



FACULTY OF MATHEMATICS AND NATURAL SCIENCES
DEPARTMENT OF PHYSICS
UNIVERSITY OF WUPPERTAL

Performance of the Insertable B-Layer for the ATLAS Pixel Detector during Quality Assurance and a Novel Pixel Detector Readout Concept based on PCIe

Dissertation zur Erlangung des Doktorgrades
vorgelegt von
TIMON HEIM



CERN-THESIS-2016-085
28/08/2015



July 2, 2015

Dedicated to my father.

Abstract

During the first long shutdown of the LHC the Pixel detector has been upgraded with a new 4th innermost layer, the Insertable B-Layer (IBL). The IBL will increase the tracking performance and help with higher than nominal luminosity the LHC will produce. The IBL is made up of 14 staves and in total 20 staves have been produced for the IBL. This thesis presents the results of the final quality tests performed on these staves in an detector-like environment, in order to select the 14 best of the 20 staves for integration onto the detector. The test setup as well as the testing procedure is introduced and typical results of each testing stage are shown and discussed. The overall performance of all staves is presented in regards to: tuning performance, radioactive source measurements, and number of failing pixels. Other measurement, which did not directly impact the selection of staves, but will be important for the operation of the detector or production of a future detector, are included.

Based on the experience with readout systems of the IBL, a novel readout concept has been developed. This concept is based around the idea of moving as much functionality into the software of the controlling host computer as possible. The YARR system was developed with focus on the usage of modern multi core CPU architectures, and an FPGA interfaced via a PCIe link. The initial implementation is performed for the FE-I4 chip, which is used in the IBL, but not limited to one chip type, due to the flexible software implementation. The hardware chosen for YARR, is the SPEC board, which is a low cost off-the-shelf PCIe card carrying an FPGA, which can be used as a reconfigurable I/O interface. The firmware for the FPGA, as well as the software is described in-depth and its performance is evaluated for the usage with FE-I4s and extrapolated for the usage with future detector front-ends. In comparison with existing FE-I4 readout systems, YARR shows exceptional performance for much lower cost. This makes it very interesting for the usage in laboratories and for testbeam campaigns. The software is easily scalable to higher performance systems and the necessary steps to replace the existing hardware with another one to accommodate higher link speed are discussed in the conclusion.

Contents

Introduction	3
1 Particle Physics	5
1.1 The Standard Model	5
1.1.1 Matter Particles	5
1.1.2 Fundamental forces	6
1.1.3 Broute-Englert-Higgs Mechanism	10
1.2 Physics Beyond the Standard Model	12
2 The ATLAS Experiment	13
2.1 ATLAS Detector	14
2.1.1 Inner Detector	14
2.1.2 Calorimeter System	17
2.1.3 Muon Chambers	18
2.2 The ATLAS Detector Performance during Run 1 and Long Shutdown 1	19
2.3 Towards the High Luminosity LHC	21
3 The Insertable <i>B</i>-Layer	23
3.1 Overview	23
3.2 Motivation	26
3.3 Sensor Technologies	28
3.3.1 Planar Pixel Sensors	28
3.3.2 3D sensors	29
3.4 Front-End Technology	29
3.5 The IBL Readout System	32
3.6 Powering and Detector Control System	35
3.7 Construction and Integration	36
4 The IBL Stave Testing and Performance	39
4.1 Timeline	39
4.2 Stave Test Stand	41
4.3 Stave Test Procedure	44
4.3.1 Optical inspection	45
4.3.2 Electrical functionality	46
4.3.3 Reception Test	47
4.3.4 Calibration	47

CONTENTS

4.3.5	Source Scan	47
4.4	Database and Analysis Framework	48
4.5	Stave Performance	50
4.5.1	Tuning Analysis	50
4.5.2	Source Scan Analysis	53
4.5.3	Bad Pixel Analysis	54
4.5.4	Stave Selection	55
4.6	Other Measurements and Lessons Learned	56
4.6.1	ToT Calibration	57
4.6.2	Comparison of Pixel Defects during Module and Stave QA . .	62
4.6.3	Increased Noise on 3D FBK Modules	62
4.6.4	Double Chip Module with one dead Chip	64
4.7	Conclusion and Thoughts for Future Pixel Detectors	65
5	Development of a Novel Readout System for Pixel Detectors	67
5.1	Overview	67
5.2	Motivation	68
5.2.1	Architecture	68
5.2.2	Front-End	69
5.2.3	Developments for Future Detectors	71
5.3	Project YARR	71
5.3.1	Hardware	72
5.3.2	Firmware	73
5.3.3	Software	82
5.3.4	FE-I4 Scans	87
5.3.5	Performance	97
5.3.6	Combined Performance	101
5.4	Discussion of Results	102
5.4.1	Processing Performance	103
5.4.2	FE-I4 Implementation	104
5.4.3	Comparison to Existing Systems	104
5.5	Outlook	105
6	Conclusion	107
6.1	IBL - the new innermost tracking layer of the ATLAS detector	107
6.2	Yet Another Rapid Readout	108
A	Appendix	111
A.1	Stave Naming Scheme	111
A.2	FE-I4 Configuration Statistics	112
	List of Tables	116
	List of Figures	117

Bibliography	123
---------------------	------------

Introduction

“You will have to brace yourselves for this - not because it is difficult to understand, but because it is absolutely ridiculous: all we do is draw little arrows on a piece of paper - that’s all!” - Richard P. Feynman [23]

What does particle Physics look like to a scientist? In the quote above Richard Feynman speaks of a theory which aims at describing all fundamental particles and their interactions. Although current attempts till now have yielded precise answers to many questions, we are still far away from a ‘theory of everything’ that could answer all of them. There are many suggestions as to how the current particle physics models could be extended to accommodate for the unanswered questions and only experimental results can begin to hint at which, if any, are the correct approaches.

Currently, the most successful is the Standard Model, discussed in Chapter 1, which provides a near complete description of fundamental particles accounting for many of the phenomena we see today. Only a small fraction of the particles described by the standard model are stable and the rest, typically of much larger mass than the stable particles, can only be produced at specialised facilities. Since these particles decay so quickly only their remnants and decay products may be detected directly in special detectors and the original particle’s presence, inferred.

One such experiment is the ATLAS detector at the Large Hadron Collider (LHC) at Cern, Geneva. The ATLAS experiment, which will be discussed in Chapter 2, analyses remnants of collisions produced by the LHC’s acceleration of very high energy proton beams - it is currently the most energetic particle beam produced. The protons collided head-on have a centre of mass energy of up to $\sqrt{s} = 14$ TeV which is available for creation of new particles. Of these new particles, more than 99% of are accounted for in the standard model and are well described and understood; the challenge lies in producing and detecting new fundamental particles. Since these are so rare, a high number of collisions and a very high centre of mass energy are required - both of these the LHC can deliver but it is with the ATLAS experiment that the detection, digitisation, and ultimately analysis of this particle data can be performed.

The first major accomplishment of the ATLAS experiment was the discovery of the Higgs boson, proving an essential mechanism of the Standard Model, during the first operation period of the LHC from 2010 to 2012. For many physicists the discovery was bitter-sweet, for whilst it confirmed a vital piece of the standard model it left many questions unanswered, and so the motivation for a second run was clear: to probe the standard model at an even higher precision and higher energy to see to what extent it can continue to account for physical observations. This next data run at the LHC begins in Summer 2015 after a long shutdown of one and a half years. During the long shutdown the ATLAS detector was prepared for the challenges of the run 2 environment. The pixel detector, the innermost tracking system in ATLAS, had a new layer added to it to increase the detector’s performance and secure the high quality tracking that it already provides. This new layer now sits

the closest to the beam pipe and is called the Insertable B-Layer (IBL), deploying state-of-the-art silicon detector technology to provide an even closer insight to the collisions occurring in ATLAS.

The IBL, specifically the production and testing of the detector, is in the focus of Chapter 3 and 4. It consists of sensor modules which are loaded onto carbon foam structures, called staves, to provide support and cooling for the electronics. The IBL itself consists of 14 of these staves and in total 20 staves have been produced for the IBL. The staves are tested in the stave quality assurance test stand, where they are operated in a detector-like environment. This thesis presents the quality assurance test stand and the testing procedure which has been performed for every stave. In the scope of this thesis, not only is the analysis of data and the selection procedure shown, but also additional measurements, their subsequent analysis, and their significance for the operation of the detector are discussed.

The second major topic and content of Chapter 5 of this thesis is the development of a novel readout concept called YARR. The novel concept uses a high-speed PCIe link to minimise the amount of processing which needs to be performed in FPGA firmware and uses the high performance of modern computer systems to perform all data processing in software. This poses a paradigm shift for readout systems which, conventionally perform as much processing as is possible in FPGAs since computer systems till recently were not capable of coping with the amount of data and custom interface protocols. A PCIe card with an FPGA is used to build a flexible I/O interface to aggregate the data from detector modules and send them to the CPU for processing. This thesis will present the work performed to implement this concept with the goal to read out the IBL readout chip and benchmark modern multi core CPU architectures for readout purposes. The results which have been obtained with YARR are very promising and makes it an attractive alternative to currently existing readout systems, as it is relatively cheap, readily available, and outperforms the current systems. The usage of YARR for future readout chips and the advantages which could be gained from using a more software heavy implementation are discussed.

Chapter 1

Particle Physics

In an endeavour to identify the building blocks of nature, particle physics is searching for elementary particles and their interactions with each other. Experimental searches in this field are performed at high energy particle collider experiments, which could be described as being analogous to microscopes looking deeper and deeper into the structure of nature. This chapter is intended to present a very a brief overview of the currently dominant theoretical model, the Standard Model, and special attention to the implications of the latest addition to the model: the Higgs boson. It is included, but by no means to be taken as a complete overview of this rich and complex model. Instead to provide one path and motivation for the developmental hardware work, and main focus of the thesis, that comes after it.

1.1 The Standard Model

The theoretical description of elementary particles and their interactions is contained in the Standard Model (SM) [20], which has been a very successful theory and made many predictions which were later confirmed by experiment. However, the SM is an incomplete theory, there are many phenomena which it does not explain: the existence of dark matter, gravity, neutrino masses and matter anti-matter asymmetry, to name a few. The SM can be divided into fermions and bosons, the former being the constituents of matter and the latter, mediating the fundamental interactions between the particles - distinguished by their half-integer, and integer spin, respectively. Three fundamental forces are described in the SM: electromagnetic, weak, and strong interaction. The latest addition to the SM is the Higgs boson, a long postulated particle emerging from the Higgs field, which solves the issue of specific bosons having mass, which used to be a major flaw in the SM.

1.1.1 Matter Particles

The SM contains 3 generations of fermions divided into quarks and leptons, listed in Tab. 1.1 and Tab. 1.2, respectively. One of the distinct differences between quarks and leptons is, that quarks have colour charge (“red”, “blue”, “green”) and can

interact via the strong force whilst leptons do not and cannot. The electromagnetic force can interact with all charged particles and the weak force interacts with all fermions. Each fermion has an anti-particle having the exact same characteristics as its partner, except for the charge whose sign is inverted. The only particle of the second or third generation, which can be observed in nature, is the muon, which is produced in the Earth's atmosphere when cosmic rays collide with it. All visible matter in the universe is made of up and down quarks building protons and neutrons, with electrons orbiting the nucleus they comprise. The other particles can only be produced in particle accelerators, but as they all have very short lifetimes only their decay products can be observed. The particles of each generation behave similarly to the ones from other generations, except for the increasing mass which also allows for more decay channels. The SM makes no prediction for the number of generations: a fourth generation of particles is possible, but as of yet has not been observed.

Quarks can, due to a phenomenon called confinement, only be observed in colour neutral ("white") compound particles, known as hadrons. Hadrons can be made up of two quarks, known as mesons, and three quarks, known as baryons. The number of observed hadrons led, initially, to the belief in a wild particle zoo, until their quark substructure was discovered.

Table 1.1: List of quarks in the SM [20].

	1. Generation		2. Generation		3. Generation	
Charge	u	$2/3 e$	c	$2/3 e$	t	$2/3 e$
Spin	(up)	$1/2$	(charm)	$1/2$	(top)	$1/2$
Mass		$\approx 2.3 \text{ MeV}$		$\approx 1.28 \text{ GeV}$		$\approx 173 \text{ GeV}$
Charge	d	$-1/3 e$	s	$-1/3 e$	b	$-1/3 e$
Spin	(down)	$1/2$	(strange)	$1/2$	(bottom)	$1/2$
Mass		$\approx 4.8 \text{ MeV}$		$\approx 95 \text{ MeV}$		$\approx 4.18 \text{ GeV}$

Table 1.2: List of leptons in the SM [20].

	1. Generation		2. Generation		3. Generation	
Charge	e	$-1 e$	μ	$-1 e$	τ	$-1 e$
Spin	(electron)	$1/2$	(muon)	$1/2$	(tau)	$1/2$
Mass		0.51 MeV		106 MeV		1.78 GeV
Charge	ν_e	0	ν_μ	0	ν_τ	0
Spin		$1/2$		$1/2$		$1/2$
Mass		$< 2.2 \text{ eV}$		$< 17 \text{ keV}$		$< 15.5 \text{ MeV}$

1.1.2 Fundamental forces

According to the SM, interactions are mediated by the exchange of a gauge boson. The gauge bosons of the three fundamental forces are listed in Tab. 1.3. Although

being the oldest known fundamental force, gravitation is not yet included in the SM, primarily due to coupling to mass and therefore being too weak to be measured with particles. In a similar manner to identifying elementary particles, the SM strives to unify all forces in one theory, for instance Maxwell unified electricity and magnetism with his theory of electromagnetism. Later it was then discovered that electromagnetism can be combined with the weak force to form the electroweak force. Each fundamental force has, at least, one boson associated with it and these bosons couple to an equivalent ‘charge’ carried by the fermions. The electroweak force is mediated by four gauge bosons: W^0 , W^+ , W^- and B^0 . Two of these gauge bosons coalesce to the two observed gauge bosons γ and Z^0 :

$$\begin{pmatrix} \gamma \\ Z^0 \end{pmatrix} = \begin{pmatrix} \cos \theta_W & \sin \theta_W \\ -\sin \theta_W & \cos \theta_W \end{pmatrix} \begin{pmatrix} B^0 \\ W^0 \end{pmatrix} \quad (1.1.1)$$

With θ_W being the weak mixing angle. The electroweak unification is a big success for particle physics, but is inherently flawed. On its own it does not allow the $W^\pm$ and Z to have any mass. This problem is solved with the addition of the Brout-Englert-Higgs mechanism to the SM, which is discussed in Section 1.1.3.

Table 1.3: Mediating particles of the fundamental forces [20].

Gauge boson	Charge	Spin	Mass	Force
γ (photon)	0	1	0	electromagnetic
Z (Z boson)	0	1	91.2 GeV	weak interaction
$W^\pm$ (W boson)	$\pm 1 e$	1	80.4 GeV	
g (gluon)	0	1	0	strong

All interactions conserve the following physical properties: energy (E), momentum (p), angular momentum (L), charge (Q), colour, baryon number (B) and the lepton number (L). Some of these are deductions from experimental observation or non-observation, for instance the conservation of baryon number is deduced from the proton being a stable particle. Furthermore there is a set of symmetries which require an interaction to be indistinguishable if the following action is performed:

- C-symmetry: Every particle in an interaction can be exchanged with its anti-particle.
- P-symmetry: Inversion of coordinate system.
- T-symmetry: Direction of time is reversed.

Electromagnetic and strong force conserve the above mentioned symmetries and the combinations of them, the weak force however violates the P and CP symmetry, but conserves the CPT symmetry, i.e. that an interaction via the weak force can favour a matter final state over an anti-matter final state. This might be a possible explanation for the vast matter anti-matter asymmetry in the universe, but the experimentally measured fraction with which CP violation occurs is too low. The

next big step in particle physics would be the unification of the electroweak and the strong force, commonly called the Grand Unified Theory (GUT), but the masses of novel particles predicted in some GUT models are far beyond the energies current particle colliders are capable to produce.

Electromagnetism

Electromagnetism is the oldest known force and, with the later addition of Quantum Electro Dynamics (QED), can be applied at the quantum and relativistic level. The most basic vertex in electromagnetism is shown in Fig. 1.1. This seemingly simple diagram can be rotated and combined with other vertices to produce a description of any electromagnetic interaction. A feynman diagram is actually a tool which allows physicists to formulate the complex matrix elements that describe the interaction of fundamental particles. These matrix elements have a fundamental strength associ-

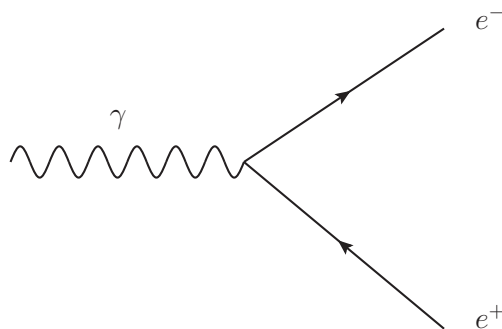


Figure 1.1: Feynman diagram of electron positron pair production.

ated with the interaction at each vertex and are composed of exchanges between a series of virtual particles which conserve energy/momentum at each of the vertices. In QED the spin of the particles is also accounted for and only charged particles are considered.

The coupling strength of one of these electromagnetic vertices α , also known as fine-structure constant, is one of the most precisely theoretically calculated and experimentally measured values in physics and is approximately $1/137$. The theory of QED agrees with the experimentally measured value with an uncertainty of less than 0.25 parts per billion.

Strong Force

The strong force is described in the theory of Quantum Chromo Dynamics (QCD) and couples to particles with colour charge. A quark can have one colour charge in addition to its electric charge and a gluon, the force carrier for QCD, has one colour and one anti-colour, see Fig. 1.2. Although they are also massless like photons, the strong force only interacts over a very short distance of approximately 10^{-15} m, but, as the name suggests, its coupling constant is very strong, around

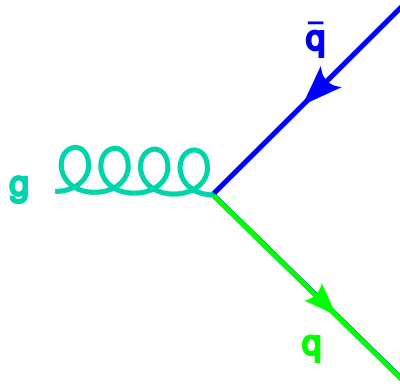


Figure 1.2: Primitive Feynman diagram of quark-antiquark to gluon vertex. Here a blue antiquark and a green quark produce a green-antiblu gluon, the strong interaction force carrier.

137 times stronger than the electromagnetic. It is strong enough to overcome the electromagnetic repulsion of quarks and making it possible to form hadrons and nuclei.

For example at high energies it is possible to scatter an electron off a quark inside a proton. At even higher energies the proton does not survive the interaction and instead new particles, other hadrons, are formed from the proton constituents, known as partons. There exist distributions representing the probability of finding a given flavor of parton after such an interaction. From this it is found that gluons and a ‘sea of quarks’ carry a significant amount of the protons momentum. Most of the physics done at the pp -colliders involves interactions understood with these basic principles. Parton interactions are one of the reasons that hadron accelerator interactions are so complicated to untangle compared to purely leptonic colliders. The QCD potential is also more complex than in QED. This leads to the interesting property of asymptotic freedom and confinement, which in turn leads to particle jets in high energy interactions.

Weak Force

The weak force can couple to every particle and is responsible for the decay of quarks to lighter quarks. For instance the radioactive β -decay, shown in Fig. 1.1, is possible with the transition from a d quark to an u quark and emitting a virtual W^- boson, which decays to an e^- and $\bar{\nu}_e$. The transition from one quark to another is only possible by emitting a $W^\pm$ boson, which is called a charged current. A transition from one quark to another via a Z boson, a neutral current, is suppressed in the SM and has not been observed yet. The range of the weak force is very short, due to the mass of the $W^\pm$ and Z boson, and the coupling constant α_w is approximately 10^{-6} , therefore weak interactions occur less often than electromagnetic or strong interaction. This is especially prominent for neutrinos, which can only interact via the weak interaction. As a result of this, one lightyear of lead is needed to stop half of the neutrinos traveling through it.

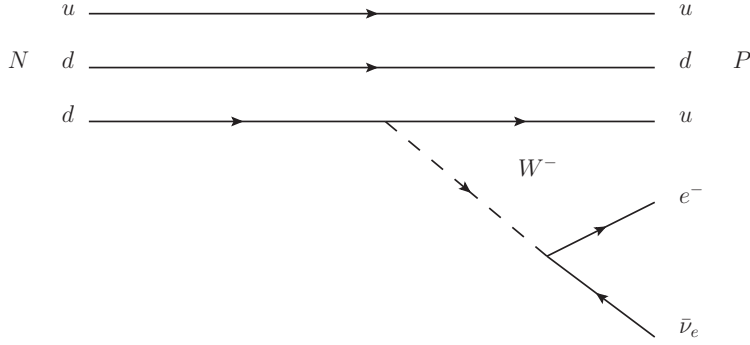


Figure 1.3: Feynman diagram of a β^- -decay.

1.1.3 Broute-Englert-Higgs Mechanism

The Broute-Englert-Higgs (BEH) mechanism is necessary to explain why the gauge bosons of the weak force are so massive. It introduces a mechanism called spontaneous symmetry breaking, which allows the electroweak force to be symmetric at high energies, but spontaneously, i.e. for no specific reason, break at lower energies. Thereby allowing the photon of the electromagnetic force to be massless and the $W^\pm$ and Z boson to be massive. This mechanism can be added to the SM by the addition of another field, called the Higgs field, and the quantum of this field is the Higgs boson. Many characteristics of the Higgs boson emerge from this theory except its mass, hence particle colliders have, since the postulation of the Higgs boson been searching for it.

At the Tevatron, a pp -collider with a centre of mass energy of $\sqrt{s} = 2$ TeV, no Higgs boson has been found in the mass range up to $m_H = 114.4$ GeV [14]. Therefore the hope to find the Higgs boson at the LHC, a pp -collider with a centre of mass energy of $\sqrt{s} = 14$ TeV, were big and in 2012 the discovery of a new boson consistent with the properties of a Higgs boson has been announced by two different experiments at the LHC [43][45]. Shown in Fig. 1.4 is the result of the Higgs boson search at the ATLAS experiment, showing a 6σ excess at a mass of 126 GeV. The highest signal strength has been observed in the two photon and four lepton decay channel, which is in accordance to the decay channels of a Higgs boson at this mass.

Since its discovery in 2012 many measurements have been performed to test if it is the Higgs boson predicted by the SM, because other theories also predict Higgs bosons but with slightly different properties. So far all measurements point to a Higgs boson compatible with the SM, but more data is needed to perform the measurements with higher precision to give a conclusive answer to this question.

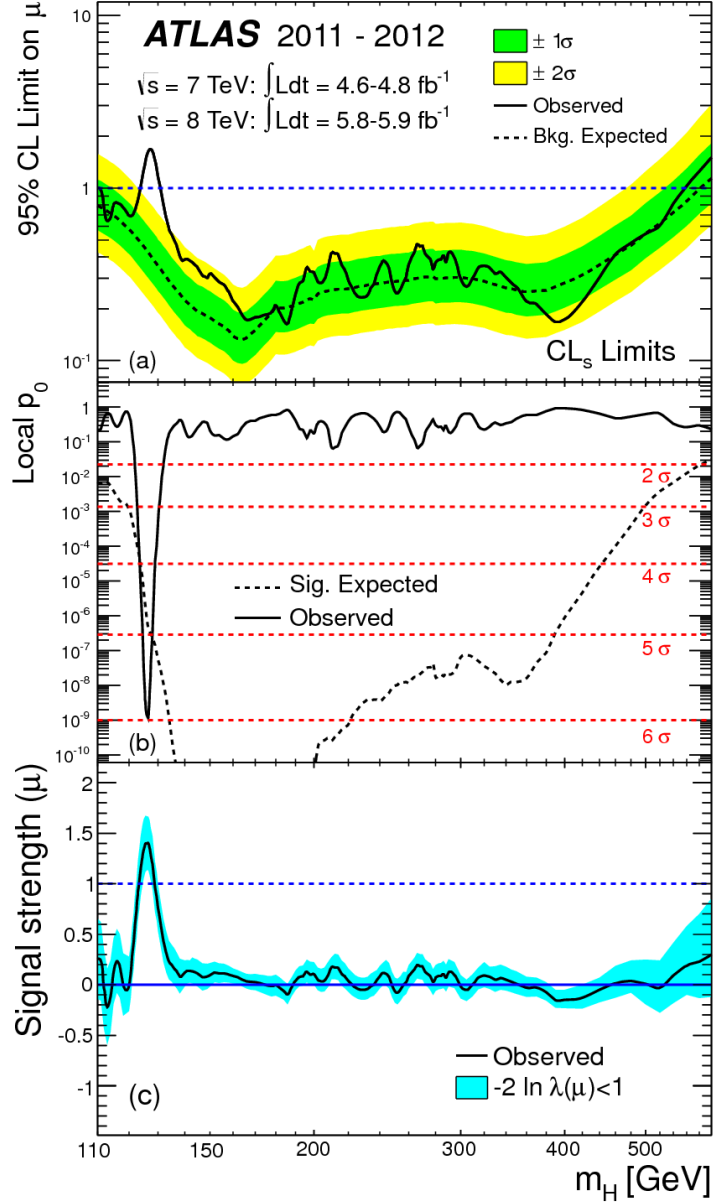


Figure 1.4: Combined search results of the ATLAS experiment: (a) The observed (solid) 95% CL upper limit on the signal strength as a function of m_H and the expectation (dashed) under the background-only hypothesis. The dark and light shaded bands show the plus/minus one sigma and plus/minus two sigma uncertainties on the background-only expectation. (b) The observed (solid) local p_0 as a function of m_H and the expectation (dashed) for a SM Higgs boson signal hypothesis ($\mu = 1$) at the given mass. (c) The best-fit signal strength $\hat{\mu}$ as a function of m_H . The band indicates the approximate 68% CL interval around the fitted value [43].

1.2 Physics Beyond the Standard Model

As discussed the experimental result of the SM is very convincing and the reason it is used in particle physics. But it lacks explanations for many phenomena observed in nature:

- Dark matter: from its gravitational influence it can be deduced that visible matter only makes up 5% of the energy in the universe, 27% is dark matter, and 68% dark energy. The SM does not predict any kind of particle or mechanism which could be responsible for this kind of phenomena, therefore there must be a new kind of group of particles, which only very weakly interacts with itself and ordinary matter. At the LHC, searches for exotic particles, many fitting the description of dark matter, are performed, but so far have not shown signs of any new particles.
- Neutrino masses and hierarchy: neutrinos in the SM are massless particles, though in order to make the observation of neutrino oscillation possible, neutrinos need to have mass. It was not yet possible to measure the mass or the mass hierarchy of neutrinos, but different experiments are performing either direct or indirect mass measurements to answer this question.
- Matter anti-matter asymmetry: the visible universe is made out of mostly matter, although the SM would predict there to be no large matter anti-matter asymmetry after the big-bang. The phenomena of CP violation might explain this asymmetry, but so far the experimentally measured fraction with which matter is produced more than anti-matter is too small. The LHCb experiment at the LHC is specialised for measurements on rare CP violating decays.

These points just represent a fraction of the open questions, which are not answered by the SM. Therefore it is known that the SM needs to be extended or even completely reformulated in some way. Many new theoretical models have been proposed, but only experimental evidence can give answers to which ones are the right direction. One of the most promising experiments to find new particles is the LHC at CERN, which will produce particle collisions with never reached energies and rates. A diverse physics program is performed and many theories for new physics are tested.

Chapter 2

The ATLAS Experiment

The ATLAS¹ experiment is one of the four large experiments located at the Large Hadron Collider (LHC) [21] at CERN, near Geneva in Switzerland. The LHC accelerates proton or ion beams to energies of the order of TeV and collides them head on. The collisions are recorded by the ATLAS experiment and their analysis can give rise to the discovery of new particles and interactions as well as deepen our understanding of known phenomena. The experiment is designed with the search for the Higgs boson in mind, but is not solely restricted to it. A wide variety of measurements and searches are performed, from precision measurements of SM parameters, to the TeV scale search for new particles beyond the SM. This is possible due to the key design aspects of the ATLAS experiment:

- High resolution of charged particle momentum
- Precise vertexing for jet flavour tagging
- Very good electron and photon identification
- Accurate jet and missing energy measurement
- Good muon identification and momentum resolution

The first proton physics run of the LHC was completed after 3 years, at the end of 2012. The first year was used to optimise the accelerator and first collisions were successfully produced, but the amount of integrated luminosity is not noteworthy. In 2011 and 2012 collisions were produced with $\sqrt{s} = 7$ TeV and 8 TeV, respectively. Although running at only 50% of the design beam energy, collisions were produced with a luminosity peaking at $7 \times 10^{33} \text{ cm}^{-2}\text{s}^{-1}$ delivering a total of 28 fb^{-1} . This excellent performance of the LHC and also of the ATLAS experiment then eventually led to the discovery of the Higgs boson in 2012.

¹Abbreviation for "A Toroidal LHC ApparatuS"

2.1 ATLAS Detector

The ATLAS detector [6] is shown in Fig. 2.1 and is divided into multiple sub detectors. Each sub detector is optimised for a specific function. The three innermost detectors - the Pixel detector, the Semiconductor Tracker (SCT), and the Transition Radiation Tracker (TRT) - are encased in a superconducting solenoid magnet and perform the tracking of charged particles for the ATLAS detector. The electromagnetic and hadronic calorimeter follow the tracking systems. Muons passing through the calorimeter system are detected in the muon chambers, which are located in a toroidal magnet. These magnets combined with the large muon chambers give the ATLAS Detector its characteristic look. The ATLAS experiment is equipped with a 3 level trigger system [5], which reduces the event rate from 40 MHz to 100 kHz on the first, 3.5 kHz on the second and 200 Hz on the third trigger stage.

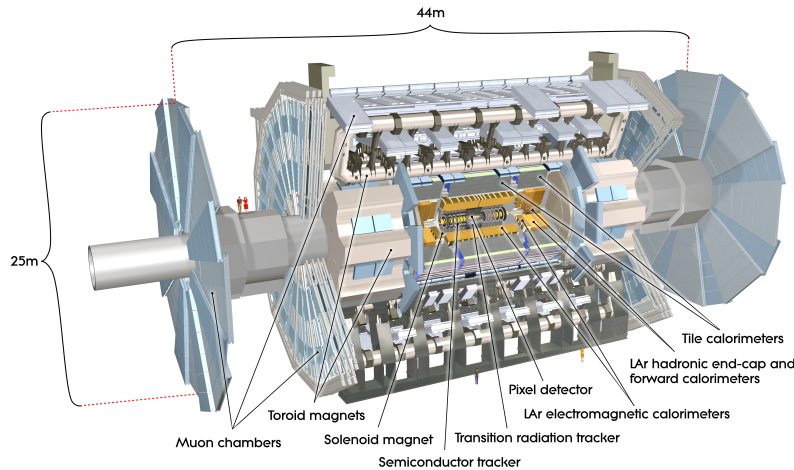


Figure 2.1: Computer generated drawing of the ATLAS detector layout [34].

The coordinate system the ATLAS experiment uses, is chosen to be R , ϕ , and η . In mechanical descriptions also z is used to describe the axis along the beam pipe. The origin is in the middle of the detector and R being the radius and ϕ the azimuthal angle around the z -axis. The pseudorapidity η is expressed by the polar angle θ :

$$\eta = -\ln \left(\tan \left(\frac{\theta}{2} \right) \right) \quad (2.1.1)$$

2.1.1 Inner Detector

The task of the Inner detector [39], shown in Fig. 2.2, is to track the trajectory of charged particles emerging from collisions in the centre of the detector. As the Inner detector is encased in a solenoid magnet, which produces a 2 T magnet field, charged particles curve on a helical trajectory as they move outwards from the interaction point. The radius of this curvature has to be measured precisely in order to determine the momentum of the particles. As every interaction of the particle

with the material of the detector can change its trajectory, all three detector systems are optimised to be as lightweight as possible. The active area, tracking resolution,

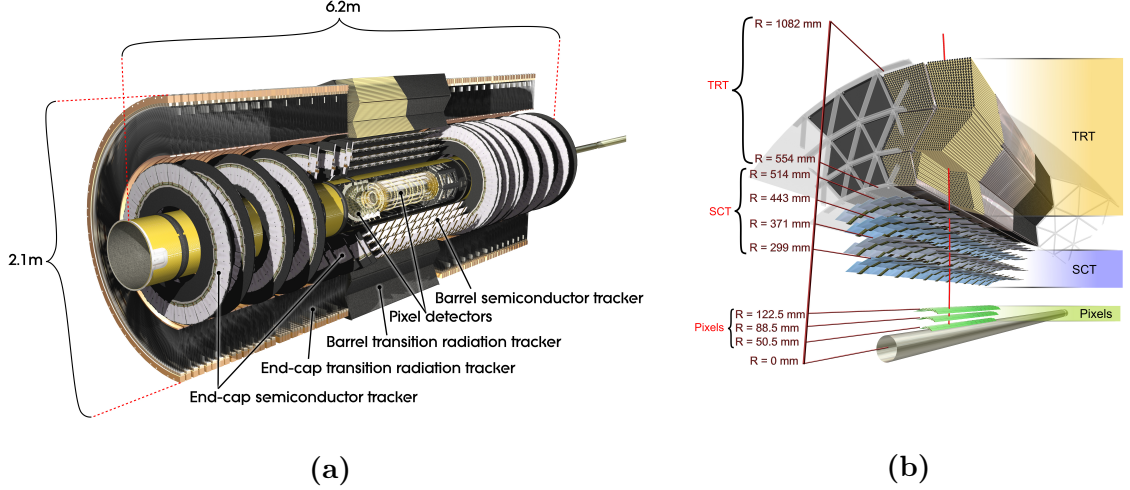


Figure 2.2: Computer generated cutaway and cross-section drawing of the Inner detector layout. [33]

number of channels, and η coverage are listed in Table 2.1.

Table 2.1: Parameters of the inner detector. [4]

System	Position	Area [m ²]	Resolution [μ m]	Channels [10 ⁶]	η coverage
Pixel	1 b layer	0.28	$R\phi = 15, z = 115$	13	± 2.5
	2 barrel layers	1.16	$R\phi = 15, z = 115$	54	± 1.7
	3 end cap disks	0.14	$\phi = 15, R = 115$	13	1.7 - 2.5
	on each side				
SCT	4 barrel layer	34.4	$R\phi = 16, z = 580$	3.2	± 1.4
	9 end cap disks	26.7	$\phi = 16, R = 580$	3.0	1.4 - 2.5
	on each side				
TRT	Axial barrel straws		170 (per straw)	0.1	± 0.7
	Radial end-cap straws		170 (per straw)	0.32	0.7 - 2.5
	36 straws per track				

Pixel Detector

The Pixel detector [24] shown in Fig. 2.3a consists of three barrel layers and three disks on each side. The 80 million pixels of the detector are located on 1744 Pixel modules, shown in Fig. 2.3b. Each module consists of a 2×6 cm² n⁺-in-n planar silicon sensor tile of 250 μ m thickness, bump bonded to 16 FE-I3 readout chips.

Each readout chip is connected to 2880 pixels of size² $400 \times 50 \mu\text{m}^2$ in $z \times R\phi$. A flexible PCB³ is glued on top of the sensor and readout chip assembly, allowing connection to the services for powering, control and, readout. The Module Control Chip (MCC) combines the 16 readout chips onto a single 40 Mb/s control link and up to two 40/80 Mb/s data links.

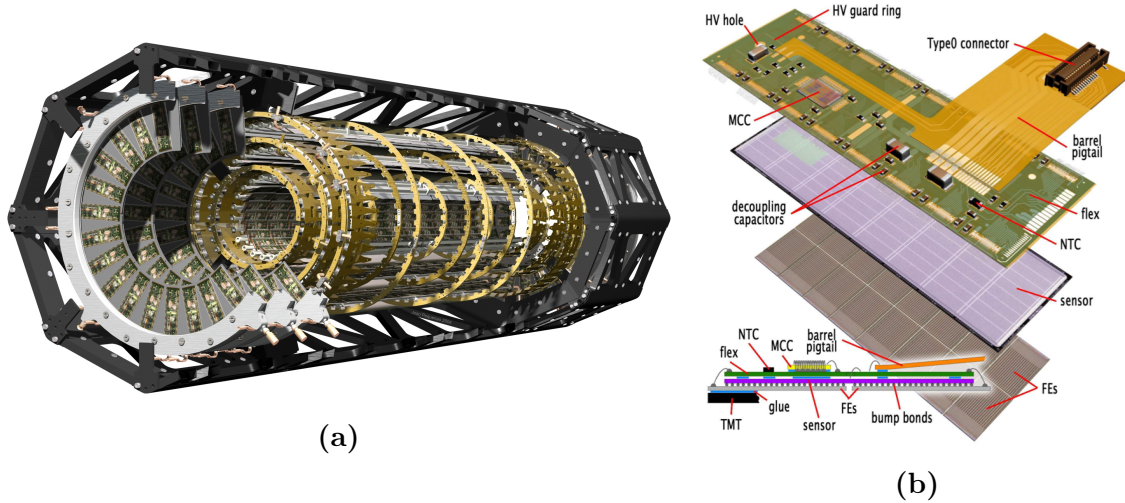


Figure 2.3: Computer generated cutaway drawing of the Pixel detector layout [35] and exploded drawing of a Pixel module [24].

Being the closest detector to the interaction point, the Pixel detector is exposed to a high dose of radiation, requiring it to be built out of a radiation tolerant technology. The sensors are radiation tolerant up to $1 \times 10^{15} \text{ 1 MeV n}_{\text{eq}} \text{ cm}^{-2}$ and the FE-I3 to 50 Mrad ionising dose, which is equivalent to the run time with a luminosity of $1 \times 10^{34} \text{ cm}^2 \text{ s}^{-1}$ of five years for the inner detector layer and the lifetime dose of the outer layers. The Pixel detector is optimised for vertexing with a resolution of the order of $10 \mu\text{m}$ and will typically deliver 3 space-points per particle.

Semiconductor Tracker

The SCT [26] has 4 barrel layers and 9 disks on each side, containing 6 million strips on 4088 modules. These sensors are produced in a single sided p^+ -in-n process on $285 \mu\text{m}^2$ thick silicon, which makes them very cost effective to cover the active area of 60 m^2 . Each SCT module has two layers of sensors which are interlaced by 40 milliradians, which increases the resolution in the plane along the strips. A total of 12 readout chips are connected with each strip sensor via wire bonds, which each chip being connected to 128 channels. Each channel has a pre-amplifier, shaper and, tuneable discriminator circuit to read out binary hit information. The hit

²In addition, there are two other sizes of pixels located on the edges of the sensor tile and in region between readout chips.

³Printed Circuit Board

information from the two sensors of one module gives one space-point per track and the SCT will typically deliver 4 space-points in the tracking region of $|\eta| < 2.5$.

Transition Radiation Tracker

The TRT [19] uses a different detection technique to the Pixel detector and the SCT. It consists of 298,304 drift tubes (straws) with a diameter of 4 mm and a length of 144 cm. These straws are distributed over a barrel and end-cap wheel section covering the pseudo rapidity up to $|\eta| < 2.5$ so that a particle typically hits 36 straws. Each straw has a charge collecting gold-plated tungsten wire anode at a bias of 1530 V and is filled with a gas mixture of 70% Xe, 27% CO₂ and 3% O₂. In between the straws are fibres (barrel) or foil (end-cap) of polypropylene. Particles traversing the straw-polypropylene boundary produce "transition" radiation, giving the detector its name. The energy of the transition radiation is directly proportional to the Lorentz factor $\gamma = \frac{E}{mc^2}$, which is used by the TRT for electron identification. Ionising particles deposit energy of the order of eV into a single straw, whilst electrons, from LHC collisions, also produce transition radiation of the order of multiple keV, which can be used to discriminate between electrons and other charged particles. The drift-time of electrons to the anode can be used to increase the precision of each straw by providing the radial distance from the anode with a resolution of up to 170 μm .

2.1.2 Calorimeter System

The ATLAS calorimeter system uses two types of calorimeters: the liquid argon electromagnetic calorimeter and tile hadronic calorimeter. As shown in Fig. 2.4 the liquid argon calorimeter is the inner cylinder-shaped system and the tile calorimeter the outer. The calorimeter system is an integral part of the ATLAS trigger system, providing a specialised fast energy reconstruction, over a defined set of cells called trigger towers, to enable the ATLAS experiment to produce a first level trigger depending on the observed energy of electrons, photons, jets and missing energy.

