Technical Design Report: A High-Granularity Timing Detector for the ATLAS Phase-II Upgrade

Draft version: 3.0
Created: 3rd April 2020 – 09:59
Prepared by: The ATLAS Collaboration

Abstract

The large increase of pileup interactions is one of the main experimental challenges for the HL-LHC physics programme. A powerful new way to mitigate the effects of pileup is to use high-precision timing information to distinguish between collisions occurring close in space but well-separated in time. A High-Granularity Timing Detector, based on low gain avalanche detector technology, is therefore proposed for the ATLAS Phase-II upgrade. Covering the pseudorapidity region between 2.4 and 4.0, this device will improve the detector physics performance in the forward region. The typical number of hits per track in the detector was optimized so that the target average time resolution per track for a minimum-ionising particle is 30 ps at the start of lifetime, increasing to 50 ps at the end of HL-LHC operation. The high-precision timing information improves the pileup reduction to improve the forward object reconstruction, complementing the capabilities of the upgraded Inner Tracker (ITk) in the forward regions of ATLAS and leading to an improved performance for both jet and lepton reconstruction. These improvements in object reconstruction performance translate into sensitivity gains and enhance the reach of the ATLAS physics programme at the HL-LHC. In addition, the HGTD offers unique capabilities for the online and offline luminosity determination, an important requirement for precision physics measurements.

Contents

1 Introduction **1**

2 Detector Requirements and Layout **5**
2.1 Beam conditions at the HL-LHC 5
2.2 Detector overview and requirements 7
2.3 Detector layout and optimisation 8
2.4 Radiation hardness 15

3 Performance and Physics Benchmarks **19**
3.1 Simulation and hit-level detector response 20
3.1.1 Detector geometry 21
3.1.2 Sensor simulation 23
3.1.3 Simulation of digitised readout signals 24
3.1.4 Occupancy 27
3.1.5 Detector-level reconstruction 29
3.2 Reconstruction and detector performance 30
3.2.1 Association of HGTD timing measurements to tracks 30
3.2.2 Determination of the time of the primary vertex 38
3.3 Physics object performance 39
3.3.1 Challenges for associating tracks to vertices 41
3.3.2 Strategy of improving physics object performance 42
3.3.3 Suppression of pileup jets 43
3.3.4 Lepton track isolation 44
3.3.5 Additional applications 47
3.4 Physics 49
3.4.1 Vector boson fusion Higgs production 50
3.4.2 Vector boson scattering 54
3.4.3 Measurement of $\sin^2\theta_{\text{eff}}$ 57
3.4.4 Impact of the luminosity uncertainty 60
3.4.5 Trigger for magnetic monopole searches 62
3.4.6 Limitations in the Monte Carlo modeling 63

4 Technical Overview **65**
4.1 Introduction 65

4.2 Detector overview and key requirements . . . 65
4.2.1 Expected Radiation levels 66
4.2.2 Key requirements 68
4.2.3 Read-out bandwidth 68
4.3 Hybrid HGTD module . . . 69
4.3.1 Sensors 70
4.3.2 Front End ASIC 72
4.3.3 Module assembly 75
4.4 Module loading on support structure . . . 77
4.5 Off detector electronics, calibration and luminosity . . . 78
4.5.1 Peripheral Electronics Boards 79
4.5.2 Luminosity 79
4.5.3 t_0 time calibration 80
4.6 Power distribution and detector control system . . . 81
4.6.1 HV system 81
4.6.2 LV system 81
4.6.3 Monitoring and Controls 82
4.7 Mechanics, Services and Infrastructure . . . 82
4.8 Assembly, Installation and Commissioning . . . 83
4.9 Next steps towards construction . . . 84

5 Sensors **87**
5.1 Sensor parameters and requirements . . . 87
5.2 Low Gain Avalanche Detectors . . . 88
5.2.1 Overview 88
5.2.2 LGAD productions 91
5.3 Radiation damage and irradiations . . . 92
5.4 Sensor tests: methodology and experimental techniques . . . 94
5.5 LGAD performance before and after irradiation . . . 95
5.5.1 Electrical characterisation: I-V and C-V 95
5.5.2 Operating bias Voltage and self-triggering 100
5.5.3 Collected charge and gain 101
5.5.4 Efficiency 103
5.5.5 Time resolution 103
5.5.6 Uniformity, inter-pad gap and edge region 105
5.5.7 Long term and stability tests 107
5.5.8 Leakage current and power after irradiation 111
5.6 Operational aspects and bias voltage evolution in HGTD . . . 113
5.7 Summary of present sensor design . . . 113
5.8 Roadmap for future sensor productions and activities . . . 115

6 Front-end Electronics **117**
6.1 General requirements . . . 118
6.2 ASIC architecture . . . 120
6.2.1 Channel architecture 121
6.2.2 Readout architecture 123
6.3 Single-channel readout electronics . . . 125
6.3.1 Preamplifier 125
6.3.2 Discriminator 127
6.3.3 TDC 128
6.3.4 Internal pulser 130
6.3.5 Hit processor 131
6.3.6 Luminosity processing unit 133
6.4 End of Column logic and digital blocks . . . 134
6.4.1 Clock generation unit 134
6.4.2 Matrix readout process 136
6.4.3 Slow control 139
6.5 Radiation tolerance . . . 140
6.6 Power distribution and grounding . . . 141
6.7 Prototype performance . . . 142
6.7.1 Test bench measurements 142
6.7.2 Test beam measurements 144
6.7.3 Irradiation tests 146
6.8 Monitoring . . . 147
6.8.1 Temperature monitoring 147
6.8.2 Supply voltages monitoring 148
6.8.3 Complete monitoring system 149
6.9 Roadmap towards production . . . 150

7 Module Assembly and Loading **153**
7.1 Introduction . . . 153
7.2 The bare module . . . 153
7.2.1 Bare module assembly 153
7.2.2 First bare module prototypes: process and results 154
7.2.3 Hybridization specifications 158
7.2.4 Quality assurance / quality control 159
7.2.5 Production hybridization strategy 160
7.3 Module design and assembly . . . 160
7.3.1 General description 160
7.3.2 Voltage distribution and signal readout: flex cables 161
7.3.3 Gluing and wire-bonding 170
7.3.4 Specifications, quality assurance / quality control 173
7.3.5 Production strategy for flex cables and module assembly 174

7.4 Module loading . . . 174
7.4.1 General description 174
7.4.2 Support units and detector units 176
7.4.3 Gluing studies 177
7.4.4 Procedure for loading and qualification 178
7.4.5 Detector unit assembly strategy 179
7.5 Thermal Simulations . . . 180
7.6 Roadmap towards production . . . 186

8 Power distribution, Grounding and Shielding 189
8.1 High voltage . . . 189
8.2 Low voltage . . . 190
8.3 Grounding/shielding . . . 191
8.4 Roadmap for power system . . . 192

9 Peripheral Electronics 193
9.1 Data transmission . . . 194
9.1.1 LpGBT 198
9.1.2 Optical links 200
9.2 DC-DC converters . . . 200
9.3 Control and monitoring . . . 201
9.4 Connectors . . . 202
9.5 Peripheral boards . . . 202
9.5.1 Physical limitations 202
9.5.2 Layout considerations 203
9.5.3 Layout 204
9.6 Power dissipation . . . 206
9.7 Roadmap towards PEB production . . . 206

10 DAQ, calibration, luminosity and control 209
10.1 DAQ interface . . . 209
10.1.1 Off-detector electronics 209
10.1.2 Calibration and timing 211
10.2 Timing correction . . . 213
10.2.1 Sources of clock jitter 214
10.2.2 Timing correction procedure 216
10.3 Luminosity . . . 218
10.3.1 HGTD as a luminometer 219
10.3.2 Linearity of the luminosity determination 220
10.3.3 Statistical precision of the luminosity determination 221
10.3.4 Noise and afterglow subtraction 222
10.3.5 Systematic uncertainties affecting the luminosity determination 223

10.3.6 Occupancy readout at 40 MHz 226
10.3.7 Luminosity back-end electronics 229
10.3.8 Processing of the occupancy data in the Luminosity Software 230
10.3.9 Per-event luminosity information stored in the ATLAS raw data 230
10.3.10 Operation in non-Stable Beams conditions 231
10.3.11 Minimum-bias trigger at Level-0 232
10.4 Detector Control System . . . 232
10.4.1 High Voltage 233
10.4.2 Low voltage 233
10.4.3 Environmental monitoring 234
10.4.4 Peripheral electronics 236
10.4.5 Configuration 237
10.4.6 DCS software 238
10.4.7 External systems 239
10.4.8 Interlock system 240
10.5 Roadmap for DAQ, luminosity and control system . . . 241

11 Detector Mechanics 243
11.1 Engineering design overview . . . 243
11.2 Detector overall layout . . . 244
11.3 CO_2 cooling system . . . 246
11.3.1 Requirements 247
11.3.2 Cooling design 249
11.3.3 Cooling plant demonstrator 252
11.4 Moderator . . . 254
11.5 Hermetic vessel . . . 255
11.5.1 Requirements 255
11.5.2 Front cover 256
11.5.3 Back cover 258
11.5.4 Inner ring design 259
11.5.5 Outer ring design 261
11.6 Support and cooling disks . . . 262
11.6.1 Requirements 262
11.6.2 Geometry and design 263
11.6.3 Assembly criteria 264

12 Detector Infrastructure 265
12.1 Specifications . . . 265
12.2 Services layout . . . 267
12.3 Patch panels in PP-EC area . . . 268
12.4 Services routing on the calorimeter front wall . . . 269
12.5 Services connection to outer ring and inside the vessel . . . 271

12.6 Services installation 272

13 Detector Assembly, Installation, and Commissioning **277**
13.1 Assembly and commissioning on surface 277
13.1.1 Half disks instrumentation 277
13.1.2 Detector assembly on hermetic vessel 278
13.1.3 Quality Assurance after assembly 278
13.2 Installation in the cavern and commissioning 279
13.2.1 Access and maintenance scenarios 279
13.2.2 Transport and installation the HGTD in the cavern 280
13.2.3 Services connection and commissioning 282
13.2.4 Radiation environment, and radio protection 283

14 Demonstrator **287**
14.1 Introduction 287
14.2 Heater demonstrator 287
14.3 Peripheral and back-end electronics, data acquisition 291
14.3.1 Peripheral electronics demonstrator 291
14.3.2 DAQ demonstrator 292
14.3.3 HGTD module 293
14.4 Full demonstrator 294
14.5 Full demonstrator tests 294
14.6 Schedule and organisation 295

15 Project Organization, Costs, and Schedule **297**
15.1 Organization and management 297
15.1.1 Upgrade organisation in ATLAS 297
15.1.2 HGTD project organisation 298
15.1.3 Technical milestones 301
15.1.4 Deliverables and WBS 303
15.2 Schedule and main milestones 305
15.3 Resources 321
15.3.1 CORE Costs 323
15.3.2 Required Labour Effort 325
15.4 Risk Analysis and Risk Mitigation 327
15.4.1 Risk Management Process and Plan 328
15.4.2 Major risks and mitigation strategies 329

A Expected Energy Spectra **335**

B Monte Carlo samples **337**

C Two-ring Layout Used in Simulation **339**

257 **D Technical Drawings** **341**

258 **Bibliography** **351**

1 Introduction

The high-luminosity (HL) phase of the Large Hadron Collider (LHC) at CERN [1] aims to deliver an integrated luminosity of up to 4000 fb^{-1}. The instantaneous luminosity of the HL-LHC will reach up to $7.5 \times 10^{34}\,\text{cm}^{-2}\,\text{s}^{-1}$, a large increase from the $2.1 \times 10^{34}\,\text{cm}^{-2}\,\text{s}^{-1}$ obtained during Run 2 of the LHC. Two extended periods without physics operation are anticipated prior to the HL-LHC operation during which upgrades will be made to the ATLAS experiment. Long Shutdown 2 (LS2) which began in 2019 facilitates the Phase-I upgrades, and Long Shutdown 3 (LS3) which is currently planned to last from 2025 until mid 2027 will be used for the extensive Phase-II upgrades, which will allow ATLAS to cope with the higher luminosities expected at the HL-LHC, and provide new capabilities. Due to the current Covid-19 emergency, potential changes to the schedule and cost of the machine will be discussed in the next year.

This report describes the technical design of a High-Granularity Timing Detector (HGTD), a novel detector introduced to augment the new all-silicon Inner Tracker (ITk) [2] in the forward region, adding the capability to measure charged-particle trajectories in time as well as space. The HGTD will measure the times of minimum-ionising particles with an average time resolution of approximately 30 ps per track at the beginning of the operation of HL-LHC, increasing to 50 ps at the end of the operation of HL-LHC. The HGTD will provide high-precision time measurements for charged particles, enhancing the performance of physics object reconstruction, complementing the ITk in the forward region. The object-level reconstruction improvements increase the physics potential of ATLAS at the HL-LHC.

The past decade of LHC running has been highly successful, despite the challenging experimental environment, with projects such as the discovery of the Higgs boson, the continued precision measurements of physics at the electroweak scale, and the broad search programs. The legacy of the last few years of LHC studies is therefore a stronger confidence in the potential of the LHC to push the reach in both precision and sensitivity well beyond what was originally assumed possible. It can be also argued that, with new abilities to determine the times of the interactions within one bunch crossing, new techniques beyond what are elaborated in this document will be developed to exploit this new capability in Run 4.

The scope of the HL-LHC physics programme is vast, covering many important areas of active research in terms of Standard Model (SM) precision measurements, study of the Higgs boson properties, as well as continuing searches of physics signatures beyond the SM at the

TeV scale. The experimental challenges of the HL-LHC will be more difficult than those at the LHC, and improvements in our detectors are required to meet these demands.

Many of the precision SM measurements as well as the Higgs properties measurements are already limited by systematic uncertainties using the Run-1 and Run-2 datasets. It is therefore important to improve the detector capabilities to limit the impact of systematic uncertainties related to the reconstructed physics objects as well as the modelling of backgrounds. To this end, improvements on the signal-to-background ratio, in addition to the statistical significance gains, are needed to increase the precision of these measurements. The HGTD will play a significant role in improving the reconstruction of physics objects in the forward region of ATLAS, restoring the performance achieved there to levels similar to the ones in the central detector regions. This will lead to a reduction in systematic uncertainties in this new phase space which has been largely unexplored in the first decade of the ATLAS physics programme.

Precision measurements of SM processes benefit from access to new regions of phase space, going beyond on-shell signal strength determination. Measurements of Higgs pseudo-observables as well as its couplings and production cross-sections are key measurements planned for the HL-LHC. These measurements will require well-understood detector performance in the forward region of ATLAS. Similarly, HL-LHC will allow ATLAS to measure to high precision differential distributions of SM processes, and with the HGTD, the precision and phase-space available for these measurements will be increased.

Without increasing the centre-of-mass energy of the collisions in the HL-LHC, many searches for new physics will shift from analyses in the style of bump-hunting to analyses looking for broad off-shell discrepancies in the tails of distributions and other new methods, such as long-lived particles giving rise to displaced vertices. These searches will be extremely challenging and any hope in finding new physics, just beyond the reach of the collision energies, will also require precise understanding of all reconstructed objects in the increased acceptance ($|\eta|$ up to 4) with the new detector.

A critical aspect of precision measurements is the precise determination of the luminosity. The HGTD is uniquely positioned to measure both the online luminosity on a bunch-by-bunch basis during HL-LHC running, and the high-precision determination of the integrated luminosity offline. The luminosity uncertainty is already one of the leading uncertainties in measurements of Higgs couplings during the first two runs of the LHC, and thus the HGTD will contribute to determine an accurate luminosity measurement for measurements of the Higgs properties with ATLAS.

This document is organised as follows. A detector overview and its requirements (e.g. expected radiation levels) are presented in Chapter 2. Chapter 3 presents simulation-based studies showing how the detector will improve ATLAS object reconstruction and physics sensitivity. The technical design of the HGTD is summarized in Chapter 4. The HGTD will consist of many silicon-based Low Gain Avalanche Detectors (LGADs), placed in front of

the end-cap and forward calorimeters at $2.4 < |\eta| < 4.0$ and arranged such that a charged particle traverses two or three sensors. Chapter 5 describes the LGAD sensors and their expected performance, based on measurements of prototype devices that include irradiation at the levels expected at the HL-LHC. Chapter 6 describes the front-end electronics, a low-noise, radiation-hard custom ASIC called the ALTIROC, and the performance of the analog front end. Chapter 7 discusses the hybridization of the LGAD and ALTIROC into modules of a single LGAD sensor bump-bonded to two ALTIROC chips, their assembly into detector units which are mounted onto cooling discs, and their connection via flex cables to peripheral electronics boards at the outer radii. Chapter 7 discusses the module hybridization with a single LGAD sensor bump-bonded to two ALTIROC ASICs, the assembly of the modules into detector units, and their connection via flex cables to peripheral electronics boards at the outer radii. Chapter 8 describes the powering and control of the detector. Chapter 9 describes the peripheral electronics boards, and Chapter 10 summarizes the connection of the detector to the ATLAS data acquistion system, the real-time inter-calibration of the arrival time within the readout path and the 40 MHz readout of highly-granular hit multiplicity data for real-time luminosity measurement. Chapter 11 provides the engineering design of the cooling system for the LGADs and front-end electronics Chapter 12 presents the mechanical design of the overall detector, the necessary services and their routing. Chapter 13 describes the assembly and commissioning of the detector. Chapter 14 describes a set of intermediate prototypes that will integrate elements of the full detector during the remaining R&D period into a demonstrator, in order to validate key aspects of the design. Finally, Chapter 15 documents the organisation, schedule and resources of the project to deliver and commission the detector for the start of the HL-LHC operations in Run 4.

2 Detector Requirements and Layout

2.1 Beam conditions at the HL-LHC

Pileup is one of the main challenges at the HL-LHC. The exact beam-spot characteristics of the HL-LHC have not yet been determined. Several scenarios are under study for the nominal operation scheme [3]. This report assume that an average of 200 simultaneous pp interactions ($\langle\mu\rangle = 200$, Phase-2 planned maximum pileup) will occur within the same bunch crossing interval. A major challenge for the ITk is the pileup suppression in the primary vertex and in the object reconstruction in this high pileup environment, especially in the end-cap region. The luminous region will have an estimated Gaussian spread of 30 to 60 mm along the beam axis (z direction[1].) The width in time could range from 175 to 260 ps. The case considered in this report is the "nominal" scenario, with Gaussian standard deviation of approximately 50 mm along the beam axis and spreads of 175 ps in time.

The spatial pileup line density, i.e. the number of collisions per length unit along the beam axis during one bunch crossing, is a key quantity for evaluating the performance of ATLAS with and without the HGTD. For $\langle\mu\rangle = 200$ an average pileup density of 1.8 vertices/mm is expected. However this average masks the effect of the local variations which are illustrated in Figure 2.1. In the same plot the distribution for $\langle\mu\rangle = 30$ is shown for comparison. The local pileup vertex density then is calculated by computing the average number of interactions per unit length in a window of ± 3 mm around the signal vertex for $\langle\mu\rangle = 200$. This window is large enough to avoid quantisation effects and small enough to probe the tails of the distribution. The most probable local pileup density for this scenario is 1.44 vertices/mm for $\langle\mu\rangle = 200$.

The ITk measures the longitudinal impact parameter of a track with respect to perigee. This can be combined with the corresponding high precision time measurement of all the tracks associated to the primary vertex in order to exclude those that are not compatible with one-another. Figure 2.2 gives an illustration of this technique, showing the distribution of

[1] The ATLAS experiment uses a right-handed coordinate system with its origin at the nominal interaction point (IP) in the centre of the detector and the z-axis along the beam pipe. The x-axis points from the IP to the centre of the LHC ring, and the y-axis points upward. Cylindrical coordinates (r, ϕ) are used in the transverse plane, ϕ being the azimuthal angle around the z-axis. The pseudo-rapidity is defined in terms of the polar angle θ as $\eta = -\ln\tan(\theta/2)$.

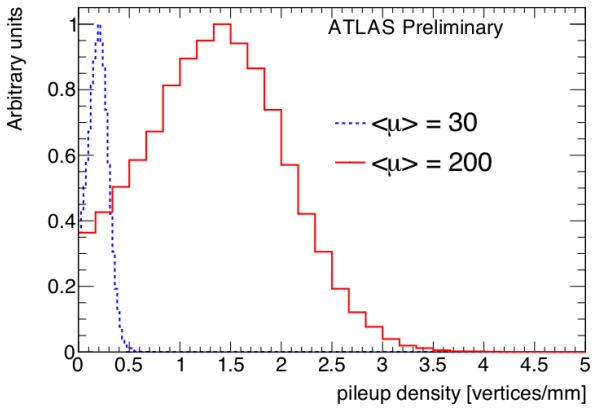


Figure 2.1: Local pileup vertex densities at generator level for two values of $\langle\mu\rangle$: $\langle\mu\rangle = 30$ and $\langle\mu\rangle = 200$.

the truth interaction time as a function of the z position, for one single Hard Scatter (HS) $t\bar{t}$ event with $\langle\mu\rangle$=200.

While the tracker resolution in z_0 is better than the typical distance between two vertices, the vertex reconstruction with ITk allows vertices to be separated. This is mainly the case in the central region (see Figure 2.6). When the z_0 resolution degrades, and becomes larger than the distance between two vertices, precision timing allows these vertices to be separated, reducing the density of vertices which are considered for a given track. The dispersion in time, for a given z, is visible in Figure 2.2.

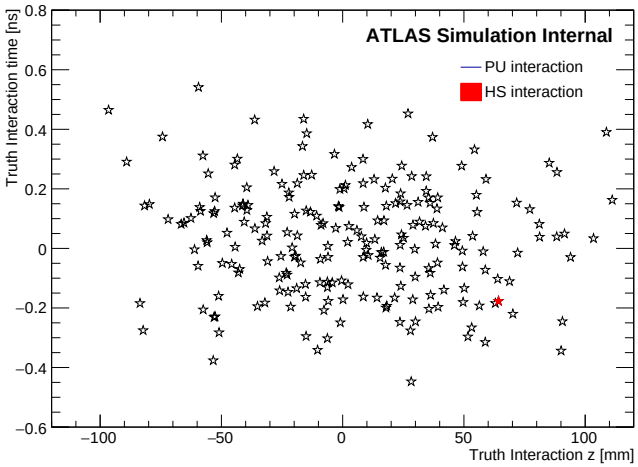


Figure 2.2: Visualisation of the truth interactions in a single bunch crossing in the z–t plane, showing the simulated Hard Scatter (HS) $t\bar{t}$ event interaction (red) with pileup interactions superimposed (black) for $\langle\mu\rangle = 200$.

2.2 Detector overview and requirements

The HGTD is being designed for operation with $\langle\mu\rangle = 200$ and a total integrated luminosity of 4000 fb^{-1}. Taking into account the space constraints of the existing ATLAS Experiment, including the more advanced planning for the tracker upgrade when R&D on the HGTD began, the HGTD will be located in the gap region between the barrel and the end-cap calorimeters, at a distance in z of approximately $\pm$ 3.5 m from the nominal interaction point. This region lies outside the ITk volume and in front of the end-cap and forward calorimeters, in the volume currently occupied by the Minimum-Bias Trigger Scintillators, which will be removed. The position of the two vessels for the HGTD within the ATLAS detector is shown in Figure 2.3.

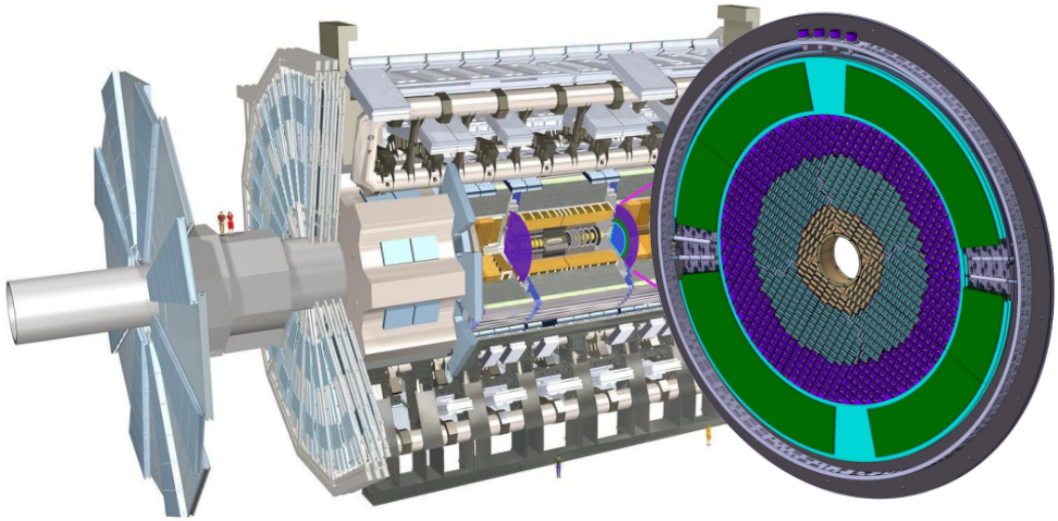


Figure 2.3: Position of the HGTD within the ATLAS Detector. The HGTD acceptance is defined as the surface covered by the HGTD between a radius of 120 mm and 640 mm at a position of $z = \pm 3.5$ m along the beamline, on both sides of the detector.

The envelope of the detector vessel has a radial extent of 110 to 1000 mm. The envelope in z is 125 mm, including the neutron moderator, supports, and front and rear vessel covers. A 50 mm-thick moderator is placed behind the HGTD to reduce the back-scattered neutrons created by the end-cap/forward calorimeters, protecting both the ITk and the HGTD. A silicon-based timing detector technology is chosen due to the space limitations. The sensors must be thin and configurable in arrays. In close collaboration with RD50 [4] and few manufacturers, an extensive R&D program is still ongoing. However baseline sensors that can provide the required timing resolution in the harsh radiation environments were already produced by three different vendors. LGAD [5] pads of 1.3 mm $\times$ 1.3 mm with an active thickness of 50 µm fulfil these requirements. This pad size ensures occupancies below 10%

at the highest expected levels of pileup, small dead areas between pads, and low sensor capacitance, which is important for the time resolution. The sensors will be operated at low temperatures (−30 °C) to mitigate the impact of irradiation.

A custom ASIC (ALTIROC), which will be bump-bonded to the sensors, is being developed to meet the requirements on time resolution and radiation hardness. The ASIC will also provide functionality to count the number of hits registered in the sensor and transmit it at 40 MHz to allow unbiased, bunch-by-bunch measurements of the luminosity and the implementation of a minimum-bias trigger. After optimising the layout for timing performance, cost and radiation tolerance, the detector described in this document is based on three active regions 120 mm $< r <$ 230 mm, 230 mm $< r <$ 470 mm, and 470 mm $< r <$ 640 mm providing in average 2.6, 2.4 and 2.0 hits per track respectively. Beyond $r >$ 640 mm are the peripheral electronics. The active area covers the pseudo-rapidity range $2.4 < |\eta| < 4.0$. A description of the detector layout optimisation is presented in Section 2.3.

Each HGTD end-cap is the integration of one hermetic vessel, two instrumented double-sided layers (mounted on two cooling/support disks), and two moderator pieces placed inside and outside the hermetic vessel. Each cooling/support disk is physically separated in two half circles. Furthermore, the layers are rotated in opposite directions with respect to one another by 15 to 20° in order to maximize the hit efficiency.

A global view of the various components of the detector and its main parameters are shown in Figure 2.4 and Table 2.1. The time resolution parameters have been optimised using information from the sensor (Chapter 5) and front-end electronics (Chapter 6) performance from lab and test beam measurements.

2.3 Detector layout and optimisation

The goal of the detector design is to provide the best possible time resolution in order to effectively suppress the effects of pileup in the forward region. The ability to associate tracks to primary vertices depends on the longitudinal impact parameter resolution of the ITk. The ITk layout is shown in Figure 2.5. Figure 2.6 shows the resolution, σ_{z_0}, of the longitudinal track impact parameter, z_0, measured by the ITk as a function of η, for muons with $p_{\mathrm{T}} = 1$ GeV and $p_{\mathrm{T}} = 10$ GeV. In this report, performance studies have been performed with an ITk layout and simulation [6] including a sensor pitch of 50×50 µm.

For good spatial separation of the HL-LHC collision vertices, σ_{z_0}, should be significantly better than the inverse of the average pileup density, 600 µm. Figure 2.6 shows that, in the central region, σ_{z_0} is well below this limit. In the forward region, however, the resolution exceeds the limit by a large factor, reaching 3 mm for particles with low transverse momentum at $|\eta| \approx 4$, due to the combination of geometric projection and, as shown in

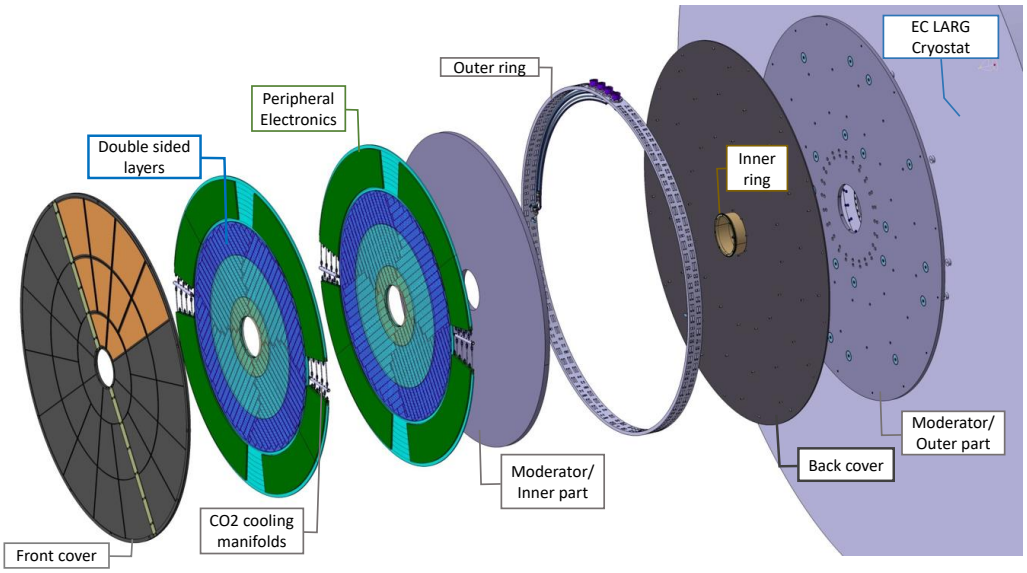


Figure 2.4: Global view of the HGTD to be installed on each of two end-cap calorimeters. The various components are shown: hermetic vessel (front and rear covers, inner and outer rings), two instrumented double-sided layers (mounted in two cooling disks with sensors on the front and back of each cooling disk), two moderator pieces placed inside and outside the hermetic vessel.

Pseudo-rapidity coverage	$2.4 < \|\eta\| < 4.0$
Thickness in z	75 mm (+50 mm moderator)
Position of active layers in z	± 3.5 m
Weight per end-cap	350 kg
Radial extension:	
Total	$110\,\text{mm} < r < 1000\,\text{mm}$
Active area	$120\,\text{mm} < r < 640\,\text{mm}$
Pad size	$1.3\,\text{mm} \times 1.3\,\text{mm}$
Active sensor thickness	50 µm
Number of channels	3.6 M
Active area	6.4 m^2
Module size	30 x 15 pads (4 cm × 2 cm)
Modules	8032
Collected charge per hit	> 4.0 fC
Average number of hits per track	
$2.4 < \|\eta\| < 2.7$ $(640\,\text{mm} > r > 470\,\text{mm})$	≈2.0
$2.7 < \|\eta\| < 3.5$ $(470\,\text{mm} > r > 230\,\text{mm})$	≈2.4
$3.5 < \|\eta\| < 4.0$ $(230\,\text{mm} > r > 120\,\text{mm})$	≈2.6
Average time resolution per hit (start and end of operational lifetime)	
$2.4 < \|\eta\| < 4.0$	≈ 35 ps (start), ≈ 70 ps (end)
Average time resolution per track (start and end of operational lifetime)	≈ 30 ps (start), ≈ 50 ps (end)

Table 2.1: Main parameters of the HGTD.

