ALICE MC in Run3/4

S. Wenzel (CERN) for the ALICE collaboration

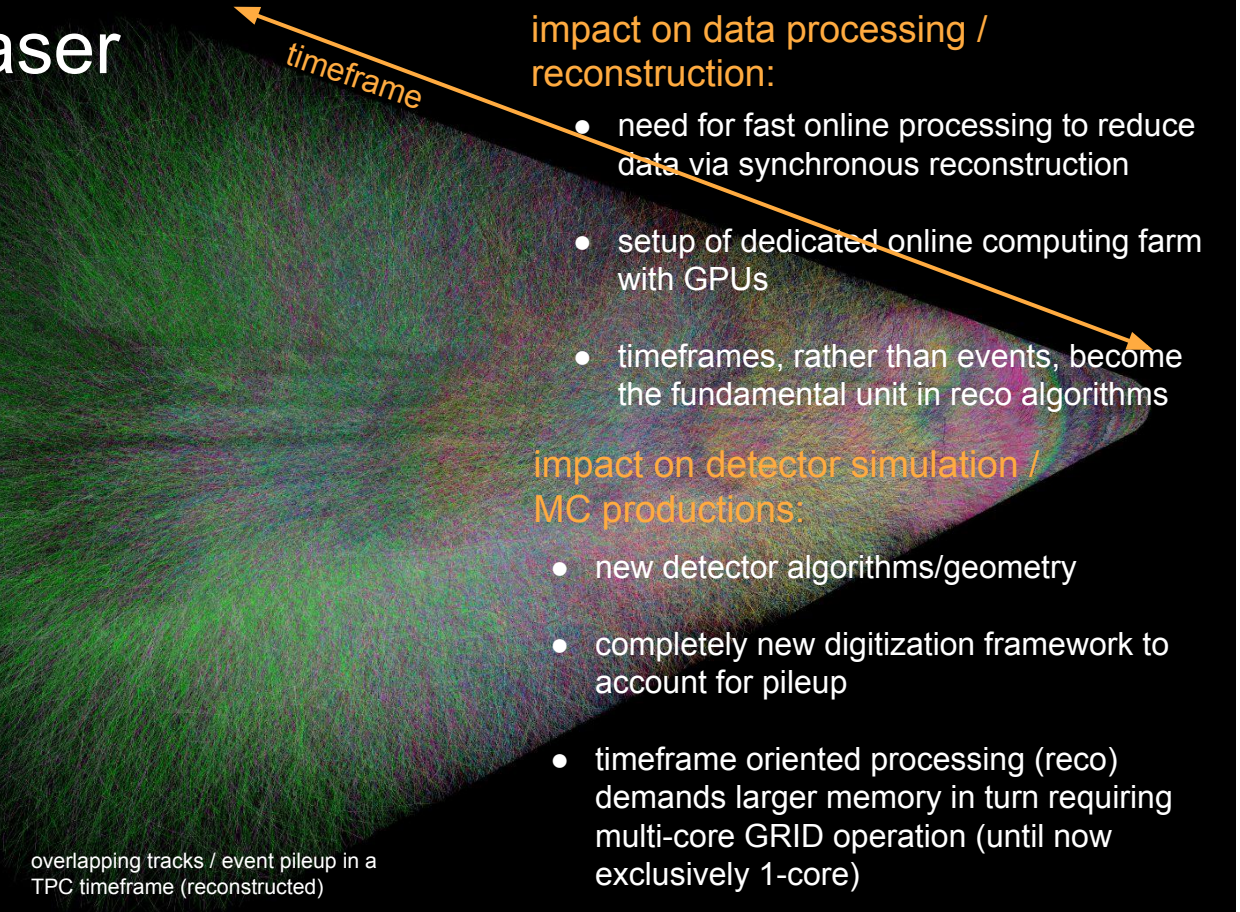
(input from A. Morsch / S. Piano / R. Preghenella)

Disclaimer

- Upgrades for ALICE already imminent so this is less of “looking into the future”
- Picked out a few subjects / highlights within our upgrade transformation, in the hope that it might be interesting for others
  - follow a holistic approach on the overall simulation ecosystem, not only on the transport engines
- Raise a few points that might be beneficial to us in the Geant4 ecosystem

