Particle Flow at CMS & the ILC

JAMES A C BALLIN

Enrolled at:
High Energy Physics, Blackett Laboratory
Imperial College, London, SW7 2AZ, United Kingdom

And a member of:
European Organization for Nuclear Research (CERN)
Geneva, CH-1211, Switzerland

CERN-THESIS-2010-086
06/05/2010

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

March 2010

Abstract

This thesis describes hadron reconstruction at the Compact Muon Solenoid (CMS) experiment at the Large Hadron Collider (LHC) at CERN, Geneva. The focus is on the *particle flow* reconstruction of these objects. This thesis revisits the subject of the CMS calorimeters' non-linear response to hadrons. Data from testbeam experiments conducted in 2006 & 2007 is compared with simulations and substantial differences are found. A particle flow calibration to correct the energy response of the testbeam data is evaluated. The reconstructed jet response is found to change by ~ ±5% when a data-driven calibration is used in place of the calibration derived from simulation. Collision data taken at the early stage of CMS' commissioning is also presented. The hadron response in data is determined to be compatible with testbeam results presented in this thesis.

This thesis also details the use of neural networks to improve the energy measurement of hadrons at CMS. The networks are implemented in a functional and concurrent language (Erlang). The advantages to using a concurrent language to solve problems which can be parallelized are demonstrated.

The Monolithic Active Pixel Sensor (MAPS) is described. This device, with 50 × 50μm^2 pitch and binary readout, is designed for sampling calorimeters at the proposed International Linear Collider (ILC). Such calorimeters are optimized for a particle flow reconstruction. The MAPS benefits from a novel industrial process whereby a deep p-well implant separates the epitaxial silicon layer from parasitic pixel electronics, to increase the charge collection efficiency. Data acquired from testbeam in 2007 is analysed. The low efficiency of the prototype is attributable to operating the sensor at an incorrect working point; this has subsequently been addressed.

Keywords: hadrons, particle flow, neural networks, Erlang, pixel sensor

To my parents

Preface & Contributions by the Author

> Men nearly always follow the tracks made by others and proceed in their affairs by imitation, even though they cannot entirely keep to the tracks of others or emulate the prowess of their models. So a prudent man should always follow in the footsteps of great men and imitate those who have been outstanding. If his own prowess fails to compare with theirs, at least it has an air of greatness about it.
>
> *Niccolò Machiavelli — The Prince [1]*

This thesis is presented in four parts. The common theme is particle flow. The presentation is slightly unusual, inasmuch as you won't find a complete description of the standard model or extensive theoretical musings. The focus is on the characterization of 'software instrumentation', the use of real data to provide insights, and an exposition of cutting edge sensor technology.

The real 'physics' is given in Part II, where I present the outcome of my efforts to better understand CMS' response to hadrons with the application of the particle flow technique. I am extremely grateful for the advice and guidance of my supervisor, Dr David Colling, and my unofficial supervisor, Dr Michel Della Negra, in conducting this research.

I originally joined the CALICE project (Calorimetry for a Linear Collider Experiment), working under the supervision of Prof. Paul Dauncey on the MAPS for a year. This project was regrettably terminated by STFC in December 2007. The MAPS is a novel pixel sensor designed for particle flow, and I hope that you find its description in Part III stimulating. This section of the thesis also presents work 'off the beaten track' looking at how neural networks may—or may not—improve hadron energy reconstruction at CMS. I've particularly enjoyed working on projects with a distinctly 'R&D' feel, and while Chapter 6 may not strictly complement earlier material it has formed an important part of my development as a researcher in the last 3 years. I consider it essential in this thesis.

Consequently, this thesis presents introductory material and the relevant theory in Part I, followed by the established (CMS) in Part II. The focus shifts to detector research and development in Part III. A short and, with hope, coherent summary is provided in Part IV.

Objectives of this thesis

PART I: PROLOGUE

- **Chapter 1:** Describe why the LHC exists, and its main features. Deliver a concise overview of the CMS experiment and the particle flow technique.
- **Chapter 2:** Present the theory necessary to understand the CMS calorimeters' response to hadrons, and discuss the implications for designing a calorimeter at the ILC.

PART II: PARTICLE FLOW AT CMS

- **Chapter 3:** Discuss CMS' response to hadrons in testbeam, in preparation for applying the particle flow reconstruction. Extensive comparisons with simulations will be made.
- **Chapter 4:** Understand the particle flow reconstruction of hadrons and the algorithm's performance when applied to testbeam data.
- **Chapter 5:** Consider particle flow's response to hadrons from data acquired in early collisions at CMS in December 2009.

PART III: FUTURE TECHNOLOGY IN CALORIMETRY

- **Chapter 6:** Discuss a *multivariate analysis* (MVA) technique for improving CMS' reconstruction of hadrons in the hadronic calorimeter. Consider the benefits of using a concurrent, functional language to implement a neural network.
- **Chapter 7:** Describe a novel sensor designed for calorimetry at a future linear collider and present results from testbeam.

PART IV: EPILOGUE

- **Chapter 8:** Provide a coherent summary of the material discussed in this thesis.

Contributions by the Author

CERN is a large, cross-border organization and, at the last count (June 2008), CMS boasted an author list of some 3,600 people. Clearly the work presented in this thesis is but a wafer-thin contribution to this massive undertaking, but the Reader can expect the following to represent original work by the Author:

Chapter 3 details an independent analysis of two sets of CMS testbeam data taken in 2006 and 2007, providing a complete and rigorous cross-check of many important results regarding the performance of the CMS calorimeters. The comparison of testbeam data with simulations is the original work of the Author.

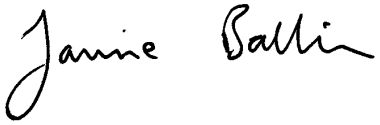
Chapter 4 exhibits an original analysis of the testbeam datasets using particle flow. Note that particle flow is developed by the *Particle Flow Physics Object Group* at CERN.

Chapter 5 presents a first look at the data acquired by CMS in late 2009. The reconstruction was performed by the Author's expert colleagues in the Particle Flow Group, but the analysis is original.

Chapter 6: while neural networks are well understood, the implementation described is completely the work of the Author. The application to hadron reconstruction is—to the Author's knowledge—original and novel.

Chapter 7: the Author contributed to operating the testbeam experiment, and the efficiency analysis was conducted under the supervision of Prof. Paul Dauncey.

Work attributable to other authors has been referenced where possible.



Jamie Ballin
Imperial College London, 2010

Thank you!

When it became clear that I couldn't continue work on the MAPS project, one year into my PhD, I was amazed by how easily I was able to join in with Imperial's CMS team. I am grateful for the support of the department in helping me make the transition, and for the opportunity to read for a PhD with the support of an STFC studentship.

I have had several masters during my time at Imperial. It was a great pleasure to work with Paul on the MAPS, but if anyone else is reading this and wonders how to slow Paul's mind to speeds comprehensible by mere mortals, I can advise that waking him from sleep before 6am to discuss problems with data acquisition at a CERN testbeam may induce the desired effect.

It has been an honour to work under the supervision of Michel at CERN: this is a superb institution, and I shall be forever proud to have been associated with it. Also, thank you to David for his questions, support at times of despair (!) and counsel.

Three people in particular deserve a special mention, for they have taught me more than any books have: to Matt, Anne-Marie and Andy, I thank you for your wisdom, company, and friendship.

And finally, I am grateful for the love of my family, and the sacrifices my parents have made to give me this education. A special salute also to Martin, my long-suffering flatmate and friend, for enduring my musings and rants on particle physics.

Contents

Abstract 2

Preface & Contributions by the Author 4

Acknowledgements 7

Glossary 16

I Prologue 18

1 Introduction to CMS & LHC 19
- 1.1 Overview . . . 19
- 1.2 The Large Hadron Collider (LHC) . . . 19
  - 1.2.1 Why build a hadron collider? . . . 19
  - 1.2.2 The machine . . . 22
- 1.3 The CMS Experiment . . . 24
  - 1.3.1 Physics goals . . . 24
  - 1.3.2 Design & implementation . . . 25
  - 1.3.3 A silicon tracker . . . 29
  - 1.3.4 Electromagnetic calorimetry: the ECAL . . . 29
  - 1.3.5 Hadronic calorimetry: the HCAL . . . 32
  - 1.3.6 Muon systems . . . 33
  - 1.3.7 Trigger & computing . . . 34
- 1.4 Why study hadrons? . . . 35
- 1.5 Particle flow reconstruction at CMS . . . 35
- 1.6 Summary . . . 37

2 Aspects of calorimetry in HEP 38
- 2.1 Introduction . . . 38
- 2.2 Calorimetry . . . 38
  - 2.2.1 Practicalities . . . 39
  - 2.2.2 The interaction of particles with detection media . . . 39
  - 2.2.3 Electromagnetic showers . . . 43
  - 2.2.4 Hadronic cascades . . . 47
  - 2.2.5 Compensating calorimeters . . . 49

2.2.6 Sampling and homogeneous calorimeters . . . 52
2.2.7 Energy resolution . . . 52
2.3 Implications for the CMS calorimeters . . . 55
2.3.1 Specific design implications . . . 55
2.3.2 The HF design . . . 56
2.3.3 Differences in response between pions, protons etc. . . . 56
2.4 Implications for ILC calorimetry . . . 57
2.4.1 The CALICE collaboration . . . 58
2.5 Summary . . . 59

II Particle flow at CMS **60**

3 The CMS calorimeters in testbeam **61**
3.1 Introduction . . . 61
3.2 Description of the testbeam . . . 61
3.2.1 Beam cleaning & pion selection . . . 63
3.3 The simulations . . . 64
3.3.1 Full simulation . . . 64
3.3.2 Fast simulation . . . 64
3.3.3 Reconstruction . . . 65
3.4 Differences between CMS, testbeam, and simulations . . . 65
3.5 Calorimeter energy response . . . 66
3.5.1 Calo rechits . . . 66
3.5.2 HCAL noise . . . 67
3.5.3 ECAL Activity . . . 67
3.5.4 Standard response . . . 68
3.5.5 Comparison with simulations . . . 70
3.6 Summary . . . 76

4 The particle flow reconstruction of hadrons **77**
4.1 Introduction . . . 77
4.2 Elements of the particle flow reconstruction . . . 77
4.2.1 Tracking . . . 77
4.2.2 Clustering . . . 79
4.2.3 The PFBlock: geometrical linking of the elements . . . 81
4.2.4 PFAlgo: analyzing the PFBlocks . . . 83
4.2.5 Excess of HCAL energy: a closer look . . . 86
4.3 Preparing to apply particle flow to testbeam . . . 87
4.3.1 Modifications required for testbeam . . . 87
4.4 Evaluating a particle flow calibration for hadrons . . . 88
4.4.1 Testbeam calibration coefficients . . . 89
4.5 Particle flow with testbeam . . . 91
4.5.1 Event displays . . . 91
4.5.2 Response to charged hadrons . . . 92
4.5.3 Hadronic energy resolution (calorimetry) . . . 97
4.5.4 Particle flow cluster multiplicity . . . 98
4.6 Implications for jet reconstruction . . . 100
4.7 Summary . . . 105

5 Particle flow hadron response in first collisions **107**
5.1 Introduction . . . 107
5.2 Early collision data . . . 107
5.2.1 Minimum bias events . . . 107
5.2.2 A reminder of differences with the testbeam environment . . . 108
5.2.3 Event selections . . . 108
5.3 Energy response of charged hadrons . . . 111
5.3.1 Global energy response . . . 111
5.3.2 Detailed energy response . . . 111
5.4 Strategies for improving PF hadron reconstruction . . . 114
5.4.1 PF hadrons without HCAL clusters . . . 114
5.4.2 Surplus PF neutral hadrons and PF photons . . . 119
5.5 Summary . . . 120

III Future technology in calorimetry & the ILC **121**

6 MVA techniques applied to hadron reconstruction **122**
6.1 Introduction . . . 122
6.2 Motivation for using neural networks . . . 123
6.2.1 Application to hadron reconstruction . . . 123
6.2.2 Neural networks . . . 124
6.2.3 The multilayer perceptron . . . 126
6.2.4 Generalized Regressive Neural Networks . . . 128
6.3 Implementation & methods . . . 128
6.3.1 Use of a functional language . . . 128
6.3.2 A brief introduction to Erlang . . . 129
6.3.3 Modifying the HCAL fast simulation . . . 130
6.3.4 The toy models . . . 131
6.3.5 Network implementation . . . 131
6.3.6 Acquiring data . . . 132
6.3.7 CaloWindows . . . 132
6.4 Results . . . 133
6.4.1 Network response with the standard implementation . . . 133
6.4.2 Varying the toy model parameters . . . 135
6.4.3 Network performance . . . 140
6.5 Computing performance . . . 142
6.5.1 Theory . . . 142
6.5.2 Practice . . . 142
6.6 Conclusions . . . 144

7 Terapixel Calorimetry **145**
7.1 Introduction . . . 145
7.2 Future linear colliders . . . 145
7.3 The International Linear Collider (ILC) . . . 146
7.3.1 Design challenges . . . 146
7.3.2 Precision measurements at a linear colander . . . 147
7.3.3 Differences in particle flow from LHC to ILC . . . 148
7.4 The Monolithic Active Pixel Sensor . . . 149
7.5 Prototype sensor: characterization . . . 152
7.6 Prototype sensor: beam test . . . 156
7.6.1 Overview . . . 156
7.6.2 Aligning the system . . . 157
7.6.3 Making tracks . . . 159
7.6.4 Efficiency results . . . 160
7.6.5 Understanding the low efficiency . . . 162
7.7 Outlook & summary . . . 163
7.7.1 The second-round sensor . . . 163
7.7.2 Summary . . . 163

IV Epilogue **164**

8 Conclusions **165**

Appendices **167**

A CMS Testbeam reconstruction details **168**
A.1 Optimization of energy reconstruction . . . 168
A.2 Vetos to be applied . . . 169
A.3 Effect of selective readout . . . 170
A.4 Stabilized gaussian fits . . . 171

B Particle flow details **172**
B.1 Effect of seed thresholds on reconstruction efficiency . . . 172
B.2 Particle flow calibration functions . . . 174
B.2.1 Testbeam calibration functions . . . 174
B.2.2 Standard fast simulation calibration functions . . . 174

C MAPS Appendix **175**
C.1 Influence of temperature on prototype sensor noise . . . 175

References **176**

Colophon **183**

List of Figures

1.1 Higgs production at the LHC . . . 21
1.2 Higgs decay modes . . . 21
1.3 The CERN accelerator complex . . . 24
1.4 Schematic of the CMS experiment . . . 27
1.5 CMS assembly and installation . . . 28
1.6 CMS tracker . . . 30
1.7 CMS tracker material budget . . . 30
1.8 ECAL resolution for electrons in testbeam . . . 32
1.9 CMS HCAL layout . . . 33
1.10 Muon system performance . . . 34
1.11 Properties of PYTHIA/GEANT4 generated jets . . . 36

2.1 Coulomb interaction: Feynman diagram . . . 39
2.2 dE/dX examples . . . 41
2.3 The Landau distribution . . . 42
2.4 Landau-Bichsel straggling functions for silicon . . . 43
2.5 Bremsstrahlung: Feynman diagram . . . 44
2.6 Electromagnetic shower . . . 45
2.7 Critical energy for electrons in copper . . . 46
2.8 Simulation of longitudinal shower development . . . 46
2.9 Hadronic cascade . . . 48
2.10 F_0 fraction variation . . . 50
2.11 π/e as a function of e/h and F_0 . . . 51
2.12 Effect of compensation . . . 51
2.13 π/e response for CMS HCAL Barrel . . . 56
2.14 CMS calorimeter response . . . 57
2.15 The Large Detector Concept design . . . 58
2.16 The CALICE prototype . . . 59

3.1 The H2 testbeam environment in 2006 . . . 62
3.2 HCAL noise distributions for standard rechits . . . 68
3.3 ECAL activity for 50 GeV pions . . . 69
3.4 Barrel energy spectra . . . 71
3.5 Endcap energy spectra . . . 72
3.6 Summary of rechit spectra . . . 73

3.7 Simulation/data comparison . . . 74
3.8 Sample energy distributions . . . 75
3.9 Sample energy distributions II . . . 76

4.1 Particle flow reconstruction . . . 78
4.2 Particle flow clustering . . . 80
4.3 Linking clusters to tracks . . . 82
4.4 PFAlgo logic . . . 84
4.5 Multiple HCAL clusters, with one linked to the track . . . 86
4.6 Fake tracks . . . 87
4.7 Satellite clusters vs. excess energy . . . 90
4.8 PFCandidate calibration . . . 91
4.9 Event display for a 100 GeV charged pion event: single cluster . . . 91
4.10 Event display for a 100 GeV charged pion event: multiple clusters . . . 93
4.11 Barrel PFCandidate energy spectra . . . 94
4.12 Endcap PFCandidate energy spectra . . . 95
4.13 Single hadron response . . . 96
4.14 PFCandidate calibration and response (full simulation, barrel) . . . 96
4.15 Barrel PFCandidates— charged hadrons, with ECAL activity . . . 97
4.16 Hadronic energy resolution in testbeam . . . 98
4.17 Cluster multiplicity . . . 99
4.18 Endcap cluster multiplicity at 50 GeV . . . 99
4.19 Extra HCAL cluster contamination . . . 101
4.20 PFJet Benchmark . . . 102
4.21 PF Jet Benchmark: resolution . . . 104
4.22 Summary of Chapters 3 & 4 . . . 106

5.1 η distributions of PF hadrons from data . . . 110
5.2 $E_{\text{reco}}/p_{\text{track}}$ for PF hadrons from collisions (I) . . . 112
5.3 $E_{\text{reco}}/p_{\text{track}}$ for PF hadrons from collisions (II) . . . 113
5.4 Barrel PFCandidate energy spectra from collisions: no clusters in the ECAL . . . 115
5.5 Barrel PFCandidate energy spectra from collisions: with clusters in the ECAL . . . 116
5.6 Endcap PFCandidate energy spectra from collisions: no clusters in the ECAL . . . 117
5.7 Endcap PFCandidate energy spectra from collisions: with clusters in the ECAL . . . 118
5.8 Calibrating only the charged hadron's clusters ('Strategy 3') . . . 120

6.1 Rosenblatt's original perceptron model . . . 125
6.2 The multilayer perceptron . . . 126
6.3 Intel processor clock speeds . . . 130
6.4 CaloWindows . . . 133
6.5 Neural network performance with standard rechits . . . 134
6.6 Neural network performance with the toy models . . . 136
6.7 Network performance for toy models with low noise and low QE . . . 137

6.8 Linearity residuals . . . 138
6.9 Comparison of resolutions . . . 139
6.10 Correlations between inputs and outputs . . . 140
6.11 Training the networks . . . 141
6.12 Erlang parallelization performance gains for GRNN . . . 143

7.1 ILC schematic . . . 146
7.2 ILC beam parameters . . . 147
7.3 Vector Boson Fusion and the role of the Higgs . . . 148
7.4 The *Pandora* particle flow algorithm . . . 149
7.5 Event displays . . . 151
7.6 The INMAPS process . . . 152
7.7 The prototype MAPS . . . 153
7.8 Noise spectra . . . 155
7.9 DESY beam test configuration . . . 157
7.10 Residuals (r_x, r_y) for two sensors . . . 158
7.11 χ^2 probabilities for x, y spatial dimensions . . . 160
7.12 θ_z for candidate tracks . . . 161
7.13 Residuals of the fourth hit for a candidate track . . . 161
7.14 Sensor efficiency . . . 162

A.1 The 'banana' correction . . . 169
A.2 Effect of selective readout . . . 170

B.1 PFCandidate species . . . 173
B.2 PF Candidate Calibration . . . 174

C.1 A threshold scan . . . 175

List of Tables

1.1 The Large Hadron Collider 23
1.2 CMS' design and implementation 26

2.1 Radiation and nuclear interaction lengths 48

3.1 Beam purity and statistics 63

4.1 PF Clustering parameters 81
4.2 Constant *a* term 90
4.3 Example PF analysis for testbeam data 93
4.4 Jet energy resolutions 104

5.1 Quality cuts applied to charged hadron PFCandidates in collisions . . . 109
5.2 Statistics available after applying quality cuts 109

6.1 Logical layer readout 123
6.2 Summary of linearity residuals 137
6.3 Summary of resolutions 137
6.4 Inter-process chatter 143

7.1 Diode pad versus MAPS: parameters 152

A.1 ADC cut values for beam halo, scintillator, muon veto, *etc.* counters at the H2 testbeam 2006 and 2007 setups 170

B.1 Efficiency for reconstructing at least one hadronic neutral (barrel) for various energies and seed/cell thresholds 172

Glossary

Selected acroymns and terms

ADC	Analogue to Digital Converter
ALEPH	Apparatus for LEP (*q.v.*) Physics
ALICE	A LHC (*q.v.*) Ion Collider Experiment
ASIC	Application Specific Integrated Circuit
ATLAS	A Toroidal LHC (*q.v.*) Apparatus
BT	Bunch train
BX	Bunch crossing
CALICE	Calorimetry for a Linear Collider Experiment
CERN	European Organization for Nuclear Research
CLIC	Compact Linear Collider
CMS	Compact Muon Solenoid
CMSSW	CMS (*q.v.*) Software
DESY	Deutsches Elektronen-Synchrotron
EB, EE	ECAL (*q.v.*) barrel and endcap calorimeters
ECAL	Electromagnetic calorimeter
FPGA	Field Programmable Gate Array
GRNN	Generalized Regressive Neural Network
HCAL	Hadronic calorimeter
HB, HE, HF, HO	HCAL (*q.v.*) barrel, endcap, forward, and outer calorimeters
HPD	Hybrid photodiode
ILC	International Linear Collider
IP	Interaction point
IPC	Inter-process chatter
LEP	Large Electron Positron Collider
LHC	Large Hadron Collider
LHCb	LHC (*q.v.*) Beauty Experiment
MAPS	Monolithic Active Pixel Sensor
MIP	Minimum ionizing particle
MLP	Multilayer perceptron
MSSM	Minimally Supersymmetric Standard Model
MVA	Multivariate analysis
PF	Particle Flow

PS	Proton Synchrotron *or* Preshower detector
QCD	Quantum Chromodynamics
QE	Quantization error
SM	Standard Model
SPS	Super Proton Synchrotron
SR	Selective readout
SST	Silicon Strip Tracker
SUSY	Supersymmetry
TOF	Time of flight
TU	Threshold units
USBDAQ	Universal Serial Bus Data Acquisition Board
ZSP	Zero suppression

Selected mathematical symbols

$\oplus$	Addition in quadrature
ϕ	Azimuthal angle
e/h	Compensation ratio
E_c	Critical energy (electromagnetic showers)
F_{o}	Electromagnetic energy fraction
$\in$	Member of
$\not{E}_{\mathrm{T}}$	Missing transverse energy
R_M	Molière radius
λ_I	Nuclear interaction length
π/e	Pion to electron response ratio
η	Pseudorapidity
$\mathcal{R}$	Radial coordinate in $\eta\phi$-space
X_{o}	Radiation length
p_T	Transverse momentum

Part I

Prologue

Chapter 1

Introduction to CMS & LHC

> The Higgs boson remains as elusive today as when it was predicted over fifteen years ago.
>
> *Duane Dicus and Scott Willenbrock, 1985 [2]*

▷ 1.1 Overview

This introductory chapter outlines the reasons for building a hadron collider. We consider how the *Large Hadron Collider* (LHC)—a high-luminosity, 7 + 7 TeV, proton-proton collider—and the *Compact Muon Solenoid* (CMS) experiment are designed for making physics discoveries at the TeV scale. An introduction to the particle flow reconstruction method is presented.

▷ 1.2 The Large Hadron Collider (LHC)

1.2.1 Why build a hadron collider?

The Standard Model (SM) [3–8] has been a victim of its own success: based on quantum field theory, it has been extremely successful at describing the spectrum of fundamental particles observed and their interactions. In essence, it accounts for 3 pairs of quarks and 3 pairs of leptons. The interactions among them are governed by quantum gauge fields with mediating bosons, the $W^{\pm}$, Z^{o}, and the γ for the electroweak interactions and gluons, g, for strong interactions between quarks. Nevertheless, problems exist:

(i) It (sometimes) lacks mathematical consistency, has nearly 20 free parameters, and some predictions break down at energy scales $\mathcal{O}$(TeV) without elementary extensions, with more serious flaws appearing at the Planck scale;

(ii) The origin of mass has not yet been explained or verified;

(iii) It has no coherent and/or cogent explanation for the spectrum of particles observed;

(iv) Neutrinos have mass [9], a fact which is not accommodated by the theory;

(v) Gravity has not been (*cannot ever be?*) successfully integrated into the theory.

Furthermore, problems at the fundamental level of particles manifest themselves at the cosmological level: conventional Newtonian dynamics [10] and General Relativity [11] account for the motions of stars, galaxies and other cosmological objects, but more than 90% of the mass which accounts for these motions cannot be attributed to visible material seen in astronomical observations. So the business of particle physics also extends to searching for dark matter [12] and dark energy [13].

A panoply of theories has been concocted to resolve these issues, but only experiments can determine which theories are valid, appropriate, and useful. Here follows a brief description of two particular extensions that have proven popular with theorists in recent years.

The Higgs boson

The Higgs boson [14] is the mediator of a scalar field with a non-zero expectation value: it therefore has zero spin and couples to every particle that has mass, including itself. The fermionic coupling is proportional to the mass of the fermion too. At high energies $\mathcal{O}(Z \text{ mass})$, the SM, specifically the Weinberg-Salam model [5, 8], correctly predicts the unification of the electromagnetic and weak interactions in terms of their U(1) and SU(2) gauge groups. The model consists of a weak isospin triplet, **W**, an isospin singlet field, B, and an isospin doublet of scalar Higgs particles. At energies well below m_W, the electromagnetic and weak interactions decouple and the symmetry is broken. In the process of electroweak symmetry breaking, mixing among the original **W** and B fields gives rise to the massless photon—an unpredicted fact—while the Z and two charged W bosons appear and all acquire mass. The Higgs is predicted to be a free, neutral particle as a remnant from this process. Its mass is neither predicted by theory, nor has it been observed to date. At the time of writing, the LEP[1] and Tevatron experiments together with other measurements of the SM have determined a lower limit of $114\,\text{GeV}/c^2$ for the Higgs mass [15], and excluded the range 163 to $166\,\text{GeV}/c^2$, both with a 95% confidence limit[2] [16].

The various LHC Higgs production mechanisms are shown in Figure 1.1 [17]. At the LHC, an SM Higgs of light to moderate mass $\mathcal{O}(100\text{–}500\,\text{GeV}/c^2)$ will predominantly be produced by gluon-gluon fusion. The predicted branching ratios for its decays are exhibited in Figure 1.2.

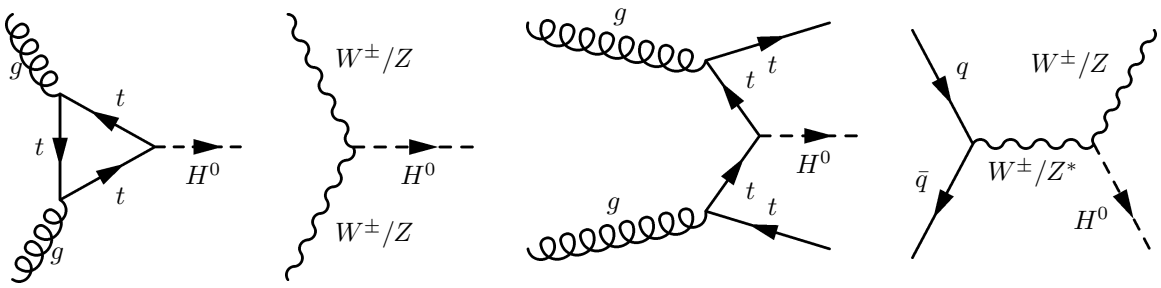
Supersymmetry (SUSY)

Supersymmetry [20] (introductions in [21, 22]) attempts to resolve the *hierarchy problem* [23, 24]: why are each of the fundamental forces' couplings so different in magnitude? Furthermore, if very high mass particles are admissible at an energy where the strong

[1] Large Electron Positron Collider, CERN, 1989–2000

[2] If we repeated our experiments many times over, for some given value of the Higgs mass, 95% of the time we would conclude that the Higgs mass was not in this range.

FIGURE 1.1 **Higgs production at the LHC**
Gluon fusion is the dominant process for Higgs production at the LHC. Top quarks form a virtual loop, and the Higgs couples to these massive quarks rather than to the massless gluons. The Higgs can decay to two photons by a similar loop process. Also shown are the *vector boson fusion* (VBF) process, and the less common associated production processes, namely $t\bar{t}$ fusion and 'Higgsstrahlung' $q\bar{q} \to W^{\pm}/Z^{*} \to W^{\pm}/Z + H$.

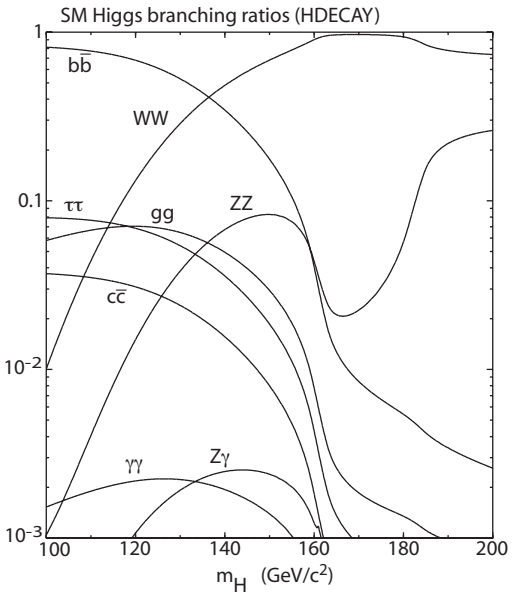


(a) gg fusion *(b) VBF fusion* *(c) $t\bar{t}$ fusion* *(d) 'Higgsstrahlung'*

and electroweak interactions unify ($E_{\text{uni}} \sim 10^{16}$ GeV), or even higher at 10^{19} GeV—the Planck scale—where gravity joins in, then the particles that exist at these scales may be indirectly experienced at lower energy scales, $\mathcal{O}(m_W)$, through loop corrections. These problems indicate 'unnatural' behaviour in our current understanding of the fundamental particles.

These *radiative corrections* make calculations of the W and Higgs masses divergent [22] without carefully constructed cancellations at the level $m_W c^2/E_{\text{uni}} \sim 10^{-14}$. SUSY arranges this by providing a broken fermion-boson symmetry: each fermion has a new supersymmetric boson partner and each boson has a new supersymmetric fermion partner, and the supersymmetric partners are called *sparticles*. Even though the symmetry is broken, implying the pairs of particles have different masses, the radiative

FIGURE 1.2 **Higgs decay modes**
For a light, $\mathcal{O}(< 200\ \text{GeV}/c^2)$, Higgs, computed by [18,19].



corrections from the supersymmetric partners are of opposite sign and suitable for regulating our otherwise divergent calculations.

The masses of the supersymmetric particles are anticipated to be $< 1\,\text{TeV}/c^2$ based on qualitative 'naturalness' arguments, but the cross-sections for producing them are expected to be substantial, $\mathcal{O}(0.01 - 0.1\,\text{pb})$, and should be readily detected at the LHC experiments [21]. Even in the most economical edition of SUSY, the *minimally supersymmetric* model (MSSM, [25]), a much richer Higgs spectrum is predicted, with no fewer than two CP-even scalars, h^0, H^0, a CP-odd scalar, A^0, and two charged Higgs, $H^\pm$, *and* their superpartners. Nevertheless, this comes at a cost as the number of free parameters in the SM balloons to over 100.

Physics beyond the Standard Model & the case for high luminosity

A hadron collider operating at a centre of mass energy an order of magnitude greater than accelerators previous to the LHC should resolve many of the shortcomings of the SM. Note that high-luminosity is required to extract the beyond-the-standard-model signals, which typically have small cross-sections and ambiguous event signatures from the SM backgrounds: since protons have quark and gluon substructure, the QCD backgrounds are expected to be substantial.

The final part of this thesis will examine the case for building a linear e^+e^- collider optimized for making precision measurements based on the anticipated discoveries to be made by the LHC, but without the complications of proton substructure and large backgrounds.

1.2.2 The machine

So much for theory and speculation. Now for the machine. The primary parameters of the LHC are its nominal beam energy of 7 TeV per beam, yielding a centre of mass energy of $\sqrt{s} = 14$ TeV, and nominal luminosity of $10^{34}\,\text{cm}^{-2}\,\text{s}^{-1}$: it is a 26.7 km circumference, superconducting machine, containing 11 GJ of stored-energy, resting between 50 and 100 m underground outside Geneva, Switzerland, with much of it extending into France. The technical challenges of building a machine to meet these requirements are terrific, and detailing the machine's implementation is well beyond the scope of this document. Many fascinating parameters are detailed in Table 1.1.

The LHC is injected with protons from the existing CERN accelerator complex, shown in Figure 1.3: after a proton source and short linear accelerator ('linac') and 'booster', the protons are injected into the *Proton Synchrotron* (PS), shaped into the required bunch structure with 25 ns spacing and accelerated to 26 GeV. They are then injected into the *Super Proton Synchrotron* (SPS) to be accelerated to 450 GeV before transfer to the LHC. This operation is repeated 12 times for each of the two beams, to fill the machine with the required bunch structure.

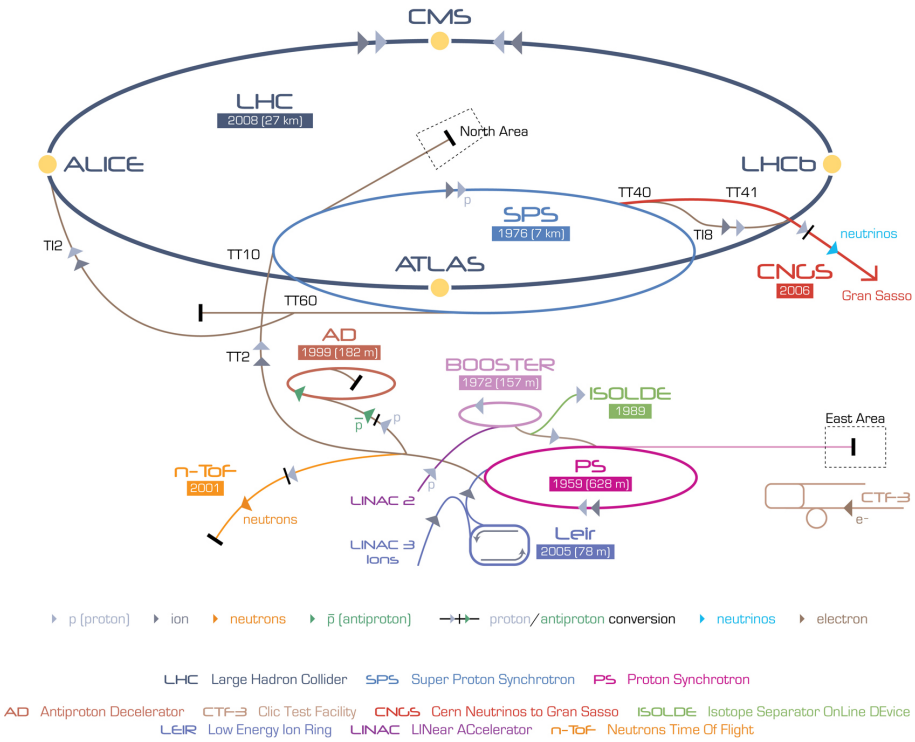
The LHC will take ~ 20 minutes to accelerate the beams to 7 TeV after injection. There are four collision points on the ring for the four main experiments: CMS, ATLAS [28], ALICE [29] and LHCb [30]. The LHC will also collide heavy ions at energies an order of magnitude higher than encountered at contemporary experiments, to further our understanding of nuclear matter.

TABLE 1.1 **The Large Hadron Collider**
Essential design parameters for the LHC, from a variety of sources [26, 27]. All values correspond to *pp* collisions where appropriate.

Collisions	
Energy per nucleon(1)	7 TeV
Dipole field at 7 TeV operation	8.3 T
Target luminosity	10^{34} cm^{-2} s^{-1}
Bunch crossing rate(2)	40 MHz
Revolution rate	11,000 s^{-1} or 660,000 RPM
Power & Energy	
Energy stored per beam(3)	350 MJ
B-field energy stored in magnets	11 GJ
Beam power loss	3.9 kW or 1.5 W m^{-1}
Energy loss per turn	6.7 keV
Acceleration per turn	0.5 MeV
Beam current	0.56 A
Power consumption(4)	120 MW
Hydrogen consumption (proton source)(5)	2 ng day^{-1}
Construction	
Number of dipole magnets × length	1,232 × 15 m
Number of quadrupole magnets × length	392 × 5–7 m
Total number of magnets	9,593
Superconducting strand and cable lengths	270,000 km ⇒ 7,600 km
Circumference(6)	26.7 km
Nitrogen mass	10,080 × 10^3 kg
Helium mass (inc. turbines and systems)	120 × 10^3 kg
Operating temperature	1.9 K
Vacuum pressure	10^{-13} atm
Cost (materials and people)	5 GCHF or 3 G€

Notes: (1) Equivalent to the kinetic energy of a typical (2 mg) *Culicidae* (mosquito) travelling at a few cm per second (*Author's calculation*). (2) For practical reasons (kicker magnet rise times, *etc.*), the collision rate will be 31.6 MHz. (3) The total stored beam energy is equivalent to the kinetic energy of a TGV 'Duplex' train travelling at 200 km h^{-1} (*Author's calculation*). (4) Equivalent to the domestic power consumption of the Canton of Geneva. LHC running cost ~ 19 M€ year^{-1}. CERN consumes ~ 230 MW total. (5) At this rate, it would take the LHC 1 million years to accelerate 1 gram of hydrogen. (6) ± 1 mm, depending on the phase of the Moon. Ground tides cause the Earth's crust to rise by some 25 cm in the Geneva area.

FIGURE 1.3 **The CERN accelerator complex**



Status at the end of 2009/start of 2010

The LHC was commissioned in September 2008 but, after a few days of running with stable beams, a catastrophic quench at a busbar junction between two superconducting dipoles caused severe damage to the machine: a large quantity, $\mathcal{O}$(5 tonnes), of helium was lost in the explosion, with some dipoles in sector '34' moving by up to 50 cm. An extensive programme of characterization, repair, and mitigation has been undertaken in the last year and the machine successfully restarted operations at the end of 2009. The general consensus has been to take a more conservative and cautious attitude to increasing the beam energy and luminosity, and this implies the machine may take several years before reaching its design goals.

▷ 1.3 The CMS Experiment

1.3.1 Physics goals

CMS' design concentrates on excellent lepton and photon reconstruction, with near complete 4π steradian coverage. This is achieved by means of a superconducting 4T solenoid containing the calorimetry and a high-precision tracking system. The decision to concentrate on lepton and photon reconstruction was driven by the physics we want to investigate and the fact that QCD processes at the LHC will form huge backgrounds to many processes.

Example: Higgs searches

Consider Figure 1.2: a low mass Higgs, $m_H < 130\ \mathrm{GeV}/c^2$, is expected to decay hadronically $H \to b\bar{b}$, but this is dominated by the QCD background, 10^5 greater in magnitude. It will be easier to discover a Higgs in this region by reconstructing the lepton channels: $H \to \tau\tau$ is attractive, but it will be hard to discriminate between real taus and narrow QCD jets faking taus. The most promising mode is $H \to \gamma\gamma$, despite its small branching ratio. A small diphoton mass resolution will allow us to recognize this mode relatively cleanly:[3] this motivated CMS' crystal electromagnetic calorimeter.

For higher Higgs masses, where the Higgs is likely to decay via W and Z bosons, it will generally be easier to reconstruct the final states with leptons.

Example: SUSY searches

SUSY is expected to manifest itself by the presence of missing transverse energy in the event, as the lightest stable SUSY particle is expected to interact very weakly. This may be accompanied by a large flux of both leptons and jets in the event. τ and b jets are especially likely, so these must be well recognized.

Detector requirements

For these reasons, and others related to §1.2.1, CMS is based around the requirements and implementation summarized in Table 1.2. Many aspects of the detector's implementation are described forthwith.

1.3.2 Design & implementation

Overview

CMS is a 21.6 m long, 14.6 m diameter, 12,500 tonne device located at Point 5 on the LHC ring at Cessy, France. It is a general purpose detector based on interlocking cylindrical 'barrel' subdetectors closed with 2 endcaps, all placed coaxially with the beam and centred on the beam interaction point. An illustration of its design, which is almost entirely driven by the 4T solenoidal field, is shown in Figure 1.4. The 13 m long, 5.9 m diameter solenoid contains 19.5 kA of current and 2.7 GJ of stored energy when running at its nominal field, and sustains 64 atm of stress across its windings. Images illustrating CMS' construction and installation are shown in Figure 1.5.

Coordinate system

This thesis uses the standard CMS coordinate system, defined in §1.2 of [26]: the origin is centred at the nominal collision vertex of the experiment. The y-axis points vertically upwards and the x-axis points to the centre of the LHC ring. The z-axis points to the Jura mountains, along the beam direction, and the polar angle, θ, is measured relative to the z-axis. The pseudorapidity is defined by $\eta = -\ln\tan(\theta/2)$. The azimuthal angle, ϕ, is measured relative to the x-axis in the xy-plane.

[3] The instrumental mass resolution dominates the natural width of the Higgs in this range.

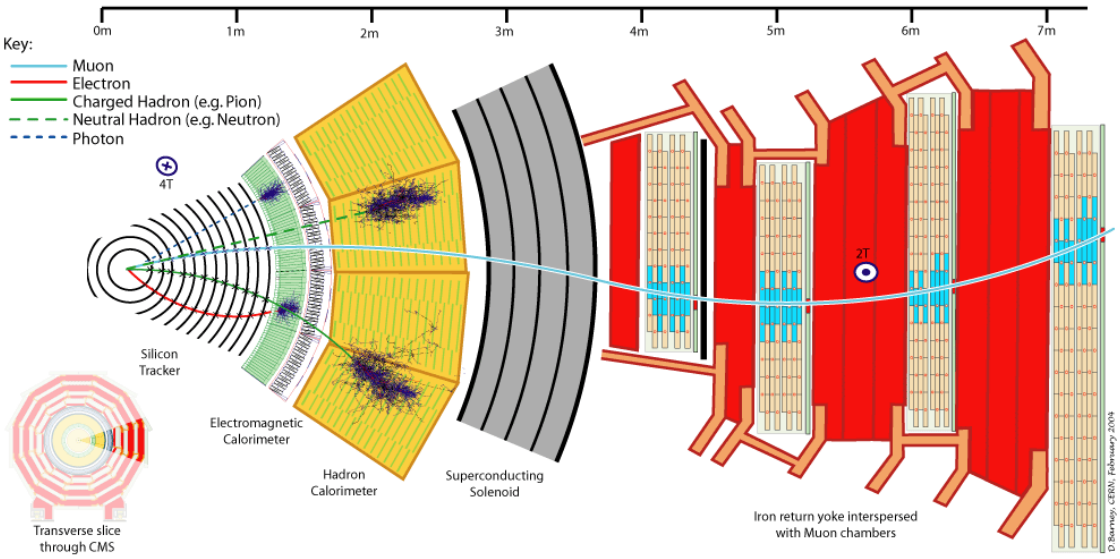
TABLE 1.2 **CMS' design and implementation**

How the physics we want to investigate (Higgs, SUSY, massive vector bosons, extra dimensions, standard model tests, and heavy ion physics) affects the design and how the implementation realizes these targets. Aspects of the calorimetry (radiation lengths and Molière radius) are extensively discussed in Chapter 2.

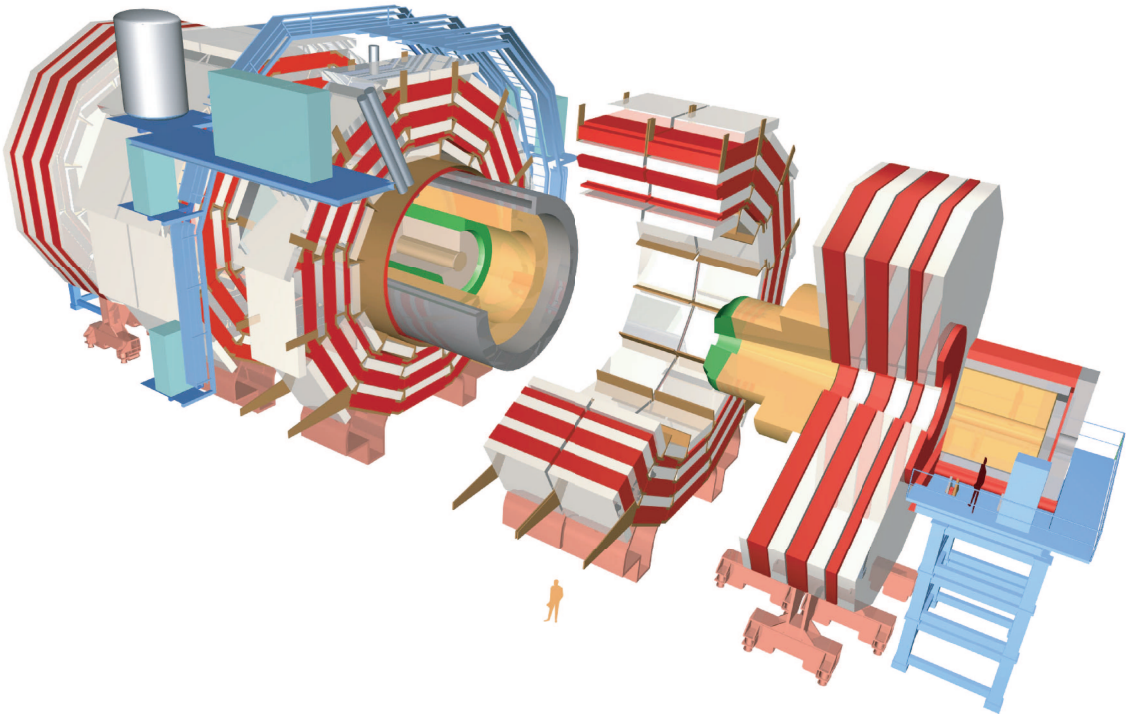
System	Objectives	Implementation
Muons	Coverage up to $\|\eta\| < 2.5$	4 layers of muon chambers in 1.5 m saturated iron
	Dimuon mass resolution ~ 1% at 100 GeV/c^2	4T field from magnet
	Identify muon charge up to $p = 1$ TeV/$c \Rightarrow \Delta p/p \sim 10\%$ at 1 TeV/c	3 types[1] of muon chambers + source for trigger
Tracking	Good resolution and reconstruction efficiency	5.8 m long, 2.6 m diameter silicon tracker
	High efficiency tagging of b, τ jets	10 layers of Si-μstrip detectors (pitch = 80–120 μm)
		3 layers of pixel detector (pitch = 100 × 150 μm^2) around IP
		Vertex location to within 20 μm
		4T field
Electromagnetic	Good EM resolution and coverage up to $\|\eta\| < 2.5$	ECAL coverage up to $\|\eta\| < 3.0$; 25 radiation lengths
	Diphoton/dielectron mass resolution ~ 1% at 100 GeV/c^2	61,200 (barrel) 2 × 7,324 (endcap) $PbWO_4$ crystals
	Direction measurement and/or localization of primary vertex	Crystal size ~ 1 Molière radius
	π° identification	π° preshower detector in endcaps (pitch = 1.9 mm)
	Efficient γ and lepton isolation at high luminosity	Source for trigger
		Two light detection technologies[2] (barrel and endcap)
		Compact, radiation hard & fast
Hadronic	Good $\not{E}_T$ and dijet mass resolution	HCAL sampling calorimeter of brass[3] & scintillator
	Coverage up to $\|\eta\| < 5.0$	11 interaction lengths of material in the barrel
	Fine lateral granularity $\Delta\eta \times \Delta\phi < 0.1 \times 0.1$	Radiation hard forward calorimeter
		$\not{E}_T$ measurement + source for trigger
		Tile segmentation $\Delta\eta \times \Delta\phi = 0.087 \times 0.087$
Trigger & electronics	Trigger attenuation rate of 10^9 events s^{-1} to 100 events s^{-1}	Hardware (Level 1) and software (High Level Trigger) triggers
	25 ns bunch spacing	Triggers process muon and calorimetry signals
		LHC Trigger Timing & Control (TTC) gives < 1 ns synchronization[4]
General	Radiation hard detector and electronics	*e.g.* Calibration of $PbWO_4$ crystals from rad. exposure
	Finite cost	~ 500 MCHF

Notes: (1) 3 types required for B-field considerations, triggering capability, and precision. (2) Require photodetectors that can operate in high B-field and with low light yield from $PbWO_4$: Silicon Avalanche Photodiodes in the barrel and Vacuum Phototriodes in the endcaps. Requires careful monitoring of temperature too. (3) Brass is non-magnetic. (4) The TTC system [31] operates across the LHC ring and within the detector. Timing of this precision is essential to stable running and triggering.

FIGURE 1.4 **Schematic of the CMS experiment**
Partially exploded view [32].



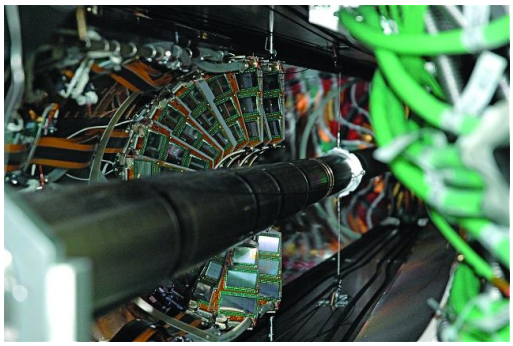
(a) A slice through the experiment



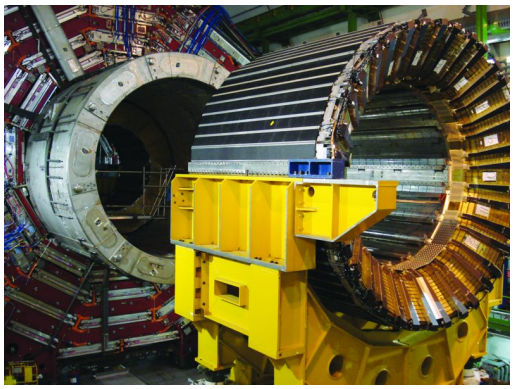
(b) Overall configuration

FIGURE 1.5 **CMS assembly and installation**

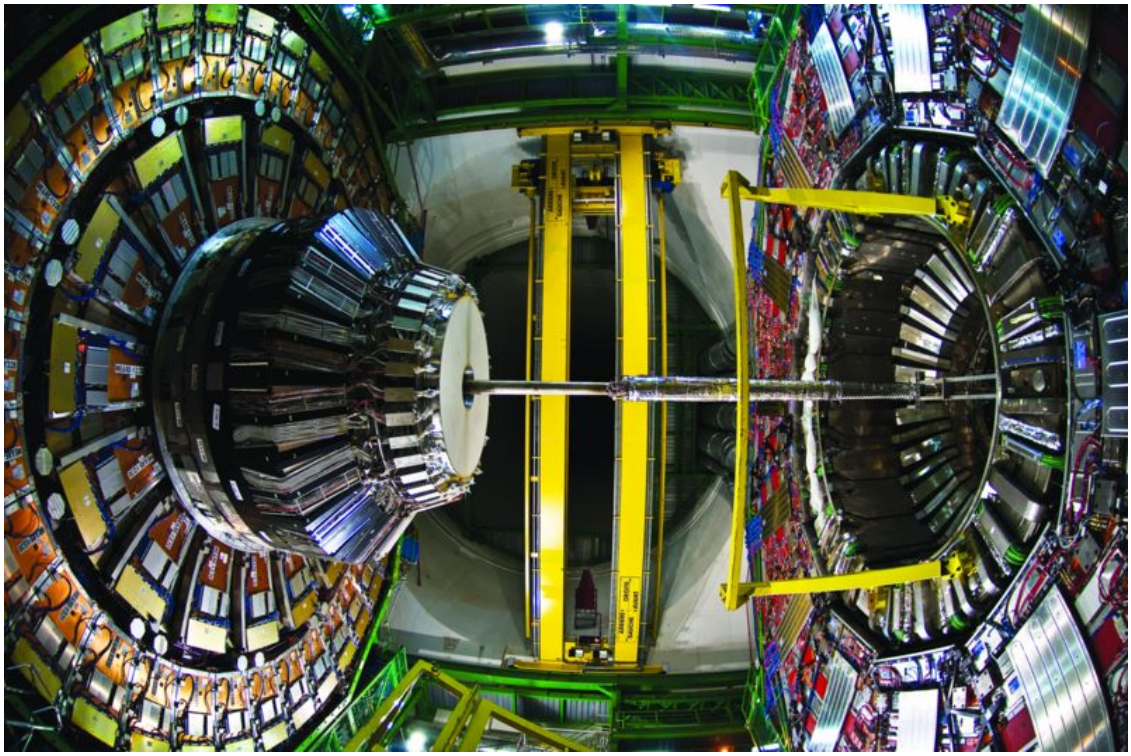
From the interaction point, to hadron calorimetry, and installation of the beam pipe before the experiment is closed by pushing the endcaps into the barrel. The CMS experiment weighs more than the combined weight of the ATLAS, ALICE, and LHCb experiments. Many other excellent and splendid pictures of CMS' construction and installation are available from [33].



(a) The 'forward' discs of the pixel detector at the interaction point. The inner radius of the pixel discs is 4 cm.



(b) The HCAL barrel prepared for installation inside the superconducting solenoid. The layers of brass/scintillator are visible.



(c) Installation of the beam pipe. An endcap is on the left, the barrel on the right. This view shows the shaft to the surface too. The white disc on the endcap is the preshower detector. A number of muon stations can be seen surrounding the endcap calorimetry. Straight lines appear distorted in this fish-eye photograph. For scale, the inner diameter of the solenoid on the right is 5.9 m.

1.3.3 A silicon tracker

Pixel detector

A pixel detector is at the very centre of CMS, with radii extending from 4.3 cm to 10.2 cm. It consists of 3 barrel layers of 53 cm length and 2 endcap discs at each end placed 34.5 cm and 46.5 cm from the interaction point. Here the radiation and particle fluence is tremendous, with a charged particle flux of 10^8 cm^{-2} s^{-1} at $r = 4$ cm. Data from the 66 million 100 × 150 μm^2 hybrid pixels—covering a square metre of silicon—form the 'seeds' for track reconstruction. The single point position resolution is 10 μm in $r\phi$ and 20 μm in z.

Strip tracker

After the pixel detector, the *silicon strip tracker* (SST) covers the region between the pixels and the ECAL: it is composed of several parts, the *tracker inner barrel* (TIB) has four layers ranging up to $|z| < 65$ cm and $r < 54$ cm and the *tracker outer barrel* (TOB) ranges up to $|z| < 110$ cm and $r < 116.5$ cm with 6 layers [34]. The pitch of the silicon strips varies between 80 μm and 120 μm in the TIB and 120–180 μm in the TOB, and both subdetectors have their first two layers placed at a stereo angle of 100 mrad to improve the measurement in the $r\phi$ and rz coordinates.

Each *tracker endcap* section (TEC) has 9 discs covering 120 cm $< |z| <$ 280 cm, and the *tracker inner discs* (TID) are 3 small discs covering the region between the TIB and the TEC. Both modules are arranged in rings concentric with the beam line, with the strips pointing towards the beam line too. Resembling petals of a flower, they therefore have a variable pitch ranging from 96 μm (TEC) and 97 μm (TID) to 143 μm.

All the sensor elements are tilted relative to the beam line to increase the Lorentz angle of the charged particles' trajectories through the silicon. In doing so, charge sharing between the pixels/strips is increased. The CMS tracker is an analogue-based device so charge-sharing is exploited to improve the overall single point resolution in $r\phi$.

In view of the much reduced particle fluence, the signal to noise ratio is improved by increasing the thickness of the silicon layer between the inner and outer tracker sections from 320μm to 500 μm. The TIB has 230 μm resolution in z and 23–34 μm resolution in $r\phi$. The TOB has 530 μm z resolution and 35–52 μm $r\phi$ resolution.

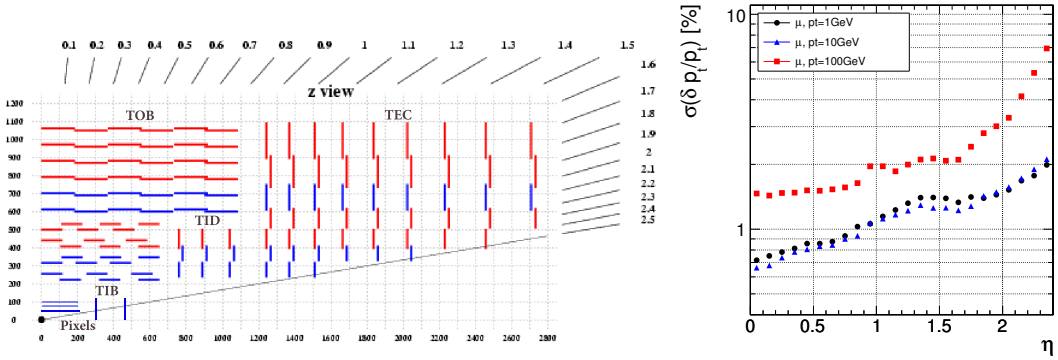
In total, there are 9.6 million strips on nearly 15,400 modules mounted on a carbon fibre support structure, cooled to −20° C, comprising an unprecedented 200 m^2 of silicon. Coverage extends to $|\eta| < 2.4$. The performance of the tracking system is summarized and exhibited in Figure 1.6. Finally, we should note that the tracker is composed of a substantial amount of material which can induce showering by the particles traversing the volume. The 'material budget' plots are shown in Figure 1.7.

1.3.4 Electromagnetic calorimetry: the ECAL

In the next two subsections, special attention is given to the materials and design of the calorimeters. In anticipation of more careful definitions to be given in Chapter 2, some terms are assumed, at this stage, to be 'common knowledge'.

FIGURE 1.6 **CMS tracker**

Layout and performance [26]. The single point resolutions of the barrel subsystems are summarized thus: pixel detector — 10 μm $r\phi$, 20 μm z; TIB — 23–34 μm $r\phi$, 230 μm z; TOB — 35–52 μm $r\phi$, 530 μm z.

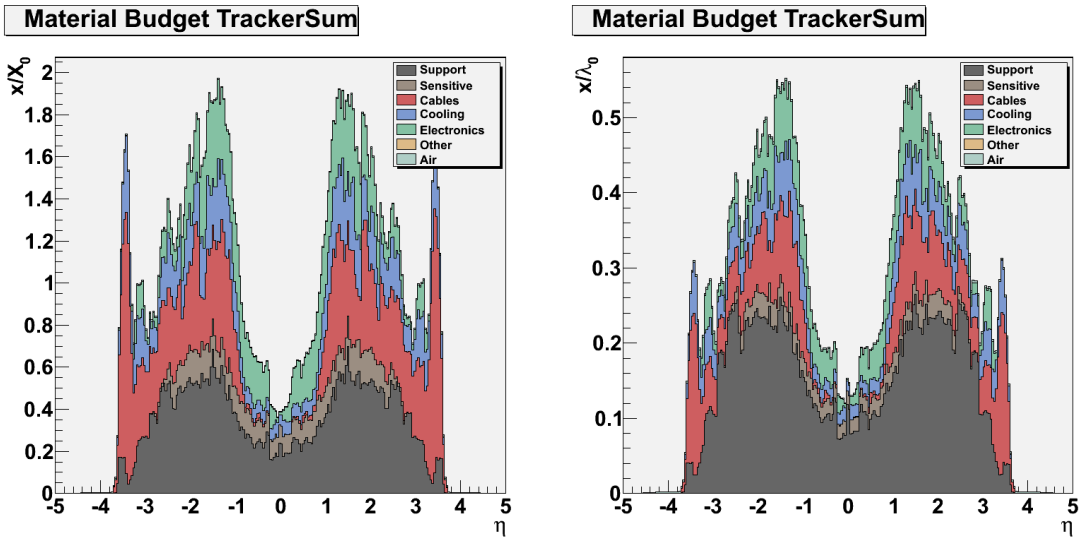


(a) Layout of 1/4 in z of the tracker.

(b) p_T resolution (simulation) for muons.

FIGURE 1.7 **CMS tracker material budget**

The silicon tracker presents a substantial amount of material to particles traversing the tracking volume, which leads to non-negligible probabilities for electromagnetic showering and hadronic interactions inside the tracker. The x-axis corresponds to the pseudorapidity. The y-axis expresses the amount of material as a fraction of a radiation length or interaction length. Plots credited to [35].



(a) Radiation lengths

(b) Interaction lengths

Material & design

The ECAL is a homogeneous calorimeter composed of 61,200 (barrel) and 2 × 7,324 (endcaps) lead tungstate ($PbWO_4$) crystals instrumented with *avalanche photodiodes* (APDs, barrel) and *vacuum phototriodes* (VPTs, endcaps). These devices have been selected for their high gain in view of lead tungstate's relatively low light yield and because they can function despite the high magnetic field experienced inside the solenoid. Lead tungstate was chosen for its (a) short radiation length, thus making for a compact calorimeter totalling 25.8 and 24.7 radiation lengths in the barrel and endcap respectively, (b) fast response, and (c) resistance to radiation. The disadvantage of the $PbWO_4$/APD combination is that both require careful temperature regulation to maintain a stable response.

