# The CBM Triggerless Readout - Concepts and Prototyping

*Volker Friese*

GSI

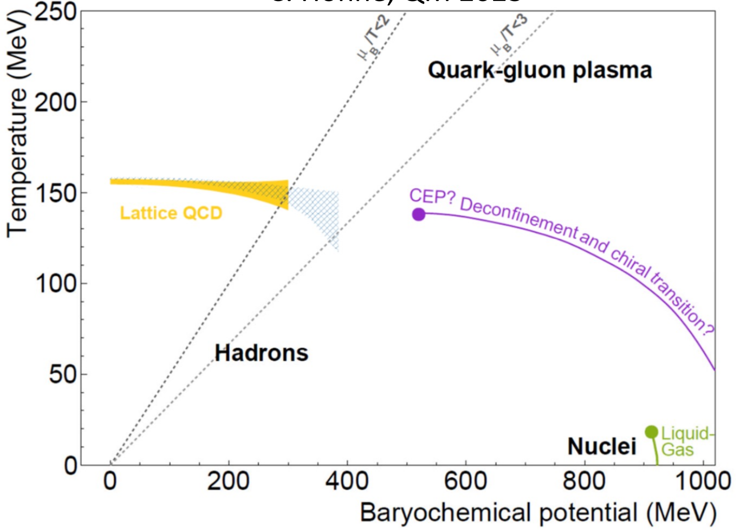
*Helmholtzzentrum für Schwerionenforschung*

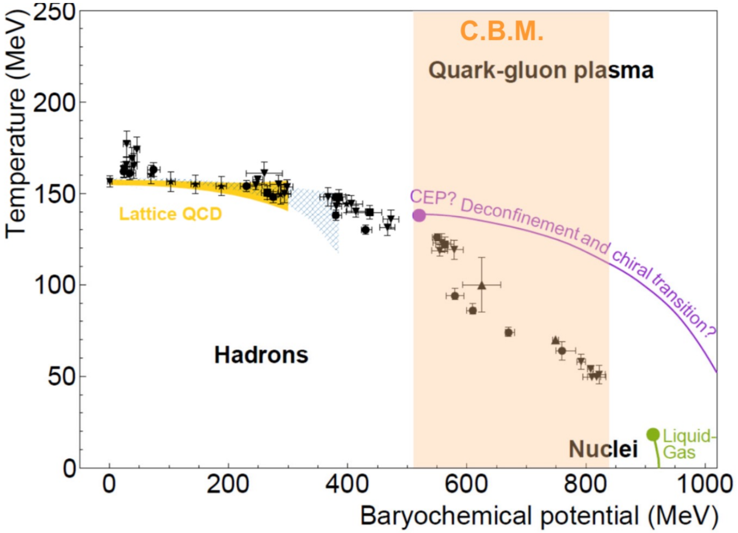
Workshop on Fast Realtime Systems

Gießen, 8 April 2024

# QCD Matter Physics: Theory and Experiment

C. Höhne, QM 2023

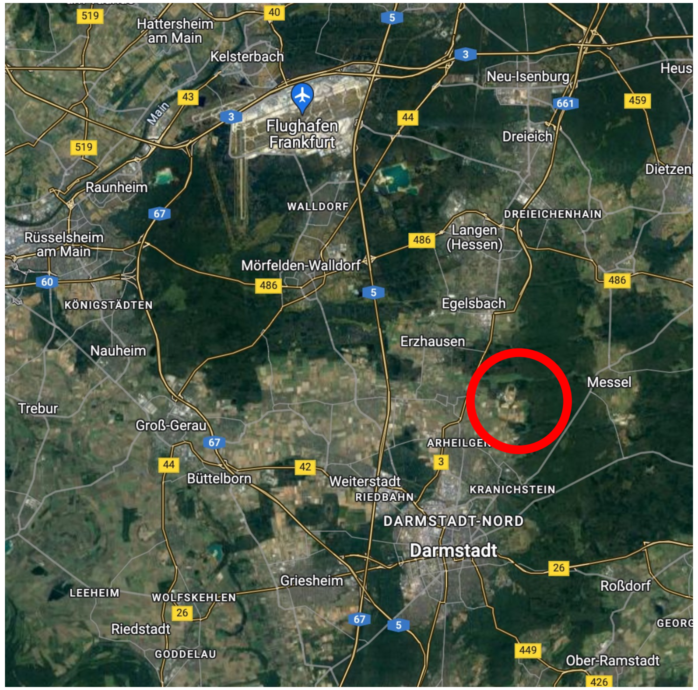



- Phase structure: 1st-order phase transition? Critical point? If yes, where?
- Equation of state?
- New QCD phases?

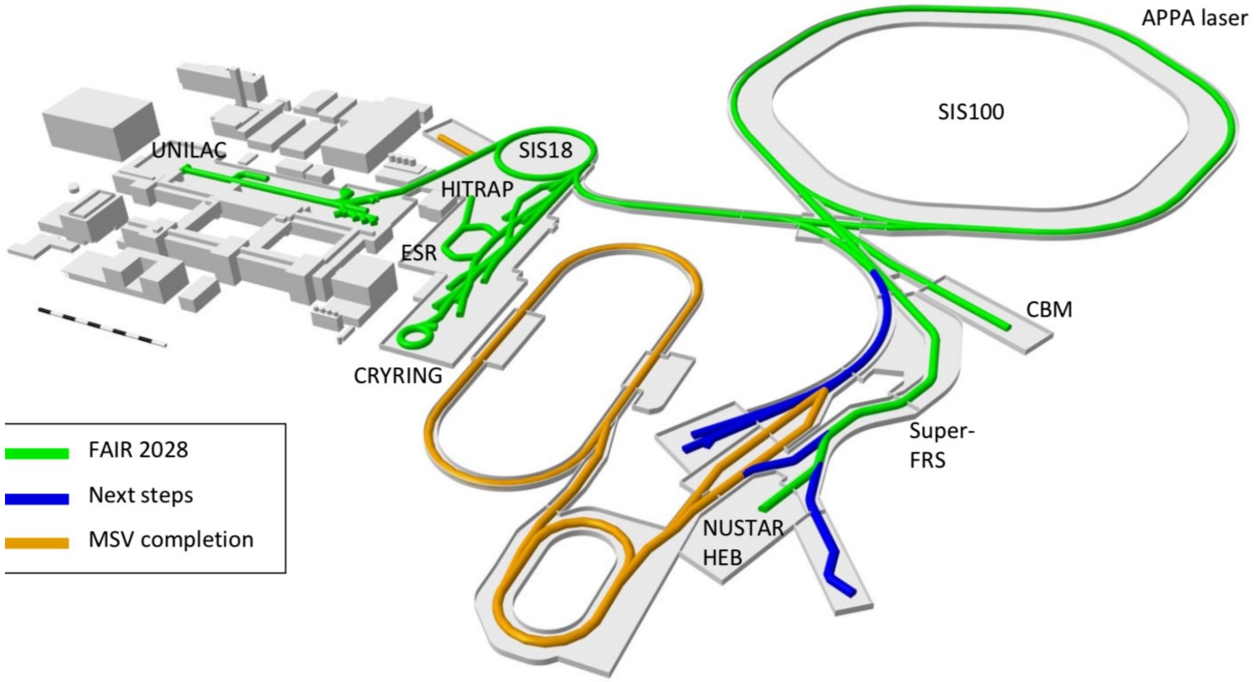
- Freeze-out points map the phase diagram
- Variation of collision energy allows to study different regions:
  - High energies (LHC, RHIC): vanishing density, high temperature
  - Lower energies (AGS, **FAIR**, GSI): high density, moderate temperature

# What’s FAIR?



- Facility for Anti-proton and Ion Research
- A major new infrastructure for basic and applied research at GSI in Darmstadt, Germany
- 11 international partners
- Currently under construction - start of operations 2028

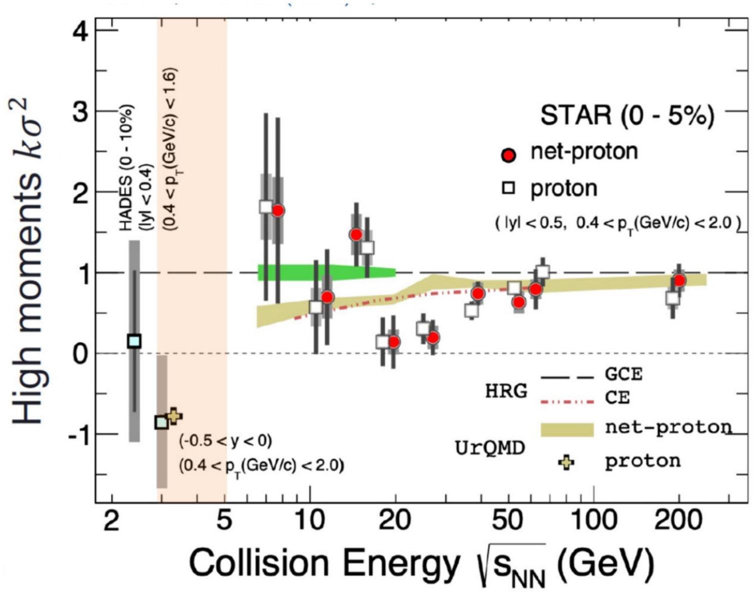
# What's FAIR?

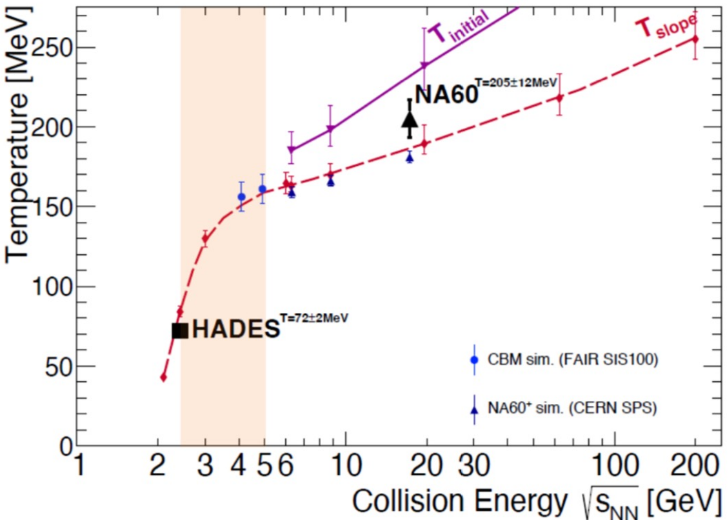


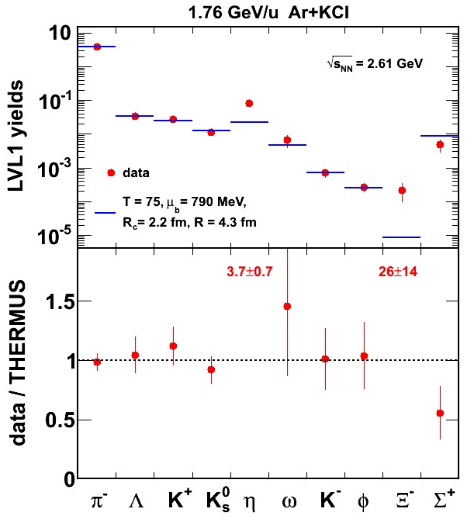

- Complex accelerator facility for a variety of physics
- Including slowly extracted nuclear beams for fixed-target experiments
- Full stripped heavy ions up to p = 12 GeV/c with intensities up to $10^{10}$ ions / s.
- Symmetric nuclei up to 14 GeV/c; protons up to 30 GeV/c.

