

Imperial College London
Department of Physics

Design and implementation of on-detector and off-detector electronics for the high-luminosity upgrade of the CMS endcap calorimeter system within LHC at CERN

Miloš Vojinović

A dissertation submitted to Imperial College London
for the degree of Doctor of Philosophy

August 2024





The copyright of this thesis rests with the author. Unless otherwise indicated, its contents are licensed under a Creative Commons Attribution-Non Commercial 4.0 International Licence (CC BY-NC). Under this licence, you may copy and redistribute the material in any medium or format. You may also create and distribute modified versions of the work. This is on the condition that: you credit the author and do not use it, or any derivative works, for a commercial purpose. When reusing or sharing this work, ensure you make the licence terms clear to others by naming the licence and linking to the licence text. Where a work has been adapted, you should indicate that the work has been changed and describe those changes. Please seek permission from the copyright holder for uses of this work that are not included in this licence or permitted under UK Copyright Law.

Statement of originality

I declare that the work presented in this thesis is my own. At times, the nature of the work required close collaboration with others and using devices which I did not design or create. This declaration explicitly specifies which of the mentioned contributions are entirely mine, partially mine, and not mine. The contributions of others have been appropriately cited throughout the thesis, where suitable. Chapters 1 and 2 are not my work, but are written in my own words giving an overview of (and the motivation for) the LHC, CMS, high-luminosity LHC, and various CMS upgrades required by the increased luminosity, focusing on HGCal and its complex electronics system. The design and creation of the various individual items (e.g. ASICs, PCBs, the Serenity board and its firmware) used in the vertical HGCal electronics test system described in Chapter 3 is not my work and are contributions from several different institutes. However, the design and implementation of this test system, as well as ensuring that this variety of devices operate together is my contribution. These items come from various international institutes and were only, if at all, tested individually at their respective home institutes before being sent to CERN. Prior experience in operating them together was minimal or non-existent. Therefore, this work necessitated communication with experts and institutions from where these devices originated. The vertical test system upgrades summarised in Chapter 4 are contributions made by the HGCal back-end teams (from CERN and Imperial College London). This work is only partially mine. Specifically, I provided the L1A Throttle firmware block and defined the design specifications for the e-Link monitoring firmware and software. I do not claim personal ownership of the beam test results shown in Section 4.1.3 and published in [1] and [2], but rather attribute them to the HGCal Collaboration. The discovery and the mitigation of the VTRx+ saturation issue presented in Section 4.3 was a collaborative effort with the CERN VTRx+ and lpGBT teams. Setting the BER upper limits in Section 4.3 on the DAQ and TPG paths is my own work. The analysis of the system-wide digital consistency in Section 4.4 is my work which was an extension of the previous studies and was based on an existing software infrastructure provided by the HGCal BE group. The work reported in Chapter 5 is my own. The simulations of deadtime are an extension to the previous work of the Imperial College HGCal BE group who also first specified the requirement for an L1A throttling block for the TCDS2. Regarding Chapter 6, I defined the: modified TC record; Cluster Sum record, Cluster Property record; Accumulator “+” operator; general three-step algorithm sequence and studied its physics performance and quantisation effects. When it comes to firmware development, my work is limited to the Property Calculator. The results of the HGCal $\rightarrow$ L1T interface tests come from a collaboration of the corresponding HGCal TPG BE and the L1T teams who drove these activities, and of which I was a part of.

Miloš Vojinović

Acknowledgements

My heartfelt appreciation goes to United Kingdom's Research and Innovation (UKRI) Science and Technology Facilities Council (STFC) and CERN, without whose joint funding this research would not have been possible.

I would like to extend my sincere gratitude to the CERN EP-CMX group and the HEP group within the Department of Physics at Imperial College London for providing me the opportunity to pursue my PhD. Access to their resources, facilities, and academic environment has greatly contributed to both my research and personal development.

My deepest gratitude goes to my university and CERN supervisors, Prof. P. Dauncey and Prof. A. David. Thank you both for believing in me as well as giving me this amazing opportunity, as well as your invaluable patience, support and advice. I could not have chosen better mentors. I will always cherish all the little wins throughout this journey and remember them as ours.

I sincerely thank Dr. R. Shukla, Dr. A. Steen, Dr. S. Mallios, Mr. M. Rosado, Ms. A. Alpana, and Mr. M. A. Şarkışla from the HGCAL back-end and system validation groups for your support, stimulating discussions, and for creating the most pleasant working environment. My special thanks go to Dr. R. Shukla - I will always remember the remarkable achievements we accomplished together. I am also grateful to Dr. J. Langford and Dr. I. Das for providing the data used for the cluster property studies and to Dr. Ö. Sahin for his role in the clock phase noise measurements in the HGCAL vertical test system. Ms. Ana, thank you for all the coffees. Additionally, I extend this gratitude to all other colleagues from the CMS and HGCAL collaborations who supported this research.

Dr. D. Barney, thank you for being the first to present the possibility of this PhD.

Prof. A. David, Prof. P. Dauncey, Ms. C. Knight, Mr. Y. A. Hakim and Ms. A. Alpana, thank you for all the hours spent in helping this electronics engineer become a physicist.

Harry, thanks for everything mate. The following text is in Serbian for my Serbian-speaking tutors, friends and family.

Хвала свим мојим учитељима, наставницима и професорима. Посебно хвала Проф. П. Михаиловићу и Проф. Д. Ел Мезенију са Електротехничког факултета у Београду.

Мојим другарима из детињства: Нешо, Маре, Уки, Ранко, Фићо, Јована, Бели, Стефане, Срећко, Немања и Дуле; као и другарима из студентских дана у Београду: Бојана, Јелена, Мико, Тања, Лазо и Данило - хвала вам на свој подршци, разумевању мог неприсуства, а највише на бескрајно занимљивим разговорима и незаборавним искуствима која граде карактер.

Мојој најдражој екипи из ЦЕРН-а: Пеђа, Амаре, Иване, Ники, Томо, Јелена, Ана, Лазаре Ц., Ирена, Драгана, Лазаре Н., Михаило и Марија - хвала вам на свим незаборавним искуствима и подршци. Нарочито хвала Амару и Пеђи, који сте били ту кад је било најтеже.

Милана, највише хвала теби, за сву пружену доброту и љубав, као и бескомпромисну и неограничену подршку - поготово у тренуцима када ми је била најпотребнија.

Мама, Тата, бака Кајо, бака Ино, деда Миле и деда Мида, хвала вам за бескрајну љубав, подршку и неговоње моје радозналости од малих ногу. Ујка Деки и Даре, као и тета Весна хвала вама и вашим породицама на свој љубави и подршци.

Викторе и Машо - ви сте следећи!

Бака Kaju.

To my grandmother Kaja.

*“Знаш ли ти како се постаје прави шампион?
Тако што изађеш на терен кад је најтеже... и победиш.”
Из филма “Небеска удица” (Режисер: Љубиша Самарџић, 2000.)*

*“Do you know how to become a true champion?
By stepping onto the court when it is the hardest... and winning.”
From the film “Sky Hook” (Director: Ljubiša Samardžić, 2000.)*

Abstract

The CMS experiment requires various technical upgrades to sustain its performance whilst coping with a harsher radiation environment and having to resolve a greater number of proton-proton interactions per beam cross-over, both brought by the increased luminosity projected for the high-luminosity LHC. One key upgrade is replacing the existing endcap calorimeter with a novel device, the High Granularity Calorimeter (HGCAL). This thesis will focus on two mutually related aspects considering the design and implementation of its electronics system.

The first aspect is about a test system that prototypes a full vertical slice (start-to-end) of the detector electronics chain. As the final system, this includes both on-detector (front-end) and off-detector (back-end) components, as well as trigger and DAQ sub-systems. Key operational aspects including fast control, timing distribution, slow control, trigger and DAQ were activated and tested. This allowed operating the system in a beam test, measuring a data transfer bit error rate upper limit of $O(1 \times 10^{-16})$, and evaluating the digital agreement between the different devices in the chain. These results mark the first-ever implementation of an HGCAL readout system prototype of such complexity.

The second aspect involves the design and FPGA implementation of two back-end algorithms, including simulations to estimate their physics performance. The first algorithm emulates front-end readout chip buffers and occasionally throttles CMS Level-1 trigger decisions to prevent data loss due to buffering limitations. An analysis showing this algorithm results in approximately 1.5×10^{-8} deadtime will be presented. The second algorithm calculates 3D cluster properties (energy, position, shape) from the front-end energy deposits for use in the CMS Level-1 trigger. It will be shown that this algorithm has excellent and very good ability to identify photons and Vector Boson Fusion jets, respectively. Both blocks were implemented in software and firmware, and successfully tested in hardware.

Contents

List of Figures	i
List of Tables	xi
List of Abbreviations	xiii
1 Introduction	1
1.1 The Large Hadron Collider	1
1.2 The Compact Muon Solenoid	3
1.3 The LHC High Luminosity Upgrade	8
1.4 Upgrading CMS for HL-LHC	10
1.5 The High Granularity Calorimeter	12
1.6 Scope of the Thesis	15
2 The HGCal Electronics System	17
2.1 General System Overview	17
2.1.1 Timing Distribution and Fast Control	18
2.1.2 Slow Control	18
2.1.3 Data Acquisition	19
Trigger Data Acquisition	19
DAQ Data Acquisition	19
2.1.4 Different layers of data transfer in HGCal	19
The Very FE Layers (HGCal Specific)	20
The Layers in the FE ↔ BE Interface (not HGCal Specific)	21
2.1.5 A Catalogue of Different Hardware Parts	21
2.2 The HGCal Electronics FE	23
2.2.1 The Hexagonal Silicon Modules	23
2.2.2 The HGCROC ASIC	24
Timing and Fast Control	25
Slow Control	25

	The ADC, ToT and ToA measurements	25
	Functionality Overview - TPG Path	27
	Functionality Overview - DAQ Path	28
2.2.3	The ECON-T ASIC	30
	Timing and Fast Control	31
	Slow Control	31
	Functionality Overview	31
2.2.4	The ECON-D ASIC	33
	Timing and Fast Control	34
	Slow Control	34
	Functionality Overview	34
2.2.5	The lpGBT ASIC	36
	The Down-link Data Path	37
	The Up-link Data Path	38
2.3	The HGCAL Electronics DAQ BE	39
2.3.1	System Structure and Organisation	40
	Data Acquisition Aspects	40
	The Usable Part of the DAQ Bandwidth	41
	Timing, Fast Control and Slow Control Aspects	41
2.3.2	Functionality Overview	41
	Slow Control	41
	Timing Distribution and Fast Control	42
	Capture Blocks	42
	Readout Blocks	43
2.4	The HGCAL Electronics TPG BE	44
2.4.1	System Structure and Organisation	44
	Organisation into Sectors and Time Multiplexing	45
	The FE $\rightarrow$ Stage 1 Interface	45
	Stage 1 and Stage 2 Responsibilities	45
	Duplication Regions at Sector Boundaries	46
	The Stage 1 $\rightarrow$ Stage 2 and Stage 2 $\rightarrow$ L1T Interfaces	46
	Latency Considerations	46
2.4.2	Stage 1 Functionality Overview	47
	Data Unpacker	47
	Trigger Cell Processing	48
	Formation of Partial Tower Energies	48

	Data Formatting and Time Multiplexing	48
2.4.3	Stage 2 Functionality Overview	49
	Data Unpacker	50
	Complete Tower Energies	50
	Cluster Finding	50
	Cluster Properties	51
	Filter/Buffer	51
	Data packer	51
2.5	A Note on Different Latency Contributions	52
3	Implementing a Vertical Slice of the HGCAL Electronics	53
3.1	The Architecture of the System	54
3.2	Fast Control and Timing Distribution	55
3.2.1	Fast Command Generation	57
3.2.2	Synthesis of the Required BE Clock Sources	57
3.2.3	The Main Interfaces in the Signal Flow	57
	Locking the FE $\leftrightarrow$ BE lpGBT Interface	57
	Timing Distribution	58
	Fast Control Distribution	59
3.2.4	Evidence of Fast Control and Timing Distribution	59
	Monitoring the ECON-D Registers	59
	General Notes on Eye Diagrams and their Sampling Phase	61
	Oscilloscope Measurements	61
3.3	Slow Control	63
3.3.1	The Main Interfaces in the Signal Flow	63
3.3.2	The Software Architecture	63
3.3.3	Evidence of Slow Control	65
	Writing to and reading from the FE ASIC Registers	65
	Oscilloscope Measurements	66
3.4	Data Acquisition of DAQ and TPG Data	68
3.4.1	The Main Interfaces in the Signal Flow	69
3.4.2	e-Link Phase and Word Alignment Procedures	69
	The Initial Phase Training Method	69
	The HGCROC $\rightarrow$ ECON e-Link Alignment	72
	The ECON $\rightarrow$ BE e-Link Alignment	73
3.4.3	Evidence of Data Flow through the whole Vertical Chain	74
	The Main FE ASIC Configuration Choices	74

	The FE DAQ Data brought to and output by the DAQ BE	75
	The FE TPG Data brought to and output by the TPG BE	78
	Oscilloscope Measurements	80
3.5	The System Initialisation Procedure	81
3.6	Results Summary	81
4	Gaining a Deeper Understanding of the HGCal Vertical Slice	83
4.1	The Beam Test	84
4.1.1	The SPS North Area Beam Parameters	84
4.1.2	Upgrading the Vertical System for the Beam Test	85
4.1.3	The Most Notable Beam Test Results	87
4.2	The Key Interfaces for Studying Error Rates and Digital Consistency	91
4.3	The Error Rate Measurements	93
4.3.1	The VTRx+ Saturation Impact on the lpGBT $\rightarrow$ BE Path	94
4.3.2	The BER Upper Limits in the HGCROC $\rightarrow$ BE Path	98
	The TPG Path	99
	The DAQ Path	101
4.4	System-wide Digital Consistency Analysis	102
4.4.1	The TPG Path	103
4.4.2	The DAQ Path	103
	General Parameters of the DAQ Run	103
	The SLink Headers and Trailers	104
	The DAQ BE Headers	107
	The ECON-D Headers	107
	The ECON-D Sub-packet Headers	110
	The L1A Timestamps	112
	Closing the DAQ System-wide Digital Consistency Analysis	117
4.5	Results Summary	117
5	The HGCal L1A Throttle for TCDS2	119
5.1	The Motivation for L1A Throttling	119
5.1.1	The purpose of the HGCROC RAM 2 FIFO	120
5.1.2	The RAM 2 Limitation and CMS Deadtime	121
5.1.3	The Deadtime Theoretical Minimum	121
5.1.4	Proposing L1A Throttling for TCDS2	122
5.2	The Software Model	124
5.2.1	The Simulation Flowchart	124

5.2.2	The Most Significant Results	127
	Agreement with the HGCROC ASIC	127
	Deadtime vs. L1A frequency	129
5.3	The Firmware Implementation	130
5.3.1	General Overview	131
5.3.2	Agreement with the Software Model	132
	Standalone behavioural simulation and FPGA hardware tests	132
	Protecting HGCROCs in the Vertical Test System	133
	Deadtime vs. L1A frequency	134
5.4	Results Summary	134
6	Calculating Cluster Properties in the HGCal TPG back-end	137
6.1	The Motivation for Cluster Properties	137
6.2	A General Overview of the Algorithm	140
6.2.1	The Core Mechanism	140
6.2.2	The Type Modifier	141
6.2.3	The Accumulator	143
6.2.4	The Property Calculator	146
	Cluster $\mathbf{E_T}$ in various longitudinal regions	146
	Cluster $\mathbf{E_T}$ fractions in various longitudinal regions	146
	Cluster $\mathbf{E_T}$ weighted mean positions	146
	Cluster shape properties	146
	Miscellaneous cluster properties	148
6.3	The Software Model	148
6.3.1	Agreement between the Emulator and the Simulator	150
6.3.2	Physics Performance of the Cluster Properties	151
	General workflow in the study	151
	Limitations and possible optimisations of the study	152
	Dropping some clusters and properties before training the BDT	153
	Selecting the clusters by $\mathbf{p_T^{gen}}$ and $\mathbf{E_T^{reco}}$	153
	Bin-based sample re-weighting in cluster $\mathbf{E_T^{reco}}$	155
	Organising the clusters into $\mathbf{p_T^{gen}}$ sub-ranges	157
	Importance of features in the trained BDT	157
	Discrimination power of the top-performing features	158
	Testing the BDT	160
	Quantisation effects	161
6.4	The Firmware Implementation	164
6.4.1	General Overview	164
6.4.2	Agreement with the Software Model	166
6.5	Results Summary	166

7 Summary and Conclusions	169
Bibliography	175
A Examples of BX-level Behaviour of the HGCAL L1A Throttle	183
B Cluster Property Quantisation Effects	187
C Dependence of Cluster Properties on k	191

List of Figures

1.1	An illustration of the LHC collider below the Geneva area with its four IP and the SPS as an injector [5]. The acronyms used are explained in the text.	1
1.2	A simplified diagram of the stages of particle acceleration for the proton-proton collisions at the LHC (from [6], after [3]). The acronyms used are explained in the text.	2
1.3	An overview of CMS's densely packed architecture [13].	3
1.4	The CMS coordinate system [14].	4
1.5	A cross-section view of the various detector layers showing different particle trajectories and their types of signatures in each detector [15].	4
1.6	A cross section view of the CMS Tracker [16].	5
1.7	A 3D schematic view of the CMS ECAL [9].	6
1.8	A cross section view of the CMS HCAL [9].	6
1.9	A cross section view of the CMS Muon system [9].	7
1.10	Block diagram of the architecture of the CMS L1T system [9].	8
1.11	Illustration of the interaction region in support of the luminosity definition in a collider experiment [5].	9
1.12	Schedule of interleaved periods of data collection and long shutdowns with a projection for the instantaneous and integrated luminosities and collision energies at the LHC [21].	9
1.13	(a) Expected radiation levels absorbed in the existing calorimetry systems if they were to be used in the HL-LHC conditions; shown here, at an integrated luminosity of 3000 fb^{-1} . The 1 MeV-neutron equivalent fluence represents the number of 1 MeV neutrons per unit area that would cause the same amount of damage as the actual mixed radiation sources. Provided by the CMS FLUKA simulator [29]. (b) An illustration of a typical signature (simulated) of energy deposits inside HGCal and general vertex density in a single HL-LHC BX [32].	12
1.14	HGCal endcap calorimeter layout in 3D (Left) [29] and its longitudinal cross section (Right) [20].	13
1.15	Assembly of silicon-based HGCal modules (Left) [33], arrangement of silicon and scintillator parts of HGCal in its front and back layers (Middle) [29] and assembly of scintillating tiles in the scintillator-based HGCal modules (Right) [33].	13
2.1	The HGCal electronics system and the complex variety of its interfaces with the CMS global BE systems illustrated in a simplified block diagram [1]. The F/O link counts are given per endcap (i.e. there are two copies of the HGCal Electronics layout in the final system). The acronyms used are explained in the text.	18

2.2	A logical block diagram of the key layers of data transfer in the HGICAL Electronics System. Any degree of parallelism that will exist in the final system is abstracted. The acronyms used are explained in the text.	20
2.3	A catalogue of different hardware parts existing in the very complex HGICAL Electronics System, from the FE silicon modules, to the BE boards. Acronyms used are explained in the text.	22
2.4	(a) An example layout of a silicon-only HGICAL layer showing different module density and thickness choices [29]. Assembly of an (b) LD and (c) HD hexagonal module [46]. In the layers of the vertical stack-up in (b) and (c) from the bottom up are the copper base-plate, resistive kapton foil, Si module, Hexaboard PCB, respectively. Acronyms used are explained in the text.	24
2.5	A block diagram of the key DAQ and TPG data manipulation steps (Colour) and other blocks related to timing, control and readout within the HGCROC ASIC (Gray) [36]. The relevant acronyms used are explained in the text.	25
2.6	Output waveforms of the HGCROC (a) Pre-Amplifier in a range of charge deposits and (b) Shaper in a range of charge deposits [36].	26
2.7	An illustration of several close BXs with energy deposits and their ToA and ToT measurements [36]. The acronyms used are defined in the text.	26
2.8	An illustration of how TCs are formed in the LD (Left) and HD (Right) Hexaboard layouts [36]. More details can be found in text.	28
2.9	The format of the HGCROC trigger frame for one of the output e-Links [36]. The acronyms used are explained in the text.	28
2.10	Generalised 32-bit HGCROC channel format [38]. The “ROC Fields” are specified by the three “x” symbols from the relevant row in Table 2.1.	29
2.11	The format of the HGCROC DAQ data about an event, as present on one output e-Link across 41 BX [36]. The acronyms used are explained in the text.	30
2.12	The block diagram of ECON-T internal algorithms [37].	30
2.13	An illustration of how STCs are formed in the LD Hexaboard layouts for the STC4 (Left) and STC16 (Right) ECON-T algorithms [37]. More details can be found in text.	32
2.14	An example of the ECON-T STC4 general data format where the ECON-T has 5 available output e-Links [37]. The individual fields are explained in the text. . . .	32
2.15	An example of the ECON-T BC high occupancy data format where the ECON-T has 4 or more available output e-Links and is transmitting 9 or more TCs per BX [37]. Individual fields are explained in the text.	33
2.16	The block diagram of ECON-D internal algorithms [38].	33
2.17	The standard ECON-D data format for N active eRx [38]. The individual fields are explained in the text.	35
2.18	The Pass-through ECON-D data format for N active eRx [38].	36
2.19	The truncated ECON-D data format [38].	36
2.20	A block diagram of the lpGBT architecture [40]. Relevant acronyms are explained in the text.	37
2.21	The lpGBT down-link 64-bit frame format [40]. Individual fields are explained in the text.	38

2.22	The lpGBT up-link 256-bit frame format specific to the HGCAL use case [40]. Individual fields are explained in the text.	39
2.23	Structure of the DAQ BE system for one endcap. At the time the block diagram was drawn it was taken as a worst case scenario that there will be 10 000 links (50% of them dark) routed between the FE and the DAQ BE and that a stage of patch panels will also be needed. Since then matters have improved and the BE may not even need any patch panels. Regardless, the DAQ BE is prepared to connect to as many as 5184 FE lpGBT links [34].	40
2.24	A block diagram of the main algorithm blocks present in one DAQ BE board. The coloured blocks are HGCAL specific [34].	42
2.25	DAQ BE Header indicating the L1A timestamp (bunch counter, event counter and orbit counter, according to the DAQ BE notion of time) and status of the 12 ECON-D packet status (e.g. normal, missing, rejected due to counter mismatch, unconnected ECON-D, etc.).	43
2.26	SLink packet format [47]. More details can be found in text.	44
2.27	Structure of the TPG BE system for one endcap, corrected from [34].	45
2.28	A block diagram of the main algorithms on a TPG BE Stage 1 FPGA, corrected from [34].	47
2.29	A block diagram of the main algorithms on a TPG BE Stage 2 FPGA [34]. . . .	49
2.30	Latency requirements for different parts of the system responsible for the generation and distribution of the trigger decision [49]. Acronyms used other than the ones defined in text: pTTs (Partial Towers), TP (Trigger Primitives).	52
3.1	A block diagram of the implemented vertical test system.	54
3.2	A block diagram of relevant components and interfaces in the vertical HGCAL electronics test system from the perspective of timing distribution and fast control. Details can be found in text.	56
3.3	An example of the Serenity BE EMP TX Capture buffer snapshot, showing the hexadecimal format of the 8-bit binary codes of (a) BCR (0b00011101 = 0x1d) followed by (b) IDLE (0b00110110 = 0x36) and (c) L1A (0b01001011 = 0x4b) fast commands intercepted on their way towards the FE. Due to BE implementation details, the relation between the frame numbers and the DAQ BE BX counter BC for a given fast command is $(BC - 1) \times 8 + 35 = \text{Frame}$, with the exception for $\text{Frame} = 27$ when $BC = 3564$, meaning that the fast commands in the examples (a), (b) and (c) are generated at $BC = 3564$, $BC = 1$ and $BC = 111$, respectively.	59
3.4	A snapshot of the terminal of the Serenity BE Board Controller where the values of the fast control and timing monitoring registers from Table 3.2 were read from the ECON-D ASIC. The DAQ BE fast control firmware was configured to generate each one of the fast commands (that can be counted by the ECON-D) once (with the exception of the BCR and the L1A which were being generated once per orbit) which is reflected in the ECON-D register values.	60
3.5	A qualitative representation of an eye diagram and its main parameters.	61
3.6	Waveforms of the 320 Mbits^{-1} fast command data stream (Top Right), the 320 MHz fast control clock (Bottom Right) and the eye diagram of the fast control clock (Left) measured with an oscilloscope at the input of an HGCROC ASIC on the Hexaboard PCB.	62

3.7	Measured phase noise distribution between the BE external clock generator and the output of the Rafael ASIC on the Hexaboard PCB [51] with a standard deviation of $\sigma_{\text{PhaseNoise}} = 8.5$ ps. More details can be found in text.	62
3.8	A block diagram of relevant components and interfaces in the vertical HGCAL electronics test system from the perspective of slow control. Details can be found in text.	64
3.9	A simplified and generalised representation of the data transfer layers in the slow control path. The software architecture is driven by and tries to mirror the hardware architecture [52].	65
3.10	A snapshot from the terminal of the BE Board Controller, illustrating <i>write</i> and <i>read</i> slow control transactions with the FE ASICs [52]. For every ASIC, the value written to the arbitrarily chosen register is also read back, as expected. Transactions with an HGCROC, the ECON-T, the ECON-D, the TPG lpGBT and the DAQ lpGBT are highlighted in red, yellow, green, blue and purple, respectively.	66
3.11	Scoped I ² C (a) Step 1 (b) Step 2 and (c) Step 3 transaction waveforms when writing 0x0A with the TPG lpGBT into an internal register of an HGCROC at the address 0x0B07 in the vertical test system and (d) a general time diagram of an I ² C bus to help interpret the meaning of individual bits in the waveforms [39]. The operating frequency of the transactions was set to 200 kHz. More details can be found in text.	67
3.12	A block diagram of relevant components and interfaces in the vertical HGCAL electronics test system from the perspective of DAQ data acquisition. Details can be found in text.	70
3.13	A block diagram of relevant components and interfaces in the vertical HGCAL electronics test system from the perspective of TPG data acquisition. Details can be found in text.	71
3.14	A snapshot from the terminal of the Board Controller reporting the status of the ECON-D (similar applies to ECON-T) input e-Link Word Alignment procedure. More details can be found in text.	73
3.15	A snapshot from the terminal of the Board Controller showing the contents of the EMP TX Capture buffer for one BX worth of the ECON-T (similar applies to ECON-D) data after successful word alignment in the BE. Each row represents the aligned contents of each of the four used ECON-T output e-Links. With respect to Equation 3.4, the BC field (e.g. BC = 0b01000 = 8) and the expected 11-bit constant pattern 0b00110101010 = 0x1aa are clearly exposed for every ECON-T output e-Link.	74
3.16	One BX worth of data at the output of the DAQ BE captured in the EMP TX Capture buffer with highlighted DAQ BE Header, ECON-D Packet Header, ECON-D Sub-Packet Header, HGCROC Channel Data and the ECON-D CRC. More details can be found in text.	76
3.17	One BX worth of data at the output of the DAQ BE that has flowed through the HGCROC → ECON-D → DAQ lpGBT → DAQ BE steps in the vertical test system. The HGCROC ADC vs. Channel ID plot was built from the raw data of a single event captured by the EMP TX Capture Buffer. The ADC vs. channel dependence is as expected given Equation 3.5.	78

3.18	One BX worth of data at the output of the TPG BE captured in the EMP TX Capture buffer in hexadecimal format (Top Left) and the relevant data in binary format (Top Right) and a table for decoding the meaning of individual bits (Bottom). Highlighted are the twelve STC Energies, Addresses of Maximum Energy TCs within a given TC and corresponding BX counters. More details can be found in text.	79
3.19	One BX worth of data at the output of the TPG BE that has flowed through the HGCROC $\rightarrow$ ECON-T $\rightarrow$ TPG lpGBT $\rightarrow$ TPG BE steps in the vertical test system. The STC Value vs. STC ID plot was built from the raw data of a single BX captured by the EMP TX Capture Buffer. The STC Value vs. STC ID dependence is as expected given Equation 3.7.	80
3.20	Eye diagram of (a) HGCROC $\rightarrow$ ECON-T 1.28 Gbit s ⁻¹ e-Link and (b) VTRx+ $\rightarrow$ Serenity BE 10.24 Gbit s ⁻¹ TPG F/O up-link.	80
3.21	A simplified block diagram of the vertical test system initialisation flow.	82
4.1	The North Area (a) layout and (b) particle “spill” time profile [61]. More details can be found in text.	85
4.2	A block diagram of the implemented beam test system and its major interfaces [1]. The system consists a vertical slice of the HGCAL electronics presented in Chapter 3 and a variety of upgrades (more details can be found in text).	85
4.3	Module histograms of (a) the mean STC values using TPG data and (b) the standard deviation of the ADC count per channel using DAQ data. The two plots come from different electron beam runs [1]. Generated using offline analysis tools processing real physics data acquired during the beam test. More details can be found in text.	88
4.4	Signal amplitude as a function of time on (a) the TPG path for a single Hexaboard STC [2] (b) the DAQ path for a single Hexaboard cell [1]. Generated using offline analysis tools processing real physics data acquired during the beam test. More details can be found in text.	89
4.5	Arrival time distributions when using external / scintillator trigger (blue) and internal / ECON-T trigger (red) L1A generation logic [2]. Generated using offline analysis tools processing real physics data acquired during the beam test. More details can be found in text.	90
4.6	Distribution of the HGCROC channel retention by the ECON-D ASIC when configured in (a) Pass-through and (b) Zero-suppression mode from an electron beam run [1]. Generated using offline analysis tools processing real physics data acquired during the beam test. More details can be found in text.	91
4.7	A general block diagram of relevant components and interfaces in the vertical HGCAL electronics test system used for the error rate measurements and the DAQ data analysis. Details can be found in text.	92
4.8	In focus (not greyed-out parts) of this block diagram are the components and interfaces of the vertical HGCAL electronics test system used for the error rate measurements on the TPG (green) and DAQ (purple) paths and the relevant slow control interface to monitor the results (yellow) when the origin of data is in the lpGBT. Details can be found in text.	95

4.9	DAQ lpGBT e-Link word error counter (Top) word error counter difference (Middle) and word error rate (Bottom) monitored over a period of 15 min when the down-link optical attenuation <i>was not</i> applied to prevent the VTRx+ from occasionally entering the saturation regime. More details can be found in text.	96
4.10	Eye diagram of the DAQ F/O up-link when the down-link optical attenuation was not (a) and was (b) applied to prevent the VTRx+ from occasionally entering the saturation regime. More details can be found in text.	97
4.11	A block diagram of the relevant components in the system architecture related to measuring the DAQ lpGBT F/O down-link eye diagram quality. More details can be found in text.	97
4.12	In focus (not greyed-out parts) of this block diagram are the components and interfaces of the vertical HGCAL electronics test system used for the error rate measurements on the TPG (green) and DAQ (purple) paths and the relevant slow control interface to monitor the results (yellow) when the origin of data is in the HGCROC. Details can be found in text.	100
4.13	In focus of this block diagram (not greyed-out parts) are the components and interfaces on the DAQ (purple) path of the vertical HGCAL electronics test system used for the DAQ data analysis. Details can be found in text.	105
4.14	Distribution of pattern validity of (a) BOE and (b) EOE fields in SLink packet Headers and Trailers, respectively. More details can be found in text.	106
4.15	Distribution of the agreement between the DAQ BE CRC and CRC fields in SLink packet Trailers. More details can be found in text.	106
4.16	Distribution of the BX ID field in SLink packet Trailers. More details can be found in text.	107
4.17	Distribution of the ECON-D Packet Status field in DAQ BE packet Headers. More details can be found in text.	107
4.18	Distribution of the Payload Length field in ECON-D packet Headers. More details can be found in text.	108
4.19	Distribution of (a) P and (b) T flags in ECON-D packet Headers. More details can be found in text.	108
4.20	Distribution of the validity of the ECON-D Header, protected by the CRC for Event Header field. More details can be found in text.	109
4.21	Distribution of RR field values in the ECON-D packet Headers. More details can be found in text.	109
4.22	Distribution of the E flag in ECON-D packet Headers. More details can be found in text.	110
4.23	Distribution of (a) H/T and (b) E/B/O fields in ECON-D packet Headers. More details can be found in text.	110
4.24	Distribution of the Stat field in ECON-D Sub-packet Headers corresponding to one input ECON-D e-Link (eRx 0) or one HGCROC half. More details can be found in text.	111
4.25	Distribution of (a) Stat and (b) Ham fields in ECON-D Sub-packet Headers corresponding to one input ECON-D e-Link (eRx 0) or one HGCROC half. More details can be found in text.	112

4.26	A diagram showing the ideal case of relative bunch and orbit counter values for the SLink, HGCROC, ECON-D and the DAQ BE for a given BX. More details can be found in text.	112
4.27	A diagram showing the used relative bunch and orbit counter values for the SLink, HGCROC, ECON-D and the DAQ BE for a given BX. More details can be found in text.	113
4.28	Distribution of (a) the M flag in ECON-D packet Headers and (b) the ECON-D M flag values as a function of the SLink bunch counter values, focused on the part of the orbit where M was 0. More details can be found in text.	114
4.29	Distribution of (a) the ECON-D event counter as a function of the DAQ BE event counter and (b) the ECON-D bunch counter as a function of the DAQ BE bunch counter. More details can be found in text.	114
4.30	Distribution of (a) the ECON-D orbit counter as a function of the DAQ BE orbit counter and (b) the results of the ECON-D and DAQ BE orbit counter match comparison as a function of the SLink bunch counter values, focused on the part of the orbit where mismatches were observed. More details can be found in text.	115
4.31	Distribution of (a) the DAQ BE event counter as a function of the SLink event counter and (b) the DAQ BE bunch counter as a function of the SLink bunch counter. More details can be found in text.	115
4.32	Distribution of (a) the DAQ BE orbit counter as a function of the SLink orbit counter and (b) the results of the DAQ BE and SLink orbit counter match comparison as a function of the SLink bunch counter values, focused on the part of the orbit where mismatches were observed. More details can be found in text.	116
4.33	Distribution of (a) the ECON-D event counter as a function of the SLink event counter and (b) the ECON-D bunch counter as a function of the SLink bunch counter. More details can be found in text.	116
4.34	Distribution of (a) the ECON-D orbit counter as a function of the SLink orbit counter and (b) the results of the ECON-D and SLink orbit counter match comparison as a function of the SLink bunch counter values, focused on the part of the orbit where mismatches were observed. More details can be found in text.	117
5.1	A block diagram showing the major parts of the HGCROC logic related to DAQ and control paths [36]. In focus (highlighted in blue) are the two HGCROC buffers (RAM 1 and RAM 2) and its readout logic. More details can be found in text.	120
5.2	Simulated distribution of the consecutive L1A distances when $\langle f_{L1A} \rangle = 750$ kHz.	120
5.3	The theoretical minimum (HGCROC hard limit) for Dendtime as a function of $\langle f_{L1A} \rangle$ derived from the limitation of the HGCROC output bandwidth for y-axis in (a) linear and (b) logarithmic scale. More details can be found in text.	122
5.4	A conceptual block diagram in support of the proposed TCDS2 L1A throttling scheme. More details can be found in text.	123
5.5	A functional time diagram of the signal waveforms in support of the proposed TCDS2 L1A throttling scheme. More details can be found in text.	124
5.6	A flowchart of the software simulation. Each iteration of the loop corresponds to one BX being processed, starts by receiving the final L1A value from the previous BX and ends by optionally throttling the L1A candidate value of the current BX. Black arrows represent the chronological order of steps, whereas brown arrows represent dependencies between steps. More details can be found in text.	126

5.7	A block diagram illustrating the main workflow when the agreement between the software model of the HGCAL L1A throttle and the HGCROC ASIC were tested one against the other. More details can be found in text.	128
5.8	Software simulated deadtime (blue) as a function of the average L1A frequency $\langle f_{L1A} \rangle$ over a (a) wide range of frequencies (b) narrow range of frequencies in the vicinity of the nominal 750 kHz. The former also has the HGCROC hard limit (grey) from Section 5.1 drawn for reference. More details can be found in text. . .	130
5.9	A simple schematic of the HGCAL L1A Throttle Generator firmware block with its interfaces and their types. More details can be found in text.	131
5.10	A typical time diagram with input and output waveforms of the HGCAL L1A Throttle Generator firmware block. The 40 MHz clock <code>clk_40</code> is not part of the block interface, but is drawn to illustrate its time dependence to <code>i_clk_40_en</code> and <code>i_clk_320</code> . More details can be found in text.	131
5.11	A resource utilisation report for the implemented HGCAL L1A Throttle Generator firmware block (highlighted in blue). Extracted from the vertical test system BE firmware project implemented in Vivado (the official Xilinx FPGA algorithm development framework) for the Xilinx KU15P FPGA. More details can be found in text.	132
5.12	A timing report for the implemented HGCAL L1A Throttle Generator firmware block in Vivado. More details can be found in text.	132
5.13	Software simulated (blue) and hardware measured (brown) Deadtime as a function of the average L1A frequency $\langle f_{L1A} \rangle$ over a (a) wide range of frequencies (b) narrow range of frequencies in the vicinity of the nominal 750 kHz. The former also has the HGCROC hard limit from Section 5.1 drawn for reference. More details can be found in text.	135
6.1	Qualitative examples of clusters of TCs originating from (a) PU (b) photon (c) VBF jet illustrating their usual signatures in the HGCAL TPG path. The sizes of spherical entries in the plots are proportional to the E_T of the corresponding TCs. To give a more balanced comparison, the three shown clusters all come from similar p_T and η ranges. Data is extracted from CMSSW simulation. More details can be found in text.	138
6.2	An illustration of the discrimination power of two example cluster properties when used to distinguish clusters originating from (a) photons and (b) VBF jets against clusters originating from PU. Data is extracted from CMSSW simulation. The explicit property definitions, details of the conducted study and x-axis units can be found later in the chapter. More details can be found in text.	139
6.3	The three main steps in the Stage 2 Cluster Properties block. More details can be found in text.	140
6.4	Agreement between the calculated quantities by the software emulator (firmware implementation/reconstructed) with the software simulator (true) for several cluster properties. For axis units, see Table 6.4. More details can be found text.	150
6.5	Distributions of cluster E_T (reconstructed) as a function of p_T (generator-level) for (a) photon and (b) VBF jet clusters. More details can be found in text. . . .	154
6.6	Distributions of cluster p_T^{gen} (Top) and E_T^{reco} (Bottom) for the photon (Left) and VBF jet (Right) identification studies after applying p_T^{gen} and E_T^{reco} cuts. More details can be found in text.	155

6.7	Plots demonstrating the mechanism of bin-based sample weighting of cluster E_T^{reco} distributions so that the one of signal matches the one of pileup clusters. Shown are the original E_T^{reco} distributions (Top), computed bin-based sample weights (Middle), and E_T^{reco} distributions after applying weights (Bottom) for the photon (Left), and the VBF jet (Right) identification studies. More details can be found in text.	156
6.8	Distributions of cluster p_T^{gen} for the (a) photon and (b) VBF jet identification studies after organising the p_T^{gen} into sub-ranges with non-linear binning. More details can be found in text.	157
6.9	Breakdown of importance of different cluster properties used as features when training a BDT for (a) photon and (b) VBF jet identification studies. More details can be found in text.	158
6.10	Pair plots allowing to visualise the discrimination power of some of the top-performing cluster properties in the: (a) photon (b) VBF jet identification studies.	159
6.11	The ROC curves as a function of p_T^{gen} and their AUC scores listed for (a) photon and (b) VBF jet identification binary classifiers. The meaning of dashed lines is clarified in the discussion of the subsequent figure. More details can be found in text.	160
6.12	The TPR (green) and FPR (red) as a function of p_T^{gen} (for their respective FPR and TPR requirements) for the (a) photon and (b) VBF jet identification binary classifiers. More details can be found in text.	161
6.13	The ROC curves as a function of p_T^{gen} and their AUC scores (Top) and the corresponding TPR as well as FPR plots (Bottom) for the photon (Left) and VBF jet (Right) identification binary classifiers. Calculated using the software Emulator (the k parameter is specified in the text box). More details can be found in text.	162
6.14	Visualising the effect of the k parameter on the performance of cluster properties when: $k = 4$ (Top), $k = 7$ (Middle), and $k = 15$ (Bottom). Shown for the two studied binary classification problems: photon vs. pileup (Left) VBF jet vs. pileup (Right).	163
6.15	A simple schematic of the Property Calculator firmware block with its interfaces and types. More details can be found in text.	164
6.16	A simple time diagram with input and output waveforms of the Property Calculator firmware block. More details can be found in text.	165
6.17	A simplified algorithm flow diagram illustrating the key sequence of steps taking place in the firmware responsible block for calculating the E_T weighted $\mu(c)$ cluster property, where $c \in \{\phi, r/z, z\}$. More details can be found in text.	165
6.18	A resource utilisation report for the an implemented firmware project which houses the Property Calculator block in Vivado (the official Xilinx FPGA algorithm development framework). Highlighted are the values of used resources for the Property Calculator itself, relative to the total available Xilinx VU13P FPGA resources. More details can be found in text.	166
6.19	A timing report for an implemented firmware project which houses the Property Calculator block in Vivado. More details can be found in text.	166
B.1	Agreement between the calculated quantities by the software emulator (firmware implementation/reconstructed) with the software simulator (true) for several cluster properties (Subset 1). For axis units, see Table 6.4.	187

B.2	Agreement between the calculated quantities by the software emulator (firmware implementation/reconstructed) with the software simulator (true) for several cluster properties (Subset 2). For axis units, see Table 6.4.	188
B.3	Agreement between the calculated quantities by the software emulator (firmware implementation/reconstructed) with the software simulator (true) for several cluster properties (Subset 2). For axis units, see Table 6.4.	189
C.1	Visualising the effect of the k parameter on the precision of $\sigma_{\phi\phi}$ for (a) $k=1$, and (b) $k=4$. For axis units, see Table 6.4.	191
C.2	Visualising the effect of the k parameter on the precision of $\sigma_{\phi\phi}$ for (a) $k=7$, (b) $k=15$, and $\mu(\phi)$ for (c) $k=1$, (d) $k=4$, (e) $k=7$, (f) $k=15$. For axis units, see Table 6.4.	192

List of Tables

1.1	Several LHC beam parameters (by design) for proton-proton collisions [5]. The LHC is currently operating above the design luminosity, particles per bunch, and pileup shown here.	3
1.2	Several notable HL-LHC design beam parameters for proton-proton collisions [22]. It should be mentioned that while the given peak luminosity value here corresponds to the technically achievable maximum, the nominal value is $7.5 \times 10^{34} \text{ cm}^{-2} \text{ s}^{-1}$ (as given in the text).	10
2.1	Selection of three 10-bit digital values for the DAQ data for a given BX on one channel. The choices depend on which thresholds the signal has passed in the given BX and the mode the HGCROC is configured to be in [36]. The “x” marks the set of selected values for any given signal criteria.	28
3.1	The HGCal Fast Command Dictionary [35].	58
3.2	The ECON-D fast control and timing monitoring registers [38]. The 1-bit registers convey that the description is <i>true</i> when their value is 1 and <i>false</i> when 0. The 8-bit counter register can monitor the number of received fast commands (or fast command errors) in the range between 0 to 255. All of the registers are sub-nodes of the <code>FCtrl.Global</code> ECON-D register.	60
4.1	The configuration of the TPG e-Link Monitor firmware blocks for each of the four ECON-T output e-Links (eTx). More details can be found in text.	101
6.1	Trigger cell definition. More details can be found in text.	142
6.2	Modified Trigger cell definition, where $c \in \{\phi, r/z, z\}$. The bit-widths of W , W^2 , Wc and Wc^2 are listed assuming $k = 0$ in the W definition. More details can be found in text.	143
6.3	Cluster Sum Definition, where $c \in \{\phi, r/z, z\}$. The bit-widths of W , W^2 , Wc and Wc^2 are listed assuming $k = 0$ in the W definition. More details can be found in text.	144
6.4	Cluster Properties definition as per the HGCal→L1T interface specification. . .	147
A.1	A BX-level emulation of the internal counters, states and status flags of the HGCal L1A throttle block when presented with a burst sequence of 45 consecutive L1A requests when the throttling decision is not taken into account (i.e. disabled). It should be noted that the event counter <code>EC</code> is incremented for every L1A received <code>l1a_in</code> . However, when the occupancy limitation is hit (i.e. <code>events_RAM2 = 33</code>) RAM 2 could not accept additional events. Generated using the HGCal L1A throttle software model. Different colours encode different FSM states.	184

- A.2 A BX-level emulation of the internal counters, states and status flags of the HGCAL L1A throttle block when presented with a **burst sequence** of 45 consecutive L1A requests when the throttling decision **is taken into account** (i.e. enabled). It should be noted that the event counter **EC** is not incremented for every L1A received **l1a_in**, but only for those events that actually entered RAM 2. Generated using the HGCAL L1A throttle software model. Different colours in the FSM column encode different FSM states. The highlighted cells in red in the throttle column indicate BXs when the throttle signal was raised. 185
- A.3 A BX-level emulation of the internal counters, states and status flags of the HGCAL L1A throttle block when presented with L1A requests arriving when **special words are being streamed out of the HGCROC**, to demonstrate some edge cases of the readout logic. Specifically, these L1A requests arrive when the output words are: IDLE due to empty buffer (**output_word** = 0), CRC (**output_word** = 40), the last event word (**output_word** = 39), mandatory idle (**output_word** = 41), both CRC and mandatory IDLE back-to-back (**output_word** = 40 and **output_word** = 41). Generated using the HGCAL L1A throttle software model. Different colours in the FSM column encode different FSM states. 186

List of Abbreviations

ADC	Analogue to Digital Converter
ALICE	A Large Ion Collider Experiment
ASIC	Application Specific Integrated Circuit
ATCA	Advanced Telecommunications Computing Architecture
ATLAS	A Toroidal LHC Apparatus
AUC	Area Under Curve
BC	Best Choice algorithm in ECON-T / Bunch Counter (context dependant)
BCR	Bunch Counter Reset
BDT	Boosted Decision Tree
BE	Back-End
BER	Bit Error Rate
BX	Bunch Crossing
cDAQ	Central DAQ
CDR	Clock and Data Recovery
CE-E	Calorimeter Endcap - Electromagnetic section
CE-H	Calorimeter Endcap - Hadronic section
CERN	European Organisation for Nuclear Research
CL	Confidence Level
CM	Common Mode
CMS	Compact Muon Solenoid
CMSSW	CMS offline software framework
CRC	Cyclic Redundancy Check
CSC	Cathode Strip Chamber
CSP	CMS Standard Protocol
DAQ	Data Acquisition

DCS	Detector Control System
DOI	Domain Of Influence
DQM	Data Quality Monitoring
DT	Drift Tube
DTH	DAQ and TCDS2 Hub
EBR	Event Buffer Reset
EC	External Control / Event Counter (context dependant)
ECAL	Electromagnetic Calorimeter
ECON-D	DAQ e-Link Concentrator
ECON-T	TPG e-Link Concentrator
ECR	Event Counter Reset
e-Links	Electrical Links
EMP	Extensible Modular data Processor
EMTF	Endcap Muon Track Finder
eRx	e-Link Receiver
eTx	e-Link Transmitter
FE	Front-End
FEC	Forward Error Correction
FIFO	First In First Out
FNR	False Negative Rate
F/O	Fibre Optic
FPGA	Field Programmable Gate Array
FPR	False Positive Rate
FSM	Finite State Machine
GEM	Gas Electron Multipliers
GPIO	General-Purpose Input / Output
HB	HCAL Barrel
HCAL	Hadronic Calorimeter
HD	High Density
HE	HCAL Endcap
HF	HCAL Forward
HGCAL	High Granularity Calorimeter

HGCROC HGCAL Readout Chip
HL-LHC High Luminosity LHC
HLT High-Level Trigger
HO HCAL Outer
I²C Inter-Integrated Circuit
IC Internal Control
IP Interaction Point
L1A Level-1 Accept
L1T Level 1 Trigger
LAN Local Area Network
LD Low Density
LEP Large Electron-Positron collider
LHC Large Hadron Collider
LHCb Large Hadron Collider beauty
LINAC Linear Accelerator
lpGBT Low Power Gigabit Transceiver
LS Long Shutdown
LSB Least Significant Bit
LUT Look Up Table
MIP Minimum Ionising Particle
ML Machine Learning
MSB Most Significant Bit
MTD MIP Timing Detector
NIM Nuclear Instrumentation Module
NZS Non-Zero-Suppressed
OC Orbit Counter
OCR Orbit Counter Reset
PA Pre-Amplifier
PbWO₄ lead tungstate
PCB Printed Circuit Board
PLL Phase Locked Loop
PS Proton Synchrotron

PSB Proton Synchrotron Booster
PU Pileup
RCC Run Control and Configuration
RF Radio Frequency
ROC Receiver Operating Characteristic
RPC Resistive Plate Chamber
RX Receiver
SCL I²C Serial Clock
SDA I²C Serial Data
SEU Single Event Upset
SH Shaper
SM Standard Model
SNR Signal-to-Noise-Ratio
SPS Super Proton Synchrotron
STC Super Trigger Cell
TC Trigger Cell
TCDS Timing and Control Distribution System
TCDS2 Phase-2 Timing and Control Distribution System
TCS Trigger Control System
TDC Time Digital Converter
TEC Tracker Endcap
TIB Tracker Inner Barrel
TID Tracker Inner Disk
TNR True Negative Rate
ToA Time of Arrival
TOB Tracker Outer Barrel
ToT Time Over Threshold
TPG Trigger Primitive Generator
TPR True Positive Rate
TX Transmitter
VBF Vector Boson Fusion
VTRx+ Versatile Link+ Transceiver
WER Word Error Rate

Chapter 1

Introduction

This chapter is organised as follows. Firstly, it introduces the Large Hadron Collider (LHC) accelerator complex (Section 1.1). Secondly, it gives a high-level overview of the existing Compact Muon Solenoid (CMS) detector (Section 1.2). Thirdly, it introduces the high-luminosity upgrade plans for the LHC (Section 1.3). This is followed by a brief summary of the main ongoing upgrades of the various sub-systems of the CMS detector, in regard to the luminosity increase plans (Section 1.4). It then specifically summarises the replacement of its current endcap calorimeter system with a new high-granularity detector (Section 1.5). Lastly, the remainder of the thesis is laid out, giving an overview of its structure and contents (Section 1.6).

1.1 The Large Hadron Collider

At the European Organisation for Nuclear Research (CERN), situated approximately 100 m beneath the Franco-Swiss border lies the largest and most powerful accelerator complex built to this date, the LHC (Figure 1.1) [3]. Since it was first turned on in September 2008, the LHC continues being the leading tool for studying the fundamental particles and forces that make up the universe. The primary aims of the LHC are to explore the Standard Model (SM) in the TeV range, as well as the revelation of the origin of particle mass and possible new physics signatures [4].

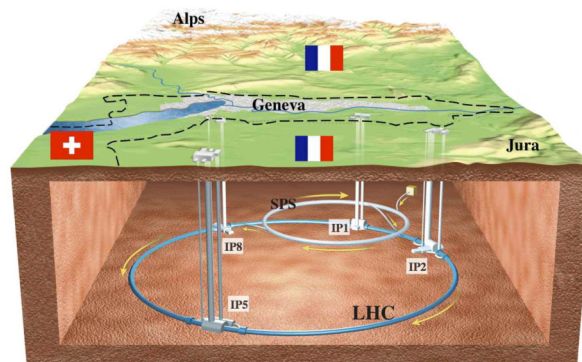


Figure 1.1: An illustration of the LHC collider below the Geneva area with its four IP and the SPS as an injector [5]. The acronyms used are explained in the text.

The LHC occupies a tunnel that has a circumference of 27 km, previously used by the Large Electron-Positron (LEP) collider. The two beam pipes allow the two beams of protons (or sometimes heavy ions) of the same charge to be accelerated in opposite directions and collided in the four Interaction Points (IPs), where the beam pipes cross over and the main particle detectors are placed.

The most important parameters to consider about any particle collider are the maximum collision energy and the beam luminosity (the number of beam particles going through an area per second; more details can be found in Section 1.3). The former puts a limit on the mass scale of the particles that can possibly be produced in collisions, whereas the latter limits the probability of particle interactions in a given beam crossover. By design, the LHC can collide beams of protons with a centre-of-mass energy of 14 TeV and beam luminosities of $10^{34} \text{ cm}^{-2} \text{ s}^{-1}$.

The beam characteristics in the LHC prior to particle collisions are a result of a process that happens in stages of linear and circular acceleration, and beam shaping (Figure 1.2) [3]. The journey starts by injecting negative hydrogen ions (H^-) into the linear accelerator (LINAC) where they reach energies of 160 MeV. During injection from the LINAC into the Proton Synchrotron Booster (PSB), the electrons are stripped of the ions, leaving only the positively charged protons. The PSB accelerates the protons to an energy of 2 GeV. This stage is followed by the Proton Synchrotron (PS) where they reach 26 GeV. Following this is the Super Proton Synchrotron (SPS) which accelerates the particles up to energies of 450 GeV. Ultimately, the protons enter the two beam pipes of the LHC where, by design, they can reach energies of up to 7 TeV per beam (14 TeV collision energy). The current record for proton-proton collision energies is 13.6 TeV. Throughout these stages, a variety of sophisticated and precisely controlled Radio Frequency (RF) instrumentation is used to accelerate, but also organise the protons into close “bunches”, focus the beam, direct the beam, etc., until the nominal beam parameters are reached (Table 1.1).

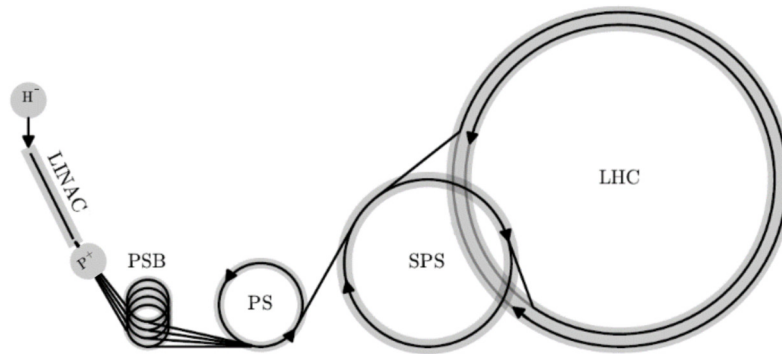


Figure 1.2: A simplified diagram of the stages of particle acceleration for the proton-proton collisions at the LHC (from [6], after [3]). The acronyms used are explained in the text.

The whole system is carefully synchronised so that, at any given IP, the Bunch Crossing (BX) period is 25 ns. Consequently, the BX frequency is 40 MHz. On average there are about 40 simultaneous proton-proton interactions per BX, commonly referred to as “pileup” (PU) ¹. Around the four IPs, four detectors are put in place: A Large Ion Collider Experiment (ALICE) [7]; A Toroidal LHC ApparatuS (ATLAS) [8]; CMS [9] and Large Hadron Collider beauty (LHCb) [10]. The ALICE and LHCb experiments specialise in

¹A PU of 40 is the nominal value, however presently this number goes as high as 60 to 65.

heavy-ion physics and the slight differences between matter and antimatter, respectively. On the other hand, CMS and ATLAS are the two general-purpose detectors, widely known for their joint discovery of the Higgs boson in 2012 [11][12].

Table 1.1: Several LHC beam parameters (by design) for proton-proton collisions [5]. The LHC is currently operating above the design luminosity, particles per bunch, and pileup shown here.

Parameter	Value
Momentum at collisions	7 TeV
Peak luminosity	$10^{34} \text{ cm}^{-2} \text{ s}^{-1}$
Number of bunches	2808
Particles per bunch	1.15×10^{11}
Maximum PU	40

1.2 The Compact Muon Solenoid

The CMS [9] is a general-purpose detector at the LHC. The 14 000 t device spans about 30 m in length and 15 m in diameter, whilst having cylindrical geometry. It has a compactly layered (“onion-like”) structure (Figure 1.3), consisting of multiple sub-detectors; each one designed for their specific task. The detector is built around the 18 000 A superconducting solenoid that produces a magnetic field inside the solenoid of 3.8 T. The solenoid houses the Tracker, the Electromagnetic Calorimeter (ECAL), and Hadronic Calorimeter (HCAL). The magnetic field outside, with an intensity of up to 2 T, returns through (and is confined by) a 12 500 t steel “yoke” that contains the muon chambers. This section presents the CMS detector in several steps. Firstly, it defines its commonly used coordinate system. Secondly, it gives a high level summary of the usual particle signature types in the detector. This is followed by describing each of the major sub-detectors in a little more detail. Lastly, the CMS triggering system is briefly summarised.

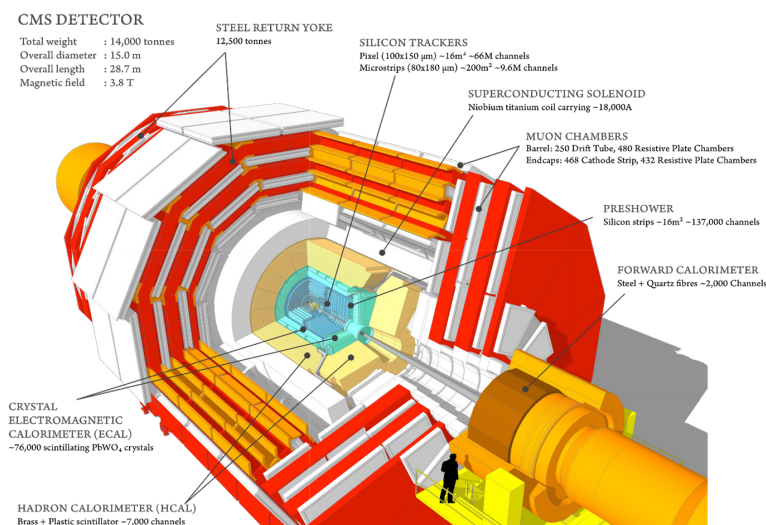


Figure 1.3: An overview of CMS’s densely packed architecture [13].

The Coordinate System The CMS coordinate system (Figure 1.4) is defined with the origin at the nominal interaction point inside the experiment. The x -axis points towards the centre of the LHC ring, the y -axis points vertically upwards, and the z -axis is aligned along the beam direction, with the positive direction corresponding to the counterclockwise beam. The azimuthal angle ϕ is measured from the x -axis in the xy -plane, and the polar angle θ is measured from the z -axis. Pseudorapidity, defined as $\eta = -\ln [\tan(\theta/2)]$, is often used instead of the polar angle. The radial distance r is defined as $r = \sqrt{x^2 + y^2}$, representing the distance from the z -axis in the transverse plane. The transverse momentum p_T is the momentum component perpendicular to the beam axis.

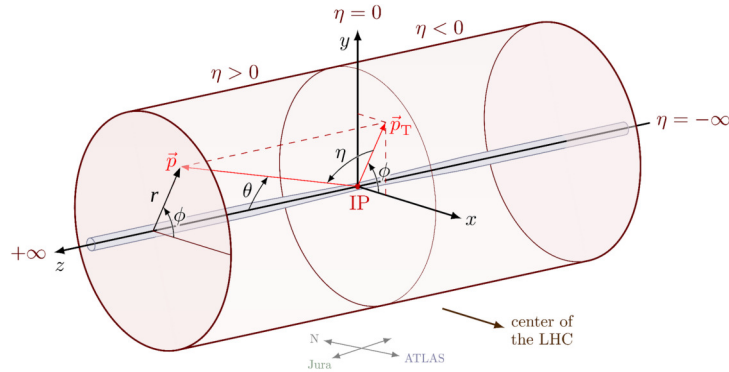


Figure 1.4: The CMS coordinate system [14].

Different Signature Types When a particle is produced in a collision (Figure 1.5), it first goes through the silicon layers of the Tracker module. If it is charged, its trajectory is bent due to the presence of the magnetic field. Given the bending direction, one can identify the sign of the particle's charge, whereas the curvature of the track allows measurement of the particle's momentum. Next in its path is the ECAL, designed to measure the energy of photons and electrons. It is followed by the HCAL, used to measure the energy of hadrons. Photons and electrons completely deposit their energy in the ECAL, whereas hadrons mainly do so in the HCAL layers. However, this is not the case for muons. They pass through the two calorimeters without being stopped.

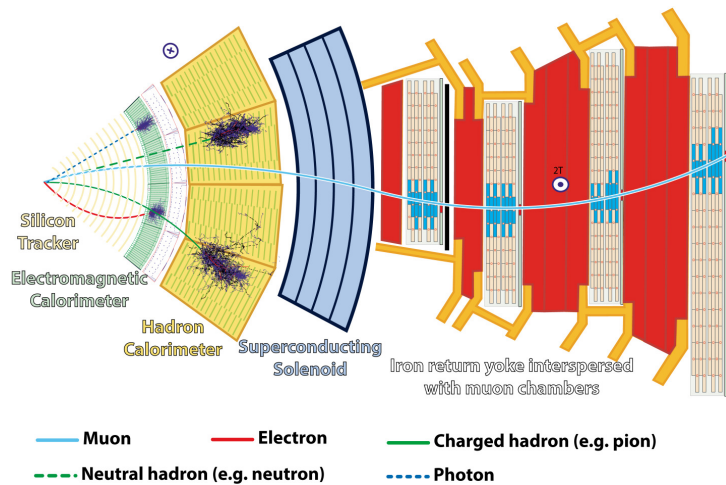


Figure 1.5: A cross-section view of the various detector layers showing different particle trajectories and their types of signatures in each detector [15].

Therefore, specialised muon chambers are used to detect these particles. In fact, CMS was designed to be particularly good at detecting muons (hence the middle name). Particles of the same charge going through the Tracker, ECAL and HCAL have their trajectories bent in one direction, whereas that is the opposite in the muon chambers. This is because of the contrasting directions of the magnetic field lines inside and outside of the superconducting solenoid.

The Tracker The Tracker [9] is designed to measure the trajectories of charged particles with high precision (Figure 1.6). The choice of the used active material is silicon, ensuring that the design requirements regarding spatial granularity, speed and radiation hardness are met. The device consists of an inner and outer section, overall contributing to a highly granular detection system. The inner section, known as the Pixel Detector, contains about 100 million silicon pixels, providing excellent spatial resolution close to the interaction point. The outer section, the Silicon Strip Tracker, is made up of about 10 million silicon strips, ensuring precise tracking over a larger volume. It is made up of several sub-systems: the Tracker Inner Barrel (TIB); Tracker Inner Disk (TID); Tracker Outer Barrel (TOB); and Tracker Endcap (TEC). As charged particles traverse the Tracker, electron-hole pairs

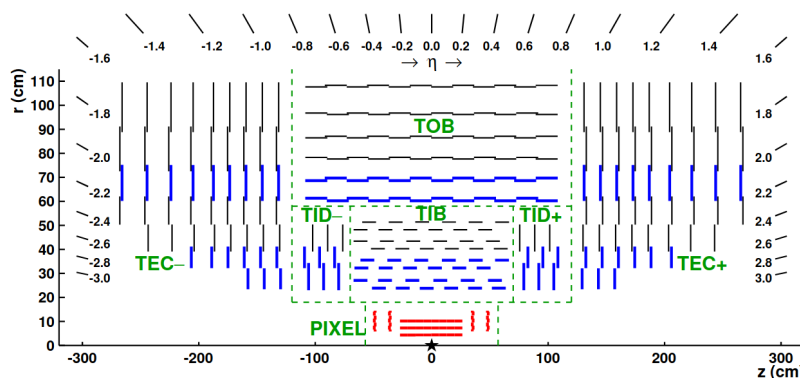


Figure 1.6: A cross section view of the CMS Tracker [16].

are created in the silicon. These charge carriers are collected by the electrodes on the silicon strips or pixels, generating signals that allow for the reconstruction of the paths of the particles. A set of hits can then be used to reconstruct the curved paths of charged particles, allowing to determine their transverse momentum p_T by analysing the radius of curvature. The acceptance range of the Tracker is $|\eta| < 2.5$.

The ECAL The ECAL [9] is a homogeneous calorimeter, meaning that it is made of a single, uniform material that serves as both the absorber and the active medium (Figure 1.7). It is composed of 61 200 lead tungstate (PbWO_4) crystals in the central barrel section, complemented by 7324 crystals in each of the two endcaps. A preshower detector, which mainly specialises in detecting $\pi^0 \rightarrow \gamma\gamma$ decay, is also placed in front of each of the two endcaps. As high-energy electrons or photons pass through, they collide with the heavy nuclei of the ECAL crystals. Consequently, showers of electrons, positrons, and photons are generated, not only from the initial interactions but also from secondary, tertiary, and further subsequent interactions. These particles excite the atoms inside the crystals causing scintillation light to be produced. This light is “sensed” using photo-detectors which produce electrical signals proportional to the energy of the passing

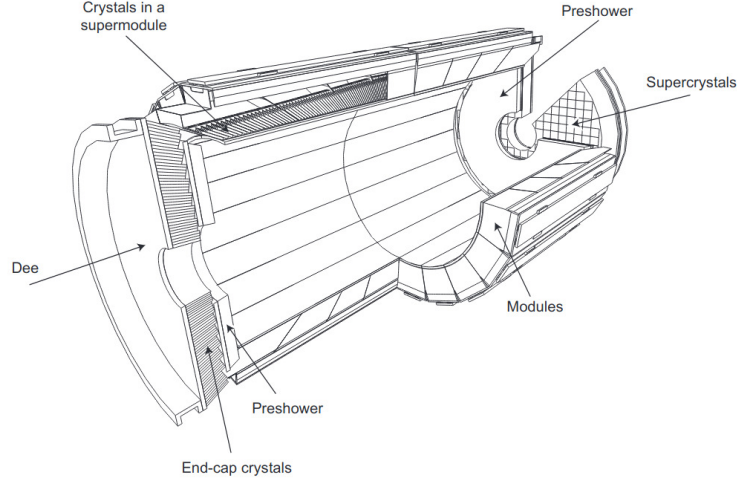


Figure 1.7: A 3D schematic view of the CMS ECAL [9].

particles. The focus of the ECAL endcap is to accurately measure the energy deposited by the electromagnetic showers and, thanks to the small Molière radius (about 2 cm [9]) and the large number of crystals, characterise their lateral shape in fair detail. The ECAL also provides timing information, however it does not provide longitudinal information about electromagnetic showers. The acceptance range of the ECAL is $|\eta| < 3.0$.

The HCAL The HCAL [9] is a sampling device, meaning that it has interleaved layers of absorber (steel/brass) and scintillating material (plastic). There are 18 scintillating layers in total. It sits behind the Tracker and the ECAL, looking from the IP (Figure 1.8). The HCAL is comprised of 9072 readout channels distributed across four sub-systems: the barrel part (HB; 2592 channels), the endcap part (HE; 2592 channels), the outer part (HO; 2160 channels), and the forward part (HF; 1728 channels) [17]. Given that the envelope of its volume is constrained by other sub-systems, so is the total amount of material that stops the hadronic showers. Hence, the outer part serves as a “tail catcher” and sits outside the solenoid to complement the barrel part. When hadrons (i.e. kaons, pions, protons, etc.) hit the absorber plates, they create showers of secondary particles.

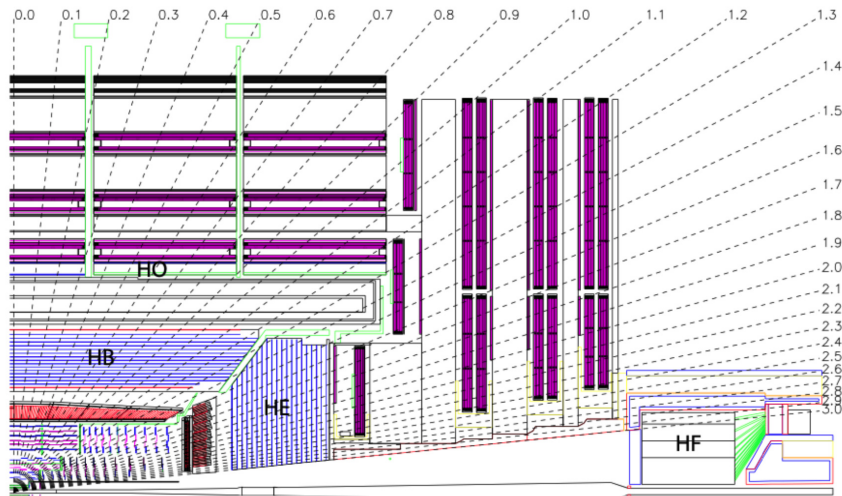


Figure 1.8: A cross section view of the CMS HCAL [9].

These products continue to shower as they go through the next absorber layer and so the shower develops. When the particles move through the scintillating material, they excite it. The emitted light is “sensed” in photo-detectors and converted to an electrical signal. The HCAL endcap measures the timing, energy and position (both lateral and longitudinal) of hadronic showers in fair detail. The acceptance range of the HCAL is $|\eta| < 5.2$.

The Muon Chambers The muon system [9] utilises three kinds of gaseous detectors: Drift Tubes (DTs) in the barrel region; Cathode Strip Chambers (CSCs) in the endcap regions; and Resistive Plate Chambers (RPCs) in both the barrel and the endcap regions (Figure 1.9). The system has about 600 000 readout channels in total. When a muon passes through, it ionises the gas, and the liberated electrons drift towards the anode wires under an electric field. The time taken for the electrons to reach the anode wires is then measured. Consequently, this allows to extract precise information on the position of the muon. A collection of hits can then be utilised to reconstruct the curved trajectories of charged particles, allowing to determine their transverse momentum p_T by examining the radius of the curvature. The acceptance range of the muon system is $|\eta| < 2.4$.

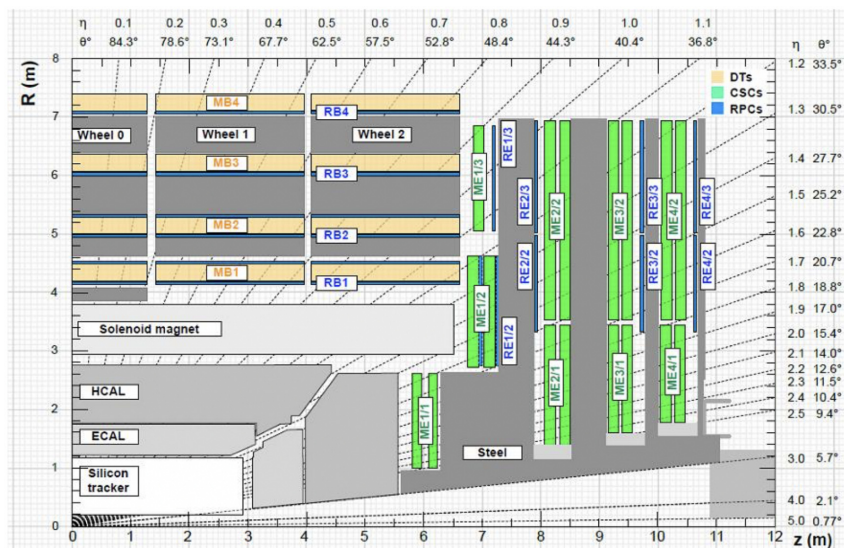


Figure 1.9: A cross section view of the CMS Muon system [9].

The CMS Trigger System Another important aspect of the CMS detector is its Trigger system. The accelerated protons collide every 25 ns, leaving tracks of where and when the collision products went and how much energy they deposited inside the detector. Hence, one can think of CMS as a digital camera that takes a 5D “picture” of a collision with a frequency of 40 MHz. Here, the five dimensions are the three spatial coordinates, time and energy (although not all dimensions are measured in every part of the detector volume). Every such “picture” has approximately 100×10^6 pixels (detector channels) [18] and its average size is about 1 MB [19] (assuming zero-suppression). However, not all BXs produce interesting events and data about every BX cannot be stored for offline analysis. Therefore, the task of the trigger system is to scan through the information generated by each collision and decide which can be discarded and which should be analysed in greater detail. More specifically, the CMS Trigger system is a two-level system consisting of the

Level 1 Trigger (L1T) and the High-Level Trigger (HLT), where the former operates with data of coarse granularity (trigger data / trigger primitives) and the latter does so with data of fine granularity (event data). Moreover, the L1T receives trigger primitives from ECAL, HCAL and the Muon Chambers but does not do so from the Tracker (Figure 1.10). These quantities are the energy deposits (the so called trigger tower energies) in the bins of a 2D (η, ϕ) grid as well as the hit patterns coming from the muon chambers.

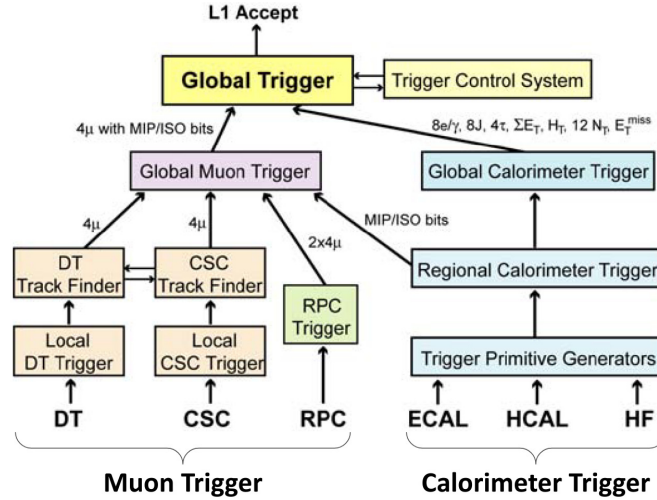


Figure 1.10: Block diagram of the architecture of the CMS L1T system [9].

The first-level decision is made within 4 μ s after which the L1T lowers the average event acceptance rate to no more than 100 kHz [20]. On the other hand, the HLT system further lowers this rate down to about 1 kHz, which corresponds to the rate at which data are written to mass storage [20].

1.3 The LHC High Luminosity Upgrade

This section is organised into three parts. Firstly, it defines the instantaneous and integrated luminosity. Secondly, via the LHC schedule it shows the plans for gradually increasing luminosity and when will the machine be maintained / upgraded over the years. Lastly, it uses an example physics process to demonstrate how the increased luminosity will generate more data.

Instantaneous and Integrated Luminosity Being a central performance measure of any accelerator or a particle collider experiment, the luminosity L (or instantaneous luminosity) is the number of beam particles going through an area per second (commonly expressed in $\text{cm}^{-2} \text{s}^{-1}$). For head-on collisions in a collider experiment, the luminosity [5] is given by (Equation 1.1)

$$L = \frac{N_1 N_2 n_b f_{rev}}{A} \quad (1.1)$$

where N_1 and N_2 are the number of particles in each of the bunches within the two beams going in opposite directions, n_b is the number of bunches, f_{rev} is the frequency of the revolutions and A is the effective beam overlap area at the IP (Figure 1.11).

The physics of the reactions is condensed down to the cross-section σ (commonly expressed in fb, where $1 \text{ b} = 10^{-24} \text{ cm}^2$), which gives the effective “size” of the target for a given reaction to occur. In a collider experiment with luminosity L , a process of cross section σ is occurring with a rate given by $L\sigma$. Therefore, the total number of events N expected on average can be found by solving the time integral (Equation 1.2).

$$N = \sigma \int L dt \quad (1.2)$$

where $\int L dt$ is another important measure being proportional to N , also known as the integrated luminosity (commonly expressed in fb^{-1}).

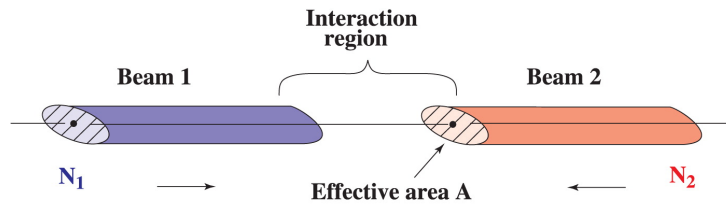


Figure 1.11: Illustration of the interaction region in support of the luminosity definition in a collider experiment [5].

LHC Schedule Over the years, CERN plans to interleave data-taking periods (Runs) with long shutdowns (LS) [21]. For the times when the laboratory is not collecting data, technical upgrades are planned such that the operating luminosity of LHC would gradually increase (Figure 1.12).



Figure 1.12: Schedule of interleaved periods of data collection and long shutdowns with a projection for the instantaneous and integrated luminosities and collision energies at the LHC [21].

The nominal instantaneous luminosity of the LHC was designed to be $10^{34} \text{ cm}^{-2} \text{ s}^{-1}$. However, after LS1 and by the end of Run-2, the instantaneous luminosity has been

increased to $2 \times 10^{34} \text{ cm}^{-2} \text{ s}^{-1}$. CERN is currently in the 3rd data-taking stage (Run-3) after which the 3rd period of upgrades (LS3) will take place. After LS3, the Phase-1 of CERN's operation will be completed and Phase-2 will start. The Phase-2 is also known as the High Luminosity era of LHC (HL-LHC). By the end of Phase-1, the instantaneous luminosity will not change and the integrated luminosity will have reached 450 fb^{-1} . By the end of Phase-2, the instantaneous luminosity will increase up to $7.5 \times 10^{34} \text{ cm}^{-2} \text{ s}^{-1}$ and the integrated luminosity will have reached as much as 4000 fb^{-1} [21]. A few additional general beam parameters during HL-LHC are listed in Table 1.2.

Table 1.2: Several notable HL-LHC design beam parameters for proton-proton collisions [22]. It should be mentioned that while the given peak luminosity value here corresponds to the technically achievable maximum, the nominal value is $7.5 \times 10^{34} \text{ cm}^{-2} \text{ s}^{-1}$ (as given in the text).

Parameter	Value
Momentum at collisions	7 TeV
Peak luminosity	$8.1 \times 10^{34} \text{ cm}^{-2} \text{ s}^{-1}$
Number of bunches	2760
Particles per bunch	2.2×10^{11}
Maximum PU	200

Higher Luminosity $\rightarrow$ More Data To bring together Equation 1.2 and the LHC schedule (Figure 1.12) one can estimate the total number of certain events that have occurred or will occur in a certain time period. For example, in 2019, the ATLAS Collaboration published a value [23] for the cross-section of the Higgs boson decaying into a pair of τ -leptons in proton-proton collisions at centre-of-mass energy of 13 TeV of $\sigma = 3.77_{-0.59}^{+0.60}(\text{stat})_{-0.74}^{+0.87}(\text{syst})\text{pb}$. Given that the LHC will deliver an integrated luminosity of 450 fb^{-1} by the end of Phase-1, it can be very coarsely estimated that these events will have happened approximately $3.77 \text{ pb} \times 450 \text{ fb}^{-1} \approx 1.7 \times 10^6$ times, since the LHC was turned on. Similarly, by the end of Phase-2, these events will have happened approximately $3.77 \text{ pb} \times 4000 \text{ fb}^{-1} \approx 1.5 \times 10^7$, which is an increase by an order of magnitude. Thereby, higher luminosity means a greater number of collisions, which in turn means more data to study. This will certainly allow the LHC experiments to understand the previous discoveries with greater confidence and can potentially open doors to new discoveries in particle physics.

1.4 Upgrading CMS for HL-LHC

The LHC requires numerous technical upgrades to operate effectively under the conditions brought by the high luminosity and this is true for both the RF instrumentation in the LHC ring, and for the detector experiments. For CMS in particular, this means being able to resolve a greater vertex density, under increased levels of radiation due to a greater PU of collisions per BX (i.e. up to 200). To achieve this, the approach is to upgrade the existing detector to have: higher radiation tolerance; better 3D granularity; a timing resolution of about 50 ps; more information to base the trigger decision on; computing

that can exploit all this information; and store these data to permanent storage whilst dealing with increased data rates [24]. Hereby, any enhancements in both the overall spatial granularity and timing information will aid in distinguishing different signatures in a single BX despite high levels of PU. As concluded in 2002 [25], for the innermost detector systems (i.e. the Tracker and the endcap calorimeters) the only viable solution is to completely rebuild them. Other sub-systems, however, do not require an overhaul of that scale, but will still be upgraded. This section provides a high-level overview of the major upgrades planned for each of the main parts of the CMS detector. Because the sub-systems will generally occupy the same volume within CMS as they used to, their corresponding figures to the ones shown in Section 1.2 will not be given.

The New Tracker The Tracker [26] will be replaced by a novel device, designed to be more granular and lighter, extend the acceptance region to $|\eta| < 4.0$ and provide trigger primitives to the L1T (i.e. tracking information for tracks above a configurable p_T threshold). The Inner Tracker will contain about 2×10^9 pixels, whereas the Outer Tracker will have about 42×10^6 strips and 170×10^6 macro-pixels. Both the Inner and the Outer Tracker will be based on silicon, as per the choice of the active material.

The New MIP-precision Timing Layer A new detector [27] will be added between the Tracker and the calorimeters allowing to measure the particle time of arrival with a resolution of 30 ps. This will be possible even for the Minimum Ionising Particles (MIPs), hence this novel device was named the MIP Timing Detector (MTD).

The Barrel Calorimeter Upgrades There will not be any radical changes in the barrel system [28], given that it will use the same layout of the crystals. Hence, they will contribute with the same granularity and the number of readout channels when it comes to event data. However, there are two significant points to note arising from the changes in the ECAL barrel readout electronics. Firstly, this will allow 25 times better granularity at the L1T level. Previously, the data from a set of crystals were sent together on the trigger path, whereas during HL-LHC this will be done on the level of individual crystals. Secondly, it will be possible to measure the timing information of the arriving particles (e.g. photons in $H \rightarrow \gamma\gamma$) with a resolution of 30 ps.

The New Endcap Calorimeter The ECAL and HCAL endcap systems will be replaced by a novel combined High Granularity sampling Calorimeter (HGCal) [29], which is the CMS upgrade in the focus of this thesis. The HGCal will have 47 active layers interleaved with absorber material. This device will bring high transverse and longitudinal spatial granularity in the endcap region whilst also providing precise timing measurements of electromagnetic and hadronic showers. It will use exclusively silicon sensors in its electromagnetic part, whilst the hadronic section will use a mixture of silicon and plastic scintillators. The HGCal will have around six million readout channels in total. It will also add a novel kind of a trigger primitive for the L1T - physical properties about 3D clusters. More details about HGCal in general can be found in Section 1.5.

The Muon System Upgrades The Muon systems [30] will upgrade the electronics of the existing DT, CSC and RPC detectors to sustain their performance during HL-LHC.

New resistive plate chambers, Gas Electron Multipliers (GEM) and a cathode strip chamber will also be added. The upgraded system will generally be able to measure an increased number of hits per a muon track and its acceptance region will be extended to $|\eta| < 2.8$.

The New Trigger System During HL-LHC, the CMS L1T [20] system will have a higher average event acceptance rate of 750 kHz and a latency of 12.5 μ s. The L1T will receive trigger primitives from various sub-systems with a latency of 5 μ s with respect to the corresponding BX they originate from. For the first time, tracking and highly granular calorimeter information will be used as inputs to the L1T decision making algorithms. The architecture of the L1T will increase in complexity, requiring to implement a new system using the state-of-the-art Field Programmable Gate Array (FPGA) devices. The average HLT rate will also be increased to 7.5 kHz and the projected CMS event size will be between 6 MB to 9 MB [31].

1.5 The High Granularity Calorimeter

Regarding the existing endcap calorimeter systems, there are two driving factors as to why the HL-LHC conditions require their replacement with a novel solution. Similarly to the general CMS upgrades, these have to do with radiation hardness and increased PU. Firstly, the accumulated effects of radiation in regard to the projected levels of integrated luminosity are expected to significantly degrade the performance of the existing endcap calorimeters (Figure 1.13a). Hence, a more radiation-tolerant endcap calorimeter design is necessary. Secondly, the original design of CMS expected an average PU of 20 collisions per BX [9]. However, during Phase-2, the PU is expected to be as high as 200 [21]. To be able to distinguish and reconstruct the collision products coming from all these proton-proton collisions within a BX (Figure 1.13b), a design with high spatial (both longitudinal and transverse) granularity and time resolution is needed.

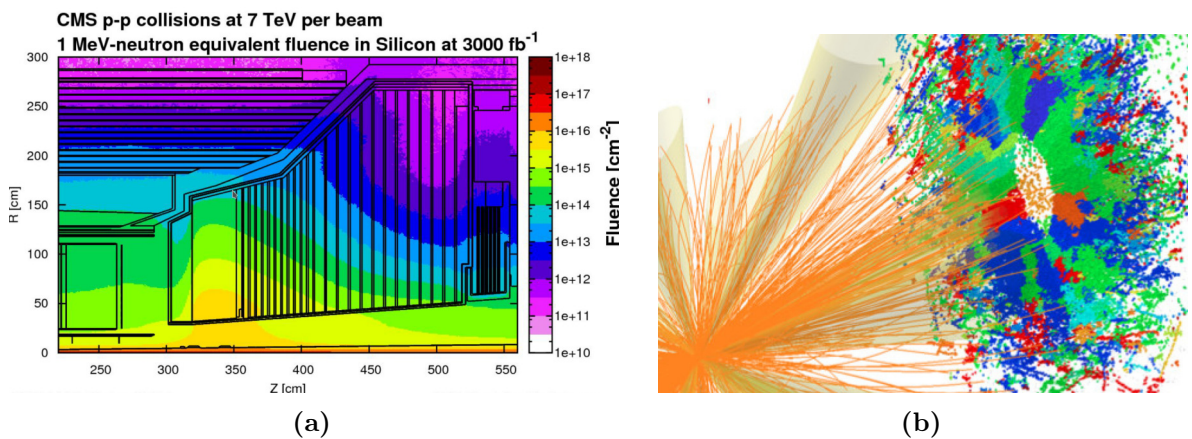


Figure 1.13: (a) Expected radiation levels absorbed in the existing calorimetry systems if they were to be used in the HL-LHC conditions; shown here, at an integrated luminosity of 3000 fb^{-1} . The 1 MeV-neutron equivalent fluence represents the number of 1 MeV neutrons per unit area that would cause the same amount of damage as the actual mixed radiation sources. Provided by the CMS FLUKA simulator [29]. (b) An illustration of a typical signature (simulated) of energy deposits inside HGCal and general vertex density in a single HL-LHC BX [32].

To tackle these challenges, the CMS collaboration has proposed replacing the two existing ECAL and HCAL endcaps with a new combined high-granularity sampling calorimeter, the HGCAL [29] (Figure 1.14). This section introduces the HGCAL by providing a high-level overview of this future CMS sub-system summarising some of its general parameters, contributions to the L1T, and its main physics goals.