Construction

The crystals have a square front face of 1° (0.0174 in η, ϕ), corresponding to approximately[4] 22 × 22 mm² and a length of 230 mm in the ECAL barrel (EB), which has an inner radius of 129 cm covering pseudorapidities up to 1.479. The crystals are placed on an $\eta\phi$ grid and point 3° away from the nominal interaction point, so particles are less likely to be 'lost' in the glass/carbon-fibre alveolar support structures which hold the crystals. The two endcaps (EE) are located ± 3.14 m from the interaction point, with their crystals arranged on an xy grid. Each crystal is identical with a front face of 28.6 × 28.6 mm² and 220 mm length. The coverage extends to $|\eta| = 3.0$.

The electronics permit a dynamic range of 15 bits for each channel. As the analogue pulse from a given channel extends over several bunch crossings, the signal is a weighted sum over multiple time frames. The noise has been measured to be 40 MeV per channel.

Preshower

Two identical *preshower* detectors cover much of the same $\eta\phi$ range as each of the endcaps: comprised of two discs of lead accounting for 2 and 3 radiation lengths respectively, their purpose is to initiate showering from π^0 mesons, electons/positrons, and photons. In this way, we can better identify π^0 mesons and improve the position measurement of electrons and photons, which is useful in the endcaps where the particle flux is expected to be higher. The showers are sampled by 2 mutually orthogonal planes of silicon strip detectors, one behind each lead layer, which have a pitch of 1.9 mm and a length of 63 mm.

Performance

The EB's crystals are arranged in 36 identical 'supermodules', one of which has been evaluated in a testbeam. With electrons impinging on the centre of a given crystal, and taking the sum of amplitudes from an array of 3 × 3 crystals with e^- directed at the central crystal, the energy resolution has been found to be:

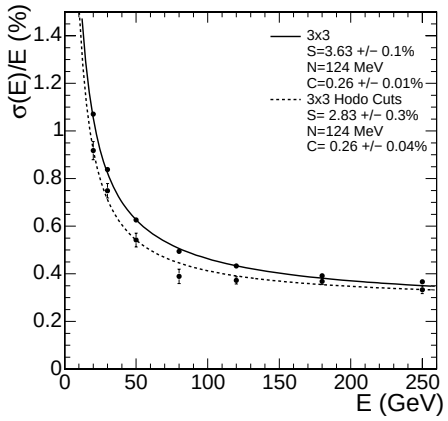
$$\frac{\sigma}{E} = \frac{2.83\%}{\sqrt{E}} \oplus \frac{124\,\text{MeV}}{E} \oplus 0.26\% \tag{1.1}$$

[4] Equal to the Molière radius—see Chapter 2.

where the ⊕ symbol indicates the terms should be added in quadrature. This curve is illustrated in Figure 1.8 and the significance of each of these terms will be discussed in Chapter 2.

FIGURE 1.8

ECAL resolution for electrons in testbeam
The lower line considers electrons which are known to shower in the central ± 2 mm of a crystal. The higher line considers events where the electrons impinge within a 20 × 20 mm² region—*i.e.* ~ one crystal's transverse extent—with a correction subsequently applied to make the optimal addition of crystal amplitudes based on the electron shower's transverse profile over the 3 × 3 crystal array. The fit parameters in the legend correspond to Eq. 1.1. Plot taken from [26].



1.3.5 Hadronic calorimetry: the HCAL

Material & design

The CMS HCAL is a sampling calorimeter providing over 10 interaction lengths of material over its entire pseudorapidity coverage of $|\eta| < 5.0$. Such a substantial depth prevents shower leakage which would otherwise degrade the energy resolution that can be attained. Furthermore, large coverage in η is essential for reliable 'missing energy' ($\not\!\!E_T$) measurements. The HCAL barrel (HB) and endcap (HE) detectors are composed of sandwiches of brass absorber instrumented with 3.7 mm thick plastic scintillators. The detectors are divided into a 'tower' geometry, with each of the 2,804 towers occupying 0.087 × 0.087 of $\eta\phi$-space (barrel). Light from each of the scintillating tiles in a given tower is coupled to wavelength shifting (WLS) fibres, which are spliced together into a single clear fibre, in turn joined to a *hybrid photodiode* (HPD). In the current implementation, no longitudinal information regarding the shower development is preserved in the barrel.

The 'hadron outer' system (HO) between the solenoid and muon system provides further containment of hadronic showers. It is composed of 5 rings, each 2.5 m long in z, nevertheless the scintillator geometry is the same as the HB's.

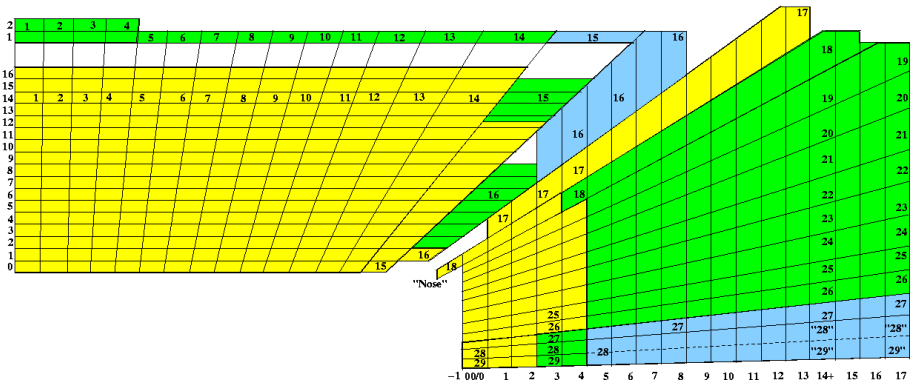
Finally, the 'hadron forward' system (HF) covers $3.0 < |\eta| < 5.0$ at $z = \pm 11$ m. It is composed of steel and quartz to make it extremely radiation tolerant. The absorber depth is 1.65 m, interleaved with 0.6 mm diameter quartz fibres, in a 5 mm square grid arrangement, with the fibres running *parallel* to the beam line. There are two sets of fibres, each with lengths of 1.43 m and 1.65 m which gives us two samples in depth—both start at the *back* of the calorimeter. The implications of this for calorimetry are discussed in §2.3.2, page 56.

Construction

The HB has fifteen ~ 5 cm thick brass plates and two external stainless steel plates for strength. The HB's coverage extends to $|\eta| < 1.4$. The HE covers $1.3 < |\eta| < 3.0$; the HB and HE overlap to allow shared services for the HCAL and ECAL to run between the HB and HE without any 'dead' areas. Each endcap tower has a $\Delta\phi$ size of either 5° or 10°, and varying $\Delta\eta$ size representing a compromise between the smallest tower size that can be accommodated and the increase in particle flux with pseudorapidity. The HF's segmentation is typically 0.175×0.175 in $\eta\phi$. A schematic of the HCAL's layout is shown in Figure 1.9.

FIGURE 1.9

CMS HCAL layout
A schematic showing the HCAL HB, HO, and HE segmentation in 1/4 of the rz-plane. The colours indicate depth segmentation used by the electronics. Plot taken from [26].



Signals from hadronic cascades may extend over many bunch crossings. As with the ECAL, the overall signal is a weighted sum of analogue samples from multiple time frames, digitized by a non-linear 7-bit ADC. The electronics noise is substantial, and will be shown in Chapter 2 to be 200 MeV and 350 MeV for the HB and HE respectively.

Performance

The response of the CMS calorimeters to hadrons is discussed in detail in Chapter 2. For now, it suffices to cite the energy resolution seen in testbeam [36] for the HB,

$$\frac{\sigma}{E} = \frac{111.5\%}{\sqrt{E}} \oplus 8.6\%, \tag{1.2}$$

but after applying the sophisticated 'banana corrections' of [36] (summarized in Appendix A.1) to correct for the system's non-linearity (again, see Chapter 2), we get:

$$\frac{\sigma}{E} = \frac{94.3\%}{\sqrt{E}} \oplus 8.4\%. \tag{1.3}$$

We use the last result as the standard benchmark resolution when comparing with other calibration methods discussed in this thesis.

1.3.6 Muon systems

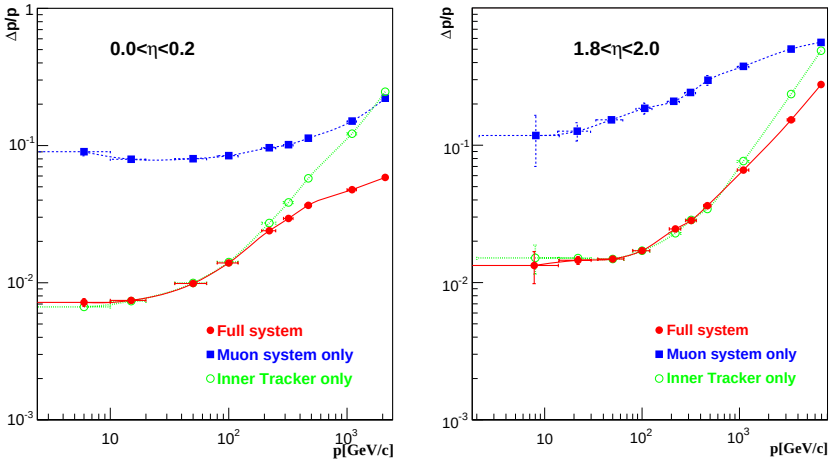
A dedicated muon detection and momentum measurement system surrounds the superconducting coil. It uses three types of gaseous chamber for reasons of (a) the large area

to be covered, (b) variations in the level of irradiation, particle flux, and magnetic field, (c) cost, (d) precision, and (e) speed for triggering. In the barrel, where the magnetic field is less intense, high precision *drift tube* (DT) chambers are employed. *Cathode strip chambers* (CSC) are used in the endcaps because they can operate under the higher magnetic field experienced there and under higher rates owing to (a) a higher muon flux in the forward direction, and (b) a higher neutron-induced background. Finally, *resistive plate chambers* (RPC) are employed in both regions because they offer a fast response with good time resolution, making them invaluable for triggering, but at the expense of a coarse position measurement. The CSCs and DTs are also used by the Level-1 Trigger.

A muon's momentum is best measured with a combination of the muon system and the inner tracking system: if only the muon system is used, then the measurement is dominated by multiple Coulomb scattering (see Chapter 2) up to transverse momenta $\mathcal{O}(200\,\text{GeV}/c)$. For high momenta, the chambers' spatial resolution dominates the measurement. The performance of the system is summarized by Figure 1.10. In total, the system provides 25,000 m² of active detection plates and ~ 1M channels.

FIGURE 1.10

Muon system performance
Momentum resolution $\Delta p/p$ as a function of p using various combinations of the inner tracking system and the muon system. The DT drift coordinate resolution is ~ 200 μm, which allows the direction in each of θ, ϕ to be known to ~ 1 mrad per station. The CSCs have similar position resolution in $r\phi$, but the angular resolution (θ) is only known to ~ 10 mrad. Plots taken from [26].



(a) Barrel *(b) Endcap*

1.3.7 Trigger & computing

A trigger is essential at a hadron collider because the LHC's collision rate of 31.6 MHz and 17 pp interactions per collision will yield $\mathcal{O}(10^9)$ events per second. Most collisions will not be of interest and can be discarded. In any case, it is not feasible to store all the events and only $\mathcal{O}(10^2)$ events per second can be archived. At CMS, this is accomplished by a two-level trigger. The *Level-1 Trigger* is hardware based, in turn heavily based on custom and flexible logic, such as Field Programmable Gate Arrays (FPGAs) and ASICs. It considers information from the calorimeters and muon systems to search for 'interesting' physics objects (muons, electrons, photons, jets, $\not{E}_T$) and makes a coarse decision based

on the information available and the 3.2 μs allocated to make the decision, before the on-detector buffers lose the event under consideration. Of the 3.2 μs total time, less than 1 μs is available for computation as the rest is consumed in cable latency between the cavern and the counting room where much of the trigger is located.

Following a Level-1 'accept', the software-based *High Level Trigger* (HLT) reduces the output rate from 100 kHz to 100 Hz for offline analysis, and a processor farm and switching network has been designed for this purpose. All subdetector information is available to the HLT for it to make its decision.

Finally, events are reconstructed 'offline', (*i.e.* after the event) in a distributed, multi-tiered, and worldwide computing model [37].

▷ 1.4 Why study hadrons?

Hadron reconstruction is the focus of this thesis, so it is worth justifying the topic. The earlier sections described how the (rare) lepton decay channels of physics processes are often the most easy to recognize, and this explains CMS' overall design. But QCD will nevertheless form a very large background to interesting physics processes, particularly where b and τ jets and $\not{E}_T$ are to be studied. Accurate jet reconstruction is required to exploit the statistics available.

Figure 1.11 illustrates the importance of charged hadrons in jet reconstruction at CMS. On average, over 60% of a jet's energy is transported by charged hadrons. Furthermore, the energy spectrum of the hadrons is exponentially falling, with most hadrons found at very low energies $\mathcal{O}(< 5\,\text{GeV})$. Observe, also, that neutral hadrons (excluding π^0 mesons) generally account for less than 10% of the total hadronic energy.

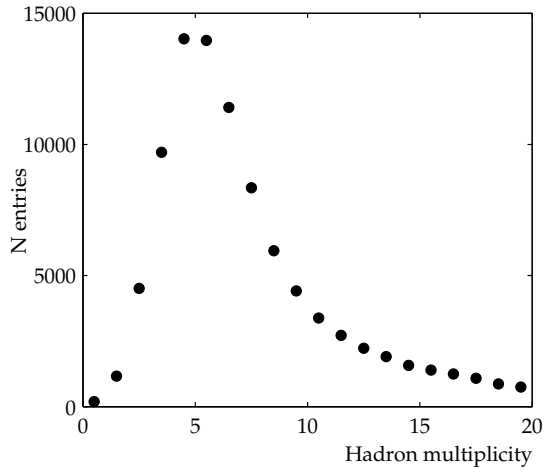
▷ 1.5 Particle flow reconstruction at CMS

Overview

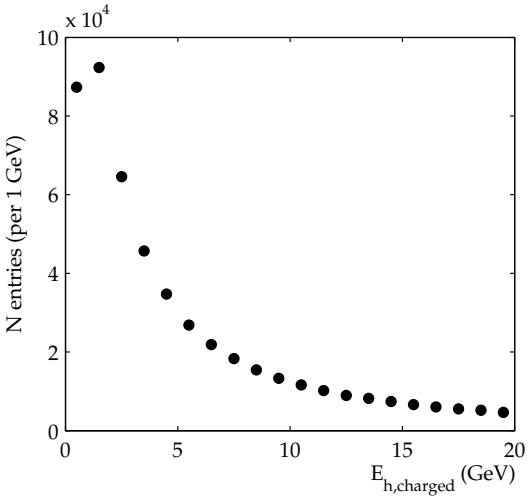
So, having built this detector, powered the magnets and pushed the protons around, what next? Most physics analyses consider an event in terms of basic visible physics objects, namely, electrons, photons, muons and jets, and, indeed, $\not{E}_\text{T}$. These objects are the final manifestations of the original hard scattering process between the quarks and gluons, *i.e.* the *partons*. By reconstructing the four-momentum vectors of the visible objects, we can distinguish between events of interest (signal) and events containing physics processes which are ostensibly already understood (background).

The idea of *particle flow* (PF) is not new. It was originally pioneered at ALEPH [38] and has deployed, albeit in a limited sense, at hadron colliders in the past [39, 40]. Nevertheless, the CMS PF reconstruction is general in its approach and suitable for widespread deployment in the collaboration. The essential idea is to analyse the event using *all* the subdetectors available, together with their known performance and limitations. The CMS PF [41] algorithm's output is a list of 'candidates': electrons, muons, photons, taus, and charged and neutral hadrons. Each candidate's type should be faithful to the real particle it represents, with a useful and accurate four-momentum description. Composite objects such as jets and missing energy can be built from these.

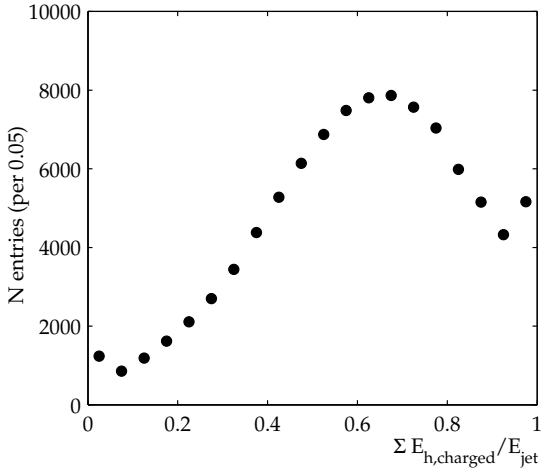
FIGURE 1.11 **Properties of PYTHIA/GEANT4 generated jets**
9,000 QCD multijet events were produced with a pseudo-flat $\hat{p}_T$ spectrum ranging from 15 to 3000 GeV/c for pp collisions at $\sqrt{s}$ = 14 TeV. The plots show various aspects of each jet's composition. π^0 mesons decay electromagnetically, and are not considered 'neutral hadrons' in these plots. (Jets were made using the iterative cone algorithm, with $\Delta\mathcal{R}$ = 0.5. This will be formally discussed in §4.6.)



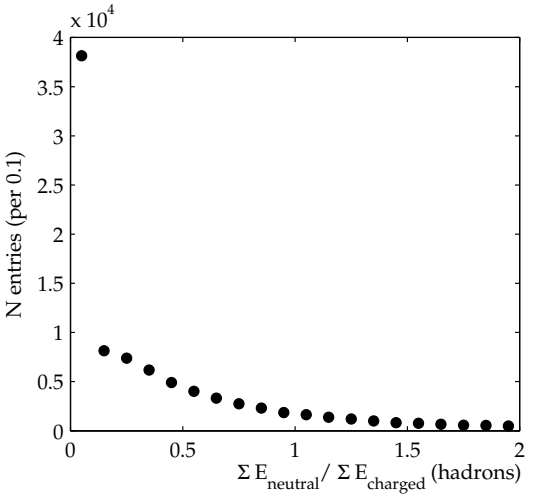
(a) Hadron multiplicity



(b) Energy spectrum of charged hadrons



(c) Fraction of jet energy transported by charged hadrons (the remainder is delivered by photons/π^0 mesons and neutral hadrons)



(d) Ratio of energies carried by neutral hadrons to those carried by charged hadrons

Combining tracking with calorimetry

Reconsider CMS' design, viz. Figure 1.4, subfigure (a). This diagram illustrates how each of the stable final state particles interacts with each subdetector. The logic *you* use to interpret this diagram is akin to that employed by PF at CMS. For instance, a neutral hadron is inferred by the presence of localized HCAL activity without the presence of a track linked to it. Indeed, much of this effort, and this thesis, may be summarized by one question:

Do the tracker and calorimeters agree?

While the answer is perhaps binary in nature, the solution and its implications are complex. With ~ 17 pp interactions for each LHC beam crossing at high luminosity, 'pile-up' and backgrounds complicate the situation because the final state particles—or their interactions in the detector at least—will often overlap. Furthermore, we must account for the resolutions of each subdetector, which are invariably dependent on energy, momentum and spatial coordinates, and perhaps accumulated radiation dose. We will examine how PF reconstructs hadrons in Part II of this thesis.

Is CMS suitable for the task?

To reconstruct the final state particles accurately requires fine detector granularity and good resolution. CMS is well suited to such a microscopic treatment.[5] The high magnetic field and performant silicon tracker can reconstruct tracks with high efficiency and resolution. Photons are measured by the very high resolution crystal ECAL, whose transverse segmentation is usually fine enough to distinguish between individual photons and electrons, if their shower maxima are not too close. Bremsstrahlung photons emitted by electrons as they accelerate in the B-field (see Chapter 2) are reliably recognized and attributed to reconstructed electron candidates. Photons which pair-convert before reaching the ECAL can usually be recognized and correctly reconstructed. Nuclear interactions and particles which undergo multiple scattering in the tracking volume are also accounted for.

The HCAL's transverse segmentation is 25 times poorer than the ECAL's with somewhat poorer resolution, and is therefore not *ideal* for reconstructing hadrons on an individual basis. Nevertheless, neutral hadrons can be inferred from an excess of calorimeter energy with respect to track momentum in the vicinity of HCAL energy clusters. Furthermore, PF corrects for the non-linear response of the calorimeters.

▷ 1.6 Summary

This chapter covered:

- ▷ The background and motivation for building the LHC and the CMS experiment.
- ▷ A description of the CMS detector, and discussion of how its design will help search for new physics at the TeV scale.
- ▷ A brief description of the particle flow technique and the advantages CMS should expect from its deployment.

[5] 'Microscopic', in the sense that individual particles are to be reconstructed.

Chapter 2

Aspects of calorimetry in HEP

▷ 2.1 Introduction

Objectives

The first part two parts of this thesis describe particle flow's response to hadrons using data acquired from CMS calorimeter testbeams conducted in 2006 and 2007. Particle flow transforms elementary subdetector information into physics objects, so we need to understand the raw inputs to the algorithm in order to judge the quality and performance of the algorithm's outputs. In this chapter a brief overview of the theory underpinning the design of calorimeters in high energy physics is described, followed by an examination of the implications for CMS and ILC-based calorimeters. We examine the drawbacks associated with a non-compensating calorimeter system.

Disclaimer

This is not intended to be an *exhaustive* review of calorimetry for high energy physics. Some elements not directly relevant to calorimetry at CMS and the ILC are excluded (*e.g.* Čerenkov radiation).

▷ 2.2 Calorimetry

In high energy physics experiments, calorimeters generally seek to reconstruct the energy of particles emanating from the collision vertex with both high accuracy and precision, which are assessed in terms of the devices' *responses* and *resolutions* respectively. Calorimeters can measure the energy of a particle independent of the particle's charge, though differences in the specific showering processes may allow the species of particle to be identified. In contrast to measuring the momentum of a particle with tracking, the relative energy resolution improves with incident energy. A hermetic design with full 4π

solid angle coverage allows for the detection of an energy imbalance in the transverse plane, making calorimeters the essential tool for $\not{E}_T$ analyses. Furthermore, they tend to be fast devices which makes them invaluable for triggering.

2.2.1 Practicalities

Designing a calorimeter system involves compromise and technical constraints (such as shower containment, cooling, response time, power consumption, and hermeticity). CMS' design is optimized for the reconstruction of leptons and photons and the excellent crystal-based ECAL should achieve this goal, but this results in an extremely imbalanced response to hadrons over electrons and photons, which therefore affects jet reconstruction.

2.2.2 The interaction of particles with detection media

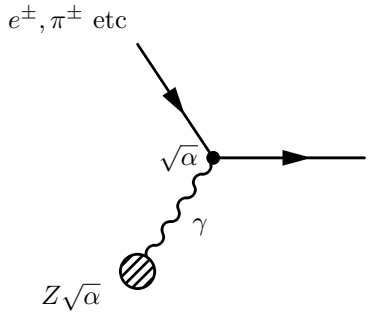
Particles generally interact with calorimeters either through (a) elastic scattering, (b) inelastic atomic ionization and excitation interactions, or (c) showering (including pair production and annihilation). We will examine each effect in turn. Electrons and photons will be given particular attention in §2.2.3. There are many excellent sources of information on calorimetry, for example [21, 42, 43].

The Coulomb interaction

While its relevance to calorimetry is negligible, relativistic charged particles may interact with the target medium's nuclei via the Coulomb interaction (Figure 2.1). This is an elastic interaction because the nuclei are massive, so it causes only slight changes to the direction of the charged particle. Also note that the nuclei are generally shielded by the atomic electron cloud.

FIGURE 2.1

Coulomb interaction: Feynman diagram
The electron's path is deflected due to the electromagnetic interaction between the electron and the nucleus, which has nuclear charge Z. A virtual photon is exchanged between the two. The strength of the interaction will be proportional to $\sqrt{\alpha} \times Z\sqrt{\alpha}$, where α is the fine structure constant.



Ionization processes. The Bethe-Bloch formula

Energy loss may occur from atomic electrons in the target medium being either excited or ionized, and the rate of energy loss per unit distance dE/dX is described by the

Bethe-Bloch formula [44]:

$$-\frac{dE}{dX} = 4\pi N_A r_e^2 m_e c^2 z^2 \frac{Z}{A}\frac{1}{\beta^2}\left[\ln\frac{2m_e c^2\gamma^2\beta^2 T_{\max}}{I^2} - \beta^2 - \frac{\delta}{2}\right] \tag{2.1}$$

where,

- ▷ z is the charge of the incident particle;
- ▷ A is the nucleon number and Z the atomic charge;
- ▷ I is the mean ionization potential of the medium empirically determined from measurements of dE/dX ($\sim \mathcal{O}(10) \times Z$ eV);
- ▷ The classical electron radius is,

$$r_e = \frac{1}{4\pi\epsilon_0}\frac{e^2}{m_e c^2} = 2.82 \times 10^{-13}\text{cm}; \tag{2.2}$$

- ▷ δ is a density effect, which accounts for polarization of atoms in the electric field of the incident particle. More distant atoms are shielded by their intermediate neighbours, and so interact less strongly with the particle and contribute less to its loss of energy;
- ▷ $T_{\max}$ is the maximum kinetic energy which can be imparted to a free electron in a single collision (see [43] for details);
- ▷ The remaining symbols N_A, γ, β, m_e carry their standard meanings.

Examples of the rate of energy loss are shown in Figure 2.2. Note that the dependence on the material is quite weak since $Z/A \sim 0.5$ for many materials and only deviates from this substantially for the very heavy elements.

The minimum ionizing particle

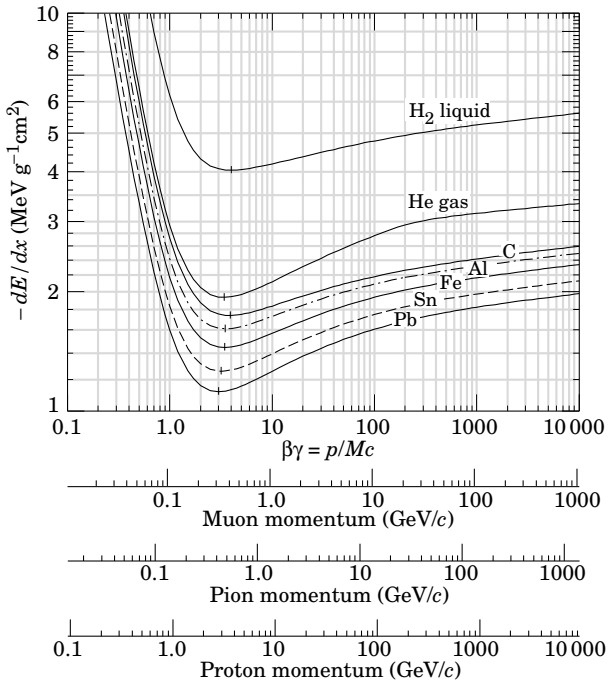
At low energies where $\beta\gamma \lesssim 2$, $dE/dX \sim 1/\beta^2$ and ionization of the material is the dominant energy loss mechanism. At moderate energies ($10 \lesssim \beta\gamma \lesssim 100$), $dE/dX \sim \ln\beta^2\gamma^2$, and we observe the *slow relativistic rise*, which turns into the *Fermi plateau*[1] at very large β. The minimum occurs at $\beta\gamma \sim 3$–4 where $dE/dX \sim 1$–$2\,\text{MeV g}^{-1}\,\text{cm}^2$, *regardless* of the species of particle (provided the charge is the same). Particles travelling in this range are called *minimum ionizing particles* (MIPs). Note that the slow relativistic rise is, indeed, very slow (the plot is shown on a double log scale): therefore most particles with momenta $\beta\gamma > 3$ are usually deemed to be MIPs.

Fluctuations & the Landau distribution

Note that Eq. 2.1 gives the *average* energy loss. Nevertheless, in general the processes involved are stochastic and fluctuations around the average result. For example, collisions with small energy transfers are more probable than those with high energy transfers, so

[1] The Fermi plateau is explained (highly non-trivially!) by restricting energy losses to $T \leq T_{\text{cut}} \leq T_{\max}$. dE/dX then approaches a constant.

FIGURE 2.2 dE/dX **examples**
For several materials, extracted from [43]. Radiative effects (see below), which affect muons and pions with $\beta\gamma \gtrsim 1000$ (though this decreases for materials with high Z) are not included. Note the units express dE/dX in a density-independent way.



the most probable energy loss is shifted to lower values. Landau described a quantitative model [45] of these fluctuations by solving the integral transport equation,

$$\frac{\partial f}{\partial x} = \int_0^\infty w(\epsilon)\,[f(x, \Delta - \epsilon) - f(x, \Delta)]\,d\epsilon, \tag{2.3}$$

where $f(x, \Delta)$ is the probability density function for a particle to lose an amount of energy between Δ and $\Delta + d\Delta$ when traversing a layer of thickness x. The function $w(\epsilon)$ is the probability per unit path length of a particle losing energy ϵ in a collision, in turn assumed to be small compared to the particle's original energy. Equation 2.3 can be solved analytically using Laplace transformations to give,

$$f(x, \Delta) = \frac{1}{2\pi i} \int_{c-i\infty}^{c+i\infty} e^I dp \tag{2.4}$$

where,

$$I = p\Delta - x \int_0^\infty w(\epsilon)(1 - e^{-p\epsilon})d\epsilon. \tag{2.5}$$

To evaluate this integral requires an explicit form for $w(\epsilon)$. Landau assumed the free electron cross section,

$$w(\epsilon) = \frac{\xi}{x}\frac{1}{\epsilon^2} \tag{2.6}$$

where,

$$\frac{\xi}{x} = \frac{2\pi e^4 \rho N_A Z}{m_e c^2 \beta^2 A} \tag{2.7}$$

and the symbols used have the same definitions as those given above. Landau showed that the integrand of Eq. 2.5 can be expanded and written as,

$$I = p(\Delta - \bar{\Delta}) - x \int_0^{E_{\max}} w(\epsilon)(1 - e^{p\epsilon} - p\epsilon)d\epsilon \tag{2.8}$$

with terms $\mathcal{O}(p^2)$ and higher discarded. $E_{\max}$ is a kinematic constraint on the maximum value of energy transfers. $\bar{\Delta}$ is the average energy loss in the layer (see below). Substituting this result into Eq. 2.4 one gets,

$$f(x, \Delta) = \frac{1}{\xi}\phi(\lambda) \tag{2.9}$$

where the universal function $\phi(\lambda)$ is given by,

$$\phi(\lambda) = \frac{1}{2\pi i} \int_{-i\infty+\sigma}^{i\infty+\sigma} e^{u \ln u + \lambda u} du \tag{2.10}$$

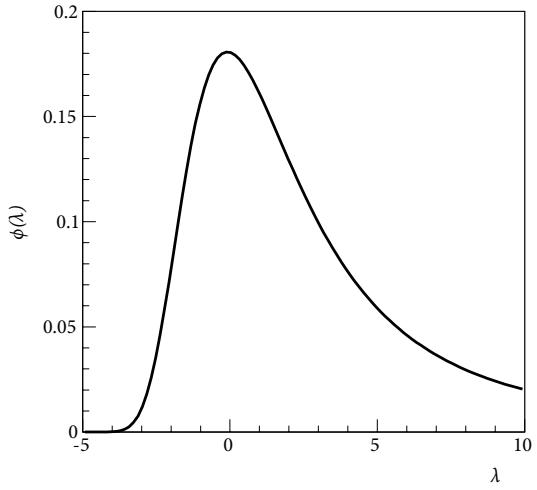
for σ taking any positive real value and,

$$\lambda = \frac{\Delta - \xi\left(\ln \frac{\xi}{\epsilon'} + 1 - C\right)}{\xi} \tag{2.11}$$

where $C = 0.557\ldots$ is Euler's constant. Eq. 2.9 is known as the *Landau distribution* and its functional dependence on λ is exhibited in Figure 2.3. It is asymmetric with a long tail extending to high energies. The function $\phi(\lambda)$ has a maximum at $\lambda \simeq -0.223$ with a full width at half-maximum of 4.02ξ.

FIGURE 2.3

The Landau distribution
The functional dependence of Eq. 2.9 is given by the universal function $\phi(\lambda)$ of Eq. 2.10 which is plotted here.



The original Landau theory exhibited good agreement with data but has subsequently been improved upon by Vavilov [46], Bichsel [47], and others: by expanding Eq. 2.5 to higher orders and using more realistic (*i.e.* band theory) models of ionization for $w(\epsilon)$, the theory better models experimental results. The most probable energy loss is given

by [43, 47],

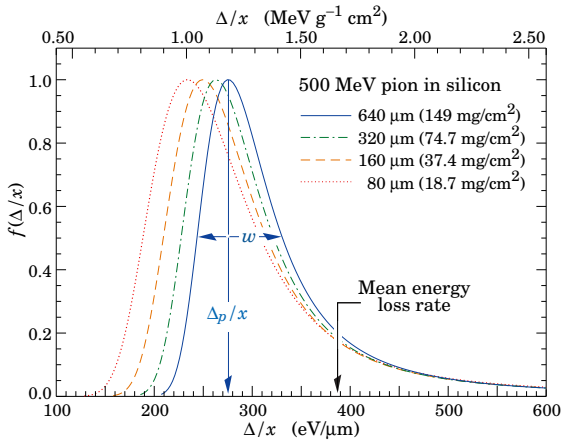
$$\bar{\Delta} = \xi \left[\ln \left(\frac{2m_e c^2 \beta^2 \gamma^2}{I} \right) + \ln \frac{\xi}{I} + j - \beta^2 - \delta \right] \tag{2.12}$$

where $j = 0.200$. The tail of large single-collision energy transfers evident in the Landau distribution (confirmed in data) implies that the *mean* energy loss given by the Bethe-Bloch equation is not appropriate for modelling the energy loss of single particles and that the most probable energy loss, $\bar{\Delta}$, should be taken instead.

For very thin media typical of silicon detectors, or volumes with particularly low densities such as gas trackers, the Landau distribution greatly underestimates the width of the fluctuations, and extensive corrections are required [43, 47, 48]. Characterizing the extent of the fluctuations in these implementations has been important to their adoption in high energy physics experiments. Straggling functions computed for silicon are shown in Figure 2.4.

FIGURE 2.4

Landau-Bichsel straggling functions for silicon
Distributions of energy loss for 500 MeV pions in silicon, normalized to unity at the most probable value $\bar{\Delta}$ (Δ_p on the plot). The width, w, is the full width at half maximum; observe that the width increases as the layer thickness decreases. From [43].



2.2.3 Electromagnetic showers

Energy loss by photons & electrons

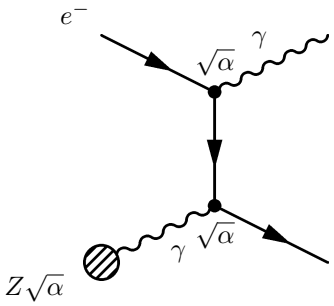
At energies $\mathcal{O}(< 1\,\text{MeV})$, photons lose energy via the photoelectric effect and Compton scattering. At higher energies, pair production is dominant.

The rate of energy loss by ionization for high energy electrons $\mathcal{O}(>1\,\text{MeV})$ can be approximated by Eq. 2.1 where $T_{\text{max}} = E$, $\beta = 1$ and $z = 1$. Energy loss by ionization therefore goes as $Z \ln E$. Nevertheless, this process is soon dominated by electrons radiating energy in the electric field of the target nuclei though *bremsstrahlung*. The Feynman diagram for this process is shown in Figure 2.5.

While all charged particles experience bremsstrahlung, electrons (by comparison with muons, hadrons *etc.*) are very light, so the deceleration is more severe. The Feynman diagram contains 3 vertices with couplings governed by $\sqrt{\alpha}$; note that the propagator will be of order $1/m^2$. In particular, it can be shown [42] that energy loss through

FIGURE 2.5

Bremsstrahlung: Feynman diagram
The electron interacts with the electric field of the nucleus via a virtual photon (shown bottom left). A real photon (top right) is radiated by the electron, as it decelerates in the electromagnetic field of the nucleus.



bremsstrahlung is governed by,

$$-\frac{dE}{dX} = 4\alpha N_A r_e^2 \frac{Z^2}{A} E \ln \frac{183}{Z^{1/3}}. \tag{2.13}$$

A numerical factor $4\ln(183/Z^{1/3})$ is introduced in the derivation of this result to account for the range of impact parameters available to the incoming electron impinging on the target nucleus. At very large impact parameters, the electron will only see the (outer) atomic electrons.

An electron may occasionally cause an atomic electron to be ejected with sufficient energy to cause further ionization of its own accord. These *delta rays* may be of comparable energy to the original electron.

Radiative effects for high energy muons

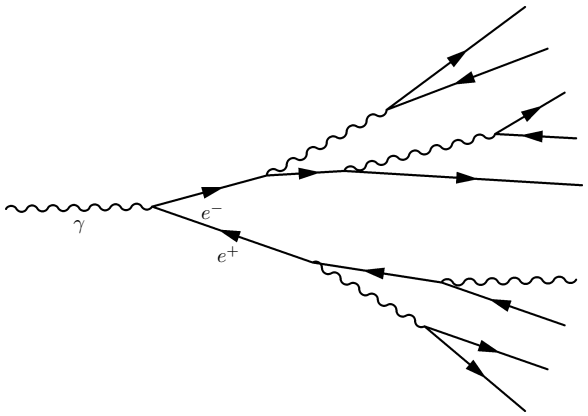
Radiative effects for muons of very high energy $\mathcal{O}(> 0.1$ TeV) are substantial and require careful modelling: photons radiated in high Z materials in this energy regime are usually hard $\mathcal{O}(10$ GeV), and subject to fluctuations. Failure to consider these effects can reduce the tracking efficiency. Further discussion may be found in [43, 49], but as muons are not the prime focus of this thesis, they will not be discussed in more detail.

Showering

Electrons and photons therefore lose energy through a continuous showering of secondary particles: photons pair produce, with the electrons and positrons subsequently radiating photons via bremsstrahlung and so on. Figure 2.6 illustrates the development of a typical electromagnetic shower. As the depth of the shower increases, the number of secondary particles increases too. Evidently, the average energy of each secondary particle must also decrease with depth. The process continues up to the *critical energy*, E_c, and the scale of each generation is given by the *radiation length*, X_0, both of which are now discussed.

FIGURE 2.6

Electromagnetic shower
In this example, a photon experiences pair production $\gamma \to e^+e^-$. The electrons and positrons subsequently experience bremsstrahlung, and the process repeats until the fermions reach the critical energy E_c. The typical generation length at each step is (ideally) one radiation length, X_0 (see text).



Critical energy

Note that the rate of energy loss via bremsstrahlung increases as Z^2E, much faster than the losses from ionization ($\sim \ln \beta^2\gamma^2$ at moderate energies). The critical energy, E_c, is the energy at which the two are, on average over many events, equal:

$$\left(\frac{dE}{dX}\right)_{\text{rad}} = \left(\frac{dE}{dX}\right)_{\text{ion}} \tag{2.14}$$

and it can be shown that [42],

$$E_c \simeq \frac{560\,\text{MeV}}{Z} \tag{2.15}$$

which for electrons travelling through copper is approximately 20 MeV, as shown in Figure 2.7. Electrons of critical energy will not travel far $\mathcal{O}(< 1X_0)$ after production.

Radiation length

Aggregating the ugly factors of Eq. 2.13 together, we can write,

$$-\frac{dE}{dX} = \frac{E}{X_0} \tag{2.16}$$

where,

$$\frac{1}{X_0} = 4\alpha_{\text{em}} N_A r_e^2 \frac{Z^2}{A} \ln \frac{183}{Z^{1/3}}. \tag{2.17}$$

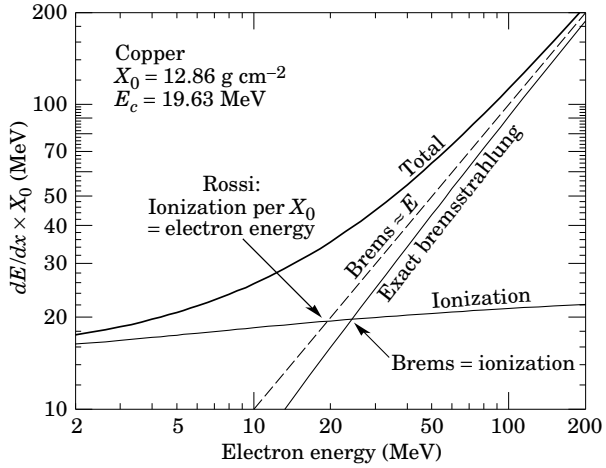
X_0 is the radiation length, and its units are g cm^{-2} (when A is in gm mol^{-1} and r_e in cm). Dividing this by the density gives X_0 in cm. From Eq. 2.16, one can write $E = E_0 e^{-x/X_0}$, defining the radiation length as the distance over which an electron will (on average) lose all but $1/e$ of its energy. Several materials' properties are presented in Table 2.1 on page 48.

Longitudinal & transverse development

Let us define the variables $t = x/X_0$ and $y = E/E_c$. A simple shower model might reasonably say that an electron (or positron) will radiate a photon, or that a photon will

FIGURE 2.7

Critical energy for electrons in copper
This figure, from [43], shows two definitions of the critical energy: the standard definition is given by Eq. 2.14 but an alternative given by Rossi [50] defines the critical energy as the energy at which the energy loss *per radiation length* is equal to the electron energy. It is equivalent to the first when $(dE/dX)_{\text{Bremm}} \approx E/X_0$ is taken as an approximation: this is shown as the dashed line on the plot (beware the units on the y-axis). The Rossi definition of critical energy is better for modelling the transverse extent of electromagnetic showers (see below).



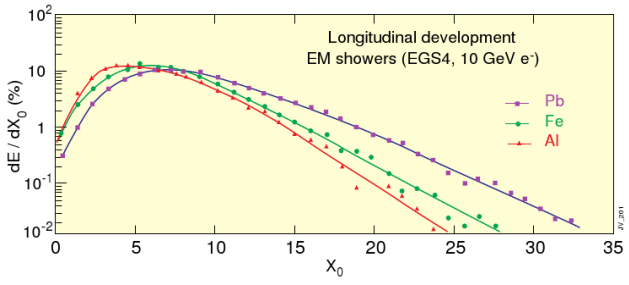
pair convert approximately every radiation length travelled.[2] The number of particles doubles after each step, so after t steps we have $n(t) = 2^t$ particles, and the energy in each is simply $E(t) = E/2^t$. The maximum number of particles will be in the final step where $n(t_{\text{final}}) = E/E_c = y$ and the maximum shower particle density will occur at,

$$t_{\text{final}} = \log_2 \frac{E}{E_c} = \log_2 y, \tag{2.18}$$

which indicates that, for two materials with the same X_0, deeper showers will occur in that material with the smaller critical energy. Materials with higher Z generally have lower critical energies—see Table 2.1. Some examples of typical longitudinal development from simulation are presented in Figure 2.8.

FIGURE 2.8

Simulation of longitudinal shower development
Reproduced from [51]. Observe that materials with higher Z exhibit both (a) a deeper shower maximum and (b) extended shower development overall. Both effects are attributable to the critical energy $E_c \propto 1/Z$.



The transverse development of an electromagnetic shower arises from multiple scattering of electrons away from the axis of the shower. The *Moliére radius* R_M [52]

[2] For photons the mean free path is actually ~ 9/7 X_0 because they travel deeper before their first conversion.

indicates the average deflection of electrons of critical energy after travelling one X_0. It can be modelled with [53, 54],

$$R_M = X_0 \frac{21\,\text{MeV}}{E_c} \tag{2.19}$$

if the Rossi definition of E_c (from Figure 2.7) is used. Approximately 90% of the shower energy is contained within $1\,R_M$ (integrated along the entire extent of the shower). A shower is generally composed of a narrow core of high energy particles, accompanied by a soft halo surrounding the core of particles scattered by the Coulomb interaction, the extent of which increases with depth as the number density of particles in the shower increases.

2.2.4 Hadronic cascades

Hadrons deposit their energy in matter via similar processes to those described for electromagnetic interactions. Nevertheless, the strong interaction is responsible and this leads to complications:

- ▷ The strong interaction of the original hadron breaks up the target material's nuclei, leading to further high energy hadrons.
- ▷ These participate in further interactions of their own.
- ▷ Nuclear breakup and spallation occur. These processes can be deeply inelastic and random, and energy is released from the nuclear binding energy difference between mother and daughter nuclei. The signals from these processes may also be delayed relative to the main cascade if the target nuclei are excited.
- ▷ The process continues as far as the di-pion production threshold.
- ▷ π^0 mesons decay to $\gamma\gamma$, which then shower electromagnetically. So the total cascade is a convolution of an electromagnetic interaction and a strong interaction.

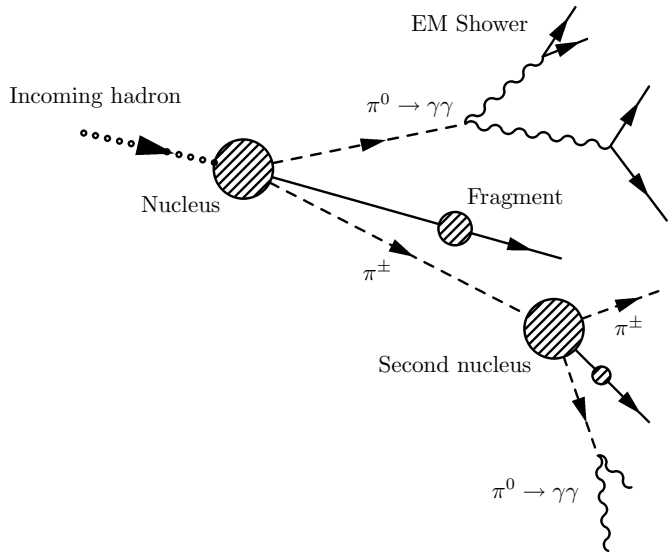
In this thesis the phrase *hadronic cascade* is used to emphasize the difference in complexity between hadronic interactions and electromagnetic showering.

Nuclear interaction length

Compared to electromagnetic showers, hadronic cascades vary enormously on an event-by-event basis. Consequently they are very difficult to simulate [55]: even state of the art methods depends heavily on Monte Carlo models featuring little determinism and empirically parametrized results from experiment. Once again however, we can get a quantitative grip on the situation by introducing a scaled variable $v = x/\lambda_I$ and the critical energy $E_{c\prime} = 2m_\pi = 280$ MeV, where the prime on the c distinguishes this from the electromagnetic critical energy. λ_I is the *nuclear interaction length*, which describes the path length for each generation in a hadronic cascade. It is very roughly approximated by [43],

$$\lambda_I \simeq 35\,\text{g}\,\text{cm}^{-2} A^{1/3}. \tag{2.20}$$

FIGURE 2.9 **Hadronic cascade**
Note the mixing of electromagnetic and hadronic components: hadronic cascades typically have a prompt electromagnetic component. See text for details.



Longitudinal & transverse development

Longitudinal energy depositions typically exhibit a prompt and large fraction of energy deposited at the start of the cascade due to localized electromagnetic showering from π^o mesons produced in the first interactions. If we assume that $\langle n \rangle$ secondaries are produced at each step in the shower, then the energy of each particle in that step is $E(v) = E/\langle n \rangle^v$. At the critical energy, $E(v_{\max}) = E_{c'}$ so,

$$n^{v_{\max}} = \frac{E}{E'_c} \Rightarrow v_{\max} = \frac{\ln(E/E_{c'})}{\ln\langle n \rangle} \tag{2.21}$$

TABLE 2.1 **Radiation and nuclear interaction lengths**
For a sample of materials, from [43]. The CMS HCAL is composed of *cartridge brass*, (type C26000 – an alloy of 70% Cu and 30% Zn by weight).

	Z	E_c (MeV)	X_o (cm)	ρ (g cm^{-3})	λ_I (cm)
Air			30,429		70,000
Si	14		23.6	2.32	
$PbWO_4$		0.67	0.89	8.28	20.7
Fe	26	24	1.76	7.9	16.8
W	74	8.1	0.35	19.3	9.6
U	92	6.2	0.32	19.0	10.5
C26000	~ 29.3		1.5	8.83	16.4

There is an important consequence: since the critical energy for hadronic cascades is much higher than the electromagnetic counterpart, the number of final state particles that can be detected will be reduced by a factor $> 2m_\pi/E_c^{\text{em}}$ so the resolution will be worse by the square root of this factor (see §2.2.7). Following the first 'hard' scatter with associated π^o production, 'soft' collisions follow (which may involve further π^o production), which generally describe an exponential decrease in the amount of energy

deposited per unit length, characterized by the scale λ_I. Furthermore, since the processes involved in hadronic cascades are subject to much greater stochastic variation, many interaction lengths are required to reliably contain the shower. Typical hadronic calorimeters tend to be physically large and massive — see the example parameters for a selection of materials in Table 2.1.

Transverse development is usually described by a hard core of activity led by the initial π^0 production (with a characteristic width of one R_M, since this activity is electromagnetically led) surrounded by a hadronic halo of scale λ_I. The transverse extent is not highly correlated with the incident particle energy.

2.2.5 Compensating calorimeters

On the interaction of electromagnetic and hadronic components in hadronic cascades

What are the implications for calorimetry given that hadronic cascades contain an electromagnetic interaction component driven by early π^0 mesons? Let us denote the total fraction of energy transported electromagnetically as F_0. Assuming that pions are the most commonly produced particles in the hadronic interactions, then approximately a third of them will be neutral (π^0, π^+, π^- are produced in similar ratios). In free space, charged pions decay weakly mainly via $\pi^+ \to \mu^+ + \nu_\mu$, but the neutral pions will decay electromagnetically $\pi^0 \to \gamma\gamma$, and frequently do so before reaching the front surface of the calorimeters. If the energy of the secondary pions is high enough, they may interact with other nuclei before decaying.

The total electromagnetic fraction, F_0, is the sum of electromagnetic fractions, f_0, generated at each step, t, in the shower:

$$\begin{aligned} &(t=1) \quad f_0 \\ &(t=2) \quad f_0 + f_0(1-f_0) \\ &(t=3) \quad f_0 + f_0(1-f_0) + f_0(1-f_0)^2 \quad \textit{etc.} \end{aligned}$$

which implies,

$$F_0 = 1 - (1-f_0)^t. \tag{2.22}$$

For low energies $\mathcal{O}$(a few GeV) where t is small, $F_0 \to f_0$ and the electromagnetic fraction is determined only by the first π^0 production. For high energies, the electromagnetic fraction will be much larger ($F_0 \to 1$). Consider Figure 2.10 as an example.

Compensation

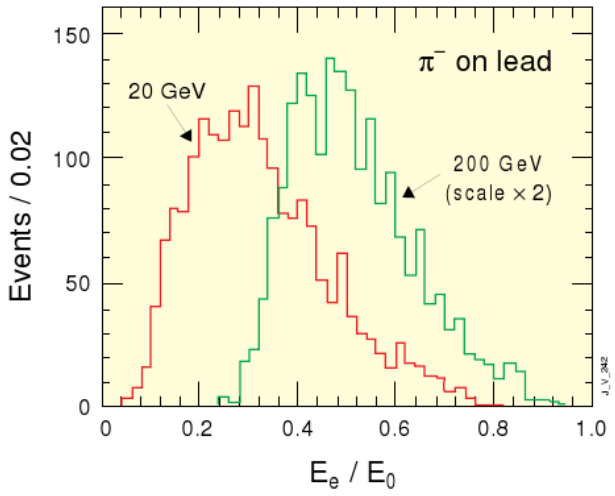
In general, the responses to the electromagnetic and hadronic components of a hadronic cascade are different. The electromagnetic response is called e and the hadronic response h. For some incident particle energy E, if the particle is an electron, the response E_e is,

$$E_e = eE \tag{2.23}$$

but for pions we have a response E_π,

$$E_\pi = \left[eF_0 + h(1-F_0)\right] E \tag{2.24}$$

FIGURE 2.10 **F_0 fraction variation**
Distribution of electromagnetic energy fraction F_0 for π^- incident on lead. In this figure, the x-axis represents the ratio of shower energy delivered electromagnetically E_e to the original electron energy E_0, which corresponds to F_0 in the text. Reproduced from [56].



so the ratio of the pion to electron response is,[3]

$$\frac{E_\pi}{E_e} = \frac{1 + F_0(e/h - 1)}{e/h}. \tag{2.25}$$

It is conventional to refer to this ratio as just 'π/e'.

Figure 2.11 shows the evolution of Eq. 2.25 for various values of F_0. When $e/h = 1$, the calorimeter is said to be *compensating*. Revisiting the subject of event-to-event fluctuations that are important in hadronic cascades, it is clear from Figure 2.11 that variations in F_0 imply a variation in the response π/e. So, if $e/h \neq 1$, there are consequences:

- ▷ The energy resolution will not necessarily be Gaussian any longer, even for monoenergetic hadrons;
- ▷ $\pi/e \neq 1$ *and* it varies with energy;
- ▷ The energy resolution will be degraded due to event-to-event fluctuations in F_0.

From Figure 2.11, one can estimate that if $e/h \gtrsim 1.1$, then the fluctuations in F_0 will affect π/e substantially.

Overall, the hadronic response is non-linear (Figure 2.12) and the resolution degraded. In general, $e/h > 1$ (under-compensating) since much of the energy from nuclear breakup is undetected (as much as 40% or more of the energy is rendered invisible). Ways to boost the hadronic response to restore compensation include (a combination of) (a) using Uranium-238 to improve the hadronic energy component from the fission of nuclei, (b) suppressing the electromagnetic response with thicker absorber plates (effectively decreasing the ratio X_0/λ_I), and/or (c) improving the response to low energy neutrons. It is clear from Figure 2.12 that the HELIOS experiment [57, 58] has excellent compensation, and that pion response is linear across the entire energy range.

[3] This formula is valid for a homogeneous calorimeter. Modifications are required for a multi-system structure like CMS's calorimetry.

FIGURE 2.11

π/e as a function of e/h and F_o
The contours show lines of constant F_o. The e/h ratio for the CMS HCAL has been added. Event-to-event fluctuations in F_o therefore induce large event-to-event fluctuations in the response. Note that e on the y-axis refers to 'electron response', while the e of the x-axis corresponds to the electromagnetic response, in accordance with the conventions of the text and general literature.

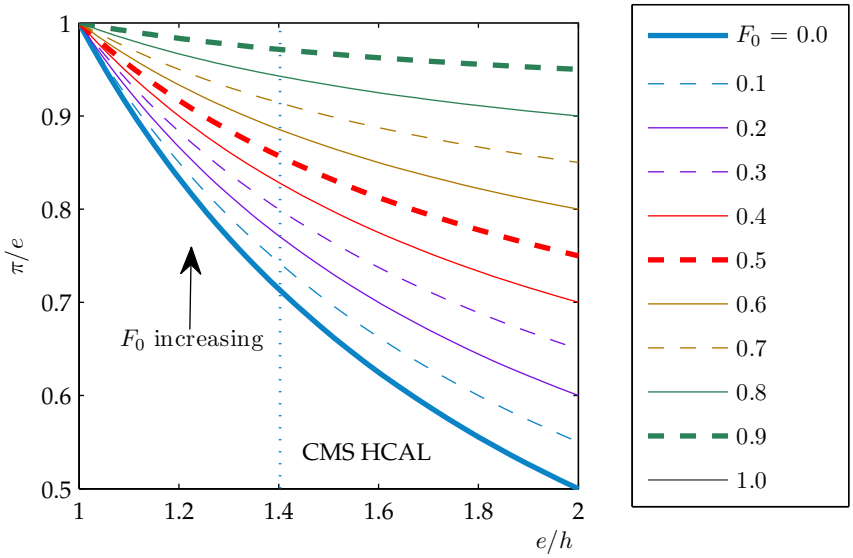
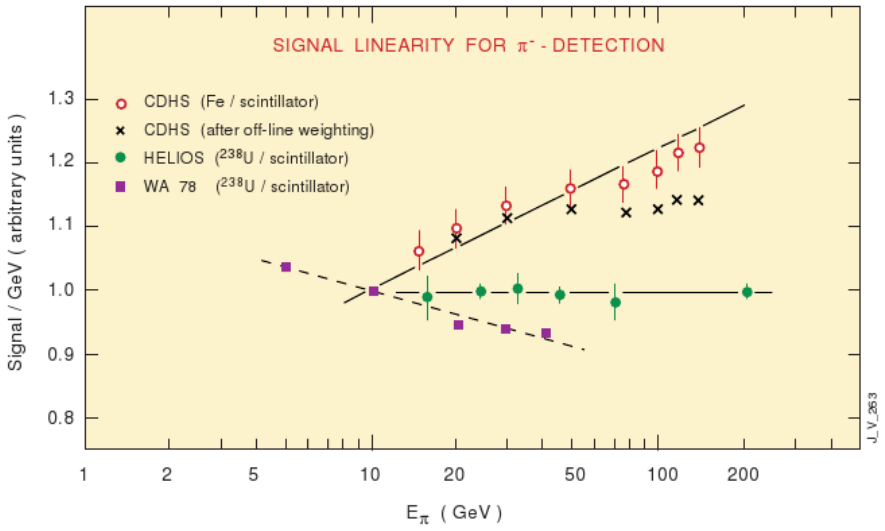


FIGURE 2.12

Effect of compensation
This shows the non-linearity observed in reconstructed energy for π^- particles for a few experiments. The 'Signal' of the y-axis has been normalized to original energy of the incident pion. The compensating HELIOS calorimeter exhibits a linear response. Reproduced from [51].



Note that the ratio e/h is not explicitly dependent on energy, and cannot normally be measured directly; all the energy dependence is assumed to be in F_0. Two parametrizations are generally available, either Groom's [56],

$$F_0 = 1 - (E/0.76)^{-0.13}, \tag{2.26}$$

or Wigmans' [59],

$$F_0 = 0.11 \ln E. \tag{2.27}$$

2.2.6 Sampling and homogeneous calorimeters

Two general calorimeter designs exist: the material in a *homogeneous* calorimeter both absorbs the energy of the incident particle and generates a signal. By contrast, *sampling* calorimeters typically consist of a sandwich of a passive absorber and active medium. The absorber causes the particle to shower and the secondaries typically give rise to scintillation light in the active layer, for which many possible implementations are possible, such as scintillators, gas detectors, and silicon detectors.

The mechanical design of a calorimeter should be such that (a) the showers are well contained in the detector volume (requiring a large number of X_0, λ), (b) there are enough samples to give an accurate energy measurement in spite of the systematic contribution due to random shower fluctuations, and (c) it fits within the detector structure and operation. Generally, the last point is non-negotiable. Note that it also includes constraints on the timescale over which the detector reconstructs a signal. We will now consider the effect of (b).

2.2.7 Energy resolution

For reasons that will become evident, the energy resolution of a calorimeter is usually expressed in the form,

$$\frac{\sigma}{E} = \frac{a}{E} \oplus \frac{b}{\sqrt{E}} \oplus c \tag{2.28}$$

where the symbol $\oplus$ indicates that these terms are added in quadrature, and the square root of the total taken to evaluate the result. The following interpretations are conventionally attributed to each term:

a — **Noise term** Represents sources of energy which contribute to the measurement regardless of the event under question, such as noise and *pileup*. Pileup refers to particles contributing to the energy measurement from outside the space-time volume under consideration.

b — **Stochastic/sampling term** The contribution to measurement error attributable to statistical fluctuations in showering (such as photostatistics). If showers from particles of a given energy generally provide $\langle N \rangle$ detectable secondaries (such as photons), the variation in this number will generally go as $\sqrt{N}$, so the measurement error will therefore be proportional to $\sqrt{N}$. Therefore, it is generally preferable to choose materials with low critical energies E_c to maximize N, and hence reduce the fractional error which is proportional to $1/\sqrt{N}$.

c — **Constant term** Independent of energy, this term is attributed to the non-uniformity of signal generation or collection, errors of calibration *between* detector elements, and fluctuations which are directly proportional to energy, such as the electromagnetic component F_o of hadronic cascades. If the quality of a detection element degrades with time or otherwise, (radiation damage and colour centre development, for example), then the shower maximum's position t_{max} may also vary. This may affect the detection efficiency of the secondary particles. Poor containment of the shower in the detection media (longitudinal, and to some extent transverse leakage) and dead areas in front of the system also contribute to the constant term.

Note that Eq. 2.28 is simply a convenient *parametrization* of the detector's performance and real-world devices may not follow it well. Substantial interplay arising from the calorimeter design may exist between each of the terms.

The stochastic term and the Fano factor

Note that statistical fluctuation may not only occur at showering, but also in the detection of the secondaries. For example, while the CMS ECAL is homogeneous, the energy is sampled in the sense that (a) the number of secondary photons *produced* by the incoming photon/electron is subject to statistical variation, and (b) the number of these photons which get *detected and converted* into photo-electrons is also subject to statistical variation. At 1 GeV, ~ 1400 secondary photons are produced,[4] so $1/\sqrt{N}$ indicates variation at the level of 2.7%. But the photodiode will only detect and convert approximately 1,000 of these photons to measurable photo-electrons, which increases this to 3.1% and, indeed, the overall crystal ECAL resolution is measured to be $\sim 3\%/\sqrt{E}$.