Liquid Argon Calorimeter

The liquid argon calorimeter [41] (LAr) is a sampling calorimeter and composed of accordion-shaped electrodes and lead absorbers in liquid argon. The barrel region (EMB) covers a pseudorapidity of $|\eta| < 1.475$ in R from 1500 mm to 1970 mm and the end-caps (EMEC) reach $1.375 < |\eta| < 3.2$. The liquid argon hadronic calorimeter (HEC) at $1.5 < |\eta| < 3.2$ uses copper as the absorber material and the liquid argon forward calorimeter (FCal) at $3.1 < |\eta| < 4.9$ uses copper/tungsten. In total there are 182,468 readout cells distributed over the entire liquid argon calorimeter. The trigger towers of the liquid argon calorimeter cover an area of $\Delta\eta \times \Delta\phi = 0.1 \times 0.1$ for $|\eta| < 2.5$ and $\Delta\eta \times \Delta\phi = 0.4 \times 0.4$ for $3.1 < |\eta| < 4.9$. The analog signal of each tower is sent to a first level trigger processor to make the

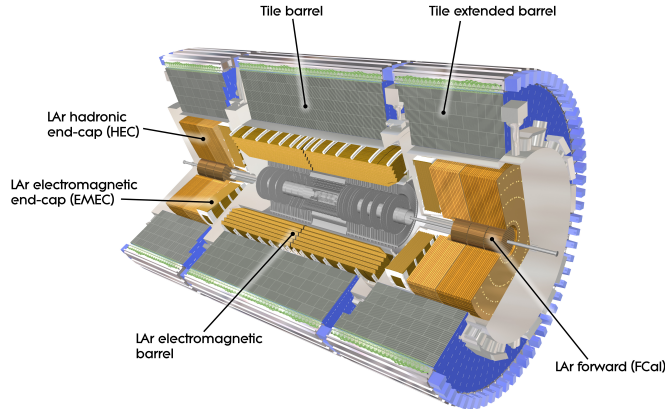


Figure 2.4: Computer generated drawing of the ATLAS detector layout. [32]

trigger decision. This trigger can then be used to issue the readout of the analog signal on a per cell level, which operates slower but with much higher precision.

Tile Calorimeter

The tile calorimeter [42] (Tile) is composed of a barrel region covering $|\eta| < 1.0$ and an extended barrel region covering $0.8 < |\eta| < 1.7$. The barrel and extended barrel are divided in 64 wedges in ϕ and three layers in R from 2280 mm to 3865 mm, furthermore the wedges are organised in towers each covering $\Delta\eta \times \Delta\phi = 0.8 \times 0.1$ and $\Delta\eta \times \Delta\phi = 0.7 \times 0.1$ in the central barrel and extended barrel respectively. Each wedge consists of a sandwich of scintillator material and steel absorber, the scintillator are trapezium, oriented radially and normal to the beam line, and read out with a fibre connected to the scintillator on the legs of the trapezium. In total the tile calorimeter has 9836 readout channels and 2080 trigger outputs.

2.1.3 Muon Chambers

The muon spectrometer [40] of the ATLAS experiment consists of four different systems with their parameters summarised in Tab. 2.2: Monitor Drift Tubes (MDT), Cathode Strip Chambers (CSC), Resistive Plate Chambers (RPC) and Thin Gap Chambers (TGC). The MDT system has three double barrel layers on a radius R of around 4 m, 7 m and 10 m covering $|\eta| < 1.1$, and three end-cap wheels on both sides at z of approximately 7.5 m, 13.5 m and 22 m covering $|\eta| < 2.7$. It is used for precision measurement of the muon momentum in conjunction with the magnet field from the superconducting toroidal magnet and achieves a resolution of $50 \mu\text{m}$ while covering an active area of 5500 m^2 . The MDT system is divided in 1088 chambers each having three or four layers of tubes with a diameter of 30 mm and an anode wire in the middle. The high resolution of the MDT system is achieved by three factors: the position of each MDT chamber is measured precisely with a laser positioning system, the distance of the particle trajectory to the anode can be determined by

the drift-time relation and the position along the tube is determined by the time difference of each signal arriving at the end of the tube. In the high- η forward region ($|\eta| > 2$) a different kind of system is used to perform precision measurement, the CSCs, which work like multiwire proportional chambers and different layers with orthogonal wires help resolving hit ambiguity. The muon trigger system is made out of the RPC and TGC systems, which, despite not being as precise as the MDT and CSC, specialise in the fast identification of high p_T muons. The muon trigger system is of high importance to the ATLAS experiment as it delivers the trigger for most decay channels with the lowest background.

Table 2.2: Parameters of the four different system of the muon spectrometer [6].

Type	Active Area [m ²]	Resolution		Space-points	Channels
		z/R	ϕ		
MDT	5500	35 μm (z)	-	20	339k
CSC	27	40 μm (R)	5 mm	4	30.7k
RPC	3650	10 mm (z)	10 mm	6	359k
TGC	2900	2-6 mm (R)	3-7 mm	9	318k

2.2 The ATLAS Detector Performance during Run 1 and Long Shutdown 1

The LHC has operated from 2010 to the end of 2012, this period is called Run 1, with increasing beam energy and luminosity. A total of 28.3 fb^{-1} have been delivered, as shown in Fig. 2.5a, and the ATLAS detector recorded 93% of these collisions. The peak luminosity, shown in Fig. 2.5b, was up to $7 \times 10^{33} \text{ cm}^2\text{s}^{-1}$, 70% of the nominal instantaneous luminosity at only 57% of the nominal beam energy. An important figure of merit to evaluate the performance of the detector is the number of collisions per bunch crossing, shown in Fig. 2.6. The higher the number of interactions in a single bunch crossing, the higher the density of particles coming from the interaction points. The number of interactions in one bunch crossing at nominal luminosity of $1 \times 10^{34} \text{ cm}^2\text{s}^{-1}$ and nominal beam energy with $\sqrt{s} = 14 \text{ TeV}$ was estimated to be 23. The particle density during Run 1 measured to be up to 40 interactions per bunch crossing, which is already very close to the one for the ATLAS detector is designed for. The performance of the ATLAS detector in these conditions has been excellent and the relative fraction of good quality data is listed in Tab. 2.3.

Especially for the tracking system, the number of interactions, and thereby the number of vertices which need to be reconstructed, influences the reconstruction efficiency. Because of this, in particular parts of the detector were upgraded in the shutdown after Run 1 (“phase” upgrade):

- nSQPs: The Pixel detector services have been exchanged to move the optical drivers into an accessible region of the experiment. This also allowed to deploy

2.2. The ATLAS Detector Performance during Run 1 and Long Shutdown 1

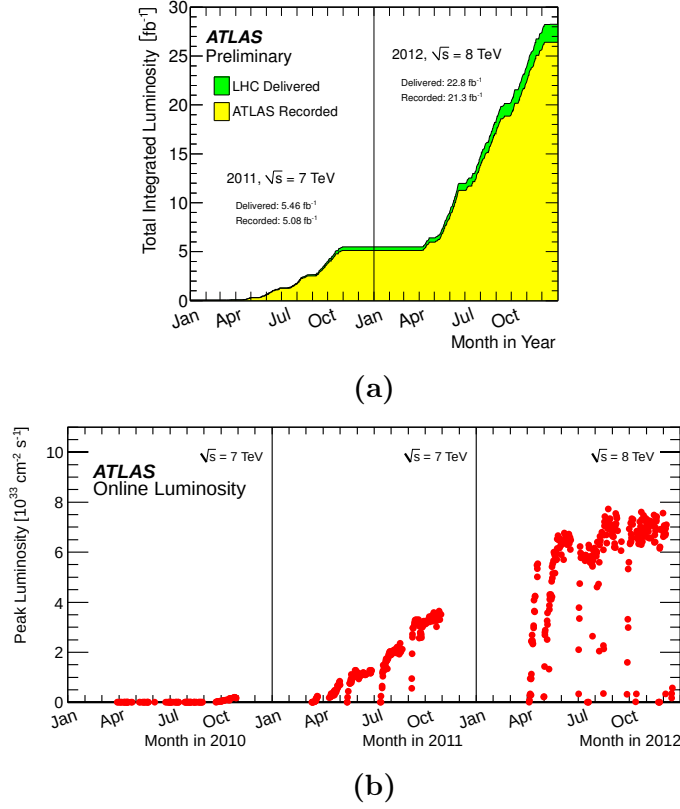


Figure 2.5: Total integrated luminosity and peak luminosity during Run 1. [9]

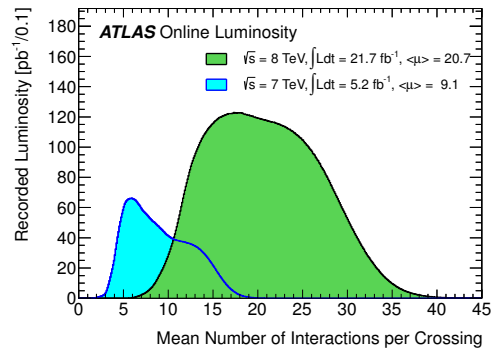


Figure 2.6: The mean number of interactions per bunch crossing during Run 1. [9]

2.3. Towards the High Luminosity LHC

Table 2.3: Luminosity weighted relative fraction of good quality data delivery in percent by the various ATLAS subsystems during LHC fills with stable beams in pp collisions [38]

Subsystem		2011 (5.2 fb ⁻¹)	2012 (21.3 fb ⁻¹)
Inner Tracker	Pixel	99.8	99.9
	SCT	99.6	99.1
	TRT	99.2	99.8
Calorimeters	LAr	96.9	99.1
	Tile	99.2	99.6
Muon Spectrometer	MDT	99.4	99.6
	RPC	98.8	99.8
	CSC	99.4	100
	TGC	99.1	99.6
Magnets	Solenoid	99.8	99.8
	Toroid	99.3	99.5
All good for physics		89.9	95.5

double channel readout of the Pixel modules in the second layer instead of a single readout channel, which will prevent expected readout inefficiencies.

- **Insertable B-Layer:** An additional layer of pixelised silicon detector installed inside of the Pixel detector, improving and securing the tracking quality of the Pixel detector during future operation.
- **Fast Track Trigger:** The data from the ATLAS tracking system is fed into new processing hardware, performing pattern matching algorithm to find high p_T tracks in preselected region of interests and provide tracking information for the L2 trigger.
- **Liquid argon trigger system:** A readout upgrade of the trigger system transforms trigger towers into higher granularity super cells, which provide the first level trigger system with higher precision energy reconstruction.
- **Muon spectrometer:** The CSC system of the muon small wheel upgrades its readout system to cope with higher particle occupancy during Run 2, until it gets changed for new small wheel during the second shutdown. Additional muon chambers have been installed in the detector to fill gaps in key positions in between the wheels and the barrel.

2.3 Towards the High Luminosity LHC

The current schedule of the LHC is shown in Fig. 2.7 and foresees two more runs before the upgrade to the High-Luminosity LHC (HL-LHC). In Run 2 from the

beginning of 2015 to mid 2018 the LHC will aim to run at nominal luminosity of $1 \times 10^{34} \text{ cm}^2\text{s}^{-1}$ and provide around 150 fb^{-1} of collisions. After a second long shutdown (LS2) of 1.5 years, Run 3 will try to reach two times the nominal luminosity and provide additional 300 fb^{-1} of integrated luminosity. The third long shutdown (LS3) will start in the beginning of 2023 for period of 2.5 years and after the upgrade to the HL-LHC 5 to 7 times the nominal luminosity is expected, which is needed to provide integrated luminosity up to 3000 fb^{-1} . Many of the sub detector systems will undergo extensive upgrades during LS2 [7] and during LS3 many systems will need to be exchanged. For instance the tracking system - the Pixel detector, SCT and TRT - will be exchanged with the Inner Tracker (ITK), a sub detector under development, which will consist of cutting edge pixel and strip tracking technology.

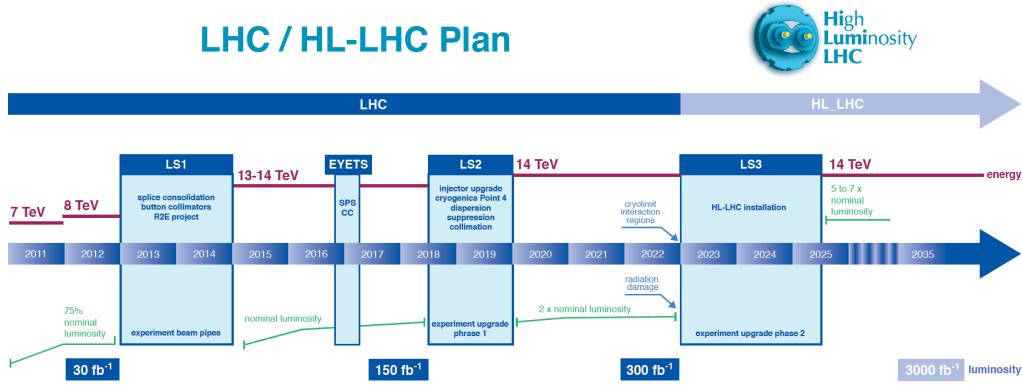


Figure 2.7: Prospective timeline of the LHC towards HL-LHC with the estimated integrated luminosity and beam energy.[15]

Chapter 3

The Insertable B -Layer

As part of the first upgrade of the Pixel detector an additional layer has been inserted into the current detector, the Insertable B -Layer (IBL). This new innermost layer benefits from state-of-the-art sensor and readout chip technology and makes it possible to successfully operate a detector very close to the interaction point, within high particle density environment, requiring both precise tracking and radiation hard electronics.

3.1 Overview

The IBL [13] is single barrel layer with a radius of 33 mm and has a sensitive region up to $|\eta| < 3.0$. There are 14 mechanical support structures made out of carbon foam, called staves, which form the IBL and supply cooling through an integrated titanium pipe, whilst also bearing the sensor modules. They are mounted on the Inner Protection Tube (IPT) which encloses the new beam pipe of radius of 24.3 mm. Each staff holds 12 planar silicon double chip modules and 8 3D silicon single chip

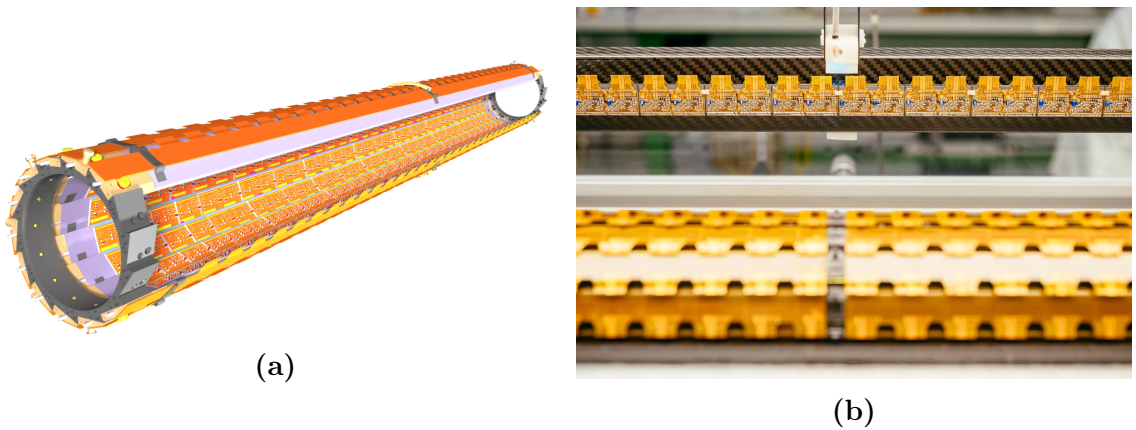


Figure 3.1: (a) Computer generated drawing of the IBL and (b) a picture of the last staff being integrated onto the IPT, also visible are the IBL modules facing towards the viewer.

modules, with the single chip modules located either end of the stave. As the name suggests a double chip module is read out by two readout chips and a single chip module by a single readout chip. The readout chip used for the IBL is the FE-I4¹, the successor of the FE-I3 which is used in the Pixel detector, and is capable of reading out 26880 pixels per chip. A total of 448 FE-I4s are used for the IBL, adding 12 million additional pixels to the ATLAS tracking system.

The IBL was inserted into the Pixel detector in April 2014 which can be seen in Fig. 3.2. Since then the detector has been commissioned using cosmic ray particles to integrate IBL into the existing ATLAS detector data acquisition and detector control system, as well as measuring the alignment with respect to the other sub detectors. By the end of November 2014, during Milestone Run 7, multiple IBL staves were operating with the ATLAS detector and cosmic particles passed through the inner detector, shown in Fig. 3.3, leaving hits in all layers of the tracking system including IBL.

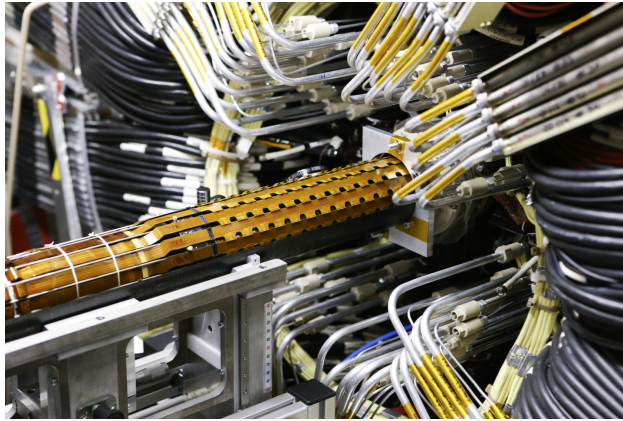
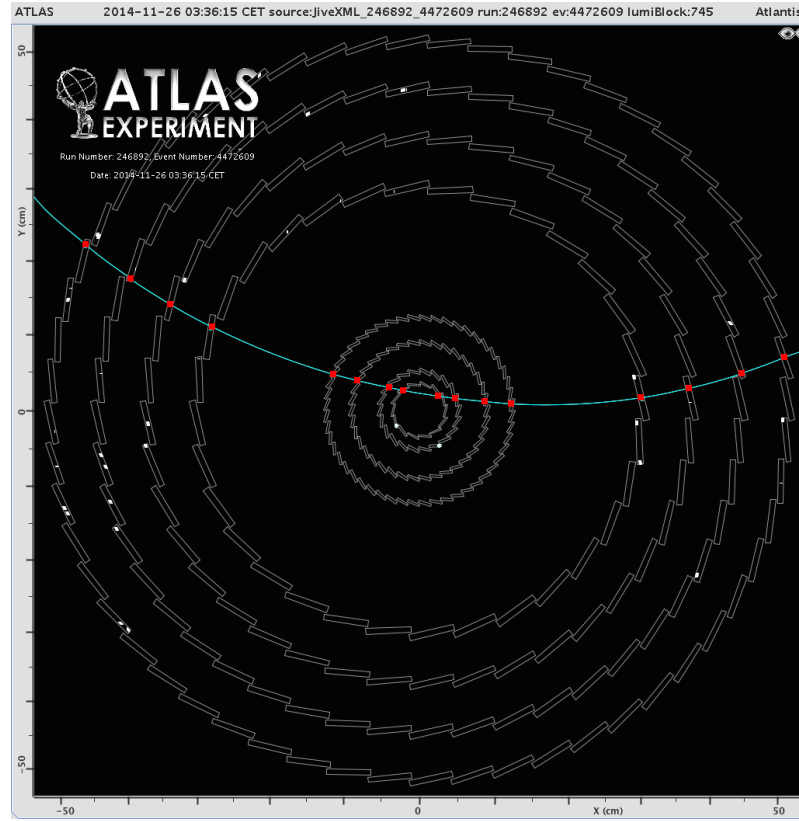


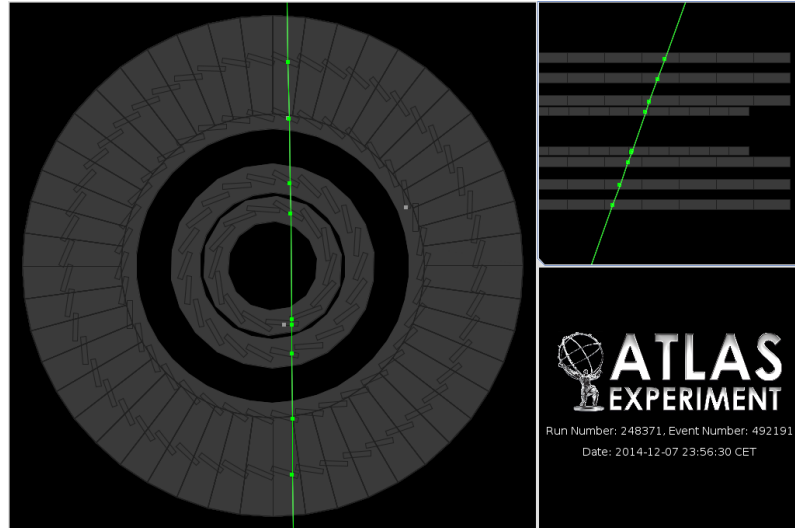
Figure 3.2: Picture of the insertion of the IBL into the Pixel detector from the C-Side of ATLAS in April 2014. [29]

¹Two types of FE-I4 exist, FE-I4A and FE-I4B, the former was only used during the prototyping phase and the latter one is used for the IBL.

3.1. Overview



(a)



(b)

Figure 3.3: The Atlantis event displays showing a track of a cosmic particle passing through the inner detector, including two hits in the IBL, (a) with and (b) without B-field [8].

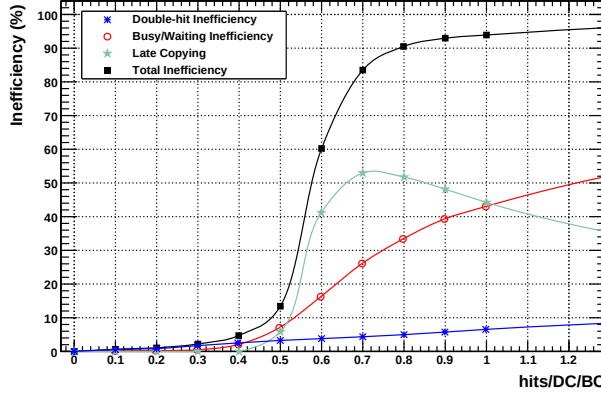


Figure 3.4: Simulation of FE-I3 inefficiencies, the occupancy for nominal luminosity is 0.16 hits per double column per bunch crossing [13].

3.2 Motivation

The original Pixel detector is designed to operate at a luminosity of $1 \times 10^{34} \text{ cm}^2\text{s}^{-1}$ and around 25 interactions per bunch crossing. Until the Phase 2 upgrade of the tracker is performed in 2022, inefficiencies in the tracking will arise from the radiation damage induced in the Pixel detector B -Layer and the higher number of interaction per bunch crossings will start to saturate the detector electronics. As the luminosity of the LHC may increase to more than twice the nominal value and the number of interactions per bunch crossing, called pile-up, having already reached a maximum of 40 in Run 1 an upgrade of the Pixel detector is needed. This upgrade is the IBL and it will facilitate either the current tracking performance or even increase it. Even if the Pixel detector B -Layer does not suffer from inefficiencies, the addition of a fourth Pixel detector layer closer to the interaction point, will, for the most part, increase the tracking performance and precision.

The FE-I3 readout chip which is used in the Pixel detector stores up to 64 hits in buffers shared by the 360 pixels of a double column. This architecture can produce inefficiencies under certain circumstances, as shown in the simulation in Fig. 3.4. The inefficiency is below 1.7% for nominal luminosity (0.16 hits/DC/bc), 3% for twice the nominal luminosity (0.36 hits/DC/bc) and 9% for three times the nominal luminosity. The values strongly depend on the beam structure and intensity, but the 50 ns bunch spacing and higher proton count per bunch produces higher pile-up and hence a higher occupancy per bunch crossing.

The smaller pixel pitch of $250 \text{ }\mu\text{m}$ in the z direction increases vertex resolution in z and is advantageous to differentiate between interactions in high pile-up collisions. The resolution of the transverse impact parameter d_0 is a direct reflection of the precision of the tracking system. In Fig. 3.5a the resolution is shown for different single muon momenta p_T with and without the addition of the IBL. The higher precision with IBL is due its smaller radius and a smaller impact parameter resolution does directly influence the b-tagging performance and, as can be seen in Fig. 3.5b,

3.2. Motivation

the light jet rejection is overall better with the IBL. At the common working point of 60 % b-tagging efficiency, the light jet rejection is better by a factor of 1.9. These are results from simulations without pile-up; if pile-up is taken into account the overall vertex reconstruction efficiency still drops, but with the addition of IBL it is kept above 99 % (98 % for tight cuts) for up to 50 interactions per bunch crossing. However, it is to be expected that the number of interactions per bunch crossing might increase up to 75 before the Phase 2 upgrade, in this case the reconstruction gains a major advantage with the IBL.

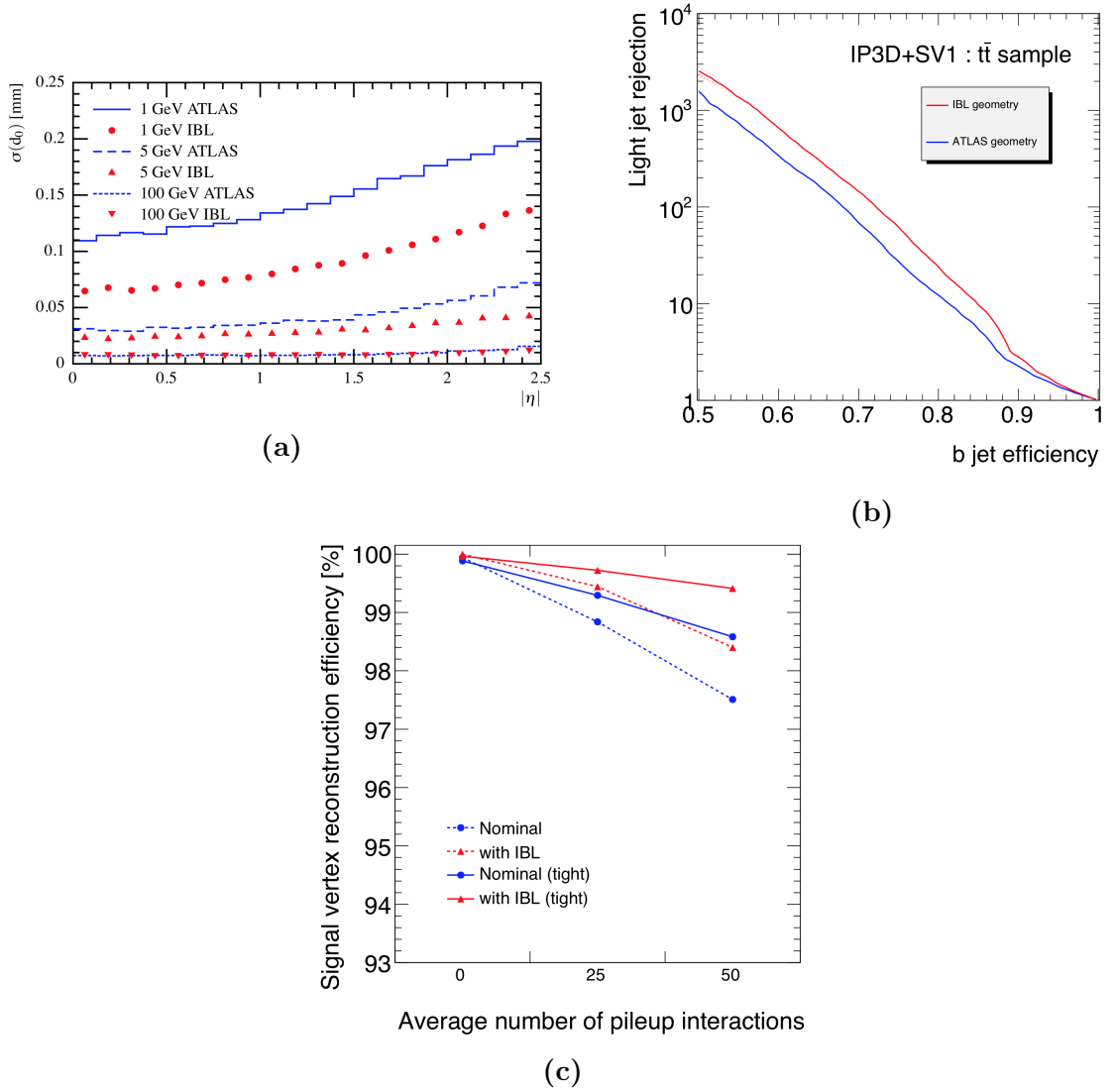


Figure 3.5: (a) Transverse impact parameter d_0 resolution for single muons at different p_T as a function of $|\eta|$ with and without the IBL. (b) Light jet rejection factor as a function of b-tagging efficiency with and without the IBL. (c) Primary vertex reconstruction efficiency in $t\bar{t}$ events as a function of average number of pile-up [13].

3.3 Sensor Technologies

Two different sensor technologies are used in the IBL [10], an improved design of the planar pixel sensor technology and new 3D sensor technology. The planar pixel sensor design is based on the planar pixel sensor currently used in the Pixel detector but with specific attributes altered for the IBL use-case. The 3D sensor technology is used for the first time in a high energy physics detector and, due to its different design concept, has advantages in the high- η region and for shallow track incidence angles. An important design consideration for the IBL sensors was to minimise the inactive region in z . This is because, unlike in Pixel, there is not enough space to overlap modules.

3.3.1 Planar Pixel Sensors

The planar pixel sensors used for double chip modules, have an area of $18.3 \times 41.3 \text{ mm}^2$ in $R\phi \times z$ and are $200 \text{ }\mu\text{m}$ thick. The sensor has 2×26880 pixelated n^+ implants in an n -bulk, with three different pixel sizes listed in Tab. 3.1. The inter-chip pixels are located in the two columns between each readout chip and the edge pixels are located in the two outer columns, whilst normal pixels make up the rest of the matrix. Not only is the pixel size smaller compared to the planar pixel sensor used in the Pixel detector but the inactive edge region has also been reduced to $200 \text{ }\mu\text{m}$ from $1,100 \text{ }\mu\text{m}$. The edge design is shown in Fig. 3.6 where it can be seen that the slim edge is made possible by extending the edge pixels underneath the guard rings. All sensors needed for IBL are manufactured by CiS².

Table 3.1: The three different pixel sizes in an IBL planar pixel sensor.

Pixel type	Size in $R\phi \times z$	Number of pixels
Normal pixels	$50 \times 250 \text{ }\mu\text{m}^2$	52416
Long inter-chip pixels	$50 \times 450 \text{ }\mu\text{m}^2$	672
Long edge pixels	$50 \times 500 \text{ }\mu\text{m}^2$	672

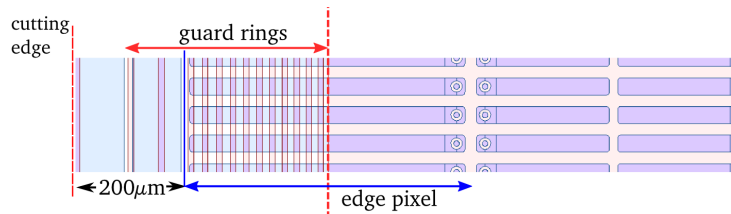


Figure 3.6: Illustration of the edge region of a planar pixel sensor showing the long pixel reaching underneath the guard rings. [10]

²CiS Forschungsinstitut für Mikrosensorik und Photovoltaik GmbH, Konrad-Zuse-Strasse 14, 99099 Erfurt, Germany, <http://www.cismst.org>.

3.3.2 3D sensors

In the 3D sensor design pillars of n^+ and p^+ material are etched through the p-bulk and so the full thickness of the sensor need not be depleted, but rather only the distance in between the pillars. In this case the electrons and holes are drifting pillar-to-pillar, parallel to the surface of the sensor and not perpendicular, i.e. across the bulk, as in planar designs. A sketch of this concept is shown in Fig. 3.7. Although the two manufacturers, FBK³ and CNM⁴ of the 3D sensors have slightly different implementations of the design, the broader operating concept is the same. The sensor measures $18.8 \times 20.7 \text{ mm}^2$ and has 26880 pixels of two different dimensions listed in Tab. 3.2. It is $230 \text{ }\mu\text{m}$ thick and the inactive edges are $230 \text{ }\mu\text{m}$ slim.

Table 3.2: The two different pixel sizes in an IBL 3D pixel sensor.

Pixel type	Size in $R\phi \times z$	Number of pixels
Normal pixels	$50 \times 250 \text{ }\mu\text{m}^2$	26208
Long edge pixels	$50 \times 500 \text{ }\mu\text{m}^2$	672

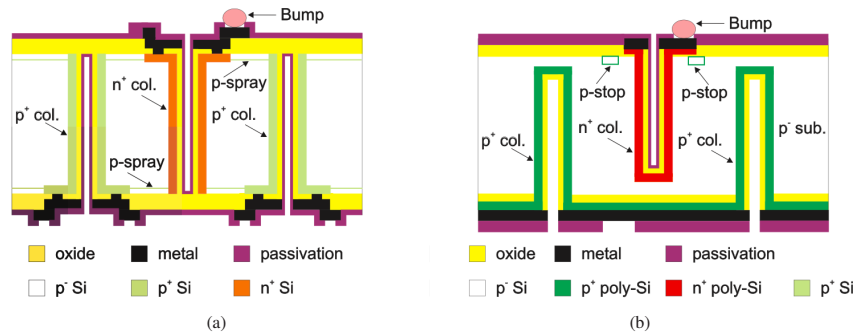


Figure 3.7: Crosssection of a 3D sensor manufactured by (a) FBK and (b) CNM. [10]

3.4 Front-End Technology

The FE-I4 readout chip [52] is built in a 130 nm CMOS process and connects to 26880 pixels, so a double chip module has two FE-I4 and a single chip module has one FE-I4. Its functionality is similar to its predecessor the FE-I3, which is used in the Pixel detector, but has been improved for operation in an environment with high particle density such as the IBL will experience.

The main parameters of the FE-I4 are summarised in Tab. 3.3 and a block diagram of the chip is shown in Fig. 3.8. The pixels are arranged in a grid with 80 columns at $250 \text{ }\mu\text{m}$ pitch and 336 rows at $50 \text{ }\mu\text{m}$ pitch. Each pixel is connected

³Fondazione Bruno Kessler (FBK), Via Sommarive 18, 38123 Povo di Trento, Italy, <http://www.fbk.eu>.

⁴Centro Nacional de Microelectronica (CNM-IMB-CSIC), Campus Universidad Autonoma de Barcelona, 08193 Bellaterra (Barcelona), Spain, <http://www.imbcnm.csic.es>.

to its own shaper and discriminator circuit, as shown in Fig. 3.9a, which have fine-tuning capabilities. The discriminator output is fed into a digital region counting the Time over Threshold (ToT) in units of 25 ns (bunch-crossings/bc), this region is shared by 2×2 pixels and can store up to 5 events each containing up to 4 ToT values. There are 118, 2×2 pixel regions connecting two columns into a digital double column, each of the 40 digital double columns connects to the end of digital column logic, which distributes digital and analog signals from and to the double columns. An important feature of the FE-I4 is that each 2×2 pixel region has its own buffers, in comparison to the FE-I3 where the buffers are located in the end of column logic. This was a necessary improvement to deal with the high hit occupancies which are expected for the IBL. The first bottleneck the FE-I4 has is the data output bandwidth of 160 Mb/s, which limits the trigger rate to 200 kHz for a hit occupancy of 10^{-3} bc^{-1} per pixel.

Table 3.3: Main parameters of the FE-I4 readout chip [22].

Parameter	Value
Pixel size ($R\phi \times z$)	$250 \times 50 \mu\text{m}^2$
Pixel array size	80×336 (Col $\times$ Row)
Pixel input	DC-coupled
Radiation tolerance	300 Mrad
Operation temperature	-40 to +60 °C
Readout initiation	Trigger
Max. trigger latency	6.4 μs
Hit-time resolution	25 ns
ADC method	Time over Threshold (ToT)
ADC resolution	4 bit
Maximum sustained trigger rate	200 kHz
External clock input	40 MHz
Serial command input	40 Mb/s
Command encoding	NRZ
Serial data output	160 Mb/s
Data encoding	8b10b [2]
Nominal regulator input voltage	1.8 V
Maximum regulator input voltage	2.5 V
Analog voltage	1.5 V
Digital voltage	1.2 V
Current at operation	$\approx 0.55 \text{ A}$

To equalise the charge digitisation over the whole detector, each FE-I4 has the capability to adjust the preamplifier feedback current and threshold on a global chip-level and a fine pixel-level. The effect of threshold and feedback current on the charge-to-ToT conversion can be seen in Fig. 3.9b. The global threshold can be adjusted with a 16 bit DAC⁵ and fine tuned with a 5 bit DAC per pixel, reaching

⁵Two 8 bit DACs, one fine and one coarse with slight overlap.

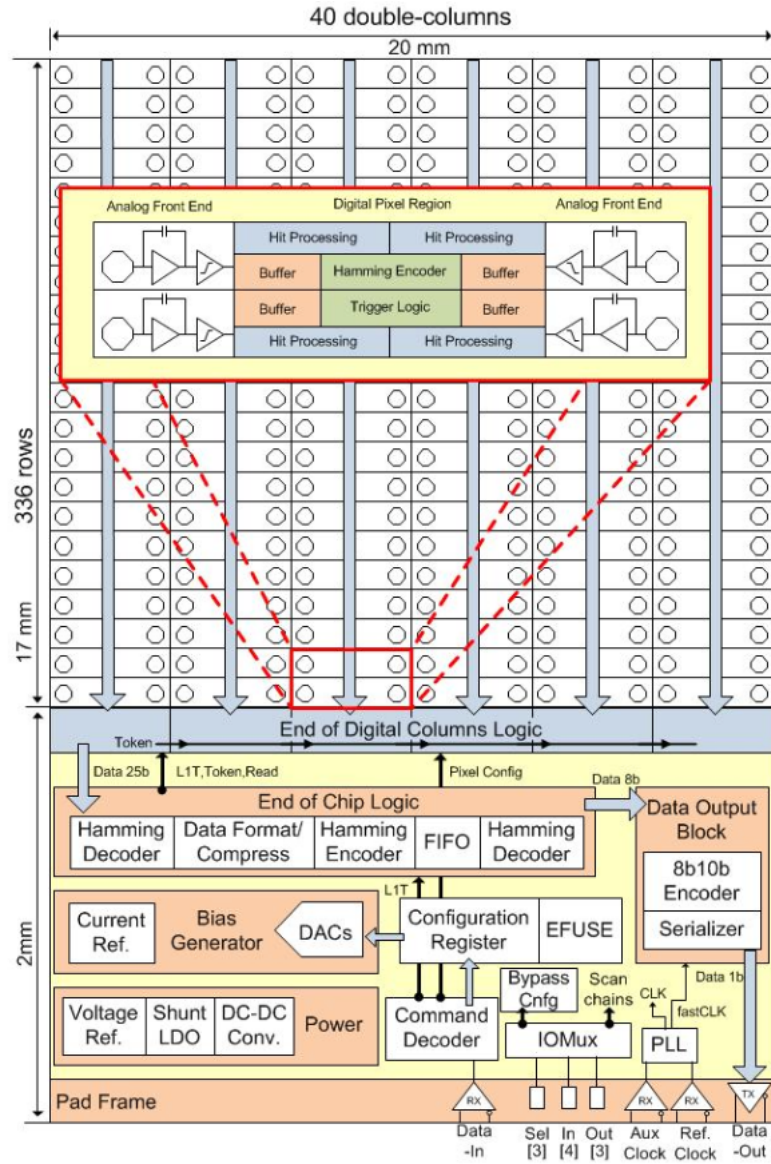


Figure 3.8: Block diagram of an FE-I4 showing the different functional blocks and the 2×2 pixel region which contains a buffer shared by all pixels in the region [52].

a dispersion of less than $100 e$ over all pixels for a threshold which can be as low as $1000 e^6$. The preamplifier feedback current is adjusted by an 8 bit global DAC and fine tuned with a 4 bit pixel DAC, rendering it possible to reach a dispersion of less than $0.5 bc$ for a given charge to ToT conversion. During calibration of threshold and charge conversion it must be considered that the threshold directly influences the charge to ToT conversion and also that the feedback current influences the threshold. Hence the tuning of one DAC is performed at least one more time following the tuning of the other DAC, first on a global level and then on the per pixel level.

