Design, Implementation, and Validation of the Barrel Calorimeter Trigger Firmware for the High-Luminosity LHC

Abhinav Mohan* and Bhawna Gomber* on behalf of the CMS Collaboration

*Center for Advanced Studies in Electronics Science and Technology, School of Physics, University of Hyderabad, Hyderabad, India

doi: https://doi.org/10.37285/bsp.SACAD2025.53

***Abstract*—The High-Luminosity upgrade of the Large Hadron Collider (HL-LHC) will increase event pileup from approximately 50 to 140-200 simultaneous interactions per bunch crossing, requiring a complete redesign of the Compact Muon Solenoid (CMS) Level-1 (L1) trigger system. This paper presents the barrel calorimeter trigger firmware developed for the CMS Phase-2 L1 trigger upgrade, demonstrating readiness for HL-LHC deployment. We describe the Regional Calorimeter Trigger (RCT) and Global Calorimeter Trigger (GCT) subsystems, which process high-granularity data from the barrel calorimeters using a new 17×6 architecture implemented with High-Level Synthesis (HLS) IP cores. The RCT firmware performs physics object reconstruction such as cluster finding, shower shape determination, and real-time energy calibration with software-controlled threshold calibration for e/γ tagging without reprogramming the FPGA. The GCT firmware aggregates data and reconstructs higher-level objects such as Particle flow clusters, jets and taus while using an efficient TMUX18 time-multiplexing scheme to improve interconnection efficiency with correlator layer 1. The development workflow was improved through automated physical optimization techniques, incremental synthesis, and modular Out-of-Context synthesis flows. Both subsystems operate on Xilinx Virtex UltraScale+ VU13P-2 FPGAs at 360 MHz clock frequency. The integrated system achieves 1.01 µs end-to-end processing latency within the 12.5 µs L1 trigger budget.**

Keywords*—High-Luminosity LHC (HL-LHC), CMS Detector, Level-1 Trigger, Field-Programmable Gate Array (FPGA), High-Level Synthesis (HLS), Calorimeter Trigger.

I. Introduction

The High-Luminosity Large Hadron Collider (HL-LHC) upgrade will increase instantaneous luminosity by a factor of 5-7 compared to the original LHC design. This enhancement presents significant challenges for the Compact Muon Solenoid (CMS) experiment, particularly due to increased event pileup from approximately 50 to an average of 140-200 simultaneous interactions per bunch crossing [1][2]. The current Phase-1 Level-1 (L1) trigger system cannot efficiently operate under these conditions due to its reliance on coarse-granularity detector data, lack of tracking information, and limited algorithmic complexity [9]. These limitations would result in either excessive trigger rates or prohibitively high energy thresholds, which compromise the physics program's sensitivity.

To address these challenges, the CMS collaboration has developed a Phase-2 upgrade of the L1 trigger system, shown in Fig.1. The upgraded system extends the L1 latency budget from 3.8µs to 12.5µs and increases the L1 accept rate from 100 kHz to 750 kHz. The system is built on custom Advanced Telecommunications Computing Architecture (ATCA) boards featuring Xilinx Virtex UltraScale+ FPGAs.

Table I summarizes key differences between the Phase-1 and Phase-2 trigger systems.

This paper describes the barrel calorimeter trigger firmware, comprising two cascaded subsystems: the Regional Calorimeter Trigger (RCT) and the Global Calorimeter Trigger Barrel (GCT-B). These systems process high-granularity data from the barrel calorimeters, implementing firmware architectures and algorithmic pipelines to meet the performance requirements of the HL-LHC era.

