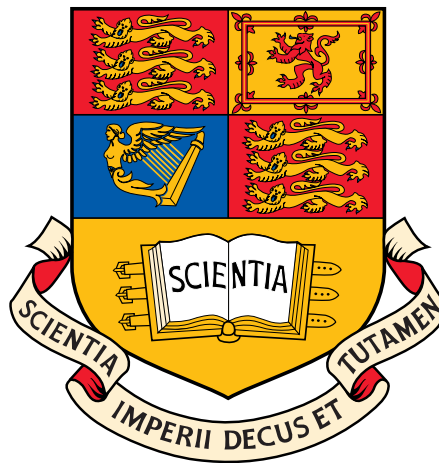


The Level-1 Trigger of the CMS experiment at the LHC and the Super-LHC

Andrew W. Rose

High Energy Physics
Blackett Laboratory
Imperial College
London



A thesis submitted for the degree of
Doctor of Philosophy
of the University of London
and the Diploma of Imperial College

2009



Abstract

The Compact Muon Solenoid experiment at the Large Hadron Collider at CERN observes proton-proton collisions at a centre of mass energy of 14 TeV and a frequency of 40 MHz. This results in a raw data rate exceeding 40 PByte per second, which must be reduced to a rate of 100 MByte per second before storage. This is achieved in two stages, a hardware trigger (Level-1 Trigger) and a software trigger (Higher Level Trigger).

The Global Calorimeter Trigger is a central component in the level-1 trigger, selecting interesting ‘physical’ events from the background of events by identifying calorimetric energy patterns and features. The development of the Global Calorimeter Trigger required the use of several modern telecommunication technologies and is discussed here, with particular emphasis on the design and testing of the Source Card.

Although the Large Hadron Collider has yet to start operation, design of an upgraded accelerator (called the Super Large Hadron Collider) is already under way. This upgrade will improve on the design luminosity of the LHC by a factor of 10, increasing the number of collisions per bunch crossing by the same factor. Such an increase would require changes to both the algorithms and the architecture of the level-1 trigger, in particular requiring the inclusion of tracking information. Several possible tracker and trigger architectures for high luminosity running are shown here. By introducing an on-detector p_T -cut, one of the schemes offers a reduction in the tracker data rate by approximately two orders of magnitude, this being sufficient to allow the inclusion of tracking data in the level-1 trigger. A case study into the identification of electrons in high pile-up environments using this kind of tracking information is also presented; it is shown that the use of tracking information may be used to recover the current level-1 trigger rate.

Acknowledgements

This thesis would not have been possible without the time, thoughts, help and advice of many different people.

In particular, my thanks go to...

Costas, my supervisor, for all the help and guidance he has provided over the last few years and for introducing me to High Energy Physics in the first place. John Jones, for the enormous amount of time, advice and education given when learning to use FPGAs, VHDL, USB, the IDAQ and everything else. Gregory Iles, Magnus Hansen, Matt Stettler and Matt Noy, their advice, firmware and hardware experience were invaluable! Alex Tapper, Rob Frazier and Jim Brooke, again, their experience with all things software was invaluable! Mark Raymond for his ability to listen to any problem and draw a circuit diagram that will solve it. Sarah Greenwood, Maria Khaleeq, Vera Kasey, Dave Price and 'Oz' in the electronic workshop for their help and advice, and having the patience to make all the various fiddly hardware modifications! Jan Troska for his help testing the Source Card optical links. Pam Klabbers, for her help in the RCT to Source Card integration tests. Geoff Hall, Anders Ryd, Harry Cheung, Mark Pesaresi, Laura Fields and everyone else working on the tracker upgrade. Jamie Ballin and Adam Dobbs for always being free to discuss a problem, an idea, CMS software or anything else over a cup of coffee. Dave Lee, for letting me stick my nose in and design the FETS laser-wire readout board and for also being ready with a cup of coffee. Everyone else at Imperial who had an answer when I needed it.

And final my thanks to Bekka and to my family for all their support, both emotional and physical, over the last four years, but especially for those occasions when I was too tired, busy or stressed to notice it at the time.

Again, to you all, my deepest thanks,

A.R.

Author's Contributions

The first half of this document relates to the design and development of the Global Calorimeter Trigger at the CMS experiment at the LHC, with particular focus on the source card. The author was involved in the Global Calorimeter Trigger project from its inception. Of the work described in chapters 2-5, the following are the author's contribution:

- Certain design choices of the source card.
- Development and maintenance of the source card firmware and software, including the interface to the trigger supervisor.
- Verification of the RCT-GCT cables and of the suitability of optical components.
- Hardware, firmware and software design of the RCT emulator card and associated system.
- Standalone and integration testing of the source cards.
- Design of the GCT data-flow architecture.
- Design of the on-line USB control-system, including hardware and software design of the USB-hub carrier card and source card monitor panel.
- Design of the distribution of hardware and of the RCT cable routing.

The second half of this document describes studies into possible upgrades to the calorimeter trigger and studies into the inclusion of tracking information in the level-1 trigger. Of the work described in chapters 6-8, the following are the author's contribution:

- Studies into the replacement of the regional calorimeter trigger.
- Studies into calorimeter-seeded tracker readout.
- Co-architect of the stacked tracker concept.
- Validation of the “fast” and “full” simulation.
- Implementation of the first stacked tracker geometry in the simulation framework.
- Implementation of a framework for the study of tracking triggers within the existing simulation framework.
- Studies of object multiplicities, efficiencies, rates, purities and resolutions for tracking triggers.
- Studies into single electron triggers using tracking trigger information.

Included in Appendix A is a discussion of the Front End Test Stand experiment. The authors contribution to this work was the design and the standalone (non-beamline) testing of the readout hardware.

0.0	Contents	9
------------	-----------------	----------

Contents

Abstract	3
Acknowledgements	5
Author's Contributions	7
Contents	9
List of Figures	14
List of Tables	22
Chapter 1. Introduction	25
1.1 The Standard Model of Particle Physics	26
1.1.1 The Standard Model Lagrangian	27
1.1.2 Searches for the Higgs Boson	34
1.1.3 Objections to the standard model	36
1.2 The Large Hadron Collider (LHC)	40
1.3 The Compact Muon Solenoid (CMS)	44
1.3.1 The Silicon Tracker	45
1.3.2 The Electromagnetic Calorimeter (ECAL)	48

1.3.3	The Hadronic Calorimeter (HCAL)	49
1.3.4	The Muon Detectors	50
1.3.5	The CMS Trigger System	50
1.3.6	The CMS Level-1 Trigger	52
1.3.7	The CMS Level-1 Trigger Software	53
Chapter 2. The CMS Calorimeter Trigger System		55
2.1	The Calorimeter Trigger Algorithms	57
2.1.1	The Electromagnetic Candidate Algorithm	57
2.1.2	The Jet Candidate Algorithm	59
2.1.3	Other calorimeter trigger tasks	60
2.2	The Calorimeter Trigger System	62
2.2.1	The Calorimeter Trigger-Primitive Generators (TPGs)	62
2.2.2	The Regional Calorimeter Trigger (RCT)	63
2.2.3	The Global Calorimeter Trigger (GCT)	65
2.2.4	The Global Trigger (GT)	66
2.3	Design choices for the GCT	68
Chapter 3. The Global Calorimeter Trigger Source Card		77
3.1	The GCT Source Card	77
3.2	Design of the Source Card	80
3.2.1	High level hardware design	80
3.2.2	Serial link design	81
3.2.3	Clock system design	82
3.2.4	Power system design	85
3.2.5	USB interface	86

0.0	Contents	11
------------	-----------------	-----------

3.2.6	JTAG interface	87
3.2.7	The Source Card Firmware Design	89
3.2.8	The Source Card Front Panel Design	92
3.3	Stand-alone Testing of the GCT Source Cards	93
3.3.1	Bit Error Rate (BER) Tests and Poisson Statistics	94
3.3.2	Cable testing	95
3.3.3	Source Card ECL front end	98
3.3.4	Optical Link Tests	102
3.3.5	TTCrx and QPLL tests	108
3.4	Integration of the GCT Source Cards	111
3.4.1	Integration of the GCT Source Cards with the RCT	111
3.4.2	Integration of the GCT Source Cards with the GCT Leaf Cards	114
Chapter 4. The Global Calorimeter Trigger Schema		117
4.1	Data-Flow Architecture	117
4.2	Control Architecture	125
4.2.1	USB Architecture	125
4.2.2	TTC Architecture	130
Chapter 5. The Global Calorimeter Trigger Software		131
5.1	Software Architecture	131
5.1.1	Software Structure	131
5.1.2	The Trigger Supervisor	136
5.1.3	Source Card Capture-file Interface	140

Chapter 6. The Super-LHC, Super-CMS and the Level-1 Trigger	145
6.1 Replacement of the Regional Calorimeter Trigger	147
6.2 Calorimeter-seeded tracker readout	153
6.3 A Self-contained Tracking Trigger for CMS at the SLHC	162
6.3.1 Reducing the tracker data volume	163
6.3.2 The theoretical basis of the stacked tracking concept	167
 Chapter 7. A Monte-Carlo study of Tracking Triggers	 175
7.1 Introduction	175
7.1.1 CMSSW	175
7.2 Validation of “Fast” Simulation against “Full” Simulation	178
7.3 A framework for the study of Tracking Triggers	182
7.3.1 Clustering algorithms	188
7.3.2 Hit-matching algorithms	190
7.3.3 Tracklets	194
7.4 Initial Studies with Tracking Triggers	194
7.4.1 Multiplicities	197
7.4.2 Efficiencies	199
7.4.3 Rates	204
7.4.4 Purities	207
7.4.5 Incorrect reconstruction of secondary tracks	209
7.4.6 Removing the vertex constraint	211
7.4.7 p_T -resolutions	212
7.4.8 Vertex reconstruction	218
7.5 Summary	222

0.0	Contents	13
------------	-----------------	-----------

Chapter 8.	A Super-LHC Electron Trigger	225
8.1	Two Level-1 Electron Trigger Philosophies	226
8.2	The “Inside-Out” Electron Trigger	228
8.2.1	Preparatory study	228
8.2.2	Background Rates	231
8.2.3	Efficiency	232
8.2.4	Balancing efficiency and rate reduction	234
8.3	Summary	235
Chapter 9.	Conclusions	237
Appendix A.	The Front End Test Stand laser-wire	241
A.1	The Front End Test Stand (FETS)	243
A.2	A readout system for the FETS laser-wire	245
A.3	Tests of the FETS laser-wire readout system	246
Appendix B.	Circular Geometry	251
Appendix C.	2-D Clustering Algorithm	253
Appendix D.	Tracker upgrade geometries	255
	Glossary	257
	References	261

List of Figures

1.1	Feynman diagrams of the trilinear and quadrilinear couplings of the standard model gauge bosons	28
1.2	Measurement of the cross section for the channel $e^+e^- \rightarrow W^+W^-$ taken by the LEP experiments	29
1.3	Feynman diagram of the coupling between fermions and gauge-bosons in the standard model	30
1.4	Feynman diagrams of the $F_\mu F^\mu h$ and $F_\mu F^\mu hh$ vertices	31
1.5	Theoretical bounds on the mass of the Higgs boson	35
1.6	Latest results of searches for the Higgs boson	36
1.7	Feynman diagrams for the three sources of ultraviolet divergences in the standard model	37
1.8	Feynman diagrams for the process $WW \rightarrow X \rightarrow WW$	37
1.9	Higgs production channels at the LHC	41
1.10	Higgs decay modes in the mass range accessible at the LHC	42
1.11	The production cross-sections of a selection of possible production modes at the LHC as a function of interaction energy	43
1.12	A comparison of $H \rightarrow \gamma\gamma$ signal and background for a Higgs mass of 130 GeV and 100 fb ⁻¹ of data recorded	45
1.13	Diagram of the CMS Detector	46
1.14	Layout of a quarter of the CMS tracking detector	47

1.15	The CMS Trigger and Data Acquisition System	51
1.16	The CMS level-1 trigger	52
2.1	The calorimeter trigger e/γ algorithm	59
2.2	The calorimeter trigger jet algorithm	60
2.3	The calorimeter TPG and DAQ architecture	64
2.4	The RCT architecture	65
2.5	The GCT architecture	66
2.6	The GT architecture	67
2.7	The GCT source card	70
2.8	The GCT architecture	70
2.9	The GCT leaf card	71
2.10	The GCT concentrator card	72
2.11	The GCT wheel card	73
2.12	The GCT matrix processor Card	75
3.1	Block Diagram of the CMS GCT source card	80
3.2	Key features of the CMS GCT source card	82
3.3	The source card clock system	83
3.4	The source card JTAG chain	88
3.5	Schematic of data flow through the CMS GCT source card	94
3.6	The RCT-GCT cable testing apparatus	96
3.7	The RCT Emulator card	98
3.8	Photograph of the experimental set-up used in the testing of the CMS GCT source cards	99
3.9	Schematic layout of the two tests of the source card ECL inputs	101

0.0	List of Figures	17
3.10	Schematic layout of the initial tests of the source card optical links	103
3.11	Schematic layout of the two physical tests of the source card optical links	105
3.12	Eye diagrams resulting from the first two tests of the source card optical links	106
3.13	Schematic layout of the indirect test of the source card optical links	107
3.14	Extrapolated BER as a function of fibre attenuations on the optical links	108
3.15	Test configuration used for the production testing of the optical links	109
3.16	Four plots of word error rate against the TTCrx fine-delay setting made during the RCT-source card integration tests	113
4.1	The bit structure of the six RCT outputs	120
4.2	The chosen routing scheme for electron and muon data	121
4.3	The chosen routing scheme for the RCT regional data for regions far from the $\eta = 0$ boundary	122
4.4	The chosen routing scheme for the RCT regional data for regions near to the $\eta = 0$ boundary	123
4.5	The prototype (V2) USB-hub carrier card for the GCT source card system	126
4.6	The monitor panel for the GCT source card system	128
4.7	Schematic of the GCT source card USB control system	129
5.1	The source card software structure	132
5.2	The trigger supervisor structure	137
5.3	Screen shots of the XDAQ and Trigger Supervisor “at-a-glance” source card summary tables	141
5.4	Screen shots of the trigger supervisor master panel	141
6.1	Single muon trigger rates for CMS at the LHC as a function of p_T threshold for the level-1 and higher level triggers	147

6.2	The split stage duplication/sharing architecture chosen for the modular implementation of the RCT electron-finder algorithm	149
6.3	The method used for calorimeter-seeded tracker readout studies	155
6.4	Distribution of the distance of tracker hits from a calorimeter seed point in the ϕ -direction	157
6.5	Distribution of the distance of tracker hits from a calorimeter seed point in the z -direction	159
6.6	Number and density of hits and “digis” plotted against radius for 200 pile-up events	163
6.7	Particle count per event related to track- p_T for 200 pile-up events	165
6.8	Tracker Combinatorials under high pile-up conditions	166
6.9	Demonstration of the equivalence of an angular cut and a p_T -cut	167
6.10	The size and offset of the track-finding window in the z -direction	169
6.11	Sensor offset in the z -direction with increasing pseudorapidity	171
6.12	Geometric method of measuring the r_{curve} , and thus p_T , from two tracker points and the nominal interaction point	173
7.1	Data flow within the CMSSW software package	176
7.2	Comparison of the densities of simulated and digitized tracker hits predicted by the “full” and “fast” simulation	180
7.3	The proof of principle geometry used to demonstrate the feasibility of the stacked tracking concept	183
7.4	The StackedTracker utility classes for interfacing the tracker geometry with trigger simulations	186
7.5	The SLHC L1Trigger classes: data-types and tools for describing a level-1 tracking trigger	188
7.6	Mean frequency of occurrence of a particular number of simulated hits, digitized hits and clusters per sensor for each clustering algorithm	191
7.7	p_T distribution of tracks accepted and rejected by each clustering algorithm	192

0.0	List of Figures	19
------------	------------------------	-----------

7.8	The effective mis-reconstruction of tracks with large crossing-angle	193
7.9	The distinct crenellations in the stub density which appeared when the geometry was changed to the Long-Barrel tracker geometry	195
7.10	Simulated estimates of stub multiplicities in the Strawman-B geometry, for four different clustering algorithms	198
7.11	Simulated estimates of tracklet multiplicities in the Strawman-B geometry, for four different clustering algorithms	200
7.12	Simulated estimates of single particle stub-finding efficiencies as a function of the p_T of the Monte Carlo-truth track	202
7.13	Failure of digitized hits to pass a p_T -cut that is passed by the equivalent simulated hits	203
7.14	Simulated estimates of single particle stub-finding efficiencies of the local geometry hit-matching algorithm as a function of the p_T of the Monte Carlo-truth track	203
7.15	Simulated estimates of single particle stub-finding efficiencies as a function of the pseudorapidity of the Monte Carlo-truth track	204
7.16	Simulated estimates of single particle tracklet-finding efficiencies as a function of the p_T of the Monte Carlo-truth track	205
7.17	Simulated estimates of tracker hit and stub rates as a function of the pseudorapidity	206
7.18	Simulated estimates of tracklet rates as a function of the pseudorapidity	207
7.19	Simulated estimates of purities of digitized hits and stubs, constructed from digitized hits as a function of the pseudorapidity	208
7.20	Simulated estimates of purities of tracklets constructed from digitized hits as a function of the pseudorapidity	209
7.21	Reconstruction of secondary tracks as primary tracks	209
7.22	Simulated distribution of accepted tracks as a function of the Monte Carlo p_T and by the radial position of the track's origin	210
7.23	Simulated estimates of single particle three-point tracklet-finding efficiencies as a function of the p_T and of the pseudorapidity of the Monte Carlo-truth track	212

7.24	Simulated estimates of three-point tracklet rates as a function of the pseudorapidity	213
7.25	Simulated estimates of purities of three-point tracklets, constructed from digitized hits as a function of the pseudorapidity	213
7.26	Simulated p_T -resolution calculated using two tracker points + the nominal vertex, three tracker points and three tracker points + the nominal vertex	215
7.27	Simulated p_T -resolution vs. track direction in ϕ for 2-point tracklets with the nominal vertex set at $x=400\ \mu\text{m}$ and $x=0\ \mu\text{m}$	216
7.28	Simulated p_T -resolution calculated using two tracker points + the adjusted vertex, three tracker points and three tracker points + the adjusted vertex and using three different fitting techniques	217
7.29	Simulated p_T -resolution calculated using two tracker points + the adjusted vertex, three tracker points and three tracker points + the adjusted vertex and using three different fitting techniques	219
7.30	Simulated p_T -resolution calculated using two tracker points + the adjusted vertex, three tracker points and three tracker points + the adjusted vertex and using three different fitting techniques	219
7.31	Simulated estimates of the distance of closest approach calculated using three-point tracklets plotted against the ϕ -position of the inner stub	220
7.32	Simulated resolution of the calculated z-position of the interaction point for both two- and three-point tracklets	221
8.1	The HLT e/γ algorithm	226
8.2	Two electron trigger philosophies: “outside-in” and “inside-out”	227
8.3	Simulated resolution with which two- and three-point tracklets can identify the location that tracks hit the calorimeter in the ϕ -direction	230
8.4	Resolution with which two- and three-point tracklets can identify the location that tracks hit the calorimeter in the η -direction	231
8.5	Resolution with which upgraded level-1 e/γ candidates can identify the location that tracks hit the calorimeter in the η - and ϕ - directions	232

0.0	List of Figures	21
------------	------------------------	-----------

8.6	Mean rate of calorimeter candidates, two-point+calorimeter candidates and three-point+calorimeter candidates as a function of applied E_T and p_T threshold for background events	233
8.7	Efficiency/Multiplicity of calorimeter candidates, two-point+calorimeter candidates and three-point+calorimeter candidates as a function of applied E_T and p_T threshold for single-electron events	234
8.8	Background rate as a function of single-particle efficiency/multiplicity for calorimeter candidates, two-point+calorimeter candidates and three-point+calorimeter candidates	235
A.1	The upgraded injector chain proposed for SLHC	242
A.2	Damage done to the scintillator screens at the FETS	244
A.3	The system of transposable and rotatable mirrors used to control the FETS laser-wire	245
A.4	The readout board of the FETS laser-wire system	247
A.5	Results of charge injection tests of the FETS readout board	247
A.6	Background charge collection using the FETS readout board running synchronously with the ion beam	249
B.1	The geometry relevant to the stacked tracker	251
C.1	The 2-D clustering algorithm	253
D.1	The Strawman-B, Long-Barrel and Hybrid tracker upgrade geometries	256

List of Tables

5.1	Field structure of the source card capture file	140
6.1	Estimation of the number of logic devices required for an upgraded calorimeter trigger based on bandwidth requirements alone	152
6.2	Distribution of the distance of tracker hits from a calorimeter seed point in the ϕ -direction	156
6.3	Distribution of the distance of tracker hits from a calorimeter seed point in the z -direction	158
7.1	Comparison of the densities of simulated and digitized tracker hits predicted by the “full” and “fast” simulation	179
7.2	Simulated estimates of stub multiplicities by layer, clustering algorithm and hit-matching algorithm	198
7.3	Simulated estimates of tracklet multiplicities by layer, clustering algorithm and hit-matching algorithm	200
7.4	Computational complexity of four methods of measuring r_{curve}	214
8.1	The chosen window sizes in η and ϕ for the association of tracker candidates with calorimeter candidates	232

Chapter 1

Introduction

“In the attitude of silence the soul finds the path in a clearer light,
and what is elusive and deceptive resolves itself into crystal clearness.
Our life is a long and arduous quest after Truth.”

- Mahatma Gandhi (1869 - 1948)

The search for a coherent description of the universe in which we live has been a constant feature throughout mankind’s history, from the five classical ‘elements’ [1] of antiquity, through the middle ages with the formation of the scientific method in the Middle East [2] and through the scientific revolution of the 17th and 18th centuries. In modern times, the emphasis in fundamental physics has been on the unification of seemingly disparate aspects of nature into a single framework, starting with Maxwell’s unified description of the electric and magnetic fields and culminating in the Standard Model of Particle Physics [3, 4, 5, 6, 7, 8].

Each generation of physicists has brought about advancement and understanding but, with each step taken, the bar is raised for future generations so that new discovery and insight requires more precise measurements, more complicated mathematical descriptions and the need to study nature on ever more extreme scales. The rare and fleeting nature of the particles currently being sought requires the current generation of experiments in particle physics to look for ever

smaller, but significant, features in an ever larger background of less significant events. In addition the accumulation and process of data needs to be at ever higher speed. This requirement for speed necessitates the use of dedicated processing hardware rather than software algorithms. These studies considers the hardware used for this purpose and how the lessons learned can be applied to future experiments in particle physics.

1.1 The Standard Model of Particle Physics

The most detailed and fundamental model of nature is the Standard Model of Particle Physics. The model describes all matter as being composed of spin- $\frac{1}{2}$ particles whose interactions arise as a direct consequence of insisting upon local gauge invariance, a prerequisite [9] to making a quantum field theory renormalizable and thus meaningful. By insisting that the Lagrangian describing any spin-0 or spin- $\frac{1}{2}$ field is invariant under rotation in $SU(3)_{colour}$ -, $SU(2)_{isospin}$ - and $U(1)_{hypercharge}$ - spaces, it is necessary to respectively introduce 8, 3 and 1 spin-1 gauge fields to account for all possible local variations in phase. These fields are observed as gluons, the $W^\pm$ & Z^0 bosons and the photon (γ). Whilst the $SU(3)_{colour}$ symmetry is thought to be exact, the $SU(2)_{isospin} \times U(1)_{hypercharge}$ ‘electroweak’ symmetry is said to be broken. Without symmetry breaking, the massless gauge bosons required by quantum field theory cannot be reconciled with the massive $W^\pm$ and Z^0 bosons seen in nature without breaking the gauge invariance of the theory. The simplest known mechanism for electroweak symmetry breaking is the Higgs mechanism.

Whilst it is legitimate to add terms on an ad hoc basis for the mass of the scalar particles to the Lagrangian, Nature’s asymmetric treatment of left-handed and right-handed spinor states precludes a similar treatment of the spinor fields; instead, the fermion masses must be introduced by the Yukawa coupling of the Higgs field to the fermion field.

1.1.1 The Standard Model Lagrangian

The Standard Model Lagrangian is constructed broadly of four terms, equation 1.1:

$$\mathcal{L}_{Standard Model} = \mathcal{L}_{Gauge Boson} + \mathcal{L}_{Fermion} + \mathcal{L}_{higgs Boson} + \mathcal{L}_{Yukawa Coupling} \quad (1.1)$$

with each term defined as follows

1.1.1.1 $\mathcal{L}_{Gauge Boson}$

$$\mathcal{L}_{Gauge Boson} = -\frac{1}{4}B^{\mu\nu}B_{\mu\nu} - \frac{1}{4}W_a^{\mu\nu}W_{\mu\nu}^a - \frac{1}{4}G_a^{\mu\nu}G_{\mu\nu}^a \quad (1.2)$$

where

$$B_{\mu\nu} = \partial_\mu B_\nu - \partial_\nu B_\mu \quad (1.3)$$

$$W_{\mu\nu} = \partial_\mu W_\nu^a - \partial_\nu W_\mu^a + g_2 f_2^{abc} W_\mu^b W_\nu^c \quad (1.4)$$

$$G_{\mu\nu} = \partial_\mu G_\nu^a - \partial_\nu G_\mu^a + g_3 f_3^{abc} G_\mu^b G_\nu^c \quad (1.5)$$

where B_μ , W_μ and G_μ are the $U(1)_{hypercharge}$, $SU(2)_{isospin}$ and $SU(3)_{colour}$ gauge fields and f_2^{abc} and f_3^{abc} are, respectively, the structure constants of the $SU(2)_{isospin}$ and $SU(3)_{colour}$ gauge group. For the field W_μ the indices a, b, c run from 1 to 3 and for the field G_μ , from 1 to 8, representing the 3 and 8 generators of the $SU(2)_{isospin}$ and $SU(3)_{colour}$ groups, respectively. The terms $g_2 f_2^{abc} W_\mu^b W_\nu^c$ and $g_3 f_3^{abc} G_\mu^b G_\nu^c$ must be included to maintain gauge invariance when taking into account the non-Abelian nature of the $SU(2)_{isospin}$ and $SU(3)_{colour}$ groups. Their inclusion creates Lagrangian terms which are cubic and quartic with respect to the field and which are not seen in the Abelian $U(1)_{hypercharge}$ group. These terms manifest themselves as trilinear and quadrilinear couplings of the weak and strong gauge bosons, shown in figure 1.1. Such couplings were demonstrated by the experiments at **LEP** (Large Electron-Positron Collider). Figure 1.2 is one of the clear indications that the standard model provides the most accurate description of the data [10].

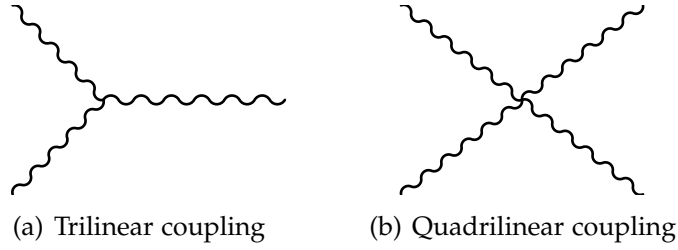


Figure 1.1: Feynman diagrams of the trilinear and quadrilinear couplings introduced by the necessary inclusion of cubic and quartic terms in the Lagrangian of non-Abelian fields required to maintain gauge invariance. Here the lines may represent the photon & weak gauge bosons ($\gamma/W^\pm/Z^0$), or the strong gauge bosons, the gluons.

1.1.1.2 $\mathcal{L}_{Fermion}$

$$\mathcal{L}_{Fermion} = \bar{Q}_i i \not{D} Q_i + \bar{u}_i i \not{D} u_i + \bar{d}_i i \not{D} d_i + \bar{L}_i i \not{D} L_i + \bar{e}_i i \not{D} e_i \quad (1.6)$$

where Q and L respectively represent the left-handed Quark- and Lepton- $SU(2)_{isospin}$ doublet and u , d and e represent the right-handed u -type quark, d -type quark and charged-lepton singlets and the index, i , runs from 1 to 3 representing the three generations. $\not{D}$ is the gauge covariant derivative, D_μ , in Feynman slash notation, $\not{D} = \gamma^\mu D_\mu$. The gauge covariant derivative, rather than the standard derivative, is required to ensure gauge invariance under $U(1)_{hypercharge}$, $SU(2)_{isospin}$ and $SU(3)_{colour}$ local gauge transformations. In field theories, a gauge covariant derivative of the form $D_\mu = \partial_\mu - i\alpha F_\mu$ is necessary, and sufficient, to account for any arbitrary gauge transformations imposed upon the field on which the derivative is acting. The necessary inclusion of the vector field in making the derivative gauge covariant results in the introduction of terms describing field-field interactions. If we consider, for example, the Lagrangian term $\bar{L}_i i \gamma^\mu D_\mu L_i$ and the $SU(2)_{isospin}$ gauge covariant derivative, $D_\mu = \partial_\mu - i\frac{g_2}{2} \sigma^a W_\mu^a$, substituting the latter into the former results in a Lagrangian term of the form $\bar{L}_i W_\mu^a L_i$ which can be seen to describe the interaction of a fermion with a gauge boson as shown in figure 1.3.

As it is now known that neutrinos have mass and thus a right handed state, an additional term of the form $\bar{n}_i i \not{D} n_i$, where n indicates the right-handed neutral-lepton singlet, must be included. It is also now known that the interactions of

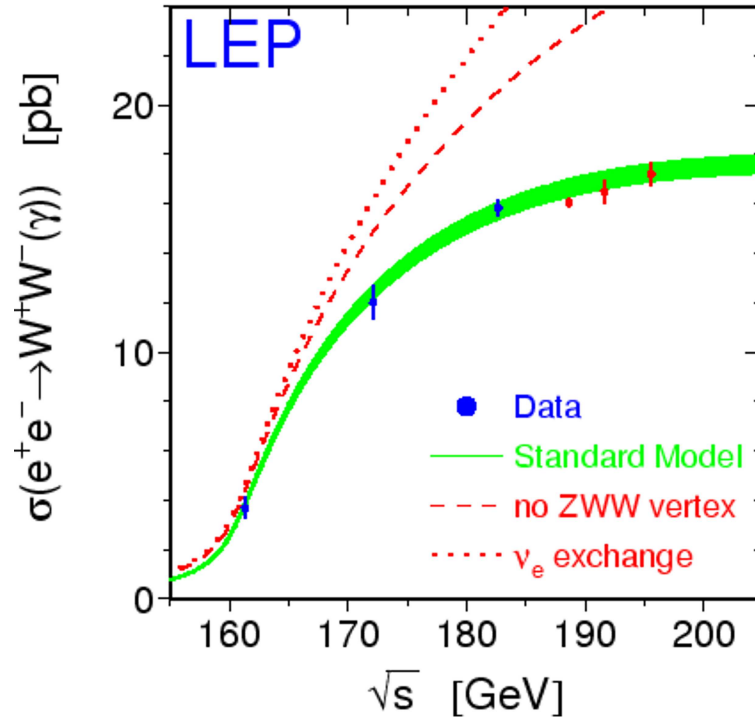


Figure 1.2: Measurements of the cross section for the channel $e^+e^- \rightarrow W^+W^-$ taken by the LEP experiments, compared with three hypotheses: the standard model (solid), the standard model with no $Z^0W^+W^-$ vertices (dashed) and the standard model with no trilinear vertices (dotted).

the three generations are not independent, but interact by the **CKM** (Cabibbo-Kobayashi-Maskawa) mechanism in the quark sector and **PMNS** (Pontecorvo-Maki-Nakagawa-Sakata) mechanism in the lepton sector. These interactions may be included by modification of the above expression, such that the terms $\bar{Q}_i i \not{D} Q_i$ and $\bar{L}_i i \not{D} L_i$ take the form $\bar{Q}_i i \not{D}_{ij} Q_j$ and $\bar{L}_i i \not{D}_{ij} L_j$ where $\not{D}_{ij}$ indicates the inclusion of the relevant mixing matrices into the $SU(2)_{isospin}$ component of the gauge covariant derivative. No modifications are required to right-handed singlet terms as there is no $SU(2)_{isospin}$ coupling to right-handed states.

1.1.1.3 $\mathcal{L}_{higgs\ Boson}$

$$\begin{aligned} \mathcal{L}_{higgs\ Boson} &= |D_\mu \{h+v\}_2|^2 - \lambda |\{h+v\}_2|^4 + \lambda v^2 |\{h+v\}_2|^2 \\ &= |D_\mu \{h+v\}_2|^2 - \frac{\lambda}{4} \cdot h^4 - \lambda v \cdot h^3 - \lambda v^2 \cdot h^2 + constant \quad (1.7) \end{aligned}$$

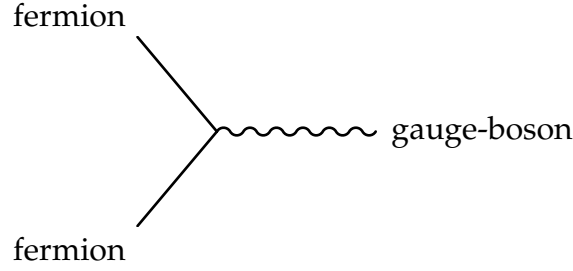


Figure 1.3: An example of a fermion/gauge-boson vertex of the type arising from the use of the gauge covariant derivative, rather than the standard derivative, as is required to ensure gauge invariance under local gauge transformations.

where h is the Higgs field and v is the vacuum expectation valueⁱ of the field. The notation, $\{h+v\}_2$, is used to represent the $SU(2)_{isospin}$ doublet, $\begin{pmatrix} 0 \\ \frac{1}{\sqrt{2}}(h+v) \end{pmatrix}$.

The h^4 and h^3 terms respectively represent quadrilinear and trilinear couplings of the Higgs field, analogous to those in figure 1.1, while the h^2 term, represents a mass term. Direct comparison of equation 1.7 with the Klein-Gordon Lagrangianⁱⁱ yields the identity:

$$mass_h = \sqrt{2\lambda} v \quad (1.8)$$

As before, the gauge covariant derivative, rather than the standard derivative, is required to ensure gauge invariance under $U(1)_{hypercharge}$ and $SU(2)_{isospin}$ local gauge transformations.

If the $SU(2)_{isospin} \times U(1)_{hypercharge}$ gauge covariant derivative, $D_\mu = \partial_\mu - i\frac{g_2}{2}\sigma^a W_\mu^a - i\frac{g_1}{2}B_\mu$ is considered, substitution into the first term of equation 1.7 may be seen to result in Lagrangian terms of the form:

$$\begin{aligned} \frac{1}{2}\partial^\mu h \partial_\mu h &+ \frac{g_2^2}{4}|W_\mu^1 + iW_\mu^2|^2 (h+v)^2 \\ &+ \frac{1}{8}|g_1 B_\mu - g_2 W_\mu^3|^2 (h+v)^2 \\ &+ 0|g_1 B_\mu + g_2 W_\mu^3|^2 (h+v)^2 \end{aligned} \quad (1.9)$$

which, upon identification of $\frac{W_\mu^1 \mp iW_\mu^2}{\sqrt{2}}$, $\frac{g_1 B_\mu - g_2 W_\mu^3}{\sqrt{g_1^2 + g_2^2}}$ and $\frac{g_1 B_\mu + g_2 W_\mu^3}{\sqrt{g_1^2 + g_2^2}}$ as $W_\mu^\pm$, Z_μ^0 and

ⁱalso known as the condensate

ⁱⁱ $\mathcal{L}_{Klein-Gordon} = \frac{1}{2}\partial_\mu \phi \partial^\mu \phi - \frac{1}{2}m_\phi^2 \phi^2$ where ϕ is a scalar field and m_ϕ is the mass of the boson associated with that field.

A_μ (the photon) respectively, become:

$$\begin{aligned} \frac{1}{2}\partial^\mu h \partial_\mu h &+ \frac{g_2^2}{4} W_\mu^+ W^{-\mu} (h^2 + 2hv + v^2) \\ &+ \frac{1}{8} (g_1^2 + g_2^2) Z_\mu^0 Z^{0\mu} (h^2 + 2hv + v^2) \\ &+ 0 (g_1^2 + g_2^2) A_\mu A^\mu (h^2 + 2hv + v^2) \end{aligned} \quad (1.10)$$

Further recognizing that the mass of a charged boson is given by a term of the form, $mass_{F^\pm}^2 F^+ F^-$ and for an electrically neutral boson, $\frac{1}{2} mass_{F^0}^2 F^0 F^0$, the mass of the bosons may be expressed as:

$$mass_{W^\pm} = \frac{g_2 v}{2} \quad (1.11)$$

$$mass_{Z^0} = \frac{v}{2} \sqrt{g_1^2 + g_2^2} \quad (1.12)$$

$$mass_{A^0} = 0 \quad (1.13)$$

and by comparison with equation 1.8 it is possible to redefine

$$\frac{g_2^2}{4} = 2\lambda \left(\frac{mass_{W^\pm}}{mass_h} \right)^2 \quad (1.14)$$

$$\frac{g_1^2 + g_2^2}{8} = \lambda \left(\frac{mass_{Z^0}}{mass_h} \right)^2 \quad (1.15)$$

implying that the coupling strength of the gauge bosons to the Higgs boson is dependent on the ratio of the mass of the gauge boson to the mass of the Higgs boson. The vertices predicted by the theory are indicated in figure 1.4.

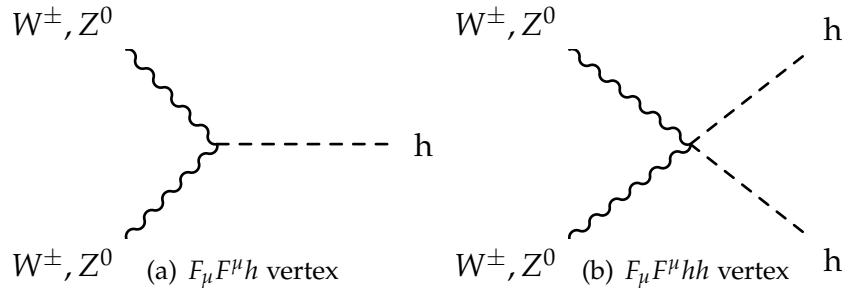


Figure 1.4: Feynman diagrams of the $F_\mu F^\mu h$ and $F_\mu F^\mu hh$ vertices introduced by the necessary use of the gauge covariant derivative, rather than the standard derivative, acting upon the Higgs field.

Prior to the discovery of the $W^\pm$ and Z^0 bosons, comparison of the electroweak theory to the Fermi theory of weak interactions allowed the vacuum expectation

value, v , to be expressed as $v^2 = 2^{-\frac{1}{2}} G_F^{-1}$, where G_F is the Fermi constant, and thus to be shown to have a value of approximately 246 GeV. This value, along with the value for the Weinberg mixing angleⁱⁱⁱ and the electromagnetic coupling constant, gave the predicted masses of the $W^\pm$ and Z^0 bosons to be 80 and 90 GeV respectively [11]; in excellent agreement with the current best-fit values of 80.398 ± 0.025 GeV and 91.1876 ± 0.0021 GeV [12]. These predictions were strongly supportive of the standard model and a similar prediction for the mass of the Higgs boson would be highly desirable. As it is not possible to separate the terms λ and $mass_h$, it is impossible to predict either value individually.

1.1.1.4 $\mathcal{L}_{Yukawa}$ Coupling

$$\begin{aligned} \mathcal{L}_{Yukawa \text{ Coupling}} = & -G_i^d \{h+v\}_2 (\bar{Q}_i d_i + Q_i \bar{d}_i) \\ & -G_i^u \{h+v\}_2^\dagger (\bar{Q}_i u_i + Q_i \bar{u}_i) \\ & -G_i^e \{h+v\}_2 (\bar{L}_i e_i + L_i \bar{e}_i) \end{aligned} \quad (1.16)$$

where $\{h+v\}_2$ again represents the $SU(2)_{isospin}$ doublet form of the Higgs field and the vacuum expectation value. Q and L respectively represent the left-handed Quark- and Lepton- doublet, u , d and e represent the right-handed u -type quark, d -type quark and charged-lepton singlets and the index, i , runs from 1 to 3 representing three generations of particles. The term, $\{h+v\}_2^\dagger$, is used to indicate the conjugate Higgs field, defined to be $-i\sigma_2 \{h+v\}_2^*$, so as to include interactions with the upper half of the doublet. The terms, G_i , are the coupling constants.

Direct comparison with the Dirac Lagrangian^{iv} yields the identity:

$$mass_x = \frac{G^x v}{\sqrt{2}} \quad (1.17)$$

ⁱⁱⁱThe Weinberg mixing angle, θ_W , is defined by $\cos(\theta_W) = \frac{g_2}{\sqrt{g_1^2 + g_2^2}} = \frac{mass_{W^\pm}}{mass_{Z^0}}$ or $\tan(\theta_W) = \frac{g_2}{g_1}$

^{iv} $\mathcal{L}_{Dirac} = \frac{i}{2} \{ \bar{\psi} \gamma^\mu \partial_\mu \psi - (\partial_\mu \bar{\psi}) \gamma^\mu \psi \} - m_\psi \bar{\psi} \psi$ where ψ is a spinor field and m_ψ is the mass of the fermion associated with that field.

or by comparison with equation 1.8:

$$G^x = 2\sqrt{\lambda} \frac{mass_x}{mass_h} \quad (1.18)$$

Noting also that the strength of the interaction between the Higgs boson and the spinor fields (that is, terms of the form $hX_i x_i$) is dependent only on the coupling constant, G_i^x , implies that the strength of the interaction is linearly dependent on the ratio of the fermion mass to the Higgs mass, and as such the Higgs Boson interacts preferentially with heavier particles.

The addition of a neutrino mass term depends on the nature of neutrinos themselves. If neutrinos are Dirac particles, then their mass may be taken into account by the addition of a term of the form $-G_i^n \{h+v\}_2^\dagger (\bar{L}_i n_i + L_i \bar{n}_i)$, where n indicates the right-handed neutral-lepton singlet. If, on the other hand, neutrinos are Majorana particles, their mass may be included by the addition of a term of the form $-\frac{m_i}{2} \hat{n}_i n_i$, where the term $\hat{n}_i$ is used to indicate the conjugate neutrino field, $-i\sigma_2 n_i^*$ and m must be generated by either a Higgs triplet or an effective operator involving two Higgs doublets arranged to transform as a triplet. As Majorana type mass terms couple left-handed states to left-handed states and similarly right-handed states to right-handed states, the masses of left-handed and right-handed Majorana neutrinos are decoupled and may possess different values. Furthermore, a Majorana neutrino may also interact with the Higgs Boson through the Yukawa mechanism, thereby also introducing a Dirac mass term for the neutrinos. If this is the case then a value of zero for the left-handed Majorana mass, a value at the electro-weak scale for the Dirac mass and a value at the Planck scale for the right-handed Majorana mass results “naturally” in an effective neutrino mass which is very much smaller than the electro-weak scale, as observed. Such a mechanism is known as a type-I see-saw mechanism. As yet, the true nature of neutrinos is unknown; namely, whether they are Dirac or Majorana particles, whether left- and right-handed neutrinos have different masses and whether the see-saw mechanism is an accurate description of neutrino mass generation.

As previously stated, the three generations of particles are now known not to be independent, but to interact via the CKM and PMNS mechanisms. This requires modifying terms of the form, $G_i^x \{h+v\}_2 (\bar{X}_i x_i + X_i \bar{x}_i)$ to $G_{ij}^x \{h+v\}_2 (\bar{X}_i x_j + X_i \bar{x}_j)$, where G_{ij} indicates the inclusion of the relevant mixing matrix element into the coupling constant.

1.1.2 Searches for the Higgs Boson

Given that the Higgs mechanism is currently the simplest known method of electroweak symmetry breaking and given that the mechanism requires the existence of at least one (as yet unobserved) massive scalar boson, most particle physicists are focused on the search for the Higgs boson or one of its supersymmetric equivalents [13, 14, 15, 16, 17, 18].

Although the existence of the Higgs boson has still to be demonstrated, should it conform to the standard model description, its properties and couplings are well understood; namely, it is massive, spin-0, uncharged and uncoloured. Although the mass of the Higgs boson is not directly predicted by the standard model, it is theoretically constrained to be less than the order of 1 TeV (figure 1.5) by the requirement that the scalar field theory is noninteracting^v, the so-called unitarity bound. Exclusion of the Higgs boson with a mass of less than 1 TeV will directly imply that the standard model is fundamentally flawed. The mass is also constrained by the fact that the Yukawa couplings act to oppose the vacuum potential: if the Yukawa couplings are too strong or, equivalently, the Higgs mass is too low, the potential is modified such that the vacuum expectation value, v , vanishes. This is forbidden as the main purpose for the inclusion of a Higgs field in the standard model is the generation of a non-zero vacuum expectation value. As such, the mass of the Higgs boson is constrained from below; the so-called vacuum-stability bound.

Conversely, the relationships indicated in figure 1.5 may be inverted such that, should the Higgs boson be found, its mass may be used to provide insight as to

^vequivalently described as “trivial”

the energy scale, Λ , at which “new physics” is required by the standard model. For a Higgs mass of between 130 GeV and 180 GeV, this energy scale is the Planck energy (10^{19} GeV). For masses outside this range, “new physics” is required at energies lower than the Planck scale.

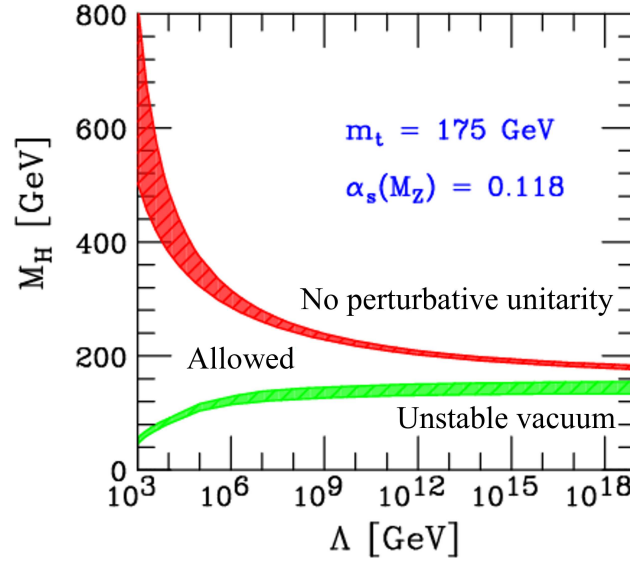
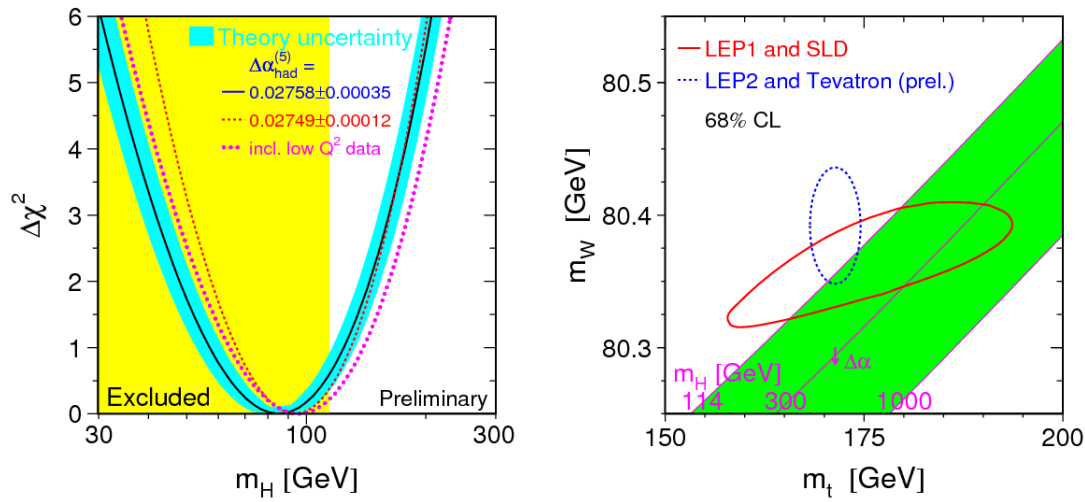


Figure 1.5: Bounds on the mass of the Higgs boson following from the requirement that the electroweak theory is consistent up to the energy scale Λ , where Λ is the energy scale at which “new physics” appears.

Although, as stated previously, the mass of the Higgs boson is not directly predicted by the standard model, the probability of any particular mass can be inferred using measurements of the radiative corrections to the $W^\pm$ and Z^0 bosons in the channels $W^+ \rightarrow \bar{b}t \rightarrow W^+$, $W^- \rightarrow \bar{t}b \rightarrow W^-$, $Z \rightarrow \bar{t}t \rightarrow Z$, $Z \rightarrow HZ \rightarrow Z$ and $W^\pm \rightarrow HW^\pm \rightarrow W^\pm$. Hence, accurate measurement of the $W^\pm$, Z^0 , t and b masses made at the LEP [19] and Tevatron [20] facilities can be used to infer the most likely Higgs mass [21]. Direct measurements at LEP [22] exclude the possibility of a Standard Model Higgs with a mass less than 114.1 GeV. Measurements at Tevatron [23] exclude at 95% confidence level, the possibility of a Standard Model Higgs with a mass of 170 GeV.

Figures 1.6(a) and 1.6(b) indicate the current status of Higgs searches. The yellow region in 1.6(a) indicates the 95% confidence level exclusion as demonstrated by LEP. These results also exclude at 95% confidence the probability of a Higgs



(a) χ^2 plot of m_H using the Tevatron and the results of direct LEP searches. (b) 68% confidence level region given the current best measurements of m_W and m_t .

Figure 1.6: Latest results of searches for the Higgs boson. Taken from [21].

mass greater than 185 GeV. Not included in these plots is the latest exclusion at 170 GeV.

1.1.3 Objections to the standard model

The significance of the result by 't Hooft [9] is that, for gauge invariant theories, higher order divergent diagrams such as those in figures 1.7(a), 1.7(b) and 1.7(c) can be 'tamed' by the processes of regularization and renormalization. The processes shown in figures 1.8(a)-1.8(e), however, diverge as $\frac{s}{M_W^2}$ as $s \rightarrow \infty$ without the systematic cancellations introduced by adding scalar diagrams of the type in figures 1.8(f) and 1.8(g). The standard model Higgs boson is a suitable candidate for such a scalar particle and, as such, provides an elegant solution to the so-called "unitarity problem" [24]. If the standard model Higgs is demonstrated not to exist, detailed studies of WW scattering in the energy range around M_W^2 will provide alternative insight into the true behaviour of nature.

The cancelling of W scattering divergences by the standard model Higgs boson raises extra problems; the first is that the cancellation is 'conspiratorial', requiring that the various couplings are finely tuned. The second is that direct calculations of the standard model Higgs mass suffer their own divergences. One

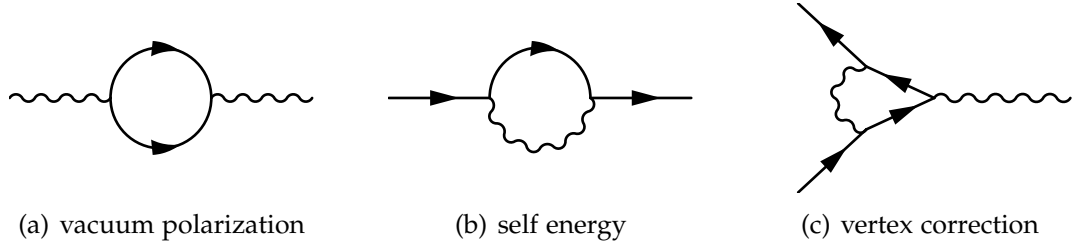


Figure 1.7: Feynman diagrams for the three sources of ultraviolet divergences in the standard model.

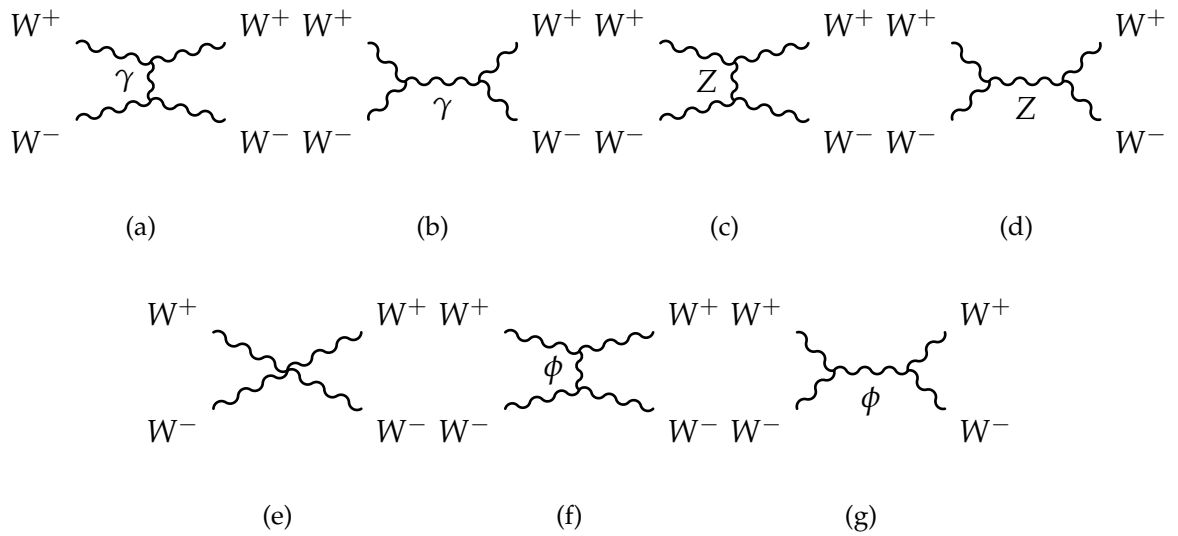


Figure 1.8: Feynman diagrams for the process $WW \rightarrow X \rightarrow WW$. Here, ϕ represents an unknown scalar propagator.

possible solution to this problem is the introduction of a **SuSy** (SuperSymmetric) partner for every standard model particle; a fermionic partner for each standard model boson and a bosonic partner for each standard model fermion. The high degree of symmetry allows for the systematic cancellation of all divergences at the expense of adding a plethora of, as yet, unseen particles: 1 photino, 2 winos, 1 zino, 8 gluinos, 6 sleptons, 6 squarks and at least 5 Higgs bosons and their fermionic partners. Extensions to the minimal model also include the graviton and gravitino. To date, no evidence of supersymmetry has been found [12].

A further objection to the standard model is the large number of parameters which cannot be determined from theory, but must be constrained by observation, namely the 3 fundamental constants (c , G and $\hbar$) and 19 parameters (6 quark masses, 3 CKM angles, 1 CKM phase, 3 charged lepton masses, 3 gauge coupling strengths, the **QCD** (Quantum ChromoDynamics) vacuum angle, the Higgs vacuum expectation value and the Higgs coupling strength^{vi}). More recent observations have demonstrated the need for at least 7 further parameters: 3 neutral lepton masses, 3 PMNS angles and 1 PMNS phase. The introduction of supersymmetry worsens this situation by adding at least 105 further parameters. Furthermore, the weak-hypercharge, weak-isospin and colour-charge of each field is assumed to be a fundamental property of the field and are neither explained, nor their values predicted.

By excluding a large section of the available SuSy parameter space, direct searches up to the TeV range will provide the strongest possible evidence as to whether supersymmetry does, in fact, apply to the universe that we observe. Even without direct production (or lack thereof) of supersymmetric particles, the exclusion of a Higgs boson with mass less than 1 TeV automatically excludes both the standard model and all supersymmetric models, as all rely on the Higgs mechanism for the generation of mass.

Although the CP-violating QCD Lagrangian term of the form $\frac{\theta_{QCD}}{64\pi^2}\epsilon^{\mu\nu\rho\sigma}G_{\mu\nu}G_{\rho\sigma}$ is gauge invariant and hence not excluded from the standard model Lagrangian,

^{vi}or alternatively, as stated previously, the Higgs mass

current limits on the QCD vacuum angle, θ_{QCD} , place an upper bound of 10^{-10} on its value. As θ_{QCD} is treated as a parameter, the standard model offers no explanation as to why this type of interaction is so much weaker than all other interactions or is possibly excluded altogether. At the same time, the standard model offers no explanation for the source of the CP-asymmetry required for baryogenesis.

Although not an insuperable objection, much of the search for a good theory of particle interactions has also been a search for an aesthetically pleasing theory: a theory that “feels” right^{vii}. Generally, in physics, beauty is synonymous with simplicity of form and with the fact that key features should not happen by chance, but should have some underlying cause. In this sense, the standard model is blemished; in particular —

- The $SU(3)_{colour} \times SU(2)_{isospin} \times U(1)_{hypercharge}$ gauge group structure is included by hand to explain observations; there is no deeper principle behind this choice or explanation as to why higher groups are excluded.
- The formation of left-handed $SU(2)_{isospin}$ doublets and right-handed singlets is, again, unexplained. Left-right symmetric models are generally regarded as more satisfactory.
- The number of generations, $N_g = 3$, is unexplained.
- The quantization of electric charge, $Q_d = -\frac{1}{2}Q_u = \frac{1}{3}Q_e$, is unexplained.

The standard model both lacks a mechanism for the inclusion of gravity, and worse, produces results directly at odds with general relativity: the vacuum energy density of the Higgs field (10^8 GeV^4) is vastly different from the cosmological constant (10^{-46} GeV^4) predicted by the latter theory [25]. Furthermore, cosmological results indicate the presence of dark matter [26] and dark energy [27], for which the standard model provides no candidates.

^{vii}Such an attitude is sometimes referred to as Paul Dirac’s religion of mathematical beauty

Finally, the standard model fails to explain the discrepancy between the observed mass scale (1 eVc^{-2} to 170 GeVc^{-2}) and the natural (Planck) mass scale — $10^{19} \text{ GeVc}^{-2}$; the so-called “Hierarchy problem”.

1.2 The Large Hadron Collider (LHC)

The **LHC** (Large Hadron Collider) [28] is a proton-proton collider based at CERN, operating at a maximum centre-of-mass energy of 14 TeV, a collision rate of 40 MHz and a nominal luminosity of $10^{34} \text{ cm}^{-2}\text{s}^{-1}$. The principle aims of the LHC are (1) to elucidate the nature of electroweak symmetry breaking (for which the Higgs mechanism is assumed to be responsible) and (2) to search for physics beyond the standard model [29]. As a hadron collider, the LHC cannot be used to make precision measurements allied to those made at LEP, but rather provides both a window to access phenomena at the energy frontier and those with ultra-low cross sections. In order to do so, the LHC is designed to operate at unprecedented energies and to produce large numbers of inelastic collisions during every **BX** (Bunch Crossing). A second mode of operation allows the LHC to collide lead ions at a centre of mass energy of 5.5 TeV per nucleon, with the express purpose of studying quark-gluon plasmas.

Due to the composite nature of the proton, the available energy in each collision is only a fraction of the 14 TeV centre of mass energy; the LHC is expected to permit discovery (5σ signal) of new phenomena up to an energy of 1 TeV, with access to physics above this energy being limited by statistics. A lack of statistics does not exclude discovery above 1 TeV with some signals being “unmissable” [30, 31].

The LHC hosts 6 experiments: the two general purpose detectors, **CMS** (Compact Muon Solenoid) and **ATLAS** (A Toroidal LHC ApparatuS), are designed to verify the standard model and to search for a wide range of new physics phenomena; the **ALICE** (An LHC Ion Collider Experiment/A Large Ion Collider Experiment) [32] experiment is designed for heavy ion studies; and LHCb

for studying CP-violation in the b-quark sector. Two further small experiments, **Totem** (Total Cross Section, Elastic Scattering and Diffraction Dissociation Measurement) and LHCf study diffractive physics in the extreme-forward regions on either side of the CMS and Atlas experiments.

Although the general purpose detectors were designed to look for any signs of ‘new’ physics, the most keenly anticipated is the discovery of the standard model Higgs boson. The five dominant production channels for the standard model Higgs at LHC are gluon-gluon fusion (via a heavy quark loop), vector-boson fusion, vector-boson bremsstrahlung^{viii}, $t\bar{t}$ fusion and $b\bar{b}$ fusion. The dominance of these modes is due to the fact that the Yukawa and Higgs couplings in the Standard Model are proportional to a particle’s mass; this fact similarly controls the ratio of the various decay modes. Figures 1.9 and 1.10 demonstrate the cross-sections and branching ratios of each mode.

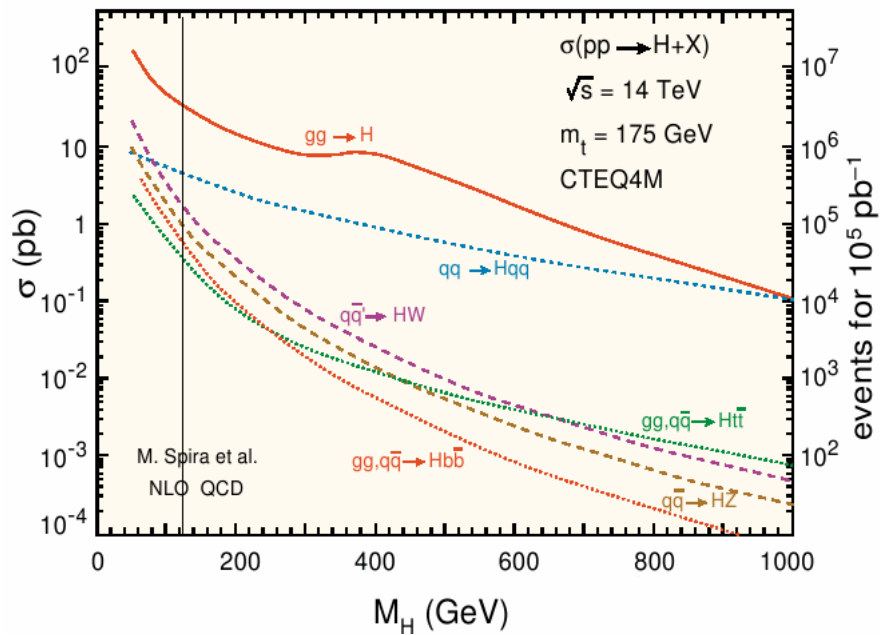


Figure 1.9: Higgs production channels at the LHC. The channel labelled $gg \rightarrow H$ represents gluon-gluon fusion via heavy quark loop, $qq \rightarrow Hqq$ represents vector-boson fusion, $qq' \rightarrow HW^\pm$ and $q\bar{q} \rightarrow HZ^0$ represents vector-boson bremsstrahlung (higgsstrahlung). $gg, q\bar{q}' \rightarrow Ht\bar{t}$ and $gg, q\bar{q}' \rightarrow Hb\bar{b}$ respectively represents $t\bar{t}$ and $b\bar{b}$ fusion. The region to the left of the vertical line at $M_H = 120 \text{ GeV}$ is excluded by LEP.

^{viii}also known as Higgsstrahlung

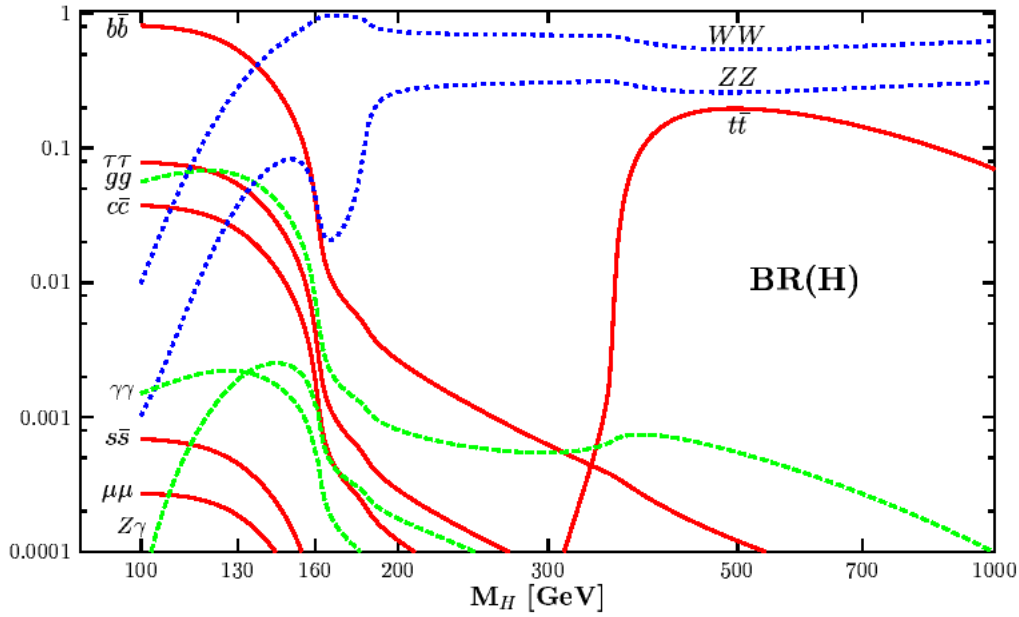


Figure 1.10: Higgs decay modes in the mass range accessible at the LHC. The decay of the Higgs boson to massless particles is necessarily mediated by heavy fermion loops.

The production cross section of the Higgs boson cannot be taken in isolation, but must be taken in relation to that of all possible processes. As can be seen in figure 1.11, the cross section (and hence production probability) of Higgs processes are between 9 and 11 orders of magnitude (dependant on the mass of the Higgs boson) lower than the total inelastic cross section and some 3 to 5 orders of magnitude lower than $W^\pm$ and Z^0 production cross section, with the general effect that the signal is “swamped”, not only by the dominant QCD processes but also by the electroweak signals which LEP was built to study.

In order to accommodate such low production probabilities, it is necessary to have a very high interaction rate, this requiring a very high beam luminosity. At LHC energy scales the total proton-proton cross-section is estimated to be 100 mb, of which 30 mb is expected to be elastic and therefore not relevant to the experiments. At the design luminosity of $10^{34} \text{ cm}^{-2} \text{ s}^{-1}$, the 70 mb inelastic scattering cross-section equates to an event rate of $7 \times 10^8 \text{ s}^{-1}$. The beam is not continuous, but bunched with a crossing rate of 40 MHz, and only around 80% of bunches are filled, this resulting in an average event rate of 22 events per crossing.

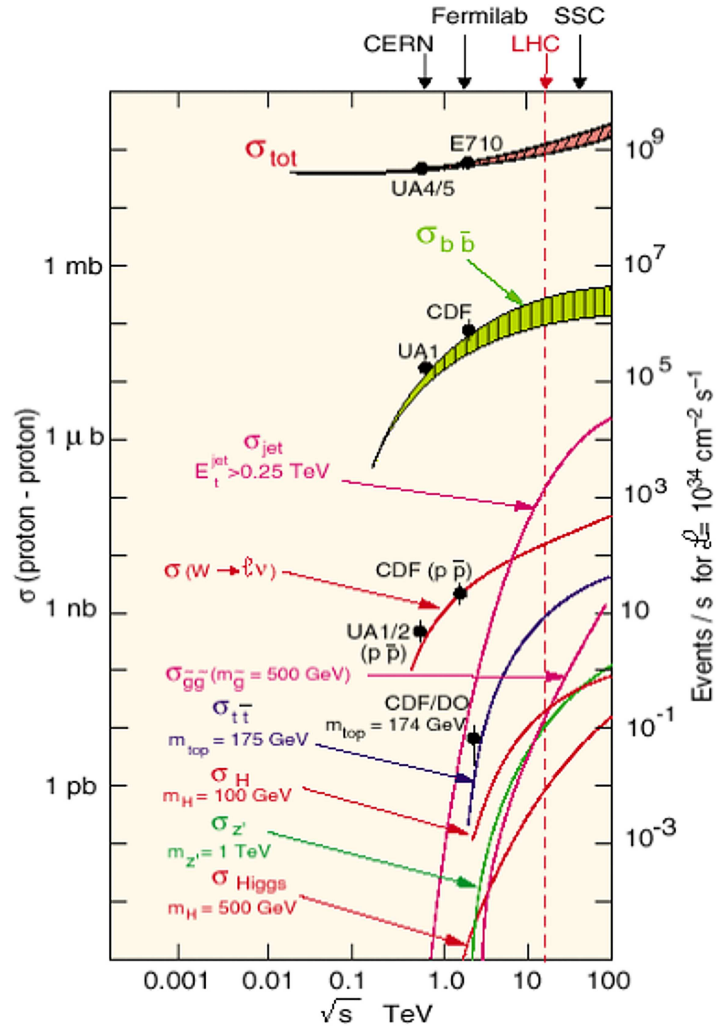


Figure 1.11: The production cross-sections of a selection of possible production modes at the LHC as a function of interaction energy. Adapted from [33].

The number of charged tracks per crossing at the LHC is in the order of one-thousand [34] and the event data size around 1 MByte per crossing [35]. As such the experiments face many challenges not faced at previous experiments.

The high crossing rate (40 MHz) places strict demands on the speed at which the sensors must collect data and the speed at which the data must be read out from the detectors, this limiting the choices of available technologies. The large number of interactions per crossing requires precision tracking in order to separate vertices and to locate secondary vertices. Both the high crossing rate and high interaction rate result in “pile-up” effects in two forms: “in-time” pile-up, which is the result of uninteresting (at least for the LHC) QCD events (termed minimum-

bias), and “out-of-time” pile-up, where particles from previous crossings remain in the detector, looping in the magnetic field. The large number of particles also result in a massive radiation dose exceeding 30 MRad (300 kGy) and 10^{15} cm^{-2} neutron equivalent dose in the forward regions of the detectors over the expected lifetime of the tracker [36]. This not only requires radiation tolerance of the active components, both in terms of resilience to long-term damage and tolerance to **SEUs** (Single Event Upsets)^{ix}, but also careful choice of materials to minimize material radioactivation over the lifetime of the experiments [39].

Not only must the detectors withstand the environment, but they require excellent resolution in order to separate signal from background: for instance, the cleanest decay signal for a low mass Higgs boson is the decay $H \rightarrow \gamma\gamma$, but even this requires excellent calorimetry in order to identify the signal over the background (see figure 1.12).

The other ‘golden’ decays, $H \rightarrow \tau\tau$, $H \rightarrow ZZ \rightarrow l^+ l^- l^+ l^-$ and $H \rightarrow ZZ \rightarrow jj l^+ l^-$ (where j represents a jet), similarly require both excellent calorimetry and particle tracking to separate them from the background.

1.3 The Compact Muon Solenoid (CMS)

CMS [41] is the smaller of the two general-purpose detectors at the LHC. It is a hermetic detector covering the rapidity range from $-3 < \eta < +3$ at full detector resolution and having an extended rapidity range of $-5 < \eta < +5$. Moving radially from the beampipe, the detector consists of a pixel tracker, a silicon microstrip tracker, a homogeneous lead tungstate electromagnetic calorimeter, a brass-plastic sampling hadronic calorimeter, a four Tesla superconducting magnet, a second hadronic calorimeter^x and an iron flux return yoke interspersed

^{ix}An SEU occurs when a charged particle passes through an electronic component, changing the state of a bit in a digital logic circuit or in a memory element. This is typically mitigated by using triple-redundant logic where each circuit element is triplicated and the state determined by a majority rule. [37, 38]

^xThe second hadronic calorimeter — called the “outer HCAL” or simply “HO” — is small, but provides a means of identifying hadronic punch-through. Because of its limited size and function, it will not be discussed further.

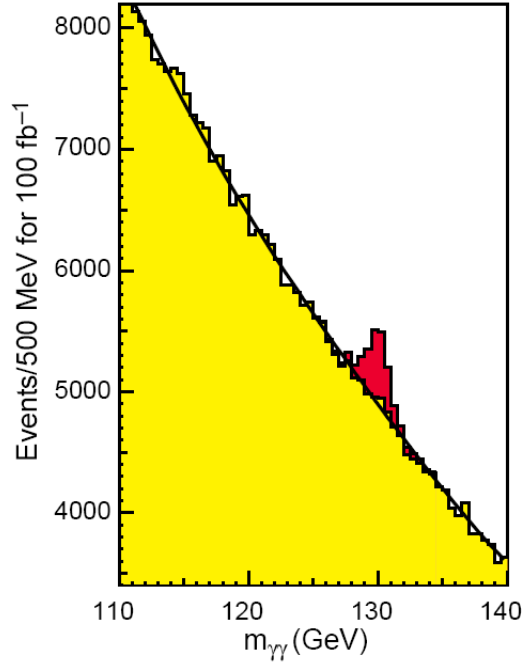


Figure 1.12: A comparison of $H \rightarrow \gamma\gamma$ signal (red) and background (yellow) for a Higgs mass of 130 GeV and 100 fb^{-1} of data recorded. Adapted from [40].

with muon detectors. The detector is shown in figure 1.13.

