

Development of a novel alignment system for
the ATLAS Inner Detector and an
investigation of the effect of alignment
inaccuracies on tracker performance

Adrian Francis Fox-Murphy

Wolfson College,
University of Oxford



Thesis submitted in partial fulfilment
of the requirements for the degree of
Doctor of Philosophy
University of Oxford, Trinity Term 1996

Development of a novel alignment system for the ATLAS Inner Detector and an investigation of the effect of alignment inaccuracies on tracker performance

Adrian Francis Fox-Murphy

Wolfson College, University of Oxford

Thesis submitted in partial fulfilment of the
requirements for the degree of Doctor of Philosophy.

University of Oxford, Trinity Term 1996

Abstract

The Large Hadron Collider at CERN will offer an unparalleled opportunity to probe fundamental physics at an energy scale well beyond that reached by current experiments. The ATLAS detector is being designed to fully exploit the potential of the LHC for revealing new aspects of the fundamental structure of nature.

In order to meet the stringent tracking requirements of ATLAS, it will be necessary to determine the positions of over 100 million tracking elements to very high precision during operation of the detector. The principles of the alignment and survey techniques used to do this are introduced and the current activities concerning the development of an alignment strategy for the ATLAS Inner Detector are presented. After consideration of the motivation and requirements, descriptions of several of the candidate technologies are given, together with explanations of how they might be applied in the various stages of the alignment process.

A fast remote measurement system known as Frequency Scanned Interferometry (FSI) which is capable of making precise measurements of absolute lengths has been developed. This novel technique is likely to be used as the basis of a run-time survey system for the ATLAS Inner Detector. The basic principles are explained and a detailed design and laboratory test results are presented.

An element common to all types of survey system is the need to combine a number of measurements to form a three-dimensional picture of the positions of all the detector elements. An introduction to measurement combination using geodetic networks is given, and the results of a study of networks suitable for use with the FSI measurement technique are presented.

As part of the process of deriving a detailed set of requirements for the survey system, a full Monte Carlo simulation study has been performed to investigate how the Inner Detector track fitting resolutions vary as a function of the alignment precisions of the SCT barrel.

A physical signal which is important for defining the required momentum resolution is the forward-backward asymmetry of a heavy analogue of the charged electroweak gauge boson. The sensitivity to this signal is investigated as a function of alignment precision.

Finally, work done during the development of the current layout of the ATLAS Inner Detector is presented.

To my parents.

© Adrian Francis Fox-Murphy, 1996

This thesis was typeset in Times Roman using \LaTeX on a Digital AlphaStation.

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording, or otherwise, without the express permission of the author.

Published in the United Kingdom.

Acknowledgements

It is a pleasure to be able to thank the many people who have helped make the last four years so enjoyable and rewarding. First and foremost, I would like to thank Richard Nickerson, my supervisor, for guiding my studies with precision, enthusiasm, good humour and endless patience. He has been an enormous source of encouragement and inspiration, and is largely responsible for teaching me how to do research. Heartfelt thanks must also go to Tony Weidberg, my group leader, whose clear and precise comments and explanations over the years have been very much appreciated. Special thanks must go to David Howell, without whom I would have achieved only a small fraction of what I have, and with whom I spent many interesting moments discussing everything from optics to sailing and the rock music business.

All the other members of the Oxford ATLAS group deserve recognition for their help and advice on the countless occasions when I have needed assistance. In particular I would like to thank Jim Loken, Alan Holmes, Roy Wastie, Barney Brooks and Brian Hawes. Thanks too to my fellow ATLAS students Richard Hawkings, Armin Reichold, Anish Grewal, Stephen Hunt and Daniel Buira-Clark for making the office an interesting, lively and enjoyable place to work.

Many members of the ATLAS collaboration deserve recognition: Kees Daum for kicking off the alignment ball; Andy Parker and Mike Tyndell for sharing their considerable experience in many useful discussions over the last few years; Geoff Tappern for explaining some of the intricacies of the Inner Detector mechanics; and Daniel Froidevaux and Stephen Haywood for many interesting conversations about tracking detectors and simulations. Thanks must go also to Stephen Haywood and Bill Murray for their comments on the final draft.

To the other students from around the lab (especially those of the ZEUS and DELPHI groups), and to my friends at Wolfson College, thanks for making the last few years so much fun. If I started naming names I'm sure I would forget someone, but you know who you are, and you know where I will be when I have finished writing these acknowledgements. See you there.

I would like to acknowledge the subsistence and travel grants provided by the Particle Physics and Astronomy Research Council.

Last, but far from least, I would like to thank my parents and siblings for far more than I could possibly describe here.

Foreword

Early in the next century, the Large Hadron Collider will start to collide protons at an energy of 14 TeV, so beginning a new era of investigation of the structure of nature at an energy scale well beyond that presently attainable. The physical motivation for the LHC and the designs of the LHC and the ATLAS detector are introduced in Chapter 1.

In order to meet the stringent tracking requirements of ATLAS, it will be necessary to determine the positions of each tracking element in the Inner Detector to very high precision. Chapter 2 serves as an introduction to alignment and survey systems and to the current activities concerning the development of the Inner Detector alignment strategy. After consideration of the motivation and requirements, descriptions of several of the candidate technologies are given, together with explanations of how they might be applied in the various stages of the alignment. This is followed by a brief outline of an overall alignment strategy for the ATLAS Inner Detector.

The measurement technique likely to be used as the basis of the Inner Detector survey system is known as Frequency Scanned Interferometry (FSI). The basic principles, the detailed design and laboratory test results are presented in Chapter 3.

An element which is common to all types of survey system is the need to combine of a number of individual measurements to form an overall three-dimensional picture of the positions of detector elements. An introduction to measurement combination based on geodetic networks and a study of networks suitable for use with the FSI measurement technique are given in Chapter 4.

As part of the process of deriving a detailed set of requirements for the survey system, a full Monte Carlo simulation study has been performed to investigate how the Inner Detector track fitting resolutions vary as a function of the Inner Detector alignment precisions. The results are presented in the first half of Chapter 5.

An important signal for defining the required momentum resolution is the forward-backward asymmetry of a heavy analogue of the charged electroweak gauge boson. The sensitivity to this signal is investigated as a function of alignment precision in the latter part of Chapter 5.

In Chapter 6, some work done during the development of the current layout of the Inner Detector is presented.

Finally, in Chapter 7, the main results are summarised and some conclusions are drawn.

Contents

Abstract	i
Acknowledgements	iii
Foreword	iv
Contents	iv
1 Introduction	1
1.1 Physical context	1
1.1.1 The Standard Model	1
1.1.2 Inadequacies of the Standard Model	4
1.1.3 New directions	5
1.1.4 The significance of the TeV scale	6
1.2 The Large Hadron Collider	8
1.3 The ATLAS experiment	12
1.3.1 Basic design considerations	12
1.3.2 Overall detector concept	13
1.3.3 Inner Detector	17
1.3.4 Calorimetry	20
1.3.4.1 Liquid Ionisation Calorimetry	20
1.3.4.2 Hadronic Scintillator Calorimeter	22
1.3.5 Muon Spectrometer	25
1.3.5.1 Precision chambers	25
1.3.5.2 Trigger system	26
1.3.6 Trigger, DAQ and computing	27
1.3.6.1 Requirements	27
1.3.6.2 Architecture	27

1.3.6.3	Offline computing	27
1.3.7	Physics at ATLAS	29
1.3.7.1	Higgs search	29
1.3.7.2	Top-quark physics	30
1.3.7.3	B-physics	31
1.3.7.4	Supersymmetric particles	31
1.3.7.5	Other searches	32
1.4	Conclusion	32
	References	33
2	Alignment of the ATLAS Inner Detector	35
2.1	Introduction	35
2.2	Inner Detector alignment requirements	36
2.2.1	Physics requirements	36
2.2.2	Engineering requirements	37
2.2.3	Alignment requirements for the Level 2 trigger	37
2.3	Alignment techniques and technologies	39
2.3.1	Alignment using charged particle tracks	39
2.3.2	Optical metrology	40
2.3.2.1	Frequency-Scanned Interferometry (FSI)	40
2.3.2.2	Electronic Speckle Pattern Interferometry (ESPI)	41
2.3.2.3	X-ray alignment	44
2.3.2.4	RASNIK	44
2.3.2.5	MPI multipoint laser straightness monitor	46
2.4	Overview of the Inner Detector alignment process	48
2.4.1	Design-phase metrology	48
2.4.2	Module prefabrication / preassembly	48
2.4.3	Surface assembly	48
2.4.4	Installation	49
2.4.5	Running	49
2.5	Conclusions	49
	References	50
3	Frequency-Scanned Interferometry	51
3.1	Introduction	51

3.2	Frequency Scanned Interferometry	55
3.2.1	Introduction	55
3.2.2	Interferometer design	57
3.2.3	Refractive index	60
3.2.4	The frequency scan	62
3.2.5	Tunable laser selection	66
3.2.6	Optical frequency measurement	66
3.2.7	Retroreflectors and jewels	69
3.3	Vibration rejection	73
3.3.1	The phase-shift technique	73
3.3.2	Fringe analysis	75
3.3.3	Numerical example	77
3.4	Reference interferometer with direct phase detection	79
3.5	Laboratory demonstration system	81
3.5.1	Introduction	81
3.5.2	Remote interferometer	81
3.5.3	Semiconductor diode laser	84
3.5.4	Optical frequency measurement	85
3.5.5	Photodetection	86
3.5.6	Data acquisition and scan control	87
3.5.7	Results	87
3.6	Conclusions	91
3.A	Optical path calculations	93
3.B	Detecting the interference signal	98
3.C	Single mode fibres	99
	References	101
4	Geodetic networks	109
4.1	Introduction	109
4.2	Mathematical geodesy	110
4.2.1	Analysis phases	110
4.2.2	Model description notation	111
4.2.3	Adjustment	112
4.2.3.1	Unconstrained adjustment	115

4.2.3.2	Adjustment with constraints	115
4.2.3.3	Explicit observation equations	116
4.2.4	Quality testing	118
4.2.5	Basis-independent precision estimates	119
4.2.6	Combination of networks	120
4.3	SCAN-3	121
4.4	FSI network design	122
4.4.1	Precision	122
4.4.2	Mechanical constraints	122
4.4.3	Reliability requirements	123
4.5	Calculation results	124
4.6	Conclusions	131
	References	132
5	Track fitting and physics performance vs. alignment	133
5.1	Software	133
5.1.1	ATLAS Monte Carlo environment	133
5.1.2	iPatRec	134
5.2	Track fitting resolutions and alignment	136
5.2.1	Introduction	136
5.2.2	Technique	136
5.2.3	Simulation results	137
5.2.3.1	R-alignment	139
5.2.3.2	Φ -alignment	140
5.2.3.3	Z-alignment	141
5.2.4	Discussion	144
5.2.5	Conclusions	150
5.3	W' asymmetry signal vs. alignment	152
5.3.1	New gauge bosons	152
5.3.1.1	Models	152
5.3.1.2	Present experimental limits	154
5.3.2	Event generation software	156
5.3.3	W' signals at ATLAS	157
5.3.3.1	Discovery potential	157

5.3.3.2	Mass measurement precision	158
5.3.4	Forward-backward asymmetry of $W' \rightarrow e\nu$	160
5.3.4.1	Introduction	160
5.3.4.2	Reconstruction	161
5.3.4.3	Charge sign identification	163
5.3.5	Simulation results	166
5.3.5.1	W' production and decay	167
5.3.5.2	Lepton transverse momentum distributions	167
5.3.5.3	Feynman-x reconstruction	170
5.3.5.4	Forward-backward asymmetries	172
5.3.5.5	Momentum resolution and CSI vs. Φ -alignment	173
5.3.5.6	Asymmetry and integrated asymmetry vs. Φ -alignment	174
5.3.5.7	Particle-level integrated A_{FB} vs. momentum resolution	177
5.4	Conclusions	179
	References	180
6	Tracker layout development	186
6.1	Introduction	186
6.2	Software	186
6.3	Layout design work	187
6.3.1	The Cosener's House layout	189
6.3.2	A Panel-style layout	191
6.3.2.1	Barrel layout	191
6.3.2.2	Forward layout	193
6.3.2.3	Resolutions	197
6.4	Conclusions	201
	References	203
7	Conclusions	206
	Index	209

Chapter 1

Introduction

1.1 Physical context

1.1.1 The Standard Model

The experimental discoveries and theoretical developments of the last 50 years have led to the development of the Standard Model (SM) of particle physics. Since it is described in great detail in many textbooks [13]–[16], the discussion given here will be brief.

The uniting principle of modern fundamental physics is that of symmetry. In addition to the familiar space-time symmetries and discrete symmetries, it appears that Nature exhibits local gauge symmetries. There exists a deep connection between local gauge symmetries and physical interactions : the phase variation permitted by a local symmetry is reconciled by a field which is considered to be the mediator of interactions between matter particles. The Standard Model is a gauge field theory with local gauge invariance based on the group $SU(3)_C \times SU(2)_L \times U(1)_Y$. It describes all of the known fundamental particles and their interactions via three of the four known forces of nature : electromagnetism, the weak force and the strong force. Gravity is by far the weakest of the four forces, and is not described in the SM.

There are three types of particle described in the Standard Model : spin- $\frac{1}{2}$ fermions (matter particles), spin-1 gauge bosons (which mediate interactions between matter particles), and a spin-0 Higgs boson (which is a consequence of spontaneous symmetry breaking). The up and down quarks, the electron, and the electron neutrino are all that is needed to make up normal matter. For reasons unknown, this pattern is repeated three times, that is, for three ‘generations’, each generation containing two quarks, a lepton and a neutrino. See Table 1.1.

gen 1	gen 2	gen 3	T	T^3	Y	Q
$\begin{pmatrix} \nu_e \\ e \end{pmatrix}_L$	$\begin{pmatrix} \nu_\mu \\ \mu \end{pmatrix}_L$	$\begin{pmatrix} \nu_\tau \\ \tau \end{pmatrix}_L$	$+1/2$	$+1/2$	-1	0
e_R	μ_R	τ_R	0	0	-2	-1
$\begin{pmatrix} u \\ d' \end{pmatrix}_L$	$\begin{pmatrix} c \\ s' \end{pmatrix}_L$	$\begin{pmatrix} t \\ b' \end{pmatrix}_L$	$+1/2$	$+1/2$	$+1/3$	$+2/3$
u_R	c_R	t_R	0	0	$+4/3$	$+2/3$
d'_R	s'_R	b'_R	0	0	$-2/3$	$-1/3$

Table 1.1: *Fermion multiplet assignments and quantum numbers T (weak isospin), T^3 (3rd component of weak isospin), Y (hypercharge) and Q (electric charge).*

The electromagnetic force is described with extremely good agreement with experiment by quantum electrodynamics (QED), a gauge field theory with local $U(1)$ gauge invariance. It describes the interaction between particles carrying electric charge via a massless spin-1 gauge boson — the photon.

The electromagnetic and weak forces are described in the Standard Model in the ‘unified’ framework of the $SU(2)_L \otimes U(1)_Y$ Glashow Salam Weinberg (GSW) electroweak model. Experimental evidence suggested that the weak force should be mediated by three massive vector bosons W^\pm and Z^0 (to explain its short range) with a $V - A$ (vector minus axial vector) interaction structure (to explain the observed parity violation among other things). One of the major stumbling blocks in the development of a field theory description of the weak force was discovering a way of permitting the gauge bosons to be massive whilst maintaining the renormalisability essential for a predictive theory. The only known solution to this problem is the so-called Higgs mechanism. In certain field theories, when the ground state does not possess the full symmetry of the Lagrangian, there inevitably exist a number of scalar ‘Goldstone’ bosons. When this happens in a local gauge theory however, the would-be Goldstone bosons are manifested as the longitudinal degrees of freedom of the gauge bosons of the theory, which thereafter behave as massive vector bosons with three spin components. In the Standard Model, four independent scalar fields arranged as a complex doublet are postulated. The Goldstone bosons arising from the non-zero vacuum expectation value of this doublet are manifested as longitudinal degrees of freedom of the three previously massless electroweak

gauge bosons W^\pm and Z^0 . The photon by contrast is chosen to remain massless, to agree with empirical evidence. There remains one massive scalar particle, the Higgs boson, whose mass is not specified in the theory. The quarks and leptons gain mass (whilst maintaining gauge invariance) through Yukawa couplings to the Higgs field.

Perhaps as much can be learned from broken symmetries as from those which are upheld. As mentioned above, the weak force violates parity conservation. The observed particle spectrum is not left-right symmetric — only left handed neutrinos (and right handed anti-neutrinos) have been detected. This asymmetry is reflected in the SM multiplet assignment. Another precious symmetry, charge conjugation followed by parity (CP), has been observed to be violated in the kaon system. CP violation can be parametrised (though not explained) by means of a phase in the Cabbibo-Kobayashi-Maskawa (CKM) matrix which is required to describe the coupling of the W to linear combinations of quarks.

The strong force is described in the SM by quantum chromodynamics (QCD), a gauge field theory with unbroken local $SU(3)$ symmetry. There are eight gauge bosons (gluons) which mediate the interactions of particles carrying a quantum number known as colour. The quarks have three possible colour states and the gluons have eight. The fact that the gauge bosons themselves are coloured results in three-gluon and four-gluon vertices which have highly non-trivial consequences for the dynamics of the theory. At high energy scales (and hence smaller distances), the strong force becomes weaker ('asymptotically free'). Conversely, at lower energies (and hence longer length scales), the force between particles becomes stronger, which may explain the fact that coloured objects have not been observed in isolation. It is suspected that this so-called 'colour confinement' is exhibited by QCD, although a proof remains elusive. Calculations involving the strong force are complicated by the large value of the coupling constant, which makes the application of perturbation theory problematic.

At the time of writing, no experimental result has been conclusively shown to disagree with the predictions of the Standard Model. The most significant vindication of the approach taken in the SM is surely the discovery in 1984 of the W^\pm and Z^0 at the expected masses. The top quark was discovered in 1995 at Fermilab, leaving the Higgs boson as the only remaining particle of the SM which has not been found. Resolution of the puzzle of electroweak symmetry breaking is currently one of the most burning issues in High-Energy Physics, and is one of the main reasons for constructing the Large Hadron Collider.

1.1.2 Inadequacies of the Standard Model

The remarkable agreement between theory and experiment notwithstanding, the Standard Model has a number of deficiencies and unexplained features. These indicate that the SM cannot be a complete theory of fundamental particle physics.

The electroweak sector:

The most serious structural problem is associated with the Higgs sector of the electroweak theory. This sector is responsible for the most noticeable feature of electroweak symmetry, namely that it is broken. What is the mechanism of electroweak symmetry breaking ?

19 free parameters:

There are 19 ‘free’ parameters in the SM which are not specified by the theory, and which must therefore be put in by hand. This is at odds with our prejudice, fostered by a history of repeated simplifications, that the world should be comprehensible in terms of a few simple laws. A more fundamental theory would be expected to specify their values.

Further unification:

In the SM, the strong and electroweak interactions are not unified. Gravity must also be incorporated before we can truly claim to understand all fundamental interactions. String theories are perhaps the precursor to a theory uniting gravity and the gauge theories. However, their logical foundations remain, for the moment, clouded in mystery.

Number of generations:

For an unknown reason, nature repeats herself (at least) three times. Why ?

Origin of P violation:

Why is the SM multiplet assignment of particles not left-right symmetric ? How is the left-right symmetry broken ? Are there right-handed neutrinos ?

Origin of CP violation:

The SM has the ‘facility’ to incorporate CP violation phenomenologically by having a non-zero phase in the CKM matrix. Even if this is the origin of the CP violation observed in the kaon system, the Standard Model gives no explanation of the values of the CKM matrix elements. An explanation would be expected from a more fundamental theory.

Strong CP problem:

θ_{QCD} must be very small to be compatible with the experimentally measured neutron electric dipole moment. Why is this 'free' parameter forced to be so small ?

Solar neutrino problem:

The observed number of neutrinos reaching the earth from the sun is significantly less than that predicted by models of the interior of the sun. Neutrino mixing may solve this problem, but this requires neutrinos to have a non-zero mass. Are neutrinos Majorana or Dirac particles ?

Baryon asymmetry of the universe:

The universe appears to be made of matter rather than antimatter. If equal amounts of both type of matter were created at the Big Bang, then the present asymmetry implies the existence of a baryon number violating process.

Dark matter:

From observations of rotational velocity profiles within galaxies, astronomers have shown that the amount of radiative matter accounts for perhaps only 10% of the total mass of the universe.

The above list may be long, but the fact that we can pose such wide-ranging and fundamental questions about nature and expect someday to find at least some of the answers is at least encouraging.

1.1.3 New directions

In their quest for the final theory, theorists have explored many different models. The element common to all these attempts is unification.

Unification :

Grand Unified Theories share some of the problems of the Standard Model (i.e. too many arbitrary parameters, no interaction with gravity). However, they are genuine unified field theories because they have only one gauge group (such as SU(5) or SO(10)) and hence only one coupling constant. Furthermore they make the prediction that the proton will decay. The existence of a desert of 12 orders of magnitude in energy containing no new physics is one of the main criticisms of the theory, but the proton decay prediction holds out some hope of testability.

being strictly valid only in the MSM (one Higgs doublet), and the upper bound being fairly model independent. If the M_H exceeds 1 TeV, weak interactions must become strong on the 1 TeV scale. This is perhaps the most compelling argument that there should be new physics at or before the energy scale of $\mathcal{O}(1 \text{ TeV})$.

- If M_H is greater than 1 TeV the tree level amplitude for $W^+W^- \rightarrow W^+W^-$ scattering diverges, leading to strongly interacting W and Z bosons.
- The coupling constant is not asymptotically free — at some energy Λ there is a Landau pole where the coupling constant becomes infinite. The value of Λ decreases as M_H increases and for the theory to make sense, Λ must be greater than M_H . This imposes an upper limit of $\sim 1 \text{ TeV}$ on M_H .
- It is possible to create construct a low-energy effective theory which has no Higgs particle but has instead symmetry-breaking mass terms in the effective Lagrangian. The cutoff scale for such an effective theory takes the place of the Higgs in the calculation of radiative corrections, and so LEP limits on the Higgs mass translate to limits on the cutoff scale. The available data points to a cutoff scale below 1 TeV.
- Supersymmetric and other theories predict particles with masses below the 1 TeV scale.
- Whatever the mechanism for solving the naturalness problem, it must involve new physics at or below the scale of 1 TeV.

Although theoretical speculation is valuable and necessary, it is unlikely that we will advance without new observations. The experimental results needed to answer our most fundamental questions will come from a number of sources, the most important of which is likely to be the next high-energy, high-luminosity hadron collider, namely the Large Hadron Collider.

One of the theoretical problems facing GUT theories is the ‘hierarchy’ or ‘naturalness’ problem : renormalisation effects mix the two mass scales (M_W and M_X) of the theory, destroying the hierarchy. Even if the theory is fine-tuned to one part in 10^{12} the scales still mix, ruining the separation. It turns out that an infinite number of fine-tunings is necessary at each level of perturbation, which is clearly undesirable.

Supersymmetry :

One appealing solution to the hierarchy problem is to include both local and global supersymmetry. There are powerful non-renormalisation theorems in supersymmetric theories which show that higher order interactions do not renormalise the mass scale, so removing the need for fine-tunings at every order of perturbation theory. One fine-tuning at the beginning is enough.

One of the main problems in building unified field theories is the inability to find a gauge group which can combine the particle spectrum and quantum gravity. The problem comes from the so-called ‘no-go theorem’ (actually a set of theorems of which the Coleman-Mandula theorem is the most powerful), which states that a group which non-trivially combines both the Lorentz group and a compact Lie group cannot have finite-dimensional unitary representations. When this theorem was discovered it was thought that the implication was that any attempt to build a ‘master group’ combining both gravity and the particle spectrum must be doomed to failure. However, there does exist a way to evade the Coleman-Mandula theorem which takes advantage of the fact that the theorem breaks down if non-commuting (or Grassman) numbers are permitted in the equations.

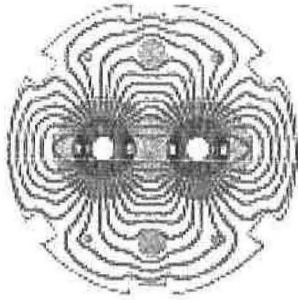
Not only does supersymmetry give us a plausible solution to the hierarchy problem, it also gives us theories of gravity in which the divergences are partially or even completely cancelled.

1.1.4 The significance of the TeV scale

There are a number of theoretical reasons why it may be expected that new effects will be seen below the energy scale of $\mathcal{O}(1 \text{ TeV})$:

- In the SM, interactions of the Higgs are not prescribed in the same manner as those for the intermediate gauge bosons. The mass of the Higgs is constrained only to lie within the range 7 GeV (vacuum stability) to 1 TeV, the lower bound

1.2 The Large Hadron Collider



In December 1994, delegates representing the 19 member states of CERN unanimously approved a resolution permitting the construction of the 14 TeV Large Hadron Collider (LHC). The LHC, a high-luminosity proton-proton collider constructed from more than one thousand superconducting magnets, will be installed in CERN's existing 27-kilometre circular tunnel built for the Large Electron-Positron (LEP) collider. It will recreate conditions believed to have prevailed in the universe just 10^{-12} seconds after the Big Bang. In doing so the LHC will launch High Energy Physics into the next millennium promising to deepen our understanding of nature.

The construction schedule will depend upon the available financial resources. At the time of the formal approval of the LHC, no non-member states had committed to finance the project. Assuming no financial contribution from non-member states, the machine would be brought into operation in two stages: first with two-thirds of the magnets (and a c.m. energy of 10 TeV) in 2004, and then completed to full energy (14 TeV) in 2008. It is currently expected that financial participation from non-member states will permit construction of the 14 TeV machine in one stage, with the first physics run in 2005. The construction schedule will be reviewed in 1997.

Two general-purpose experiments, ATLAS [3][6] and CMS [5][6], are being designed for the LHC. Both were approved by CERN in January 1996. A dedicated B-physics experiment, LHC-B [7], is planned. The LHC machine may also be used to provide heavy ion (Pb-Pb) collisions at 5.4 TeV per nucleon pair with a luminosity of up to $10^{27} \text{ cm}^{-2} \text{ s}^{-1}$ using the existing CERN ion facility. For this a single detector, ALICE [8], is envisioned.

In September 1995, following several years of extremely successful operation at the Z^0 mass, the LEP beam energy was increased by inserting a number of superconducting cavitites into the ring. The CM energy will be increased further during 1996, ramping up to around twice its former energy. It is expected that LEP2, as it is now known, will be operated until 2000, when it will be removed from the tunnel to make way for the installation of the LHC. For the first three years following commissioning, the LHC will operate at "low" luminosity ($10^{33} \text{ cm}^{-2} \text{ s}^{-1}$). Following this, the machine will run at a higher luminosity ($10^{34} \text{ cm}^{-2} \text{ s}^{-1}$) for an expected period of at least 10 years.

If physically justified on the basis of new discoveries, the LEP machine may be later reinstalled in the tunnel on top of the LHC in order to provide electron-proton collisions at unprecedented energy.

Considerable technical innovation is needed to fit the two rings of the LHC into the tunnel cross section whilst leaving enough space for an eventual lepton ring. In order to achieve the design energy within the constraint of the 27 km circumference LEP tunnel, the magnet system must operate in superfluid helium at below 2 K. In addition, space and cost limitations have resulted in a two-in-one magnet design, in which the two rings are incorporated into the same cryostat. Some of the machine parameters are listed in Table 5.10.

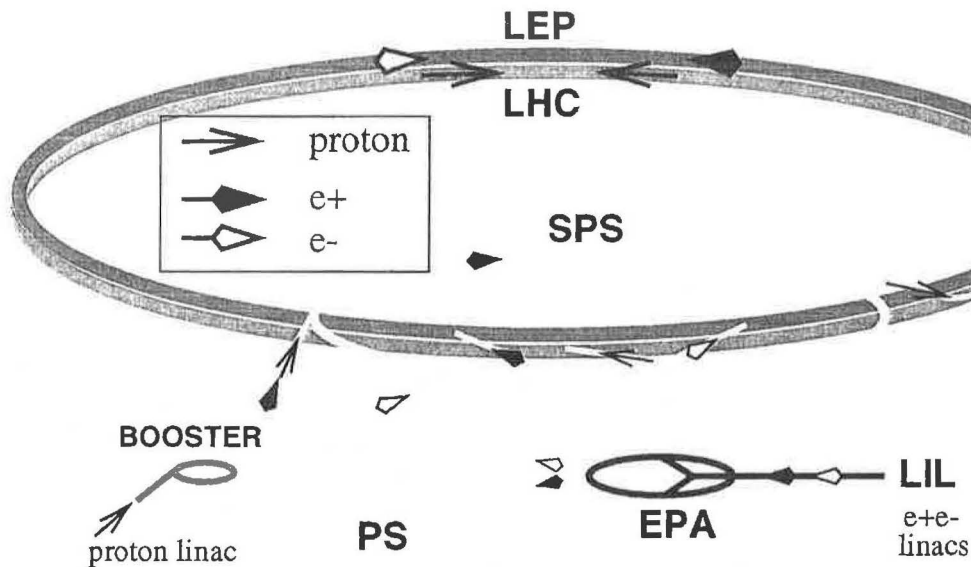


Figure 1.1: *The Large Hadron Collider.*

The LHC will be filled with protons delivered from the SPS and its pre-accelerators at 450 GeV. Two superconducting magnetic channels will accelerate the protons to 7-on-7 TeV, after which the beams will counter-rotate for several hours, colliding at the experiments, until the beam degradation is such that the machine has to be emptied and refilled. The magnetic channels will be housed in the same yoke and cryostat, a unique configuration that not only saves space but also gives a 25% cost saving over separate rings.

LHC machine parameters (v4.1 22/03/95)	
Ring circumference	26658.883 m
Proton energy	7000 GeV
Stored energy per beam	334 MJ
Expected luminosity (\mathcal{L})	$1.00 \times 10^{34} \text{ cm}^{-2} \text{ s}^{-1}$
Luminosity lifetime	10 hours
Average no. interactions per crossing	18.84
Bunch collision frequency	40 MHz
Number of bunches (k_b)	2835
Number of protons per bunch	1.05×10^{11}
Bunch separation	7.49 m (25 ns)
R.M.S. (x,y) beam size at IP	$15.9 \mu\text{m}$
R.M.S. bunch length	7.7 cm (0.257 ns)
Crossing angle at IP (Φ)	$200 \mu\text{rad}$
Depth of focus at IP (β^*)	0.5 m
Number of main bends	1232
Main bend magnetic field	8.386 T
Main bend length	14.200 m

Table 1.2: *LHC machine parameters.*

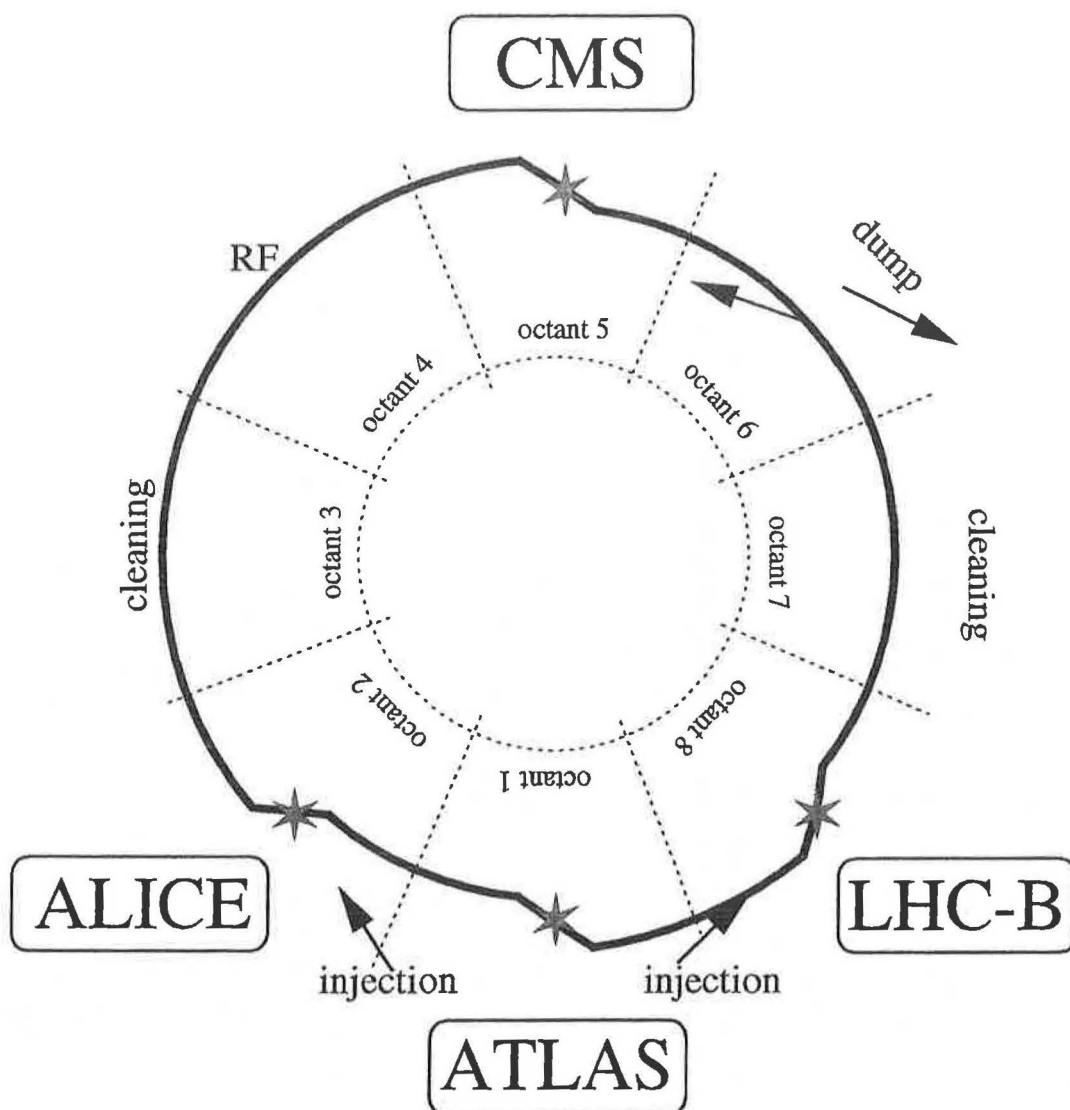


Figure 1.2: *Locations of the experiments on the LHC ring.*

1.3 The ATLAS experiment



The ATLAS detector ¹ is a general-purpose pp experiment designed to exploit the full discovery potential of the Large Hadron Collider. It is being designed by a collaboration of over 1500 physicists working at 140 institutions in 21 countries. A Letter of Intent [3] was submitted to the LHC Committee (LHCC) by the ATLAS Collaboration in October 1992. A more detailed design was presented in the ATLAS Technical Proposal [6] which was published in December 1994. The ATLAS project was formally approved by CERN in January 1996.

Given the schedule described in Section 1.2, the construction of the surface buildings for the ATLAS facilities will start in 1997. Following the termination of LEP operations, ATLAS will be installed at LHC Point 1 over a period of 2.5 years.

1.3.1 Basic design considerations

The LHC offers a large range of physics opportunities, amongst which the quest for the origin of spontaneous symmetry breaking in the electroweak sector of the SM is a major focus of interest for ATLAS. The Higgs search is therefore a prime benchmark for the detector optimisation. Other important goals are the searches for supersymmetric particles, for compositeness of the fundamental fermions, and for new heavy gauge bosons, as well as the investigation of CP violation in B-decays, and detailed studies of the top quark.

In order to maximise the chance of observing new physics, ATLAS will be designed to operate at high luminosity ($10^{34} \text{ cm}^{-2} \text{ s}^{-1}$) with as many different physics signatures as possible (e, μ, γ , jet, E_T^{miss} , b-tagging...). Emphasis is also put on the performance necessary for the physics accessible during the initial low-luminosity ($10^{33} \text{ cm}^{-2} \text{ s}^{-1}$) period, using in addition, more complex signatures such as τ and heavy-flavour tags from secondary vertices.

The basic requirements for ATLAS may be summarised as follows :

- Excellent electromagnetic calorimetry for e and γ identification and measurements, complemented by hermetic jet and E_T^{miss} calorimetry
- Efficient tracking at high luminosity for lepton momentum measurements, for b-

¹A Toroidal LHC ApparatuS

quark tagging, and enhanced e and γ identification, as well as τ and heavy-flavour vertexing and good reconstruction capability for a number of B-decay final states at low luminosity

- Standalone high-precision muon momentum measurements up to the highest luminosity and very low p_T trigger capability at lower luminosity

The physics reach will be maximised by constructing a detector which has coverage over a large η range and which has low- p_T thresholds for triggering and for particle momentum measurement.

1.3.2 Overall detector concept

The geometry is the familiar tracker-calorimeter-muon chamber onion skin configuration. See Figure 1.3.

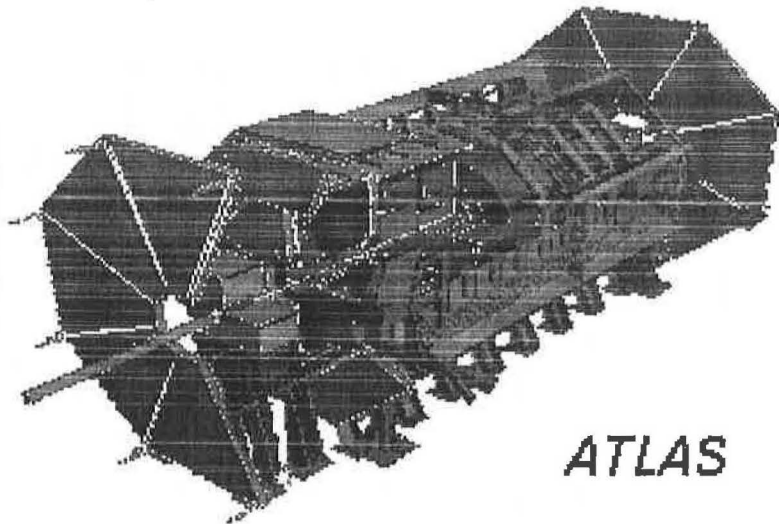


Figure 1.3: *The ATLAS detector.*

The global layout is largely determined by the configurations of the magnetic fields. A superconducting solenoid surrounds the inner tracking cavity. On the outside of this is the calorimeter which is itself surrounded by superconducting air-core toroids consisting of independent coils with eight-fold ϕ -symmetry. This magnet configuration makes it possible to build a high-resolution, large-acceptance, robust stand-alone Muon Spectrometer with minimal constraints on the calorimeter and Inner Detector.

The Inner Detector is contained within a cylinder of full length 6.80 m and radius 1.15 m and is surrounded by a solenoid providing a 2 T magnetic field parallel to the beam axis. High performance pattern recognition, momentum and vertex mea-

surements, and enhanced electron identification are achieved by combining ‘discrete’ high-resolution pixel and strip detectors in the part of the tracking volume closest to the interaction point with ‘continuous’ straw-tube tracking with transition radiation capability in the outer part.

Highly granular liquid argon (LAr) electromagnetic sampling calorimetry is employed in the pseudorapidity range $|\eta| < 3.2$. This will give excellent performance in terms of energy and position resolutions. The LAr technology is also used for the end-cap hadronic calorimeter, sharing cryostats with the e.m. end-caps. The same end-cap cryostats also house LAr forward calorimeters which cover the pseudorapidity range $3.2 < |\eta| < 4.9$. The LAr calorimetry is contained in a cylinder with an outer radius of 2.25 m and extends to ± 6.65 m along the beam axis. The bulk of the hadronic calorimetry is based on a scintillating tile technique and is situated in a cylinder outside the LAr cryostats, extending to an outer radius of 4.25 m and a length of ± 6.10 m. It is divided into a barrel cylinder and two ‘extended barrel’ cylinders. The whole calorimeter system contributes to the very good jet and E_T^{miss} performance of the detector. The total weight of the calorimeter system, including the solenoid flux return iron yoke, which is integrated into the tile calorimeter support structure, is about 4000 tons.

The Muon Spectrometer surrounds the calorimeters. The air-core toroid system, with a long barrel and two inserted end-cap magnets, generates a large field volume and strong bending power with a light and open structure. Multiple scattering effects are therefore minimal, and an excellent muon momentum resolution is achieved with three stations of high-precision tracking chambers. For triggering, the muon instrumentation is equipped with fast resistive plate chambers (RPCs).

The overall dimensions of the ATLAS detector are defined by the Muon Spectrometer. The outer chambers of the barrel are at a radius of about 11 m. The length of the barrel toroid coils is 13 m, and the third layer of the forward muon chambers, mounted on the cavern wall, is located at 21 m from the interaction point. The total weight of the ATLAS detector is about 7000 tons.

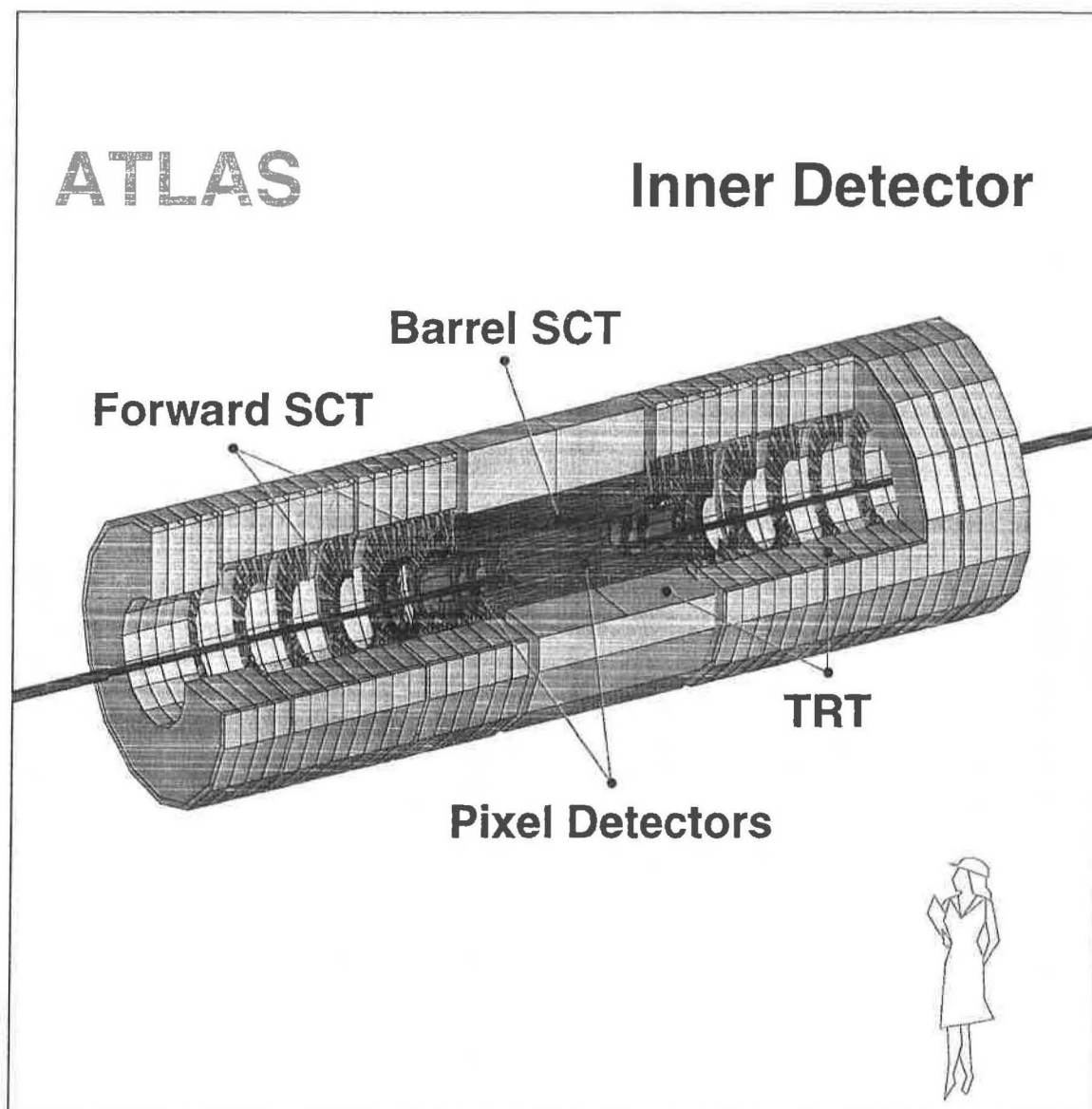


Figure 1.4: *The ATLAS Inner Detector (Morges layout).*

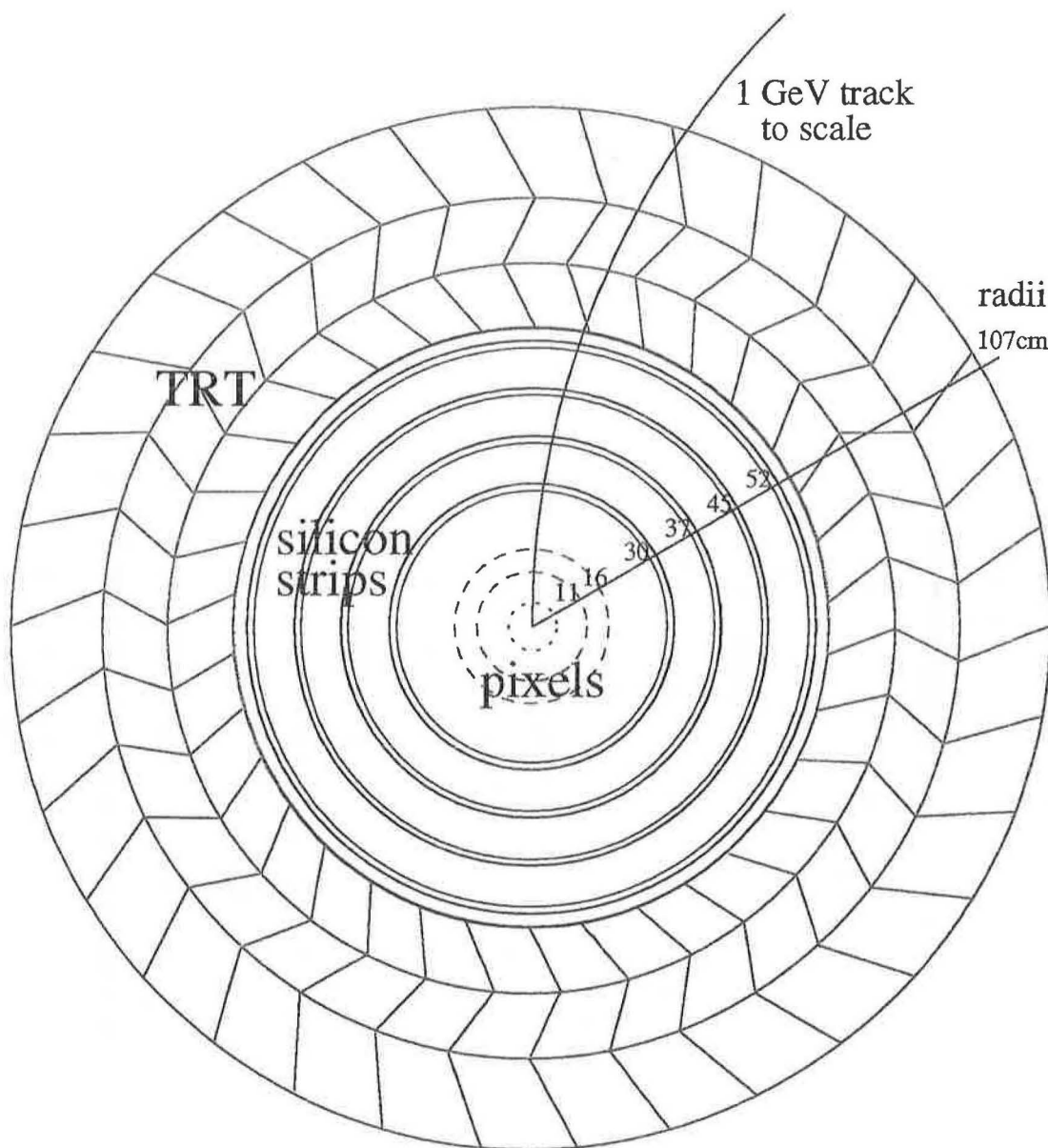


Figure 1.5: Section through the ATLAS Inner Detector at $\eta = 0$. (Morges layout).

1.3.3 Inner Detector

The purpose of the Inner Detector is to make high-precision measurements of the kinematic parameters of charged tracks moving in a solenoidal magnetic field with maximal capability for pattern recognition, particle identification and triggering. These requirements must be met in conditions of high track density at a high bunch crossing rate. The physics goals define a set of requirements for the Inner Detector. The following is a brief summary of the major requirements :

- coverage over the pseudorapidity range $|\eta| = 2.5$
- transverse momentum resolution of better than 30% at $p_T = 500 \text{ GeV}$
- high-efficiency track finding for isolated and associated tracks
- good electron and photon identification capability
- high-efficiency b-tagging
- precise secondary vertexing
- tracking trigger at Level 2

The high track density and the stringent momentum and spatial resolution targets require the use of high-granularity tracking. In order to maximise the capability for resolving the inevitable ambiguities caused by overlapping tracks, secondary interactions and detector inefficiencies, it is desirable to make a large number of measurements along the length of a track. It is believed that a combination of high-precision ‘discrete’ (i.e. few point) and low-precision ‘continuous’ (i.e. many point) tracking will offer the best possible track finding and track fitting capabilities. The discrete tracking will be composed of semiconductor pixel and strip detectors at radii close to the beam axis. The continuous tracking will be provided by straw drift tubes situated at higher radii.

Semiconductor strip detectors will be used at radii between 30 cm and 60 cm from the beam axis. In the barrel and at higher radii in the forward region, the substrate will be silicon. In the region of highest fluence in the forward region, a gallium arsenide substrate may be used if this is shown to confer higher radiation tolerance. The measurement in the non-bending coordinate will be obtained through the use of a small stereo angle, since the precision requirement for this coordinate is less stringent.

The radiation levels at the LHC preclude the long-term operation of silicon strip detectors within a radius of about 30 cm from the beam axis. Pixel detectors are expected to have a resistance to radiation damage of the order of 10 times better than silicon strip detectors. This and the high track density have led to the choice of pixel detectors for use close to the interaction point.

The relatively high cost per unit area of semiconductor layers and their high radiation length mean that the number of precision layers must be limited. Straw tubes will be used at radii greater than 60 cm where the track density will be relatively low and the > 36 track points produced by about 64 closely-spaced layers may be used to best advantage for pattern recognition. The position resolution will be inferior to that of the semiconductor detectors, but the cost per point and the power dissipation will be much lower. Radiator material will be included to generate transition X-rays to give an improved electron identification capability independent of the energy-momentum matching between the calorimeter and the tracker. This information will be particularly useful at lower transverse momenta (< 5 GeV). The straw tube system is known as the Transition Radiation Tracker (TRT).

The layout of the precision tracking layers will be chosen to ensure that a track crosses two pixel layers and four silicon strip superlayers over the pseudorapidity range $|\eta| < 2.5$. The TRT will provide at least 36 straw hits over the same angular range. A transition from barrel to disc geometry is made starting at $\eta \sim 1$ in order to minimise the amount of material traversed.

A ‘vertexing’ layer at low radius is included in the Inner Detector design for high-performance vertexing for B-physics studies during the initial period of low-luminosity running. It will have a short lifetime and will be removed for high-luminosity phase, but will significantly enhance the B-physics potential of the experiment by virtue of its smaller distance from the beamline and low multiple scattering. This layer is foreseen as either an additional pixel layer at a radius of 4 cm or a double-sided silicon strip layer at $r \sim 6$ cm.

In order to meet the tracking precision requirements it is likely that a survey system capable of determining positions of detector elements at a level of around $10 \mu\text{m}$ during operation of the detector will be required. Much of this thesis is concerned with the specification and design of such a survey system.

ATLAS Calorimetry

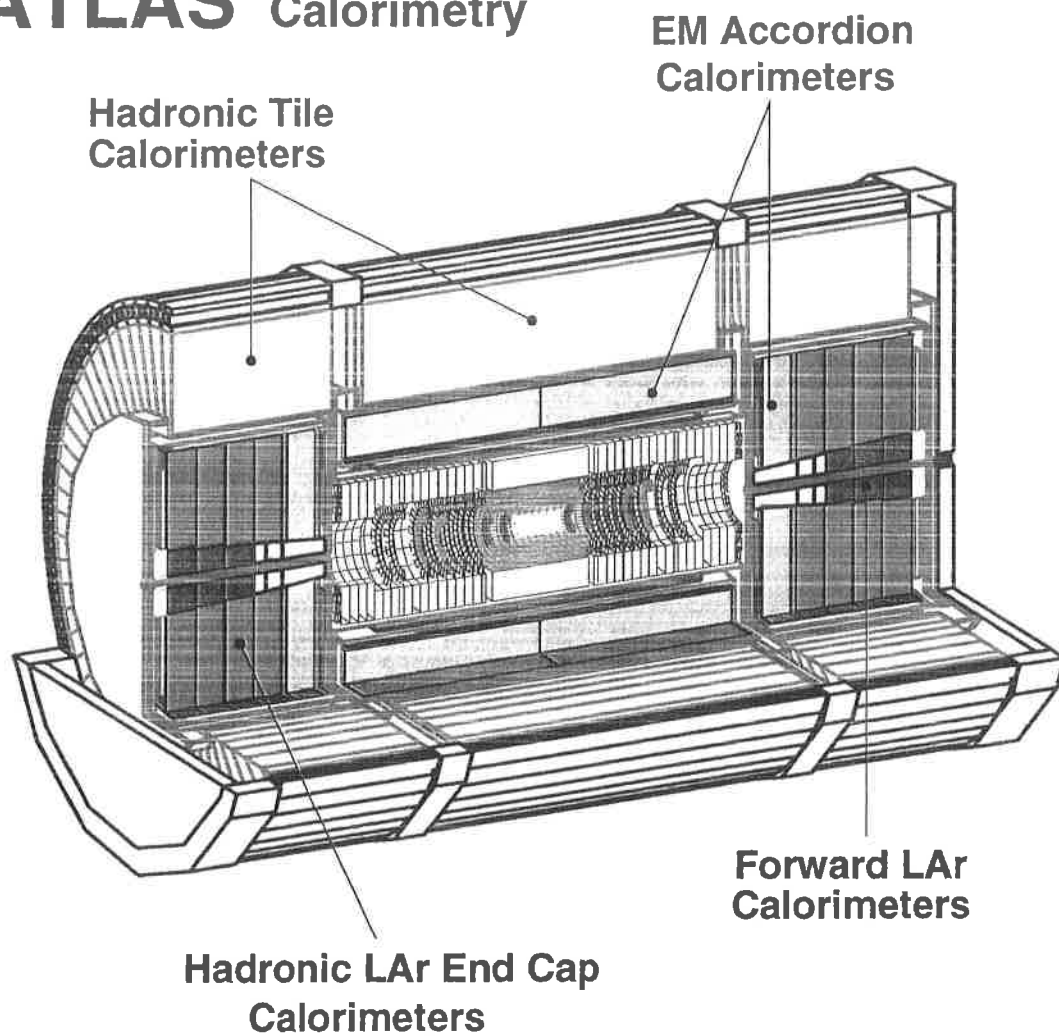


Figure 1.6: *The ATLAS calorimeter system.*

1.3.4 Calorimetry

Many important physics processes require the identification and reconstruction of the energy of electrons, photons and jets as well as measurement of missing transverse energy (E_T^{miss}). The ATLAS calorimetry system is designed to meet these requirements in the very high luminosity environment of the LHC, with an acceptance covering the region $|\eta| < 4.9$.

Many of the important requirements for the electromagnetic calorimetry come from the Higgs processes and from decays of new heavy gauge bosons W' , Z' to electrons. The channels $H \rightarrow \gamma\gamma$ and $H \rightarrow ZZ \rightarrow 4e$ place the most stringent requirements on the energy resolution. The goal is an e.m. energy resolution of $\frac{10\%}{\sqrt{E}} \oplus 0.7\%$. Fine segmentation of the e.m. calorimeter is required to reach a rejection of 10^4 against single jets for sensitivity to $H \rightarrow \gamma\gamma$. A wide dynamic range is required : from 2 GeV (for i.d. of electrons from semileptonic b decays) to 5 TeV (for $Z' \rightarrow ee$ and $W' \rightarrow e\nu$). The channel $H \rightarrow ZZ \rightarrow 4e$ also requires low-energy electron reconstruction down to a transverse energy of 5 GeV.

Requirements for the hadronic calorimetry are the identification and measurement of the energy and direction of jets, and precise measurement of missing transverse energy. In addition, the hadronic calorimetry will enhance the performance of the e.m. calorimeter with measurements of quantities such as leakage and isolation. A hadronic calorimeter energy resolution of $\frac{50\%}{\sqrt{E}} \oplus 3\%$ in the barrel (with segmentation 0.1×0.1) and $\frac{100\%}{\sqrt{E}} \oplus 10\%$ in the forward regions (with segmentation 0.2×0.2) is considered sufficient.

The layout of the ATLAS calorimeters is shown in Figure 1.6. An annular cryostat surrounding the Inner Detector cavity contains the barrel electromagnetic calorimeter and the solenoidal coil which provides the magnetic field for the tracker. Two end-cap cryostats enclose the electromagnetic and hadronic end-cap calorimeters as well as the integrated forward calorimeter. The three cryostats are surrounded over their full length by a scintillating-tile hadronic calorimeter which is divided into three sections (a barrel and two ‘extended barrels’) and is contained within an outer support cylinder which doubles as the the main solenoid flux return.

1.3.4.1 Liquid Ionisation Calorimetry

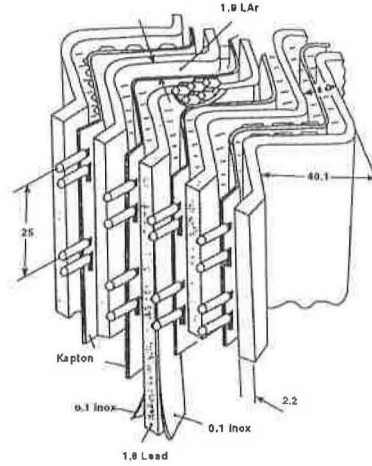
The inner barrel and end-caps are based on an intrinsically radiation-resistant technology using liquid argon (LAr) as the active material. In the electromagnetic barrel

Calorimeter	$ \eta $	absorber/active	X_0/λ	$\Delta\eta \times \Delta\phi$	channels
LAr calorimetry					
Barrel e.m. PS	0–1.4	Pb/LAr	$< 2 X_0$	0.003×0.100	28 000
Barrel e.m.	0–1.4	Pb/LAr	$4.5 X_0$	0.003×0.100	57 000
			$13\text{--}20 X_0$	0.025×0.025	29 000
			$7.5\text{--}12 X_0$	0.025×0.050	14 000
End-cap e.m.	1.4–3.2	Pb/LAr	$29X_0$	various	82 000
End-cap had.	1.5–3.2	Cu/LAr	10λ	$4 \times 0.1 \times 0.1$	8 800
Forward e.m.	3.1–4.9	Cu/LAr	$26X_0/2.5\lambda$	$(\eta = 3.1\text{--}3.6) 0.120 \times 0.100$ $(\eta = 3.6\text{--}4.6) 0.140 \times 0.200$ $(\eta = 4.6\text{--}4.9) 0.150 \times 0.400$	6060
Forward had.	3.1–4.9	W+Fe/LAr	$94X_0/3.4\lambda$	as forward e.m.	4040
			$93X_0/3.4\lambda$		3030
Scintillator tile calorimetry					
Barrel	0–1.0	Fe/scintillator	1.5λ	0.1×0.1	6 000
			4.2λ	0.1×0.1	6 000
			1.9λ	0.2×0.1	6 000
Ext. barrel	1.0–1.6	Fe/scintillator		$3 \times .1 \times .1$	4 000

Table 1.3: Summary of the ATLAS calorimetry.

calorimeter, lead absorber plates and Kapton electrode boards are placed in LAr in an ‘accordion’ configuration. The end-cap cryostats contain a Pb/LAr e.m. accordion calorimeter with the so-called ‘Spanish fan’ geometry, a Cu/LAr hadronic calorimeter with flat absorber plates, and a Cu,Fe,W/LAr forward calorimeter with a tube electrode structure.

