System design and prototyping of the CMS Level-1 Trigger at the High-Luminosity LHC

Tom Williams (RAL), on behalf of the CMS Level-1 Trigger project

CONTEXT

High-Luminosity LHC: 2 → 7.5 × 10^{34}cm^{-2}s^{-1}, pile-up 60 → 200

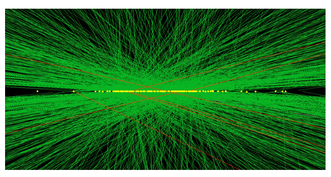
All CMS detector systems will be completely replaced or significantly upgraded to cope with harsher conditions.

Retain two-level trigger paradigm, in which the level-1 trigger selects which events will be read out, within fixed latency. With latest technology in readout chain, new design requirements:

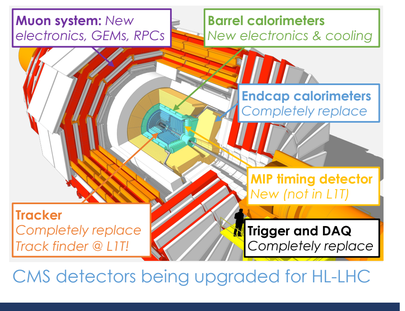
- Latency: 3.8 → 12.5 μs
- Max. readout rate: 100 → 750 kHz

Inputs (2 → 70 Tbps): Finer granularity info. from calorimeters & muon detectors. Particle trajectories from silicon tracker *(new!)*.

Aims: Maintain thresholds from current system despite harsher conditions; extend physics reach targeting specific topologies.



Vertices & tracks from HL-LHC collision

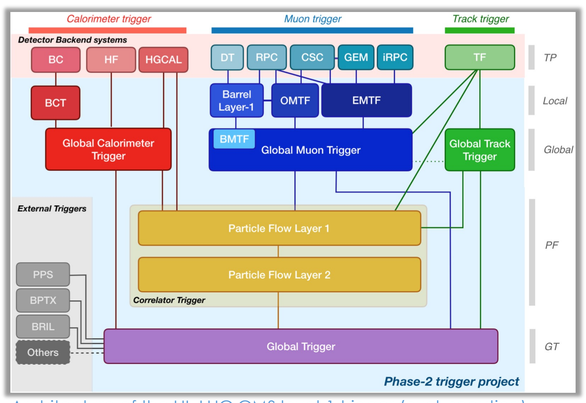


CMS detectors being upgraded for HL-LHC

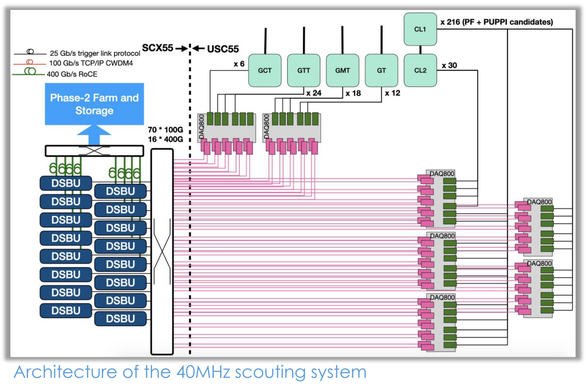
SYSTEM ARCHITECTURE

The level-1 trigger system is organised into several subsystems [1]:

- **Calorimeter**, **muon** and **Global Track Trigger** implement independent triggers from separate detector systems.
- **Correlator** combines information from all detector systems using particle-flow algorithms, reconstructing e, γ, τ, jets & sums with optimal performance.
- **Global trigger (GT)** applies menu of up to approx. 1700 trigger paths (single-/multi-object kinematic and quality cuts, or NNs) to decide whether to accept events.
- **40MHz scouting system**
  - Receives copy of all objects sent to GT, as well as PUPPI objects from correlator.
  - Performs analysis in real-time on server farm without latency constraint (not part of the readout decision).
  - Stores analysis results (i.e. histograms).
  - Also stores scouting inputs for O(%) of events, to develop new analyses.
  - Unique possibilities for multi-BX analyses, higher-stats real-time diagnostics & more.