1.3.1 The Silicon Tracker

The CMS tracker [42, 43] (figure 1.14) consists of three layers of silicon pixel sensors and ten layers of silicon microstrip sensors in the barrel, and twelve layers in either endcap. The inner layer of the pixel detector is at a 4 cm radius from the nominal interaction point and, as such, has to cope with high particle fluences; a $3.2 \times 10^{15} \text{ cm}^{-2}$ neutron equivalent dose is expected over the lifetime of the experiment. It must be situated close to the interaction region in order to identify the (approximately 20) separate vertices, tag ‘long-lived’ particles such as b and c quarks and τ leptons, and to identify light quark and gluon jets. The microstrip tracker provides many widely spaced points for precision momentum and vertex measurements [44]. The entire silicon tracker is 5.4 m long, has a diameter of 2.4 m and is operated at -10°C to minimize the effects

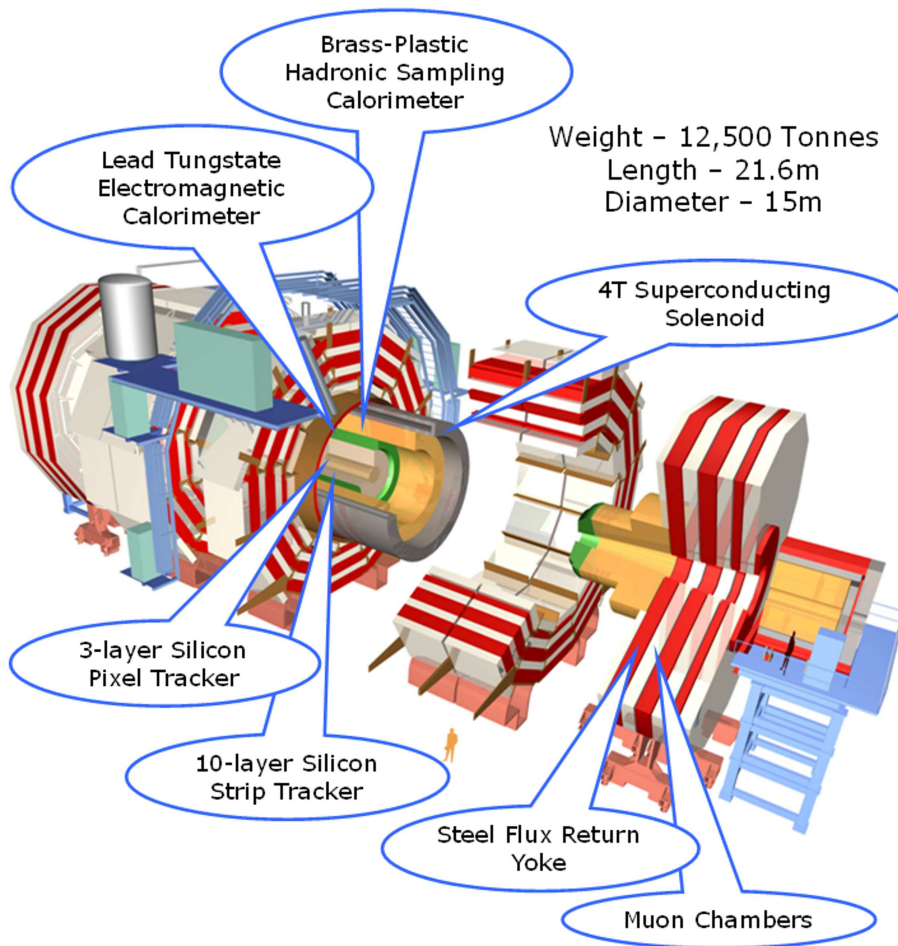


Figure 1.13: Diagram of the CMS Detector. The figure includes two “people” to provide an indication of scale.

of radiation damage. The pixel tracker has ~ 45 million channels and the strip tracker a further 10 million.

The CMS pixel detector is a hybrid design consisting of a silicon sensor with rectangular pixels of pitch $100\ \mu\text{m}$ ($r\phi$) by $150\ \mu\text{m}$ (z) and a separate readout chip. The sensors require a large external bias voltage (up to 600 volts), a large depletion region and fast charge collection to ensure a useable signal after heavy irradiation and in order to be read out within a single bunch-crossing. Each pixel contains a set of n-on-n diodes and a contact pad for bump bonding it to the corresponding channel on the readout chip. The readout chip [45] provides an analogue readout of the pixels, performs a comparison to threshold on a pixel-by-pixel basis, stores the data in a pipelined buffer while waiting for a

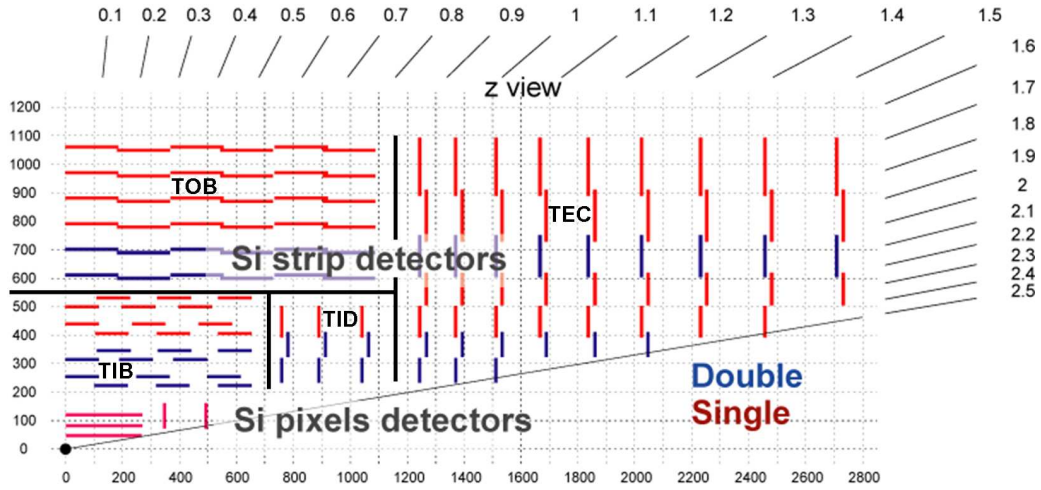


Figure 1.14: Layout of a quarter of the CMS tracking detector. This image is mirrored along both axes to make the full detector layout. The interaction point is marked at $z=0$.

level-1 trigger decision and provides a sparsified analogue readout via a high-speed token-ring. Because the readout is analogue, interpolation may be used to improve the spatial resolution to approximately $10 \mu\text{m}$ in $r\phi$ and $15\text{-}20 \mu\text{m}$ in z [46].

The CMS microstrip tracker is similarly a hybrid design with the strips being read out by the **APV25** (Analogue Pipeline Voltage, $0.25 \mu\text{m}$) [47] readout chip. The sensors consist of either 512 or 768 strips with pitch ranging from $80 \mu\text{m}$ to $205 \mu\text{m}$ and length ranging from 9 cm to 20 cm [48, 49].

In the readout chip, each channel is preamplified, shaped and sampled before being stored in a pipelined buffer whilst waiting for a level-1 trigger decision. The APV25 has two sampling modes; peak mode where the signal is sampled once, at the peak, and deconvolution mode where the signal is sampled before, at and after the peak, and the signal reconstructed using an Analogue Pulse Shape Processor. The mode is selected according to the beam luminosity with the former mode at low luminosity (less susceptible to noise) and the latter at high luminosity (less susceptible to pile-up). Each APV25 chip reads out 128 channels. For the microstrip tracker, zero-suppression is not performed on-detector.

Upon receipt of a level-1 trigger signal, the buffered analogue signals are time-division multiplexed (in order to reduce the cabling requirements) and trans-

mitted optically. This results in approximately 45,000 analogue optical readout fibres leaving the detector. The use of analogue signals was driven primarily by power considerations, namely that the power consumed by digitization increases with the speed of digitization and that within the tracker access for both power and cooling (to compensate for the increased heat dissipation) are severely restricted. There are however additional benefits;

- The ability to interpolate between channels thereby increasing the effective resolution
- A reduced susceptibility to noise by minimizing the number of manipulations made on-detector, mitigating the possibility of ‘bit-flipping’ upsets.
- The possibility of monitoring degradation of the sensors and readout chips by studying the signal quality.

Track reconstruction by combinatorial Kalman filtering gives a track reconstruction efficiency of approximately 100% for muons in the acceptance range, whilst electron/hadron reconstruction efficiencies range from 95% in the central region to 80% in the forward region. For $p_T=100\text{ GeV}$ muons in the central region, the p_T resolution is between 1% and 2% and improves at lower track momentum. The tracker can determine the impact parameter with a resolution of $20\text{ }\mu\text{m}$ for $p_T=100\text{ GeV}$ tracks, although this performance degrades at lower momentum due to multiple scattering [50].

1.3.2 The Electromagnetic Calorimeter (ECAL)

The **ECAL** (Electromagnetic Calorimeter) [40] is a homogeneous scintillating crystal calorimeter. The requirement that the ECAL and HCAL be located within the solenoid severely limits the size of the calorimeter. In order to achieve high containment in the limited space available, lead tungstate was chosen due to its small Molière radius (2.2 cm) and short radiation length (0.89 cm) [46]. Additionally, lead tungstate has a high radiation tolerance and a rapid light collection

time ($> 80\%$ in one bunch-crossing). It does, however, have a relatively low light yield and, as such, the light signal must be amplified. This is achieved using **APD** (Avalanche PhotoDiodes) in the barrel and radiation hard **VPTs** (Vacuum PhotoTriodes) in the endcaps, where the radiation dose is greater and the magnetic field is close to axial. The barrel ECAL uses two APDs per crystal for the purpose of noise mitigation. Each crystal has a front face of $2.2\text{ cm} \times 2.2\text{ cm}$ in the barrel section (approximately 0.0175×0.0175 in $\eta \times \phi$) and a depth of 230 mm. The energy resolution is $\frac{\sigma_E}{E} = \frac{0.125}{E} \oplus \frac{2.9\%}{\sqrt{E}} \oplus 0.3\%$, better than the design value of 0.5% for electrons and photons at 100 GeV.

The ECAL signals are digitized on-detector and trigger primitives generated for each trigger tower (5×5 crystals in the barrel and approximations to η - ϕ segments in the endcaps). The trigger primitives are sent off-detector to the calorimeter trigger (see chapter 2), while the raw data is stored in a pipelined buffer pending a level-1 trigger decision.

1.3.3 The Hadronic Calorimeter (HCAL)

Heavier particles and their decay products are not contained in the ECAL and penetrate through to the **HCAL** (Hadronic Calorimeter) [51]. In the barrel this is a sampling calorimeter with a brass absorber and plastic scintillator which is read out by **HPD** (Hybrid PhotoDiode) sensors coupled via wavelength-shifting fibres. In the forward endcap region, quartz Cerenkov fibres, rather than plastic scintillator, are embedded into steel absorber due to their greater radiation tolerance. The energy resolution is measured to be $\frac{\sigma_E}{E} = \frac{75\%}{\sqrt{E}} \oplus 8\%$, compared to the design value of $\frac{\sigma_E}{E} = \frac{100\%}{\sqrt{E}} \oplus 4.5\%$; the poor energy resolution^{xi} is currently the limiting factor in CMS' calorimetry and is closely related to the lack of longitudinal shower information from the HCAL towers and noise in the readout electronics. Upgrades to both are currently under investigation.

^{xi}For comparison, the ATLAS Copper-Liquid Argon HCAL has a resolution of $\frac{\sigma_E}{E} = \frac{45\%}{\sqrt{E}} \oplus 1.3\%$.

1.3.4 The Muon Detectors

The CMS detector intersperses the iron magnetic-flux return circuit with three types of muon stations. In the barrel, **DTs** (Drift Tubes) provide precision tracking while **RPCs** (Resistive Plate Chambers) provide charge collection sufficiently quickly to create tracks for the purpose of triggering at level-1. In the endcaps **CSCs** (Cathode Strip Chambers) provide the precision tracking while, again, RPCs provide triggering information. The spatial resolution of the muon detectors ranges from 50 to 200 μm , translating to a momentum resolution of 15% for a muon with 10 GeV p_T and 40% at 1 TeV when taken independently of the silicon tracker [52, 53].

1.3.5 The CMS Trigger System

As stated in section 1.2, the LHC crossing rate is 40 MHz and the CMS event size is 1 MByte per crossing, resulting in a raw data rate of 40 TByte per second. As such, there is neither the readout bandwidth nor the storage capability to store all of the data produced. The peak storage rate for CMS is approximately 10 TByte per day (100 Hz) and, therefore, the data volume must be reduced by a factor of 4×10^5 before anything can be written to disk. This data reduction is achieved by triggering; the selective storage of events based on a coarse-grained analysis of the data.

The CMS trigger [54, 55, 56] is achieved in two stages; the level-1 trigger and the **HLT** (Higher Level Trigger). The level-1 trigger [57] is a high bandwidth, fixed latency system based on **FPGAs** (Field Programmable Gate Arrays) and **ASICs** (Application Specific Integrated Circuits). The enormous volume of data from the detector and the requirement of a trigger decision in a very short (and guaranteed) time period of 128 BX excludes the use of iterative algorithms, requiring instead pipelined algorithms and massive parallelization. The level-1 trigger uses only calorimetric and muon data, as the bandwidth requirement of the tracker is too large to allow read out of every bunch-crossing; instead, it is read

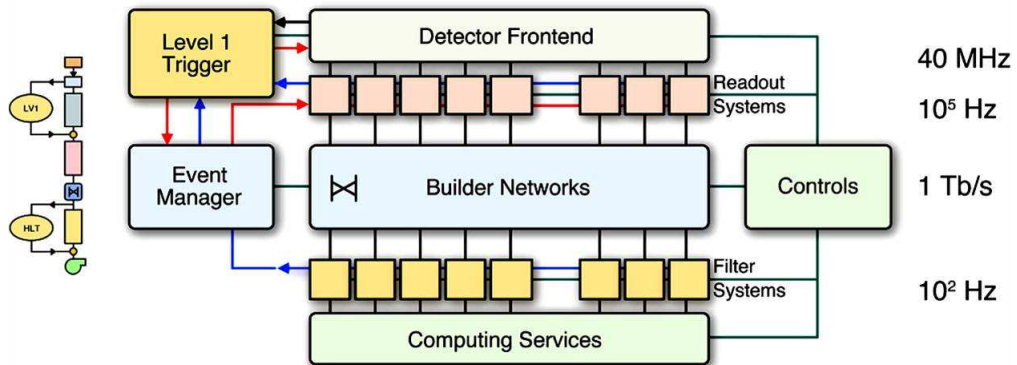


Figure 1.15: The CMS Trigger/DAQ System. Data from a subset of the detector is first sent to the level-1 trigger for processing. Upon a level-1 accept signal, the remaining detector data is read out to the Higher Level Trigger for further processing.

out only once the level-1 accept signal is issued. The aim of the level-1 trigger is to reduce the data rate by an average factor of 400, from a rate of 40 MHz to 100 kHz. The Higher Level Trigger is a multi-stage iterative algorithm conducted on a farm of 3000 PCs: it uses full detector information to reproduce the level-1 trigger decision and then to iteratively improve on this decision by the staged introduction of fine-grained calorimetry information and tracking information. If, at any stage, the event is rejected, the process is halted and the processing resources freed for handling the next event. A time cut-off (the value of which depends on the process) is applied to prevent resources becoming locked. The aim of the HLT is to reduce the data rate by a further factor of 1,000 so that the data can then be recorded to disk and tape storage at approximately 100 Hz.

The purpose of the trigger system, particularly of the level-1 trigger, is not (as is often described) to identify interesting events from the background of uninteresting events. Such a definition would imply that it is already known which interesting events exist to be found. Instead, the trigger is designed to discard events that lie within our current understanding, while simultaneously having a threshold for acceptance that is sufficiently low to allow all possible signatures of new physics to be recorded. In CMS this is achieved by identifying the type of decay products and the transverse energy (particles from heavy states are emitted isotropically, while less interesting events are focussed in the forward regions) of those products; by this means, as few constraints as possible

are placed on the discovery potential. This type of triggering is described as ‘inclusive triggering’.

1.3.6 The CMS Level-1 Trigger

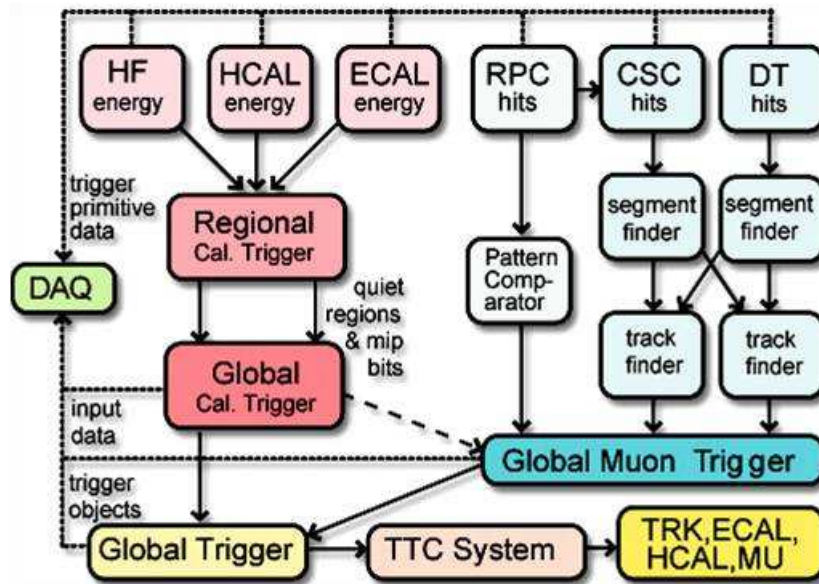


Figure 1.16: The CMS level-1 trigger.

As stated previously, the level-1 trigger only uses data from the electromagnetic and hadronic calorimeters and tracks from the muon systems. Most of the trigger infrastructure is located in the **USC** (Underground Sorting Cavern) to minimize the cable latency: two exceptions to this are the ECAL trigger tower primitives, which are generated on the ECAL front-end electronics, and the muon local track finder which is performed on the racks on the outside of the detector.

The muon trigger creates tracks from hits, sorts all the tracks in an event by their transverse momentum and forwards the top four candidates to the global trigger.

The calorimeter trigger creates two different types of trigger object, reflecting the two fundamental types of energy deposition in the calorimeter; electron/photon candidates and jets. The former are spatially-compact objects mainly confined to the ECAL, while the latter are larger and primarily in the HCAL. The objects are

classified by various criteria and sorted, with the four highest ranked candidates of each type being forwarded to the global trigger. The calorimeter trigger is discussed in detail in chapter 2.

The global trigger combines candidates from the muon and calorimeter triggers and, based on these combinations, decides whether to discard an event or whether to read out the full granularity information and forward it to the Higher Level Trigger. The trigger can be based either on simple objects, such as a single particle with a transverse momentum greater than a certain threshold, composite objects, such as “two electrons $\vee$ two muons $\vee$ (electron + muon)”, or any other conceivable combination, although the object should be kept simple in order to avoid introducing bias.

Tracker information is not used in the level-1 trigger, because it was not believed to be necessary for triggering under LHC conditions. As such, the tracker was not designed to be capable of being read out for every event; the bandwidth requirement, the choice of analogue optical link technology [58, 59] and off-detector zero-suppression [60, 61] prohibiting this. Under conditions of higher luminosity this requirement changes and is discussed in chapter 6.

1.3.7 The CMS Level-1 Trigger Software

Given the many different types of hardware and the vast number of control and processing subsystems, both on- and off-detector, the system used to control the experiment requires a great deal of flexibility and the ability to be scaled as required, while retaining complete reliability. The size and international nature of the collaboration also requires that the user interface is both standardized and clear, as those operating the experiment cannot be expected to be experts in all subsystems. This is achieved using a framework called **XDAQ** (Distributed Data Acquisition) [62] which provides a platform-independent environment into which each subsystem dynamically loads its own software modules. The framework thus provides a standardized infrastructure for core tasks, such as

message passing and data transfer via **SOAP** (Simple Object Access Protocol) or **I²O** (Intelligent Input/Output) over TCP/IP, which minimizes the amount of code duplication and thereby reduces the probability of errors. The framework also provides a web-based interface through which the user interacts by means of a standard internet browser.

The modules used to control each subsystem are written as C++ classes based on standard XDAQ templates and are compiled as libraries that can be dynamically loaded at runtime. This allows individual subsystems to be started, stopped and recompiled without affecting the rest of the modules on that node.

Although the control system of the level-1 trigger is clearly very different from the control system of, for example, the silicon pixel tracker, within the level-1 trigger there are many tasks which are common to all subsystems. The Trigger Supervisor is a framework built upon XDAQ, specializing it for the task of controlling the subsystems of the CMS level-1 trigger and implementing the common functionality.

Chapter 2

The CMS Calorimeter Trigger System

“A man is too apt to forget that in this world he cannot have everything.
A choice is all that is left him.”

- Harry Mathews (1930 -)

The calorimeter trigger system lies in the trigger chain between the calorimeter front-end electronics and the level-1 global trigger system. The primary task of the calorimeter trigger is to convert the raw data signals into sorted lists of electron/photonⁱ, henceforth e/γ , and jet candidates. The main difficulty in performing this task is the raw data rate: the CMS calorimeter consists of 76,832 ECAL crystals and 15,096 HCAL channels all of which are digitized, read out every bunch-crossing and converted to, and summarized by, a small list of meaningful “physical” objects. This is achieved in two stages; first, by summing groups of 5×5 ECAL crystals into trigger-towers which have a one-to-one mapping to the HCAL towers and, secondly, summing these towers into regions whose size is closer to the natural scale of the trigger objects. These objects are classified according to geometric criteria, each type sorted by its rank (a measure

ⁱor more correctly electromagnetic-type candidates as, since the level-1 trigger does not include tracking information, electron and photon signatures (and in certain cases pion signatures) are not clearly distinguishable

of its transverse energy and ‘quality’) and the top four of each type forwarded to the global trigger for further processing. Furthermore, the calorimeter trigger is required to count the number of jet candidates and to calculate the total energy (and any energy imbalances) in the transverse plane and to send this data to the global trigger. The calorimeter trigger is also required to provide calorimetry information for the muon trigger and to monitor luminosity through studies of event rates. Finally, for every event accepted by the level-1 trigger, the calorimeter trigger must also read out not only the raw ECAL and HCAL data and the final trigger candidates but also all intermediary information generated from the trigger-tower level objects upward. The applied geometric criteria are that, e/γ candidates passing an isolation criteria are identified as isolated, whilst those that fail the criteria are identified as non-isolated. For jets, those passing a τ -veto are identified as τ -jets, those where the pseudorapidity is greater than a programmable threshold (currently $|\eta_{thresh.}| = 3$) are defined as forward-jets and those where the pseudorapidity is lower than the threshold are defined as central-jets. A further consideration is that all subsystems in the level-1 trigger must obey several strict design requirements. As stated previously, the raw data rate of 40 Tbyte/s precludes reading out the entire detector on every bunch-crossing. Instead, the full granularity data is stored in on-detector pipelined buffers of length 128 BX to be read out at a maximum rate of 100 kHz while coarse grained information is read out every bunch-crossing and used for triggering. If no trigger decision is received before the data reaches the end of the buffer then the data is irretrievably lost. As such 128 BX, or $3.2 \mu\text{s}$, represents a hard limit on the time available not only for all level-1 trigger computation, but also for cable delaysⁱⁱ. Furthermore, every bunch-crossing must be analysed and the analysis of one may not interfere with the analysis of the next; the system must be free of dead-time. This excludes the possibility of iterative algorithms and requires a pipelined processing architecture. It is also required that the trigger should have sufficient flexibility and programmability that triggering conditions may be modified should the need arise.

ⁱⁱThe underground counting room is 90 m from the experimental cavern, however, this distance must be counted twice, once for the data to be read out and once for the trigger decision to be sent to the detector. This corresponds to approximately 36 BX or $0.9 \mu\text{s}$

The complete list of objects required by the global trigger from the calorimeter trigger is:

- $4 \times$ isolated e/γ candidates
- $4 \times$ non-isolated e/γ candidates
- $4 \times$ forward-jets
- $4 \times$ central-jets
- $4 \times \tau$ -jets
- $12 \times$ jet counts
- Total transverse energy
- Total transverse energy in jets
- Missing transverse energy
- Missing transverse energy in jets
- Energy in forward ($|\eta| > 3$) hadronic rings — “ring triggers”

2.1 The Calorimeter Trigger Algorithms

The primary task of the calorimeter trigger is the identification of e/γ and jet candidates. The two types of object are very different and this is reflected in the methods used to search for them.

2.1.1 The Electromagnetic Candidate Algorithm

Electromagnetic-type objects are ‘compact’ objects, being well contained within a single trigger tower (5×5 ECAL crystals). Furthermore their size in η is even

more tightly constrained, being typically just two crystals wideⁱⁱⁱ, whilst the extent in ϕ is larger (typically 5 crystals) due to bremsstrahlung, pair-production and the dispersive effect of the 4T magnetic field. The ECAL crystals are also 26 radiation-lengths in depth, resulting in near-complete longitudinal containment of the shower.

In order to reduce the data volume, therefore, the smallest object considered in the level-1 trigger is a single trigger tower. In order that finer information is available for the identification of e/γ candidates, however, pattern matching is performed to determine whether the ratio of the highest energy pair of adjacent strips (where a strip is defined to be 5×1 crystals in $\phi \times \eta$), to the total energy of the tower is lower than some programmable fraction, R . For tower energies below 10 GeV, $R=0.95$ in order to improve rejection of fake signatures and above 10 GeV, $R=0.9$ to improve acceptance of high energy candidates. If the ratio is lower than R , the so-called fine-grain veto bit is set, otherwise it is unset.

As the ECAL crystals offer near complete longitudinal containment of the shower, for each tower a hadronic veto flag is set if the ratio of transverse energy in the HCAL to that in the ECAL is greater than a programmable threshold (nominally 0.05): such a flag identifies calorimeter towers whose signals are likely to have been produced by hadronic interactions or which contain significant background contamination.

The e/γ algorithm [63] is based on windows of 3×3 trigger tower (figure 2.1) tiled repeatedly across the entire span of the ECAL ($|\eta| < 2.5$). In each window it is required that the sum of the energy of the central tower and that of its highest energy neighbour (excluding diagonal neighbours) is greater than a programmable threshold (nominally 15 GeV at low luminosity and 30 GeV at high luminosity). By doing so, candidates of interest are identified even if they are centred on the boundary between towers. It is further required that neither the fine-grain veto bit nor the hadronic veto bit is set for the central tower; this provides restrictions on both the lateral and longitudinal shape of the shower.

ⁱⁱⁱdue to the small Molière radius (2.2 cm) of lead tungstate

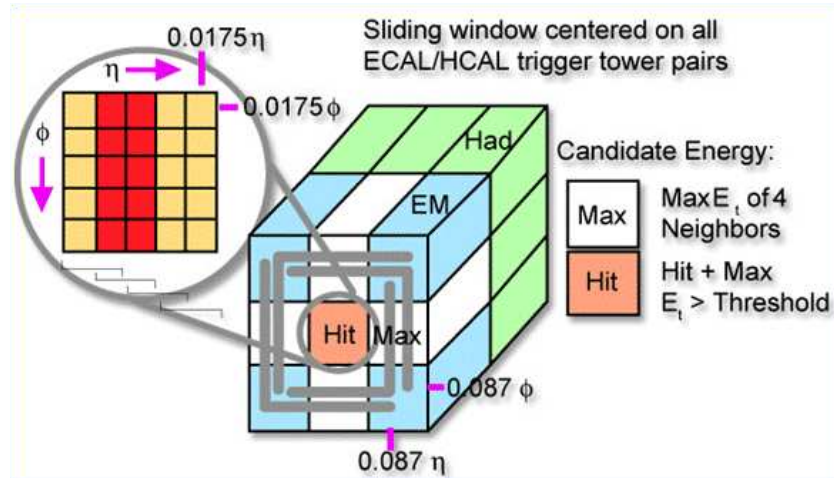


Figure 2.1: The calorimeter trigger e/γ algorithm.

If a window passes the above criteria, it is then classified according to two further 'isolation' criteria: if each of the 8 trigger towers surrounding the central tower also pass the fine-grain and hadronic vetos and at least one of the four 5-tower corners (marked as grey 'L' shapes in figure 2.1) has all towers below a programmable threshold, then the candidate is considered to be an isolated electron. If not, it is considered non-isolated.

The e/γ candidates are sorted by rank (a measure of its transverse energy and 'quality') and the four highest of each type are forwarded to the global trigger.

2.1.2 The Jet Candidate Algorithm

Whereas e/γ objects are compact and well contained, hadronic jets are much larger and, as such, must be dealt with differently. Groups of 4×4 trigger towers (ie 20×20 ECAL crystals) are summed into calorimeter regions. These 4×4 regions are the same size as the individual regions in the forward hadronic calorimeter. For each region a τ -veto bit is set if the pattern of energy deposition matches none of a set of predefined patterns (as shown in figure 2.2).

Jet candidates are defined by the criterion that the transverse energy in a region is greater than or equal to the transverse energy in each of the eight surrounding

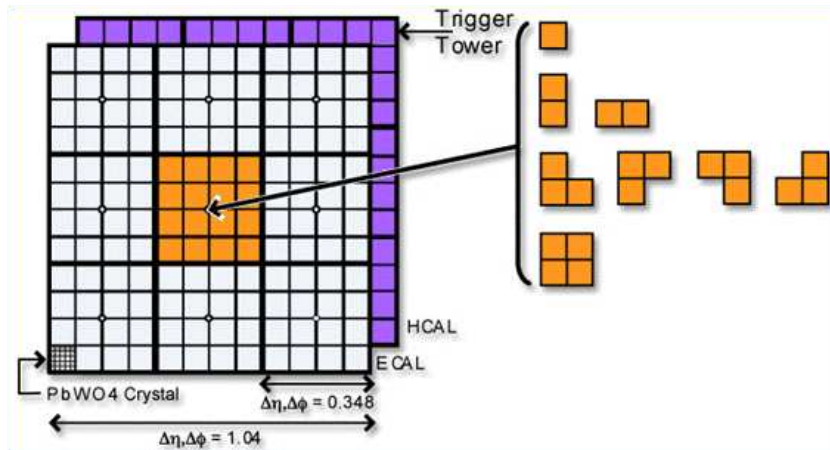


Figure 2.2: The calorimeter trigger jet algorithm.

regions. If this criterion is met, then the energy of the jet is defined to be the sum of the transverse energy of the region and its 8 neighbours. Depending on the location of the central region in η , the candidate is labelled as being either a forward- or central-jet.

Because τ leptons are relatively light and likely to be highly boosted, the hadronic decays of τ leptons result in jets which are more tightly collimated than jets produced by QCD processes. By insisting that none of the regions within the jet candidate (3×3 regions) have the τ -veto bit set (that is, at least one region matches one of the predefined patterns), the lateral extent of the jet can be inferred to be small and the jet labelled as a τ -jet.

The jet candidates are sorted and the four highest of each type, central-, forward- and τ -, are forwarded to the global trigger.

2.1.3 Other calorimeter trigger tasks

In addition to the e/γ - and jet-candidates, the level-1 trigger was designed to accept and trigger on a variety of other calorimeter signals, including:

- **Jet count** - In order to improve the trigger efficiency of rare multi-jet events, the calorimeter trigger counts the number of central, forward and τ -jet

objects fulfilling each of twelve sets of rank and position criteria and sends these twelve counts to the Global Trigger. This prevents the limit of four jet-candidates of each type from missing events with larger numbers of jets.

- **Total transverse energy** - The energies of all calorimeter regions are added as scalar quantities to calculate the total transverse energy in an event.
- **Total transverse energy in jets** - The energies of all jet candidates are added as scalar quantities to calculate the total transverse energy in an event carried by jets.
- **Missing transverse energy** - The energies of all calorimeter regions are also added as vector quantities in order to calculate the asymmetry in the transverse energy which may be used to infer the presence of neutrinos or other 'invisible' particles.

During the development of the trigger, two further quantities were added to the trigger specification:

- **Missing transverse energy in jets** - The energies of all jet candidates are added as vector quantities, to calculate the asymmetry of the transverse energy carried by jets which, again, may be used to infer the presence of neutrinos or other 'invisible' particles.
- **Energy in forward hadronic rings (ring triggers)** - In order to understand the detector, it is desirable to obtain a large sample of background (minimum bias) data, for which the low-luminosity start-up conditions are ideal. Such events are dominated by low momentum particles in the forward region and, as such, forward- (rather than transverse-) energy must be used to identify them. The ring triggers calculate the total energy in concentric rings of the forward HCAL and the number of towers over threshold within each ring. Both values are sent to the global trigger. Under normal running conditions this trigger is not used.

2.2 The Calorimeter Trigger System

By the nature of the algorithms described, they can not only be divided into discrete chronological steps, but each step may also be neatly subdivided geometrically, allowing each region to be analysed in parallel. Only when creating the final sorted lists of candidates must all the information be brought together. The calorimeter trigger system uses three stages, with each stage handling successively larger regions of the detector. The final stage of the calorimeter trigger is to bring together candidates from the entire detector, to sort them and to send the highest ranked candidates to the level-1 Global Trigger.

2.2.1 The Calorimeter Trigger-Primitive Generators (TPGs)

The first stage in the calorimeter trigger chain is the generation of calorimeter trigger-primitives. The **TPGs** (Trigger-Primitive Generators) interface directly with the ECAL and HCAL front-end electronics and receive raw calorimetric data. For the ECAL, the crystals are summed into towers and, for both the ECAL and the HCAL, the data is converted from raw energies into transverse energies and information relating to the timing of the bunch-crossing is extracted. For each tower, the ECAL TPGs additionally compute the fine-grain veto bit for use in e/γ identification and the HCAL TPGs produce a **MIP** (Minimum Ionizing Particle) consistency flag to aid muon identification. The TPGs must also synchronize the data and flag the trigger-primitive corresponding to LHC **BXo** (Bunch Crossing Zero).

The initial stage of ECAL trigger-primitive generation is performed on-detector, at the “front-end board”, by the “FENIX” [64] radiation-hard ASIC. The FENIX chip performs the summation of ECAL crystals into towers and the identification of the fine-grain veto bit. The trigger information is transmitted off-detector at a rate of 40 MHz on a dedicated 90 m optical fibre per tower, while the high-resolution data is stored in a pipelined buffer for readout on receipt of a level-1

accept signal. Sixty-eight trigger fibres from the barrel region or forty-eight trigger fibres from the endcap region are received by a **TCC** (Trigger concentrator card), which finalizes the trigger-primitive format by conversion of the endcap coordinate systems and compression to a non-linear energy scale. The finalized candidates are passed to a mezzanine card^{iv}, the **SLB** (Synchronization and Link Board), which identifies the LHC bunch-crossing zero by comparing the “observed” bunch structure with the “known” bunch structure, serializes the data and forwards it to the regional calorimeter trigger.

HCAL trigger-primitive generation is performed off-detector by the **HTR** (HCAL Trigger and Readout) card. Each board receives 16 optical fibres, corresponding to 48 channels, from the front-end electronics and performs linearization, peak-finding, MIP-finding and finalization of the trigger-primitives. The finalized candidates are again aligned and passed, via an SLB, to the regional calorimeter trigger. The HTR card also stores the raw data in pipelined buffers for readout on receipt of a level-1 accept signal.

Both the ECAL and the HCAL trigger-primitive generators have a dedicated **DCC** (Data concentrator card) through which, upon receipt of a level-1 accept signal, the raw data and the trigger-primitives are sent to the **DAQ** (Data Acquisition) system. The calorimeter TPG system is summarized in figure 2.3.

2.2.2 The Regional Calorimeter Trigger (RCT)

The **RCT** (Regional Calorimeter Trigger) receives tower level information from the trigger-primitive generators through the receiver card. Each receiver card has eight receiver mezzanine cards which provide the physical interface to the cables from the synchronization and link boards and performs deserialization of the data. The receiver card deskews and linearizes the data, sums towers into regions, performs τ -veto identification and forwards the regions across a custom backplane to the jet/summary card. The receiver card also performs data

^{iv}there are actually 9 SLBs per Trigger concentrator card

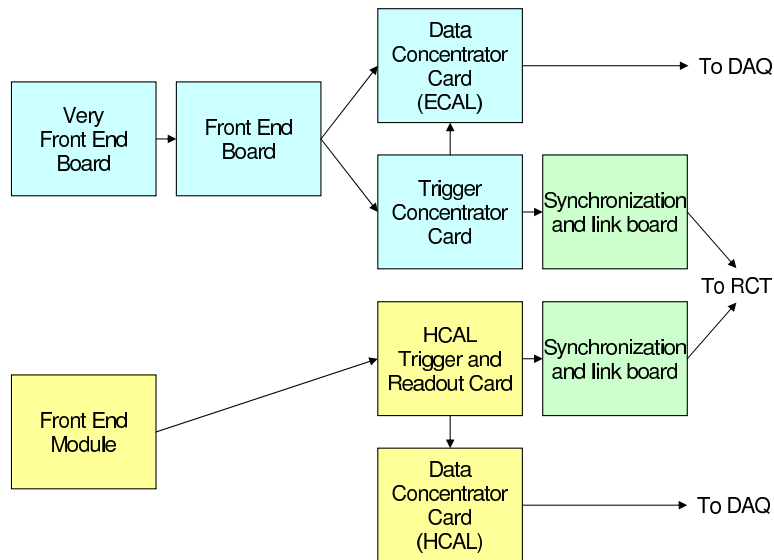


Figure 2.3: The calorimeter TPG and DAQ architecture. The ECAL TPG/DAQ system (blue) consists of the very-front-end and front-end electronics on-detector and the data and trigger concentrator cards off detector. The HCAL TPG/DAQ system (yellow) consists of the front-end electronics on-detector and the trigger readout and concentrator cards off detector. The links from both systems to the RCT are provided by the Synchronization and Link Boards.

sharing to handle the boundaries between sectors and transmits the raw tower information to the electron identification card. The electron identification card performs the e/γ algorithm and forwards the candidates to the jet/summary card for sorting. The jet/summary card sorts the electron candidates and forwards them, with the regional sums, to the global calorimeter trigger. It also acts as a host to its own receiver mezzanine cards to receive data from the forward hadronic calorimeter, which is simply forwarded to the global calorimeter trigger. The MIP bits are also simply forwarded to the global calorimeter trigger.

As the tower level information is read out by the calorimeter TPG system and all data generated by the regional calorimeter trigger is sent to the global calorimeter trigger, the RCT does not interact directly with the DAQ system. The regional calorimeter trigger system is summarized in figure 2.4.

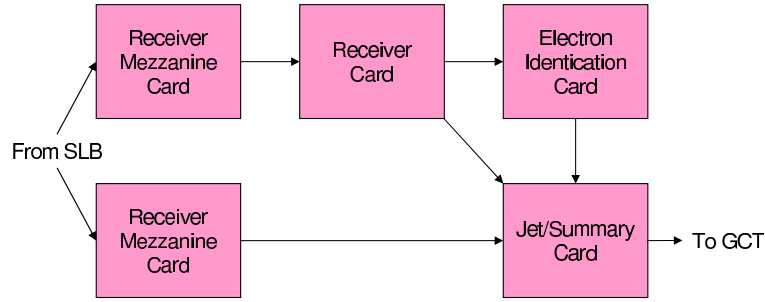


Figure 2.4: The RCT architecture.

2.2.3 The Global Calorimeter Trigger (GCT)

The formation of jet-candidates from regions and the sorting of e/γ -candidates is performed by the **GCT** (Global Calorimeter Trigger). The cables from the regional calorimeter trigger are received by the GCT source cards. The source cards separate the e/γ -candidates, MIP bits and the regional sum data and transmits each type on a separate optical fibre for processing by different subsystems of the GCT. The optical fibres for the e/γ -candidates and regional sums are received by leaf cards while the optical fibres containing MIP bits are handled by the matrix processor card. The e/γ -candidates are sorted in two stages; separately for each end of the detector and finally globally. The first sorting is performed on two leaf cards and the final sorting on the concentrator. Jet-candidates are also found by the leaf card and the candidates again sorted separately for each end of the detector and finally globally. For jet-candidates, the initial sorting is performed by the wheel card and the final sorting is again performed on the concentrator card. The calculations of total and missing transverse energies and the counting of jets are also performed systematically through the leaf, wheel and concentrator cards. The concentrator card sends the energy sums, jet counts and the four highest ranking candidates of each object type to the global trigger via the **GTi** (Global Trigger interface) card. The global calorimeter trigger stores all event data in ring buffers and, on receipt of a level-1 accept signal, sends this data to the DAQ system through the concentrator card. The global calorimeter trigger is summarized in figure 2.5 and discussed in depth in section 2.3.

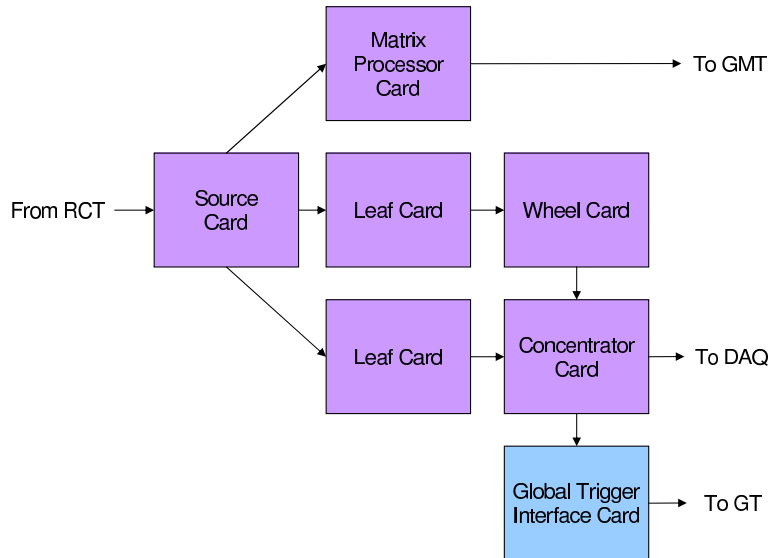


Figure 2.5: The GCT architecture. The source card separates MIP, e/γ and regional sum signals and sends them to their respective processing. The source, wheel and concentrator cards are **VME** (Versa Module Eurocard) cards, the leaf and GTi cards are dual PMC cards and the muon matrix processor is μ **TCA** (Micro-Telecommunications Computing Architecture).

2.2.4 The Global Trigger (GT)

From the global calorimeter trigger and the **GMT** (Global Muon Trigger), the **GT** (Global Trigger) receives four isolated e/γ 's, four non-isolated e/γ 's, four central-jets, four forward-jets, four τ -jets, four muons, total and missing transverse energies and jet counts. In addition, the global trigger may be supplied with 64 technical triggers for the purpose of testing and calibration, allowing triggering on events independent of the existence of physics candidates, for instance, randomly.

The calorimeter objects for the global trigger and the MIP data for the global muon trigger are received through global trigger interface cards mounted on **PSB** (Pipeline Synchronizing Buffer) cards, which align the data with the other parts of the trigger system. The muon objects from the global muon trigger and the calorimeter objects from the global calorimeter trigger (via the PSB) are sent across a custom backplane to the global trigger logic card. The global trigger logic card provides 128 configurable logic pathways which may be set to trigger on any of the incoming signals, or may combine any of the physics candidates

into composite objects, based on separation in η and ϕ , and to trigger on events containing these new objects. The output from the global trigger logic card is 128 bits, one per channel, indicating whether that channel has passed or failed the specified cut. These 128 bits and the 64 technical trigger bits are received by the final decision logic card, which monitors the trigger rate for each channel individually and pre-scales channels as necessary. The trigger control system card and level-1 accept output card, respectively, monitor the trigger status of the entire detector and, if the detector is ready to be read out, broadcasts a level-1 accept signal to all subsystems through the **TTC** (Trigger Timing and Control) system. As with the trigger-primitive generators and the global calorimeter trigger, the global trigger stores the event data in a ring buffer and, on receipt of a level-1 accept signal, sends this data to the DAQ system via the global trigger front end card.

The global trigger system is summarized in figure 2.6.

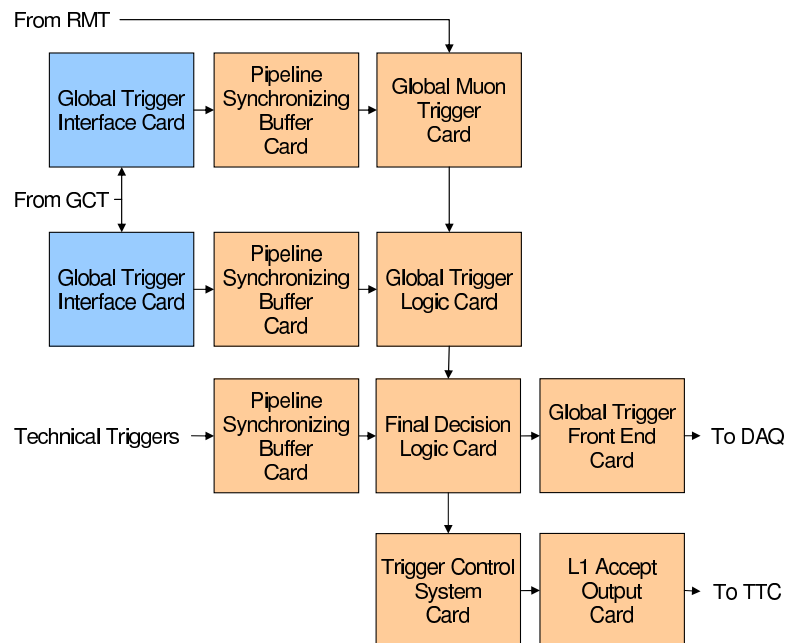


Figure 2.6: The GT architecture. The Pipeline Synchronizing Buffers align the data from all subsystems. The Level-1 Accept Output card issues the level-1 accept signal to the TTC system which broadcasts it to the entire detector and trigger system.

2.3 Design choices for the GCT

In January 2006 an alternative design for the global calorimeter trigger was proposed, to replace the existing design. The proposed LHC start date at that time was August 2007, leaving very little time for development and commissioning and the time available allowed no more than one design-revision cycle. The hardware was, therefore, developed partly around designs that were well understood and thoroughly tested.

In searching for spatially-extended objects and in sorting large lists of candidates, each algorithm must simultaneously handle data from multiple regions; in order to achieve this, the data may either be duplicated and forwarded to multiple processors, or the data may be selectively shared between the processors handling nearest neighbours. Given that the raw data rate for the GCT is of the order 250 Gbit/s [65], the former option, which massively increases the flow-forward data rate, is undesirable and so the latter option was chosen. During the construction of the GCT, however, it was discovered that the algorithms could not be fully accommodated by the scheme of data-sharing alone, and the system design was modified to include a small amount ($\sim 17\%$) of data duplication.

The complexity of the GCT is compounded by several other factors:

- The RCT was developed early in the history of CMS and, as such, utilized ASICs and discrete-logic devices, this completely prohibiting any change to the specification.
 - Again, as the RCT was designed early, the high-speed 40 MHz **DDR** (Double Data Rate) output from the RCT was necessarily differential **ECL** (Emitter-Coupled Logic); a near-redundant standard by the time the alternative GCT design was being proposed.
 - The RCT uses 68-pin SCSI-III (HD68) connectors with a non-standard 1-1 pin-pair mapping [66]. The non-standard mapping forced the use of
-

custom-made, rather than off-the-shelf, cables — this increasing the risk of production errors. HD68 connectors were also near-redundant by the time of the GCT design, having been universally replaced by more compact standards.

- The RCT crates are located 12 m from the GCT crate, leading to a high risk of electrical interference.
- The RCT crates and GCT crate are powered separately and must be kept electrically isolated to prevent ground return loops.
- The existing links to the Global Trigger were at the limits of their specification and unstable, requiring replacement.
- The GCT must provide a readout pathway for the calorimeter trigger, this requiring an interface with the data acquisition (DAQ) system.

For the sake of simplicity it was necessary to minimize the amount of data sharing between processors by concentrating the data in as few processors as possible. It was decided that this could most easily be achieved by utilizing the on-board **SerDes** (Serializer/Deserializer) technology on modern, large FPGAs and using high-speed serial links, rather than large parallel buses.

The spatial compression of data from the RCT to the GCT could then be achieved in several steps: custom cables convert the SCSI-III (HD68) connectors of the RCT to **VHDCI** (Very High Density Cable Interconnect) connectors, the latter being small enough to fit on a VME card of single unit (1U) width. The data is then serialized and transmitted optically on a single channel fibre to an optical patch-panel and retransmitted, on 12-channel fibres, to receivers where it is converted back to electrical signals for deserialization inside the processor. Not only does this scheme spatially compress the data, but the use of optical links, rather than 12 m copper links, mitigates the possibility of electrical interference or grounding problems.

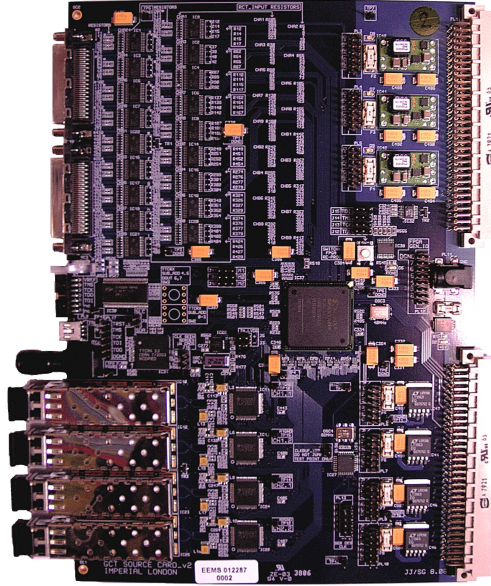


Figure 2.7: The GCT source card. Full details are given in figure 3.2.

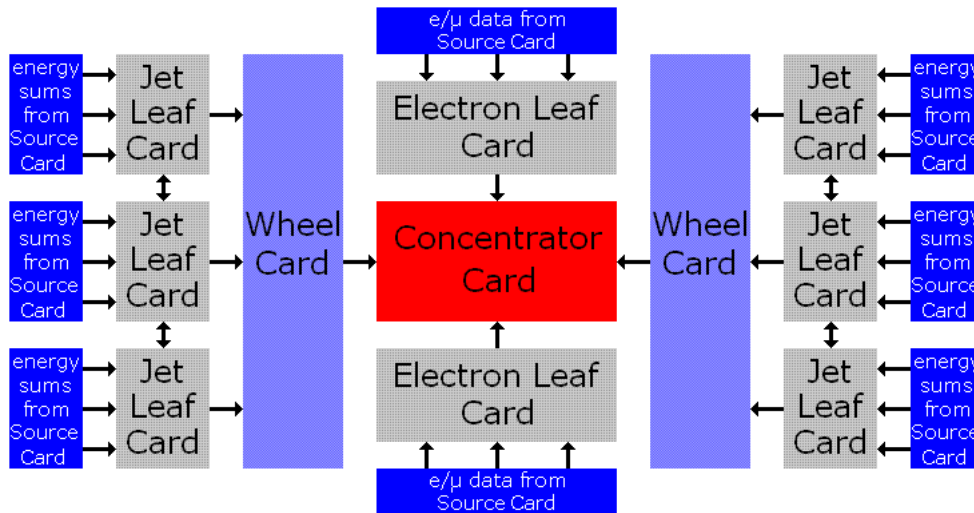
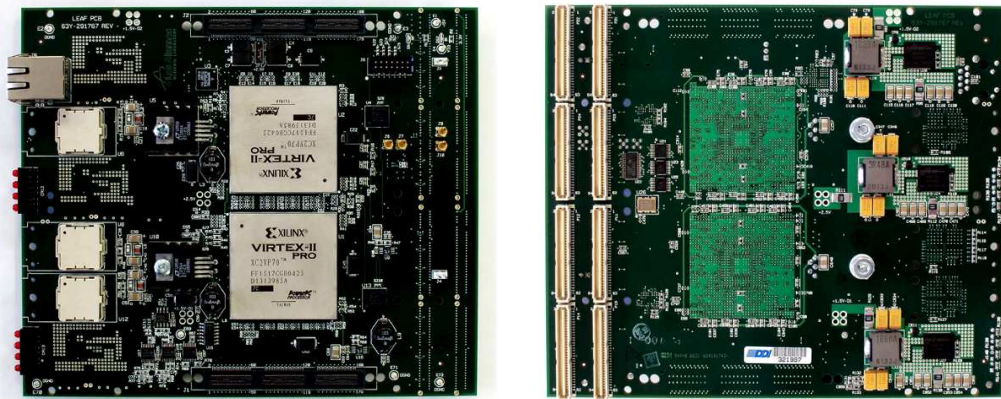


Figure 2.8: The GCT architecture. The mirror symmetry of the detector along its length is reflected both in the Electron leaf cards (here top to bottom) and in the wheel cards (left to right). The rotational symmetry may be seen in the Jet leaf card triplets. The arrow indicating sharing between the top and bottom member of each triplet is omitted here for the sake of clarity. The minimal data duplication across the $\eta = 0$ boundary is also omitted for clarity.

The electro-optical conversion is performed by the GCT source card, which may be seen in figure 2.7 and is discussed in chapter 3.

The CMS calorimeter is symmetric along its length and has 72-fold rotational symmetry. As the calorimeter trigger algorithms consider only contiguous detector regions, the algorithms must respect these symmetries, and, for simplicity, it is highly desirable that the trigger hardware should also reflect them. The GCT was designed to do this. The e/γ -trigger pathway handles each end of the detector symmetrically before finally merging the data from both ends, to perform the final sorting of candidates. The jet-finder pathway is also symmetric across the $\eta = 0$ boundary; due to the larger volume of data, however, each end may not be treated as a single entity. Instead, the hardware used a 3-fold rotational symmetry with data sharing across the ϕ -boundaries, before the lists of candidates are merged and sorted; first within their own end of the detector and finally globally. The aforementioned data duplication is also symmetric across the $\eta = 0$ boundary. The symmetry may be seen in figure 2.8.



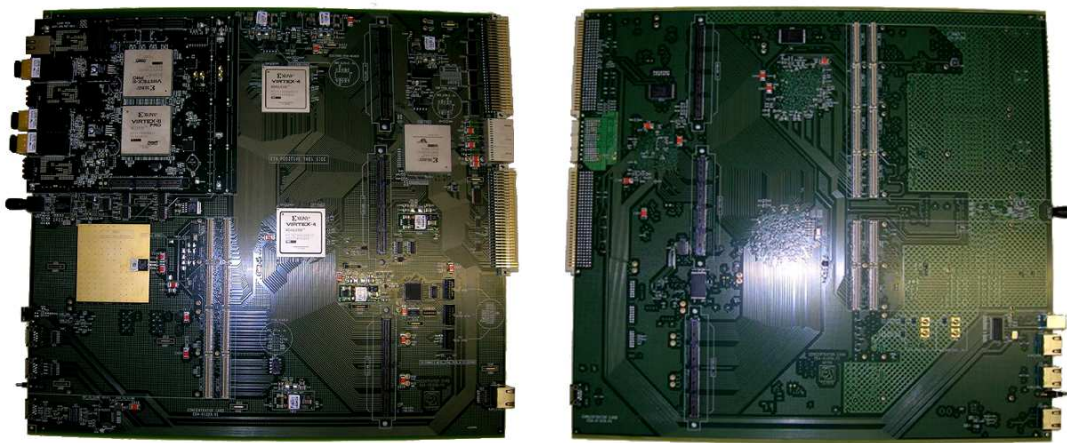
(a) Upper surface of the leaf card

(b) Lower surface of the leaf card

Figure 2.9: The GCT leaf card. The key features on the upper surface are 2 Xilinx Virtex-2 Pro FPGAs (centre), two high-speed connectors (top and bottom) and three 100-pin meg-array connectors for SNAP-12 optical links. The key feature on the lower surface is the dual PMC connector (left).

The main processing of the GCT is performed by the leaf card (figure 2.9). Each leaf card houses two Xilinx Virtex-II Pro 70 -7 FPGAs [67], at the time of design

the largest Xilinx FPGAs with functional serializers, and three SNAP-12 optical receivers^v. The leaf card has a **dual-PMC** (dual PCI Mezzanine Card) form factor and is additionally equipped with two high-density, high-performance connectors for up to 96 high-speed differential signals, 48 per connector. Through these connectors it is possible to interconnect any number of leaf cards in a daisy-chain topology, including the possibility of connecting the cards in a loop.



(a) Upper surface of the concentrator card (b) Lower surface of the concentrator card

Figure 2.10: The GCT concentrator card. On the upper surface, from left to right, are two dual-PMC sites (one with leaf card fitted), two Xilinx Virtex-4 FPGAs, three high-speed connectors, a Xilinx Virtex-2 FPGA and VME64x (SLINK) [68, 69, 70] connectors. On the lower surface, from left to right, are three high-speed connectors and a single- and dual- PMC site.

The concentrator card, figure 2.10, is a 9U-high VME card based around two 1500-pin Xilinx Virtex-4 LX FPGAs; these FPGAs were chosen for their extensive **I/O** (Input/Output resources) and high logic density. The card may host up to 3 dual-PMC cards and one additional **PMC** (PCI Mezzanine Card). For the GCT, the dual-PMC sites are occupied by two electron-algorithm type leaf cards and the global trigger interface card, while the additional PMC slot is unused. The concentrator card is equipped with six high-density, high-performance connectors for up to 400 high-speed differential signals and it is through these that it connects to the two wheel cards. The same connectors additionally provide the programming, clock and control pathways to the wheel cards, which have no

^vAlthough only 32 of the 36 channels are routed due to the limited number of embedded deserializers in the FPGAs.

external control interface of their own. An additional Xilinx Virtex-II FPGA acts as a communications manager, implementing an interface to the TTC system, a VME-64x interface to control the entire system, an S-LINK [68] interface for reading out the system on receipt of a level-1 accept signal and both Ethernet and **USB** (Universal Serial Bus) interfaces for standalone testing. The spare I/O on the communications FPGA is routed to a collection of 8P8C (“RJ45”) connectors for use should the need arise.

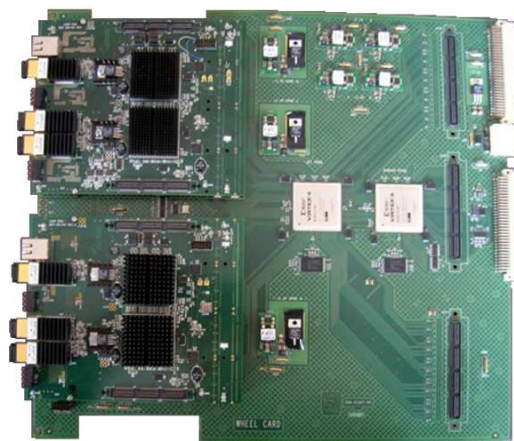


Figure 2.11: The GCT wheel card. From left to right, are two dual-PMC sites (with leaf card fitted but no interconnecting ribbon cables), two Xilinx Virtex-4 FPGAs, three high-speed connectors and VME64x connectors (power only). At the top and bottom of the card are two cutouts through which the ribbon cables pass.

The wheel card, figure 2.11, has many design features in common with the concentrator: it is also a 9U-high VME card based around two 1500-pin Xilinx Virtex-4 LX FPGAs, hosting 3 dual-PMC cards and equipped with high-density, high-performance connectors. There are, however, some key differences: the wheel card uses the VME form factor for power and mechanical support only, receiving control signals from the concentrator card. Similarly, the wheel card relies on the concentrator for management of both the TTC and JTAG systems. Whereas the concentrator hosts two leaf cards which are essentially independent, the wheel card hosts three leaf cards which are linked in a closed loop by ribbon cables. In order to allow interconnection of the leaf cards, which are mounted on both surfaces of the wheel card, required that “cutouts” were made in the PCB through which the high-performance ribbon cables could pass.

Not shown on figure 2.8 is the GCT muon subsystem. This system runs in parallel with the above system and forwards MIP- and Q-bits from the RCT to the Global Muon Trigger (GMT). This is performed by the muon matrix processor card, figure 2.12. The matrix card uses an entirely different design philosophy to any other system in CMS. Based on the successful demonstration of the use of the integrated serializers in high-speed links between the source and leaf cards, the board is designed around a high-speed serial link — rather than a large parallel bus — architecture. The design is based around a large Xilinx Virtex-5 LX110T FPGA^{vi} with 16 integrated multi-gigabit transceivers operating at 3.6 Gbit/s and 16+16 optical links. A novel feature is the use of a 72×72 protocol-agnostic gigabit crosspoint switch (by Mindspeed [71]), through which all of the serial signals are routed. As the switch provides capability for signal duplication and signal broadcasting, it was desirable that the serial signals should be transmissible to other cards in the system across a back plane. A VME backplane is not compatible with such a philosophy and, instead, a design based around the modern μ TCA standard was chosen. As the board was no longer restricted to a VME control bus, a modern 10/100 Mbit/s Ethernet control bus was chosen; the physical connection for the bus being through the backplane for standard usage or through an onboard 8P8C socket^{vii} for standalone testing. A full Ethernet stack is implemented in an on-board microcontroller to minimize resource usage in the processing FPGA. The board is fitted with 512 MByte of DDR2-666 SDRAM memory for flexibility, although this is not used in the GCT.

^{vi}chosen for its high logic-density and high-speed serializers.

^{vii}8P8C Ethernet connectors are commonly but incorrectly referred to as “RJ45” connectors. Although 8P8C connectors are used for both, the electrical connections for the RJ45 specification and for the T568A/B Ethernet/telephone specification are different.

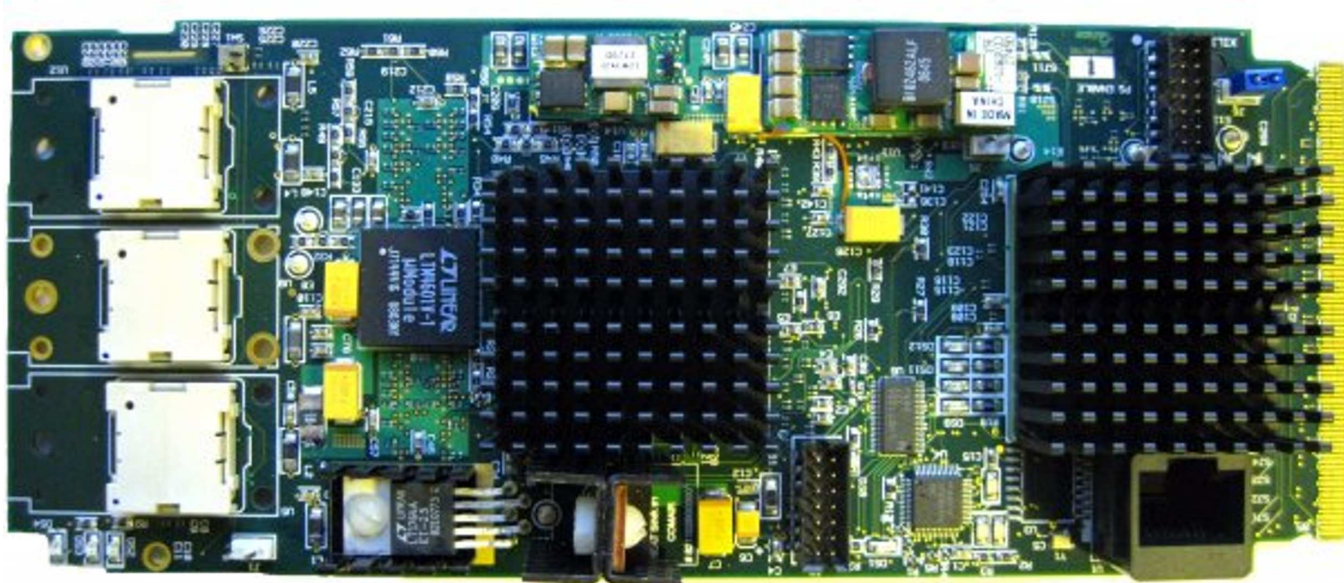


Figure 2.12: The GCT matrix processor Card. From left to right are three 100-pin meg-array connectors for SNAP-12/POP-4 optical links, high current linear power supplies, a Xilinx Virtex-5 LX110T FPGA (with heatsink), a Mindspeed 72×72 protocol agnostic crosspoint switch (with heatsink) and a μ TCA backplane connector. On the reverse surface are 512 MByte of DDR2-666 SDRAM memory and an NXP LPC2366 microcontroller. A single width μ TCA/AMC card is $74 \text{ mm} \times 183.5 \text{ mm}$ and is shown at approximately full size. The card has been taken as a template for the direction in which the future development of the trigger will be taken, because of its small size, high-bandwidth and task-independent design.

Chapter 3

The Global Calorimeter Trigger Source Card

“However far the stream flows, it never forgets its source.”

- Nigerian Proverb

3.1 The GCT Source Card

Beyond the basic requirement that the source card converts the differential ECL signals from the RCT into an optical format, there is a stricter requirement that this must be accomplished with a latency of less than 2 LHC bunch-crossings (50 ns) in order that there is sufficient time to perform the GCT algorithms. Although not in the original remit, the source card functionality was extended during development to include the ability to inject variable test patterns for the testing of the downstream system and to provide a readout pathway for the testing of the upstream system.

As different data types (e/γ , jet, muon) are processed by different subsystems within the GCT, it is necessary for the source cards to separate the various data streams, which are both merged onto and shared across the output cables from the RCT.

It was originally intended that the 6 cables from each RCT crate, each 34-bit wide, would be handled by 3 source cards, retransmitting the data serially on 12 optical links. Commercial serializers, including the Texas Instruments TLK2501 serializers used on the source cards, operate on 16-bit wide basis: it is, however, possible to reconcile the 68-bit wide inputs with the 64-bit (4×16 -bit) wide outputs as there are unused bits on each RCT cable that can be discarded.

It was realized during the construction of the GCT, that there was insufficient bandwidth to perform the jet finding algorithm across the $\eta = 0$ boundary using data sharing alone. Although various solutions were proposed, the chosen solution was to duplicate the $\eta = 0$ data at the source cards. However, the duplicated data increased the data volume, necessitating the introduction of two further optical links per RCT crate, this, in turn, requiring the introduction of an extra source card per RCT crate.

This solution was the least technologically ‘risky’, in that it required no new designs or modifications to existing hardware, relying instead on an increased number of a proven card (the source card) and utilizing the spare optical channels available on the leaf cards.

Noting that each source card has four optical links, a saving could be made by pairing RCT crates and sharing one source card between two RCT crates effectively reducing the number of additional source cards from one per RCT crate to half. As such a total of 63 source cards are used in the GCT.

It was initially intended that the source cards be a “passive” card, simply capturing the data from the RCT and relaying it via optical links. The relative flexibility of FPGA-based source card to the ASIC-based RCT and to the more tightly constrained leaf cards, coupled with the source cards USB control system meant that, during construction of the GCT, significant extra requirements were made of the source cards.

The RCT has no readout path of its own and the high-speed USB controls provided by the source cards allows direct capture of the RCT data straight into

software. The RCT also has only a limited ability to inject test data into the system. To remove any dependence on external systems, the source cards can load patterns via USB and transmit the data to the leaf cards for the purposes of verifying correct connection of the optical fibres and correct operation of the leaf card algorithms. During testing it was found that the gigabit deserializers on the leaf card failed to lock to the optical links from the source cards when the source card links were clocked from the CERN TTC clock, due to jitter on the signal. To overcome this deficiency of the leaf card, the source cards were modified to operate the optical links from a standalone 100 MHz oscillator. The leaf cards require the BCo marker in the data stream in order to configure and align the data across the optical links. This limited the GCT by requiring the RCT to be configured before the GCT could be started; a dependency which was clearly undesirable. The source card firmware was modified in order to allow insertion of an internally generated BCo marker onto the optical link to remove this dependence.

Because the system was developed late in the history of the experiment, the source cards were highly constrained by the existing hardware. The available space for locating the cards restricted their size to 6 rack units and the short time scale for hardware development required the maximal reuse of proven technology. As such, the source card has a 6U VME form factor, (although uses the VME crate only for power and mechanical support) and was partly derived from the IDAQ card, a card previously developed at Imperial College for the CMS micro-strip tracker readout system. Because of the large number of source cards and the limited space available on the source card, it was neither physically nor economically possible to use the large Xilinx Virtex-II FPGA's as used on the IDAQ and so the smaller, cheaper Xilinx Spartan-3 FPGA was used instead. This had several implications for the source card; because of the limited FPGA I/O, limited PCB space and limited experience using the VME interface, and because the source card was originally intended not to be read out during running, it was neither possible nor desirable to implement the low speed, I/O intensive VME interface. Instead, it was decided to implement a USB 2.0 interface, requiring

fewer I/O, capable of considerably higher data rates and providing enormous benefits for testing in terms of ubiquity; any modern PC can operate a USB 2.0 device, whereas a specialist hardware-controller is needed to test a VME device. The later requirements, that the source cards should provide readout for the RCT and be capable of injecting test patterns, subsequently necessitated the implementation of USB controls in the online system.

3.2 Design of the Source Card

3.2.1 High level hardware design

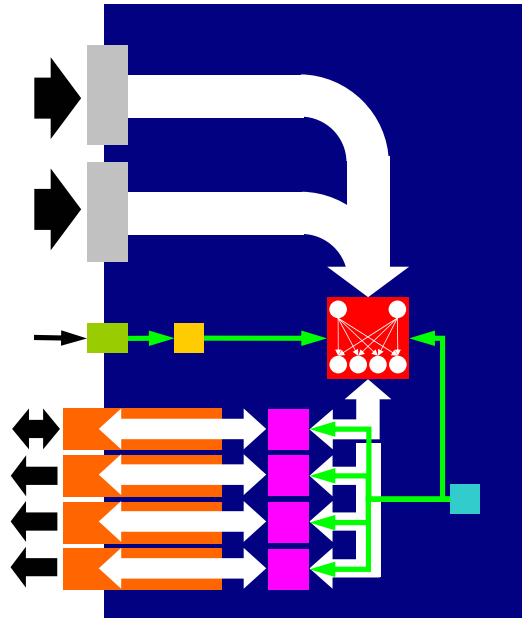


Figure 3.1: Block Diagram of the CMS GCT source card. The white arrows indicate the data path through the board, the green arrows represent clock signals on the board and the black arrows indicate data/clock direction into or out of the board. The green and orange box respectively represent the TTC receiver and CERN TTCrx which decodes the TTC clock and the turquoise box indicates the standalone clock. The FPGA (red) acts as a configurable multiplexer for the data path. The clock and datapath components are indicated in their correct relative locations on the board.

As stated previously, the primary goal of the source card is to convert the differential ECL signals from the RCT into an optical format. This is achieved as shown in figure 3.1. Data enters the card at the top of the front panel, the two

channels are merged, then split into four, and the data transmitted optically to the rest of the GCT through the bottom half of the front panel. The top optical channel is bidirectional, allowing receipt of optical data. The RCT data is synchronous with the TTC system and, as such, a TTC clock is required to capture it. This is received optically via a front panel connection, before being extracted electronically from the mixed data signal. The optical links are run asynchronously with respect to the TTC system, using instead a standalone clock; this clock is distributed in parallel to the FPGA and to the four serializers.