TABLE I: COMPARISON OF CMS PHASE-1 AND PHASE-2 LEVEL-1 TRIGGER SYSTEMS

Parameter	L1 Trigger Upgrade Phase	
	Phase 1 (current)	*Phase 2*
Latency Budget (µs)	3.8	12.5
Accept Rate (KHz)	100	750
Average Pileup	□50	140-200
Data Granularity	Coarse	High
Tracker information	No	Yes
Algorithm Complexity	Limited	Advanced

II. Barrel Calorimeter Trigger Architecture

The barrel calorimeter trigger consists of two primary hardware stages implemented on the APx-F `Falcon' ATCA board, which hosts a Xilinx Virtex UltraScale+ VU13P-2 FPGA. This FPGA's four Super Logic Regions (SLRs) and 124 high-speed multi-gigabit transceivers provide the required logic capacity and I/O bandwidth. The system includes 24 identical RCT boards that receive data directly from detector front-ends, and 3 GCT Barrel boards that aggregate RCT outputs, as illustrated in Fig.1.

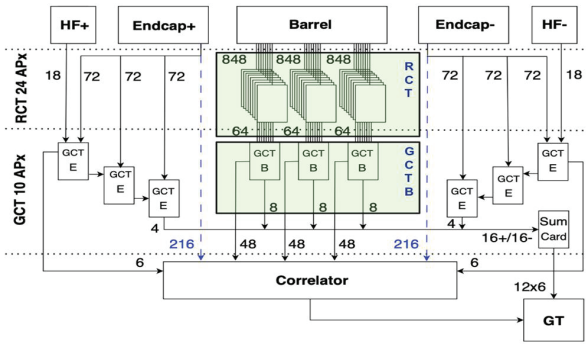


Fig 1. High-level overview of the CMS Phase-2 Level-1 Trigger architecture, showing the dataflow from detector front-ends through the calorimeter and muon triggers to the Global Trigger. RCT and GCT Barrel are highlighted.

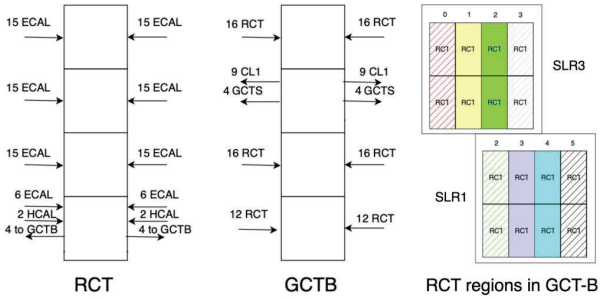


Fig. 2. (left) Per SLR IO configuration for RCT, (middle) Per SLR IO configuration for GCT-B, (right) RCT regions processed by each GCT-B board, overlap regions are shaded..

A. Hardware Platform

The APx-F board shown in Fig. 3 integrates 124 high-speed optical links operating at 25 Gb/s, and are tested for endurance with a bit error rate below 10^{-15}. Thermal management is done through an efficient heatsink design that maintains FPGA temperatures below 100°C during 200 W operation, for reliable performance in the high-radiation environment of the HL-LHC.

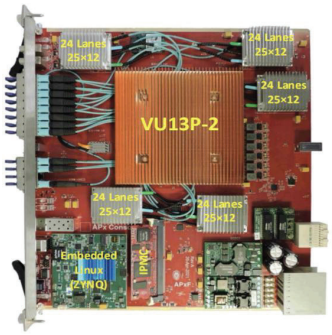
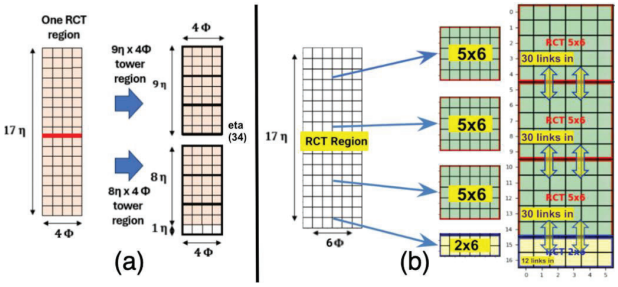


Fig. 3. The APx 'Falcon' ATCA Board based on the VU13P-2 FPGA.