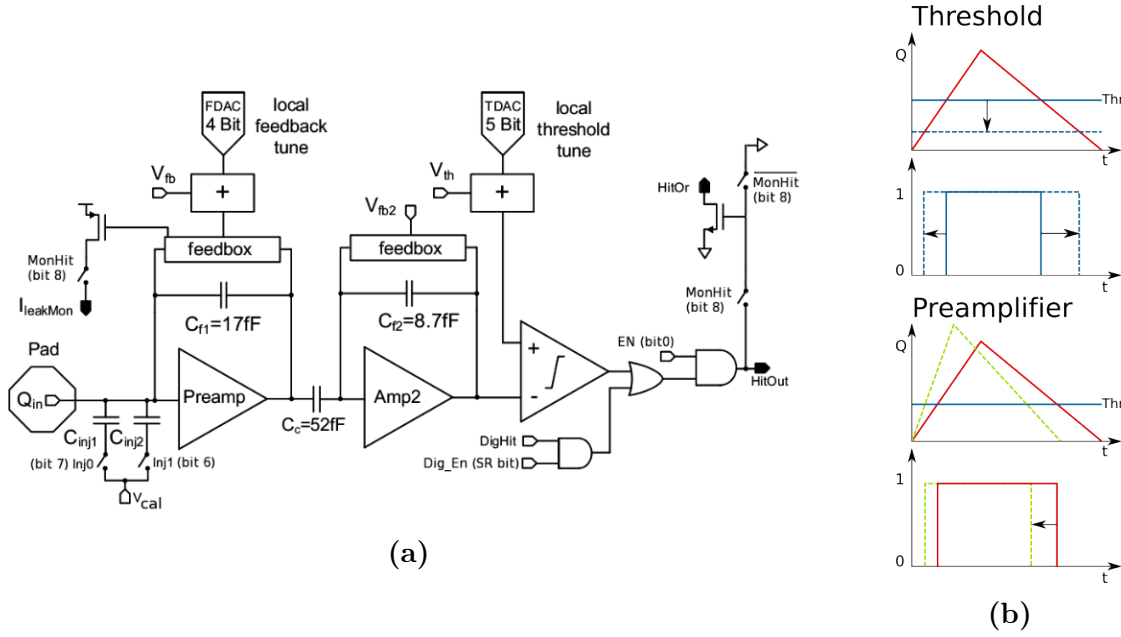


Figure 3.9: (a) Schematic of the pixel pre-amplifier and discriminator circuit [22], also shown are the 13 pixel configuration bits. Shown in (b) is the effect of the threshold of the discriminator and the preamplifier feedback current onto the charge to ToT conversion. Lowering the threshold increases the ToT of the same charge (shown in blue), increasing the feedback current lowers the ToT (shown in green).

3.5 The IBL Readout System

The IBL readout system is based on a pair of VME cards connected to the modules via an optical link. As seen in the Schematic in Fig. 3.10, the readout system is made up of four hardware pieces with different functionality. Housed in a VME crate are multiple pairs of Read Out Driver (ROD) and Back Of Crate (BOC) cards and a single Timing Interface Module (TIM). The optical and electrical conversion on

⁶Lowest possible threshold depends on noise, which is strongly influenced by the sensor capacitance.

3.5. The IBL Readout System

the detector is performed by the Optoboard, which is directly connected to the IBL modules via an electrical LVDS signal. The ROD and BOC are new developments for the IBL, as they have to deal with a higher bandwidth data output from the modules and a new data format. Each FE-I4 has its own link to send data to the off-detector readout, but the Timing, Trigger and Control (TTC) link, one clock and one command signal, is shared by two FE-I4s readout chips.

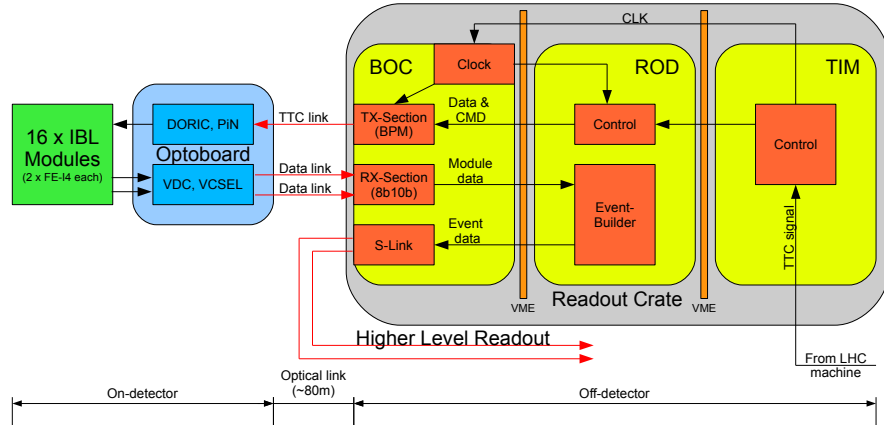


Figure 3.10: Schematic of the IBL readout system showing the data flow from the Off-detector site to the detector and back.

The functionality of the hardware can be summarised as follows:

- The TIM [12] receives and distributes the LHC 40 MHz clock, which is synchronous to the bunch crossings, and the ATLAS Level 1 trigger signals to all cards in the VME crate.
- The ROD card [11], shown in Fig. 3.11a, is used to steer the chips and frame the received data. During normal operation it sends the Level 1 trigger commands to the chips, builds the received event which is then sent to the higher level readout. It also checks that all chips are still synchronised and if not it can issue a resynchronisation to a specific chip. The card is controlled by a Xilinx Virtex 5 FPGA running a PPC⁷ which can be interfaced via ethernet. In total there are two slave Xilinx Spartan 6 FPGAs, each connected to 16 chips, responsible for building the event and checking the data integrity. During calibration the PPC steers the different scans necessary for tuning the modules, the two slave FPGAs make use of SRAMs⁸ to histogram the received data. These histograms are sent by the slaves via ethernet to a computer farm that fits and analyses the data.
- The BOC card [46], shown in Fig. 3.11b, performs the optical-electrical conversion of the signals sent to and received from the detector. Two Spartan 6

⁷Power PC: A microprocessor architecture.

⁸Synchronous Random Access Memory

FPGAs, each connected to 16 chips, perform encoding and decoding of the outgoing and incoming data. These two FPGAs are controlled by another Xilinx Spartan 6 FPGA, which can be interfaced via ethernet or via a custom bus from the ROD. Bi-PhaseMark (BPM) encoding is performed for the TTC link, to encode the 40 MHz LHC clock together with the commands from the ROD into a single 40 Mb/s serial link. The data coming from the chips is an 8b/10b encoded 160 Mb/s serial stream, after decoding and deserialisation the data is forwarded to the ROD slaves. Snap-12⁹ optical plugins are used to connect to the optical link. The higher level readout can receive data from the BOC via the S-Link [30], a custom FIFO-like data link used throughout the ATLAS DAQ system.

- The Optoboard uses two chips to receive and decode the TTC link and forward the data stream via the optical link. The optical signal of the TTC link is received by a PiN diode, converting the optical to an electrical signal. This BPM encoded signal is then decoded by the DORIC chip into one clock and one command stream, which are shared by the two FE-I4s. The other chip on the Optoboard is the VDC, a Laser driver circuit which converts the electrical data stream into an optical one.

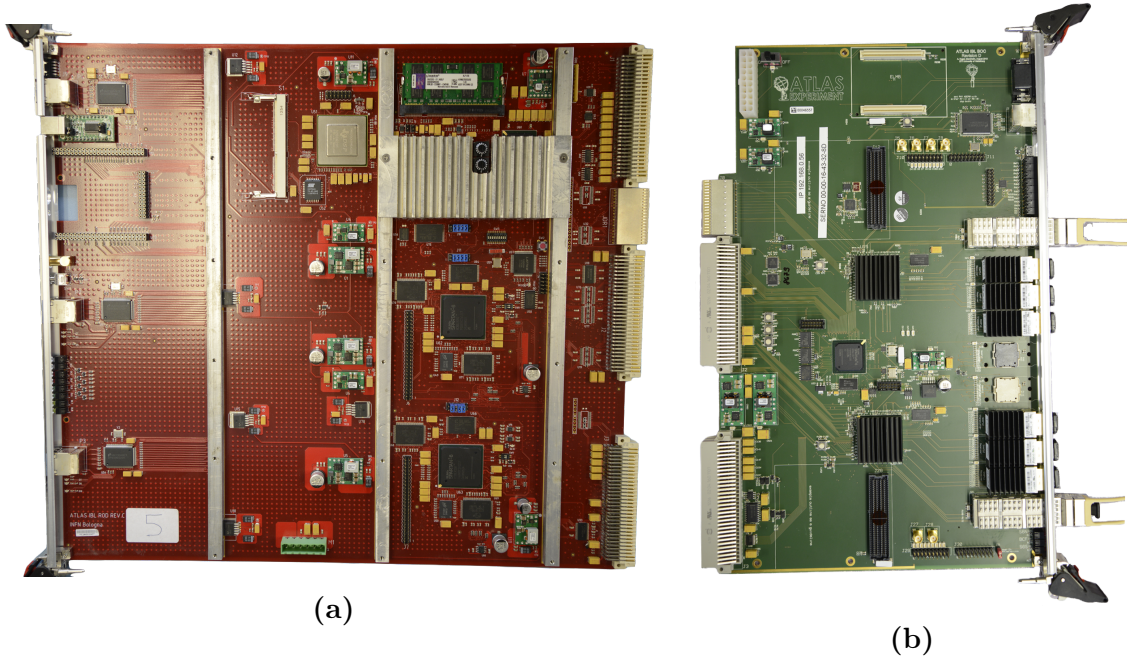


Figure 3.11: Pictures of the (a) IBL ROD card and (b) the IBL BOC card.

⁹A standardised optical plugin interconnect with 12 optical channels [1].

3.6 Powering and Detector Control System

In the IBL powering and control scheme [27], two double chip or four single chip modules, i.e. four FE-I4 chips, are connected in parallel as one unit, called a powering group. Each powering group is supplied with a low voltage (LV) to power the FE-I4s and a high voltage (HV) to bias the sensors. Each powering group also has one NTC temperature sensor. Each half stave is connected via a cable board to the DCS services, as shown in Fig. 3.12, a total of four powering groups are connected there. The cable board itself holds either another NTC¹⁰ temperature sensor or a humidity sensor. The Type 1 services connect the cable board to the PP1¹¹, which is approximately 3.5 m away from the cable board and just outside the inner detector, to Type 2 services. The HV, temperature, and humidity sensors are just routed through the PP2 to the counting rooms, but for the LV the PP2 has sensing regulators. The regulators need to be as close as possible to the detector and sensed, as the voltage drop for the nominal 2 V supply would be too high if delivered directly from the counting room. These PP2 regulators also supply the power to the Opto-boards, which perform the electro-optical conversion of the TTC and data link. The

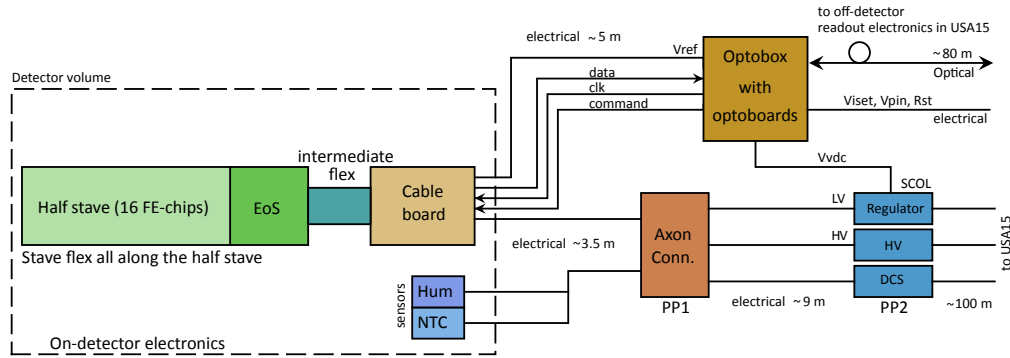


Figure 3.12: Schematic of the IBL powering and control scheme.

DCS Finite State Machine (FSM) [28] takes care of operating all power supplies and monitoring the voltage, current, temperature and humidity readings. It assures that each powering group is switched on using the appropriate procedure and all monitor values are in a nominal range. In case of a value being too high or too low, there is a two stage interlock. The first stage is the soft interlock, which is performed by the DCS FSM, and acts in the order of seconds e.g. if a current is too high the interlock will switch off the respective module. The second stage is a hardware interlock and it acts only upon specific signals supplied to the Interlock Matrix Crate. Normally the soft interlock should always act before the hard interlock takes place, hence the hard interlock plays a failsafe role and prevents all power supplies from being turned on until the interlock is released and acknowledged.

¹⁰Negative Temperature Coefficient: A component whose resistance is dependant on temperature.

¹¹Patch Panel: a specific location in the experiment where services are interconnected.

3.7 Construction and Integration

The focus of the IBL production is 14 staves with 280 sensor pixel modules. The production sequence of the different components is outlined in Fig. 3.14 up to the point where the stave is integrated on the Inner Positioning Tube (IPT). Sensor and FE-I4 chips are bump bonded in a flip-chip process, which is a very delicate procedure due to the small pixel pitch and low thickness of the sensor and chip. At this stage sensors and FE-I4 chips are already tested, so that only functioning modules are produced. In the module assembly at the University of Bonn and INFN Genova a flex is glued to the module which then undergoes a thorough testing procedure. As the initial wafer probing of the chips cannot test every operational aspect, modules can still fail this quality check. The overall yield per sensor type and batch number is shown in Fig. 3.13. The first batch has a higher number of failures, due to a problem during bump bonding, which was fixed for the following batches resulting in a yield of 75%, 63%, and 62% for Planar Pixel Sensor (PPS) double chip, CNM single chip modules, and FBK single chip modules, respectively. In total, more than 700 modules were produced for the IBL and many of these were left over as spares and can be used for future detector development.

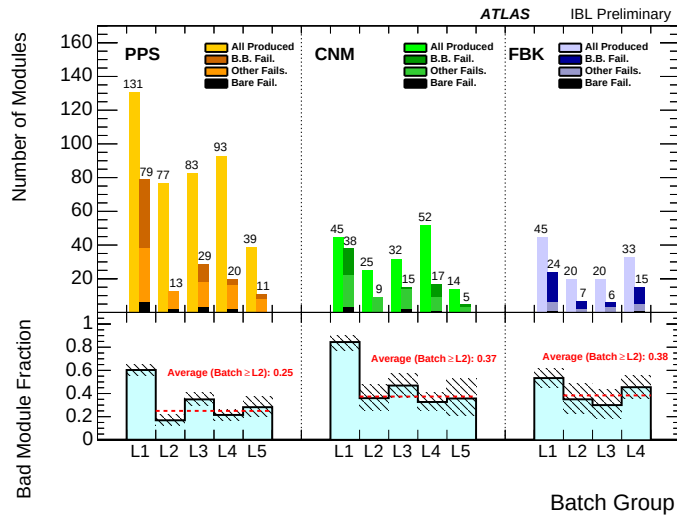


Figure 3.13: Module production yield for the 4/5 different module batches assembled in Bonn and Genova.

Modules which were of IBL quality were sent to the University of Geneva, where they were loaded onto a stave-flex assembly. After the loaded stave has been checked in a brief electrical test for any crude failures, it was delivered to CERN to be fully qualified. Before a stave can be integrated on the IPT the cooling pipes need to be extended to their full length by 2.75 m on each side, in a process called brazing. After brazing, integration on the IPT, and connection of services, the staves are electrically tested and checked for any deviation from the results of the stave testing.

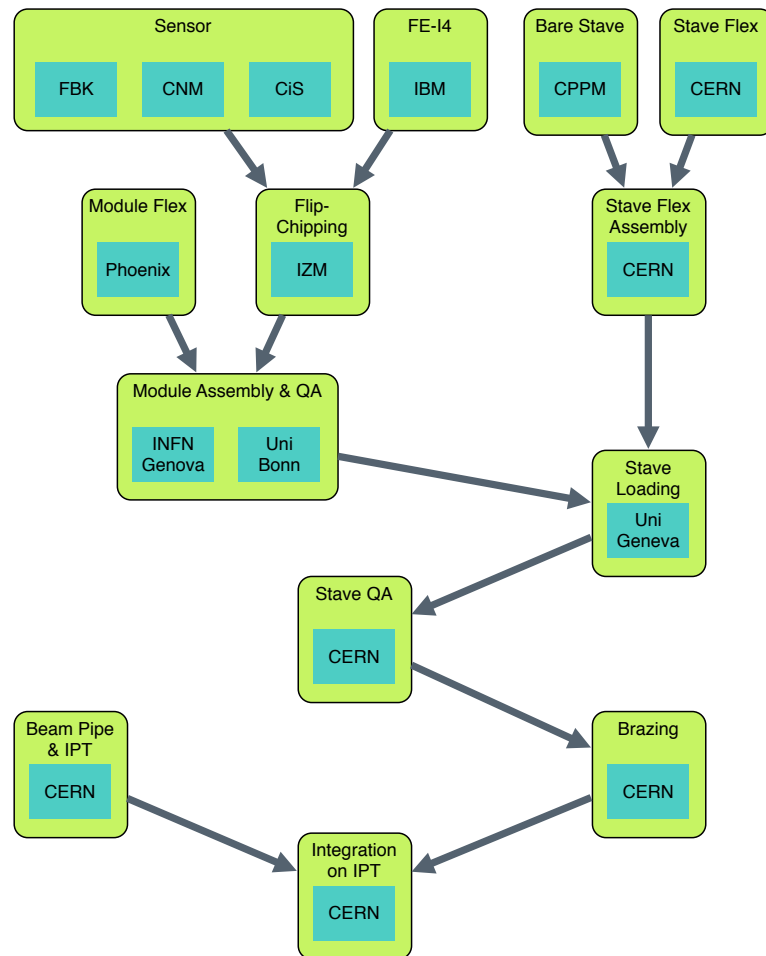


Figure 3.14: Flow chart showing the path of each part of IBL up to the point where the detector was fully assembled on surface. Each stage has its own quality control, to be able to trace back to where the source of problems occur.

Chapter 4

The IBL Stave Testing and Performance

A total of 20 staves have been produced for the IBL. These were tested for their performance, to select the best 14 staves for integration around the beam pipe. The experience gained and data gathered during the stave quality assurance (QA) procedure was an important milestone towards the assembly of the IBL, as it was the first operation of staves in a detector-like environment. Although single and double chip modules have been operated for quite some time before the first stave was built, the table-top operation of module is very different to the operation inside the detector. It was important not only to verify that the whole IBL powering and control scheme worked, but also to verify that the modules deliver the expected performance on stave. Many of the the presented results have already been published in [44].

The naming scheme, which is used in the following to identify the different chips on a stave, is described and shown in Appendix A.1.

4.1 Timeline

The stave QA can be divided in three stages: setup and commissioning with two prototype staves, initial production QA up to the time wire bond corrosion was discovered, and final production QA of all 20 staves. The experience gained during each stage led to improvements of the setup, a better understanding of the electrical properties, and adjustments to the procedure. The two prototype staves (ST00A and ST00B) were the first staves to be assembled using the production procedure, ST00A having modules with prototype FE-I4A chips and ST00B having modules with FE-I4B chips, which are also the chips used for the IBL production.

In July 2012 ST00A arrived in SR1 and was tested with a two week procedure, mimicking the plan of production QA. The two week procedure was interrupted by the discovery and subsequent investigation of low voltage oscillations in the FE-I4A chips, but allowed the commissioning of the whole test setup and readied it for the production QA. ST00A was followed up with ST00B, the first production stave with

FE-I4B modules, but the modules were of too low quality, due to the bump bonding issue (see Section 3.7), to be considered for IBL. In contrast to the FE-I4A, the FE-I4B chips use on-chip regulators to generate their analog and digital voltage, the phenomenon of low-voltage oscillations is suppressed, which solved many problems faced during the QA of ST00A.

The first production stave, ST01, arrived in April 2013 and by then the two week QA procedure was finalised based on the experience gained from the prototype staves. ST01, and the six production staves which followed until August 2013, showed excellent performance and the schedule allowed for extra investigations outside of the QA procedure to be performed. Whilst ST07 and ST08 were under test at $-25\text{ }^{\circ}\text{C}$, the cover of the environmental box in which the staves were tested, and which is flushed with dry air, was not properly closed. Due to outage of the dry air supply, a generally humid environment inside the lab, and the cover not being properly closed, humid air got inside the environmental box to the cold staves. This led to the formation of ice around the stave, which was not discovered until the staves were warmed up and the subsequent melting of the ice produced high leakage currents on the sensors. After leaving the staves for a week in a dry environment, an optical inspection revealed not only that a white residue from the evaporated water had formed, but also that the wire bonds had corroded. This led to the optical inspection of all staves, which were tested in the QA setup and could have suffered from water condensation during cold tests. Corrosion was found, though in much earlier state, on all staves, including ST12 which had arrived in SR1 but was not yet tested. The subsequent investigation discovered that during the thermal cycling of the loaded stave performed at the loading site, the climate chamber ramped the temperature faster up than the stave could follow. Leading to a point in time when the dew point of the surrounding warm air was higher than the stave temperature, hence allowing water to condense on the modules during each thermal cycle. This was confirmed by ST11, which had not yet been thermal cycled and showed no traces of corrosion.

The wire bonds are made of an aluminum (99%) and silicon (1%) alloy with mass spectrometry of the corroded wire bonds revealing the existence of halogens (F and Cl). Together with water these elements are likely to chemically attack the wire bond which was reproduced with deionised water on samples, suggesting that the halogens are already localised on the module flex. Even with thorough plasma cleaning of the module flexes, corrosion could not be prevented. This is still not fully understood, but it is possible that the halogens are inside the flex material.

In order to avoid any corrosion it is crucial to keep the staves as dry as possible at all times. The stave QA test setup was adjusted to accommodate for this. The humidity monitoring inside the environmental box was improved and coupled to an interlock system, preventing cooling if the margin between the cooling point and the dew point is too low. Furthermore the staves are kept in a smaller volume which is constantly flushed with nitrogen, acting as a safeguard if the dry air supply and interlock system fails. All staves which had been thermal cycled after loading needed a thorough cleaning and if corrosion had already started, the wire bonds needed to be replaced. This was performed by the PH Department Silicon Facility

(DSF) located at CERN.

The stave QA restarted in November 2013 after a review of the setup and the QA procedure. A new QA was put in place as the time to test a stave had decreased from two weeks to 4 days, to meet the schedule for the installation of IBL. As ST07 and ST08 suffered major damage, they were used for testing the brazing and integration procedure. The QA finished testing all staves in February 2014 and the brazing and integration of the best staves started. Integration of all 14 staves finished by the beginning of April 2013.

In Fig. 4.1 the timeline of each stave is shown, the data does not necessarily represent the total number of days worked on each stave, as for some staves the production may have paused and been restarted later. It is notable, though, that the time spent in the stave QA drastically changed before and after the incident.

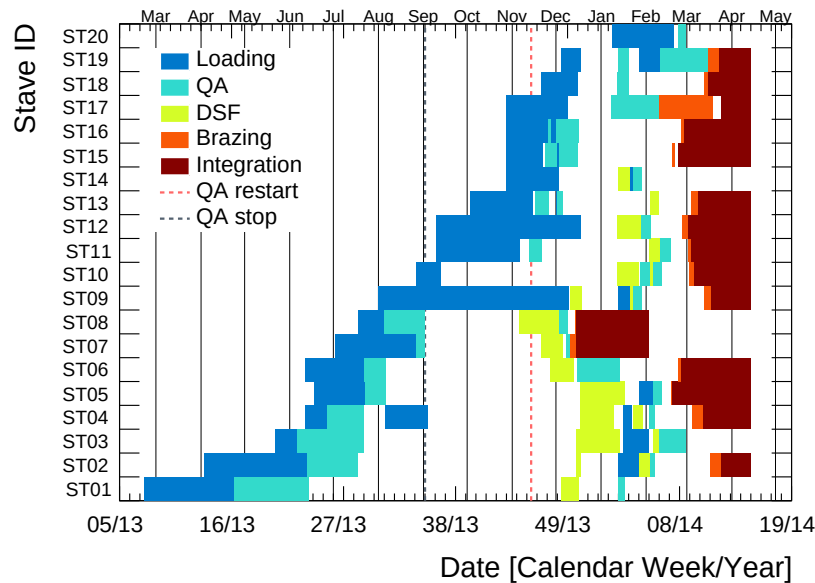


Figure 4.1: Graph showing each step of the production from loading, up to integration on the IPT for every stave. This does not necessarily represent the number of days worked on each stave, as the work on some staves was paused for specific reasons. For example, ST07 and ST08 were used to test the bring and integration procedure and were dismantled from the IPT afterwards.

4.2 Stave Test Stand

The stave test stand is located in the SR1 clean room at CERN and its purpose is to operate and test two loaded IBL staves under conditions close to the ones they will experience in the detector later. These conditions include CO₂ cooling and a powering scheme similar to the one used in the detector. At the time of building the setup, not all services for IBL were available, hence many services from the Pixel detector were used to complete the setup, but this has no effect on the comparability

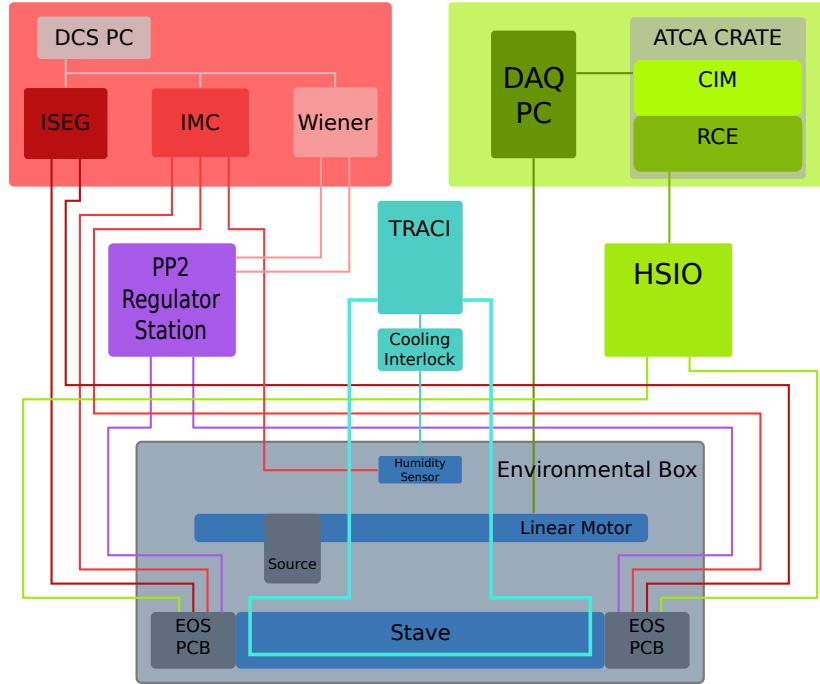


Figure 4.2: Schematic of the stave test stand, showing the different components needed for powering, control, readout, and cooling [44].

to the operation in the detector.

A schema of the setup is shown in Fig. 4.2. The main component is an environmental box which allows mounting and connection of two staves. This box (see picture in Fig. 4.3) measures $2 \times 1 \times 1 \text{ m}^3$ and is flushed with dry air, which permits the TRACI (Transportable Refrigeration Apparatus for CO₂ Investigation) cooling system to cool the staves down to -30°C . As it is very important to keep the staves dry and prevent any condensation of water upon it, the humidity in the box is constantly monitored and prevents powering and cooling if the dew point is less than 20°C away from the environmental temperature. The stave itself is kept in a smaller volume which is flushed with nitrogen as a further failsafe mechanism. The Staves are connected via an End of Stave (EOS) PCB to the powering and readout, this PCB is only mounted for testing purposes and is replaced by a cable board during the integration of the staves on the IPT.

The readout of the staves is performed by the RCE readout system, an ATCA¹ based readout which consists of three components:

- The Reconfigurable Cluster Element (RCE) is a board in the ATCA standard and is equipped with a PPC processor running a real-time operating system, in charge of performing the main processing during calibration and data-taking. It runs a custom software, developed to perform scans on FE-I4 chips.

¹Advanced Telecommunications Computing Architecture

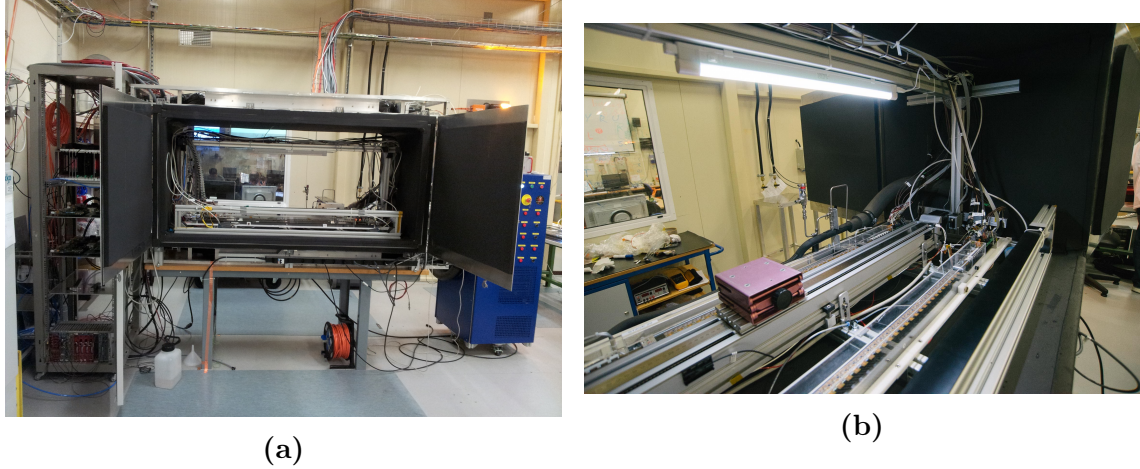


Figure 4.3: (a) The environmental box seen from the outside, with the TRACI (blue) on the right, parts of the powering and readout in the rack on the left. (b) The environmental box seen from the inside with two staves mounted and connected, the linear stage to move a radioactive source is in between the two staves (source not mounted).

- The Cluster Interconnect Module (CIM) interfaces the RCEs in one crate with an external computer, which controls execution of scans.
- The High Speed I/O (HSIO) card is a reconfigurable I/O interface and connects electrically to the FE-I4 chips. It is connected via an optical link to the RCE running a custom protocol at 3.125 Gb/s, which is used to send FE-I4 configuration to the HSIO and data to the RCE.

Powering and monitoring is supplied by the following DCS components:

- The Wiener² power supply is used to supply the low voltage to operate the FE-I4. For protection of the detector the Wiener power supply can be hardware interlocked by an external source.
- The ISEG³ modules are used to provide high voltage to bias the sensors, it can supply up 1000 V or 500 V depending if it used for planar or 3D sensors, respectively.
- The PP2 regulator station is a sensing power supply, which provides the low voltage to the readout chips and compensates for the voltage drop in the cables to the chip. It is supplied by the Wiener power supply and the PP2 finely regulates the provided voltage.

²Worldwide-Industrial Electronics-Nuclear Electronics-Resources W-IE-NE-R, Plein & Baus GmbH, Müllersbaum 20, 51399 Burscheid Germany, <http://www.wiener-d.com/>.

³iseg Spezialelektronik GmbH, Bautzner Landstrasse 23, 01454 Radeberg / OT, Rossendorf, Germany, <http://www.iseg-hv.com>.

- The Interlock Matrix Crate (IMC) is system which can receive a variety of inputs, which are monitored and can be used to interlock other DCS components. It provides interlock signals on a hardware level and the last safeguard to protect the detector from damage.

4.3 Stave Test Procedure

The stave test procedure was chosen such that each stave would be fully characterised over the course of 4 days. Two staves can be tested at the same time and the tests run each day are shown in Fig. 4.4. On arrival in SR1, and before the stave is installed in

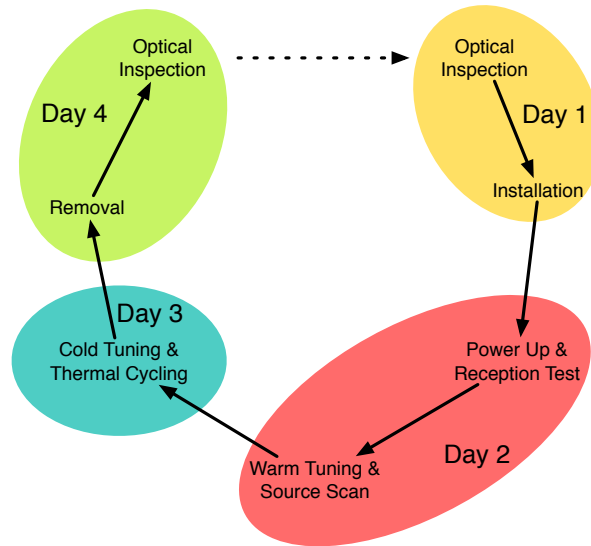


Figure 4.4: Flow of the stave testing procedure, outlining which tests are performed on which day. Two staves go through this procedure simultaneously and the actions of the last day and the first day of a new stave pair can be performed in parallel.

the test setup, a thorough optical inspection is performed which documents the state of the stave upon arrival. After installation the power up of the modules is checked and a short reception test scan is executed. The performance of the stave is then measured in regards to tuning the modules and response to a radioactive source. Thus far in the QA, all measurements have been performed at approximately 22 °C, however on the third day performance measurements are made at -15 °C. In general the performance should not be influenced by temperature, but it has been observed that specific problems are only in scans performed at cold temperatures, and, since the nominal operation temperature of IBL is -30 °C, these have to be excluded. Subsequently, the stave is warmed up and cooled down three times and undergoes a short set of scans to test if the temperature cycles changed its behaviour in any

way. After the stave has been removed from the test stand, it undergoes a second optical inspection to determine its state after testing and a create a reference point for the next production step.

4.3.1 Optical inspection

The optical inspection is twofold: a high resolution photograph is taken of each module and the stave is inspected with a microscope with special attention to the wire bonds. An example of one of these pictures is shown in Fig. 4.5 and also an example of an issue found in the inspection, a loose wire bond. Pictures taken of

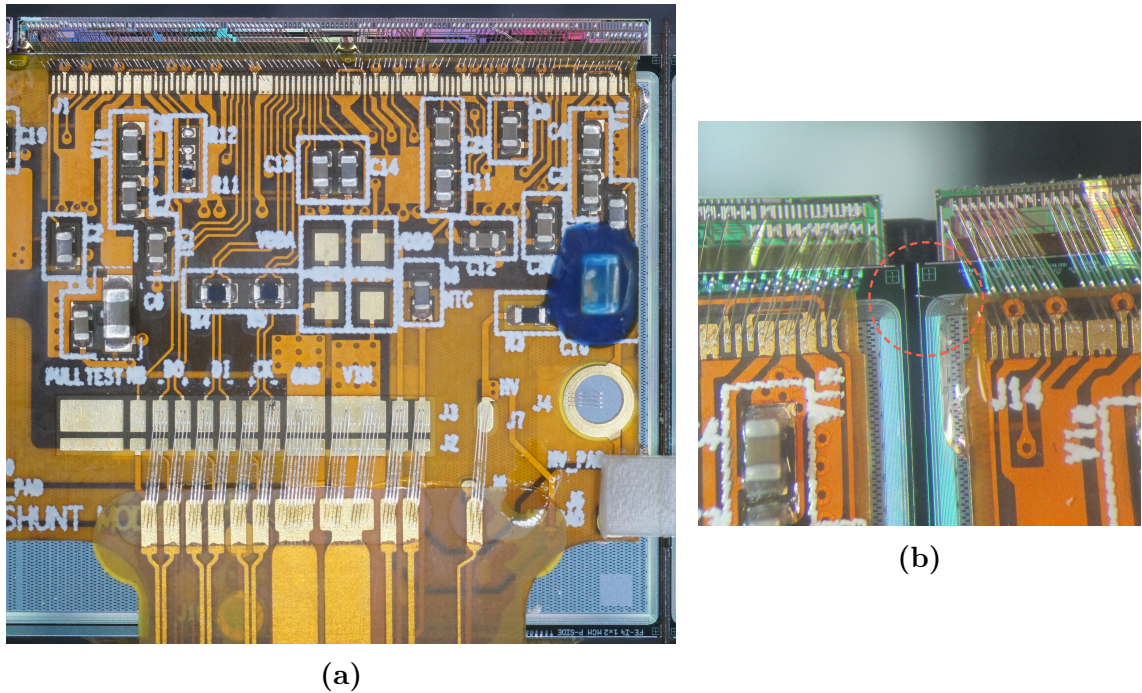


Figure 4.5: (a) A example photograph of a module taken during the optical inspection and (b) a possible issue discovered during the optical inspection, a loose wire bond close to the sensor edge which needed to be removed.

every module are compared to pictures taken after the stave has been loaded, to identify if distinctive features were already present on the module and are evaluated to see if they could be problematic. The stave is then thoroughly inspected through a microscope, each wire bond and wire bond pad is checked for issues or signs of corrosion. Possible issues could be: two wire bonds touching, a bond foot which has lifted off or a bond foot touching another pad. In general the delicate nature of the wire bonds and the complex and difficult sequence to replace them, led to a meticulous inspection of each of over 150 wire bonds. This procedure is repeated after the stave has been tested and is removed from the box, ensuring it was not damaged during testing and documents the state after testing for the following steps to come.

4.3.2 Electrical functionality

The first electrical check performed on the stave is a simple startup check. With this check two things can be verified: the stave has been properly connected to the sensing power supplies and the modules power up nominally and consistently. The former is especially important as the modules will be damaged if supplied with a voltage higher than 2.5 V. The regulators located in the PP2 crate sense the voltage supplied to the stave to precisely adjust the low voltage, incorporating the voltage drop over the cable and supply the set voltage at the end of the stave. In the case that the sense lines are not connected properly, the voltage would be adjusted higher and higher as it would assume the voltage drop is too high. The sense line check performed before the power up of the modules, ensures that the sense lines are connected properly and the voltage is adjusted correctly. The stave is then powered on 10 times in a row and the drawn current is monitored. The current should not change much over the 10 power cycles, as seen in Fig. 4.6a, although it is possible that current fluctuations of the chip do not perform a proper power-on reset. This is verified for abnormal modules by monitoring the current after configuration. The currents directly after power-up and after configuration are stored for all modules and are important figures of merit for the DCS of the detector after installation in the ATLAS experiment.

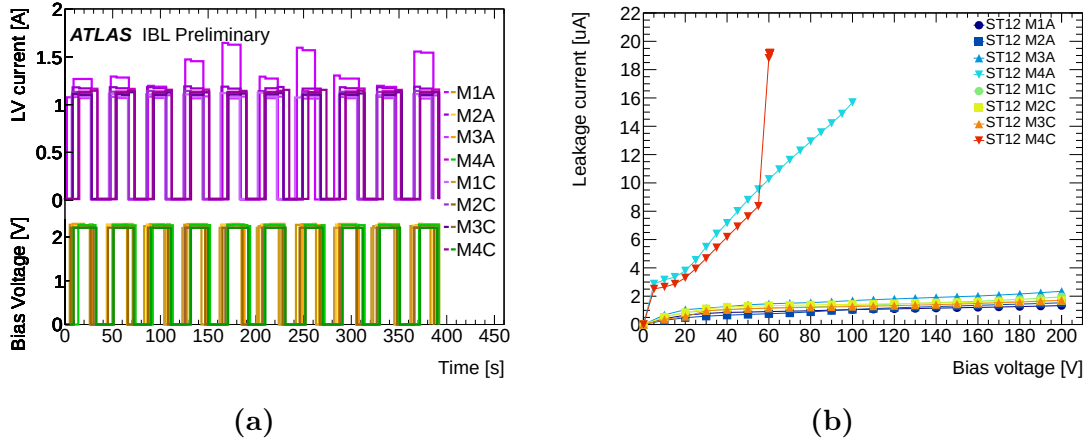


Figure 4.6: Results of one stave of the (a) power cycles and (b) sensor IV characteristic [44].

The current-voltage (IV) characteristic of the sensor is verified and an exemplary result is shown in Fig. 4.6b. Planar sensor modules are tested up to 200 V and 3D sensor modules up to 100 V. For planar sensors the breakdown should not occur before 100 V and for 3D sensors not before 20 V, the leakage current before breakdown should be below 15 μA for both types of sensor. The IV characteristic, which changes over the course of the production, can point to problems like mechanical stress, scratches, cracks or problematic glue.

4.3.3 Reception Test

The first set of scans performed on the stave is called the reception test and it is similar to the scans which have already been performed directly after stave loading with the same configuration. These include: a Digital Scan, Analog Scan, ToT Scan, Threshold Scan and Threshold Scan without biasing the sensor. This does not yet allow the performance of the stave to be evaluated, but the results can be directly compared to the ones taken after loading. If inconsistencies are observed, the stave must have been damaged during handling or, more likely, the inconsistency is based upon the differences in the setup. For example, this method helped to identify the issue of increased threshold noise on FBK 3D modules, see Sec. 4.6.3.

4.3.4 Calibration

How well the chips can be calibrated is a mayor indicator for the performance of the stave. The existing calibration was performed during the module QA and is not accurate enough anymore after the module has been loaded on the stave. The calibration is performed with the target parameters being: a 10 bc ToT response for 16000 e and a threshold of 3000 e and 1500 e at 22 °C and -15 °C respectively. It should be possible for every pixel to be tuned close to these settings and although outliers have some kind of issue they are not necessarily completely defective. Important for the operation of the detector is the noise hit occupancy for the specific tunings, as this can seriously influence the data taking performance. During calibration each pixel is also tested for its digital and analog functionality. If this is then combined with the results from the tunings the number of electrically working pixels can be determined.