Architecture of the HL-LHC CMS level-1 trigger (excl. scouting)



Architecture of the 40MHz scouting system

203 boards, of 4 different designs, implementing 20 functions, connected by thousands of links.

TRIGGER ALGORITHMS

With finer granularity inputs from the calorimeters and muon detectors, particle trajectories from the tracker, and newer FPGAs, we can significantly improve the trigger performance [1]:

- **Pile-up mitigation:** Improved through reconstruction of the primary vertex from charged particle tracks.
- **Particle reconstruction:** Efficiencies and resolutions closer to offline reconstruction (i.e. sharper turn-on curve), through use of particle-flow techniques that optimally combine calorimeter & track information.
- **New event topologies:** Developing sophisticated triggers to select specific topologies that could not target with current system, e.g. b-jets & rare B-meson decays (using tracks); forward μ triggers for τ → 3μ; displaced jets and muons.

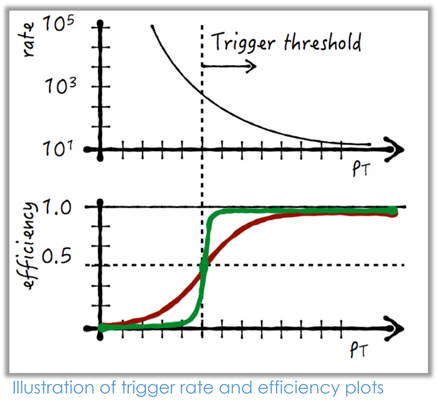
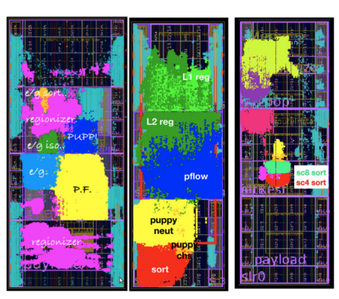


Illustration of trigger rate and efficiency plots

Implementation

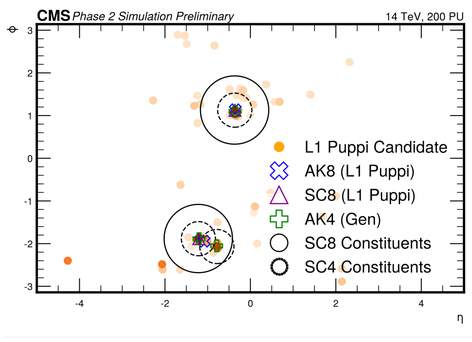
- Mixture of languages & tools, depending on the task:
  - **VHDL** for routing data between blocks, buffering etc.
  - **HLS** (High-Level Synthesis, C++) for computational core
  - **HLS4ML** [2,3]: Converts NN models → HLS inference logic
  - **Conifer** [4]: Toolkit for BDT evaluation
- Only use 50% of FPGA resources → room for more algos!
- Automated builds of firmware with GitLab CI!



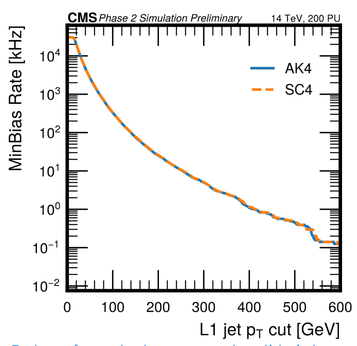
Floorplans of correlator firmware

Example: Seeded-cone jets [5]

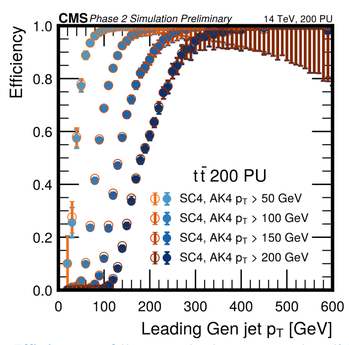
- Inputs: PUPPI candidates
- Two variants: SC4 & 8 with radius 0.4 & 0.8
- Iterative clustering, based on distances between each particle & jet radius
- Performance comparable to AK4/8
- Implemented for VU9P & VU13P, tested in hardware; latency < 1μs



Event display of L1 PUPPI particle candidates from a sample of Higgs bosons produced in associated with a W or Z boson; seeded-cone & anti-k_T jets superimposed.



