

# Level-1 Trigger Online Software

---

Simone Bologna<sup>1</sup>, Tom Williams<sup>2</sup>

<sup>1</sup> University of Bristol

<sup>2</sup> Rutherford Appleton Laboratory

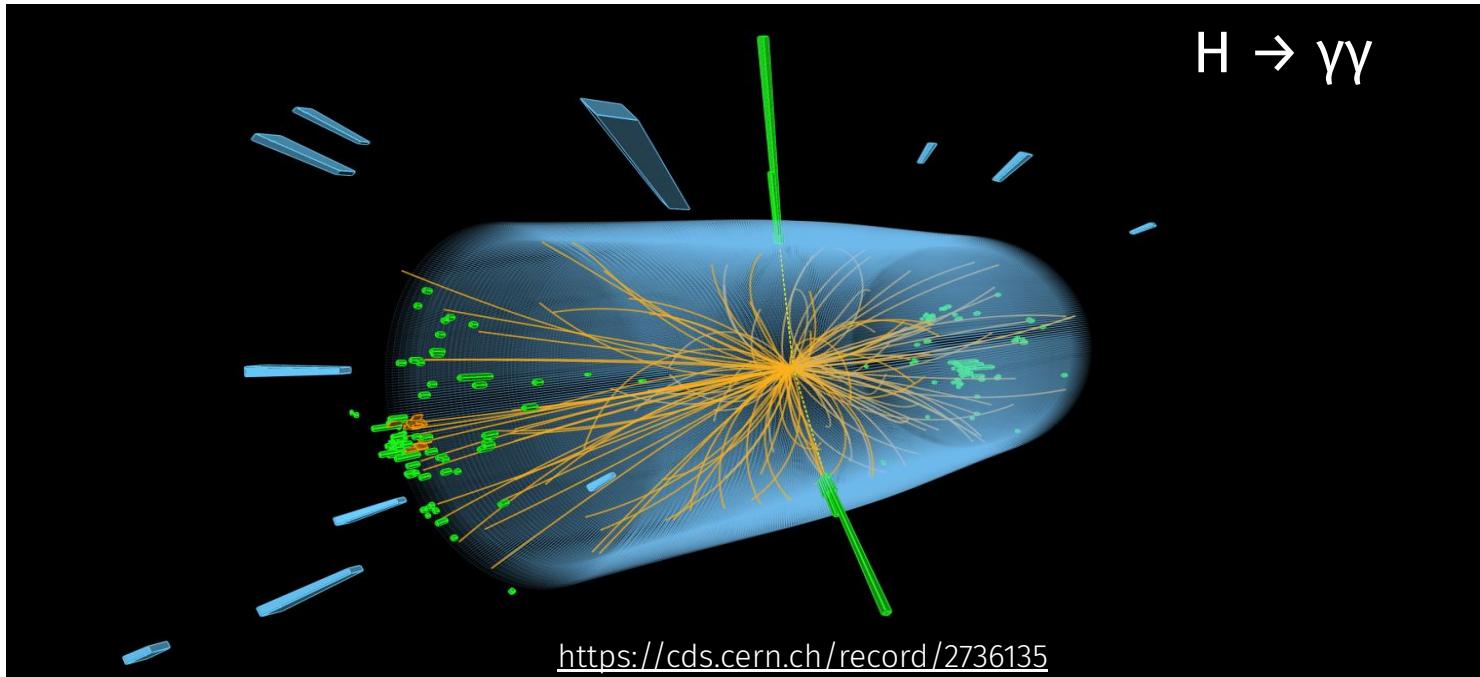
CMS induction course, 21/09/2021

# Outline

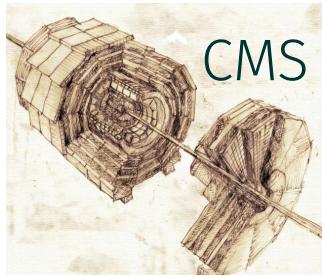
- CMS level-1 trigger
- Online software tools
  - Control
  - Configuration management
  - Monitoring
- A glance into the future

# Goal of CMS

*Record results of LHC collisions, to understand fundamental laws of physics*



# CMS data flow

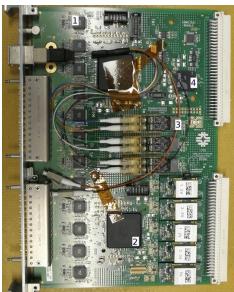


CMS

40 MHz

40-80 TB/s

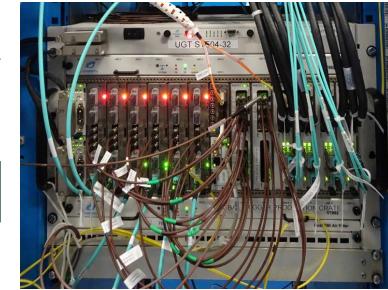
Readout



40 MHz

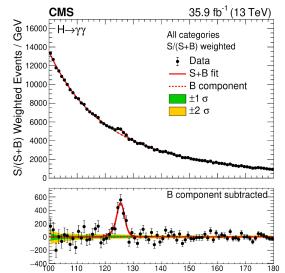
~ 1 TB/s

Accept/reject



Level-1 trigger

100 kHz  
100 GB/s



300 nHz

100 papers/y



1 kHz

~ 1-2 GB/s



Event builder &  
High-level trigger

# CMS data flow



40 MHz

40-80 TB/s

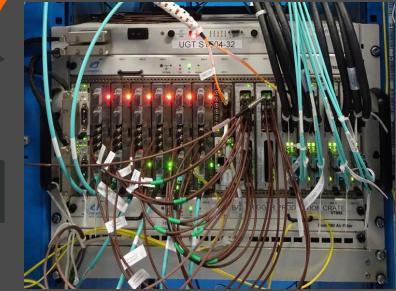
Readout



40 MHz

~ 1 TB/s

Accept/reject



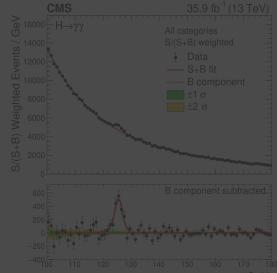
Level-1 trigger

Focussing on level-1 trigger in  
this talk

700 kHz

100 GB/s

Event builder &  
High-level trigger



300 nHz

100 papers/y



1 kHz

~ 1-2 GB/s



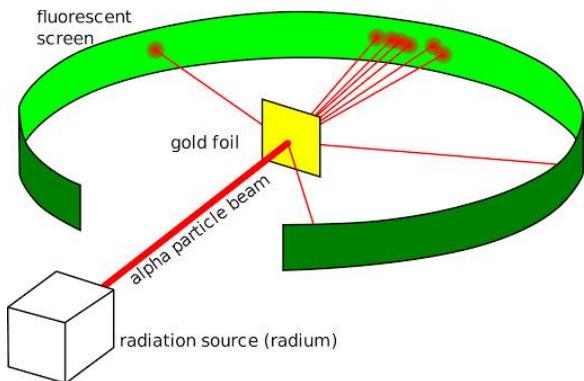
# Triggers & online SW

- Trigger recognises interesting activity in the detector
- Online Software is responsible for ...
  - Configuration
  - Control
  - Monitoring
- If an event isn't accepted by the trigger, it's **lost forever**
  - L1T online SW: Critical role in data taking – must be designed carefully to ...
    - Ensure reliable, robust behaviour
    - Allow on-call experts to rapidly resolve problems
    - Maintain high uptime  
*(in addition to other requirements)*
  - Achieve the above despite complexity of the level-1 trigger (& upstream systems)

*First, for contrast: Brief interlude on some trigger systems from previous generations of particle/nuclear experiments ...*

