CMS ECAL: Meeting the trigger performance challenges of Run 3

John Dervan[1,*] on behalf of the CMS collaboration

[1]Northeastern University

Abstract. The CMS electromagnetic calorimeter (ECAL) at the CERN LHC is a high-resolution crystal calorimeter equipped with fast digital signal processing algorithms for measuring the energy and timing of calorimeter deposits. These trigger primitives are sent to the Level-1 trigger system at a rate of 40 MHz to help decide whether to store the event data offline. This paper summarizes the ECAL trigger performance during LHC Run 3 (2022+), highlighting enhanced calibration methods for trigger primitives. A key focus is the commissioning of a new automated procedure for deriving, validating, and deploying time-dependent response corrections using in-situ measurements from a dedicated laser monitoring system. Additionally, the rejection of unwanted large signals ("spikes") caused by hadronic deposits on the APD photodetectors in the ECAL barrel region ($|\eta| < 1.48$) has been a persistent issue since 2009. The evolution of spike rejection techniques in ECAL's on-detector and off-detector electronics, along with the potential of currently unused features of the ECAL electronics to further enhance spike rejection, will be discussed.

1 Introduction

The Compact Muon Solenoid (CMS) experiment is a general-purpose detector operating at the Large Hadron Collider (LHC) at CERN on the French-Swiss border. The CMS physics program combines precision measurements with searches for physics beyond the Standard Model, including supersymmetry, dark matter, warped extra dimensions, and a wide range of other phenomenology. CMS combines input from its subdetectors to identify and measure the momenta and energies of objects passing from the interaction point.

Vital to this program is the CMS electromagnetic calorimeter (ECAL), a homogeneous crystal calorimeter whose primary task is the measurement of electron and photon energies but which also contributes to measurements of the electromagnetic fractions of jets, taus, and energy sums. ECAL consists of 75,848 lead-tungstate ($PbWO_4$) crystals divided into the barrel and endcap regions which combined provide hermetic coverage up to a pseudorapidity of $|\eta| < 3$. The barrel region, covering $|\eta| < 1.479$, consists of 61,200 crystals with readout via avalanche photodiodes (APDs); the two endcaps, covering the remaining $1.479 \leq |\eta| < 3$, consist each of 7,324 crystals affixed to vacuum phototriodes (VPTs). In addition, a silicon-based preshower detector in the endcaps enhances the neutral pion detection capabilities of CMS.

1.1 The CMS trigger system

In order to optimize the physics of interest given readout and processing bandwidth constraints, CMS utilizes a two-tiered trigger system. The first tier, the Level-1 (L1) trigger [1], consists of hardware-level algorithms running on custom electronics with a fixed latency of 3.8 μs, and reduces the event rate from the 40 MHz delivered by the LHC down to around 110 kHz. At L1, information from ECAL, the hadronic calorimeter, and the muon system are fed to the global trigger.

The second tier, known as the high-level trigger (HLT), consists of high-level physics algorithms processed on an event filter farm comprised of commercial computers, and incorporates tracking information. In Run 3, HLT operates with a processing time of up to 500 ms per event, and further reduces the rate from L1 down to 5 kHz [2].

The front-end electronics of ECAL sum energies in groups of neighboring crystals known as trigger towers and send trigger primitives (TPs) to the L1 trigger to form L1 electron/photon, jet, tau, and energy sum candidates [3]. In addition to a per-tower (in the barrel) or 5-crystal pseudostrip (in the endcaps) measurement of transverse energy, TPs consist of up to two feature bits to characterize the inter-crystal energy spread along with an LHC bunch-crossing assignment. The broad goal of ECAL in the CMS trigger is to keep efficiencies high, thresholds low, and rates stable with time.

1.2 The challenges of Run 3

Over the course of LHC operation, the instantaneous luminosity delivered to CMS has increased, resulting in larger data sets and opportunities for improved sensitivity to rare processes. However, this has also come with an increase in pileup, or the number of overlapping interactions per bunch crossing, and these two factors pose challenges for maintaining good trigger performance in CMS. Figure 1 shows the substantial increase in pileup compared to Run

*e-mail: john.dervan@cern.ch

2; moreover, the total integrated luminosity for Run 3 is expected to reach 260 fb^{-1}.