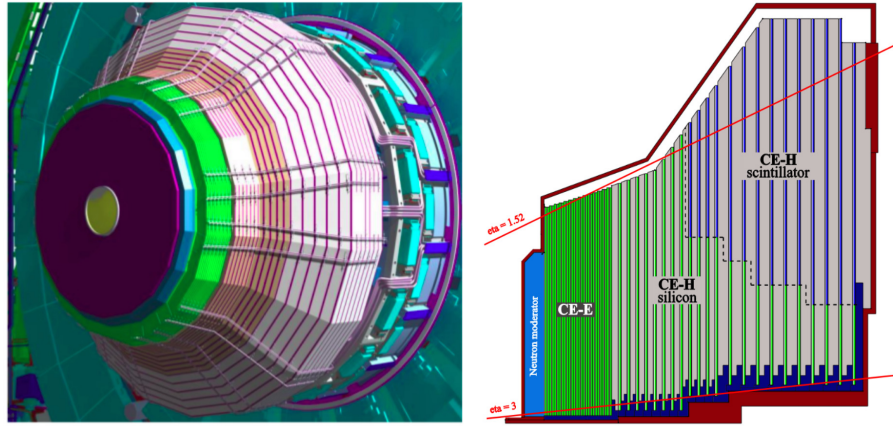


Figure 1.14: HGCAL endcap calorimeter layout in 3D (Left) [29] and its longitudinal cross section (Right) [20].

General HGCAL parameters The device will have a total of 47 sensitive layers: 26 electromagnetic (CE-E) and 21 hadronic (CE-H). The absorber layers will be made of tungsten and copper in the CE-E part, and of stainless steel in the CE-H part. The entire CE-E and part of the CE-H will use silicon sensors with cell areas of 0.5 cm^2 or 1 cm^2 , whereas the rest will use plastic scintillators, with cell areas of 4 cm^2 to 30 cm^2 (Figure 1.15). This will result in a total of 6×10^6 channels coming from the silicon cells and 240×10^3 channels from the scintillator tiles. In the choice of the active material used

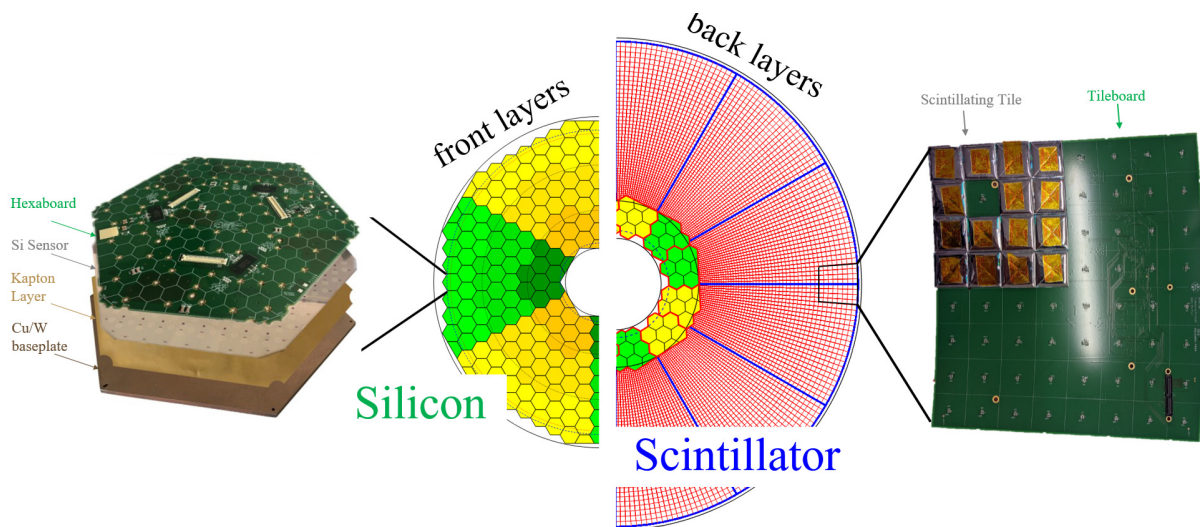


Figure 1.15: Assembly of silicon-based HGCAL modules (Left) [33], arrangement of silicon and scintillator parts of HGCAL in its front and back layers (Middle) [29] and assembly of scintillating tiles in the scintillator-based HGCAL modules (Right) [33].

for the sensors, silicon is selected for the areas where levels of radiation will be higher, compared to where the plastic scintillators will be used. The whole device will be cooled down to -30°C to reduce the dark currents arising due to radiation effects. In addition, the “on-detector” electronics will be made of radiation-hard Application Specific Integrated Circuits (ASICs). The HGCal will weigh about 215 t and consume 125 kW of electrical power per endcap. Throughout its volume (unlike the existing endcap systems), the HGCal will be able to provide precise information about the spatial location (lateral and longitudinal), timing, and energy deposition of the collision products (i.e. electromagnetic and hadronic showers), with resolutions as good as 1 cm, 10 ps, and 0.3 MIP, respectively. It will also be able to provide measurements of energy deposits over large dynamic ranges. Therefore, the HGCal itself can be compared to a 5D digital camera with high granularity. Both the highly granular spatial and timing information will help distinguishing products from different collisions within a given BX.

HGCal inputs to the L1T Each of the HGCal endcaps will contribute to the L1T by providing trigger [29] primitives for every BX in the form of: tower energies from each of the bins in the 20×72 (η, ϕ) grid; and a record of physical properties for each one out of about a hundred 3D clusters on average. The towers, already provided by the existing endcaps with similar granularity, will each have an (η, ϕ) coverage of 0.087×0.087 . The cluster properties, however, are a novel kind of input for the L1T. They will provide the L1T insights into the endcap showers by utilising both the transverse and longitudinal granularity of HGCal, a capability unavailable with the previous endcap architecture. These quantities will mainly describe 3D clusters in terms of their: energy, position and shape.

HGCal Physics Goals There are several aspects regarding the main physics goals of HGCal. Using its fine spatial (transverse and longitudinal) and timing granularity will allow to disentangle signatures from the collision products of different proton-proton interactions despite the high levels of PU in each BX. This will be the case both when analysing the coarse-detailed trigger data as well as when doing so with fine-detailed event data (although it is only the latter that has the timing information available). Therefore, this will be crucial in the trigger path to efficiently select relevant events, minimise background noise, and optimise event reconstruction. Specifically, the fine granularity will allow to clearly measure (and distinguish) the electromagnetic and hadronic showers, critical for analysing complex event topologies such as those involving photon showers, high p_T jets, Vector Boson Fusion (VBF) jets, and missing transverse energy – pivotal for Higgs boson studies and new physics searches, including super-symmetry. The point about the Higgs boson is particularly important because one of the major goals of CMS during HL-LHC is to gain a deeper knowledge about it. Namely, detailed measurements of electromagnetic showers are significant because one of the main decay modes of the Higgs which has a clean signature in the electromagnetic calorimeter is via the $H \rightarrow \gamma\gamma$ channel. Similarly, detailed measurements of hadronic showers are important because they can stem from jets originating from VBF which is one of the primary production mechanisms for the Higgs boson. Additionally, the high granularity of particle signatures in HGCal will connect with the ones from the upgraded tracker (both for the trigger and event data), enhancing the general CMS track reconstruction capabilities. This improvement will also aid in the identification of MIPs, thereby enhancing muon triggering and reconstruction.

1.6 Scope of the Thesis

This thesis is organised into seven chapters, with some recurring themes in it being the design, implementation, testing, and evaluation of the physics performance of various parts of HGCal electronics (or CMS electronics directly related to HGCal). The second chapter (Chapter 2) provides an in depth overview of the entire HGCal electronics system. It will therefore be a pivotal point that connects all other chapters to one another and serves as a foundation to which each one of them will often refer to. The third chapter (Chapter 3) summarises how a small-scale prototype of the HGCal electronics system (that has the same variety of devices and interfaces as the final system will) was implemented and all of its major interfaces enabled. The fourth chapter (Chapter 4) is about gaining a deeper understanding of the previously implemented system prototype. This involves using it in a beam test to acquire real physics data, but also setting upper limits on the (trigger and event) data acquisition bit error rates and doing various low-level consistency checks between the different devices in the data acquisition chain. This chapter will also report on a discovered signal integrity issue (and how it was mitigated) in the optical interface between the on-detector and the off-detector electronics, which is relevant for all experiments using the same technology (i.e. not just HGCal and CMS). The fifth chapter (Chapter 5) is about an FPGA algorithm to reside in the future Phase-2 Timing and Control Distribution system (TCDS2) of CMS. Its functionality is to prevent CMS from requesting some Level-1 triggered events from all of its sub-detectors, because in these rare cases the HGCal readout chips would not be able to send out any event data. This would consequently complicate event reconstruction both in HLT and in offline analysis and that is what this algorithm is trying to protect from. The sixth chapter (Chapter 6) considers an FPGA algorithm which calculates various physical properties about the HGCal 3D clusters formed in the off-detector electronics on the trigger path. Lastly, a conclusion chapter summarises the most important results of this thesis and their significance, reflects on the biggest challenges encountered, and provides an outlook into potential future research activities. Each main chapter (Chapter 3, 4, 5 and 6) is closed with a brief section summarising its most important results.

At the request of various members of the HGCal Collaboration, this thesis was written with the goal to serve as a reference for newcomers, and also as a user's manual or an application note for the more experienced in the project. Hence, some parts of it (especially in Chapters 2, 3 and 4) may seem repetitive (e.g. having similar paragraphs in the presentation of different HGCal ASICs) or too detailed (e.g. exposing the exact ASIC register names or their bit-level content). These choices were made to allow different logical units of text to be self-consistent (e.g. not requiring one to read the whole thesis in order to develop a quick understanding of a device or a procedure) and to allow reproducible results with little effort.

Chapter 2

The HGCal Electronics System

Being a pivotal point to which most of the chapters in this thesis will often refer to, this chapter explores and describes the HGCal Electronics System. It starts with a general overview (Section 2.1), describing the on-detector / front-end (FE) and the off-detector / back-end (BE) parts and their most relevant interfaces. There, a particular focus is put on the Trigger Primitive Generator (TPG) and the Data Acquisition (DAQ) sub-systems that work in parallel to one another and in both the FE and the BE parts. Additionally, the three major kinds of data exchange in the FE $\leftrightarrow$ BE interface are emphasised: timing distribution and fast control, slow control, and data (TPG and DAQ) acquisition. This is followed by a presentation of the FE (Section 2.2), the DAQ BE (Section 2.3), and the TPG BE (Section 2.4) of the HGCal electronics system on a deeper level ¹. Here, the TPG and the DAQ data are traced from their origins (e.g. the FE sensors) through all the key FE algorithms implemented in ASICs and the 100 m distant BE algorithms implemented in FPGAs, respectively. Lastly, having covered all the different parts of the system and how they interface with one another, a short note on their individual contributions to the overall system latency is given in Section 2.5.

2.1 General System Overview

The final HGCal electronics system is highly specialised and complex, involving multiple layers of data transfer and control and having a complex variety of interfaces with the CMS global systems (Figure 2.1) [34]. There are two distinct parts to the HGCal electronics system: the FE and the BE electronics. Both of them house two co-existing sub-systems, the TPG and the DAQ system. Considering both endcaps, the data from the six million FE readout channels is pushed across 20 000 fibre optical (F/O) links to the 100 m distant BE systems. This will result in a data rate of about 110 Tbit s^{-1} of usable TPG and DAQ data between the detector FE and the detector BE (more details behind this rate can be found in Section 2.3 and Section 2.4). While the FE ASICs need to be radiation-hard, the BE FPGAs do not because they are far enough from the hostile radiation environment.

This section is organised into several parts. To start with, it introduces the basic concepts behind: timing distribution and fast control; slow control; and the acquisition of TPG and

¹This chapter highlights the specifics of the silicon part of the detector, but similar concepts apply for the scintillator part, too.

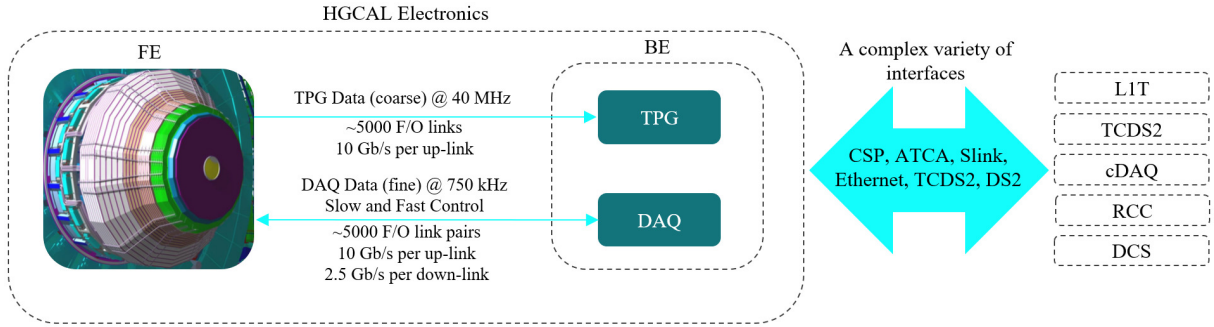


Figure 2.1: The HGCAL electronics system and the complex variety of its interfaces with the CMS global BE systems illustrated in a simplified block diagram [1]. The F/O link counts are given per endcap (i.e. there are two copies of the HGCAL Electronics layout in the final system). The acronyms used are explained in the text.

DAQ data. This is followed by exposing the variety of the different layers of data transfer on a logical level (i.e. abstracting any degree of parallelism that exists in these layers). Lastly a catalogue of the different hardware parts that exist in the system is given.

2.1.1 Timing Distribution and Fast Control

In order to ensure that the FE and the BE electronics are synchronised and operate based on a common understanding of time, HGCAL is using a “time-relative” system, allowing for data collected from different parts of the detector to accurately correspond to the same moment in time. From the TCDS2, a precise 40 MHz LHC clock reference is distributed to the HGCAL BE and through the DAQ BE to the FE ASICs. The 40 MHz LHC clock reference is then recovered in the FE ASICs which allows both the FE and the BE to synchronise with the crossings of particle bunches at the IP. From the TCDS2, a variety of fast control signals (commonly referred to as fast commands) is also brought to the DAQ BE from where it is further distributed to the FE ASICs. In addition to the widely known L1A, there are also other fast commands used for various levels of system synchronisation, initiating reset procedures, clearing memory buffers, etc. There is an HGCAL dictionary of 8-bit wide fast commands to which the different parts of the system need to react as per definition [35]. A fast command is distributed every 25 ns. A special example is the IDLE fast command that is present in every BX when no other specific fast command is being distributed and is used by the FE ASICs to recover the phase of the 40 MHz LHC clock. What is common to most fast commands is that they have to be distributed while respecting precise timing (e.g. BX-level precision) constraints. Also, their typical distribution latency from the TCDS2 to the detector FE is within $O(1\ \mu\text{s})$. Exceptionally for L1As, there is an $O(10\ \mu\text{s})$ latency constraint for their generation and distribution.

2.1.2 Slow Control

The Run Control and Configuration (RCC) system performs slow control (e.g. configuration, monitoring, resets, etc.) of the HGCAL BE and the HGCAL FE, where the latter happens through the DAQ BE. Although the details are not entirely clear at this time, there is a possibility that the Detector Control System (DCS) may also use the DAQ BE as an entry point for monitoring certain FE parameters (e.g. temperature). The configuration

part usually happens during the system initialisation procedure before the actual data acquisition starts and needs to be completed in $O(1 \text{ min})$ whereas the monitoring happens from time to time and usually not more often than $O(1 \text{ s})$. Unlike fast control, it is not critical to precisely time the execution of slow control.

2.1.3 Data Acquisition

Unlike some other sub-detectors (e.g. ECAL Barrel), the HGCAL cannot afford to send the “picture” of every BX in full available detail from the detector FE to the detector BE. Instead, for every BX, the HGCAL FE produces two “pictures” of the collision products traversing through its volume: the coarse-detailed (TPG) data and the fine-detailed (DAQ) data for the given BX (more details in Section 2.2).

Trigger Data Acquisition

For every BX occurring at 40 MHz, the TPG BE acquires trigger data from the detector FE. The TPG BE is responsible for organising the data coming from different parts of the detector FE, unpacking, calibrating and processing it to form the final trigger primitives for each BX. These trigger primitives are then sent to the L1T system where they are combined with trigger primitives coming from other CMS sub-systems (e.g. Barrel calorimeters, Tracker, etc.) and taken into account when a decision is made about which BXs are interesting and should be examined in fine detail and which can be discarded. The decision is made and distributed to the FE in the form of a fast command, commonly referred to as the Level-1 Accept (L1A) signal.

DAQ Data Acquisition

Due to a hardware limitation (i.e. buffer depth), the FE readout ASICs can hold on to DAQ data for any given BX for no longer than 512 BX so it is critical for the L1A fast command to reach them sooner than $512 \times 25 \text{ ns} = 12.8 \mu\text{s}$. The DAQ BE acquires the full-granularity data for the selected BX from the detector FE. After it organises the data coming from different parts of the detector, the DAQ BE sends fully built events to the Central DAQ (cDAQ) system for further selection and analysis at the HLT. The projected average L1A rates during the HL-LHC operation of CMS are as high as 750 kHz, whereas the selection in the HLT is expected to bring this number further down to 7.5 kHz [20]. The F/O links between the FE and the BE (up-link direction) for both the TPG and the DAQ systems run at 10 Gbit s^{-1} . The F/O links between the DAQ BE and the FE (down-link direction) run at 2.5 Gbit s^{-1} . With 5000 F/O links per endcap between the detector FE and BE for each of the TPG and the DAQ systems, the projected usable rates of DAQ and TPG data coming from the HGCAL FE are 9 Tbit s^{-1} and 45 Tbit s^{-1} per endcap, respectively [34].

2.1.4 Different layers of data transfer in HGCAL

To further (but still on a high level) expose the different layers of data transfer that exist in HGCAL (more details can be found in Chapter 3), the parallelism that will exist in the final system can be temporarily abstracted to a great extent. In this context, only the main devices that exist on the TPG and the DAQ data paths are logically considered (Figure 2.2).

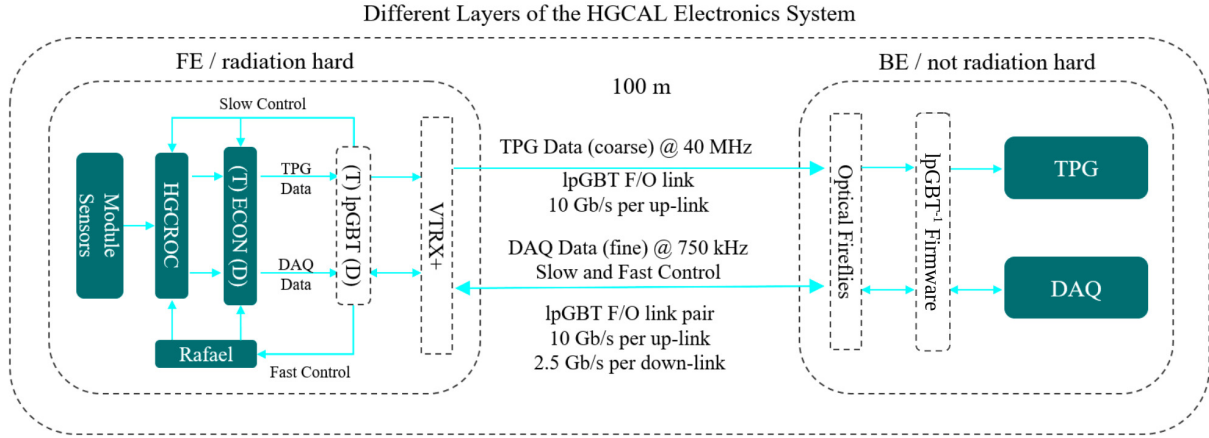


Figure 2.2: A logical block diagram of the key layers of data transfer in the HGCAL Electronics System. Any degree of parallelism that will exist in the final system is abstracted. The acronyms used are explained in the text.

The Very FE Layers (HGCAL Specific)

When a particle traverses through the HGCAL volume and interacts with a FE module sensor, an electrical signal is formed. This analogue signal is then amplified, shaped, digitised and processed in the HGCAL Readout Chip (HGCROC) [36], forming the TPG and the DAQ variants of the data for a given BX. After formatting, the TPG data are streamed out of the HGCROC on every BX, whereas the DAQ data are temporarily stored and streamed out only if the BX is selected by the L1T. The TPG and the DAQ data are transmitted over a number of serial differential Electrical Links (e-Link) towards the TPG and the DAQ e-Link Concentrator ASICs (ECON-T and ECON-D, respectively) [37], [38]. There, coming from several HGCROCs, they are aggregated, concentrated (with a general aim to occupy a smaller bandwidth than originally) and formatted. For both ECON ASICs, the level of data concentration is highly configurable by the user and is mainly driven by what part of the detector a given ECON is designated to service (e.g. different parts of the detector have different output bandwidth limitations due to the available number of ECON output e-Links). Each output e-Link from either an HGCROC or an ECON are designed to stream out one 32-bit word every 25 ns (i.e. at 40 MHz), hence each output e-Link data rate is $32 \text{ bit} \times 40 \text{ MHz} = 1.28 \text{ Gbit s}^{-1}$. The Rafael ASIC fans out the fast control data stream (has fast commands modulated within it) and the fast control clock to the HGCROC and ECON ASICs. Again, both signals are coming to the FE targets through a serial e-Link interface. The fast control clock stream is a 320 MHz periodic signal. From it, the FE ASICs can generate various clock frequencies that they need to run their internal algorithms (e.g. 40 MHz, 80 MHz, 160 MHz, 320 MHz, 640 MHz, 1.28 GHz). Each fast command is an 8-bit word which needs to be received by the HGCROC and ECON ASICs every 25 ns, hence the fast command e-Link data rate is $8 \text{ bit} \times 40 \text{ MHz} = 320 \text{ Mbit s}^{-1}$. The HGCROC and ECONs have an Inter-Integrated Circuit (I²C) [39] interface through which their register space can be accessed (e.g. read from or written to) from outside. The HGCROC, ECON-T and ECON-D are all highly configurable devices, each with thousands of I²C registers. Also they all have triplicated I²C registers in order to be more robust against Single Event Upsets (SEU).

The Layers in the FE $\leftrightarrow$ BE Interface (not HGCAL Specific)

Coming from several ECONs and over a smaller (than input) number of 1.28 Gbit s^{-1} output e-Links (due to data concentration), the TPG and DAQ data are brought to the Low Power Gigabit Transceiver (lpGBT) ASICs [40]. There, they are aggregated together, de-serialised, zero-balanced, formatted, protected and serialised to be sent through the High-Speed Line Driver serial interface towards the Versatile Link+ (VTRx+) [41]. The VTRx+ is where the conversion of the electrical into optical signals takes place. After the data complete their 100 m long journey to the HGCAL BE, they reach the Optical Fireflies [42] that convert the optical signal back into the electrical domain. The electrical signals then go into the BE FPGA fabric, where they are first processed by the lpGBT-specific algorithms that perform inverse operations to the ones done in the lpGBTs so that the data at the output of these blocks is the same as it was at the lpGBT e-Link inputs. The TPG BE and the DAQ BE algorithms then receive the TPG and DAQ data in the ECON-T and ECON-D formats (as if the real ECON outputs were in the BE itself) and start processing them. In other words, the lpGBT and the VTRx+, with their BE inverses, are transparent to the HGCAL algorithms and the same concepts apply in the down-link direction. The data rate in the up-link serial interface is $10.24 \text{ Gbit s}^{-1}$ (per up-link fibre) and that bandwidth is mostly consumed by the transmission of DAQ and TPG data, but a small fraction of it is used for the “reading” part of slow control. The data rate in the down-link serial interface is 2.56 Gbit s^{-1} (per down-link fibre) and that bandwidth is mostly consumed by the transmission of fast commands, but a small fraction of it is used for the “writing” part of slow control. The DAQ lpGBT transmits the fast control data and fast control clock streams to the Rafael ASIC, over 320 Mbit s^{-1} serial e-Links. The slow-control transactions with the FE ASIC slaves are orchestrated using lpGBT built-in I²C masters.

The above description exposed the various layers of data transfer in both the up-link and the down-link directions of the data flow on a logical level. It motivated a need to explore the FE modules, the HGCROC, the ECON-T, the ECON-D and the lpGBT in more detail. Just before doing that and to complement the given logical with a physical perspective, a catalogue of different hardware parts and how they fit into the big picture is given.

2.1.5 A Catalogue of Different Hardware Parts

Stepping away from the logical level the system was previously discussed in terms of, one can consider an exploded view of all the main hardware parts from the detector FE to the detector BE (Figure 2.3). The Hexaboard Printed Circuit Board (PCB) [43] is home to the HGCROC and the Rafael ASICs. Regarding Hexaboard cell density, there are two choices: Low Density (LD) and High Density (HD) Hexaboards, for LD and HD silicon modules (more details can be found in Section 2.2.1), respectively. The Concentrator Mezzanine PCB is home to the ECON-T and the ECON-D ASICs and connects to the Hexaboard. For practical purposes, one can consider the Concentrator Mezzanine as an extension of the Hexaboard (e.g. if one wants to interface with the Concentrator Mezzanine, one can only do so through the interface with the Hexaboard). A number (depending on position) of Hexaboards connect to the Wagon PCB, which does not change the data, but only routes them between the Hexaboards and the closest Engine PCB. The Engine PCB is home to the lpGBT ASICs and the VTRx+.

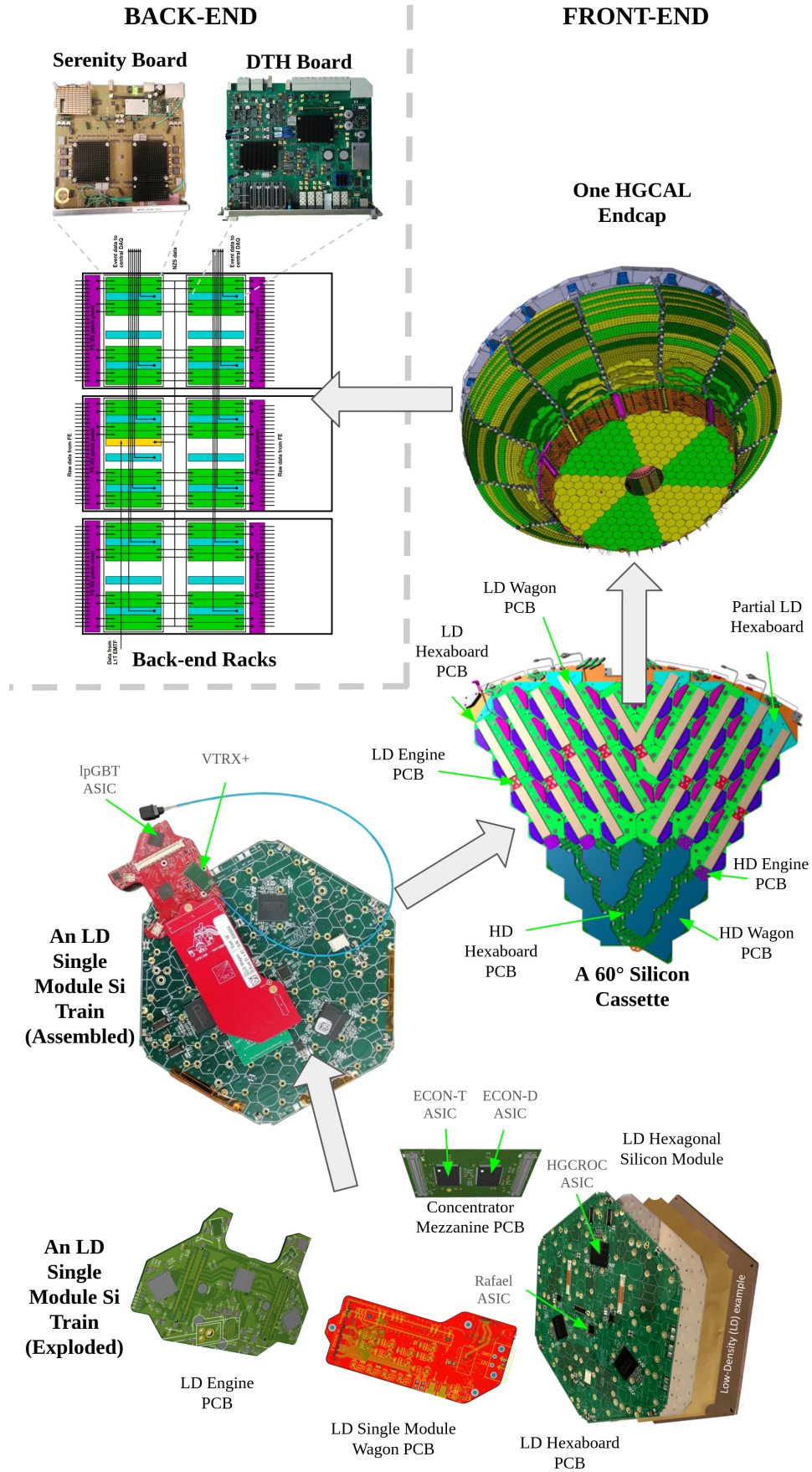


Figure 2.3: A catalogue of different hardware parts existing in the very complex HGCal Electronics System, from the FE silicon modules, to the BE boards. Acronyms used are explained in the text.

A FE structure that brings together a number (depending on position) of Hexaboards, Wagon PCBs (one or two), and one Engine PCB is commonly referred to as a FE Train. Two varieties of Trains exist: the LD Trains and the HD Trains, depending on whether the given Train services LD or HD silicon modules. An HD Train comes with a special HD Wagon PCB (always one) and an HD engine PCB. Putting many Trains together, one can assemble a cassette of an HGAL Layer.

Implementation, operation and performance optimisation of a LD single module silicon Train are some of the central topics of this thesis (more details can be found in Chapter 3 and Chapter 4). Around the cassette outer edges, specially shaped Partial Hexaboards are put in place. Bringing together cassettes allows assembling HGAL layers, eventually building an HGAL endcap. Both endcaps will together have around 30 000 FE modules, 1000 m² worth of active material and consume 250 kW of electrical power [33]. The HGAL BE is located in a subset of CMS BE racks (shared by other sub-detector BE algorithms) which are home to the BE crates. The BE crates house the custom Advanced Telecommunications Computing Architecture (ATCA) Serenity boards [44]. The Serenity boards accommodate the Xilinx VU13P FPGA devices where the BE DAQ and TPG algorithms are implemented. These data are then acquired from the Serenity boards by the DAQ and TCDS2 Hub (DTH) boards [45]. Additionally, through the DTH boards, the timing and control signals are distributed to the DAQ BE which then further distributes these to the rest of the HGAL electronics system.

2.2 The HGAL Electronics FE

The HGAL FE has a large degree of complexity due to its variety (e.g. there are no two of the same layers throughout the longitudinal depth of an endcap; each layer has Trains of different shapes, numbers of modules, module densities, etc.). Following the data-flow from its creation onward, this section abstracts most of that variety. The focus is on the HGAL hexagonal silicon modules (that is for full hexagons and not partial hexagons, although similar principles apply for the latter) and further exposing the main functionalities of the HGCROC, the ECON-T, the ECON-D as well as the lpGBT ASICs.

2.2.1 The Hexagonal Silicon Modules

The following description of the hexagonal silicon modules is based on material in [29]. Each with their longest distance between any two corners being 20 cm, the hexagonal silicon modules are the central building blocks of the silicon parts of the HGAL FE (Figure 2.4). The silicon cell density and the module thickness vary depending on the position of a given module in the detector (Figure 2.4a). Regarding cell density, there are the LD (Figure 2.4b) and HD (Figure 2.4c) variations. An LD module has 192 silicon cells each of 1.26 cm², whereas an HD module has 432 silicon cells, each of 0.56 cm². These cells are commonly referred to as channels. Higher density provides finer spatial segmentation when trying to distinguish nearby particles in the volume of the detector where there is a higher density of traversing particles. Regarding the module thickness, there are three choices: 120 μ m, 200 μ m and 300 μ m used in areas of respectively decreasing radiation exposure. As the detector ages, the negative radiation effects such as the ability to collect charge or the leakage currents will worsen. Ways to combat these negative effects are to

increase the reverse bias on the p-n junction and to cool the detector down (e.g. HGCal FE will be kept at -30°C), respectively. At the start of detector life, the bias voltage for the 120 μm , 200 μm and 300 μm sensor thicknesses will be approximately 20 V, 100 V and 300 V, but can never exceed 800 V for any given sensor thickness by the end of life. Over the years of operation, the reverse-bias voltage will be gradually increased, and in greater steps for modules with smaller thickness while never exceeding 800 V for any given thickness. This logic is the main driver of the choices for the silicon module thicknesses. The general choices of the size ballpark in which all silicon sensor cells are in were made to optimise physics performance in terms of the Signal-to-Noise-Ratio (SNR) and the spatial resolution requirements.

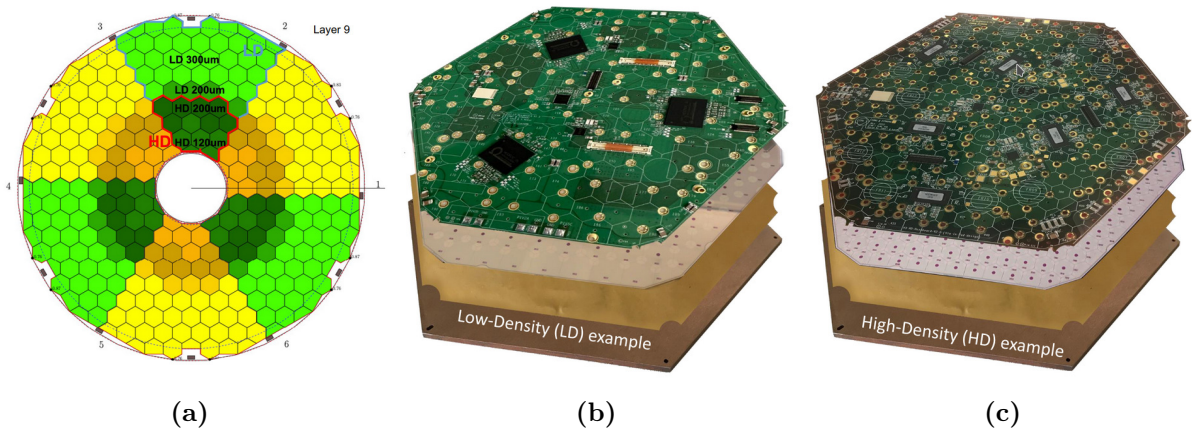


Figure 2.4: (a) An example layout of a silicon-only HGCal layer showing different module density and thickness choices [29]. Assembly of an (b) LD and (c) HD hexagonal module [46]. In the layers of the vertical stack-up in (b) and (c) from the bottom up are the copper base-plate, resistive kapton foil, Si module, Hexaboard PCB, respectively. Acronyms used are explained in the text.

In the vertical stack-up of each hexagonal module are the copper base-plate (used for cooling and mechanical support), a layer of kapton foil (to provide galvanic isolation between the silicon and the baseplate), the silicon layer and lastly the Hexaboard PCB [43].

2.2.2 The HGCROC ASIC

The following description of the HGCROC ASIC is based on material in [36]. The HGCROC is the first ASIC in the HGCal FE readout chain (Figure 2.5). It first amplifies and shapes the analogue signals coming from the module sensors. These analogue signals are then digitised and processed, forming the TPG and the DAQ variants of the data for a given BX. The HGCROC can receive a total of 78 input channels, out of which 72 are standard (designated to be connected to the FE module silicon sensor cells). To mitigate the effect of interference that affects all channels uniformly, there are 4 special channels designated for Common Mode (CM) noise subtraction. Lastly, there are 2 channels designated for MIP calibration. The LD Hexaboard will have three HGCROCs mounted (each with 64 connected standard channels, i.e. $64 \times 3 = 192$), whereas the HD Hexaboard will have six HGCROCs (each with all 72 connected standard channels, i.e. $72 \times 6 = 432$). The HGCROC outputs TPG data on every BX across two or four (HD or LD case) differential 1.28 Gbit s $^{-1}$ e-Links and DAQ data only for BXs selected by

the L1T across two differential 1.28 Gbit s^{-1} e-Links. The HGCROC also has designated internal blocks for clock recovery and fast control, as well as an I²C interface for slow control of the device.

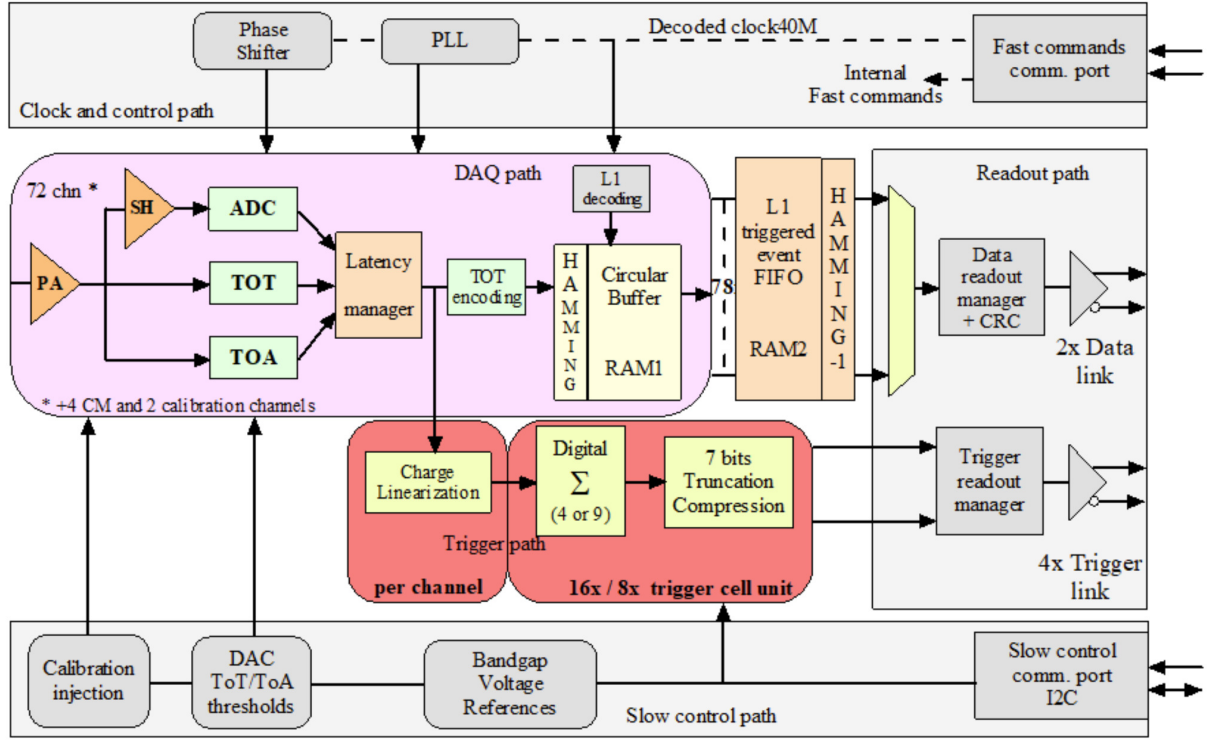


Figure 2.5: A block diagram of the key DAQ and TPG data manipulation steps (Colour) and other blocks related to timing, control and readout within the HGCROC ASIC (Gray) [36]. The relevant acronyms used are explained in the text.

Timing and Fast Control

Once presented with the 320 MHz fast control reference clock and the 320 Gbit s^{-1} fast control data stream, both coming from the Hexaboard Rafael ASIC, the internal clock recovery logic, based on a Phase Locked Loop (PLL), recovers the 40 MHz LHC clock frequency and phase. A variety of clock frequencies is made available for the internal algorithms (e.g. 40 MHz , 160 MHz , 320 MHz , 640 MHz , 1.28 GHz). The Phase Shifter logic allows for fine-tuning of the recovered clock phase (e.g. to ensure analogue signals are sampled at optimal times). The fast control interface de-serialises the 8-bit fast commands from the incoming 320 Mbit s^{-1} fast control data stream, so that the HGCROC can react to them as per the HGCAL fast control specification [35] (more details about using fast commands with the HGCROC can be found through working examples in Chapter 3).

Slow Control

The HGCROC has an I²C [39] interface for slow control purposes. It is a highly configurable device with about 8000 I²C 8-bit registers. The internal I²C registers are triplicated to be robust against SEUs. There are also a couple of General-Purpose Input / Output (GPIO) pins through which the HGCROC can be taken through a couple of levels of reset.

The ADC, ToT and ToA measurements

The HGCROC receives the analogue signal coming from the FE module sensor cell and takes it through the analogue Pre-Amplifier (PA) which converts the deposited charge into

a voltage signal (Figure 2.6a).

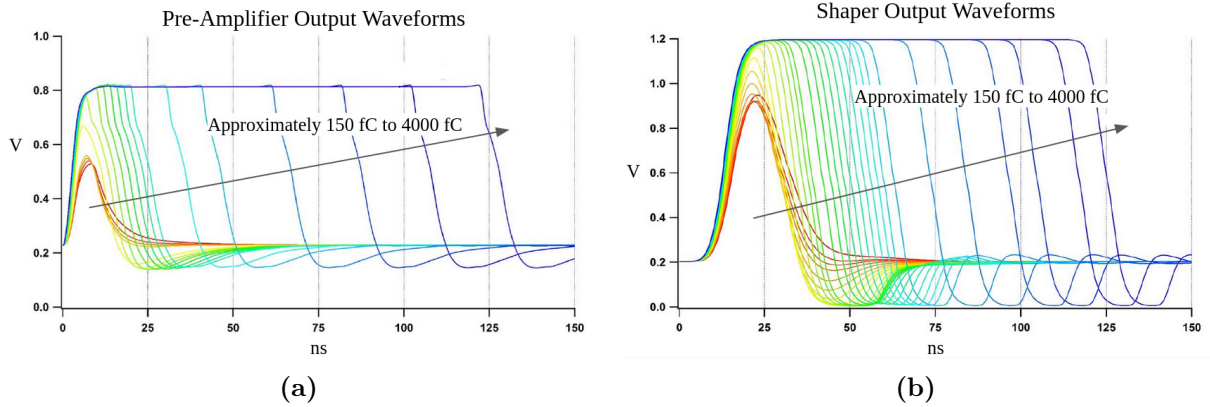


Figure 2.6: Output waveforms of the HGCROC (a) Pre-Amplifier in a range of charge deposits and (b) Shaper in a range of charge deposits [36].

There are two methods for measuring the amplitude of this signal, depending on the size the signal. If the signal is small enough it will be sampled in the HGCROC 10-bit Analogue to Digital Converter (ADC). This happens after the signal is taken through the analogue Shaper (SH) step (Figure 2.6b), mainly dedicated to shape the original analogue signal into covering the full dynamic range of the ADC (≈ 1 V) and ensuring optimal peaking time between 20 ns to 25 ns. The Phase Shifter block can offset the recovered 40 MHz clock reference in steps of $25 \text{ ns}/16 = 1.5625 \text{ ns}$ to ensure that the signal is sampled at the peak. As shown in the waveforms (Figure 2.6), large signals can cause the Pre-Amplifier (and in turn the Shaper) output to saturate. If the signal is too big (i.e. outside of the operating range of the ADC), the HGCROC can measure the time during which the Shaper output was above a threshold, which is also known as the Time Over Threshold (ToT) measurement. In addition to the measurement of the deposited energy, the HGCROC also performs a timing measurement of the incoming particles, also known as the Time of Arrival (ToA). An example with energy deposits in close BXs illustrates when ToA and ToT measurements start (Figure 2.7). The user can configure the ToA and ToT threshold

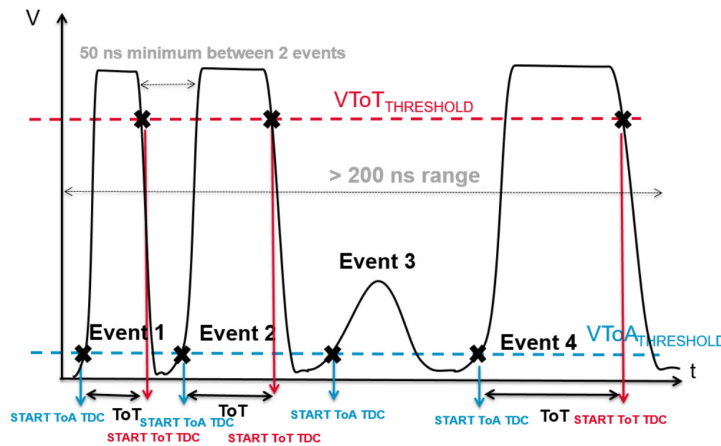


Figure 2.7: An illustration of several close BXs with energy deposits and their ToA and ToT measurements [36]. The acronyms used are defined in the text.

values via slow control. When the signal goes above the programmed ToA threshold, the ToA Time Digital Converter (TDC) completes the measurement it had previously started

at the beginning of the given BX in which the excitation occurred. That is also the time when the ToT TDC starts. However, that measurement is only complete once the Shaper output falls below the programmed ToT threshold.

Regarding HGCROCs ability to measure deposited charge, the dynamic range is between $O(1 \text{ fC})$ and $O(1000 \text{ fC})$, with the MIP being represented by $O(10)$ ADC counts. Regarding the ToA, the HGCROC provides a 10-bit measurement across the 25 ns time window, resulting in 25 ps time granularity. With a minimum time between hits of 25 ns, the HGCROC provides a 12-bit (later compressed to 10 bits) measurement of the ToT over a range between 2 ns to 200 ns achieving a time resolution smaller than 50 ps. Hence, there can be cases when there are multiple energy deposits in a channel in nearby BXs and it would not be possible to obtain their individual ToA and ToT measurements. In other words, they would have one combined signature that would appear as a single hit that has one ToA and one relatively large ToT value. For similar reasons, it is not possible to distinguish multiple sensor hits that happen in a single channel within one BXs (e.g. coming from PU). The HGCROC normally sends out the ADC value from the previous BX, together with the ADC (or TOT) and ToA value for the selected BX. When the HGCROC completes the ADC, ToT and ToA measurements, the results are taken through a Latency manager which makes sure that all three measurements for a given BX are available in the same clock cycle, even though they individually complete at different clock cycles from one another. It is at the output of the Latency manager where the data first splits into the TPG and the DAQ paths. From here onward, already within the HGCROC ASIC, the TPG and the DAQ sub-systems start their parallel co-existence in the HGCal electronics system.

Functionality Overview - TPG Path

After the Latency manager block (Figure 2.5), the information from the ADC and ToT for every BX is streamed towards a sequence of three main steps in the HGCROC TPG path. The first step is the charge linearisation step. Here, both the ADC and the ToT value go through a pedestal subtraction stage and a selection versus threshold stage (each having their own circuitry for each of the two stages). The ToT, however goes into an extra step of converting time units into ADC units. The second step sums individual HGCROC channel values from a neighbourhood of four or nine cells into trigger sums for the LD (Figure 2.8, Left) and the HD (Figure 2.8, Right) layouts, respectively. These trigger sums are most commonly referred to as Trigger Cells (TCs). The newly formed TCs are 19-bit and 21-bit positive integers in a linear range for the LD and the HD cases, respectively. However, in the third step, they get compressed into a 7-bit unsigned floating point representation to save the output bandwidth.

Finally, the Trigger Readout Manager (Figure 2.5) formats the trigger data and serialises it so that it is streamed out of the HGCROC across two or four (HD or LD cases) 1.28 Gbit s^{-1} serial differential output e-Links towards the ECON-T ASIC. One HGCROC is responsible for generating a total of 8 or 16 TCs (HD or LD cases) for every BX. Each of the output e-Links streams out a 32-bit word within the 25 ns time window. Each output 32-bit word includes a Header (H) and the four 7-bit TCs (Figure 2.9) for which a given e-Link is responsible.

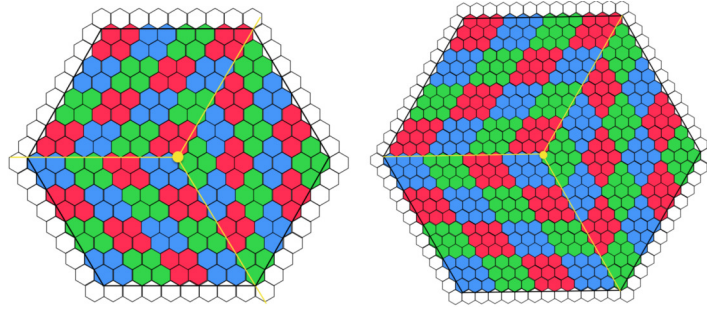


Figure 2.8: An illustration of how TCs are formed in the LD (Left) and HD (Right) Hexaboard layouts [36]. More details can be found in text.

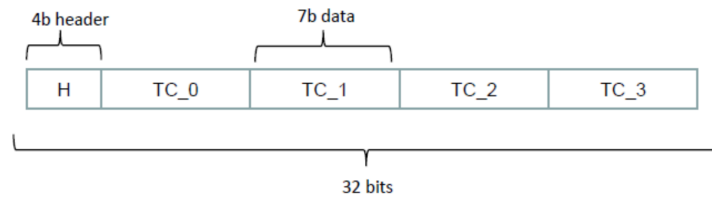


Figure 2.9: The format of the HGCROC trigger frame for one of the output e-Links [36]. The acronyms used are explained in the text.

Functionality Overview - DAQ Path

After the Latency manager block (Figure 2.5), the information about the ADC, ToT and ToA for every BX (for every one of the 78 HGCROC channels) is streamed towards a sequence of steps in the HGCROC DAQ path. The ADC measurement from the previous BX is also available at this stage. Firstly the 12-bit ToT gets compressed down to 10-bit to fit into the available output bandwidth. Following this, a set of three (out of the previously mentioned four) digital values is selected for transmission, depending on which thresholds the signal has passed in the given BX and the mode (e.g. Normal or Characterisation) the HGCROC is configured to be in (Table 2.1).

Table 2.1: Selection of three 10-bit digital values for the DAQ data for a given BX on one channel. The choices depend on which thresholds the signal has passed in the given BX and the mode the HGCROC is configured to be in [36]. The “x” marks the set of selected values for any given signal criteria.

Mode	Signal	ADC(BX-1)	ADC(BX)	ToT(BX)	ToA(BX)
Normal	< ToA threshold	x	x		x (ToA= 0)
Normal	< ToT threshold	x	x		x
Normal	> ToT threshold	x		x	x
Characterisation	any		x	x	x

Two additional 1-bit flags are added to allow the user to remove some ambiguities in the data path. The Tp flag informs if a ToT occurred in a previous BX and that the ADC value may be corrupted (e.g. it may represent a tail of the signal discharge), whereas the Tc flag tells if the second 10 bits is a ToT and not an ADC measurement. Finally, a 32-bit DAQ representation of a sensor channel of the given BX is formed (Figure 2.10).



Figure 2.10: Generalised 32-bit HGCROC channel format [38]. The “ROC Fields” are specified by the three “x” symbols from the relevant row in Table 2.1.

The DAQ data from all channels and from every BX is then taken into a circular buffer (RAM 1) (Figure 2.5) where it can stay for no longer than 512 BX which is a limitation coming from the buffer memory capacity. This is consistent with the CMS-wide constraint on the whole trigger processing and L1A distribution chain: the L1A signal must reach the HGCROCs no later than $500 \times 25 \mu\text{s} = 12.5 \mu\text{s}$ of the BX the L1T wants to select. In RAM 1, the data are protected by a Hamming encoding against SEUs.

In the case of a received L1A, the DAQ data for the selected BX is taken into the L1-triggered event First in First Out (FIFO) memory (RAM 2) (Figure 2.5). The reason for it being part of the readout is de-randomisation. Specifically, it takes the readout logic 41 BX to stream out any event. Hence, in case of a random burst of L1As, the entire burst worth of events from RAM 1 can be transferred into RAM 2 and then streamed out event by event from there. The RAM 2 and the readout logic can be home to no more than 33 events at any given time, from where another very important system limitation arises for CMS. If the CMS L1T was to request the DAQ data for a BX that could not fit into RAM 2 due to its previous occupancy, that data would never be sent out from the HGCROC. This is true for all HGCROCs. Hence, in such a scenario, the entire HGCAL would appear as a volume of the CMS detector that did not provide any data for the given BX. It is very important to understand the rate of these scenarios and how to protect from them (more details can be found in Chapter 5).

Following the DAQ data being taken out of RAM 2, a reverse of the Hamming transformation is done and the data is brought to the DAQ Readout Manager (Figure 2.5) which formats the DAQ data and serialises it so that it is streamed out of the HGCROC across two 1.28 Gbit s^{-1} serial differential output e-Links towards the ECON-D ASIC. The data from the 78 input channels is equally shared between the two output e-Links, such that each one of them is responsible for streaming out data for its share of 39 channels (or a half HGCROC worth of data). Specifically, the format of the data coming from a half HGCROC during the 41 BX long readout of a given event consists of a 32-bit Header, 38×32 -bit words consisting information about standard channels, common mode channels and calibration channels, a 32-bit Cyclic Redundancy Check (CRC) Sum (CRC-32) computed for the Header and all channel fields to allow for a signal integrity check in the ECON-D and a mandatory IDLE word to mark the end of the event (Figure 2.11). The Header contains important information for synchronising with the ECON-D, such as the 12-bit BX, 6-bit event and 3-bit orbit counters indicating when the event being transmitted is coming from. It also passes on three integrity bits (H1, H2, H3) specifying if there were any errors during the Hamming decoding step. The mandatory 4-bit Hd and Tr fields mark the beginning and the end of the event Header.

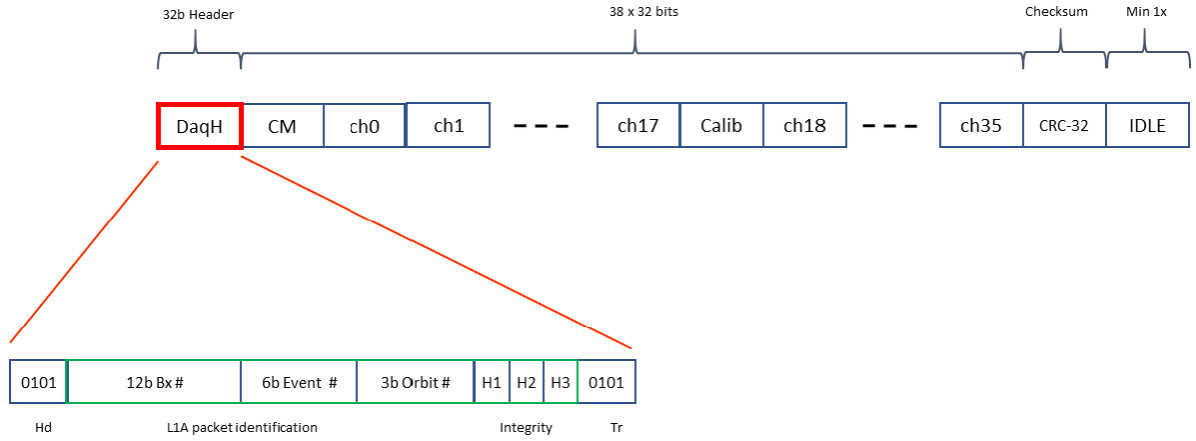


Figure 2.11: The format of the HGCROC DAQ data about an event, as present on one output e-Link across 41 BX [36]. The acronyms used are explained in the text.

2.2.3 The ECON-T ASIC

The following description of the ECON-T ASIC is based on material in [37]. The ECON-T (Figure 2.12) ASIC receives TPG data across 12 input 1.28 Gbit s^{-1} e-Links coming from 3 and 6 HGCROCs in the LD and HD Hexaboard cases, respectively. It then calibrates the TPG data and concentrates it using a selection of several configurable algorithm choices and outputs the TPG data across up to 12 output 1.28 Gbit s^{-1} e-Links. Different parts of the detector will use different algorithms and a different number of ECON-T input or output e-Links enabled. Regardless of the Hexaboard density (and unless partial modules are considered), the ECON-T always utilises all 12 input e-Links from which it receives 48 TCs for every BX. However, the main goal of the ECON-T is to generally reduce the data bandwidth or simply apply a selection with respect to a given criteria to pass some TCs and suppress others. The ECON-T also has designated internal blocks for clock recovery and fast control, as well as an I²C interface for slow control of the device.

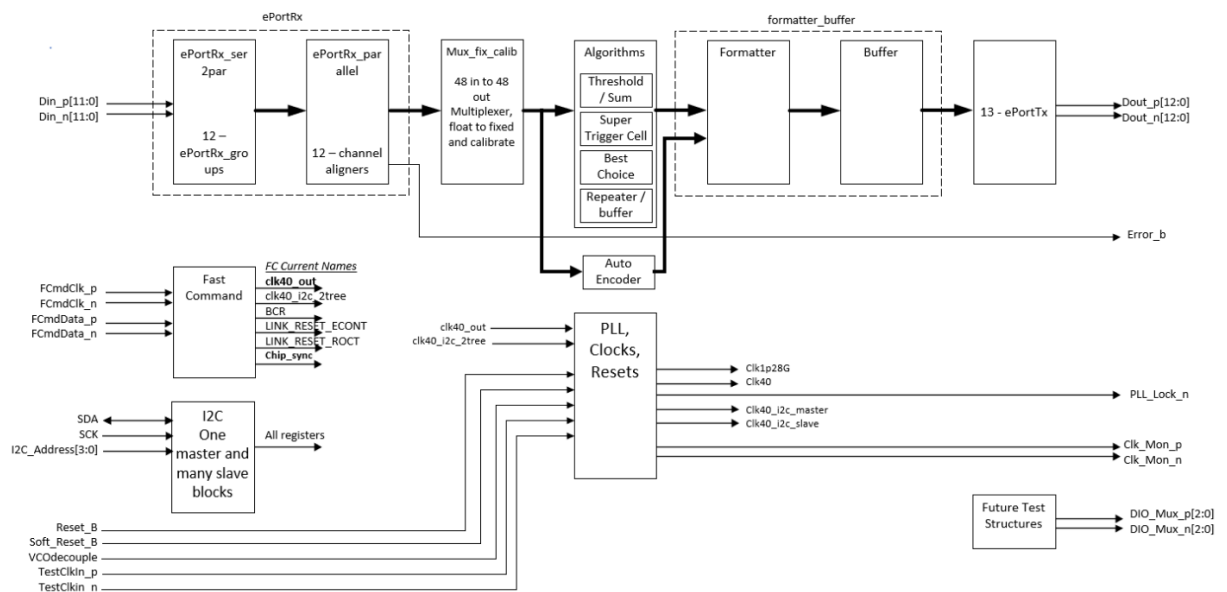


Figure 2.12: The block diagram of ECON-T internal algorithms [37].

Timing and Fast Control

When provided with the 320 MHz fast control reference clock and the 320 Gbit s⁻¹ fast control data stream from the Hexaboard Rafael ASIC, the internal clock recovery logic, which uses a PLL, retrieves the 40 MHz LHC clock frequency and phase. Additionally, it creates various clock frequencies for the internal algorithms (e.g. 40 MHz, 80 MHz, 160 MHz, 320 MHz, 640 MHz, 1.28 GHz). The fast control interface de-serialises the 8-bit fast commands from the incoming 320 Mbit s⁻¹ fast control data stream, allowing the ECON-T to respond according to the HGCal fast control specification [35] (for more details on using fast commands with the ECON-T, see the working examples in Chapter 3).

Slow Control

The ECON-T features an I²C [39] interface for slow control purposes. It is highly configurable, containing approximately 1000 tunable parameters. The internal registers are triplicated to ensure robustness against SEUs. Additionally, there are several GPIO pins that allow the ECON-T to be reset through multiple levels.

Functionality Overview

The TPG data enters the ECON-T through the e-Link Receiver Port (ePortRx) block. The first important step here is to make sure the serial data is sampled with the right phase offset relative to the internal ECON-T clock reference. This step is commonly referred to as the Phase Alignment Procedure (more details can be found in Chapter 3). In order to apply any sort of algorithm across the TCs, the received data is then de-serialised and made parallel again (i.e. the data need to become available as entire words and not individual bits). The reconstructed 32-bit words from different input e-Links are then taken through the Word Alignment Procedure, where the 32-bit word boundaries are found for every input e-Link (more details can be found in Chapter 3). The outcome is that in the end the ECON-T has all 32-bit words from all 12 input e-Links internally recovered and aligned, as they were available before exiting the HGCROCs, making the electrical interface transparent to the algorithms.

The 32-bit words coming in from any input e-Link represent TCs with indices from 0 to 3 (Figure 2.9). In order to distinguish TCs from different e-Links, the Multiplexer steps associates a unique ID to every TC to be in the range from 0 to 47. The ECON-T also performs a conversion from the custom HGCROC 7-bit floating point representation to an 21-bit fixed point representation. The TCs are then calibrated by being multiplied by a 12-bit multiplication constant after which the 11 least significant bits are dropped (resulting in the possibility to calibrate the TCs by 0 to 1.999 51). The ECON-T algorithms then operate on the 22-bit fixed point TC representation.

As of this date, the HGCal baseline is to use the Super Trigger Cell (STC) and the Best Choice (BC) algorithms. The critical constraint for these algorithms is to be able to fit the final TPG description of any given BX over a number of available output 1.28 Gbit s⁻¹ e-Links and the 25 ns time window. In other words, for a given number of available output e-Links N_{eTx} , the final ECON-T output size needs to be smaller than $N_{eTx} \times 32$ bit per BX.

Super Trigger Cell Algorithm The STC is a fixed-format algorithm that reduces the output data bandwidth by creating sums of neighbouring TCs. There are several options

available, but of interest to this thesis is the STC4 4E3M variant, which makes 12 sums of four neighbouring 22-bit TCs to form the 12 STCs (Figure 2.13, Left). For illustration, the formation of STCs when STC16 is used is also given (Figure 2.13, Right).

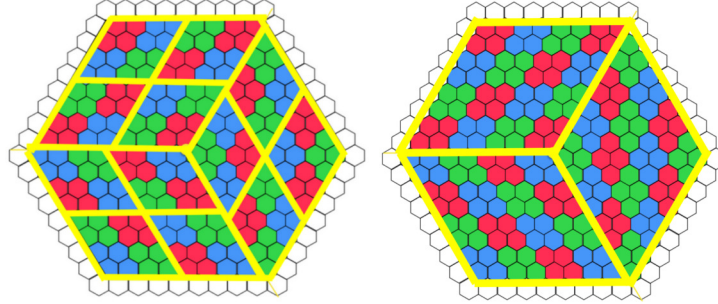


Figure 2.13: An illustration of how STCs are formed in the LD Hexaboard layouts for the STC4 (Left) and STC16 (Right) ECON-T algorithms [37]. More details can be found in text.

Before being sent out, the STCs are encoded into a 7-bit floating point representation, using a 4-bit exponent and a 3-bit mantissa (hence the 4E3M naming). As prepared by the Formatter Buffer, the ECON-T STC data format (Figure 2.14) consists of a 4-bit Header to convey the information about what BX the data packet comes from, a sequence of twelve 2-bit fields to map which TC within the STC had the highest energy, the 12 STC energies and a padding of trailing zeros.

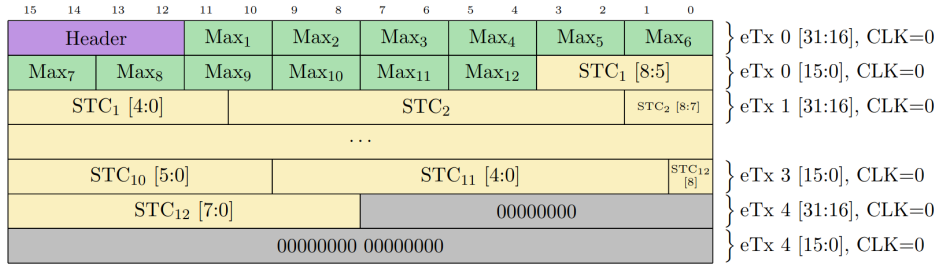


Figure 2.14: An example of the ECON-T STC4 general data format where the ECON-T has 5 available output e-Links [37]. The individual fields are explained in the text.

Best Choice Algorithm The BC is a fixed format algorithm that reduces the output data bandwidth by sending out only the number of TCs with the highest energy that can fit into the number of available output e-Links. Depending on the number of transmitted TCs, there are the Low and High Occupancy formats, but the latter is the only one presented here. As prepared by the Formatter Buffer, the ECON-T BC data format (Figure 2.15) consists of a 4-bit Header to convey the information about what BX the data packet comes from, an 8-bit module sum (e.g. the energy sum of the entire module), a 48-bit Channel map of which TCs made the selection, the corresponding TC 7-bit charges and a padding of trailing zeros.

Regardless of the selected algorithm, the ECON-T data is streamed out over a number of enabled output 1.28 Gbit s^{-1} differential serial e-Links towards the designated lpGBT.

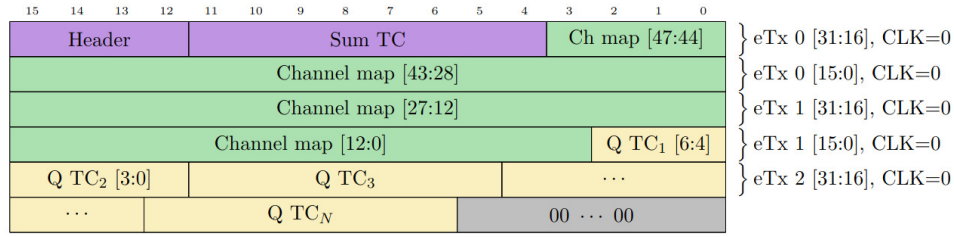


Figure 2.15: An example of the ECON-T BC high occupancy data format where the ECON-T has 4 or more available output e-Links and is transmitting 9 or more TCs per BX [37]. Individual fields are explained in the text.

2.2.4 The ECON-D ASIC

The following description of the ECON-D ASIC is based on material in [38]. The ECON-D (Figure 2.16) ASIC receives DAQ data across up to 12 input 1.28 Gbit s^{-1} e-Links, coming from 3 and 6 HGCROCs in the LD and HD Hexaboard cases, respectively. It performs different kinds of checks over the received HGCROC data, stores the data of the current event being processed into a memory buffer, processes the event, formats the event and takes it to the output interface where the data are streamed out over up to 6 output 1.28 Gbit s^{-1} e-Links. The key functionality considering the ECON-D processing step is the highly configurable Zero-suppression algorithm. Different parts of the detector will have different choices for how the zero-suppression thresholds are configured and a different number of ECON-D input or output e-Links enabled. The main goal of the ECON-D is to generally reduce the output data bandwidth or simply apply a selection with respect to

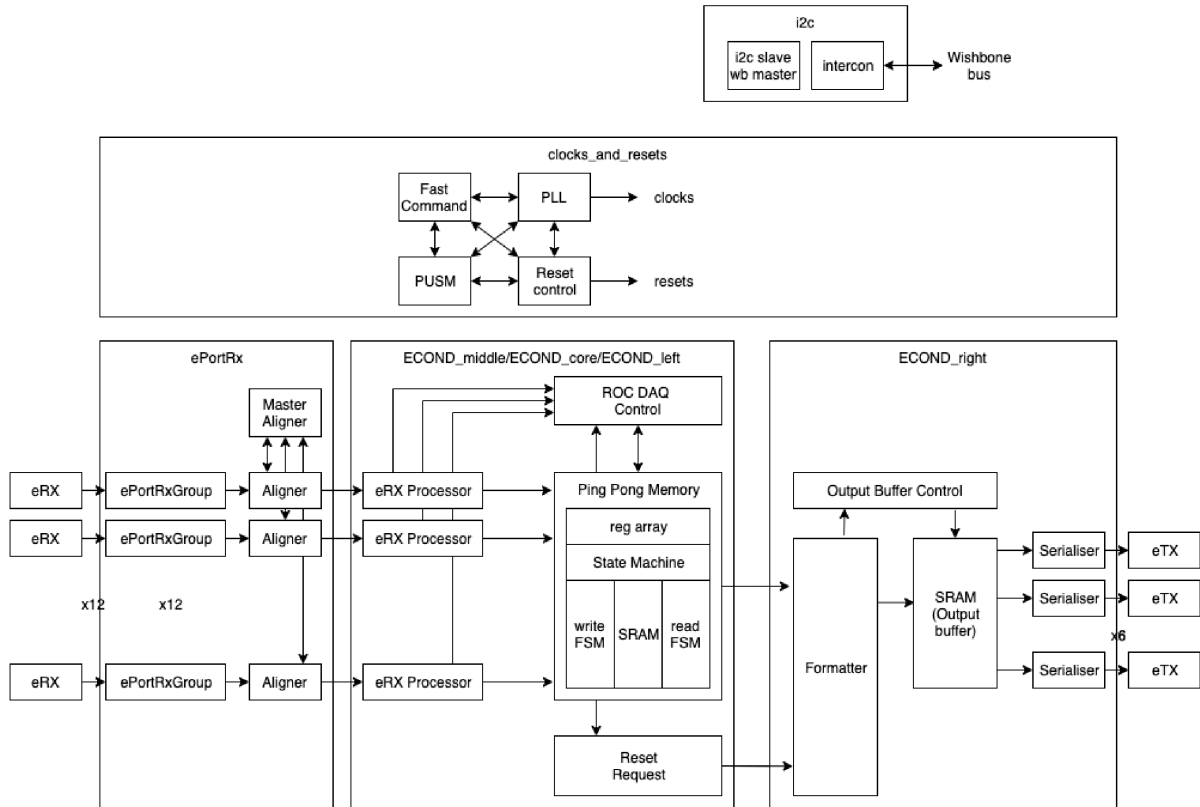


Figure 2.16: The block diagram of ECON-D internal algorithms [38].

a given criteria to pass some channels and suppress others. It is important to note that the DAQ BE can issue a special fast command called the Non-Zero-Suppressed L1A (L1A NZS) that can request the ECON-D to send a non-suppressed packet (i.e. ignoring the previously set zero-suppression thresholds). Before outputting the data, the ECON-D also introduces a few fields in the output format that contribute to monitoring the data integrity (and allowing for a limited online correction of bit errors) during its transmission to the DAQ BE. The ECON-D also has designated internal blocks for clock recovery and fast control, as well as an I²C interface for slow control of the device.

Timing and Fast Control

Upon receiving the 320 MHz fast control reference clock and the 320 Gbit s⁻¹ fast control data stream from the Hexaboard Rafael ASIC, the internal clock recovery logic, which is centred around a PLL, retrieves the 40 MHz LHC clock frequency and phase. Various clock frequencies are made available for the internal algorithms (e.g. 40 MHz, 80 MHz, 160 MHz, 320 MHz, 640 MHz, 1.28 GHz). The Fast Control interface de-serialises the 8-bit fast commands from the incoming 320 Mbit s⁻¹ fast control data stream, allowing the ECON-D to respond according to the HGCal fast control specification [35] (for more details on using fast commands with the ECON-D, see the working examples in Chapter 3).

Slow Control

The ECON-D features an I²C [39] interface for slow control purposes. It is highly configurable, containing approximately 5000 tunable parameters. The internal registers are triplicated to ensure robustness against SEUs. Additionally, there are several GPIO pins that allow the ECON-D to be reset through multiple levels.

Functionality Overview

The notes about Phase and Word Alignment presented for the ECON-T (more details in Section 2.2.3), also apply for the ECON-D. Hence, the rest of this section focuses on the specific functionalities of the ECON-D. Consider the HGCROC DAQ data frame (Figure 2.11) and the Standard ECON-D event packet data format (Figure 2.17) to explore the main ECON-D functionality.

Each of the ECON-D e-Link inputs coming from the HGCROCs corresponds to data from half of an HGCROC (i.e. 36 standard channels, 2 common mode channels and 1 calibration channel). Hence, the ECON-D is structured to process HGCROC halves individually. Over each half-HGCROC worth of data, the ECON-D computes the CRC (as the HGCROC did) and compares the result with the CRC-32 field that the given half HGCROC had transmitted. The result of this CRC check is flagged in the Most Significant Bit (MSB) of the Stat field in the ECON-D sub-packet header (green). Looking into all the half HGCROC headers, the ECON-D investigates what the different half HGCROCs sent in the header start (Hd) and header stop (Tr) fields, the event counter (Event) field, the BX counter (Bx) field and the orbit counter (Orbit) field. The ECON-D then uses a vertical reconstruction method (with a programmable channel count threshold) in order to attempt on converging to what the actual values are, even if not every half HGCROC necessarily agrees. In normal operation, the ECON-D is configured to then send out the results of this vertical reconstruction when describing the given event through its Event Packet Header (purple) BX number, L1A number and Orb fields. The ECON-D also informs the DAQ BE about how well did the vertical reconstruction go through the Event Packet Header H/T and EBO fields (for either of the two it can be Perfect, Good, Failed

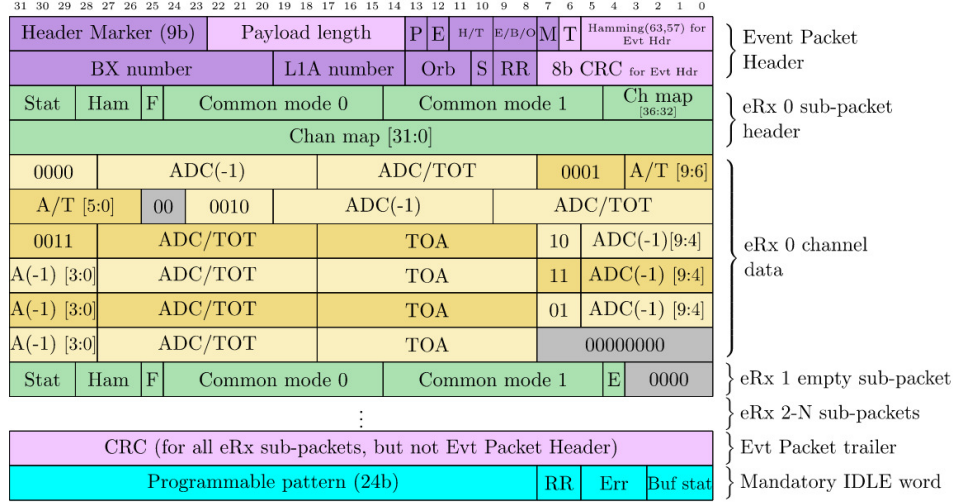


Figure 2.17: The standard ECON-D data format for N active eRx [38]. The individual fields are explained in the text.

or Ambiguous depending on how many half HGCROCs agree). Through the M field in the Event packet Header the ECON-D also marks whether the transmitted BX, event and orbit counters agree with the ECON-D internal counters (i.e. this allows understanding if the HGCROCs and ECON-D agree about these counters). Through its middle and Lowest Significant Bit (LSB) of the Stat field, the ECON-D also flags whether or not there was a match between the ECON-D transmitted event, BX and orbit counters with the corresponding individual HGCROC half values that were received. The ECON-D provides a 6-bit Hamming field computed for the entire 64-bit Header (except the Header CRC field) which can be used to perform a single bit error correction in the DAQ BE in case of a bit error in the ECON-D Header. The ECON-D also computes two CRCs and packs them into the output data format: one for the ECON-D Event Packet Header itself (except the Hamming field) and the other for all HGCROC sub-packets, but not the Event Packet header. The bulk of the HGCROC payload information (e.g. ADC, ToT and ToA information about the HGCROC channels that made the zero-suppression cut) is present in the channel data portion of the data format (yellow). The HGCROC common mode channels are present in the designated sub-packet headers. The sub-packet headers also have a 37-bit channel bit map to convey information about which standard channels make the zero-suppression cut or not (in doing so the calibration channel is treated as a standard channel). The sub-packet headers also pass through the 3-bit result of the HGCROCs internal Hamming check.

There are two more ECON-D data formats (Pass-through and Truncated) other than the Standard data format relevant for this thesis that will be briefly mentioned. The ECON-D can be configured to pass (most of) the HGCROC data through as it was when being received by the ECON-D (Figure 2.18). This is useful when it is important to debug the HGCROC $\rightarrow$ ECON-D interface (more details can be found in Chapter 4).

The ECON-D output buffer can hold no more than 12×256 32-bit words at any given moment, from where an important system limitation arises. If the CMS L1T was to issue a request for a BX that could not fit into the ECON-D Output buffer due to its previous occupancy, the data for that BX would be output in the truncated event data format (Figure 2.19). The truncated format is generally not desired as it contains only a minimal

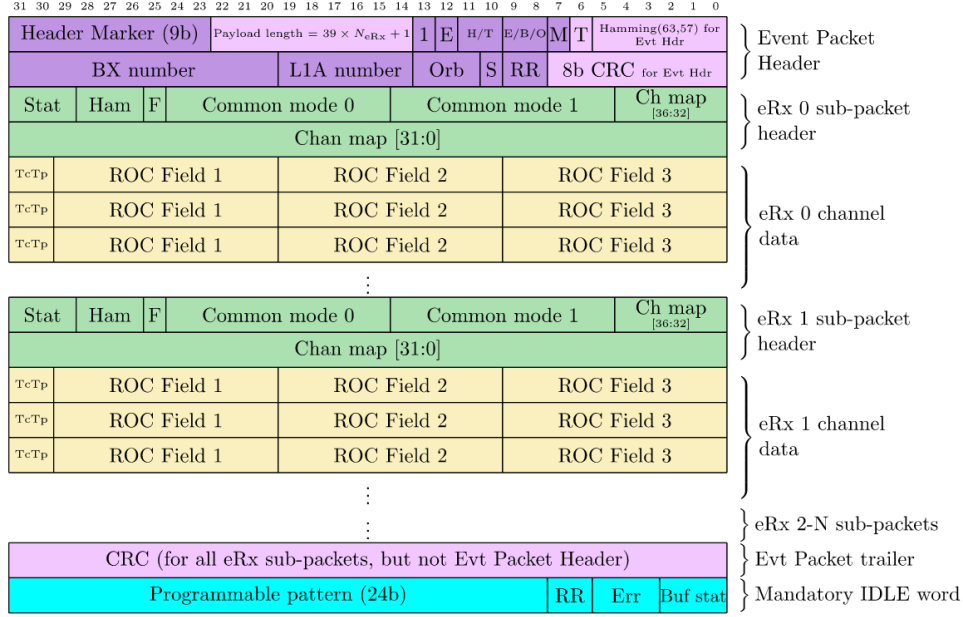


Figure 2.18: The Pass-through ECON-D data format for N active eRx [38].

set of information for the given event. This type of ECON-D packet is bound to occur when the DAQ BE is issuing L1As at a moderately high average rate.

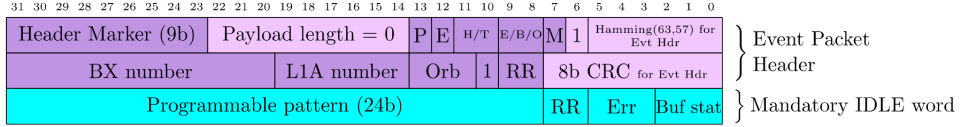


Figure 2.19: The truncated ECON-D data format [38].

The P and the T fields in the ECON-D Event Packet header flag whether the ECON-D is in Pass-through mode and whether the packet is truncated, respectively. The Header Marker field specifies the beginning of an ECON-D packet, the Payload length specifies the ECON-D packet length in terms of the number of 32-bit words including the HGCROC sub-packet headers (green), ROC fields (yellow) and the CRC trailer. Every ECON-D word must end with a Mandatory Idle word (cyan). The E field specifies whether the event Hd and Tr fields are received from HGCROCs when they were expected to. Through the RR field, the ECON-D can request a reset from the DAQ BE.

Regardless of the selected algorithm, the ECON-D data is streamed out over a number of enabled output 1.28 Gbit s^{-1} differential serial e-Links towards the designated lpGBT.

2.2.5 The lpGBT ASIC

The following description of the lpGBT ASIC is based on material in [40]. It has already been demonstrated how the lpGBT and the VTRx+, together with their BE inverses make the FE $\leftrightarrow$ BE interface transparent for the FE and BE algorithms (more details can be found in Section 2.1.4). Here, however, the lpGBT is in focus (Figure 2.20). While the lpGBT ASIC is by design a multipurpose device (e.g. applicable for various high energy physics experiments), its main functionalities are presented here on the example of its

use-case in HGCAL. Being the entry point for the BE algorithms in the FE and vice versa, the lpGBT should be explored with respect to its roles in HGCAL when it comes to: timing and fast control distribution, slow control and data acquisition. To achieve that, the down-link and up-link lpGBT directions are summarised. In HGCAL, lpGBTs are generally used in two main modes of operation. One is the Transceiver mode, where the lpGBT implements a bi-directional interface between FE $\leftrightarrow$ BE (e.g. lpGBTs used in the DAQ sub-system, commonly referred to as DAQ lpGBT). The other is the Transmitter mode, where the lpGBT implements a uni-directional interface between FE $\rightarrow$ BE (e.g. lpGBTs used in the TPG sub-system, commonly referred to as the TPG lpGBTs).

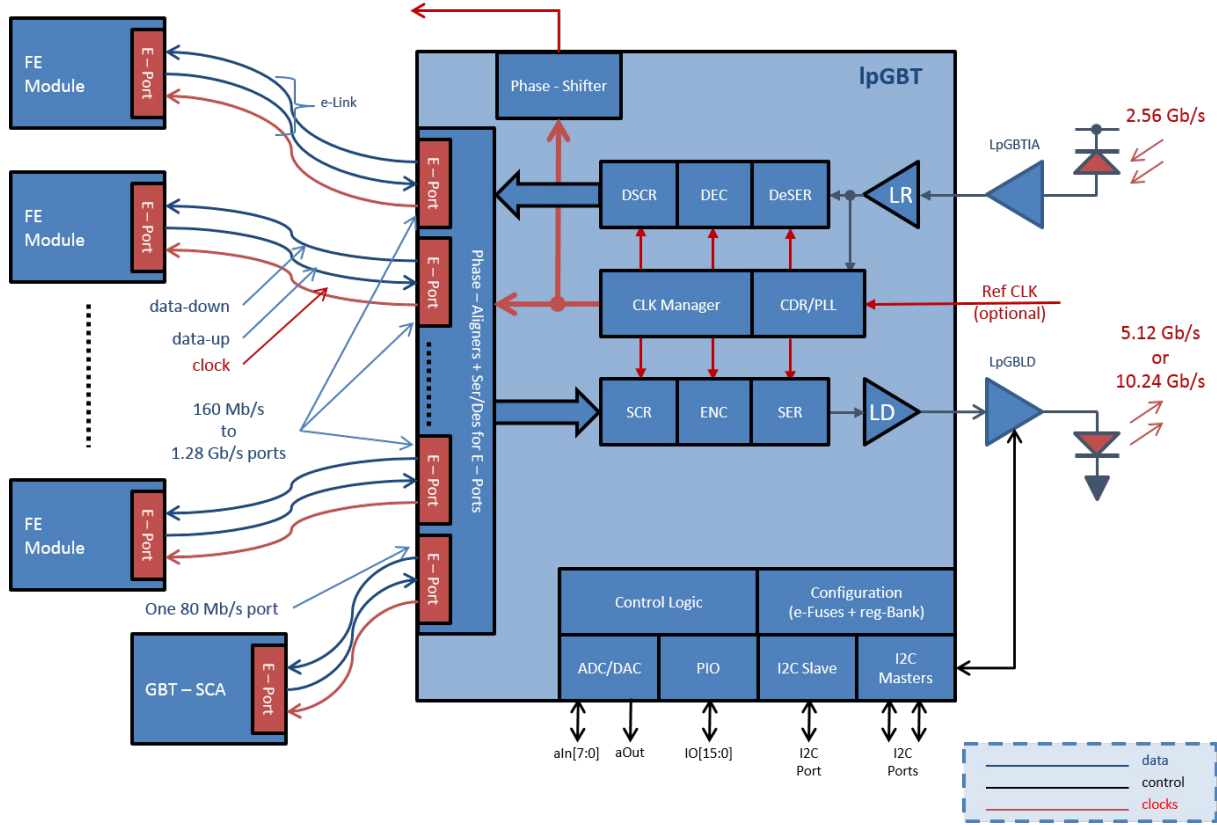


Figure 2.20: A block diagram of the lpGBT architecture [40]. Relevant acronyms are explained in the text.

The Down-link Data Path

The DAQ BE sends data to the FE DAQ lpGBTs in the format of the 64-bit down-link lpGBT frames (Figure 2.21) every 25 ns (or at 40 MHz) and that is why the data rate in a down-link fibre is $64 \text{ bit} \times 40 \text{ MHz} = 2.56 \text{ Gbit s}^{-1}$. Because the data was scrambled in the BE while the down-link frame was formed, it is ensured that there are enough logical "0" $\leftrightarrow$ "1" transitions in the down-link in order for the lpGBT Clock and Data Recovery (CDR) circuits to recover the clock reference for the lpGBT. The down-link is then de-serialised, decoded and de-scrambled (in DeSER, DEC and DSCR lpGBT blocks, respectively) so that the information that was originally modulated in the BE is now also available in its original form in the FE. In this process, the 4-bit Header (H) field in the down-link frame was important to mark the beginning of the frame and the 24-bit Forward Error Correction (FEC) field was used to recover from up to 12 consecutive wrong bits during frame transmission from the BE to the FE. The fields of the data frame that carry

the information intended for the FE ASICs to receive are embedded in the 2-bit Internal Control (IC), 2-bit External Control (EC) and the 32-bit Data Payload (D) fields.



Figure 2.21: The lpGBT down-link 64-bit frame format [40]. Individual fields are explained in the text.

The D field can carry four 8-bit fast commands. Focusing on one of these, it is then serialised and output through a serial differential e-Link interface towards the Rafael ASIC. Because there is a new 8-bit fast command re-appearing in this part of the frame at 40 MHz, the corresponding e-Link needs to be configured to run at $8 \text{ bit} \times 40 \text{ MHz} = 320 \text{ Mbit s}^{-1}$. Similarly, the lpGBT is configured to output a 320 MHz clock on a separate e-Link interface. Depending on where the used DAQ lpGBT is in the detector, it needs to be configured accordingly to output a specific number of fast control data and clock streams.

Through the IC field, the BE can configure the internal registers of the DAQ lpGBT. That kind of action does not only allow configuration of the lpGBT to a baseline mode of operation, but it can also lead to slow control transaction with other FE ASICs. For example, by writing to the lpGBT, it is possible to configure one of the lpGBT I²C Masters, requesting a read or write transaction to be performed with one of the FE ASICs (in the HGCal silicon part, that is the interface used for configuring the VTRx+). Alternatively, it can be used to configure one of the lpGBT GPIOs to output a logical "0" or "1" in the output, allowing to power cycle or reset FE ASICs connected to it. The bandwidth of the IC field is $2 \text{ bit} \times 40 \text{ MHz} = 80 \text{ Mbit s}^{-1}$, meaning that it takes a number of down-link lpGBT frames until a full slow control message is assembled.

The EC field is directly forwarded to an output e-Link, designated to output a data rate of 80 Mbit s^{-1} . The format of the data in EC is exactly the same as in the IC field. The EC interface is meant for conveying slow control information to ASICs external to the DAQ lpGBT. In the silicon part of HGCal, this interface of the DAQ lpGBT is used to configure TPG lpGBTs. This effectively allows accessing TPG lpGBTs as if they had an IC interface, even though they are transmitters that do not have a down-link. This “trick” is widely used in HGCal and allows TPG lpGBTs to be also used as I²C masters when writing to the HGCal ASICs (i.e. HGCROC, ECON-T and ECON-D) or driving the logical level of a GPIO pin.