# Triggers

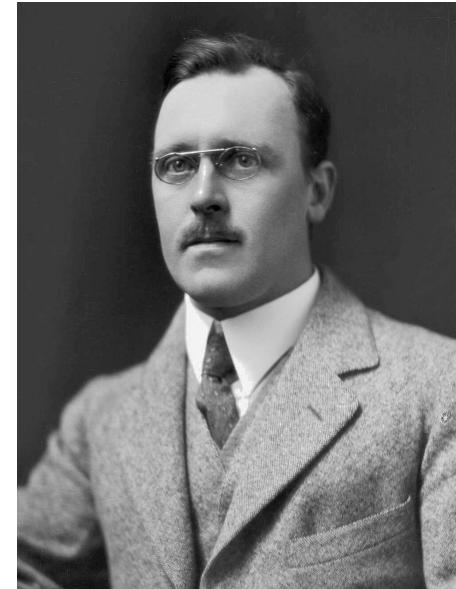
- Rutherford scattering expt: Geiger and Marsden manually counted alpha particles scattered at given angles
- Acted as organic TDAQ (trigger & data acquisition) system
  - Low event rate



Picture of an early TDAQ system



Hans Geiger  
Post-doc



Ernest Marsden  
PhD student

# Triggers

- Human brains are not designed for boring, repetitive tasks
  - Machines do not get bored!
    - And can process readings from detectors quite a bit faster than us
  - Idea already considered in 1919
    - “visual or audible methods of counting are quite trying on the nerves ... A self-recording device would therefore be an obvious improvement”
      - Kovarik on automatic recording device for the Rutherford scattering experiment
- Phys. Rev. 13, 272, 1st April 1919

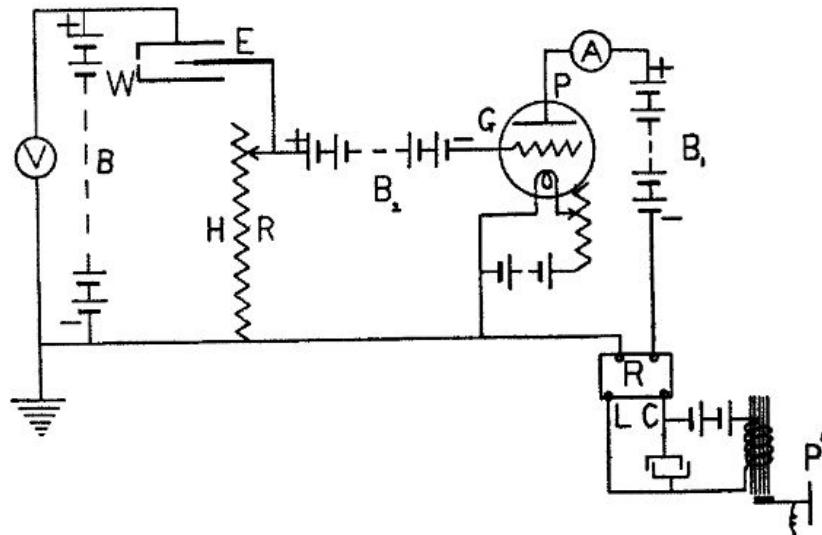


Fig. 1.

Recording system proposed by Kovarik

Phys. Rev. 13, 272, 1st April 1919

# Automated triggers

- More sophisticated TDAQ systems developed as experiment goals became more ambitious
- Early systems: Relatively simple, based on coincidences and thresholds of analogue signals
- UA1 & 2 @ SppS, 1981 — first example of experiment with broad physics program and generic triggers
  - 250kHz → ~ 5Hz, three level trigger
- Complex trigger → sophisticated control & monitoring software

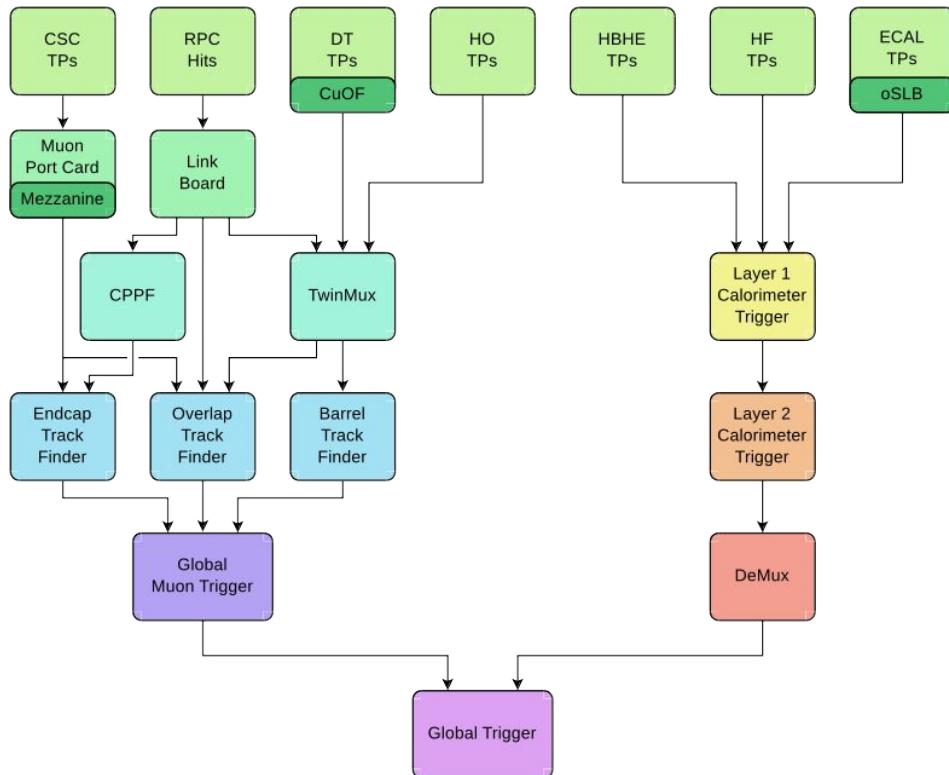
NIM crate



Signal  
discriminator

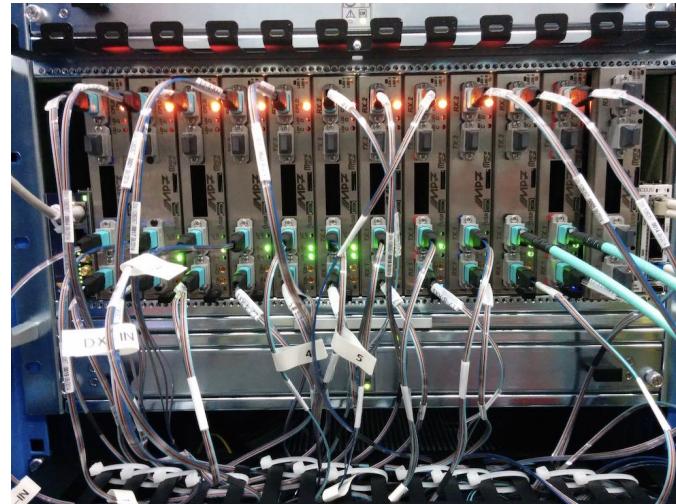
# CMS level-1 trigger

- ~ 150 boards
- ~ 3000 optical links
  - Up to 10Gbps
- Complex system
  - Input from 5 detector systems
    - 9 subsystems → different data-processing algorithms (calo clusters, local  $\mu$  reco ...)
  - 5 different board designs



# CMS level-1 trigger

- Configuration settings
  - Must be easy to edit and manage
- Status continually monitored
  - Intelligent analysis
    - Concise error summaries & useful suggestions
  - Rich user interfaces
  - ✓ **Rapidly solve problems**