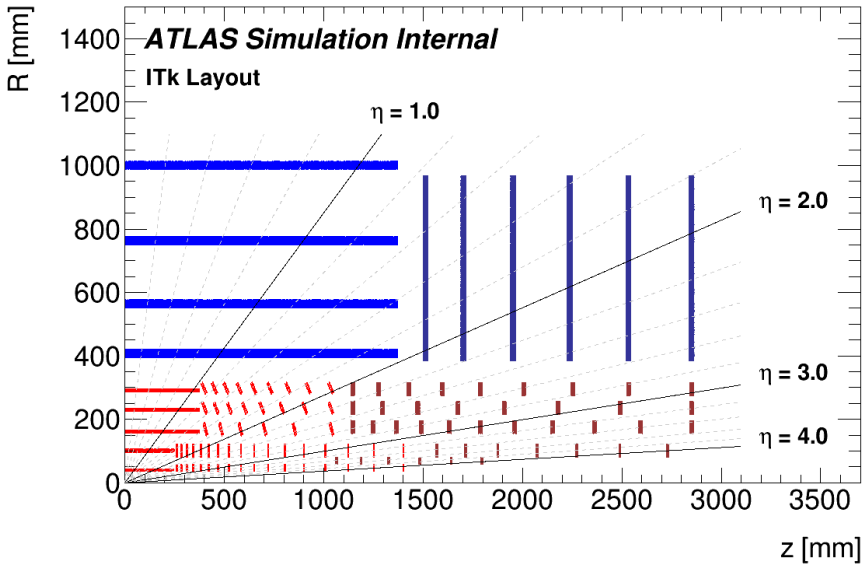


Figure 2.5: Schematic layout of the ITk for the HL-LHC phase of ATLAS. The active elements of the barrel and end-cap ITk Strip detector are shown in blue, for the ITk Pixel detector the sensors are shown in red for the barrel layers and in dark red for the end-cap rings. Here, only one quadrant and only the active detector elements are shown.

Figure 2.7, increased material. As a result, there is increased residual pileup contamination when assigning reconstructed objects to the reconstructed vertex.

The main contributions to the time resolution of a detector element are:

$$\sigma^2_{\text{total}} = \sigma^2_{\text{L}} + \sigma^2_{\text{elec}} + \sigma^2_{\text{clock}} \tag{2.1}$$

where σ^2_{L} are Landau fluctuations in the deposited charge as the charged particle traverses the sensor, σ^2_{elec} represents the contributions from the readout electronics, and σ^2_{clock} is the clock contribution. Beam tests and sensor simulations show that thinner silicon sensors reduce the contribution from Landau fluctuations. With a 50 µm thick LGAD sensor, this contribution amounts to approximately 25 ps. This is further discussed in Chapter 5. With fast detector signals and a high signal-to-noise ratio, the contribution from the electronics can be kept to approximately 25 ps. This is achievable only if applying corrections for the time walk induced by different signal amplitudes, using small bins in the time-to-digital conversion and applying precise in-situ inter-calibration. The details of the design of the readout electronics to achieve this are described in Chapter 6. The clock contribution should be kept below 10 ps; its distribution is discussed in more detail in Chapter 10.

For simplicity, the size used for the pads (single active pixel sensor) is the same for the entire HGTD, 1.3 mm × 1.3 mm. This pad size balances several characteristics. For smaller pad

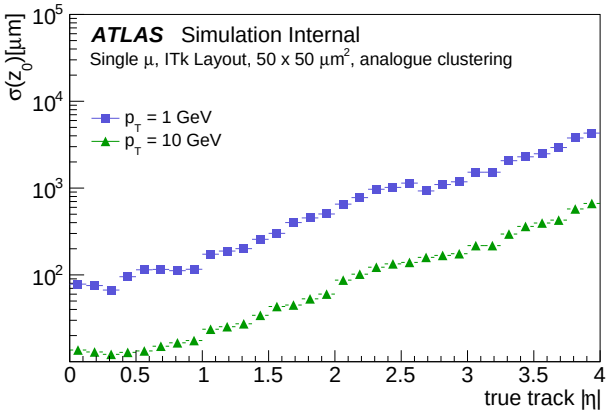


Figure 2.6: Resolution of the longitudinal track impact parameter, z_0, as a function of η for muons of $p_{\mathrm{T}} = 1\,\mathrm{GeV}$ and $p_{\mathrm{T}} = 10\,\mathrm{GeV}$ using ITk alone.

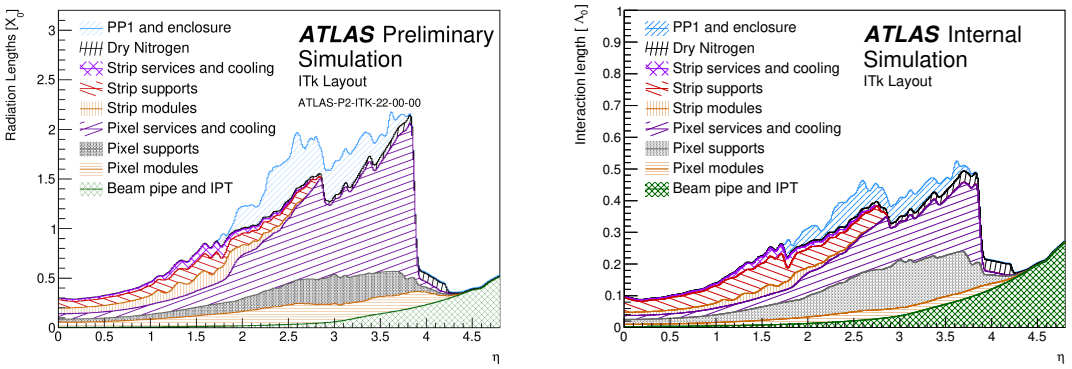


Figure 2.7: Material budget in radiation length X_0 (left) and nuclear interaction length λ_0 (right) as a function of pseudo-rapidity η, broken down by sub-system and material category for the ITk Layout [6] and beam pipe.

sizes, both electronic noise and physics occupancy are smaller, while the number of channels to be instrumented and the cumulative area of inter-pad dead zones are larger. The size was chosen to give a maximum occupancy of less than 10% in the modules exposed to the highest particle fluxes near the smallest instrumented radius. The choice also ensures a low double-hit probability for a single pad in one bunch crossing. Unifying the pad size across the entire detector also simplifies the production of sensors and assembly of the detector.

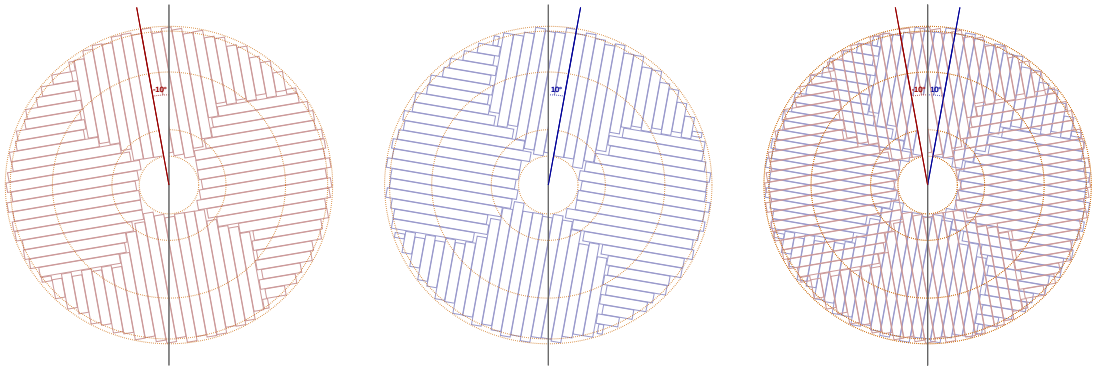
Each LGAD module contains 30×15 pads, for a total area of 4×2 cm^2. There are in total 8032 modules in the HGTD. The layout of modules was defined by maximising the coverage and minimising the effect of non-instrumented regions. The overlap between modules on the front and back of the disk was then optimised to give approximately uniform performance

as a function of radius.

The geometry of the detector has been optimised to approximate a flat timing resolution as a function of η. Due to radiation damage, the timing resolution of the detector will be degraded as the integrated luminosity delivered by the LHC increases. This radiation depends strongly on r, with higher radiation closer to the beam axis. The radiation levels expected for the full lifetime of the HL-LHC, including safety factors, are discussed in Section 2.4.

The readout rows are sets of modules whose flex cables (flexible PCB cables) are guided together towards larger radii to the peripheral on-detector electronics. They extend as a line of modules from lower to higher radii. The maximum length of the readout rows is limited by the manufacturing capabilities for the flexible circuits used for the data transmission. Their disposition for the first and second layer is shown as rectangles in Figure 2.8. The active width of a module is 39 mm which limits how well the area near the circular opening at 120 mm can be covered, however, for $r > 150$ mm the coverage is complete. The non-instrumented zone is 0.5 mm for each row edge to account for mechanical tolerances, adding up to 1 mm. Furthermore, an inactive region of 0.3 to 0.5 mm at the edge of each LGAD arrays is present. Adding up conservatively row and sensor edges, a dead region of 2 mm between rows is expected. The total effective width of a readout row is therefore 41 mm. These constraints lead to the helix structure shown in Figure 2.8. A particle transiting the detector should encounter multiple sensors as it passes through the two layers. Figure 2.8(a) shows the geometry of the first layer and Figure 2.8(b) shows the geometry of the second layer. The first and second layer are arranged to mirror the geometry of one another. Each of the layers is rotated in opposite directions by 15 to 20°. The baseline angle of 20° rotation between disks is shown in Figure 2.8(c). Any angle of rotation beyond 10° results in similar performance in terms of the number of simulated hits and dead regions. The baseline angle is chosen largely due to detector services considerations, which are further discussed in Chapter 12 and Chapter 13. Along with optimising the coverage, the rotation frees sufficient room at 640 mm to install the cooling equipment between the peripheral electronics.

Each layer of the HGTD is double-sided, i.e., the modules with sensors and on-detector electronics are mounted on the front and back sides of a common cooling disk. As illustrated in Figure 2.9, the modules on the two sides of a disk are arranged to overlap. A study using full simulation was performed to determine the optimal overlap between modules in three rings to achieve the required timing resolution via the average number of simulated hits given the expected time resolution of the pads. The maximal overlap is limited by the need for sufficient space between the modules to allow the readout of the data. For $r > 470$ mm, an overlap of 20%, for 230 mm $< r <$ 470 mm an overlap of 54% and for $r <$ 230 mm an overlap of 70% was the result of the optimisation. The HGTD acceptance is defined as the surface covered by the HGTD between a radius of 120 mm and 640 mm. The number of simulated hits as a function of radius and transverse plane position is shown in Figure 2.10.



(a) First layer (b) Second layer (c) Overlay

Figure 2.8: The orientation of the readout rows for the first and second layer encountered by a particle, separately and with the overlay of both. Each layer is rotated in alternating directions by 20°. In the figures the staves of the three rings are separated by the circular lines.

The relative fraction of tracks as a function of simulated hits per track for each ring can be found in Figure 2.11.

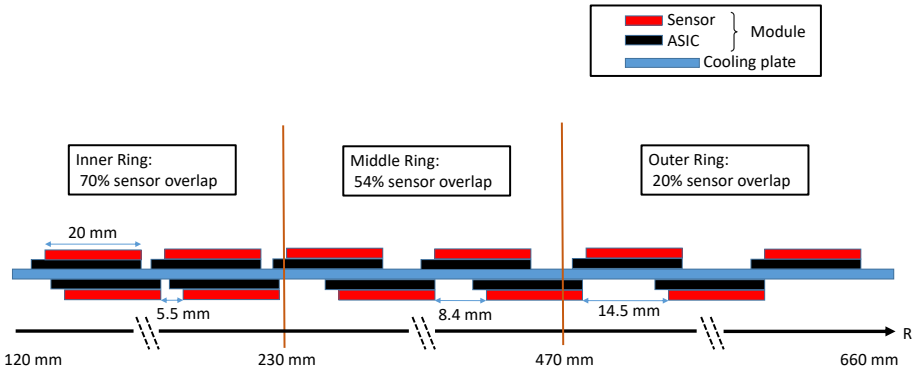


Figure 2.9: The schematic drawing shows the overlap between the modules on the front and back of the cooling disk. There is a sensor overlap of 20% for $r > 470\,\text{mm}$, 54% for $230\,\text{mm} < r < 470\,\text{mm}$ and 70 % for $r < 230\,\text{mm}$.

The material for the HGTD is highlighted in Figure 2.12, which includes the material for the moderator located behind the HGTD active sensor area.

Beyond pileup mitigation, HGTD can play an important role in the ATLAS HL-LHC physics programme as a luminometer. An accurate luminosity determination will be a critical input for precision measurements. The luminosity uncertainty can be a limiting factor to

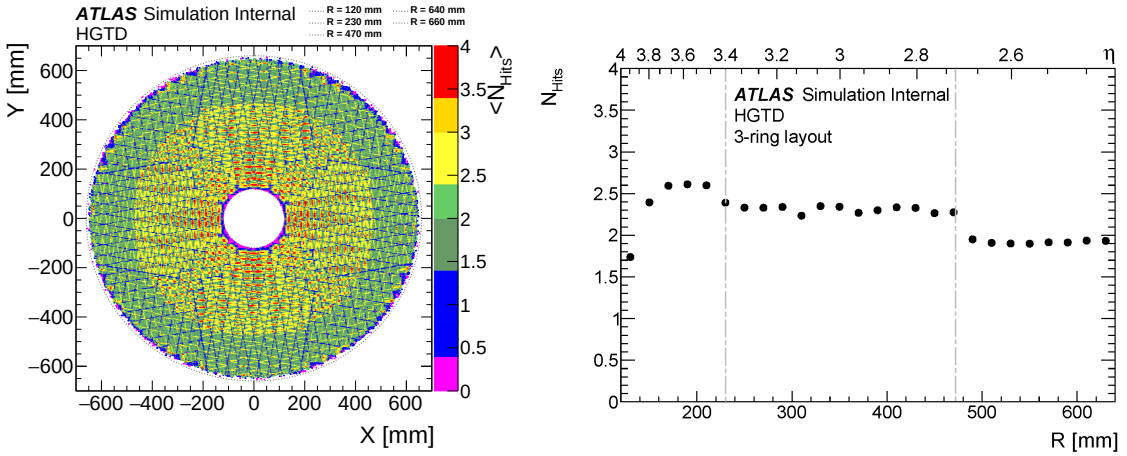


Figure 2.10: Hit multiplicity as function of x, y (left) and r (right). The figures were made using simplified simulations, resulting in an uncertainty of roughly 10 % compared to the full simulation studies of the HGTD, discussed in Section 3.1. The vertical grey dashed lines in the right plot shows the separation between the three rings.

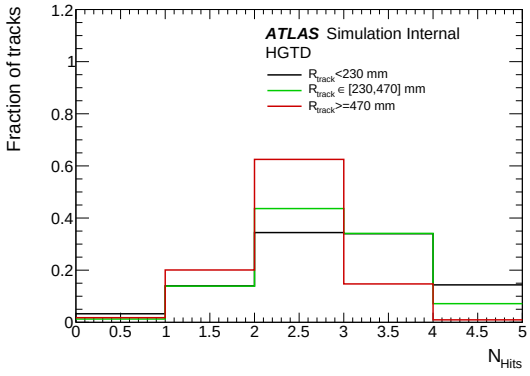


Figure 2.11: Fraction of tracks as a function of number of associated simulated hits, separated for tracks in the inner, middle and outer ring.

many precision cross-section measurements, including achieving $\mathcal{O}(1\%)$ accuracy on certain measurements of the Higgs boson production and couplings. It is therefore important to be able to determine the luminosity as accurately as in Run 2, which will be a challenge in the harsh HL-LHC environment. The HGTD provides unique and pileup-robust capabilities for measuring the luminosity at the HL-LHC and will be an essential part of the combined ATLAS luminosity measurement.

Taking advantage of the high granularity of the detector, the luminosity can be measured by counting the mean number of simulated hits in the detector, a quantity linearly proportional

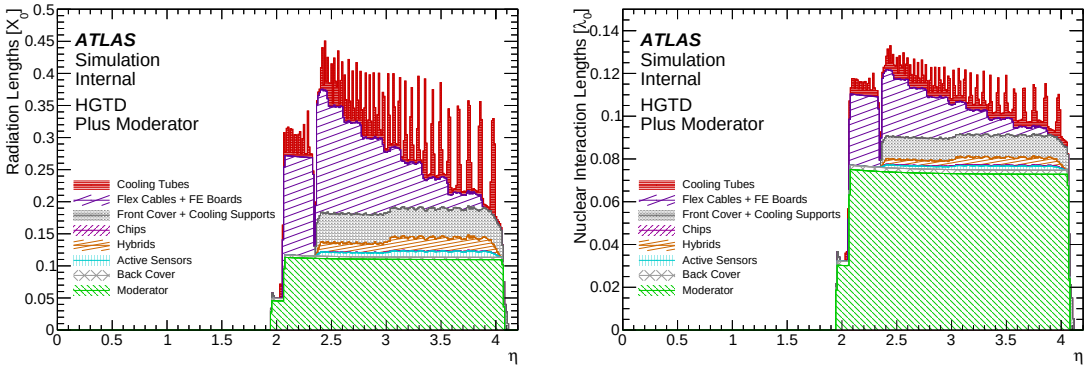


Figure 2.12: Radiation length X_0 (left) and nuclear interaction length λ_0 (right) as a function of pseudo-rapidity η, broken down by type of material for the HGTD, using the simulation of the two ring detector geometry described in Section 3.1.1. The moderator is included as it is within the hermetic vessel, although it is situated behind the active area of the HGTD. The baseline cooling pipes will be made with titanium instead of stainless steel as used in the simulation and material plots shown in this figure. The resulting radiation and nuclear interaction lengths will also be reduced with titanium cooling pipes.

to the average number of interactions per bunch crossing. The counting will be done over two time windows, one centred at the bunch crossing and with a width of 3.125 ns, the other with both width and relative position tuneable with a step of 3.125 ns. The application of these capabilities and their implementation are further discussed in Chapter 6 and Chapter 10.

The time resolution per track as a function of the radius is shown in Figure 2.13, for various integrated luminosity during the HL-LHC runs, corresponding to the replacements of the two inner rings during the lifetime of the detector. This replacement strategy is due to the expected radiation damage to the detector, described in detail in Section 2.4.

2.4 Radiation hardness

One of the most important parameters of the HGTD will be the radiation hardness of the sensors and electronics. Since the HGTD will be installed with a pseudo-rapidity coverage of $2.4 < |\eta| < 4.0$, it is essential that the detector can withstand the radiation levels throughout the HL-LHC operations. The neutron-equivalent fluence at a radius of 120 mm, is expected to reach $5.6 \times 10^{15}\,\mathrm{n_{eq}\ cm^{-2}}$ and the total ionising dose (TDI) about $5.6 \times 10^{15}\,\mathrm{n_{eq}\ cm^{-2}}$ as shown in Figure 2.14. To account for uncertainties in the simulation, a safety factor of 1.5 is applied to both estimates. An additional factor of 1.5 is applied to the TID due to uncertainties in the behaviour of the electronics after irradiation, primarily for low-doses-rate effects, which have not been fully qualified as of today. This leads to a total safety factor of 1.5 for the sensors that are most sensitive to the particle fluence, and 2.25 for the

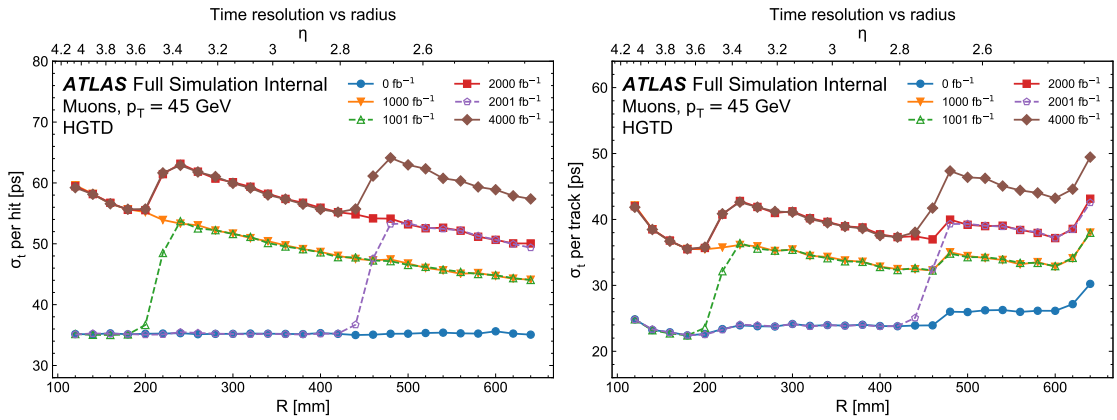


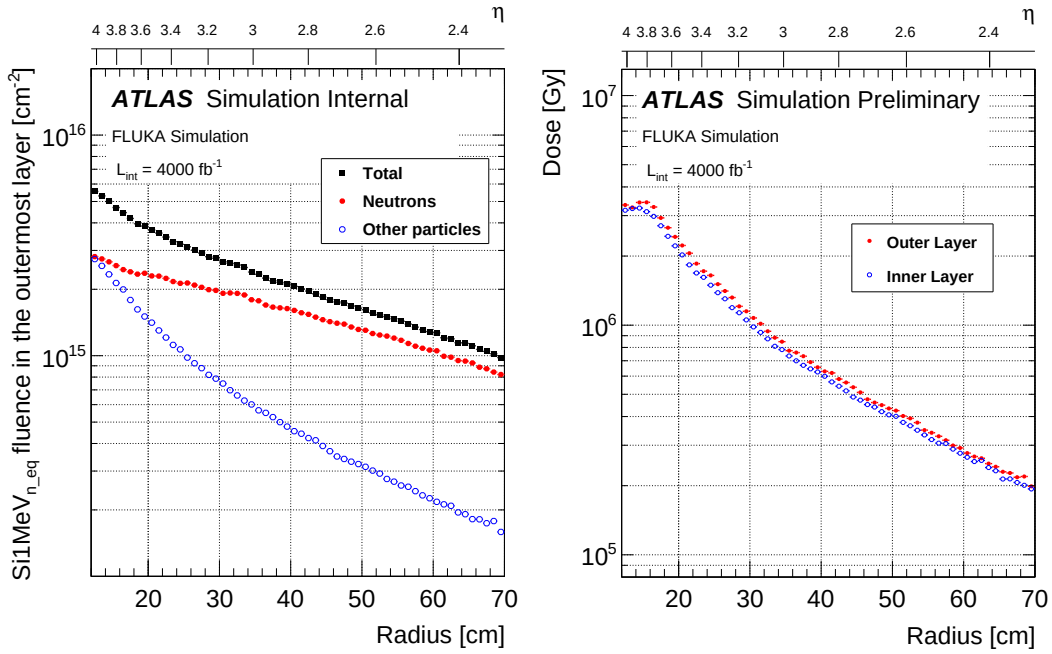
Figure 2.13: Time resolution per hit (left) and per track (right) within HGTD acceptance as a function of the radius. The time resolution is shown for various integrated luminosities. The time resolution is improved at higher luminosities corresponding to the replacements of inner-most rings during the lifetime of the detector.

electronics which are more sensitive to the TID. After applying these, the detector would need to withstand $8.3 \times 10^{15}\,n_{eq}\,cm^{-2}$ and 7.5 MGy.

To achieve sufficient performance of the sensors and ASICs, the detector layout has been designed considering a replacement scenario during the HL-LHC. Through an intensive R&D campaign described further in Chapter 5 and Chapter 6, a minimum charge of 4 fC is required to obtain a high efficiency signal. This can be achieved up to a radiation damage of $2.5 \times 10^{15}\,n_{eq}\,cm^{-2}$ and 2.0 MGy. As a result, the sensors and electronics within the lowest-radius ($r < 230$ mm) will be replaced after each 1000 fb^{-1} and the sensors and ASICs within 230 mm $< r <$ 470 mm should be replaced at half of the data-taking (2000fb^{-1}) during the HL-LHC program. This corresponds to about 52% of the sensors and ASICs which will need to be replaced. The maximum fluence and total ionising dose as a function of the radial position including the replacement of the rings can be found in Figure 2.15. In the resulting three-ring layout, the maximal TID and fluence, using the Fluka estimations of September 2019, does not exceed 2 MGy and 2.5×10^{15} neq/cm^2. In the inner ring the total Si 1MeV neq has a similar contribution from neutrons and charged particles while in the middle and outer rings the dominant effect comes from neutrons.

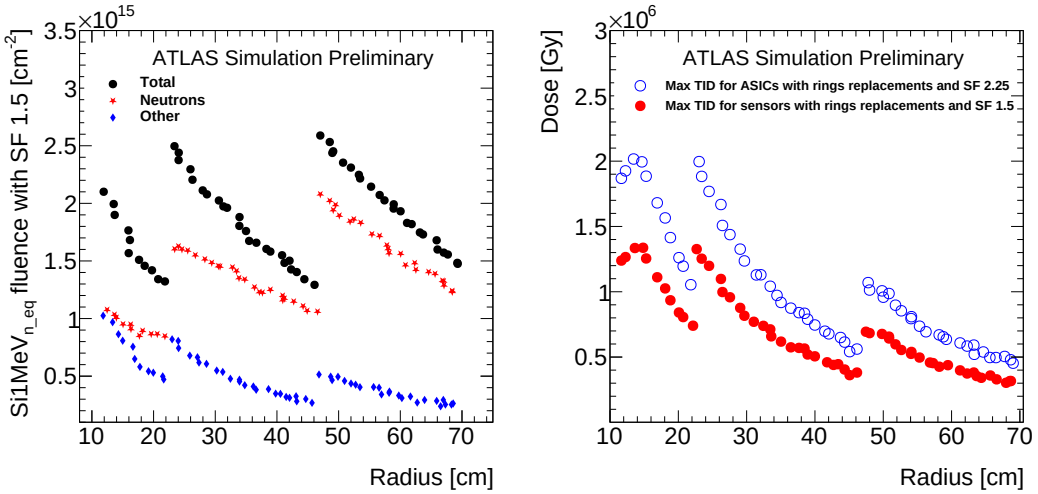
The exact radial transition between the three rings will be tuned for the final detector layout, once the FLUKA simulations will be updated with the final ITk layout, and the radiation hardness of the final sensors and ASICs are re-evaluated.

More details can be found in Chapter 5 to Chapter 6. The expected proton, neutron, and pion energy spectra in the HGTD front and rear layer after 4000 fb^{-1} are shown in Figure A.1, Figure A.2, and Figure A.3.



(a) Nominal Si1MeV$_{n_{eq}}$ fluence for HL-LHC. (b) Nominal ionising dose for HL-LHC.

Figure 2.14: Expected nominal Si1MeV$_{n_{eq}}$ fluence and ionising dose as functions of the radius in the outermost sensor layer of the HGTD for 4000 fb^{-1}, i.e. before including safety factors. The contribution from charged hadrons is included in 'Others'. These estimations used Fluka simulations using ATLAS Fluka geometry 3.1Q7 (from December 2019).



(a) (b)

Figure 2.15: Expected Si1MeV$_{n_{eq}}$ radiation levels in HGTD, using Fluka simulations, as a function of the radius considering a replacement of the inner ring every 1000 fb^{-1} and the middle ring replaced at 2000 fb^{-1}. For the radiation levels, the particle type is included and the contribution from charged hadrons is included in 'Others'. These curves included a factor of 1.5 to account for simulation uncertainty. An additional factor of 1.5 is applied to the TID to account for low dose rate effects on the electronics, leading to a SF = 2.25.

3 Performance and Physics Benchmarks

One of the most significant ATLAS detector upgrades for HL-LHC is the replacement of the inner tracker system and its extension in pseudorapidity coverage up to $|\eta| = 4$. The forward extension provides object-level improvement of jets from vector boson fusion (VBF), vector boson scattering (VBS), and many other key signatures for the HL-LHC physics program. However, exploiting forward tracks is challenging. As η increases beyond the acceptance of the current tracker ($|\eta| < 2.5$), tracks become more collinear to the beam and are subject to large multiple scattering effects due to increased material from services relative to the central barrel region. While the ATLAS Phase 2 inner detector is able to reconstruct charged particles up to $\eta = 4$ with very high precision, the above effects are particularly relevant for the case of soft pileup forward tracks that may contaminate the hard-scatter vertex. At very low track p_T, there is a large and rapid degradation of the longitudinal impact parameter z0 resolution as a function of η to the point in which the z0 resolution of pileup tracks is larger than the typical separation between primary vertices. This effect significantly weakens the ability of the tracking detector to unambiguously associate low p_{T} tracks to vertices, resulting in reduced physics performance capabilities in the forward region. In other words, for the first time, a high luminosity hadron collider will operate a forward tracker in an environment in which the pileup density is higher than its spatial resolution in z for low p_{T} tracks. HGTD has been designed primarily to overcome this challenge, ensuring that the physics performance, particularly the pileup suppression, does not degrade in the forward region. This is achieved by leveraging the time spread of the LHC beam spot with a fast timing detector that can associate time stamps to forward tracks. The capability to provide high-precision time measurements for charged particles allows the HGTD to enhance the performance of physics object reconstruction in the forward region, complementing the ITk in the forward region. Those object-level improvements can then, in turn, increase the physics potential of ATLAS.

The reconstruction of track times in the forward region, on the other hand, is also a challenging task. The two main experimental challenges of a forward timing detector are the large amount of material in front of it, and the limited η acceptance of the HGTD in the space available between the ITk and the end-cap calorimeter. The former limits the rate of forward tracks that can be associated to a time with high confidence. The latter impacts the ability to determine the global event-vertex time. The hard-scatter interaction needs to produce enough particles within the HGTD acceptance, and separate them from forward activity from pileup interactions. In this section the focus is on both the event reconstruction and

physics improvements introduced by HGTD, as well as on the challenges associated to the reconstruction and use of timing information in the forward region.

This chapter is organised in four sections. In the first section, the HGTD simulation and the modeling of low-level performance are described, such as the timing distribution and detector occupancy at the level of simulated HGTD hits. The second section describes the reconstruction algorithms developed to associate HGTD hits to tracks found by the Inner Tracker (ITk) and thereby assign them times, and how those times are used to determine the times of primary pp vertices. Detailed studies are performed using simulated samples of single-particle events as well as from full physics events with an average of 200 additional pileup interactions overlaid. Since the ability to correctly assign times to tracks is key to the higher-level performance for physics objects, particular attention is devoted to the understanding of the hit-to-track matching efficiency, mis-tag rate and to identify the main factors that limit performance. After describing the low-level performance, the third section discusses the application of the newly available track and vertex times to improve the reconstruction of jets and leptons in the forward region. In particular, the focus is on the improvements in pileup-jet rejection, and lepton isolation efficiency, but possible additional applications are also discussed. The final section illustrates how the improvements in object-level performance can enhance the sensitivity of the ATLAS physics programme through a few example studies. Two main broad classes of physics analyses are considered, motivated by the specific physics object performance improvements studied: Vector Boson Fusion final states, which benefit from the increased pileup-jet rejection in the forward region, and the measurement of the weak mixing angle, which leverages the improved forward lepton isolation efficiency. Additional physics applications, including the potential to significantly constrain the luminosity uncertainty are also discussed.

3.1 Simulation and hit-level detector response

The full simulation of the HGTD is performed using a software release dedicated to the HL-LHC ATLAS upgrade programme. The production of simulated samples follows the same steps as the regular ATLAS offline software chain [7]: event generation, detector simulation, digitisation of simulated energy depositions into detector read-out data, and event reconstruction[1].

The detector simulation uses a layout of the HGTD ("two-ring") which is very close to the baseline layout ("three-ring") summarized in Chapter 2 and described in more detail in the other sections of this report. The layout described in this chapter, and included in the simulations, is not identical due to the layout optimization undertaken in parallel to these simulation studies. The number of modules in the two scenarios is within 1% of each other (with slightly fewer modules in the simulation in comparison to the detector description

[1] Details about the list of samples prepared for the studies in this document can be found in Appendix B.

in Table 2.1) with the same η coverage. Therefore it is expected that the performance and physics potential will be similar for both layouts. Plots of the detector geometry used in the simulation will be shown in Section 3.1.1, similar to those in Section 2.2, for a direct comparison.