The Up-link Data Path

In the up-link, the same concepts generally apply as for the down-link, but in the other direction of the data flow. Hence only some up-link specifics are covered here.

While the down-link format is always the same, the up-link format can vary depending on the lpGBT configuration. In HGCal, a FE (TPG or DAQ) lpGBT sends data to the (TPG or DAQ) BE algorithms in the format of the 256-bit up-link lpGBT frames (Figure 2.22) every 25 ns (or at 40 MHz) and that is why the data rate in an up-link fibre is $256 \text{ bit} \times 40 \text{ MHz} = 10.24 \text{ Gbit s}^{-1}$. The fields of the data frame that carry the information to be sent from the FE ASICs to the BE algorithms are embedded in the 2-bit IC, 2-bit EC and the 224-bit D fields. The 6-bit LM field can be used for latency measurements, although it will not be discussed in more detail. Before being sent to the BE, the data is

scrambled, encoded and serialised (in SCR, ENC and SER lpGBT blocks, respectively). The 20-bit FEC field is used in the BE to recover up to 10 consecutive wrong bits during frame transmission from the FE to the BE.

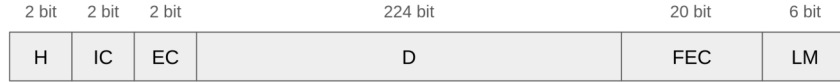


Figure 2.22: The lpGBT up-link 256-bit frame format specific to the HGCALE use case [40]. Individual fields are explained in the text.

The D field can carry seven 32-bit ECON words. Each one of them comes after de-serialising the data coming from a serial differential 1.28 Gbit s^{-1} input e-Link. Similarly to the ECONs in the HGCROC $\rightarrow$ ECON interface, the lpGBT needs be carefully configured to appropriately sample the input e-Links (more details can be found in Section 2.2.3). However, the Word Alignment procedure must be handled by the BE (more details can be found in Chapter 3), because the lpGBT is agnostic of the 32-bit word boundaries in the HGCALE specific ECON outputs.

The IC field can be used to read internal registers of the lpGBT. Some of these registers store the results of I²C transactions the lpGBT I²C Masters carry out with the FE ASICs or the logical level of an lpGBT GPIO pin. Hence, the IC interface is the path used for the “reading” part of slow control, orchestrated by the DAQ BE.

The EC field directly transfers what is brought to the EC input of the lpGBT to the BE. It should be noted that the HGCALE EC “trick” mentioned for the down-link data path is also applicable for the up-link data path, allowing TPG lpGBTs to be also used as I²C masters when reading from the HGCALE ASICs (i.e. HGCROC, ECON-T and ECON-D) or reading the logical level of a GPIO pin.

In HGCALE, neither IC nor EC fields are populated with meaningful content in TPG lpGBT up-links, only in the DAQ lpGBT up-links.

2.3 The HGCALElectronics DAQ BE

The following description of the DAQ BE is based on material in [34]. There are two effectively orthogonal systems that share some of the DAQ BE hardware. The first major part is the DAQ BE system itself which is related to the acquisition of fine granularity data from the FE, fast control and timing distribution to the FE and slow control and monitoring of the FE. The other minor part is related to the acquisition of NZS data from some specifically selected parts of the detector controlled through the L1T Endcap Muon Track Finder (EMTF) interface. This section will only discuss the DAQ BE system, exploring it in two main steps. Firstly, the focus is on the structure and organisation of the system (e.g. organisation of multiple BE boards together, numbers of BE boards and links, bandwidth considerations, etc.). Secondly, the focus shifts to one DAQ BE board and the main firmware algorithms implemented within its FPGA.

2.3.1 System Structure and Organisation

The DAQ BE system is logically split into two halves, with each one servicing one of the two FE endcaps. The DAQ BE hardware responsible for a single endcap (Figure 2.23) is further organised into three parts, such that each one is responsible for servicing a $120^\circ\phi$ sector of an endcap. Such an organisation is chosen based on the symmetries in the hardware of the HGCAL detector FE. The DAQ boards implement the HGCAL DAQ BE algorithms while the DTH boards are used for data aggregation (DAQ800) from multiple DAQ boards and as hubs for the distribution of timing and control (DTH400) signals to the DAQ boards. The DAQ and DTH boards together occupy six crates held together by three racks, per endcap. In the following, the system structure and organisation is presented from the perspective of the data acquisition as well as timing and control perspective. A short note is given on what part of the DAQ bandwidth is actually useful.

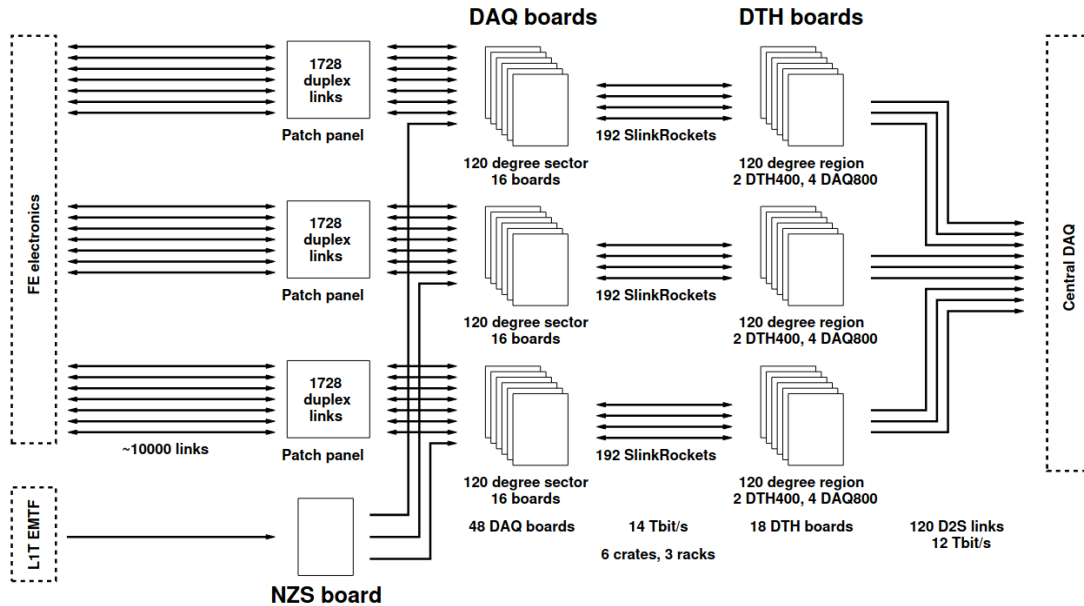


Figure 2.23: Structure of the DAQ BE system for one endcap. At the time the block diagram was drawn it was taken as a worst case scenario that there will be 10 000 links (50% of them dark) routed between the FE and the DAQ BE and that a stage of patch panels will also be needed. Since then matters have improved and the BE may not even need any patch panels. Regardless, the DAQ BE is prepared to connect to as many as 5184 FE lpGBT links [34].

Data Acquisition Aspects

The DAQ BE is prepared to connect to as many as 5184 lpGBT link pairs (i.e. $10.24 \text{ Gbit s}^{-1}$ up-link and 2.56 Gbit s^{-1} down-link) routed from the DAQ lpGBTs in the FE, per endcap. Each DAQ BE board will be a Serenity board with a single VU13P FPGA having ten pairs of 12 channel bi-directional modules. Nine pairs of these modules are used to implement the interface with $9 \times 12 = 108$ DAQ lpGBTs in the FE. Hence, $5184/108 = 48$ boards are needed per endcap (16 per sector). The remaining bi-directional module pair on a DAQ BE board is used to implement the interface to a DAQ800 DTH board via 12 SLinkRocket 25 Gbit s^{-1} streams. Hence, there are a total of $16 \times 12 \times 3 = 576$ SLinkRocket links in the interface with the 12 DAQ800 DTH boards [45], per endcap. These DTH boards further aggregate the data to be sent across a smaller number of links and at yet again

higher rates, ultimately sending 12 Tbit s^{-1} across 120 D2S 100 Gbit s^{-1} links towards the cDAQ system, per endcap. This is significantly smaller when compared to the full data bandwidth of $5184 \times 10.24 \text{ Gbit s}^{-1} \approx 50 \text{ Tbit s}^{-1}$ at the input of the DAQ BE, per endcap. To understand why, a closer look will be given into what portion of the full input bandwidth to the DAQ BE is actually useful.

The Usable Part of the DAQ Bandwidth

With each DAQ lpGBT up-link sending a user rate (i.e. the rate coming from the 224-bit D field in the lpGBT up-link frame as shown in Figure 2.22) of $40 \text{ MHz} \times 224 \text{ bit} = 8.96 \text{ Gbit s}^{-1}$ that allows to approximate the total user data rate coming from the detector FE to the DAQ BE to 45 Tbit s^{-1} , per endcap. However, the DAQ event data rate excludes the times when the system is idling (e.g. not transmitting L1T selected DAQ data between the FE and the BE). The total data size per event is expected to be around 3 MB or 24 Mbit per event. At 750 kHz average L1A frequency, this gives a **total average usable DAQ data rate** of $24 \text{ Mbit} \times 750 \text{ kHz} = 18 \text{ Tbit s}^{-1}$ (9 Tbit s^{-1} per endcap). This implies that the DAQ lpGBTs are sending idles around 80% of the time on average and that the usable data rate coming from a single DAQ lpGBT up-link is approximately 2 Gbit s^{-1} on average. The estimated number of DAQ lpGBT up-links coming from the FE is actually 1250 per sector (this number is not final), meaning that the DAQ BE has left a margin of about 25% of links possibly being unused. Therefore, the usable data rate going into a DAQ BE board can be estimated to $0.75 \times 108 \times 2 \text{ Gbit s}^{-1} \approx 160 \text{ Gbit s}^{-1}$ on average. The SLinkRocket output streams provide a bandwidth of $12 \times 25 \text{ Gbit s}^{-1} = 300 \text{ Gbit s}^{-1}$ per DAQ BE board. The goal is to load balance the DAQ board inputs to the average input rate. Fluctuations around this rate are then handled internally in the boards by buffering (e.g. for de-randomisation of the L1A rate).

Timing, Fast Control and Slow Control Aspects

The six (per endcap) DTH400 boards [45] implement the TCDS2 system. Each one uses the ATCA crate back-plane to distribute the clock reference and fast commands to the DAQ BE boards. The DTH400 board also provides a 10 Gbit s^{-1} Ethernet interface to an ATCA crate, granting the DAQ BE access via the Local Area Network (LAN). This gives an entry point for system slow control. Through their interface with the FE lpGBTs, the DAQ BE boards take part in timing distribution to the FE ASIC as well as fast control and slow control of the FE ASICs.

2.3.2 Functionality Overview

One Serenity DAQ BE board is home to a VU13P FPGA where the DAQ BE firmware is implemented (Figure 2.24). Some of these blocks are not specific to HGCal (grey) while there are four major HGCal-specific (colour) blocks: Slow Control, Timing and Fast Control Distribution (TCDS Distribution Logic and Autorecovery), the Capture Blocks and the Readout Blocks (Interconnect and SLink Interface blocks). Each Serenity board has a Board Controller, external to the FPGA.

Slow Control

Through the CPU Link, the board controller (which is also accessible via LAN) can configure and monitor the DAQ FPGA algorithms (through AXI or IPBus interfaces)

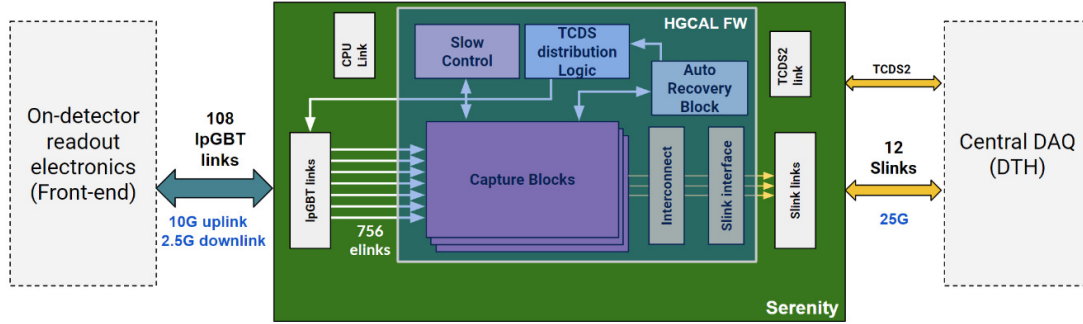


Figure 2.24: A block diagram of the main algorithm blocks present in one DAQ BE board. The coloured blocks are HGCAL specific [34].

and also the FE ASICs, which happens through the Slow Control Firmware block, closely coupled to the lpGBT links. Taking into account all the different interface rules that apply when wanting to reach a particular FE ASIC (e.g. ASIC addressing, using intermediate ASICs to act as slow control masters, etc.), the software prepares a slow control command which is then brought to the Slow Control firmware block via the CPU Link. Each slow control command is stored into a buffer after which the Slow Control block transforms it into streams of 2-bit IC or EC lpGBT fields. These streams are then brought as inputs to the lpGBT links where they take part in the formation of lpGBT frames. Similar concepts apply in the up-link direction. The slow control block needs to support an interface with 108 DAQ lpGBT bi-directional links per DAQ BE board.

Timing Distribution and Fast Control

Through the TCDS2 link, the DAQ BE FPGA receives the TCSDS2 fast control commands and the LHC clock reference, coming from the DTH400 board (via the ATCA back-plane). These commands are CMS global (e.g. L1A) and are translated into their equivalent words from the HGCAL fast command dictionary [35]. The HGCAL foresees adding some fast commands independent of the global TCDS2 stream. Some examples are: fast commands for regular FE $\leftrightarrow$ BE synchronisation maintenance (e.g. regular orbit or event counter resets), special fast commands requested by the board controller via software (e.g. local commands specifically targeting particular detector regions for NZS) and fast commands requested by the Auto Recovery block (e.g. irregular FC reset commands). The Auto Recovery block “spies” on the event data in the Capture Blocks, in particular the flags indicating errors for the given event and issues fast commands attempting to recover from the causes of these errors. Fast commands from different sources are taken through a prioritisation logic that decides which fast commands are more important (e.g. L1A) than others (e.g. Idle) and tries to combine multiple requests if permitted (e.g. L1A and Bunch Counter Reset, etc.). Similarly to how the slow controls contribute to the IC and EC field, the 8-bit fast control commands contribute to the D field bandwidth of the down-link lpGBT frames. The critical difference is that a fast control command is sent as part of one lpGBT down-link frame, whereas a slow control command can take many frames to be transmitted.

Capture Blocks

An important driver of this part of the system architecture is that the DAQ BE needs to handle situations when the output e-Links of a single ECON-D ASIC are shared by two DAQ lpGBTs. For example, there can be an ECON-D which has more than seven

output e-Links enabled in some places in the detector, which cannot be transmitted by a single DAQ lpGBT alone. Despite the latest FE mapping not anticipating such cases, the DAQ BE still does take that as a specification parameter (e.g. for added flexibility if the mapping plans change). Hence, the basic chunk of data the DAQ BE operates with is two DAQ lpGBT's worth of data (i.e. 14 e-Links). Logically, there are four major steps within a DAQ BE Capture Block: the ECON-D Packet Assemblers, the Main Buffer, the Event Builder and the Event Buffer. There are 54 Capture blocks in the DAQ BE (i.e. one for each DAQ lpGBT pair), each containing 12 ECON-D Packet Assemblers, one Main Buffer, one Event Builder and one Event Buffer.

The 12 ECON-D Packet Assemblers receive detector event data from the two DAQ lpGBTs (from up to 12 ECON-Ds). This step is configurable to accommodate the exact ECON-D $\rightarrow$ DAQ lpGBT mapping. The data corresponding to each ECON-D is assembled together as an arranged group of words belonging to that ECON-D only. Up to 12 valid ECON-D packets are put into the Main Buffer (if allowed by the occupancy). The start of an ECON-D packet is recognised by the mandatory idle word before each packet followed by the ECON-D Header Marker in the ECON-D packet header (more details about ECON-D data format can be found in Section 2.2.4). Each ECON-D packet is considered valid if it has the correct CRC Header value (because the correct payload length is critical for the DAQ BE algorithm). The Event Builder aggregates the assembled ECON-D packets (by reading from the Main Buffer) into event packets (each not having data from more than 12 ECON-Ds) and stores these in the Event Buffer. At this point in the algorithm the ECON-D orbit, bunch and event counters are checked versus their expected DAQ BE values (the ECON-D packet CRC check is not implemented yet). For every event packet, a header (Figure 2.25) is also written to indicate the DAQ BE L1A timestamp and the status of the 12 individual ECON-D packets.

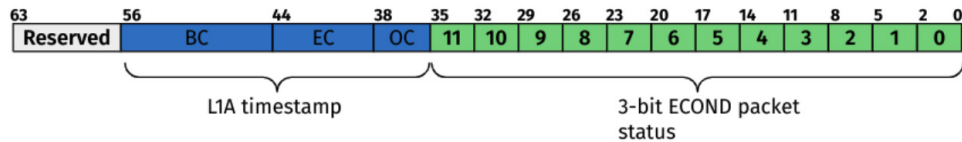


Figure 2.25: DAQ BE Header indicating the L1A timestamp (bunch counter, event counter and orbit counter, according to the DAQ BE notion of time) and status of the 12 ECON-D packet status (e.g. normal, missing, rejected due to counter mismatch, unconnected ECON-D, etc.).

Readout Blocks

Finally, the data from the 54 Event Buffers (up to $54 \times 12 = 648$ ECON-Ds) is aggregated and streamed out towards the cDAQ across the 12 SLink Rocket streams. In order to balance the data rates across the output streams (to minimise the need for back-pressure on any of them), an Interconnect firmware block is put in place which allows a sufficient degree of flexibility for the Event Buffer to SLink mapping. Each SLink packet (output over a single SLink output stream) for a given detector event consists of an SLink Header and an SLink Trailer, wrapped around the Payload (Figure 2.26). The BOE and EOE fields mark the beginning and the end of an SLink packet, respectively. The Global event ID is the overall L1A number since the start of the DAQ run. DAQ CRC is calculated over the full event, including the SLink Header and Trailer. The Event Length specifies the number of 128-bit words in the SLink packet. The BX ID and the Orbit ID fields mark the bunch crossing and the orbit counters for the event as per the TCDS2 notion of

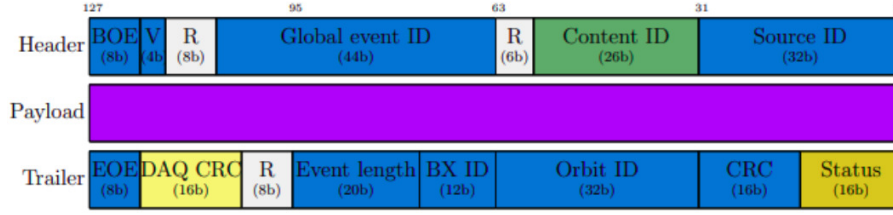


Figure 2.26: SLink packet format [47]. More details can be found in text.

time. The payload consists of a variable number (given by the Interconnect) of DAQ BE event packets (each containing up to 12 ECON-D packets). More details about the SLink packet format can be found in [47].

2.4 The HGCAL Electronics TPG BE

The following description of the TPG BE is based on material in [34]. The TPG BE receives TPG data from the detector FE, processes it to generate the final trigger primitives and sends them out to the L1T. The final trigger primitives consist of two parts: the 3D clusters formed from the selected TCs, each described by a set of physical properties (e.g. energy, position, shape) and complete module energies summed into 2D $\eta \times \phi$ energy histograms. The former and the latter are commonly referred to as *clusters* and *tower energies*, respectively. The tower energies are there so that the full energy of the module (and with fair spatial resolution in the transverse plane) is always made available for the L1T, which is not the case with the clusters that are formed only from a selection of TCs. Additionally, the number of clusters shared with the L1T is also limited by TPG BE output bandwidth. The processing is effectively split into two parts: the Stage 1 and the Stage 2 of the TPG BE. However, the Stage 1 mostly acts as a data preparation step for the main algorithm that takes place in Stage 2.

The TPG BE will receive a timing reference and fast controls from the TCDS2. It will also send some level of information to the cDAQ to complement what the DAQ BE is sending (e.g. for cross-checking the DAQ vs. TPG data). A minimal set of information for every L1T selected event (i.e. counters and error flags) will be always sent while for some L1T selected events (extended readout events) that set of information will be much more complete. Frequency and content of the extended readout events is not precisely known at this time.

This section covers the TPG BE in three steps. Firstly, the focus is on the structure and the organisation of the system (e.g. organisation of multiple BE boards together, numbers of BE boards and links, bandwidth considerations, very high-level considerations of Stage 1 and Stage 2 functionality, etc.). The second and third part focus on the main firmware algorithms implemented within the FPGAs on a single Stage 1 and Stage 2 TPG BE board, respectively.

2.4.1 System Structure and Organisation

The TPG boards are grouped together (Figure 2.27) by crates such that four and six crates are responsible for one endcap for Stage 1 and Stage 2, respectively. The crates are

held together by two racks in Stage 1 and three racks in Stage 2. The boards used in both stages of the TPG BE are Serenity boards (each housing one VU13P Xilinx FPGA). For the interface to the cDAQ and TCDS2, the DTH infrastructure provided by the DAQ BE will be used as a common platform.

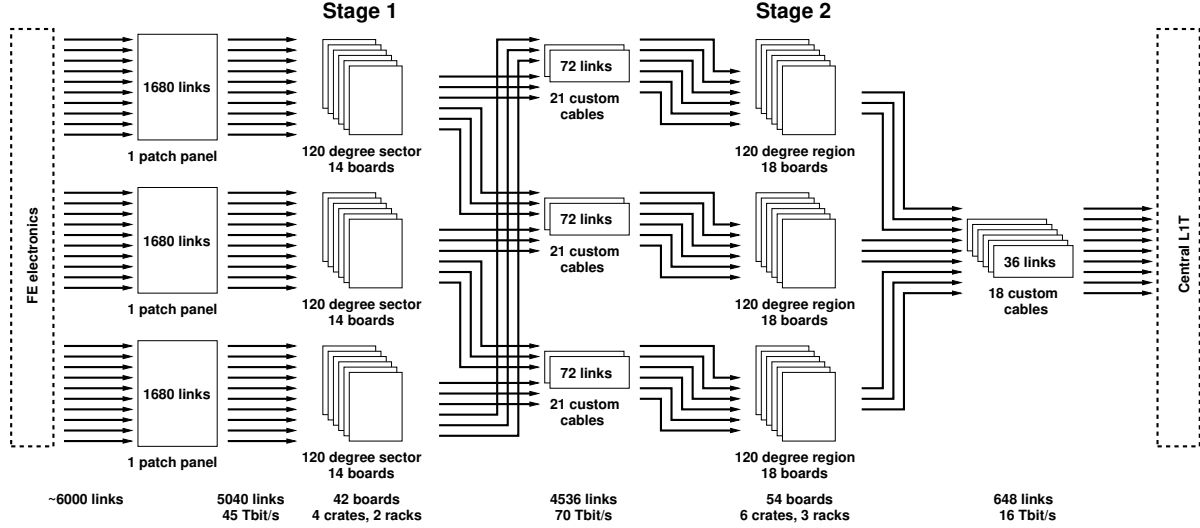


Figure 2.27: Structure of the TPG BE system for one endcap, corrected from [34].

Organisation into Sectors and Time Multiplexing

There are two main drivers for the choices made regarding the structure and the organisation of the TPG BE. Firstly, aside their organisation into crates or racks, the boards in both stages are organised on the algorithm level into three sectors. Each of these is responsible for processing one third of one endcap (a $120^\circ\phi$ sector). Such an organisation is chosen based on the symmetries in the hardware of the HGCal detector FE. On that note, the two endcaps can be treated identically as well as any of the sectors within an endcap. Thus, there is no need for six variations of firmware design. Secondly, the Stage 2 will be working with a time multiplex of 18 BX. In other words, it will process any set of consecutive 18 BXs in parallel, processing each of them only on its designated board. Such an organisation was chosen based on link constraints into Stage 2, latency constraints, matching time-multiplexing periods with the L1T, etc.

The FE $\rightarrow$ Stage 1 Interface

The estimated number of TPG lpGBT links (i.e. $10.24 \text{ Gbit s}^{-1}$) coming from the detector FE is 1500 per $120^\circ\phi$ sector. Each Serenity TPG BE board has a capacity to receive 10 pairs of 12 channel bi-directional modules. Hence, a single Stage 1 board can receive up to 120 FE TPG lpGBT up-links. For added flexibility, each Stage 1 sector will have 14 boards, being able to interface with up to $14 \times 120 = 1680$ FE TPG lpGBTs. With each TPG lpGBT up-link sending a user rate (i.e. the rate coming from the 224-bit D field in the lpGBT up-link frame as shown in Figure 2.22) of $40 \text{ MHz} \times 224 \text{ bit} = 8.96 \text{ Gbit s}^{-1}$ that allows to approximate the **total user data rate coming from the detector FE to the TPG BE** to 45 Tbit s^{-1} , per endcap.

Stage 1 and Stage 2 Responsibilities

The Stage 1 receives ECON-T data from the detector FE, finalises any TC energy calibration (if needed), geometrically sorts (needed for Stage 2 clustering) the TCs, sums

the module energies into partial tower energies and sends the results (selected TCs and partial towers) to Stage 2 while implementing a time-multiplex of 18 BX.

The Stage 2 implements a clustering algorithm which associates the TCs originating from the same particle shower into a cluster of TCs, calculates various physical properties about the clusters (e.g. energy, position, shape) which help the L1T to discriminate between PU, e/γ or hadronic showers, sums the partial tower energies into complete tower energies and sends the final trigger primitives (clusters and tower energies) to the L1T.

The links from the FE are connected so that every one of the 14 Stage 1 FPGAs responsible for a given sector receives the data from 1/14 worth of FE TPG lpGBT links of that sector. A Stage 1 FPGA performs its share of processing of the TPG data and routes it to a particular Stage 2 FPGA. It is required to temporarily buffer the incoming data and implement a readout control in Stage 1 such that the data corresponding to one particular BX is routed to one particular Stage 2 FPGA responsible for the given BX and the given sector. The TPG data corresponding to 1/14 of a sector and from every BX are processed by a single FPGA within Stage 1, whereas a single FPGA in Stage 2 processes the trigger data corresponding to a whole sector and only one out of every 18 BXs.

Duplication Regions at Sector Boundaries

A peculiar problem in the TPG BE is forming clusters in Stage 2 FPGAs at the boundaries of nominal $120^\circ \phi$ sectors. Thus, each Stage 2 FPGA receives 180° worth of TCs in ϕ (additional 30° on every side of a nominal $120^\circ \phi$ sector). This results in some of the TCs being sent from a single Stage 1 FPGA towards two Stage 2 FPGAs instead of only one. However, each Stage 2 FPGA still outputs only the clusters originating from its nominal $120^\circ \phi$ sector.

The Stage 1 $\rightarrow$ Stage 2 and Stage 2 $\rightarrow$ L1T Interfaces

Due to duplication regions at sector boundaries, the data rate between Stage 1 and Stage 2 is increased by a half with respect to the Stage 1 input (plus some overhead) giving about 70 Tbit s^{-1} , per endcap. Each of the 42 Stage 1 boards outputs 108 and each of the 54 Stage 2 boards receives 84 Gbit s^{-1} links, per endcap. That totals 4536 links in the Stage 1 to Stage 2 interface, per endcap. The protocol used is the CMS Standard Protocol (CSP) [48]. Because the Stage 2 is designed to work with a time multiplex of 18 BX, it has 18 boards per one $120^\circ \phi$ sector (54 boards per endcap). A Stage 2 FPGA processes the TPG data for a particular sector and BX at a time, as received from Stage 1. The 108 output links from a Stage 1 FPGA are divided into 18 groups of six so that each such group transmits trigger data to two out of 18 Stage 2 FPGAs matched by the BX. Hence, each Stage 2 FPGA has 84 input links (six links coming from every one of the 14 Stage 1 FPGAs). There are three groups of four output 25 Gbit s^{-1} CSP links per Stage 2 board. The factor of three comes from the L1T requirement for the same data to be triplicated for its three internal sub-systems. Ignoring the triplicating, this gives $4 \times 18 = 72$ such links per sector. That totals in 216 output links and 5.4 Tbit s^{-1} , per endcap.

Latency Considerations

The $12.5 \mu\text{s}$ latency window of the CMS's trigger targeted for the HL-LHC era needs to be shared between multiple systems in sequence. Of importance to the HGCAL's TPG is that the latency window available for the data being acquired in the FE to the trigger primitives being output by the BE must not be greater than $5 \mu\text{s}$. The current understanding is that

the FE acquisition and trigger data transmission will take approximately $1.5\ \mu\text{s}$. Therefore, this puts a latency constraint on the development of the BE systems: Stage 1 and 2 together must not exceed a latency window of $3.5\ \mu\text{s}$.

2.4.2 Stage 1 Functionality Overview

The Stage 1 (Figure 2.28) unpacks the ECON-T data received from the FE, aggregates all the TPG data fragments of a given BX together, optionally calibrates the energy of TCs, geometrically sorts the TCs (as needed for the clustering algorithm in Stage 2), selects the highest energy TCs to fit the available output bandwidth, sums the module energies into partial tower energies, packs the resulting TCs and partial towers and sends them to Stage 2 while implementing a time-multiplex of 18 BX. The coloured elements are not specific to the core TPG Stage 1 algorithms, hence they will not be discussed.

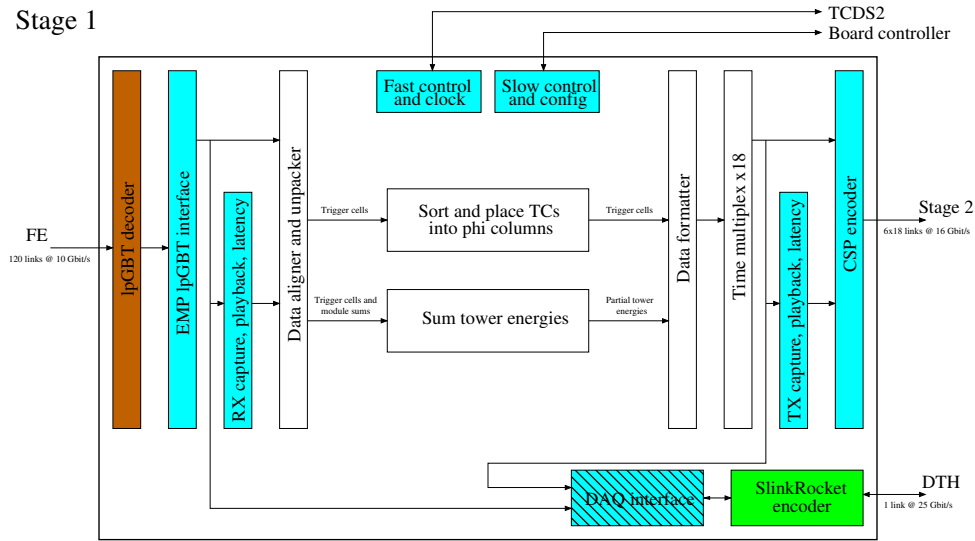


Figure 2.28: A block diagram of the main algorithms on a TPG BE Stage 1 FPGA, corrected from [34].

Data Unpacker

The Data Unpacker is designed to interface with 120 TPG lpGBT up-links. It needs to absorb a lot of variety coming from the FE so that it is not visible by the downstream algorithms. Some ECON-Ts have more output e-Links than others and some ECON-Ts require more than one lpGBT to sample their e-Links. The ECON-T packet headers can appear in any of the lpGBT input e-Link indices. The data format in the ECON-T output varies depending on the ECON-T position in the detector FE. For example, ECON-Ts will operate with at least two algorithm choices (BC, STC). Moreover, the ECON-T BC algorithm yet again has variety in the output format (e.g. low and high occupancy). Some variety in the output format may come from using different choices of the STC algorithm as well (STC4 4E3M will not be used everywhere). The Data Unpacker block is structured to process lpGBT link pairs to handle ECON-T data being sampled by two TPG lpGBTs, resulting in 60 parallel copies of the block. The Data Unpacker also performs checks on the received data (e.g. BX counter) and can automatically re-synchronise with the ECON-T by finding a special header which marks the beginning of an orbit. Because of all these specifications, the data unpacker is the most resource intensive block in a Stage 1 FPGA.

The Data Unpacker outputs the unpacked TCs and the module sums for every BX across two parallel paths to the downstream algorithms.

Trigger Cell Processing

Because there are several ECON-T algorithms used in the FE, the TCs can be represented as either 7-bit or 9-bit binary words. Hence, it should be noted that right before the Trigger Cell Processor, the TCs are normalised to a 9-bit representation. Inside this block, the TCs are sorted in ϕ in this step. This is to ensure the least amount of variety with respect to this coordinate when they are unpacked in Stage 2. In turn, that relaxes the complexity of the Stage 2 algorithm which has to route the TCs into their designated ϕ streams. The trigger cell processing step also associates a unique (per output link) 6-bit label to every TC from which Stage 2 can look-up the physical position of the TC. This results in having 15 bits per TC.

Formation of Partial Tower Energies

After being unpacked, the STCs (for parts of the detector where ECON-Ts are in STC mode) and unselected module sums (for parts of the detector where ECON-Ts are in BC mode) for each BX are streamed into the Sum Tower Energies block, where two partially complete 2D maps in (η, ϕ) of 20×24 tower energies are produced: one for the CE-E and the other for the CE-H part. Every one out of 14 Stage 1 boards responsible for a given sector generally is designed to process data coming from any position within that sector. However, the processing load of the sector is balanced between different Stage 1 boards (e.g. each board processes a subset of modules from that sector). Hence, the produced tower energies in a given Stage 1 board are referred to as partial and require merging with others, which is done in Stage 2. The 2D (η, ϕ) grid is such that single (η, ϕ) bins do not map well onto the Hexaboards (which is not an issue with scintillating tiles). For example, the inner edge of the detector has modules that overlap with about 20 towers, whereas the outer edge of the detector has modules that overlap with about 4 towers. The firmware will probably (i.e. this is to be decided) be designed so it can be configured to split any given module sum into 8 or 16 smaller sums, which are then associated to their designated (η, ϕ) bin. There are $20 \times 24 = 480$ partial towers (CE-E or CE-H) per sector, each one in an 8-bit floating point representation (i.e. 4E4M). Because every partial tower energy is sent further downstream on every BX, there is no need to add more bits into their representation which would convey information about their position.

Data Formatting and Time Multiplexing

The resulting TC and partial tower energy (CE-E and CE-H) streams are combined, formatted and time-multiplexed such that every Stage 2 board receives its BX worth of TCs and partial tower energies. A $120^\circ \phi$ worth of TCs are sent to the designated Stage 2 board of the same sector, but a $60^\circ \phi$ worth of TCs is also sent to the pair of designated Stage 2 boards which are nominally in other sectors (i.e. $30^\circ \phi$ each) because of duplication regions. There is no duplication when it comes to tower energies, so each Stage 1 board sends out only the $120^\circ \phi$ worth of tower energies to the designated Stage 2 board. The resulting data is streamed out of a Stage 1 FPGA across 108 CSP 16 Gbit s⁻¹ links, with 6 links being designated for each of the 18 BX. Each link sends 64-bit words with a frequency of 240 MHz which allows six such words over one BX or 108 words over 18 BX to be sent out, per link.

One 64-bit word can pack three 15-bit TCs and two 8-bit partial towers and some 3-bit synchronisation flags. This means that a single Stage 1 FPGA can send out up to 324 TCs and 216 partial towers for a given BX, per link. Each Stage 1 FPGA will send four of its output links to a designated Stage 2 FPGA (which is nominally in the same sector) and two of its output links to a different Stage 2 FPGA (which is nominally in a different sector, for duplicated regions). Thereby, each pair of links approximately corresponds to a $60^\circ \phi$ region (be it nominal or duplicated). Hence the limit on the number of TCs each Stage 2 can receive for a given BX for a $180^\circ \phi$ region is $6 \times 324 = 1944$ TCs. There are two groups (CE-E and CE-H) of 480 (960 in total) partial towers for a given BX for the $120^\circ \phi$ region. That cannot fit the available four-link budget limit of $4 \times 216 = 864$ partial towers. However, with load balancing at the Stage 1 inputs, it is expected that there will be no more than two groups (CE-E and CE-H) of 360 (720 in total) partial towers actually calculated by a single Stage 1 FPGA, which does fit the output link bandwidth.

2.4.3 Stage 2 Functionality Overview

A single Stage 2 FPGA receives TPG data (TCs and partial tower energies) from Stage 1 for a particular BX and ϕ sector of an endcap, forms clusters, calculates their properties and completes the tower energy summation. The information calculated for the given BX and sector is then formatted and sent out to the L1T. This process repeats every 18BX, given the time multiplex period. Some form of filtering in terms of which clusters get sent out is foreseen for the rare scenarios when the output bandwidth turns out to be a limitation. There are two data paths in the algorithm - one of the trigger cells on which the 3D clusters are formed and the other of the partial tower energies on which the total tower energies are formed. The architecture is broken down into a sequence of modules (Figure 2.29). The coloured elements are not specific to the core TPG Stage 2 algorithms, hence they will not be discussed. The processing done in Stage 2 is the most resource intensive step in the HGCal BE, both hardware resource and latency wise. For the latter in particular, recent estimates project that Stage 2 must not exceed about $2\mu\text{s}$ for its share of processing, which is by far the most significant part of the total $2.5\mu\text{s}$ latency budget for the TPG BE.

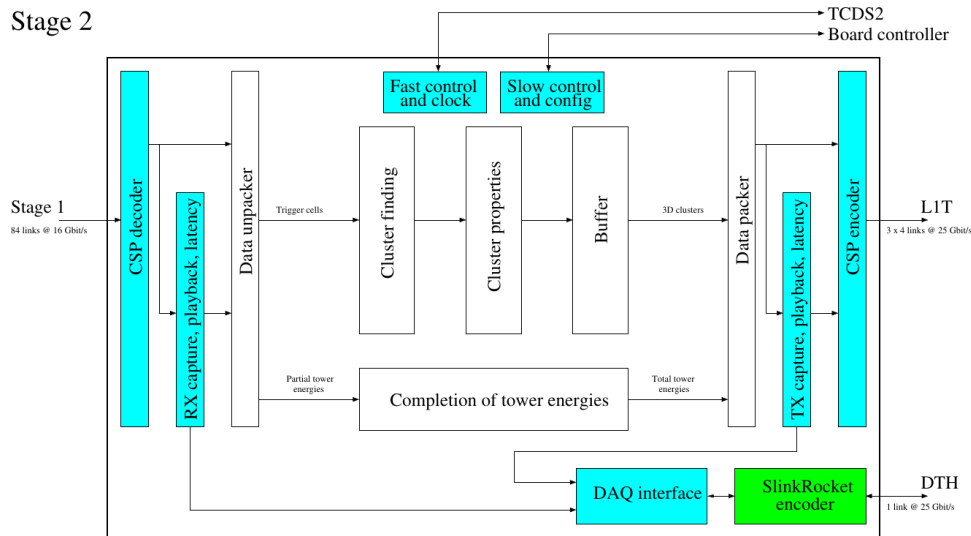


Figure 2.29: A block diagram of the main algorithms on a TPG BE Stage 2 FPGA [34].

Data Unpacker

The data about TCs and partial towers is extracted into two separate streams and sent across one of the two parallel data processing paths in the algorithm. In addition, the 6-bit TC label is used to look-up the exact positional information about a trigger cell (layer ID and $r/z, \phi$ coordinates) from a Look Up Table (LUT) and the energies will be transformed to a linear range. The current estimates are that at least 12 bits and 19 bits should be sufficient for coding each of the coordinates and linear energies, respectively. Except if explicitly specified otherwise, these two numbers are assumed for the rest of this thesis.

Complete Tower Energies

The Complete Tower Energies module sums the partial tower energies into final tower energies. In doing so, it combines the contributions from different Stage 1 boards but also the previously separate CE-E and CE-H towers. In addition, it calculates the ratio of the total CE-H and CE-E contributions (e.g. for e/γ discrimination) for every tower. The result is a set of 480 final towers per $120^\circ \phi$ sector, where each final tower is represented by 16 bits. The Complete Tower Energies module has the same time window as all the algorithms on the cluster path combined to do so, hence it is not a challenging element to implement.

Cluster Finding

Depending on its ϕ coordinate each TC from the $180^\circ \phi$ region gets routed into one of 108 streams, with each stream corresponding to a $180^\circ/108$ slice of the $180^\circ \phi$ range. Similarly, based on its r/z coordinate, each one of the TCs get allocated a row ID (0 to 41 integer) which codes which slice of the r/z range it belongs to. The hard-coded stream index (column ID) and the row ID define a 2D 42 by 108 ($r/z, \phi$) grid. The TCs get streamed into the Histogramming module and depending on their position, their energies get accumulated in the bins of the 2D histogram grid. Hence, the energy histogram above the 2D grid “grows” as time passes and more TCs contribute to its formation.

After the 2D energy histogram is formed, a Gaussian 2D smearing kernel is applied to “smooth” it out. Then, the local energy maxima of the histogram are found. Each such maximum will be referred to as a *seed* in further text. The information about all of the histogram seeds is stored in a memory block. In parallel to the Histogramming step, all the TCs of the BX are also saved in a memory block. Having both the TCs and the seeds stored, the algorithm tries to associate every TC (hit) with a seed. The condition for the hit-seed association is that the TC falls into an ($r/z, \phi$) neighbourhood of a seed, which will be referred to as the *domain of influence* (DOI) of a seed in further text. The DOI takes on the form of an ellipse, defined by ± 3 rows/columns in the ($r/z, \phi$) space. This DOI size can be calibrated, per layer. In the case of a TC falling into multiple DOIs, it gets associated to the closest seed. If a TC does not fall into any DOI it does not get associated. The association to a seed is explicitly done by setting a field in the definition of the TC object which codes the ID of the seed it belongs to. This module outputs the TCs of the given BX that got associated to seeds. This is done across 78 output streams, with each stream corresponding to a subset of the 108 histogram ϕ columns, centred to the middle of the sector. The central 72 outputs stream out the associated TCs from the nominal $120^\circ \phi$ sector. The outer six (three on each end) outputs do so for TCs whose seed happened to be in one of the three (per side) columns closest to the nominal $120^\circ \phi$ sector edges. Unlike at the input to the Histogram, each cell now has a seed ID associated

to it. In other words, the TC is “clustered”. This is how the main clustering algorithm of Stage 2 groups the TCs into clusters; there are as many clusters of TCs as there are seeds and the seed ID can be thought of as a cluster ID instead. There will be up to around 100 clusters per Stage 2 board and $O(100)$ TCs per cluster (about 100 on average).

Cluster Properties

Each cluster of TCs gets accumulated down to a single data packet. Such a data packet will be referred to as a *Cluster Sum* in further text. Cluster Sums come in the form of a structured data type, whose various fields are used as inputs to the algorithm that calculates various cluster properties. For every one Cluster Sum data packet, another structured data packet, whose fields are the final properties about the cluster is output from this module. Each sum or final properties are represented in firmware with about 500-bit and 250-bit objects, respectively. Both of these data packet types are one-to-one representations of a group of TCs associated to a seed. The final properties can be broken down into several categories: *E_T in various longitudinal regions*; *E_T fractions in various longitudinal regions*; *E_T weighted mean positions*; *shape properties*; and *miscellaneous properties*. These quantities are then used by the L1T in order to discriminate between PU, e/γ and hadronic showers when a trigger decision is being made. More details about the Cluster Properties block can be found in Chapter 6.

Filter/Buffer

Some form of filtering in terms of which clusters get sent out is foreseen for the situations when the output bandwidth turns out to be a limitation. More details about the data rates can be found in the description of the Data packer block. For example, some clusters can be prioritised over others based on their transverse energy. Likewise, having all the properties about a cluster available, a more sophisticated selection can be done as well. For example, the available properties at this point will already allow some level of discrimination between the cluster coming from PU, e/γ or hadronic showers. The Cluster properties are also buffered before the data packer so that it can read them out in its own time and at its own working frequency.

Data packer

The role of this module is to combine what is calculated by the Stage 2 algorithm on the clustering and the tower energy completion paths and output it over the three groups of four 25 Gbit s^{-1} CSP links. The L1T requires all Stage 2 output data to be triplicated, hence the three groups. However, the focus here is on a single group of four links. Each link outputs one 64-bit word at 360 MHz. Therefore, nine such words can be sent out per one BX across one link. Hence, a total of $9 \times 18 = 162$ words per 18 BX can be sent, per link. Finally, the output bandwidth can be expressed as 648 output words per 18 BX, across all four output links together. There will be one final tower for every bin in a 20 by 24 (η, ϕ) grid, hence 480 towers in total. Each final tower will be a 16-bit word. Hence, for the towers, 120 output words are needed. That leaves $648 - 120 = 528$ output words available for the clusters. The current size of a single cluster going out of Stage 2 being 256-bit wide, or four output words. Thus, a limitation in the number of clusters that can go out of Stage 2 per BX is $528/4 = 132$ (or 131 if one word is needed to indicate the start of a new BX). Putting a transverse energy (E_T) condition (e.g. in the Filter/Buffer module) on what clusters will go out of Stage 2 at $E_T > 2 \text{ GeV}$, results in an average number of ≈ 200 clusters per one BX for both endcaps ([20], page 45). Having one Stage 2

FPGA responsible for only one out of six 120° ϕ sectors, gives $200/6 \approx 34$ clusters on average. Hence, the 131 cluster limit is well above the average 34 clusters. However, the high end tails in the distribution of possible number of clusters suggest that there may be times when there are more clusters than can fit the output bandwidth ([20], page 45). This is particularly the case for low E_T thresholds and cases when the majority of clusters inside an endcap come from a single 120° ϕ sector.

2.5 A Note on Different Latency Contributions

Consider the estimate of latency contributions of different parts of the system on the path of TPG data acquisition, processing and L1A signal distribution (Figure 2.30) [49]. From the moment of its creation in the FE, it takes $1.5\mu\text{s}$ for the TPG data to arrive to Stage 1. Somewhat more than $3\mu\text{s}$ is the time it takes for the TPG data, processed by Stage 1 to reach the Stage 2. The Stage 2 alone (i.e. measured between its input and output) has to consume less than $2\mu\text{s}$ to meet the constraint of the TPG data being available at the Stage 2 outputs sooner than $5\mu\text{s}$ since its creation in the FE. By the time the L1T processes the final trigger primitives prepared by Stage 2 and the possible L1A fast command gets distributed by the TCDS2 to the interface with the DAQ BE that number gets increased to somewhat less than $11.5\mu\text{s}$. The distribution of the L1A fast command by the DAQ BE and to the FE HGCROC ASIC takes an additional $1\mu\text{s}$, completing the whole loop in $12.5\mu\text{s}$, which is the requirement.

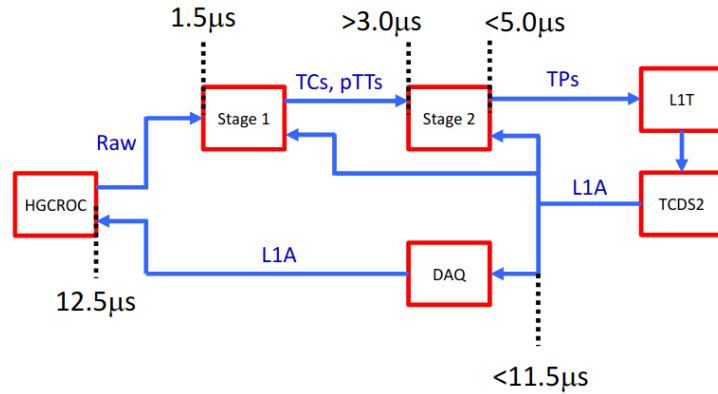


Figure 2.30: Latency requirements for different parts of the system responsible for the generation and distribution of the trigger decision [49]. Acronyms used other than the ones defined in text: pTTs (Partial Towers), TPs (Trigger Primitives).

Chapter 3

Implementing a Vertical Slice of the HGCal Electronics

Chapter 2 demonstrated two aspects that reflect the complexity of the HGCal electronics system. Firstly, one can consider the system from a **vertical** point of view (e.g. focusing on the different layers of data transfer and control within the FE and the BE). In that case, one finds a variety of interfaces and algorithms distributed between the FE (sensor cells, HGCROC, ECON-T, ECON-D, lpGBT, VTRx+ ASICs, each associated with their designated PCB) and the BE (Fireflies, lpGBT firmware, TPG and DAQ BE algorithms). Secondly, when considering the system from a **horizontal** point of view, one can imagine vertical copies of the system implemented in parallel to one another. This act of horizontal scaling, of course, happens in both the FE (e.g. implementing multi-module Trains, single-layer $60^\circ\phi$ cassettes consisting Trains of different shapes and densities, multi-layer $60^\circ\phi$ cassettes, $120^\circ\phi$ sectors or even entire endcaps) and the BE (e.g. implementing BE algorithms into multiple boards, crates, racks, etc.). Obviously, with an increase in the degree of horizontal scaling, there is also a corresponding increase in the number of lpGBT F/O links and in the bandwidth at the FE $\leftrightarrow$ BE interface.

Due to this twofold complexity of the HGCal electronics system, it was decided to put together a vertical test system at CERN which would consist of a single module Train in the FE, an optical FE $\leftrightarrow$ BE interface, and a Serenity BE board. The FE would contain the full variety of real ASICs and PCBs, while the BE would mainly consist of a small number of Stage 1 TPG BE unpacker and DAQ BE capture block instances (providing sufficient complexity to interface with such a FE) and a few other infrastructure blocks.

After having enabled all the major interfaces so that all the different kinds of data exchange (timing distribution and fast control, slow control, TPG and DAQ data acquisition) could be exercised between the FE and the BE parts in the lab (e.g without trying to read signals from real particles hitting the module), the project could consider taking the vertical test system into a beam test and use it to acquire real physics data (more details about this can be found in Chapter 4) or move towards horizontal scaling of the detector.

This chapter starts by presenting the chosen architecture of the vertical test system (Section 3.1). That is followed by a detailed overview of how each of the different major interfaces in the system was enabled: timing distribution and fast control (Section 3.2), slow control (Section 3.3), and TPG and DAQ data acquisition (Section 3.4). Thereby, it presents evidence in support of each of the interfaces being enabled and working, and

quantifies their measured performance. However, the quality of the acquisition of DAQ and TPG data will be examined again and in more detail in Chapter 4. While aiming to cover these three kinds of data exchange in separate sections, some of them are impossible to entirely separate from others. Hence, the reader should not be surprised by a degree of slow control or fast control aspects mentioned in the section designated for the data acquisition, or similar. The knowledge acquired about enabling all the different interfaces naturally resulted in a concrete system initialisation procedure, which is then summarised to bring together the most important points discussed previously (Section 3.5).

3.1 The Architecture of the System

Historically, there were two versions of the FE and three versions of the BE implementation choices in the vertical test system activities throughout this research. The FE choices revolved around two versions of the silicon low-density Trains (Train V2 and Train V3). The BE choice was at first a Xilinx development board (ZCU102) that housed only the lpGBT firmware, slow control and fast control firmware, and TPG and DAQ data capture buffers. For a while, the BE consisted of a hybrid solution where a vertically scaled TPG BE was implemented on the Serenity v1.1 board and the slow control and fast control aspects were orchestrated from the ZCU102. Later on, the slow control and fast control firmware were both migrated into the Serenity board, and a vertically scaled DAQ BE was also added there. For a substantial part of this research, the ECON ASICs were not available and when that was the case, the vertical system used special interposer PCBs and ECON algorithms were emulated in FPGAs in the FE. Different major achievements happened for the first time in different system architectures. However, only a specific version of the vertical system (Figure 3.1) to which the project finally converged is presented here and is used to summarise the major achievements.

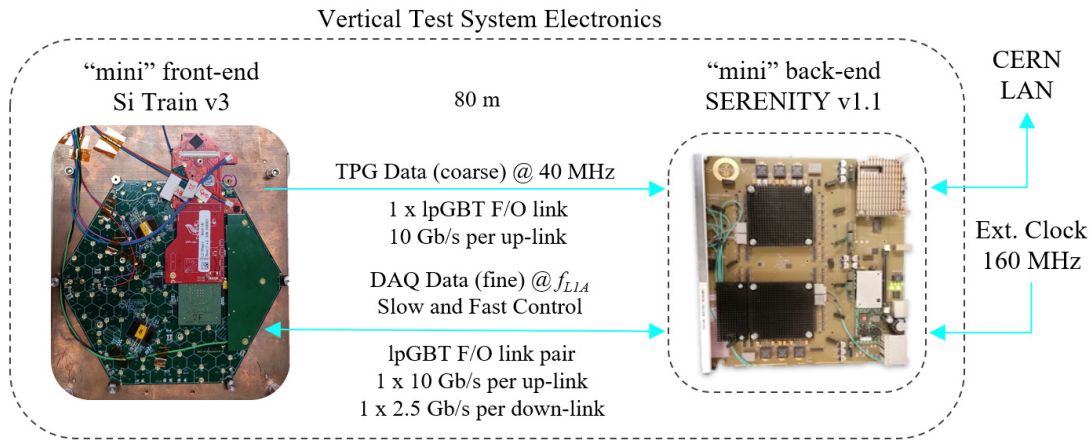


Figure 3.1: A block diagram of the implemented vertical test system.

This system consists of a single-module LD silicon Train V3 in the FE and a Serenity v1.1 BE board in the BE. The FE of the setup is mounted on a copper base plate. The silicon module has $192 \times 1.26 \text{ cm}^2$ sensor cells, above which an LD Hexaboard PCB is mounted. The Hexaboard houses three HGCROC ASICs and the Rafael ASIC. A Concentrator Mezzanine with ECON-T-P1 and ECON-D-P1 ASICs is connected to the Hexaboard. The Hexaboard further connects to an Engine V3 PCB through an LD single-module Wagon

PCB. The engine PCB contains a DAQ lpGBT and a TPG lpGBT ASIC, as well as a VTRx+.

There is a degree of complexity to the power distribution system which will not be explored in great detail. The Hexaboard as well as the engine PCB have interfaces to the low voltage power supplies so that the HGCROCs, ECONs, Rafael as well as the lpGBTs and the VTRx+ ultimately receive their low voltage power requirements. For the VTRx+ the required input voltage is 2.5 V, while that requirement for all other ASICs is 1.2 V. The Hexaboard also has a high voltage interface necessary for the reverse bias of the silicon module. The reverse bias voltage is typically set to -250 V. While this chapter does not explore the use of silicon sensors (e.g. the HGCROCs will be programmed to use fixed ADC values and not sample the signals from the physical sensors) these details are only mentioned for completeness. More details on signals acquired by sampling real sensors can be found in Chapter 4.

The optical interface mainly consists of an 80 m long F/O cable (to emulate realistic latencies between the FE and the BE) and several passive optical components between the FE and the BE (e.g. F/O cassettes), which will be considered transparent in the following text. While the F/O cable has twelve of each of the up-link and down-link fibres, only two up-link fibres running at $10.24 \text{ Gbit s}^{-1}$ (for the TPG and the DAQ sub-systems) and one down-link fibre running at 2.56 Gbit s^{-1} (for the DAQ sub-system) are used.

The Serenity BE board implements a vertically scaled (e.g. to interface with one single-module Train) TPG BE Stage 1 Unpacker and DAQ BE Capture blocks as well as other necessary blocks (e.g. the lpGBT firmware, e-Link Aligner firmware, input and output data capture buffers, TCDS2 emulator, etc.) in its Xilinx KU15P FPGA. Additionally, the Serenity board is also home to optical fireflies, a board controller accessible from the CERN LAN via Ethernet and clock synthesiser modules. The Serenity board requires a 48 V power supply for normal operation. An external clock generator provides a stable 160.32 MHz (for simplicity will be referred to as 160 MHz) clock to the Serenity BE board.

3.2 Fast Control and Timing Distribution

In the final system (more details can be found in Chapter 2), the HGCAL DAQ BE fast control firmware combines fast commands coming from different sources (e.g. the TCDS2 or the HGCAL DAQ BE itself) into the resulting fast command stream which is then brought to the designated parts of the HGCAL FE (e.g. FE ASICs) and the HGCAL BE (e.g. DAQ BE Capture blocks). Specifically, there are the TCDS2 fast control stream, fast commands specified by the HGCAL DAQ BE Board Controller and encoded internally in the HGCAL DAQ BE fast control firmware, the fast commands requested by the HGCAL DAQ BE Autorecovery block, etc. In doing so, the HGCAL DAQ BE fast control firmware needs to be the source of HGCAL-specific fast commands to which the TCDS2 may be agnostic. Furthermore, the HGCAL DAQ BE fast control firmware needs to have the option of generating all possible fast commands on its own (e.g. being independent from TCDS2 when performing internal HGCAL fast control tests). By design, the TCDS2 will be implemented in the DTH400 boards and from it the HGCAL Serenity BE boards will receive the precise LHC clock reference and the TCDS2 fast command stream. In the absence of an implementation of this interface, a hybrid solution was adopted for the vertical test system (Figure 3.2). On one hand, together with the DAQ BE fast control



Figure 3.2: A block diagram of relevant components and interfaces in the vertical HGCal electronics test system from the perspective of timing distribution and fast control. Details can be found in text.

firmware, the TCDS2 is emulated within the FPGA on the Serenity board. On the other hand, a precise 160 MHz clock reference is brought to the Serenity board from an external clock generator. For simplicity, the HGCal-specific DAQ BE fast control firmware and the TCDS2 emulator in the vertical test system can be thought of as a single fast command source and will be collectively referred to as the DAQ BE fast control firmware. This section gives more details about how the fast commands are generated and the required clock sources are synthesised in the BE, what are the main interfaces in the signal flow, and evidence of working fast control and timing distribution in the vertical test system.

3.2.1 Fast Command Generation

The DAQ BE fast control firmware can be accessed from the CERN LAN through the Board Controller (Figure 3.2) and can be configured via slow control to generate a variety of fast commands, scheduled to occur at specific BXs within the orbit. For example, it can be programmed to produce a generally arbitrary fast command or a sequence of fast commands and output them once at a specified BX or a sequence of BXs, respectively. Additionally, it can be programmed to do so periodically (with a specified BX distance) or at pseudo-random times. However, there are certain limitations as to what fast commands can be sent out and in which of the three mentioned scheduling mechanisms (e.g. the L1A fast command is the only one that can be sent out at pseudo-random times, the Bunch Crossing Reset (BCR) fast commands must appear and only appear at BX= 3564, etc.). By design, HGCal foresees using a number of different fast commands (Table 3.1).

3.2.2 Synthesis of the Required BE Clock Sources

A variety of clock frequencies used by different parts of the BE are derived from the external 160 MHz reference and are made available internally on the Serenity board (e.g. 40 MHz, 80 MHz, 160 MHz, 320 MHz, 640 MHz, 1.28 GHz, 2.56 GHz and 10.24 GHz). The Serenity board has a clock synthesiser device that can be programmed through the Board Controller to produce the required clock sources.

3.2.3 The Main Interfaces in the Signal Flow

There is a variety of interfaces in the signal flow in both the FE and the BE of the HGCal electronics vertical test system that are involved in fast control and timing distribution (Figure 3.2).

Locking the FE $\leftrightarrow$ BE lpGBT Interface

Once the external clock reference is stable and the required clock references are synthesised, the lpGBT firmware starts to output lpGBT frames (without specific information modulated into the frames yet) which get serialised and sent towards the FE via the serial F/O 2.56 Gbit s^{-1} down-link. Even though the lpGBT frames do not contain any meaningful content at this point in time, they do contain enough logical "0" $\leftrightarrow$ "1" transitions to allow the lpGBT CDR circuits to recover the clock reference within the lpGBT. In turn, the lpGBT starts sending similar frames via the serial F/O $10.24 \text{ Gbit s}^{-1}$ up-link to the BE which again recovers the clock reference embedded in the up-link. If the original clock

Table 3.1: The HGCAL Fast Command Dictionary [35].

Commands	Binary Code	Description
Idle	0b00110110	Used for LHC clock phase recovery
PreL1A	0b11001100	Used to signal an incoming L1A
L1A	0b01001011	Level-1 Accept
L1A+PreL1A	0b11100001	Combined PreL1A and L1A
L1A+NZS	0b00001111	Level-1 Accept requesting NZS data
L1A+NZS+PreL1A	0b00101011	Combined PreL1A, L1A and NZS request
L1A+BCR	0b01110001	Combined L1A and BCR
L1A+BCR+PreL1A	0b10100101	Combined L1A and BCR and PreL1A
L1A+CalPulseInt	0b00111001	Combined L1A and CalPulseInt
L1A+CalPulseExt	0b10000111	Combined L1A and CalPulseExt
L1A+CalPulseInt+PreL1A	0b11100010	Combined L1A, CalPulseInt and PreL1A
L1A+CalPulseExt+PreL1A	0b11110000	Combined L1A, CalPulseExt and PreL1A
BCR	0b00011101	Resets BX counters
BCR+PreL1A	0b10100011	Combined BCR and preL1A
BCR+OCR	0b10010101	Resets orbit and BX counters
CalPulseInt	0b00101101	HGCROC Internal Calibration Pulse
CalPulseExt	0b01111000	HGCROC External Calibration Pulse
CalPulseInt+PreL1A	0b01010101	Combined CalPulseInt and PreL1A
CalPulseExt+PreL1A	0b10010011	Combined CalPulseExt and PreL1A
ChipSync	0b11010010	Clears buffers, resets counters
EBR	0b11010001	Resets event buffers
ECR	0b10101001	Resets event counters
LinkResetROCT	0b10011001	HGCROC outputs special TPG pattern
LinkResetROCD	0b10011010	HGCROC outputs special DAQ pattern
LinkResetECONT	0b10101010	ECON-T outputs special TPG pattern
LinkResetECOND	0b10111000	ECON-D outputs special DAQ pattern
ECOND-Watchdog-Clear	0b10001110	Clears the ECON-D watchdog timer
ROC-Serialiser-Reset	0b10011100	Resets the HGCROC serialiser

reference provided by the BE and the reference (coming from the FE) recovered in the BE are synchronised, it is considered that the FE $\leftrightarrow$ BE lpGBT interface is *locked*. This step in the FE $\leftrightarrow$ BE synchronisation is commonly taken for granted as it is usually not mentioned in the fast control and timing distribution context, but is a critical first step required for basic system operation.

Timing Distribution

Having locked the FE $\leftrightarrow$ BE lpGBT interface, the BE can perform slow control of the FE lpGBTs and the VTRx+. Relevant for fast control and timing distribution, it is necessary to configure the DAQ lpGBT to output a 320 MHz clock signal at a particular lpGBT e-clock (similar to e-Link but specialised for clock distribution) interface. The 320 MHz clock is then output by the DAQ lpGBT as a differential pair and routed through the Engine PCB, the Wagon PCB, and the Hexaboard PCB where it reaches the Rafael ASIC. The Rafael receives the 320 MHz clock and produces five of its copies that are fanned out to each of the three HGCROC ASICs, the ECON-T ASIC, and the ECON-D ASICs. Additionally, the DAQ lpGBT is configured to output a 40 MHz clock to the TPG lpGBT. At this point, all the FE ASICs have recovered a clock reference to which their internal PLLs have locked. This allows all the FE ASICs and the BE algorithms to be synchronised

to the same clock reference in terms of the working clock frequency (not yet in phase). Moreover, it allows slow control to be exercised with any FE ASIC, opening doors for complete FE configuration and monitoring (more details about the vertical test system slow control can be found in Section 3.3).

Fast Control Distribution

The DAQ BE fast control firmware generates the 8-bit fast commands that go through the Transmitter (TX) Capture buffer of the Extensible Modular data Processor (EMP)¹, where their snapshot can be taken for debugging purposes (Figure 3.3).

Link 013	Link 013	Link 013
Frame 0027 0001 000000001d1d1d1d	Frame 0035 0001 0000000036363636	Frame 0915 0001 000000004b4b4b4b
(a)	(b)	(c)

Figure 3.3: An example of the Serenity BE EMP TX Capture buffer snapshot, showing the hexadecimal format of the 8-bit binary codes of (a) BCR (0b00011101 = 0x1d) followed by (b) IDLE (0b00110110 = 0x36) and (c) L1A (0b01001011 = 0x4b) fast commands intercepted on their way towards the FE. Due to BE implementation details, the relation between the frame numbers and the DAQ BE BX counter BC for a given fast command is $(BC - 1) \times 8 + 35 = \text{Frame}$, with the exception for Frame = 27 when BC = 3564, meaning that the fast commands in the examples (a), (b) and (c) are generated at BC = 3564, BC = 1 and BC = 111, respectively.

Going out of the EMP TX Capture buffer, the 8 bit fast command words enter the lpGBT firmware with a frequency of 40 MHz. This fast command stream occupies the 320 Mbit s^{-1} portion of the 2.56 Gbit s^{-1} bandwidth of the lpGBT down-link by populating an 8-bit portion of the 32-bit D field in the lpGBT down-link frame (more details about the lpGBT protocol can be found in Chapter 2). It is necessary to configure the DAQ lpGBT to extract the fast control stream from the lpGBT down-link frame and output it at a particular e-Link interface configured to operate at 320 Mbit s^{-1} . The fast command stream is then output by the DAQ lpGBT as a differential pair and routed through the Engine PCB, the Wagon PCB and the Hexaboard PCB where it reaches the Rafael ASIC. The Rafael receives the 320 Mbit s^{-1} fast command stream and produces five of its copies that are fanned out to each of the three HGCROC ASICs, the ECON-T ASIC and the ECON-D ASICs. As per the dictionary [35], each of the fast commands provokes a particular action by the components of the vertical system (as will be demonstrated through several examples in Section 3.4). However, the IDLE fast command is particularly interesting as its start marks the beginning of the 25 ns period of the 40 MHz clock. Hence, it allows the FE ASICs to also recover the phase of the clock reference. This step completes the distribution of the BE clock reference to all system components.

3.2.4 Evidence of Fast Control and Timing Distribution

Monitoring the ECON-D Registers

There is yet another handle which allows monitoring the quality of the distribution of the 320 MHz clock and the 320 Mbit s^{-1} fast command stream. Namely, the ECON-D ASIC

¹EMP provides various infrastructure blocks when developing FPGA algorithms for Serenity boards (e.g. in the interface between the core algorithm and the FPGA physical inputs/outputs such as the lpGBT firmware or Receiver (RX) and Transmitter (TX) buffers, etc.).

has several registers implemented for this purpose which can be read via slow control (Table 3.2).

Table 3.2: The ECON-D fast control and timing monitoring registers [38]. The 1-bit registers convey that the description is *true* when their value is 1 and *false* when 0. The 8-bit counter register can monitor the number of received fast commands (or fast command errors) in the range between 0 to 255. All of the registers are sub-nodes of the `FCtrl.Global` ECON-D register.

Register	bit-width	Description
<code>command_rx_inverted</code>	1	current fast command polarity
<code>locked</code>	1	current lock status
<code>lock_count</code>	8	lock condition counter
<code>fc_error_fc_cmd_count</code>	8	fast command error counter
<code>nzs_fc_cmd_count</code>	8	NZS request counter
<code>l1a_fc_cmd_count</code>	8	L1A counter
<code>bcr_fc_cmd_count</code>	8	BCR counter
<code>ocr_fc_cmd_count</code>	8	BCR+OCR counter
<code>cal_pulse_int_fc_cmd_count</code>	8	CalPulseInt counter
<code>cal_pulse_ext_fc_cmd_count</code>	8	CalPulseExt counter
<code>ebr_fc_cmd_count</code>	8	EBR counter
<code>ecr_fc_cmd_count</code>	8	ECR counter
<code>link_reset_roc_t_fc_cmd_count</code>	8	LinkResetROCT counter
<code>link_reset_roc_d_fc_cmd_count</code>	8	LinkResetROCT counter
<code>link_reset_econ_t_fc_cmd_count</code>	8	LinkResetROCT counter
<code>link_reset_econ_d_fc_cmd_count</code>	8	LinkResetROCT counter

Every one of the fast commands whose reception can be counted by the ECON-D (as indicated by Table 3.2) was sent from the DAQ BE and its proper receipt was verified by reading back the values in the corresponding registers via slow control (Figure 3.4).

```
[cmx@serenity-2368-03-i5 swamp]$ python3 econ_control.py -s e0 -a 0x6F -rw read -f configs/econd_fc_status.yaml -p emp -c 13
Elapsed time: 0.0976 seconds
FCtrl:
  Global:
    bcr_fc_cmd_count: 11
    cal_pulse_ext_fc_cmd_count: 1
    cal_pulse_int_fc_cmd_count: 1
    command_rx_inverted: 0
    ebr_fc_cmd_count: 1
    ecr_fc_cmd_count: 1
    fc_error_fc_cmd_count: 0
    l1a_fc_cmd_count: 14
    link_reset_econ_d_fc_cmd_count: 1
    link_reset_econ_t_fc_cmd_count: 1
    link_reset_roc_d_fc_cmd_count: 1
    link_reset_roc_t_fc_cmd_count: 1
    lock_count: 1
    locked: 1
    nzs_fc_cmd_count: 1
    ocr_fc_cmd_count: 1
```

Figure 3.4: A snapshot of the terminal of the Serenity BE Board Controller where the values of the fast control and timing monitoring registers from Table 3.2 were read from the ECON-D ASIC. The DAQ BE fast control firmware was configured to generate each one of the fast commands (that can be counted by the ECON-D) once (with the exception of the BCR and the L1A which were being generated once per orbit) which is reflected in the ECON-D register values.

Moreover, a test was conducted where a variety of fast commands was sent from the DAQ BE to the FE and the `fc_error_fc_cmd_count`, `lock_count` and `locked` registers were

found to have the expected values (0, 1, and 1, respectively) for a duration of $O(10\text{ d})$. This allows to set an upper limit [50] with a 95% confidence level (CL) on the fast command word error rate WER_{FC} when distributing fast commands between the DAQ BE and the FE of the vertical test system (Equation 3.1).

$$\text{WER}_{\text{FC}} < \frac{3.0}{40\text{ MHz} \times 10\text{ d}} \approx 9 \times 10^{-14}, \text{ with a 95\% CL} \quad (3.1)$$

General Notes on Eye Diagrams and their Sampling Phase

An eye diagram (Figure 3.5) is a common measurement that allows the evaluation of the quality of a serial signal, used to transmit binary information (e.g. logical "0" and "1") either electrically or optically. Essentially, it is a 2D histogram drawn above the Amplitude $\times$ Time bin space in which the physical signal appears. Critically, the device receiving the serial signal needs to be configured to evaluate (sample) the signal amplitude value at the “right” moment (phase offset measured from the beginning of the bit period) during the bit period when the eye height is big enough (the eye is “open” enough). In the given example (Figure 3.5), phase choices 0,1,14 and 15 should be avoided as the receiving device would often have issues distinguishing a logical "0" from a logical "1". Of course, the exact specification of the required eye height (also eye width) is specific to the receiving device.

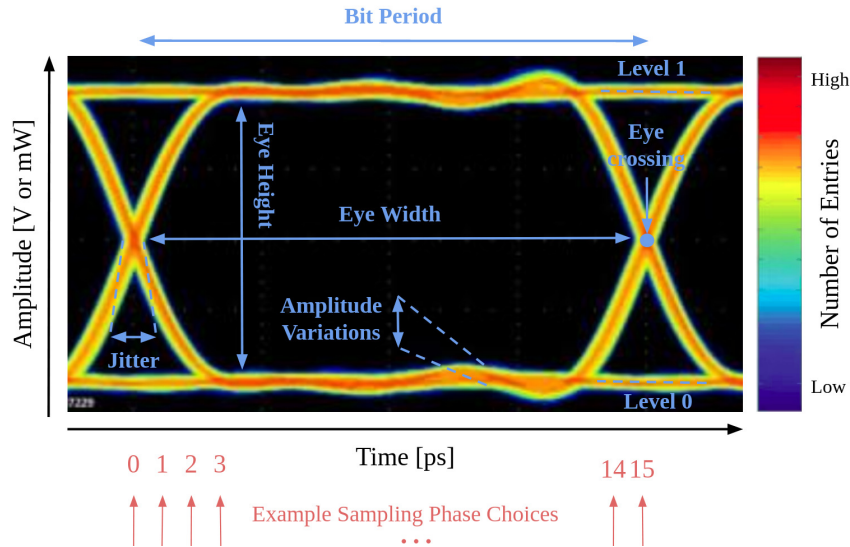


Figure 3.5: A qualitative representation of an eye diagram and its main parameters.

Oscilloscope Measurements

Waveforms of both a 320 Mbit s^{-1} fast command data (Figure 3.6, Top Right) signal and a 320 MHz fast control clock (Figure 3.6, Bottom Right) signal have been measured with an oscilloscope on the Hexaboard PCB at the input of an HGCROC ASIC, demonstrating that the two physical streams are indeed available in the FE. Moreover, they are of the expected electrical quality as can be observed in the eye diagram of the 320 MHz fast control clock (Figure 3.6, Left), but a similar quality applies for the fast command data signal. Specifically, anywhere except in the region around the eye crossings, the eye height is measured to be well above the 140 mV e-link receiver requirement [40]. Because the fast command data stream is accompanied by the fast control clock stream, the latter can be

used to sample the former in the device receiving the two streams. Hence, it is not difficult to ensure that the fast command data stream is sampled at the right phase. Specifically, if the fast control clock edge at which the fast control data stream is being sampled happens to fall in the vicinity of the eye crossing, the user can invert the polarity of the fast control clock signal via slow control. This effectively delays the sampling phase of the data signal by a half-period which is also when the eye height is satisfactory. Otherwise, there is no need for inverting the polarity of the fast control clock.

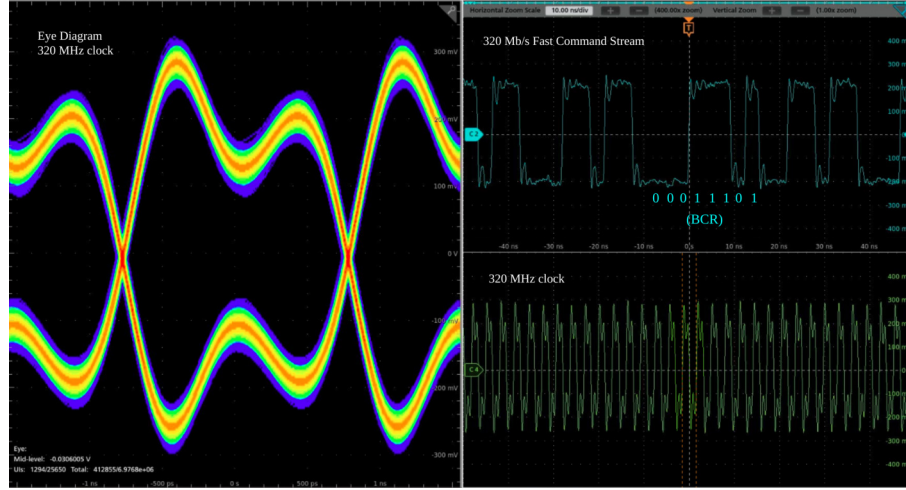


Figure 3.6: Waveforms of the 320 Mbits s^{-1} fast command data stream (**Top Right**), the 320 MHz fast control clock (**Bottom Right**) and the eye diagram of the fast control clock (**Left**) measured with an oscilloscope at the input of an HGCROC ASIC on the Hexaboard PCB.

The phase noise distribution between the BE external clock generator and the clock output by the Rafael ASIC was also measured using an oscilloscope (Figure 3.7). Specifically, the standard deviation of the measured phase noise was $\sigma_{\text{PhaseNoise}} = 8.5 \text{ ps}$ [1][51]. The impact of the internal clock recovery circuits of the HGCROC (e.g. that come after the Rafael output) was measured to be less than 2 ps in addition to the $\sigma_{\text{PhaseNoise}}$, which puts the quality of the clock distribution in the vertical test system well within the 15 ps requirement [51].

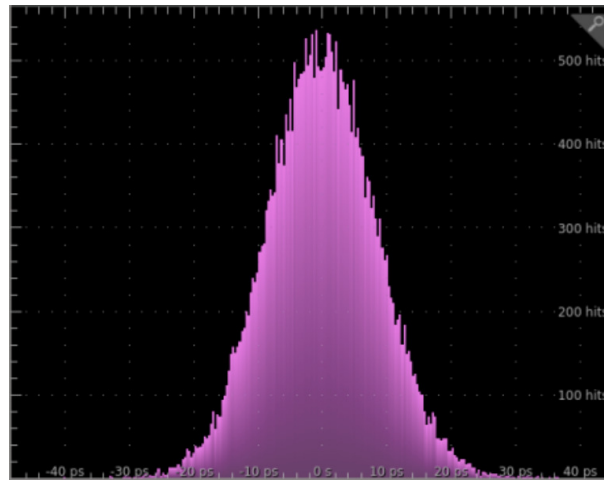


Figure 3.7: Measured phase noise distribution between the BE external clock generator and the output of the Rafael ASIC on the Hexaboard PCB [51] with a standard deviation of $\sigma_{\text{PhaseNoise}} = 8.5 \text{ ps}$. More details can be found in text.

3.3 Slow Control

This section describes the main interfaces in the signal flow, the software architecture and presents evidence of working slow control in the vertical test system.

3.3.1 The Main Interfaces in the Signal Flow

There are a variety of interfaces in the signal flow in both the FE and the BE of the HGCAL electronics vertical test system that take part in slow control (Figure 3.8).

From the CERN LAN, the user can access the Board Controller which implements the software layer for slow control. From there, the user can request write or read transactions with any of the FE ASIC registers (or BE registers). From the Board Controller it is also possible to initiate data captures in the EMP RX and TX Capture buffers, allowing the user to spy on the data going into or out of the implemented BE algorithms. The EMP capture buffers are very useful for debugging purposes but are limited (due to their hardware implementation) to being able to capture no more than 128 BX worth of data in a single snapshot. This in turn limits the data volume that can be acquired in this system architecture when trying to evaluate the quality of system operation.

A down-link slow control transaction message is output by the DAQ BE firmware across the 2-bit IC and the 2-bit EC streams at a frequency of 40 MHz, forming two 80 Mbit s⁻¹ data streams. These are received by the lpGBT firmware and each pair of two IC and EC bits contributes to the formation of the corresponding fields in a down-link lpGBT frame (more details about the lpGBT protocol can be found in Chapter 2). The formed lpGBT frames get serialised and sent through the lpGBT F/O down-link interface, where they occupy a 160 Mbit s⁻¹ portion of the 2.56 Gbit s⁻¹ bandwidth. The lpGBT frames are reconstructed in the DAQ lpGBT which then acts as the main FE hub for further distribution of the slow control transactions. Similar concepts apply in the up-link direction.

The DAQ BE can directly access the DAQ lpGBT via the IC interface. To access the VTRx+, the DAQ BE needs to configure the DAQ lpGBT through the IC interface to act as a master which performs I²C transactions with the VTRx+. The DAQ BE access to the TPG lpGBT is implemented using the EC interface. This happens via the DAQ lpGBT which is transparent for the EC interface. To access the register space of any of the HGCROCs or ECONs, the DAQ BE needs to configure the TPG lpGBT through the EC (which again happens via the DAQ lpGBT) interface to act as a master which performs I²C transactions with HGCROCs and ECONs. Similarly, the TPG lpGBT can be configured to bring a logical "0" or "1" to one of its GPIO pins. This is the interface through which the HGCROCs and ECONs can be power cycled or taken through a (soft or hard) reset.

3.3.2 The Software Architecture

There are several different physical components (e.g. DAQ BE slow control block, lpGBT, ECON, etc.) and different “languages” spoken between them (e.g. IC, EC, I²C) that take part in slow control in the vertical test system (Figure 3.8). It is the responsibility of the software running on the Board Controller to take into account all this complexity and

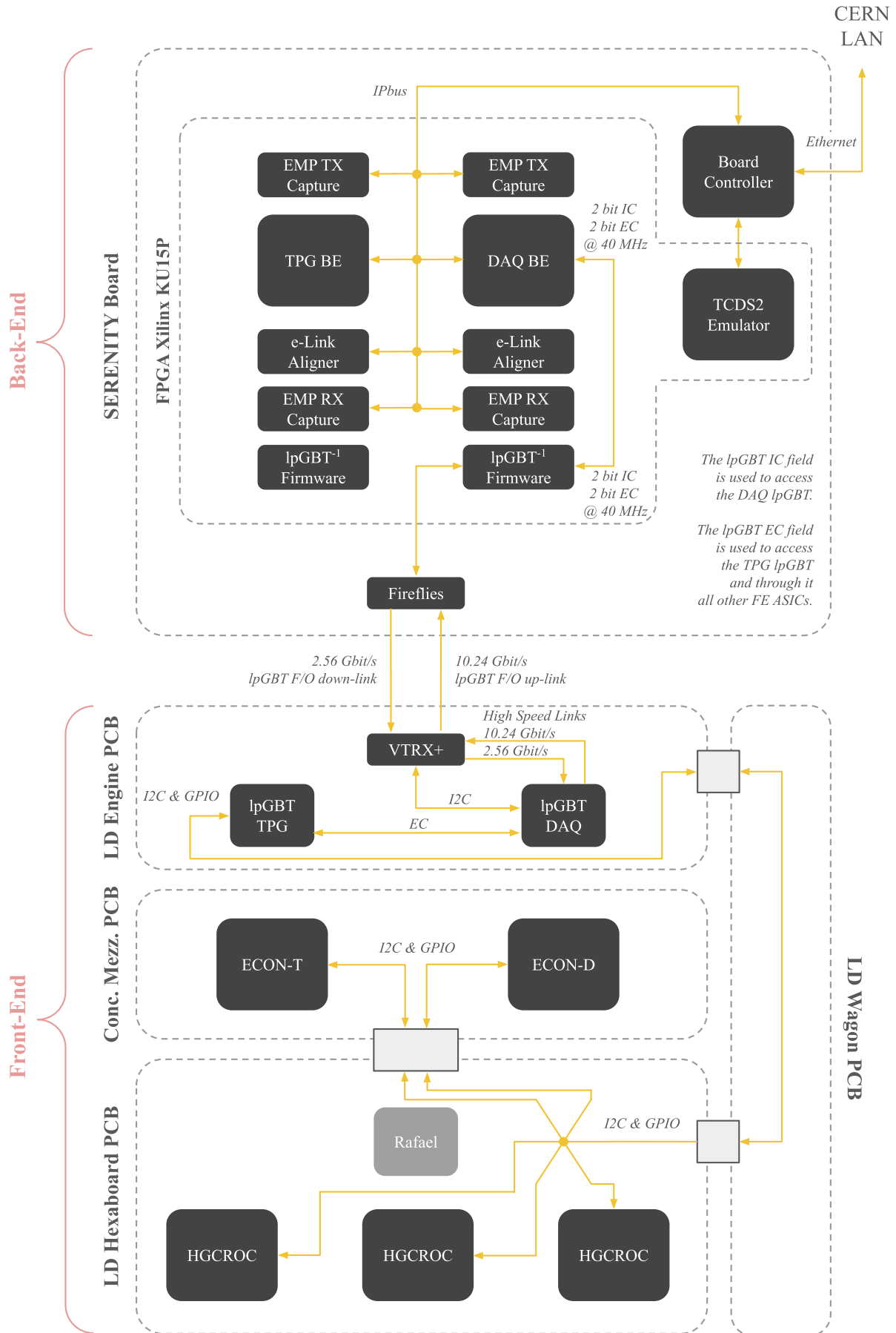


Figure 3.8: A block diagram of relevant components and interfaces in the vertical HGCAL electronics test system from the perspective of slow control. Details can be found in text.

either put together (e.g. in the writing direction) or interpret (e.g. in the reading direction) a transaction message to be sent to or received from the FE, respectively. The transaction message is essentially just a sequence of binary "0" and "1" that contains all the info required for realising a given register read or write access with an FE ASIC target. In the writing direction, the software sends the transaction messages to the DAQ BE slow control firmware which buffers and sends them to the lpGBT firmware where they are used to construct the IC and EC fields of down-link lpGBT frames. In the reading direction, the lpGBT firmware sends the IC and EC fields of the up-link lpGBT frames to the DAQ BE slow control firmware which then assembles the transaction messages and buffers them for collection by the software. Pushing the interface variety handling into software allows all the hardware downstream to be agnostic of it. For that, it is essential to choose an appropriate way of structuring the software layer.

The chosen architecture of the software layer (Figure 3.9) considers implementing a software object (and its interfaces to other software objects) that mirrors the corresponding hardware device (and its interfaces to other hardware devices) [52].

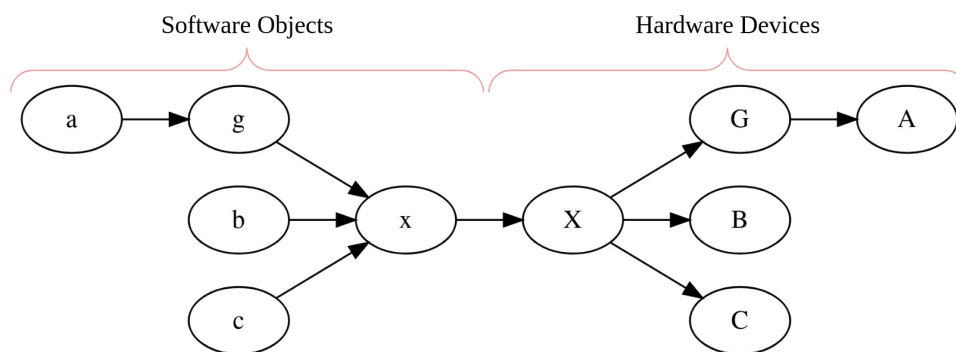


Figure 3.9: A simplified and generalised representation of the data transfer layers in the slow control path. The software architecture is driven by and tries to mirror the hardware architecture [52].

While referring to the Figure 3.9 and Figure 3.8, consider wanting to write a value to the VTRx+ register space as an example. The hardware components that will take part in this transaction are the slow control firmware (X), the DAQ lpGBT (G) and the VTRx+ itself (A). The software VTRx+ object (a) prepares the VTRx+ I²C address, the internal VTRx+ register address and the value to be written. The software DAQ lpGBT object (g) adds information about the DAQ lpGBT address and how and which DAQ lpGBT I²C Master should be configured to perform the physical transaction with the VTRx+. Lastly, the software counterpart of the slow control firmware (x) assembles and formats the final transaction messages. Similar approach applies in the reading direction.