ALICE Run 3 teaser

- PbPb collisions at max 50kHz (100x higher data rate through novel detector technology)
- substantial inter bunch crossing pileup (like here in the TPC)
- (untriggered) continuous readout
- data comes in unfiltered readout buffers, called timeframe, accumulated over ~11ms (~550 min bias PbPb events)



overlapping tracks / event pileup in a TPC timeframe (reconstructed)

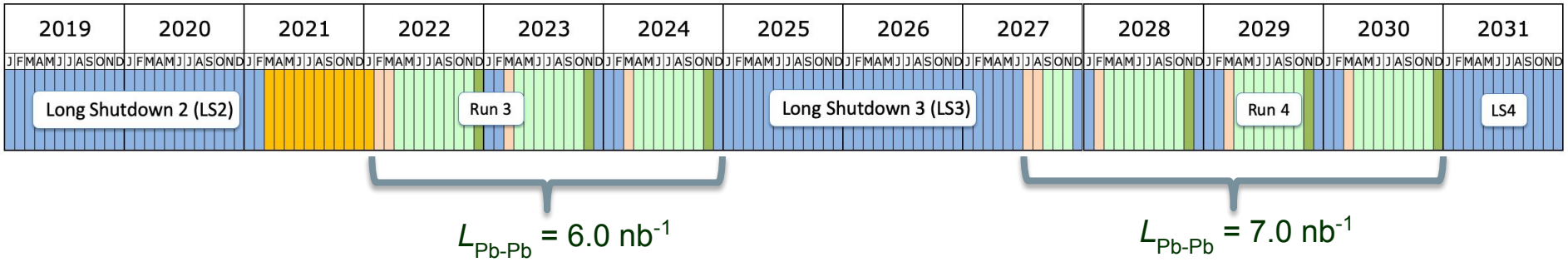
impact on data processing / reconstruction:

- need for fast online processing to reduce data via synchronous reconstruction
- setup of dedicated online computing farm with GPUs
- timeframes, rather than events, become the fundamental unit in reco algorithms

impact on detector simulation / MC productions:

- new detector algorithms/geometry
- completely new digitization framework to account for pileup
- timeframe oriented processing (reco) demands larger memory in turn requiring multi-core GRID operation (until now exclusively 1-core)

ALICE physics goals for Run 3 and 4



- Heavy-flavour mesons and baryons (down to very low p_T)
  - ⇒ mechanism of quark-medium interaction
- Charmonium states
  - ⇒ dissociation/regeneration as tool to study de-confinement and medium temperature
- Dileptons from QGP radiation and low-mass vector mesons
  - ⇒ χ-symmetry restoration, initial temperature and EOS
- High-precision measurement of light and hyper (anti-)nuclei
  - ⇒ production mechanism and degree of collectivity

- Focus on PbPb physics !
- (smaller pp processing than before)
- ~100x more min bias + ~10x increase in triggered events (interesting events) compared to Run1/2

CERN Yellow Report arXiv:1812.06772

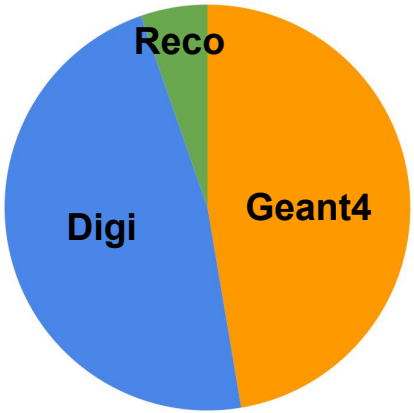
Full MC needs for Run 3 (similar for Run4)

- Run3-TDR: At least 500 million MC PbPb events to serve physics analysis
  - e.g., driven by requirement to obtain reco efficiencies in whole pT range
- In Run2, 1 min-bias PbPb MC collision costed 24k HS06s on average
  - ~40min on average GRID CPU
- Folding in computing resources, estimated ability to simulate ~1 10^9 events in Run3.