Rates of seeded-cone and anti-k_T jets above a p_T threshold (R=0.4), both using the same L1T PUPPI candidates as inputs.



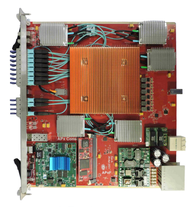
Efficiency of the seeded-cone and anti-k_T algorithms as a function of p_T, both using the same L1T PUPPI candidates as inputs.

HARDWARE

Level-1 trigger decision chain implemented with four cutting-edge generic data-processing ATCA boards:

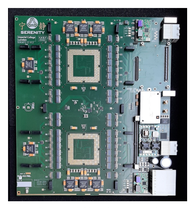
- **Data-processing engine:** VU13P FPGA
  - A2577 package; evaluating lidded & lidless
- **Optical I/O:** ≤ 124 25Gb/s in & 124 25Gb/s out
  - Samtec FireFly (4+4 and x12); QSFPs
  - Require bit error rate (BER) < 10^{-12}
  - Extensive QA tests of optical parts, incl. x12 25G FireFly beta parts
    - Measuring BER vs OMA (Optical Modulation Amplitude); tuning optical module settings for electrical interface.
- **On-board computer:** Xilinx Kria (System on Module)
  - Handles control, monitoring and management tasks, including interface to off-board software.
- **IPMC (shelf management):** Implemented either on dedicated mezzanine, or in the Zynq PL.

Synchronization, readout and scouting implemented using DTH-400 & DAQ-800 boards from DAQ group [6].

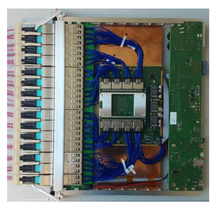


APx board

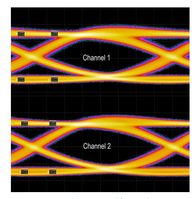
BMT-L1 board



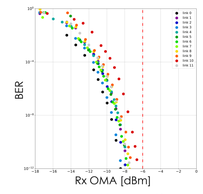
Serenity board



X2O board



Example optical eyes for a 25Gb/s part



Example attenuation scan of a 25Gb/s part

After extensive tests of prototypes (e.g. thermal performance, signal quality, functional tests individually and in systems), we will be starting final production runs in the next year.

INTEGRATION TESTS

For each algorithm, a bit-accurate CMSSW-based emulator is developed alongside the firmware; this emulator software must completely reproduce the firmware output at bit level.

To validate the system design in an efficient manner, we adopted a strategy that factorises the tests via well-defined I/O interfaces and the use of capture/playback buffers on each board:

1. **Single-board tests**
   - Goal: Validate algorithm firmware for each element in the baseline design of the system (independently)
   - Inject Monte Carlo simulated data via buffers, then compare captured outputs with emulator, bit-by-bit
   - N.B. Data in buffers has the exact same format as that sent to/from link cores in multi-board tests
2. **Multi-board slice tests**
   - Start off with 2-board slices to test each interface
     - Some inputs from links, rest from local buffers
   - Once overlapping 2-board slice tests successful, merge them together to form 3-, 4-, N-board slices.