CBM energy range:
P = 3.5 - 12 GeV/c
$\sqrt{s_{NN}}$ = 2.3 - 5.3 GeV

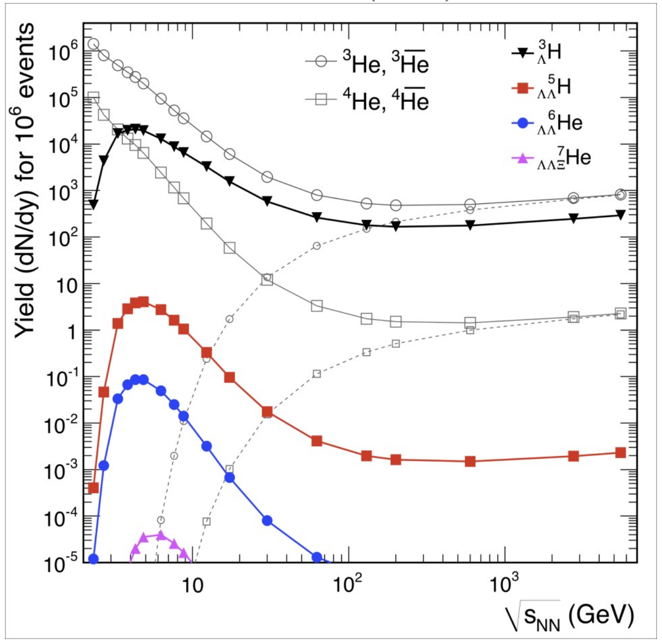
# A rich menu of observables




*A. Andronic et al., PLB 697 (2011) 203*

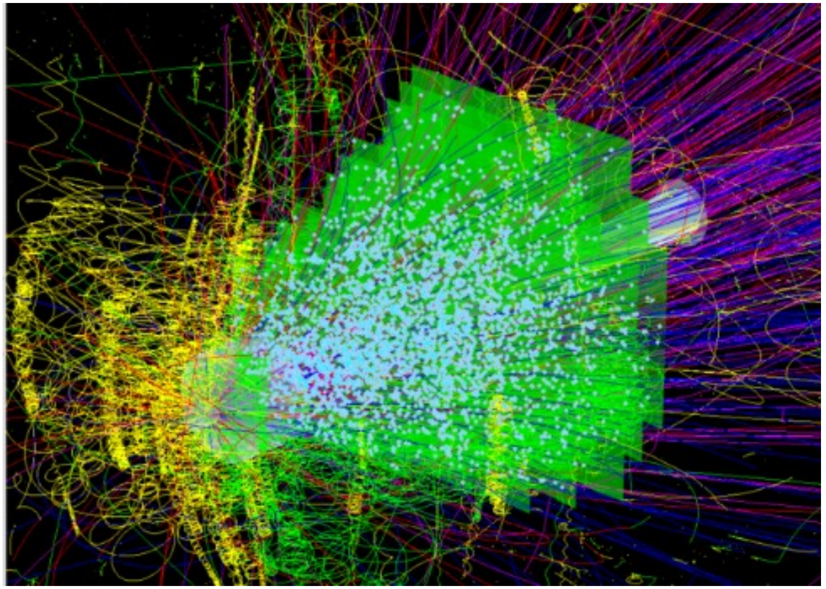



... some of which are really rare!

# CBM in a Nutshell

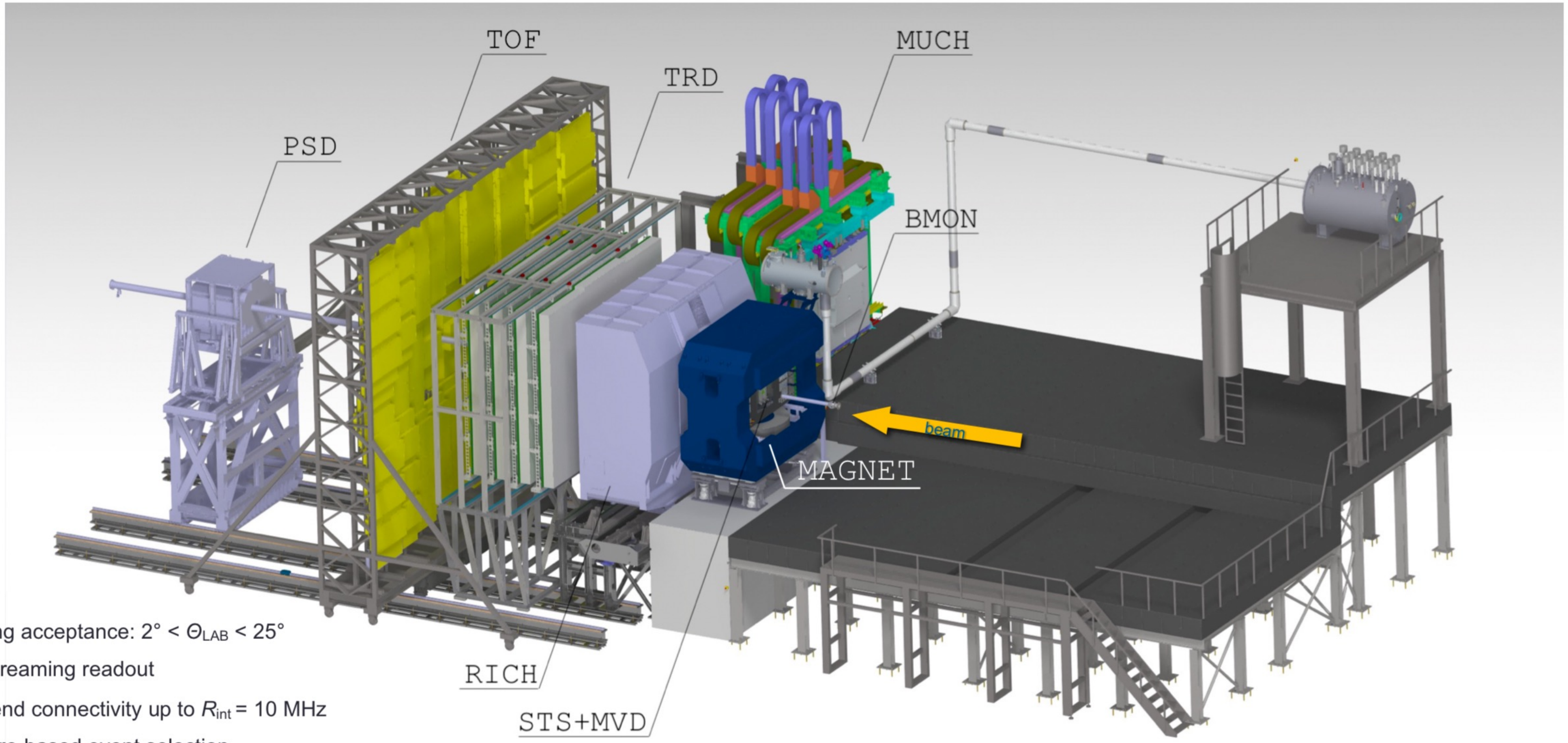
- Measure charged pions, kaons, protons
- Measure electrons and muons
- Reconstruct weak decay topologies (hyperons, hyper-nuclei, charm)
- In fixed-target collisions in the FAIR-SIS100 energy range
- In a large acceptance
- With high precision
- With high statistics

Simulated Au+Au collision @ 12 AGeV (UrQMD)



Punchline: Moderate collision energy, very high rates

# CBM as it will be

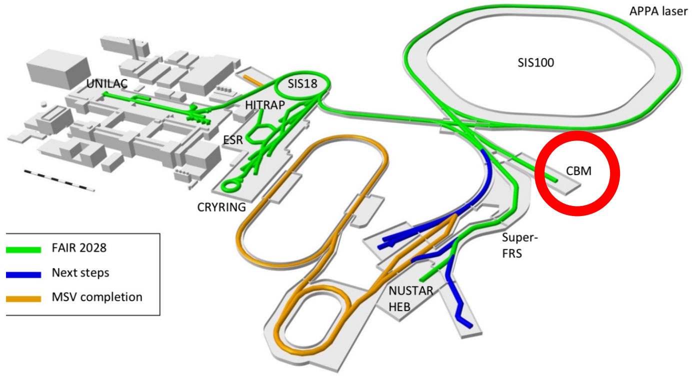



Tracking acceptance: 2° < $\Theta_{LAB}$ < 25°

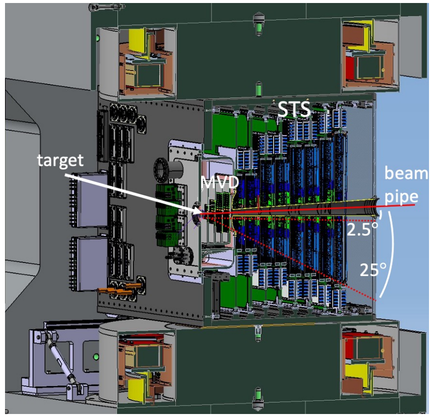
Free streaming readout

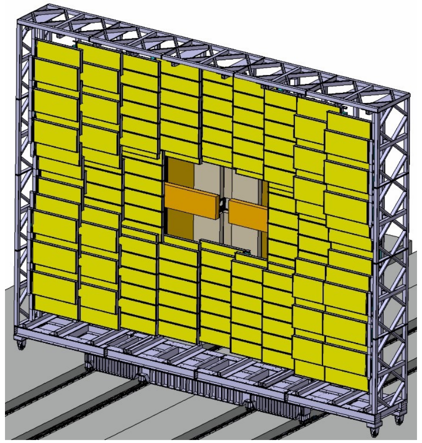
Front-end connectivity up to $R_{int}$ = 10 MHz