- **Ok for operation**
  - … but small space for errors or unforeseen needs
- **Challenging for some analysis**
  - through drastic reduction of ratio MC / collected data to ~ 1-2% on average, in comparison to Run1 (40%)
- **However not “hopeless”**
  - Often systematics rather than MC-statistics effected
  - Plan to increase quality of MC-sample for specific analysis by strengthening event selection/biasing or signal enhancement approaches
  - Run3 software refactoring not yet account

A word on relative CPU share in Run2

- Basic observation driving developments

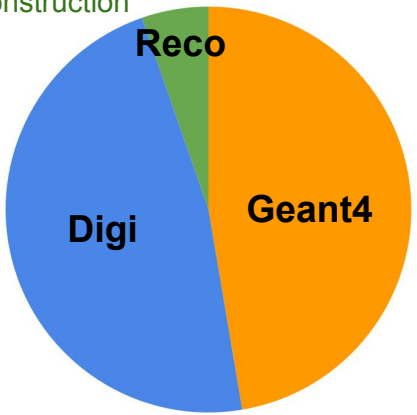


Run2 distribution (schematic)
1 event = 24000 HS06s
MC budget only

- transport **and** digitization expensive in Run2 MC
  - pp: digi >= transport
  - PbPb: digi ~ 1/2 transport
- Unusual corner of phase-space compared to other LHC experiments
- In ALICE, substantial need to optimize beyond G4
  - improvements in G4 will be limited unless rest is improved too

Overview of MC upgrade developments

- Try to bridge the gap … target high quality MC samples for analysis beyond bare minimum
- Apply holistic multi-level strategy for MC (incorporate within the anyway necessary transformation of Run2->Run3 software stack)



New (GPU) reconstruction

New pileup digitization framework with optimized + parallelised algorithms

Sub-event parallel simulation framework

Simulator-as-service

Tune engine and detectors settings (cuts, geometry)

Optimize physics response code

Fast sim in certain detectors (ZDC)

→ On an integrated / global level

Improved event biasing / selection capabilities

Signal enhancement strategies (embedding, like LHCb ReDecay) needing less background simulation

Task parallelism and multicore readiness on a global level

- efficient overall GRID algorithm scheduler

Decrease error rate through CI and performance monitoring efforts

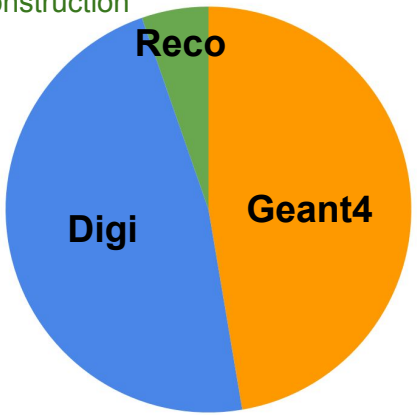
Improved MC production planning and setup

= more details later

Overview of MC upgrade developments

- Try to bridge the gap … target high quality MC samples for analysis beyond bare minimum
- Apply holistic multi-level strategy for MC (incorporate within the anyway necessary transformation of Run2->Run3 software stack)

New (GPU) reconstruction



New pileup digitization framework with optimized + parallelised algorithms

Sub-event parallel simulation framework

Simulator-as-service

Tune engine and detectors settings (cuts, geometry)

Optimize physics response code

Fast sim in certain detectors (ZDC)

On an integrated / global level

Improved event biasing / selection capabilities

Signal enhancement strategies (embedding, like LHCb ReDecay) needing less background simulation

Task parallelism and multicore readiness on a global level
- efficient overall GRID algorithm scheduler