# CMS level-1 trigger

- Configuration settings
  - Must be easy to edit and manage
- Status continually monitored
  - Intelligent analysis
    - Concise error summaries & useful suggestions
  - Rich user interfaces
  - ✓ **Rapidly solve problems**
- Online software is crucial
  - Design & implementation far from simple

## Example scenario: Anomalous signals from ECAL electronics

- CMS running stably
- In the middle of the night, L1T rate suddenly increases to 40MHz  
⇒ *can't record data until problem fixed*
- Shift crew phone on-call expert
- To solve this, they need to quickly ...
  - Identify which of the several hundred trigger paths have high rates & which don't (e.g. electron paths)
  - Optical links: Check whether there are any CRC errors, and check for anomalous data values

*Losing data for every second of delay*

# Level-1 online software: Key components

- Configuration and control
  - SWATCH & SWATCH cell
  - Level-1 configuration editor
- Monitoring
  - L1page
  - SWATCH & SWATCH cell
  - Data Quality Monitoring (DQM)
    - Not part of online software itself, but essential for operating trigger

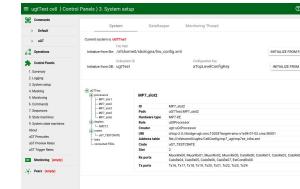
# Where are the tools?

- Online software, together with all software required to run CMS, is installed in a separate network
  - Network not directly reachable from outside
- Internal resources accessible via tunnel
  - Much safer and easier to protect

P5  
network

CERN  
network

External  
network



Online  
software

cmsusr

Lxplus

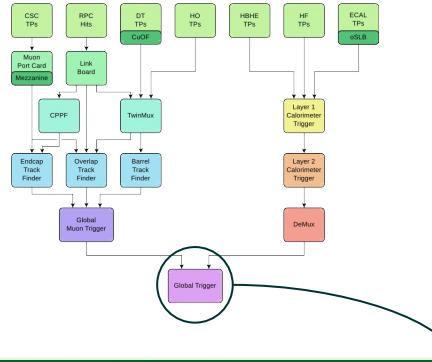


You

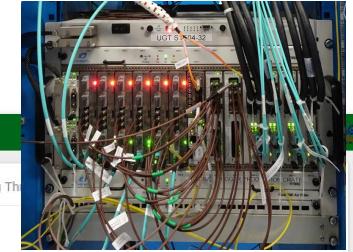
# Configuration & control

# SWATCH & SWATCH cell

- SWATCH = SoftWare for Automating the conTrol of Common Hardware
- Centrally-developed library that provides unified software interface to control, configure and monitor trigger hardware
- Each subsystem is managed by a SWATCH cell



crate setup



ugtTest cell > Control Panels > 3. System setup

Commands

- > Default
- > uGT

Operations

Control Panels

- 1. Summary
- 2. Logging
- 3. System setup
- 4. Masking
- 5. Monitoring
- 6. Commands
- 7. Sequences
- 8. State machines
- 9. System state machines
- About
- uGT Prescales
- uGT Preview Rates
- uGT Trigger Rates

Monitoring (empty)

Peers (empty)

System GateKeeper Monitoring Th...

Current system is 'uGTTest'  
Initialize from file: /nfshome0/sbologna/hw\_config.xml  
Subsystem ID  
Initialize from DB: ugtTest Configuration Key aTopLevelConfigKey

INITIALIZE FROM FILE  
INITIALIZE FROM DB

uGTTest

- processors
  - MP7\_slot1
  - MP7\_slot2
  - MP7\_slot3
  - MP7\_slot4
  - MP7\_slot5
  - MP7\_slot6
- daqptc
  - AMC13
- crates
  - UGT\_TESTCRATE
- links
  - connected FEDs

MP7\_slot2

ID	MP7_slot2
Path	uGTTest.MP7_slot2
Hardware type	MP7-XE
Role	uGProcessor
Creator	ugt:uGTProcessor
URI	chtc-2.0://bridge-ugt.cms:10203?target=amc-s1e04-07-02.cms:5001
Address table	file:///nfshome0/ugtis/CellConfig/mp7_ugt/mp7xe_infra.xml
Crates	uGT_TESTCRATE
Slot	2
Rx ports	MuonRx00, MuonRx01, MuonRx02, MuonRx03, CaloRx00, CaloRx01, CaloRx02, CaloRx03, CaloRx04, CaloRx05, CaloRx06, CaloRx07, ExtCondRx00
Tx ports	Tx16, Tx17, Tx18, Tx19, Tx20, Tx21, Tx22, Tx23, Tx24

# SWATCH

- Individual subsystems can be configured from a browser via the SWATCH cell web panels
  - E.g. for manual tests/debugging
- Each SWATCH cell can also be controlled by a central cell, that coordinates the entire trigger
  - I.e. 1 central cell instance for whole L1T
  - Configuration settings taken from central database & edited via L1CE

Subsystem configuration page

The screenshot shows a web-based interface for managing system state machines. On the left, a sidebar lists various control panels and monitoring options. The main area displays a state machine named 'runControl' in 'Synchronized' mode. It shows three steps of a transition process:

- Step 1:** Device: AMC13, Transition: clockSetup, Progress: 100%, Time: 0.007289s.
- Step 2:** Devices: MP7\_slot1, MP7\_slot2, MP7\_slot3, MP7\_slot5, MP7\_slot6, all in setup state with 100% progress and times between 1.181s and 1.176s.
- Step 3:** Device: AMC13, Transition: cfgDaq, Progress: 100%, Time: 0.002929s.

On the right, a legend defines states and transitions:

State	Transition → End state
Halted	coldReset → Halted ⓘ setup → Synchronized ⓘ
Error	configure → Configured ⓘ
Synchronized	align → Aligned ⓘ start → Running ⓘ stop → Configured ⓘ
Configured	start → Running ⓘ stop → Configured ⓘ
Aligned	pause → Paused ⓘ stop → Configured ⓘ
Running	resume → Running ⓘ stop → Configured ⓘ
Paused	resume → Running ⓘ stop → Configured ⓘ

# Level-1 Configuration Editor

- Configuration settings are stored in a database, with specific configuration chosen by selecting key from shift crew web interfaces before the start of a data-taking run
  - Top-level key made of subkeys (2 per subsystem), each containing XML

Subsystem selection      Key selection      Sub key      Sub key content

The screenshot illustrates the workflow for selecting a configuration key:

- Subsystem selection:** The "Subsystem" dropdown is set to "uGT".
- Key selection:** The "Configuration" dropdown is set to "testing2020" and "v17".
- Sub key:** The "UGT\_KEYS" panel shows the selected key "testing2020 - v17" and its subkeys: "HW" and "ugt\_hw\_base2018/v2".
- Sub key content:** The "ugt\_hw\_base2018/v2" subkey is expanded, displaying its XML content.

