of the ROF.
  - TPC tracks have uncertain Z due to the unknown interaction time (t0)
    - We have to assign some plausible t0 to evaluate cluster corrections and errors during track fit (Z-dependent!)
      - if the track can be propagated to the beamline →assume primary and assign t0 such that the track converges to Z=0 at the beamline



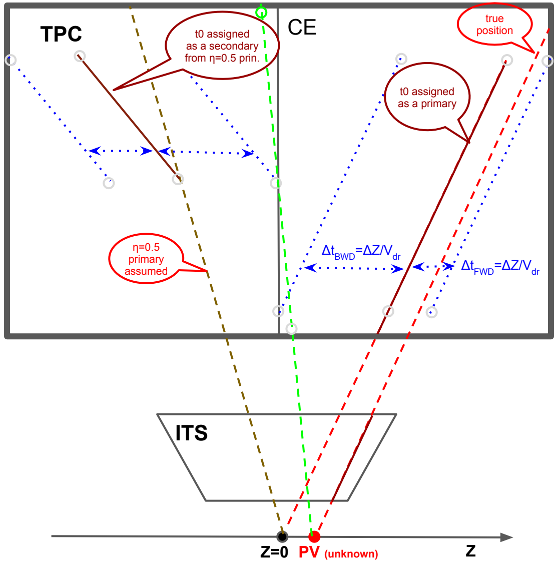
ITS, TPC tracks time assignment

- Independent standalone tracking in ITS and TPC
  - ITS tracks have precise positions. Time uncertainty corresponding to ITS readout frame (ROF) length (~5μs in pp, ~15μs in PbPb) is assigned to timestamp defined as a middle of the ROF.
  - TPC tracks have uncertain Z due to the unknown interaction time (t0)
    - We have to assign some plausible t0 to evaluate cluster corrections and errors during track fit (Z-dependent!)
      - if the track can be propagated to the beamline →assume primary and assign t0 such that the track converges to Z=0 at the beamline
      - otherwise, assume track comes from decay of |η|=0.5 primary
      - define time bracket by a decrement Δt_{BWD} and an increment Δt_{FWD} to assigned t0 from the condition that under any Z or t assignment edge clusters remain within TPC volume.



ITS, TPC tracks time assignment

- Independent standalone tracking in ITS and TPC
  - ITS tracks have precise positions. Time uncertainty corresponding to ITS readout frame (ROF) length (~5μs in pp, ~15μs in PbPb) is assigned to timestamp defined as a middle of the ROF.
  - TPC tracks have uncertain Z due to the unknown interaction time (t0)
    - We have to assign some plausible t0 to evaluate cluster corrections and errors during track fit (Z-dependent!)
      - if the track can be propagated to the beamline →assume primary and assign t0 such that the track converges to Z=0 at the beamline
      - otherwise, assume track comes from decay of |η|=0.5 primary
      - define time bracket by a decrement Δt_{BWD} and an increment Δt_{FWD} to assigned t0 from the condition that under any Z or t assignment edge clusters remain within TPC volume.
      - Special treatment for TPC tracks having clusters in both TPC sides: they are well constrained in Z, a time error of ~1 μs is assigned.