Decrease error rate through CI and performance monitoring efforts

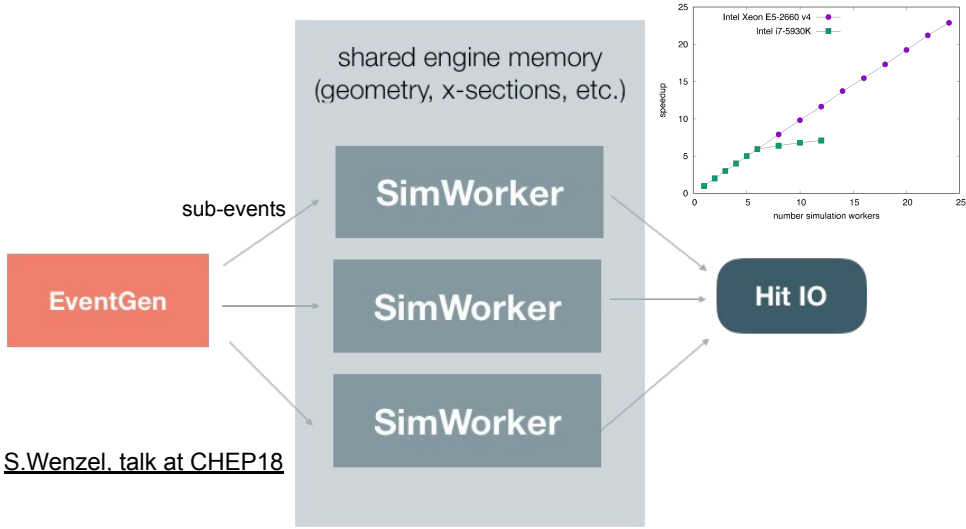
Improved MC production planning and setup

New Run3/4 transport framework

- ALICE remains in Virtual Monte Carlo (VMC) ecosystem and uses TGeo/ROOT as geometry system
  - Geant4 as main backend (optional and seamless switch to Fluka, Geant3, ...)

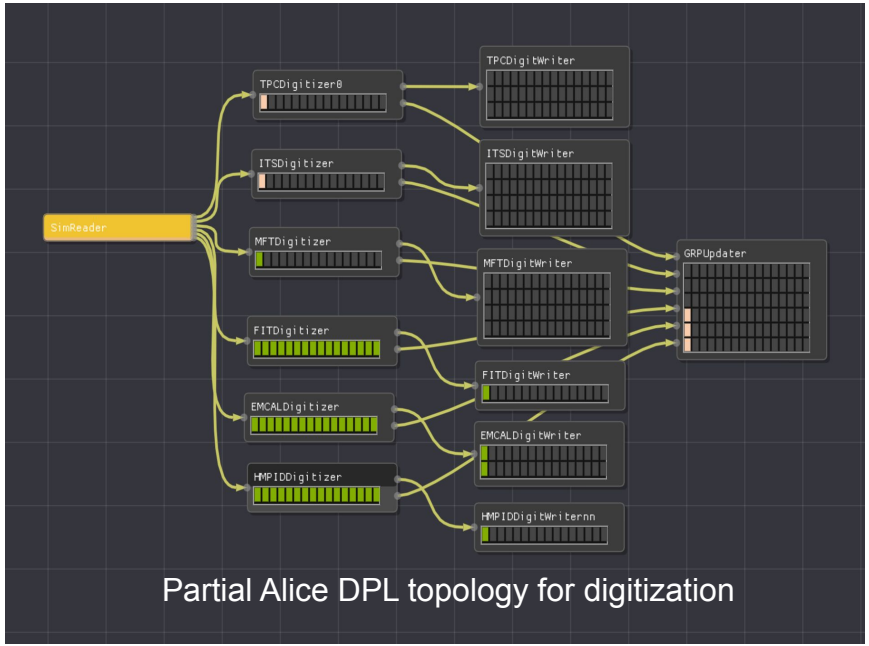
new

- **Distributed multi-process components based on FairMQ**
  - asynchronous generation, transport, IO
- **Backend-independent sub-event parallel simulation**
  - works for G4 and Fluka, G3
- **HPC / multicore ready**
  - nice CPU scaling and low mem-footprint (late worker forking)
- Developers/CI-testing appreciate ability to get few PbPb on the laptop in minutes rather than hours



Run3 developments/architecture for digitization

- **Complete rewrite or new development of digitization**
  - due to novel continuous readout + pileup
- **Chance to use modern C++ features and target optimizations from scratch**
  - Major gain in TPC digitization (6 to 20x faster than before)
- **Implemented using ALICE Run3 data processing layer (DPL)**
  - reactive data-flow framework also used in data taking / synchronous reconstruction / analysis
- **New parallelism features for multicore GRID:**
  - parallel within detector
  - … or across detectors