For completeness, we should also acknowledge the importance of the *Fano factor*, $\mathcal{F}$ [60]. The total number of secondaries, n, produced by some particle of energy E might be given by,

$$n = \frac{E}{W} \tag{2.29}$$

for W the mean energy required to produce the detected secondary. So the fractional error on the detected energy is,

$$\frac{\Delta E}{E} = \frac{\sqrt{n}}{n} = \sqrt{\frac{W}{E}}, \tag{2.30}$$

but because E doesn't fluctuate the resolution will be better than this. Furthermore, the energy loss in the collisions generating the secondaries is not driven by purely statistical effects. For example, the secondaries in the case of ionization are produced from atoms with *discrete* energy levels, so the secondaries' energies *are* correlated.[5] The Fano factor is dependent on the material in question, and included to account for this effect and the subsequent improvement to the resolution calculated on purely statistical arguments:

$$\frac{\Delta E}{E} = \sqrt{\frac{\mathcal{F} W}{E}}. \tag{2.31}$$

[4] From Table 2.1, page 48, $E_c = 0.7$ MeV

[5] The Fano factor can also be interpreted as how Poisson the processes are.

The importance of the Fano factor must be appreciated for high resolution calorimetry, for example, in the spectroscopy of low energy gamma rays. Theoretical estimates of $\mathcal{F}$ range from 0.08 to 0.13 [61].

Interaction between the stochastic term and sampling calorimeter design

The stochastic term also depends on the frequency of the sampling (or *granularity*) and the fraction of energy that is deposited in the active material. If the fraction of energy that is deposited in the active medium is small, then the energy resolution will be dominated by the variation in this fraction. If the energy loss in the active layers is small relative to that deposited by a MIP in the absorber layers ΔE_{abs}, then for each sandwich of absorber and active layer, the number of particles entering the active layer is approximately,

$$N \sim \frac{E}{\Delta E_{\text{abs}}}. \tag{2.32}$$

But E_{abs} is just the path length in the absorber t_{abs} multiplied by dE/dX,

$$\Delta E_{\text{abs}} = t_{\text{abs}} \frac{dE}{dX} \tag{2.33}$$

so the sampling error $b/E \sim 1/\sqrt{N}$ varies as

$$\frac{b}{E} \propto \frac{\sqrt{t_{\text{abs}}}}{\sqrt{E}}. \tag{2.34}$$

We would conclude from this that selecting thin absorbers would improve the energy resolution but, for detectors with thin passive layers or low density material, there are generally few collisions per unit length travelled in the medium. However some will carry a large energy transfer. Long Landau tails are the result at high energy losses. It is better therefore to construct a detector with thick layers and/or high density material, so many collisions result. The central limit theorem better applies (since we are taking the mean of a sample of collisions), and therefore the overall energy loss distribution is more Gaussian.

Note also that Eq. 2.34 is *not* valid if there are correlations in the amount of energy deposited in successive active layers—which there are. A more sophisticated analysis can give a generally valid result [42]:

$$\frac{b}{E} = \frac{5\%}{\sqrt{E}} (1 - f_{\text{samp}}) \Delta E_{\text{cell}}^{0.5(1-f_{\text{samp}})} \tag{2.35}$$

where ΔE_{cell} is the typical energy deposition in one absorber/active pair and f_{samp} is the *sampling fraction* representing the fraction of energy deposited in the active medium. As $f_{\text{samp}} \to 1$, the b-term (in principle) tends to zero. As the fraction tends to zero, the error goes as $\sqrt{\Delta E_{\text{cell}}}$.

The stochastic term is therefore reduced by (a) increasing f_{samp} and (b) increasing the sampling frequency, so ΔE_{cell} is reduced. The latter is particularly important for HCAL designs to avoid being susceptible to event-to-event fluctuations.

▷ 2.3 Implications for the CMS calorimeters

2.3.1 Specific design implications

Design & containment of hadronic cascades

Recall from Chapter 1 that the CMS ECAL is a homogeneous calorimeter, whereas the HCAL is a sampling calorimeter, instrumented with a cartridge brass absorber and plastic scintillator. The entire CMS HCAL system covers 11 λ_I for $|\eta| < 1.26$, and contains no fewer than 18 sampling layers in the barrel region, which reduces susceptibility to event-to-event fluctuations in f_{samp} in Eq. 2.35. Because showers are generally well-contained with the large number of interaction lengths provided, the contribution to the constant term, c, of the resolution is mitigated.

Containment of electromagnetic showers

The radiation length of CMS' lead tungstate crystals is 0.89 cm, so a single 23 cm crystal presents 25.8 X_0. Electromagnetic showers are therefore very well contained, which implies that the contribution to the constant term in the parametrization of resolution is small. R_M = 2.2 cm for $PbWO_4$ and the CMS ECAL crystals have a transverse extent approximately equal to this. This fine granularity is useful for separating showers from non-overlapping photons.

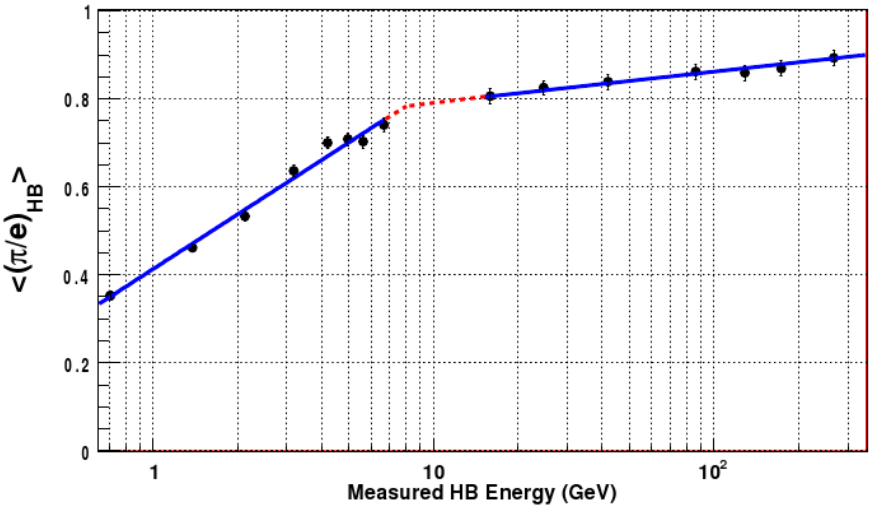
Material interactions in the tracker

There would ideally be zero interaction lengths' worth of material between the collision vertex and the calorimeter. At CMS however, the tracker presents $\mathcal{O}(1X_0)$ of material to electrons (see Figure 1.7, page 30), and on average 50% of electron energy is radiated in the tracker. Significant efforts are expended in the offline reconstruction to account for the 'brem photons' radiated by the hard electrons.

Non-compensation at CMS

Figure 2.13 exhibits this non-linearity characteristic of the CMS calorimetry. For this figure the HCAL was calibrated to electrons, where the response is generally linear. Compensating for this severe non-linearity is a non-trivial task at CMS, but there are good reasons for building a non-compensating calorimeter at the LHC, particularly since the calorimeter structure is simplified. The energy regime under investigation is much higher than that considered in experiments to date, so the variation that occurs in the electromagnetic fraction (for example) of a low energy hadron shower will not affect the reconstruction of TeV scale jets to any appreciable degree. Wigman's parametrization indicates $e/h = 1.4$ above 8 GeV. A function of the form $\pi/e = 0.179 \log E_{\text{HB}} + 0.413$ describes the data below 8 GeV, but the origin and interpretation of this result are currently not well understood. Groom's parametrization yields $e/h = 1.3$, but Wigmans' parametrization is generally preferred since it restricts π/e exceeding unity.

FIGURE 2.13 **π/e response for CMS HCAL Barrel**
After [36], "π/e vs $\log(E_{\mathrm{HB}})$ for HB [HCAL Barrel] only. The data are fit to two separate log functions with the break point at 8 GeV/c. Above 8 GeV/c the value is $e/h = 1.4$." This plot exhibits the non-linearity characteristic of the CMS calorimeter system.



2.3.2 The HF design

Recall from Chapter 1 that the forward HCAL calorimeter is a steel block instrumented with quartz fibres of two different lengths. In this way, the signal (arising from Čerenkov light generated in the fibre) has two components: the short fibres measure the cascades after 22 cm of steel which, on average, samples the hadronic component of a hadronic cascade, while the long fibres sample both the hadronic and electromagnetic components. While the system is *extremely* non-compensating with $e/h \sim 5$ (!), the electromagnetic and hadronic components can, in principle, be separated and re-weighted to improve the energy resolution.

2.3.3 Differences in response between pions, protons etc.

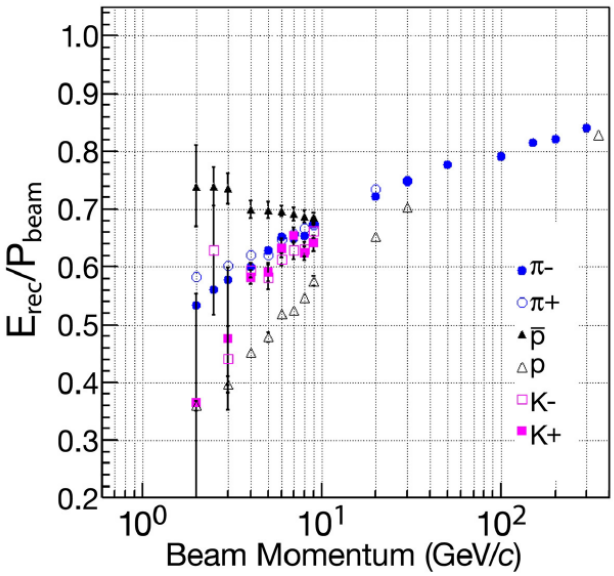
We can now anticipate that the response of the calorimeter to hadrons is different from the electron response: particles of different species present different amounts of energy available for showering and, subsequently, detection. For instance, for kaons, pions, and protons, the energy available to detect is the kinetic energy of the incident particle. For antiprotons, which will annihilate in the detection media, twice the rest mass m_p of the antiproton is also available for showering. For low energies, $E \sim m_p$, so this implies a substantial change to the response. So much is visible from Figure 2.14

We also observe a systematic increase in the response of π^+ to π^-: the *charge exchange* reactions are $\pi^+ + n \to \pi^o + p$ and $\pi^- + p \to \pi^o + n$ and they compete with nuclear breakup and spallation processes. The calorimeters contain some heavy nuclear species which contain ~ 50% more neutrons than protons, so the contribution of the charge exchange processes (which give an enhanced electromagnetic signal from the prompt π^o) to the overall response is slightly boosted for π^+s. Charge exchange is also more dominant at low energies, so the difference is exaggerated for lower beam momenta. The response for protons is slightly suppressed relative to pions because proton interactions will favour reactions with baryon number conservation. The charged pions, on the other

hand, readily produce neutral pions which decay electromagnetically with a subsequent increase in response in CMS' under-compensating HCAL.

FIGURE 2.14

CMS calorimeter response
From [36]: "The response of the combined calorimeter system to six different particles is shown as a function of the beam momentum. Both the EB [ECAL Barrel] and HB [HCAL Barrel] have been calibrated with 50 GeV electrons."



▷ 2.4 Implications for ILC calorimetry

A more complete introduction to the ILC will be made in Part III.

A detector for a future linear collider will have to distinguish between background events and final states of interest with a high multiplicity of jets. Furthermore, as we are concerned with investigating physics processes with cross sections of the order of femtobarns we require a detector with excellent jet energy resolution: every available channel must be exploited, including those with hadronic final states. For instance, while the Z in the 'Higgsstrahlung' process,

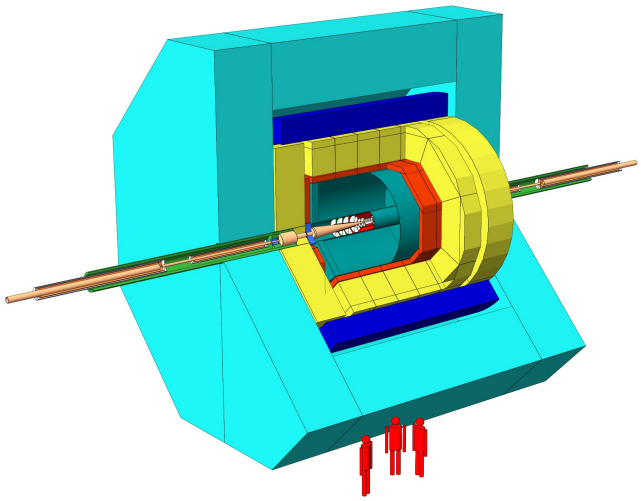
$$e^+e^- \to ZH \to f\bar{f}H, \tag{2.36}$$

can be cleanly reconstructed from its e^+e^-, $\mu^+\mu^-$ decays, the hadronic final states are more numerous.

The challenge is to build a detector that provides good jet reconstruction at a reasonable cost. The future linear collider community decided that particle flow is the best way to get both. The benchmark performance process is to separate the WW and ZZ events (further described in §7.3.2). This demands a jet energy resolution of $30\%/\sqrt{E}$, a factor of 2 better than was achieved at ALEPH [38].

One ILC concept detector design is presented in Figure 2.15. This design is used as the basis for simulation studies presented in this thesis.

FIGURE 2.15 **The Large Detector Concept design**
This design is similar to CMS, with a solenoidal magnetic field (dark blue) enclosing the calorimetry (ECAL in red, HCAL in yellow). The LDC has a gas tracker. The flux return yoke is shown in cyan. The LDC project merged with other similar concepts to form the *International Large Detector* [62] Concept in 2008.



2.4.1 The CALICE collaboration

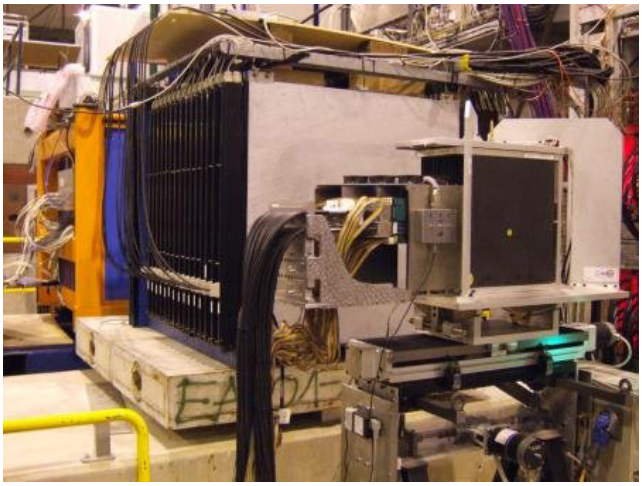
It has been proposed that the best way of making precision measurements in this multi-jet environment is to build a highly granular calorimeter optimized for a particle flow measurement of the jets in the final state: CALICE is an international collaboration studying how to build such a device. The essential observation to be made is that the overall attainable jet energy resolution is a convolution of detector performance and the performance of the available reconstruction algorithms. The jet energy resolution that can be attained is therefore *not necessarily* driven by minimizing the ECAL and HCAL energy resolutions.

A sampling calorimeter

The UK part of CALICE was primarily charged with the electromagnetic calorimetry. The baseline detector [62] (Figure 2.16) is of a sampling calorimeter design, composed of a sandwich of tungsten (absorber) and silicon (active layer): it has 30 layers overall and 24 radiation lengths in total depth (X_0 = 3.5 mm in Tungsten from Table 2.1, page 48). As an electromagnetic shower develops, the electrons and positrons in the shower pass through the silicon wafers between the tungsten sheets. Measuring the charge deposited by ionization in the silicon therefore provides a measurement of the shower energy at that point. Two technologies are proposed for the silicon layers: either (a) 0.5 cm^2 analogue readout silicon diode pads, where we consider the level of ionization in the pad to be proportional to the shower particle density, or (b) the monolithic active pixel sensor (Chapter 7), with pixels 50 × 50 μm^2 and digital readout. When a charged particle travels through the MAPS pixel, the pixel should detect the ionization that results and register a hit. Counting the hit pixels is therefore a measure of the shower energy.

A compact—and therefore cheaper—calorimeter requires an ECAL with a small radiation length and a large interaction length. This will ensure that the electromagnetic and hadronic components of a jet are well separated. Shower spread between layers

FIGURE 2.16 **The CALICE prototype**
The ECAL module forms the smaller box at the front of the assembly. The HCAL is behind this. The orange/blue construction at the rear is the 'tail-catcher/muon-tracker'. The ECAL front face is ~ 0.4 m square.



is reduced by keeping the depth of each silicon sheet small at ~ 500 μm. So while the fraction of energy sampled by the silicon, f_{samp}, (Eq. 2.35) is relatively small, which contributes to a larger stochastic term in the energy resolution, particle flow will be less confused by shower spread between layers. A similarly structured HCAL is essential for an effective particle flow reconstruction, especially for recognizing neutral hadrons. Ref. [62] is an excellent source of further information.

Is the Molière radius important?

The Molière radius in tungsten is 9 mm (see Table 2.1, page 48) and this sets the scale for the diode pad construction. The Molière radius is not a *directly* relevant parameter for the MAPS since the MAPS is a digital system where counting pixel hits from the secondary charged particles in a shower is the way to measure the original particle's energy. The scale of the MAPS pixel is therefore determined by the shower particle density.[6] In any case, maintaining a narrow Molière radius will allow more showers to be disentangled should they overlap.

▷ 2.5 Summary

This chapter covered:

- ▷ The physics governing the interaction of particles with calorimetric media;
- ▷ The essential mechanics of electromagnetic showers and hadronic cascades;
- ▷ The implications from a physics perspective of having a non-compensating calorimeter;
- ▷ The physics contributing to energy resolution;
- ▷ The specific implications for CMS and ILC calorimeters.

[6] Indirectly related to the Molière radius, however!

Part II

Particle flow at CMS

Chapter 3

The CMS calorimeters in testbeam

▷ 3.1 Introduction

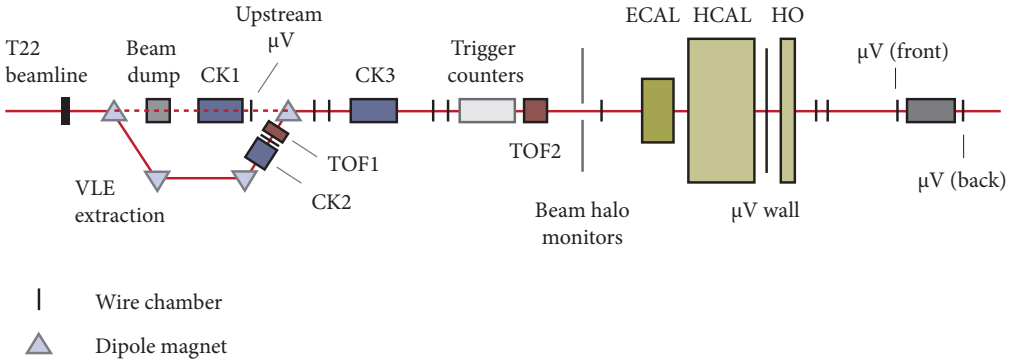
The relevant principles of calorimetry which affect the design and performance of the CMS calorimetry system have been presented. In this chapter, two CMS testbeam experiments conducted in 2006 and 2007 are described. These experiments sought to characterize the response of the CMS ECAL and HCAL to hadrons. A description of the CMS simulations is given and we consider the differences between the simulations, the testbeam experiments, and the CMS collision environment. This is followed by an analysis of the CMS calorimeters' response to hadrons both in data and simulation, in preparation for applying the particle flow reconstruction in the next chapter.

▷ 3.2 Description of the testbeam

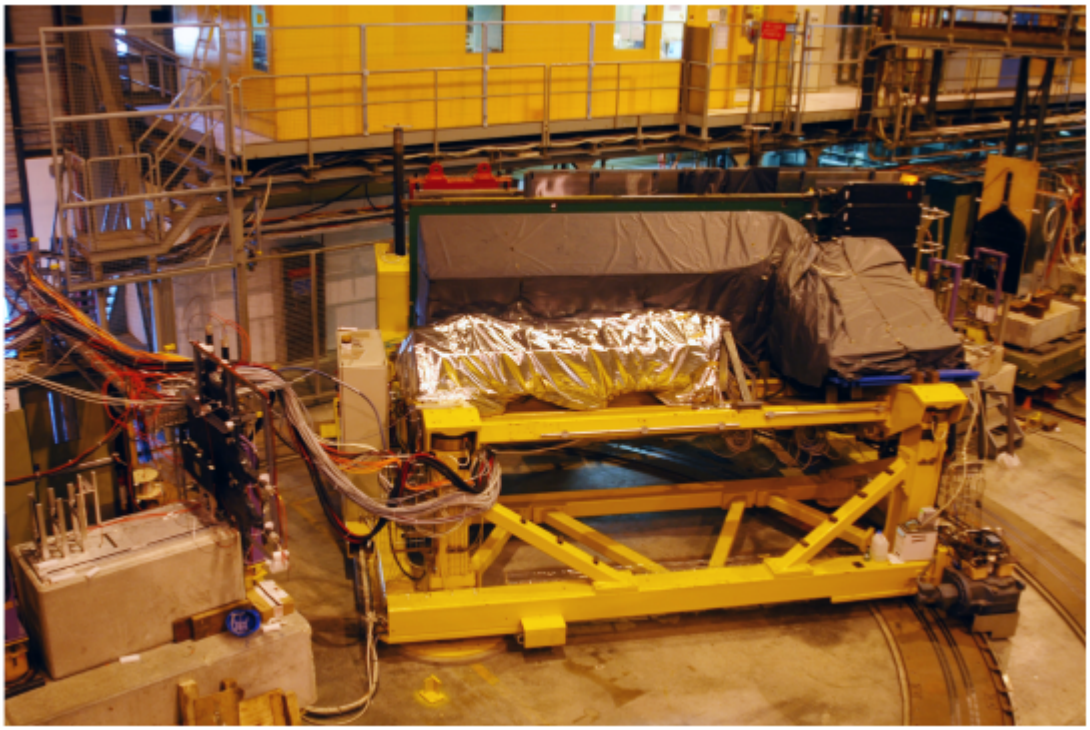
The dataset considered in this chapter was analysed and described in detail in the paper [36] but a brief synopsis of the experiment is presented here for the reader's convenience. Complete wedges of each of the ECAL and HCAL barrel calorimeters, each covering $\Delta\phi = 20°$ and up to $\eta = 1.3$, were fixed to a movable stage at the H2 beamline at CERN. Figure 3.1 shows a schematic of the testbeam. In 2007, prototype HCAL endcap modules were placed in the beamline, corresponding to towers 17 to 21 in Figure 1.9. Towers 17 to 20 have the same $\eta\phi$ granularity as the barrel, while tower 21 has $\Delta\eta = 0.09$ and $\Delta\phi = 10°$. The beam was positioned at the centre of towers 19 and 20.

Particles were provided in the momentum range 2 to 350 GeV/c by the CERN Super Proton Synchrotron: the accelerated protons impinge on a target, producing the required secondary particles. This is followed by a particle momentum spectrometer and then species selection and identification. The desired particles were directed onto the calorimeter wedges in the same manner as in CMS collisions — that is, from a nominal vertex of $(x, y, z) = (0, 0, 0)$, but in the absence of a magnetic field and tracker.

FIGURE 3.1 **The H2 testbeam environment in 2006**



(a) H2 beamline apparatus (not to scale). The various Čerenkov counters (CK1, CK2, CK3) and time of flight (TOF) chambers are shown. Muon detectors are indicated with labels 'μV'.



(b) The detector modules on a movable stage. The ECAL module is under a foil blanket, and the HCAL is behind the ECAL under a black blanket. The beam enters at the bottom left. For scale, the stage is approximately 4 m wide.

The outer HCAL calorimeter (HO) was also present but has not been considered in this analysis. The setup was identical for the endcap testbeam experiment (2007) but a section of the preshower ECAL calorimeter was also in place. We do not use the preshower in this chapter.

3.2.1 Beam cleaning & pion selection

The dataset used follows elementary reconstruction by the HCAL team. The HCAL response is calibrated to 50 GeV π^+ depositing a MIP-equivalent in the ECAL, and considers the sum of energy deposited in an area of 3×3 HCAL towers. The ECAL is separately calibrated using an electron beam. The dataset contains all triggered events, so it must be cleaned to isolate charged pions from electrons, muons, and protons or kaons. There may also be spurious beam triggers to remove. Finally, we require single isolated beam particles: particles with a beam halo are removed by means of beam halo discriminators, and triggers with more than one MIP present are removed using scintillation counters. Pions are disentangled from other species using a combination of Čerenkov counters and time of flight (TOF) chambers. Muons are removed using muon vetos surrounding the calorimeters. The specific cut values are detailed in Appendix A.2.

TABLE 3.1

Beam purity and statistics
Fraction of the beam determined to contain single isolated pions (barrel, 2006), determined by the author and according to [36], and the number of events passing similar selections for the endcap dataset (2007). The beam purity for the endcap dataset was very similar. The beam is relatively pure from 20 GeV; only muons and double-particle vetos are applied.

p_{beam}	π **fraction (%) (2006)**		**Clean π events**	
(GeV/c)	*Ego scriptor*	Ref [36]	Barrel	Endcap
2	10	10	10, 763	—(1)
3	24	26	23, 113	164(2)
4	40	40	26, 365	1, 559
5	52	53	40, 243	2, 217
6	55	58	39, 880	3, 465
7	60	62	35, 134	12, 558
8	55	57	18, 985	25, 629
9	70	73	27, 765	21, 205
20			14, 969	69, 819
30			7, 471	55, 740
50			12, 559	55, 541
100			14, 518	30, 860
150			10, 657	55, 235
200/225(3)			9, 420	53, 287
300			27, 768	54, 752
Σ			319, 610	442, 031

Notes: (1) No data available at this energy. (2) The beam quality cuts were tighter in 2007, but this had the adverse effect of removing many events at low energy. (3) Data were taken at 225 GeV/c in 2007.

These independently implemented cuts reproduce the values quoted in [36], and are presented for comparison in Table 3.1. The purity of the pion beam increases with energy. The slight variations between the values are due to the choice of a stabilized Gaussian fit (see Appendix §A.4) used to evaluate the pedestal for the beam halo counters on a run-by-run basis.

▷ 3.3 The simulations

3.3.1 Full simulation

A full simulation [63] of the CMS experiment is available in the *CMS Software* (CMSSW) software framework. It uses the GEANT4 simulation toolkit [64] and a number of similar packages to simulate beam transport and particular detector effects. The simulation covers all physical regions of the detector, and models the magnetic field's effect on the detector's response. The effect of dead channels and poorly calibrated components is introduced via a 'conditions' database. The full simulation is estimated to contain some 1.3 million geometrical volumes and uses over 400 materials in GEANT4's database.

A particle gun or a full LHC parton-parton hard scatter is generated in a dedicated tool such as PYTHIA [65]. The full simulation takes these particles and transports them through the detector on a 'delta time' approach. The original particles are decayed according to their known branching ratios and kinematics. At each step, the interaction of particles with the detector material is computed on this minute basis, exploiting models derived from the physics outlined in the first part of this chapter. Consequently, the simulation is *extremely* detailed, including detector noise, Landau fluctuations, jitter and cross-talk between channels, and the digitization steps applied by each sub-detector's electronics.

Both in-time and out-of-time pile-up events can be randomly sampled from pre-generated source files according to a user-specified Poisson distribution (*i.e.* $\mathcal{O}(20)$ events per bunch crossing at LHC full luminosity) and mixed in with the primary event. A full emulation of the Level-1 and high level triggers is included.

Many aspects of the simulation have been tuned to experimental data, but there are, however, difficult challenges in modelling such highly stochastic phenomena as hadron cascades, for reasons alluded to earlier: a number of models for simulating hadronic cascades is available (the so-called 'physics lists' of the *hadronic physics parametric model* LHEP and the *microscopic models* QGSP, QGSP_BERT [66]).

3.3.2 Fast simulation

The fast simulation [26, 67] takes the same generated particles as the full simulation and decays them in a similar way. It too provides an emulation of the Level-1 and high level triggers and pile-up. The fast simulation exploits a simplified geometrical description of CMS, and makes extensive use of parametrized physics processes to gain a speed increase of $\mathcal{O}(\times 100\text{–}1000)$ per event over the full simulation. It proves extremely useful in studying background processes and systematic errors and will likely be of high value when characterizing the detector's response in early data as quick turnarounds will be essential. Unlike some other 'fast' simulations, the CMS fast simulation aims to provide data of suitable quality for analysis.

The fast simulation considers a subset of the possible interactions with matter a particle may have, namely (a) bremsstrahlung, (b) photon conversion, (c) ionization and multiple scattering, and (d) showering. The first 3 are considered for particles traversing the tracking volume. The last one applies to interactions in the ECAL and HCAL: in this respect, the fast simulation takes a major time-saving step, whereby the showers are

simulated based on parametrized shower characteristics, using information extracted from testbeams and the full simulation.

Finally, the fast simulation permits a physicist to generate several thousand particle gun events in a matter of minutes. This is *invaluable* to reconstruction groups: developers can rapidly improve and understand their code. For studies of *Super-LHC* [68], where $\mathcal{O}(400)$ minimum bias events are expected per crossing, the fast simulation is currently the *only* tool available for characterizing new tracking-trigger models [69].

3.3.3 Reconstruction

The full simulation generally proceeds in two distinct steps: (1) depositing energies, charges, hits, *etc.* in subdetectors that are (2) subsequently digitized. The fast simulation proceeds straight to digitization—thus saving time—though users may request the digitization be unwound according to some approximations.

Following digitization, the reconstruction is *identical* for both simulations' output and collisions. Analyses can therefore be agnostic of the data's provenance.

▷ 3.4 Differences between CMS, testbeam, and simulations

For technical reasons, there are some problems in providing a direct comparison between the testbeam data, simulations, and CMS operating in collisions. Specific controls applied to the simulation will be given in §3.5.5.

Limitations to the testbeam datasets

The 2006 & 2007 datasets only characterized samples of the ECAL and HCAL modules. In particular we have only one η point for each of the barrel and endcap experiments. Calibration procedures [70] exist to extrapolate the performance of ECAL and HCAL components tested at the beamline to those in other modules, towers, η and ϕ, *etc.*

Material effects

The testbeam environment directs particles onto the ECAL front surface without them passing through any tracking material (with the exception of thin scintillators and air). At CMS we have $0.3\,X_0$ and $0.1\,\lambda_I$ (Figure 1.7, 30) of tracker material at $\eta = 0$ so the probability of early showering/interactions is non-negligible. Also, material between the ECAL and HCAL (for services and readout) increases substantially with $|\eta|$ in the barrel; this will modulate the HCAL response as a function of η.

Magnetic field

It is known that the plastic scintillators of the HCAL have an increased light yield in the presence of a strong magnetic field, such as that found at CMS, but the mechanisms are not well understood [43]. This *scintillator brightening* is a non-linear effect and, while no magnetic field was applied at the testbeam setups, neither the fast nor full simulation models this in CMS running with collisions at 4T. Both the fast and full simulations however account for the deflection of charged particles that result from shower development: we can expect an increased signal (energy) with the field on

because the trajectories taken by the secondaries will generally be curved, and so they will pass through more scintillating material. For the simulation studies presented in this chapter, the magnetic field is set to 0T. With the field set at 4T, the magnitude of the scintillator brightening effect is such that the response increases $\mathcal{O}(< 4\%)$ at high energy in the barrel, and $\mathcal{O}(< 8\%)$ in the endcaps [26].

Zero Suppression and Selective Readout

These two modes will operate during CMS running with collisions. The specifics of these modes are complex [26] but, to summarize, both drastically reduce the bandwidth required by the calorimeters during normal running. *Zero Suppression* (ZSP) implies HCAL towers are not read out if their signals are less than that channel's calibrated pedestal value. This means 'negative' energies (due to downward noise fluctuations) do not contribute to the sum of energies found in a given region of the calorimeter, and a positive bias results which is exaggerated for lower energies $\mathcal{O}$(a few GeV).

Selective Readout (SR) means only a limited region of the ECAL is read out where the Level-1 Trigger has determined some activity to have taken place. This however does not, on average, result in a bias to the total energy measurement from the calorimeter. A brief example of selective readout effects is shown in Appendix A.3.

In particle flow we expect to be immune from these effects: we calibrate the energy response using simulation and/or an *in situ* data-driven method where ZSP and SR are applied. In any case, PF applies energy thresholds to each calorimeter channel to reduce noise contributions, so the effect of any bias from these two modes is therefore diminished.

▷ 3.5 Calorimeter energy response

3.5.1 Calo rechits

Calo 'rechits' are the basic reconstruction objects. They simply represent some amount of energy collected in a given calorimeter channel. For the ECAL, rechits are the result of the standard ECAL reconstruction (where the ECAL has been calibrated to electrons depositing energy in an array of 7×7 crystals) and rechits from the HCAL are calibrated to a 50 GeV pion not showering in the ECAL, collecting energy in a region defined by 3×3 HCAL cells in the barrel, and 4×4 cells in the endcap in view of the finer granularity at higher η. No threshold is applied to their energies; only the pedestals have been subtracted.

Energy collected in $\Delta\mathcal{R}$ and comparison with previous definitions

So that we may eventually compare calibrations from testbeam with those from collisions, we have chosen to select hits from a cone around the particle's impact point, rather than in a fixed array of crystal or cell size as this approach is independent of the geometry. This cone size is effectively the same as collecting hits from the HCAL in 3×3 cells at $\eta = 0$. From the material presented in the first part of this thesis—specifically §2.2.4—we know that the interaction length approximates the lateral extent of a hadronic shower. The cartridge brass of CMS' HCAL has an interaction length of $\lambda_I = 16.4$ cm. One HCAL

cell is ~ 15 cm.[1] A hadronic shower therefore has an equivalent transverse extent of $\Delta R \sim 0.09$. Collecting energies from a 3×3 matrix of HCAL cells (in the barrel) or from a cone of size $\Delta\mathcal{R} = 0.15$ (cone half-angle) therefore comfortably contains a hadronic shower. *Be advised that the calo rechit comparisons presented in this document refer to the sum of ECAL and HCAL rechit energies in a cone of $\Delta\mathcal{R} = 0.15$ around the track impact point.*

Testbeam peculiarities

Due to miscabling at the experiment setup, the translation of some HCAL rechit IDs to geometrical η, ϕ is incorrect. A patch is applied to fix this at all places where standard rechits are used or referred to forthwith.

For the ECAL, each crystal channel provides a signal amplitude. This is converted to an energy by first removing the channel's pedestal value, followed by application of a gain factor also specific to the channel. In the barrel experiment, rechits were already available where this process had been applied by ECAL experts. For the endcap experiment, calibrated rechits were created using pedestal and gain values supplied by [71]. Rechits were also translated in $\eta\phi$-space so the beam impact point, and ECAL and HCAL beamspots were all aligned.

Finally, despite the standard CMS definition of calibrating the HCAL to 50 GeV pions, the 2007 endcap dataset had the HCAL calibrated to electrons. The conversion to a pion calibration is straightforward. From Figure 2.13, $\pi/e = 0.85$ at 50 GeV for the HCAL barrel. For the endcaps, this is reported to be 0.88 [72]. The HCAL endcap module was, unfortunately, incomplete and some transverse leakage $\mathcal{O}(< 10\%)$ of hadronic cascades was observed. The Author has determined that scaling all rechits by $\pi/e = 0.85$ is sufficient to recover a rechit response of 50 GeV to 50 GeV pions; this will be demonstrated in the following sections.

3.5.2 HCAL noise

Events where no particles were incident on the experimental setup were selected to acquire a noise sample. To estimate the HCAL noise, we plot the *individual* rechit energies *not summing them in* 3×3 *towers,* $\Delta\mathcal{R}$ *or otherwise*. Individual HCAL rechits (with no ZSP applied, as in all testbeam data) have had their pedestals subtracted. The rechits for many noise events and many towers are histogrammed (Figure 3.2); the noise values are:

Barrel $\sigma = 200$ MeV **Endcap** $\sigma = 350$ MeV

The simulations have been tuned to provide similar values. These values are used in the next chapter to set the seed and cell threshold for creating and growing PF clusters.

3.5.3 ECAL Activity

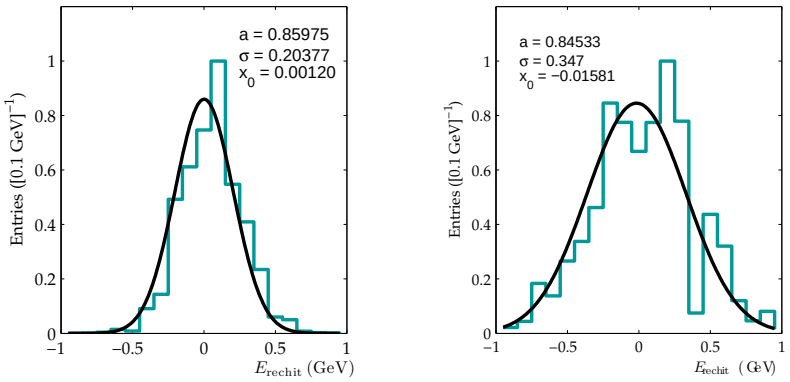
Definition of 'MIP' in ECAL

In anticipation of the different π/e responses for the ECAL and HCAL, it will prove useful to disentangle pions into two classes:

MIP in ECAL: refers to pions that behave as minimum ionizing particles in the ECAL and don't start showering there.

[1] $\delta\phi = 5^\circ \Rightarrow 15.5$ mm at the inner barrel radius of 1777 mm.

FIGURE 3.2 **HCAL noise distributions for standard rechits**
The histograms have been normalized to their peak values. As in many other plots in this thesis where a Gaussian of the form $a \exp\left(\frac{-(x-x_0)^2}{2\sigma^2}\right)$ has been fitted, the corresponding fit parameters are shown in the diagram.



(a) Barrel σ = 200 MeV *(b) Endcap σ = 350 MeV*

Interacting in ECAL: refers to pions that start showering in the ECAL.

In Figure 3.3, the sum of ECAL rechit energies collected in $\Delta\mathcal{R} = 0.15$ around the particle impact point is presented. Two classes of pions are evident and we can identify the MIP in ECAL class as depositing 450 MeV in the barrel. A combination of intrinsic physics processes (multiple Coulomb scattering, bremsstrahlung) and extrinsic detector effects (noise, miscalibration) smears this value, giving rise to the substantial Gaussian width of 860 MeV.

The original testbeam analysis [36] places a MIP cut at 1.2 GeV. The results presented here are in agreement with the original analysis, but the Author believes this cut to be too aggressive and has revised its definition to a more conservative 2 GeV. As the simulations reproduce the data well (in the barrel at least), this should have no effect in the comparisons that follow. Secondly, for the purposes of particle flow reconstruction and this thesis, the cut is not used explicitly.

On pions interacting in the ECAL

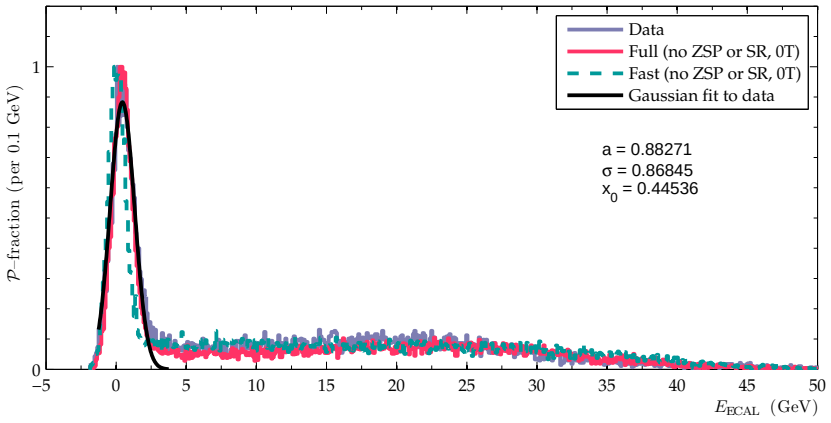
Consider the class of pions interacting in the ECAL in Figure 3.3: for the barrel, there is a broad peak of energies peaking at ~ 20 GeV and extending all the way to the full pion energy of 50 GeV. The simulations and data agree well. For the endcap however, there is reasonable agreement between the fast simulation and the full simulation with a broad peak at ~ 22 GeV (recall that the fast simulation is in many respects tuned to the full simulation), but the simulations do not reproduce the data well which peaks at ~ 15 GeV: it is understood that the model of hadronic interactions used by GEANT4 (the 'physics lists' described above) would benefit from some optimization, and work is underway to this effect.

3.5.4 Standard response

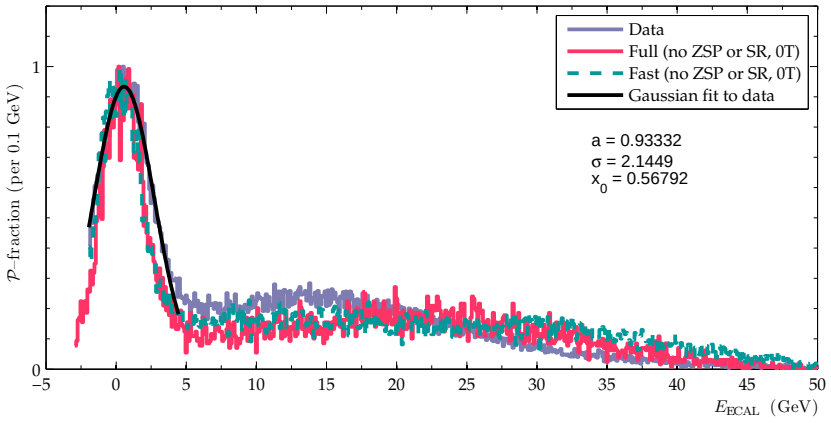
The raw rechit energy distributions for each available beam momentum in the 2006 barrel experiment are presented in Figure 3.4, and for the 2007 endcap experiment in

FIGURE 3.3

ECAL activity for 50 GeV pions
ZSP and SR have been disabled. The Gaussian fit to the portion of the data curve representing pions depositing a MIP in the ECAL indicates a cut for the barrel systems should be placed at $x_o + 2\sigma \sim 2$ GeV, and 3 GeV for the endcaps. The histograms have each been normalized to their peak value (hence '$\mathcal{P}$-fraction' on the y-axis).



(a) Barrel



(b) Endcap

Figure 3.5. The x-axis is labelled by the response R where,

$$R \equiv \frac{E_{\text{reco}}}{E_{\text{true}}} = \frac{\sum_{\Delta\mathcal{R}=0.15} \left[E_{\text{rechits}}\right]}{E_{\text{true}}}, \tag{3.1}$$

and the rechits are taken from both the ECAL and HCAL.

In these plots, the class 'MIP in ECAL' is shown. The other class 'all pions', is just that: One could plot the class 'interacting in ECAL', *i.e.* the subset of events excluding 'MIP in ECAL events', but at very low energies this introduces a substantial bias so it has been decided that showing the fully inclusive case is more enlightening.

These data are summarized in Figure 3.6, where the fitted Gaussian mean has been plotted as a function of the beam momentum. The error bars refer to the error on the fitted mean given by ROOT/MINUIT [73,74]. At 50 GeV, the HCAL is indeed calibrated to give a response of 50 GeV, for a pion depositing a MIP in the ECAL.

Interpretation

From these spectra, we observe values of $E_{\text{reco}}/E_{\text{true}} < 0$. The ECAL and HCAL are running without ZSP and SR, so for each beam trigger all channels are readout regardless of their values. Each channel has had its pedestal subtracted, so any downward noise fluctuation may result in a negative signal relative to the pedestal value, which therefore appears as 'negative' energy. This effect becomes increasingly dominant at low energies (2, 3, … GeV).

Second, we observe that the distributions are broadly Gaussian though with some interesting features. At 2 GeV, the signal barely exits the noise. At intermediate energies of 3 – 8 GeV, the effect of non-compensation becomes apparent for both the 'all pions' and MIP in ECAL classes because we observe substantial asymmetry in the distributions.

Third, all the distributions have substantial widths. At low energies, one can estimate the resolution to be of $\mathcal{O}(100\%/\sqrt{E})$. A further analysis of the resolution will be made in the next chapter.

Fourth, for the 'all pions' class, the response is generally lower than for the MIP in ECAL class, and a Landau tail is still visible at 20 GeV (b) of Figure 3.4. This would indicate that the ECAL has a lower π/e response (Eq. 2.25) than the HCAL, for if it were the same then the responses would be equal for the two cases, assuming F_0 was, on average, also equal. We miss more energy from hadronic cascades interacting in the ECAL so the overall response is diminished.

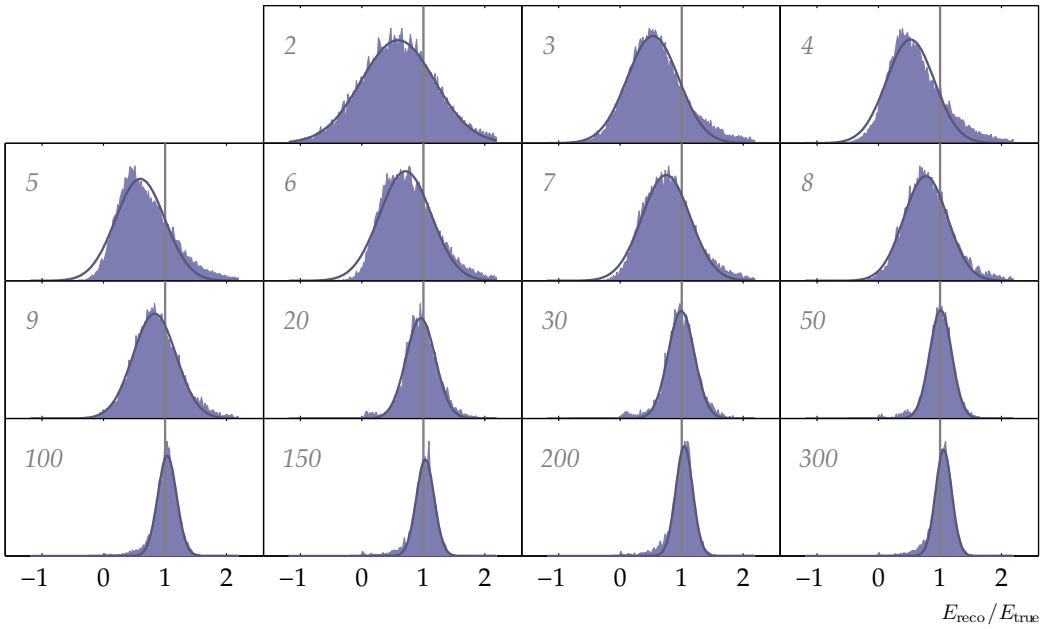
Given that π/e is diminished for the ECAL, and the electromagnetic fraction F_0 is the same for both ECAL and HCAL, then Eq. 2.25 implies that e/h must be higher for the ECAL. We conclude that the ECAL is even more undercompensating than the HCAL.

3.5.5 Comparison with simulations

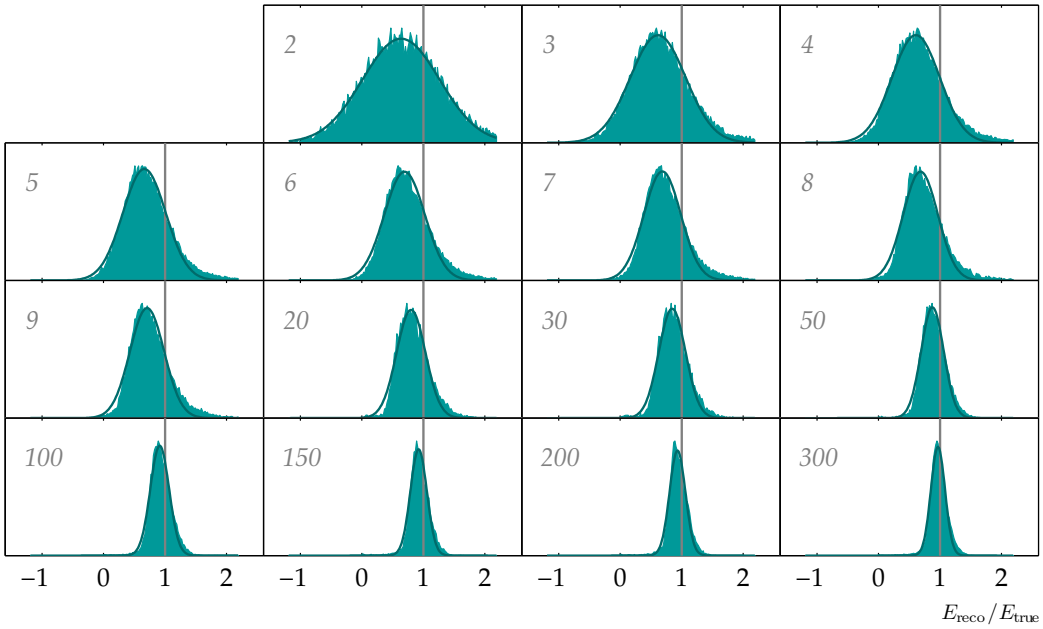
Prescription

Subsection 3.4 listed the differences between the testbeam and the collision environment of CMS. While the simulations are designed to emulate the latter, the following effects have been enabled/disabled to facilitate a fair test:

FIGURE 3.4 **Barrel energy spectra**
The x-axis corresponds to Eq. 3.1. Each histogram is labelled with the beam momentum (in units of GeV/c), and normalized to a peak value of unity.

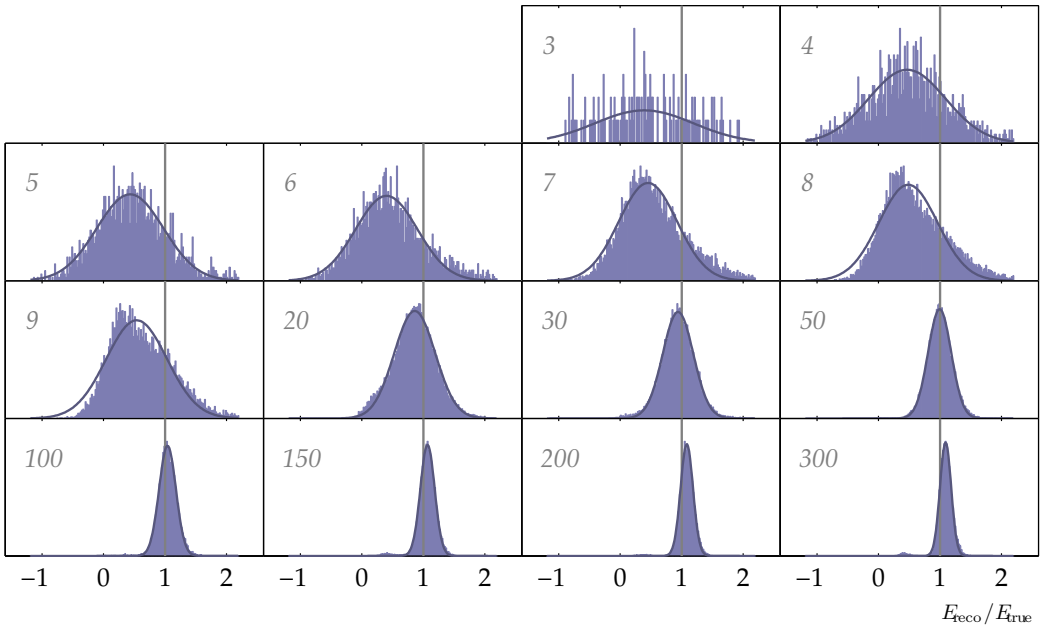


(a) MIP *in* ECAL

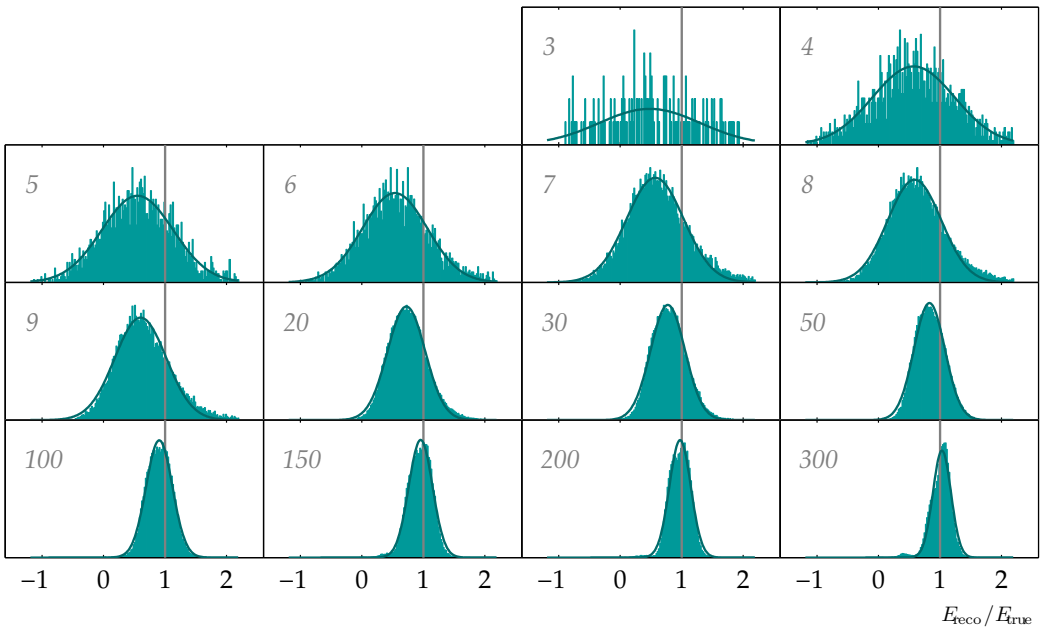


(b) All pions

FIGURE 3.5 **Endcap energy spectra**
The x-axis corresponds to Eq. 3.1. Each histogram is labelled with the beam momentum (in units of GeV/c), and normalized to a peak value of unity.

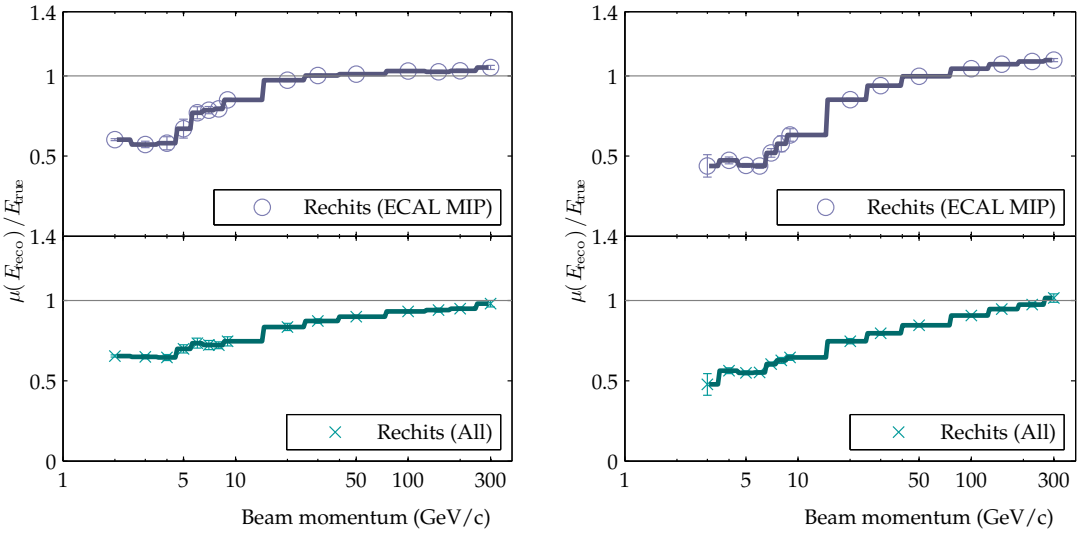


(a) MIP in ECAL



(b) All pions

FIGURE 3.6 **Summary of rechit spectra**
Fitted Gaussian means for Figures 3.4 and 3.5 as a function of beam momentum. The error bars are the errors on the fitted means.



(a) Barrel *(b) Endcap*

(i) The magnetic field is set to 0T.

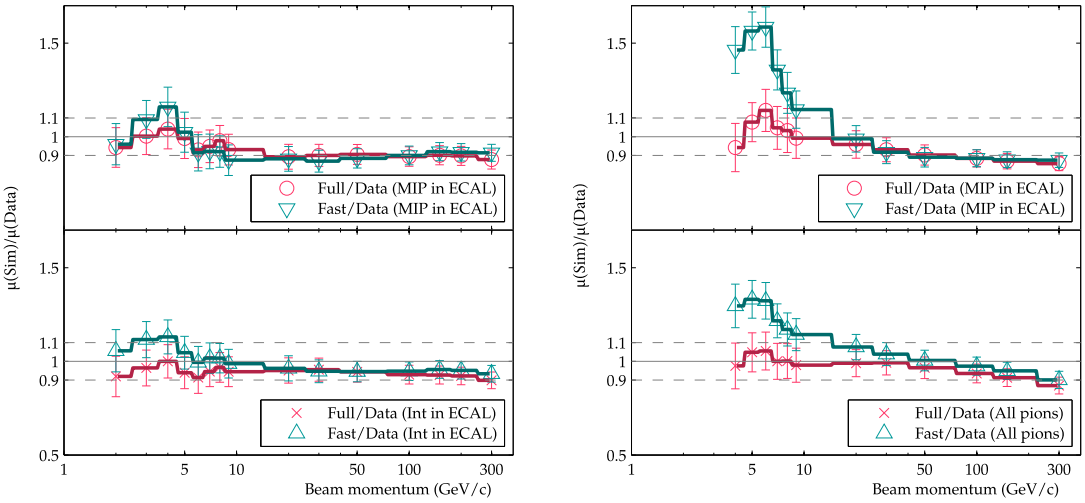
(ii) In the fast simulation, the material effect of the tracker and subsequent nuclear interactions is disabled. Pions therefore behave as though the tracker were not present – *i.e.* like the testbeam environment. In the full simulation it is not straightforward to do this, but events are rejected if the primary Monte Carlo particle has daughter particles when it reaches the ECAL front surface. If daughters are present then we conclude the particle interacted in the tracking volume. This also rejects the occasional $\pi \to \mu + \nu$ decay.

(iii) The effects of ZSP and SR are disabled.

(iv) The particles are directed to the same η point as in the testbeams (η = 0.2 in the barrel and 1.6 in the endcap experiments), but sprayed over all ϕ angles. No ϕ variation is expected. Since there was no accurate tracking at the testbeam setup, the beam's impact point has an uncertainty of ± 2 cm in x, y, which is approximately the same size as an ECAL crystal. By spraying the simulated particles over all ϕ we avoid resonating with a crack or specific material effect.

Ten thousand particles are generated at each beam momentum point for both simulations. The rechits are collected in the same cone as for the testbeam analysis. The resultant spectra are similar to those shown above (*i.e.* broadly Gaussian) but there is some variation between the means of the simulations and those of the data; the results are presented in Figure 3.7, which plots the ratio of the fitted means *to the data* (NOT to the beam momentum).

FIGURE 3.7 **Simulation/data comparison**
Ratio of fitted means (μ) of total rechit energy between simulation and data (barrel). For a given error ϵ on the fitted mean, the error bars shown are $\sqrt{\epsilon_{\text{sim}}^2 + \epsilon_{\text{data}}^2}$.



(a) Barrel *(b) Endcap*

Interpretation

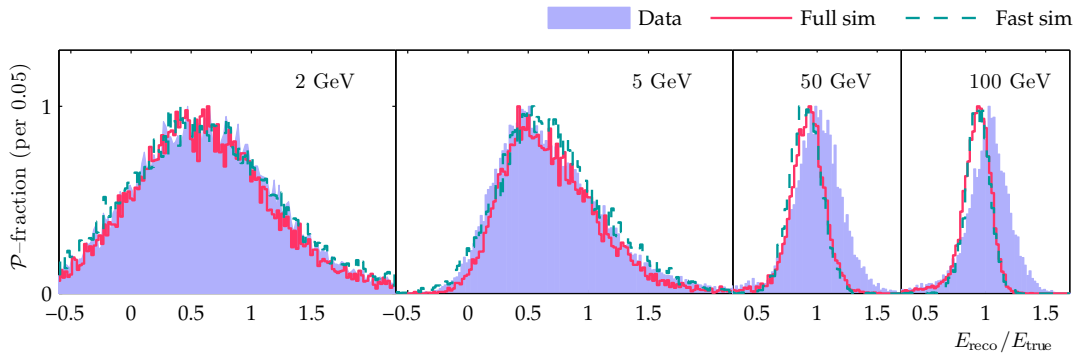
First, we assume that the beam is purely composed of pions following the selections detailed in §3.2.1. We conclude from Figure 3.7 that in the barrel, both simulations underestimate the response by $\mathcal{O}(10\%)$ for $p > 10\,\text{GeV}/c$. The simulations are tuned to reproduce a 50 GeV response for a 50 GeV pion but only when *all* HCAL energy depositions are included in the energy sum. The reasons for this are involved and technical [75,76], but amount to the issue of *sampling factors* in scaling the simulated energy *sampled* in the HCAL scintillators to a simulated *deposited* energy. In the endcaps, for $p > 20\,\text{GeV}/c$ the simulations underestimate the response in data by $\mathcal{O}(10\%)$ for the same reason as the barrel. For $p < 20\,\text{GeV}/c$, the full and fast simulations disagree, with the fast simulation overestimating the response by up to 50%.