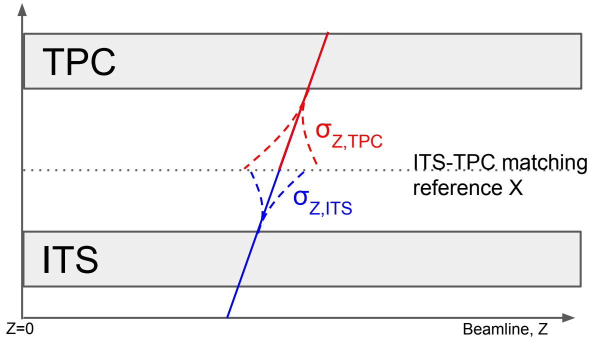
ITS-TPC matching, time assignment

- Only TPC tracks contributed by clusters at or below certain pad-row (20 for apass4/Pb23, 78 for apass5/Pb23 and apass3/Pb24) are considered for matching. A pool of ITS and TPC tracks propagated to X_{Ref}= 60 cm is created.
- Only pairs of ITS and TPC tracks in overlapping time brackets defined by ITS ROF and TPC $\{t0-\Delta t_{BWD} : t0+\Delta t_{FWD}\}$ (augmented by a safety margin of ~20cm/VDfrift) respectively are considered.
- For each pair the TPC track at X_{Ref} is moved by distance ΔZ to match the Z of ITS track at the same X (except for CE-crossing tracks which are fixed in Z).
  This fixes the ITS-TPC track candidate time to $t0+\Delta Z/Vdrift$: if goes outside of the the TPC time bracket, the pair is rejected, otherwise the χ^2 between two track is calculated and stored in the match candidates for ITS and TPC tracks.
- Multiple matching partners are allowed for each ITS and TPC track but at most 1 unique winner match for each ITS or TPC track is selected in a few iterations as pairs of ITS and TPC tracks having mutually best χ^2.
- ITS-TPC track time uncertainty assignment is derived from the uncertainty of ΔZ, which is defined by the intrinsic uncertainties of TPC and ITS track extrapolation to the matching reference X and systematic error:

$$\sigma_t = \sqrt{\sigma^2_{Z,ITS} + \sigma^2_{Z,TPC}} + \epsilon_{t,syst}$$

  with $\varepsilon_{t,syst}$ = 0.2μs.

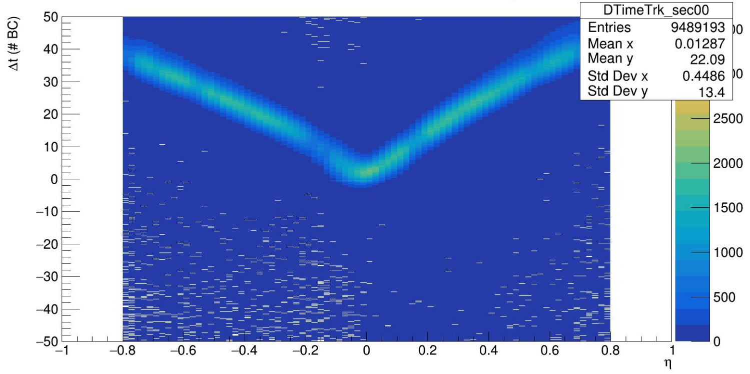
  Note that the errors $\sigma_{Z,TPC}$ are defined not only by the extrapolation of the TPC track fit intrinsic errors but also by the "systematic" errors matrices added on the TPC boundaries (~0.25-0.5 cm depending on the interaction rate).

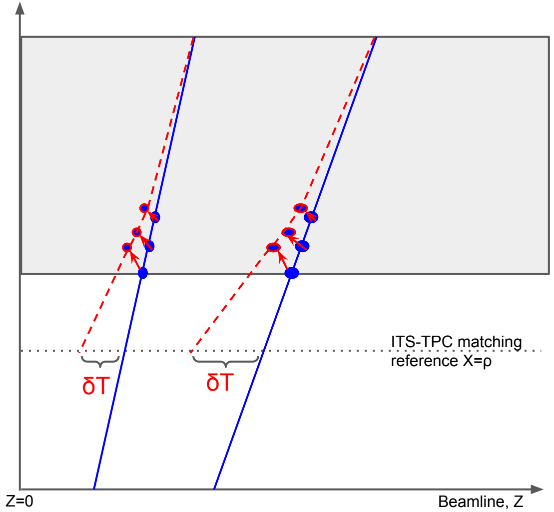


ITS-TPC matching, time assignment

- Note that the assigned time error is sensitive not only to the correctness of the nominal TPC Vdrift: $\delta_t = T^{true}_{drift} \frac{V_{true}-V_{assumed}}{V_{true}}$ (which introduces its dependence on the track η) but also to the residual radial distortions which make the η-error (hence the track extrapolation error in Z) to depend on track η.