4.3.5 Source Scan

To evaluate the performance of the stave under realistic conditions particles should be detected with it. A Sr^{90} radioactive source is used, which emits 0.5 MeV and 2.3 MeV electrons with an activity of 28.8 MBq, where the 2.3 MeV electron is close to the energy of a minimum ionising particle (mip) and should deposit approximately 16000 e . The modules are operated with a threshold of 3000 e at 22 °C and a ToT response of 10 bc ToT for 16000 e . There are two available radioactive sources and the linear stage positions them above two chips, where the data is then collected for 400 s. The chips are running in self trigger mode, where the output of the hitbus is used as feedback to trigger the chip. I.e. there is no need for an external trigger (e.g. a scintillator) and instead the chips will trigger on every energy deposition above threshold, which drastically increases the chance to trigger on particles which get stuck in the sensor and deposit less energy than a mip particle. A source scan for two staves is performed in around 5 hours, this time is optimised to fit into the four day testing procedure and still give conclusive results.

An example of a source scan of a single chip is shown in Fig. 4.7. The structures seen in the hit map are the passive components mounted on the module flex. The

average numbers of hits per pixel is 172 ± 37 and, considering a failing pixel should have no hits, this is a decent number of hits to distinguish between a good and a bad pixel but there is a long tail towards a smaller number of hits due to the passive components and pixels which are functional but disconnected from the sensor, can have noise hits. As the noise occupancy after masking noisy pixels is below 10^{-6} hits/pixel/bc and during one source scan around 5×10^6 triggers are read out, noisy pixels should produce less than 5 hits. This is very close to the occupancy observed below the large HV filter capacitance and the cut value is tuned to not include pixels with low occupancy under these components, resulting in a cut of 1% mean occupancy equal to more than 1 to 2 hits. Although there are indicators from other

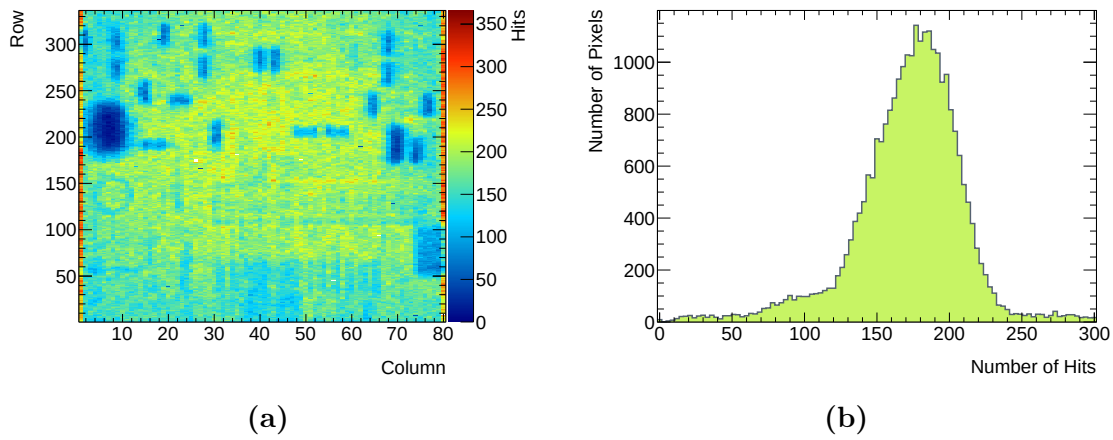


Figure 4.7: Results of a source scan performed on ST11, (a) the hit map with the passive components lowering the occupancy and (b) the distribution of number of hits per pixel.

scans, a source scan is the only conclusive way to identify a disconnected bump, e.g. if a pixel is declared functional during the previous tests but does not show any or not enough hits during a source scan, the bump is disconnected. This information, combined with calibration, allows the determination of the total number of working pixels, the primary indicator of the performance of a stave.

4.4 Database and Analysis Framework

It is vital to every QA to store the results in an accessible format, so that the data is, on the one hand, easily accessible for direct analysis, but also available in the long term as a reference point during the commissioning. During the stave QA it is possible to perform tests which are not possible after installation, but could help in solving issues which might appear during operation. Hence the data which is taken during the QA is organised in two databases: one - the “RCE Run DB” - containing the information of every scan performed with the RCE readout and the second one - the “Stave QA DB” containing the analysis and test results of every production stave. The databases are implemented in MySQL, which is a very

popular open source relational database management system used frequently for web applications.

An analysis framework was developed to interact with both these databases, analysing raw data from the RCE and pushing the results into the QA database. The block diagram in Figure 4.8 outlines how the framework interacts with the databases. As the framework has access to all scans performed on the RCE and provides a common interface to analyse different types of scans in modular way, it is used for a number of tasks ranging from analysis of stave performance, comparison to another QA, generation of summary plots, or on-the-fly analysis during a QA shift. The results are saved in tables which summarise the most important values and parameters, but additionally a binary ROOT file is contained in the table containing all histograms and graphs the results were taken from. Access to the databases is made via network, enabling users to run the framework and develop analyses on their own laptop, which increases the accessibility compared to running it from a centralised server. All data which is not contained in the RCE database is added

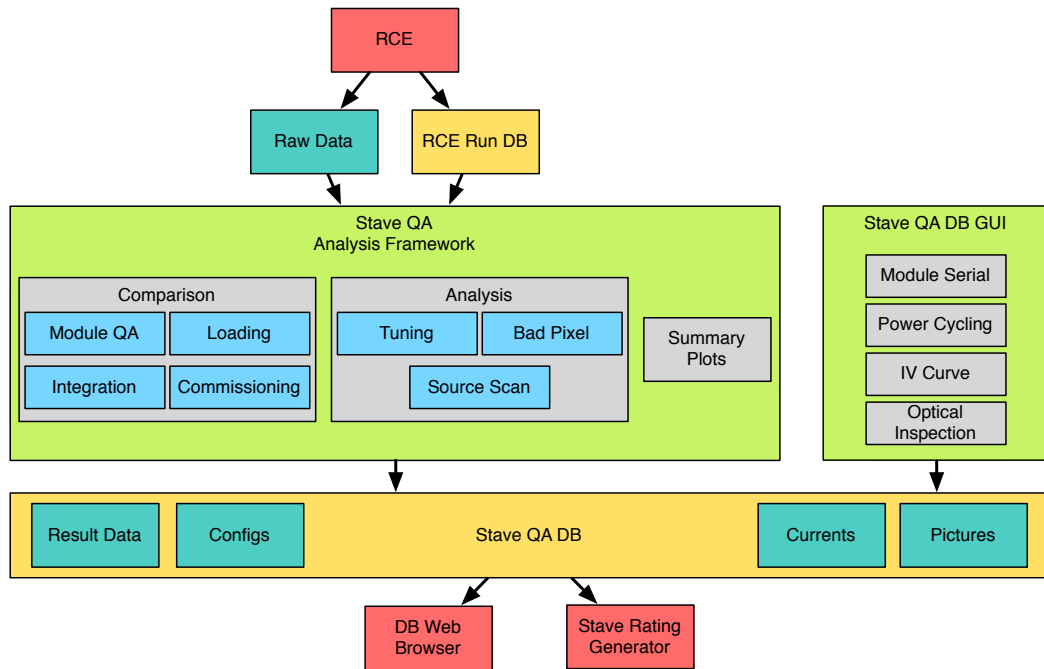


Figure 4.8: Block diagram of the stave QA framework, showing the interaction between RCE database and stave QA database.

by a user via a special application, this includes: pictures and comments from the optical inspections, LV power cycling data, IV characteristics and the module serial number map.

The data contained in the QA database is made accessible via a web interface, enabling a user to browse through the results of every stave in a natural way. An example of an overview table produced by the web interface can be seen in Fig. 4.9, data can be viewed down to a per chip level. This makes the access to the data very

easy and, more importantly, independent of any local software, which is usually a big problem as, in time, software often has problems if it is not maintained to stay up to date with novel devices and platforms.

Chip Summary and QA Status

DCS Grp.	Module	Chip Addr.	Chip Serial	Module ID	Repl.	Type	gAddr	Prod.	Production Summary				SR1 QA Summary							
									Rank	V _{bd} (GE)	N _{bad_bump}	N _{bad_flex}	Opt.Inspc	LV	DCS	Tuning	Src.Scan	Bad.Pixel	Total Bad Pixels	Comment
M4C	C8	C8-2	10-06-03-59	F10-06-03	--	3D	7	Genova	230	44	0	180	✓			✓	✓	✓	18	--
		C8-1	12-09-07-31	F12-09-07	--		6	Genova	212	42	3	112	✓	✓		✓	✓	✓	10	--
	C7	C7-2	10-18-04-23	F10-18-04	--	3D	7	Genova	174	45	117	164	✓			✓	✓	✓	9	--
		C7-1	11-39-07-23	F11-39-07	--		6	Genova	153	40	0	153	✓			✓	✓	✓	32	--
M3C	C6	C6-2	92-19-03-33	D92-19-03	Yes	Planar	7	Bonn	153	250	134	438	✓			✓	✓	✓	8	--
		C6-1	92-19-03-34	D92-19-03	--		6	Bonn	153	250	134	438	✓	✓		✓	✓	✓	31	--
	C5	C5-2	93-12-04-46	D93-12-04	--	Planar	7	Genova	259	230	96	259	✓			✓	✓	✓	53	--
		C5-1	93-12-04-49	D93-12-04	--		6	Genova	259	230	96	259	✓			✓	✓	✓	114	--
M2C	C4	C4-2	93-22-02-60	D93-22-02	--	Planar	7	Genova	262	280	9	202	✓			✓	✓	✓	33	--
		C4-1	93-22-02-57	D93-22-02	--		6	Genova	262	280	9	202	✓	✓		✓	✓	✓	5	--
	C3	C3-2	93-19-02-15	D93-19-02	--	Planar	7	Bonn	251	>350	1	201	✓			✓	✓	✓	11	--
		C3-1	93-19-02-14	D93-19-02	--		6	Bonn	251	>350	1	201	✓			✓	✓	✓	8	--
M1C	C2	C2-2	93-15-04-24	D93-15-04	--	Planar	7	Genova	220	285	72	170	✓			✓	✓	✓	92	--
		C2-1	93-15-04-22	D93-15-04	--		6	Genova	220	285	72	170	✓	✓		✓	✓	✓	12	--
	C1	C1-2	60-18-04-58	D60-18-04	--	Planar	7	Genova	154	200	245	154	✓			✓	✓	✓	50	--
		C1-1	60-18-04-55	D60-18-04	--		6	Genova	154	200	245	154	✓			✓	✓	✓	392	--
M1A	A1	A1-1	93-23-03-03	D93-23-03	--	Planar	7	Genova	105	>350	8	105	✓			✓	✓	✓	11	--
		A1-2	93-23-03-01	D93-23-03	--		6	Genova	105	>350	8	105	✓	✓		✓	✓	✓	2	--
	A2	A2-1	93-20-04-30	D93-20-04	--	Planar	7	Bonn	213	>350	44	213	✓			✓	✓	✓	80	--
		A2-2	93-20-04-16	D93-20-04	--		6	Bonn	213	>350	44	213	✓			✓	✓	✓	18	--
M2A	A3	A3-1	93-23-02-54	D93-23-02	--	Planar	7	Genova	214	>350	18	214	✓			✓	✓	✓	17	--
		A3-2	93-23-02-53	D93-23-02	--		6	Genova	214	>350	18	214	✓	✓		✓	✓	✓	66	--
	A4	A4-1	93-20-02-40	D93-20-02	--	Planar	7	Bonn	223	>350	24	223	✓	✓		✓	✓	✓	63	--
		A4-2	93-20-02-38	D93-20-02	--		6	Bonn	223	>350	24	223	✓			✓	✓	✓	14	--
M3A	A5	A5-1	61-07-03-08	D61-07-03	--	Planar	7	Bonn	367	340	3	167	✓			✓	✓	✓	3	--
		A5-2	61-07-03-10	D61-07-03	--		6	Bonn	367	340	3	167	✓	✓		✓	✓	✓	7	--
	A6	A6-1	61-04-04-47	D61-04-04	Yes	Planar	7	Genova	568	340	163	368	✓	✓		✓	✓	✓	9	--
		A6-2	61-04-04-50	D61-04-04	--		6	Genova	568	340	163	368	✓			✓	✓	✓	8	--
M4A	A7	A7-1	10-01-08-58	F10-01-08	--	3D	7	Bonn	110	35	0	110	✓			✓	✓	✓	5	--
		A7-2	12-08-05-41	F12-08-05	--		6	Genova	109	35	15	109	✓	✓		✓	✓	✓	17	--
	A8	A8-1	12-04-01-18	F12-04-01	--	3D	7	Genova	140	40	0	130	✓			✓	✓	✓	23	--
		A8-2	10-25-05-15	F10-25-05	--		6	Genova	236	41	0	186	✓			✓	✓	✓	14	--

Figure 4.9: Example of an overview table produced by the web interface for ST03.

4.5 Stave Performance

The performance of a stave is determined in three different analyses: the tuning analysis evaluates the quality of the calibration, the total number of operating pixels is counted in the bad pixel analysis, and the source scan analysis is used to assess the response to real particles. The results of these analyses are then used to rank the staves and select the best 14 staves for integration. The following results include the data from 18 of the 20 produced staves, ST07 and ST08 are excluded as they could not be repaired after the condensation accident.

4.5.1 Tuning Analysis

The cold 1500 e and warm 3000 e tunings are analysed for how well the target threshold and ToT is reached and in general all staves have shown no severe issues

during tuning. The overall distribution of the threshold and threshold noise for a 1500 e tuning of all 18 staves is shown in Fig. 4.10. The standard deviation of the threshold is less than 45 e and the mean threshold noise is below 150 e . For the threshold, a second peak around 2000 e is observed, which is related to an issue in a scan during the tuning and it was possible to fix it upon inspection, but due to the minor effect the overall performance was not updated. The 3D FBK modules show an increased threshold noise and a second peak around 280 e , investigations have shown a systematic influence of the test setup on 3D FBK modules, which is discussed in Sec. 4.6.3. This is confirmed in the distribution of mean threshold and mean threshold noise per chip position in Fig. 4.11, showing an increased noise on the A-Side. No other systematic deviation is observed for any of the other positions. The overall results are summarised for each pixel type in Tab. 4.1. The noise is expected to be higher for long planar pixels due to their size and for 3D pixels due to their higher sensor capacitance. The quality of the tuning from an operational standpoint is reflected in the number of noisy pixels, which is measured to be 0.03% of all pixels for the 1500 e threshold tuning.

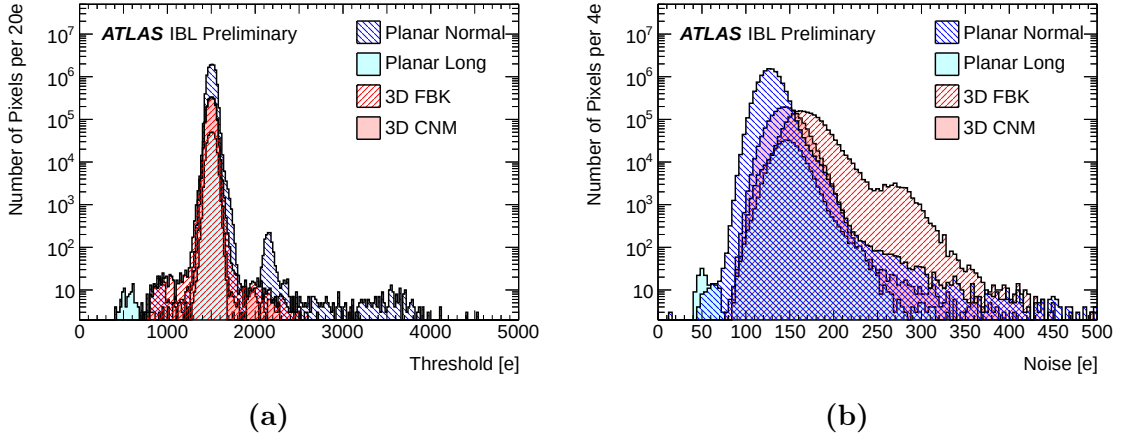


Figure 4.10: Resulting (a) threshold and (b) threshold sigma distribution for the cold tuning to a threshold of 1500 e of all 18 staves [44].

The preamp is calibrated to a response of 10 bc ToT for 16000 e , a mip particle should deposit on average 16000 e when it travels straight through a sensor of this thickness, i.e. the mode of the landau distribution of mip particles should be around 10 bc ToT, separating it from the background in the low ToT region. The mean ToT distribution is shown together with the mean ToT per chip in Fig. 4.12. The mean ToT favours integer ToT values as the charge interval where a pixel jumps between two ToT values is small compared to the charge interval where it is stable, this behaviour will be discussed in detail in Sec. 4.6.1. The dispersion for the reference charge is 0.2 bc and no dependance on the position on the stave is observed.

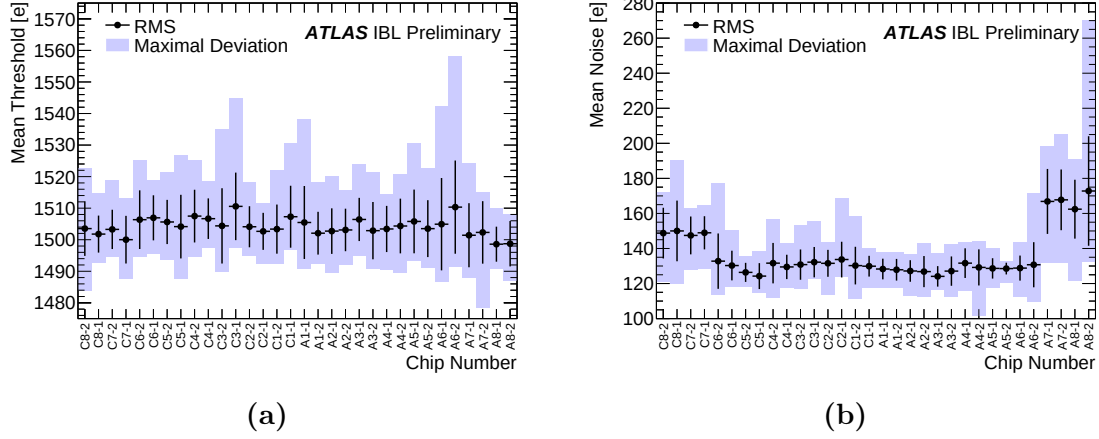


Figure 4.11: Mean (a) threshold and (b) threshold sigma of each position on a stave for the cold tuning to a threshold of 1500 e of all 18 staves [44].

Table 4.1: Threshold calibration summary for different pixel types. Listed values are the standard deviation of the threshold, mean noise and its standard deviation, and mean threshold over noise and its standard deviation [44].

Tuned Threshold	Pixel Type	Std. Dev. [e]	Noise [e]	Threshold over Noise
3000 e at 22°C	Planar Normal	37	123 ± 10	25 ± 2
	Planar Long	58	146 ± 15	21 ± 2
	3D FBK	39	171 ± 25	18 ± 2
	3D CNM	40	149 ± 15	20 ± 2
1500 e at -12°C	Planar Normal	42	129 ± 13	12 ± 1
	Planar Long	47	149 ± 16	10 ± 1
	3D FBK	46	171 ± 25	9 ± 1
	3D CNM	41	146 ± 16	10 ± 1

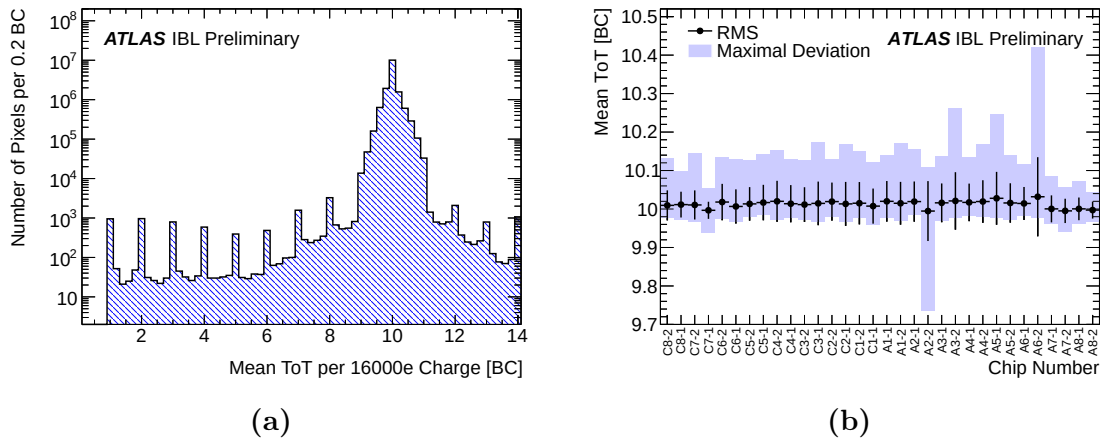


Figure 4.12: (a) resulting mean ToT distribution and (b) mean ToT of each position on a stave for a tuning of 10 bc ToT at 16000 e [44].

4.5.2 Source Scan Analysis

The source scan analysis serves two purposes: determine working and failing pixels and analyse the response to mips. The latter is discussed in this section, while the former is explored in the bad pixel analysis later. An important bottle neck in the analysis of source scan is that due to data bandwidth and memory size it is only possible to save histograms on the RCE during a source scan. I.e. the raw data is not available for post-processing and all important information has to be contained in histograms filled during the scan. This made it necessary to perform online clustering to measure the total charge deposited by a particle. As it is not possible to convert ToT to charge on the RCE, the sum of ToT per cluster is calculated. Technically this is not correct as the sum of ToT does not translate to the sum of charge, due to the threshold and the effects of non-linearity, but it is the best option which is feasible for implementation. The cluster algorithm looks for adjacent hits with a gap of max. one hit, clusters with a hit of 14 bc ToT are rejected, as this represents the overflow value. The resulting cluster ToT distribution of one chip is shown in Fig. 4.13. It has a high ratio of low energy hits which distort the expected Landau distribution. If only clusters with more than one hit are taken into account, the distribution allows fitting with a Landau function convoluted with a Gauss function. The resulting most probable value (MPV) does not give a quantitative answer to how close the result is to the theoretical value, because it is biased to higher values. However, it is possible to compare the result to other distributions and determine the overall continuity of the calibration and module performance.

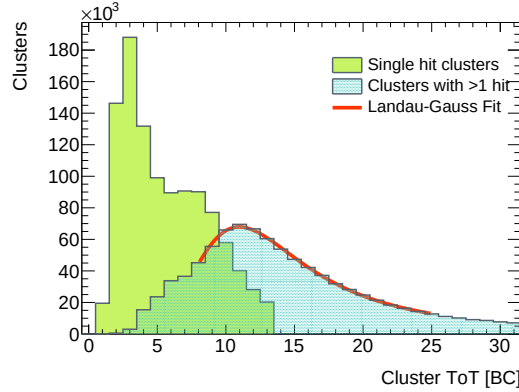


Figure 4.13: Cluster ToT distribution of one chip for a cluster size of one and clusters with more than one hit. A Landau-Gauss fit is shown for the clusters with more than hit, as most of the background is contained in the single hit clusters.

The result of this qualitative comparison is shown in Fig. 4.14 and the average cluster ToT MPV is 10.5 ± 0.3 bc, which is consistent with the 10 bc ToT at 16000 e tuning. The cluster ToT MPV for 3D modules is slightly higher than the one from planar modules, as the 3D module is 30 μm thicker and hence collects more charge. No significant deviation from the average is observed for a specific position on a

stave. This leads to the conclusion that the charge calibration of all staves can be performed uniformly over the whole detector.

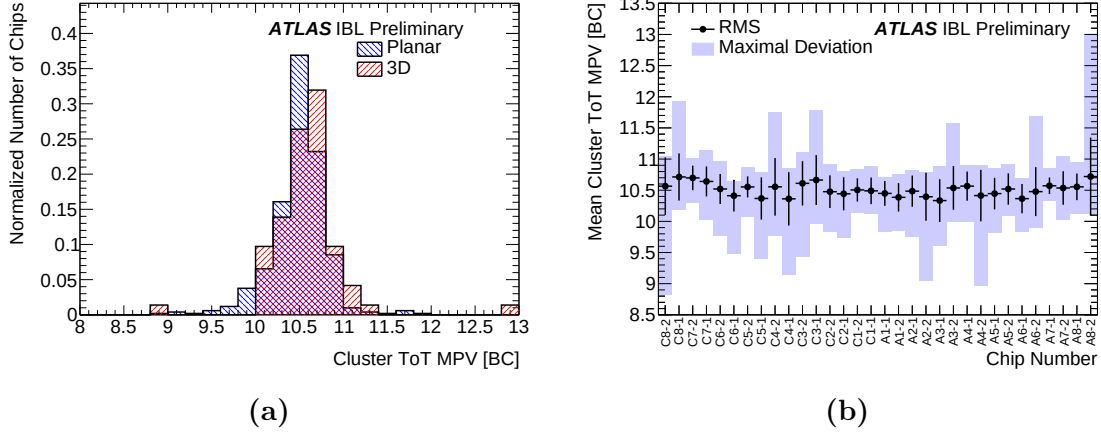


Figure 4.14: (a) distribution of cluster ToT MPV per chip and (b) mean cluster ToT MPV of each position on a stave [44].

4.5.3 Bad Pixel Analysis

By combining the results of the different scans performed during the calibration and from the source scan, it is possible to identify bad pixels and their failure mode. The different exclusive failure modes and their identification criteria are listed in Tab. 4.2 and an example of a bad pixel analysis for one stave is shown in Fig. 4.15. The most common failure, making up around 50% of all bad pixels, are disconnected bumps, the remaining 50% is equally divided amongst analog failures and pixels failing the tuning. All other failure modes do not occur very frequently, but have been observed sporadically without any systematic behaviour. This confirms that the flip-chip process for modules of this size, with small pixel pitch, and a low thickness, is the major source of pixel failures. The total number of bad pixels is shown in

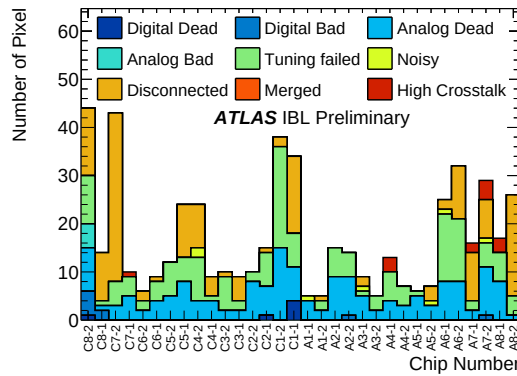
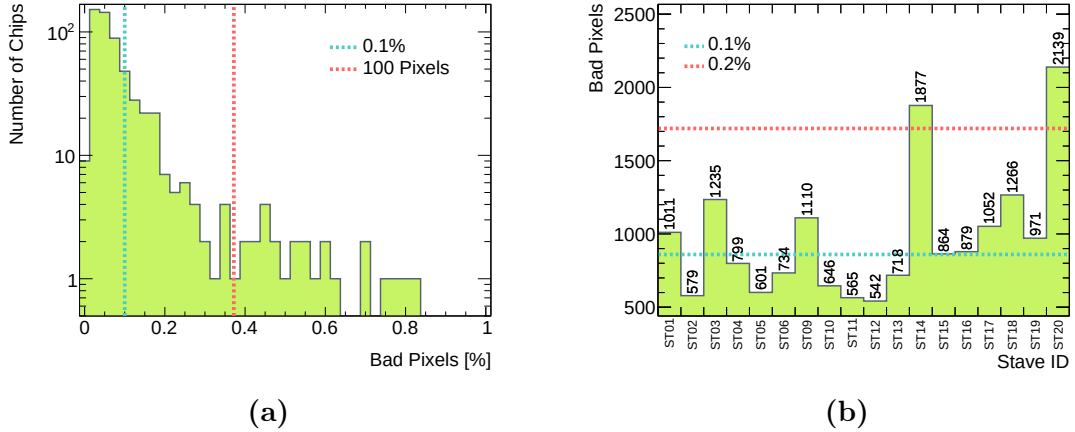


Figure 4.15: Results of a bad pixel analysis from one stave[44].

Table 4.2: Classification of pixel failures [44].

Failure Name	Scan Type	Criteria
Digital Dead	Digital Scan	Occupancy < 1% of injections
Digital Bad	Digital Scan	Occupancy < 98% or > 102% of injections
Merged Bump	Analog Scan	Occupancy < 98% or > 102% of injections
	Crosstalk Scan	Occupancy > 80% of 25 ke injections
Analog Dead	Analog Scan	Occupancy < 1% of injections
Analog Bad	Analog Scan	Occupancy < 98% or > 102% of injections
Tuning Failed	Threshold Scan	s-curve fit failed
	ToT Test	ToT response is 0 or 14 BCs
Noisy	Noise Scan	Occupancy > 10^{-6} hits per BC
Disc. Bump	Source Scan	Occupancy < 1% of mean Occupancy
High Crosstalk	Crosstalk Scan	Occupancy > 0 with 25 ke injection

Fig. 4.16 and the target for the IBL is to stay below 1% bad pixels. The average number of bad pixels per chip is 0.1% and 73% of all chips have less than 0.1% bad pixels. This very high module quality is also reflected in the total number of bad pixels per stave, 50% of all staves have less than 0.1% and 89% have less than 0.2%. Furthermore modules of higher quality are placed in the centre of a stave, i.e. modules with more bad pixels are placed in a region where the cluster size is increasing and the effect of single bad pixel is negated.

**Figure 4.16:** Total number of bad pixels per (a) chip and (b) stave.

4.5.4 Stave Selection

To select the 14 best staves out of the produced 18 the staves are ranked. The results from the analyses discussed in the prior sections, suggest that the main driver of the rank of a stave will be the number of bad pixels, as all staves performed equally well during tuning and an operational pixel is more important than a pixel which does

not tune well. The ranking weights every pixel failure by its η position to reflect its importance during operation and generates a score V as follows:

$$V = \frac{\sum_{i \in \text{bad pixels}} \cosh^{-1}(\eta_i)}{\sum_{i \in \text{all pixels}} \cosh^{-1}(\eta_i)} \quad (4.5.1)$$

This weighting enhances the central region of the detector and is preferred over a simple sum of bad pixels. As the integration of staves around the IPT started while the stave QA was still in progress, the selection was a continuous process of picking the best staves from a smaller sample size until the QA was finished. Due to the overall excellent performance of the staves, a stave with low planarity can be favoured instead of a better ranked stave. This is especially important for the integration of the last stave, which is inserted in between two staves. The selected staves are furthermore optimised for the position around the IPT, to uniformly distribute the bad pixels. The result of this selection can be seen in Fig. 4.17, which shows how the selection enhanced the centre region and the final operational fraction of pixels of the IBL, which is better than 99.9%.

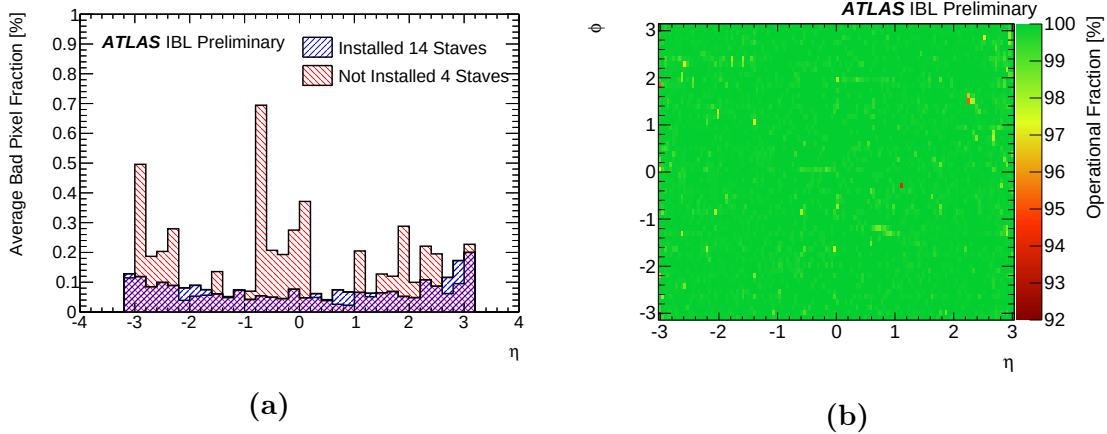


Figure 4.17: (a) number of bad pixels with respect to η for selected and rejected staves and (b) $\eta - \phi$ map of operational pixels of the IBL [44].

4.6 Other Measurements and Lessons Learned

The main purpose of all measurements performed during the QA is to qualify staves for the IBL and even though it had to be done in a short amount of time, the first operation of staves led to multiple, additional observations. Particularly the two prototype staves and first production staves were used for many measurements and investigations which are not part of the QA. In the following section some of these measurements and investigations are presented, as they might be of importance during the operation of the detector.

4.6.1 ToT Calibration

To simulate or reverse the digitisation of the charge to ToT, the offline software of the IBL detector needs to model the conversion. During the early stage of the stave testing a set of scans were performed to understand how the modelling can be performed and in which way it will differ from the current ToT calibration of the Pixel detector. It is also important to understand how the ToT calibration can be performed with minimal effort, i.e. with the least number of measurements and least number of parameters.

The investigation is performed on a data set of ToT scans performed on stave ST01 injecting charges from 1000 e to 23500 e in 250 e steps, with the chips on the stave being tuned to a threshold of 3000 e and a response of 10 bc ToT for a charge of 16000 e . Pixels which fail a digital or analog test are excluded from the analysis. As a range of charges from threshold up to around 23000 e are represented by only 14 ToT values⁴ the conversion from ToT to charge can not be very precise. In an ideal case there is a linear dependency between charge and ToT, but in reality the width of the charge interval given by one ToT increases with charge.

Four different functions are evaluated for how well they describe the digitisation:

$$f_1(Q) = \sum_{i=0}^{14} \frac{1}{2} \left(2 - \operatorname{erf} \left(\frac{Q - P_i}{\sqrt{2}} \right) \right) \quad (4.6.1)$$

$$f_2(Q) = \sum_{i=0}^{14} \frac{1}{2} \left(2 - \operatorname{erf} \left(\frac{Q - P_{2i}}{P_{2i+1}\sqrt{2}} \right) \right) \quad (4.6.2)$$

$$f_3(Q) = P_0 \frac{P_1 + Q}{P_2 + Q} \quad (4.6.3)$$

$$f_4(Q) = P_0 + P_1 Q + P_2 Q^2 \quad (4.6.4)$$

The first two functions $f_1(Q)$ (Eq. 4.6.1) and $f_2(Q)$ (Eq. 4.6.2) are step functions, $f_1(Q)$ with an ideal transition and $f_2(Q)$ with a smeared edge. $f_3(Q)$ (Eq. 4.6.3), is the function currently used for the ToT to charge conversion of the Pixel detector, and $f_4(Q)$ (Eq. 4.6.4) is a simple polynomial of 2nd order. The result of these 4 functions fitted onto the data of a single pixel is shown in Fig. 4.18. It is important to note that the quality of a fit cannot solely be determined by the residual, the specific application has to be considered. The reconstruction will receive integer ToT values from particle hits, the best charge estimation is represented by the mean value for that particular ToT value. The fit function will give the most accurate result if it crosses these points. The simulation will generate particles which deposit a specific charge into a pixel, which then needs to be converted to an integer ToT value to mimic the detector response. Also in this case the fit function will give the best result if it crosses the points in the middle of the plateau and the resulting ToT value for a given charge is rounded to the next integer.

⁴ToT = 14 represents the overflow value. Maximum ToT value depends on the FE-I4 configuration, but the commonly used setting gives 14 values from 1 to 14.

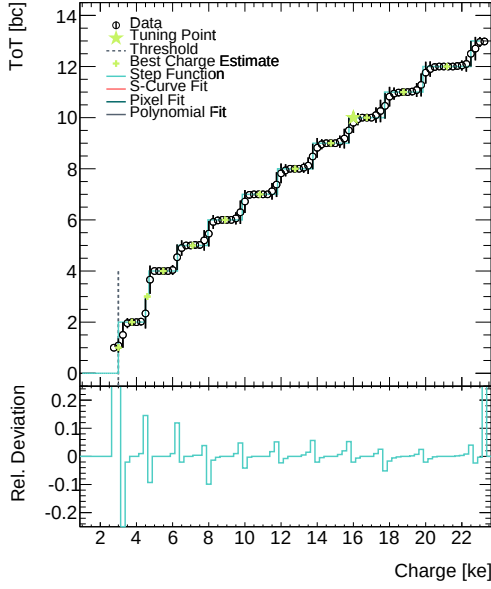
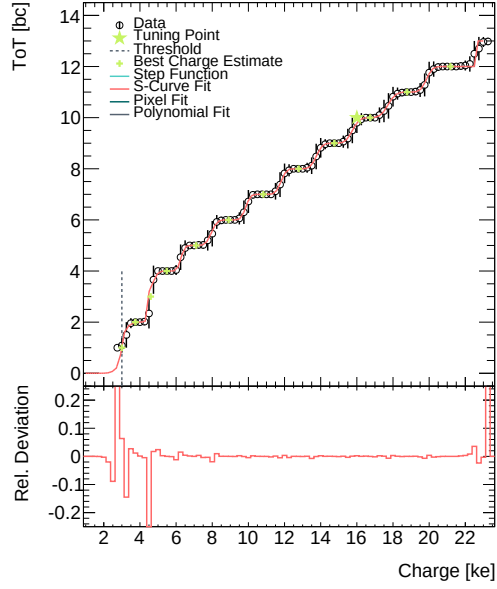
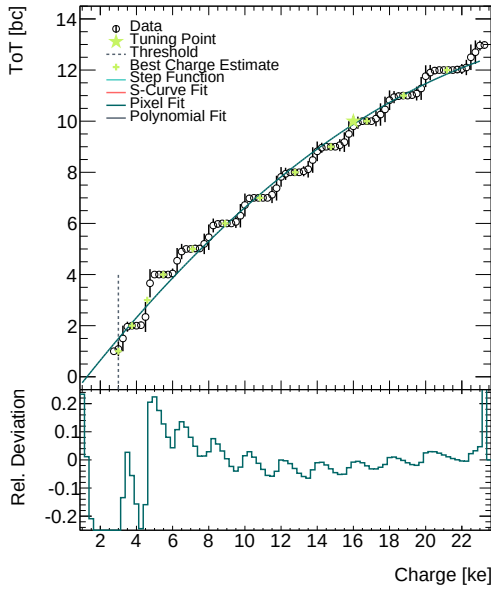
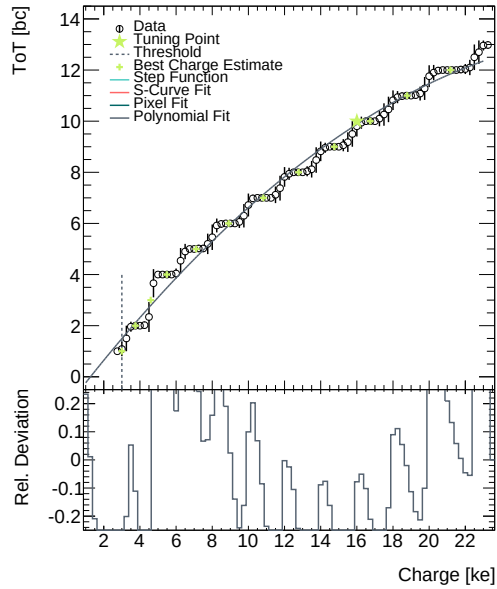

 (a) Step function $f_1(Q)$ (Eq. 4.6.1)

 (b) Smeared Step function $f_2(Q)$ (Eq. 4.6.2)

 (c) Current Pixel detector function $f_3(Q)$ (Eq. 4.6.3)

 (d) 2nd order polynomial $f_4(Q)$ (Eq. 4.6.4)

Figure 4.18: Comparison of different functions fitted to the ToT response of a single pixel.

There are two characteristics which are not modelled by $f_3(Q)$ (Eq. 4.6.3) and $f_4(Q)$ (Eq. 4.6.4): there should be no response below threshold and the response can not be higher than the maximum ToT value. As this is not the case for these functions, they have to be used with boundary conditions. It is also clear that these functions do not model the real pixel behaviour, but rather are approximations which are giving the right answer for the charge conversion. An important factor for the usability of the fit function is also the number of parameters and the modularity (per pixel, module or stave), because the parameters have to be stored in a database and need to be received every time the ToT of a hit is converted to charge. Saving any number of parameters on a per pixel basis, would cost too much performance during offline reconstruction and is therefore not suitable. For the Pixel detector the fit parameters are also stored on a per module basis, with two sets for the two different pixel types⁵. This is not needed for the IBL modules as the difference between normal and long pixels is negligible with respect to the overall spread of all pixels on one module.

Fig. 4.19 shows an example of a single chip with the ToT response of all pixels superimposed. The plateau regions are still visible, but the whole distribution is more continuous than it was for a single pixel alone. The residual charge from the four different fit functions fitted to the data of one chip is shown in Fig. 4.20, the RMS value for y-axis can be used to judge the goodness of the fit and is 793 e , 605 e , 625 e and 626 e for the fit functions $f_1(Q)$, $f_2(Q)$, $f_3(Q)$ and $f_4(Q)$ respectively. I.e. if the functions $f_3(Q)$ and $f_4(Q)$ are used with a boundary condition to include threshold and saturation behaviour, they are as good as the two functions $f_1(Q)$ and $f_2(Q)$ which are closer to modelling the real behaviour of a pixel.