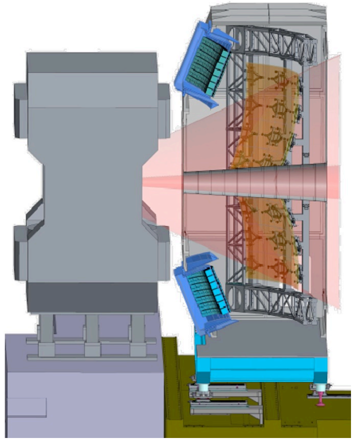
Software-based event selection

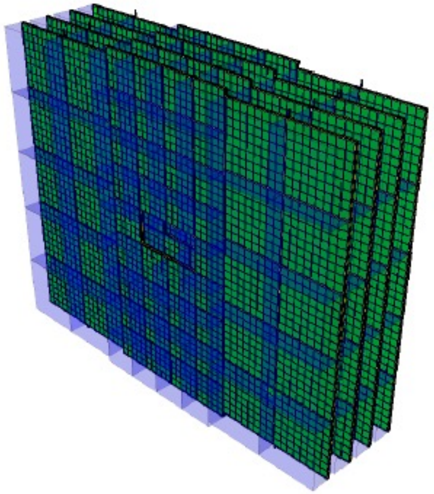


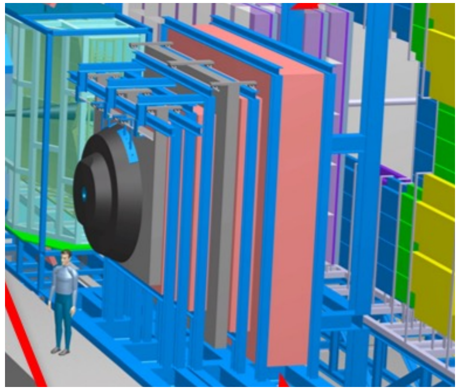

# Detector Systems









Silicon Tracking System for tracking of charged particles

Micro-Vertex Detector for measurement of displaced vertices

Time-of-Flight detector for identification of hadrons

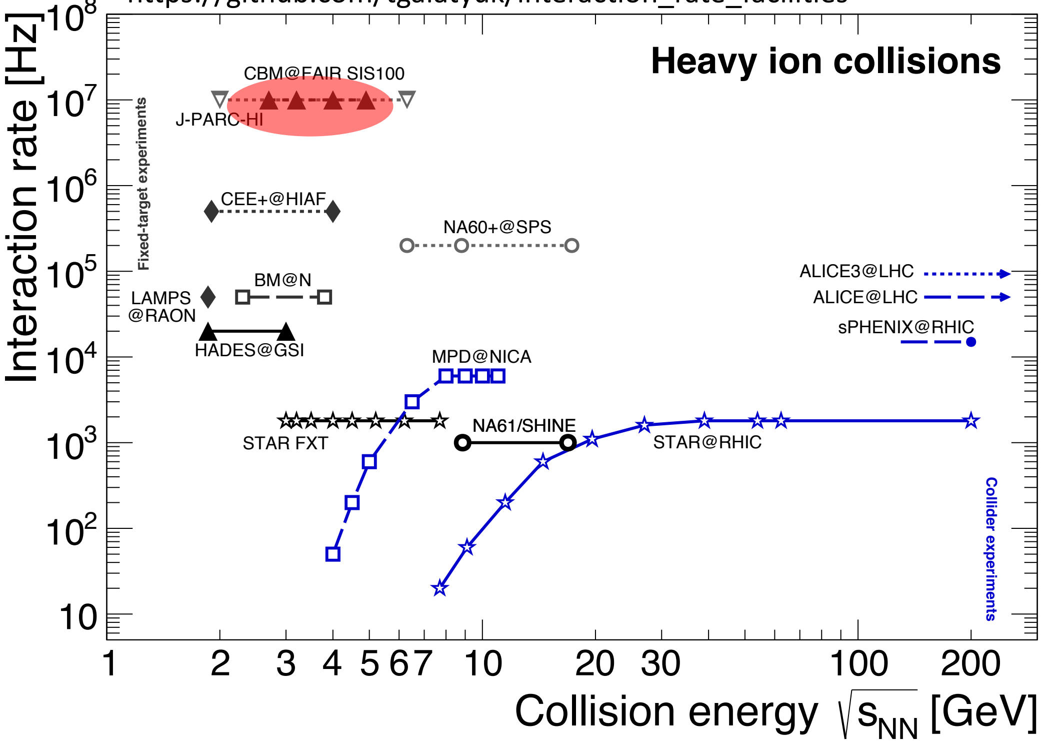
Ring-Imaging Cherenkov detector for identification of electrons

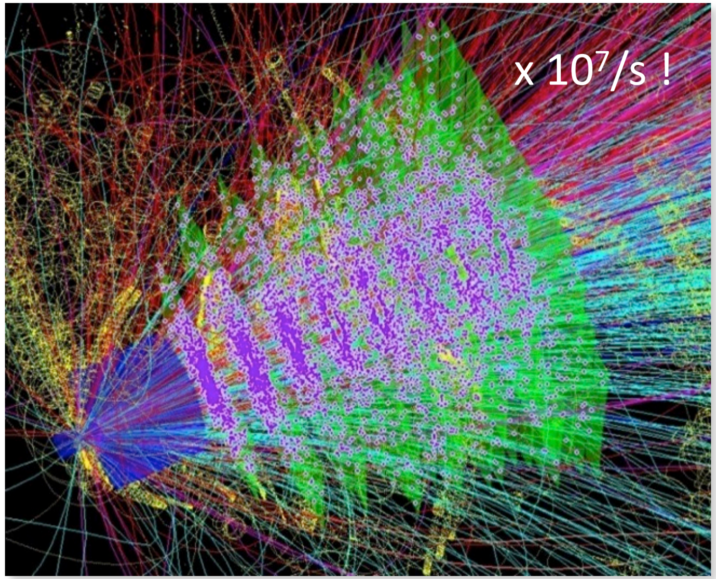
Transition-Radiation Detector for identification of electrons and intermediate tracking

Muon System for identification of muons

# The High-Rate Frontier

Compilation by T. Galatyuk:
https://github.com/tgalatyuk/interaction_rate_facilities




CBM design rate: up to 10 million collisions per second.

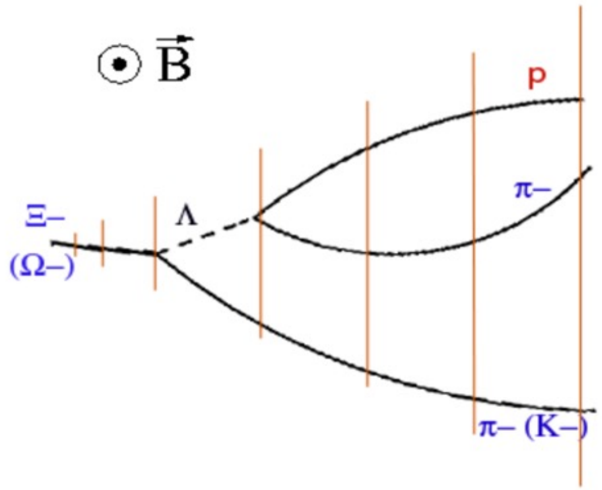
# Rate issues

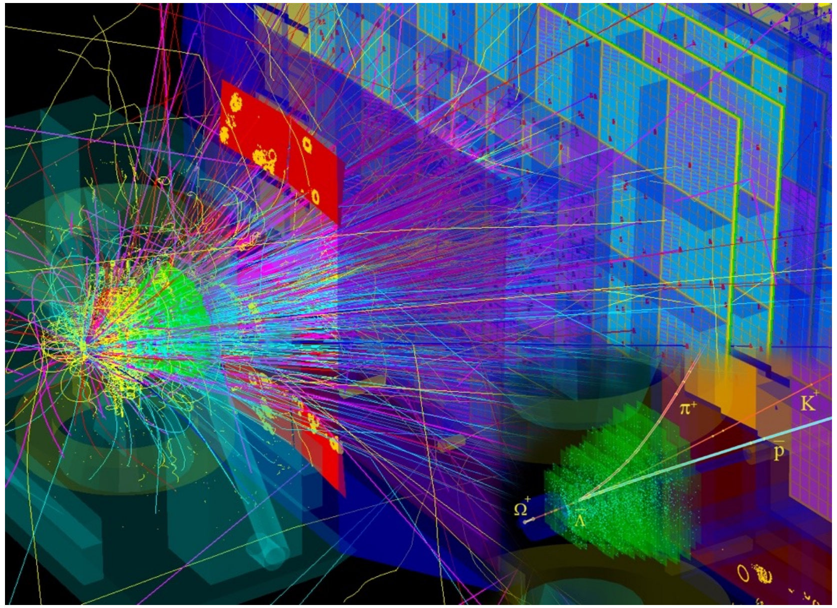
- Radiation load on detectors and front-end electronics. Difficult, but not unsurmountable.
- Fast timing response of detectors and electronics
- HUGE data rate coming off the detectors

- Raw data size: ~100 kB / event (min. bias Au+Au @ 12 GeV/u).
- $10^7$ events/s -> 1 TB/s raw data rate from the detector.
- Archival rate: 100 GB/s surely possible; 1 TB/s on the high side
- Data volume, assuming 2 months of operation per year ($5 \times 10^6$ s):
  5 EB / year!

Data selection in real-time (by about 3 OOM) is base of the CBM experimental concept.

# Trigger issues 1: Signatures are complicated



Complicated event topology!

Trigger decision requires track reconstruction and detection of the decay topology.

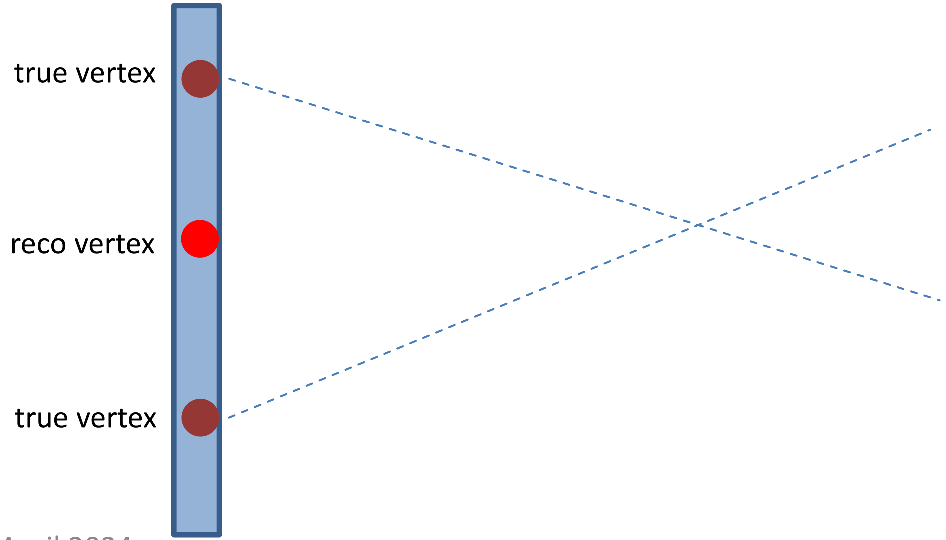
No chance to realize that in hardware logic. Has to be done in software.

# Trigger challenges 2: Event overlap (pile-up)

- Average time interval between events: 100 ns (Poisson process)
- Event duration: about 40 ns
- Close-by events will overlap in time
- Hard to detect by simple means; disentangling pile-up events requires close inspection (tracking)
- Not to be realized in a hardware trigger; to be done in software.

Pile-up can spoil the physics - and the trigger!

- Two peripheral events on top of each other may look like one central event.
- Two primary tracks from different events my look like secondary tracks -> may fake a decay trigger



# Note bene: Fixed-target vs. collider experiments

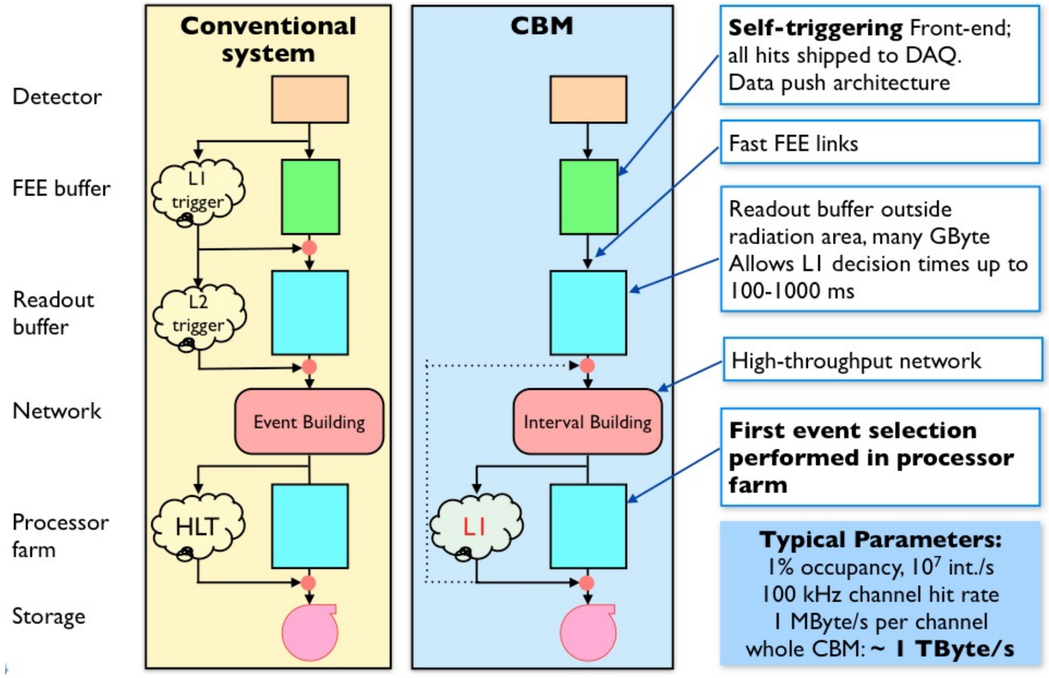
*Collider*

- Collisions at well-defined points in time (bunch crossing): external trigger from the machine clock.
- Macroscopic bunch crossing area: pile-up events can be resolved in space.