3.3.3 Evidence of Slow Control

Writing to and reading from the FE ASIC Registers

Enabling all the slow control interfaces in the vertical test system has made possible to *write to* and *read from* any register of any of the FE ASICs, while orchestrating these transactions from the BE Board Controller. To illustrate this, consider an example where (for every ASIC) an arbitrary value is written to an arbitrary register and then the same value is read back from the given register (Figure 3.10).

```
[cmx@serenity-2368-03-i5 swamp]$ python3 roc_control.py -s e0 -a 0x08 -rw write -pname DigitalHalf.0.IdleFrame -p emp -c 13 -v 66666666
[cmx@serenity-2368-03-i5 swamp]$ python3 roc_control.py -s e0 -a 0x08 -rw read -pname DigitalHalf.0.IdleFrame -p emp -c 13
DigitalHalf:
0:
IdleFrame: 66666666

[cmx@serenity-2368-03-i5 swamp]$ python3 econ_control.py -s e0 -a 0x6F -rw write -pname FormatterBuffer.Global.link_reset_pattern -p emp -c 13 -v 666
[cmx@serenity-2368-03-i5 swamp]$ python3 econ_control.py -s e0 -a 0x6F -rw read -pname FormatterBuffer.Global.link_reset_pattern -p emp -c 13
FormatterBuffer:
Global:
link_reset_pattern: 666

[cmx@serenity-2368-03-i5 swamp]$ python3 econ_control.py -s e0 -a 0x27 -rw write -pname Aligner.Global.match_pattern -p emp -c 13 -v 66666666
[cmx@serenity-2368-03-i5 swamp]$ python3 econ_control.py -s e0 -a 0x27 -rw read -pname Aligner.Global.match_pattern -p emp -c 13
Aligner:
Global:
match_pattern: 66666666

[cmx@serenity-2368-03-i5 swamp]$ python3 lpGBT_control.py -t daq -rw write -p emp -c 13 -reg DPDATA_PATTERN0 -v 66
[cmx@serenity-2368-03-i5 swamp]$ python3 lpGBT_control.py -t daq -rw read -p emp -c 13 -reg DPDATA_PATTERN0
{'DPDATA_PATTERN0': 66}

[cmx@serenity-2368-03-i5 swamp]$ python3 lpGBT_control.py -t trg_e -rw write -p emp -c 13 -reg DPDATA_PATTERN0 -v 66
[cmx@serenity-2368-03-i5 swamp]$ python3 lpGBT_control.py -t trg_e -rw read -p emp -c 13 -reg DPDATA_PATTERN0
{'DPDATA_PATTERN0': 66}
```

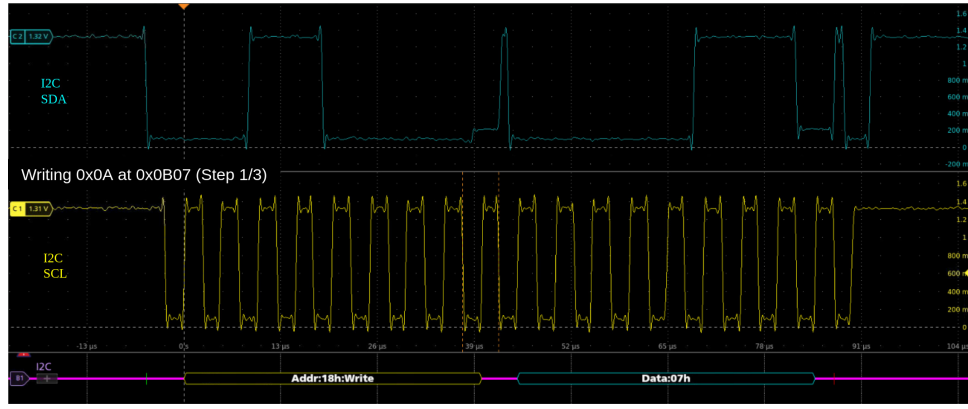
Figure 3.10: A snapshot from the terminal of the BE Board Controller, illustrating *write* and *read* slow control transactions with the FE ASICs [52]. For every ASIC, the value written to the arbitrarily chosen register is also read back, as expected. Transactions with an HGCROC, the ECON-T, the ECON-D, the TPG lpGBT and the DAQ lpGBT are highlighted in red, yellow, green, blue and purple, respectively.

To further evaluate the quality of the slow control interfaces in the vertical system, a test was conducted where $O(1 \times 10^6)$ of write and read slow control transactions was performed with each of the FE ASICs. Specifically, there were three steps in each of iterations of the test. Firstly, a value was written into a chosen register of a given ASIC. Secondly, the same register was read back. Lastly, the value that was originally intended to be written was compared to the value which was read back. If there was a difference between the two, a counter would have been incremented. No errors were observed for any of the slow control transactions in the described test. This allows to set an upper limit [50] with a 95% CL on the slow control transaction error rate TER_{SC} when either reading or writing a register of any of the FE ASIC registers from the DAQ BE in the vertical test system (Equation 3.2).

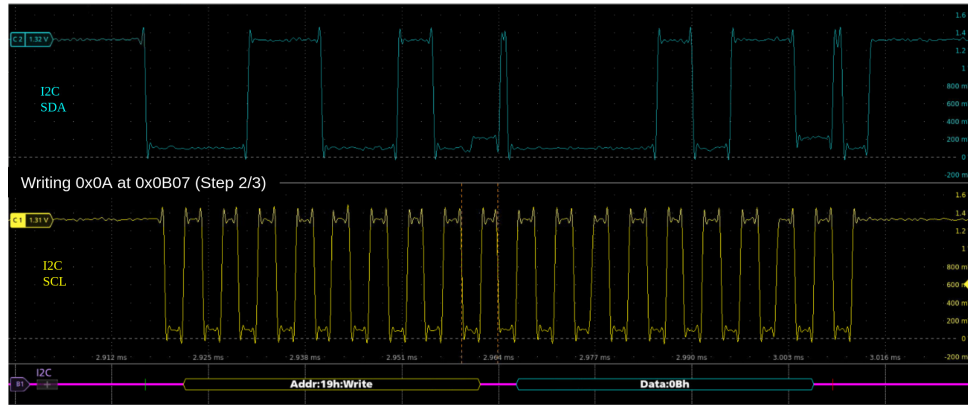
$$TER_{SC} < \frac{3.0}{1 \times 10^6} = 3 \times 10^{-6}, \text{ with a 95\% CL} \quad (3.2)$$

Oscilloscope Measurements

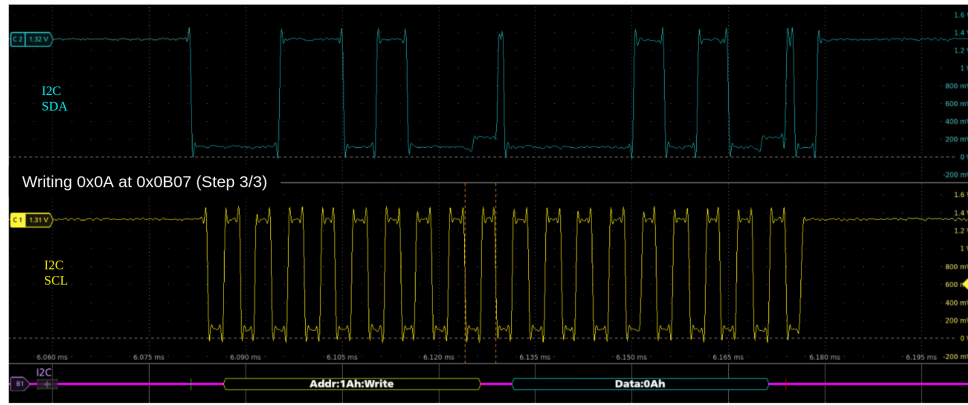
The lowest level of evidence of having enabled the slow control interface in the vertical system is being able to observe the expected shape of the waveforms of the most relevant electrical signals in the FE while orchestrating these transactions from the DAQ BE. Consider an example of using the TPG lpGBT to write a value (e.g. 0x0A) to an HGCROC internal register (e.g. 0x0B07). This, in fact, requires the TPG lpGBT to perform three I²C transactions with the HGCROC (Figure 3.11). Firstly, the TPG lpGBT master writes the lower part of the 16-bit address (i.e. 0x07) into the HGCROC register designated to store it at the address 0x18 (Figure 3.11a). Secondly, the TPG lpGBT master writes the higher part of the 16-bit address (i.e. 0x0B) into the HGCROC register designated to store it at the address 0x19 (Figure 3.11b). Lastly, the TPG lpGBT master writes the value (i.e. 0x0A) into the HGCROC register designated to store it at the address 0x1A (Figure 3.11c). The HGCROC internal I²C logic then makes sure that the value 0x0A gets



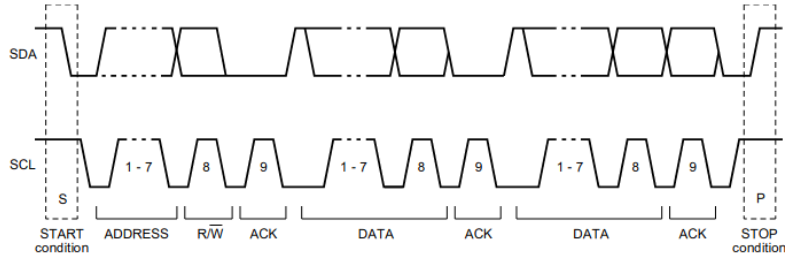
(a)



(b)



(c)



(d)

Figure 3.11: Scoped I²C (a) Step 1 (b) Step 2 and (c) Step 3 transaction waveforms when writing 0x0A with the TPG lpGBT into an internal register of an HGCROC at the address 0x0B07 in the vertical test system and (d) a general time diagram of an I²C bus to help interpret the meaning of individual bits in the waveforms [39]. The operating frequency of the transactions was set to 200 kHz. More details can be found in text.

written into the register with the address 0x0B07, as previously specified.

To understand each of the logical bits in the waveforms, one can refer to the general time diagram (Figure 3.11d). For example, consider 300 mV and 900 mV being the threshold voltages for the logical "0" (V_{IL}) and the logical "1" (V_{IH}), respectively. Also, consider that the rising edge of a logical "1" pulse of the Serial Clock (SCL) line (yellow) being the moment when the logical value on the Serial Data (SDA) line (cyan) is evaluated. Looking at the first step of the transaction (Figure 3.11a), the SDA bus decodes as 0b00110000000001110 binary string. The first seven bits are driven by the master and specify the address 0b0011000 which is equivalent to 0x18 (as expected). The next bit is driven by the master and specifies the direction of the transaction giving "0" which flags that this is a write transaction (as expected). The next bit is driven by the receiver, its value is "0" which means that the receiver acknowledges the first eight bits of the transaction which were specified by the master (as expected). The following eight bits are driven by the master and specify the value (to be written into the previously specified register address) which is 0b000000111, equivalent to 0x07 (as expected). Finally, the last bit is driven by the receiver, its value is "0" which means that the receiver acknowledges the second eight bits of the transaction which were specified by the master (as expected). The I²C START and STOP conditions (Figure 3.11d) are also clearly observable in the considered waveforms (Figure 3.11a).

The chosen slow control transaction example is arguably one of the most complex in the vertical test system as it is indirect (e.g. three transactions needed to achieve a single write transaction to an internal register) and requires the highest number of components in the vertical chain (Figure 3.8) to have beforehand communicated with one another (and over a number of interfaces). Similar checks were done in the reading direction of the data flow and for all I²C (and GPIO) interfaces with all the HGCROCs, the ECON-T and the ECON-D.

The electrical quality of the I²C signals (SDA and SCL) was found to be satisfactory, considering the lpGBT electrical specifications [40]. Specifically, the bus voltage level corresponding to the logical "1" is measured to be 1.32 V which is well above the $V_{IH} = 900$ mV requirement [40]. Similarly, the bus voltage level corresponding to the logical "0" is measured to be 200 mV (in the worst case), which is below the $V_{IL} = 300$ mV requirement [40]. Lastly, the signal rise (10-90%) and fall (90-10%) times are both measured to be less than 5% of the 200 kHz SCL period.

3.4 Data Acquisition of DAQ and TPG Data

This section starts by presenting the main interfaces in the vertical test system on its DAQ and TPG paths. This is done following the signal flow from start (e.g. the FE HGCROCs) to end (e.g. the BE algorithms). Whenever an e-Link is used to transmit data between two ASICs, it is necessary to perform e-Link Phase and Word Alignment procedures. These procedures are next described in detail, for every such interface in the system. This is followed by presenting evidence of successful transmission of both DAQ and TPG data flowing through the whole vertical chain.

3.4.1 The Main Interfaces in the Signal Flow

There is a variety of interfaces in the signal flow in both the FE and the BE of the HGCal electronics vertical test system that take part in the acquisition of DAQ (Figure 3.12) and TPG (Figure 3.13) data.

The DAQ and the TPG data flow out of each of the three HGCROC ASICs across two and four 1.28 Gbit s^{-1} serial differential e-Links towards the ECON-D and the ECON-T ASICs, respectively. The complete LD Hexaboard worth of DAQ and the TPG data enter the ECON-D and the ECON-T ASICs across six and twelve 1.28 Gbit s^{-1} serial differential e-Links, respectively. Assuming the e-Link Phase and Word alignment procedures (more details can be found in Section 3.4.2) were conducted in the interface between the HGCROC and the ECONs, the DAQ and the TPG data are then processed as per the working ECON-D and ECON-T configuration choices, respectively. The DAQ and TPG data flow out of the ECON-D and the ECON-T across two and four 1.28 Gbit s^{-1} serial differential e-Links towards the DAQ lpGBT and the TPG lpGBT ASICs, respectively. Assuming the Phase Alignment procedure (more details can be found in Section 3.4.2) was conducted in the interface between the ECONs and the lpGBTs the phase-aligned ECON data are transmitted to BE. Here, the ECON-D data from the two 1.28 Gbit s^{-1} e-Links takes up 2.56 Gbit s^{-1} of the total $10.24 \text{ Gbit s}^{-1}$ DAQ lpGBT F/O up-link bandwidth. Similarly, the ECON-T data from the four 1.28 Gbit s^{-1} e-Links takes up 5.12 Gbit s^{-1} of the total $10.24 \text{ Gbit s}^{-1}$ TPG lpGBT F/O up-link bandwidth. In both cases, data corresponding to each of the ECON output e-Links occupies a 32-bit portion of the D-field in the corresponding lpGBT up-link frame. The BE lpGBT firmware performs the inverse transformation to the one done in the FE lpGBTs. This makes the phase-aligned ECON data available in the BE. Each of the ECON e-Links is first brought to the EMP RX Capture buffer where it can be checked by the user from the CERN LAN via the Board Controller. Going out of the EMP RX Capture buffer, the ECON data flow into the e-Link Aligner firmware, where the Word Alignment (more details can be found in Section 3.4.2) procedure is done. The fully aligned ECON data then go into their designated BE algorithm. For the DAQ and the TPG case that means two and four 32-bit words at 40 MHz, respectively. The outputs of the DAQ and the TPG BE can then be captured in the EMP TX Capture Buffer, from where they can be read from the CERN LAN via the Board Controller. This allows cross-checking of the received data versus their expectations, given the configuration of the FE ASICs.

3.4.2 e-Link Phase and Word Alignment Procedures

Having in mind the general notes on eye diagrams and the importance of choosing the right phase to evaluate the logical levels of their corresponding serial signals (more details can be found in Section 3.2.4), this part firstly describes *Initial Phase Training* - the method of choice at the heart of the Phase Alignment procedure in the vertical test system. Secondly, both e-Link Phase and Word Alignment procedures are described in detail for both HGCROC and ECON output data.

The Initial Phase Training Method

There are several options (Fixed Phase, Initial Phase Training, Continuous Phase Tracking, Continuous Phase Tracking with Initial Phase) for the e-Link inputs of the ECONs and

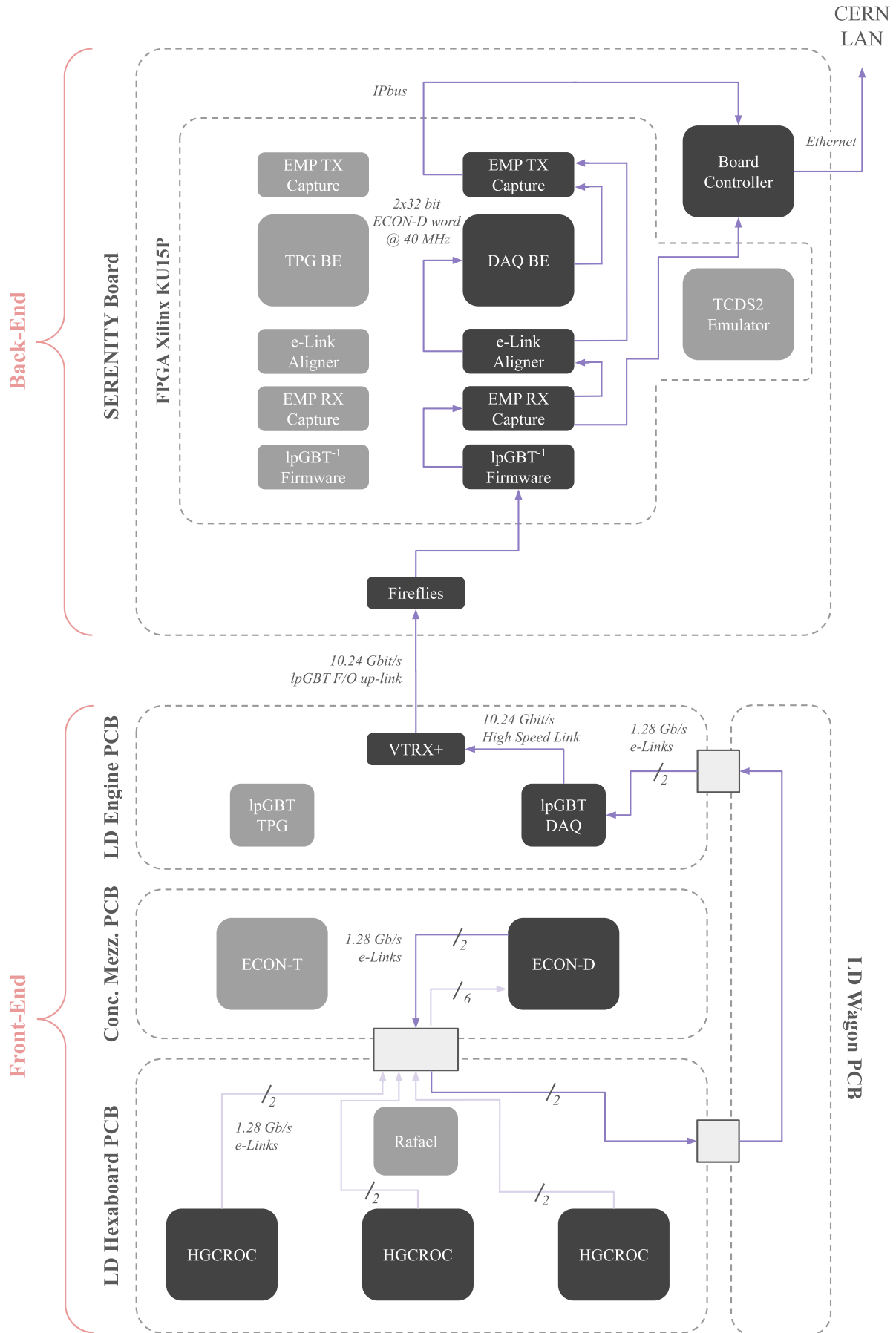


Figure 3.12: A block diagram of relevant components and interfaces in the vertical HGCAL electronics test system from the perspective of DAQ data acquisition. Details can be found in text.

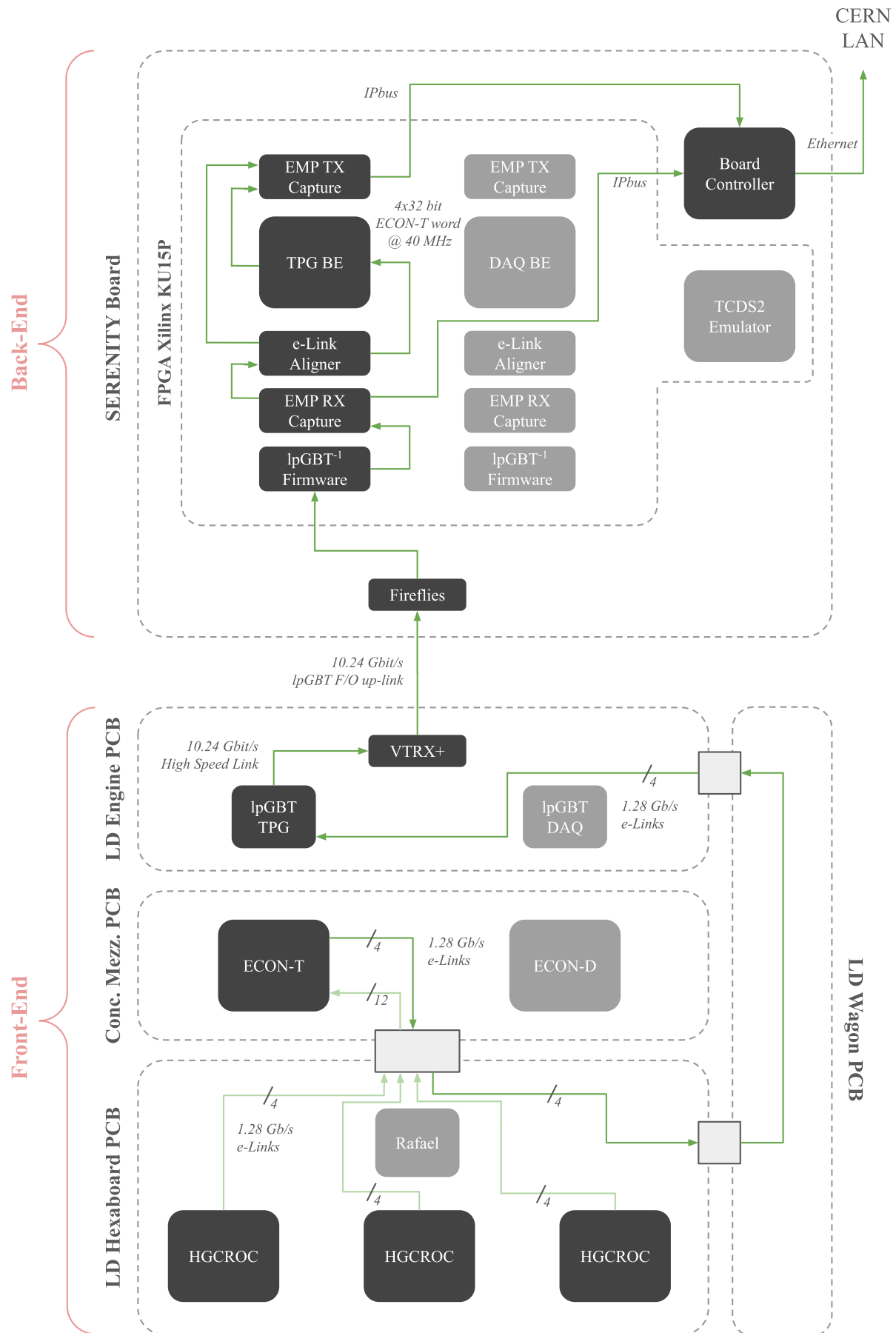


Figure 3.13: A block diagram of relevant components and interfaces in the vertical HGCAL electronics test system from the perspective of TPG data acquisition. Details can be found in text.

the lpGBTs which allow the user to choose optimal phase sampling settings [40]. The chosen technology for the devices used in the vertical test system is Initial Phase Training. This procedure consists of several steps. Firstly, the user needs to make sure that there are enough logical "0" $\leftrightarrow$ "1" transitions on a given e-Link. Secondly, the user initiates the phase training procedure by writing to the appropriate register via slow control. The given e-Link channel phase locking state-machine finds the optimum phase value by scanning through phase increments of $\pm T/8$, where T is the bit period [40]. Once 64 transitions are detected in the expected region, the device updates a register value indicating the corresponding channel is *locked* [40]. The user can monitor the status of this value via slow control. When the *locked* condition is met, the e-Link receiving device can be safely used to sample the relevant e-Links without needing to change the phase sampling settings for the rest of system operation.

The HGCROC $\rightarrow$ ECON e-Link Alignment

Each HGCROC is configured such that its internal registers `DigitalHalf.h.IdleFrame` where $h \in \{0, 1\}$ is the identifier specifying the HGCROC half are configured to store the value `0x555555`. Consequently, when the HGCROC receives the `LinkResetROCT` or the `LinkResetROCD` fast command, it will output a special 32-bit `LRROCIdePattern` pattern for 256 consecutive BXs on all of its TPG or DAQ e-Links, respectively. Generally, the `LRROCIdePattern` depends on the internal HGCROC BX counter (BC) value at the time it is being output (Equation 3.3). For the most part of the orbit it has the same value, except in the special case when it marks the beginning of an orbit.

$$\text{LRROCIdePattern}(\text{BC}) = \begin{cases} 0xA5555555, & \text{if } \text{BC} \in 2 \dots 3564 \\ 0x95555555, & \text{if } \text{BC} = 1 \end{cases} \quad (3.3)$$

Phase Alignment For both ECONs the Initial Phase Training procedure is chosen as the Phase Alignment procedure of its input e-Links. When performing this procedure on the TPG path or the DAQ path, the DAQ BE fast control firmware is configured to send a sequence of `LinkResetROCT` or `LinkResetROCD` fast commands to the FE, respectively. The fast command sequence is sent periodically every orbit and it is chosen to provoke the HGCROCs to keep sending the `LRROCIdePattern` throughout the orbit. This removes any possibility of there not being enough "0" $\leftrightarrow$ "1" transitions on any given e-Link while the user is conducting the Initial Phase Training procedure.

Word Alignment The Phase Alignment procedure ensures that the ECONs will interpret the logical value of the e-Link correctly for every bit of information being sent from the HGCROCs. However, due to e-Link routing length differences in the HGCROC $\rightarrow$ ECON interface, it is not guaranteed that one BX worth of HGCROC data (32-bit per e-Link) from all e-Links will arrive to (and be sampled by) the ECON in the same BX, observed from the ECON's time reference frame. To mitigate this effect, the e-Link Word Alignment procedure needs to be conducted. For every e-Link, the ECONs implement a six BX deep snapshot buffer in the channel aligner block, able to store six consecutive 32-bit words. The DAQ BE fast control firmware sends the appropriate `LinkResetROC` (`LinkResetROCT` or `LinkResetROCD`) fast command. The channel aligner searches for the special `0x9` pattern that marks the start of the orbit in the `LRROCIdePattern` (Equation 3.3) received from the HGCROC. Of course, the ECON also needs to be configured to

expect the right `LRR0CIdlePattern` and that is done by writing the `0x95555555` value to the `Aligner.Global.match_pattern_val` internal ECON register. Having found these special patterns for every e-Link (Figure 3.14), each of the ECON channel aligners then applies an e-Link specific bit offset to each of the e-Link data streams. Specifically, for the

```
[cmx@serenity-2368-03-i5 temp_dir]$ cat rocs_to_econd_snapshot_west.txt
eRx 00: snapshot_dv=1 pattern_match=1 select=0x45 : aaaaaab4aaaaaab4aaaaaab4aaaaaab2aaaaaab4aaaaaab4
eRx 01: snapshot_dv=1 pattern_match=1 select=0x45 : aaaaaab4aaaaaab4aaaaaab4aaaaaab2aaaaaab4aaaaaab4
eRx 02: snapshot_dv=1 pattern_match=1 select=0x44 : 5555555a5555555a5555555a555555595555555a5555555a
eRx 03: snapshot_dv=1 pattern_match=1 select=0x44 : 5555555a5555555a5555555a555555595555555a5555555a
eRx 04: snapshot_dv=1 pattern_match=1 select=0x44 : 5555555a5555555a5555555a555555595555555a5555555a
eRx 05: snapshot_dv=1 pattern_match=1 select=0x44 : 5555555a5555555a5555555a555555595555555a5555555a
```

Figure 3.14: A snapshot from the terminal of the Board Controller reporting the status of the ECON-D (similar applies to ECON-T) input e-Link Word Alignment procedure. More details can be found in text.

bottom four e-Link receivers (i.e. `eRx 03, 04, 05`), the special `0x9` pattern was found in the 6 BX deep snapshot buffer at the bit-position `0x44 = 68` counted from right to left in the printout. However, due to e-Link routing differences, for the upper two e-Link receivers (i.e. `eRx 00, 01`) the special pattern was found at bit-position `0x45 = 69`. This is also why the pattern in the printout (far right column) appears different for these two e-Links when compared to the other four. However, the content is actually the same after applying a 1-bit shift.

The ECON-D is reporting success of the procedure by flagging that `pattern_match = 1` for all six receivers. Thus, the ECON algorithms downstream of the channel aligners have the HGCROC data from all the input e-Links of a given BX available together and at the same time. Getting the special `0x9` pattern in the ECON snapshot buffer is a very delicate task as it requires choosing the right time in the orbit to send the `LinkResetROC` fast command, as well as correctly configuring several registers in the HGCROC and the ECON but will not be discussed here in further detail.

The ECON → BE e-Link Alignment

The e-Link Phase and Word Alignment procedures of the ECON data are performed in the FE lpGBT and the BE e-Link Aligner firmware, respectively.

The internal register `FormatterBuffer.Global.link_reset_pattern` is configured in the ECON to store the 11-bit value `0x1AA`. Consequently, when the ECON-T or the ECON-D receives the `LinkResetECON` or the `LinkResetECOND` fast command, it will output a special 32-bit `LRECONIdlePattern` pattern for 256 consecutive BXs on all of its TPG or DAQ e-Links, respectively. The pattern `LRECONIdlePattern` changes with every BX (Equation 3.4) as part of it conveys the information about the ECON internal 5-bit BX counter BC. However, a 22-bit portion is always the same, as specified by the user.

$$\text{LRECONIdlePattern}(\text{BC}) = \text{BC} \ \& \ 0\text{x}1\text{AA} \ \& \ \text{BC} \ \& \ 0\text{x}1\text{AA} \quad (3.4)$$

e-Link Phase Alignment For both the TPG lpGBT and the DAQ lpGBT the Initial Phase Training procedure is chosen as the Phase Alignment procedure of its input e-Links. When performing this procedure on the TPG path or the DAQ path, the DAQ BE fast control firmware is configured to send a sequence of `LinkResetECON` or `LinkResetECOND` fast commands to the FE, respectively. The fast command sequence is sent periodically

every orbit and it is chosen such to provoke the appropriate ECON to keep sending the `LRECONIdlePattern` throughout the orbit. This removes any possibility of there not being enough "0" $\leftrightarrow$ "1" transitions on any given e-Link while the user is conducting the Initial Phase Training procedure.

e-Link Word Alignment The Phase Alignment procedure ensures that the lpGBTs will interpret the logical value of the e-Link correctly for every bit of information being sent from the ECONs. However, due to e-Link routing length differences in the ECON $\rightarrow$ lpGBT interface, it is not guaranteed that one BX worth of ECON data (32-bit per e-Link) from all e-Links will arrive at (and be sampled by) the lpGBT in the same BX, observed from the lpGBT's time reference frame. To mitigate this effect, the e-Link Word Alignment procedure needs to be conducted in the BE. The DAQ BE fast control firmware sends the appropriate `LinkResetECON` (`LinkResetECON_T` or `LinkResetECON_D`) fast command, periodically (once per orbit). From the Board Controller, the user takes a data snapshot using the EMP RX Capture buffer. Of course, the capture needs to be delayed by the right offset of the 128 BX capture window so that the `LRECONIdlePattern` falls within it. A software application on the Board Controller then processes the data from the EMP RX Capture buffer, determines if and what number of bits need to be offset for every ECON e-Link and programs the BE e-Link Aligner firmware to implement the required transformation. Thus, the BE algorithms downstream of the channel aligners have the ECON data from all the e-Links of a given BX available together and at the same time. The word aligned ECON data is also brought to the EMP TX Capture Buffer so that the user can perform a snapshot to see if the data is indeed aligned (Figure 3.15).

Link	015
Frame 0288	0101 0000000041aa41aa
Frame 0289	0001 0000000041aa41aa
Frame 0290	0001 0000000041aa41aa
Frame 0291	0001 0000000041aa41aa

Figure 3.15: A snapshot from the terminal of the Board Controller showing the contents of the EMP TX Capture buffer for one BX worth of the ECON-T (similar applies to ECON-D) data after successful word alignment in the BE. Each row represents the aligned contents of each of the four used ECON-T output e-Links. With respect to Equation 3.4, the BC field (e.g. $BC = 0b01000 = 8$) and the expected 11-bit constant pattern $0b00110101010 = 0x1aa$ are clearly exposed for every ECON-T output e-Link.

3.4.3 Evidence of Data Flow through the whole Vertical Chain

The Main FE ASIC Configuration Choices

In order to verify whether the TPG (Figure 3.13) and the DAQ (Figure 3.12) data going through the components and their interfaces in the vertical system successfully reaches the Serenity BE, a simplified configuration of the FE ASICs was chosen.

Configuring the HGCROCs Most importantly, the HGCROCs (more details can be found in Chapter 2) were configured not to perform measurements of charge deposits in the silicon sensors. Instead, they were configured to send out fixed ADC values, specified

by the user. Specifically, the ADC values were chosen as a function of the HGCROC identifier r and the HGCROC standard channel identifier c (Equation 3.5).

$$(\forall r \in \{0, 1, 2\} \wedge \forall c \in \{0, 1, 2, \dots, 71\}) \text{ADC}[r, c] = 100 \times r + c \quad (3.5)$$

The $100 \times r$ offset allows to easily distinguish channels belonging to different HGCROCs and the c offset allows to easily distinguish different channels within an HGCROC. Additionally, the HGCROCs were configured to always mask the ToT and ToA measurements.

With respect to Equation 3.5, only the subset of 64 standard channels (out of 72) is used for creating the 48 TC values, while the remaining channels (i.e. for every HGCROC these are $c \in \{8, 17, 18, 27, 44, 53, 54, 63\}$) do not enter the sums. For consistency, the subset of 64 channels will be referred to as $\text{ADC}^*[r, c^*]$ where $c^* \in \{0, 1, 2, \dots, 63\}$. The TC values can be specified as a function of the TC identifier t and the given HGCROC identifier r (Equation 3.6). The TC pedestal subtraction mechanism was not enabled.

$$(\forall r \in \{0, 1, 2\} \wedge \forall t \in \{0, 1, 2, \dots, 15\}) \text{TC}[r, t] = \sum_{n \in \{0, 1, 2, 3\}} \text{ADC}^*[r, t \times 4 + n] \quad (3.6)$$

Configuring the ECON-T The ECON-T (more details can be found in Chapter 2) is configured to use the STC4 4E3M algorithm. With respect to Equation 3.6, each one of the 12 STC values can be consequently specified as a function of the STC identifier s and the given HGCROC identifier r (Equation 3.7). The ECON-T was configured such that the energy calibration factor is equal to one. The ECON-T was also configured to output data across four 1.28 Gbit s^{-1} differential e-links, which gives sufficient bandwidth to output the ECON-T Header, and all 12 highest TC maps and 12 STCs of the LD module.

$$(\forall r \in \{0, 1, 2\} \wedge \forall s \in \{0, 1, 2, 3\}) \text{STC}[r, s] = \sum_{n \in \{0, 1, 2, 3\}} \text{TC}[r, s \times 4 + n] \quad (3.7)$$

Configuring the ECON-D The ECON-D (more details can be found in Chapter 2) was configured to be in Pass-through mode allowing the data from all HGCROC channels to be sent to the DAQ BE (i.e. rather than being zero-suppressed). Sending an isolated L1A to the FE is then expected to result in the ECON-D outputting a packet whose total size (e.g. Header, Sub-packet Headers, ROC Fields, CRC and Mandatory Idle) is 238×32 -bit words. The ECON-D was also configured to output data across two 1.28 Gbit s^{-1} differential e-links, so that gives 119×32 -bit words per output e-Link. Hence, the ECON-D is expected to certainly truncate some number of packets for average L1A rates greater than $40 \text{ MHz} / 119 = 336.1 \text{ kHz}$. The vertical reconstruction threshold was set to three in the ECON-D, meaning that at least three (out of six) HGCROC halves had to agree in terms of bunch, orbit, and event counters for the ECON-D to evaluate the vertical reconstruction procedure as “good”.

The FE DAQ Data brought to and output by the DAQ BE

The DAQ BE fast control firmware was configured with a sequence of fast commands that was being sent periodically on every orbit. The sequence consisted of the following fast commands: OCR+BCR at BE BC = 3564, ECR at BE BC = 50 and L1A at BE BC = 100. Having one L1A per orbit means an L1A frequency of $1/3564 \times 40 \text{ MHz} = 11.22 \text{ kHz}$, which

is well below the 336.1 kHz ECON-D truncation threshold. The EMP TX capture buffer 128 BX time window was offset at the appropriate position within the orbit to capture an occurrence (Figure 3.16) of the corresponding periodic output of the DAQ BE. The choice of the fast command sequence allows for predicting the timestamps of the captured event, specifically: the Orbit Counter (OC), Bunch Counter (BC) and Event Counter (EC).

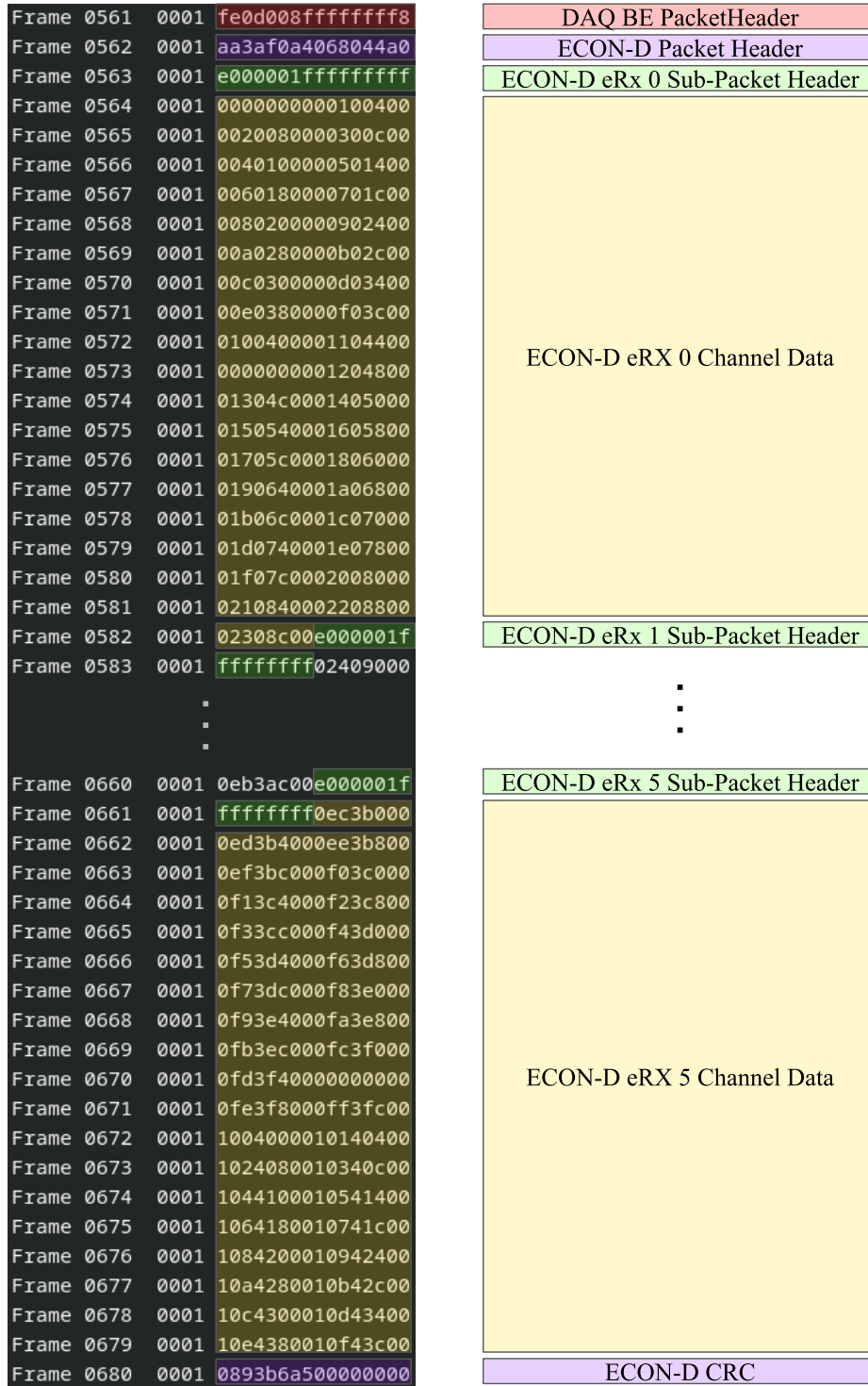


Figure 3.16: One BX worth of data at the output of the DAQ BE captured in the EMP TX Capture buffer with highlighted DAQ BE Header, ECON-D Packet Header, ECON-D Sub-Packet Header, HGCROC Channel Data and the ECON-D CRC. More details can be found in text.

The used HGCROC version (HGCROCv3A) has a fault (which will not be discussed here) and it should be noted that in order to compensate for it, these ASICs were configured to have their internal BX counter consistently four BXs ahead of the BE reference. Hence, the ECON-D and the DAQ BE Capture blocks were also configured accordingly so that the HGCROC, ECON-D, and the DAQ BE Capture blocks agree with each other in terms of what BC an event belongs to. From the perspective of the DAQ BE time reference, the expected L1A timestamps for the event in the working example are $(BC, EC, OC) = (104, 1, 0)$.

The DAQ BE Capture Block can serve up to twelve ECON-D ASICs. However, in the vertical test system it is serving only one ECON-D. Having in mind the DAQ BE output data format (more details can be found in Chapter 2), one can interpret the information encoded in the captured DAQ BE header to be as expected. Specifically, the DAQ BE header `0xfe0d008fffffffff8` reads that the DAQ BE L1A timestamp of the event is $(BC, EC, OC) = (104, 1, 0)$. The ECON-D(0) packet status is `"0b000"`, meaning that the DAQ BE interprets this event as a **normal packet**. For example, the packet was not discarded due to too big of a packet size, an EC mismatch, a BC or an OC mismatch, etc. The ECON-D(i) for $i \in \{1, 2 \dots, 11\}$ packet status decodes as unconnected ECON-D input.

Having in mind the ECON-D output data format for the Pass-through mode (more details can be found in Chapter 2), one can interpret the information encoded in the captured ECON-D header and Sub-Packet headers to be as expected. Specifically, the ECON-D header `0xaa3af0a4068044a0` most importantly reads that: the **Packet Length** of the packet is 235, **E** is 1 meaning that the event Hd and Tr field of the packets were received from the HGCROCs when expected to, the **H/T** and **E/B/O** fields are both `0b00` indicating perfect vertical reconstruction results, the **M** flag is 1 indicating that the bunch, orbit and event counters transmitted by the ECON-D do agree with the HGCROC ones, the **P** flag is 1 indicating that the ECON-D is in Pass-through mode, the **T** flag is 0 indicating that the packet is not truncated, the L1A timestamp of the event is $(BC, EC, OC) = (104, 1, 0)$, the **RR** field is `0b00` indicating there is no reset request by the ECON-D. The ECON-D Header CRC check was conducted in firmware and the ECON-D Packet CRC was conducted in software, both resulted as valid. Each of the six ECON-D Sub-Packet headers is `0xe000001fffffffff` which most importantly reads that the **Stat** field is `0b111` meaning that: there was a match between the ECON-D transmitted H/T, E/B/O to each one of the corresponding values reported by individual HGCROC halves and the valid result of the corresponding half-ROC CRC check. The 37-bit **Channel Map** field has all ones indicating that all standard HGCROC channels of a given HGCROC half are being transmitted.

Having in mind the ECON-D output data format for the Pass-through mode and the HGCROC DAQ output data formats (more details can be found in Chapter 2), one can interpret the information encoded in the captured ECON-D payload (HGCROC output channel data) and to be as expected. Specifically, the **Tc** and **Tp** flags are both `"0"` for every channel, meaning that each HGCROC channel worth of data consists of the ADC(BX-1), ADC(BX) and ToA. The ADC measurements are not corrupted by a ToT event from the previous BX and the ToA readings are always 0, which is in line with the aforementioned HGCROC configuration that has the ToA and ToT results masked. For any HGCROC channel **ADC(BX-1)=ADC(BX)**. Once the **ADC value of every channel** is decoded from the EMP Capture, it can be plotted versus the HGCROC channel ID (Figure 3.17). The values in the plot clearly match the expectation (explicitly given by

Equation 3.7) as per the chosen configuration of the HGCROCs and the ECON-D.

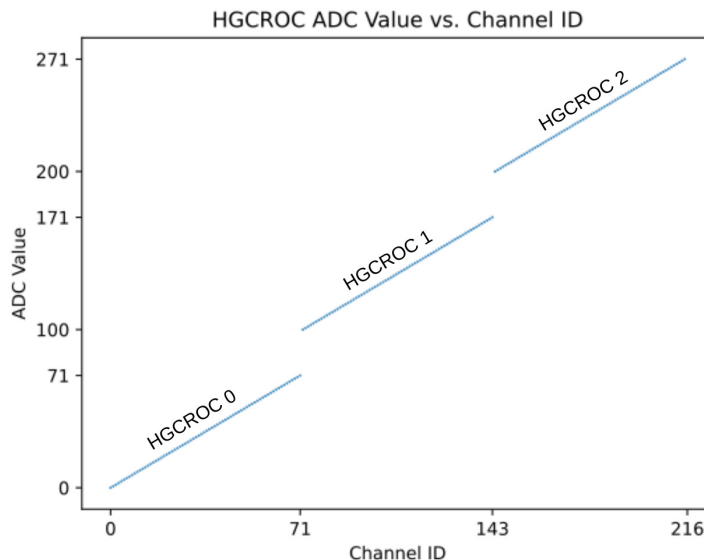


Figure 3.17: One BX worth of data at the output of the DAQ BE that has flowed through the HGCROC → ECON-D → DAQ lpGBT → DAQ BE steps in the vertical test system. The HGCROC ADC vs. Channel ID plot was built from the raw data of a single event captured by the EMP TX Capture Buffer. The ADC vs. channel dependence is as expected given Equation 3.5.

A note on other operated DAQ FE ASIC modes While only a simplified configuration of the FE ASICs was used to illustrate the evidence of successful data flow in the DAQ path of the vertical test system, other ASIC configurations were tried out as well. For example, the HGCROC was also used to sample data arising from the HGCROC receiving CalPulse fast commands, as well as real physics data from the module sensors in a beam test. The ECON-D was also operated in the Zero-suppression mode during the beam test. More details can be found in Chapter 4.

A note on FE → DAQ BE interface tests The DAQ BE currently has the CRC check of the ECON-D header implemented, which allowed for a simple FE → DAQ BE interface quality test to be conducted. No CRC errors were observed while monitoring the system for O(1d) long period. There, the ECON-D was in Pass-through mode and the DAQ BE fast control firmware was generating random L1A fast commands at an average rate of 750 kHz (nominal HL-LHC CMS L1A rate). More details about thorough Bit Error Rate (BER) measurements and the agreement of different components in the vertical DAQ path will be presented in Chapter 4.

The FE TPG Data brought to and output by the TPG BE

The information embedded into the ECON-T output data format (more details can be found in Chapter 2) is mainly the STC energies, with the exception of the ECON-T Header and the twelve 2-bit fields specifying the most energetic TC used to form a given STC. The information embedded into the output of the TPG BE Stage 1 STC Unpacker is quite similar the ECON-T output. Consider an example EMP TX Capture (Figure 3.18) of a BX worth of TPG data of the output of the TPG BE Stage 1 STC Unpacker in the vertical test system.

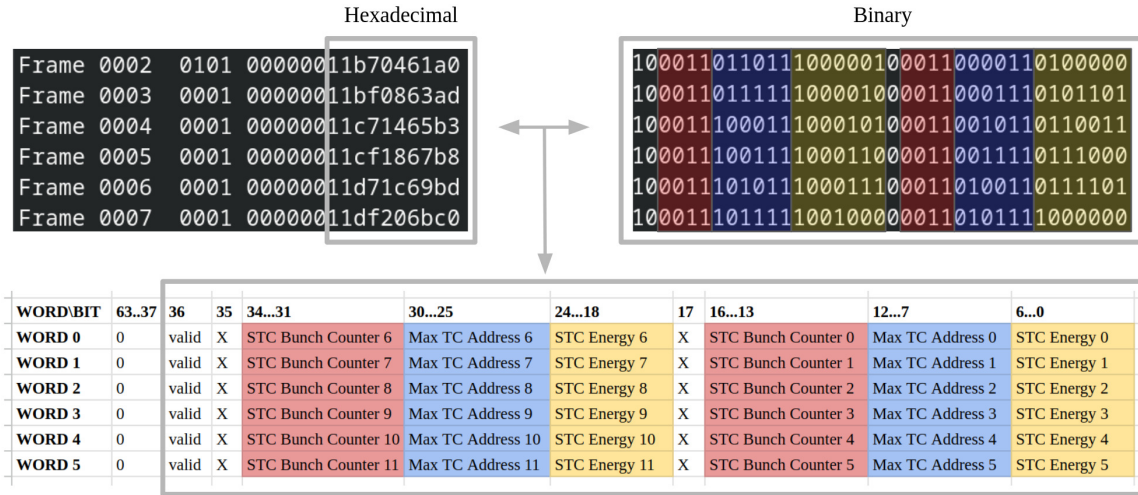


Figure 3.18: One BX worth of data at the output of the TPG BE captured in the EMP TX Capture buffer in hexadecimal format (**Top Left**) and the relevant data in binary format (**Top Right**) and a table for decoding the meaning of individual bits (**Bottom**). Highlighted are the twelve STC Energies, Addresses of Maximum Energy TCs within a given TC and corresponding BX counters. More details can be found in text.

Specifically, there are six 37-bit words which encode the twelve 7-bit STC energies in the 4E3M encoding, 6-bit addresses of the most energetic TCs used to form a given STC and the 4-bit BX counters specifying from what BX a given STC originates from. Two bits are unused (X) and the most significant bit out of 37 is a valid flag. The **BX counters** for all twelve STCs decode as originating from the third BX within the given orbit. This number is, of course, arbitrary (e.g. TPG data is being streamed out on every BX and the EMP capture window was positioned such that the TPG data from the third BX was captured). The twelve **addresses of the most energetic TCs used to form a given STC** decode as 3, 7, 11, 15, 19, 23, 27, 31, 35, 39, 43 and 47, as expected given Equation 3.6 and Equation 3.7. Once the **energy values of STCs** are decoded out of the 4E3M encoding back into the linear range they can be plotted versus the STC ID (Figure 3.19). The values in the plot (Figure 3.19) clearly match the expectation (explicitly given by Equation 3.7) as per the chosen configuration of the HGCROCs and the ECON-T.

A note on other operated TPG FE ASIC modes While only a simplified configuration of the FE ASICs was used to illustrate the evidence of successful data flow in the TPG path of the vertical test system, other ASIC configurations were tried out as well. For example, the HGCROC was also used to sample data arising from the HGCROC receiving CalPulse fast commands, as well as real physics data from the module sensors in a beam test. The ECON-T was also operated using the Threshold sum and the Best choice algorithm. More details can be found in Chapter 4.

A note on FE → TPG BE interface tests The TPG BE Stage 1 Unpacker monitors the received ECON-T header value and compares it versus its expectation. If there are no issues, the status of this interface is considered *locked*. This allowed for a FE → TPG BE interface test to be conducted. The *locked* condition was not lost in a O(1 d) long test. More details about thorough BER measurements and the agreement of different components in the vertical TPG path will be presented in Chapter 4.

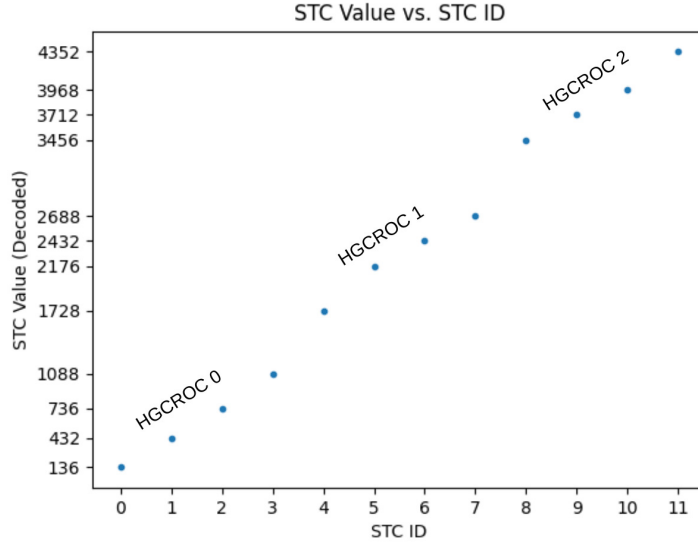


Figure 3.19: One BX worth of data at the output of the TPG BE that has flowed through the HGCROC $\rightarrow$ ECON-T $\rightarrow$ TPG lpGBT $\rightarrow$ TPG BE steps in the vertical test system. The STC Value vs. STC ID plot was built from the raw data of a single BX captured by the EMP TX Capture Buffer. The STC Value vs. STC ID dependence is as expected given Equation 3.7.

Oscilloscope Measurements

The electrical quality of the 1.28 Gbit s^{-1} serial differential e-Links in the vertical system was examined by measuring the corresponding eye diagrams (Figure 3.20). For example, consider the eye diagram of e-Link 0 in the HGCROC $\rightarrow$ ECON-T interface (Figure 3.20a). Clearly, this is a 1.28 Gbit s^{-1} signal, whose eye height is greater than the 140 mV specification of the e-Link receiver [40].

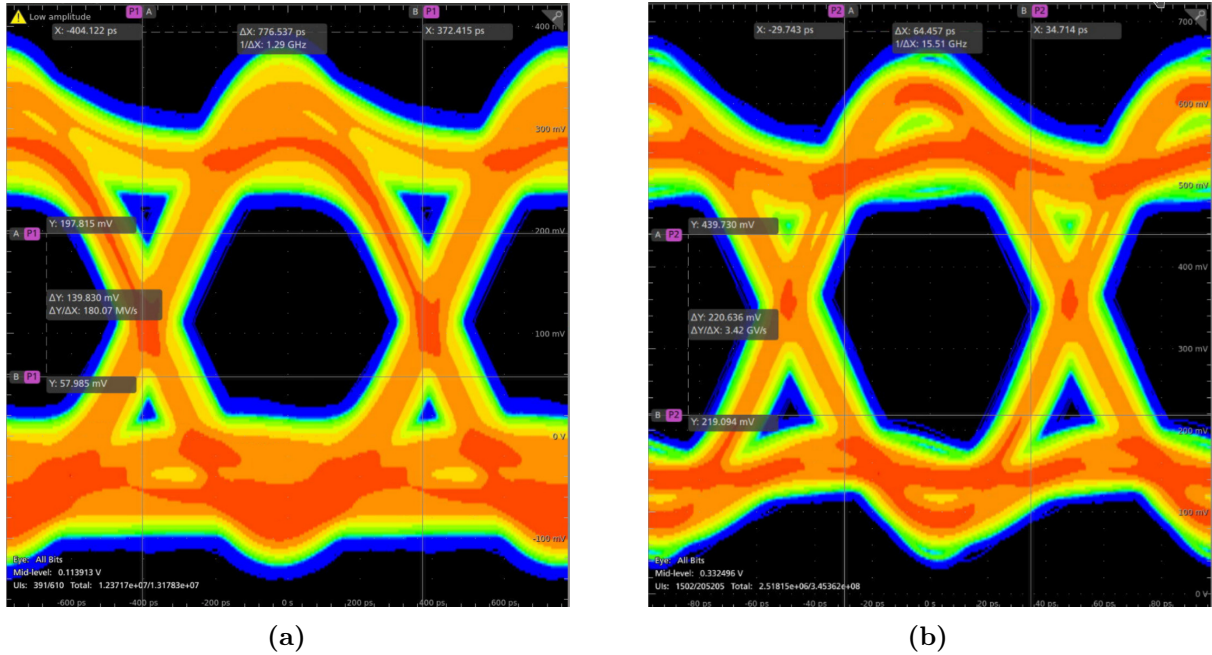


Figure 3.20: Eye diagram of (a) HGCROC $\rightarrow$ ECON-T 1.28 Gbit s^{-1} e-Link and (b) VTRx+ $\rightarrow$ Serenity BE $10.24 \text{ Gbit s}^{-1}$ TPG F/O up-link.

Being able to choose the phase sampling offsets in steps of $\pm T/8$, where T is the bit period

ensures that there is a range of phase offset choices where the eye height is satisfactory. Similar conclusions were drawn for the eye diagram of the VTRx+ $\rightarrow$ Serenity BE 10.24 Gbit s⁻¹ TPG F/O up-link (Figure 3.20b). However, that is not as easy to illustrate because the optical signal had to be converted into an electrical signal to perform the measurement.

Similar measurements were made for other e-Links and lpGBT F/O links in the system. In fact, a critical system issue was discovered in the DAQ lpGBT F/O interface which was causing a non-zero BER in the DAQ lpGBT F/O up-link. The details of how this was identified and mitigated are described in Chapter 4.

3.5 The System Initialisation Procedure

The previous sections describe how the major interfaces in the vertical test system (e.g. fast control and timing, slow control, TPG and DAQ data acquisition) are enabled. To bring all of that together, a procedure for initialising the whole vertical test system (Figure 3.21) was proposed. It starts by turning on the FE and the BE power supplies and ends with having both the DAQ and the TPG available in the Serenity BE. The procedure is condensed into 19 main steps where moving from one step to another is generally forbidden unless a certain condition is met. A version of this procedure was automated and used in the 2023 HGCAL beam tests (more details about these activities can be found in Chapter 4). The automated procedure is executed from the Board Controller and is what was referred to as *the user* in the previous sections of this chapter.

3.6 Results Summary

This chapter detailed the assembly of the vertical HGCAL electronics test system and the activation of its key operational aspects: fast control and timing distribution; slow control; and acquisition of TPG and DAQ data. Each of these system aspects was examined from multiple perspectives.

Regarding fast control and timing distribution, an upper limit was set on the fast command word error rate to 9×10^{-14} (at 95 % CL). Additionally, the phase noise between the BE external clock generator and the Rafael ASIC output was measured to be 8.5 ps, which, together with the 2 ps impact of the internal HGCROC clock recovery circuits puts the quality of the clock distribution in the vertical test system well within the 15 ps requirement.

For slow control, an upper limit of 3×10^{-6} (at 95% CL) was set on the slow control transaction error rate when either reading or writing a register of any of the FE ASICs.

When it comes to data acquisition, the chapter showed examples of data originating from the HGCROCs being streamed through all the real devices and their interfaces in the vertical chain and being available in the Serenity BE (both TPG and DAQ data). The content of the received data in the BE precisely matched the expectations given the overall system configuration. As a simple test of the quality of the interfaces on the DAQ path, the DAQ BE CRC error counters were monitored over a period of O(1 d) and showed no issues while operating the system at an average L1A rate of 750 kHz. Similarly for the

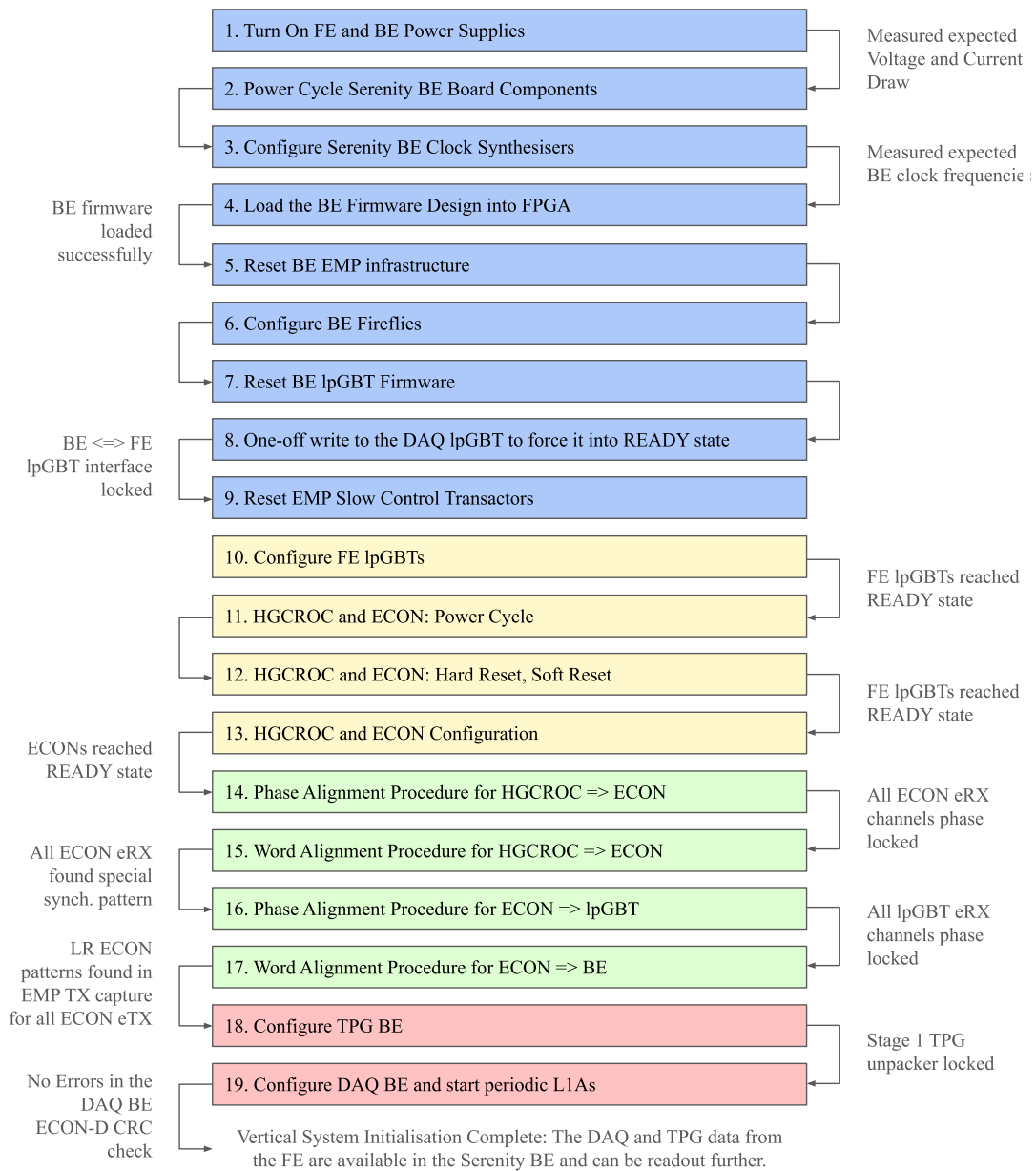


Figure 3.21: A simplified block diagram of the vertical test system initialisation flow.

TPG path, the TPG BE Stage 1 unpacker maintained the *locked* state over a period of $O(1\text{ d})$. More details regarding the quality of the data acquisition will be described in Chapter 4.

The quality of the physical signals involved in each of these system operation aspects was also examined and found to be satisfactory.

Lastly, transmitting data from the HGCROCs to the BE algorithms requires precise tuning of all devices and their interfaces in the vertical path. This process has been incorporated into an automated system initialisation procedure presented at the end of the chapter.

Chapter 4

Gaining a Deeper Understanding of the HGCal Vertical Slice

Having implemented the HGCal vertical test system (more information can be found in Chapter 3) and having enabled the major interfaces (timing and fast control, slow control, and TPG and DAQ data acquisition) within it, the natural next step was to examine its operation in greater detail. There are three important aspects to this.

Firstly, it should be possible to use the vertical test system to acquire true physics data in a realistic environment such as a *beam test*. The HGCal Collaboration had already conducted various beam tests in the past [53]–[59] with prototype precursor systems. This chapter presents the most notable results from the first ever HGCal beam test ¹ where the whole vertical electronics chain of real components were used, including a DTH-based readout [1][2].

Although Chapter 3 does mention successful interface tests between the vertical test system FE and the corresponding BE algorithms, it does not qualify this beyond the ECON-D Header CRC error counter monitoring in the DAQ BE, nor the monitoring of the locked condition of the Stage 1 TPG BE unpacker with the incoming ECON-T data stream. Hence, the second aspect considers explicit *error rate measurements* and setting the corresponding *BER upper limits* for the whole vertical chain for both the TPG and the DAQ path by running the system for an extended period of time $O(10\text{ d})$. During the BER measurements, an issue regarding the saturation of the VTRx+ in the Firefly $\rightarrow$ VTRx+ down-link F/O interface was discovered and mitigated.

The third aspect considers studying the system-wide digital consistency (e.g. consistency of various component status fields, agreement of cross-component L1A timestamps, etc.). It is rather simple to examine this on the TPG path, hence this part is only briefly summarised. However, the DAQ interfaces in the system are significantly more complex, hence a thorough *analysis of the DAQ data* over a collection of $O(10^8)$ events was done. Specifically, in the focus of this analysis are the DAQ header quality flags and the L1A timestamps at various levels in the vertical system (e.g. HGCROC, ECON-D, DAQ BE, etc.).

¹In fact, two beam tests happened during this research (August and September 2023) but are hereby presented as one. The reader is kindly invited to read the blog [60] written to capture some of the most memorable moments during the August 2023 beam test.

To gain a deeper understanding of the vertical test system for any of these aspects, the system had to be upgraded to what was previously presented in Chapter 3. Most importantly, this considers adding an ability to generate trigger signals based on real physics data as well as implementing a readout system able to save large amounts of data to permanent storage.

To put matters into chronological order, it was the beam test that first raised suspicions about the integrity of the interfaces in the vertical test system and motivated the error rate measurements and the analysis of system-wide digital consistency.

This chapter starts by briefly summarising the activities related to the beam test (Section 4.1). These include the vertical system upgrade notes and the most notable beam test results. This is followed by an overview of the key system interfaces (Section 4.2) that take part in the error rate measurements and the system-wide digital consistency analysis. Subsequently, a presentation is given about the BER measurements on the TPG and the DAQ vertical paths and how the VTRx+ saturation issue was discovered and overcome in the process (Section 4.3). Lastly, this chapter presents the results of the system-wide digital consistency analysis (Section 4.4).

4.1 The Beam Test

This section starts by presenting the SPS North Area beam parameters, where the beam test took place. That is followed by a summary of the upgrades made to the vertical test system (previously presented in Chapter 3) in order to comply with these parameters. Lastly, this section presents some of the most notable beam test results.

4.1.1 The SPS North Area Beam Parameters

The HGCAL beam test involving the vertical test system took place at the SPS North Area [61] beam test H4 facility (Figure 4.1a). The primary beam extracted from the SPS (TT20 and latter TDC2) carries 450 GeV protons which are then made to interact with a target (T2) after which the secondary beam is created. The secondary beam is made up of various particles (electrons, pions, muons and photons) which cover a large range of energies (e.g. 10 GeV to 400 GeV) [61]. A selection can be applied to the secondary beam so that only the particles from the desired part of the energy range (e.g. 100 GeV on average) and type (e.g. electrons) reach the experimental setup. Rather pure beams can be obtained (e.g. 95% for electrons). Furthermore, a layer of absorber material can be added in front of the detector sensitive material to provoke a particle shower, thereby creating tertiary particles, etc. The available energy range is sufficient to exercise all kinds of measurements (ADC, ToT and ToA) in the HGCROC ASIC. Additionally, the expected ballpark for the beam transverse width is about 3 cm in diameter in the worst case (e.g. electron showers arising due to interaction with the absorber material), hence it is bound to be well contained within the area of an HGCAL hexagonal module.

The particles arrive to the experimental setup in the so-called particle “spills” (Figure 4.1b). On average, each particle spill lasts about 5 s and reappears periodically with a period of about 35 s, thus introducing about 30 s worth of pause between two spills [61]. Each particle from the secondary beam results in an event that needs to be measured

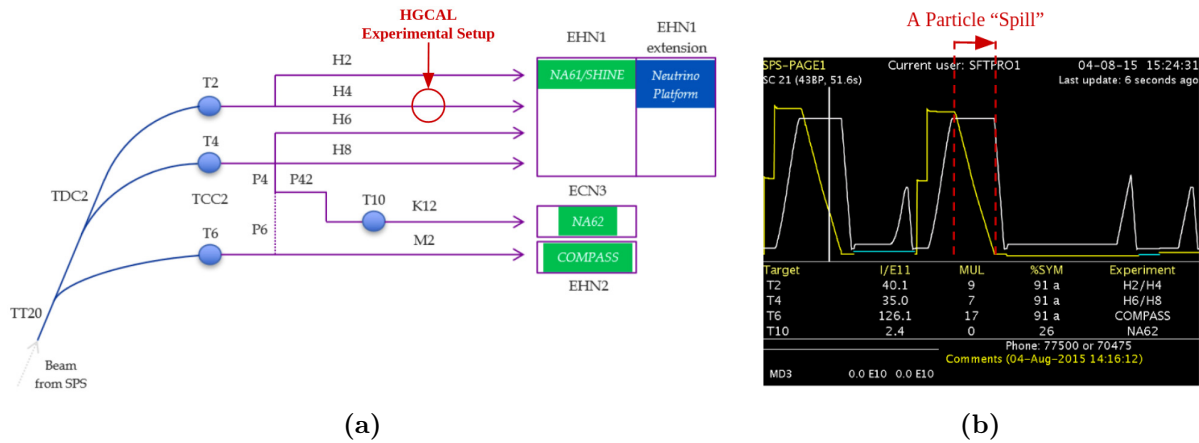


Figure 4.1: The North Area (a) layout and (b) particle “spill” time profile [61]. More details can be found in text.

and read out by the experimental setup. Therefore, there are as many events as there are particles in a given spill. The number of particles per spill generally is not greater than about 500 000. This corresponds to an event rate of about 100 kHz during particle spills, which is well below the nominal 750 kHz L1A rate that the HGCal electronics are designed for. Therefore, using the HGCal readout electronics to acquire data under the aforementioned beam parameters is an adequate choice.

4.1.2 Upgrading the Vertical System for the Beam Test

The vertical test system presented in Chapter 3 had to be upgraded in order to meet the beam test operational requirements at the H4 facility as well as the HGCal project goals (Figure 4.2) in a variety of ways. The most significant ones are regarding the FE, the triggering, the BE firmware, the BE readout and the BE software upgrades.

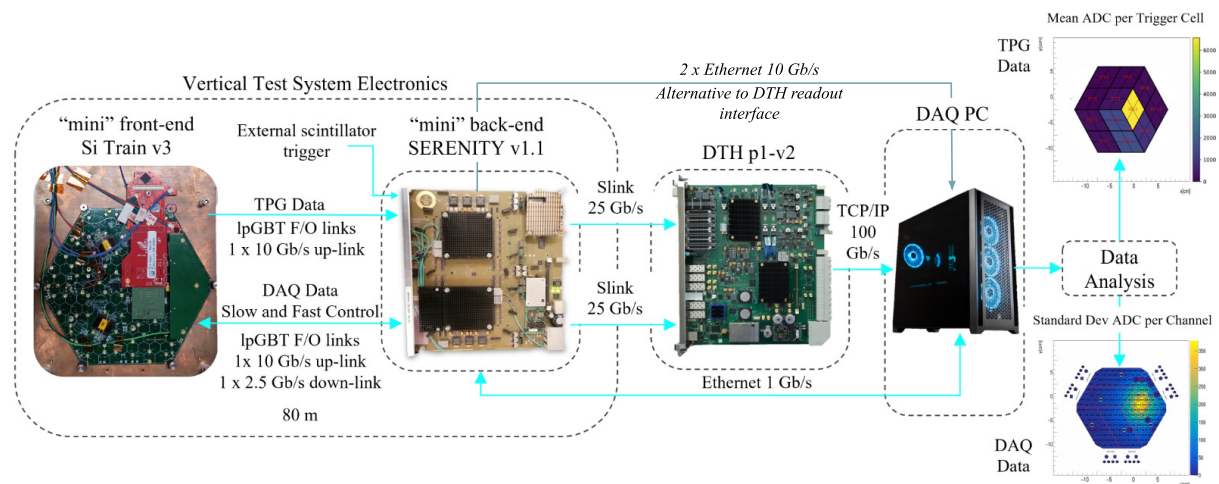


Figure 4.2: A block diagram of the implemented beam test system and its major interfaces [1]. The system consists a vertical slice of the HGCal electronics presented in Chapter 3 and a variety of upgrades (more details can be found in text).

The FE Upgrades A second single module LD silicon Train was added to the vertical system FE so that the data acquired from each can be cross-checked against the other. However, this thesis focuses on a single vertical slice of the HGCAL electronics system. Hence, this system will be generally discussed as such in further text (and it will be explicitly stated if otherwise). The hexagonal module was mounted on a copper base-plate which was water-cooled with a chiller to avoid overheating. The whole LD silicon Train was put in a dark box to reduce the noise arising from ambient light. The box had a dry air source to maintain a constant ambient humidity level.

Trigger Related Upgrades The beam test system implemented two triggering options: the external and the internal (self) trigger. Regarding the external trigger, several scintillator layers were put in front of the HGCAL-specific FE hardware in order to generate an external trigger signal together with a standard Nuclear Instrumentation Module (NIM) crate. An additional lpGBT (not related to the HGCAL-specific FE hardware) was used to sample this signal and bring it through a VTRx+ to the Serenity BE via the usual F/O interface. The external trigger signal had 32 samples for every 25 ns period of the 40 MHz clock reference and its use was twofold. Firstly, it conveyed if there was a particle traversing the FE, thus it was possible to use it as a source for the L1A generation logic in the DAQ BE fast control firmware. Secondly, it was used to reconstruct the phase between the particles traversing through the scintillator and the 40 MHz system clock reference. This timing information was combined with the signal amplitude measurements for every event and allowed to compensate for the asynchronicity between the SPS beam structure and the 40 MHz system clock reference during offline analysis (more details can be found in Section 4.1.3). Regarding the self trigger, the BE implemented a firmware block that spied on the incoming TPG data stream. If the TPG data met a condition, this firmware block generated an internal signal that could be used as a source for the L1A generation logic in the DAQ BE fast control firmware. In the absence of a TPG Stage 2 placeholder in the beam test system, the final trigger primitives and the condition for self triggering were rather simple (e.g. trigger if any of the TCs is above a user programmable threshold). The external trigger also allowed to measure the efficiency and purity of the self trigger [1].

The BE Firmware Upgrades In addition to the firmware responsible for the self trigger and the external trigger, the Serenity BE firmware was upgraded in several other ways. Firstly, a placeholder of the TC Processing block was added to the output of the TC unpacker firmware. Secondly, the data output by the TC Unpacker, TC Processor, DAQ Capture Block, the external trigger timing data, etc. was all aligned together (i.e. latency matched) in special firmware blocks to form a single data object which contains the TPG, DAQ and external trigger timing information for each event. The motivation for this is so that these three kinds of data could be saved downstream (i.e. in the DAQ PC) to permanent storage so that they could be cross-checked in offline analysis. Thirdly, special firmware was added to read out these data while implementing the SLink protocol (more details about the SLink data format can be found in Chapter 2). Also, a prototype of the Extended Readout firmware block was added that allowed reading out a $\pm N$ BX window (before and after an issued L1A) worth of TPG data. Lastly, the HGCAL L1A throttle firmware block (more details about it can be found in Chapter 5) was added to the BE TCDS2 emulator in order to prevent the HGCROCs from causing readout errors (i.e. not

sending event data for some L1As) if the throttling was not applied.

The BE Readout Upgrades There were, in fact, two separate readout data paths from the Serenity BE to the DAQ PC (both implementing the SLink data format and each requiring specialised readout firmware). The main path was across two 25 Gbit s^{-1} optical links, through the DTH p1-v2 board and via the 100 Gbit s^{-1} TCP/IP link to the DAQ PC. The alternative data path was across two 10 Gbit s^{-1} Ethernet links and directly into the DAQ PC. Only one of the two paths could be used at any given time. Each of the two readout infrastructures was successfully used to acquire beam data. In each approach one readout link was used for TPG data and the external trigger timing information while the other was used for the DAQ data. It is out of the scope of this thesis to explicitly break down how the different fragments of the readout data populate the available output bandwidth of the Serenity board in either readout approach. The reader is kindly reminded that Chapter 3 did demonstrate that the data coming into the Serenity BE has a volume of about 5 Gbit s^{-1} and about 2.5 Gbit s^{-1} (in the worst case) for the ECON-T and the ECON-D, respectively. This illustrates that both the 10 Gbit s^{-1} and 25 Gbit s^{-1} links have sufficient bandwidth for that (or a small variation of it if introduced by the BE algorithms) but also for some additional information (as mentioned earlier). Once the data were saved to the DAQ PC permanent storage, they were moved through an Ethernet interface of the DAQ PC to the CERN EOS [62] storage system. From there, the acquired data was checked by the Data Quality Monitoring (DQM) system and other tools (some examples can be found in Section 4.1.3).

The BE Software Upgrades The Serenity BE and the DAQ PC together implemented software to emulate the RCC system. The Serenity Board Controller and the DAQ PC used a standard 1 Gbit s^{-1} Ethernet link for the RCC interface. The RCC was mainly used to initialise the FE, setup the readout parameters of any given run (e.g. duration, trigger type, interface type, etc.), monitor the input data rates on the readout links and save the data to disk. Additionally, the RCC allowed for special non-physics runs where it could specify software-generated fast command sequences (e.g. L1A or CalibPulse for latency-matching the TPG and DAQ data in the Serenity BE or similar). These non-physics runs are particularly important for the error rate measurements and the DAQ data analysis, where data acquisition runs were setup to use random or periodic L1As at specific rates (more details can be found in Section 4.3 and Section 4.4).

4.1.3 The Most Notable Beam Test Results

For the first time, an implementation of a vertically-scaled HGCAL electronics system (Figure 4.2) was successfully operated in the September 2023 beam test [1]. The vertical system consisted of all the critical components (e.g. PCBs, ASICs, optics, BE algorithm placeholders, etc.) and their interfaces on its readout and control paths, demonstrating that such an electronics system architecture can indeed be used to acquire real physics data. Several kinds of particle beams were used (e.g. pion, muon, electron) covering a large range of energies (e.g. from a MIP up to 200 GeV). This allowed to exercise all the major measurements (e.g. ADC, ToT, ToA) as well as calibration procedures (e.g. pedestal subtraction and pedestal trimming) in the HGCROCs. Regarding the TPG path, the ECON-T and the Stage 1 TPG BE Unpacker were used in BC and STC4E3M modes of

operation (which are the current HGCAL baseline). Regarding the DAQ path, the ECON-D was used in the Pass-through and the Zero-suppression modes of operation. Depending on the beam type and system configuration choices, the L1A rate and the useful data volume varied but were not greater than $O(100\text{ kHz})$ and $O(10\text{ Gbit s}^{-1})$, respectively [1]. The system was able to save data reliably from long $O(8\text{ h})$ data acquisition runs, writing as much as $O(100 \times 10^6)$ worth of events per run to the DAQ PC storage and doing so autonomously (i.e. without human intervention) [1]. A total of 2.2 TB worth of data was acquired during the beam test [2]. The inclusion of a DTH readout board into the vertical readout chain for acquiring real physics data also marked the first instance of such an integration among all CMS upgrade sub-detectors [1] [2]. There are various data analyses currently being conducted and only some of their most notable results will be briefly mentioned in this section: *TPG and DAQ Module Histograms*; *TPG and DAQ Signal Amplitude Over Time*; *the Comparison of Internal and External Trigger*; and *ECON-D Channel Retention*. These are presented only as a high-level illustration that the acquired data meets the expectations. In order to interpret some of the results more easily, a note on asynchronicity will also be given.

TPG and DAQ Module Histograms The first line of evidence of data acquired by the beam test system are the module histogram plots (Figure 4.3). The TPG module histogram [63] (Figure 4.3a) shows a typical distribution of the mean STC values for every STC on the hexagonal module. The *red dot* approximates the beam location. The DAQ module histogram [1] (Figure 4.3b) shows a typical distribution of the standard deviation of the ADC values for every channel on the hexagonal module. Both module histograms were generated using the data from electron beam runs and with absorber material in front of the experimental setup. Both plots are generated to qualitatively cross-check the beam position and the beam profile. The two module histograms also clearly demonstrate the spatial granularity difference when comparing TPG to DAQ data.

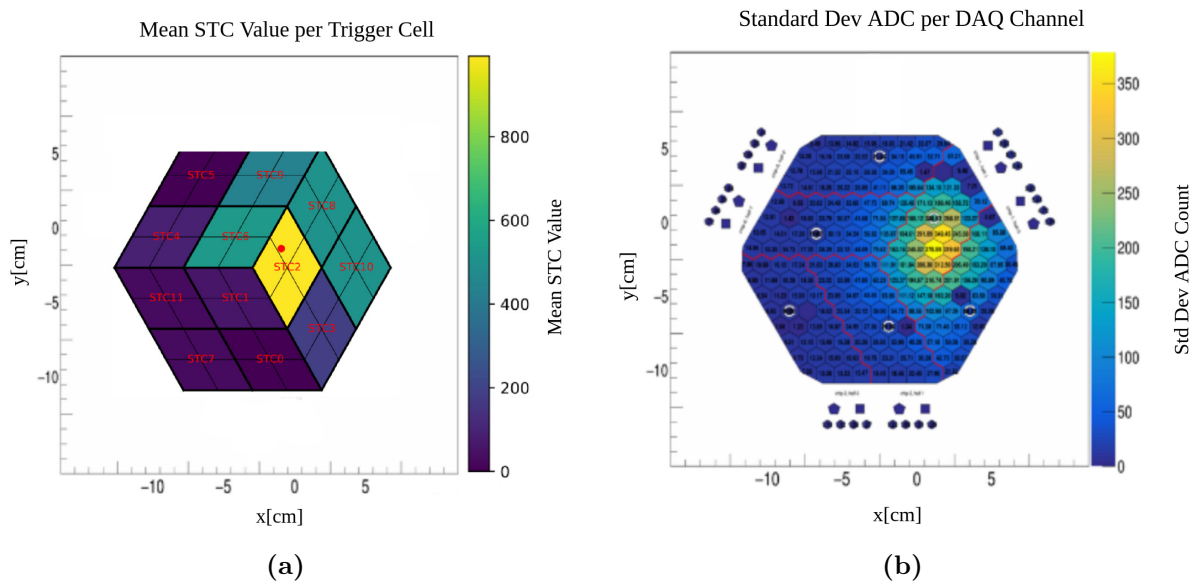


Figure 4.3: Module histograms of (a) the mean STC values using TPG data and (b) the standard deviation of the ADC count per channel using DAQ data. The two plots come from different electron beam runs [1]. Generated using offline analysis tools processing real physics data acquired during the beam test. More details can be found in text.

A Note on Asynchronicity Unlike in the final HGCal system, the particles hitting the FE during the beam test were not doing so synchronously with the BE external clock reference brought to the Serenity BE. Instead, they would appear asynchronously (with a random phase offset) to the system 40 MHz rising clock edge. This is due to the SPS beam not having a 40 MHz structure.

TPG and DAQ Signal Amplitude Over Time A second important check is regarding the shape of the signal amplitude changing as a function of the particle phase with respect to the 40 MHz clock (Figure 4.4). For the TPG case, an electron beam run example where

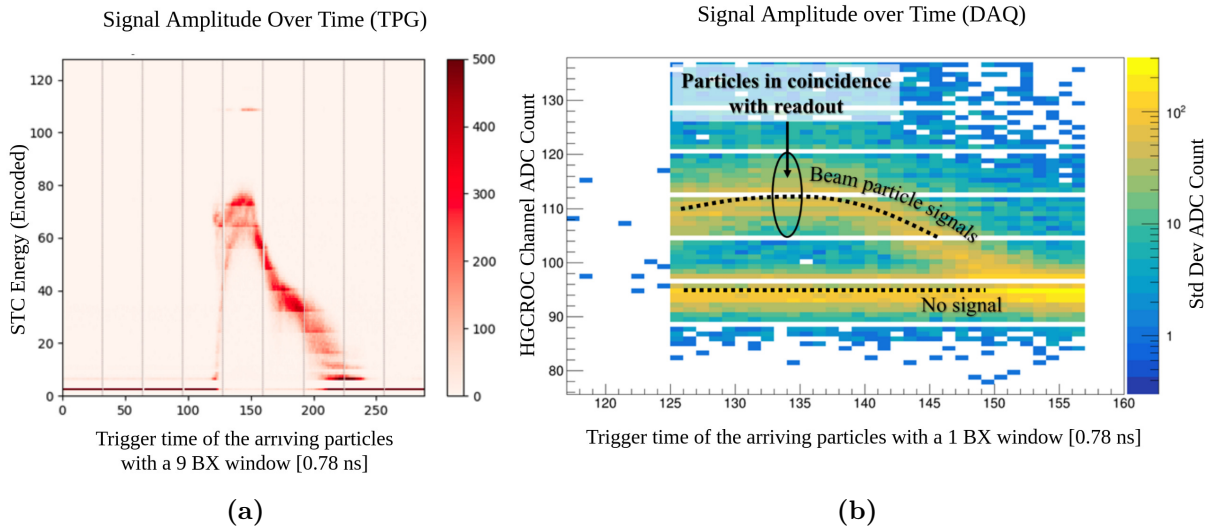


Figure 4.4: Signal amplitude as a function of time on (a) the TPG path for a single Hexaboard STC [2] (b) the DAQ path for a single Hexaboard cell [1]. Generated using offline analysis tools processing real physics data acquired during the beam test. More details can be found in text.

the major part of the charge is deposited within an area of a STC can be considered. Combining the STC energies and the trigger time of the arriving particles allowed to plot a typical energy over time signal (Figure 4.4a) from the data on the TPG path [2]. Setting the N parameter of the Extended Readout block to $N = 4$ allowed gaining a deeper insight into the TPG signal shape over a larger time window (i.e. 9 BX). In this example, this was necessary because the beam particles were energetic enough to cause the STC energy pulse to last for several BXs. For the DAQ case, a pion MIP beam run example where the major part of the charge was deposited within a single Hexaboard cell can be considered. Combining the HGCROC ADC energy measurements of the given cell and the trigger time of the arriving particles allowed to plot a typical energy over time (Figure 4.4b) MIP signal (also known as the “MIP peak”) from the data on the DAQ path [1]. In this example, the duration of the signal pulse almost completely falls within the one BX time window. When the beam particles are in coincidence with the readout (pulse is peaking), the beam particle signal size is about 20 ADC counts above the corresponding pedestal value when there is no beam. This falls within the $O(10)$ ADC count ballpark with respect to the MIP signal size expectation. The step size on the x-axis in both TPG and DAQ plots are $\frac{25}{32} \text{ ns} = 0.78 \text{ ns}$ because there are $157 - 125 = 32$ bins for the 25 ns time window when measuring trigger time.