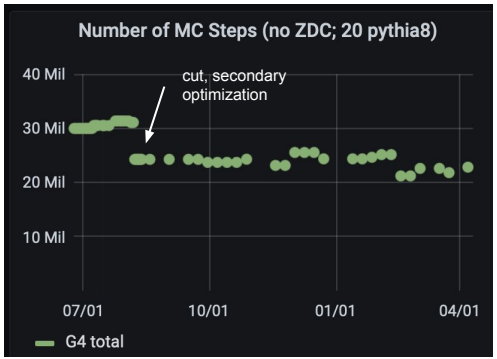


Partial Alice DPL topology for digitization

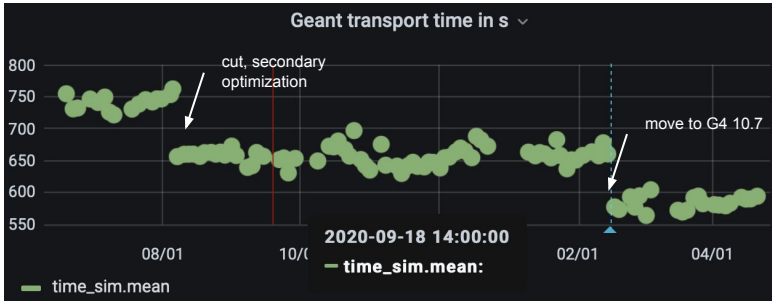
Run2-Run3 vanilla perf comparison (1-core)

- For vanilla full event sim, **Run3 upgrade refactoring + tuning already very impactful**

	G4 trans	digitization
PbPb	**~2x**	**~5x**
pp	**~1.8x**	**~20x**

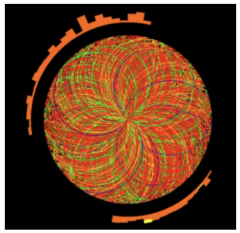


- **Potential factor ~2(PbPb) to ~5(pp) of more Run3 statistics for physics analysis!**
- Novel continuous perf monitoring an important aspect in this
- For vanilla mode, now **increased weight on Geant4 performance (less digitization)**

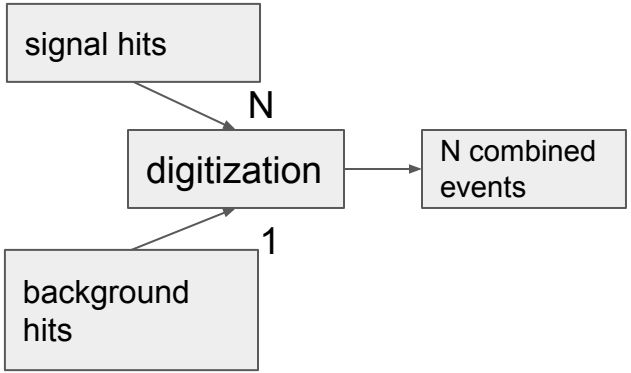


Embedding for signal (statistics) enhancement

- **Identified signal enhancement strategies** as key to provide **larger sample to physics analysis in Run3/4**
- **Reuse as much as possible expensively simulated background with multiple injected signals (similar to LHCb ReDecay)**
  - combine signal with same background during digitization
  - saves time from less background Geant transport
- Currently **yields another factor of ~4-5x overall !**
- Even faster and more sophisticated schemes in preparation (“background on signal filtering”)



