Vertical Test / Data Challenge 2025

V. Friese
on behalf of F. Linz, K. Piasecki, F. Uhlig and many others

CBM Collaboration Meeting, 20 October 2025

Recap - Vertical Test: Why?

Slide taken from Collaboration Meeting February 2025

We have been developing software for CBM over many years in various areas:

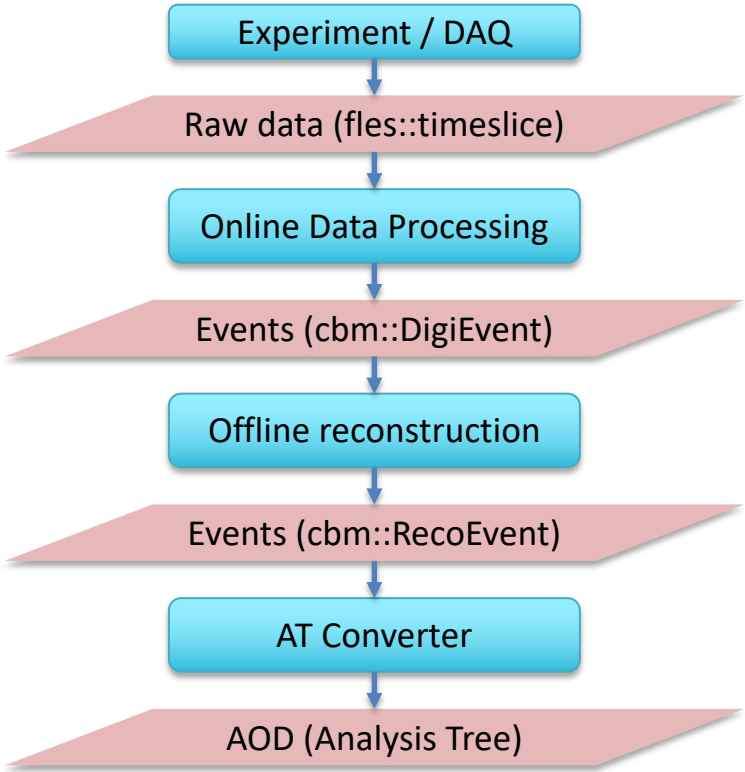
- Simulation (event-by-event)
- Simulation (time-based)
- Event reconstruction
- Time-based reconstruction
- Online data processing (unpacking, triggering, event selection)
- Physics analysis / feasibility studies / performance evaluation

In some three years from now, we will be faced with several PB of experiment data per year, to be analysed and physics results to be extracted and published.

How well are we prepared for that?

(and: which computing resources will we need?)

Our Data Processing Model – Coarse View



data stream; transient

define and select events

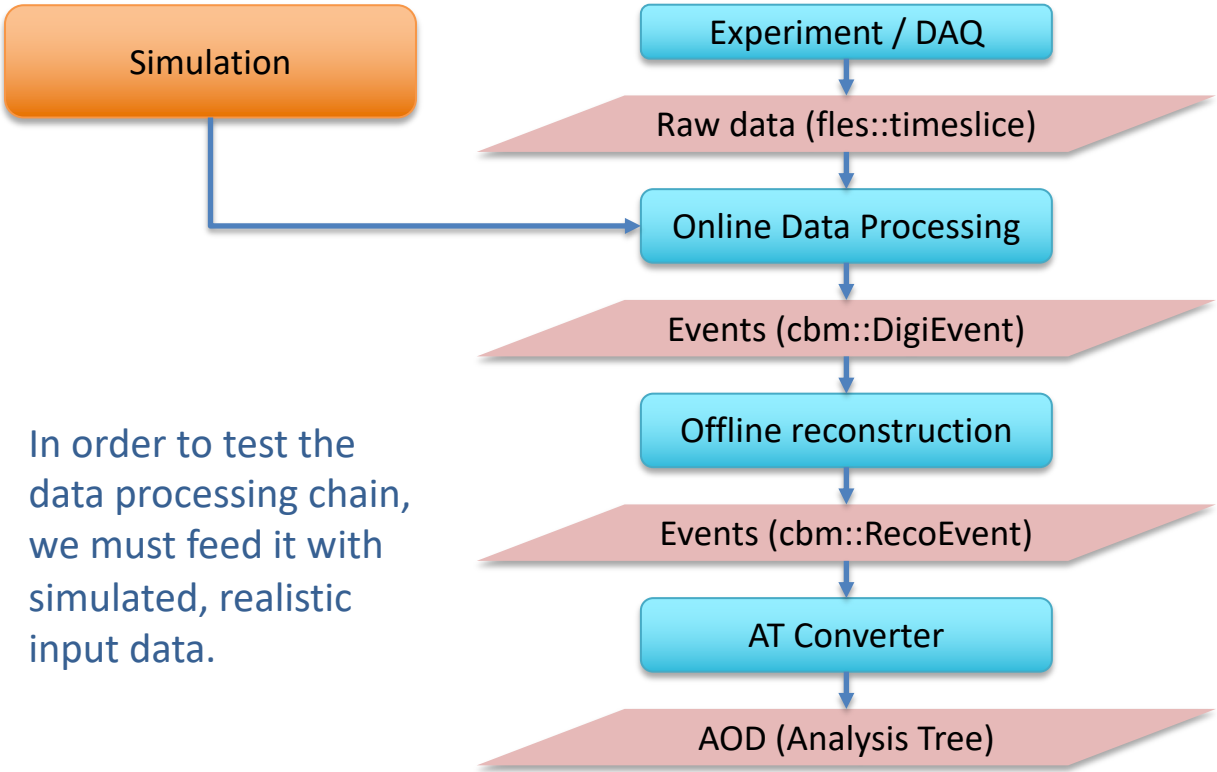
persistent (experiment data)