**Browsing key tree**

```
<system id="uGT">
  <creator>ugt:uGtSystem</creator>
  <crates>
    <create id="UGT_CRATE">
      <description>Vadatech Crate for uGT</description>
      <location>Point5, SIE04-32</location>
    </create>
  </crates>
  <processors>
    <processor id="MP7_slot1">
      <creator>ugt:uGtProcessor</creator>
      <hw-type>MP7-XE</hw-type>
      <role>uGtProcessor</role>
      <uri>http://bridge-ugt.cms:10203?target=amc-s1e04-32-01.cms:50001</uri>
      <address-table>file:///nfshome0/ugts/CellConfig/mp7_ugt/mp7xe_infra.xml</address-table>
      <crate>UGT_CRATE</crate>
    </processor>
    <processor id="MP7_slot2">
      <creator>ugt:uGtProcessor</creator>
      <hw-type>MP7-XE</hw-type>
      <role>uGtProcessor</role>
      <uri>http://bridge-ugt.cms:10203?target=amc-s1e04-32-02.cms:50001</uri>
      <address-table>file:///nfshome0/ugts/CellConfig/mp7_ugt/mp7xe_infra.xml</address-table>
      <crate>UGT_CRATE</crate>
    </processor>
    <processor id="MP7_slot3">
      <creator>ugt:uGtProcessor</creator>
      <hw-type>MP7-XE</hw-type>
```

# Level-1 Configuration Editor

- Configuration settings are stored in a database, with specific configuration chosen by selecting key from shift crew web interfaces before the start of a data-taking run
  - Top-level key made of subkeys (2 per subsystem), each containing XML

Subsystem selection      Key selection      Sub key      Sub key content

The screenshot illustrates the workflow for selecting a configuration key:

- Subsystem selection:** The "Subsystem" dropdown is set to "uGT".
- Key selection:** The "Configuration" dropdown is set to "testing2020" and "v17".
- Sub key:** In the "UGT\_KEYS" panel, the "testing2020 - v17" key is selected. Its subkey "ugt\_hw\_base2018/v2" is highlighted.
- Sub key content:** The XML content of the selected subkey is displayed, showing details like system ID, creator, crates, and processors.

**Browsing key tree**

**Mission-critical GUI — must be available 24/7, easy to use (minimise risk of human error), and protect against mistakes (e.g. verify L1 vs HLT consistency)**

```
<system id="uGT">
  <creator>ugt::uGtSystem</creator>
  <crates>
    <create id="UGT_CRATE">
      <description>Vadatech Crate for uGT</description>
      <location>Point5, SIE84-32</location>
    </create>
  </crates>
  <processors>
    <processor id="MP7_slot1">
      <creator>ugt::uGtProcessor</creator>
      <hw-type>MP7-XE</hw-type>
      <role>uGtProcessor</role>
      <uri>http://bridge-ugt.cms:10203?target=amc-s1e04-32-01.cms:50001</uri>
      <address-table>file:///nfs/home0/ugts/CellConfig/mp7_ugt/mp7_xe_infra.xml</address-table>
      <crate>UGT_CRATE</crate>
    </processor>
  </processors>
  <slot1>
    <processor id="MP7_slot1">
      <creator>ugt::uGtProcessor</creator>
      <hw-type>MP7-XE</hw-type>
      <role>uGtProcessor</role>
      <uri>http://bridge-ugt.cms:10203?target=amc-s1e04-32-01.cms:50001</uri>
      <address-table>file:///nfs/home0/ugts/CellConfig/mp7_ugt/mp7_xe_infra.xml</address-table>
      <crate>UGT_CRATE</crate>
    </processor>
  </slot1>

```

# Monitoring

# Level-1 page

- Main dashboard for L1T status

The screenshot shows the main dashboard for the L1T status. At the top, it displays the current key "l1\_hit\_cosmics2017/v150" and a red circular icon with a checkmark indicating "Status running #294398 for 6 minutes". It also shows a connection status to the API.

The central part of the dashboard features a hierarchical status tree diagram. The root node is "Central Cell", which branches into "μGT" and "μGMT". "μGT" further branches into "Calo L2" and "BMTF". "Calo L2" branches into "ECAL" and "HOAL". "BMTF" branches into "DT" and "RPC". "μGMT" branches into "OMTF" and "EMTF". "OMTF" branches into "TWINMUX" and "CPPF". "TWINMUX" branches into "CSC". Each node has a small icon above it, such as a green triangle for Calo L2 or a red square for EMTF.

To the right of the tree, there is a "logs" section containing several error and warning messages:

- EMTF SWATCH monitoring Error components : numberOfBrokenInputLinks, totalNumberOfBrokenInputLinks
- OMTF SWATCH monitoring Error components : CSCports
- UGMT uGMT SWATCH monitoring Warning components : algo
- TWINMUX TWINMUX SWATCH monitoring Warning components : algo

Below the logs, there is a "suggestions" section with a cartoon illustration of a beach scene and the text "you deserve a coffee break".

At the bottom of the dashboard, there is a search bar with dropdown menus for "Type" (set to "Cosmics") and "Section" (set to "General"). A "Search" button and a "New" button (+) are also present.

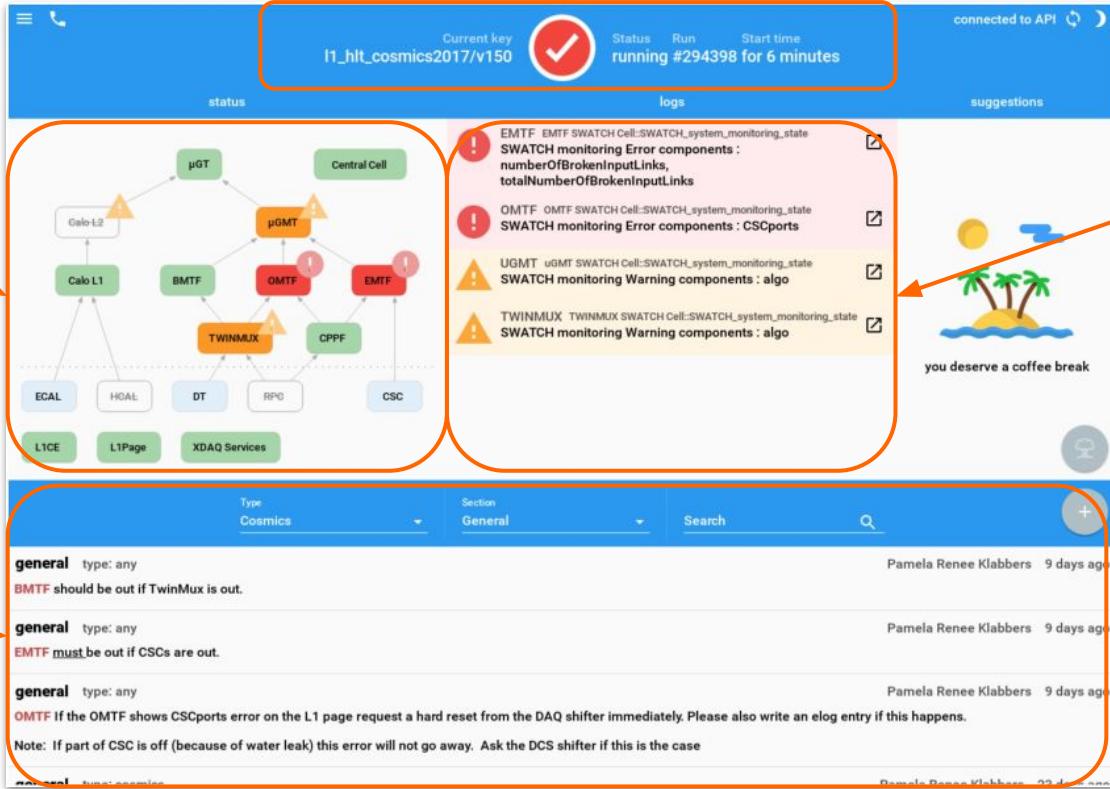
The log section at the bottom contains several entries from Pamela Renee Klabbers:

- general type: any 9 days ago
- BMTF should be out if TwinMix is out. 9 days ago
- general type: any 9 days ago
- EMTF must be out if CSCs are out. 9 days ago
- general type: any 9 days ago
- OMTF If the OMTF shows CSCports error on the L1 page request a hard reset from the DAQ shifter immediately. Please also write an elog entry if this happens. 9 days ago
- Note: If part of CSC is off (because of water leak) this error will not go away. Ask the DCS shifter if this is the case 9 days ago

# Level-1 page

- Main dashboard for L1T status

Trigger overview



Run status and configuration

Current alerts

Information logging

# SWATCH cell: Monitoring

SWATCH → generic interface for subsystem/board monitoring info.