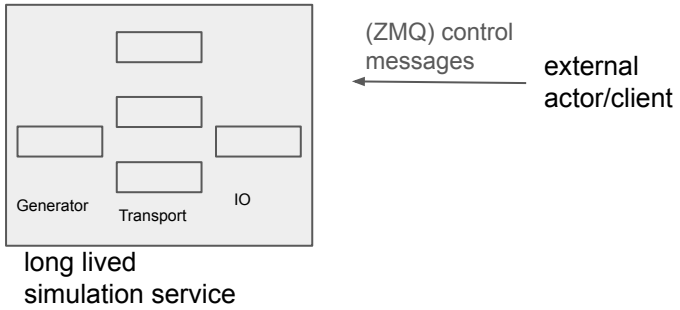
+ N x



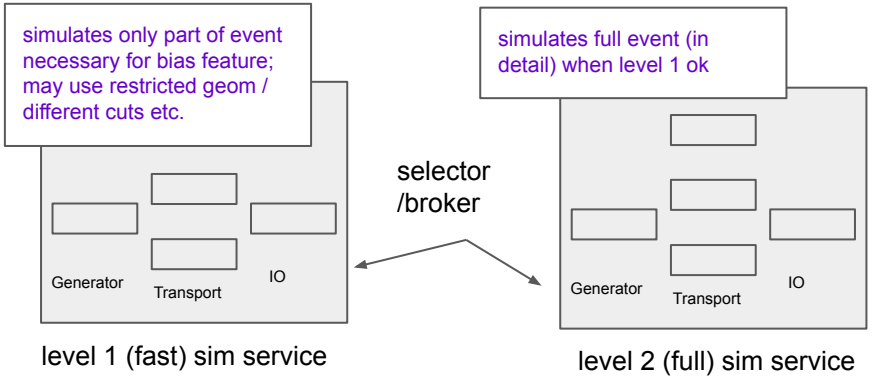
work in progress

Simulator as-a-service / improved biasing

- Offer new **“daemonized” simulator service** with client communication/interface via ZMQ
  - Enables “just-time-time” event requests from outside actor
  - May save overhead of repeated geometry/x-section initialization



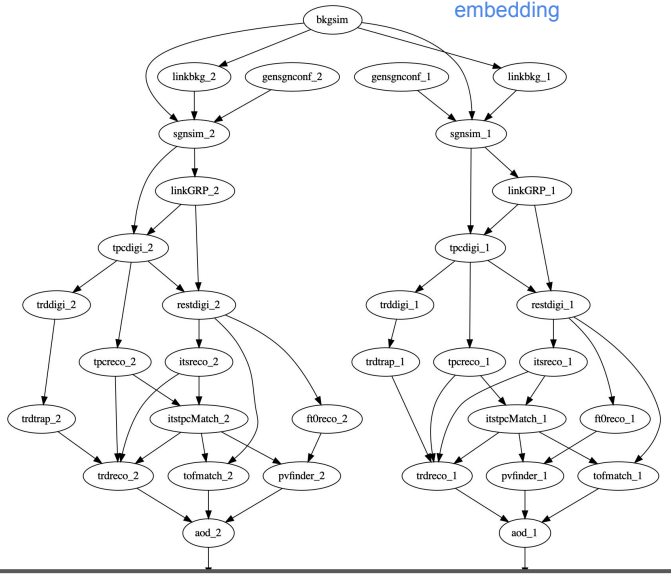
- **Building block** to construct **complex biasing (purity enhancing) schemes** via external event inspector (**start-stop-inspect-continue**)
- Enables **multi-stage simulation schemes** with crude and detailed sim



Novel multicore GRID executor “framework”

- ALICE globally based on multiprocessing rather than big multithreaded framework
  - sim, digi, reco different processes
- Previously (1-core world), a simple shell script ok to call processes in order
- Move to **8-core GRID** execution brings **a new degree of freedom,** requiring a more sophisticated approach to achieve high CPU efficiency

Toy graph pipeline for a 2-timeframe simulation with embedding



New **graph-oriented workflow modelling** - and resource aware **GRID executor/scheduler**

Last component to **achieve parallel execution of**
sub-events
timeframes
algorithms (Geant4, digitization, reco)

… and brings all **good features of an (ETL) build system such as**
- incremental execution
- restart from where last fail, ...