This leads to the conclusion, that if it is not feasible to store a pixel by pixel ToT calibration, the same procedure as for the Pixel detector can be used for IBL, if the results are interpreted with the correct boundary condition. As the ToT calibration of the IBL will not be performed with as many sampling points as this analysis has been done, it is important to note that the test charges should not be equally spaced, but rather more frequent close to the threshold and less frequent for higher ToTs. The ToT to charge conversion of the IBL will never give precise results, which originates from the low ToT resolution of the FE-I4, a feature to cope with the high occupancy so close to the interaction point. For the Pixel detector long and ganged pixels have their own ToT calibration, this is not necessary for the IBL as there is no visible difference between the different pixel types on one module. 3D CNM modules seem to have a slightly different shape, less like a step function more like a continuous distribution, but as the ToT calibration is performed per module, they are separated anyway.

⁵In total there are three different pixel types, but edge and ganged pixels can be combined.

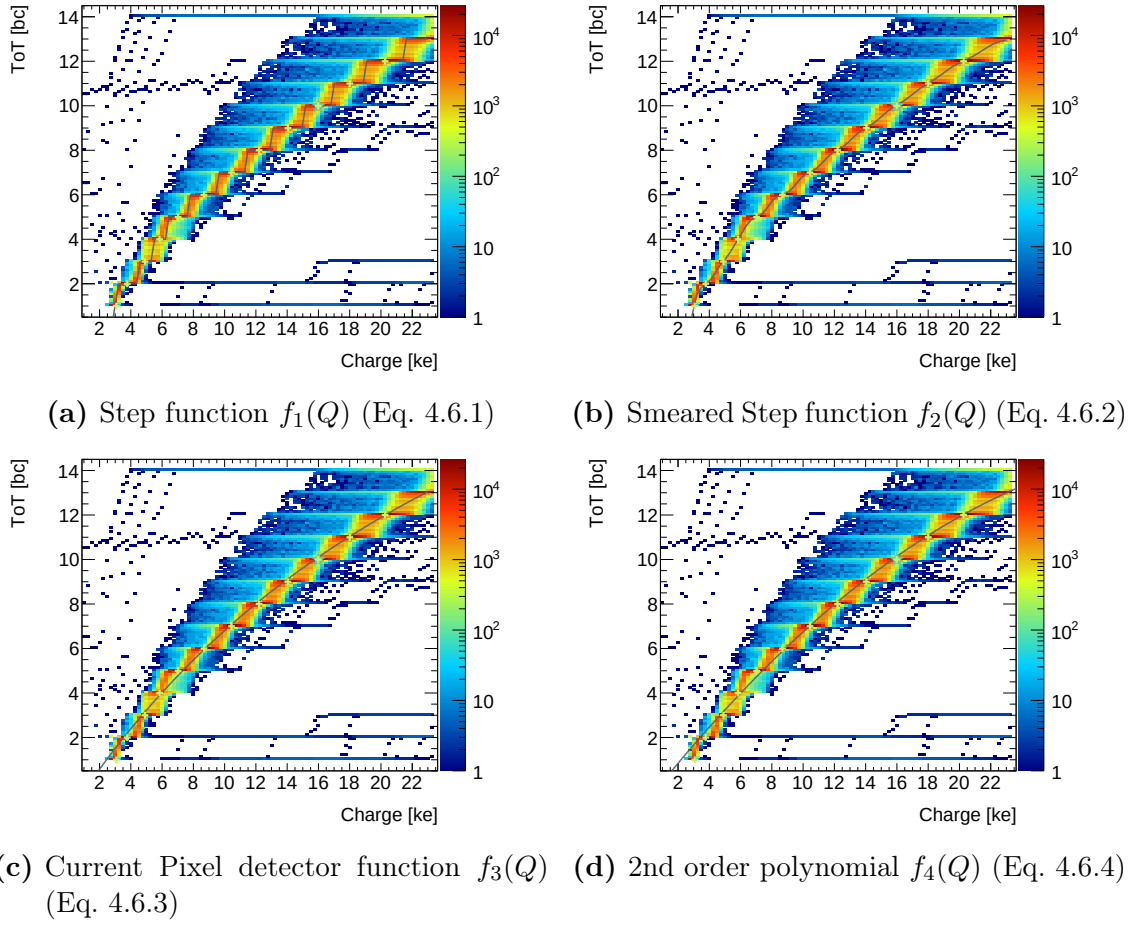


Figure 4.19: Comparison of different functions fitted to the ToT response of one chip.

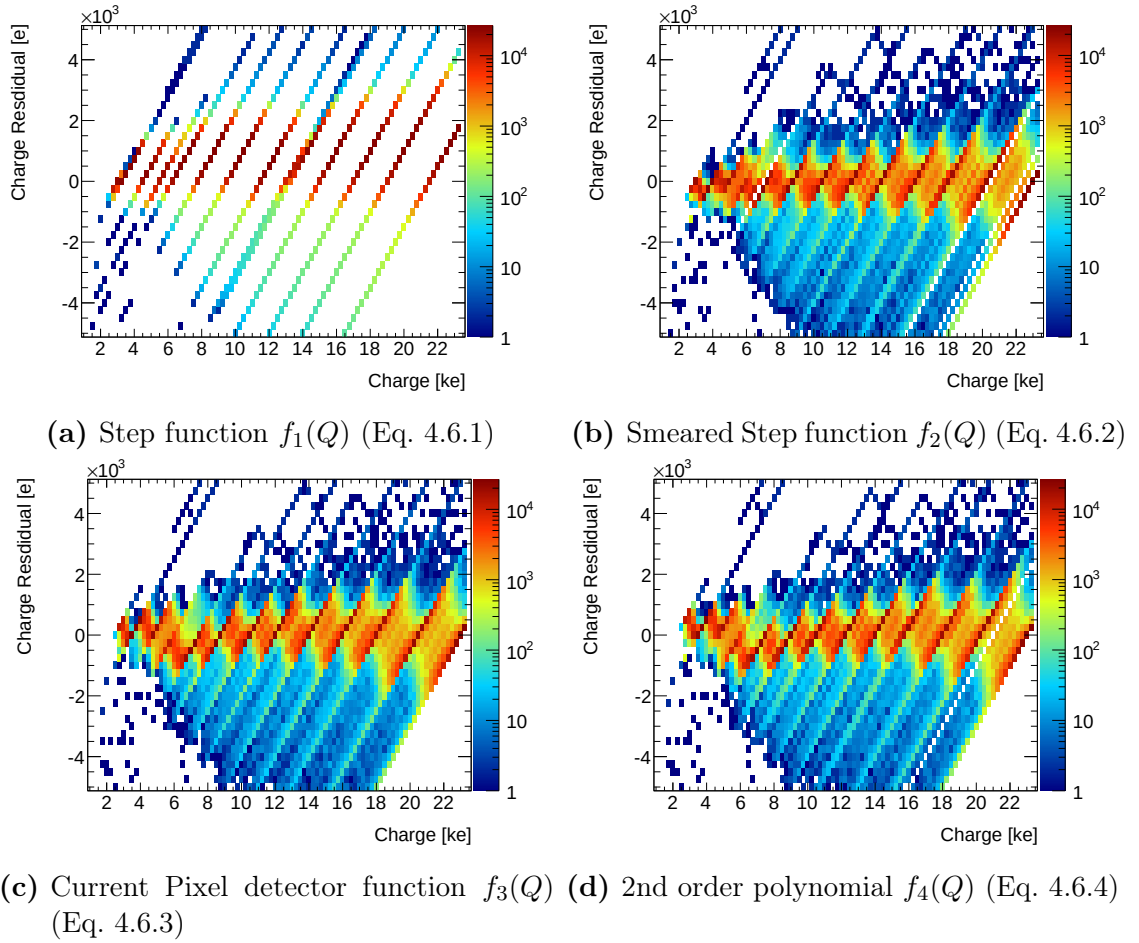


Figure 4.20: Comparison of the charge residual from the data from one chip fitted with four different functions.

4.6.2 Comparison of Pixel Defects during Module and Stave QA

By comparing the bad pixel analysis during the module production QA and stave QA, the modules can be tested for systematic damage which happened between the QAs. For simplicity a bad pixel in the following analysis is defined as a pixel which has seen less than 1% or more than 500% mean occupancy during a Sr^{90} source scan. The correlation of the total number of bad pixel per chip is shown in Fig. 4.21. There tend to be more bad pixels during the production QA than in the stave QA, the reason for this is that more vigorous cuts are made during calibration in the production QA whereby more pixels are masked during the source scan. The general tendency for there not being more bad pixels in the stave QA is a good sign and implies that the modules are not systematically damaged during transportation or stave loading.

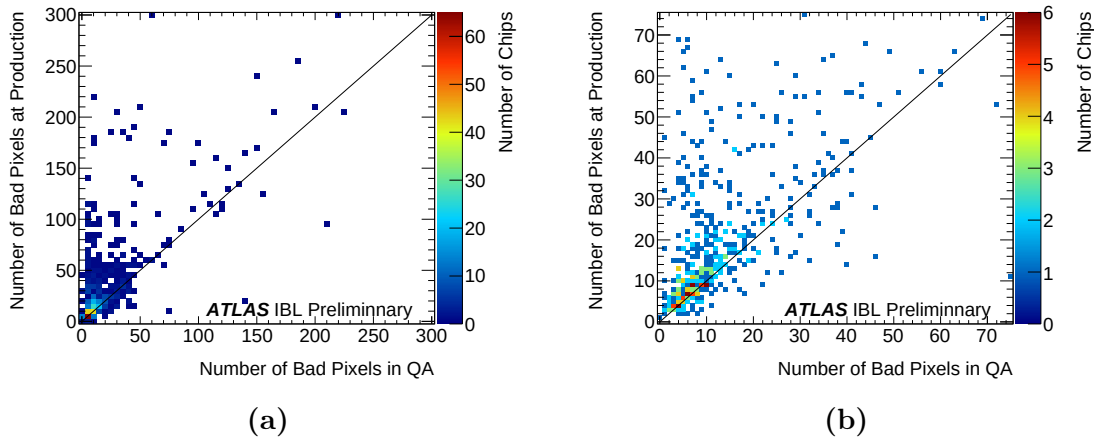


Figure 4.21: (a) correlation of number of bad pixels as identified by the module QA and stave QA, (b) magnified view.

A more detailed analysis compares on a pixel-by-pixel basis, but variations within this, due to the different calibrations and setups between the two QA's, are expected. Fig. 4.22 shows the result of this comparison, for both module types the bin for each pixel was incremented if the pixel failed during stave QA but not during production QA and vice versa. By averaging over all modules statistical fluctuations introduced by the different calibrations and setups should average out and a systematic damage would appear as a cluster of pixels with positive entries. This is not visible with statistical significance, leading to the conclusion that the modules were not systematically damaged during transport or stave loading, which is important feedback for the stave loading procedure.

4.6.3 Increased Noise on 3D FBK Modules

From the observation of increased noise on 3D FBK modules in the stave QA setup, further investigation has shown this to be characteristic of the FBK sensor's sus-

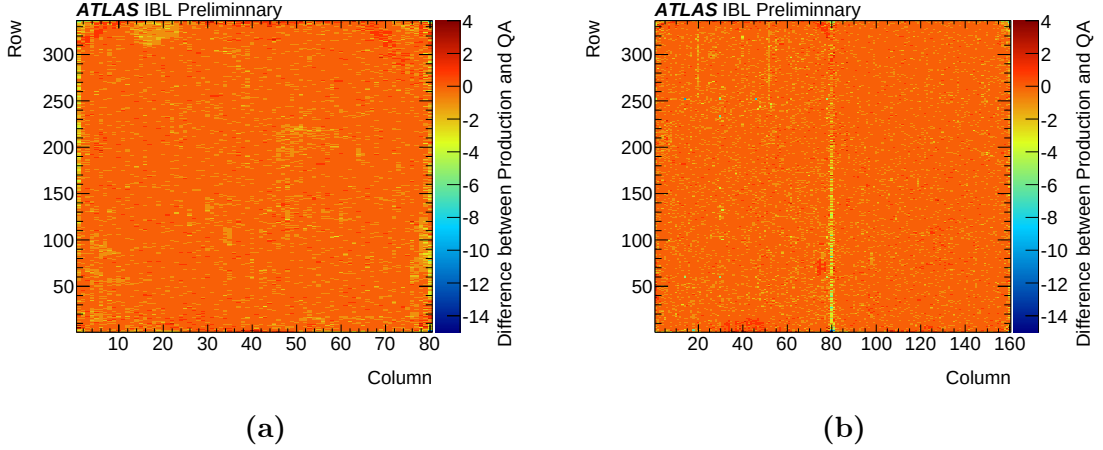


Figure 4.22: For each pixel tagged bad during the stave QA but not during module QA the respective bin is increased and for each pixel tagged bad during the module QA and not during the stave QA the respective bin is decreased. The resulting histogram is shown for (a) 3D and (b) planar modules.

ceptibility to external noise. The increased noise is traced back to one specific HV channel in the stave QA setup, shown in Fig. 4.23, though the peak-to-peak noise ripple voltage is not much higher than on other channels, 100 mV instead of 50 mV on not noisy channels. Operating planar sensor modules or 3D CNM modules with this noisy HV channel did not increase the threshold noise of these types of modules. Furthermore not all 3D FBK modules seem to be affected in the same magnitude, most of the modules show an increase from 150 e to 200 e , while a small number increases up to around 280 e . The specific susceptibility of 3D FBK modules to noise

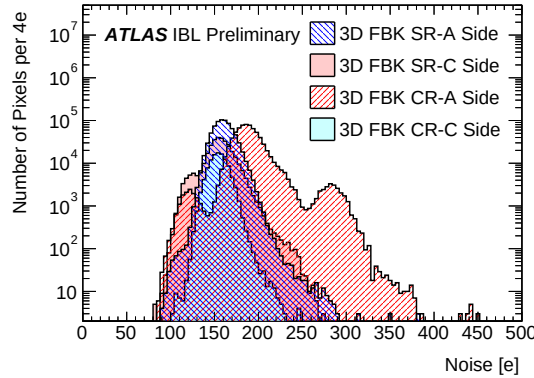


Figure 4.23: Noise distribution of 3D FBK modules for the different ends of a stave and the two different test setup positions, called "SR" and "CR" [44].

is also observed through two other noise sources. The NTC temperature sensor is not electrically connected to the module, it is directly connected through services to the off-detector DCS. To enable a correct temperature measurement, the NTC is sitting on a copper plane which has vias to the bottom side of the flex, thermally coupling it directly to the sensor. The hitbus output of the chip is connected to the

flex and routed out for testing during production, this trace is running on the bottom side of the flex for a specific section. If it is enabled by the configuration, each hit in the chip produces a pulse on the hitbus line. Both the NTC and the hitbus trace, are visible in the threshold noise on 3D FBK modules, shown in Fig. 4.24. The magnitude of the effect once again differed per module and the ones shown here are the more severe cases. The noise induced by the hitbus line can easily be disabled in the configuration of the FE, as the trace is not connected to the services anyway. For the noise induced through the NTC, the DCS side needs to decrease the noise ripple on the temperature sense lines. The reason for this behaviour is not

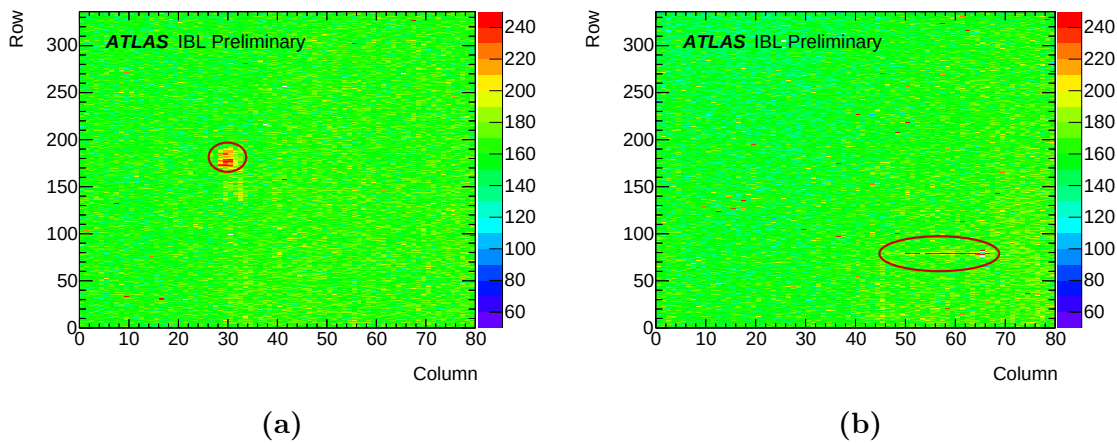


Figure 4.24: Noise injected into an 3D FBK sensor from (a) the NTC and (b) the Hitbus line [44].

fully understood, as it not the same for all FBK modules. In comparison to CNM modules, FBK modules do not have a full metallisation layer on the top side of the sensor (see Fig. 3.7). The metallisation layer could potentially act as shield for the CNM modules and reduce the susceptibility to external noise.

4.6.4 Double Chip Module with one dead Chip

An interesting observation is made on the ST07 after it was damaged by condensation and corrosion. One of the double chip modules on ST07 has one working chip and one dead chip, the threshold noise map of the working chip shows an increasing noise gradient towards the side where the dead chip is located, as shown in Fig. 4.25. Only the chip which is electrically connected to the dead chip via the sensor shows this kind of behaviour, not the neighbouring chip of the next double chip module. Further investigation has not led to a solution to this issue, but it is an important effect to be considered for the long term operation in the detector. At some point in the life time of the IBL one chip of a double chip module can die for various reasons and the shown behaviour has to be expected from the chip on the same double chip module.

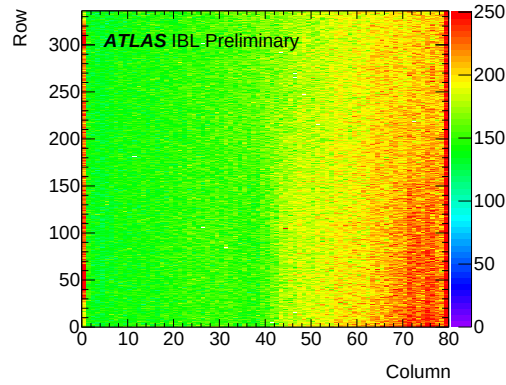


Figure 4.25: Threshold noise map of double chip planar module on ST07 where the neighbouring chip is dead [44].

4.7 Conclusion and Thoughts for Future Pixel Detectors

In conclusion the IBL stave QA has been an essential part of the IBL production and the data which has been taken during the QA is used as a reference point during commissioning and operation of the detector. Although facing a major change in schedule and severe issues, the stave QA has delivered a steady output of data to be used by the selection algorithm. The components which are used in the test stand are very close to the ones used in the detector, making it possible to gather experience of how a stave will behave in the detector prior to the installation of the IBL. One of the most important tools during the QA has been the QA database, which makes it possible to easily store and receive data. The framework, which was build around this database, enables the user to analyse data on-the-fly and store it in the QA database, where it can be viewed in a web browser. The analysis of the QA shows that the IBL staves have an excellent quality and by selecting the 14 best staves out of 18 it is possible to build a detector with more than 99.9% working pixels.

In retrospect, two aspects are not given enough attention during the production QA in the environment of particle detectors. One is the optical inspection, once an electrical component has passed the first functional testing burn-in its is unlikely to fail. Most issues or failures observed in the stave QA, were discovered during the initial optical inspection before the stave was even powered up. The optical inspection in the stave QA was done manually and very tedious, for the future ITk Pixel detector there will be over hundreds of staves and this step should not be done manually. There should rather be an automatic optical inspection, which performs an image analysis to a reference and reports any abnormalities, a very common production step in industry. The second one is the data storage in the stave QA. An effort was made to save all data in a database making it easily accessible. The module production or any other production step of the IBL did not use a database to store data, they relied on spreadsheets and folder structures to organise their

testing results. Hence the results were not easily transferable from one production step to another, making it complicated to compare the results. The database used for the stave QA is much more versatile in this regard and for the production of particle detectors care should be taken beforehand to define the basis of a database used throughout the testing.

Chapter 5

Development of a Novel Readout System for Pixel Detectors

Around 15 years ago when the specifications for the readout system for the current Pixel detector of the ATLAS experiment were set, it was necessary to develop custom hardware to cope with the data bandwidth from the detector. In many regards the requirements set by the Pixel detector upon the readout system were ahead of what off-the-shelf hardware was able to cope with, e.g. a high-end PC of the year 2000 had a 1 GHz single core processor, 256 MB memory and a 100 GB hard disk. It was clearly necessary to use FPGAs and DSPs¹ to process the massive amount of data in parallel, but technology has come a long way since the year 2000 and it is necessary to rethink the concepts of previous readout systems. This chapter will describe the realisation of a novel concept and discuss its performance with respect to the needs of different applications in the field of detector readout systems.

5.1 Overview

The goal of the novel concept, discussed and presented in this section, is to test in which way modern day computers can be utilised to their full extent for detector readout systems. The concept is realised in project YARR² and the choices and design decisions for hardware, firmware, and software will be discussed and presented. The target of project YARR is the control and readout of FE-I4 chips, but the modular design opens up the possibility for implementation of other FE technologies. The advantages of this approach over currently existing systems and the prospect for usage with future FE technology will be analysed.

¹Digital Signal Processor

²Yet Another Rapid Readout

5.2 Motivation

The IBL readout system (as described in Sec. 3.5) is highly complex, with multiple master processor units communicating with multiple slave processors. Most of the functionality is contained inside the firmware and software running on FPGAs located on the ROD and BOC cards. The computer controlling the cards is performing tasks which mainly consist of managing configurations and downloading scan loops to the scan engine in the ROD. Most of the processing power is contained in the ROD and an external computer farm, which is only used to perform the fitting of histogram data. The high complexity is partly due to backwards compatibility to the Pixel detector readout system, but also a necessity to give the ROD the capability to perform scans standalone. As such there is a very steep learning curve for new developers, as both firmware and software elements need to be understood in great detail. Effectively, less work can be done because of this, as the time to get started with the system reduces that available for development. Also, the strong interconnection of firmware and software hinders development in the long run, as both fields have to wait until the respective counterpart is implemented.

How can these challenges be tackled to ease the development of future readout systems? A possible solution is to decouple firmware and software and move as much intelligence as possible into software: a much more accessible regime for developers. In what follows, the question of if modern computer systems can cope with the requirements set by particle detectors, will be explored.

Different considerations must be taken into account for the design of a readout system and are not solely driven by the requirements set by the FE to be read out. There is no way to avoid the usage of FPGAs as the first interface to the detector FE, their flexibility and performance in parallel digital signal processing is not matched by any other type of hardware. Hence the primary interface to the detector is based on an FPGA, but the way in which the FPGA will be utilised may differ to previous systems.

5.2.1 Architecture

Besides chip specific requirements, which can be deduced from Sec. 3.4, a readout system has to handle two different states: operation and calibration. The main task during operation is to send a clock synchronous to the global system clock and trigger commands with a deterministic timing to the chip. The data which is read back is checked for errors e.g. desynchronisation of a chip, framed in specific format and sent to a higher level readout system. Data loss due to buffer shortage has to be reported via busy logic. All these tasks can be performed inside an FPGA, which can deal with all data streams in parallel. On the other hand, the purpose of calibration is to find the best parameter set for a given target of threshold and ToT conversion. Parameters are tested via repeatedly injecting test charges and analysing the response. This requires an intricate set of commands to be executed in a specific order and timing and the configuration of the chip to be altered in

response to the data received. The implementation of these scans is best to be done in software, allowing for higher modularity and complexity. Thus a link is required to ship data from the FPGA to a processor running the analysis. In the past this link was too slow to ship all the data, so the data was compressed in the form of histograms. The histograms contain enough information with respect to the target of the scan in which they are created.

If the link between FPGA and processor is fast enough, all aggregated data can be shipped to the processor and the analysis is not bound to use only histogrammed data. For many scans histograms contain enough information, but most scans are developed with this limitation in mind. Having access to the raw data stream would, for example, make it possible to determine the threshold and in-time threshold in a single threshold scan. This new architecture concept is depicted in Fig. 5.1, showing the difference between the current concept used for the Pixel detector and the concept which will be established in this chapter. The changes seem minor, two blocks have been moved from the FPGA to the CPU, but the implications of this change are major from a developer perspective.

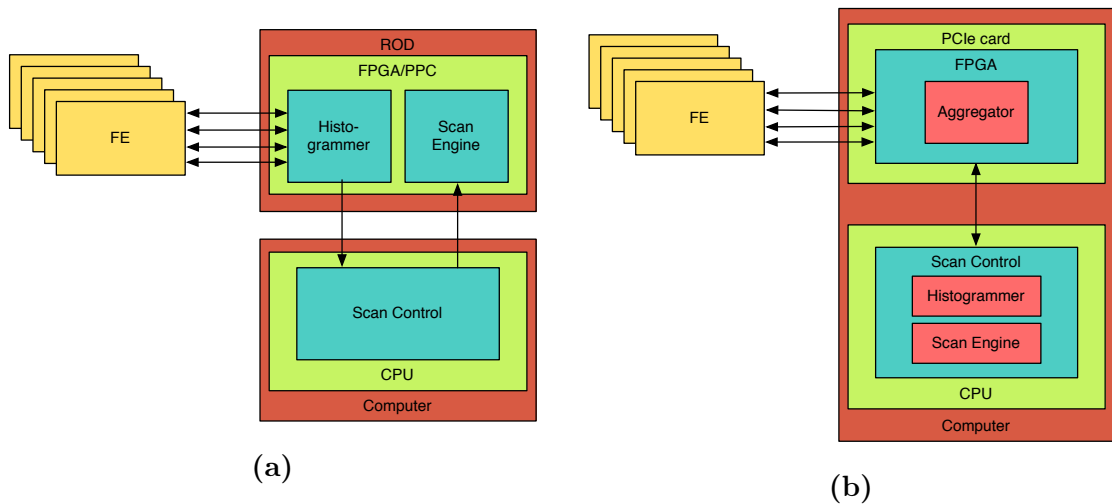


Figure 5.1: Comparison of current readout concept (a) as used for the Pixel detector and (b) the new concept.

5.2.2 Front-End

The type of FE which is used for proof of concept, the FE-I4B, is a state-of-the-art sensor readout chip and is optimised for usage in the IBL, but the general architecture from a readout system perspective is similar to many other chips, e.g. FE-I3, Timepix3³ [3] or PSI46⁴ [17]; they have a matrix of pixels, each having an analog and digital region with adjustable threshold and shaper settings. Serial links are used for communication to transmit clock, commands, and data output. If most

³To be used for the LHCb Velo upgrade.

⁴Used for the CMS Pixel detector.

of the functionality of the readout is done in software, a high degree of flexibility can be achieved with respect to the FE. The specific implementations might differ, but the common software infrastructure can be FE independent.

Scope of Application

Readout systems for particle detectors are used in three different environments: laboratory, testbeam, and the detector. Each of these environments require different attributes from the readout system and the key ones are listed in Table 5.1.

Table 5.1: List of requirements and how they apply to different application scenarios. $\checkmark$ means important and necessary, $\sim$ applicable to some degree and $\times$ not applicable or important.

Requirement	Laboratory	Testbeam	Detector
Accessibility	$\checkmark$	$\times$	$\times$
Flexibility	$\checkmark$	$\sim$	$\times$
Cost	$\checkmark$	$\sim$	$\times$
Multi Module	$\sim$	$\checkmark$	$\checkmark$
Timing	$\times$	$\sim$	$\checkmark$
Scalability	$\times$	$\times$	$\checkmark$

In a laboratory environment, typically small scale systems are encountered and it is very important that these kinds of systems are accessible for new users, are of low cost so many of them can be bought without hesitation and, in general it is a useful addition, if they can be used for more than one type of FE. USBpix [18] is an example of such a system, though it lacks the possibility to operate multiple FE-I4s at one time, which comes in handy for the development of multi-chip modules or if multiple assemblies are to be characterised at the same time. Calibration is generally more important than operation, as there are limited possibilities to operate an FE in lab., e.g. radioactive sources or cosmic rays.

In a testbeam there are two main components: the telescope (multiple FEs are arranged in planes) and the device under test. The FEs are usually chips which are already in use in other experiments, e.g. the FE-I4 or Timepix, but in comparison to the laboratory environments the operation is more important than the calibration. The higher the sustainable trigger rate during operation is, the more data can be taken during a testbeam campaign, which results in higher statistics and more parameters to be tested. It is also advantageous if the telescope and the device under test can be operated by the same readout system, which requires a certain degree of flexibility as the devices under test are often early prototypes with only rudimentary feature sets.

In a full detector the requirements shift very radically to scalability and timing. Many systems have to operate synchronously and with high reliability. As the FE are optimised for the particle hit rate which is expected during detector operation, the readout system has to be able to cope with the full data output bandwidth of

the FE. The RCE system, for example, performs very well in the laboratory and the testbeam, but as it was designed to be used in the detector, the cost to build one system for these uses is very high.

5.2.3 Developments for Future Detectors

The development of readout systems is driven by the development of new FEs, which are increasing in pixel density and data output bandwidth. This is a requirement to cope with the high particle densities to be expected after the LHC upgrade. The serial link bandwidth will determine the hardware decision: the I/Os of a low cost Xilinx Spartan 6 FPGA [49] can be used up to 1088 Mbit/s [47], I/Os of a high-end Xilinx Virtex 7 FPGA [50] can be operated up to 1.25 Gbit/s [51]. If higher transfer rates are needed, specific gigabit transceivers should be used, which only exist in a limited number on the FPGA. The firmware will have to be adjusted but, if it is possible to process the data stream in a CPU, this will not change that much for future detectors.

In the past, the effort put into laboratory systems for characterising and testing the detector was separated from the development performed for the operation of the full detector. Especially for the IBL and its short timescale, this led to the DAQ not being ready when the detector was being tested nor when it was installed in the ATLAS experiment. Part of the reason is the high complexity and coupling of firmware and software, but a major drawback is that a large amount of knowledge was contained in the laboratory readout system, which was not documented and not easy to transfer into the IBL DAQ. A readout system aimed at, for example ITk modules, for laboratory applications and testbeams, should have a rigid and scalable software framework which renders it possible to use the same software cores in the laboratory and for operation in the detector. Laboratories will most certainly not use the hardware which will be deployed for operation, as a consequence of this the software should be independent of the hardware or at least modular enough that certain parts of the software can be exchanged easily.

If it is possible to build such a system, its development will not only occur after installation and in the detector, but much earlier in a multitude of environments. This will not only make the system more stable, but will also build an experienced user base, who are all able to help with development of the detector DAQ whilst they are improving the system in their own lab. Another advantage of using the same system in the laboratory is that the laboratory is a controlled environment and perfect for testing patches to the code. In general this concept might result in a longer initial design period, but the advantages from it will outweigh this in the long run.

5.3 Project YARR

The goal of this project is the PCI express (PCIe) [37] based readout of an FE-I4B module, with emphasis on keeping the firmware simple and moving most functional-

ity into software. In the following the different building blocks - hardware, firmware and software - are discussed. The results of the first demonstrator setup are presented and compared to already existing FE-I4 readout systems.

5.3.1 Hardware

The SPEC⁵ board, as shown in the picture in Fig. 5.2, was chosen for multiple reasons over other solutions:

- The Xilinx Spartan 6 FPGA is a cheap but powerful FPGA and has already been used on the IBL BOC to interface FE-I4 chips.
- The card has been developed by the CERN BE department and is used in the LHC, i.e. hardware experts are readily available.
- It can be bought from three⁶ different commercial producers, making it easy for users to buy one and reducing the price.
- The Gennum GN4124 is a PCIe local bus bridge and eases the usage of PCIe with an FPGA and has advantages during prototyping, like FPGA reconfiguration via PCIe.

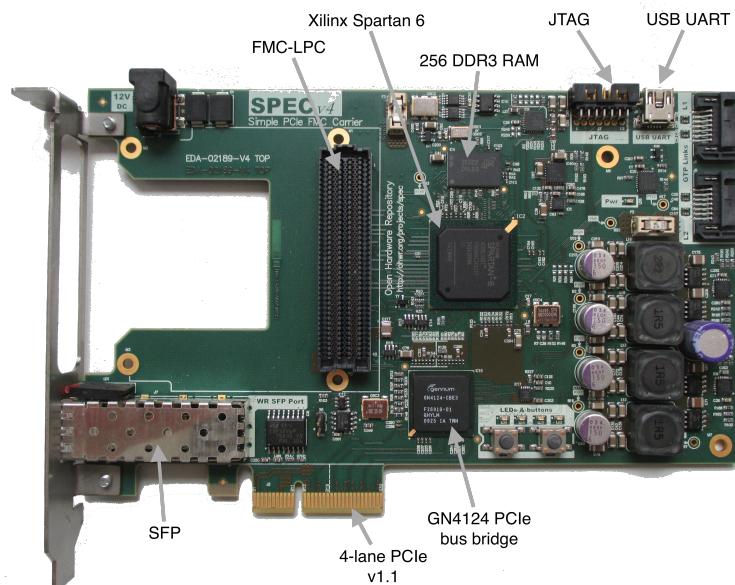


Figure 5.2: Picture of the SPEC board, pointing out the main components.

⁵Simple PCIe Carrier: <http://www.ohwr.org/projects/spec/wiki>

⁶At the time of writing.

Not all features of the board are used, the most important components are: the Xilinx Spartan 6 FPGA (LX45T), the 256 MB DDR3⁷ memory, the FMC-LPC⁸ connector, and the PCIe bus bridge (Gennum GN4124 [25]). The PCIe bus bridge uses up to 4 PCIe lanes with PCIe v.1.1, which has a bandwidth of 250MB/s per lane, but is specified up to 800MB/s. The card is plugged into an off-the-shelf computer with the following components: an Intel Core i5 3570K processor, an Asus P8Z77-V LX (Intel Z77 chipset) motherboard, and 2×8 GB DDR3-1333 (CL9-9-9-24) memory. All tests are performed with a standard Scientific Linux CERN 6⁹ installation.

5.3.2 Firmware

A block diagram of the YARR firmware is shown in Fig. 5.3, the different blocks are implemented around a single master Wishbone[31] bus, which is used to control multiple slave blocks. A second Direct Memory Access (DMA) Wishbone bus is used exclusively for data transfer in and out of the memory. Most blocks are of generic functionality and not specialised on the usage with FE-I4, except the TxCore and RxCore which contain functionality specifically for FE-I4 chips, e.g. a serialiser and 8b10b[2] decoder. As the TxCore and RxCore are connected to the bus as a Wishbone slave, it is easily possible to exchange these blocks or add more to communicate with a different kind of FE. The Wishbone bus and all masters and slaves attached to it are operated with a system clock of 200 MHz; only the FE-I4 serialiser and deserialiser are operated at 40 MHz and 160 MHz respectively. As the Wishbone bus is 32 bits wide, the maximum bandwidth is 800 MB/s with a 200 MHz clock, which is also the maximum bandwidth of the GN4124. A higher bandwidth in the FPGA can be achieved by either increasing the frequency of the system clock or increasing the bus width, but both need to be taken into account in the timing simulation and can make the design violate the timing constraints.

GN4124 Core and DMA Controller

The GN4124 core implements the interface to the GN4124 local PCIe bus bridge as a Wishbone master. It is based on an existing design, but optimised for the transfer of data via DMA with high bandwidth¹⁰. A block diagram of the GN4124 core is shown in Fig. 5.4, with two Wishbone bus interfaces, one a pipelined¹¹ bus used for DMA and the other a standard bus for control. The GN4124 has two 16 bit DDR

⁷Double Data Rate signals use both clock edges for sampling, hence increasing the data rate by factor two.

⁸Abbreviation for: FPGA Mezzanine Connector - Low Pin Count. An interconnect standardised under Vita 57.

⁹<http://linux.web.cern.ch/linux/scientific6/>

¹⁰The changes and optimisation are to be merged with the original GN4124-core project on <http://www.ohwr.org/projects/gn4124-core/wiki>.

¹¹Pipelined means that a series of read or write requests can be executed before the request is acknowledged. I.e. that a pipelined interface is capable of block transfers, which increase the bandwidth with respect to single transfers.

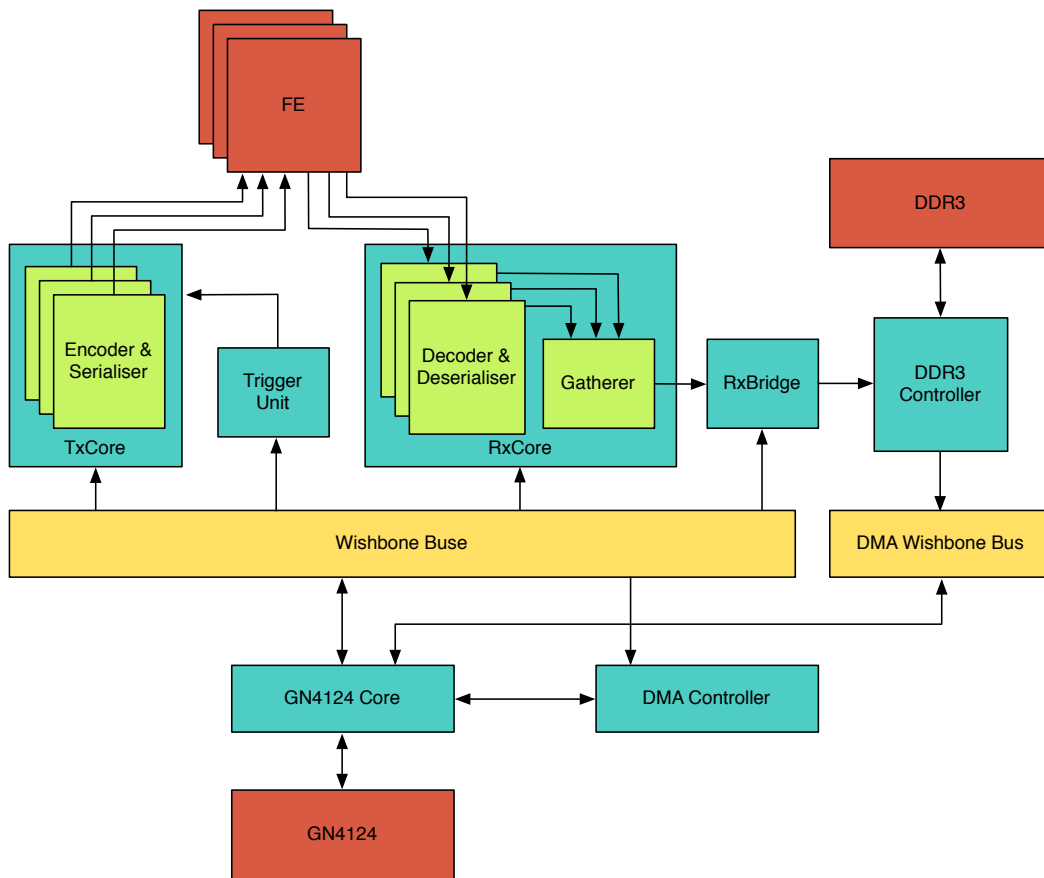


Figure 5.3: Block diagram of the YARR firmware. Red blocks depict interfaces to hardware, yellow blocks the communication busses, blue blocks are the main common firmware blocks and green blocks are FE specific.

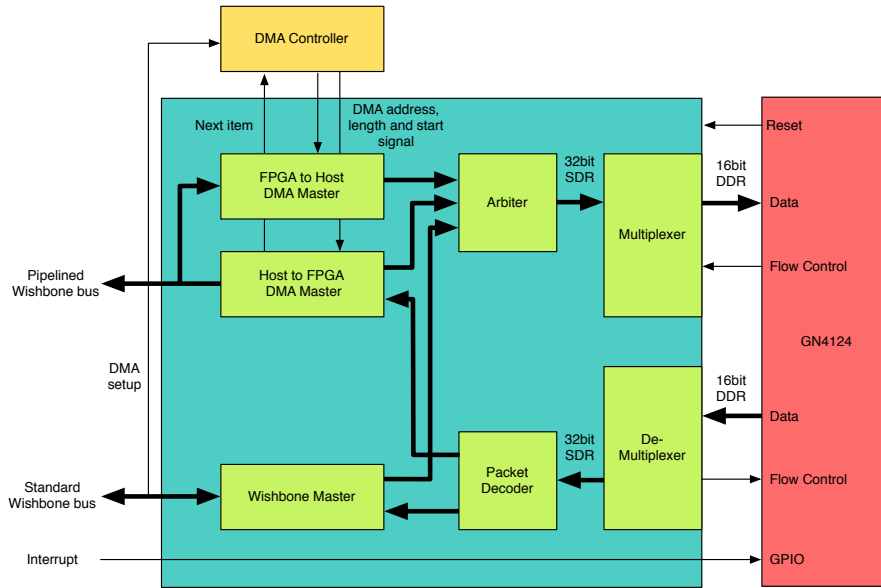


Figure 5.4: Block diagram of the GN4124 core. The standard Wishbone master interface is controlled via the host computer. During a DMA transfer the GN4124 core acts as the master and steers the DMA transfer on both ends. The data path is shown with bold arrows and control signals with thin arrows.

ports which are operated with a custom data protocol consisting of a 32 bit header, and 64 bit address for a read transaction and for a write transaction additionally up to 16 32 bit data packets, hence each transaction can carry up to 512 bits of data. The different transaction types can be initiated from both sides, i.e. the host computer can read/write from/to the FPGA and vice versa. When using the standard Wishbone bus the host computer is the master, but when a DMA transfer is started the FPGA takes control of the transfers. A normal PCIe transaction can be directly translated to a single wishbone read/write cycle, if multiple data packets are transferred or requested, they are read from/written to the FPGA in single cycles. The multiplexer and demultiplexer contain buffer FIFOs¹² which absorb data pressure in both directions.