Sample energy distributions

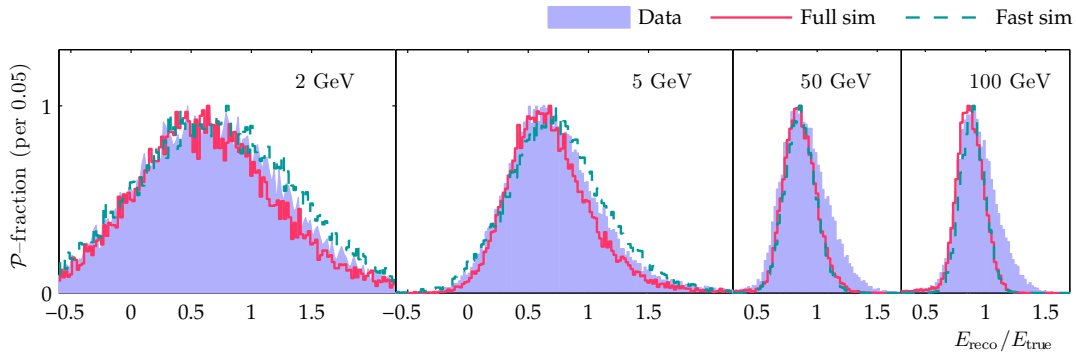
To understand the variation of the simulation with respect to the data exposed in Figure 3.7, we should examine a few energy points in detail. Figure 3.8 displays the raw histograms for 2, 5, 50, and 100 GeV for the data and simulations for the barrel.

From these spectra, we can conclude that the simulations do not reproduce the data well when the $\Delta\mathcal{R} = 0.15$ calibration cone definition is used at 50 GeV. The full simulation response is in very good agreement however at low energies. The fast simulation's endcap response requires further tuning, as exhibited in Figure 3.9. The Author has found that the full simulation does reproduce a 50 GeV response to a 50 GeV pion, but only when the rechit collection cone is enlarged far beyond $\Delta\mathcal{R} = 0.15$, in accordance with expectations—but not desires—attributable to the sampling factors in the simulation of the HCAL response. It is the view of the Author and his colleagues that the simulations should be modified to reproduce the data, and reproduce the containment of hadronic

FIGURE 3.8 **Sample energy distributions**
Comparing data and simulation at a selection of energies in the barrel.



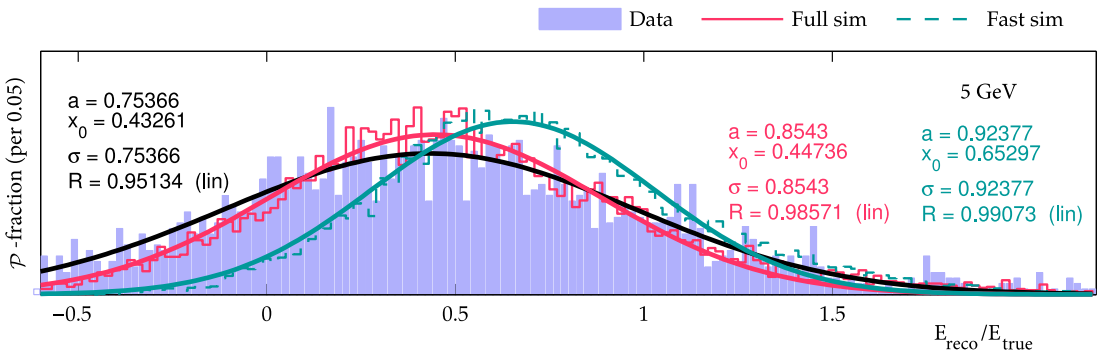
(a) MIP in ECAL



(b) All pions

showers seen in the data.

FIGURE 3.9 **Sample energy distributions II**
Comparison of data and simulation for 5 GeV pions depositing less than 3 GeV (MIP value) in the ECAL endcap.



▷ 3.6 Summary

This chapter put the introductory material given in the Part I of this thesis into practice. Aside from providing an independent analysis of the work presented in [36]:

- ▷ A systematic comparison between barrel and endcaps has been shown. We have also recognized the difference in response between pions that start showering in the ECAL and those that do not;
- ▷ An original comparison with the two simulations was also presented; the simulations underestimate the calorimeters' response for pion energies > 20 GeV.

Chapter 4

The particle flow reconstruction of hadrons

▷ 4.1 Introduction

CMS' undercompensating and non-linear calorimeters must be calibrated to recover a linear response for single isolated hadrons. Furthermore, many analyses will require accurate reconstruction of jets. The particle flow method, introduced in Chapter 1, seeks to reconstruct the original parton-parton collision in terms of electrons, muons, photons, taus and hadrons, the last of which are the focus of this chapter. A particle flow-specific calibration of hadrons is evaluated using the testbeam data and compared with simulations. We estimate the systematic uncertainty affecting PF jet energy response and resolution resulting from an inappropriate hadron calibration.

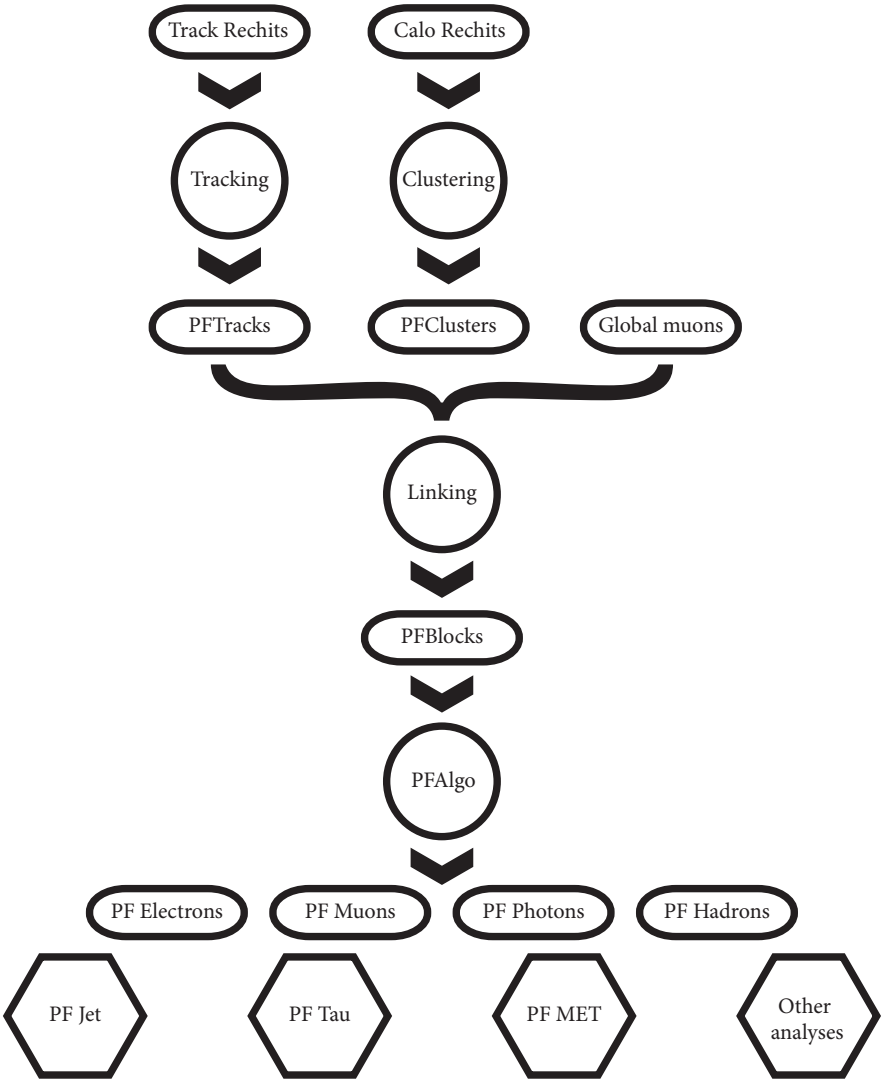
▷ 4.2 Elements of the particle flow reconstruction

The recently published *CMS Analysis Note* [41] describes the particle flow algorithm in detail and also describes the anticipated performance of the reconstruction method based on Monte Carlo experiments. Here we start with the essential building blocks of the algorithm, namely tracks, clusters and blocks. A high-level overview of the PF reconstruction is presented in Figure 4.1.

4.2.1 Tracking

Approximately two-thirds of the particles produced in collisions are charged (Figure 1.11), so it is imperative that these particles are correctly reconstructed. The CMS tracker's resolution outperforms the calorimeters for charged hadrons up to transverse momenta

FIGURE 4.1 **Particle flow reconstruction**
Various terms used in the diagram are explained in the text.



(p_T) of $\mathcal{O}(300\,\text{GeV}/c)$. It also gives the direction of the charged particles emanating from the collision vertex. It is therefore heavily relied upon by particle flow.

The track reconstruction must be extremely efficient and offer very high purity. Low momentum hadrons $\mathcal{O}(< 10\,\text{GeV}/c)$ will not be well reconstructed based on raw calorimetric information alone—so much is obvious from Chapter 3—so it is essential to absolve ourselves as much as possible from the calorimetry. But if so much responsibility is to be assigned to the tracker then the reconstruction must be very pure because fake tracks could have a degrading effect on the overall outcome [41].

Particle flow uses an iterative tracking strategy [77] that creates tracks using very strict quality criteria which are subsequently loosened to increase efficiency while maintaining a negligible fake rate. The net result is that particles with momentum as low as 150 MeV/c, and created as far as 50 cm from the beam axis are reconstructed with a fake contamination rate as low as $\mathcal{O}(1\%)$. Isolated muons inside the tracker acceptance are reconstructed with an efficiency $\mathcal{O}(99.5\%)$.

Electrons and muons

As alluded to in Chapter 2, electrons radiate 50% of their energy in the tracker, so much effort is devoted to ensuring these electrons are accurately and precisely reconstructed. Muons receive a specialized treatment combining information from the tracking and muon systems, giving rise to the jargon of 'global muons' [78]. These objects form another ingredient of the particle flow reconstruction. These are not however the main subject of this thesis and will not be discussed further.

4.2.2 Clustering

Clustering recognizes the individual energy depositions from charged and neutral particles. By exploiting the high granularity of the calorimeters, particle flow can separate energy depositions that are close together ($\pi^0 \to \gamma\gamma$, for example) provided the particle showers don't overlap.

Algorithm

Each subdetector is considered separately, giving rise to a collection of clusters specific to it. The clustering procedure is summarized thus:

(i) Find all rechits greater than the *seed threshold*. Neighbouring rechits exceeding the threshold (see Table 4.1) contest each other for seed status; the rechit with the greatest energy becomes the seed. In the ECAL, crystals which share a corner qualify as 'neighbouring'; in the HCAL the towers must share an edge. In other words, cells must be a local maximum above the seed threshold to qualify for seed status.

(ii) Grow clusters: connect the remaining rechits with energies greater than the *cell threshold* to the seeds, where the proposed rechit and any of the seed and already connected cells are neighbours (according to the definition above). Cells may belong to more than one cluster. Note, however, that in practice clusters do not grow larger than 2 or 3 cells across even in the most dense environments.

(iii) Determine the energy and position of the clusters with an iterative procedure. First, each cluster is assigned a position equal to that of its original seed. Second, each rechit contributes energy to each of its parent clusters with a weight,

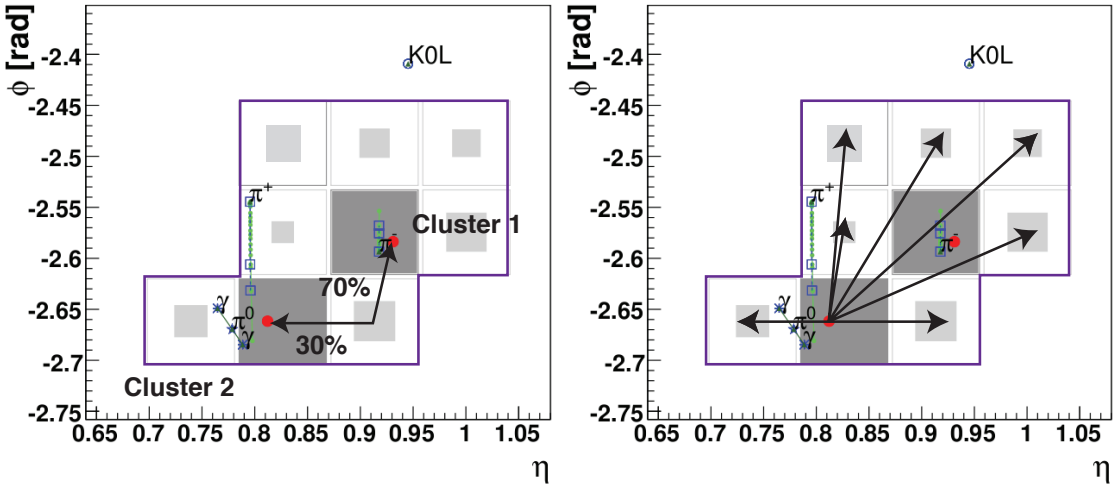
$$w_{ij} \exp(-d_{ij}^2/R^2) \tag{4.1}$$

where d_{ij} is the distance between the cluster i and cell j, R is given by Table 4.1, and w_{ij} is a normalization to prevent double-counting of energy. The position of each cluster is then re-computed as the average position of its rechits, weighted by a factor $\log(E_j/E_{\text{cell}})$. The energy of the cluster is then re-evaluated. This position/energy reassignment is repeated until the cluster's position does not move more than a small fraction of that subdetector's position resolution.[1]

Note that clusters can overlap as they share rechits. Current PF clustering parameters are shown in Table 4.1, and an example of clustering in the HCAL is shown in Figure 4.2.

[1] A further correction is applied in the ECAL since the crystals point away from the nominal CMS vertex.

FIGURE 4.2 **Particle flow clustering**
An example of particle flow clustering [79]. The seeds are shown in dark grey. In the HCAL clustering algorithm they are not considered as neighbours because they don't share a common edge. The green points correspond to the track intersection points. The blue boxes correspond to the simulated (Monte Carlo truth) particles. The red points denote the cluster centres. (This illustration shows the cluster positions after the complete clustering process, so they are not positioned on the seeds.)



(a) Finding the seeds in the HCAL. Two clusters overlap with an envelope shown in purple. The bottom right rechit contributes its energy to the two seeds with a weight according to Eq. 4.1.

(b) Sharing energy. Seeds do not contribute to other cluster's energies. This stabilizes the algorithm in highly asymmetric cases.

Optimizing the seed and cell thresholds with data

The seed and cell thresholds are based on HCAL noise values derived from testbeam data and implemented in Monte Carlo. The values are a compromise between high efficiency (low thresholds) and high purity (high thresholds — rejecting fake rechits created from noise).

Given the HCAL noise exhibited in Figure 3.2, the threshold seed values are 0.8 GeV for the barrel and 1.1 GeV for the endcaps. It is not meaningful to set a cell threshold higher than the seed threshold but, again, in order to exclude noise fluctuations, the cell thresholds are set to 0.8 GeV for both barrel and endcaps. These values, based on real data, represent a substantial change for particle flow, as the seed and cell thresholds were previously[2] 1.4 and 0.8 GeV (barrel) respectively, having been derived from an old simulation. The effect of varying the seed and cell thresholds was found to be quite weak when varied within a reasonable range (see Appendix B.1).

Particle flow's treatment of calorimetric energy & application to testbeam

At this point, it is appropriate to discuss how the particle flow reconstruction exploits calorimeter energy:

Calo Rechits As detailed in Chapter 3, §3.5.1. These are inputs to PF rechits (see below).

[2] Versions `CMSSW < 3XX`

TABLE 4.1 **PF Clustering parameters**
Reproduced from [41]. No clustering is applied in the HF. Components are: EB, EE – ECAL barrel & endcap; HB, HE – HCAL barrel & endcap; PS – preshower.

	EB	EE	HB	HE	PS
Cell threshold (GeV)	0.08	0.30	0.80	1.10	7×10^{-6}
Seed threshold (GeV)	0.23	0.80	0.80	0.80	5×10^{-4}
# cells to compare to candidate seed	8	8	4	4	8
R (cm) (Eq. 4.1)	5.0	5.0	10.0	10.0	0.2
# cells for position calculation	9	9	5	5	all

Calo Towers These objects combine ECAL and HCAL rechits into *towers* whose geometrical size is defined by the HCAL geometry. Rechits contribute to a Calo Tower according to a threshold scheme [80]. Particle flow only uses Calo Towers for technical reasons in the HCAL PF clustering.

PFRechits These are standard calo rechits with the addition of (η, ϕ) and (x, y, z) spatial coordinates, rather than just the raw '`CellID`s' found in the calo rechits. They form inputs to the PF clustering.

PFClusters Described above; are inputs to the PFBlock (see below) and PF reconstruction algorithms.

PFCandidates The outcome of particle flow is a list of PFCandidates. The four momentum of these PFCandidates should be an accurate description of the original particle's state. The possible type assignments are charged hadrons, neutral hadrons, electrons, photons, and muons. The PFCandidates are inputs to higher level physics objects, such as jets, missing energy, and taus [81].

The only relevant measurements for this document are (a) the total uncalibrated calorimeter energy recovered from calo rechits found in a cone of size $\Delta\mathcal{R} = 0.15$ around the track impact point, and (b) the calibrated calorimeter energy associated with PFCandidates reconstructed from the same single particle event.

In Figure 3.6 of Chapter 3, we saw that that the total uncalibrated calo rechit energy is a poor estimate of the true particle energy. The corresponding *CaloJets* [82]—which are built from calo rechits alone—therefore need large corrections. The calibration employed by particle flow ensures the PFCandidate's calorimetric energies are, on average, a good approximation of the true particle energy. This will be demonstrated in the sections that follow.

4.2.3 The PFBlock: geometrical linking of the elements

Particles will generally interact with more than one subdetector at CMS: given tracks and clusters and global muon reconstruction, the next step is to link together those elements from different subdetectors which are due to the same original particle. The PFBlock exists for this purpose.

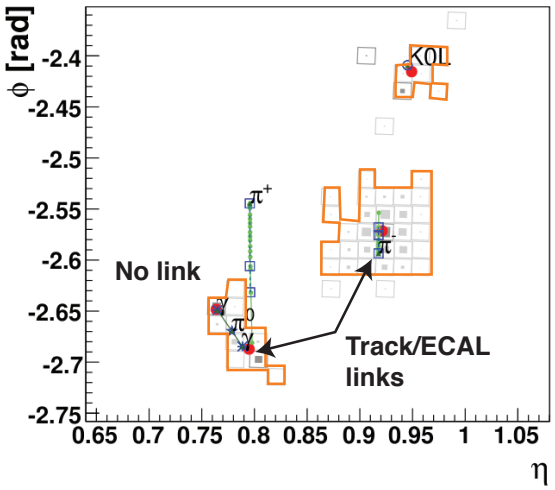
The block algorithm considers all the elements in the event and links elements together based on their spatial separations. Tracks are linked to clusters as follows:

(i) The track is extrapolated one X_0 depth into the ECAL and one λ_I depth into the HCAL. In the endcaps, it is also extrapolated to the preshower in front of the ECAL.

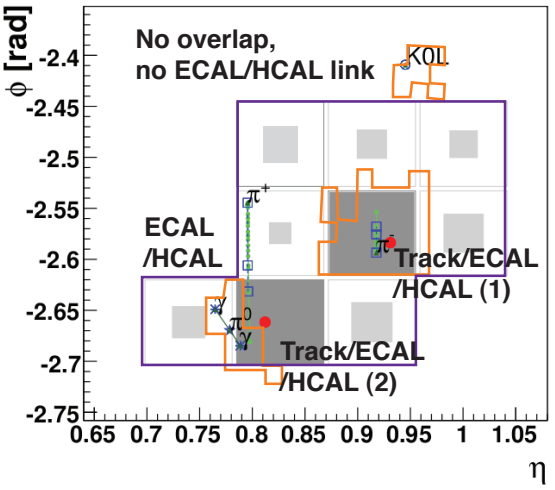
(ii) A cluster is associated to the track if the $\eta\phi$ projection of the cluster contains the track. The envelope can be enlarged to consider cracks in the calorimeters, uncertainty in the position of the shower maximum, and multiple scattering of very low p_T particles. In each case, the link distance is computed in the $\eta\phi$ plane. Recall that clusters may overlap with each other, so more than one cluster from any one subdetector may be connected to the track.

Clusters in each of the ECAL and HCAL, or preshower and ECAL are associated with each other when the cluster in the less granular calorimeter completely contains the cluster in the more granular calorimeter, unless a track can link them both. Figure 4.3 presents an example of the linking.

FIGURE 4.3 **Linking clusters to tracks**
Example from [79]. The combined envelopes of the ECAL clusters are shown in orange, and the HCAL envelopes are in purple. The reconstructed tracks, and their extrapolations to the ECAL and HCAL, are shown in green.



(a) Four ECAL clusters have been found. The π^+ does not give rise to a cluster in the ECAL. ECAL clusters are linked to the track if the track intersects a rechit in a cluster.



(b) The left hand ECAL (2) and HCAL (2) clusters get associated with each other, and the HCAL cluster is associated with a track. The central clusters get associated with each other (1), but only the HCAL (1) cluster also gets associated to the track. The top-right ECAL cluster remains orphaned.

The PFBlock is intended to be a small object relevant to a single particle (but including, for example, the photons from electron bremsstrahlung). This treatment is possible due to the granularity of the CMS detector, particularly in the tracker, preshower, ECAL, and muon systems. The number of PF blocks is therefore proportional to the complexity of the event, and a jet from a typical QCD event will usually give rise to a large number of blocks with a few elements each, rather than one large block with many elements. This also makes the algorithm easier to understand, especially in scaling from sparse to dense events.

4.2.4 PFAlgo: analyzing the PFBlocks

Overview

And so the gravamen cometh. Figure 4.4 illustrates how PFAlgo reconstructs PF hadron candidates, and it may be helpful to refer to this figure in the discussion that follows. The algorithm examines the properties of each PFBlock in turn to give a list of PFCandidates. The essential objective is to estimate the true energy of a particle using the calorimeters and compare it with the tracking measurement where available. The poor resolution calorimetric measurement can then be replaced with the generally accurate and precise momentum measurement.

Leptons

The rarity of leptons in hadronic collisions makes them easier to recognize. They are the first objects to be reconstructed from the event and, in doing so, some clusters and tracks will be removed from a block before the following treatment.

Poor quality tracks

Tracks that remain in the block which have poor quality (high 'χ^2s') and carry less p_T than the corresponding calorimeter energy measurement, within the calorimeter's resolution (more of which will come in §4.4), are discarded. Approximately 0.2% of tracks in hadronic jets meet this fate, of which ~ 90% are fakes. The remaining 10%'s energy will be accounted for by the calorimeters, which can be expected to have better precision in these cases. Meanwhile, the remainder of the block will generally become photons, charged or neutral hadrons or, occasionally, new muons.

Reconstructing photons & hadrons (charged and neutral)

If the block contains a track linked to a number of ECAL and HCAL clusters, then the energy of the clusters is compared to the momentum of the track. First, the PFCluster energies are calibrated using,

$$E_{\text{reco}} = a + b(E_{\text{eval}}, \eta)E_{\text{ECAL}} + c(E_{\text{eval}}, \eta)E_{\text{HCAL}} \tag{4.2}$$

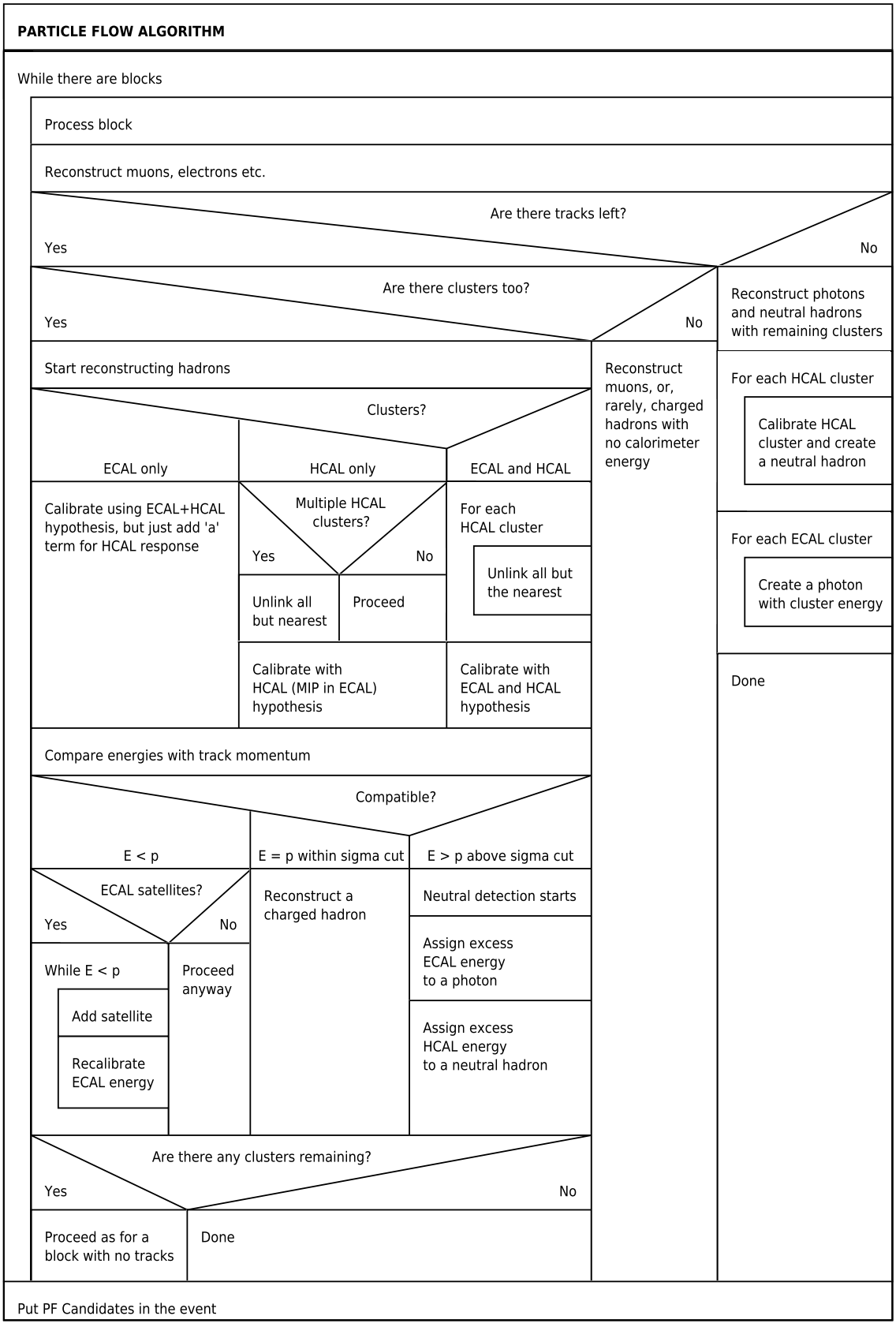
where E_{ECAL} and E_{HCAL} are the (total) cluster energies linked in the block, and a, b, c are calibration coefficients. (The determination of these coefficients will be described in Section 4.4.) The calibration coefficients seek to convert the non-linear calorimeter response shown, for example, in Figure 3.6 to a reasonably accurate estimate of the true particle energy, which can then be compared with the track momentum. In Eq. 4.2, the b and c are evaluated at E_{eval}, which is taken as,

$$E_{\text{eval}} = \max(E_{\text{ECAL}} + E_{\text{HCAL}}, p). \tag{4.3}$$

Since $\pi/e < 1$ it is generally the track momentum that is taken.

Several tracks may link to any one HCAL cluster, and the sum of their momenta is compared to the calibrated cluster energy. If the track links to several HCAL clusters, only the closest cluster in the $\eta\phi$ plane is kept and the remaining HCAL clusters become

FIGURE 4.4 **PFAlgo logic**
Specifically for hadron reconstruction. This is a *Nassi-Shneiderman* diagram [83], which illustrates a structured program's flow in a much more economical way than the usual flowchart.



PF neutral hadrons. A similar procedure is applied to the ECAL clusters, but because fluctuations in the electromagnetic fraction F_0 discussed in Chapter 2 might be the reason for these clusters, a *satellite recovery algorithm* reconnects these extra clusters to the track in order of distance, provided the total calibrated calorimeter energy at each step is less than the total track momentum.

If the total calibrated calorimeter energy is still less than the track momentum by more than 3σ of the expected calorimeter energy resolution, then a new search for muons is performed subject to relaxed criteria. If no muons can be found which satisfy the requirements, which is rare $\mathcal{O}(< 0.03\%$ of all tracks), then the remaining tracks in the block are discarded in descending order of the uncertainty of their measured p_T, until either (a) all those with an uncertainty in excess of 1 GeV/c have been examined, or (b) the total track momentum would become smaller than the calibrated calorimeter energy.

Tracks which remain will become PF charged hadrons. The calibrated calorimeter energy is compared to the total track momentum p and a hypothesis test is made:

> H_0: the excess of calorimeter energy is due to a shower fluctuation and/or detector effects.
>
> H_1: the excess of calorimeter energy implies the existence of a neutral particle coincident with the original charged hadron's impact point.

The default hypothesis, H_0, is accepted when the calibrated calorimeter energy does not exceed the total track momentum by more than 1σ. If H_0 is accepted, then the charged hadrons' momenta p_i are determined by recombining the tracker and calorimeter measurements to minimize,

$$\chi^2 = \sum_i \frac{(p_i - t_i)^2}{\sigma_i^2} + \frac{(\sum_i p_i - E)^2}{\sigma_E^2} \tag{4.4}$$

where t_i, σ_i are the ith track's momentum and uncertainty, and E, σ_E are the corresponding calorimeter measurement and error. This optimization becomes important at high η and high energies, where the tracks are generally measured with poorer resolution and quality.

When H_1 is accepted because the calibrated calorimeter energy exceeds the track momentum by at least 1σ, a further test is conducted to ascertain which PFCandidates should be created to absorb the excess, and with what energy. In principle, either a photon or a neutral hadron could be inferred. Photons are preferentially created in the algorithm because 25% of a jet's energy is transported, on average, by photons. By contrast, only 3% of a jet's energy is deposited in the ECAL by neutral hadrons. If the total excess energy is larger than the calibrated ECAL energy, a PF photon is created with the ECAL energy, and a PF neutral hadron is assigned the remainder. Alternatively, a PF photon is created with the *uncalibrated* excess.

The cut of 1σ excess is arbitrary. Note that $\sigma = \sigma(E) \oplus \sigma(p)$ and at 100 GeV is approximately 17 GeV: this depends on the expected energy resolution of PFCandidates, which will be discussed in due course and includes the increased uncertainty that results from subtracting any contributions due to muons previously reconstructed in the block. So, in the standard particle flow algorithm, neutral hadrons and photons are *only resolved*

above this cut. Furthermore, the energy of the excess particle is,

$$E_{\text{neutral}} = E_{\text{calo}} - p_{\text{track}}. \tag{4.5}$$

The minimum energy of the neutral is therefore 1σ, which manifests itself as a sudden bump in the energy spectrum of these 'excess' particles. For the case of single isolated particles, one could adopt,

$$E_{\text{neutral}} = E_{\text{calo}} - p_{\text{track}} - 1\sigma. \tag{4.6}$$

While neutrals are only resolved when there is a 1σ excess, the smallest energy is now zero and the excess energy spectrum becomes smooth. However this is not the case in PFAlgo, as internal studies have shown that this degrades the PF jet performance.

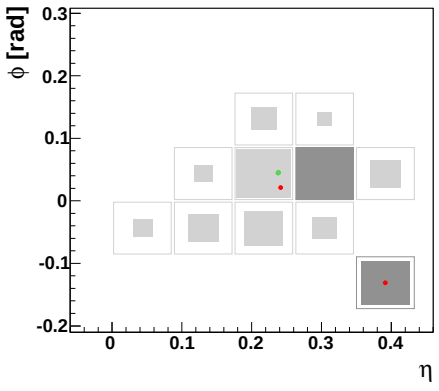
Finally, with all the tracks accounted for, any remaining ECAL clusters will give rise to PF photons, and the remaining HCAL clusters will create PF neutral hadrons. The neutral hadron PFCandidates are assigned calibrated energies. PF photons are assigned the raw ECAL energy with a small photon-specific calibration to account for ECAL cracks, thresholds, the ECAL's calibration to electrons, and the primary vertex's position.

4.2.5 Excess of HCAL energy: a closer look

For the case of two distinct HCAL clusters, an excess neutral hadron will be created. An HCAL cluster is only linked to a track in the PFBlock if the track directly intersects it. The effect of this on a 100 GeV event from testbeam is presented in Figure 4.5, where an excess neutral of energy 13.4 GeV has been made because one HCAL cluster does not overlap with the track. In this event, there are two separate blocks because the bottom-right cluster's envelope does not contain the track. But even if an association is made at the block level, PFAlgo will unlink all but the closest HCAL cluster to the track. A further examination of this scenario is made in §4.5.4.

FIGURE 4.5

Multiple HCAL clusters, with one linked to the track
From a 100 GeV testbeam event. Two HCAL PFClusters imply an excess fake hadronic neutral may be created. The track impact point is shown by the green dot. The cluster at bottom-right leads to a neutral hadron PFCandidate with 13.4 GeV of calibrated energy.



▷ 4.3 Preparing to apply particle flow to testbeam

4.3.1 Modifications required for testbeam

Subsection 3.5.1 detailed some of the technical obstacles to be overcome in applying the standard reconstruction to the testbeam dataset. In addition to these, the following modifications are required to enable the proper functioning of the particle flow algorithm:

Preshower (endcap dataset only)

Preshower hits are available but are not aligned by default with the ECAL and HCAL geometry. The preshower is not directly used in this analysis, but some pions will interact with the preshower material or scintillator material before them. These events are usually correlated with an (unusual) absence of HCAL activity. To simplify the analysis, events which deposit more than a MIP equivalent in the preshower detector are removed.

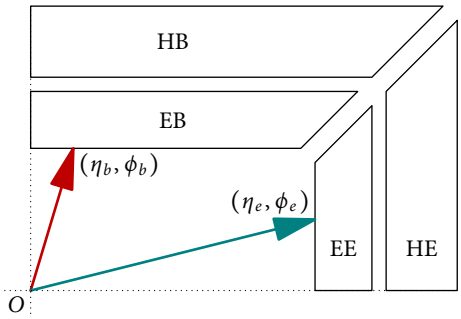
Providing tracking

Accurate tracking for the beam particles was not available at the testbeams of 2006 and 2007. Fortunately the trigger counters define a window $4\times4\,\mathrm{cm}^2$ wide so, neglecting the effects of air scattering, the beam particles hit the ECAL front face with coordinates[3] $\eta = 0.2\pm0.03$, $\phi = 0.02\pm0.03$ in the barrel experiment, and $\eta = 1.6\pm0.03$, $\phi = 0.37\pm0.03$ for the endcaps.

PFTracks are built starting from $(x, y, z) = (0, 0, 0)$ propagating in the direction $\eta = 0.2$, $\phi = 0.02$ using the known beam momentum (Figure 4.6). Because the size of an ECAL crystal at $2.2 \times 2.2\,\mathrm{cm}^2$ is smaller by a factor of 2 than the uncertainty on the track impact point, the use of this dataset in evaluating the ECAL's spatial resolution for pion reconstruction is compromised. The HCAL cells however at $\sim 15\times 15\,\mathrm{cm}^2$ are much larger than this uncertainty.

FIGURE 4.6

Fake tracks
Providing fake tracks for PFBlockAlgo and PFAlgo to analyse.



The PFBlock algorithm has been modified to account for the uncertainty on the track impact point. In the barrel, all ECAL clusters within $\Delta\mathcal{R} = 0.02$ of the track impact point are associated with the track. For the endcaps, reconstructing the geometry consistently in many runs has been problematic. Associating all ECAL clusters within $\Delta\mathcal{R} = 0.04$

[3] In some runs this was changed, and analysis described here accounts for this.

suffices to fix this problem in the endcaps. In the absence of these modifications, fake photons with very large energies were formed.

▷ 4.4 Evaluating a particle flow calibration for hadrons

Overview

The calibration of hadrons is described in [41]. In summary, coefficients a, b, c defined in Eq. 4.2 are to be determined to minimize the quantity,

$$\chi^2 = \sum_{i=1}^{N} \frac{\left(E^i_{\text{reco}} - E^i_{\text{true}}\right)^2}{\sigma_i^2}, \tag{4.7}$$

for N sample hadrons, and σ_i the *expected* calorimeter energy resolution for the ith test hadron. E_{reco} is the sum of calibrated cluster energies associated to the track (Eq. 4.2), while E_{true} is taken from Monte Carlo. The b, c coefficients are evaluated at energy E_{eval} (Eq. 4.3) taken as an estimate of the true energy of the particle. This χ^2 is separately defined for the barrel and endcap regions, and furthermore for particles interacting in the ECAL *as defined by the existence of a cluster there*[4] and those that do not. Dependence of the coefficients on ϕ is assumed to be negligible, because ϕ variation should have been removed at the earlier stages of calorimeter calibration.

As the calorimeter system exhibits a non-linear energy response, we cannot calibrate a non-linear system with a linear solution. These a, b, c coefficients are therefore determined in bins of energy, and their evolution with energy is then parametrized with *ad-hoc* functions (detailed below). The energy resolution, σ_i, is determined iteratively in the calibration procedure by evaluating the Gaussian sigma of $E_{\text{reco}}/E_{\text{true}}$ for each bin in energy. Finally, the a coefficient is also determined iteratively, so as to minimize the variation of b and c on E_{true}: this reduces the systematic uncertainly in evaluating b, c for neutral hadrons, for which no accurate estimate of the true energy is available, in contrast to charged hadrons whose total momentum is generally well-measured. In all cases, the parametrizations of b, c are evaluated at E_{eval} from Eq. 4.3.

Reference [41] describes how the standard calibration coefficients were extracted from particle gun K^0_L events generated in the fast simulation, over a continuum of energies from 1 GeV to 1 TeV and over the full range of CMS' η, ϕ coverage. An η dependence is observed in simulation (due to geometrical and material effects), so a residual correction is applied after the a, b, c calibration. The testbeam dataset allows us to cross-check the coefficients at each of the beam energies and, while we can also verify the coefficients for the barrel and endcap, we cannot verify the η dependence because the beam was not scanned over the entire range of ECAL and HCAL supermodules. In the interests of consistency however, comparisons with simulations presented forthwith are repeated in the same manner as done for the data, *i.e.* with discrete energy points and for just two η/ϕ points (barrel and endcap).

[4] This is NOT the same as the definition described in Chapter 3, because that applies to calo rechits.

Prescription

We start by anticipating some variation in the calorimeter and PF response between data and simulation. At the start of this procedure, the default set of fast simulation-derived calibration coefficients is in place. The procedure for calibrating the testbeam datasets is as follows:

(i) Neutral hadrons and photons are created where the *calibrated* energy of the calorimeters is more than 1σ in excess of the hadron's track momentum. To make an accurate and precise calibration for single particles, we must therefore loosen this cut (raised to 5σ) so that we do not create excess neutrals which would otherwise bias the sample for calibration to those pions with energies $< +1\sigma$ of the true distribution's mean (this would otherwise exclude ~ 17% of the available data).

(ii) We select all events where there was at least an HCAL cluster. All the clusters within a cone of $\Delta\mathcal{R} = 0.15$—chosen to be consistent with the HCAL calibration definition used in Chapter 3—of the track impact point are selected for calibration. This strategy, which we call a 'global calibration', aims to make the total reconstructed event energy match the incoming track momentum, and not just the energy associated with the PF charged hadron. Other strategies are possible and are discussed in the next chapter. We produce a preliminary calibration based on this dataset.

(iii) The entire dataset is then reprocessed using the preliminary calibration, and the same selection of events is made.

(iv) The calibration procedure is then repeated to check for convergence.

A further reason for loosening the σ cut is that if the fast simulation derived calibration produces a response that is too strong (*i.e.* $E_{\text{reco}}/E_{\text{true}}$ is overestimated), then many events will result in a charged hadron with some neutral hadrons or photons created in excess. Figure 4.7 illustrates the scenario. If we were then to only select events with one charged hadron, we would be biasing the sample for calibration. On this occasion, neither of the yellow shaded clusters would be included in the calibration because they are not within $\Delta\mathcal{R} = 0.15$ of the track impact point. While this particular event indicates that the cone size chosen may provide inadequate containment of the hadronic shower, events which exhibit this level of outside-cone activity are very rare $\mathcal{O}$(a few percent). Furthermore, even if the (arbitrary) HCAL calibration definition meant that, say, 10% of shower energy fell outside the cone, this energy would be distributed over a larger area and is unlikely to exceed PFCluster thresholds in any one cluster cell.

In all the cases presented here, there was a negligible difference between the preliminary and final calibration.

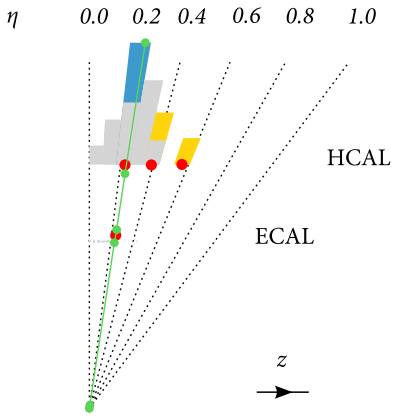
4.4.1 Testbeam calibration coefficients

The calibration coefficients are parametrized by,

$$\alpha + \left(\beta + \frac{\gamma}{\sqrt{x}}\right) e^{-x/\delta} - \epsilon e^{-x^2/\zeta}. \tag{4.8}$$

FIGURE 4.7

Satellite clusters vs. excess energy
In this 100 GeV testbeam event, 3 HCAL clusters are seen. Satellite HCAL clusters are shown in yellow. The cluster attached to the track has a raw energy of 103 GeV (shown in blue). When calibrated using E_{eval} = 103 GeV, the calibrated energy (using the fast simulation calibration) is more than $1\sigma \sim 17\%$ from the track momentum, indicating that a neutral particle should be resolved (of 25 GeV in this case). In the testbeam calibration procedure, such an excess particle is *not* formed and the entire blue cluster is left attached to the PFTrack for calibration.



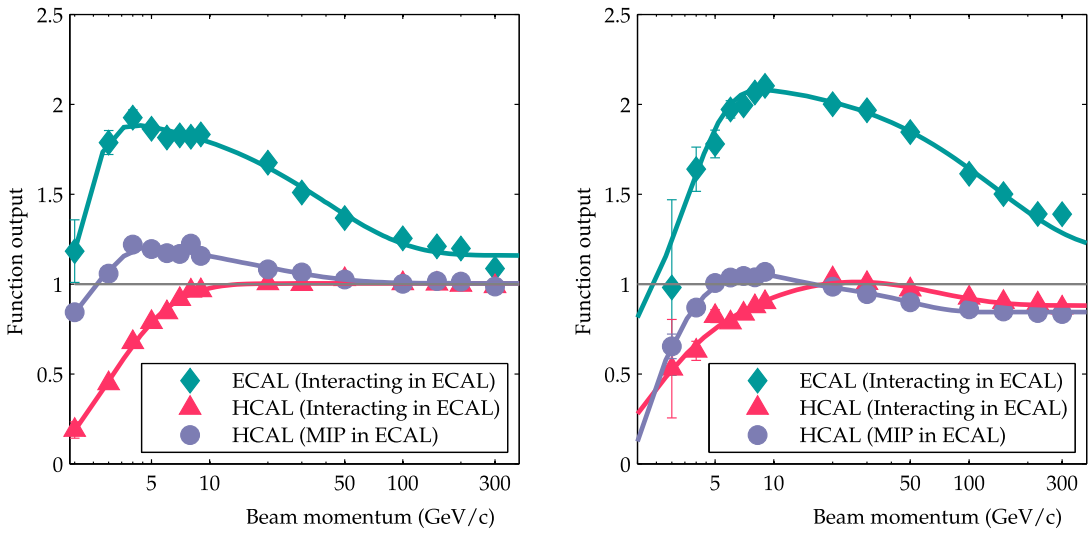
This function has no particular or meaningful physical interpretation. Generally MINUIT [74] can find a stable fit with errors on each of the parameters $\mathcal{O}(< 10\%)$, though some human interaction is occasionally required to refine the fit. In particular, the function should not become negative at any positive energy. Figure 4.8 shows the calibration coefficients obtained for testbeam data. The error bars are given by the least-squares minimization. Functions obtained from simulation are similar, but are different from the data due to the variations exhibited at the rechit level (see Chapter 3, §3.5.5). The standard fast simulation functions are shown in Appendix B.2, together with the testbeam's numerical fit values for Eq. 4.8. Two separate a terms are computed: one for the interacting in ECAL class, and one for the MIP in ECAL class; the values are shown in Table 4.2.

TABLE 4.2

Constant a term
For the two classes of pions, barrel and endcap. The units are GeV. The equivalent constants for the standard fast simulation calibration are also indicated. (The fast simulation calibration functions are shown in Appendix B.2.)

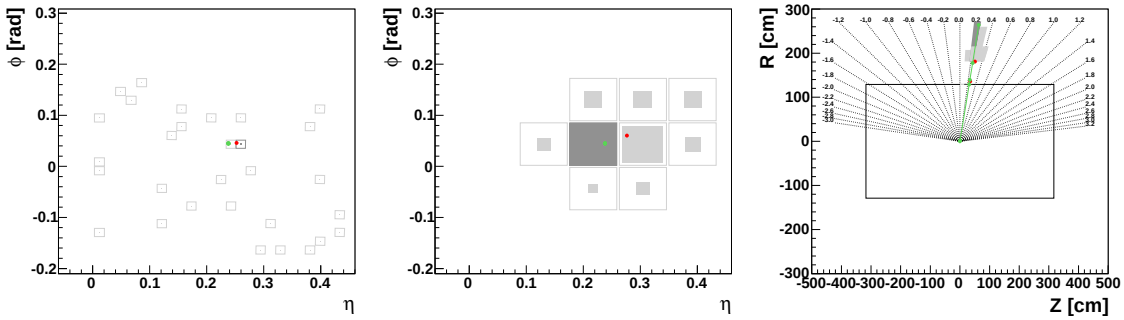
	0 ECAL clusters	≥ 1 ECAL cluster
Barrel	0.6	1.2
Endcap	1.1	0.1
c.f. Fastsim (Barrel & Endcap)	2.7	3.7

FIGURE 4.8 **PFCandidate calibration**
Calibration coefficients extracted from testbeam data. The data are fit with the function of Eq. 4.8.



(a) Data, barrel *(b) Data, endcap*

FIGURE 4.9 **Event display for a 100 GeV charged pion event: single cluster**
In this example, there is only one HCAL cluster and one ECAL cluster.



(a) ECAL *(b) HCAL* *(c) RZ*

▷ 4.5 Particle flow with testbeam

4.5.1 Event displays

Figure 4.9 shows the particle flow reconstruction applied to a 100 GeV testbeam event in the barrel. In this example, there is only one cluster in each of the ECAL and HCAL, each of which is linked to the track. The HCAL cluster carries 90 GeV of energy, and the ECAL cluster carries 1.7 GeV. When the calibrations are applied, the ECAL cluster's energy is multiplied by 1.25 and the HCAL's energy is barely changed. The *a* term is 1.2 GeV in this case, and the one charged hadron candidate that particle flow finds in this event has 93.3 GeV of calibrated energy.

A second, more complex example is shown in Figure 4.10, with two clusters in the ECAL and three in the HCAL. A full breakdown of the particle flow analysis is shown

in Table 4.3. One of the HCAL clusters arises from a noisy channel at $\eta = 1.35$ (this can be seen in the corresponding RZ display of Fig. 4.10). When calibrated this noisy cluster yields an energy of 1.4 GeV (shown as the second of the h° candidates). The most energetic ECAL and HCAL clusters are linked to the track and give the charged hadron candidate 96.9 GeV of calibrated energy. An extra ECAL cluster gives rise to a fake photon with energy of 1 GeV after the residual photon calibration. Finally, the HCAL cluster at $\eta = 0.4$ and energy 10.5 GeV yields a 13.4 GeV neutral hadron after calibration.

4.5.2 Response to charged hadrons

In this section, we often plot the quantity,

$$E_{\text{reco}}/E_{\text{true}} = \frac{\sum_{\forall \text{Cands}} \left(E_{\text{ECAL}'} + E_{\text{HCAL}'}\right)}{E_{\text{true}}}, \tag{4.9}$$

where the quantities $E_{\text{ECAL}'}$ and $E_{\text{HCAL}'}$ are the *calibrated* PFCandidate ECAL and HCAL energies found in a cone of $\Delta\mathcal{R} = 0.15$ around the track, using the calibration coefficients appropriate for testbeam shown in Figure 4.8. Note that we sum over *all* candidates, to assess the performance of the entire reconstruction, in keeping with the global calibration strategy described above. If we want to consider the response from charged hadrons alone, we could plot,

$$E_{\text{reco}}/E_{\text{true}} = \frac{\sum_{\forall \text{Cands}} \left(E_{\text{ECAL}'} + E_{\text{HCAL}'}\right) - \sum_{\forall \gamma} E_{\gamma} - \sum_{\forall h^{\circ}} E_{h^{\circ}}}{E_{\text{true}}}. \tag{4.10}$$

Figures 4.11 and 4.12 present the outcome of the reconstruction applied to both barrel and endcap datasets, subject to the requirement that the track is at least linked to an HCAL cluster. Here we show two classes of pions, namely those with little ECAL interaction and therefore, no cluster there, and those that do interact and leave at least one cluster there.

These data are summarized in Figure 4.13 where the fitted Gaussian mean has been plotted as a function of the beam momentum. The error bars refer to the error on the fitted mean given by ROOT/MINUIT [73, 74].

The same procedure has been applied to the full and fast simulations. Chapter 3 detailed how the simulations are used to model the testbeam experiments. The magnetic field was set to 4T in the simulations so the standard particle flow tracking could be applied. The equivalent calibration functions and response for the full simulation are shown in Figure 4.14 which demonstrate that the performance (using calibrations appropriate to simulation) in simulation is similar to that with data: no substantial differences between simulation and data were found in the energy spectra after the appropriate calibration was applied.

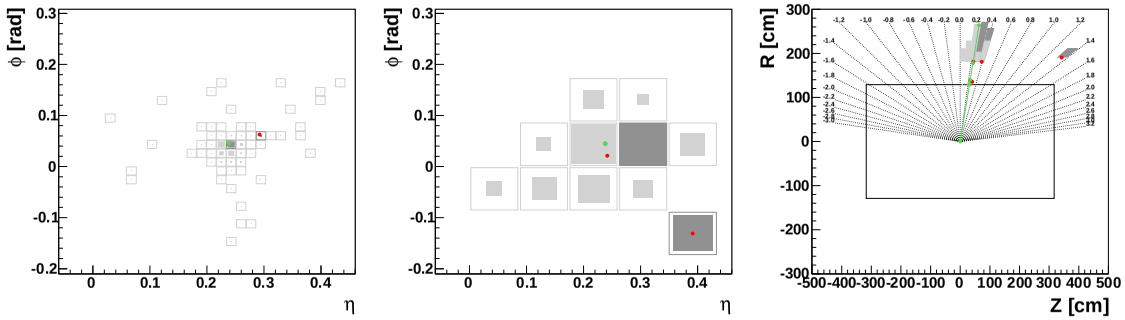
Interpretation

In contrast with the results shown in Chapter 3, we do not observe values of $E_{\text{reco}}/E_{\text{true}} < 0$. That the ECAL and HCAL were running with ZSP or SR disabled is irrelevant because the PF clustering will only allow energies above the substantial thresholds to contribute to a cluster's energy. Only events which are linked to an HCAL cluster are shown.

Like the raw calo rechit spectra, the distributions are broadly Gaussian where there is energy to calibrate. The distributions above 10 GeV are broadly centred at unity. For

FIGURE 4.10 **Event display for a 100 GeV charged pion event: multiple clusters**

In this example, HCAL activity has given rise to two ECAL clusters (one is centred directly beneath the track impact point), and two HCAL clusters. Some activity is also seen in the rz plane at $\eta = 1.4$ due to a noisy HCAL channel.



(a) ECAL *(b) HCAL* *(c) RZ*

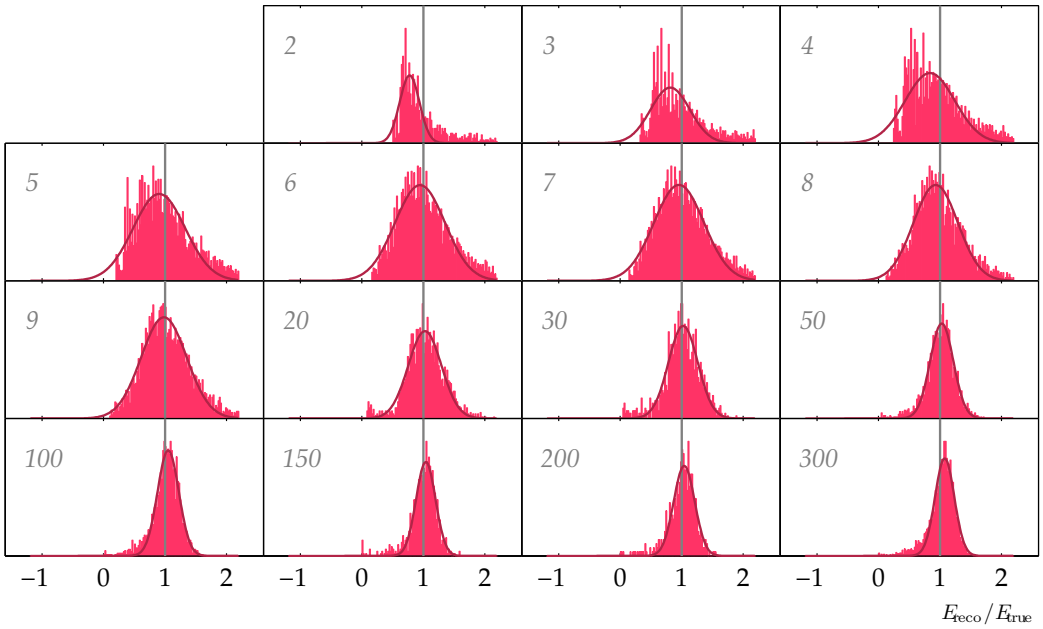
TABLE 4.3 **Example PF analysis for testbeam data**

Summary of PFCandidates found for Figure 4.10, with clusters and calibrations. See text for details.

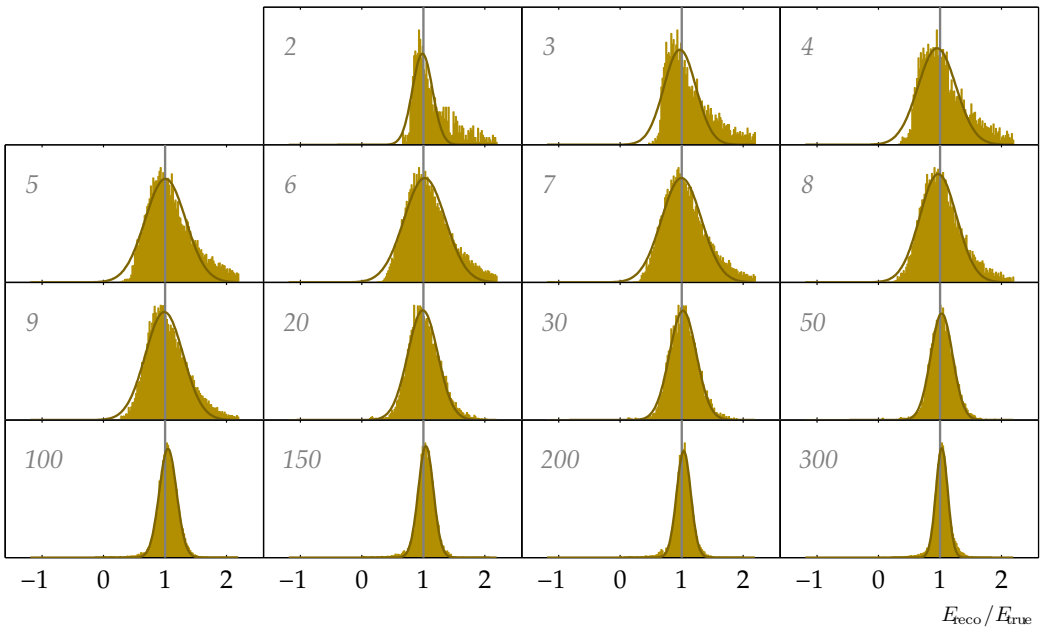
ECAL Clusters	E (GeV)	Attaches to…	**HCAL Clusters**	E (GeV)	Attaches to…
$\eta = 0.237$	34.1	$h^{\pm}$	$\eta = 0.241$	53.1	$h^{\pm}$
$\eta = 0.292$	0.914	γ	$\eta = 0.392$	10.5	h^{0} (1)
			$\eta = 1.35$	0.911	h^{0} (2)
PFCandidates					
$h^{\pm}$	$E^{\ddagger}$ (GeV)	γ	E (GeV)	h^{0}	E (GeV)
$\eta = 0.0238$, $\phi = 0.045$	$1.0^{(\$)} \times 53.1$	$\eta = 0.292$, $\phi = 0.063$	$f_{\gamma}(0.914)^{(*)}$	(1) $\eta = 0.391$, $\phi = -0.031$	$10.5 \times \sim 1.3 + a$ [0.6]
	$+ 1.25 \times 34.1$		$\Rightarrow 1.06$		$\Rightarrow 13.42$
	$+ a$ [1.2]			(2) $\eta = 1.35$, $\phi = -0.13$	$0.911 \times 0.9 + a$ [0.6]
	$\Rightarrow 96.9$				$\Rightarrow 1.40$

Notes: (‡) This is the total calibrated ECAL + HCAL energy. ($) 1.0 and 1.25 are the b, c coefficients. (*) f_{γ} is a residual photon correction.

FIGURE 4.11 **Barrel PFCandidate energy spectra**
The x-axis corresponds to Eq. 4.9. The inset numbers indicate the beam momentum in units of GeV/c.

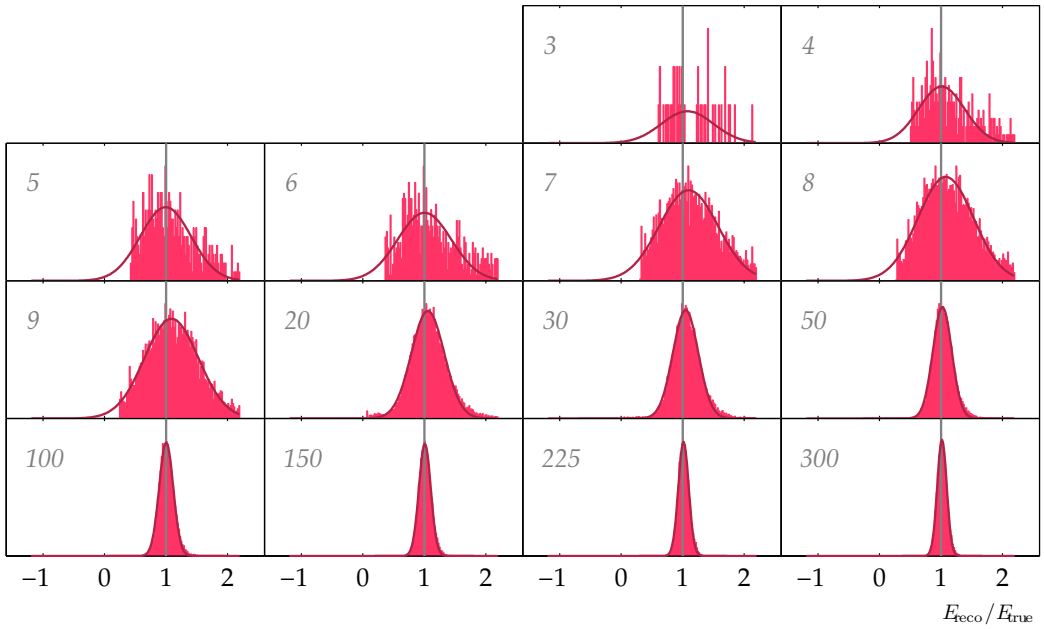


(a) No ECAL clusters present

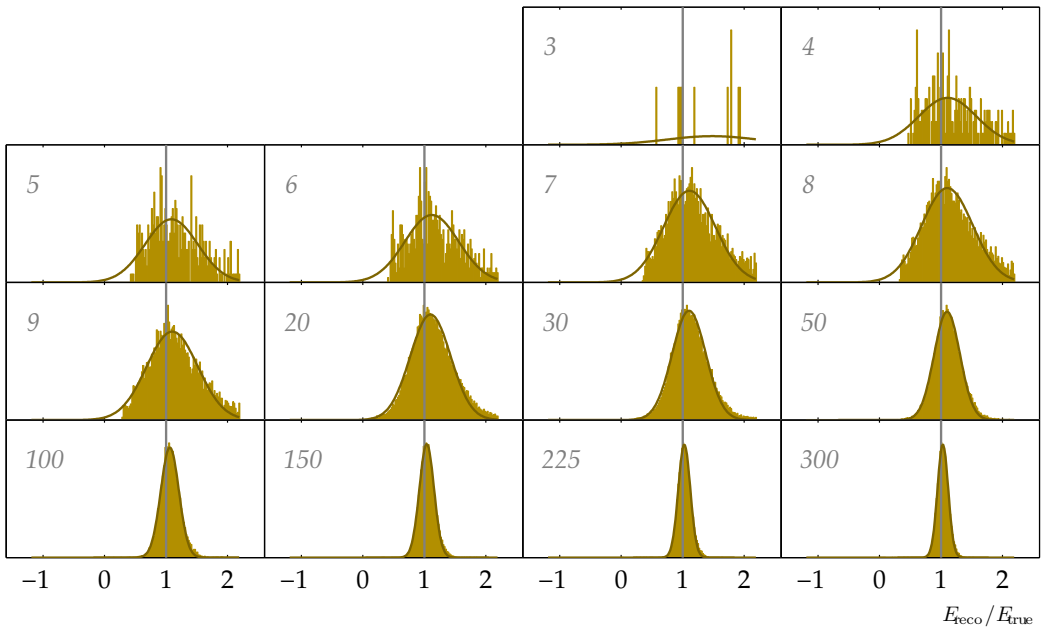


(b) With ≥ 1 ECAL cluster

FIGURE 4.12 **Endcap PFCandidate energy spectra**
The x-axis corresponds to Eq. 4.9. The inset numbers indicate the beam momentum in units of GeV/c.