B. Regional Calorimeter Trigger (RCT) Subsystem

The RCT provides the first processing layer. Each RCT board processes a detector region spanning 17η × 6Φ in trigger tower granularity. The firmware architecture matches the physical four-SLR structure of the VU13P FPGA. The 17η × 6Φ region is divided into four sub-regions (three 5η × 6Φ and one 2η × 6Φ) which are processed by four parallel High-Level Synthesis (HLS) IP cores, to minimize latency-critical cross-SLR communication. This 17x6 design has improved logic distribution and resource balancing compared to a previous 17x4 architecture based on the VU9P FPGA [9], which would require 36 boards to cover the barrel region, while the 17x6 architecture needs only 24 VU13P FPGAs. A comparison of 17x6 and 17x4 is shown in Fig. 4. The proposed floorplan for the RCT, which maps processing logic to different SLRs, is shown in Fig. 5.

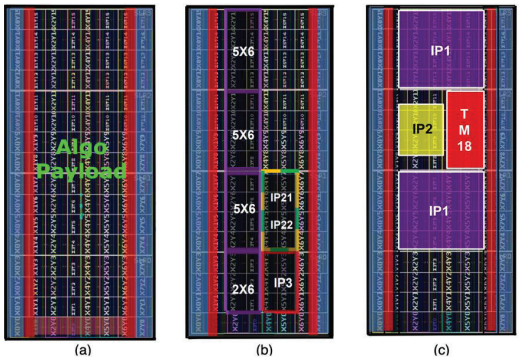
Fig. 4. Comparison of logic distribution for the new 17x6 architecture (b) versus the previous 17x4 architecture (a).



C. Global Calorimeter Trigger (GCT) Subsystem

The GCT Barrel subsystem consists of three boards, each receiving data from eight RCT boards. The GCT performs higher-level object reconstruction by correlating information across multiple RCT regions. To ensure continuous object reconstruction across boundaries, each GCT board receives overlapping data from adjacent RCTs, as seen in Fig. 2 (right). Processing within each GCT FPGA is split across two SLRs (SLR 3, 1), each handling data from four primary and four adjacent RCTs. The proposed floorplan for this two-SLR processing division is shown in Fig.5. The GCT barrel firmware also reconstructs jets and tau candidates, refines e/γ objects, and computes barrel-wide energy sums (SLR2). A time-multiplexing (TMUX18) stage serializes the parallel outputs for transmission to the downstream Correlator and Global Trigger systems for superior interconnect efficiency.

Fig. 5. (a): Area available for algorithm payload after subtracting Firmware Shell (blue) and DAQ (red). (b): Proposed RCT floorplan



(c): Proposed GCT floorplan.

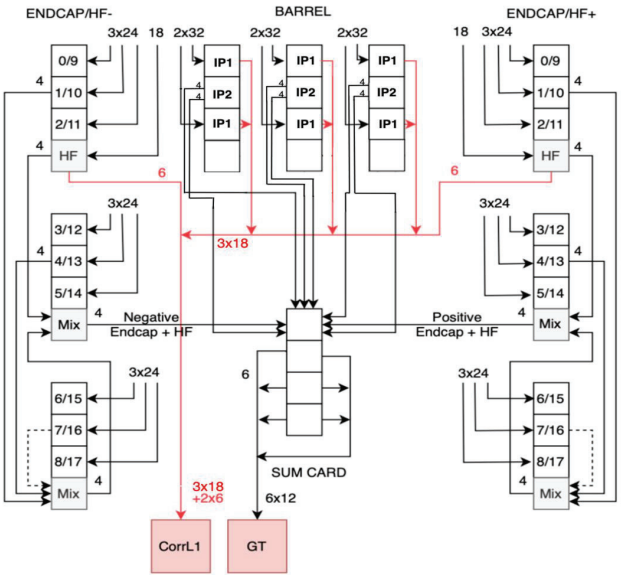


Fig. 6. Global Calorimeter Trigger architecture diagram.