In DMA, data is directly transferred between a source and a destination memory, without any other buffering of the data. In a normal transfer the data is typically buffered by the CPU or some other processor. Bypassing this buffering increases the performance of the transfer. As the CPU does not take part in a DMA transfer, the host memory needs to be prepared: modern operating systems use virtual memory, a memory management where a contiguous virtual memory space does not necessarily need to be contiguous in physical memory. This scheme is called paging and the smallest memory region which is contiguous in virtual and physical memory is called a page and typically has a size of 4 kB. If a contiguous space of memory is allocated on a user level, this memory can be scattered over the physical memory.

¹²First In, First Out: A commonly used memory structure for buffering and pipelining.

As the FPGA only has access to physical memory and does not know of the virtual memory mapping, the CPU needs to instruct the FPGA where in physical memory to write/read each page to have read a contiguous space in virtual memory. This procedure is called scatter/gather DMA and for this the FPGA needs to hold a list of addresses of each page (scatter/gather list), this is shown in Fig. 5.5. In this specific implementation the DMA controller is configured with a linked list consisting of seven 32 bit words, which are listed in Tab. 5.2. Each list item contains the information for one block transfer of a given length and a pointer to the next item. After a transfer has been completed and if there is more data to be transferred, the next item is read from host memory and the next transfer is initiated. When the last item has been processed the CPU is informed via an interrupt that the DMA transfer has been completed.

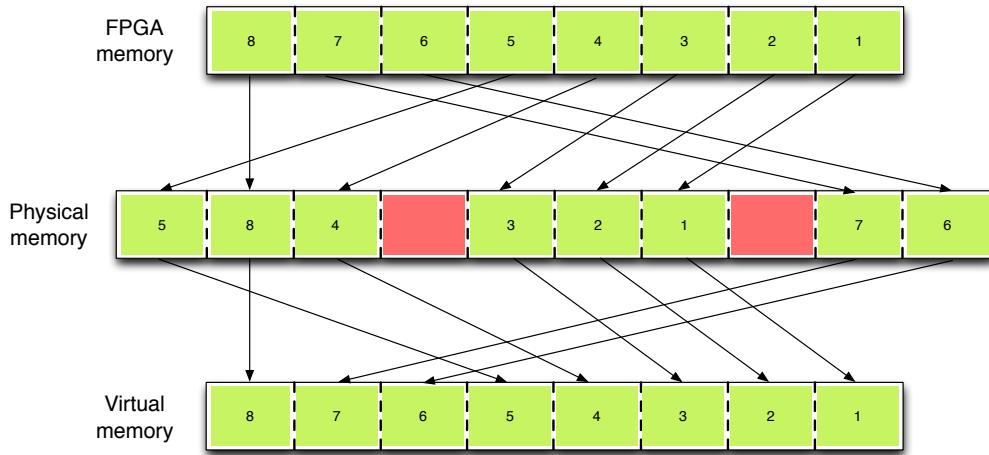


Figure 5.5: Schematic of how physical memory is mapped into virtual memory. The memory is segmented in pages and the representation in physical memory is not necessarily contiguous and can also contain pages owned by other processes (red blocks).

Table 5.2: Items of a DMA linked list.

#	Type	Function
0	Address[31:0]	FPGA memory start address
1	Address[31:0]	Host memory start address
2	Address[63:32]	Host memory start address
3	Length[31:0]	Number of words to transfer
4	Address[31:0]	Host memory address of next list
5	Address[63:32]	Host memory address of next list
6	Attribute[31:0]	Type of transaction (read/write) and last item

The creation of the linked list, setup of the memory for DMA and configuration of the DMA controller produce an overhead to the actual amount of data that will

be transferred. i.e. that a DMA transfer is only worthwhile when large data packets are transferred. Another observation with this particular implementation is that each block transfer of a linked list is typically the size of one page (4 kB), i.e. the total amount of data is transferred in chunks of 4 kB. After each chunk the DMA controller needs to request the next item from the host memory, the time it needs to receive the next item is roughly equal to the time to transfer one 4 kB chunk, hence only around 50% of the bandwidth is actually used (shown in Fig. 5.6). To use the full bandwidth the full linked list has to be stored in the FPGA memory (reducing the list by 2 words), assuming the transfer of up 500 kB packages: 2.5 kB memory is needed to store the linked list. This is a possible solution in case the bandwidth is not sufficient anymore, which is not the case in the current stage of development and there might be a bottleneck earlier in a different firmware block.

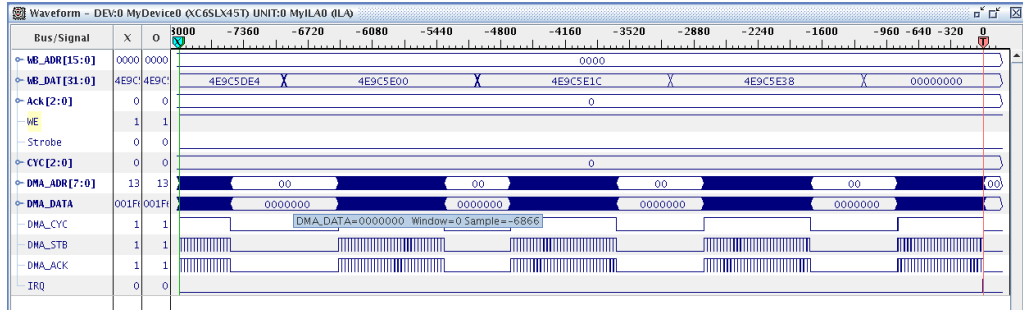


Figure 5.6: Measurement of the DMA Wishbone bus with an FPGA internal logic analyser. The pauses between the data transfers (blue regions) are the time until the next list item has been received.

DDR3 Controller

To interface the DDR3 memory on the SPEC board, a memory controller is used based on an existing design¹³ and optimised for performance. The memory controller is a wrapper around a Xilinx Spartan 6 memory controller IP core[48] to two pipelined Wishbone bus slave interfaces. A block diagram of the controller is shown in Fig. 5.7 and most of the complex functionality like calibration is contained in the memory controller from Xilinx. DDR memory is Direct Random Access Memory (DRAM) and due to its internal construction there is a delay between setting an address and receiving the data, hence DRAM is slow when random addresses are accessed. However the impact of the delay can be minimised by transferring many consecutive data words, called a burst transfer. A Wishbone bus block transfer is translated by the wrapper to a command which instructs the Xilinx memory controller to read or write a specific number of words to a given start address. If a write command is issued the data has to be already written to the data port and if a read command is issued the data will appear after a short delay on the data port. The maximum length of a burst transfer is configured to be 16×32 bit words,

¹³<http://www.ohwr.org/projects/ddr3-sp6-core/wiki>

i.e. two burst transfer are needed to produce on PCIe package in the GN4124 core, making the transfer from DDR3 to PCIe very efficient. The peak bandwidth of the DDR3 memory, when operated with 333 MHz and a 16 bit wide data bus, is 1.3 GB/s, but this calculation completely ignores the latency and therefore the actual bandwidth will be lower. Even though there are two Wishbone slave interfaces,

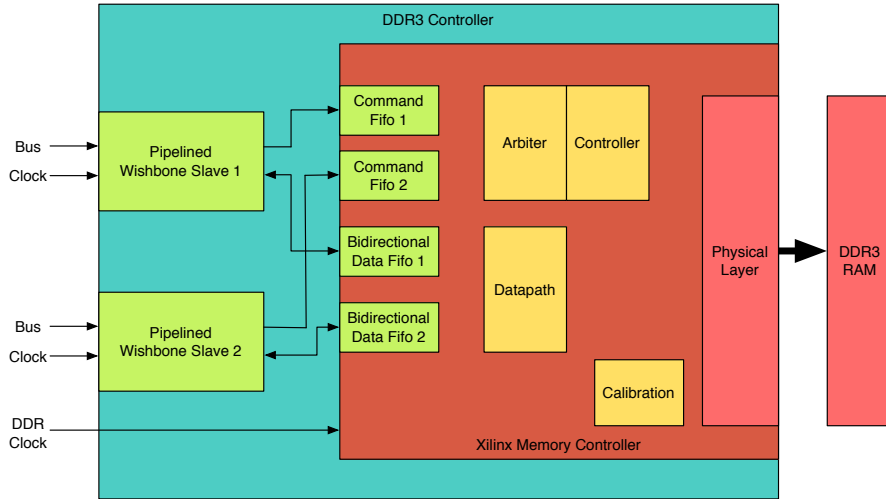


Figure 5.7: Block diagram of the DDR3 controller showing the wrapper around a Xilinx memory controller.

the DDR3 memory is not a true dual port memory. An arbiter¹⁴ will swap, after each command, in a round robin¹⁵ manner between the ports, which will impact the performance as the memory is used in a FIFO manner, with one port writing data consecutively and the other port reading it consecutively. The pointer to the memory location where the next word will be written is always in front and different to the read pointer (except when the memory is empty), i.e. when the arbiter is switching between the write and read port the latency of the data has to be added to nearly every transfer. The DDR3 memory used on the SPEC board can perform 666 million 16 bit transfers/s and has a latency of 13.5 ns, if the latency is applicable to every full burst transfer the effective data rate is 1040 MB/s. As both ports will use the memory equally, each port can use up to 520 MB/s. In this calculation it is assumed that always full burst transfers are performed, if less than 16 words are transferred the performance can be as low as 240 MB/s for single word transfers. In the actual implementation the reading is performed in large chunks (DMA transfer) while the writing depends on the data coming from the FE.

TxCORE

The TxCore is responsible for sending clock and commands to the FE-I4 and, in general terms, can be described as a serialiser. Fig 5.8 shows the block diagram of

¹⁴A combinatorial selection mechanism controlling the multiplexer.

¹⁵No input link is prioritised over the other.

the TxCore. It is interfaced via a single Wishbone slave interface and serialises data with a second clock, which, in the case of an FE-I4, is 40 MHz. The clock domain crossing is carried out by a FIFO and commands which are to be serialised are simply written into it. The Wishbone slave interface is not pipelined, but commands usually don't have to be written with high bandwidth, as long as they are written faster into the FIFO than they are serialised, e.g. in the case of an FE-I4, commands have to be written with at least 5 MB/s to avoid a FIFO underrun¹⁶. For calibration of an

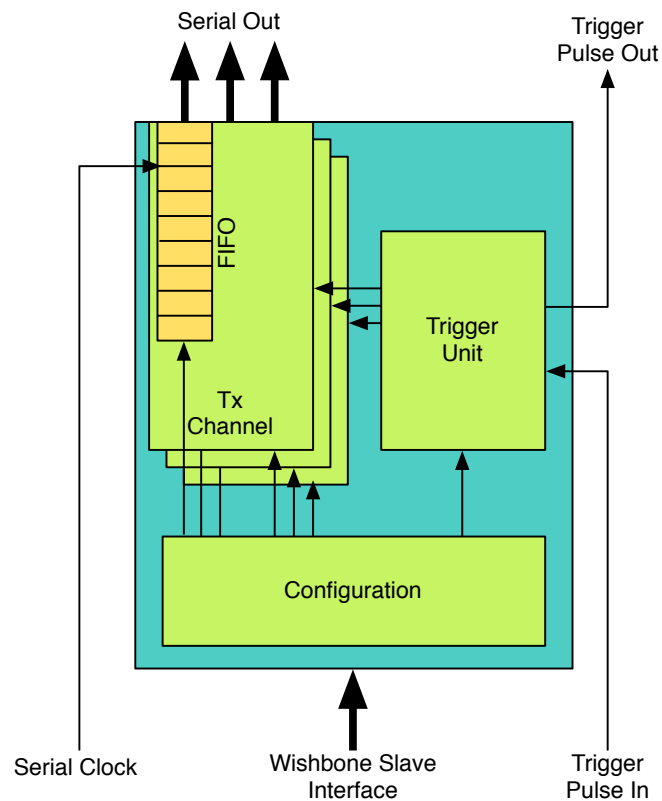


Figure 5.8: Block diagram of the TxCore, which sends clock and commands in a serial stream to the FE. The number of output channels, 3 in this example, is scalable.

FE-I4 it is necessary to repeatedly send the same command, e.g. an injection and trigger command with a specific fixed delay between the two. For this purpose the TxCore can be configured with a fixed command pattern of up to 256 bits, which on reception of a trigger pulse is loaded into the FIFO and sent out. The trigger pulse is generated by the trigger unit, which can be configured in different modes:

- Fixed time: pulses are generated for a fixed amount of time with a variable frequency but are not limited to a fixed set of possible times. The fixed amount can be set at configuration.

¹⁶Reading from a FIFO although it is already empty.

- Fixed count: a variable number of pulses is generated with a variable frequency.
- External: a pulse is generated for each rising edge received on a specific signal.
- Pseudo random: pulses are generated pseudo randomly with the mean frequency being variable.

Times and frequencies are measured in multiples of the serial output clock frequency, the time is stored in a 64 bit register setting the maximum time higher than multiple years. The fixed time, count, and pseudo random modes are important for the calibration, while the external trigger system is mainly used for operation.

RxCore

The RxCore is responsible for receiving the data from the FE and preparing it for storage in memory. The implementation which is shown in Fig. 5.9 is FE-I4 specific, as it is designed to receive a 160 MHz serial data stream which is 8b10b encoded. The phase alignment to the data is done automatically and deserialisation synchronises itself to the comma symbols¹⁷ of the 8b10b data. After the data has been decoded, it is stripped of K-words and three bytes are collected to form a 24 bit data word, which is stored together with an 8 bit channel identifier in a 32 bit word, as seen in Fig. 5.10. These 24 bit data words represent the smallest unit the FE-I4 can send out. Each framed word is saved in a FIFO, which is also the clock domain crossing from the receiver clock to the system clock domain, where an arbiter multiplexes the data in a round-robin fashion into the next logical block. The input data rate is 16 MB/s and, when including the 33% overhead when the data is framed with the channel, it is 21.3 MB/s, i.e. the system can read the data from up to 37 channels without the FIFOs filling up. As this number is higher than the possible number of I/Os which could be used to connect FE-I4s, no busy logic is deployed in this block.

RxBridge

The RxBridge is the element which connects the RxCore and the DDR3 memory controller. As it is shown in the block diagram in Fig. 5.11, the data coming from the RxCore is passed into a FIFO which is used to buffer the data until there is enough such that it can be transferred in an optimal manner via a Wishbone master interface, which is connected to one port of the DDR3 Controller. The back pressure logic tries to collect data of the size of at least one burst transfer, but if not enough data is received in a certain time frame (10 μ s) it empties the FIFO. The packet builder counts the number of words which have been transferred into memory and will store the start address and word count when the packet gets large enough. As the data will be read via DMA out of the memory, the minimum packet size has to be optimised to outweigh the effect from the overhead of the DMA transfer. The packet is chosen to be 250 kB for reasons discussed in Sec. 5.3.5 and a packet is always built after 0.1 ms. The different timeouts and data sizes are parameters

¹⁷The 8b10b encoding allows for special 10 bit words, comma symbols, which can not be translated into 8 bit data words, commonly used to control the link.

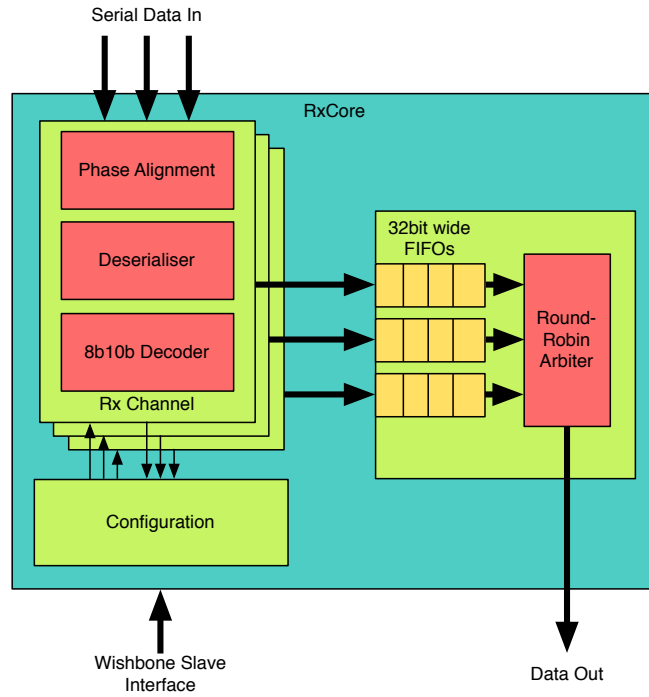


Figure 5.9: Block diagram of the RxCore, each channel is deserialised, decoded, and buffered in a FIFO. From there a round robin arbiter multiplexes the data into the following logic. The number of input channels is variable, here shown for 3.

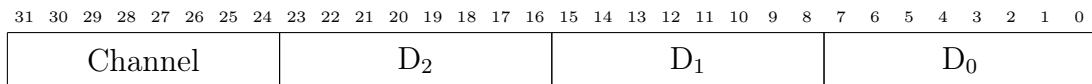


Figure 5.10: Composition of framed FE-I4 data in a 32 bit word, showing the encoding of the channel identifier and the three bytes of data (D₀, D₁, D₂).

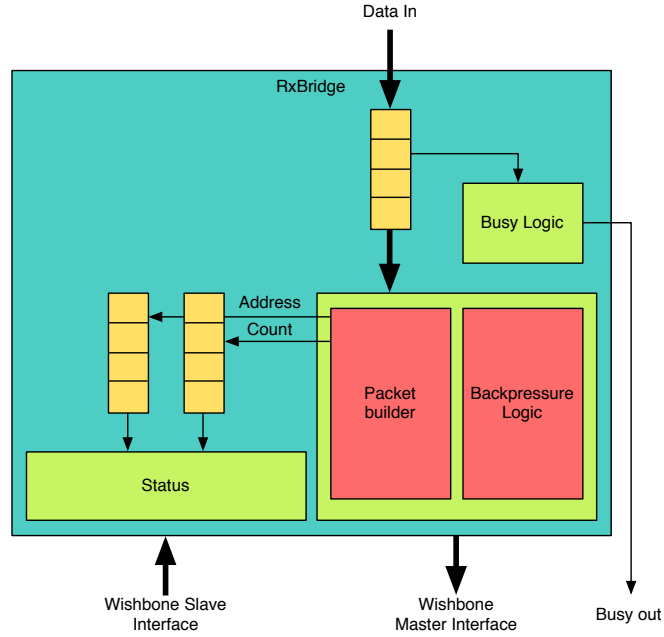


Figure 5.11: Block diagram of the RxBridge, which receives the data from the RxCore and builds packages which are written to the memory.

which have to be carefully tuned for the specific implementation, as it has been shown that the bottleneck of the system is writing and reading from the memory simultaneously. Writing and reading into the memory should be bunched in fewer but larger transfers, rather than many small transfers. This way the performance of the memory can be increased and the system should be able to transfer as much raw data into the host, as it is possible with the performance of the PCIe link. The main limitation in how many chips can be read out simultaneously is the number of I/Os, therefore the firmware was designed with the read out of 16 FE-I4s in mind.

5.3.3 Software

The YARR software needs three core pieces to perform the readout: a kernel driver to communicate with the firmware via PCIe, an engine driving the scans, and processors which analyse the received data. Moving the scan engine and data processing into software has certain advantages: the scan engine is not limited in its structure anymore and the data processing can provide the analysis with more information than only a certain number of histograms. The main concept of the software is data driven, rather than being driven by functionality, i.e. the software is interlinked by a buffer structure which hands data from one processor to another.

Kernel Driver

A custom Linux kernel device driver [16] is used in conjunction with the SPEC board and YARR firmware and its structure is shown in Fig. 5.12. An important part of

this driver is to make it possible to allocate memory for DMA, an action which can not be done from user space. Usually memory cable of DMA is only allocated in kernel space, where direct access to the physical memory is possible, but user space is operating on virtual memory. Therefore the driver has two modules taking care of translation from user space memory into physical memory and if needed allocating dedicated consecutive physical memory in kernel space and translating it to a user level. The former being used for data transfer via DMA from the FPGA into host memory and vice versa and the latter being used to save the linked list for the DMA controller in the FPGA. As the DMA engine is located on the FPGA, no extra functionality is needed for that. The PCIe card is mapped into virtual memory and

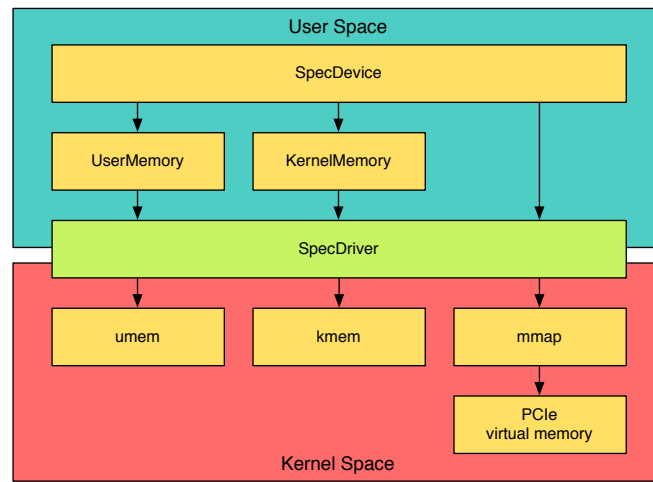


Figure 5.12: Internal structure of the SPEC kernel driver and how it is distributed over user and kernel space.

can be accessed from software like any other data structure in the memory. The kernel understands that specific memory locations are located in the PCIe card and executes the necessary steps to perform a read or write instruction. On a software level this makes reading or writing registers extremely simple, as registers can be accessed via a pointer. The software communicates with the kernel modules via a character device, which knows certain commands to initialise the software with the correct memory mapping. With the DMA being controlled from the FPGA, the host computer is not aware of the transfer and does not know when it's finished. For this reason the kernel driver allocates an interrupt, which can wake up a sleeping process. A DMA is handled as following in software:

- The user allocates memory where the data is written to or read from.
- The user memory module in the kernel driver will translate the allocated memory into a scatter/gather list.
- Via the kernel memory module a consecutive chunk of physical memory is allocated and initialised with the linked list of the transaction.

- The DMA controller in the FPGA is configured with the first item of the linked list.
- The DMA controller is instructed to start the transfer and the host system waits for an interrupt.
- When the transfer is finished or an error occurs the DMA controller sends an interrupt to the host system.

As each SPEC board has its own character device, the driver is capable of handling multiple boards at the same time.

Scan Engine and Loop Structure

The kernel driver mentioned in the previous paragraph is used by user software to control the calibration and configuration of FE-I4 and similar chips. Commands are executed in nested loops, which set parameters necessary for the specific scan, activate a portion of all pixels, and inject test pulses. Due to hardware limitations the scan engine of former Pixel readout system were limited in their capabilities. They consisted of a fixed number of nested loops, which were only capable of performing specific tasks. In the YARR software a modular loop model is implemented, an abstract LoopBase object is defined which has four main functions:

- *init()*: executed once when the loop is entered.
- *execPart1()*: first part of the loop.
- *execPart2()*: second part of the loop.
- *end()*: executed once when the loop is left.

These LoopBase objects can be added to a LoopEngine, which will execute them in a manner shown in Fig. 5.13. The engine will switch into the next inner Loop, if it exists, from the *execPart1()* function. Once it return from the inner loop it will execute the *execPart2()* function. If a done flag is set, it will move to the *end()* function else it will go back to the *execPart1()*. This structure allows for a modular loop design, with as many loops as the user wants and also the loop order can easily be changed. While a scan in prior implementations used to be a configuration of the fixed loop structure, in this implementation a scan is a collection of arbitrary loops. Multiple loops have been implemented for FE-I4 scans:

- Trigger Loop: configures and starts the trigger unit in firmware.
- Data Loop: transfers data received by the firmware via DMA into memory, typically until the trigger unit in the FPGA signals that it is finished.
- Mask Staging: Activates and cycles through a variable portion of pixels in each column.

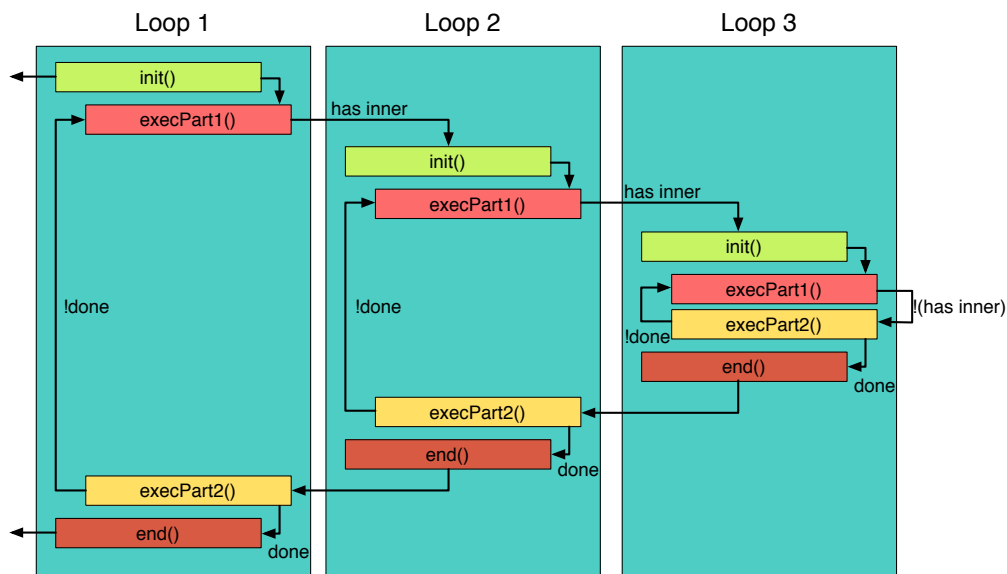


Figure 5.13: Diagram showing in which order the loop engine executes the different function of a loop object.

- Double Column Loop: Activates a variable number of double columns for injection and configuration.
- Parameter Loop: Cycles through a list a values for a variable global register of the FE-I4.
- Global Feedback Loop: Performs a search algorithm to find the optimal setting of a global register, steered from an analysis outside the loop.
- Pixel Feedback Loop: Performs a search algorithm to find the optimal setting for a pixel register of all pixels, steered from an analysis outside the loop.

With the loops it is possible to build most of the basic FE-I4 scans and tunings necessary for calibration. The exact configuration of loops needed for the different types of scans are described in Sec. 5.3.4.

Data Processing

The data from the FE is transferred into memory and as it is raw data it needs to be processed and analysed in order to present the results of a scan. The data processing structure of YARR is shown in Fig. 5.14 and is based on the handling of specific data containers. Each processing step only knows the input and output format of the data, which makes the different steps independent of each other.

Each iteration of the Data Loop produces one raw data fragment in memory, which is picked up by a data processor specifically implemented for FE-I4, it breaks the data up by channel and builds events which store all hits to a specific time point.

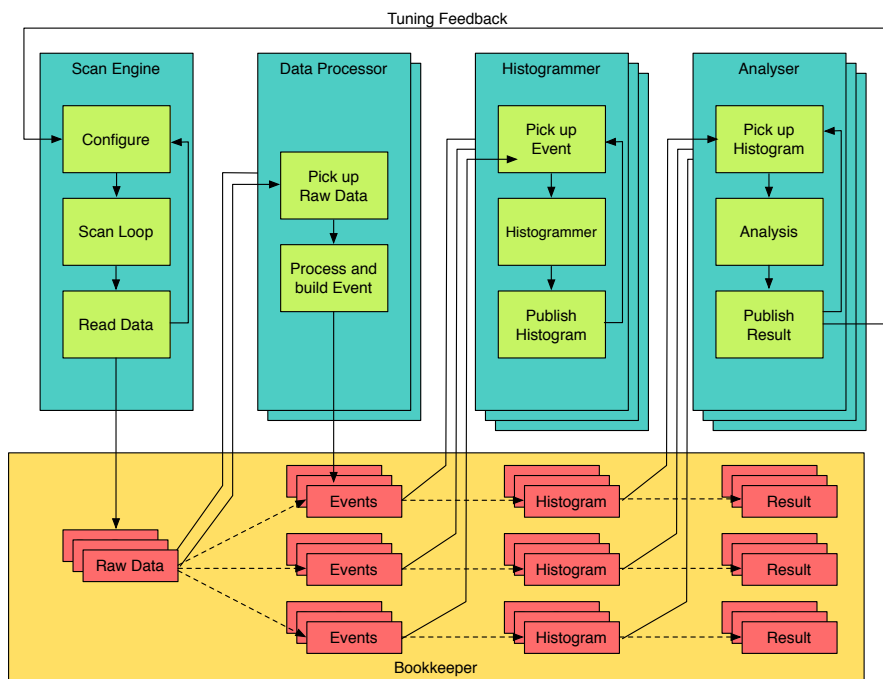


Figure 5.14: Diagram of the data processing structure of the YARR software. The scan engine is running in a single thread and the data received by it is picked up by multiple data processors. The data processor splits up the data by channel identifier and builds events, each channel then has its own pair of threads for the histogrammer and analyser.

The scan engine is single threaded and will occupy one CPU core, the highest processing performance can be achieved if as many data processor threads are running in parallel to the scan engine as there are cores left. This ensures that the host computer processes the data while the scan is running.

Every active FE-I4 has its own pair of threads which take care of histogramming and analysing the data. A feature of the histogrammer is, that it sacrifices memory for not knowing from which pixels it can expect data, i.e. even though it will only receive the data of one mask stage, the histogram is not optimised for a reduced set of data, but rather is a full size histogram capable of storing the histogram of any mask stage. This requires (depending on the mask staging) 128 times more memory, but simplifies the histogrammer drastically. An FE-I4 sized histogram with 26880×32 bit bins requires 105 kB in memory, any modern machine has multiple GB of memory available, which makes it possible to optimise the code for simplicity rather than memory consumption.

One raw data container is constructed for every possible stage of the nested loop structure and the state of the loop structure when the data was produced is stored in the container. Hence it is possible for the analysis to track which data fragments or histograms it has received and how many are missing to complete the scan. To give two examples for this:

1. Digital scan analysis: 32 mask stages and 4 double columns loops produce 128 raw data fragments. Then the analysis adds up all 128 histograms which were produced and can then publish the result of the full chip.
2. Threshold scan s-curve fitting: 32 mask stages and 4 double column loops over 100 different parameter settings, to fit the s-curve of a single pixel the analysis only has to collect the data from the 100 different parameter settings and not wait for all other loops to complete, which increases the amount of data being analysed while the scan is still running.

Overall the analysis relies on specific loops to be present in a scan and the histogrammer to produce certain histograms. This way the whole structure is modular and flexible, but does not prevent developers making errors when putting together a scan and how it is processed. However, this is acceptable as users will use pre-made scans and developers should know enough about the system to make the right decisions. In any case, this does not lead to the software crashing, an error is simply produced or there is no output.

5.3.4 FE-I4 Scans

In this section the basic FE-I4 scans and tunings will be described which were implemented into YARR. Though the exact implementation might be different, the scan procedure and goal of all scans is the same for every readout system. Therefore it is safe to assume that the following scan description applies to all other existing FE-I4 readout systems. In general a scan does not alter the chip configuration, while a tuning starts from either the existing configuration or a standard configuration and tunes a parameter to a certain target.

Common Procedures

All the different scans and tunings have a common procedure: mask staging, double column loop, and trigger injection loop. In the mask staging only a certain portion of pixels in the FE-I4 are enabled, typically $1/32$ or $1/16$, because it is not possible to read out hits in all pixels at the same time. The first iteration of a mask stage is shown in Fig. 5.15, where every 32nd pixel is activated and it is necessary to shift this mask 32 times by one pixel to scan over the whole FE-I4. The pattern is variable, but for convenience this pattern is chosen as it only needs to be shifted by one pixel in a double column, whilst other patterns would require a complete rewriting of the mask.

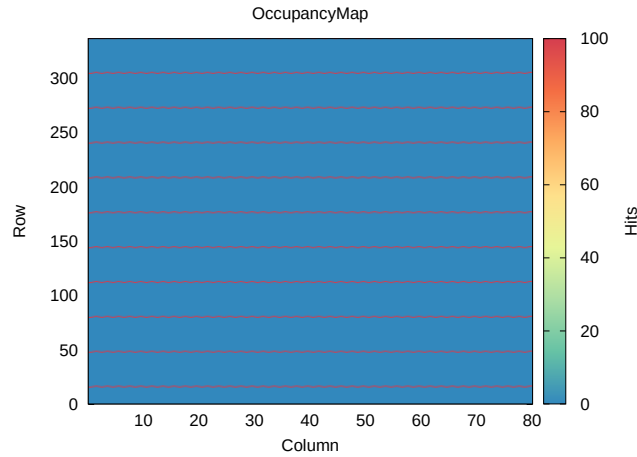


Figure 5.15: Occupancy histogram of only one mask stage.

The injection is performed on even smaller subset of pixels, only a fraction of all double columns are activated, typically $1/4$ or $1/8$, so as not to saturate the injection circuit due to too many pixels, i.e. for 32 mask stages and 4 double column loops, 210 pixels are being injected into and readout per iteration. Each injected charge needs to be triggered out of the FE-I4 with a specific timing, which is typically repeated around 100 times to gather a statistically significant data sample.

There are specific scans and tunings which do not follow this procedure, but all of the following basic scans and tunings do so.

Digital Scan

In a digital scan the analog section of a pixel cell is excluded from the scan by setting the discriminator threshold to the maximum. Instead of analog charge a digital pulse can then be injected directly into the digital region, emulating the output of the discriminator. With this method it is possible to test the digital functionality of a chip. A typical result is shown in Fig. 5.16, in this case all pixels have seen all

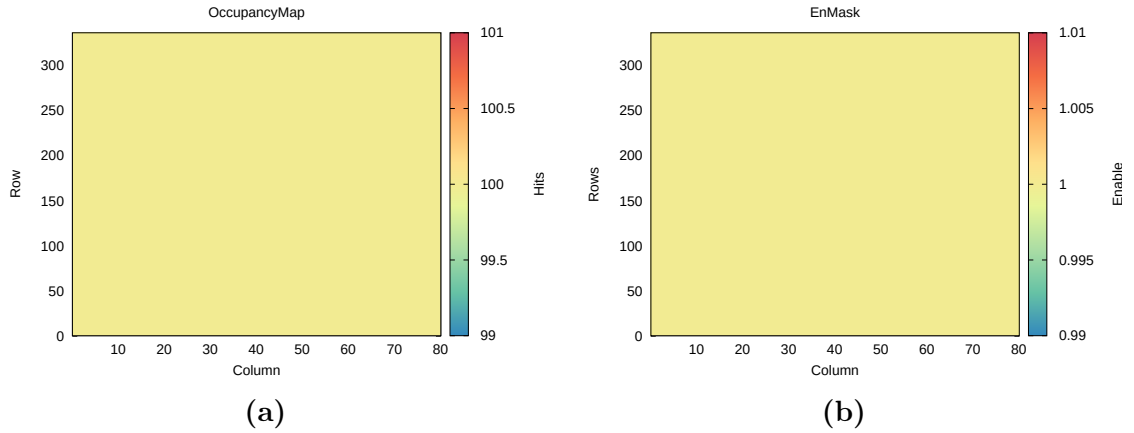


Figure 5.16: Result of digital scan with 100 injections, showing (a) the occupancy and (b) enable mask.

injections, which results in all pixels being activated in the enable mask¹⁸. Pixels which show less than 99% or more than 101% of all injections, would be disabled.

Analog Scan

An analog scan is similar to a digital scan, except that now the analog cell of a pixel is properly configured. An analog injection of a very high charge is performed and in conjunction with the result from a digital scan, it is possible to identify pixels with a failing analog region. A typical result is shown in Fig. 5.17, where a small amount of pixels have not seen all or too many injections, these pixels are disabled in the enable mask. It is common, using a good quality FE-I4, to observe very little to no digital failures, whilst small numbers of analog failures are present almost all of the time.

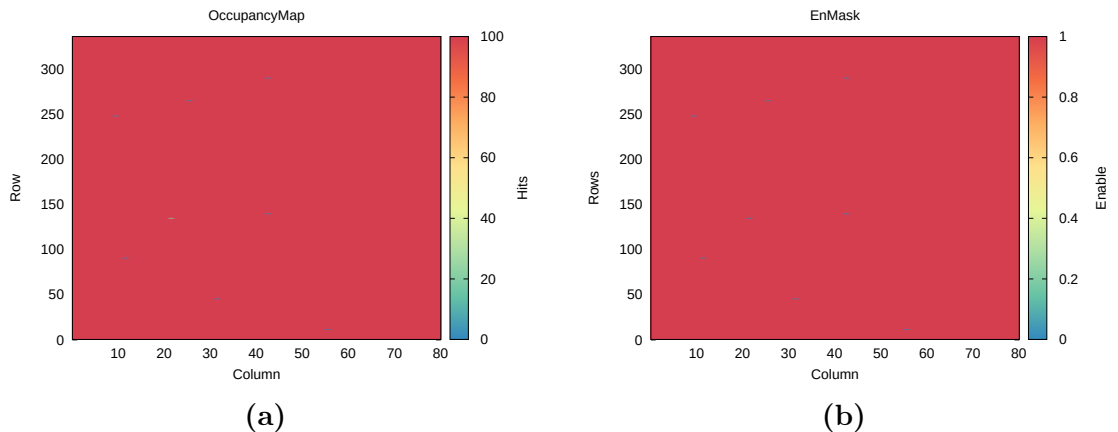


Figure 5.17: Result of analog scan with 100 injections, showing (a) the occupancy and (b) enable mask.

¹⁸1 = enabled, 0 = disabled.

ToT Scan

A ToT scan measures the ToT response to a specific charge, which is used to test the preamplifier tuning of a FE-I4. The mechanism is the same as for the analog scan, just that the injected charge is not as large and is determined by the user (usually set to be the typical charge deposited by a mip in the amount of silicon corresponding to the nominal path length in the corresponding sensor). The response is analysed in terms of ToT and not occupancy. A typical result of an uncalibrated chip is shown in Fig. 5.18 with a wide mean ToT distribution. Non-integer ToT mean values are suppressed, because the transition region from one ToT to another is small. Also measured is the per-pixel ToT sigma, which is calculated as following:

$$\sigma_{\text{ToT}} = \sqrt{\left| \frac{\left(\sum_{i=0}^{i=N} \text{ToT}_i^2 \right) - \frac{1}{N} \left(\sum_{i=0}^{i=N} \text{ToT}_i \right)^2}{N - 1} \right|} \quad (5.3.1)$$

With N being the number of injections, $\sum_{i=0}^{i=N} \text{ToT}_i^2$ and $\sum_{i=0}^{i=N} \text{ToT}_i$ represent the entry in the ToT and ToT² histograms. The σ_{ToT} can be interpreted as the stability of a ToT response, a small σ_{ToT} corresponds to high stability and a high σ_{ToT} to low stability. The value itself is the one σ value of the gaussian distribution of the ToT response of one pixel.