The barrel electromagnetic calorimeter is constructed from identical half-barrels and covers the rapidity range $|\eta| < 1.4$. It is preceded radially by a presampler layer located immediately behind the cryostat inner wall designed to correct for the energy lost in the material in front of the calorimeter (i.e. the Inner Detector, coil and cryostat) and to assist in measuring the direction of e.m. showers with the necessary accuracy in η . The first sampling of the barrel e.m. calorimeter (at a depth of $4.5 X_0$) is finely segmented in η (0.003×0.100) and plays the rôle of a preshower detector. The second and third samples have granularities of 0.025×0.025 and 0.025×0.050 respectively.



The end-cap hadronic calorimeter is a LAr calorimeter using kapton electrode boards glued to flat copper plates with 8mm gaps. The segmentation is 0.1×0.1 with four longitudinal samplings. The use of copper rather than iron allows a lower cell capacitance and hence lower electronic noise, a lower optimum integration time and reduced pile-up noise.

The LAr forward calorimetry, integrated into the end-cap cryostats, will operate in an environment of extreme particle and energy flux. Speed and radiation hardness are achieved with a metallic tube and rod electrode structure with a very small LAr gap ($250 \mu\text{m}$), embedded in a copper absorber for the electromagnetic front module, and a tungsten alloy absorber for the two hadronic modules behind it.

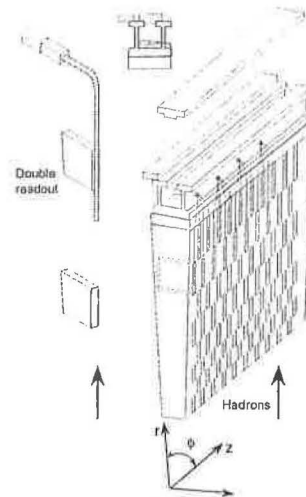
1.3.4.2 Hadronic Scintillator Calorimeter

The hadronic scintillating tile calorimeter occupies the radial interval $r = 2.3\text{--}4.2\text{ m}$. It is divided into a 5.6 m barrel and two 2.6 m extended barrels which cover the regions $|\eta| < 1.0$ and $1.0 < |\eta| < 1.6$ respectively.

The calorimeter is based on a sampling technique using a steel absorber material and 3 mm thick plastic scintillator tiles which are read out using 1 mm diameter wavelength-

shifting fibres. Photomultiplier tubes for detection of the scintillation light and associated electronics are housed in 512 identical drawers which slide inside iron girders supporting the calorimeter modules. This design allows easy access to the PMTs and the electronics with minimal interference with other parts of the detector.

An innovative feature of the design is the orientation of the scintillating tiles which are placed in planes perpendicular to the colliding beams and staggered in depth. Monte Carlo simulations have demonstrated that this orientation provides good sampling homogeneity. Radially the calorimeter is segmented into three layers with depths of 1.5, 4.2 and 1.9 radiation lengths at $\eta = 0$. The $\eta \times \phi$ segmentation will be 0.1×0.1 (0.2×0.1 in last layer). Prototype tests indicate that this system will meet the requirements.



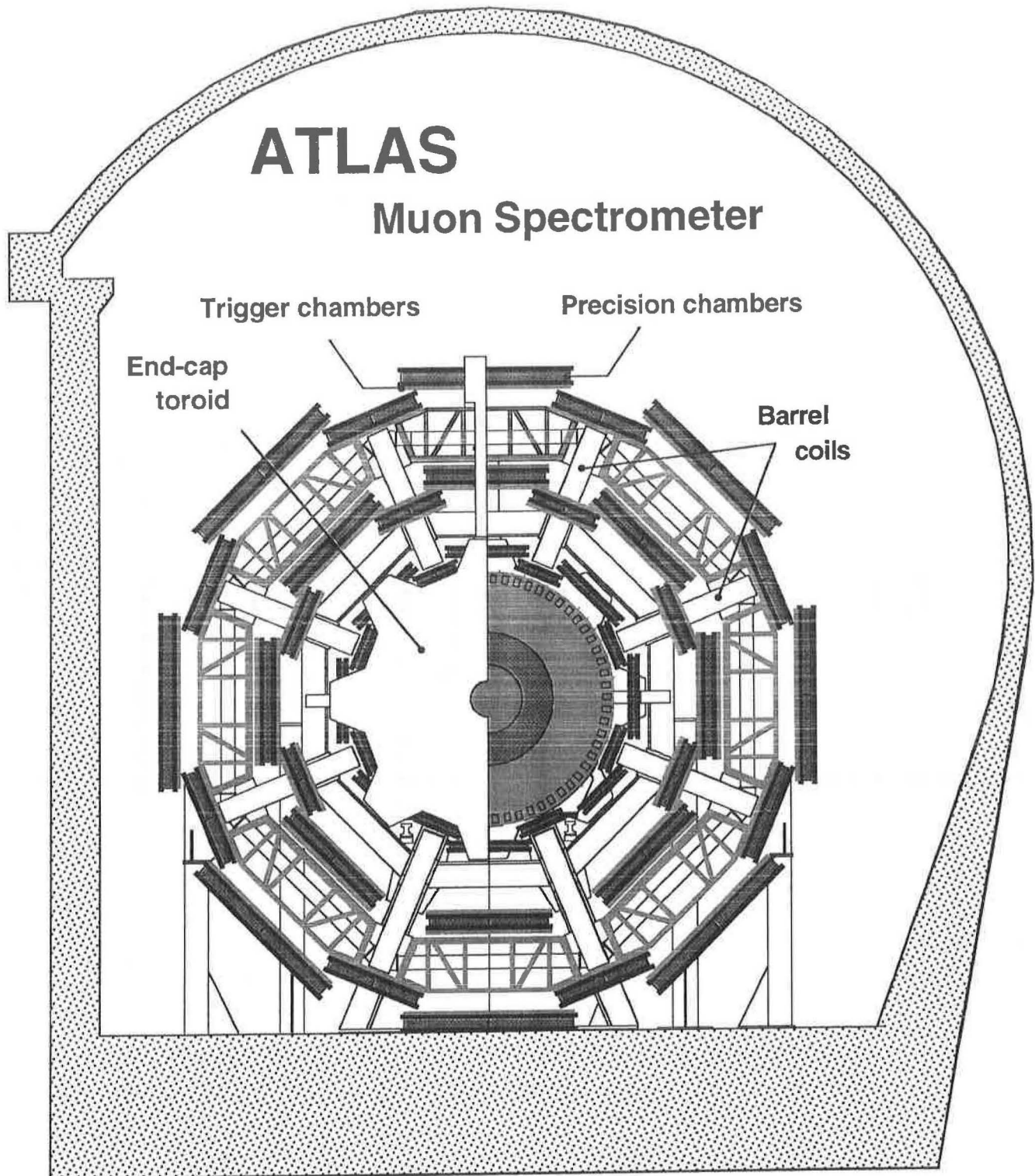


Figure 1.7: *End ($r\phi$) view of the ATLAS Muon Spectrometer.*

1.3.5 Muon Spectrometer

The layout of the Muon Spectrometer is shown in Figures 1.7 and 1.8. High precision tracking chambers are used to fully exploit the advantages offered by the open superconducting air-core muon toroid magnet system. In the barrel, the natural layout consists of three layers of chambers: at the inner and outer edges of the magnetic volume and in the mid-plane. In the forward direction the chambers are placed at the front and back faces of the toroid cryostats, with a third layer against the cavern wall, to maximise the lever-arm. The high-precision chambers are complemented with an independent fast trigger chamber system.

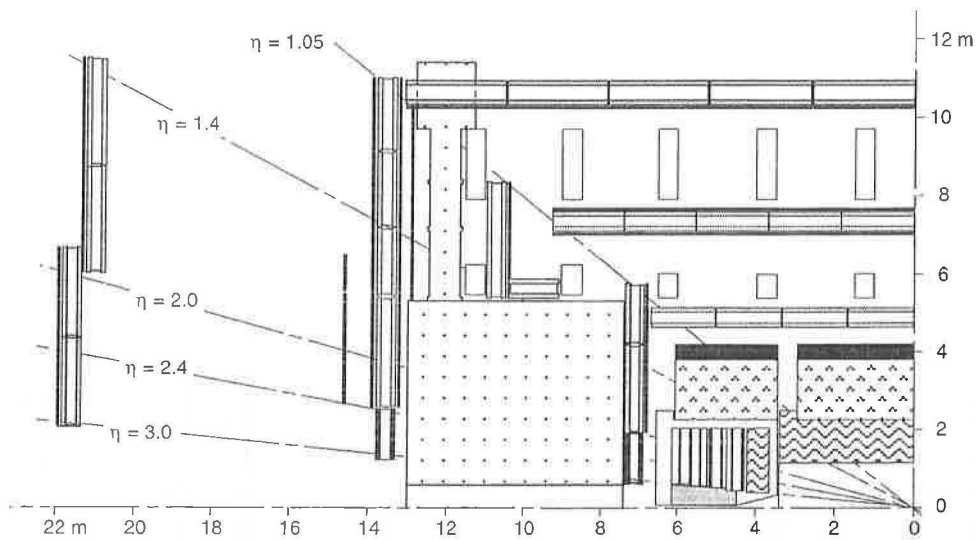


Figure 1.8: Side (rz) view of the Muon Spectrometer. The precision multilayers are shown in dark grey, their support structures in light grey. The trigger chambers are drawn in black.

1.3.5.1 Precision chambers

The precision muon tracking is based on a new muon chamber concept: the Monitored Drift Tube (MDT). MDT chambers can be used over a very large part of the η acceptance, although Cathode Strip Chambers (CSC) will be used in the highest rate environment at large η .

The MDT chambers consist of two multilayers of three or four planes of pressurised thin-wall aluminium drift tubes with a diameter of 30 mm. When operated at a pressure in the range 3–4 bar, single cell resolutions of about $60\mu\text{m}$ have been achieved, with non-flammable drift gases. The precise location of the two multilayers is provided by

three stiff comb plates of the support structure which define the wire positions at the ends of the chamber.

The CSCs are fast multiwire proportional chambers for which precise position measurements (typically $50\text{ }\mu\text{m}$) are achieved by determining the centre of gravity of the induced charge on the segmented cathode strips. High rates can be handled by appropriate fine segmentation. Multilayers of CSCs will be assembled in a similar way as for the MDTs to provide local track vectors.

The momentum resolution of the muon spectrometer relies heavily on the ability to master the alignment of the large muon chambers spaced far apart. The alignment will be monitored using optical straightness monitors. These are discussed in Chapter 4.

1.3.5.2 Trigger system

The high-rate environment combined with cost considerations make it necessary to use two trigger chamber technologies: Resistive Plate Chambers (RPC) in the barrel and Thin Gap Chambers (TGC) in the forward regions. The RPC is a gaseous parallel plate detector with a time resolution of a few nanoseconds. Position information is obtained from external pick-up electrodes segmented into strips. The TGC is a wire chamber operating in saturated mode. Capacitive readout on pads or strips is used, with a time resolution of typically less than 5 ns.

In addition to their primary function, the trigger chambers complement the MDT chambers with the bunch crossing assignment and the measurement of the ‘second coordinate’ in the non-bending plane, which is needed to cope with the ϕ -variations of the magnetic field and is required in the alignment procedure.

1.3.6 Trigger, DAQ and computing

1.3.6.1 Requirements

The ATLAS trigger system has the task of selecting interesting events from the mass of collisions at LHC. The trigger must reduce the event rate from the bunch crossing rate of 40 MHz to around 10 Hz for recording onto mass storage media. The data rates are immense: after compression the rate is 1Mb/event or 40Tb/second.

1.3.6.2 Architecture

The ATLAS trigger is organised in three trigger levels:

- **Level 1:** a synchronous (i.e. fixed latency) system of dedicated hardware acting at 40MHz on reduced-granularity data from a subset of the detector subsystems (calorimeters and muon spectrometer). During the Level 1 processing, the data from all parts of the ATLAS detector are held in pipeline memories. A yes/no trigger decision is delivered for each bunch crossing with a fixed latency of around $2.5\mu\text{s}$.
- **Level 2:** an asynchronous system of programmable processors using full-granularity, full-precision data from much of the detector, but examining only regions of interest (RoIs) of the detector identified by Level 1 as containing interesting information.
- **Level 3:** an event building and reconstruction system using farms of commercial processors ($\sim 10^6$ MIPS), operating on the assembled full event data arriving at 10–100 Hz, to make the final selection of events to be permanently recorded at around 10 Hz for offline analysis.

The overall DAQ architecture is shown in Figure 1.9.

1.3.6.3 Offline computing

In the LHC environment, the boundary between offline and online computing will be less clearly defined than in previous experiments because ‘offline’ algorithms will be applied in real-time for event selection (at Level 3) and first-pass reconstruction may be done online. With improved understanding of the detector, data and algorithms, more and more tasks may be migrated from the offline to the online environment.

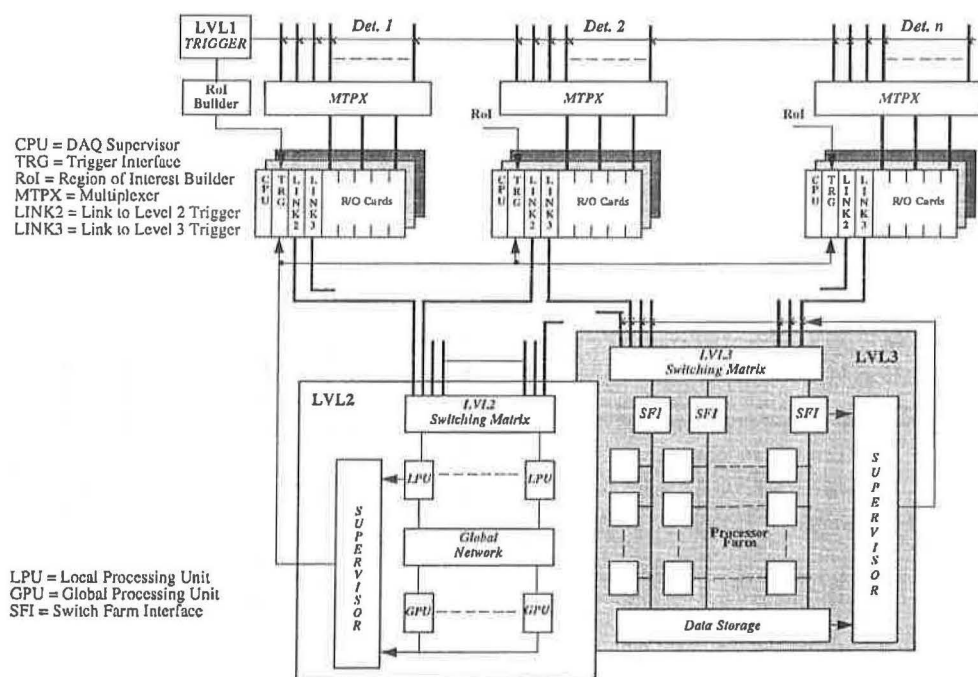


Figure 1.9: Overall trigger/DAQ architecture.

1.3.7 Physics at ATLAS

The following is a brief summary of the physics which will be studied with the ATLAS detector. The total cross-section for proton-proton collisions at 14 TeV is expected to be ~ 70 mb.

1.3.7.1 Higgs search

Since one of the prime reasons for building LHC is to solve the puzzle of electroweak symmetry breaking, considerable attention has been paid during the detector design process to the sensitivity to Higgs signatures.

low mass ($80 < m_H < 120$ GeV) :

For a SM Higgs in the mass range $80 < M_H < 120$ GeV, the decay modes important for discovery are $H \rightarrow \gamma\gamma$ and $H \rightarrow b\bar{b}$.

$H \rightarrow \gamma\gamma$ places severe demands on the performance of the electromagnetic calorimeter. Excellent energy and angular resolution will be required to observe the narrow mass peak above the irreducible prompt $\gamma\gamma$ continuum. Powerful particle identification capability will be needed to reject the large jet background and, for $m_H \sim m_Z$, the potentially dangerous resonant background from $Z \rightarrow ee$ decays.

For a SM Higgs boson in the mass region below the threshold for $H \rightarrow WW$ and $H \rightarrow ZZ$, the branching ratio for $H \rightarrow b\bar{b}$ is essentially 100%, since in this scenario the b-quark is the heaviest accessible particle.

The first years of running at low luminosity will not be sufficient for statistically significant discovery in any single channel. However, the combination of the various channels should provide enough evidence for a SM Higgs boson even in this difficult mass range.

intermediate mass ($120 < m_H < 180$ GeV) :

The $H \rightarrow ZZ^* \rightarrow 4\ell$ mode provides a very clean signature for a SM Higgs boson in the mass range from ~ 130 GeV to ~ 800 GeV. However, for $m_H < 2m_Z$ (i.e. below 180 GeV), although the event signature is very distinctive, the Higgs boson is narrow, the expected signal cross-sections are small, and the backgrounds are significant. Good mass resolution is required to find a signal.

high mass ($m_H > 180$ GeV) :

For $180 < m_H < 800$ GeV the $H \rightarrow ZZ \rightarrow 4\ell$ channel has a large rate and small

background, and is considered to be the most reliable for SM Higgs discovery. The major background from continuum ZZ production can be reduced considerably by requiring one Z to have $p_T > m_H/2$, resulting in a signal-to-noise ratio of 20:1 for $m_H < 500$ GeV.

For masses greater than 800 GeV, the channel $H \rightarrow ZZ \rightarrow \ell\ell\nu\nu$ may be considered. It benefits from a six times higher rate, but the decay cannot be completely reconstructed because of the escaping neutrinos.

SUSY Higgs :

In the MSSM, two Higgs doublets are required, resulting in five physical states, usually referred to as H^+ , H^- , h , H and A . At the tree level, their masses can be computed in terms of only two parameters, usually quoted as m_A and $\tan\beta$ (the ratio of the vacuum expectation values of the two doublets), where the parameter space is usually shown as extending over $0 < m_A < 500$ GeV and $0 < \tan\beta < 50$. Important channels are $A/H \rightarrow \tau\tau$, $h/H \rightarrow \gamma\gamma$, $H \rightarrow ZZ \rightarrow 4\ell$, and $t \rightarrow bH^+$, of which the $A/H \rightarrow \tau\tau$ mode has been studied in detail. Increasing the calorimeter coverage as far as the beam pipe (i.e. to $|\eta| < 7.6$) could significantly improve the acceptance.

Strongly interacting Higgs :

A Higgs of high mass $m_H > 1$ TeV is very broad and leads to strong scattering of gauge boson pairs, which would eventually violate unitarity for the s-wave scattering amplitude. At this energy scale, longitudinally polarised W_L s play the role of the Goldstone bosons of the Higgs symmetry-breaking mechanism. Thus, WW fusion will be dominated by the longitudinal components, and studies of the cross-section regularisation mechanism using the rate of production of W_L pairs at the LHC will provide information about the Higgs boson, if it exists, or alternatively about the nature of whatever dynamical process is responsible for electroweak symmetry breaking.

1.3.7.2 Top-quark physics

The LHC will be the world's first top factory. Even at very low initial luminosities of $10^{33} \text{ cm}^{-2} \text{ s}^{-1}$, approximately 6000 $t\bar{t}$ pairs will be produced per day for $m_t = 170$ GeV, yielding about 100 reconstructed $t\bar{t} \rightarrow (\ell\nu b)(jjb)$ decays per day and about 10 clean isolated $e\mu$ pairs per day.

The mass of the top-quark will be measured at LHC using $t \rightarrow jjb$ and multilepton events, yielding an expected ultimate accuracy of about $\pm 2 \text{ GeV}$.

In extensions to the SM with charged Higgs, such as the MSSM, the decay $t \rightarrow bH$, if kinematically allowed, can compete with the $t \rightarrow bW$ decay.

Rare decay modes such as $t \rightarrow Zc$, $t \rightarrow Ws$ or Wd , and $t \rightarrow W_R b \rightarrow jjb$ will also be studied.

1.3.7.3 B-physics

The very large cross-section for b-quark production at the LHC will allow a wide range of precision measurements to be performed in the rich field of B-physics. The ATLAS programme will include studies of CP-violation using the decays $B_d \rightarrow J/\psi K_s$, $B_d \rightarrow \pi\pi$, and $B_s \rightarrow J/\psi \phi$. Other topics are the measurement of B_s mixing, searches for rare decays such as $B_d \rightarrow \mu\mu$ and $B_s \rightarrow \mu\mu$, the study of B-baryon decay dynamics, and spectroscopy of rare B hadrons.

B-physics will be experimentally easiest during the first few years of LHC operation when the low luminosity will be low and hence the effect of pile-up will be reduced. The vertex detector, located close to the beam axis, is expected to survive this initial phase. Although much of the B-physics studies will be performed during this period, B-physics may also be usefully pursued at ATLAS during the high luminosity phase of the LHC.

Strong features of the ATLAS detector for the B-physics programme are a powerful and flexible trigger system, high-resolution secondary vertex measurement, and efficient track reconstruction and electron identification down to low p_T .

1.3.7.4 Supersymmetric particles

Supersymmetric (SUSY) extensions of the SM predict a wide spectrum of new particles. The many theoretical advantages of supersymmetric extensions of the SM require that the scale for the masses of the supersymmetric partners of ordinary particles cannot be significantly larger than the scale for electroweak symmetry breaking. Thus the masses of the SUSY particles (and also the production rates) are such that the ATLAS detector could discover them over a large fraction of the parameter space.

In addition to the SUSY Higgs signatures, three main signatures allowing the exploration of a significant region in parameter space will be pursued, namely multijets + E_T^{miss} , same-sign dileptons from squark and gluino production, and three-lepton events from chargino/neutralino production.

1.3.7.5 Other searches

New neutral or charged vector gauge bosons occur naturally in several models, e.g. some minimal extensions of the Standard Model and models for electroweak symmetry breaking through compositeness. The sensitivity of ATLAS to these signals is discussed in Chapter 2.

Leptoquarks (LQs) are predicted in many theoretical models beyond the Standard Model (SM), inspired by the symmetry between the quark and lepton generations.

Recent results from the Tevatron, though not statistically significant, have tantalisingly raised the prospect of quark compositeness. If quarks and/or leptons contain more fundamental constituents, deviations from Standard Model predictions would be expected to be seen by ATLAS.

Measurements of gauge-boson pair production provide tests of the SM $SU(2)$ triple-gauge-boson vertices, and so could be useful probes of the structure of the Higgs sector.

1.4 Conclusion

About ten years from now, the Large Hadron Collider will offer unmatched opportunities for the discovery of new physics. The ATLAS detector promises to exploit its potential to the full.

References

- [1.1] CERN, le Conseil Européen pour la Recherche Nucléaire, has its headquarters near Geneva, Switzerland. At present, the Member States of CERN are Austria, Belgium, the Czech Republic, Denmark, Finland, France, Germany, Greece, Hungary, Italy, the Netherlands, Norway, Poland, Portugal, the Slovak Republic, Spain, Sweden, Switzerland and the United Kingdom. The following countries and organisations have observer status: Israel, Japan, the Russian Federation, Turkey, Yugoslavia (status suspended after UN embargo, June 1992), the European Commission and UNESCO.
- [1.2] The European Committee for Future Accelerators (ECFA), *Large Hadron Collider Workshop*, CERN 90-10, ECFA 90-133, Aachen, 3 December 1990.
- [1.3] The ATLAS Collaboration, *Letter of Intent by the ATLAS Collaboration for a General-Purpose pp Experiment at the LHC*, CERN/LHCC/92-4 LHCC/l 2, 1 October 1992.
- [1.4] The ATLAS Collaboration, *Technical Proposal for a General-Purpose pp Experiment at the Large Hadron Collider at CERN*, CERN/LHCC/94-43 LHCC/P2, 15 December 1994.
- [1.5] The CMS Collaboration, *The Compact Muon Solenoid Technical Proposal*, CERN/LHCC/94-38 LHCC/P1, 15 December 1994.
- [1.6] William Shakespeare was perhaps the earliest critic of CMS: "Thou art no Atlas for so great a weight.", *King Henry VI Part 3*, Act 5 Scene 1 (1595).
- [1.7] The LHC-B Collaboration, *LHC-B Letter of Intent: A Dedicated LHC Collider Beauty Experiment for Precision Measurements of CP-Violation*, CERN/LHCC 95-5 LHCC/I 8, 25 August 1995.

- [1.8] The ALICE Collaboration, *Technical Proposal for A Large Ion Collider Experiment at the CERN LHC*, CERN/LHCC/95-71 LHCC/P3, 15 December 1995.
- [1.9] The CDF Collaboration, *CDF Results on Top*, CDF/TOP/PUBLIC/3317, 20 September, 1995.

Books and review articles

- [1.10] P. Fayet and S. Ferrara, *Supersymmetry*, Phys. Rep. 32 (1977) 249–334.
- [1.11] E. Eichten, I. Hinchcliffe, K. Lane and C. Quigg, *Supercollider physics*, Rev. Mod. Phys. 56 (1984) 579.
- [1.12] H.E. Haber and G.L. Kane, *The search for supersymmetry: probing physics beyond the Standard Model*, Phys. Rep. 117 (1985) 75–263.
- [1.13] M. Guidry, *Gauge Field Theories*, John Wiley and Sons, Inc., 1991.
- [1.14] I.J.R. Aitchison and A.J.G. Hey, *Gauge Theories in Particle Physics*, 2nd edition, IOP Publishing, Bristol, 1989.
- [1.15] M. Kaku, *Quantum Field Theory*, Oxford University Press, 1993.
- [1.16] E. Leader and E. Predazzi, *An introduction to gauge theories and modern particle physics*, Cambridge University Press, 1996.



Chapter 2

Alignment of the ATLAS Inner Detector

2.1 Introduction

For many of the interesting channels expected at the LHC, good sensitivity will require precise measurement of the kinematic parameters of high- p_T charged tracks. In order to do this, the Inner Detector is designed to make precise position measurements of points on these tracks as they curve through a solenoidal magnetic field. The precision with which this can be done is dependent upon both the intrinsic measurement precision of the detector elements employed and the accuracy with which their locations are known. The intrinsic precision of a detector is simply a property of its design. The positioning accuracy of a detector element is a combination of the accuracy with which it was positioned at assembly and the stability of its location subsequently. The overall process of accurately determining the positions of detector elements both before and after assembly is known as **alignment**.

Since the precision with which the positions of detector elements must be known during running will in general be higher than that which can be inferred from knowledge of the detector just after assembly, it will be necessary to determine the location of each element during operation of the detector. This may be done either indirectly using the track data, or directly using a **survey system**. Survey measurements must be made with a precision sufficient for the total combined precision to be within the specifications derived from the physics requirements.

$$\sigma_{\text{total}} = \sigma_{\text{intrinsic}} \oplus \sigma_{\text{survey}} < \sigma_{\text{required}}$$

Since the detectors are optimised for maximal physics performance with minimal channel count, the choice of the intrinsic detector precisions is influenced by the expected alignment precision — in some cases good alignment may permit slightly lower intrinsic precisions and hence savings in cost.

Survey measurements will be combined geometrically to produce a three-dimensional picture of the locations of the detector elements.

It is useful to factorise the alignment problem into the **local** or **internal alignment** of individual subdetectors which are then spatially related to each other with a **global alignment** of these large blocks. If these blocks are internally aligned then they may be considered as rigid bodies, leaving only 6 degrees of freedom for the relative location of any pair of subdetectors.

The detector will be referenced to the beam pipe, the interaction point, and to selected points on the wall of the underground cavern.

Eventually the survey measurements will be made available for use in the offline data analysis software (and to some extent in the trigger system) as a set of corrections to the ideal element locations.

Precise knowledge of detector positions will be required at every stage of the construction of ATLAS. There are at least five distinct phases during the design and construction of the Inner Detector during which precise position measurements will be necessary :

- at the design stage for choosing materials and evaluating mechanical designs
- during module construction, testing and calibration
- during barrel/wheel preassembly
- after installation of the detector into the underground area
- periodically during running

2.2 Inner Detector alignment requirements

2.2.1 Physics requirements

High sensitivity to a wide range of physics processes places stringent requirements on the performance of the Inner Detector. The physics requirements may be translated

into three classes of performance specifications: those for track finding (i.e. pattern recognition), track fitting and triggering.

As discussed above, the effective track fitting resolutions of the Inner Detector are combinations of the intrinsic detector resolutions and the alignment precisions. The alignment requirements for acceptable track fitting resolutions are investigated in detail in Chapter 5. The physically-motivated specifications for each of five track fitting resolutions are reviewed in Section 5.2.4.

There is currently little information available on the alignment requirements which may be derived from a need for good pattern recognition performance.

2.2.2 Engineering requirements

Ultimately the need for a run-time alignment system and the performance required of it will be determined by the positional stability of the detector elements over time. This involves detailed knowledge of the performance of both the modules and the support structures under the expected mechanical and thermal loads. The innovation required in the mechanical design and the use of new materials has made it difficult to predict this stability. The detailed design of the alignment system will depend on the degree of stability expected. Prototyping will be used to tune the FEA models.

The luminosity of the LHC will vary considerably during each machine fill. The variation in power dissipation in the front-end electronics that this will cause (perhaps 10% of the total) may be large enough to cause significant changes in the shape of the detector. In such a case, a direct survey system would need to operate fast enough to permit several complete surveys during a run.

2.2.3 Alignment requirements for the Level 2 trigger

In the offline track finding, alignment errors will be corrected in software using survey and other measurements. However, it is desirable that the Level 2 tracking trigger can be implemented in hardware without the requirement that it include alignment corrections. This means that the correct functioning of the tracking trigger may impose requirements on the assembly tolerances of the Inner Detector.

The Level 2 tracking trigger uses the four $r\phi$ silicon layers at radii of 30–60 cm from the beam pipe. The readout granularities for the trigger are given in Table 2.1.

The dependence of the tracking trigger performance on the assembly tolerances of the silicon barrel has been investigated using the high- p_T track trigger and the ‘Panel’

Layer	readout pitch	radius
1	150 μm	30 cm
2	200 μm	40 cm
3	200 μm	50 cm
4	200 μm	60 cm

Table 2.1: *Tracking trigger granularity.*

layout [3]. The effect of alignment errors was simulated by increasing the effective strip widths in the four silicon layers. The electron efficiency and the jet rejection of the trigger were investigated.

- **Electron efficiency :** increasing the strip width reduces the p_T resolution of the trigger algorithm and so smears out the p_T threshold of the electron trigger. This will cause more fake triggers.
- **Jet rejection :** Table 2.2 shows the jet rejection performance of the Level 2 track trigger on jets which pass Level 1 in the presence of a mean of 20 minimum bias events. Any reduction in resolution results in deterioration of the standalone track trigger performance. At $\frac{1}{2}$ -resolution the effect may be reduced by requiring tracker-calorimeter matching. However, for $\frac{1}{4}$ -resolution, the effect of this matching is not sufficient to prevent deterioration of the global rejection.

Si resolution	% events passing trigger (excluding genuine high p_T)		
	trk	trk+calo	trk+calo+match
full	4.7	1.5	0.4
$\frac{1}{2}$	27	6	0.4
$\frac{1}{4}$	39	9	3.0

Table 2.2: *Jet rejection performance of the L2 track trigger.*

At $\frac{1}{2}$ -resolution ($300\text{--}400\mu\text{m}$) the global Level 2 trigger performance remains satisfactory, but at $\frac{1}{4}$ -resolution ($600\text{--}800\mu\text{m}$) it is inadequate. Hence the assembly precision requirement from the Level 2 track trigger is that the silicon barrels be placed within about $300\mu\text{m}$ of their correct positions.

2.3 Alignment techniques and technologies

There are three basic types of alignment technique

- charged particle tracks
- non-optical position sensors
- optical metrology

It will be seen that of these methods, optical metrology is the most attractive for making the high precision measurements needed to determine the locations of element positions in the Inner Detector.

Non-optical methods are not favoured for the alignment of the Inner Detector and will not be discussed here.

In the following sections a few of the techniques which are candidates for various stages of the alignment of the ATLAS Inner Detector are introduced. In some cases the techniques have been suggested for use in other parts of the detector, and their adaptation for use in the Inner Detector is under consideration.

2.3.1 Alignment using charged particle tracks

Historically the run-time alignment of particle detectors has been done predominantly using the detected tracks themselves for lining up detector elements. However, in many cases this has been a very slow process. In addition, it is possible that there could be significant movement of the detector elements during the integration time required to accumulate enough data for an alignment using tracks.

The following is a simple calculation of the time required to collect enough track data to be able to perform a track-based alignment.

Consider a module with dimensions 12 cm by 6 cm ($\Delta\eta = 0.2$, $\Delta\phi = 0.1$) at a radius of 60 cm from the beam pipe. The total charged particle fluence through such a module is about 0.02 tracks per proton-proton event. If 100 tracks are required per module for an alignment run, then, assuming a Level 3 trigger rate of 10 Hz , at low luminosity (i.e.

1 collision per trigger) one obtains the figures given in Table 2.3 for the time required for a low- p_T track alignment as a function of the p_T threshold. It can be seen that even for low-momentum tracks, the collection times are long.

p_T threshold	% tracks over threshold	100-track alignment time
0.5 GeV/c	44%	20 mins
1.0 GeV/c	14%	60 mins
2.0 GeV/c	2.9%	280 mins

Table 2.3: *Data acquisition time for a track-based alignment vs. p_T threshold of the tracks. The data collection time grows exponentially with the p_T threshold. Calculated using GENCL.*

Due to the high amount of material in the Inner Detector, it is possible that a precise alignment could only be done using tracks with momenta of at least 40 GeV where the effect of multiple scattering becomes small. Since the track multiplicity falls exponentially with momentum, the collection times required for even a first pass alignment are likely to be very long. This and the complexity of the data analysis required for track-based alignment make it attractive to examine direct ways of determining the positions of detector elements.

2.3.2 Optical metrology

2.3.2.1 Frequency-Scanned Interferometry (FSI)

Frequency Scanned Interferometry is a novel measurement technique which is currently the baseline for the SCT run-time alignment system. The basic principles, the detailed design and laboratory test results are presented in Chapter 3.

The technique requires that the Inner Detector be instrumented with a large number of specially-designed interferometers. These novel interferometers are constructed from two millimetre-dimensioned objects and have low volume, very low mass, no moving parts, no active components and require no careful adjustment. Pairs of single mode optical fibres are used for remote laser illumination and for remote detection of the interference fringe signal. This means that all the complex optical components are located at the surface facility where they may be accessible.

The absolute distance between the two components of the interferometer is measured by counting interference fringes seen at the detector whilst the optical frequency of the laser is scanned over a measured range. The measurement is directly referenced to a length standard. A single laser serves many interferometers and many measurements are made simultaneously. The precision is expected to be of the order of $1\text{ }\mu\text{m}$ for a measured length of 1 m.

The one-dimensional length measurements are combined using a highly overconstrained, three-dimensional geodetic network. In addition to permitting the calculation of the locations of the nodes in three dimensions, a highly redundant geodetic network allows internal cross-checking of the measurement data in order to identify measurement biases caused by measurement mistakes (such as miscounted fringes) and localised gas refractive index changes. The performances of geodetic networks suitable for alignment of the SCT barrel using FSI are investigated in Chapter 4.

Clearly the alignment of a tracker as complex as the ATLAS Inner Detector requires the reduction of a very large number of degrees of freedom. Finite element analysis (FEA) and laboratory measurements of the behaviour of modules and barrels under thermal and mechanical loads will provide a parametrisation of their static deformation modes. By monitoring the movement of a number of carefully selected reference points in the Inner Detector it will be possible to infer the positions of individual detector elements by interpolation.

2.3.2.2 Electronic Speckle Pattern Interferometry (ESPI)

When an optically rough surface is illuminated with coherent light from a laser, the resulting optical field is a complex yet coherent pattern known as a speckle pattern. **Electronic speckle pattern interferometry (ESPI)** or **TV holography** is a technique for electronically extracting information contained in a laser speckle pattern in order to measure microscopic motions of macroscopic objects [10]. Two speckle patterns are recorded using a CCD camera: one before distortion of the object under study and one after distortion. The two patterns, which individually are apparently featureless, when electronically subtracted yield a fringe contour map of the distortion of the object. Each fringe corresponds to motion of half a wavelength either out of the plane of the object or in-plane motion in a direction defined by the setup of the interferometer. See Figure 2.1.

The main application of ESPI is during the design process for analysis of the be-

behaviour of the materials and structures which will be used in the Inner Detector and for development and tuning of finite element analysis (FEA) models. It is currently being used to analyse the behaviour of silicon detector modules and parts of the TRT support structure under mechanical and thermal loads. The current system is capable of making both in-plane and out-of-plane measurements of the deformation of objects with dimensions from a few centimetres square to about 1.8 m square. The precision is better than $0.25\text{ }\mu\text{m}$.

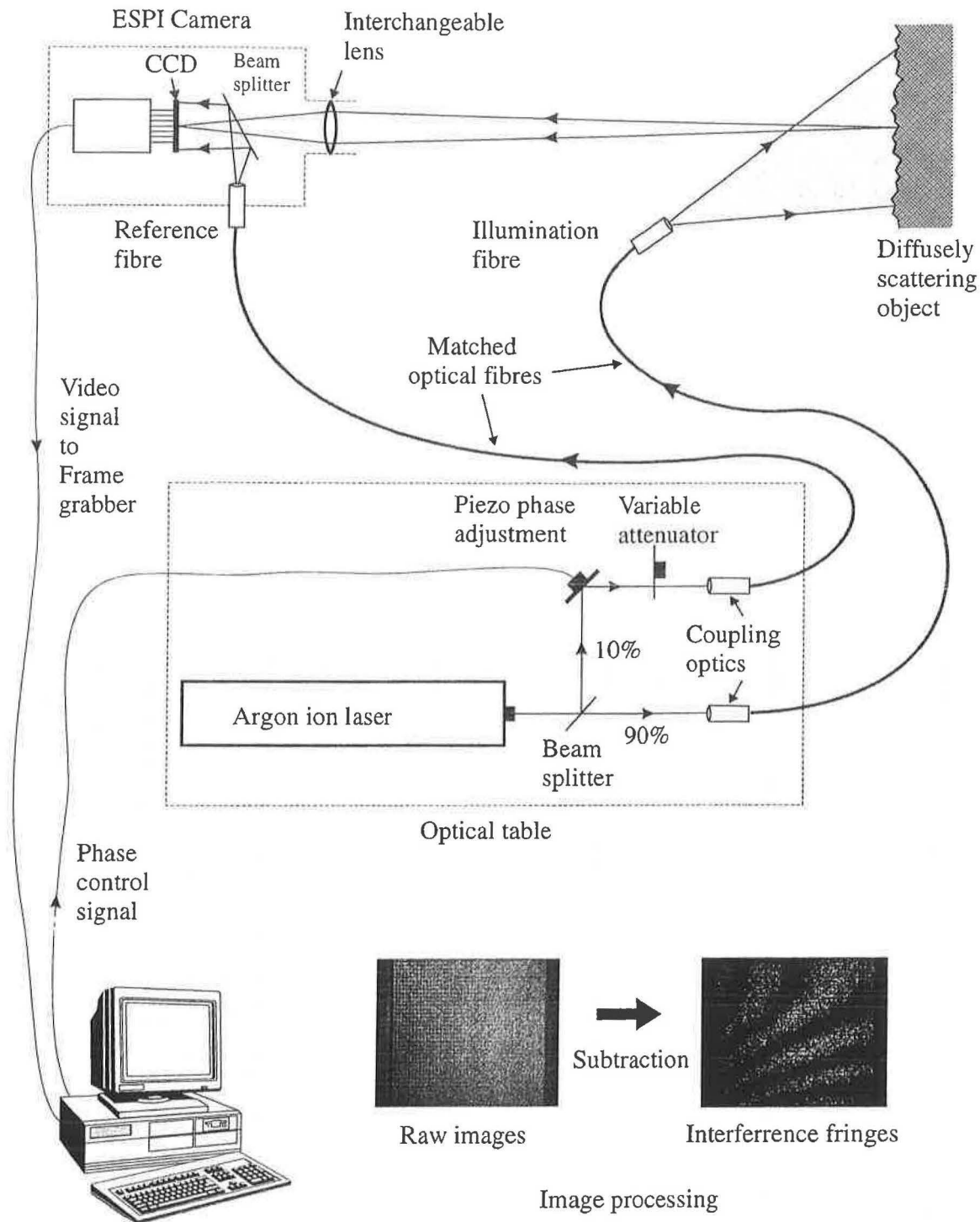


Figure 2.1: An electronic speckle pattern interferometer, set up for sensitivity to motions out of the plane of the object.

2.3.2.3 X-ray alignment

X-ray techniques may be used for a survey just after assembly of the detector. A narrow beam is scanned across the detector and sensed in one of two ways: either by shadowing on to a layer of scintillator placed behind the module under examination, or detection of the X-rays using the detector elements themselves.

A system for a post-assembly survey of the Transition Radiation Tracker has been demonstrated [4][5]. A finely collimated ($10\text{--}50\mu\text{m}$) monochromatic X-ray beam (36 keV) is scanned across a TRT cell and is detected using the TRT straws themselves. It has been shown that this technique can be used to measure both the positions of the anode wires in space and the positions of the wires relative to the tube walls. The position of the anode wires is measured with a precision better than $1\mu\text{m}$.

An X-ray survey system for the muon drift tubes using a scintillator detection method has been demonstrated. See Figure 2.2.

Work is currently in progress to investigate whether the Semiconductor Tracker could be surveyed with X-rays. In semiconductor detectors the absorption does not vary across a wafer, so the X-rays must be detected using the strips themselves. The low energy of the X-ray photons (a Bremsstrahlung distribution with a maximum photon energy of 50 keV) means that a low threshold is needed, and the narrow collimation means that the photon rate is low (10^6 s^{-1}), so the first tests will focus on checking that the signal-to-noise ratio will be adequate. If the initial proof-of-principle is successful, post-assembly survey methods will be investigated. It is possible that an X-ray tube could be placed at the beam axis and swept over the SCT using a (θ, ϕ) or a (z, ϕ) scanning device. It may be useful to emit two collimated beams back-to-back.

Although an X-ray technique could be used for run-time surveys of the Muon Spectrometer, it is likely that an X-ray system would require too much space to be permanently installed in the Inner Detector. It is most likely that such a technique would be used for a post-assembly survey of the SCT and the TRT.

2.3.2.4 RASNIK

A **straightness monitor** is a device for measuring the collinearity of three or more objects: if an axis is defined between the first and last object, a straightness monitor measures the transverse distance(s) of the middle object(s) from this axis.

The RASNIK system is a straightness monitoring technique which has been used

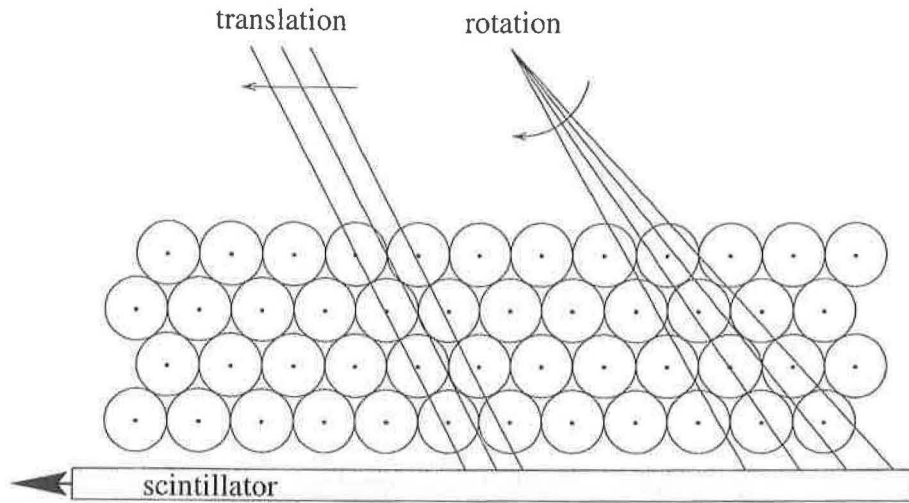


Figure 2.2: *X-ray survey of a drift tube detector using the shadow detection method.*

successfully applied in the L3 experiment [9]. A more developed version, the CCD-RASNIK system, shown in Figure 2.3, has been proposed for use in the ATLAS Muon Spectrometer [6]. It has three components: an illuminated coded mask, a lens, and a CCD sensor. Since movement of the lens by a distance d in a direction perpendicular to the axis defined by the mask and the CCD causes displacement of the image of the mask by a distance $2d$, the transverse position of the lens can be calculated from the image position by means of image processing of a CCD frame. Since only a small section of the mask is seen, the coded non-repeating pattern shown in Figure 2.4 is used to obtain a unique position. The observed magnification gives the longitudinal position of the lens.

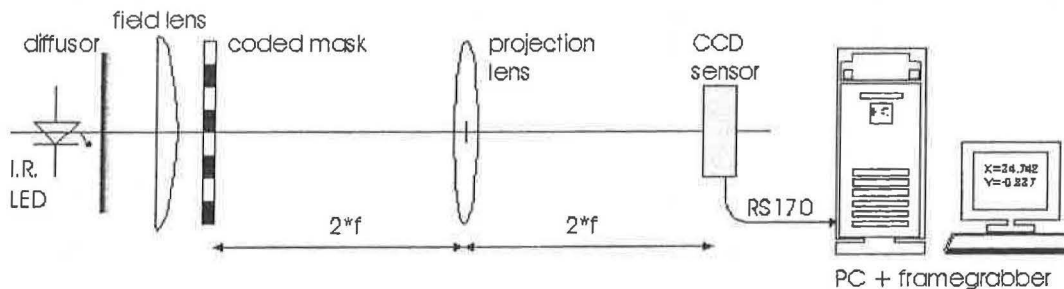


Figure 2.3: *Schematic of the CCD-RASNIK system.*

The resolution of the CCD-RASNIK system in terms of lens displacements perpendicular to the axis is better than $1\text{ }\mu\text{m}$ for a system with a mask-to-CCD distance of 5.5 m. The resolution in the longitudinal direction is better than $30\text{ }\mu\text{m}$. The maximum transverse measurement range of the system is limited only by the diameter of the mask.

Although the RASNIK system is simple and is well tested, the size of the lens (and the material and amount of required space that this implies) is a major limitation in its

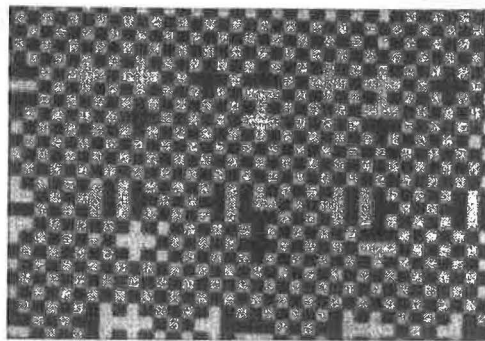


Figure 2.4: *The non-repeating coded mask used in the CCD-RASNIK system.*

application to the alignment of the Inner Detector. In addition, it is unlikely that the CCD would be sufficiently radiation hard for use in the Inner Detector. The techniques which could be used to overcome this (such as fibre bundle detection) are not very attractive.

2.3.2.5 MPI multipoint laser straightness monitor

A multipoint straightness monitor is currently under development at MPI Munich [8]. It has been proposed for use in the run-time survey of the ATLAS Muon Spectrometer [7], and is currently under consideration for use in the Inner Detector. Collimated laser beams are used as alignment references, and are detected using multiple layers of semi-transparent optical position sensors. See Figure 2.5.

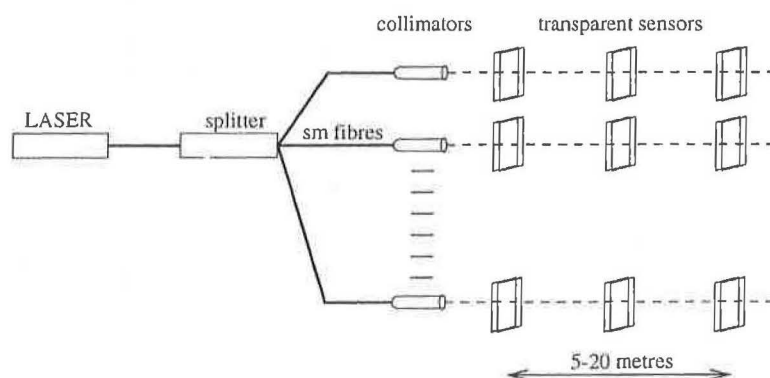


Figure 2.5: *The principle of the MPI multipoint laser straightness monitor using transparent silicon photodiodes. The laser light is distributed using single mode fibres.*

A novel type of silicon strip detector has been developed for this application. This provides very precise and uniform position information with a precision of the order of $1\text{ }\mu\text{m}$ over a wide measurement range. At suitable wavelengths (such as those of common laser diodes), the sensors have an optical transmission of over 90%, which allows more

than 30 sensors to be positioned in a line along one laser beam. See Figure 2.6.

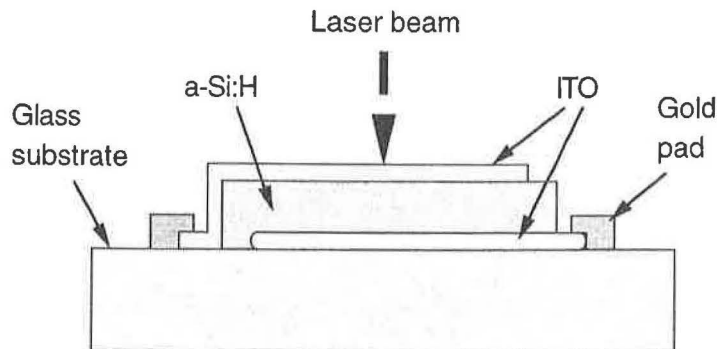


Figure 2.6: *Structure of the amorphous silicon-strip photodiodes. The indium-tin oxide (ITO) layers are segmented into orthogonal strip rows, shown parallel here for the sake of simplicity.*

If the readout requirements and radiation hardness of the silicon photodiodes prove to be satisfactory, the multipoint straightness monitor may be used in the Inner Detector, probably in combination with another technique such as FSI.

2.4 Overview of the Inner Detector alignment process

As explained above, there are at least four distinct phases during which precision metrology will be used for alignment purposes — during assembly of modules and barrels, at assembly of prefabricated units at the Surface Assembly Facility, at installation underground, and during operation of the detector during running. In addition, metrological techniques are required during the design phase for evaluation of the behaviour of materials and support structures under the mechanical and thermal loads expected during operation. The following is a brief outline of current thinking on the overall alignment process for the ATLAS Inner Detector.

2.4.1 Design-phase metrology

The most important metrological technique for evaluating the performance of materials and structures at the design is likely to be ESPI. Its capability in measuring small multidirectional distortions over large surfaces and complex structures is now well established. At the time of writing it has already been used for measuring the distortions of a number of different module designs, and has been used to make unprecedented measurements of the distortions of carbon fibre under the very small loads expected in the TRT support structure. The ESPI facility at Oxford now has planned a vigorous programme of measurement work.

2.4.2 Module prefabrication / preassembly

The assembly of modules and barrels will be done at a number of institutes in Europe, Japan and the U.S.. There are a number of technical and organisational issues which must be resolved before a detailed assembly and alignment procedure can be drawn up. The technical issues include the handling of wafers and modules, choice of temperature and the degree of automation of the assembly process. Suitable technologies for this stage of the alignment include optical microscopy, X-ray and laser systems and FSI.

2.4.3 Surface assembly

The techniques suitable for surveys during the surface assembly are similar to those envisaged for module construction.

2.4.4 Installation

The Inner Detector will be one of the last subdetectors to go underground. Most of the pre-assembly will be done at institutes other than CERN. The final assembly of the prefabricated barrels and wheels will be done in a surface building near the pit. When the solenoid tests have been completed the Inner Detector barrel will be lowered into the pit and slid into the solenoid on rails. At this time a number of techniques will be used to ensure correct positioning, for which optical microscopy, X-ray alignment and FSI are among the candidates. It seems reasonable to expect that the most useful survey would be a full X-ray scan of the fully assembled Inner Detector including services. Whether this would be best done after installation or just before is under consideration.

2.4.5 Running

It is likely that the run-time survey system for the precision layers of the Inner Detector will be based on FSI, with the possible addition of a straightness monitor such as the MPI device. Data from such a survey system would be combined with a relatively small amount of information from tracks and discrete sensors to provide a real-time map of the positions of a large number of carefully selected points within the Inner Detector. Track-based alignments will undoubtedly be attempted, and if successful will offer precise direct measurements of every module, but for the reasons given above they may be slow and limited by the required integration time.

2.5 Conclusions

In this chapter the motivation and requirements for an alignment system for the Inner Detector have been examined. A number of the available and developing technologies which are candidates for various stages of the alignment process of the Inner Detector have been described. An overview of the overall alignment process of the ATLAS Inner Detector has been presented. There remains much work to be done before a detailed alignment strategy can be developed.

References

- [2.1] D.P. Eartly, P.T. Johnson
- [2.2] J. Govingnon, *SDC muon measurement system: conceptual design of the alignment system*, SDC Collaboration Internal Note SDC-92-381, October 1992.
- [2.3] R. Hawkings, *Alignment requirements for the track trigger*, presentation, Oxford, 27 May 1994.
- [2.4] O. Fedin, S. Muraviev, A. Smirnov, *The First Results of a TRT Cell Prototype Alignment Test with 36 keV X-ray Beam*, ATLAS Collaboration Internal Note INDET-NO-057, 27 July 1994.
- [2.5] O. Fedin, S. Muraviev, Yu. Platonov, A. Smirnov, *An X-ray Test Station for ATLAS TRT Calibration: Technical Proposal*, ATLAS Collaboration Internal Note INDET-NO-073, 22 November 1994.
- [2.6] see Section 4.6.2 in the ATLAS Technical Proposal.
- [2.7] see Section 4.6.2.2 in the ATLAS Technical Proposal.
- [2.8] W. Blum, H. Kroha, P. Widmann, *A Novel Laser Alignment System for Tracking Detectors Using Transparent Silicon Strip Sensors*, Max Plank Institut preprint MPI-PhE/95-05, February 1995.
- [2.9] J. Alcaraz, M.I. Josa, J.C. Pinto, *Global alignment of the L3 silicon microvertex detector* L3 Collaboration Internal Note 1724, 23 February 1995.
- [2.10] A. Reichold, *Development and Performance Measurements of a Silicon Tracker Prototype for ATLAS and a Study of Gluino Cascade Decays into same-sign Dileptons using the ATLAS Inner Detector*, PhD thesis, Dortmund University, July 1996.



Chapter 3

Frequency-Scanned Interferometry

here thou viewest, beholdest, surveyest, or seest

— LOVE’S LABOURS LOST

3.1 Introduction

Motivation

Traditionally, once an HEP experiment has been assembled, the positions of the tracking elements have been determined by long and labour-intensive software analyses of large quantities of tracking data. Such processes, besides being difficult and laborious, require the assumption that the detector positions were reproducible over the data collection period.

There are a number of reasons why it is possible that the ATLAS Inner Detector will not be stable over the periods of time necessary for the collection of sufficient data for such track-based alignment calculations. One possible cause of significant movements in the silicon subdetector is the dependence of the power dissipation of the front-end electronics on the occupancy levels and trigger rates, and hence on the varying luminosity of the LHC during a run. Overall, the average temperature of the detector should be held constant by the cooling system, but it is possible that the changes in the temperature distributions between the heat sources (front-end chips) and the heat sinks (cooling pipes) will be large enough to cause significant changes in the shape of the silicon subdetector.

Reasons such as these make it attractive to investigate direct, fast, automatic, remote and computationally simple ways of surveying the positions of critical parts of the Inner Detector during operation. The survey system described here is one possible solution.

Requirements

The essential requirement is for a system capable of producing a 3D map of the positions of strategically chosen parts of the detector in a time short enough for any mechanical or thermal drifts of the detector during that period to be small compared with the measurement precision. The alignment precision must be below about $10\mu\text{m}$ in order to meet the physics specifications (see Chapter 5).

Due to shut-downs and maintenance, continuous operation is likely to be impossible, so a system is needed which can make absolute measurements of position with no *a priori* knowledge. Many existing metrological techniques are only capable of monitoring variations in position and are therefore ruled out by this requirement.

As for any system used in ATLAS, the in-detector elements must have low radiation length, very low mass, small physical size, must be radiation hard, must have no moving parts, must require no maintenance or adjustment during operation, and must communicate with the outside world via wires, pipes or fibres.

The proposed solution

The solution which has been proposed to satisfy the requirements given above is a direct survey system based on a large number of simultaneous interferometric length measurements of the absolute distances between selected points (nodes) in the Inner Detector, each point being common to several such measurements. The 1D measurements are combined using a highly overconstrained, 3D geodetic network for computation of the positions of the nodes in three dimensions. The measurements will be made using a metrological technique known as Frequency Scanned Interferometry (FSI).

The FSI technique requires that the Inner Detector be instrumented with a large number of interferometers. A novel, very low mass interferometer, which will have no moving parts or active components and will require minimal pre-alignment has been designed for the task. See Figure 3.1. The interferometers will be remotely illuminated and the resulting interference patterns remotely detected via pairs of single mode optical fibres, possibly several hundred metres long. One fibre is coupled to a tunable laser located outside the detector, and delivers light to the interferometer. The other fibre

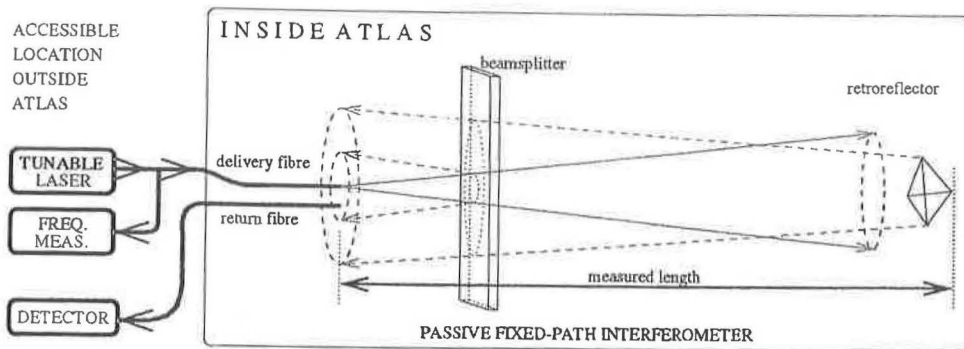


Figure 3.1: *The basic concept of the FSI system.*

is used to collect light for delivery to a photodetector, also located externally. The actual length determination is made by scanning the optical frequency of the laser and calculating the ratio of the change in frequency to the number of fringe oscillations seen at the photomultiplier.

In order to meet the specification of a 3D precision of $10\mu\text{m}$, the 1D precision must be of the order of $1\mu\text{m}$ to take account of the errors associated with the positioning of the measurement endpoints and the triangulation. This precision is determined by the length of the frequency scan and the precision with which the phase of the fringe pattern is measured. The possibility that within the detector there could be mechanical vibrations with an amplitude greater than $\frac{\lambda}{2}$ has led to the development of a technique which should permit both rejection and analysis of vibrations.

The interferometer is designed to be capable of measuring to micron precision the positions of nodes in a 3D network with a shape known to only the millimetre level. To this end it employs a retroreflector (see Section 3.2.7). This minimises the pre-alignment of the optics during assembly of the Inner Detector and ensures that no adjustment is needed during operation or normal maintenance cycles. In each interferometer, the measured distance is between a known point near the closely positioned ends of the fibre pair and the corner-point of a retroreflector mounted some distance away. A single interferometer design is capable of making measurements in the range $\sim 10\text{ cm}$ to $\sim 1.5\text{ m}$.

By using optical splitters, many interferometers can be served simultaneously by one tunable laser and one frequency measurement system. All of this is located outside the detector and is the same regardless of the number of measurements made. Since the external optics is the major cost in the FSI system, the incremental cost of additional measurements is comparatively low, which makes FSI a suitable choice for the present

application in which many (possibly several hundred) simultaneous length measurements will be required.