The moderators downstream of the active HGTD detector elements are included, with a total thickness of 50 mm. The front and back covers, as well as the heaters, are also included. The cooling plates are modeled in detail, including the cooling loops modeled as steel tubes filled with liquid CO_2, and the support plates for the modules.

3.1.1 Detector geometry

The GEANT4 toolkit [8] is used to simulate the ATLAS detector. The simulation uses dedicated `GeoModel` packages [9] to implement the detector geometry and convert it to GEANT4 volumes. As particles are propagated through this geometry, the various interactions between the particles and the detector material are simulated. In sensitive detector elements, processes ranging from energies of a few eV, such as the ionisation in gases, up to TeV energies, are simulated to provide a detector-response model that is as realistic as possible. The simulation propagates particles step-wise through the material of the volumes in the detector model and produces energy depositions at specific points in space and time.

In each HGTD end-cap, there are two cooling plates with silicon sensors mounted on both sides totaling to four individual active layers. The detector description of the HGTD has been extended to include approximate volumes representing the peripheral electronics at radii greater than 640 mm. For illustration, a simulated event is shown in Figure 3.1 visualising the placement of the individual modules along with tracks representing the trajectories of simulated charged particles.

The detector modules each consist of one sensor and two readout ASICs (see Chapter 4). These are simulated as boxes of size 22 mm $\times$ 40 mm with silicon sensors of size 20.5 mm $\times$ 40 mm placed flat in the x–y plane, corresponding closely to the actual sensor size which is discussed in Chapter 5. The modules are larger than the sensors in one dimension to provide the margin needed for wire-bonding the ASIC to the flex. The total thickness of the silicon sensor is 250 µm of which the active part makes up 50 µm and the passive part 200 µm in agreement with the chosen LGAD technology.

The flex PCB cables connecting the ASICs to the peripheral electronics beyond 640 mm have also been implemented in the simulation. As the total thickness of these cables increases as a function of radius, the contribution of the flex cables can be seen clearly in the HGTD material distribution in Figure 2.12.

As mentioned at the beginning of this section, due to the evolving detector design, the module layout used in the full simulations differs from the nominal layout described in

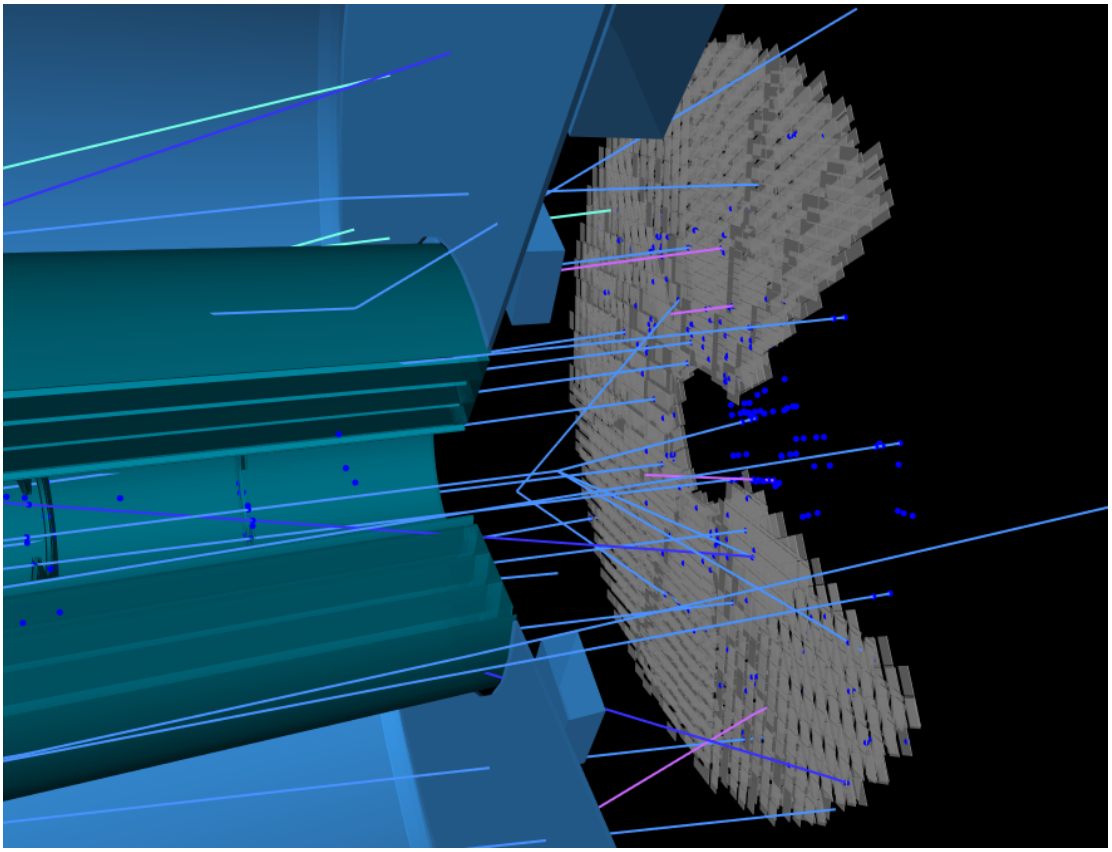


Figure 3.1: Visualization of a simulated QCD dijet event showing the trajectories of charged particles and the resulting simulated hits in the HGTD. A wedge in ϕ and volumes representing services and support structures have been removed to expose the individual detector modules of the HGTD. No pileup interactions were overlaid in this simulated event.

Section 2.3. The geometry implemented in the GEANT4 detector description is a two-ring layout, Figure C.1 in Appendix C shows the readout row orientation of the two-ring layout used for the full simulations. The detector description includes 80% overlap between sensors on front and back sides of a cooling plate at $R < 320$ mm, and 20% outside, as shown in Figure C.3 in Appendix C . This can be compared with the overlap regions described in Figure 2.9 for the three-ring layout.

Figure 3.2 shows the average number of HGTD simulated hits per track in simulated events for the two-ring layout, as a function of the X and Y position as well as the radial distance from the beam axis, which is close to the one shown for the nominal layout in the previous chapter (Figure 2.10). The minor difference in the detector geometry description will only have a minor influence on the studies of physics objects and analyses, due to the nearly identical geometric coverage and the very similar numbers of simulated HGTD hits per track across different pseudorapidity regions. The three-ring layout will in the future be

propagated to the GEANT4 detector description and allow simulation studies with a layout fully consistent with the nominal design.

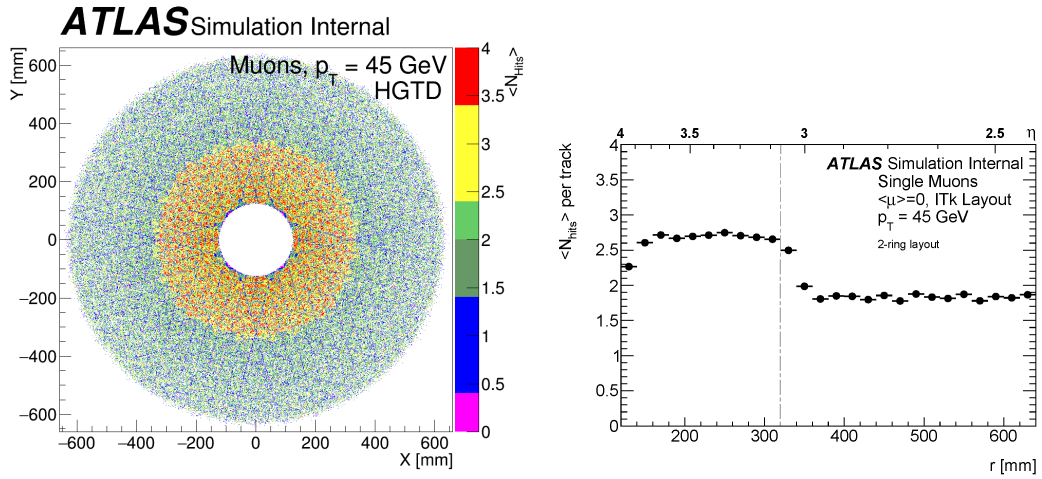


Figure 3.2: The average hit multiplicity as a function of the X and Y position as well as radius (and pseudorapidity) is shown for the two-ring detector layout used in the simulation, for muons with a p_{T} of 45 GeV. The vertical grey dashed line in the right plot shows the separation between the two rings. This can be compared to Figure 2.10 for the three-ring layout.

3.1.2 Sensor simulation

A pad size of 1.3 mm × 1.3 mm is used in the simulations. Two sources of inefficiency over the surface of the modules are implemented in the simulations:

- the guard ring of 0.5 mm surrounding the edge of the sensor, and
- the inter-pad dead zones of 50 µm between active pads.

As a result, 79% of the total silicon area is active. The different zones of the sensors are illustrated in Figure 3.3 showing the positions of energy depositions from single-particle simulations in active and non-instrumented regions.

The sensor simulation, just like the digitisation and reconstruction described below, is implemented using software developed for the pixel-based tracking detectors in ATLAS, which also provides functionality for associating truth information to the simulated detector hits.

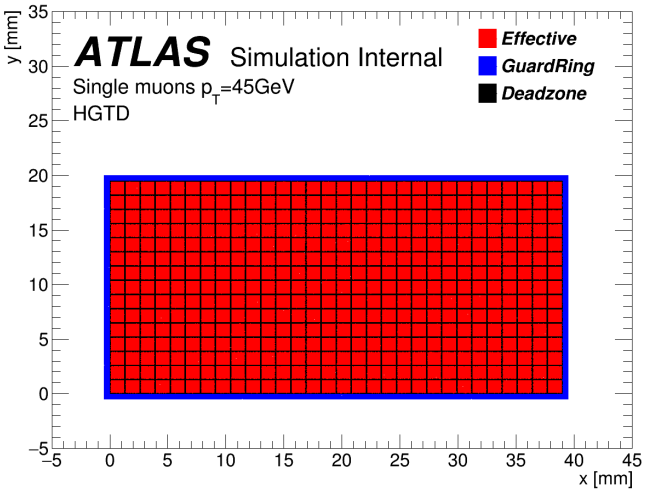


Figure 3.3: Positions of simulated energy deposits in active detector regions (red), inter-pad dead zones (black) and the guard ring surrounding the edges of a sensor (blue), given in local module coordinates x and y. This figure is made through drawing points for individual GEANT4 energy depositions from single-particle simulations and shows the level of detail implemented in the geometry model used.

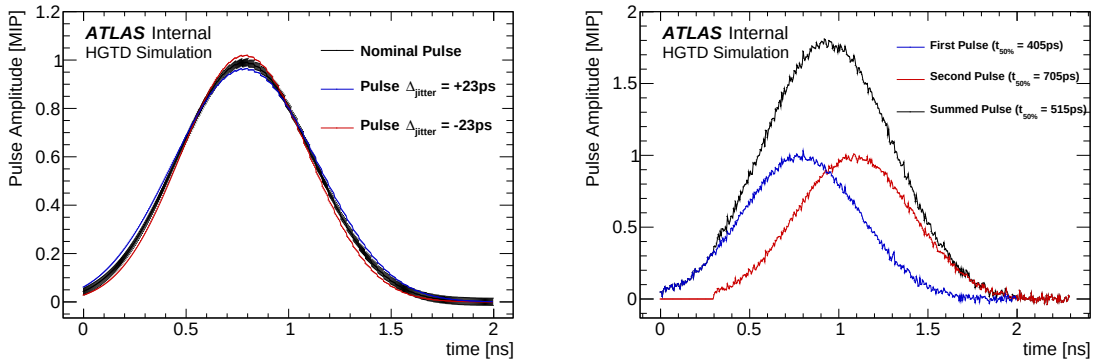
3.1.3 Simulation of digitised readout signals

The GEANT4 energy depositions are processed in a digitisation step in order to emulate the detector electronics and generate the detector readout signals. The LGAD sensors are described as planar n-in-p pixel sensors with electron carriers. The channel efficiency is simulated as perfect for energy depositions above threshold, and no defects from radiation damage or defective hardware are included.

During digitisation, the energy deposited for each GEANT4 step in the active silicon volume is used to evaluate the free charge and the drift time to the readout surface based on the sensor thickness, carrier mobility, depletion and bias voltages, and Lorentz shift. Given the characteristics of the sensors, the capacitive coupling to nearby pixels (i.e. cross-talk) is considered small, matching conclusions from LGAD beam tests and discussed in Section 5.5.4. Any cross-talk effects are therefore neglected for now.

In the digitisation step, each energy deposition is also used to generate a pulse following a shape extracted from beam tests of LGAD sensors, see Figure 3.4. Figure 3.4(a) shows the nominal pulse shape and Figure 3.4(b) shows the result of two particles passing through the same pad, separated by 300 ps. A convolution of a Gaussian and a Landau distribution was found to give a good description of the pulse shape. The amplitude of the simulated pulse and its location in time are determined by the magnitude and time of the GEANT4 energy depositions. On top of this pulse, the electronic noise as measured in test beam studies [10]

is added to each pulse bin, as variations randomly sampled from a Gaussian centered at zero and with a standard deviation corresponding to 1.5% of the mean pulse amplitude of a MIP. The pulse time is extracted from the leading edge of the pulse, thereby modelling one component contributing to the total timing resolution. The impact of Landau fluctuations on the overall timing resolution is modelled via an additional Gaussian smearing of the pulse shape derived after the previous steps, which contributes to a timing resolution of about 20 ps. The resolution contributed by the electronics is modelled by smearing the signal time with two Gaussian functions, one reflecting the clock jitter and time-walk contribution from the readout system ($\sigma = 25$ ps with no irradiation) and the other for the clock distribution ($\sigma = 15$ ps). The simulation of timing resolution is set up according to the specification detailed in Section 2.3, and the total resolution for the case of no irradiation contributed from the above sources is about 35 ps per hit. The impact due to the overlapping hits on the expected occupancy is found to be minor and illustrated later in Section 3.1.4.



(a) (b)

Figure 3.4: The simulated pulse shape in a pad of the HGTD is shown for (a) one particle and (b) two particles passing through the same pad separated by 300 ps. The impact on the shape due to the clock jitter variations of $\pm$ 23 ps is given in (a). The particles under study are traversing the HGTD perpendicularly.

Samples of single muons and pions with $\langle\mu\rangle = 0$ and flat η distributions have been simulated to study the expected time distribution for HGTD hits. The distribution of the time measurements of the energy depositions from single-muon simulations in the HGTD sensors is shown in Figure 3.5, corresponding to the timing performance expected before any radiation damage. The time distribution is obtained by taking the time of the deposition, subtracting the time-of-flight (TOF) expected for a particle with $\beta = 1$ travelling from the production vertex to the sensor in a straight line, then subtracting the true time of the primary vertex. This simplified approach was taken to demonstrate the overall timing structure with the available information at the level of hit simulation. The general TOF correction used for track-hit assignment at the reconstruction level and for the later studies of objects and physics analyses is detailed in Section 3.2.1.

For all layers, the depositions originating from primary and secondary particles are in time for the bulk of the distribution, where primary and secondary particles are those produced from hard-scattering vertices and material interactions, respectively. However, the distribution for secondaries also features a pronounced tail in the timing distribution, arising from secondary particles with low momentum and/or longer path length.

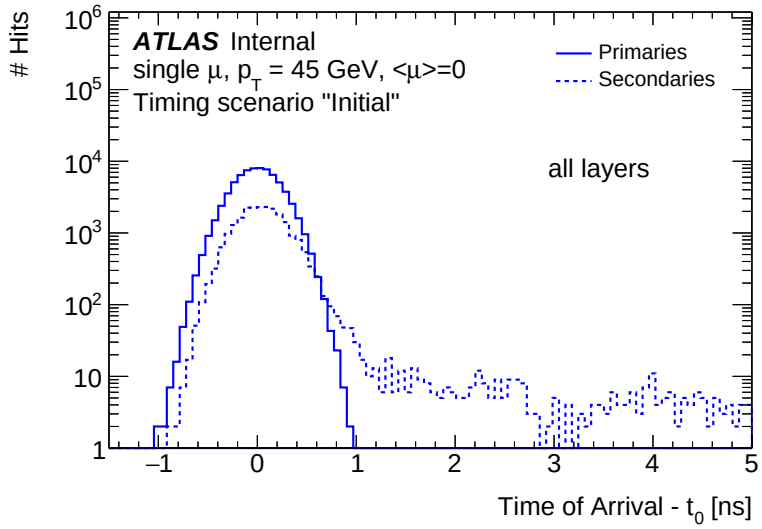
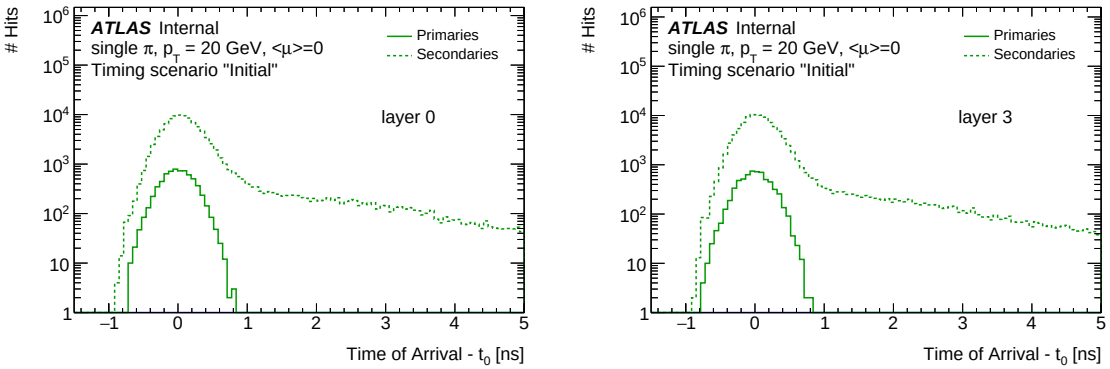


Figure 3.5: The distribution of times of simulated energy depositions after applying corrections for the expected time-of-flight from the origin to the sensor and the time of the primary vertex (t_0). Simulated hits originating from primary and secondary particles are shown from a single-muon sample without pileup.

The simulated hit-time in the first and last layer of the HGTD is shown in Figure 3.6 for single-pion events. The structure is similar in all layers, and simulated hits for the bulk of the primary and secondary particles are within a narrow time window with respect to the true arrival time. The distributions from secondary particles exhibit significant tails towards larger times, more pronounced for the pions which undergo hadronic interactions within the ITk and the material upstream of the HGTD than for the single-muon events in Figure 3.5. The secondary-hit distributions have larger magnitudes compared with those from primary hits, due to the fact that all hits were plotted without any selection. For a realistic evaluation of performance for physics objects and analyses, the impact from secondary hits can be much reduced, with a proper selection criterion (see Section 3.2) .

The digitisation software may also be used to emulate the expected timing performance at any point during the HL-LHC programme. With increased integrated luminosity, the detector gradually suffers more radiation-induced damage which degrades the timing performance. To mitigate this, modules are replaced according to the replacement scheme discussed in Section 2.4. As a consequence, the per-hit and per-track timing resolution will generally vary as a function of radius and integrated luminosity in a non-trivial way, which



(a) Layer 0 (the first layer)

(b) Layer 3 (the last layer)

Figure 3.6: The distribution of times of simulated energy depositions after applying corrections for the expected time-of-flight from the origin to the sensor and the time of the primary vertex (t_0). Simulated hits originating from primary and secondary particles are shown from a single-pion sample without pileup.

was demonstrated in Section 2.3.

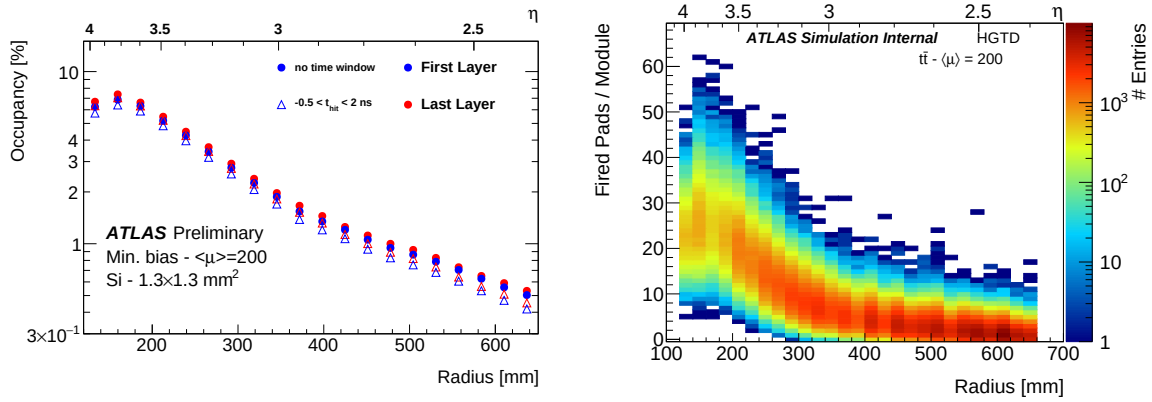
The contribution to the timing resolution due to radiation damage is taken into account as a function of the position of the sensor and the accumulated integrated luminosity with a Gaussian smearing. The doses used are those computed using FLUKA simulations in Section 2.4, then data from test bench measurements of sensors determine the corresponding gain for the sensor and the resulting degradation in per-hit timing resolution, based on measurements with ALTIROC0.

For all studies in this chapter, unless specified otherwise, the timing resolution (35 ps per hit) corresponding to the initial running condition at HL-LHC is used in the simulation, hereafter referred to as the "initial" timing scenario.

3.1.4 Occupancy

The hit occupancy is studied using simulated minimum-bias and $t\bar{t}$ events with a pileup of $\langle\mu\rangle = 200$. As expected from the particle flow as a function of rapidity in hadron collisions, the probability to have a hit in a pad (with fixed pad size as a function of radius) decreases as a function of the distance from the beam axis. To reduce the probability that an individual pad is traversed by several particles in the same event, a maximal occupancy of less than 10% is required (Section 2.3). This is achieved, but with the smallest margin around $R = 160$ mm, where an occupancy of 8% is observed. Figure 3.7(a) shows the hit occupancy expected for the minimum-bias events, defined as the percentage of pads in the HGTD registering a hit, for the HGTD baseline pad size of 1.3 mm $\times$ 1.3 mm. Compared to the innermost layer there is a slight increase for the outermost layer, primarily caused by the increased probability of

initiating showers due to hadronic interactions as more material is traversed. In Figure 3.7(b) the distribution of the number of pads in a module with signal is shown as a function of the radius for $t\bar{t}$ events. The variation of the number of pads with signal in a module has to be taken into account in the calculation of the bandwidth for the data transfer to the peripheral electronics.



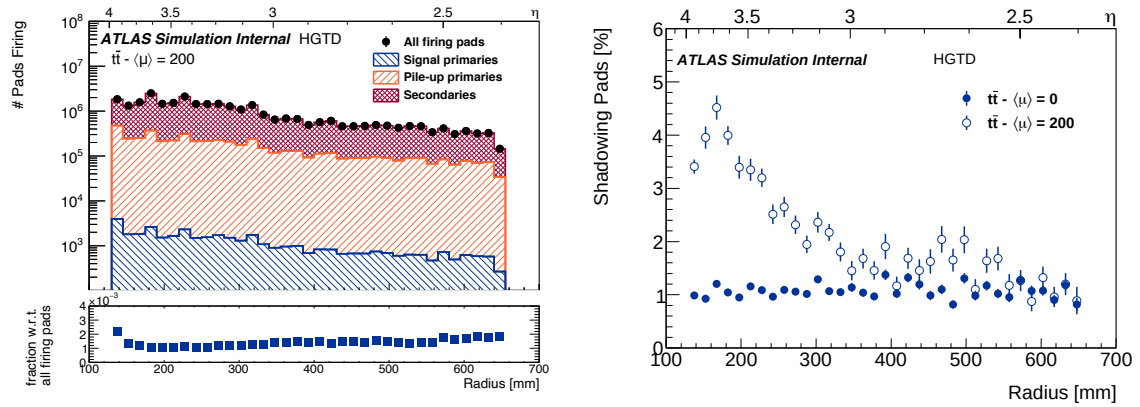
(a) Occupancy

(b) Number of fired pads per module

Figure 3.7: The occupancy (a) and the number of fired pads (b) per module are shown as a function of the radius for a pad size 1.3 mm × 1.3 mm at a pileup of $\langle\mu\rangle = 200$.

By studying the GEANT4 truth information it is possible to study the types of particles giving rise to hits. In $t\bar{t}$ events with $\langle\mu\rangle = 200$, pileup particles and secondaries from showers created in the upstream detector material dominate the occupancy. Figure 3.8(a) shows the breakdown of the origin of the hits measured in the HGTD within a window of ± 1 ns centered around the time of the primary particles as a function of the radius.

If two particles deposit energy in the same pad, the signal of one can be missed or be deformed by a signal from another particle that arrives earlier, and this effect is referred to as "shadowing". It is therefore important to evaluate the number of hits from primaries masked by particles arriving earlier (Figure 3.4(b)). Figure 3.8(b) shows the percentage of pads fired by secondaries and pileup particles (also consisting of primary and secondary particles from pileup) shadowing a primary particle with respect to the number of pads where at least one primary particle has deposited energy within the 2 ns window. For this high-pileup sample the percentage of shadowed pads is 4.5% at low radii where the occupancy is maximal, decreasing to 1% at larger radii. Performing the same analysis for $\langle\mu\rangle = 0$ shows that the level of 1% is due to secondary particles originating from the same hard-scatter interaction and characterised by a time of arrival compatible within the timing resolution. In this limit the shadowing effect does not bias the time measurement.



(a) Signal origin

(b) Shadowing percentage

Figure 3.8: (a) The origin of the hits detected in the HGTD, in the lower panel the ratio of the primaries with respect to all firing pads is shown, and (b) the percentage of shadowed pads with respect to all firing pads as a function of the radius.

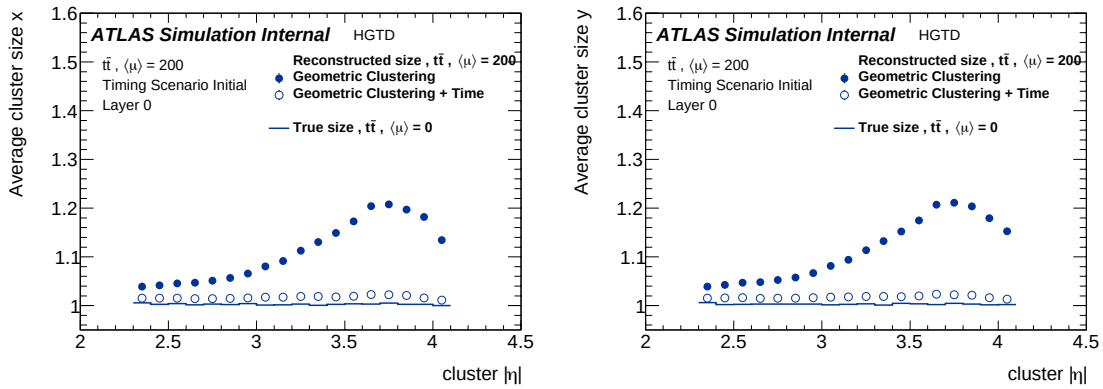
3.1.5 Detector-level reconstruction

Three different clustering approaches were studied:

- Geometric clustering: this approach groups adjacent fired pads to form clusters neglecting the time measurement associated to them. No timing information is exploited in this method, and it is identical to the method used in the ITk pixel reconstruction.
- Geometric clustering with time filtering: as above but also using time measurements. Adjacent pads are grouped if the time difference of the considered channels is smaller than 30 ps, corresponding to an overly tight constraint to accentuate any effect from adding timing.
- No clustering: each fired pad is converted into a cluster object.

Figure 3.9 shows the reconstructed cluster size (number of adjacent fired pads) in module-local[2] x and y using the geometric clustering algorithm with and without time filtering for simulated $t\bar{t}$ events with a pileup of $\langle\mu\rangle = 200$. The reconstructed cluster size is compared to the "true" cluster size, defined as the size of clusters arising from the same simulated particle reconstructed with the geometric clustering with time filtering for $t\bar{t}$ events without pileup. When time consistency is required, smaller clusters are obtained with an average size close to one pad in local x and y coordinates. Despite the 30 ps time window corresponding to a very tight requirement, the size of the resulting clusters still deviate more from that of the true clusters than when using single-pad clusters (i.e. no clustering). Figure 3.10

[2] The local x and y coordinates represent the two coordinates along the sensor grid. Local x is in the $R\phi$ plane perpendicular to the beam line while local y points radially in R.



(a) Average cluster width in local x (b) Average cluster width in local y

Figure 3.9: Average cluster width (number of adjacent fired pads) in module-local (a) x and (b) y coordinates for clusters obtained with the geometric clustering algorithm with and without time filtering, in $t\bar{t}$ events with $\langle\mu\rangle = 200$. The reconstructed cluster sizes are compared to the "true" cluster sizes, defined with the clusters arising from the same simulated particle reconstructed with the geometric clustering with time filtering for $t\bar{t}$ events without pileup.

shows the probability of merging contributions originating from multiple particles into the same cluster with the geometric clustering algorithm with and without time filtering in the same $t\bar{t}$ sample. In order to correctly associate clusters to tracks and provide good timing measurements, the rate of merging multiple contributions into one cluster should remain as low as possible. Based on these results, the no-clustering option was chosen as input to the track reconstruction in the studies for this TDR.

3.2 Reconstruction and detector performance

This section discusses the performance of the HGTD relating to reconstructed tracks and primary vertices. The assignment of times to tracks is discussed in Section 3.2.1, after which the methodology and performance of assigning times to primary vertices is presented in Section 3.2.2. A thorough understanding of these basic ingredients is critical for the later discussion of improvements on object performance and physics results.

3.2.1 Association of HGTD timing measurements to tracks

This section describes the techniques developed to assign a time stamp to reconstructed tracks. The algorithm is based on the progressive extrapolation of tracks to the active HGTD surfaces, in each surface the association to the tracks of nearby hits in the HGTD is performed.

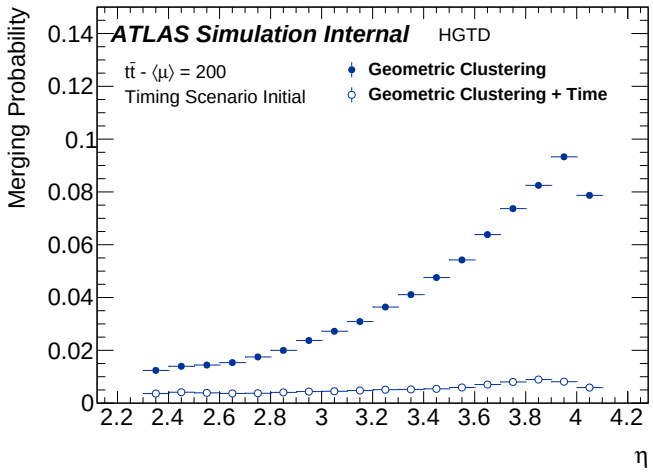


Figure 3.10: Probability of merging contributions originated by multiple particles in the same cluster with the geometric clustering algorithm with and without time filtering in $t\bar{t}$ events with a pileup of $\langle\mu\rangle = 200$.

In a second step, properties of the ITk part of the track and the associated hits in HGTD are used to remove incorrect assignments to improve the purity of correctly assigned times.