The screenshot shows the BMTF SWATCH Cell Control Panels interface. The top navigation bar includes a menu icon, the title 'BMTF SWATCH Cell > Control Panels > 1. Summary', a help icon, and the Xdaq logo. On the left, a sidebar lists 'Commands' (Default), 'Operations', 'Control Panels' (selected), and 'Monitoring'. Below these are 'Peers', 'DB', and 'MON'. A mouse cursor hovers over the first port of the 'Top\_Crate' chart, which is highlighted with an orange circle. The main content area displays two charts: 'Top\_Crate fed-id: 1376' and 'Bottom\_Crate fed-id: 1377', both titled 'wedge01 - inputPorts'. Each chart has 13 ports numbered 1 to 13. The bars in the charts are colored red, green, and yellow. Above the charts, the status is shown as 'Status: Error' and 'State machine: runControl::Error'. At the bottom, there is a dropdown menu set to 'bmtf' and a status bar indicating 'Monitoring: Enabled'.

Hovering mouse reveals component name

bmtf System overview Status: Error State machine: runControl::Error

System Processors Object Details Ports

Top\_Crate fed-id: 1376

Bottom\_Crate fed-id: 1377

wedge01 - inputPorts

T = 5000ms last: 3s ago

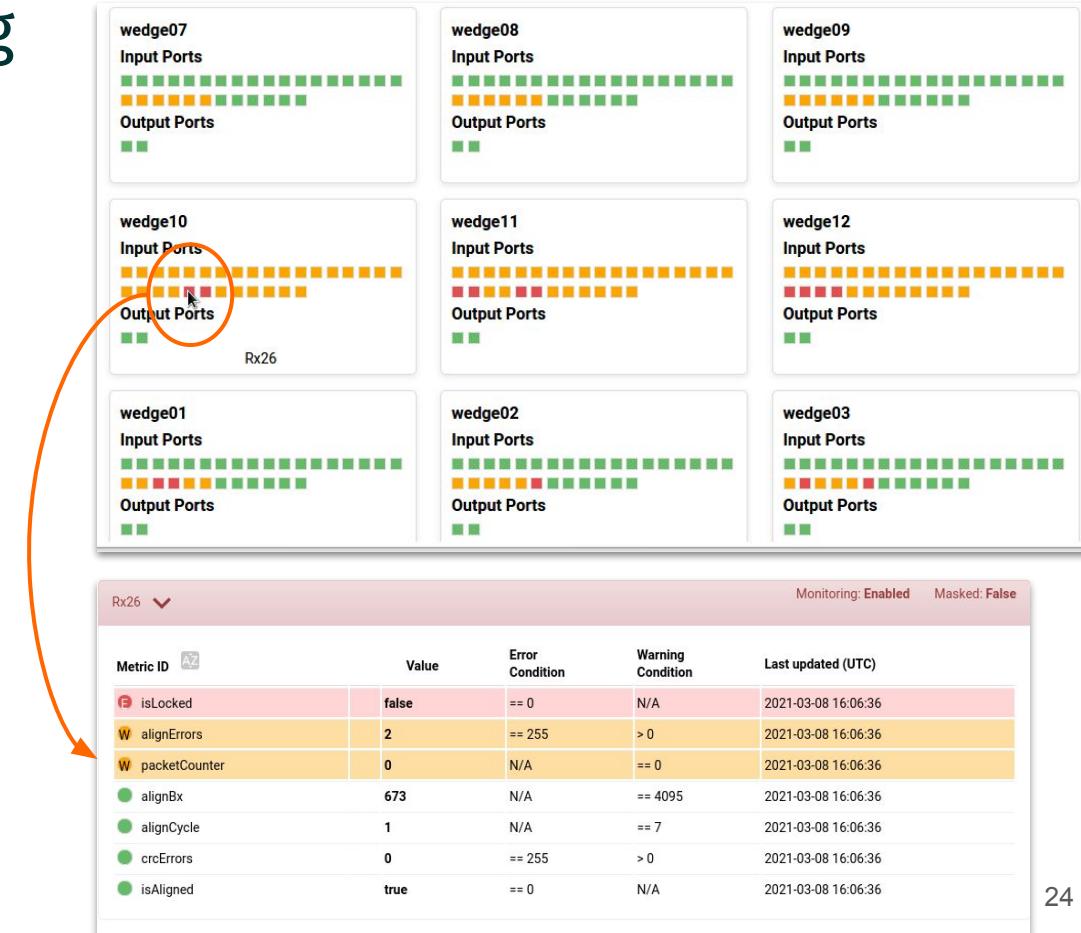
bmtf ▾ Monitoring: Enabled

# SWATCH cell: Monitoring

## I/O port dashboard

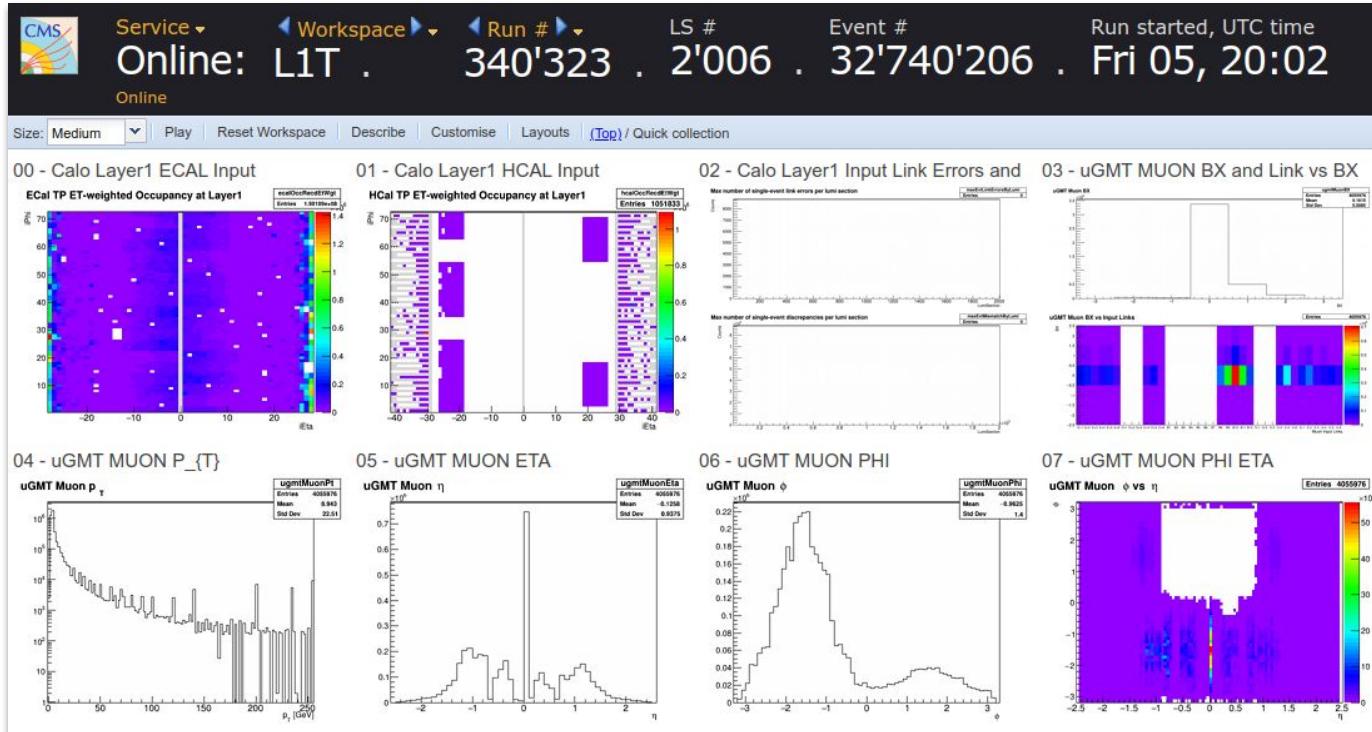
Highlights problematic components

Gives immediate access to *in-depth info*  
so that on-call experts can rapidly  
diagnose and solve the problem