It is important to note that the desired quantities, namely spatial positions, are measured directly, with no coupling of other variables such as the (non-uniform) B-field. This means that there are no intrinsically preferred directions, which is not the case with track-based alignment techniques.

The control system, the data acquisition and the geometrical analysis will be automated to provide a continuous quasi-real-time survey of the detector with minimal human intervention. Since it makes absolute measurements of distance, the system may be powered down at any time without penalty.

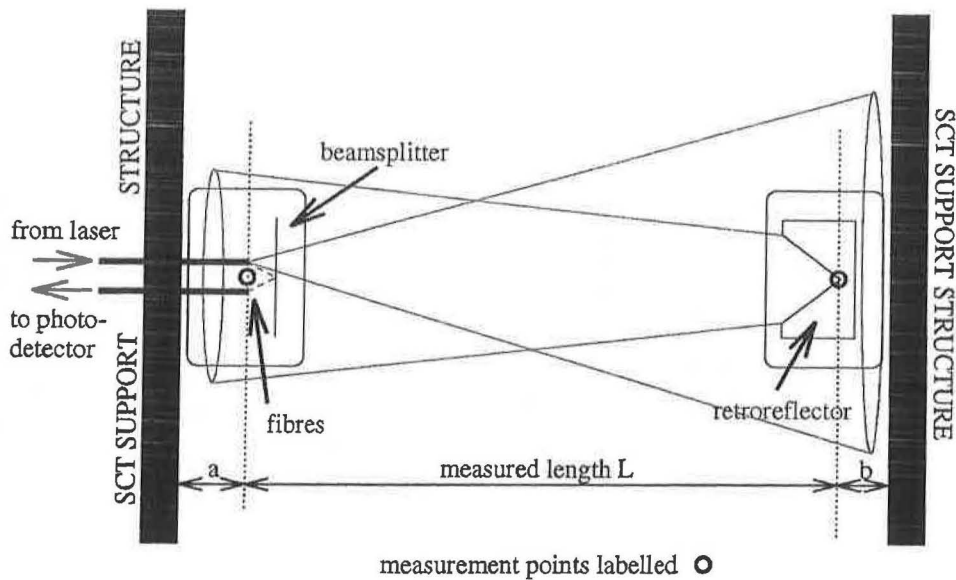


Figure 3.2: A measurement is made between two sections of the SCT support structure. Part of the diverging laser beam is reflected back towards the return fibre by a corner-cube retroreflector. A glass sliver placed in the beam acts as a weak beamsplitter to return a small fraction directly to the return fibre. The fundamental measurement is made between the point midway between the two fibres and the corner-point of the corner-cube. Corrections are applied to relate it to points on the support structure: the pathlength of the reference beam is measured in a laboratory calibration and the lengths such as 'a' and 'b' are measured during assembly.

3.2 Frequency Scanned Interferometry

3.2.1 Introduction



In any interferometer a light beam is divided into two or more parts which travel along different paths and subsequently recombine to form an interference pattern. The

form of the interference at any point in the pattern is determined by the vector sum of the electric fields at that point, and so depends on the amplitudes and phases of the electric fields in each of the interfering beams.

If there are just two beams then the combined intensity $I(\mathbf{r})$ at a point \mathbf{r} is given by

$$I(\mathbf{r}) = I_1 + I_2 + 2\sqrt{I_1 I_2} \cos(\phi_1(\mathbf{r}) - \phi_2(\mathbf{r})) \quad (3.1)$$

where I_1 and I_2 are the individual intensities of the two beams, and $\phi_1(\mathbf{r})$ and $\phi_2(\mathbf{r})$ are the spatial parts of the phases of the two electric fields.

With no loss of generality, the phases of both beams may be taken as zero at the point where they are split from one common beam. The total change in the phase of the electric field over the path of beam i is proportional to the **optical path** \mathcal{D}_i (the equivalent path length *in vacuo*), which may be expressed as

$$\mathcal{D}_i = \sum_{\text{subpaths } j} n_{ij} l_{ij} = \bar{n}_i L_i \quad i = 1, 2 \quad (3.2)$$

where n_{ij} is the refractive index for each section of the path, l_{ij} is the geometric length of each section, \bar{n}_i is the overall path-averaged refractive index, and L_i is the total geometric path length.

The phase difference in Equation 3.1 may now be written as

$$\Phi \equiv \phi_1 - \phi_2 = 2\pi(\mathcal{D}_1 - \mathcal{D}_2) \frac{\nu}{c} \quad (3.3)$$

where \mathcal{D}_1 and \mathcal{D}_2 are the optical paths of the two beams, ν is the optical frequency of the light, and c is the speed of light. Thus the combined intensity depends on the **optical path difference** (OPD) and on the optical frequency of the light ¹.

The phase terms at the heart of the description of any interferometer have the same form, namely

¹Since the wavelength of light changes depending on the refractive index of the medium in which it is propagating, it is more convenient to talk in terms of optical frequency since it is invariant in this respect.

$$\text{phase} \propto (\text{length}_1 * \text{index}_1 - \text{length}_2 * \text{index}_2) * \text{frequency} \quad (3.4)$$

By looking at the the manner in which the interference pattern varies as one of the five quantities in the above equation is changed, one may obtain information about any one of the other four provided that the other three are known. Clearly there are three physical quantities which are measurable with an interferometer (either singly or in combination): length, refractive index and frequency. Commonly it is either the geometric length or the refractive index of one of the paths (arms) which is varied, with the optical frequency remaining constant. If an interferometer is not permitted to have any moving parts and all the refractive indices are constant, i.e. the optical paths are fixed, then the only remaining variable is the optical frequency.

In the past, a commonly used technique for obtaining length measurements from a fixed-path interferometer was to record the detected intensity for a number of different wavelengths and to solve the resulting set of simultaneous equations using the **method of exact fractions** [10][11]. The use of such a technique, known as **multiple discrete wavelength interferometry**, requires some *a priori* knowledge of the length which is to be measured, and is not suitable for the simultaneous measurement of many widely different lengths since the choice of wavelengths depends to some extent on the lengths to be measured.

Recent advances in tunable laser technology have made it feasible to consider the simultaneous, absolute measurement of many different lengths with no such prior knowledge. If the optical frequency in a fixed-path interferometer is varied (scanned) continuously over the range $\Delta\nu$, then the phase Φ also varies continuously, resulting in oscillations in the combined intensity I , with the number ΔN of oscillations or **fringes** being given by

$$\Delta N = \frac{\Delta\Phi}{2\pi} = \frac{\Delta\nu}{c} |D_1 - D_2| \quad (3.5)$$

$$= \frac{\Delta\nu}{c} |\bar{n}_1 L_1 - \bar{n}_2 L_2| \quad (3.6)$$

where ΔN is not necessarily an integer.

In order to extract one of the the geometric paths L_1 , it is necessary to know the values of \bar{n}_1 and $\bar{n}_2 L_2$. Then the measured length is given by

$$L_1 = \frac{1}{\bar{n}_1} \left(\frac{c\Delta N}{\Delta\nu} + \bar{n}_2 L_2 \right) \quad (3.7)$$

If the interferometer is constructed in such a way as to make the reference path L_2 stable *by design* then L_2 need only be determined once by means of a calibration measurement. The required stability may be achieved by choosing L_2 to be much shorter than L_1 so that for mechanical or thermal changes $\frac{\Delta L_2}{L_2}$ is much less than $\frac{\Delta L_1}{L_1}$. This also reduces the sensitivity to changes in \bar{n}_2 . If the error on $\bar{n}_2 L_2$ is negligible then the precision of the measurement of L_1 is given by

$$\frac{\sigma_{L_1}}{L_1} = \left(\frac{\sigma_{\Delta N}}{\Delta N} \right) \oplus \left(\frac{\sigma_{\Delta \nu}}{\Delta \nu} \right) \oplus \left(\frac{\sigma_{\bar{n}_1}}{\bar{n}_1} \right) \quad (3.8)$$

where σ_{L_1} , $\sigma_{\Delta N}$, $\sigma_{\Delta \nu}$ and $\sigma_{\bar{n}_1}$ are the measurement errors of L_1 , ΔN , $\Delta \nu$ and \bar{n}_1 respectively, and \oplus represents summation in quadrature. The errors are discussed in more detail in Section 3.2.4.

It should be noted that multiple lengths may be measured simultaneously by using light from a single laser to illuminate multiple interferometers and recording the detected fringe intensity for each one concurrently. If the the phase measurement accuracy $\sigma_{\Delta N}$ and the required length precision σ_{L_1} are both independent of the length measured, then a single frequency sweep $\Delta \nu$ is suitable for all the interferometers.

Thus the geometric path lengths of one of the arms of each of a set of fixed-path two-beam interferometers may be determined simultaneously by scanning the optical frequency of a single laser and recording the fringes seen at a set of detectors. This is the basic principle behind FSI. The next step is to design an interferometer in which the measured geometric path length L_1 is related in a simple way to a desired length (R say), and to examine whether the required accuracy can be attained, whilst satisfying the constraints imposed by the use of such a technique within ATLAS.

3.2.2 Interferometer design

Since the structure to be measured will be inaccessible and in a hostile environment, it will not be possible to place a laser source, photodetectors or any moving parts inside the detector. The only viable method of constructing interferometers in such conditions is to use optical fibres for the delivery and reception of light. Each interferometer is therefore connected to a pair of radiation hard fibres (single mode for reasons explained below). The laser is coupled into one fibre of each pair, using a series of splitters, whilst the second fibre of each pair returns light from the interferometer to a small photomultiplier or avalanche photodiode. The fibres may be several hundred metres long, meaning that all the complex optics may be accessible and in a non-hostile environment.

Figures 3.1 and 3.2 illustrate schematically the interferometer design which is proposed for use in ATLAS. The measurement is made between two accurately known points on two millimetre-sized objects precisely positioned within the Inner Detector. The first is a mounting block in which the two fibres are fixed parallel to each other a short distance apart with a small sliver of glass mounted in front them to act as a beamsplitter. The second object is a retroreflector (see Section 3.2.7) mounted some distance away on another part of the structure.

There are no lenses in front of the fibres, so the beam emitted from the delivery fibre is conical, as is the acceptance region of the return fibre. The intensity distribution within the laser cone is Gaussian with a usable width of about $\pm 3^\circ$ and a sharp cutoff at about $\pm 5^\circ$. See Appendix 3.C for details of the coupling properties of single mode fibres.

A small fraction of the light emitted from the delivery fibre is reflected directly into the return fibre by the beamsplitter. Most of the light in the laser cone passes through the beamsplitter towards the retroreflector. The retroreflector returns a fraction of this forward-going light cone back towards the fibres, of which a fraction is coupled into the return fibre. Thus there are two paths within the interferometer and hence two interfering beams incident on the return fibre. Detailed calculations of the optical path lengths are given in Appendix 3.A.

It is essential that both optical fibres are monomode. The phase relationship between the different modes of a multimode fibre is not constant, depending strongly on the temperature and mechanical state of the fibre. For this reason, a multimode laser delivery fibre would result in a light field with insufficient spatial coherence. Use of a monomode delivery fibre ensures that there is a constant phase relationship between the light at all points within the interferometer independent of the state of the delivery fibre. Similarly, the return fibre must be monomode. It can be shown that, using a monomode return fibre, the intensity at the photodetector is simply proportional to the intensity at the centre of its input face (see Appendices A and D). Thus the use of monomode fibres ensures that the measured optical path difference is intrinsically independent of the fibres.

The optical efficiency of the interferometer is calculated in Appendix 3.B. If photomultipliers or avalanche photodiodes are used to detect the interferometer fringe signals then the required fringe phase accuracy should be attained with an input optical power of around 10 mW for each interferometer.

It is possible that there may be vibrations of the Inner Detector support structure with an amplitude greater than $\frac{\lambda}{2}$. Immunity to such vibrations can be achieved by utilising a technique known as phase-shift interferometry (PSI). This is explained in detail in Section 3.3. An acousto-optic modulator is used to produce a small, rapid dithering of the laser frequency (in addition to the long, slow, monotonic FSI frequency sweep). Provided that the modulation is performed at a sufficient rate (determined by the maximum vibratory velocity), the phase of the fringe pattern may be determined directly for each point in the FSI scan, which allows analysis and hence rejection of the vibration. See Figure 3.5.

It is essential to examine the geometry of the interferometer in detail in order to find out precisely how the measured optical path length is related to the desired length. It is shown in Appendix 3.A that if R_c is the radial distance from the point midway between the fibres to the corner-point of the corner-cube, n_g is the refractive index of the gas near the beamsplitter, \bar{n}_g is the path averaged refractive index of the gas, n_b is the refractive index of the beamsplitter material and d is the thickness of the beamsplitter, then

$$\mathcal{D}_1 = 2 \left(\bar{n}_g R_c + d \left[\frac{n_b}{\cos \theta_2} - \frac{n_g}{\cos \theta_1} \right] \right) \quad (3.9)$$

where θ_1 is the angular position of the corner-cube with respect to the axis defined by the fibres and $n_g \sin \theta_1 = n_b \sin \theta_2$. See Figures 3.16–3.19.

Note that Equation 3.9 contains two angular terms. These terms account for the fraction of the path \mathcal{D}_1 which is within the (flat) beamsplitter and so they vary according to the angle of the measurement axis with respect to the axis defined by the fibres. It is likely that the angular position of the corner-cube with respect to the fibres will always be known to within $\pm 1^\circ$ (i.e. within ± 1.7 cm at 1 m distance). If $d = 1$ mm, $n_g = 1.0$ and $n_b = 1.5$, then if the position of the corner-cube is imagined to be swept around the arc $R_c = 1$ m, the fractional variation in the optical path \mathcal{D}_1 may be expressed as

$$\frac{\mathcal{D}_1(\theta_1 = 0) - \mathcal{D}_1(\theta_1 = 1^\circ)}{\mathcal{D}_1(\theta_1 = 0)} = 1.6 \times 10^{-7}$$

This ratio rises to 4.6×10^{-7} for $\sigma_{\theta_1} = 3^\circ$. The beamsplitter thickness d will in fact be much less than 1 mm and so the small angular effect may be ignored, giving

$$R_c = \frac{1}{\bar{n}_g} \left[\frac{1}{2} \left(\frac{c \Delta N}{\Delta \nu} + \mathcal{D}_2 \right) - d(n_b - n_g) \right] \quad (3.10)$$

The value of \mathcal{D}_2 is a constant for each fibre/beamsplitter unit, and will be measured in a laboratory calibration so that each measured length R_c may be related to a set of fiducial marks on the outer faces of the two endpoint blocks. These fiducials will be used as positioning guides during the microscopic assembly of the Inner Detector, with the result that each measured R_c will be related to a distance $R_{\text{structure}}$ between two known points on the support structure, i.e.

$$R_{\text{structure}} = R_c + \text{end corrections} \quad (3.11)$$

Given that the (constant) end corrections are known, it only requires knowledge of the path-averaged refractive index of the gas to be able to calculate the distance between two sections of support structure from an FSI scan.

3.2.3 Refractive index

In order to be able to extract spatial dimensions from measurements of optical path lengths in a gaseous medium, an accurate estimate of the path-averaged refractive index of the gas is required. The refractive index of a gas is a function of its temperature and pressure and of the partial pressures of water vapour (humidity), carbon dioxide, and of any other significant gaseous constituent of the local atmospheric environment which is active at the wavelength used. Temperature sensors isolated from the structure will be used to measure the gas temperature at a large number of selected points within the Inner Detector.

A widely used parametrisation [12] for the refractive index of dry air under standard conditions of temperature and pressure (15 °C, 1013.25 mbar) is

$$(n_{\text{std}} - 1) \cdot 10^8 = 8342.13 + \frac{2406030}{130 - \sigma^2} + \frac{15997}{38.9 - \sigma^2} \quad (3.12)$$

where $\sigma = \frac{1}{\lambda_{\text{vac}}}$ is the vacuum wavenumber in μm^{-1} . This formula is correct to within 0.01 ppm over the range $\lambda_{\text{vac}} = 200\text{--}1000\text{ nm}$. In Figure 3.3 this variation is shown over a wavenumber range typical of an FSI scan. The variation is about 0.05 ppm over the whole scan.

The refractive index, n_{TP} at any temperature T °C and pressure P mbar may then be computed using the formula

$$(n_{\text{TP}} - 1) = (n_{\text{std}} - 1) \cdot \left(\frac{100 P [1 + P(61.3 - T)10^{-8}]}{96095.4(1 + 0.003661 T)} \right) \quad (3.13)$$

Further small corrections need to be made for the partial pressures of water vapour and CO_2 . The variation of $(n_{\text{TP}} - 1)/(n_{\text{std}} - 1)$ as a function of temperature is shown in Figure 3.4. This translates to a variation of n_{TP} of about 1 ppm per $^\circ\text{C}$.

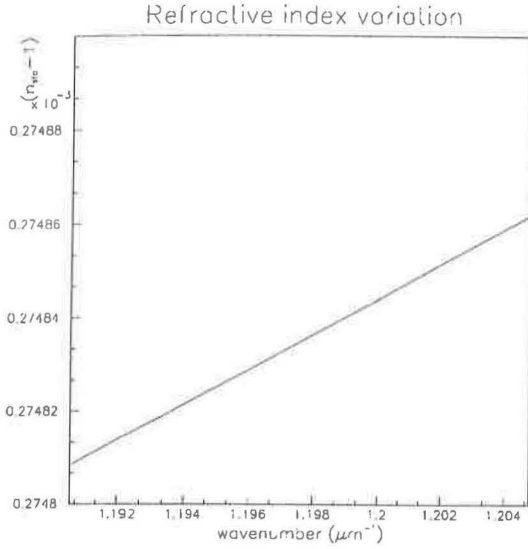


Figure 3.3: Variation of $(n_{\text{std}} - 1)$ with optical frequency over a typical FSI scan length, from $\sigma = 1/0.840\mu\text{m}$ to $\sigma = 1/0.830\mu\text{m}$.

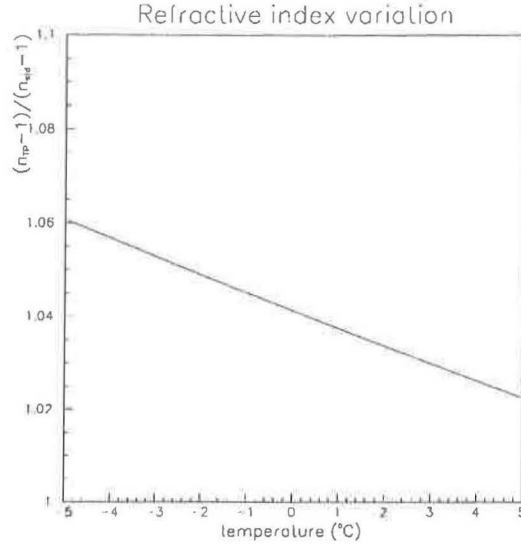


Figure 3.4: Variation of $(n_{\text{TP}} - 1)/(n_{\text{std}} - 1)$ with temperature over the range -5.0 to $+5.0^\circ\text{C}$. $P = 1000\text{ mbar}$.

It can be shown that a bias in the measured optical path difference of 0.1 ppm will be induced by each of the following

- a temperature change of 0.1°C
- a pressure change of 0.3 mbar
- a humidity change of 10 %
- a CO_2 level change of 600 ppm

The figures for a dry nitrogen environment are very similar. Those for helium are better by about a factor of ten.

Pressure variations will be global, at least within each subdetector, but for temperature a distinction should be made between global variations (which affect the overall scale of the measurements but not the measured 3D shape), and local variations (which may affect individual measurements differently).

The path-averaging means that even quite large local temperature changes are not

serious provided that the path lengths through them are small compared with the total path lengths.

It may be possible to distinguish local gas temperature changes from mechanical movement of the inner detector using the 3D geodetic analysis. In the language of Chapter 4, localised heating or cooling will cause a bias in one or more measurements. It may be possible to use the internal redundancy of the geodetic network to identify measurements which have been affected in this way, and hence to exclude them from the analysis. See Section 4.4.

3.2.4 The frequency scan

In order to meet the specification of a 3D precision of around $10\text{ }\mu\text{m}$, it is expected that the 1D length measurement precision must be of the order of $1\text{ }\mu\text{m}$ to take account of the trigonometric combination and the positioning precisions of the measurement endpoints. From Equation 3.8 it can be seen that the precision σ_{R_c} of the length measurement extracted from an FSI scan is a function of the fractional errors of the measured change in fringe order ΔN and of the measured optical frequency change $\Delta\nu$. Assuming that the error on the refractive index is small, the measurement precision is given by

$$\left(\frac{\sigma_{R_c}}{R_c}\right)^2 = \left(\frac{\sigma_{\Delta N}}{\Delta N}\right)^2 + \left(\frac{\sigma_{\Delta\nu}}{\Delta\nu}\right)^2 \quad (3.14)$$

From Equation 3.10 it can be seen that if $d, D_2 \ll R_c$ and $n_g \sim 1$, then for a measured length R_c the number of fringes seen over a frequency sweep $\Delta\nu$ is given by

$$\Delta N \sim \frac{2R_c\Delta\nu}{c} \quad (3.15)$$

Eliminating ΔN from Equation 3.14, the frequency sweep required to obtain a length measurement precision σ_{R_c} is given by

$$\Delta\nu \simeq \sqrt{A^2 + B^2} \quad (3.16)$$

$$\text{where } A = \left(\frac{c\sigma_{\Delta N}}{2\sigma_{R_c}}\right) \quad \text{and} \quad B = \left(\frac{R_c\sigma_{\Delta\nu}}{\sigma_{R_c}}\right) \quad (3.17)$$

If the required measurement precision is $\sigma_{R_c} = 1\text{ }\mu\text{m}$, and if the signal-to-noise ratio of the detected fringes is sufficient to be able to determine the phase Φ to within $\frac{2\pi}{100}$ (i.e. $\sigma_{\Delta N} = \frac{1}{100}$), then $A = 1.5\text{ THz}$, independent of R_c . A represents the lowest possible value of $\Delta\nu$ for a given measurement precision and fringe signal-to-noise ratio. For

reasons of scan time and laser limitations it is desirable to make the scan as short as possible, and so the value of B should be comparable to or less than A .

From the form of B it can be seen that for a given scan length $\Delta\nu$ the frequency measurement requirement becomes more stringent as R_c increases. For the long length $R_c \sim 1$ m, if B is required to be similar to or less than A , then the frequency measurement error $\sigma_{\Delta\nu}$ must be of the order of 1.5 MHz. Thus a reasonable choice of scan parameters is :

$$\Delta\nu \sim 3 \text{ THz} \qquad \sigma_{\nu} \sim 0.75 \text{ MHz}$$

$$\Delta N \sim 20,000 \times R_c[\text{m}] \qquad \sigma_{\Delta N} \leq \frac{1}{100}$$

The fractional frequency measurement error ($\frac{\sigma_{\Delta\nu}}{\Delta\nu}$) is a constant of the optical equipment (i.e. independent of R_c) and is required to have a value of 0.25×10^{-6} or better.

The fractional phase error ($\frac{\sigma_{\Delta N}}{\Delta N}$) may vary with the measured length. The number of fringes ΔN is proportional to R_c , and thus is lower for shorter lengths. The received power will be inversely proportional R_c^2 , and so $\sigma_{\Delta N}$ is expected to decrease for lower values of R_c .

Since the exact dependence of the overall measurement error on R_c requires experimental investigation, for the remainder of this thesis a conservative assumption² will be made that the absolute length error σ_{R_c} is constant at $1 \mu\text{m}$ for all values of R_c .

Tunable lasers are available which will sweep over the required 3 THz, although not continuously : during tuning, the optical frequency of most tunable lasers goes through discontinuities due to **mode hops** between different cavity modes. A scan must therefore be composed of a number of shorter frequency sweeps, or **sub-scans**, which are then linked together using absolute optical frequency measurements. See Figure 3.5. The sub-scans must be close enough in frequency to be able to determine exactly the integer number of fringes between one sub-scan and the next by extrapolation of Φ vs. ν . This use of several short sub-scans is anyway highly desirable since it considerably reduces the amount of data collected and hence the time required for the full scan.

The need to combine several short scans necessitates the measurement of absolute laser frequencies and not just relative frequencies. In order to extract relative optical frequency changes $\Delta\nu$ with an accuracy of 1.5 MHz from the difference of two absolute

²The most optimistic position would be to assume that there are no unforeseen systematic limits and that the signal to noise ratio will improve sufficiently to maintain a constant value for the fractional phase error ($\frac{\sigma_{\Delta N}}{\Delta N}$) and thus a constant fractional length measurement error ($\frac{\sigma_{R_c}}{R_c}$).

frequency measurements, the precision of the latter must be around 0.75 MHz.

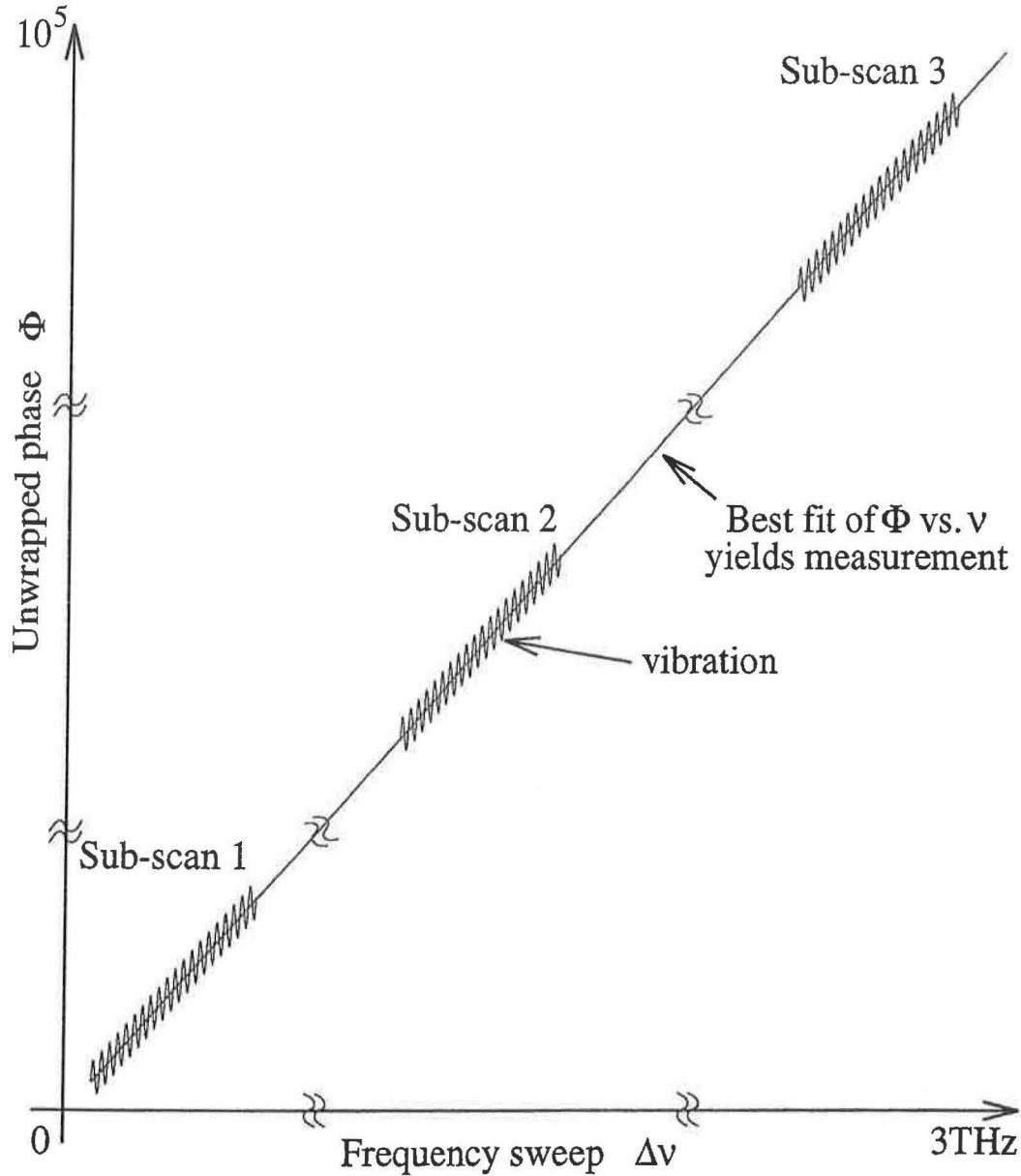


Figure 3.5: Several sub-scans are needed to cover the required scan range. If the **unwrapped phase** Φ (see Section 4) is plotted against laser frequency ν then the gradient of the best fit line yields the OPD and hence the measured length. A single frequency component vibration is shown, appearing as a sinusoidal deviation from the best fit straight line. For a 1.0 m measured length $\Delta\Phi = 2\pi \times 20,000$ and $\Delta\nu = 3 \text{ THz}$.

3.2.5 Tunable laser selection

The choice of laser is determined by the required frequency stability, power, and tuning capability. The frequency stability must be less than the frequency measurement precision of 0.75 MHz. The optical power required for each individual interferometer is around 10 mW (see Appendix 3.B), so a laser power of the order of 1 W would be suitable. The total tuning range must be greater than 3 THz, with a mode-hop-free tuning range which is as large as possible. An insufficiently smooth rate of change of laser wavelength would result in degradation of the measurement precisions for both the fringe phase and the laser wavelength. One way to achieve smooth tuning is by mechanically scanning the length of the laser cavity using a piezo-electric actuator locked to a variable reference cavity.

Commercial tunable single frequency titanium-sapphire lasers exist which meet the specifications given above. A typical example [44][45] delivers 1 W/2 W optical power when pumped by a 7 W/13 W argon ion laser and is tunable over the range 700–1050 nm. The mode-hop-free tuning range is 30 GHz. The laser cavity is piezo-stabilised to a reference cavity, resulting in a frequency stability better than 100 kHz RMS. A servo-controlled piezo-electric scan provides better than 0.5% scan linearity. The scan is digitally controlled up to 30 GHz, and the scan time can be set digitally from 5–4000 seconds.

Whilst not prohibitive, the cost and complexity of titanium-sapphire lasers has prompted the establishment of a research programme at Oxford to investigate cheaper and more convenient alternatives. Due to considerable commercial pressure, the rate of development of tunable lasers is high. It is therefore likely that the performance and cost will improve before the final laser selection for the ATLAS FSI system has to be made.

3.2.6 Optical frequency measurement

As explained above, the FSI scan is composed of a number of sub-scans linked together using measurements of absolute optical frequency.

A commonly used technique for measuring the *absolute* optical frequency of a laser is to compare the unknown laser wavelength with that of a known laser using a two-beam interferometer in which the path difference between the two arms is varied periodically, typically by several centimetres several times a second. The ratio of the rate of fringe

detection for each of the two sources is equal to the ratio of the two wavelengths. Such a device is known as a **wavemeter**. Commonly available wavemeters have a maximum accuracy of about one part in 10^6 and typically give a reading several times per second. Higher precision is possible at the cost of a longer measurement time. The reference laser is typically a frequency-stabilised He-Ne laser which is integrated with the interferometer into a single benchtop unit.

Very precise measurements of *relative* optical frequency may be obtained by using a **Fabry-Pérot étalon** [11]. This is an interferometric device constructed from two partially reflecting optical flats which are mounted parallel a short distance apart to form an optical cavity. The interference caused by multiple reflections between the inner surfaces results in a transmission pattern which is an infinite set of equally spaced sharp peaks. See Figure 3.6. The optical frequency spacing between the transmission maxima is known as the **free spectral range** (FSR), and is given by $\text{FSR} = \frac{c}{2nd \cos \theta}$ where c is the speed of light, d is the width of the gap between the two glass sheets, n is the refractive index of the material in the gap (often air), and θ is the angle of incidence of the laser beam on the étalon. The **finesse** of an étalon is a measure of the sharpness of the peaks. It is numerically equal to the ratio of the FSR to the **full width at half maximum** (FWHM) of the peaks. For good quality étalons the finesse is generally of the order of 100.

By recording the intensity of the transmitted beam whilst sweeping the optical frequency, and fitting the data to find the positions of the transmission peaks, it is possible to obtain a set of precisely known markers of relative frequency vs. time.

The frequency measurement technique proposed for FSI combines the absolute frequency measurement capability of a wavemeter with the high precision of a set of Fabry-Pérot étalons. Each device is fed with a beam extracted from the main laser beam. The transmitted signals of the étalons are all recorded continuously during the scan. The wavemeter is used for a single stationary measurement of frequency at the beginning and/or end of each sub-scan.

The basic idea is to make use of an étalon (E_1) with a free spectral range (FSR) greater than twice the frequency uncertainty of a wavemeter measurement. If one transmission peak of this étalon is arbitrarily taken as order zero and its frequency is measured with the wavemeter, then the order of any other peak relative to this zeroth peak can be determined with a single wavemeter measurement, provided that the FSR of the étalon E_1 is known sufficiently accurately. Hence the frequency relationship between sub-scans

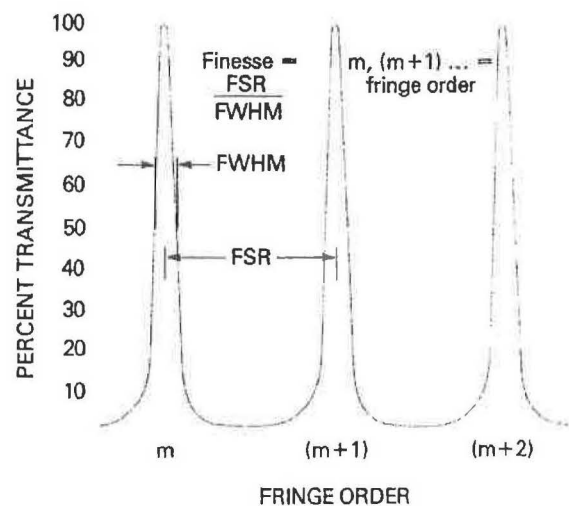


Figure 3.6: The transmission pattern of a Fabry-Pérot étalon. The horizontal axis is light frequency, or wavelength, or fringe order number, or time when frequency-scanning. The **free spectral range (FSR)** of an étalon is the spacing of the transmission peaks, usually quoted in Hz. The **fineness**, \mathcal{F} , is a measure of the sharpness of the peaks and is given by $\mathcal{F} = \frac{FSR}{FWHM}$, where **FWHM** is the **full width at half maximum** of the peaks.

can be determined by using the wavemeter to measure the optical frequency of an E_1 peak in each sub-scan.

A second étalon (E_2) may be used to increase the accuracy and number of the frequency markers within a sub-scan. As for E_1 the transmitted signal is recorded continuously during the scan. The FSR of E_2 is chosen to be more than twice the frequency error which arises in fitting the peaks of the E_1 trace. If the ratio of the FSRs of E_1 and E_2 is very stable and is measured independently to high precision, it is possible to use a fitted E_1 peak to determine the order of an E_2 peak relative to an arbitrarily defined zero order peak in a previous sub-scan. The frequency measurement error is then the fitting error for the narrower E_2 peaks.

The final precision of the FSI measurement is determined by the precision and stability of the last element in the optical frequency measurement chain. Since this device must provide an optical frequency for each recorded point in the scan, it is advantageous to use a device which will provide a continuous reading of optical frequency, as opposed to the periodic frequency measurements which are obtained from étalon peaks. This may be done using a two-beam interferometer by employing a novel instantaneous phase readout technique. The operation of such a **reference interferometer** will be explained in detail in Section 3.4 following the introduction of the basic principles

of its operation in Section 3.3.

It is likely that the étalons and the reference interferometer will be placed in a temperature-controlled and evacuated enclosure to ensure their stability. The reference interferometer may be actively stabilised by locking its length to a small reference laser such as a He-Ne laser, itself locked to a saturated absorption feature of an iodine vapour cell. In this way, the ratios of the cavity lengths will be held constant to within 0.05 ppm or better. The FSRs of each of the étalons will be measured optically, and recalibrated as necessary.

3.2.7 Retroreflectors and jewels

A **retroreflector** is a reflecting object which returns a light ray parallel to its incident direction. One of the best known devices of this type is the **corner-cube** or **cube corner retroreflector** [39, 40, 41, 42, 43]. It is formed from three reflecting surfaces arranged as in one corner of a cube. Any light ray hitting the inside surface of such a cube corner will be reflected up to three times before being returned parallel to the incident ray. A corner-cube may either be constructed from three mirrored surfaces, in which case the reflection takes place externally, or it may be a prism such as the one shown in Figure 3.7, in which case there is total internal reflection. For the purposes of a system intended for use in the ATLAS environment, an externally reflecting corner-cube is preferred. It is expected that this will result in lower optical losses, lower mass, and elimination of any problems which might be associated with loss of transparency due to radiation damage or variation of refractive index with wavelength.

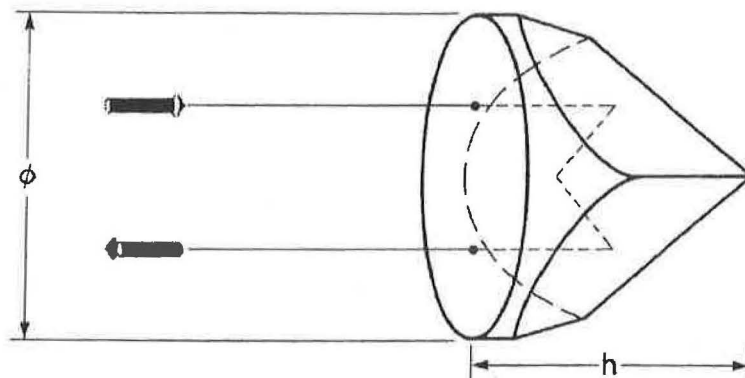


Figure 3.7: A prism type corner-cube. The reflecting surfaces are internal. Externally reflecting corner-cubes are preferred for use in the ATLAS environment.

The geometrical properties of the corner-cube retroreflector are discussed in detail in

Appendix 3.A where it is shown that the use of retroreflectors in FSI makes the system intrinsically insensitive to the exact relative positions and orientations of the end points of the the measurement. This has a number of important advantages :

- No pre-alignment of the interferometers is necessary, and changes in the shape of the detector have no effect on the performance of the optics.
- More than one measurement may be made simultaneously to a single corner-cube. That is, a corner-cube may be the target of more than one laser light cone. If, conversely, there is more than one corner-cube within a laser light cone, then provided that the lengths are sufficiently different, it may be possible to extract more than one length from a single fringe pattern.
- The above features result in a certain amount of freedom in the design of the network nodes since the number of corner-cubes may be less than the number of lengths measured.

As already explained, each interferometer is composed of two basic units, the first being a pair of fibres and a beamsplitter (i.e. a **fibre head**), and the second being a retroreflector. It is intended that several retroreflectors and/or fibre heads will be combined to form a node of the 3D geodetic network. This complex multifaceted object is known as a **jewel**. Two configurations are under consideration : a 'half-half' scheme in which fibre heads and retroreflectors are combined in a single object, and an 'AB' scheme in which each jewel consists of either fibre heads or retroreflectors. See Figures 3.8 and 3.9. A conceptual design for a half-half type jewel is shown in Figure 3.10. The symmetry of the network will be used to keep the number of different jewel designs to a minimum. It is possible that the jewel may be made of reflection-coated injection-moulded plastic or constructed from silicon. The design of geodetic survey networks is discussed in detail in Chapter 4.

A half-half scheme

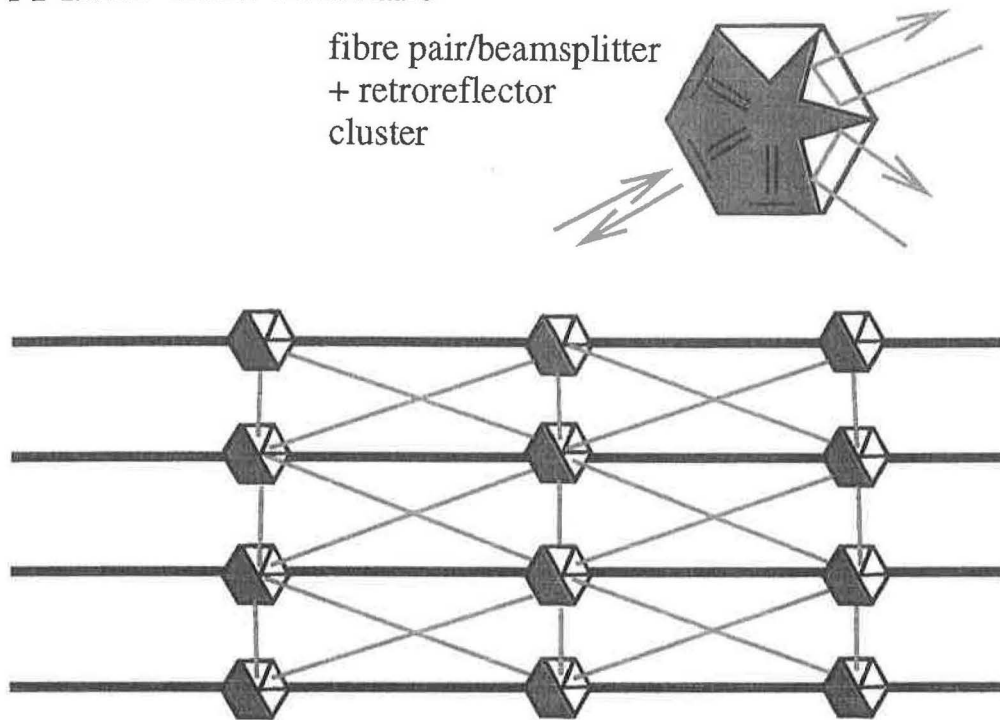


Figure 3.8: A half-half scheme, with a single type of jewel containing both fibre heads and retroreflectors.

An AB scheme

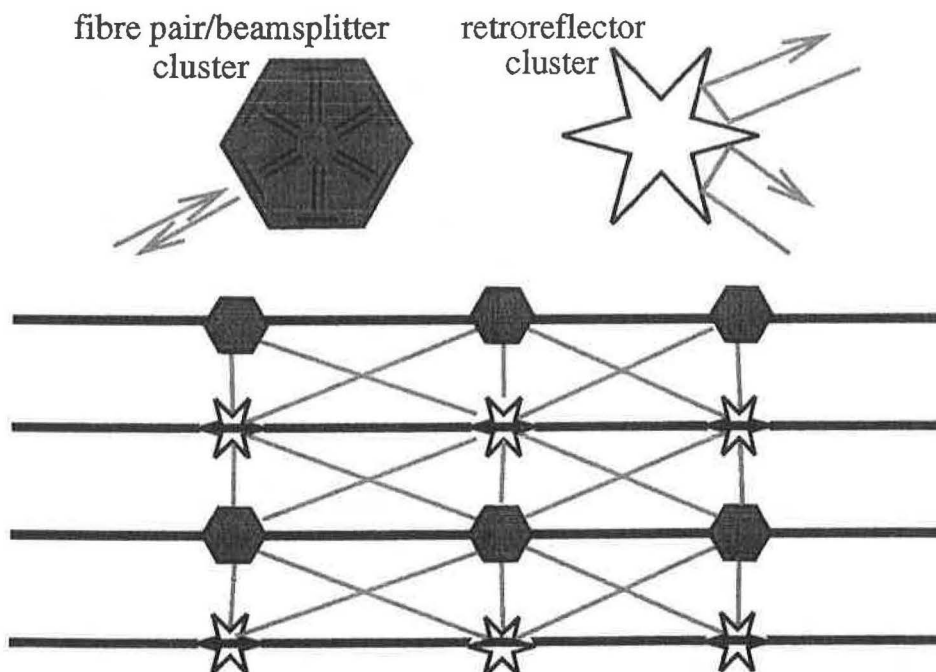


Figure 3.9: An AB scheme, with two basic types of jewels, one consisting of a number of fibre heads, the other composed of retroreflectors.

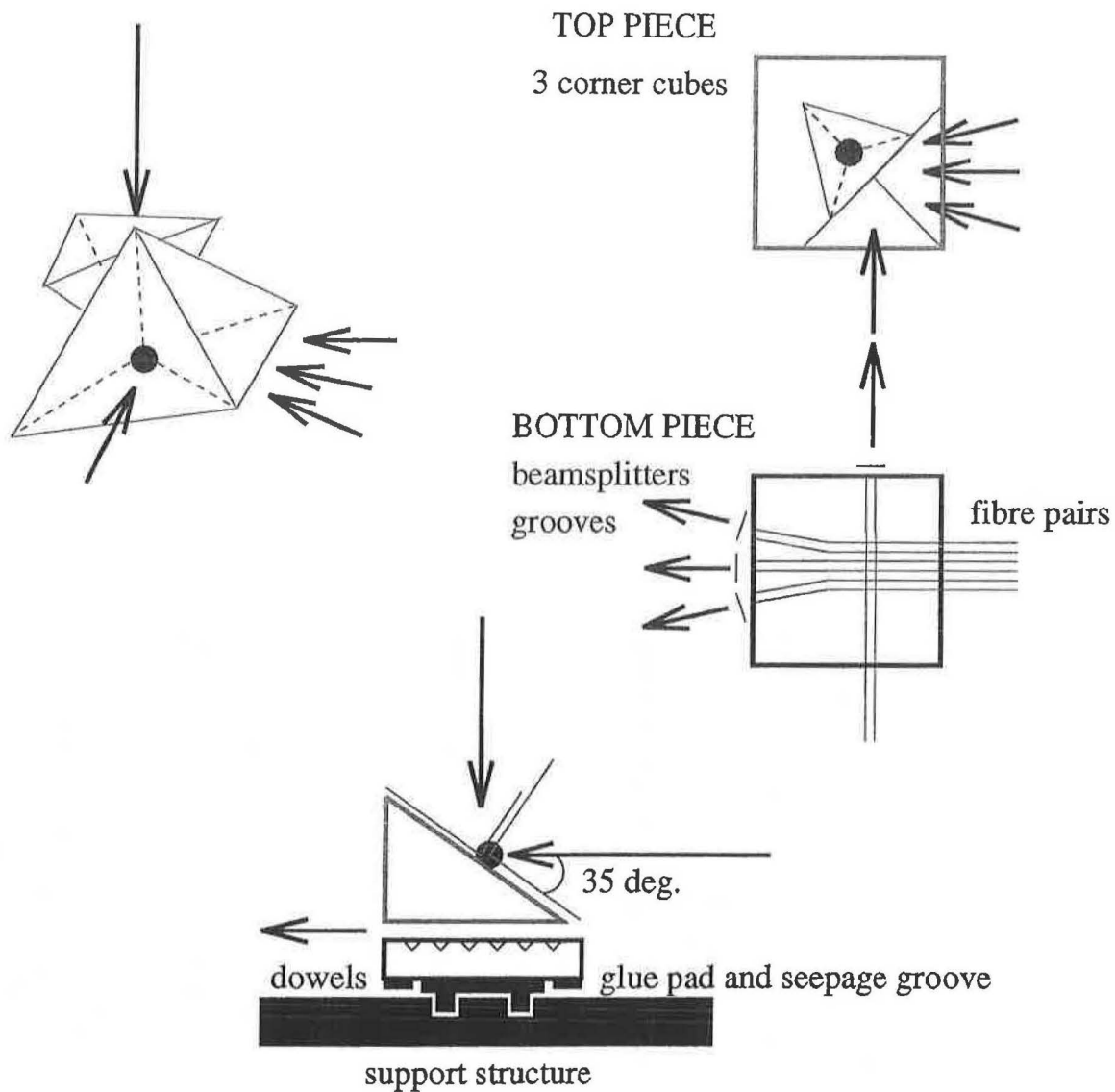


Figure 3.10: *Conceptual design for a half-half jewel. Three corner-cubes and four fibre/beamsplitter units are combined in a manner appropriate to form one node of an end-plane alignment grid in which all nodes are identical. The corner points of the retroreflectors are common, which simplifies the post-scan geometrical analysis.*

3.3 Vibration rejection

It is possible that there could be mechanical vibrations within the Inner Detector during running. With the FSI system as described so far, vibration of the interferometer length with an amplitude greater than half the laser wavelength would result in loss of the fringe signal. Smaller vibrations would reduce the possible measurement precision. The critical dependence of the measurement precision upon the phase accuracy and the possibility that the interferometer signal could be rendered useless by the presence of vibrations with an amplitude smaller than the required length measurement precision have led to the invention of the vibration rejection technique described below. It is designed to provide immunity to vibrations whilst maintaining the phase accuracy. In the sections which follow, the basic principle is introduced and the basic method for extracting the fringe phase from the received signal is described.

3.3.1 The phase-shift technique

In the absence of significant vibration it would be possible to simply scan the laser wavelength monotonically and to record and fit the received fringe intensity in order to extract the phase of any given point. However the possible presence of vibrations larger than $\frac{\lambda}{2}$ has led to consideration of a modified version of a technique known variously as **direct phase detection**, **phase-shift interferometry (PSI)**, **digital wavefront measurement** or **phase-stepping** [65]–[81]. Conventionally the technique involves directly changing the fringe phase ϕ of the received signal I by altering the optical path length of either the reference beam or the measurement beam by a known amount. Since the received fringe intensity varies cosinusoidally with the optical path difference, it is possible to calculate ϕ if I is measured for several known values of $\delta\phi$.

The change in path length is commonly achieved by using a piezo-electric actuator to move a mirror by a fraction of a wavelength. In the present system however, there is no access to either the reference beam or the measurement beam, both being remote and having rigid path lengths. Nevertheless, the phase-shift technique may still be applied by using a small change in the laser frequency to produce the required phase change. It is expected that an **acousto-optic modulator** will be used to shift the laser wavelength by a variable amount under electronic control [70].

The received intensity is given by

$$I(t) = P \left\{ f_b^2 + f_c^2 + 2f_b f_c \cos \left[\frac{2\pi(\mathcal{D}_{12} + 2x_{vib}(t))}{\lambda} \right] \right\} \quad (3.18)$$

where P is proportional to the power delivered into the interferometer, f_b and f_c are the fractions of the incident power returned from the beamsplitter and the corner-cube respectively, \mathcal{D}_{12} is the optical path difference in the absence of vibration and $x_{vib}(t)$ represents any vibration of the interferometer dimensions or turbulence in the gaseous medium. This is more conveniently written as

$$I = I_0 \{1 + \gamma_0 \cos \Phi(t)\} \quad (3.19)$$

where I_0 is the d.c. intensity, γ_0 is the magnitude of the fringe modulation, and the fringe phase $\Phi(t)$ is given by

$$\Phi(t) = \frac{2\pi(\mathcal{D}_{12} + 2x_{vib}(t))}{\lambda} \quad (3.20)$$

Whatever the form of the vibration/turbulence/noise represented by x_{vib} , there exists a mean interferometer phase $\langle \Phi(t) \rangle$, corresponding to a mean optical path difference, and to a mean measured length. The cosine in Equation 3.19 complicates the extraction of the mean phase $\langle \Phi(t) \rangle$, since any technique applied to extract a phase from an interferometer fringe pattern returns not Φ but ϕ where

$$\phi = \Phi \bmod 2\pi \quad (3.21)$$

Simple averaging of the received intensity I or of the phase ϕ is useless since taking the average of the modulus of a distribution will not return the correct value for the average value of the distribution, that is

$$\langle \phi(t) \rangle \neq \langle \Phi(t) \rangle \quad (3.22)$$

In order to extract Φ and hence $\langle \Phi \rangle$, the 2π modulus must be resolved by tracking ϕ across the boundaries where it changes abruptly from 2π to 0 and *vice versa* and correcting subsequent measurements by adding or subtracting 2π . This process is known as **phase unwrapping** and Φ is known as the **unwrapped phase**. Clearly the rate of measurement of ϕ must be sufficiently high for ϕ not to have changed by more than π between measurements. Hence if phase-shift interferometry is to be used, the phase-shifting must operate faster than the vibration or, put another way, for a given rate of phase-shifting there is a maximum possible vibration frequency which can be analysed.

3.3.2 Fringe analysis

The following analysis shows the general technique for extracting the phase ϕ from a set of phase-shifted measurements of fringe intensity I .

If $\delta\nu_i$ is the i th frequency shift in a series of N shifts, then the interferometer phase ϕ and phase-shift $\delta\phi_i$ are given by

$$\begin{aligned}\phi &= \frac{2\pi\nu}{c}(\mathcal{D}_{12} + 2x_{vib}) \\ \phi_i &= \phi - \delta\phi_i = \frac{2\pi(\nu - \delta\nu_i)}{c}(\mathcal{D}_{12} + 2x_{vib}) \\ \delta\phi_i &= \frac{2\pi\delta\nu_i}{c}(\mathcal{D}_{12} + 2x_{vib}) \approx \frac{2\pi\delta\nu_i}{c}\mathcal{D}_{12}\end{aligned}\quad (3.23)$$

So for a measurement range R_c of 1 m a phase shift $\delta\phi$ of π requires a frequency shift $\delta\nu$ of 75 MHz.

Unless truly discrete phase steps are used, the detector will integrate the fringe intensity over a phase shift range Δ . The recorded intensity is then

$$\begin{aligned}I_i &= \frac{1}{\Delta} \int_{\delta\phi_i - \frac{\Delta}{2}}^{\delta\phi_i + \frac{\Delta}{2}} I_0 \{1 + \gamma_0 \cos[\phi + \delta\phi(t)]\} d(\delta\phi(t)) \\ &= I_0 \{1 + \gamma_0 \operatorname{sinc} \frac{\Delta}{2} \cos[\phi + \delta\phi_i]\}\end{aligned}\quad (3.24)$$

If discrete steps are used then $\Delta = 0$ and the sinc function has a value of unity. Consider an alternative representation of this received fringe intensity [66][79]

$$I_i = a_0 + a_1 \cos \delta\phi_i + b_1 \sin \delta\phi_i \quad i = 1, 2, \dots, N \quad (3.25)$$

where

$$\begin{aligned}a_0 &= I_0 \\ a_1 &= I_0 \gamma_0 \operatorname{sinc} \frac{\Delta}{2} \cos \phi \\ a_2 &= I_0 \gamma_0 \operatorname{sinc} \frac{\Delta}{2} \sin \phi\end{aligned}\quad (3.26)$$

The least-squares solution to this set of simultaneous equations is then

$$\begin{pmatrix} a_0 \\ a_1 \\ a_2 \end{pmatrix} = A^{-1}B \quad (3.27)$$

where

$$A = \begin{pmatrix} N & \sum_i \cos \delta\phi_i & \sum_i \sin \delta\phi_i \\ \sum_i \cos \delta\phi_i & \sum_i \cos^2 \delta\phi_i & \sum_i \cos \delta\phi_i \sin \delta\phi_i \\ \sum_i \sin \delta\phi_i & \sum_i \cos \delta\phi_i \sin \delta\phi_i & \sum_i \sin^2 \delta\phi_i \end{pmatrix} \quad (3.28)$$

and

$$B = \begin{pmatrix} \sum_i I_i \\ \sum_i I_i \cos \delta\phi_i \\ \sum_i I_i \sin \delta\phi_i \end{pmatrix} \quad (3.29)$$

Hence the phase may be calculated to within modulus 2π :

$$\tan \phi = \frac{a_2}{a_1} = \frac{I_0 \gamma_0 \operatorname{sinc} \frac{\Delta}{2} \sin \phi}{I_0 \gamma_0 \operatorname{sinc} \frac{\Delta}{2} \cos \phi} \quad (3.30)$$

$$\Phi = \operatorname{ATAN2}(a_2, a_1) + 2n\pi \quad (3.31)$$

where n is any integer and $\operatorname{ATAN2}$ is the FORTRAN function which returns ϕ to within modulus 2π (not just π) by checking the signs of the numerator and denominator. Phase unwrapping then returns Φ by keeping track of n .

Since there are three unknowns ($\{I_0, \gamma_0, \phi\}$ or $\{a_0, a_1, a_2\}$) the minimal number of phase steps which can be used is three. The use of a greater number of steps leads to greater accuracy in the phase determination, or equivalently, the same accuracy for ϕ for a less precise knowledge of the phase shifts $\delta\phi_i$. It is expected that 5–10 steps will be used, arranged in such a way as to ensure three or more useful steps for each of the measurement lengths, which are likely to vary in value by about a factor three.

Once the phase unwrapping has been used to extract Φ from ϕ , values of Φ from the whole FSI scan may be plotted against the laser frequency and fitted to a straight line.

See Figure 3.5. The optical path difference is simply given by the gradient of this line :

$$\mathcal{D}_{12} = | \mathcal{D}_1 - \mathcal{D}_2 | = \frac{c}{2\pi} \frac{d\Phi}{d\nu} \quad (3.32)$$

In this way the effect of vibration is eliminated and the path difference corresponding to the average positions of the optical components is found.

3.3.3 Numerical example

For a vibration with a single frequency component, the displacement x_{vib} may be written as

$$x_{\text{vib}}(t) = \frac{a_{\text{vib}}}{2} \cos(2\pi f_{\text{vib}}t + \phi_{\text{vib}}) \quad (3.33)$$

where f_{vib} , a_{vib} and ϕ_{vib} are the frequency, peak-to-peak amplitude and phase of the vibration respectively. If the intensity signal at the detector is observed, an oscillating signal is detected with the form of Figure 3.11(a) which has an average frequency f_{det} given by

$$f_{\text{det}} = \frac{2a_{\text{vib}}f_{\text{vib}}}{\lambda/2} \quad (3.34)$$

The maximum velocity during the motion is

$$\left. \frac{dx_{\text{vib}}(t)}{dt} \right|_{\text{max}} = \pi f_{\text{vib}} a_{\text{vib}} \quad (3.35)$$

For a vibration with a single frequency component $f_{\text{vib}} = 100 \text{ Hz}$, an amplitude $a_{\text{vib}} = 1.0 \mu\text{m}$, and a wavelength of $\lambda = 800 \text{ nm}$, the detected fringe frequency would be $f_{\text{det}} = \frac{2 \times 100 \times 1.000}{0.800/2} = 500 \text{ Hz}$. Thus a sampling frequency of $f_{\text{sample}} = 5 \text{ kHz}$ ($f_{\text{step}} = 25 \text{ kHz}$) would be sufficient to detect vibrations with a frequency-amplitude product up to around $f_{\text{vib}} a_{\text{vib}} = 100 \text{ Hz } \mu\text{m}$, which corresponds to a maximum velocity of about 0.3 mm s^{-1} .