The algorithm that extrapolates the track and associates hits proceeds as follows. First, tracks reconstructed in the ITk are extrapolated to the HGTD by using the last measurement of the track in the ITk as the starting point. The tracks are extended to the HGTD surfaces using a progressive Kalman filter. In each sensor layer (two per HGTD layer since they are double sided), HGTD clusters found around the extrapolated crossing location are evaluated for compatibility with the track by attempting to add them to the track in a forward filtering step. Each hit in the active layer that is spatially compatible with the extrapolated position is considered, and the one with the lowest χ^2 is accepted as an extension of the track. The extended tracks must satisfy a requirement of χ^2/n.d.f of less than 5.0. In case of a successful extension, the track parameters are updated and extrapolated to the next sensor layer in the HGTD. This is done progressively for the four sensor layers of the HGTD. At each step, the track information from the last step is used as the starting point of the extension.

To compare reconstructed track times with truth track times (defined as the true times of the production vertices corresponding to the tracks under consideration), the individual hit times need to be corrected. This is achieved by a TOF correction done for each hit in the HGTD. The path length of the particle's track is assumed as a straight line between the origin of the track and the position of a given hit. The origin of the track is defined as (0,0,z_0), where z_0 is the longitudinal impact parameter of the track (defined as the z coordinate of the point on the track closest to the primary vertex in the transverse plane). The TOF is then

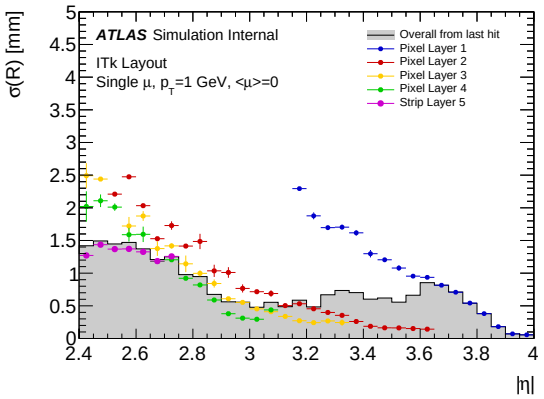
calculated by dividing the path length by the speed of light[3] and subtracted from the hit time. The track-time is then calculated as the arithmetic mean of the times of the individual associated HGTD hits used in the track extrapolation.

The precision of the extrapolation is affected by the material in the ITk and between the ITk and the HGTD. In Figure 3.11 the precision of the extrapolation (resolution per extrapolated HGTD hit) as a function of η to the HGTD surface is shown. Single muons with transverse momenta of 1 GeV and 10 GeV were analysed. The extrapolation is performed from the last hit in the ITk associated to the track. For the majority of tracks $p_T > 1$ GeV, the precision of the extrapolation is better than the pad size used in the HGTD (1.3 mm $\times$ 1.3 mm).

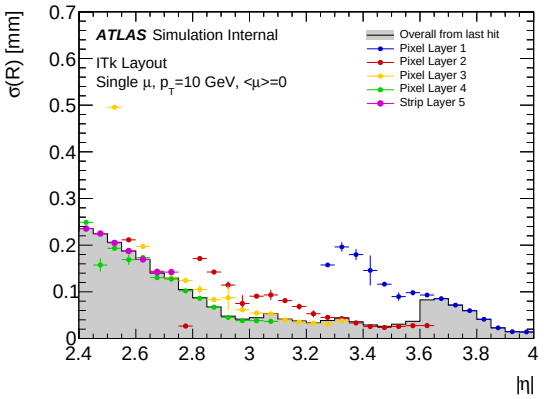
The performance of track-time determination has been evaluated using single-muon and single-pion samples at $\langle\mu\rangle = 0$, generated with a flat distribution in η and ϕ, and the physics sample simulated for VBF $H \to Z(\nu\nu)Z(\nu\nu)$ at $\langle\mu\rangle = 200$. The choice of a VBF sample was motivated by the fact that this final state contains forward jets within the HGTD acceptance. Furthermore, VBF is a broad class of topology particularly suitable for HGTD improvements. The single-particle samples are used to compare the performance for the ideal case of high-p_{T} muons that undergo less material interactions, and the more challenging low-p_{T} pions with hadronic interactions. The high-pileup VBF sample is used as an example to show the more realistic performance for physics studies. All reconstructed tracks with $p_{\mathrm{T}} > 1$ GeV within the HGTD acceptance are used, exactly corresponding to the set of tracks defined for the denominator in the later efficiency calculations. The detector timing resolution corresponding to the start of HL-LHC running is considered.

The overall efficiency of associating a timing measurement from the HGTD to a track is shown as function of pseudorapidity for the two single-particle samples in Figure 3.12. Given the fact that the timing measurement is available when at least one HGTD hit is associated to the ITk track, the overall efficiency is therefore identical to the efficiency of the track extrapolation. It shows the overall rate for determining a track-time (black line), with a bin-by-bin breakdown categorising the origins of the HGTD hits providing track-time measurements. The categories shaded in green to cyan represent "correct" assignments where a faction of HGTD hits genuinely originate from the same primary particle (true particle from hard scattering) as for the ITK track, with the fractions noted in the legends. The case where the primary particle giving rise to the track does not produce any HGTD hits, but timing measurements from hits caused by other particles are labelled "misassignment" and shown in magenta. Primary particles which do produce at least one hit in HGTD but get unrelated hits associated to its track constitute the category labelled "confusion" which is shown in red (but hardly visible in this figure). The impact of the upstream material on the efficiency of assigning a time is apparent for the pions, for which a small fraction of tracks are observed to get times assigned stemming from hits produced by secondary particles.

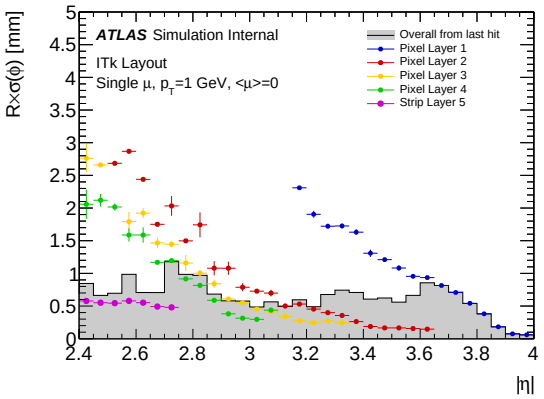
[3] Further iterations of this algorithm will take into account the actual path length of the track and the measured momentum.



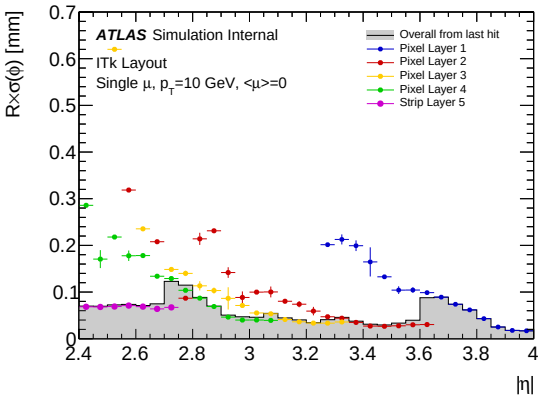
(a) Extrapolation error in R for $p_{\mathrm{T}} = 1\,\mathrm{GeV}$



(b) Extrapolation error in R for $p_{\mathrm{T}} = 10\,\mathrm{GeV}$



(c) Extrapolation error in $R \times \phi$ for $p_{\mathrm{T}} = 1\,\mathrm{GeV}$

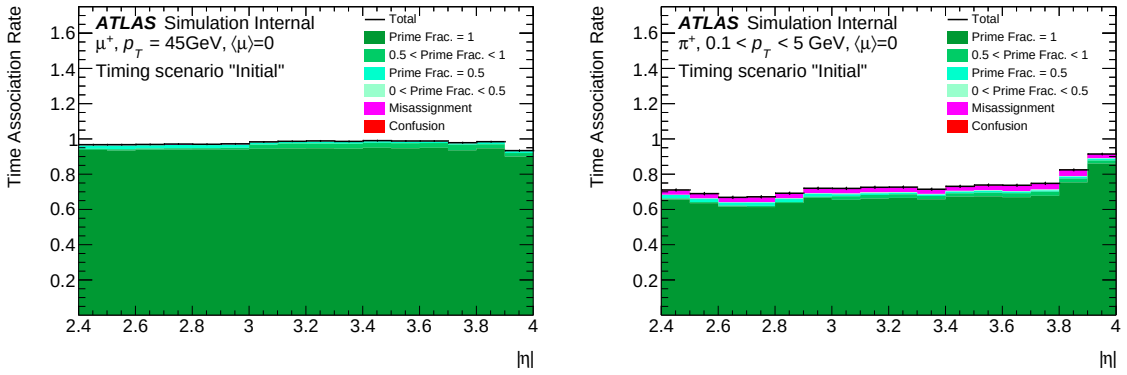


(d) Extrapolation error in $R \times \phi$ for $p_{\mathrm{T}} = 10\,\mathrm{GeV}$

Figure 3.11: The extrapolation resolution in radius r and in the product $r \times \phi$ for tracks with $p_{\mathrm{T}} = 1\,\mathrm{GeV}$ and $p_{\mathrm{T}} = 10\,\mathrm{GeV}$. The resolution is plotted as a function of η for the extrapolation of the track from the last hit in the ITk. The actual layers (segmentation in radius) in the ITK where the last hits are located at are indicated by different colors. The resolution is better than the size of a single pad in the HGTD.

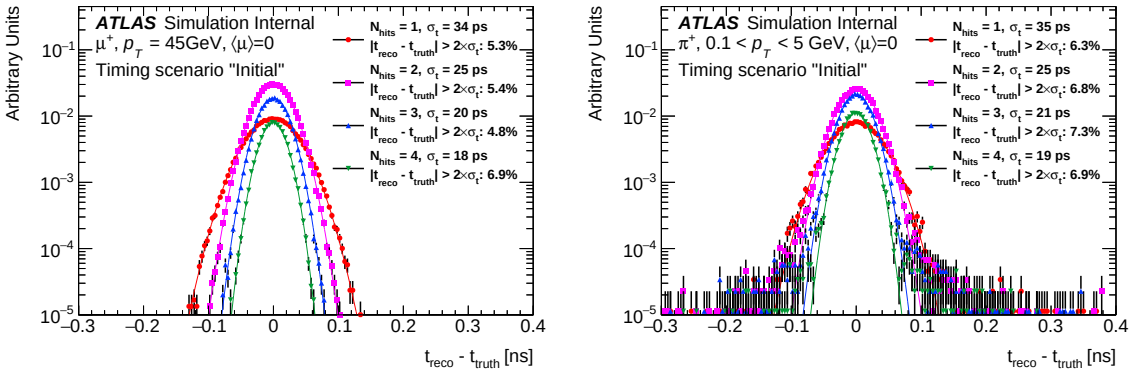
The resulting track-time resolution (shown in Figure 3.13), i.e. the difference between the measured and true track-times ($t_{\mathrm{reco}} - t_{\mathrm{truth}}$), is calculated for tracks extrapolated with one, two, three and four associated HGTD hits separately. Fits to the Gaussian core of each distribution yield σ values which are consistent with the expectations, i.e. $\sigma_{\mathrm{hit}}/\sqrt{n_{\mathrm{hits}}}$. The pions that undergo hadronic interactions give rise to tails that are not visible for the muon events. The slight asymmetry in the tails of these distributions is caused by low-p_{T} particles which travelled a longer path than the assumed straight line between the track origin and the hit position.

Figure 3.14 shows distributions for the same track-time residual variable, but split up into the categories indicating the number of correct and incorrect hits assigned to the tracks,



(a) Single-muon events (b) Single-pion events

Figure 3.12: Overall time association rate for tracks as function of pseudorapidity for (a) single-muon and (b) single-pion events without pileup. A bin-by-bin breakdown of correct (green shades) and incorrect (red/magenta) hit associations is also shown.



(a) Single-muon events (b) Single-pion events

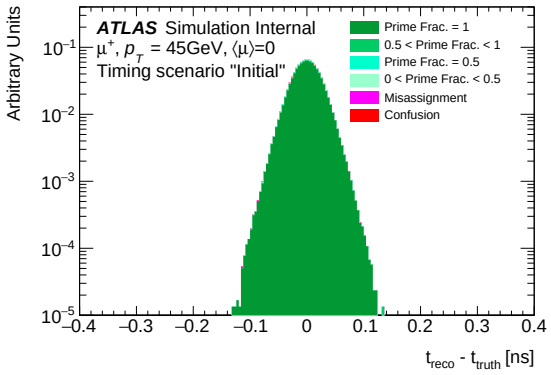
Figure 3.13: Difference between the measured and true track-time for extrapolated tracks with HGTD extensions single-particle events of (a) single-muon and (b) single-pion events without pileup. Distributions corresponding to tracks with different numbers of associated HGTD hits are shown separately. For each distribution, the fitted Gaussian σ and the fraction of tracks outside 2σ are given in the legend.

showing how each category contributes to the tails. The non-Gaussian tail for tracks in which the assigned HGTD hits all originated from primary particles is due to the resolution effect of track z_0 used for TOF correction, which is more profound in the cases of pion and VBF events (consistent with expectation). The fraction of tracks with misassigned hits is close to negligible for both single-particle samples, but Figure 3.14(c) shows that the VBF $H(\text{inv})$ sample suffers from dramatically increased tails due to misassignment, in addition to the Gaussian core with about a resolution of 30 ps . This misassignment occurs when a track undergoes a material interaction before reaching the HGTD such that no hits from the primary track are found in the extrapolated path of the track-hit association algorithm. When this happens in a high-pileup environment, the track-hit assignment algorithm frequently assigns a nearby hit from an unrelated particle, e.g. secondary or pileup particles, resulting in an incorrect track-time.

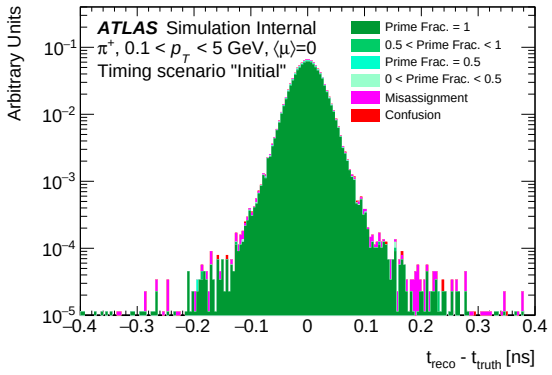
In order to address the challenge of misassigned hits, a hit-cleaning procedure is applied as a second step to improve the purity of the determined track-times. First, one requires the last ITk hit on the track to be on a detector layer closest to the HGTD in the longitudinal direction. This requirement suppresses (wrong) time assignment to tracks that underwent hadronic interactions earlier in the tracker volume. Second, at least two associated hits in the HGTD are required for tracks within $3.5< |\eta| < 3.9$. This is a region with substantial material upstream of the HGTD, but that also features larger overlaps between the HGTD modules, allowing for a more stringent number of hits requirement. These two quality requirements aim at identifying tracks undergoing material interactions, for which the time assignment is likely incorrect. This of course reduces the fraction of tracks that will be assigned a time. The final cleaning step ("outlier removal") applies only to tracks that have at least two hits assigned. It consists of checking the consistency in time among all the associated hits and then removing the track under consideration if the times assigned to the HGTD hits have significant inconsistency.

The impact of the cleaning procedure on the track-time association rate is shown in Figure 3.15 as a function of track η for the VBF $H(\text{inv})$ sample. The cleaning procedure is effective at reducing the number of tracks with incorrect times, at the expense of a slight reduction of the overall efficiency to correctly attach a time to a track. In the future, more sophisticated versions of track-hit reconstruction and cleaning can be developed and are expected to further improve efficiencies and reduce misassignments.

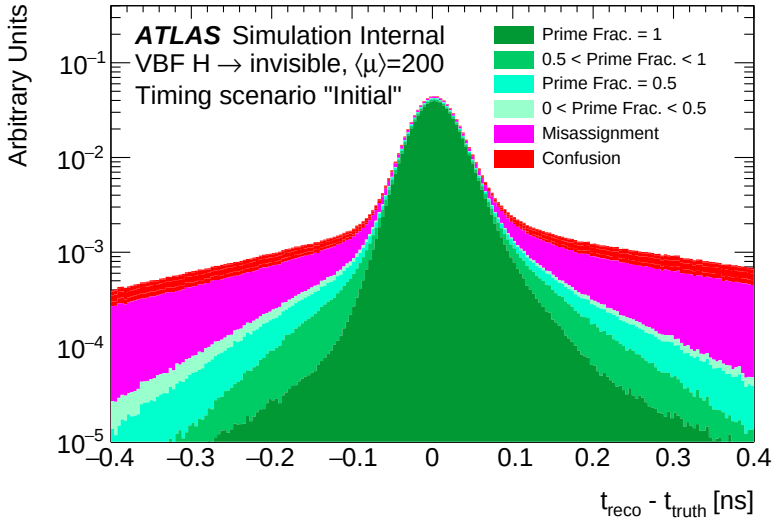
Based on the above studies (especially the timing resolution presented in Figure 3.14), the track-time is considered to be assigned correctly, if the number of associated HGTD hits corresponding to the same primary particle is more than 50% of the total; otherwise, it is considered as a misassignment. Figure 3.16 shows the rates of correct assignment and misassignment of track-times, after the cleaning procedure, as a function of track η and p_{T}. The overall efficiency to correctly assign a time to a reconstructed track is around 50% at 1 GeV and plateaus at 60% for $p_{\text{T}} > 4$ GeV, with a misassignment rate of approximately 10%. The impact of the cleaning procedure on the track-time resolution, separately for cases



(a) Single-muon events

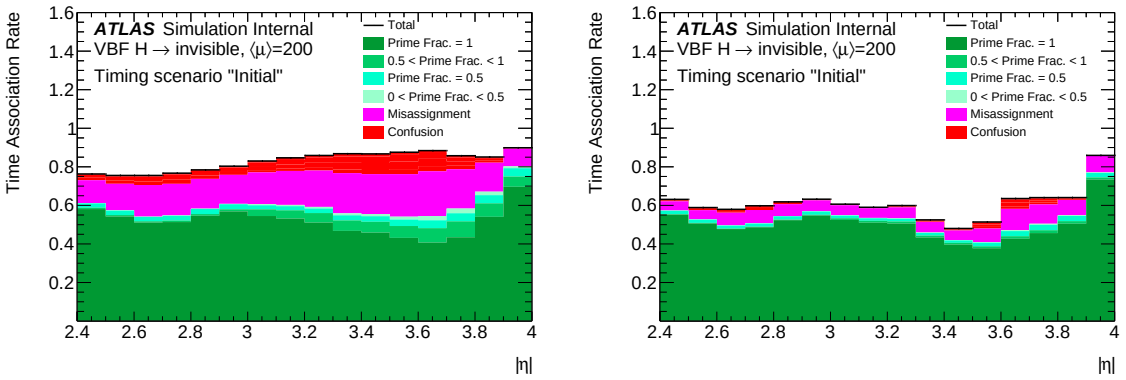


(b) Single-pion events



(c) VBF $H(\text{inv})$ events with $\langle\mu\rangle = 200$

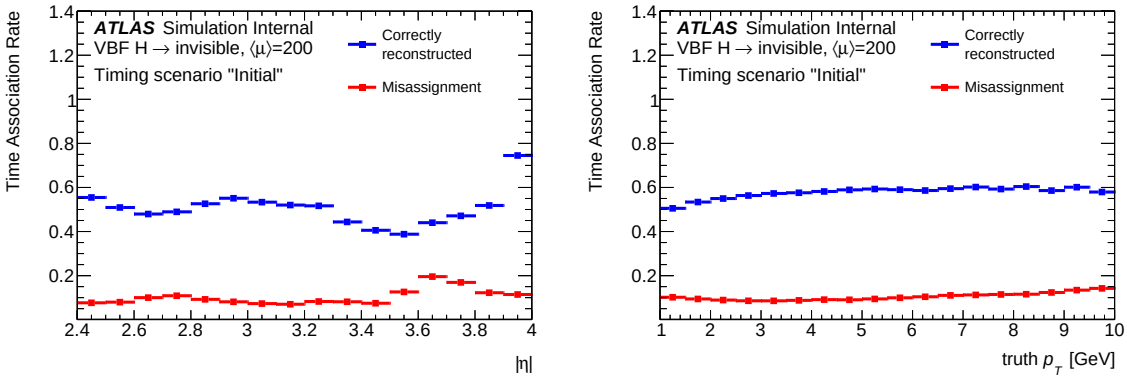
Figure 3.14: Difference between the measured and true track-time for extrapolated tracks with HGTD extensions single-particle events of (a) single-muon and (b) single-pion events without pileup, and (c) VBF $H(\text{inv})$ events with $\langle\mu\rangle = 200$. A breakdown of how correct (green shades) and incorrect (red/magenta) hit associations contribute in each bin is shown. These distribution are before hit cleaning.



(a) (b)

Figure 3.15: Rate of correctly assigned and misassigned times as a function of track η before (left) and after (right) the cleaning procedure.

with different fractions of primary hits, is shown in Figure 3.17. Comparing this figure with Figure 3.14(c) one can see how the fraction of non-Gaussian tails has been reduced significantly. The track and time hit requirements of the cleaning procedure help reduce the red and magenta components of the time resolution distribution due to tracks with no primary hits, whereas the outlier removal step reduces the contribution of the green components by removing out-of-time hits that degrade the time resolution.



(a) (b)

Figure 3.16: Rate for correct and incorrect assignment of track-times as a function of track η (left) and p_{T} (right). The sum of two rates gives the inclusive efficiency of track-time assignment.

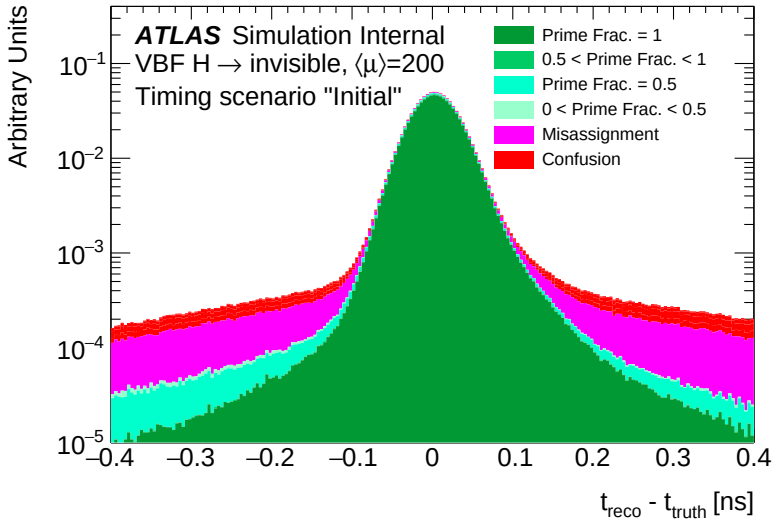


Figure 3.17: Difference between the measured and true track-time for extrapolated tracks with at least one HGTD hit for VBF $H \to Z(\nu\nu)Z(\nu\nu)$ events after the cleaning procedure, separately for cases with different fractions of primary hits that are left by true particles not originating from material interactions

3.2.2 Determination of the time of the primary vertex

Due to the large uncertainty of the longitudinal impact parameter for tracks in the forward region (Figure 2.6), the association of tracks to nearby vertices purely based on spatial information is ambiguous in high-pileup environments, especially for low transverse momentum tracks. The ability to determine the time of the primary vertex of the hard-scatter process, here denoted as t_0, provides a new handle to enhance the capability of the ATLAS detector to remove pileup tracks contaminating physics objects originating from the hard-scatter vertex.

The experimental determination of the vertex t_0, however, is challenging. There are two key factors that affect the accurate determination of the hard-scatter vertex time. First, due to the limited pseudorapidity acceptance of the HGTD, the hard-scatter interaction needs to have enough high-p_{T} tracks with $|\eta| > 2.4$. Second, the limited efficiency for correct track-time association efficiency for hadrons further reduces the number of tracks available to determine t_0. It will be shown that these two effects limit the availability of a global vertex time to approximately 65% of the events in a VBF $H(\mathrm{inv})$ sample.

This section focuses on the implementation and performance of a relatively simple vertex t_0 technique. The algorithm proceeds as follows. First, an iterative time-clustering algorithm is used to find clusters of tracks that are within a window in z around the selected hard-scatter vertex in the event and have consistent times. The window is defined using a 2-dimensional

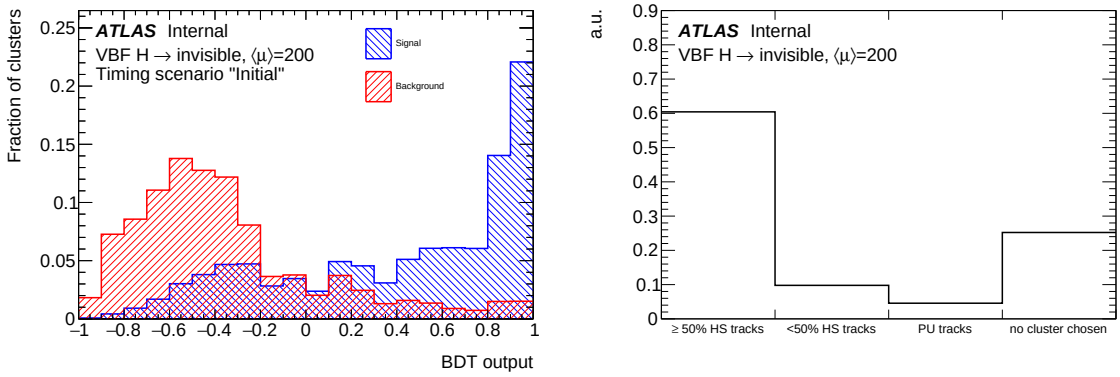
parameterisation of the track z_0 resolution as a function of track p_{T} and η, as shown in Figure 2.6 for only two p_{T} bins. To ensure tracks inside a cluster are consistent in time, the track-time of a given track must agree with that of any other track within a window of $3 \times \sigma_t$, where σ_t is the square root sum of track-time errors of the two tracks under consideration. Next, a Boosted Decision Tree (BDT) algorithm was trained to identify the most likely hard-scatter cluster among the various clusters, taking eight variables as input. The eight variables are: the weighted averages (taking into account the corresponding track parameter errors) of both the transverse impact parameter and $1/p_{\mathrm{T}}$ as well as the uncertainties on these two averages, the uncertainty on the weighted average of the longitudinal impact parameter, the distance and significance in z between the cluster's averaged z_0 and the position of the primary vertex, and the total sum of p_T^2 of the tracks. Signal (hard-scatter-like) clusters were defined as those clusters containing more than or equal to 50% of hard-scatter tracks in them, as determined using truth information in VBF $H(\mathrm{inv})$ events. Background (pileup-like) clusters were those with less than 50% of hard-scatter tracks.

Figure 3.18 shows the BTD output for signal and background clusters, as well as a detailed description of how often it selects the correct in-time cluster. The best cluster is determined as the one having at least three tracks, and the maximum BDT output that passes a cut of 0.2. The cut was chosen to keep the background efficiency below $\sim$ 10%. In about 60% of the vertices, the BDT selects the correct cluster. Whereas approximately 25% of the cases, no cluster is selected (either because the BDT output is less than 0.2 or because the cluster has less than three tracks) and the algorithm does not provide a t_0 for the event. The remaining 25% of the cases correspond to mixed clusters that have various fractions of pileup and hard-scatter tracks. In particular, 5% of the time, this algorithm will calculate the time purely based on pileup tracks and therefore result in an incorrect time for the hard scattering vertex.

After a time-compatible track cluster is chosen, the vertex t_0 is defined as the weighted average time of all the tracks belonging to the cluster. Figure 3.19 shows the distribution of reconstructed t_0 minus truth t_0 for all vertices for which a t_0 was found, separated into various categories based on the fraction of hard-scatter tracks. The vertices with at least 50% HS tracks has an RMS spread of 22 ps, while the ones with some but smaller fraction of HS tracks and only pileup tracks have spreads of 70 ps and approximately 200 ps, respectively.

3.3 Physics object performance

The new capability introduced by the HGTD to provide a vertex t_0 as well as time information for forward tracks can be exploited to mitigate the impact of pileup in high-level physics object reconstruction. In this section the "initial" timing scenario was used. After a detailed description about the ways in which timing information can be leveraged to improve the association of tracks to vertices, this section focuses on how the HGTD can improve the



(a) (b)

Figure 3.18: (a) BDT output distribution for in-time clusters containing more than or equal to (signal) or less than (background) 50% of hard-scatter tracks; (b) Fraction of events as a function of the fraction of hard-scatter tracks in each cluster. The first bin correspond to the cases in which the BDT selects the correct cluster as hard-scatter. The last bin are cases in which the BDT does not select any cluster. The intermediate bins show the various ways in which the BDT picks an incorrect cluster, where "HS" ("PU") stands for hard-scatter (pileup).

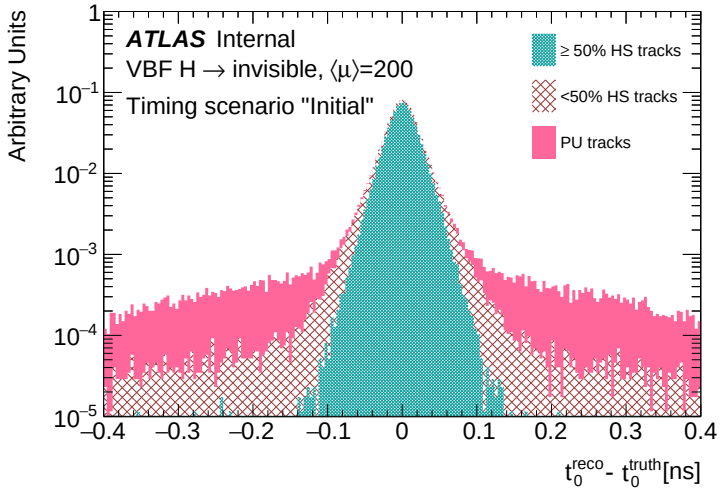


Figure 3.19: Vertex t_0 resolution separately for various cases, where "HS" ("PU") stands for hard-scatter (pileup).

suppression of forward pileup jets and the efficiency of forward lepton isolation based on full simulation studies. At the end of the section, a brief description of additional applications of the HGTD left for future work is also included.

3.3.1 Challenges for associating tracks to vertices

The precise assignment of tracks to primary vertices (track-to-vertex association) is one of the key elements to mitigate the effects of pileup on the full suite of event reconstruction algorithms at hadron colliders. Jet reconstruction and calibration, pileup mitigation for jets, b-tagging, lepton isolation, and jet substructure measurements rely strongly on the correct assignment of tracks to primary vertices and jets.

A track is associated to a vertex if its origin is geometrically compatible in z with the vertex position. The compatibility can be determined by the resolution on the track z_0 impact parameter such that

$$\frac{\left|z_0 - z_{\text{vertex}}\right|}{\sigma_{z_0}} < s, \tag{3.1}$$

where σ_{z_0} is the per-track resolution on the longitudinal impact parameter which depends primarily on the track η and p_{T}, and s is a significance cut. Typical values for s are 2.5 or 3.

While the longitudinal impact parameter resolution is relatively constant and small ($\leq 30\mu m$) for $|\eta| < 1.5$, it grows rapidly with pseudorapidity, reaching several millimetres for $|\eta| \gtrsim 3.2$ for low-p_{T} tracks. The η dependence of the impact parameter resolution is mostly determined by the geometry of the inner detector. As η increases, tracks become more collinear to the beam line.

Based on Figure 2.6, a 1 GeV track with $|\eta| = 3$ has a z_0 resolution of approximately 1 mm. With a most probable average vertex density (at $\langle\mu\rangle = 200$) of 1.8 vertices/mm at $z = 0$, this means that, on average, a low p_{T} forward track can be compatible with up to about 9 near-by vertices on average. This means that the association of low p_{T} tracks to vertices becomes ambiguous at large pseudorapidity and high luminosity, leading to a high level of pileup track contamination. Or, in other words, track-to-vertex association will suffer significantly from pileup contamination, reducing the efficiency of track-based pileup suppression methods.

Another way to understand this challenge is by comparing the z_0 resolution of a few millimetres for forward low p_{T} tracks with the average separation between vertices, given by the inverse of the average vertex density $1/\langle\rho(z)\rangle$ ~0.6 mm. This means that the tracker longitudinal impact parameter resolution in the forward region is significantly larger than the typical separation between vertices. This is an intrinsic challenge of forward trackers in hadron colliders.