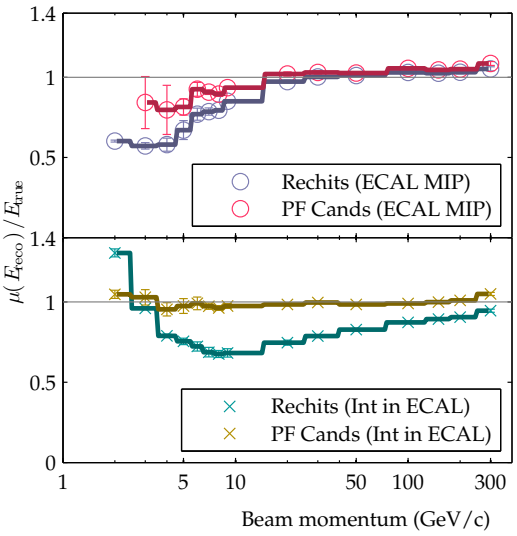


(a) No ECAL clusters present

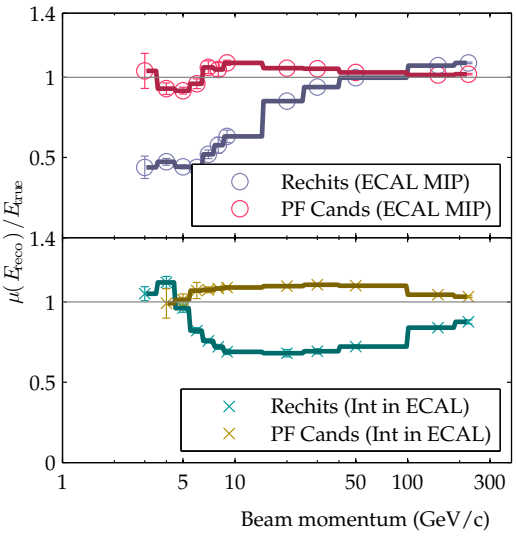


(b) With ≥ 1 ECAL cluster

FIGURE 4.13 **Single hadron response**
Fitted Gaussian means to the calo rechit and reconstructed PFCandidate energy distributions. The error bars indicate the error on the fitted mean.

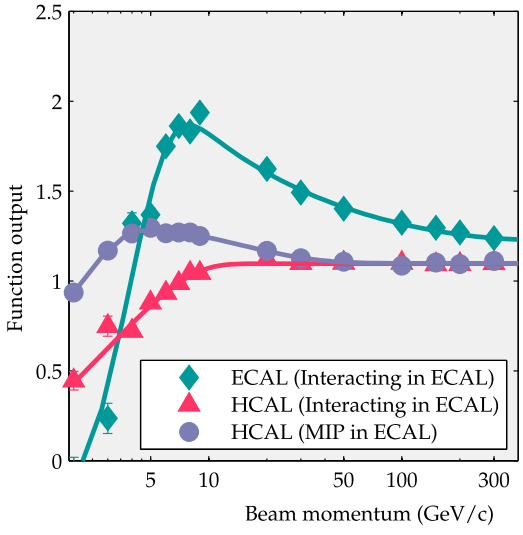


(a) Data, barrel

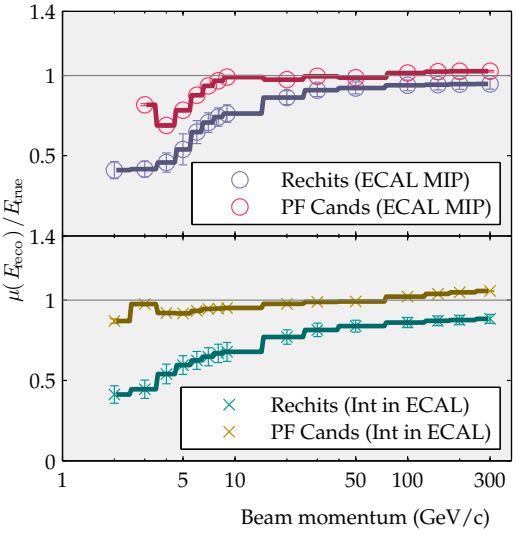


(b) Data, endcap

FIGURE 4.14 **PFCandidate calibration and response (full simulation, barrel)**
For example, for pions in the barrel. The calibration coefficients shown on the left give the response shown on the right. Note that the transformation from rechits to PFCandidate energies includes both clustering and an *a* term (not shown here).

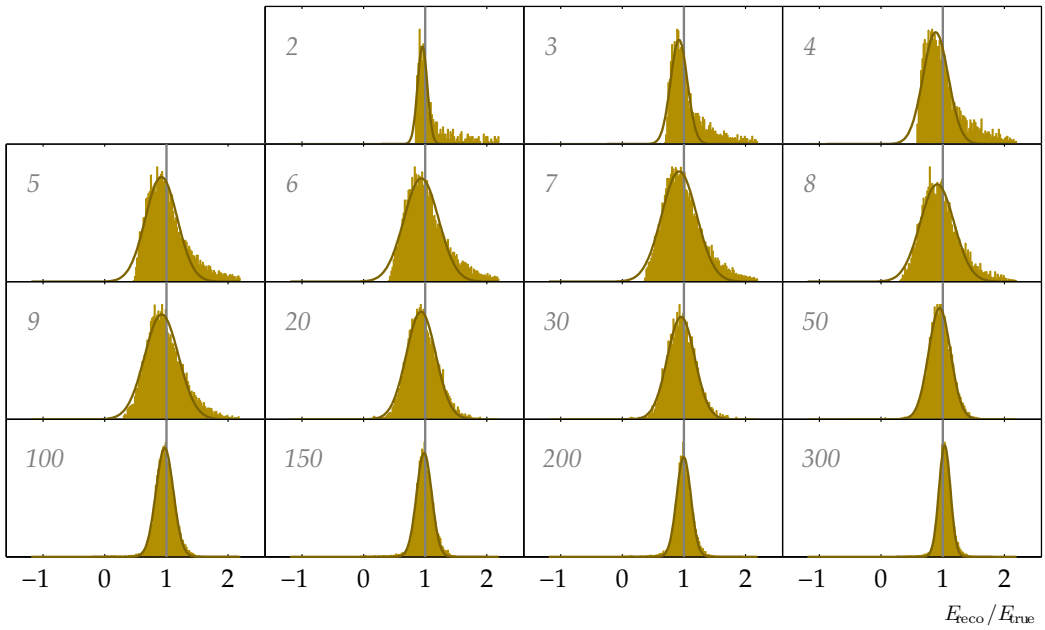


(a) Calibration coefficients



(b) Response

FIGURE 4.15 **Barrel PFCandidates— charged hadrons, with ECAL activity**
The x-axis corresponds to Eq. 4.10, so energy contributed by 'fake' photons and neutral hadrons is excluded. The inset numbers indicate the beam momentum in units of GeV/c.



energies less than 10 GeV, a Gaussian is not strictly appropriate to the data, but the calibration has undoubtedly improved the response. The over-amplification of energies seen in the endcap (Figure 4.13 (b)) is the result of calibrating *all* the clusters in the $\Delta\mathcal{R} = 0.15$ cone, rather than the clusters directly associated with the charged hadron's track. For while the calibration algorithm *does* restore unity response, PFAlgo calibrates the candidate charged hadron satellite ECAL and HCAL clusters separately. The extra clusters are generally (a) not attached to the track, and (b) of lower energy than the track momentum. They therefore get calibrated with higher a, b, c coefficients than they would were they attached to the track. An exploration of this phenomenon—which affects the endcap more than the barrel—is considered in §4.5.4.

Consider the response due to just charged hadron candidates, expressed by Eq. 4.10, and shown in Figure 4.15. The distributions are broadly unaffected. At low energy, some clusters do not get linked to the track. In the absence of accurate tracking, it is not possible to draw firm conclusions about the efficiency of the PFBlock algorithm at linking tracks to clusters.

4.5.3 Hadronic energy resolution (calorimetry)

The resolution of the calo rechits and PFCandidates is exhibited in Figure 4.16 for the barrel. The raw rechit analysis presented in Chapter 3 gives a resolution for pions in the barrel of,

$$110\%/\sqrt{E} \oplus 12\% \quad \text{(barrel)}, \tag{4.11}$$

while the same analysis for the endcap yields a rather awful,

$$200\%/\sqrt{E} \oplus 8.8\% \quad \text{(endcap)}. \tag{4.12}$$

These values are consistent with the raw rechit resolutions quoted by [36]. The calibration and particle flow procedure gives a slight improvement. Fitting the same function to those pions which do not create an ECAL cluster gives a resolution of,

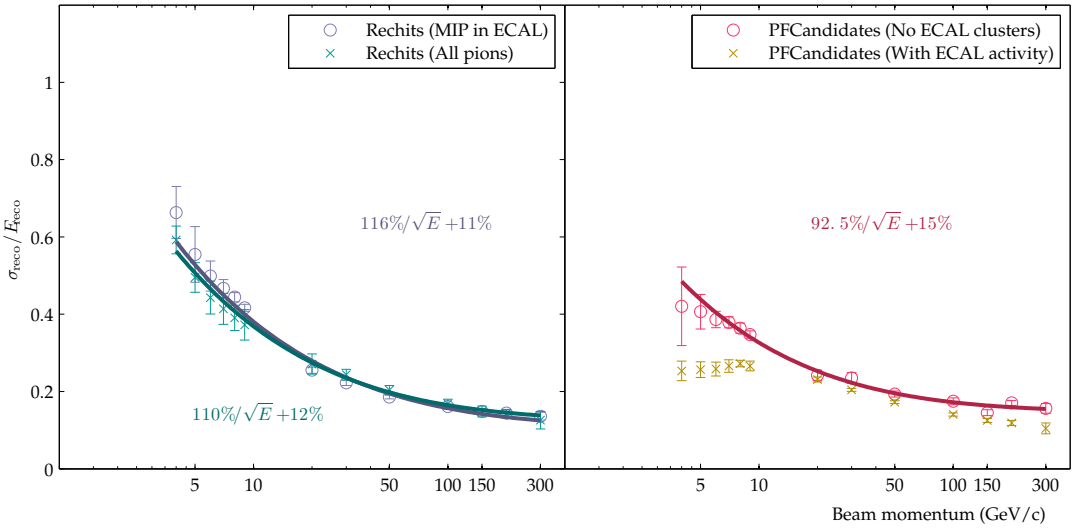
$$92.5\%/\sqrt{E} \oplus 15\% \quad \text{(barrel)}, \tag{4.13}$$

and for the endcap,

$$105\%/\sqrt{E} \oplus 3.9\% \quad \text{(endcap)}. \tag{4.14}$$

FIGURE 4.16

Hadronic energy resolution in testbeam
The left-hand side shows the raw rechit resolution following the analysis given in Chapter 3, while the right-hand side shows the resolution for PFCandidates calibrated using a testbeam-specific calibration (using Eq. 4.9). The odd low-energy behaviour for PFCandidates with ECAL clusters present (gold points) is symptomatic of a non-Gaussian response. The large error bars on the magenta points are the consequence of low statistics and large error on the fitted mean.



(In view of the non-Gaussian shapes characteristic of the low energy response for pions which do form ECAL clusters, it has been decided not to fit the resolution function to these data, but it is qualitatively clear that the overall resolution is acceptable.)

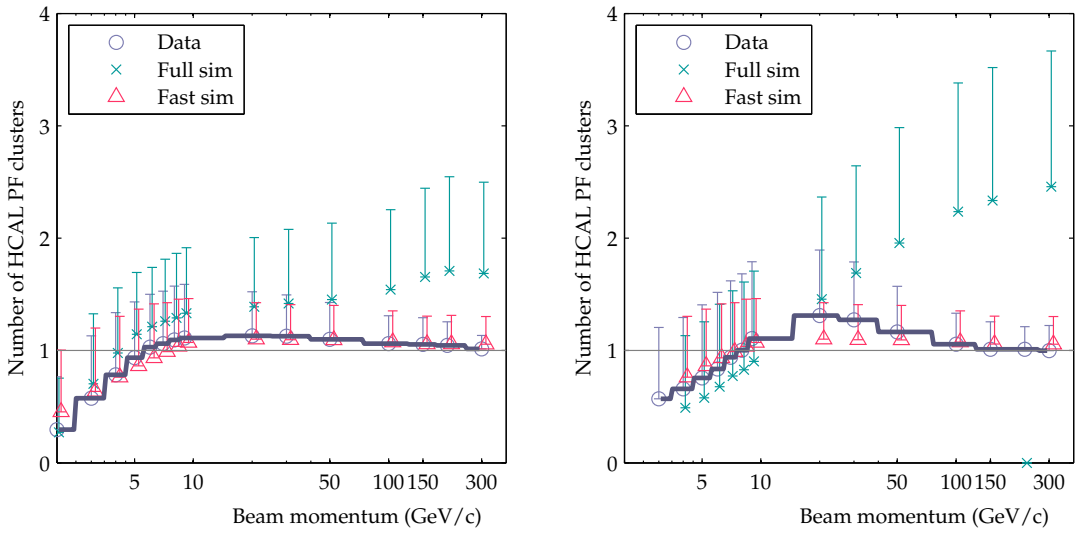
The parametrizations of the energy resolution in Eqs. 4.13, 4.14 are fed back into PFAlgo for the purposes of evaluating the σ cut (Eqs. 4.5, 4.6) discussed above. While the calibration does indeed improve the energy response, it is worth reiterating that the goal of PFAlgo is to replace the poor energy measurement given by the calorimetry with the comparatively excellent momentum measurement given by the tracking.

4.5.4 Particle flow cluster multiplicity

Figure 4.17 shows the arithmetic mean number of HCAL clusters per event in data, full, and fast simulations. Only events without ECAL clusters are selected. First, note that for the barrel and endcap, the average number of clusters found is less than unity for low energies.

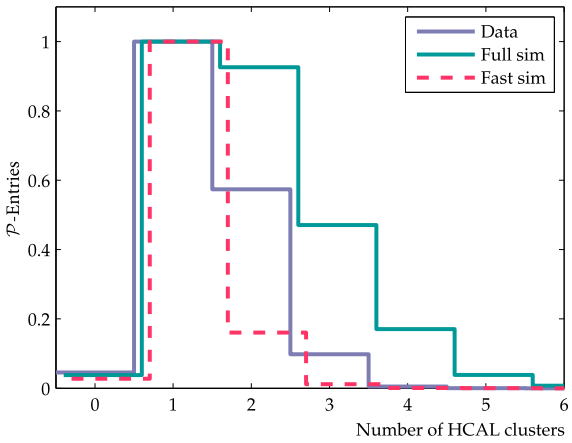
Consider the distribution, for example, at 50 GeV in the endcaps (Figure 4.18). This shows that neither the fast nor full simulations model the data's cluster multiplicity well.

FIGURE 4.17 **Cluster multiplicity**
Number of clusters per event in data and simulation for the HCAL barrel and endcap. In this plot, the error bars refer to the RMS of the source distribution; for legibility, these have only been plotted in the positive direction. Also, data and simulation are shown with their x coordinates slightly offset, so that the three points at each beam momentum may be clearly seen.



(a) Barrel *(b) Endcap*

FIGURE 4.18 **Endcap cluster multiplicity at 50 GeV**
Spectrum of the number of HCAL PFClusters for 50 GeV pions in the endcap. Data and simulation are shown with their x coordinates slightly offset, so that the three points in each bin may be clearly seen.



The cause of these effects has not yet been ascertained, but Figure 4.19 aims to quantify their implications. Recall that (a) HCAL cells only contribute to growing a cluster when an edge is shared with cells already in the cluster, and (b) particle flow unlinks all but the nearest HCAL cluster to the track if more than one cluster is present. If the original calibrated cluster energy is less than the track momentum, the satellite recovery algorithm will attach as many ECAL clusters as possible to the track while keeping $E < p$. No such algorithm currently exists for the HCAL: previous attempts (by the Author and others) to implement it can work at the level of single particle reconstruction but ruin the jet response. This indicates that the HCAL's transverse segmentation is too coarse for such an approach to work.

At very low energies it is quite rare that two or more HCAL clusters are found in data, but when it does happen, these clusters carry as much as 40% of the total cluster energy. These extra clusters will give rise to fake PF neutral hadrons; the fraction of calibrated calorimeter energy that is carried by these neutral PFCandidates is presented in the same figure. The clusters will generally be of lower energy than the track momentum, and the HCAL clusters will therefore be calibrated with a higher c coefficient than if they were attached to the track. This leads to the bulge in the total calibrated event energy in Figure 4.13 (b). For the endcaps, we see that these extra PF neutral hadron candidate energies form a substantial contribution to the total calibrated calorimeter energy. For both barrel and endcap, the fraction transported by photons is relatively small $\mathcal{O}(3\%)$.

The differences in cluster multiplicity between data and simulation may be due to variations in the way that hadronic cascades are simulated. There are encouraging indications that the full simulation reproduces the neutral PFCandidate energy contribution—which will allow the phenomenon to be studied—but a full comparison is deferred until the simulations' rechit responses are well tuned to the data.

▷ 4.6 Implications for jet reconstruction

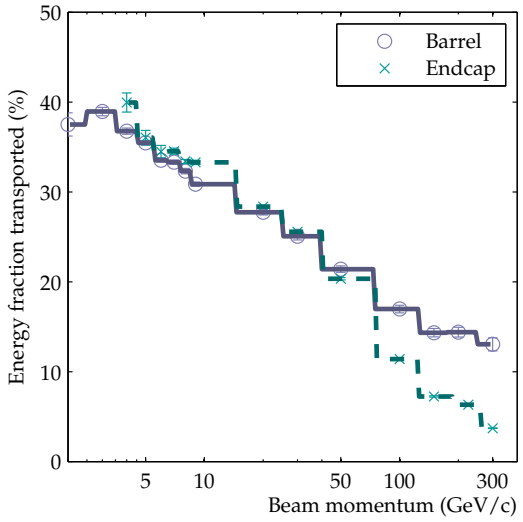
The PF Jet Benchmark

Let us now consider the implications for jet reconstruction. We want to know the systematic uncertainty that arises from using an inappropriate calibration. The calibration presented in this chapter corrects the non-linearity of the CMS calorimeter system that was exhibited at the rechit level; from Chapter 3, Figure 3.7 we know that the simulations and data are at odds with each other at $\mathcal{O}(10\%)$ for beam momenta greater than 10 GeV/c. What happens when a calibration derived from data is applied to jets simulated in the fast and full simulations?

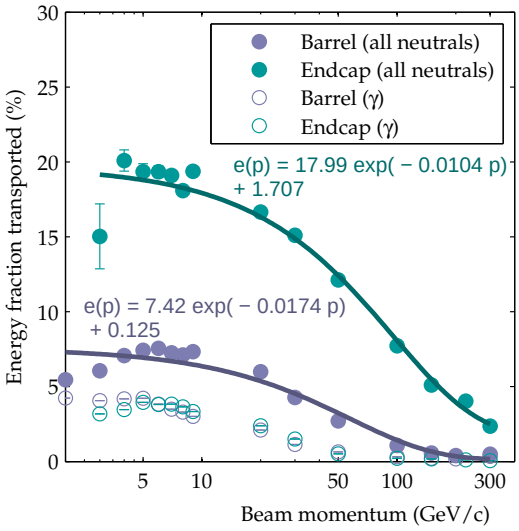
The *PF Jet Benchmark* is detailed in [41], but for our purposes the following description should suffice. QCD multijet samples are generated with PYTHIA and processed by the full and fast simulations according to §3.3. These QCD samples have a pseudo-flat $\hat{p_T}$ spectrum up to 1.5 TeV/c (where the hat on the p_T implies a reference to the transverse momentum of the parton-parton hard scatter). Jets are reconstructed using the *iterative cone algorithm* [26] with a cone size of $\Delta\mathcal{R} = 0.5$ from both the visible Monte Carlo truth particles[5] and from the PFCandidates reconstructed by the particle flow algorithm. The former are called *GenJets*, the latter *PFJets*.

[5] 'Visible' ⇒ excluding neutrinos.

FIGURE 4.19 **Extra HCAL cluster contamination**
Properties of non-nearest HCAL clusters to the track and implications for energy reconstruction. Note that the right-hand plot includes ECAL clusters that did not get used by the satellite recovery algorithm. A parametrization of the PF neutral candidate energies $e(p)$ is shown, for p the track momentum in GeV/c and e in percent.



(a) Fraction of total cluster energy transported by non-nearest HCAL clusters ($E^{\text{clusters}}_{2,3,\ldots}/E^{\text{clusters}}_{\text{all}}$) when there is more than one cluster.



(b) Energy transported by neutral PFCandidates ($E^{\text{cands}}_{\sum n^0,\gamma}/E^{\text{cands}}_{\text{all}}$) in all events.

The PFJets are compared with the GenJets with the quantity $(E_{\text{reco}} - E_{\text{true}})/E_{\text{true}}$, histogrammed as a function of $\hat{p_T}$. A Gaussian is fitted to each bin of p_T. The response is defined as the Gaussian mean, μ, and the resolution is $\sigma/(1+\mu)$.

Comparing data with simulation

In this section, the PF jet response of

▷ full simulation jets with a testbeam calibration applied,

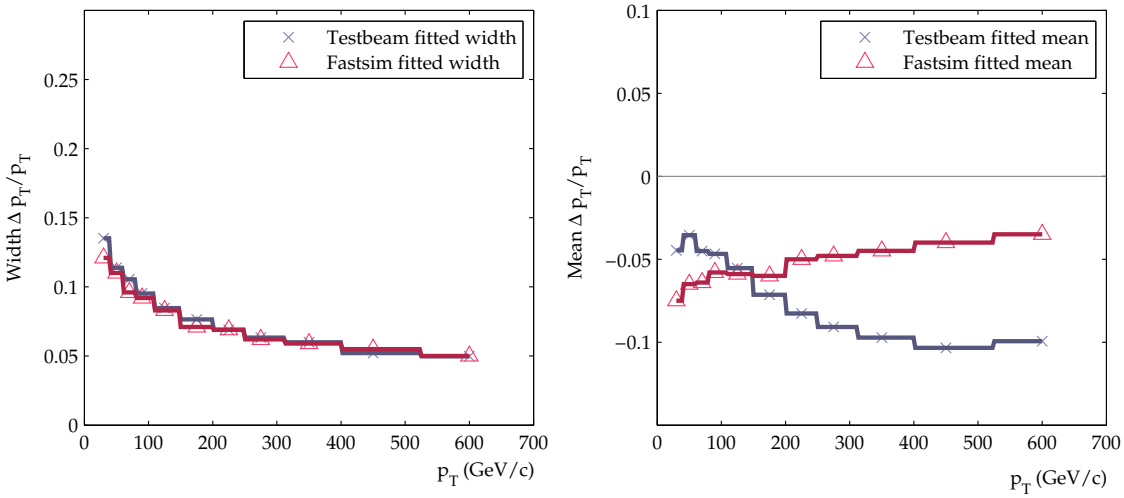
is compared with

▷ fast simulation jets with the standard fast simulation calibration applied.

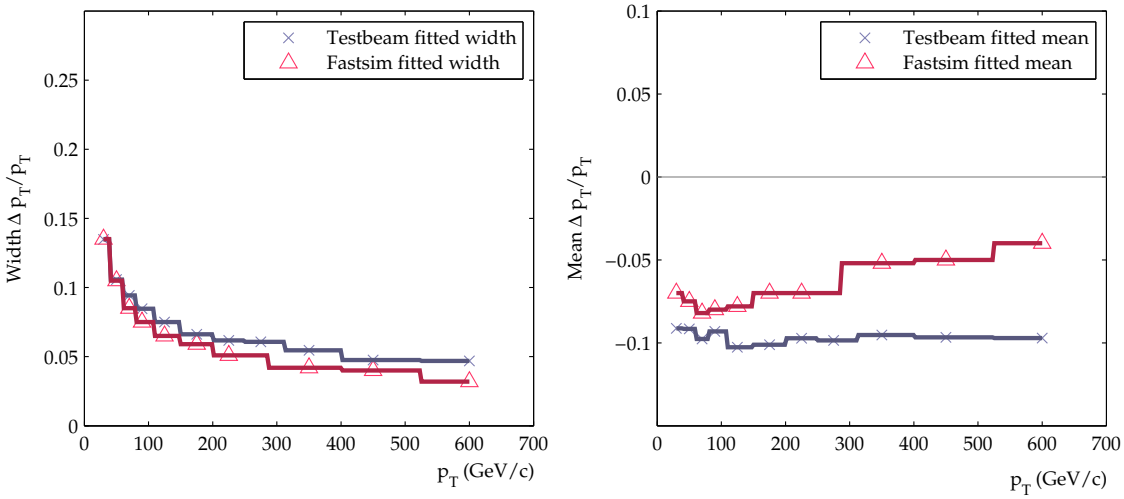
This comparison is somewhat inconsistent, inasmuch as we are comparing a miscalibrated (testbeam calibration on full simulation) system with a well calibrated system, but there are some interesting features nonetheless. The results of this exercise are presented in Figure 4.20. We observe that the provenance of the calibration does not substantially alter the width of the distributions in the barrel, but a slight degradation is seen in the endcap. (The PF jet response is predominantly Gaussian across the entire p_T range, whether using the testbeam calibration or the fast simulation calibration.)

The response is strongly affected: in §3.5.5, Figure 3.7, we saw that the full simulation underestimated the raw response by 10% for pions not interacting in the ECAL barrel, and a little less than this for those that do. Applying a testbeam calibration to full simulation rechits therefore produces an underestimate of the true hadron energy. Consider Figure

FIGURE 4.20 **PFJet Benchmark**
Comparison of full simulation QCD jets with a testbeam-derived calibration applied and fast simulation QCD jets with the standard fast simulation calibration applied.



(a) Barrel



(b) Endcap

4.8 and Figure 4.14: from 20 GeV onwards, the HCAL testbeam coefficients tend to 1.0, while the full simulation HCAL coefficients tend to 1.1. The testbeam's calibration applied to full simulation jets produces PFJets with less calibrated energy than fast simulation jets with the fast simulation calibration applied.

In the endcaps, again, the PFJet response using the testbeam calibration is consistent with the behaviour seen in Figure 3.7. It has been noted however that there are problems with the full simulation's modelling of cluster multiplicity, so it may be ill-advised to draw firm conclusions until this is addressed.

Finally, applying the fast simulation calibration to the same full simulation jets (not shown here, see [41]) is relatively consistent, because the full and fast simulations agree save at very low energies in the endcap. The PFJet Benchmark looks broadly similar for full and fast simulation jets when the fast simulation calibration is applied.

Comparison with purely calorimetric methods

CaloJets were previously mentioned as a purely calorimetric source of finding jets. These use uncalibrated calo rechits as inputs, and therefore require substantial corrections to restore the response $E_{\text{reco}}/E_{\text{true}}$ to unity across the range of generated p_T. The so-called *Level-2/Level-3 Jet Corrections* [84] have been applied to the CaloJets in the analysis that follows.

CMS adopts a factorized scheme to correct jets: the first step compensates for calorimeter noise and pile up, the second equalizes the response in η, and the third normalizes the reconstructed jet response to the generated jet. The corrections are small for particle flow, but substantial for CaloJets because they need to correct for the calorimeter's non-linearity.

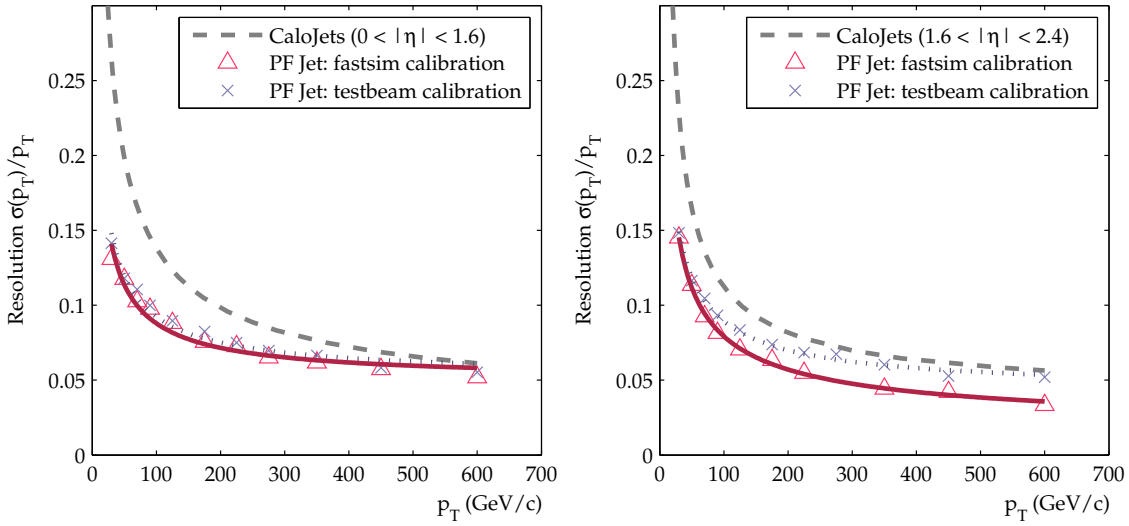
The resolution of CaloJets can be parametrized [82, 85] for the iterative cone algorithm, and this has been plotted alongside the PFJet resolution in Figure 4.21. Particle flow's treatment improves the resolution substantially across the entire range of p_T.

Resolution functions describing each scenario are shown in Table 4.4. The table quotes three sets of parametrizations:

- ▷ PFJet resolution in the full simulation using the testbeam-derived calibration;
- ▷ PFJet resolution in the fast simulation;
- ▷ CaloJet resolution with corrections.

Evidently, the resolution is seriously degraded in the endcaps where the testbeam calibration has been applied to the full simulation: for jets with $p_T > 200\,\text{GeV}/c$, the jet resolution is compromised as a result of mismatched calibrations. To reiterate, (a) the calibration is inappropriate to the rechits being calibrated, (b) the problem of energetic HCAL satellite clusters—and subsequent impact on calibration—has not been addressed, and (c) the difference between simulation and data is not limited to the *scale* of the response, but will also include a difference in the *relative weights* of the calibration assigned to the ECAL and HCAL components. We should not be surprised by this result.

FIGURE 4.21 **PF Jet Benchmark: resolution**
The testbeam-derived calibrations detailed in §4.4 were applied in the reconstruction of the full simulation jets. The inconsistency of this method is exposed in the endcaps where the simulations do not model the data's cluster multiplicity and energy response well. The fast simulation jets, calibrated with a fast simulation calibration, show a substantial improvement on the CaloJet method.



(a) Barrel *(b) Endcap*

TABLE 4.4 **Jet energy resolutions**
Parametrized as a function of generated p_T. Jets were made using the iterative cone algorithm with $\Delta\mathcal{R} = 0.5$. A fair comparison can be made between the last two columns, whereas the first column is shown for completeness.

	PFJet (TB calib on full sim)	PFJet (fast sim)	CaloJet (L2L3 corrected)
Barrel	$\frac{\sigma}{p_T} = \frac{0.0}{p_T} \oplus \frac{0.759}{\sqrt{p_T}} \oplus 0.052$	$\frac{\sigma}{p_T} = \frac{0.0}{p_T} \oplus \frac{0.722}{\sqrt{p_T}} \oplus 0.050$	$\frac{\sigma}{p_T} = \frac{3.41}{p_T} \oplus \frac{1.30}{\sqrt{p_T}} \oplus 0.03$
Endcap	$\frac{\sigma}{p_T} = \frac{0.0}{p_T} \oplus \frac{0.782}{\sqrt{p_T}} \oplus 0.0430$	$\frac{\sigma}{p_T} = \frac{1.07}{p_T} \oplus \frac{0.766}{\sqrt{p_T}} \oplus 0.0170$	$\frac{\sigma}{p_T} = \frac{4.22}{p_T} \oplus \frac{0.96}{\sqrt{p_T}} \oplus 0.04$

▷ 4.7 Summary

A summary of the last two chapters is shown in Figure 4.22.

The particle flow method

- ▷ Particle flow exploits information in each CMS subdetector to give a list of PFCandidates which describe the outcome of the original parton-parton collision;
- ▷ We considered how clustering applies thresholds to rechits and groups them topologically. These clusters are then calibrated to correct their response. Clusters are then compared to tracks to determine whether the cluster energy(ies) are statistically consistent with the track momentum. In this way, we substitute the poor calorimetric energy measurement with a more precise and accurate momentum measurement.

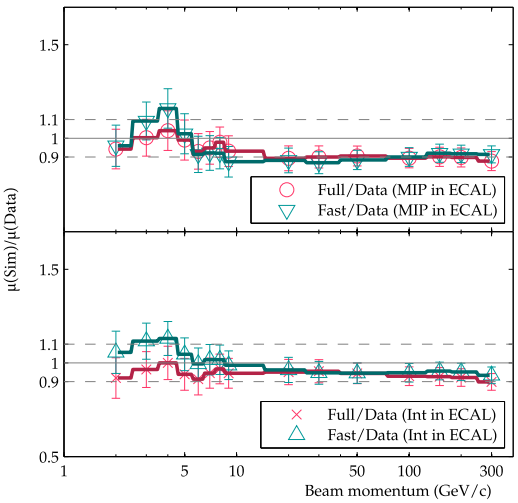
Particle flow with testbeam

- ▷ A calibration specific to the two testbeams was created and shown to be effective in restoring linearity and $E_{\text{reco}}/E_{\text{true}} \to 1$;
- ▷ The calibration improves the energy resolution of the combined ECAL/HCAL system;
- ▷ Various features of the particle flow output were described, such as fake photons and extra neutral hadrons. Satellite clusters which are not attached to the charged hadron's track give rise to a substantial amount of 'fake' neutral PFCandidate energy.

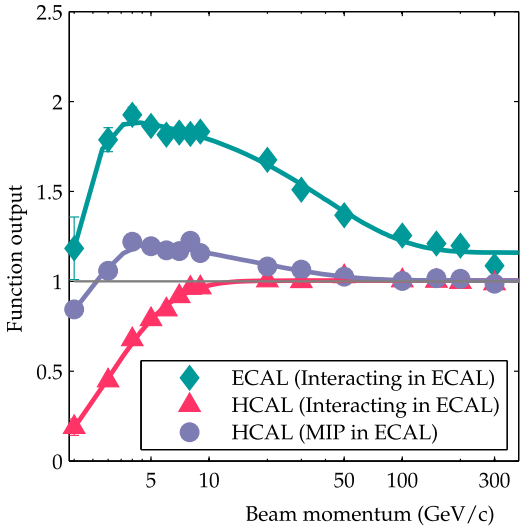
Jet energy response

- ▷ Calibrations to correct PFClusters specific to data & simulation isolate us from differences in data & simulation at the rechit level;
- ▷ In contrast to CaloJets, PFJets require only modest corrections to the overall jet energy scale. The resolution of correctly calibrated PFJets is superior to CaloJets;
- ▷ The effect of using a testbeam calibration on full simulation QCD jet events was evaluated. The change in jet energy response was in agreement with expectations. The simulations require retuning both of their high-energy responses and reproduction of cluster multiplicities, particularly in the endcap where the differences and inconsistencies degrade the energy resolution substantially.

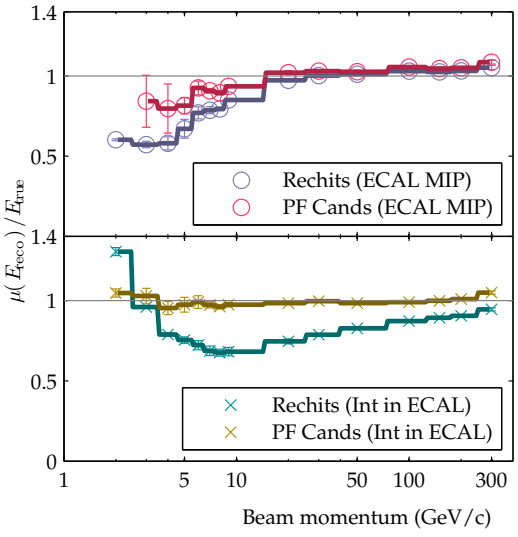
FIGURE 4.22 **Summary of Chapters 3 & 4**
How variations at the level of rechits in simulations and data propagate through to calibration and to jet energy response. All plots relate to the barrel.



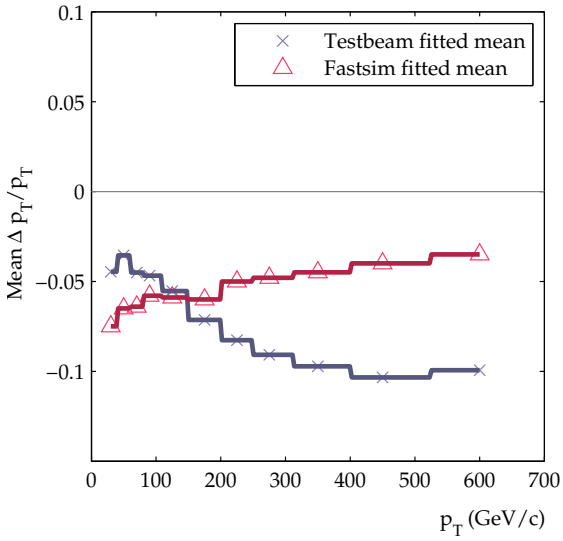
(a) The simulations do not reproduce the calorimeters' response in testbeam. For the barrel, the variation is of $\mathcal{O}(10\%)$ underestimation for pions of 20 GeV and above.



(b) We need a calibration specific to each of the simulations and to the testbeam…



(c) …to get calibrated PFCandidates which correct the calorimeters' non-linearity.



(d) Differences between simulation and data rechit response manifest themselves as differences in jet energy response when a calibration specific to data is used to calibrate hadrons in simulation.

Chapter 5

Particle flow hadron response in first collisions

▷ 5.1 Introduction

This short chapter provides a first look at particle flow charged hadrons collected from collision data taken at the end of 2009. The analysis compares the response using the standard set of calibration coefficients derived from the fast simulation with the calibration derived by the Author based on testbeam data.

▷ 5.2 Early collision data

The LHC collided beams in 2009 for 4 weeks, giving each of the 4 main experiments an opportunity to record data. Most of the data were taken at a centre of mass energy of $\sqrt{s}$ = 900 GeV, but some were also taken at a world record beating $\sqrt{s}$ = 2.36 TeV. The tracking system was aligned using cosmic muon events recorded over several months [86]. Over 97% of the tracking system was functional.

5.2.1 Minimum bias events

Most pp collisions at the centre of mass energy of 900 GeV are merely soft glancing collisions (described by either elastic or 'diffractive' processes), where any hadrons produced are very soft [87]. The *minimum bias* trigger accepts all inelastic collisions including hard scattering events, but rejects elastic collisions and single diffractive (SD) events of little interest; this trigger provides the dataset considered in this chapter.

Monte Carlo simulation

A dataset of 2M fully simulated minimum bias events has also been made available using realistic alignment and miscalibrations. This is analysed alongside the real data in this chapter where appropriate.

5.2.2 A reminder of differences with the testbeam environment

In collisions, the CMS detector operated with a magnetic field of 3.8T. Scintillator brightening increases the HCAL response by several percent, but this effect has been accounted for in the rechit calibration and subsequently removed. Selective readout was active in the ECAL readout. Particles emanating from minimum bias collisions are uniformly distributed over the central rapidity region $|\eta| < 2.5$ covered by the tracker acceptance [88]. Finally, it is worth noting that the amount of dead material between the ECAL and HCAL barrel increases substantially with $|\eta|$, which will change the overall HCAL response.

5.2.3 Event selections

Technical selection of events from collision data

Events were reconstructed by the Particle Flow Group. Selections required to clean the data and apply the particle flow reconstruction are detailed in [89]. These selections yield a total of 412,517 potential collision events for analysis.

Selecting charged hadron PFCandidates

In contrast to many other reconstruction techniques, PF does not strictly require conventional selections of single isolated particles from events: rather, we rely on the PF technique to separate calorimeter energy due to leptons and hadrons to obtain a sample of charged hadrons. Work is underway in the PF group to estimate the contamination of the reconstructed sample arising from neutral particles overlapping with the charged hadron calorimeter deposits. While only $\mathcal{O}(10\%)$ of jet energy is transported by neutral hadrons, implying the subsequent contamination to be small, the main source of contamination at high momenta is due to overlapping π^0 mesons. When comparing with testbeam, we must consider only the energy associated with the charged hadron PFCandidates, given by Eq. 4.10, for we cannot know in the collision environment whether photons or neutral hadron PFCandidates in the vicinity are a consequence of satellite clusters or the original pp collision. Finally, as in Chapter 4, the σ cut in this analysis has been relaxed to 5σ to enable direct comparison with testbeam data.

Quality cuts

The charged hadron PFCandidates presented here represent *all* such candidates found in collisions, subject to the following quality cuts:

- A $p_T > 1\,\text{GeV}/c$ cut has been applied to remove very soft particles;
- The track has at least two hits in the vertex detector and at least 14 hits overall. This ensures that the tracks are well reconstructed.
- Define: barrel $\Rightarrow |\eta| < 1.0$, endcap $\Rightarrow 1.4 < |\eta| < 2.5$;

▷ We require all hadrons to have a link to at least one HCAL cluster, as in Chapter 4. There is a substantial number of 'ECAL-only' pions, and in the medium term their situation should be better addressed (see §5.4).

Clearly the track quality cut and calorimeter energy association cuts are important to guarantee the quality of the sample. Figure 5.1 shows the number spectrum as a function of $|\eta|$ and the effect of the track quality and calorimeter energy cuts. The cut 'survival rates' are shown in Table 5.1. Table 5.2 shows the number of charged hadron PFCandidates collected from the analysed runs after the cuts have been applied.

TABLE 5.1 **Quality cuts applied to charged hadron PFCandidates in collisions**
Values are normalized to the first basic cut of $p_T > 1\,\text{GeV}/c$. See text for details.

	Barrel		Endcap	
Cut	Number	%	Number	%
$p_T > 1\,\text{GeV}/c$	177,383	100.0	185,260	100.0
Track quality	124,208	70.0	111,388	60.1
At least one PF cluster	34,402	19.3	49,812	26.9
At least one HCAL cluster	12,947	7.2	23,411	12.6

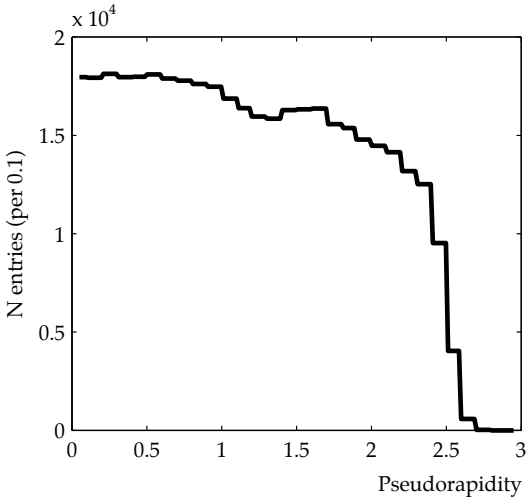
Figure 5.1 exhibits some interesting features: evidently, the track quality cut attenuates the distribution in the region of $0.7 < |\eta| < 1.2$, for this is at the join of the tracker barrel and endcap regions (Figure 1.6). We also observe more charged hadron tracks with clusters in the endcap regions. This is mostly due to the p_T cut, whereby particles with a p_T of $1\,\text{GeV}/c$ in the endcaps have more energy than a hadron of the same p_T in the barrel. The decline of the distribution in subfigure (d) in the barrel region is likely attributed to the decrease in HCAL noise[1] from 300 GeV to 180 GeV over the range $|\eta| = 0$ to $|\eta| = 1.4$ [90], the increase in the number of interaction lengths presented by tracking volume material over that range (see Figure 1.7), and the increase in dead material between the ECAL and HCAL barrel.

TABLE 5.2 **Statistics available after applying quality cuts**
Total number of charged hadron PFCandidates available from runs taken in 2009.

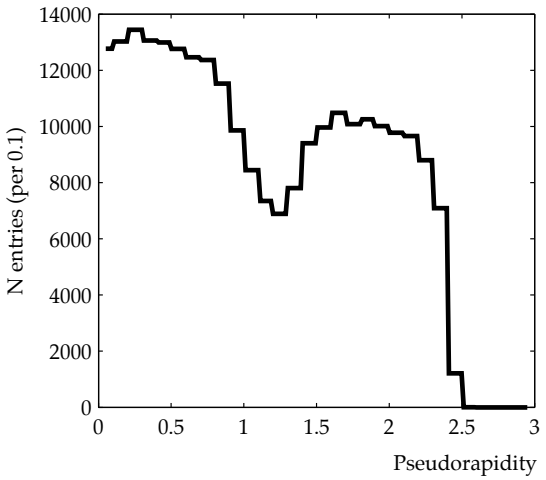
p_{track} (GeV/c)	**Barrel**	**Endcap**
1–2	8,200	21
2–3	3,071	1,888
3–4	1,025	4,227
4–5	370	4,842
5–6	165	4,413
6–7	61	2,893
7–8	21	1,795
8–10	22	1,894
10–13	9	931
13–17	2	367
17–25	1	140
$\sum$	12,947	23,411

[1] The HCAL noise measurement from testbeam is still valid, but is only specific to one η point.

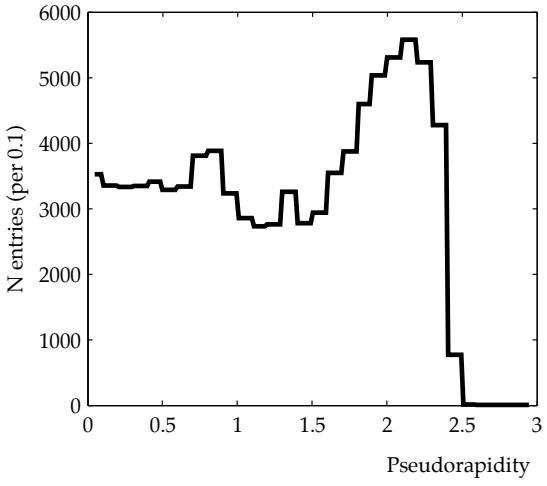
FIGURE 5.1 **η distributions of PF hadrons from data**
Requiring at least one HCAL cluster reduces the number of events substantially. The same pattern is seen in Monte Carlo. See text for details.



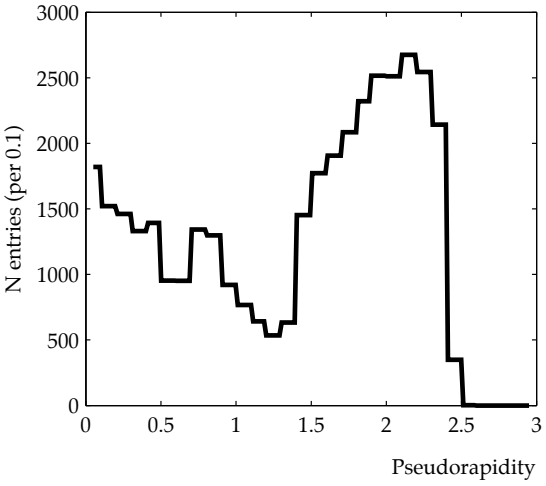
(a) $|\eta|$ spectrum without quality cuts



(b) $|\eta|$ spectrum with track quality cut



(c) $|\eta|$ spectrum with track quality & at least one cluster cut



(d) $|\eta|$ spectrum with track quality & at least an HCAL cluster

▷ 5.3 Energy response of charged hadrons

In interpreting the results that follow, please refer to Figure 3.7 on page 74.

5.3.1 Global energy response

Figures 5.2 and 5.3 shows the ratio $E_{\text{reco}}/p_{\text{track}}$ for PF charged hadrons collected from collisions for barrel and endcap and under four scenarios: Figure 5.2 (a, b) shows collision data with the application of standard fast simulation calibration coefficients while (c, d) shows the equivalent result from the full simulation of the event. Figure 5.3 (a, b) shows collision data with the application of the calibration derived from testbeam while (c, d) plots the results from testbeam according to Eq. 4.10, *where the satellite neutral PFCandidate contributions have been removed.*

Variable bin sizes have been employed to better exploit the statistics available.[2]

Interpretation

Given the limited statistics available we can only draw conclusions about the low energy response. In this regime, it is clear from Figure 5.2 that the full simulation reproduces the data well, which is in agreement with our expectations from Figure 3.7. From Figures 5.2 (b, d) it is evident that the fast simulation calibration does not correct the energy response satisfactorily in the endcaps for data or the full simulation. Again, from Figure 3.7, the fast simulation overestimates the hadron response so the calibration coefficients are smaller in magnitude than they should be, and the overall response is undercorrected.

Figure 5.3 compares the testbeam-derived calibration of the collision data with the equivalent testbeam results: subfigure (a) should be compared to (c) for the barrel, and (b) is to be compared to (d) for the endcap. In this comparison, the collision data are subject to a large amount of material between the collision vertex and the calorimetry (Figure 1.7), which particularly affects the endcap region. Because the satellite neutral PFCandidates have been removed, we do not expect to recover $E_{\text{reco}}/p_{\text{track}} = 1$. In the barrel, for $p_{\text{track}} < 20\,\text{GeV}/c$, we can expect the ratio to be diminished by ~ 6%. In the endcaps, the ratio will be ~ 18%.

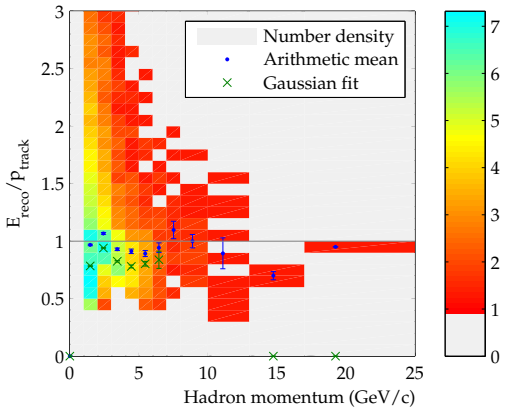
In the barrel, the mean Gaussian response of $\bar{\mu} = 0.95$ in collisions should be compared to $\bar{\mu} = 0.80$ in testbeam. For the endcap, compare $\bar{\mu} = 0.84$ in collisions with $\bar{\mu} = 0.79$ in testbeam data. From Figure 5.3, it is apparent that the ratios are in qualitative agreement with each other and with expectations.

5.3.2 Detailed energy response

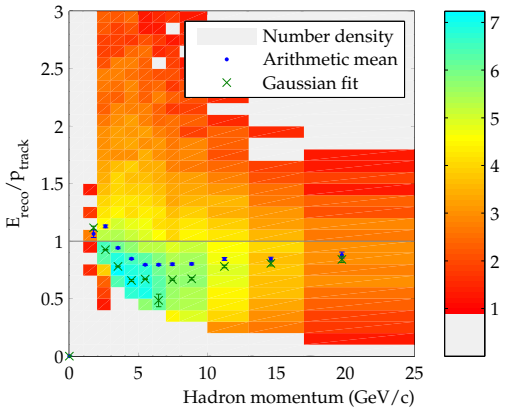
Figures 5.4–5.7 on the following pages show the detailed energy response of charged hadrons at low energies. Following the example of the previous chapter, we consider separately barrel and endcap, and hadrons with and without at least one ECAL cluster. Four plots of $E_{\text{reco}}/p_{\text{track}}$ are shown on each page, according to the following scheme:

[2] Note that the x-axis position of points plotted in these figures is at the arithmetic mean of the entries in that bin, and *not* at the standard bin centre. The energy spectrum of pions is exponentially falling over the extent of the bins, so making this correction is important.

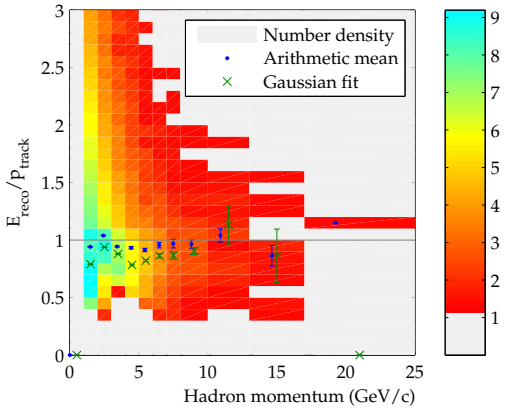
FIGURE 5.2 **$E_{\text{reco}}/p_{\text{track}}$ for PF hadrons from collisions (I)**
These plots show the calibrated PFCandidate energies using the standard fast simulation-derived coefficients. The arithmetic and Gaussian fitted mean for each bin in p has been plotted on top of the distribution. The error bars indicate the error of means. *Note:* The z-scale in these figures is logarithmic.



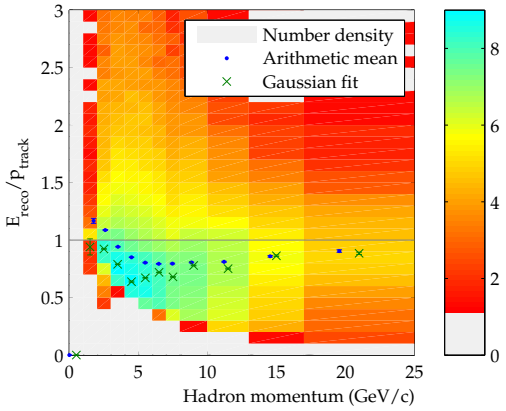
(a) Barrel collision data, with the fast simulation calibration



(b) Endcap collision data, with the fast simulation calibration

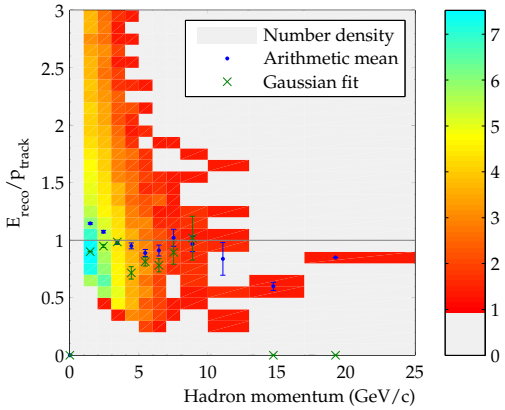


(c) Barrel full simulation, with the fast simulation calibration

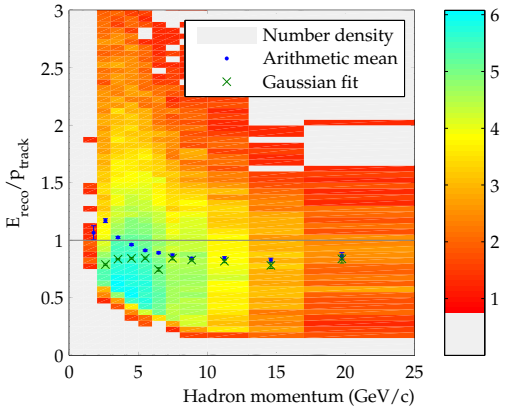


(d) Endcap full simulation, with the fast simulation calibration

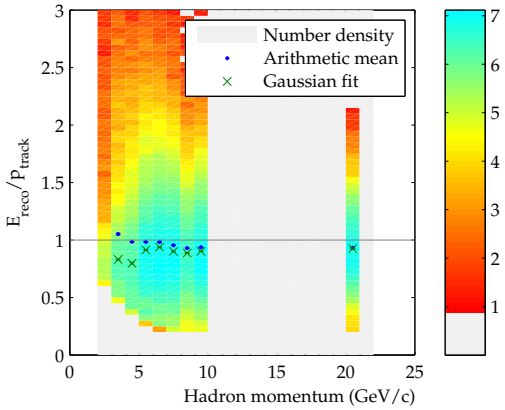
FIGURE 5.3 $E_{\text{reco}}/p_{\text{track}}$ **for PF hadrons from collisions (II)**
These plots show the calibrated PFCandidate energies using the testbeam-derived coefficients. The arithmetic and Gaussian fitted mean, μ, for each bin in p has been plotted on top of the distribution. The error bars indicate the error of means. A cut of $E_{\text{reco}}/p_{\text{track}} > 0.2$ has been applied. *Note:* The z-scale in these figures is logarithmic. The number of events in each bin of p_{track} is not suited for comparison between plots.



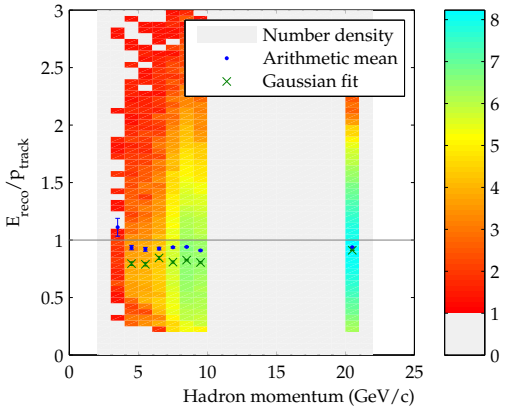
(a) Barrel collision data with a testbeam calibration. $\bar{\mu} = 0.95$



(b) Endcap collision data with a testbeam calibration. $\bar{\mu} = 0.84$



(c) Barrel testbeam data with a testbeam calibration, plotting charged hadron response only. $\bar{\mu} = 0.80$



(d) Endcap testbeam data with a testbeam calibration, plotting charged hadron response only. $\bar{\mu} = 0.79$

(a) Collision data with the standard fast simulation calibration coefficients applied;

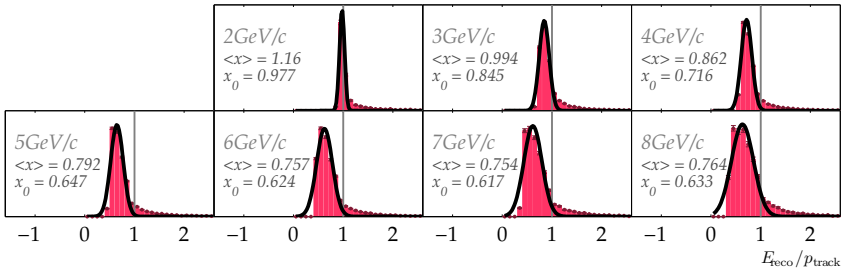
(b) Collision data with the calibration coefficients derived from testbeam data applied;

(c) A reproduction of the testbeam charged hadron PFCandidate spectra, with the testbeam calibration applied;

(d) The full simulation of the minimum bias events, where the standard fast simulation calibration coefficients have been applied.

Interpretation

There are some features of the response where the standard fast simulation calibration has been applied (subfigures (a, d) in Figures 5.4–5.7) that should be noted: first, the apparently narrow width is an artefact of a high a term in the calibration and lower b, c coefficients. The calibrated energy is more often dominated by a fixed constant, rather than the cluster energies. This is evidenced by a systematic decrease—which in turn decreases in magnitude—in the (Gaussian) response as the pion momentum increases. For example, Figure 5.6 (d) (endcap, simulation, no ECAL clusters),



Second, the fast simulation b, c calibration coefficients are negative at $E_{\text{eval}} \lesssim 3$ (Eq. 4.3), which gives an odd double-peak shape to the 2 GeV/c plot of Figure 5.4 (a).

The testbeam calibration produces shapes which are broadly similar to the testbeam response (compare subfigures (b, c) in Figures 5.4 to 5.7) which is encouraging given the known differences between the collision environment and the testbeams. While the Gaussian fits are lop-sided (the p_T cut removes low values of $E_{\text{reco}}/p_{\text{track}}$, particularly in the endcap), the arithmetic and Gaussian means are in better agreement with the testbeam response. In some of the plots, the contribution of the a term is more pronounced in testbeam because the pion energy is monoenergetic in that bin (so a/p_{track} is single valued), whereas in collision data each plot covers $p_{\text{plot}} \pm 0.5$ GeV/c so the a term contribution is smeared over this range too.

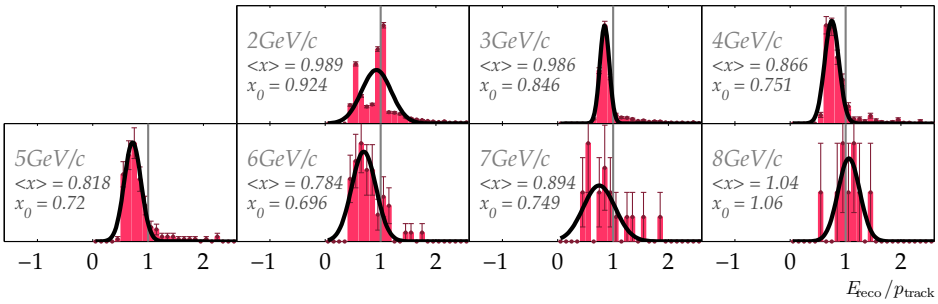
An estimation of the contamination arising from π^0 mesons and neutral hadrons will also be enlightening.