*Fixed-target*

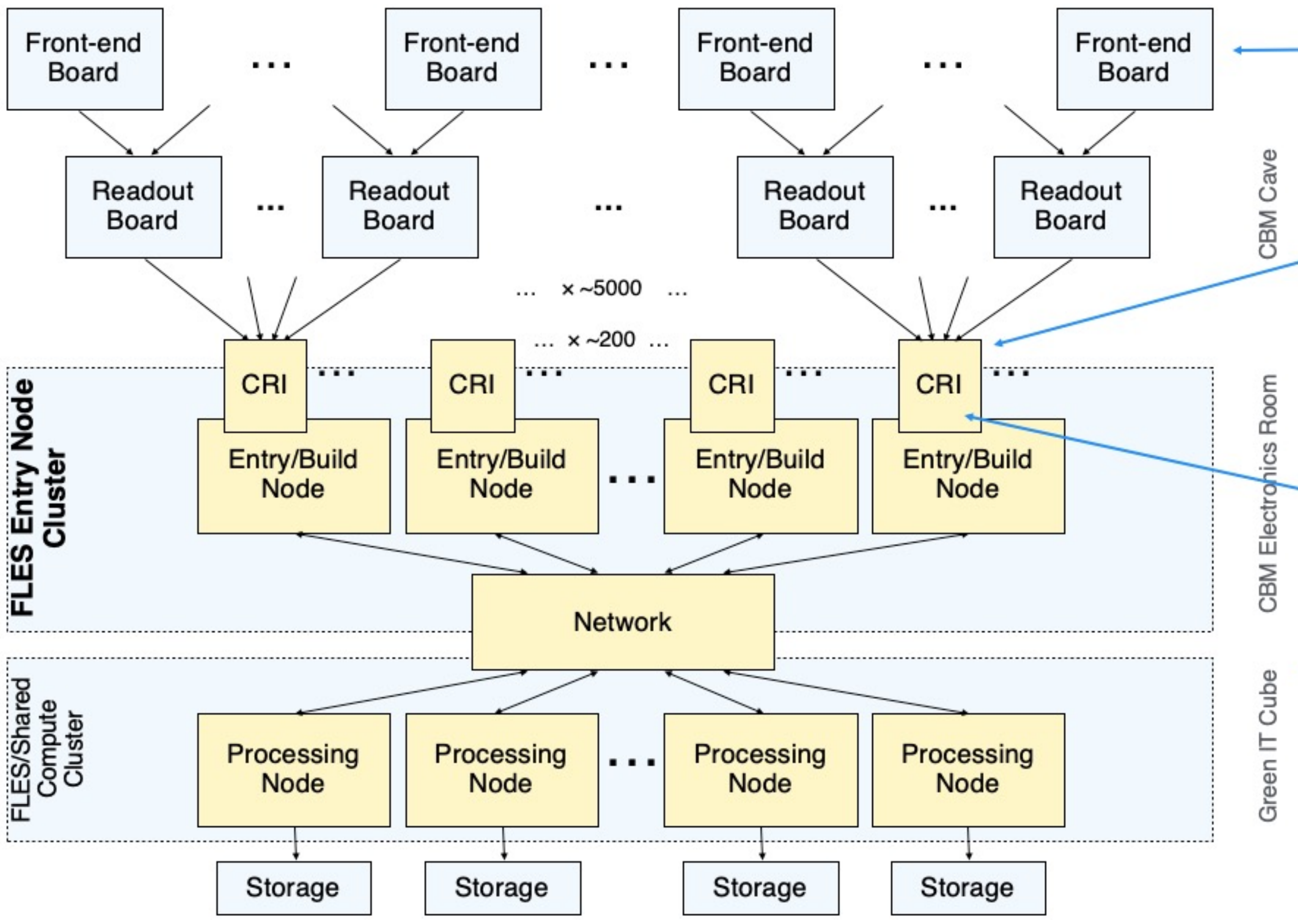
- Continuous beam on target; continuous collisions.
- No beam trigger.
- All interactions take place in a thin target.
- Collision must be resolved in time.

# Meet the Challenges: Free-Streaming Readout




- No hardware trigger at all
- Continuous readout by autonomous FEE
- FEE sends data message on each signal above threshold (“self-triggered”)
- Hit message come with a time stamp; readout system is synchronised by a central clock
- DAQ aggregates messages based on their time stamp into “time slices”
- Time slices are delivered to the online computing farm (FLES)
- “Event building” and storage decision on data selection is done in software in the FLES.

# CBM Readout Architecture




**Self-triggered** readout electronics

Sends timestamped hit messages

**Common Readout Interface (CRI)**

FPGA-based PCIe board

Interface to COTS nodes

**FLES Interface Module (FLIM)**

HDL module as subsystem interface

Interface between detector specific and generic DAQ data path

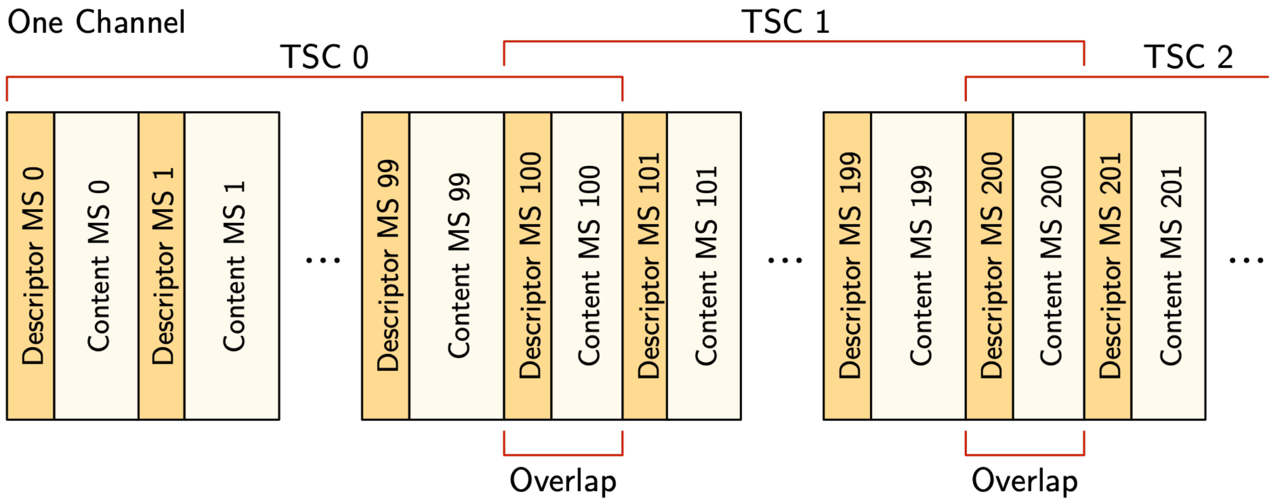
**First-level Event Selector (FLES)**

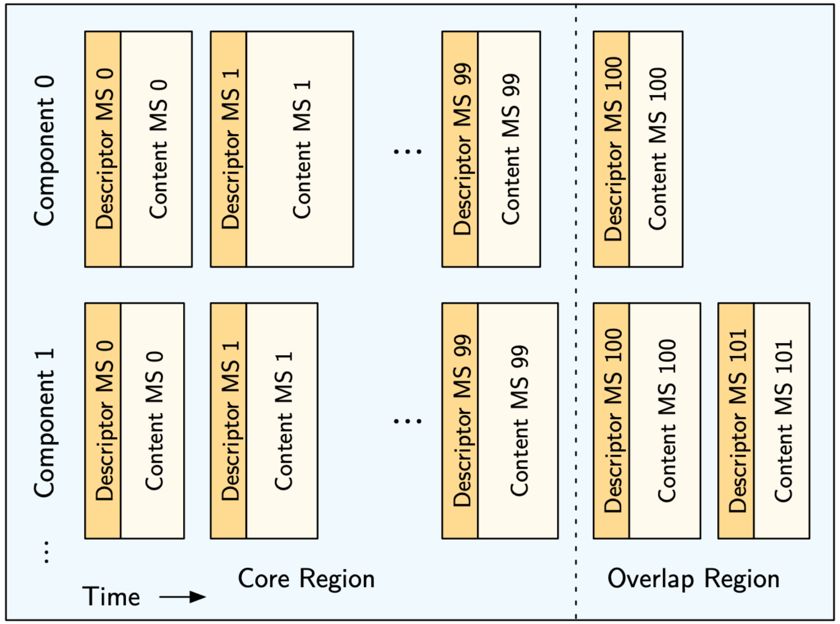
Dual HPC cluster

Entry stage design data rate > 1 TB/s

High-speed, RDMA-enabled network

# Timeslice Building: Data Model

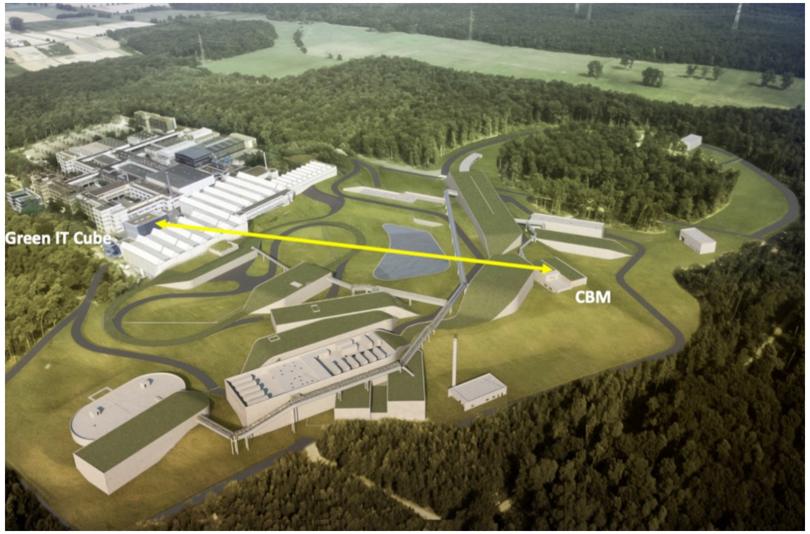



- Partitioning of the data streams into short, context-free time intervals and encapsulating them into data transport containers called microslices
- Meta data provides all necessary information for data handling, e.g., start time
- Partitioning can be done locally by the subsystems
- Free choice of optimal detector data format for each subsystem