Timing information constitutes a powerful new way to address this challenge. By requiring all tracks within a z window around the primary vertex [11] to have a common time compatible with the time of the hard-scatter vertex, the additional pileup tracks from nearby interactions can be significantly reduced. The hard-scatter vertex is reconstructed from all possible tracks in the full inner detector, while its time is derived from the associated tracks within the HGTD acceptance. In this way, the use of timing information provided by the HGTD can therefore improve the performance of physics object reconstruction, which are detailed in following sections.

3.3.2 Strategy of improving physics object performance

This section describes how the use of timing information can improve the reconstruction of physics objects, such as jets and leptons, by reducing the impact of forward tracks from pileup interactions that cannot be unambiguously associated to the hard-scatter vertex of the event.

There are two main approaches. In one approach, the hard-scatter vertex time t_0 is determined so that it can be used as a global reference to check the time compatibility of tracks associated to jets or other physics objects in the event. This is the most powerful, and intuitive, way to utilise timing information, and it is a natural extension of the track-to-vertex association in 4 dimensions (space-time). Once a vertex t_0 is found, tracks are required to satisfy

$$\frac{t_{trk} - t_0}{\sigma_t} < s \tag{3.2}$$

where σ_t is the sum in quadrature of the vertex t_0 and the track-time (t_{trk}) errors, and s is a significance cut, such as 2, or 3.

As discussed in Section 3.2.2, however, the experimental challenges associated to the determination of the vertex t_0 limit the full power of this approach.

A second approach, denoted *self-tagging*, does not require the knowledge of the hard-scatter time. The key idea is to check the consistency of the measured production time for all tracks associated to the same physics object (such as a jet) among themselves. For example, if a jet consists of four tracks but one of them has a significantly different time, then this fourth track which is incompatible in time can be filtered out. More generally, the self-tagging method consists of finding clusters of tracks within a jet that have compatible times, and splitting the jet into smaller sub-jets with consistent times. Specific algorithms can then use the sub-jets in different ways, as will be shown in the next subsections with particular examples.

The self-tagging approach is limited by several elements. First, it requires physics objects to have at least two tracks with time assigned. In the case of pileup jets, the majority of them have only one track in the acceptance of HGTD, reducing the power of this method

compared to the global t_0 approach. In other applications, like b-tagging or particle-flow jet reconstruction, where more tracks are available, this approach can be important. Second, the self-tagging approach can only address the case of *stochastic* pileup contamination, as opposed to hard-QCD pileup interactions. It assumes that a jet consists of a group of tracks with a common time origin plus additional tracks out-of-time from nearby, uncorrelated, pileup interactions. In a hard-QCD pileup jet, on the other hand, all its tracks will have a common time, making this method not applicable. As the fraction of hard-QCD pileup jets increases with jet p_{T}, the self-tagging method will work best at low jet p_{T}.

Both the global t_0 and the self-tagging approaches are complementary to each other, and can be combined for maximum performance across jet p_{T}. The following subsections show how these two techniques can be used to improve the rejection of pileup jets and lepton isolation. Other applications, like b-tagging, missing transverse energy, and particle flow are outside the scope of this TDR due to their complexity, but are also expected to benefit from the use of track-time information to mitigate the impact of pileup.

3.3.3 Suppression of pileup jets

Pileup jets can reduce the precision of Standard Model measurements and the sensitivity to discover new physics. For example, additional jets can increase the number of background events passing a selection, as well as reduce the efficacy of kinematic variables or discriminants to separate signals from backgrounds. Hence, the efficient identification and rejection of pileup jets are essential to enhance the physics potential of the HL-LHC. These pileup jets can be produced as the result of a hard QCD process (QCD jets) from a pileup vertex, or by random combinations of particles from multiple vertices. At low jet p_{T}, the latter mechanism is dominant, whereas at high jet p_{T}, the majority of pileup jets are QCD jets.

The key element to suppress pileup in jets is the accurate association of jets with tracks and primary vertices. A simple but powerful discriminant is the $R_{p_{\mathrm{T}}}$ jet variable, defined as the scalar sum of the p_{T} of all tracks that are inside the jet cone and originate from the hard-scatter vertex PV_0, divided by the fully calibrated jet p_{T}, i.e.

$$R_{p_{\mathrm{T}}} = \frac{\Sigma p_{\mathrm{T}}^{\mathrm{trk}}(\mathrm{PV}_0)}{p_{\mathrm{T}}^{\mathrm{jet}}}.$$

The tracks used to calculate $R_{p_{\mathrm{T}}}$ fulfill the quality requirements defined in Ref. [11] and are required to have $p_{\mathrm{T}} > 1\,\mathrm{GeV}$. The matching criteria are defined in Ref. [12]. In this study, jets are reconstructed from clusters of calorimeter energy deposits using the anti-k_t algorithm [13, 14] with radius parameter $R = 0.4$. Reconstructed hard-scatter jets are required to be within $\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2} < 0.3$ of a truth jet with $p_{\mathrm{T}} > 10\,\mathrm{GeV}$. Hard-scatter and pileup jets for simulated events are defined by their matching to truth jets, which are reconstructed

from stable and interacting final state particles coming from the hard interaction. The pileup jets must be at least $\Delta R > 0.6$ away from any truth hard scattering jet with $p_{\mathrm{T}} > 4\,\mathrm{GeV}$.

At moderate levels of pileup, where track impact parameter measurements can be used to assign tracks to vertices with relatively little ambiguity, small values of $R_{p_{\mathrm{T}}}$ correspond to jets which have a small fraction of charged-particle p_{T} originating from the hard-scatter vertex PV_0. These jets are therefore likely to be pileup jets. However, at high-pileup conditions, and particularly in the forward region, the power of this discriminant is reduced, because the longitudinal impact parameter resolution becomes worse and the pileup tracks could have more chance to be incorrectly included in the numerator of $R_{p_{\mathrm{T}}}$.

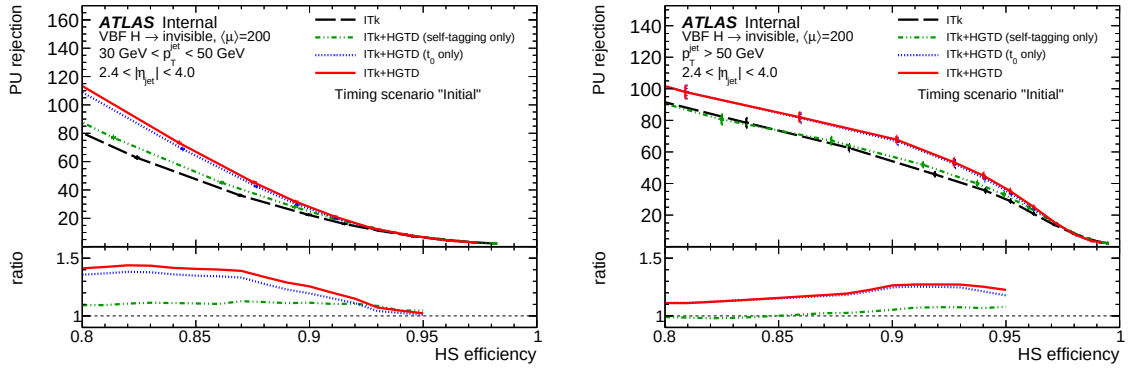
As described in Section 3.3.2, there are two approaches to incorporate the time information of tracks inside jets: self-tagging, and global vertex t_0. In the self-tagging approach, jets with at least two tracks with time information are split into sub-jets of in-time track clusters. The $R_{p_{\mathrm{T}}}$ variable is then recomputed for each in-time cluster plus the additional tracks that have no time assigned. Then, the jet $R_{p_{\mathrm{T}}}$ is defined as the maximum $R_{p_{\mathrm{T}}}$ of all sub-jets. Since each cluster contains a subset of tracks, by construction, the self-tagging $R_{p_{\mathrm{T}}}$ will have a smaller value (as expected) under the presence of pileup, improving the discrimination power of the method. When a global vertex t_0 is available, $R_{p_{\mathrm{T}}}$ can be recomputed after removing tracks outside a $2\sigma_t$ window around the reconstructed time of the hard-scatter vertex. It is also possible to combine both approaches such that when no t_0 is found, the self-tagging method is used.

Figure 3.20 shows the rejection, i.e., the inverse of the mis-tag efficiency, of pileup jets as a function of the efficiency for selecting hard-scatter jets using the $R_{p_{\mathrm{T}}}$ discriminant for jets with low and high p_{T} in VBF Higgs invisible events with $\langle\mu\rangle = 200$ without and with the HGTD using the three approaches described above: self-tagging, t_0, and combined. The combined method improves the rejection of pileup jets with $30 < p_{\mathrm{T}} < 50$ GeV in the forward region up to a factor of approximately 1.5 at a signal efficiency of 85%. The presented performance for the ITK-only case is largely consistent with that presented in Ref. [15]. which could be due to the difference in the physics processes and in the material budget being simulated.

Figure 3.21 shows the relative pileup-jet rate for relatively low-p_{T} jets, as a function of pseudorapidity using the combined timing reconstruction algorithms described in this section. A significant improvement is observed at larger values of η where the z impact parameter resolution is worse and timing information becomes more important to associate tracks to vertices.

3.3.4 Lepton track isolation

The ability to assign a time to leptons can be exploited to reduce the impact of pileup in the case of applying track-isolation criteria to leptons in the forward region. The efficiency of



(a) Jets with $30\,\text{GeV} < p_\text{T} < 50\,\text{GeV}$.

(b) Jets with $p_\text{T} > 50\,\text{GeV}$.

Figure 3.20: Pileup jet rejection as a function of hard-scatter jet efficiency in the $2.4 < |\eta| < 4.0$ region, VBF H to invisible sample, for the ITk-only and combined ITk + HGTD reconstruction.

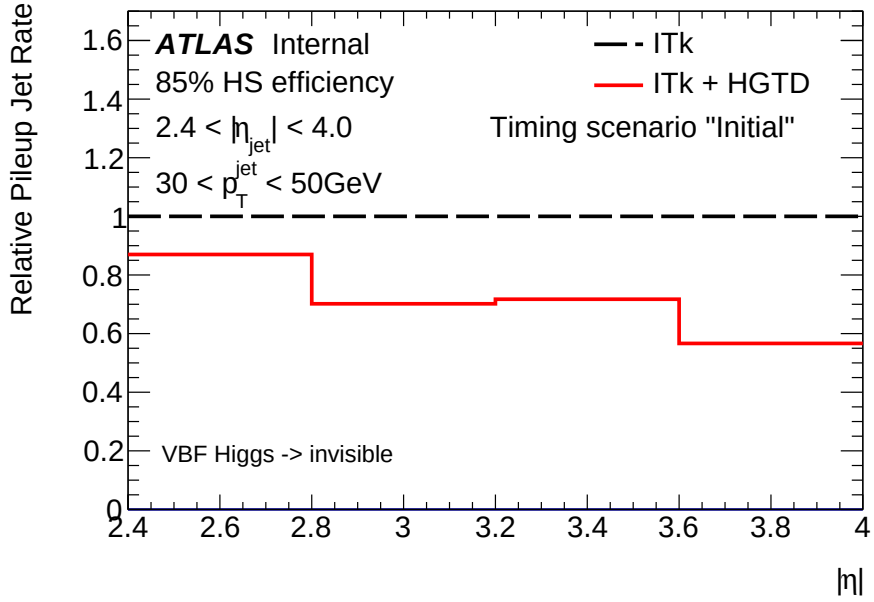


Figure 3.21: Relative pileup jet rate as a function of jet pseudorapidity, for jets with $30\,\text{GeV} < p_\text{T} < 50\,\text{GeV}$.

the "track-based" lepton isolation is defined as the probability that no additional tracks with $p_{\mathrm{T}} > 1\,\mathrm{GeV}$ are reconstructed within $\Delta R < 0.2$ of the lepton track. In the forward region, the relatively large z window required to associate tracks to the primary vertex results in increased pileup track contamination, which consequently degrades the isolation efficiency. The association of a time to the lepton track can be utilised to reject tracks within the isolation cone which come from pileup interactions spatially close to the hard-scatter vertex.

As an example, this is studied using electrons from Z boson events. Similar results are expected for tau leptons decays in acceptance. Forward electrons with $p_{\mathrm{T}} > 20\,\mathrm{GeV}$ passing the standard ATLAS "medium" identification criteria are selected [.] The electron track is defined as the track closest to the calorimeter cluster of the electron, out of those that have a ratio of track p_{T} to transverse cluster energy greater than 0.1.

In order to improve the lepton isolation definition, the time of all tracks with $p_{\mathrm{T}} > 1\,\mathrm{GeV}$ which are within the $\Delta R < 0.2$ isolation cone are compared with the time of the electron track. If the time difference between the two is larger than twice the quadratic sum of the timing resolution of both tracks, the track is discarded. This procedure allows the recovery of the cases in which a nearby pileup track will cause the lepton isolation to fail.

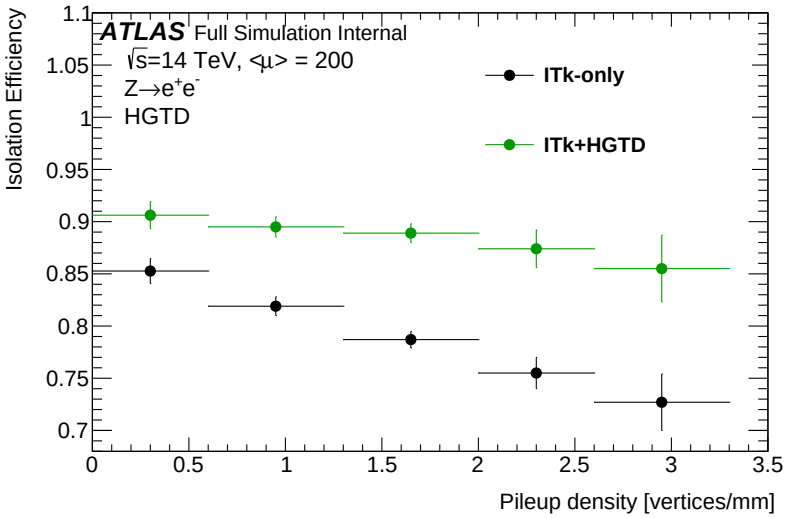


Figure 3.22: The efficiency for electrons to pass track-isolation criteria, denoted as $\epsilon(p_{\mathrm{T}}^{\mathrm{iso}})$, as function of the local vertex density, for the ITk-only and ITk+HGTD scenarios.

The isolation efficiency as a function of the pileup density is shown in Figure 3.22 for the ITk-only and HGTD scenarios, where all electrons no matter whether a timing measurement is available or not are included in the denominator.

While the efficiency drops with increased pileup vertex density when using only the ITk, the addition of the HGTD timing information reduces this drop, keeping an efficiency above

85% even at high pileup density, i.e. with up to three additional vertices per mm around the hard-scatter vertex on average. For a local pileup density of the order of 1.6 vertices/mm the electron isolation efficiency is improved by about 10%, which corresponds to a factor of two reduction of the inefficiency. These results show that the expected HGTD performance is sufficient to achieve a forward lepton track isolation efficiency essentially independent of the pileup vertex density and at a level similar to that achieved in the central region. The study of the impact on the isolation efficiency for background electrons produced inside jets or from misidentification is beyond the scope of the study in this document. The result from this study is later applied in the study of the sensitivity improvement in the weak mixing angle measurement.

3.3.5 Additional applications

While improvements in the performance of jets and forward leptons have been demonstrated in the previous sections, the incorporation of timing information into the full suite of ATLAS physics object event reconstruction is expected to bring additional improvements in other areas not yet considered in this document. In particular, the HGTD is expected to enhance the performance of particle-flow jet energy reconstruction, transverse missing energy, and forward b-jet tagging. Discussions about potential improvements in the understanding of the pileup activity and the dimension of beam spots are given at the end.

Particle-flow jet reconstruction

Particle-flow jet reconstruction relies on the ability to match charged particle tracks with calorimeter signals and primary vertices. In particular, a key component of this approach, leading to the improved jet energy resolution, is the removal of calorimeter energy deposited by tracks originating from pileup vertices. The ambiguity to accurately associate forward tracks to nearby vertices is expected to limit the capability of particle-flow algorithms to reduce the impact of energy fluctuations due to pileup within jets in the forward region. This effect can be addressed by the use of timing information to correctly identify hard-scatter and pileup tracks within the jet. For jets with low track-multiplicities, on the other hand, one may take advantage of the vertex t_0 identification. Improved jet energy resolution can lead to further improvements of sensitivity in many key physics channels, such as Vector Boson Fusion analyses. The full integration of HGTD into the particle-flow reconstruction chain, however, is a long-term goal, involving many steps at the level of calorimeter reconstruction and calibration that are still under development.

Missing E_{T}

There are three ways in which the HGTD can be utilised to potentially enhance the resolution of the missing E_{T}. The first means is by leveraging the improved jet energy resolution from particle-flow reconstruction using HGTD. Second, by reducing the (forward) pileup contamination in the track-based *soft-term* of the missing ET. The soft-term is defined by all

charged tracks associated to the primary vertex that do not belong to hard physics objects. The knowledge of the track-time will enable a more pure selection of forward tracks in the soft-term component of the missing E_{T}. Third, the improved forward pileup jet suppression will directly translate into improvements in the missing E_{T} resolution, by the rejection of pileup jets that appear to come from the primary vertex but do not belong to the hard-interaction. A full propagation of the pileup jet suppression and the incorporation of timing information to the soft-term reconstruction will be pursued in the future.

b-tagging

The HGTD can be particularly useful to mitigate the impact of pileup track contamination on b-tagging. The presence of pileup tracks with relatively large z impact parameter with respect to the hard-scatter vertex can create fake secondary vertices leading to a reduction in light-quark jet rejection. A combination of self-tagging (for high multiplicity track jets) and vertex t_0 could potentially enhance the rejection of pileup tracks to compensate for the lost b-tagging performance at high vertex densities. The full incorporation of timing information within the software framework for heavy-flavour tagging requires major infrastructure changes and is left for a future study.

Additional pileup applications

Section 3.3.3 discussed how HGTD can address the challenge of pileup by mitigating the impact of pileup jets in the forward region. But the HGTD can also help control pileup activities in different ways. For example, the ability to access the time of charged tracks can serve as a robust way to isolate and estimate pileup contributions in data and constrain systematic uncertainties related to pileup itself. Timing information can also be used to create dedicated (orthogonal) control regions to increase the understanding of pileup effects on track reconstruction in dense environments, possibly leading to reduced systematic uncertainties in track-jet observables for physics. These are some of the most difficult experimental uncertainties limiting jet shape measurements and tagging techniques which will be a major element of the Run 4/5 physics programme. Furthermore, the HGTD adds robustness and redundancy, as well as complementarity to ITk, to ensure the full exploitation of the forward region for physics.

Four-dimensional beam spot

Knowledge about the shape and characteristics of the luminous regions at the interaction points of the experiments is valuable information. With the timing capabilities of the HGTD, the beam spot can be determined in four dimensions, adding a time profile in addition to the distributions of where the interactions happen along the three spatial directions. This provides an extra handle for understanding the beams. Accurate determination of the beam spot is also of high importance for several ATLAS applications, e.g. tracking and flavor-tagging in the online trigger system, the offline reconstruction and calibration processes, and etc. Adding a fourth dimension to the determination of the beam spot can result in improvements for all of these uses.

3.4 Physics

This section describes the impact of the HGTD on a set of selected physics analyses. Each of these were chosen as representative examples of broader classes of analyses in final states of particular interest, such as Vector Boson Fusion (VBF) and Vector Boson Scattering (VBS) processes, and precision measurements with leptons in the forward region.

There are three main ways in which the HGTD will enhance the physics capabilities of ATLAS by exploiting the new dimension of timing information that is orthogonal to any other detector measurement:

- by improving the reconstruction of physics objects such as forward jets and leptons, key in VBF, VBS, and lepton-based forward-backward asymmetry measurements;
- by providing new features on the data from the use of timing information uncorrelated with other detector measurements;
- and by providing a new, powerful capability for precise online and offline luminosity measurements at ATLAS to help achieve the goal of 1% luminosity uncertainty for the Higgs precision physics programme of the HL-LHC.

The broad class of physics analyses benefiting from improved jet and lepton reconstruction are exemplified by a search for VBF-produced Higgs bosons decaying invisibly, and with Standard Model (SM) measurements of a VBS process and the weak mixing angle $\sin^2\theta_{\text{eff}}$. VBF final states constitute a major component of the HL-LHC physics programme, both in terms of precision measurements and new physics searches. In the specific case of Higgs decaying invisibly, the Higgs boson could be the portal to dark matter, or, in the context of Hidden Valley models, a rich dark sector beyond the SM [16, 17]. This particular decay mode, however, is meant to provide an example of how the HGTD can improve the relevant VBF analyses. The primary way in which the HGTD can enhance VBF physics event reconstruction is by reducing the impact of pileup. VBF final states are characterised by two tagged jets with a large rapidity gap such that most of the time at least one jet is within the HGTD acceptance. One of the dominant backgrounds is due to QCD $Z+$ jet production, where the final state often contains a hard-scatter jet plus at least one additional forward pileup jet produced in a different interaction close to the hard-scatter vertex. Utilising the improved pileup jet rejection provided by HGTD, it will be shown that the signal-over-background ratio can increase by 7-15%, depending on the event selection categories considered. Additionally, the impact of forward pileup jets and the potential impact of HGTD on a VBS WZjj analysis is discussed. The measurement of the weak mixing angle $\sin^2\theta_{\text{eff}}$ exemplifies the potential of the HGTD to improve the broader class of precision measurements containing at least one forward lepton. As it will be shown, the improved lepton isolation efficiency, allowed by better pileup track rejection, enables the signal acceptance increasing for this analysis resulting in a 13% increase on the $\sin^2\theta_{\text{eff}}$ sensitivity in the dominant central-forward category.

As introduced beforehand the power of the HGTD as a new luminometer can significantly improve the luminosity measurement uncertainties. A reduced luminosity uncertainty is considered to be one of the keystones to enable precision measurements at the HL-LHC. In fact, it is known that in order to achieve the HL-LHC goals for Higgs coupling precision (percentage level for the main couplings), significant improvements in the precision of the luminosity measurement (with a target of 1%) are required [18]. The HGTD provides several unique capabilities to reach this goal: very high granularity and low occupancy, timing information to enable afterglow background removal, and additional redundancy complementing the primary ATLAS luminosity detector for Run 4 (LUCID).

The results presented here are only meant to provide a few representative examples of how timing information and the HGTD can impact the physics potential at the LHC. There are more opportunities for HGTD to improve final states containing low-p_{T} objects in the forward region which are particularly sensitive to the impact of pileup, as well as completely new possibilities. One example in which the HGTD could provide entirely new, future, opportunities that can expand the scope of the HL-LHC physics programme at ATLAS is the search for magnetic monopoles, discussed at the end of this section (although it requires ALTIROC modification, so it's not part of the baseline).

In addition it has to be pointed out that precision timing information is a completely new feature at hadron collider experiments, different and uncorrelated to any existing measurement. It is expected that the use of more sophisticated machine learning algorithms and physics analyses using timing variables will result in both further improvements and new applications. Moreover, many potential improvements from improved particle-flow jet and missing transverse energy, jet vetos, forward b-tagging, etc. have not yet been considered in time for this TDR.

3.4.1 Vector boson fusion Higgs production

The analysis of Vector Boson Fusion (VBF) Higgs production is a major component of the HL-LHC physics programme. This production mechanism has the highest cross section after gluon-gluon fusion and provides key features in the trigger and offline to separate the signal from backgrounds. The main characteristic of VBF events is the presence of two jets with a large rapidity gap. Since most of the time at least one of those jets is within the HGTD acceptance, this final state can benefit from the improved jet reconstruction and pileup jet suppression provided by the HGTD. There are several ways in which the HGTD can increase the sensitivity to VBF topologies. First, the HGTD can reduce the impact of pileup. Depending on the decay final states of the Higgs boson, the major backgrounds usually originate from the production of one or two bosons in association with two jets, where one of the two could be a forward pileup jet. While tracking-based pileup jet suppression algorithms are powerful at removing pileup jets, this task is more challenging in the forward region and timing information can overcome this limitation. In addition to reducing the

impact of pileup, improved jet and transverse missing momentum reconstruction can lead to improvements in the signal to background ratio (S/B).

The search for invisible decays of the Higgs produced through VBF was chosen as a representative analysis to illustrate the impact of the HGTD on VBF topologies. This analysis is particularly challenging because of the lack of high p_{T} features in the central region [19]. However, it is likely that the conclusions from this specific study are relevant for the broader set of VBF and vector-boson-scattering physics analyses planned for the HL-LHC. The dominant backgrounds in this case are the production of $W/Z+$ jet, where the resulting final states often contain one hard-scatter jet plus an additional forward pileup jet, or two forward pileup jets. The relative improvement on the sensitivity is demonstrated by showing the gain on S/B, which is particularly relevant for an analysis dominated by background systematic uncertainties.

This study was performed using a full simulation based on GEANT4 [7, 8] with the goal of accounting for all pileup effects in a more detailed and complete way than in a fast simulation approach. The signal VBF $H \rightarrow$ invisible and dominant QCD $Z(\rightarrow \nu\bar{\nu})$+jets backgrounds were simulated with POWHEG-BOX [20–24] (v1_r2856) interfaced with PYTHIA8 [25, 26] (v8.186). This study only considers QCD Z+jet background for simplicity, but Z+jet and W+jet have the same jet structure, so it is likely that the conclusions obtained for the pileup jets in Z+jets will be relevant also for W+jets. Other considerations such as lost leptons in W+jets will scale differently than $Z(\rightarrow \nu\bar{\nu})$+jets with the extended tracking coverage of the ITk, but that is not considered further here. Further details of the analysis strategy can be found in Ref. [27].

Due to the challenges of Monte Carlo generation at $\langle\mu\rangle = 200$, the number of signal and background events is limited. This required a loosening of some of the selection criteria typically used in VBF analyses. Therefore, the conclusions of this study apply to a VBF preselection instead of a fully emulated Run 4-5 VBF Higgs to invisible analysis. This will illustrate potential gains that could be achieved until a more sophisticated analysis with larger Monte Carlo statistics is available.

Events are required to have at least two jets with leading and second-leading jet with $p_{\mathrm{T}}^1 > 75$ GeV and $p_{\mathrm{T}}^2 > 50$ GeV, respectively. Furthermore, the two leading jets are required to have $\Delta\eta(j_1, j_2) > 3$. All jets are required to pass an ITk-only $R_{p_{\mathrm{T}}}$ pileup jet tagger [28] operating at 85% hard-scatter efficiency. The looser selection requirements on $\Delta\eta$ and the lack of m_{jj} cut results in topologies with less forward activity than what is expected with a realistic (tighter) selection on these variables. Hence, it is expected that timing information will lead to larger improvements than what is reported in this study.

In the Run-2 VBF Higgs to invisible analysis, an additional selection is made requiring that the $\Delta\phi$ angle between the two leading jets is less than 2. This cut helps to reduce the impact of background events consisting of forward dijet pileup interactions. The effect of this requirement was checked and found to be insignificant within the statistical precision

of this study. This is likely due to the fact that the pileup jet structure at high luminosity is different from that in Run 2 such that simple extrapolations from Run 2 pileup expectations are not necessarily accurate. For example, the relative fraction of stochastic vs QCD pileup jets depends on the luminosity, with the former being larger at high μ, and the $\Delta\phi$ angle requirement is mostly useful for rejecting the case with back-to-back QCD jets. The Run 2 analysis additionally applies selections on the scalar and vector sum of event momenta (missing H_T and missing E_{T}, respectively). These selections have been shown to suppress events with two forward jets in the Run 2 selection. Further studies are required to see if a similar benefit is possible in the higher pileup environment at the HL-LHC.

There are two main ways in which the HGTD can be used to enhance the suppression of Z+jet pileup background and increase the signal acceptance, depending on the number of jets within the HGTD acceptance. First, by the use of the jet-by-jet pileup suppression technique described in Section 3.3.3. This means finding the event vertex time, and using it as a reference when comparing it with the times of tracks within each jet. As studied using the signal events and presented in Figure 3.20(b), the rejection of forward pileup jets with $p_{\mathrm{T}} > 50$ GeV is improved by approximately a factor of 1.2 with the HGTD, assuming a hard-scatter jet efficiency of 85%. The performance improvement is largely limited by the ability to find the correct vertex time. For events in which both jets are forward (FF), however, the knowledge of the vertex time is not so critical. This is because it is enough to compare the relative time of both forward jets to determine if they are compatible with each other (i.e. both jets originate from the same interaction vertex) or not (each jet comes from a different vertex). In this context, HGTD is expected to provide higher levels of improvements in the FF category, compared to the simpler approach consisting of applying the pileup jet tagging algorithm to both jets independently. The use of a dedicated pileup *e*vent tagging algorithm for the FF category, however, was not considered in this document due to lack of time. Such method is expected to be developed as a next step. The relative fraction of CF and FF events for signal and Z+jet background event as a function of m_{jj} is shown in Figure 3.23. At low and moderate values of m_{jj}, where S/B is low, the majority of the events are in the central-forward category, followed by forward-forward, and central-central. As m_{jj} (and, as result, S/B) increases, the relative fraction of events with two forward jets within the HGTD acceptance also increases. This increase is more prominent for background events and, as a result, of larger pileup-jet contributions.

Whereas this analysis is based on full ATLAS detector simulation, the impact of HGTD was estimated in a parametrised way, taking as reference the ROC curves of pileup jet suppression obtained from full simulation. Assuming a fixed ITk-based hard-scatter selection efficiency of 85%, events were reweighted as a function of the pileup jet rejection efficiency gain relative to the ITk-only scenario.

Figure 3.24 shows the expected gain in S/B as a function of the improvement in pileup jet suppression efficiency, normalized to the ITk-only performance ($S/B = 1$ when the x-axis is unity, by construction). For a particular pileup jet efficiency gain from the application

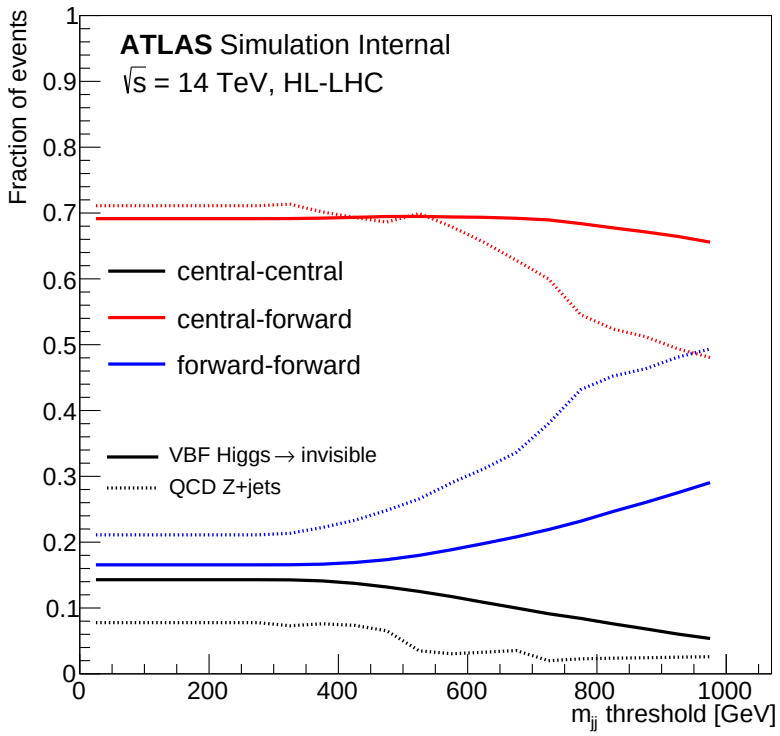


Figure 3.23: The dashed line shows the fraction of signal VBF $H \rightarrow$ invisible and Z+jet background events as a function of a m_{jj} threshold after a loose VBF preselection. Forward jets are those with $|\eta| > 2.4$. Solid (dotted) lines correspond to VBF $H \rightarrow$ invisible (Z+jet) events. The fraction of central-central, central-forward, and forward-forward events are shown in black, red, and blue colors respectively.

of HGTD relative to ITk, the dotted blue line shows the corresponding gain in S/B. For example, a 20% increase in forward pileup rejection for jets above 50 GeV (without any loss in hard-scatter efficiency), would correspond to a S/B gain of approximately 15% (corresponding to the y-axis when the horizontal value is 0.80). Figure 3.24 includes two additional curves that are useful to understand the contribution of the CF and FF topologies separately. The black (red) curve corresponds to the case where HGTD is *o*nly used in CF (FF) events. For the particular event selection cuts used in this analysis, and for a jet-by-jet pileup suppression improvement of 20%, the S/B gain obtained from using HGTD on FF events only is comparable to that of CF events only. Figure 3.24 and the relevant conclusions were made for the loose VBF selection, where the $m(jj)$ threshold is small. Higher $m(jj)$ requirements are expected to further enhance the FF contribution.