III. Methodology And Algorithmic Implementation

The firmware development used a hybrid approach, with C++ based High Level Syntehsis (HLS) for core algorithms and VHDL for top-level integration and infrastructure. This enabled rapid development of complex algorithms while maintaining fine-grained hardware control.

A. RCT Algorithmic Pipeline

The RCT firmware executes a multi-stage pipeline for high-throughput, low-latency processing.

- **IP1 - Parallel Sub-Region Processing**: Four IP1 cores (3 x IP_5x6 and 1 x IP_2x6) process their sub-regions in parallel. Each core receives ECAL crystal data subdivided into smaller 11×30 crystal regions and executes a cluster-finding algorithm to identify up to nine e/γ candidates. The algorithm locates high-energy "seed" crystals and sums energy from 3×5 crystal matrices around them. For each candidate, shower shape variables are calculated, including energies in 2×5 (E_{2x5}) and 5×5 (E_{5x5}) crystal matrices, along with Bremsstrahlung flags.
- **IP21 - Aggregation, Stitching, and Calibration**: This core receives up to 30 cluster candidates from the IP1 instances. It applies an inter-region stitching algorithm to merge duplicate clusters on sub-region boundaries, as shown in Fig.7. Resulting clusters are sorted by energy using a hardware-optimized 32-element bitonic sorter. The core then implements a p_T-dependent calibration by comparing each cluster's shower shape (SS = E_{2x5} / E_{5x5}) against a threshold from a 25-bin look-up table. This table is accessible via an AXI4-Lite interface, allowing real-time software updates.
- **IP22 & IP3 - Energy Conservation and HCAL Integration**: IP22 reintegrates energy from rejected clusters back into ECAL tower sums to ensure energy conservation. IP3 processes the final ECAL tower data, the top e/γ clusters, and HCAL tower data, calculating the hadronic-to-electromagnetic energy ratio (H/E) for each object. IP3 outputs are duplicated in the top level wrapper to account for the overlap required in the GCT Barrel as shown in Fig. 2 (right).

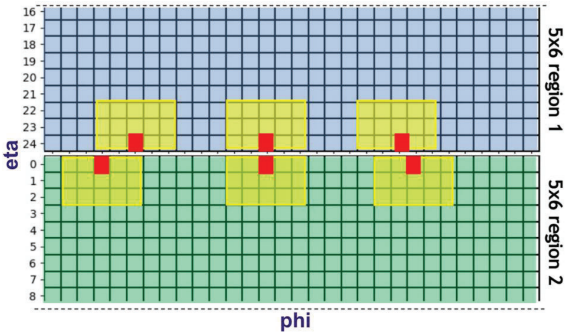


Fig. 7. Illustration of cluster stitching conditions across RCT5x6 region boundaries.

B. GCT Algorithmic Pipeline

The GCT firmware processes aggregated RCT data to perform Particle Flow clustering and generation of higher level physics objects.

- **IP1 - Inter-RCT Stitching and PF Cluster Generation**: Two IP1 core instances (one per SLR) first stitch e/γ clusters across the boundaries of eight RCT regions. Using the assembled tower grid, an iterative algorithm generates Particle Flow (PF) clusters by seeding on the highest-energy tower and forming 3×3 clusters.
- **IP2 - Jet, Tau, and Energy Sum Reconstruction**: A single IP2 core aggregates data from both IP1 instances, stitches information across the η=0 boundary, and reconstructs jets using a 3×3 sliding window algorithm over the SuperTower grid. Tau candidates are identified from isolated SuperTowers. The core also computes barrel-wide scalar and vector energy sums, including total transverse energy and Missing Transverse Energy (MET).
- **TMUX18 Time Multiplexing**: A TMUX18 scheme was developed to reduce the output fiber count. An initial TMUX6 design required 144 fibers going to the correlator layer 1. The improved TMUX18 scheme uses 18 time slots to serialize all objects from two SLRs onto a single fiber path, reducing the fiber count to 54 and eliminating the need for downstream data duplication. The TMUX18 has a modular design with two composite TMUX9 modules along with switching and control logic to get the desired TMUX18 functionality while being resource efficient.