Latest multicore GRID performance

Where are we today in terms of cost and relative CPU share (PbPb)?

vanilla full sim

G4 (in %)	digi (in %)	reco (in %)
~85	~10	~5

full sim with signal enhancement through embedding (planned standard for physics analysis)

G4 (in %)	digi (in %)	reco (in %)
~24	~54	~22

cost per event ~12 kHS06ss

2x improvement

24 kHS06s (Run2)

(8-10) x improvement

cost per event ~ 3 kHS06s or less

- Hypothetical factor 2 G4 improvement => ~1.73 overall throughput increase

- Future Factor 2 G4 improvement => 1.14 overall impact
- Limited impact unless rest further optimized

Where we could gain or profit from in future

- Geant4:
  - VecGeom navigation (estimated 15-20% in transport time)
  - Fast ML GAN sim for certain detectors
    - clearly interested in standard toolkit and API within Geant4
    - limited to a few % since since no “outstanding” detectors except ZDC
  - Better community “tooling”, e.g.: automatic optimization frameworks
    - e.g., automatic cut/secondary tuning according to user objective would be tremendously useful

- Digitization on GPU:
  - digitization share still very relevant when using signal enhancement
  - good candidate for GPU; often natural massive parallelism
  - will become important as soon as GPUs become ubiquitous in computing resources

Summary

- ALICE Run3/4 simulation program is challenged by a much higher data rate
- We cannot keep the ratio MC / recorded data at the same level as before, using Run2 approach and software
- To bridge gap, managed to undertake transformations / optimization on a holistic level, achieving
  - HPC ready transport simulation
  - Considerably faster digitization
  - Efficient multicore GRID task scheduler
  - Signal embedding and event biasing framework to enhance MC samples in number and purity
- Future optimization in Geant4 now more impactful to us in standard full simulation (minimum bias)