Example of the η-dependent ITS-TPC track time bias (difference to TOF matched time) due to the Vdrift and residual radial distortions miscalibration



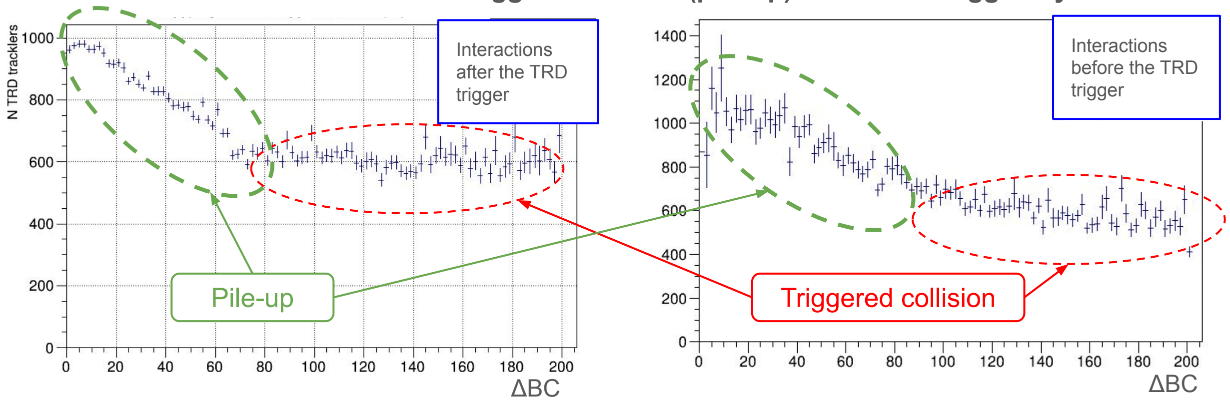


- **Afterburner**: remaining TPC tracks are matched to unused ITS tracks in the outer 4 layers (at least 2 outermost layers are required, holes are not allowed).
  - For that every TPC track is cloned and constrained in time (and Z) by every FT0 interaction time (compatible with track's $\{t0-\Delta t_{BWD} : t0+\Delta t_{FWD}\}$), extrapolated to the outermost layer, then matches hypotheses tree is build for every clone.
  - Again, at most 1 winner per TPC track is allowed.
  - Since the AB matched track time is provided by the FT0, a nominal 1 ns time error is assigned.

Matches to TRD and TOF and their timing

- ITS-TPC tracks have well defined Z position: directly propagated to TRD radius.
- Every TPC-only track is cloned and time-(Z-)constrained by every TRD trigger compatible with its $\{t0-\Delta t_{BWD} : t0+\Delta t_{FWD}\}$
- Similar to ITS-TPC afterburner, a hypotheses tree is build for every candidate and at most one longest / best winner per ITS-TPC or TPC-only track is selected; the time of the TRD trigger is assigned as the matched track timestamp.
- Being a drift detector, TRD sees tracklets not only from the triggered collision but also from the pile-up in ± ~2μs

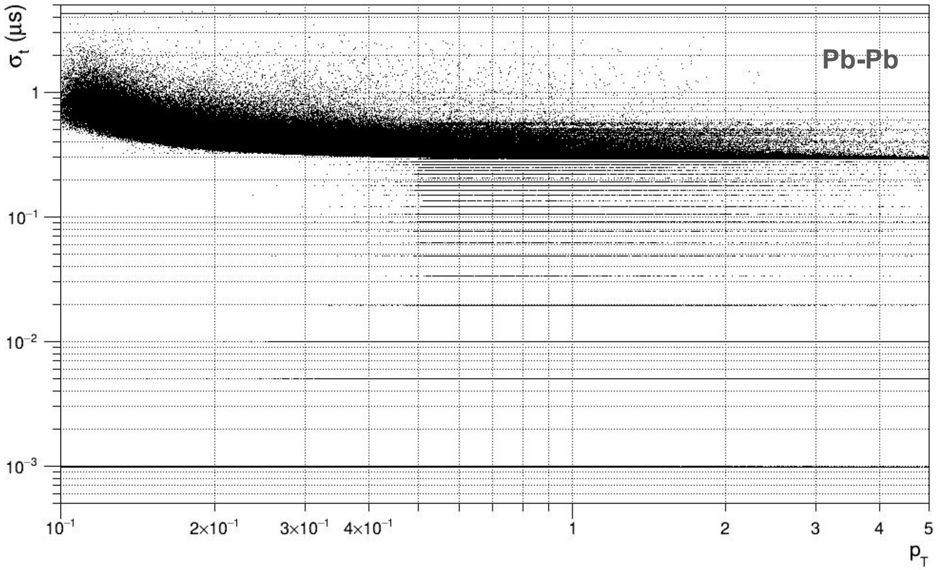
Distance in BC from TRD trigger to closest (pile-up) interaction tagged by FT0