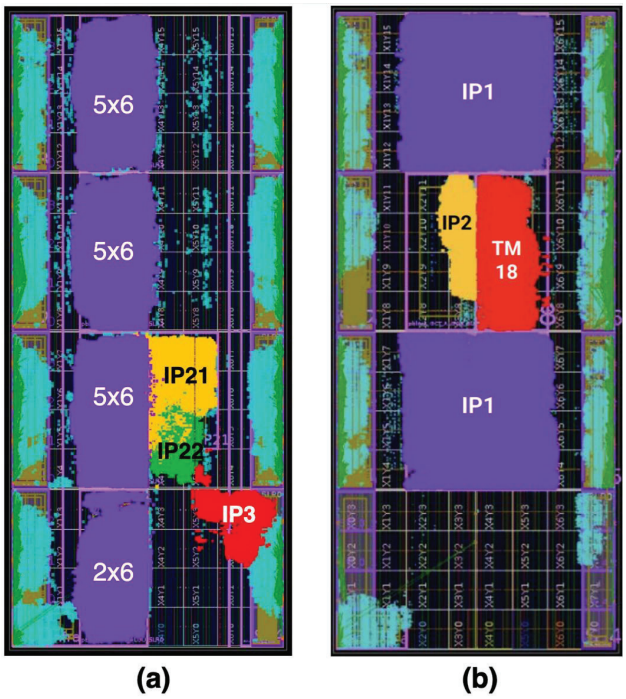


Fig. 8. Final achieved floorplans post-implementation for the RCT (a) and GCT (b) designs, showing logic placement within the specified pblocks.

IV. Implementation And Performance Results

Both firmware designs underwent hardware implementation on APx VU13P-2 platform and underwent validation to meet Phase-2 trigger performance targets.

A. Firmware Implementation Flow

HLS IP cores were synthesized in Out-of-Context (OOC) mode to ensure modularity and reproducible results. A floorplanning methodology using placement blocks (pblocks) was enforced to constrain logic to specific SLRs, avoiding cross-SLR timing issues and meeting DAQ keep-out constraints. The future integration of DAQ firmware necessitates the requirement to keep space along the edge of the FPGA free between the firmware shell and the algo payload, therefore we have a corridor kept empty as shown in Fig. 5 (a). The final, implemented floorplans for both RCT and GCT are shown in Fig. 8. To achieve the 360 MHz timing target, an iterative "PhysOpt looping" strategy was employed, which repeatedly applies physical optimization steps post-placement to reach positive slack before routing. Incremental synthesis and implementation flows were also utilized to accelerate iterations.

B. Performance Metrics

Timing Closure and Resource Utilization: Both RCT and GCT designs achieved timing closure at the 360 MHz target frequency. Per-SLR resource utilization is detailed in Table II. The utilization remained within the sufficient margin for stable timing closure and future modifications.

TABLE II: SLR-WISE RESOURCE UTILIZATION FOR RCT AND GCT-B DESIGNS

Design	SLR	Resource Type			
		LUT(%)	*CLB(%)*	*Regs(%)*	*BRAM(%)*
RCT	0	20.91	51.32	21.26	12.57
	1	29.60	54.24	27.64	13.39
	2	23.36	41.75	21.07	13.39
	3	23.27	42.38	20.91	13.39
GCT	0	3.46	9.61	3.32	1.86
	1	37.06	68.47	32.10	14.29
	2	12.90	39.53	15.44	11.61
	3	37.01	67.84	31.65	14.29

Processing Latency: The pipeline latencies for each IP core and subsystem are listed in Table III. The total RCT pipeline latency measured 0.655 µs, with the GCT pipeline adding 0.444 µs. The combined end-to-end processing latency for the barrel calorimeter trigger chain is 1.1 µs, which is well within the 12.5 µs total L1 trigger budget.