# Online Data Selection

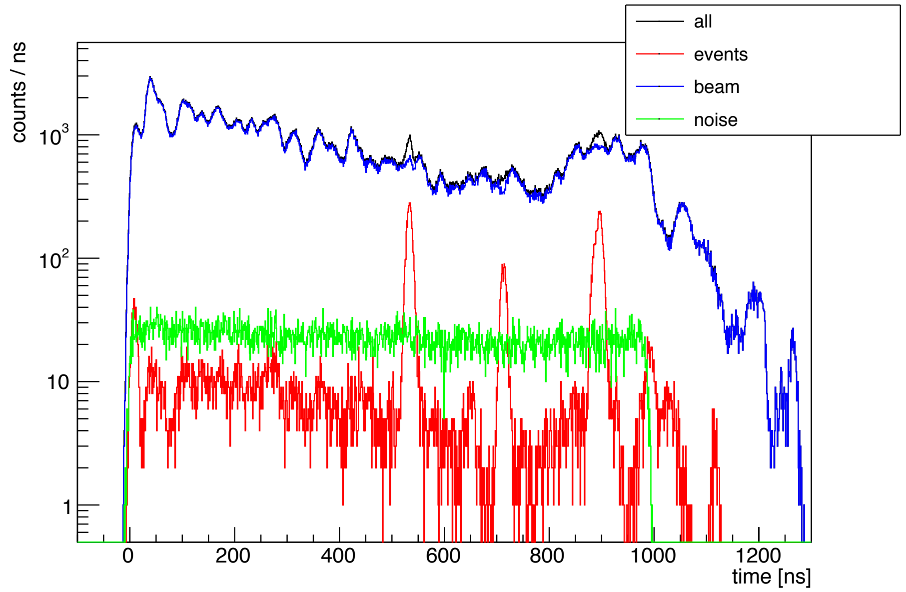
- Time and charge calibration
- Cluster and hit reconstruction
- Track reconstruction
- Event construction
- Event analysis (w.r.t. trigger signature) and trigger decision




- Traditionally done offline
- To be done in real-time
- Needs very fast algorithms
- Massively parallel computing (multi-core, GPU) in the GSI Data Centre

# Issues of a Continuous Readout System

- The task of identifying events and inspecting them for signatures of interesting physics (rare observables) is shifted from electronic hardware to software.
- Precise understanding the timing behaviour of the detector and their readout electronics is mandatory!
- Detector noise and other backgrounds need to be understood and treated as well.




Simulated STS data (w/o thermal shielding),
Au+Au @ 10A GeV, beam rate $10^9$/s, event rate $10^7$/s

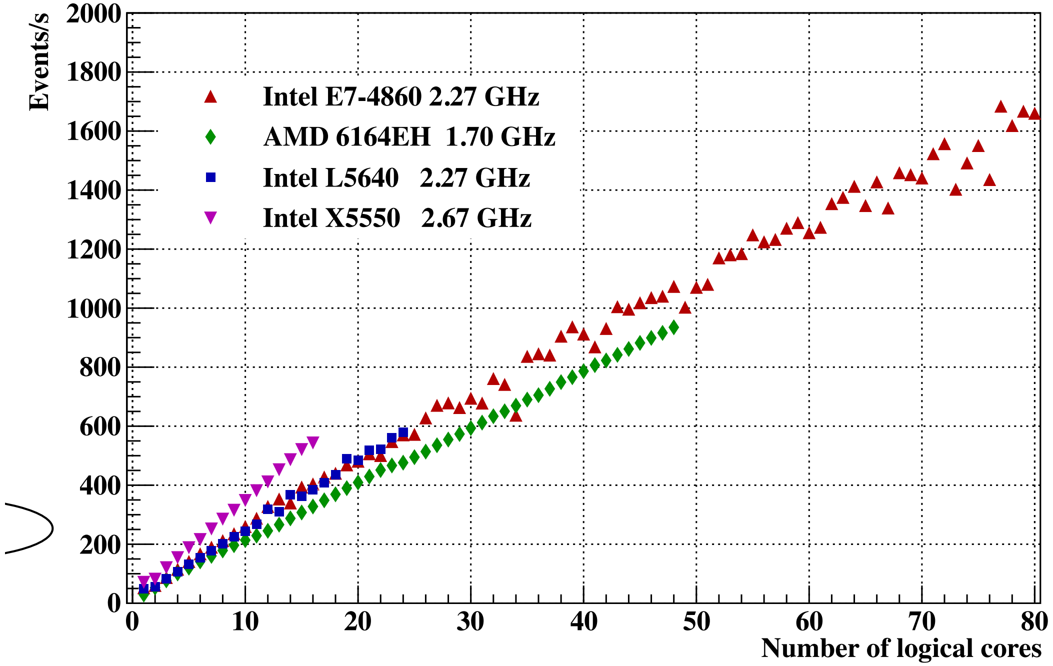
# Data Inspection in Real-Time

- For a given event / data rate, the speed of the algorithms determines the required size of the online compute farm.
- For a given financial budget / size of the online farm, the speed of the algorithms determine the physics output of the experiment.
- High-performance online software is a pre-requisite for the successful operation of CBM.
  - Make optimal use of available parallel computer architectures: many-core, GPU, accelerators
  - Be flexible to upcoming new architectures
- Parallelism is the key word
  - Data-level parallelism: one timeslice per compute node (if GPUs are used: one timeslice per GPU)
  - Task-level and data-level parallelism within time slice
  - Simplest approach: one process per timeslice
  - Multi-threading on CPU within timeslice using OpenMP
  - Mult-threading on GPU using the XPU framework

# The Key: Finding Tracks Online

100 AuAu minimum bias events at 10 AGeV

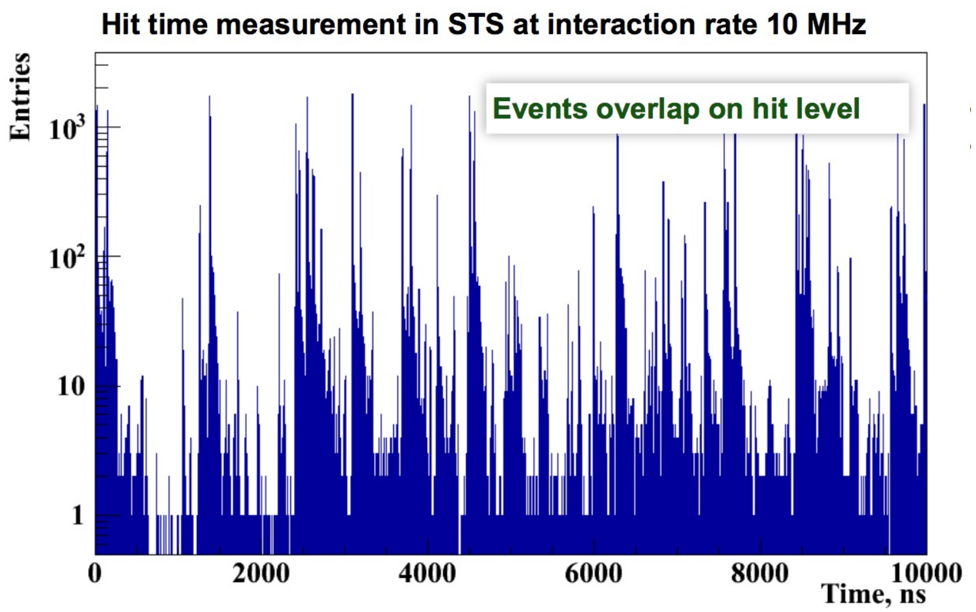
| Efficiency, % | 3D | 4D 0.1MHz | 4D 1MHz | 4D 10MHz |
|---|---|---|---|---|
| All tracks | 92.5 % | 93.8 % | 93.5 % | 91.7 % |
| Primary high-p | 98.3 % | 98.1 % | 97.9 % | 96.2 % |
| Primary low-p | 93.9 % | 95.4 % | 95.5 % | 94.3 % |
| Secondary high-p | 90.8 % | 94.6 % | 93.5 % | 90.2 % |
| Secondary low-p | 62.2 % | 68.5 % | 67.6 % | 64.3 % |
| Clone level | 0.6 % | 0.6 % | 0.6 % | 0.6 % |
| Ghost level | 1.8 % | 0.6 % | 0.6 % | 0.6 % |
| True hits per track | 92% | 93 % | 93 % | 93% |
| Hits per MC track | 7.0 | 7.0 | 6.97 | 6.70 |



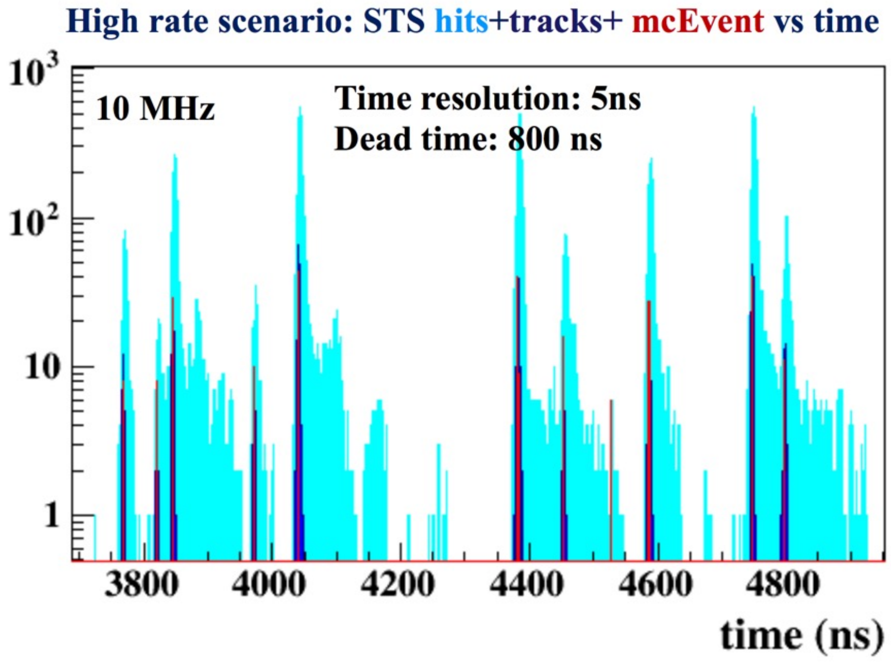

High efficiency for primary tracks
Rate effects become visible above 1 MHz interaction rate

Good scaling behaviour: well suited for many-core systems

# Resolving event pile-up

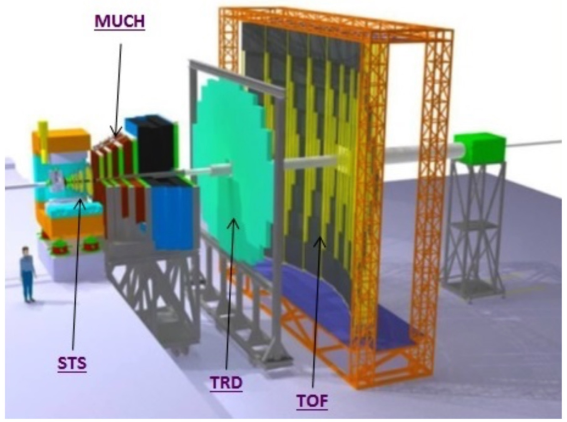


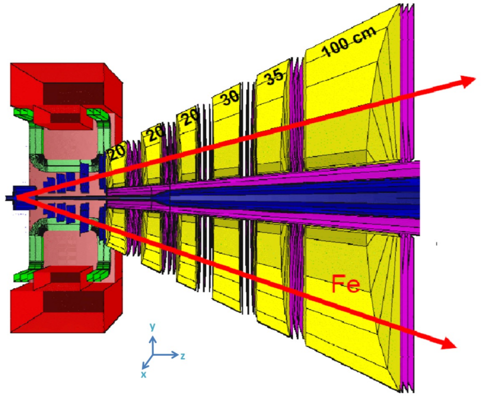

track finder




- Reliable event resolution requires track reconstruction.
- Tracking operates on stream of detector hits, without event association.
- Needs a 4d track model (x, y, z, t, q/p); sophisticated algorithm.
- Reduces the temporal event extension from ~40 ns to ~3 ns.
- Runs in software - no way to have it in hardware.