digis to hits, tracks, vertex

transient

persistent (limited period)

Our Data Processing Model – Coarse View



data stream; transient

define and select events

persistent (experiment data)

digis to hits, tracks, vertex

transient

persistent (limited period)

In order to test the data processing chain, we must feed it with simulated, realistic input data.

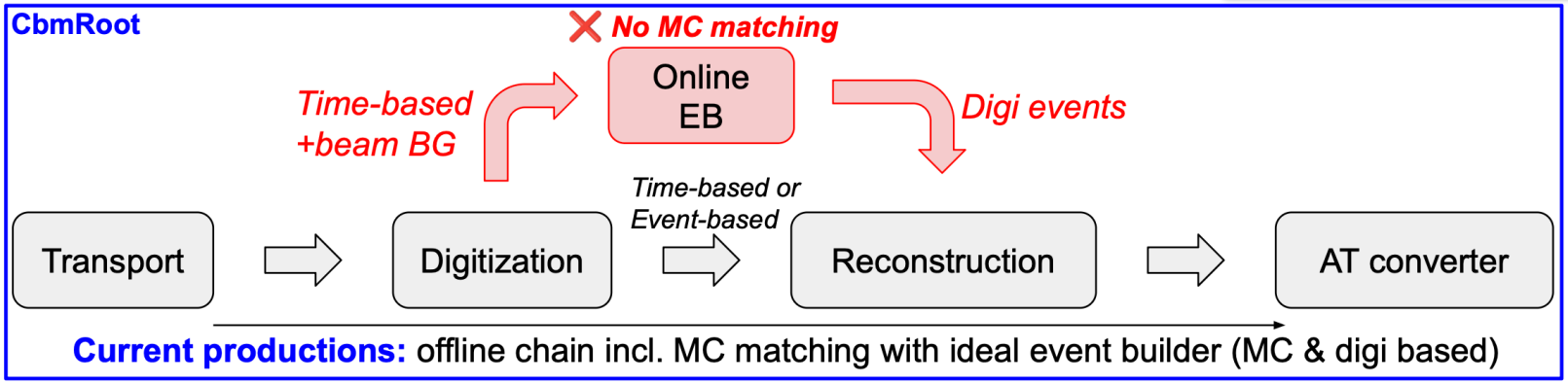
Vertical Test – Scope and Purposes

- General: Verify our data processing concepts (online and offline) on the base of realistic input data and of the current software stack (JUL25).
- Online: Apply strategies developed on the base of mCBM data for full CBM. Assess the required online computing resources. Identify performance bottlenecks for further optimisation.
- Offline: Identify missing pieces, inconsistencies, incoherences, (inconveniences).
- Analysis: Train physics analysis on large data samples w/o MC truth. Assess which amount of simulation statistics is needed to obtain the proper corrections. Compare finally obtained results with MC truth (generator level).

Vertical Test – What's Different?

Scope: Execute the entire software chain: simulation - **online** - offline - analysis with the same software stack. Compare physics results with MC truth.