▷ 5.4 Strategies for improving PF hadron reconstruction

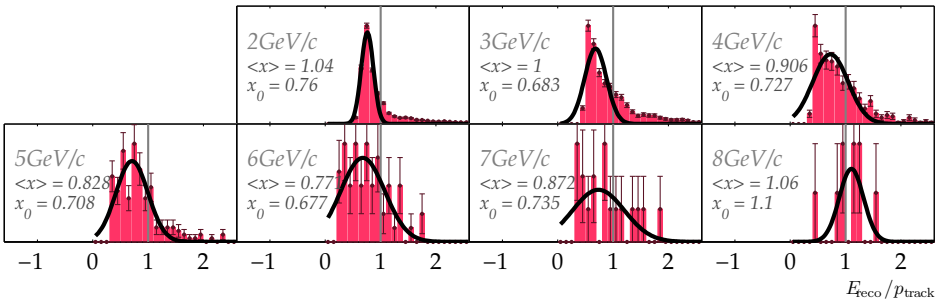
5.4.1 PF hadrons without HCAL clusters

In this chapter and the previous one, it was noted that a substantial fraction $\mathcal{O}(25\%)$ of hadrons at low energy have only ECAL clusters linked to the track. These hadrons are

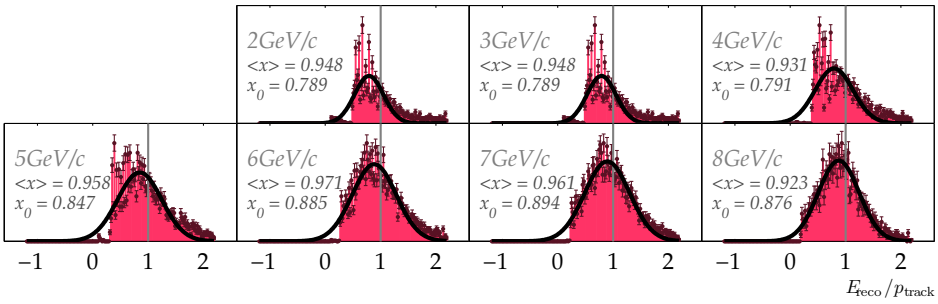
FIGURE 5.4 **Barrel PFCandidate energy spectra from collisions: no clusters in the ECAL**
As in Chapters 3 and 4, the histograms have all been normalized to their peak values.



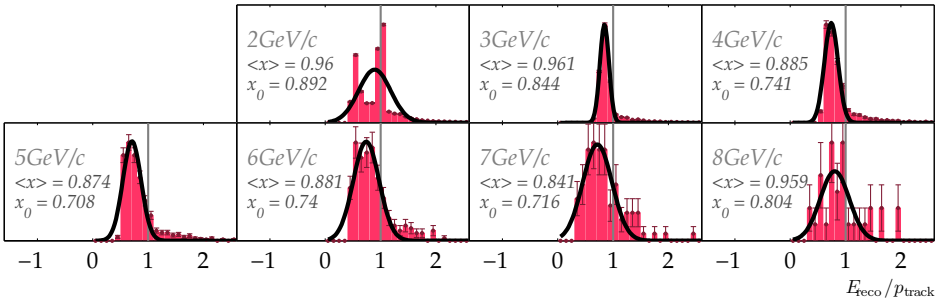
(a) Barrel PFCandidates with 0 ECAL clusters, collision data with the fast simulation calibration coefficients applied



(b) Barrel PFCandidates with 0 ECAL clusters, collision data with the testbeam calibration coefficients applied

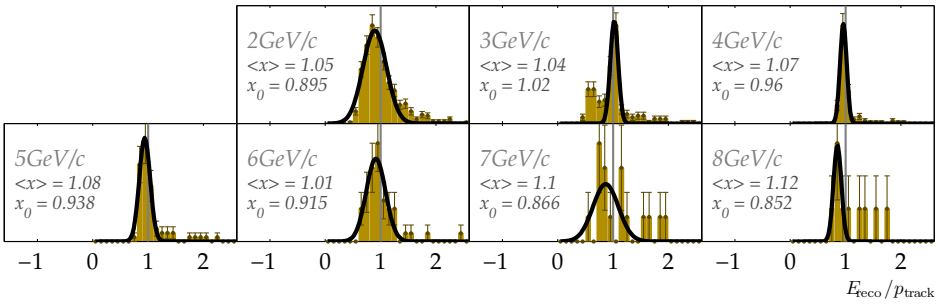


(c) Barrel PFCandidates with 0 ECAL clusters, testbeam data with the testbeam calibration coefficients applied

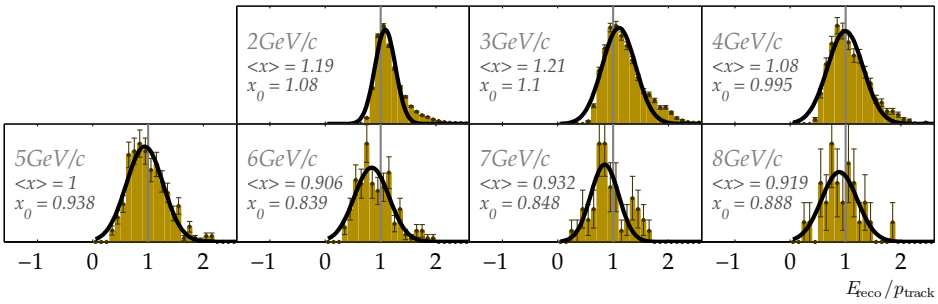


(d) Barrel PFCandidates with 0 ECAL clusters, full simulation with the fast simulation calibration coefficients applied

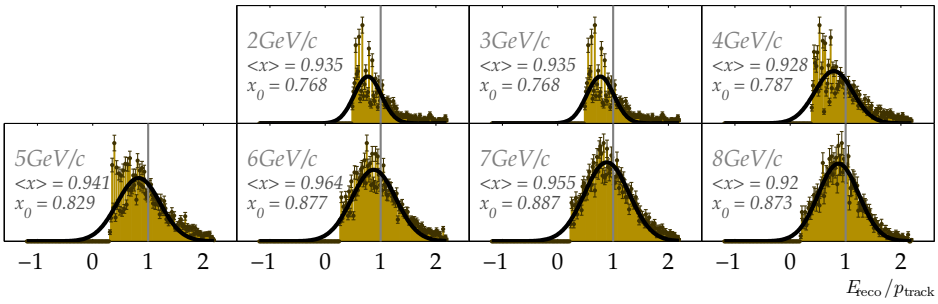
FIGURE 5.5 **Barrel PFCandidate energy spectra from collisions: with clusters in the ECAL**
As in Chapters 3 and 4, the histograms have all been normalized to their peak values.



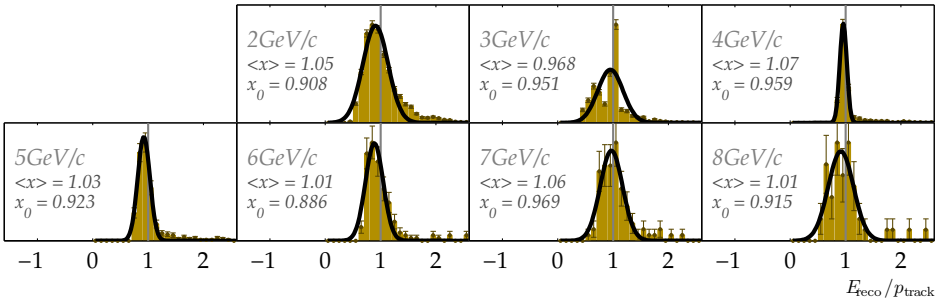
(a) Barrel PFCandidates with > 0 ECAL clusters, collision data with the fast simulation calibration coefficients applied



(b) Barrel PFCandidates with > 0 ECAL clusters, collision data with the testbeam calibration coefficients applied

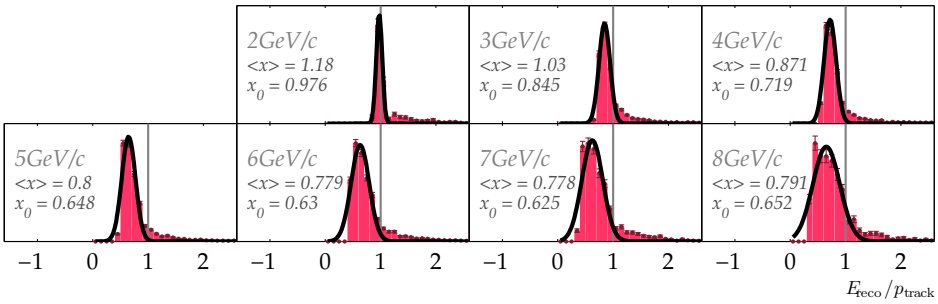


(c) Barrel PFCandidates with > 0 ECAL clusters, testbeam data with the testbeam calibration coefficients applied

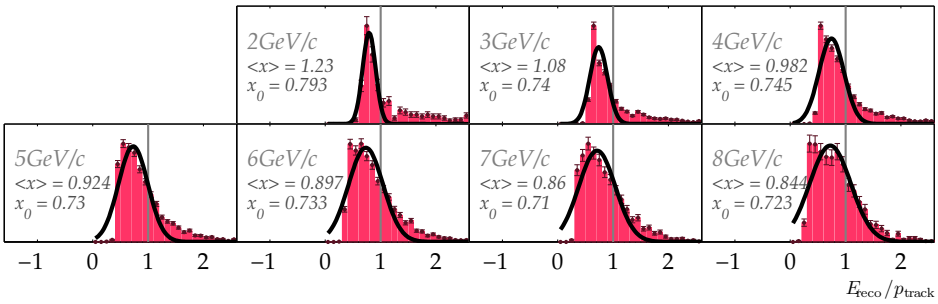


(d) Barrel PFCandidates with > 0 ECAL clusters, full simulation with the fast simulation calibration coefficients applied

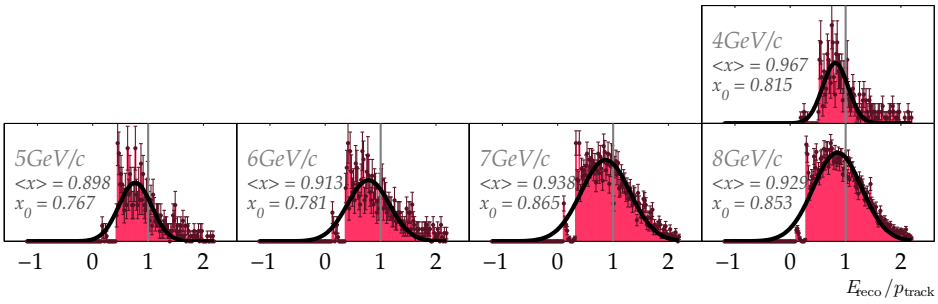
FIGURE 5.6 **Endcap PFCandidate energy spectra from collisions: no clusters in the ECAL**
As in Chapters 3 and 4, the histograms have all been normalized to their peak values.



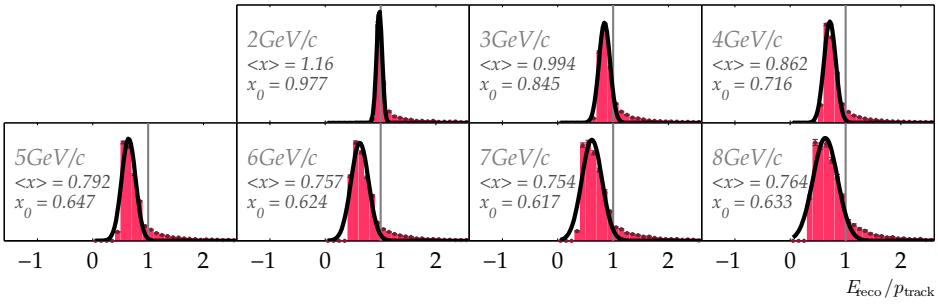
(a) Endcap PFCandidates with 0 ECAL clusters, collision data with the fast simulation calibration coefficients applied



(b) Endcap PFCandidates with 0 ECAL clusters, collision data with the testbeam calibration coefficients applied

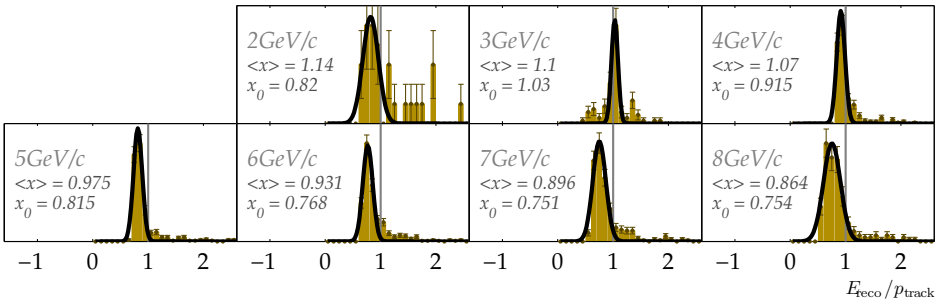


(c) Endcap PFCandidates with 0 ECAL clusters, testbeam data with the testbeam calibration coefficients applied

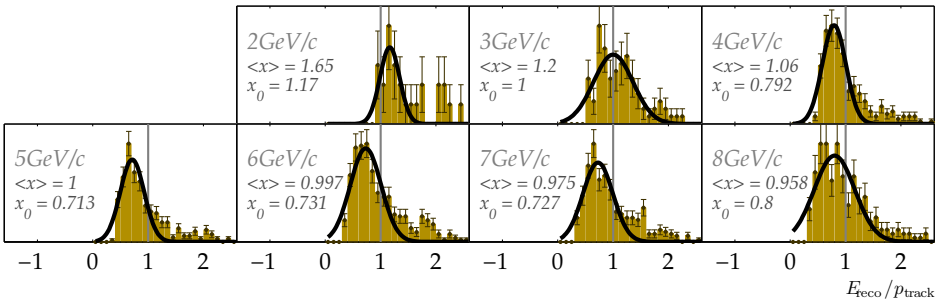


(d) Endcap PFCandidates with 0 ECAL clusters, full simulation with the fast simulation calibration coefficients applied

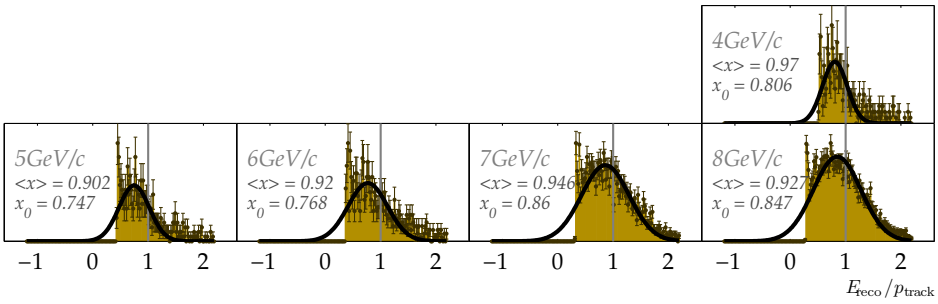
FIGURE 5.7 **Endcap PFCandidate energy spectra from collisions: with clusters in the ECAL**
As in Chapters 3 and 4, the histograms have all been normalized to their peak values.



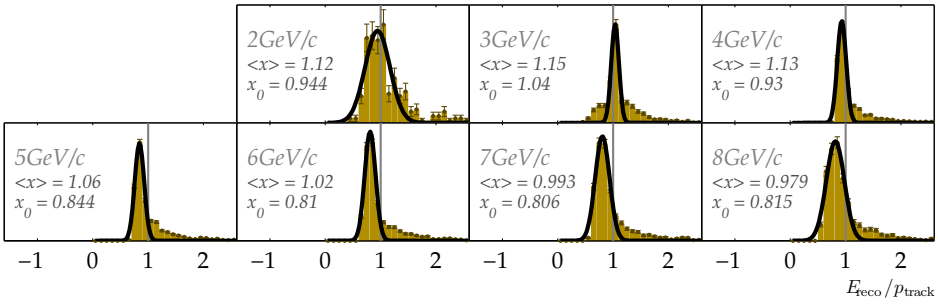
(a) Endcap PFCandidates with > 0 ECAL clusters, collision data with the fast simulation calibration coefficients applied



(b) Endcap PFCandidates with > 0 ECAL clusters, collision data with the testbeam calibration coefficients applied



(c) Endcap PFCandidates with > 0 ECAL clusters, testbeam data with the testbeam calibration coefficients applied



(d) Endcap PFCandidates with > 0 ECAL clusters, full simulation with the fast simulation calibration coefficients applied

currently calibrated according to the 'ECAL and HCAL' calibration coefficients, but with $E_{\mathrm{HCAL}} = 0$, and at low energy the reconstructed energy is underestimated by $\mathcal{O}(20\%)$ in the endcap. In the barrel, the difference is negligible. Preliminary investigations have shown that the standard 'ECAL and HCAL' calibration coefficients suffice, but a larger a term is appropriate $\mathcal{O}(1\,\mathrm{GeV})$, to account for the energy below threshold in the HCAL.

5.4.2 Surplus PF neutral hadrons and PF photons

The contamination arising from surplus clusters not linked to the charged track, and their subsequent manifestation as PF neutral hadrons and PF photons is substantial. Figure 4.19 parametrizes the energy fraction $e(p)$ transported by these candidates as a function of beam momentum p,

$$e(p) = 7.42\% e^{-0.0174p} + 0.125\% \quad \text{(barrel)} \tag{5.1}$$

and,

$$e(p) = 17.99\% e^{-0.0104p} + 1.707\% \quad \text{(endcap)}. \tag{5.2}$$

The electromagnetic contribution is small for both, and the problem is attributed to secondary HCAL clusters. Four possible strategies are:

(1) Modify the HCAL clustering algorithm: allow cells sharing a corner with a growing cluster to contribute to the cluster, rather than requiring an edge to be shared;

(2) Implement a form of HCAL cluster recovery: rather than unlink all but the nearest HCAL clusters from the track, PFAlgo should add them to the calibrated energy calculation to satisfy the E/p comparison as it does in the ECAL;

(3) Calibrate the charged hadron component alone, rather than all the clusters in $\Delta\mathcal{R} = 0.15$. Modify jet reconstruction on some macroscopic basis to accommodate the 'fake' neutrals;

(4) Preserve the *status quo*, where all clusters within $\Delta\mathcal{R} = 0.15$ are calibrated on a global basis, but adjust the E/p comparison in PFAlgo so that, for a given charged hadron, one expects its calibrated calorimeter energy to not satisfy $E = p$ but,

$$E = [1 - e(p)]p \tag{5.3}$$

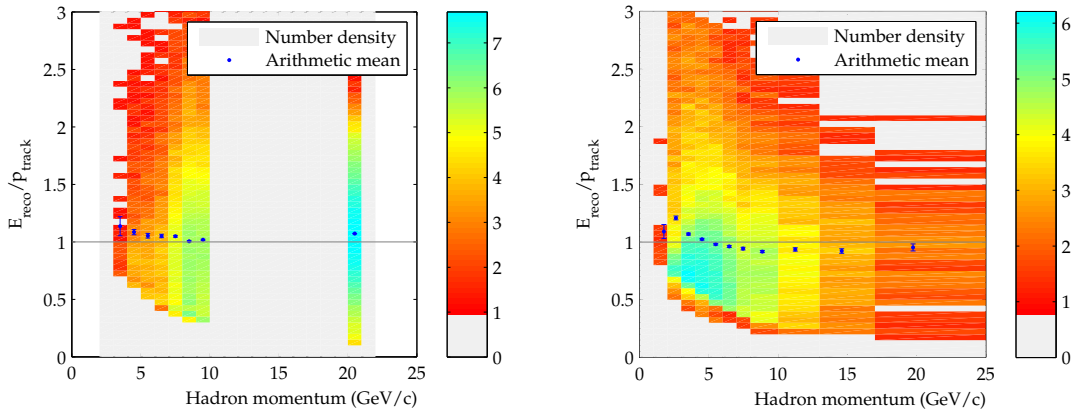
for $e(p)$ given by Eqs. 5.1, 5.2.

The first two strategies are likely to have serious side-effects for PF jet reconstruction. As mentioned earlier, previous attempts to implement a form of HCAL cluster recovery did not succeed. Strategy (3) is certainly feasible. Figure 5.8 shows what happens when only the clusters attached to the charged track are used for calibration (the procedure is otherwise identical to that described in Chapter 4). On average, the calibration coefficients are slightly higher than under the global calibration strategy because less cluster energy is to be calibrated. Figure 5.8 plots only reconstructed charged hadron energy (Eq. 4.10) and it is clear that such a procedure is successful in both testbeam and collision data in recovering $E/p = 1$. The advantage to such a procedure is that it is consistent with the way in which PFAlgo *applies* the calibration—there is no bulge in

the mid range of beam momenta evidenced by Figure 4.13 (b) when only the charged hadron component is considered. The disadvantages are that (a) the surplus neutral PFCandidates are still there, (b) their energies are higher than they were before, which exacerbates the situation, and (c) it is not clear how jet reconstruction should be modified to account for this fake excess. It is the Author's view that Strategy (4) is likely to be the most fruitful for study. It will, however, require extensive modification to the PF algorithm and subsequent testing and validation.

FIGURE 5.8

Calibrating only the charged hadron's clusters ('Strategy 3')
Only the clusters linked to the charged hadron's track are selected for calibration, and this restores $E_{\text{reco}}/p_{\text{track}} = 1$ when only the charged hadron's calorimeter energies are plotted (Eq. 4.10). (Discounting the satellite clusters and subsequent PF neutral hadron component makes the distributions lop-sided; the Gaussian fit has therefore been omitted.)



(a) Calibrating only the charged hadron clusters in testbeam (endcaps)

(b) Using the same calibration on collision data (endcaps)

▷ 5.5 Summary

The collision data collected in 2009 has proven useful for cross-checking the conclusions made in Chapters 3 and 4 of this thesis. Evidently, the fast simulation, from which the standard particle flow calibration is derived, has shortcomings that need to be addressed.

The particle flow calibration coefficients derived from testbeam work well in correcting the energy response despite lacking a correction for η variation. The shapes of the response seen in collisions are qualitatively compatible with the shapes exhibited by testbeam data.

Part III

Future technology in calorimetry & the ILC

Chapter 6

MVA techniques applied to hadron reconstruction

> They're good, but it's not like they're going to clean your flat or anything like that.
>
> *Freya Blekman*

6.1 Introduction

This chapter describes how neural networks can calibrate the CMS HCAL's response to hadrons. Using such a multivariate technique makes it relatively straightforward to evaluate the HCAL's performance to pions (resolution, response) if a longitudinal profile of the hadron shower is made available in each HCAL tower. Such an upgrade to the HCAL is expected in the coming years, and results presented in this chapter seek to anticipate what improvements may be expected, making extensive use of the fast simulation.

This chapter also introduces a concurrent, functional programming language called Erlang, in which the neural networks are implemented. The performance gains that result from using such a language are presented and discussed.

Questions

Three questions will be considered in this chapter:

(i) Can a multivariate technique offer benefits when reconstructing hadrons? What are the drawbacks?

(ii) Do such techniques benefit from extra spatial information?

(iii) Is a functional language useful? Can we exploit concurrency better?

Disclaimer

The work presented in this chapter is of an academic and hypothetical nature. The results presented forthwith should not be taken as a commentary on the CMS HCAL upgrade plans: the models of depth segmentation and 'toy models' for upgraded electronics are emphatically not serious efforts to model the upgrade, but merely provide a motivation for studying neural network techniques.

▷ 6.2 Motivation for using neural networks

6.2.1 Application to hadron reconstruction

Review of the current HCAL implementation

The HCAL's construction was described in Chapter 1, §1.3.5. There it was observed that while the HCAL is a sampling calorimeter, with up to 15 samples per tower in the barrel, all the scintillating tiles in a tower are optically joined and read out by one HPD.

An HCAL upgrade?

The LHC experiments anticipate a long shutdown $\mathcal{O}$(9 months) in several years' time, during which various on-detector electronics and components can be replaced and upgraded as necessary. The 'Phase 1' HCAL upgrade proposes to group the physical scintillating layers into several logical readout layers per tower, where at the time of writing 'several' equals four [91,92]. These four logical layers are readout with electronics improved relative to the current design, and will likely use silicon photomultipliers (SiPMs) as the detection method rather than the HPD.

The optimal correspondence of physical to logical layers is still under investigation by other parties, but they are likely to be more dense towards the front of the HCAL to increase the fraction of energy that is sampled electromagnetically. The current scheme is shown in Table 6.1. Note the synergy of this scheme with the forward HCAL calorimeter with its long and short fibres (§2.3.2, page 56).

TABLE 6.1 **Logical layer readout**
Correspondence of physical layers to logical layers in the Phase 1 HCAL upgrade — subject to revision.

Physical layer	Logical layer
0	0
1, 2	1
3, 4	2
≥ 5	3

Advantages of depth segmentation

It is clear from Chapters 2 to 4 that the HCAL offers poor resolution for low energy hadrons. The intrinsic noise of each channel's electronics degrades the energy measurement. Furthermore, the lack of depth information means it is impossible to directly

detect neutral hadrons cascading in the vicinity of charged hadrons. Depth information will *in principle* allow:

(i) Weighting of the energy measurements in depth to optimize the enegy resolution;

(ii) Improved spatial resolution of hadronic cascades. Particle flow can create PF clusters in depth, not just in $\eta\phi$;

(iii) Decorrelation of noise in neighbouring (in depth) channels. Assuming that the noise is not correlated between channels, this reduces the susceptibility of the energy measurement to noise fluctuations *if* the noise is small relative to the signal being measured;

(iv) Improved redundancy should one channel fail;

(v) Recognition of highly energetic showers leaking from the ECAL into the HCAL.

6.2.2 Neural networks

What is a multivariate technique?

Strictly speaking, 'multivariate analysis' simply implies that the solution to the problem considers more than one variable. More usefully, correlations between each input are fully exploited. The linear least-squares minimization technique for calibrating the hadron response barely qualifies for this definition because we merely have two classes of hadrons: those that interact in the ECAL and those that do not. Furthermore, only two pieces of information are considered—the ECAL and HCAL energies.

More conventionally, a multivariate analysis implies a *connectionist* approach is employed to *learn* about a given problem. A neural network is one example of connection-based machine learning.

Overview & history

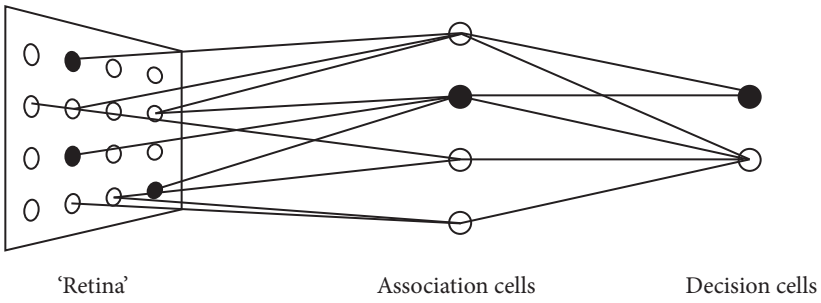
Artificial neural networks (see [93] for an introduction) were originally developed to model biological neural networks such as the human brain: the elementary network component is the neuron which fires or produces some output subject to inputs received from other neurons. An artificial model of such a structure also appears to learn, in the sense that its organization and connections modify in response to evidence accrued over time. This *connectionist* approach to modelling emergent and complex behavior is in contrast to the *symbolic* approach. Rather than solve a problem by means of a hierarchical application of patterns and rules, the connectionist approach learns and recognizes facts by the constructive interference of signals within a network. Knowledge in such a network is distributed, whereas the symbolic approach generally delegates the work to a single node. Needless to say both have their advantages and disadvantages and the symbolic approach has induced much research—some of it successful—into artificial intelligence.

Early neural models (1943–1956) [94–96] started by calculating basic logical functions (such as XOR), and could eventually learn and recognize basic patterns presented, culminating in the *perceptron* shown in Figure 6.1. This was followed by high hopes and

expectations of the power of such an approach, but the system was shown to be mathematically flawed [97] (and subsequently ignored) until work by Hopfield in 1982–1985 [98] combined with earlier work by Widrow and Hoff [99] brought rigour and formalism to the field. Second, the development of the back-propagation technique [100, 101] generalized Widrow and Hoff's work, and provided a rigorous and systematic method for training layered networks.

FIGURE 6.1

Rosenblatt's original perceptron model
This was the first concrete model which was (a) sufficiently specified for claims regarding its performance to be tested, (b) complex enough to show emergent behaviour, (c) sufficiently simple for analysis, and (d) produced results in agreement with reality. Redrawn from [93].



Associative networks

Multilayered networks are sometimes known as associative networks: a set of input neurons is connected to a set of outputs via another set of neurons with modifiable connections. The output is interpreted as the sum of signals received at the output from all the neurons in the network. Associativity is implied by the way that a problem, presented in terms of a pattern, shape, or situation is paired with another object—another pattern, shape, or situation, but which isn't necessarily of the same type. The network accomplishes this by building a multidimensional surface describing the non-linear mapping of inputs to outputs.

Advantages & disadvantages of these approaches

The advantage of such an associative approach is that it can handle large multivariate datasets without requiring an explicit form of a solution. No assumptions are made, and no preconceptions are implied. It provides a non-linear response surface without specification of derivatives, bounds or exceptions. In spite of this, these are emphatically *not* black boxes and reasonable performance will only be achieved subject to wise and careful deployment.

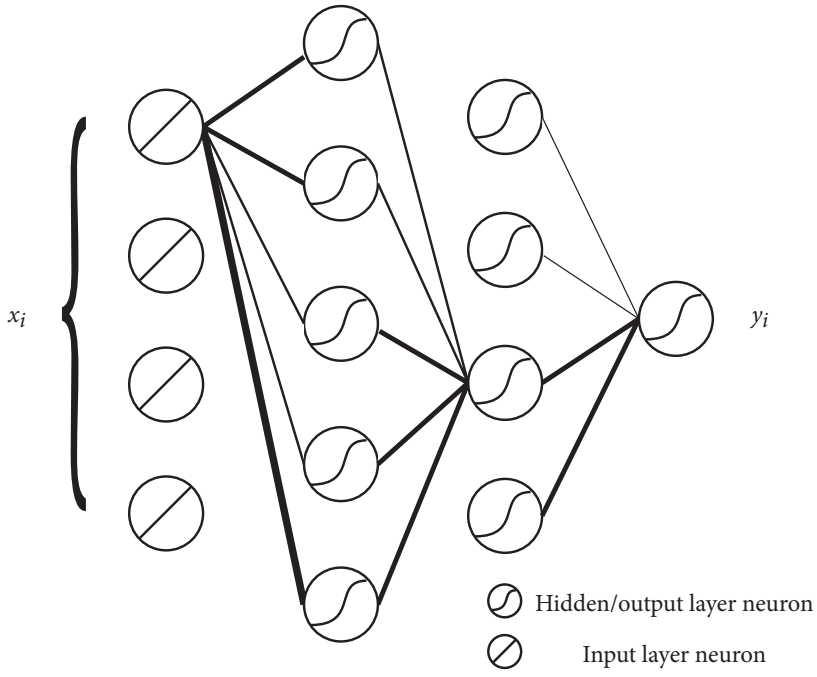
The flexibility offered by this associative approach is tempered by both the cost of implementation and optimization of the specific network under consideration, which may prove difficult to characterize. Finally, because the network's knowledge is distributed, with many neurons responding to any input, it is hard to (a) assign a physical *interpretation* to the neurons in the hidden layers, and (b) extract generic *rules* about the underlying system's behaviour.

6.2.3 The multilayer perceptron

Conventional neural networks are organized in layers and, in a *feed forward* network, direct connections between neurons are only permitted between neurons in adjacent layers. The network is specified by an input layer, with one neuron for each of the input variables, $x_1, \ldots, x_n$, followed by an arbitrary number of hidden layers, each with an arbitrary number of neurons, and a single output layer containing as many neurons as parameters required of the solution, $y_1, \ldots, y_m$. These outputs are often called the *neural network estimator(s)*. In classification problems the output denotes the likelihood that the test example is of class A or B, whereas in regression the output has a direct correspondence with the regression target variable. Most problems only require one output neuron. A typical scheme for a multilayer perceptron is shown in Figure 6.2.

FIGURE 6.2

The multilayer perceptron
Showing an input layer where each neuron emits an input, x_i, two hidden layers, and an output layer which emits the neural network estimator, y_i. The hidden and output layer neurons apply the non-linear activation function to the dot product of the weight and input vectors received (see text). Observe that connections in this scheme only exist between neurons in directly adjacent layers. Some weights have been drawn between neurons, with the line thickness indicative of the weight assigned between two neurons.



Each neuron specifies an *activation function*, f, which transforms the weighted sum of its inputs received from neurons in the layers before it, or merely the x_is for the input layer neurons, into one output. The network can accommodate a non-linear response by means of a non-linear activation function[1] and the hyperbolic tangent function is typically used. If the neuron labelled by j has Q inputs, q_i, each weighted by W_{ij}, the neuron's output p_j is,

$$p_j = f\left(\sum_{i=1,Q} W_{ij} q_i\right), \tag{6.1}$$

[1] In principle, the mapping is non-linear if at least one neuron has a non-linear activation function, but this is not usually sufficient to solve the problem well!

where in this document,

$$f(x) = \tanh(x). \tag{6.2}$$

Training the network with back-propagation

The weights, W, between neurons are determined by training the network. The most common and simple method is the *back-propagation* algorithm, a full description of which may be found in [93]. The central idea is to operate the network backwards, propagating a correction to the weights from the output neuron to the input neurons. This is a supervised learning method implying the error on each output neuron can be determined for every training example available: let the total error, $E^k(W)$, on the outputs when the k-th training example is presented be,

$$E^k(W) = (y^k - t^k)^2 = \sum_{i=1,m} (y_i^k - t_i^k)^2 \tag{6.3}$$

for y_i being the ith neuron's output and t_i its desired output. So the total error for the entire training set is,

$$E(W) = \sum_k E^k(W). \tag{6.4}$$

The weights that will minimize this quantity can be found by the *gradient descent* method, provided that the neuron response can be differentiated with respect to the input weights. If at some example, k, the weights are W^k, then at the next step the weights are adjusted by stepping in the direction $-\nabla_W \cdot E^k(W)$, for that is the direction where E decreases most:

$$W^{k+1} = W^k - \rho \nabla_W \cdot E^k(W), \tag{6.5}$$

where ρ is prescribed as the *learning rate*. The back-propagation method derives a rule from this result for changing the weights between two neurons i and j. After substantial manipulation, it re-expresses Eq. 6.5 as,

$$W_{ij}^{k+1} = W_{ij}^k - \rho(k) d_i O_j, \tag{6.6}$$

where O_j is the output of neuron j. For the output neurons the steps, d_i, are,

$$d_i = 2(y_i - t_i) f'(I_i) \tag{6.7}$$

while for the hidden neurons,

$$d_i = \sum_h W_{hi} f'(I_i), \tag{6.8}$$

where h labels all the neurons connected directly ahead of the ith neuron. I_i is the argument of the function in Equation 6.1. The back-propagation rule requires that f is differentiable; for our networks $f'(x) = 1 - \tanh^2(x)$. The networks can either be trained in *batch learning*, whereby all the training examples are presented and the total error computed before the networks' weights are adjusted; alternatively, the errors and weights may be adjusted after each example presented in *online learning*. This requires a well randomized training set, otherwise the network will quickly forget previous

examples. The ρ parameter can be increased to increase training speed and prevent the algorithm getting stuck in a local minimum of the error function, but doing so may induce oscillations and subsequent failure to converge.

Problems with back-propagation

The back-propagation algorithm does not offer a formal proof of convergence, and may get 'stuck' in local minima or 'lost' on plateaus of the error function. One popular alternative is the Broyden-Fletcher-Goldfarb-Shanno algorithm (BFGS) [102–105] which uses the second derivatives of the error function to optimize the amount by which the weights are changed at each iteration. For networks with $\mathcal{O}$(tens, not hundreds) of neurons, the BFGS method generally converges faster to a solution.

6.2.4 Generalized Regressive Neural Networks

The *Generalized Regressive Neural Network* (GRNN) belongs to a general class of networks designed for regression rather than classification outcomes. The network architecture is very different to the MLP, though it preserves an input and output layer for the same function. There are usually two intermediate layers, the first of which, the *pattern layer*, contains a neuron for every training example available. The second intermediate *summation layer* has two neurons. When a test example is presented, the output of each neuron in the pattern layer is the sum of all known training inputs, each transformed by a *kernel function* representing how similar the test example's inputs are to that neuron's training example. The first neuron of the summation layer sums the outputs of the pattern layer while the second neuron merely exists for normalization so the output is the ratio of the two neurons in the summation layer.

The kernel function controls the specificity of the network, and is usually governed by a smoothing parameter, γ. If this parameter is small in the context of the problem under study then a spiky response surface results and the network will likely fail to provide a reasonable solution. If the γ factor is too large then the surface will be too smooth and the network will offer poor resolution for a given test example. The kernel function is usually taken as a Gaussian, so the γ factor is the width, σ, of the Gaussian. In conventional implementations the γ factor is the same for all input variables but more sophisticated implementations may tune this.

While GRNNs are easy to implement they tend to be slow and despite the simplicity of governing the kernel function, the best network performance is unlikely to be attained without tuning the γ factor, perhaps for each input variable available. Intermediate neurons (*i.e.* training examples) should also be pruned out where possible if they offer no new information. Finally, the GRNN is widely acknowledged to be poor at extrapolating beyond the limits of the known input parameter space.

▷ 6.3 Implementation & methods

6.3.1 Use of a functional language

Functional languages, sometimes called 4^{th} generation languages, differ from imperative and procedural languages such as C, C++ and Java in important respects:

- ▷ Functional programs solve problems by the evaluation of functions which have a much closer relationship to mathematical functions than their procedural counterparts. Once defined, the value of a variable cannot be changed, so statements such as `x = x + 1` are not legal.[2] This property of functional programs is called *single assignment*. By contrast, imperative programs usually solve problems through iterative changes of state.

- ▷ The result of a function is only determined by its arguments and its definition. Calling the same function with the same arguments repeatedly is guaranteed to give the same result. This property is called *referential transparency*.

- ▷ The combination of these two characteristics implies that functional programs are free of *side effects*: no state of the program is altered by function evaluation.

- ▷ In some languages, expressions are not evaluated if the value of the expression has no effect on the value of the encompassing expression. This quality is called *lazy evaluation*.

In practice, functional languages often encourage the user to solve problems recursively. Secondly, most functional languages provide mechanisms which permit side effects, such as input/output, in a controlled manner.

6.3.2 A brief introduction to Erlang

Erlang [106, 107] is one example of a functional language. It was developed at Ericsson in 1986 and now exists as open source software. Concurrency is at the heart of its design: in the absence of side effects in such a language, Erlang *processes* communicate asynchronously by passing messages. There is no need for memory or variables to be shared, and no synchronized code, locks or mutexes. Processes are spawned with much smaller overheads than the operating system level threads common to procedural languages. The Erlang virtual machine schedules the processes across the processors and cores locally available. Virtual machines which may be geographically separated can be linked, so the entire computation becomes distributed.

Erlang was originally developed for telephony applications and, aside from concurrency, Erlang programs are made extremely robust by enabling processes to monitor and trap abnormal exits of fellow processes. The runtime also supports hot-swapping of code, so faulty code can be replaced with patched code without needing to restart the global system.

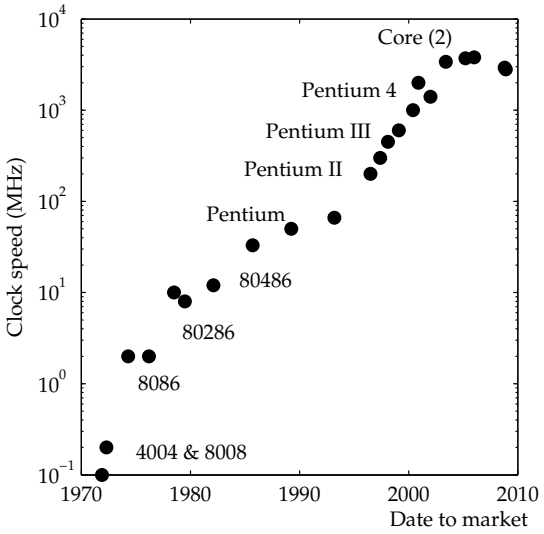
Message passing concurrency

Until recently, much of the world has implemented concurrency through the sharing of state. Writing correct, fault-tolerant multi-threaded programs is notoriously hard. With multiple threads sharing memory and modifying it, the scope for data corruption and error increases. In a message passing language, *there is no shared or mutable state*. To first order, because variables are immutable, no locks are necessary and the problem at

[2] Strictly speaking, this is an expression that will evaluate to `false`. In most imperative languages, or indeed, popular languages, the `=` operator is equivalent to assignment. So the statement `x = x + 1` says '`x` receives the value `x + 1`'. In a functional language with single assignment, this expression implies a test of equality, which must fail and evaluate to `false`. In Erlang, this is a *pattern match* which fails.

FIGURE 6.3

Intel processor clock speeds
Data from [108, 109]. Processor speeds have not increased from their peak in 2008: processors entering the market today are likely to have 2 or more cores on each chip. Technical reasons for the plateau in clock speeds are that (a) memory speeds are not increasing, (b) power consumption increases, and (c) difficulties are encountered when designing a processor—itself a distributed entity—that can interpret a stream of instructions at such high speeds.



large becomes easier to parallelize. With some thought, many problems which have been solved sequentially can be rapidly parallelized through judicious use of independent, concurrent processes such as those available in Erlang.

The role of multicore CPUs

Clock speeds of conventional desktop processors have reached a plateau in recent years, as evidenced in Figure 6.3. The trend followed by major manufacturers has been to increase the number of cores on each chip. Programs will not run faster unless they take advantage of this technology. Interest in Erlang has been renewed of late because the performance of programs written in Erlang can scale with the number of cores available to the Erlang virtual machine. These claims will be verified forthwith.

These observations are not of purely academic interest: Google and Yahoo! have enjoyed great technological success by building applications based on these principles [110–112].

6.3.3 Modifying the HCAL fast simulation

For this academic study, the fast simulation (see Chapter 3) provides a convenient tool for simulating potential HCAL upgrades despite its shortcomings in modelling the calorimeters' response to real data. In the standard software release, the hadronic cascades are simulated for each physical layer in the HCAL. The energies deposited in each layer are then summed before digitization is applied. It was a straightforward matter to sum the energies according to the logical layer scheme described above, but implementing a full digitization simulation is beyond the scope of this study. Rather, we will consider a set of toy models of digitization.

6.3.4 The toy models

The simulated energy in each physical layer is converted to a reconstructed energy by a toy model of noise and smearing. Consider the following definitions:

▷ E_l: The reconstructed energy in the l-th logical layer.

▷ T_p: The simulated energy in the p-th physical layer.

In general,

$$E_l = Q\left(\sum_p^{p\in l} R(T_p)\right) \tag{6.9}$$

for Q, R unknown functions of physical and logical layer energies respectively, and the notation $p \in l$ denoting all those physical layers p contributing to the logical layer l. Here we describe a toy model to provide some reasonable conversion of the T_ps to the E_ls. In the following, a random sample from a Gaussian distribution with mean, μ, and width, σ, is represented by $G(\mu, \sigma)$.

The current HPD-based implementation is modelled with:

$$E_l = \left[\sum_p^{p\subset l} T_p + G(0, 200\,\text{MeV})\right] \times G(1, X\%), \tag{6.10}$$

and l is single-valued. The 200 MeV noise term is based on the HCAL barrel noise obtained in Chapter 3. The last multiplicative smearing term relates to the accuracy and precision offered by the digitization. In this document, we define it as the *quantization error* (QE). Possible values for X are detailed forthwith. Eq. 6.10 represents a very simple emulation of the real digitization that takes place. For example, the effects of saturation, non-linearity in the conversion of collected charge/photons to ADC counts, digitization and rounding errors, temperature fluctuations, and subsequent gain dependence can all be expected to affect real-world systems. This toy model brazenly wraps all these subtle effects into the smearing term.

For a multi-layer HCAL which uses the same electronics, l ranges over each logical layer so the total reconstructed energy is just,

$$E_{\text{reco}} = \sum_l E_l. \tag{6.11}$$

In this thesis we will only consider the 4 logical layer structure given by Table 6.1.

6.3.5 Network implementation

A multilayer perceptron was built in Erlang according to the general principles of §6.2.3. A back-propagation trainer was also written to present training examples to the network, compute the output error function, and adjust the network weights. The implementation consists of four basic Erlang elements:

`neuron`: Models the activation function, Eq. 6.2. Maintains a list of weights relevant to each neuron connected behind it. When its input changes it notifies neurons ahead of it of its new output. Broadcasts errors to neurons behind it when its weights change, according to Eq. 6.8. A neuron recomputes its weights according

to the back-propagation algorithm when a neuron ahead of it broadcasts an error backwards.

`layer`: Holds neurons. When the network is initialized the layer initializes its neurons and connects them to each and every one of the neurons in the layer ahead.

`mlp`: Contains the layers. Computes the output vector $y_1, \ldots, y_m$ for a given vector of inputs $x_1, \ldots, x_n$.

`trainer`: Given a set of training examples, this presents each one to the `mlp` in turn, computes the quadratic error for the kth example according to Eq. 6.3, and tells the output neurons how much to adjust their weights. The output neurons broadcast the error backwards.

Each one of these components exists as a distinct Erlang process. For a network with 9 inputs, two hidden layers of 10 and 9 neurons respectively and one output, there are 29 `neuron` processes, 4 `layer` processes, one `mlp` process and one `trainer`, totalling 35 processes. A neuron initializes each of its weights to a random number sampled from the interval $\left[-\frac{1}{n}, \frac{1}{n}\right]$ for n being the number of neurons connected behind the neuron.

6.3.6 Acquiring data

π^+ PYTHIA [65] particle gun events were processed by the fast simulation, in a manner similar to that described in earlier chapters. The beam momentum was sampled from a uniform distribution ranging from 2 to 100 GeV/c. The simulation of nuclear interactions and material effects due to the presence of the tracker were disabled in this study. As in previous chapters, the pions originated from the nominal interaction point of CMS and ranged over all possible values of ϕ. The pseudorapidity was restricted to $0.2 < \eta < 0.25$ (the central barrel). Networks were trained to predict the true energy of the pion given calorimeter data as input.

Data for training and testing were read from a text file using functions provided in the standard Erlang libraries; the file contains events simulated in CMSSW. The data used to evaluate the network's performance were kept distinct from the training set.

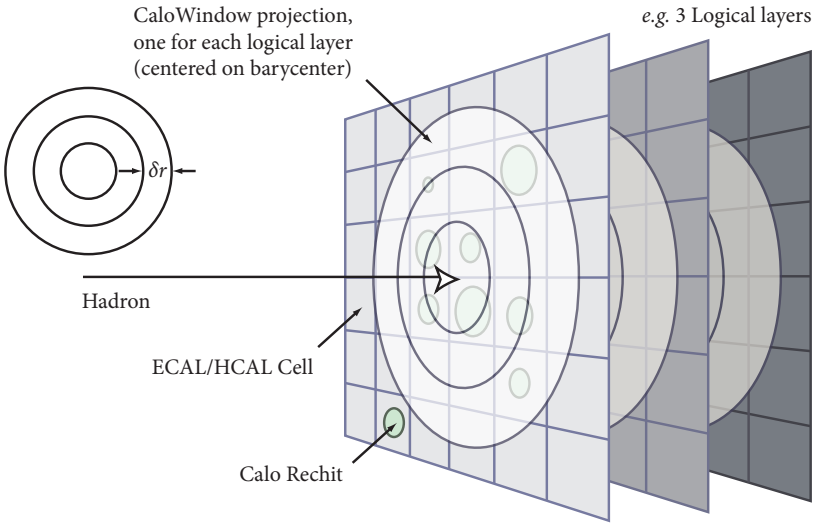
6.3.7 CaloWindows

The neural network accepts transverse shower information as input according to the following prescription: a geometric transformation was employed to convert HCAL calo rechits into a mapping which is independent of rechit η, ϕ. A `CaloWindow` was created, which looks like a bullseye target. It is centred at the geometric centre of the rechits collected from a cone around the projection of the charged hadron's track on the HCAL, in a similar manner to previous work detailed in earlier chapters.

The `CaloWindow` contains a set of concentric rings, called `CaloRings`, (Figure 6.4) each separated by a distance $\delta\mathcal{R} = 0.04$. Four such rings therefore cover the region up to $\Delta\mathcal{R} = 0.16$. The energies of the rechits are then projected onto the `CaloWindow`. The neural network receives the total energy of each `CaloRing`. In the case of an HCAL with layers, a `CaloWindow` is created for each layer.

`CaloWindows` are also employed in the ECAL: in view of the ECAL's much finer granularity, the `CaloRings` are separated by $\delta\mathcal{R} = 0.01$, and there are five per `CaloWindow`.

FIGURE 6.4 **CaloWindows**
The `CaloWindows` are projected onto the pion shower. A `CaloWindow` is created for each logical readout layer.



▷ 6.4 Results

6.4.1 Network response with the standard implementation

Figure 6.5 shows the performance of the neural network when standard rechits from the fast simulation are used — *i.e.* not a toy model, and with just 1 layer. The resolution achieved by the 'banana corrections' of [36] is shown by the grey line; this was previously described in Eq. 1.3. Using the neural network improves the resolution to,

$$\frac{\sigma}{E} = \frac{90.3\%}{\sqrt{E}} \oplus 5.6\%, \tag{6.12}$$

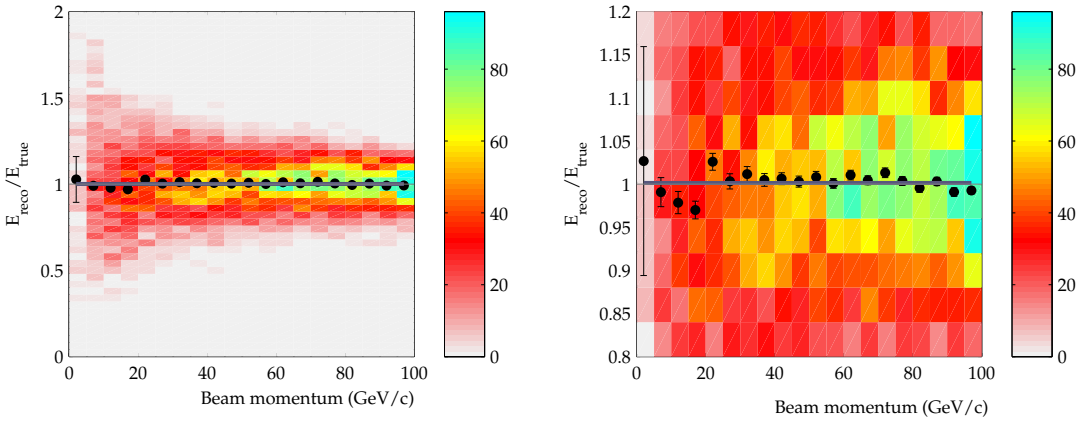
representing a small improvement in both terms, *but* this study is based on the fast simulation, whereas the banana corrections are derived from testbeam data. By visual inspection of Figure 6.5 it appears that the results do not substantially improve on the benchmark banana correction, despite the fit values, though there is a systematic improvement at higher energies. At 100 GeV the banana correction achieves $\sigma/E = 12\%$ (data) and the network produces 10% (simulation).

The linearity of the network is excellent, with none of the points of fitted $E_{\text{reco}}/E_{\text{true}}$ deviating more than 3% from unity, and to within 1% above 25 GeV. The banana correction achieves mean responses in the range 96%–107%.

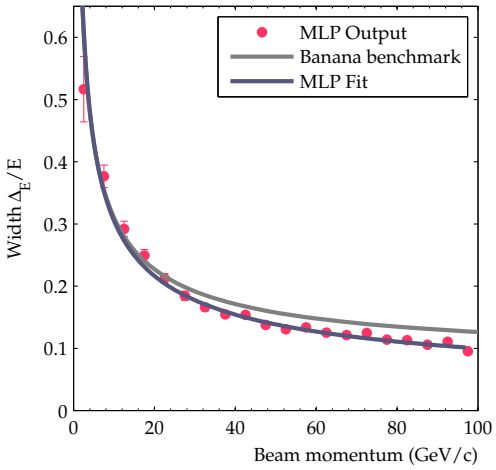
The benchmark toy model

Regarding an upgrade, we should compare like with like. A toy model (Eq. 6.10) with 1 layer, 200 MeV noise and $X = 20\%$ QE yields a network whose response is $\sigma/E = 73.3\%/\sqrt{E} \oplus 6.7\%$. This value of X was intended to yield a network performance comparable to that obtained using standard rechits, but in hindsight this toy

FIGURE 6.5 **Neural network performance with standard rechits**
Reconstructed energy ratios and resolutions using the neural network technique. The network was trained with 2,000 events and tested with 10,000 events. The Gaussian fitted mean is plotted as a function of beam momentum. The coloured z-scale represents the number density of events in each bin.



(a) Response *(b) Response — zoomed*



(c) Resolution $\sigma/E = b/\sqrt{E} \oplus c$

model is too good, and a larger value of X may be more appropriate.[3]

Figure 6.6 demonstrates the performance of the networks using this toy model of the current electronics for both the current case of no depth segmentation and where 4 logical layers are provided. Again, the banana correction's resolution is shown by the grey dashed line.

Fitting the resolution σ_E/E using the usual parametrization is problematic because there is a substantial interaction between the a and b terms. We will therefore generally consider a fit of the form $\sigma_E/E = b/\sqrt{E} \oplus c$, to avoid a compensating for b and vice versa. For Figure 6.6, the network produces $b \sim 80\%$, and this term does *not* improve when 4 logical layers are implemented. The c term improves slightly.

6.4.2 Varying the toy model parameters

We will consider the following variations of the toy model:

- ▷ Noise of 200 MeV, 20% QE;
- ▷ Noise of 100 MeV, 20% QE;
- ▷ Noise of 100 MeV, 5% QE;
- ▷ Noise of 25 MeV, 5% QE.

Each model was evaluated for the case of no depth segmentation and for the case of 4 logical layers. These toy model parameters are purely hypothetical, chosen to span the spectrum of possibilities between the status quo and some near-perfect system. Let us consider the network performance for the case of 25 MeV noise and 5% quantization error, as shown in Figure 6.7. Excellent linearity is observed, and the resolution b term is much reduced to 50.5%, while the c term is zero (within error).

Linearity

Consider how the fitted Gaussian mean of $E_{\text{reco}}/E_{\text{true}}$ in each bin of E_{true} varies in Figure 6.8, shown for the two most extreme toy models, each with 1 or 4 layers. Good linearity is observed with the best performance, unsurprisingly, coming from the model with lowest noise and lowest QE factor.

To assess the linearity systematically consider the sum of the residuals squared. This is defined as,

$$\sum_{i}^{i \in \text{bins}} (\mu_i - \bar{\mu})^2. \tag{6.13}$$

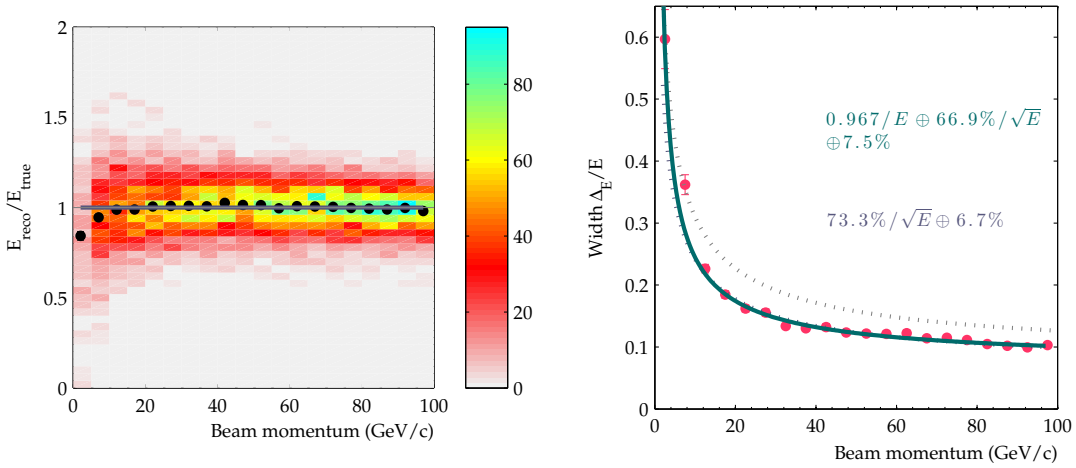
Networks with better linearity will have lower values of this quantity. The values for each of the toy model and layer combinations are shown in Table 6.2. There is a general trend that the networks with lower noise and less QE perform better. There are also indications that a 4 layer HCAL produces better results than a 1 layer HCAL.

Resolution

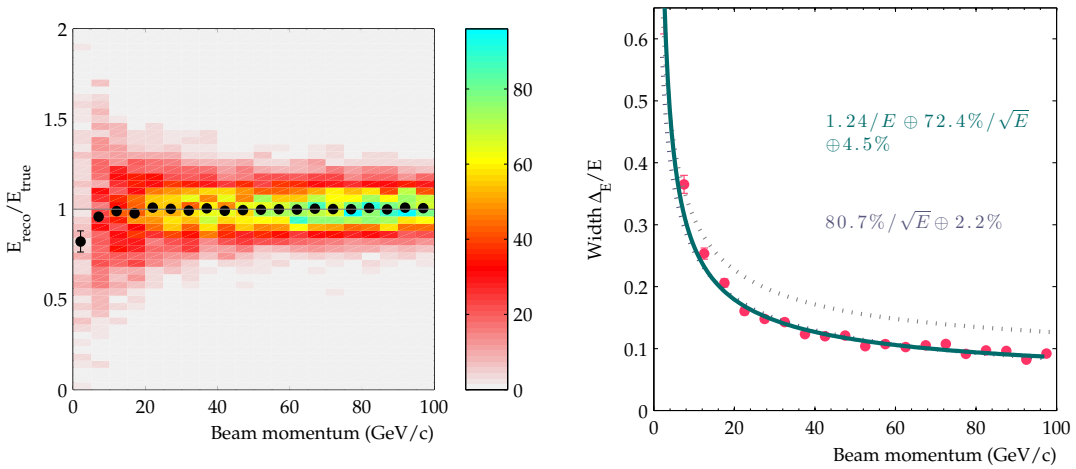
A summary of fitted resolution functions is exhibited in Table 6.3. Comparisons between contributions to the resolution from electronics and layering are shown in Figure 6.9.

[3] The X parameter was estimated using an earlier version of the simulation, which in turn overestimated the HCAL's performance.

FIGURE 6.6 **Neural network performance with the toy models**
Reconstructed energy ratios (left) and resolutions (right) using the neural network technique. All networks were trained with 2,000 events and tested with 10,000 events. The standard parametrization of the banana correction's resolution [36] is shown by the grey dashed line.

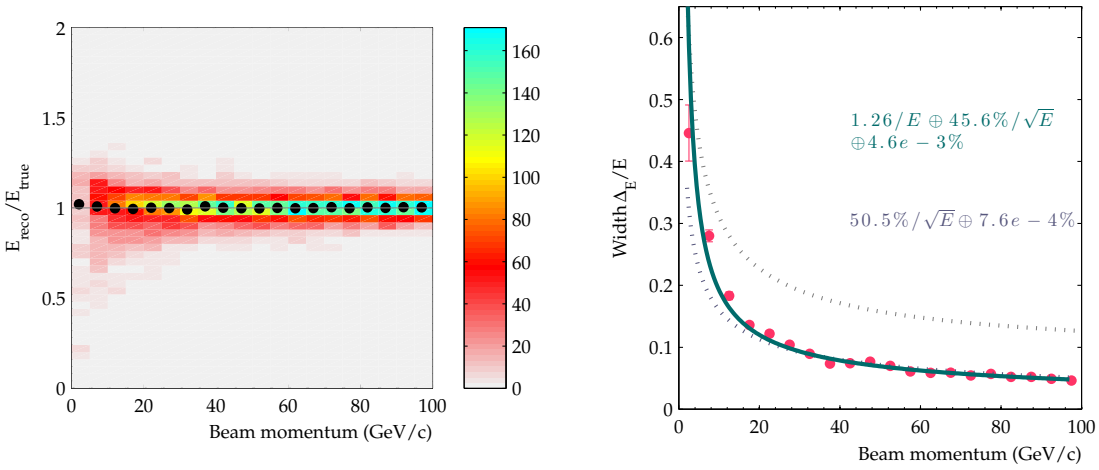


(a) HPD-emulation, 1 logical layer — 200 MeV noise per layer and 20% quantization error



(b) HPD-emulation, 4 logical layers — same parameters as above

FIGURE 6.7 **Network performance for toy models with low noise and low QE**
25 MeV noise and 5% quantization error were used in the toy model. See text for details.



(a) Response *(b) Resolution*

TABLE 6.2 **Summary of linearity residuals**
Sum of residuals squared (Eq. 6.13, multiplied by 100) for each toy model and layer combination. Lower values indicate better linearity. The results for 100 MeV noise and 20% QE appear anomalous.