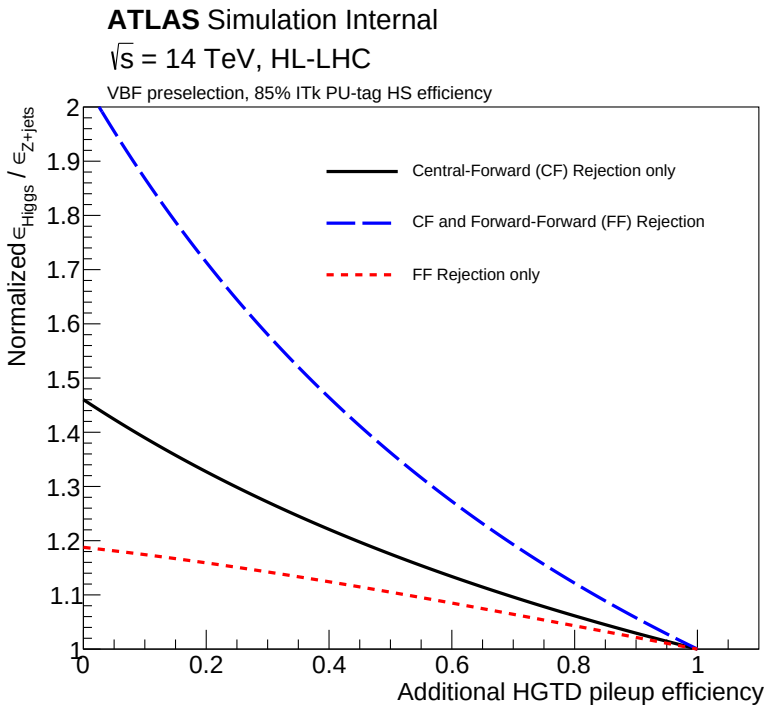


Figure 3.24: Normalized signal over background gain relative to ITk-only pileup jet suppression performance, as a function of the additional pileup jet suppression from HGTD. The solid black (dotted red) line represents the HGTD improvement from the CF (FF) event topologies separately. The dotted blue line shows the total improvement when the combined HGTD+ITk pileup suppression algorithm is applied to all jets in the event.

3.4.2 Vector boson scattering

The study of VBS diboson production is a salient piece in the physics programme for the HL-LHC, due to the sensitivity it provides to probe the nature of electroweak symmetry breaking [29–31]. At the LHC, the electroweak (EW) production of diboson and two jets is the main channel used to study VBS, where the irreducible background typically originates from the QCD production of the same final state. In this section, the potential improvement which the HGTD could bring is discussed using the example of VBS $WZjj$. The relevant measurements of EW $WZjj$ production are currently limited by the data statistics [32, 33], and a more precise measurement is foreseen at the HL-LHC [34], due to the larger integrated luminosity and the upgraded detector capabilities. Improvements to VBS analyses could increase the reach to higher q^2 (higher m_{jj}) and potentially give access to longitudinally polarised diboson production which is of particular interest for studying electroweak symmetry breaking. The HGTD can help improve the measurement by further rejecting pileup jets in the forward region, and this is particularly important for certain phase spaces in the charac-

teristic distributions which are largely contaminated by background events with forward pileup jets. The studies in this section are limited to showing the fraction of background due to pileup, thereby illustrating the areas where HGTD can provide improvement.

The study in this section was performed using a fast simulation of the ATLAS detector [35]. The trigger, reconstruction and identification efficiencies, the energy and transverse momentum resolutions of leptons and jets are computed (as a function of η and p_{T}) from tabulated values that were evaluated using full simulation. The signal sample generation and event selections outlined below follow closely the study done for 2019 CERN Yellow Report which was part of the 2019 CERN Yellow Report documenting the expected performance of the ATLAS and CMS experiments at the HL-LHC [34, 36]. The signal EW $WZjj$ was simulated using Sherpa 2.2.2 [37], while the irreducible background from QCD $WZjj$ production was simulated using Sherpa 2.2.1. Other small background contributions are not considered in this study. Pileup interactions are generated with PYTHIA8 and with an average of 200 interactions per bunch crossing.

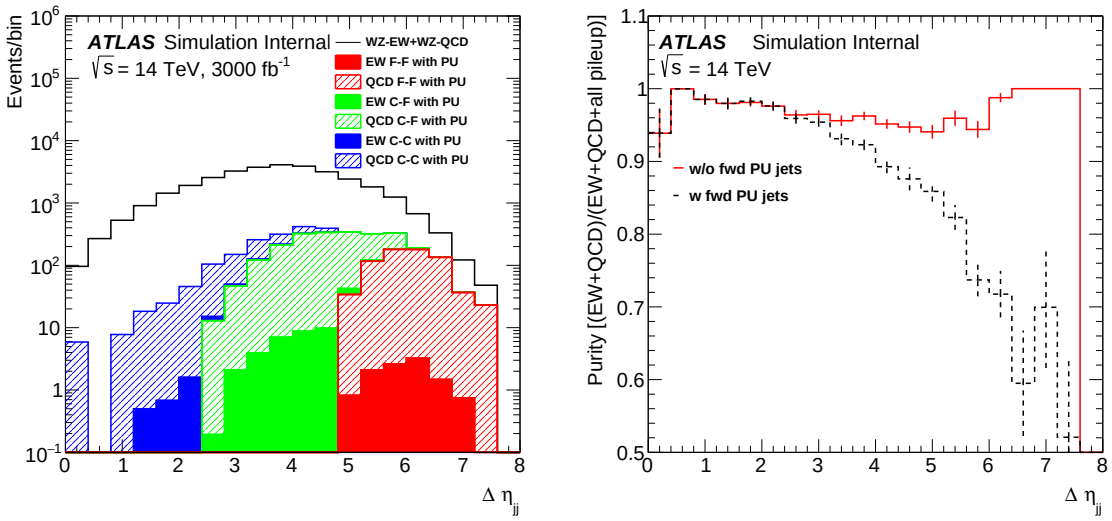
Events with exactly three leptons (e or μ) are selected, where e (μ) must have $p_{\mathrm{T}} > 15$ GeV, $|\eta| < 4$ (2.7), and pass the loose (tight) identification criteria. At least one lepton candidate is required to have $p_{\mathrm{T}} > 25$ GeV. The event must have at least one same-flavour opposite-sign lepton pair, with an invariant mass that is consistent within 10 GeV of PDG m_Z, and the Z boson candidate is formed by the pair which gives the invariant mass closest to m_Z. The third lepton is assigned to the W boson and its p_{T} is required to be greater than 20 GeV. Finally, the transverse mass of the W candidate, computed using the $E_{\mathrm{T}}^{\mathrm{miss}}$ and the p_{T} of the third lepton, is required to be above 30 GeV. At least two jets with $p_{\mathrm{T}} > 30$ GeV in opposite hemisphere and with $|\eta| < 3.8$ are required, with an invariant dijet mass $m_{jj} > 500$ GeV.

The opportunity for improvement by using information from the HGTD is investigated by removing the contribution of forward pileup jets. The focus is on two different aspects. First, the purity of the sum of the EW $WZjj$ and QCD $WZjj$ processes which interfere and cannot be computed separately at higher orders. Second, two differential distributions are studied, selected for their ability to probe the theoretical modelling, to discriminate between EW $WZjj$ and QCD $WZjj$, and their sensitivity to anomalous quartic gauge couplings (QGC) [38].

For the first case, one example is the $\Delta\eta_{jj}$ distribution shown in Figure 3.25. The contamination originating from forward pileup jets constitutes about 50% of the selected events for $\Delta\eta_{jj} > 5$. If completely suppressed, the purity (Figure 3.25(b)) of the sum of EW $WZjj$ and QCD $WZjj$ would approach 100% over the entire kinematic range, providing increased sensitivity to new physics at large momentum transfer.

For the second case, a variable considered for its discriminating power between EW and QCD WZ production is the centrality[4] shown in Figure 3.26(a). The pileup component of the EW and QCD $WZjj$ distributions mainly concentrates under the EW distribution (not

[4] The centrality is here defined as $\min(\Delta\eta^{\mathrm{min}}, \Delta\eta^{\mathrm{max}})$, where $\Delta\eta^{\mathrm{min}} = \eta_{\ell}^{\mathrm{min}} - \eta_{j}^{\mathrm{min}}$ and $\Delta\eta^{\mathrm{max}} = \eta_{\ell}^{\mathrm{max}} - \eta_{j}^{\mathrm{max}}$ are calculated using the minimum and maximum lepton and jet pseudorapidities.



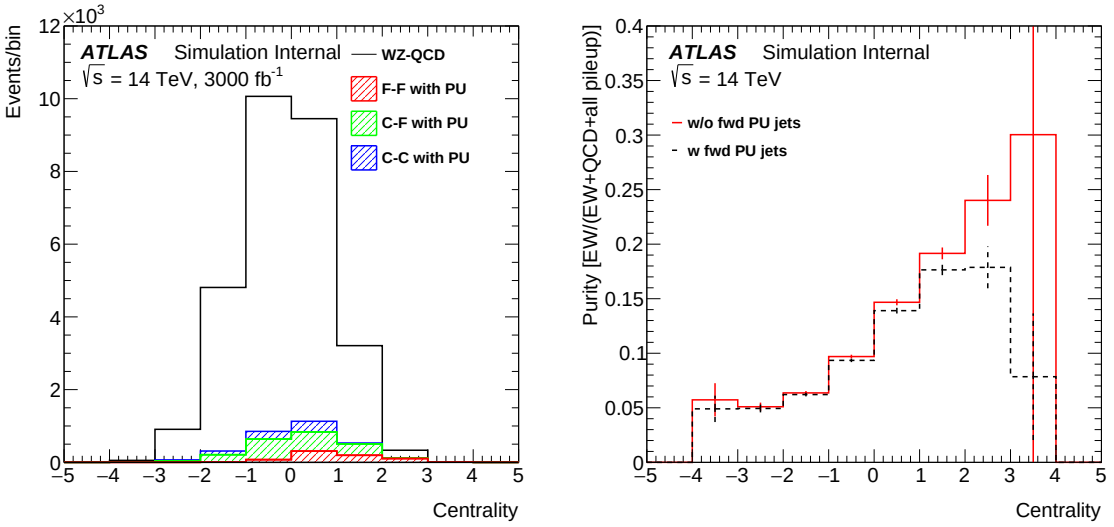
(a) $\Delta\eta_{jj}$ with breakdown of pileup backgrounds. (b) Impact of removing forward pileup jets.

Figure 3.25: Distribution of $\Delta\eta_{jj}$ for (a) EW $WZjj$ plus QCD $WZjj$ where backgrounds with at least one pileup jet are highlighted. FF, CF and CC refer to the two jets being either both forward ($|\eta| > 2.4$), one central and one forward, or both central. (b) is the ratio of EW+QCD $WZjj$ events divided by all selected events vs $\Delta\eta_{jj}$ with and without events of CF/FC and FF jets with at least a forward pileup jet.

shown here), at slightly higher value of the centrality, while the QCD distribution is centered around zero. By removing the forward pileup jets contribution the EW $WZjj$ purity increases by 30% for the values of centrality over 2, so that the EW $WZjj$ signal becomes larger than that contaminated by pileup (Figure 3.26(b)). The high centrality region is more EW $WZjj$ enriched and is also expected to be more sensitive to new physics effects. Therefore a higher purity in this region is beneficial as it isolates a pileup free EW $WZjj$ set of events.

The differential distributions shown here illustrate how improved rejection of forward pileup jets can improve the purity of the sum of EW+QCD $WZjj$ as well as the separation power between the EW $WZjj$ and QCD $WZjj$ processes. The improved purity allows for more sensitive tests of $WZjj$ production in spite of the high-pileup environment.

In summary, the above studies illustrate the importance of background from pileup jets in forward VBS topologies. The extended tracker acceptance of the ITk together with the HGTD timing information will help identify and remove such backgrounds and increase the purity of the event selection in crucial regimes for the study electroweak scattering processes.



(a) Pileup contributions for QCD WZ.

(b) Purity of EW WZ with and without backgrounds with forward pileup jets.

Figure 3.26: Centrality distributions for (a) QCD $WZjj$ with shown fractional contributions from pileup jets, and (b) is the ratio of EW $WZjj$ divided by all selected events in bins of centrality

3.4.3 Measurement of $\sin^2\theta_{\text{eff}}$

In the Standard Model, the Z boson couplings differ for left- and right-handed fermions due to the mixing between the neutral states associated to the $U(1)$ and $SU(2)$ gauge groups. The difference leads to an asymmetry in the angular distribution of positively and negatively charged leptons produced in Z boson decays and depends on the weak mixing angle, $\sin^2\theta_{\text{eff}}$ [39]. Experimentally, this asymmetry can be expressed by

$$A_{\text{FB}} = \frac{N(\cos\theta^* > 0) - N(\cos\theta^* < 0)}{N(\cos\theta^* > 0) + N(\cos\theta^* < 0)},$$

where θ^* is the angle between the negative lepton and the quark in the Collins-Soper frame [40] of the dilepton system. In this formalism, the quark is always assumed to moves in the direction of the boost of the dilepton system. This asymmetry is enhanced by Z/γ^* interference and exhibits significant dependence on the dilepton mass.

The weak mixing angle is one of the fundamental parameters of the SM. Several measurements of $\sin^2\theta_{\text{eff}}$ have been made at previous and current colliders, and the current world average of $\sin^2\theta_{\text{eff}} = 0.23153 \pm 16 \times 10^{-5}$ is dominated by the combination of measurements at LEP and at SLD, which, however, exhibit a tension. At LHC, the best sensitivity to $\sin^2\theta_{\text{eff}}$ is at high Z rapidity when at least one lepton is present in the forward region [41]. Only Z bosons decaying to electrons are considered in this analysis since this final state provides the best experimental precision within the largest acceptance.

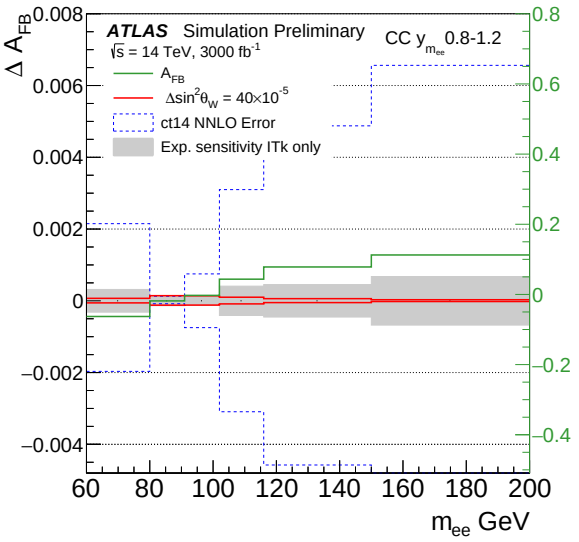
Simulated $Z/\gamma^* \to ee$ signal samples at $\sqrt{s} = 14$ TeV are smeared to match the expected detector response. The performances of the upgraded ATLAS detector [42] in the high pileup environment of the HL-LHC are emulated using the physics object performance recommendations in Ref. [43]. The fiducial acceptance of $Z/\gamma^* \to ee$ events is split into three independent channels depending on the electron $|\eta|$: CC, CF, FF when C represents electron reconstructed in the central region ($|\eta| < 2.47$) and F represents electron reconstructed in the forward region ($2.5 < |\eta| < 4.2$). Both electrons are required to have $p_{\mathrm{T}} > 25\,\mathrm{GeV}$. The invariant mass of the electron pair is required to be loosely consistent with the Z boson mass, $60 < m_{\ell\ell} < 200\,\mathrm{GeV}$, and the events are further categorised in 10 equal-size bins in absolute dilepton rapidity up to $|y_{ee}| = 4.0$.

The contribution of jets misidentified as electrons is suppressed using a tight electron identification and a track isolation requirement. The use of the calorimeter-based isolation and its potential improvements at HL-LHC is not considered in this study. In the forward region, the timing information provided by the HGTD is used to improve the electron isolation by rejecting additional tracks from interactions close in space, but separated in time from the hard-scatter vertex. The purity of the candidate sample is determined with simulation, and is found to be greater than 99% in the CC channel, between 90 and 98% in the CF, and between 60 and 90% in the FF channel. In the FF channel, there is a possibility that both electrons can get their charges measured wrongly, which will introduce ambiguity to the determination of A_{FB}. This effect is not considered in this study, and is not expected to affect the conclusion drawn in this study, which is expressed as the relative improvement due to the inclusion of the HGTD. The signal-over-background ratio with HGTD is up to 20% higher with respect to the case of ITk only in the CF channel.

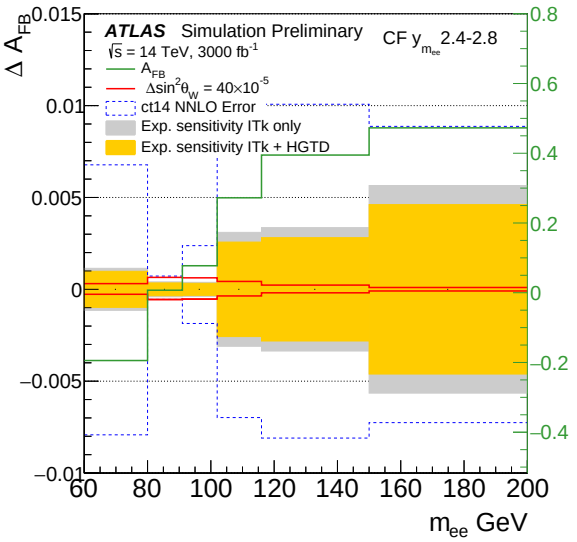
A_{FB} is calculated from the selected electron pairs, and unfolded to correct for detector effects and migrations in $m_{\ell\ell}$ and $|y_{ee}|$ bins. In the CF and FF channels migrations in the $m_{\ell\ell}$ are up to 50% and 60% respectively. Various sources of uncertainty are considered. Those associated with backgrounds are mostly relevant in CF and FF channel and are estimated to be 5% on the background yield and considered uncorrelated among the $m_{\ell\ell}$ and $|y_{ee}|$ bin.

Significant uncertainties arise from knowledge of the momentum scale and resolution for the electrons. Following Reference [44] a systematic of 0.5% (0.7%) is considered to account for possible non-linearity in the energy scale of electron reconstructed in the central (forward) region with $E_{\mathrm{T}} < 55\,\mathrm{GeV}$ and up to 1.5% (2.1%) for central (forward) electron with $E_{\mathrm{T}} > 100\,\mathrm{GeV}$.

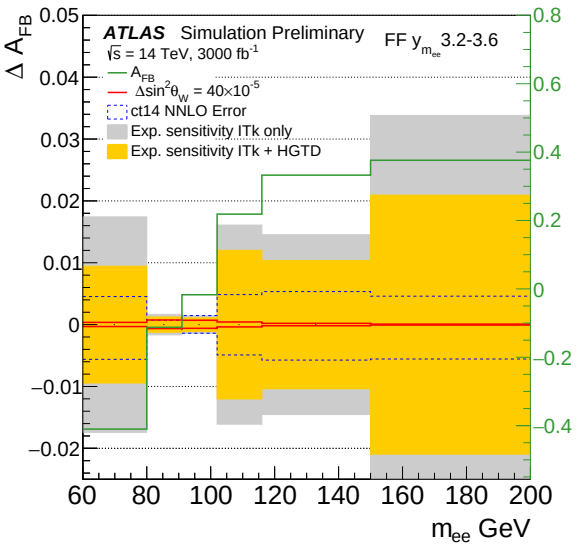
The expected sensitivity to particle level A_{FB} as a function of m_{ee}, for an integrated luminosity of 3000 fb^{-1}, is shown in green in Figure 3.27 for each channel for the chosen rapidity bin. As expected the largest asymmetry is observed in the CF channel. The extraction of $\sin^2\theta_{\mathrm{eff}}$ is done by minimising the χ^2 value between particle-level A_{FB} distributions with different weak mixing angle hypotheses, at LO in QCD, with the NNLO CT14 parton distribution function (PDF). As shown in Figure 3.27, the imperfect knowledge of the PDF results in sizeable uncertainties on A_{FB}, in particular in regions where the absolute values of the



(a) Central-central (CC)



(b) Central-forward (CF)



(c) Forward-forward (FF)

Figure 3.27: Distribution of ΔA_{FB} as a function of mass for the CC, CF and FF channels. The filled bands correspond to the experimental sensitivity with and without the HGTD. The solid red lines correspond to a variations of $\sin^2\theta_{\mathrm{eff}}$ corresponding to 40×10^{-5}. The dashed blue lines illustrate the total error from CT14 NNLO PDF. Overlaid green line shows the particle-level A_{FB} distribution.

asymmetry is large, i.e. at high and low $m_{\ell\ell}$. On the contrary, near the Z boson mass peak, the effect of varying $\sin^2\theta_{\text{eff}}$ is maximal, while being significantly smaller at high and low masses. Thus, in this projection a global fit is performed where $\sin^2\theta_{\text{eff}}$ is extracted while constraining at the same time the PDF uncertainties [41]. With this analysis, the expected sensitivity of the extraction of $\sin^2\theta_{\text{eff}}$ are respectively 25×10^{-5}, 21×10^{-5} and 40×10^{-5} for the CC, CF and FF channel. The uncertainty of the results is dominated by the currently limited knowledge of the PDFs. If looking purely at the experimental uncertainties, including the HGTD in the ATLAS forward region brings a 13% improvement on the $\sin^2\theta_{\text{eff}}$ sensitivity in the CF channel. Combining the three channels together the expected sensitivity reaches a precision of $\Delta\sin^2\theta_{\text{eff}} = 18 \times 10^{-5}$ ($\pm 16 \times 10^{-5}$ (PDF) $\pm 9 \times 10^{-5}$ (exp.)) which exceeds the precision achieved in all previous single-experiment results so far.

3.4.4 Impact of the luminosity uncertainty

Many high-precision cross section measurements at the HL-LHC will be limited by the uncertainty in the integrated luminosity. That uncertainty affects not only the normalisation of the signal, but also that of any background not determined from data, thus enters cross-section measurements in a twofold way. At the ECFA HL-LHC Experiments Workshop in Aix-Les-Bains in 2016, it was stated that "Experimental progress on luminosity determination may be the keystone for precision physics at the HL-LHC". This applies to the Higgs boson physics programme, where the luminosity uncertainty could be the dominant source of systematic uncertainty unless the error is reduced to approximately 1% (compared to the best measurement of 1.7% for ATLAS in Run-2 [45]), despite the much harsher environment for the luminosity measurement at the HL-LHC. The percentage precision of the luminosity measurement is also critical for measuring important SM processes, such as W and Z boson production, and single and pair production of top quarks. Further information about the expected performance of the luminosity determination at the HL-LHC can be found in Ref. [35].

The HGTD has been designed with luminosity determination capabilities in mind from the beginning, and a few unique capabilities will provide important information that can help constrain the total luminosity uncertainty to a low level. With its relatively low occupancy even at the highest anticipated luminosities, the detector response is nearly perfectly linear as a function of $\langle\mu\rangle$ (see Section 10.3). This characteristic, combined with being able to operate both at the low interaction rates during a van der Meer scan and at $\langle\mu\rangle = 200$, can help constrain the significant uncertainty component otherwise incurred when extrapolating from low to high $\langle\mu\rangle$. Secondly, the timing resolution of the HGTD allows measuring and subtracting difficult transient backgrounds in the high-radiation environment. The uncertainties of such backgrounds (e.g. so-called *afterglow*) can otherwise limit the precision of methods relying on hit- or track-counting techniques. Finally, the HGTD will have a dedicated readout path for sending occupancy data for each module at 40 MHz, allowing

bunch-by-bunch luminosity measurements online without trigger bias. This is important for promptly feeding back luminosity information to the machine for luminosity-levelling purposes which are of increased importance at the HL-LHC. The technical details of the implementation of the luminosity capabilities, including the method and treatment of systematic uncertainties, are given in Section 10.3. The description of the occupancy readout in the ASIC is given in Chapter 6.

As input to the update to the European Strategy for Particle Physics in 2020, the expected performance for many analyses at the HL-LHC was studied both in ATLAS and CMS [46]. Unfortunately, there is no breakdown of the impact of individual sources of systematic uncertainties for any of the ATLAS combined Higgs boson measurements. Such breakdowns exist however for some of the single-channel measurements. Since the Higgs boson analyses cannot constrain the uncertainty on the luminosity, it is straightforward to compare any value for the luminosity uncertainty to the magnitude of the other uncertainties affecting these analyses.

Table 3.1 lists the largest sources of uncertainty, aside from integrated luminosity, affecting three important Higgs boson cross section measurements; gluon-fusion (ggH) production of Higgs bosons with decays to $\gamma\gamma$ and ZZ^*, and combined gluon-fusion and vector boson fusion (VBF) production of Higgs bosons with decay to $\tau\tau$. For all these measurements, an uncertainty of 2% on the integrated luminosity would be the single largest source of uncertainty on the results.

Analysis channel	**Largest uncertainty**	$\Delta\sigma/\sigma_{\mathrm{SM}}$
Cross section for $ggH(\rightarrow \gamma\gamma)$	Photon isolation efficiency	1.9%
Cross section for $ggH(\rightarrow ZZ^*)$	Electron eff. reco. total	1.5%
Cross section for $ggH + \mathrm{VBF}$, $H \rightarrow \tau\tau$	QCD scale ggH, $p_{\mathrm{T}}^{H} \geq 120\,\mathrm{GeV}$	1.7%

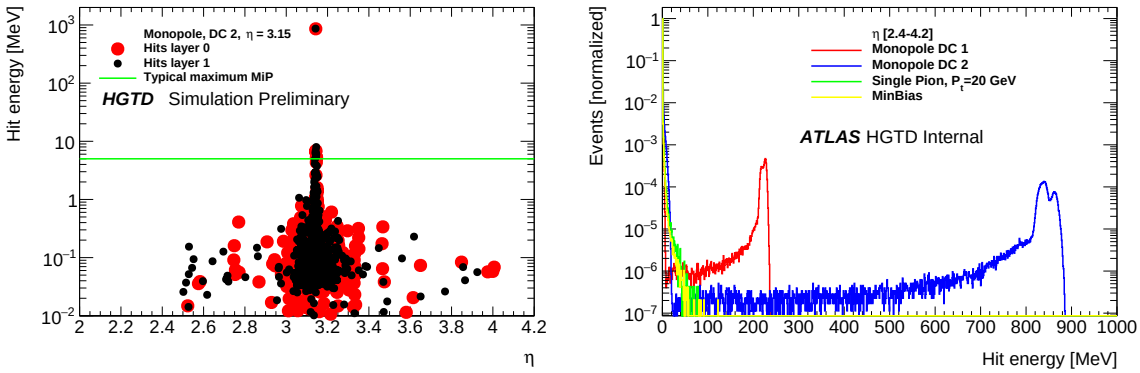
Table 3.1: List of dominant uncertainties (excluding the uncertainty on the integrated luminosity) affecting various expected Higgs boson cross section results at the HL-LHC using 3000 fb^{-1} of data. An uncertainty on the luminosity measurement of 2% would be the dominant source of uncertainty for all these measurements.

The above considerations illustrate the importance of a precise luminosity measurement for the Higgs boson physics programme at the HL-LHC. The same concerns apply to any measurement of processes with similar, or larger, cross sections compared to the Higgs boson.

Finally, it is important to stress that precision in the luminosity programme can only be achieved by having several independent luminosity detectors, and that any single detector will not be able to achieve the precision goals. With readout at 40 MHz, and with occupancy determination in a dedicated sideband time window (further described in Section 6.2.1), the HGTD has unique capabilities compared to other silicon detectors. These are aimed at constraining the sources of systematic uncertainties affecting the luminosity determination to the percent level or better.

3.4.5 Trigger for magnetic monopole searches

Magnetic monopoles (single pole of magnetic charge) are hypothetical elementary particles which appear in several models beyond the Standard Model [47–50]. Monopoles would appear as long-lived particles which would give dramatic ionisation since a monopole with one Dirac unit of magnetic charge charge, DC 1, is ionisation-wise equivalent to an electric charge of 68.5e. Searches for monopoles have been conducted with the ATLAS detector for central signature in the electromagnetic calorimeter [51–53], and with the dedicated MoEDAL experiment at LHCb [54].



(a) (b)

Figure 3.28: (a) Energies of simulated hits in HGTD in one single monopole event. (a) Distributions of hit energies in simulated events of minimum bias interactions and single-particle samples with pions and monopoles.

However, according to a recent review, the exclusion limits presented in the previous sentence do not apply for scenarios with monopoles with $m < 200$ GeV [55]. The HGTD could enhance the monopoles discovery capability of ATLAS in HL-LHC by providing an online trigger to highly ionising particle. The HGTD response to single monopoles with a mass of 200 GeV and magnetic charges of 1 DC and 2 DC was simulated.The distribution of simulated hit energies in the HGTD for a single monopole event is shown in Figure 3.28(a). Single high-energy HGTD hits from monopoles are clearly separated from the deposits from MIPs as seen in Figure 3.28(b). This can provide a clear and unique signature that can be exploited by HGTD electronics to recognise candidate Monopole events at trigger level. The HGTD will send hit summary data at 40 MHz to a dedicated luminosity processing system (Section 10.3) so a special bit can be added to flag high-energy depositions in a single pad. If one channel reports such a high-energy deposit, the corresponding readout ASIC will report a reserved word in place of the luminosity hit counts. If the off-detector luminosity processing electronics receives such a signal, it will send a special trigger signal to the Central Trigger Processor, and dedicated algorithms implemented in software in the Event Filter can investigate the event further. The detection of very high-energy hits requires

modifications to the ALTIROC (Chapter 6) readout chip (addition of a second very high threshold discriminator, Section 6.2), and is not included in the current baseline design. If not realised for the nominal ALTIROC, the proposed functionality could be implemented in future versions of the HGTD readout and be installed with the scheduled module-ring replacements.

3.4.6 Limitations in the Monte Carlo modeling

Finally, two aspects of the simulated Monte Carlo samples used for these studies are worth discussing.

Modeling the jet activity in the forward region is challenging as the current trackers in ATLAS and CMS only extend to $|\eta| < 2.5$. Charged-particle multiplicities in inelastic proton–proton collisions have been measured in this region using the *Minimum-Bias Trigger Scintillator* detector [56], and transverse energy flow using calorimeter measurements [57], both indicating the shortcomings of the standard ATLAS PYTHIA tunes when it comes to describing forward activity. Studies in high-pileup data in Run 2 show significant discrepancies with more forward jet activity than predicted by the simulations tuned to data recorded at low μ. Several mechanisms could contribute to this discrepancy, ranging from poor modeling of the particle-level process to effects of calorimeter calibration affecting seeding of energy clusters in the forward region during reconstruction. While it is likely that several factors conspire to give rise to this discrepancy, which grows rapidly with μ, it cannot be excluded that the minimum-bias events used to simulate pileup interactions in the samples used for the studies in this document severely underestimate the forward jet activity expected at the HL-LHC. This would imply more background, and a stronger need for the HGTD.