From: F. Linz, Monday 11:00



Size: 1 – 5 M events -> 100 M events

Conditions for VT25 (Day-1 Operation)

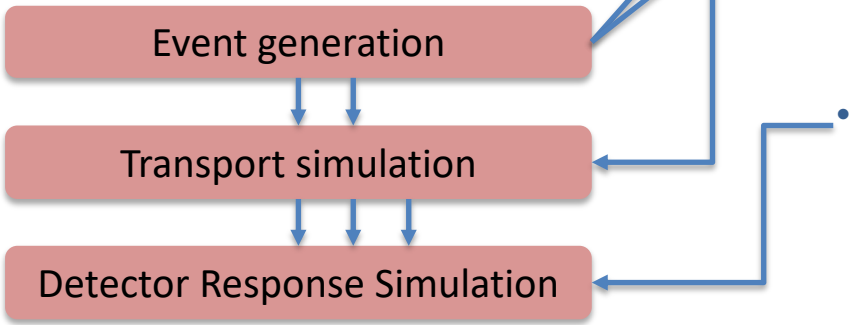
- Au+Au, p = 12 AGeV, 5 x 10^4 events/s
- Electron setup
- 10^8 events, corresponding to ~40 minutes of data taking at 80% duty cycle.
- Ideal geometry (no mis-alignment)
- Ideal response (no mis-calibration, no inefficiencies)
- Simple min. bias event trigger (digi multiplicity in TOF: not compute-intensive)
- No high-level event selection (physics trigger)
- Physics: bulk observables (di-electrons, hadrons, EbE)

More complexity (mis-alignment, mis-calibration, rare signals, high-level trigger) will be added in future VTs.

1st Step: Preparation of Input Data

A realistic data input shall comprise:

- Collision data (Au+Au)
- (optionally) signals to be embedded (electron pairs from PLUTO)
- Background from (non-interacting) beam passing through target and setup
- Noise from detectors and read-outs.



Event Generation

- Au+ Au collisions
  - Generator: UrQMD 4.0 (released June 2025)
  - Includes coalescence and some hyper-nuclei
  - Needed adaption of converter to Unigen (changed output format)
  - Settings: see Wiki or Redmine
  - Status: **done (10^8 events).**
- Signals for embedding
  - Electron pairs from vector meson decays, generated with PLUTO
  - Whether and how to embed: **still in discussion**
  - Current procedure of PWG Dilepton (embed in each event and apply weight in analysis) not applicable (no link to MC will be transported).

Transport Simulation

- Settings
  - Engine: Geant3 (still)
  - Geometry: sis100_electron (JUL25), except for FSD and beam pipe
  - Beam: Au, p=12 AGeV/c, centred on target centre, width 1 mm, no emittance
  - For more details, see Wiki.
- Status
  - UrQMD: 20,000 files with 5,000 events each, **in progress (20% ready)**
  - Beam: 1 file with 500,000 events (representative), **running**
  - PLUTO: in discussion

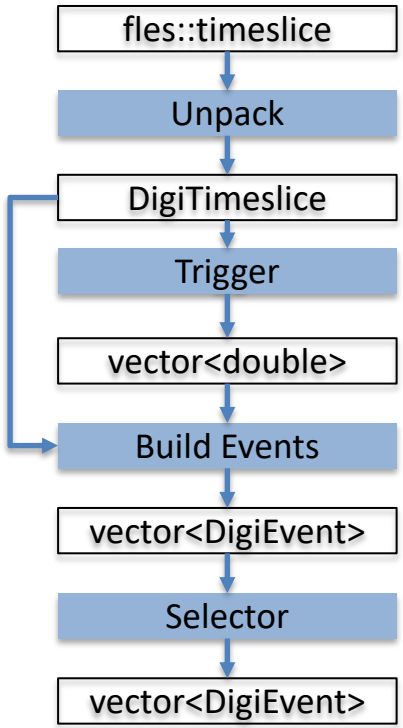
Detector Response Simulation (Digitization)

- Time-based with data streams:
  - UrQMD (+ embedded PLUTO) at 5 x 10^4 / s
  - beam events at 5 x 10^6 / s
    - use representative sample of 5 x 10^5 events (draw random)
  - Detector noise (continuous)
  - Strategy: create one timeslice of 0.2 s per file (should comprise all 5,000 events plus some pause).
- Status
  - Pilot (1 file, 5,000 events, w/o embedding, w/o beam background) concluded.
  - Is fast compared to transport, but consumes a lot of RAM (16 GB).
  - Will be done without MC matching (will not be recognised by online processing anyhow).