Figure 3.2 shows a source card with the key features and components indicated. Each card receives data from the RCT via the two VHDCI SCSI connectors and retransmit it through the four **SFP** (Small form-factor, Pluggable) optical links. Each optical link is driven by a Texas Instruments TLK2501 serializer/deserializer, capable of operating at an 8b/10b coded data rate between 1.5 and 2.5 Gbit/s. The data rate used was originally intended to be 1.6 Gbit/s, synchronous with the TTC clock, but this was later increased to 2.0 Gbit/s to improve link operation. The optical links used on the board are Avagotech HFBR-5720AL fibre channel transceivers [72] which can operate at a rate of up to 2.125 Gbit/s. In this configuration, the FPGA essentially acts as a bit-configurable multiplexer for a pair of RCT cables.

3.2.2 Serial link design

Although both the optical components (HFBR-5720AL) and the SerDes (TLK2501) are fully bidirectional, there are insufficient I/O pins and digital clock managers on the Spartan-3 FPGA to implement four fully bi-directional optical channels and, as such, only a single receiver channel was fully implemented for the purposes of loop-back testing. Although this is limiting for the purposes of self-testing the cards, the limitation can be overcome: the Texas Instrument TLK2501 SerDes include an internal **PRBS** (Pseudo-Random Bit Stream) generator and tester, allowing testing of the high-speed (2.5 GHz) differential traces between the SerDes and the SFP and also FPGA-independent verification of the optical

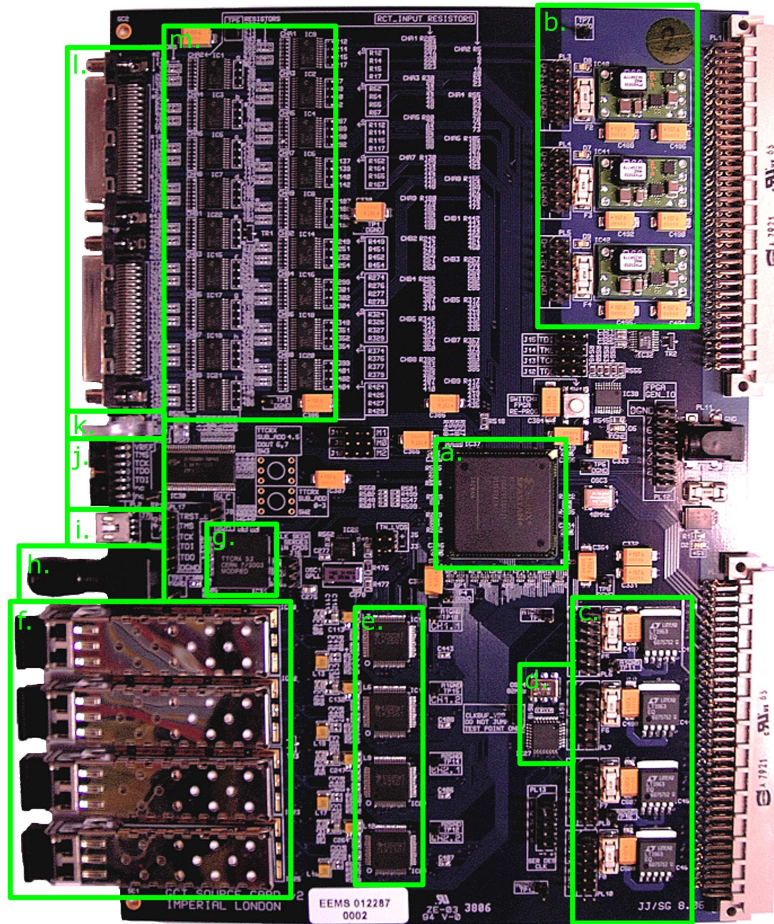


Figure 3.2: The CMS GCT source card. Of particular interest are: (a) Xilinx Spartan-3 FPGA, (b) Switch-mode (digital) power supply, (c) Linear (analogue) power supply, (d) 100 MHz LVDS oscillator and 1:10 Low-Jitter LVDS Clock buffer, (e) Texas Instrument TLK2501 Serializer/Deserializers, (f) Avagotech HFBR-5720AL fibre channel transceivers (SFPs), (g) CERN TTCrx, (h) Pin Diode, (i) USB mini-b connector, (j) JTAG connector, (k) Tri-colour LEDs, (l) VHDCI receptacles and (m) Differential ECL-LVTTL level translators.

links, either by coupling two optical channels on the same board or between two different boards.

3.2.3 Clock system design

The use of high-speed serializers requires careful selection of components so that the additive jitter from different sources does not exceed the maximum jitter limit for the serializers, in the case of Texas Instruments TLK2501 serializers, 50 ps peak-peak[73]. This limit on the jitter is considerably lower than

250 ps peak-peak jitter on the output from a digital clock manager on the Xilinx Spartan-3 FPGA [74] and, as such, the FPGA can not directly clock the serializers. Instead, the source card uses a dedicated low jitter clock buffer driven by low jitter inputs and both the FPGA and the serializer are slaves to this clock. Using dedicated clocking components has the added advantage that the signal-signal skew between different clocks can be better controlled than if the clocks were driven by the FPGA. The source card clock tree is shown in figure 3.3.

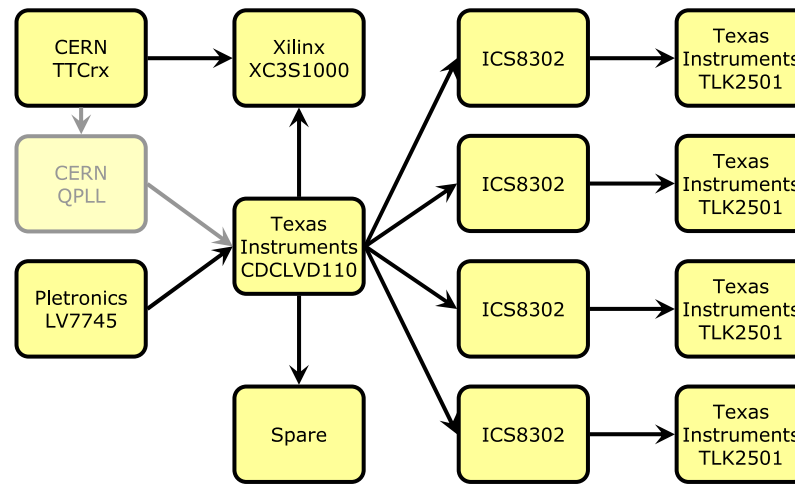


Figure 3.3: The source card clock system. Although the CERN QPLL is physically included in the clock system, it is unused due to the change from TTC-synchronous to asynchronous optical links. Not included in the diagram is the continually running 40 MHz system clock nor the 24 MHz USB clock.

The source card has two low jitter clock sources: a 100 MHz standalone LVDS clock by Pletronics and a CERN **QPLL** (Quartz Phase-Locked Loop). The CERN QPLL provides a low jitter **LVDS** (Low-voltage differential signalling) clock phase-aligned with a clock provided by the CERN TTCrx and therefore phase-aligned with every other component in the TTC sub-system. Running synchronously with the TTC system is desirable as both the input (from the RCT) and the output (to the Global Trigger) were designed to be TTC synchronous. In running the GCT in a TTC synchronous mode, the total latency is reduced by eliminating costly clock-domain changes in the data path: it was originally intended that the standalone oscillator would be used only when the TTC clock was unavailable. Due to the inability of the leaf card's **MGTs** (Multi-Gigabit Transceivers) to lock to the optical links when clocked by the QPLL, however, it

became necessary to run the links from the oscillator — thereby becoming asynchronous with the rest of the level-1 trigger — and to later realign the data to the TTC clock for transmission to the global trigger.

The TTCrx offers two clock outputs with programmable skew as well as the zero-phase master output. The programmable skew clocks allow the clocks on each board to be shifted relative to the TTC clock of the entire TTC system in 104 ps steps. This allows the phase offset between two TTC synchronous subsystems to be optimized to account for cable (propagation) delay and to maximize signal integrity. One of these clocks is used to drive the QPLL, while the other is passed directly to the FPGA.

The low latency clock used is selected by the CDCLVD₁₁₀ 1:10 clock fanout buffer [75], which produces 10 duplicate clocks with an expected maximum channel-channel skew of 30 ps. Because both the data path from the FPGA to the serializers and the serializers themselves are clocked directly from the clock buffer, it is necessary that the variation in the lengths of the PCB traces from the clock buffer to the serializers exactly matches the variation in length of the data lines from the FPGA to that serializer. In order to introduce a phase offset between channels with the purpose of minimizing transient demand on the power supplies, the absolute lengths of both the clock and data traces is different for each serializer. An additional complication is the necessary conversion of the clocks from LVDS (as used by the buffer) to the **LVTTL** (Low Voltage Transistor-Transistor Logic) used by the serializers; a conversion not required for the FPGA clock or the data path. The level translation is achieved using an ICS8302 LVDS-LVTTL translator with a quoted part-part skew of 500 ps. On the prototype source cards, the phase offset between the outputs on the fanout buffer and the clock input pins on the serializers was measured to be 2.6, 2.6, 2.8 and 3.0 ± 0.1 ns, going from the serializer nearest to the FPGA to the one furthest from it. As the offsets were identical (within experimental limits) for both source cards, this indicates that the variation in propagation delay is dominated by the differences in PCB trace lengths for each clock signal.

3.2.4 Power system design

A further consideration in the use of high-speed links is the quality and stability of the power supply. High-speed serial links use differential signal standards for which the “no-man’s land” between a binary ‘0’ and ‘1’ is only a few tens of millivolts, and, as such, noise of this magnitude on the power supply can affect the distinction between binary states. Even without causing a misrepresented bit, variation in the supply voltage can change the transition point, such that consecutive bit transitions are not evenly spaced in time. This effect can usually be safely ignored in systems with separate clock and data lines, as the clock should be aligned such that the transitions are centred on a stable point in the data signal. On high-speed serial links, however, the clock signal is embedded into the datastream itself and is recovered at the receiver by a phase-locked loop that is used to synchronize the decoder to the data stream. A large edge-edge jitter will result in the phase-locked loop being unable to lock and the datastream being indecipherable.

Modern electronic equipment relies on three types of voltage regulators: switch-mode regulators, linear regulators or **SCR** (Silicon-controlled rectifier) regulators. As the names suggest, switch-mode regulators operate with the component transistors in the saturation region and, hence, are either “on” or “off”, whereas linear regulators operate with the component transistors working in the linear region. The two, fundamentally different, modes of operation produce very different behaviours: because of the restricted size of the linear region, linear regulators are limited both in the amount of current that they can output and in the required input and available output voltages; switch-mode regulators are not limited in this way and can drive considerably larger currents, as well as offering considerably more flexibility in the choice of input and output voltages. Because switch-mode regulators employ pulse-width modulation when switching between conducting and non-conducting states, however, each transition produces switching-noise of several tens of millivolts and the transition frequency is of the order 100 kHz to 1 MHz. Such noise can also feed back into

switch-mode regulators and cause multiple switch-mode regulators to oscillate in phase, whereas a linear regulator acts to filter out high frequency input noise.

Because of superior noise performance, Linear Technologies' LT1963 linear regulators [76] are used to power the source card serializers and optical drivers, while Texas Instruments' PTH05050 [77] switch-mode regulators are used to power the digital logic and ECL front-end due to their greater efficiency (and thus thermal and power performance) and current-handling capability. To further minimize the likelihood of power supply noise, the power planes are split so that the linear and switch-mode supplies do not overlap, with noise-sensitive components kept well separated from the switch-mode plane.

3.2.5 USB interface

Although the source card has a 6U VME form factor, there was neither sufficient space on the PCB nor sufficient FPGA resources to implement the VME protocol, the card relying on the VME connectors only for power. In order to minimize FPGA resource usage, it was desirable that as much of the communications protocol should be kept off the FPGA as possible and, to this end, it was decided that a USB interface was most suitable as a number of one-chip solutions are available. Furthermore, the IDAQ board (previously developed at Imperial College) had already successfully demonstrated the use of one such chip, the Cypress EZ-USB SX2 [78] and, in the interest of minimizing design risk, the existing design was reused.

Although USB communication is physically based on a packet structure, it is logically implemented as a set of unidirectional pipes with the device (as opposed to the host) end of each pipe referred to as an endpoint. The Cypress SX2 provides four endpoints, two of which receive data from the host PC (called EP2 and EP4) and two of which transmit data in the opposite direction (EP6 and EP8). The SX2 transparently handles all of the intricacies of the USB protocol and provides the user with a "glueless" **FIFO** (First-In-First-Out) interface,

consistent with the requirement of placing minimal demand on the FPGA. The SX2 is USB-2.0 compliant and supports transfer speeds of up to 480 Mbit/s, 50% higher than that of VME64.

Although the IDAQ board uses a standard USB-B connector, there was insufficient space on the source card front panel to do the same requiring, instead, the use of a USB Mini-B connector; this change did not represent a technical risk as the electrical standard and pin assignments are identical.

3.2.6 JTAG interface

As FPGAs utilize volatile RAM-based “soft logic”, they cannot retain their state when powered down and thus require external programming every time the device is powered up. On the source card, programming the FPGA may be achieved in two ways: directly using the **JTAG** (Joint Test Action Group — IEEE 1149.1) interface or indirectly using the onboard serial **EEPROM** (Electrically Erasable Programmable Read Only Memory), which is itself programmed by a JTAG interface. JTAG is a daisy-chained serial interface; every JTAG device must support four lines; TDI (data in), TDO (data out), TCK (clock) and TMS (mode select), with a daisy chain formed by connecting the data-out line of each device to the data-in line of next. The input to the first device in the chain and the output from the final device, as well as the clock and mode select lines, must be connected to a JTAG source. Two such sources exist on the source cards and are selected by the placing of jumpers: the first is a 14-pin header on the front panel for use with the Xilinx Platform-USB programming device. JTAG headers are usually “keyed” to ensure that the connector may only be fitted one way; a layout mistake resulted, however, in the pinout being rotated through 180° with respect to the key. Given the tight development schedule, it was decided that, rather than making a design revision, the flaw would be accepted and the board programmed using a specially modified JTAG cable with the connector key removed. The second JTAG source is the FPGA itself, with four of the general-purpose I/O pins connected to the four JTAG lines: by this means, a

JTAG datastream may be generated within the FPGA, passed along the JTAG chain (which itself includes the FPGA) and returned to the FPGA. In practice, the JTAG stream is not generated within the FPGA but received by the FPGA from the controller software and simply forwarded through the chain. The EEPROM could, by this means, be reprogrammed remotely whilst the FPGA was running, rather than having to be manually reprogrammed on every board. The only time when remote programming is not available is the initial programming of the EEPROM, before which the FPGA has no firmware with which to control the JTAG chain. The source card JTAG chain is shown schematically in figure 3.4.

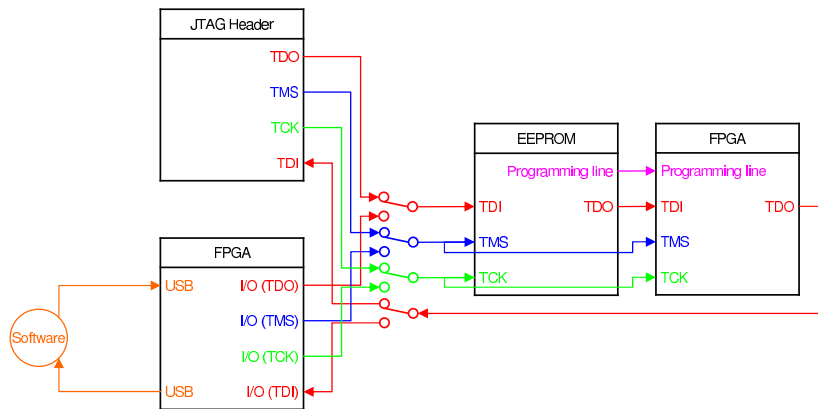


Figure 3.4: The source card JTAG chain. The JTAG chain is indicated in red (data), blue (mode select) and green (clock). The bank of switches indicate the bank of jumpers used to select the JTAG source. Not part of the JTAG chain but included for completeness is the serial programming line (pink) through which the FPGA is programmed by the EEPROM on power up. The orange lines are not part of the standard JTAG chain but indicate the interface with the software via the USB connection. Although there is only a single FPGA on the board it is represented here twice to indicate the two different roles it plays, as a JTAG source (left) and as a JTAG slave (right).

The concept of having a **PLD** (Programmable Logic Device)ⁱ reprogramme itself indicates clearly the blurring of lines between software, firmware and hardware: the JTAG datastream flows transparently from software through the PLD, by means of the implemented firmware, to the hardware bus. The stream is considered by the FPGA both as data to be manipulated (when it is received from the PC) and as the programme which manipulates the data (when it is received

ⁱAlthough PLD used to refer to a particular class of programmable logic devices it is now commonly used, and is used here, to refer generically to any kind of configurable logic including FPGAs.

through the JTAG chain or through the serial programming line). Such a blurring places even greater emphasis on good system design.

3.2.7 The Source Card Firmware Design

The Xilinx Spartan-3 FPGA provides four clock manager blocks that are fully utilized on the source card in the interfacing of the FPGA with the four on-board clock domains:

- The **system clock (40 MHz)** is permanently running, being driven by a local, on-board oscillator. It provides a clock for the core system functionality (for example, the USB link, temperature monitoring and indicator LEDs), such that the board is always accessible even without the global TTC system clock.
- The **TTC clock (40.08 MHz)** is driven directly from one of the fine-skewable outputs of the TTCrx, and thus has a well-defined phase relationship with the global TTC system. This is used both in capturing the data from the RCT and in capturing the signals embedded into the TTC stream (such as the BCo strobe and L1 RESYNC).
- The **Receiver clock (100 MHz)** is driven by the **PLL** (Phase-Locked Loop) clock output of the TLK2501 deserializer, providing a clock with a well-defined phase relationship with the signals from the TLK2501 and hence from the optical receiver.
- The **Transmitter clock (100 MHz)** is driven from one of the outputs from the CDCLVD110 clock buffer, which is itself driven by the 100 MHz output from an onboard oscillator. It is driven in parallel to the TLK2501 serializers and thus has a fixed phase relationship with the serializer input register.

Within the FPGA, the bulk data path (that runs from the RCT through to the transmitters) is clocked by a doubled TTC clock, so as to capture the DDR RCT

input, and maintains synchronization throughout; as stated previously, it was originally intended that the optical links should also run at this frame rate. In order to minimize the number of firmware changes required by the transition from 80.16 MHz to 100 MHz links, a clock bridge was added immediately prior to the output IOB. A four-phase clock bridge design was chosen in order to minimize the latency of the transition, with the data-valid signal to the serializer being cleared when the queue is empty.

The data from the RCT includes no provision for error checking or correction. The phase bit in the data stream is monitored in order to verify the integrity of the signal: it is expected that the phase bit should change twice for every TTC clock count, except on LHC BXo where the phase bit is held high for the entire clock count; this is called the **BCo** (RCT Bunch Crossing Zero) marker. The BCo is extracted from the stream and the timing checked both against the previous BCo marker and against the TTC BXo marker, with both quantities being expected to remain constant and fixed; should either value change, then both an error-flag and a latched error-flag are set. These flags are polled regularly by software to monitor the status of the RCT.

The data from both inputs is duplicated onto two data paths: the primary path forwards the data to a routing table where it is split into four 16-bit streams, the content of each stream depending on the context (or mode) of the card. The four outputs from the routing table are registered to guarantee timing on the transmitter side of the firmware. The mode of the source card (equivalently the routing table) is programmable by software. The secondary data path forwards the data to a 64-bit wide, 4096-element deep ring buffer. The data is written continuously in one direction until the receipt of a trigger signal that stops the write process and starts reading the data back in the opposite direction, reading it into the dedicated high data-volume USB FIFO. As the data is recorded continuously, data both preceding and following a trigger may be read out, a feature which is useful for debugging in the event of a system upset. In order to take advantage of this capability, however, the signal which triggers the readout requires a large degree of flexibility. The trigger signal is generated upon the receipt of one or

more of the following; a TTC L1RESET, a TTC BXo, an RCT BCo, an input pattern matching a 64-bit software-programmable pattern or a software commandⁱⁱ and is subjected to a software-programmable delay. The reset sources, the number of captures allowed between resets and the length of each capture are also software-programmable parameters.

On the transmitter side of the firmware a **CRC16** (16-bit Cyclic Redundancy Check) is calculated per channel and the data multiplexed with various test patterns (software programmable fixed pattern, software programmable data stream, counter, **LFSR** (Linear Feedback Shift Register), A-5) and control patterns (alignment comma, status register, the aforementioned CRC16). The multiplexer is controlled by a **FSM** (Finite State Machine) and a software programmable instruction list. The trigger signal for the transmitter FSM is generated identically to that of the RCT capture buffer. The control patterns may be used during normal operation by insertion into the LHC bunch gaps, where the RCT data is null and vetoed by the leaf card anyway. The final stage of the data path is the transition from 80.16 MHz to 100 MHz.

Because of limited FPGA resources, the capture buffer was given a secondary purpose of storing a software programmable data stream: data corresponding to a maximum of 4096 half bunch-crossings may be loaded by software, triggered and transmitted as though it were real data. Clearly the two uses of the buffer are mutually exclusive, the usage being determined by software settings.

A third high data-volume path exists for the purpose of capturing data from one of the optical receivers. Although not required for normal use, the inclusion of the optical receiver allowed testing of the optical links without the necessity of developing additional hardware (see section 3.3.4). When not in use, the DCM is deactivated, thereby disabling all of the logic in the optical receiver block. When in use, the data and control lines are first registered and the data subsequently written into a FIFO by the TLK2501 data-valid strobe, thereby automatically

ⁱⁱThe use of a software trigger is only for the purposes of testing as the software, USB and control buses are asynchronous with respect to the TTC clock and the RCT.

stripping the control and error codes from the data stream. The optical receiver path is then multiplexed with the RCT capture readout path and the data read into the high data-volume USB FIFO. The optical receiver system lacks a feedback mechanism and, as such, care must be taken to limit the size of the packet to be captured to less than the FIFO depth (256 elements). This is trivial, as the receiver system is only used in controlled tests where the software has control over both what is being sent and received.

3.2.8 The Source Card Front Panel Design

All cards must be fitted with front panels in order that vertical airflow be maintained in the VME crates. In the case of the source cards this represented a particular challenge and four factors were of particular importance: first, the front face of the source card is particularly congested, with approximately half of the available area occupied by components requiring a particularly durable material. Secondly, in order to reduce strain on the VHDCI connectors, it was required that the front panel mechanically supports the connector; to prevent ground return loops, electrical isolation must be maintained between the grounded outer shell of the connector and the ground plane of the source cards, this prohibiting the use of bare metal panels. Thirdly, the small size of the mini-USB and VHDCI sockets do not allow the connectors to span the gap between the edge of the card and a flat front panel, instead it was decided that the panel should be bowed to allow for easy connections, requiring a flexible material. Finally, CERN regulations require the use of flame-retardant, halogen-free materials in underground areas. To fulfil the four requirements, the panels were machined from transparent, halogen-free, flame-retardant polycarbonate (which has excellent strength, flexibility and electrical characteristics) and held at either end with standard VME-64x handles and in the middle by a nut on the shaft of the pin-diode.

3.3 Stand-alone Testing of the GCT Source Cards

The GCT source cards were manufactured and assembled by Exception PCB [79] and Exception EMS [80] respectively. For the purpose of staged testing, the production was split into a prototype run (2 cards), a pre-production run (8 cards) and two production runs (each of 36 cards).

Self-testing of the board (beyond power-up tests) is limited due to the unidirectional and specializedⁱⁱⁱ nature of the links on the board. Instead, a comprehensive series of tests was devised to separately test each subsystem of the board and finally as an entire system.

The datapath may be considered to be split as follows

- RCT cables
- ECL front-end
- FPGA
- Serializers/Deserializers
- Optical transmitters and fibres

or equivalently shown schematically in figure 3.5.

The other critical components are

- Power supply
- Clocking

ⁱⁱⁱThe SFPs are in-fact bidirectional and the TLK2501 SerDes provide loopback functionality. However, as the control sequences are complicated and the results of such a test of limited value when compared to the full test suite, it was considered that implementing an additional set of tests was not a good use of the time and resources available.

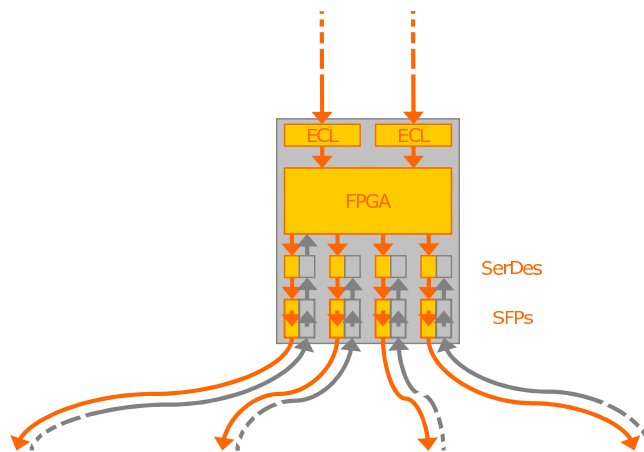


Figure 3.5: Schematic of data flow through the CMS GCT source card. From top to bottom, the data passes through the RCT cables, the ECL front-end, the FPGA, the SerDes, the optical transmitters (SFPs) and the optical fibres. In this figure, and in the figures which follow, active components are indicated in orange while disabled features are indicated in grey.

If any one of the above systems fails, the source card would fail to perform its designated task and would risk crippling the calorimeter trigger. Demonstration that all of the subsystems work individually is, however, insufficient to guarantee correct operation when working together: as such, once each system had been demonstrated individually, the components were systematically chained until the entire data path was being tested. The testing of power supplies was limited to testing for short circuits between the various power and ground planes and testing of voltages at the output of the regulators.

3.3.1 Bit Error Rate (BER) Tests and Poisson Statistics

In order to evaluate the quality of a copper or optical link carrying a digital signal, it is necessary to determine the rate at which bits are misrepresented. Furthermore, given their random nature, an absence of observed errors is not the same as a proof of the absence of error but may be used to infer a limit on the **BER** (Bit Error Rate) using Poisson statistics.

Given a binary hypothesis, in this case the occurrence or non-occurrence of a bit-error, where the probability, μ , of occurrence is very much lower than the

probability of non-occurrence, then for a large number of repetitions, the probability of exactly n occurrences is expressed by the Poisson distribution, equation 3.1.

$$p(n) = \frac{e^{-\mu} \mu^n}{n!} \quad (3.1)$$

In this context, the Poisson mean, μ , is defined as the absolute probability of occurrence, P , multiplied by the total number of samples, M .

$$\mu = MP \quad (3.2)$$

It is necessary to assert a confidence level to which we wish to declare the non-occurrence of errors. It is usual to assert a 95% confidence on the non-occurrence of errors requiring that the probability, $p(n)$, of the occurrence of precisely $n = 0$ errors is 5%. This may be stated mathematically in the form of equation 3.3.

$$p(0) = \frac{e^{-\mu} \mu^0}{0!} = e^{-\mu} = 0.05 \quad (3.3)$$

which may be inverted to obtain $\mu = 3.00$.

By equation 3.2, in order to assert a limit on the bit-error rate of P with 95% confidence, therefore, requires demonstration of non-occurrence in $M = 3/P$ samples. Modern telecommunications standards are typically qualified to a bit error-rate of 10^{-12} , requiring the capture of 3×10^{12} bits of data for a 95% confidence level.

3.3.2 Cable testing

The RCT uses 68-pin SCSI-III (HD68) connectors with a non-standard 1-1 pin-pair mapping [66], the non-standard mapping forcing the use of custom-made [81], rather than off-the-shelf, cables, thereby increasing the risk of production errors. The source cards were also restricted to the use of VHDCI connectors, by the limited space available, thus requiring different connectors on each end of the customised cable. To further complicate matters, HD68 connectors were not available for the source card test system (see section 3.3.3), instead requiring the use of VHDCI connectors and, thus, a second type of cable. In order to test both

types of cables to verify the non-standard pinout and characteristic impedance^{iv}, **TDR** (Time Domain Reflectometry) [82] methods were used: a fast, single-ended to differential signal driver was constructed to the target cable impedance specifications and a 175 mV, 10 ns **FWHM** (Full Width at Half Maximum) Gaussian pulse injected into the cables and monitored on a high-bandwidth oscilloscope (figure 3.6).

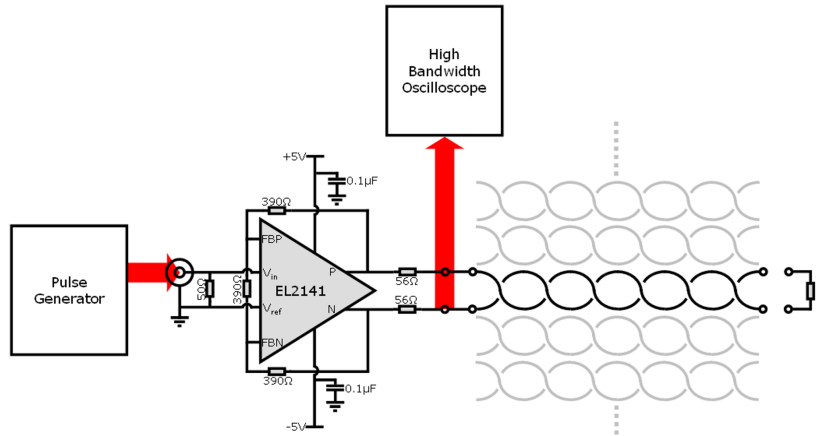


Figure 3.6: The RCT-GCT cable testing apparatus. A 10 ns FWHM pulse was generated using a commercial pulse generator and converted from single-ended to differential form using custom-built electronics. The outgoing and reflected signals were observed using a high-bandwidth oscilloscope at the input of the cable. Both the driver's connection to the cable and the termination resistance, R , were free to be coupled to any pair of pins in the cable, not only those expected to form a differential pair.

The following configurations were tested

- Cable unconnected
- Open-ended assuming the required, non-standard pinout
- 0Ω termination assuming the required, non-standard pinout
- 120Ω termination assuming the required, non-standard pinout
- Open-ended assuming the unwanted, standard pinout
- 0Ω termination assuming the unwanted, standard pinout
- 120Ω termination assuming the unwanted, standard pinout

^{iv}The RCT was constructed to expect a characteristic cable impedance of 125Ω

- Open-ended using widely-spaced, uncorrelated pins
- 0Ω termination using widely-spaced, uncorrelated pins
- 120Ω termination using widely-spaced, uncorrelated pins

Assuming the required, non-standard pinout, tests showed that the cables displayed the characteristic reflection when the cable was unterminated, anti-reflection with 0Ω termination and near-zero reflection with 120Ω termination, as expected of a good-quality twisted pair. The results of tests assuming the unwanted, standard pinout matched those of the tests using widely-spaced, uncorrelated pins, both showing a large degree of cross-talk and multiple reflections. These results indicate both that the pins were correctly paired and the pairs correctly shielded, and that the characteristic impedance was indeed slightly greater than the 120Ω available for testing. For the purpose of verification of the reflectometry setup, the characteristic delay of the reflection for the SCSI-III/VHDCI cables was measured to be 50 ± 5 ns. Taking the **NVP** (Nominal Velocity of Propagation) of a shielded twisted pair to be $\sim 0.66c$ (approximately 2×10^8 m/s), this indicated a double cable length of approximately 10 m or a cable length of approximately 5 m. For the VHDCI/VHDCI cables, the characteristic reflection had a delay of 20 ± 5 ns, corresponding to a length of approximately 2 m. Both of these values agreed with the physically measured lengths.

Once the relative positions of the source cards and the RCT crates had been defined, it was shown that the maximum path length between the RCT output and the source card input was 2.75 m, including a routing overhead. As such, the production length of the RCT cables was reduced from 5 m to 3 m to both reduce the cable latency by approximately 10 ns^v and reduce the probability of electromagnetic interference.

^vFor comparison, recall that the source cards are allowed a maximum of 50 ns processing latency

3.3.3 Source Card ECL front end

In order to test the ECL front end of the source card, a test card to emulate the output from the RCT was developed (figure 3.7).

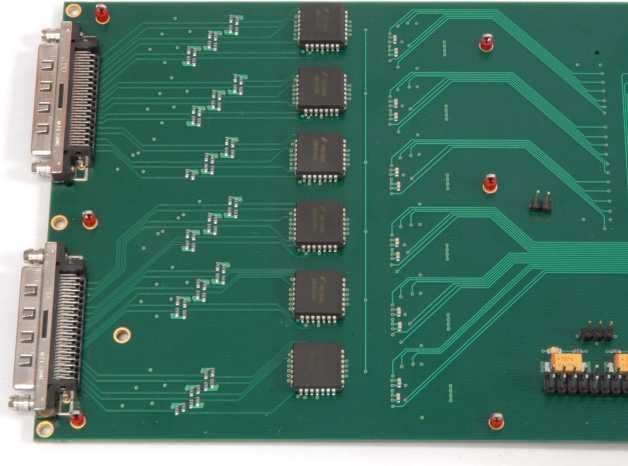


Figure 3.7: The RCT Emulator card. The card is mounted on an IDAQ card by means of 0.1 inch pitch header and acts as a logic-level translator. The central integrated circuits are six-channel TTL-ECL converters driven by the FPGA on the IDAQ. The ECL output is subjected to a -5 V bias and output on two VHDCI SCSI connectors.

In order to reduce the amount of development, the RCT Emulator card was designed as a mezzanine card for the IDAQ; a board previously developed at Imperial College. The IDAQ was designed for use in the readout system of the CMS silicon microstrip tracker and is a 6U VME card housing a large Xilinx Virtex-II FPGA. The unused FPGA I/O resources are routed to two banks of high-density, high-performance connectors to allow the expansion of functionality. The board is programmed using a Compact-Flash card boot-loader and is equipped with an Ethernet interface, a full VME interface and a USB-2.0 controller. The RCT Emulator card uses the IDAQ card to provide the core infrastructure, in particular the FPGA and USB communications, while the mezzanine card itself provides the physical interface and performs the necessary level translation to match the output from the RCT, and thus the input to the source cards. The IDAQ does not, however, have a TTC input and cannot, in itself, drive a TTC synchronous signal. In order to minimize the amount of additional infrastructure it was decided that — rather than adding a TTC input and decoder to the RCT Emulator card — an

LVDS clock would be fed from the source card under test^{vi}, via the Emulator to the IDAQ (figure 3.9(a)). Due to the unavailability of a CERN TTCi [83], the TTC clock and control signals were provided by a CERN TTCvi MkII [84] configured by an SBS VME crate controller [85] and a CERN TTCex [86]. A programmable pattern generator by Agilent was used to provide a precision reference clock. The experimental set up is shown in figure 3.8.

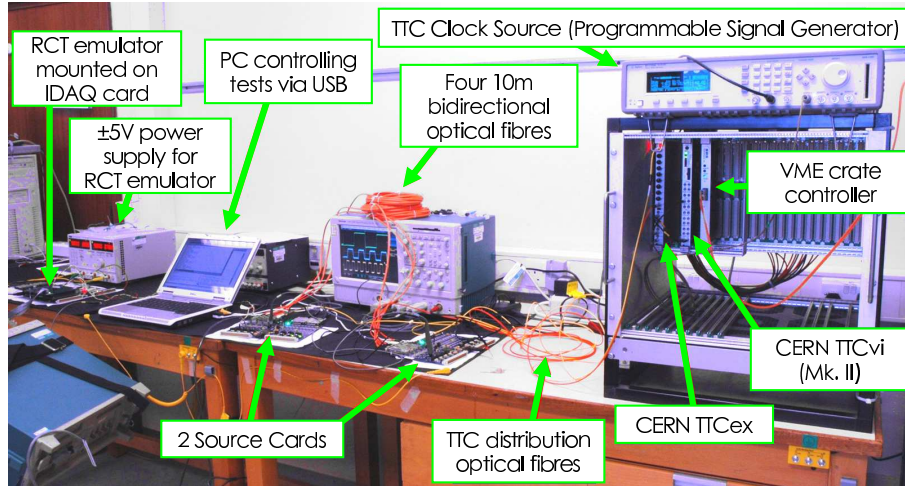


Figure 3.8: Photograph of the experimental set-up used in the testing of the CMS GCT source cards. This particular test set-up was used to implement the tests shown in figures 3.9(a) and 3.10(b).

The RCT Emulator firmware is minimal, consisting of a 4096-element deep buffer (equivalent to the buffer depth in the source card), a USB interface identical to that in the source cards and a minimal core-system: clock management and temperature monitoring. The buffer is loaded from software via the USB interface and, upon receipt of a software trigger, read out to the RCT Emulator card in a TTC-synchronous fashion. A single bunch-crossing zero marker (BCo) is inserted into the datastream in firmware and it is upon this signal that the source card is set to trigger the data capture. In order to verify the data, the source card buffer was read out and compared to the data loaded into the Emulator. As the effective throughput of the USB system is around 20 Mbyte/s (160 Mbit/s),

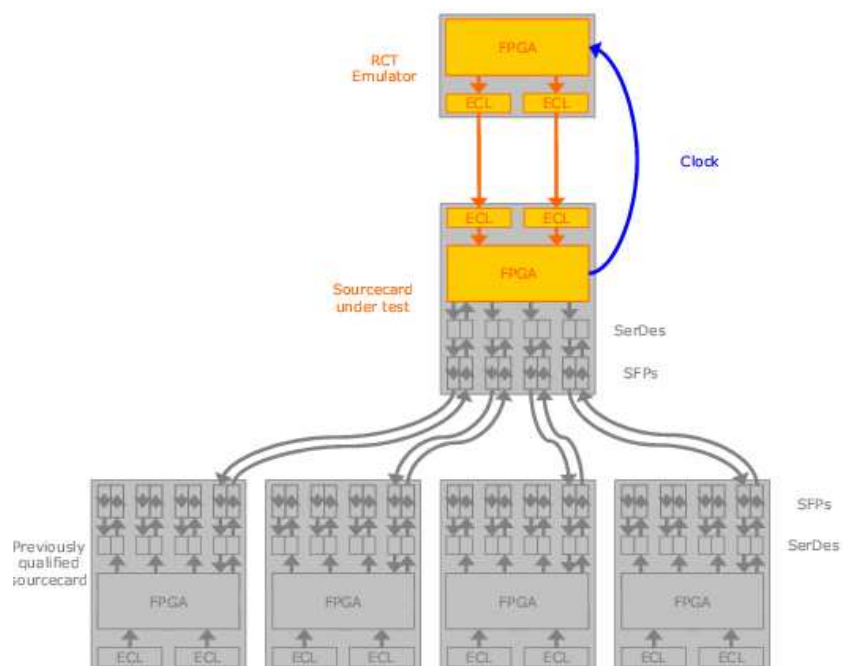
^{vi}This was the standard mode of testing. In fact, any source card which had an active TTC input could be used; this cross-coupled mode of operation was used to check that the quality of the capture was not due to the fact that both the Emulator and the source card were being clocked by the “same” clock.

it takes at least 5.2 hours to qualify the RCT inputs to a bit error rate of 10^{-12} . This test was performed several times for various patterns including A-5, F-o^{vii}, counters and a pseudo-random bit pattern based on a Mersenne twister [87].

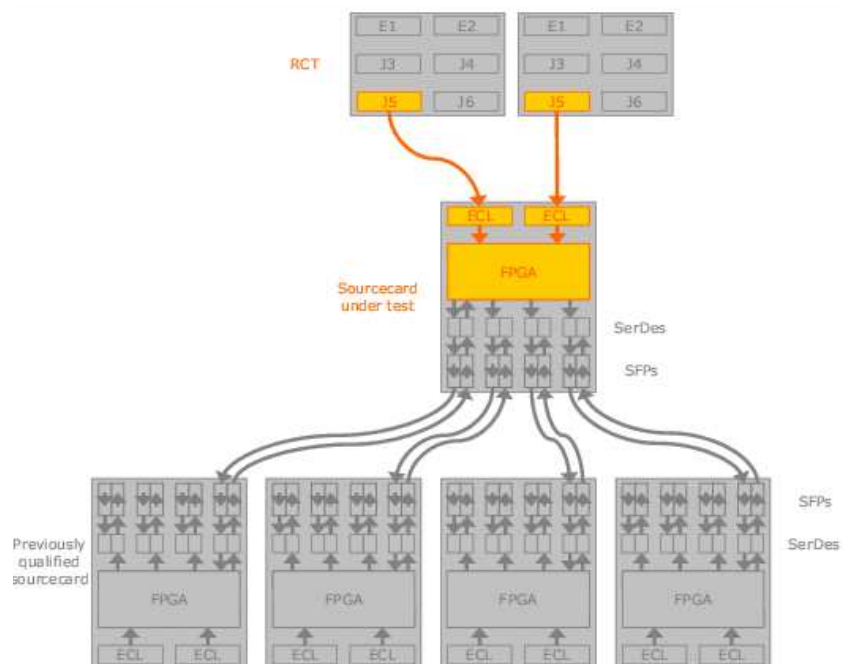
Once it was shown that the source card could successfully capture data from the Emulator, an analagous test was performed at CERN: data from the JETSUM 5 output on two RCT crates was captured by the source card, with both RCT crates and the source card being clocked by a common TTCi (figure 3.9). Verification of the functional equivalence of the RCT Emulator to the RCT meant that validation of later batches of source cards could be performed using the RCT Emulator only and that the results would apply equally to the RCT. It is interesting to note that the source card performance with the RCT was actually better than with the RCT Emulator, as the latter had been designed with variations in the trace lengths to marginally exceed the design limits of variation in the RCT trace length. The actual variation in track length of the RCT is, however, superior to the target variation, this resulting in lower bit-to-bit skew.

The testing with the RCT was limited by the available modes of operation of the RCT; in this case the RCT was operated in stand-alone mode, allowing a programmable bit-shift pattern to be produced from the outputs of the RCT several clock cycles after a TTC reset. While the capability to stress the links was restricted by the limited selection of patterns, the full test allowed a measurement of the absolute timing between the two systems, something which is not possible with the RCT Emulator: the common TTCi, shared by both the RCT and the source cards, not only ensures that the clocks in both systems are synchronized, but also allows TTC commands to be sent both simultaneously to the two systems and synchronously with the TTC clock. By triggering the source card data capture from the TTC L1-reset and observing the period between the start of the captured block and the start of valid data, the timing relationship between the RCT and source card could be deduced and, thus, the latency of the RCT output, cable assembly and source card ECL input could also be determined.

^{vii}The A-5 test involves sending $A...A_{16} = 1010...1010_2$ followed by $5...5_{16} = 0101...0101_2$ on each successive clock cycle to maximize the probability of channel-channel cross talk. The F-o test involves sending $F...F_{16} = 1111...1111_2$ followed by $0...0_{16} = 0000...0000_2$ on each successive clock cycle to maximize the number of bit transitions and the transient current drawn.



(a) The RCT Emulator test



(b) The RCT test

Figure 3.9: Schematic layout of the two tests of the source card ECL inputs. The paths in orange indicate the data path used in the test. For the RCT Emulator test, a TTC clock signal must be fed from the source card to the RCT Emulator as the Emulator has no TTC clock input.

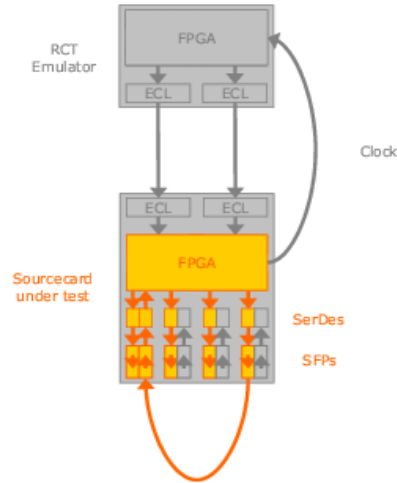
3.3.4 Optical Link Tests

The initial tests of the source card optical links were made in the form of loop-back tests (figure 3.10(a)) and board-to-board testing (figure 3.10(b)).

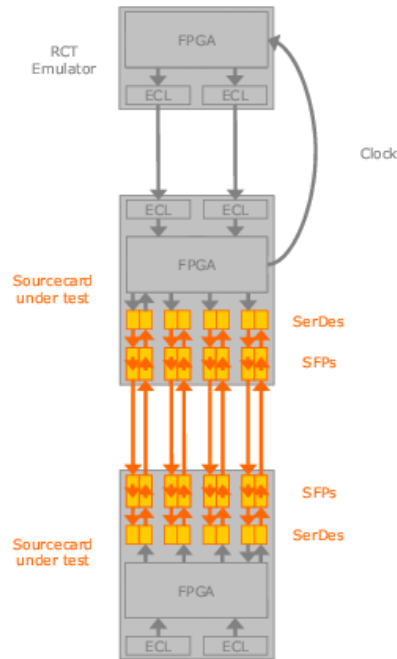
The loop-back test was used to verify the ability of each source card optical transmitter channel to correctly send, serialize and transmit optical data, and of the source card optical receiver channel^{viii} to receive, deserialize and correctly capture data. The board-to-board test allowed for optical isolation of the transmitting and receiving system, to ensure that the correct capture of data was not an artefact of the transmitter and receiver being in the same clock domain. Although three of the deserializer channels are not connected to the FPGA, simultaneous testing of the eight optical channels (four bidirectional channels per board) is possible due to the test functionality built into the TLK2501 SerDes. In test mode, each serializer produces a PRBS pattern and each deserializer provides automated verification of the receipt of such a pattern; on receipt of an erroneous dataframe, a flag is raised and the error logged by the FPGA, from which the error-count is accessible to the controller PC. Unlike the optical loop-back tests or the capture from the RCT Emulator, this scheme does not require the data to be read into the PC for verification: this has the advantage that the test may be performed at the full 1.6 Gbit/s/channel (6.4 Gbit/s/board), rather than being limited by the maximum 480 Mbit/s (or effective 160 Mbit/s) of the USB connection and the time taken for the subsequent checks. As such, to quantify each link to a BER of less than 10^{-12} at 95% confidence takes approximately 32 minutes and verification of a BER of less than 10^{-14} at 95% confidence (as was performed for the two prototype source cards) takes about 55 hours.

Before the final decision was made on the choice of optical components, testing was required to ensure optical compatibility between the proposed parts and to ensure that there was sufficient contingency in the specifications. This was achieved in three stages: by examining the quality of the optical output from

^{viii}recalling that although each SFP and TLK2501 SerDes is bidirectional, only one of the four deserializers was connected to the FPGA due to a lack of resources in the FPGA.



(a) Loop-back test configuration



(b) Board-to-board test configuration

Figure 3.10: Schematic layout of the initial tests of the source card optical links. The paths in orange indicate the data path used in the test. Although only one of the four deserializers is connected to the FPGA, the SerDes can automatically generate and verify a PRBS pattern, allowing simultaneous testing of all channels. To test the serial link between FPGA and SerDes requires that each link is tested separately.

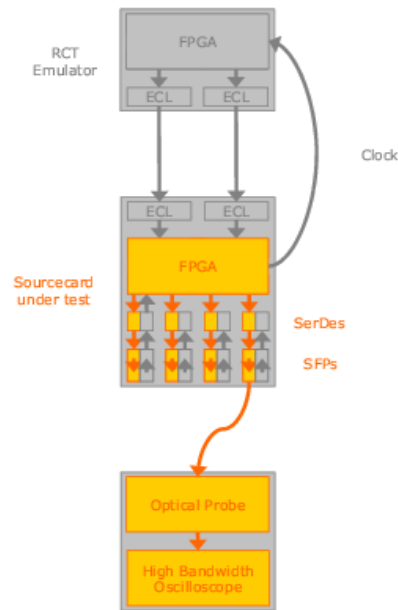
the source card, by examining the quality of the electrical output of a SNAP-12 receiver^{ix} when driven by a source card and by measuring the bit-error rate as a function of link attenuation. These tests are indicated in figures 3.11(a), 3.11(b) and 3.13 respectively.

For the first two tests (figures 3.11(a) and 3.11(b)) eye diagrams were produced and the results overlaid with a standardized “window” to indicate whether the signal was within, or whether it broke, the specification. The results, shown in figure 3.12(a), indicated problems with the SFP chosen for the source cards; namely, that the signal was being over-driven, causing “ringing”. To determine the cause of this effect, the SFP was replaced with an alternative part of higher specification and it was found that the ringing did not occur. This indicated that the problem was not inherent to the source card, but either in the SFP itself or in the relationship between the board and that type of SFP. When the SFP (whose signal was outside the specification) was coupled to the SNAP-12 receiver, however, the signal quality was regained (figure 3.12(b)) as the receiver contains a low-pass filter designed to suppress the high-frequency transients. As it is the electrical signal, not the optical signal, which is relevant to the receipt of the data, the SFP/SNAP-12 pairing was considered acceptable.

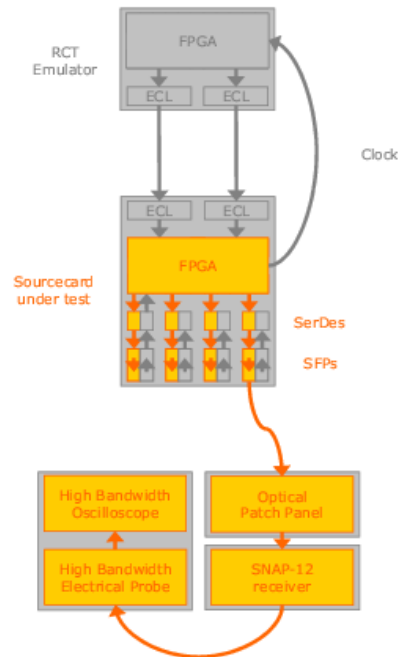
The quoted minimum expected **OMA** (Optical Modulation Amplitude) of the transmitter output and quoted minimum acceptable OMA for the receiver input are -10 dBm and -13 dBm, respectively. Assuming the worst case scenario, this leaves a maximum allowance of -3 dBm for fibre losses. The expected fibre losses, calculated from the manufacturer’s specifications, indicated losses of -2 dBm and, as such, there was little safety margin. Although the eye diagrams qualitatively prove compatibility of the parts in this configuration, a more comprehensive test was required to quantify the stability of the behaviour and the margins of the system.

Using an optical probe, the average optical power and OMA of the transmitter were measured to be -6.23 ± 0.1 dBm and -4.71 ± 0.1 dBm, respectively, while an

^{ix}the type intended for use on the leaf card



(a) Test of transmitter optical quality



(b) Test of transmitter and receiver signal quality

Figure 3.11: Schematic layout of the two physical tests of the source card optical links. The paths in orange indicate the data path used in the test. The link quality was tested (a) directly at the output of the optical links and (b) electrically immediately after the receiver.

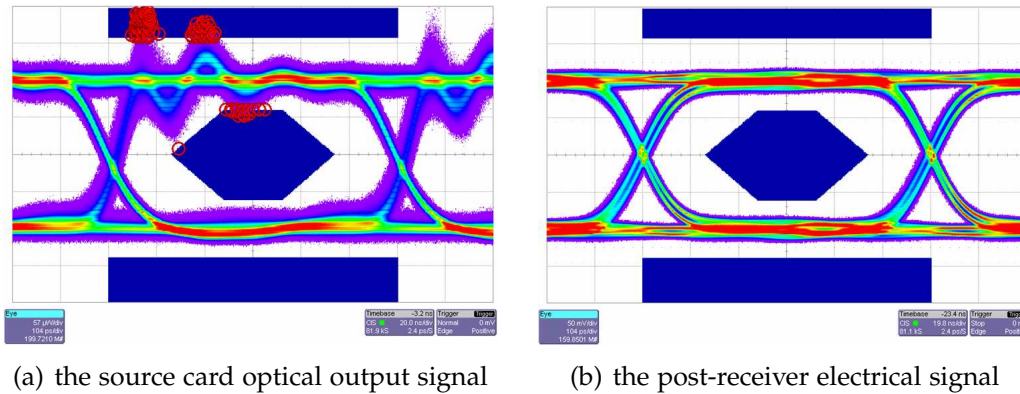


Figure 3.12: Eye diagrams resulting from the first two test of the source card optical links. Points circled in red lie outside the link specification and correspond to “undefined” behaviour (and hence failure of the test).

optical power meter recorded a slightly higher average power of -5.64 ± 0.1 dBm. While the two measures of optical power differ, both measurements lie within the manufacturer’s stated operating range: as the difference between the power measurements was $\mathcal{O}(10\%)$, it was decided that the measured value of the OMA would provide a reasonable estimate of the true value and that alternative methods of measuring the OMA would only be necessary if the results indicated that there was likely to be insufficient contingency in the links. Furthermore, the OMA agreed well with the manufacturer’s quoted typical value of -4.6 dBm [72].

In order to verify the validity of the data, the SerDes on the source cards were configured to repeatedly send and capture the automated PRBS test signal and the number of word errors counted by the FPGA. A variable optical attenuation was introduced between the transmitter and receiver, and the received signal retransmitted to the source cards (shown in figure 3.13). This test was performed for greater than 10^{11} bits of data for a range of values of attenuation, and the bit error rates estimated from the number of errors relative to the amount of data sent. The results of this test and the quoted tolerances of the components are indicated in figure 3.14.

The bit error rates can not be calculated by simply counting the error flags because the SerDes do not indicate the number of bit errors per received word, only whether the word was received correctly or not. If it is assumed that the

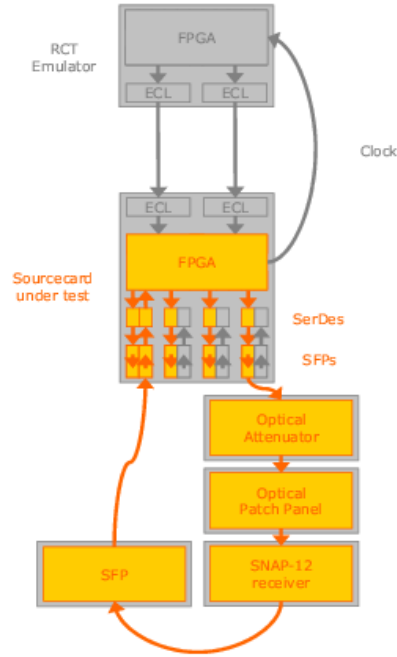


Figure 3.13: Schematic layout of the indirect test of the source card optical links. The paths in orange indicate the data path used in the test. The link quality was tested by performing bit error measurements as a function of signal attenuation.

number of bit errors in a word is given by a binomial distribution, then the probability of no bit errors occurring in a word is $P(r = 0) = (1 - p)^n$ where n is the word size and p is the probability of a bit error. The probability of at least one bit error is, therefore, $P(r > 0) = 1 - (1 - p)^n$, which may be estimated as the fraction of words received incorrectly, E . The relationship may then be inverted to estimate the bit error rate as $p \approx 1 - (1 - E)^{1/n}$.

Qualitative tests, in which the attenuation was adjusted until errors were observed, were performed for several different batches of SNAP-12 receivers and some batch-to-batch variation was observed: for the worst, errors began to appear at an attenuation of 16.75 ± 0.05 dB (corresponding to an OMA of -19.86 dBm) and for all batches there was a turn-on in the error rate at 17.30 ± 0.05 dB (or an OMA of -20.37 dBm). The channel-to-channel variation within a receiver was limited, with the point where errors were first seen showing a standard deviation of approximately 0.2 dB. The very large buffer region between the quoted tolerances of the components and the observed error turn-on suggested that there was sufficient contingency in the link to successfully use the two parts together.

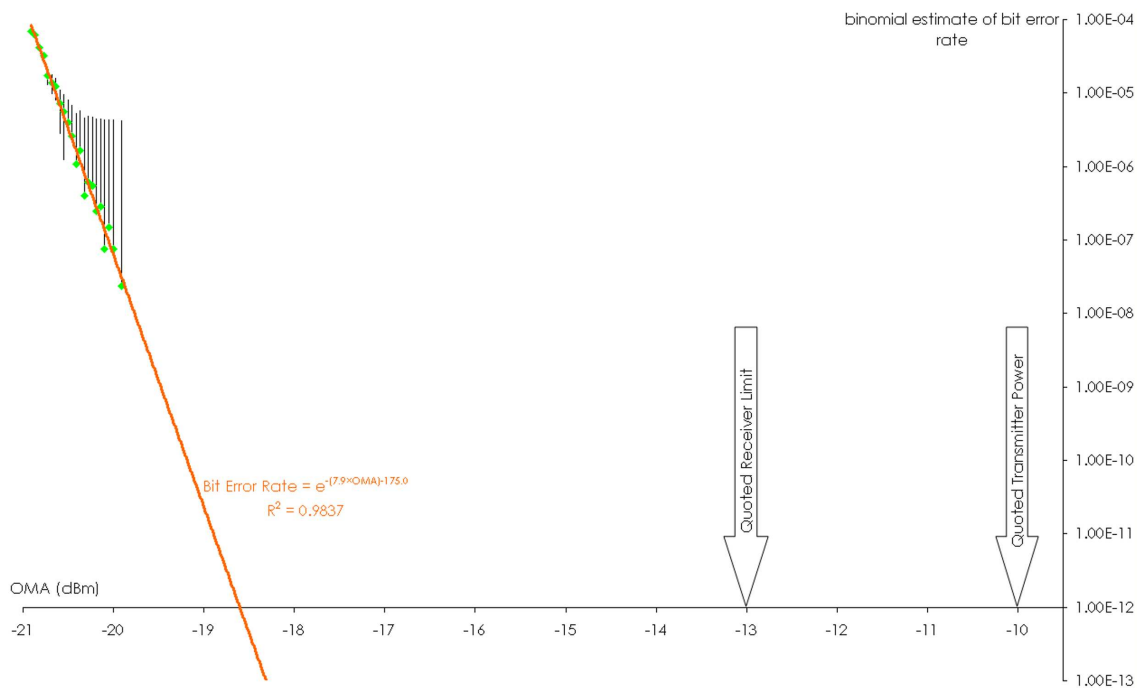


Figure 3.14: Extrapolated BER as a function of the fibre attenuation of the optical links expressed in terms of the Optical Modulation Amplitude. Above the point at -20.4 dBm the error bars drop below the axis due to limited statistics and are omitted here for clarity.

Once it had been shown that the parts were suitable for use and that the bit-error rate of an SFP/SNAP-12 pair was of the same order as the SFP-to-SFP tests, it was deemed sufficient that, for production testing, each board may be verified by SFP-to-SFP tests rather than requiring a separate test fixture with a SNAP-12 receiver.

During production testing of the optical links, it was necessary to test data transmission through the FPGA-Serializer-SFP chain simultaneously for all four channels: this was achieved by the inclusion of four “pre-qualified” source cards (as indicated in figure 3.15), the four pre-qualified boards providing four receiver channels and thereby allowing the simultaneous capture of all transmitted data.

3.3.5 TTCrx and QPLL tests

In order that the clocks of all systems are coordinated, the variation in clocking frequency across each subsystem must be tightly controlled. Although the

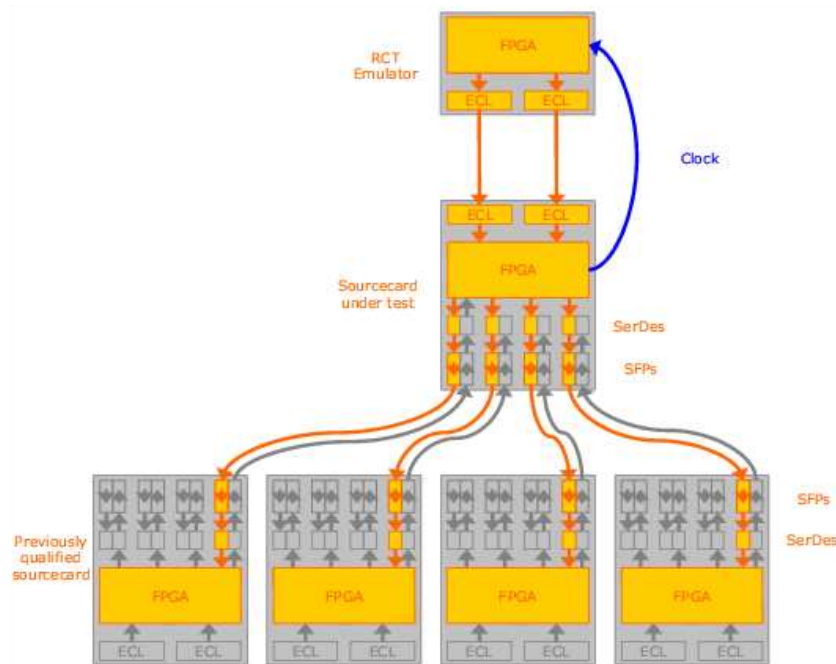


Figure 3.15: Test configuration used for the production testing of the optical links.

machine clock frequency is 40.08 MHz, the frequency within each clock domain is controlled by the use of a QPLL with a nominal locking range of 40.0749–40.0823 MHz. The locking range is, however, affected by the manufacturing quality of the crystals, the drive strength of the QPLL and the effects of parasitic capacitance around the crystal. Despite following the recommended design precautions, it was necessary to quantify the actual locking range and to verify that the source cards could lock to the TTC system. By driving the TTC system from an extremely high-precision clock generator, it was possible to iterate across (and beyond) the entire locking range in 100 Hz steps. Direct measurements showed that the source card QPLL locked over an input range of 40.0728–40.0819 ± 0.0001 MHz indicating that, although slightly lower than expected, the range was acceptable and sufficiently wide to lock to the machine clock. For the production testing of the cards, the extremely high-precision clock generator was not available and, instead, a programmable pattern generator (with a precision of 10 kHz) was used: the nominal signal frequency of 40.08 MHz being sufficiently accurate to be locked to.

While the 2 prototype and 8 pre-production source cards showed no problems

with their TTC system, numerous errors were observed in the production batches, namely:

- Failure of the QPLL to lock to the TTC clock
- Failure of the TTCrx **I²C** (Inter-Integrated Circuit) communication
- Failure of the TTCrx to activate
- Failure of the TTCrx to enter the correct boot mode
- Failure of the FPGA **DCM** (Digital Clock Manager) to lock to the TTC clock

Furthermore, because of the way the TTC ID is shared between the TTCrx and the FPGA (where it is used as the board hardware ID) with certain failures of the TTCrx, the hardware ID was also unreadable; this meaning that, when multiple boards were present in the system, the board which failed could not necessarily be identified. Two solutions were put in place to resolve this problem: rather than raising an exception when a TTCrx failed, the software was modified so as to evaluate all of the remaining cards and to so deduce which ID was expected, but not found. This method could not, however, handle all possible cases and the system was further modified to store the board's hardware identifier in the EEPROM's usercode space, thereby separating the identifiability of the hardware from the TTCrx system.

While these precautions did not solve the underlying problems, they did allow for consistent debugging. What initially appeared to be many different failures were shown, in fact, to be different manifestations of problems with the TTCrx: the failure of the QPLL or FPGA DCM to lock onto the TTC clock was due to poor-quality or non-existent clock signals from the TTCrx; this was demonstrated to be closely related to the TTCrx failing to enter the correct boot mode, whilst failure of the **I²C** bus was shown to be associated with the failure of the TTCrx to activate. Various explanations of the TTCrx problems were considered, including the possibility of a short circuit on the address line incorrectly setting

the boot mode, a firmware fault causing the same problem, an incorrect or floating input on one of the reset pins, slow power-up causing the device to boot incorrectly, noise or a marginal voltage on the power supply or a poor-quality differential input signal. These causes were successively excluded to the point where it was decided that the only remaining sensible option was replacement of the TTCrx itself: whilst this improved the stability of many boards, it did not eliminate the errors on all boards, this indicating that the problem was, most likely, external to the TTCrx and that some components were less sensitive to this problem than others.

The cause of the errors was later shown to be due to a poor-quality differential input signal: this was missed in the initial tests as the signal was measured differentially and showed a “clean” signal whose peak-to peak voltage (although lower than expected) was well within the range accepted by the TTCrx. Later tests revealed that there was a fault in the pin diode (whether present from manufacture or due to the assembly process) causing the diode to drive only a single member of the differential pair; this fault was hidden by the original choice of measurement technique, but was of significance in driving the TTCrx. Replacement of the pin diode was sufficient to resolve the problem.

3.4 Integration of the GCT Source Cards

3.4.1 Integration of the GCT Source Cards with the RCT

Upon installation of the GCT source cards and their interconnection with the RCT, it was necessary to validate both the connections themselves and the timing relationship between the two systems. An automated test was devised, based on the already-proven RCT to source card timing validation test. A 64 bunch-crossing (128 half bunch-crossing) test pattern was loaded into the test registers at the output of the RCT and transmission of the patterns triggered by a TTC L1-reset. The source cards were configured to capture the packet based on the same

trigger and, upon triggering, the data was read out and compared in software. The test was performed for each available setting of the TTCrx fine delay and the number of word errors observed on each channel related to the delay: figure 3.16 shows a selection of such plots. The transmitted pattern was chosen such that the transitions of neighbouring bits were uncorrelated and that, for each bit, the total time spent high was approximately equal to the time spent low. The quality of transmission and reception of data is indicated by the width of the regions where the error count drops to zero. Figure 3.16(a) indicates that the source card in question has two correctly functioning inputs with the word error rate dropping to zero for a wide range of TTC fine delay settings. Figure 3.16(b) indicates that the source card has one good input, the error rate dropping to zero as before, and one bad channel where the word error rate never falls to zero. Noting that the word error rate is around 50% for the period where timing is expected to be good indicates the presence of a “stuck” bit; those words where the bit is stuck to the “correct” value being registered as correct and those where the bit is stuck to the “incorrect” value being registered as wrong. The figure of approximately 50% reflects the distribution of bits on the damaged channel; this channel being damaged in the process of installation. In figure 3.16(c), the word error rate drops as expected but never reaches zero on either channel. The occurrence of errors on both channels and the error rate not being around 50%, excludes the possibility of stuck bits and the consistency of the error rate excludes the possibility of random errors. The problem was traced to a fault in the pattern injection subsystem of the RCT hardware and was remedied by replacing the faulty circuit board. Figure 3.16(d) indicates two distinct faults: the signal in red never drops to zero (c.f. figure 3.16(b)) in this case due to a poor connection caused by an unscrewed connector. More significantly, neither channel has a wide, stable capture window; repeating this test produced a different set of timings on each occasion. The cause of this problem was found to be a timing contention in the source card firmware.

The faults diagnosed by this method may be summarized by

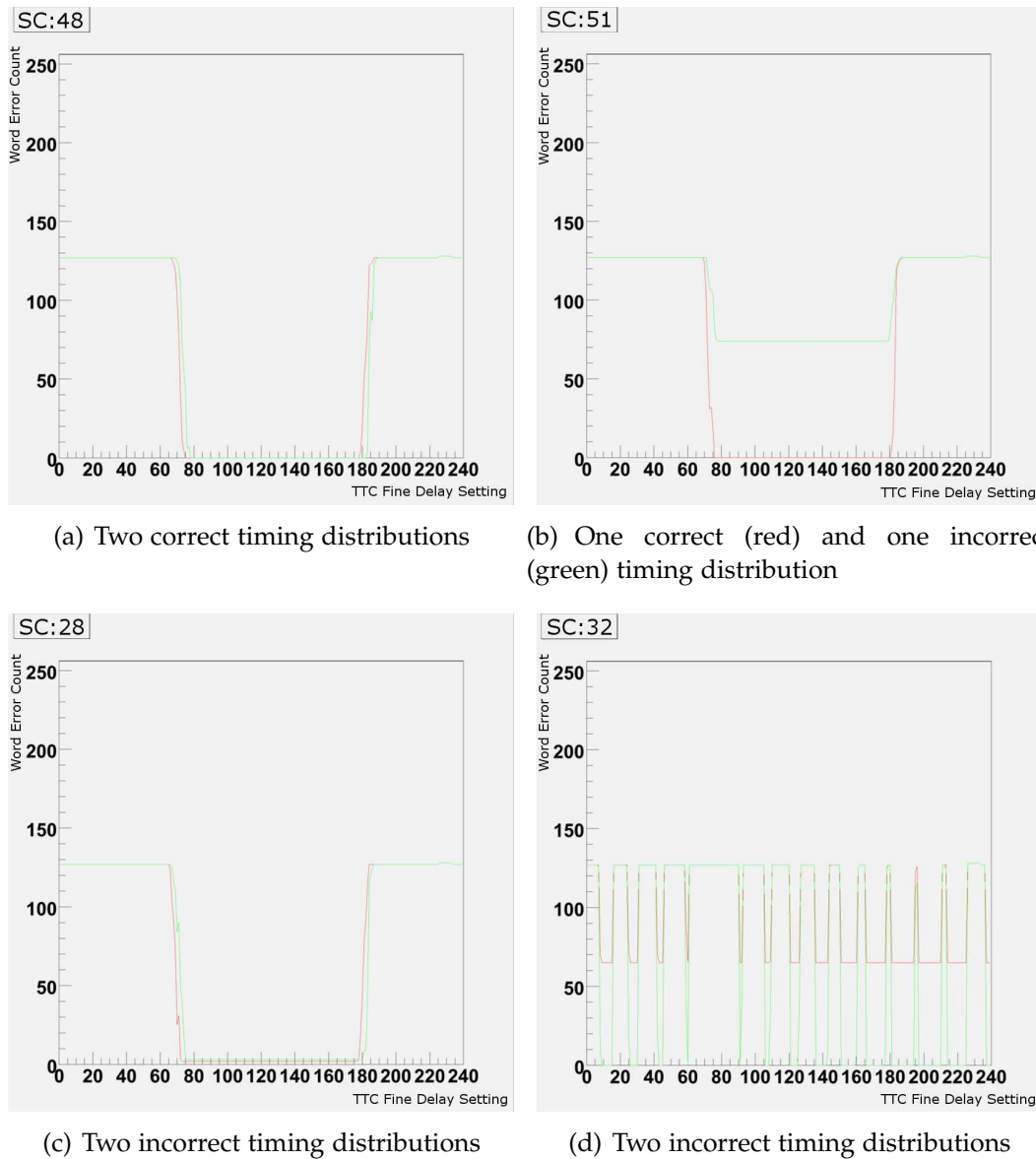


Figure 3.16: Four plots of word error rate against TTCrx fine-delay setting made during the RCT-source card integration tests. A separate plot was made for each source card in the system. The red and green lines separately represent the two inputs of the source card.

- cards which were unable to find a good fine-delay parameter with which to lock to the RCT (firmware error)
- loose connectors
- connectors damaged during installation
- electrical faults in the RCT

Not indicated in these plots, but of equal importance, was the discovery of errors in documentation (primarily related to undocumented changes made to the RCT layout) and software errors in both the RCT and source card systems.

A further complication arose when it was later discovered that the heavy duty cables were suffering damage each time they were being connected or disconnected. This resulted in intermittent bit errors which would disappear upon replacement of a board, but which could later reappear if the cable was moved slightly.

3.4.2 Integration of the GCT Source Cards with the GCT Leaf Cards

The initial test of compatibility between the source card and the leaf card, required that the leaf card should correctly capture a counting pattern of fixed length transmitted by the source card; it was found that the leaf card was not capable of doing so. It was shown that the pattern could be correctly captured when transmitted from a source card to a second, independently clocked, source card; this indicating that both the data and the clock signal were, in principle, extractable from the serial data. The optical components had already been demonstrated to be compatible, therefore it was considered that the most likely cause of the error was either a failure of the Multi-Gigabit Transceivers on the leaf card, or of the pairing between the source card's serializers and the leaf card's deserializers. The source cards were designed to run the optical links synchronously

with the TTC system but included a standalone oscillator for testing the links when a TTC clock was unavailable; when the optical links failed to operate in the standard, TTC-synchronous, mode of operation, the standalone mode was also tested and shown to work correctly. This was strongly indicative that the cause of the problem was the use of the CERN TTCrx and QPLL to clock the source card TLK2501 SerDes. The positive result of the source card to source card tests, where the source card to leaf card test failed, can be attributed to the use of dedicated SerDes on the source card and integrated SerDes (MGTs) on the leaf card; the former have superior jitter acceptance and clock recovery to the latter.

The existing standalone oscillator operated at 80.00 MHz compared to the 80.16 MHz TTC clock and was, therefore, not suitable for use in the trigger system, as the output rate from the source card would be lower than the input rate. Instead, the oscillators on the source cards were replaced with 100.00 MHz components and a four-phase clock-bridge added at the output of the FPGA. When no valid data was available for transmission, the clock-bridge signalled to the serializers that “comma” characters^x should be sent.