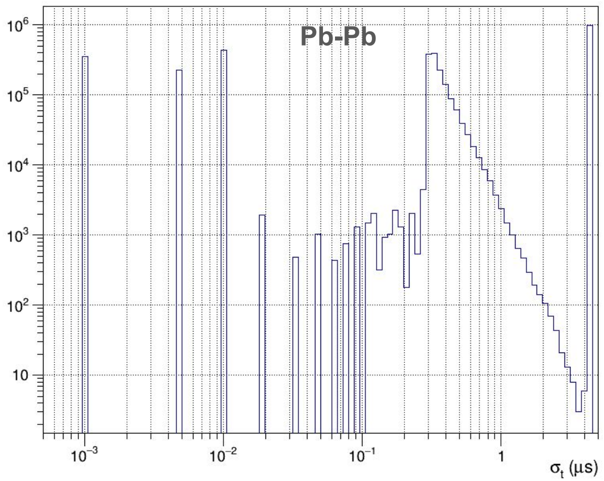


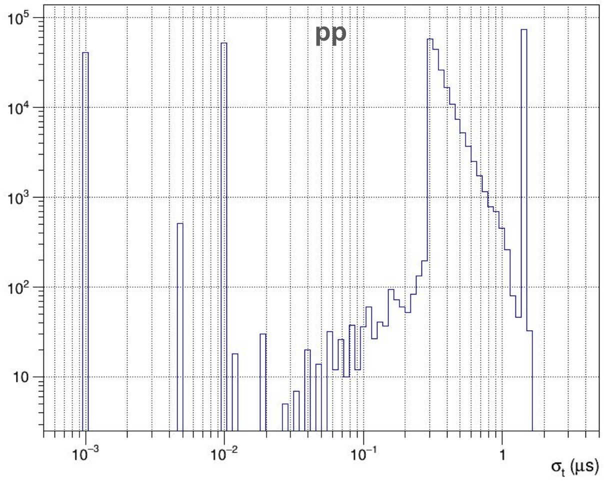
⇒ to account for this, the furthest pule-up in ± 80 BCs from the TRD trigger is sought using FT0:
If found, its distance to the trigger, scaled by $1/\sqrt{12}$ is assigned as a time error, otherwise a nominal 5ns error is used.

- Exactly the same procedure is repeated for TOF layer, except that the TOF cluster time (corrected for the expected time-of-flight to the TOF radius) is used to constrain TPC-only tracks and assigned as a final timestamp.
Nominal error of 10ns is assigned to account for eventual error in the time-of-flight estimate.