Backup section

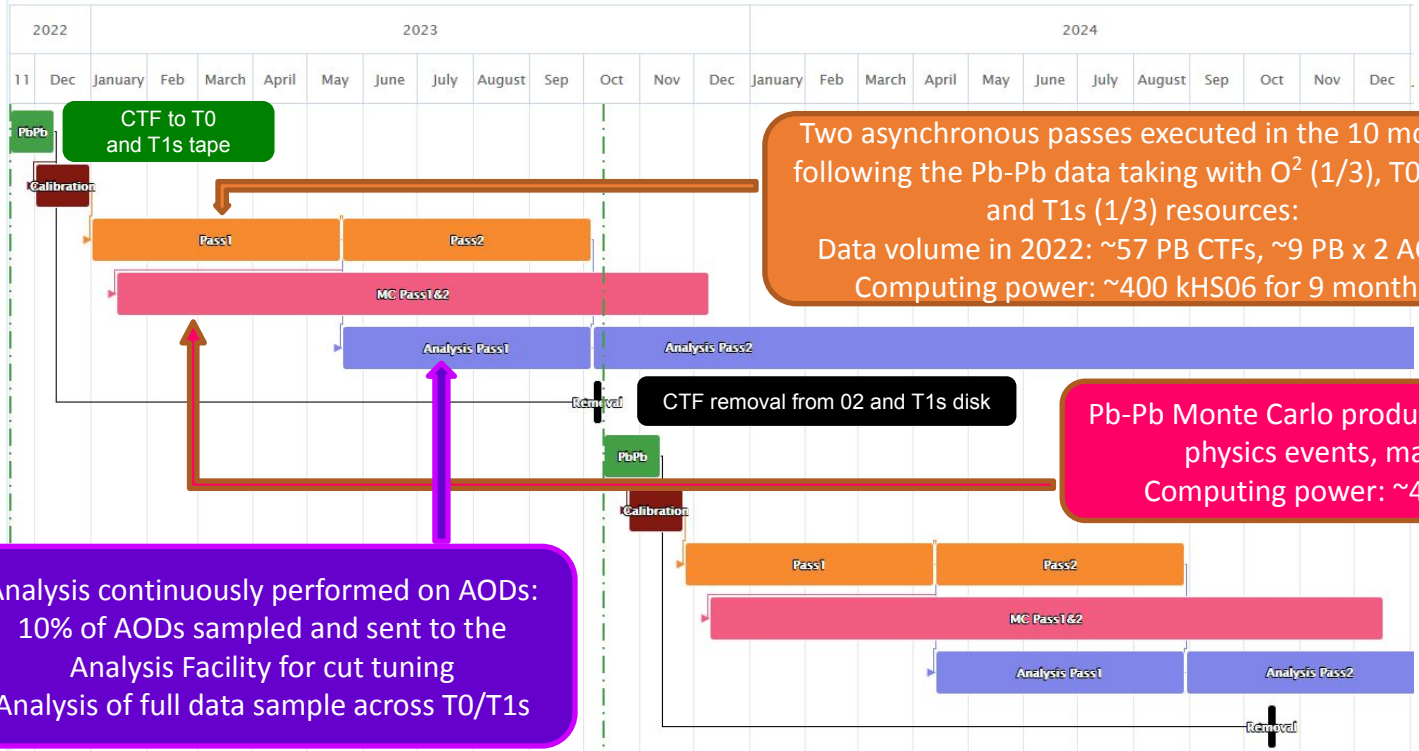
ALICE and Fast-Sim

- Embedding / biasing as a means to accelerate vanilla full sim
- Have a group working on ML method’s in shower generation

- Custom “fast sims” have been used in various detector studies (typically for upgrade TDRs)
- To this date, no systematic demand for “frameworked” parametric fast-sim by physics groups/analysis

- Currently preparing integration of Delphes into Run3 framework to enlarge offering

Pb-Pb processing plan



CTF to T0 and T1s tape

Two asynchronous passes executed in the 10 months following the Pb-Pb data taking with O^2 (1/3), T0 (1/3) and T1s (1/3) resources:
Data volume in 2022: ~57 PB CTFs, ~9 PB x 2 AODs
Computing power: ~400 kHS06 for 9 months

CTF removal from 02 and T1s disk

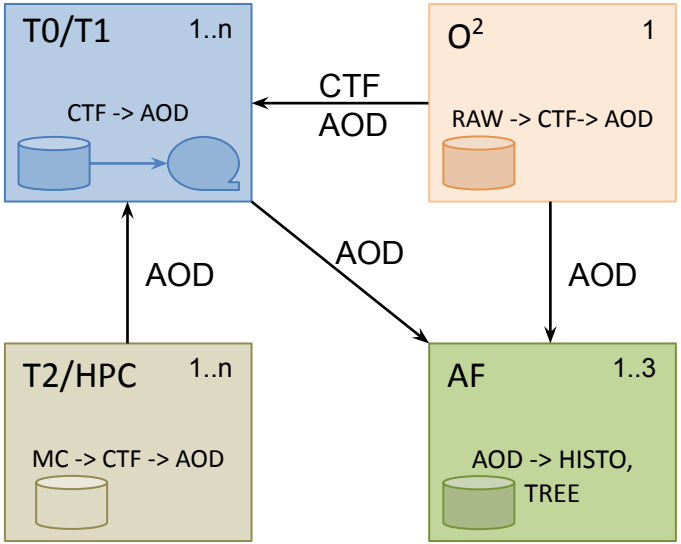
Pb-Pb Monte Carlo production: 2.0% of the number of physics events, mainly generated at T2s
Computing power: ~400 kHS06 for 11 months

Analysis continuously performed on AODs:
10% of AODs sampled and sent to the Analysis Facility for cut tuning
Analysis of full data sample across T0/T1s

ALICE Run 3 Computing Model

Grid Tiers will be mostly specialized for given role

Reconstruction
Calibration
Archiving
Analysis



Calibration
Reconstruction
Compression

Simulation

Analysis

CPU share	O^2	T0	T1s	T2s	AF
Synch.	100%	0%	0%	0%	0%
Asynch.	33%	33%	33%	0%	0%
MC	0%	0%	0%	100%	0%
Analysis	0%	30%	20%	0%	50%

- Subject to fine tuning
- MC can be run as a backfill

2/3s of CTFs processed by O^2 + T0 and archived at T0;
1/3 of CTFs exported, archived and processed on T1s;

One calibration (sync.) and two reconstruction passes (async.) over raw data each year;

The goal is to minimize data transfer and optimize processing efficiency

10% of AODs sampled and sent to the Analysis Facility for quick analysis and cut tuning;
Analysis of full data sample across T0/T1s only performed upon Physics Board approval.

O^2 TDR and addendum