In addition, the samples used for the studies in this chapter have a spatial beam-spot spread of $\sigma_z \approx 50$ mm. In the scenarios under discussion for HL-LHC operation, the beam spot extension in z would plateau at slightly lower values around 40 mm, yielding $\sim$10% higher vertex densities. A beam spot that is more spatially compressed makes it more challenging to do pileup suppression with the ITk alone, and the relative improvement from HGTD would likely be larger.

4 Technical Overview

4.1 Introduction

This chapter summarizes the fundamental aspects of the design of the HGTD. The main requirements that drive the design and the proposed technical solutions are discussed in this chapter. Measurements from the on-going R&D program are presented, especially on sensors and electronics, that demonstrate the achieved performance.

4.2 Detector overview and key requirements

The detector has been designed for operation with 200 proton-proton collisions per bunch crossing and a total integrated luminosity of 4000 fb^{-1}. The HGTD will be located in the gap region between the end of the ITK and the end-cap calorimeter, at a distance of approximately ±3.5 m from the interaction point. Figure 4.1 shows a transverse view of the detector, without the front cover of the vessel, where the front layer of the first double-sided active layer (in blue) and the peripheral electronics boards (PEBs) location(in green) can be seen. The envelope of the detector vessel has a radial extent of 110 to 1000 mm. The envelope in z, including the moderator, supports and front and rear vessel covers is 125 mm. This includes the moderator that is placed behind the HGTD with a total thickness of 50 mm, to reduce the back-scattered neutrons created in the end-cap/forward calorimeters, protecting both the ITk and the HGTD. Each end-cap is made of one hermetic vessel, two instrumented double-sided layers (mounted in two cooling/support disks), and two moderator pieces placed inside and outside the hermetic vessel. The weight of an end-cap is approximately 350 kg. The moderator, whose mass is equally distributed between the pieces located inside and outside the vessel, contributes to 150 kg.

The front vessel cover and each cooling/support disk are physically separated in two half circular disks to enable the opening of the detector in the presence of the beam pipe.

The active detector element is made of Low-Gain Avalanche Silicon Detectors (LGADs) read-out by dedicated front-end electronics ASICs (ALTIROC). It covers the pseudo-rapidity range $2.4 < |\eta| < 4.0$ (120 mm $< R <$ 640 mm). The active area is divided into three rings (inner, middle and outer ring). The inner ring covering the region $3.5 < |\eta| < 4.0$ (120 mm $< R <$ 230 mm) is equipped with modules mounted on the front and back sides of a given

cooling plate, with 70 % overlap along the readout row direction, in order to provide on average 2.7 hits per track in the most irradiated and highest occupancy region.

The middle ring covering the region $2.7 < |\eta| < 3.5$ (230 mm $< R <$ 470 mm) is equipped with modules overlapping 54% providing on average 2.5 hits per track. The outer ring covering the region $2.4 < |\eta| < 2.7$ (470 mm $< R <$ 640 mm) is equipped with modules overlapping only 20% providing on average 2.1 hits per track.

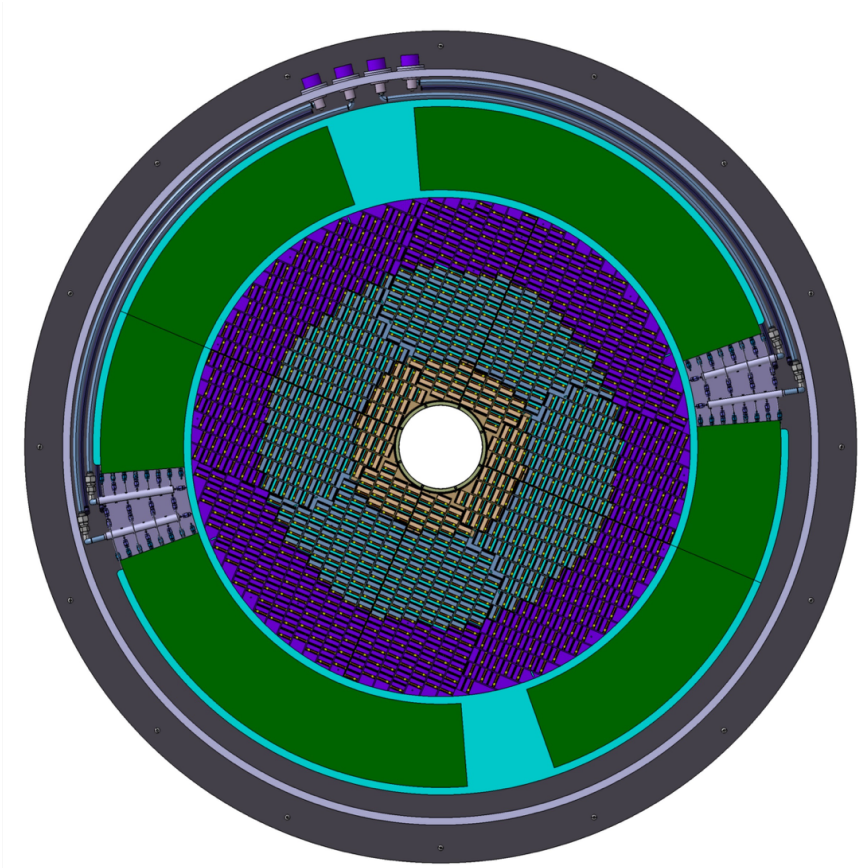


Figure 4.1: Front view of the detector. The active detector region is defined by the three rings and is surrounded by a green area where the PEBs are located, except in two areas needed for the CO_2 cooling manifolds.

4.2.1 Expected Radiation levels

As discussed in Chapter 2, the radiation levels in the forward region exceed the radiation hardness of both the sensors and the front-end electronics, especially at low radius. In order to assure high performance operation during the full life time of the HL-LHC, the

plan is to replace the innermost ring after each 1000 fb^{-1} (3 times in total) and the middle ring after 2000 fb^{-1} (once), during long shutdowns. It is expected that improvement of the LGAD performance with respect to radiation hardness is possible and the plan would be to target slightly more demanding specifications for the inner ring replacement detector to be delivered during LS5. Section 4.2.1 and Section 4.2.1 show respectively the expected 1 MeV neutron-equivalent fluence and TID as a function of the detector radius with the most updated simulation of the ATLAS detector. A safety factor of 1.5 is applied to account for uncertainties in the simulation[1]. An additional safety factor of 1.5 is applied to the total ionising dose (TID) to account for low-dose rate effects on the ASICs. In the inner ring the total Si 1MeV neq has a similar contribution from neutrons and charged particles while in the middle and outer rings the dominant effect comes from neutrons, as seen in Figure 2.14.

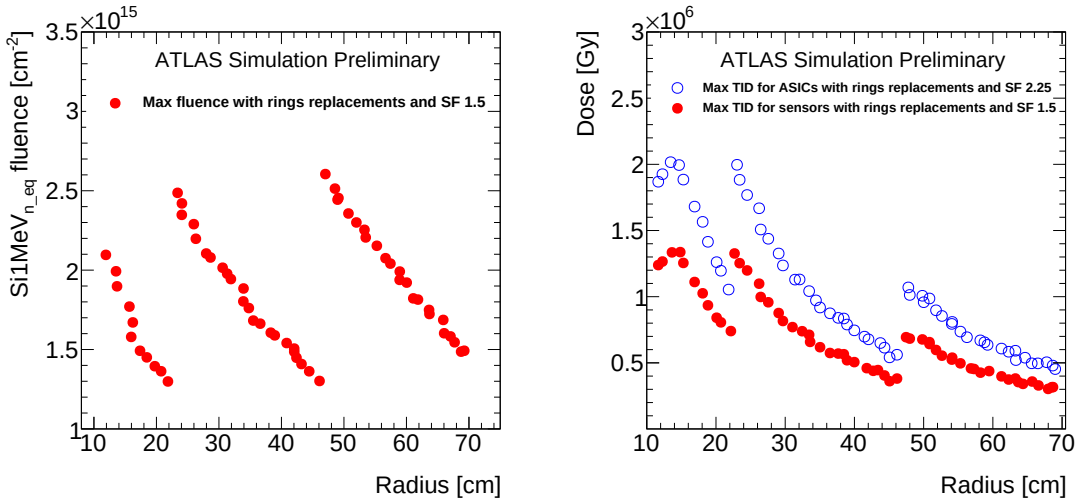


Figure 4.2: Expected Si1MeV$_{n_{eq}}$ radiation levels in HGTD, using Fluka simulations, as a function of the radius considering a replacement of the inner ring every 1000 fb^{-1} and the middle ring replaced at 2000 fb^{-1}. These curves included a safety factor (SF) of 1.5 to account for simulation uncertainty. An additional safety factor of 1.5 is applied to the TID to account for low dose-rate effects on the electronics, leading to a safety factor of 2.25.

The exact radial transition between the three rings will be optimised for the final detector layout, once the FLUKA simulations will be updated with the final ITk layout, and the radiation hardness of the final sensors and ASICs are re-evaluated. The requirement is a maximal fluence and TID not exceeding 2 MGy and 2.5x10^{15} n_{eq} /cm^2 respectively.

[1] This safety factor takes into account the comparison between the Run-1 and Run-2 fluences and the simulation, as well as a possible inaccurate detector description used by FLUKA/GEANT4

4.2.2 Key requirements

A high intrinsic single hit efficiency is essential throughout the lifetime of the HGTD. This puts stringent constraints on the smallest charge to be delivered after irradiation. Taking into account the lowest expected threshold of the electronics discriminator, a minimal charge of 4 fC is required to be delivered by the sensors after irradiation. In addition, a signal-over-noise (S/N) larger than 7 (from testbeam studies) , while keeping a low rate of fake hits induced by the electronics noise (< 0.1 %), is needed. As measured with testbeam, in these conditions an efficiency larger than 95% can be obtained.

The target time resolution per track, combining multiple hits, is from 25 ps at the start of lifetime to 50 ps after 4000 fb^{-1}. To achieve this performance, the time resolution per hit should be about 35 ps at the start of lifetime and not worse than 70 ps at the end of lifetime over the full surface of the detector.

The main contributions to the time resolution of a hit are given by:

$$\sigma^2_{\text{hit}} = \sigma^2_{\text{Landau}} + \sigma^2_{\text{elec}} + \sigma^2_{\text{clock}} \tag{4.1}$$

- where σ_{Landau} is the time resolution contribution induced by the Landau fluctuations in the deposited charge as the charged particle traverses; the sensor
- σ_{elec} is the contribution from the electronics read-out (jitter and time-walk). It is required to be about 25 ps for a particle at its minimum of ionisation (MIP) with a LGAD gain of 20 (corresponding to a charge of 10 fC) at the start of the HL-LHC, and at most 70 ps after 4000 fb^{-1} for a charge of 4 fC. The TDC contribution is expected to be negligible;
- σ_{clock} is the non-deterministic jitter contribution from the clock distribution, expected to be smaller than 15 ps after calibration.

In addition, the detector should be able to distinguish hits in the same pad that come from consecutive bunch crossings and should provide the sum of the number of hits per ASIC for each bunch crossing. The latter is used in the luminosity measurement.

4.2.3 Read-out bandwidth

With the baseline ATLAS architecture, the ATLAS detector is read-out with a single Level-0 (L0) trigger at an maximum rate of 1 MHz, with a maximum latency of 10 µs [58]. The time information of the HGTD hit cells will be read out on reception of this L0 trigger signal. In the evolved scheme considered by ATLAS, called L0–L1, the HGTD will be read-out on the reception of a L1 trigger signal with a maximum frequency of 800 kHz and a maximum latency of 35 µs. The time information from each ASIC is read-out by only one data e-link to the lpGBT [59] (Section 9.1.1). Therefore the maximal bandwidth is limited to 1.28 Gbit s^{-1}.

The luminosity data is transmitted at each bunch crossing to dedicated lpGBTs, requiring a 640 Mbit s^{-1} e-link bandwidth (Table 6.1).

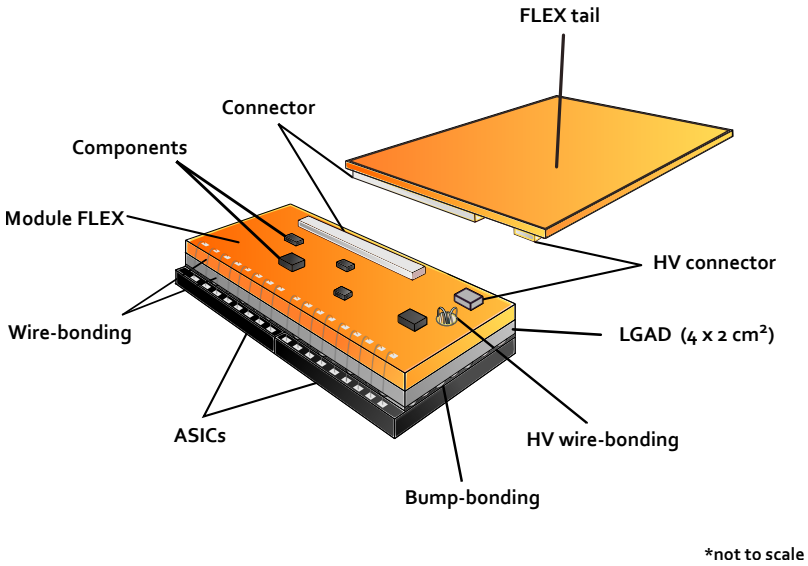


Figure 4.3: View of an HGTD hybrid module equipped with its read-out flex cable tail. The bare module, glued on the flex cable, is made of a $4 \times 2\,\text{cm}^2$ sensor with two bump bonded ASICs. The signal lines of the ASIC are wire bonded on one side of the module flex, while the bias voltage of the sensor is provided to the back-side of the sensor through a hole in the module flex.

4.3 Hybrid HGTD module

Figure 4.3 shows a view of a hybrid module made of two parts: a LGAD sensor and two ASICs, called a bare module, and the flexible printed circuit board (flex cables). The flex is made of two pieces, a small flex board permanently glued to the bare module and a long flex tail whose length, of up to about 60 cm, depends on the module position in the detector. The sensor and the ASICs are connected through a flip-chip bump bonding process called hybridization. All connections between the ASIC and the peripheral electronics are routed through the flex cable. The bare module is glued on the back side of the sensor to the flex module small piece, and all the signals are wire bonded between the ASIC and the flex cable and for the high voltage between the sensor and the flex.

The characteristics of the bare modules are:

- The size of the bare modules, all identical, is approximately $2 \times 4\,\text{cm}^2$ and each bare module contains 450 pads (15x30). Its size has been defined to optimize the coverage

at the inner radius and to provide a good yield for the hybridization process. The nominal total number of bare modules is 8032.

- The size of the pad, $1.3 \times 1.3\,\mathrm{mm}^2$, is the result of a compromise between smaller pads, leading to lower occupancy and smaller capacitance and thus low electronics jitter, and larger pads, which provide better geometric coverage with large fill factors and less power dissipation from the ASIC.
- The sensor is connected to two ASICs, each of them reading a matrix of 225 (15x15) pads. The size of the ASIC is about $20 \times 22\,\mathrm{mm}^2$.

The status of the R&D of key components is discussed briefly below, with more technical details in subsequent chapters.

4.3.1 Sensors

The sensors are based on LGAD technology, pioneered six years ago by the Centro Nacional de Microelectrónica (CNM) Barcelona [5] in close collaboration with the RD50 collaboration.

LGADs are n-on-p silicon detectors containing an extra highly-doped p-layer below the n-p junction to create a high field which causes internal amplification as displayed in Figure 4.4(a). When a charged particle crosses the detector, an initial current is created by the drift of the electrons and holes in the silicon. When the electrons reach the amplification region, new electron/hole pairs are created and the holes drift towards the p^+ region and generate a large current. This charge amplification is referred as the gain of the LGAD. This current, much larger than in a standard diode, is the key ingredient to get an excellent time resolution for energy deposited by Minimum Ionizing Particles (MIP). The expected currents for different irradiation levels (therefore different gains) are presented in Figure 4.4(b). For large gain, the rise time is about 500 ps and the signal duration is approximately 1 ns. When the irradiation neutron fluence increases, the charge is smaller and the rise time and the signal duration are shorter.

After amplification in the gain layer, the height of the LGAD signal is proportional to the gain. On the other hand, the slope dV/dt depends on the thickness of the sensor, favouring thin sensors since the electronics jitter scales as the inverse of the slope. However, the jitter also depends linearly on the detector pad capacitance, therefore limiting the potential use of very thin sensors. Consequently, the optimal thickness relies strongly on the performance of the read-out ASIC. The baseline active thickness has been chosen to be 50 µm while the total thickness is 250 µm. The pad size is $1.3 \times 1.3\,\mathrm{mm}^2$ which resulted from an optimization discussed in the previous section.

Over the last five years LGAD sensors have been produced by CNM/Spain, HPK/Japan, FBK/Italy and recently NDL/China with different doping level, active thickness, pad size,

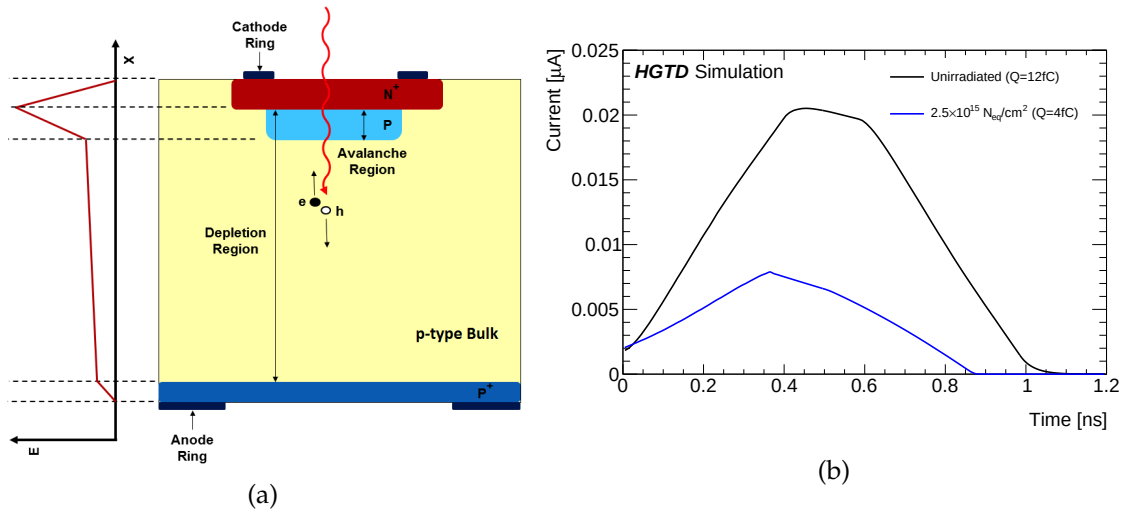


Figure 4.4: Cross section of an LGAD (a) and simulated signal current in LGADs at start and after full integrated neutron fluence (b).

and inter-pad gaps. These detectors have been exposed to protons, neutrons and X-rays up to the expected maximum radiation levels (including the safety factors) and intensively characterized in the laboratory (with probe station, β source, laser) or in beam tests (at CERN, DESY, FERMILAB).

Under irradiation, the expected decrease of the charge yield can be mitigated by increasing the bias voltage (up to 750 V operation voltage) and operating at low temperature (–30 °C). Figure 4.5 summarizes results obtained in the laboratory, with dedicated electronics, for sensors from different producers exposed to a neutron fluence up to 2.5×10^{15} n_{eq} cm^{-2}. A charge of 4 fC can be reached up to a fluence of 3×10^{15} n_{eq} cm^{-2} (Section 5.5.3), providing a time resolution smaller than 70 ps per hit (Section 5.5.5). The performance of sensors from all manufacturers is similar, even if before irradiation the optimal operating bias voltage might be different because the doping profile is different. With a minimal charge of about 4 fC and a discriminator threshold of about 2 fC, a hit efficiency of at least 95% is expected. For the largest fluence, the Boron doping in the gain region has been mostly inactivated and half of the remaining reduced gain is supplied by the bulk diode, due only to the large high bias voltage applied. The time resolution in this domain is fully dominated by the electronics jitter, thus dominated by the ASIC performance at low charge.

Intense R&D is still ongoing to improve the radiation hardness with deep narrow doping implantation and carbon (C) implantation. Depending on the results of these studies, discussed in detail in Chapter 5, the exact radius of the inner and middle rings might be optimized. The operating voltage (V_{op}) needs to be adjusted with respect to the radiation flux.

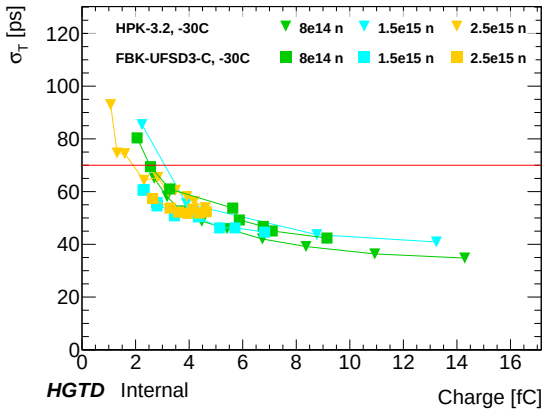


Figure 4.5: Time resolution as a function of the collected charge for neutron irradiated LGADs from different producers (HPK, FBK) with a 50 µm active thickness. These measurements have been made at −30 °C, in the laboratory with a β source using a custom electronics read-out board (not the ALTIROC therefore optimistic with respect to the expected electronics jitter contribution). The typical error bar is about 3 ps. The red line shows that in these conditions better than 70 ps time resolution is obtained at 4 fC

Already, many single pads (> 1000), small arrays of 2×2 and 5×5 pads from various companies have been measured in the laboratory and test beams, showing an excellent yield. The first matrices of 15×15 pads, delivered by HPK, have also been characterized. They show an excellent uniformity both for the operating bias voltage (i.e breakdown voltage) and the low leakage current (see Figure 5.4).

Following almost four years of R&D activities, shared in part with the CMS timing detector and RD50, a first set of criteria for the parameters of the final sensor design has been established, constituting our baseline. These include: 50 µm active thickness, narrow and deep doping profile and a 300 µm slim edge distance with two guard rings. However, some of the parameters will need to be further validated up to the Final Design Review, scheduled for Q4 2021.

4.3.2 Front End ASIC

Taking into account the expected TID radiation levels and needed low jitter, the CMOS TSMC 130 nm technology has been selected. The global architecture of the ASIC, called ALTIROC, is similar to the ASICs developed for pixel detectors but with a significantly reduced number of channels and a quite different single pixel Front End optimized for the time measurement. Figure 4.6 presents the general architecture with a matrix of 225 channels organized along columns for the read-out and with common digital electronics at the bottom.

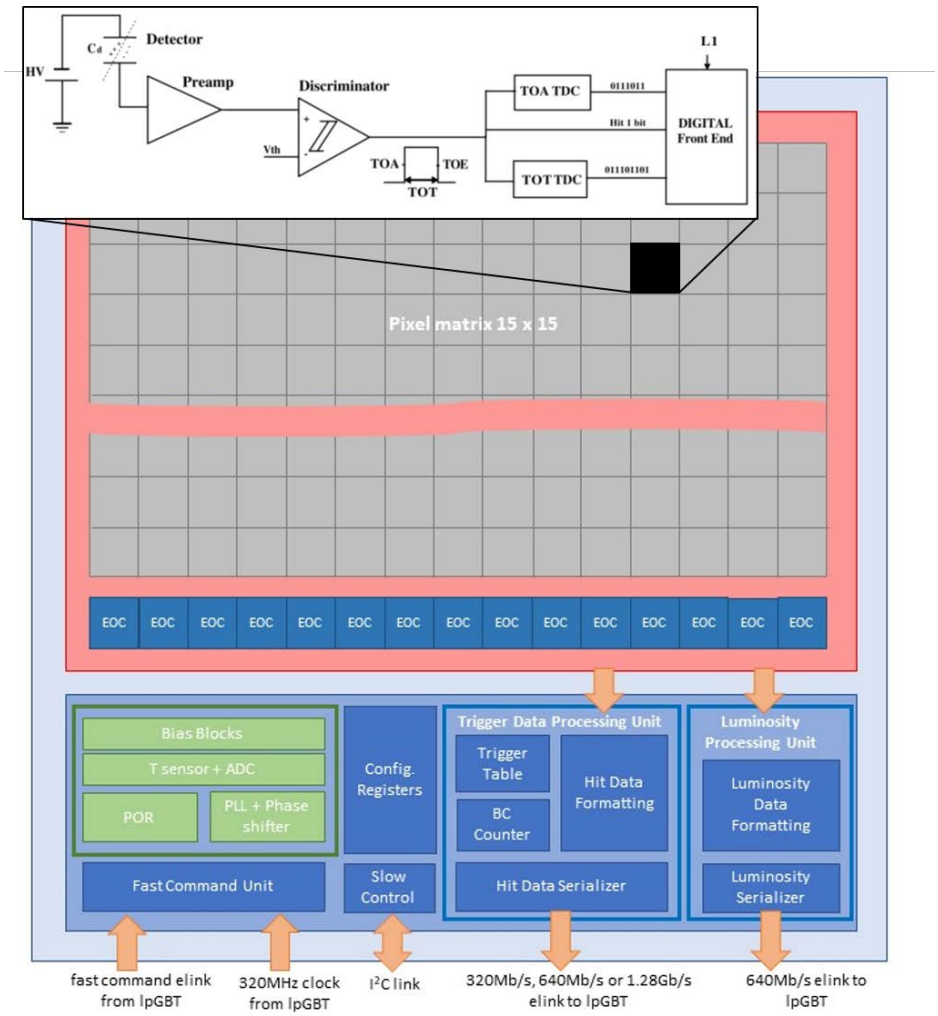


Figure 4.6: Global architecture of the ALTIROC ASIC. The schematic of one Front End electronics channel is displayed on top of the channels matrix, with the preamplifier followed by a discriminator, two TDCs, and a digital front end block.

The analog Front End electronics of each channel is the most critical element to reach low jitter. The sensor signal is amplified using a voltage preamplifier. Taking into account the non-negligible duration of the LGAD signal (approximately 1 ns), a preamplifier with about a 1 GHz bandwidth, is enough. The preamplifier is followed by a fast discriminator. The leading edge of the output (Time Of Arrival, or TOA) provides the start of a Time to Digital Converter (TDC) using a Vernier delay line configuration. The stop is given by the clock. This start-stop structure minimizes the power dissipation when hits are absent. The quantisation step is 20 ps, which does not contribute significantly to the expected time resolution. The TOA measurements are restricted to a 2.5 ns window centered on the bunch

crossing. The expected time dispersion of the hits has a r.m.s of 300 ps so that such a window contains all the hits of the collisions if centered with about 100 ps accuracy with a phase shifter. The falling edge of the discriminator output provides the stop of a second TDC (which uses also the leading edge as start) , with 40 ps quantisation step, in order to measure the Time Over Threshold (TOT), which is used as an estimate of the signal amplitude. The TOT information is used offline to correct the TOA for time walk effect. After correction, the residual variations are well within ±10 ps. The digital Front End is used to store the time data up to the reception of a trigger and buffers the data in order to be read by the End Of Column cells. This buffer is needed to cope with event to event fluctuations in the number of hits and random arrivals of the triggers.

A four-channel prototype (ALTIROC0), bump-bonded to a sensor of 2×2 pads, has first been characterized [60] to select the preamplifier and discriminator architecture. A second 25 channel prototype, including the complete pixel read-out with the TDC and SRAM, has been produced. This second ASIC has been characterized in the laboratory first with the ASIC wire bonded to a specific board (see Figure 4.7), or later bump bonded to a 5×5 pad sensor. This figure also shows preliminary measurements of the jitter and of the TOA as a function of an injected calibration charge. With the ASIC alone, an input capacitance of 4 pF and injecting calibration signal, the threshold can be as low as 2 fC, allowing a measurement of an input charge down to 4 fC. The jitter for a calibration injected charge of 10 fC (4 fC) is about 15 ps (25 ps) with a pad capacitance of 4 pF. With a LGAD sensor connected to the ASIC, the measured jitter on testbench is about 55 ps at 4 fC (see Section 6.7) due to the different input LGAD signal shape. The variation of the TOA versus the input charge, about 300 ps, is compatible with the preamplifier bandwidth. To achieve the target time resolution, this time walk effect needs to be corrected using the TOT information. The TOT has also been measured (see Section 6.7) but it is quite sensitive to any coupling preventing its use in testbeam condition in November 2019[2]. Preliminary measurements with beam show that a time resolution of 46 ps, i.e a jitter of 39 ps, can be reached with non irradiated sensors with a charge about 20 fC. This performance is largely dominated by a noise source coming from the DAQ board discovered after the beam period : with a filtering interface board developed recently, the noise has been been reduced by 35-40 %, so that the testbeam jitter is expected to be to < 30 ps. Testbeam campaign with a new ASIC version and the interface board is scheduled in 2020 to validate this expectation.

The common digital electronics must satisfy a wide variety of requirements. It first retrieves the time information of the matched hits and the luminosity hits sum computed in the End of Column. The luminosity hits are summed in two different windows, a 3.125 ns window centered on the bunch crossing and a second one with a larger size configurable by slow control. In a second step, it formats these data, and provides them to the serializer, which

[2] A output signal of the ASIC (TOA busy) dedicated to the testbeam was used to trigger the external SiPM. This signal induced a strong coupling on the falling edge of the preamplifier output, therefore a distorted TOT distribution)

transfers the data on the e-link to the lpGBT. The speed of the serializer can be selected through slow control at 320 Mbit s^{-1}, 640 Mbit s^{-1} or 1.28 Gbit s^{-1}, in order to maximize the use of the bandwidth. A control unit receives the fast commands from the lpGBT (clock, BCID, L01/L1,...) and through I^{2}C the slow control parameters. A phase-locked loop (PLL) and a phase shifter are used to clean the jitter of the clock and adjust the clocks with a 100 ps step. This allows the time and luminosity windows to be centered on the bunch crossing clock for each individual ASIC. Finally, monitoring blocks are included to measure the temperature and the leakage current.

The next major ASIC iteration, ALTIROC2, will integrate all the functionality of the final ASIC and will have its final size. Triple redundant registers will mitigate against SEE and will be implemented for all control and signals registers but not for the read-out data. The first iteration should be submitted in 2020 and a second iteration one year later. The Final Design Review is planned in Q4 2022.

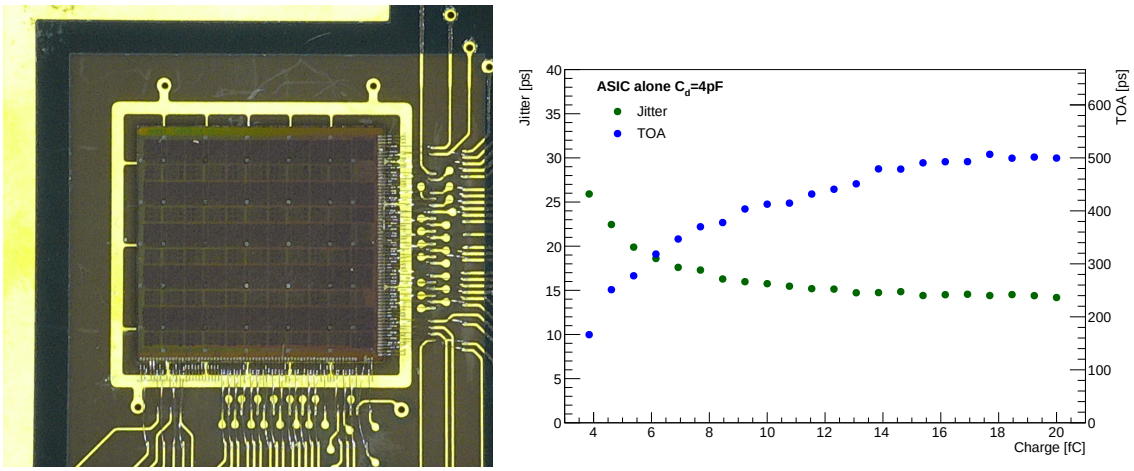


Figure 4.7: Picture of a ALTIROC1 die with 5 × 5 channels wire-bonded on the test board (left) and preliminary measurements of the average time and jitter as a function of the injected charge using calibration injection with one channel of ALTIROC1 (right).