# Data Quality Monitoring

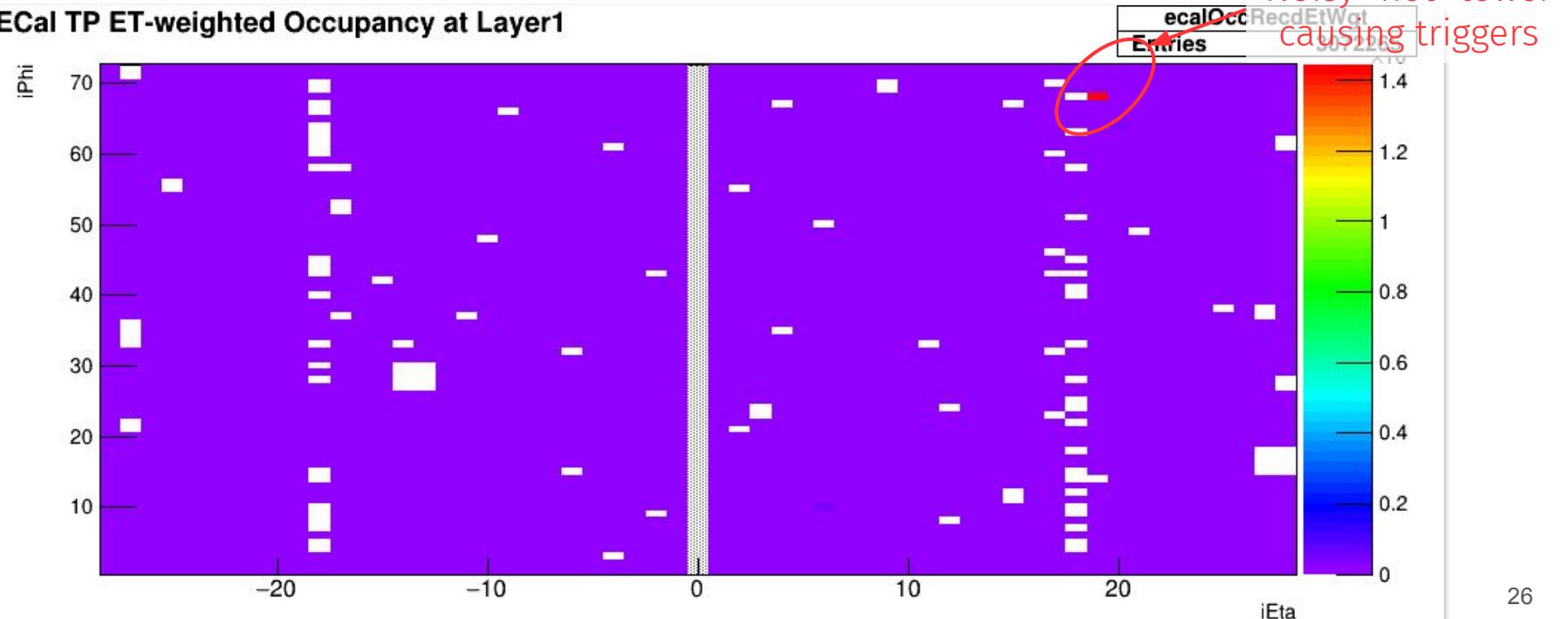
DQM  $\Rightarrow$  dashboard for prompt feedback on the quality of the collected data



# Data Quality Monitoring

E.g. we sometimes get noisy calorimeter towers  $\Rightarrow$  high L1 rates

ECal TP ET-weighted Occupancy at Layer1

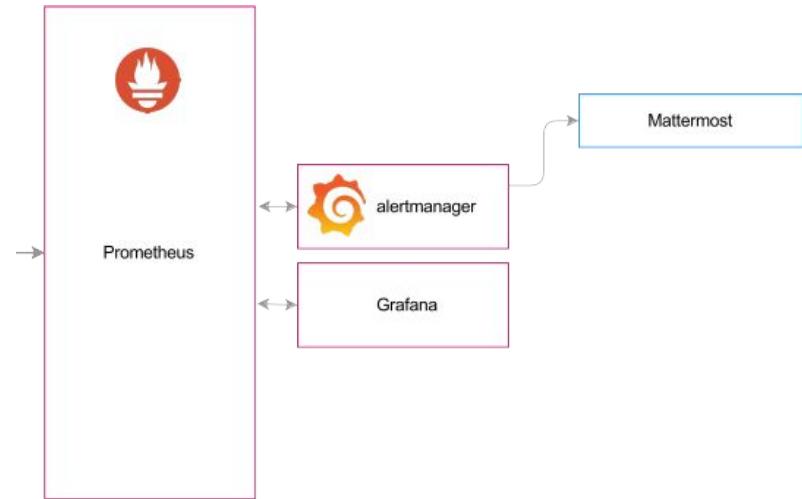


# Updating the monitoring system

- Software tools shown so far got us through Run 2, but continuous improvement with new tools essential to further reduce downtime and further increase reliability
- Over last few years, have been integrating new industry-standard tools
- Goals
  - Improve monitoring dashboards
  - Develop richer and smarter alerting
  - Reduce overall stress on experts
    - E.g. by suppressing warnings that we know on-call team doesn't need to react to

# New monitoring tools

- All metrics fed into a central Prometheus monitoring DB instance
- Alertmanager for alerting
- Grafana for system-wide dashboards
- Early testing and commissioning of the new infrastructure on OpenShift cluster (Kubernetes customisation)
  - Components are containerised for easy deployment
- Commissioning to production environment started in June