TABLE III: PROCESSING LATENCY FOR RCT AND GCT IP CORES

Subsystem	*IP Core*	*Latency (µs)*
RCT	IP1_5x6	0.364
	IP21	0.133
	IP22	0.075
	IP3	0.083
RCT Total		0.0655
GCT-B	IP1	0.367
	IP2	0.078
GCT Total		0.444
End-to-End Total		**1.099**

a) IP1_2x6 operates in the latency shadow of IP1_5x6, therefore its latency is ignored.

V. System Level Validation

A hierarchical validation strategy confirmed firmware functional correctness from block level to full system level.

A. Standalone and Virtualized System Tests

Standalone single-board tests were conducted on APx hardware by injecting predefined test vectors for limited bunch crossings. Captured hardware output matched RTL simulation output bit-perfectly, validating the individual board implementation. For integrated multi-board validation, the FPGA Environment for Algorithm Slice Tests (FEAST), illustrated in Fig.9 was used [12]. FEAST is an environment for system level FPGA validation that enables virtualized testing of the large, interconnected systems on minimal hardware, as shown in Fig.10. FEAST supports mixed physical and virtual link test systems including link alignment. In the case of all virtual links, one FPGA can evaluate all positions sequentially. The user defines the system and test conditions, and bitfiles run on target FPGAs with target MGT channel allocations. This allows constructing full-scale, multi-layer, virtualized installations and testing on actual FPGAs using the FEAST engine. We defined the 27-board barrel trigger architecture (24 RCTs + 3 GCTs) in a FEAST configuration. The framework emulated the system by sequentially running each board's firmware on physical FPGAs while managing data flow between virtual links. GCT-B outputs from this full-scale test matched expected standalone test results, verifying system integrity.

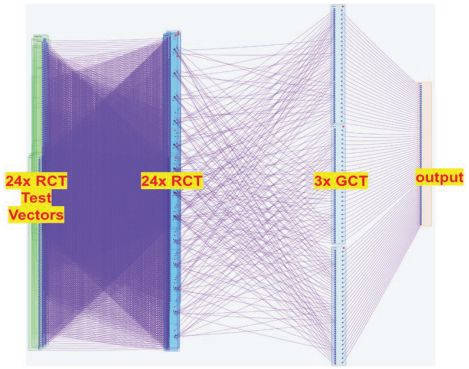


Fig. 9. FEAST for virtual and hybrid end to end system integration and testing.(illustration)

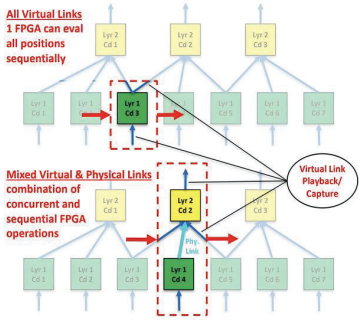


Fig. 10. A FEAST config file showing the 24 RCT - 3 GCT-B system generated and visualized using the CHEF tool.

B. Automation in Verification

The complexity of generating configurations for these large-scale tests required the development of automation tools due to the large number of fiber connections that have to be mapped. These included CHEF (Configuration Helper for Easy FEASTing) for building FEAST configurations, MGT mapping utilities for transceiver configuration, visual floorplanning constraint generators, and test vector generation and formatting tools. These tools reduced manual effort, eliminated errors, and accelerated the design and verification cycle.

VI. Conclusion

We have developed functional, optimized firmware for the RCT and GCT-B subsystems suitable for the high-rate, high-pileup HL-LHC environment. Architectural innovations, including a 17x6 RCT architecture based on the VU13P-2, a real-time software-controlled calibration mechanism, a TMUX18 scheme for the GCT, and automation tools, which were critical to meeting the system's performance goals.

The designs achieve timing closure at 360 MHz with balanced resource utilization as seen in Table II. The 1.1 µs combined processing latency is within the 12.5 µs L1 trigger budget. Multi-layered validation, culminating in full-scale virtualized system testing, provides confidence in the functional correctness and hardware efficaccy of the developed firmware. This work establishes a foundation for the deployment of the CMS barrel calorimeter trigger in the HL-LHC era.