Comparison of Internal and External Trigger Considering a collection of events that triggered data acquisition in a beam test run, one can plot the distribution of particle arrival times (Figure 4.5). The arrival times are always measured from the external scintillator signal, but the trigger decision in the BE is made using the external (blue) or internal (red) trigger L1A generation logic. Saving both the external and internal trigger decision for every acquired event, it is possible to evaluate the purity of the internal trigger versus the external trigger as a reference [1]. All the acquired events that have the internal (red) trigger condition met also have the external (blue) trigger condition met. It also never happened that there was an internal trigger but there was no external trigger. Hence, for the chosen internal trigger logic in the BE, the internal trigger purity was 100% in the considered example. As described for the previous measurements, the x-axis range and step size are 9 BX (due to the extended readout) and 0.78 ns, respectively.

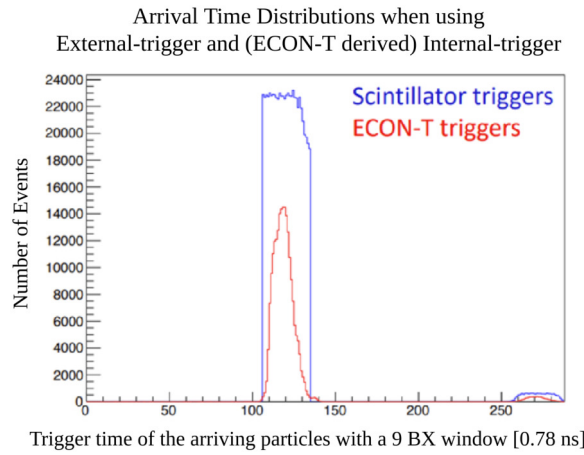


Figure 4.5: Arrival time distributions when using external / scintillator trigger (blue) and internal / ECON-T trigger (red) L1A generation logic [2]. Generated using offline analysis tools processing real physics data acquired during the beam test. More details can be found in text.

ECON-D Channel Retention The next notable result (Figure 4.6) considers using the ECON-D ASIC in Pass-through and Zero-suppression modes of operation [1]. Both plots indicate how often the signal passed the threshold for each HGCROC channel. As expected, signals coming from all channels are always observed when the ECON-D is in Pass-through mode (Figure 4.6a). Conversely, only the signals from a subset of channels which are exposed to the beam and above a programmable threshold are retained when the ECON-D is in Zero-suppression mode (Figure 4.6b).

The last and arguably most important (from the system validation point of view) beam test result is that it raised a suspicion regarding the signal integrity in the TPG and the DAQ vertical readout paths. This manifested as rare and unexpected instances of corrupted bits in the transmitted data on the DAQ path which suggested that there was a data integrity problem somewhere in this vertical chain. Furthermore, it motivated a systematic study on both the DAQ and TPG (for completeness, even though no similar issues were observed here) paths. More details about how these issues were confirmed and corrected for can be found in Section 4.3.

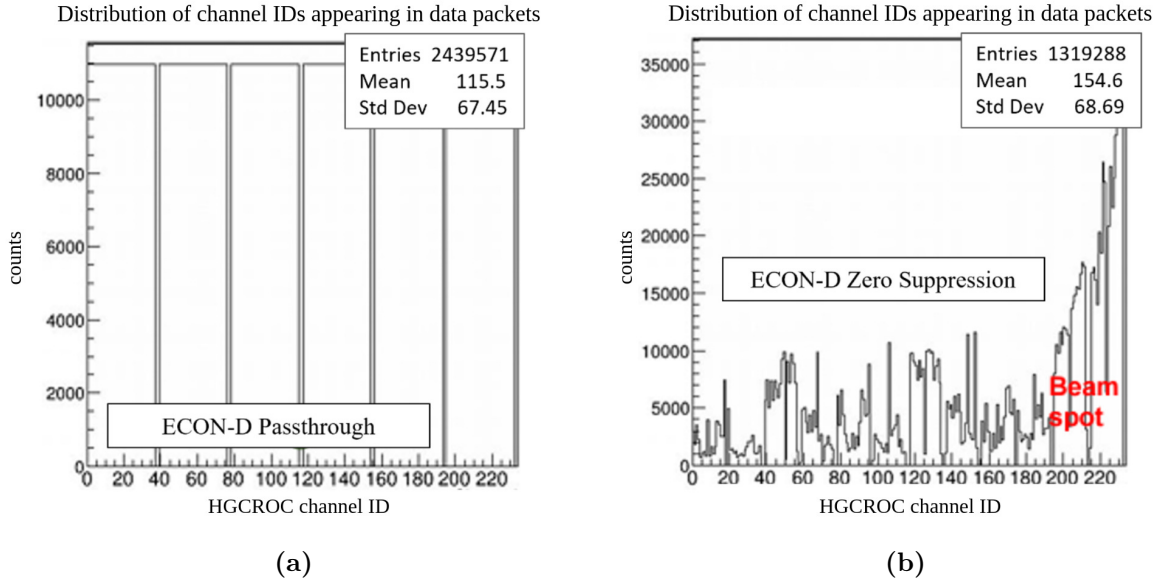


Figure 4.6: Distribution of the HGCROC channel retention by the ECON-D ASIC when configured in (a) Pass-through and (b) Zero-suppression mode from an electron beam run [1]. Generated using offline analysis tools processing real physics data acquired during the beam test. More details can be found in text.

4.2 The Key Interfaces for Studying Error Rates and Digital Consistency

The HGCAL electronics test system has multiple layers of components (and their interfaces) on its vertical readout and control paths. Hence, if there is an issue with the readout or control, it generally can originate from any of these steps in the vertical chain and often requires meticulous testing strategies in order to gain any deeper understanding. The approaches for the error rate measurements (more details can be found in Section 4.3) and the DAQ data analysis (more details can be found in Section 4.4) use several subsets of the components (and their interfaces) in this chain. All of these are put into perspective together here (Figure 4.7) while the individual subsets will be put into focus when each of the different studies is discussed in detail (i.e. in Section 4.3 and Section 4.4).

In addition to the vertical system architecture described earlier (more details can be found in Chapter 3) and the beam test related upgrades (as described in Section 4.1.2) to that architecture, there are two additional system upgrades made as a requirement for the error rate measurements. The first are the *TPG* and *DAQ e-Link Monitor firmware* blocks added to the FPGA in the Serenity BE. The second is the *DAQ e-Link Monitor software* application added to the RCC system running on the DAQ PC.

The TPG and DAQ e-Link Monitor firmware blocks in the BE FPGA compare a programmable subset of the received ECON-T and ECON-D 32-bit words versus a programmable expectation and increment a counter if this expectation is not met. These counters can be read from the Board Controller via the standard slow control interface. The DAQ e-Link Monitor software application in the RCC system compares the contents of the ROC fields in a received ECON-D pass-through event against a programmable expectation. The event is ignored if the expectation is met and otherwise saved to disk.

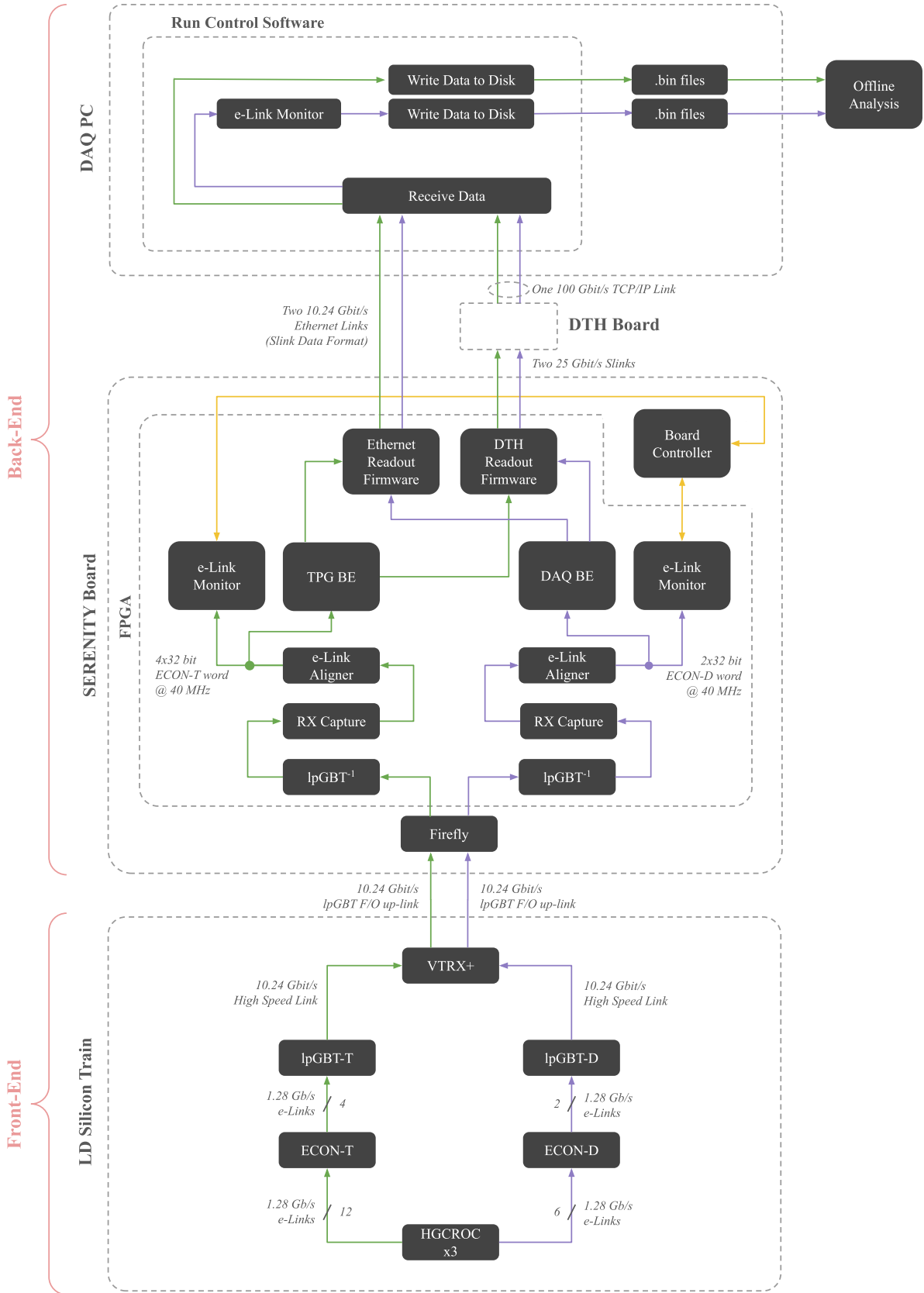


Figure 4.7: A general block diagram of relevant components and interfaces in the vertical HGCAL electronics test system used for the error rate measurements and the DAQ data analysis. Details can be found in text.

The exact subsets of components from the vertical chain used in different studies change depending on what ASIC is chosen as the FE data source (e.g. lpGBTs, ECONs or HGCROCs), where (and when) is the received data being checked in the BE (e.g. online in the e-Link Monitor firmware, online in the e-Link monitor software or in offline analysis) and what is the examined data type (i.e. TPG or DAQ). For example, when measuring the error rates in the path between (TPG or DAQ) lpGBTs and the core (TPG or DAQ) BE algorithms the data source is the (TPG or DAQ) lpGBT and the integrity checks are done in the (TPG or DAQ) online e-Link Monitor firmware blocks. That is similar to when the TPG path is extended to reach out to HGCROCs as the data source. However, the corresponding DAQ measurement does not use a similar architecture and the data is checked in the online e-Link Monitor software application in that case. Lastly, the DAQ data analysis is conducted using the HGCROCs as the data source and the data are saved to the DAQ PC and analysed offline. The nuances behind this variety will clarify when each of the different studies is discussed in detail (e.g. in Section 4.3 and Section 4.4).

4.3 The Error Rate Measurements

In order to locate and measure the suspected signal integrity issue in the multi-layered vertical system, the strategy was to conduct a series of tests where every next test would involve a greater number of components from the vertical chain, thus “elongating” the vertical path under test in each step. The data monitoring would always happen in the BE, but the data source would be moved deeper into the FE from one test to another (e.g. starting in the lpGBT, moving into the ECONs and ultimately into the HGCROCs). This strategy allowed to expose the effects of adding the individual FE ASICs (e.g. lpGBT, ECON and HGCROC) on the error rates in the vertical readout path.

The suspicion about the signal integrity issue was confirmed already in the first test, when the lpGBT was used as the data source. The measured non-zero Word Error Rate (WER) manifested only on the DAQ path and not the TPG path. Elongating the vertical chain with other FE ASICs (e.g. ECONs and HGCROCs) did not worsen the measured error rates on either the TPG or the DAQ paths.

The reason for the non-zero WER measured between the DAQ lpGBT and the BE e-Link Monitor firmware was rather well hidden because the source of the issue did not have to do with the up-link, but with the down-link direction of the DAQ path instead. Specifically, it turned out that the optical power of the used BE Firefly module driving the F/O down-link was too high, causing the FE VTRx+ to saturate and induce a non-zero error rate in the F/O up-link of the DAQ path. The issue was completely resolved by adding a simple optical attenuator between the BE Firefly module and the FE VTRx+ which moved the VTRx+ operating point well away from saturation.

After resolving the VTRx+ saturation issue, the focus of these studies shifted to setting an upper limit on the BER for both the TPG and DAQ paths. In doing so, the data transfer from the FE to the BE was monitored for an extended period of time $O(10\text{ d})$ using all the components in the vertical chain (e.g. starting with the HGCROCs and ending with the designated BE e-Link Monitors).

This section starts by presenting the impact of VTRx+ saturation on the WER measurements in the vertical path between the lpGBT and the BE firmware e-link Monitors.

The focus here is on the DAQ path which was discovered to be susceptible to the VTRx+ saturation. This section ends by presenting the work related to setting the BER upper limits on the TPG and the DAQ paths when all the components in the vertical chain (e.g. starting with the HGCROCs and ending with the designated BE e-Link Monitors) were used and the VTRx+ saturation was avoided.

4.3.1 The VTRx+ Saturation Impact on the lpGBT → BE Path

To measure the error rates between the TPG lpGBT and the DAQ lpGBT and their respective core BE algorithms (TPG BE and DAQ BE), a special subset of components in the vertical HGCAL electronics test system was used (Figure 4.8). The data sources were the lpGBTs and the data brought to the BE was monitored in the e-Link Monitor firmware blocks which “spy” on the data stream entering the core BE algorithms. The lpGBTs were not configured to sample the physical signals on their input e-Links coming from the ECONs. Instead, the lpGBTs were configured to hard-code a constant 32-bit pattern into every input lpGBT e-Link. As for the real ECON data (more details can be found in Chapter 3), these constant patterns would flow through the standard vertical system interfaces and again become available at the inputs of the core BE algorithms.

There are a total of four 1.28 Gbit s^{-1} ECON-T e-Links and a total of two 1.28 Gbit s^{-1} ECON-D e-Links received by the TPG and DAQ lpGBTs, respectively. Consequently, a total of four and two 32-bit words with lpGBT constant patterns and with a frequency of 40 MHz (i.e. on every BX) would be presented to the TPG and DAQ e-Link Monitor firmware blocks, respectively. The BE e-Link monitor firmware blocks were configured to compare each received “ECON” 32-bit word against the same pattern that was programmed in the lpGBT for any given lpGBT input e-Link. If any of the bits in a given lpGBT e-Link 32-bit word would not meet the expectation, the corresponding e-Link word error counter would be incremented by one. Through the standard slow control interface from the Board Controller, the user could read the latest error counter values in the e-Link Monitor firmware blocks.

Immediately after starting to monitor the acquisition of the lpGBT data, the suspicions raised by the beam test were confirmed: *A non-zero Word Error Rate was observed on the DAQ path.* Completing a short 15 min monitoring run, where the corresponding e-Link word error counters were sampled periodically every 5 s allowed to plot several representative parameters changing over time (Figure 4.9). Specifically, these are: the Word Error Counter value (Figure 4.9, Top), the Word Error Counter Difference between any two consecutive measurements (Figure 4.9, Middle) and the WER (Figure 4.9, Bottom). The erroneous words appeared in bursts, where each burst did not have either a predictable duration nor latency from the previous one. The 15 min run was sufficient to see the WER converge to an $O(1 \times 10^{-8})$ average rate.

These example measurements (Figure 4.9) are presented only for one DAQ lpGBT input e-Link. However, similar results were reproduced for the other DAQ e-Link. Moreover, similar results were reproduced using an entirely different LD single module Si Train v3. No signal integrity issues were observed on the TPG path.

Having observed errors in the DAQ up-link path between the lpGBT and the BE, a communication channel was immediately opened with the lpGBT [40] and the VTRx+ [41] groups at CERN. Naturally, examining this issue became a collaborative effort. For the

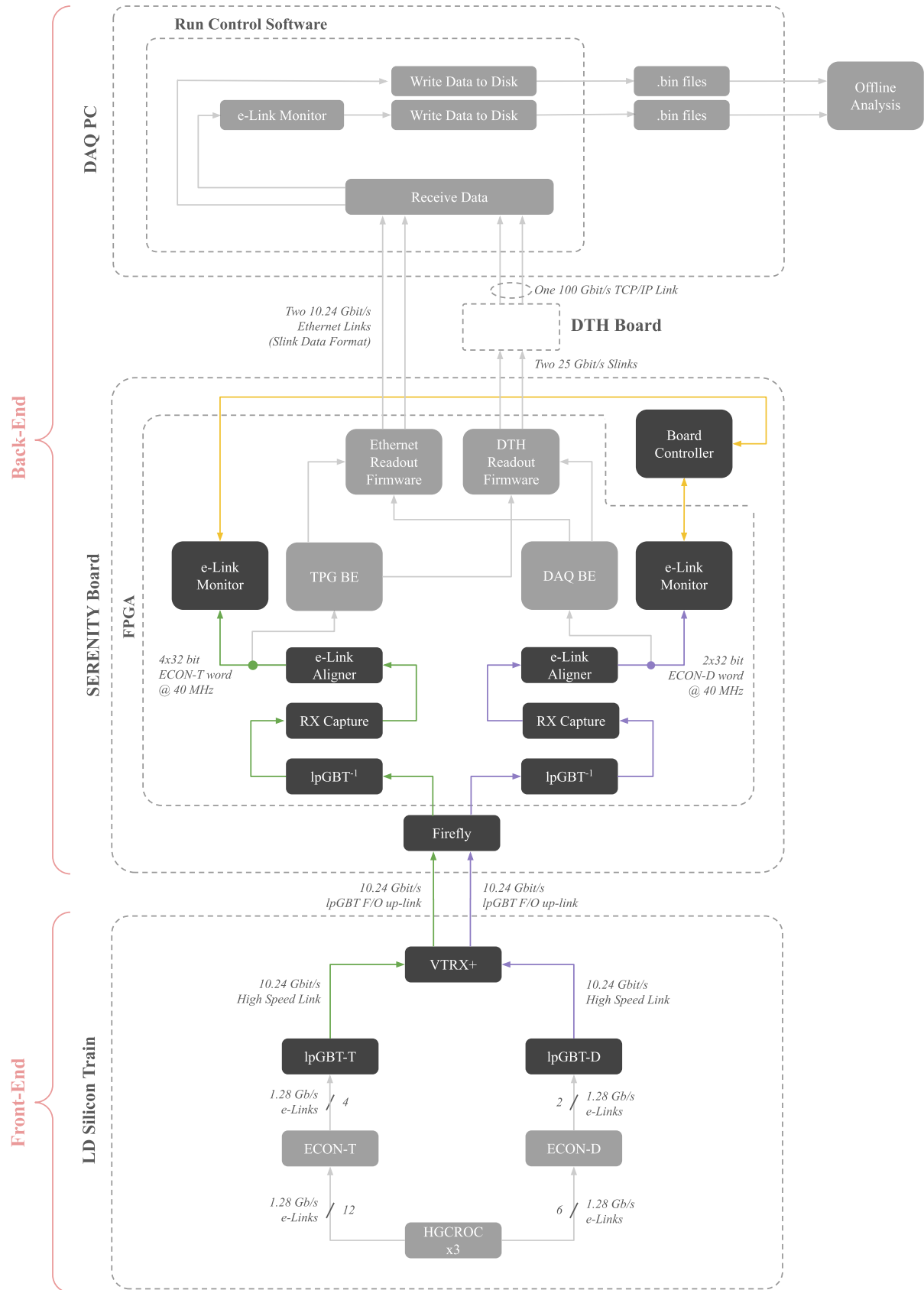


Figure 4.8: In focus (not greyed-out parts) of this block diagram are the components and interfaces of the vertical HGCAL electronics test system used for the error rate measurements on the TPG (green) and DAQ (purple) paths and the relevant slow control interface to monitor the results (yellow) when the origin of data is in the lpGBT. Details can be found in text.

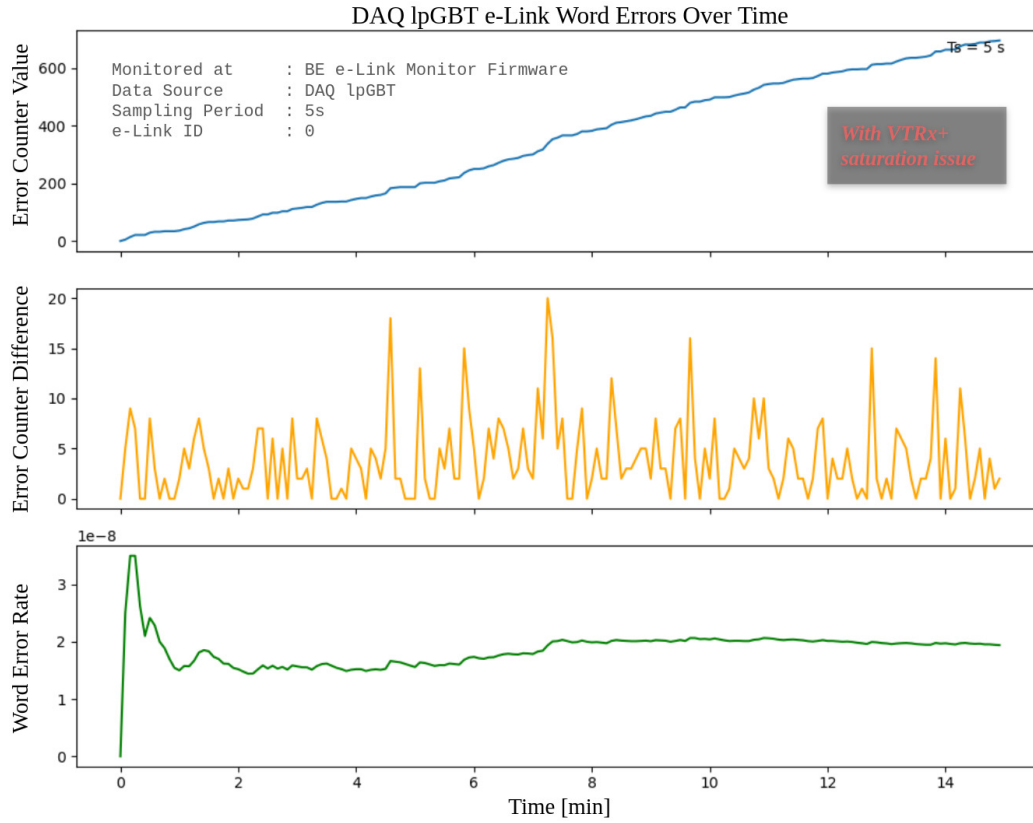


Figure 4.9: DAQ lpGBT e-Link word error counter (**Top**) word error counter difference (**Middle**) and word error rate (**Bottom**) monitored over a period of 15 min when the down-link optical attenuation *was not* applied to prevent the VTRx+ from occasionally entering the saturation regime. More details can be found in text.

components and corresponding interfaces under test, these groups expected a BER better than $O(1 \times 10^{-12})$. The measured WER of $O(1 \times 10^{-8})$ meant that the underlying BER was either equally bad or even worse than $O(1 \times 10^{-8})$. Regardless of the exact BER value, it was clear that the measured non-zero BER was **unacceptable**.

One of the most significant tests that contributed to better understand this issue was to measure the quality of the eye diagram of the DAQ $10.24 \text{ Gbit s}^{-1}$ F/O up-link (Figure 4.10a). This helped immediately spot an abnormality: *The serial signal rising and falling edges would occasionally get delayed with respect to the position of the eye-crossing where they should nominally appear*. This manifested as the blue 2D tail on the left side of the eye (Figure 4.10a).

To develop a more thorough understanding, eye diagrams of the TPG $10.24 \text{ Gbit s}^{-1}$ lpGBT F/O up-link and the DAQ 2.56 Gbit s^{-1} lpGBT F/O down-link were also measured. There was nothing obviously abnormal about the quality of either. However, something rather peculiar happened: the act of measuring the DAQ F/O down-link caused the DAQ F/O up-link eye diagram to take on the expected shape (Figure 4.10b). Moreover, the errors previously observed in the DAQ F/O up-link (Figure 4.9) disappeared.

To better understand how did the act of measuring the DAQ 2.56 Gbit s^{-1} lpGBT F/O down-link help, one should consider a critical change to the system architecture made for this measurement (Figure 4.11). The DAQ 2.56 Gbit s^{-1} lpGBT F/O down-link was brought to an optical splitter. The splitter outputs two logical copies of the input signal

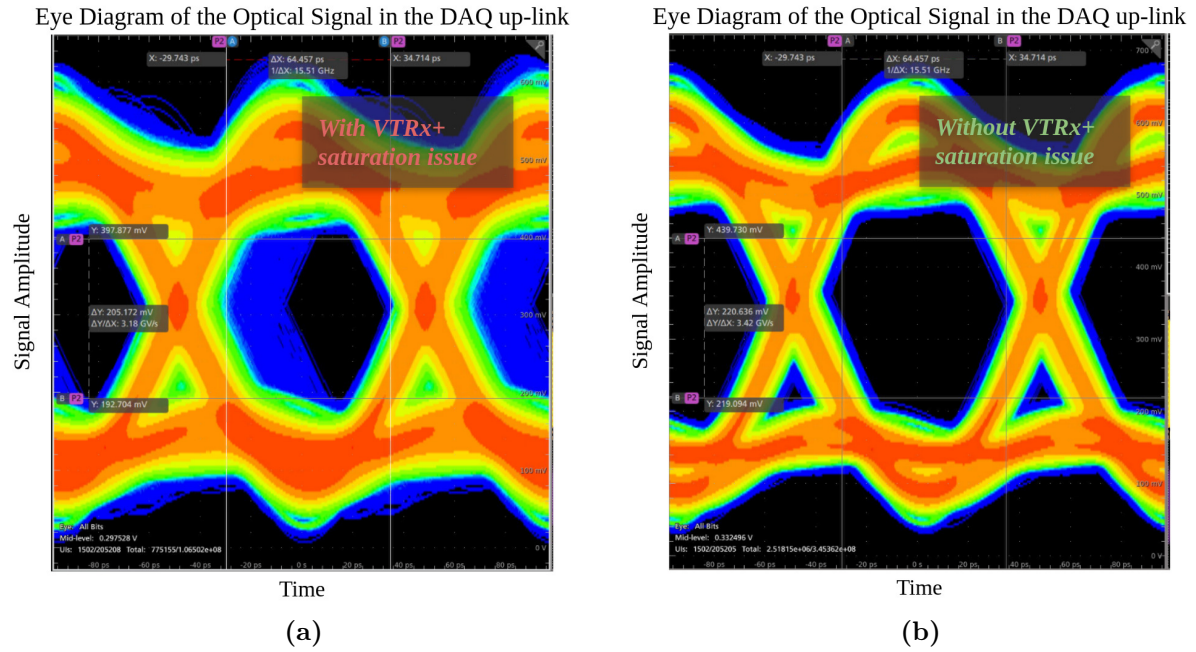


Figure 4.10: Eye diagram of the DAQ F/O up-link when the down-link optical attenuation was not (a) and was (b) applied to prevent the VTRx+ from occasionally entering the saturation regime. More details can be found in text.

(e.g. they carry the same information as the one modulated in the input signal). One output branch normally went to the VTRx+ while the other went into an oscilloscope used to measure the optical eye diagram of the signal. The subtlety here is that the optical splitter introduced a degree of attenuation to the BE Firefly output optical power in the DAQ 2.56 Gbit s⁻¹ lpGBT F/O down-link. These insights motivated a comprehensive study [64] conducted by the lpGBT and the VTRx+ groups at CERN. They also wrote an application note summarising the details about this effect and how to mitigate it [65].

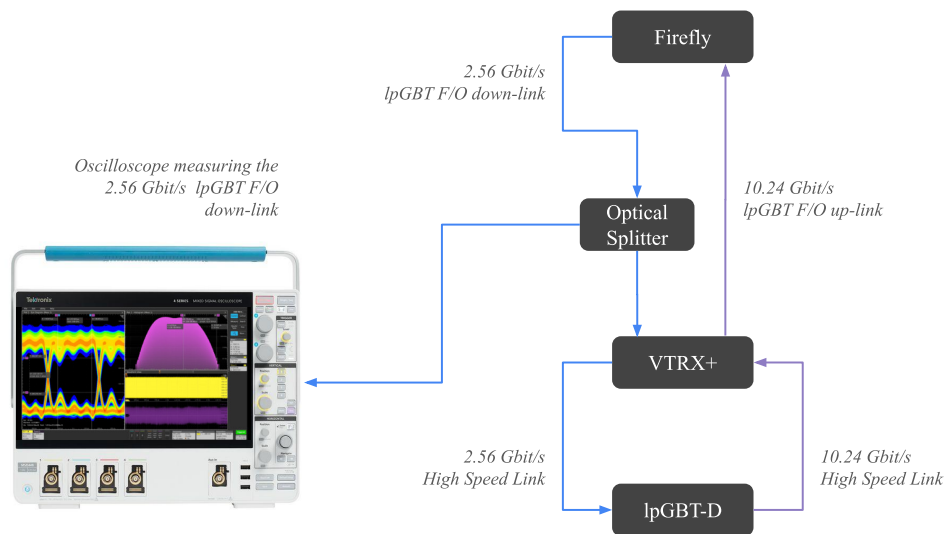


Figure 4.11: A block diagram of the relevant components in the system architecture related to measuring the DAQ lpGBT F/O down-link eye diagram quality. More details can be found in text.

As explained by the VTRx+ group, after the optical signal gets converted into a differential electrical signal in the internal photo-diode, the electrical signal goes through a sequence of differential amplifiers. The relationship of the resulting (e.g. at the output of the final amplifier stage) electrical signal differential amplitude and the input optical power is generally linear. However, that is not the case if the differential amplitude at any of the amplifier stages inputs is too big. This causes the signal in the output of the amplifier stages to be non-linearly transformed (or saturated). Before introducing optical attenuation to the DAQ down-link an amplifier stage would occasionally saturate its output. In turn, the differential electrical signal output by the VTRx+ over the High-Speed LD interface to the DAQ lpGBT would appear as noisy. This noise would also propagate to the clock recovered from the down-link signal in the DAQ lpGBT. Ultimately, using the noisy clock to serialise the data going out of the DAQ lpGBT would manifest in the DAQ up-link having a noisy eye-diagram and a non-zero BER when measured. The attenuation introduced with the optical splitter in turn reduced the VTRx+ internal differential signal amplitudes, preventing the amplifier stages from ever saturating their outputs. The reason why there were no issues in the TPG up-link, even when the down-link optical power was not attenuated, is because the TPG lpGBT has a block that filters out the noisy component from the clock reference it receives from the DAQ lpGBT.

The discovery of the issue with signal integrity related to the VTRx+ saturation has non-negligible significance because it affects all CERN experiments (i.e. not just HGCAL) that plan to use the lpGBT, the VTRx+ and the Firefly components in their FE $\leftrightarrow$ BE interfaces. While different experiments are preparing for HL-LHC, it was HGCAL who had first brought this issue to the attention of the relevant groups. An implemented short-term solution (recommended by the VTRx+ group) was to install a standard off-the-shelf 3 dB optical attenuator on the DAQ down-link path. A long-term solution will be based on fine-tuning the output optical power of Firefly modules from the Serenity BE Board Controller through a standard I²C interface.

Regardless of the observed VTRx+ saturation issue, the Firefly I²C interface will allow to gradually increase the Firefly output optical power as the detector FE “ages”. Specifically, this is needed because due to negative radiation effects the VTRx+ will become less sensitive to the input optical power level [41].

For posterity, one can conceptualise a general procedure for fine-tuning the Firefly parameters throughout the lifetime of the detector. The procedure would use custom DAQ BE firmware and software that scans the configuration space of every output channel of every DAQ BE Firefly, finds and then saves the settings that correspond to optimal operating points of the FE VTRx+ ASICs by measuring DAQ BE up-link WERs. This would allow normal data acquisition using the found optimal Firefly settings. However, the procedure should be periodically repeated (e.g. every $O(1\text{ w})$) in case that the optimal VTRx+ operating point changed since the previous iteration.

4.3.2 The BER Upper Limits in the HGCROC $\rightarrow$ BE Path

Having resolved the VTRx+ saturation issue, the origin of the monitored data was moved deeper into the FE: first into the ECONs and later into the HGCROCs. There were no issues observed for any chosen FE ASIC as the data source and that was the case for both the TPG and the DAQ paths. The focus here is on the tests involving the HGCROCs, as these tests are the most significant. This is because the addition of the HGCROCs allowed

to monitor the signal integrity and ultimately set an upper limit on the BER of the full vertical chain of components (e.g. HGCROC $\rightarrow$ ECON $\rightarrow$ lpGBT $\rightarrow$ VTRx+ $\rightarrow$ BE) for both the TPG and DAQ paths. To perform these measurements, two subsets of specifically configured components from the vertical HGCAL electronics test system were used (Figure 4.12). These had to be different for the TPG and the DAQ cases, because there are some fundamental differences between the TPG and DAQ data formats.

Regarding FE ASIC configuration, each HGCROC was configured to hard-code a 10 bit value $682 = 0b1010101010$ into each of its ADC channels, ensuring plenty of "0" $\leftrightarrow$ "1" transitions in the data. The HGCROCs were also configured to mask the ToA and ToT measurements. Both ECONs were configured to sample the physical e-Links coming from the HGCROCs. The ECON-T was configured to be in the STC4E3M mode. Given the HGCROC configuration, the TCs reaching the ECON-T were also constant. Ultimately, each 7-bit STC value formed in the ECON-T was $0x52 = 0b1010010$ which decodes as decimal 10752². The ECON-D was configured to be in Pass-through mode. In turn, the bulk of each ECON-D output packet consisted of the hard-coded HGCROC data. The lpGBTs were configured to sample the physical e-Links coming from the ECONs and send these data to the BE via the F/O interface.

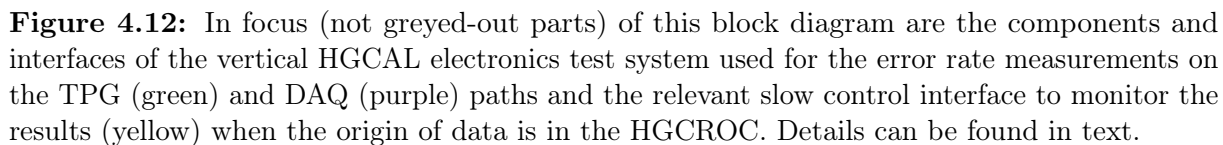
Regarding the TPG path, there were four ECON-T output e-Links, each giving a 32-bit word per BX (more details can be found in Chapter 3). Considering the STC4E3M data format (more details can be found in Chapter 2), each of these words had a constant part (e.g. STC energies) and a part that was changing from one BX to another (e.g. header consisting of the BX counter). Hence, it was possible to use the e-Link Monitor firmware blocks by configuring them to compare the non-changing subset of each of the four words against an expectation for every BX.

Regarding the DAQ path, even though there were two ECON-D output e-Links, each giving a 32-bit word per BX (more details can be found in Chapter 3), it was not possible to monitor the ECON-D data in the BE e-Link Monitor firmware block. This was because there is no part of the 32-bit ECON-D e-Link words that is guaranteed to not change from one BX to another. There are two reasons for this. Firstly, the content of the ECON-D e-Links changes depending on the presence of L1A fast commands. For example, in the absence of L1As, there is no DAQ event data being streamed out. The ECON-D only outputs Idle words in that case. Secondly, even if the ECON-D output e-Links were always busy outputting Pass-through events, there would still be BXs when each e-Link would be bound to transmit data that does not have to do with the constant HGCROC fields. For example, these are the ECON-D packet header, sub-packet headers, etc. (more details can be found in Chapter 2). This is why it was decided to monitor the data on the DAQ path in the RCC e-Link Monitor software application on the DAQ PC.

The TPG Path

Considering the FE ASIC configuration choices, along with the format of one BX worth of TPG data at the output of the ECON-T, and the output of the BE e-Link Aligner firmware (more details can be found in Chapter 2 and Chapter 3), it is possible to predict the contents of each BX worth of TPG data that reaches the e-Link Monitor firmware blocks. These blocks were configured to monitor only the parts of the four 32-bit ECON-T

²The expected ballpark for each TC value is approx. $4 \times 682 = 2728$. In turn, for each STC that value is approx. $4 \times 2728 = 10912$. However, due to conversions between different number systems on the TPG path, the exact STC value is 10752.



words that were not changing from one BX to another (Table 4.1). The *Word* and *Mask*

Table 4.1: The configuration of the TPG e-Link Monitor firmware blocks for each of the four ECON-T output e-Links (eTx). More details can be found in text.

	ECON-T eTx[0]	ECON-T eTx[1]	ECON-T eTx[2]	ECON-T eTx[3]
Word	0xXFFFFFFFA	0x54A952A5	0x4A951A54	0xA9520000
Mask	0x0FFFFFFF	0xFFFFFFFF	0xFFFFFFFF	0xFFFFFFFF

fields specify the expected content and which bits to check, respectively. For example, the leading 0x0 in the ECON-T eTx[0] *Mask* field means that the leading four bits should not be checked in ECON-T eTx[0] (i.e. these encode the 4-bit BX number which changes from one BX to another). Similarly, an expected 7-bit STC value (e.g. for STC[11]) can be seen at bit positions 22...16 of the ECON-T eTx[3] *Word* field. At the same positions the *Mask* field specifies a logical "1", meaning that these bits should be checked on every BX. The same concepts apply to choosing all the remaining bits in the four pairs of Words and Masks.

Monitoring the TPG path happened in a number $O(10)$ of runs, with each run having a different duration. Note that two single module Trains were used in parallel to double the total number of bits checked. In terms of the total number of checked bits, using the two FE Trains is logically the same as if a single Train was used throughout an effective period of 56 d during which trigger data appeared with a 40 MHz frequency (e.g. in every BX). Not a single bit error was found in the total $40 \text{ MHz} \times 56 \text{ d} = 1.935 \times 10^{14}$ BXs checked. The absence of errors in long monitoring runs of the four 32-bit output ECON-T e-Links means that there were also no errors in the HGCROC $\rightarrow$ ECON-T interface. In other words, explicit monitoring of the four 32-bit output ECON-T e-Links allowed implicit monitoring of the transmission of twelve 32-bit worth of digital information in the TPG path (i.e. there are twelve e-Links in the HGCROC $\rightarrow$ ECON-T interface). This allowed to set an upper limit [50] with a 95% CL on the TPG bit error rate BER_{TPG} (Equation 4.1).

$$\text{BER}_{\text{TPG}} < \frac{3.00}{1.935 \times 10^{14} \text{ BXs} \times 12 \text{ words/BX} \times 32 \text{ bits/word}} \approx 4 \times 10^{-17}, \text{ at 95\% CL} \quad (4.1)$$

The DAQ Path

Considering the FE ASIC configuration choices, along with the format of one L1A worth of DAQ data (or one DAQ event) at the output of the ECON-D, the DAQ BE and the Serenity BE SLink interface (more details can be found in Chapter 2 and Chapter 3), it is possible to predict the contents of each DAQ event reaching the DAQ PC. In turn, this allows to compare every received DAQ event versus an expectation in the RCC e-Link Monitor software. To keep the e-Link Monitor software application fast and simple but still manage to check plenty of bits for every event, it was chosen to monitor only the words in the DAQ event that originated from hard-coding the HGCROC ADCs and masking the ToA and ToT measurements.

Specifically, after one drops the SLink header and trailer, the DAQ BE packet header, the ECON-D packet header, the ECON-D sub-packet headers, the ECON-D CRC as well as the HGCROC calibration and common mode channels from the ECON-D payload, one is left with 204 hard-coded 32-bit words per DAQ event whose content is always the

same and equals to 0x2AAAA800. Each one of these 32-bit patterns corresponds to an HGCROC field in the ECON-D payload and reads that the single bit **Tc** and **Tp** flags are both 0, the 10-bit **ADC-1** and **ADC** fields are both 0b1010101010 and the 10-bit **ToA** field is 0b0000000000, as expected given the FE ASIC configuration. There are a total of 124 64-bit words on the DAQ path per BX in the SLink format, which means that the monitoring efficiency for each event is $(204 \times 32 \text{ bits}/2)/(124 \times 64 \text{ bits}) \approx 82\%$. One could also check the remaining 18% in the BE software application, but the checking logic would become significantly more complicated. This is mainly because some of the words change from one BX to another (e.g. L1A timestamps, ECON-D CRC, etc.). Because this checking is done in online software, over-complicating it may hit limitations of the computation power of the DAQ PC (e.g. not being able to compute all the parameters about the current event before the next event arrives). That is why it was decided to monitor only the 82% of the bits per DAQ event.

The e-Link Monitor software can only work with full ECON-D Pass-through packets (e.g. truncated packets are not supported). To comply with this, the DAQ BE fast control firmware was configured to send out periodic L1As and with a frequency smaller than the 336.1 kHz limit (more details about this limit can be found in Chapter 3) beyond which truncation would appear. Periodic L1As were also chosen instead of random L1As to avoid the possibility of some events being truncated.

Monitoring the DAQ path happened in a number $O(10)$ of runs, with each run having a different duration and different periodic L1A frequencies. For example, some runs were operated with an L1A frequency of 100 kHz while others were done so with an L1A frequency of 300 kHz. Note that two single module Trains were used in parallel to double the total number of bits checked. In terms of the total number of checked bits, using the two operating L1A frequencies and two FE Trains is logically the same as if a single module Train was used throughout an effective period of 108 d while being operated with a 100 kHz effective periodic L1A frequency. Not a single bit error was found in the total $100 \text{ kHz} \times 108 \text{ d} = 9.331 \times 10^{11}$ events checked, allowing to set an upper limit [50] with a 95% CL on the DAQ bit error rate BER_{DAQ} (Equation 4.2).

$$\text{BER}_{\text{DAQ}} < \frac{3.00}{9.331 \times 10^{11} \text{ events} \times 204 \text{ words/event} \times 32 \text{ bits/word}} \approx 5 \times 10^{-16}, \text{ at 95\% CL} \quad (4.2)$$

4.4 System-wide Digital Consistency Analysis

Having set the respective TPG and DAQ BER upper limits to 4×10^{-17} and 5×10^{-16} gave high confidence in the quality of the electrical and optical interfaces in the vertical HGCal electronics test system (more details can be found in Section 4.3). Additionally, seeing the data originating from the HGCROC channel ADCs in the BE e-Link monitoring blocks does give confidence that the different components from the vertical chain logically work well together. However, the BER results do not give insights into some subtle details regarding the digital consistency of different components in the vertical chain. These are rather simple on the TPG path, but are significantly more complex on the DAQ path. Hence, the results of system-wide digital consistency checks regarding the TPG path will be only briefly summarised, whereas the corresponding results for the DAQ path will be presented in detail.

4.4.1 The TPG Path

The only information related to system-wide digital consistency not covered by the TPG BER measurements is in the HGCROC and the ECON-T headers. These headers allow BX-to-BX synchronisation in the HGCROC $\rightarrow$ ECON-T $\rightarrow$ TPG BE vertical chain. A test was conducted where two kinds of status flags were monitored over a 5 d period. To address the HGCROC $\rightarrow$ ECON-T interface, the ECON-T internal registers `ChAligner.i.hdr_mm_cntr` were monitored via slow control. Here, $i \in \{00, 01, 02, \dots, 09, 10, 11\}$ specifies each of the twelve input ECON-T e-Links. The ECON-T checks the HGCROC headers, and these registers allow the reading of the count of how many times it detects an incorrect HGCROC header for two BX in a row on a given input e-Link. To address the ECON-T $\rightarrow$ TPG BE interface, the *locked* status of the Stage 1 TPG BE Unpacker blocks was monitored. Having used two single-module Trains in the FE and two BE unpacker blocks, this is equivalent to using one vertical system over a 10 d period. No evidence of synchronisation loss was observed in either of the two interface layers. This was true for the total $10 \text{ d} \times 40 \text{ MHz} = 3.456 \times 10^{13}$ TPG packets monitored during this test. This allowed to set an upper limit [50] with a 95% CL on the TPG digital consistency error rate DCER_{TPG} (Equation 4.3).

$$\text{DCER}_{\text{TPG}} < \frac{3.00}{3.456 \times 10^{13}} \approx 9 \times 10^{-14}, \text{ at 95\% CL} \quad (4.3)$$

4.4.2 The DAQ Path

There are several aspects on the DAQ path that require careful analysis in order to validate system-wide digital consistency in the vertical test system. For every received L1A fast command, each component in the vertical chain generally: receives a packet from the previous component and checks the received packet for issues³; performs a data aggregation or selection algorithm; formats an output packet containing an L1A timestamp, various status or quality flags and a number of protection bits in its header or trailer; and sends out the resulting packet to the next component in the sequence. Specifically, the analysis explores: the relevant status and quality flags about the different DAQ interfaces, the vertical reconstruction results in the ECON-D, and the cross-component L1A timestamp agreement between the HGCROC, ECON-D, DAQ BE, and the SLink⁴ readout. This section first summarises the general parameters of the analysed data acquisition run (e.g. exact hardware used, component configuration, L1A rates, run duration, etc.), and then presents the results of the analysis conducted over the acquired data. The results are mainly organised into parts corresponding to the Headers or Trailers introduced by the different components (e.g. SLink, DAQ BE, ECON-D) in the vertical chain and accessible by the BE⁵. Although the L1A timestamps from different components exist in component Headers or Trailers, they are discussed all together in a separate part.

General Parameters of the DAQ Run

The parameters of the data acquisition run for this analysis were mainly driven by three factors: to use the closest to final components in the vertical chain; to operate the system

³Except the HGCROC which is the first ASIC in the vertical path.

⁴The SLink timestamps are directly taken from the TCDS2 reference.

⁵HGCROC Header information gets absorbed into the ECON-D Header.

in the closest to final conditions; and to save as much data as feasible in these conditions. Regarding the first factor, the full vertical readout path: HGCROCs $\rightarrow$ ECON-D $\rightarrow$... $\rightarrow$ DAQ BE $\rightarrow$ DTH $\rightarrow$ DAQ PC was used (Figure 4.13). Specifically, the DAQ BE readout implementation revolved around the DTH and not the Ethernet-oriented readout alternative. Regarding the second factor, the DAQ BE fast control firmware was configured to send out random-spaced L1As with an average frequency of 750 kHz throughout the duration of the data acquisition run ⁶, which is the nominal HL-LHC average L1A rate. A peculiarity of how the RCC system and the Serenity BE readout firmware were developed for the beam test was to save both DAQ and TPG data (as well as additional debugging data) of each selected BX (e.g. via the L1A fast command). In turn, the desired operating L1A rate of 750 kHz also motivated the use of the DTH-oriented readout instead of the Ethernet-oriented readout, because the latter did not have sufficient bandwidth to work at so high L1A frequencies. Additionally, the DTH output packet rate and the overall data throughput going into the DAQ PC hit the limitations of the ability to write data to permanent storage (e.g. to the DAQ PC hard-drive). To address this issue: the HGCROCs were configured to send constant ADC patterns so that only one channel in each HGCROC half is above a zero-suppression threshold configured in the ECON-D; the ECON-D was configured to be in the Standard (i.e. zero-suppression) mode of operation; the amount of non-DAQ data being saved (e.g. TPG data or debug information) was significantly reduced; and the chosen DAQ-PC storage unit was a RAM block instead of the usual hard-drive. The RAM unit can save data at high packet rates and at a high throughput, but is limited in the total amount of data one can save during a single data acquisition run. The RAM memory constraint directly influenced the third driving factor of the DAQ run: it was possible to save up to 3 min worth of DAQ data during a single run operated at an average L1A rate of 750 kHz. This allowed to acquire about $3 \text{ min} \times 750 \text{ kHz} = 135 \times 10^6$ events, which is considered to be a sufficient amount of data for this analysis. A subtle detail important for this analysis is that the ECON-D was configured to vertically reconstruct the event, bunch and orbit counters coming from the six HGCROC halves and transmit those to the DAQ BE.

The SLink Headers and Trailers

Each SLink packet has a Beginning of Event (BOE) and an End of Event (EOE) field in its Header and Trailer, respectively. These allow the DTH to “know” when each SLink packet transmission by the DAQ BE started and is finished, respectively. These fields should always have the same values, from one SLink packet to another. Comparing the BOE (Figure 4.14a) and EOE (Figure 4.14b) fields of every saved SLink packet against the respective expectations allowed to check the validity of each pattern. There were no issues observed in any of the acquired events. This gives confidence that the readout interfaces in the DAQ BE $\rightarrow$ DTH $\rightarrow$ DAQ PC chain work as expected, regarding these aspects.

⁶This does not entirely correspond to the HL-LHC beam profile. For example, there will be L1T-accepted DAQ events from some BXs and not others. In turn, the instantaneous L1A rate will at times be below and at other times above the nominal 750 kHz L1A rate.

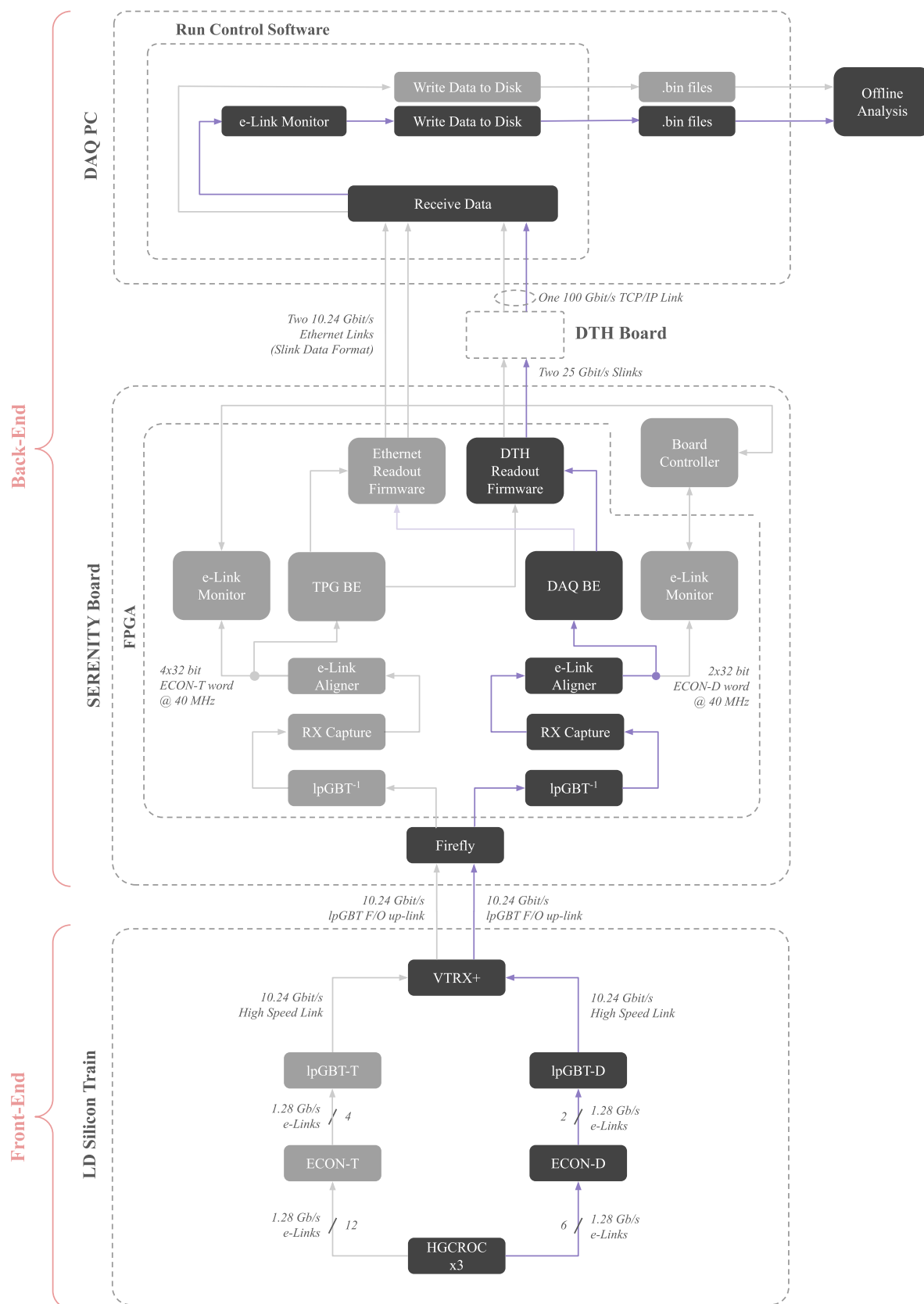


Figure 4.13: In focus of this block diagram (not greyed-out parts) are the components and interfaces on the DAQ (purple) path of the vertical HGCAL electronics test system used for the DAQ data analysis. Details can be found in text.

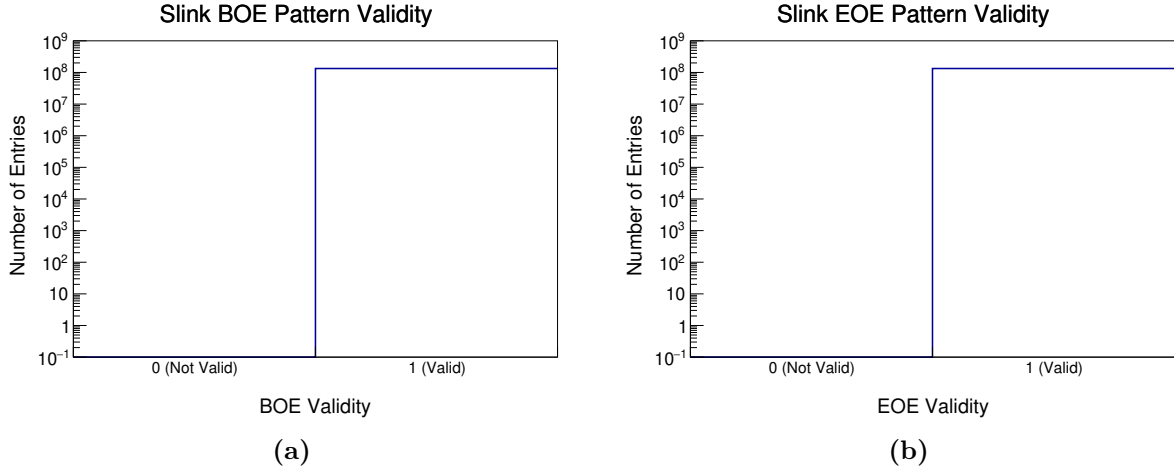


Figure 4.14: Distribution of pattern validity of (a) **BOE** and (b) **EOE** fields in SLink packet Headers and Trailers, respectively. More details can be found in text.

Each SLink packet has a DAQ BE CRC and a CRC field in its Trailer. These are calculated by two sequential components (SLinkRocket and FED) in the DTH readout firmware on the Serenity FPGA and over the full SLink packet, including the Header and Trailer [47]. For every SLink packet, these two have to be the same. There were no differences observed in any of the acquired events (Figure 4.15). This gives confidence that the readout interfaces in the DAQ BE $\rightarrow$ DTH $\rightarrow$ DAQ PC chain work as expected, regarding this aspect.

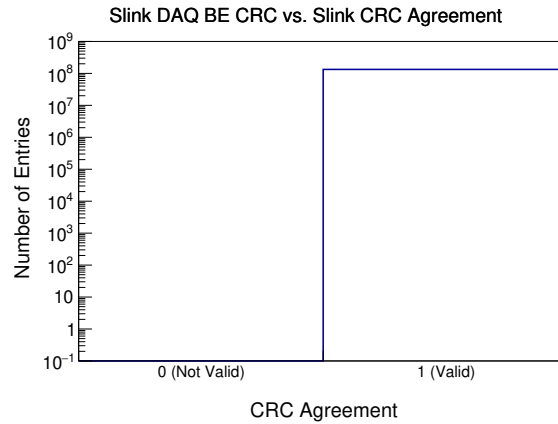


Figure 4.15: Distribution of the agreement between the **DAQ BE CRC** and **CRC** fields in SLink packet Trailers. More details can be found in text.

Each SLink packet has the BX ID field in its Trailer. This field conveys the BX counter value (taken from the TCDS2 reference) from which each SLink packet originates from. The distribution of the SLink packet BX counter values (Figure 4.16) showed to be uniform across the orbit, taking on values from $\{1, 2, 3, \dots, 3564\}$. This is exactly as expected given that the DAQ BE fast control firmware was configured to generate random L1As (having random distance between any two L1As).

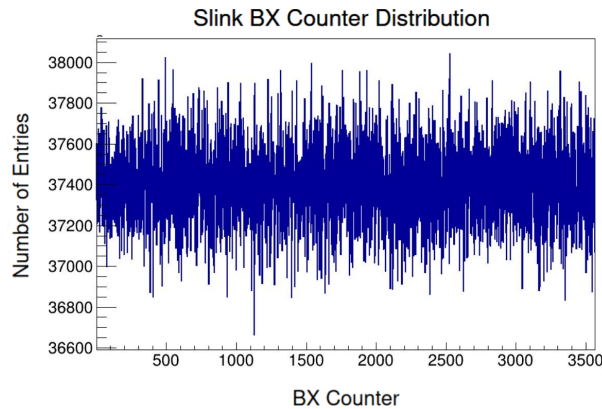


Figure 4.16: Distribution of the **BX ID** field in SLink packet Trailers. More details can be found in text.

The DAQ BE Headers

Each DAQ BE packet has the ECON-D Packet Status field in its Header. This field generally conveys the quality of each ECON-D packet it is supposed to receive. The DAQ BE ECON-D Packet Status field can take on eight possible 3-bit values: 000 - Normal packet (OK), 001 - No packet because it was discarded due to over-size (OS), 010 - Packet with CRC Error (CRC), 011 - Packet with event counter mismatch (EC), 100 - No packet due to Event Builder timeout (TO), 101 - No packet due to bunch counter or event counter mismatch (BC/OC), 110 - No packet because it was discarded due to Main Buffer overflow (OF) and 111 - Inactive ECON-D input (OFF) [34]. The DAQ BE marked all packets it received from the ECON-D as Normal for all acquired events (Figure 4.17). This gives confidence that the ECON-D $\rightarrow$ DAQ BE chain works as expected, regarding this aspect.

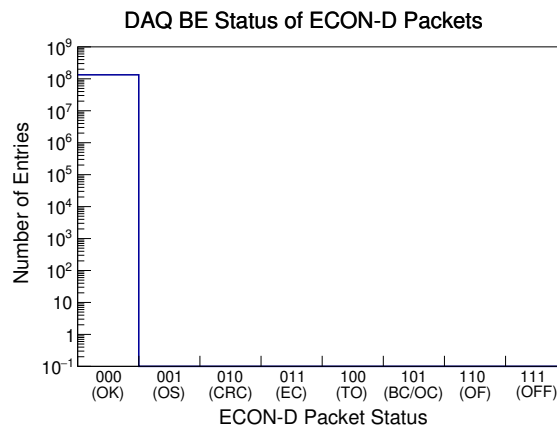


Figure 4.17: Distribution of the **ECON-D Packet Status** field in DAQ BE packet Headers. More details can be found in text.

The ECON-D Headers

Each ECON-D packet has the Payload Length field in its Header. With the ECON-D being in Standard mode (Zero-suppression), this generally varies from one event to another (e.g. depending how many HGCROC channels pass the ECON-D selection criteria). However, considering the HGCROC and ECON-D configuration used in this data acquisition run, all the ECON-D packets are expected to be of the same size (i.e. 21×32 -bit words).

This field had the expected value (21) in all the acquired events (Figure 4.18). This gives confidence that the HGCROC $\rightarrow$ ECON-D chain works as expected, regarding this aspect.

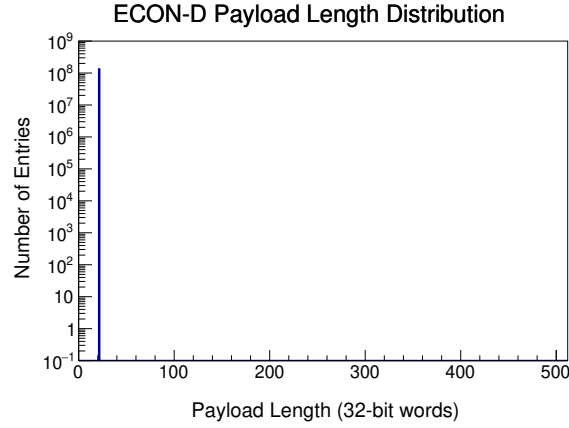


Figure 4.18: Distribution of the **Payload Length** field in ECON-D packet Headers. More details can be found in text.

Each ECON-D packet has the P and T flags in its Header. These indicate the mode the ECON-D was configured to operate in and whether it sent out a truncated packet, respectively. Given the HGCROC and ECON-D configuration choices for this run, these fields are expected to always convey that the ECON-D was in Standard Mode and that the packets were never truncated. This was the case in all the acquired events (Figure 4.19a and Figure 4.19b), which gives confidence that the HGCROC and the ECON-D work as expected regarding these aspects.

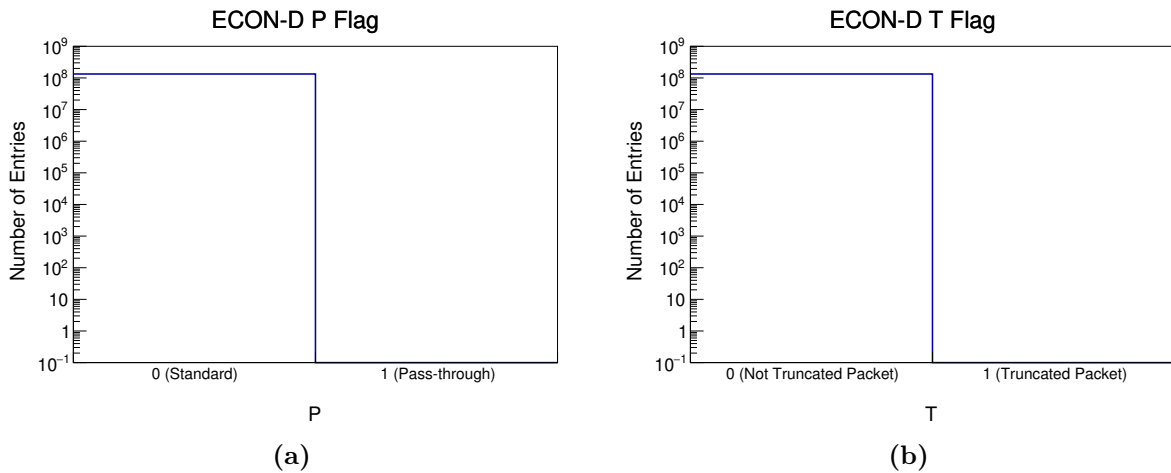


Figure 4.19: Distribution of (a) **P** and (b) **T** flags in ECON-D packet Headers. More details can be found in text.

Each ECON-D packet has the CRC for Event Header field in its Header. This value is calculated by the ECON-D itself over the ECON-D packet Header (except the Hamming field). Calculating this value offline and comparing that expectation to what the ECON-D had sent allows to evaluate the validity of the ECON-D Header protected by this CRC. The result of this check was valid for all the acquired events (Figure 4.20). This gives confidence that the ECON-D $\rightarrow$ DAQ BE chain works as expected, regarding this aspect.

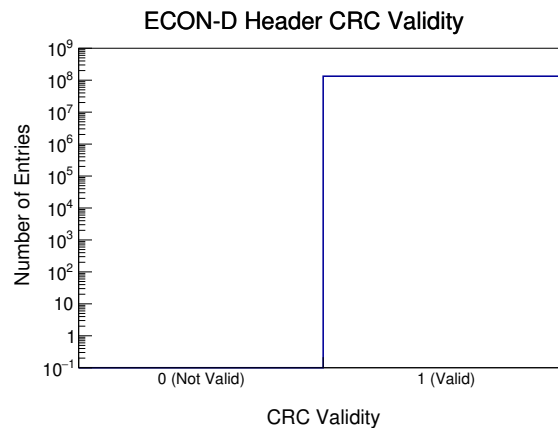


Figure 4.20: Distribution of the validity of the ECON-D Header, protected by the **CRC for Event Header** field. More details can be found in text.

Each ECON-D packet has the RR field in its Header. This 2-bit field is used to indicate up to two simultaneous (or individual) reset requests raised by the ECON-D. Each of the two bits in the RR field corresponds to one of the two possible reset requests. For each bit, a logical "1" means there is a reset request active. Otherwise, the corresponding bit is set to a logical "0". The user can configure the underlying logic that drives the value for each of the two reset requests. In the conducted run, this logic was determined by the default ECON-D configuration. The ECON-D did not request either of the two kinds of resets for all the acquired events (Figure 4.21). Regarding this aspect, these results give confidence that the ECON-D remained in a valid operating state throughout this data acquisition run.

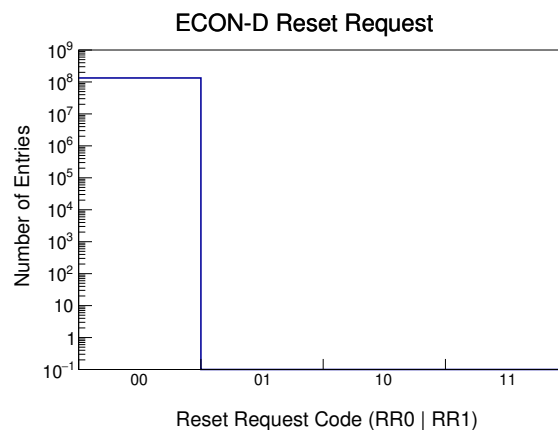


Figure 4.21: Distribution of **RR** field values in the ECON-D packet Headers. More details can be found in text.

Each ECON-D packet has the E flag in its Header. This flag is used to indicate whether the ECON-D receives the HGCROC Headers and Trailers when expected. This was the case for all the acquired events (Figure 4.22). In turn, this gives confidence that the HGCROC → ECON-D chain works as expected, regarding this aspect.

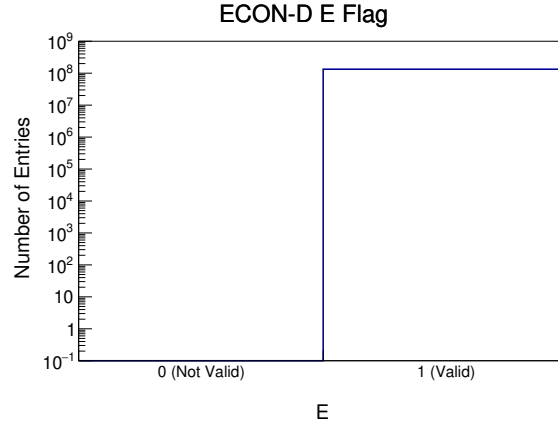


Figure 4.22: Distribution of the **E** flag in ECON-D packet Headers. More details can be found in text.

Each ECON-D packet has the H/T and E/B/O fields in its Header, through which it indicates the vertical reconstruction results from its interface with the six HGCROC halves. Through the H/T and E/B/O fields, the ECON-D conveys how well do the HGCROC halves agree amongst one another when sharing the Header and Trailer (H/T) patterns and event, bunch and orbit (E/B/O) timestamps with the ECON-D. Both H/T and E/B/O fields are 2 bit wide, being able to encode four levels of vertical reconstruction quality: 00 - Perfect, 01 - Good, 10 - Failed and 11 - Ambiguous. Both H/T (Figure 4.23a) and E/B/O (Figure 4.23b) fields indicated Perfect reconstruction, meaning that all the six HGCROC halves always conveyed the same information in the corresponding fields. This gives confidence that this aspect of the HGCROC $\rightarrow$ ECON-D interface works in an optimal manner, as expected.

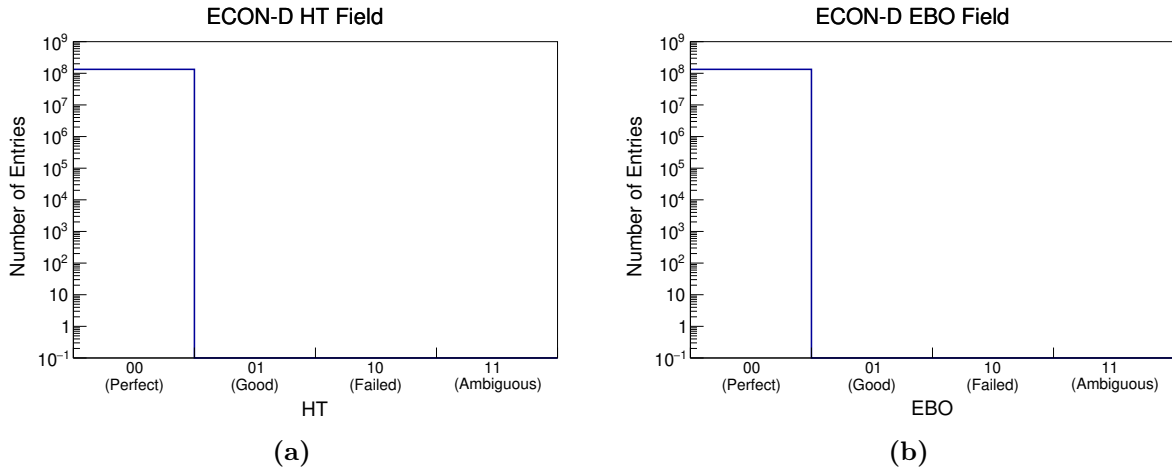


Figure 4.23: Distribution of (a) **H/T** and (b) **E/B/O** fields in ECON-D packet Headers. More details can be found in text.

The ECON-D Sub-packet Headers

Each ECON-D packet has the Stat field in each of the six Sub-packet Headers. For simplicity, only the results of one out of six Stat fields (e.g. for ECON-D input e-Link 0 or eRx 0) are presented here (Figure 4.24). Each Sub-packet corresponds to one HGCROC half (e.g. one ECON-D input e-Link). The ECON-D is configured to transmit the result

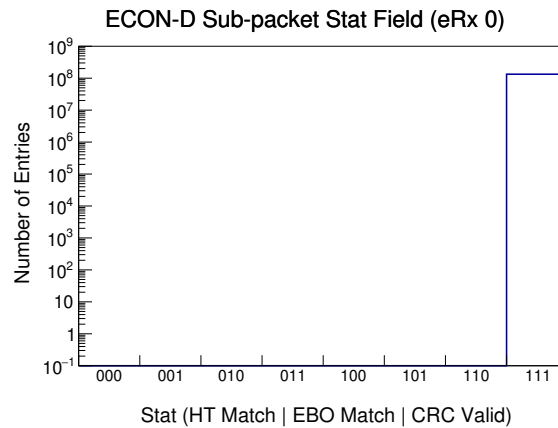


Figure 4.24: Distribution of the **Stat** field in ECON-D Sub-packet Headers corresponding to one input ECON-D e-Link (eRx 0) or one HGCROC half. More details can be found in text.

of the vertical reconstruction in its H/T and E/B/O fields. Additionally, for every event, the ECON-D performs a CRC check for each HGCROC half. Even though the vertical reconstruction of the H/T and E/B/O fields may have been successful, it is possible that the resulting transmitted values do not agree with H/T or E/B/O values on individual e-Links coming from the HGCROC halves. This information, together with the result of the result of the CRC check is transmitted in the 3-bit Stat field for every Sub-packet. The most significant bit of the Stat field indicates whether the vertical reconstruction result transmitted by the ECON-D matches the H/T patterns received from the corresponding HGCROC half. Similarly, the second bit indicates if the received E/B/O values match the transmitted ones. Finally, the least significant bit of the Stat field indicates if the CRC check of the corresponding HGCROC half was valid. The ECON-D indicated no issues regarding the H/T match, E/B/O match and the CRC validity in all the acquired events, which gives confidence that the HGCROC and the ECON-D work as expected regarding these aspects.

Each ECON-D packet has the Ham and F fields in each of the six Sub-packet Headers. The Ham field is the 3-bit result of the HGROC's internal hamming check and is simply forwarded through the ECON-D. The most significant bit covers the Header of the HGCROC half. The next leading bit covers the first quarter of channels (e.g. from CM to CH17). The least significant bit covers the second quarter of channels (e.g. from Calib to CH35). For each of the three bits, a logical "1" and a "0" indicate the presence and the absence of a Hamming Error, respectively. The F flag indicates the format of the ECON-D sub-packet, which can be either Regular (logical "1") or Empty (logical "0"). The Regular sub-packet format consists of a two-word header, data for at least one channel (which is expected given the HGCROC and ECON-D configuration) and a 37-bit channel map. The Empty sub-packets consist only of a single-word header and no channel data nor an address map. For all acquired events, no Hamming errors were observed (Figure 4.25a) and all Sub-packets had the Regular (Figure 4.25b) format. This gives confidence that the HGCROC and the ECON-D work as expected, regarding these aspects.

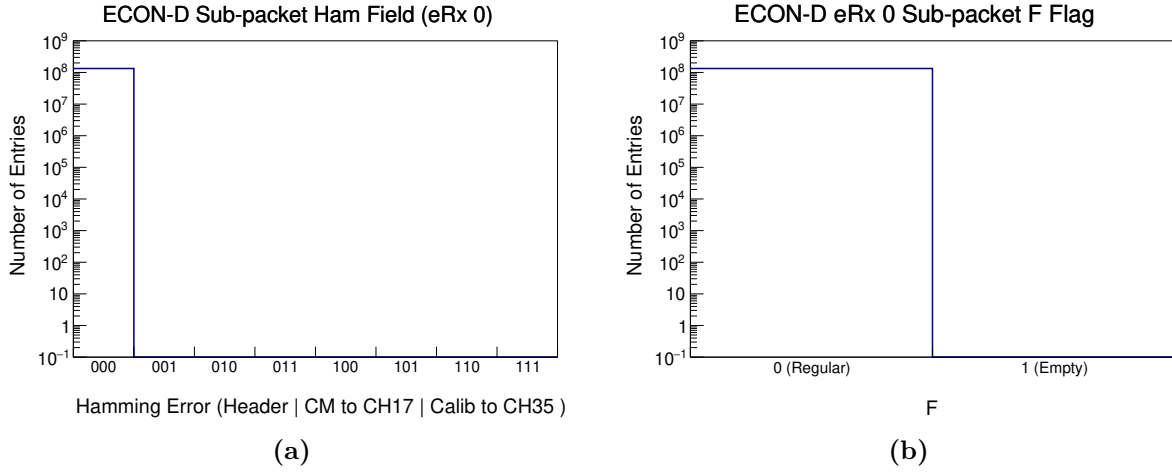


Figure 4.25: Distribution of (a) **Stat** and (b) **Ham** fields in ECON-D Sub-packet Headers corresponding to one input ECON-D e-Link (eRx 0) or one HGCROC half. More details can be found in text.

The L1A Timestamps

Each of the major components in the vertical chain (HGCROC, ECON-D, DAQ BE and the BE SLink readout firmware) have their own notion for the current event, bunch and orbit counters. It is the responsibility of the user to synchronise the different components via fast control commands (ECR, BCR, OCR) and by fine-tuning certain parameters in these components via slow-control (e.g. BCR reset value) so that they agree on the L1A timestamp of the current event. The HGCAL electronics system should ensure that being possible. However, it turned out that the used version of the HGCROC has two design flaws which do not completely allow that. The first issue will not be discussed in detail here. However, for the purpose of this analysis, it is important to note that to compensate for this problem, the HGCROC was configured to have its internal BX counter consistently four BXs ahead of the SLink BX counter. Second, the HGCROC orbit counter does not roll to the next value when the HGCROC BX counter transitions from 3564 to 1, but it does so one BX later. These have been communicated to the HGCROC design team, and a new version of the ASIC is expected not to suffer from the same issues. Ideally, the relative values for a given BX of the bunch and orbit counters of the HGCROC, ECON-D, DAQ BE and the BE SLink readout firmware should all be the same (Figure 4.26).

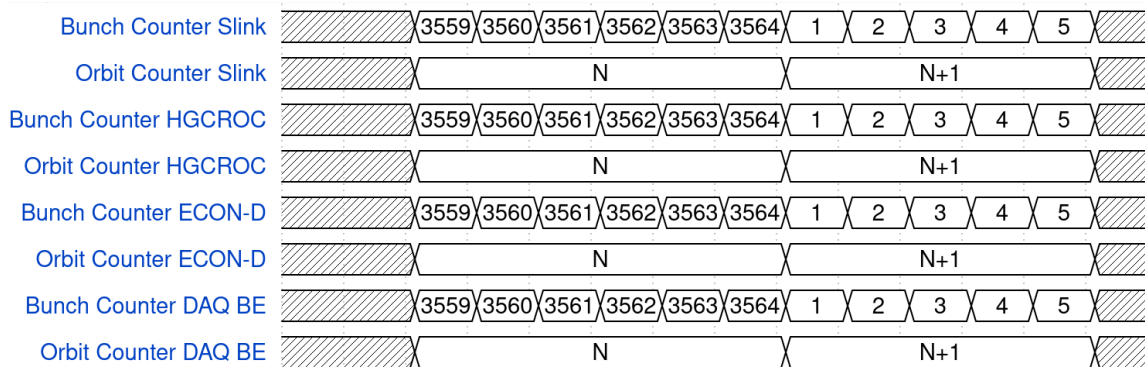


Figure 4.26: A diagram showing the ideal case of relative bunch and orbit counter values for the SLink, HGCROC, ECON-D and the DAQ BE for a given BX. More details can be found in text.

Because of the two HGCROC issues, it was decided to try to compensate for these effects (Figure 4.27). Regarding the first issue, the DAQ BE and the ECON-D were configured to offset their internal bunch counters to match the HGCROC bunch counter. Nothing could be done to address the second issue. Consequently, several synchronisation issues are bound to systematically occur in this analysis: the HGCROC, ECON-D and the DAQ BE bunch counters will be four BXs ahead of the SLink bunch counter; the HGCROC orbit counter will disagree with the DAQ BE and the ECON-D when the SLink bunch counter BC_{SLink} is $BC_{SLink} = 3561$; the DAQ BE orbit counter and SLink orbit counter will disagree when $BC_{SLink} \in \{3561, 3562, 3563, 3564\}$; and the HGCROC orbit counter will disagree with the SLink orbit counter at $BC_{SLink} \in \{3562, 3563, 3564\}$. While these are unacceptable for the final system operation, the rest of this analysis will treat them otherwise and try to spot any additional issues.

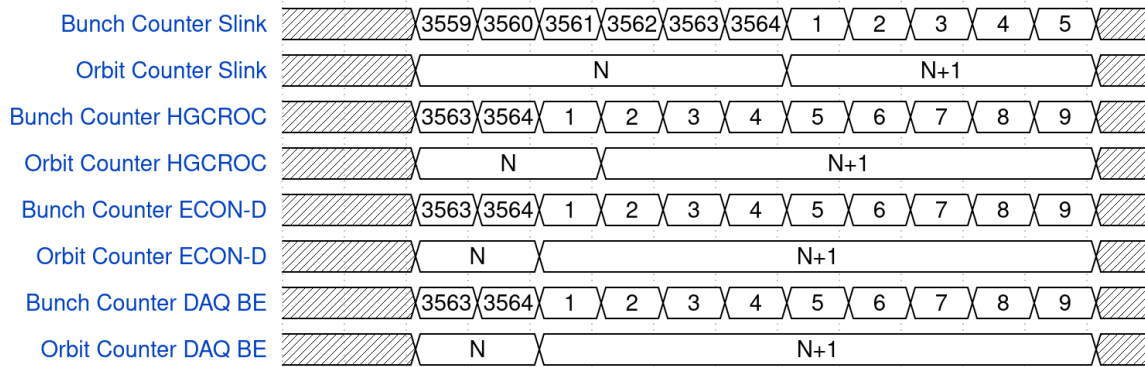


Figure 4.27: A diagram showing the used relative bunch and orbit counter values for the SLink, HGCROC, ECON-D and the DAQ BE for a given BX. More details can be found in text.

The HGCROCs and the ECON-D

Each ECON-D packet has the M Flag in its Header. This indicates if the event, bunch and orbit numbers transmitted by the ECON-D match the same set of ECON-D internal values. Given that the ECON-D is configured to transmit the vertically reconstructed values from the HGCROCs, the M flag specifies if the ECON-D event, bunch and orbit numbers match with the HGCROC ones (Figure 4.28a). For the vast majority of the acquired events, the M flag had the value of a logical "1", meaning that the ECON-D and the HGCROC agree with one another regarding these L1A timestamps. However, for a minority of events the M flag had the value of a logical "0", indicating otherwise. Explicitly, there were 37 253 events without a valid M flag. Having a closer look at the M Flag as a function of the SLink bunch counter BC_{SLink} (Figure 4.28b), showed that this mismatch does not appear at random times but that it appears systematically at $BC_{SLink} = 3561$, as previously anticipated. The fact that the ECON-D is configured to transmit the L1A timestamps taken from the HGCROCs should be kept in mind in further text.

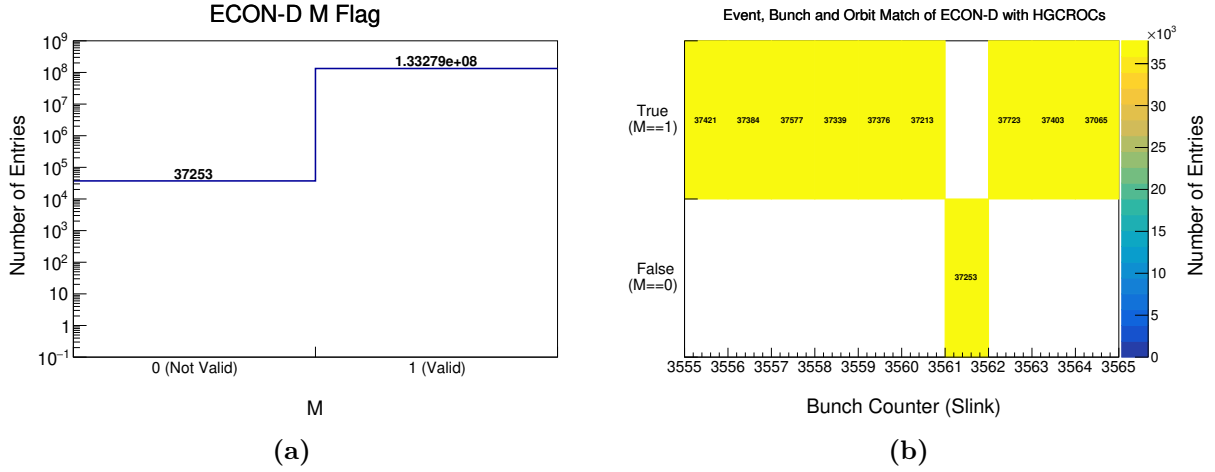


Figure 4.28: Distribution of (a) the M flag in ECON-D packet Headers and (b) the ECON-D M flag values as a function of the SLink bunch counter values, focused on the part of the orbit where M was 0. More details can be found in text.

The ECON-D and the DAQ BE

Plotting the ECON-D event counter values as a function of the DAQ BE event counter values shows the expected linear dependency with no offset on the y-axis in all acquired events (Figure 4.29a). Similar observation can be made for the ECON-D and the DAQ BE bunch counter values (Figure 4.29b), as expected.

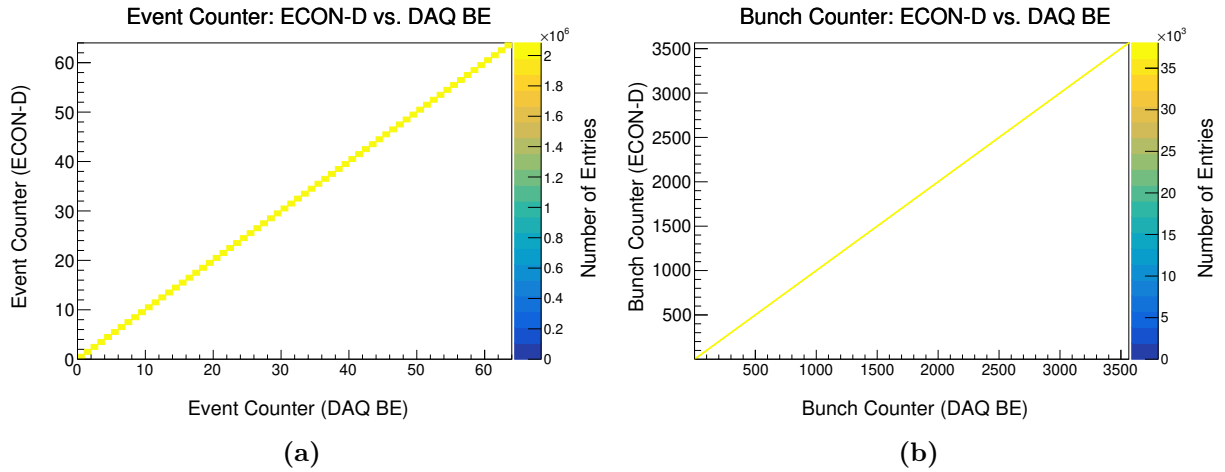


Figure 4.29: Distribution of (a) the ECON-D event counter as a function of the DAQ BE event counter and (b) the ECON-D bunch counter as a function of the DAQ BE bunch counter. More details can be found in text.