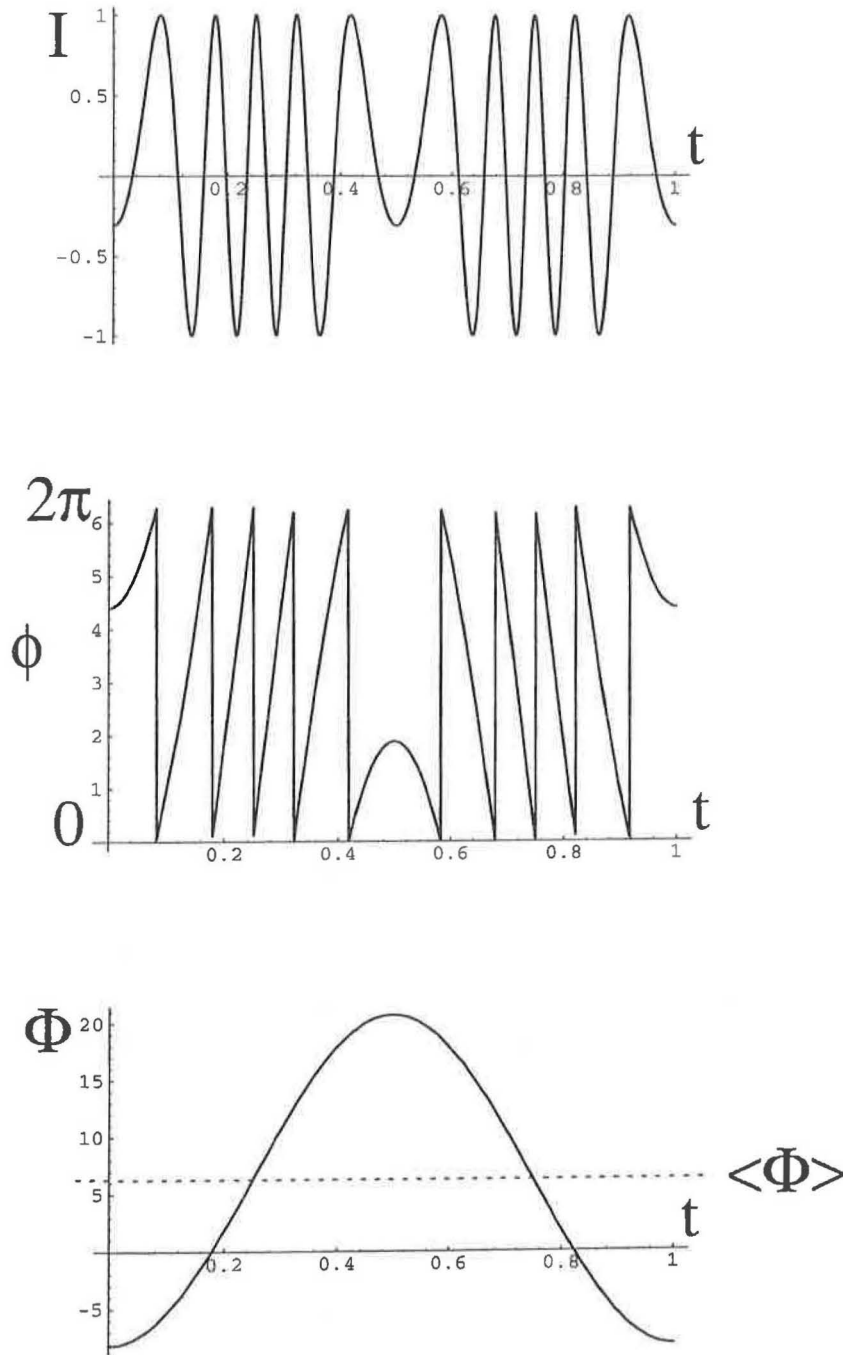


Figure 3.11: Example plots of (a) the fringe intensity $I(t)$ (ignoring the d.c. offset I_0), (b) the fringe phase $\phi(t)$, and (c) the unwrapped phase $\Phi(t)$, for a single frequency (i.e. sinusoidal) vibration. $\Phi(t)$ is related to $x_{\text{vib}}(t)$ by Equation 3.20. The vibratory motion reverses direction every half-unit on the horizontal (time) axis. The peak-peak amplitude of the vibration is 4.6 wavelengths ($4.6 \times 2\pi = 28.9$).

3.4 Reference interferometer with direct phase detection

As already mentioned in Section 3.2.6, the most suitable device for the final element in the frequency measurement chain is a stabilised two-beam interferometer referenced to an absolute length standard. Whereas étalon peaks are very sharp and are separated by large featureless gaps, the signal from a two beam interferometer is a continuous sinusoidal curve which permits a continuous measurement of optical frequency. This is important for the final frequency measurement stage because high precision tracking of the optical frequency will allow reduction of several possible sources of error (laser jitter, modulator behaviour etc.).

Using a direct phase detection technique similar in principle to the frequency-shift technique introduced in Section 3.3, it is possible to extract the phase of the reference interferometer fringes instantaneously with no temporal fringe fitting. Temporal fringe fitting is based on an assumption of continuity in the frequency between samples. Instantaneous phase calculation requires only the assumption that the phase changes by less than 2π between samples.

A Michelson interferometer in which the longer arm has a length of 0.5–1.0 m will be a suitable reference interferometer, the optimal length being of the same order as the measured distance. See Figure 3.12.

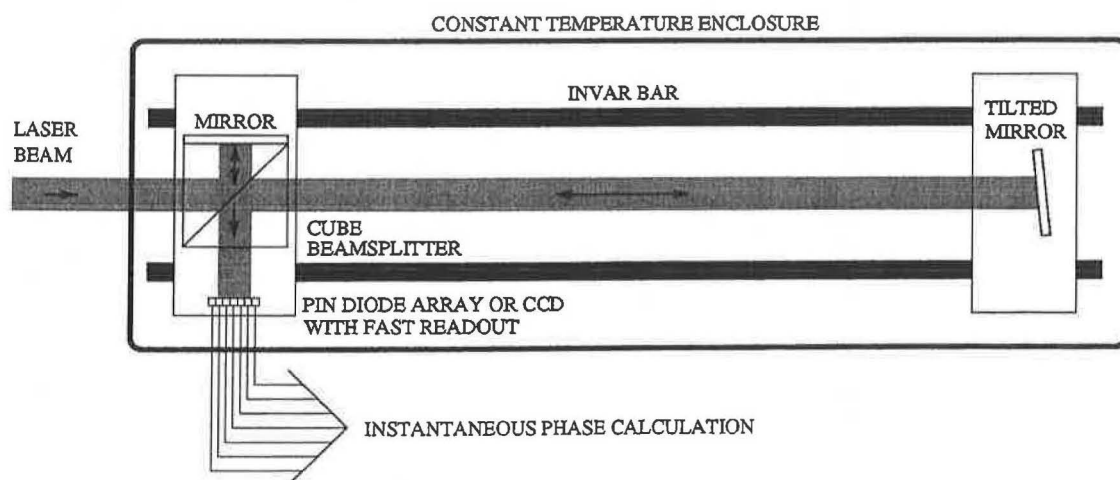


Figure 3.12: Reference interferometer with direct phase detection and fast readout (100–500 kHz). The length of the long arm will be of the order of 0.5–1.0 m.

If one component of the interferometer is slightly misaligned, the measured phase will have a spatial dependence. If several detectors are used to measure the intensity at several different positions, each one will sit at a different point on the sinusoidal

curve of intensity vs. frequency, i.e. there will be a constant set of phase-shifts between the detectors. Using the calculational technique of Section 3.3, a single phase may be extracted mathematically from the set of phase-shifted intensities. Because the phase-shifts are obtained by spatial separation of the detectors (rather than by shifting the optical frequency as in the previous section), the set of intensities is obtained simultaneously. Since the laser beam in the reference interferometer will be collimated and may have a high optical power, the signal-to-noise ratio of the received intensity signals will be extremely good, and will allow a very accurate determination of the fringe phase with a very short integration time. This will permit very high rate monitoring of the optical frequency delivered to the measurement interferometers.

Precise tracking of the optical frequency will be of particular importance when an optical frequency modulator is used for vibration rejection since it will be important for reasons of error reduction to measure the behaviour of such a device as it is stepped or ramped. In the previous section it was shown that in order to be able to measure vibrations with a frequency-amplitude product of up to $100 \text{ Hz } \mu\text{m}$, a step frequency of the order of 25 kHz is needed. Thus readout rates for the reference interferometer of $100\text{--}500 \text{ kHz}$ would allow precise evaluation of the integral of Equation 3.24 for high accuracy in the phase calculation.

A fast reference interferometer will also be useful for analysis of the stability of the laser as it is tuned. The optical frequency stability of a laser is a combination of three distinct components: the natural linewidth of the laser transition (very narrow), drifts of the optical frequency (monotonic and slow), and optical frequency jitter (bi-directional and fast). For this reason the quoted 'linewidth' of a laser is generally accompanied by an integration time to give an indication of the maximum deviation from a given frequency to be expected over a given time period. A fast readout rate would correspond to a short integration time and hence a smaller instability-induced frequency error.

More information may be obtained by using two input beams to the reference interferometer, one directly from the laser and one after the frequency modulator. The beams would propagate side-by-side and would hit different detectors or different parts of a single CCD. Another possibility is that in order to give optimum performance for a range of measured lengths, the reference interferometer could have several mirrors along its length at different transverse positions. Again several input beams would be used.

3.5 Laboratory demonstration system

3.5.1 Introduction

A laboratory demonstration system was built in order to demonstrate the validity of the basic principles introduced in this chapter and to investigate the issues important in the development of an FSI system for ATLAS.

The system was based around a single fully-remote interferometer illuminated with a tunable semiconductor diode laser operating in the near-infrared part of the spectrum. A photomultiplier was employed to detect the interference fringes. Three étalons and a wavemeter were used to measure the laser frequency. The scan control and data acquisition were based on a Pentium PC with a CAMAC interface. A realistic optical power level and data acquisition bandwidth were used. See Figure 3.13 and Plate 3.1.

3.5.2 Remote interferometer

The basic design of the interferometer was discussed in Section 3.2.2. The interferometer used in the demonstration system was intended to be optically similar to the one proposed for use within ATLAS and therefore employed all of the important elements of the device envisaged for the final system (namely a pair of long single-mode optical fibres, a plate glass beamsplitter and a retroreflector) used within similar operational parameters (namely those of range, power level, and data acquisition bandwidth).

Two fibre ends, one from each of two standard 100 m-long single-mode optical fibres, were mounted parallel about 1 mm apart in a metal block which was secured to a vibration-damped optical table. The two free fibre ends were connected to the laser and to the PMT. The half angle of the Gaussian laser cone emitted from the delivery fibre (and conversely the acceptance angle of the return fibre) was about 5° .

The beamsplitter was a 2 mm-thick glass plate mounted 1–5 cm from the fibre mounting block, rotatable about the vertical axis.

The retroreflector was a 7.16 mm prism-type corner-cube mounted at a variable range of between 10 cm and 1.5 m from the fibres. See Plate 3.2.

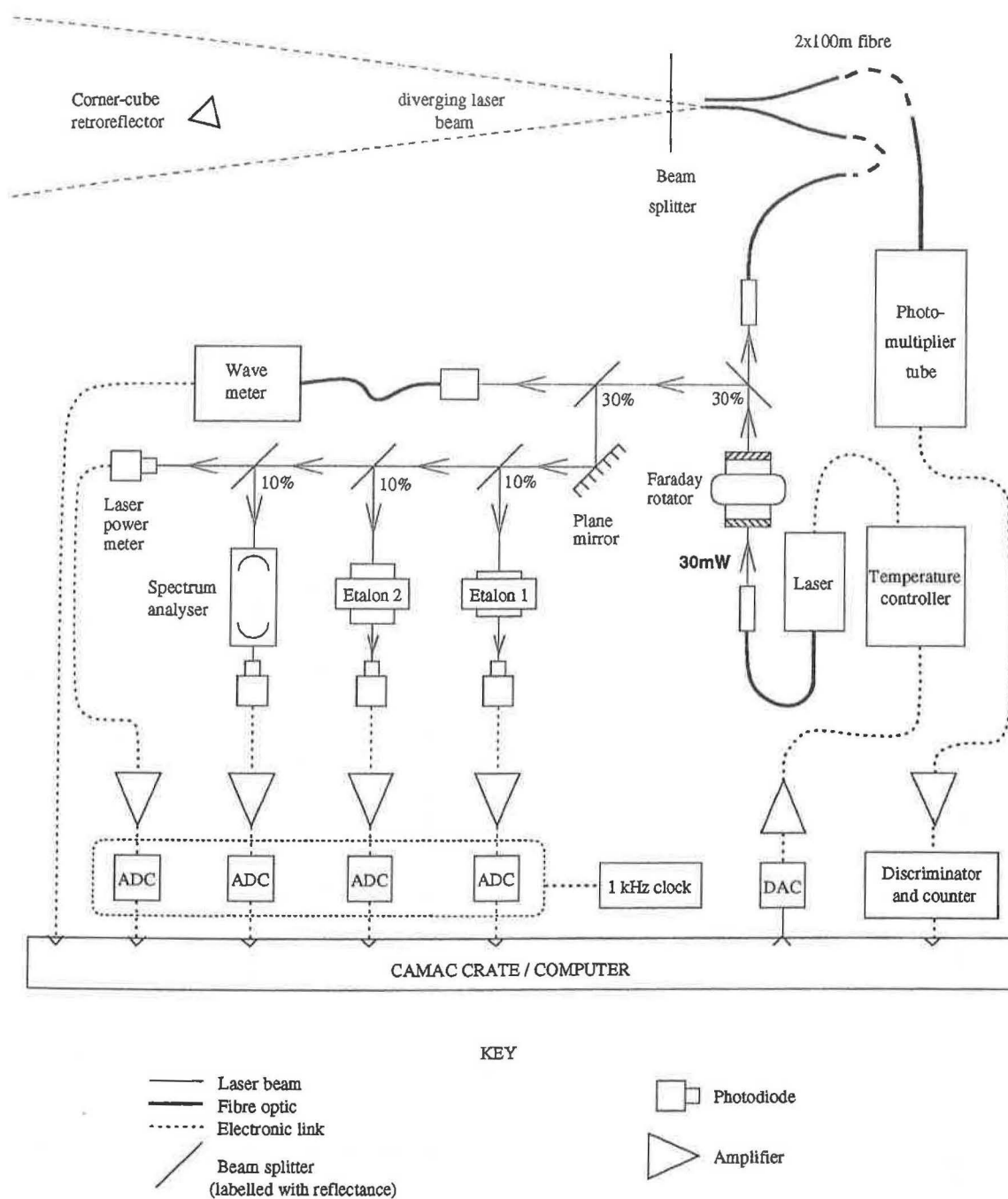


Figure 3.13: The FSI laboratory demonstration system.

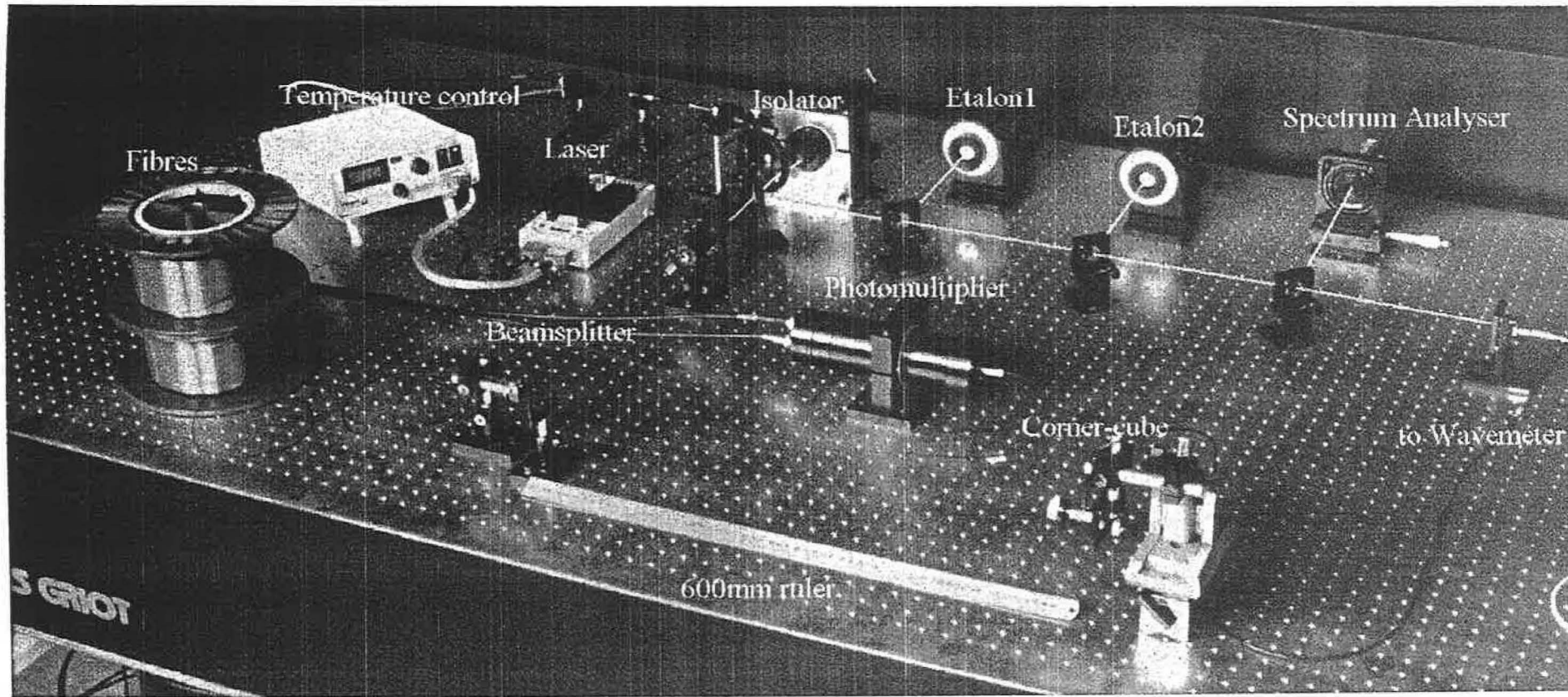


Plate 3.1: The FSI demonstration system. In the foreground is the interferometer, set up to measure a 600 mm length. The two drums to the left hand side each hold 100 m of single-mode optical fibre. The smaller white box contains the laser. The larger box to the left of it holds the temperature control electronics. In the background, from left to right, are the étalons, E_1 and E_2 , and the spectrum analyser, SA (E_3). The wavemeter (out of view) is connected by an optical fibre just visible on the right hand side. The black cylinder in the centre of the table is a light-tight magnetic shield case containing the photomultiplier tube. The laser beam is normally invisible, but has been added electronically to this photograph.

3.5.3 Semiconductor diode laser

The choice of laser was forced by the technology which was available at low cost and within a short period of time. The expense and complexity of a Ti:sapphire laser and its associated pump laser would have been unjustified at this early stage, and since high power was not a priority for the demonstration of a single interferometer, a semiconductor laser diode was chosen.

There are two basic types of semiconductor diode lasers: those based on an optical cavity formed by the cleaved ends of the semiconductor die which contains the laser gain structure, and those in which the optical cavity contains components which are external to the semiconductor. External-cavity diode lasers have the advantage that the tuning is independent of the properties of the semiconductor crystal, and is typically achieved by precise movement of the external cavity components using piezoelectric actuators under microprocessor control. This technology is at a relatively early stage of development, so the laser which was chosen for the demonstration system is of the non-external-cavity variety. Tuning in this case is rather more crude and is effected by simply varying the temperature of the diode and hence, by thermal expansion, the length of the optical cavity. The device selected for use in the demonstration system had a tuning range

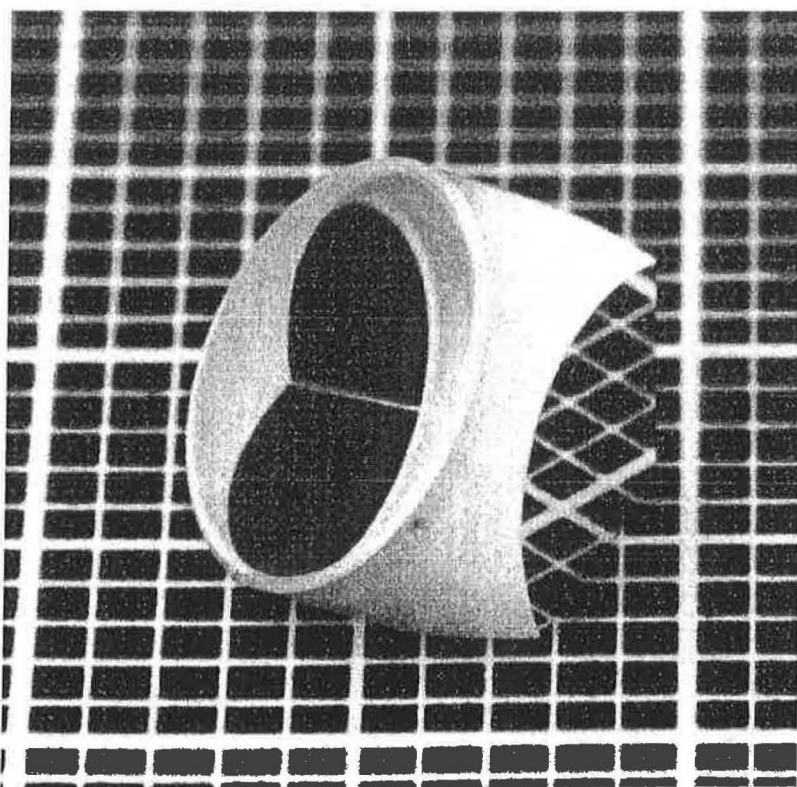


Plate 3.2: The corner-cube used in the demonstration system. The diameter of the front face (LHS) is 7.16 mm. The backing squares have side 1 mm.

of 823–832 nm (5–45 °C), although the mode-hop-free tuning range was much less than this (of order GHz). It delivered an optical power of 30 mW into a fibre — enough to serve the frequency measurement equipment and one interferometer.

A well known problem with semiconductor diode lasers is their hypersensitivity to back-reflected light. In systems in which optical components (for example fibre couplers and confocal cavities) are aligned perpendicular to the laser beam there is often a considerable amount of light reflected back into the laser. This was found to be the case in the FSI demonstration system, so a **Faraday isolator** was placed in the beam. This device rejects light reflected back from the frequency measurement equipment using a Faraday polarisation rotation cell and a pair of polarisers.

It was found that when the temperature was ramped, the laser did not tune continuously but in a series of small jumps which resulted in small steps appearing in the fringes and in the étalon traces. These frequency skips were attributed to interaction of the laser with light reflected back from the glass-air interface at the end of the short fibre attached directly to the laser.

3.5.4 Optical frequency measurement

Three étalons were used to provide high-precision measurements of relative optical frequency. A wavemeter was employed to link these measurements in terms of absolute frequency. The equipment specifications are shown in Table 3.1. It was considered unnecessary to use a reference interferometer for this preliminary stage of the proof-of-principle.

Item	Air gap length	FSR	Finesse	Width of peaks
Wavemeter	—	—	—	eff. 30GHz
Etalon 1	0.6mm	250GHz	100	2.5GHz
Etalon 2	15mm	10GHz	100	100MHz
Spectrum Analyser (E ₃)	eff. 75mm	2GHz	100	20MHz

Table 3.1: *Frequency measurement equipment used in the demonstration system, in order of increasing precision.*

The uncertainty of the optical frequency obtained with each device was required

to be less than half the peak spacing (free spectral range, FSR) of the next. The wavemeter (accurate to 1 part in 10^4 , i.e. ± 30 GHz uncertainty) provided an absolute optical frequency measurement which was sufficiently accurate for the relative order of E_1 peaks (250 GHz apart) in different sub-scans to be determined. Fitting of the E_1 peaks (better than the peak width of 2.5 GHz) then allowed determination of the relative orders of the E_2 peaks (10 GHz apart). Similarly, fitting the E_2 peaks yielded the orders of the E_3 peaks (2 GHz apart). The final frequency measurement precision was the peak-fitting accuracy of the last étalon – better than 20 MHz in this case.

3.5.5 Photodetection

As explained in Appendix 3.B, for a measured length $R_c = 1$ m, a supplied optical power of the order of 10 mW would result in a peak returned power of the order of 3 pW. Detection of signals at this level requires a sensitive detector and so a photomultiplier tube (PMT) was chosen for the demonstration system. It was used in photon counting mode, with a pulse height discriminator and a CAMAC counter module.

The selected device (a Hamamatsu R943-02) had a GaAs(Cs) photocathode rather than one of the more usual multialkali materials, and as a result was sensitive into the infra-red part of the spectrum and not just the visible region. At 850 nm the radiant sensitivity was 65 mA W^{-1} and the quantum efficiency was about 10%, so for an incident optical power of 3 pW (about 10^7 photons sec^{-1}), the cathode current I_k was around 200 fA (1.2×10^6 electrons sec^{-1}).

For a photomultiplier used in photon counting mode (rather than measuring the cathode current), the signal-to-noise ratio is given by

$$S/N = \frac{N_{sig} \sqrt{T}}{\sqrt{N_{sig} + 2(N_{bkg} + N_{dark})}}$$

where T is the measurement time, and N_{sig} , N_{dark} and N_{bkg} are the count rates for the signal, the dark current and the background respectively.

For the chosen PMT, the dark count rate N_{dark} was around 3×10^4 counts sec^{-1} at 20 °C. With this dark count rate, an integration time of $T = 0.01$ s and a signal of $N_{sig} = 10^6$ counts sec^{-1} , the expected signal-to-noise ratio would be of the order of 100. Since the signal rate was much larger than the dark count, there would have been little advantage in operating the PMT at a lower temperature.

3.5.6 Data acquisition and scan control

The PMT signal was amplified and passed to a pulse height discriminator, the output of which was counted with a CAMAC counter module. PIN photodiodes were used to detect the intensities of the transmitted beams of the three étalons. In addition a beam was extracted and fed to a fourth photodiode to provide a signal proportional to the laser power. All four signals were amplified and digitised using a CAMAC ADC module. The wavemeter output was transmitted directly to the PC using a serial data link.

The FSI scan was effected by ramping the set-point temperature of the laser temperature controller under computer control whilst simultaneously recording data from the PMT, the étalons, the wavemeter and the laser level meter. The laser set-temperature was driven by a DAC card in the PC. See Figure 3.13.

The scan control and data acquisition were both synchronised to an external quartz oscillator which was used to gate the ADC module. The PC waited for the CAMAC look-at-me (LAM) signal corresponding to the end of the ADC conversion, at which point the DAC was incremented and the ADCs and the PMT pulse counter were read out and reset ready for the next cycle.

3.5.7 Results

Interferometer

The first tests were concerned with verifying that the interferometer functioned as expected and checking the returned signal power, the fringe visibility and the signal-to-noise ratio.

In order to demonstrate that the received signal was indeed sinusoidally dependent on the length R_c between the fibres and the retroreflector, a piezoelectric transducer driven by a triangular voltage ramp was employed to move the corner-cube repeatedly towards and away from the fibres over a distance of a few microns, whilst the laser was held at a fixed wavelength. The received signal was displayed on an oscilloscope. As expected, the detected signal varied sinusoidally with the range of the corner-cube.

The piezo setup was also used to investigate how the received signal depended on the positions of the corner-cube and the beamsplitter with respect to the fibre mounting block. As expected, the corner-cube could be moved around freely within the Gaussian laser cone resulting in a fringe signal with an amplitude dependent upon the fibre-to-retroreflector distance and upon the transverse position of the retroreflector in the laser

cone. No precise angular positioning of the corner-cube was needed in order to obtain fringes: due to its large angle of acceptance ($\pm 35^\circ$) it could simply be placed in the beam pointing roughly in the right direction.

It was found that the visibility of the fringes could be optimised by rotating the beamsplitter, although good fringes could be obtained with a relatively large range of beamsplitter positions and angles. An interference effect due to reflections from both sides of the beamsplitter was also observed. This caused modulation of the received interference signal with a sinusoidal envelope. The beamsplitter thickness was later chosen to minimise the unwanted modulation.

With an estimated input optical power of 10 mW, a clear fringe signal was detectable up to a range R_c in excess of 1.2 m. An acceptable signal-to-noise ratio could be maintained at sampling rates of more than 1 kHz, depending on the range. See Figures 3.14 and 3.15.

Independence of the interference signal from mechanical deformation of the fibres was demonstrated experimentally. (The polarisation of a beam transmitted through a fibre was shown to be very sensitive to the shape of the fibre. However, as explained in Section 3.2.2, the use of single-mode fibres makes the fringe signal intrinsically insensitive to the fibre shape.)

The received signal was found to be relatively insensitive to the light reflected from objects placed in the laser cone (but not obscuring the line of sight to the retroreflector). It was found that even white objects could be placed in the beam with an acceptably small effect on the received fringe signal.

Operation of the interferometer through a tube was demonstrated successfully: when a piece of 25 mm diameter black plastic pipe was placed between the fibre head and the retroreflector, an acceptable fringe visibility was maintained — indicating that FSI measurements could be made through tubes if required. Light thin tubes or screens could be used to protect lines of sight or to prevent light leakage between one interferometer and another if this proves to be excessive.

System test

Following the successful demonstration of the behaviour of the interferometer, the full system was tested by scanning the laser under computer control whilst recording data in the manner explained in Section 3.6.6.

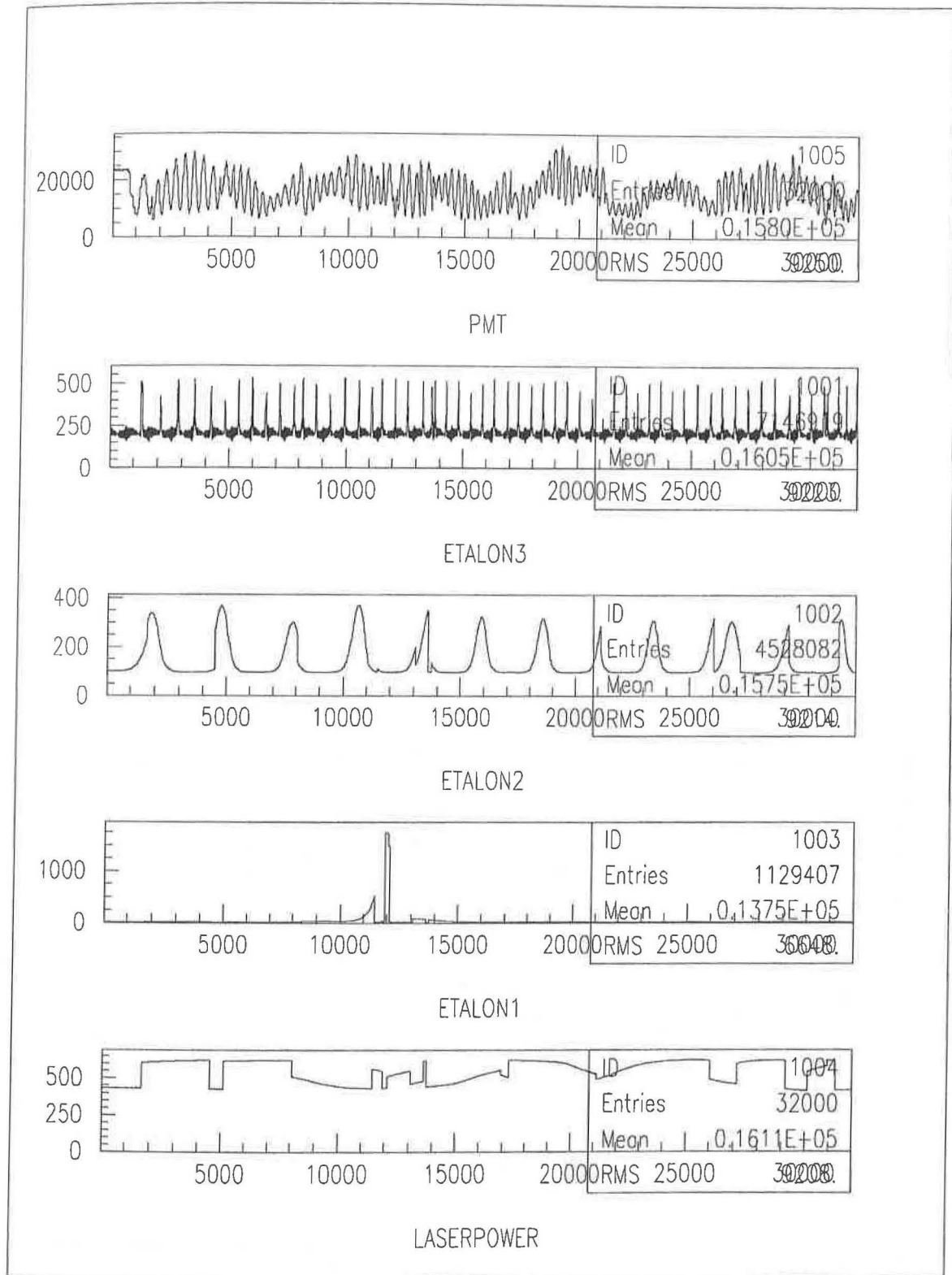


Figure 3.14: Data taken with the demonstration system. The temperature was scanned from 20.500°C to 24.406°C over a period of 320 seconds. Data was digitised at 100Hz. The étalons have been misaligned to reduce the sharpness of the peaks for clarity. The discontinuities in the laser power level correspond to mode-hops — i.e. discontinuities in optical frequency. The measured length was $R_c = 14.1\text{ cm}$.

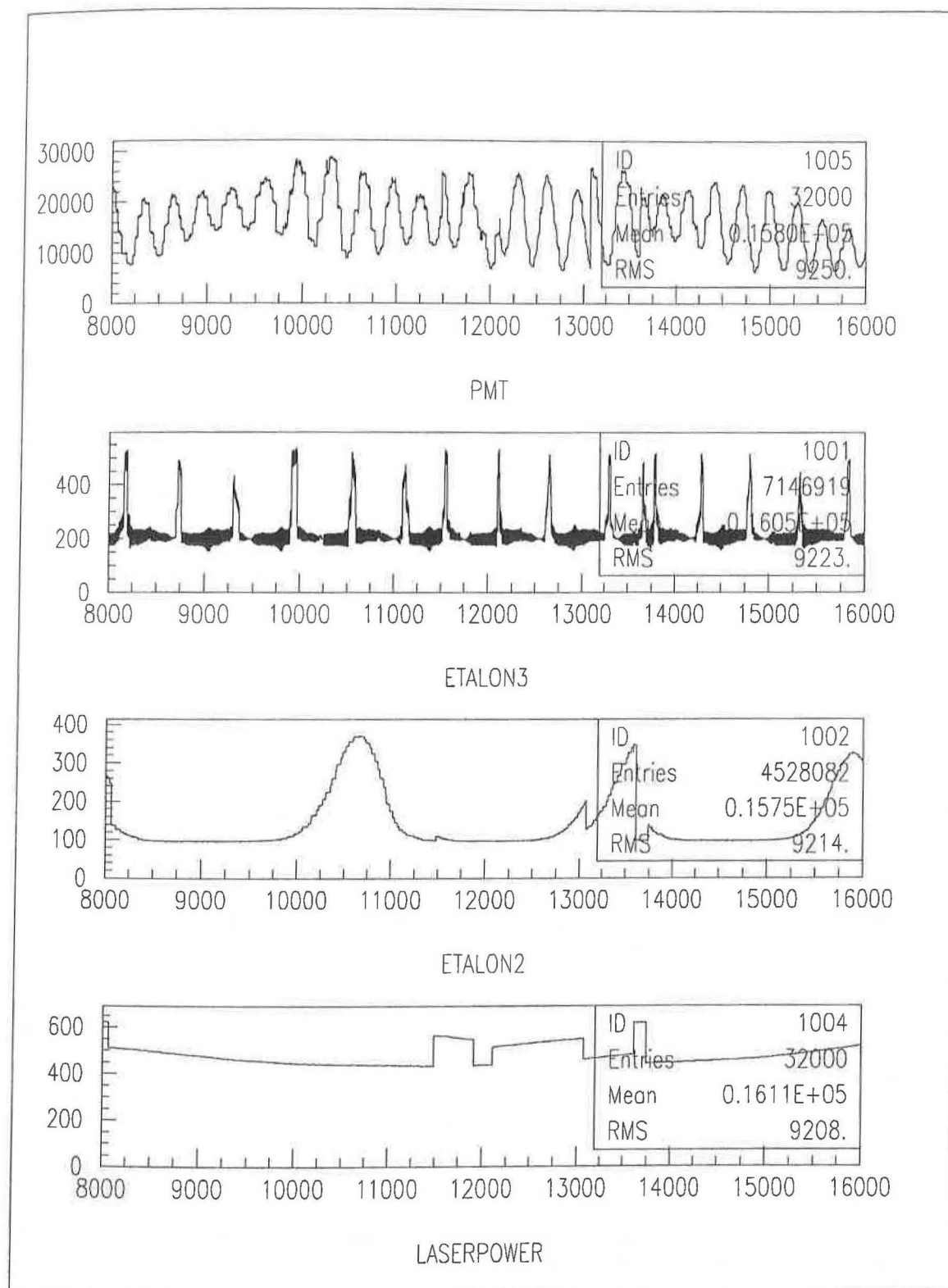


Figure 3.15: A section of the data shown in Figure 14. The small steps visible in the plots (particularly the PMT and ETALON2 traces) are due to a frequency stepping effect of the laser (see Section 3.6.3). The digitisation and the temperature ramp steps take place on a much shorter time scale.

Figures 3.14 and 3.15 show the fringe intensity I (PMT), the transmitted signals for étalons E_1 , E_2 and E_3 , and the laser power, all digitised at 100 Hz over a period of 320 s. The total change in laser temperature during this period was 3.906 °C corresponding to a change in optical frequency of just over 110 GHz. The length measured was $R_c = 14.1$ cm.

It can be seen from the figures that the laser exhibits a large number of mode-hops as it is tuned over a subscan. These are visible as discontinuities in the data which appear simultaneously in all the signals, most visibly in the laser power trace. The small frequency-skipping effect explained in Section 3.5.3 is also visible in the data as very small steps in the traces (see Figure 3.15).

Despite these laser limitations, length measurements were extracted from data such as that shown in Figure 3.14, albeit with a precision unrepresentative of the precision which could be obtained with the same apparatus using laser which did not hop during the subscans. At each mode hop, the phase change in the FSI interferometer takes a random value in the range $[-\pi, \pi]$, with a mean value of zero. By simply counting the number of fringes seen as the the frequency is scanned, and ignoring the mode hops, it is possible to obtain a measurement with a precision of approximately 1%. Lengths in the range 5cm to 50cm were measured successfully in this way. For longer lengths the small stepping of the laser tuning became comparable to the free spectral range of the interferometer.

As explained in Section 3.5.3, the cost of a high-performance laser was not considered to be justified for this preliminary investigation of FSI. There is good reason to expect that when such a laser is used, and it becomes possible to apply all the analysis techniques described in this chapter, a resolution meeting the ATLAS requirement of of 1 part in 10^6 will be attained.

3.6 Conclusions

The survey system presented here is designed to be capable of making high-precision measurements of absolute position of a large number of points within the ATLAS Inner Detector during operation. The underlying measurement technique (FSI) has been described in detail and initial findings from a laboratory-based demonstration system have been presented.

The remote passive interferometer and the fibre-based laser delivery and detection

technique have been tested successfully. Thus correct operation of all the in-detector elements has been demonstrated. The design of the interferometer, the geodetic network and the jewels will be developed as more information becomes available about the requirements and constraints for the use of FSI in ATLAS.

The measurement precision attained with the demonstration system was limited by the semiconductor laser used. However, the results are otherwise extremely encouraging. Although the FSI technique cannot yet be claimed as proven at the required precision, the results presented here allow a degree of confidence that the final desired accuracy will be attained when the laser is upgraded. It is known that currently available titanium-sapphire lasers provide the tuning performance and optical power necessary for the final ATLAS alignment system.

In the near future, the laser used currently in the demonstration system will be replaced by a high-performance external cavity semiconductor diode laser capable of tuning over subscans without mode hops. The frequency measurement equipment will be extended to incorporate a stable reference interferometer of the type described in Section 3.4. Given the otherwise excellent performance of the demonstration system, there is good reason to expect that the upgraded system will permit the attainment of a length measurement precision at the level of 1 part in 10^6 required for the use of FSI in the ATLAS Inner Detector alignment system.

3.A Optical path calculations

It is important that the geometry of the interferometer be examined in detail in order to investigate the factors influencing the value of the measured length. In doing so, one may note an important consequence of the use of single mode fibres: since the core diameter of the fibres is small compared to the wavefront curvature and the angle of incidence of the rays on the fibre ends is small, the phase variation across the surface of the receiving fibre is also small, and so the phase of the signal transmitted down the fibre can be approximated by the phase of the single ray which hits the centre of the fibre end. The same principle applies for both sending and receiving fibres. Hence to calculate the phase of the interference signal it is only necessary to evaluate the optical path lengths for one ray in each of the interfering beams.

In the following sections, the total optical path lengths of light rays in the measurement beam and in the reference beam are calculated. It is assumed here that the corner-cube is of the externally reflecting variety and that the three reflecting faces are accurately plane and at exactly 90° to each other.

Measurement beam

Figure 3.16 illustrates the reflection of a point object in a corner-cube. If α is the angle the incident beam makes with the first reflective surface that it hits, then it can be seen that

$$\begin{aligned} |\vec{IF}| &= |\vec{IC}| \cos \alpha = |\vec{IE}| \cos^2 \alpha \\ |\vec{EG}| &= |\vec{EC}| \sin \alpha = |\vec{IE}| \sin^2 \alpha \\ \Rightarrow |\vec{IE}| &= |\vec{IF}| + |\vec{EG}| \end{aligned} \quad (3.36)$$

So the emerging ray has the same path length as it would have had it continued to F and emerged anti-parallel from the point G . That is, the corner-cube can be said to be equivalent to a reflection from the plane perpendicular to the incident ray which passes through the corner-point C , combined with an inversion of the reflection point through C , i.e. F is translated to G .

Now if A' is the reflection of A in C , it is clear that

$$\vec{AF} = -\vec{A'G} \quad (3.37)$$

Therefore the emerging ray has a direction and phase consistent with it having been emitted from the image point A' , and so the calculation of the total path length of a ray emitted at A and received at B after multiple reflections in a corner-cube requires only the calculation of the single length $A'B$.

$$|\vec{AI}| + |\vec{IE}| + |\vec{EB}| = |\vec{A'B}| \quad (3.38)$$

Figure 3.17 illustrates the use of the corner-cube in the FSI interferometer.

$$\begin{aligned} \vec{OA} &= (0, y_{fib}, 0) \\ \vec{OB} &= (0, -y_{fib}, 0) = -\vec{OA} \\ \vec{OC} &= (R_c, \theta_c, \phi_c) \text{ in spherical polar coordinates} \\ &= (R_c \sin \theta_c \cos \phi_c, R_c \sin \theta_c \sin \phi_c, R_c \cos \theta_c) \text{ in cartesian coordinates} \\ \vec{AC} &= \vec{OC} - \vec{OA} \\ \vec{OA'} &= \vec{OA} + 2\vec{AC} \\ \vec{A'B} &= \vec{OB} - \vec{OA'} \\ &= \vec{OB} - (\vec{OA} + 2\vec{AC}) \\ &= \vec{OB} - \vec{OA} - 2(\vec{OC} - \vec{OA}) \\ &= \vec{OB} - \vec{OA} - 2\vec{OC} + 2\vec{OA} \\ &= -2\vec{OC} \\ \Rightarrow |\vec{A'B}| &= 2R_c \end{aligned} \quad (3.39)$$

That is, if the effect of the beamsplitter is ignored, the total optical path length of a ray which leaves the delivery fibre, which is reflected by the corner-cube, and which enters the receiving fibre is equal to twice the distance between the corner point and the point midway between the two fibres, independent of the angular position of the corner-cube with respect to the fibres, and independent of the separation of the fibres.

Figure 3.18 illustrates the refractive effect of the beamsplitter. The total optical path length ($OP\mathcal{L}$) now involves angular terms related to the distance the light travels inside the beamsplitter :

$$\begin{aligned} x_c &= x_b + d + x_{SC} \\ n_2 \sin \theta_2 &= n_1 \sin \theta_1 \end{aligned}$$

$$\begin{aligned}
\mathcal{D}_1 = \mathcal{OPL}(A \rightarrow (C) \rightarrow B) &= 2(n_1 |\vec{AT}| + n_2 |\vec{TP}| + n_1 |\vec{SC}|) \\
&= 2 \left(\frac{n_1 x_b}{\cos \theta_1} + \frac{n_2 d}{\cos \theta_2} + \frac{n_1 x_{RC}}{\cos \theta_1} \right) \\
&= 2 \left(\frac{n_2 d}{\cos \theta_2} + \frac{n_1 (x_c - d)}{\cos \theta_1} \right) \\
&= 2 \left(n_1 R_c + d \left[\frac{n_2}{\cos \theta_2} - \frac{n_1}{\cos \theta_1} \right] \right) \quad (3.40)
\end{aligned}$$

Reference beam

See Figure 3.19. As is the section above there are two similar triangles in the ratio 2:1, AOM and ABA'' , with common angle OAM . Clearly $A''B = 2OM$.

$$\begin{aligned}
\mathcal{D}_2 = \mathcal{OPL}(A \rightarrow R \rightarrow B) &= n_1 |\vec{A''B}| \\
&= 2n_1 |\vec{OM}| \\
&= 2n_1 |\vec{OA}| \cos(\delta - \alpha) + 2n_1 |\vec{AM}| \cos \alpha \\
&= 2n_1 r_b (\sin \alpha \cos(\delta - \alpha) + \sin(\delta - \alpha) \cos \alpha) \\
&= 2n_1 r_b \cos \theta \quad (3.41)
\end{aligned}$$

Where θ is the angle of deviation of the beamsplitter away from being normal to the fibre axis, $\delta = \frac{\pi}{2} - \theta$, and $r_b^2 = |\vec{AN}|^2 = x_b^2 + y_{fib}^2$, i.e. the distance from the point A to the point where the beamsplitter intersects the fibre axis. The total optical path length of the reference beam is effectively what is measured in the laboratory calibration. It is a constant for any given fibre/beamsplitter unit.

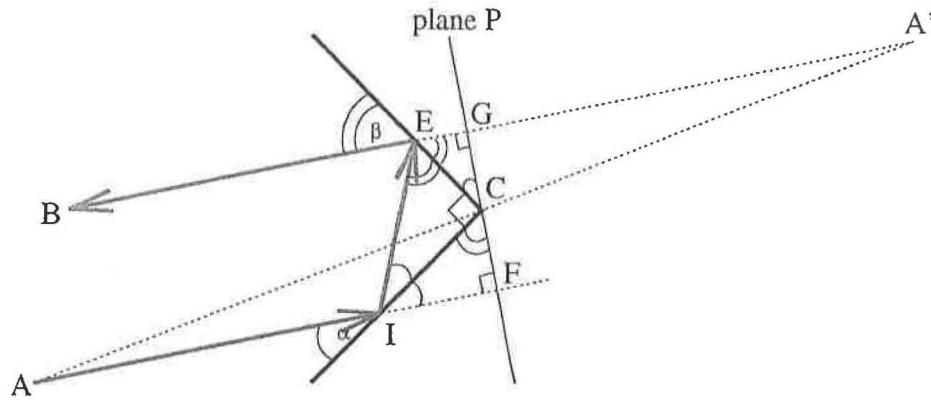


Figure 3.16: Geometry of the reflection of a single ray from a corner-cube. The reflected ray is parallel to the incident ray but reversed in direction. The plane P is perpendicular to the incident and reflected rays and passes through the vertex of the corner-cube C . The path length is the same as it would be had the ray continued to F and emerged anti-parallel from G , i.e. the corner-cube is equivalent to a reflection from the plane P combined with an inversion through C . Alternatively the image A' of A may be used. It is found by reflecting A in C .

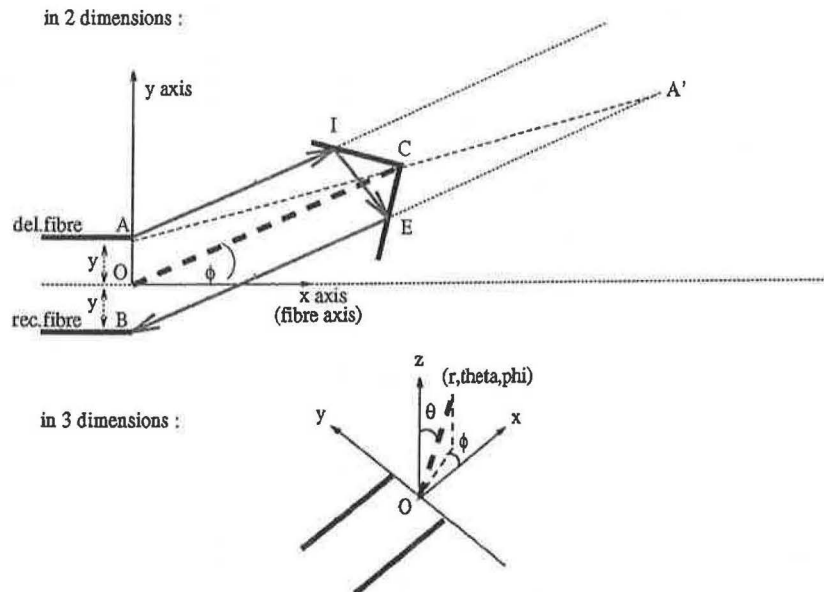


Figure 3.17: Geometry of the reflection of a ray emitted from the delivery fibre at A and hitting the return fibre at B , in the simplified case where the effect of the beamsplitter is considered negligible. The total path length $|A \rightarrow I \rightarrow E \rightarrow B| = 2|\vec{OC}| = 2R_c$ independent of the angular position of C in space and independent of the fibre separation, $2y_{fib}$.

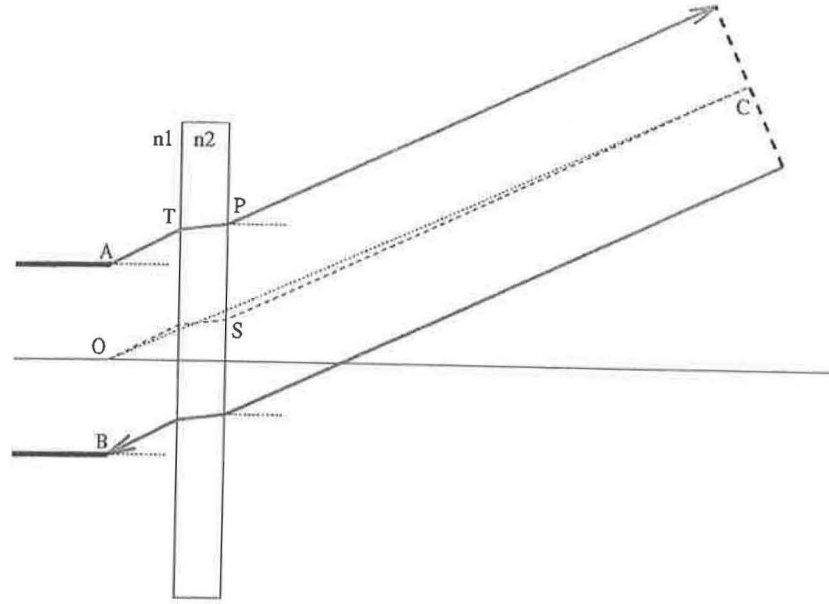


Figure 3.18: Geometry of the corner-cube ray including refraction due to the beamsplitter. The total optical path length $\text{OPL}(A \rightarrow B) = 2 \left(n_1 R_c + d \left(\frac{n_2}{\cos \theta_2} - \frac{n_1}{\cos \theta_1} \right) \right)$ where $R_c = |\vec{OC}|$ and $n_1 \sin \theta_1 = n_2 \sin \theta_2$.

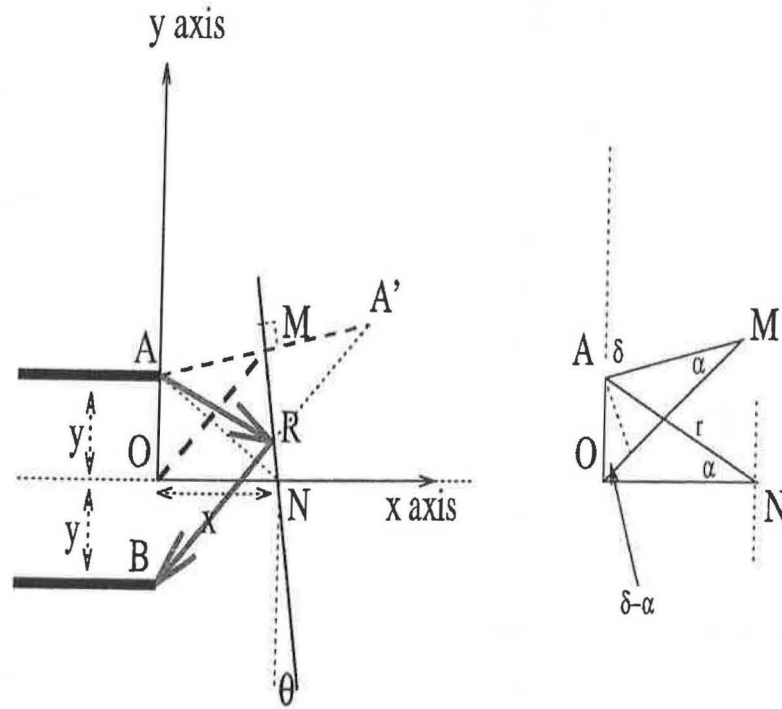


Figure 3.19: Geometry of the beamsplitter ray. The total optical path length $\text{OPL}(A \rightarrow R \rightarrow B) = 2n_1 |\vec{OM}| = 2n_1 r_b \cos \theta_b$ where $r_b^2 = |\vec{AN}|^2 = x_b^2 + y_{fb}^2$ and θ_b is the angle by which the beamsplitter deviates from parallel to the y axis. This is a constant for any given fibre/beamsplitter unit and will be measured during the laboratory calibration.

3.B Detecting the interference signal

The measurement precision depends upon the accuracy with which the phase of the interferometer fringes can be determined. This is a matter of laser power, geometrical efficiency, fringe visibility, and detector signal-to-noise ratio.

It is useful to define a **geometrical efficiency** g for the interferometer as the fraction of the light emitted from the delivery fibre which hits the return fibre, ignoring all non-geometrical losses. The overall optical efficiency is then given by the product of the geometrical efficiency and a factor representing the reflection losses at the boundaries between materials and the losses due to scattering.

The geometrical efficiency may be written as $g = g_1 g_2$ where g_1 is the fraction of the emitted light cone which hits the corner-cube, and g_2 is the fraction of the light returned from the corner-cube which hits the return fibre. The dimensions important in the calculation of g are the core radius a and the numerical aperture $\text{NA} \stackrel{\text{def}}{=} \sin \theta_{\text{beam}}$ of the fibres, the active diameter d of the corner-cube, and the measurement range R_c . The relationships between the numerical aperture, the core radius and the single mode operation of an optical fibre are given in Appendix 3.C.

If the fibre pair is positioned at $x = 0$ pointing along the positive x -axis, and the retroreflector is situated on the x -axis at $x = R_c$, then the circle made by the laser cone in the plane of the corner-cube has a diameter $2R_c \tan \theta_{\text{beam}}$. The fraction of the light emitted by the laser delivery fibre which is returned towards the fibre pair by the corner-cube is given by

$$g_1 = \left(\frac{d}{2R_c \tan \theta_{\text{beam}}} \right)^2$$

where the d is the diameter of the corner-cube aperture.

The light intercepted by the corner-cube is returned to the fibre head as if coming from a point source at $x = 2R_c$ through a circular aperture with the same diameter (d) as the corner-cube situated at $x = R_c$, resulting in a spot with diameter $2d$ in the plane $x = 0$. If the return fibre has radius a then

$$g_2 = \left(\frac{a}{d} \right)^2$$

Hence the overall geometrical efficiency is given by

$$g = g_1 g_2 = \frac{a^2}{4R_c^2 \tan^2 \theta_{\text{beam}}}$$

Note that the corner-cube diameter cancels out: a smaller retroreflector would intercept less of the incident light, but would return it towards the fibres in a proportionately more concentrated beam. Values of g for different measured distances R_c are given in Table 3.2.

Measurement length R_c	Geometrical efficiency g
50 cm	$6.25 \cdot 10^{-10}$
100 cm	$1.56 \cdot 10^{-10}$
150 cm	$0.69 \cdot 10^{-10}$

Table 3.2: Geometrical efficiency, g (defined in the text), for two bare fibres with numerical aperture, $NA=0.1$, and core radius $a = 2.5 \mu\text{m}$

Non-geometrical losses account for about a factor two reduction in the optical power of each of the two beams. See Table 3.3. Maximal fringe modulation is attained when the position and angle of the beamsplitter are such that the returned signal due to the reference beam alone is equal to that of the measurement beam. In this case the fraction of the input power received at the photodetector at a fringe peak is $4 \times 0.49 \times g = 1.96 \times g$. For a measurement range $R_c = 1 \text{ m}$ the overall signal attenuation factor would be $3.2 \cdot 10^{-10}$. So for an injected laser power of 10 mW , the peak received signal power would be 3.2 pW .

3.C Single mode fibres

The numerical aperture, NA , of an optical fibre is given by

$$NA \stackrel{\text{def}}{=} \sin \theta_{\max} = \sqrt{n_{\text{core}}^2 - n_{\text{cladding}}^2}$$

where θ_{\max} is the half angle of a beam emerging from the fibre (or conversely the maximum acceptance angle for an incoming ray), and n_{core} and n_{cladding} are refractive indices. A typical value of the numerical aperture for a single mode fibre is $NA = 0.1$.

The **V-value** or **normalised frequency** of a fibre is given by

$$V = ka(NA)$$

where $k = \frac{2\pi}{\lambda}$ and a is the core radius. It can be shown that the condition for single mode operation is

Item	Efficiency
transmission through 100 m fibre (@4dB/km loss)	0.91
back reflection at fibre end	0.96
beam splitter forward transmission	0.96
delivery fibre to corner-cube geometrical factor	g_1
corner-cube triple-reflectivity	0.70
beam splitter reverse transmission	0.96
corner-cube to return fibre geometrical factor	g_2
back reflection at fibre end	0.96
transmission through 100 m fibre (@4dB/km loss)	0.91
overall efficiency	$0.49 \times g$

Table 3.3: Overall optical efficiency at a measurement length $R_c = 1\text{m}$ for two fibres with numerical aperture $NA=0.1$, and core diameter $a = 2.5\mu\text{m}$.

$$V < V_{crit} = 2.405$$

or alternatively for a given wavelength and numerical aperture

$$a < a_{crit} = \frac{2.405\lambda}{2\pi(NA)}$$

If $\lambda = 830\text{ nm}$ and $NA=0.1$ then $a_{crit} = 3.2\mu\text{m}$. A typical value for the core radius of a single mode fibre is $a = 2.5\mu\text{m}$.

References

- [3.1] A. Fox-Murphy, D. Howell, R. Nickerson, A. Weidberg, *Frequency Scanned Interferometry (FSI): the basis for an alignment system for the ATLAS Inner Detector using fast remote interferometry*, ATLAS Collaboration Internal Note INDET-NO-112, 6 October 1995.
- [3.2] A. Fox-Murphy, D. Howell, R. Nickerson, A. Weidberg, *Frequency Scanned Interferometry (FSI): the basis for an alignment system for the ATLAS Inner Detector using fast automated remote interferometry*, Proceedings of the Second International Symposium on the Development and Application of Semiconductor Tracking Detectors, Hiroshima, 10th–13th October 1995. To be published in Nuclear Instruments and Methods in Physics Research (Section A).
- [3.3] Monitoring of the Alignment of the Inner Detector, C. Daum, NIKHEF, INDET-TR-076, 12 Nov. 1993.
- [3.4] The performance specifications of the ATLAS Inner Detector, D. Froidevaux, A. Parker, ATLAS Collaboration Internal Note INDET-NO-046, 6 May 1994.
- [3.5] Alignment of the ATLAS Inner Detector, C. Daum, NIKHEF, ATLAS Collaboration Internal Note INDET-NO-060, 12 Sep. 1994.
- [3.6] The ATLAS Collaboration, Technical Proposal for a General-Purpose pp Experiment at the Large Hadron Collider at CERN, CERN/LHCC/94-43 LHCC/P2, 15 December 1994.
- [3.7] Alignment of the ATLAS detector from Inner Detector to muon spectrometer, C. Daum, NIKHEF, ATLAS Internal Note, TECH-NO-013, 7 Nov. 1994.
- [3.8] Least square fit and covariance matrix for an octagonal alignment grid for the ATLAS detector, C. Daum, NIKHEF, ATLAS Internal Note, TECH-NO-014, 23 Feb. 1995.

- [3.9] Alignment of the DELPHI vertex detector, V. Chabaud, CERN.