Model [Noise (MeV)/QE (%)]	**1 layer** [×100]	**4 layers** [×100]
200/20	2.96	3.16
100/20	3.46	0.32
100/5	2.24	0.97
25/5	1.47	0.07

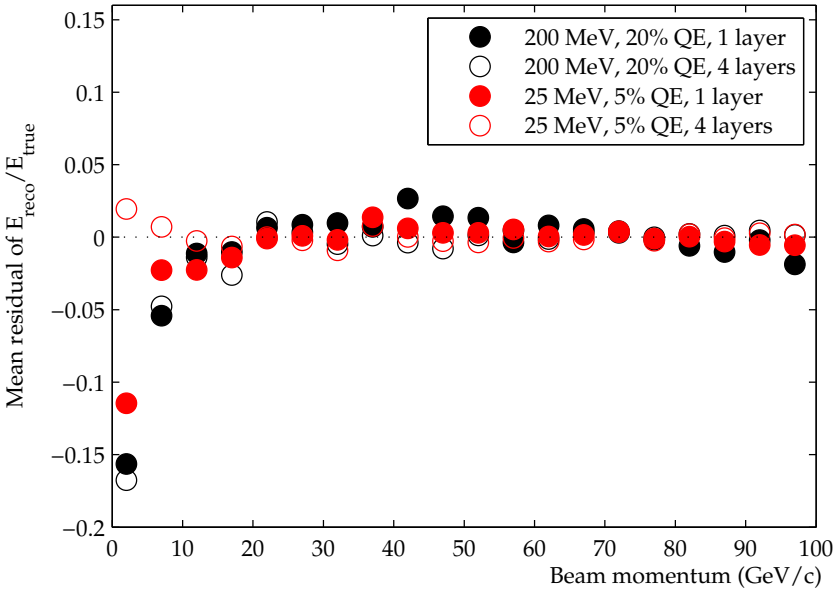
It is interesting to note that increasing the depth segmentation does not *a priori* improve the resolution b term. Specifically, the performance at low energies $\mathcal{O}(< 50$ GeV) is not improved, but the curves are qualitatively compatible. The absence of a dramatic improvement in resolution with depth segmentation may be due to:

- ▷ A flaw in the network implementation: unlikely, because the networks behave sensibly, and similar results have been found with other network implementations;
- ▷ Domination of signal by noise in each logical layer. Less energy is collected by each channel and, while the energy collected is given by Eq. 6.11, if the spread attributable to each energy measurement is σ_E (assuming they are all equal), then the total error is $\sqrt{4\sigma_E^2} = 2\sigma_E$, implying the result is worse in the noise-dominated

TABLE 6.3 **Summary of resolutions**
Excluding the a term, $\sigma_E/E = b/\sqrt{E} \oplus c$. The errors on each term are small $\mathcal{O}(1\%)$.

Model [Noise (MeV)/QE (%)]	**1 layer**	**4 layers**
200/20	$73.3\%/\sqrt{E} \oplus 6.7\%$	$80.7\%/\sqrt{E} \oplus 2.2\%$
100/20	$73.9\%/\sqrt{E} \oplus 6.5\%$	$69.2\%/\sqrt{E} \oplus 5.4\%$
100/5	$50.6\%/\sqrt{E} \oplus 0\%$	$52.0\%/\sqrt{E} \oplus 0\%$
25/5	$50.9\%/\sqrt{E} \oplus 0\%$	$50.5\%/\sqrt{E} \oplus 0\%$

FIGURE 6.8 **Linearity residuals**
Residuals between a line fitted to the Gaussian mean of bins of $E_{\text{reco}}/E_{\text{true}}$, and the points themselves, for four toy models.



regime. At high energy, the signal dominates the noise and an improvement in resolution is observed;

- ▷ The smal longitudinal depth penetration of low energy $\mathcal{O}(< 20\,\text{GeV}/c)$ hadrons, which limits the bulk of the energy deposited to just one or two logical layers. The networks improve the c term of the resolution at high energies;
- ▷ The fast simulation's model of hadronic energy deposition in each tower being physically inaccurate.

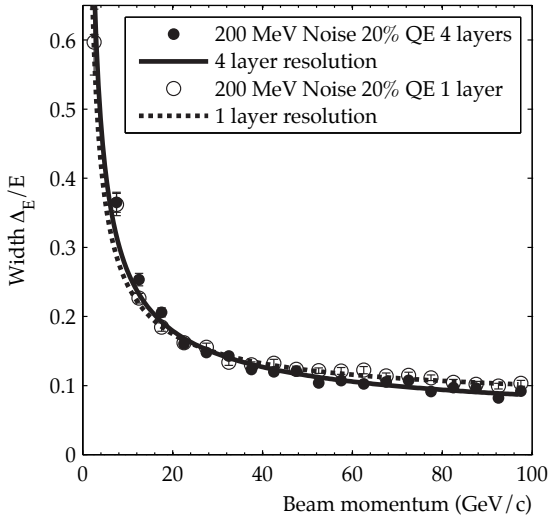
Studies by other members of the CMS collaboration have found a small improvement $\mathcal{O}$(several percent) in the resolution that can be attained with a multilayer HCAL when the full simulation is used [113], but these studies produce different results depending on the model of hadronic cascades used.

Correlation between inputs, truth and neural network output

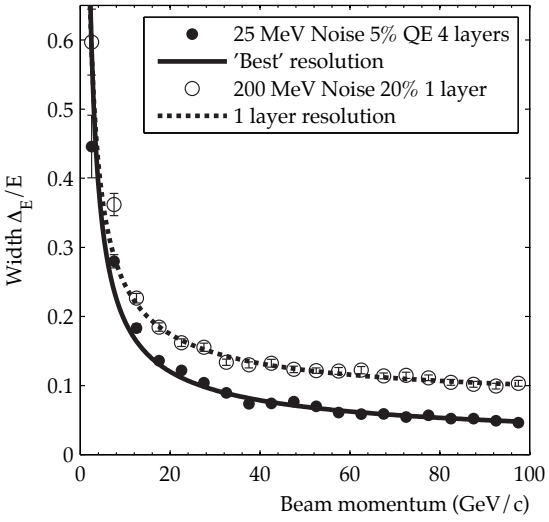
The correlation matrix R with elements $R(i, j)$ defines the correlation coefficient between any two of the input and/or output variables. In Figure 6.10, the correlation matrices for two toy models are presented. Evidently the matrices are symmetric. Consider the 1 layer toy model example. The i, j indices for this network are:

- ▷ 1–5: ECAL `CaloRing` energies, with the most central first;
- ▷ 6–9: HCAL `CaloRing` energies, with the most central first;
- ▷ 10: the network's output;
- ▷ 11: the truth/target value.

FIGURE 6.9 **Comparison of resolutions**
A 4 layer HCAL versus a 1 layer HCAL, where the same toy model of digitization is used on each channel, and where the 'ideal' model of digitization is used. 4 layers does not improve the stochastic b term of the resolution, but there is a slight improvement in the c term, visible at high energy. The same is true when an improved model of HCAL digitization is used. Improving the digitization is the way to improve the resolution at all energies.



(a) Preserving the current digitization while increasing the depth segmentation does not improve the resolution substantially



(b) The digitization dominates the resolution

Observations to be made here are:

(i) *The top left corner of 5 × 5 boxes:* the ECAL variables are closely correlated with each other. Hadronic cascades in the ECAL are large compared to the scale of ECAL crystals.

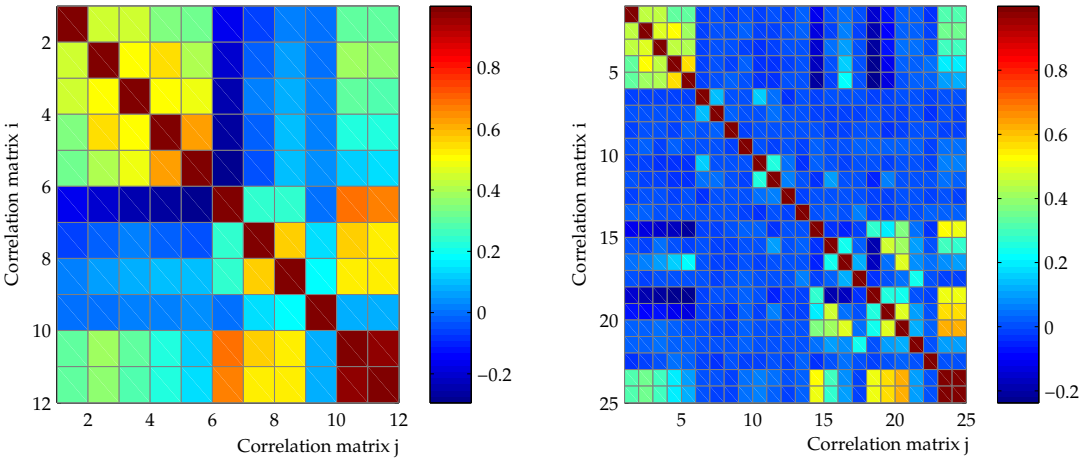
(ii) *Consider how column 6 matches rows 1–5:* the first HCAL variable ($i = 6$) corresponds to the most central `CaloRing`. It is anticorrelated with the ECAL inputs. This is indicative of the fact that pions that shower in the ECAL will leave less energy in the HCAL.

(iii) *Consider how columns 6–9 correlate with rows 6–9:* the HCAL variables are barely correlated with each other. With the exception of a modest correlation between two `CaloRings` (read column $j = 7$ with row $i = 8$) indicating that there is occasionally more than 1 rechit, the colours vary between pea green and blue indicating correlation coefficients no larger than 0.3. This explains why the neural network does not perform better than the standard banana correction method. This matrix indicates that there is little transverse spatial information for the neural networks to exploit.

(iv) The most central HCAL `CaloRing` ($i/j = 6$) is the most correlated with the Monte Carlo truth ($i/j = 11$).

(v) The neural network output ($i/j = 10$) is closely correlated with the target variable ($i/j = 11$) as both hoped for and expected.

FIGURE 6.10 **Correlations between inputs and outputs**
See text for details.



(a) Toy model with 1 layer, 200 MeV Noise, 20% QE

(b) Toy model with 4 layers, 100 MeV Noise, 5% QE

A similar analysis can be made of the 4 layer case shown in the same Figure: $i, j = 1-5$ denotes the ECAL, $i, j = 6-21$ are 4 `CaloWindows` each with 4 `CaloRings`. The η coordinate was also included ($i, j = 22$) to see if that could improve the performance; evidently it has no correlation with any other variable. The network output and truth are $i, j = 23$ and 24 respectively. The last logical layer ($i, j = 18-21$), which sums most of the physical layers, exhibits modest correlations between each of its rings similar to the 1 layer HCAL case, and with the innermost ring of the previous logical layer. The first 3 layers provide little extra information to the neural networks which, in addition to earlier arguments, explains why the neural networks do not benefit from the extra longitudinal segmentation.

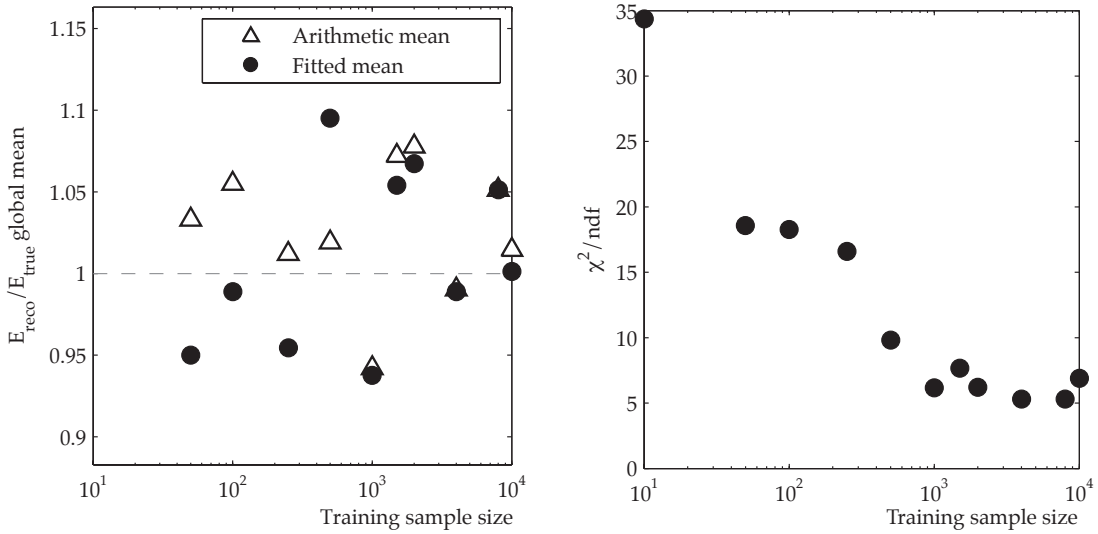
6.4.3 Network performance

Training sample size

There is no single test of a network's convergence to a solution (see remarks above concerning back propagation). The performance of a network can be judged in terms of its linearity and resolution, but we should consider whether the output is Gaussian too.

The 'at-a-glance' acid test for convergence used by the Author is given by the quantity $E_{\text{reco}}/E_{\text{true}}$. When a Gaussian is fitted to this distribution, a well-trained network presents a Gaussian with near unity mean. Furthermore, differences between (a) the arithmetic mean and standard deviation, and (b) the fitted mean and width indicate a non-Gaussian distribution—*i.e.* tails/asymmetry. A good fit will also have a low χ^2, a quantity determined by ROOT considering the residuals between the fit, the data points, and the number of degrees of freedom available to the fit. An example of this treatment is presented in Figure 6.11. Similar behaviour has been found with the other networks. On the basis of this information the networks were all trained with 2,000 events.

FIGURE 6.11 **Training the networks**
Output quality, defined by matching the arithmetic and Gaussian means, and a χ^2 measure, as a function of the number of training examples presented. The example is for the toy model with parameters 200 MeV noise, 20% QE, and 4 layers.



(a) Matching of arithmetic and Gaussian means indicates a good fit

(b) The χ^2 of the fit converges too

Dependence on network architecture

The network architecture did not affect the network performance substantially provided the number of hidden neurons was at least half as many as the number of input neurons. There are no theoretical results or general rules to suggest the optimal network architecture for a given problem: in some applications the best architecture (for some definition of 'best') can be found by a *genetic search*. Such a search was implemented by the Author but no obvious behaviours or advantages to using any one particular architecture were determined.

The results presented in this thesis concern networks with two hidden layers. For N inputs (N varies according to the `CaloWindow` and number of logical layers specified), there were $N + 1$ neurons in the first hidden layer and N neurons in the second. This follows the example set by the ROOT-based TMVA packages [114].

Validation of network implementation

Neural networks implemented in the TMVA package and MATLAB [115] application were tested using the same data. Results were in broad agreement with the Author's implementation. Substantial speed gains in training were achieved by use of the optimized learning methods made available in these implementations.

▷ 6.5 Computing performance

> In theory, there is no difference between theory and practice. But, in practice, there is.
>
> *Jan L. A. van de Snepscheut*

6.5.1 Theory

Most development was carried out on the Author's laptop, a 32-bit 1.8 GHz Intel Pentium Celeron, with 1 GB of memory running Ubuntu Linux 9.04 [116]. This is a single-core machine so there is little advantage to using Erlang when looking for speed gains in parallelization. Training and testing a typical network to produce the results shown in this thesis took $\mathcal{O}$(1 minute). The complex banana correction described in [36] and Appendix A.1 involves a number of steps that require careful optimization by hand and, subsequently, substantially more time to reproduce.

While the neural network is written using Erlang processes, the training (specifically, the back propagation of errors) and evaluation of a given test example are not optimized for parallel processing: the network can only evaluate one example at a time and the intermediate layers of hidden neurons are blocked until the output neuron(s) have performed the final transformation of their inputs. In a more performant implementation the evaluation of a set of examples would be pipelined, with each example being evaluated by the network on a layer-by-layer basis. At any one time as many examples propagate through the network as there are layers and, with each layer running as an Erlang process on a dedicated process core, the performance can scale appropriately.

6.5.2 Practice

Training the MLP; inter-process chatter

In principle, the Author's implementation should scale well on a multicore machine, but a flaw in the design has revealed a serious bottleneck: the atomic process unit of the implementation is the elementary neuron. The Author had envisaged that training a network (or a number of networks) simultaneously on a multicore machine would allow many neurons, layers and networks to use all available CPU resources. Indeed, this is the case. But in both training and testing the network, the number of messages exchanged between neurons is so large relative to the actual computation being performed by each neuron, that the performance is swamped by *inter-process chatter* (IPC). Measurements with the `cprof` tool [117] showed that the `neuron` module used 16,903,301 calls to solve a toy model problem with 4 layers and 1000 training and test values—see Table 6.4. Overall, only a modest speed improvement $\mathcal{O}$(1–5%) has been realized.

To avoid IPC, a redesign would have the network as the elementary Erlang process in the computation. While a network wouldn't be *trained* any faster, since training is a sequential process, a trained network could be cloned across each core available, so the dataset used for evaluation could be split up and evaluated in parallel in much reduced time.

TABLE 6.4

Inter-process chatter
Measurement of communications between Erlang processes by the `cprof` tool reveals that the Author's implementation of a multi-layer perceptron implies a vast number of calls to the `neuron` module.

	`trainer`	`mlp`	`layer`	`neuron`
Calls	7,020	89,068	359,0578	16,903,301

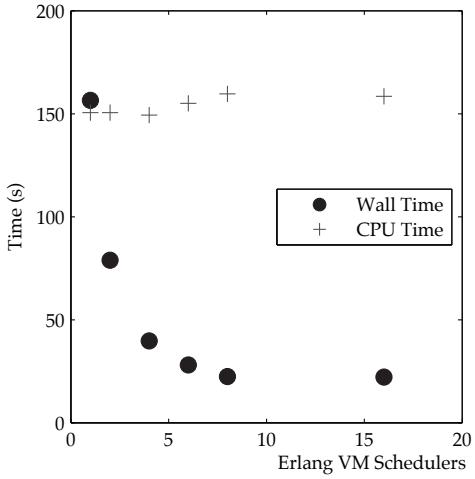
Performance with GRNNs

In any case, let us consider the computing performance of the GRNN described in §6.2.4. This is relatively easy to implement with $\mathcal{O}$(100 lines) of Erlang in the Author's prototype implementation. For tests, the Author used one of the batch processing (`lxbatchXX`) machines at Imperial College. Each of these 64-bit machines has 2 quad-core, 2.5 GHz Intel Xeon processors (8 cores), and 16 GB of memory and runs Linux CentOS v5 [118].

The Erlang virtual machine accepts a switch to govern its use of multicore machines, specifically the number of *schedulers*, S, which can run Erlang processes. In principle (and by default) S is as large as the number of CPU cores, N, available: increasing S beyond N has no effect to first order; lower S values restrict processing to S cores. In this test, a scan of 14 γ values (see §6.2.4) was to be performed, so 14 network processes were 'spawned', each with a different γ parameter. Each network was trained with 1,000 examples, and was evaluated with 1,000 tests. Two statistics were measured as provided by the Erlang `statistics(runtime, wall_clock)` functions. The *wall time* measures the total time elapsed, while the *CPU time* measures the total time spent using each CPU. Each test was repeated 3 times using a fresh Erlang virtual machine instance on each occasion since these times vary slightly due to operating system interference. Figure 6.12 shows the average results of this test. Times for each test had a spread of ~ 1 s.

FIGURE 6.12

Erlang parallelization performance gains for GRNN
How CPU and wall time scale for training and testing a GRNN implementation as a function of the number of CPU schedulers S enabled in the Erlang virtual machine.



This figure clearly demonstrates the drastic reduction in wall time achieved by using a multicore machine. Note that the CPU time increases slightly with S: this is a

consequence of the overhead in inter-core communication and management by the Erlang runtime.

▷ 6.6 Conclusions

The findings and results described in this chapter are summarized thus:

HCAL depth segmentation

- ▷ Noise between logical layers was not (anti-)correlated in this model. Although this would imply that excessive fluctuations on any one channel in a tower could be vetoed by neighbouring channels, the energy measured by each channel in a tower is much reduced because the energy is shared between them. There is no net improvement to the resolution that can be achieved, at least based on the approach adopted in this chapter.
- ▷ We conclude that depth segmentation *a priori* does not help the neural networks if the electronics and readout are the same.
- ▷ Reducing the electronics noise and QE clearly improves the linearity and resolution of the neural networks' performance.

Multivariate analysis

- ▷ Once a multivariate approach has been implemented, it offers quick turnarounds for investigating such ideas as an HCAL upgrade. Training and evaluating a new model of HCAL design only takes a few minutes.
- ▷ The networks described here have not succeeded in improving HCAL performance substantially: the HCAL's coarse transverse segmentation implies there is no extra information for the network to exploit. The longitudinal segmentation scheme evaluated here also provides little new information for the networks to exploit.

Using Erlang & concurrency

- ▷ Erlang has proven to be a useful and interesting language for implementing the multilayer perceptron. There are clear performance gains to be had when the computation under question can be parallelized: for the user this is particularly easy to achieve in Erlang where concurrency is a natural part of the language, and where there are no side effects.
- ▷ Erlang code is concise and highly expressive. It is generally received wisdom that the number of bugs in a program scales with the number of lines of code, so writing concise code is advantageous.
- ▷ Inter-process chatter will compromise performance if elementary Erlang processes communicate frequently relative to the amount of computation they are required to perform.
- ▷ Analysing large datasets in a functional, concurrent language such as Erlang may prove most fruitful in future analysis frameworks.

Chapter 7

Terapixel Calorimetry

The first draft of anything is [terrible].

Ernest Hemingway

▷ 7.1 Introduction

In this chapter we move from the LHC and CMS to prospects for a future linear collider. A description of the *International Linear Collider* project is given, followed by a brief outline of the requirements of the detectors for such a machine. The differences between the collision environments of the LHC and ILC are considered. The motivation for using a particle flow reconstruction technique is outlined. A monolithic active pixel sensor technology designed for calorimetry is described, and results from testbeam experiments conducted at DESY in 2007 are presented. A calculation of the prototype sensor's efficiency is made. We will consider the faults of the prototype design and discuss how their rectification has contributed to a viable second-round prototype. Finally, prospects for future research in this area are discussed.

▷ 7.2 Future linear colliders

The LHC and its associated experiments are designed for making discoveries at the TeV scale. Hadron colliders make for excellent discovery machines but there are substantial disadvantages in using them to conduct precision measurements: first, while the proton centre of mass energy E_{CM} is well specified, the hard scattering process between the partons of the proton does not have a known centre of mass energy; second, initial, and final state radiation, and multi-quark fragmentation make reconstruction of the initial hard scattering process more complicated; third, there are severe constraints on the design of the detectors due to the extreme particle flux in which they operate.

In the context of these issues, an e^+e^- collider becomes an extremely attractive proposition: colliding leptons makes for a simplified reconstruction because the centre of mass energy is precisely specified without the complication of parton density functions. Two next generation colliders are currently under development. For the

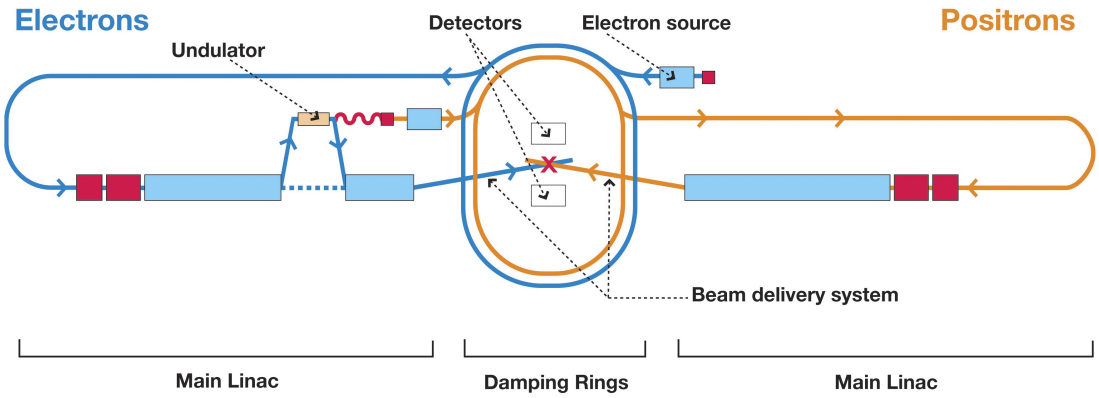
International Linear Collider (ILC), the chief objective is a centre of mass energy of 500 GeV (upgradable to 1 TeV) which may be tuned. There is also the very promising *Compact Linear Collider* (CLIC) project [119]. CLIC uses an entirely different acceleration structure to collide leptons at a nominal energy of 3 TeV. Detectors for the ILC and CLIC will be very similar in any case, and there is the possibility that the ILC and CLIC projects will merge in the near future. The CLIC project is at a less advanced stage than the ILC and will not be discussed any further in this document.

▷ 7.3 The International Linear Collider (ILC)

Authoritative sources for details of the ILC may be found in [120, 121]. A diagram of the ILC is given in Figure 7.1. The machine exploits superconducting RF cavities (1.3 GHz) made from niobium, which give low power loss and higher plug-to-machine efficiency. The average accelerating gradient is 31.5 MV m^{-1}, which is a major challenge for design and production. The initial electron source comes from a photocathode gun, and positrons are generated by *undulators*, which wiggle the main electron beam at 150 GeV and extract the emitted photons which subsequently pair-produce. The main linacs are 11 km long for 500 GeV operation, and the final beam delivery system is 4 km long. A linear collider provides pulsed bunches, and we expect a 5 Hz pulse rate and several thousand bunches per pulse.

FIGURE 7.1

ILC schematic
Placement of major ILC components and structures [122].



7.3.1 Design challenges

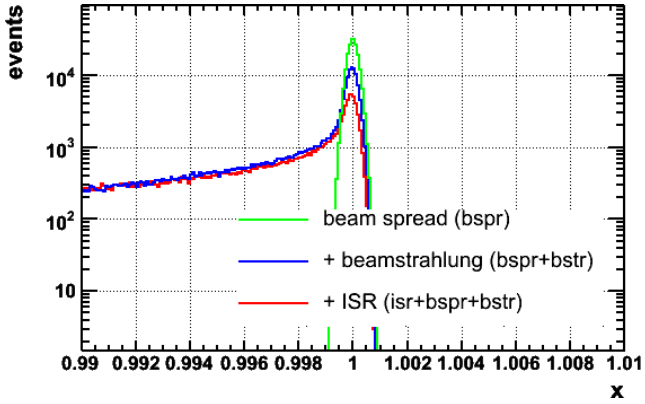
Acceleration

Electrons and positrons are vulnerable to synchrotron radiation under acceleration because they are light in mass. Accelerating muons might be a possibility in the far future, but it would be tricky to achieve the high luminosity required with today's technology. To exceed the energies achieved by LEP without resorting to a collider ring

of larger radius we need a linear collider.[1] Even considering the effect of beamsstrahlung (defined as radiation in the presence of the electromagnetic field of the other beam), we can expect 99% of the events to be at energies greater than 99.5% of the nominal E_{CM} (see Figure 7.2[2]).

FIGURE 7.2

ILC beam parameters
Plot of ILC beam parameters [123], demonstrating that beamsstrahlung and ISR will be well managed at the ILC, so that 99% of events will be delivered at 99.5% of the nominal beam energy. The x-axis indicates the collision energy as a fraction of the nominal energy of 500 GeV.



Luminosity

The luminosity will be comparable to the LHC, at $2 \times 10^{34}\ \mathrm{cm}^{-2}\ \mathrm{s}^{-1}$: this requires tiny beam dimensions at the interaction point, with small emittance and high bunch charge. Also included in the ILC's design are the damping rings, whose objective is to reduce the emittance of the beam. In the final few kilometres of the accelerating structure, a complex beam delivery system rotates the bunches so that they collide head-on and focusses the beam spot to the order of nanometres in each spatial dimension.[3] The final focussing quadrupoles must be stabilized to 0.2 nm for vibrations greater than 4 Hz, which presents a serious engineering challenge. The beam delivery system therefore represents a colossal cost component for any linear collider.

Minimizing beamsstrahlung is essential to providing the nominal (and tunable) E_{CM} with a small width. Indeed, uncertainty in the luminosity spectrum $\mathrm{d}E/\mathrm{d}\ell$ translates into a systematic uncertainty on many precision measurements.[4]

7.3.2 Precision measurements at a linear colander

Polarizable beams

The ILC will also support polarizable beams, with up to 80% polarization for electrons, and 60% for positrons. This adds to the ILC's credentials as a machine for making

[1] Choosing stronger magnetic fields for the bending dipoles would cause the electrons to radiate even more synchrotron radiation.

[2] Measuring E_{CM}: the centre of mass energy can be accurately measured with beam spectrometers, Bhabha scattering and Z boson production.

[3] *c.f.* The LHC beam radius is 16.7 μm [26].

[4] The luminosity uncertainty can be measured from the large angle bias in Bhabha scattering, which is sensitive to a momentum imbalance between the beams (this demands excellent tracking resolution in the forward direction).

precision measurements because the availability of polarizable beams allows spin studies to be conducted.

Higgs physics

The relatively clean environment would allow us to look for rare Higgs decays, investigate the quantum numbers of the Higgs and its self-couplings, and search for other new particles. It is not *impossible* to do many of these measurements at the LHC, but the backgrounds and systematic errors involved will have to be understood to an exacting degree. This will take a very long time given the intrinsic detector performance available [124].

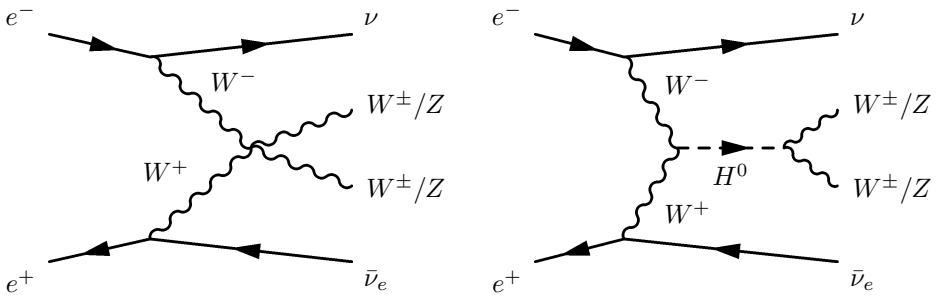
Separating WW and ZZ events

Let us consider the channel $e^+e^- \to \nu\bar{\nu}WW/\nu\bar{\nu}ZZ \to$ jets to illustrate the use of particle flow and the need for good jet energy resolution. We wish to distinguish between the W and Z events by reconstructing the events from the di-jets observed in the detector. The LEP experiments at CERN were able to reconstruct these masses provided there were no neutrinos in the final state. The masses were then reconstructed with the knowledge of kinematic constraints (consult [125] for details of this 'recoil method'). When, however, there is more than one invisible particle, all we have is the energy and angles of the jets.

This channel is particularly important for physics; the relevant Feynman graphs are shown in Figure 7.3. The cross section for WW scattering violates quantum mechanical unitarity without the presence of the scalar boson and its associated propagator, though other mechanisms may regulate this, such as strong electroweak symmetry breaking at the TeV scale [126].

FIGURE 7.3

Vector Boson Fusion and the role of the Higgs
Feynman diagrams for Vector Boson Fusion, where the bosons subsequently decay to jets. The presence of two neutrinos in the final state means that kinematic constraints cannot be used to reconstruct the event; in order to deduce the mass of the W or Z, the di-jets' energies and angles must be accurately known. The second diagram, with the addition of a Higgs propagator, keeps the overall process from being divergent.



(a) No Higgs *(b) With Higgs*

7.3.3 Differences in particle flow from LHC to ILC

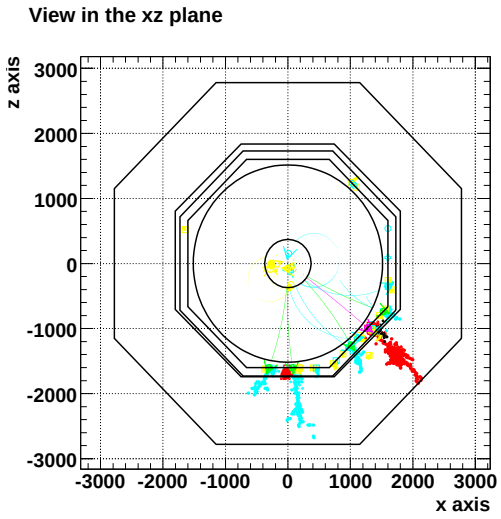
It is worth noting at this point that the particle flow technique at the ILC will likely be very different to the one developed for CMS. The state of the art ILC algorithm is provided

by *Pandora* [127]. An illustration of this algorithm in a typical ILC event is shown in Figure 7.4, together with a benchmark for jet energy resolution.

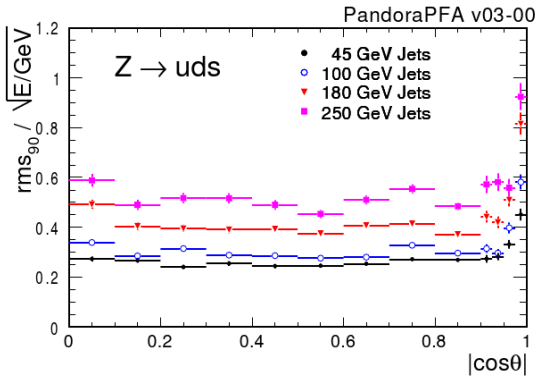
At CMS, event reconstruction is a continuous fight for high efficiency and purity: both in-time and out-of-time pile up (see earlier chapters) affect pattern recognition and degrade the resolution of individually reconstructed particles, *i.e.* the PFCandidates of earlier chapters. Furthermore, the detectors need to be very radiation tolerant and operate at high rate. In particular, the CMS HCAL's coarse granularity reduces the sophistication that can be deployed. For example, we saw that neutral hadrons are inferred from a statistically significant energy excess with respect to the momentum of the track which is linked to the HCAL cluster.

A particle flow reconstruction at the ILC can be much more microscopic in its treatment: for example, the calorimeters are of a sampling design, so we have high spatial granularity in both the transverse and longitudinal directions. The relatively clean environment enhances both the efficiency of detecting the actual particles and the purity of the reconstructed candidates.

FIGURE 7.4 **The *Pandora* particle flow algorithm**
Examples of event reconstruction and jet energy resolution. In the event display, the linking of clusters and tracks is clearly evident, and individual clusters are resolved due to the high transverse and longitudinal spatial resolution available. The jet energy resolution target of $\alpha = 30\%$ at the Z-pole is already attainable in simulation.



(a) An end-on view of a ZH event in the 'LDCo1' detector concept (see Figure 2.15, page 58) where the Higgs decays invisibly



(b) Jet energy resolution defined as α in $\sigma/E = \alpha\sqrt{E}$ attainable with Pandora and the International Large Detector concept in simulation

▷ 7.4 The Monolithic Active Pixel Sensor

Chapter 2 introduced the Monolithic Active Pixel Sensor (MAPS) as technology for instrumenting the active layers of a silicon-tungsten sampling calorimeter. The MAPS is a particle detection device with binary readout. It consists of a sensitive epitaxial silicon layer 12 μm thick mounted on 300 μm of support silicon substrate with an in–pixel comparator and logic. A formal proposal may be found in [128]. The MAPS is designed

to be a swap–in alternative to the standard diode pad, and is made of standard—and hence cheap—CMOS silicon, rather than the high-resistivity (expensive) silicon often encountered in other high energy physics experiments. Monolithic active pixel sensors are generic devices, commonly used for light detection (handheld cameras and suchlike), but this is the first usage of the technology for calorimetry. Figure 7.5 demonstrates the fine granularity of the MAPS compared to the standard diode pad technology.

The INMAPS process

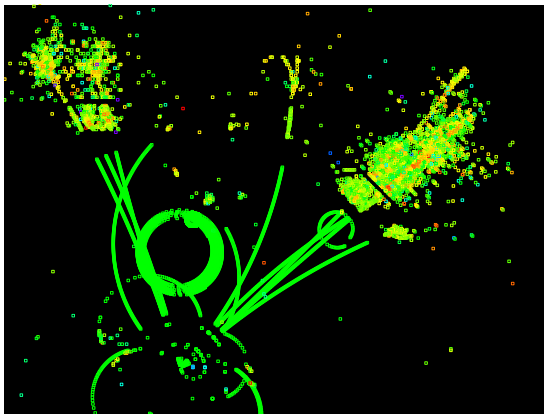
The MAPS takes advantage of a new industrial process formulated specifically for the sensor: the n-wells of the embedded electronics normally suffer from the unfortunate side effect of collecting some fraction of the signal charge deposited in the sensitive epitaxial layer, which reduces the efficiency of the charge collection by the 4 diodes on the pixel. With the novel INMAPS process, a deep p-well layer, 1 μm thick, is embedded under the n-wells to shield the epitaxial layer from the n-wells. This increases the charge collection efficiency of the chip. Figure 7.6 illustrates the concept. While this process was developed in collaboration with a particular industrial partner, it can be exploited at many modern CMOS fabrication facilities.

Sensor parameters

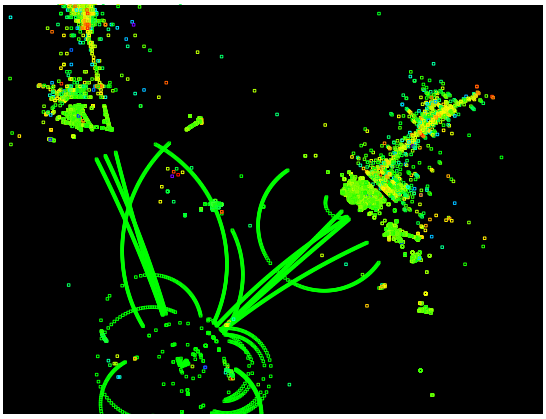
Given a shower density of 100 particles mm^{-2}, the size of the pixel (50×50 μm^2) has been selected to minimize the probability of more than one hit per pixel occurring, while not increasing the number of pixels—and therefore channels to read out—beyond some tractable number. The target noise rate is 10^{-6} per pixel, but with an ECAL of 10^{12} such pixels, one can therefore expect 10^6 noise hits per bunch crossing [130]. The threshold for the MAPS is set on a global basis with an in-pixel trim. The size of the 4 charge collecting diodes on the pixel maximizes the efficiency as a function of charge collection, collection time, and signal to noise ratio, so as to be comparable to the diode pad with $S/N \sim 10$. The design anticipates the noise to be ~ 70 eV (energy equivalent), while a MIP will typically deposit 4.3 keV [128], but the signal to noise ratio is attenuated by charge sharing between pixels, and shall be discussed forthwith.

Table 7.1 gives a short overview of the standard diode pad compared with the prototype MAPS pixel. Subsequent designs will need to address power consumption: the standard ECAL expects a power dissipation of 1 μW mm^{-2}, and no active cooling components. The prototype MAPS design expends 40 μW mm^{-2}, but this is only a proof of concept implementation. We may elect to have a larger sensor; moving to 100 × 100 μm^2 gives a 4 fold reduction in power per unit area. Increasing the integration time of the charge collection may give a factor of 2 reduction. It should also be possible to reduce the operating voltage and move to a smaller feature size, therefore reducing the power consumption even further. Finally, the MAPS does not require a custom designed ASIC chip as the diode pad does so power dissipation will be much more uniform with a MAPS ECAL as the ASIC chip is a 'hotspot' on the standard diode pad.

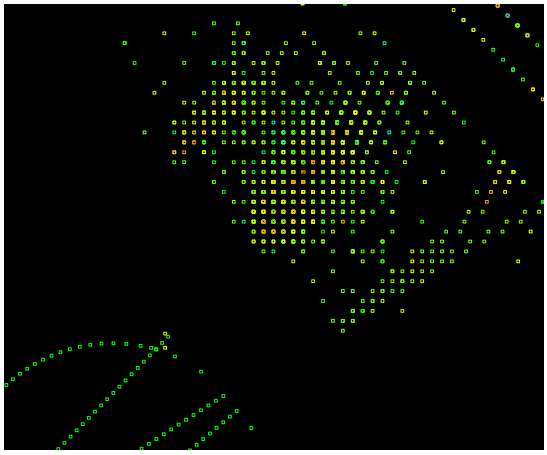
FIGURE 7.5 **Event displays**
Comparison of the standard diode pad technology and the MAPS for a $e^+e^- \to Z \to 2$ jets event using the GEANT4 simulation and the LDCSc01 detector concept [129]. The high granularity afforded by the MAPS is immediately apparent.



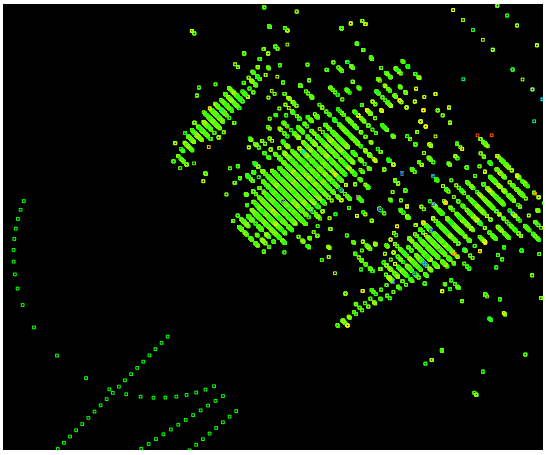
(a) Standard view, no zoom



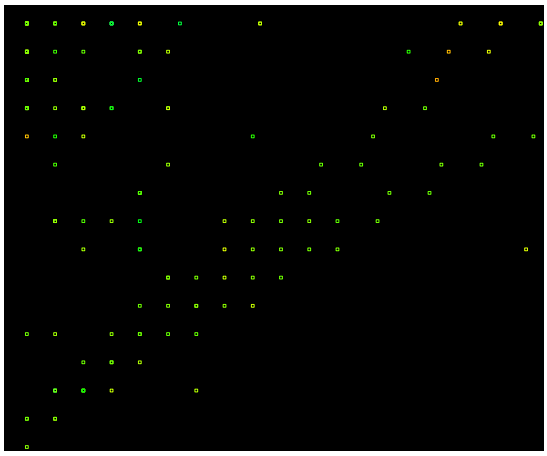
(b) MAPS view, no zoom



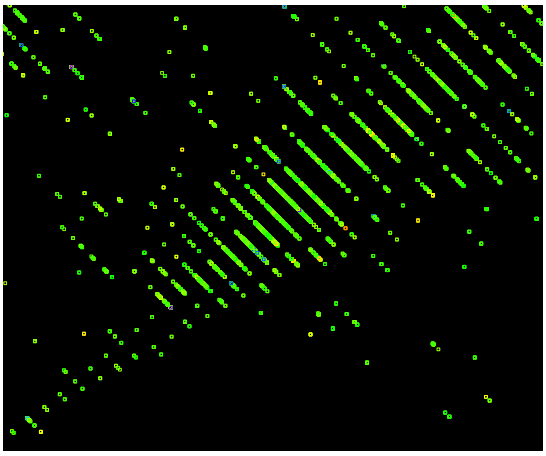
(c) Standard view, zoom. Note the two tracks from two electrons seen at the bottom right.



(d) MAPS view, zoom. More spatial detail in the clusters is visible.

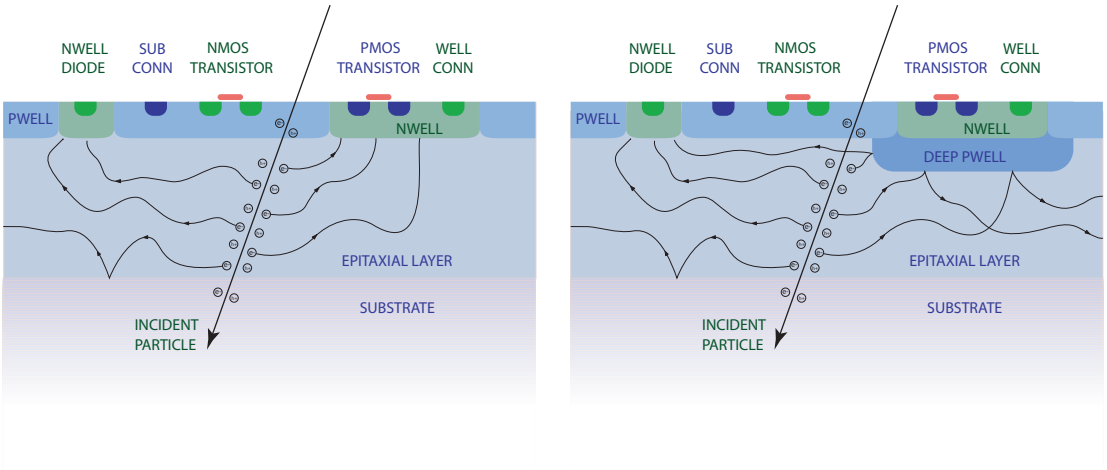


(e) Standard view, more zoom. The two electron's showers overlap and cannot be individually resolved.



(f) MAPS view, more zoom. The particle flow algorithm may resolve the clusters from the two electrons.

FIGURE 7.6 **The INMAPS process**
Conceptual illustration of a charged particle crossing a CMOS sensor epitaxial layer. The n-well diode at the upper left is a signal collecting diode. The one at the upper right encloses a PMOS transistor which is part of the in-pixel circuitry. *Diagrams courtesy of J. Crooks (RAL).*



(a) With no deep p-well the PMOS diode parasitically collects charge.

(b) With the deep p-well implant, more charge is collected by the n-well diode.

TABLE 7.1 **Diode pad versus MAPS: parameters**
Comparison of parameters between the diode pad and the MAPS.

	Diode pad	**MAPS**
Nominal size	0.5×0.5 cm^2 (1)	50×50 μm^2
Construction	high resistivity Si	0.18 μm CMOS
Attachment	conductively glued to PCB	bump bonded to PCB
Readout	digitization by VFE ASIC to ~ 14 bits	binary; 1 bit ADC
Number of channels	$\sim 8 \times 10^7$	$\sim 10^{12}$
Power consumption	1 μW mm^{-2}	< 40 μW mm^{-2}
Heat dissipation	uneven due to ASIC hotspot	even
S/N (2)	~ 10	~ 10

Notes: (1) The test beam implementation uses diode pads 1 cm square, but the design has since been revised to this new size. (2) This is the design target.

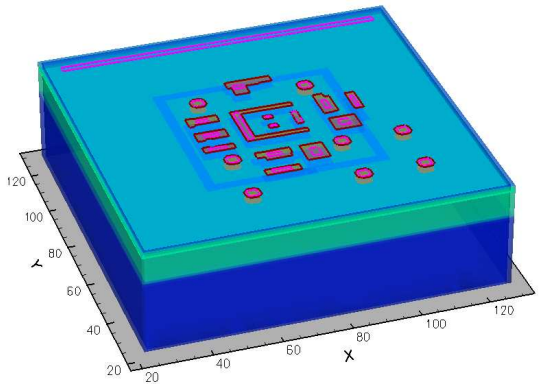
▷ 7.5 Prototype sensor: characterization

Sensor structure

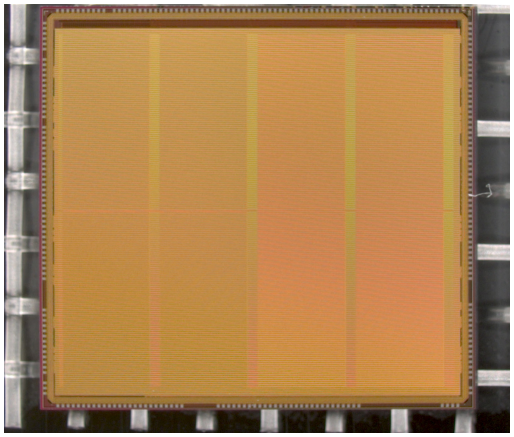
Images of the prototype sensor are shown in Figure 7.7. The sensor was laid out in 4 vertical regions $Q_1, \ldots, Q_4$, each delineated by a memory/readout column. The sensor was also divided in half horizontally with readout logic running underneath each half. These memory and readout sections are not sensitive. Two pixel architectures, *shapers* and *samplers*, were evaluated in the prototype sensor, with the shapers in sections $Q_{1,2}$ and the samplers in $Q_{3,4}$. There were slight variations in the pixel design between the top and bottom half of the sensor.

The pixel threshold is set with a global sensor threshold G, applied by a digital-to-analogue control (DAC) with a range of ± 3,000 *threshold units* (TU). The sensor was designed for readout in an ILC scheme: a *bunch train* (BT) consists of up to 8,000 *bunch*

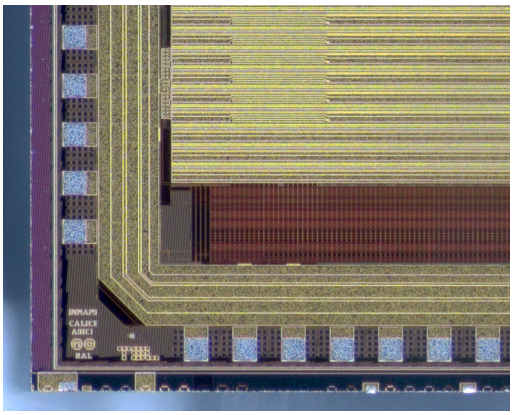
FIGURE 7.7 **The prototype MAPS**
August 2007. From pixel to readout.



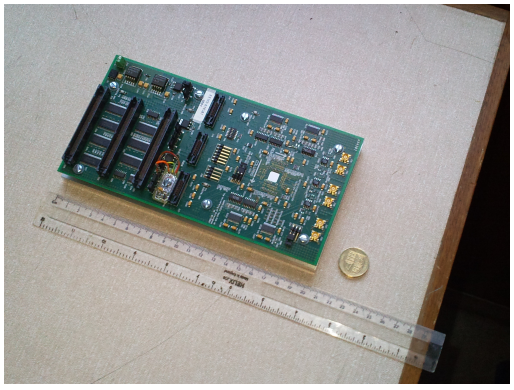
(a) Simulation schematic of a single MAPS pixel: the silicon substrate is shown in dark blue, the epitaxial layer is green, and the INMAPS implantation is pink. The x, y-axis units are microns (for a 50 × 50 μm² pixel).



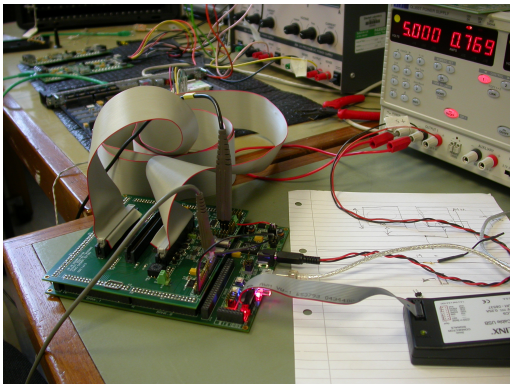
(b) A MAPS array of approximately 27,000 pixels. The four readout columns are visible.



(c) A zoomed-in view of a corner of the array. The rows of pixels are visible in the top right hand corner of the photograph.



(d) Photograph of the readout PCB, with the sensor appearing as a silvery square near centre-right. The CALICE budget is shown on the right of the PCB.



(e) DAQ system designed by M. Noy. The USBDAQ is fitted with a MAPS-specific daughter card on top, which is connected to the readout PCB with ribbon cables.

crossings (BX). If a pixel fires in a BX, then all the hits in a group of 6 pixels in that row are read out for that BX. The memory can only accommodate 19 hits per row per BT: if this limit is reached then all the pixels in that row are marked as 'full' for the remainder of that BT.

Charge sharing & digitization

The charge deposited in the epitaxial layer of the silicon will diffuse and spread across the pixel boundary. Therefore the charge collected in the hit pixel is reduced and may cause neighbouring pixels to fire. This requires careful modelling and a simulation has been implemented in Sentaurus [131]; a schematic of the model is shown in Figure 7.7. On average, the charge collected by the hit pixel is diminished to 35% of what it would be were charge sharing not an effect. Charge sharing and digitization are modelled in the GEANT4-based detector simulation according to the following:

(i) Simulate energy deposition in a 5 μm × 5 μm grid in GEANT4;

(ii) Apply the model of charge sharing evaluated by the Sentaurus simulation. The total energy contained in neighbouring pixels is typically $\mathcal{O}$(50–80%) of the original energy deposited in the central pixel;

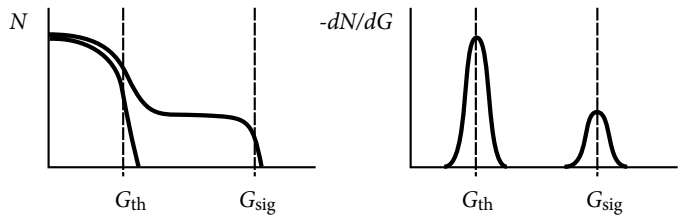
(iii) Sum the energy in each pixel over the 5 μm × 5 μm grid. Gaussian noise is added with width σ = 70 eV;

(iv) For a full detector simulation, add hits solely due to noise (with probability 10^{-6} per pixel[5]);

(v) Keep those pixels with energies above a given threshold, G_{th}.

Noise and ^{55}Fe calibration

Extraction of noise and signal spectra is performed as follows: the global threshold, G, is scanned over a range of values and the number of pixels that fire, N, in each BT is measured. As G is raised past some threshold value, G_{th}, pixels will cease to fire from noise. In the presence of a source of signal, pixels will continue to fire up to some G_{sig}. The spectrum of N can be differentiated with respect to TU to determine the noise and signal curves, as shown here:



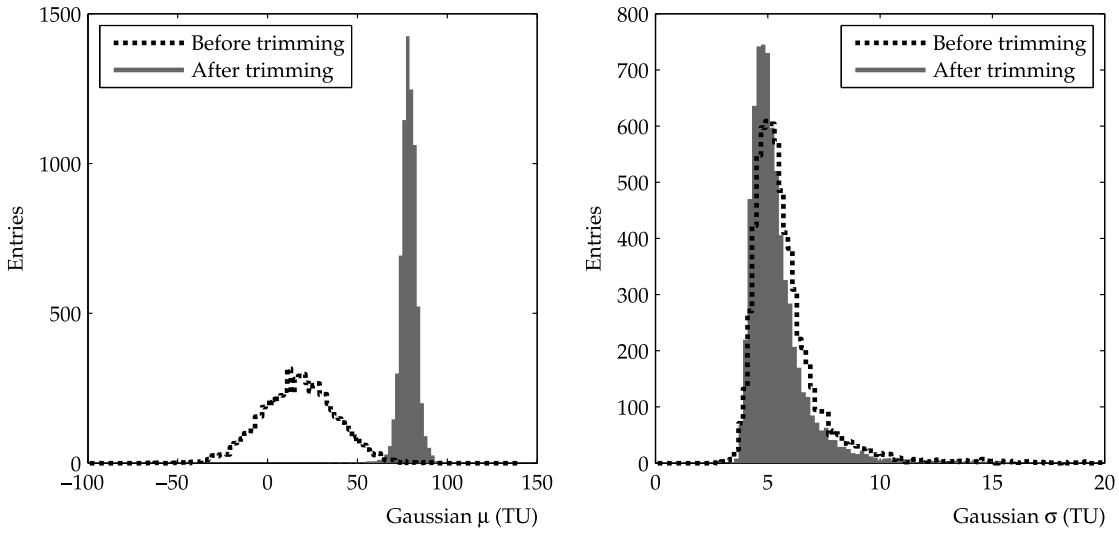
A conversion of TU to electrons or eV is possible by calibrating the sensor using Iron-55. This source emits X-ray photons with a sharp peak at 5.9 keV. The photons interact with the silicon by the photoelectric effect, depositing all their energy in approximately

[5] This is dependent on the sensor threshold.

1 μm^3. This process yields just more than a MIP-equivalent, with an average of 1,640 photoelectrons generated (silicon's average excitation energy is 3.6 eV). But, as noted above, charge sharing attenuates this value so that only a third of these photoelectrons will be collected, giving ~ 540 e^-. The signal for this source was found at $G_{\mathrm{Fe-55}}$ = 200 TU above the noise pedestal: so we conclude that 1 TU corresponds to 2.7 e^-, or 10 eV. A similar measurement of the noise spectrum indicates noise of 75 eV (21 e^-) in agreement (within errors) with the target value of 70 eV. An example is illustrated in Figure 7.8.

FIGURE 7.8

Noise spectra
Each pixel in a quarter of a sensor (7,096 pixels) has its threshold scanned. The process of extracting the noise spectrum as described in the text is applied, which yields 7,096 Gaussian distributions. The left hand plot shows the distribution of the Gaussian means, (*i.e.* the noise pedestals and *not* the individual Gaussians), and the right shows the distribution of Gaussian widths ('before trimming'). The pixel trims are adjusted to unify the noise pedestals ('after trimming'). The mean width indicates the mean noise level: with 1 TU = 10 eV, the noise in this sensor is approximately 60 eV. Improving the trim precision would narrow the width of the distribution of means. To first order, trimming does not change the noise, although the overall working point of the sensor will alter slightly.



(a) Distribution of fitted means

(b) Distribution of fitted widths

^{90}Sr β source test

Early tests conducted by the Author involved placing a Strontium-90 (β emitter) on the pixel. While the source was visible, it manifested itself as a long tail extending to all threshold values. Second, noise hits were still seen even at the highest threshold values. An example of this behaviour is shown in Appendix C.1. This indicated problems in the sensor's design and operation, which were not diagnosed until much later in the project (see §7.6.5). The values given in the section above on Fe-55 calibration were acquired using a 'fixed' sensor.

Gain measurement

The sensor's gain is found using a fixed-intensity laser pulse. Once again, the threshold is scanned and the signal found by differentiating the curve of the number of hits acquired.

This system was automated at Imperial College and allowed many hundreds of pixels to be automatically scanned. The gain was found to be uniform to 12%.

Signal to noise ratio

The total charge available from a MIP is ~ 1,200 e^- [128], so if an average of 420 e^- is collected this yields a signal to noise ratio of $420/21 = 20$. But the worst case scenario is when the charge is deposited at the very corner of the pixel, leading to only 25% of the charge being collected, making the S/N ratio ~ 15. The charge collected would be attenuated by a factor of 10 without the deep p-well implant, and in the worst case only 13 e^- would be collected. The deep p-well implantation is invaluable to the success of the pixel as a particle detector.

▷ 7.6 Prototype sensor: beam test

The results detailed in this section were taken from a beam test conducted at DESY, Hamburg, Germany in December 2007, for which the Author contributed to the analysis.

7.6.1 Overview

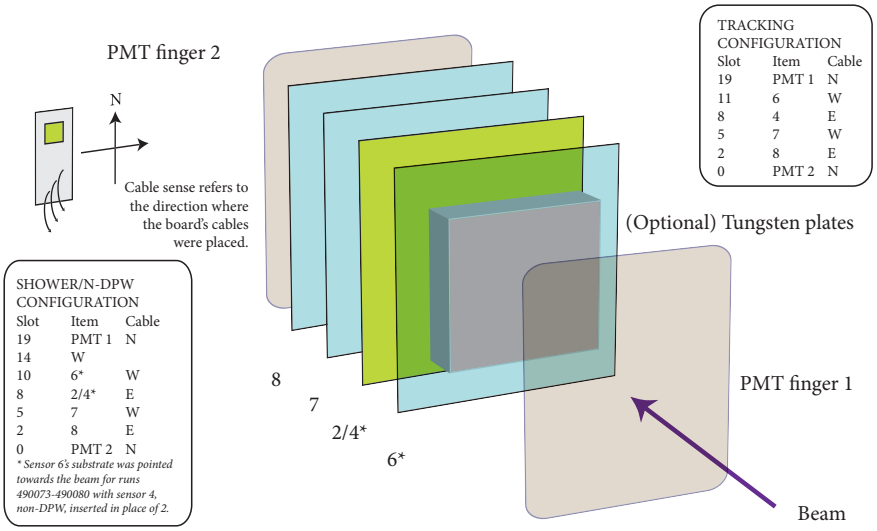
Four prototype sensors were taken to DESY in December 2007. Each PCB was placed in a rigid structure, and the relative orientation of the sensors was periodically altered to suit the investigation required. Sheets of tungsten scintillator were placed between the sensors for some runs to investigate the response of the sensor to electron showers. Finger photo multiplier tubes (PMTs) were placed at the front and rear of the apparatus. These were unreliable and have not been used in the analysis which follows. A diagram illustrating the physical setup is shown in Figure 7.9. The objectives of the experiment were to,

- ▷ Demonstrate that the sensor could be operated successfully, and could see the electron beam;
- ▷ Measure the efficiency of the sensor when operated as a tracker with no tungsten between the sensors. In this chapter we will only consider this mode;
- ▷ Characterize the response of the sensor to electron showers.

The beam facility provided monoenergetic electron beams at energies up to 6 GeV. All the data presented in this chapter concern just two (!) runs taken at the end of the beam test where the sensors were considered to be running relatively well. In both runs, a 6 GeV electron beam impinged on the sensors. At any one time, three of the four sensors were held at a nominal threshold of 120 TU, while the threshold of the 'fourth sensor' was swept from 80 to 200 TU in steps of 10 TU. This process was repeated for each of the four sensors in the stack repeatedly, with stable operation seen over the entire time alloted for the run. In total, ~ 1.5M bunchtrains were recorded (though the actual event rate is approximately an order of magnitude smaller than this number).

FIGURE 7.9

DESY beam test configuration
A schematic of the sensor arrangement for the data taken at DESY. For the runs considered in this chapter, the 'Tracking' configuration was in place. The tungsten plates were omitted for these runs.



7.6.2 Aligning the system

The first task is to align the system (in software): the two outer sensors of the system define a world coordinate system to which the two innermost sensors are to be aligned. We will ignore the possibility that the sensors are skewed, which is reasonable given the quality and stability of the frame in which they were mounted.