Plotting the ECON-D orbit counter values as a function of the DAQ BE orbit counter values shows the expected linear dependency with no offset on the y-axis in the vast majority of acquired events (Figure 4.30a). However, for a small minority of events the ECON-D orbit counter is offset by -1 from the DAQ BE orbit counter. Explicitly, summing the total number of outlier entries in the plot gives 37 253. Additionally, plotting the orbit counter match between the ECON-D and the DAQ BE as a function of the SLink bunch counter BC_{SLink} (Figure 4.30b), showed that this mismatch does not appear at random times but that it appears systematically at $BC_{SLink} = 3561$. Both of these plots

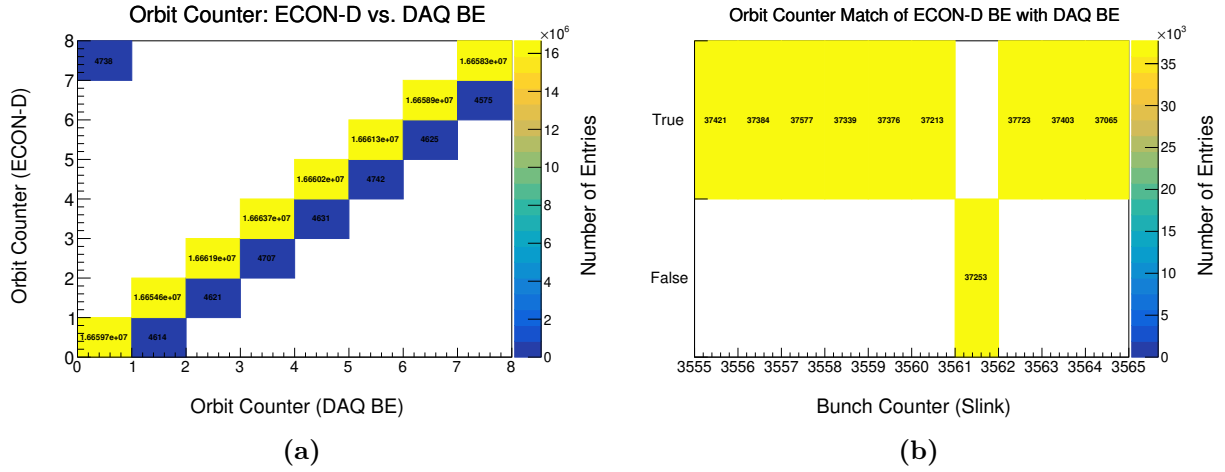


Figure 4.30: Distribution of (a) the ECON-D orbit counter as a function of the DAQ BE orbit counter and (b) the results of the ECON-D and DAQ BE orbit counter match comparison as a function of the SLink bunch counter values, focused on the part of the orbit where mismatches were observed. More details can be found in text.

are consistent with the previously discussed ECON-D M flag results. There are no issues introduced by the ECON-D itself. The ECON-D transmits the values coming from the HGCROC, hence the HGCROC errors propagate through the ECON-D.

The DAQ BE and the SLink

Plotting the DAQ BE event counter values as a function of the SLink event counter values shows the expected linear dependency with no offset on the y-axis in all acquired events (Figure 4.31a). A similar observation can be made for the DAQ BE and the SLink bunch counter values, although the DAQ BE is always ahead by an offset of +4 from the SLink (Figure 4.31b), as previously anticipated.

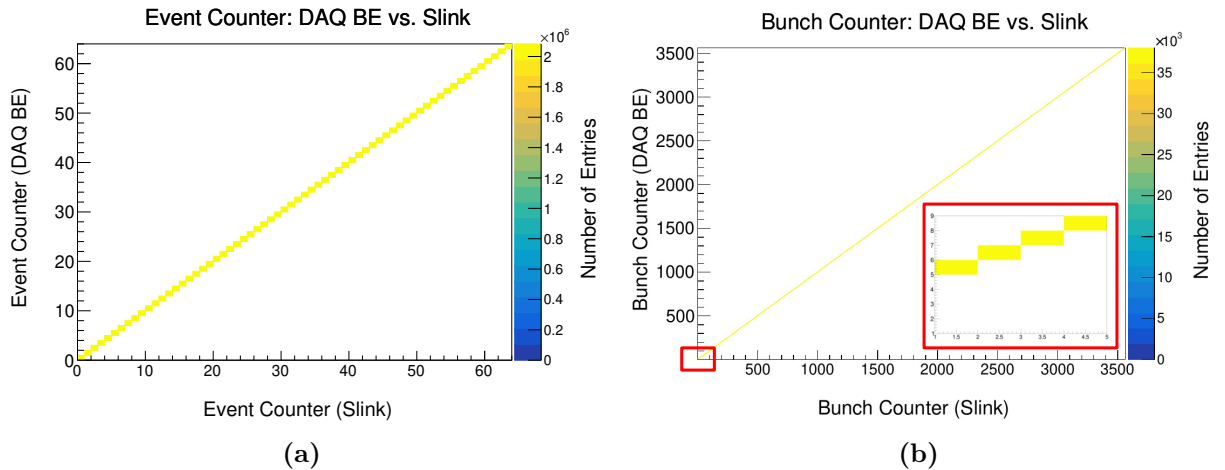


Figure 4.31: Distribution of (a) the DAQ BE event counter as a function of the SLink event counter and (b) the DAQ BE bunch counter as a function of the SLink bunch counter. More details can be found in text.

Plotting the DAQ BE orbit counter values as a function of the SLink orbit counter values shows the expected linear dependency with no offset on the y-axis in the vast majority

of acquired events (Figure 4.32a). However, for a small minority of events the DAQ BE orbit counter is offset by +1 from the SLink orbit counter. Explicitly, summing the total number of outlier entries in the plot gives 149 444. Additionally, plotting the orbit counter match between the DAQ BE and the SLink as a function of the SLink bunch counter BC_{SLink} (Figure 4.32b), reveals the same total number of outliers as the previous plot (149 444) and that this mismatch does not appear at random times but systematically for $BC_{\text{SLink}} \in \{3561, 3562, 3563, 3564\}$, as previously anticipated.

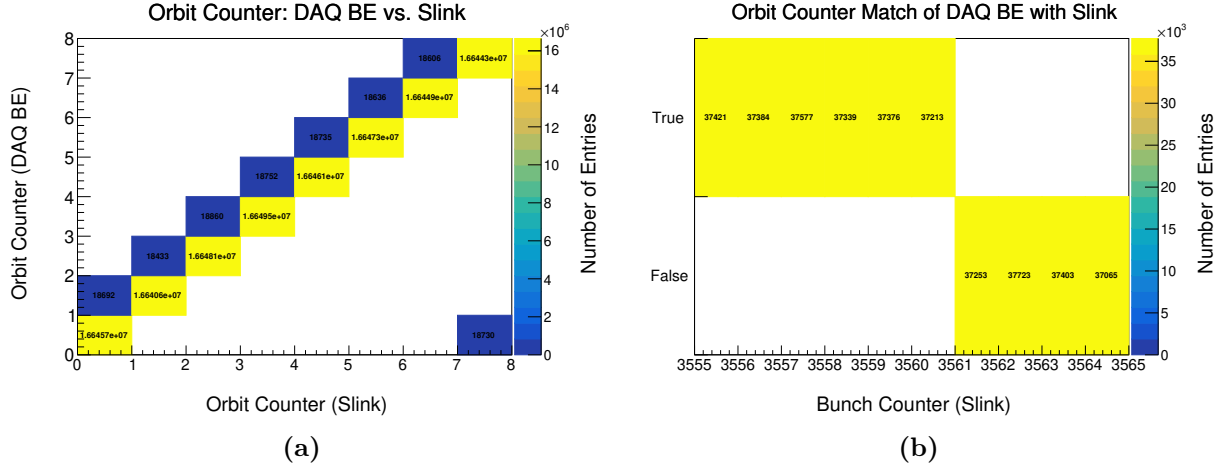


Figure 4.32: Distribution of (a) the DAQ BE orbit counter as a function of the SLink orbit counter and (b) the results of the DAQ BE and SLink orbit counter match comparison as a function of the SLink bunch counter values, focused on the part of the orbit where mismatches were observed. More details can be found in text.

The ECON-D and the SLink

Plotting the ECON-D event counter values as a function of the SLink event counter values shows the expected linear dependency with no offset on the y-axis in all acquired events (Figure 4.33a). A similar observation can be made for the ECON-D and the SLink bunch counter values, although the ECON-D is always ahead by an offset of +4 from the SLink (Figure 4.33b), as previously anticipated.

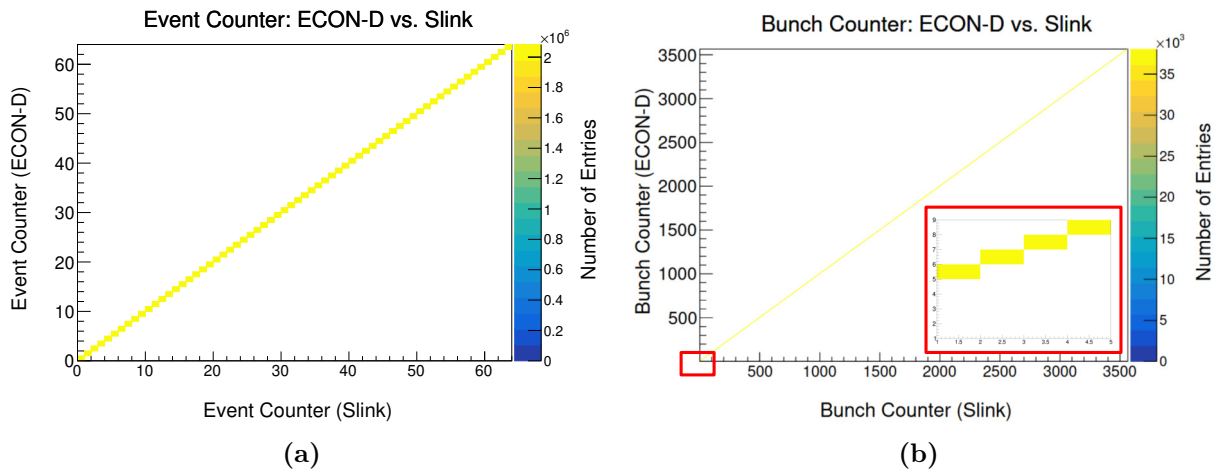


Figure 4.33: Distribution of (a) the ECON-D event counter as a function of the SLink event counter and (b) the ECON-D bunch counter as a function of the SLink bunch counter. More details can be found in text.

Plotting the ECON-D orbit counter values as a function of the SLink orbit counter values shows the expected linear dependency with no offset on the y-axis in the vast majority of acquired events (Figure 4.34a). However, for a small minority of events the ECON-D orbit counter is offset by +1 from the SLink orbit counter. Explicitly, summing the total number of outlier entries in the plot gives 112 191. Additionally, plotting the orbit counter match between the ECON-D and the SLink as a function of the SLink bunch counter BC_{SLink} (Figure 4.34b), reveals the same total number of outliers as the previous plot (112 191) and that this mismatch does not appear at random times but systematically for $BC_{\text{SLink}} \in \{3562, 3563, 3564\}$, as previously anticipated.

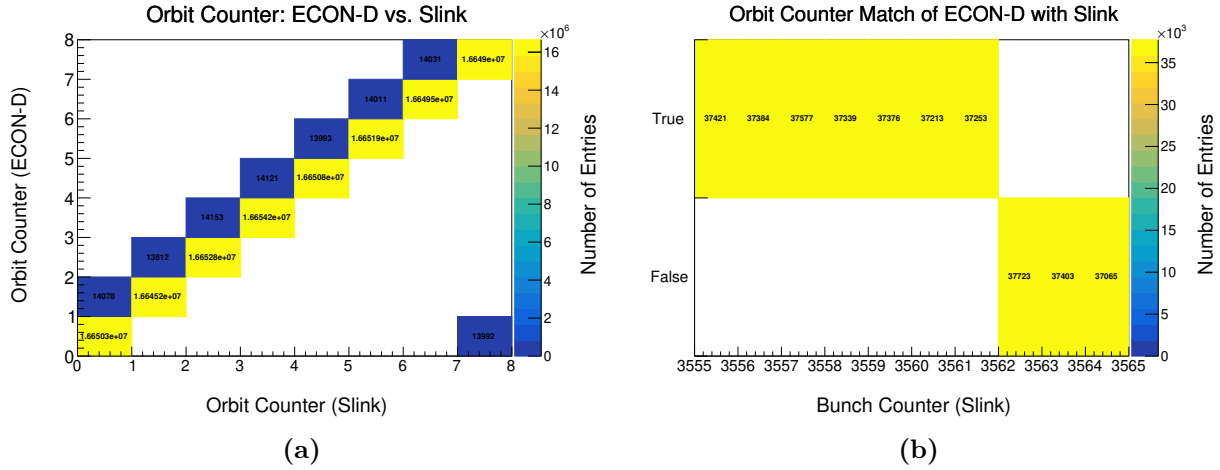


Figure 4.34: Distribution of (a) the ECON-D orbit counter as a function of the SLink orbit counter and (b) the results of the ECON-D and SLink orbit counter match comparison as a function of the SLink bunch counter values, focused on the part of the orbit where mismatches were observed. More details can be found in text.

Closing the DAQ System-wide Digital Consistency Analysis

With the exception of the two digital design flaws found in the HGCROC ASICs (expected to change in HGCROCv3B), the tested digital aspects of all components in the vertical chain work as expected when operated at the nominal average L1A rate of 750 kHz in ambient conditions. This was the case for the total 133 315 819 DAQ packets analysed during this test. This allowed to set an upper limit [50] with a 95% CL on the DAQ digital consistency error rate $DCER_{\text{DAQ}}$ (Equation 4.4).

$$DCER_{\text{DAQ}} < \frac{3.00}{133\,315\,819} \approx 2 \times 10^{-8}, \text{ at 95\% CL} \quad (4.4)$$

4.5 Results Summary

This chapter provided deeper insights into the operation of the vertical test system, focusing on using this system in a beam test, identifying and mitigating the VTRx+ saturation issue, setting BER upper limits, and evaluating the degree of system-wide digital consistency.

The selection of beam test results (module histograms, signal amplitude over time, comparison of internal and external trigger, ECON-D channel retention) indicates that this vertical readout system architecture is indeed viable for the task of acquiring real physics

data. As a result, both the TPG and DAQ sub-systems were tested in multiple operational modes: BC and STC4E3M for the TPG, and Pass-through and Zero-suppression for the DAQ. Additionally, both external and self-triggering methods were employed. Various types of particle beams (pion, muon, and electron) were utilised, spanning a wide range of energies (from a MIP up to 200 GeV). This comprehensive testing enabled the evaluation of major measurements (including ADC, ToT, and ToA) and calibration procedures (such as pedestal subtraction and trimming) within the HGCROCs. Depending on the beam type and configuration settings, the L1A rate and data volume varied but remained below approximately $O(100 \text{ kHz})$ and $O(10 \text{ Gbit s}^{-1})$, respectively. These beam test activities also marked the first time that any CMS subdetector upgrade used such a long chain of components in the readout path, including a DTH-based readout.

The beam test activities raised suspicions about the signal integrity in the FE $\rightarrow$ BE interface. This was confirmed when the issue of VTRx+ saturation was found and later mitigated by reducing the output optical power of the Serenity BE Fireflies in the DAQ F/O down-link by 3 dB. Having communicated these findings to the VTRx+ and lpGBT teams lead to a comprehensive study of this effect. Consequently, an application note was produced to benefit all experiments using VTRx+ and Firefly components.

After mitigating the VTRx+ saturation issue, extended monitoring of TPG and DAQ data acquisition set BER upper limits to 4×10^{-17} and 5×10^{-16} (at 95 % CL), respectively. Consequently, this gave high confidence in the quality of the electrical and optical interfaces in the vertical system when operated in ambient conditions.

The BER results could not fully capture the subtle details regarding the digital consistency of different components in the vertical chain (e.g. indicated by different status and quality flags, agreement of L1A timestamps, etc.). Hence, an analysis was done which found the digital consistency error rates to be below 9×10^{-14} and 2×10^{-8} (at 95 % CL and an L1A rate of 750 kHz) for the TPG and DAQ devices, respectively.

Chapter 5

The HGCAL L1A Throttle for TCDS2

This chapter explores a limitation that arises from the size and operation of the HGCROC RAM 2 and its readout logic. Additionally, it considers the implication of this limitation on the CMS operation during HL-LHC. Furthermore, it proposes a solution by introducing an L1A throttling block to the TCDS2 firmware. Moreover, it presents a functional prototype of such a block, providing detailed evidence of its successful operation. Ultimately, it demonstrates that this block can exist in a real DAQ system (i.e. the HGCAL vertical test system) as a working proof-of-principle. The chapter is organised into three sections. Firstly, it presents the fundamental reasons behind the motivation for L1A throttling (Section 5.1). Secondly, it summarises the aspects surrounding the software model of the proposed L1A throttling block (Section 5.2). Lastly, it presents the corresponding firmware block candidate for TCDS2 (Section 5.3). The chapter also compares the major results from the different sections (e.g. firmware against the software, software against theoretical predictions, software behavioural predictions of the HGCROC against the real HGCROC ASIC behaviour, etc.). Although this chapter will not provide a more detailed discussion on this topic, it should be noted that this L1A throttling firmware block was successfully used in the TCDS2 emulator of the vertical readout system during the beam test activities previously outlined in Chapter 4.

5.1 The Motivation for L1A Throttling

In the case of an L1A signal arriving at the HGCROC (more details about the HGCROC can be found in Chapter 2), the DAQ data for the triggered event moves through two main interfaces before being streamed across HGCROC output e-Links (Figure 5.1). Firstly, the selected event (which is Hamming-encoded) is taken out of the Circular Buffer (RAM 1) into the L1 Triggered FIFO (RAM 2). Secondly, these data are then moved out of RAM 2 (Hamming decoding is also done here) and through the serialising block in the readout logic. This section is organised into four parts. Firstly, it gives a general overview of the purpose of the RAM 2 FIFO. Secondly, it describes a limitation arising from the HGCROC RAM 2 and its readout logic as well as its impact on CMS operation during HL-LHC. Thereby, it defines a useful measure of the corresponding negative effects called *deadtime*. Thirdly, it derives a theoretical minimum for *deadtime*. This hard limit will then serve as a useful cross-check for software simulations (Section 5.2) and hardware measurements (Section 5.3) of *deadtime* as a function of the average L1A frequency.

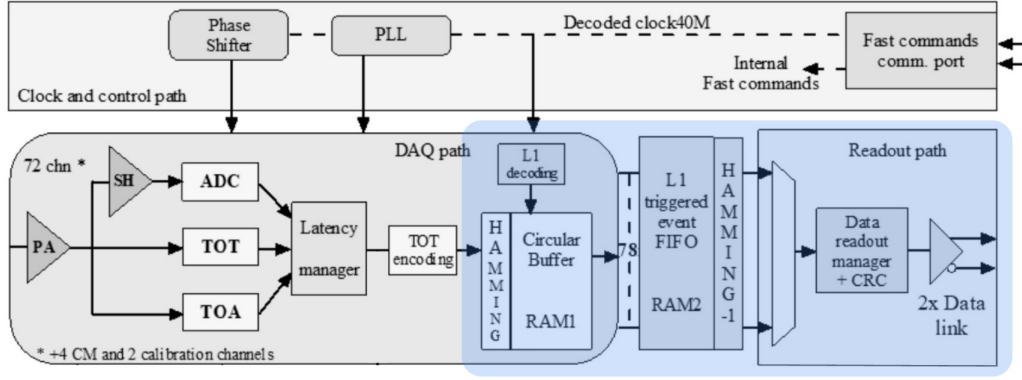


Figure 5.1: A block diagram showing the major parts of the HGCROC logic related to DAQ and control paths [36]. In focus (highlighted in blue) are the two HGCROC buffers (RAM 1 and RAM 2) and its readout logic. More details can be found in text.

5.1.1 The purpose of the HGCROC RAM 2 FIFO

The RAM 2 FIFO is used to de-randomise the DAQ event readout in the case of the HGCROC receiving bursts of L1A signals. To better understand this point, one should take into consideration that it takes 41 BX to stream any triggered event data out of the HGCROC. Having in mind the relationship between the average L1A frequency $\langle f_{L1A} \rangle$ and the average consecutive L1A BX distance $\langle D_{L1A} \rangle$ (Equation 5.1)

$$\langle D_{L1A} \rangle = \frac{40 \text{ MHz}}{\langle f_{L1A} \rangle} \quad (5.1)$$

and the HL-LHC nominal average L1A frequency of $\langle f_{L1A} \rangle = 750 \text{ kHz}$, one can find that the corresponding average consecutive L1A distance is $\langle D_{L1A} \rangle = 53.33$. This means that the request for the next triggered event will, on average, come after the previous event has already been streamed out of the HGCROC (i.e. $53.33 \text{ BX} > 41 \text{ BX}$). However, the consecutive L1A distances in real operation will take on random values, following a discretised exponential distribution (Figure 5.2). Consequently, there will be times when

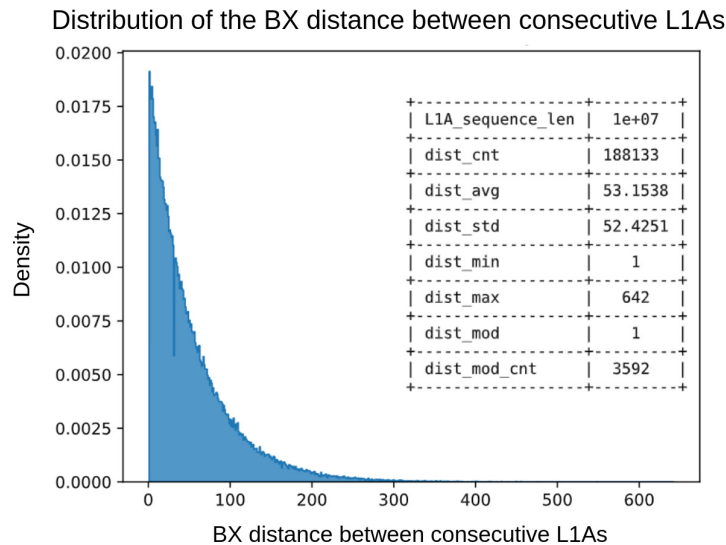


Figure 5.2: Simulated distribution of the consecutive L1A distances when $\langle f_{L1A} \rangle = 750 \text{ kHz}$.

the next L1A arrives at the HGCROC while the event corresponding to the previous L1A is still being streamed out. To be able to accept more L1As in parallel to streaming out the current event, the HGCROC implements the RAM 2 FIFO. Specifically, the RAM 2 FIFO and its readout logic memory capacities are designed such that together they can contain no more than 33 DAQ events at any given time. Any possible number (i.e. ≤ 33) of random-distanced DAQ events loaded into the RAM 2 FIFO would be streamed out of the HGCROC as if they originated from equal-distanced¹ L1As (i.e. $D_{L1A} = 41$).

5.1.2 The RAM 2 Limitation and CMS Deadtime

An important system limitation for CMS arises from the combined capacity of the RAM 2 FIFO and its readout logic being 33 DAQ events: *If the CMS L1T is to request an event that cannot fit into RAM 2 due to its current occupancy, that event will not be saved and will never be sent out from the HGCROC.* Hence, in such a scenario, the entire HGCAL would appear as a volume of the CMS detector that did not provide any data about the problematic event². To effectively qualify the rate of these undesired events, one can use a quantity often referred to as deadtime. This is given by the ratio of the number of BXs when a potential L1A signal would be ignored N_{DeadBX} and the total number of BXs N_{BX} in a given measurement (Equation 5.2).

$$\text{Deadtime} = \frac{N_{\text{DeadBX}}}{N_{\text{BX}}} \quad (5.2)$$

An important goal of this chapter is to estimate the average deadtime as a function of $\langle f_{L1A} \rangle$ and specifically in the vicinity of $\langle f_{L1A} \rangle = 750 \text{ kHz}$. This will, of course, depend on the implementation details of RAM 2 and its readout logic as well as the random nature of D_{L1A} . The first precise estimates of this relationship considering both of those dependencies will be presented in Section 5.2. Nevertheless, it is already possible to derive a theoretical minimum for $\text{Deadtime}(\langle f_{L1A} \rangle)$ which will serve as a useful cross-check of the precise estimates later on.

5.1.3 The Deadtime Theoretical Minimum

To derive the deadtime theoretical minimum, $\text{Deadtime}_{\text{min}}$, as a function of $\langle f_{L1A} \rangle$, one can abstract the random nature of D_{L1A} and the existence of RAM 2 and focus only on the average consecutive L1A distance $\langle D_{L1A} \rangle$ and the fact that it takes the HGCROC 41 BX to stream out a single event. In this simplified exercise, the number of BXs when a potential L1A signal would be ignored N_{DeadBX} can be expressed as $N_{\text{DeadBX}} = (41 - \langle D_{L1A} \rangle) \times N_{L1A}$, where N_{L1A} is the total number of L1As. From here, one can find $\text{Deadtime}_{\text{min}}(\langle f_{L1A} \rangle)$

¹That is why it is commonly said that RAM 2 is used for DAQ event readout de-randomisation.

²All HGCROC ASICs in the HGCAL FE receive the same L1A stream, hence they all have the same RAM 2 occupancy at any given moment. Consequently, they altogether ignore problematic L1As.

(Equation 5.3).

$$\begin{aligned}
 N_{\text{DeadBX}} &= (41 - \langle D_{\text{L1A}} \rangle) \times N_{\text{L1A}} , \text{ divide by } 41 \times N_{\text{L1A}} \\
 \text{Deadtime}_{\text{min}} &= 1 - \frac{\langle D_{\text{L1A}} \rangle}{41} , \text{ substitute } \langle D_{\text{L1A}} \rangle = \frac{40 \text{ MHz}}{\langle f_{\text{L1A}} \rangle} \\
 &= 1 - \frac{40 \text{ MHz}}{41 \times \langle f_{\text{L1A}} \rangle}
 \end{aligned} \tag{5.3}$$

The deadtime theoretical minimum is 0 when $\langle f_{\text{L1A}} \rangle \leq \frac{40 \text{ MHz}}{41} \approx 975.6 \text{ kHz}$ and asymptotically approaches the value 1 when $\langle f_{\text{L1A}} \rangle \rightarrow +\infty$ (Figure 5.3). The former says that

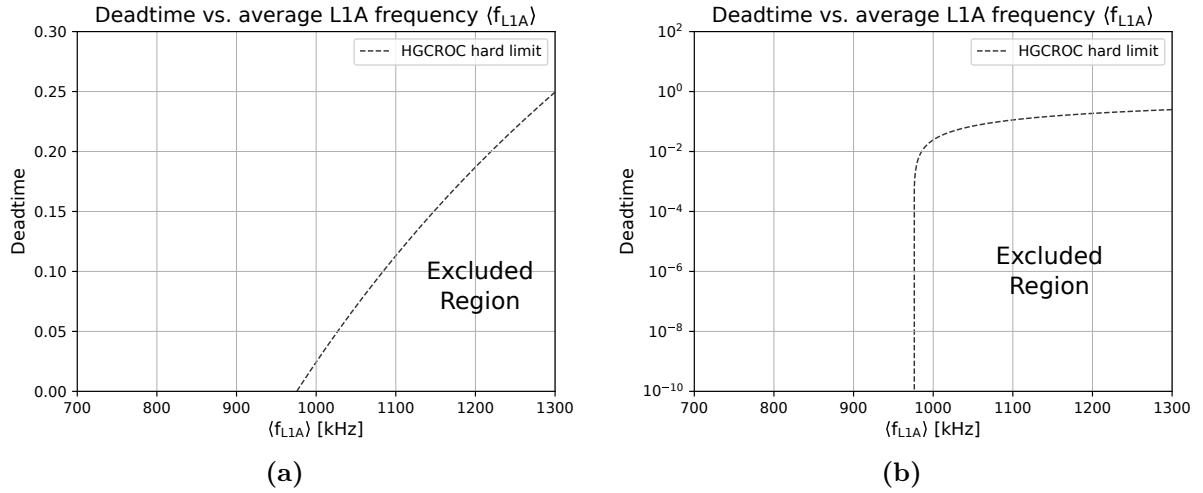


Figure 5.3: The theoretical minimum (HGCROC hard limit) for Deadtime as a function of $\langle f_{\text{L1A}} \rangle$ derived from the limitation of the HGCROC output bandwidth for y-axis in (a) linear and (b) logarithmic scale. More details can be found in text.

the system is not poorly designed since $\langle f_{\text{L1A}} \rangle = 975.6 \text{ kHz}$ is well above the nominal $\langle f_{\text{L1A}} \rangle = 750 \text{ kHz}$. However, stepping back from the previous theoretical abstractions, the random nature of D_{L1A} will inevitably lead to L1A sequences with consecutive BX distances smaller than 41. In turn, the HGCROC will de-randomise them by temporarily storing the corresponding events in the RAM 2 FIFO. However, there will not always be enough room for some of these events due to the RAM 2 33-event occupancy limit. Although the theoretical minimum ensures that the actual average deadtime will not fall below it for any given $\langle f_{\text{L1A}} \rangle$, the finite capacity of the RAM 2 buffer will add a non-zero contribution to the real deadtime in addition to the theoretical minimum.

5.1.4 Proposing L1A Throttling for TCDS2

Generally, other CMS subdetectors may have similar buffering limitations to the one of HGCal. To protect the CMS event reconstruction, one can introduce the ability to deactivate (commonly referred to as “throttling”) the L1As that would otherwise be problematic for any downstream device. This throttling logic should be implemented in TCDS2 (Figure 5.4), because its output is the common point from which the same L1A streams are taken towards the different CMS subdetectors.

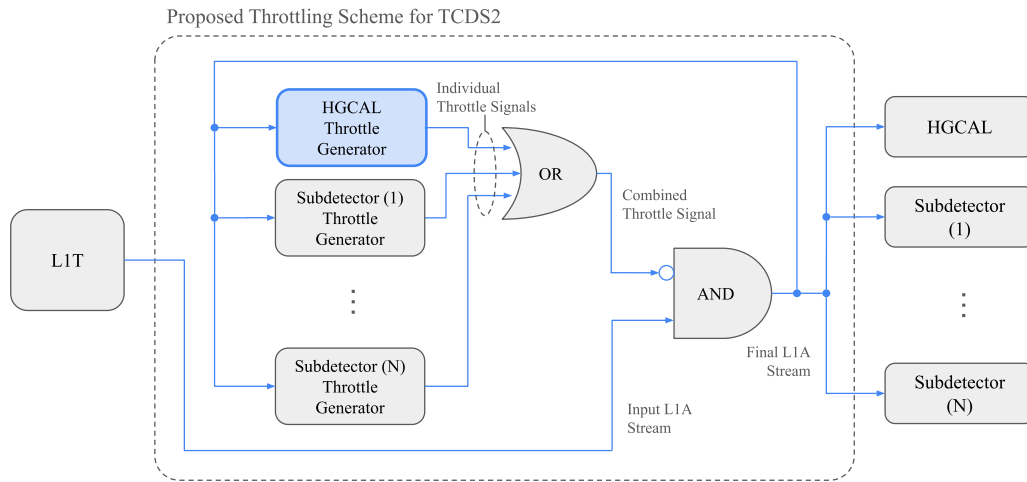


Figure 5.4: A conceptual block diagram in support of the proposed TCDS2 L1A throttling scheme. More details can be found in text.

Knowing their buffering limitations, the CMS subdetectors can consider implementing a throttle generator block to be instantiated in the TCDS2 firmware. These blocks would: spy on the output L1A stream (which is also the one brought to the downstream devices); implement a BX-level emulation of the first-to-overflow buffers (e.g. for HGCAL that is RAM 2 and its readout logic); and raise a throttle signal (with the active logical level being "1") requesting to deactivate the next L1A which would otherwise cause a buffer overflow. The results of each of the throttle generators would be brought to a logic OR gate, producing a combined throttle signal. That way if at least one of the subdetectors requests a throttle, that request would be visible in the combined throttle decision. The combined throttle signal and the input L1A stream coming from the L1T would then be brought to an AND gate with an inverted input of the former. This would ensure that if the combined throttle is inactive (i.e. logical "0") the AND gate would simply pass through the input L1A stream as is, or deactivate it if otherwise. The output of the AND gate would be the final L1A sent to the subdetectors but also fed back to the inputs of the throttle generators to decide about throttling the next BX, etc.

One can also conceptualise the BX-level behaviour of the relevant signals through functional time diagrams (Figure 5.5). It is assumed that the TCDS2 internally has faster derivations of the 40 MHz clock that the throttle generators can use when preparing the throttle decision for the next BX. The duration of the throttle signal relative to the 40 MHz clock period is chosen arbitrarily in the drawn waveform. However, the throttle signal and in turn, the final L1A signal must be set up well in advance of the next rising edge of the 40 MHz clock. This means that the throttle blocks must be fast in generating their decisions with a sub-BX (< 25 ns) latency requirement. Moreover, to not interfere with the rest of TCDS2 firmware, these blocks must be light in terms of FPGA resources. Although a general scheme for throttling was hereby shown, the rest of this chapter further explores only an HGCAL-specific throttle generator.

Before moving on, it is worthwhile noting that there is an abundance of material covering the general theory about queues, properties of queuing systems and issues arising due to their congestion (e.g. available in [66]). Moreover, CERN motivated a general study of "Dead-time losses in a buffered data recording system", back in 1972 [67]. Furthermore, the concept of throttling L1As in the BE as a means of preventing FE ASIC buffer overflow is

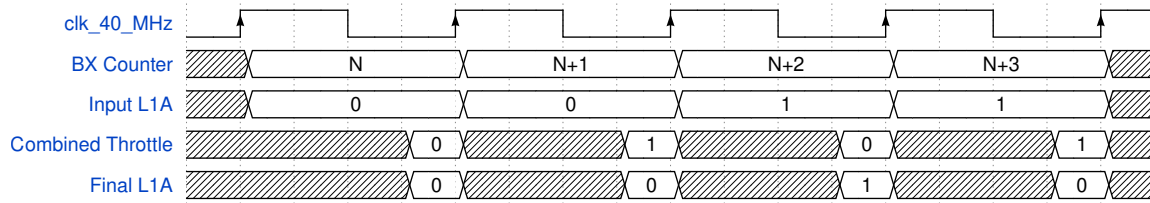


Figure 5.5: A functional time diagram of the signal waveforms in support of the proposed TCDS2 L1A throttling scheme. More details can be found in text.

also not alien to the CMS experiment. Namely, in 2002, the Tracker designed a custom BE APVE board [68] which emulated the buffer occupancy of the APV25 ASIC [69] on an FPGA. This board resided close to the Trigger Control System (TCS) crate to which it was then able to report the Tracker FE buffer occupancy status in a timely manner. Eventually, this emulator was incorporated into the Timing and Control Distribution System (TCDS), when the APVE board was abandoned [70]. However, this chapter focuses on the throttling requirements specific to the implementation details of the HGCROC ASIC and the TCDS2.

5.2 The Software Model

The software model of the HGICAL L1A Throttle block is only a part of a larger software ecosystem developed around it. When exposed to an arbitrary sequence of the L1A stream, the software model provides a BX-level emulation of the: HGCROC RAM 2 and its readout logic; throttle decision-making; and creation of the final L1A stream. The surrounding software also provides utilities to generate various L1A sequences (periodic, burst and random) of arbitrary BX length; stimulate the software model or the real HGCROC ASIC with arbitrary L1A sequences; store outputs; compare the behaviour of the software model and the real ASIC; and conduct deadtime simulations. Using this software ecosystem allowed extensive testing of the HGCROC ASIC until the exact BX-level behaviour of the RAM 2 and its readout logic were fully understood. The behavioural evidence gathered in these hardware tests ultimately got absorbed into the part of the software model responsible for emulating the RAM 2 and its readout logic. It is worthwhile noting that a detailed description regarding the internal operation of these HGCROC internal blocks was not available, hence it had to be reverse-engineered.

This section is organised into two parts. Firstly, it presents a simulation flow chart when the software model is stimulated with an arbitrary L1A sequence, demonstrating the key behavioural aspects being emulated. Secondly, it presents the most significant results coming out of the implemented software model.

5.2.1 The Simulation Flowchart

A single iteration of the HGICAL L1A throttle simulation essentially receives the final L1A stream value from the previous BX, updates various counters (and status flags) to emulate the corresponding status in a real HGCROC ASIC and produces a throttle decision for the current BX. This signal is then used to potentially deactivate the L1A candidate value of the current BX and produce the final L1A value. Bringing sequences of L1A candidates to

the simulation results in sequences of counter (and status flag) values, throttle decisions, and final L1A values being output by the simulation. The simulation can be presented in the form of a flow chart (Figure 5.6). Each iteration of the loop corresponds to processing a single BX. Black arrows in the flowchart represent the chronological order of steps, whereas brown arrows represent dependencies between steps. An example of the former is that after the final L1A stream value from the previous BX is received, the simulation bunch counter is incremented. An example of the latter is that updating the HGCROC event counter depends on the previous final L1A value and the previous value of the event counter itself. The simulation, of course, begins by resetting various simulation variables to initial values. In the following, each of the major simulation steps and their dependencies will be discussed in more detail.

The iteration starts by receiving the final L1A value from the previous BX, followed by the simulation bunch counter BC being incremented. Next, the HGCROC 6-bit event counter EC is updated. If the previous BX had an L1A this means incrementing the EC, otherwise holding on to its previous value. Its reset value is defined to be 1 after which it is incremented according to $EC = (EC + 1) \bmod 64$. It is critical to emulate the EC properly because it is the key parameter that can indicate whether the HGCROC RAM 2 buffer overflowed. Specifically, if the HGCROC receives an L1A but cannot fit the corresponding event into RAM 2, it will still increment its EC. In other words, when consecutive events going out of the HGCROC have their EC being incremented in steps greater than 1, that indicates RAM 2 overflow ³.

The next four iteration steps decide whether the current state of the HGCROC allows draining RAM 2 as several conditions can prevent that, even at times when the FIFO is not empty ⁴. Specifically, whenever RAM 2 is empty, and then a new event is moved into it, there will be a 3 BX timeout before the first word of that new event starts being drained out of the FIFO. Moreover, whenever the 40th (CRC) or the 41st (mandatory IDLE) word of an event is being output by the HGCROC, the draining of the RAM 2 FIFO stops. Hence, these four steps include: updating the current value of a drain timeout counter due to the previously empty buffer; generating a 1-bit flag that prohibits/allows draining of the RAM 2 based on the timeout counter value; generating a 1-bit flag that prohibits/allows draining of the RAM 2 based on the ID of the output word; generating the final 1-bit global flag that combines the previous two and generates the final decision about allowing the RAM 2 to drain.

The global flag indicating paused draining of RAM 2, the occupancy of RAM 2 and the received value of the L1A stream for the previous BX, all contribute to defining the next state of the readout logic. Specifically, the next iteration step accordingly updates the core Finite State Machine (FSM) state to one of the four possible states: `FILL_N_DRAIN_N`, `FILL_N_DRAIN_Y`, `FILL_Y_DRAIN_N` and `FILL_Y_DRAIN_Y` (Y/N are short for YES/NO). This defines how to update the buffer occupancy in the following step. If the state indicates that the RAM 2 should be filled, the buffer will accept an additional 39 event words (excluding the CRC and the IDLE from the total of 41 words). Similarly, if the state indicates that the RAM 2 should be drained, the buffer will output the next event word out. The next step is to update the ID of the HGCROC output word. This can be either: 0 (IDLE due

³Except when the EC rolls over from 63 to 0, which is allowed given the $EC = (EC + 1) \bmod 64$ rule.

⁴Discovering these several rules was arguably the most difficult part of reverse-engineering the behavioural model of the HGCROC RAM 2 and its readout logic since it required exposing the HGCROC to specifically tailored sequences of L1As.

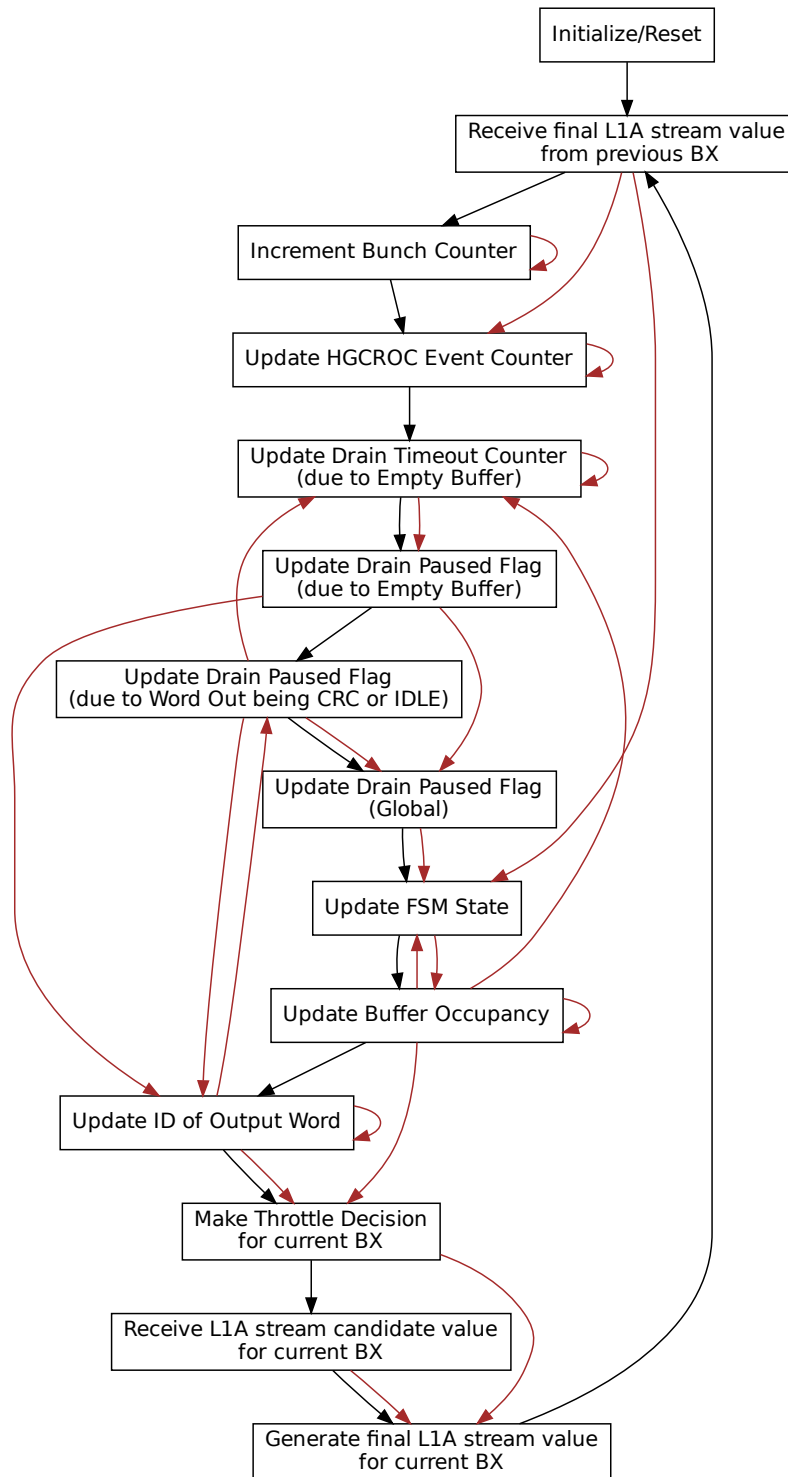


Figure 5.6: A flowchart of the software simulation. Each iteration of the loop corresponds to one BX being processed, starts by receiving the final L1A value from the previous BX and ends by optionally throttling the L1A candidate value of the current BX. Black arrows represent the chronological order of steps, whereas brown arrows represent dependencies between steps. More details can be found in text.

to Empty Buffer); 1-39 (event data); 40 (event CRC); or 41 (event mandatory IDLE).

Based on the updated buffer occupancy, the following simulation step is to make the throttle decision for the current BX. The throttle becomes active if: the current RAM 2 occupancy is greater than 32×39 words; or exactly 32×39 words and it is the CRC or the IDLE word that is being output by the HGCROC. In any other case, the throttle signal is made inactive.

After receiving the next L1A stream candidate value for the current BX, the simulation combines it with the prepared throttle decision and generates the final L1A stream value for the current BX. This completes one iteration of the simulation loop.

Describing the major steps of the simulation flowchart gave a reasonable level of detail of the HGCAL L1A throttle algorithm ⁵. For more details, the reader is invited to consult three examples (presented through tables in Appendix A) carefully chosen to cover all relevant cases and explicitly demonstrate the BX-level behaviour of this block.

5.2.2 The Most Significant Results

The two most notable results arising from the implemented software model presented here address: the agreement of the software model and the HGCROC ASIC behaviour; and the simulated deadtime as a function of the average L1A frequency.

Agreement with the HGCROC ASIC

The agreement between the implemented HGCAL L1A throttle software model and the HGCROC ASIC was tested before the vertical test system was upgraded for the beam test. Hence, the architecture of the vertical test system mostly resembled the one presented in Chapter 3 and not the one in Chapter 4. Most importantly, there was no ability to save large amounts of data to disk, and no ability to program L1A sequences longer than about 2000 BX at a time. An important architectural system detail is that the ECON Mezzanine PCB (with the ECON ASICs) was not available at the time. Instead, an FPGA emulator of the ECON algorithms was used in the system FE (running on a Xilinx ZCU102 board). This FPGA design had capture blocks (similar to the Serenity EMP capture blocks) which allowed saving the HGCROC data going into the ECON-D emulator in response to the 2000 BX long L1A sequence. Hence, the system architecture allowed to: load an arbitrary 2000 BX long sequence into the DAQ BE fast control firmware; fire it once; capture the corresponding output of the HGCROCs; and save it in the ZCU102 board controller.

There are several key aspects in the workflow used when testing the agreement of the software model and the HGCROC ASIC behaviour (Figure 5.7). Firstly, a 2000 BX long L1A sequence is generated in software. This sequence is then played through the software model which provides a BX-level emulation of the HGCROC's response to this sequence. For any event that is accepted into RAM 2, the software model output provides information about: the event counter EC; the bunch counter of the event BC; and the BC value when each one of the 41 HGCROC words is being streamed out. On the other hand, the same L1A sequence is sent to the real HGCROC ASIC, its response to it saved, and the same set of information extracted for any event output by the HGCROC. Finally, the software compares the EC and BC values coming from the real HGCROC and the software model,

⁵And consequently gave an understanding of the underlying logic in the HGCROC.

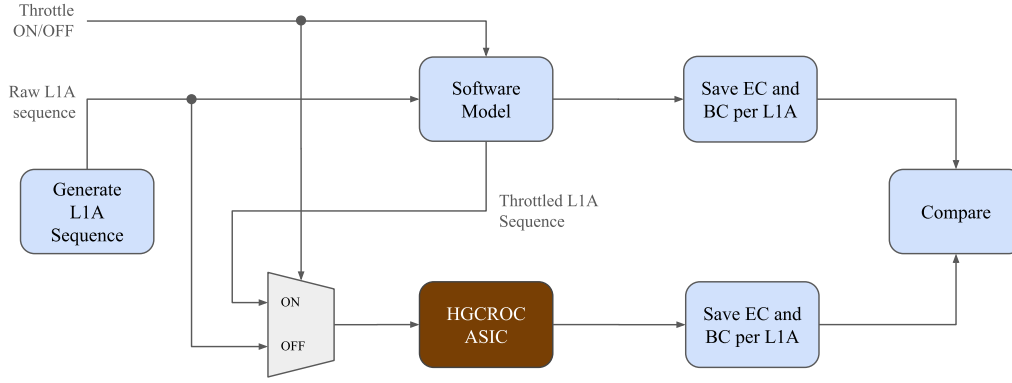


Figure 5.7: A block diagram illustrating the main workflow when the agreement between the software model of the HGCAL L1A throttle and the HGCROC ASIC were tested one against the other. More details can be found in text.

when exposed to the same input L1A sequence. These considerations assume that the raw L1A sequence (as generated in software) is directly sent to the software model and the HGCROC ASIC, and that there is no active throttling logic (i.e. Throttle “OFF” or disabled) that could alter the generated raw L1A stream. However, the software also allows to take the throttling decision into account (e.g. Throttle “ON” or enabled) which allows deactivating the problematic L1As in the raw L1A stream. Accordingly, it would then be the optionally throttled L1A sequence that would be emulated in the software model and the one sent to the HGCROC ASIC. In summary, the presented test-bench allows a BX-level comparison of the software model expectation against the real HGCROC output for either case of the throttling logic being enabled or disabled. Moreover, additionally observing the EC incrementing continuously in the case of enabled throttling is a sign of the throttling logic successfully protecting the system from problematic L1As.

The first result of these tests came from exposing the software model and the HGCROC ASIC to the special L1A sequences (shown in Appendix A) meant to cover all the corner cases of the RAM 2 and its readout logic behaviour. No disagreement between the two was observed for either having the throttling logic disabled or enabled. No discontinuities in the EC were observed when the throttling logic was enabled.

The second result of these tests came from exposing the software model and the HGCROC ASIC to random L1A sequences of various average consecutive L1A distances:

$$\langle D_{\text{L1A}} \rangle \in \{1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 30, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 50, 60, 70, 80, 90, 100\}.$$

meant to uncover any previously unknown corner cases of the RAM 2 and its readout logic behaviour. The distances were chosen to emphasise probing situations when the HGCROC would be: very “busy” (i.e. for $\langle D_{\text{L1A}} \rangle \leq 10$); and in the vicinity of the HGCROC Deadtime theoretical limit (i.e. for $\langle D_{\text{L1A}} \rangle$ within 41 ± 5). For each $\langle D_{\text{L1A}} \rangle$ choice, 50 random L1A arrays were generated, each being 2000 BX long. Having 29 different $\langle D_{\text{L1A}} \rangle$ choices, this resulted in a total of $29 \times 50 = 1450$ individual tests and is equivalent to a single L1A sequence with a combined effective BX length of $1450 \times 2000 \approx 3 \times 10^6$. No disagreement between the two was observed for either having the throttling logic disabled or enabled. No discontinuities in the EC were observed when the throttling logic was enabled. No additional edge case of the RAM 2 or its readout logic behaviour was hereby discovered.

These results gave confidence that the behaviour of the HGCROC is fully understood and properly modelled. Moreover, they opened doors to studying the $\text{Deadtime}(\langle f_{\text{L1A}} \rangle)$ and implementing the HGCAL L1A throttle block in firmware.

Deadtime vs. L1A frequency

The theoretical limit $\text{Deadtime}_{\min}(\langle f_{\text{L1A}} \rangle)$ derived from the fact that it takes the HGCROC 41 BX to stream out a single event (more details can be found in Section 5.1) did not consider the random nature of L1As and the occupancy limitations of the HGCROC RAM 2 and its readout logic. Having the exact BX-level behaviour of their filling/draining logic condensed in the software model, allowed to conduct a study of Deadtime as a function of the average L1A frequency, $\text{Deadtime}(\langle f_{\text{L1A}} \rangle)$.

There are several simple steps to the method behind simulating deadtime for a particular $\langle f_{\text{L1A}} \rangle$. Firstly, a random L1A sequence was generated in the software. Secondly, this sequence was played through the software model. The total number of BX when the throttle signal was raised, N_{DeadBX} , throughout the experiment was recorded. Together with N_{DeadBX} , knowing the total BX length of the simulation N_{BX} allowed deadtime to be calculated (as per Equation 5.2).

Specifically, the simulation was repeated for a wide (Figure 5.8a) and a narrow (Figure 5.8b) range of $\langle f_{\text{L1A}} \rangle$. The former considered the range $\langle f_{\text{L1A}} \rangle \in (700 \text{ kHz}, 1300 \text{ kHz})$ in steps of 50 kHz and the latter did so for the range in the vicinity of the nominal 750 kHz rate with $\langle f_{\text{L1A}} \rangle \in (744 \text{ kHz}, 756 \text{ kHz})$ in steps of 2 kHz. The simulated BX length behind each point in either of the plots was $N_{\text{BX}} = 1 \times 10^{12}$. The error bands and the error bars are given by 2σ (95%) confidence intervals. Due to low measurement uncertainties, these errors become clearly visible only in the narrow range (Figure 5.8b). On one hand, the higher frequency range (e.g. $\langle f_{\text{L1A}} \rangle > 975.6 \text{ kHz}$) shows very good agreement between the simulated deadtime and the theoretical HGCROC limitation, meaning that the latter is the dominant contribution to the overall Deadtime for this $\langle f_{\text{L1A}} \rangle$ range. On the other hand, in the lower frequency range (e.g. $\langle f_{\text{L1A}} \rangle < 975.6 \text{ kHz}$), a non-zero deadtime emerges due to the random nature of D_{L1A} and the occupancy limitations and implementation details of the HGCROC RAM 2 and its readout logic.

The software simulation resulted in an explicit value for the mean deadtime under the nominal $\langle f_{\text{L1A}} \rangle = 750 \text{ kHz}$ being $\text{Deadtime}_{\text{SW}}(750 \text{ kHz}) = (1.51 \pm 0.01) \times 10^{-8}$ (the errors are given as $\pm 1\sigma$).

It should be noted that the throttle signal is usually not raised for only one BX at a time. Instead, once raised it will on average stay raised for the next $41/2 = 20.5$ BX, depending on the phase of transmission of the event currently being sent from the HGCROC. Despite it being a rare effect, the CMS deadtime introduced by HGCAL will account for about 36 “dead” BX per one minute worth of running CMS at the nominal $\langle f_{\text{L1A}} \rangle = 750 \text{ kHz}$. Not using such a block to throttle problematic L1As would: cause CMS event reconstruction issues (i.e. HGCAL appearing as a volume that does not provide any HGCROC event data); and it would create synchronisation problems for the downstream HGCAL devices (e.g. the ECON-D and the DAQ BE would be missing some event data from the HGCROCs).

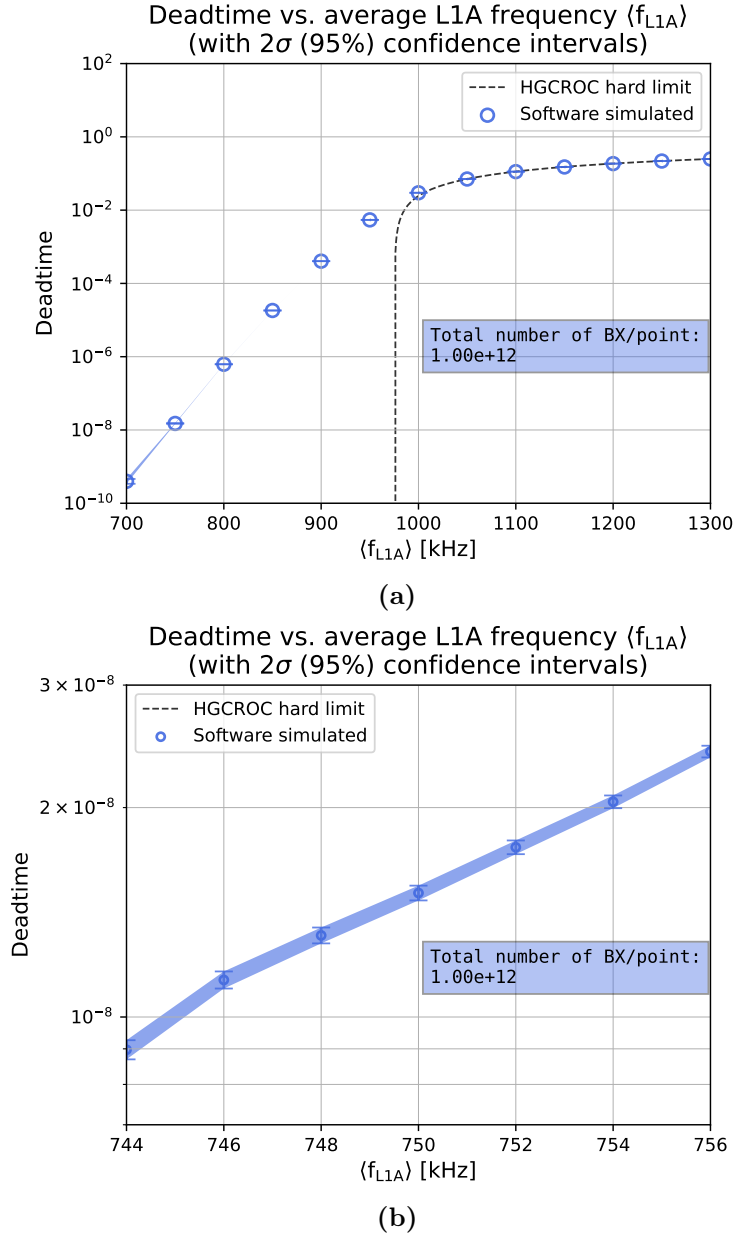


Figure 5.8: Software simulated deadtime (blue) as a function of the average L1A frequency $\langle f_{L1A} \rangle$ over a (a) wide range of frequencies (b) narrow range of frequencies in the vicinity of the nominal 750 kHz. The former also has the HGCROC hard limit (grey) from Section 5.1 drawn for reference. More details can be found in text.

5.3 The Firmware Implementation

After understanding the BX-level behaviour of the HGCROC RAM 2 and its readout logic, the HGCAL L1A Throttle Generator firmware block was implemented. Its underlying algorithm is essentially the same as the one of the software model so it will not be discussed in further detail here. This section is organised into two parts. Firstly, a general overview of the implemented firmware block is given. This considers a summary of its top-level interfaces accompanied by a typical time diagram of its inputs and outputs ⁶. Secondly,

⁶In other words, it acts as a manual for anyone wanting to use the firmware block.

this section presents several tests that validate the agreement of the firmware block to the software model (and consequently the HGCROC ASIC, given the results presented in Section 5.2).

5.3.1 General Overview

The implemented firmware block has four inputs and one output (Figure 5.9). All inputs and outputs are of VHDL⁷ `std_logic` type meaning they are single-bit signals that can take on logical "0" and "1" levels. The active level of the reset input `i_reset` is a logical "1". With it, the user can initialise the internal state machine, counters and status flags.

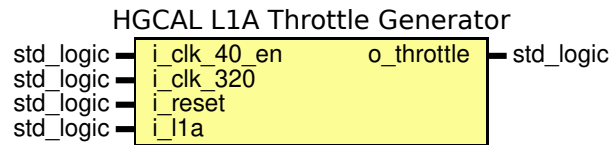


Figure 5.9: A simple schematic of the HGCAL L1A Throttle Generator firmware block with its interfaces and their types. More details can be found in text.

The block requires a 320 MHz clock input `i_clk_320`, synchronous to the 40 MHz LHC clock. It also needs an enable signal input `i_clk_40_en`, synchronous to `i_clk_320`, to mark the beginning of the 40 MHz LHC clock period. The block also expects a copy of the final L1A stream as an input. It is critical for this stream to be the same as the one sent to the FE. Otherwise, the HGCAL L1A Throttle Block would not be emulating the true state of the HGCROC. After a latency of seven 320 MHz clock periods, the final throttle decision is made available via `o_throttle` and is valid for one 320 MHz clock period⁸. This signal should be used in a combinatorial process to pass or deactivate (as proposed in Section 5.1) the current L1A candidate, brought to the `i_l1a` input. The explicit interplay of these signals in time can be presented through their time diagrams (Figure 5.10).

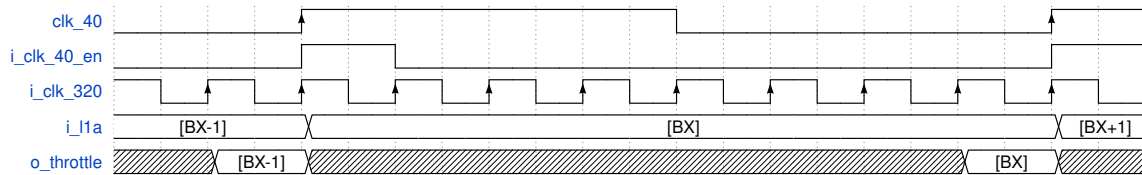


Figure 5.10: A typical time diagram with input and output waveforms of the HGCAL L1A Throttle Generator firmware block. The 40 MHz clock `clk_40` is not part of the block interface, but is drawn to illustrate its time dependence to `i_clk_40_en` and `i_clk_320`. More details can be found in text.

For correct use of this block, it is very important to note that any condition that would reset the FE HGCROC RAM 2 and its readout logic (e.g. a power cycle or a GPIO reset of the HGCROCs, the EBR fast command, etc.) would have to be accompanied by the reset of this block in TCDS2 as well (and vice versa). This is a limitation, because, by design, TCDS2 does not “know” about the specific slow-control and fast-control decisions that the HGCAL makes and HGCAL does not need to be informed about TCDS2 resetting

⁷VHDL is a hardware description language used for FPGA algorithm development.

⁸The latency of the device is $7 \times 3.125 \text{ ns} = 21.88 \text{ ns}$, which well bellow the 25 ns requirement.

its internal blocks. Hence, using this throttling scheme can be done by following two sets of rules, one during CMS initialisation and the other during data acquisition. During CMS initialisation (i.e. before sending L1As to the FE), it is necessary to: reset this block in TCDS2, and ensure that HGCROC event buffers are empty. During data acquisition (i.e. when L1As are being sent to the FE), it is necessary to prevent: the HGCAL DAQ BE from performing any slow or fast control action that would reset the HGCROC RAM 2 and its readout logic; and TCDS2 from again resetting this block. If any of these rules are not followed, the HGCAL L1A Throttle Block would consequently not be emulating the true state of the HGCROC.

The overall resource utilisation of the HGCAL L1A Throttle Generator firmware block is extremely light (Figure 5.11), requiring only about a few tens of FPGA LUTs and registers. These correspond to less than 0.01 % of the total available resources on a KU15P FPGA and will occupy even a smaller fraction of the VU35P FPGA, used on the DTH400 boards, where the final TCDS2 system will reside.

Name	CLB LUTs (522720)	CLB Registers (1045440)	LUT as Logic (522720)	CLB (65340)	CARRY8 (65340)	F7 Muxes (261360)	F8 Muxes (130680)	LUT as Memory (161280)	Block RAM Tile (984)	URAM (128)
top	117893	158469	115292	24770	4931	4579	1291	2601	360	4
payload (emp_payload)	56345	68863	55674	10663	3674	3223	1084	671	126	4
fastcommand_gen (fc_top)	1887	2388	1827	426	99	89	10	60	8	0
tcds2_emu (tcds2_fake)	1227	1392	1189	243	74	81	8	38	4	0
hgcroc_throttle_model (l1a_throttle_generator)	41	34	41	9	0	0	0	0	0	0

Figure 5.11: A resource utilisation report for the implemented HGCAL L1A Throttle Generator firmware block (highlighted in blue). Extracted from the vertical test system BE firmware project implemented in Vivado (the official Xilinx FPGA algorithm development framework) for the Xilinx KU15P FPGA. More details can be found in text.

Developing high-speed and low-latency algorithms in FPGAs requires careful planning when it comes to meeting the setup and hold time requirements (t_{setup} and t_{hold}) of internal signals. The former puts a constraint on how early should a given digital signal transition to a stable state relative to the next clock edge when it would be evaluated. The latter puts a constraint on how long should a given digital signal remain stable after the clock edge for its evaluation is meant to arrive. The implemented firmware block met all specified timing constraints (Figure 5.12).

Design Timing Summary			
Setup	Hold	Pulse Width	
Worst Negative Slack (WNS): 0.116 ns	Worst Hold Slack (WHS): 0.010 ns	Worst Pulse Width Slack (WPWS): 0.000 ns	
Total Negative Slack (TNS): 0.000 ns	Total Hold Slack (THS): 0.000 ns	Total Pulse Width Negative Slack (TPWS): 0.000 ns	
Number of Failing Endpoints: 0	Number of Failing Endpoints: 0	Number of Failing Endpoints: 0	
Total Number of Endpoints: 123569	Total Number of Endpoints: 122991	Total Number of Endpoints: 54108	
All user specified timing constraints are met.			

Figure 5.12: A timing report for the implemented HGCAL L1A Throttle Generator firmware block in Vivado. More details can be found in text.

5.3.2 Agreement with the Software Model

Standalone behavioural simulation and FPGA hardware tests

When checking the agreement between the software model and the HGCROC ASIC (more details can be found in Section 5.2), it was shown that the L1A sequences from Appendix A

are probing all the corners of the HGCROC RAM 2 and its readout logic behavioural model. Hence, to test the agreement of the implemented firmware block with the software model (and consequently the HGCROC), it was sufficient to stimulate it with these L1A sequences. This was done on two levels. The first level of testing considered a standalone behavioural simulation of the implemented firmware block in a Vivado simulation. The second level considered testing the standalone firmware block loaded on a physical Serenity board KU15P FPGA. The chosen L1A sequences were used as input stimuli for the simulated or physical firmware blocks and the BX-level outputs of the blocks were then captured to output files in either case. The respective output files were then compared (on a BX-level) between each other and against the expectations of the software model ⁹. Similarly to the agreement in tests between the software model and the HGCROC, these tests were repeated for both the throttle decision being taken into account and ignored. No issues were observed in any of the tests of the firmware block.

Protecting HGCROCs in the Vertical Test System

Successful standalone tests of the firmware block opened doors to its integration into the TCDS2 emulator of the vertical HGCAL electronics test system Serenity BE (more details about the vertical system can be found in Chapter 3 and Chapter 4). This was done to protect the HGCROCs from problematic L1As generated during data acquisition runs orchestrated from the DAQ PC (e.g. in the lab or during beam tests), but also to demonstrate the block feasibility in a realistic application to the TCDS2 firmware engineers. Three kinds of tests were done to check if the HGCAL L1A throttle firmware block was properly instantiated in the TCDS2 emulator of the vertical test system.

The first test considered firing the special L1A sequences ¹⁰ (shown in Appendix A) from the DAQ BE fast control firmware block and counting the total number of L1As that were sent to the HGCROCs. This was repeated for both cases of the throttling logic being disabled or enabled. The total number of L1As sent out to the HGCROCs was identical to the prediction of the software model in both cases.

The second test was conducted during a 3 min DTH data acquisition run with an average random L1A rate of 750 kHz. This run was previously described in Chapter 4. Its total duration was $3 \text{ min} \times 40 \text{ MHz} \approx 7 \times 10^9 \text{ BX}$, where about 100 BX were predicted to be throttled by the firmware block software model. Nevertheless, throughout the 133×10^6 events acquired during that run, there were no discontinuities observed in the ECON-D event counter, which was configured to pass through the vertically-reconstructed values from the HGCROCs. This indicates that the HGCAL throttle block has successfully protected the HGCROCs from the problematic L1As in this DAQ run.

The third test consisted of doing two 30 s DTH data acquisition runs with an average random L1A rate of 1 MHz (to ensure being in the problematic zone on average). In both runs, the ECON-Ds were configured to: pass the HGCROC L1A timestamps from the vertical reconstruction results (so that the HGCROC event counter information is available); and to be in Zero-suppression mode (to ensure that the packets are small enough so they can be saved to the DAQ-PC, even at such high rates). Throughout the first and second run, the HGCAL throttle firmware block was kept disabled and enabled,

⁹These checks included the inputs/outputs of the firmware block, but also the low-level internal signals such as the current number of words in RAM 2, etc.

¹⁰Each L1A sequence was of course preceded by an EBR fast command to reset both the HGCROC buffers as well as the logic of the throttle block to synchronise them to one another.

respectively. In the first test, about 600×10^3 discontinuities were observed in the ECON-D (hence, the HGCROC) event counter. This is in agreement with the theoretical predictions from the beginning of this chapter (Figure 5.3a). No event counter discontinuities were observed in the second test, again indicating correct operation of the implemented firmware block.

Deadtime vs. L1A frequency

The final test of agreement between the firmware block and the software model was to measure deadtime over a range of average L1A frequencies $\langle f_{L1A} \rangle$, similar to the $\text{Deadtime}(\langle f_{L1A} \rangle)$ simulation results (shown in Section 5.2). The RCC system on the DAQ PC of the vertical test system allowed configuring the DAQ BE fast control firmware to produce random L1As with the $\langle f_{L1A} \rangle$ specified at the beginning of a run. During initialisation, the RCC would issue the EBR fast command and a reset of the throttle firmware block, ensuring both parts are operating with the same start conditions. The TCDS2 emulator has a counter which could be read from the Serenity board controller. This counter conveys how many BX were throttled N_{DeadBX} during a run. Together with N_{DeadBX} , knowing the total BX length of the experiment N_{BX} allows a calculation of Deadtime (as per Equation 5.2). Specifically, the measurements were repeated for the wide (Figure 5.13a) and the narrow (Figure 5.13b) range of $\langle f_{L1A} \rangle$, as previously chosen during $\text{Deadtime}(\langle f_{L1A} \rangle)$ simulations. The BX length behind the points in the wide-frequency plot is $N_{\text{BX}} = 4 \times 10^{10}$ whereas in the narrow-frequency plot that is $N_{\text{BX}} = 4.32 \times 10^{11}$. The latter was chosen such to give a smaller measurement uncertainty in the region of most interest. The error bands and the error bars are again given by 2σ (95%) confidence intervals. These become clearly visible only in the narrow range (Figure 5.13b) due to low measurement uncertainties.

The hardware measurement resulted in an explicit value for the mean deadtime under the nominal $\langle f_{L1A} \rangle = 750 \text{ kHz}$ being $\text{Deadtime}_{\text{HW}}(750 \text{ kHz}) = (1.56 \pm 0.02) \times 10^{-8}$ (the errors are given as $\pm 1\sigma$).

The above results indicate good agreement between the hardware measured and software simulated $\text{Deadtime}(\langle f_{L1A} \rangle)$, giving yet another degree of confidence about the overall agreement between the HGCal L1A Throttle firmware block and its corresponding software model.

5.4 Results Summary

This chapter demonstrated the feasibility of the overall implementation of the HGCal L1A Throttle Generator firmware block which can be summarised through several most important points.

The critical step in this work was to reverse-engineer the exact BX-level behaviour of the HGCROC RAM 2 buffer and its readout logic. This was confirmed once the software model and the HGCROC ASIC were observed to agree when exposed to the special L1A sequences (tailored to cover all the known corner cases of the RAM 2 and its readout logic) and random L1A sequences of various average consecutive L1A distances $\langle D_{L1A} \rangle$ between 1 to 100 having a combined effective duration of $3 \times 10^6 \text{ BX}$. This opened the door to study Deadtime as a function of the average L1A frequency, $\text{Deadtime}(\langle f_{L1A} \rangle)$, in software, but also to proceed with the firmware implementation of the block.

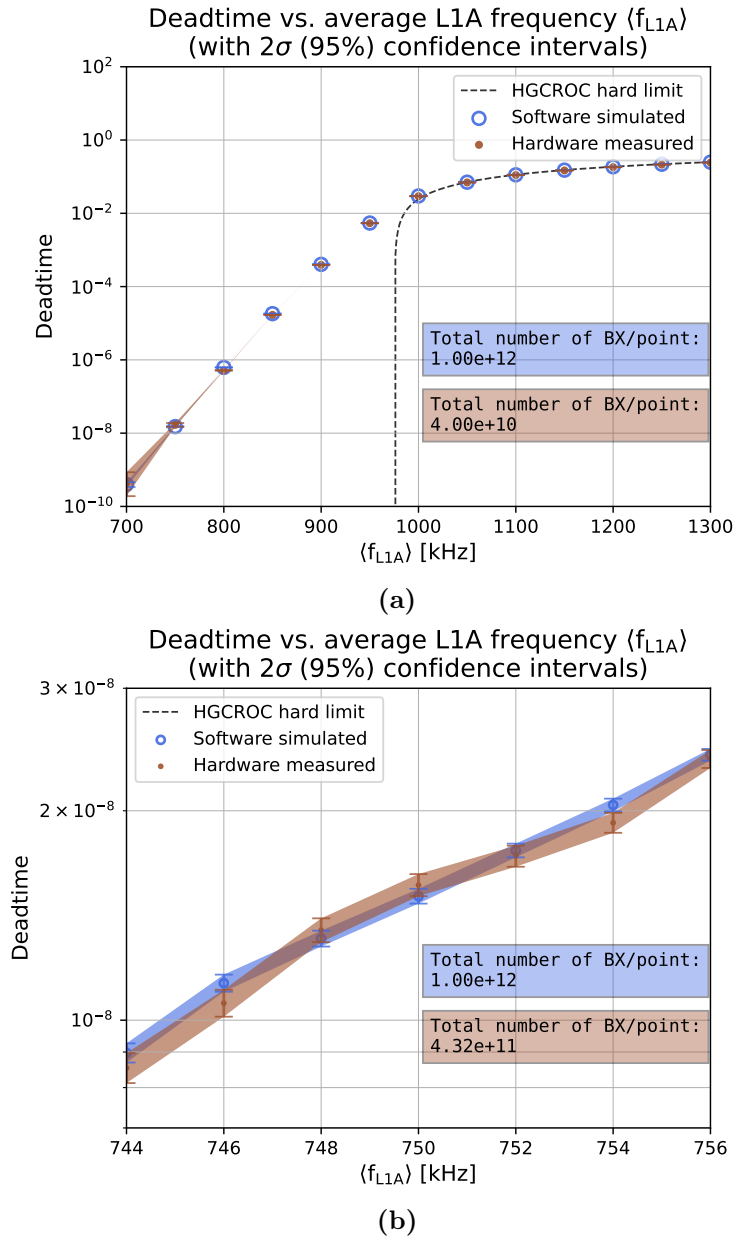


Figure 5.13: Software simulated (blue) and hardware measured (brown) Deadtime as a function of the average L1A frequency $\langle f_{L1A} \rangle$ over a (a) wide range of frequencies (b) narrow range of frequencies in the vicinity of the nominal 750 kHz. The former also has the HGCROC hard limit from Section 5.1 drawn for reference. More details can be found in text.

The software model was used to simulate $\text{Deadtime}(\langle f_{\text{L1A}} \rangle)$ in the $\langle f_{\text{L1A}} \rangle$ range between 700 kHz to 1300 kHz and allowed to predict an explicit value for the mean deadtime under the nominal $\langle f_{\text{L1A}} \rangle = 750 \text{ kHz}$ being $\text{Deadtime}_{\text{SW}}(750 \text{ kHz}) = (1.51 \pm 0.01) \times 10^{-8}$.

When testing the implemented firmware block standalone, its behaviour was one-to-one matched with the software model (and thus with the HGCROC ASIC) when using the aforementioned special L1A sequences as input stimuli. This was done in two levels of tests: FPGA simulation and standalone hardware tests of the block.

Having instantiated the implemented firmware block in the TCDS2 emulator of the vertical HGCal test system allowed performing DAQ runs where acquired data could be saved and examined offline (specifically, checking for the HGCROC event counter discontinuities). No HGCROC event counter discontinuities were observed when exposing the system to the aforementioned special L1A sequences and a random L1A sequence with an average rate of 750 kHz for a total of about 133×10^6 events. Both of these results indicate that the block successfully protected the HGCROCs in the system FE from problematic L1As while running at the nominal HL-LHC L1A rate of 750 kHz.

Using the implemented firmware block in the TCDS2 emulator of the vertical HGCal test system allowed measuring $\text{Deadtime}(\langle f_{\text{L1A}} \rangle)$ in the $\langle f_{\text{L1A}} \rangle$ range between 700 kHz to 1300 kHz. Additionally it allowed measuring the explicit value for the mean deadtime under the nominal $\langle f_{\text{L1A}} \rangle = 750 \text{ kHz}$ being $\text{Deadtime}_{\text{HW}}(750 \text{ kHz}) = (1.56 \pm 0.02) \times 10^{-8}$. Both of these measurements indicated good agreement between the implemented firmware block and the corresponding predictions by the software model.

Lastly, although this chapter did not provide a more detailed discussion on this topic, it should be noted that this L1A throttling firmware block was successfully used in the TCDS2 emulator of the vertical readout system during the beam test activities previously outlined in Chapter 4.

Chapter 6

Calculating Cluster Properties in the HGCAL TPG back-end

A significant portion of Chapter 2 gave a general overview of the TPG path in the HGCAL electronics system. This chapter focuses on calculating the physical properties of 3D clusters in Stage 2 of the TPG BE firmware. It is organised into four parts. Firstly, it motivates why the HGCAL cluster properties are desired as inputs to the CMS L1T during HL-LHC (Section 6.1). Secondly, it gives a general overview of the algorithm (Section 6.2). Thirdly, it presents the key perspectives that emerge from the software model of the implemented cluster properties block (Section 6.3). These involve evaluating the degree of quantisation-induced errors to the resulting properties and demonstrating the usefulness of the implemented cluster properties in terms of their physics performance. Specifically, the last point is made by demonstrating the ability of the properties to distinguish clusters originating from photons (or VBF jets) and PU in the transverse momentum p_T range below 100 GeV. Lastly, this chapter summarises the implementation details of the corresponding firmware algorithm (Section 6.4).

6.1 The Motivation for Cluster Properties

Having in mind the LHC BX rate of 40 MHz and the projected average rate of L1T decisions during HL-LHC of 750 kHz, one can find that only 1 in about 50 BXs (2 %) are allowed to pass the selection in this first level of CMS triggering. In contrast, the rest must be discarded (98 %). To achieve this, the L1T receives trigger primitives from all its sub-detectors. These primitives are then processed by the algorithms implemented in the L1T FPGAs, where a combined global trigger decision is made. The HGCAL will provide two kinds of such inputs to the L1T for each BX from each of the nominal 120° ϕ sectors of an endcap: one 16-bit energy tower for every bin in a 20 by 24 (η, ϕ) grid; and on average 34 clusters each described by a 256-bit record of its physical properties (the focus of this chapter).

Generally for CMS, it is essential for the trigger primitives (standalone or combined) to be powerful discriminators between the different kinds of physics signatures in the detector. In other words, knowing *what* happened in a given BX allows the L1T to make an informed decision about whether it is interesting to trigger. This is particularly difficult due to the high PU levels (i.e. as high as 200) during HL-LHC. Therefore, it is crucial for the trigger

primitives to exploit and reflect the available granularity of the detector FE. However, a recurring theme when implementing any TPG algorithm is balancing the level of detail conveyed by the trigger primitives with FPGA resource utilisation, algorithm latency, and bandwidth limitations. An optimal trigger primitive may be theoretically ideal, but if its implementation exceeds the resource constraints of the target FPGA, incurs excessive computation time, or surpasses the available bandwidth budget, it becomes infeasible.

Specifically for HGCal, some interesting physics signatures to distinguish from the usual ones coming from a PU of 200 simultaneous (i.e. within a single BX) proton-proton interactions are: the clusters arising from the e/γ showers; and *high p_T jets*. Motivated by the HGCal TDR from 2019 [29], this chapter focuses on clusters arising from *photons* and *VBF jets* when evaluating the physics performance of the implemented cluster properties (Figure 6.1). It is such objects (i.e. clusters of TCs; as shown in Figure 6.1) about which a record of physical properties is calculated in Stage 2 of the HGCal TPG BE. The shown clusters are extracted from CMS Software (CMSSW) simulation.

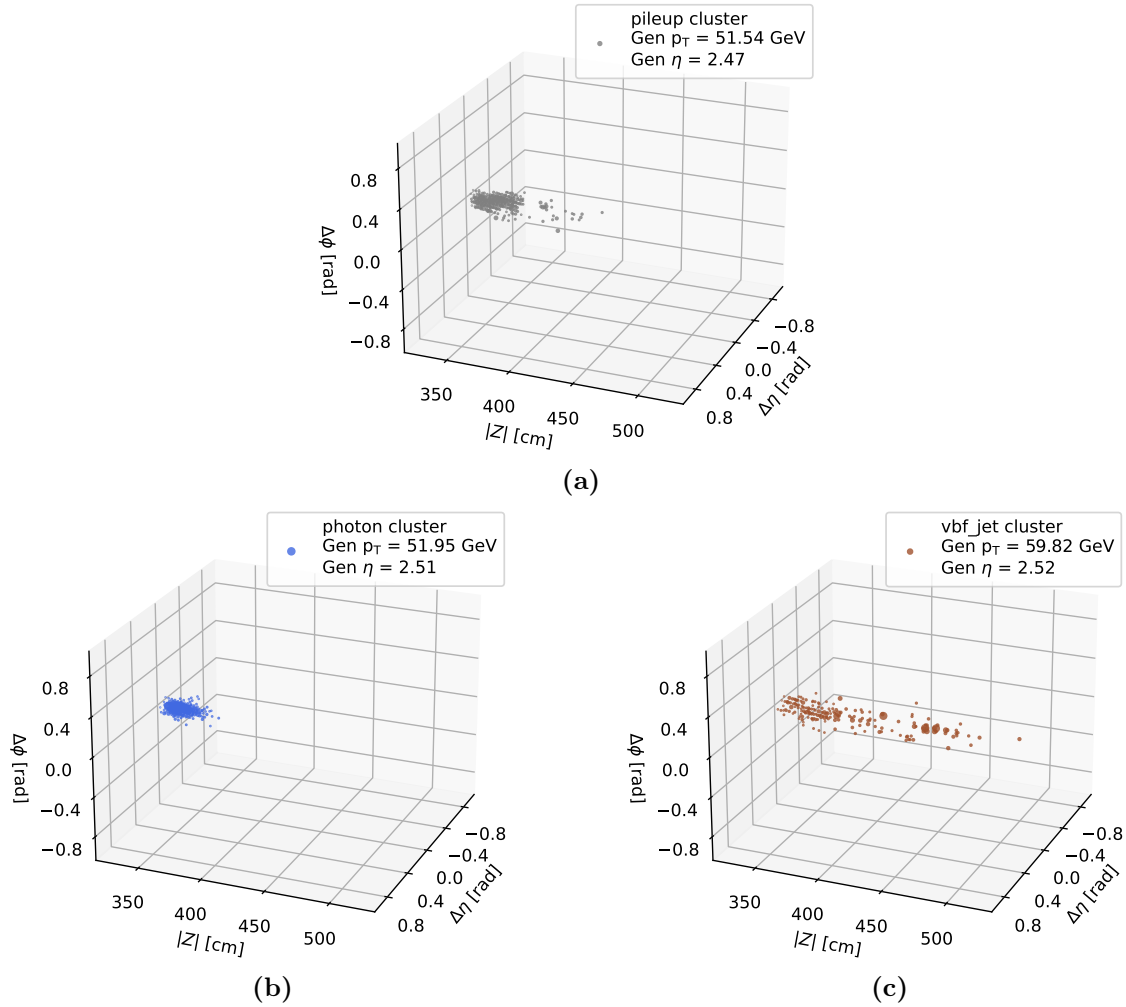


Figure 6.1: Qualitative examples of clusters of TCs originating from (a) PU (b) photon (c) VBF jet illustrating their usual signatures in the HGCal TPG path. The sizes of spherical entries in the plots are proportional to the E_T of the corresponding TCs. To give a more balanced comparison, the three shown clusters all come from similar p_T and η ranges. Data is extracted from CMSSW simulation. More details can be found in text.

Furthermore, because it is considered that the L1T will trigger on any HGCAL cluster above a certain p_T threshold, the focus of this chapter is to evaluate the performance of a cluster classifier algorithm (which uses the implemented cluster properties) for $p_T < 100$ GeV, where the discrimination task is more difficult than beyond such a p_T cut (more details can be found in Section 6.3).

To explicitly motivate the usefulness of having HGCAL cluster properties as L1T trigger primitives, one can consider the probability distributions of two highly performing quantities (Figure 6.2). The example property for the photon identification task is the fraction of the cluster energy deposited in the core of the HGCAL CE-E section (Figure 6.2a). Unlike PU clusters, the electromagnetic showers arising from the photons are expected to usually deposit all of their energy in the CE-E part of HGCAL. Hence, the mean of this probability distribution takes on higher values in the case of photon clusters when compared to the mean of PU clusters. The example property for the VBF jet identification task is the cluster standard deviation in ϕ (Figure 6.2b). The jets are on average expected to have a more narrow width in ϕ (and η) when compared to the PU clusters. Hence, the mean of this probability distribution takes on lower values in the case of VBF jet clusters compared to the mean of PU clusters.

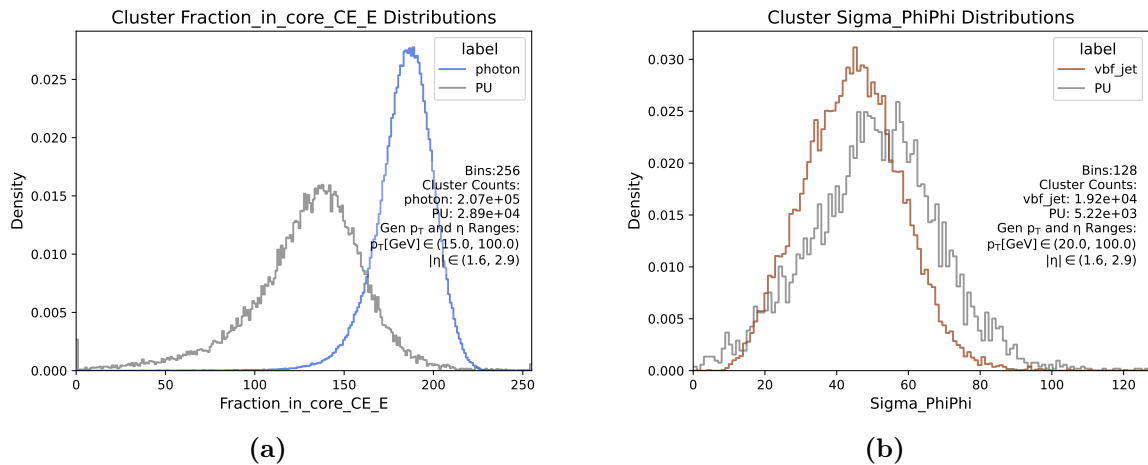


Figure 6.2: An illustration of the discrimination power of two example cluster properties when used to distinguish clusters originating from (a) photons and (b) VBF jets against clusters originating from PU. Data is extracted from CMSSW simulation. The explicit property definitions, details of the conducted study and x-axis units can be found later in the chapter. More details can be found in text.

There are several high-level remarks to note early on, regarding how the signal identification performance of the cluster properties will be studied in this chapter. Firstly, Figure 6.2 only illustrates the discrimination power of one variable for each of the two mentioned identification problems (i.e. photon vs. PU, and VBF jet vs. PU). However, the overall discrimination power will become apparent only when the contributions from various individual properties are combined. Similarly, the study will focus only on the HGCAL TPG primitives (single clusters, specifically) from one Stage 2 ϕ sector of a single HGCAL endcap. In real operation, the final L1T decision will rely on significant extra information. Specifically, the decision will be based on additional HGCAL TPG information (e.g. multiple clusters from both endcaps and tower energies) and also on trigger primitives from other CMS sub-detectors (e.g. the Tracker, when doing track-to-cluster matching).

In this chapter, the discrimination power will be studied by training a simple Boosted Decision Tree (BDT) model as a binary classifier for each of the two binary classification problems. Specifically, the goal of this chapter is not to explore how a BDT (or another Machine Learning (ML) algorithm for binary classification) can be optimised but rather to evaluate the performance of the cluster properties on a very high level (e.g. even when using the default BDT parameters). Furthermore, these BDT models will be only implemented in software and their potential firmware implementation will not be discussed. In the final CMS trigger system, such algorithms will be implemented in the L1T FPGAs and are not the responsibility of the HGCAL TPG.

6.2 A General Overview of the Algorithm

This section is organised into four parts. Firstly it explains the core mechanism behind cluster property calculation. This occurs in a sequence of three steps (Type Modifier, Accumulator and Property Calculator) which are then separately discussed in more detail. It should be noted that the algorithm described here is essentially common to both the software model (Section 6.3) and the firmware implementation (Section 6.4) which are discussed after this section.