Upon installation of both the source cards and the leaf cards in the underground counting rooms, it was necessary to verify the correct interconnection of the two. Four possible problems were identified:

- There are 252 optical channels on the source cards; these are routed to the patch panel using 12-fibre bundles. Manual connection of 252 fibres at both ends introduces the risk of errors.
- The source cards are located in racks on the floor above the rest of the GCT and the fibres can not, therefore, be manually traced along their entire length.
- The fibres from the source cards are plugged into the back of the patch-panel, a location which is not easily accessible.

^xplace-holder characters which are understood to indicate a lack of data, but whose presence maintains synchronization of the serial links.

- At the output from the patch-panel, the 252 optical channels must be manually routed onto 12-channel adapter fibres, again introducing the risk of errors.

The source cards optical links were configured to send a static test-pattern, consisting of the source card logical ID (8bits) concatenated with the optical link number (8bits), running from 1 (indicating the top optical link) to 4 (the bottom optical link). The leaf cards captured this pattern and the connections were examined. Due to the staged installation of hardware, the first attempt at connection resulted in only 61 of the 234 source card to leaf card fibres^{xi} being routed correctly. The fibres were unplugged from the patch-panel and systematically reinstalled; this resulted in all fibres being correctly routed.

^{xi}The remaining 18 fibres go to the muon system and were not part of this test

Chapter 4

The Global Calorimeter Trigger Schema

“Design can be art. Design can be aesthetics.
Design is so simple, that’s why it is so complicated.”

- Paul Rand (1914 - 1996)

4.1 Data-Flow Architecture

Rather than the large, monolithic boards that characterize much of the CMS trigger, the Global Calorimeter Trigger was designed to consist of many simpler, highly interconnected, modular parts. This posed many questions as to the optimal choice of data-flow architecture.

The output from each of the 18 RCT crates is on six 34-bit wide SCSI cables and it was intended that each RCT crate should be handled by three source cards. The bit-content of the six outputs is summarized in figure 4.1.

As the data is transmitted at double-data rate (DDR), each clock cycle corresponds to two data cycles: labelled “cycle 0”, which corresponds to the leading clock edge, and “cycle 1”. To distinguish these phases, a phase marker is embedded into the data stream, being high on “cycle 1” and low on “cycle 0”, except

for the first bunch-crossing of each orbit, where the marker is held high on both cycles.

The data can be broadly split into five categories:

- MIP candidate/Quiet region information (coloured red in figures 4.1 - 4.4)
- e/γ candidates (coloured in shades of blue and purple in figures 4.1 - 4.4)
- Forward hadronic calorimeter (HF) energies (coloured peach and orange in figures 4.1 - 4.4)
- Regional energy sums (coloured in shades of green and brown in figures 4.1 - 4.4)
- The phase marker (coloured black in figures 4.1 - 4.4)

This colour scheme is consistent throughout figures 4.1 - 4.4.

The e/γ candidates and MIP candidate/Quiet region information are distributed across the first two cables (figures 4.1(a) and 4.1(b)). On each cycle the data consists of 14 bits of MIP (Quiet) information and, for each of the 2 isolated electron candidates and 2 non-isolated electron candidates, there are 6 bits of “rank” (a measure of energy) and 4 bits identifying location. The HF energies are split across the third and fourth cables (figures 4.1(c) and 4.1(d)) and, on each cycle, consist of 8 bits of energy and a single “quiet” bit for each of four regions. The regional energy sums are distributed across the fourth, fifth and sixth cables (figures 4.1(d), figures 4.1(e) and 4.1(f)) and, on each cycle, each of the receiver cards (seven per RCT crate) sends 10 bits of energy, an overflow bit and a tau veto bit.

As the implementation of GCT algorithms is distributed across 8 leaf cards (2 for e/γ and 6 for jets) and the entirely separate muon system, and, as there is limited communication between the components implementing the different algorithms, it is necessary that the source cards pre-sort the data such that all

output 1				output 2			
Pin Pair	Bit	Cycle 0	Cycle 1	Pin Pair	Bit	Cycle 0	Cycle 1
1,2	33	0 Isolated e/v rank <0>	2 Isolated e/v rank <0>	1,2	33	0 Non-Isolated e/v rank <0>	2 Non-Isolated e/v rank <0>
3,4	31	0 Isolated e/v rank <1>	2 Isolated e/v rank <1>	3,4	31	0 Non-Isolated e/v rank <1>	2 Non-Isolated e/v rank <1>
5,6	29	0 Isolated e/v rank <2>	2 Isolated e/v rank <2>	5,6	29	0 Non-Isolated e/v rank <2>	2 Non-Isolated e/v rank <2>
7,8	27	0 Isolated e/v rank <3>	2 Isolated e/v rank <3>	7,8	27	0 Non-Isolated e/v rank <3>	2 Non-Isolated e/v rank <3>
9,10	25	0 Isolated e/v rank <4>	2 Isolated e/v rank <4>	9,10	25	0 Non-Isolated e/v rank <4>	2 Non-Isolated e/v rank <4>
11,12	23	0 Isolated e/v rank <5>	2 Isolated e/v rank <5>	11,12	23	0 Non-Isolated e/v rank <5>	2 Non-Isolated e/v rank <5>
13,14	21	0 Isolated e/v region id	2 Isolated e/v region id	13,14	21	0 Non-Isolated e/v region id	2 Non-Isolated e/v region id
15,16	19	0 Isolated e/v card id <0>	2 Isolated e/v card id <0>	15,16	19	0 Non-Isolated e/v card id <0>	2 Non-Isolated e/v card id <0>
17,18	17	0 Isolated e/v card id <1>	2 Isolated e/v card id <1>	17,18	17	0 Non-Isolated e/v card id <1>	2 Non-Isolated e/v card id <1>
19,20	15	0 Isolated e/v card id <2>	2 Isolated e/v card id <2>	19,20	15	0 Non-Isolated e/v card id <2>	2 Non-Isolated e/v card id <2>
21,22	13	unused	unused	21,22	13	unused	unused
23,24	11	1 Isolated e/v rank <0>	3 Isolated e/v rank <0>	23,24	11	1 Non-Isolated e/v rank <0>	3 Non-Isolated e/v rank <0>
25,26	9	1 Isolated e/v rank <1>	3 Isolated e/v rank <1>	25,26	9	1 Non-Isolated e/v rank <1>	3 Non-Isolated e/v rank <1>
27,28	7	1 Isolated e/v rank <2>	3 Isolated e/v rank <2>	27,28	7	1 Non-Isolated e/v rank <2>	3 Non-Isolated e/v rank <2>
29,30	5	1 Isolated e/v rank <3>	3 Isolated e/v rank <3>	29,30	5	1 Non-Isolated e/v rank <3>	3 Non-Isolated e/v rank <3>
31,32	3	1 Isolated e/v rank <4>	3 Isolated e/v rank <4>	31,32	3	1 Non-Isolated e/v rank <4>	3 Non-Isolated e/v rank <4>
33,34	1	1 Isolated e/v rank <5>	3 Isolated e/v rank <5>	33,34	1	1 Non-Isolated e/v rank <5>	3 Non-Isolated e/v rank <5>
35,36	32	1 Isolated e/v region id	3 Isolated e/v region id	35,36	32	1 Non-Isolated e/v region id	3 Non-Isolated e/v region id
37,38	30	1 Isolated e/v card id <0>	3 Isolated e/v card id <0>	37,38	30	1 Non-Isolated e/v card id <0>	3 Non-Isolated e/v card id <0>
39,40	28	1 Isolated e/v card id <1>	3 Isolated e/v card id <1>	39,40	28	1 Non-Isolated e/v card id <1>	3 Non-Isolated e/v card id <1>
41,42	26	1 Isolated e/v card id <2>	3 Isolated e/v card id <2>	41,42	26	1 Non-Isolated e/v card id <2>	3 Non-Isolated e/v card id <2>
43,44	24	unused	unused	43,44	24	unused	unused
45,46	22	0 receiver card MIP <0>	0 receiver card quiet <0>	45,46	22	4 receiver card MIP <0>	4 receiver card quiet <0>
47,48	20	0 receiver card MIP <1>	0 receiver card quiet <1>	47,48	20	4 receiver card MIP <1>	4 receiver card quiet <1>
49,50	18	1 receiver card MIP <0>	1 receiver card quiet <0>	49,50	18	5 receiver card MIP <0>	5 receiver card quiet <0>
51,52	16	1 receiver card MIP <1>	1 receiver card quiet <1>	51,52	16	5 receiver card MIP <1>	5 receiver card quiet <1>
53,54	14	2 receiver card MIP <0>	2 receiver card quiet <0>	53,54	14	6 receiver card MIP <0>	6 receiver card quiet <0>
55,56	12	2 receiver card MIP <1>	2 receiver card quiet <1>	55,56	12	6 receiver card MIP <1>	6 receiver card quiet <1>
57,58	10	3 receiver card MIP <0>	3 receiver card quiet <0>	57,58	10	unused	unused
59,60	8	3 receiver card MIP <1>	3 receiver card quiet <1>	59,60	8	unused	unused
61,62	6	unused	unused	61,62	6	unused	unused
63,64	4	unused	unused	63,64	4	unused	unused
65,66	2	unused	unused	65,66	2	unused	unused
67,68	0	BC0	'1'	67,68	0	BC0	'1'

(a) RCT output 1

(b) RCT output 2

output 3				output 4			
Pin Pair	Bit	Cycle 0	Cycle 1	Pin Pair	Bit	Cycle 0	Cycle 1
1,2	33	HF:0 η :0 <1>	HF:0 η :2 <1>	1,2	33	RC:5 Region:0 <0>	RC:5 Region:1 <0>
3,4	31	HF:0 η :1 <1>	HF:0 η :3 <1>	3,4	31	RC:5 Region:0 <1>	RC:5 Region:1 <1>
5,6	29	HF:0 η :0 <2>	HF:0 η :2 <2>	5,6	29	RC:5 Region:0 <2>	RC:5 Region:1 <2>
7,8	27	HF:0 η :1 <2>	HF:0 η :3 <2>	7,8	27	RC:5 Region:0 <3>	RC:5 Region:1 <3>
9,10	25	HF:0 η :0 <3>	HF:0 η :2 <3>	9,10	25	RC:5 Region:0 <4>	RC:5 Region:1 <4>
11,12	23	HF:0 η :1 <3>	HF:0 η :3 <3>	11,12	23	RC:5 Region:0 <5>	RC:5 Region:1 <5>
13,14	21	HF:0 η :0 <4>	HF:0 η :2 <4>	13,14	21	RC:5 Region:0 <6>	RC:5 Region:1 <6>
15,16	19	HF:0 η :1 <4>	HF:0 η :3 <4>	15,16	19	RC:5 Region:0 <7>	RC:5 Region:1 <7>
17,18	17	HF:0 η :0 <5>	HF:0 η :2 <5>	17,18	17	RC:5 Region:0 <8>	RC:5 Region:1 <8>
19,20	15	HF:0 η :1 <5>	HF:0 η :3 <5>	19,20	15	RC:5 Region:0 <9>	RC:5 Region:1 <9>
21,22	13	HF:0 η :0 <6>	HF:0 η :2 <6>	21,22	13	RC:5 Region:0 overflow	RC:5 Region:1 overflow
23,24	11	HF:0 η :1 <6>	HF:0 η :3 <6>	23,24	11	RC:5 Region:0 tau	RC:5 Region:1 tau
25,26	9	HF:0 η :0 <7>	HF:0 η :2 <7>	25,26	9	RC:6 Region:0 <0>	RC:6 Region:1 <0>
27,28	7	HF:0 η :1 <7>	HF:0 η :3 <7>	27,28	7	RC:6 Region:0 <1>	RC:6 Region:1 <1>
29,30	5	HF:1 η :0 <0>	HF:1 η :2 <0>	29,30	5	RC:6 Region:0 <2>	RC:6 Region:1 <2>
31,32	3	unused	unused	31,32	3	unused	unused
33,34	1	unused	unused	33,34	1	unused	unused
35,36	32	HF:1 η :1 <7>	HF:1 η :3 <7>	35,36	32	HF:0 η :1 <0>	HF:0 η :3 <0>
37,38	30	HF:1 η :0 <7>	HF:1 η :2 <7>	37,38	30	HF:0 η :0 <0>	HF:0 η :2 <0>
39,40	28	HF:1 η :1 <6>	HF:1 η :3 <6>	39,40	28	HF:1 η :1 quiet	HF:1 η :3 quiet
41,42	26	HF:1 η :0 <6>	HF:1 η :2 <6>	41,42	26	HF:1 η :0 quiet	HF:1 η :2 quiet
43,44	24	HF:1 η :1 <5>	HF:1 η :3 <5>	43,44	24	HF:0 η :1 quiet	HF:0 η :3 quiet
45,46	22	HF:1 η :0 <5>	HF:1 η :2 <5>	45,46	22	HF:0 η :0 quiet	HF:0 η :2 quiet
47,48	20	HF:1 η :1 <4>	HF:1 η :3 <4>	47,48	20	RC:6 Region:0 tau	RC:6 Region:1 tau
49,50	18	HF:1 η :0 <4>	HF:1 η :2 <4>	49,50	18	RC:6 Region:0 overflow	RC:6 Region:1 overflow
51,52	16	HF:1 η :1 <3>	HF:1 η :3 <3>	51,52	16	RC:6 Region:0 <9>	RC:6 Region:1 <9>
53,54	14	HF:1 η :0 <3>	HF:1 η :2 <3>	53,54	14	RC:6 Region:0 <8>	RC:6 Region:1 <8>
55,56	12	HF:1 η :1 <2>	HF:1 η :3 <2>	55,56	12	RC:6 Region:0 <7>	RC:6 Region:1 <7>
57,58	10	HF:1 η :0 <2>	HF:1 η :2 <2>	57,58	10	RC:6 Region:0 <6>	RC:6 Region:1 <6>
59,60	8	HF:1 η :1 <1>	HF:1 η :3 <1>	59,60	8	RC:6 Region:0 <5>	RC:6 Region:1 <5>
61,62	6	HF:1 η :0 <1>	HF:1 η :2 <1>	61,62	6	RC:6 Region:0 <4>	RC:6 Region:1 <4>
63,64	4	HF:1 η :1 <0>	HF:1 η :3 <0>	63,64	4	RC:6 Region:0 <3>	RC:6 Region:1 <3>
65,66	2	unused	unused	65,66	2	unused	unused
67,68	0	BC0	'1'	67,68	0	BC0	'1'

(c) RCT output 3

(d) RCT output 4

output 5				output 6			
Pin Pair	Bit	Cycle 0	Cycle 1	Pin Pair	Bit	Cycle 0	Cycle 1
1,2	33	RC:0 Region:0 <0>	RC:0 Region:1 <0>	1,2	33	RC:2 Region:0 <6>	RC:2 Region:1 <6>
3,4	31	RC:0 Region:0 <1>	RC:0 Region:1 <1>	3,4	31	RC:2 Region:0 <7>	RC:2 Region:1 <7>
5,6	29	RC:0 Region:0 <2>	RC:0 Region:1 <2>	5,6	29	RC:2 Region:0 <8>	RC:2 Region:1 <8>
7,8	27	RC:0 Region:0 <3>	RC:0 Region:1 <3>	7,8	27	RC:2 Region:0 <9>	RC:2 Region:1 <9>
9,10	25	RC:0 Region:0 <4>	RC:0 Region:1 <4>	9,10	25	RC:2 Region:0 overflow	RC:2 Region:1 overflow
11,12	23	RC:0 Region:0 <5>	RC:0 Region:1 <5>	11,12	23	RC:2 Region:0 tau	RC:2 Region:1 tau
13,14	21	RC:0 Region:0 <6>	RC:0 Region:1 <6>	13,14	21	RC:3 Region:0 <0>	RC:3 Region:1 <0>
15,16	19	RC:0 Region:0 <7>	RC:0 Region:1 <7>	15,16	19	RC:3 Region:0 <1>	RC:3 Region:1 <1>
17,18	17	RC:0 Region:0 <8>	RC:0 Region:1 <8>	17,18	17	RC:3 Region:0 <2>	RC:3 Region:1 <2>
19,20	15	RC:0 Region:0 <9>	RC:0 Region:1 <9>	19,20	15	RC:3 Region:0 <3>	RC:3 Region:1 <3>
21,22	13	RC:0 Region:0 overflow	RC:0 Region:1 overflow	21,22	13	RC:3 Region:0 <4>	RC:3 Region:1 <4>
23,24	11	RC:0 Region:0 tau	RC:0 Region:1 tau	23,24	11	RC:3 Region:0 <5>	RC:3 Region:1 <5>
25,26	9	RC:1 Region:0 <0>	RC:1 Region:1 <0>	25,26	9	RC:3 Region:0 <6>	RC:3 Region:1 <6>
27,28	7	RC:1 Region:0 <1>	RC:1 Region:1 <1>	27,28	7	RC:3 Region:0 <7>	RC:3 Region:1 <7>
29,30	5	RC:1 Region:0 <2>	RC:1 Region:1 <2>	29,30	5	RC:3 Region:0 <8>	RC:3 Region:1 <8>
31,32	3	unused	unused	31,32	3	unused	unused
33,34	1	unused	unused	33,34	1	unused	unused
35,36	32	RC:2 Region:0 <5>	RC:2 Region:1 <5>	35,36	32	RC:4 Region:0 tau	RC:4 Region:1 tau
37,38	30	RC:2 Region:0 <4>	RC:2 Region:1 <4>	37,38	30	RC:4 Region:0 overflow	RC:4 Region:1 overflow
39,40	28	RC:2 Region:0 <3>	RC:2 Region:1 <3>	39,40	28	RC:4 Region:0 <9>	RC:4 Region:1 <9>
41,42	26	RC:2 Region:0 <2>	RC:2 Region:1 <2>	41,42	26	RC:4 Region:0 <8>	RC:4 Region:1 <8>
43,44	24	RC:2 Region:0 <1>	RC:2 Region:1 <1>	43,44	24	RC:4 Region:0 <7>	RC:4 Region:1 <7>
45,46	22	RC:2 Region:0 <0>	RC:2 Region:1 <0>	45,46	22	RC:4 Region:0 <6>	RC:4 Region:1 <6>
47,48	20	RC:1 Region:0 tau	RC:1 Region:1 tau	47,48	20	RC:4 Region:0 <5>	RC:4 Region:1 <5>
49,50	18	RC:1 Region:0 overflow	RC:1 Region:1 overflow	49,50	18	RC:4 Region:0 <4>	RC:4 Region:1 <4>
51,52	16	RC:1 Region:0 <9>	RC:1 Region:1 <9>	51,52	16	RC:4 Region:0 <3>	RC:4 Region:1 <3>
53,54	14	RC:1 Region:0 <8>	RC:1 Region:1 <8>	53,54	14	RC:4 Region:0 <2>	RC:4 Region:1 <2>
55,56	12	RC:1 Region:0 <7>	RC:1 Region:1 <7>	55,56	12	RC:4 Region:0 <1>	RC:4 Region:1 <1>
57,58	10	RC:1 Region:0 <6>	RC:1 Region:1 <6>	57,58	10	RC:4 Region:0 <0>	RC:4 Region:1 <0>
59,60	8	RC:1 Region:0 <5>	RC:1 Region:1 <5>	59,60	8	RC:3 Region:0 tau	RC:3 Region:1 tau
61,62	6	RC:1 Region:0 <4>	RC:1 Region:1 <4>	61,62	6	RC:3 Region:0 overflow	RC:3 Region:1 overflow
63,64	4	RC:1 Region:0 <3>	RC:1 Region:1 <3>	63,64	4	RC:3 Region:0 <9>	RC:3 Region:1 <9>
65,66	2	unused	unused	65,66	2	unused	unused
67,68	0	BC0	'1'	67,68	0	BC0	'1'

(e) RCT output 5

(f) RCT output 6

Figure 4.1: The bit structure of the six RCT outputs. The colour scheme indicates blocks of data of the same type and is carried over to figures 4.2 - 4.4. Adapted from [66].

data of similar types is routed on as few optical fibres as possible. The sorting and separation of data is especially important in the separation of e/γ data from muon data and was achieved using the scheme shown in figure 4.2, which shows the bitwise output on each of the four optical links on each clock cycle. Under this scheme, the MIP- and Quiet- bits are placed onto a separate optical fibre and the e/γ data distributed as evenly as possible across the remaining three outputs.

The other four outputs from each RCT crate, each containing broadly similar kinds of data, were originally intended to be forwarded by the source cards and the regional ϕ -boundaries handled by interconnections between leaf cards and the η -boundary handled by the wheel and concentrator cards. Later, however, it was imposed that the algorithm's implementation should be precisely that specified in the CMS trigger TDR [54]. This required modification such that the RCT regions on either side of the $\eta = 0$ boundary were duplicated and forwarded to the leaf cards representing both ends of the calorimeter. The handling of the

Cycle 0				
Bit	Output 0	Output 1	Output 2	Output 3
0	0 rec card MIP <0>	0 Isolated e/v rank <0>	0 Non-Isolated e/v rank <0>	1 Non-Isolated e/v rank <0>
1	0 rec card MIP <1>	0 Isolated e/v rank <1>	0 Non-Isolated e/v rank <1>	1 Non-Isolated e/v rank <1>
2	1 rec card MIP <0>	0 Isolated e/v rank <2>	0 Non-Isolated e/v rank <2>	1 Non-Isolated e/v rank <2>
3	1 rec card MIP <1>	0 Isolated e/v rank <3>	0 Non-Isolated e/v rank <3>	1 Non-Isolated e/v rank <3>
4	2 rec card MIP <0>	0 Isolated e/v rank <4>	0 Non-Isolated e/v rank <4>	1 Non-Isolated e/v rank <4>
5	2 rec card MIP <1>	0 Isolated e/v rank <5>	0 Non-Isolated e/v rank <5>	1 Non-Isolated e/v rank <5>
6	3 rec card MIP <0>	0 Isolated e/v region id	0 Non-Isolated e/v region id	1 Non-Isolated e/v region id
7	3 rec card MIP <1>	0 Isolated e/v card id <0>	0 Non-Isolated e/v card id <0>	1 Non-Isolated e/v card id <0>
8	4 rec card MIP <0>	0 Isolated e/v card id <1>	0 Non-Isolated e/v card id <1>	1 Non-Isolated e/v card id <1>
9	4 rec card MIP <1>	0 Isolated e/v card id <2>	0 Non-Isolated e/v card id <2>	1 Non-Isolated e/v card id <2>
10	5 rec card MIP <0>	1 Isolated e/v rank <0>	1 Isolated e/v rank <3>	1 Isolated e/v card id <0>
11	5 rec card MIP <1>	1 Isolated e/v rank <1>	1 Isolated e/v rank <4>	1 Isolated e/v card id <1>
12	6 rec card MIP <0>	1 Isolated e/v rank <2>	1 Isolated e/v rank <5>	1 Isolated e/v card id <2>
13	6 rec card MIP <1>	unused	1 Isolated e/v region id	unused
14	unused	unused	unused	unused
15	BC0	BC0	BC0	BC0

Cycle 1				
Bit	Output 0	Output 1	Output 2	Output 3
0	0 rec card quiet <0>	2 Isolated e/v rank <0>	2 Non-Isolated e/v rank <0>	3 Non-Isolated e/v rank <0>
1	0 rec card quiet <1>	2 Isolated e/v rank <1>	2 Non-Isolated e/v rank <1>	3 Non-Isolated e/v rank <1>
2	1 rec card quiet <0>	2 Isolated e/v rank <2>	2 Non-Isolated e/v rank <2>	3 Non-Isolated e/v rank <2>
3	1 rec card quiet <1>	2 Isolated e/v rank <3>	2 Non-Isolated e/v rank <3>	3 Non-Isolated e/v rank <3>
4	2 rec card quiet <0>	2 Isolated e/v rank <4>	2 Non-Isolated e/v rank <4>	3 Non-Isolated e/v rank <4>
5	2 rec card quiet <1>	2 Isolated e/v rank <5>	2 Non-Isolated e/v rank <5>	3 Non-Isolated e/v rank <5>
6	3 rec card quiet <0>	2 Isolated e/v region id	2 Non-Isolated e/v region id	3 Non-Isolated e/v region id
7	3 rec card quiet <1>	2 Isolated e/v card id <0>	2 Non-Isolated e/v card id <0>	3 Non-Isolated e/v card id <0>
8	4 rec card quiet <0>	2 Isolated e/v card id <1>	2 Non-Isolated e/v card id <1>	3 Non-Isolated e/v card id <1>
9	4 rec card quiet <1>	2 Isolated e/v card id <2>	2 Non-Isolated e/v card id <2>	3 Non-Isolated e/v card id <2>
10	5 rec card quiet <0>	3 Isolated e/v rank <0>	3 Isolated e/v rank <3>	3 Isolated e/v card id <0>
11	5 rec card quiet <1>	3 Isolated e/v rank <1>	3 Isolated e/v rank <4>	3 Isolated e/v card id <1>
12	6 rec card quiet <0>	3 Isolated e/v rank <2>	3 Isolated e/v rank <5>	3 Isolated e/v card id <2>
13	6 rec card quiet <1>	unused	3 Isolated e/v region id	unused
14	unused	unused	unused	unused
15	'1'	'1'	'1'	'1'

Figure 4.2: The chosen routing scheme for electron and muon data.

RCT regions far from the $\eta = 0$ boundary was unchanged by this modification (figure 4.3).

As each RCT output requires two optical fibres for retransmission, duplication of the $\eta = 0$ boundary region necessarily required four optical fibres and, hence, one whole source card. The remaining output from each RCT crate therefore required an extra two optical links, or “half” a source card. In order to minimize the number of source cards needed, it was decided that a single half-used source card should not be dedicated to each RCT crate: instead, a source card was shared (and thus fully utilized) between the outputs from the RCT crates representing the positive and negative η sections of each ϕ segment. In performing the routing, the positive and negative η sections were handled separately and the distribution of signals performed at the optical patch-panel. The routing scheme used for the duplicated-type and shared-type of source cards is indicated in figures 4.4(a) and 4.4(b), respectively.

Each source card may be thought of in two complimentary ways: either as a

Cycle 0				
Bit	Output 0	Output 1	Output 2	Output 3
0	RC:5 Region:0 <0>	RC:6 Region:0 <0>	HF:0 η :0 <1>	HF:1 η :0 <1>
1	RC:5 Region:0 <1>	RC:6 Region:0 <1>	HF:0 η :0 <2>	HF:1 η :0 <2>
2	RC:5 Region:0 <2>	RC:6 Region:0 <2>	HF:0 η :0 <3>	HF:1 η :0 <3>
3	RC:5 Region:0 <3>	RC:6 Region:0 <3>	HF:0 η :0 <4>	HF:1 η :0 <4>
4	RC:5 Region:0 <4>	RC:6 Region:0 <4>	HF:0 η :0 <5>	HF:1 η :0 <5>
5	RC:5 Region:0 <5>	RC:6 Region:0 <5>	HF:0 η :0 <6>	HF:1 η :0 <6>
6	RC:5 Region:0 <6>	RC:6 Region:0 <6>	HF:0 η :0 <7>	HF:1 η :0 <7>
7	RC:5 Region:0 <7>	RC:6 Region:0 <7>	HF:0 η :1 <0>	HF:1 η :1 <0>
8	RC:5 Region:0 <8>	RC:6 Region:0 <8>	HF:0 η :1 <1>	HF:1 η :1 <1>
9	RC:5 Region:0 <9>	RC:6 Region:0 <9>	HF:0 η :1 <2>	HF:1 η :1 <2>
10	RC:5 Region:0 overflow	RC:6 Region:0 overflow	HF:0 η :1 <3>	HF:1 η :1 <3>
11	RC:5 Region:0 tau	RC:6 Region:0 tau	HF:0 η :1 <4>	HF:1 η :1 <4>
12	HF:0 η :0 Q	HF:1 η :0 Q	HF:0 η :1 <5>	HF:1 η :1 <5>
13	HF:0 η :1 Q	HF:1 η :1 Q	HF:0 η :1 <6>	HF:1 η :1 <6>
14	HF:0 η :0 <0>	HF:1 η :0 <0>	HF:0 η :1 <7>	HF:1 η :1 <7>
15	BC0	BC0	BC0	BC0

Cycle 1				
Bit	Output 0	Output 1	Output 2	Output 3
0	RC:5 Region:1 <0>	RC:6 Region:1 <0>	HF:0 η :2 <1>	HF:1 η :2 <1>
1	RC:5 Region:1 <1>	RC:6 Region:1 <1>	HF:0 η :2 <2>	HF:1 η :2 <2>
2	RC:5 Region:1 <2>	RC:6 Region:1 <2>	HF:0 η :2 <3>	HF:1 η :2 <3>
3	RC:5 Region:1 <3>	RC:6 Region:1 <3>	HF:0 η :2 <4>	HF:1 η :2 <4>
4	RC:5 Region:1 <4>	RC:6 Region:1 <4>	HF:0 η :2 <5>	HF:1 η :2 <5>
5	RC:5 Region:1 <5>	RC:6 Region:1 <5>	HF:0 η :2 <6>	HF:1 η :2 <6>
6	RC:5 Region:1 <6>	RC:6 Region:1 <6>	HF:0 η :2 <7>	HF:1 η :2 <7>
7	RC:5 Region:1 <7>	RC:6 Region:1 <7>	HF:0 η :3 <0>	HF:1 η :3 <0>
8	RC:5 Region:1 <8>	RC:6 Region:1 <8>	HF:0 η :3 <1>	HF:1 η :3 <1>
9	RC:5 Region:1 <9>	RC:6 Region:1 <9>	HF:0 η :3 <2>	HF:1 η :3 <2>
10	RC:5 Region:1 overflow	RC:6 Region:1 overflow	HF:0 η :3 <3>	HF:1 η :3 <3>
11	RC:5 Region:1 tau	RC:6 Region:1 tau	HF:0 η :3 <4>	HF:1 η :3 <4>
12	HF:0 η :2 Q	HF:1 η :2 Q	HF:0 η :3 <5>	HF:1 η :3 <5>
13	HF:0 η :3 Q	HF:1 η :3 Q	HF:0 η :3 <6>	HF:1 η :3 <6>
14	HF:0 η :2 <0>	HF:1 η :2 <0>	HF:0 η :3 <7>	HF:1 η :3 <7>
15	'1'	'1'	'1'	'1'

Figure 4.3: The chosen routing scheme for the RCT regional data for regions far from the $\eta = 0$ boundary.

piece of hardware or as a processing element within the system. In order to identify and track an individual piece of hardware, each board was given a unique hardware identifier. This identifier is set by means of soldering links onto the board such that, once set, the ID should never be modified. When considering the card as a processing element, we wish to ignore the physical implementation and instead define the device by the role it plays within the system: to do so, a system of logical identities was introduced, that uniquely identifies each board according to the task it performs and the origin of the data upon which it operates. The logical identity of the board is independent of the hardware identifier in that, all boards being identical, any board can be used to fill any logical identity. The logical identity of the board both determines, and is determined by, its interconnections with the RCT but is inferred from the hardware identity by means of a software lookup tableⁱ. Given that the source cards handle, at most, two RCT crates and that the source cards never

ⁱClearly care must be taken to maintain this record when hardware is swapped as each board has no inherent way of knowing its expected functionality or location within the system

Cycle 0				
Bit	Output 0	Output 1	Output 2	Output 3
0	RC:0 Region:0 <0>	RC:1 Region:0 <0>	RC:0 Region:0 <0>	RC:1 Region:0 <0>
1	RC:0 Region:0 <1>	RC:1 Region:0 <1>	RC:0 Region:0 <1>	RC:1 Region:0 <1>
2	RC:0 Region:0 <2>	RC:1 Region:0 <2>	RC:0 Region:0 <2>	RC:1 Region:0 <2>
3	RC:0 Region:0 <3>	RC:1 Region:0 <3>	RC:0 Region:0 <3>	RC:1 Region:0 <3>
4	RC:0 Region:0 <4>	RC:1 Region:0 <4>	RC:0 Region:0 <4>	RC:1 Region:0 <4>
5	RC:0 Region:0 <5>	RC:1 Region:0 <5>	RC:0 Region:0 <5>	RC:1 Region:0 <5>
6	RC:0 Region:0 <6>	RC:1 Region:0 <6>	RC:0 Region:0 <6>	RC:1 Region:0 <6>
7	RC:0 Region:0 <7>	RC:1 Region:0 <7>	RC:0 Region:0 <7>	RC:1 Region:0 <7>
8	RC:0 Region:0 <8>	RC:1 Region:0 <8>	RC:0 Region:0 <8>	RC:1 Region:0 <8>
9	RC:0 Region:0 <9>	RC:1 Region:0 <9>	RC:0 Region:0 <9>	RC:1 Region:0 <9>
10	RC:0 Region:0 overflow	RC:1 Region:0 overflow	RC:0 Region:0 overflow	RC:1 Region:0 overflow
11	RC:0 Region:0 tau	RC:1 Region:0 tau	RC:0 Region:0 tau	RC:1 Region:0 tau
12	RC:2 Region:0 <0>	RC:2 Region:0 <3>	unused	unused
13	RC:2 Region:0 <1>	RC:2 Region:0 <4>	unused	unused
14	RC:2 Region:0 <2>	RC:2 Region:0 <5>	unused	unused
15	BC0	BC0	BC0	BC0

Cycle 1				
Bit	Output 0	Output 1	Output 2	Output 3
0	RC:0 Region:1 <0>	RC:1 Region:1 <0>	RC:0 Region:1 <0>	RC:1 Region:1 <0>
1	RC:0 Region:1 <1>	RC:1 Region:1 <1>	RC:0 Region:1 <1>	RC:1 Region:1 <1>
2	RC:0 Region:1 <2>	RC:1 Region:1 <2>	RC:0 Region:1 <2>	RC:1 Region:1 <2>
3	RC:0 Region:1 <3>	RC:1 Region:1 <3>	RC:0 Region:1 <3>	RC:1 Region:1 <3>
4	RC:0 Region:1 <4>	RC:1 Region:1 <4>	RC:0 Region:1 <4>	RC:1 Region:1 <4>
5	RC:0 Region:1 <5>	RC:1 Region:1 <5>	RC:0 Region:1 <5>	RC:1 Region:1 <5>
6	RC:0 Region:1 <6>	RC:1 Region:1 <6>	RC:0 Region:1 <6>	RC:1 Region:1 <6>
7	RC:0 Region:1 <7>	RC:1 Region:1 <7>	RC:0 Region:1 <7>	RC:1 Region:1 <7>
8	RC:0 Region:1 <8>	RC:1 Region:1 <8>	RC:0 Region:1 <8>	RC:1 Region:1 <8>
9	RC:0 Region:1 <9>	RC:1 Region:1 <9>	RC:0 Region:1 <9>	RC:1 Region:1 <9>
10	RC:0 Region:1 overflow	RC:1 Region:1 overflow	RC:0 Region:1 overflow	RC:1 Region:1 overflow
11	RC:0 Region:1 tau	RC:1 Region:1 tau	RC:0 Region:1 tau	RC:1 Region:1 tau
12	RC:2 Region:1 <0>	RC:2 Region:1 <3>	unused	unused
13	RC:2 Region:1 <1>	RC:2 Region:1 <4>	unused	unused
14	RC:2 Region:1 <2>	RC:2 Region:1 <5>	unused	unused
15	'1'	'1'	'1'	'1'

(a) Duplicating source card routing scheme.

Cycle 0				
Bit	Output 0	Output 1	Output 2	Output 3
0	RC:3 Region:0 <0>	RC:4 Region:0 <0>	sister RC:3 Region:0 <0>	sister RC:4 Region:0 <0>
1	RC:3 Region:0 <1>	RC:4 Region:0 <1>	sister RC:3 Region:0 <1>	sister RC:4 Region:0 <1>
2	RC:3 Region:0 <2>	RC:4 Region:0 <2>	sister RC:3 Region:0 <2>	sister RC:4 Region:0 <2>
3	RC:3 Region:0 <3>	RC:4 Region:0 <3>	sister RC:3 Region:0 <3>	sister RC:4 Region:0 <3>
4	RC:3 Region:0 <4>	RC:4 Region:0 <4>	sister RC:3 Region:0 <4>	sister RC:4 Region:0 <4>
5	RC:3 Region:0 <5>	RC:4 Region:0 <5>	sister RC:3 Region:0 <5>	sister RC:4 Region:0 <5>
6	RC:3 Region:0 <6>	RC:4 Region:0 <6>	sister RC:3 Region:0 <6>	sister RC:4 Region:0 <6>
7	RC:3 Region:0 <7>	RC:4 Region:0 <7>	sister RC:3 Region:0 <7>	sister RC:4 Region:0 <7>
8	RC:3 Region:0 <8>	RC:4 Region:0 <8>	sister RC:3 Region:0 <8>	sister RC:4 Region:0 <8>
9	RC:3 Region:0 <9>	RC:4 Region:0 <9>	sister RC:3 Region:0 <9>	sister RC:4 Region:0 <9>
10	RC:3 Region:0 overflow	RC:4 Region:0 overflow	sister RC:3 Region:0 overflow	sister RC:4 Region:0 overflow
11	RC:3 Region:0 tau	RC:4 Region:0 tau	sister RC:3 Region:0 tau	sister RC:4 Region:0 tau
12	RC:2 Region:0 <6>	RC:2 Region:0 <9>	sister RC:2 Region:0 <6>	sister RC:2 Region:0 <9>
13	RC:2 Region:0 <7>	RC:2 Region:0 overflow	sister RC:2 Region:0 <7>	sister RC:2 Region:0 overflow
14	RC:2 Region:0 <8>	RC:2 Region:0 tau	sister RC:2 Region:0 <8>	sister RC:2 Region:0 tau
15	BC0	BC0	BC0	BC0

Cycle 1				
Bit	Output 0	Output 1	Output 2	Output 3
0	RC:3 Region:1 <0>	RC:4 Region:1 <0>	sister RC:3 Region:1 <0>	sister RC:4 Region:1 <0>
1	RC:3 Region:1 <1>	RC:4 Region:1 <1>	sister RC:3 Region:1 <1>	sister RC:4 Region:1 <1>
2	RC:3 Region:1 <2>	RC:4 Region:1 <2>	sister RC:3 Region:1 <2>	sister RC:4 Region:1 <2>
3	RC:3 Region:1 <3>	RC:4 Region:1 <3>	sister RC:3 Region:1 <3>	sister RC:4 Region:1 <3>
4	RC:3 Region:1 <4>	RC:4 Region:1 <4>	sister RC:3 Region:1 <4>	sister RC:4 Region:1 <4>
5	RC:3 Region:1 <5>	RC:4 Region:1 <5>	sister RC:3 Region:1 <5>	sister RC:4 Region:1 <5>
6	RC:3 Region:1 <6>	RC:4 Region:1 <6>	sister RC:3 Region:1 <6>	sister RC:4 Region:1 <6>
7	RC:3 Region:1 <7>	RC:4 Region:1 <7>	sister RC:3 Region:1 <7>	sister RC:4 Region:1 <7>
8	RC:3 Region:1 <8>	RC:4 Region:1 <8>	sister RC:3 Region:1 <8>	sister RC:4 Region:1 <8>
9	RC:3 Region:1 <9>	RC:4 Region:1 <9>	sister RC:3 Region:1 <9>	sister RC:4 Region:1 <9>
10	RC:3 Region:1 overflow	RC:4 Region:1 overflow	sister RC:3 Region:1 overflow	sister RC:4 Region:1 overflow
11	RC:3 Region:1 tau	RC:4 Region:1 tau	sister RC:3 Region:1 tau	sister RC:4 Region:1 tau
12	RC:2 Region:1 <6>	RC:2 Region:1 <9>	sister RC:2 Region:1 <6>	sister RC:2 Region:1 <9>
13	RC:2 Region:1 <7>	RC:2 Region:1 overflow	sister RC:2 Region:1 <7>	sister RC:2 Region:1 overflow
14	RC:2 Region:1 <8>	RC:2 Region:1 tau	sister RC:2 Region:1 <8>	sister RC:2 Region:1 tau
15	'1'	'1'	'1'	'1'

(b) Shared source card routing scheme.

Figure 4.4: The chosen routing scheme for the RCT regional data for regions near to the $\eta = 0$ boundary.

handle data from more than one ϕ -segment, the logical identity is defined by the concatenation of the RCT ϕ segment number (0 to 8, four bits) with the source card local identity (three bits). The 3-bit source card local identity is defined by

- local identity 0, electron/muon type, $\eta -$
- local identity 1, regional data type, $\eta -$
- local identity 2, $\eta = 0$ boundary duplicating type, $\eta -$
- local identity 3, shared type, $\eta \mp$
- local identity 4, electron/muon type, $\eta +$
- local identity 5, regional data type, $\eta +$
- local identity 6, $\eta = 0$ boundary duplicating type, $\eta +$
- local identity 7, unused due to sharing in type 3 cardⁱⁱ

The logical identity provides the system control software with sufficient information to set the board routing mode, to “build” an event from captured data and to load simulated events into the cards for the purposes of testing the upstream hardware.

To further emphasize the symmetric and self contained grouping of the source cards, the boards representing each ϕ segment are physically located next to each other and each group spatially separated from each other. Within each group, the boards are arranged symmetrically by type to reflect the η symmetry.

In order that all source cards might have the same firmware, the four possible routing modes are all encoded into the **VHDL** (VHSIC Hardware Description Language)ⁱⁱⁱ and the routing mode to be used selected at runtime by the control

ⁱⁱthe presence of an unused type means that the logical identities are not contiguous. However, the formation of identities by concatenation is preferable, in terms of hardware, to a multiplication by 7 followed by an addition

ⁱⁱⁱ**VHSIC** (Very High Speed Integrated Circuit)

software. A macro was created in Microsoft Excel [88] to automatically construct the VHDL routing matrix from a tabular representation to assist in the construction of the routing matrix^{iv}; this having the double advantage of allowing colour coded (and thus more intuitive) design and the automatic generation of documentation (for example, figures 4.1 - 4.4).

4.2 Control Architecture

4.2.1 USB Architecture

While the RCT and the concentrator card use a VME control bus, on the source cards there was insufficient space on the PCB and insufficient FPGA resources to implement the VME protocol. The card instead provided a USB-2.0 link for the purpose of standalone testing. During a review of the GCT design, however, it was decided that the source cards should provide a readout pathway for the RCT and that the cards should be capable of injecting test patterns for the rest of the GCT. This necessitated the implementation of USB controls in the online system. The scale — 63 cards — and distributed nature of the system presented complexities in the implementation of a USB system. The USB standard specification is strict: of particular importance to the current system, the standard specifies a maximum of 127 devices per host controller and a maximum single run length of 5 m. Several options were considered, including the use of a per-crate VME **SBC** (Single-Board Computer). This proposition was rejected due to the requirement that either all PCs in the CMS domain should be, from the system perspective, identical and under control of a central computing authority, or that the SBC should be considered a piece of custom hardware, for which the maintenance and implementation of compliance with the CERN computing rules would remain the responsibility of the subsystem. Both of these options were deemed unacceptable by an external review and the concept rejected. Instead, it was decided that the system should be implemented using the standard

^{iv}There are 17408 permutations of bit routings: 68-bit input $\times$ 64-bit output $\times$ 4 routing modes.

CERN rack PCs and a cascade of USB hubs: this led to two further complications. The row of racks containing the RCT and source card contained only vertical-airflow racks. These are suitable for VME crates but prohibit the placement of rack PCs which require horizontal airflow and completely block vertical airflow. The nearest suitable rack was in the row of racks facing the RCT. The row faces are spaced two metres apart and within the row itself, the two furthest source cards are separated by nearly 8 rack widths, or, nearly 4.4 m. Allowing a 50 percent overhead in the estimated required length to take into account routing, it is clear that running cables from each source card to the PC rack (the preferred location for the USB hubs) was prohibited by the length specification in the USB standard and, as such, there was no alternative but to locate the USB hubs within the source card crates themselves. Rather than introduce further custom hardware into the system, it was decided to be technologically safer to use commercial USB hubs; in order to provide power to and mechanical support for the devices, and in order that the devices complied to rack airflow rules, the hubs required a VME carrier card (figure 4.5).

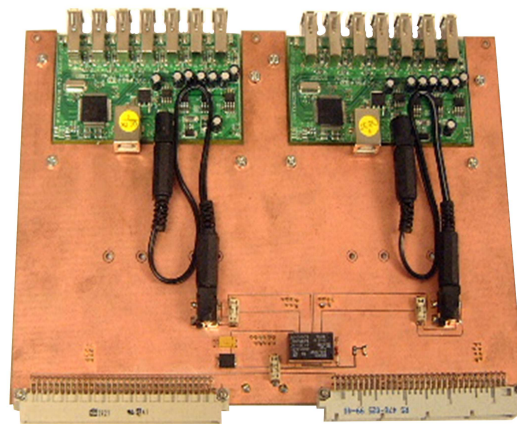


Figure 4.5: The prototype (V2) USB-hub carrier card for the GCT source card system. The board provides mechanical support and power for the two commercial 7-port USB hubs. The final version differs in that it is lacquered and equipped with a front panel.

It was initially planned that the USB hubs should be unmodified and powered directly from the VME+5 V rail by means of a “flying-lead” cable from the carrier to the on-board 5.5 mm power socket; on construction of a prototype carrier board, however, it was discovered that the construction of the hubs was such

that the USB+5 V line from the host PC was not isolated from the +5 V line from the external supply, with the result that, should the host PC be powered while the VME crate was not, the VME+5 V rail was pulled high by the USB+5 V line, and the current sunk by any other devices using the VME+5 V rail. Whereas the VME crate power supply is specified to 100 A at 5 V^v and typical usage is at the level of several amps, the USB bus is specified to supply only 500 mA and, as such, a current drain of the type expected for a VME crate risks serious damage to the USB host. On the production model of the carrier board, the 5 V supply was configured to be switched by the VME+12 V rail, thereby isolating the PC from the crate when the crate was not powered. It was later discovered, however, that the VME crates provided for the source cards were non-standard and did not include a 12 V supply: as a result, rather than redesigning the carrier cards, the cards were modified to bypass the switching mechanism and the hubs modified to isolate the USB+5 V line from the external power supply. Whilst not satisfying the original desire that the hub be unmodified, this proved effective none-the-less. Although the chosen hubs had a 7-to-1 port ratio and, recalling that the source cards are logically organized as nine groups of seven cards, it was decided that the hubs should be organized as one carrier card per crate rather than the one hub per logical group. The six VME crates in which the source cards are located are split into two groups of three crates, with the RCT clock and control crates separating the two groups: the nine groups of source cards are distributed in such a way that the “central” group is split either side of the two RCT crates. If a single USB hub had been used for each group, the USB cables for the central group would have needed to traverse the two RCT crates and, as such, the concept was rejected. A per-crate carrier card provides 14 channels, of which only 10 or 11 are used, but eliminates crate-to-crate communications. Although the USB specification states that up to 127 devices may be used with each USB host, USB is a point-to-point protocol making no provision for broadcast instructions and, as such, the bandwidth per device is inversely proportional to the number of devices. In order that performance should not be degraded significantly, it

^vAlthough at CERN the total current per crate is restricted to 16 A at 5 V by the external infrastructure [89]

was decided that control of the system should be split across multiple PCs and, as the layout of the crates naturally formed two groups, two PCs offered a high degree of symmetry and a reasonable trade-off between bandwidth per card and cost. As, due to the cable length restrictions, it was not possible to run a cable directly from the PCs to the hubs, an additional pair of “master” hub carrier cards were added in the VME crate that is immediately opposite the PC. The six hubs in each group of three crates were then connected to the PCs via these intermediate hubs. Hubs are operationally transparent to the operating system and, therefore, the exact physical structure is irrelevant to the software architecture.

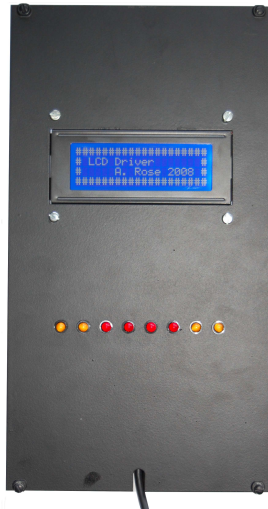


Figure 4.6: The monitor panel for the GCT source card system. The panel is 7 VME slots wide although the VME interface is not used. The card is powered directly from the USB bus.

Although, with experience, the colour-encoded **LEDs** (Light Emitting Diodes) on the source card front panel provide sufficient information to understand the key status indicators and the configurations of certain subsystems, to those with less experience the signals were unintelligible. To overcome this, two USB controlled monitor panels (one for each PC) were constructed based around a 20×4 character array **LCD** (Liquid Crystal Display) and eight indicator LEDs (figure 4.6). These panels were located in the centre of each group of three crates and connected to the PCs via the same USB hub tree as the source cards.

As will be discussed for the source card in the next chapter, the control software for the monitor panel consists of a hardware interface layer which represents a

single instance of the hardware. The interface class provides templated streaming operators so that the LCD monitor may be written to exactly as to the `c++ std::out`, as well as more specialized functions such as a progress bar and functions to control the LEDs.

The USB tree may be represented schematically as in figure 4.7, showing the two PCs, the hierarchy of hubs and hub mounting cards and the distribution of source cards and monitor panels.

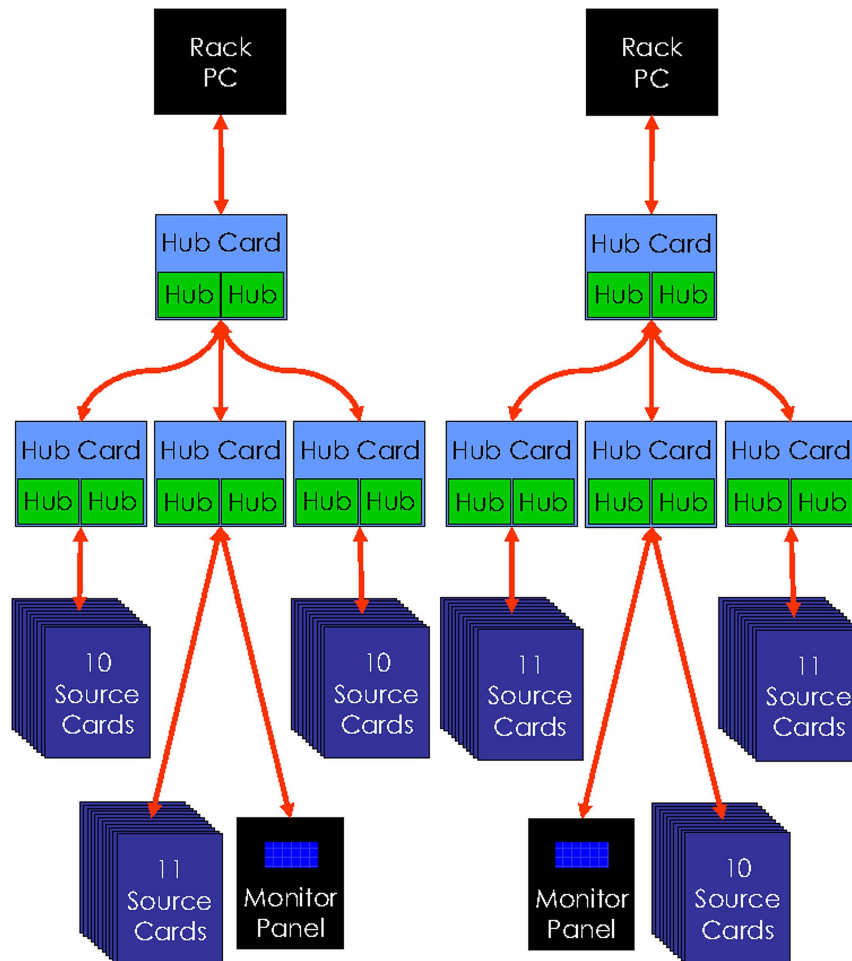


Figure 4.7: Schematic representation of the USB system used to control the CMS GCT source cards. The six hub cards in the centre of the image are one per source card VME crate and the devices below them are all in the same crates as the hub card. The two rack PCs are in the row of racks facing the source cards and the two “master” hub cards are located in the the VME crate closest to the rack PCs; that is, in the rack directly opposite the PCs.

4.2.2 TTC Architecture

In order to distribute the TTC clock and control signal to all of the source cards, a passive splitter (TTCoc) was included in each source card crate. The TTC signals from TTCoc to source card are distributed by 0.5 m optical fibres and do not cross the crate boundaries. The input signals for each TTCoc originate from the RCT Master Clock Crate and, as such, the possibility of phase change due to misconfiguration of the TTC system is negated, either between source card crates or between the RCT and the source cards^{vi}.

^{vi}The RCT Master Clock Crate performs optical to electrical conversion to provide electrical, rather than optical, clock signals for the RCT crates. If this crate is misconfigured, then there is still the possibility of phase-related errors; this is, however, a problem of the RCT and not of the TTC system.

Chapter 5

The Global Calorimeter Trigger Software

“You have no control over what the other guy does;
you only have control over what you do.”

- A.J. Kitt (1968 -)

5.1 Software Architecture

It was initially intended that the source cards be a “passive” card, simply capturing the data from the RCT and relaying it via optical link. The externally-imposed increase in the complexity of the source card subsystem required, however, the implementation of USB controls in the online system.

5.1.1 Software Structure

The source card software is subdivided into many layers, each deriving from the layer below it and each abstracting the controls away from the hardware intricacies towards a purely functional interface. As is standard in the CMS online software, the libraries are written in C++, compiled with **GCC** (Gnu Compiler Collection) [90] and run in **SLC4** (CERN Scientific Linux, release 4). It is usual

in the CMS online systems for the lowest level of the software to be the standardized CMS **HAL** (Hardware Access Library) [91] in order to remove the repeated necessity of each subsystem writing its own hardware interface layer, thereby reducing the probability of coding errors. USB is not supported by the CMS HAL, however, and it was necessary to develop an equivalent USB hardware access layer. The structure of the source card software is shown in figure 5.1.

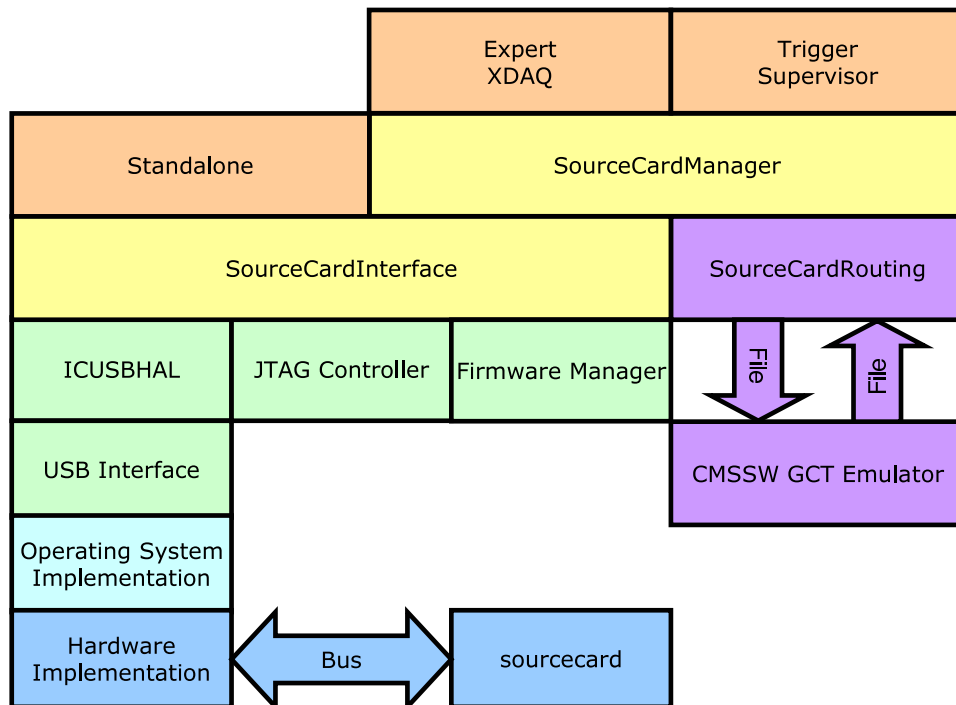


Figure 5.1: The source card software structure.

At the lowest level, USB is implemented by passing data between devices using unidirectional transmission lanes called endpoints. With the exception of the control endpoint (used for enumeration of the device with the host PC), the number of endpoints is unspecified by the USB specification and depends only on the implementation in the slave device: the Cypress SX2 USB driverⁱ used on the source card provides four endpoints, two of which receive data from the host PC and two of which transmit data. On the source card these endpoints are organized so that, for the two receiver endpoints, there is a control pathway and a master reset pathway and, for the two transmitter endpoints, a control pathway

ⁱrecall section 3.2.5

and a DAQ pathway. This arrangement of receiver endpoints allows for control of the board while ensuring that, should the control endpoint fail, the board can be remotely reset. The separation of the transmitter endpoints is optimal for high-bandwidth applications, ensuring that the small packet size commands do not interfere with the throughput of the high bandwidth payload. The throughput of the USB subsystem is typically 20-30 Mbyte/s (160-240 Mbit/s).

In software, the endpoint structure (the zeroth-order abstraction) is represented by the “UsbDevice” class: an abstract base class that prototypes the member functions needed to access the source card, the use of pure virtual functions forcing higher level interfaces to implement all of the required functionality expected from the device. The first-order abstraction customizes the USB interface for a given platform: in this case, the implementation is based on Libusb [92] as the code is run on a Linux distribution but the equivalent library for Microsoft platforms was previously developed for the IDAQ. The second-order abstraction is the ICUSBHAL layer, the direct analogue of the standard CMS HAL. Within the ICUSBHAL, two basic types of function exist: those to read and write 16-bit data to and from the 16 bit address space within the FPGA (the control buses) and those to implement bulk data transfers, where the targets are not registers but FIFOs (and so do not need addressing). At this level of abstraction there is no dependence on the type of device, only on the structure of the endpoints. The third-order abstraction becomes hardware-dependent, by encoding “standard” operations, such as enabling and disabling components, as sequences of read and write operations. In doing so this completely hides the structure of the firmware and the USB endpoints from the software controlling the boards. The GCT uses three such classes, one for the source card, one for the RCT Emulator (IDAQ) and one for the source card LCD monitor panels. Each instance of such a class controls one USB device of that type: one of the numerous advantages of USB is that it can automatically discover how many and what type of devices are connected to the PC and as such, by creating the appropriate number of instances of each of the relevant classes, all (or any) devices connected to the system can be controlled simultaneously. This should be considered in contrast

to other control bus standards, such as VME, where the number and types of devices must be specified and adhered to rigidly.

The abstraction layers are compiled as dynamically-loadable libraries and require a top level executable. For test configurations where the exact numbers and types of board are rigidly specified by the interconnections between them, it is sufficient for the control software executables to refer to each board explicitly. For large-scale systems, however, another layer of abstraction (fourth-order) is needed to automatically handle an arbitrary number of devices. For the RCT Emulator or LCD status displays this is an unnecessary abstraction, as there is never more than one device per controller, but for the source cards (where each PC has either 31 or 32 devices) the SourceCardManager layer is essential.

The top level executable requires both a large degree of user configurability (and thus an intuitive user interface) and the capability of integration into a much wider system to allow for automated and coordinated configuration across the experiment. At the start of the source card software design, it was envisaged that the top level executable would be a standalone XDAQ [62] application with the ability to receive and parse SOAP messages from a central control cell. It was later decided that, as the tasks of each of the level-1 subsystems were all broadly similar, even if the fine details of those tasks are very different, that every subsystem of the level-1 trigger should be derived from the same object type; this making the entire system uniform and making messaging between the objects transparent. This uniformity was achieved by means of the trigger supervisor. The source card trigger supervisor layer, located between the SourceCardManager and the trigger supervisor executive is the fifth-order abstraction.

One additional software class is the Firmware manager. Producing the source card firmware involves two stages of processing: compilation of VHDL code into a “netlist” and building this netlist into a device-specific bit-file. The two processes must occur in this order, but the time taken between the two is unpredictable. The bit-file includes a header region, which includes information about the firmware file and about the build process, and is human readable allowing

for simple machine extraction; of particular importance to the sourcecards, the bit-file header stores the date and time that the file was created. In order to track the expected firmware version against the implemented firmware version, the source cards originally stored a copy of the build-time timestamp in the user-code space of the EEPROM and compared this to the timestamp in the header of a local copy of the bitfile. This scheme, although simple, was deficient in two ways: first, the code expected a file of a fixed name in the local directory. This was generally achieved using a symbolic link to the required file, which, although effective when working, is both prone to configuration error (i.e. the user points the link to the wrong place) and is awkward to manage. Secondly, it occupies the EEPROM usercode space that it later turned out was required for storing the board's unique identifier.

To overcome these problems, the firmware was modified to include two registers that were set to a constant at compile time: one containing an auto-incrementing firmware revision number and one containing a compile-time timestamp. The difference between the compile-time and the build-time timestamp is two-fold: first, the compile-time timestamp is embedded directly into the firmware and, thus, into the fabric of the FPGA and may be accessed by software, whereas the build-time timestamp is not. Secondly, because the compile-time timestamp is embedded directly into the firmware, it is deeply encoded into the body of the bitfile and so not easily accessible for simple machine extraction, as opposed to the build-time timestamp. Although the two stamps are closely related, they cannot be determined from each other due to the indeterminate interval between the two.

The Firmware manager class was created to keep a register of all available firmware files, their compile-time timestamp, revision number and build-time timestamp. This scheme was implemented so as to be backward compatible with older firmwares (saving the build-time timestamp in the EEPROM's user-code space), to distinguish old- from new-type firmwares (those with and without a compile-time timestamp) and to determine which firmware version should be implemented.

As discussed in section 3.2.6, the source cards provide an interface to the JTAG programming chain through the FPGA itself, this allowing the boards to be reprogrammed across the USB interface. This represents an inherent risk as, should transmission of the JTAG datastream be interrupted such that only some fraction of the firmware is loaded onto the EEPROM, then upon the next reboot (either hard or soft), the FPGA will be loaded with a corrupt firmware thereby making a further remote update impossible and necessitating manual reprogramming. This weakness of the system was illustrated when a user, who was unfamiliar with the firmware updating system, terminated the software process during an update and thereby corrupted the board. To prevent this happening again, the software was modified so that the POSIX system signals “INTERRUPT”, “HANGUP”, “TERMINATE”, “QUIT”, “ABORT” and “TTY STOP” are redefined before an update took place, to instead call a function which alerts the user that a firmware upgrade is underway and sets a “kill” flag. The system signals are reset to their default behaviour upon completion of the update and, if the “kill” flag is set, the process is terminated. The two POSIX calls, “STOP” and “KILL” cannot be caught and still represent a risk, as do the two cases of power failure (of the card, PC or both) and USB failure (such as should the cables be unplugged) during an update. These remaining risks may be reduced by educating users and by good design of infrastructure.

5.1.2 The Trigger Supervisor

The trigger supervisor [93] provides a standardized and extensible, platform-independent framework [94] through which all configuration commands are issued and all status monitoring performed; thus allowing a central system to maintain a complete and continuous record of the state of all subsystems and to create a history of performance, errors, rates and system behaviour.

As the system operates across large numbers of widely-distributed PCs and, given the requirement of standardization, a web-based (and thus inherently distributed) architecture was chosen, built on the recognized **XML** (Extensible

Markup Language) standard [95]. The trigger supervisor is not itself responsible for managing the network processes, nor the messages passed between them; instead it uses the XDAQ framework to provide the domain-specific middlewareⁱⁱ. Within this framework the trigger supervisor provides “cells” which are the nodes of the network. These cells may contain an instance of the controller class for each type of hardware, thereby both controlling the hardware and representing that hardware to the rest of the system. It is also possible that a cell may contain no hardware controller existing, instead, solely to send messages to other cells. An example of the trigger supervisor structure is shown in figure 5.2. In the case of the source card cells, they contain an instance of the Source-CardManager class and therefore have control over, and know the status of, all source cards connected to the PC on which the source card cell is running.

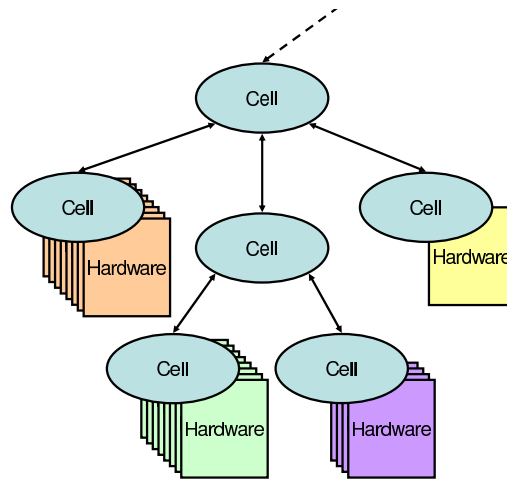


Figure 5.2: The trigger supervisor structure.

For each cell it is possible (and in fact necessary) to create a selection of objects which operate on that cell to achieve a particular result and which automatically manage the creation of a web-based GUI (Graphical User Interface) and SOAP interface. The most important type of object is the CellOperation that implements a Finite State Machine, whose transitions are controlled via GUI/SOAP.

ⁱⁱMiddleware is software designed to facilitate the connection of software components, processes or applications. These components, processes or applications may be located on one or more machines, the middleware providing a set of services to allow hardware- and system-agnostic interactions.

An object of this type is used to configure, enable and suspend operation of each subsystem within the level-1 trigger. All but the simplest of systems require not only an instruction to configure, but also an instruction of which configuration to use. This functionality is achieved in the trigger supervisor by means of an Oracle database [96]: upon receipt of a configure command, the CellOperation requests (from the cell which issued the command) which database key to use and then retrieves from the database the parameters specifying the desired configuration. By making the database entries immutable, storing the database key used for a particular configuration cycle has the same effect as storing the entire parametrization and thereby provides a complete and continuous history of the settings used for all hardware.

While it is necessary to maintain a record of the configuration used, this is not sufficient to guarantee that the hardware is operating as expected. A separate record of the hardware status is required and, in order to record the status of the hardware, it is clearly necessary to have a mechanism to pass messages upwards from each cell to the cell above it in the trigger supervisor hierarchy. This is achieved in two ways: in the case of a hardware or software malfunction (be that a warning, error or exception), a warning message with associated severity level may be created, which is automatically propagating up the chain, where it is handled according to its severity. In order to create a continuous, coherent record of the entire system during normal operation, a command (called a pulse) is regularly propagated from the central cell down through the hierarchy to where it is caught by an object of type "MonitorSource". A MonitorSource object responds by constructing a table of variables summarizing the status of the cell (the status being defined by the cell itself), and passing it back to the cell above. A table's entries may be Boolean, numeric, string or even another table, and each column within a table may be of different type.

One of the most useful features of the standalone XDAQ application was the "at-a-glance" summary table, giving a complete, colour-coded overview of the status of all source cards under the control of the XDAQ application. The importance of the summary table in the installation and integration testing made it

highly desirable that the same feature should be included in the trigger supervisor. Although, in principle, the two tables are very similar, their implementation required very different approaches: for the trigger supervisor, the table must be generated at the master cell level, rather than by the controller PC and this necessitated that the summary table be produced by the merging and interpreting of the data tables returned from the MonitorSource of each source card cell. In contrast, the XDAQ summary table was constructed directly from the class object representing each hardware device. Secondly, the trigger supervisor constructs its web interface dynamically using AJAX[97] while the source card XDAQ application acts as a static web-server, constructing a new, static HTML document each time the page is reloaded; the former has several advantages, but suffers from increased complexity and has tighter specifications. The latter has the advantage of simplicity and speed. Thirdly, as the trigger supervisor represents twice as many hardware devices, the summary table contains twice as many entries. The combined effect of cross-network data manipulation, the processing overhead of dynamic page construction and the effect of more data meant that a “brute-force” approach to constructing the table — whereby the markup language for displaying the table was produced in a single pass (analogous to the static construction in XDAQ) — failed to meet the timing constraints imposed by the AJAX framework. Instead, a more refined approach was required, using the advanced features available in AJAX: upon initial loading of the page, each row of the table was represented by a placeholder object containing an AJAX timer and, when the timer triggered, a call-back was made to the server requesting the data for that row of the table. The timers were configured so that the delay associated with each was proportional to its position in the table, resulting in a staggered loading; this thereby reducing the peak bandwidth and processing demand and conforming to the timing constraints. As periodic self-updating of the page was required, rather than reloading the entire page, which would redraw the placeholders every time, each row was configured to update independently at a set interval after loading. The result was a table with very low peak usage of resources and — excluding the time taken for initial loading — a near-zero dead-time.

Figures 5.3(a) and 5.3(b) show screen shots of the source card “at-a-glance” summary tables for the standalone XDAQ application and for the trigger supervisor, respectively. Within the trigger supervisor, the same data may be displayed in many different ways because all data is passed as tables. For example, figure 5.4 shows a screen capture of the GCT master cell status panel which includes, at the bottom of the image, source card status information interpreted from the same data as the source card summary table. Instead of showing all available information, the data is displayed as a single entry per card with details of any errors being made available by hovering the mouse pointer over each entry. There is no equivalent GCT master status panel in the standalone XDAQ application, as the XDAQ applications are written on a “per subsystems” basis and each subsystem is unaware of any other.

5.1.3 Source Card Capture-file Interface

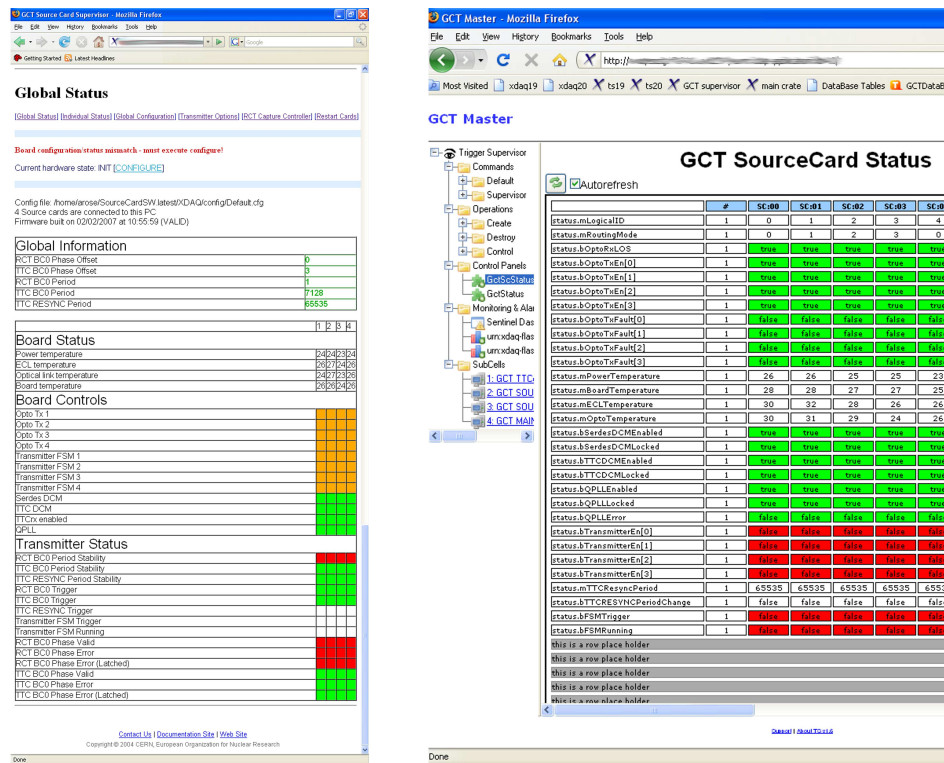
The source cards were required to both capture data from the RCT and to transmit “physics”-like events to the downstream system and, as the hardware may be controlled by several different controllers (standalone per board, standalone all boards, XDAQ, Trigger supervisor), it was decided that the most flexible method of data-storage would be a flat file formatⁱⁱⁱ. For ease of debugging, a tab-delimited file structure was used with the field structure indicated in table 5.1.