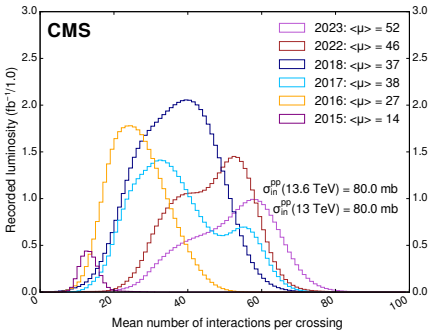

Figure 1. Distribution of the pileup for proton-proton collisions in Run 2 (2015–2018) and the first two years of Run 3 (2022, 2023).

Two primary challenges arise from the increased luminosity in Run 3 which will be discussed here.

2 ECAL response losses

Lead tungstate was chosen as a scintillation medium for its short radiation length, small Molière radius, and fast scintillation time [4]. However, crystals lose response due to irradiation during LHC operation, and this effect must be taken into account. Radiation damage in ECAL is categorized as either electromagnetic damage, which is dose rate-dependent and partially recoverable via annealing at room temperature, and hadron damage, which is dose-dependent and non-recoverable. Figure 2 shows the response loss since 2011, which includes both electromagnetic and hadron damage as well as photodetector response losses.

To account for the crystal response losses in Run 3, ECAL has adopted improved tracking of response changes

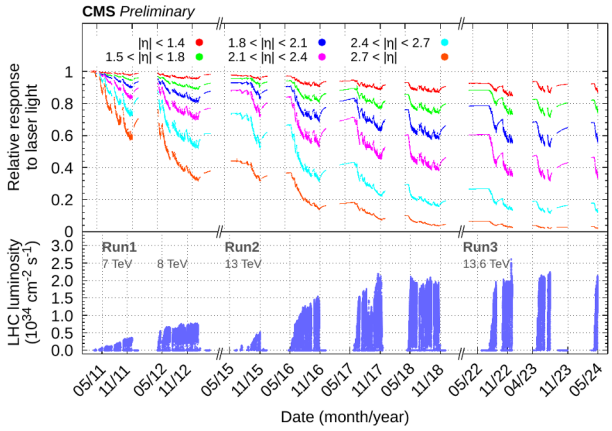

Figure 2. Top panel: response histories for different pseudorapidity regions in CMS since Run 1. Partial recovery of electromagnetically-induced damage can be seen during periods without beam. Bottom panel: instantaneous luminosity delivered by the LHC from 2011 to mid-2024.

at L1 and HLT. ECAL uses a dedicated laser system which fires at each crystal once per 40-minute cycle during LHC collisions. Response measurements are then processed and provide corrections to stabilize the per-channel response over time, with lower response amplitudes corresponding to larger corrections. Corrections are then validated and inserted into an offline database, where they are used for prompt reconstruction within 48 hours of data-taking.

During Run 1 (2012), corrections at L1 and HLT were applied once per week in 22 pseudorapidity regions, and were applied only in the endcap regions. At the start of Run 2 (2015–2016), the corrections were applied at per-crystal granularity, and extended into the barrel region. Near the end of Run 2 (2017–2018), the frequency of these corrections was increased to twice per week. In Run 3, ongoing since 2022, a new scheme has been developed to validate and upload corrections at per-fill frequency, in order to better capture the substructure of laser response changes. This can be seen visually in the left and right-hand sides of Figure 3. Using the $Z \rightarrow ee$ mass ratio, the new update scheme has been validated and shown to result in a more stable energy scale (Fig. 4). This new scheme has been made possible through a revamped laser processing framework, which reduces the update latency from 6–7 hours down to 1.5 hours, as well as the full automation of the validation and deployment process.