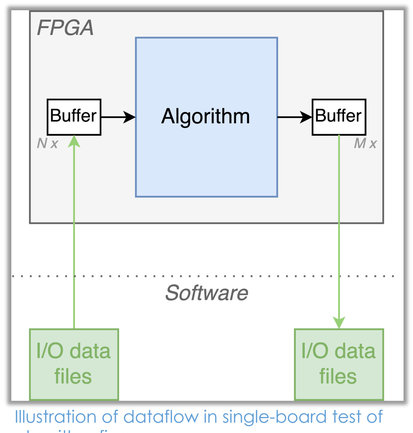


Illustration of dataflow in single-board test of algorithm firmware

The **majority of algorithms** have been **verified in single-board tests**, notably the most challenging designs. **Several interfaces** have been **tested with multi-board slices. Latency is measured routinely** in both single- and multi-board tests; current results imply a system latency of **8.6 μs, less than our target of 9.5 μs**.

Tests performed around the globe, with main facility at CERN in CMS Electronics Integration Centre. Also, a slice of phase-2 prototype electronics is processing real pp data from DTs.

A **low-latency asynchronous link protocol** has been defined to transfer LHC-synchronous trigger candidates between the boards. **Implementations** of this protocol have recently been **extensively tested to verify their compatibility and robustness against errors**. This included both the injection of well-defined errors in firmware, and induction of random errors on the link.

Integration test crate at CERN

ONLINE SOFTWARE

The online software provides a single, uniform interface for controlling, configuring and monitoring the various boards and algorithms in the trigger system:

- **Architecture based on proven framework + plugin approach** of current level-1 trigger system, which maximises the fraction of common code
  - Re-use core C++ framework (SWATCH) [7]
  - Update architecture for on-board CPU
  - Enhance flexibility to support test stands & labs
- Common on-board application: **HERD**
  - Loads plugins, which bring board-/subsystem-specific procedures and monitoring data
- Off-board supervisor: **Shep**
  - Implemented web-based and command-line user interfaces, using latest open-source libraries
- **Leveraging GitLab CI** to streamline development
  - E.g. run single-board test from browser in 1 click!

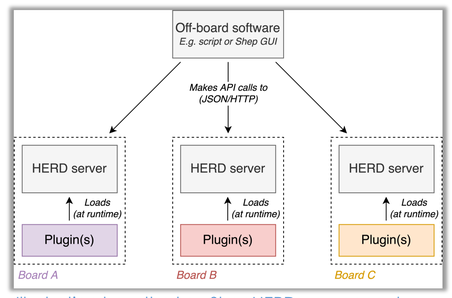
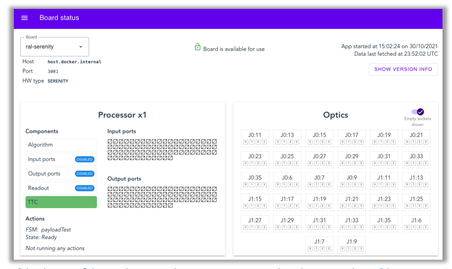


Illustration how the key ShepHERD components interact in a multi-board test



Status of key board components shown by Shep

REFERENCES

[1] The CMS collaboration, "The Phase-2 Upgrade of the CMS Level-1 Trigger", CERN, Geneva, 2020. CERN-LHCC-2020-004.
[2] The FastML Team, *fastmachinelearning/hls4ml*. Zenodo, 2023. DOI: 10.5281/zenodo.1201549.
[3] J. Duarte et al., "Fast inference of deep neural networks in FPGAs for particle physics", *JINST*, vol. 13, no. 7, P07027, 2018.
[4] S. Summers et al., "Fast inference of Boosted Decision Trees in FPGAs for particle physics", *JINST*, vol. 15, no. 5, P05026, 2020.
[5] The CMS collaboration, "Jet Reconstruction with the Seeded Cone algorithm in the CMS Phase-2 Level-1 Trigger", CERN, Geneva, 2023. CERN-CMS-DP-2023-023.
[6] The CMS collaboration, "CMS Phase-2 DAQ and Timing Hub – Prototyping results and perspectives", *JINST*, vol. 17, no. 5, C05003, 2022.
[7] S. Bologna et al., "SWATCH: Common software for controlling and monitoring the upgraded level-1 trigger of the CMS experiment" Proceedings of 20th IEEE-NPSS Real Time Conference, 2016. DOI: 10.1109/RTC.2016.7543077.