Event	Logical	VHDCI 0	VHDCI 0	VHDCI 1	VHDCI 1
Number	Card ID	cycle 0	cycle 1	cycle 0	cycle 1
(dec)	(dec)	(hex)	(hex)	(hex)	(hex)

Table 5.1: Field structure of the source card capture file.

Not only is such a structure easy to read, but it also allows fast manual modification which proved invaluable for the purpose of testing. When storing multiple events for multiple cards, all entries from a single event were grouped and the

ⁱⁱⁱAs opposed to a hierarchical format such as a ROOT [98] tree



(a) Standalone XDAQ

(b) Trigger Supervisor

Figure 5.3: Screen shots of the standalone XDAQ and Trigger Supervisor “at-a-glance” source card summary tables.

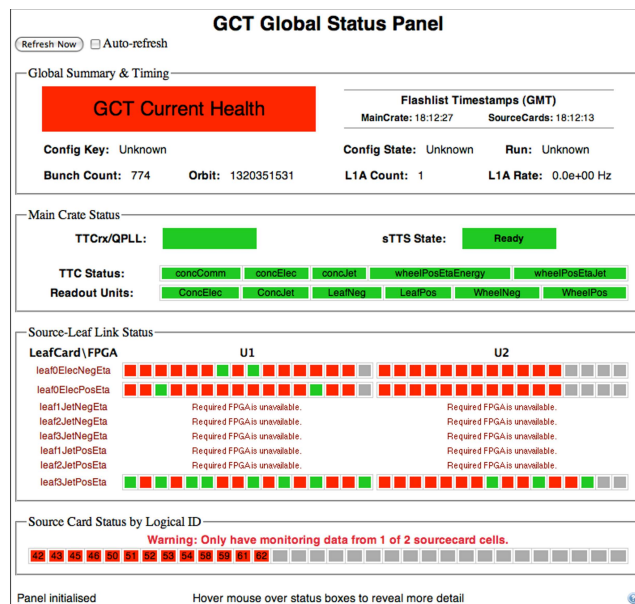
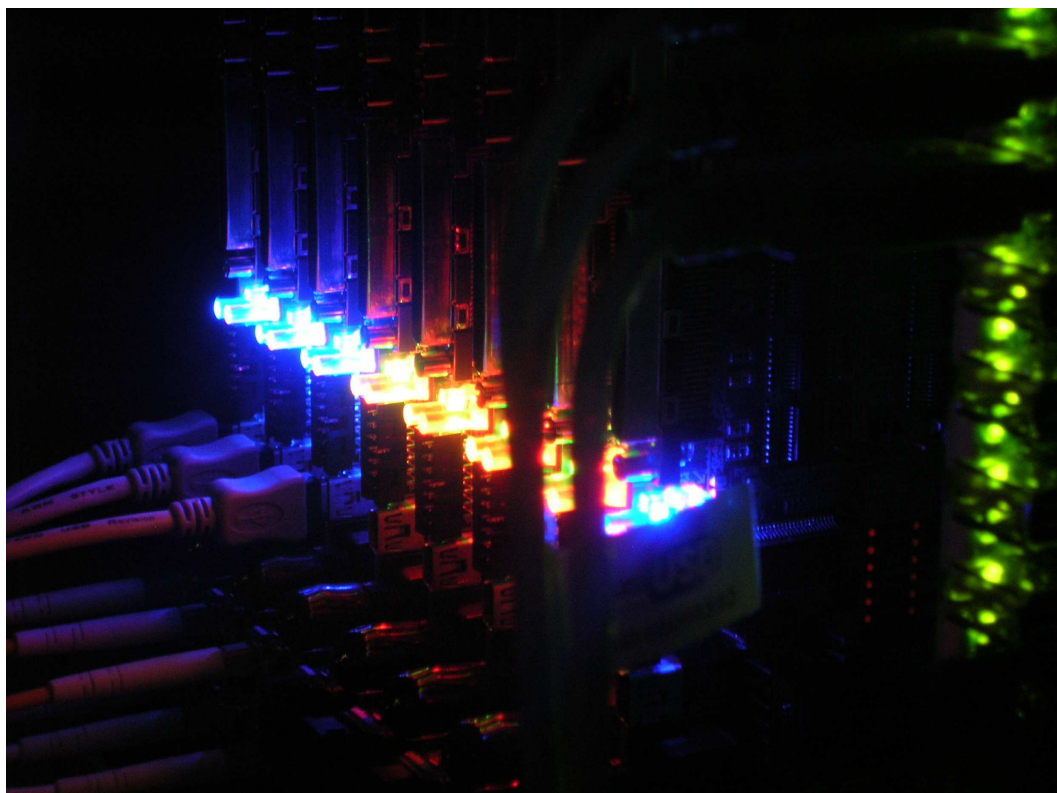


Figure 5.4: Screen shots of the trigger supervisor master panel. This screen shot was taken using a test setup consisting of only a part of the system and, hence, reports many missing devices.

entries sorted by the logical card ID. Given that the final system has 63 cards and that each card may capture or transmit up to 2048 bunch-crossings (4096 half bunch-crossings), the size of these files is clearly large; the structure is, however, such that they may be efficiently compressed by standard compression utilities such as gzip^{iv} [99]. The core functionality for the file interface is provided by the SourceCardRouting library, which emulates the source card's bit manipulation. Several automated tools were written to produce and analyse large samples. A standalone application was produced to construct files where a well-specified test pattern was wanted in a particular bunch-crossing or where the pattern changed in a well-specified manner with each bunch crossing. More importantly, an interface to the **CMSSW** (CMS Software) framework [100] was written that converted the file format into the CMS event format, and vice-versa, thereby allowing a direct comparison between the behaviour and simulation of the hardware. This interface also permitted the use of “physics”-type signals to probe the performance of both the Global Calorimeter Trigger and the Global Trigger.

^{iv} An uncompressed file containing 2000 events is approximately 5.3 MByte but may be compressed to around 350 kByte, a compression ratio of $\sim 15:1$



Overnight integration test of pre-production source cards with a prototype leaf card (not shown).
The green LEDs of the USB hub card may be seen on the right of the figure.
CERN, 31st October 2006.

Chapter 6

The Super-LHC, Super-CMS and the Level-1 Trigger

“We ought not be anxious to encourage innovation, in case of doubtful improvement, for an old system must ever have two advantages over a new one; it is established and it is understood.”

- C.C. Colton (1780 - 1832)

While full operation of the LHC and current experiments is yet to commence, it is anticipated that both the accelerator and the experiments will be upgraded after a decade of running and the beam luminosity raised by an order of magnitude from $10^{34} \text{ cm}^{-2} \text{ s}^{-1}$ to $10^{35} \text{ cm}^{-2} \text{ s}^{-1}$ [101]. Proposals for how this will be achieved vary, but include increasing the crossing rate from 40 MHz to 80 MHz or decreasing it to 20 MHzⁱ. There are also proposals to raise the collision energy (some even suggesting 28 TeV), although these have now largely been ruled out unless there is compelling evidence from LHC results [102]. It is generally understood that the upgrade would be achieved in two phases: phase 1, after 5 years, would consist of a doubling of the luminosity while most infrastructure remains unchanged. Phase 2, after 10 years running, would see a further 5-fold

ⁱIn doubling the crossing rate the number of particles per bunch needs to be increased 5-fold to achieve the increased luminosity, whereas halving the rate requires a 20-fold increase. The number of interactions per crossing increases similarly, while the time available for processing is halved in the former scenario but doubled in the latter.

increase in luminosity, requiring a long shutdown and major modification of the infrastructure. In all of the final state scenarios, the increased flux presents a challenge both to the design of the detector (in terms of radiation damage) and to the design of the data acquisition and triggering systems.

Much of the current CMS data acquisition system is based on a flexible bandwidth approach, relying on the fact that the data rate is governed by Poissonian statistics: rather than building a system capable of handling the peak data rate, where much of the system lies unused at any one time, the system is built to handle the mean data rate — buffering data when the rate is higher than average and emptying the buffers when the rate is lower. A 10-fold increase in the luminosity (and thus in the interaction rate) is matched by a proportionate increase in the number of tracks (and thus of hits) and also data rate; given the recent advancement in semiconductor (particularly FPGA) and telecommunication technologies, however, a 10-fold increase in bandwidth represents a logistical, rather than technological, challenge.

The current level-1 trigger defines events in terms of isolated high- p_T electrons and photons, large missing/transverse energy and jets (including τ -jets), and muon candidates [54], the four ‘highest quality’ of each type of object being found in the global calorimeter and muon trigger before being forwarded to the global trigger. The increased particle density associated with an upgrade affects the level-1 trigger in two fundamental ways: first, due to the increase pile-up there are *NO* isolated objects and, therefore, inherent spatial separation cannot be used to separate objects; for example, the photon rate in the ECAL is expected to be 12 per trigger towers per crossing, with a mean p_T of 3 GeV [103]. Secondly, the increased muon rate requires increased rejection power, rejection power that cannot be achieved by simply raising the p_T threshold (figure 6.1). Only the inclusion of tracker information in the level-1 trigger can resolve this issue.

It is clear that in order to lift these limitations, major modifications of the level-1 trigger system are required: as well as the necessary upgrades to the trigger, it is also likely that the HCAL front-end system will be replaced to improve both the energy resolution and the signal-to-noise ratio.

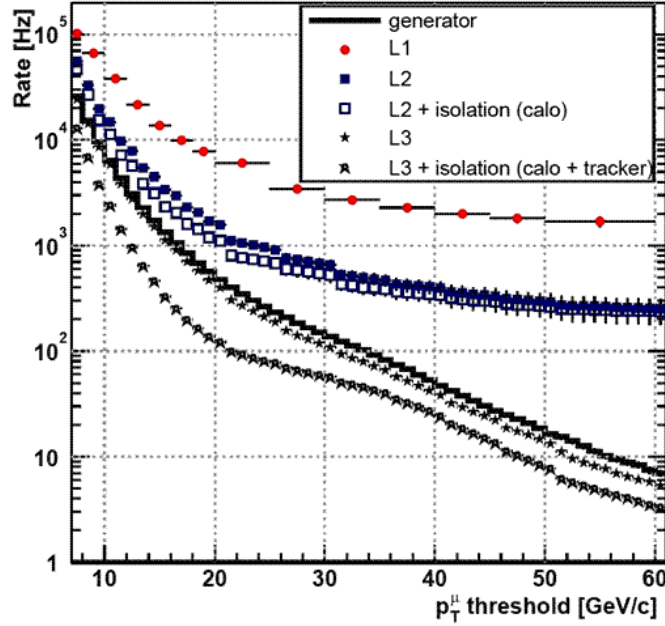


Figure 6.1: Single muon trigger rates for CMS at the LHC as a function of p_T threshold for the level-1 and higher level triggers [56]. Here L1 indicates the rate of CSC+RPC tracks from the Global Muon Trigger, L2 indicates extrapolation of tracks to a valid vertex (with and without calorimeter isolation) and L3 indicates matching of a muon to at least 5 tracker hits (again with and without isolation tests). At SLHC the rates should be scaled by a factor of 10. The flattening of the L1 and L2 rates at high- p_T means that in order to maintain the same trigger rate at SLHC as LHC excludes merely raising the momentum threshold, but instead, requires the inclusion of tracking information.

6.1 Replacement of the Regional Calorimeter Trigger

Due to its reliance on discrete components and ASIC technology, the current Regional Calorimeter Trigger is highly limited in terms of scope for modification and, as such, will almost certainly require replacing at SLHC. In anticipation of such an eventuality, a preliminary study was performed on the resource requirements of such an upgrade; in particular, the processing requirements that would be needed. It was decided that implementation of the current RCT electron-finder algorithm — described in section 2.1.1 — in an FPGA/ μ TCA based system, like the current GCT matrix processor system, would provide a suitable benchmark. The other tasks of the RCTⁱⁱ (namely, deskewing and linearizing the tower data and summing towers into regions) are considered simple, relative to

ⁱⁱsee section 2.2.2

the task of electron-finding, and are not considered here.

Although the study was speculative, certain assumptions were required about the level-1 trigger data-flow architecture at SLHC. The RCT electron-finder operates across 4032 trigger towers, where each tower consists of an ECAL and HCAL pair. Whereas the trigger optical links from the ECAL operate at a rate of $16+4$ bitsⁱⁱⁱ per bunch-crossing (a rate of 800 Mbit/s) and operate at a wavelength of 1310 nm, the high density optics and serializer/deserializers of the type used on the matrix processor operate at 850 nm and rates of up to 3.75 Gbit/s. The routing of the trigger tower fibres directly to the processor card would, therefore, not only require custom optical components but would also be a very inefficient use of the available bandwidth. Instead, it is envisaged that an inexpensive adapter card would be developed to convert between the old style links and the newer optical links, much like the role of the source cards in the current trigger-system. As the current RCT operates on a compressed data format and not on raw data, it seemed reasonable that, for the current study, the adapter card could also be considered to perform such a compression, reducing the number of bits per tower from 16 to 8. Taken alongside a 4-fold rate increase, from 800 Mbit/s to 3.2 Mbit/s, a working assumption of 8 ECAL towers per fibre per bunch-crossing was made. As the nature of the upgraded HCAL and the format of the upgraded links was unknown, it was assumed that the HCAL too would be preprocessed to 8 bits per tower per bunch-crossing, or equivalently 8 towers per fibre per bunch-crossing. It is assumed that, apart from data compression, no other preprocessing is performed as this maximizes the processing load (and thus the resource usage) of the main processor card.

As with the current GCT, somewhat unsurprisingly given the similarities between the schemes, the most significant complication in implementing the algorithm is boundary-sharing between regions: given the success of the split-stage duplication and sharing method in the current scheme, the same technique was applied; those fibres where all 8 towers about a boundary being duplicated prior

ⁱⁱⁱThe links are 16b/20b encoded

to processing and sent to both sides of the boundary and, where the towers abutting the boundary are not on a single fibre, the data is deserialized and only the minimal data volume necessary shared between FPGAs (figure 6.2).

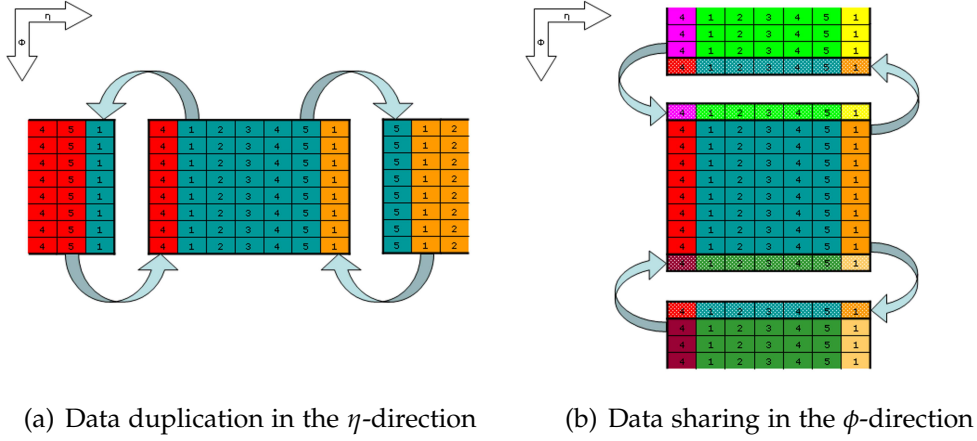


Figure 6.2: The split stage duplication/sharing architecture chosen for the modular implementation of the RCT electron-finder algorithm. The dotted regions indicate that, when shared between FPGAs, the data need not simply be duplicated but may consist of derived quantities. In this case, there is insufficient bandwidth on a single fibre to transfer the raw data, as 8 bits ECAL + 8 bits HCAL for 7 tower = 112 bits per bunch-crossing, exceeding the limit of 64 bits per bunch-crossing. It is not necessary to transfer all 16 bits per tower: instead, all that is required is 7 bits ECAL energy and 2 veto bits, or 9 bits per tower in total, for which there is sufficient bandwidth to share data from 7 towers.

Such a scheme would allow for a maximum region size of 8 towers in ϕ by 5 towers in η (or 6 towers at extreme η where boundary sharing is only along the inner edge) which, when mapped onto the calorimeter, is equivalent to 9 regions in ϕ by 11 regions in η , or 99 processor cards. The interconnection of the hardware is based on three considerations:

- The data duplication across the η -boundaries and the data-sharing across the ϕ -boundaries may be equally performed across the front-panel optical links or across the backplane.
- Of the data duplication and the data-sharing, one must necessarily use the front-panel optical links.
- A μ TCA crate provides 12 card slots, that are fully interconnected across the backplane

As the limitations of the hardware are not restrictive in this case, then the optimal layout is represented by the scenario with minimal hardware usage: this is achieved by the use of 9 μ TCA crates^{iv} each containing 11 processor cards, and each crate representing $\frac{1}{9}$ of the detector in ϕ .

For the purpose of comparison, two different firmware implementations were created: the first consisting of combinatorial logic similar to that used in the current RCT and the second using a deep micropipeline architecture more suitable for FPGA implementation. The implicit assumption of the former type of architecture is that, upon initial receipt and deserialization of the data, it is “unfolded” into a 2-dimensional array and the array processed as a monolithic block with each stage of processing being performed in parallel on each member of the array. The output from each processing stage was synchronously registered and all stages chained into a macroscopic processing pipeline, the final stage being a single-pass matrix-sorting algorithm^v to return the maximum energy candidate. The micropipelined architecture performs branching immediately on initial receipt and deserialization of the data, so that each “local” region (a region of the minimal size of the object in which we are interested) runs separately and the threads merged bitonically, this automatically resulting in a sorted list. Performing a delta-time simulation of both models indicated that the unoptimized latency of both designs was 8 bunch-crossings, excluding the initial deserialization and allowing 3 bunch-crossings for data-sharing. Synthesis of the two designs for the Xilinx XC5VLX110T FPGA indicated a resource usage of 20% and 5% respectively, the former design making more extensive use of look-up table resources, the latter design of registers, but both designs clearly indicating that the limiting factor in implementation of the algorithm being the number of serial links and not in the processing resources. Compilation and routing

^{iv} As the vertically oriented μ TCA standard evolved from the horizontally oriented **AMC** (Advanced Mezzanine Card) standard, the somewhat confusing nomenclature exists whereby the vertical dimension is referred to as the width and the left-right horizontal dimension is referred to as the height. A single width μ TCA crate occupies 2U of rack space in a standard 19” rack (excluding external power). In contrast a single VME crate of the type used in the current RCT occupies 9U. Therefore 9 μ TCA crates occupy the same amount of rack-space as 2 VME crates.

^v Identical to that used on the GCT leaf card.

of the two designs indicated that the most significant difference between the two design philosophies was the maximum acceptable clock speed: the combinatorial design was limited to 17 MHz, not the 160 MHz required for a 40 MHz crossing rate, while the maximum acceptable clock speed of the pipelined design was in excess of 260 MHz. Both the difference in resource usage and in the maximum clock speed agree with what was expected: namely that, in FPGAs, combinatorial trees require large amounts of routing resources and that this causes problems with routing signals; in particular, in the length-matching of the traces and this consequently results in performance which is dominated by wire delays. By using a deep micropipelined design, only signals local to a thread need be length-matched and significant improvement in performance may be achieved.

Taking, as a figure of merit, an unoptimized resource usage of 5% (as above) for a link usage of 100%, it is reasonable to suppose that the internal resources of the FPGA will not (assuming good firmware design) be the limiting factor in the design of any future trigger and, as such, preliminary calculations of hardware requirements for designs based around alternative devices may be made in terms of bandwidth alone. These calculations may be found for three top-specification current generation FPGAs, along with that for the Xilinx XC5VLX110T, in table 6.1. In this table the total bandwidth estimate is calculated as

$$\text{number of towers} \times \text{number of bits per tower} \times \text{encoding factor} \times \text{crossing rate}$$

Assuming LHC crossing rates (40 MHz) and 8b/10b encoding, for a compressed (8+8 bits) input the aggregate bandwidth is 3.2256 Tbit/s and, for uncompressed (16+16 bits) inputs, the rate is 6.4512 Tbit/s. Allowing 50% overhead for boundary sharing and rounding up, this results in approximate data rates of 5 and 10 Tbit/s, respectively. It can be seen that the calculation based on bandwidth indicates the need for 98 Xilinx XC5VLX110T FPGAs, while the detailed calculation discussed in the text indicated 99, giving confidence that — for the purpose of approximation — calculations based on bandwidth are suitable.

Device	Xilinx Virtex-5 LX110T (XC5VLX110T)	Xilinx Virtex-5 TX240T (XC5VTX240T)	Altera Stratix-IV GX (EP4SGX530)	Achronix Speedster (SPD180)
Serial links	16	48	48	40
Link speed (LHC compatible)	3.75 Gbit/s (3.2 Gbit/s)	6.5 Gbit/s (6.4 Gbit/s)	6.5 Gbit/s (6.4 Gbit/s)	10.375 Gbit/s (9.6 Gbit/s)
Serial Bandwidth (LHC compatible)	60 Gbit/s (51.2 Gbit/s)	312 Gbit/s (307.2 Gbit/s)	312 Gbit/s (307.2 Gbit/s)	415 Gbit/s (384 Gbit/s)
Logic Cells normalized to XC5VLX110T	1	2.166	1.992	1.48
Logic clock speed	550 MHz	550 MHz	600 MHz	1500 MHz
General Purpose I/O (Channel Data Rate)	340 (1.25 Gbit/s)	340 (1.25 Gbit/s)	196 (1.6 Gbit/s)	850 (1.066 Gbit/s)
Number of FPGAs				
Compressed input	98	17	17	13
Uncompressed input	196	34	34	26

Table 6.1: Estimation of the number of logic devices required for an upgraded calorimeter trigger based on bandwidth requirements alone.

All results stated so far have been based on the assumption that the upgraded trigger system will be based on a one FPGA per board philosophy, with all communications between FPGAs performed by high speed serial links. As listed in table 6.1, however, modern FPGAs have considerable I/O resources that are ignored in such a philosophy. If, instead, one considers larger boards with multiple FPGAs, with inter-chip communication performed using only the general purpose I/O and the high-speed serial links reserved for inter-board communication, the bandwidth requirement for boundary sharing drops considerably and, consequently, so does the number of devices and boards: taken to the extreme (for the purpose of illustration only), performing the entire electron trigger on a single board would avoid any boundary sharing across the high-speed serial links and would require only 21 Xilinx Virtex-5 TX240T to handle all of the uncompressed (16+16 bits) calorimeter data. Such a design would, however, be

against the generally accepted philosophy towards smaller, modular, application non-specific hardware. Based on the ever increasing bandwidth requirements of commercial products and aggressive competition between FPGA manufacturers, and also on the current rate and direction of developments in the capabilities of FPGAs, it is reasonable to suppose that the next generation of FPGAs will raise both the number and speed of integrated serializers, further reducing the number of processors that will be required [104].

6.2 Calorimeter-seeded tracker readout

As the CMS pixel tracker will need replacing after ten years operation under LHC conditions, it has been suggested [105] that the future tracker should be designed such that it may be selectively read out, based on objects found at the calorimeter or muon systems, and the tracking information “used”^{vi} to refine and more tightly classify the primitives. As such a scheme requires two separate readout stages — but so as not to cause confusion with the higher level trigger, which already contains a stage referred to as a level-2 trigger — such a scheme is referred to as a level-1.5 trigger. To evaluate the feasibility of such a scheme, a Monte Carlo study was performed using the current detector geometry to estimate the rates and data volumes that might be expected.

The electron-finding algorithm in the current CMS level-2.5 (HLT) trigger [106, 107] has been shown to significantly reduce the level-1 isolated electron rate, having an electron-finding efficiency of greater than 95% and reducing the background from jets by a factor of 10. This background rejection is achieved by the required association of the level-2 calorimeter clusters with hits in the inner pixel tracker. The hit-matching is performed in fixed windows centred on the helical projection from the nominal interaction point (0,0,0) to the centre of the calorimeter cluster. If a compatible hit is found in the inner pixel layer, the longitudinal (z) vertex position is calculated, windows projected onto the remaining

^{vi}At the time of the proposal it was not suggested how the information should be used, it was merely known that it was required

two pixel layers and hits which are compatible with the vertex, inner hit and calorimeter cluster are located. If no compatible hit is found in the inner pixel layer, the initial search is performed on the second layer, with the third layer providing the confirmatory hit. If a projection exists with two or more compatible hits, it is accepted as an electron; if not, it is rejected as a jet.

It was decided that porting of the higher level trigger algorithm to a scheme implementable at level-1 would provide a suitable benchmark for the calculation of the data volume requirements. The higher level trigger, however, has certain advantages over the level-1 trigger: first, to achieve the best possible position resolution, the higher level trigger uses crystal level information in the formation of clusters and calculates the direction and the centre of the cluster by means of a logarithmically-corrected energy-weighted sum. Secondly, because the higher level trigger clusters are not of fixed size or shape, the higher level trigger can perform energy/momentum corrections based on the size of the cluster. Thirdly, the $\mathcal{O}(1\text{ s})$ latency of the higher level trigger permits multiple iterations over data.

For the purpose of this study, it was assumed that the ECAL front-end electronics would remain unchanged and that, as such, the finest granularity trigger primitive information would be at the trigger tower level^{vii}. Furthermore, such coarse granularity provides no information about the energy distribution of the showers contained within a tower and insufficient information (about the energy distribution within larger objects) to reconstruct the parent particle's charge.

Given the uncertainty in particle charge and the uncertainty in the energy and position measurement associated with the fixed-size clusters, and given the computational complexity of helical projection, it was decided that linear projection from calorimeter to nominal interaction point would be the most likely scenario. Three scenarios for calorimeter-seeding were considered and, ranked in order of increasing complexity, these were:

^{vii}More recent work indicates that the granularity could be improved to half a trigger tower in each dimension by performing energy-weighted interpolation. As will be shown, however, such an improvement does not significantly affect the present results.

1. **Threshold Method** - Any tower with a reconstructed EM energy over a threshold is used as a seed point. Threshold levels between 0 GeV and 50 GeV were tested in steps of 5 GeV.
2. **Maxima Method** - The central tower of any 3×3 region, identified by the current RCT electron-finder algorithm as being an e/γ -candidate, is used as a seed point.
3. **Weighted Method** - The energy-weighted mean position of all towers in any 3×3 region, identified by the current RCT electron-finder algorithm as being an e/γ -candidate, is used as a seed point.

Using a single particle simulation (termed a “particle gun”) with no pile-up in CMSSW 1.8.4 (“fast” simulation) for electrons and positron, at both 5 and 50 GeV, the separation in the ϕ - and z - directions between the position of the projected point and the positions of all simulated hits in the pixel tracker were calculated for each calorimeter-seed (figure 6.3). Exclusion of pile-up events and examination of all tracker hits allows for a Monte Carlo study of the effects of multiple scattering, charge-sharing and bremsstrahlung without the necessity of associating hits to individual tracks in the analysis.

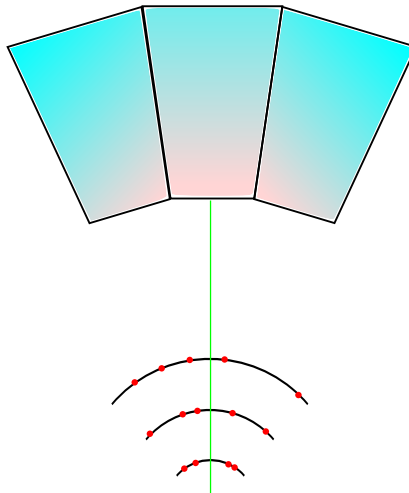


Figure 6.3: The method used for calorimeter-seeded tracker readout studies. A linear projection was made from the calorimeter seed point to the nominal vertex and, for each tracker layer, the distance in the ϕ - and z - directions of each hit from the projected point was measured.

5 GeV electron (positron) gun, no pile up																		
Calorimeter Seed Method	Pixel layer 0						Pixel layer 1						Pixel layer 2					
	Entries	median $\Delta\phi$ (radians)	Lower Bound (radians)	Upper Bound (radians)	Capture Fraction	"width" (radians)	Entries	median $\Delta\phi$ (radians)	Lower Bound (radians)	Upper Bound (radians)	Capture Fraction	"width" (radians)	Entries	median $\Delta\phi$ (radians)	Lower Bound (radians)	Upper Bound (radians)	Capture Fraction	"width" (radians)
Threshold (0 GeV)	415285 (410597)	-0.12 (0.12)	-0.90 (-0.77)	0.77 (0.90)	0.99 (0.99)	1.80	333081 (330268)	-0.12 (0.13)	-0.89 (-0.76)	0.76 (0.89)	0.99 (0.99)	1.78	213968 (214140)	-0.14 (0.14)	-0.90 (-0.75)	0.75 (0.89)	0.99 (0.99)	1.79
Threshold (25 GeV)	271 (294)	-0.05 (0.06)	-0.13 (-0.05)	0.05 (0.14)	0.99 (0.99)	0.27	214 (208)	-0.05 (0.06)	-0.12 (-0.02)	0.04 (0.13)	0.99 (0.99)	0.25	0 (0)	N/A (N/A)	N/A (N/A)	N/A (N/A)	N/A (N/A)	N/A
Maxima	136812 (135668)	-0.12 (0.12)	-0.46 (-0.25)	0.26 (0.48)	0.99 (0.99)	0.94	109343 (108777)	-0.13 (0.13)	-0.46 (-0.25)	0.26 (0.47)	0.99 (0.99)	0.93	69017 (69276)	-0.15 (0.15)	-0.50 (-0.28)	0.28 (0.52)	0.99 (0.99)	1.02
Weighted	136812 (135668)	-0.13 (0.13)	-0.34 (-0.10)	0.10 (0.34)	0.99 (0.99)	0.68	109343 (108777)	-0.13 (0.14)	-0.34 (-0.09)	0.09 (0.34)	0.99 (0.99)	0.68	69017 (69276)	-0.15 (0.15)	-0.37 (-0.11)	0.10 (0.38)	0.99 (0.99)	0.75
50 GeV electron (positron) gun, no pile up																		
Calorimeter Seed Method	Pixel layer 0						Pixel layer 1						Pixel layer 2					
	Entries	median $\Delta\phi$ (radians)	Lower Bound (radians)	Upper Bound (radians)	Capture Fraction	"width" (radians)	Entries	median $\Delta\phi$ (radians)	Lower Bound (radians)	Upper Bound (radians)	Capture Fraction	"width" (radians)	Entries	median $\Delta\phi$ (radians)	Lower Bound (radians)	Upper Bound (radians)	Capture Fraction	"width" (radians)
Threshold (0 GeV)	632831 (629757)	-0.02 (0.02)	-0.82 (-0.79)	0.80 (0.81)	0.99 (0.99)	1.63	504718 (503376)	-0.02 (0.02)	-0.81 (-0.79)	0.80 (0.81)	0.99 (0.99)	1.62	319193 (320031)	-0.02 (0.02)	-0.82 (-0.80)	0.81 (0.82)	0.99 (0.99)	1.64
Threshold (25 GeV)	163336 (162586)	-0.01 (0.01)	-0.12 (-0.11)	0.11 (0.12)	0.99 (0.99)	0.24	128389 (127864)	-0.01 (0.01)	-0.12 (-0.11)	0.11 (0.12)	0.99 (0.99)	0.24	74598 (74558)	-0.01 (0.01)	-0.09 (-0.06)	0.06 (0.09)	0.99 (0.99)	0.18
Maxima	137744 (136905)	-0.01 (0.01)	-0.11 (-0.10)	0.10 (0.11)	0.99 (0.99)	0.22	110667 (110193)	-0.01 (0.01)	-0.11 (-0.10)	0.10 (0.11)	0.99 (0.99)	0.22	69507 (69567)	-0.01 (0.01)	-0.11 (-0.07)	0.07 (0.11)	0.99 (0.99)	0.22
Weighted	137744 (136905)	-0.01 (0.01)	-0.09 (-0.07)	0.07 (0.09)	0.99 (0.99)	0.18	110667 (110193)	-0.01 (0.01)	-0.08 (-0.07)	0.07 (0.08)	0.99 (0.99)	0.16	69507 (69567)	-0.01 (0.01)	-0.08 (-0.05)	0.04 (0.08)	0.99 (0.99)	0.16

Table 6.2: Distribution of the distance of tracker hits from a calorimeter seed point in the ϕ -direction. The four methods of calorimeter seeding are in order of increasing complexity. The upper and lower bounds are defined to be the points capturing 49.5% of events above and below the median, respectively. The "width" of the window is the size of window required to capture 99% of tracks, given that the charge is unknown.

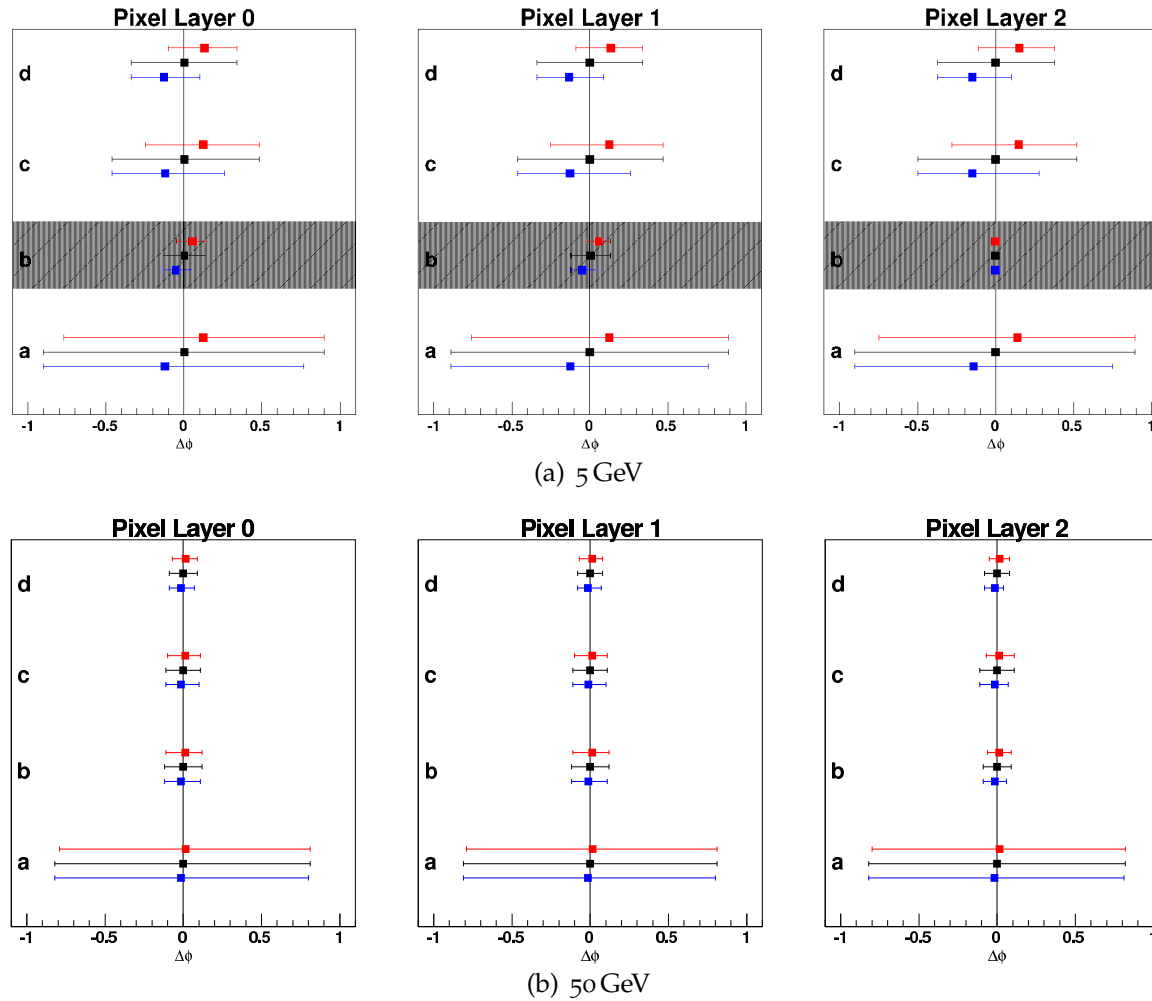


Figure 6.4: Distribution of the distance of tracker hits from a calorimeter seed point in the ϕ -direction. The spot represents the median $\Delta\phi$, whilst the lines represent the upper and lower bounds. The calorimeter seeding methods are labelled a)Threshold(0 GeV), b)Threshold(25 GeV), c)Maxima and d)Weighted. The blue points represent electrons, red points represent positrons and black points represent the window required for unknown charge. The hashed regions indicate the non-representative, highly-boosted case with very low statistics.

5 GeV electron (positron) gun, no pile up																		
Calorimeter Seed Method	Pixel layer 0						Pixel layer 1						Pixel layer 2					
	Entries	median Δz (cm)	Lower Bound (cm)	Upper Bound (cm)	Capture Fraction	"width" (cm)	Entries	median Δz (cm)	Lower Bound (cm)	Upper Bound (cm)	Capture Fraction	"width" (cm)	Entries	median Δz (cm)	Lower Bound (cm)	Upper Bound (cm)	Capture Fraction	"width" (cm)
Threshold (0 GeV)	415285 (410597)	-0.02 (-0.04)	-14.20 (-14.20)	14.20 (14.30)	0.99 (0.99)	28.50	333081 (330268)	-0.02 (-0.02)	-15.20 (-15.20)	15.20 (15.20)	0.99 (0.99)	30.40	213968 (214140)	-0.03 (-0.02)	-14.60 (-14.40)	14.40 (14.30)	0.99 (0.99)	29.00
Threshold (25 GeV)	271 (294)	-0.46 (-0.13)	-13.30 (-10.30)	12.10 (12.50)	0.99 (0.99)	25.80	214 (208)	-0.06 (-0.23)	-12.90 (-9.60)	9.90 (10.30)	0.99 (0.99)	23.20	0 (0)	N/A (N/A)	N/A (N/A)	N/A (N/A)	N/A (N/A)	N/A
Maxima	136812 (135668)	-0.01 (-0.02)	-13.00 (-13.00)	12.90 (13.00)	0.99 (0.99)	26.00	109343 (108777)	0.00 (-0.00)	-13.20 (-13.20)	13.00 (13.10)	0.99 (0.99)	26.30	69017 (69276)	-0.02 (-0.00)	-12.70 (-12.60)	12.30 (12.50)	0.99 (0.99)	25.20
Weighted	136812 (135668)	-0.01 (-0.03)	-13.00 (-13.10)	13.00 (13.00)	0.99 (0.99)	26.00	109343 (108777)	-0.01 (-0.00)	-13.20 (-13.20)	13.00 (13.10)	0.99 (0.99)	26.30	69017 (69276)	-0.03 (-0.01)	-13.10 (-13.00)	12.70 (12.80)	0.99 (0.99)	25.90
50 GeV electron (positron) gun, no pile up																		
Calorimeter Seed Method	Pixel layer 0						Pixel layer 1						Pixel layer 2					
	Entries	median Δz (cm)	Lower Bound (cm)	Upper Bound (cm)	Capture Fraction	"width" (cm)	Entries	median Δz (cm)	Lower Bound (cm)	Upper Bound (cm)	Capture Fraction	"width" (cm)	Entries	median Δz (cm)	Lower Bound (cm)	Upper Bound (cm)	Capture Fraction	"width" (cm)
Threshold (0 GeV)	632831 (629757)	0.01 (-0.03)	-14.20 (-14.10)	14.20 (14.20)	0.99 (0.99)	28.40	504718 (503376)	0.00 (-0.03)	-15.60 (-15.40)	15.30 (15.50)	0.99 (0.99)	31.10	319193 (320031)	0.01 (-0.05)	-14.40 (-14.50)	14.30 (14.30)	0.99 (0.99)	28.80
Threshold (25 GeV)	163336 (162586)	0.00 (-0.02)	-12.90 (-12.90)	13.00 (13.00)	0.99 (0.99)	25.90	128389 (127864)	0.01 (-0.03)	-13.30 (-13.20)	13.30 (13.40)	0.99 (0.99)	26.70	74598 (74558)	0.03 (-0.06)	-12.30 (-12.30)	12.50 (12.50)	0.99 (0.99)	24.80
Maxima	137744 (136905)	0.01 (-0.02)	-13.00 (-13.00)	13.00 (13.00)	0.99 (0.99)	26.00	110667 (110193)	0.01 (-0.03)	-13.20 (-13.00)	13.20 (13.20)	0.99 (0.99)	26.40	69507 (69567)	0.02 (-0.07)	-12.40 (-12.30)	12.40 (12.50)	0.99 (0.99)	24.90
Weighted	137744 (136905)	0.01 (-0.02)	-12.90 (-12.90)	12.90 (12.90)	0.99 (0.99)	25.80	110667 (110193)	0.01 (-0.03)	-13.00 (-12.90)	13.00 (13.10)	0.99 (0.99)	26.10	69507 (69567)	0.02 (-0.07)	-12.50 (-12.40)	12.50 (12.50)	0.99 (0.99)	25.00

Table 6.3: Distribution of the distance of tracker hits from a calorimeter seed point in the z-direction. The four methods of calorimeter seeding are in order of increasing complexity. The upper and lower bounds are defined to be the points capturing 49.5% of events above and below the median, respectively. The "width" of the window is the size of window required to capture 99% of tracks, given that the charge is unknown.

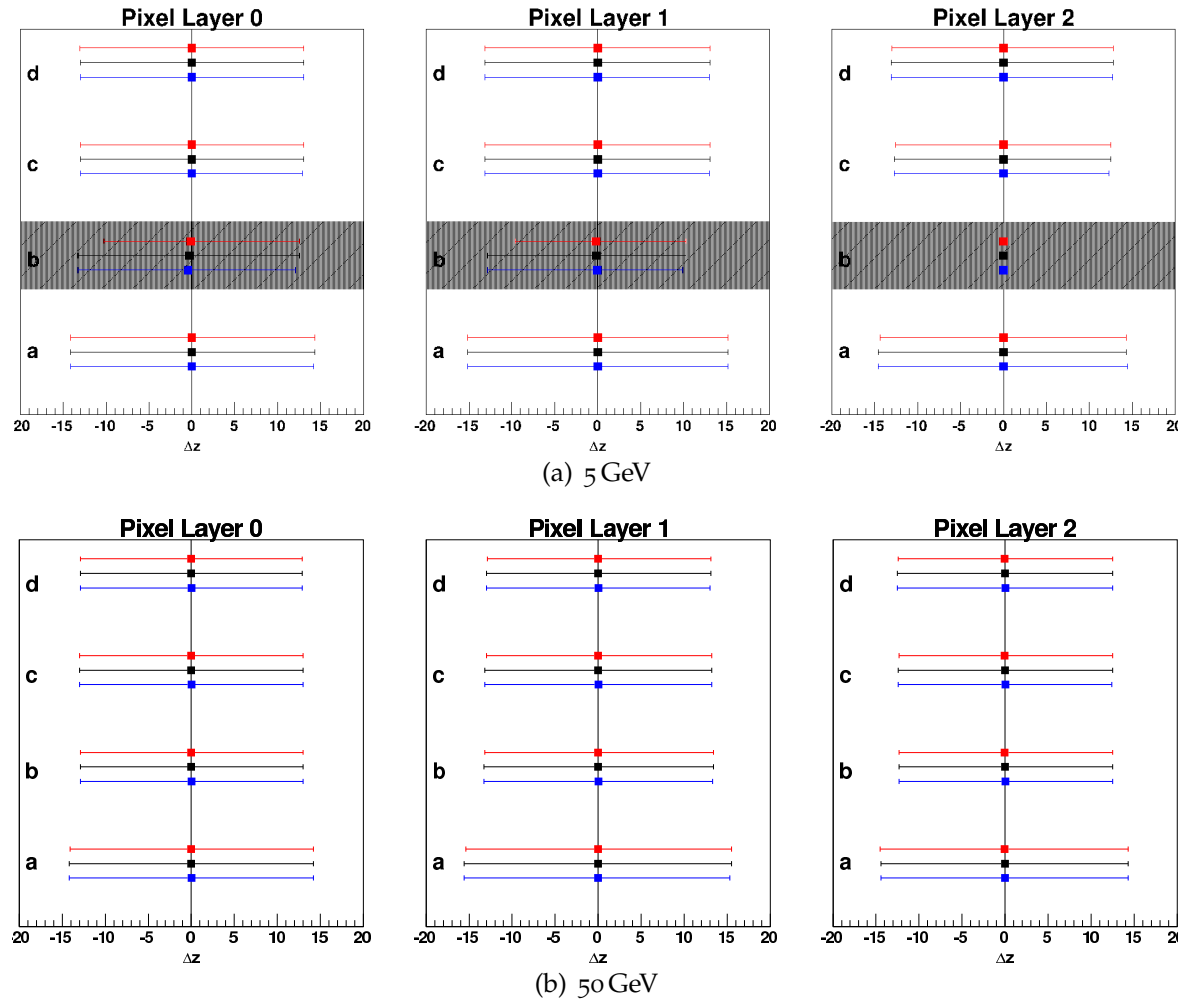


Figure 6.5: Distribution of the distance of tracker hits from a calorimeter seed point in the z -direction. The spot represents the median Δz , whilst the lines represent the upper and lower bounds. The calorimeter seeding methods are labelled a)Threshold(0 GeV), b)Threshold(25 GeV), c)Maxima and d)Weighted. The blue points represent electrons, red points represent positrons and black points represent the window required for unknown charge. The hashed regions indicate the non-representative, highly-boosted case with very low statistics.

For the scenarios listed above, table 6.2 and figure 6.4 summarize the distributions of the angular separation of hits in the pixel tracker from the straight track hypothesis. As the distributions are highly non-Gaussian, rather than fitting a curve, the results have been integrated numerically to determine the median of the distribution and the upper and lower bounds defined to be the points which contain 49.5% of the entries above and below the median, respectively (such that 99% of entries lie between the upper and lower bound). First, considering the equivalent measurements for the electron and positron cases: it can be seen that, as expected, the sign of the median angular separation and that of the bounds are opposite, but both the magnitude of median angular deviation and magnitude of the bounds of the distributions are similar for positrons and electrons. Furthermore, in all cases the bounds of the distribution are as large or larger than the magnitude of the median of the distribution indicating a significant overlap of the positron and electron distributions. Without charge identification in the calorimeter, the window to be considered must therefore extend between the outer bounds of both the electron and positron distributions; this quantity is shown in the table 6.2 under the entry “width”. It can be seen that (with the exception of the case of candidates with $p_T=5$ GeV and a calorimeter cut of $E=25$ GeV, ie the highly-boosted case) the size of the window required falls with the increasing complexity of the calorimeter trigger algorithm reflecting the improved measurement of the “seed” point. The angular width of the window remains approximately constant with increasing tracker radius, due to the limited separation of the layers compared to the radius of the calorimeter.

These values may be compared to previous studies based on calorimeter-seeding using ECAL crystal level information[108]; these studies suggested that the required angular window sizes were of the order 0.05, 0.043 and 0.036 radians for radii of 4, 7 and 11 cm, respectively (based on the channel $H \rightarrow ZZ \rightarrow e^+e^-e^+e^-$ at $M_H = 200$ GeV). These values are, respectively, 3.6, 3.7 and 4.4 times smaller than the values predicted using the weighted method of calorimeter-seeding for 50 GeV candidates, this being consistent with the calorimeter having a 5-fold better position resolution in the previous study.

For the scenarios listed above, table 6.3 and figure 6.5 summarize the distributions of separation in the z -direction of hits in the pixel tracker from the straight track hypothesis. As expected, the candidate momentum has no noticeable effect on either the median separation or on the upper or lower bounds of the distributions; in all cases the median separation is close to zero, while the bounds are approximately ± 13 cm. Similarly, the positive and negative charge cases produce no noticeable differences. Increasing the complexity of calorimeter-seeding methods provides only negligible improvement, indicating that the distribution shape is dominated not by uncertainty in the calorimeter hit position, but by uncertainty in the true interaction point. This is in agreement with what is already known: while the beam-size is tightly constrained in the x - y plane, the size of the luminous region in the z -direction is large, the interaction profile in the z -direction being typically described as Gaussian (with a standard deviation, σ , of between 4.5 cm and 5 cm). The “width” is again shown and represents the distance between the maximum upper bound and the minimum lower bound. In all cases the window-width is around 26 cm, consistent with the $\pm 3\sigma$ length of the luminous region. These results are also consistent with the window size in the z -direction found in the previous study [108], further confirming that the window-size is determined by the size of the luminous region and not by the calorimeter resolution.

Approximating each pixel layer as a cylinder of length 50 cm, the window sizes calculated using the weighted method correspond to rectangular windows of approximately 1.5% of the tracker area for the case of 50 GeV candidates and 5.5% of the tracker area for 5 GeV candidates. Assuming that the upgraded calorimeter trigger identifies eight regions of interest^{viii}, such a scheme would require reading out up to 12% of the tracker for the case of 50 GeV candidates and up to 45% for 5 GeV candidates. While reducing the area to be read out by an order of magnitude may seem like a significant reduction, it is actually insufficient to reduce the data rate. As the pixel tracker is zero-suppressed, the data rate is directly proportional to occupancy and, as such, the effect of reducing the

^{viii}eight is the number of isolated + non-isolated e/γ candidates found by the current system

tracker area to be read out is completely cancelled by the ten-fold increase in occupancy expected from the luminosity upgrade: therefore, for the same reason that the current tracker may not be read out every bunch-crossing, an upgraded pixel tracker may not be selectively read out every bunch-crossing. It could be argued that the speed of optical links will be increased in an upgrade from the current 0.8 Gbit/s to multi-gigabit links, but the highest payload rate considered implementable in radiation-hard electronics is 3.36 Gbit/s^{ix}, this providing only a factor of four increase in the link speed, corresponding to a four-fold saving in the required number of optical links; this being insufficient improvement to handle the data associated with every bunch-crossing.

6.3 A Self-contained Tracking Trigger for CMS at the SLHC

In order to evaluate the magnitude of the problems faced in reading out the tracker at SLHC, conditions of 200 pile-up events were simulated in CMSSW “full” simulation and the rate and densities of tracker hits recorded (figure 6.6(b)).

As seen in figure 6.6(b), under conditions of 200 pile-up events, the density of simulated hits per crossing in the detector is around 2.5 cm^{-2} at radius 10 cm: this figure is consistent with, but slightly lower than, previous studies[108] (3 cm^{-2} at $r=11 \text{ cm}$) and with extrapolation from the track-density expected in the current tracker^x(4 cm^{-2} at $r=10 \text{ cm}$).

^{ix}Rather, it is envisaged that the maximum link speed will be 4.8 Gbit/s, but shared between data and control pathways with a maximum of 3.36 Gbit/s for data [109].

^xThe expected track-density in the current pixel system is 0.4 hits per cm^2 at a radius of 10 cm [42]. A 10-fold increase in luminosity would be expected to give a corresponding 10-fold increase in tracker occupancy, corresponding to a track-density of 4 cm^{-2} .

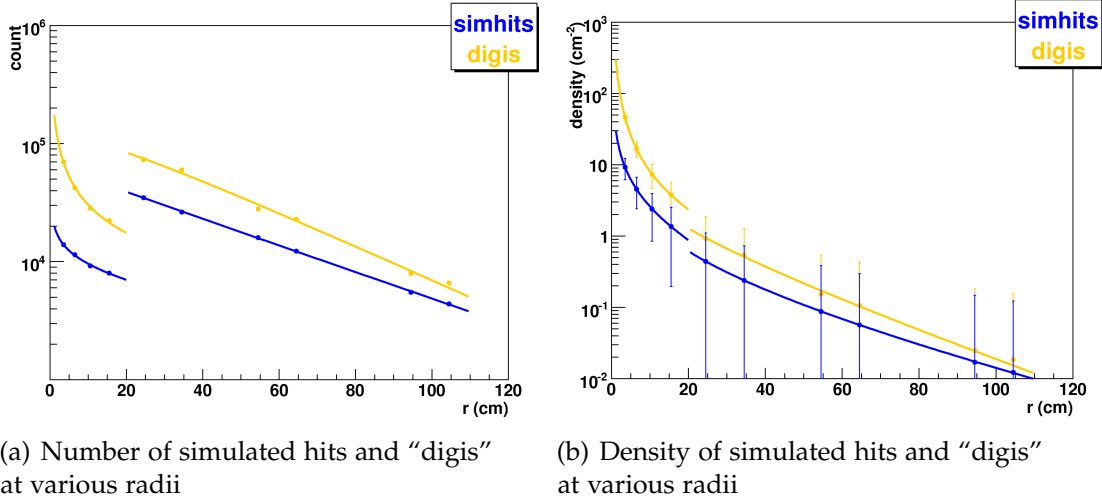


Figure 6.6: Number and density of simulated hits and **digis** (digitized hits) plotted against radius for 200 pile-up events (without “out-of-time” pile-up) using the “Strawman-B” SLHC-like tracker geometry. The discontinuity at $R=20$ cm is due to the change in topology between the inner 4 pixel layers and the outer layers, such that the outer layers have twice the surface area of silicon per unit arc length as compared to the inner layers. The density plots are also discontinuous, due to the different pseudo-rapidity coverage of the inner and outer layers.

6.3.1 Reducing the tracker data volume

If tracker data is to be included in triggering, it must be read out every bunch-crossing, this precluding an analogue readout scheme (as is used in the current outer tracker) and requiring on-detector zero-suppression. Assuming a 16-bit addressing scheme, the quoted hit density of 2.5 cm^{-2} corresponds to an average raw data rate of 40 bit/cm^2 per crossing or $1.6 \text{ Gbit/cm}^2/\text{s}$. Optical links, however, require an error detection scheme such as 8b/10b encoding (20% overhead), or an error correction scheme such as Hamming encoding [110] (75% overhead) and, allowing for the insertion of “extra” information onto the fibre, this results in a bandwidth of $3\text{--}4 \text{ Gbit/cm}^2/\text{s}$. Although radiation-hard transceivers in the multi-gigabit range are now thought feasible, the physical space and enormous power required to read out even one layer would certainly exclude such a scheme. Furthermore, as seen in figure 6.6(b), due to charge sharing, the density of digitized hits is 3-fold greater than the density of simulated hits, this either raising the bandwidth requirement by a factor of 3 or requiring on-detector clusterization, which itself, would increase the power requirements of the detector.

Even worse, when considering schemes to reduce the data rate, it must be remembered that these signals are not “noise” or a detector effect, but genuine data with physical significance.

Apart from vertexing of b - and τ -jets, the tracker plays two other roles in the current higher level trigger: as a classifier and to provide a measurement of momentum. In the former role, the presence (or lack) of tracker hits in a window around a “seed” point is used to distinguish electrons from photons, primary muons from muons generated by hadronic decays, and τ -jets from QCD jets. In the latter role, the tracker’s measurement of momentum allows for tight cuts to be placed on the transverse momentum of muon candidates and on the content of a τ -jet. The “softest” objects which are currently considered in the higher level trigger are the constituents of τ -jets, which need only pass a 1 GeV cut, while the thresholds for electrons and muons are $\mathcal{O}(10 \text{ GeV})$ and that of photons, $\mathcal{O}(100 \text{ GeV})$. This should be contrasted to figure 6.7, where it can be seen that, both at the point of production and at the outer surface of the tracker volume, the vast majority of tracks are below 1 GeV; the bin 0-1 GeV contains an order of magnitude more hits than the bin 1-2 GeV and 3 orders of magnitude more hits than the bin 5-6 GeV. By excluding these low- p_T tracks from a readout scheme, not only could the data rate be reduced by several orders of magnitude, but also the complexity of performing track association would be greatly simplified by the reduced number of combinatorials.

Two approaches are possible in order to distinguish tracks with high transverse momentum from those with low transverse momentum, both dependent on the fact that the curvature of a charged particle’s track is inversely proportional to its transverse momentum.

The first approach is based on the principle that, if a track crosses a pixelated sensor at a shallow angle to the surface, charge will be deposited across many pixels, resulting in wide clusters; tracks crossing nearly perpendicular to the surface will, however, leave only narrow clusters. In principle, tracks with high transverse momentum originating at the interaction point should be distinguishable

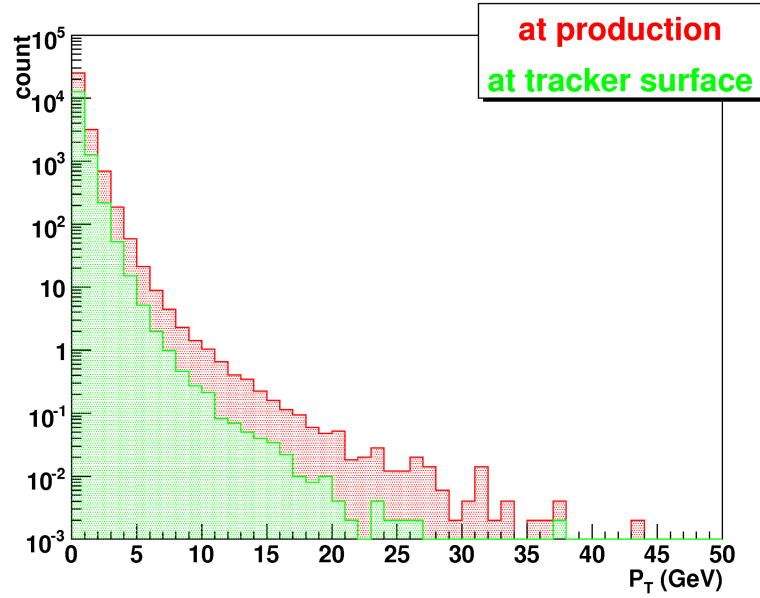


Figure 6.7: Particle count per event related to track- p_T for 200 pile-up events. The red curve includes both primary and decay particles, whilst the green curve includes only those particles which escape the tracker volume.

from the background of low-momentum tracks, as the high momentum tracks propagate approximately radially. It is unclear at this stage, however, whether this approach will be feasible under SLHC pile-up conditions or whether sufficient reduction in rate may be achieved. Furthermore, it is known that this method cannot be performed in the inner tracker (where the cluster size is always small and, so, there is little ability to discriminate) and would need to be performed in the outer region of the tracker, where many of the benefits of tracking (such as separating primary vertices) would be limited.

The second approach is loosely based on the traditional approach to tracking, whereby multiple points are used to define a curve and the transverse momentum deduced from the curvature. Traditionally, the curvature is measured by calculating the sagitta, a process requiring a long track segment and thus widely-spaced points. Communication of data is required between different detector layers to perform such a reconstruction; as tracker points provide neither measurement of momentum nor constraint of the interaction point, the association of hits between layers requires large windows and, under conditions of high

occupancy, this results in a large number of misassociated combinations (figure 6.8).

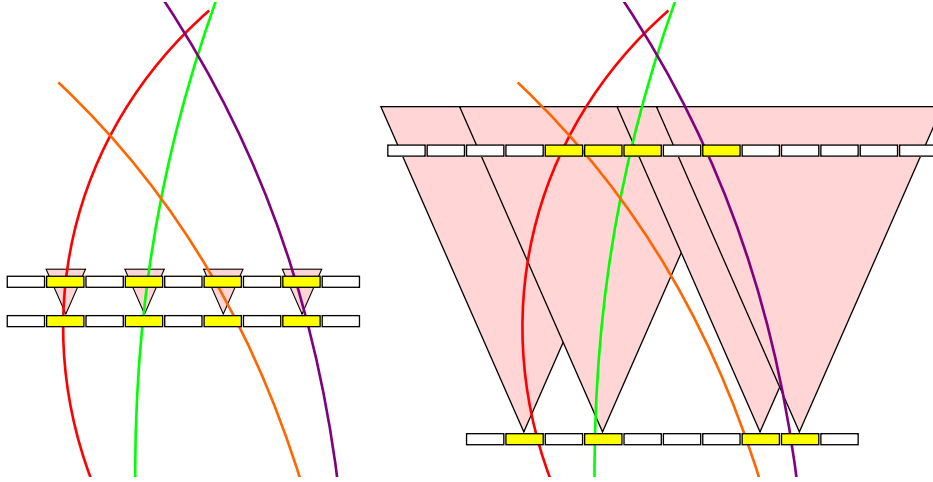


Figure 6.8: Tracker Combinatorics under high pile-up conditions. For closely spaced points, only small regions need be considered giving little probability of overlap between tracks. Between widely spaced layers, the distances over which tracks bend is large and, under conditions of high occupancy, the probability of tracks overlapping is high. In order to reconstruct a track, large regions must be searched for a matching hit; this results in a large number of fake associations. These fake tracks must be eliminated using complex filtering algorithms.

Multiple-pass reconstruction methods must be applied to eliminate these fake tracks. This is currently achieved by means of Kalman filtering (or Gaussian-Sum Filtering to compensate for non-linear energy loss [111]), but such an approach is too slow and too computationally-intensive for use in reducing the rate of tracker readout.

Measurement of the angle of the track, relative to the radial vector, provides an alternative to measuring the sagitta: at all points, highly-curving tracks form a large angle to the radial vector through that point, whilst tracks that bend little form only a small angle. While such a method provides inferior resolution of momentum as compared to the sagitta method, the advantage of measuring the angle, rather than the sagitta, is that it is localized, being performed at a single point. While conceptually the process involves a point and an angle, in practice — to make a measurement of track angle — two points are still required: as the two points need not be widely-spaced, however, the probability that they originate from different tracks can be minimized by bringing the two layers close

together and, as such, the need for complex filtering algorithms is eliminated. By imposing a restriction on the maximum acceptable track crossing-angle, the probability of finding erroneous tracks is reduced further and an inherent cut on transverse momentum introduced; a cut which limits the track to a small crossing-angle is equivalent to requiring the track to have high transverse momentum (figure 6.9). If the two layers could be brought sufficiently close together that they may be built into a single “hard-wired” module, then the module’s self-contained ability to eliminate low- p_T tracks is a most attractive feature, as it allows the much-needed reduction in data rate to be performed internally, thereby minimizing the volume of data produced.

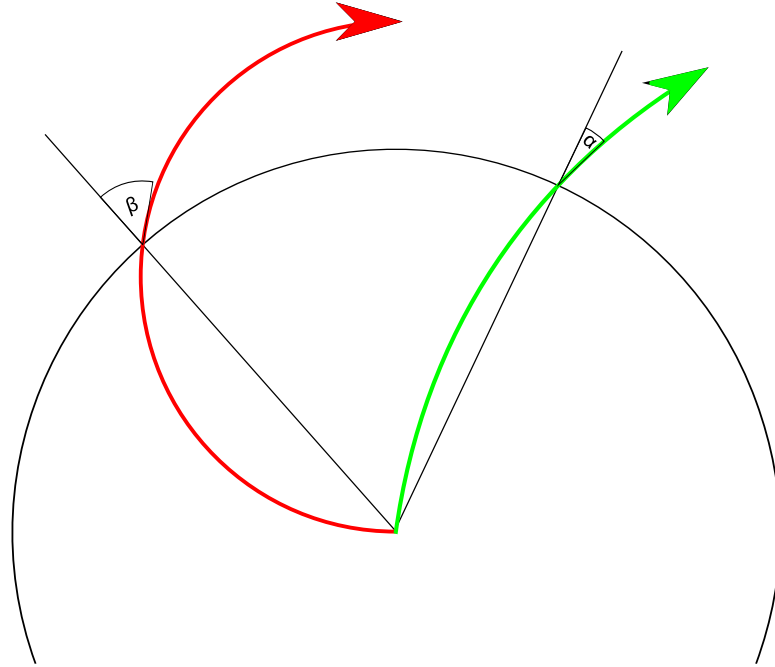


Figure 6.9: Demonstration of the equivalence of an angular cut and a p_T -cut. A high- p_T track (green) crosses at a shallow angle (α) to the radial vector at all radii while a low- p_T track (red) crosses at a larger angles (β).

6.3.2 The theoretical basis of the stacked tracking concept

The transverse momentum, p_T , may be related to the radius of curvature by the equation:

$$p_T = 10^{-9} \cdot cB \cdot r_{curve} \quad (6.1)$$

where r_{curve} is measured in metres, p_T is in GeV/c, B is the magnetic field strength in Tesla and c is the speed of light in m/s.

By fundamental geometry (illustrated in appendix B), the angle, ϕ , between the tangent and the chord (labelled α and β in figure 6.9) is defined by the relationship:

$$\sin(\phi) = \frac{0.5 \cdot r_{chord}}{r_{curve}} \quad (6.2)$$

Substituting for r_{curve} using equation 6.1 and relabelling r_{chord} as r_{layer} in recognition that r_{chord} is the radial distance of the sensor layer from the interaction point results in:

$$\sin(\phi) = \frac{0.5 \times 10^{-9} \cdot cB \cdot r_{layer}}{p_T} \quad (6.3)$$

or, substituting for c with the approximation 3×10^8 m/s and for B with the nominal strength of the CMS magnetic field, $4T$,

$$\sin(\phi) \approx \frac{0.6 \cdot r_{layer}}{p_T} \quad (6.4)$$

Although the conceptual design of this algorithm is in the form of a point plus an angle, any physical implementation is necessarily in the form of two points. As the intention of this design is to bring the two points close together to eliminate mismatched tracks, the separation in $r\phi$ between the two points, s , may be approximated using the relationship:

$$\tan(\phi) = \frac{s}{\Delta r} \quad (6.5)$$

where Δr is the radial separation between the two points.

As, for the purpose of triggering, we are interested only in high- p_T tracks, we may restrict ourselves to the case where ϕ is small, and so use the small angle approximation, $\sin(\phi) \approx \tan(\phi) \approx \phi$, to derive the relationship

$$s \approx \frac{0.6 \cdot r_{layer} \cdot \Delta r}{p_T} \quad (6.6)$$

If we consider the two points with a radial spacing of 2 mm centred at 25 cm along a track of momentum $2 \text{ GeV}^{\text{xi}}$, then from 6.6 the expected deviation of

^{xi}We drop the GeV/c notation for convenience here since the explicit reference to the speed of light in equations 6.1 and 6.3 has been absorbed into the constant term.

such a track in the $r\phi$ direction is $150\ \mu\text{m}$. This figure is similar in scale to the pixel width of the current pixel tracker ($100\ \mu\text{m}$) and to the strip width of the current inner strip tracker (80 and $120\ \mu\text{m}$) and, as such, a pair of sensors with an $r\phi$ resolution of this scale should (according to these calculations) be able to distinguish high- p_T tracks from low- p_T tracks.

In the z -direction, the region over which hits must be associated is constrained, albeit relatively loosely, by the luminous region around the interaction point (figure 6.10). For any hit in the inner sensor, the hypothesis that the track origi-

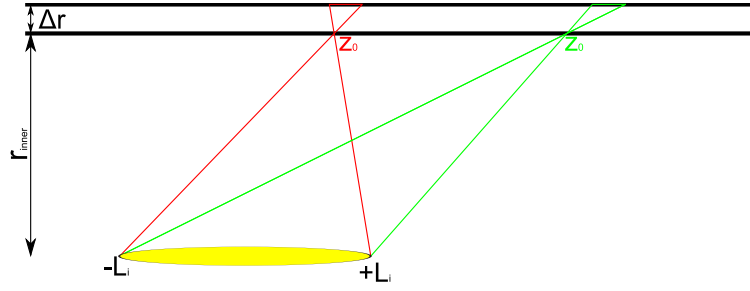


Figure 6.10: The size and offset of the track-finding window in the z -direction. By constraining the track to originate from the luminous region and to pass through a particular tracker point, the allowed location of a hit in any other layer (and thus the region which must be searched) is limited.

nated within the interaction point can be used to define a window in the outer sensor within which the second hit must lie. As the magnetic field at CMS is oriented along the z -axis and exceedingly uniform throughout the tracking volume, tracks do not bend in the rz -plane and, ignoring material effects, may be treated as straight lines. As such, by fundamental trigonometry, the bounds on the window, z_1 and z_2 , may be calculated to be

$$\frac{z_1 - L_I}{r_{inner} + \Delta r} = \frac{z_0 - L_I}{r_{inner}} \quad (6.7)$$

$$\frac{z_2 + L_I}{r_{inner} + \Delta r} = \frac{z_{inner} + L_I}{r_{inner}} \quad (6.8)$$

where $\pm L_I$ are the bounds of the luminous region and, subsequently, the length of such a window may be calculated to be

$$|z_2 - z_1| = 2L_I \frac{\Delta r}{r_{inner}} \quad (6.9)$$

For the LHC, the luminous region is adequately described by a Gaussian of $\sigma = 5$ cm and, as such, for simplicity we may substitute for L_I with the 3σ limit of 15 cm,

$$|z_2 - z_1| = 0.3 \frac{\Delta r}{r_{inner}} \quad (6.10)$$

Again, considering two points with a radial spacing of 2 mm centred at 25 cm along a track originating from anywhere within the luminous region gives a window length of approximately 2.4 mm. This should be compared to the current pixels which have a length of $150 \mu\text{m}$ in z and the silicon strips which are of the order 10 cm; the resolution required in z is more like that of a pixel system than that of a strip tracker. It can also be seen that the size of the window is independent of η which, although slightly counter-intuitive, is most clearly demonstrated by considering the (hypothetical) case where $\Delta r = r_{inner}$; in which case the window is shown, by equation 6.9, to be an exact mirror of the luminous region.

It may also be shown that the distance in the z -direction of the centre of the window, $\bar{z}_{window} = \frac{1}{2}(z_1 + z_2)$, from the inner hit is given by

$$|\bar{z}_{window} - z_{inner}| = \Delta r \frac{z_{inner}}{r_{inner}} \quad (6.11)$$

Identifying $\frac{r_{inner}}{z_{inner}}$ as $\tan(\theta)$ and using the tangent double-angle identity and the definition of the pseudorapidity^{xii}, η ,

$$|\bar{z}_{window} - z_{inner}| = \frac{\Delta r}{\tan(\theta)} \quad (6.12)$$

$$= \Delta r \cdot \sinh(\eta) \quad (6.13)$$

The current tracker covers the range $-2.5 < \eta < 2.5$ and, in order to be useful for forward triggering, the tracking trigger should cover at least this range. Thus, again considering a radial spacing of 2 mm, at the extremes of η the window offset is approximately 1.2 cm. In order to avoid inefficiencies, such an offset must be accounted for both in the algorithm for pairing hits and in the physical structure of the detector (figure 6.11).

^{xii} $\tan(\theta) = \frac{2 \tan(\frac{\theta}{2})}{1 - \tan^2(\frac{\theta}{2})}$ and $\eta = -\ln(\tan(\frac{\theta}{2}))$

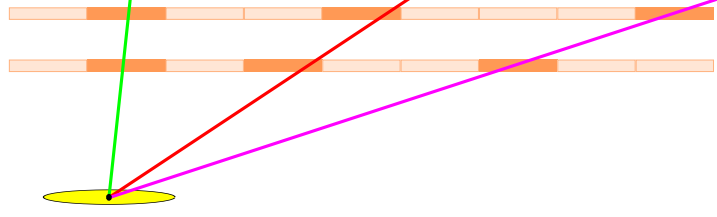


Figure 6.11: Schematic representation of the necessary sensor offset in the z -direction with increasing η . The scales here are exaggerated.

While bringing two sensor layers close together allows for (relatively) unambiguous matching of points and thus to place a cut on momentum, the penalty to be paid is the loss in resolution of momentum. An error analysis on equation 6.6 implies that, given a perfectly understood detector (that is, no error on Δr and r),

$$\frac{\sigma_s}{s} = \frac{\sigma_{p_T}}{p_T} \quad (6.14)$$

where σ_s and σ_{p_T} are the respective errors on s and p_T . Rearranging and substituting for s gives

$$\sigma_{p_T} = \frac{1}{0.6 \cdot r_{layer}} \cdot \frac{\sigma_s}{\Delta r} \cdot p_T^2 \quad (6.15)$$

Taking for σ_s a nominal value of $150 \mu\text{m}$ and using the same geometric parameters as previously leads (for a track with $p_T = 2 \text{ GeV}$) to an error in the measurement of transverse momentum of 2 GeV (or 100%). In order to keep the resolution to 100% at 10 GeV , the $r\phi$ position resolution must be lowered to $30 \mu\text{m}$. Noting that s is, itself, a composite measurement implies the $r\phi$ position resolution for each sensor must be a factor $\frac{1}{\sqrt{2}}$ better than this.