Detector Response Simulation: Open Issues

- Are the current (default) settings of the digitizers appropriate / realistic?
  - In particular: noise model (STS). At such low rates, noise is the dominating data source.
  - Need to describe reality in order to see a realistic data stream in online data processing.
  - Feedback from detector groups welcome, before we start the machinery…

Online Data Processing



- Interface from simulated data to fles::timeslice is prepared (wrap digis in micro-slices; no data operation in the unpacker)
- Options:
  - “offline”: process files in batch mode with online binary. Tested with pilot run (5,000 events).
  - ”online”: use replay mode to distribute time slices to Virgo nodes.
- Data processing: simple trigger, minimum bias (no event selector); i.p.: no online tracking.
- Status: online binary (formally) tested with pilot run.
- Todo: Adjust trigger settings (TOF digi multiplicity; trigger window, threshold) and settings (windows) for event builder. See presentation by A. Senger (Tuesday 12:10).

Offline Reconstruction

- Event-by-event reconstruction using run_reco.C
- Not so much to say; we are doing this since long.
- N.b.: will be replaced by updated reconstruction scheme suitable for parallelization on event level (also for online purposes) in future.
- Conversion into AT format.
- Will keep reco data since not all analyses yet use AT as input.

Progress Status

Step	Pilot (5,000 evts)	Test (5 x 10^6 evts.)	Production (10^8 events)
Event generation (UrQMD)	completed	completed	completed
Event generation (PLUTO)			
Transport (UrQMD)	completed	completed	running
Transport (PLUTO)			
Transport (beam)	completed	completed	completed
Digitization	evaluating		
Conversion to tsa	evaluating		
Online Data Processing	evaluating		
Offline Reconstruction			
Conversion to AT			
Analysis			

Resource Estimates (Update)

Stage	CPU h	Disk [TB]	Data Level
Event Generation (UrQMD)	120,000	12.5	f13
Event Generation (-> Unigen)		1.3	gen
Event generation (PLUTO)			gen
Transport (UrQMD)	35,000	44.0	tra
Transport (PLUTO)			tra
Transport (beam)	30	0	tra
Digitisation	5,500	17.0	digi_ts
Online Data Processing			digi_ev
Offline Data Processing	50,000	32.0	reco
AT converter	25,000	7.5	aod
Total		114.3	

Features of the VT

- Go step-by-step with data integrity checks and basic data QA.
- Balance care and progress speed.
- Document procedures and (configurable) scripts -> ease future VTs and other productions.
- Along the process: Note and document experiences and categorize found issues.

- Fatal: To be fixed for this VT. Requires patches to the JUL25 release.
  (none so far)

- Necessary: missing feature, to be addressed after the VT

- Convenience: Address once free resources are available.

Documentation: Follow the Progress

[Redmine]

[Wiki]