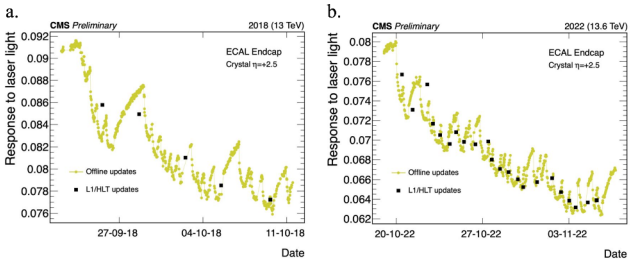

Figure 3. Response tracking for a single ECAL crystal over three-week periods for the twice-per-week (a.) and once-per-fill (b.) update scheme at L1/HLT. Yellow points represent offline laser response measurements at the time of recording; black points show L1/HLT laser response measurements at the time of deployment.

3 Spike killing

Since Run 1, ECAL has been shown to exhibit large spurious energy pulses in the barrel region induced by direct hadronic ionization of the APDs. If untreated, these "spikes" would saturate the trigger bandwidth at L1 and risk introducing significant biases in the energy reconstruction of electrons, photons, jets and missing transverse energy. Moreover, spikes become more prevalent with higher instantaneous luminosity and pileup[1], necessitating effective mitigation at L1, HLT, and offline reconstruction.

[1]At $\sqrt{s}$ = 7 TeV, an average of one spike with E_T > 3 GeV is observed per 370 minimum bias triggers. This rate scales with pileup and the logarithm of $\sqrt{s}$.

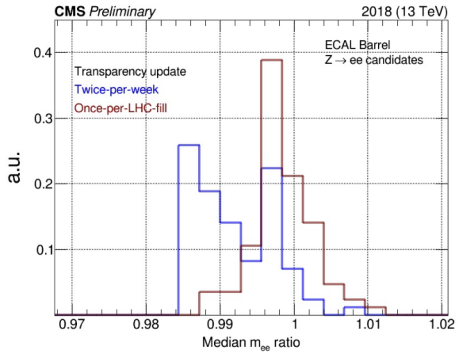

Figure 4. Comparison of invariant mass m_{ee} ratio distributions from $Z \rightarrow ee$ candidates for Run 2 and Run 3 update schemes, evaluated on a portion of data from 2018. The ratio is taken between the median invariant mass computed with respective online response correction schemes and full offline laser corrections applied every 45 minutes. Once-per-fill updates produce a more stable energy scale with a smaller bias.

3.1 Online and offline spike killer

ECAL uses a combination of online and offline spike killing to reduce the impact of spikes. The on-detector electronics make use of a feature bit known as the strip fine-grain veto bit (sFGVB), which exploits the fact that spikes tend to exhibit only a single crystal in a trigger tower with significant energy deposited, in contrast to the inter-crystal energy sharing typical of electromagnetic showers (Fig. 5). The sFGVB uses the distribution of crystals above a configurable single-channel threshold in a trigger tower to determine if a deposit is spike-like or shower-like. If the tower E_T is above a tunable threshold and the sFGVB returns 0, the deposit does not trigger. During Run 1, the sFGVB achieved better than 95% efficiency for spikes above 8 GeV with minimal impact on electron trigger efficiency [5].

Figure 5. Schematic of sFGVB logic. In each strip, if there are more than one crystal above a configurable energy threshold, the strip bit is set to 1. The strip bits are ORed to give the per-tower sFGVB. An sFGVB of 1 is considered shower-like, while a value of 0 is considered spike-like.

The spike killing algorithm used at HLT and offline utilizes a "Swiss cross" variable which similarly exploits anomalous energy sharing between crystals (Fig. 6). A cut above 0.95 of this variable results in 99% spike rejection above 10 GeV with minimal impact on electromagnetic shower efficiency [5]. A timing cut orthogonal to the Swiss cross is also used offline which helps to reject additional non-isolated spikes. The performance of the L1 spike killer can be evaluated by comparison against spikes flagged offline by the Swiss-cross algorithm and timing cut.

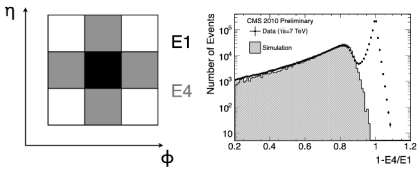

Figure 6. Schematic of the offline spike killer, which utilizes the Swiss cross variable.