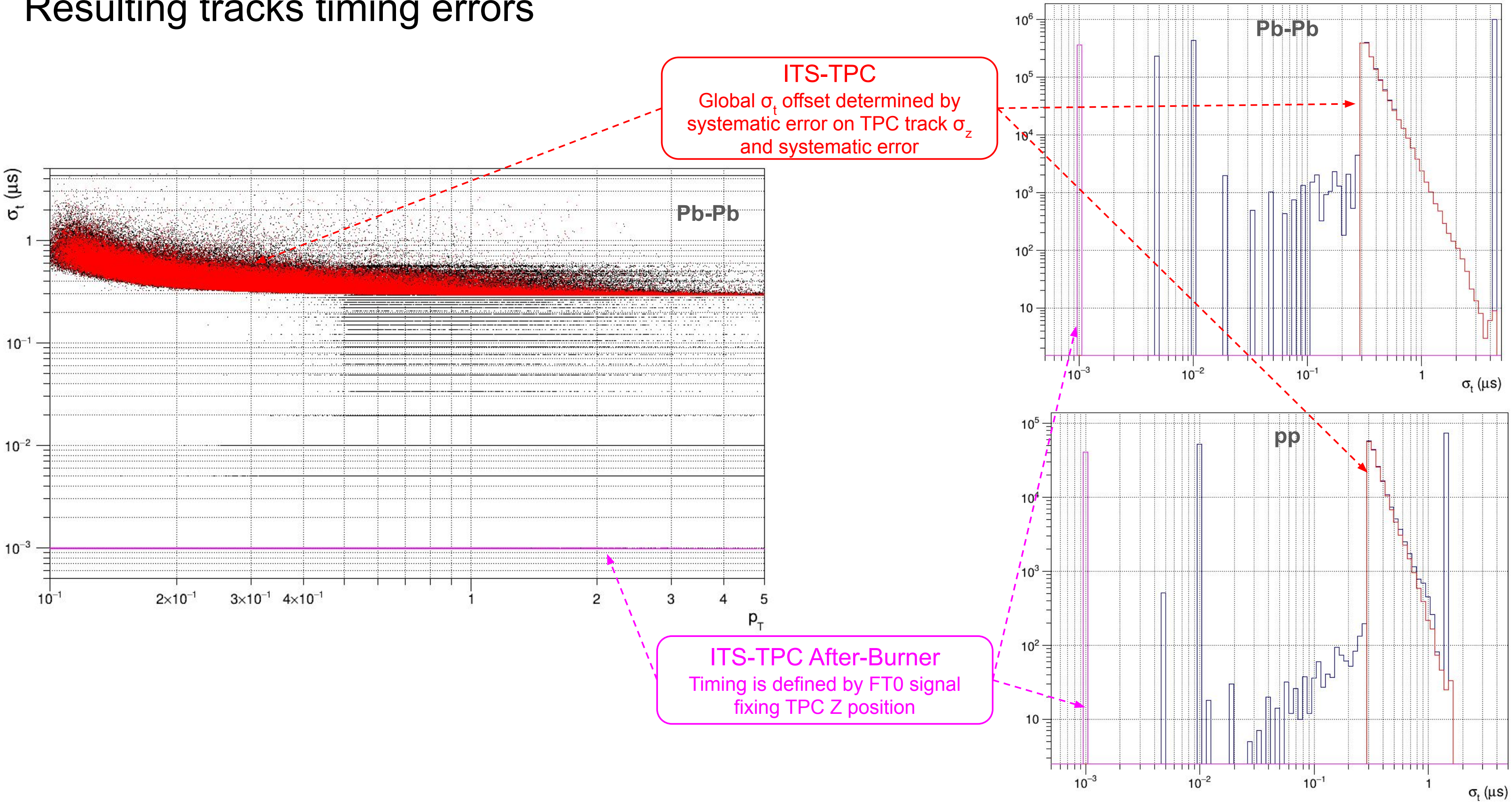
Resulting tracks timing errors



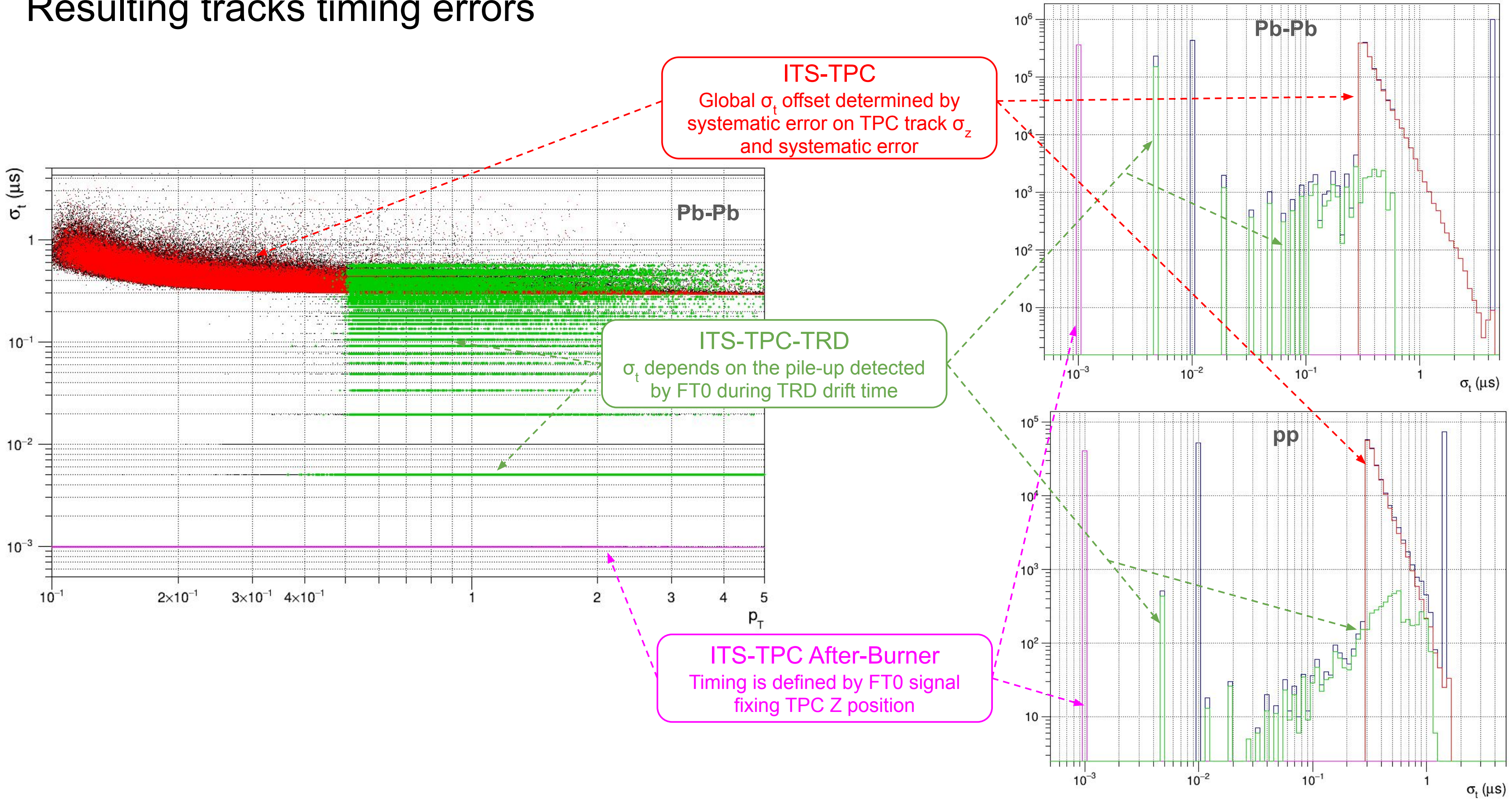




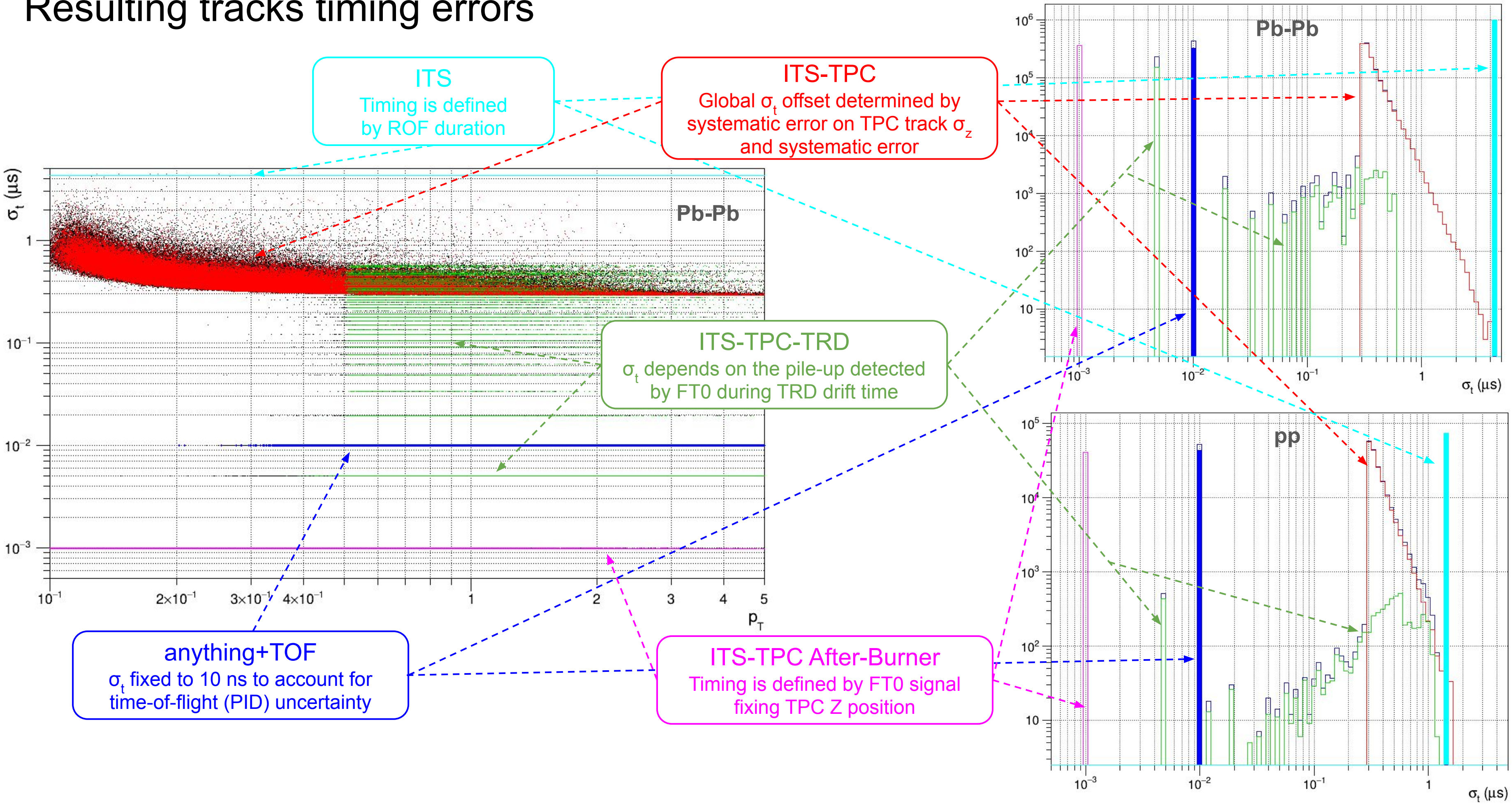
Resulting tracks timing errors



Resulting tracks timing errors



Resulting tracks timing errors



Primary vertexing and its timing, track-to-vertex association

- Only tracks with contribution from ITS IB and approaching to certain fiducial distance to the beam-line (MeanVertex) are allowed to contribute to the primary vertex. The tracks errors are fixed at the extrapolation to the beam line
- Seeding starts with custom version of DBSCAN, operating in time and Z (beam) dimensions.
- Every found cluster is subjected to a few iterations of histogram-based peak finding in time and z, with subsequent vertex fit with iterative outliers rejection. The final point of contributors point of closest approach both in space and in time is done with least square: the PV time is effectively a weighted sum of the track times, except the that for the ITS-only tracks only 1 contribution (with $\sigma_t = \Delta T_{ROF}/\sqrt{12}$) is accounted, since all ITS tracks have identical timestamp of the middle of the ROF (essentially, ITS provides single time measurement).
- Tracks used in vertex fit are flagged as “contributors” (-> track.isPVContributor() for AO2D tracks). Such a track is allowed to contribute to only one PV.
- Every non-contributing tracks is associated with every PVs provided they are compatible in time (4-σ for Gaussian errors and uniform distribution for ITS, MFT, and TPC tracks). PV 4-σ bracket are augmented by 1.3 μs and track brackets are augmented by 0.5 μs safety margins.
- Non-contributor track can be associated with any number of PVs, but in the AO2D the reference only the 1st compatible PV (can be the least probable!) is stored: analyses should take care of testing all possible track-vertex associations.