Multiple-frequency interferometric metrology

- [3.10] J.R. Benoit, *Journ. de Phys.* (3)7 (1898) 57.
- [3.11] *Principles of Optics*, Max Born and Emil Wolf, page 290, Pergamon Press, 1959.
- [3.12] The refractive index of air, B. Edlén, *Metrologia* 2 (1966) 71–80.
- [3.13] Two-reference-beam holographic interferometry, R. Dändliker, E. Marom and F.M. Mottier, *J.Opt.Soc.Am.* 66 (1976) 23.
- [3.14] Absolute distance measurements by CO₂ laser multiwavelength interferometry, G.L. Bourdet and A.G. Orszag, *Applied Optics* 18 (1979) 225–227.
- [3.15] Absolute distance measurement by variable wavelength interferometry, F. Bien, M. Camac, H.J. Caulfield and S. Ezekiel, *Applied Optics* 20 (1981) 400–403.
- [3.16] Infrared He-Xe laser interferometry for measuring length, Hirokazu Matsumoto, *Applied Optics* 20 (1981) 231–234.
- [3.17] Synthetic millimeter-wave signal generation for length measurement, Hirokazu Matsumoto, *Applied Optics* 23 (1984) 973–974.
- [3.18] Two-wavelength phase shifting interferometry, Yeou-Yen Cheng and James C. Wyatt, *Applied Optics* 23 (1984) 4539–4543.
- [3.19] Optical ranging by wavelength multiplexed interferometry, C.C. Williams and H.K. Wickramasinghe, *J. Appl. Phys.* 60 (1986) 1900–1903.
- [3.20] Distance measurement by the wavelength shift of laser light, Hisao Kikuta, Koichi Iwata and Ryo Nagata, *Applied Optics* 25 (1986) 2976–2980.
- [3.21] Fiber-optic interferometer using frequency-modulated laser diodes, Glenn Beheim, *Applied Optics* 25 (1986) 3469–3472.
- [3.22] Absolute distance measurement by wavelength shift interferometry with a laser diode: some systematic error sources, Hisao Kikuta, Koichi Iwata, and Ryo Nagata, *Applied Optics* 26 (1987) 1654–1660.
- [3.23] Interferometer for measuring displacement and distance, Toshihiro Kubota, Makoto Nara and Toshihiko Yoshino, *Optics Letters* 12 (1987) 310–312.

- [3.24] Two-wavelength laser interferometry using superheterodyne detection, R. Dändliker, R. Thalmann and D. Prongué, *Optics Letters* **13** (1988) 339–341.
- [3.25] Heterodyne interferometry with a frequency-modulated laser diode, Jun Chen, Yukihiro Ishii and Kazumi Murata, *Applied Optics* **27** (1988) 124–128.
- [3.26] Absolute optical ranging with 200 nm resolution, C.C. Williams and H.K. Wickramasinghe, *Optics Letters* **14** (1989) 542–544.
- [3.27] Interferometric absolute distance measurement utilizing a mode-jump region of a laser diode, K. Seta and B.K. Ward, *Opt. Comm.* **77** (1990) 275–278.
- [3.28] Optical-heterodyne displacement measurement using a frequency-ramped laser diode, Masaaki Imai and Koji Kawakita, *Optics Communications* **78** (1990) 113–117.
- [3.29] Interferometric determination of a static optical path difference using a frequency swept laser diode, Ryoji Ohba, Ichiro Uehira and Sei-ich Kakuma, *Meas. Sci. Technol.* **1** (1990) 500–504.
- [3.30] Synthetic wavelength stabilization for two-colour laser-diode interferometry, Peter de Groot and Stanley Kishner, *Applied Optics* **30** (1991) 4026–4032.
- [3.31] Recent advances in displacement measuring interferometry, N. Bobroff, *Meas. Sci. Technol.* **4** (1993) 907–926.
- [3.32] External frequency modulation of a laser source for non-incremental interferometric measurements, Letizia De Maria and Mario Martinelli, *Meas. Sci. Technol.* **4** (1993) 1228–1231.
- [3.33] Laser interferometry for precision engineering metrology, P. Gill, Chapter 5, *Optical Methods in Engineering Metrology*, edited by D.C. Williams, Chapman & Hall, London 1993, ISBN 0 412 39640 8.
- [3.34] Length metrology using optically-narrowed swept frequency laser diodes, G.P. Barwood, P. Gill and W.R.C. Rowley, Presented at Laser Dimensional Metrology, Photonex '93, Brighton, UK, 5–7 October 1993.
- [3.35] Laser diodes for length determination using swept-frequency interferometry, G.P. Barwood, P. Gill and W.R.C. Rowley, *Meas. Sci. Technol.* **4** (1993) 988–994.

- [3.36] Tunable, double-wavelength heterodyne detection interferometer for absolute-distance measurements, E. Gelmini, U. Minoni, and F. Docchio, *Optics Letters* **19** (1994) 213–215.
- [3.37] Laser triangulation : fundamental uncertainty in distance measurement, Rainer G. Dorsch, Gerd Häusler and Jürgen M. Hermann, *Applied Optics* **33** (1994) 1306–1314.
- [3.38] Active optical feedback in a dual-diode laser configuration applied to displacement measurements with a wide dynamic range, W.M. Wang, K.T.V. Grattan, W.J.O. Boyle and A.W. Palmer, *Applied Optics* **33** (1994) 1795–1801.

Corner-cube retroreflectors

- [3.39] A New Principle in Interferometer Design, E.R. Peck, *J.Opt.Soc.Am.* **38** (1948) 66.
- [3.40] Theory of the Corner-Cube Interferometer, E.R. Peck, *J.Opt.Soc.Am.* **38** (1948) 1015–1024.
- [3.41] Polarization Properties of Corner Reflectors and Cavities, E.R. Peck, *J.Opt.Soc.Am.* **52** (1962) 253–257.
- [3.42] Retroreflecting mirror for dynamic compensation of optical inhomogeneities, V.K. Orlov *et al.*, *Sov. J. Quantum Electronics* **8** (1978) 799–800.
- [3.43] Retroreflecting arrays as approximate phase conjugators, H.H. Barrett and S.F. Jacobs, *Optics Letters* **4** (1979) 190–192.

Lasers

- [3.44] Titan-CW Titanium Sapphire Laser, manufacturer's data sheet, Schwartz Electro-Optics (SEO), 3404 N. Orange Blossom Trail, Orlando, FL 32804, USA. Tel : 407/298-1802. Fax : 298-9889.
- [3.45] Single Frequency Ti:sapphire Laser MBR-110, manufacturer's data sheet, Micro-lase Optical Systems Ltd., 141 St. James Road, Glasgow G4 OLT, UK. Tel : +44 141 552 8205. Fax : +44 141 552 3906.
- [3.46] Tunable Lasers, manufacturer's data sheet, New Focus, Inc.

- [3.47] Continuously Tunable Diode Lasers, *Lasers and Optronics*, June 1993.
- [3.48] New dawn beckons for semiconductor lasers, D. Welch, *Physics World*.
- [3.49] Spectral characteristics of external-cavity controlled semiconductor lasers, Mark W. Fleming and Aram Mooradian, *IEEE Journal of Quantum Electronics*, Vol. QE-17, No. 1, January 1981, 44–59.
- [3.50] Frequency stabilization of semiconductor lasers by resonant optical feedback, B. Dahmani, L. Hollberg and R. Drullinger, *Optics Letters* **12** (1987) 876–878.
- [3.51] External-cavity frequency-stabilization of visible and infrared semiconductor lasers for high resolution spectroscopy, M.G. Boshier, D. Berkeland, E.A. Hinds and V. Sandoghdar, *Optics Communications* **85** (1991) 335–359.
- [3.52] Longitudinal mode control in laser diodes, G.P. Barwood, P. Gill and W.R.C. Rowley, *Meas. Sci. Technol.* **3** (1992) 406–410.
- [3.53] Self-mixing interference in a diode laser: experimental observations and theoretical analysis, W.M. Wang, W.J.O. Boyle, K.T.V. Grattan and A.W. Palmer, *Applied Optics* **32** (1993) 1551–1557.

Fibre optics

- [3.54] High-noise-rejection fibre-optic probe for interferometric applications, M. Martinelli, *Optics Letters* **7** (1982) 189–191.
- [3.55] Fibre-end interferometric sensor using cooperative retroreflectors, M. Johnson, *Optics Letters* **8** (1983) 593–595.
- [3.56] Stabilized fibre-end reftroreflecting interferometer, M. Johnson, *Applied Optics* **23** (1984) 2629–2632.
- [3.57] Remote interferometer using an optical fibre, Toshihiro Kubota, Toshihiko Yoshino and Teruji Ose, *Optics Letters* **9** (1984) 31–33.
- [3.58] Prefocused optics simplify fiberoptic laser-beam delivery, David Pointer, *Laser Focus World*, March 1990.

Frequency measurement

- [3.59] Accurate laser wavelength measurement with a precision two-beam scanning Michelson interferometer, J.-P. Monchalin, M.J. Kelley, J.E. Thomas, N.A. Kuzmit, A. Szöke, F. Zernike, P.H. Lee and A. Javan, *Applied Optics* **20** (1981) 736.
- [3.60] Frequency measurements on optically narrowed Rb-stabilised laser diodes at 780 nm and 795 nm, G.P. Barwood, P. Gill, and W.R.C. Rowley, *Applied Physics B* **53** (1991) 142–147.
- [3.61] The New Wavemeter, Internal Report, The Clarendon Laboratory, Oxford, U.K.
- [3.62] The calibration cavity for Parity Non-Conservation work, Internal Report, The Clarendon Laboratory, Oxford, U.K..
- [3.63] Optically Narrowed Rb-stabilised GaAlAs Diode Laser Frequency Standards with 1.5×10^{-10} Absolute Accuracy, G.P. Barwood, P. Gill and W.R.C. Rowley, *SPIE Proceedings* **1837** (Frequency-stabilized lasers and their applications), pages 262–70 (1992).
- [3.64] Fringe-counting technique used to lock a suspended interferometer, F. Barone *et al.*, *Applied Optics* **33** (1994) 1194–1197.

Phase-shift techniques

- [3.65] Direct phase detecting system, Y. Ichioka and M. Inuiya, *Applied Optics* **11** (1972) 1507.
- [3.66] Digital wavefront measuring interferometer for testing optical surfaces and lenses, J.H Bruning *et al.*, *Applied Optics* **13** (1974) 2693.
- [3.67] Use of an AC heterodyne lateral shear interferometer with real-time wavefront correlation systems, J.C. Wyant, *Applied Optics* **14** (1975) 2622.
- [3.68] Phase determination of an amplitude modulated complex wavefront, Demetri Psaltis and David Casasent, *Applied Optics* **17** (1978) 1136.
- [3.69] Optical phase measurement in real time, Lee M. Frantz, Alexander A. Sawchuk and Werner von der Ohe, *Applied Optics* **18** (1979) 3301.
- [3.70] Chapter 7, *The Laser Doppler Technique*, L.E. Drain, Wiley (1980).

- [3.71] Digital wave-front measuring interferometry: some systematic error sources, J. Schwider *et al.*, *Applied Optics* **22** (1983) 3421.
- [3.72] Multichannel phase-shifted interferometer, Osuk Y. Kwon, *Optics Letters* **9** (1984) 59.
- [3.73] Two-wavelength phase shifting interferometry, Yeou-Yen Cheng and James C. Wyant, *Applied Optics* **23** (1984) 4539.
- [3.74] Contouring aspheric surfaces using two-wavelength phase-shifting interferometry, Katherine Creath, Yeou-Yen Cheng and James C. Wyant, *Optica Acta* **32** (1985) 1455.
- [3.75] Computer-aided analysis of holographic interferograms using the phase-shift method, Bernd Breuckmann and Werner Thieme, *Applied Optics* **24** (1985) 2145.
- [3.76] Phase shifter calibration in phase-shifting interferometry, Yeou-Yen Cheng and James C. Wyant, *Applied Optics* **24** (1985) 3049.
- [3.77] Phase-shifting speckle interferometry, Katherine Creath, *Applied Optics* **24** (1985) 3053.
- [3.78] Homogeneity testing by phase sampling interferometry, Johannes Schwider, R. Burow, K.-E. Elssner, R. Spolaczyk and J. Grzanna, *Applied Optics* **24** (1985) 3059.
- [3.79] Phase-Measurement Interferometry Techniques, Katherine Creath, *Progress in Optics XXVI*, E. Wolf (1988) 349–393.
- [3.80] pp316–319, *Holographic and Speckle Interferometry*, 2nd. edition, Robert Jones and Catherine Wykes, *Cambridge studies in modern optics: 6*, Cambridge University Press (1989), ISBN 0-521-34878-1.
- [3.81] Direct Spatial Reconstruction of Optical Phase from Phase-Modulated Images, Valentin I. Vlad and Daniel Malacara, *Progress in Optics XXXIII*, E. Wolf (1994) 261–317.

Vibration analysis

- [3.82] Remote vibration measurement of rough surfaces by laser interferometry, Robert A. Bruce and Gerald L. Fitzpatrick, *Applied Optics* **14** (1975) 1621–1626.

- [3.83] Laser Doppler vibration analysis measuring system using bispectral analysis, Osami Sasaki, Takuso Sato, and Tetsuya Oda, *Applied Optics* **19** (1980) 151–153.
- [3.84] Method of vibration measurements in heterodyne interferometry, Karl A. Stetson, *Optics Letters*, **7** (1982) 233–234.
- [3.85] High-speed laser interferometry applied to transient vibration measurements, Richard S. Eng, Kristine A. Langdon, William A. Richardson, and John Lapato, *Applied Optics* **23** (1984) 2956–2960.
- [3.86] Fringe-counting technique used to lock a suspended interferometer, Fabrizio Barone, Enrico Callone, Rosario De Rosa, Luciano Di Fiore, Francesco Fusco, Leopoldo Milano, and Guido Russo, *Applied Optics* **33** (1994) 1194–1197.



Chapter 4

Geodetic networks

4.1 Introduction

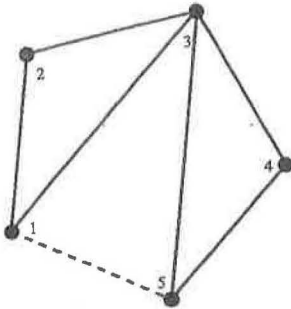
The techniques applicable to real time surveying of a particle detector provide one- or two- and three-dimensional measurements at a number of localised positions. In order to produce a consistent, distributed survey of detector element positions across a whole (sub)detector, it is necessary to combine a number of individual measurements. This may be done by configuring the measurements as a three-dimensional **geodetic network** — a network composed of a number of nodes and measurements between nodes, each node generally being common to several measurements ¹.

The complexity of the survey requirements and the low headroom between the intrinsic measurement precision of suitable techniques and the required precision mean that careful optimisation of the configuration of the network is necessary. A suitable network will permit determination of the node locations with the required precision, and will satisfy a number of quality (reliability) criteria defined in this chapter. In addition, there are HEP requirements such as redundancy which must be met.

In the sections which follow, a mathematical framework for analysis of geodetic alignment networks is developed, initially for the general case and subsequently for specific networks suitable for surveying the Inner Detector precision layers using FSI.

¹The word **geodetic** was coined for use in **geodesy** (the science of Earth surveying) and refers to a network of imaginary lines drawn on the surface of the Earth for measurement purposes. It now has currency in many other areas where a network of nodes and lines is used. Its application to a three-dimensional survey of a particle detector seems appropriate.

4.2 Mathematical geodesy



The purpose of this section is to present a mathematical framework for the analysis of geodetic networks which may be used for both the analysis of the performance of a geodetic network (in order to optimise the design) and for the processing of survey data (to determine the coordinates of the nodes from the measured quantities, and to spot bad data). The philosophy of the description given here is not to explain the mathematics comprehensively but rather to give an overview of the techniques and their capabilities. The mathematics presented here is coded in a software package called SCAN-3 [1]. Results of calculations made with this program for a number of FSI networks are presented in Section 4.5.

The techniques described are very general. The observations may be any of the types likely to be encountered in the survey of a particle detector : one-dimensional (e.g. lengths from FSI or 1D transverse MPI measurements), two-dimensional (e.g. RASNIK transverse measurements) or three-dimensional (e.g. RASNIK with magnification measurement). Combinations of any number of different types of measurement are permitted, although a full description of how this can be done will not be given here.

4.2.1 Analysis phases

As mentioned above, there are two distinct stages in a geodetic analysis :

Design : Before measurements are available, the performances of trial geodetic network configurations are evaluated in order to test whether they give precision and quality measures which satisfy the requirements.

Data analysis : Following the acquisition of a measurement set, the data is checked and the desired parameters are evaluated.

At present the ATLAS Inner Detector geodetic survey network is in the planning phase. The next stage in the development of the FSI system will permit multiple simultaneous measurements which will allow an investigation of the data analysis.

4.2.2 Model description notation

A measurement campaign of any kind makes use of a number of measurements (observations) to determine the values of a number of desired quantities (parameters). In the case under consideration here, the desired parameters are related to the spatial locations of a set of points, and the observations made are stochastic quantities with a known functional relationship to the parameters. In addition there may be a number of deterministic or stochastic constraints. For a geodetic network with n parameters, and with which m observations are made and b constraints are imposed then the redundancy is $r = m + b - n$. Clearly the redundancy must be greater than or equal to zero to permit determination of the parameters from the observations ².

The notation used in this chapter is that of Reference [1]. Stochastic quantities are underlined, least squares solutions are denoted with a hat, and the subscript a is used for unconstrained solutions.

If the observations are represented by the $(m \times 1)$ vector \underline{y} , and the parameters by the $(n \times 1)$ vector x then the relationship between them may be expressed as

$$\begin{array}{l} E\{\underline{y}\} = A(x) ; \quad B^*(x) = c \\ D\{\underline{y}\} = Q_y ; \quad D\{\underline{c}\} = Q_c \end{array}$$

where

$E\{.\}$: mathematical expectation value of a stochastic quantity

$D\{.\}$: mathematical dispersion of a stochastic quantity

x : $(n \times 1)$ vector of parameters (i.e. node coordinates)

\underline{y} : $(m \times 1)$ vector of observations (i.e. measured lengths, angles etc.)

c : $(b \times 1)$ vector of constraint values

$A(x)$: m non-linear observation equations in the parameters x

$B^*(x)$: b non-linear constraint equations in the parameters x

Q_y : $(m \times m)$ covariance matrix for observations

Q_c : $(b \times b)$ covariance matrix for constraints

A set of equations such as this is referred to as a **model**. Changes to a model are made for various treatments of constraints and for quality testing of a network using

²For the 5-node 2D geodetic network at the beginning of this section, $n = 10$ and $m = 7$. It is a rigid network, but 3 coordinate constraints are required to determine the θ and (x,y) position of the whole network uniquely (i.e. to reach $r = 0$) within a planar coordinate system.

the methods of hypothesis testing described in Section 4.2.4.

In nearly all cases \mathcal{A} and \mathcal{B} will be non-linear functions of the parameters. The first term of the Taylor expansion about the point x_0 in the parameter space yields a system of linearised equations to which the techniques of linear algebra (matrix algebra) may be applied. The point x_0 , which will be referred to hereafter as the **nominal parameter values**, should be an approximate estimate of the parameters obtained either from *a priori* knowledge (such as the ideal network shape) or from a previous measurement set.

$$\begin{array}{l} E\{\underline{\Delta y}\} = \partial_x \mathcal{A} \Delta x \quad ; \quad \partial_x \mathcal{B}^* x_0 = \Delta c \\ D\{\underline{\Delta y}\} = Q_y \quad ; \quad D\{\underline{\Delta c}\} = Q_c \end{array}$$

For simplicity of notation the first derivatives $\partial_x \mathcal{A}$ and $\partial_x \mathcal{B}^*$ will be written in the sections which follow as the matrices A and B^* (the adjoint now being interpreted as the transpose). These derivative matrices are known as the **design matrices**.

4.2.3 Adjustment

Once a measurement data set has been acquired, the values of the parameters are determined. The general technique is to start with a nominal set of parameters, and to use an **adjustment** procedure to obtain a set of parameters (i.e. node positions) which are the ‘best fit’ to the set of observations. See Figure 4.1.

Adjustment requires a minimisation procedure. There are a number of techniques available: least squares, maximum likelihood, L1, L2 and others. The least squares minimisation technique yields Best Linear Unbiased Estimates (BLUEs) — parameter estimates which are unbiased and compared to which no other estimates have smaller variances (under the assumption that the model is correct and provided that the covariance matrix of observations is used in the adjustment to weight the observations).

It is helpful to consider constrained least squares minimisation in geometrical terms. See Figure 4.2. The unconstrained least squares solution is the point \hat{x}_a in \mathbb{R}^n for which the squared norm of the vector $(\underline{y} - A\hat{x}_a)$ is minimal in the metric of \mathbb{R}^m . In a constrained least squares adjustment, the solution \hat{x}_a is projected on to the b -dimensional subspace defined by the constraint equations.

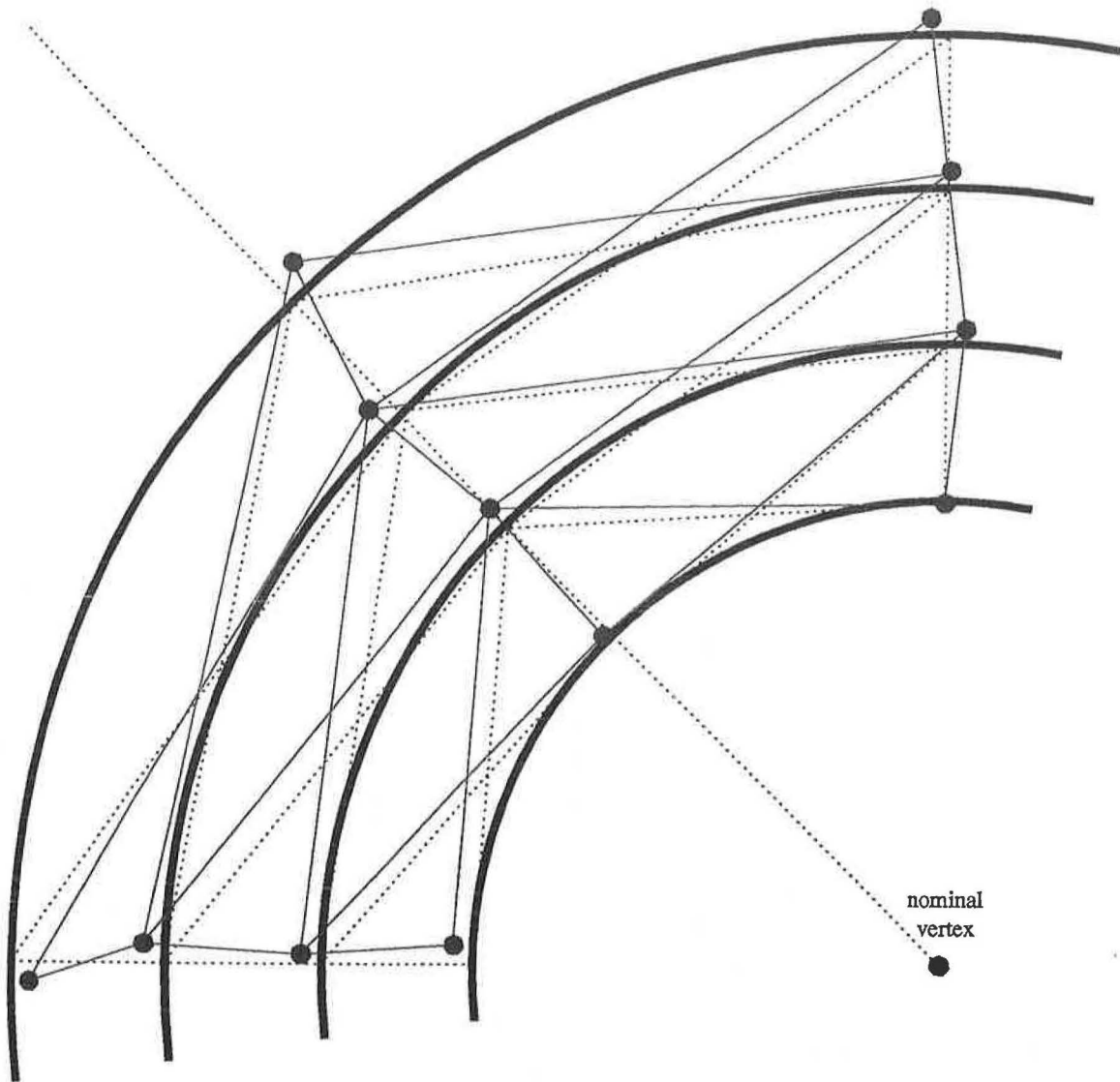


Figure 4.1: An illustration of adjustment. An end view of the barrel SCT layers is shown. The barrels and the dotted lines represent the nominal network and its measurement set. The actual positions of the nodes (circles) and the corresponding measurements (thin solid lines) will differ from nominal by an amount depending on barrel shape distortions and end-point positioning constants. The adjustment procedure is started at the nominal (or last survey) positions and is iterated until convergence, at which point the solution \hat{x} should correspond to a best fit approximation to the actual node positions.

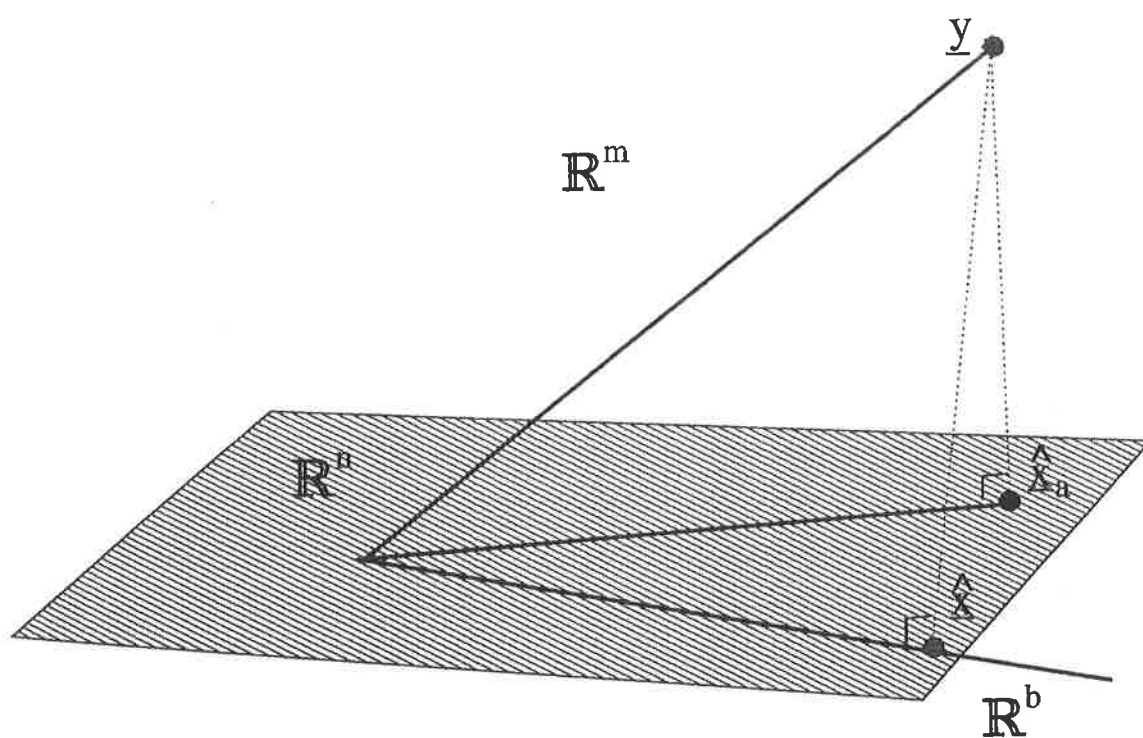


Figure 4.2: A geometrical picture of constrained least squares adjustment as a double projection. Here there are 3 measurements, 2 parameters and 1 constraint.

4.2.3.1 Unconstrained adjustment

The least squares technique is to minimise the functional

$$F(x) = (\underline{y} - Ax)^* Q_y^{-1} (\underline{y} - Ax) \quad (4.1)$$

by requiring $\partial_x F(x) = 0$. This gives the system of **normal equations**:

$$(A^* Q_y^{-1} A) \hat{x}_a = A^* Q_y^{-1} \underline{y} \quad (4.2)$$

which yields **unconstrained solution** \hat{x}_a and its **covariance matrix** $Q_{\hat{x}_a}$

$$\hat{x}_a = (A^* Q_y^{-1} A)^{-1} A^* Q_y^{-1} \underline{y} \quad (4.3)$$

$$Q_{\hat{x}_a} = (A^* Q_y^{-1} A)^{-1} \quad (4.4)$$

where the elements of the covariance matrix are

$$Q_{ij} = \langle (x_i - \bar{x}_i)(x_j - \bar{x}_j) \rangle \quad (4.5)$$

Q is the inverse of the matrix of second derivatives of $F(x)$ with respect to pairs of parameters, calculated at its global minimum:

$$(Q^{-1})_{ij} = \frac{1}{2} \frac{\partial^2 F}{\partial x_i \partial x_j} \quad (4.6)$$

Due to the inversion, the diagonal elements of Q contain contributions from all the elements of the second derivative matrix, meaning that the diagonal elements of the covariance matrix are the squares of the individual parameter errors, *including the effects of correlations*.

4.2.3.2 Adjustment with constraints

Least squares adjustment with constraints may be achieved by minimising the quantity

$$W(x, \lambda) = F(x) - (B^* x - c)^* \lambda$$

where λ is a $(b \times 1)$ vector of **Lagrange multipliers**. It can be shown that $F(x)$ is minimal under the constraints when $\partial_x W(x, \lambda) = 0$ which leads to a set of **extended normal equations**

$$\begin{pmatrix} A^* Q_y^{-1} A & B \\ B^* & 0 \end{pmatrix} \begin{pmatrix} \hat{x} \\ \frac{1}{2} \lambda \end{pmatrix} = \begin{pmatrix} A^* Q_y^{-1} \underline{y} \\ c \end{pmatrix} \quad (4.7)$$

whence the Lagrange multipliers and the **constrained solution** $\hat{\underline{x}}$

$$\frac{1}{2}\hat{\lambda} = (B^*Q_{\hat{x}_a}B)^{-1}(B^*\hat{x}_a - c) \quad (4.8)$$

$$\hat{\underline{x}} = \hat{x}_a - Q_{\hat{x}_a}B\frac{1}{2}\hat{\lambda} \quad (4.9)$$

The precision of the parameter determination is given by the covariance matrices

$$R = B^*Q_{\hat{x}_a}B$$

$$Q_{\frac{1}{2}\hat{\lambda}} = R^{-1}(R - Q_c)R^{-1} \quad (4.10)$$

$$Q_{\hat{x}} = Q_{\hat{x}_a} - Q_{\hat{x}_a}BQ_{\frac{1}{2}\hat{\lambda}}B^*Q_{\hat{x}_a} \quad (4.11)$$

The above equations relate to a single iteration of the least squares adjustment procedure. Repeated iteration should result in convergence of the solution $\hat{\underline{x}}$ to a best-fit approximation of the actual network parameters.

4.2.3.3 Explicit observation equations

There follows a simple 2D explicit derivation of the form of the unconstrained least squares solution.

The distance ℓ_{ik} between any two nodes i and k is given by

$$\ell_{ik}^2 = r_i^2 + r_k^2 - 2r_i r_k \cos(\phi_k - \phi_i) \quad (4.12)$$

where (r_i, ϕ_i) and (r_k, ϕ_k) are the node coordinates. Note that the above equation is a non-linear observation equation. Linearising by taking the first term of the Taylor expansion about the nominal coordinates ℓ_{ik}^0 one obtains

$$\begin{aligned} \Delta\ell_{ik} &= \ell_{ik} - \ell_{ik}^0 \\ &= \sum_j \frac{\partial\ell_{ik}}{\partial x_j} \Delta x_j \end{aligned} \quad (4.13)$$

where the partial derivatives are given by

$$\begin{aligned} \frac{\partial\ell_{ik}}{\partial r_i} &= \frac{1}{\ell_{ik}}(r_i - r_k \cos(\phi_k - \phi_i)) \\ \frac{\partial\ell_{ik}}{\partial \phi_i} &= -\frac{1}{\ell_{ik}}(r_i r_k \sin(\phi_k - \phi_i)) \end{aligned} \quad (4.14)$$

and similar expressions for $\frac{\partial\ell_{ik}}{\partial r_k}$ and $\frac{\partial\ell_{ik}}{\partial \phi_k}$. If it is assumed that the deviations are small, numerical values for these partial derivatives may be evaluated for the nominal network.

Equation 4.13 yields a set of m simultaneous observation equations which are linear in the n parameters x_j (the (r, ϕ) coordinates of the nodes):

$$E\{\Delta y_i\} = \sum_{j=1}^n A_{ij} \Delta x_j \quad i = 1, 2, \dots, m \quad (4.15)$$

Then, dropping the Δ s as before, for any particular measurement set

$$y \sim Ax \quad (4.16)$$

where y is a vector of measured deviations from the nominal network (i.e. a particular instance of a stochastic quantity), and A is the matrix of partial derivatives, The least squares functional $F(x)$ is given by

$$F(x) = \sum_{i=1}^m \left(\frac{y_i - \sum_{j=1}^n A_{ij}x_j}{\sigma_i} \right)^2 \quad (4.17)$$

where the σ_i is the standard deviation of each 1D measurement y_i . The single-iteration unconstrained least squares solution is the set of parameters satisfying

$$\begin{aligned} 0 &= \frac{\partial F}{\partial x_k} \\ &= \frac{\partial}{\partial x_k} \sum_{i=1}^M \left(\frac{y_i - \sum_{j=1}^n A_{ij}x_j}{\sigma_i} \right)^2 \\ &= \sum_{i=1}^M \frac{-2A_{ik}}{\sigma_i} \left(\frac{y_i - \sum_{j=1}^n A_{ij}x_j}{\sigma_i} \right) \quad \forall k = 1, 2, \dots, n \end{aligned} \quad (4.18)$$

Expanding the bracket

$$0 = \sum_i \frac{A_{ik}y_i}{\sigma_i^2} - \sum_i \sum_j \frac{A_{ik}A_{ij}x_j}{\sigma_i^2} \quad (4.19)$$

If all the measurements are made to the same precision then this reduces to

$$A^T y = A^T A x \quad (4.20)$$

which yields the unconstrained solution for a set of measurements with the same precision (c.f. Equation 4.3)

$$x = (A^T A)^{-1} A^T y. \quad (4.21)$$

4.2.4 Quality testing

Aside from the precision, there are other criteria which characterise a good network. Quality (or reliability) testing of a network centres around the idea of supposing that there is a mistake (or **bias**) in one of the observations, and then calculating what the probability of detecting it is and what the effect of an undetected bias would be on the values of the parameters determined using such a network.

The **minimum detectable bias** (MDB) ∇ of an observation is the smallest bias in the observed value which could be detected with probability γ by systematic cross-checking of the measurements at the data analysis stage.

The **bias-to-noise ratio** (BNR) $\sqrt{\lambda_{\hat{x}_1}}$ of an observation expresses the effect on the whole network of an undetected bias in the observation with a magnitude equal to the minimum detectable bias. With probability γ , an undetected error ∇ will not cause an error in any of the parameter estimates $\Delta \hat{x}_i$ greater than $\sigma_{\hat{x}_i} \sqrt{\lambda_{\hat{x}_1}}$. Put another way,

$$\frac{\Delta x_i}{\sigma_{\hat{x}_i}} \leq \sqrt{\lambda_{\hat{x}_1}} \quad \forall i \quad (4.22)$$

where $\sigma_{\hat{x}_i}$ are the standard deviations of the parameters, i.e. the square roots of the diagonal elements of the covariance matrix. A reliable network is one in which the minimum detectable bias in any one of the observations is small, and for which the effect of an undetected bias in any observation on any parameter is likewise small. Most questions about the calculational quality of a network can be expressed in terms of these two quantities and the covariance matrix $Q_{\hat{x}}$. Quality testing is thus used

- (a) during the design phase to analyse the quality of a candidate network
- (b) during the data processing phase to systematically examine a set of observations in order to reveal biases which could have resulted from erroneous measurements cause by mistakes, poor calibrations etc. and to decide whether these biases are significant. This type of systematic checking is known as **data snooping**.

The general technique used to make these calculations is known as **hypothesis testing**. Stages (a) and (b) both involve consideration of firstly the **null hypothesis**, denoted H_0 , that the form of the original model is correct ³

$$H_0 : E \left\{ \begin{pmatrix} y \\ c \end{pmatrix} \right\} = \begin{pmatrix} A \\ B^* \end{pmatrix} x ; D \left\{ \begin{pmatrix} y \\ c \end{pmatrix} \right\} = \begin{pmatrix} Q_y & 0 \\ 0 & Q_c \end{pmatrix} \quad (4.23)$$

³For test purposes constraints are treated in the same way as observations.

and secondly an **alternative hypothesis**, H_A , that the form of the model differs in a chosen way from the original (systematically different for each observation i)

$$H_{A_i} : E \left\{ \begin{pmatrix} y \\ c \end{pmatrix} \right\} = \begin{pmatrix} A \\ B^* \end{pmatrix} x + \begin{pmatrix} c_i \\ 0 \end{pmatrix} \nabla ; \quad D \left\{ \begin{pmatrix} y \\ c \end{pmatrix} \right\} = \begin{pmatrix} Q_y & 0 \\ 0 & Q_c \end{pmatrix} \quad (4.24)$$

where c_i is an $(m \times 1)$ **unit vector** in \mathbb{R}^m ($\dots 0 \ 1 \ 0 \dots$) and ∇ is a (scalar) unknown model error.

During the evaluation of a candidate network at the design stage, the MDB and BNR are calculated for each of the observations using the mathematical model. In the data analysis stage, in order to spot biases in the measured data set, a quantity known as a **test statistic** is evaluated for each observed value. These test statistics are mutually correlated, and the observation with the largest test statistic is the one most likely to contain a bias.

4.2.5 Basis-independent precision estimates

In general, extra information has to be added to a model (in the form of constraints) to locate the whole network within a coordinate system — i.e. it is **rank-deficient**. In the previous sections a full-rank model was assumed. The analysis for a rank-deficient model is very similar to that presented above. However, the covariance matrix is not basis-independent : the values of the elements of $Q_{\hat{x}}$ are dependent on which points were fixed within the coordinate system — points close to the fixed points will have smaller standard deviations than those further away.

It is common to require measures of precision which are independent of the particular coordinate choice made. Such basis-independent precision estimates may be obtained in two ways :

- One way is not to judge the precision of the coordinates but on derived quantities which are independent of the coordinate definition. The precisions of angles and distances (derived from the coordinate precisions by the law of propagation of variances) are independent of the S-basis chosen.
- A second way to obtain basis-independent precision estimates is by comparing the covariance matrix with a specially constructed **criterion matrix** H by means of the generalised eigenvalue equation :

$$Q_{\hat{x}} v = \lambda H v \quad (4.25)$$

The eigenvalues λ are basis-independent measures of the variances of the 3D distances between the determined node positions and their true values. Note that the eigenvectors v are not basis-independent.

The interpretation is that if the largest eigenvalue from this general eigenvalue problem is less than one, then for all linear functions derived from the coordinates, the precision based on the covariance matrix Q is better than that based on the criterion matrix H . In practice the eigenvalue analysis is rather complicated.

4.2.6 Combination of networks

In general, a survey network may be formed from a number of sub-networks. Three possible scenarios are

- (a) a single standalone network
- (b) full integration of two or more networks by applying a full least squares minimisation to the combined network formed from their sum
- (c) separate networks adjusted separately and then fixed together using peripheral constraints (a pseudo-least squares minimisation)

At present it is foreseen that the Inner Detector will be equipped with a standalone alignment network which will be spatially related to the other subdetectors using tracks. The geodetic analysis techniques outlined above will be useful for analysing how best to combine Inner Detector surveys with survey measurements made in the outer parts of the detector and with measurements of the underground area which will relate the position of ATLAS to the cavern wall and to the LHC.

4.3 SCAN-3

SCAN-3 [1] is a software package for the design of geodetic networks (and combinations of networks) and the analysis of recorded data. It is a mature system which is the result of 30 years of development in the field of land surveying. It has the following features :

- networks may be one- two- or three-dimensional
- precision and reliability calculations are performed in the framework of constrained least squares explained in Section 4.2
- any number of deterministic and/or stochastic constraints may be defined
- a number of types of hypothesis test, including those discussed in Section 4.2.4, may be made for reliability analysis of networks.
- basis-independent precision estimates may be calculated using the method of Section 4.2.5.
- a large number of possible measurement types are available in one, two or three dimensions and combinations thereof
- a network may be constructed from a number of observation groups, with different types of measurements in each group
- multiple networks may be analysed separately before full or pseudo-least squares combination
- parameters related to measurement instruments are included, permitting integration of calibration uncertainty (i.e. scale of whole network) and other instrumental parameters into the design and analysis
- extensive use of coordinate transformations means that the software can handle input and output in a number of different representations such as Cartesian, ellipsoidal and map coordinates. Extensions to the range of coordinate systems are possible.

SCAN-3 has been used to evaluate the precision and reliability of a number of geodetic network designs for an FSI-based survey system for the ATLAS Inner Detector. The results are presented in Section 4.5.

4.4 FSI network design

Geodetic network design bears similarities to the engineering problem of designing a maximally stiff structure from a number of compressible or extensible rods fixed together with universal joints. The requirements of high precision and unique node location lead naturally to rigid engineering-like structures of triangles and struts. The challenge is to design a network which meets a set of precision criteria within set of mechanical constraints.

The design of a geodetic FSI-based alignment network for the ATLAS Inner Detector is currently in the preliminary stages. Information about which kinds of network meet the alignment requirements will be very important for defining the line-of-sight requirements, designing the jewels and organising the multiplexing of groups of measurements.

4.4.1 Precision

As explained previously, the precision requirement for the survey system is set by the need to be able to determine the positions of individual detector elements to a precision sufficient for track reconstruction which meets the physics requirements. The most stringent requirement is in the Φ direction for which a total alignment precision of $10\mu\text{m}$ or better is required. If an allowance is made for the positioning tolerances of the FSI jewels and for inaccuracies arising from the interpolation between jewels, the Φ -precision required from the geodetic network itself will be about $5\mu\text{m}$. The radial and longitudinal alignment specifications are more than ten times weaker than the azimuthal requirement. It will be seen that suitable networks will have comparable precisions in all three directions, and so the R and Z specifications will be met comfortably by any network meeting the Φ requirement.

4.4.2 Mechanical constraints

A brief account is given below of the current understanding of the mechanical constraints which must be taken into consideration for the design of a geodetic FSI network which could be used for alignment of the SCT barrel. The networks which will be studied in Section 4.5 are not intended to be full designs, but are rather intended as an initial investigation of the number and configuration of the measurements needed to meet the specifications. The alignment of the forward regions will not be considered here.

The design currently favoured for the SCT barrel is composed of four solid carbon fibre cylinders with a coefficient of thermal expansion smaller than $0.5 \text{ ppm}/^\circ\text{C}$ [10]. Double-sided detector modules are pre-assembled in lengthwise rows on local support structures which are then attached to machined metal inserts glued into holes in the barrels. Cooling tubes may be part of the detachable local support units or may be permanently fixed to the barrel. A flange with a width of about 3 cm will be attached to the end of each barrel. The barrels are joined together at each end with an interlink plate. Services run from the cylinders through this plate before bending towards the outer radius in their path out of the detector. The modularity of the services in this region is under discussion.

The primary mechanical requirement of FSI is for lines of sight. Since the silicon layers will be positioned to give 100% coverage, it is only likely to be possible to place jewels in planes at the ends of the barrels. These jewels would then be used for making transverse measurements in the end plane and longitudinal measurements from one cylinder to another through the gap between planes. Provided that the services can be positioned so as to provide lines of sight, it is expected that there will be spaces for transverse measurements in the plane between the ends of the barrels and the interlink plate. The lines of sight would then have widths of 5 mm transverse to the measurement axes.

The radial space occupied by the modules increases the effective thickness of the cylinders. As well as restricting the space available for making longitudinal measurements, this may mean that it will be necessary for the jewels to be designed to span the ends of the cylinders so that longitudinal measurements can be made on both sides.

The amount of space required for services beyond the interlink plate (i.e. in the transition region) is likely to preclude the use of FSI in this region. It is therefore assumed that the barrel and the two end-caps will be internally aligned and then linked as rigid bodies using track information.

4.4.3 Reliability requirements

The requirement for minimum detectable bias in the context of FSI requires some explanation. Since FSI measurements are obtained by counting interference fringes seen at a set of photodetectors as the optical frequency of a single laser is scanned, a possible measurement bias is that a fringe is missed in the counting process. This is most likely to occur for the longer measured lengths because the separation of the fringes

(in optical frequency) is inversely proportional to R_c , and because the signal-to-noise ratio decreases with increased measured distance. The fringe counting will be more robust for short measurements since the fringes become wider and the signal-to-noise ratio improves.

From Equation 3.10 the relationship between the measured length, the scan range and the number of fringes counted is $R_c \sim \frac{c\Delta N}{2\Delta\nu}$. An error of one fringe would cause a bias in the measured value of R_c of $\nabla = \frac{1}{\Delta N} \times R_c = \frac{c}{2\Delta\nu}$. For $\Delta\nu = 3 \text{ THz}$ this corresponds to a bias of $\nabla = 50 \mu\text{m}$. Thus a minimum detectable bias of less than $50 \mu\text{m}$ for $\gamma = 0.999$ for all measurements would mean that it would be likely that any single fringe counting error in an FSI data set of ~ 500 measurements would be detected during the data analysis.

Another possible source of a measurement bias is a localised change in the refractive index of the gaseous medium due to an undetected hot- or cold-spot. As explained in Section 3.2.3, a change in the path-averaged temperature of 0.1°C would cause a 0.1 ppm change in the refractive index, and since the measured length is inversely proportional to the value of the refractive index (see Equation 3.10), this would result in a 0.1 ppm measurement bias. There is no lower limit to the MDB which would be derived from this type of bias, but as explained already, it is hoped the gas flow rate and the mixing will be sufficient to make any index-induced biases small compared to the desired accuracy.

4.5 Calculation results

SCAN-3 was applied to several preliminary FSI networks to perform calculations of the three precision components for each node (from the covariance matrix Q_x), and minimum detectable biases and bias-to-noise ratios for each observation.

The networks considered included only the four silicon layers of the Inner Detector barrel, i.e. the SCT barrel. Extension to include the barrel pixel layers was not considered necessary for this preliminary investigation.

The SCT barrel network is conveniently factorised into two planar ‘transverse’ networks of measurements made within the two $r\phi$ planes at the two ends of the barrel and a ‘longitudinal’ network of measurements joining the two transverse networks with measurements running parallel or nearly parallel to the z -axis. These two types of networks are almost orthogonal and relatively independent.

Two designs (A and B) for the transverse network are shown in Figures 4.3 and 4.4.

They both exhibit octagonal symmetry but have different numbers of measurements in order to investigate the dependence of the performance on the number of measurements made.

Two designs (I and II) for the longitudinal network are shown in Figures 4.5 and 4.6. Not all the measurements are in the plane of the diagrams : the lines schematically indicate both 'straight' measurements made parallel to the z -axis and 'skew' measurements running round in ϕ by plus or minus one node. In order to investigate the dependence of the performance on the number of longitudinal measurements, calculations were performed with either one straight and one skew measurement, or one straight and two skew measurements. Thus the lines represent doublets or triplets of measurements.

The lengths of the transverse measurements lie in the range 10–40 cm, whereas the longitudinal measurements are all of comparable length, being about 90 cm in layouts with a middle plane, or 180 cm with no middle plane. The variation of the FSI precision as a function of the measured range R_c was discussed in Section 3.2.4. Although it is possible that the precision σ_R may be better for shorter measured lengths, for the purposes of the calculations presented here, the conservative assumption was made that the precision will be the same for every measurement. The precisions, the minimum detectable bias and the bias-to-noise ratio are quoted below in multiples of this common measurement precision.

SCAN-3 was employed to perform calculations of the precisions, MDBs and BNRs for various combinations of the longitudinal and transverse measurement configurations.

The technique used to calculate the precisions requires some explanation. As explained in Section 4.2.5, calculation of precision estimates for a network requires the addition of a number of constraints. Typically this is done by fixing one or more of the three coordinates of each of a number of nodes. The precision estimates obtained in such a way (for a given coordinate) have a range of values, varying from zero (for nodes constrained in that coordinate) to a maximum value (for nodes most distant from the constrained points). Since calculation of basis-independent precisions involves a rather complicated procedure, the precision estimates presented here were calculated in the simple basis-dependent fashion. In order to minimise and regularise the effect of the constraints, the choice of constrained points was made so as to minimise the standard deviations of the least well determined points. It is believed that this results in precision estimates which are pessimistic by a factor two. The standard deviations of the least well determined points obtained using this procedure are shown in Table 4.1. The carte-

sian (x, y, z) precisions obtained from SCAN-3 were converted into (r, ϕ, z) cylindrical coordinates which are more natural for a cylindrically symmetric particle detector.

The MDB and BNR results are shown in Figures 4.7 to 4.10. Because they are effectively 3D lengths, the MDB and BNR values are basis-independent for the reasons explained in Section 4.2.5. In addition, the longitudinal MDB and BNR figures are almost independent of the transverse measurement configuration employed, and *vice versa*. The values shown correspond to a hypothesis test ‘power’ of $\gamma = 0.8$. The figures for $\gamma = 0.99$ and $\gamma = 0.999$ are 36% and 55% larger respectively. Clearly for any of the networks, including those with only double measurements (not shown), a single fringe skip in one measurement out of a full FSI data set would be detected during the data analysis.

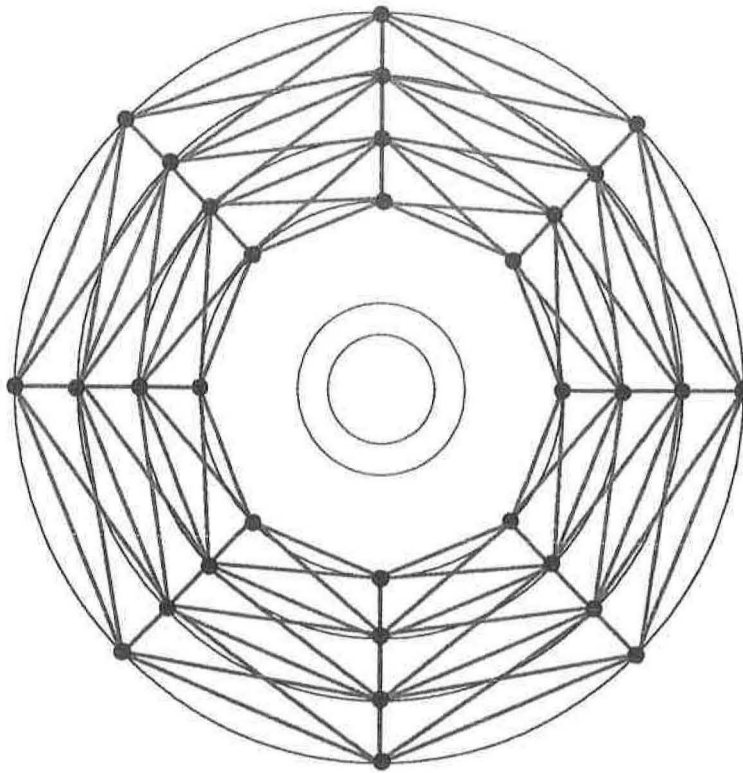


Figure 4.3: *Type A transverse network. The dots and lines are jewels and measurement axes respectively. All measurements are in the plane.*

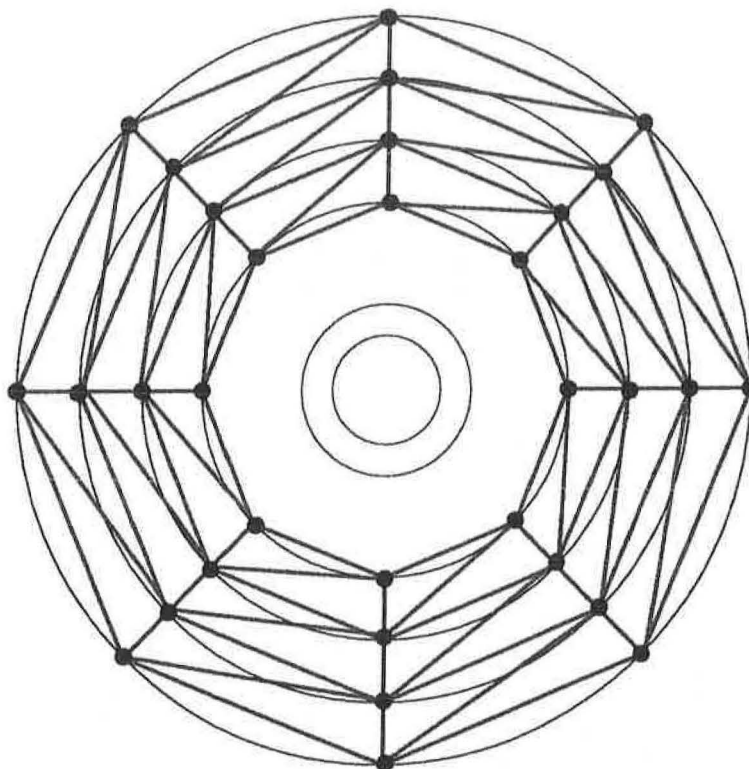


Figure 4.4: *Type B transverse network. The dots and lines are jewels and measurement axes respectively. All measurements are in the plane.*

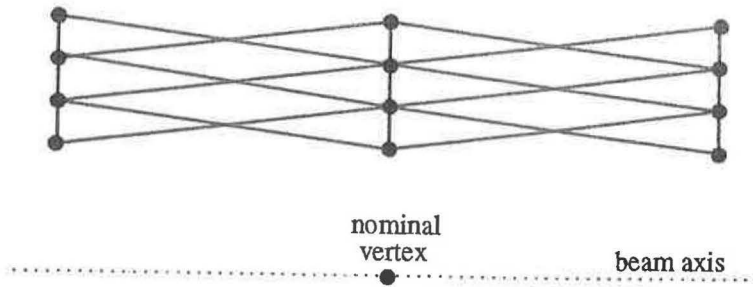


Figure 4.5: *Type I longitudinal network. The dots are jewels. The lines correspond to doublets or triplets of measurements [see text].*

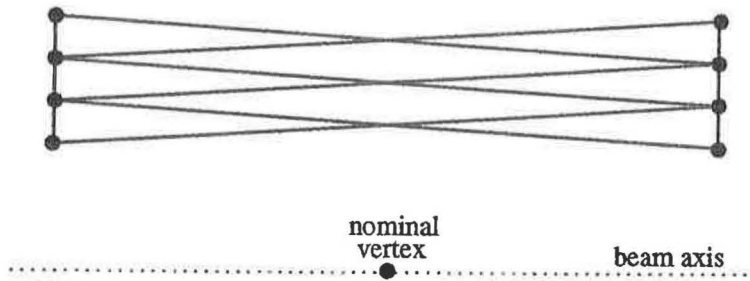


Figure 4.6: *Type II longitudinal network. The dots are jewels. The lines correspond to doublets or triplets of measurements [see text].*

Model	T type	L type	m	n	b	r	$\frac{\sigma_r^{max}}{\sigma_{FSI}}$	$\frac{(r\sigma_\phi)^{max}}{\sigma_{FSI}}$	$\frac{\sigma_z^{max}}{\sigma_{FSI}}$
7C	A	I (triple)	600	288	7	319	2.8	3.2	1.6
8C	A	II (triple)	352	192	9	169	4.1	5.0	2.4
8F	A	II (double)	304	192	9	121	5.9	7.5	3.6
9A	B	II (double)	256	192	8	72	16.5	22.0	8.7

Table 4.1: *Worst-node precisions (in the constraint basis) for SCT barrel networks as calculated with SCAN-3. The precisions are quoted in multiples of the common FSI 1D measurement precision. For the longitudinal measurements “double” and “triple” refer to the number of measurements (1 radial + 1 or 2 cross-braces). There are m measurements, n parameters, b constraints, and the redundancy is $r = m - n - b$.*

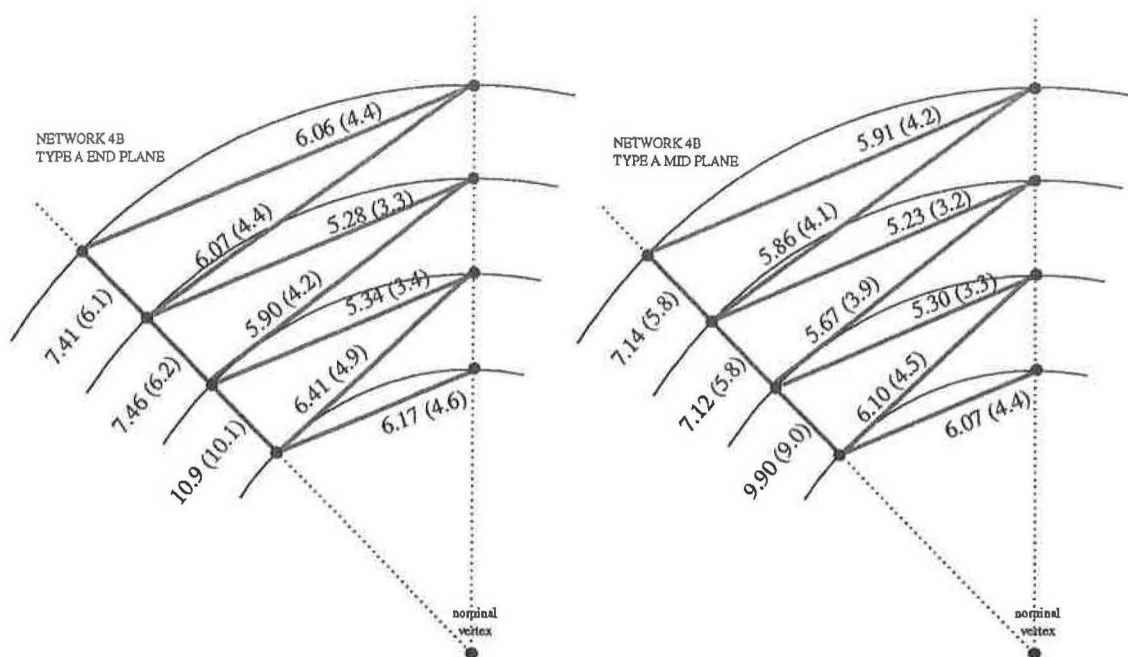


Figure 4.7: Results of the SCAN-3 MDB (BNR) calculation for the Type A (double cross-brace) transverse end plane and mid plane ($\gamma = 0.8$). For clarity, only one cross-brace of each pair is shown.

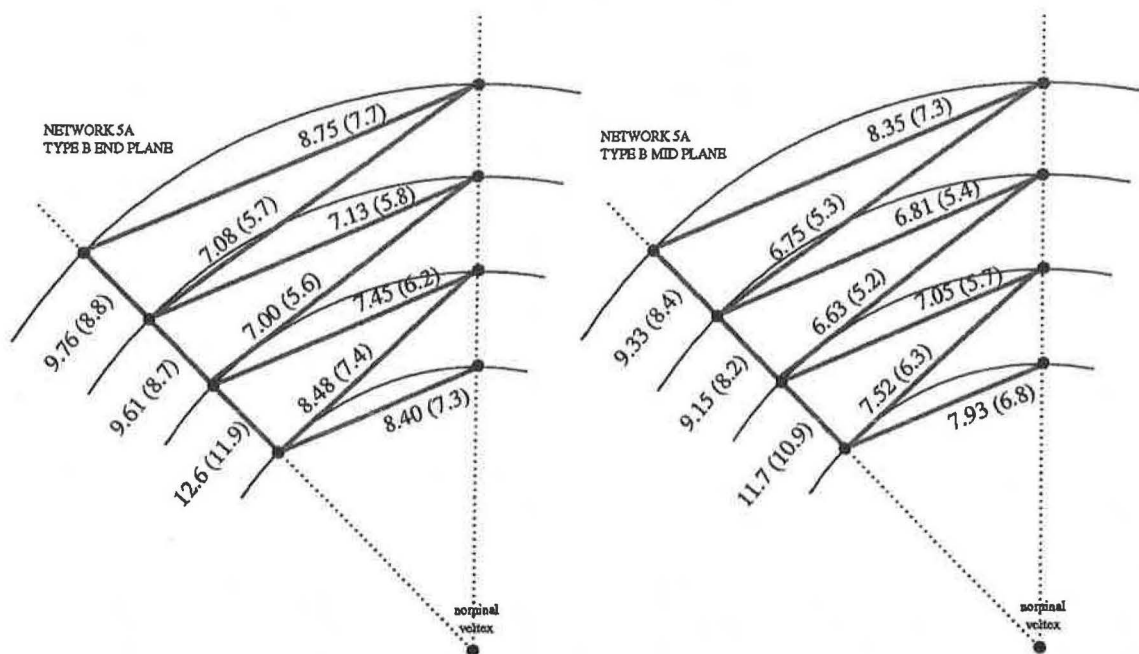


Figure 4.8: Results of the SCAN-3 MDB (BNR) calculation for the Type B (single cross-brace) transverse end plane and mid plane ($\gamma = 0.8$).