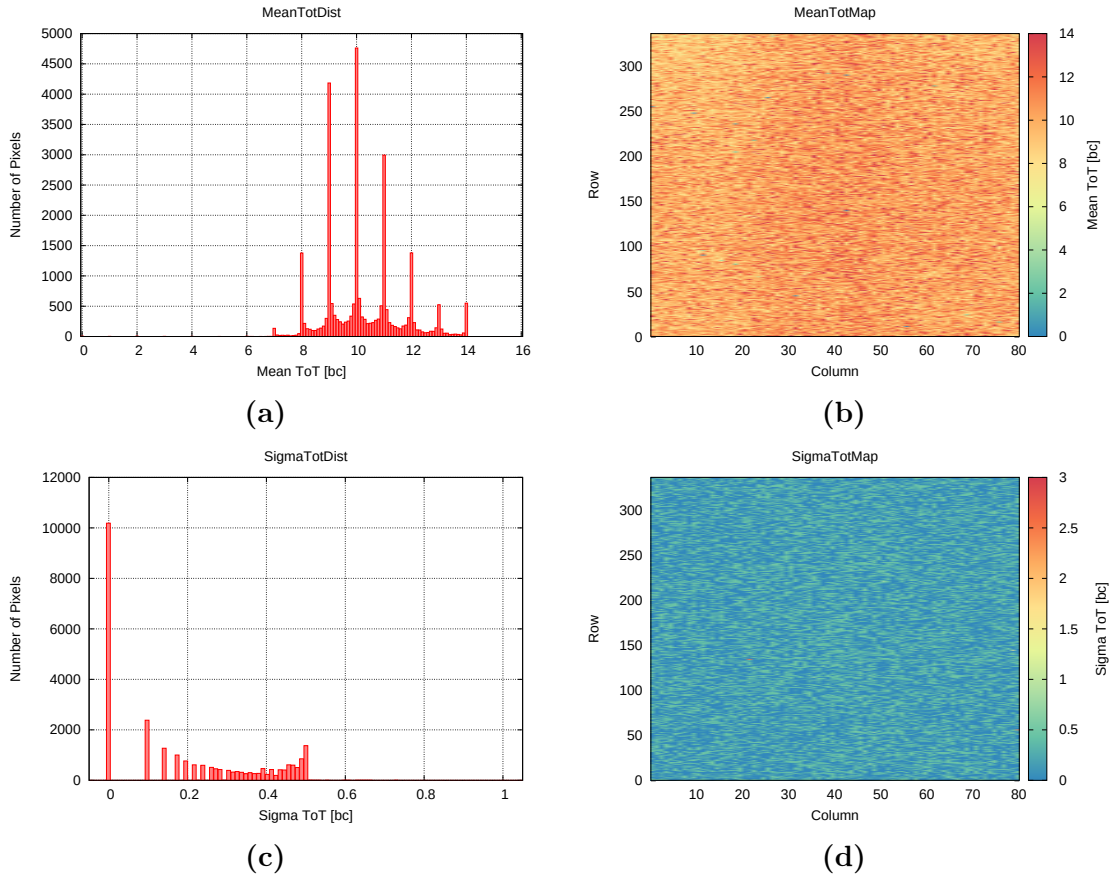


Figure 5.18: Result of a ToT scan with a charge of 16000 e and 100 injections, showing (a) the ToT distribution, (b) the ToT map, (c) the ToT sigma distribution and (d) the ToT sigma map.

Threshold Scan

A threshold scan is a series of analog scans where the charge is varied in a specific interval. The occupancy for each step and pixel is registered and forms an s-curve, shown in Fig. 5.19. The mean of the s-curve is the threshold of the pixel and the sigma is equivalent to the electronic noise. These values are extracted by fitting an error function to data points for each pixel. The parameter which controls the injected charge is VCAL and though the results here are shown in VCAL steps, this setting can be translated into charge, Q , using:

$$Q = \frac{C(a_0 + a_1 \text{VCAL})}{e} \quad (5.3.2)$$

With C being the capacitance of the injection capacitors, a_0 the VCAL offset and a_1 the VCAL slope. A typical value for one VCAL step is $50 e$, but this needs to be calibrated on a chip-by-chip basis.

The result of a threshold scan is shown in Fig. 5.20 for an uncalibrated chip, hence the threshold distribution is very wide. The mean threshold is roughly $3000 e$ and the noise is around $120 e$, where the threshold depends on the tuning and the noise on the particular assembly and its state.

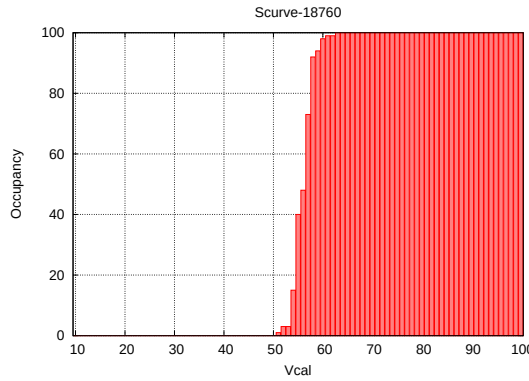


Figure 5.19: S-curve of a single pixel for 90 VCAL steps.

Global Threshold Tuning

The global threshold tuning is efficiency based, which means that it injects the target threshold as charge and measures the occupancy. The best threshold DAC setting is found when the occupancy is closest to 50%. This corresponds to moving the threshold distribution, without changing its shape, so that its mean is centered on the desired target threshold. A special search algorithm must be used for the global threshold, as if it is set below the noise level the threshold may be assumed to be too high and the generic algorithm would continuously decrease it. This is because below the noise level pixels continuously fire so becoming ‘stuck’ according to the scan which is looking for an edge trigger and not a constantly high level. Therefore the search algorithm starts with a high threshold setting and decreases it in constant

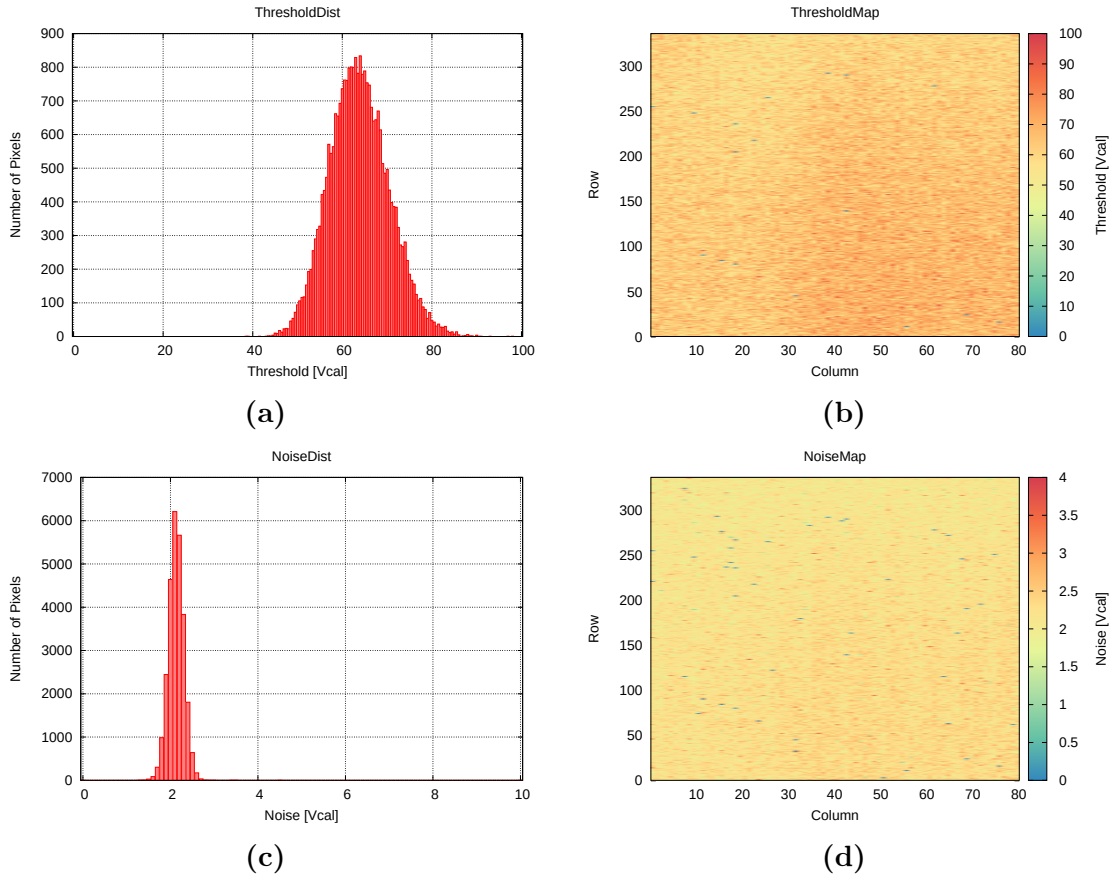


Figure 5.20: Result of a threshold scan with 90 steps, showing (a) the threshold distribution, (b) the threshold map, (c) the threshold sigma distribution and (d) the threshold sigma map. One VCAL step is roughly equal to 50 e charge.

steps until it reaches an occupancy larger than 50%, from then on it will perform a binary search like algorithm. This process and the resulting occupancy distribution is shown in Fig. 5.21. For a new tuning, the pixel thresholds are either all set to the value in the middle of the register or left at the setting from an old tuning in the case of a retune. The latter is often used when tuning to low thresholds, as it decreases the width of the threshold distribution and thereby the number of pixels which could fall below the noise level. The total number of iterations until the best setting is found depends on the target threshold, a lower target threshold will take more iterations due to constant stepping from the high threshold setting. The occupancy distribution in Fig. 5.21b, shows a bathtub-like curve, which is expected as no fine tuning of the pixel threshold has been done yet. Therefore most pixels have a threshold just above the injected charge, hence showing no response, or just below, showing full response.

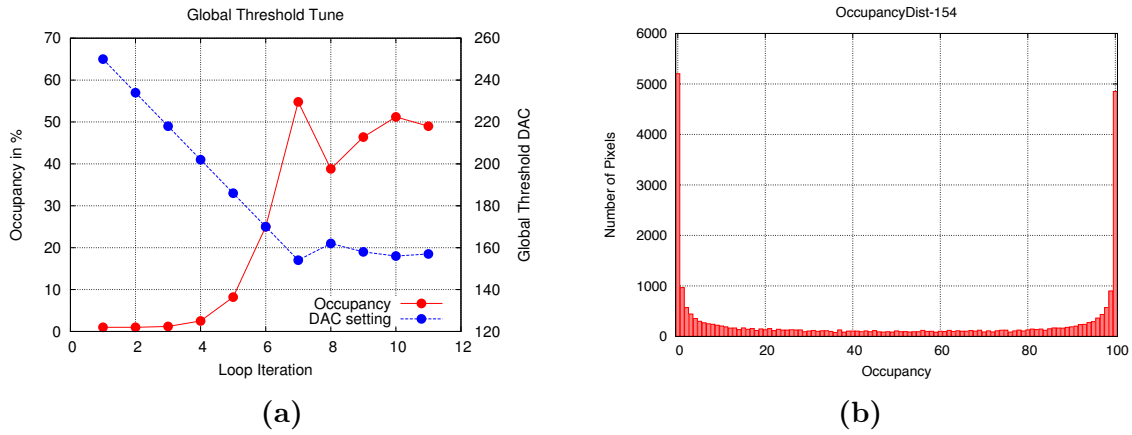


Figure 5.21: Shown are (a) the threshold DAC setting and occupancy throughout a global threshold tuning and (b) the occupancy histogram with the final threshold DAC setting.

Global Preamplifier Tuning

To tune the global preamplifier feedback current DAC setting a binary search algorithm is used. This is shown in Fig. 5.22, in this case only five iterations were needed, the maximum number of necessary iterations is 8 iterations. The tuning injects charge at the conversion target and will adjust the mean ToT value until it is closest to the chosen target ToT for this charge.

Pixel Threshold Tuning

In the pixel threshold tuning a binary search algorithm is performed to set the threshold DAC in each pixel. Just like the global threshold tuning, this one is efficiency based. This has great advantages, as a tuning based on s-curve fitting needs to loop over injections at many different charges, greatly increasing the time needed to perform such a tuning. The result after a pixel threshold tuning is shown

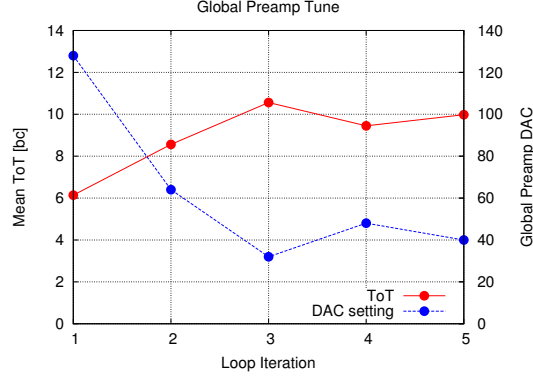


Figure 5.22: Preamplifier DAC setting and mean ToT throughout a global preamplifier tuning.

in Fig. 5.23, the threshold distribution is taken from a threshold scan performed after the tuning. The distribution is much narrower when compared to the untuned distribution in Fig. 5.20, the standard deviation of the distribution depends on the specific module, but values around 40 e are to be expected after tuning. A successful tuning can also be seen in the Occupancy distribution, the bathtub-like distribution from the prior global tuning collapsed around the 50% value, i.e. the mean value of the s-curves of all pixels overlaps at the desired threshold.

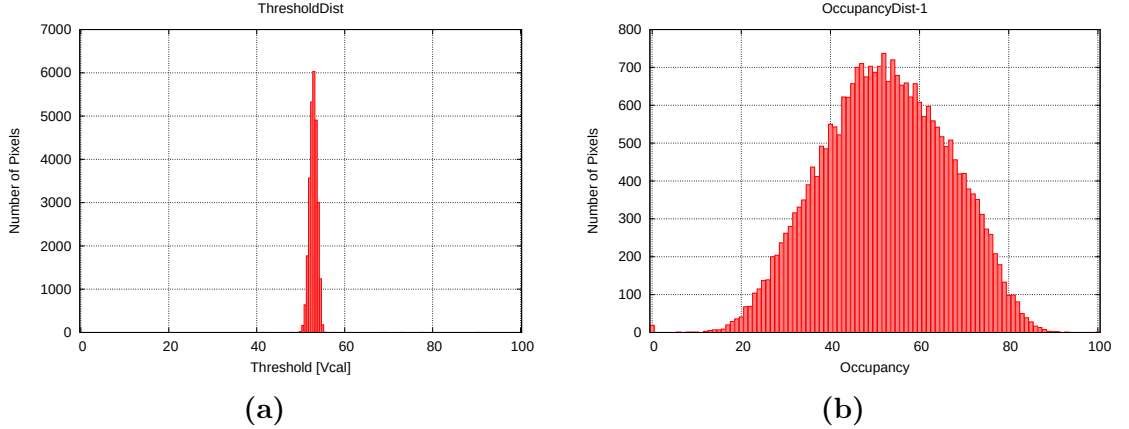


Figure 5.23: (a) Threshold distribution and (b) occupancy distribution after global and pixel threshold tuning.

Pixel Preamplifier Tuning

For the pixel preamplifier tunings, a binary search algorithm is performed on the pixel feedback current DAC setting. In a similar manner to the global preamplifier tuning the DAC of each pixel is adjusted until it gives the correct ToT response for the target charge. The resulting mean ToT distribution is shown in Fig. 5.24, a successful ToT tuning to an integer ToT is also visible in the ToT sigma distribution,

it should be closer to 0 as each pixel should give a constant rather than varying ToT response for the target charge.

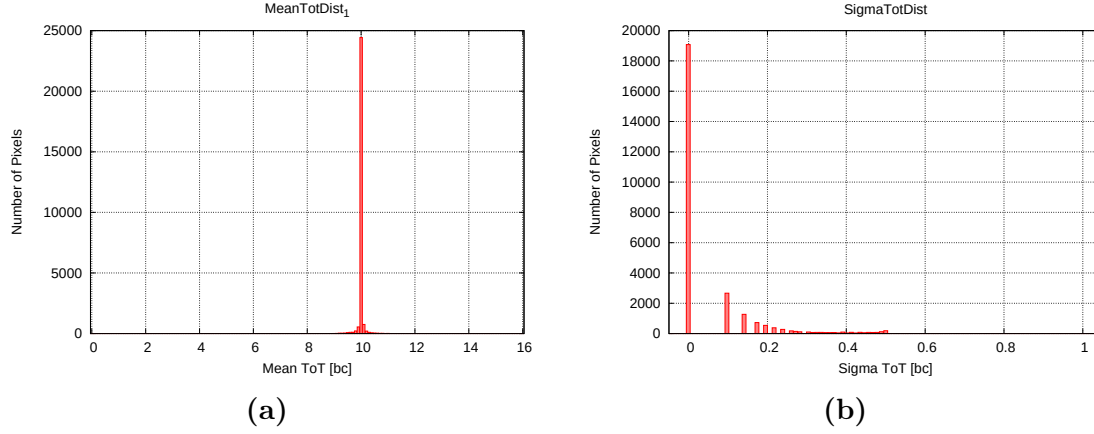


Figure 5.24: Shown is (a) the mean ToT distribution and (b) the ToT sigma after global and pixel preamp tuning.

5.3.5 Performance

So far it has been shown that YARR is a low-cost, off-the-shelf readout system for FE-I4 readout, but how does its performance compare? In the following section different benchmarks and measurements will be shown which evaluate the system for the usage of multiple FE-I4 modules, how it compares to existing FE-I4 readout systems and in which way a system with this specific performance can be used for future detector modules.

PCIe DMA Benchmark and Stability

All data send by the FE is transferred via the PCIe link into the host memory, the bandwidth of this transfer determines the number of modules which can be read out at the same time. Earlier it was also mentioned that the commands written into the FIFO of the TxCore have to be written with at least 5 MB/s to avoid an underrun. Two benchmarks are performed, one using standard Wishbone bus reading and writing to a dummy register in firmware with single PCIe transfers and the other one transferring data from the DDR3 memory via the DMA Wishbone bus with PCIe block transfers. The results of this benchmark are shown in Fig. 5.25 for single and DMA read/write transactions with varying package size. As expected the bandwidth of the single read/write transfer speed does not change with package size, as each word is transferred on its own. The write speed is 100 MB/s and read speed 2 MB/s, reading is much slower due to the time the host computer has to wait for the data to arrive in the CPU. The write speed is much faster than the TxCores' required 5 MB/s and would, in this concept, allow serial link output speeds of up to 800 MHz. The DMA transfer speed changes with increasing package size, because the overhead due to the setup of the DMA takes up a smaller ratio compared to the actual time needed to transfer data. The transfer speed saturates after a package size of around 200 kB at 400 MB/s and reading from the FPGA memory into the host memory is slightly faster than the other way around. This confirms the estimation that the DMA controller in its current design will use around 50% of the PCIe bandwidth, which is 800 MB/s with the GN4124. A useable bandwidth of 400 MB/s renders it possible to read out a maximum of 20 FE-I4 Chips at the same time, which means that the bottleneck of the system is located in simultaneously reading and writing from the DDR3 memory.

Although the average DMA transfer speed is high enough, it is possible that the time to perform a single transfer might vary a lot due to the host system performing other tasks first. Fig. 5.26 shows the time needed to complete a single transfer via DMA for a package size of 150 kB, which, for reading and writing, is very stable at around 375 μ s. Only a very small number of transfers tend to take longer and this is true for the host system being under no load or heavy load. Even the usage of the graphics card which is also connected via a PCIe link with a graphics benchmark, does not change this result.

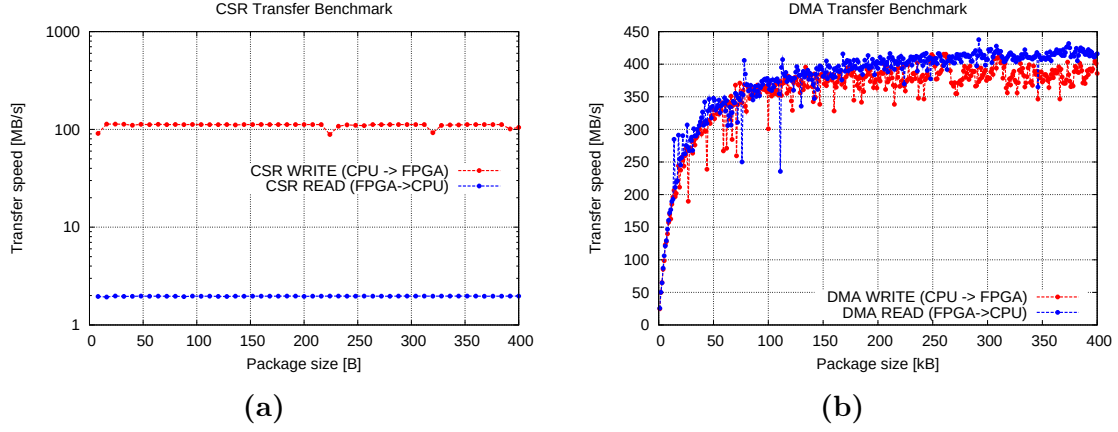


Figure 5.25: Performance benchmark of (a) single and (b) DMA read/write transactions. Each data point is the average of the transfer speed measured over 100 transactions.

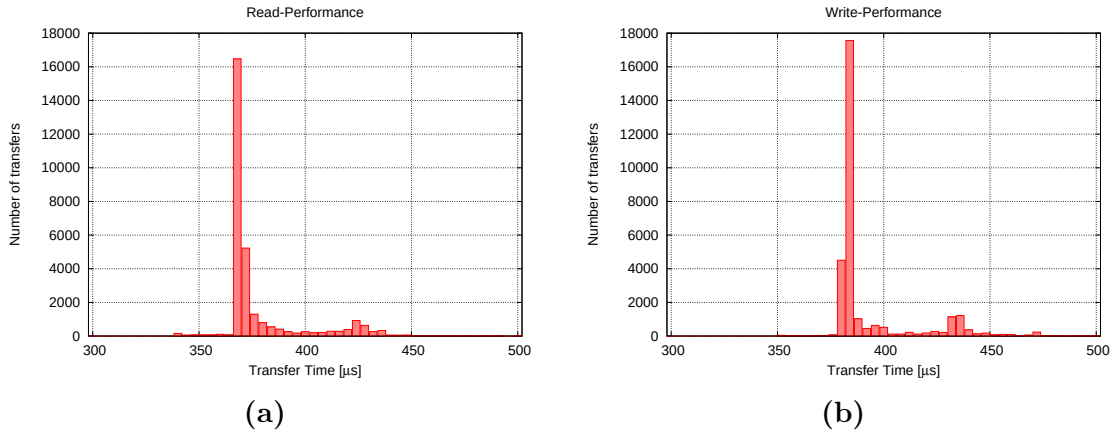


Figure 5.26: Histogram of the time needed to complete a single transfer via DMA with a package size of 150 kB for (a) reading from and (b) write to the FPGA.

Scan Performance

The time needed to perform a scan can be determined by calculating the number of data frames sent out by the FE-I4. A data frame is a response to one trigger and consists of a start of frame, data header, one data record per hit and end of frame, with start and end of frame being 8 bit and the data header and each data record being 24 bit. For each mask stage a portion of the pixels are activated and the corresponding double columns have charge injected into them. Let n be the number of pixels active and being injected into, m the trigger multiplier setting, i the number of triggers sent during one loop iteration and j the total number of loop iterations. Then the total amount of data D the FE-I4 needs to send during one scan is calculated as following:

$$D = ((i \cdot m \cdot 40 \text{ bit}) + (i \cdot n \cdot 24 \text{ bit})) \cdot j \quad (5.3.3)$$

The time t needed by the FE-I4 to send all this data is given by:

$$t = \frac{10}{8} D \cdot 6.25 \frac{\text{ns}}{\text{bit}} \quad (5.3.4)$$

For a typical scan with 32 mask stages, injection in a quarter of all columns, a trigger multiplier of 10 and 100 triggers sent to the FE-I4, the total amount of data is $D = 69$ Mbit and the time needed by sent this data with a 160 Mbit/s link is $t = 0.54$ s. This is the absolute minimum time needed for such a scan where an answer is expected from each injection into a pixel. In reality time is lost at different points in the scan loop: this calculation ignores the configuration of the pixel mask stages and the setting of various registers in order to shift the mask stage. The amount of configuration send to the module also strongly depends on the type of scan. But this calculation can help to determine the maximum trigger frequency f the trigger unit needs to be configured with:

$$f = \frac{1}{((m \cdot 40 \text{ bit}) + (n \cdot 24 \text{ bit})) \cdot 6.25 \frac{\text{ns}}{\text{s}}} \quad (5.3.5)$$

In the case above, the maximum frequency¹⁹ is $f = 23.5$ kHz, if the trigger frequency is higher the chip is not able to send out all the data and hits will be missed. The trigger frequency in the scan is chosen to be 20 kHz, which is lower than the maximum trigger frequency to leave tolerance for possible service records²⁰. A digital scan which has this kind of configuration takes 0.85 s with YARR, which is as expected higher than the theoretical value, but very close to it. These results are reproducible with longer scans and it would be necessary to calculate the exact amount of time needed for configuration during the scan to analyse how the scan performance can be increased.

¹⁹Not to be confused with the maximum trigger frequency during operation which is given by the physics occupancy.

²⁰Data frames with which the FE-I4 reports errors.

Data Processor Benchmark

The data processor picks up the raw data fragments and translates them into timing information, hits, or service records for each channel. The time needed to process the data from one digital scan is shown in Fig. 5.27, and measured for running a varying number of processor threads and the amount of data to be analysed. A digital scan delivers 128 data fragments in total which can easily be distributed over multiple threads and as expected this speeds up the processing of the data until there are more threads than available CPU cores. The amount of data from the digital scan is directly proportional to the number of triggers sent. Hence the time needed to process the data is also proportional to the amount of data, it is not observed that there is an initial overhead or saturation effect, which is to be expected.

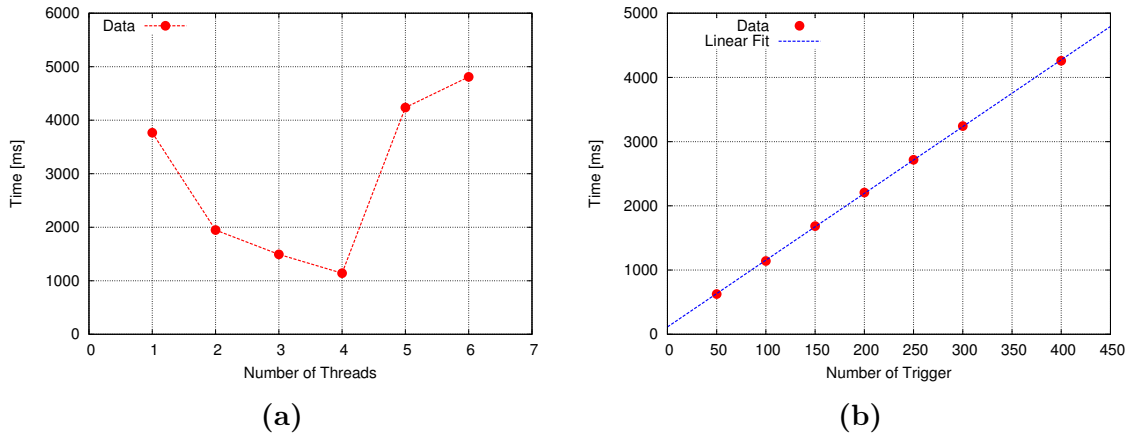


Figure 5.27: Time needed to process the data from 16 modules of one digital scan, for (a) different number of threads and (b) an increasing number of triggers while running 4 processor threads. The measurement is done on a Intel Core i5 CPU with 4 cores.

The effective bandwidth of 4 processor threads in parallel is thereby 172 MB/s, which is surprisingly low. The overall processing time can be minimised by running it in parallel to the scan, but due to its low performance, depending on the scan, it takes some extra time after the scan has finished. The processor which is used in this measurement is not optimised for performance and performs different error checks on the data, it is very likely that the bandwidth can be increased by optimising the processor for speed. Furthermore each thread should process data with the same bandwidth, therefore if run on a CPU with 8 cores the performance should double.

Histogramming in DRAM

Another critical processing step in software might be the histogramming. Histogramming may be defined as reading a value from memory, incrementing or adding a number to it and writing it back into the memory. Computers have DRAM memory which are good at writing/reading chunks of memory, but less than ideal due to their latency when randomly accessing data. This is why the histogramming in the

IBL DAQ is performed with SRAM, which can perform one read/increment/write operation in as low as 3 clock cycles. For this reason a special benchmark was developed which can test the histogramming performance in software, especially for generating multiple histograms in parallel. The benchmark creates a variable number of threads, each creates a histogram with 26880 bins (FE-I4 sized) and a data vector which gets randomly filled with 10^8 hits from 0 to 26880. Then the time is measured to fill the histogram with this data vector. Each thread performs the same operation, the data does not get divided between the threads. This simulates the histogramming of the data received from multiple modules at the same time and it is to be expected that the more modules trying to access the memory, the time needed for histogramming one hit increases. The result of the benchmark is shown in Fig. 5.28 for three different systems: a Intel Core i7 laptop, a Intel Core i7 desktop machine, and a AMD Opteron server. The Intel Core i5 and Core i7 machines are very similar except for the operating system, which is Linux on the AMD and Intel Core i5 machine and OS X on the Intel Core i7 machine. The operating system could also be the cause that these systems diverge when 10 or more histogrammer threads are running at the same time. In general the performance is expected to get worse when the number of threads is higher than the number of available CPU cores. The server is special in the sense that it has two AMD Opteron CPUs each containing 6 cores, therefore it was to be expected that the performance of this system is higher than the standard systems, but this does not seem to be the case. The reason for this could be that the memory latency of the AMD machine is higher than the latency from the Intel Core i7, but there are too many variables in a computer to specifically make out the reason for this behaviour.

The results should not be seen as a comparison of the systems, but rather to get a feeling for what performance can be expected for performing histogramming on a normal computer. Assuming, for example, that, as in the case for YARR, the histogramming is performed for 16 modules at the same time, each hit of a module takes 5 ns to be processed. Hence each module could send hits with a rate of 200 MHz and the data should not build up. The rate with which the FE-I4 can send hits is given by the size of a hit, which is 24 bit, and the output data rate of 160 MHz, resulting in a rate of 5.3 MHz. Therefore the histogramming, even if many histograms are used at the same time, should lead to no performance issues.

5.3.6 Combined Performance

To test the performance of the entire framework, the time needed to complete the various stages of a full scan is measured. Four data processor threads and one pair of histogramming and analysis threads are run in parallel to the scan engine, this allows for on-the-fly analysis of the received data. So the time to complete the data processing, histogramming or analysis is the extra time needed of the cumulative time of the stages beforehand. It is desired that the full scan is as short as possible, which is limited by the data output bandwidth of the FE-I4. The results for the time needed to complete a threshold scan with a varying number of FE-I4's being readout is shown in Tab. 5.3. Here the compiler optimisation level is set to the lowest level

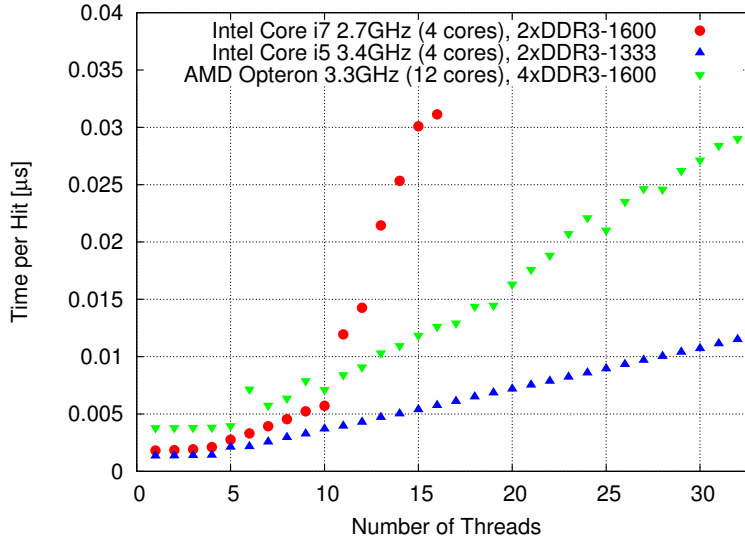


Figure 5.28: Results of a histogramming benchmark from three different types of systems for a varying number of threads, each performing 10^8 random increment operation on a 26880 bins wide histogram. The time per hit is described as the time needed to read a value from memory, increment it and write to memory. Quantitative behaviour might differ from system to system due to the many varying parameters in a computer.

to slow down the processors and see what would happen if the processing could not be completed within the scan run-time. The time needed to complete the scan loop is essentially constant and slight variations are likely due to the different threads blocking each other, because they are not busy enough when there are only a small number of modules. When there are more than three modules included in the scan, the data processor needs extra time after the scan has finished and this time increases with the number modules included. This also means that data fills up the memory, which could, in extreme cases, lead to the memory completely filling up. This can be mitigated by activating the first level of compiler optimisation, which leads to all data being processed in the time the scan is running, which is the case for all performance measurements. Without compiler optimisation this does not represent the real performance, but is able to show which stage of the scan could be a bottleneck and, as was determined earlier, the data processor code is not optimised for speed. The other stages do not contribute massively to the overall time, even though the threshold scan has one of the more complicated analyses with an s-curve fit being performed for each pixel.

5.4 Discussion of Results

The YARR readout system shows an overall convincing performance paired with high flexibility. The system has been successfully tested with 16 FE-I4 modules and is able to read out all of them at full speed. In this section the system will be

Table 5.3: Time needed to complete each stage of a full threshold scan for varying number of modules being read out. The measurement, denoted with a *, is performed with the compiler optimisation increased from level 0 to level 1.

#	Config. [ms]	Scan Loop [ms]	Data Proc. [ms]	Histogrammer & Analysis [ms]	Total [ms]
1	22	72236	93	98	72449
2	44	80196	92	98	80430
3	66	86589	6539	213	93407
4	88	82369	33804	191	116452
8	176	78829	148330	299	223634
12	264	75836	275748	394	352242
16	352	75590	407283	483	483708
16*	352	83438	90	90	83970

evaluated in terms of performance and software implementation, and compared to existing FE-I4 readout systems.

5.4.1 Processing Performance

Although the overall performance is very good with regards to reading out multiple FE-I4s, certain parts of the firmware and software stick out as possible bottlenecks. Using the DDR3 memory as a FIFO in firmware is not optimal. Addresses are, in the worst case scenario, constantly swapped between the write and read pointer, introducing the DRAM specific latency in every write or read cycle. This can be optimised by accumulating data in the FPGA to write bigger chunks into the memory or by using a priority arbiter, which could prioritise the transfer into host memory over FE data because the DMA will read a big chunk in one go. But this would also mean that the buffer in the FPGA has to be able to store all data from the FE while the DMA is running. If the data transferred in a DMA is a chunk of 250 kB with a speed of 400 MB/s, the FPGA has to buffer data for 625 μ s. Already for the data output rate of a single FE-I4 this is 13.6 kB, for 16 FE-I4s the needed buffer memory would take up too much space in the FPGA and complicate the routing of the firmware. A more intelligent solution is a priority arbiter in the memory controller, which generally prefers the DMA to read its chunk in one go, but knows the amount of data buffered by the receiver logic and gives control to it when a certain FIFO fill level has been reached. If a higher bandwidth needs to be achieved with the current hardware, this process of producing controlled back pressure needs to be carefully implemented and optimised for the given implementation.

On the software side it is surprising to find the bottleneck in a process which runs over the incoming data and converts it into FE-I4 event objects. This is a fairly simple process with a for-loop over an array of data, each data element is handed through a sequence of if-statements to decode the raw data. This is probably also where the bottleneck partially rooted in, many checks on data integrity are performed in this early stage of the system. It is possible to decrease the number of error checks

on the data, to increase the effective bandwidth to 245 MB/s, an increase of 142%. This represents the general advantage of software, it gives high flexibility during development and debugging and as soon as the functionality has been proven the code can be optimised for performance. Therefore the numbers quoted for the YARR software performance represent an early state of the software, which most likely does not represent the absolute maximum. But most certainly they give an idea about the suitable environments that this kind of readout concept can be expected to operate in.

5.4.2 FE-I4 Implementation

After the implementation of the basic concept in firmware and software, the development for the FE-I4 was very fast. It was possible, over the course of three months, to implement most of the basic FE-I4 scans and tunings and as the scan and tuning mechanism now exists new ones can be implemented with ease. The overall speed of the scans is satisfying. It is possible to make them faster but only by complex structures in the firmware, which in return increases the susceptibility to bugs. Stability of the system has been a major advantage during development, as the firmware does not depend on the incoming data it can not get stuck if something wrong is done at the software level. This increases the turnaround time to test a changes in software and thereby the overall development speed.

5.4.3 Comparison to Existing Systems

There are multiple existing FE-I4 readout system: USBpix, RCE, ROD/BOC and SEABAS²¹. When comparing them to YARR their scope of application has to be considered, USBpix and SEABAS are targeting the usage in laboratories, RCE is used mostly for multi module readout of detector test systems and telescopes and the ROD/BOC system targets the readout of the full detector. Three aspects of these systems will be used to compare them: cost, number of modules simultaneously read out and performance. The performance is evaluated by the time it takes to perform one threshold scan and a full tuning. The full tuning should consist of two iterations of global tunings and then followed by pixel tunings of threshold and preamplifier feedback current. Only fast tuning routines are considered, if available for the specific system, which do not use a lengthy iterative procedure. In general the comparison of scan times is not accurate, because the implementation of scan and tuning will differ. The comparison is summarised in Tab. 5.4, the times for the threshold scan and tuning should not be used for a quantitative comparison of performance. YARR excels in every aspect, it has a very low price point but can still read out many modules with high speed, which makes it an optimal choice for laboratories and testbeams. Though the numbers look promising, for the read out of a full detector the DAQ systems has to be scalable and this has not yet been proven for YARR or PCIe based readout system in general.

²¹A readout system based on an FPGA board connected via Gigabit Ethernet to the host computer.

Table 5.4: Comparison of existing FE-I4 readout systems.

System	# of Modules	Time Threshold Scan [min]	Time Full Tuning [min]	Cost [€]
USBpix	4	10	20	1500
RCE	16	5	15	4000
SEABAS	4	6.5	20	2000
ROD/BOC	32	20	120	8000
YARR	16	1.5	1	800

5.5 Outlook

More and more groups are taking interest in the newly developed YARR readout system. It is an attractive alternative to existing FE-I4 readout systems and seals the deal with a very low cost and high performance. The development for FE-I4 is ongoing in mainly two aspects: increasing usability by implementation of a graphical user interface and high speed readout of a testbeam telescope.

An upcoming research field are active pixel sensors, which are sensors made in a High Voltage CMOS process [36], and allow the placement of analog circuitry and digital logic inside the sensor. It is currently being explored how well these types of chips perform as a replacement for the current passive sensors, for this they are bump bonded or glued with an FE-I4 to perform the readout of the sensor. YARR will be used to simultaneously control and read out the active sensor and FE-I4, for example YARR enables the HVCMOS sensor to inject charge into the sensor element and measure the response to the FE-I4. This requires delicate timing of injection command to the HVCMOS and triggering of the FE-I4, a slight modification in the YARR software would make it possible to implement this in software very easily. The YARR scan loop structure allows for mixing HVCMOS loops with FE-I4 loops, enabling rapid development of mixed scans.

Care has been taken that the firmware and software are not specifically tailored to the hardware. If YARR shall be used for the new FE generation, which is currently under development for the ITk, the SPEC board needs to be exchanged for new, higher performance hardware. The adaptation for the new hardware should be simple and the software does not depend on the specific hardware at all. Even if it is chosen to perform certain steps of the data processing in the FPGA, or a different hardware accelerator like a GPU, the software delivers flexibility to deal with this without changing the core functionality. For the ITk it is desired to use the same software in the laboratory and for the detector, because then everyone who uses the system in the laboratory can also operate the detector and furthermore new developments, for instance scans or tunings, are also available for the detector without the need for an expert to implement them. An effort will be made to generalise the YARR software so it can be used as starting point to develop a common, hardware independent, ITk DAQ software construct.

Chapter 6

Conclusion

This thesis presents and discusses the work carried out during my PhD: the production and testing of IBL staves and the development of YARR, a novel readout concept. The IBL is a state-of-the-art detector and its presence in the ATLAS detector will be of great benefit to physics performance. YARR takes the experience acquired from the development of the IBL DAQ system and utilises modern computer hardware to realise a flexible readout system.

6.1 IBL - the new innermost tracking layer of the ATLAS detector

The quality assurance setup performed a crucial task during the IBL production: staves were operated in a detector-like environment which gave a realistic indication of their expected performance in the detector and enabled operators hands-on experience of how the IBL would behave in-situ. Multiple challenges have been faced during the production and testing, e.g. a bump-bonding problem delayed the schedule, wire bond corrosion required a swift investigation and action. The test results from the QA illustrate that the staves have excellent performance in calibration, physics response, and yield of functional pixels. A total of 18 staves qualified for the IBL and the best 14 of these have been chosen on the basis of the measured performance during QA, to be used in the actual detector. The data gathered on these staves is stored in a database from where it can be retrieved and analysed with a software framework, developed specifically for the QA. This has proven particularly useful when comparing data from the commissioning of the detector with the data obtained during qualification. The database is also easily accessed via web interface, allowing the operators of the detector a quick way to reference their observations. Aside from the QA procedure, additional data was gathered with the intention of understanding long-time operation of the detector: specific behaviours have been identified and the underlying issues, investigated. If modules show similar behaviour during operation of the detector, this background information can be utilised as reference since it is not possible to perform a similar level of in-depth debugging after insertion of the IBL into the ATLAS detector.

The IBL has left the surface in excellent condition and now its journey really begins as an addition to the ATLAS detector, performing the increased resolution tracking which is needed for high pile-up collisions in the LHC during future runs. Despite the wealth of experience and information gained about constructing state-of-the-art detectors in this period it should be noted that the production volume of the IBL was relatively small and so, in designing and manufacturing the new ITk Pixel detector over the next few years, not only will the performance of sensors and readout chips have to be improved but care must also be taken that a high volume production can be performed with high quality. Notably, during IBL production, there were cases where problems delayed the schedule but this did not stop the success of the detector. Conversely, if a problem in a high volume production is identified too late, the quality or lifetime of the detector might be at stake. Therefore it is vital to build system test setups with all detector components early in the production cycle, and then even if no issues are found the system the test setups provide hubs for scientists to actively gather experience with a new detector.