A. Future Developments

Future developments include optimizing the CLB usage below 50% per SLR, utilizing spare capacity on the RCT and GCT barrel to add new algorithms, multi-board tests with physical optical interconnects, benchmarking the design across Vivado / HLS tool versions, improving build times and larger scale FEAST tests including more subsystems.

References

[1] L. Evans and P. Bryant, “LHC Machine,” J. Instrum., vol. 3, no. 08, p.S08001, 2008.

[2] CMS Collaboration, “The CMS experiment at the CERN LHC,” J.Instrum., vol. 3, no. 08, p. S08004, 2008.

[3] CMS Collaboration, “Performance and operation of the CMS electro-magnetic calorimeter,” J. Instrum., vol. 5, no. 03, p. T03010, 2010.

[4] CMS Collaboration, “Performance of the CMS hadron calorimeter withcosmic ray muons and LHC beam data,” J. Instrum., vol. 5, no. 03, p.T03012, 2010.

[5] A. M. Sirunyan et al. (CMS Collaboration), “Performance of the CMS muon detector and muon reconstruction with proton-proton collisions at$\sqrt{s}$ = 13 TeV,” J. Instrum., vol. 13, no. 06, p. P06015, 2018.

[6] B. Acar et al., “Response of a CMS HGCAL silicon-pad elec tromagnetic calorimeter prototype to 20-300 GeV positrons,” 2022,arXiv:2111.06855.

[7] S. Piperov, “Geant4 validation with CMS calorimeters test-beam data,” 2008, arXiv:0808.0130.

[8] Xilinx, “Vivado Design Suite User Guide,” UG896 Vivado Design Suite User Guide: Designing with IP.

[9] P. Kumar and B. Gomber, “The CMS Level-1 Calorimeter Trigger for the HL-LHC,” in EPJ Web of Conf., vol. 295, 2024, p. 02022.

[10] C. Savard on behalf of the CMS Collaboration, “CHEP 2023 Overview of the HL-LHC Upgrade for the CMS Level-1 Trigger,” in EPJ Web of Conf., vol. 295, 2024, p. 02022.

[11] P. Kumar and B. Gomber, “System Design and Prototyping of the CMS Level-1 Calorimeter Trigger at the High-Luminosity LHC,” IEEE Trans. Nucl. Sci., vol. 72, no. 3, pp. 385–391, 2025.

[12] I. Ojalvo, “The APx Board for the CMS Phase 2 L1 Calorimeter trigger: Testing and Performance,” presented at TWEPP, 2024.

[13] A. Mohan, “FPGA Automation Toolkit,” 2022. [Online]. Available: https://abhinavm2000.github.io/fpga-automation-toolkit/.

[14] G. Daughtry, “Top 5 Timing Closure Techniques,” presented at the Club Vivado Users Group, Paris, France, 2015.

[15] CMS Collaboration, “The Phase-2 Upgrade of the CMS Level-1 Trigger,” CERN-LHCC-2020-004; CMS-TDR-021, 2020.

[16] CMS Collaboration, Technical Proposal for the Phase-II Upgrade of the CMS Detector,” CERN-LHCC-2015-010; LHCC-P-008, 2015.

[17] G. Apollinari et al., High-Luminosity Large Hadron Collider (HL-LHC): Technical Design Report V. 0.1,” CERN-2017-007-M, 2017.

[18] Xilinx, “UltraScale Architecture and Product Data Sheet: Overview,” DS890, 2021.

[19] D. Acosta et al., “The CMS Level-1 Trigger for the High Luminosity LHC,” J. Instrum., vol. 12, no. 01, p. C01065, 2017.

[20] S. Dasu et al., “CMS Level-1 Calorimeter Trigger Upgrade for the HL-LHC,” in Proc. IEEE Nucl. Sci. Symp. Med. Imag. Conf., 2022, pp.1-3.