Similarly, performing an error analysis on equation 6.13 and assuming that the z position resolutions of the inner and outer points are identical, the resolution in η is given by

$$\sigma_\eta^2 = 2 \left(\frac{\partial \eta}{\partial z_{inner}} \right)^2 \sigma_z^2 \quad (6.16)$$

or

$$\sigma_\eta = \sqrt{2} \cdot \frac{\sigma_z}{\Delta r} \cdot \text{sech}(\eta) \quad (6.17)$$

where σ_z and σ_η are the respective errors on z_{inner} (and also z_{outer}) and η . Taking a nominal z position resolution of 2 mm and the same geometric parameters

as used previously, leads to an uncertainty in $\sigma_\eta \approx 1.4$ at $\eta = 0$, falling to $\sigma_\eta \approx 0.23$ at $\eta = \pm 2.5$. The improved resolution in the forward direction is due to the increased path length between the inner and outer points, which has the effect of reducing the angle subtended by the error on measurement of the outer point. In order to achieve an η resolution equivalent to that of the calorimeter ($\sigma_\eta = 0.087$ in the barrel region), such a tracker would require a z position resolution of better than $120 \mu\text{m}$ on each measurement.

It is clear, therefore, that in using a pair of closely spaced sensors, the intrinsic resolution required for precision measurements of p_T and η is prohibitive. From equations 6.15 and 6.17 it is also clear that the value of $\Delta r \ll 1 \text{ m}$ is strongly responsible for the poor resolution and that the resolution may be improved by allowing $\Delta r \rightarrow 1 \text{ m}$. Such a scheme would, however, result in larger search windows, thereby increasing the number of hit-pair combinations and this, entirely, negating the fundamental argument for closely-spaced sensors.

To retain a reduction in data-rate, through a cut on track angle, and to achieve good p_T and η resolutions, therefore, would require at least four sensor layers arranged as sets of closely-spaced pairs, with the sets themselves being widely-spaced.

If we consider the minimal design of two sets of sensors (four sensors in total) and consider that each set can provide only a position measurement and no directional measurement in the $r\phi$ direction^{xiii}, then there is insufficient information available to fit a unique curve; instead, we must use the beam-spot as an additional constraint. When making a measurement of momentum using widely-spaced points, the assumptions used in deriving, for example equation 6.4, are no longer valid and must be replaced by a more complete, geometric calculation of the radius of curvature (r_{curve}) and hence of the transverse momentum.

^{xiii} As the stacks are designed to cut tracks with low p_T , only tracks which cross approximately radially are considered and it is for these tracks that the angular resolution is worst.

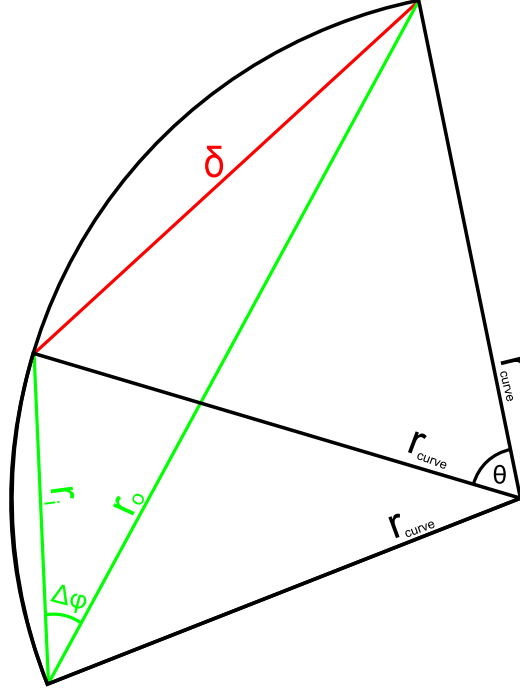


Figure 6.12: Geometric method of measuring the r_{curve} , and thus p_T , from two tracker points and the nominal interaction point. The lengths r_i and r_o are the radii of the inner and outer tracker points, respectively, and $\Delta\phi$ is the angular separation of the two point.

Using the geometry indicated in figure 6.12, the relationships

$$\theta = 2 \cdot \Delta\phi \quad (6.18)$$

$$\delta^2 = r_i^2 + r_o^2 - 2 \cdot r_i r_o \cdot \cos(\Delta\phi) \quad (6.19)$$

$$\frac{\delta}{2} = r_{curve} \cdot \sin(\Delta\phi) \quad (6.20)$$

may be combined, together with equation 6.1, to yield

$$p_T = 10^{-9} \cdot cB \cdot \frac{\sqrt{r_i^2 + r_o^2 - 2 \cdot r_i r_o \cdot \cos(\Delta\phi)}}{2 \cdot \sin(\Delta\phi)} \quad (6.21)$$

$$\approx 0.6 \cdot \frac{\sqrt{r_i^2 + r_o^2 - 2 \cdot r_i r_o \cdot \cos(\Delta\phi)}}{\sin(\Delta\phi)} \quad (6.22)$$

If more than three points are known then the problem is overconstrained and must be solved using an alternative method, for instance, by using a least-squares or “average of intersections” approach [112].

Whilst associating hits between two closely-spaced, hard-wired sensors may be performed on-detector to reduce the data rate, the operations required to form

track-like objects are too certainly complex and would need to be performed off-detector in the level-1 trigger. In the chapters which follow, the following nomenclature will be used; a pair of hits in closely-spaced sensors will be called a “local stub”, this is the type of object which may be formed on-detector. The representation of this object within the global reference frame of the detector will be called a “global stub”, this type of object would be formed at the level-1 trigger. Track-like objects constructed from global stubs will be called “tracklets”, these would also be formed at the level-1 trigger.

Chapter 7

A Monte-Carlo study of Tracking Triggers

“In theory, there is no difference between theory and practice. But, in practice, there is.”

- Jan L.A. van de Snepscheut (1953 - 1994)

7.1 Introduction

While the geometric calculations presented in the previous chapter indicate that the principle of the stacked tracker is sound, there are many practical aspects that cannot be taken into consideration by the simplifying assumptions. To better understand the physical processes which could affect the performance of a tracking trigger, Monte Carlo simulation studies were necessary.

7.1.1 CMSSW

The CMS experiment has a comprehensive software framework for the Monte Carlo simulation of the experiment called “CMSSW”. The software embeds functionality for every stage of simulation, all the way through to analysis, as a series of tasks within a framework (figure 7.1).

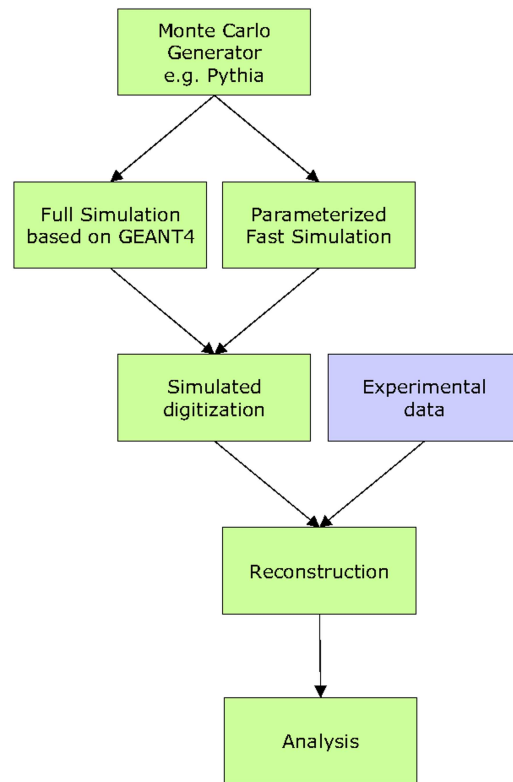


Figure 7.1: Data flow within the CMSSW software package. The software elements are indicated in green and, for comparison, the experimental data is indicated in purple. Diverging or converging data pathways indicate that either, but not both, pathways may be taken. GEANT4 [113] and Pythia [114] are software external to the CMS experiment.

Although the arrows in figure 7.1 indicate a linear flow of data between each of the processing steps, data flow is actually mediated through an “event” object which may be written to file and, as such, each step of the process may occur at widely separated times or locations. While more traditional simulation and analysis methods use a distinct executable to perform each step, CMSSW uses a software bus model, in which there exists a single executableⁱ that dynamically loads plug-in modules to perform each task. The executable itself performs no simulation or analysis and its sole purpose is to perform “book-keeping” operations: to interpret the user’s configuration file, to manage data objects and to load and execute each task as it is required. In this sense, figure 7.1 is misleading as each task may represent many tens, or even hundreds, of plug-in modules rather than a single executable. This approach has two main advantages: first, as each module is self-contained and very small relative to the whole

ⁱIn CMSSW, this executable is called cmsRun

framework, division of tasks between developers is greatly simplified. Secondly, each module may be compiled without recompilation of the entire project, this significantly reducing the time taken between revision cycles.

The two simulation paths, “full” simulation and “fast” simulation, represent alternative approaches to simulation of the detector. In the “full” simulation, particle tracks are propagated, using a delta-time approach, through both a field map of the detector’s magnetic field and through a full material model of the detector, taking into account “all” physical processes to create a list of particle decays and interactions, together with a map of energy deposition in the finest possible detail. Such an approach is, however, computationally intensive and hence very slow. The “fast” simulation takes an alternative approach: rather than simulating every detail of a particle’s behaviour for all time, it relies on an assumption that the path of a track between interactions is deterministic, that similar interactions will not vary significantly in their behaviour (and that small variations may be taken into account by a stochastic correction), and that decays, bremsstrahlung, photon conversions and Coulomb scattering may all be treated as random modifications to a track. While the track “modifying” interactions and energy loss through ionization may be modelled in this way, nuclear interactions can not: instead, the “fast” simulation uses a look-up table approach with, as a reference, interactions modelled in the “full” simulation. By using these simplifying assumptions, the “fast” simulation has been shown to run between 2 to 3 orders of magnitude faster than the “full” simulation. This is a significant advantage in modelling, as datasets containing in the order of 1 billion events are required to understand the backgrounds and systematic errors of the experiment; a value beyond the capacity of the “full” simulation [115]. The “fast” simulation parameters are currently tuned so that the “fast” simulation matches the “full” simulation, however, when experimental data becomes available, the parameters will be adjusted to reflect the true behaviour of the detector.

As simulations of the SLHC require between 10 and 20 times more interactions per crossing than simulations of LHCⁱⁱ, the use of “full” simulation would be

ⁱⁱthis factor is dependent on the upgrade scenario

prohibitively slow, taking up to two hours per eventⁱⁱⁱ and, as such, the use of “fast” simulation becomes a requirement rather than just an option. Before studies were made into tracking triggers at SLHC, it was decided that the validity of the “fast” simulation should be independently verified by a comparison with the “full” simulation.

7.2 Validation of “Fast” Simulation against “Full” Simulation

Because “full” simulation is slow in modelling events, large numbers of pile up events are created and stored centrally in files so that, instead of generating and simulating all pile-up events for each Monte Carlo crossing, the user may simply superimpose the ready-made background onto a signal event, without the necessity of regenerating events that are not actually of direct interest. As these background data sets were generated for the standard geometry, however, they cannot be used with an SLHC-type geometry and a decision was made that, instead of generating large background samples for a new geometry, the standard geometry would be used for the purpose of comparing the two simulation methods. As the “fast” simulation generates and models everything in the event, including pile-up, at runtime there is no equivalent consideration for the “fast” simulation.

Two measures were used in the comparison of the “fast” and “full” simulations: the number of simulated hits (“simhits”) and the number of digitized hits (“digis”). The former object type is produced when any Monte Carlo track crosses a sensing element, whereas the latter corresponds to the digitized response to such a hit, as would be expected from the detector (figure 7.1). For conditions of 22 minimum-bias events (equivalent to a luminosity of $10^{34} \text{ cm}^{-2}\text{s}^{-1}$), the numbers of each hit-type were recorded separately in each barrel layer, and

ⁱⁱⁱ ~ 40 s per minimum bias event [116], which corresponds to ~ 133 minutes for 200 minimum bias events.

Default settings							
		FullSim (with Out-Of-Time Pile-Up)		FastSim (300MeV p_T cut, looping tracks culled)		Ratio of FullSim/FastSim	
Layer number	Radius (cm)	“Simhits” (cm^{-2})	“Digis” (cm^{-2})	“Simhits” (cm^{-2})	“Digis” (cm^{-2})	“Simhits”	“Digis”
0	4.38	$3.22 \times 10^0 \pm 1.59 \times 10^{-2}$	$2.52 \times 10^0 \pm 2.49 \times 10^{-2}$	$2.62 \times 10^{-1} \pm 2.73 \times 10^{-3}$	$1.27 \times 10^0 \pm 1.32 \times 10^{-2}$	12.29	1.98
1	7.29	$1.69 \times 10^0 \pm 8.39 \times 10^{-3}$	$1.08 \times 10^0 \pm 1.08 \times 10^{-2}$	$1.28 \times 10^{-1} \pm 1.36 \times 10^{-3}$	$4.72 \times 10^{-1} \pm 5.02 \times 10^{-3}$	13.18	2.28
2	10.16	$1.07 \times 10^0 \pm 5.48 \times 10^{-3}$	$6.19 \times 10^{-1} \pm 6.21 \times 10^{-3}$	$7.80 \times 10^{-2} \pm 8.39 \times 10^{-4}$	$2.46 \times 10^{-1} \pm 2.64 \times 10^{-3}$	13.71	2.51
3	25.61	$1.78 \times 10^{-1} \pm 8.89 \times 10^{-4}$	$5.43 \times 10^{-1} \pm 2.67 \times 10^{-3}$	$1.24 \times 10^{-2} \pm 1.35 \times 10^{-4}$	$4.15 \times 10^{-2} \pm 3.26 \times 10^{-4}$	14.30	13.11
4	33.99	$1.12 \times 10^{-1} \pm 5.84 \times 10^{-4}$	$3.72 \times 10^{-1} \pm 1.86 \times 10^{-3}$	$7.81 \times 10^{-3} \pm 8.62 \times 10^{-5}$	$3.32 \times 10^{-2} \pm 2.40 \times 10^{-4}$	14.32	11.22
5	41.89	$7.68 \times 10^{-2} \pm 4.17 \times 10^{-4}$	$2.07 \times 10^{-1} \pm 1.07 \times 10^{-3}$	$4.80 \times 10^{-3} \pm 5.57 \times 10^{-5}$	$1.92 \times 10^{-2} \pm 1.36 \times 10^{-4}$	15.99	10.83
6	49.86	$5.46 \times 10^{-2} \pm 2.99 \times 10^{-4}$	$1.53 \times 10^{-1} \pm 7.94 \times 10^{-4}$	$3.36 \times 10^{-3} \pm 4.01 \times 10^{-5}$	$1.60 \times 10^{-2} \pm 1.03 \times 10^{-4}$	16.25	9.59
7	60.80	$4.16 \times 10^{-2} \pm 2.27 \times 10^{-4}$	$1.09 \times 10^{-1} \pm 5.87 \times 10^{-4}$	$1.87 \times 10^{-3} \pm 2.28 \times 10^{-5}$	$8.68 \times 10^{-3} \pm 6.78 \times 10^{-5}$	22.21	12.61
8	69.20	$3.20 \times 10^{-2} \pm 1.81 \times 10^{-4}$	$8.41 \times 10^{-2} \pm 4.63 \times 10^{-4}$	$1.35 \times 10^{-3} \pm 1.70 \times 10^{-5}$	$7.17 \times 10^{-3} \pm 5.20 \times 10^{-5}$	23.78	11.72
9	78.00	$2.45 \times 10^{-2} \pm 1.41 \times 10^{-4}$	$6.39 \times 10^{-2} \pm 3.67 \times 10^{-4}$	$9.71 \times 10^{-4} \pm 1.28 \times 10^{-5}$	$6.13 \times 10^{-3} \pm 4.16 \times 10^{-5}$	25.24	10.44
10	86.80	$1.95 \times 10^{-2} \pm 1.13 \times 10^{-4}$	$5.01 \times 10^{-2} \pm 2.93 \times 10^{-4}$	$7.14 \times 10^{-4} \pm 9.72 \times 10^{-6}$	$5.34 \times 10^{-3} \pm 3.27 \times 10^{-5}$	27.36	9.38
11	96.50	$1.57 \times 10^{-2} \pm 8.92 \times 10^{-5}$	$5.13 \times 10^{-2} \pm 2.84 \times 10^{-4}$	$5.19 \times 10^{-4} \pm 7.36 \times 10^{-6}$	$6.69 \times 10^{-3} \pm 3.26 \times 10^{-5}$	30.29	7.67
12	108.00	$1.24 \times 10^{-2} \pm 7.01 \times 10^{-5}$	$3.86 \times 10^{-2} \pm 2.12 \times 10^{-4}$	$3.65 \times 10^{-4} \pm 5.44 \times 10^{-6}$	$6.06 \times 10^{-3} \pm 2.59 \times 10^{-5}$	33.83	6.37
After Modification							
		FullSim (no Out-Of-Time Pile-Up)		FastSim (no p_T cut, looping tracks included)		Ratio of FullSim/FastSim	
Layer number	Radius (cm)	“Simhits” (cm^{-2})	“Digis” (cm^{-2})	“Simhits” (cm^{-2})	“Digis” (cm^{-2})	“Simhits”	“Digis”
0	4.38	$7.75 \times 10^{-1} \pm 7.85 \times 10^{-3}$	$2.38 \times 10^0 \pm 2.39 \times 10^{-2}$	$3.47 \times 10^{-1} \pm 3.47 \times 10^{-3}$	$1.63 \times 10^0 \pm 1.63 \times 10^{-2}$	2.23	1.46
1	7.29	$4.06 \times 10^{-1} \pm 4.19 \times 10^{-3}$	$1.01 \times 10^0 \pm 1.04 \times 10^{-2}$	$1.82 \times 10^{-1} \pm 1.83 \times 10^{-3}$	$6.53 \times 10^{-1} \pm 6.60 \times 10^{-3}$	2.23	1.55
2	10.16	$2.54 \times 10^{-1} \pm 2.66 \times 10^{-3}$	$5.71 \times 10^{-1} \pm 6.04 \times 10^{-3}$	$1.16 \times 10^{-1} \pm 1.19 \times 10^{-3}$	$3.70 \times 10^{-1} \pm 3.78 \times 10^{-3}$	2.19	1.54
3	25.61	$4.23 \times 10^{-2} \pm 4.43 \times 10^{-4}$	$1.46 \times 10^{-1} \pm 1.39 \times 10^{-3}$	$2.07 \times 10^{-2} \pm 2.19 \times 10^{-4}$	$6.96 \times 10^{-2} \pm 6.11 \times 10^{-4}$	2.05	2.09
4	33.99	$2.66 \times 10^{-2} \pm 2.86 \times 10^{-4}$	$1.01 \times 10^{-1} \pm 9.49 \times 10^{-4}$	$1.28 \times 10^{-2} \pm 1.39 \times 10^{-4}$	$5.05 \times 10^{-2} \pm 4.20 \times 10^{-4}$	2.08	2.00
5	41.89	$1.81 \times 10^{-2} \pm 2.02 \times 10^{-4}$	$5.72 \times 10^{-2} \pm 5.48 \times 10^{-4}$	$7.87 \times 10^{-3} \pm 8.86 \times 10^{-5}$	$2.82 \times 10^{-2} \pm 2.31 \times 10^{-4}$	2.30	2.02
6	49.86	$1.29 \times 10^{-2} \pm 1.49 \times 10^{-4}$	$4.33 \times 10^{-2} \pm 4.08 \times 10^{-4}$	$5.48 \times 10^{-3} \pm 6.45 \times 10^{-5}$	$2.25 \times 10^{-2} \pm 1.77 \times 10^{-4}$	2.36	1.93
7	60.80	$9.81 \times 10^{-3} \pm 1.11 \times 10^{-4}$	$2.95 \times 10^{-2} \pm 2.96 \times 10^{-4}$	$3.11 \times 10^{-3} \pm 3.73 \times 10^{-5}$	$1.31 \times 10^{-2} \pm 1.18 \times 10^{-4}$	3.16	2.25
8	69.20	$7.59 \times 10^{-3} \pm 8.99 \times 10^{-5}$	$2.33 \times 10^{-2} \pm 2.37 \times 10^{-4}$	$2.20 \times 10^{-3} \pm 2.74 \times 10^{-5}$	$1.03 \times 10^{-2} \pm 8.96 \times 10^{-5}$	3.45	2.26
9	78.00	$5.76 \times 10^{-3} \pm 6.91 \times 10^{-5}$	$1.81 \times 10^{-2} \pm 1.79 \times 10^{-4}$	$1.55 \times 10^{-3} \pm 2.01 \times 10^{-5}$	$8.48 \times 10^{-3} \pm 7.13 \times 10^{-5}$	3.72	2.13
10	86.80	$4.53 \times 10^{-3} \pm 5.40 \times 10^{-5}$	$1.46 \times 10^{-2} \pm 1.44 \times 10^{-4}$	$1.10 \times 10^{-3} \pm 1.48 \times 10^{-5}$	$6.96 \times 10^{-3} \pm 5.44 \times 10^{-5}$	4.13	2.09
11	96.50	$3.67 \times 10^{-3} \pm 4.43 \times 10^{-5}$	$1.61 \times 10^{-2} \pm 1.48 \times 10^{-4}$	$7.60 \times 10^{-4} \pm 1.05 \times 10^{-5}$	$8.14 \times 10^{-3} \pm 5.31 \times 10^{-5}$	4.83	1.97
12	108.00	$2.86 \times 10^{-3} \pm 3.52 \times 10^{-5}$	$1.28 \times 10^{-2} \pm 1.09 \times 10^{-4}$	$4.95 \times 10^{-4} \pm 7.28 \times 10^{-6}$	$6.85 \times 10^{-3} \pm 3.69 \times 10^{-5}$	5.78	1.87

Table 7.1: Comparison of the densities of simulated tracker hits (“simhits”) and digitized tracker hits (“digis”) predicted by the “full” and “fast” simulation. The upper section of the table indicates the densities found with the default settings, whilst the lower section of the table indicates the densities found when the two simulations are adjusted to be suitable for direct comparison.

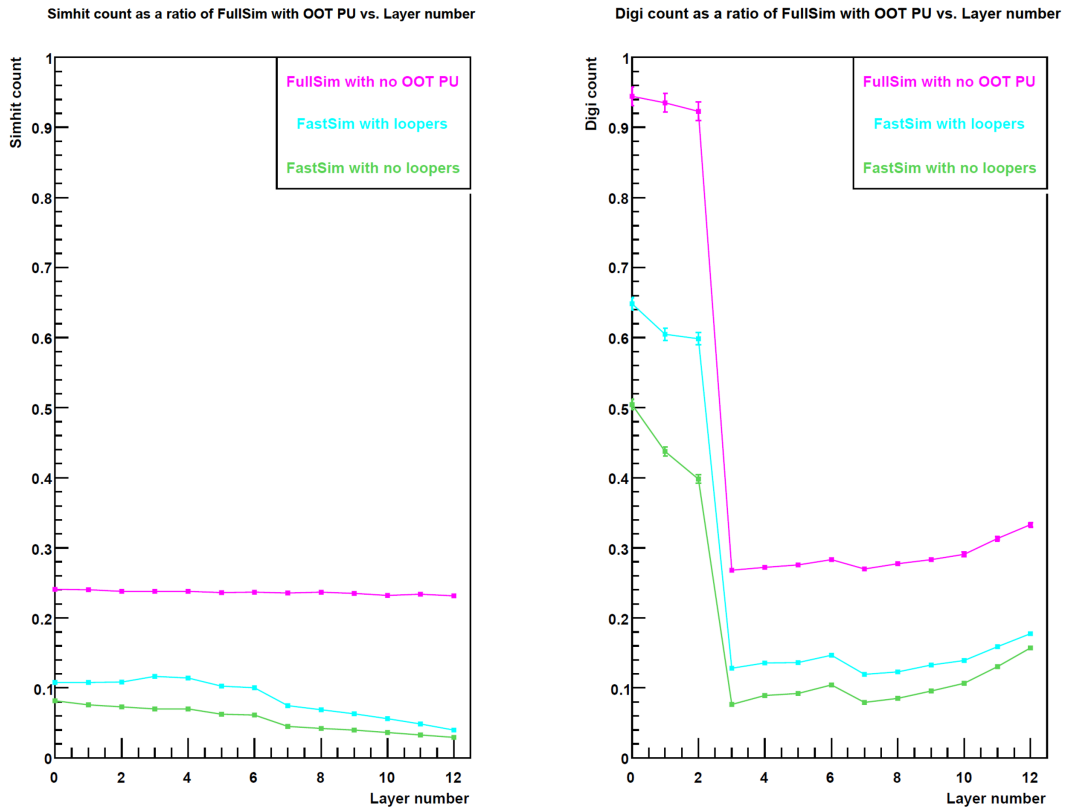


Figure 7.2: Relative densities of simulated tracker hits (“simhits”) and digitized tracker hits (“digis”) predicted by the “full” and “fast” simulation, normalized to “full” simulation with Out-Of-Time Pile-Up.

the “fast” and “full” simulations compared directly. These results were converted to densities (to remove the radial dependence) and are summarized in the upper half of table 7.1. The ratio of the mean number of hits found in the “full” simulation as compared to the mean number of hits found in the “fast” simulation is also shown in table 7.1. The most striking feature of these results is the discrepancy between the number of “simhits” produced in “fast” and “full” simulation; the latter having between ten and thirty times more hits at all radii. There are many differences in the simulation mechanisms of the “fast” and “full” simulations and so an inconsistency in the number of simulated hits is expected. The difference in the number of simulated hits between the two simulations is even acceptable, as long as the number of digitized hits are in agreement, since both methods are trying to simulate the detector — which produces digitized and not “truth” data — and it is thus the digitized data which should be used for analyses. In the digitized data, however, we also observe a disparity of up to an order of magnitude.

Four particular differences between the “fast” and “full” simulations were identified as likely to lead to significant discrepancies:

- The default setting for “fast” simulation is to cull any tracks with $p_T < 0.3$ GeV
- The default setting for “fast” simulation is to cull any tracks once it has performed more than half a loop in the magnetic field
- “Fast” simulation has no mechanism to account for out-of-time pile-up
- “Fast” simulation cannot model delta rays

While it is possible in the “fast” simulation both to lower the p_T threshold at which tracks are culled and to disable the culling of looping tracks, the lack of out-of-time pile-up and delta rays are inherent to the way that “fast” simulation is designed and so instead, for the purpose of comparison, the “full” simulation was modified to exclude out-of-time pile-up. As delta rays cannot be disabled

in “full” simulation, they continued to represent a difference between the two simulation methods. The analysis was then re-run to include these modifications and the results are shown in the lower half of table 7.1: while the ratio of the number of “simhits” in the full and “fast” simulations has improved, such that they are within an order of magnitude, the ratio of digitized hits has improved such that the two simulation methods differ only two-fold.

For the purpose of exploratory studies into tracking triggers, a two-fold difference between the two methods of modelling was considered acceptable.

7.3 A framework for the study of Tracking Triggers

The concept of the stacked tracker includes many features for which there is no equivalent in the current detector and, as such, much development of the framework was required before simulation studies could be performed. The problem of creating a suitable framework can be split broadly into:

- The lack of a suitable detector element to simulate a stack
- The hard-coding of simulation parameters to the values required for the current tracker
- The lack of a mechanism to associate detector elements
- The lack of data formats to describe the new trigger objects
- The lack of software infrastructure to construct the trigger objects

The initial feasibility study for a stacked tracker necessitated the creation of a simulation geometry which includes an object whose behaviour (in terms of both simulation and digitization) is that of a stack. The main practical difficulty in performing this task was that the software was written only for the implementation of the current geometry and, as such, it was never intended that it should

be modified: consequently, minimal documentation existed, many of the parameters were ill-defined, and numerous “shortcuts” and optimizations (which were relevant only to the current geometry) had been made. As the feasibility study was to be a proof of principle, rather than a study of detailed effects, it was decided that (to minimize the number of changes required to the geometry) it would be sufficient to modify the current pixel tracker geometry such that every ladder would be duplicated and the clone translated in the direction perpendicular to the original sensor. The cooling structures were not duplicated, but arranged so as to be shared between each ladder and its clone. The modified geometry is shown in figure 7.3.

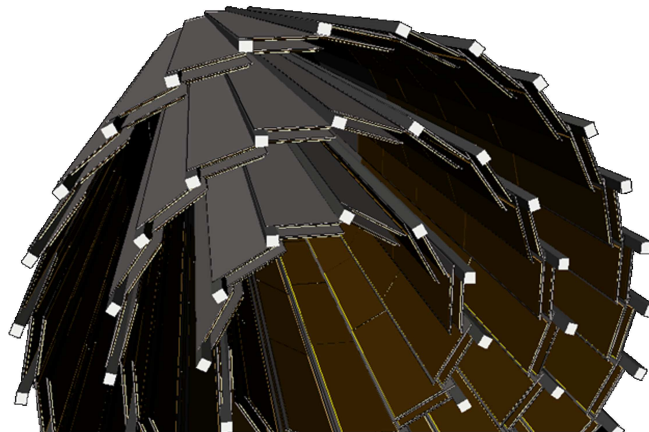


Figure 7.3: An event display (produced by the CMS visualization software, **IGUANA** (Interactive Graphics for User ANALysis) [117]), of the proof of principle geometry used to demonstrate the feasibility of the stacked tracking concept. The gold boxes represent silicon detector elements, the black boxes represent carbon fibre support structures and the silver boxes are cooling pipes. The layers are at radii 4.5 cm, 7.5 cm and 10.5 cm. Not indicated here is the beampipe which is at radius 4 cm, and is omitted for clarity.

The geometry was further modified to remove the **TIB** (Tracker Inner Barrel) and **TOB** (Tracker Outer Barrel), and to allow the stacked pixel tracker layers to be scaleable to any radius in the tracker volume. In order that the geometry be as open to change as possible, many of the parameters which had previously been “hard-coded” into the original XML configuration files were replaced by relational statements, so that the entire geometry was encoded using a minimal set of parameters.

Initial tests with the modified geometry returned inconsistent results, with large geometric sections of the detector recording simulated hits but not digitized hits. It was later shown that the effect only occurred when the number of ladders in a layer exceeded 256 and, subsequently, was shown to be caused by the “hard-coding” of field widths in the construction of the detector element identifiers (called “DetIds”). A similar “hard-coding” of field widths was determined to be the underlying cause of problems associated with changing the number of pixels per sensor.

To demonstrate that such a system could be used as described in section 6.3.2 (without the introduction of any further assumptions) it was shown that, when using a single electron signal from 1 GeV to 10 GeV (the range in which a momentum cut would be expected to lie), it was possible to reconstruct the track crossing-angle from the digitized hits and, from this, to obtain an estimate of the track momentum.

Because this basic model was based around the ladders of the current pixel geometry, its suitability for exploring the parameter space of possible future tracker upgrade was severely restricted and it became clear that a more flexible model was required. The prototype geometry, described here, provided the basis for the development of such a model (the so-called “Strawman-B” [118]), which was sufficiently flexible to both explore large regions of parameter space and to be later refined into two, more realistic, upgrade scenarios [119, 120].

To minimize the maintenance requirements of the new geometries, the existing software machinery was reused as much as possible, with modifications only being made when the existing code contained “hard-coded” parameters or placed limitations relevant only to the current geometry. This approach of minimal modification was taken with both the simulation and digitization modules and, as such, the sensors in a stack were treated independently as two separate sensing elements: whilst this was sufficient for studies of tracking performance (where each point in the tracker is considered in the global reference frame), for the purpose of trigger studies it was necessary to associate pairs of sensors

into a stack object, so that the hits in each sensor may be considered within the reference frame of the stack itself. Such an approach relies on the assumption that hits will only be compared within a stack and not between sensors in different stacks; given the close spacing of sensors, as discussed in section 6.3.1, and the need for simplicity (in order that the system should be implementable on-detector), this assumption is considered reasonable.

As there is no analogue for the paired sensors in the current detector, both a new detector-element type and a new type of detector-element identifier were required. For ease of adoption, the classes were designed to be as similar to the current detector-element and identifier classes as possible, through the use of shared base classes and the re-use of method names. In an exact analogy with the standard tracker, the stack elements are stored in a “geometry” object, allowing iteration over — or random access to — the stacks themselves.

Due to inconsistent numbering of the sensors and the regular nature of the layout, the association of sensors into stacks was initially performed geometrically, with any set of sensors falling within a configurable (r, ϕ, z) window being grouped. The later, more complicated, stack layouts subsequently made purely geometric matching insufficient and, coupled with corrections to the sensor numbering, matching sensors into stacks, using the sensor identifier numbers was implemented; this resulted in a simpler and faster algorithm. Figure 7.4 shows the software infrastructure implemented so that the stacks may be treated as fundamental detector elements.

In the same way that the software framework lacked a way of describing the detector element, the existing software provided neither an object suitable for describing the trigger candidates, nor a mechanism for the construction of such trigger candidates. As with the detector units, a new class was created to describe these trigger candidates and several considerations governed the design of this new data container: first, as the hit objects already exist within the event it is not desirable to duplicate data by storing a copy of the hits within the trigger candidate. The data was, instead, saved by means of a persistent pointer to the

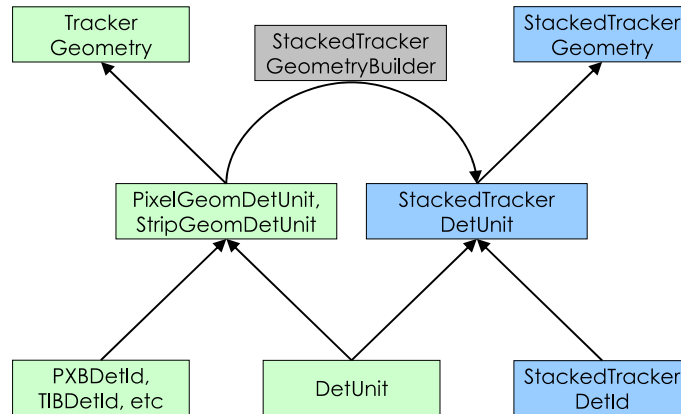


Figure 7.4: The StackedTracker utility classes for interfacing the tracker geometry with trigger simulations. The green boxes show existing CMSSW code to describe the tracker. The blue boxes show the StackedTracker classes designed to be as similar as possible to the equivalent tracker classes. The StackedTrackerDetUnits are automatically “constructed” from the tracker’s GeomDetUnits by the StackedTracker geometry builder (grey box).

original hit, such that all information in the original hit was accessible through the trigger candidate. Secondly, for the purpose of testing, it was considered desirable to be able to construct the trigger objects from both the simulated and digitized hits, in order to separate algorithmic effects from detector effects. It was also considered desirable that trigger objects should be capable of being constructed from an independent hit format, the “TrackTriggerHit”, to allow their use with stand-alone software: to accommodate the different types of hits, the container class was templated and difference in the underlying types masked through template specialization. Thirdly, as it was unknown if clustering of hits was required, it was necessary to provide for such a scenario.

In the CMSSW framework, plug-in modules which produce data are declared to be of type “Event Data Producer” and it is usual that each method of data production has its own Event Data Producer module. As the tracking trigger is under development and has, therefore, neither a well-defined method for clustering, nor one for matching hits between layers, such an approach would require an Event Data Producer module for each conceivable combination of algorithms — this resulting in large amounts of code duplication, thereby increasing the number of opportunities for the introduction of errors. Instead, a novel approach was taken: a single Event Data Producer module was defined independently of

an explicit hit-matching or clustering algorithm and was, instead, provided with “handles” to the two types of algorithm. Handles are framework objects which provide an interface between the software module and the event object: if the object type is unique, or given an identifying “tag”, the handle accesses that object in the event and behaves as a pointer to that object. Because handles behave like a pointer to an object, it is possible to design them with respect to a base class (with virtual methods), rather than with respect to a specific implementation of the object. By taking this approach, the specification of algorithms was moved from a compile-time to a run-time issue: each specific implementation of an algorithm could be written as a stand-alone plug-in module (inherited from the relevant base class) and, at run-time, the algorithm of choice inserted into the event object. The framework ensures that only a single algorithm of each type may be present in the event object and the Event Data Producer module is guaranteed to use this algorithm. If multiple algorithms of the same base type are presented to the event, a run-time error is raised and the simulation terminated.

Because the afore-described trigger candidate object acts as a container for hits, these candidates are defined without reference to the global coordinate system and are therefore called “local stubs”. A second, equivalent, data type was implemented to represent the “vector” object in the global geometry and called a “global stub”. A distinction was made between the two types of stub, even though the vector representation could have been stored in the container class, or a helper method provided to calculate it: this was because it was believed that the on-detector representation would be in the form of a pair of hits, whereas the off-detector (trigger) representation would be in terms of the global coordinate system; the separate data-types reflect this distinction.

A third data container was implemented to represent extended track-like objects called “tracklets”: these tracklets are produced from global stubs and may be desirable at the level-1 trigger. While the stubs contain no momentum information, tracklets provide a means to measure the track curvature and thus also the momentum. The relationships between the data types are shown in figure 7.5.

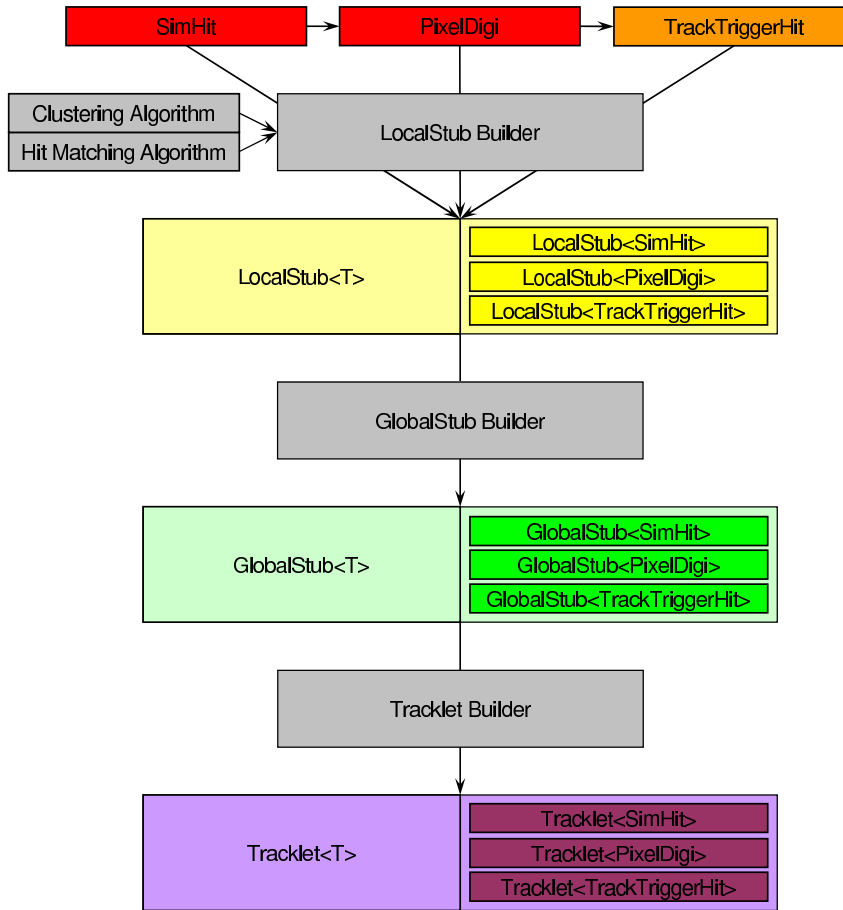


Figure 7.5: The SLHC L1Trigger classes: data-types and tools for describing a level-1 tracking trigger.

Although not strictly necessary, the templating introduced into the local stubs was carried across to the other trigger candidate types so that the source of the data was explicit through the object type itself. Maintaining the type-distinction was highly desirable, as, due to the differences in the intrinsic resolutions of the underlying hit types, confusing the sources of data could have a major impact on the choice of algorithms and parameters.

7.3.1 Clustering algorithms

The framework described is sufficiently flexible to allow for the implementation of clustering algorithms of any complexity, but implementing complex algorithms “on-chip” requires considerable logic resources. Digital logic is, however,

costly in terms of both the power and physical space required and, for this reason, it is desirable to minimize its usage. Four algorithms were written, based on schemes considered simple enough to be implemented on-detector.

The first scheme is the trivial case where every hit pixel is considered to be a cluster of 1 hit; this being equivalent to no clustering at all.

The second scheme is 1-dimensional clustering. It is believed that the most likely pixel topology for use in a tracking trigger layer will be $\mathcal{O}(100\ \mu\text{m}) \times \mathcal{O}(1\ \text{mm})$ in $r\phi \times z$, with the sensors being read out in columns of constant z . Due to bremsstrahlung and tracks looping in the $4T$ magnetic field, and due to the small width of the pixels, there is a high probability of charge-sharing in the $r\phi$ -direction and, consequently of a single track triggering multiple pixels. In a scheme where the pixels are being read out column-wise, it is imaginable that, as each pixel is clocked out, it is filtered through an edge-finding algorithm and all contiguous groups of triggered pixels considered as clusters. This method was referred to as “broadside clustering”.

The third scheme is a variation on the second: as the $r\phi$ -width of the energy deposited in a tracker layer is dependent on the curvature of the track, tracks with low transverse momentum create wide deposits and tracks with high transverse momentum create narrow deposits. As we are already trying to eliminate tracks with low transverse momentum, it is possible to place a restriction on the maximum acceptable width of a cluster; the default value chosen for this maximum was 3 pixels. This scheme may again be considered a simple addition to a column-wise readout.

The fourth scheme involves 2-dimensional clustering. In the low $|\eta|$ region of the tracker, the only contribution to creating long clusters in the z -direction is charge-sharing from tracks passing close to the boundary; the relatively long, $\mathcal{O}(1\ \text{mm})$, pixels reduces this possibility. In the forward regions of the detector, however, tracks may cross the detector at very shallow angles making the triggering of multiple pixels in the z -direction more likely and, as such, a 2-dimensional

clustering scheme was considered. For each pixel, three “kill”-bits are calculated through Boolean combinations of the neighbouring pixel states or of the neighbouring pixel’s “kill”-bits (illustrated in appendix C): if any of the “kill”-bits are set, then the pixel is not considered for readout. By this means, the cluster is limited to a maximum width of 2 pixels in the $r\phi$ -direction and 2 pixels in the z -direction. This scheme has a known limitation that, if a cluster is longer than two pixels in the z -direction, the cluster is split into two; the additional complexity required to handle longer clusters was considered as providing insufficient benefit to be worthwhile. This scheme is not compatible with the previously-described column-wise readout scheme and would require implementation either within the pixel itself, or in a readout chip with access to all the pixels.

To evaluate the four schemes, the rate of clusters was compared to the rates of the simulated and digitized hits (figure 7.6) and the momentum distributions of the accepted and rejected tracks plotted (figure 7.7).

The use of even the simplest clustering (1-D with no width cut) helps to control the most extreme cases, in which there are many tens of pixels triggered (figure 7.6). The introduction of a width cut offers a slightly greater improvement, however, imposing a cut improves the rejection of tracks with p_T less than 1 GeV by 2-3 orders of magnitude while maintaining good acceptance above 6 GeV (figure 7.7). The high- p_T tracks that are rejected by the 2-D clustering (and not by the 1-D clustering with cluster width cut) represent the difference between a width cut of 2 pixels and a width cut of 3 pixels.

7.3.2 Hit-matching algorithms

When matching hits into stubs, two distinct approaches may be taken: first, for each pair of hits in the inner and outer sensor of a stack, the separation in the z -direction and the angular separation in the x - y plane may be determined. For any specific value of transverse momentum there is, associated with each point

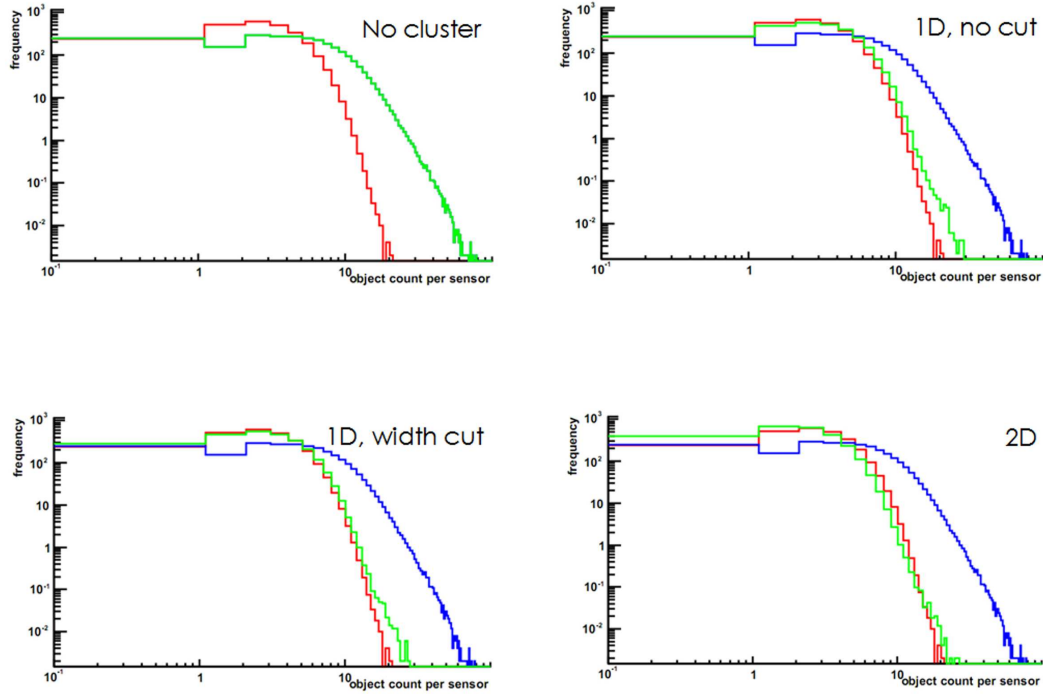


Figure 7.6: Mean frequency of occurrence of a particular number of simulated hits (red), digitized hits (blue) and clusters (green) per sensor for each clustering algorithm. The distributions of simulated hits and digitized hits are the same in all cases. The figure labelled “No cluster” represents the case where every pixel is considered to be a cluster of 1 hit and so corresponds precisely to the digitized hits.

in the tracker, an angle^{iv}; by requiring that the magnitude of the angle calculated in the x-y plane is less than this threshold angle, a p_T -cut is introduced. There is, similarly, associated with each point in the tracker, a positive and negative threshold in the z-direction related to the size of the luminous region^v; by requiring that the separation in the z-direction is within these thresholds, hit combinations which point outside the luminous region may be rejected. The implementation of this algorithm was called the “Global Geometry” algorithm as, in order to calculate the z-separation and angular separations in the global coordinate system, the absolute positions of the hits must be known. While this method works well when the position of the hits in the sensor is known exactly (for example, when using the simulated hits), the use of digitized hits causes problems. By reconstructing the hit at the centre of the pixel, tracks which ac-

^{iv}c.f. equation 6.4

^vc.f. equation 6.10

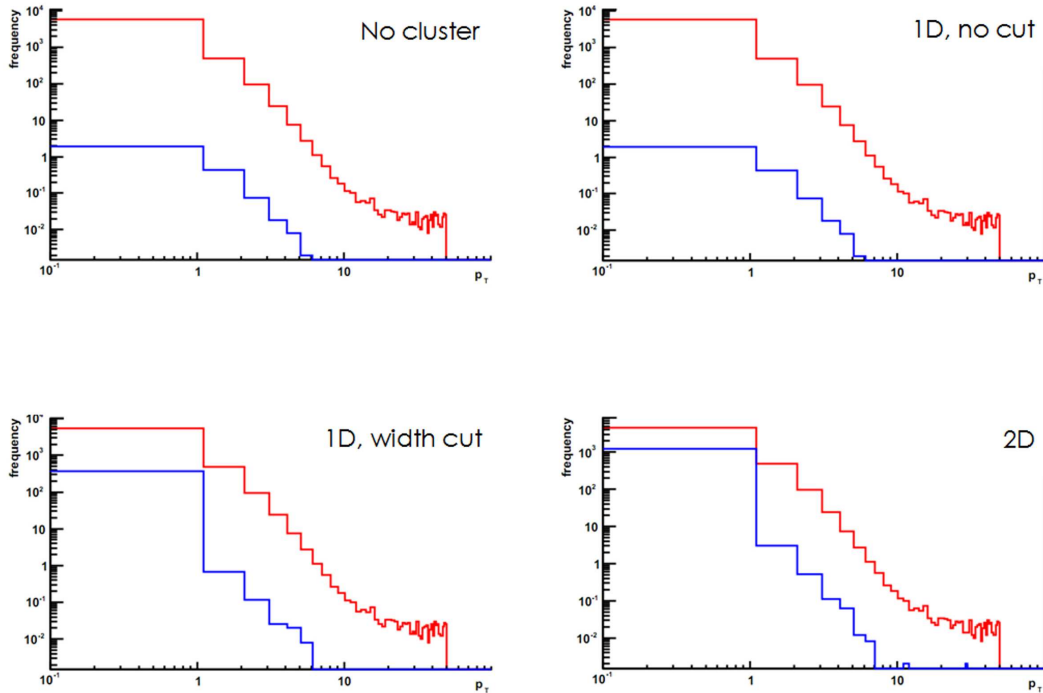


Figure 7.7: p_T distribution of tracks accepted (red) and rejected (blue) by each clustering algorithm. The rejected tracks in the top two plots are those which failed an ADC cut of 30/255 in the conversion from digital to binary pixels.

tually passed close to the edges of the pixels were lost, as the vertex was, in effect, being reconstructed outside the luminous region (figure 7.8) or the track reconstructed at the wrong p_T .

A variation on the Global Geometry algorithm was produced (called the “Pixel Ray” algorithm) which took into account the finite pixel size by calculating the minimum and maximum possible crossing-angles and z-separations, rather than assuming that the track passed through the centre of a pixel. The maximum and minimum were then compared to the threshold values and, if either value satisfied the crossing-angle condition and either value satisfied the z-separation condition, then the pair of hits was accepted as a stub.

Although the described algorithms accurately match the geometric arguments laid out in section 6.3.2, they are not an accurate description of how such a scheme would be implemented in hardware; the flaws being three-fold:

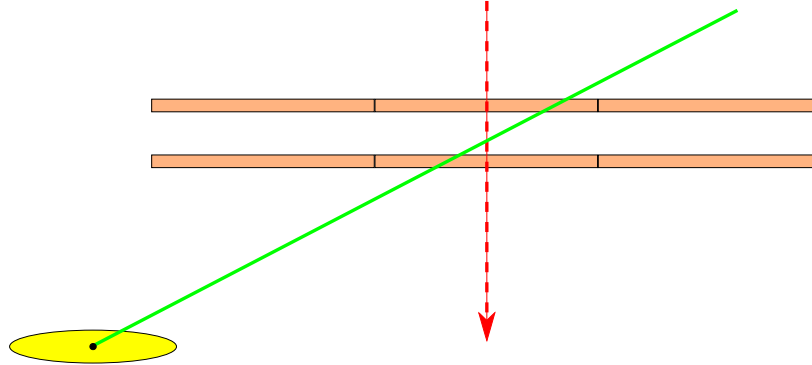


Figure 7.8: The effective misreconstruction of tracks with large crossing-angles. A genuine primary track (solid line) may be reconstructed (dashed line) as being invalid, if the location of the hit is assumed to be the centre of the pixel.

- In order to calculate $\Delta\phi$ and Δz requires either a coordinate transformation from the sensor frame of reference to the global frame, or a look-up table which stores the value for the various pairings of rows and columns.
- The threshold values are different for every point in the tracker and would need to be stored in a per-pixel look-up table.
- There is no means of limiting the number of pixel pairings before the comparison has taken place, resulting in a large number of combinatorials.

Such a scheme is too computationally intensive for implementation on-detector and another approach is required: for each pixel in the inner sensor of each stack, the aforementioned ϕ - and z -thresholds are calculated, projected onto the second stack member and associated with pixel row and column numbers. Each pixel in the inner sensor of a stack has, therefore, associated with it a group of pixels in the outer sensor within which any hit from a high- p_T primary track must lie — thereby restricting the number of combinations. Furthermore, as the window sizes are predetermined and defined to be an integer number of rows and columns, the comparison between sensors is trivial and may be performed in a look-up table or even constructed using “hard-wired” logic. This algorithm may be termed a “Local Geometry” algorithm, as it performs all matching in the local frame of reference. As all calculations are done in terms of rows and

columns, it is only possible to use this algorithm with digitized hits, and, therefore, there can be no direct comparison with simulated hits.

7.3.3 Tracklets

While the provision to match hits into stubs was provided by means of an algorithm plug-in module, no such provision was made for the tracklet class; the matching algorithm, instead, being “hard-coded”. It is envisaged that tracklets would be constructed off-detector as part of the trigger-chain, rather than as part of the tracker hardware, and that these objects would be constructed in the global reference frame^{vi}. Once in the global reference frame, the number of methods for matching hits is limited and all are conceptually similar; a scaled-up version of the global geometry algorithm is an adequate approximation to most of the cases: the separation of hits in the ϕ -direction provides a means of measuring the track’s transverse momentum and, by requiring that the separation of the hits in the z -direction is within a set of bounds, the track may be constrained to originate from the luminous region. In calculating the p_T using the separation in the ϕ -direction, the nominal beam-spot ($x=0, y=0$) is implicitly assumed as a third point on the track.

7.4 Initial Studies with Tracking Triggers

Initial studies were performed using “fast” simulation in CMSSW, version 1.8.4, using simulated particle guns, particle guns with background and pure background samples. The studies were initially performed using the Strawman-B geometry, but later used the “FermiLab Long-Barrel” geometry as this was considered a more realistic design. In order to introduce as little bias as possible,

^{vi}Even if the hits were to be matched in the frame of reference of one of the stacks, rather than a true global reference frame, a geometric transformation of the second stack into the frame of the first would still be required. The computational complexity of the two scenarios is similar, although more bits are required to describe a true global reference frame.

hit-matching was performed geometrically in the global reference frame (as described in section 7.3.2) and no clustering was used. The two geometries, as well as the Hybrid geometry, may be compared in appendix D.

In moving from the Strawman-B geometry to the Long-Barrel geometry, it was noted that, although the distribution of the number of hits per layer remained broadly similar, the number of stubs was markedly different, changing from a smoothly falling distribution to a distribution with distinct crenellations: every odd-numbered layer having more stubs than the even-numbered layers either preceding or following it (figure 7.9).

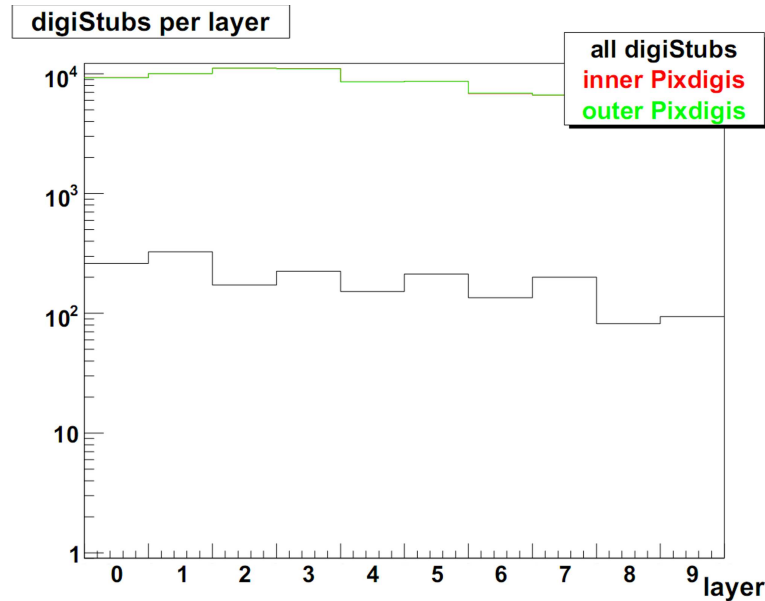


Figure 7.9: The distinct crenellations in the stub density which appeared when the geometry was changed to the Long-Barrel tracker geometry. No comparable pattern was observed when using Strawman-B tracker geometry. The high-low pattern is seen in neither the inner or outer sensors of the stack (red and green lines, overlapping).

Because these crenellations are not seen in the distribution of hits from which the stubs are formed, this necessarily implies that the origin of the effect is in the probability of stub formation: there is a higher probability of the hits in odd-numbered layers being associated into stubs, as compared to hits in even-numbered layers. The probability of forming stubs depends not on the total number of hits in each stack member, but rather on the local density of hits; this implies that the density of hits should be higher in the odd-numbered layers

than in the even-numbered layers: by calculating the average separation of every pair of hits within each silicon sensor^{vii}, this hypothesis was demonstrated to be correct.

One of the key differences between the Strawman-B and the Long-Barrel geometries is the radial spacing of the stacked layers: Strawman-B has layers separated by 10 cm and 20 cm alternately, while the Long-Barrel uses alternate 4 cm and 12 cm separations. This difference in spacing was suggestive that the number of stubs in a layer is inversely proportional to the distance to the preceding layer and this, in itself, suggests the cause is a material effect^{viii}. If a track were to interact with a tracking layer in the detector, the products of that interaction would define a conical volume as they propagate. If the following layer was close to the point of interaction, the cross-sectional area of the cone would be small and, therefore, the tracks and hits in the tracking layer densely packed; this leads to a high probability of hits from different tracks being incorrectly matched into stubs. If the following layer was far from the point of interaction, the cross-sectional area of the cone would be large, the tracks and hits in the tracking layer less densely packed and the probability of stub formation low.

To test the hypothesis that layer spacing was affecting the probability of stub formation, the distribution of stubs in each layer was measured as the Long-Barrel geometry was systematically modified so as to increase the radius of the odd-numbered layers, whilst leaving the even-numbered layers unchanged. As the separation of the layers was increased, the scale of the crenellations diminished until, at a separation of around 7 cm, the effect was no longer observed and the smoothly falling distribution recovered. As such, the discrepancy with the Strawman-B geometry was understood — the 10 cm spacing was not susceptible to the material effect — and the first insights into the importance of material effects were made.

^{vii}sensor, not stack

^{viii}A material effect is any effect which is introduced when the assumption that the detector is simply a passive, non-interacting observer is relaxed (for example, coulomb scattering or nuclear interactions).

7.4.1 Multiplicities

The hypothesis that material interactions were causing an increase in the number of stubs associated with a single track, raised the issue of the number of stubs and tracklets that might be expected. Clearly, a scenario where the number of stubs exceeds the number of hits is unacceptable, as the very purpose of producing stubs is to reduce the tracker data-rate. Based on the discussion in section 7.3.1, it is clear that the choice of clustering algorithm should have a significant impact on the number of stubs.

Using 50,000 events consisting of an “electron gun” ($1 \leq p_T \leq 20 \text{ GeV}$, $|\eta| \leq 1.5$) with 0 pile-up, for each event, the number of stubs per layer and tracklets per pair of layers were estimated. The performance of local-geometry and global-geometry hit-matching algorithms were compared using the four different clustering algorithms and, for comparison, stubs were also constructed from simulated hits using the global geometry hit-matching algorithm^{ix}. By creating stubs from simulated hits, the detrimental effects of pixelization and charge-sharing are eliminated and, as such, these represent the “cleanest” possible signal. All algorithms were set to use a nominal p_T -threshold of 1 GeV. The shape of the distributions (which are broadly similar in all cases) may be seen in figure 7.10 and the results are summarized in table 7.2.

As expected, the stub multiplicity falls as the complexity of the clustering algorithm rises: whilst the 2-D clustering results in multiplicities which closely match the results obtained using simulated hits, the unclustered case results in considerably more events where multiple stubs are produced by a single track.

It is expected that, if the cuts with which stubs are matched into tracklets are insufficiently tight, then each stub located close to the track in one layer will be matched to *ALL* stubs close to the track in the next layer; as such, the number of tracklets should be proportional to the product of the number of stubs in each layer. As the mean number of stubs per layer is greater than one, it is expected

^{ix}as clustering is not available for simulated hits

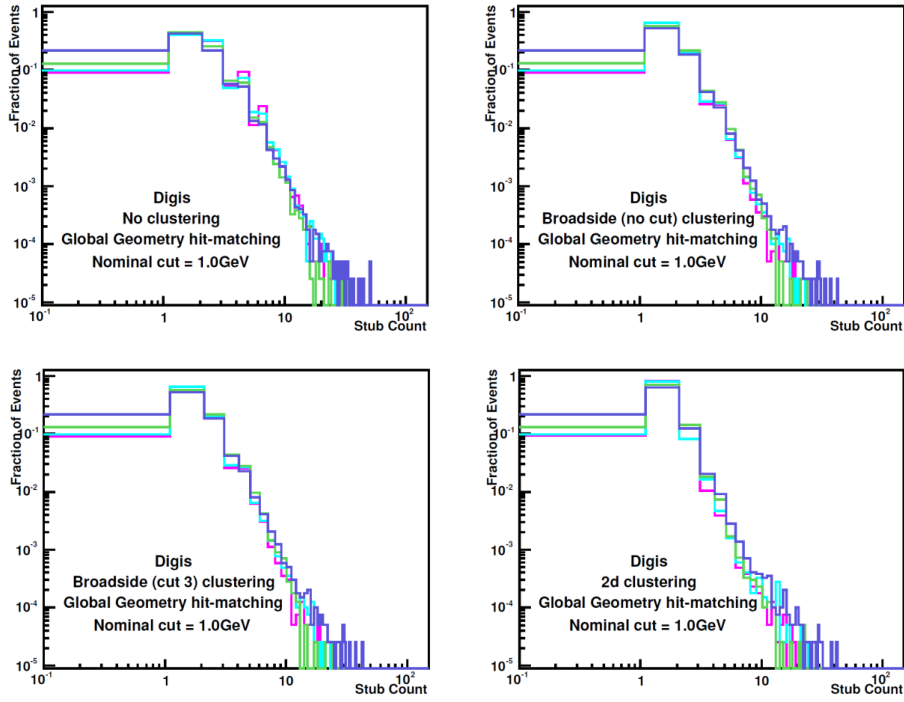


Figure 7.10: Simulated estimates of stub multiplicities in the Strawman-B geometry, for four different clustering algorithms. The four colours correspond to four stacked tracker layers at radii 25 cm (pink), 35 cm (turquoise), 55 cm (green) and 65 cm (purple).

Hit Matching Algorithm		Global Geometry				Local Geometry				Simhits
Clustering Algorithm		No	Broadside (no cut)	Broadside (cut 3)	2d	No	Broadside (no cut)	Broadside (cut 3)	2d	
Layer 0	$f(N_{stub} = 0)$	9.0%	9.1%	9.1%	9.1%	9.0%	9.2%	9.2%	9.2%	9.3%
	$f(N_{stub} = 1)$	39.8%	64.4%	64.4%	81.1%	40.2%	65.0%	65.0%	81.9%	81.8%
	$f(N_{stub} > 1)$	51.2%	26.5%	26.5%	9.8%	50.8%	25.8%	25.8%	8.9%	8.9%
	$\overline{N}_{stub}$	2.4	1.8	1.8	1.5	2.4	1.8	1.8	1.5	1.5
Layer 1	$f(N_{stub} = 0)$	9.7%	9.8%	9.8%	9.8%	9.7%	9.9%	9.9%	9.9%	10.0%
	$f(N_{stub} = 1)$	40.2%	63.8%	63.8%	79.5%	40.7%	64.5%	64.5%	80.5%	80.5%
	$f(N_{stub} > 1)$	50.0%	26.4%	26.4%	10.7%	49.6%	25.6%	25.6%	9.7%	9.5%
	$\overline{N}_{stub}$	2.4	1.8	1.8	1.6	2.3	1.8	1.8	1.5	1.5
Layer 2	$f(N_{stub} = 0)$	12.9%	13.0%	13.0%	13.0%	12.6%	12.8%	12.8%	12.8%	13.1%
	$f(N_{stub} = 1)$	45.0%	56.7%	56.7%	70.1%	45.2%	57.0%	57.0%	70.6%	70.8%
	$f(N_{stub} > 1)$	42.1%	30.4%	30.4%	16.9%	42.2%	30.2%	30.2%	16.6%	16.0%
	$\overline{N}_{stub}$	2.2	1.8	1.8	1.6	2.2	1.8	1.8	1.6	1.6
Layer 3	$f(N_{stub} = 0)$	21.6%	21.7%	21.7%	21.7%	21.3%	21.5%	21.5%	21.5%	21.8%
	$f(N_{stub} = 1)$	41.9%	51.7%	51.7%	62.3%	42.2%	51.9%	51.9%	62.6%	63.0%
	$f(N_{stub} > 1)$	36.5%	26.6%	26.6%	16.0%	36.5%	26.6%	26.6%	15.8%	15.2%
	$\overline{N}_{stub}$	2.0	1.7	1.7	1.5	2.0	1.7	1.7	1.5	1.5

Table 7.2: Simulated estimates of stub multiplicities by layer, clustering algorithm and hit-matching algorithm. The fraction of events, f , with “0”, “1” or “greater than 1” stubs per layer is quoted and the mean number of stubs, $\overline{N}_{stub}$, is also given.

that the multiplicity of tracklets will be greater than that of stubs. It is, similarly, expected that the fraction of events with no tracklets should be higher than the fraction of events with no stubs. The shapes of the distributions (which are again all broadly similar) may be seen in figure 7.11 and the results are summarized in table 7.3.

It should be noted that the spikes which may be seen in figure 7.11 are not statistical fluctuations, but correspond to composite (non-prime) numbers of tracklets. Such composite numbers of tracklets is to be expected as, in the simplest case, to produce a prime number of tracklets, N , requires a single stub in one layer and N stubs in another. If — as suggested previously (section 7.4) — the density of stubs is determined by closely-packed tracks resulting from material interactions, then it is more likely that the number of stubs in each layer will be similar and, thus, the number of tracklets more likely to be composite.

Whilst the cases with a large number of stubs and tracklets per signal track represent one extreme, the cases with zero candidate objects represent the other extreme and raise questions of the efficiency with which valid tracks are found by stubs or tracklets.

7.4.2 Efficiencies

The efficiency with which tracks are reconstructed could reasonable be expected to be a function of the p_T , the pseudorapidity and the radius at which the reconstruction is taking place. In the ideal scenario where the positions of hits within the tracker are known exactly, where tracker coverage is 100% and where the tracker material-budget is negligible, the efficiencies might be expected to have an infinitely-sharp turn-on at the selected threshold and to be flat in both η and radius. An imperfect position resolution will translate to an imperfect momentum resolution, this causing a loss of gradient of the turn-on curve, whilst imperfect coverage, material effects and bremsstrahlung will also limit the maximum efficiency.

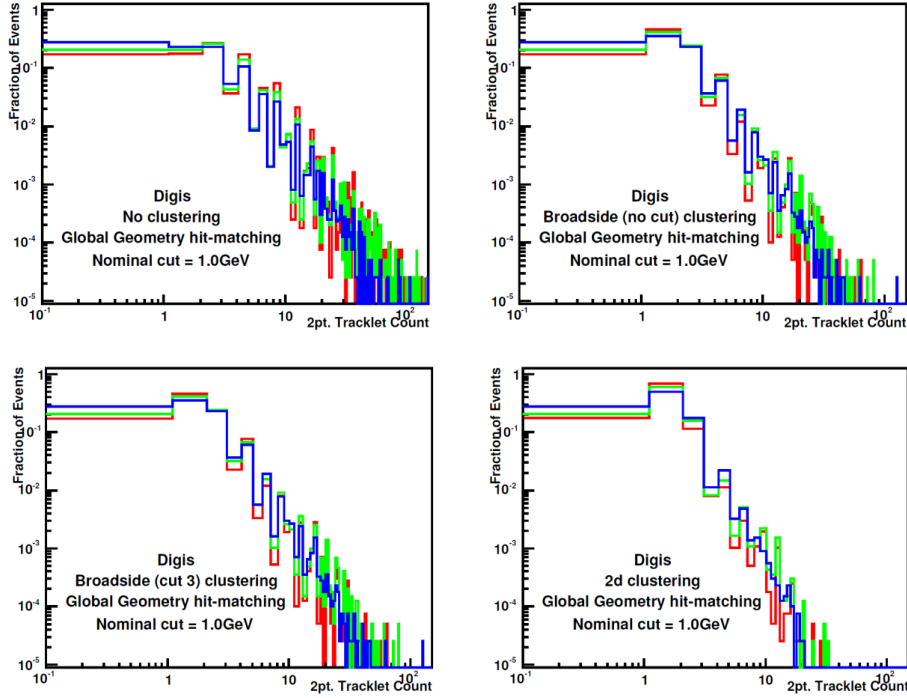


Figure 7.11: Simulated estimates of tracklet multiplicities in the Strawman-B geometry, for four different clustering algorithms. The three colours correspond to three pairs of stacked tracker layers at radii 25-35 cm(red), 35-55 cm(green) and 55-65 cm(blue).

Hit Matching Algorithm		Global Geometry				Local Geometry				Simhits
Clustering Algorithm		No	Broadside (no cut)	Broadside (cut 3)	2d	No	Broadside (no cut)	Broadside (cut 3)	2d	
Layer 0 to 1	$f(N_{tracklet} = 0)$	17.2%	17.3%	17.3%	17.3%	17.1%	17.3%	17.3%	17.3%	17.2%
	$f(N_{tracklet} = 1)$	17.7%	45.2%	45.2%	68.5%	17.7%	45.2%	45.2%	68.4%	68.2%
	$f(N_{tracklet} > 1)$	65.1%	37.5%	37.5%	14.2%	65.2%	37.5%	37.5%	14.3%	14.5%
	$\bar{N}_{tracklet}$	4.0	2.2	2.2	1.6	4.0	2.2	2.2	1.6	1.6
Layer 1 to 2	$f(N_{tracklet} = 0)$	20.6%	20.6%	20.6%	20.6%	20.5%	20.6%	20.6%	20.6%	20.6%
	$f(N_{tracklet} = 1)$	20.4%	40.6%	40.6%	60.0%	20.4%	40.5%	40.5%	59.8%	59.8%
	$f(N_{tracklet} > 1)$	59.0%	38.8%	38.8%	19.4%	59.2%	38.8%	38.8%	19.6%	19.7%
	$\bar{N}_{tracklet}$	3.5	2.3	2.3	1.6	3.6	2.3	2.3	1.6	1.6
Layer 2 to 3	$f(N_{tracklet} = 0)$	27.5%	27.6%	27.6%	27.6%	27.4%	27.5%	27.5%	27.5%	27.5%
	$f(N_{tracklet} = 1)$	22.8%	35.1%	35.1%	50.0%	22.7%	35.1%	35.1%	49.9%	49.9%
	$f(N_{tracklet} > 1)$	49.7%	37.3%	37.3%	22.4%	49.9%	37.5%	37.5%	22.6%	22.6%
	$\bar{N}_{tracklet}$	2.9	2.1	2.1	1.6	2.9	2.1	2.1	1.6	1.6

Table 7.3: Simulated estimates of tracklet multiplicities by layer, clustering algorithm and hit-matching algorithm. The fraction of events, f , with '0', '1' or 'greater than 1' tracklets per pair of layers is quoted and the mean number of tracklets, $\bar{N}_{tracklet}$, also given.

If pile-up events are again excluded, the stub-finding efficiency may be defined as the probability that at least one stub is found in a layer and the tracklet-finding efficiency may be defined as the probability that at least one tracklet is found per pair of consecutive layers.

To demonstrate the p_T -dependence of the algorithms, events containing at least one stub (or, separately, tracklet) were classified according to the p_T of the Monte Carlo-truth track and the distribution compared with the p_T distribution of all events, based on the same conditions as for the multiplicity studies. The performance of the algorithms was compared using nominal p_T -thresholds of 1 GeV and 5 GeV and a selection of results are shown in figure 7.12.