Untergeordnete Tickets 30 (19 offen — 11 geschlossen) Hinzufügen

Feature #3526: VT25 - Event Generation	New	Florian Uhlig	05.02.2025	10.10.2025	
> ~~Feature #3527~~: Choice of event generator	Closed	Florian Uhlig	05.02.2025	13.05.2025	
> ~~Feature #3528~~: Settings of the event generator	Closed	Volker Friese	14.05.2025	14.05.2025	
> ~~Feature #3529~~: Small-scale test production	Closed	Volker Friese	20.05.2025	29.05.2025	
> ~~Feature #3530~~: Planning of large-scale production	Closed	Volker Friese	30.05.2025	30.05.2025	
> Feature #3531: Perform large-scale data production (GEN)	In Progress	Volker Friese	19.09.2025	10.10.2025	
> Feature #3562: Embedding of signals (PLUTO)	In Progress	Partha Pratim Bhaduri	26.02.2025	07.03.2025	
> ~~Feature #3629~~: Adaption of Unigen UrQMD converter	Closed	Volker Friese	14.05.2025	19.05.2025	
> ~~Development #3658~~: Unigen converter for UrQMD v4	Closed	Volker Friese	30.06.2025		
Feature #3532: VT 25 - Transport Simulation	New		05.02.2025	17.10.2025	
> ~~Feature #3533~~: Geometry setup	Closed	Eoin Clerkin	05.02.2025		
> ~~Feature #3534~~: Choice of transport engine	Closed	Volker Friese	05.02.2025	15.04.2025	
> ~~Feature #3535~~: Small-scale test production and resource estimate	Closed	Volker Friese	30.05.2025	30.05.2025	
> ~~Feature #3536~~: Production planning	Closed	Volker Friese	02.06.2025	02.06.2025	
> Feature #3537: Documentation	In Progress	Florian Uhlig	05.02.2025		
> Feature #3538: Large-scale data production (MC)	In Progress	Frédéric Julian Linz	13.10.2025	17.10.2025	
> Feature #3541: Beam-target simulation	In Progress	Volker Friese	07.02.2025		
Feature #3542: VT 25 - Digitization (Detector Response Simulation)	New	Alberica Toia	07.02.2025	24.10.2025	
> Feature #3543: Parameters and configuration	New		07.02.2025		
> Feature #3544: Small scale test production (digi)	In Progress	Volker Friese	02.06.2025	02.06.2025	
> Feature #3545: Resource estimate and production planning	New		03.06.2025	03.06.2025	
> Feature #3546: Large scale data production (digi)	New		20.10.2025	24.10.2025	
> Feature #3607: Software issues with time-based digitization	New	Frédéric Julian Linz	09.04.2025		
Feature #3591: VT 25 - Online Data Processing	New	Volker Friese	31.03.2025	26.12.2025	
> Feature #3592: Enable cbmreco to take Digitimeslice input	In Progress	Bartosz Sobol	31.03.2025	17.10.2025	
> ~~Feature #3593~~: Distribution of Digitimeslices to the processing nodes	Closed	Bartosz Sobol	20.10.2025	26.12.2025	
> Feature #3594: Choice of trigger	New	Volker Friese	14.04.2025	06.06.2025	
> Feature #3595: Small-scale test production and resource estimate	New	Volker Friese	09.06.2025	12.06.2025	
> Feature #3596: Resource allocation	New	Volker Friese	13.06.2025	20.06.2025	
> Feature #3597: Online Data Challenge	New	Volker Friese	27.10.2025	31.10.2025	

Vertical Test 2025

The Vertical Test 2025 (VT25) is the first data challenge comprising the entire chain from simulation to physics analysis, including all stages of data processing anticipated to be used in the actual experiment. The aim of this enterprise is to assess how well prepared we are for the start of actual data taking currently foreseen for 2028, to verify the data procession concept as such, to identify missing pieces and define the next work steps and milestones, and to check the overall consistency.

VT25 will be performed on the base of 10^8 Au+Au events at p_{beam} = 12 GeV/nucleon in the electron setup, which is the one that will be used for first data taking. This event statistics corresponds to about 40 minutes of data taking at 5 x 10^4 events/s at 80% duty cycle. Data will be taken minimum-bias (no high-level physics trigger). The software base is the JUL25 release of cbmroot.

The final objective is to derive physics results for a smaller number of benchmark observables and to compare them to the MC truth. In the process, the used computing resources shall be measured and monitored in order to extrapolate to the needs for CBM operation in day-1, updating and replacing the current resource estimates.

The processing stages are:

- event generation,
- transport simulation,
- digitization (time-based, full data stream including beam-target events and noise),
- online data processing (emulation),
- offline reconstruction and AOD production,
- offline analysis.

These pages are to document the processing stages, findings and results.

Other information sources

- Presentations
- Planning and discussions in Redmine

Data processing steps

1. Event generation

2. Transport simulation

Outlook

- The current VT25 will be completed (data given to PWGs for analysis) this year.
- After the VT is before the VT! Possible next issues (based on priorities and findings in this VT):
  - High rates with muon setup and selective (di-muon) trigger
  - High rates with hadron setup and selective (hyperon) trigger
  - Insert mis-alignment and inefficiencies in the simulation
  - Higher statistics if use case is present
- We can re-use the event generator data (for the same collision system) and transport data (for the same geometry setup) -> skip compute-expensive steps.
- We should target two VTs / Data Challenges per year until 2028.

Outlook

- The current VT25 will be completed (data given to PWGs for analysis) this year.
- After the VT is before the VT! Possible next issues:
  - High rates with muon setup and s
  - High rates with hadron setup and
  - Insert mis-alignment and inefficie
  - Higher statistics if use case is pres

- We can re-use the event generato stem) and transport data (for the same geometry setup) -> skip compute-
- We should target two VTs / Data C

Let‘s work together
to be prepared for real data in 2028!