6.2.1 The Core Mechanism

There are three main steps in the core algorithm of the Stage 2 Cluster Properties block (Figure 6.3). The output of the Stage 2 Cluster Finding algorithm has 78 parallel streams across which TCs with a cluster ID are sent out (more details can be found in Chapter 2). These streams are input to the three-step cluster property calculation algorithm. In the Type Modifier, each TC gets converted to a different representation favourable for the Accumulator step. In the Accumulator, the TCs get grouped by their cluster ID and aggregated through a linear process into data packets called Cluster Sums. There can be many TCs, but only one Cluster Sum per cluster. The output of the Accumulator is a serial (i.e. not parallel like the input) stream of Cluster Sums. Each Cluster Sum is a representation of a cluster, which comes as a structured data type. In the Property Calculator block, the individual fields of Cluster Sums are used as input primitives when calculating the final properties (i.e. as expected by the L1T) about the cluster. It is worthwhile noting that the Cluster Properties block uses TCs from a 180° ϕ sector of an endcap to describe clusters from the nominal 120° ϕ sector (this organisation in ϕ is

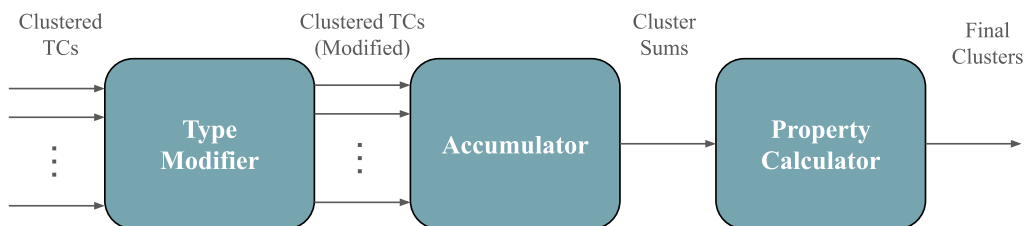


Figure 6.3: The three main steps in the Stage 2 Cluster Properties block. More details can be found in text.

described in Chapter 2). Additionally, the TCs are represented in a coordinate system local ¹ to the 180° ϕ sector.

How a cluster property is calculated in these three steps can be illustrated through an example of a simple property - the weighted mean of the cluster ϕ coordinate. Here, the relevant fields describing each TC are its transverse energy E_T and the ϕ coordinate. At the output of the Type Modifier block, the generalised weight W and the weight-coordinate product $W\phi$ fields about the TC are formed. Thereby, W can be thought of as an approximation of E_T which relaxes the requirements on the overall FPGA resource utilisation required by the Cluster Properties block (the explicit rule and the motivation for transforming E_T into W is clarified later on). The Accumulator sums these fields for all the TCs with the same Cluster ID, forming the $\sum W$ and $\sum W\phi$ fields. The summed quantities are used as inputs to the Property Calculator which performs a division to obtain $\mu(\phi)$ (Equation 6.1).

$$\mu(\phi) = \frac{\sum W\phi}{\sum W} \quad (6.1)$$

Because this property is at first calculated in the Stage 2 local coordinate system and the L1T require it represented in a specific format ², an offset and a multiplication by a scaling factor are applied before it is finalised.

Coming from FPGA development considerations, there are two important constraints on the complexity of the Cluster Properties block. Firstly, each property needs to be computed after a single pass (i.e. nested loops are not allowed) through the TCs of a cluster. Secondly, the bit widths of the data packets at the inputs and outputs of the Accumulator block need to be kept within a few hundred bits. If either of these two constraints is not respected, the property would be infeasible to implement in firmware, requiring FPGA resources outside of the available budget. Hence, it is generally important to balance the level of detail conveyed by the cluster properties with their FPGA resource requirements.

6.2.2 The Type Modifier

At the input of the Type Modifier step, each TC is represented as a structured data type defined by several fields (Table 6.1). The listed ranges come from: HGCAL endcap geometry first-principles (e.g. the spatial information from the mechanical layout); interface with the L1T (e.g. there can be no more than 131 clusters output by a single Stage 2 FPGA per one BX); and CMSSW simulation (e.g. the useful TC energy range). The bit-width and Least Significant Bit (LSB) choices stem from the current understanding of the required TC level of precision across the corresponding ranges. It is worth noting that although an HGCAL endcap will have 47 layers, only the alternate layers of the CE-E and all of the layers in the CE-H section are used by the trigger. That totals in 34 trigger layers per endcap. The *Cluster* field specifies the ID of the cluster that the given TC is associated with.

In the first steps of the Type Modifier block, the local z coordinate is looked up from a dictionary, knowing the *Layer* field. The looked-up TC z coordinate is represented with

¹Except the TC r/z coordinate. This coordinate is only local to one endcap (i.e. always positive), but it ranges 0.0822 to 0.575 (i.e. it is not shifted to start from 0).

²The interface between HGCAL and L1T defines $\mu(\phi)$ to be represented in the signed 2's complement binary system, with a resolution of $\pi/720$, using 9 bits and covering the range $\phi \in (-64^\circ, 64^\circ)$.

Table 6.1: Trigger cell definition. More details can be found in text.

Field	Bit-width	LSB	Range	Description
E_T	19	2^{-10} GeV	0 GeV to 512 GeV	Transverse energy
ϕ	12	$\pi/(108 \times 32)$	0 to π	Local ϕ coordinate
r/z	13	7.017×10^{-5}	0.0822 to 0.575	Local r/z coordinate
Layer	6	1	1 to 34	Layer ID
Cluster	8	1	0 to 130	Cluster ID
Total bit-width	58			

12 bits, which is sufficient to cover the local HGCal z range between 0 mm to 1930 mm with an LSB of 0.5 mm. The η information is not available on the TC level and the final η related cluster properties (e.g. cluster η mean position or standard deviation) are looked up from the corresponding r/z properties at the end of the algorithm.

The output of the Type Modifier consists of TCs, each represented in a structured data type (Table 6.2) which is favourable for the Accumulator step. In fact, the bit-level representation of TCs and Cluster Sums at the output of the Type Modifier and the Accumulator are exactly the same, respectively. It is only the interpretation of the bits at the output of these two blocks that is different. Namely, a data packet output by the Type Modifier and the Accumulator represents a single TC and a cluster of TCs, respectively. In other words, one can consider the TCs going into the Accumulator as partial (i.e. incomplete) Cluster Sums, each of them being finalised after the aggregation of TCs originating from the same cluster is complete. This allows the Accumulator to define an accumulation operator “+” which operates within a single data type.

The transverse energy of a TC is kept in the E_T field, but there are other energy fields added. For example, $E_{T(CE-E)}$, $E_{T(CE-E-Core)}$ and $E_{T(CE-H-Early)}$ copy the E_T of the TC if it originates from the electromagnetic section, core of the electromagnetic section, or the early layers of the hadronic section, respectively. Otherwise they are set to zero. Before the E_T of the TC is copied into any of the $E_{T(CE-E)}$ or $E_{T(CE-E-Core)}$ fields it is multiplied by a layer-dependent constant with values ≈ 1 , to compensate for the slight differences in the response of HGCal to electromagnetic and hadronic sources.

The TCs are also described by fields such as W , W^2 , Wc and Wc^2 , where c symbolises any of the three spatial coordinates (i.e. $c \in \{\phi, r/z, z\}$) and W is the generalised weight (Equation 6.2)

$$W = \max(E_T/2^k, 1) \quad (6.2)$$

where k can be any positive integer smaller than the bit-width of the E_T field. In an FPGA, performing a division by 2^k is equivalent to a bit-shift to the right by k bits. Choosing the maximum of the two values ensures that the weight W is never zero because the result of this E_T scaling is only the quotient of the division. The k parameter is a “handle” in the algorithm that can reduce the bit-width requirements of the corresponding fields at the cost of precision of the cluster properties whose calculation depends on W (as it will be clarified later, these are the weighted mean positions and the standard deviation properties). This is useful for finding a compromise between the FPGA resource requirements in the Accumulator step and physics performance. This chapter usually

Table 6.2: Modified Trigger cell definition, where $c \in \{\phi, r/z, z\}$. The bit-widths of W , W^2 , Wc and Wc^2 are listed assuming $k = 0$ in the W definition. More details can be found in text.

Field	Bit-width	Description (TC-level)
E_T	22	Transverse energy
$E_{T(CE-E)}$	22	E_T in CE-E
$E_{T(CE-E-Core)}$	22	E_T in core CE-E (layers 4-8)
$E_{T(CE-H-Early)}$	22	E_T in early CE-H (layers 14-17)
W	19	Generalised weight
W^2	38	Generalised weight squared
Wc	31, 32, 31	Product of W and the respective c
Wc^2	43, 45, 43	Product of W and the respective c^2
<i>LayerBits</i>	34	One hot encoding of TC layer ID (counting from 1, not 0)
<i>SatTC</i>	1	1 if the TC energy saturated, else 0
N_{TC}	10	Number of TCs (hard-coded to 1)
$N_{TC(W)}$	10	Number of TCs used to form W (hard-coded to 1)
<i>ShapeQ</i>	1	Shape Quality Flag (hard-coded to 1)
<i>Cluster</i>	8	Cluster ID
Total bit-width	434	

assumes the ideal case when $k = 0$ (i.e. $W = E_T$) as the working baseline. However, notes on the impact of the k parameter on the precision of the calculated properties and its impact on the resulting physics performance will be briefly mentioned in Section 6.3).

There are a few more TC fields to mention here. The *LayerBits* field is a 34-bit binary vector, with a “one-hot” encoded layer ID. For example, a *Layer* = 3 field of a TC gets transformed into *LayerBits* = 0b001000...0. The *SatTC* flag is added to indicate if the transverse energy of the TC has saturated to all ones in the 19-bit E_T field. The N_{TC} , $N_{TC(W)}$ and *ShapeQ* fields are all hard-coded to 1 at this step of the algorithm. Lastly, the *Cluster* field is copied directly from the Type Modifier input.

The complete meaning of the mentioned fields (and their bit-width choices) in the Type Modifier output will clarify after describing the Accumulator.

6.2.3 The Accumulator

The Accumulator block receives TCs from the Type Modifier. Here, the TCs originate from various clusters. This module distinguishes the ones originating from different clusters, and accumulates the ones coming from the same cluster. The accumulation happens within every one of the fields that describe TCs at the output of the Type Modifier. At the output of the Accumulator, a single Cluster Sum is streamed out for every cluster. These data packets come in the form of a structured data type (Table 6.3). The accumulation operator

“+” is mainly defined through the comparison of the previous two tables (Table 6.2 and Table 6.3) but some additional notes are still needed.

Table 6.3: Cluster Sum Definition, where $c \in \{\phi, r/z, z\}$. The bit-widths of W , W^2 , Wc and Wc^2 are listed assuming $k = 0$ in the W definition. More details can be found in text.

Field	Bit-width	Description (Cluster-level)
$\sum E_T$	22	Sum of E_T of TCs
$\sum E_{T(CE-E)}$	22	Sum of E_T of TCs in CE-E
$\sum E_{T(CE-E-Core)}$	22	Sum of E_T of TCs in core CE-E (layers 4-8)
$\sum E_{T(CE-H-Early)}$	22	Sum E_T of TCs in early CE-H (layers 14-17)
$\sum W$	19	Sum of generalised weights of TCs
$\sum W^2$	38	Sum of generalised weights squared of TCs
$\sum Wc$	31, 32, 31	Sum of products of W and the respective c of TCs
$\sum Wc^2$	43, 45, 43	Sum of products of W and the respective c^2 of TCs
$\bigcup LayerBits$	34	Bitmap of layer IDs of the cluster that have TC hits
$\bigcup SatTC$	1	True if a TC of the cluster saturated in E_T , else False
N_{TC}	10	Number of TCs of the cluster
$N_{TC(W)}$	10	Number of TCs of the cluster used to form $\sum W$
$ShapeQ$	1	False if there are more TCs than used to form $\sum W$
$Cluster$	8	Cluster ID
Total bit-width	434	

Regarding the four transverse energy sums, they are programmed to saturate at 22 bits which corresponds to 4096 GeV. Regarding the fields depending on the generalised weight W , there are a couple of remarks remaining. Firstly, it is considered that the clusters whose $E_T > 500$ GeV will cause a trigger regardless of other properties. Hence, the properties depending on W do not need to be calculated well in such a scenario. This means that with such a consideration and for $k = 0$, 19 bits are sufficient for the $\sum W$ field (i.e. it is possible to encode up to 512 GeV³ with an LSB of 2^{-10} GeV). Secondly, the $\sum W$ keeps being incremented by new TC entries until 80% of the 512 GeV range is reached and then the summation gets stopped. This condition stops not only the summation of the W field, but all other fields defined by the weight W . Namely, these are W , $N_{TC(W)}$, W^2 , Wc and Wc^2 fields. If this condition is met, that is indicated by the *ShapeQ* flag. The idea is that in the case of a very energetic cluster (higher transverse energy than 512 GeV), it is not desired to allow the mentioned fields associated to the weight W to saturate. Instead, the goal is to form these fields from a “random-enough” sample of TCs from the cluster. Because these fields are used as inputs when calculating various cluster properties (weighted mean positions and standard deviations), their saturation would introduce non-linear effects into the final results. The bit-width choices of $\sum W^2$, $\sum Wc$

³Hence, the choice of $k = 0$ is compatible with the clusters considered in the physics performance study (i.e. their $E_T < 100$ GeV) in Section 6.3.

and $\sum Wc^2$ are a direct consequence of the bit-widths of W and c . It is expected that a single cluster cannot have more than a thousand TCs, hence 10 bits are sufficient for encoding the N_{TC} and the $N_{TC(W)}$ fields. The $\bigcup LayerBits$ encodes a bitmap of the layers of a cluster that have at least one TC hit. It is a bit-wise logical OR of all the $LayerBits$ fields of individual TCs of the cluster. For example, if there are TCs from layers 1, 2, and 3, this would be represented as $\bigcup LayerBits = 0b111000\dots0$. The $\bigcup SatTC$ encodes if there is at least one E_T saturated TC within the cluster. It is a logical OR of all the $SatTC$ fields of individual TCs of the cluster.

A deeper look into how the Cluster Sums are formed reveals that the Accumulator block implementation requires a lot of FPGA resources (i.e. block RAM and input data bandwidth) and that this is susceptible to the total bit-width of the Cluster Sum data type (Table 6.3). Regarding the block RAM requirement, this is mainly due to the formation of a Cluster Sum data packet for every cluster happening in several steps. At first, only partial Cluster Sums get formed and these later get combined into the final Cluster Sums. Specifically, the Stage 2 Cluster Finding block outputs clustered TCs across multiple streams (78) into the Accumulator and it is possible that there are more than one stream with TCs that have the same cluster ID. Hence, within every stream it is needed to form partial Cluster Sums for all the IDs that occur in that stream. The partial Cluster Sums within various streams at the output of the first accumulation step still have the same IDs so they need to be combined together, forming “more complete” Cluster Sums. Merging these together, again within the same ID eventually converges to the final Cluster Sums. Specifically, when any partially completed Cluster Sum is calculated, it needs to be temporarily stored in a memory block before it is combined with its ID pair. For example, with this data type having the total bit-width in the range from 100 to 700 bits, the block memory requirement relative to what is available on a VU13P FPGA ranges from 36 % to 71 % (in total with other modules in Stage 2 that are requesting some of this resource). The current system configuration (for $k = 0$) is requesting about 58 % of block RAM. Regarding the input bandwidth, there are about 100 input streams, where each one propagates type-modified TCs with a bit-width of about 500 bits at a clock frequency of 360 MHz. This results in the Accumulator having to accept an input bandwidth as high as $78 \times 434 \text{ bit} \times 360 \text{ MHz} \approx 11 \text{ Tbit s}^{-1}$, which is a non-trivial amount (e.g. it is comparable to the interface of the HGCAL DAQ BE system with an entire HGCAL endcap where the usable data throughput is 9 Tbit s^{-1}). The output bandwidth of the Accumulator (same as the Property Calculator input bandwidth) is $1 \times 434 \text{ bit} \times 360 \text{ MHz} \approx 150 \text{ Gbit s}^{-1}$, given that Cluster Sums are being streamed across a single stream at a digital system clock frequency of 360 MHz. These points motivate exploring the effect of the k parameter on the physics performance of cluster properties. Increasing k relaxes the block RAM (and the input/output bandwidth) requirement of the Accumulator, but degrades the precision of the calculated properties that depend on W (i.e. weighted mean positions and standard deviations).

For completeness, it is worth noting that it is possible that Stage 2 will not be able to aggregate all TCs (but only a subset) belonging to a cluster in the Accumulator block, but that this process will have to stop after a latency limit is reached. This chapter does not study the implications of this effect.

6.2.4 The Property Calculator

From the Accumulator, the Property Calculator block receives Cluster Sums from whose fields it derives the final quantities about the clusters (Table 6.4). The HGCal $\rightarrow$ L1T interface bandwidth constraints allow having a cluster property record with 256 bits per cluster. Coming from the details of the interface specification, the complete list of implemented properties is organised into four 64-bit words. There are still 61 bits available (marked as *Spare*) for possible future quantities in this definition. Regardless, the cluster properties can be broken down into several categories: E_T in various longitudinal regions; E_T fractions in various longitudinal regions; E_T weighted mean positions; shape properties; and miscellaneous properties. More details about each category can be found below.

Cluster E_T in various longitudinal regions

The total cluster transverse energy E_T and the e/γ E_T are copied from the corresponding $\sum E_T$ and $\sum E_{T(CE-E)}$ Cluster Sum fields. These properties are then scaled and rounded to have a specific bit-width and LSB in the specified range.

Cluster E_T fractions in various longitudinal regions

The different cluster E_T fractions are formed as per Equations 6.3, 6.4, and 6.5

$$\text{Fraction in CE-E} = \frac{\sum E_{T(CE-E)}}{\sum E_T} \quad (6.3)$$

$$\text{Fraction in core CE-E} = \frac{\sum E_{T(CE-E-Core)}}{\sum E_{T(CE-E)}} \quad (6.4)$$

$$\text{Fraction in early CE-H} = \frac{\sum E_{T(CE-H-Early)}}{\sum E_T} \quad (6.5)$$

These properties are then rounded to have a specific bit-width and LSB in the specified range.

Cluster E_T weighted mean positions

The different cluster E_T weighted mean positions for each coordinate c are formed as per Equation 6.6

$$\mu(c) = \frac{\sum Wc}{\sum W}, \text{ where } c \in \{\phi, r/z, z\} \quad (6.6)$$

These properties are then generally scaled, rounded and offset to having a specific bit-width and LSB in the specified range. The E_T weighted $\mu(\eta)$ is looked up from a dictionary of pre-calculated values that correspond to the calculated $\mu(r/z)$. Thereby, the relationship between $\mu(\eta)$ and $\mu(r/z)$ is given by Equation 6.7.

$$\mu(\eta) = \sinh^{-1} [(\mu(r/z))^{-1}] \quad (6.7)$$

Cluster shape properties

The different cluster E_T weighted standard deviations for each coordinate c are formed as per Equation 6.8

$$\sigma_{cc} = \sqrt{\frac{\sum W \sum Wc^2 - (\sum Wc)^2}{(\sum W)^2}}, \text{ where } c \in \{\phi, r/z, z\} \quad (6.8)$$

Table 6.4: Cluster Properties definition as per the HGCal→L1T interface specification.

Word	Bits	Field	#bits	Format	LSB	Range	Notes
0	13-0	E_T	14	unsigned	0.25 GeV	0 to 4.0975 TeV	Without PU subtraction
0	27-14	$e/\gamma E_T$	14	unsigned	0.25 GeV	0 to 4.0975 TeV	Without PU subtraction
0	31-28	GCT e/γ select bits	4	unsigned	-	-	placeholder definition, TBC At bit indices $\in \{0, 1, 2, 3\}$ (1 if true else 0) 0 : Fraction in CE-E > 0.5 1 : Fraction in core CE-E > 0.5 2 : Fraction in front CE-H > 0.5 3 : $e/\gamma E_T > 16$ GeV
0	39-32	Fraction in CE-E	8	unsigned	2^{-8}	0 to 0.996	Fraction of E_T in CE-E trigger layers (1-13)
0	47-40	Fraction in core CE-E	8	unsigned	2^{-8}	0 to 0.996	Fraction of E_T in core CE-E trigger layers (4-8)
0	55-48	Fraction in front CE-H	8	unsigned	2^{-8}	0 to 0.996	Fraction of E_T in front CE-H trigger layers (14-17)
0	61-56	First layer	6	unsigned	1	1 to 34	HGCal trigger layer number
0	63-62	Spare	2	-	-	-	-
1	9-0	E_T weighted $\mu(\eta)$	10	unsigned	$\pi/720$	≈ 1.4 to 3.0	Corresponding unsigned range 320 to 687
1	18-10	E_T weighted $\mu(\phi)$	9	signed 2's complement	$\pi/720$	$\approx \pm 64^\circ$	W.r.t. centre of sector Sector edges defined in text Corresponding unsigned range (-256 to 255)
1	30-19	E_T weighted $\mu(z)$	12	unsigned	0.5 mm	0 to 1929.5 mm	Measured from first layer
1	31-31	Spare	1	-	-	-	-
1	41-32	Number of TCs	10	unsigned	1	0 to 1023	Number of TCs in cluster
1	48-42	Quality flags	7	unsigned	-	-	At bit indices $\in \{0, 1, 2, \dots, 6\}$ (1 if true/valid else 0) 0 : Existence of E_T saturated TC (at ≈ 512 GeV) 1 : Quality of Fraction in CE-E 2 : Quality of Fraction in core CE-E 3 : Quality of Fraction in front CE-H 4 : Quality of σ and μ properties 5 : Saturated ϕ ($\mu(\phi)$ outside of $\approx \pm 64^\circ$) 6 : Nominal ϕ ($\mu(\phi)$ within $\approx \pm 60^\circ$)
1	63-49	Spare	15	-	-	-	-
2	6-0	E_T weighted $\sigma(E_{T(TC)})$	7	unsigned	$Range_{max}/2^7$	0 to 13.915 GeV	placeholder LSB and range, TBC $Range_{max}$ coming from a CMSSW data sample
2	12-7	Last layer	6	unsigned	1	1 to 34	HGCal trigger layer number
2	18-13	Shower length	6	unsigned	1	1 to 34	Difference between the last and the first trigger layer
2	31-19	Spare	13	-	-	-	-
2	38-32	E_T weighted σ_{zz}	7	unsigned	$Range_{max}/2^7$	0 to 778.098 mm	placeholder LSB and range, TBC $Range_{max}$ coming from a CMSSW data sample
2	45-39	E_T weighted $\sigma_{\phi\phi}$	7	unsigned	$Range_{max}/2^7$	0 to 0.12822	placeholder LSB and range, TBC $Range_{max}$ coming from a CMSSW data sample
2	51-46	Core shower length	6	unsigned	1	1 to 34	Longest consecutive number of trigger layers
2	56-52	E_T weighted $\sigma_{\eta\eta}$	5	unsigned	$Range_{max}/2^5$	0 to 0.148922	placeholder LSB and range, TBC $Range_{max}$ coming from a CMSSW data sample
2	63-57	E_T weighted $\sigma_{\frac{z}{z_{max}}}$	7	unsigned	$Range_{max}/2^7$	0 to 0.024584	placeholder LSB and range, TBC $Range_{max}$ coming from a CMSSW data sample
3	31-0	Layer bits (Low)	32	unsigned	-	-	Bitmap of layer IDs of the cluster that have TC hits (Layers 32 down to 1)
3	33-32	Layer bits (High)	2	unsigned	-	-	Bitmap of layer IDs of the cluster that have TC hits (Layers 34 down to 33)
3	63-34	Spare	30	-	-	-	-

These properties are then scaled, rounded and saturated (i.e. clipped) at both edges of the specified range. This results in these properties having a specific bit-width and LSB in the specified range. The E_T weighted $\sigma_{\eta\eta}$ is looked up from a dictionary of pre-calculated values that correspond to the calculated $\sigma_{\frac{r}{z}\frac{r}{z}}$. Hereby, the approximate relation between $\sigma_{\eta\eta}$ and $\sigma_{\frac{r}{z}\frac{r}{z}}$ is given by Equation 6.9.

$$\sigma_{\eta\eta} \approx \left| \frac{d\eta}{d(r/z)} \right| \sigma_{\frac{r}{z}\frac{r}{z}} \quad (6.9)$$

Because $\left| \frac{d\eta}{d(r/z)} \right|$ depends on r/z , the $\mu(r/z)$ is used together with $\sigma_{\frac{r}{z}\frac{r}{z}}$ to form the look-up address.

The cluster E_T weighted $\sigma(E_{T(TC)})$ is formed as per Equation 6.10.

$$\sigma(E_{T(TC)}) = \sqrt{\frac{N_{TC(W)} \sum W^2 - (\sum W)^2}{N_{TC(W)}^2}} \quad (6.10)$$

Based on the $\bigcup LayerBits$ vector, several properties describing the longitudinal shower progression are derived. These include the ID of the first and last layers with hits, the length of the shower as well as the core length of the shower. The last one is calculated as the longest array of consecutive ones in the $\bigcup LayerBits$ vector. For example, looking at a cluster having $\bigcup LayerBits = "001011011101111000000000000000000000"$ its first and last layers are layers 3 and 15, respectively. Similarly, its shower length and core shower length would be calculated as 13 and 4, respectively. The Layer bits property copies the $\bigcup LayerBits$ field from the corresponding Cluster Sum.

Miscellaneous cluster properties

The *GCT e/ γ select bits* are formed as requested by the L1T Global Correlator Trigger. They simply convey whether the three fraction properties are above the threshold of 0.5 and if the e/ γ E_T is above 16 GeV. The *Number of TCs* is copied from the N_{TC} Cluster Sum field. Remaining are the various single-bit *Quality Flags*. The *Existence of E_T saturated TC* flag is simply copied from the $\bigcup SatTC$ Cluster Sum field. The *Quality of Fraction* flags inform of potential saturation of input operands in the calculation of fractions in longitudinal regions. The *Quality of σ and μ properties* flag copies the $\bigcup ShapeQ$ Cluster Sum field. The *Nominal ϕ* flag informs the receiver if the cluster $\mu(\phi)$ is in the nominal 120° Stage 2 ϕ sector. If outside of the nominal 120° Stage 2 ϕ sector, the $\mu(\phi)$ saturates at $\pm 64^\circ$ which is indicated by *Saturated ϕ* flag⁴.

The Property Calculator achieves an output bandwidth as high as $1 \times 256 \text{ bit} \times 360 \text{ MHz} \approx 85 \text{ Gbit s}^{-1}$, given that Cluster Property records are being streamed across a single stream at the system clock frequency of 360 MHz.

6.3 The Software Model

The software model essentially implements the three step core algorithm (described in Section 6.2) and comes in two “flavours”: the *simulator* and the *emulator*. The simulator

⁴This should rarely happen, given that the cluster seeds are always found within the $\pm 60^\circ$ ϕ range.

works with full-precision float numbers, whereas the emulator works with limited precision, emulating the exact behaviour of the algorithm implemented in firmware (i.e. they are one-to-one matched on the LSB level). Additionally the simulator (unlike the emulator) is not susceptible to the choice of the k parameter, i.e. $W = E_T (\forall k)$. Thereby, one of the major use-cases of the software simulator is to have its outputs serve as a reference to what the *true* value (i.e. the most precise and accurate one) of any given cluster property should be. Specifically, the software emulator (and the firmware implementation) of the cluster property algorithm is always bound to be less precise than the simulator due to computation noise introduced by quantisation effects and implementations of the algorithm when $W \neq E_T$ (i.e. $k > 0$). Hence, an important aspect in the development of the Cluster Properties block is to bring the behaviour of the software emulator (and the firmware implementation) as close as possible to the one of the software simulator while abiding to the usual FPGA design constraints (e.g. algorithm resource utilisation, latency and bandwidth; as mentioned in Section 6.1).

There are a couple of additional reasons why having such an organisation of the development environment of a firmware algorithm is useful. Firstly, the software model (both simulator and emulator) is very flexible, allowing: rapid testing of novel cluster properties; quickly processing relatively large data samples (e.g. $O(100 \times 10^3)$ of clusters); and studying the physics performance of the algorithm, without having to go through all the different steps specific to firmware development (e.g. FPGA algorithm simulation, synthesis, implementation, timing closure, hardware tests, etc.). Secondly, having the software emulator available allows behavioural comparison of the implemented firmware to it (e.g. they both need to act the same on the LSB level). This is very useful for reassuring that the algorithm implemented in firmware is indeed the same as originally conceptualised in the software model.

This section is organised into two parts, reporting on the most significant results coming from the software model. Firstly, it presents the level of agreement between the software simulator and the software emulator. In other words, it characterises the degree of quantisation-induced computation noise in the final cluster properties. Secondly, it evaluates the physics performance of the cluster properties when they are used as features when training/testing two simple BDTs. Each of these two BDTs is used for binary classification to distinguish signal (photon and VBF jet, respectively) clusters from PU clusters. Of course, there was extra software developed in addition to the central software model of the Cluster Properties core algorithm to reach the findings discussed in any of these parts. For example, this includes: an emulator vs. simulator comparison tool; and the BDT models.

This section focuses on the optimal possible performance that the Cluster Properties block can deliver, hence k parameter will usually be set to $k = 0$ when working with emulated clusters. Since the BDT scores (i.e. TPR and FPR; both explicitly defined later) are chosen as *the metrics* used to evaluate the overall physics performance of the implemented properties, a short note will be given to illustrate the impact of the k parameter choice on the overall performance. This can serve as a useful reference in case it is required to reduce the size of the Modified TC and Cluster Sum records to save Stage 2 FPGA resources.

6.3.1 Agreement between the Emulator and the Simulator

Plotting 2D histograms of several cluster properties calculated by the software emulator (reconstructed) vs. the software simulator (true) allows evaluating the degree of quantisation-induced computation noise (Figure 6.4). The shown properties come from each one of the four main categories (as described in Section 6.2, excluding the miscellaneous group). Below each 2D histogram there is a Δ plot showing the difference between the emulated and the simulated values as a function of the simulated value. Furthermore, to the right of each Δ plot is a marginal histogram of the Δ values. Each plot also has a statistics summary box, giving the values of the mean, the standard deviation and the median of the Δ . About 800×10^3 clusters in total were used to produce each plot. The number of bins per axis is usually chosen as 2^b , where b is the bit-width of any given property or 2^7

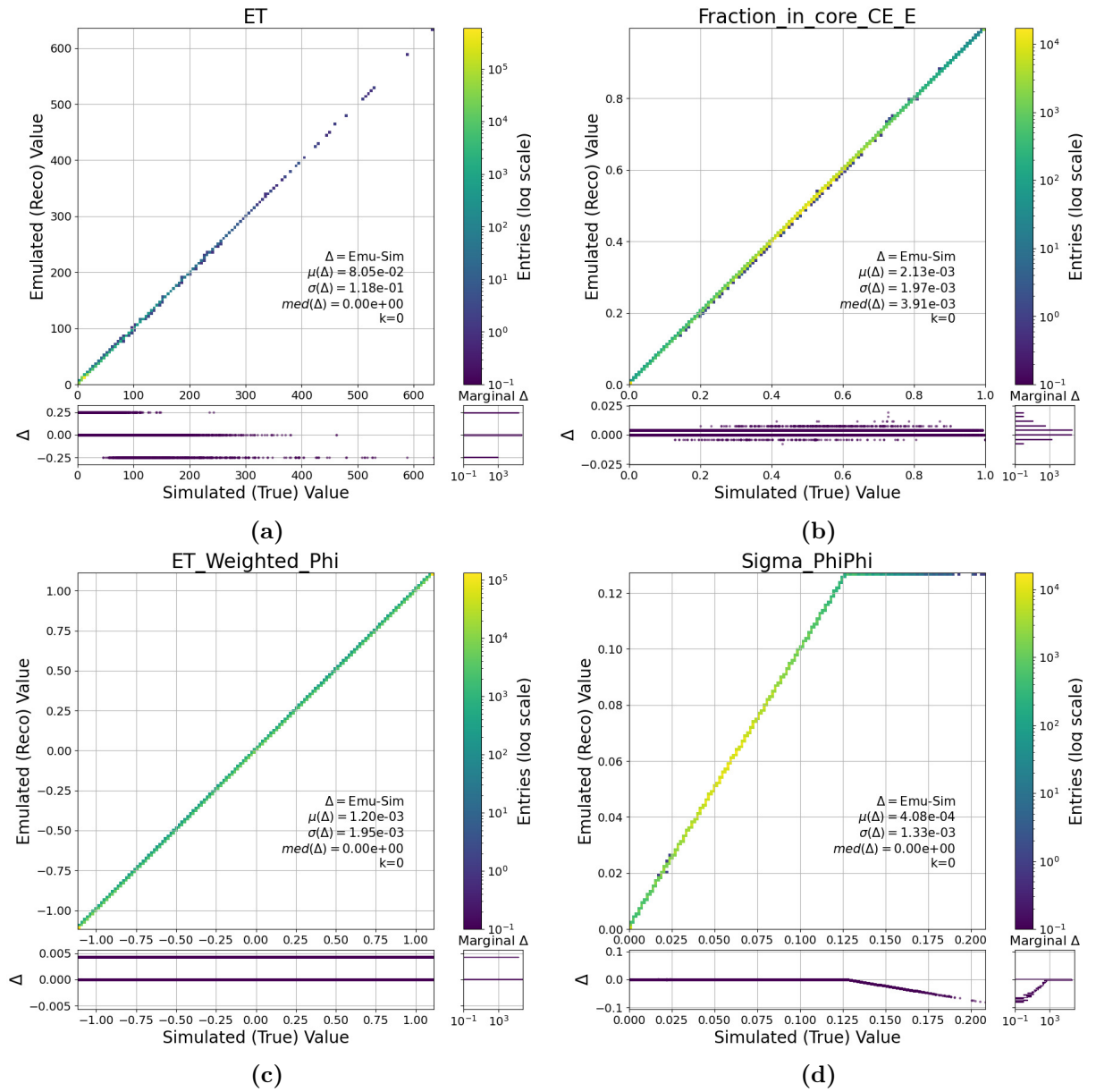


Figure 6.4: Agreement between the calculated quantities by the software emulator (firmware implementation/reconstructed) with the software simulator (true) for several cluster properties. For axis units, see Table 6.4. More details can be found text.

for properties whose $b > 7$ (about 100 bins give sufficient resolution for the axis of values of a histogram). Similar plots from this study about additional cluster properties can be found in Appendix B.

All plots show expected dependence between the emulated and the simulated values. This dependence is mostly linear with the exception of standard deviations which are designed to saturate when wider than a programmable threshold (as specified in Table 6.4). Occasionally, the former are offset from the latter by up to several LSBs of a given property. Specifically, for more than 99 % of the studied clusters, the emulated values are not further away from the simulated values by more than one LSB for most of the properties. For the E_T weighted $\mu(z)$ and $\sigma_{\eta\eta}$ properties, this difference was slightly worse, but it still was not greater than two and three LSBs in more than 95 % of the studied clusters. This allows qualifying the behaviour of the software emulator (and firmware implementation) in terms computation noise induced by quantisation effects as very good. Visualising the additional negative effects in terms of computation noise caused by $k > 0$ is available in Appendix C for an example cluster shape property and several values of k .

6.3.2 Physics Performance of the Cluster Properties

The goal of this study is to evaluate the physics performance of the implemented cluster properties when used as features in two simple binary classification models used to distinguish: single photon from PU clusters; and single VBF jet from PU clusters. Specifically, two BDT classifiers are used without optimising any of their hyper-parameters (e.g. the number of estimators, maximum tree depth, maximum number of leaves, learning rate, etc.) for performance. Although the L1T may not use BDTs, the performance of the models presented here can serve as a proxy for whatever the final system implements. Three data samples are used in total, where each stems from CMSSW and simulates the HGCal TPG response to: only the PU of 200 parallel proton-proton interactions (PU200); both PU200 and a photon gun; both PU200 and VBF jets stemming from a VBF process (with $H \rightarrow \tau\tau$ in the vertex). The reason for simulating the photon and VBF jets (signal) together with PU200 (background) is to ensure working with more realistic signal clusters. For example, when there is a cluster in real detector operation, it is highly likely that it will have background contributions in addition to the ones of the signal itself. The study uses PU clusters only from the first file, whereas the photon and the VBF jet clusters are taken from the latter two files. Both a photon or a VBF jet cluster is defined as the cluster whose distance Δr (defined as $\Delta r = \sqrt{\Delta\phi^2 + \Delta\eta^2}$) of its reconstructed mean position in the (η, ϕ) space from the generator-level position of the original photon and the VBF jet is $\Delta r < 0.2$. If multiple clusters pass such a requirement, then the most energetic cluster is chosen. The following presents a parallel analysis of photon and VBF jet identification studies, summarising the key aspects of each. The main narrative will lean on the ideal case when working with full precision floats (i.e. when using the software simulator) but the final results will also be shown for the limited precision case (i.e. when using the software emulator).

General workflow in the study

The format of the input data of the study are clustered TCs (as seen in Table 6.1) with a couple of fields added to the TC record. The first is the *label* field, indicating the “origin” of the corresponding cluster (i.e. PU, photon or VBF jet). The second is the generator-level

p_T (p_T^{gen}) of the simulated photon or the VBF jet. The clustered TCs are played through the three-step algorithm (described in Section 6.2) of the software model (both simulator and emulator flavours) calculating the various physical properties about the clusters (as seen in Table 6.4). The *label* and the p_T^{gen} fields are also saved for each cluster. Afterwards, some clusters (e.g. the ones with invalid values of *Quality Flags*) and properties (e.g. $\mu(\eta)$ and $\mu(\phi)$) are dropped from the study. Next, it is ensured that both PU and signal clusters appear in the same p_t^{gen} and p_t^{reco} ranges by applying appropriate cuts to the data. This is followed by re-weighting the E_T^{reco} distributions so that the one of signal matches the one of PU clusters. Both re-weighting or property dropping is generally done to ensure that the BDT does not learn to discriminate based on some shapes of the distributions of these properties which are not realistic (i.e. distribution shapes that are a consequence of simulation specifics). Two BDT binary classifiers are then trained (using 20 % of the data) and tested (using 80 % of the data), each trying to distinguish signal (i.e. photon or VBF jet / interesting to trigger on) from background (i.e. PU / not interesting to trigger on) clusters, while using the cluster properties as discriminating features. The training is done using the clusters across the entire E_T^{reco} range of interest, while the testing is done in p_T^{gen} sub-ranges. The *label* field is required for training/testing the two BDT models. The p_T^{gen} field is used to check the correlation between the reconstructed cluster E_T (E_T^{reco}) against cluster p_T^{gen} and to select subsets of clusters when testing the BDTs in different p_T^{gen} sub-ranges. Finally, the performance of the BDTs (hence, the implemented cluster properties) is evaluated relying on usual ML metrics such as the: Area Under Curve (AUC) score of the BDT Receiver Operating Characteristic (ROC) curve; True Positive Rate (TPR); True Negative Rate (TNR); False Positive Rate (FPR); and False Negative Rate (FNR). Hereby, the notion of a “positive” and “negative” outcome is equivalent to classifying a cluster as a signal and PU, respectively. The four rates are defined as (Equations 6.11, 6.12, 6.13 and 6.14)

$$TPR = \frac{TP}{(TP + FN)} \quad (6.11)$$

$$TNR = \frac{TN}{(TN + FP)} \quad (6.12)$$

$$FPR = \frac{FP}{(FP + TN)} \quad (6.13)$$

$$FNR = \frac{FN}{(FN + TP)} \quad (6.14)$$

where TP, FN, TN and FP are the total counts of true positive, false negative, true negative and false positive outcomes of the binary classification tests, respectively. Given that $TPR + FNR = 1$ and $TNR + FPR = 1$, it is sufficient to study only two out of these four rates. Namely, this study focuses on the TPR and the FPR .

Limitations and possible optimisations of the study

There are several limitations of this study. Firstly, due to the absence of generator-level information in the available data about PU clusters, this study *does not* filter out photon or VBF jet clusters that appear when simulating PU from either of the three samples, but always treats them as background. On the other hand, the photon and VBF jet clusters from the latter two files are always treated as signal. This means that the BDT binary classifier may flag a true photon or a VBF jet cluster stemming from simulated PU as signal. However, that cluster would be treated as misclassified because it is not labelled

by the BDT as background, but instead labelled as signal when testing the BDT. These situations would consequently lead to an increase in the measured FPR (even though they should have contributed to the increase in the measured TPR) when evaluating the performance of the BDT. Additionally, to allow testing the BDT in p_T^{gen} sub-ranges, the unavailable p_T^{gen} information about PU clusters is copied from the available cluster E_T^{reco} . Secondly, the correlation between the p_T^{gen} and E_T^{reco} of VBF jets could be optimised by either using a different Δr cut or by allowing more than just the most energetic cluster that makes the Δr cut into the VBF jet cluster definition. Thirdly, this study does not use trigger primitives from other CMS sub-systems (e.g. the Tracker or the Barrel Calorimeter) and is focused only on single photon or single VBF jet clusters that occur in one HGCAL Stage 2 ϕ sector (of a single HGCAL endcap). The second to last limitation is that the re-weighting is done only in E_T , whereas multi-dimensional re-weighting should ideally be applied to the samples and some of the dropped properties kept. For example, one could consider performing the re-weighting in both E_T and $\mu(\eta)$ and allowing both of these properties to enter the study, instead of re-weighting only in E_T , allowing E_T to be used and dropping $\mu(\eta)$. Lastly, the study uses a simple (i.e. without hyper-parameter optimisation) BDT as the binary classification model, although other models or even more careful optimisation of a BDT could bring better overall performance of the model. Hence, the performance of the algorithm presented here can be considered as a viability proof-of-principle and conservative. In other words, the final performance can probably be improved if any of these limitations is addressed more carefully.

Dropping some clusters and properties before training the BDT

There are two categories of conditions for which clusters would be dropped from the study before proceeding to training/testing the BDT model. Firstly, to avoid working with clusters in the vicinity of the edge of a Stage 2 ϕ sector, the clusters with *Nominal* $\phi = 0$ or *Saturated* $\phi = 1$ are dropped (such a cluster would be calculated well in a different Stage 2 FPGA responsible for the neighbouring sector). Secondly, to avoid working with clusters whose E_T fractions as well as the E_T weighted mean positions and standard deviations are possibly not calculated well, the clusters whose *Quality of Fraction* = 0 and *Quality of σ and μ* = 0 are also dropped.

There are several categories of fields from the cluster property record that are dropped from the clusters before training/testing the BDT model. Firstly, these are the various flags under the *Quality flags* field that generally inform if a given cluster property was calculated well and are not intended to directly contribute when discriminating between signal and background clusters. The second category consists of the fields that are directly related to cluster energy such as E_T , $e/\gamma E_T$ and the third *GCT e/γ select bit*. However, the E_T field itself is not dropped because the samples were re-weighted by it. In the third category are the remaining *GCT e/γ select bits* which do not bring any new information given that the E_T fractions are there. Lastly, the cluster E_T weighted $\mu(\eta)$ and $\mu(\phi)$ are also dropped. The reasoning for dropping the properties in the second and the last category are the same as the ones given for performing the re-weighting in E_T . Ideally, multi-dimensional re-weighting would be applied to the samples and these properties kept. However, this was considered as fine-tuning and out of the scope of this study.

Selecting the clusters by p_T^{gen} and E_T^{reco}

This study focuses on clusters whose $p_T^{gen} < 100$ GeV for two reasons. Firstly, this range is considered to be more difficult to separate signal from background clusters. For example,

with an increase in cluster p_T^{gen} , the signal clusters generally become narrower in η and ϕ . This simplifies their separation from average p_T^{gen} PU clusters (which are generally wider in η and ϕ than signal clusters). Moreover, with an increasing E_T of the clusters, the L1T is expected to base its trigger decision more on the cluster E_T itself, and less on other cluster properties (e.g. in the extreme case, the L1T will issue a trigger for any cluster with $E_T > 500$ GeV). Secondly, the number of PU clusters in the available data sample becomes very small (e.g. in the $O(100)$) already beyond 50 GeV. Although there are plenty of available signal clusters beyond 100 GeV, it was decided to limit the study to below 100 GeV in order to avoid having very imbalanced numbers of signal and background clusters.

By looking into 2D histograms of cluster p_T^{gen} vs. E_T^{reco} (Figure 6.5), it is possible to check the quality of definitions of clusters with respect to the generator-level particles. Ideally,

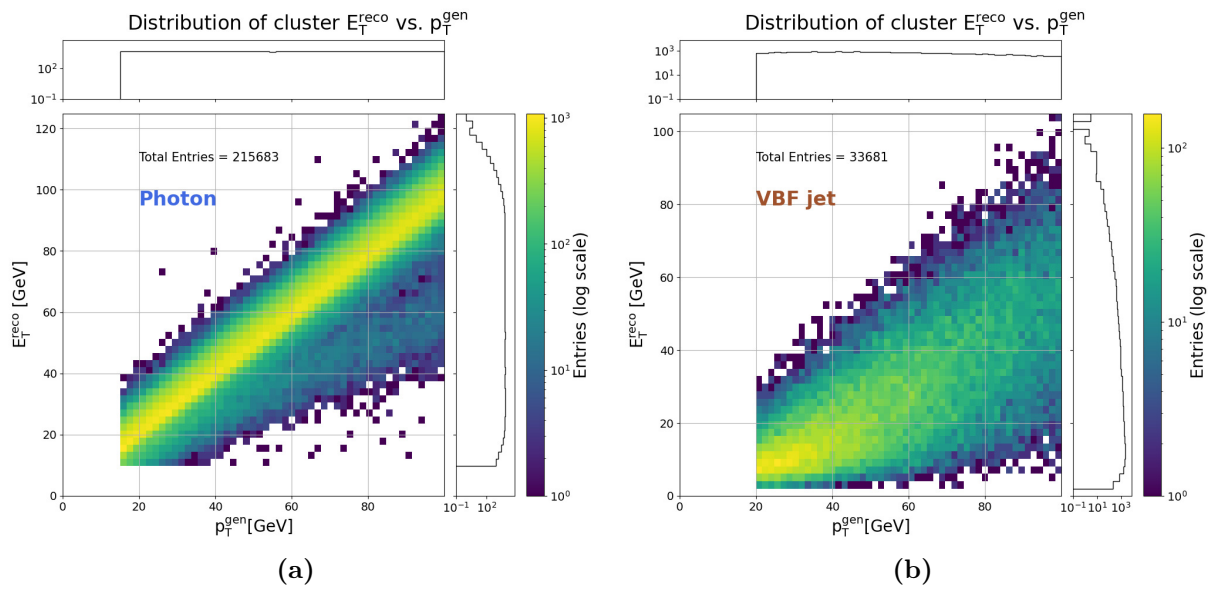


Figure 6.5: Distributions of cluster E_T (reconstructed) as a function of p_T (generator-level) for (a) photon and (b) VBF jet clusters. More details can be found in text.

the two would be one-to-one matched. For photons, this correlation is very good, whereas for the VBF jets it is worse but still considered acceptable for the purpose of this study. Due to the PU clusters lacking generator-level information, the approach was to artificially add p_T^{gen} by copying the available E_T^{reco} values. As it will be shown later, this information about clusters (of any class) is required when testing the BDT model in p_T^{gen} sub-ranges.

When compared to the PU samples whose E_T^{reco} and p_T^{gen} range from 10 GeV to 100 GeV, the E_T^{reco} vs. p_T^{gen} histograms in Figure 6.5 also show an imbalance of entries between the signal and PU clusters in the lower parts of the p_T^{gen} and E_T^{reco} ranges. To resolve this, the strategy was to apply cuts to the data such that signal clusters appear in the same p_T^{gen} and E_T^{reco} range as pileup clusters. This resulted in constraining the photon identification study to $p_T^{gen} \in (15 \text{ GeV}, 100 \text{ GeV})$ and $E_T^{reco} \in (10 \text{ GeV}, 100 \text{ GeV})$. Similarly, the VBF jet identification study was constrained to $p_T^{gen} \in (20 \text{ GeV}, 100 \text{ GeV})$ and $E_T^{reco} \in (10 \text{ GeV}, 100 \text{ GeV})$. For reference, the CMS-TDR-021 [20] mentions working with transverse energies above 15 GeV and 30 GeV for e/γ and jet triggers, respectively.

One can plot the distributions of p_T^{gen} and E_T^{reco} for clusters that pass these cuts and enter the study (Figure 6.6). The photon cluster distributions (both in p_T^{gen} and E_T^{reco})

are generally flat (because the generator gun covered the dynamic range uniformly in the simulation). For VBF jet clusters, the distribution in p_T^{gen} is mostly flat (except in the lower end of the range). The corresponding E_T^{reco} distribution is not matching the p_T^{gen} as well as in the photon case but this is an artefact of the VBF jet cluster definition. The PU distributions both take on the expected physical shape (i.e. falling exponentially with an increase in p_T^{gen} or E_T^{reco}).

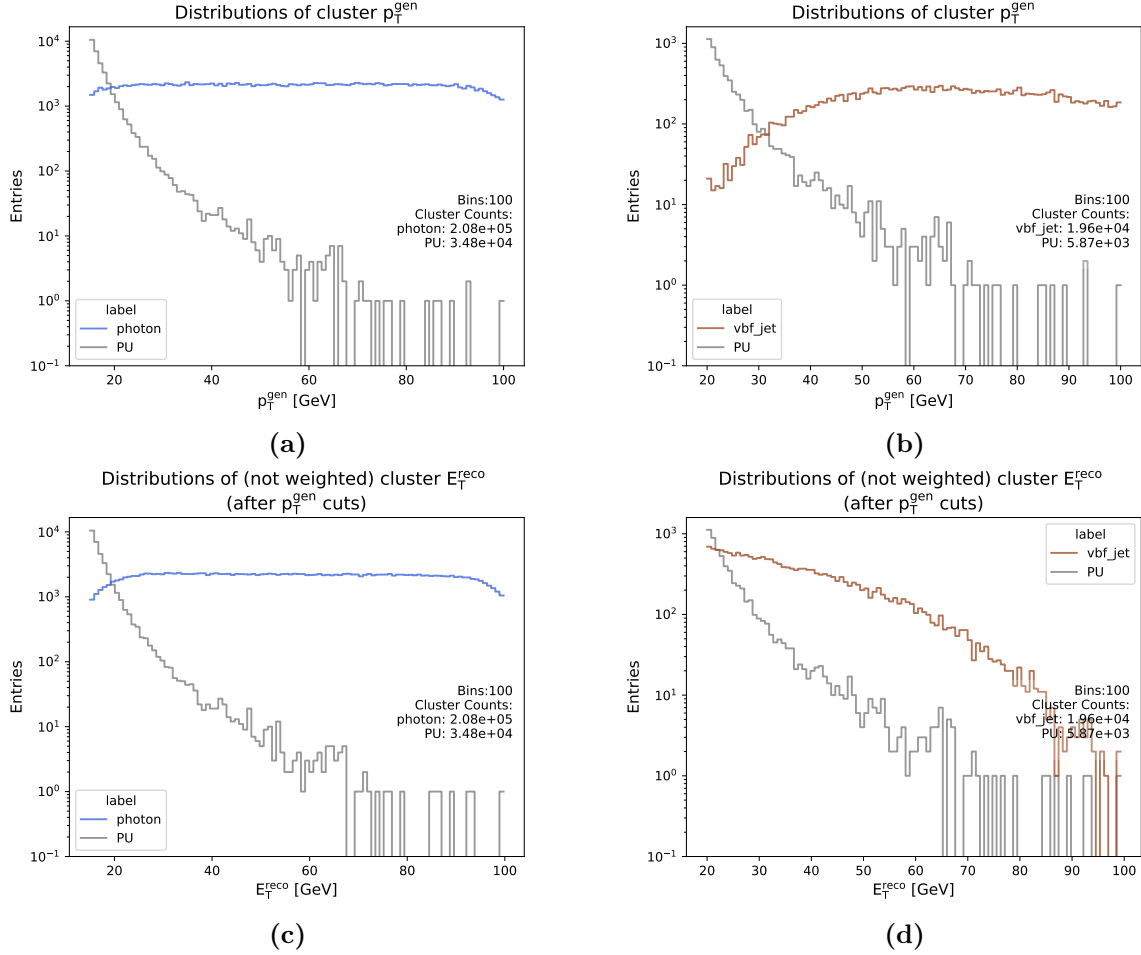


Figure 6.6: Distributions of cluster p_T^{gen} (Top) and E_T^{reco} (Bottom) for the photon (Left) and VBF jet (Right) identification studies after applying p_T^{gen} and E_T^{reco} cuts. More details can be found in text.

Bin-based sample re-weighting in cluster E_T^{reco}

To prevent the BDT model from learning about the specific shapes of E_T^{reco} distributions when separating signal from background, bin-based sample weights are applied to match the E_T^{reco} distributions of signal to those of the background (Figure 6.7). This approach ensures that the trained BDT model: does not rely on the E_T^{reco} distribution differences between signal and background clusters in simulated data; allows for E_T^{reco} variations in real physical distributions.

Specifically, the `hep_ml.reweight.BinsReweigher(n_bins, n_neighs)` python class was used for the bin-based sample re-weighting [71]. The first argument `n_bins` specifies the number of bins to be used in the range of interest. The second argument `n_neighs`

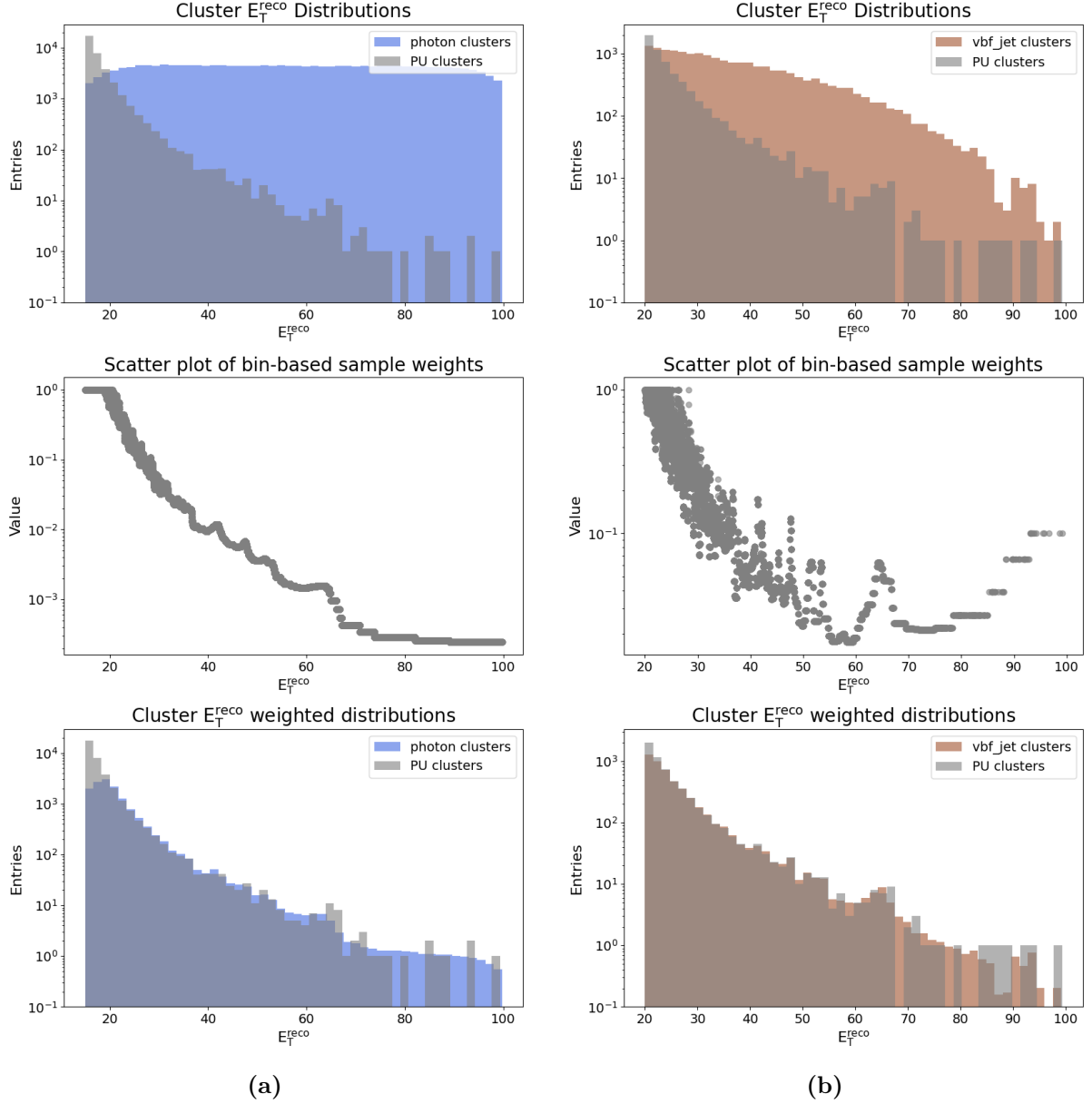


Figure 6.7: Plots demonstrating the mechanism of bin-based sample weighting of cluster E_T^{reco} distributions so that the one of signal matches the one of pileup clusters. Shown are the original E_T^{reco} distributions (**Top**), computed bin-based sample weights (**Middle**), and E_T^{reco} distributions after applying weights (**Bottom**) for the photon (**Left**), and the VBF jet (**Right**) identification studies. More details can be found in text.

specifies the size (in bins) of the Gaussian filter for the smoothing step (so that the re-weighting coefficient includes information from other bins in the neighbourhood of the nominal bin). Also, the bin sizes are not even but they use quantiles instead. Due to using both quantiles and Gaussian smoothing, the bottom pairs of distributions do not give an exact match.

Organising the clusters into p_T^{gen} sub-ranges

The chosen strategy for testing the BDT model is to do so in sub-ranges of p_T^{gen} . However, their organisation will not include equidistant sub-range edges, due to data sparsity of the PU sample. For example, there are bins with very few or no PU cluster samples, especially in the higher halves of the relevant p_T^{gen} ranges (as visible in the distributions of Figure 6.6 from earlier). Hence, the chosen approach was to slice the p_T^{gen} range into bins of monotonically increasing width (with the increase of p_T^{gen}) and having each of the bins be wide enough to have at least O(100) entries. Specifically, this resulted in having seven and five p_T^{gen} sub-ranges in total for the photon and VBF jet identification studies, respectively (Figure 6.8).

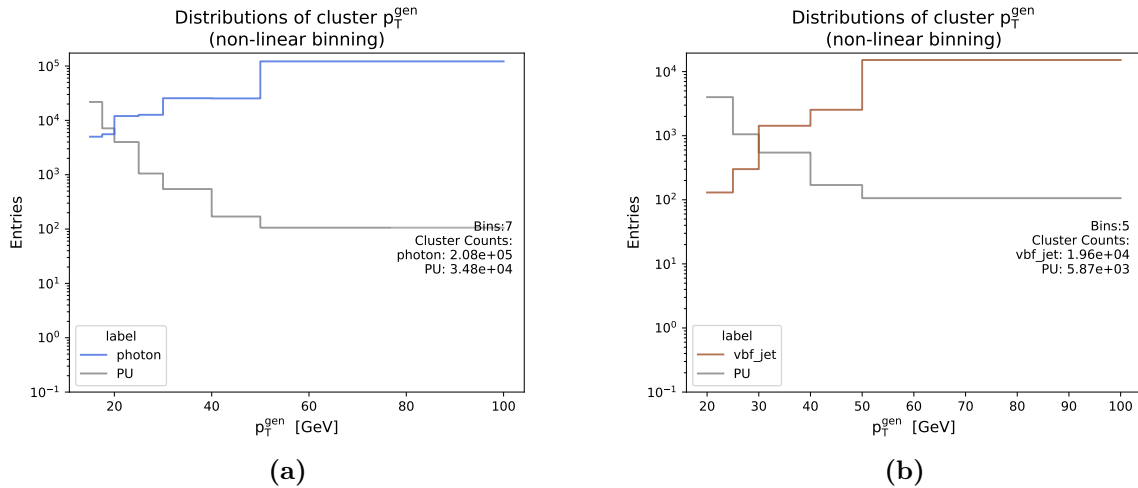


Figure 6.8: Distributions of cluster p_T^{gen} for the (a) photon and (b) VBF jet identification studies after organising the p_T^{gen} into sub-ranges with non-linear binning. More details can be found in text.

Importance of features in the trained BDT

There was one BDT binary classifier model trained for each one of the two conducted signal identification studies. This allows to visualise the relative feature importance⁵ that the BDT gave to different cluster properties when identifying signal vs. background (Figure 6.9). In the photon identification study, there are several cluster properties that were given non-negligible importance amongst which the E_T weighted $\mu(z)$ was given by far the biggest importance. In the VBF jet identification study, it is more difficult to distinguish signal clusters from PU clusters. This caused the BDT to give non-negligible importance to most of the available cluster properties.

⁵The feature importance (gain) represents the improvement in the error reduction brought by a feature. High relative importance values (close to 1) indicate that the feature is very important in the model, whereas low relative importance values (close to 0) indicate low importance.

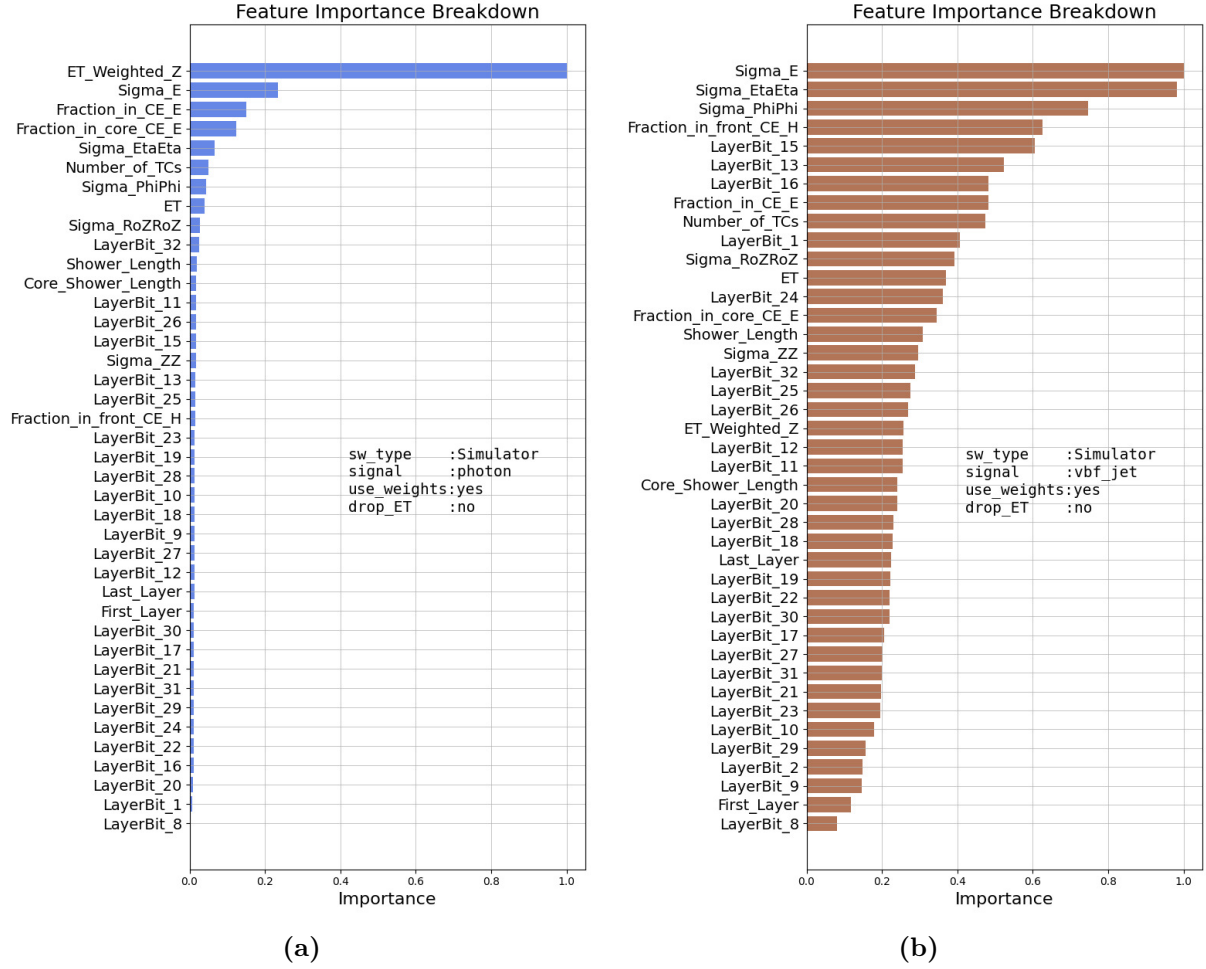


Figure 6.9: Breakdown of importance of different cluster properties used as features when training a BDT for (a) photon and (b) VBF jet identification studies. More details can be found in text.

Discrimination power of the top-performing features

One can also plot a correlation pair-plot matrix to visualise the discrimination power of some of the top performing cluster properties from each of the two studies (Figure 6.10). The plots on the principle diagonal are 1D histograms of cluster properties that allow visualising the discrimination power of individual properties. The plots off the principle diagonal are correlation plots that allow visualising how any pair of properties separates the two populations (e.g. PU from signal).

Unlike PU clusters, the electromagnetic showers initiated by a photon usually deposit all of their energy in the electromagnetic section of HGCal. Additionally, the photon clusters are expected to be narrower in η and ϕ than PU clusters. Therefore, it is more likely to be seeing photon signatures than the ones of PU for: low values of the E_T weighted $\mu(z)$; large values of *Fraction in CE-E* and *Fraction in CE-E-Core*; and low values of $\sigma_{\eta\eta}$. All of these expectations are in good agreement with the content shown in the plots. Additionally, a good discriminator is $\sigma(E_{T(TC)})$, which on average does not take on higher values if the cluster stems from photons.

Separating VBF jet clusters from PU clusters is more difficult because their underlying constituents are very similar. Some key expected differences are that the VBF jet clusters

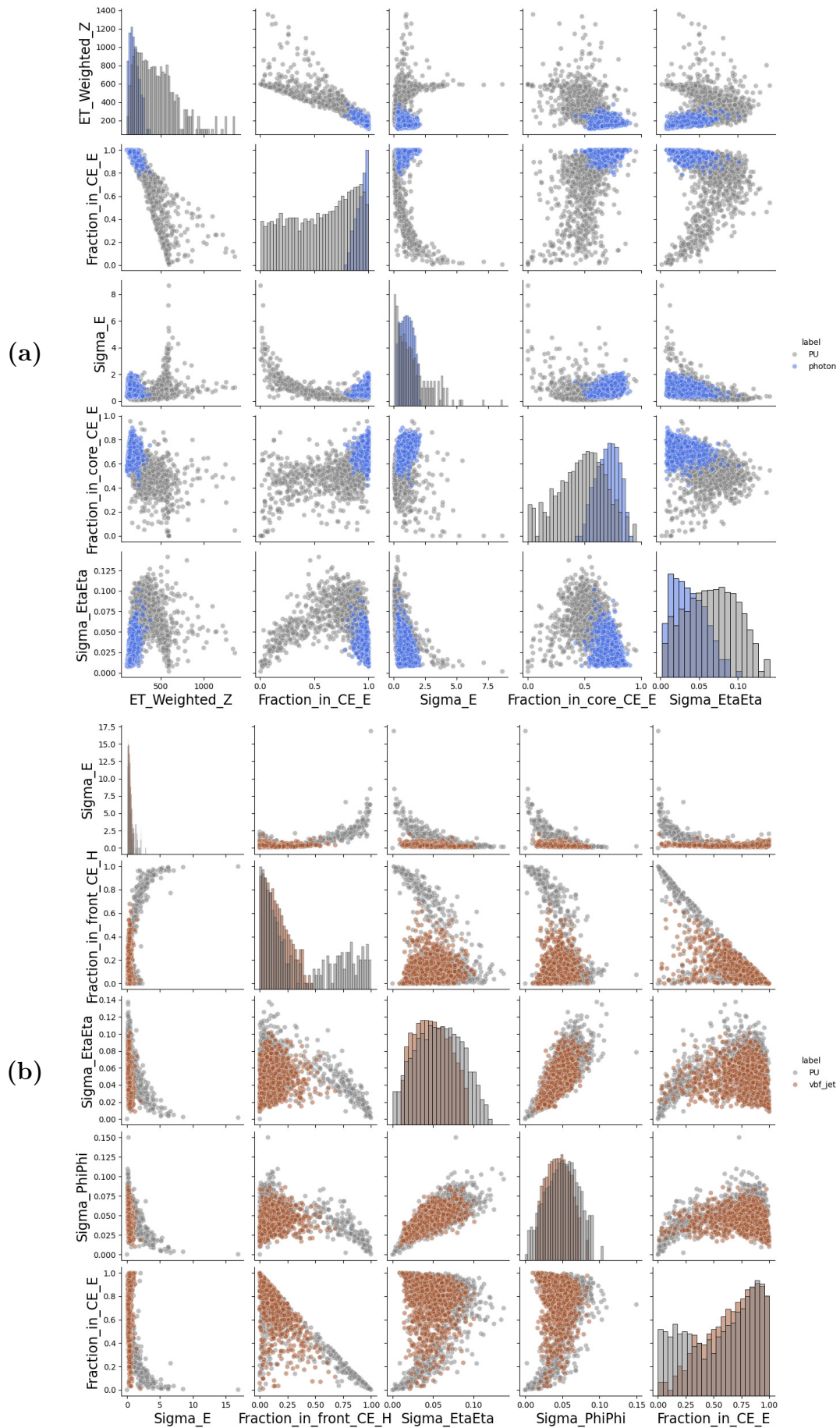


Figure 6.10: Pair plots allowing to visualise the discrimination power of some of the top-performing cluster properties in the: (a) photon (b) VBF jet identification studies.

will on average be more narrow in η and ϕ and that the PU clusters would much more rarely reach the latter layers of the HGCAL hadronic section. Therefore, it is more likely to be seeing VBF jet signatures than the ones of PU for: low values of $\sigma_{\eta\eta}$ and $\sigma_{\phi\phi}$; low values of *Fraction in CE-H-Front*. All of these expectations are in good agreement with the content shown in the plots. Additionally, good discriminators are $\sigma(E_{T(TC)})$ and *Fraction in CE-E*, which on average does not take on higher and lower values if the cluster stems from VBF jets, respectively.

Testing the BDT

A common metric for evaluating the performance of a BDT model is the quality of its ROC curve, specifically its AUC score. The ROC AUC score can range between 0.5 and 1.0, where the former corresponds to a random model and the latter to a perfect model. One can visualise the ROC curves (and their corresponding AUC scores) of the two binary classifiers as a function of the cluster p_T^{gen} (Figure 6.11). The p_T^{gen} sub-ranges correspond

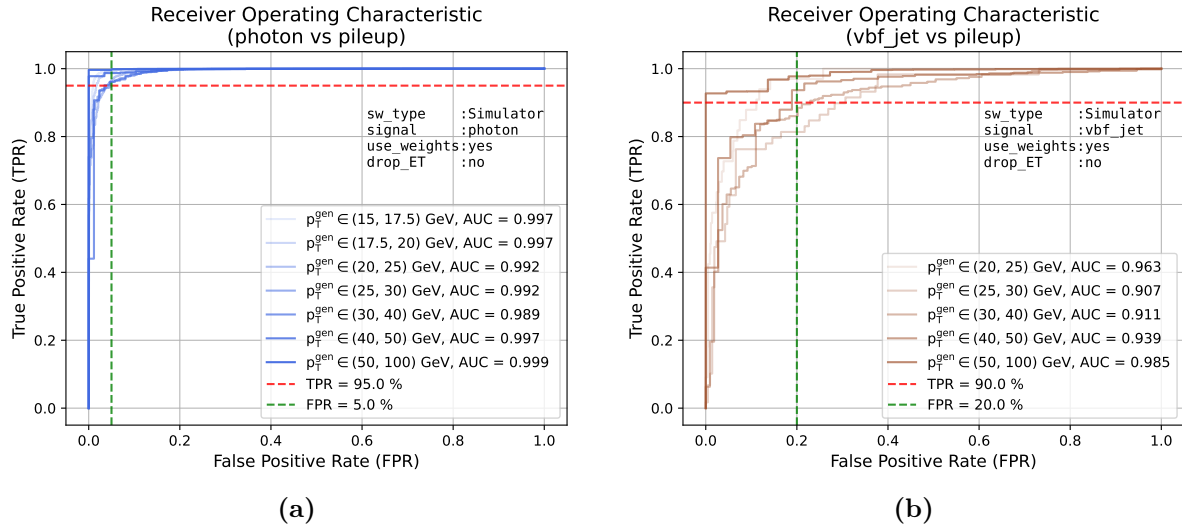


Figure 6.11: The ROC curves as a function of p_T^{gen} and their AUC scores listed for (a) photon and (b) VBF jet identification binary classifiers. The meaning of dashed lines is clarified in the discussion of the subsequent figure. More details can be found in text.

to non-linear binning specified earlier. For any cluster p_T^{gen} , the ROC AUC scores in the photon identification binary classifier are excellent (i.e. $AUC > 0.95$), whereas the ones in the VBF jet identification classifier are very good (i.e. $AUC > 0.90$).

It is important to note that a model can have an excellent AUC score, but that this score alone does not provide sufficient information for characterising the quality of the model in the context of triggering. The critical metrics to know are the TPR and the FPR of the model. These depend on the operating point of the model, which is the choice of the user of the model and can be any one of the points on the ROC curve. For example, one can set the operating point in the upper right corner of a ROC curve where the TPR is excellent, but that does not mean much since the FPR there is bad (and vice versa, if the operating point is set in the bottom left part of a ROC curve). Hence, the choice of the operating point is a compromise between achieving a high enough TPR while retaining a low enough FPR. Therefore, the final metrics chosen for evaluating the performance of the implemented BDT models (thus the cluster properties used as their features) in this study are the TPR and the FPR as a function of the cluster p_T^{gen} and for specified FPR and

TPR values, respectively (Figure 6.12). Given the excellent shape of the ROC curve of the photon identification model, the specified FPR and TPR values used in the plots are $\text{FPR} = 5\%$ and $\text{TPR} = 95\%$. On the other hand, with the VBF binary classifier having somewhat worse shapes of the ROC curves, the specified FPR and TPR values used in the plots are $\text{FPR} = 20\%$ and $\text{TPR} = 90\%$. These plots can be interpreted together with the previously shown ROC curves. Specifically, the TPR and FPR entries in Figure 6.12 correspond to the intersections of the dashed green (vertical) and red (horizontal) line with the ROC curves in Figure 6.11, respectively.

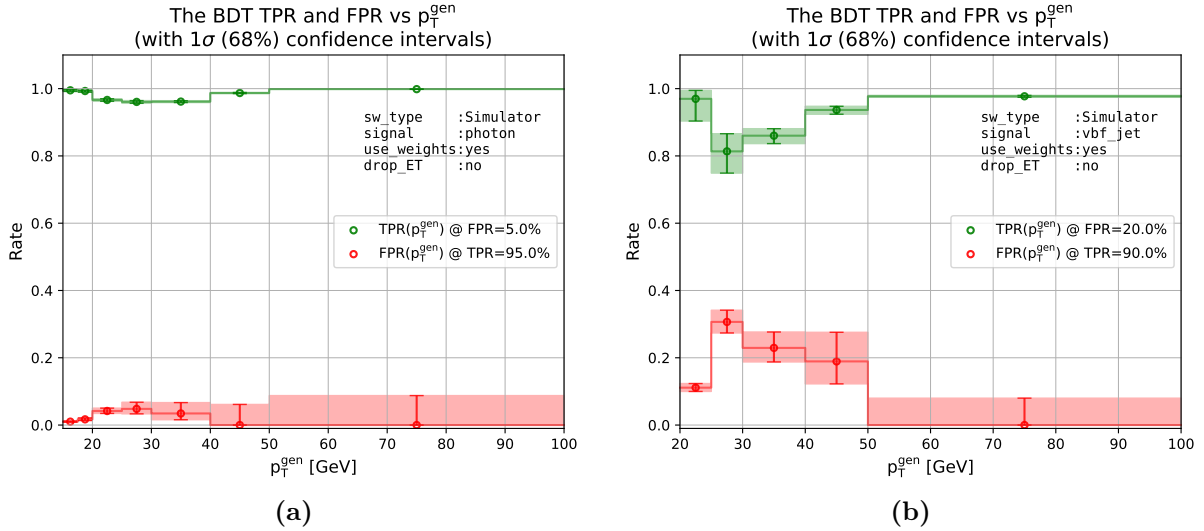


Figure 6.12: The TPR (green) and FPR (red) as a function of p_T^{gen} (for their respective FPR and TPR requirements) for the (a) photon and (b) VBF jet identification binary classifiers. More details can be found in text.

The TPR and FPR values are excellent across the whole tested p_T^{gen} range in the case of the photon identification study. The corresponding values in the VBF jet identification study are very good across the whole tested p_T^{gen} range. It is interesting to notice that there is a performance minimum for each of the binary classifiers in the first bin below 30 GeV, but understanding this effect in further detail is beyond the scope of this study. A general comment can be made that this is where the signatures of PU and signal clusters in HGCAL (as per their definitions in this study) are the most like one another.

Quantisation effects

Everything reported up to this point related to the physics performance of the cluster properties assumes they are calculated with full-precision (i.e. using the software simulator). This part reports only the final results of the study, when the properties are calculated with limited-precision (i.e. using the software emulator / firmware implementation).

To begin with, the previous four plots are reproduced for $k = 0$, when the quantisation noise induced while calculating the cluster properties is the lowest (Figure 6.13). In comparison to the corresponding results when the full-precision was used, these results are effectively the same (i.e. generally within the one σ uncertainty intervals of one another). This demonstrates excellent agreement between the software model and the software emulator (thus the real firmware implementation) in terms of the resulting physics performance of the cluster properties, and is as expected given the results of Section 6.3.1.

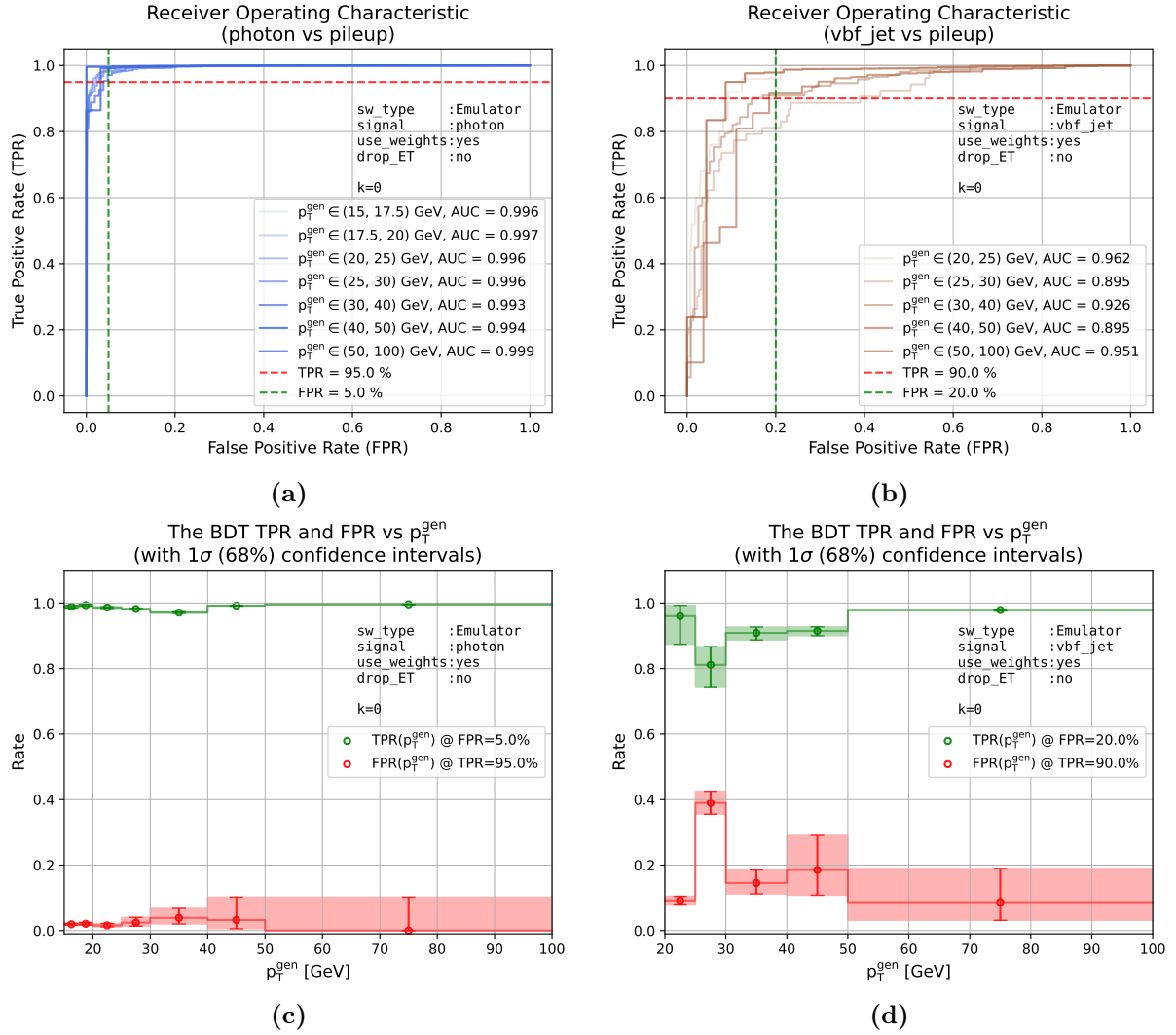


Figure 6.13: The ROC curves as a function of p_T^{gen} and their AUC scores (**Top**) and the corresponding TPR as well as FPR plots (**Bottom**) for the photon (**Left**) and VBF jet (**Right**) identification binary classifiers. Calculated using the software Emulator (the k parameter is specified in the text box). More details can be found in text.

The other important aspect in terms of quantisation effects is to explore what happens to the physics performance of the algorithm when the software emulator (real firmware implementation) is configured to work with $k > 0$. The performance study was repeated for $k \in \{1, 2, \dots, 15\}$ (i.e. all the possible values of k were tried out). The final results of the performance study (i.e. TPR and FPR plots as a function of p_T^{gen}) for several values of k are available in Figure 6.14. In comparison to the corresponding results when $k = 0$, these still do show very comparable performance. This demonstrates that (at least for the binary classification problems explored in this chapter) it is possible to retain similar physics performance of the implemented cluster properties, while working with Cluster Sum records whose bit-widths span between 254 (for $k = 15$) and 434 (for $k = 0$) bits. For example, this allows reducing the overall block RAM utilisation in the HGCAL TPG BE Stage 2 target FPGA (VU13P) to about 45 % (for $k = 15$) down from the original 58 % (for $k = 0$). Similarly, the highest possible Accumulator input and output bandwidths would be reduced to about 6.5 Tbit s⁻¹ and 90 Gbit s⁻¹ (for $k = 15$) down from the original 11 Tbit s⁻¹ and 150 Gbit s⁻¹ (for $k = 0$), respectively. Consequently, this can simplify the

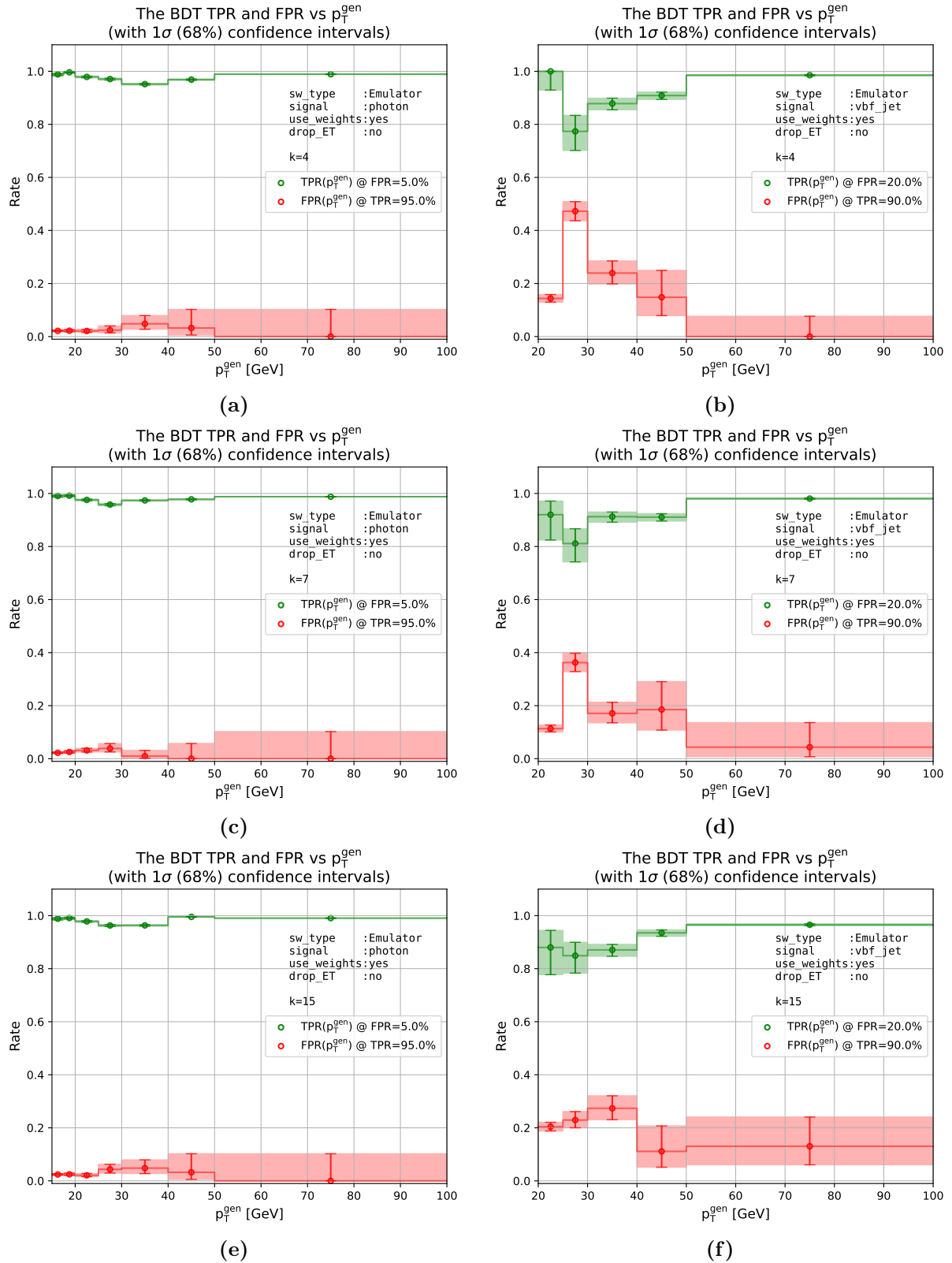


Figure 6.14: Visualising the effect of the k parameter on the performance of cluster properties when: $k = 4$ (Top), $k = 7$ (Middle), and $k = 15$ (Bottom). Shown for the two studied binary classification problems: photon vs. pileup (Left) VBF jet vs. pileup (Right).