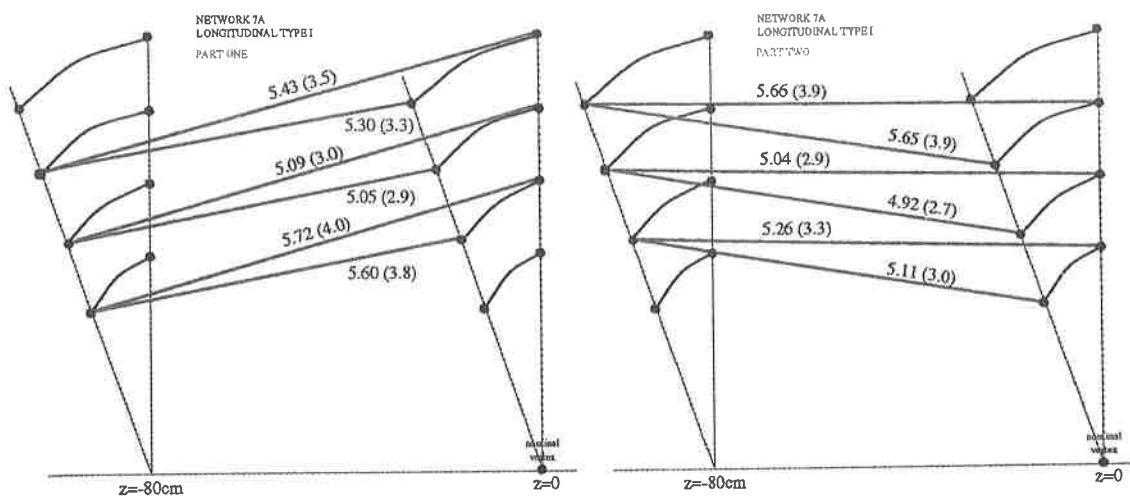


Figure 4.9: Results of the SCAN-3 MDB (BNR) calculation for the Type I(triple) longitudinal measurement configuration ($\gamma = 0.8$).

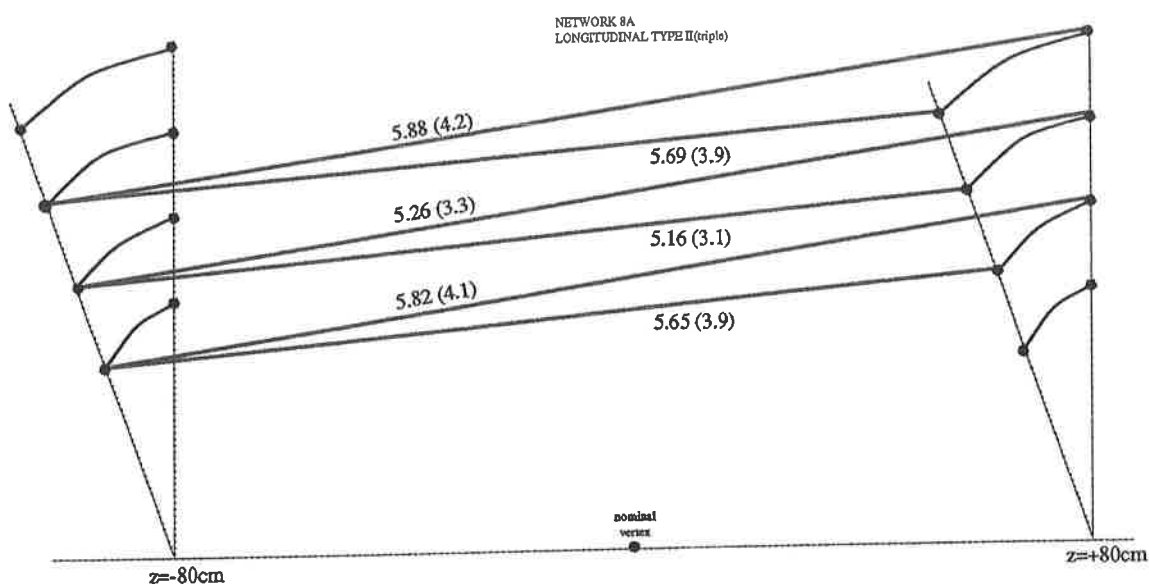


Figure 4.10: Results of the SCAN-3 MDB (BNR) calculation for the Type II(triple) longitudinal measurement configuration ($\gamma = 0.8$). Other measurements are the same by symmetry.

4.6 Conclusions

The use of any measurement technique for alignment requires a carefully planned combination strategy. A geodetic alignment network is proposed for the ATLAS Inner Detector. In this chapter, the mathematics used for performance analysis of a geodetic network and processing of recorded data has been introduced and results of precision and reliability calculations for a number of FSI networks have been presented. Initial indications are that the specifications can be met with a raw length measurement precision of $1\mu\text{m}$ for all measurements and a network such as those analysed above.

It was seen that the reliability requirements are met comfortably with the FSI networks analysed here, and so these networks are sufficiently reliable that optimisation could be performed simply on the basis of their precisions.

In the near future, a grid layout will be designed with more realistic mechanical constraints. Optimisation of the design will be continued as more information becomes available about the positioning of services (and hence the available lines of sight), and as a fuller understanding is developed of the stability and deformation modes of the support structure and the modules (and hence of the survey requirements). It is expected that the next iteration will include alignment grids for the forward precision layers and the pixel layers. In the longer term it may be possible to extend the calculations to include the Muon Spectrometer alignment system.

In the next step in the development of the FSI system, a high performance tunable laser will be used to demonstrate measurements at the required precision. This will provide information about the expected variation of the FSI precision with the distance measured, which could then be included in the network model.

Eventually the calculations in SCAN-3 will become part of the ATLAS offline software and will be augmented with code to interpolate between nodes to find individual detector element positions.

References

- [4.1] SCAN-3, System for Analysis of Geodetic Networks, Release 2.1, August 1995, Geodetic Computing Centre, Delft University of Technology, the Netherlands.
- [4.2] C. Daum, *Alignment of the ATLAS Inner Detector*, ATLAS Collaboration Internal Note INDET-NO-60, 12 September 1994.
- [4.3] C. Daum, *Alignment of the ATLAS detector from Inner Detector to Muon Spectrometer*, ATLAS Collaboration Internal Note TECH-NO-13, 7 November 1994.
- [4.4] C. Daum, *Least square fit and covariance matrix for an octagonal alignment grid for the ATLAS detector*, ATLAS Collaboration Internal Note TECH-NO-14, 1 February 1995.



Chapter 5

Monte Carlo study of track fitting and physics performance vs. alignment

In this chapter the effects of detector misalignment are investigated in detail by means of a full simulation of the ATLAS detector. The results are presented in two parts. In the first, the variation of the resolutions obtained by the track fitting algorithm are evaluated as a function of the alignment precision of the barrel pixel and silicon layers of the Inner Detector. In the second part, the same full simulation technique is applied to investigate the effect of alignment inaccuracies on the sensitivity of ATLAS to the forward-backward asymmetry of a new heavy gauge boson.

5.1 Software

5.1.1 ATLAS Monte Carlo environment

The ATLAS Monte Carlo software is based on GEANT and provides a system for detailed simulation and reconstruction of LHC events within the ATLAS detector. It is divided into three parts : SLUG, DICE and ATRECON.

- SLUG (Simulations for LHC Using GEANT) is the main program of the simulation system. It contains an interface to GEANT, handling all the low level tasks such as initialising ZEBRA, GEANT and HBOOK, performing input/output and basic detector graphics.

- DICE (Detector Integration for a Collider Experiment) is a package of routines containing all the ATLAS-specific code required to simulate and digitise events. It provides an interface to several particle generators, including the general-purpose event generator PYTHIA. The Inner Detector layout currently implemented is that of the Technical Proposal.
- ATRECON is a package of reconstruction code. The pattern recognition and track fitting for the study presented here were done using a modified version of the *iPatRec* routine. This employs a track finding algorithm which uses information from both the discrete detectors (silicon and pixels) and the continuous tracking detector (TRT).

5.1.2 *iPatRec*

iPatRec is a road-based track finding and fitting package for the reconstruction of charged tracks in the ATLAS Inner Detector [10]. Track identification is not restricted to isolated tracks but may also be performed within jets and (slowly) in the presence of the minimum bias pile-up expected at the highest LHC luminosities.

The pattern recognition algorithm is described in Figure 5.1. A ‘seed’ direction for the track search is sourced from the calorimeter or Muon Spectrometer code, or (cheating) from generator-level track information. The seed parameters are used to define a ‘road’ in real and momentum space. This includes the full z -extent of the vertex. Starting in the low-occupancy Si-barrel layers, track hits in a road are associated in a combinatorial fashion to form ‘skeleton’ tracks. Skeleton tracks not matching the road parameters (within cuts) are rejected. Following this, a histogram-based pattern recognition stage is initiated in TRT roads with a transverse width of approximately 1 mm. The results of the TRT stage are used to veto wrong combinations of Si hits and to solve the left-right ambiguity of the TRT drift time measurement. A candidate track is then fitted and finally accepted as a found track if the quality of the fit meets certain criteria.

iPatRec Si+TRT track finding algorithm

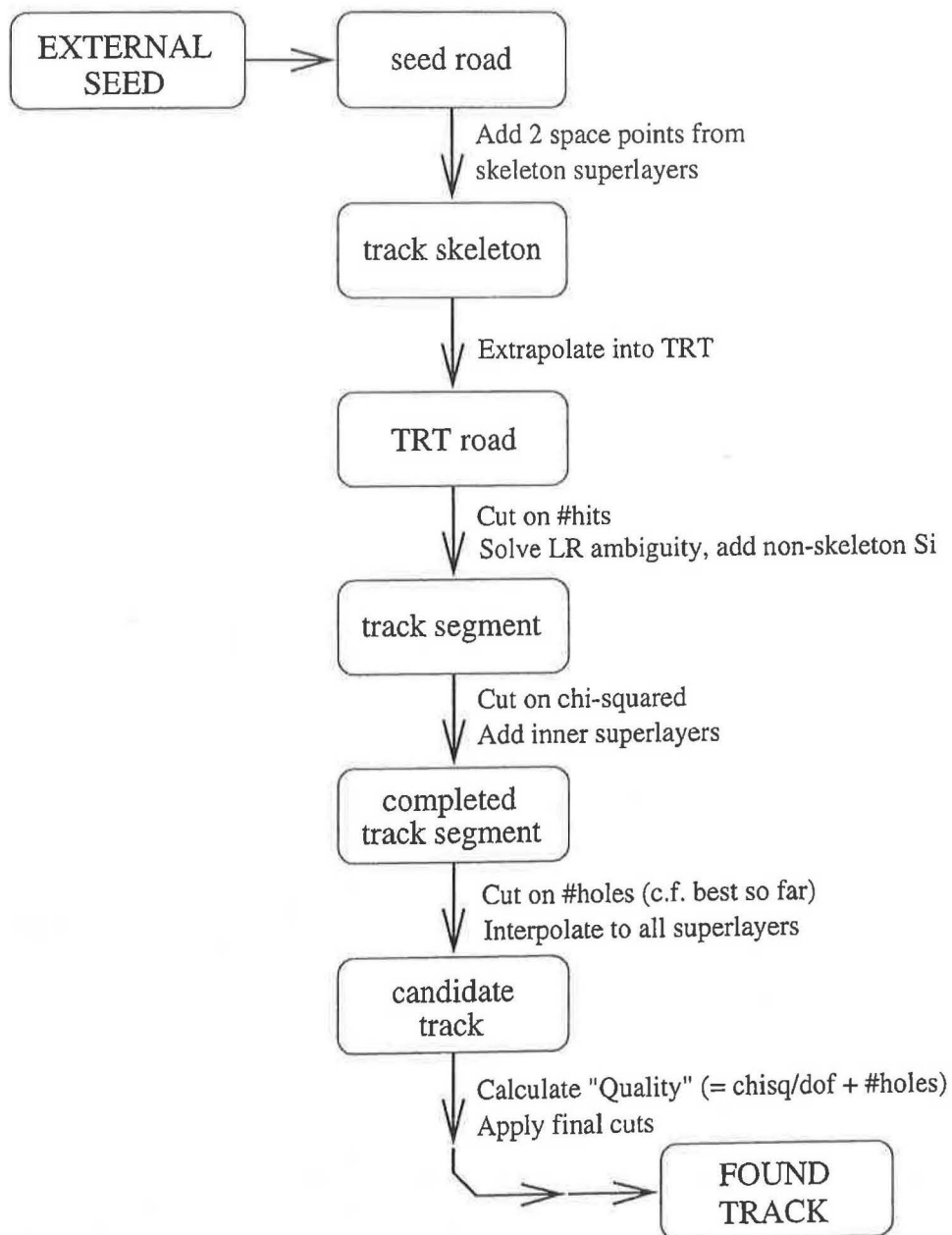


Figure 5.1: The iPatRectrack finding algorithm

5.2 Track fitting resolutions and alignment

5.2.1 Introduction

The precision with which the position of a hit on a track is known involves a combination of the intrinsic precision of the detector and the precision with which its spatial position has been determined. As explained in previous chapters, a survey system uses direct measurements to reduce the effect of the assembly tolerances and the detector stability on the location precision. The requirement which defines the specifications of such a survey system is that the total effective spatial precision of the hits on a track is sufficient to allow determination of the five helix parameters with precisions meeting the physics specifications.

Although a suitably defined analytic technique could be used to calculate the track fitting accuracies as a function of the alignment precisions *given the points to be fitted*, a full simulation is required to take the pattern recognition algorithm into account, since the performance of a such an algorithm (and all the combining, cutting and fitting that this implies) is in general too complex for an analytic calculation. In addition, the performance depends on the detailed structure of the events to be analysed.

For these reasons a full Monte Carlo simulation technique has been used to investigate how the error on the determined track parameters varies as a function of the alignment precisions of elements of the ATLAS Inner Detector.

A note on notation: In the ATLAS collaboration, it is conventional to use a 3D cylindrical coordinate system (r, ϕ, z) , in which the z -axis is collinear with the beam pipe, and the r and ϕ directions are respectively radial and azimuthal with respect to the z -axis. The five helix parameters of a track are taken to be the transverse momentum (p_T), the 2D impact parameter (a_0), the initial angle with respect to the z -axis (θ_0), and the initial values of the coordinates ϕ and z . In order to avoid confusion with the helix parameters, in the following sections, parameters related to alignment will be labelled using the upper case letters R , Φ and Z .

5.2.2 Technique

The simulation environment described in Section 5.1 was used, with the standard particle generation, detector simulation and digitisation. The reconstruction was performed using a version of *iPatRec* which was modified to permit the simulation of misalignments.

In order to factorise the alignment analysis into manageable components, it was

thought expedient for this first analysis to begin with a simulation of the effects of misalignment of only the barrel SCT layers. Another factor in this decision was the fact that the forward layer positions and resolutions currently implemented in the software (those of the TP layout) differ considerably from those of the current (fine-tuned Morges) layout. By contrast, the barrels of all Panel-style designs (see Chapter 6) are rather similar.

The technique used to simulate the effect of lack of knowledge of the element positions due to finite alignment precision was to randomly dither the positions of the SCT superlayers ¹ according to Gaussian distributions with widths corresponding to the alignment precisions. A set of three Gaussian widths $\sigma_R, \sigma_\Phi, \sigma_Z$ was defined for each of the six superlayers in the datacard file. A new (R, Φ, Z) position was selected from the defined distributions for each of the superlayers for each event.

Since the intention was to use only single track events (single muons in this section and single electrons from $W' \rightarrow e\nu$ in Section 5.3) it was not considered necessary to move different modules within a superlayer by different amounts since, ignoring overlaps, a single track only crosses one module per superlayer. Superlayers were moved rather than single layers because the positions of the elements in the two layers of a back-to-back pair will be strongly correlated.

In order to investigate the three spatial components of the alignment in the R, Φ and Z directions separately, only one of the three coordinates was dithered at a time, the other two Gaussian widths being set to zero. The five track-fitting resolutions were then evaluated as a function of the widths of the dither distributions.

Pile-up was not included in this study because the very long reconstruction time needed for events with pile-up (200 seconds per event) would not have permitted high enough statistics in the time available. The effect of pile-up on the Inner Detector performance has been investigated with low statistics [10]. A degradation of the p_T resolution of about 10% was seen in the barrel at 500 GeV. Further development of the reconstruction software may yield sufficient improvement in execution time to permit an alignment study to be done with pile-up in the future.

5.2.3 Simulation results

The technique described above was applied to three samples of 10,000 single muon events with transverse momenta of 10, 500 and 3000 GeV. These momenta were chosen

¹Superlayer : for the pixels a single layer, and for the silicon strips, a stereo pair of layers.

in order to span most of the useful momentum range of the ATLAS Inner Detector, including the high- p_T range considered in the asymmetry study in Section 5.3.

The track fitting resolutions given below are averaged over the pseudorapidity range $|\eta| < 0.7$. All resolutions except the impact parameter resolution σ_{a_0} were obtained using vertex constrained fits. A beam constraint is inappropriate for evaluation of the impact parameter resolution since this is most important for locating secondary vertices.

The results for all three track momenta are displayed in the same plots, normalised to their values for perfect alignment. These are given in Table 5.1 for reference.

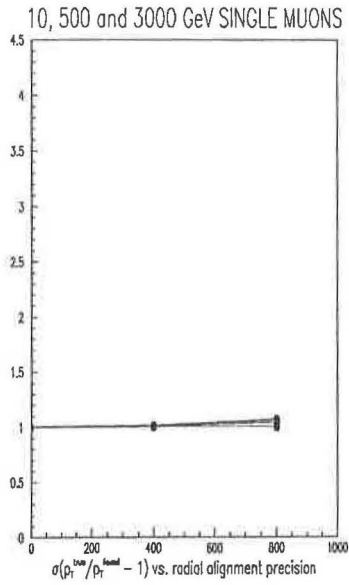
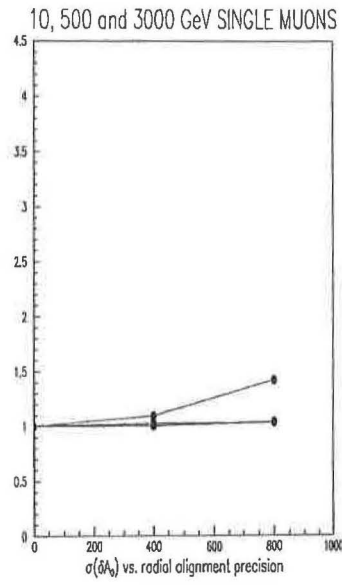
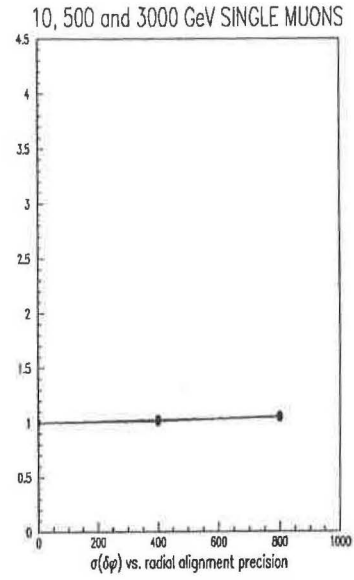
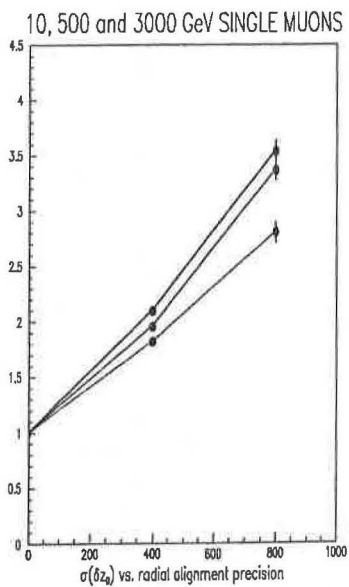
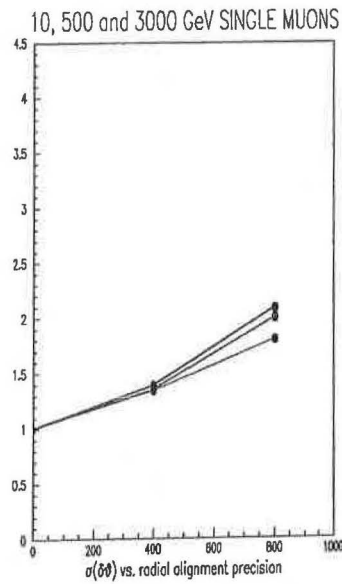
Resolution	unit	10 GeV	500 GeV	3000 GeV
p_T	%	1.52	15.3	93.1
a_0	μm	34.9	16.3	16.7
ϕ	mrad	0.177	0.0714	0.0734
z_0	μm	108	98.7	97.4
θ	mrad	0.600	0.516	0.500

Table 5.1: *Perfect alignment track fitting resolutions obtained with the iPatRec pattern recognition routine for single muon tracks with transverse momenta of 10, 500, 3000 GeV.*

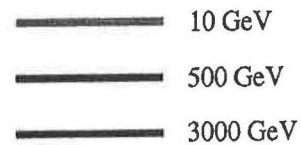
Whereas the ϕ -resolutions of the barrel pixel and strip detectors are comparable, the z -resolutions of the barrel pixels and barrel silicon strip detectors are quite different: $770\mu\text{m}$ and $300/\sqrt{12} = 87\mu\text{m}$ respectively. In order to investigate the relative importance of the Z -alignment of the pixels and the silicon strips, σ_Z was varied in three different ways : firstly for with equal distribution widths for all layers as in the previous sections, then for pixels only (keeping $\sigma_Z = 0$ for strips) and lastly for strips only (keeping $\sigma_Z = 0$ for pixels).

5.2.3.1 R-alignment

In the barrel, the r -coordinate of track hits is obtained purely by virtue of the radial positioning of the layers. Results are given below for $\sigma_R = 0, 400, 800 \mu\text{m}$, the radial alignment precision of each of the six barrel precision superlayers being taken as equal.

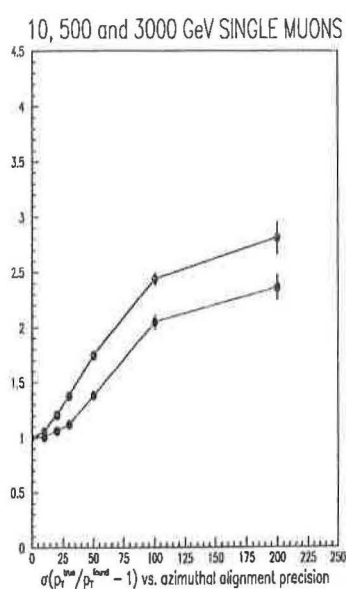
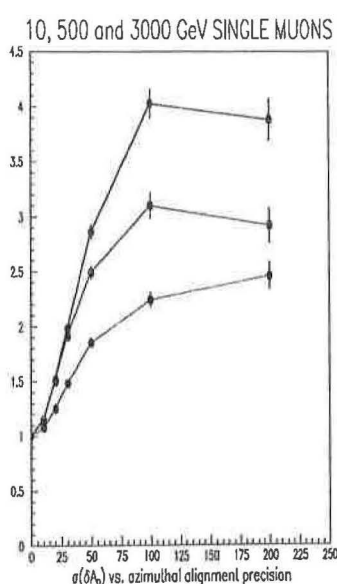
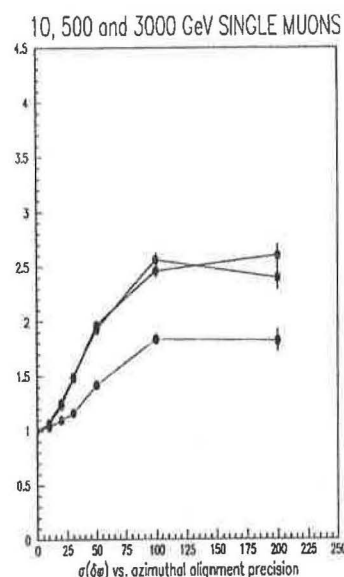
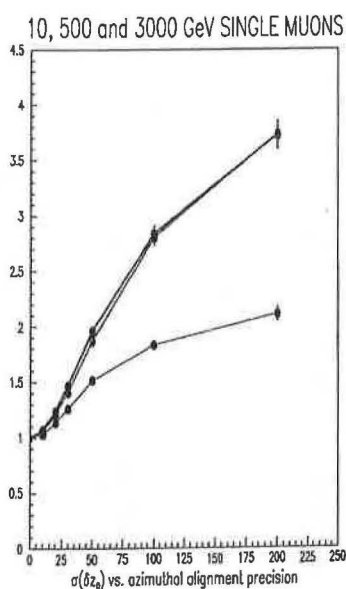
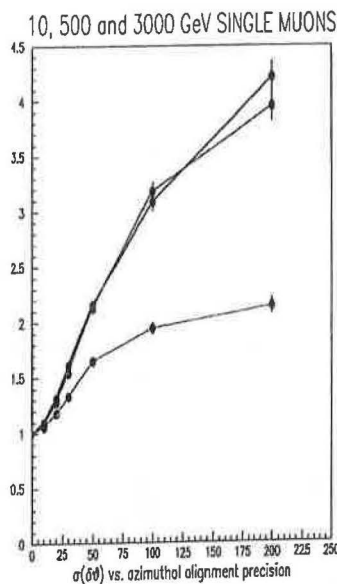
Figure 5.2: σ_{p_T} vs. σ_R Figure 5.3: σ_{a_0} vs. σ_R Figure 5.4: σ_ϕ vs. σ_R Figure 5.5: σ_{z_0} vs. σ_R Figure 5.6: σ_θ vs. σ_R

Key :

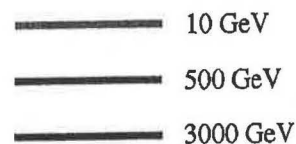


5.2.3.2 Φ -alignment

In an effort to model a plausible alignment scenario, the value of $R\sigma_\Phi$ (rather than σ_Φ say) was chosen to be the same for all layers. Results are given below for $R\sigma_\Phi = 0, 10, 20, 30, 50, 100, 200 \mu\text{m}$. When generating Figure 5.7, it was found that for 3000 GeV tracks, the p_T resolution became too poor to permit a valid Gaussian residual fit and so the corresponding line is not shown.

Figure 5.7: σ_{p_T} vs. $R\sigma_\Phi$ Figure 5.8: σ_{a_0} vs. $R\sigma_\Phi$ Figure 5.9: σ_ϕ vs. $R\sigma_\Phi$ Figure 5.10: σ_{z_0} vs. $R\sigma_\Phi$ Figure 5.11: σ_θ vs. $R\sigma_\Phi$

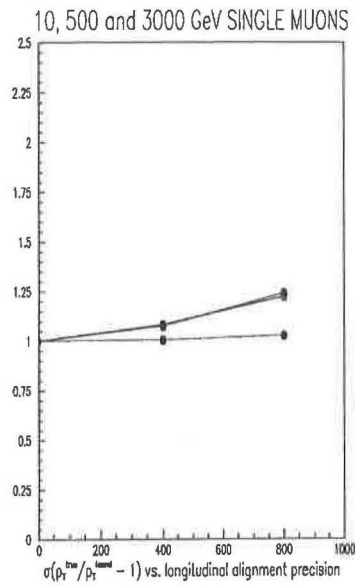
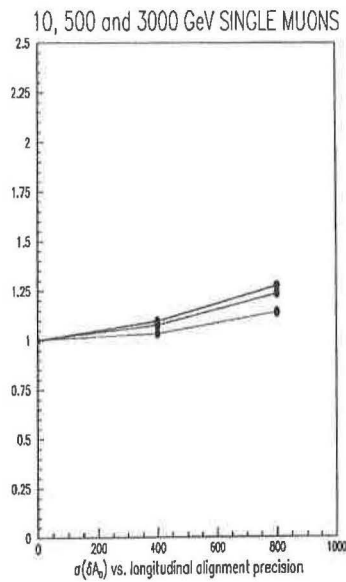
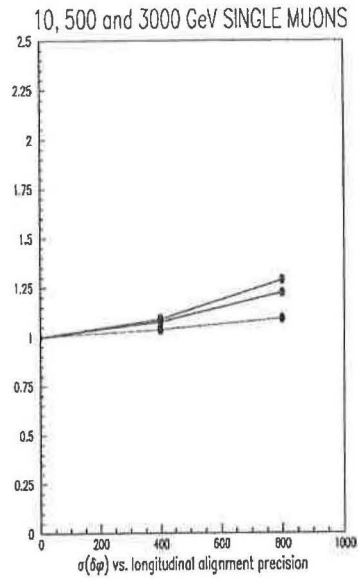
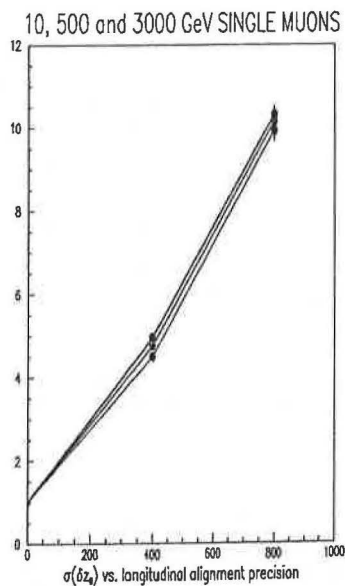
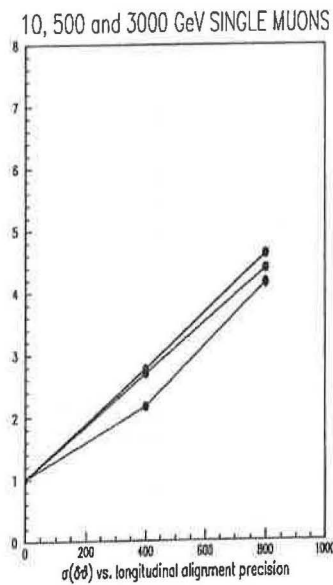
Key :



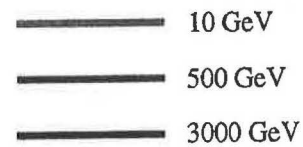
5.2.3.3 Z-alignment

5.2.3.3.1 Pixels and strips dithered equally

Results are given below for the fitting resolutions as a function of the Z-alignment precision of the barrel pixels and silicon strips, all 6 layers having the same distribution widths.

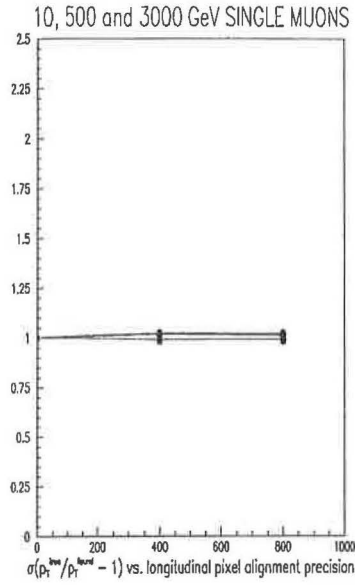
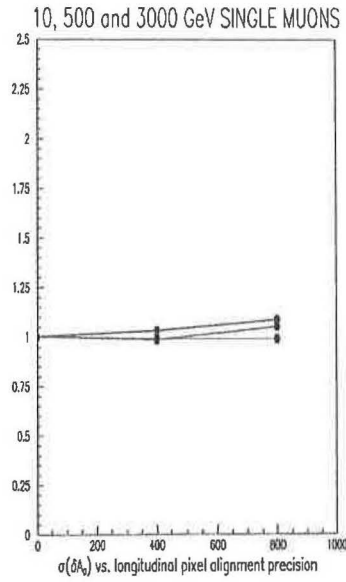
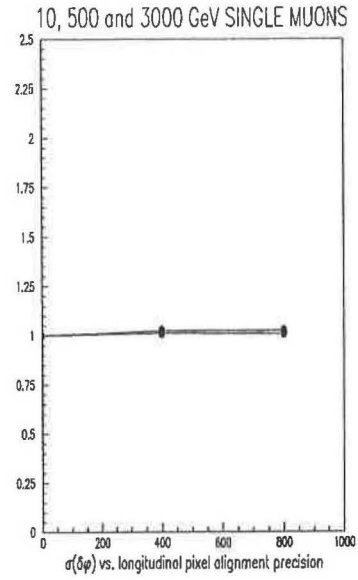
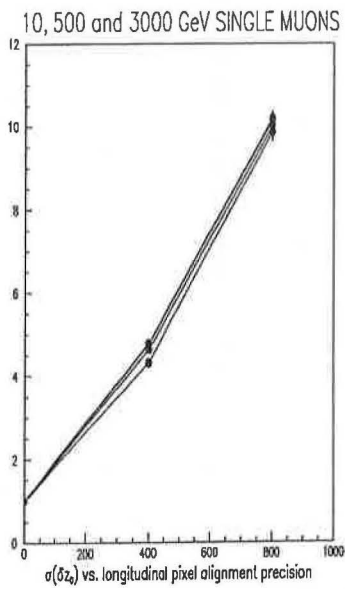
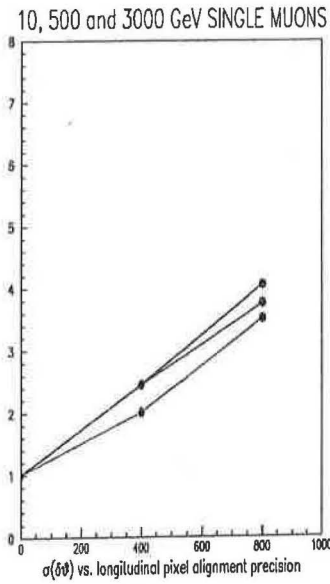
Figure 5.12: σ_{p_T} vs. σ_Z Figure 5.13: σ_{a_0} vs. σ_Z Figure 5.14: σ_ϕ vs. σ_Z Figure 5.15: σ_{z_0} vs. σ_Z Figure 5.16: σ_θ vs. σ_Z

Key :

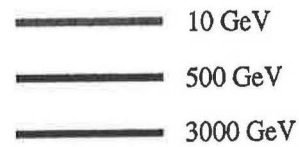


5.2.3.3.2 Dither pixels only

Results are given below for the fitting resolutions as a function of the Z -alignment precision of the two pixel layers, both layers having the same distribution widths. The silicon strips were assumed to be perfectly aligned.

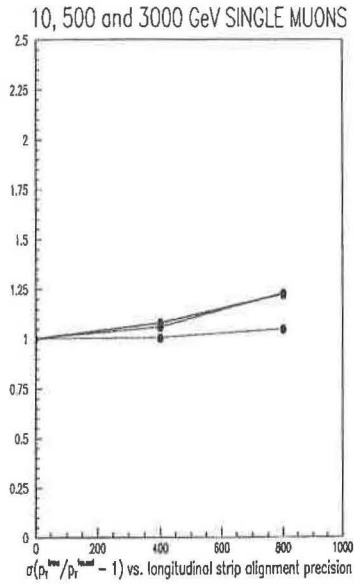
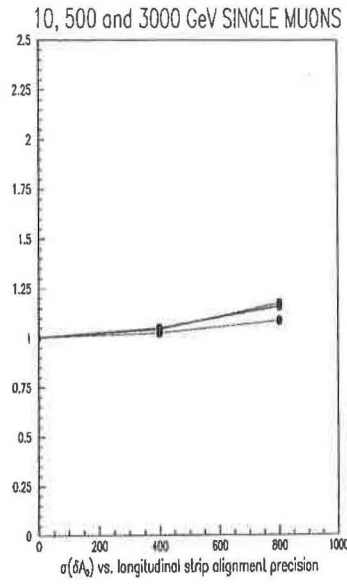
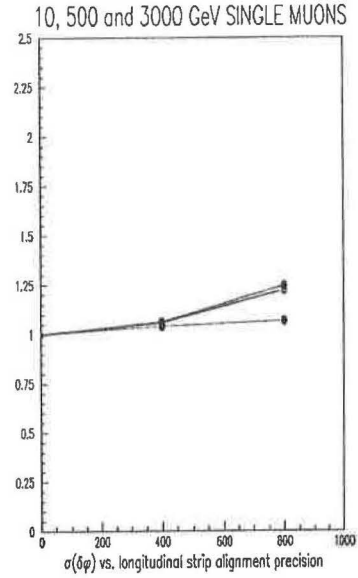
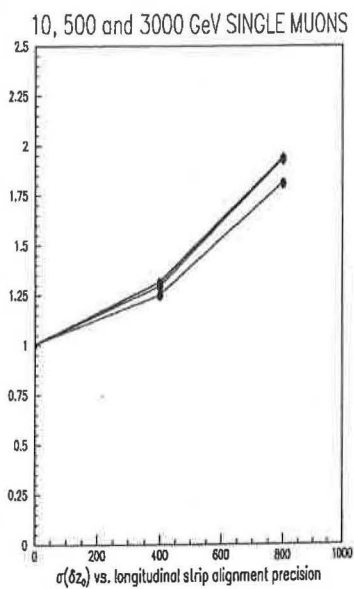
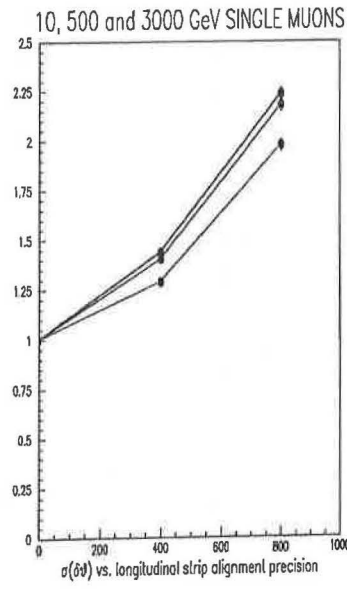
Figure 5.17: σ_{p_T} vs. σ_Z^{pix} Figure 5.18: σ_{a_0} vs. σ_Z^{pix} Figure 5.19: σ_ϕ vs. σ_Z^{pix} Figure 5.20: σ_{z_0} vs. σ_Z^{pix} Figure 5.21: σ_θ vs. σ_Z^{pix}

Key :

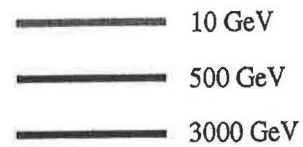


5.2.3.3.3 Dither silicon strips only

Results are given below for the fitting resolutions as a function of the Z -alignment precision of the four silicon superlayers, all layers having the same distribution widths. The pixels were assumed to be perfectly aligned.

Figure 5.22: σ_{p_T} vs. σ_Z^{Si} Figure 5.23: σ_{a_0} vs. σ_Z^{Si} Figure 5.24: σ_ϕ vs. σ_Z^{Si} Figure 5.25: σ_{z_0} vs. σ_Z^{Si} Figure 5.26: σ_θ vs. σ_Z^{Si}

Key :



5.2.4 Discussion

Before interpreting the results presented above, it is worth reviewing the requirements for the track fitting resolutions [9]:

p_T and ϕ resolutions

It is desirable to have as high a momentum resolution as possible in order to be maximally sensitive to unexpected signals and to reduce backgrounds as far as possible. It must be remembered however, that the calorimeter and muon systems will inevitably outperform the Inner Detector for momentum measurements over much of the momentum range. A requirement comes from asymmetry measurements of the decay $W' \rightarrow e\nu$, investigated later in this chapter, which require the identification of the charge sign of very high energy leptons.

θ resolution

The specification comes from the B meson mass resolution in the channel $B_d^0 \rightarrow J/\Psi K_s^0$. The specification for the Inner Detector is 2.0 mrad. Better resolution than this will slightly reduce the background under the mass peak. With perfect alignment, the resolution obtained with the TP layout is 0.6 mrad at 10 GeV, 3.3 times smaller than the specification.

z resolution

In order to help separate primary vertices from pile-up, the resolution of the z -coordinate of the primary interaction should be better than 1 mm. B-physics mass resolutions place a similar requirement on the z resolution. A point with a z precision of less than $300\mu\text{m}$ at any radius inside the calorimeter will improve the resolution of the Muon Spectrometer. However, the standalone performance of the muon system is considered sufficient for all physics processes. The resolution obtained with a perfectly aligned TP layout was found to be about 0.1 mm.

a_0 resolution

Good impact parameter resolution is required for a variety of physics topics. At high luminosity these include b-tagging both for top-quark physics and for rejection of backgrounds to $H \rightarrow 4l$ and searches for new physics. There is no optimum specification: the rejection obtained improves monotonically with improved a_0 resolution, and hence this should be as low as possible. Studies have assumed $\sigma_{a_0} \sim 30\text{--}40\mu\text{m}$. The perfectly aligned TP layout gives a resolution of $35\mu\text{m}$.

A note on TRT alignment

In the sense that there will inevitably be alignment errors in the TRT, the results given here can be considered optimistic as far as the TRT precision is concerned. However, since the TRT measures only r and ϕ coordinates and not z , and since it is a relatively low precision detector at high radius, its influence on the results of this study is limited to the p_T resolution and the Φ -alignment. The reasons are as follows:

- Since the barrel TRT measures only r and ϕ coordinates and not z , the alignment of the TRT has no influence on the z resolution.
- With z and r positions of the layers of the order of tens of centimetres, a change in r of the order of tens of micrometres changes the angle θ by an amount much less than the required precision and so TRT alignment errors will not affect the θ resolution.
- Since good impact parameter resolution requires high-precision $r\phi$ measurements at low radius, the a_0 resolution is dominated by the pixels and inner silicon layers, and will not be influenced strongly by misalignments in the TRT.

Because the perfectly aligned TRT gives unrealistically good momentum performance, any Φ -alignment requirement derived from the results must be considered as an upper limit.

The p_T resolution obtained for perfect alignment was slightly better than the corresponding resolution given in the ATLAS Technical Proposal, for which an estimated alignment component was added to the intrinsic $r\phi$ precision of the TRT.

Overview of results

In general the variation of fitting resolutions with alignment precision is stronger for high- p_T tracks low- p_T ones, the exception being the dependence of the impact parameter resolution on radial alignment, which is strongest for low-momentum tracks.

In the following, resolutions will often be referred to in terms of the increase (i.e. degradation) compared to the perfectly aligned detector. A 20% degradation, for example, corresponds to a y -value of 1.20 in the figures given above.

R -alignment

R -alignment affects primarily the z_0 and θ resolutions. The 20% degradation point is $\sigma_R = 70\,\mu\text{m}$ for the z_0 resolution and $\sigma_R = 200\,\mu\text{m}$ for the θ resolution. The

effect of R -alignment on the other resolutions is small, the most significant being the a_0 resolution for the low-momentum 10 GeV track which degrades by 20% at $\sigma_R \sim 530 \mu\text{m}$.

Φ -alignment

Accurate Φ -alignment is essential for all the track fitting resolutions, of which the requirement for the 2D impact parameter resolution is the most stringent, at $R\sigma_\Phi = 11 \mu\text{m}$ for 20% degradation. From Figures 5.7 to 5.11 it can be seen that the resolutions tend towards the TRT standalone resolutions as the alignment of the precision layers is degraded. The TRT standalone precision is reached for the a_0 and ϕ resolutions at $R\sigma_\Phi \sim 100 \mu\text{m}$.

Note that the z_0 resolution is affected by the Φ -alignment of the silicon layers. This is because of the use of stereo strips (see below).

Z -alignment

Z -alignment affects the z_0 and θ resolutions most strongly, the other resolutions being only weakly dependent on σ_Z . The figures for the three types of Z -alignment considered show that Z -alignment of the pixels has a much greater impact than the Z -alignment of the strips (as expected, since the pixels have a much better z -precision than the strips).

Note that the p_T , ϕ and a_0 resolutions are affected by the Z -alignment of the silicon layers. This is because of the use of stereo strips (see below).

A note on stereo strips

The four silicon superlayers consist of pairs of layers : one with strips running parallel to the z -axis and the other with a small stereo angle. In the TP layout used in the simulation, the stereo angle is $\alpha = 40 \text{ mrad}$ giving a z resolution of $\sigma_z = 770 \mu\text{m}$. The $r\phi$ resolution of both layers is $\sigma_{r\phi} = 15 \mu\text{m}$.

Although misalignments of the axial strips in the Φ -direction do not affect the longitudinal track fitting resolutions (z_0 and θ), the Φ -alignment of the stereo strips affects them strongly, since an azimuthal alignment error $R\Delta\Phi$ causes longitudinal measurement error of $\Delta z = R\Delta\Phi / \tan \alpha = 25R\Delta\Phi$. See Figures 5.10 to 5.11.

Similarly, the precision of the Z -alignment of the stereo strips affects the transverse track fitting resolutions (p_T , ϕ and a_0). In this case the dependence is weaker, a Z -alignment error of ΔZ giving an azimuthal error of $r\Delta\phi = \Delta Z \tan \alpha = \Delta Z / 25$.

See Figures 5.22 to 5.24 (in which the Z -alignment of the silicon layers is varied), and compare with Figures 5.22 to 5.24 (in which the Z -alignment of the pixel layers is varied).

Thus Φ -alignment errors of the order of the (small) $r\phi$ resolution are amplified to become z fitting errors comparable to the (large) z resolution. Similarly, Z -alignment errors of the order of the required z precision cause azimuthal fitting errors of the order of the required $r\phi$ precision. Thus it is possible that a meaningful alignment specification in the Z -direction may come from the specification for p_T resolution and that a meaningful Φ -alignment specification may be derived from the low z -resolution specification.

In passing, it is worth mentioning that since the motions of the two layers in a superlayer will be almost identical, if the strips were oriented to measure the u and v coordinates rather than u and ϕ coordinates (u and v strips having stereo angles $+\alpha$ and $-\alpha$ with respect to the z -axis), the fitting biases caused by a misalignment of a module in either the Φ or Z direction would be in opposite directions for the two layers in a module, so cancelling the effect of the misalignment when fitting a track which produced a hit in both layers.

Interpretation

The results are interpreted below in two ways:

- In the first section the results are interpreted under the assumption that the detector performance should be as good as possible, and that the detector should be aligned well enough that the resolutions are close to the perfect alignment scenario. The alignment requirements for a degradation of no more than 20% and 50% compared to the perfectly aligned detector are given.
- The second interpretation is that the alignment requirements are set merely by the need to meet the specifications given above.

(a) Minimum degradation interpretation

The alignment precisions for which each of the track fitting resolutions is degraded by 20% and 50% compared to the perfect alignment situation are shown for the R , Φ and Z -alignments in Tables 5.2 to 5.4. If one insists that none of the tracking resolutions may be allowed to be degraded by more than 20% or 50% from its perfect alignment value then one obtains the set of alignment requirements given in Table 5.5. Where

an entry is not given, the figure is beyond the maximum value used in this study, i.e. greater than $800\text{ }\mu\text{m}$ for σ_R and σ_Z , or $200\text{ }\mu\text{m}$ for σ_Φ .

resolution	degraded by 20% at $\sigma_R = (\text{ }\mu\text{m})$	degraded by 50% at $\sigma_R = (\text{ }\mu\text{m})$
a_0	530 (10 GeV track)	—
θ	200	470
ϕ	—	—
z_0	70	180
p_T	—	—

Table 5.2: Effect of R -alignment on the five track-fitting resolutions. The values of σ_R for which each resolution is degraded by 20% and 50% are given.

resolution	degraded by 20% at $R\sigma_\Phi = (\text{ }\mu\text{m})$	degraded by 50% at $R\sigma_\Phi = (\text{ }\mu\text{m})$
a_0	11	20
θ	$14^{(s)}$	28
ϕ	17	30
z_0	$17^{(s)}$	32
p_T	20	36

Table 5.3: Effect of Φ -alignment on the five track-fitting resolutions. The values of $R\sigma_\Phi$ for which each resolution is degraded by 20% and 50% are given. $^{(s)}$ indicates that the dependence is a consequence of the stereo angle of the silicon strips.

resolution	degraded by 20% at $\sigma_Z = (\mu\text{m})$	degraded by 50% at $\sigma_Z = (\mu\text{m})$
a_0	650 ^(s)	—
θ	40 (pix)	120
ϕ	640 ^(s)	—
z_0	20 (pix)	50
p_T	720 ^(s)	—

Table 5.4: *Effect of Z-alignment (all layers) on the five track-fitting resolutions. The values of σ_Z for which each resolution is degraded by 20% and 50% are given. ^(s) indicates that the dependence is a consequence of the stereo angle of the silicon strips.*

alignment type	20% (μm)	50% value (μm)	set by resolution
R	70	180	z_0
Φ	11	20	a_0
Z (all)	20	50	z_0
Z (pixels)	20	50	z_0
Z (strips)	180	430	θ

Table 5.5: *Requirements for the alignment of the barrel precision layers in the R , Φ and Z directions obtained by insisting that no track fitting resolution may degrade by more than 20% or 50% from its perfect alignment value.*

(b) “Just meet specs” interpretation

A second way of looking at the alignment requirements is to determine the points at which the resolutions pass their specifications. For the specifications given above, the alignment requirements given in Table 5.6 are obtained.

resolution	specification	perfect \times	R	Φ	Z
a_0	max. 20% degrad.	1.20	530	11	650 ^(s)
θ	2 mrad	3.3	> 800	115	520(all)/610(pix)
z_0	0.3[1.0] mm	3.0[10]	650[> 800]	120[> 200]	200(pix)[780(pix)]
p_T	30% for 500 GeV	1.9	> 800	< 60	> 800

Table 5.6: *Maximum barrel SCT alignment precisions compatible with meeting the tracker specifications. Since there is no lower limit for the desired impact parameter resolution, the 20% degradation point is given. The $60\mu\text{m}$ requirement from p_T resolution for the Φ -alignment precision must be considered as an upper limit, since misalignments of the TRT were not included in this study. ^(s) indicates that the dependence is a consequence of the stereo angle of the silicon strips.*

5.2.5 Conclusions

Clearly the Φ -alignment sets the most stringent requirements for the Inner Detector survey system. A Φ -alignment requirement of $11\mu\text{m}$ is set by insisting that the impact parameter resolution is degraded by no more than 20% from its value for perfect alignment. Consideration of the specification for p_T resolution with a perfectly aligned TRT sets a upper limit of $60\mu\text{m}$. Since the precision layers and the TRT make comparable contributions to the p_T resolution (standalone figures at 500 GeV of $\sim 25\%$ and $\sim 46\%$ respectively), it seems probable that when the TRT alignment is taken into consideration, the Φ -alignment requirement from p_T resolution will be reduced considerably.

A Z -alignment requirement of $520\mu\text{m}$ for all layers (pixel and strips) is set by the θ resolution specification. Alignment of the pixels with a precision of $\sigma_Z^{\text{pix}} = 200\mu\text{m}$ would result in a z_0 resolution sufficient to improve the performance of the Muon Spectrometer. As shown in Table 5.4, the Z -alignment of the silicon layers affects ϕ , p_T and a_0 resolutions.

An R -alignment requirement of $530\ \mu\text{m}$ is set by insisting that the impact parameter resolution is degraded by no more than 20% from its value for perfect alignment.

It should be borne in mind that misalignments were considered in only one of the three spatial directions at a time. In reality any given element of the detector has three alignment precision components. The combined effect of these has not been considered. Thus it remains to be seen what would be the effect on the z_0 resolution of, for example, simultaneous alignment errors of $\sigma_R = 70\ \mu\text{m}$, $R\sigma_\Phi = 17\ \mu\text{m}$ and $\sigma_Z = 20\ \mu\text{m}$, each of which alone would give 20% degradation compared to the perfectly aligned detector. Investigation of 3-component alignment precisions would involve no changes to the software, and so it is possible that this study could be conducted in the near future

5.3 W' asymmetry signal vs. alignment

According to current theoretical predictions, a new heavy analogue of the SM charged vector boson, generically named the W' ² could be very heavy. Current cross-section estimates allow its discovery up to a mass of $m_W \sim 5 \text{ TeV}$ in one year of running at the LHC. The forward-backward asymmetry of its semi-leptonic decay is important for model discrimination, but requires the identification of the charge sign of a lepton with a p_T of up to $\sim 3 \text{ TeV}$. In the electron channel $W' \rightarrow e\nu$, this may place a requirement on the momentum resolution of the Inner Detector and consequently on the Φ -alignment.

In this section the full simulation code of Section 5.2 is used to investigate the alignment requirements for good sensitivity of the ATLAS detector to this asymmetry signal. The assumption made previously that the alignment shift may be the same for all modules in a superlayer cylinder is still valid for this event since the event reconstruction involves only one charged track.

The specifications for the performance of the Inner Detector, including those of alignment, should ideally be derived directly from the physics requirements. This is one of the first studies to examine the performance as detector parameters are changed dynamically in a full simulation.

5.3.1 New gauge bosons

5.3.1.1 Models

Its undoubted phenomenological success at energies of order 10^2 GeV notwithstanding, the Standard Model suffers from a number of theoretical problems which indicate that it cannot be a complete description of the fundamental particles and their interactions. New neutral or charged vector bosons occur in many attempts to extend the gauge group of the SM because extra $U(1)$ and $SU(2)$ symmetries arise when larger groups are broken down to the SM gauge group $SU(3)_C \times SU(2)_L \times U(1)_Y$. The main classes of extended gauge models are as follows:

E_6 models :

E_6 models, whose roots lie in the ideas of grand unification, aim to solve the puzzle of the existence of three different gauge coupling constants associated with the gauge groups $SU(3)_C$, $SU(2)_L$ and $U(1)_Y$, i.e. the 'gauge problem'. An attractive

²Wherever there is a possibility of notational confusion with the Standard Model electroweak gauge bosons, W^\pm and Z^0 , the new gauge bosons will be written as V^\pm and V^0 instead of W'^\pm and Z'^0 .

possibility is offered by superstring theories from which E_6 may emerge as the low-energy (compared to the Planck scale) grand unified group. Since the breaking of E_6 into the SM is accompanied by at least one extra $U(1)$, one expects the existence of one or more heavy neutral bosons which can mix with the Standard Model Z . E_6 superstring-inspired models are relatively constrained since there are only a few possible breaking patterns of E_6 .

Left-right symmetric (LRS) models :

LRS models based on the group $SU(2)_L \times SU(2)_R \times U(1)_{B-L}$ are motivated by the attempt to understand the origin of parity violation and small neutrino masses. In minimal LR models where parity violation is spontaneously broken one predicts the existence of an additional neutral boson and two charged bosons, W_R^\pm , which mix with the left-handed Standard Model W s. Low-energy phenomenology and especially the $K_L - K_S$ mass difference constrain the mixing angle ξ and the W_R mass. A possible way to evade the mixing is to consider the alternative left-right models (ALRMs) from string theory.

The BESS model (for Breaking Electroweak Symmetry Strongly) :

This offers a mechanism of possible strong breaking of the electroweak symmetry, with a Higgs which is no longer a physical particle. Here the principle of local $SU(2)_L \times U(1)$ gauge invariance is maintained but the breaking of the electroweak symmetry is realised in a non-linear way. Weak interactions are described by an effective Lagrangian which is $SU(2)_L \times SU(2)_V \times U(1)$ invariant. A triplet of spin-1 resonances associated with the $SU(2)_V$ gauge group is predicted. This triplet is the analogue of the ρ meson in hadron physics.

Composite models :

Composite models take into account the possibility that the W and Z are not elementary. A typical model is one in which the the low-energy phenomenology of the SM is incorporated into an effective Lagrangian which is invariant under a global $SU(2)$ weak isospin symmetry and which satisfies local $U(1)$ QED-like gauge invariance. The (W^\pm, Z^0) triplet is coupled to fermions in an invariant way with strength g_w whereas its neutral component couples to the photon with strength λ_w such that $e = \lambda_w g_w$. In this model one expects a rich spectrum of new heavy bosons.

The Extended Gauge Model :

The Extended Gauge Model (EGM) [2] is not a purely physically motivated (i.e. ‘natural’) model, but is one designed to match the phenomenology of a number of natural models in a manner useful for general explorative studies of signals and backgrounds at new colliders.

A common phenomenological feature of models with gauge groups which are extensions of the SM group is that although the Vqq couplings are generally of similar strength to those of the Standard Model, the triple boson couplings, VWW and VWZ , are suppressed with respect to the Standard Model. In fact these couplings are generally absent in the symmetric (i.e. high-energy) limit and only arise after symmetry breaking through a mixing between the new vector bosons and the standard vector bosons, which is typically of order M_W^2/M_V^2 .

Thus, in the Extended Gauge Model the Vqq couplings are chosen to be simple copies of the corresponding SM couplings, whereas the strength of the VWZ coupling is the SM strength multiplied by a suppressing mixing factor ξ with the form $\xi = K \left(\frac{M_W^2}{M_V^2} \right)$ where K is a parameter of order 1. The $V \rightarrow WW$ and $V \rightarrow WZ$ decay widths then increase as M_V .

The EGM is implemented in PYTHIA and was used in the Monte Carlo study described in this chapter.

5.3.1.2 Present experimental limits

The present limits for the masses of heavy gauge bosons and their mixing with the standard model bosons are given in Tables 5.7, 5.8 and 5.9. [41]

Species	lower mass limit
W' — with standard couplings ^(a)	520 GeV CL=95%
W_R — a right-handed W ^(a)	406 GeV CL=90%
W_R ^(b)	1–2 TeV
(ν_R) ^(c)	460 GeV)

Table 5.7: Mass limits from searches for new charged gauge bosons. ^(a) [Particle Data Group, July 1994] ^(b) refs 11,12 in Mohapatra95, ^(c) ref 13 in Mohapatra95.

Species	lower mass limit from $p\bar{p}$ direct search	lower mass limit from electroweak fit
Z'_{SM} — with standard couplings	412 GeV	779 GeV
Z_{LR} of $SU(2)_L \times SU(2)_R \times U(1)$	310 GeV	389 GeV
Z_χ of $SO(10) \rightarrow SU(5) \times U(1)_\chi$	340 GeV	321 GeV
Z_ψ of $E_6 \rightarrow SO(10) \times U(1)_\psi$	320 GeV	160 GeV
Z_η of $E_6 \rightarrow SU(3) \times SU(2) \times U(1) \times U(1)_\eta$	340 GeV	182 GeV

Table 5.8: Mass limits from searches for new neutral gauge bosons. All figures correspond to a 95% confidence limit [Particle Data Group, July 1994]

Species	Z - Z' mixing limit
Z_{LR}	$-2.5 \leq \zeta \leq 6.2 \text{ mrad}$
Z_χ	$-4.7 \leq \zeta \leq 8.5 \text{ mrad}$
Z_ψ	$-5.0 \leq \zeta \leq 8.1 \text{ mrad}$
Z_η	$-16.2 \leq \zeta \leq 22.1 \text{ mrad}$

Table 5.9: Limits on Z - Z' mixing [Bright95].

5.3.2 Event generation software

Events containing heavy analogues of the standard model gauge bosons may be generated using the PYTHIA event generator [8]. The following decays are implemented (default coupling models in brackets):

$$\begin{aligned}
 W'^{\pm} &\longrightarrow q_i \bar{q}_j \quad (\text{def. : EGM} = \text{SM}) \\
 &\longrightarrow \ell \bar{\nu}_\ell \quad (\text{def. : EGM} = \text{SM}) \\
 &\longrightarrow WZ \quad (\text{def. : EGM} = \text{SM} \times m_W^2/m_{W'}^2) \\
 &\longrightarrow W^{\pm} H^0 \quad (\text{def. : LRS}) \\
 \\
 Z' &\longrightarrow q_i \bar{q}_i \quad (\text{def. : EGM} = \text{SM}) \\
 &\longrightarrow \ell^+ \ell^- \quad (\text{def. : EGM} = \text{SM}) \\
 &\longrightarrow W^+ W^- \quad (\text{def. : EGM} = \text{SM} \times m_W^2/m_{Z'}^2) \\
 &\longrightarrow Z^0 H^0 \quad (\text{def. : LRS})
 \end{aligned}$$

The PYTHIA implementation of the W' does not at present include interference with the Standard Model W . The couplings of the W' to quarks and leptons are set using four free parameters : the vector and axial couplings for the hadronic decays $W' \rightarrow q\bar{q}$ and for the leptonic decays $W' \rightarrow \ell\nu_\ell$. The defaults correspond to the $V - A$ structure of the Standard Model W , but it is possible to simulate a wide range of models by setting the couplings appropriately. One possible limitation is that the same CKM quark mixing matrix is assumed as for the standard W .