kubernetes



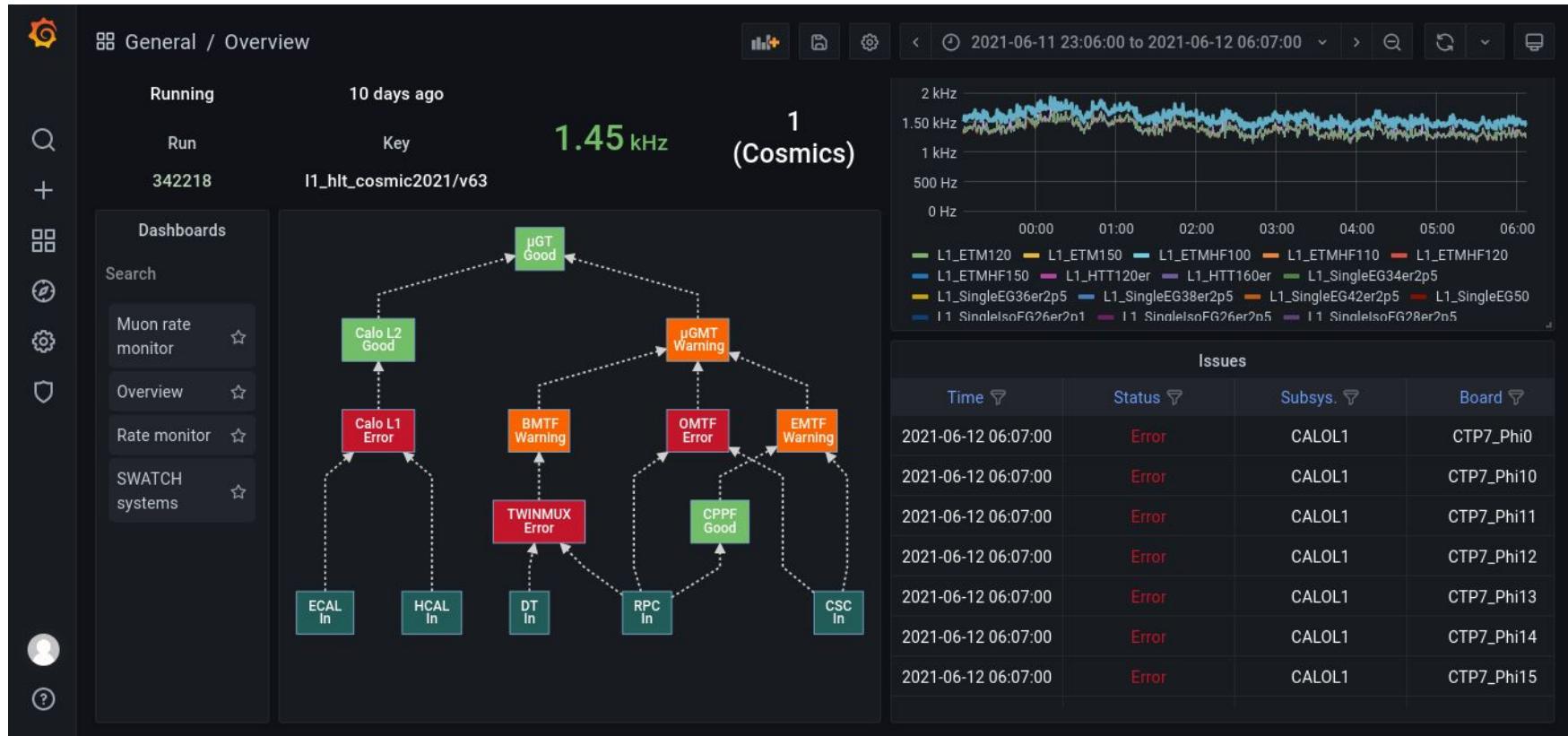
OPENSHIFT



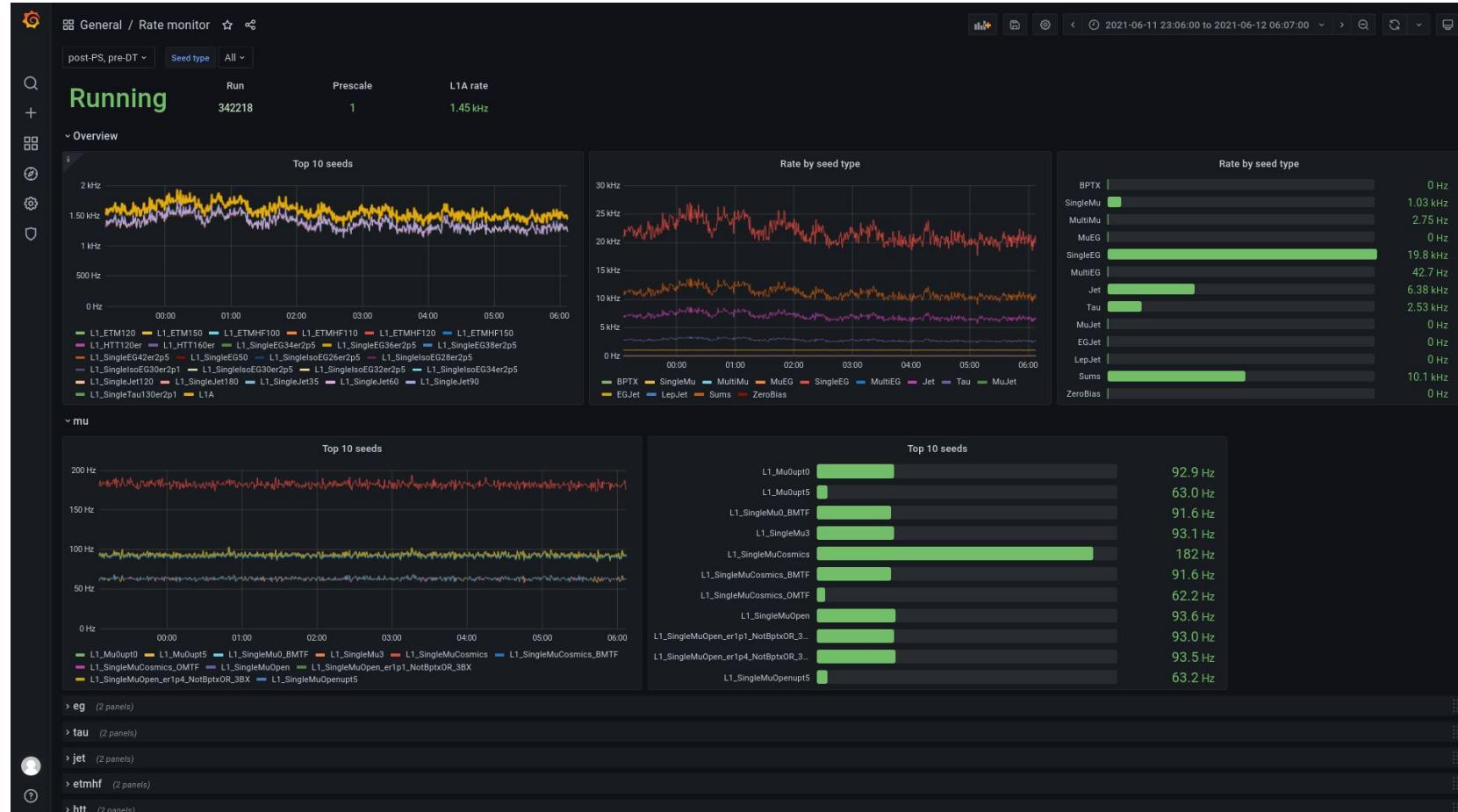
docker

# New monitoring dashboards

- Deployed new monitoring dashboards over the Summer



# New dashboards – rate monitoring



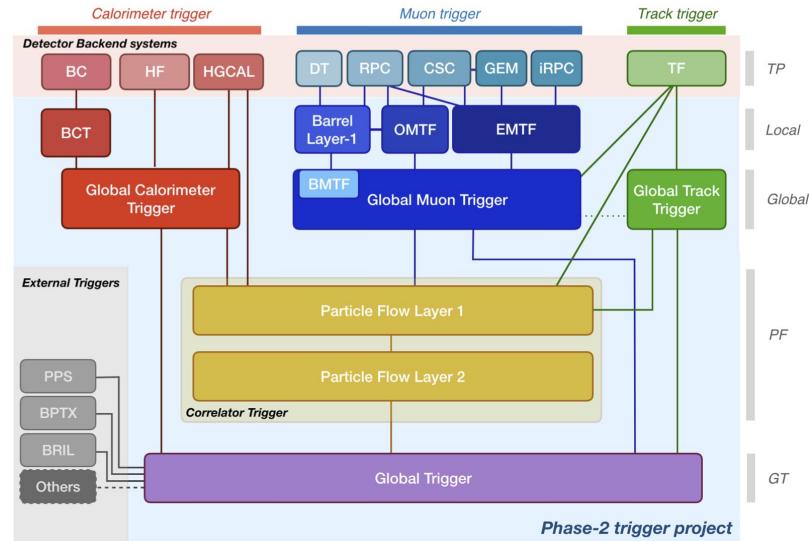
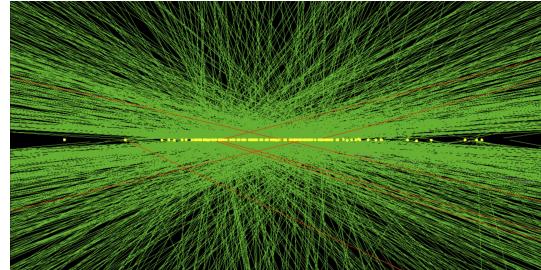
# Phase-2 upgrade