Using a nominal threshold of 1 GeV, the efficiency turn-on curves are as expected (having a sharp turn-on and high overall efficiency) and match well with the efficiency curves of stubs constructed from “simhits”: the same is not true, however, when using a nominal threshold of 5 GeV. The cause of this anomaly is quantization errors introduced by digitization; by quantizing the positions on each of the sensors, a high- p_T track producing “simhits” which pass the hit-matching criteria, may trigger pixels whose centres correspond to a vector that does not pass the hit-matching criteria. As the nominal p_T -cut is raised, the acceptance cone narrows and the probability that the centre of the triggered pixel lies within this cone falls; this results in a lower efficiency, as demonstrated in figure 7.13. Conversely, it is also possible that, whilst a track may produce “simhits” which lie outside the acceptance cone, the digitized hits may correspond to a vector which is acceptable and this results in a low- p_T tail of objects that are incorrectly accepted.

Hit-matching based in the local geometry (that is, based on row and column windows in the digitized sensors) should not suffer the inefficiencies described above, as the cone size cannot drop below the size of a pixel. Such a system does not, however, allow tight control of the threshold and the penalty paid is a slower turn-on curve (figure 7.14). Clearly, narrowing the pixels in the r - ϕ direction would reduce the amount of quantization error and allow for tighter control of

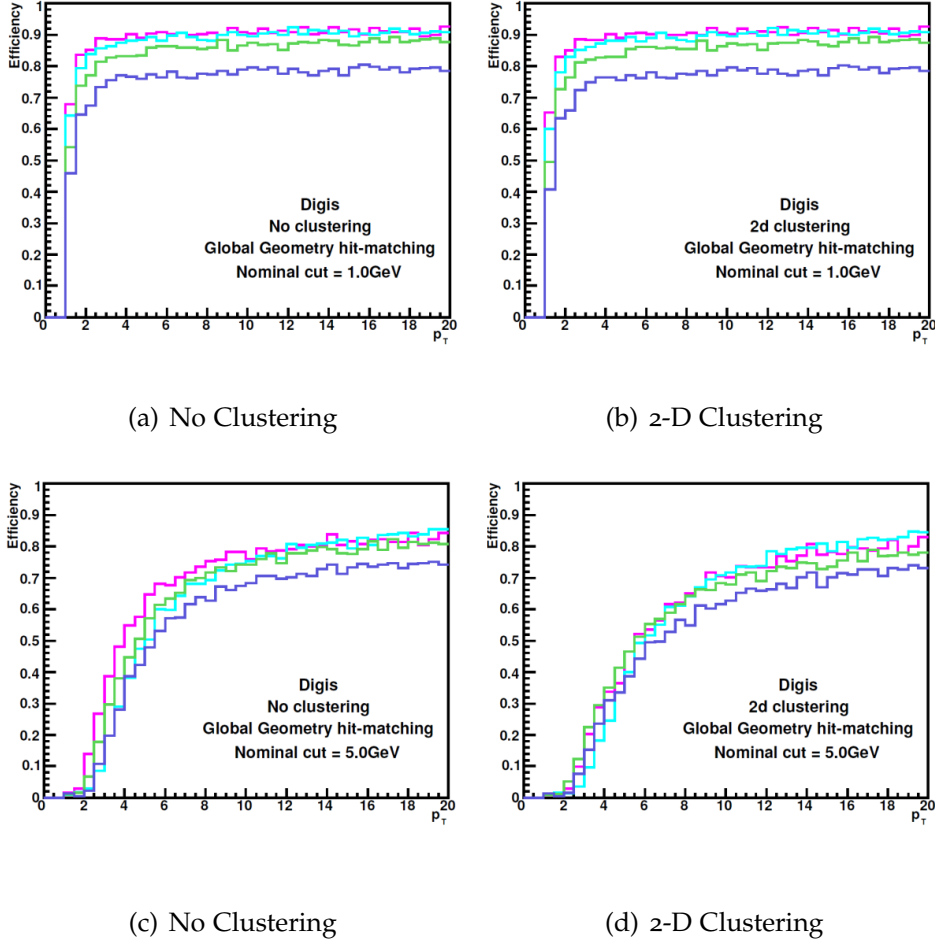


Figure 7.12: Simulated estimates of single particle stub-finding efficiencies as a function of the p_T of the Monte Carlo-truth track for a nominal p_T -cut of 1.0 GeV (figures 7.12(a), 7.12(b)) and 5.0 GeV (figures 7.12(c), 7.12(d)). The four colours correspond to four stacked tracker layers at radii 25 cm (pink), 35 cm (turquoise), 55 cm (green) and 65 cm (purple).

the p_T -threshold; the penalty for doing so would, however, be an increased number of channels, higher power-consumption, increased risk of triggering multiple pixels and increased cost.

A further concern is the $\mathcal{O}(15\%)$ lower efficiency in the layer at 65 cm compared to that at 25 cm: the difference comes in the forward region, where the pseudorapidity coverage of the outer layers does not match that of the inner layers (figure 7.15).

The reason that the η -coverage of the test is defined to be larger than the η -coverage of the detector, is that the test was designed as part of a suite of tools

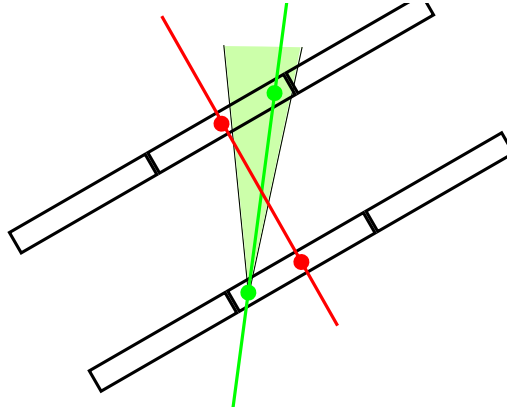


Figure 7.13: Failure of digitized hits to pass a p_T -cut that is passed by the equivalent simulated hits. A high- p_T track (green line) produces simulated hits (green circle) which lie within a cone (light green) corresponding to a geometric p_T -cut. Due to quantization of the hit position, however, the centres of the two hit pixels (red circles) may correspond to a vector (red line) which lie outside the cone. Raising the p_T -cut corresponds to narrowing the geometric cone.

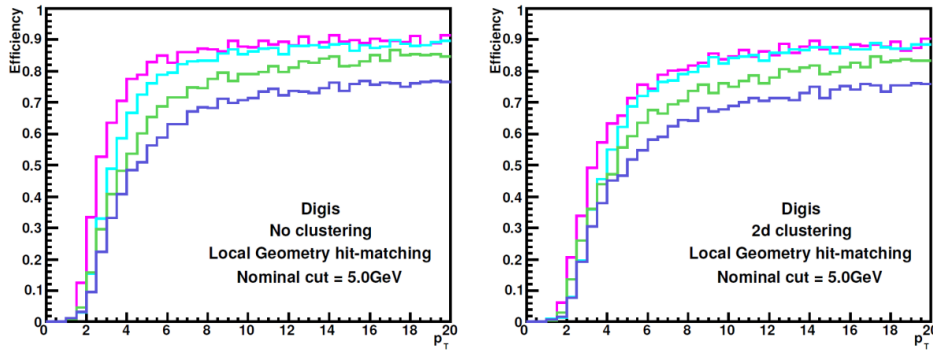


Figure 7.14: Simulated estimates of single particle stub-finding efficiencies of the local geometry hit-matching algorithm as a function of the p_T of the Monte Carlo-truth track, for a nominal p_T -cut of 5.0 GeV. The four colours correspond to four stacked tracker layers at radii 25 cm (pink), 35 cm (turquoise), 55 cm (green) and 65 cm (purple).

to allow direct comparison between different upgrade geometries: although the Strawman-B geometry does not cover the forward region, other upgrade geometries do have extended pseudorapidity coverage.

The estimated efficiency of stub production may be defined by $\epsilon_{stub} = \rho_{stub} \epsilon_{hit}^2$, where ϵ_{hit} is the efficiency with which a track leaves a hit in a single layer of silicon and ρ_{stub} is the probability of two hits being matched into a stub; both are functions of pseudorapidity, radius and p_T . The efficiency of tracklet production may then be, similarly, defined as $\epsilon_{tracklet} = \rho_{tracklet} \epsilon_{stub}^2 = \rho_{tracklet} \rho_{stub}^2 \epsilon_{hit}^4$, where $\rho_{tracklet}$ is the probability of two stubs being matched into a tracklet and is,

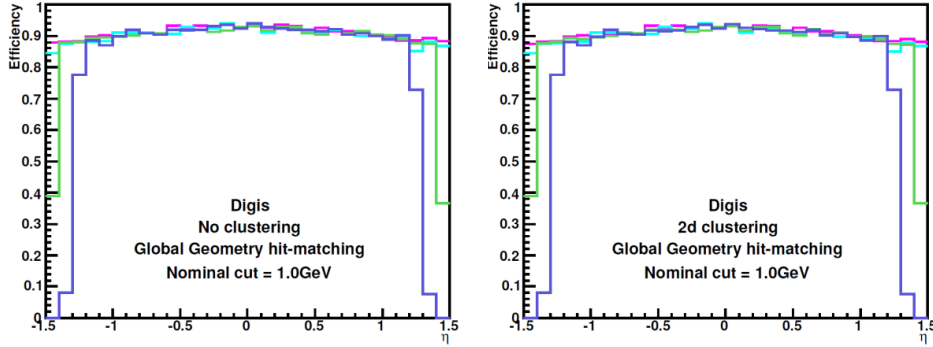


Figure 7.15: Simulated estimates of single particle stub-finding efficiencies as a function of the pseudorapidity of the Monte Carlo-truth track. The four colours correspond to four stacked tracker layers at radii 25 cm (pink), 35 cm (turquoise), 55 cm (green) and 65 cm (purple).

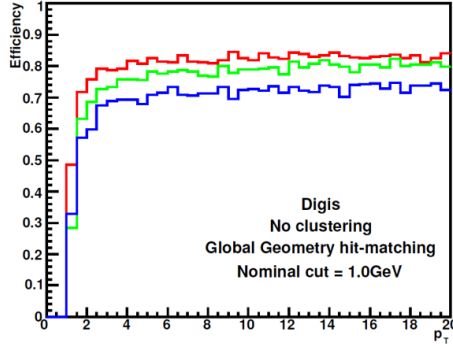
again, a function of pseudorapidity, radius and p_T . As the tracklet efficiency is proportional to the square of the stub efficiency, the inefficiencies described for stubs may be expected to be magnified for tracklets, as observed (figure 7.16).

7.4.3 Rates

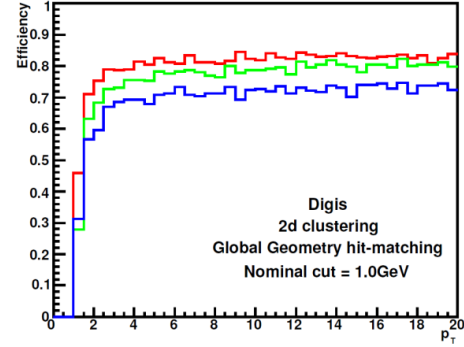
As stated in section 7.4.1, a scenario where the number of stubs exceeds the number of hits is unacceptable, as the very purpose of producing stubs is to reduce the tracker data-rate. While the efficiency curves show that the stacked tracking scheme eliminates tracks with low transverse momentum (as required), the effect of multiplicities would counteract any reduction in data-rate.

To determine the achievable reduction in rates, for each layer, all hits and stubs were classified by the pseudorapidity of their location, for 2000 events consisting of an “electron gun” ($1 \leq p_T \leq 20 \text{ GeV}$, $|\eta| \leq 1.5$) with 200 pile-up. The distribution of tracklets was, similarly, determined for each pair of consecutive layers. For this study, the local geometry hit-matching algorithm (with a nominal p_T -threshold of 5 GeV) and the 2-D clustering algorithm were chosen, as the local geometry hit-matching algorithm is both realistic and maintains the highest efficiency, and the 2-D clustering algorithm results in the lowest multiplicities.

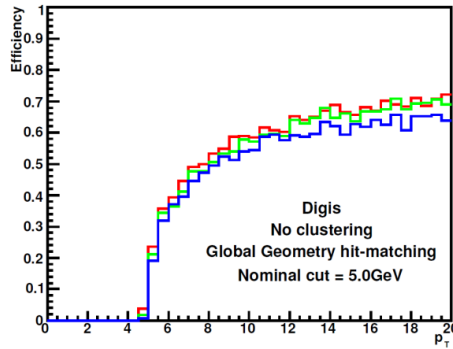
In the central region, the number of stubs is two orders of magnitude lower than the number of hits, whilst in the forward region the difference is three orders of



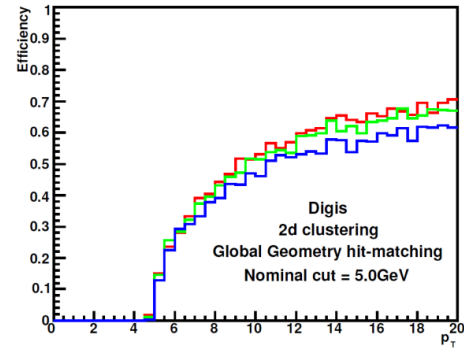
(a) Nominal p_T -cut of 1.0 GeV,
Global Geometry Algorithm



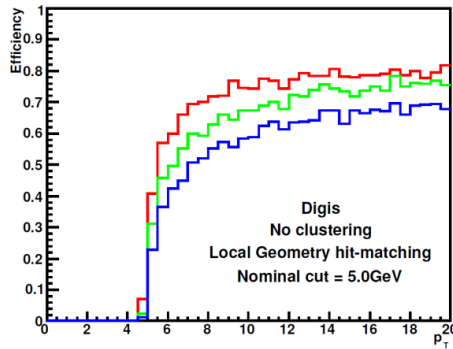
(b) Nominal p_T -cut of 1.0 GeV,
Global Geometry Algorithm



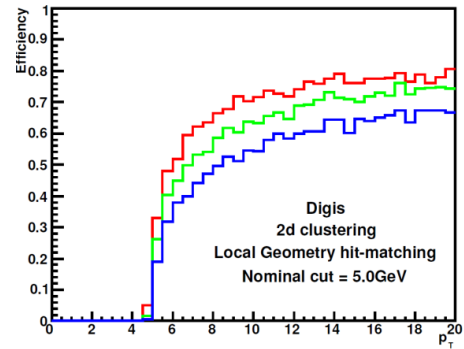
(c) Nominal p_T -cut of 5.0 GeV,
Global Geometry Algorithm



(d) Nominal p_T -cut of 5.0 GeV,
Global Geometry Algorithm



(e) Nominal p_T -cut of 5.0 GeV,
Local Geometry Algorithm



(f) Nominal p_T -cut of 5.0 GeV,
Local Geometry Algorithm

Figure 7.16: Simulated estimates of single particle tracklet-finding efficiencies as a function of the p_T of the Monte Carlo-truth track for a nominal p_T -cut of 1.0 GeV and 5.0 GeV. The three colours correspond to three pairs of stacked tracker layers at radii 25-35 cm (red), 35-55 cm (green) and 55-65 cm (blue).

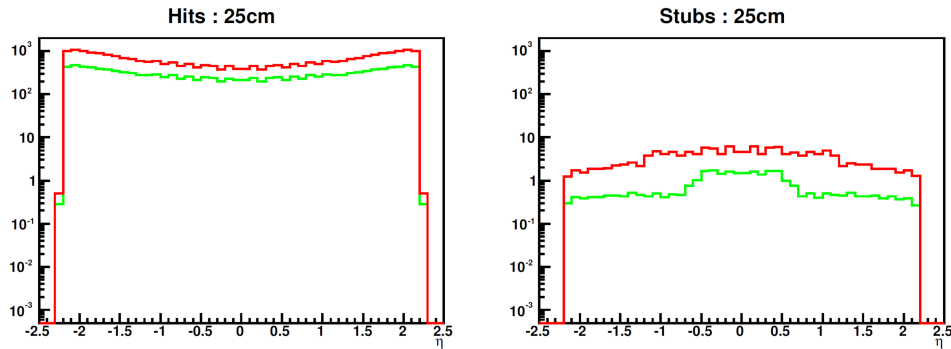


Figure 7.17: Simulated estimates of tracker hit (left) and stub (right) rates as a function of the pseudorapidity. The green line represents the number of simulated hits and stubs constructed from simulated hits, the red line represents digitized hits and stubs constructed from digitized hits.

magnitude (figure 7.17). This is to be expected: although the number of particles in the forward region is higher, their momentum in the transverse-plane would be expected to be lower, due to their forward-boosted nature.

For a single layer at a radius of 25 cm, converting the rates of hits and stubs into true data rates — assuming a 16-bit addressing scheme and 8b/10b encoding — results in rates of ~ 28 Tbit/s and ~ 100 Gbit/s respectively. Further assuming the use of the aforementioned 3.36 Gbit/s serializers would require ~ 8200 links to transmit all hits off detector, whilst ~ 30 links would be required to transmit the stubs; in terms of power and space required, the latter is conceivable, the former is not. Changing the nominal threshold will, of course, change the rate of accepted candidates and hence the data rate; the data rate is, therefore, a key consideration in the setting of the threshold.

While a rate-reduction by two orders of magnitude is significant, it is not as large as had been — possibly naïvely — hoped for: the distribution of track momenta (figure 6.7, for example), indicates that the number of tracks with $p_T < 5$ GeV is approximately 4 orders of magnitude higher than the number of tracks with $p_T \geq 5$ GeV. Although figure 7.18 indicates that the formation of tracklets achieves a less than ten-fold further reduction in the “object” count as compared to the number of stubs, it should be remembered that — because tracklets are formed between pairs of tracker layers — there is one less “layer” of tracklets to be considered as compared to stubs.

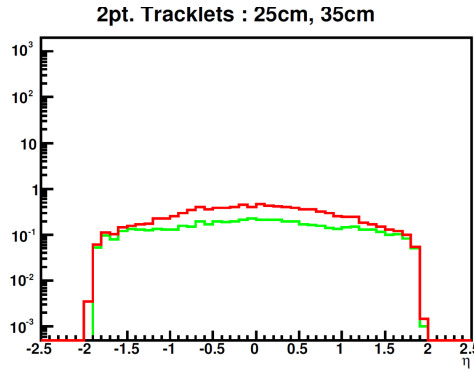


Figure 7.18: Simulated estimates of tracklet rates as a function of the pseudorapidity. The green line represents the number of tracklets constructed from simulated hits, the red line represents tracklets constructed from digitized hits.

7.4.4 Purities

To further understand the rates of the stubs and tracklets, the objects were classified based on their Monte Carlo-truth. The classifications were:

- **“Good”** - all hits in the stub or tracklet originate from a single track, whose Monte Carlo $p_T \geq 5$ GeV (the nominal p_T -threshold).
- **“Low- p_T ”** - all hits in the stub or tracklet originate from a single track, whose Monte Carlo $p_T < 5$ GeV (the nominal p_T -threshold).
- **“Bad”** - the hits in the stub or tracklet originate from at least two different tracks.
- **“Parallel”** - at least two different tracks contribute to every hit in the stub or tracklet.

The “good” or “low- p_T ” tracks, were then further classified as either lepton, photon or “other” (in effect, hadrons).

Under the same simulation conditions as the rate studies, the proportion of objects under each classification were histogrammed according to their pseudorapidity (figure 7.19) and, as expected, the digitized tracker hits are dominated

by “low- p_T ” hadrons (65-75%) and “low- p_T ” leptons (25-35%) at all pseudorapidities. The formation of stubs has some effect in reducing the rate of “low- p_T ” leptons and boosting the presence of both “good” leptons and “good” hadrons, but the dominance of “low- p_T ” hadrons is still clear. Furthermore, the formation of stubs introduces an additional background of “bad” hits that also dominates the “good” signals.

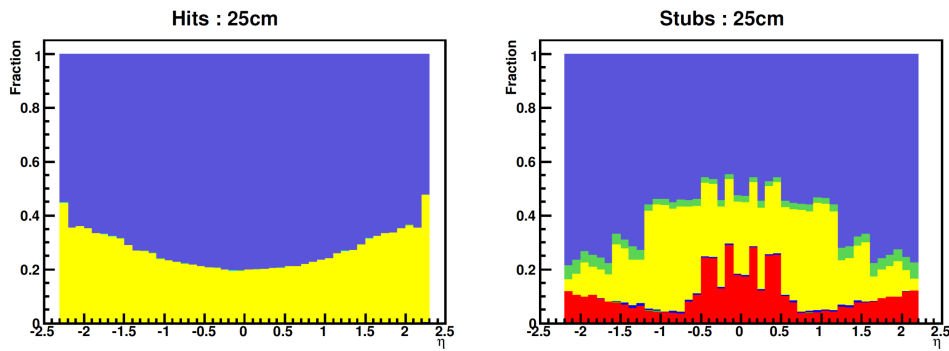


Figure 7.19: Simulated estimates of purities of digitized hits (left) and stubs (right) constructed from digitized hits as a function of the pseudorapidity. From the bottom of the stack upwards, the object types are “bad” (red), “parallel” (lime green), “good” lepton (blue), “low- p_T ” lepton (yellow), “good” photon (pink), “low- p_T ” photon (turquoise), “good” other (dark green) and “low- p_T ” other (purple).

Based on the efficiency turn-on curves^x, it is expected that tracklets should provide superior rejection of low- p_T tracks. As may be seen in figure 7.20, this is indeed the case: because of the large inter-stack spacing (10 cm), however, large matching windows are required to associate stubs into tracklets and, as such, the background of mismatched hits predominates.

It is not possible to draw conclusions about “bad” objects from the efficiency turn-on curves, as these are based on the response to single particles: the number of “low- p_T ” objects which pass the cut is, however, greater than expected. As this can not be explained by the single particle response and as these objects are — by their classification — genuine tracks, it was surmized that the low- p_T tracks being reconstructed were, in fact, secondary tracks. Both the formation of stubs from hits and the formation of tracklets from stubs use, either implicitly

^xfor example, comparing figures 7.12(d), 7.14 and 7.16

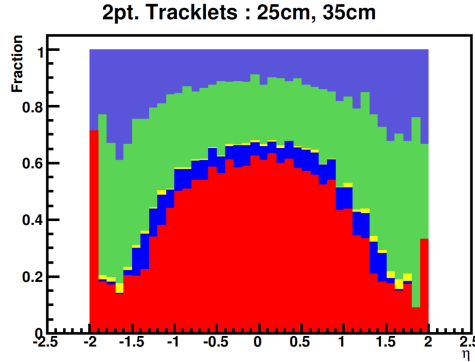


Figure 7.20: Simulated estimates of purities of tracklets constructed from digitized hits as a function of the pseudorapidity. From the bottom of the stack upwards, the object types are “bad” (red), “parallel” (lime green), “good” lepton (blue), “low- p_T ” lepton (yellow), “good” photon (pink), “low- p_T ” photon (turquoise), “good” other (dark green) and “low- p_T ” other (purple).

or explicitly, the beamspot as a constraint in the reconstruction of the object. For secondary tracks, such a constraint is fictitious and can cause tracks with low transverse momentum to be reconstructed at much higher p_T (figure 7.21).

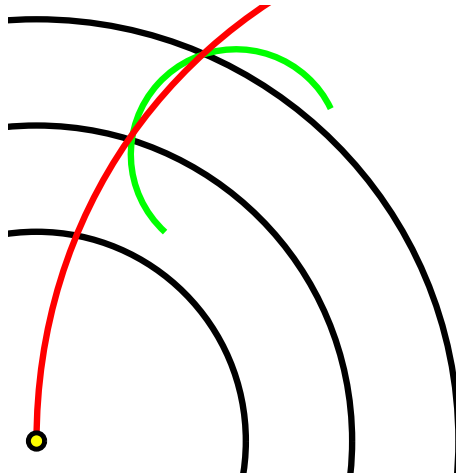


Figure 7.21: Reconstruction of secondary tracks as primary tracks. A genuine low- p_T secondary track (green) may be reconstructed incorrectly as a high- p_T primary track (red) due to the incorrect assumption that the beam constraint applies.

7.4.5 Incorrect reconstruction of secondary tracks

To test the hypothesis that secondary tracks were responsible for impurities, for each stub (and, similarly, for each tracklet) that was not classified as “bad”, the

Monte Carlo track that created the object was found and the radial position of the track's origin plotted against the Monte Carlo p_T (figure 7.22).

As expected, those tracks which pass the momentum cut despite having a Monte Carlo p_T below the nominal threshold, originate predominantly away from the interaction region. Furthermore, three definite substructures exist within the plots: first, tracks with p_T below approximately 0.8 GeV spiral in the 4T magnetic field and cannot reach the calorimeter; these looping tracks having a disproportionate presence due to their multiple passes through the tracker. Secondly, above $p_T \approx 0.8$ GeV, tracks enter the ECAL and are lost from the tracker; if such tracks are generated at a radius smaller than a particular tracker layer, then they only pass through that layer once; if generated at a radius larger than a particular tracker layer, they cannot cross that layer at all. Thirdly, a distinct abundance of tracks originate at radii 4, 7, 11, 15, 25, 35 and 55 cm, corresponding to the radii of the pixel and stacked tracker layers; this demonstrates that material interactions are a significant source of fake signatures — a result that was not unexpected, considering the results shown earlier in this section^{xi}.

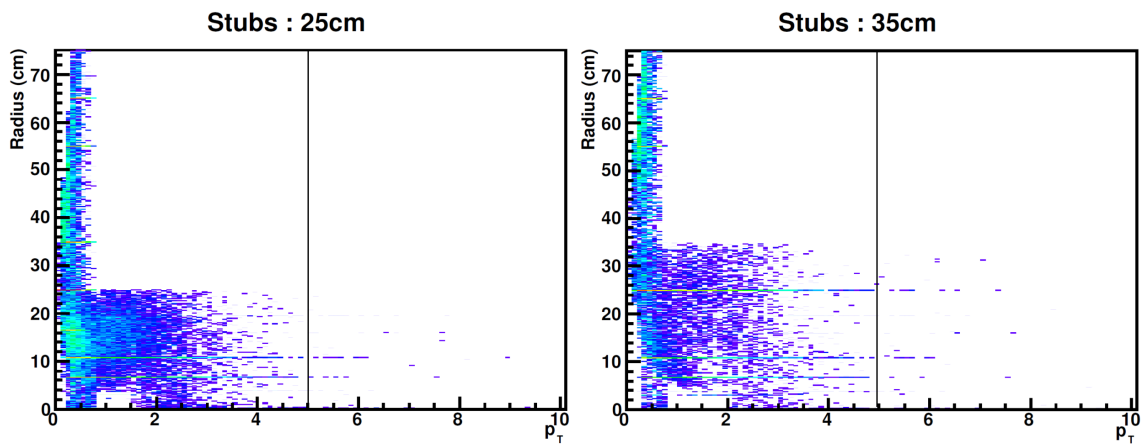


Figure 7.22: Simulated distribution of accepted tracks as a function of the Monte Carlo p_T and by the radial position of the track's origin, for tracker layers at 25 cm (left) and 35 cm (right).

What was not expected is that the same results apply to tracklets; it having been thought that the better p_T measurement available with tracklets would be able to reject these low- p_T tracks.

^{xi}for example, section 7.4

7.4.6 Removing the vertex constraint

As one source of misreconstructed tracks was the assumption of a primary vertex, it was natural to examine whether performance could be improved by removal of this constraint: whilst this cannot be done for stubs (as we are constrained to the use of two sensors in a stack), it can be done for tracklets. As shown in figure 7.21, although a low- p_T track may fake a high- p_T track in two layers, it cannot do so in three layers and may, therefore, be vetoed. Rather than modifying the existing tracklet software, a “third pass” was introduced, whereby the tracklets built from two stubs, henceforth two-point tracklets, were daisy-chained into “three-point” tracklets. Any track which can create a three-point tracklet capable of passing a p_T -cut can necessarily create a two-point tracklet capable of passing a p_T -cut and, as such, this approach does not introduce a bias.

While the efficiency of stubs falls as $\epsilon_{stub} = \rho_{stub} \epsilon_{hit}^2$ and the efficiency of two-point tracklets falls as $\epsilon_{2pt} = \rho_{2pt} \rho_{stub}^2 \epsilon_{hit}^4$, the efficiency of three-point tracklets falls as $\epsilon_{3pt} = \rho_{3pt} \rho_{2pt}^2 \rho_{stub}^3 \epsilon_{hit}^6$ (where ρ_{2pt} is the probability of two stubs being matched into a tracklet, ρ_{3pt} is the probability of two two-point tracklets being matched into a three-point tracklet and all other terms are defined in section 7.4.2) and, as such, is expected to be low. Furthermore, the efficiency is limited by the restricted pseudorapidity coverage of the outer layers. Both of these effects may be seen in 7.23.

The number of three-point tracklet objects is illustrated in figure 7.24. While the number of tracklets constructed from simulated hits is approximately the same as that of two-point tracklets, the number of tracklets constructed from digitized hits is further reduced five-fold. The total reduction in data-rate is, however, greater than this as there are fewer layers to be read out.

The real gain in constructing three-point tracklets is the improvement in purity (figure 7.25). Two two-point tracklets can be matched into a three-point tracklet without any assumption of the nominal vertex position (which is only loosely

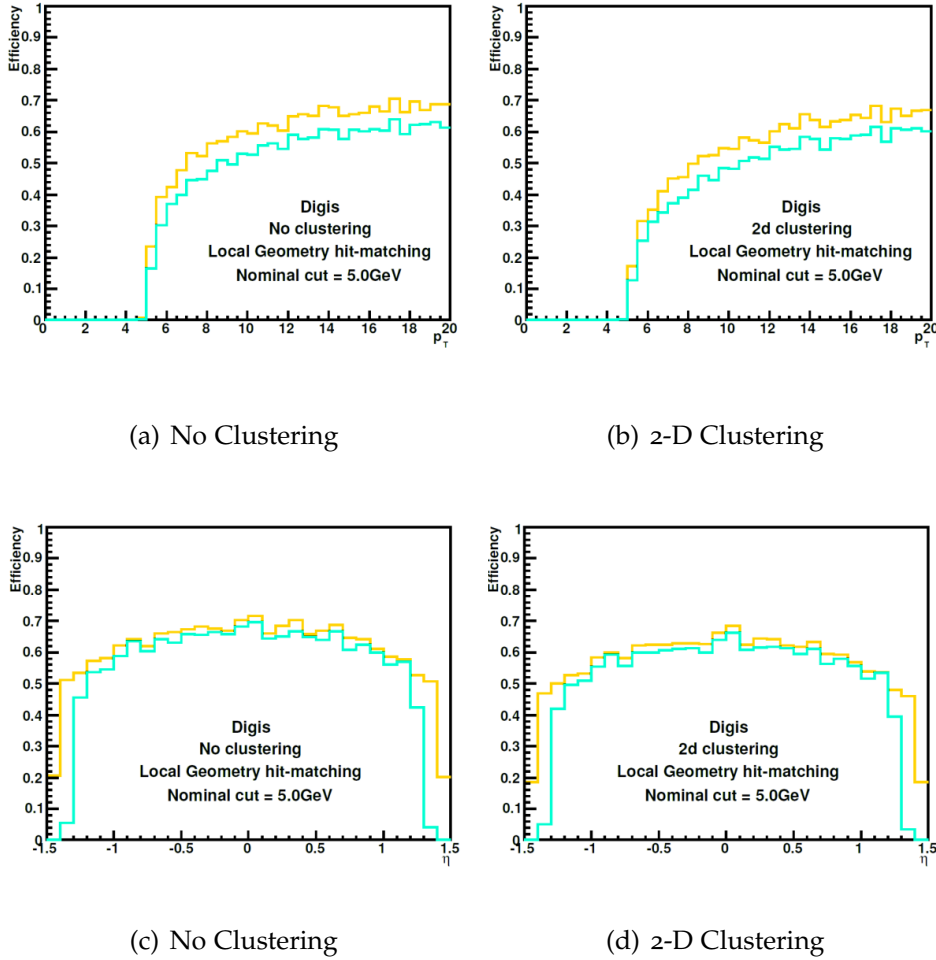


Figure 7.23: Simulated estimates of single particle three-point tracklet-finding efficiencies as a function of the p_T (figures 7.23(a), 7.23(b)) and of the pseudorapidity (figures 7.23(c), 7.23(d)) of the MonteCarlo-truth track for a nominal p_T -cut of 5.0GeV. The two colours correspond to tracklets formed from stubs at radii 25, 35, 55 cm (orange) and 35, 55, 65 cm (aquamarine).

constrained) and, as such, the number of mismatched tracks is minimized. The three-point tracklets found are dominated by “good” leptons (20%) and “good” hadrons (70%), while the number of “low- p_T ” tracks is reduced to the $\mathcal{O}(1\%)$.

7.4.7 p_T -resolutions

An estimate of the transverse momentum may be made geometrically if three points in the tracker, or two points in the tracker together with the beam-spot

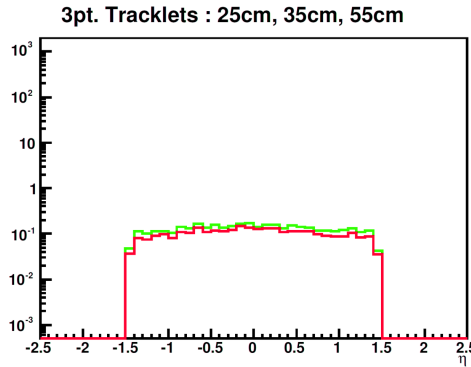


Figure 7.24: Simulated estimates of three-point rates as a function of the pseudorapidity. The green line represents the number of three-point tracklets constructed from simulated hits, the red line represents three-point tracklets constructed from digitized hits.

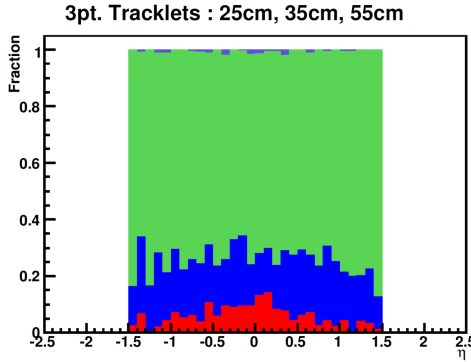


Figure 7.25: Simulated estimates of purities of three-point tracklets constructed from digitized hits as a function of the pseudorapidity. From the bottom of the stack upwards, the object types are “bad” (red), “parallel” (lime green), “good” lepton (blue), “low- p_T ” lepton (yellow), “good” photon (pink), “low- p_T ” photon (turquoise), “good” other (dark green) and “low- p_T ” other (purple).

are known^{xii}. If more than three points are known, however, the circle-fitting problem is over-constrained. To compare the momentum resolutions of tracker objects consisting of more than three points (for example, three tracker points plus beam-spot), two circle-fitting methods were created, based on two different “least-squares” approaches for which closed-form solutions exist [112].

The first approach is termed the “average of intersection” fit or “perpendicular bisector” fit. By noting that the perpendicular bisector of any chord on a circle must necessarily pass through the centre of the circle, two chords may be used to define the centre. If chords are constructed from data points (with an asso-

^{xii}for reference, see figure 6.12 and equation 6.22

ciated error) then every chord has an associated uncertainty and, therefore, so does every estimate of the centre: a best estimate of the centre may be achieved by taking an averaging over the estimated centres. An estimate of the radius of curvature (and thus of the transverse momentum) may be found by averaging the distance between the estimated centre of curvature and the data points. Such an approach is mathematically similar (although computationally very different) to the Hough transform [121] used at the MINOS [122] and ALICE [32] experiments, although when two points are very close together, the current method suffers from mathematical instabilities that are absent from the latter method.

The second approach is called the “modified least-squares” fit. It uses a traditional “least-squares” approach to minimize the distance from the estimated centre of the circle to the known perpendicular bisectors, but weighted so as to favour bisectors calculated from widely-spaced points. The radius of curvature (and thus the transverse momentum) are, again, estimated by averaging the distance between each of the data points and the estimated centre of curvature.

The computational complexity of the two least-squares approaches may be compared with the geometric approach and Hough transform methods of calculating the radius of curvature (table 7.4).

Fit Method	Computational Complexity (Elemental Operations)					
	For n points	2 tracker points + vertex	3 tracker points	3 tracker points + vertex	4 tracker points	4 tracker points + vertex
Geometric	-	14	32	-	-	-
Perpendicular Bisector	$6n^3 - 17n^2 + 18n$	63	63	184	184	415
modified least-squares	$23n + 46$	115	115	138	138	161
M-bit Hough Transform	$2^{3M} + 7n 2^{2M}$	$> 2^{3M} = 16,777,216$ for $M = 8$ -bit resolution				

Table 7.4: Computational complexity of four methods of measuring r_{curve} . It is assumed here that the data has been pre-formatted into the most convenient coordinate systems; for geometric fitting of 2 tracker points + vertex this is a polar coordinate system, for all others it is a Cartesian system. Elemental operations are defined to be: add, subtract, multiply, divide, square root, sine and cosine. The list of elemental operations and the complexity of the Perpendicular Bisector and modified least-squares methods are defined in [112]. The resolution of the Hough Transform refers to the number of address bits required to label the bins along each axis in accumulator space: 8 bits $\equiv$ 256 bins.

The p_T -resolution is defined by equation 7.1, where the superscripts *reco.* and

M.C. denote “reconstructed” and “Monte Carlo-truth”, respectively.

$$\sigma_{p_T} = \frac{p_T^{reco.} - p_T^{M.C.}}{p_T^{M.C.}} \quad (7.1)$$

The p_T -resolution was calculated, using the above methods, for three scenarios: two tracker points + the nominal vertex; three tracker points; and three tracker points + the nominal vertex (figure 7.26).

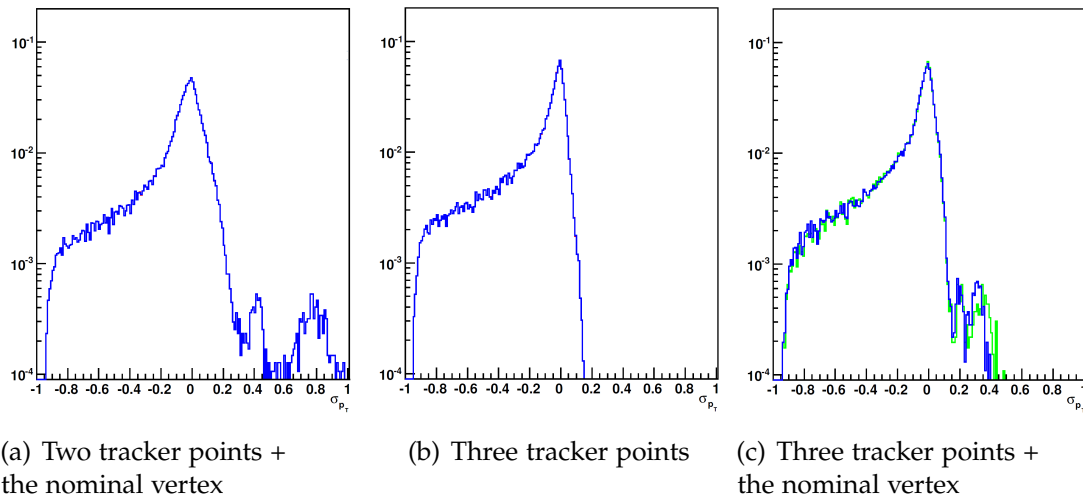


Figure 7.26: Simulated p_T -resolution calculated using two tracker points + the nominal vertex, three tracker points and three tracker points + the nominal vertex. Three different fitting techniques were used: the geometric method (red, plots 7.26(a) and 7.26(b) only) is in excellent agreement with — and almost completely overlapped by — the modified least-squares method (blue), while the perpendicular bisector method (green) is similar.

All of the distributions exhibit three notable features:

- A peak at $\sigma_{p_T} = 0$
- A long shoulder at $\sigma_{p_T} < 0$
- Spikes at $\sigma_{p_T} > 0$

The central peak is as expected: demonstrating that both the process of track-finding and the process of p_T measurement appear to function correctly. Although the peaks are all centred on zero, in the case of two tracker points + the nominal vertex, the peak is considerably broader than that derived from three

tracker points + the nominal vertex, and that of the latter is marginally broader than that derived from three tracker points alone. Furthermore, inclusion of the nominal vertex constraint would also appear to be responsible for the introduction of the spikes at $\sigma_{p_T} > 0$. The fact that, for all methods of measurement, the momentum resolution is worse when including the vertex constraint, as compared to when excluded would suggest that there was a systematic error in the assumption that the interaction point is at the nominal beam-spot location of ($x=0$, $y=0$); if the true interaction point was offset from the nominal beam-spot, then the (incorrectly) assumed interaction point would skew any fit. Moreover, it would be expected that such an offset would create a ϕ -dependency of the reconstructed p_T : this effect was clearly demonstrated using both the nominal vertex location and a vertex offset by $400\text{ }\mu\text{m}$, and plotting the p_T -resolution against the reconstructed direction in ϕ . It was later revealed that a little-documented offset in the x-direction (of approximately $350\text{ }\mu\text{m}$) is included in the simulation, to thereby force analyses to take into account any possible beam offset and to remove the necessity of having a perfectly-controlled accelerator. The offset of $400\text{ }\mu\text{m}$ was determined by “stepping” the assumed vertex location across the x-y plane in steps of $100\text{ }\mu\text{m}$. The figure of $400\text{ }\mu\text{m}$ was used as the “adjusted vertex location” — and it is this value that is used for the rest of this chapter — prior to the discovery of the true offset.

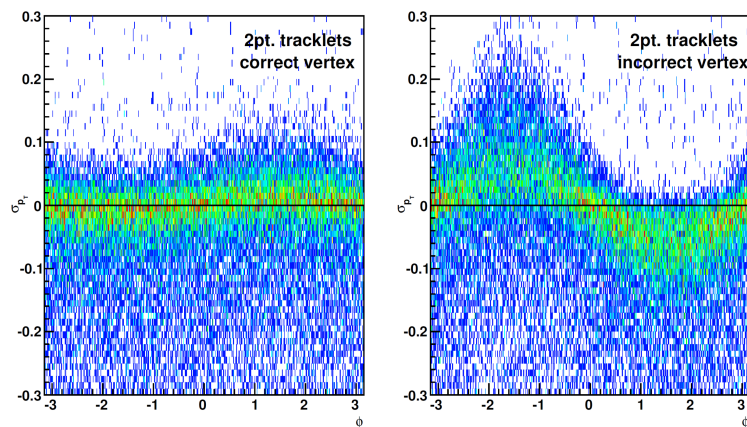


Figure 7.27: Simulated p_T -resolution vs. track direction in ϕ for 2-point tracklets with the nominal vertex set at $x=400\text{ }\mu\text{m}$ (left) and $x=0\text{ }\mu\text{m}$ (right). It can be seen that using $x=400\text{ }\mu\text{m}$ results in a slight over-compensation.

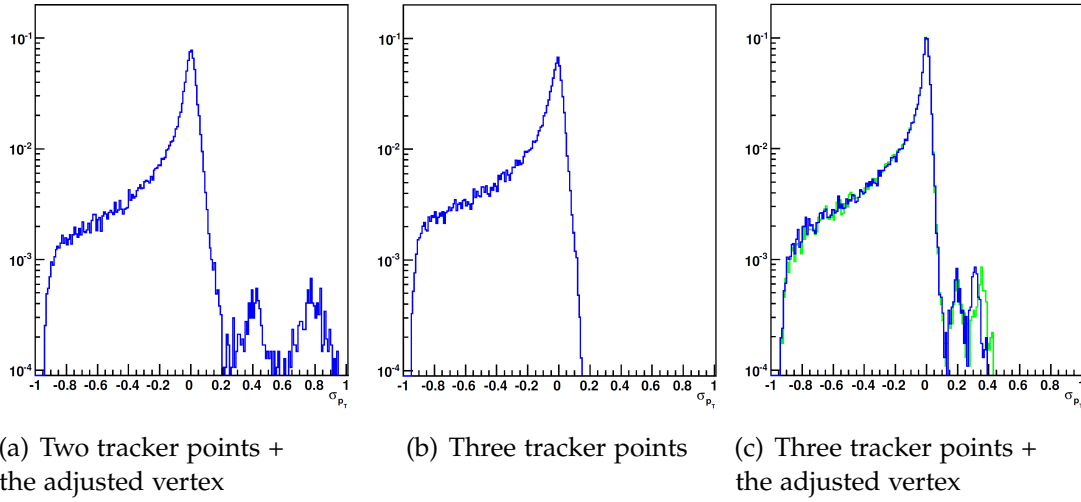


Figure 7.28: Simulated p_T -resolution calculated using two tracker points + the adjusted vertex, three tracker points and three tracker points + the adjusted vertex. Three different fitting techniques are used: the geometric method (red — overlapped by blue), modified least-squares method (blue) and perpendicular bisector method (green).

The p_T -resolutions were recalculated using the adjusted location of the “assumed” vertex (figure 7.28). Whilst adjusting the assumed location of the vertex results in a narrower peak, representing an improved p_T -resolution, it fails to remove the spikes at $\sigma_{p_T} > 0$: these spikes are caused by high- p_T secondary tracks, whose momentum measurement is being “pulled” upward by the incorrect assumption that the track passes through the primary vertex — using the adjusted vertex cannot rectify these cases as the assumption that the track passes through the primary vertex (whether its location is corrected or not) is entirely false.

Whether including the vertex constraint or not, in all cases there exists a broad shoulder at $\sigma_{p_T} < 0$, this signifying a systematic underestimate of the true p_T . It was thought likely that this effect was genuine and due to bremsstrahlung and material interactions; if this were the case, then what would appear to be an erroneous estimation of the p_T was not due to a failure of the track fitting, but attributable to the chosen method for calculating the momentum resolution. To test this hypothesis, the resolution was recalculated in two ways: first, using the Monte Carlo p_T at the calorimeter face, rather than at the point of production (figure 7.29) and, secondly, considering only those tracks where the p_T at the

calorimeter face was greater than 90%^{xiii} of the production p_T , in which case the Monte Carlo p_T was defined as the average of the value at production and the value at the calorimeter face (figure 7.30).

Consistent with the hypothesis of bremsstrahlung and material interactions, the track fitting consistently over-estimates the track momentum observed at the calorimeter face (figure 7.29). By further considering only the “clean” signals, both the over- and under-estimated cases are eliminated. Even in this “clean” scenario, however, it can be seen that the vertex assumption still results in badly-fitted tracks.

7.4.8 Vertex reconstruction

The demonstrated sensitivity of track reconstruction to a beam offset raised the issue of whether the tracking trigger could be used to detect such an offset and the more general issue of how well a track trigger could identify primary vertices within the luminous region.

The **DOCA** (distance of closest approach) was calculated from three-point track-lets, according to the relationship:

$$r_{DOCA} = \sqrt{x_{curve}^2 + y_{curve}^2} - r_{curve} \quad (7.2)$$

where r_{curve} is the radius of curvature and (x_{curve}, y_{curve}) is the centre of curvature of the circle fitted to the three tracker points.

The absolute value of the DOCA cannot, by itself, indicate the location of beam-spot: a track originating from a displaced vertex may pass through the nominal vertex, resulting in a DOCA of zero. Instead, the DOCA should be considered as a function of ϕ : in the case of a correctly-centred vertex, the DOCA should be

^{xiii} An arbitrary figure that was chosen so as to limit the “acceptable” energy loss, whilst maintaining reasonable statistics

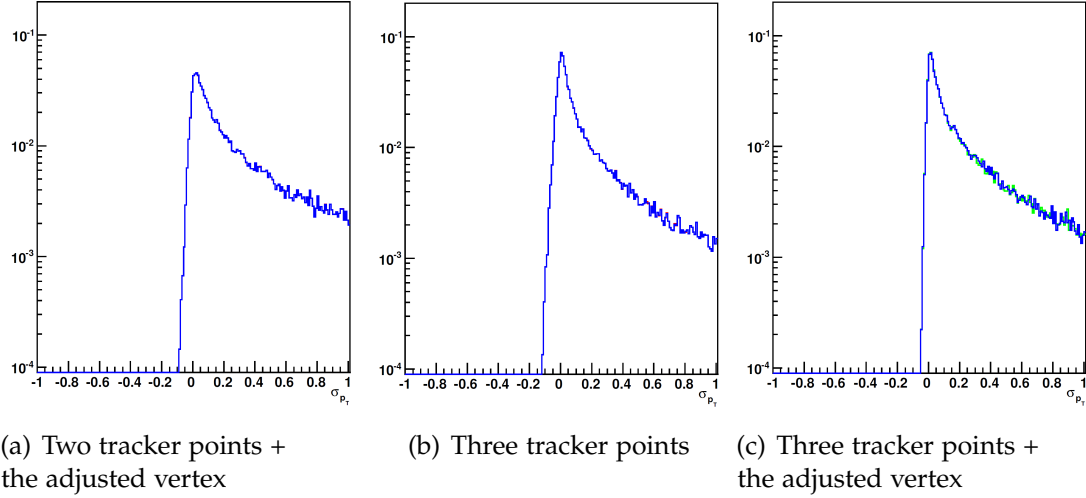


Figure 7.29: Simulated p_T -resolution calculated using two tracker points + the adjusted vertex, three tracker points and three tracker points + the adjusted vertex. Three different fitting techniques were used: the geometric method (red — overlapped by blue), modified least-squares method (blue) and perpendicular bisector method (green). The Monte Carlo p_T is here calculated at the calorimeter face.

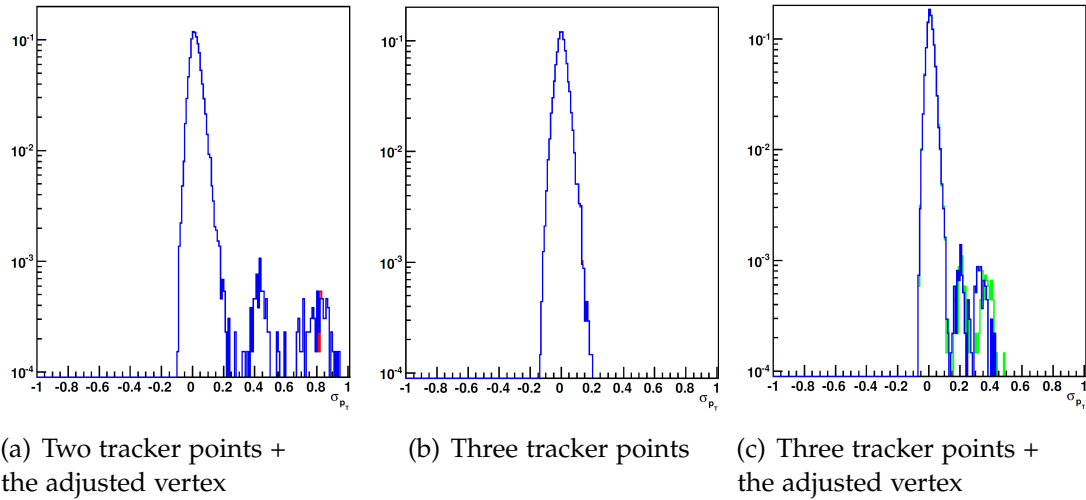


Figure 7.30: Simulated p_T -resolution calculated using two tracker points + the adjusted vertex, three tracker points and three tracker points + the adjusted vertex. Three different fitting techniques were used: the geometric method (red — overlapped by blue), modified least-squares method (blue) and perpendicular bisector method (green). Here considering only those tracks where the p_T at the calorimeter face was greater than 90% of the production p_T and the Monte Carlo p_T was then defined as the average of the value at production and the value at the calorimeter face.

uniformly zero in ϕ . In the case of a displaced vertex, the DOCA should vary sinusoidally: from zero for tracks passing through the nominal vertex, to the true offset for tracks at right angles to this, to twice the true offset for tracks pointing away from the nominal vertex. The phase-offset of the sinusoidal curve should indicate the direction of the physical offset of the beamspot. To test this conjecture, the DOCA was calculated for three-point tracklets (formed from stubs generated using 2-D clustering and the local-geometry hit-matching algorithm) with a nominal cut of 5 GeV, using 50,000 events consisting of an “electron gun” ($1 \leq p_T \leq 20$ GeV, $|\eta| \leq 1.5$) with 0 pile-up. For the sake of simplicity, rather than calculating the ϕ -direction of the track at the DOCA, the ϕ -location of the inner hit of the tracklet was used; this is a reasonable assumption provided that the p_T of the tracks being considered is sufficiently high that the tracks are approximately straight. The results are shown in figure 7.31, in which the fitted curve ignores any phase-offset and is defined by the equation:

$$r_{DOCA}(\phi) = \zeta \cdot (1 - \sin(\phi)) \quad (7.3)$$

where ζ is the predicted location of the beam offset.

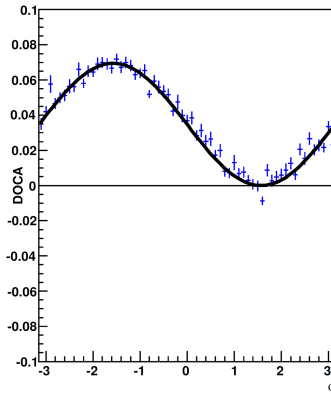


Figure 7.31: Simulated estimates of the distance of closest approach calculated using three-point tracklets plotted against the ϕ -position of the inner stub.

For all three methods of circle fitting, the location of the beam offset, ζ , was found to be $347 \pm 3 \mu\text{m}$ — in good agreement with the known offset. If such a measurement were to be made in the level-1, the implementation would undoubtedly be very different to the scheme presented here, however, such a

scheme could provide a useful, complementary measurement to the beam luminosity monitor already used in the calorimeter trigger.

To estimate the z -position of the interaction point, z_{IP} , a straight-line fit was projected to $r = 0$ in the rz -plane, this being performed for both two- and three-point tracklets. Although the significance of an offset beam-spot in the xy -plane has already been demonstrated, due to the length of pixels in the z -direction and the small size of a beam-spot offset as compared to the radii of the tracker layers, it was decided that taking the offset into account would provide only a minor correction and any benefit would be outweighed by the increased computational complexity.

In order to avoid problems at $z_{IP}^{M.C.} = 0$, the resolution of the z -position of the vertex, Δz_{IP} , was defined by

$$\Delta z_{IP} = z_{IP}^{reco.} - z_{IP}^{M.C.} \quad (7.4)$$

where the superscripts *reco.* and *M.C.* denote “reconstructed” and “Monte Carlo-truth”, respectively.

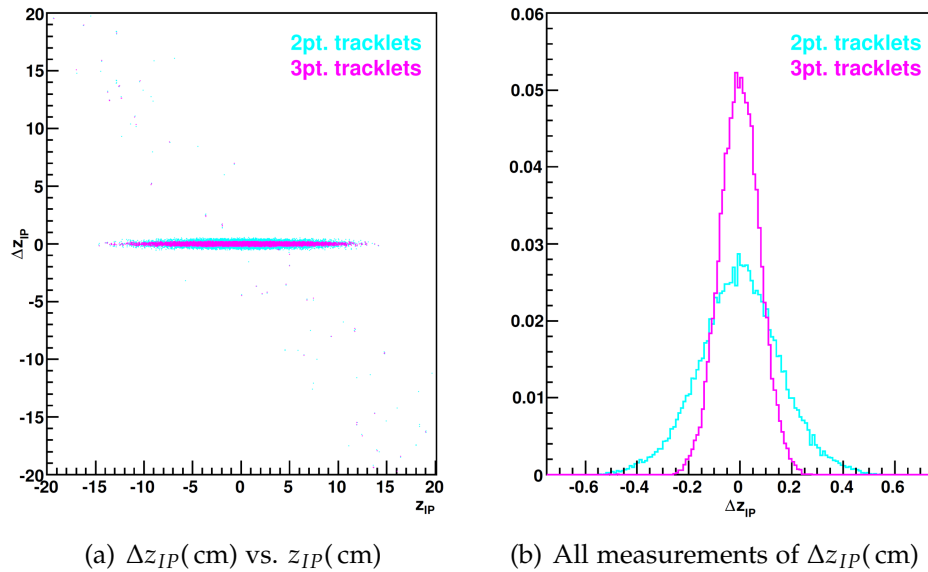


Figure 7.32: Simulated resolution of the calculated z -position of the interaction point for both two-point (turquoise) tracklets and three-point (pink) tracklets.

As may be seen in figure 7.32(a), there are two distinct features: a central region of hits with small Δz_{IP} and $|z_{IP}^{M.C.}| < 15$ cm and a small number of hits around

the line $\Delta z_{IP} = -z_{IP}^{M.C.}$. The central feature is due to tracks originating from the luminous region being reconstructed correctly. The latter feature is due to tracks originating from secondary vertices: although the vector is being correctly reconstructed, the projection to $r = 0$ (rather than to the true radial position of the vertex) results in an under-estimation of the true z-position. Considering all points and ignoring the z-dependence (figure 7.32(b)) results in Gaussian distributions, whose widths were found to be $\sigma_{2pt.} = 1.56 \pm 1.23$ mm and $\sigma_{3pt.} = 0.768 \pm 0.525$ mm.

7.5 Summary

It has been shown that the stacked tracker is, in principle, an effective method for reducing the tracker data rate, although this is dependent on several key considerations.

If no clustering is performed prior to hit-matching, large numbers of stubs are found for a single track. The multiplicities are worse still for tracklets, due to the large search windows needed to take into account the widely-spaced layers. Several clustering algorithms have been designed which may be implementable on-detector: the most advanced of these (2-D clustering) produces results which match well with the results that can be achieved using simulated, rather than digitized, hits.

The efficiency with which stubs are produced from hits is limited by geometric and algorithmic factors. Quantization errors result in a trade-off between the maximum achievable efficiency and the sharpness of the turn-on curve: the sharpness may be recovered by reducing the pixel size in the ϕ -direction, but this will result in an increased number of channels, more charge-sharing, higher power-consumption and increased cost. Inefficiencies in the production of stubs are multiplied when producing two- and three-point tracklets, resulting in low over-all efficiencies.

Under conditions of 200 pile-up events, the number of stubs in the central region is two orders of magnitude lower than the number of hits, whilst in the forward region the difference is three orders of magnitude. Formation of two-point tracklets achieves a further reduction in rate of less than ten-fold (although, there is an additional reduction associated with fewer “layers” to be read) and formation of three-point tracklets results in an additional reduction in rate by a factor of five (although, again, there is an additional rate reduction due to fewer “layers”).

In forming stubs from hits, the proportion of tracker objects associated with tracks above the nominal p_T threshold is boosted, but is still dominated by the background of low- p_T tracks. There is, in addition, a new background of stubs from mismatched hits, which also dominates the signal. Formation of two-point tracklets strongly suppresses the low- p_T background (due to the longer lever-arm and, thereby superior p_T resolution), but at the cost of an increased proportion of tracklets from mismatched hits: these mismatched tracklets may be suppressed by the formation of three-point tracklets.

The assumption of a primary vertex causes several problems: first, low- p_T secondary tracks may be misinterpreted as a high- p_T primary track and, secondly, an offset in the true interaction point results in a ϕ -dependent skewing of the p_T measurement, this resulting in a degraded p_T -resolution.

The use of three points in the tracker was shown to permit estimation of the distance of closest approach and plotting this value against the ϕ -position of the inner tracker point allows identification of a beam offset: whether such a scheme is useful (or implementable) in the level-1 trigger has not been considered in this investigation. The z -position of the primary vertex is considered likely to be useful in the level-1 trigger, by allowing improved association of multiple candidates. The resolution with which identification of the z -position of the primary vertex can be achieved using the Strawman-B geometry is approximately 1.5 mm for two-point tracklets and 0.77 mm for three-point tracklets.

The high material-budget of the tracker design is responsible for many of the observed problems: in particular, impurities (due to the production of secondary

tracks) and a limited ability to predict the transverse momentum of the track, either at production or at the calorimeter face (due to energy-loss). The efficiencies presented here (based on electrons) are lower than those presented elsewhere [123], these other studies being based on muon response and, as such, are (relatively) unaffected by the tracker material-budget.

Chapter 8

A Super-LHC Electron Trigger

“If Edison had a needle to find in a haystack, he would proceed at once with the diligence of the bee to examine straw after straw until he found the object of his search; I was a sorry witness of such doings, knowing that a little theory and calculation would have saved him ninety per cent of his labour.”

- Nikola Tesla (1857 - 1943)

“Results! Why, man, I have gotten a lot of results! I know several thousand things that won't work!”

- Thomas A. Edison (1847 - 1931)

The e/γ algorithm of the level-1 trigger is described in section 2.1.1 and the HLT algorithm in section 6.2. The HLT refines the level-1 calorimeter trigger objects by the use of ECAL crystal-level information and then iteratively includes tracking information (figure 8.1). The use of tracking information in the e/γ algorithm offers three key benefits over the sole use of calorimetric information:

- At the LHC, the level-1 trigger measures isolation of e/γ -candidates based on calorimetric information alone: at the HLT, isolation of energy deposits in the calorimeter becomes less significant, instead being replaced by isolation criteria defined by the tracker. At the SLHC, there are no isolated candidates in the calorimeter and, therefore, tracker information must play a more significant role.
-

- Minimum bias events generate large numbers of π^0 's, which may look like e/γ hits in the ECAL: tracking information not only distinguishes electrons from photons, but also allows for rejection of π^0 's. At SLHC, a ten-times higher pile-up will make the contamination from π^0 's even more significant and the ability to reject them even more important.
- The silicon tracker and associated infrastructure results in a very large material-budget (up to 1.75 radiation lengths at $\eta = 2$ [124]); this, along with the 4T magnetic field, resulting in a large amount of bremsstrahlung and pair production. Within the calorimeter, it is impossible to separate these secondary tracks from tracks originating at the primary vertex. The use of vertex information from the pixel layers in the HLT provides, through rejection of secondary tracks, a 10-fold reduction in trigger rate [125].

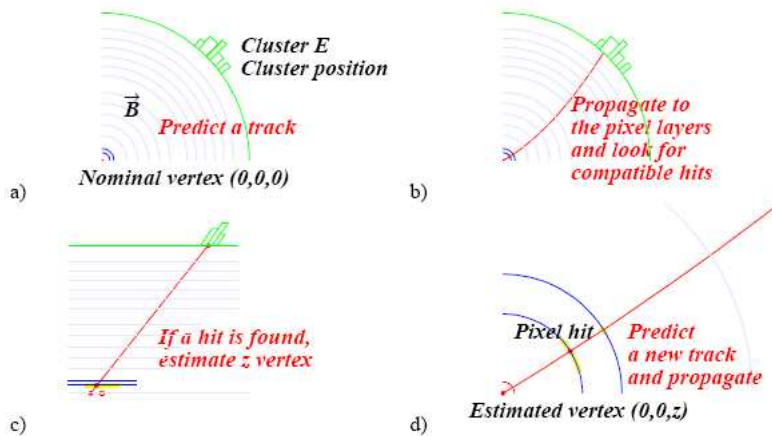


Figure 8.1: The HLT e/γ algorithm. Taken from [107].

8.1 Two Level-1 Electron Trigger Philosophies

An evident starting point for the use of tracking in a level-1 electron trigger is implementation of the HLT algorithm, with the HLT clusters replaced by level-1 e/γ candidates and the tracker hits replaced with stubs. However, an alternative

option also exists: for each tracker object consistent with a high- p_T track, a path is projected to the calorimeter and associated with a calorimeter object. The two philosophies are referred to as “outside-in” and “inside-out”, respectively (figure 8.2).

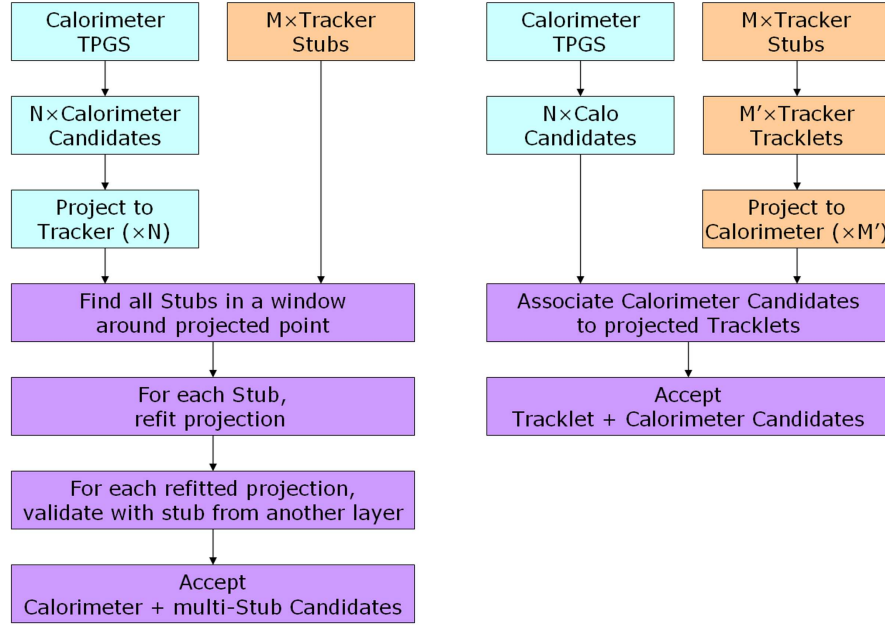


Figure 8.2: Two electron trigger philosophies: the HLT-like “outside-in” algorithm (left) and the alternative “inside-out” algorithm (right).

The “inside-out” philosophy offers several advantages over simply reimplementing the HLT algorithm:

- As discussed for calorimeter-seeded tracker readout (section 6.2), the size of the search window required in the tracker to achieve a good acceptance is large when projecting from calorimeter to tracker; this resulting in multiple association. In contrast, tracklets provide a pointing resolution similar in scale to the calorimeter candidates, resulting in a one-to-one, or one-to-zero, association.
- For each stub found in the window by the “outside-in” scheme, a second projection must be made between the stub and the calorimeter candidate and a second window constructed in a different tracker layer. Requiring a

second window opens the possibility of further multiple associations, even though the second window is far more tightly constrained than the first. The “inside-out” scheme requires no “second-pass”.

- Tracklets may be produced on-detector within the “inside-out” scheme, to reduce the bandwidth requirements for the optical links from the experimental cavern to the underground counting room.
- The “inside-out” scheme is closer to the design philosophy of the level-1 trigger: each subsystem generates its own trigger primitives and the primitives combined according to a simple set of criteria. Not only does the “outside-in” scheme use an iterative approach which is forbidden in the current system, but it renders the tracking trigger dependent on the calorimeter trigger.

While studies into the “outside-in” scheme are continuing [126], only the “inside-out” scheme is considered here.

8.2 The “Inside-Out” Electron Trigger

8.2.1 Preparatory study

From previous investigations (for example, section 7.4.7), it is known that both bremsstrahlung and the high material-budget of the tracker have a significant impact on electrons: of particular importance is the discrepancy between the transverse momentum measured from the tracklets and the transverse momentum measured at production and at the calorimeter faceⁱ. The calorimeter candidates are designed to recover as many of the secondary tracks and bremsstrahlung photons as possible to recover the original track momentum: if the secondary tracks are created before the p_T is measured, the tracklet will underestimate the

ⁱfigures 7.28 and 7.29

energy observed by the calorimeter. Secondary tracks produced after the p_T -measurement should be recovered by the calorimeter candidates and, therefore, should have less effect on the correlation of the E_T and p_T measurements: they will, however, result in a loss of ability to correctly predict the ϕ -location that the primary-track entered the calorimeter. Energy loss due to material interactions will have a similar effect.

The large size of the calorimeter candidates will, furthermore, result in contamination from pile-up events, increasing the discrepancy between the E_T and p_T measurements. The combined effect of energy-loss, secondary tracks and pile-up would then be expected to be that the calorimeter candidate will consistently measure a higher transverse energy than is measured by the tracklet.

The resolution with which tracklets could point to hits in the calorimeter was investigated, prior to the tracklets being associated with calorimeter trigger objects. For electron gun events ($1\text{ GeV} \leq p_T \leq 20\text{ GeV}$) with no pile-up, both two- and three-point tracklets were projected to the face of the ECAL and the position compared to the locations predicted by the Monte Carlo-truth tracks. To separate the effect of energy-loss from the effect of secondary tracks, two measurements were made: the first considering all tracks which reach the calorimeter and the second considering only the “true” track.

As expected from geometric calculations, both two- and three-point tracklets could identify the location of the true hit to within approximately 0.05 radians (figure 8.3) — which should be compared to the trigger-tower dimensions of 0.087 radians.

Given the superior z-vertex resolution of three-point tracklets over two-point trackletsⁱⁱ, three-point tracklets might likewise be expected to be able to identify the location of the calorimeter hits more accurately than two-point tracklets: this was the observed situation (figure 8.4)

ⁱⁱfigure 7.32

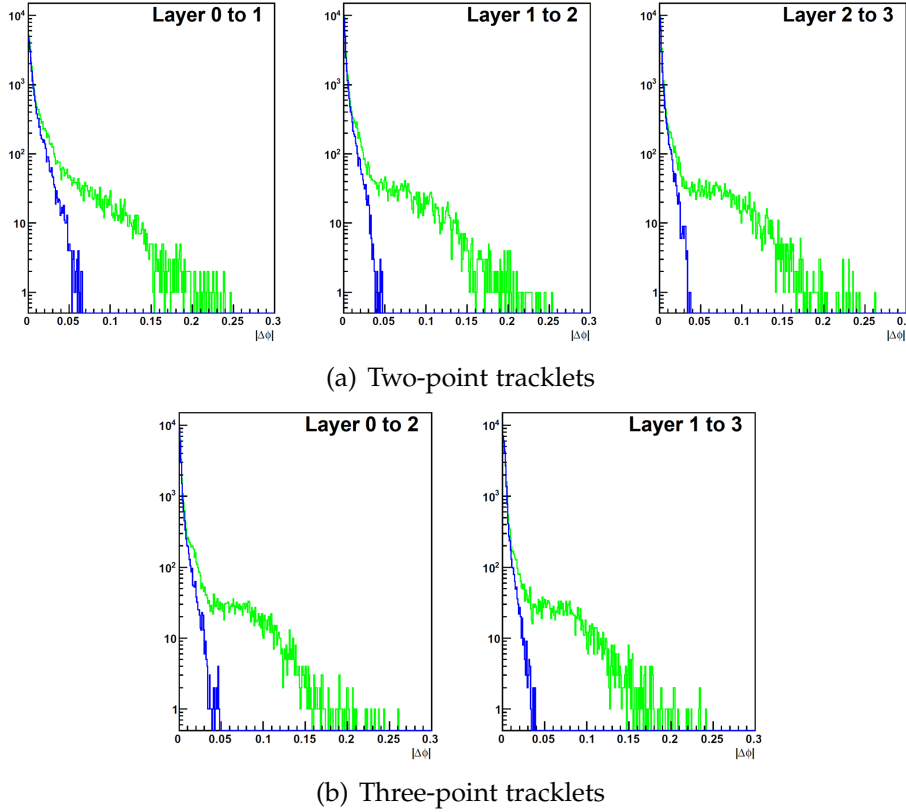


Figure 8.3: Simulated resolution with which two- and three-point tracklets can identify the location that tracks hit the calorimeter in the ϕ -direction. The green curve includes all hits in the calorimeter while the blue curve includes only the “true” track. For two-point tracklets, the projection is made using the “corrected” nominal vertex position ($x=350\ \mu\text{m}$).

The coarse granularity of the calorimeter trigger towers (0.087×0.087 in $\eta \times \phi$) results in a relatively poor position resolution. It is expected that the upgraded calorimeter trigger will use energy-weighted interpolation to improve on the resolution, but the improvement which may be achieved is limited. The position resolutions (in pseudorapidity and ϕ) of the upgraded level-1 e/γ candidates are shown in figure 8.5. Whilst the resolution with which the calorimeter trigger can identify the true hit location in η is marginally better than the size of a calorimeter tower, the resolution in ϕ is poor, as the primary track and secondary tracks originating from it are “clustered” into a single object.

Based on the measured position resolutions, the window sizes chosen for associating track and calorimeter candidates are shown in table 8.1.