# Not all triggers are complicated...




- The muon system has a very simple measurement principle.
- Trigger requires two tracks to pass the absorbers at the same time.
- Simple algorithm; implementation on various platforms (CPU / GPU) investigated at VECC.

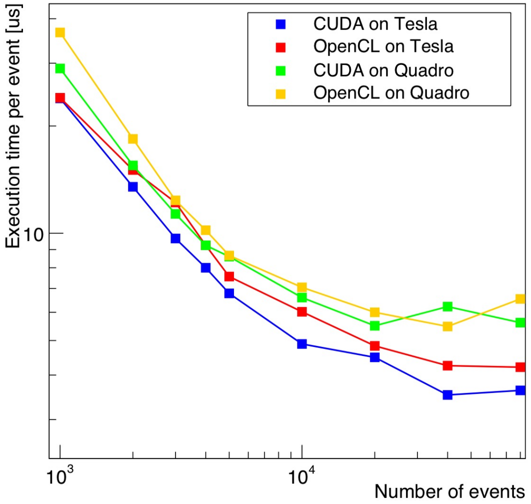



Table 1: Results for the event selection algorithm on the Tesla GPU

| # Events | # blocks | # threads | GPU Time (ms) | CPU-GPU Transfer Time(ms) | CPU Time (ms) | Speed-Up (CPU time/GPU time) |
|---|---|---|---|---|---|---|
| 1000 | 32 | 32 | 23.9 | 1.2 | 16.38 | 0.69 |
| 2000 | 64 | 32 | 27 | 2.9 | 32.40 | 1.20 |
| 3000 | 64 | 64 | 29 | 6.7 | 48.60 | 1.68 |
| 4000 | 64 | 64 | 32 | 9.4 | 64.83 | 2.03 |
| 5000 | 128 | 64 | 33.9 | 10.1 | 81.16 | 2.39 |
| 10000 | 128 | 128 | 48.9 | 12 | 161.67 | 3.31 |
| 20000 | 256 | 128 | 89.7 | 14.5 | 320.24 | 3.57 |
| 40000 | 512 | 128 | 140.7 | 19.7 | 640.25 | 4.55 |
| 80000 | 1024 | 128 | 289.8 | 28.3 | 1280.39 | 4.42 |

V. Singhal et al., Comp. Phys. Comm. 252 (2020) 107190

# Selectivity and operation modes

Not all observables require or allow a highly selective trigger:

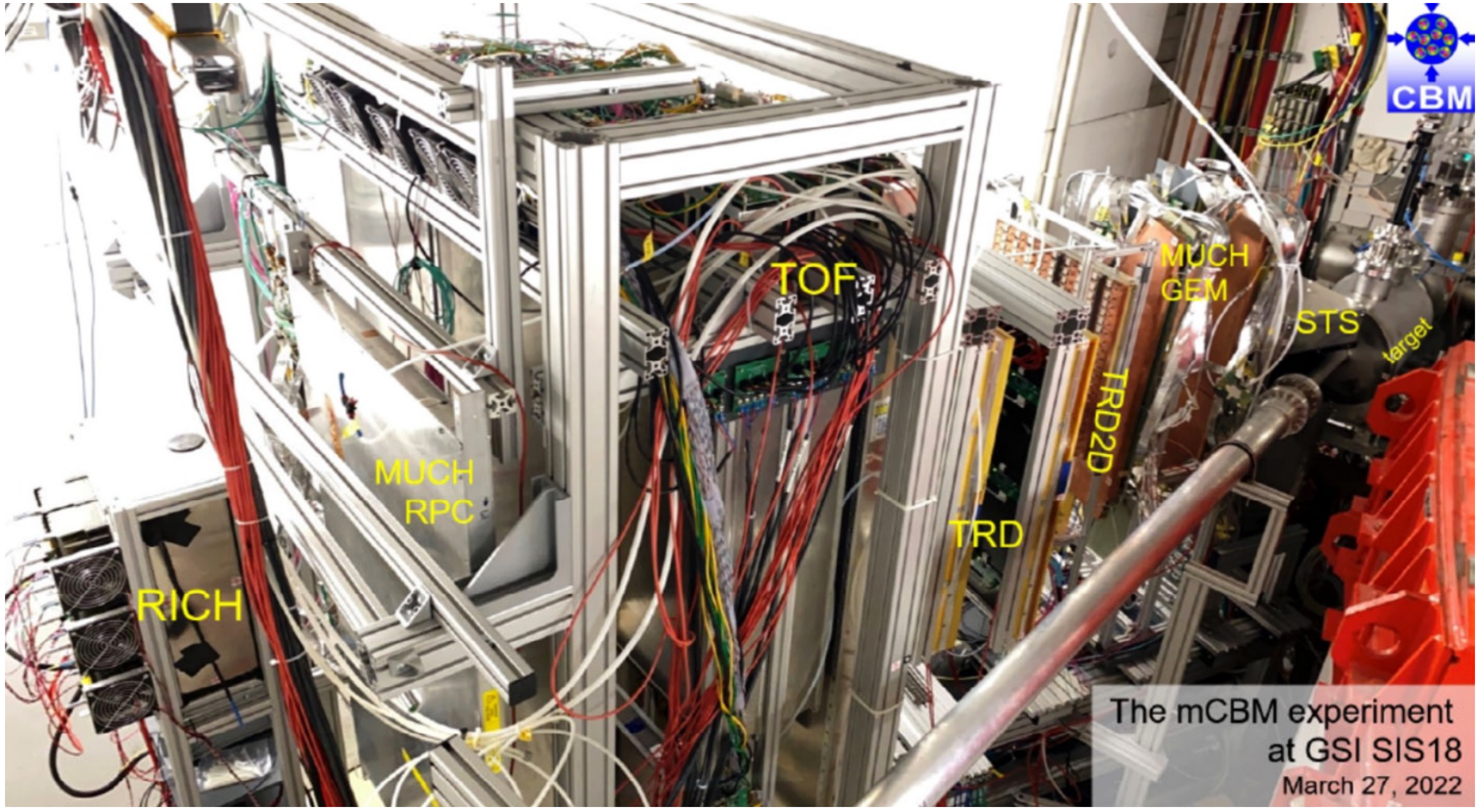
- Some have no trigger signature: fluctuations, low-mass electron pairs
- Some are not rare: all bulk observables
- Some allow lesser selectivity: high-mass electron pairs

Recordable min. bias rate (w/o online selection): $10^5$ events/s

Of course, CBM can also be operated at lower rates!

N.b.: High-rate operation allows to take selective (triggered) data together with min. bias (downscaled to saturate the bandwidth).

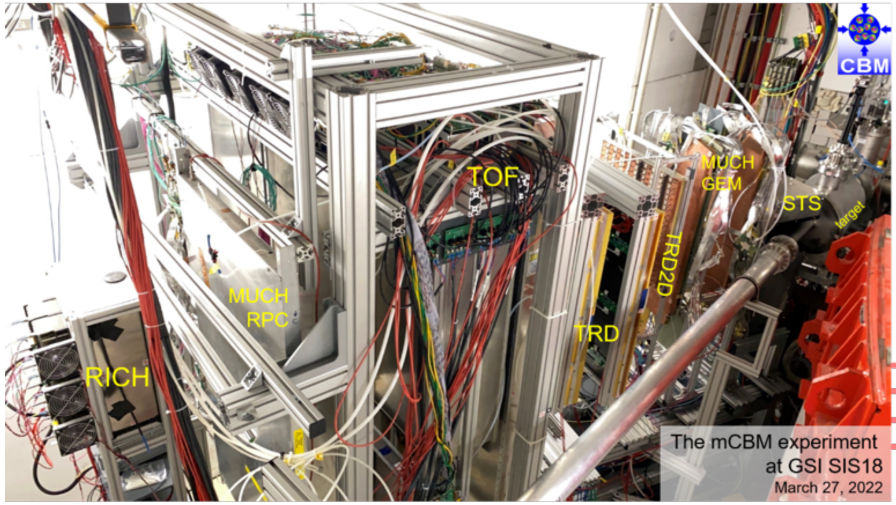
# Our testbed: mCBM




- A full-system sandbox with detector prototypes / pre-series components
- Study detectors after integration
- Verify free-streaming read-out and data transport and online data processing
- Benchmark online data processing
- Gain operational experience

- First runs 2021, 2022 (w/o online processing, all data to disk)
- Next benchmark run: 2024, applying online data selection