Method

(i) Consider the subset of all events where there is at least one hit in each of the two external sensors in one BX;

(ii) Join the hits with a straight line;

(iii) For each hit in one of the internal sensors in the same BX, determine the residual between the expected intersection of the projected line and the hit in the internal sensor;

(iv) Plot this residual (r_x, r_y). Hits for real tracks will coalesce in a peak, whereas random noise hits will produce random and uniform (r_x, r_y). The alignment is determined from the mean $(\bar{r}_x, \bar{r}_y)$ of the distribution.

Consideration of dead areas & rotations

Given the structure of the sensor outlined above, with its dead areas and suchlike, it is important to translate pixel coordinates to physical coordinates. Software was written to account for this mapping. Furthermore, the sensors were placed in alternating orientation, so the beam would see shaper-sampler-shaper-sampler or *vice versa*. For

the sake of brevity, the differences between these two architectures are not explored here.

Alignment results

The results of this process are shown in Figure 7.10: the secondary peaks in the x direction are due to readout corruption of the pixel coordinate. (This was due to a bug which was fixed in the second-round sensor.) A Gaussian is fitted to each distribution and the means, relative to the nominal positioning, are found to be (in mm):

$$r_{x,y|2} = (0.092 \pm 0.019, 0.143 \pm 0.026) \tag{7.1}$$

and,

$$r_{x,y|7} = (0.151 \pm 0.027, 0.102 \pm 0.019). \tag{7.2}$$

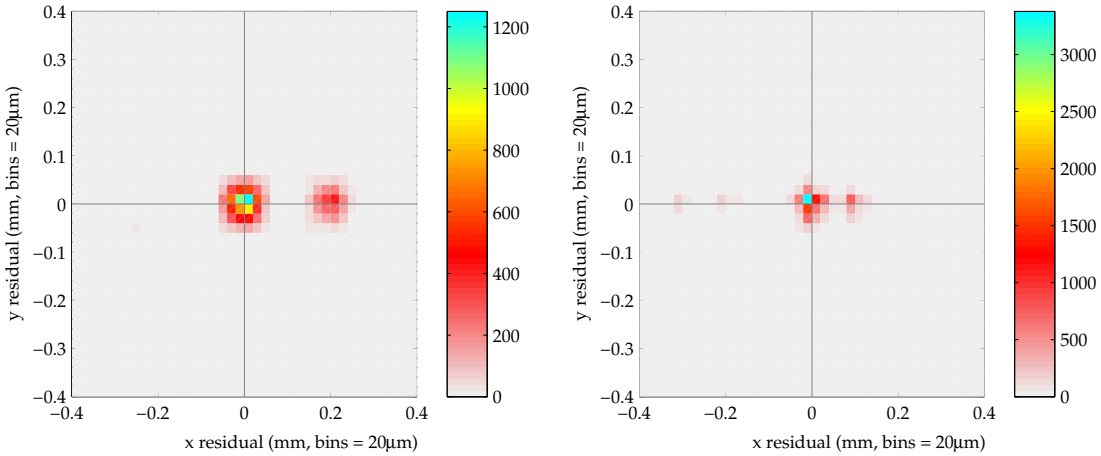
The system was not disturbed over the course of the runs considered, nor did these values change. The widths of these distributions give us the tracking error. It can be shown that the width σ_{fit} of the fitted Gaussian has the correspondence,

$$\sigma_{\text{fit}} = 1.25\sigma_{\text{o}} \tag{7.3}$$

for σ_{o} the intrinsic resolution of the sensor. The width of the residual error is a convolution of the track fit error and the sensor's intrinsic error. This σ_{o} is then used as an input to the efficiency calculation which follows. We may conclude that the tracking error is therefore,

$$\sigma_{\text{o}} = (0.018\,\text{mm}, 0.018\,\text{mm}). \tag{7.4}$$

FIGURE 7.10 **Residuals (r_x, r_y) for two sensors**
Following alignment, the residuals are zeroed. The 'ghost' spots in x are due to a corruption in the pixel column readout.



(a) The first internal sensor (# 2)

(b) The next internal sensor (# 7)

7.6.3 Making tracks

Method

We want to determine the efficiency, ϵ, of a sensor as a function of the global threshold. An electron travelling through the apparatus, in the absence of tungsten, will appear as a MIP. Let us look for events where there is at least one hit in each of the three sensors being held at a constant threshold, and look for a fourth hit in the sensor whose threshold is being scanned. For now, we will assume that the hits must be simultaneous in time (no time skew was found between the sensors). The results of aligning the system (described in the preceding section) will be applied. Furthermore, to reduce noise, and hence increase the purity of the sample, we will apply the following cut on the spatial positioning of the tracks: consider N points in the candidate track, and let a straight line of the form,

$$x(z) = p_0 + zp_1 \tag{7.5}$$

link them, for x being one of the two transverse spatial dimensions, and z the distance from the front of the apparatus in the direction of the beam. Define the χ^2 as,

$$\chi^2 = \sum_{i=1}^{N} \frac{\left[x_i - (p_0 + z_i p_1)\right]^2}{\sigma_i^2} \tag{7.6}$$

where p_j are the fit parameters (to be determined), σ_i is the error intrinsic to the measurement at z_i, and i labels each of the four layers. Let us also take $\sigma_i = \sigma_0$, so the relative positions between sensors are not correlated, though we will *not* assume $\sigma_x = \sigma_y$. In minimizing χ^2, we get a matrix equation,

$$\begin{pmatrix} N & \sum_i z_i \\ \sum_i z_i & \sum_i z_i^2 \end{pmatrix} \begin{pmatrix} p_0 \\ p_1 \end{pmatrix} = \begin{pmatrix} \sum_i x_i \\ \sum_i x_i z_i \end{pmatrix}. \tag{7.7}$$

This matrix is inverted to find the p_j fit parameters. The process is repeated for the y dimension,

$$y(z) = q_0 + zq_1. \tag{7.8}$$

We then get,

$$r(z) = \begin{pmatrix} p_0 \\ q_0 \end{pmatrix} + z \begin{pmatrix} p_1 \\ q_1 \end{pmatrix}. \tag{7.9}$$

We may therefore evaluate χ^2 for each track.

Evaluation

The probability of obtaining this χ^2 for the number of degrees of freedom (which is $N-2$, for the 2 p_j) is then used to distinguish fake tracks from real tracks.[6] The χ^2 probability for x is shown in Figure 7.11. This assesses the reliability of the alignment and error estimation described above. The observed distribution indicates the tracking error has been overestimated due to the observed bias at high probability, since an underestimate of the tracking error would reject tracks that were formed by the beam, and hence

[6] These probabilities evaluate the likelihood of obtaining a given value of the χ^2 for a given number of degrees of freedom *not* being due to chance. A standard function to evaluate this probability is provided and documented in the ROOT package, specifically the `TMath::Prob(chi2, ndf)` function.

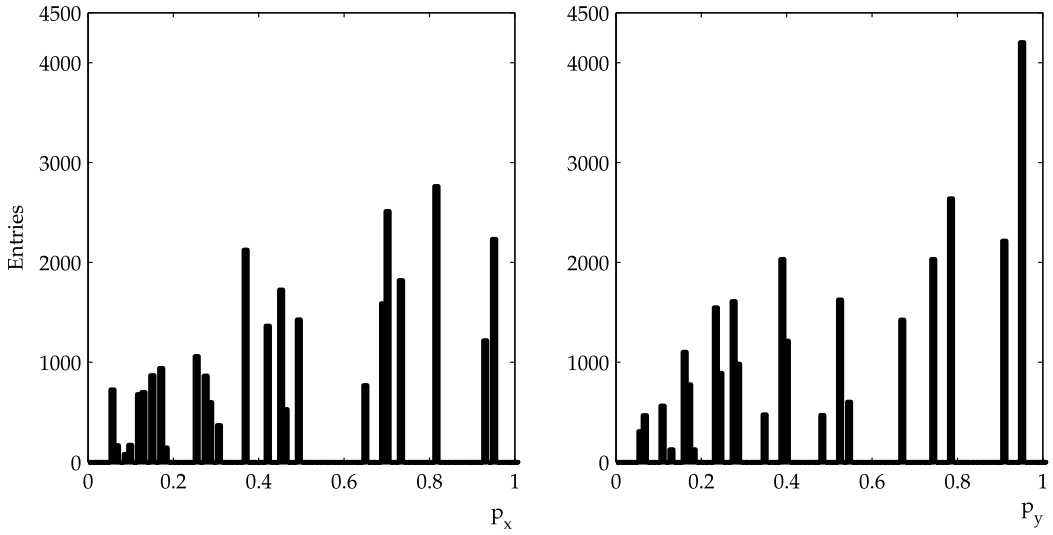
exhibit a bias towards low probabilities. Overall, we conclude that the tracking error could be reduced. Lastly, note that the overall χ^2_{tot} is just addition in quadrature,

$$\chi^2_{\text{tot}} = \chi^2_x + \chi^2_y. \tag{7.10}$$

A cut of $p_x, p_y > 5\%$ was used to isolate real tracks from noise.

FIGURE 7.11

χ^2 probabilities for x, y spatial dimensions
The entries refer to all 3 and 4 hit tracks with $p_x, p_y > 5\%$. The spectrum observed is not smooth due to the finite combinations of pixel hits relative to a fitted track.



We may also evaluate the angle, θ_z, that the track makes with the beam axis. Taking the dot product of the direction component of Eq. 7.9 with the unit vector in the z direction yields,

$$\cos\theta_z = \frac{1}{\sqrt{p_1^2 + q_1^2 + 1}}. \tag{7.11}$$

This angle is shown in Figure 7.12, and clearly demonstrates the efficacy of finding likely beam candidates. The peak at very small values of θ_z indicates that the apparatus was well aligned with respect to the beam.

7.6.4 Efficiency results

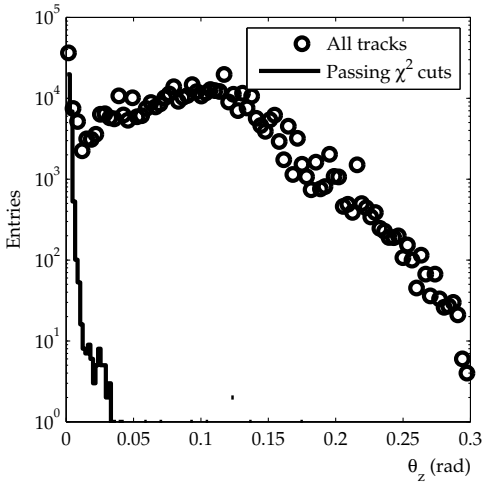
We will consider candidate tracks to be defined as having one hit in each of the three sensors held at nominal threshold, which pass the quality cuts defined above. The question is then asked: is there a hit in the sensor whose threshold is being scanned, which is consistent with a 4 hit track with a χ^2 probability that will also pass the quality cut? We count the cases where this is true. Candidate tracks are discarded where the projection would intersect with a dead area on the fourth sensor. The efficiency is then,

$$\epsilon_{\text{sensor}} = \frac{n_4}{n_3 + n_4} \tag{7.12}$$

for $n_{3,4}$ the number of 3 and 4 hit tracks respectively.

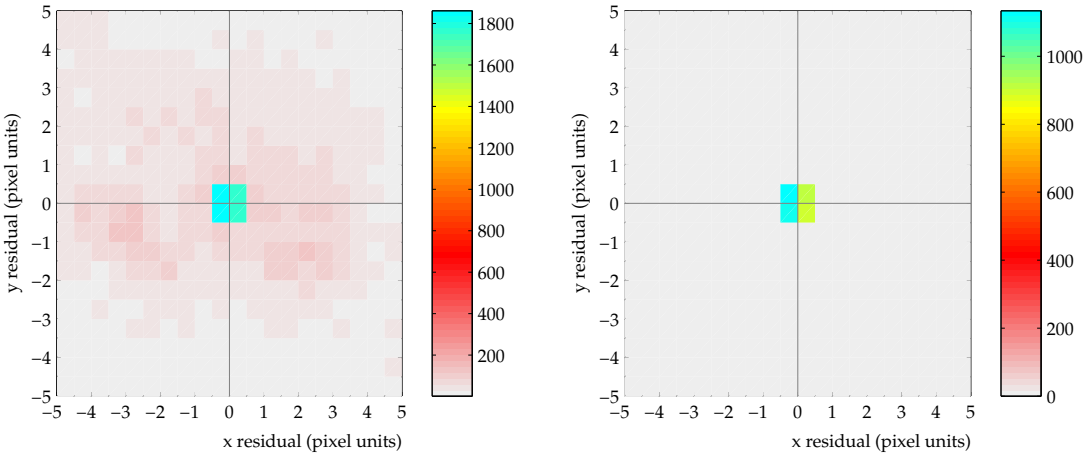
In Figure 7.13, we plot the residual between each 3 hit track's projected position to the fourth sensor and the hit found in that sensor. A peak is seen at (0, 0). We can select

FIGURE 7.12 **θ_z for candidate tracks**
Real tracks have a narrow angle w.r.t. the z-axis of the experiment, as expected.



good four hit tracks by applying the quality cuts. This successfully isolates real 4 hit tracks in the peak from fake 4 hit tracks otherwise made by random noise.

FIGURE 7.13 **Residuals of the fourth hit for a candidate track**
The residual is defined as the difference between the projected position of the candidate 3 hit track and the hit found in the 4th sensor.



(a) No track quality cuts applied *(b) Track quality cuts applied*

In total, 28,394 tracks were found which passed quality cuts. 1,405 tracks were discarded since the projected track would intersect with a dead area of the fourth sensor. Considering the efficiency for all sensors, averaged over all thresholds, we find:

$$\bar{\epsilon} = \frac{\langle n_4 \rangle}{\langle n_3 \rangle + \langle n_4 \rangle} = \frac{3,819}{23,466 + 3,819} = 14.0\% \tag{7.13}$$

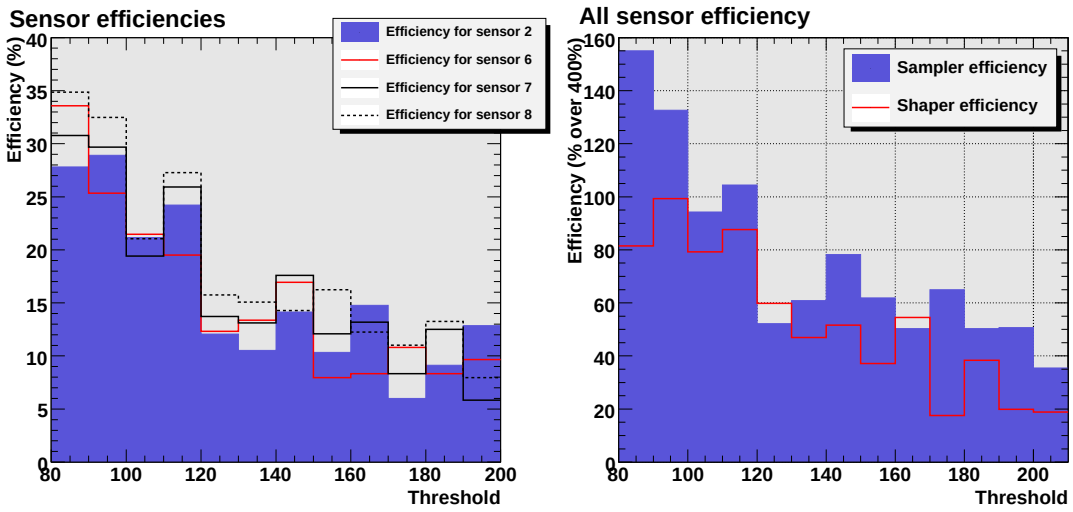
By way of cross-check, the DESY accelerator was known to provide electrons at a rate R_{DESY} $\mathcal{O}(1\,\text{kHz}) \Rightarrow\sim 3\, e^-$ BT $^{-1}$ (1 BT lasts 3.2 ms). Given that there are 3,819 4-hit tracks, and we assume that the efficiency for all sensors is the same at $\bar{\epsilon} = 14\%$, then we have

3,819 = $\tilde{\epsilon}^4 R_{\text{MAPS}} N$, where N is the number of BT recorded (1.5M). Therefore R_{MAPS} = 3,819/$(0.14^4 \times 1.5 \times 10^6)$ = 6.6 e^- BT $^{-1}$. R_{MAPS} and R_{DESY} are in disagreement because, despite removing candidate tracks which intersect with a dead area of the threshold-scanned sensor, tracks that intersect with a pixel in a group which is 'full' have not been vetoed. Accounting for this effect would improve the efficiency by a few percent, and improve the agreement between the two values.

The efficiency was also evaluated for each of the four sensors taken to the beam test, and for each threshold. The results are shown in Figure 7.14. A separate histogram is shown for each of the shaper and sampler architectures. For the case of shaper efficiency, this refers to the case where the track intersects with the shaper region of the sensor whose threshold is being scanned, and *vice versa* for the sampler plot.

FIGURE 7.14

Sensor efficiency
On the left, sensor efficiency for each sensor individually, and, on the right, the total as a function of threshold (note normalization to 400%), for each of the *shaper* and *sampler* architectures.



7.6.5 Understanding the low efficiency

The low efficiencies were an early confirmation that the working point of the sensor had not been optimized, and that the S/N ratio was lower than expected. For example, due to the sampler's design, noise accumulates during the bunch train, so the pixel is more likely to fire for later bunch crossings. This explains why the samplers appear (apparently) more efficient. This is typical of many phenomena that were unknown at the time of the beam test. To summarize:

- ▷ While the β source was visible, the spectrum observed was not sensible;
- ▷ Noise hits were sporadically seen at high thresholds;
- ▷ The sensor exhibited very low efficiency in testbeam.

Subsequent investigations have revealed that the last two of these problems are attributable to one cause: recall that a pixel's threshold is set with a global threshold G which can be adjusted by a 4-bit in-pixel trim, in turn offering adjustment in steps of 15 TU above or below G. The pixels appear to have a broad spread of natural offsets δg_i

relative to G: in the absence of applied trims therefore, while the global sensor threshold is set to G, individual pixels experience a different value of $G + \delta g_i$. Those pixels with natural offsets far below the global G are responsible for filling the memory columns early in a BT, thus reducing the pixel efficiency in testbeam.

Furthermore, the β spectrum continued to very high thresholds because G has been found to saturate at high values $\mathcal{O}(300\,\text{TU})$ despite the control extending to ± 3000 TU.

These reasons explain why noise hits were seen at very high G: those pixels whose natural offsets relative to G are highly negative ($\Rightarrow$ large $-\delta g_i$) can contribute noise hits. Given that G saturates, this may continue to very high applied G. The spread in offsets has subsequently been characterized, and the trims can be set to compensate for the natural offsets. Unfortunately, the precision and range of the trim was not sufficient in the prototype sensor to yield highly satisfactory operation; this was addressed in the second sensor's design. The distribution of natural offsets, δg_i, has been attributed to variations in the size of the pixels' components, which were minimized to fit into the available pixel area.

Despite the apparently low efficiency, the sensor does work and the remarkable result is that this sensor could, in principle, be used for particle tracking, as well as calorimetry.

▷ 7.7 Outlook & summary

7.7.1 The second-round sensor

A second-round design has been tested in a similar manner to the prototype sensor described in this chapter. The design addressed a number of issues to fix the low efficiency found at the beam test, most importantly offering improved trim precision. Preliminary investigations show that this sensor operates well and with high efficiency.

7.7.2 Summary

Clearly the MAPS is a fledgling technology. Precision measurements at a future linear collider will require precision instrumentation. The MAPS offers a promising technology for particle flow in such an environment. It demonstrates a calorimetry technique where counting the *number* of secondaries in an electromagnetic shower (or hadronic cascade) provides a measure of the original particle energy. The sensor has high spatial granularity and a good signal to noise ratio. A position resolution of 18 μm has been attained. The low detection efficiency seen at the testbeam has subsequently been characterized and addressed. Further studies will need to address how to best exploit this in physics reconstruction, particularly in calorimetry.

Part IV

Epilogue

Chapter 8

Conclusions

Now we summarize the main results of the work presented.

Calorimeter response at CMS

QCD forms a large background to many physics processes at the LHC. The CMS detector is optimized for the reconstruction of leptons and photons, because their signatures are relatively distinct. Good jet reconstruction is nonetheless important, and the CMS calorimeters play a major role in this task. The non-linear and non-compensating nature of these calorimeters' response to hadrons has been demonstrated, characterized, and compared with simulations. The hadronic calorimeter's intrinsic hadronic energy resolution is poor, at $\sigma/E = 111.5\%/\sqrt{E} \oplus 8.6\%$ (for the barrel in testbeam, Eq. 1.2). Three outstanding issues remain at the time of writing:

- The simulations require further tuning of the response for low energies $\mathcal{O}(<$ 10 GeV) (Figures 3.7 and 3.9);
- The fast simulation of the ECAL's response to hadrons in the endcaps, for hadrons which deposit more than a MIP equivalent there, is sub-optimal (Figure 3.3);
- The overall calibration point of the simulation does not match that of the data, namely that a 50 GeV pion depositing a MIP equivalent in the ECAL should produce a response of 50 GeV in the HCAL (Figure 3.8).

The CMS particle flow reconstruction

The CMS particle flow algorithm combines CMS' subdetectors in an optimal way to build a list of PFCandidates describing the outcome of the original parton-parton hard scatter. The reconstruction of hadrons forms a major part of the analysis, and Chapter 4 describes how particle flow does this by creating/using calorimeter clusters, tracks, and 'global' muons which are subsequently topologically linked and analysed to create the so-called PFCandidates (Figures 4.1 and 4.4).

Over 60% of a jet's energy is transported by charged hadrons, with an exponentially falling energy spectrum (Figure 1.11). Neutral hadrons only account for a small fraction $\mathcal{O}(< 5\%)$ of a jet's energy. The CMS calorimeters must be calibrated (Figure 4.8) to recover a linear and uniform response (Figure 4.13), which gives a modest improvement to the

energy resolution of charged hadron PFCandidates (Figure 4.16). Various aspects of the PF reconstruction of hadrons were described, in particular the inference of photons and neutral hadrons based on statistically significant excesses of calorimeter energy with respect to track momentum.

The systematic uncertainty affecting PF jet resolution and response that results from using an inappropriate hadron calibration was also assessed. In the barrel, the resolution was unaffected, but the response was found to change in accordance with expectations (Figure 4.20). In the endcaps, the resolution was degraded for high p_T $\mathcal{O}(> 200\,\mathrm{GeV}/c)$ jets. In any case, correctly calibrated PF jets offer a clear improvement in resolution compared with CaloJets built from calorimeter information alone (Figure 4.21).

A first look at collision data taken in late 2009 demonstrated that the particle flow method is readily applied to the data [89]. The calibration of hadrons should be carefully monitored, for the calibration derived from testbeam data is more appropriate when correcting the collision data than the equivalent calibration derived from the fast simulation (Figure 5.2).

Multivariate techniques applied to hadron reconstruction

An upgrade of the CMS HCAL will improve the electronics noise and potentially the overall detector resolution. The neural networks of Chapter 6 provide a convenient way to investigate the various scenarios. For a given model of electronics, no improvement to the detector resolution was found when the depth segmentation was increased: the energy deposited in each channel is reduced and therefore proportionately smaller compared to the electronics noise. Despite offering good linearity (Figure 6.8), the neural networks did not substantially improve the HCAL resolution: the HCAL's transverse segmentation is too coarse to offer any extra information for neural networks to exploit (Figure 6.10). This is because the lateral size of an HCAL tower is similar to the typical transverse extent of a hadronic cascade.

The use of Erlang, a functional and concurrent language, was demonstrated. For problems that can be parallelized, Erlang offers the possibility of excellent gains in computing performance on machines with many processor cores (Figure 6.12). This is important given the current trend of (a) stagnation of microprocessor clock cycle speeds, and (b) the increasing prevalence of multicore processors.

The Monolithic Active Pixel Sensor

The last chapter of this thesis described a novel pixel sensor technology. Detectors at the proposed International Linear Collider are built with particle flow reconstruction in mind, to better conduct precision measurements. A sampling calorimeter design is advantageous because the calorimeter clustering may be more sophisticated. The small pitch of the MAPS at $50 \times 50\mu\mathrm{m}^2$ together with its binary readout can provide fine grained calorimetry where the secondaries of the particle showers are directly sampled. The MAPS also uses a deep p-well implant between the epitaxial layer and the chip's electronics to reduce the charge parasitically collected. Despite the prototype sensor's low efficiency in testbeam, for reasons now understood, the position resolution of the sensor was found to be 18 μm (Eq. 2.28). The noise was measured to be 75 eV, and the signal to noise ratio determined to be in excess of 10.

Appendices

Appendix A

CMS Testbeam reconstruction details

▷ A.1 Optimization of energy reconstruction

Also advertised as 'The Banana Corrections'

This appendix summarizes the method detailed in [36] for correcting the ECAL and HCAL energies in testbeam. In all stages, energies are taken from a 7×7 matrix of ECAL crystals and 3×3 matrix of HCAL cells centred on the beam impact point. The correction proceeds in three stages. First, pions which deposit the equivalent of a MIP in the ECAL are used to correct the HCAL energy. Second, the ECAL correction is then derived using the corrected HCAL energies and the beam constraint for all events. Finally, the overall ECAL and HCAL response is corrected as a function of the ECAL energy by means of a cubic polynomial.

For those events where a MIP equivalent is seen in the ECAL (see Chapter 3), the quantity π/e, exhibited in Figure 2.13, is parametrized as a function of the HCAL response. Substituting Wigman's parametrization (Eq. 2.27) into Eq. 2.25, and fitting to the data gives,

$$(\pi/e)_{\mathrm{HB}} = 1.4. \tag{A.1}$$

The corrected HCAL response is then,

$$E^{*}_{\mathrm{HB}} = \frac{E_{\mathrm{HB}}}{(\pi/e)_{\mathrm{HB}}} \tag{A.2}$$

from which we deduce the expected ECAL energy deposit to be,

$$\langle E_{\mathrm{EB}} \rangle = p_{\mathrm{beam}} - E^{*}_{\mathrm{HB}}. \tag{A.3}$$

This then gives the correction factor $(\pi/e)_{\mathrm{EB}}$ as,

$$(\pi/e)_{\mathrm{EB}} = \frac{E_{\mathrm{EB}}}{p_{\mathrm{beam}} - E^{*}_{\mathrm{HB}}}. \tag{A.4}$$

This quantity is also parametrized, but with an *ad-hoc* function of the form,

$$(\pi/e)_{\mathrm{EB}} = a \log\left(E_{\mathrm{EB}}/1\,\mathrm{GeV}\right) + b, \tag{A.5}$$

with values $a = 0.057 \pm 0.006$ and $b = 0.49 \pm 0.04$. The corrected ECAL energy is therefore,

$$E^{*}_{\mathrm{EB}} = \frac{E_{\mathrm{EB}}}{(\pi/e)_{\mathrm{EB}}}. \tag{A.6}$$

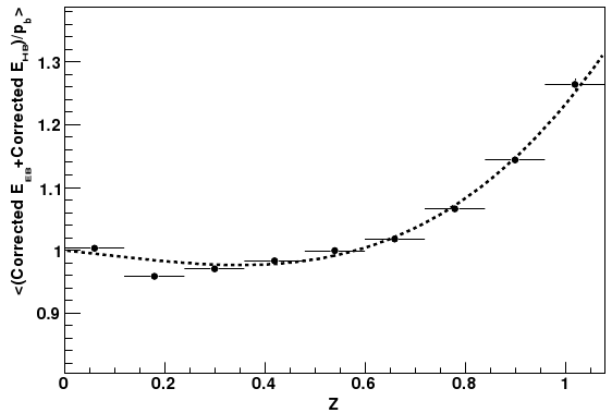
At this point, one could terminate the procedure and let $E_{\mathrm{tot}} \equiv E^{*}_{\mathrm{EB}} + E^{*}_{\mathrm{HB}}$. It is observed however that E_{tot} overestimates the true energy when there is a substantial fraction of energy deposited in the ECAL, defined by $Z \equiv E_{\mathrm{EB}}/(E_{\mathrm{EB}} + E_{\mathrm{HB}}) \geq 70\%$. These cases indicate that the shower fluctuated to a relatively large fraction of π° mesons, for which the ECAL response is enhanced. The variation of E_{tot} with respect to the beam momentum can be parametrized as,

$$\left\langle \frac{E^{*}_{\mathrm{EB}} + E^{*}_{\mathrm{HB}}}{p_{\mathrm{beam}}} \right\rangle = \alpha Z^{3} + \beta Z^{2} + \gamma Z + 1, \tag{A.7}$$

which gives the characteristic 'banana' shape exhibited in Figure A.1.

FIGURE A.1

The 'banana' correction
From [36]: " The π/e corrected response ratio for 100 GeV/c pions of the combined system as a function of the EB fraction. The Z value is defined as $Z \equiv E_{\mathrm{EB}}/(E_{\mathrm{EB}} + E_{\mathrm{HB}})$, [the] ratio of raw energy deposit in the EB with respect to the total in the calorimeter. The smooth curve is a third order polynomial fit to the data [See Eq. A.7]."



The fitted values are $\alpha = 0.412 \pm 0.045$, $\beta = -0.096 \pm 0.058$ and $\gamma = -0.084 \pm 0.018$. This final linearization is not correlated with the beam momentum. The final corrected energy is then,

$$E^{*}_{\mathrm{tot}} = \frac{E_{\mathrm{tot}}}{\alpha Z^{3} + \beta Z^{2} + \gamma Z + 1}. \tag{A.8}$$

▷ A.2 Vetos to be applied

Muon vetos should be used for all energies. CK2 can be used to reject electrons and the TOF (time of flight) to reject protons and kaons for energies of 9 GeV and less. For 2 and 3 GeV, CK3 can be used to reject events with double electrons ('double electron tagging'). For energies greater than 5 GeV, CK3 is used to remove protons and kaons.

The cut values to be applied are detailed in the table below, and were obtained from [132] and [71]. *Interested parties are advised that cleaned collections of events are available on the Author's CASTOR space, and that they should contact the Author directly to save themselves the hassle of reimplementing them!*

TABLE A.1 **ADC cut values for beam halo, scintillator, muon veto, *etc.* counters at the H2 testbeam 2006 and 2007 setups**
All vetos require the signal value to be less than the veto value to pass (unless otherwise stated).

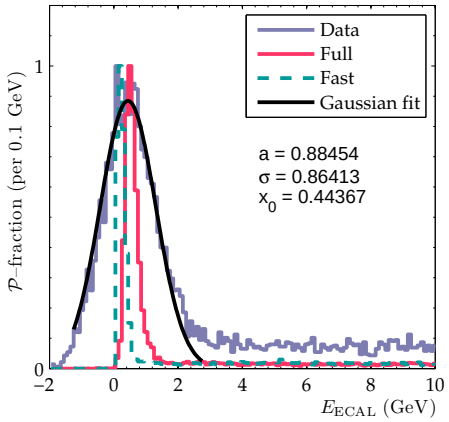
Scintillators and Beam halo vetos									
S1	S2	S4	BH1- BH4						
210	360	370	Per run(1)						
Muon vetos									
VM1	VM2	VM3	VM4	VM5	VM6	VM7	VM8	VMF	VMB
127	125	100	185	170	154	175	135	120	160
Čerenkov		**Time of flight(2)**							
CK2	CK3	≤ 6 GeV	7 GeV	8 GeV	9 GeV				
48	480(3)	-545	-550	-555	-565				

Notes: (1) See § A.4. (2) In ADC units. In the endcap experiment, the TOF's sign appears reversed. The time of flight is computed from the *Saleve* (S1, S2) and *Jura* (J1, J2) times as TOF= [(S1 + J1) – (S2 + J2)]/2. (3) 2, 3 GeV require < 480; otherwise > 480.

▷ A.3 Effect of selective readout

For technical reasons, it is only feasible to study this in simulation or apply ZSP and SR effects at the hardware level while taking data. We saw that the simulations accurately model the ECAL response for 50 GeV pions in the barrel (Figure 3.3). In Figure A.2, we see the effect of applying selective readout to the ECAL components in simulation and it clearly has a substantial impact, for hadrons behaving as minimum ionizing particles are barely resolved.

FIGURE A.2 **Effect of selective readout**
ZSP and SR were *enabled* in the simulation. To be compared with Figure 3.3 (a).



▷ A.4 Stabilized gaussian fits

While the specific ADC values for these cuts were the same as those used in [36], some had to be computed dynamically on a run-by-run basis: for the beam halo counters, with varying pedestals, a stabilized gaussian was fitted to a histogram of ADC values. Given the histogram's arithmetic mean, X, and RMS, R, the gaussian's initial fit range f_1 was restricted to,

$$X - 3R < f_1 < X + 3R. \tag{A.9}$$

This gaussian then has mean, μ_1, and width, σ_1. A second gaussian was then fitted to the histogram with the range f_2 restricted to,

$$\mu_1 - 3\sigma_1 < f_2 < \mu_1 + 3\sigma_1. \tag{A.10}$$

For the second gaussian fit with mean, μ_2, and sigma, σ_2, we can choose the ADC threshold value to be $\mu_2 + 2\sigma_2$. The factor of 2 is arbitrary but represents a conservative cut.

Appendix B

Particle flow details

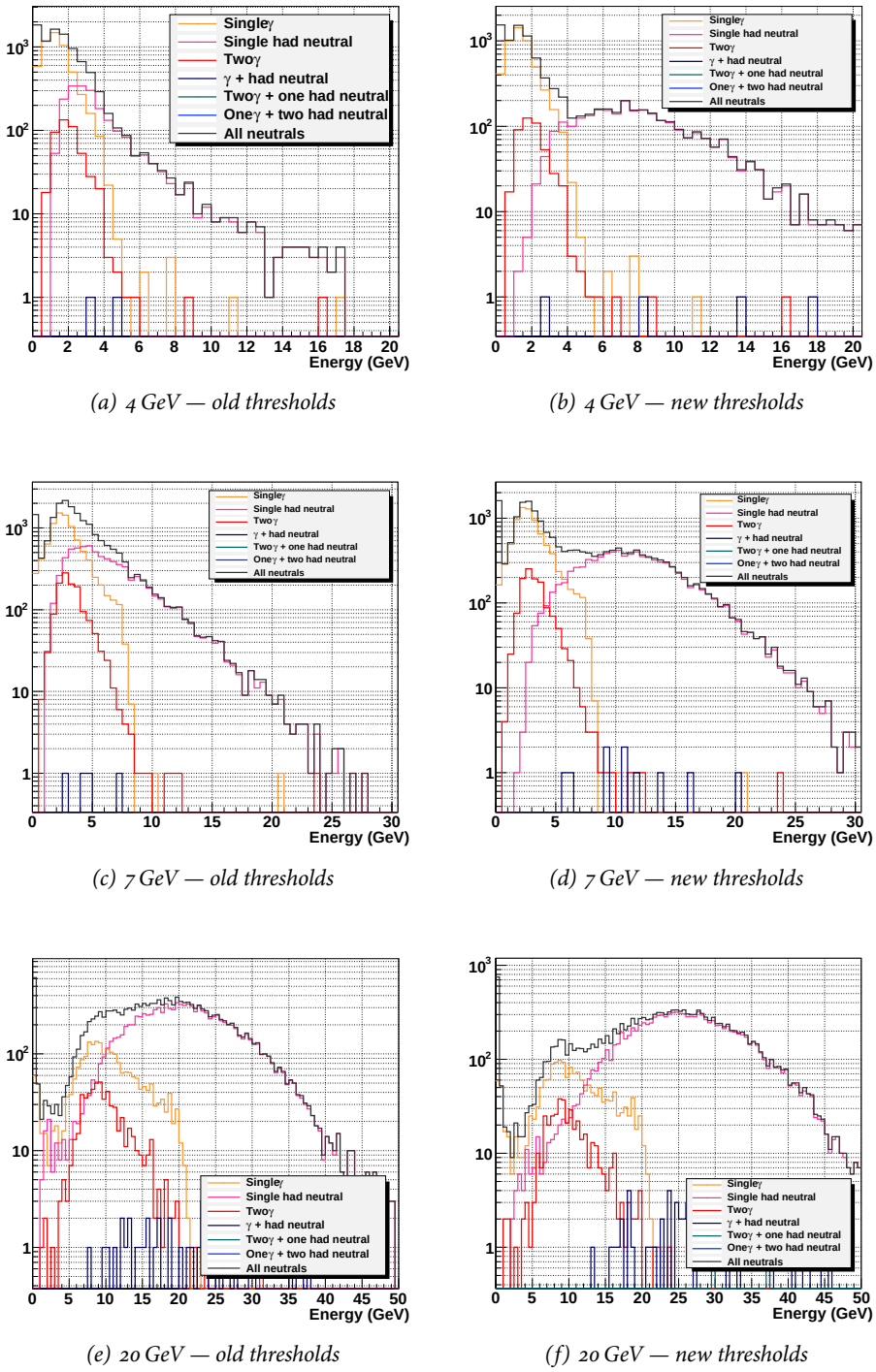
B.1 Effect of seed thresholds on reconstruction efficiency

Table B.1 below presents the efficiency for reconstructing at least one hadronic neutral in the barrel testbeam data as a function of seed and cell threshold at 4 energies. The objective is to maximize the energy fraction associated with these hadronic neutrals. Figure B.1 shows the effect on PFCandidate energy spectra.

TABLE B.1 **Efficiency for reconstructing at least one hadronic neutral (barrel) for various energies and seed/cell thresholds**
Changing the cell threshold has little effect. Lower values of the seed threshold yield greater efficiency, particularly at low energy. The column headings γ and $2\,\gamma$ refer to the number of PFCandidates of type γ reconstructed.

Energy	**Seed (GeV)**	**Cell (GeV)**	γ	2γ	**Total**	$\epsilon_{\text{neutral}}$	$\frac{E_{\text{hadronic}}}{E_{\text{total}}}$
2 GeV	0.6	0	2055	40	3627	0.42	0.14
	1.4	0.8	2139	40	3627	0.40	0.08
4 GeV	0.6	0	4881	453	9196	0.42	0.29
	0.6	0.6	4881	453	9196	0.42	0.29
	1.4	0.8	5329	466	9196	0.37	0.23
7 GeV	0.6	0	7855	1207	19429	0.53	0.45
	1.4	0.8	8990	1324	19429	0.47	0.40
20 GeV	0.6	0	1762	408	15275	0.86 (0.82)	0.81
	1.4	0.8	2319	589	15275	0.81 (0.80)	0.77

FIGURE B.1 **PFCandidate species**
Neutral energy spectrum for testbeam pions at three energies, running at the old seed and cell thresholds (left column), and the new optimized thresholds (right column). Data presented relates to the barrel calorimeters.



(a) 4 GeV — old thresholds

(b) 4 GeV — new thresholds

(c) 7 GeV — old thresholds

(d) 7 GeV — new thresholds

(e) 20 GeV — old thresholds

(f) 20 GeV — new thresholds

▷ B.2 Particle flow calibration functions

B.2.1 Testbeam calibration functions

Enumerated here 'for the record' are the calibration functions derived from testbeam. These are drawn in Figure 4.8, page 91. First, the barrel:

```
f(ecal) = 1.15921+(0.81047)*exp(-x/39.6)-exp(-x^2/2.68)*3.319
f(hcal_int) = 1.00336+(-1.37298+0/sqrt(x))*exp(-x/2.52328)-exp(-x^2/10.7134)*0.286525
f(hcal_only) = 1.0043+(0.171518+0.208563/sqrt(x))*exp(-x/22.1065)-exp(-x^2/5.41617)*0.956523
```

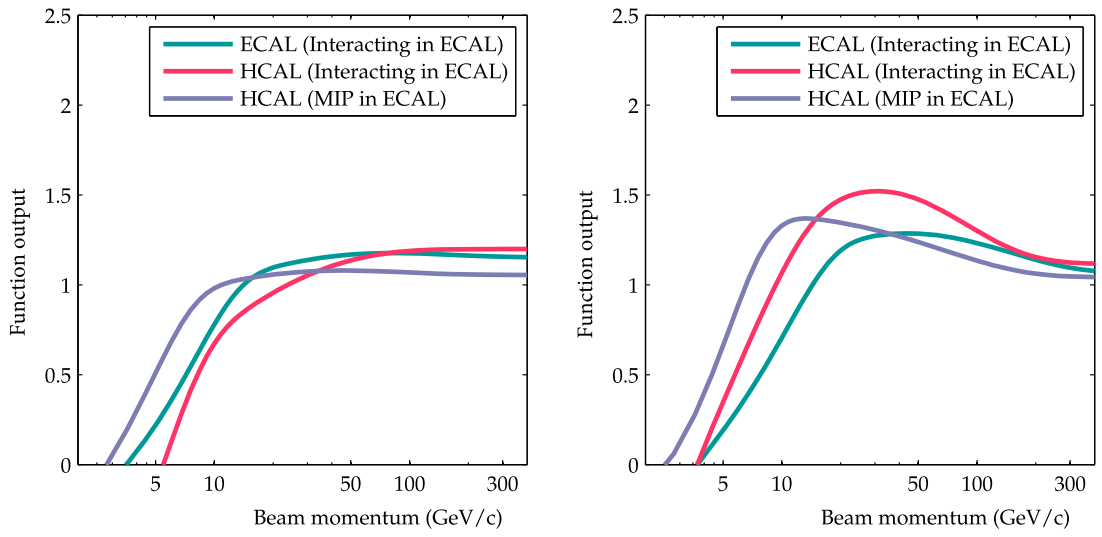
And the endcaps:

```
f(ecal) = 1.17625+(0.96815+0/sqrt(x))*exp(-x/137.084)-exp(-x^2/12.5064)*1.81195
f(hcal_int) = 0.880581+(0.5306+-1.49689/sqrt(x))*exp(-x/43.6451)-exp(-x^2/6.82207)*0.176657
f(hcal_only) = 0.844046+(2.42667+-5.48193/sqrt(x))*exp(-x/3.92954)-exp(-x^2/2474.56)*-0.151833
```

B.2.2 Standard fast simulation calibration functions

The standard fast simulation calibration functions used by particle flow look very different to the testbeam-based functions. Figure B.2 shows their shape in `CMSSW_3_3_6`. The thresholds are given in Table 4.2, page 90.

FIGURE B.2 **PF Candidate Calibration**
Calibration coefficients extracted from the fast simulation. The curves are defined by the function of Eq. 4.8.



(a) Barrel

(b) Endcap

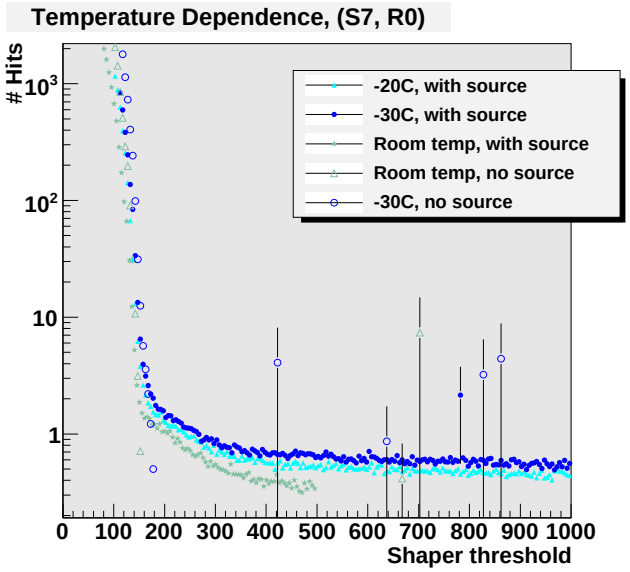
Appendix C

MAPS Appendix

C.1 Influence of temperature on prototype sensor noise

FIGURE C.1

A threshold scan
This typical threhold scan shows the number of hits, N, recorded (for a given number of bunch crossings) as a function of global threshold, G. The various series show the effect of the Strontium-90 source and temperature. Even at -30° C, noise hits were visible at high thresholds. Furthermore, the source spectrum did not stop at high thresholds. The reasons for this are explained in the main text of Chapter 7.



References

[1] N. Machiavelli, *The Prince*, (Modern translation by G. Bull from Penguin), first published 1532.

[2] D. A. Dicus and S. S. D. Willenbrock, Phys. Rev. D **32**, 1642 (1985).

[3] S. L. Glashow, Sci. Am. **246**, 42 (1982).

[4] D. J. Gross, Physics Today **40**, 39 (1987).

[5] A. Salam, Science **210**, 723 (1980).

[6] F. Englert and R. Brout, Phys. Rev. Lett **13**, 321 (1964).

[7] G. 't Hooft, Sci. Am. **243**, 90 (1980).

[8] S. Weinberg, Sci. Am. **231**, 50 (1974).

[9] Y. Fukada *et al.*, Phys. Rev. Lett. **81 (6)**, 1158 (1998).

[10] I. Newton, *Principia*, volume 3, 1687.

[11] A. Einstein, Sitzungsberichte der Preussischen Akademie der Wissenschaften zu Berlin *(Reports of the Prussian Academy of Sciences in Berlin)*, 844 (1915).

[12] F. Zwicky, Astrophysical Journal **86**, 217 (1937).

[13] P. J. E. Peebles and B. Ratra, Rev. Mod. Phys. **75**, 559 (2003).

[14] P. Higgs, Phys. Rev. **13**, 508 (1964).

[15] The LEP Electroweak Working Group, Precision electroweak results and constraints on the Standard Model, `http://lepewwg.web.cern.ch/LEPEWWG/`.

[16] The CDF Collaboration, the Dϕ Collaboration, the Tevatron New Physics, Higgs Working Group, *arXiv:0911.3930v1*, FERMILAB-CONF-09-557-E.

[17] M. Spira, Proceedings of the International Europhysics Conference in High Energy Physics, 1144 (1997), Jerusalem, Israel. Also under CERN-TH/97-323.

[18] J. Conway, Standard Model Higgs; Tevatron Run 2 Cross Sections and Branching Ratios, `http://www.physics.ucdavis.edu/~conway/research/higgs/smhiggs-tev.html`.

[19] A. Djouadi, J. Kalinowski, and M. Spira, Comp. Phys. Commun. **108 C**, 56 (1998), hep-ph/9704448.

[20] P. Fayet and S. Ferrara, Phys. Rep. **32**, 251 (1977).

[21] D. H. Perkins, *Introduction to High Energy Physics*, Cambridge, 4th edition, 2000.

[22] J. Ellis, Lectures presented at the European School of High-Energy Physics (1998), *arXiv:hep-ph/9812235v1*.

[23] G. 't Hooft *et al.*, *Recent Developments in Field Theories*, Plenum Press, NY, 1980.

[24] E. Witten, Nucl. Phys. Lett. B **B188**, 513 (1981).

[25] H. E. Haber and G. L. Kane, Phys. Rep. **117**, 75 (1985).

[26] M. Della Negra, A. Herve, L. Foa, and D. Acosta, CMS Physics TDR, Volume 1: Detector Performance and Software, Technical Report v8.1, CERN, 2006, CERN-LHCC-2006-001.

[27] CERN Communication Group, CERN FAQ: LHC - the guide, 2009, CERN-Brochure-2009-003-Eng from `http://cdsweb.cern.ch/record/1165534/files/CERN-Brochure-2009-003-Eng.pdf`.

[28] The ATLAS Collaboration, The ATLAS Detector and Physics Performance, Technical report, CERN, 1999, CERN-LHCC-99-14.

[29] The ALICE Collaboration, Technical report, CERN, 2001, CERN-LHCC-2001-021.

[30] The LHCb Collaboration, LHCb: Technical Proposal, Technical report, CERN, 1998, CERN-LHCC-98-004.

[31] The TTC Website, `http://ttc.web.cern.ch/TTC/intro.html`.

[32] D. Barney, 2004, CMS technical drawings available from the CERN document server.

[33] The CMS Collaboration, *The CMS Photobook*, CERN, 2009, Also available at `http://cms.web.cern.ch/cms/Media/Images/PhotoBook/index.html`.

[34] H. Breuker for the CMS Tracker Collaboration, The CMS tracker parameter booklet, 2004, CERN HB/PP 1-11-2004 - available from `https://edms.cern.ch/document/367936`.

[35] The CMS Tracker Detector Performance Group, 2010, Private communication via M. Della Negra.

[36] E. Yazgan *et al.* (for the CMS ECAL/HCAL Collaborations), Journal of Physics: Conference Series **160**, 012056 (8pp) (2009).

[37] The CERN Large Hadron Collider Computing Grid, `http://lcg.web.cern.ch/LCG/`.

[38] The ALEPH Collaboration, Nucl. Instr. and Meth. Phys. Res. **A360**, 481 (1995).

[39] A. Abulencia *et al.* (The CDF Collaboration), Phys. Rev. **D75**, 092004 (2007).

[40] C. Gallea, *Measurement of* $\sigma(pp \to Z + X)\mathrm{Br}(Z \to \tau^+\tau^-)$ *and search for Higgs bosons decaying to* $\tau^+\tau^-$ *at* $\sqrt{s} = 1.96$ *TeV*, PhD thesis, NIKHEF, 2008.

[41] The Particle Flow Physics Object Group, CMS Analysis Note (2009), AN-2009/039.

[42] T. S. Virdee, Calorimetry, St Croix NATO Summer School, Year unknown, `http://cmsdoc.cern.ch/~virdee/stcroix.all.pdf`.

[43] C. Amsler *et al.* (The Particle Data Group), Physics Letters **B667** (2008), `http://pdg.lbl.gov`.

[44] H. Bethe and J. Ashkin, *Experimental Nuclear Physics*, J. Wiley, New York, 1953.

[45] L. D. Landau, J. Exp. Phys. (USSR) **8**, 201 (1944).

[46] P. V. Vavilov, Sov. Phys. JETP **5**, 749 (1957).

[47] H. Bichsel, Rev. Mod. Phys. **60**, 663 (1988).

[48] G. Hall, Nucl. Instrum. and Meth. **220**, 356 (1984).

[49] P. H. Barrett *et al.*, Rev. Mod. Phys. **24**, 133 (1952).

[50] B. Rossi, *High Energy Particles*, Prentice-Hall, Inc., Englewood Cliffs, NJ, 1952.

[51] R. Wigmans, Ann. Rev. Nucl. And Part. Sci **41**, 133 (1991).

[52] The Molière radius, Formal origin unknown. See any standard text such as those already referenced, in particular [43].

[53] W. R. Nelson *et al.*, Phys. Rev. **149**, 201 (1966).

[54] G. Bathow *et al.*, Nucl. Phys. **B20**, 592 (1970).

[55] *On the difficulties of simulating hadronic interactions:* There is no reference in particular that states this explicitly, but anyone who works in the simulation of such affairs will certainly agree. See, for example `http://geant4.web.cern.ch/geant4/UserDocumentation/UsersGuides/PhysicsReferenceManual/fo/PhysicsReferenceManual.pdf`.

[56] E. Cheu, T. Embry, J. Rutherford, and R. Wigmans, editors, *Proc. Seventh Inter. Conf. on Calorimetry in High Energy Physics, Tucson, Arizona, 9 November 1997*, World Scientific, Singapore, 1998.

[57] The HELIOS Collaboration, T. Akesson *et al.*, Zeit. fur Physik C **38**, 383 (1988).

[58] The HELIOS Collaboration, T. Akesson *et al.*, Phys. Lett. B **214**, 295 (1988).

[59] R. Wigmans, Nucl. Instr. and Meth. **A265**, 273 (1988).

[60] U. Fano, Phys. Rev **72**, 26 (1947).

[61] R.C. Alig, S. Bloom, and C.W. Struck, Phys. Rev. B **22**, 5565 (1980).

[62] The ILD Concept Group, The International Large Detector, Technical report, 2008, See http://www.ilcild.org/documents/ild-letter-of-intent.

[63] S. Banerjee (for the CMS Collaboration), Readiness of CMS Simulation towards LHC Startup, 2007, Presented at *CHEP07, Victoria, Canada*, CMS CR-2008/003.

[64] The GEANT4 Collaboration, Nucl. Instrum. and Methods **506** (2003).

[65] T. Sjostrand, S. Mrenna, and P. Skands, *PYTHIA 6.4 Physics and Manual*.

[66] GEANT4 Physics Lists, See http://www.geant4.org/geant4/support/proc_mod_catalog/physics_lists/referencePL.shtml.

[67] D. Orbaker (for the CMS Collaboration), Fast Simulation of the CMS Detector, May, 2009, Presented at *Computing in High Energy Physics (CHEP)*. CMS CR-2009/074.

[68] The Super-LHC Upgrade website, 2010, See http://project-slhc.web.cern.ch/project-slhc/.

[69] A. Rose, *The Level-1 Trigger of the CMS Experiment at the LHC and the Super-LHC*, PhD thesis, Imperial College, London, 2010.

[70] M. Adams *et al.*, Nucl. Instr. and Meth. **A511**, 311 (2003).

[71] P. Moisenz, Private communication, 2009.

[72] J. Ballin, M. Della Negra, P. Moisenz, and S. Banerjee, Private communication. Petr Moisenz reporting on the HE π/e ratio determined from results (not yet published) in testbeam, 2010.

[73] ROOT: An object oriented toolkit for analysis, http://root.cern.ch.

[74] F. James, MINUIT: Function Minimization and Error Analysis, CERN Program Library Long Writeup D506. See http://wwwasdoc.web.cern.ch/wwwasdoc/minuit/minmain.html.

[75] S. Abdoulline, HCAL Signal Evaluation, 2009, Revision 38, https://twiki.cern.ch/twiki/bin/viewauth/CMS/HcalSignalEvaluation.

[76] D. Gerbaudo, HCAL Sampling Factors, 2008, Revision 5, https://twiki.cern.ch/twiki/bin/view/CMS/HcalSamplingFactors.

[77] M. Pioppi, CMS Internal Note (2007), CMS IN-2007/065.

[78] M. Mulders *et al.*, CMS Analysis Note (2008), AN-2008/098.

[79] C. Bernet, S. Gennai, P. Janot, and R. Cavanaugh, Pre-Approval of PAS PFT-09/001, 2009, http://indico.cern.ch/conferenceDisplay.py?confId=54860.

[80] *Calo Towers:* The standard 'Scheme B' of CMSSW v312 was used here, but more information can be found at `https://twiki.cern.ch/twiki/bin/viewauth/CMS/SWGuideCaloTowers`.

[81] The Particle Flow Group for the CMS Collaboration, CMS Analysis Note (2008), CMS AN-2008/043.

[82] V. Chetluru *et al.*, (2009), CMS-AN-2009/067.

[83] I. Nassi and B. Shneiderman, SIGPLAN Notices XII (1973), `http://www.cs.umd.edu/hcil/members/bshneiderman/nsd/`.

[84] The CMS Collaboration, CMS Physics Analysis Summary (2007), JME-07-002.

[85] CMS Jet Resolution Studies: results from Summer 08 Redigi QCD di-jet sample processed using CMSSW_2_2_3, From `http://home.fnal.gov/~vasurang/ResolutionStudies`.

[86] The CMS Collaboration, Alignment of the CMS Silicon Tracker during Commissioning with Cosmic Rays, Submitted to JINST, 2009.

[87] The CMS Collaboration, (2009), In preparation. CMS Paper QCD-09-010.

[88] The CMS Collaboration, JHEP (2010), *arXiv:1002.0621v2*.

[89] The CMS Collaboration, Particle flow reconstruction of 0.9 TeV and 2.36 TeV collision events in CMS, 2010, CMS PAS PFT-10-001.

[90] CMS HCAL noise (no zero suppression) measured from data with conditions STARTUP3X_V8D_vs_MC_31X_V9, 2010, `https://cms-cpt-software.web.cern.ch/cms-cpt-software/General/Validation/SVSuite/HCAL/STARTUP3X_V8D_vs_MC_31X_V9/Noise_NZS/HcalRecHits_Emean_seq_HB1.gif`.

[91] J. Whitmore, Nucl. Instrum. and Meth. A. (2009), `doi:10.1016/j.nima.2009.08.088`.

[92] J. Nachtman, HCAL Upgrade TP simulation status, 2010, See `http://indico.cern.ch/getFile.py/access?contribId=2&resId=1&materialId=slides&confId=77377`.

[93] E. Davolo and P. Naïm, *Neural Networks*, Macmillan Computer Science Series, Macmillan, 1991, Translated by A. Rawsthorne.

[94] W. McCulloch and W. Pitts, Bulletin of Mathematical Biophysics **5**, 115 (1943).

[95] D. O. Hebb, *The Organisation of Behaviour*, J. Wiley, New York, 1949.

[96] F. Rosenblatt, Psychological Review **65**, 386 (1956).

[97] M. Minksy and S. Papert, *Perceptrons: an Introduction to Computational Geometry*, MIT Press, 1969.

[98] J. J. Hopfield, Proceedings of the National Academy of Sciences, USA **79**, 2554 (1982).

[99] B. Widrow and M. E. Hoff, IRE WESCON Conv Record, Part 4, 96 (1960).

[100] Y. Le Cun, Proceedings of Cognitiva 85, 599 (1985).

[101] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, Nature **323**, 533 (1986).

[102] C. Broyden, Journal of the Institute of Mathematics and Its Applications **6**, 76 (1970).

[103] R. Fletcher, Computer Journal **13**, 317 (1970).

[104] D. Goldfarb, Mathematics of Computation **24**, 23 (1970).

[105] D. F. Shanno, Mathematics of Computation **24**, 647 (1970).

[106] J. Armstrong, *Making reliable distributed systems in the presence of software errors*, PhD thesis, The Royal Institute of Technology, Stockholm, 2003, Available from `http://www.sics.se/~joe/thesis/armstrong_thesis_2003.pdf`.

[107] J. Armstrong, *Programming Erlang*, The Pragmatic Bookshelf, 2007.

[108] Intel Corp., Intel Microprocessor Quick Reference Guide, 2010, `http://www.intel.com/pressroom/kits/quickreffam.htm`.

[109] A Picture of the Multicore Crisis, 2007, `http://smoothspan.wordpress.com/2007/09/06/a-picture-of-the-multicore-crisis/`.

[110] J. Dean and S. Ghemawat, OSDI'04: Sixth Symposium on Operating System Design and Implementation (2004), Available from `http://labs.google.com/papers/mapreduce.html`.

[111] Yahoo! Corp., The! Yahoo! Home! Page!, See `http://info.yahoo.com/center/us/yahoo/`.

[112] The Apache Software Foundation, The Hadoop Project, See `http://hadoop.apache.org/`.

[113] M. Stein, I. Melzer-Pellmann, V. Andreev, and K. Borras, Energy Weighting with a longitudinally segmented HCAL, 2009, SLHC Upgrade meeting. See `http://indico.cern.ch/getFile.py/access?contribId=125&sessionId=13&resId=2&materialId=slides&confId=59696`.

[114] A. Hoecker *et al.*, TMVA: A Toolkit for Multivariate Analysis, v4, See `http://tmva.sourceforge.net`.

[115] The Mathworks Corporation, MATLAB v7.9, See `http://www.mathworks.com/`.

[116] Canonical Ltd., Ubuntu 9.04, See `http://www.ubuntu.com/`.

[117] cprof: An Erlang Call Count Profiling Tool, See `http://ftp.sunet.se/pub/lang/erlang/doc/man/cprof.html`.

[118] CentOS: The Community Enterprise Operating System, See `http://www.centos.org/`.

[119] The CLIC Collaboration, CLIC Study, Technical report, See http://clic-study.web.cern.ch/clic-study.

[120] M. Battaglia, *arXiv:0705:3997v1* (2007).

[121] The ILC Collaboration, Reference Design Report, Technical report, 2007, http://www.linearcollider.org/about/Publications/Reference-Design-Report.

[122] The ILC Collaboration, Linear Collider Website, See http://www.linearcollider.org.

[123] G. Blair, ILC – The Machine, 2007, A talk presented at the IOP Meeting, Oxford. Subsequently no longer available (2010).

[124] M. Moretti, S. Moretti, F. Piccinini, R. Pittau, and A. D. Polosa, JHEP02, 24 (2005), See http://iopscience.iop.org/1126-6708/2005/02/024/.

[125] The ALEPH Collaboration, Phys. Lett. B **422**, 384 (1998).

[126] C. Quigg, *Gauge Theories of the Strong, Weak and Electromagnetic Interactions*, The Benjamin/Cummings Publishing Company, Inc., 1983.

[127] M. Thomson, Particle Flow and Calorimetry at the ILC, 2006, Presented at the European ILC Software and Physics Meeting, Cambridge. See http://ilcagenda.linearcollider.org/materialDisplay.py?contribId=23\&sessionId=8\&materialId=slides\&confId=165.

[128] P. Dauncey, Concept for a Si-W ECAL for the ILC based on MAPS (draft 0.2), Technical report, 2004, Available from http://www.hep.ph.ic.ac.uk/calice/maps/mapscd/mapscd.pdf.

[129] The LDC Concept Group, Sketch for a Large Detector Concept, Technical report, 2005, Subsequently subsumed into ILD activities. See http://www.ilcldc.org; report ostensibly no longer available (2010). The TESLA Project and its report drew heavily from it, and the website is still available at http://tesla.desy.de/new_pages/TDR_CD/start.html.

[130] K. Stefanov, MAPS-based ECAL Option for ILC, 2006, Talk presented at ECFA 2006, Valencia, Spain. Available at http://www.hep.ph.ic.ac.uk/calice/generalMeetings/061103manchester/stefanov.pdf.

[131] Synopsys, Inc., Sentaurus TCAD, See http://www.synopsys.com/Tools/TCAD/Pages/default.aspx.

[132] E. Yazgan, Private communication, 2008.

Colophon

This thesis was typeset in LaTeX. The copy font is Adobe's Minion Pro. The more pleasant plots and figures were produced in MATLAB, the more unsightly ones were produced with ROOT.