## The challenge

- 5-fold increase in collision rate,  
more pile up, but retain the same  
thresholds in the level-1 trigger

## The solution

- Re-build system using latest tech.
  - 256 boards hosting UltraScale+ FPGAs
  - Connected by O(10k) links, up to 25Gbps
- Consume more data
  - Calo. & muons: Finer granularity
  - Particle trajectories from tracker



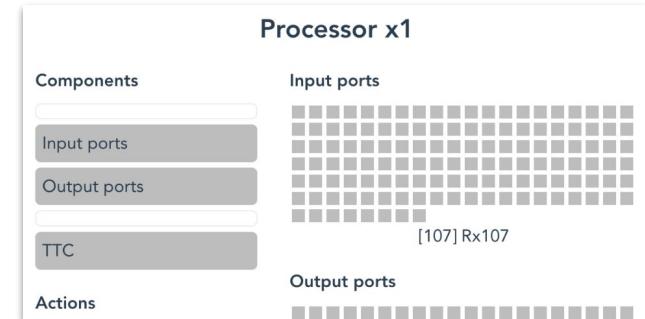
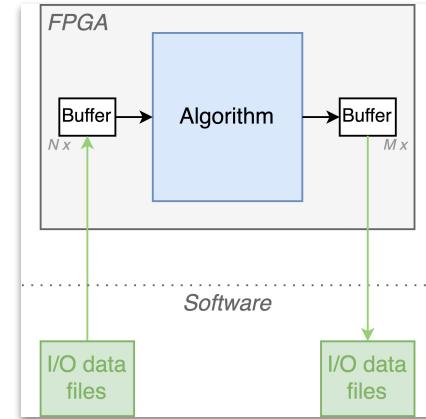
# Phase-2 upgrade

Focus over the next year: Implementing & validating all components of design from TDR

1. Test reco. & selection firmware in single boards
2. Test slices spanning multiple layers of system

⇒ Will need to control & monitor multiple boards – like in current system at P5

- Developing key components of phase-2 online software
  - Building on existing codebase, where applicable
  - Taking advantage of new tools & libraries
  - Informed by feedback from current system  
    ⇒ *Design the best system possible*
- Use for slice tests in coming months

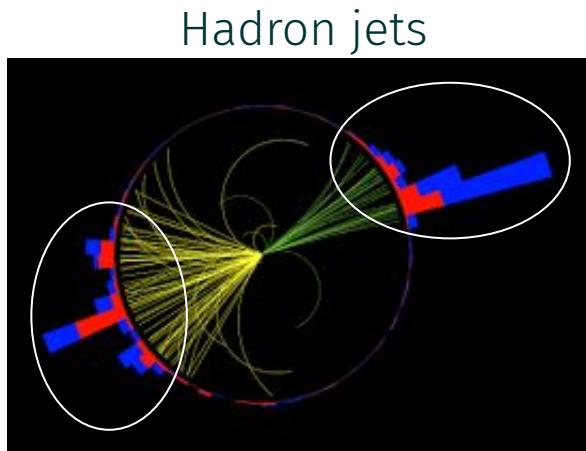


# Wrap-up

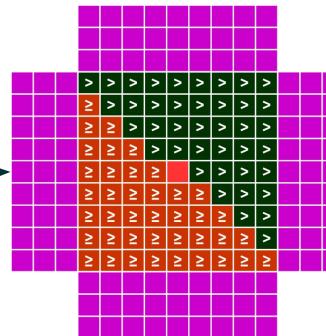
- Level-1 trigger is a complex system and a fundamental component of CMS
- Configured, controlled and monitored through the L1T online software
  - A sophisticated set of libraries and tools – L1CE, SWATCH, L1Page & DQM
- Renovating online software with modern, industry-standard tools
  - On-going work, but already useful for on call personnel
- Phase-2 upgrade
  - More complex than current system (# boards, # links, # subsystems)
  - Next generation of online software under development
- Interested in contributing?
  - Email: [simone.bologna@cern.ch](mailto:simone.bologna@cern.ch) and [tom.williams@cern.ch](mailto:tom.williams@cern.ch)

# Backup

# Example triggers



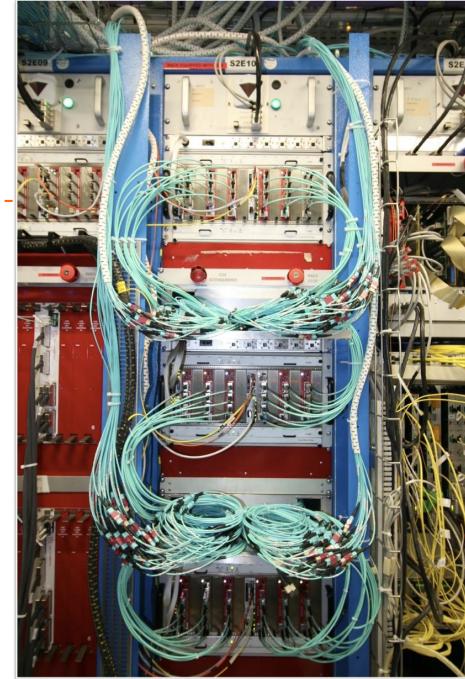
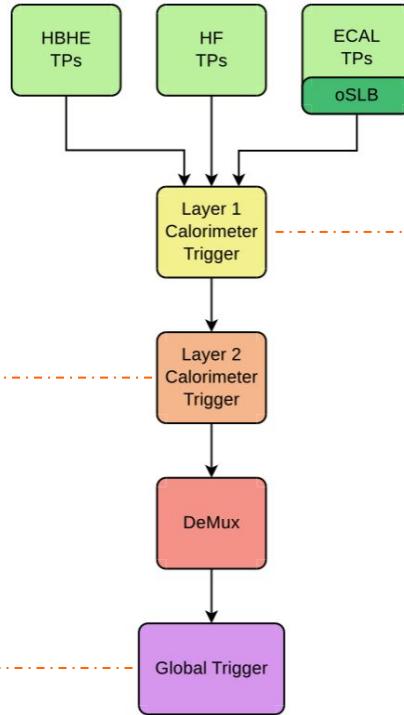
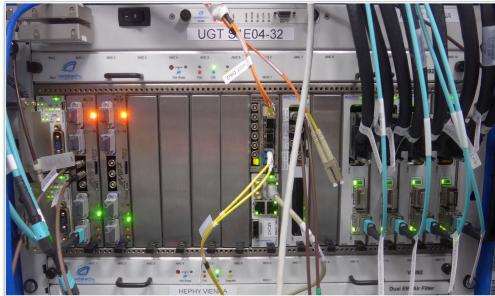
L1T, fixed-area cone



HLT, anti-kt 4

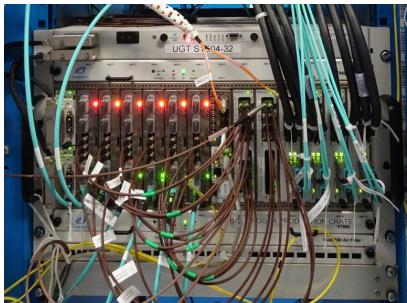
$$d_{ij} = \min(p_{\text{T},i}^{-2}, p_{\text{T},j}^{-2}) \frac{\Delta R_{ij}^2}{R^2} \quad \Delta R_{ij}^2 = (\eta_i - \eta_j)^2 + (\phi_i - \phi_j)^2$$
$$d_{iB} = p_{\text{T},i}^{-2}$$

# Level-1 trigger: Calorimeter chain



# CMS trigger

- Level-1 trigger, HW-based
- High-level trigger, SW based
- Successive selections of finer granularity



Level-1 trigger

L1T



HLT



High-level trigger