The PYTHIA implementation of the Z' contains the full $\gamma^*/Z/Z'$ interference structure for couplings to fermion pairs. By setting a parameter appropriately, it is possible to pick a subset, e.g. just the pure Z' piece. The couplings of the Z' to quarks and leptons can be set via a set of eight parameters which correspond to the vector and axial vector couplings of d-type quarks, u-type quarks, leptons and neutrinos. The default settings of these parameters correspond to the SM Z^0 , with axial couplings $a_f = \pm 1$ and vector couplings $v_f = a_f - 4e_f \sin^2 \theta_W$. This implies a resonance width which increases linearly with the mass. With a suitable choice of parameters, it is possible to simulate just about any imaginable Z' scenario, with full interference effects in cross sections and angular decay distributions.

The default value of the $Z'WW$ coupling corresponds to the Extended Gauge Model, wherein the coupling is the SM coupling multiplied by a factor $r = m_W^2/m_{Z'}^2$, giving

a Z' partial width into this channel which increases linearly with $m_{Z'}$. By manually dividing the default value by r one may obtain a model in which the partial decay width goes as $m_{Z'}^5$ (the Reference Model of [2]).

5.3.3 W' signals at ATLAS

5.3.3.1 Discovery potential

The W' production cross-section and the $W' \rightarrow \ell\nu$ cross-section are shown in Figure 5.27 for the Extended Gauge Model. The branching ratio $\text{BR}(W' \rightarrow e\nu)$ is shown in Figure 5.28.

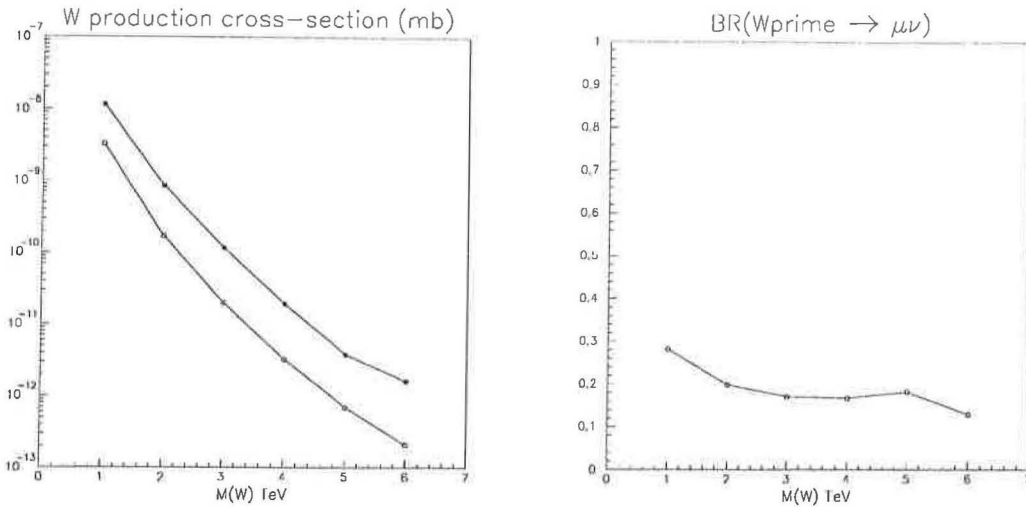


Figure 5.27: Cross-sections $\sigma(pp \rightarrow W')$, Figure 5.28: EGM branching ratio and $\sigma(pp \rightarrow W' \rightarrow \ell\nu)$ in the Extended Gauge Model. $\text{BR}(W' \rightarrow \ell\nu)$.

A distinctive kinematic feature of the two-body decay $W' \rightarrow \ell\nu$, is the so-called ‘Jacobian peak’ in the distribution of the lepton-neutrino transverse mass, m_T , defined as

$$m_T = \sqrt{2(|\mathbf{p}_T^\mu||\mathbf{p}_T^\nu| - (\mathbf{p}_T^\mu \cdot \mathbf{p}_T^\nu))}$$

$$0 \leq m_T \leq m_{W'} \quad (5.1)$$

The distribution peaks at a value of m_T near the centre of the W' Breit-Wigner distribution.

The background to this signal is mainly from semileptonic decays of the Standard Model W and from leptons from $t\bar{t}$ decays ($t\bar{t} \rightarrow XWb \rightarrow X\ell\nu b$). The following kinematic cuts have been suggested for background rejection [3]

- $|\eta(\ell)| \leq 2.5$
- $p_T(\ell) \geq 200 \text{ GeV}$
- $E_T^{miss} \geq 200 \text{ GeV}$
- $E_T^{iso}(\ell) \leq 12 \text{ GeV}$
- $m_T \geq m_T^{min}(m_W)$

It has been shown that with the application of the above cuts, a W' is clearly visible above the background, and a W' with a mass of up to about 5 TeV could be discovered at ATLAS within one year of LHC running [7]. See Figures 5.29 and 5.30.

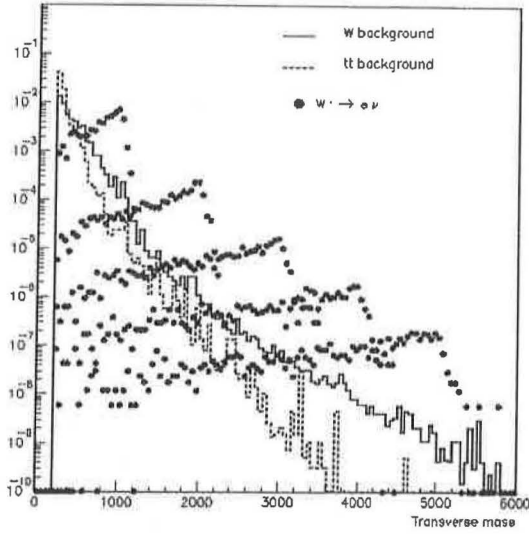


Figure 5.29: Transverse mass distribution for signal ($m_W = 1-5 \text{ TeV}$) and background (before cuts).

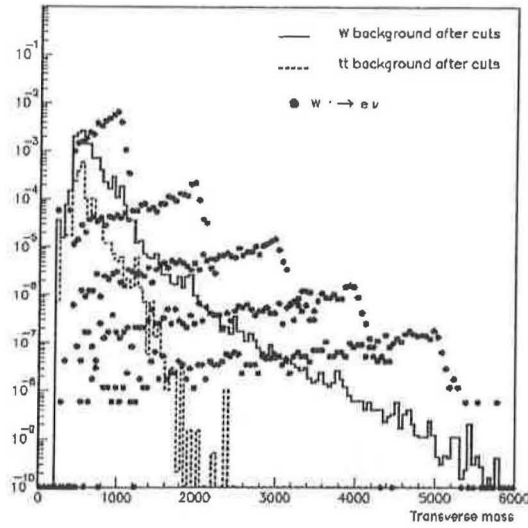


Figure 5.30: Transverse mass distribution for signal ($m_W = 1-5 \text{ TeV}$) and background (after cuts).

5.3.3.2 Mass measurement precision

If a W' is discovered, its mass may be determined by comparing the experimental transverse mass distribution with Monte Carlo distributions generated for a range of W' masses. The value of the mass is then taken as the one which minimises the quantity $S^2(M_W)$, defined as

$$S^2(m_W^{MC}) = \sum_{\text{bins } i} \frac{[N_i^{exp}(m_W^{exp}) - N_i^{MC}(m_W^{MC})]^2}{N_i^{MC}(m_W^{MC})} \quad (5.2)$$

where N_i^{exp} and N_i^{MC} are the numbers of events in histogram bin i of the experimental and Monte Carlo distributions respectively.

$M_{W'}$ (GeV)	$\Gamma_{W'}$ (GeV)	$W' \rightarrow \ell\nu$	$W_{SM} \rightarrow \ell\nu$	$t\bar{t}$
1000	27	220 285 (201 537)	32 739 (28 878)	7 190 (1 831)
2000	54	11 018 (9 682)	861 (771)	228 (67)
3000	82	1 233 (1 128)	156 (142)	43 (2)
4000	101	153 (131)	14 (13)	3 (< 1)
5000	130	22 (19)	3 (3)	< 1 (< 1)
6000	161	3 (2)	0.6 (0.6)	< 1 (< 1)

Table 5.10: *Number of signal (EGM) and background events expected in one year of running at the LHC before (after) cuts ($|\eta| < 2.8$).*

The precision with which m_W may be extracted from the transverse mass distribution may be determined by investigating the expected curvature of S^2 near its minimum value using a simulated ‘experimental’ distribution. The steepness of the curve around the minimum gives an indication of the attainable mass measurement precision. This analysis was performed in [7] and showed that the mass measurement precision which can be expected at ATLAS is of the order of 30–40 GeV for $m_W = 2$ TeV, rising to about 100 GeV for a $m_W = 4$ TeV.

5.3.4 Forward-backward asymmetry of $W' \rightarrow e\nu$

5.3.4.1 Introduction

The forward-backward asymmetry, A_{FB} , of the leptonic decay of a new gauge boson is strongly model-dependent quantity. For this reason, if a new gauge boson is discovered at LHC, asymmetry measurements will be useful in discriminating between physical models of its origin. It is thus important to assess the sensitivity of ATLAS to such an asymmetry. It will be seen that this is dependent upon the alignment of the Inner Detector.

The forward-backward asymmetry is defined as follows. For the semileptonic decay of a W' produced with rapidity y_W the forward-backward asymmetry $A_{FB}(y_W)$ is defined to be

$$A_{FB}(y_W) = \frac{F(y_W) - B(y_W)}{F(y_W) + B(y_W)} \quad (5.3)$$

where

$$F \pm B = \left[\int_0^1 \pm \int_{-1}^0 \right] \frac{d^2\sigma}{dy d(\cos\theta^*)} d(\cos\theta^*) \quad (5.4)$$

and θ^* is the angle between the direction of motion of the muon in the rest frame of the W' and the direction of motion of the W' in the lab frame. That is, θ^* is the analogue of the angle θ_μ^{lab} in the rest frame of the W' . See Figure 5.31.

Evaluation of the forward-backward asymmetry requires the reconstruction of the rapidity of the W' and the angle θ^* from the momenta of the decay products. In addition, the sign of the lepton is required for the forward/backward assignment.

For very high p_T electrons from the decay of a heavy W' , the electromagnetic calorimeter will offer better momentum resolution than the Inner Detector. However, the inner tracker is required to provide the sign of the electron charge. It was shown in Section 5.2 that the p_T resolution is sensitive to the Φ -alignment of the precision layers of the Inner Detector barrel. For this reason the variation (dilution) of the asymmetry signal as a function of the Φ -alignment of the precision layers of the barrel of the Inner Detector has been studied with a view to investigating the alignment requirements.

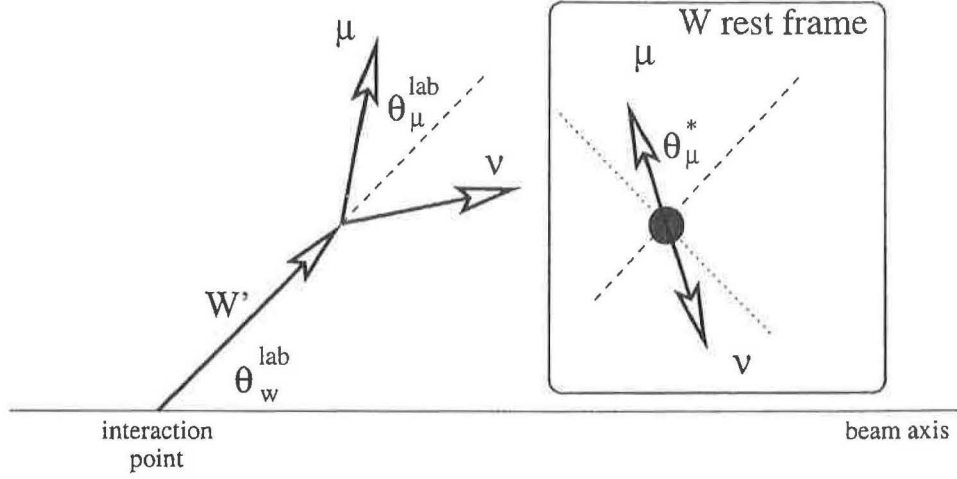


Figure 5.31: The angle used in the forward/backward assignment is the angle between the direction of motion of the muon in the W' rest frame and the projected direction of motion of the W' in the lab frame.

5.3.4.2 Reconstruction

The leptonic decay of the W' is shown in Figure 5.32. Summing the four-vectors at the

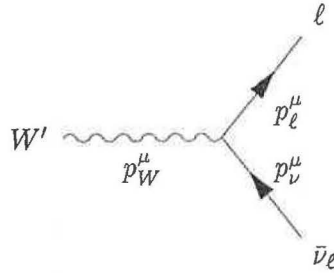


Figure 5.32: Leptonic decay of a W' , labelled with the four-momenta.

vertex and squaring to form the invariant one obtains

$$p_W = p_\ell + p_\nu \quad (5.5)$$

$$p_W^2 = (p_\ell + p_\nu)^2 = m_W^2 \quad (5.6)$$

For lepton and neutrino energies much higher than their rest masses (i.e. $E_{\ell,\nu} \approx p_{\ell,\nu}$)

$$\begin{aligned} m_W^2 &= (E_\ell + E_\nu)^2 - (\mathbf{p}_\ell + \mathbf{p}_\nu)^2 \\ &\approx 2(p_\ell p_\nu - \mathbf{p}_\ell \cdot \mathbf{p}_\nu) \\ &\approx 2p_\ell p_\nu (1 - \cos \theta_{\ell\nu}) \end{aligned} \quad (5.7)$$

where $\theta_{\ell\nu}$ is the angle between the directions of the two leptons.

Writing the transverse and longitudinal components of the neutrino momentum explicitly yields the reconstruction equation

$$m_W^2 = 2|p^\ell| \sqrt{|p_T^\nu|^2 + |p_L^\nu|^2} - (\mathbf{p}_T^\ell \cdot \mathbf{p}_T^\nu + p_L^\ell p_L^\nu) \quad (5.8)$$

Since a measurement of E_T^{miss} gives only the transverse momentum, the longitudinal momentum of the neutrino p_L^ν is unknown. Equation 5.8 thus has two unknowns: p_L^ν and m_W . If the value of m_W is taken to be the mean value of the W' mass distribution, this quadratic equation may be solved to yield two solutions for p_L . Often one of the solutions is unphysical and can be rejected. However, a considerable fraction of reconstructions yield two physically acceptable solutions for the **Feynman-x** ($x_F = p_L/E_{beam}$) of the W' . In these cases the best that can be done is to choose the solution with the lower value for $|x_F|$. The inclusion of wrong choices of x_F distorts considerably the shape of the measured asymmetry curve.

In addition to the choice of x_F , there are a number of other factors result in a measured asymmetry with a shape which differs markedly from the theoretical curve

- reconstructed W' rapidity
 - choice of Feynman-x of the W'
 - the width of the W' (cannot be forced to have the nominal mass without affecting the reconstructed x_F)
 - finite resolution of the calorimeter for the lepton and neutrino momentum measurements
- sign of the measured track curvature (tracker momentum resolution)
- track finding performance (wrong hits associated with track)
- kinematic cuts required for background rejection

The Feynman-x identification performance is examined in Section 5.3.5.3.

5.3.4.3 Charge sign identification

Since the identification of events as either ‘forward’ or ‘backward’ is dependent upon identifying the lepton sign correctly, measurement of the forward-backward asymmetry of the leptonic decay of a heavy boson requires a good charge-sign identification (CSI) performance for high- p_T leptons and hence a good momentum resolution. It was shown in Section 2.2 that the momentum resolution is a strong function of the azimuthal alignment precision.

The relationship between the p_T resolution of the tracking detector and the charge sign identification performance is as follows. To a good approximation, the distribution of $(p_T^{\text{found}}/p_T^{\text{true}})$ obtained from a tracking detector is a Gaussian with mean value 1. The normalised Gaussian centred on 1 is given as a function of x by

$$G(x, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} \exp \frac{-(x-1)^2}{2\sigma^2}$$

with the wrong-sign fraction w for such a Gaussian being given by

$$w = \int_{-\infty}^0 G(x, \sigma) dx$$

See Figures 5.33 and 5.34.

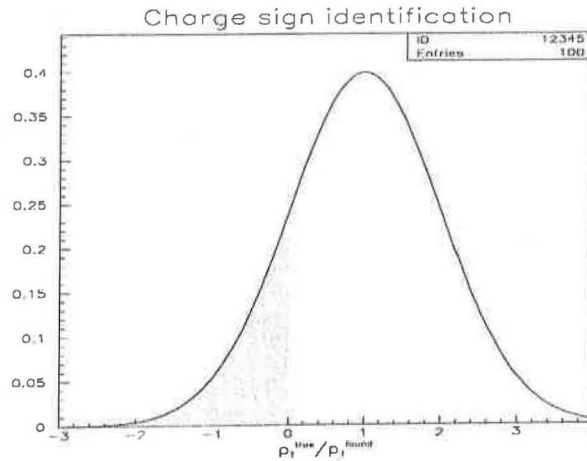


Figure 5.33: The ratio $(p_T^{\text{true}}/p_T^{\text{found}})$ has a Gaussian distribution with a width σ ($=1$ here). The shaded region indicates tracks for which the measured curvature has the opposite sign from the true value and for which the forward/backward assignment will therefore be incorrect.

If it is supposed that for a given rapidity bin, the true fraction of events which are ‘forward’ is f , and the total number of events at this rapidity is N , then if the charge

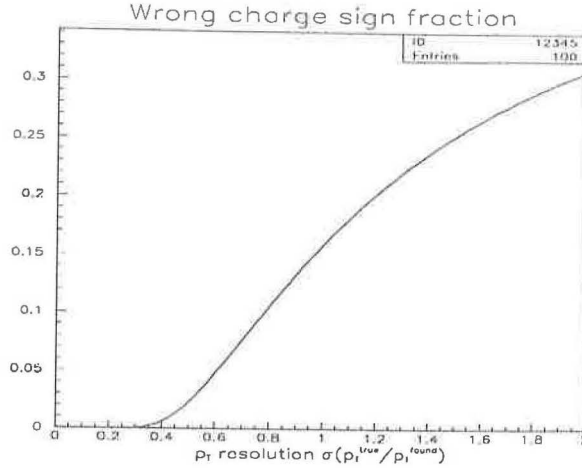


Figure 5.34: The fraction of tracks with an incorrectly assigned charge sign as a function of the p_T resolution, assuming a perfectly Gaussian distribution for $(p_T^{\text{true}}/p_T^{\text{found}})$.

sign identification were perfect, the number of events in the forward and backward bins and the asymmetry would be given by

$$N_F = Nf \quad (5.9)$$

$$N_B = N(1 - f) \quad (5.10)$$

$$\begin{aligned} A_{FB}^0 &= \frac{N_F - N_B}{N_F + N_B} \\ &= (2f - 1) \end{aligned} \quad (5.11)$$

If the effective wrong-sign fraction averaged over the lepton momenta of the events in this rapidity bin is \bar{w} (for both W^+ and W^-) then the numbers of events in the forward and backward bins for a given rapidity become

$$N_F = Nf(1 - \bar{w}) + N(1 - f)\bar{w} \quad (5.12)$$

$$N_B = N(1 - f)(1 - \bar{w}) + Nf\bar{w} \quad (5.13)$$

$$\begin{aligned} A_{FB}^{\bar{w}} &= \frac{N_F - N_B}{N_F + N_B} \\ &= (2f - 1)(1 - 2\bar{w}) \\ &= A_{FB}^0(1 - 2\bar{w}) \end{aligned} \quad (5.14)$$

so the asymmetry observed with an otherwise perfect detector which has an average wrong sign fraction \bar{w} is simply the true asymmetry scaled by a factor $(1 - 2\bar{w})$. As expected, the measured asymmetry tends to zero as \bar{w} tends to $\frac{1}{2}$.

Binomial errors for A_{FB} and CSI

Since any given event may be identified as only either backward or forward, the calculation of the error on the measured asymmetry requires a binomial analysis. It may be shown that the error on N_F is given by $\sigma_{N_F} = \sqrt{Nf(1-f)}$ which may be approximated using experimental or Monte Carlo data by $\sigma_{N_F} \sim \sqrt{\frac{N_F N_B}{N_F + N_B}}$. The error on the forward fraction f is given by

$$\sigma_f = \frac{1}{N} \sqrt{Nf(1-f)} \sim \frac{1}{N_F + N_B} \sqrt{\frac{N_F N_B}{N_F + N_B}}$$

The value of the forward-backward asymmetry in this rapidity bin is given by

$$A_{FB} = \frac{N_F - N_B}{N_F + N_B} = (2f - 1)$$

and the error on the asymmetry is given by

$$\sigma_{AFB} = 2\sigma_f = \frac{2}{N_F + N_B} \sqrt{\frac{N_F N_B}{N_F + N_B}}$$

A similar calculation is made for the error on the wrong-sign fraction. If the momentum-averaged fraction of wrong sign assignments for a rapidity bin is \bar{w} then for a total of N events in the bin, the number of right and wrong sign identifications is $N_R = (1 - \bar{w})N$ and $N_W = \bar{w}N$ respectively. The error on both the right-sign and wrong-sign fractions is given by

$$\sigma_{\bar{w}} = \sqrt{\frac{\bar{w}(1-\bar{w})}{N}} \sim \frac{1}{N_R + N_W} \sqrt{\frac{N_R N_W}{N_R + N_W}}$$

5.3.5 Simulation results

Two types of simulation were used : a particle-level study to give high statistics where possible, and a full simulation study to include the effects of pattern recognition and the alignment of the SCT barrel.

A particle-level analysis of the discovery potential, the mass resolution and the background rejection was made in [7], and was not repeated. The main focus of the study presented here was a detailed evaluation of the sensitivity to the forward-backward asymmetry of the decay $W' \rightarrow e\nu$, with particular attention to the effects of misalignments of the SCT barrel.

The particle-level event samples consisted of 250,000 $W' \rightarrow e\nu$ events for each of six different W' masses from 1 to 6 TeV. These samples were used for calculation of theoretical asymmetries (as opposed to the distorted detected asymmetries) and for a calculation of the expected dilution of the forward-backward asymmetry as a function of the overall tracker momentum resolution, under the assumption of a precisely Gaussian distribution of $(p_T^{\text{found}}/p_T^{\text{true}})$.

The fully simulated event samples consisted of around 9,000 events at each of three W' masses : 1, 3 and 6 TeV, generated with PYTHIA and fully simulated within the framework of SLUG and DICE. Track reconstruction, including simulation of the barrel SCT alignment, was performed using the modified *iPatRec* routine described in Section 5.2.2.

For the sake of speed, a full calorimeter simulation was not used in this study. Instead, the electron energy was smeared according to the nominal electromagnetic calorimeter energy resolution of $\frac{10\%}{\sqrt{E}} \oplus 0.5\%$, and the transverse energy of the neutrino was smeared according to the nominal E_T^{miss} resolution of $\frac{50\%}{\sqrt{E}} \oplus 3\%$.

The following cuts, required for background rejection, were applied

- $p_T(e) \geq 200$
- $p_T(\nu) \geq 200$
- $m_T \geq m_T^{\text{cut}}$, where m_T^{cut} was 550, 1500, 3800 GeV for $m_W = 1, 3, 6$ TeV respectively

Due to the finite calorimeter energy resolution and the use of the mean mass in the reconstruction, rather than the true Breit-Wigner distributed mass, Equation 5.8 is not necessarily soluble for every event. Hence, in addition to the cuts given above, events were rejected for which calculation problems were encountered in the reconstruction (e.g. no solution for x_F or two unphysical values).

5.3.5.1 W' production and decay

The production rapidity of the W' is shown in Figure 5.35 for $m_{W'} = 3$ TeV. The rapidity of the decay electron is shown in Figure 5.36.

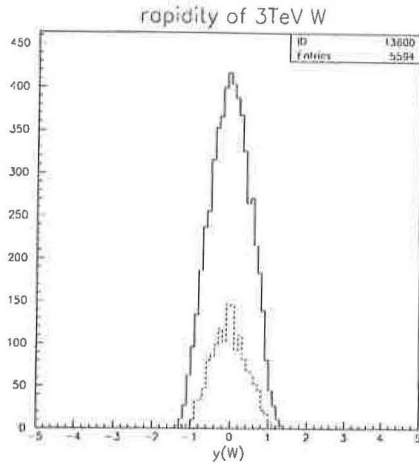


Figure 5.35: Production rapidity distribution of 3 TeV W'^+ (solid) and W'^- (dashed).

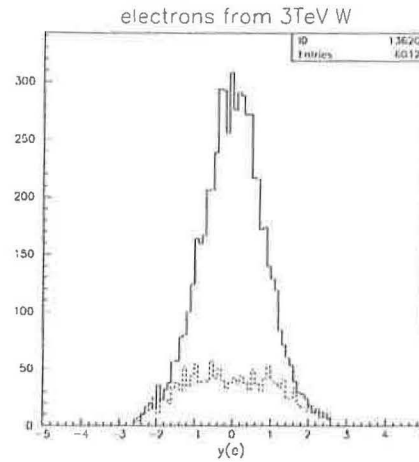


Figure 5.36: Rapidity distribution of e^+ (solid) and e^- (dashed) from the semileptonic decay of a 3 TeV W'^+ and W'^- .

5.3.5.2 Lepton transverse momentum distributions

Transverse momentum distributions (after cuts) of electrons from the semileptonic decay of a W' are shown in Figures 5.37 to 5.39 for $m_{W'} = 1, 3$ and 6 TeV. Jacobian peaks are seen centred on $p_T = m_{W'}/2$.

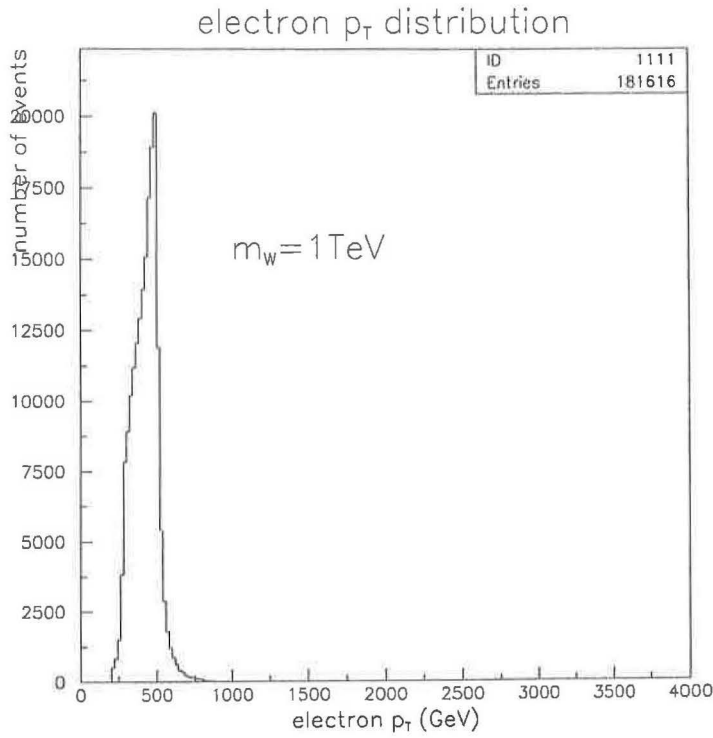


Figure 5.37: Transverse momentum distribution of electrons from the decay of a 1 TeV W' after the cuts : $p_T(e) \geq 200 \text{ GeV}$, $p_T(\nu) \geq 200 \text{ GeV}$ and $m_T \geq 550 \text{ GeV}$.

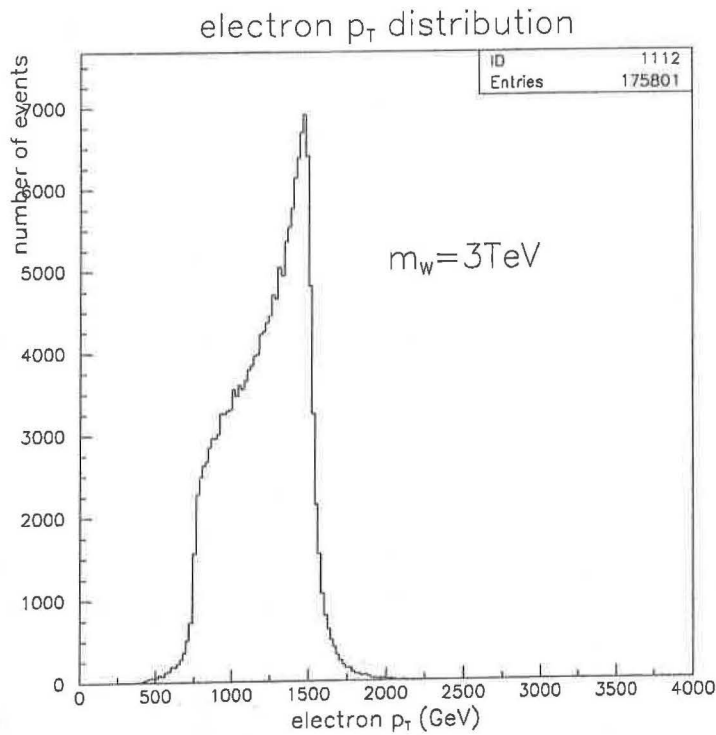


Figure 5.38: Transverse momentum distribution of electrons from the decay of a 3 TeV W' after the cuts : $p_T(e) \geq 200 \text{ GeV}$, $p_T(\nu) \geq 200 \text{ GeV}$ and $m_T \geq 1500 \text{ GeV}$.

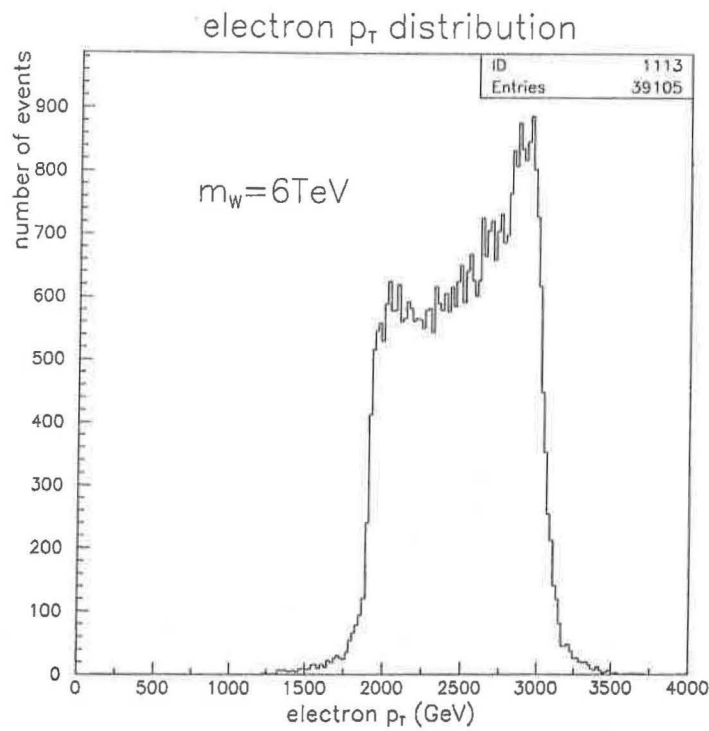


Figure 5.39: Transverse momentum distribution of electrons from the decay of a 6 TeV W' after the cuts : $p_T(e) \geq 200$ GeV, $p_T(\nu) \geq 200$ GeV and $m_T \geq 3800$ GeV.

5.3.5.3 Feynman-x reconstruction

For the reasons explained in Section 5.3.4.2, the reconstructed value for the rapidity of the W' differs from the true value. In Figure 5.40 the reconstructed rapidity is plotted against true rapidity for all the events passing the cuts. The distribution is predominantly near the diagonal, with a small contribution from wrong choices of x_F .

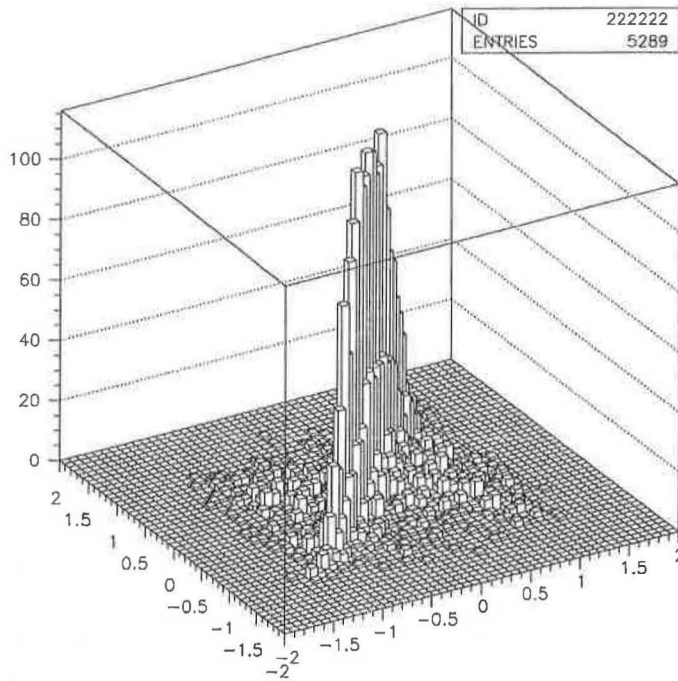


Figure 5.40: $y^{\text{found}}(W')$ vs. $y^{\text{true}}(W')$ for all events passing the cuts.

Figure 5.41 (5.42) shows the events for which the solution chosen for the Feynman-x was closer to (further from) the true value of x_F than the rejected solution. The width around the diagonal is caused by the calorimeter smearing and the fact that the reconstruction fits to the central value of the W' mass rather than the true value.

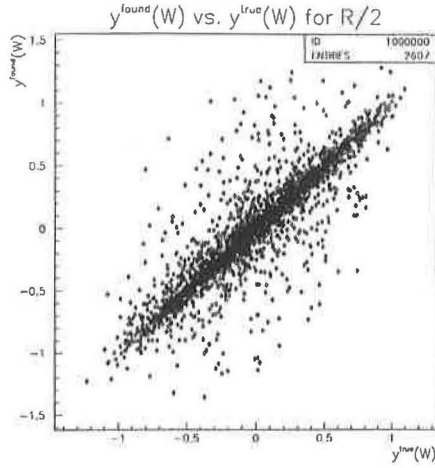


Figure 5.41: $y^{\text{found}}(W')$ vs. $y^{\text{true}}(W')$ for events in which the 'better' choice (see text) of two was made for x_F ($m_W = 3 \text{ TeV}$).

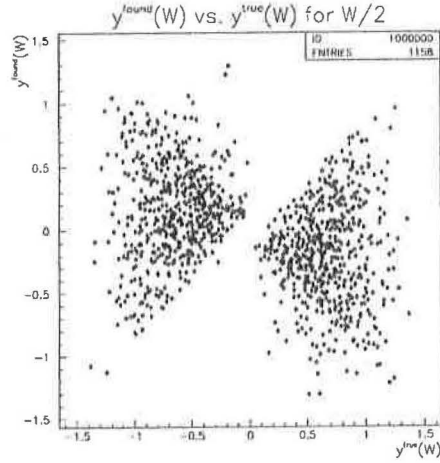


Figure 5.42: $y^{\text{found}}(W')$ vs. $y^{\text{true}}(W')$ for events in which the 'worse' choice (see text) of two was made for x_F ($m_W = 3 \text{ TeV}$).

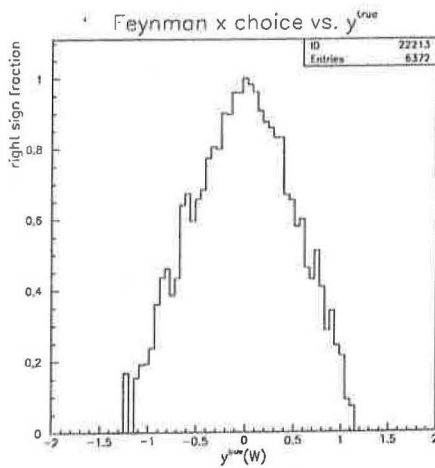


Figure 5.43: Right x_F choice fraction vs. y^{true} for events in which the quadratic equation for p_L^ν yields two physically acceptable solutions ($-1 < x_F < 1$).

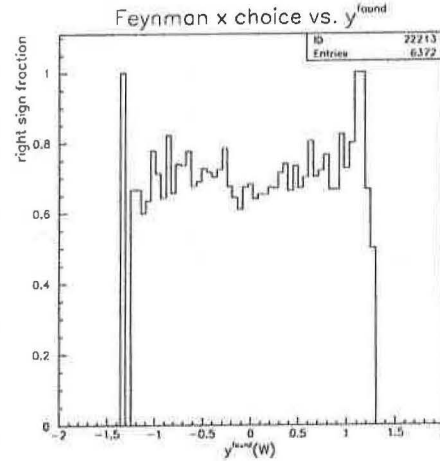
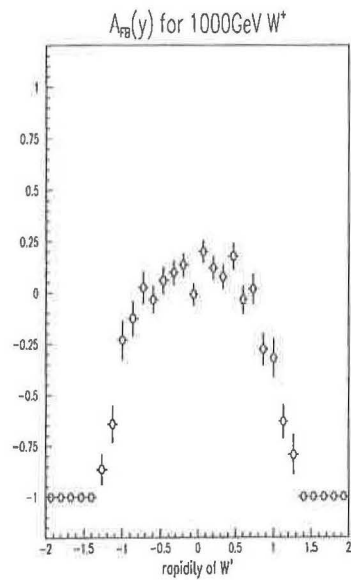
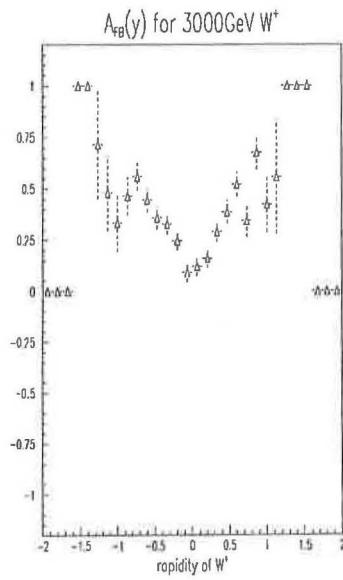
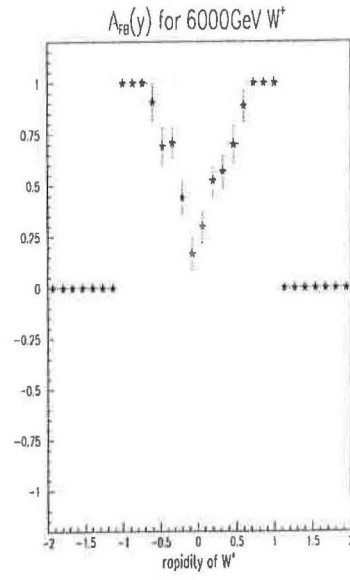
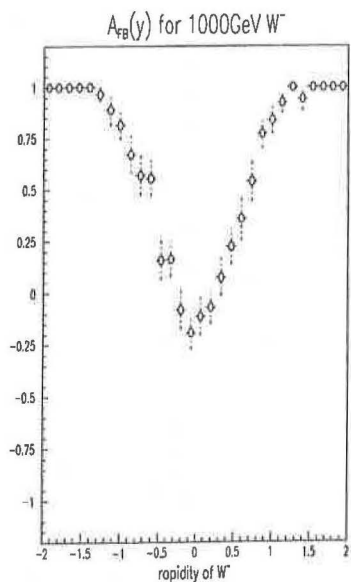
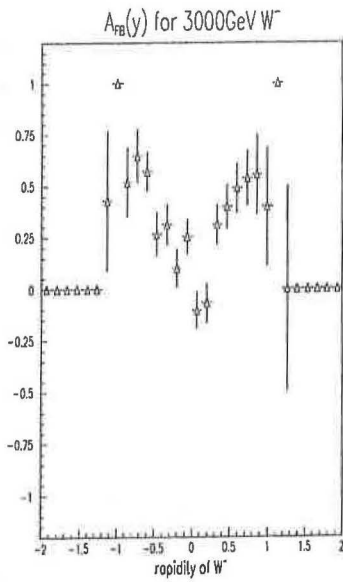
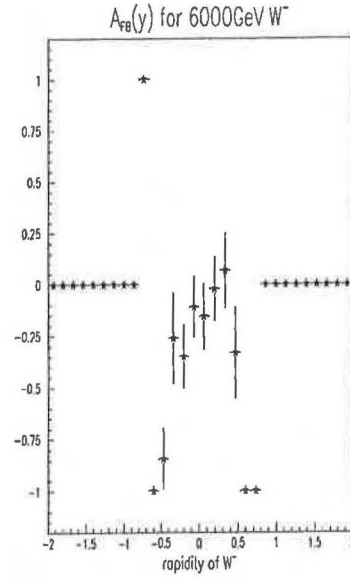


Figure 5.44: Right x_F choice fraction vs. y^{found} for events in which the quadratic equation for p_L^ν yields two physically acceptable solutions ($-1 < x_F < 1$).

5.3.5.4 Forward-backward asymmetries

The measured forward-backward asymmetries for the W'^{+} and W'^{-} with masses $m_{W'} = 1, 3$ and 6 TeV are shown in Figures 5.45 to 5.50. For the reasons given in Section 5.3.4.2, the measured asymmetries differ considerably from those which would be obtained with perfect detection.

Figure 5.45: 1 TeV W'^{+} Figure 5.46: 3 TeV W'^{+} Figure 5.47: 6 TeV W'^{+} Figure 5.48: 1 TeV W'^{-} Figure 5.49: 3 TeV W'^{-} Figure 5.50: 6 TeV W'^{-}

5.3.5.5 Momentum resolution and CSI vs. Φ -alignment

The average momentum resolution for the electrons from 3 TeV W' decays is shown as a function of the Φ -alignment of the barrel SCT in Figure 5.51. The average wrong charge sign fraction w is shown as a function of Φ -alignment in Figure 5.52.

It should be noted that for the high- p_T electrons passing the cuts, the momentum resolution is proportional to p_T . The wrong sign fraction w is not proportional to p_T or $\sigma(p_T)$, so the momentum-averaged points in Figures 5.51 and 5.52 cannot be simply related using the curve in Figure 5.34.

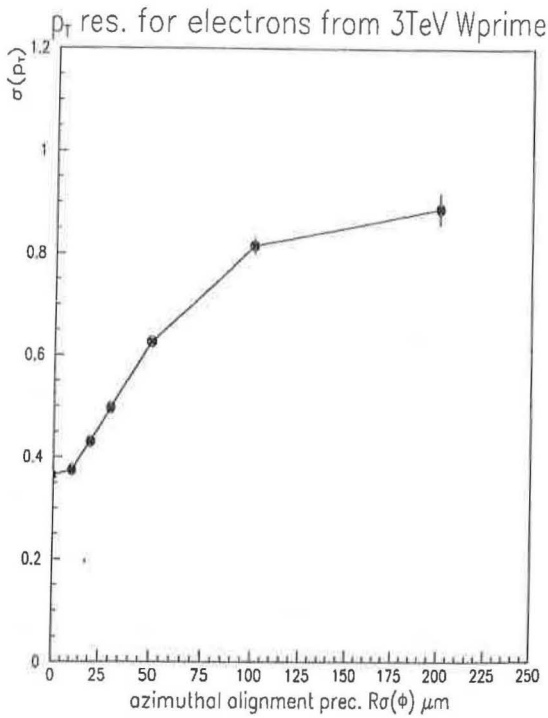


Figure 5.51: Momentum-averaged p_T resolution vs. barrel Φ -alignment precision.

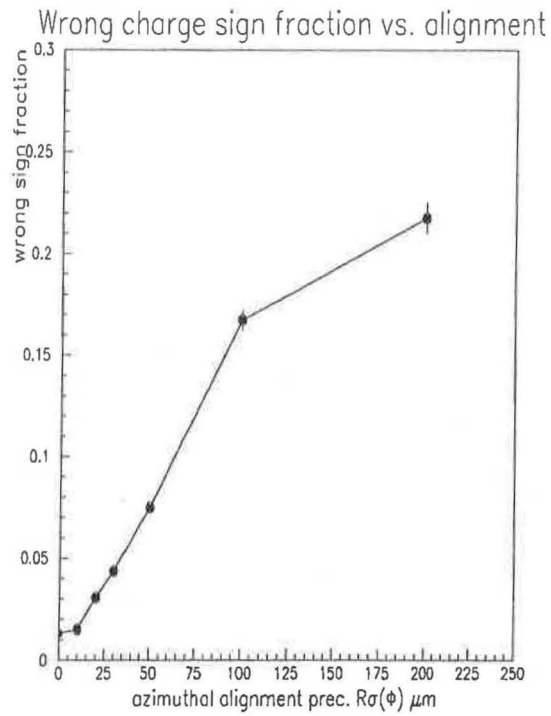


Figure 5.52: Wrong charge sign fraction vs. barrel Φ -alignment precision.

5.3.5.6 Asymmetry and integrated asymmetry vs. Φ -alignment

The forward-backward asymmetries of the 3TeV W^+ and W^- are shown as a function of the barrel SCT Φ -alignment precision in Figures 5.53 and 5.54.

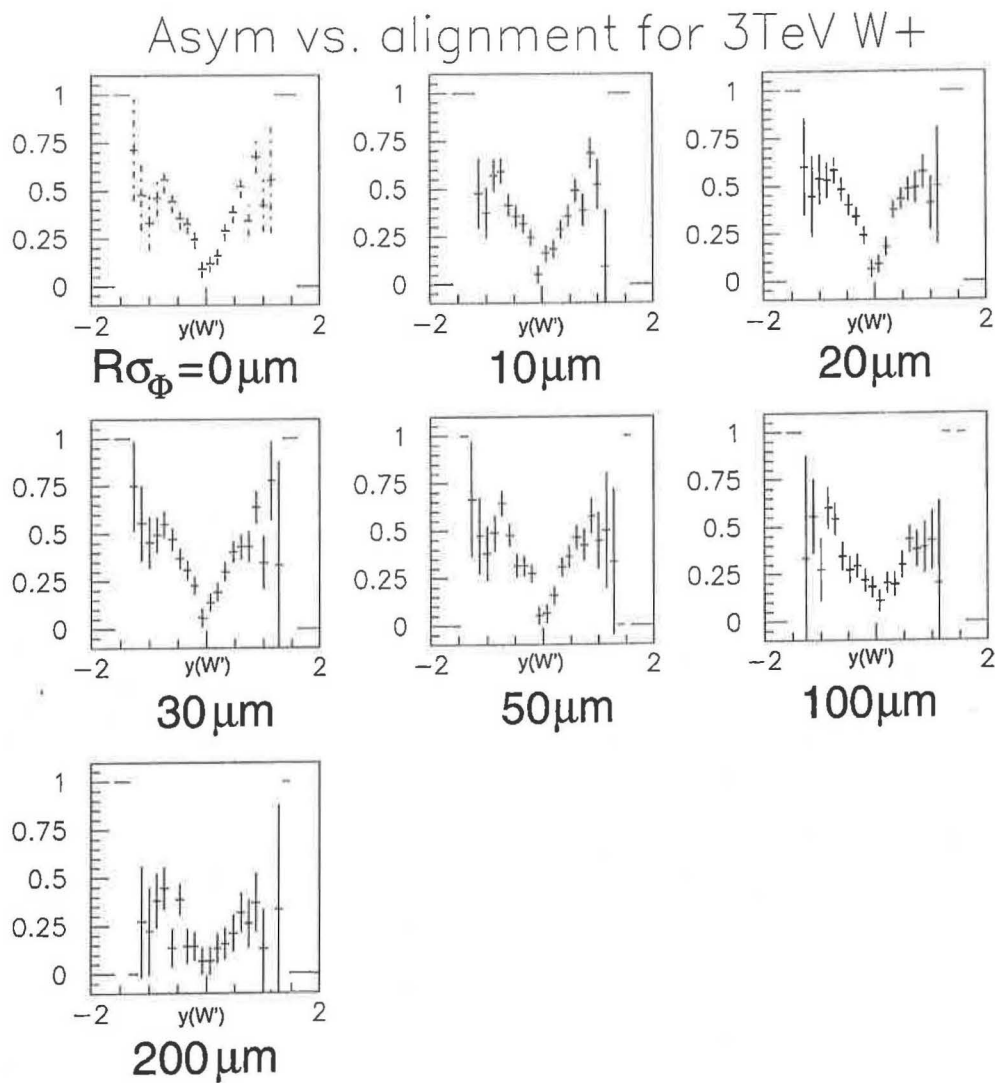


Figure 5.53: $A_{FB}(y)$ for 3 TeV W'^{\pm} vs. barrel Φ -alignment precision.

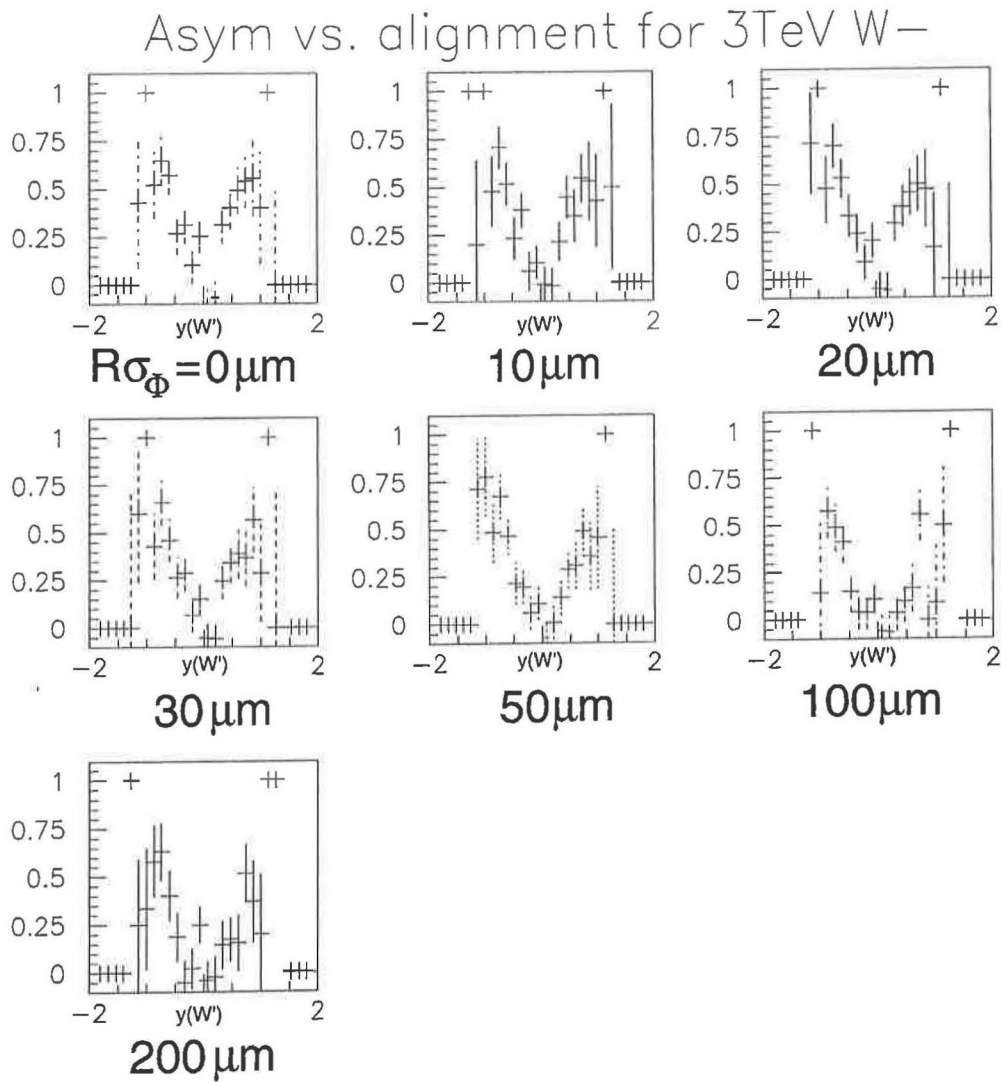


Figure 5.54: $A_{FB}(y)$ for 3 TeV W'^- vs. barrel Φ -alignment precision.

Each of the seven curves in Figures 5.53 and 5.54 was integrated with respect to the W' rapidity by summing the number of entries in the bins lying between two limits. The integration limits were chosen so as to exclude the high rapidity bins which contained low numbers of events. The integrated forward-backward asymmetries of the 3 TeV W'^+ and W'^- are shown as a function of the barrel SCT Φ -alignment precision in Figures 5.55 and 5.56.

It is seen that as the Φ -alignment precision is reduced to $200\mu\text{m}$, the integrated asymmetries drop to 0.5 ± 0.1 of their values for perfect alignment, i.e. from 9.7 to 4.7 (arbitrary units) for the 3 TeV W'^+ and from 3.2 to 1.6 for the 3 TeV W'^- .

Although the asymmetry signal is reduced, it is not lost completely as the alignment precision of the precision layers of the barrel is reduced, indicating that for the perfect TRT alignment scenario modelled here, the TRT standalone performance is sufficient for detection of this signal.

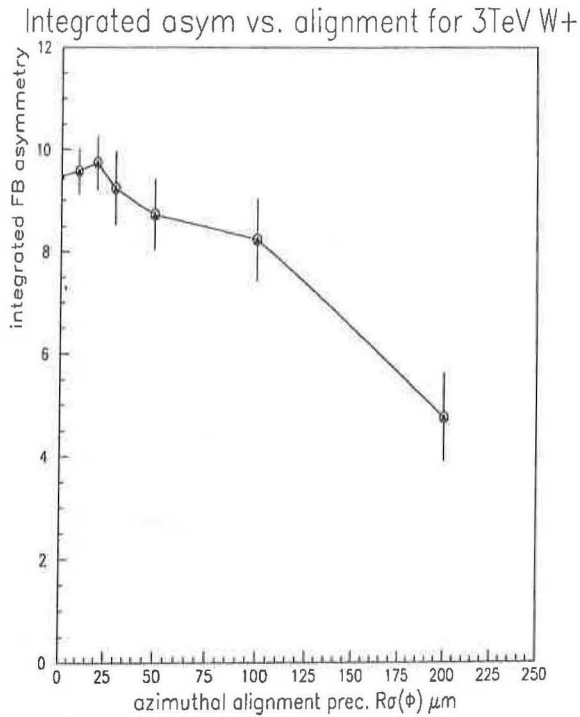


Figure 5.55: *Integrated asymmetry of 3 TeV W'^+ vs. SCT barrel Φ -alignment precision.*

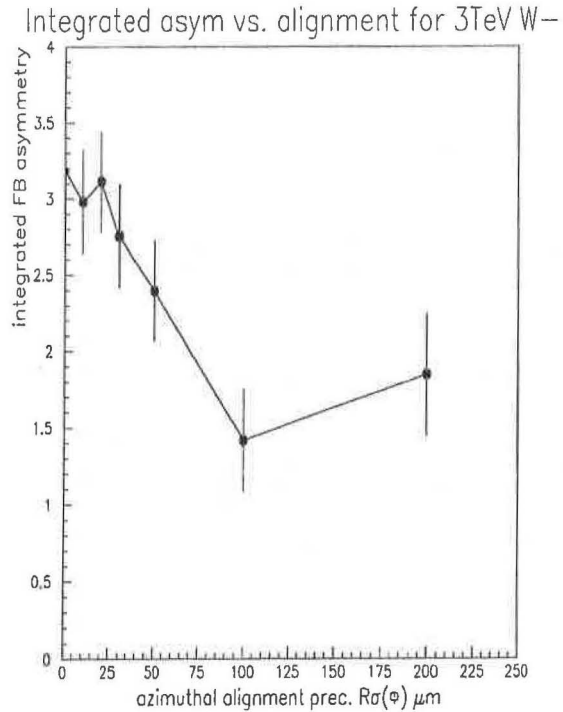


Figure 5.56: *Integrated asymmetry of 3 TeV W'^- vs. SCT barrel Φ -alignment precision.*

5.3.5.7 Particle-level integrated A_{FB} vs. momentum resolution

It was shown in Section 5.3.4.3 that the value of the measured asymmetry in any rapidity bin is proportional to the value of $(1 - 2\bar{w})$ where \bar{w} is the effective wrong-sign fraction averaged over the lepton momenta of the events in the bin. The dependence of the wrong-sign fraction w on the momentum resolution was shown in Figure 5.34 under the assumption that the distribution $(p_T^{\text{found}}/p_T^{\text{true}})$ is precisely Gaussian. Assuming a Gaussian distribution and taking the momentum resolution of the tracker as proportional to p_T , it is possible to use the electron p_T distribution to calculate an averaged value for \bar{w} and hence the average asymmetry dilution $(1 - 2\bar{w})$ as a function of the overall tracker momentum performance.

For leptons above about 40 GeV multiple scattering effects are small and the overall p_T resolution of the tracker may be assumed to be simply proportional to p_T , i.e. $\sigma(p_T) = A \cdot p_T$ where A is a parameter of the detector. The specification for the Inner Detector momentum resolution is that for a 500 GeV track, $\sigma(p_T) < 30\%$ i.e. $A^{\text{spec}} = 0.0006 \text{ GeV}^{-1}$.

The integrated asymmetry is shown in Figure 5.57 as a function of the tracker momentum constant A for $m_{W'} = 1, 3, 6 \text{ TeV}$. For $m_{W'} = 1 \text{ TeV}$, the relative asymmetry remains above 90% up to $A = 0.0014 \text{ GeV}^{-1}$, and then falls gradually, reaching 50% at $A = 0.0036 \text{ GeV}^{-1}$. The relative asymmetries for $m_{W'} = 3$ and 6 TeV remain above 90% as far as $A = 0.0005$ and 0.0002 GeV^{-1} respectively and then fall steeply, reaching 50% at $A = 0.0013$ and 0.0006 GeV^{-1} respectively.

It should be noted that the relative asymmetry is a measure of the fraction of the true asymmetry which will be observed. The true value of the integrated asymmetry may be small or large depending on the details of the physics involved.

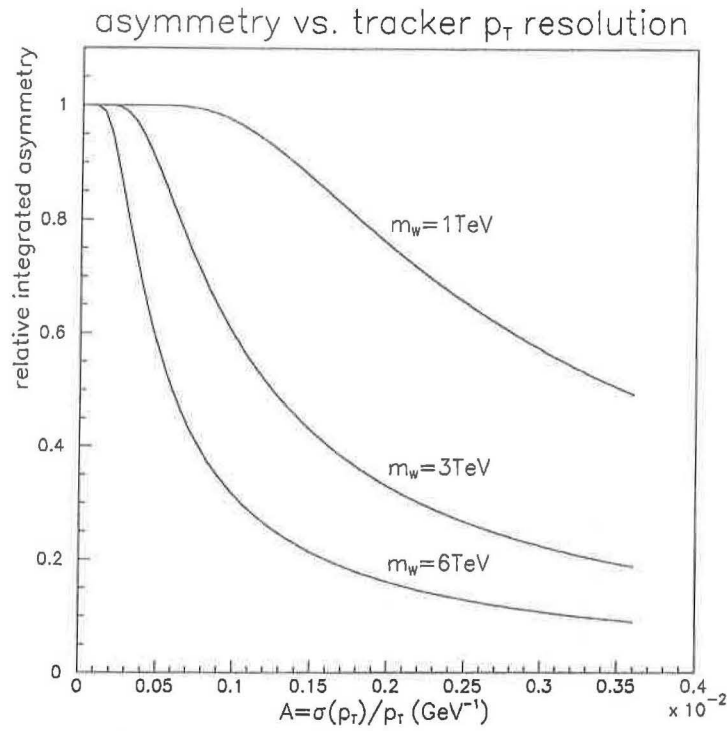


Figure 5.57: *Relative integrated asymmetry vs. tracker p_T resolution for $m_{W'} = 1, 3, 6$ TeV. This was calculated by folding the (post-cut) electron p_T distributions with the curve of w vs. σ_{p_T} and using momentum resolution $\sigma(p_T) = A \cdot p_T$ to calculate a momentum-averaged value for $(1 - 2\bar{w})$ versus A . The distribution of $(p_T^{\text{found}}/p_T^{\text{true}})$ is assumed to be precisely Gaussian. The tracker specification requires $A < 0.06 \times 10^{-2} \text{ GeV}^{-1}$.*

5.4 Conclusions

It has been shown that precise alignment of the SCT barrel is critically important in attaining the desired track fitting resolutions. The results obtained from the study of track fitting resolutions were discussed in detail in Section 5.2.4. The results indicate that the alignment precisions need to be as low as $10\mu\text{m}$ for low degradation of the resolutions compared to the perfectly aligned detector, Φ -alignment requirements being in general the most stringent. The use of stereo angles has the effect of linking the transverse track fitting resolutions to the longitudinal alignment precisions and vice versa.