The L1 spike killer has been implemented since 2011 but requires retuning with changing luminosity conditions. During LHC Long Shutdown 2 ahead of Run 3, a retuning of the single-channel threshold was performed to cope with higher pileup and photodetector noise. This retuning was shown to achieve a 12% spike contamination for TPs above 30 GeV (Fig. 7), reduced from 19% in Run 2. However, the killing threshold of 16 GeV, set in 2016, means that lower-energy spikes fail to be removed at L1.

3.2 Double weights mechanism

The time of arrival of spikes exhibits a long tail which is exploited offline, but until now could not be exploited at L1. In order to target low-energy spikes at L1, a new hardware feature that targets the timing profile of late spikes has been under study since 2018. This "double weights" mechanism exploits an unused duplicate datapath in the strip Front-End New Information eXtractor (FENIX) ASIC. During the computation of trigger primitives, FENIX performs amplitude reconstruction using a set of digital weights comprising a filter on digitized pulses. These "EVEN" weights are fully configurable to a given waveform of interest, and are optimized for in-time signals. A second, previously unused set of tunable "ODD" weights has been explored as an out-of-time tagger. The double weights relies on comparing the weighted sums of strip E_T values as in Eq. 1, where S_i are the five digitized samples for a given strip and $w_i^{EVEN,\ ODD}$ are the corresponding EVEN or ODD weights. If $E_T^{ODD} > E_T^{EVEN}$,

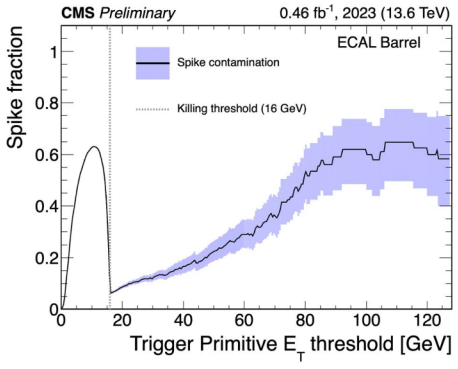

Figure 7. Spike contamination as a function of trigger primitive E_T threshold measured using 2023 data.

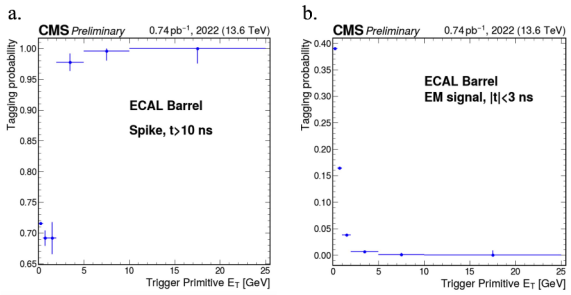

Figure 8. Spike tagging (a.) and signal mis-tagging (b.) probabilities as a function of trigger primitive E_T for 2022 data.

the result is flagged (and can be rejected); if $E_T^{ODD} \leq E_T^{EVEN}$, the result is accepted. In this way, late out-of-time spikes can be flagged and their energies suppressed.

$$E_T^{str.} = \sum_{i=0}^{4} S_i \times w_i^{EVEN,\ ODD} \tag{1}$$

The values of the ODD weights have been derived via numerical optimization to target late (+10ns) spikes. The optimization loss function was parametrized by the relative priorities of signal retention and spike rejection (λ_{sig} and λ_{spk} in Eq. 2, respectively) as well as minimum separation $\delta_{min} = E_T^{ODD} - E_T^{EVEN}$ of the E_T values calculated via the EVEN and ODD weights:

$$L = (\lambda_{\text{sig}} \times L_{\text{sigEff}}) + (\lambda_{\text{spk}} \times L_{\text{spkRej}}) + (\lambda_{\text{norm}} \times W2LN) + W2LL \tag{2}$$

with the components L_{sigEff} and L_{spkRej} given by

$$L_{\text{sigEff}} = \begin{cases} (A_{W_2,d_1} - A_{W_1,d_1}) \geq \delta_{min} : A_{W_2,d_1} - A_{W_1,d_1} \\ (A_{W_2,d_1} - A_{W_1,d_1}) < \delta_{min} : 0\,, \end{cases}$$

$$L_{\text{spkRej}} = \begin{cases} (A_{W_1,d_2} - A_{W_2,d_2}) \geq \delta_{min} : A_{W_1,d_2} - A_{W_2,d_2} \\ (A_{W_1,d_2} - A_{W_2,d_2}) < \delta_{min} : 0\,. \end{cases}$$