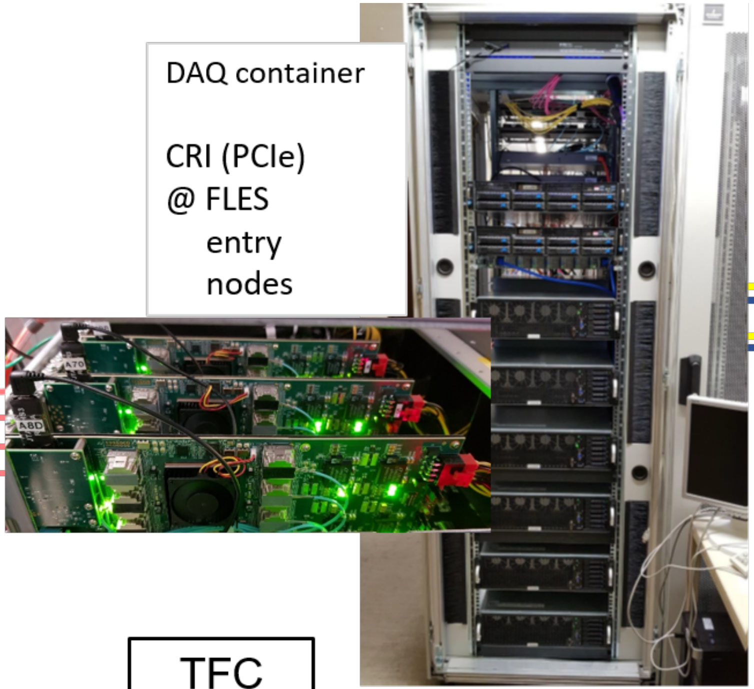
# mCBM Data Transport System




optical
fibers
50m

DAQ container

CRI (PCIe)
@ FLES
entry
nodes



optical
fibers
300 m


FLES
processing
nodes

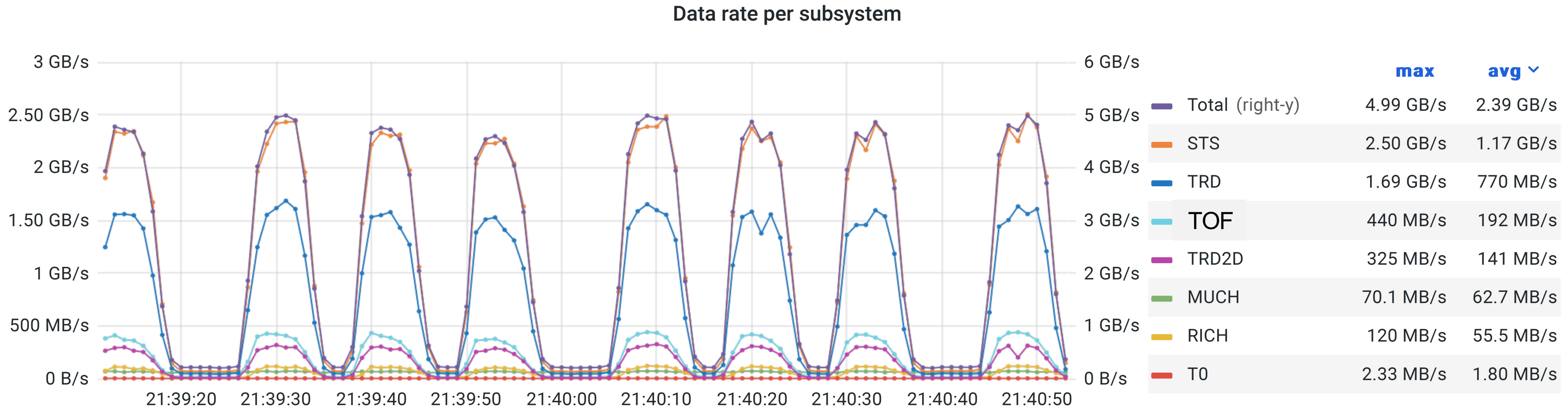
triggerless-streaming FEE **assigning time stamps to hits** → 1 m Copper → GBTx → 50 m optical → CRI FPGA **µSlice building** FLES entry nodes → 300 m optical InfiniBand → Green IT Cube

TFC (CRI based) → CRI FPGA

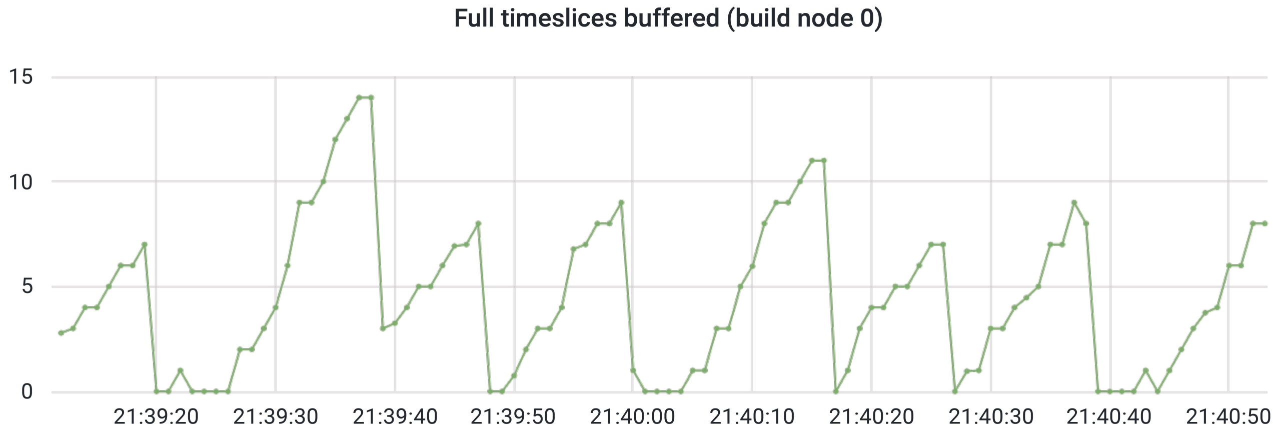
Green IT Cube:
- Timeslice Building
- Event Reconstruction
- Event Selection
- Archiving

# mCBM: Data Path Performance

FLES input



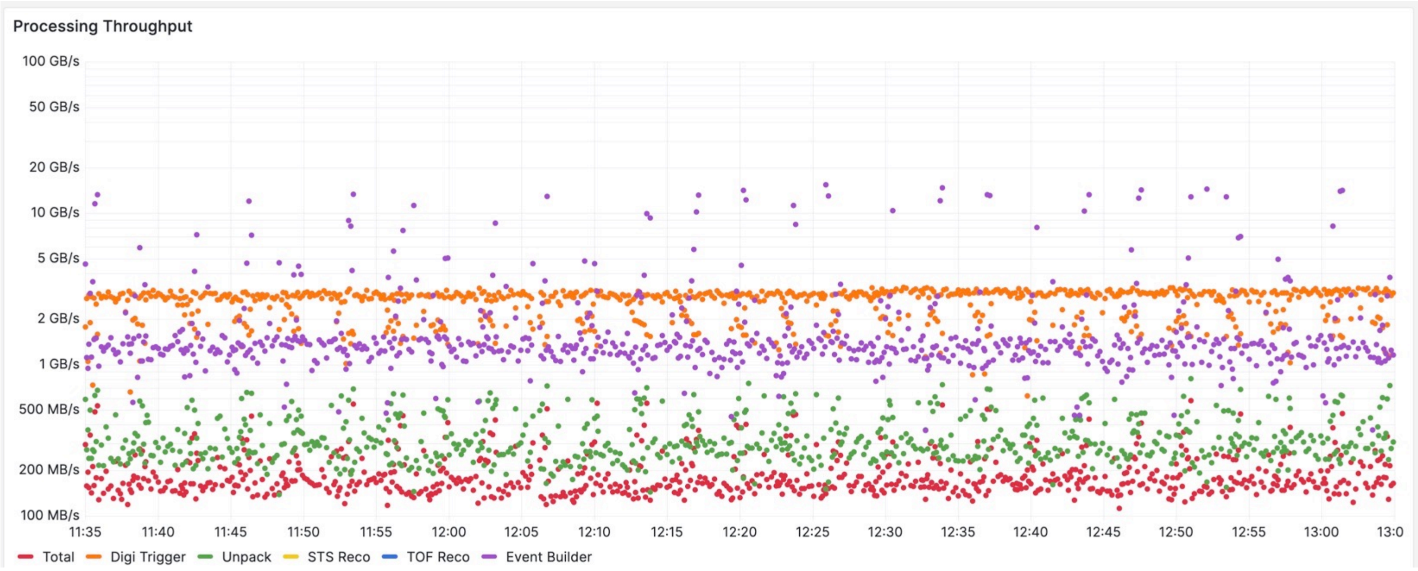

FLES output




Run 2448
June 16, 2022
Au + Au, T = 1.23 AGeV
av. collision rate: 300 - 400kHz
av. data rate 2.4 GB/s to disc

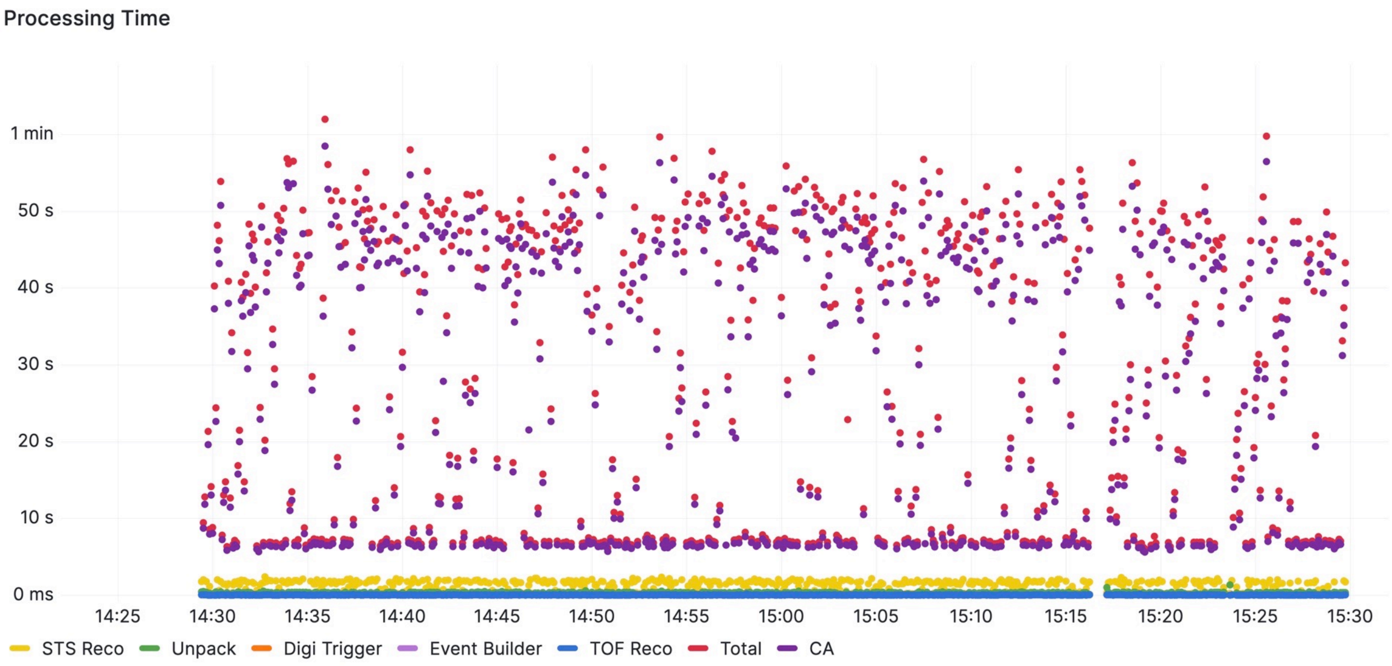
# mCBM: Data Challenges

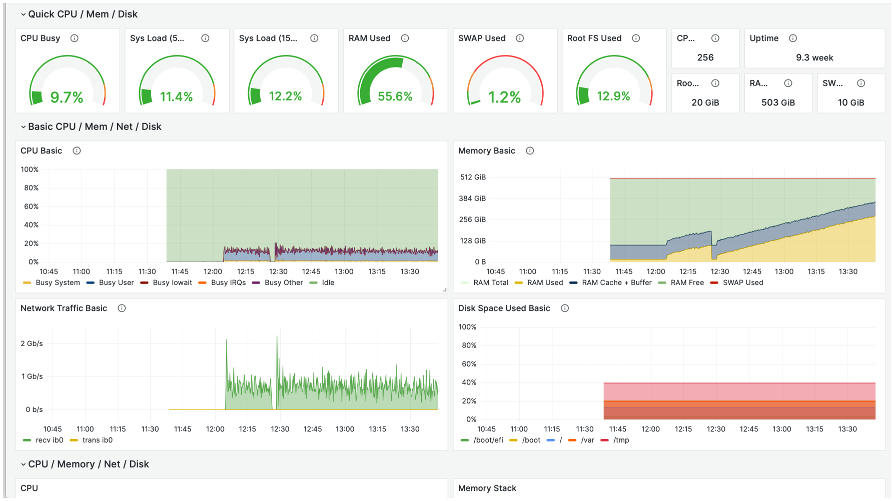
- Until 2022: No online data processing; dumping raw data (full timeslices) to disk
- Development of online processing chain with re-play of archived raw data (emulates DAQ)
- Allows to study and benchmark timeslice building and online data processing prototypes
- Gain operational experience - interplay of owned resources (entry nodes) with shared resources (compute nodes).