6.2 Yet Another Rapid Readout

The YARR project is a proof-of-concept stage readout system to demonstrate that the stronger focus on utilisation of modern computer systems can not only increase the performance of a readout system, but also simplify its development. Starting with the choice of hardware for the YARR project, a PCIe card - the SPEC board - was chosen which is considerably lower priced than any other FE-I4 readout system to date. The hardware capability is by no means high-end when compared with other PCIe cards, but for the usage in a laboratory environment the cost of a readout system is an important feature to promote its usage and the SPEC board more than fulfils the requirements for a testbench. Even more notable is the fact that SPEC board is developed under an open hardware license and produced by multiple commercial suppliers, making it readily available to users. The card features a Xilinx Spartan 6 FPGA that is used as a reconfigurable I/O interface to the FEs which often use custom protocols optimised for the detector environment. The functionality of the firmware is simplified such that the only task it performs is transferring data received from the FEs to the host computer memory. All real functionality is transferred into software, where the multi-core architecture of modern CPUs can be used to parallelise the processing and analysis of the received data.

In the beginning of this project it was not known to the extent of which it would be possible to process the data of multiple FEs in software, but the implementation for the FE-I4 demonstrates the high performance of the system and YARR has been tested with up to 16 FEs simultaneously sending data to the computer; monitoring of the processing and analysis during this has shown that the computer is not yet at the limits of performance and so processing multiple FEs in parallel is indeed realisable. The time taken for scans of the FE-I4 to complete is also considerably lower than with other FE-I4 readout system and a full tuning of multiple FEs can be performed in one minute. Whilst being a high performance readout system, YARR

is not specialised solely for FE-I4 readout. Other detector module technologies can easily be implemented into software and the concept would even allow for operation of different technologies with one readout system at the same time. This will soon be proven with the implementation of active sensor technology readout in YARR which needs to be operated in conjunction with an FE-I4 and, at the time of writing, previously two different readout system were always needed to perform this task.

YARR's excellent performance not only makes it attractive for laboratory users but also simply as a concept to be utilised in future readout systems. As the current system is not yet at its performance limit, and higher performance hardware is available, this concept could be utilised for the ITk Pixel readout. The crucial step, which needs to be proven, is the scalability of the system. It remains to be shown that multiple PCIe cards can be operated at full bandwidth inside a computer and, furthermore, where the processing limits of a CPU lies. However, even if a different solution for the hardware is demonstrated to be more effective, the flexibility of the software allows it be hardware independent, and therefore still usable. This will be used to develop a common software platform, with which the SPEC board can be used in a laboratory and testbeams and other hardware optimised for the detector for operation.

Appendix A

Appendix

A.1 Stave Naming Scheme

The different sections of an IBL stave are named in the following scheme shown in Fig. A.1. A stave is divided in two half staves, each having 4 powering groups which are numbered consecutively starting with one from the interaction point. Although physically there are single chip modules on a stave, the naming scheme groups all chips in two chip modules sequentially numbered from the interaction point. Each of these two chip modules has their own TTC link. Every chip in the two chip group is then identified by the number one or two, with one being the chip closer to the centre of the stave. Therefore the chips on a stave are denoted as for instance A2-1 or C7-2.

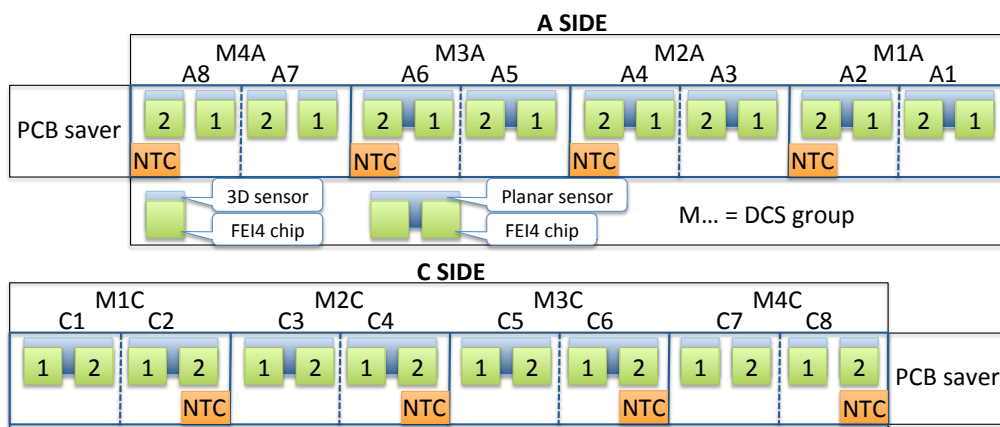


Figure A.1: Stave naming scheme [44].

A.2 FE-I4 Configuration Statistics

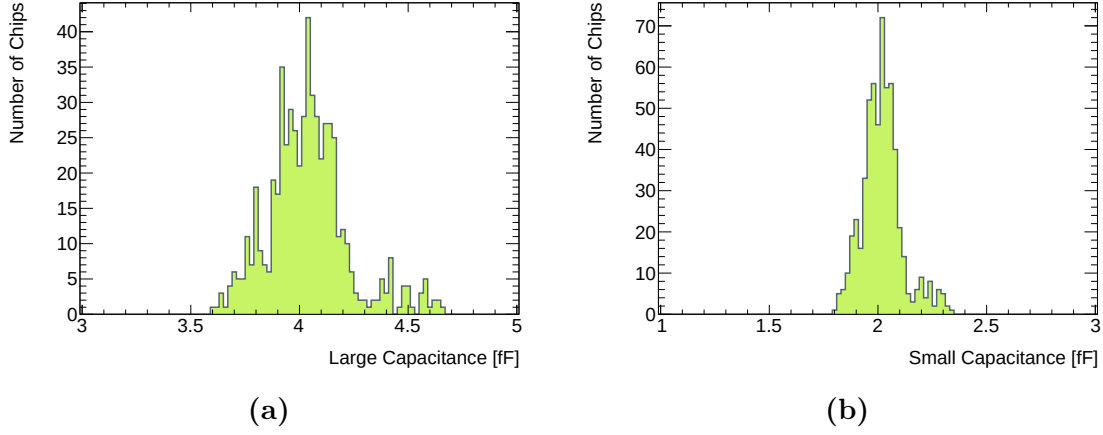


Figure A.2: Capacitance of the injection (a) large (4.04 ± 0.18 fF mean) and (b) small (2.02 ± 0.1 fF mean capacitor).

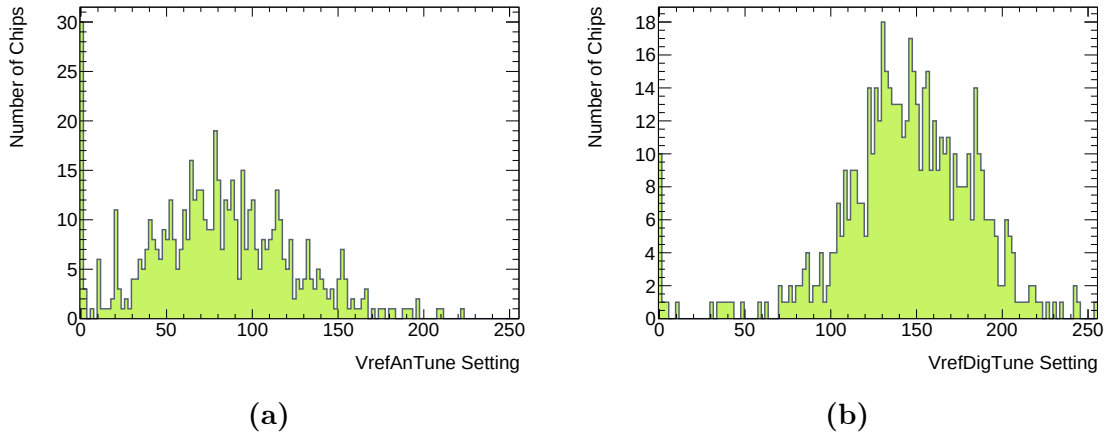


Figure A.3: DAC setting of the (a) analog (81 ± 42 mean) and (b) digital (144 ± 41) regulator adjustment.

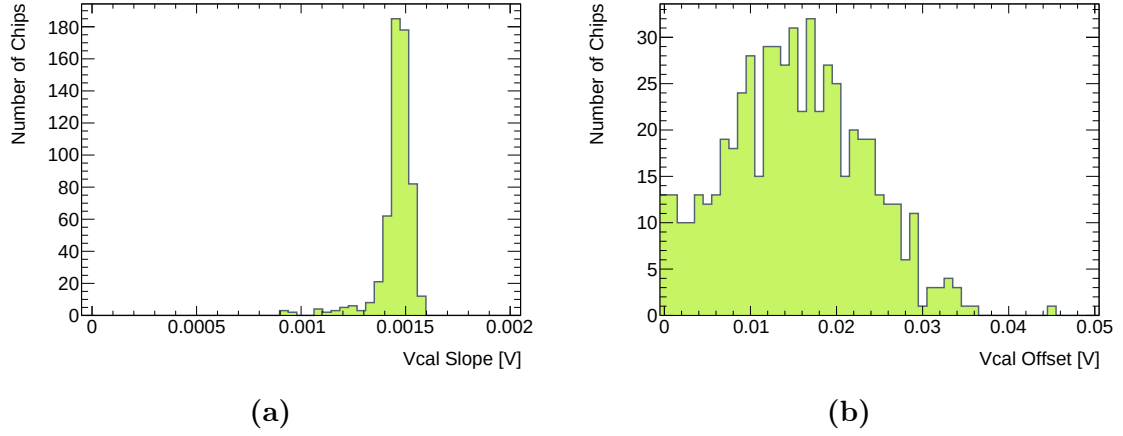


Figure A.4: (a) slope (1.46 ± 0.08 mV mean) and (b) offset (15.2 ± 7.9 mV mean) of the Vcal to injected charge formula.

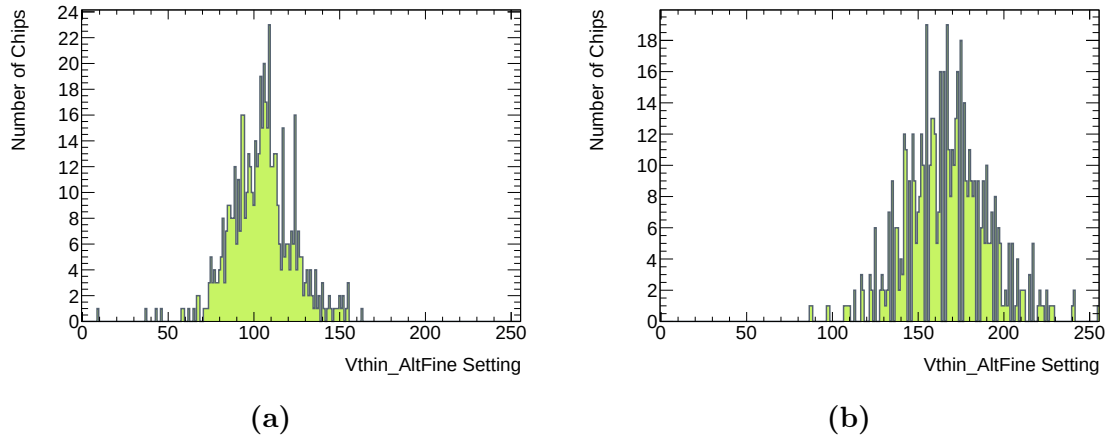


Figure A.5: Distribution of the global threshold DAC setting for a (a) 1500 e threshold tuning (105 ± 18 mean) and (b) 3000 e threshold tuning (167 ± 24 mean). The ToT response is tuned to 10 bc ToT for 16000 e in both cases.

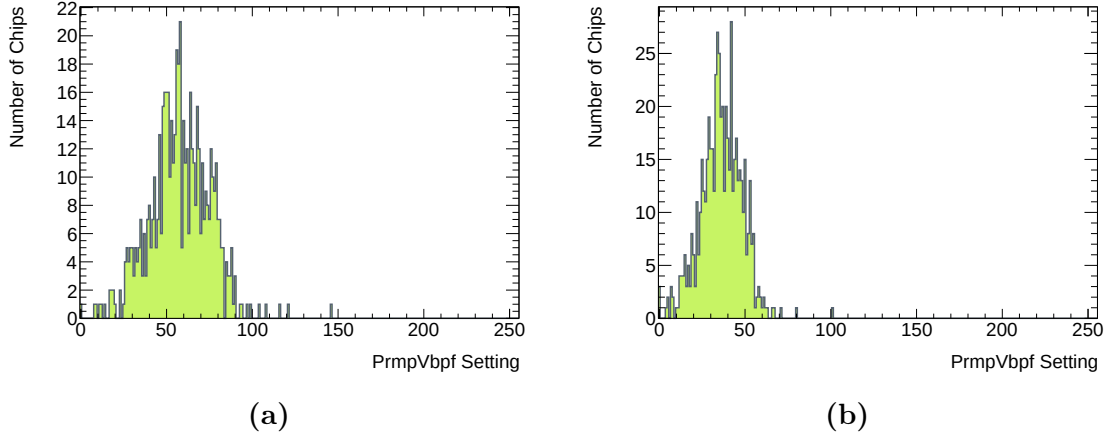


Figure A.6: Distribution of the global preamplifier feedback current DAC setting for a (a) 1500 e threshold tuning (58 ± 18 mean) and (b) 3000 e threshold tuning (36 ± 12). The ToT response is tuned to 10 bc ToT for 16000 e in both cases.

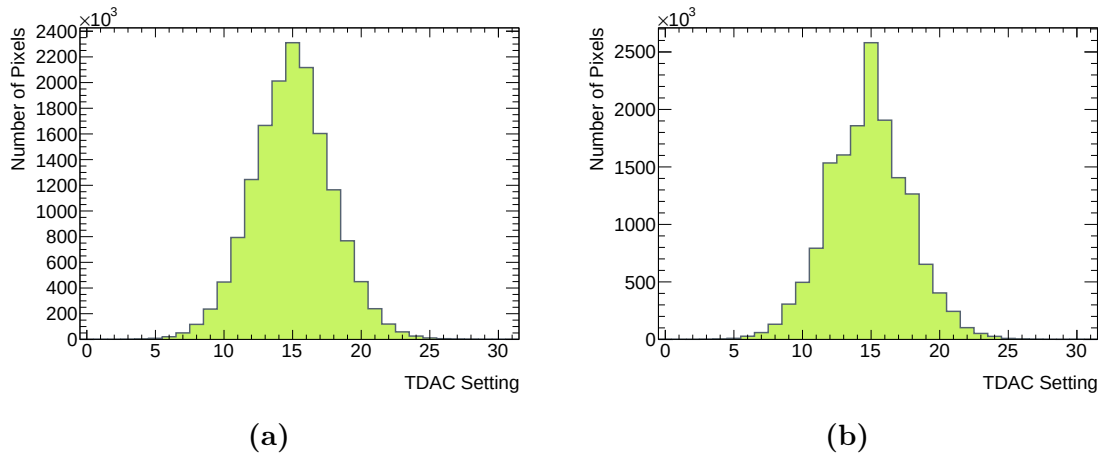


Figure A.7: Distribution of the per pixel threshold adjustment DAC setting for a (a) 1500 e threshold tuning (15.0 ± 2.9 mean) and (b) 3000 e threshold tuning (14.8 ± 2.9 mean). The ToT response is tuned to 10 bc ToT for 16000 e in both cases.

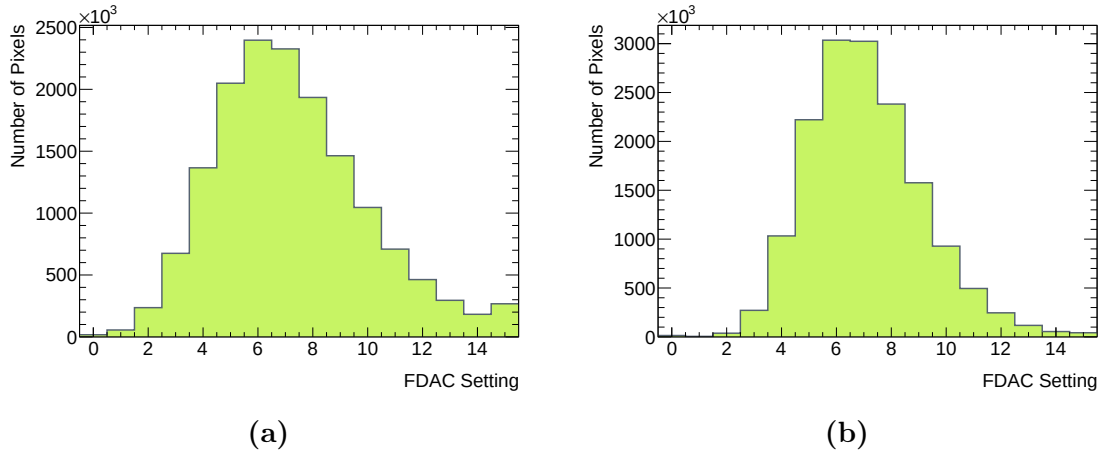


Figure A.8: Distribution of the per pixel preamplifier feedback current adjustment DAC setting for (a) 1500 e threshold tuning (7.2 ± 2.8 mean) and (b) 3000 e threshold tuning (7.1 ± 2.1 mean). The ToT response is tuned to 10 bc ToT for 16000 e in both cases.

List of Tables

1.1	List of quarks in the SM [20].	6
1.2	List of leptons in the SM [20].	6
1.3	Mediating particles of the fundamental forces [20].	7
2.1	Parameters of the inner detector. [4]	15
2.2	Parameters of the four different system of the muon spectrometer [6].	19
2.3	Luminosity weighted relative fraction of good quality data delivery in percent by the various ATLAS subsystems during LHC fills with stable beams in pp collisions [38]	21
3.1	The three different pixel sizes in an IBL planar pixel sensor.	28
3.2	The two different pixel sizes in an IBL 3D pixel sensor.	29
3.3	Main parameters of the FE-I4 readout chip [22].	30
4.1	Threshold calibration summary for different pixel types. Listed val- ues are the standard deviation of the threshold, mean noise and its standard deviation, and mean threshold over noise and its standard deviation [44].	52
4.2	Classification of pixel failures [44].	55
5.1	List of requirements and how they apply to different application sce- narios. $\checkmark$ means important and necessary, $\sim$ applicable to some degree and $\times$ not applicable or important.	70
5.2	Items of a DMA linked list.	76
5.3	Time needed to complete each stage of a full threshold scan for varying number of modules being read out. The measurement, denoted with a *, is performed with the compiler optimisation increased from level 0 to level 1.	103
5.4	Comparison of existing FE-I4 readout systems.	105

List of Figures

1.1	Feynman diagram of electron positron pair production.	8
1.2	Primitive Feynman diagram of quark-antiquark to gluon vertex. Here a blue antiquark and a green quark produce a green-antiblue gluon, the strong interaction force carrier.	9
1.3	Feynman diagram of a β^- -decay.	10
1.4	Combined search results of the ATLAS experiment: (a) The observed (solid) 95% CL upper limit on the signal strength as a function of m_H and the expectation (dashed) under the background-only hypothesis. The dark and light shaded bands show the plus/minus one sigma and plus/minus two sigma uncertainties on the background-only expectation. (b) The observed (solid) local p_0 as a function of m_H and the expectation (dashed) for a SM Higgs boson signal hypothesis ($\mu = 1$) at the given mass. (c) The best-fit signal strength $\hat{\mu}$ as a function of m_H . The band indicates the approximate 68% CL interval around the fitted value [43].	11
2.1	Computer generated drawing of the ATLAS detector layout [34]. . . .	14
2.2	Computer generated cutaway and cross-section drawing of the Inner detector layout. [33]	15
2.3	Computer generated cutaway drawing of the Pixel detector layout [35] and exploded drawing of a Pixel module [24].	16
2.4	Computer generated drawing of the ATLAS detector layout. [32] . . .	18
2.5	Total integrated luminosity and peak luminosity during Run 1. [9] . .	20
2.6	The mean number of interactions per bunch crossing during Run 1. [9]	20
2.7	Prospective timeline of the LHC towards HL-LHC with the estimated integrated luminosity and beam energy.[15]	22
3.1	(a) Computer generated drawing of the IBL and (b) a picture of the last stave being integrated onto the IPT, also visible are the IBL modules facing towards the viewer.	23
3.2	Picture of the insertion of the IBL into the Pixel detector from the C-Side of ATLAS in April 2014. [29]	24
3.3	The Atlantis event displays showing a track of a cosmic particle passing through the inner detector, including two hits in the IBL, (a) with and (b) without B-field [8].	25

3.4	Simulation of FE-I3 inefficiencies, the occupancy for nominal luminosity is 0.16 hits per double column per bunch crossing [13].	26
3.5	(a) Transverse impact parameter d_0 resolution for single muons at different p_T as a function of $ \eta $ with and without the IBL. (b) Light jet rejection factor as a function of b-tagging efficiency with and without the IBL. (c) Primary vertex reconstruction efficiency in $t\bar{t}$ events as a function of average number of pile-up [13].	27
3.6	Illustration of the edge region of a planar pixel sensor showing the long pixel reaching underneath the guard rings. [10]	28
3.7	Crosssection of a 3D sensor manufactured by (a) FBK and (b) CNM. [10]	29
3.8	Block diagram of an FE-I4 showing the different functional blocks and the 2×2 pixel region which contains a buffer shared by all pixels in the region [52].	31
3.9	(a) Schematic of the pixel pre-amplifier and discriminator circuit [22], also shown are the 13 pixel configuration bits. Shown in (b) is the effect of the threshold of the discriminator and the preamplifier feedback current onto the charge to ToT conversion. Lowering the threshold increases the ToT of the same charge (shown in blue), increasing the feedback current lowers the ToT (shown in green).	32
3.10	Schematic of the IBL readout system showing the data flow from the Off-detector site to the detector and back.	33
3.11	Pictures of the (a) IBL ROD card and (b) the IBL BOC card.	34
3.12	Schematic of the IBL powering and control scheme.	35
3.13	Module production yield for the 4/5 different module batches assembled in Bonn and Genova.	36
3.14	Flow chart showing the path of each part of IBL up to the point where the detector was fully assembled on surface. Each stage has its own quality control, to be able to trace back to where the source of problems occur.	37
4.1	Graph showing each step of the production from loading, up to integration on the IPT for every stave. This does not necessarily represent the number of days worked on each stave, as the work on some staves was paused for specific reasons. For example, ST07 and ST08 were used to test the bring and integration procedure and were dismantled from the IPT afterwards.	41
4.2	Schematic of the stave test stand, showing the different components needed for powering, control, readout, and cooling [44].	42
4.3	(a) The environmental box seen from the outside, with the TRACI (blue) on the right, parts of the powering and readout in the rack on the left. (b) The environmental box seen from the inside with two staves mounted and connected, the linear stage to move a radioactive source is in between the two staves (source not mounted).	43

4.4	Flow of the stave testing procedure, outlining which tests are performed on which day. Two staves go through this procedure simultaneously and the actions of the last day and the first day of a new stave pair can be performed in parallel.	44
4.5	(a) A example photograph of a module taken during the optical inspection and (b) a possible issue discovered during the optical inspection, a loose wire bond close to the sensor edge which needed to be removed.	45
4.6	Results of one stave of the (a) power cycles and (b) sensor IV characteristic [44].	46
4.7	Results of a source scan performed on ST11, (a) the hit map with the passive components lowering the occupancy and (b) the distribution of number of hits per pixel.	48
4.8	Block diagram of the stave QA framework, showing the interaction between RCE database and stave QA database.	49
4.9	Example of an overview table produced by the web interface for ST03.	50
4.10	Resulting (a) threshold and (b) threshold sigma distribution for the cold tuning to a threshold of 1500 e of all 18 staves [44].	51
4.11	Mean (a) threshold and (b) threshold sigma of each position on a stave for the cold tuning to a threshold of 1500 e of all 18 staves [44].	52
4.12	(a) resulting mean ToT distribution and (b) mean ToT of each position on a stave for a tuning of 10 bc ToT at 16000 e [44].	52
4.13	Cluster ToT distribution of one chip for a cluster size of one and clusters with more than one hit. A Landau-Gauss fit is shown for the clusters with more than hit, as most of the background is contained in the single hit clusters.	53
4.14	(a) distribution of cluster ToT MPV per chip and (b) mean cluster ToT MPV of each position on a stave [44].	54
4.15	Results of a bad pixel analysis from one stave[44].	54
4.16	Total number of bad pixels per (a) chip and (b) stave.	55
4.17	(a) number of bad pixels with respect to η for selected and rejected staves and (b) $\eta - \phi$ map of operational pixels of the IBL [44].	56
4.18	Comparison of different functions fitted to the ToT response of a single pixel.	58
4.19	Comparison of different functions fitted to the ToT response of one chip.	60
4.20	Comparison of the charge residual from the data from one chip fitted with four different functions.	61
4.21	(a) correlation of number of bad pixels as identified by the module QA and stave QA, (b) magnified view.	62
4.22	For each pixel tagged bad during the stave QA but not during module QA the respective bin is increased and for each pixel tagged bad during the module QA and not during the stave QA the respective bin is decreased. The resulting histogram is shown for (a) 3D and (b) planar modules.	63

4.23	Noise distribution of 3D FBK modules for the different ends of a stave and the two different test setup positions, called "SR" and "CR" [44].	63
4.24	Noise injected into an 3D FBK sensor from (a) the NTC and (b) the Hitbus line [44].	64
4.25	Threshold noise map of double chip planar module on ST07 where the neighbouring chip is dead [44].	65
5.1	Comparison of current readout concept (a) as used for the Pixel detector and (b) the new concept.	69
5.2	Picture of the SPEC board, pointing out the main components. . . .	72
5.3	Block diagram of the YARR firmware. Red blocks depict interfaces to hardware, yellow blocks the communication busses, blue blocks are the main common firmware blocks and green blocks are FE specific. .	74
5.4	Block diagram of the GN4124 core. The standard Wishbone master interface is controlled via the host computer. During a DMA transfer the GN4124 core acts as the master and steers the DMA transfer on both ends. The data path is shown with bold arrows and control signals with thin arrows.	75
5.5	Schematic of how physical memory is mapped into virtual memory. The memory is segmented in pages and the representation in physical memory is not necessarily contiguous and can also contain pages owned by other processes (red blocks).	76
5.6	Measurement of the DMA Wishbone bus with an FPGA internal logic analyser. The pauses between the data transfers (blue regions) are the time until the next list item has been received.	77
5.7	Block diagram of the DDR3 controller showing the wrapper around a Xilinx memory controller.	78
5.8	Block diagram of the TxCore, which sends clock and commands in a serial stream to the FE. The number of output channels, 3 in this example, is scalable.	79
5.9	Block diagram of the RxCore, each channel is deserialised, decoded, and buffered in a FIFO. From there a round robin arbiter multiplexes the data into the following logic. The number of input channels is variable, here shown for 3.	81
5.10	Composition of framed FE-I4 data in a 32 bit word, showing the encoding of the channel identifier and the three bytes of data (D_0 , D_1 , D_2).	81
5.11	Block diagram of the RxBridge, which receives the data from the RxCore and builds packages which are written to the memory. . . .	82
5.12	Internal structure of the SPEC kernel driver and how it is distributed over user and kernel space.	83
5.13	Diagram showing in which order the loop engine executes the different function of a loop object.	85

5.14	Diagram of the data processing structure of the YARR software. The scan engine is running in a single thread and the data received by it is picked up by multiple data processors. The data processor splits up the data by channel identifier and builds events, each channel then has its own pair of threads for the histogrammer and analyser.	86
5.15	Occupancy histogram of only one mask stage.	88
5.16	Result of digital scan with 100 injections, showing (a) the occupancy and (b) enable mask.	89
5.17	Result of analog scan with 100 injections, showing (a) the occupancy and (b) enable mask.	89
5.18	Result of a ToT scan with a charge of 16000 e and 100 injections, showing (a) the ToT distribution, (b) the ToT map , (c) the ToT sigma distribution and (d) the ToT sigma map.	91
5.19	S-curve of a single pixel for 90 VCAL steps.	92
5.20	Result of a threshold scan with 90 steps, showing (a) the threshold distribution, (b) the threshold map , (c) the threshold sigma distribution and (d) the threshold sigma map. One VCAL step is roughly equal to 50 e charge.	93
5.21	Shown are (a) the threshold DAC setting and occupancy throughout a global threshold tuning and (b) the occupancy histogram with the final threshold DAC setting.	94
5.22	Preamplifier DAC setting and mean ToT throughout a global preamplifier tuning.	95
5.23	(a) Threshold distribution and (b) occupancy distribution after global and pixel threshold tuning.	95
5.24	Shown is (a) the mean ToT distribution and (b) the ToT sigma after global and pixel preamp tuning.	96
5.25	Performance benchmark of (a) single and (b) DMA read/write transactions. Each data point is the average of the transfer speed measured over 100 transactions.	98
5.26	Histogram of the time needed to complete a single transfer via DMA with a package size of 150 kB for (a) reading from and (b) write to the FPGA.	98
5.27	Time needed to process the data from 16 modules of one digital scan, for (a) different number of threads and (b) an increasing number of triggers while running 4 processor threads. The measurement is done on a Intel Core i5 CPU with 4 cores.	100
5.28	Results of a histogramming benchmark from three different types of systems for a varying number of threads, each performing 10^8 random increment operation on a 26880 bins wide histogram. The time per hit is described as the time needed to read a value from memory, increment it and write to memory. Quantitative behaviour might differ from system to system due to the many varying parameters in a computer.	102

A.1	Stave naming scheme [44].	111
A.2	Capacitance of the injection (a) large (4.04 ± 0.18 fF mean) and (b) small (2.02 ± 0.1 fF mean capacitor).	112
A.3	DAC setting of the (a) analog (81 ± 42 mean) and (b) digital (144 ± 41) regulator adjustment.	112
A.4	(a) slope (1.46 ± 0.08 mV mean) and (b) offset (15.2 ± 7.9 mV mean) of the V_{cal} to injected charge formula.	113
A.5	Distribution of the global threshold DAC setting for a (a) 1500 e threshold tuning (105 ± 18 mean) and (b) 3000 e threshold tuning (167 ± 24 mean). The ToT response is tuned to 10 bc ToT for 16000 e in both cases.	113
A.6	Distribution of the global preamplifier feedback current DAC setting for a (a) 1500 e threshold tuning (58 ± 18 mean) and (b) 3000 e threshold tuning (36 ± 12). The ToT response is tuned to 10 bc ToT for 16000 e in both cases.	114
A.7	Distribution of the per pixel threshold adjustment DAC setting for a (a) 1500 e threshold tuning (15.0 ± 2.9 mean) and (b) 3000 e threshold tuning (14.8 ± 2.9 mean). The ToT response is tuned to 10 bc ToT for 16000 e in both cases.	114
A.8	Distribution of the per pixel preamplifier feedback current adjustment DAC setting for (a) 1500 e threshold tuning (7.2 ± 2.8 mean) and (b) 3000 e threshold tuning (7.1 ± 2.1 mean). The ToT response is tuned to 10 bc ToT for 16000 e in both cases.	115

Bibliography

- [1] *SNAP12 Multi-Source Agreement*, May 2002.
- [2] P. A. Franaszek A. X. Widmer. A dc-balanced, partitioned-block, 8b/10b transmission code. *IBM Journal of Research and Development*, 27(5), 1983.
- [3] Kazuyoshi Akiba et al. Charged particle tracking with the timepix ASIC. *NIM A*, 661(1):31 – 49, 2012.
- [4] ATLAS Collaboration. *ATLAS inner detector: Technical Design Report, 1*. Technical Design Report ATLAS. CERN, Geneva, 1997.
- [5] ATLAS collaboration. ATLAS trigger performance: Status report. 1998. CERN-LHCC-98-15.
- [6] ATLAS Collaboration. The ATLAS Experiment at the CERN Large Hadron Collider. *Journal of Instrumentation*, 3, 2008.
- [7] ATLAS Collaboration. Letter of Intent for the Phase-II Upgrade of the ATLAS Experiment. Technical Report CERN-LHCC-2012-022. LHCC-I-023, CERN, Geneva, Dec 2012.
- [8] ATLAS Collaboration. Event Displays from Non-Collision Data. <https://twiki.cern.ch/twiki/bin/view/AtlasPublic/EventDisplayRun2Start>, 3 2015.
- [9] ATLAS Collaboration. Luminosity Public Results. <https://twiki.cern.ch/twiki/bin/view/AtlasPublic/LuminosityPublicResults>, 1 2015.
- [10] ATLAS IBL Collaboration. Prototype ATLAS IBL Modules using the FE-I4A Front-End Readout Chip. *Journal of Instrumentation*, 7(11):P11010, 2012.
- [11] G. Balbi, G. Bruni, M. Bruschi, I. D’Antone, J. Dopke, et al. A PowerPC-based control system for the read-out-driver module of the ATLAS IBL. *JINST*, 7:C02016, 2012.
- [12] J. Butterworth, J.B. Lane, M. Postranecky, and M.R.M. Warren. TTC Interface Module for ATLAS Read-Out Electronics. 10th Workshop on Electronics for LHC and Future Experiments LECC, 2004.

- [13] M Capeans, G Darbo, K Einsweiler, M Elsing, T Flick, M Garcia-Sciveres, C Gemme, H Pernegger, O Rohne, and R Vuillermet. ATLAS Insertable B-Layer Technical Design Report. Technical Report CERN-LHCC-2010-013. ATLAS-TDR-019, CERN, Geneva, Sep 2010.
- [14] CDF, D0 collaboration. Combined cdf and d0 upper limits on standard model higgs boson production with up to 8.6 fb⁻¹ of data. *arXiv*, 2011. arXiv:1107.5518v2.
- [15] CERN. The HL-LHC project. <http://hilumilhc.web.cern.ch/about/hl-lhc-project>, 03 2015.
- [16] Jonathan Corbet, Alessandro Rubini, and Greg Kroah-Hartman. *Linux Device Drivers, 3rd Edition*. O'Reilly Media, Inc., 2005.
- [17] W. Erdmann. The front-end for the CMS pixel detector. *NIMA A*, 549, 2005.
- [18] M. Backhaus et al. Development of a versatile and modular test system for ATLAS hybrid pixel detectors. *Nuclear Instruments and Methods in Physics Research Section A*, 650(1), 2011.
- [19] T. Akesson et. al. Status of design and construction of the transition radiation tracker (trt) for the atlas experiment at the lhc. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 522(1-2):131 – 145, 2004. TRDs for the Third Millennium. Proceedings of the 2nd Workshop on Advanced Transition Radiation Detectors for Accelerator and Space Applications.
- [20] K.A. Olive et al. (Particle Data Group). 2014 review of particle physics. *Chin. Phys. C*, 2014.
- [21] Lyndon Evans and Philip Bryant. LHC Machine. *Journal of Instrumentation*, 3(08):S08001, 2008.
- [22] FE-I4 Collaboration. *FE-I4B Integrated Circuit Guide*. v2.3.
- [23] Richard P. Feynman. *QED: The strange theory of light and matter*. Princeton University Press, 2006.
- [24] G. Aad, et. al. ATLAS Pixe Detector electronics and sensors. *Journal of Instrumentation*, 3, 2008.
- [25] Gennum. *GN412x PCI Express Family Reference Manual*, June 2009.
- [26] J N Jackson. The ATLAS semiconductor tracker (sct). *Nucl. Instrum. Methods Phys. Res., A*, 541:89–95, 2005.
- [27] S Kersten et al. Detector Control System of the ATLAS Insertable B-Layer. *Conf. Proc.*, C111010:MOPMS021. 4 p, 2011.

- [28] Kerstin Lantzsch. *The Finite State Machine for the ATLAS Pixel Detector and Beam Background Studies with the ATLAS Beam Conditions Monitor*. PhD thesis, University of Wuppertal, 2011. To be published.
- [29] Claudia Marcelloni De Oliveira. IBL installation into the inner detector of the ATLAS Experiment side C. <https://cds.cern.ch/record/1702006>, May 2014.
- [30] E. van der Bij O. Byole, R. McLaren. the s-link interface specification. Technical report, CERN, 1997.
- [31] OpenCores Organization. *WISHBONE System-on-Chip (SoC) Interconnection Architecture for Portable IP Cores*, 9 2007.
- [32] Joao Pequena. Computer Generated image of the ATLAS calorimeter. Mar 2008.
- [33] Joao Pequena. Computer generated image of the ATLAS inner detector. Mar 2008.
- [34] Joao Pequena. Computer generated image of the whole ATLAS detector. Mar 2008.
- [35] Joao Pequena. Computer generated images of the Pixel, part of the ATLAS inner detector. Mar 2008.
- [36] Ivan Perić, Peter Fischer, Christian Kreidl, Hong Hanh Nguyen, Heiko Augustin, et al. High-voltage pixel detectors in commercial CMOS technologies for ATLAS, CLIC and Mu3e experiments. *Nucl.Instrum.Meth.*, A731:131–136, 2013.
- [37] Peripheral Component Interconnect Special Interest Group. PCI Express Base Specification, October 2014.
- [38] The ATLAS Collaboration. Data quality information for 2010 and 2011 data. <https://twiki.cern.ch/twiki/bin/view/AtlasPublic/RunStatsPublicResults2010>.
- [39] The ATLAS Collaboration. The ATLAS inner detector commissioning and calibration. *The European Physical Journal C*, 70(3), 2010.
- [40] The ATLAS Collaboration. Commissioning of the atlas muon spectrometer with cosmic rays. *The European Physical Journal C*, 70(3), 2010.
- [41] The ATLAS Collaboration. Readiness of the atlas liquid argon calorimeter for lh collisions. *The European Physical Journal C*, 70(3), 2010.
- [42] The ATLAS Collaboration. Readiness of the atlas tile calorimeter for lh collisions. *The European Physical Journal C*, 70(4), 2010.

- [43] The ATLAS Collaboration. Observation of a new particle in the search for the standard model higgs boson with the ATLAS detector at the LHC. *Physics Letters B*, 716(1):1–29, 2012.
- [44] The ATLAS Collaboration. ATLAS Pixel IBL: Stave Quality Assurance. (ATL-INDET-PUB-2014-006), Sep 2014.
- [45] The CMS Collaboration. Observation of a new boson at a mass of 125 gev with the CMS experiment at the LHC. *Physics Letters B*, 716(1):30–61, 2012.
- [46] Marius Wensing et al. Firmware development and testing of the ATLAS IBL Back-Of-Crate card. *PoS, TIPP2014*:216, 2014.
- [47] Xilinx. *UG381 - User Guide: Spartan-6 FPGA SelectIO Resources*, December 2010.
- [48] Xilinx. *UG388 - Spartan-6 FPGA Memory Controller*, August 2010.
- [49] Xilinx. *DS126 - Spartan-6 FPGA Data Sheet: DC and Switching Characteristics*, September 2011.
- [50] Xilinx. *DS180 - 7 Series FPGAs Overview*, May 2015.
- [51] Xilinx. *UG471 - 7 Series FPGAs SelectIO Resources*, May 2015.
- [52] V Zivkovic et al. The fe-i4 pixel readout system-on-chip resubmission for the insertable b-layer project. *Journal of Instrumentation*, 7(02):C02050, 2012.

Acknowledgment

This work was supported by the Wolfgang Gentner Programme of the Federal Ministry of Education and Research in Germany.

I would like to thank my university supervisor Prof. Peter Mättig and my CERN supervisor Heinz Pernegger for the opportunity to obtain a PhD in physics. I'm very grateful for the experience of working at CERN and for the support I received from the group in Wuppertal. The realisation of my own project, with all its challenges, would not have been possible without them. Furthermore I would like to thank Jens Dopke and Tobias Flick for making themselves available for the variety of issues I ran into during my studies.

A big shoutout goes to all my friends and colleges at CERN, with whom I have shared many, many hours at work. Without you this job would only be half as fun and I'm happy to not only convene with you for meetings, but also for dinner (and a dram!) afterwards. In particular I would mention Jennifer Jentzsch who has been there from the beginning of my stay at CERN and with whom I have spent countless hours in the laboratory.

I'm deeply grateful for the continuous support of my family, especially during the difficult times we faced. My brother and mother took on many tasks in order to give me the time to finish my PhD studies.

Thank you, Rebecca Carney, for being with me and helping me whenever you can. I'm truly looking forward to explore new places, and discover new physics with you!

Erklärung

Hiermit versichere ich, dass ich diese Arbeit selbstständig verfasst und keinen anderen als die angegebenen Quellen und Hilfsmittel benutzt, sowie Zitate kenntlich gemacht habe. Diese Dissertation ist in keinem weiteren Fachbereich einer wissenschaftlichen Hochschule vorgelegt worden.

Wuppertal, der 02.06.2015

(Timon Heim)