A W' with a mass up to about 5 TeV could be detected at ATLAS in one year of LHC running at a luminosity of $10^{34}\text{ cm}^{-2}\text{s}^{-1}$. If perfectly aligned, the ATLAS detector will be sensitive to the forward-backward asymmetry of the semileptonic decay of the W' over all of the accessible mass range. Although the measured integrated asymmetry is diluted by misalignments of the barrel SCT, the signal for $m_{W'} = 3\text{ TeV}$ is not lost completely, indicating that for the perfect TRT alignment scenario modelled here, the TRT standalone performance is sufficient for detection of this signal. The relative integrated asymmetry of a 6 TeV W' is likely to be a steeper function of the tracker performance, and may set stronger alignment requirements.

Future work will focus on developing a model of misalignments in the TRT and using it to extend the current study to include the effect of the TRT alignment on both single-track fitting resolutions and physics signals. Similarly, the software could be extended to include the alignment of the Muon Spectrometer. An improved implementation of the reconstruction algorithm is expected in the near future which may be sufficiently fast to allow the inclusion of pile-up.

References

- [5.1] P. Langacker, R.W. Robinett and J.L. Rosner, *New heavy gauge bosons in pp and $p\bar{p}$ collisions*, Phys. Rev. D30 (1984) 1470–1487.
- [5.2] G. Altarelli, B. Mele and M. Ruiz-Altaba, *Searching for new heavy vector bosons in $p\bar{p}$ colliders*, Z. Phys. C45 (1989) 109–121.
- [5.3] *Large Hadron Collider Workshop*, European Committee for Future Accelerators, Aachen 4–9 October 1990. CERN 90-10, ECFA 90-133.
- [5.4] Botterweck, p722 in [3].
- [5.5] Chiapetta et. al., *New vector bosons*, Volume II, Section A7 of [3].
- [5.6] J. Layssac, F.M. Renard and C. Verzegnassi, *Model independent constraints on a heavy neutral vector boson from present and future LEP and SLC data*, Z. Phys. C53 (1992) 97–114.
- [5.7] M.C. Cousinou, *Search for a W' in the $l\nu$ channel*, ATLAS Collaboration Internal Note PHYS-NO-059, 19 December 1994.
- [5.8] Torbjörn Sjöstrand, *PYTHIA 5.7 and JETSET 7.4 physics and manual*, Computer Physics Commun. **82** (1994) 74.
- [5.9] D. Froidevaux and A. Parker, *The performance specifications of the ATLAS Inner Detector*, ATLAS Collaboration Internal Note INDET-NO-046, 4 May 1994.
- [5.10] J. Loken and A. Reichold, *Description of a Pattern Recognition Algorithm using Continuous and Discrete Tracking Detectors*, ATLAS Collaboration Internal Note INDET-NO-132, 21 May 1996.
- [5.11] D. Cavalli, *Missing transverse momentum reconstruction in ATLAS*, ATLAS Collaboration Internal Note PHYS-NO-80, 26 February 1996.

Books and review articles

- [5.12] R. Slansky, *Group theory for unified model building*, Phys. Rep. 79 (1981) 1–128.
- [5.13] G.G. Ross, *Grand Unified Theories*, The Benjamin/Cummings Publishing Company, 1985, ISBN 0–8053–6967–8.
- [5.14] M.B. Green, J.H. Schwartz and E. Witten, *Superstring Theory*, Cambridge University Press, 1987, ISBN 0–521–32384–3.
- [5.15] Joanne L. Hewett and Thomas G. Rizzo, *Low-energy phenomenology of superstring-inspired E_6 models*, Phys. Rep. 183 (1989) 193–381.
- [5.16] John F. Donoghue, Eugene Golowich and Barry R. Holstein, *Dynamics of the Standard Model*, Cambridge University Press, 1992, ISBN 0–521–36288–1.
- [5.17] P. Langacker, M. Luo and A.K. Mann, *High-precision electroweak experiments : a global search for new physics beyond the Standard Model*, Rev. Mod. Phys. 64 (1992) 87.

Superstrings

- [5.18] B. Campbell, J. Ellis, M.K. Gaillard, D.V. Nanopoulos and K.A. Olive, Problems for superstring models with vacuum expectation values for conjugate sneutrinos, Phys. Lett. B180 (1986) 77–82.
- [5.19] G.G. Ross *et al.*, A three-generation superstring model, (I) Compactification and discrete symmetries, Nucl. Phys. B278 (1986) 667–693.
- [5.20] G.G. Ross *et al.*, A three-generation superstring model, (II) Symmetry breaking and the low-energy theory, Nucl. Phys. B292 (1987) 606–652.
- [5.21] G.G. Ross *et al.*, A superstring-inspired standard model, Phys. Lett. B180 (1986) 69–76.
- [5.22] G.G. Ross *et al.*, 27^3 Yukawa couplings for a three-generation superstring model, Phys. Lett. B192 (1987) 111–118.
- [5.23] J.A. Casas and G.G. Ross, A solution to the strong CP problem in superstring models, Phys. Lett. B192 (1987) 119–124.

- [5.24] M. Cvetič and P. Langacker, General signatures of superstring theories, Phys. Rev. D42 (1990) 1797–1806.
- [5.25] M. Drees and A. Yamada, A decisive test of superstring-inspired E_6 models, eprint hep-ph/9508254, August 1995.
- [5.26] M. Cvetič and P. Langacker, Implications of abelian extended gauge structures from string models, eprint hep-ph/9511378, 28 November 1995.

Extended gauge models

- [5.27] R.W. Robinett and J.L. Rosner, Prospects for a second neutral vector boson at low mass in $SO(10)$, Phys. Rev. D25 (1982) 3036–3064.
- [5.28] R.W. Robinett, Neutral currents in E_6 , Phys. Rev. D26 (1982) 2388–2395.
- [5.29] R.W. Robinett and J.L. Rosner, Mass scales in grand unified theories, Phys. Rev. D26 (1982) 2396–2419.
- [5.30] D. London and J.L. Rosner, Extra gauge bosons in E_6 , Phys. Rev. D34 (1986) 1530–1546.
- [5.31] B.W. Lynn, F.M. Renard and C. Verzegnassi, Identification of E_6 -generated Z' from combined CHARM II, LEP I and SLC high-precision measurements, Nucl. Phys. B310 (1988) 237–253.
- [5.32] P. Langacker and M. Luo, Constraints on additional Z bosons, Phys. Rev. D45 (1992) 278–292.

Left-right symmetric models

- [5.33] E. Ma, Left-right symmetry and supersymmetric unification, eprint hep-ph/9405329, 19 May 1994.
- [5.34] E. Ma, Unifiable supersymmetric left-right model with E_6 particle content, eprint hep-ph/9408237, 4 Aug 1994.
- [5.35] E. Ma, Left-right gauge symmetry at the TeV energy scale, eprint hep-ph/9411408, 28 Nov 1994.

- [5.36] K. Huitu, M. Raidal and J. Maalampi, Phenomenological implications of supersymmetry in left-right electroweak model, eprint hep-ph/9501255, 10 January 1995.
- [5.37] E. Akhmedov, M. Lindner, E. Schnapka and J.W.F. Valle, Dynamical left-right symmetry breaking, preprint TUM-HEP-222/95, September 1995.

Forward-backward asymmetries

- [5.38] V. Barger, N.G. Deshpande, J.L. Rosner and K. Whisnant, Production, decays and forward-backward asymmetries of extra gauge bosons in E_6 , Phys. Rev. D35 (1987) 2893–2896.
- [5.39] J.L. Rosner, Off-peak lepton asymmetries from new Z s, Phys. Rev. D35 (1987) 2244–2247.
- [5.40] J.L. Rosner, Observability of charge asymmetries for lepton pairs produced in present collider experiments, Phys. Lett. B221 (1989) 85–90.
- [5.41] P.G. Bright-Thomas, A measurement of the forward-backward asymmetry of $Z \rightarrow \mu^+\mu^-$ and a search for an additional neutral vector gauge boson using electroweak observables at OPAL, PhD thesis, Brunel University, London, May 1995.
- [5.42] J.L. Rosner, Forward-backward asymmetries in hadronically produced lepton pairs, eprint hep-ph/9512299, submitted to Phys. Rev. D.

Strong electroweak sector

- [5.43] R. Casalbuoni, S. de Curtis, D. Dominici and R. Gatto, Effective weak interaction theory with a possible new vector resonance from a strong Higgs sector, Phys. Lett. B155 (1985) 95–99.
- [5.44] R. Casalbuoni, S. de Curtis, D. Dominici and R. Gatto, Physical implications of possible $J = 1$ bound states from strong Higgs, Nucl. Phys. B282 (1987) 235–252.
- [5.45] R. Casalbuoni, P. Chiapetta, D. Dominici, F. Feruglio and R. Gatto, High-energy tests for a possible strong sector in the Standard Model, Nucl. Phys. B310 (1988) 181–221.

- [5.46] R. Casalbuoni, S. de Curtis, D. Dominici, P. Chiapetta, A. Deandrea and R. Gatto, Limits on the BESS model at NLC, eprint hep-ph/9306262, 11 June 1993.
- [5.47] R. Casalbuoni, S. de Curtis, A. Deandrea, N. di Bartolomeo, R. Gatto, F. Feruglio and G. Nardulli, Hadronic transitions among quarkonium states in a soft-exchange approximation. Chiral breaking and spin symmetry breaking processes, *Phys. Lett. B*309 (1993) 163–173.
- [5.48] R. Casalbuoni, P. Chiapetta, A. Deandrea, S. de Curtis, D. Dominici and R. Gatto, Vector resonances from a strong electroweak sector at linear colliders, *Z. Phys. C*60 (1993) 315–326.
- [5.49] R. Casalbuoni, P. Chiapetta, A. Deandrea, S. de Curtis, D. Dominici and R. Gatto, The BESS model at e^+e^- colliders, eprint hep-ph/9309334, 29 September 1993.
- [5.50] G. Cvetič, C. Grosse-Knetter and R. Kögerler, Two- and three-vector-boson production in e^+e^- collisions within the BESS model, *Int. J. Mod. Phys. A*9 (1994) 5313–5358.
- [5.51] P. Chiapetta, Latest developments on the BESS model, eprint hep-ph/9411284, 14 November 1994.
- [5.52] A. Deandrea, The BESS model at future colliders, eprint hep-ph/9411291, 14 November 1994.
- [5.53] R. Casalbuoni, A. Deandrea, S. de Curtis, D. Dominici and R. Gatto, Some results on the BESS model at future colliders, eprint hep-ph/9502325, 16 February 1995.
- [5.54] L. Anichini, R. Casalbuoni and S. de Curtis, Low-energy effective Lagrangian of the BESS model, *Phys. Lett. B*348 (1995) 521–529.
- [5.55] R. Casalbuoni, P. Chiapetta, D. Dominici, A. Fiandrino and R. Gatto, Top quark production in the Extended BESS model, eprint hep-ph/9505212, 02 May 1995.
- [5.56] R. Casalbuoni, A. Deandrea, S. de Curtis, R. Gatto and M. Grazzini, Degenerate BESS model : the possibility of a low-energy strong electroweak sector, eprint hep-ph/9510431, 27 October 1995.



Chapter 6

Tracker layout development

6.1 Introduction

Much of the evaluation of the performance of a tracker layout can be done using analytical calculations and simple simulation programs. In this chapter, results are presented of analytical calculations which were done as part of the development of the Inner Detector layout [1] during the transition from the Cosener's House design to a Panel-style layout. Although historical, the results remain of relevance today since the current layout is a Panel-style design.

6.2 Software

All the track fitting resolutions presented in this chapter were analytic covariance matrix calculations made with the program, TRACKERR [4]. It has the following features:

- The input is an ASCII description of the tracker layout: layer positions and shapes (i.e. cylinder, cone or disk), their precisions, stereo angles, and material types and thicknesses. The input file also describes any non-standard materials, and the graphical display of the results. Materials types are defined in terms of mass number, atomic number, density and radiation length.
- The magnetic field is assumed to be uniform and parallel to the detector symmetry axis.
- The layout description is used to determine which layers would be hit (and at which angle) by a track following a helical path starting at the nominal interaction

point. The resolutions and scattering contributions of the hit layers are used to calculate the track fitting covariance matrix for the 5 track parameters using a fast analytic algorithm [5].

- The effects of multiple scattering, energy loss, and energy loss fluctuation are included. The main limitation is that the helical trajectory is not modified by energy loss, although the effect of limited range due to energy loss is included.

6.3 Layout design work

Historical context

The layout used for the 1992/3 Monte Carlo simulations of the overall performance of the ATLAS detector is known as the Cosener's House layout. After a review in May 1993 it was suggested that the performance be evaluated for a design which differed from the Cosener's layout in a number of ways. It was proposed that the precision tracking layers be positioned within the inner radius of the continuous tracking layers (i.e. TRT). Previously the precision tracking layers (silicon and MSGC) had been positioned between TRT units. It was hoped that separating the detector technologies would result in a simplification of the mechanical design, with corresponding reductions in design complexity and in cost. The reduction of the outer radius of the precision layers was expected to bring a reduction in detector area, cost and channel count.

Following the reviewers' guidelines, two new 'Panel-style' layouts were proposed in December 1993. Both were aimed at addressing the deficiencies which had been identified in the Cosener's layout. A revised Panel-style design for the ATLAS Inner Detector was developed, incorporating elements of both previous Panel designs.

The Panel concept was accepted as the baseline in mid 1994. The layout given in the ATLAS Technical Proposal, which is the one currently coded in DICE, is one such layout.

The major change made to the Inner Detector layout was to move from the Cosener's layout to a Panel-style layout. Since then the layout has been optimised further, but with no radical change in design, reaching the current 'fine-tuned Morges' layout in a number of evolutionary steps.

This study

The work presented here is a snapshot taken in early 1994 of work which was done between late 1993 and mid 1994 as part of the development and optimisation of the Panel-style layout and comparison with the Cosener's layout. It is mostly a historical record of the design issues under consideration at the time, but the the results are still of relevance today.

The performance of one particular Panel-style layout is considered as the precisions are changed and as layers are removed. The aims were three-fold: to simulate the effect of layer loss due to radiation damage, to investigate the relative importance of the discrete and continuous tracking elements, and to determine the optimal resolutions for the inner silicon layers. Track fitting resolutions were calculated and compared with those of the Cosener's layout. Pattern recognition capability was taken into account in the development by consideration of the number of hits and their radial distribution along a track, as a function of pseudorapidity η .

Naming conventions

For want of a better name, the layout described in this chapter is called the Revised Panel Layout (RPL). This was the name used at the time of first publication, and referred to the fact that this was one of a second iteration of layouts.

At the time of the first publication of the results presented here, there were two options for the technology of the forward precision tracking: microstrip gas chambers (MSGCs) and silicon strip detectors (Forward SCT). The silicon option was selected subsequently. The calculations described here are not specific to either technology. For this reason, the discrete forward detectors will be referred to as forward precision tracking (FPT) rather than MSGC or Forward SCT.

The precisions of pixel, silicon strip and FPT layers used in this chapter are quoted in terms of the standard deviation (σ) rather than the strip pitch (p). Thus the results quoted here are presented in a form which is to some extent independent of the choice of electronic readout technology. With the binary electronics option, the standard deviation is slightly better than the top hat standard deviation ($\sigma = p/\sqrt{12}$). Thus a standard deviation of $\sigma = 20(40) \mu\text{m}$ corresponds to strip pitch of $p \sim 69(138) \mu\text{m}$. A strip pitch of $p = 80 \mu\text{m}$ corresponds to a standard deviation of $\sigma \sim 23 \mu\text{m}$.

As in previous chapters, the five track parameters will be referred to as p_T , z_0 , a_0 , ϕ and θ .

6.3.1 The Cosener's House layout

For comparison with the Panel-style layout described in the next section, the Cosener's layout is shown below, with the momentum resolution and the number of hits versus pseudorapidity. A numerical description is given in Table 6.1.

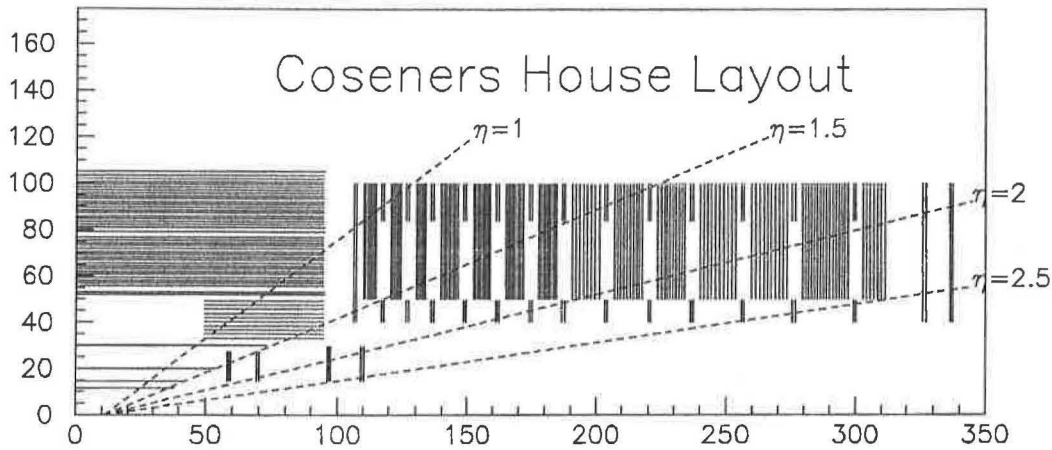


Figure 6.1: *The Cosener's House layout.*

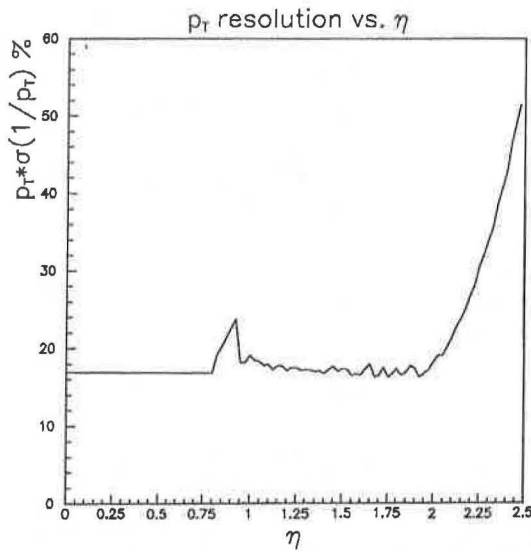


Figure 6.2: *Cosener's layout : p_T resolution for a 500 GeV track.*

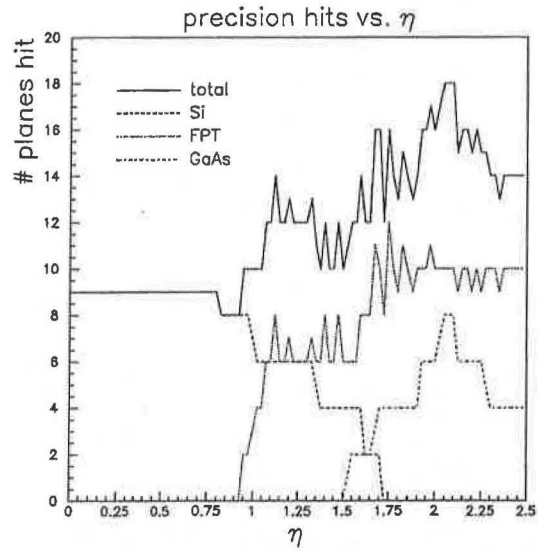


Figure 6.3: *Cosener's layout : number of layers hit by a track vs. pseudorapidity η .*

type	r cm	n	area/pl m^2	A m^2	pitch μm	sections	ch/plane	# channels
FPT	40-50	78	0.28	21.84	200	1×10 cm	12,600	980K
	84-100	78	0.92	71.76	200	1×16 cm	26,400	2.1M
	40-100	18	2.64	47.52	200	5×12 cm	100,500	1.8M
				141.1				4.8M
Silicon	20	1DS	1.34	1.34	100	18×6 cm	452K	452K
	30	1DS	2.75	2.75	100	24×6 cm	905K	905K
	52.5	2SS	6.27	12.53	200	16×12 cm	264K	528K
	80	2SS	9.55	19.10	200	16×12 cm	402K	804K
	105	1SS	12.53	12.53	200	16×12 cm	528K	528K
				48.3				3.2M
GaAs	14.5-27.0	8	0.16	1.30	20(69)	1×12.5 cm	9110	72,800
	14.5-29.5	8	0.21	1.66	20(69)	1×15 cm	9110	72,800
				2.96				146K
Pixel	11.5	1	0.55	0.55	50×200	-	55M	55M
	14.5	1	0.77	0.77	50×200	-	77M	77M
				1.32				130 M

Table 6.1: Cosener's House Layout : Areas and number of channels (no overlaps)

6.3.2 A Panel-style layout

The major differences between the Panel-style layouts and the Cosener's layout are

- 80 cm barrel half-length (reduced from 95 cm)
- precision layers at lower radius
- reduced area and channel count
- barrel TRT no longer interleaved with Si
- forward TRT no longer interleaved with FPT
- simplified TRT mechanics
- increased radial length of inner FPT wheels
- introduction of a TRT 'plug'

The precision tracking layers were placed at low to intermediate radius, leaving uninterrupted continuous tracking in the outer radial range. The intention was to reduce area and simplify the mechanics. By putting the precision layers in separate inner cylinder it was also thought possible to improve access and cooling for the precision layers.

More reliance is placed on the continuous tracking (i.e. TRT) for momentum resolution and triggering, but the changes should increase the robustness of the straws by simplifying the mechanical design.

As explained in the introduction, the results given here are for one particular Panel-style design, the Revised Panel Layout (RPL). This layout is shown graphically in Figure 6.4 and numerically in Table 6.2.

In the next two sections, a detailed description is given of the barrel and forward layout choices and their implementation in the analytical resolution calculations. In the following section calculation results are given for the basic layout and for test cases in which layer precisions are changed and layers are removed.

6.3.2.1 Barrel layout

Length

A ± 80 cm barrel was chosen to reduce silicon area and channel count, and to reduce straw occupancy. The bump in the p_T resolution in the transition region is slightly worse than for a 95 cm barrel, but it remains acceptable.

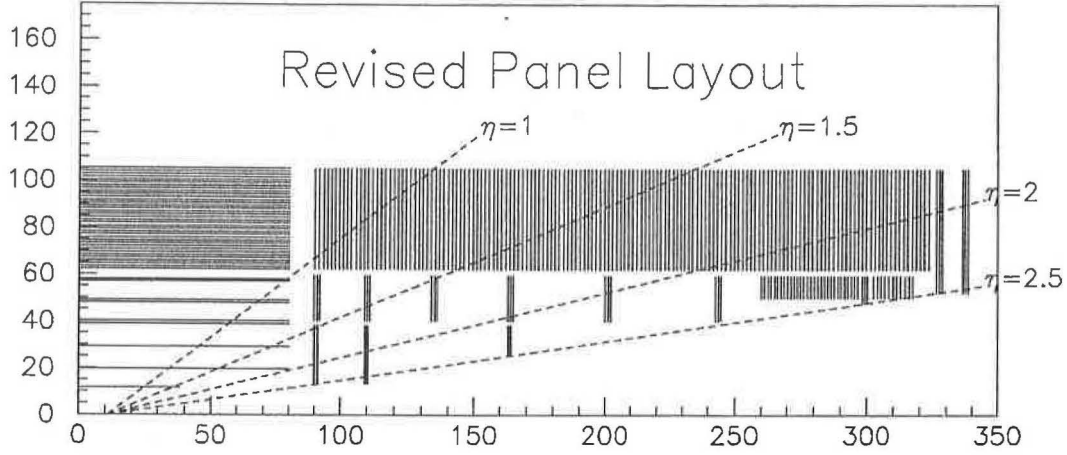


Figure 6.4: An rz quarter-plane view of the Revised Panel Layout.

Barrel silicon :

The outer radius of the barrel silicon was reduced to 60 cm. This, together with the reduction in the number layers from 7 to 6 resulted in a reduction in total silicon area from 52 to 42 m², with a corresponding expected saving in material, power dissipation and cost.

The two inner silicon layers at $r = 20, 30$ cm were entered in the calculation as double-sided silicon detectors with a single-layer precision of $\sigma = 20 \mu\text{m}$, and a stereo angle of 30 mrad, giving a z -resolution $\sigma_z = 600 \mu\text{m}$.

The three outer silicon layers at $r = 40, 50, 60$ cm were entered as pairs of single-sided strip detectors with a precision of $\sigma = 40 \mu\text{m}$, and a stereo angle of 30 mrad. (The choice of precision is explained in Section 6.3.2.3.)

Barrel pixels :

In the RPL layout there was a single layer of silicon pixel detectors at radius 11.5 cm, with a resolution of $\sigma = 20 \times 60 \mu\text{m}$ (corresponding to a pitch of $p = 50 \times 200 \mu\text{m}$). The layer extends to $z = \pm 38.0$ cm ($|\eta| = 1.9$). Due to the lack of a pixel detector type in the software, the pixel layer was simulated in the calculation as a layer of double-sided silicon strips with $\sigma = 20 \mu\text{m}$ and stereo angle 300 mrad.

Barrel TRT :

Due to the removal of the interleaved silicon layers, the RPL barrel TRT is an uninterrupted block of axial straws, which gives 36 straw hits between $r = 62$ cm

and $r = 105$ cm. A precision of $\sigma = 180 \mu\text{m}$ per layer was used in the covariance matrix calculation.

The inner radius of the barrel straws was increased from 55 to 63 cm. The decrease in barrel length and increase in inner radius was expected to reduce the straw count rate and increase the fraction of usable drift-time measurements, and hence improve the $r\phi$ precision.

Vertexing layer :

The vertexing layer proposed for the low-luminosity phase of the experiment was not included in any of the calculations presented here.

6.3.2.2 Forward layout

Forward precision tracking (FPT) :

Seven low-radius forward precision disks were used, filling the radial interval $r = 40 - 60$ cm. Two 'full' end wheels were placed at $z = 324$ cm and $z = 334$ cm covering the radial interval $r = 40 - 105$ cm.

The full disk immediately after the transition region in the Cosener's design was found to be unnecessary for good momentum resolution provided that a resolution of $\sigma = 40 \mu\text{m}$ was used for the inner FPT layers. The full wheels were found to be very important in maintaining good momentum resolution beyond $\eta = 2.0$, and so were retained in the RPL.

The radial length of the inner disks was double that of the Cosener's layout. Longer planes made it geometrically easier to design a layout with an even distribution of hits along a track for all η .

In the Cosener's layout at high η there was a long gap with no space points between the inner silicon and gallium arsenide and the end wheels. The z positions of the disks in the RPL were determined by requiring that there be a constant two FPT superlayers hits for all relevant η . Thus the z positions are given by two interleaved geometric progressions.

In the resolution calculations the FPT precision was taken to be $\sigma = 40 \mu\text{m}$ [11]. The stereo angle was set as before to 17 mrad. Since the (3-layer) MSGC option for the forward precision tracking had lower single plane efficiency than the (2-layer) semiconductor option, it was possible to model both options simultaneously

by using two planes in the calculation rather than three. This had the additional effect of normalising the hits plots to a ‘silicon equivalent’ plane.

GaAs forward disks :

Pairs of single-sided GaAs detectors, with a resolution of $\sigma = 20 \mu\text{m}$ and a stereo angle of 17 mrad, were arranged to obtain one hit in the radial interval $r = 30 - 40 \text{ cm}$. This allowed the possibility of supporting the GaAs planes from the FPT layers — the services may have been difficult for layouts in which the GaAs planes were attached to a tapered silicon barrel.

Forward TRT :

In the RPL, there are no FPT layers between TRT units, giving more continuous tracking. In the covariance matrix calculation 150 radial straws were used at uniform density in z with $\sigma = 180 \mu\text{m}$. The number of TRT hits in this layout exhibited a pronounced peak at $|\eta| \sim 1.9$ indicating that the density should be high-low-high along z in order to keep the number of TRT hits roughly constant. This was implemented in the next layout iteration.

For $|\eta| < 2.0$, the inner radius of the forward TRT was increased from 50 to 63 cm. This reduced the occupancy of the bulk of the straws, whilst maintaining a total of at least 30 hit straws per track at high η .

For $|\eta| > 2.0$ the TRT was extended to lower radius to maintain a high number of straw hits. With this so-called ‘plug’, the number of straw hits at $\eta = 2.4$ is 36.

For reasons of occupancy, the plug was envisaged as at least electrically separate from the rest of the TRT. Mechanical separation of the low-radius and high-radius units would have the additional advantage that in this section ($z = 260 - 324 \text{ cm}$) the ϕ -separation at the outer radius ($r = 105 \text{ cm}$) could remain as low as the rest of the TRT barrel, since these straws do not have to run down to $r = 50 \text{ cm}$.

In the TRACKERR implementation of the plug there were 50 radial straws with $\sigma = 180 \mu\text{m}$, with dimensions $r = 50 - 60 \text{ cm}$, $z = 260 - 324 \text{ cm}$, in two sections, one section either side of the FPT layer at $z = 303$.

type	n	measuring	r cm	z cm	σ μm	stereo mrad
Barrel TRT	36	ϕ	62-105	± 80.0	180	0
Forward TRT	150	ϕ	62-105	90.0-323.4	180	0
TRT Plug	35	ϕ	50-60	260.0-296.5	180	0
	15	ϕ	50-60	302.5-323.4	180	0
FPT	3	ϕ, u, v	40-60	90.0	40	17
	3	ϕ, u, v	40-60	109.3	40	17
	3	ϕ, u, v	40-60	134.4	40	17
	3	ϕ, u, v	40-60	163.1	40	17
	3	ϕ, u, v	40-60	200.4	40	17
	3	ϕ, u, v	40-60	243.2	40	17
	3	ϕ, u, v	47.5-60	298.5	40	17
	3	ϕ, u, v	52.1-105	326.5	40	17
	3	ϕ, u, v	52.1-105	336.5	40	17
Silicon	1 DS	ϕ, u	19.5	± 80.0	20	30
	1 DS	ϕ, u	29.0	± 80.0	20	30
	2 SS	ϕ, u	39.0	± 80.0	40	30
	2 SS	ϕ, u	48.0	± 80.0	40	30
	2 SS	ϕ, u	57.0	± 80.0	40	30
GaAs	2 SS	ϕ, u	13.0-40.0	90.0	20	17
	2 SS	ϕ, u	13.0-40.0	109.3	20	17
	2 SS	ϕ, u	25.1-40.0	163.1	20	17
Pixel	1	ϕ, z	11.5	± 38.0	20×60	-
Vertex	1 DS	ϕ, u	6.0	± 36.3	10	17

Table 6.2: Revised Panel Layout : layer positions, granularities and stereo angles.

type	r cm	n	area/pl m^2	A m^2	pitch μm	sections	ch/plane	# chann
FPT	40-60	36	0.63	22.6	200	2×10 cm	28,300	1.0M
	47.5-60	6	0.42	2.5	200	1×12.5 cm	14,900	89,500
	52.1-105	12	2.96	31.3	200	5×11 cm	115,000	1.4M
				56.4				2.5M
Silicon	19.5	1DS	1.96	1.96	100	28×6 cm	686K	686K
	29.0	1DS	2.92	2.92	100	28×6 cm	1.0M	1.0M
	39.0	2SS	3.92	7.84	200	14×12 cm	172K	343K
	48.0	2SS	4.83	9.66	200	14×12 cm	211K	422K
	57.0	2SS	5.73	11.46	200	14×12 cm	251K	501K
				33.8				3.0M
GaAs	13.0-40	8	0.45	3.60	100	3×9 cm	41,500	332K
	25.1-40	4	0.30	1.22	100	2×7.5 cm	36,200	145K
				4.82				477K
Pixel	11.5	1	0.55	0.55	50×200	N/A	55M	55M
Vertex	6.0	1	0.27	0.27	50	12×6 cm	91K	91K

Table 6.3: Revised Panel Layout: areas and numbers of channels (no overlaps)

6.3.2.3 Resolutions

The results of the momentum and impact parameter resolution calculations are given below. The RPL was also taken as the basis for a number of studies: the p_T and impact parameter resolutions were evaluated as a function of the precision of the silicon layers and the TRT, and the robustness to radiation-induced layer failures was studied.

Basic layout

The precision of the outer three silicon layers in this nominal RPL layout was chosen to be $\sigma = 40 \mu\text{m}$ as a good compromise between cost and p_T performance (see below). The p_T resolution for a 500 GeV track is shown in Figures 6.5 and 6.6 both with and without a $20 \mu\text{m}$ beam constraint. Due to the lower outer radius of the precision tracking, the momentum resolution is slightly worse than that of the Cosener's layout. It is below the target 30% up to $|\eta| = 2.3$.

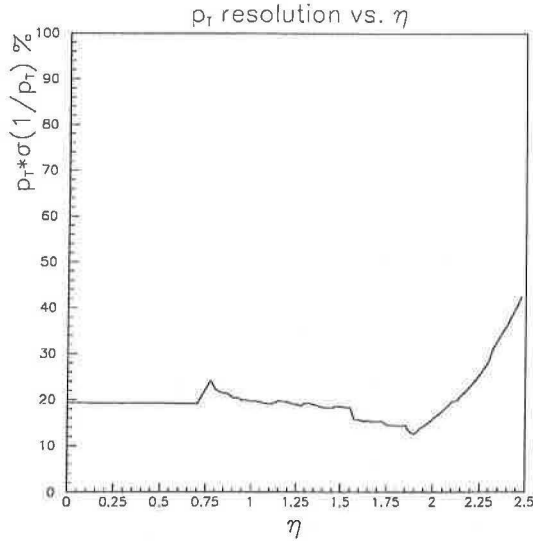


Figure 6.5: *Revised Panel Layout : p_T resolution for a 500 GeV track ($20 \mu\text{m}$ beam constraint).*

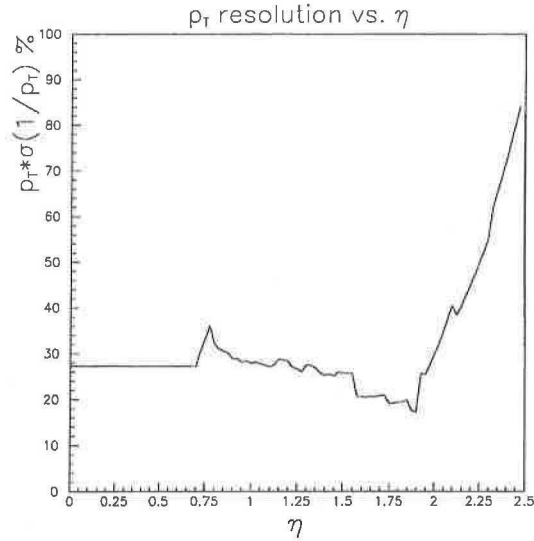


Figure 6.6: *Revised Panel Layout : p_T resolution for a 500 GeV track (no beam constraint).*

The number of hit detector planes versus η is shown in Figures 6.7 and 6.8.

Performance vs. silicon granularity

For financial reasons, the silicon strip pitch should be as large as is compatible with obtaining acceptable track fitting resolutions, vertexing and pattern recognition. In order to determine whether the outer three layers of silicon could have a granularity coarser than the inner layers without adversely affecting momentum resolution, the

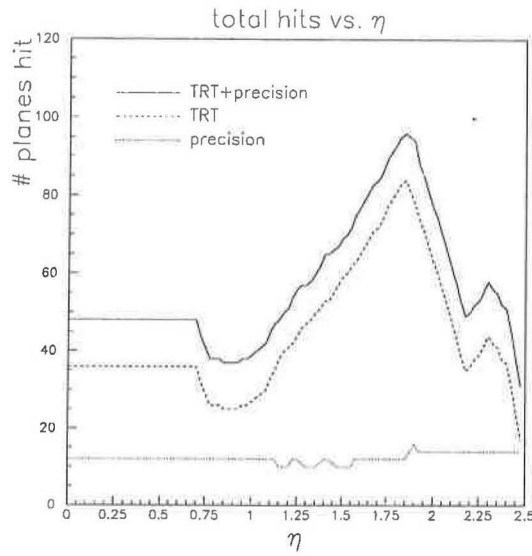


Figure 6.7: Revised Panel Layout : number of layers hit vs. pseudorapidity η .

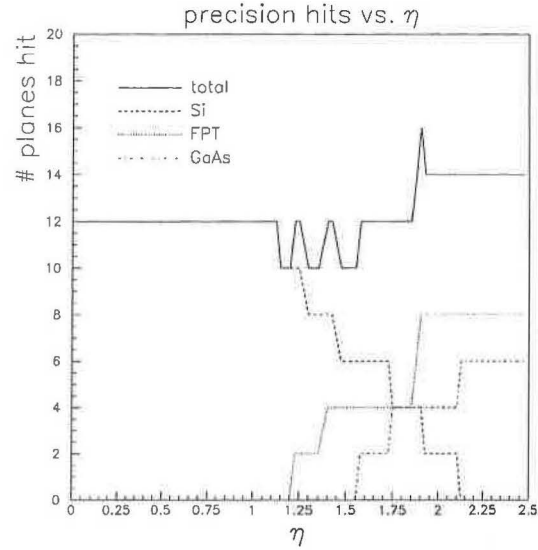


Figure 6.8: Revised Panel Layout : number of precision layers hit vs. pseudorapidity η .

track fitting resolutions were calculated with outer silicon resolutions of $\sigma = 20, 40$ and $60 \mu\text{m}$. The resolution of the inner silicon layers was taken as $\sigma = 20 \mu\text{m}$, i.e. the same order as the beam spot since this limits the ultimate momentum precision. The calculations were made with and without the presence of the TRT in order to evaluate the silicon standalone performance. See Table 6.4.

outer silicon σ	p_T resolution, $p_T \times \sigma(\frac{1}{p_T})$			exceeds 30% at $\eta =$
	for $ \eta \geq 0$	in transition region	maximum	
$20 \mu\text{m}$	17% (32%)	22%	44% (60%)	2.3
$40 \mu\text{m}$	19% (41%)	24%	44% (60%)	2.3
$60 \mu\text{m}$	20% (48%)	25%	44% (69%)	2.3

Table 6.4: p_T resolution vs. outer Si granularity. The numbers given in brackets are for no TRT.

With a full-precision TRT, the p_T resolution depends only weakly on the granularity of the outer silicon. However, the silicon standalone p_T resolution varies strongly as the outer silicon granularity is changed.

Performance with degraded TRT precision

In the RPL, more reliance is put on the continuous tracking for p_T resolution and triggering than in the Cosener's design. In order to investigate the degree of dependence of the momentum resolution on good straw performance and alignment, the calculation was performed firstly with a degraded TRT resolution and then with no TRT to show the silicon standalone performance.

(0) full RPL layout

(1) TRT resolution degraded to $250\ \mu\text{m}$

(2) no TRT

The results are summarised in Table 6.5.

Test	p_T resolution, $p_T \times \sigma(\frac{1}{p_T})$		σ_{a_0}
	at $\eta = 0$	maximum	at $\eta = 0$
	b.c. / no b.c.	b.c. / no b.c.	no b.c.
(0) : full RPL (no vtx layer)	19% / 27%	44% / 75%	$29\ \mu\text{m}$
(1) : $250\ \mu\text{m}$ TRT	22% / 31%	44% / 76%	$31\ \mu\text{m}$
(2) : no TRT	41% / 68%	60% / 96%	$42\ \mu\text{m}$

Table 6.5: p_T and impact parameter resolutions vs. TRT precision, with and without a $20\ \mu\text{m}$ beam constraint.

Performance after radiation damage

Current estimates are that radiation damage will kill layers from the inner radius outwards, reaching about 30 – 40 cm radius after 10 years of operation at $10^{34}\ \text{cm}^{-2}\text{s}^{-1}$, although estimates of the 10 year survival radius vary over the range 20 cm to 50 cm [10]. In order to determine how the track fitting resolutions will be affected as the inner layers are lost, the momentum and impact parameter resolutions were calculated for the following :

(3) Pixel layer at $r = 11.5\text{cm}$ removed.

(4) As (3), but with the additional removal of the silicon layer at $r = 20\text{cm}$.

(5) As (4), but with the additional removal of the silicon layer at $r = 30\text{cm}$.

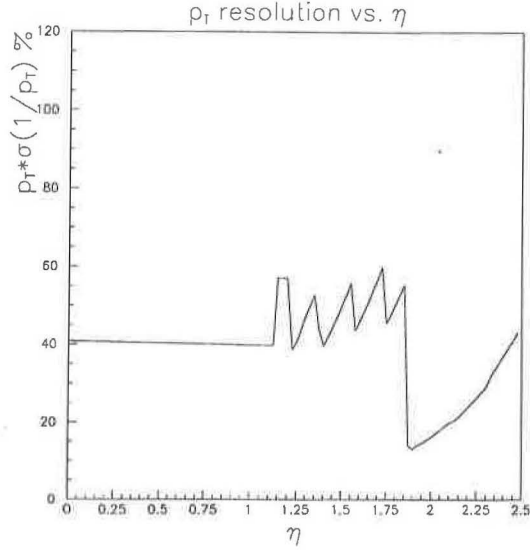


Figure 6.9: *Silicon standalone p_T resolution of the Revised Panel Layout ($20\mu\text{m}$ beam constraint).*

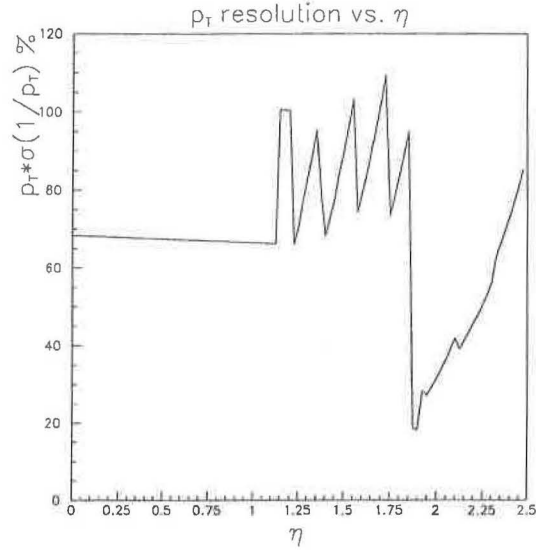


Figure 6.10: *Silicon standalone p_T resolution of the Revised Panel Layout (no beam constraint).*

Each test was done both with and without a $20\mu\text{m}$ beam constraint. The results for a 500 GeV track are summarised in Table 6.6.

Test	$\frac{\Delta p_T}{p_T}$ at $\eta = 0$	Max $\frac{\Delta p_T}{p_T}$	σ_{a_0} at $\eta = 0$
	b.c. / no b.c.	b.c. / no b.c.	no b.c.
(0) : Full RPL (no vtx layer)	19% / 27%	44% / 75%	$29\mu\text{m}$
(3) : (0) – Pix11.5	19% / 35%	44% / 88%	$50\mu\text{m}$
(4) : (0) – Pix11.5 – Si20	20% / 50%	44% / 88%	$91\mu\text{m}$
(5) : (0) – Pix11.5 – Si20 – Si30	24% / 87%	44% / 99%	$300\mu\text{m}$

Table 6.6: p_T and impact parameter resolutions with progressive removal of inner silicon layers, with and without a $20\mu\text{m}$ beam constraint.

With a beam constraint the p_T resolution remains acceptable even for (5). Without the beam point the effect of layer loss is stronger. The results confirm expectations that a good impact parameter resolution requires high-precision layers at low radius.

6.4 Conclusions

The Revised Panel Layout presented above is mechanically simpler than the Cosener's House design and appears to meet the physics requirements. The slightly lower p_T resolution in the barrel is an inevitable consequence of reducing the outer radius of the precision layers, without a corresponding increase in granularity.

The detector areas and channel counts are given for the Cosener's House layout in Table 6.1, and for the Revised Panel Layout in Table 6.3. The FPT area is 60% lower in the Revised Panel Layout than in the Coseners layout. Silicon area is reduced by 30%. Gallium arsenide area has increased from 3.0 m^2 to 4.8 m^2 . The FPT surface area is reduced from 47 to 20 m^2 , and the channel count from 4.6×10^6 to 2.3×10^6 .

It was later shown that together with the increase in the inner radius of the TRT from 55 cm to 63 cm, the decrease in the barrel length reduced the peak (average) hit probability of the barrel straws from 34% (25%) to 24% (18%). The 30% reduction in the straw count rate and an increase in the average number of hits with a usable drift time measurement from ~ 25 to ~ 28 for an average of 33 hit straws, gives a more precise and more robust $r\phi$ measurement. The above estimates are given for a luminosity of $10^{34} \text{ cm}^{-2} \text{ s}^{-1}$.

In the RPL, the target of 30 straw hits was not met in the transition region between the barrel and forward regions. In the Cosener's design extra TRT barrel modules were positioned at low radius to cover this region. Optimisation of the transition region later resulted in a smoother transition from cylinder to disk geometry.

In order to slow the effects of radiation damage to the silicon, operation at low temperature ($< 5^\circ \text{C}$) is foreseen. This is more easily achieved in a Panel-style layout since the layers are grouped into a single cold volume.

A Panel-style layout was coded for full simulation and was used as the basis for detailed studies in the run-up to the publication of the ATLAS Technical Proposal.

Continued tracker development

Following the Coseners/Panel step-change, the layout of the Inner Detector has continued its evolution towards the current (fine-tuned Morges) layout in a more continuous fashion. The silicon option for the forward precision tracking has now been chosen as the baseline technology. The number of forward precision layers has been increased from 7 to 9 for added pattern recognition robustness. The non-uniform magnetic field has been taken into account in the optimisation. Continued work on the detectors has

resulted in a choice of a pitch of $p = 80\,\mu\text{m}$ for all silicon layers. When used with the binary readout option, this will give a resolution of slightly better than $\sigma \sim 23\,\mu\text{m}$.

References

- [6.1] A. Fox-Murphy, R. Nickerson, A. Weidberg, *Tracker layout optimisation*, ATLAS Collaboration Internal Note INDET-NO-043, March 28 1994.
- [6.2] R.L. Gluckstern, *Uncertainties in track momentum and direction due to multiple scattering and measurement errors*, Nucl. Instr. and Meth. 24 (1963) 381-389.
- [6.3] A. Parker, *A Proposal for Simulation for the ATLAS Inner Detector*, University of Cambridge, December 1993.
- [6.4] TRACKERR version 1.44, written by Walt Innes, Stanford Linear Accelerator Centre. Latest version available by anonymous FTP from
<ftp://ftp.SLAC.Stanford.edu/software/trackerr>.
- [6.5] P. Billoir, *Track fitting with multiple scattering: a new method*, Nucl. Instr. and Meth. A225 (1984), 352-366.
- [6.6] C. D'Ambrosio, T. Gys, H. Leutz, U. Gensch and S. Schlenstedt, *Optimization of central particle tracking at future hadron colliders*, Nucl. Instr. and Meth. A322 (1992) 20-33.
- [6.7] A.Clark and G.Taylor, *Proposal to Study a Revised Inner Tracker Layout for ATLAS*, ATLAS Collaboration Internal Note INDET-NO-29, 25-OCT-93.
- [6.8] P. Allport, *Charge Division and Silicon Resolution*, presentation to ATLAS-UK/BSP, Nov 1993.
- [6.9] D.Froidevaux and M.Parker, *The Performance Specifications of the ATLAS Inner Detector*, ATLAS Collaboration Internal Note INDET-NO-46, 06-MAY-94.

Radiation damage

- [6.10] A. Chilingarov and S. Roe, *Radiation Damage Projections : Current Status*, ATLAS Collaboration Internal Note INDET-NO-31, 16 Nov 1993.
- [6.11] N. Jackson, Private communication, University of Liverpool.

Layouts

- [6.12] S. Gadomski et al., *A possible new layout for barrel part of ATLAS ID*, ATLAS Collaboration Internal Note INDET-NO-35, 30-NOV-93.
- [6.13] D.Froidevaux and M.Parker, *The Cosener's and Panel Layouts: Criteria for decision*, ATLAS Collaboration Internal Note INDET-NO-47, 06-MAY-94.
- [6.14] B. Van Eijk and J. Schmitz, *A proposal for the ATLAS Inner Tracker based on the Panel and Cosener's House layouts*, ATLAS Collaboration Internal Note INDET-NO-50, 21-JUN-94.
- [6.15] G.Lutz and R.H.Richter, *Detector layout and optimization of the detector front-end electronics system*, ATLAS Collaboration Internal Note INDET-NO-56, 20-JUL-94.
- [6.16] J. Baines et al., *Inner Detector Layout for Technical Proposal*, ATLAS Collaboration Internal Note INDET-NO-66, 14-OCT-94.
- [6.17] A. Clark and A. Poppleton, *Optimization of the forward SCT layout*, ATLAS Collaboration Internal Note INDET-NO-100, 10 March 1995.

Pattern recognition

- [6.18] V.V.Ammosov et al., *Pattern recognition study in the barrel inner tracker of the ATLAS*, ATLAS Collaboration Internal Note INDET-NO-104, 11-APR-95.

Track fitting resolutions

- [6.19] S. Haywood, *Variation of Momentum Resolution with Beam-Spot Size*, ATLAS Collaboration Internal Note INDET-NO-49, 01-JUN-94.
- [6.20] B. Dolgoshein et al, *Impact of the choice of TRT gas and operating point on the ATLAS Inner Detector momentum resolution*, ATLAS Collaboration Internal Note INDET-NO-51, 24-JUN-94.

- [6.21] S. Haywood, *Impact parameter resolution in the presence of multiple scattering*, ATLAS Collaboration Internal Note INDET-NO-091, January 1995.
- [6.22] A. Poppleton, *p_T resolution with the post-Annecy tracker layout*, ATLAS Collaboration Internal Note INDET-NO-077, 20 January 1995.
- [6.23] F.J.Fares and G.N.Taylor, *A Study of the Transition Region in the Continuous Tracker (TRT) of the ATLAS Inner Detector*, ATLAS Collaboration Internal Note INDET-NO-88, January 1995.
- [6.24] T. Bowcock et al., *Report on a silicon backup to the MSGC detectors of ATLAS parts A and B*, ATLAS Collaboration Internal Note INDET-NO-68/69, 27-OCT-94.
- [6.25] R. Hawkings, *Electron Triggering in the Transition Region of the ATLAS Inner Detector*, ATLAS Collaboration Internal Note INDET-NO-96, 01-FEB-95.
- [6.26] P.Luthäus et al., *Tau Identification in the ATLAS Inner Tracker*, ATLAS Collaboration Internal Note INDET-NO-99, 17-FEB-95.



Chapter 7

Conclusions

*I know my physic will
work For this night, to
bed, and dream on the event. Farewell.*

Exit

— TWELFTH NIGHT 2.3

The performance of the ATLAS detector will be critically dependent upon the accuracy with which its tracking elements are aligned. In order to meet the stringent requirements imposed by the desired sensitivity to new physics channels, it will be necessary to align components with high precision. The physics requirements, engineering constraints and a number of the available alignment technologies under consideration were presented in Chapter 2. There are a number of technologies and techniques under consideration for each of the alignment stages.

The Frequency Scanned Interferometry technique described in Chapter 3 looks very promising as the basis of a run-time survey system for the precision layers of the Inner Detector. It is designed to provide a large number of quick absolute length measurements whilst meeting the requirements of high radiation resistance, low size and mass, low cost per channel, and on-detector simplicity.

Initial laboratory tests confirm that the interferometer operates as required. In the near future, an external-cavity semiconductor diode laser will be added to the demonstration system to permit the demonstration of high precision measurements. Development of a laser suitable for the final system which will deliver high power and long range smooth tuning at low cost is underway.

Geodetic combination of survey measurements was introduced in Chapter 4. Use of a geodetic network for combination of survey measurements is attractive both to fully exploit the precision of the underlying measurement technique and because of the possibility of using the high degree of internal redundancy to eliminate certain types of measurement biases. Geodetic techniques are not specific to FSI or even one-dimensional measurements and may find wider application in the Muon Spectrometer and in the process of location of ATLAS with respect to the LHC and the cavern walls. The preliminary networks analysed in Chapter 4 appear to meet the specifications.

It is clear that the need for good track fitting resolutions places stringent requirements on the alignment precisions of the SCT layers. The single track Monte Carlo study presented in Chapter 5 has provided essential information about the effect of alignment inaccuracies on the track fitting resolutions. Of the SCT alignment modes considered, the Φ -alignment specification of $11\,\mu\text{m}$ derived from the impact parameter resolution sets the most stringent survey requirements. It is expected that when misalignments in the TRT are taken into consideration, the Φ -alignment requirement derived from the p_T resolution will be comparable.

A Z -alignment requirement of $520\,\mu\text{m}$ for all layers (pixel and strips) is set by the θ resolution specification. Alignment of the pixels with a precision of $\sigma_Z^{\text{pix}} = 200\,\mu\text{m}$ would result in a z_0 resolution sufficient to improve the performance of the Muon Spectrometer.

An R -alignment requirement of $530\,\mu\text{m}$ is set by insisting that the impact parameter resolution is degraded by no more than 20% from its value for perfect alignment.

The three alignment components R , Φ and Z were considered individually. Investigation of the effect of finite alignment precisions in all three directions simultaneously would involve no changes to the software, and so it is possible that such a study could be conducted in the near future.

A W' with a mass up to about 5 TeV could be detected at ATLAS in one year of LHC running at a luminosity of $10^{34}\,\text{cm}^{-2}\text{s}^{-1}$. If perfectly aligned, the ATLAS detector will be sensitive to the forward-backward asymmetry of the semileptonic decay of the W' over all of the accessible mass range. Although the measured integrated asymmetry is diluted by misalignments of the barrel SCT, the signal for $m_{W'} = 3\,\text{TeV}$ is not lost completely, indicating that for the perfect TRT alignment scenario modelled here, the TRT standalone performance is sufficient for detection of this signal. The relative

integrated asymmetry of a 6 TeV W' is likely to be a steeper function of the tracker performance, and may set stronger alignment requirements.

Future work will focus on developing a model of misalignments in the TRT and using it to extend the current study to include the effect of the TRT alignment on both single-track fitting resolutions and physics signals. Similarly, the software could be extended to include the alignment of the Muon Spectrometer. An improved implementation of the reconstruction algorithm is expected in the near future which may be sufficiently fast to allow the inclusion of pile-up.

The current baseline for the layout of the Inner Detector is a Panel-style layout similar to that of Chapter 6, with modifications which resulted from its continued optimisation.

The Future

As many times before in the history of science, we find ourselves at a point where further theoretical progress cannot be made without new experimental discoveries. The Large Hadron Collider will offer an unparalleled opportunity to probe fundamental physics at an energy scale beyond our current experience. The ATLAS detector promises to fully exploit the potential of the LHC for improving our knowledge of the fundamental structure of nature.



Index

- acousto-optic modulator, 59, 73
- adjustment, 112, 115
- ALICE, 8
- alignment
 - and track fitting, 136
 - introduction, 35–50
 - local and global, 35
 - Muon Spectrometer, 26
 - phases of, 35
 - requirements, 36, 150, 207
 - techniques, 39
 - TRT, 145, 208
 - with tracks, 39, 51
 - with X-rays, 44
- ATLAS, 8
- ATLAS experiment, 12
- ATRECON, 134
- B-physics, 31, 144
- basis-independent precision, 119
- beam pipe, 35
- BESS model, 153
- bias, 118
- bias-to-noise ratio, 118
- binomial errors, 165
- calorimeter, 14, 20
 - forward, 20, 22
 - hadronic tile, 22
 - liquid argon, 20
- cathode strip chambers, 25
- charge sign i.d., 163, 165, 173
- CMS, 8
- corner-cube retroreflector, 69
- covariance matrix, 115
- data snooping, 118
- DICE, 134
- digital wavefront measurement, 73
- diode laser, 84
- direct phase detection, 73, 79
- E_6 models, 152
- electroweak
 - GSW model, 2
 - symmetry breaking, 2, 4, 12, 153
- ESPI, 41
- FEA, 41, 42
- Feynman-x, 162, 170
- finite element analysis, *see* FEA
- forward-backward asymmetry, 160, 165,
172, 174, 177
- frequency scanned interferometry, *see* FSI
- fringe analysis, 75
- FSI, 40, 51–108, 206
 - demonstration system, 81
 - interferometer, 81, 87
 - optical frequency measurement, 66,

- GaAs detectors, 17, 194, 201
- gauge field theory, 1
- geodesy, 109
- geodetic networks, 41, 52, **109–132**, 207
 - calculations, 124
 - design, 122
- hadronic scintillator calorimeter, 22
- Higgs
 - boson, 3, 153
 - mechanism, 2
 - search, 29, 144
 - strongly interacting, 30
 - supersymmetric, 30
- hypothesis testing, 118
- Inner Detector, 13, 17
- integrated asymmetry, 174, 177
- interferometer
 - FSI, 52, 57
- interferometry
 - frequency scanned, *see* FSI
 - phase-shift, 73
 - speckle pattern, *see* ESPI
- iPatRec, 134, 136
- jewels, 70
- Large Hadron Collider, 8
- laser
 - linewidth, 66, 80
 - performance, 66, 79, 80, 84, 91, 92
- layouts, **186–205**
 - Coseners, 186, 187, 189
 - Morges, 201
 - Panel-style, 186–188, 191, 208
 - left-right symmetric models, 153
 - LEP, 8
 - LHC, 8, 207
 - LHC-B, 8
 - liquid argon calorimetry, 20
 - minimum detectable bias, 118
 - mode hops, 63
 - momentum resolution, 144, 177, 197
 - monitored drift tubes, 25
 - Monte Carlo software, 133
 - Muon Spectrometer, 14, 25, 144, 207
 - offline computing, 27
 - optical metrology, 40
 - parity violation, 2, 153
 - phase unwrapping, 74
 - phase-shift interferometry, 73
 - phase-stepping, 73
 - photomultiplier, 86
 - pixel detectors, 14, 17, 18, 124, 131, 133, 138
 - presampler layer, 22
 - preshower detector, 22
 - PSI, 73
 - PYTHIA, 154, 156
 - quantum
 - chromodynamics, 3
 - electrodynamics, 2
 - field theory, 1
 - radiation hardness, 17, 20, 22, 46, 47, 52, 57, 69, 199, 206
 - RASNIK, 44, 110
 - reference interferometer, 68, 79

- refractive index, 60
- reliability testing, 118
- resistive plate chambers, 25
- retroreflectors, 53, 69, 93
- Revised Panel Layout, 188
- SCAN-3, 110, 121
- semiconductor diode laser, 84
- silicon strip detectors, 14, 17, 192
- SLUG, 133
- software
 - ATLAS Monte Carlo, 133
 - PYTHIA, 156
 - SCAN-3, 121
 - TRACKERR, 186
- speckle pattern interferometry, *see* ESPI
- Standard Model, 1, 152
 - inadequacies, 4
- stereo strips, 17, 137, 146, 179
- straightness monitor
 - multipoint, 46, 110
 - RASNIK, 44, 110
- supersymmetry, 6
 - SUSY Higgs, 30
 - SUSY search, 31
- survey system, 18, 35, 52
- thin gap chambers, 25
- top quark, 3, 30, 144
- track fitting
 - specifications, 144
- tracker layouts, **186–205**
- TRACKERR, 186
- transparent Si straightness monitor, 46
- transverse mass, 157, 167
- trigger, 27
 - architecture, 27
 - Level 2 track, 37
 - Muon Spectrometer, 25
- TRT, 18, 192, 194, 198
- unification, 4, 5
- vertexing, 18, 193
- vibration, 73
- W' Monte Carlo study, **152–179**
- X-ray alignment, 44