DC 3 (December 2023)
Unpacking + event building
+ local reconstruction

# mCBM: Inclusion of Track Reconstruction

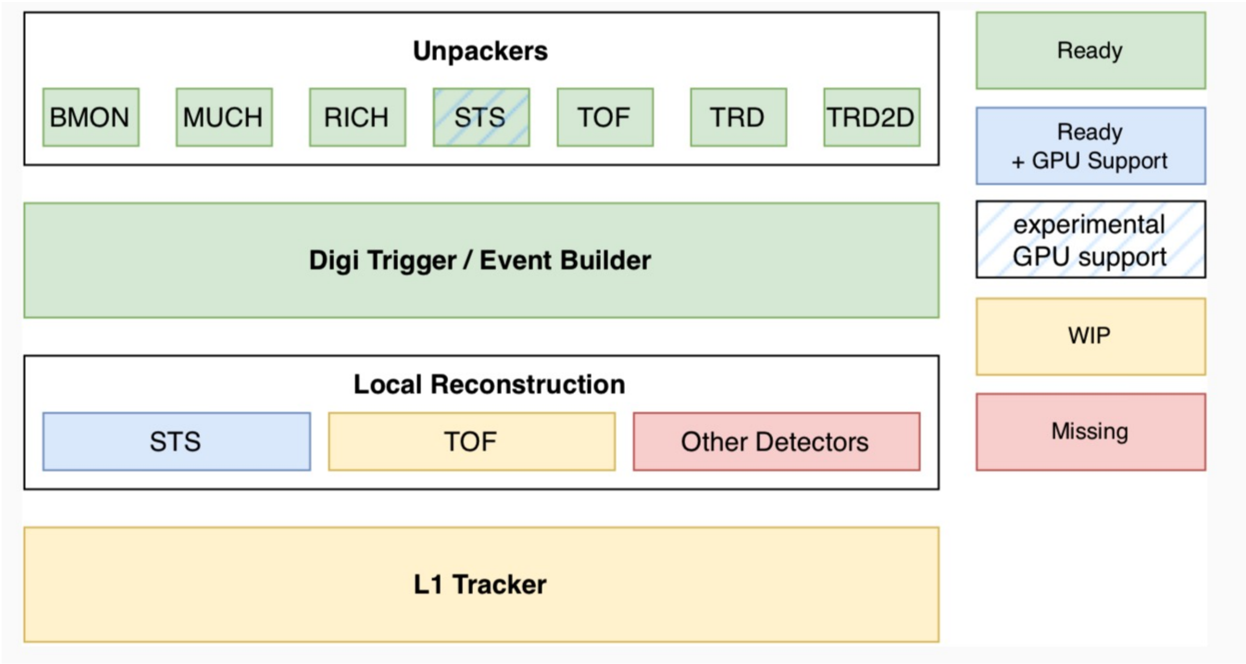


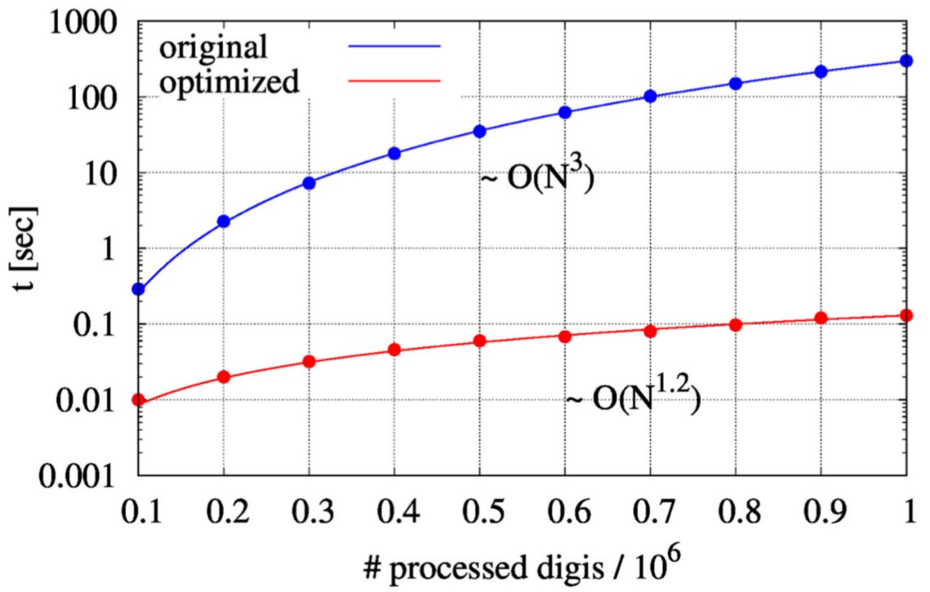




DC 4 (February 2024)
Unpacking + event building + local reconstruction + track reconstruction (not yet multi-threaded)

# The state of our art...

- Complete event reconstruction developed for offline purposes.
- Ongoing: adapting / porting / optimising for online usage.
- Development and testing on real data (mCBM) and simulated data (full CBM)

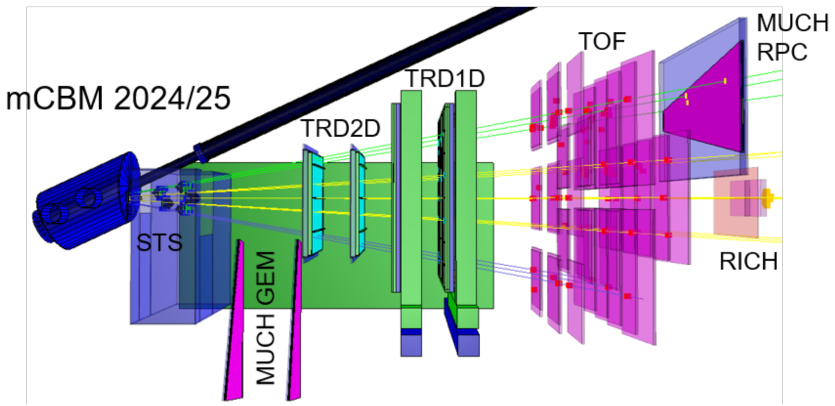


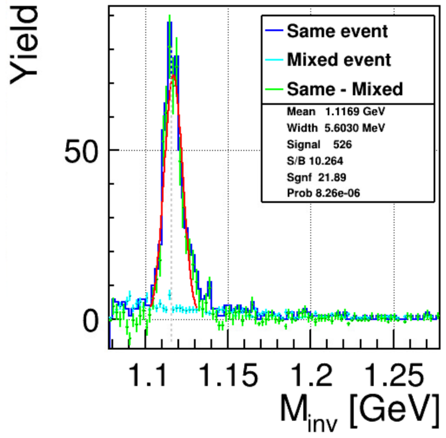




Optimisation example: STS hit finder

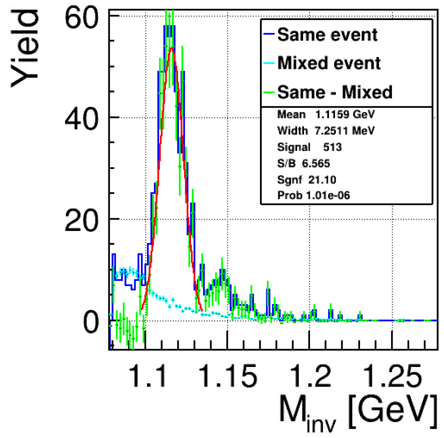
# mCBM: Achievements and Next Steps

- ✓ Multi-core parallelization (OpenMP) of all reconstruction steps (STS, TOF, tracking).
- ✓ STS unpacking and reconstruction ported to GPU and tested on VIRGO nodes.
- ✓ Commissioning beam time March 2024: applied online processing during data taking; minimum-bias trigger based on digi multiplicity; data reduction by a factor of four.
- • Benchmark beam time May 2024 (Ni+Ni): Application of full online reconstruction and trigger on displaced vertices (Lambda).



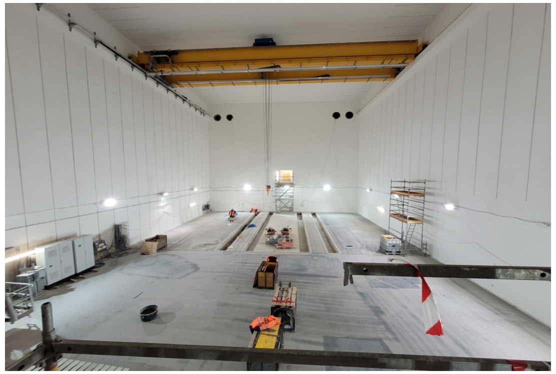



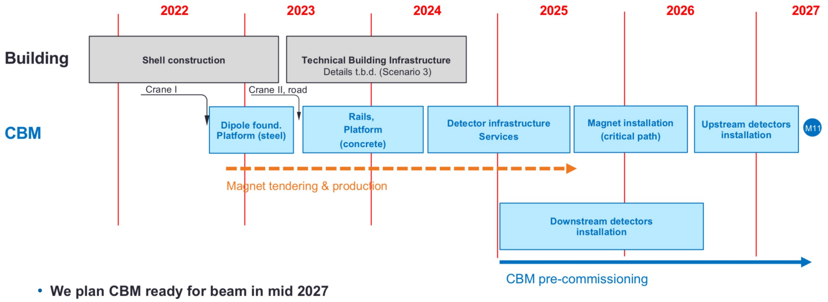

mCBM simulation
100 M events
$10^5$ events/s



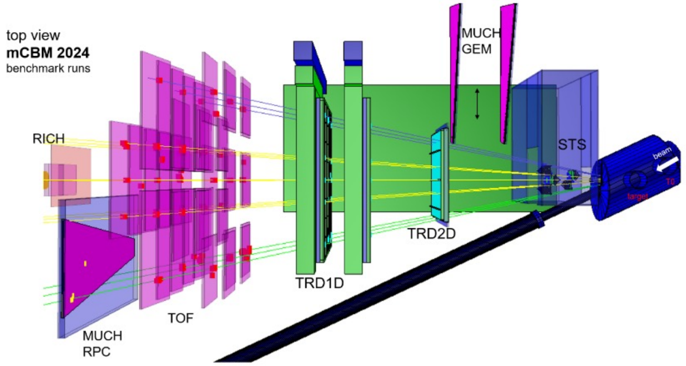

mCBM offline data analysis (preliminary)
Run 2391 (May 2022)
Av. rate 5 * $10^5$ events/s
$10^9$ events

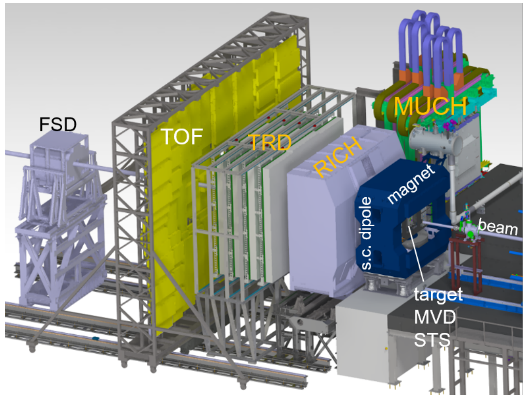
# Summary: From mCBM to CBM






- We plan CBM ready for beam in mid 2027




A long and winding road - but we are some way up already.