Because of the problems associated with energy loss, secondary tracks and pile-

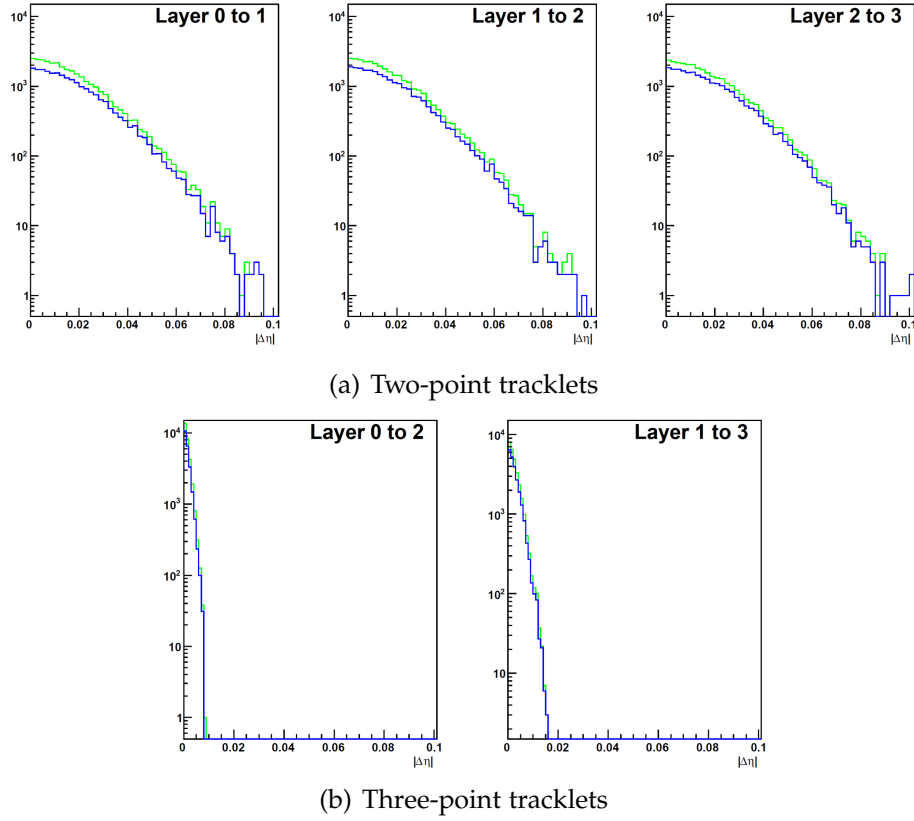


Figure 8.4: Resolution with which two- and three-point tracklets can identify the location that tracks hit the calorimeter in the η -direction. The green curve includes all hits in the calorimeter while the blue curve includes only the “true” track.

up, it was considered simpler to apply a p_T and E_T threshold to the tracklets and calorimeter candidates, respectively, rather than placing a cut on $\frac{E_T}{p_T}$, which would necessarily be loose. This threshold was set to be the same for both the tracklets and the calorimeter candidates.

8.2.2 Background Rates

To determine the rate of candidates, the applied threshold was varied and the number of calorimeter candidates, two-point+calorimeter candidates and three-point+calorimeter candidates recorded under conditions of 200 pile-up events with no signal: only the inner-most two- and three-point tracklets were considered. The mean rates are shown plotted against the threshold in figure 8.6. At all thresholds, the rate of two-point+calorimeter candidates is approximately a

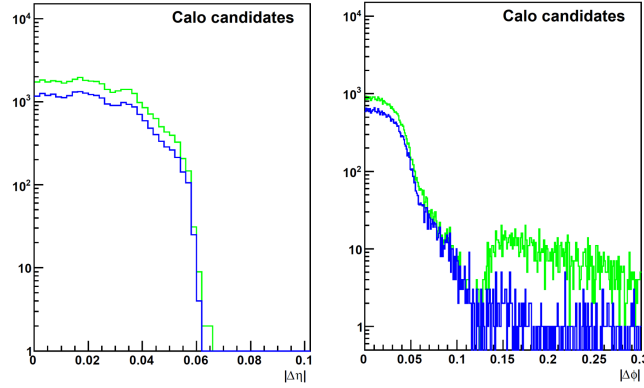


Figure 8.5: Resolution with which upgraded level-1 e/γ candidates can identify the location that tracks hit the calorimeter in the η - and ϕ - directions. The green curve includes all hits in the calorimeter while the blue curve includes only the primary track.

	$ \eta $	$ \phi $
two-point tracklets	0.1	0.25
three-point tracklets	0.07	0.25

Table 8.1: The chosen window sizes in η and ϕ for the association of tracker candidates with calorimeter candidates. All values are limited by the calorimeter trigger’s coarse granularity, apart from the $|\eta|$ window for two-point tracklets, which is limited by the length of the pixels in the z -direction.

factor of 20 lower than the rate of calorimeter candidates. The rate of three-point+calorimeter candidates is approximately a factor of 50 lower than the rate of calorimeter candidates.

8.2.3 Efficiency

It is known from section 7.4.2 that the efficiency of creating two-point tracklets is relatively low ($\sim 80\%$) and that the efficiency for three-point tracklets is even lower ($\sim 65\%$). The number of calorimeter candidates, two-point+calorimeter candidates and three-point+calorimeter candidates were, again, recorded as a function of the E_T and p_T threshold for two different scenarios: first, using a single-electron gun with $p_T^{Thresh.} \leq p_T^{M.C.} \leq 20$ GeV and 200 pile-up events (figure 8.7(a)); and secondly, using a single-electron gun with $p_T^{M.C.} = 50$ GeV and 200 pile-up events (figure 8.7(b)). The two case describe the “efficiency” when the applied

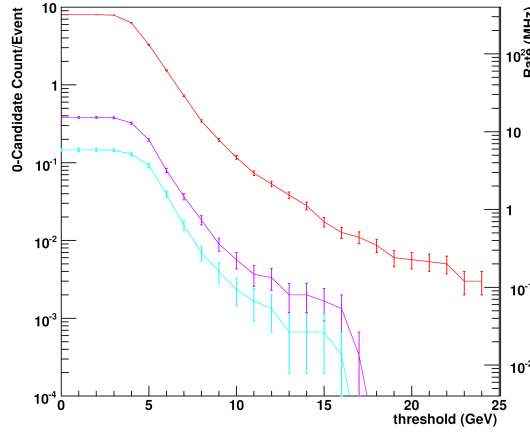


Figure 8.6: Mean rate of calorimeter candidates (red), two-point+calorimeter candidates (violet) and three-point+calorimeter candidates (cyan) as a function of applied E_T and p_T threshold for background events.

threshold is close to the true p_T and when the threshold is much lower than the true p_T , respectively. Whilst the latter scenario represents a “true” efficiency — it being expected that all candidates should be found — the former scenario is only an approximation of the efficiency, based on the caveat that, at $p_T^{Thresh.}$, the true p_T and the measured p_T differ by only a multiplicative factor, with no constant offset (that is, $\overline{p_T^{M.C.}} = \overline{p_T^{tracklet}}$) and equivalently for the measured E_T (similarly, $\overline{E_T^{M.C.}} = \overline{E_T^{calorimeter}}$). The p_T resolutions shown in section 7.4.7 indicate that this assumption is valid for tracklets; similar studies for the calorimeter trigger candidates indicate the assumption is valid for these also.

For the case where the threshold is much lower than the candidate momentum, the efficiency with which calorimeter candidates identify electrons is greater than 90%, whilst the efficiency for the three-point+calorimeter candidates and two-point+calorimeter candidates are greater than 65% and 70%, respectively. When the threshold is close to the candidate momentum, the efficiency is considerably lower: for the two-point+calorimeter and three-point+calorimeter candidates, this is associated with the slow efficiency turn-on curves described in section 7.4.2, however, a similar effect is observed in the calorimeter candidates, which will also contribute to the inefficiencies of the two-point+calorimeter and three-point+calorimeter candidates.

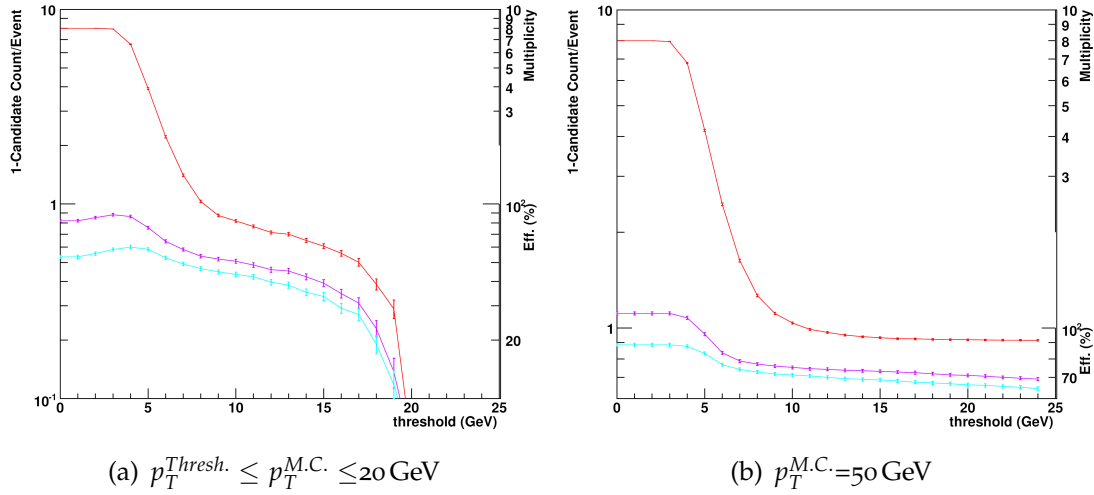


Figure 8.7: Efficiency/Multiplicity of calorimeter candidates (red), two-point+calorimeter candidates (violet) and three-point+calorimeter candidates (cyan) as a function of applied E_T and p_T threshold for: (8.7(a)) single-electron events with $p_T^{M.C.}=50$ GeV and 200 pile-up events; and (8.7(b)) single-electron gun with $p_T^{M.C.}=50$ GeV and 200 pile-up events.

For the case where the threshold is much lower than the candidate momentum then, if the threshold is too low, both the calorimeter candidates and two-point+calorimeter candidates find, on average, more than one object per event: in defining the efficiency as it has been, it is assumed that, above the threshold where less than one object is found per event, the object found is the candidate track that is being sought.

8.2.4 Balancing efficiency and rate reduction

For each E_T (p_T) threshold it is possible, to define a point in the 2-D “background rate vs. efficiency” space, the continuum of possible thresholds being described by a line in this space (figure 8.8): the efficiency in this plot being that defined by the case where the threshold is much lower than the candidate momentum.

At the level-1 trigger, the maximum acceptable rate for electron candidates is $\mathcal{O}(10 \text{ kHz})^{\text{iii}}$. This cannot be achieved using calorimeter candidates alone, with the range of thresholds currently under consideration. Furthermore, the points (which are evenly-spaced in energy) become more closely-spaced in “rate vs.

ⁱⁱⁱor $\mathcal{O}(10^{-2} \text{ MHz})$ in the units of figure 8.8

efficiency” as the threshold rises: extrapolating this trend to the acceptable background rate would indicate that a very high threshold would be required. The acceptable background rate may be achieved by using two-point+calorimeter candidates and three-point+calorimeter candidates: at this rate, the efficiency of the two-point+calorimeter candidates is $\sim 5\%$ greater than that of three-point+calorimeter candidates, however, the threshold required for three-point+calorimeter candidates is lower than two-point+calorimeter candidates.

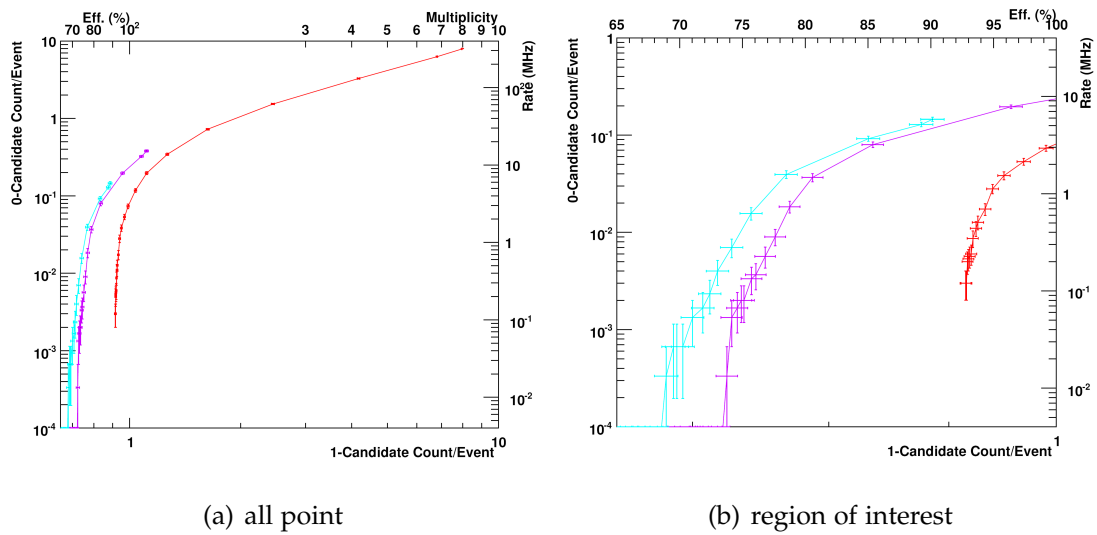


Figure 8.8: Background rate as a function of single-particle efficiency/multiplicity for calorimeter candidates (red), two-point+calorimeter candidates (violet) and three-point+calorimeter candidates (cyan). The left-hand plot shows points corresponding to all thresholds in the range 0-25 GeV, the right-hand plot shows the region of interest.

8.3 Summary

The “inside-out” electron trigger using two-point tracklets can successfully reduce the rate of “fake” e/γ -candidates by a factor of around 20 for a given threshold with respect to the calorimeter trigger alone: the penalty for doing so, however, is an efficiency of down to $\sim 75\%$ when the threshold is much lower than the candidate momentum. Using three-point tracklets results in a greater rate reduction (a factor of around 50 for a given threshold with respect to the calorimeter trigger alone) but at a more severe cost to the efficiency — down to $\sim 70\%$ — when the threshold is much lower than the candidate momentum.

Chapter 9

Conclusions

“We learn something every day,
and lots of times it’s that what we learned the day before was wrong.”

- Bill Vaughan (1915 - 1977)

The CMS detector at the LHC is fully installed, operational and awaiting the first collisions from the freshly-repaired accelerator. Until that time, the detector and data-acquisition systems are being run, calibrated and investigated by means of collecting data from cosmic-muon interactions.

The subsystems of the Global Calorimeter Trigger were successfully designed, manufactured, tested and installed, and the whole system tested, between January 2006 and August 2008. The Calorimeter Trigger was ready for data acquisition at the first circulation of beams within the LHC, on 10th September 2008 and has been included in all cosmic-muon data-taking runs since that time.

To achieve the rapid development required by the tight schedule, a new design-philosophy was used: small, modular boards were interconnected by high-speed serial links based on the integrated SerDes in modern large FPGAs. This design approach was taken to its ultimate realization in the matrix processor card, the design of which is independent of a specific purpose. Those involved in the level-1 trigger have taken the matrix processor card as a template for the

future development of the trigger: by eliminating the multitude of different hardware designs and replacing them with a single, high-performance, processing card, the problems of hardware manufacture are minimized, the problems of subsystem incompatibility eliminated, and the implementation of functionality changed from a hardware to a firmware design issue. Based on the ever-increasing bandwidth requirements of commercial products and the competition between FPGA manufacturers, it is reasonable to suppose that the next generation of FPGAs will provide not only higher logic densities, but also serial links that are both faster and more numerous, at lower cost: such devices look likely to facilitate previously impossible tasks — such as the inclusion of tracker data in the level-1 trigger. Even with such processing power, however, novel techniques — such as the stacked tracking concept presented in section 6.3 — will be required to reduce the data rate to a manageable level.

There is continuing research into understanding the scale of difficulties faced at the SLHC. It is evident that tracking information will be required to recover trigger performance under the conditions of increased pile-up. A scheme — termed stacked tracking — has been proposed which can reduce the tracker data rate by imposing an on-detector cut on transverse momentum: such a scheme would face many technical challenges in its implementation, such as the need for on-detector clusterization and hit-association. The stacked tracker also faces logistical challenges, such as minimizing the power consumption, providing sufficient cooling and minimizing the tracker material-budget. Many subtleties and trade-offs would need to be considered in the design of the stacked tracker, in particular trying to achieve a balance between the needs of the trigger and the performance of the tracker.

Further investigations into the algorithms for the upgraded trigger are continuing and it is believed that, by using a combination of tracker information and upgraded calorimeter trigger information, the level-1 electron trigger rate may be recovered: two tracker layers can, with an efficiency of $\sim 75\%$, reduce the rate of “fake” e/γ -candidates by a factor of around 20 for a given transverse energy

threshold with respect to the calorimeter trigger alone; whilst three tracker layers provide a means to reduce the background rate by a factor of around 50 (with respect to the calorimeter trigger alone) at an efficiency of $\sim 70\%$. Trigger upgrade studies with muons, τ 's and jets are also being performed and early results indicate that the performance of these trigger channels may also be recovered.

Appendix A

The Front End Test Stand laser-wire

“If you can’t stand the heat, get out of the beamline.”

- Paraphrasing Harry Vaughan (1888 - 1964)

Although the unprecedented size, energy and luminosity of the LHC undoubtedly make it an impressive machine in its own right, it does not generate its own beam; instead, accelerating an existing beam. The injection chain of the LHC consists of the H^+ DuoPlasmatron, CERN Linac 2, the **PSB** (Proton Synchrotron Booster), the **PS** (Proton Synchrotron), the **SPS** (Super Proton Synchrotron) and finally the LHCⁱ. This is an impressive achievement considering that the PS reached its 50th year of operation in 2009: although much of the PS has been upgraded, such that it now operates at energies 3 orders of magnitude higher than in 1959, it still contains some original features. For the luminosity upgrades required by the SLHC, however, the PS is running close to its limits and cannot be used; instead it is proposed that a new linac will be constructed (linac 4) followed by a further booster linac – the **LPSPL** (Low Power Superconducting Proton Linac) – and a replacement for the PS, the PS2 (figure A.1).

ⁱHistorically, the chain of the PSB, PS and SPS were comically referred to as the “Fellowship of the Rings”; this nomenclature, of course, naturally lends itself to referring to the LHC as the “Lord of the Rings” [127].

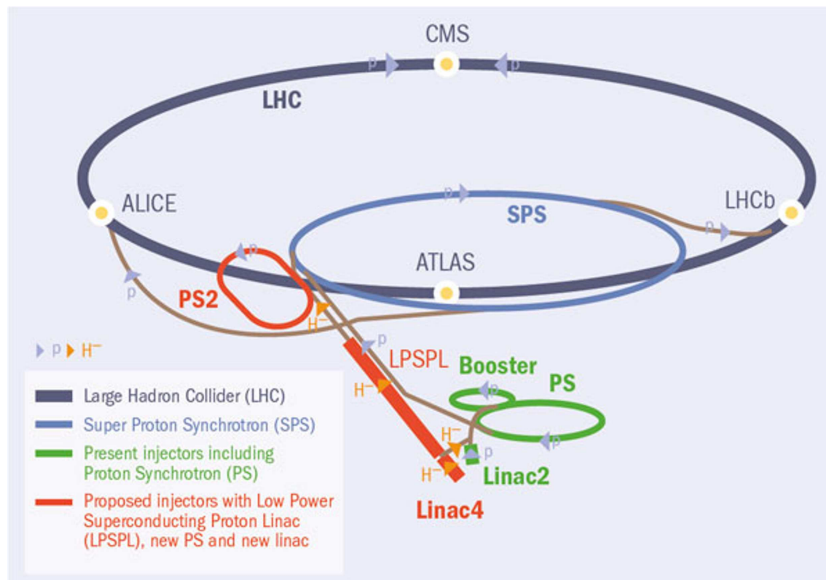


Figure A.1: The upgraded injector chain proposed for SLHC.

A limiting factor in the current design is that the injection of protons into an accelerator or proton storage ring is fundamentally limited (according to Liouville's theorem) in the size of the phase-space which may be achieved: it is not possible for the injected path and the circulating path to coincide, as the phase-space must be conserved along a path. Instead, rather than using an H^+ source (as in the current design), an H^- source will be used at the SLHC complex, with the H^- ion beam being accelerated and injected into the PS2 by means of a chicane magnet, which bends the injected ion beam and circulating proton beam into the same trajectory (which does not violate Liouville's theorem). Passing the beam through a stripper foil (to strip the ions of their electrons) results in a high-intensity proton beam. Although multiple Coulomb-scattering in the stripper foil causes increased emittance and the intensity of H^- sources is an order of magnitude lower than H^+ sources, these effects are more than compensated by the increased injection efficiency of such a scheme [128]. Although this technology is well understood at the scales required for SLHC, investigation into more extreme limits of the technology are currently under way.

A.1 The Front End Test Stand (FETS)

The **FETS** (Front End Test Stand) experiment at **RAL** (Rutherford-Appleton Laboratory) in Oxfordshire is a testbench for new technologies required for high-powered pulsed proton drivers. Pulsed proton drivers with a beam power of up to 5 MW will be required for neutrino factories and the next generation of neutron spallation sources: such beam power represents a significant challenge, the current record being held by the ISIS experiment (also at RAL) with a beam power of only 0.16 MW. The FETS, built in collaboration between ISIS, ASTeC, Imperial College and the University of Warwick, is designed to test the first stages needed to produce a very high quality, chopped H^- ion beam with an energy of 3 MeV, a beam current of 60 mA at 50 Hz repetition rate and a pulse durations of up to 2 ms. The FETS consists of an H^- ion source, a **LEBT** (Low Energy Beam Transport) section, an **RFQ** (Radio Frequency Quadrupole), a beam chopper and a **MEBT** (Medium Energy Beam Transport) section followed by a set of conventional (destructive) and novel non-destructive beam diagnostic devices.

The conventional method of beam diagnostics is a “Pepper-pot” apparatus, which allows 4-D measurement of the beam’s emittance profile. The principle behind a Pepper-pot scanner is that the beam is segmented into a 2-D array of beamlets in the x-y plane, the beamlets being allowed to diverge briefly and then intercepted by a scintillating screen. A fast high-resolution CCD camera records the pattern of scintillation light and, from this, the intensities of the beamlets identified and the profile of the original beam estimated. By placing the apparatus on a stage that may be moved in the z-direction and by using a fast camera, the longitudinal beam shape (z) and time variation may also be measured. The intensity of the proton beam, however, caused many problems with the Pepper-pot scheme, in particular a complete destruction of the scintillator materials (figure A.2).

Clearly such damage is unsustainable and having to intercept the beam in order to measure the profile undesirable. As such a novel, non-invasive, non-

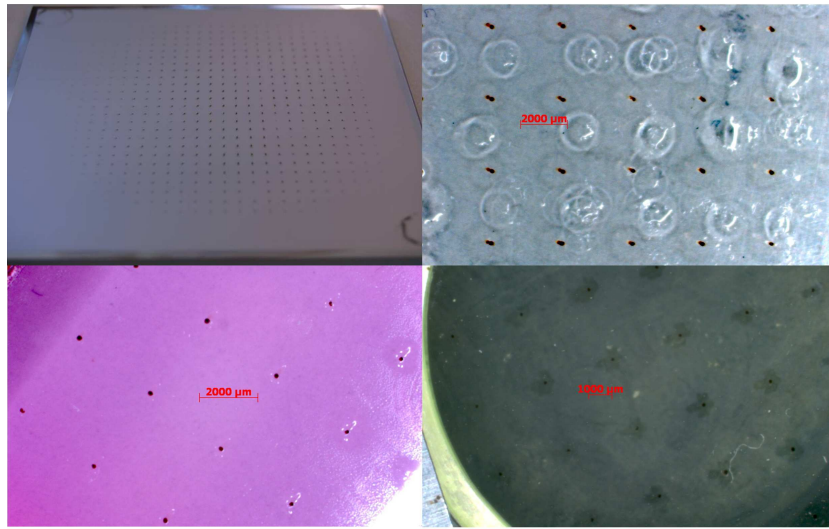


Figure A.2: Damage done to the scintillator screens at the FETS. Top left: P46 phosphor screen - instantaneous explosion. Top right: Plastic scintillator - melted . Bottom left: Ruby scintillator as used at ISIS - burned. Bottom right: Cerium-doped YAG - unusable after 3 hours beam-time. P43 & P47 phosphor screens, Cadmium Tungstate scintillator and **LYSO** (Cerium-doped Lutetium Yttrium Orthosilicate) scintillator suffered similar damage. Only pure quartz and Cerium-doped quartz showed resistance to damage.

destructive approach is being developed and tested at the FETS, called a “laser-wire” [129]. H^- ions can be made to photo-dissociate, with the maximum dissociation occurring at a wavelength of 840 nm. By subsequently passing the beam through a dipole magnet, the untouched H^- ions, the H_0 atoms and the stripped electrons may be separated and the electrons collected in a Faraday cup for diagnosis, while the H^- ion beam proceeds virtually unscathed. At FETS, the H_0 beam is directed to a conventional Pepper-pot scanner to allow for direct comparison of the beam profile as seen by the electron collection method and by traditional techniques: if proven successful, inclusion of the traditional scanner would not be necessary. The laser used at FETS is a 500 mW, 671 nm diode-pumped solid-state laser focussed to a beam diameter of 1.5 mm, the narrowest parallel beam achievable, for passing through the ion beam. Such a configuration results in an ion dissociation fraction of 10^{-7} , corresponding to a peak signal of approximately 7×10^7 electrons per 2 ms pulse. By measuring the charge collected, the number of ions within the laser beam may be estimated and a 1-D profile of the beam constructed. By using a series of rotary mirrors and linear stages, however, the system has been designed to allow the laser beam to pass

through the x-y plane of the ion beam at any angle (figure A.3). By measuring 1-D profiles from a large number of angles, a 2-D beam profile may be reconstructed using an **ART** (Algebraic Reconstruction Technique) [130].

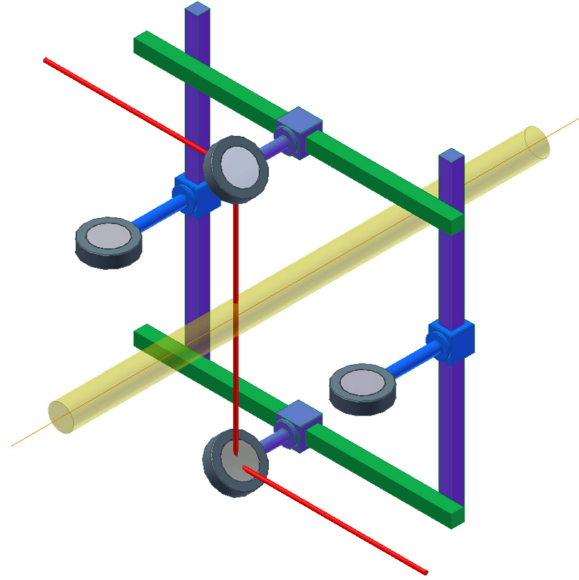


Figure A.3: The system of transposable and rotatable mirrors used to control the FETS laser-wire. The ion beam is represented in yellow and the laser beam in red. The horizontal (green) and vertical (dark blue) bars represent fixed axes along which the mirrors may be transposed. The cylinders on which the mirrors are supported indicates the axis of rotation. Beyond an angle of $\pm 45^\circ$ from the vertical, the optics change from using just two mirrors to using all four.

A.2 A readout system for the FETS laser-wire

The readout system of the laser-wire presented several major design challenges: the system must be able to accurately detect charges up to 10 pC with a 2 ms integration time and 50 Hz repetition rate. Because of the small charges involved, the readout system must be located as close to the Faraday cup as possible to minimize the introduction of noise. As the Faraday cup is located within the vacuum vessel and there is no suitable site for locating readout electronics on the outside of the vessel, it was decided that the readout system should be located inside the vessel itself: such a choice requires careful design for the management of heat, due to the lack of convectional cooling. Low-power components were

used where possible and variation in component heights minimized so that a heat-sink might be placed across the highest power components and connected to the walls of the vacuum vessel. The size of the board is also constrained by the size of the vacuum vessel.

The **ADC** (Analogue to Digital Converter) chosen was the Texas Instruments DDC112 [131]. The DDC112 a dual-channel charge-digitizing ADC with 20-bit resolution and digitally programmable range, the lowest being the range -0.2 to 50.0 pC. The readout system was controlled using an Atmel ATmega128 [132] microcontroller, clocked at 7 MHz and implementing the RS-232 serial protocol. A pair of lemo sockets are connected to interrupt-enabled inputs on the controller to synchronize the readout to the machine clock and any unused I/O pins are routed to 0.1 inch headers for use if they should be required. Although the RS-232 connection was originally intended to use a standard socket, insufficient room in the vacuum vessel necessitated connecting the cable (via flying-lead connections) to header pins. A bank of LEDs is included for debugging, although clearly these are not visible when inside the vacuum vessel and, to reduce the power consumption, can be disabled by the removing of jumpers. The board is powered by a low-noise 5 V power supply outside the vacuum tank; the only on-board power supply is a stabilized, precision, low-power, 4.096 V reference [133] for the ADC. The readout board is connected to the Faraday cup via a lemo cable; as the cup is held at a bias voltage of 747 V, however, a capacitive coupling must be included between the connector and the ADC, to isolate the DC bias of the two systems. The laser-wire readout board may be seen in figure A.4.

A.3 Tests of the FETS laser-wire readout system

Once the RS-232 communications had been successfully demonstrated by the correct capture and transmission of counter and bit-shifting patterns, both to and from the microcontroller, it was possible to test the ADC; this was achieved in two ways.

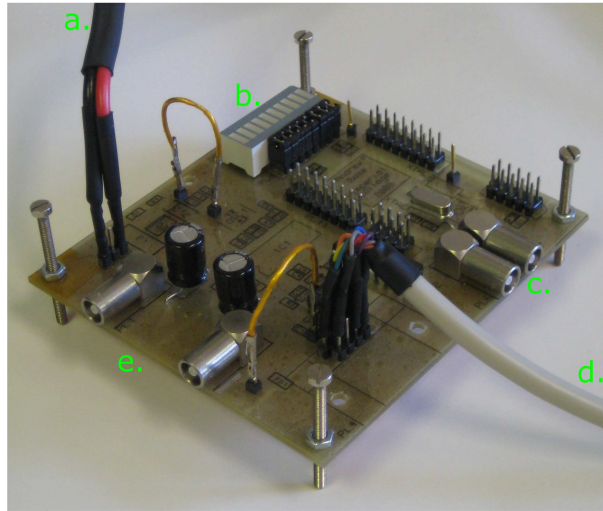


Figure A.4: The readout board of the FETS laser-wire system. All “hot” components are located on the underside of the board; the board is shown standing on the bolts which are used to hold the “hot” components in contact with a heat sink. The labelled parts are: a. power supply leads for bench-top tests, b. LED bank and jumpers for disconnecting it, c. two lemo connectors for synchronization with the machine clock, d. RS-232 connection via flying-lead connections and e. two lemo connectors for the inputs to the ADC.

The DDC₁₁₂ ADC includes a test mode allowing the injection of a 13 pC (with a stated accuracy of 10%) to be injected directly into each of the integrators. Both integrators on the ADC were subjected to 55,000 injection and readout cycles and the results histogrammed (figure A.5).

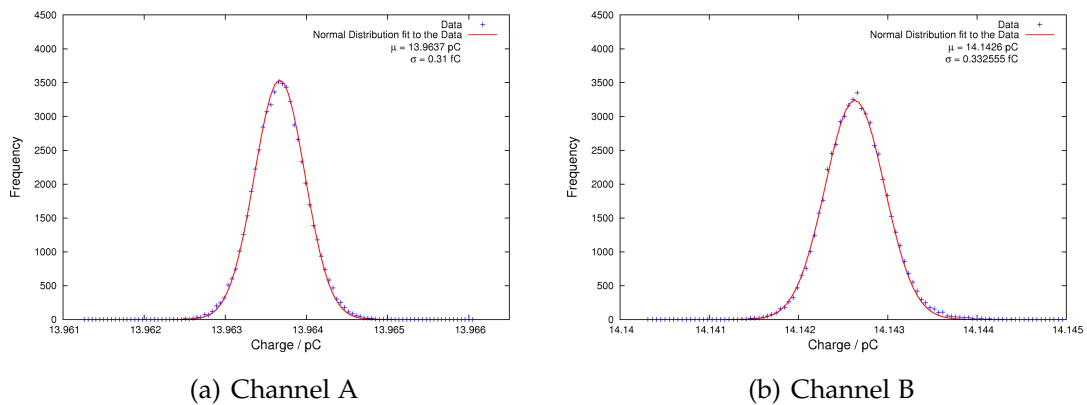


Figure A.5: Results of 55,000 charge injection and readout cycles of the DDC₁₁₂ ADC on the FETS readout board.

Both channels report a mean charge within the 10% accuracy of the nominal 13 pC injected charge. The observed resolution of around 0.3 fC should be com-

pared to the expected quantization errorⁱⁱ of 0.0239 fC: the difference is because the observed resolution is a convolution of the inherent resolution of the ADC (which is determined by the quantization error) and of the error associated with the charge injection (which is not quoted by the manufacturer). A resolution of 0.0239 fC corresponds to around 150 electrons, while a resolution of around 0.32 fC corresponds to around 2000 electrons. This should be compared with the expected maximum signal of 7×10^7 electrons and, as such, even the worst case resolution still provides an excellent dynamic range.

Once the apparatus had been installed at the FETS, the Faraday cup and readout electronics were run synchronously with the ion beam (but without using the laser) to examine backgrounds and noise performance. An unexpected result was the total saturation of the ADC: the bias voltage on the Faraday cup provided sufficient acceleration to cause inelastic collisions with the cup's surface, resulting in up to 45 nC of charge being collected; this is significantly more than the 350 pC maximum range of the readout electronics. By placing a guard mesh with an adjustable voltage over the face of the cup, the total charge collected could be controlled and brought within the dynamic range of the electronics. The source of the particles being collected is the interaction of the ion beam with residual gas in the accelerator: large volumes of hydrogen gas must be ionized to achieve the aforementioned beam intensity and it is estimated that up to 20 mL/s of unionized hydrogen atoms escape from the ion-source into the beam pipe. With a guard mesh voltage of 250 V, the readout system was tested and the distribution of charge recorded (figure A.6). The average charge collected due to gas-beam interactions far exceeds the charge expected from ion stripping by the laser-wire; worse still, the width of the distribution exceeds the charge expected from the laser-wire, hiding any excess charge in the background.

To improve the signal-to-background ratio it is proposed that the laser-wire should be moved 10 m further down the beam-line, where the residual gas pressure is two orders of magnitude lower, and to use a laser with a significantly

ⁱⁱThe quantization error is calculated as $\epsilon_{\text{quantization}} = \frac{1}{2} \frac{R_{\text{max}} - R_{\text{min}}}{2^Q}$, where R_{max} and R_{min} are the overflow and underflow values and Q is the number of bits. In this case, $\epsilon_{\text{quantization}} = \frac{1}{2} \frac{50 \text{ pC} + 0.2 \text{ pC}}{2^{20}} = 0.0239 \text{ fC}$.

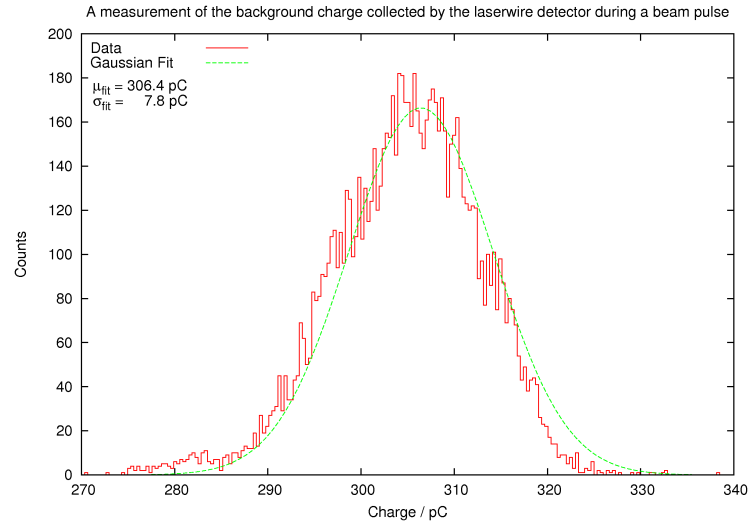


Figure A.6: Background charge collection using the FETS readout board running synchronously with the ion beam.

increased power. As a demonstration of the readout system, however, this test was considered a success and no adaptation of the readout electronics is considered necessary at the present time.

Appendix B

Circular Geometry

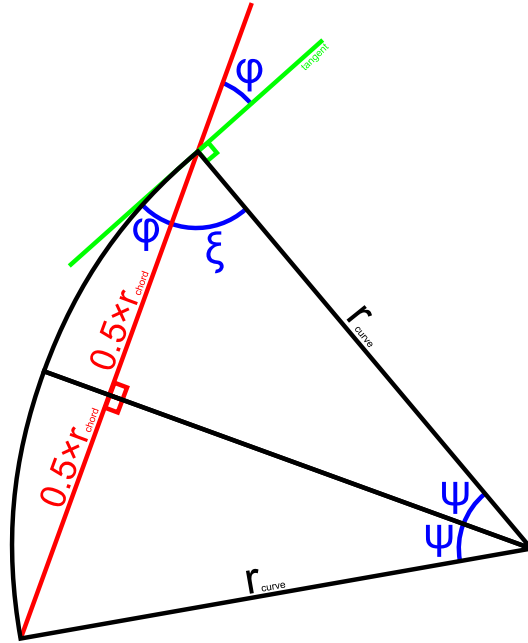


Figure B.1: The geometry relevant to the stacked tracker. The crossing-angle, ϕ , is defined to be the angle between the tangent(green) and the radial vector(red). The chord is the section of the radial vector bound by the arc.

Using the geometry shown in figure B.1: By definition, $\psi + \xi + \frac{\pi}{2} = \pi$ and $\phi + \xi = \frac{\pi}{2}$ and, therefore, $\phi = \psi$.

However, ψ is defined by the trigonometric relationship

$$\sin(\psi) = \frac{0.5 \cdot r_{chord}}{r_{curve}} \quad (\text{B.1})$$

and, therefore, we may define the crossing angle, ϕ , by

$$\sin(\phi) = \frac{0.5 \cdot r_{chord}}{r_{curve}} \quad (\text{B.2})$$

Appendix C

2-D Clustering Algorithm

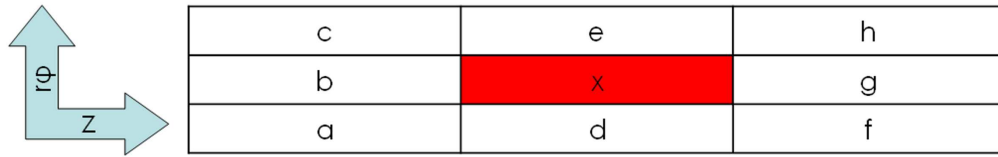


Figure C.1: The 2-D clustering algorithm. If a central pixel, x , is over threshold, it is only considered to have fired if it, and all of its neighbours (a-h) fulfil certain Boolean criteria.

Using the labelling scheme in figure C.1, three “kill” criteria are applied to pixel- x : only if all three criteria return false is it passed to the hit-matching algorithm.

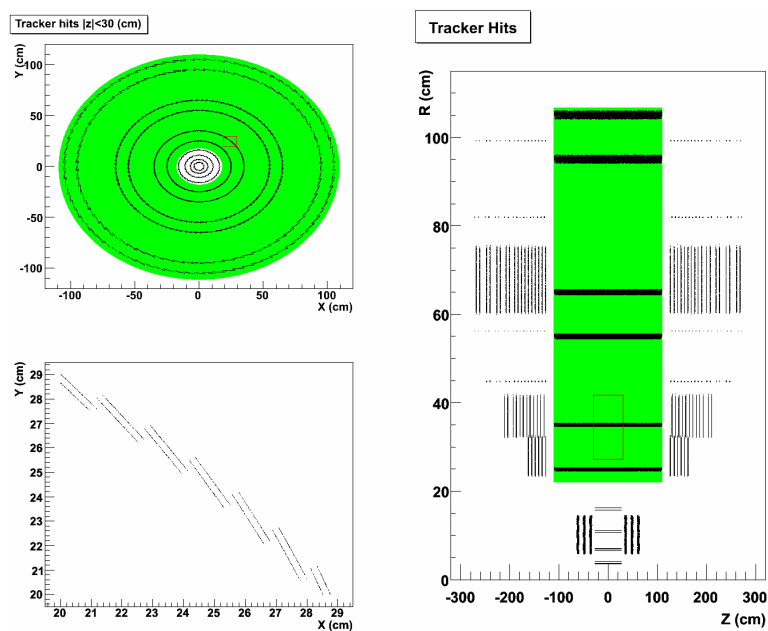
For pixel x , kill-bit 0 (x_{kb0}) is defined as $x_{kb0} = (a \vee d \vee f) \cap (c \vee e \vee h)$, where $\vee$ indicates Boolean “OR” and $\cap$ indicates Boolean “AND”: this condition restricts the cluster to being less than three pixels wide.

For pixel x , kill-bit 1 (x_{kb1}) is defined as $x_{kb1} = a \vee d \vee f \vee g$: this forces the cluster to be centred on either a , d , f or g . Although by doing so, we introduce a bias towards pixel- f , a bias will be present in all schemes without a method of representing “half”-pixels.

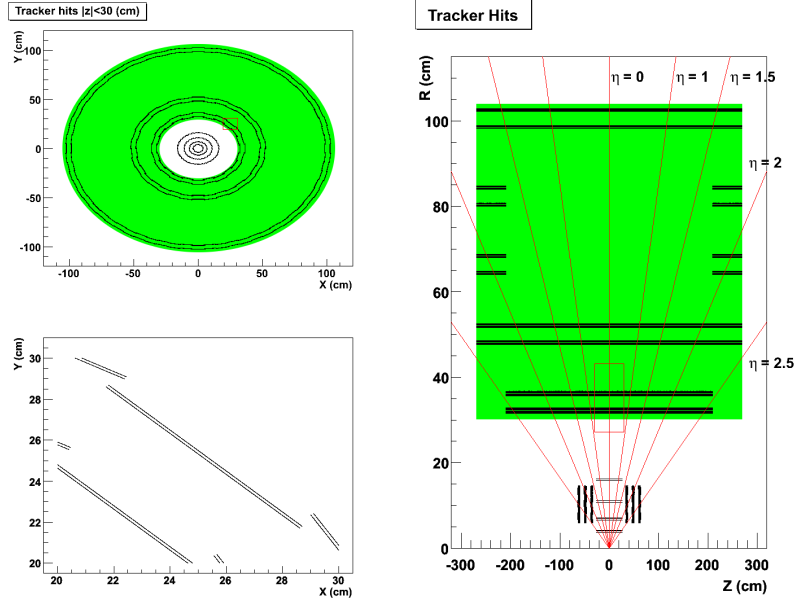
For pixel x , kill-bit 2 (x_{kb2}) is defined as $x_{kb2} = c_{kb0} \vee e_{kb0} \vee h_{kb0}$: if this condition is met, then pixel c , e or h has found a three-pixel wide cluster that includes pixel x , and therefore, pixel x should be excluded to limit the cluster width to less than three pixels wide.

Appendix D

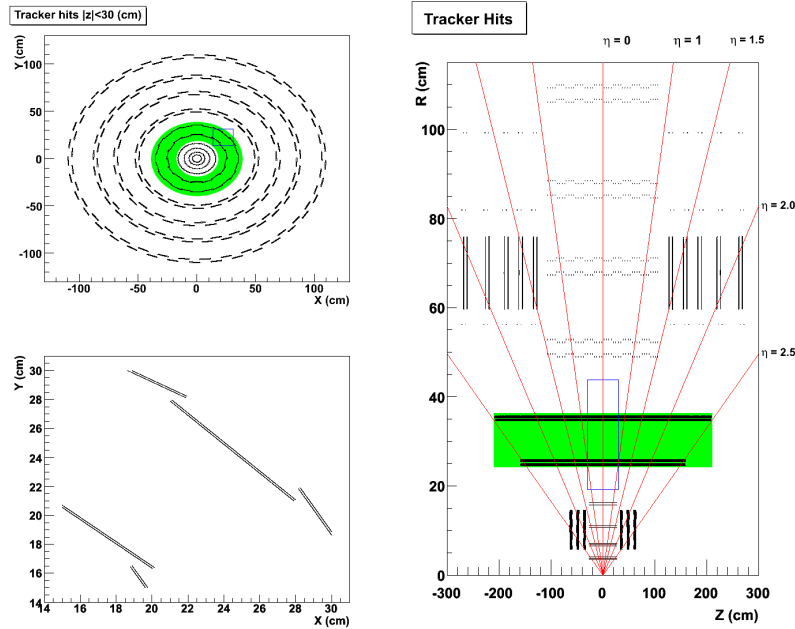
Tracker upgrade geometries



(a) Strawman-B Geometry



(b) Long-Barrel Geometry



(c) Hybrid Geometry

Figure D.1: Distribution of simulated hits showing the structure of the Strawman-B, Long-Barrel and Hybrid tracker upgrade geometries. The areas highlighted in green indicate the regions in which only stacked layers are present. Within each subfigure, the bottom-left plot shows the X-Y view in the region of interest, indicated by a box, in the top-left and righthand plots.

Glossary

μ TCA	Micro-Telecommunications Computing Architecture, 65
ADC	Analogue to Digital Converter, 246
ALICE	An LHC Ion Collider Experiment/A Large Ion Collider Experiment, 40
AMC	Advanced Mezzanine Card, 149
APD	Avalanche PhotoDiodes, 48
APV25	Analogue Pipeline Voltage, 0.25 μ m, 47
ART	Algebraic Reconstruction Technique, 243
ASICs	Application Specific Integrated Circuits, 50
ATLAS	A Toroidal LHC ApparatuS, 40
BCo	RCT Bunch Crossing Zero, 90
BER	Bit Error Rate, 94
BX	Bunch Crossing, 40
BXo	Bunch Crossing Zero, 62
CKM	Cabibbo-Kobayashi-Maskawa, 28
CMS	Compact Muon Solenoid, 40
CMSSW	CMS Software, 140

CRC16	16-bit Cyclic Redundancy Check, 91
CSCs	Cathode Strip Chambers, 49
DAQ	Data Acquisition, 63
DCC	Data concentrator card, 63
DCM	Digital Clock Manager, 110
DDR	Double Data Rate, 68
digis	digitized hits, 162
DOCA	distance of closest approach, 218
DTs	Drift Tubes, 49
dual-PMC	dual PCI Mezzanine Card, 71
ECAL	Electromagnetic Calorimeter, 48
ECL	Emitter-Coupled Logic, 68
EEPROM	Electrically Erasable Programmable Read Only Memory, 87
FETS	Front End Test Stand, 242
FIFO	First-In-First-Out, 86
FPGAs	Field Programmable Gate Arrays, 50
FSM	Finite State Machine, 91
FWHM	Full Width at Half Maximum, 95
GCC	Gnu Compiler Collection, 131
GCT	Global Calorimeter Trigger, 64
GMT	Global Muon Trigger, 65
GT	Global Trigger, 65
GTi	Global Trigger interface, 64

GUI	Graphical User Interface, 137
HAL	Hardware Access Library, 131
HCAL	Hadronic Calorimeter, 49
HLT	Higher Level Trigger, 50
HPD	Hybrid PhotoDiode, 49
HTR	HCAL Trigger and Readout, 63
I ² C	Inter-Integrated Circuit, 110
I ² O	Intelligent Input/Output, 53
I/O	Input/Output resources, 72
IGUANA	Interactive Graphics for User ANALysis, 183
JTAG	Joint Test Action Group — IEEE 1149.1, 87
LCD	Liquid Crystal Display, 128
LEBT	Low Energy Beam Transport, 242
LEDs	Light Emitting Diodes, 128
LEP	Large Electron-Positron Collider, 27
LFSR	Linear Feedback Shift Register, 91
LHC	Large Hadron Collider, 40
LP SPL	Low Power Superconducting Proton Linac, 241
LVDS	Low-voltage differential signalling, 83
LVTTL	Low Voltage Transistor-Transistor Logic, 84
LYSO	Cerium-doped Lutetium Yttrium Orthosili- cate, 243
MEBT	Medium Energy Beam Transport, 242

MGTs	Multi-Gigabit Transceivers, 83
MIP	Minimum Ionizing Particle, 62
NVP	Nominal Velocity of Propagation, 97
OMA	Optical Modulation Amplitude, 104
PLD	Programmable Logic Device, 88
PLL	Phase-Locked Loop, 89
PMC	PCI Mezzanine Card, 72
PMNS	Pontecorvo-Maki-Nakagawa-Sakata, 28
PRBS	Pseudo-Random Bit Stream, 81
PS	Proton Synchrotron, 241
PSB	Pipeline Synchronizing Buffer, 66
PSB	Proton Synchrotron Booster, 241
QCD	Quantum ChromoDynamics, 38
QPLL	Quartz Phase-Locked Loop, 83
RAL	Rutherford-Appleton Laboratory, 242
RCT	Regional Calorimeter Trigger, 63
RFQ	Radio Frequency Quadrupole, 242
RPCs	Resistive Plate Chambers, 49
SBC	Single-Board Computer, 125
SCR	Silicon-controlled rectifier, 85
SerDes	Serializer/Deserializer, 69
SEUs	Single Event Upsets, 43

SFP	Small form-factor, Pluggable, 81
SLB	Synchronization and Link Board, 62
SLC ₄	CERN Scientific Linux, release 4, 131
SOAP	Simple Object Access Protocol, 53
SPS	Super Proton Synchrotron, 241
SuSy	SuperSymmetric, 36
TCC	Trigger concentrator card, 62
TDR	Time Domain Reflectometry, 95
TIB	Tracker Inner Barrel, 183
TOB	Tracker Outer Barrel, 183
Totem	Total Cross Section, Elastic Scattering and Diffraction Dissociation Measurement, 40
TPGs	Trigger-Primitive Generators, 62
TTC	Trigger Timing and Control, 66
USB	Universal Serial Bus, 72
USC	Underground Sorting Cavern, 52
VHDCI	Very High Density Cable Interconnect, 69
VHDL	VHSIC Hardware Description Language, 124
VHSIC	Very High Speed Integrated Circuit, 124
VME	Versa Module Eurocard, 65
VPTs	Vacuum PhotoTriodes, 48
XDAQ	Distributed Data AcQuisition, 53
XML	Extensible Markup Language, 136

References

- [1] G.E.R. LLOYD, **Aristotle: The Growth and Structure of his Thought.** Cambridge University Press, 1968.
- [2] R. GORINI, **“Al-Haytham the Man of Experience: First Steps in the Science of Vision”.**
- [3] S. L. GLASHOW, **“Partial-symmetries of weak interactions”** *Nuclear Physics* **22** (February, 1961) 579–588.
- [4] S. WEINBERG , **“A Model of Leptons”** *Phys. Rev. Lett.* **19** (November, 1967) 1264–1266.
- [5] A. SALAM , **“Elementary Particle Theory: Relativistic Groups and Analyticity”** *Nobel Symposium No. 8* (May, 1968).
- [6] F. ENGLERT, R. BROUT, **“Broken Symmetry and the Mass of Gauge Vector Mesons”** *Phys. Rev. Lett.* **13** (August, 1964) 321–323.
- [7] P.W. HIGGS, **“Broken Symmetries and the Masses of Gauge Bosons”** *Phys. Rev. Lett.* **13** (October, 1964) 508–509.
- [8] G.S. GURALNIK, C.R. HAGEN, T.W.B.KIBBLE, **“Global Conservation Laws and Massless Particles”** *Phys. Rev. Lett.* **13** (November, 1964) 585–587.
- [9] G. 'T HOOFT, M.J.G. VELTMAN, **“Regularization and Renormalization of Gauge Fields”** *Nucl. Phys.* **B44** (1972) 189–213.

-
- [10] D. KARLEN FOR THE LEP ELECTROWEAK WORKING GROUP, **“Experimental Status of the Standard Model”**. 29th International Conference on High Energy Physics, Vancouver, Canada.
- [11] D. GREEN, **“Lectures in particle physics”**. Volume 55 of World Scientific lecture notes in physics.
- [12] C. AMSLER ET AL.(PARTICLE DATA GROUP), **“Review of Particle Physics”** *Physics Letters B* **667** (September, 2008).
- [13] R. KINNUNEN, **“Higgs Physics at LHC”**. CMS Conference Report 2002/020.
- [14] K. LASSILA-PERINI, **“Higgs Physics at LHC”**. CMS Conference Report 2002/018.
- [15] G. WROCHNA, **“Physics at LHC”**. CMS Conference Report 2002/015.
- [16] M. DITTMAR, **“Searching for the Higgs and other Exotic Objects”**. CMS Conference Report 1999/009.
- [17] C. E. WULZ, **“CMS Physics Overview”**. CMS Conference Report 2001/016.
- [18] J. W. ROHLE, **“Physics Reach with CMS at High and Super-High Luminosities”**. CMS Conference Report 2003/027.
- [19] THE LEP COLLABORATIONS; ALEPH, DELPHI, L3, OPAL, AND THE LEP ELECTROWEAK WORKING GROUP, **“Precision Electroweak Measurements and Constraints on the Standard Model”**. CERN-PH-EP/2007-039.
- [20] TEVATRON ELECTROWEAK WORKING GROUP AND FOR THE CDF COLLABORATION AND THE D0 COLLABORATION, **“Combination of CDF and D0 Results on the Mass of the Top Quark”**.
- [21] D. WOOD, **“Electroweak Physics”**. XXIII International Conference on High Energy Physics.
-

-
- [22] THE LEP COLLABORATIONS; ALEPH, DELPHI, L₃ AND OPAL, **“Search for the Standard Model Higgs Boson at LEP”**. International Conference on High Energy Physics 2002.
- [23] THE TEVNPH WORKING GROUP FOR THE CDF AND DØ COLLABORATIONS, **“Combined CDF and DØ Upper Limits on Standard Model Higgs Boson Production at High Mass (155 - 200 GeV) with 3 fb⁻¹ of data”**. FERMILAB-PUB-08-270-E, CDF Note 9465, DØ Note 5754.
- [24] P. SPHICAS, **“Design Principles and Performance of CMS”**. CERN Academic Training.
- [25] C. QUIGG, **“The Standard Model (Electroweak Theory)”**. European School of High-Energy Physics.
- [26] G. KANE, S. WATSON, **“Dark matter and LHC: What is the connection?”** *Modern Physics Letters A (MPLA)* **23** (2008) 2103–2123.
- [27] P. J. E. PEEBLES, B. RATRA, **“The cosmological constant and dark energy”** *Rev. Mod. Phys.* **75** (April, 2003) 559–606.
- [28] **“The Large Hadron Collider homepage”**
<http://lhc.web.cern.ch/lhc/>.
- [29] D. ACOSTA ET AL., **“CMS Physics TDR: Volume I (PTDR₁), Detector Performace and Software”**. CERN/LHCC/2006/001.
- [30] D. BOURILKOV, **“Physics with the CMS Experiment in the First Year of LHC”**. APS Meeting, Tampa, FL, USA.
- [31] I. MOCIOIU, Y. NARA, I. SARCEVIC, **“Hadrons as signature of black hole production at the LHC”** *Physics Letters B* **557-1** (March, 2003) 87–93.
- [32] **“The Alice Portal”**
<http://aliceinfo.cern.ch/index.html>.
- [33] G. FLÜGGE, **“Future Research in High Energy Physics”**. European School of High-Energy Physics.
-

-
- [34] D. TREILLE, **"The Experimental Challenges of LHC"** *Journal of Physics: Conference Series* **53** (2006) 117–162. Corfu Summer Institute on Elementary Particle Physics.
- [35] F. MEIJERS, **"The CMS Event Builder"** March, 2003. 2003 Conference for Computing in High Energy and Nuclear Physics (CHEP03), www.slac.stanford.edu/econf/C0303241/proc/pres/10008.PDF.
- [36] D. BORTOLETTO ET AL., **"Sensor Development for the CMS Pixel Detector"** *Nucl. Instr. Methods A.* **485** (2002) 89–99.
- [37] J. R. FULCHER, **"Single Event Upset Studies on the APV25 Front End Readout Chip"**. Proceedings of the 6th Workshop on Electronics for LHC Experiments.
- [38] P. BRINKLEY, C. CARMICHAEL, **"SEU Mitigation Design Techniques for the XQR4000XL"**.
- [39] I. DAWSON, A. MORAES, C. BUTTAR, V. CINDRO, I. MANDIC, **"Radioactivation of silicon tracker modules in high-luminosity hadron collider radiation environments"** *Nucl. Instrum. Methods Phys. Res., A* **515** (2003) 422–438.
- [40] CMS COLLABORATION, **"The Electromagnetic Calorimeter Project - Technical Design Report"**. CERN/LHCC/1997/033.
- [41] CMS COLLABORATION, **"The CMS Technical Proposal"** CERN/LHCC/1994/038.
- [42] CMS COLLABORATION, **"CMS Tracker - Technical Design Report"** CERN/LHCC/1998/006.
- [43] CMS COLLABORATION, **"Addendum to the CMS Tracker TDR"** CERN/LHCC/2000/016.
- [44] A. TRICOMI, **"Performance of ATLAS & CMS Silicon Tracker"**. International Europhysics Conference on High Energy Physics.
-

-
- [45] D. KOTLINSKI, R. BAUR, K. GABATHULER, R. HORISBERGER, R. SCHNYDER, W. ERDMANN, **“Readout of the CMS Pixel Detector”**. Proceedings of the 6th Workshop on Electronics for LHC Experiments.
- [46] CMS COLLABORATION, **“The CMS Physics Technical Design Report, Volume I, Detector Performance and Software”**. CERN/LHCC 2006-001.
- [47] M. FRENCH ET AL., **“Design and Results from the APV25, a Deep Submicron CMS Front-End Chip for the CMS Tracker”** *Nucl. Instr. Methods A* **466** (2001) 359–365.
- [48] K. KLEIN, **“The CMS Silicon Strip Tracker - Overview and Status”**. International Europhysics Conference on High Energy Physics, Lisbon, Portugal.
- [49] N. DEMARIA, **“Status of CMS Silicon Strip Tracker”**. 8th Topical Seminar on Innovative Particle and Radiation Detectors, Siena, Italy.
- [50] PAOLO AZZURRI, **“The CMS Silicon Strip Tracker”** *Journal of Physics: Conference Series* **41** (2006) 127–134. EPS Euroconference XIX Nuclear Physics Divisional Conference.
- [51] CMS COLLABORATION, **“The Hadron Calorimeter Project - Technical Design Report”**. CERN/LHCC/1997/031.
- [52] P. GIACOMELLI, **“The CMS Muon Detector”** *Nucl. Instr. Methods A* **478** (2002) 147–152.
- [53] CMS COLLABORATION, **“The Muon Project - Technical Design Report”**. CERN/LHCC/1997/032.
- [54] CMS COLLABORATION, **“The TriDAS Project Technical Design Report, Volume 1”** CERN/LHCC/2000/038.
- [55] C. SEEZ, **“The CMS Trigger System”**. CMS Conference Report 2003/008.
-

-
- [56] CMS COLLABORATION, **“The TriDAS Project Technical Design Report, Volume 2: Data Acquisition and High-Level Trigger”**
CERN/LHCC/2002/026.
- [57] J. VARELA, **“CMS L1 Trigger Control System”**. CMS Note 2002/033.
- [58] **“The Analog Optohybrid Homepage”**
<http://wwwhephy.oeaw.ac.at/u3w/f/friedl/www/aoh/>.
- [59] F. VASEY, **“CMS Tracker Optical Readout Link Specification”**
[http://cmstrackercontrol.web.cern.ch/cmstrackercontrol/
documents/FVasey/OpticalLinkSpecs1.1.pdf](http://cmstrackercontrol.web.cern.ch/cmstrackercontrol/documents/FVasey/OpticalLinkSpecs1.1.pdf).
- [60] J. COUGHLAN ET AL., **“The Front-End Driver card for the CMS Silicon Strip Tracker Readout”**. Proceedings of the 8th Workshop on Electronics for LHC Experiments, CERN/LHCC/2002/034.
- [61] J. COUGHLAN ET AL., **“The CMS Tracker Front-End Driver”**. Proceedings of the 9th Workshop on Electronics for LHC Experiments, CERN/LHCC/2003/055.
- [62] L.ORSINI AND J. GUTLEBER, **“The XDAQ framework”**
<http://xdaq.web.cern.ch/xdaq/>.
- [63] S. DASU, J. LACKEY, W. SMITH, W. TEMPLE, **“CMS Level 1 Calorimeter Trigger Performance on Technical Proposal Physics”** CMS Technical Note 1995/183.
- [64] M. HANSEN, **“CMS ECAL FENIX ASIC design methodology”**.
- [65] C. FOUDAS, M. HANSEN, G. ILES, J. JONES, A. ROSE, M. STETTLER, **“Proposal for an alternative design of the Global Calorimeter Trigger, Version II”**.
- [66] A. ROSE, **“Proposal for CMS Global Calorimeter Trigger naming and numbering scheme. Version 2.2”** May, 2007.
www.hep.ph.ic.ac.uk/cms/gct/Documents/GCT_refdoc_v2_2.pdf.
-

-
- [67] M. STETTLER ET AL., **“The CMS Global Calorimeter Trigger Hardware Design”**. Proceedings of the 12th Workshop on Electronics for LHC Experiments.
- [68] **“CERN S-LINK Homepage”**
<http://hsi.web.cern.ch/HSI/s-link/>.
- [69] H. VAN DER BIJ ET AL., **“The S-LINK Interface Specification”**
<http://hsi.web.cern.ch/HSI/s-link/spec/spec/s-link.pdf>.
- [70] H. VAN DER BIJ ET AL., **“The S-LINK 64 bit extension specification: S-LINK64”**
<http://edms.cern.ch/file/249683/2/slink64%20v20.pdf>.
- [71] **“Mindspeed Technologies, Inc.”**
www.mindspeed.com/.
- [72] AVAGOTECH, **“HFBR-5720AL/5720ALP Optical Transceiver Data Sheet”**
www.avagotech.com/pc/downloadDocument.do?id=4568.
- [73] TEXAS INSTRUMENTS, **“TLK2501 1.5 to 2.5 GBPS Transceiver”**
www.ti.com.
- [74] XILINX, **“Spartan-3 FPGA Family: Complete Data Sheet”**
www.xilinx.com.
- [75] TEXAS INSTRUMENTS, **“Programmable Low-Voltage 1:10 LVDS Clock Driver”** 2003.
www.ti.com.
- [76] LINEAR TECHNOLOGY, **“LT1963 Series 1.5A Low Noise, Fast Transient Response LDO Regulators”**
www.linear.com.
- [77] TEXAS INSTRUMENTS, **“PTH05050W Datasheet”**
www.ti.com.
-

-
- [78] CYPRESS SEMICONDUCTOR, **“CY7C68001 EZ-USB SX2 High-Speed USB Interface Device”**
www.cypress.com.
- [79] **“Exception PCB”**
www.exceptionpcb.com.
- [80] **“Exception EMS”**
<http://exceptionems.com>.
- [81] **“Madison Cable Corporation (TYCO electronics)”**
<https://madisoncable.tycoelectronics.com/contact.asp>.
- [82] AGILENT TECHNOLOGIES, INC., **“Time Domain Reflectometry Theory”**
May, 2006. Agilent Application Note 1304-2.
- [83] A. HOLZNER, **“TTCci User Guide”**
<http://cmsdoc.cern.ch/cms/TRIDAS/ttc/modules/ttcci/index.html>.
- [84] P. GÄLLNÖ, **“TTC-VMEbus INTERFACE, TTCvi-MkII”**
www.cern.ch/TTC/TTCviSpec.pdf.
- [85] **“SBS Technologies”**
www.sbs.com.
- [86] B. TAYLOR, **“TTC laser transmitter (TTCex, TTCtx, TTCmx) User Manual”**.
- [87] **“Mersenne Twister Home Page”**
www.math.sci.hiroshima-u.ac.jp/~m-mat/MT/emt.html.
- [88] MICROSOFT CORPORATION. 1985-2009.
<http://office.microsoft.com/en-us/excel/FX100487621033.aspx>.
- [89] F. SZONCSO, **“Application Restrictions for VME Crates manufactured by W-IE-NE-R (Plein & Baus Ges.m.b.H.)”** June, 2003. CERN Technical Note: CERN-TIS-2003-002-GS-TN,
http://ph-dep-ese.web.cern.ch/ph-dep-ese/crates/documents/VME_TueV_Internal_report_030603.pdf.
-

-
- [90] **"GCC, the GNU Compiler Collection"**
<http://gcc.gnu.org/>.
- [91] C. SCHWICK, **"The Hardware Access Libraries"**
<http://cmsdoc.cern.ch/~cschwick/software/documentation/HAL/index.html>.
- [92] **"libusb project home"**
<http://libusb.sourceforge.net/>.
- [93] I. M. DE ABRIL, C.-E. WULZ, J. VARELA, **"Concept of the CMS Trigger Supervisor"**. CMS NOTE 2005/011.
- [94] I. M. DE ABRIL, **"The CMS Trigger Supervisor: Control and Hardware Monitoring System of the CMS Level-1 Trigger at CERN"**. PhD Thesis.
- [95] T. BRAY, J. PAOLI, C. M. SPERBERG-McQUEEN, E. MALER, F. YERGEAU, **"Extensible Markup Language (XML) 1.0 (Fifth Edition)"** November, 2008.
www.w3.org/TR/xml/.
- [96] ORACLE CORPORATION, **"Oracle Database 11g Product Family"** January, 2009. Oracle White Paper,
www.oracle.com/technology/products/database/oracle11g/pdf/database-11g-product-family-technical-whitepaper.pdf.
- [97] C. ULLMAN, L. DYKES, **Beginning Ajax**. Wiley (J. Wiley and Son), March, 2007.
- [98] **"ROOT : A Data Analysis Framework"**
<http://root.cern.ch/>.
- [99] J.-L. GAILLY, M. ADLER, **"The GNU Zip Utility"**
www.gzip.org/.
- [100] C. ROVELLI, **"The detailed simulation of the CMS detector"** October, 2007. 10th ICATPP Conference, Como, Italy.
-

-
- [101] F. GIANOTTI, M.L. MANGANO, T. VIRDEE ET AL., **“Physics Potential and Experimental Challenges of the LHC Luminosity Upgrade”**.
CERN-TH/2002-078 hep-ph/0204087.
- [102] OLIVER BRUENING, **“Accelerator Upgrades - Talk at 1st Workshop for upgrades to CMS at SLHC”**.
- [103] H-C. KÄSTLI , **“CMS Upgrade Requirements and Upgrade Plans”**.
CHIPP Workshop on Detector R&D.
- [104] XILINX, **“Virtex-6 Family Overview”** June, 2009.
www.xilinx.com/support/documentation/data_sheets/ds150.pdf.
- [105] S. DASU, **“SLHC Level-1 Calorimeter Trigger & Track Correlation”** July, 2004. Second CMS Workshop on Detectors and Electronics for SLHC,
<http://indico.cern.ch/getFile.py/access?contribId=s1t2&resId=1&materialId=0&confId=a043080>.
- [106] G. DASKALAKIS, K. LASSILA-PERINI, **“Jet rejection using the pixel matching for the low and the high luminosity”** CMS-NOTE-2002-039.
- [107] K. LASSILA-PERINI, **“Jet rejection with matching ECAL clusters to pixel hits”** May, 2001. CMS NOTE 2001/021.
- [108] A. ROSE ET AL., **“A Tracking Trigger for CMS at SLHC”**. Proceedings of the 11th Workshop on Electronics for LHC Experiments.
- [109] P. MOREIRA, A. MARCHIORO, K. KLOUKINAS, **“The GBT: A proposed architecure for multi-Gb/s data transmission in high energy physics”**.
Topical Workshop on Electronics for Particle Physics, Prague, Czech Republic.
- [110] R. W. HAMMING, **“Error Detecting and Error Correcting Codes”** *The Bell System Technical Journal* **26** (April, 1950).
- [111] T. SPEER, **“A Gaussian-Sum Filter for Vertex Reconstruction”**. CMS
CR-2004/052.
-

-
- [112] D. UMBACH, K.N. JONES, **“A few methods for fitting circles to data”**
November, 2003.
- [113] J. ALLISON ET AL., **“GEANT 4 - a simulation toolkit”** July, 2003.
- [114] T. SJÖSTRAND, **“PYTHIA (and JETSET) Webpage”**
www.thep.lu.se/~torbjorn/Pythia.html.
- [115] P. JANOT, F. BEAUDETTE, **“Simulating and reconstructing events with fast simulation”**
<https://twiki.cern.ch/twiki/bin/view/CMS/WorkBookFastSimulation>.
- [116] FILIPPO AMBROGLINI, **“Full CMS simulations for SLHC studies”** March, 2007.
<http://indico.cern.ch/getFile.py/access?contribId=7&resId=0&materialId=slides&confId=13095>.
- [117] IANNA OSBORNE ET AL., **“Interactive Graphics for User ANALysis”**
<http://iguana.web.cern.ch/iguana/>.
- [118] H. CHEUNG, M. PESARESI, **“Strawman-B Tracker Upgrade Geometry”**
<https://twiki.cern.ch/twiki/bin/view/CMS/ExampleStrawmanB>.
- [119] E. BROWNSON, H. CHEUNG, I. REID, M. WEINBERGER, **“Long Barrel Tracker Upgrade Geometry”**
<https://twiki.cern.ch/twiki/bin/view/CMS/ExampleLongBarrel>.
- [120] E. BROWNSON, H. CHEUNG, C. CIVININI, M. PESARESI, I. REID, **“Hybrid Tracker Upgrade Geometry”**
<https://twiki.cern.ch/twiki/bin/view/CMS/ExampleHybridGeometry>.
- [121] R.O. DUDA, P. E. HART, **“Use of the Hough Transformation to Detect Lines and Curves in Pictures”** *Communications of the Association for Computing Machinery* **15** (January, 1972) 11–15.
- [122] **“The MINOS Experiment and NuMI Beamline”**
<http://www-numi.fnal.gov/>.
-

-
- [123] M. PESARESI, **“Development of a New Silicon Tracker for CMS at Super-LHC”**.
- [124] M. M. ANGARANO, **“The silicon strip tracker for CMS”**.
- [125] S. DASU, **“Challenges of Trigger Systems for LHC and SLHC”**.
- [126] L. J. FIELDS, **“Electron Trigger Studies using Stack TPGs - I”** July, 2009.
CMS Trigger Upgrade Workshop,
<http://indico.cern.ch/getFile.py/access?contribId=2&sessionId=3&resId=0&materialId=paper&confId=55141>.
- [127] J.R.R. TOLKIEN, **“The Lord of the Rings”**
- [128] S.Y. LEE, **Accelerator physics**. World Scientific, 2 ed., 2004.
- [129] D. A. LEE, J. K. POZIMSKI, C. GABOR, P.SAVAGE, **“A Laserwire Beam Profile Measuring Device for the RAL Front End Test Stand”**. Diagnostics and Instrumentation for Particle Accelerators Conference, Venice, Italy.
- [130] D. RAPARIA. ET AL., **“The ‘Algebraic Reconstruction Technique’ (ART)”**. Particle Accelerator Conference, Vancouver, Canada.
- [131] TEXAS INSTRUMENTS, **“Dual Current Input 20-Bit ANALOG-TO-DIGITAL CONVERTER”**
<http://focus.ti.com/lit/ds/symlink/ddc112.pdf>.
- [132] ATMEL, **“8-bit AVR Microcontroller with 128K Bytes In-System Programmable Flash : ATmega128, ATmega128L”**
www.atmel.com/dyn/resources/prod_documents/doc2467.pdf.
- [133] TEXAS INSTRUMENTS, **“CMOS VOLTAGE REFERENCE : 50ppm/°C Max, 50 μ A in SOT23-3”**
<http://focus.ti.com/lit/ds/symlink/ref3012.pdf>.
-

This document was produced using

L^AT_EX 2_ε