The amplitudes $A_{W_x,\,d_x} = \sum_{i=1}^{5} W_{x,\,i} \times d_{x,\,i}$ are the sums of the EVEN ($x = 1$) or ODD ($x = 2$) weights multiplied by the signal ($x = 1$) or spike ($x = 2$) digis, while $W2LN$ and $W2LL$ are normalization terms that help control the magnitude of the weights.

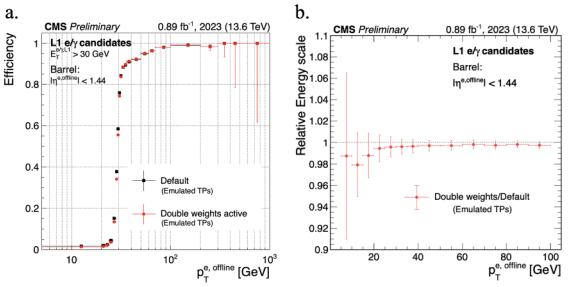

Figure 9. Impact of double weights on L1 electron/photon trigger efficiency (a.) and energy scale of L1 electron/photon candidates (b.) as a function of offline electron transverse momentum.

The results shown here utilize a δ_{min} value of 2.5 GeV, which offers the best compromise between signal efficiency and spike rejection. When tested with data from a 2022 run, the double weights mechanism flags more than 99% of out-of-time spikes above $E_T \geq 5$ GeV, while mistagging less than 1% of real electromagnetic signals above $E_T \geq 2$ GeV (Fig. 8). Meanwhile, the L1 electron/photon triggering efficiency and L1 energy scale are not significantly impacted by the double weights (Fig. 9).

4 Conclusion

The CMS ECAL plays a central role in the performance of the CMS trigger system at L1 and HLT. In Run 3, the increased instantaneous luminosity compared to earlier years poses a challenge to maintaining good trigger performance, as the impacts of crystal response losses and direct APD pulses in the barrel region of ECAL become more severe with increasing luminosity. In Run 3, crystal response losses are more precisely tracked by a revamped response validation and deployment system that allows for per-fill updates at per-crystal granularity. Meanwhile, spike rejection benefits from a retuned online spike killer. The potential for a new functionality in the strip FENIX ASIC to be used to mitigate late spikes at L1 has also been investigated. This double weights mechanism appears to improve spike killing without significant impact on signal efficiency or energy scale, but some final investigation and quantification of the potential improvements will be required before a decision is made to implement this feature online.

References

[1] A. M. Sirunyan *et al.*, Performance of the CMS Level-1 trigger in proton-proton collisions at $\sqrt{s} = 13$ TeV. JINST **15** P10017 (2020). https://doi.org/10.1088/1748-0221/15/10/P10017

[2] A. Hayrapetyan *et al.*, Development of the CMS detector for the CERN LHC Run 3. JINST **19** P05064 (2024). https://doi.org/10.1088/1748-0221/19/05/P05064

[3] P. Paganini (for the CMS collaboration), CMS Electromagnetic Trigger commissioning and first operation experiences. J. Phys.: Conf. Ser. **160** 012062 (2009). https://doi.org/10.1088/1742-6596/160/1/012062

[4] The CMS collaboration, The Electromagnetic Calorimeter Technical Design Report. CERN/LHCC 97–33 (1997).

[5] D. A. Petyt (for the CMS collaboration), Mitigation of anomalous APD signals in the CMS electromagnetic calorimeter. J. Phys.: Conf. Ser. **404** 012043 (2012). https://doi.org/10.1088/1742-6596/404/1/012043