FPGA implementation of the overall Stage 2 design (i.e. in terms of routing, timing closure, resource utilisation, bandwidth constraints, etc.)

6.4 The Firmware Implementation

There are three perspectives in the focus of this section, regarding the firmware implementation of the algorithm for calculating cluster properties. Firstly, a general overview of the implemented firmware block is given (via its top-level interfaces, time-diagrams, resource utilisation and timing closure reports, etc.) Secondly, the section reports on the level of agreement between the implemented firmware block and the software emulator. Lastly, a brief report is given on the HGCal TPG BE $\leftrightarrow$ L1T interface testing campaign (where the Property Calculator took part).

The overall implementation: defined the modified TC record; defined the Cluster Sum record, defined the Cluster Property record; defined the Accumulator “+” operator; defined the general three-step algorithm sequence; and studied its physics performance. However, when it comes to firmware development, it is limited to the Property Calculator (i.e. the third step) alone. Hence, the firmware implementation aspects reported in this section are only about the Property Calculator block. Also, it is worth noting that the implemented firmware version discussed here had the k parameter specified to $k = 3$.

6.4.1 General Overview

The implemented firmware block has two inputs, one output and one generic for algorithm parametrisation (Figure 6.15). The block is designed to operate with a 360 MHz system clock `i_clk` which is of VHDL `std_logic` type and is the first input to the algorithm. The second input to the algorithm `i_ClusterSum` is where the Cluster Sum

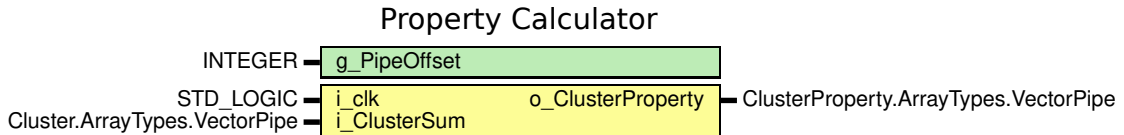


Figure 6.15: A simple schematic of the Property Calculator firmware block with its interfaces and types. More details can be found in text.

records (previously specified by Table 6.3) are being streamed into the block. The output of the algorithm `o_ClusterProperty` is where the calculated Cluster property records (previously specified by Table 6.4) of the corresponding clusters are being streamed out of the block. This input and output are of `Cluster.ArrayTypes.VectorPipe` and `ClusterProperty.ArrayTypes.VectorPipe`, respectively ⁶. The algorithm has no limit on the rate of incoming Cluster Sum records. Specifically, there is no required number of clock cycles worth of pause before it can accept a new input (i.e. it can be fully “pipelined”). Measured from the rising edge of `i_clk` when `i_ClusterSum` input is evaluated, it takes 94 clock cycles before the corresponding Cluster Property record about the same cluster should be evaluated at the `o_ClusterProperty` output. The output can be further delayed

⁶Both of these are special data types common to the RUFL development framework used in Stage 2 [72].

(if required) by specifying the desired number of clock cycles in the `g_PipeOffset` generic. The explicit interplay of the relevant signals in time can be illustrated with time diagrams (Figure 6.16).

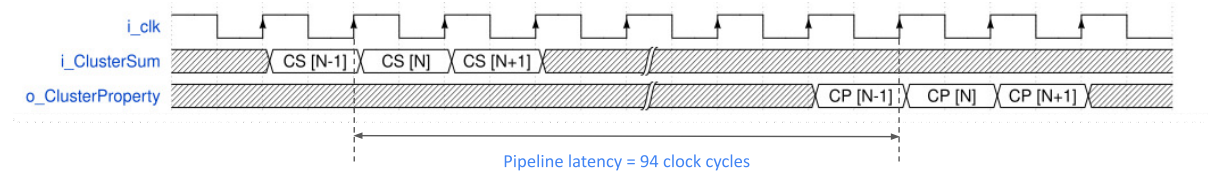


Figure 6.16: A simple time diagram with input and output waveforms of the Property Calculator firmware block. More details can be found in text.

The 94 clock cycle latency of the Property Calculator at the working clock frequency of 360 MHz corresponds to $94 \times \frac{1}{360 \text{ MHz}} \approx 260 \text{ ns}$. This consumes about 13 % of the 2 μs latency budget estimated for the entire Stage 2 algorithm. Hence, it leaves about 87 % in the latency budget for the upstream algorithms (e.g. the main cluster finding step).

The calculation of different cluster quantities happens in parallel to one another, where each property computation has its designated hardware core. This allows utilising the power of parallel processing available in FPGAs and consequently optimises the latency of the algorithm. For example, one can consider a simplified algorithm flow diagram of one such hardware core used to calculate E_T weighted $\mu(c)$, where $c \in \{\phi, r/z, z\}$ (Figure 6.17). Thereby, three such parallel cores are used for any of the three coordinate values c .

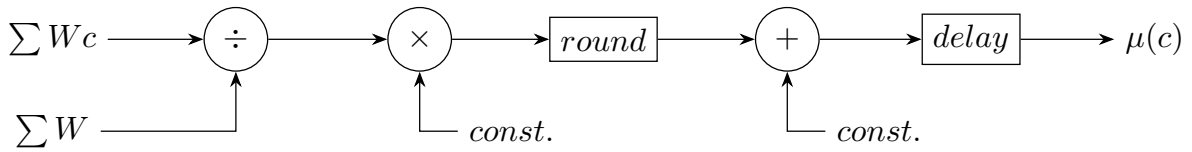


Figure 6.17: A simplified algorithm flow diagram illustrating the key sequence of steps taking place in the firmware responsible block for calculating the E_T weighted $\mu(c)$ cluster property, where $c \in \{\phi, r/z, z\}$. More details can be found in text.

Most of the steps in this flow diagram correspond to what was described in the general overview of the algorithm (as shown in Section 6.2). However, the *delay* step is unique to the firmware implementation (e.g. it does not exist even in the software emulator). Namely, the computational complexity varies from one cluster property to another and so does its computation latency. For example, it takes a significantly greater number of clock cycles to compute the standard deviations (which involve performing three multiplications, one subtraction, one division, a square root, a scaling, a rounding and a saturation step) compared to the simpler properties such as the energies in longitudinal regions (which only require a scaling and a rounding step). Therefore, each of the hardware cores can temporarily store (buffer) its result for a programmable number of clock cycles (as specified to the *delay* operation) ensuring that the final cluster property record has all of its fields available at the same time at the output of the algorithm.

The overall resource utilisation of the Property Calculator firmware block is rather light (Figure 6.18), requiring below 2 % of any resource type on a Xilinx VU13P (the target FPGA choice for the Stage 2 TPG BE Serenity boards). The implemented firmware block also met all specified timing constraints (Figure 6.19).

Name	CLB LUTs (1728000)	CLB Registers (3456000)	CARRY8 (216000)	F7 Muxes (864000)	F8 Muxes (432000)	CLB (216000)	LUT as Logic (1728000)	LUT as Memory (791040)	Block RAM Tile (2688)	DSPs (12288)
▼ N top	22.85%	18.78%	12.42%	0.59%	0.59%	34.67%	21.95%	1.96%	27.51%	5.40%
> [I] ctrl (emp_ctrl)	<0.01%	<0.01%	0.00%	0.00%	0.00%	<0.01%	<0.01%	0.00%	0.00%	0.00%
> [I] datapath (emp_datapath)	8.22%	6.32%	1.13%	0.31%	0.21%	15.17%	8.17%	0.11%	23.21%	0.00%
> [I] fabric (ipbus_fabric_sel)	<0.01%	0.00%	0.00%	0.00%	0.00%	0.01%	<0.01%	0.00%	0.00%	0.00%
> [I] info (emp_info)	0.02%	<0.01%	<0.01%	<0.01%	<0.01%	0.05%	0.02%	0.00%	0.00%	0.00%
> [I] infra (emp_infra_pcie)	0.70%	0.45%	0.13%	0.02%	<0.01%	1.24%	0.61%	0.20%	3.53%	0.00%
> [I] payload (emp_payload)	13.65%	11.84%	11.11%	0.22%	0.36%	23.45%	12.90%	1.66%	0.73%	5.40%
▼ [I] HgcStage2TopInstance (HgcDummyTop)	13.65%	11.84%	11.11%	0.22%	0.36%	23.45%	12.90%	1.66%	0.73%	5.40%
> [I] G1[1].ClusterOutput0 (ConnectToOutput_1)	0.00%	<0.01%	0.00%	0.00%	0.00%	0.05%	0.00%	0.00%	0.00%	0.00%
> [I] G1[1].PropertyCalculator (PropertyCalculator_xdcDup_)	1.03%	0.87%	0.85%	0.02%	0.03%	1.73%	0.97%	0.12%	0.06%	0.42%

Figure 6.18: A resource utilisation report for the an implemented firmware project which houses the Property Calculator block in Vivado (the official Xilinx FPGA algorithm development framework). Highlighted are the values of used resources for the Property Calculator itself, relative to the total available Xilinx VU13P FPGA resources. More details can be found in text.

Setup	Hold	Pulse Width
Worst Negative Slack (WNS): 0.060 ns	Worst Hold Slack (WHS): 0.010 ns	Worst Pulse Width Slack (WPWS): 0.000 ns
Total Negative Slack (TNS): 0.000 ns	Total Hold Slack (THS): 0.000 ns	Total Pulse Width Negative Slack (TPWS): 0.000 ns
Number of Failing Endpoints: 0	Number of Failing Endpoints: 0	Number of Failing Endpoints: 0
Total Number of Endpoints: 1090385	Total Number of Endpoints: 1088043	Total Number of Endpoints: 670455
All user specified timing constraints are met.		

Figure 6.19: A timing report for an implemented firmware project which houses the Property Calculator block in Vivado. More details can be found in text.

6.4.2 Agreement with the Software Model

The Property Calculator firmware block underwent extensive standalone testing where it achieved one-to-one agreement with the corresponding software emulator (across all properties) for about 10 000 clusters played through the algorithm. The behaviour was matched with the software emulator on two levels. Firstly, it was matched in Vivado FPGA behavioural simulation. Secondly, it was matched in a hardware test where the algorithm was operated using a real FPGA.

Additionally, the HGICAL TPG BE and the L1T groups conducted a successful campaign of Stage 2 $\leftrightarrow$ L1T interface tests, where the Property Calculator acted as a placeholder of the HGICAL TPG BE [73][74]. Clusters from about 1000 BXs were thereby run through the Property Calculator block with no reported interface issues (including one-to-one agreement with the corresponding software emulator), setting a major milestone for the HGICAL Stage 2 TPG project.

6.5 Results Summary

This chapter demonstrated the feasibility of the overall implementation of the Cluster Properties block which can be summarised as follows.

A high level of agreement between the software emulator / firmware implementation (limited precision) and the software simulator (full precision) was achieved. Specifically, for over 99% of the 800×10^3 studied clusters, the emulated values deviated from the simulated values by no more than one LSB for most properties. For the E_T weighted $\mu(z)$ and $\sigma_{\eta\eta}$ properties, this difference was slightly worse, but it still was not greater than two and three LSBs in more than 95% of the studied clusters.

Using simple BDT models (without hyper-parameter optimisation) as proxies for the future L1T algorithms, the chapter evaluated the physics performance of the cluster properties when used as features to separate signal (photon or VBF jets) from PU clusters. This was done in the challenging p_T^{gen} range below 100 GeV. For photon identification, the TPR values were excellent, exceeding 95% at an FPR of 5% across the entire tested p_T^{gen} range. For VBF jet identification, at an FPR of 20% the TPR values were very good ($> 80\%$ for $p_T^{gen} < 50$ GeV and $> 95\%$ for $p_T^{gen} > 50$ GeV). The limitations of this study are also summarised leaving the door open for fine-tuning and potential performance optimisations.

The impact of the k parameter on quantisation effects and the overall physics performance was also explored, demonstrating that similar performance (for the binary classification problems considered) can be retained while working with Cluster Sum records whose bit-widths range from 254 bits (for $k = 15$) to 434 bits (for $k = 0$). Most importantly, this flexibility can be used as an argument for reducing the Stage 2 block RAM utilisation ($58\% \rightarrow 45\%$) and Accumulator input bandwidth ($11 \text{ Tbit s}^{-1} \rightarrow 6.5 \text{ Tbit s}^{-1}$) which would consequently simplify the implementation of the overall Stage 2 design (i.e. in terms of resource utilisation, routing, bandwidth constraints, timing closure, etc.).

The firmware implementation of the Property Calculator is relatively light, requiring less than 2% of any Xilinx VU13P resource type. The pipeline latency of the block is 94 clock cycles or about 13% of the envisaged Stage 2 latency budget if using a 360 MHz system clock.

The FPGA algorithm underwent standalone hardware tests, where one-to-one agreement with the corresponding software simulator in about 10 000 clusters was achieved. Moreover, this firmware block took part in the successful HGCAL to L1T interface tests, which marked a major milestone for the HGCAL TPG BE project.

Chapter 7

Summary and Conclusions

The CMS experiment is actively preparing for the high-luminosity era of LHC operation. Enhancing the instantaneous luminosity will allow to produce data at an order of magnitude greater rate. This will consequently enable reducing uncertainties on previous measurements and also may lead to new discoveries in particle physics. However, higher luminosity also means a higher PU of proton-proton interactions in every BX. In turn, the increased PU levels create a more hostile radiation environment for the detector FE electronics as well as requiring CMS to resolve a much greater vertex density. This requires completely changing the innermost detector sub-systems and upgrading others so that they generally have greater radiation tolerance and granularity. One of these CMS upgrades is the replacement of the existing endcap calorimeters with HGCAL.

The HGCAL will leverage its fine spatial, timing and energy precision to distinguish the signatures from different proton-proton interactions despite the challenges brought with the high levels of PU during HL-LHC. This will allow efficient CMS triggering, minimising background noise and optimising event reconstruction. It will enable precise measurements of electromagnetic and hadronic showers, which is crucial for analysing (and triggering whilst having) complex event topologies, including: photon showers, high p_T jets, VBF jets, and missing transverse energy. These objects are pivotal for Higgs boson studies and new physics searches. Specifically, detailed measurements of electromagnetic showers are vital for the $H \rightarrow \gamma\gamma$ decay mode, and hadronic showers are significant for jets from VBF, one of the primary Higgs production mechanisms.

The ambitious scientific objectives of HGCAL would be impossible without first implementing its highly specialised and complex electronics system (which is an engineering feat in and of itself), involving multiple layers of data transfer and control, and a variety of interfaces with the CMS global systems. The HGCAL electronics system has two co-existing branches (the TPG and the DAQ) both split in their respective FE and BE parts. Considering both endcaps, the data from the six million FE readout channels is transmitted through 20 000 fibre optical (F/O) links to the BE systems. This will result in a data rate of about 110 Tbit s^{-1} of usable TPG and DAQ data between the detector FE and the detector BE.

The focus of this thesis revolved around two related aspects of the design and implementation of the HGCAL electronics system. The first was to implement a test system that prototypes a complete vertical slice of the future endcap electronics system, encompassing FE and BE components (e.g ASICs, PCBs, F/O interface, Serenity board, etc.), along

with the trigger and DAQ sub-systems. The second aspect focused on creating BE algorithms and implementing them in FPGA devices, alongside simulations to evaluate their performance and sensitivity to the physics processes that the CMS experiment aims to detect. This included a TPG BE algorithm that determines the physical properties of 3D clusters based on FE energy deposits (such as cluster energy, position, and shape), which are considered by the CMS L1T system when making trigger decisions. Additionally, an algorithm was developed to emulate the HGCAL FE readout chip buffers in the TCDS2 and to occasionally generate an L1A throttle signal. This throttling mechanism prevents certain L1As from reaching the CMS FE in rare cases where the HGCAL FE readout chips might otherwise ignore them due to internal buffer overflow. Without this, the HGCAL would appear as a non-responsive part of the CMS detector (i.e. it would not provide any data about the selected event), complicating the overall CMS event reconstruction.

Chapter 3 described the commissioning of the vertical HGCAL electronics test system, focusing on fast control and timing distribution, slow control, and the acquisition of TPG and DAQ data. Each aspect was examined from multiple perspectives. For fast control and timing distribution, the fast command word error rate upper limit was found to be 9×10^{-14} (at 95 % CL), and the phase noise between the BE external clock generator and the Rafael ASIC output was measured to be 8.5 ps. With the internal HGCROC clock recovery circuits contributing an additional 2 ps, this was found to be well within the 15 ps requirement. Slow control tests set an upper limit on the transaction error rate at 3×10^{-6} (at 95 % CL) when reading or writing registers of any FE ASIC. Regarding data acquisition, Chapter 3 showcased (both TPG and DAQ) data from HGCROCs streaming through all real devices and interfaces in the vertical chain, and arriving in the Serenity BE as expected. As a simple test of the quality of the interfaces on the DAQ path, the DAQ BE CRC error counters were monitored over a period of O(1 d) and showed no issues, while operating the system at an average L1A rate of 750 kHz. Similarly for the TPG path, the TPG BE Stage 1 unpacker maintained the *locked* state over a period of O(1 d). The quality of the physical signals in each system operation aspect was also found to be satisfactory. Additionally, precise tuning of devices and interfaces for data transmission from HGCROCs to BE algorithms was incorporated into an automated system initialisation procedure. All of these results opened doors to studying the vertical test system in greater detail.

It should be emphasised that the devices involved in the vertical test system came from various international institutes and were only, if at all, tested individually at their respective home institutes before being sent to CERN. Prior experience in operating them together was minimal or non-existent. Therefore, this work necessitated communication with experts and institutions from where these devices originated. Another challenging perspective to these activities is that this was a multi-layered (e.g. hardware, software, firmware) electronics system with heterogeneous signal domains (e.g. analogue, digital, optical) which made it difficult to debug. The results presented in this chapter mark the first-ever implementation of an HGCAL readout system prototype of such complexity.

Chapter 4 provided deeper insights into the vertical test system operation, addressing its use in a beam test, the VTRx+ saturation issue and its mitigation, setting upper limits on BERs, and assessing system-wide digital consistency.

During the beam test, the TPG and DAQ sub-systems were tested in various modes (BC and STC4E3M for TPG, Pass-through and Zero-suppression for DAQ) and with different

triggering methods (external and self-triggering). A variety of particle beams (pion, muon, and electron) covering an energy range from a MIP to 200 GeV were used to evaluate ADC, ToT and ToA measurements, and calibration procedures like pedestal subtraction and trimming of the HGCROCs. The L1A rates and data volumes varied with beam type and configuration, staying below approximately $O(100 \text{ kHz})$ and $O(10 \text{ Gbit s}^{-1})$, respectively. These activities allowed HGCAL to become the first CMS subdetector upgrade to use such a complete readout component chain and to implement a DTH-based readout to acquire beam test data.

Signal integrity concerns in the FE $\rightarrow$ BE interface, raised during the beam test, were confirmed with the identification of VTRx+ saturation, which was later mitigated by reducing the output optical power of the Serenity BE Fireflies in the DAQ F/O down-link by 3 dB. The HGCAL was the first system to report these findings to the VTRx+ and lpGBT teams, resulting in a detailed study and an application note for all CERN experiments using these components.

Post-mitigation, extended monitoring allowed to set the BER upper limits at 4×10^{-17} for TPG and 5×10^{-16} for DAQ (at 95 % CL), affirming the quality of electrical and optical interfaces in ambient conditions. Digital consistency error rates were also analysed, setting upper limits below 9×10^{-14} for TPG and 2×10^{-8} for DAQ (at 95 % CL and the nominal average L1A rate of 750 kHz).

These results proved that the proposed vertical architecture of the HGCAL electronics system can indeed be used to acquire real physics data, has a high integrity of the physical signals as well as the quality of the system-wide digital interfaces when operated in ambient conditions. It should be emphasised that this required a deep understanding of various system level aspects including: synchronising the readout of TPG and DAQ data, isolating the VTRx+ saturation issue in a number of possible failing interfaces, knowing the details about the various functionalities of FE ASICs and BE algorithms when conducting the digital consistency checks, etc.

Chapter 5 demonstrated the viability of the HGCAL L1A Throttle Generator firmware block through several key points. The critical step involved exhaustively validating the BX-level behaviour of the HGCROC RAM 2 buffer and its readout logic. This was completed when the software model and the HGCROC ASIC agreed under special L1A sequences, designed to cover all known corner cases of the RAM 2 and its readout logic, as well as random L1A sequences with various average consecutive L1A distances $\langle D_{\text{L1A}} \rangle$ between 1 to 100 over a combined effective duration of 3×10^6 BX. This enabled the study of Deadtime as a function of average L1A frequency, $\text{Deadtime}(\langle f_{\text{L1A}} \rangle)$, in software and allowed proceeding with the firmware implementation. The software model simulated $\text{Deadtime}(\langle f_{\text{L1A}} \rangle)$ within the $\langle f_{\text{L1A}} \rangle$ range of 700 kHz to 1300 kHz, predicting a mean deadtime of $\text{Deadtime}_{\text{sw}}(750 \text{ kHz}) = (1.51 \pm 0.01) \times 10^{-8}$. Testing the standalone firmware block matched the software model (and thus the HGCROC ASIC) when using the specially crafted L1A sequences as input stimuli, which was confirmed through both FPGA simulations and standalone hardware tests. Instantiating the firmware block in the TCDS2 emulator of the vertical HGCAL test system allowed DAQ runs with offline data examination, specifically checking for HGCROC event counter discontinuities. No discontinuities were observed with special L1A sequences and random L1A sequences at an average rate of 750 kHz over approximately 133×10^6 events. These results indicated that the block successfully protected the HGCROCs in the system FE from problematic L1A sequences at

the nominal HL-LHC L1A rate of 750 kHz. Finally, using the firmware block in the TCDS2 emulator measured Deadtime($\langle f_{L1A} \rangle$) within the $\langle f_{L1A} \rangle$ range of 700 kHz to 1300 kHz, with a mean deadtime at $\langle f_{L1A} \rangle = 750$ kHz of $\text{Deadtime}_{\text{HW}}(750 \text{ kHz}) = (1.56 \pm 0.02) \times 10^{-8}$, showing good agreement between the firmware block and software model predictions. This block also protected the vertical test system from problematic L1A sequences during the beam test activities.

Two big challenges regarding the L1A throttling block should be emphasised. Firstly, there was no detailed documentation describing the behaviour of HGCROC RAM 2 and its readout logic (hence this was reverse-engineered). Secondly, instantiating this block in the TCDS2 emulator of the vertical system required meticulous planning (since it is essential for the throttle signal to be used at a precise system clock tick and to deactivate exactly the right L1A candidate). With this work, HGCAL has become the first CMS sub-system upgrade to propose the concept of L1A throttling to the TCDS2 developers based on a working proof-of-principle prototype.

Chapter 6 demonstrated the feasibility of the overall implementation of the Cluster Properties firmware block, highlighting several key points. There was a high level of agreement achieved between the software emulator / firmware implementation (with limited precision) and the software simulator (with full precision). Specifically, for over 99 % of the 800×10^3 clusters analysed, the emulated values differed from the simulated ones by no more than one LSB for most properties. For the E_T weighted $\mu(z)$ and $\sigma_{\eta\eta}$ properties, this difference was slightly worse, but it still was not greater than two and three LSBs in more than 95 % of the studied clusters.

This work also assessed the physics performance of the cluster properties by using simple BDT models as proxies for future L1T algorithms, aimed at distinguishing signal (photon or VBF jets) from PU clusters. This was done in the particularly challenging p_T^{gen} range below 100 GeV. For photon identification, TPR values were excellent, surpassing 95% at an FPR of 5% across the entire p_T^{gen} range tested. For VBF jet identification, at an FPR of 20% the TPR values were very good ($> 80\%$ for $p_T^{\text{gen}} < 50$ GeV and $> 95\%$ for $p_T^{\text{gen}} > 50$ GeV). The limitations of the study were also noted, suggesting opportunities for fine-tuning and performance optimisation.

The impact of the quantisation effects on overall physics performance was also highlighted, showing that comparable performance for the studied binary classification problems can be maintained while using Cluster Sum records with bit-widths ranging from 254 bits to 434 bits. This flexibility is crucial for reducing Stage 2 block RAM usage (58 % $\rightarrow$ 45 %) and Accumulator input bandwidth (11 Tbit s $^{-1}$ $\rightarrow$ 6.5 Tbit s $^{-1}$), thereby simplifying the overall Stage 2 design in terms of resource utilisation, routing, bandwidth, and timing closure. The Property Calculator firmware implementation is relatively lightweight, requiring less than 2 % of any target FPGA resource type. The block's pipeline latency is 94 clock cycles, consuming about 13 % of the envisaged Stage 2 latency budget when driven by a 360 MHz system clock.

The FPGA algorithm underwent standalone hardware tests, achieving one-to-one agreement with the corresponding software simulator for $O(10^4)$ clusters. Lastly, this firmware block participated in a successful campaign of HGCAL $\rightarrow$ L1T interface tests. These were the first ever interface tests between these two systems, whose success marked a major milestone for the HGCAL BE TPG project.

The most challenging aspect of this development was to find the right balance between the level of detail conveyed by the trigger primitives with FPGA resource utilisation, algorithm latency, and bandwidth limitations.

The combined achievements concerning vertical system implementation and validation opened the doors to several future activities. Firstly, they have paved the way for a smooth implementation of different kinds of vertical test systems. Namely these would involve FE structures with HD and partial silicon modules as well as scintillating tileboards. It will be possible to transfer the bulk of the system operation “know-how” developed for the LD single-module silicon Train to these systems, where similar performance studies should be conducted. Secondly, they allow the project to move towards horizontal scaling. Specifically, the achievements of the single-module system already enabled to work with multi-module LD silicon Trains. However, together with the future achievements for HD, partial and scintillator flavoured vertical systems will allow building single-layer, followed by multi-layer cassettes, and ultimately entire HGCAL endcaps. Of course, scaling up the detector FE will require changes in the detector BE. Lastly, given that the BER measurements as well as the system-wide digital consistency checks were done in the absence of SEUs and in ambient conditions, it would be worthwhile repeating such measurements whilst having the whole variety of the FE ASICs in the vertical chain exposed to expected levels of SEUs and in nominal conditions. Consequently, this would allow setting more realistic upper limits on these measurements.

A future activity regarding the HGCAL L1A throttling would be co-locating this block in the central TCDS2. Additionally, to protect CMS event reconstruction from such scenarios, TCDS2 should also consider implementing throttling blocks for all sub-systems that suffer from similar, deterministic buffering limitations as HGCAL does.

There are several possible ways to proceed with the work regarding cluster property calculation. Firstly, it is possible to calibrate this algorithm to work with larger values of standard deviation properties, instead of saturating. Secondly, it is possible to refine the $\sigma_{\eta\eta}$ look-up mechanism to reduce the quantisation-induced noise levels. Both of these would improve the overall precision of the corresponding properties. Lastly, the evaluated physics performance of the algorithm can be considered as a proof-of-principle and conservative. In other words, better performance can probably be obtained if any of the following limitations of the study is addressed: absence of generator-level information in the available data about PU clusters; optimising the VBF jet cluster definition; using multi-dimensional re-weighting; trying out other classification models or using BDTs with optimised hyper-parameters; and combining clusters with trigger primitives from other CMS sub-systems.

In conclusion, the work presented in this thesis directly contributed to the implementation and validation of the overall HGCAL electronics system in CMS’s preparation for the HL-LHC. Additionally, this thesis presented two firmware algorithms for the CMS TCDS2 and the HGCAL TPG BE systems. These respective blocks can be used to protect CMS event reconstruction from situations when HGCAL cannot accept L1A triggers, and to identify at the L1T level some of the most important physics signatures (photons and VBF jets) that CMS aims to be sensitive to during HL-LHC.

Bibliography

- [1] M. Vojinovic on behalf of the CMS Collaboration, “CMS HGCal electronics vertical integration system tests”, en, *Journal of Instrumentation*, vol. 19, no. 02, p. C02022, Feb. 2024, Publisher: IOP Publishing, ISSN: 1748-0221. DOI: 10.1088/1748-0221/19/02/C02022. [Online]. Available: <https://dx.doi.org/10.1088/1748-0221/19/02/C02022> (visited on 01/03/2024).
- [2] R. Shukla on behalf of the CMS Collaboration, “The CMS HGCal trigger data receiver”, en, *Journal of Instrumentation*, vol. 19, no. 01, p. C01049, Jan. 2024, Publisher: IOP Publishing, ISSN: 1748-0221. DOI: 10.1088/1748-0221/19/01/C01049. [Online]. Available: <https://dx.doi.org/10.1088/1748-0221/19/01/C01049> (visited on 04/03/2024).
- [3] L. Evans and P. Bryant, “LHC Machine”, en, *Journal of Instrumentation*, vol. 3, no. 08, S08001, Aug. 2008, ISSN: 1748-0221. DOI: 10.1088/1748-0221/3/08/S08001. [Online]. Available: <https://dx.doi.org/10.1088/1748-0221/3/08/S08001> (visited on 28/11/2023).
- [4] J. Ellis, “Physics at the LHC”, en, *The European Physical Journal C*, vol. 34, no. 1, pp. 51–56, May 2004, ISSN: 1434-6044, 1434-6052. DOI: 10.1140/epjc/s2004-01766-8. [Online]. Available: <http://link.springer.com/10.1140/epjc/s2004-01766-8> (visited on 30/11/2023).
- [5] O. Brüning, H. Burkhardt and S. Myers, “The Large Hadron Collider”, *Progress in Particle and Nuclear Physics*, vol. 67, no. 3, pp. 705–734, Jul. 2012, ISSN: 0146-6410. DOI: 10.1016/j.pnnp.2012.03.001. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0146641012000695> (visited on 30/11/2023).
- [6] J. Wright, “Study of Higgs boson production through vector boson fusion at the CMS experiment using a dense convolutional neural network”, Ph.D. dissertation, Imperial College, London, 2018. DOI: 10.25560/67922.
- [7] The ALICE Collaboration, “The ALICE experiment at the CERN LHC”, en, *Journal of Instrumentation*, vol. 3, no. 08, S08002, Aug. 2008, ISSN: 1748-0221. DOI: 10.1088/1748-0221/3/08/S08002. [Online]. Available: <https://dx.doi.org/10.1088/1748-0221/3/08/S08002> (visited on 29/07/2024).
- [8] The ATLAS Collaboration, “The ATLAS Experiment at the CERN Large Hadron Collider”, en, *Journal of Instrumentation*, vol. 3, no. 08, S08003, Aug. 2008, ISSN: 1748-0221. DOI: 10.1088/1748-0221/3/08/S08003. [Online]. Available: <https://dx.doi.org/10.1088/1748-0221/3/08/S08003> (visited on 29/07/2024).

- [9] The CMS Collaboration, “The CMS experiment at the CERN LHC”, en, *Journal of Instrumentation*, vol. 3, no. 08, S08004, Aug. 2008, ISSN: 1748-0221. DOI: 10.1088/1748-0221/3/08/S08004. [Online]. Available: <https://dx.doi.org/10.1088/1748-0221/3/08/S08004> (visited on 01/12/2023).
- [10] The LHCb Collaboration, “The LHCb Detector at the LHC”, en, *Journal of Instrumentation*, vol. 3, no. 08, S08005, Aug. 2008, ISSN: 1748-0221. DOI: 10.1088/1748-0221/3/08/S08005. [Online]. Available: <https://dx.doi.org/10.1088/1748-0221/3/08/S08005> (visited on 29/07/2024).
- [11] The CMS Collaboration, “Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC”, *Physics Letters B*, vol. 716, no. 1, pp. 30–61, Sep. 2012, ISSN: 0370-2693. DOI: 10.1016/j.physletb.2012.08.021. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0370269312008581> (visited on 01/12/2023).
- [12] The ATLAS Collaboration, “Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC”, *Physics Letters B*, vol. 716, no. 1, pp. 1–29, Sep. 2012, ISSN: 0370-2693. DOI: 10.1016/j.physletb.2012.08.020. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S037026931200857X> (visited on 28/07/2024).
- [13] T. Sakuma and T. McCauley, “Detector and Event Visualization with SketchUp at the CMS Experiment”, en, *Journal of Physics: Conference Series*, vol. 513, no. 2, p. 022032, Jun. 2014, ISSN: 1742-6596. DOI: 10.1088/1742-6596/513/2/022032. [Online]. Available: <https://dx.doi.org/10.1088/1742-6596/513/2/022032> (visited on 01/12/2023).
- [14] I. Neutelings, *CMS coordinate system*, en-US, Jul. 2021. [Online]. Available: https://tikz.net/axis3d_cms/ (visited on 29/07/2024).
- [15] D. Barney, *CMS Detector Slice*, 2016. [Online]. Available: <https://cds.cern.ch/record/2120661> (visited on 01/12/2023).
- [16] The CMS Collaboration, “Description and performance of track and primary-vertex reconstruction with the CMS tracker”, en, *Journal of Instrumentation*, vol. 9, no. 10, P10009, Oct. 2014, ISSN: 1748-0221. DOI: 10.1088/1748-0221/9/10/P10009. [Online]. Available: <https://dx.doi.org/10.1088/1748-0221/9/10/P10009> (visited on 30/07/2024).
- [17] B. Bilki on behalf of the CMS Collaboration, “CMS Forward Calorimeters Phase II Upgrade”, en, *Journal of Physics: Conference Series*, vol. 587, no. 1, p. 012014, Feb. 2015, ISSN: 1742-6596. DOI: 10.1088/1742-6596/587/1/012014. [Online]. Available: <https://dx.doi.org/10.1088/1742-6596/587/1/012014> (visited on 06/12/2023).
- [18] The CMS Collaboration, “Technical proposal for the upgrade of the CMS detector through 2020”, Tech. Rep. CERN-LHCC-2011-006, CMS-UG-TP-1, LHCC-P-004, 2011. [Online]. Available: <https://cds.cern.ch/record/1355706> (visited on 03/12/2023).
- [19] A. Rizzi, G. Petrucciani and M. Peruzzi, “A further reduction in CMS event data for analysis: the NANO AOD format”, *EPJ Web Conf.*, vol. 214, p. 06021, 2019. DOI: 10.1051/epjconf/201921406021. [Online]. Available: <http://cds.cern.ch/record/2699585>.

- [20] The CMS Collaboration, “The Phase-2 Upgrade of the CMS Level-1 Trigger”, CERN, Geneva, Tech. Rep. CERN-LHCC-2020-004, CMS-TDR-021, 2020, Final version. [Online]. Available: <https://cds.cern.ch/record/2714892> (visited on 05/12/2023).
- [21] G. Apollinari, Béjar Alonso I, Brüning O, Lamont M and Rossi L, “High-Luminosity Large Hadron Collider (HL-LHC) Preliminary Design Report”, en, 2015, Publisher: CERN. DOI: 10.5170/CERN-2015-005. [Online]. Available: <http://cds.cern.ch/record/2116337> (visited on 04/12/2023).
- [22] O. Brüning and L. Rossi, “The High Luminosity Large Hadron Collider – HL-LHC”, in *The High Luminosity Large Hadron Collider*, 2024, pp. 1–53. DOI: 10.1142/9789811278952_0001. [Online]. Available: https://www.worldscientific.com/doi/abs/10.1142/9789811278952_0001.
- [23] The ATLAS Collaboration, “Cross-section measurements of the Higgs boson decaying into a pair of tau-leptons in proton-proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector”, *Physical Review D*, vol. 99, no. 7, p. 072001, Apr. 2019, ISSN: 2470-0010, 2470-0029. DOI: 10.1103/PhysRevD.99.072001. [Online]. Available: <http://arxiv.org/abs/1811.08856> (visited on 04/12/2023).
- [24] A. David, “HL-LHC CMS Detector Upgrade (from the past to the future)”, en, CMS Week, Dec 2022, CERN. [Online]. Available: <https://indico.cern.ch/event/1180059/contributions/5131637/> (visited on 31/07/2024).
- [25] F. Gianotti, M. L. Mangano, T. Virdee *et al.*, “Physics potential and experimental challenges of the LHC luminosity upgrade”, en, *The European Physical Journal C - Particles and Fields*, vol. 39, no. 3, pp. 293–333, Feb. 2005, ISSN: 1434-6052. DOI: 10.1140/epjc/s2004-02061-6. [Online]. Available: <https://doi.org/10.1140/epjc/s2004-02061-6> (visited on 31/07/2024).
- [26] The CMS Collaboration, “The Phase-2 Upgrade of the CMS Tracker”, CERN, Geneva, Tech. Rep. CERN-LHCC-2017-009, CMS-TDR-014, 2017. [Online]. Available: <http://cds.cern.ch/record/2272264> (visited on 31/07/2024).
- [27] The CMS Collaboration, “A MIP Timing Detector for the CMS Phase-2 Upgrade”, CERN, Geneva, Tech. Rep. CERN-LHCC-2019-003, CMS-TDR-020, 2019. [Online]. Available: <https://cds.cern.ch/record/2667167> (visited on 31/07/2024).
- [28] The CMS Collaboration, “The Phase-2 Upgrade of the CMS Barrel Calorimeters”, CERN, Geneva, Tech. Rep. CERN-LHCC-2017-011, CMS-TDR-015, 2017. [Online]. Available: <https://cds.cern.ch/record/2283187> (visited on 31/07/2024).
- [29] The CMS Collaboration, “The Phase-2 Upgrade of the CMS Endcap Calorimeter”, CERN, Geneva, Tech. Rep. CERN-LHCC-2017-023, CMS-TDR-019, 2017. DOI: 10.17181/CERN.IV8M.1JY2. [Online]. Available: <https://cds.cern.ch/record/2293646> (visited on 05/12/2023).
- [30] The CMS Collaboration, “The Phase-2 Upgrade of the CMS Muon Detectors”, CERN, Geneva, Tech. Rep. CERN-LHCC-2017-012, CMS-TDR-016, 2017. [Online]. Available: <https://cds.cern.ch/record/2283189> (visited on 31/07/2024).
- [31] The CMS Collaboration, “The Phase-2 Upgrade of the CMS Data Acquisition and High Level Trigger”, CERN, Geneva, Tech. Rep. CERN-LHCC-2021-007, CMS-TDR-022, 2021. [Online]. Available: <https://cds.cern.ch/record/2759072> (visited on 31/07/2024).

- [32] *Energy deposits in the HGCal for a typical HL-LHC collision*, en-GB. [Online]. Available: <https://www.imperial.ac.uk/a-z-research/high-energy-physics/research/experiments/cms/detector/cms-detector-development/hgcal/> (visited on 01/08/2024).
- [33] A. David on behalf of the CMS Collaboration, “Performance of CMS High-Granularity Calorimeter for HL-LHC”, Parallel talk presented at the ICHEP 2022, Bologna, Italy, 2022, INFN. [Online]. Available: <https://agenda.infn.it/event/28874/contributions/169080/> (visited on 06/12/2023).
- [34] The HGCal Backend Group, “The HGCal Backend TDAQ Systems for the Pre-series Design Review”, en, Tech. Rep. DN-19-032, 2022. [Online]. Available: <https://edms.cern.ch/ui/file/2220310/2/DN-19-032-V2.pdf> (visited on 08/12/2023).
- [35] J. Mans, “FAST CONTROL FOR THE CMS ENDCAP CALORIMETER”, CERN, Tech. Rep. EDMS NO. 2044200, REV. 4.8, Nov. 2023. [Online]. Available: https://edms.cern.ch/ui/file/2044200/4.4/fast-control-spec_V4_8.pdf (visited on 27/12/2023).
- [36] C. De La Taille, F. Dulucq, D. Thienpont, S. Extier and M. El Berni, “HGCROC3b”, en, CERN, Tech. Rep. EDMS No. 2954073, CMS-CE-ES-0413, Dec. 2023. [Online]. Available: <https://edms.cern.ch/ui/#!/master/navigator/document?D:101362066:101362066:subDocs> (visited on 11/12/2023).
- [37] J. Hirschauer, J. Hoff and D. Noonan, “ECON-T TRIGGER CONCENTRATOR ASIC FOR THE CMS HIGH GRANULARITY ENDCAP CALORIMETER”, en, CERN, Tech. Rep. EDMS NO. 2206779, CMS-CE-EN-0003, Jun. 2023. [Online]. Available: https://edms.cern.ch/ui/file/2206779/1/ECON-T_specification_working_doc_v10_1jun2023.pdf (visited on 11/12/2023).
- [38] J. Hirschauer, J. Hoff and D. Noonan, “DAQ CONCENTRATOR ASIC (ECON-D) FOR THE CMS HL-LHC ENDCAP CALORIMETER”, en, CERN, Tech. Rep. EDMS No. 2599620, CMS-CE-EN-0022, Oct. 2023. [Online]. Available: <https://edms.cern.ch/ui/#!/master/navigator/document?D:100904542:100904542:subDocs> (visited on 11/12/2023).
- [39] Philips Semiconductors, “The I2C-bus specification”, *Philips Semiconductors*, vol. 9397, no. 750, p. 00 954, 2000.
- [40] P. Moreira, “The LpGBT Project, Status and Overview”, *ACES 2016 - Fifth Common ATLAS CMS Electronics Workshop for LHC Upgrades*, Mar. 2016. [Online]. Available: <https://cds.cern.ch/record/2137809> (visited on 08/12/2023).
- [41] J. Troska, A. Brandon-Bravo, S. Detraz *et al.*, “The VTRx+, an Optical Link Module for Data Transmission at HL-LHC”, in *Proceedings of Topical Workshop on Electronics for Particle Physics — PoS(TWEPP-17)*, Santa Cruz, California, Mar. 2017, p. 048. DOI: 10.22323/1.313.0048. [Online]. Available: <https://cds.cern.ch/record/2312396> (visited on 08/12/2023).
- [42] *Micro Flyover On-Board Optical Engine, FireFly™, MXC®*, *MTP®* | Samtec. [Online]. Available: <https://www.samtec.com/optics/optical-cable/mid-board/firefly> (visited on 21/12/2023).

- [43] F. A. Khan on behalf of the CMS Collaboration, “Design and performance optimisation of the Hexaboards for CMS HGCAL silicon sensor readout electronics”, en, *Journal of Instrumentation*, vol. 18, no. 03, p. C03015, Mar. 2023, Publisher: IOP Publishing, ISSN: 1748-0221. DOI: 10.1088/1748-0221/18/03/C03015. [Online]. Available: <https://dx.doi.org/10.1088/1748-0221/18/03/C03015> (visited on 23/12/2023).
- [44] The Serenity Collaboration, *Serenity Home Page*. [Online]. Available: <https://serenity.web.cern.ch/serenity/> (visited on 28/12/2023).
- [45] V. Amoiridis, U. Behrens, A. Bocci *et al.*, “CMS phase-2 DAQ and timing hub prototyping results and perspectives”, en, *Journal of Instrumentation*, vol. 17, no. 05, p. C05003, May 2022, Publisher: IOP Publishing, ISSN: 1748-0221. DOI: 10.1088/1748-0221/17/05/C05003. [Online]. Available: <https://dx.doi.org/10.1088/1748-0221/17/05/C05003> (visited on 28/12/2023).
- [46] D. Barney, *Hexaboards: Readiness for production*, en, CMS HGCAL ESR1. [Online]. Available: https://indico.cern.ch/event/1325036/contributions/5595440/attachments/2749973/4828089/HexaboardsReadiness_13112023.pdf (visited on 23/12/2023).
- [47] D. Gigi, J. Hegeman and C. Schwick, “The Phase-2 CMS SLinkRocket DAQ link Protocol and format description”, CMS, Tech. Rep. EDMS No. 2502737, Rev No. 1.2, Sep. 2022. [Online]. Available: https://edms.cern.ch/ui/file/2502737/2/cms_phase2_slinkrocket.pdf (visited on 01/01/2024).
- [48] T. Gorski, “CMS Standard Protocol”, Tech. Rep., Apr. 2022. [Online]. Available: https://indico.cern.ch/event/1137504/contributions/4832473/attachments/2434223/4168920/CMS_Standard_Protocol_29Apr2022.pdf (visited on 03/01/2024).
- [49] P. Dauncey, *Concepts for Stage 1 DAQ event readout*, en, CERN. [Online]. Available: <https://indico.cern.ch/event/1240426/contributions/5679039/attachments/2755223/4799099/231121.pdf> (visited on 05/01/2024).
- [50] R. Workman et al. (Particle Data Group), *Progr. Theor. Exp. Phys.* Dec. 2023. [Online]. Available: <https://pdg.lbl.gov/2023/reviews/rpp2022-rev-statistics.pdf#page=28> (visited on 25/01/2024).
- [51] O. Sahin, M. Vojinovic, V. Lohezic and F. Couderc, *Clock Distribution Studies with Train V2 and V3 Test Stands*, CERN, Dec. 2022. [Online]. Available: https://indico.cern.ch/event/1205759/contributions/5071445/attachments/2527113/4346984/HGCAL_clock_distribution_121022.pdf (visited on 24/01/2024).
- [52] A. David, A. Becker, A. Steen and M. Rosado, *The articulated SWAMP*, en, 2023. [Online]. Available: <https://codimd.web.cern.ch/s/kkAnMGmY0#> (visited on 26/01/2024).
- [53] A. Martelli, *The CMS HGCAL detector for HL-LHC upgrade*, Aug. 2017. DOI: 10.48550/arXiv.1708.08234. [Online]. Available: <http://arxiv.org/abs/1708.08234> (visited on 20/07/2024).
- [54] T. Quast, “Construction and beam-tests of silicon-tungsten prototype modules for the CMS High Granularity Calorimeter for HL-LHC”, *Journal of Instrumentation*, vol. 13, no. 02, pp. C02044–C02044, Feb. 2018, ISSN: 1748-0221. DOI: 10.1088/1748-0221/13/02/C02044. [Online]. Available: <http://arxiv.org/abs/1712.05638> (visited on 20/07/2024).

- [55] A. Lobanov, “Precision timing calorimetry with the CMS HGCAL”, *Journal of Instrumentation*, vol. 15, no. 07, pp. C07003–C07003, Jul. 2020, ISSN: 1748-0221. DOI: 10.1088/1748-0221/15/07/C07003. [Online]. Available: <http://arxiv.org/abs/2005.13324> (visited on 20/07/2024).
- [56] B. Acar, G. Adamov, C. Adloff *et al.*, “Response of a CMS HGCAL silicon-pad electromagnetic calorimeter prototype to 20-300 GeV positrons”, *Journal of Instrumentation*, vol. 17, no. 05, P05022, May 2022, ISSN: 1748-0221. DOI: 10.1088/1748-0221/17/05/P05022. [Online]. Available: <http://arxiv.org/abs/2111.06855> (visited on 20/07/2024).
- [57] B. Acar, G. Adamov, C. Adloff *et al.*, *Performance of the CMS High Granularity Calorimeter prototype to charged pion beams of 20-300 GeV/c*, May 2023. DOI: 10.48550/arXiv.2211.04740. [Online]. Available: <http://arxiv.org/abs/2211.04740> (visited on 20/07/2024).
- [58] The CMS HGCAL Collaboration, “Timing Performance of the CMS High Granularity Calorimeter Prototype”, *Journal of Instrumentation*, vol. 19, no. 04, P04015, Apr. 2024, ISSN: 1748-0221. DOI: 10.1088/1748-0221/19/04/P04015. [Online]. Available: <http://arxiv.org/abs/2312.14622> (visited on 20/07/2024).
- [59] M. Aamir, B. Acar, G. Adamov *et al.*, *Using graph neural networks to reconstruct charged pion showers in the CMS High Granularity Calorimeter*, Jun. 2024. DOI: 10.48550/arXiv.2406.11937. [Online]. Available: <http://arxiv.org/abs/2406.11937> (visited on 20/07/2024).
- [60] M. Vojinovic, *Firing a Particle Beam into a Vertical Slice of HGCAL (Or into a slice of a cake?) | The Cylindrical Onion*, Oct. 2023. [Online]. Available: <https://cylindricalonion.web.cern.ch/blogs/firing-particle-beam-vertical-slice-hgcal-or-slice-cake> (visited on 21/02/2024).
- [61] B. Raë on behalf of CERN Experimental Areas Group, “Overview over CERN test beam facilities and LS2 modifications”, en, *8th Beam Telescopes and Test Beams Workshop*, Jan. 2020. [Online]. Available: <https://indico.cern.ch/event/813822/contributions/3648315/> (visited on 21/02/2024).
- [62] *Exabyte Open Storage Homepage*. [Online]. Available: <https://eos-web.web.cern.ch/eos-web/> (visited on 23/02/2024).
- [63] C. Knight, *HGCAL August Beam Test – ECON-T Studies*, en, CERN, Aug. 2023. [Online]. Available: https://indico.cern.ch/event/1316942/contributions/5546840/attachments/2701732/4689548/HGCAL_BE_Trigger.pdf (visited on 05/03/2024).
- [64] D. Hernandez Montesinos and L. Olantera, *The mysterious behaviour of Versatile Link and Versatile Link+ in 2023*, en, Mar. 2024. [Online]. Available: <https://indico.cern.ch/event/1386928/> (visited on 11/03/2024).
- [65] L. Olanterä, *VTRx+ Receiver Saturation and Mitigation | Document 3127793 (v.1)*, en, Mar. 2024. [Online]. Available: <https://edms.cern.ch/ui/#!/master/navigator/document?D:101573437:101573437:subDocs> (visited on 20/07/2024).
- [66] D. R. Cox, *Queues*. Chapman and Hall/CRC, 2020, ISBN: 9781003072904. DOI: <https://doi.org/10.1201/9781003072904>.

- [67] H. Øverås, “Dead-time losses in a buffered data recording system”, *Nuclear Instruments and Methods*, vol. 104, no. 1, pp. 85–91, 1972, ISSN: 0029-554X. DOI: [https://doi.org/10.1016/0029-554X\(72\)90300-X](https://doi.org/10.1016/0029-554X(72)90300-X). [Online]. Available: <https://www.sciencedirect.com/science/article/pii/0029554X7290300X>.
- [68] G. Iles, W. Cameron, C. Foudas, G. Hall and N. Marinelli, “The APVE emulator to prevent front-end buffer overflows within the CMS Silicon Strip Tracker”, in *8th Workshop on Electronics for LHC Experiments*, Sep. 2002.
- [69] M. Raymond, P. Moreira, J. Fulcher *et al.*, “The CMS Tracker APV25 0.25 μm CMOS readout chip”, *6th Workshop on Electronics for LHC Experiments*, pp.130, 2000. DOI: 10.5170/CERN-2000-010.130.
- [70] M. Pesaresi, “Development of a new Silicon Tracker at CMS for Super-LHC”, CERN-THESIS-2010-083, Ph.D. dissertation, Imperial College, London, 2010.
- [71] *Reweighting algorithms — hep_ml 0.7.0 documentation*. [Online]. Available: https://arogozhnikov.github.io/hep_ml/reweight.html#hep_ml.reweight.BinsReweighter (visited on 15/07/2024).
- [72] A. Rose, *RUFL / RuflCore · GitLab*, en, Nov. 2022. [Online]. Available: <https://gitlab.cern.ch:8443/rufl/RuflCore> (visited on 11/07/2024).
- [73] E. Clement, *L1T interface test status and results*, en, Jul. 2023. [Online]. Available: <https://indico.cern.ch/event/1240393/#2-l1t-interface-test-status-an> (visited on 11/07/2024).
- [74] G. Petrucciani, G. Cerminara, S. Summers and E. Clement, *Multi-board tests at B904 for the Correlator trigger*, en, Jul. 2023. [Online]. Available: <https://indico.cern.ch/event/1307728/contributions/5500933/attachments/2686340/4660897/multi-board%20tests%20at%20B904%20for%20the%20Correlator%20trigger.pdf> (visited on 11/07/2024).

Appendix A

Examples of BX-level Behaviour of the HGCAL L1A Throttle

This appendix presents detailed BX-level behaviour of the HGCAL L1A Throttle Generator firmware block (previously covered in Chapter 5), in response to three specially crafted L1A sequences meant to cover all the corner cases of the HGCROC RAM 2 and its readout logic behaviour. That is done through three respective tables (Tables A.1, A.2 and A.3), where each one conveys the status of the internal counters, states and status flags of the block (columns), changing from one BX to another (rows). For each row, the meaning of each one of the columns can be summarised as:

- `BC` is the current value of the BX counter;
- `l1a_in` is the current L1A candidate;
- `throttle` is the reached decision if the current L1A candidate should be deactivated;
- `l1a_fin` is the final L1A value reached based on the previous two points;
- `FSM` is the current status of the finite state machine;
- `output_word` is the index of the current word being streamed out;
- `EC` is the current event counter value;
- `words_RAM2` is the current number of words in RAM 2;
- `events_RAM2` is the current number of events in RAM 2;
- `drain_timeout` conveys for how many next BXs will the draining of data be paused;
- `no_drain_empty` indicates if the draining is stopped due to the empty buffer;
- `no_drain_crc_idle` indicates if the draining is stopped due to the word being streamed out is either the CRC or IDLE word; and
- `no_drain_global` conveys about the current global status of the draining logic.

Each binary column takes on the value of 0 and 1 if the corresponding statement is false and true, respectively.

Table A.1: A BX-level emulation of the internal counters, states and status flags of the HGICAL L1A throttle block when presented with a **burst sequence** of 45 consecutive L1A requests when the throttling decision is **not taken into account** (i.e. disabled). It should be noted that the event counter EC is incremented for every L1A received `l1a_in`. However, when the occupancy limitation is hit (i.e. `events_RAM2 = 33`) RAM 2 could not accept additional events. Generated using the HGICAL L1A throttle software model. Different colours encode different FSM states.

BC	l1a_in	throttle	l1a_fin	FSM	output_word	EC	words_RAM2	events_RAM2	drain_timeout	no_drain_empty	no_drain_crc_idle	no_drain_global
1	0	0	0	FILL_N_DRAIN_N	0	0	0	0	2	1	0	1
2	1	0	1	FILL_Y_DRAIN_N	0	1	39	1	2	1	0	1
3	1	0	1	FILL_Y_DRAIN_N	0	2	78	2	1	1	0	1
4	1	0	1	FILL_Y_DRAIN_N	0	3	117	3	0	0	0	0
5	1	0	1	FILL_Y_DRAIN_Y	1	4	155	4	0	0	0	0
6	1	0	1	FILL_Y_DRAIN_Y	2	5	193	5	0	0	0	0
...	...	...	...	...	...	...	...	...	...	...	...	...
33	1	0	1	FILL_Y_DRAIN_Y	29	32	1219	32	0	0	0	0
34	1	0	1	FILL_Y_DRAIN_Y	30	33	1257	33	0	0	0	0
35	1	1	1	FILL_N_DRAIN_Y	31	34	1256	33	0	0	0	0
36	1	1	1	FILL_N_DRAIN_Y	32	35	1255	33	0	0	0	0
...	...	...	...	...	...	...	...	...	...	...	...	...
42	1	1	1	FILL_N_DRAIN_Y	38	41	1249	33	0	0	0	0
43	1	1	1	FILL_N_DRAIN_Y	39	42	1248	32	0	0	1	1
44	1	1	1	FILL_N_DRAIN_N	40	43	1248	32	0	0	1	1
45	1	1	1	FILL_N_DRAIN_N	41	44	1248	32	0	0	0	0
46	1	0	1	FILL_Y_DRAIN_Y	1	45	1286	33	0	0	0	0
47	0	1	0	FILL_N_DRAIN_Y	2	45	1285	33	0	0	0	0
48	0	1	0	FILL_N_DRAIN_Y	3	45	1284	33	0	0	0	0
...	...	...	...	...	...	...	...	...	...	...	...	...
83	0	1	0	FILL_N_DRAIN_Y	38	45	1249	33	0	0	0	0
84	0	1	0	FILL_N_DRAIN_Y	39	45	1248	32	0	0	1	1
85	0	1	0	FILL_N_DRAIN_N	40	45	1248	32	0	0	1	1
86	0	1	0	FILL_N_DRAIN_N	41	45	1248	32	0	0	0	0
87	0	0	0	FILL_N_DRAIN_Y	1	45	1247	32	0	0	0	0
88	0	0	0	FILL_N_DRAIN_Y	2	45	1246	32	0	0	0	0
...	...	...	...	...	...	...	...	...	...	...	...	...
124	0	0	0	FILL_N_DRAIN_Y	38	45	1210	32	0	0	0	0
125	0	0	0	FILL_N_DRAIN_Y	39	45	1209	31	0	0	1	1
126	0	0	0	FILL_N_DRAIN_N	40	45	1209	31	0	0	1	1
127	0	0	0	FILL_N_DRAIN_N	41	45	1209	31	0	0	0	0
128	0	0	0	FILL_N_DRAIN_Y	1	45	1208	31	0	0	0	0
129	0	0	0	FILL_N_DRAIN_Y	2	45	1207	31	0	0	0	0
...	...	...	...	...	...	...	...	...	...	...	...	...
1354	0	0	0	FILL_N_DRAIN_Y	38	45	40	2	0	0	0	0
1355	0	0	0	FILL_N_DRAIN_Y	39	45	39	1	0	0	1	1
1356	0	0	0	FILL_N_DRAIN_N	40	45	39	1	0	0	1	1
1357	0	0	0	FILL_N_DRAIN_N	41	45	39	1	0	0	0	0
1358	0	0	0	FILL_N_DRAIN_Y	1	45	38	1	0	0	0	0
1359	0	0	0	FILL_N_DRAIN_Y	2	45	37	1	0	0	0	0
...	...	...	...	...	...	...	...	...	...	...	...	...
1395	0	0	0	FILL_N_DRAIN_Y	38	45	1	1	0	0	0	0
1396	0	0	0	FILL_N_DRAIN_Y	39	45	0	0	0	0	1	1
1397	0	0	0	FILL_N_DRAIN_N	40	45	0	0	2	1	1	1
1398	0	0	0	FILL_N_DRAIN_N	41	45	0	0	2	1	0	1

Table A.2: A BX-level emulation of the internal counters, states and status flags of the HGCAL L1A throttle block when presented with a **burst sequence** of 45 consecutive L1A requests when the throttling decision **is taken into account** (i.e. enabled). It should be noted that the event counter EC is not incremented for every L1A received `l1a_in`, but only for those events that actually entered RAM 2. Generated using the HGCAL L1A throttle software model. Different colours in the FSM column encode different FSM states. The highlighted cells in red in the throttle column indicate BXs when the throttle signal was raised.

BC	l1a_in	throttle	l1a_fin	FSM	output_word	EC	words_RAM2	events_RAM2	drain_timeout	no_drain_empty	no_drain_crc_idle	no_drain_global
1	0	0	0	FILL_N_DRAIN_N	0	0	0	0	2	1	0	1
2	1	0	1	FILL_Y_DRAIN_N	0	1	39	1	2	1	0	1
3	1	0	1	FILL_Y_DRAIN_N	0	2	78	2	1	1	0	1
4	1	0	1	FILL_Y_DRAIN_N	0	3	117	3	0	0	0	0
5	1	0	1	FILL_Y_DRAIN_Y	1	4	155	4	0	0	0	0
6	1	0	1	FILL_Y_DRAIN_Y	2	5	193	5	0	0	0	0
...	...	...	...	...	...	...	...	...	...	...	...	...
33	1	0	1	FILL_Y_DRAIN_Y	29	32	1219	32	0	0	0	0
34	1	0	1	FILL_Y_DRAIN_Y	30	33	1257	33	0	0	0	0
35	1	1	0	FILL_N_DRAIN_Y	31	33	1256	33	0	0	0	0
36	1	1	0	FILL_N_DRAIN_Y	32	33	1255	33	0	0	0	0
...	...	...	...	...	...	...	...	...	...	...	...	...
42	1	1	0	FILL_N_DRAIN_Y	38	33	1249	33	0	0	0	0
43	1	1	0	FILL_N_DRAIN_Y	39	33	1248	32	0	0	1	1
44	1	1	0	FILL_N_DRAIN_N	40	33	1248	32	0	0	1	1
45	1	1	0	FILL_N_DRAIN_N	41	33	1248	32	0	0	0	0
46	1	0	1	FILL_Y_DRAIN_Y	1	34	1286	33	0	0	0	0
47	0	1	0	FILL_N_DRAIN_Y	2	34	1285	33	0	0	0	0
48	0	1	0	FILL_N_DRAIN_Y	3	34	1284	33	0	0	0	0
...	...	...	...	...	...	...	...	...	...	...	...	...
83	0	1	0	FILL_N_DRAIN_Y	38	34	1249	33	0	0	0	0
84	0	1	0	FILL_N_DRAIN_Y	39	34	1248	32	0	0	1	1
85	0	1	0	FILL_N_DRAIN_N	40	34	1248	32	0	0	1	1
86	0	1	0	FILL_N_DRAIN_N	41	34	1248	32	0	0	0	0
87	0	0	0	FILL_N_DRAIN_Y	1	34	1247	32	0	0	0	0
88	0	0	0	FILL_N_DRAIN_Y	2	34	1246	32	0	0	0	0
...	...	...	...	...	...	...	...	...	...	...	...	...
124	0	0	0	FILL_N_DRAIN_Y	38	34	1210	32	0	0	0	0
125	0	0	0	FILL_N_DRAIN_Y	39	34	1209	31	0	0	1	1
126	0	0	0	FILL_N_DRAIN_N	40	34	1209	31	0	0	1	1
127	0	0	0	FILL_N_DRAIN_N	41	34	1209	31	0	0	0	0
128	0	0	0	FILL_N_DRAIN_Y	1	34	1208	31	0	0	0	0
129	0	0	0	FILL_N_DRAIN_Y	2	34	1207	31	0	0	0	0
...	...	...	...	...	...	...	...	...	...	...	...	...
1354	0	0	0	FILL_N_DRAIN_Y	38	34	40	2	0	0	0	0
1355	0	0	0	FILL_N_DRAIN_Y	39	34	39	1	0	0	1	1
1356	0	0	0	FILL_N_DRAIN_N	40	34	39	1	0	0	1	1
1357	0	0	0	FILL_N_DRAIN_N	41	34	39	1	0	0	0	0
1358	0	0	0	FILL_N_DRAIN_Y	1	34	38	1	0	0	0	0
1359	0	0	0	FILL_N_DRAIN_Y	2	34	37	1	0	0	0	0
...	...	...	...	...	...	...	...	...	...	...	...	...
1395	0	0	0	FILL_N_DRAIN_Y	38	34	1	1	0	0	0	0
1396	0	0	0	FILL_N_DRAIN_Y	39	34	0	0	0	0	1	1
1397	0	0	0	FILL_N_DRAIN_N	40	34	0	0	2	1	1	1
1398	0	0	0	FILL_N_DRAIN_N	41	34	0	0	2	1	0	1

Table A.3: A BX-level emulation of the internal counters, states and status flags of the HGCAL L1A throttle block when presented with L1A requests arriving when **special words are being streamed out of the HGCROC**, to demonstrate some edge cases of the readout logic. Specifically, these L1A requests arrive when the output words are: IDLE due to empty buffer ($\text{output_word} = 0$), CRC ($\text{output_word} = 40$), the last event word ($\text{output_word} = 39$), mandatory idle ($\text{output_word} = 41$), both CRC and mandatory IDLE back-to-back ($\text{output_word} = 40$ and $\text{output_word} = 41$). Generated using the HGCAL L1A throttle software model. Different colours in the FSM column encode different FSM states.

BC	l1a_in	throttle	l1a_fin	FSM	output_word	EC	words_RAM2	events_RAM2	drain_timeout	no_drain_empty	no_drain_crc_idle	no_drain_global
1399	0	0	0	FILL_N_DRAIN_N	0	45	0	0	2	1	0	1
1400	1	0	1	FILL_Y_DRAIN_N	0	46	39	1	2	1	0	1
1401	0	0	0	FILL_N_DRAIN_N	0	46	39	1	1	1	0	1
1402	0	0	0	FILL_N_DRAIN_N	0	46	39	1	0	0	0	0
1403	0	0	0	FILL_N_DRAIN_Y	1	46	38	1	0	0	0	0
1404	0	0	0	FILL_N_DRAIN_Y	2	46	37	1	0	0	0	0
...	...	...	...	...	...	...	...	...	...	...	...	...
1440	0	0	0	FILL_N_DRAIN_Y	38	46	1	1	0	0	0	0
1441	0	0	0	FILL_N_DRAIN_Y	39	46	0	0	0	0	1	1
1442	1	0	1	FILL_Y_DRAIN_N	40	47	39	1	2	1	1	1
1443	0	0	0	FILL_N_DRAIN_N	41	47	39	1	2	1	0	1
1444	0	0	0	FILL_N_DRAIN_N	0	47	39	1	1	1	0	1
1445	0	0	0	FILL_N_DRAIN_N	0	47	39	1	0	0	0	0
1446	0	0	0	FILL_N_DRAIN_Y	1	47	38	1	0	0	0	0
1447	0	0	0	FILL_N_DRAIN_Y	2	47	37	1	0	0	0	0
...	...	...	...	...	...	...	...	...	...	...	...	...
1482	0	0	0	FILL_N_DRAIN_Y	37	47	2	1	0	0	0	0
1483	0	0	0	FILL_N_DRAIN_Y	38	47	1	1	0	0	0	0
1484	1	0	1	FILL_Y_DRAIN_Y	39	48	39	1	0	0	1	1
1485	0	0	0	FILL_N_DRAIN_N	40	48	39	1	0	0	1	1
1486	0	0	0	FILL_N_DRAIN_N	41	48	39	1	0	0	0	0
1487	0	0	0	FILL_N_DRAIN_Y	1	48	38	1	0	0	0	0
1488	0	0	0	FILL_N_DRAIN_Y	2	48	37	1	0	0	0	0
...	...	...	...	...	...	...	...	...	...	...	...	...
1524	0	0	0	FILL_N_DRAIN_Y	38	48	1	1	0	0	0	0
1525	0	0	0	FILL_N_DRAIN_Y	39	48	0	0	0	0	1	1
1526	0	0	0	FILL_N_DRAIN_N	40	48	0	0	2	1	1	1
1527	1	0	1	FILL_Y_DRAIN_N	41	49	39	1	2	1	0	1
1528	0	0	0	FILL_N_DRAIN_N	0	49	39	1	1	1	0	1
1529	0	0	0	FILL_N_DRAIN_N	0	49	39	1	0	0	0	0
1530	0	0	0	FILL_N_DRAIN_Y	1	49	38	1	0	0	0	0
1531	0	0	0	FILL_N_DRAIN_Y	2	49	37	1	0	0	0	0
...	...	...	...	...	...	...	...	...	...	...	...	...
1567	0	0	0	FILL_N_DRAIN_Y	38	49	1	1	0	0	0	0
1568	0	0	0	FILL_N_DRAIN_Y	39	49	0	0	0	0	1	1
1569	1	0	1	FILL_Y_DRAIN_N	40	50	39	1	2	1	1	1
1570	1	0	1	FILL_Y_DRAIN_N	41	51	78	2	2	1	0	1
1571	0	0	0	FILL_N_DRAIN_N	0	51	78	2	1	1	0	1
1572	0	0	0	FILL_N_DRAIN_N	0	51	78	2	0	0	0	0
1573	0	0	0	FILL_N_DRAIN_Y	1	51	77	2	0	0	0	0
1574	0	0	0	FILL_N_DRAIN_Y	2	51	76	2	0	0	0	0
...	...	...	...	...	...	...	...	...	...	...	...	...
1610	0	0	0	FILL_N_DRAIN_Y	38	51	40	2	0	0	0	0
1611	0	0	0	FILL_N_DRAIN_Y	39	51	39	1	0	0	1	1
1612	0	0	0	FILL_N_DRAIN_N	40	51	39	1	0	0	1	1
1613	0	0	0	FILL_N_DRAIN_N	41	51	39	1	0	0	0	0
1614	0	0	0	FILL_N_DRAIN_Y	1	51	38	1	0	0	0	0
1615	0	0	0	FILL_N_DRAIN_Y	2	51	37	1	0	0	0	0
...	...	...	...	...	...	...	...	...	...	...	...	...
1651	0	0	0	FILL_N_DRAIN_Y	38	51	1	1	0	0	0	0
1652	0	0	0	FILL_N_DRAIN_Y	39	51	0	0	0	0	1	1
1653	0	0	0	FILL_N_DRAIN_N	40	51	0	0	2	1	1	1
1654	0	0	0	FILL_N_DRAIN_N	41	51	0	0	2	1	0	1
1655	0	0	0	FILL_N_DRAIN_N	0	51	0	0	2	1	0	1
1656	0	0	0	FILL_N_DRAIN_N	0	51	0	0	2	1	0	1

Appendix B

Cluster Property Quantisation Effects

This appendix provides information about the agreement between the software emulator and the simulator when $k = 0$ (Figures B.1, B.2, and B.3) for additional cluster properties that were not previously covered in Chapter 6. Among these, $\sigma_{\eta\eta}$ (Figure B.3a) stands out by having worse precision with respect to other properties from the same category (e.g. $\sigma_{\phi\phi}$). It should be noted that this property is only approximated based on the available $\mu(r/z)$ and $\sigma_{\frac{r}{z} \frac{r}{z}}$ which are used to form the look-up table (LUT) address. There, the number of bits used to form it is limited to 12 bits (6 bits each). Another reason for poorer precision is that only 5 bits are used in its final representation, whereas the other σ properties use 7 bits. Increasing either the number of bits in the LUT address or in the output word would enhance precision. However, that would come at the cost of FPGA resource utilisation (e.g. block memory). Lastly, the implementation details for this property should be fine-tuned since they are dependant on an old data set which did not have many samples. Overall, this implementation of the $\sigma_{\eta\eta}$ calculation can be considered as a proxy for a future implementation that may bring better precision, if needed.

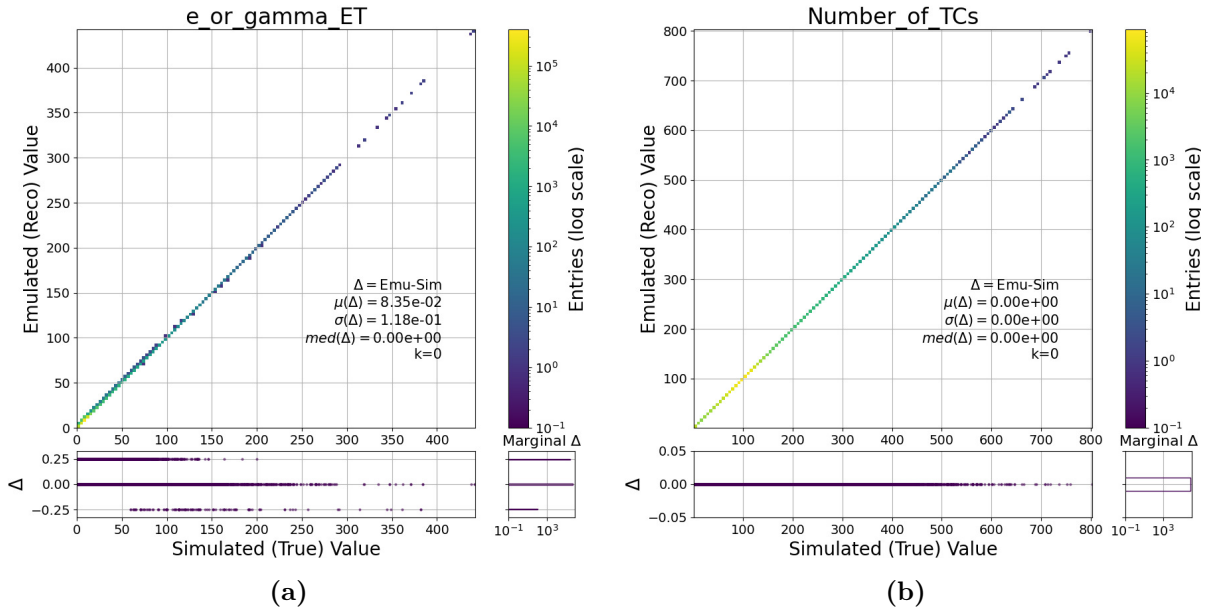


Figure B.1: Agreement between the calculated quantities by the software emulator (firmware implementation/reconstructed) with the software simulator (true) for several cluster properties (Subset 1). For axis units, see Table 6.4.

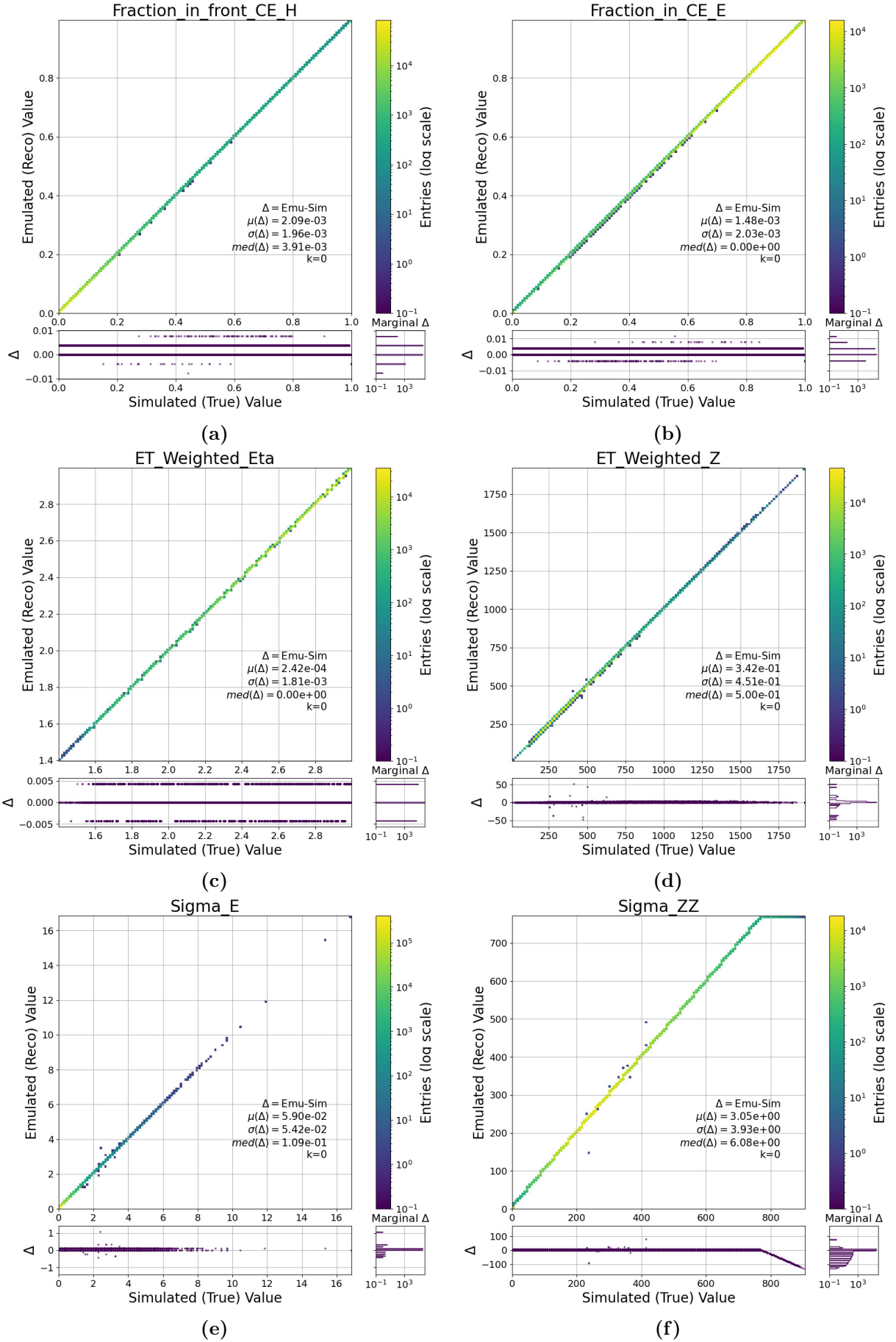


Figure B.2: Agreement between the calculated quantities by the software emulator (firmware implementation/reconstructed) with the software simulator (true) for several cluster properties (Subset 2). For axis units, see Table 6.4.

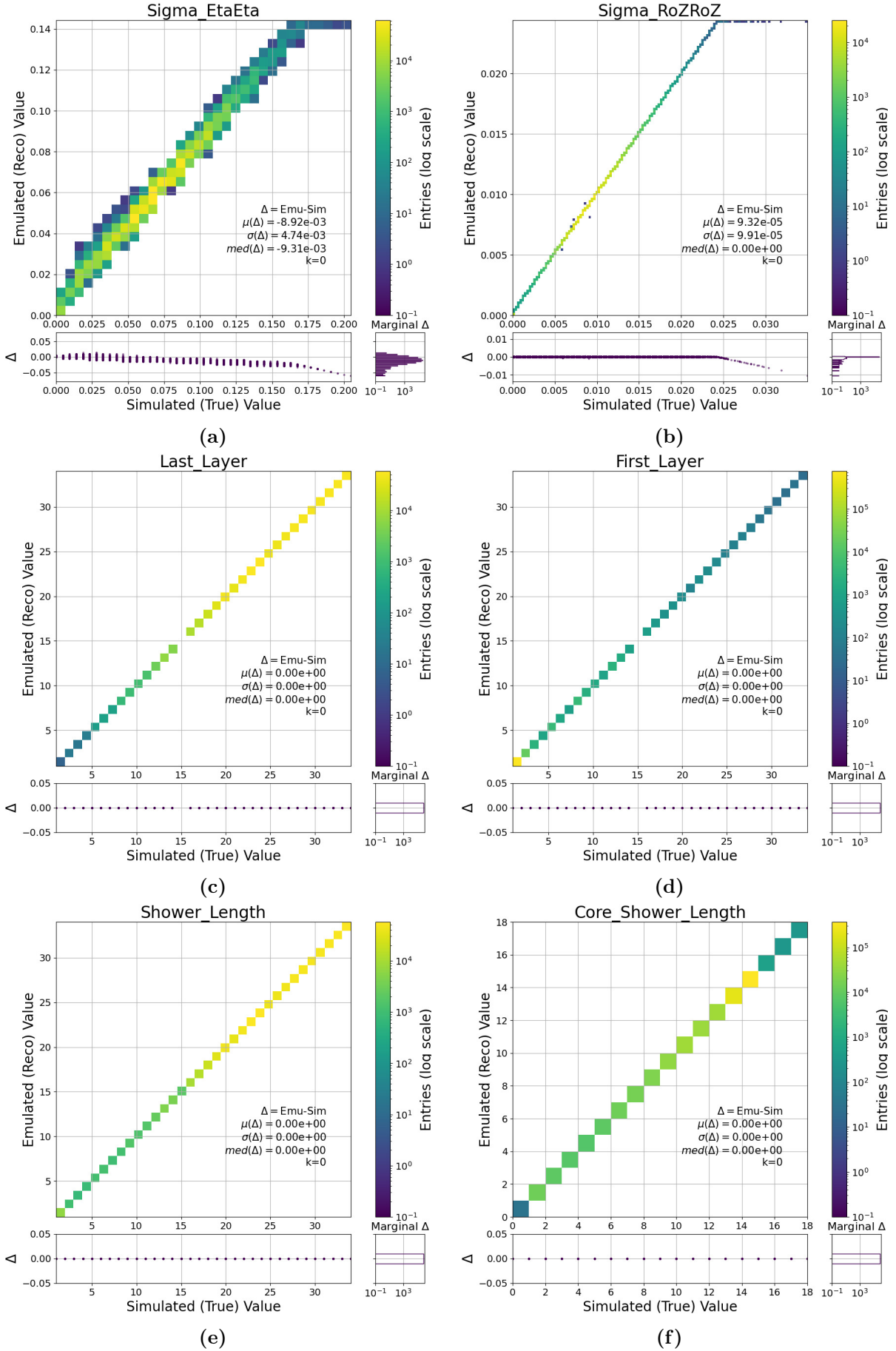


Figure B.3: Agreement between the calculated quantities by the software emulator (firmware implementation/reconstructed) with the software simulator (true) for several cluster properties (Subset 2). For axis units, see Table 6.4.

Appendix C

Dependence of Cluster Properties on k

In addition to the material covered in Chapter 6, this appendix demonstrates the degradation of the precision of two example cluster properties (Figures C.1 and C.2) for several values of the k parameter, which controls the size of the Cluster Sum record. The chosen properties in Figures C.1 and C.2 are $\sigma_{\phi\phi}$ and $\mu(\phi)$ because they cover the two main kinds of properties that involve the generalised weight W in their calculation (which depends on k). Although the precision of these properties obviously degrades with an increase in k , the overall performance of the algorithm remains similar even for $k = 15$, as was earlier appreciated in Chapter 6 (Figure 6.14).

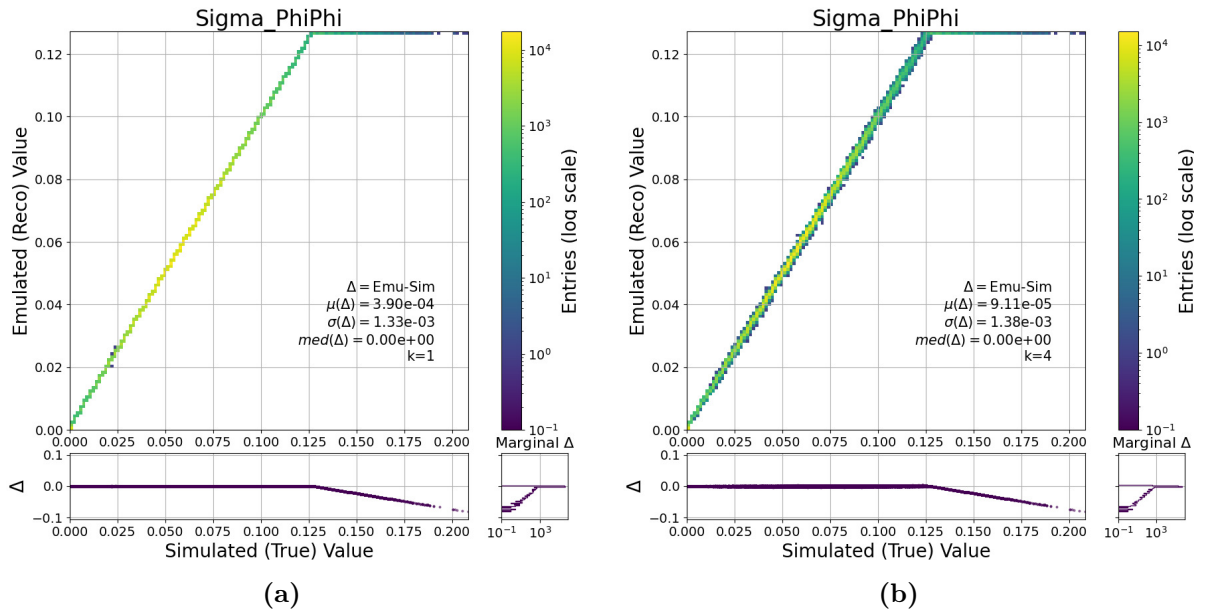


Figure C.1: Visualising the effect of the k parameter on the precision of $\sigma_{\phi\phi}$ for (a) $k=1$, and (b) $k=4$. For axis units, see Table 6.4.

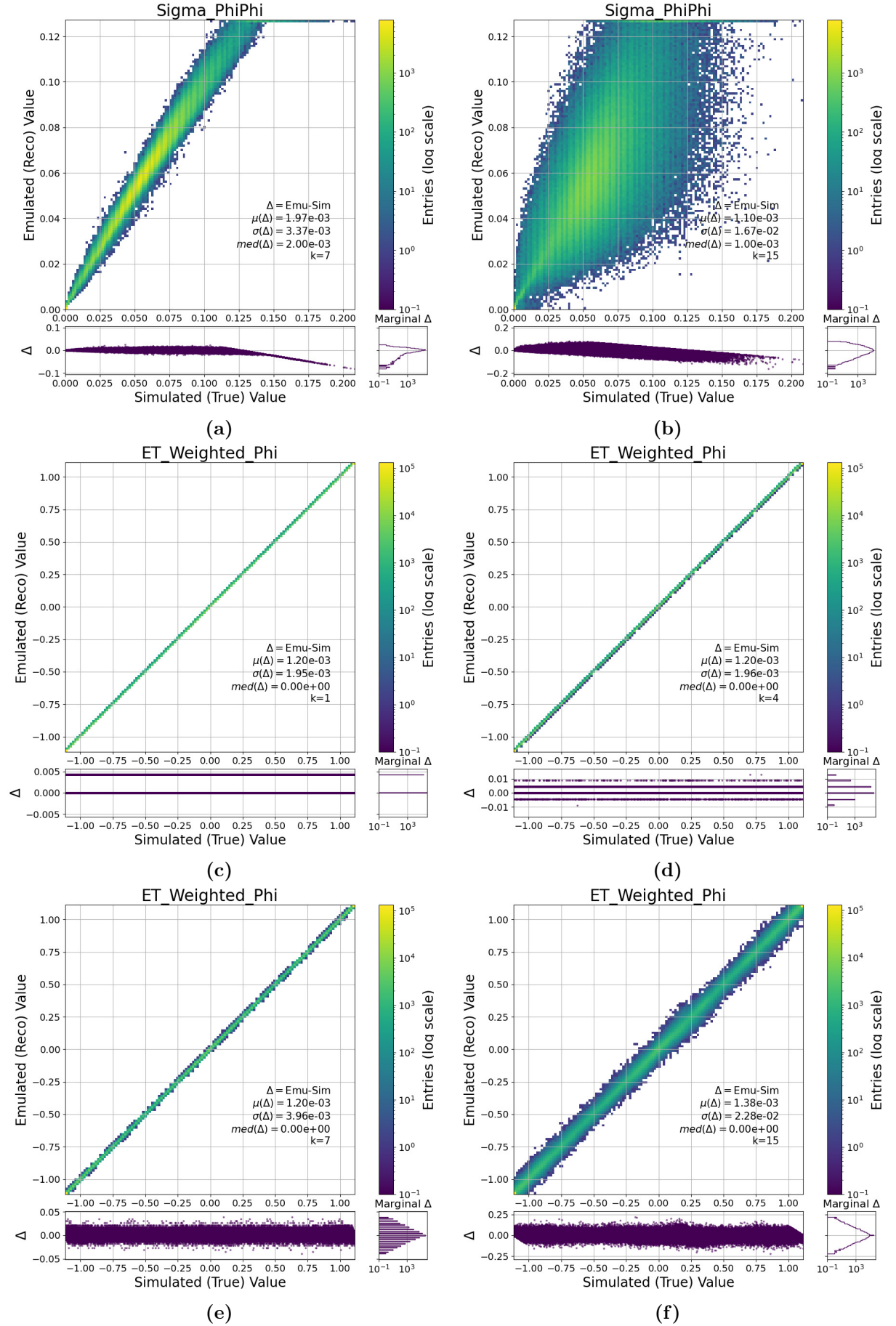


Figure C.2: Visualising the effect of the k parameter on the precision of $\sigma_{\phi\phi}$ for (a) $k=7$, (b) $k=15$, and $\mu(\phi)$ for (c) $k=1$, (d) $k=4$, (e) $k=7$, (f) $k=15$. For axis units, see Table 6.4.