4.3.3 Module assembly

After having qualified separately the sensor and the ASIC at the wafer level, they will be connected through a flip-chip bump bonding process. Under Bump Metallization (UBM) will be deposited on the sensor wafer before dicing. UBM and solder bumps will be deposited on the ASIC wafers. The next step of the hybridisation consists in the flip-chipping during which the sensor and ASIC are aligned, heated, and compressed, so that each solder bump melts and provides the electrical contact between the sensor pad and the channel readout. With the large pad size of 1.3 mm × 1.3 mm, solder bump as large as 90 µm can be used, making

the process standard for many companies, contrary to the hybridization of the ATLAS ITk pixel detector. The bump bonding of the prototypes has been done in collaborating Institutes and also in industry with ALTIROC1 and 5×5 channel sensors (including UBM, bump deposition and flip-chip). Satisfactory performance results have been obtained, both for connectivity and mechanical stress. The qualification of the bump-bonding process will be carried out with the full size ASIC and sensors when the ASICs become available in 2021. The final design review of the bump-bonding process is planned at the end of 2022.

As shown in Figure 4.3 the bare module is glued to a small flex cable PCB, called the module flex, on which the ASIC signals and the high voltage are wire bonded. The module flex is connected to a long flex cable tail through a mini-connector for the high voltage and a separate connector for the signal transmission to the PEBs.

Taking into account the space constraints, the flex tail is expected to be a two layer design with a maximum thickness of 220 µm. The longest readout row services 19 modules. As displayed in Figure 4.8, each flex transfers four types of signals:

- the data to be read out (time information or luminosity) on two differential-pair e-links per ASIC. The speed of the data transmission varies from 1.28 Gbit s^{-1} for the inner radius modules to 320 Mbit s^{-1}. For the luminosity, the speed is 640 Mbit s^{-1}.
- the fast commands from the lpGBT (clock, L0/L1 trigger, BCID and configuration parameters) and the slow control parameters through I^2C.
- the ASIC power supplies (1.2 V), setting a strong constraint on the flex plane resistance to minimize the voltage drop and the power dissipation ($< 300\,m\Omega$ for the longest flex). Digital and analog supply lines are separated.
- the bias voltage for the sensor (up to 800 V requiring excellent insulation).

The first flex cable prototypes, made of a single piece and longer than required, have been manufactured in two companies and at the CERN PCB workshop. Preliminary measurements satisfy the data transmission, bias voltage insulation and resistance requirements.

The R&D is still on-going on the design of both flex cables to ensure they satisfy the tight thickness constraint along Z and to identify/develop reliable mini-connectors for both the module and PEB connections. A few companies have been contacted for this specific R&D and the final design review should take place in 2022.

Tests of the glue used to attach the bare module to the module flex are ongoing in close collaboration with the ITk Pixel community, as the requirements are similar. In a first step, to exercise the module assembly, heaters that mimic real size modules will be mounted in summer 2020 and later with real modules (in 2021-2022). This activity will be done in the framework of the demonstrator activity (detailed in Chapter 14).

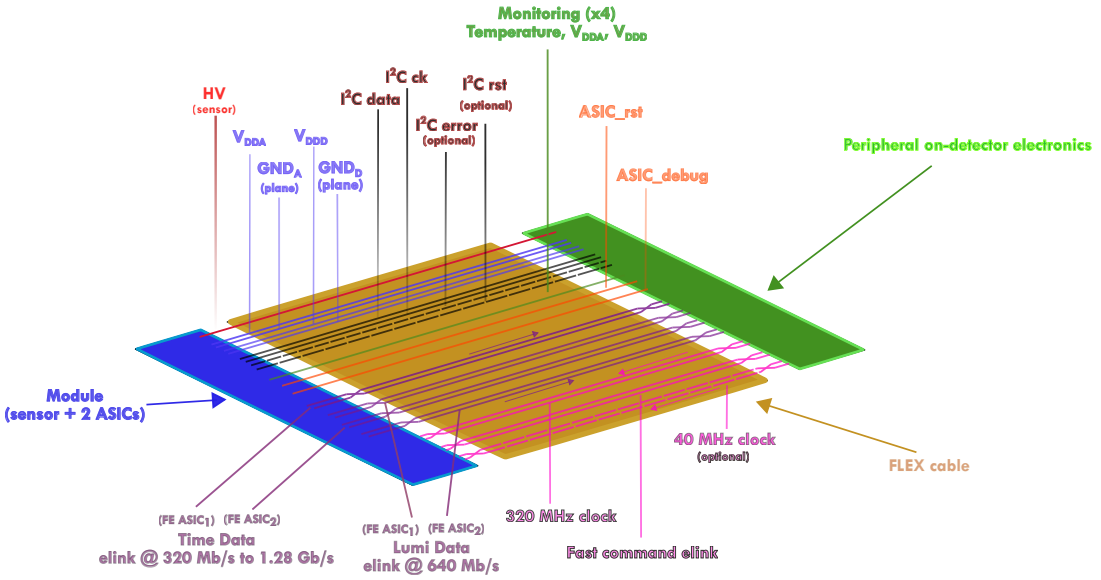


Figure 4.8: Signal transmitted from the ASICs to the peripheral electronics. Each ASIC has a dedicated e-link for luminosity and time data transmission while the other signals are common to both ASICs. The HV line is connected to the sensor.

4.4 Module loading on support structure

The modules are loaded on the cooling support plate using a support plate unit, made in carbon fibre, in which the modules are inserted in pre-defined windows and glued on each side on a small strip. This support unit, which ensures the exact position of each module and the alignment along the x and y readout row directions, as displayed in Figure 4.9, is screwed on the cooling plate. The modules have a step of 25.5 mm in a given row, corresponding to a 70% overlap between the top and bottom side modules of a layer for the inner ring. In the middle ring this step is 28.4 mm and the overlap 54%. In the outer ring the step is 34.5 mm and the overlap 20%. An independent support unit will be manufactured for each ring to allow for a fast replacement of the rings planned to take place at the surface in the long shutdowns. A thermal conductive grease is used to insure a good contact between the module and cooling plate. A simulation of the thermal behaviour of the system including the best knowledge of the thermal contacts of each material, and including the expected power dissipation of the sensor with radiation and temperature, has been done. The calculation shows that with the baseline cooling plates made of Aluminium, no thermal runaway is observed for the highest power dissipation over about a 25 °C range, guaranteeing safe operation under all conditions.

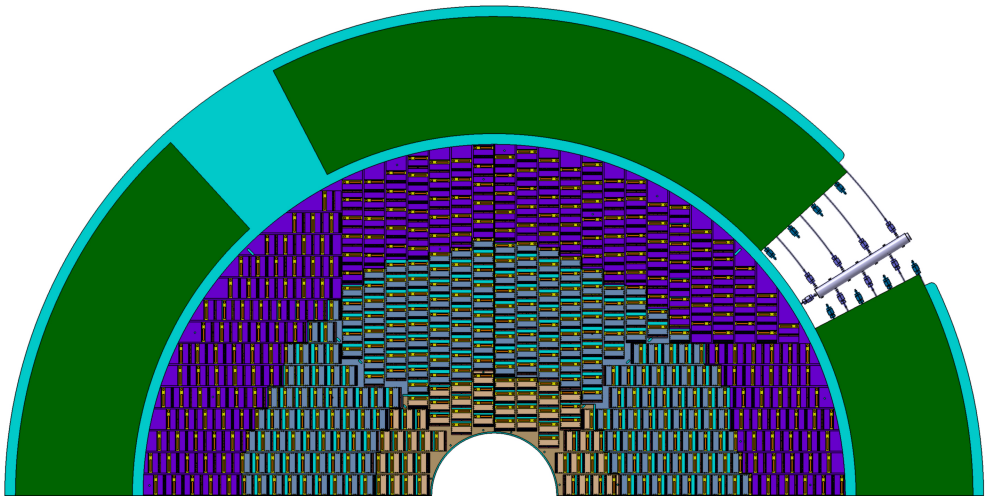


Figure 4.9: View of the modules inserted on the inner, middle and outer ring half disk support units. The modules are glued to the support units of each ring, and the support units screwed to the cooling half disk. The modules are aligned along x or y direction with a step of 25.5 mm in a given row, corresponding to a 70% overlap between the top and bottom side modules of a layer for the inner ring. In the middle ring this step is 28.4 mm and provides 54% modules overlap between top and bottom modules. In the outer ring the step is 34.5 mm and provides 20% modules overlap between top and bottom modules.

4.5 Off detector electronics, calibration and luminosity

Figure 4.10 shows the data path from the front-end ASIC to the off-detector backend. Different data and control signals from the flex cable are connected to the PEBs, where the electrical signals are encoded and transmitted via optical link to the off-detector electronics located in USA15.

The off-detector electronics consist of Front End LInk eXchange (FELIX) system and Data Handler and will be described in section Section 10.1.1. The general purpose FELIX receives event data from the on-detector electronics and transmits them to the Data Handler via multi gigabit network. In addition, FELIX interfaces to the TTC system via Local Trigger Interface (LTI) and to DCS for control, configuration and monitoring. Two different data paths are proposed: the main data stream that provides timing information per triggered event and the luminosity stream that provides bunch-by-bunch hit information. The main data stream is read out at the L0 trigger rate (about 1 MHz), while the luminosity stream is read out by dedicated FELIX boards.

4.5.1 Peripheral Electronics Boards

With the current design and taking into account the different read-out rows, five different PEB designs are needed for a layer quadrant but identical between quadrants. One PEB receives the data of up to 55 modules, encodes, aggregates and transmits them via optical links at 10.24 Gbit s^{-1} to the off detector electronics. In the down link direction, at 2.56 Gbit s^{-1}, this board transmits the trigger commands and clock to the ASIC. In addition this board distributes the DC voltage to all ASICs using DC-DC regulators and the High Voltage to the sensors. The board also handles voltage and temperature monitoring, and parameter setting in the ASICs for the detector control system. Taking into account the large numbers of signals with different properties and the high component density, the layout of this board is quite complex and a two-year development phase is still needed.

Most of the components to be used have already been developed by CERN for the LHC upgrades, namely the lpGBTs, the VTRx optical receiver and transmitter, and the bPOL12V converter. A dedicated analogue multiplexer (64 to 1) has been developed also in TSMC 130 nm to support digitization of monitoring signals to the ADC in the lpGBT. The first prototype of this ASIC has been received in December 2019. Due to the strong constraints on the envelope dimension (both in z and r), further development for the flex connection (flex tail integrated in PEB PCB) and high voltage connectors is on-going. A first functional prototype of the PEB is expected by end 2020.

4.5.2 Luminosity

Each ASIC can provide the number of hits for each bunch crossing for luminosity measurement. Due to bandwidth, space and cost limitation, these data are transmitted to specific lpGBT on the PEB at 640 Mbit s^{-1} only for the outer ring. These data are sent to dedicated

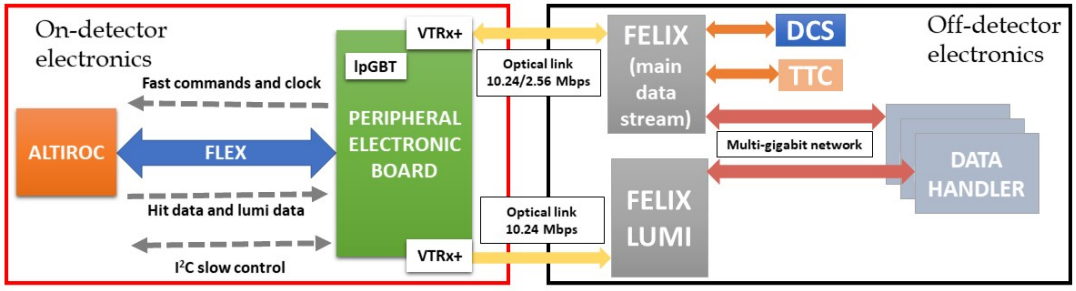


Figure 4.10: Data transmission paths for the main stream and the luminosity stream.

FELIX boards which sum the number of hits over a large enough region to provide an accurate online luminosity measurement. It is currently planned to divide the luminosity data in sixteen different regions, but the system should remain versatile enough to modify the size of these regions and remove any ASIC not working correctly if needed (for instance an ASIC with too noisy channels creating fake hits)

4.5.3 t_0 time calibration

The t_0 knowledge of each individual channel (about 3.6 million channels) is crucial to achieve the expected time resolution. The irreducible and non deterministic clock contribution to the resolution is expected to be around 15 ps, coming mainly from the lpGBT clock jitter and the additional contribution from the flex cable and ASIC. However this performance assumes that all channels are ideally timed with respect to the bunch crossing clock. The use of HGTD for physics relies strongly on the relative comparison of the time of different channels within an event. Consequently the geometrical (time static) inter-calibration of the t_0 of all channels is the most crucial while global time drifts over large regions will have smaller impact on the performance.

Geometric effects can be corrected with the signals provided by the pulser in the ASIC described in Section 6.3.4 (different flex cable length, systematic difference between channels in one ASIC due the imperfect clock tree distribution, etc.) or computed (geometrical time of flight). Regular sets of calibration runs between LHC fills will be used to monitor these calibration constants.

The shift of the 40 MHz clock phase per BCID, and therefore of the t_0, can be determined and corrected in-situ using data. Such low frequency clock phase variations can arise in the HGTD, for instance with temperature variations at module level from the CO_2 cooling, variations from one lpGBT to another (serving a few modules), or from the known day/night effect of the LHC clock, probably common to an entire HGTD end-cap. The calibration procedure will consist in measuring the inclusive average time of each channel with the data triggered at 1 MHz in the FELIX processor. Depending on the time period of these effects, and on the affected component and area (ASIC, module, group of ASICs of same lpGBT, PEB board), they may be calibrated with a good accuracy. For instance, at the ASIC level, a preliminary study shows that by computing the t_0 online, a 20 ps (50 ps) contribution can be reached at low (high) radius for periodic effects with a time period beyond 20 ms. The final calibration will need an additional offline calibration combining the information from many calibration windows.

4.6 Power distribution and detector control system

4.6.1 HV system

A schematic layout of the high voltage system is shown in Figure 8.1 and detailed in Chapter 8. Each of the 8032 modules should provide a bias voltage in a range from approximately 300 V, at the start of the HL-LHC, up to 750 V, after the detector has been exposed to the expected maximum irradiation levels of $2.5\text{x}10^{15}\text{neq/cm}^2$, as detailed in Chapter 5. The irradiation of each module will strongly depend on its radial position in the detector, seen in Figure 4.2, with an expected maximum variation smaller than 15% inside each module. The ultimate goal is to use individual adjustable voltages for each module, to allow for optimal operation. As a compromise between cost and performance, the baseline choice is nevertheless to initially share a supply between on average two sensor modules (having low leakage current at the beginning of Run-4) with an option to later supply individually each module. The multiplexing of a HV channel to modules can be done either in the USA15 services cavern, where the HV power supplies will be located or in the patch-panel region. All HV cables will be routed from the beginning to allow, at a later stage, each module to be supplied by a separate HV channel.

Consequently HV power supplies requirements are to deliver up to 800 V and up to 6 mA current in order to feed simultaneously two modules, keeping a bit of margin on the sensor leakage current. The power supplies will be based on commercial multi-channel rack mounted units located in the service cavern.

Monitoring the leakage current and the TOT as an indicator of the collected charge will give a good estimate of the sensor gain evolution during data taking (between fills), allowing to perform the necessary HV adjustments.

4.6.2 LV system

A schematic layout of the Low Voltage power system is shown in Figure 8.2 and detailed in Chapter 8. The LV power supply system needed by the front-end and peripheral electronics will deliver almost 20 kW and will be provided in a three stage system. Bulk power supplies, located in USA15, will provide 300 V DC voltages to DC-DC converters to be placed in the patch panel areas (PP-EC), located around the end-cap calorimeter outer radius surface, and accessible during technical stops and shutdowns. The second-stage multi-channel DC-DC units convert 300 V to 10 V that is distributed to the radiation hard DC-DC converters located on the PEBs. The last stage converts the power to the front-end electronics (ASICs) and the peripheral electronics boards providing mainly 1.2 V DC power but also 2.5 V for the optical receivers/transmitters . The converters of the peripheral boards are based on the bPOL12V, being developed by CERN for the HL-LHC upgrades.

4.6.3 Monitoring and Controls

A Detector Control System (DCS) will be implemented to control and monitor the various detector parameters: the power (HV, LV) supplied to the detector; the temperatures of the modules and of the peripheral electronics, the cooling system and the pressure of the N_2 used to keep a dry atmosphere inside the detector volume. A Finite State Machine (FSM) structure will be implemented and integrated in the ATLAS FSM tree during data taking, and will allow to operate in stand alone mode during commissioning and maintenance. It provides the tools to monitor the operational parameters of the detector, to bring the detector into any desired operational state, and to signal any abnormal behaviour by allowing for manual and automatic actions. More details are given in Section 10.4.

4.7 Mechanics, Services and Infrastructure

The detector mechanics and services were designed taking into account the severe constraints of space to accommodate the detector and the services that need to be routed in the gap between the barrel and end-cap calorimeters. The use of light structures was prioritized to minimize the amount of material in front of the active layers, and to minimize the potential increase of the radiation levels, leading to a weight per end-cap of approximately 350 kg.

The hermetic vessel provides a robust support structure to the detector disks in a cold and dry volume, with radial dimensions of 100 mm $< r <$ 1000 mm. It has four main components: the front and back covers, the inner ring and the outer ring (which will hold all the service feedthroughs), as illustrated in Figure 2.4. The front cover is divided in two half disks to allow its manipulation in the presence of the beam pipe. It consists of a honeycomb core placed between two thin carbon fibre reinforced panels to reduce deflection. The thickness of front and rear covers have been optimized to 13 mm and 7 mm respectively. To avoid condensation on the external face of the HGTD vessel during operation, heaters will be placed on the external face of the front and back covers, insuring a minimal temperature of at least 14 °C outside the HGTD vessel. An air gap of 3 mm will be kept between the HGTD detector and the end-cap LAr calorimeter as requested by ATLAS TC to avoid direct thermal contact with the cryostat front face.

Each double-sided layer (two per end-cap) is divided in two half circular disks of 30 kg each with 120 mm inner radius and 920 mm outer radius. This allows the detector installation to be completed later, in case of delays, even when the beam pipe is in place, provided that the back vessel cover and moderator are installed in LS3, when the beam pipe is not in place. The detector concept should facilitate rapid and safe removal of the detector to the surface while minimizing working time in the high radiation environment. This operation is envisaged at each long shutdown of the HL-LHC for the replacement of the innermost or middle rings. The rotation of the two disk layers inside the vessel by approximately 15 to 20°

with respect to each other, as seen in Figure 11.3, provides better integration of the cooling pipes inside the vessel while minimising the regions with zero hits resulting from the dead zones between the readout rows and imperfect coverage in the inner most radius.

The expected maximum power consumption of the detector, operating at −35 °C to reach the required performance, amounts to 40 kW in total (20 kW per end-cap); details of the various components are summarized in Table 11.2. An evaporative CO_2 cooling system of 50 kW will be used and part of its infrastructure and the cooling spare unit will be shared with ITk.

The evaluation of the number of services required to operate the detector, summarized in Table 12.1, and their respective routing design was subject to a careful evaluation and optimisation. This is due to the limited space in the vessel outer ring allocated to the services feedthroughs, in the barrel-end-cap calorimeter gap region and, last but not least, in the ATLAS flexible chains that allow to keep part of the services connected during the opening and closing of the end-cap calorimeters. The detector services routing on the end-cap calorimeter face is shown in Figure 4.11. The cables, exiting in four layers in the feedthroughs region, will merge into one layer at $r > 1.3$ m to fit within an envelope of 17 mm in z. At the outer radius of the end-cap calorimeter, services are routed in various layers in z but narrow slots in ϕ to pass in between the Tile fingers, a space also shared with ITk services. A dedicated slot in ϕ, on the top of the calorimeter, will be used to route four CO_2 cooling pipes with a maximum diameter of 50 mm each. The priority for services installation in flexible chains will be given to optical fibres, cooling pipes, interlock and cooling temperature sensor cables. The other services need to go through fixed cable trays and should be disconnected before the extended barrel calorimeters are moved for maintenance of the ATLAS detector. For that purpose patch panel boxes (PP-EC) will be organised on the end-cap Tile calorimeter outer surface in accessible places. The patch panel boxes will be also used for re-mapping the cables to match connectors on the detector.

4.8 Assembly, Installation and Commissioning

The final assembly of the detector and quality assurance, e.g. mounting the modules support frames and peripheral electronics boards into the half circular disks, connecting each flex cable to the respective peripheral electronics boards, and global certification, should take place at CERN from Q3 2024 to Q4 2026 with the participation of several collaborating Institutes. After the assembly, the detector will be transported to the pit. Each end-cap, HGTD A and HGTD C, will be lowered on side A and side C respectively, directly from the surface to the minivans. The final installation of the detector should take approximately 1 month per end-cap and it is planned for Q2 2026 (HGTD A) and Q1 2027 (HGTD C).

Dedicated tools are needed for assembly, lowering, and final installation of the detector. The designs for these tools are still at a conceptual stage and where possible will use synergies

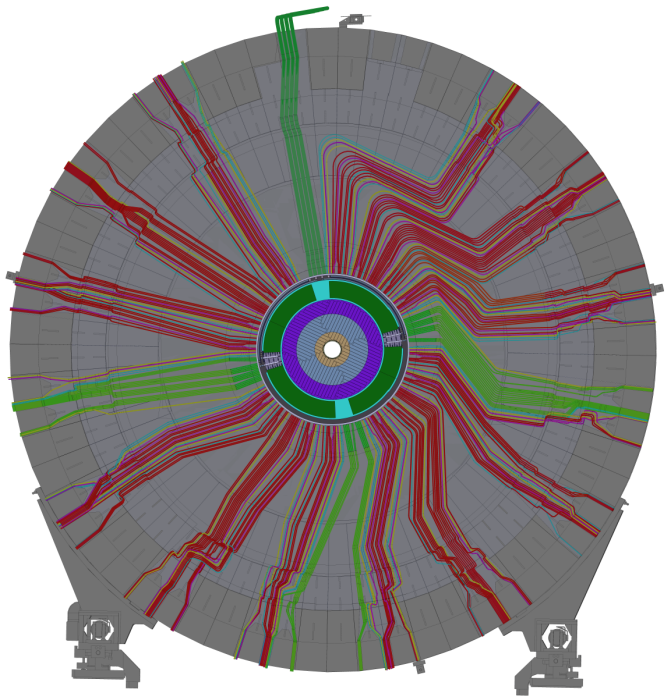


Figure 4.11: Side view of HGTD with services routed along the end-cap calorimeter face. The services design is already complete up to the the outer most radius of the end-cap calorimeter for the most difficult locations, where some Tile calorimeter fingers are blocked, not allowing an approximate radial orientation of the services. The four cooling pipes are indicated in dark green straight line. The different colours indicate different type of services (HV/LV cables, optical fibres, etc.)

with tools already developed for other sub-detectors. Extenders of cables and CO_2 cooling lines will be installed permanently to operate also the HGTD when ATLAS is in the open configuration.

The overall commissioning will start immediately after the connection of the services to the detector. Access to the detector components during the commissioning should be possible until approximately Spring 2027, close to the expected second end-cap calorimeter closure, giving about 6-months of intense commissioning for the HGTD C.

4.9 Next steps towards construction

The goal is to install the HGTD detector during LS3, in April 2026 and January 2027 for the A and C side, respectively. Three main schedule steps are planned:

- **2018-2021** R&D

- **2021-2026** Construction
- **2026-2027** Integration, installation and commissioning

The remaining key R&D steps needed to validate the HGTD design and performance are:

- Demonstrate the performance and radiation hardness of a full size ASIC. The first full size prototype will be available early 2021 (ALTIROC2), and most probably will necessitate a second iteration due to the complexity of the chip.
- Establish the performance of a full size HGTD module of 15x30 channels, that is to say a sensor bump bonded to two ASICs. With ALTIROC2, the hybridisation process will be tested, some modules will be assembled and a detector unit partially equipped during the demonstrator program in 2021.
- Conclude on the maximum fluence that the detector can sustain and consequently optimise the exact radial coverage of each of the 3 detector rings and rings replacement frequency. This tuning will depend on the outcome of the active R&D ongoing with new sensors by different companies, in particular with deep narrow doping implantation, C implantation and with real size sensors, to be delivered mid 2020.
- Validate the performance of CO2 cooling, including detailed thermal runaway studies with full detector size, including the integration and assembly of several modules in a readout row. The mechanical integration aspects will be validated with the heater demonstrator planned in 2020. The full demonstrator, planned for 2021 will include real size modules assembled in a realistic detector row, including flex cables, peripheral electronics and a FELIX Board to validate the entire readout chain up to the DAQ integration.

5 Sensors

5.1 Sensor parameters and requirements

The HGTD sensor parameters and requirements are summarized in Table 5.1. The sensors are intended to provide a fast signal in response to charged particles for a time resolution per hit of about 35 ps at the start and 70 ps at the end of lifetime (combined performance with the electronics and other contributions). The minimum charge collected for a MIP should be at least 4 fC and the hit efficiency at least 95%. The granularity should be 1.3 mm $\times$ 1.3 mm and the physical thickness 300 µm or less. The sensor should be of total active size of 39 mm $\times$ 19.5 mm with 30 $\times$ 15 pads and bump-bonded to two readout chips (ALTIROC) of 15 $\times$ 15 pads. The inactive edge around the sensor should be maximally 500 µm. The low-gain inter-pad gap should be maximally 100 µm, corresponding to a fill factor of at least 85%. In the baseline scenario, discussed in Chapter 4, the innermost part of the detector (r < 230 mm) should be replaced after each 1000 $fb^{-1}$and the middle ring within 470 mm $> r >$ 230 mm should be replaced at half lifetime (2000 fb^{-1}) of data-taking during the HL-LHC program. The sensors are then required to sustain a 1 MeV-neutron equivalent particle fluence of maximally 2.5×10^{15} n_{eq} cm^{-2} and a TID of 1.5 MGy, including a 1.5 safety factor.

The leakage current should be maximally 5 µA per pad, the applied bias voltage maximally 800 V[1] and the power density less than 100 mW/cm^2 at an operation temperature of maximally −30 °C on-sensor. The technology chosen for the HGTD sensors is Silicon Low Gain Avalanche Detectors (LGAD) with a baseline active thickness of 50 µm. The target gain[2] (charge) is 20 (10 fC) at the start and at least 8 (4 fC) at the end of lifetime.

[1] In fact, 50 µm thick sensors should be operated at maximally 750 V (see Section 5.5.2), but a margin is kept to allow for future developments of sensors with potentially different voltage requirements.

[2] the collected charge for a PiN of thickness 50 µm is around 0.5 fC

Technology	Silicon Low Gain Avalanche Detector (LGAD)
Time resolution	≈ 35 ps (start); ≈ 70 ps (end of lifetime)
Time resolution uniformity	No requirement
Min. gain	20 (start); 8 (end of lifetime)
Min. charge	4 fC
Min. hit efficiency	95%
Granularity	1.3 mm $\times$ 1.3 mm
Max. inter-pad gap	100 µm
Max. physical thickness	300 µm
Active thickness	50 µm
Active size	39 mm $\times$ 19.5 mm (30 $\times$ 15 pads)
Max. inactive edge	500 µm
Radiation tolerance	2.5×10^{15} n_{eq} cm^{-2}, 1.5 MGy
Max. operation temperature on-sensor	-30 °C
Max. leakage current per pad	5 µA
Max. bias voltage	800 V
Max. power density	100 mW/cm^2

Table 5.1: Sensor parameters and requirements.

5.2 Low Gain Avalanche Detectors

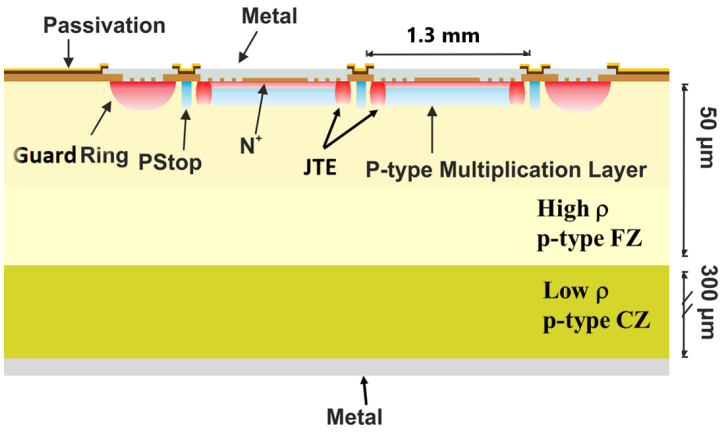
5.2.1 Overview

LGADs are segmented planar Silicon detectors with internal gain as illustrated in Figure 5.1. The gain depends on the doping dose of the multiplication layer as seen in Figure 5.2 and diminishes with radiation fluence as shown in Section 5.5.3. They have been pioneered by the Centro Nacional de Microelectrónica (CNM) Barcelona [5] and developed during the last 5 years within the CERN-RD50 community [4] including collaboration with two other LGAD vendors: Hamamatsu Photonics (HPK, Japan) and Fondazione Bruno Kessler (FBK, Italy). An introduction to the technology is given in Chapter 4. Additional background and details are given in Reference [61].

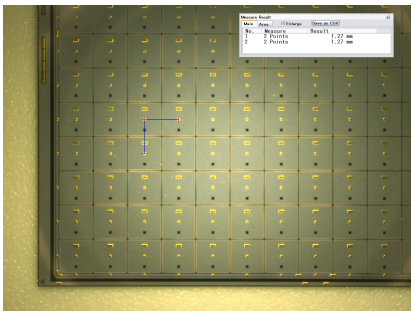
Three major effects determine the time resolution: time walk from amplitude variations, jitter from electronic noise and "Landau fluctuations" from charge deposition non-uniformities along the particle path. Time walk and noise jitter depend on the type of readout electronics chosen. Both depend inversely on the signal slope (voltage slope at the output of the amplifier) dV/dt:

$$\sigma_{\text{TimeWalk}} = \left[\frac{V_{\text{th}}}{\frac{S}{t_{\text{rise}}}}\right]_{\text{RMS}} \qquad \sigma_{\text{Jitter}} = \frac{N}{(dV/dt)} \simeq \frac{t_{\text{rise}}}{(S/N)} \tag{5.1}$$

where S refers to the signal which is proportional to the gain, N to the noise, t_{rise} to the



(a) Cross section of a 2 × 2 array.



(b) Photo of a 15 × 15 array.

Figure 5.1: (a) Cross section of a 2 × 2 array including a JTE around each sub-pad (SiSi wafer, CNM design) [62]. (b) Microscope photo of an HPK-3.1 15 × 15 array.

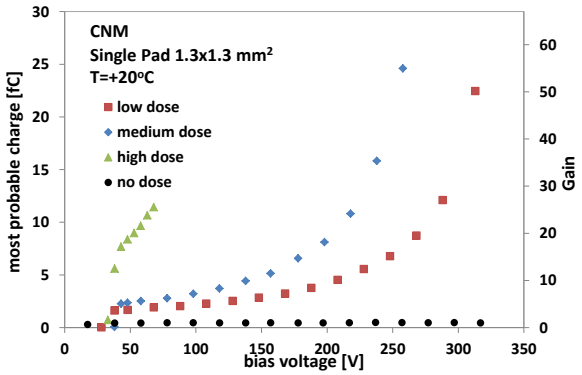


Figure 5.2: Gain and charge as a function of bias voltage for CNM LGADs with different doping concentration (in the legend "dose") of the multiplication layer. The "no dose" points are corresponding to a sensor without moltiplication layer (classic PiN).

rise time and V_{th} to the threshold voltage. It can be seen that the lowest noise jitter and time walk are achieved with sensors with high signal-to-noise ratio (S/N) and small rise time, i.e. with thin sensors and large gain. Time walk can usually be corrected using time reconstruction algorithms such as constant-fraction discrimination (CFD) or amplitude or time-over-threshold (ToT) corrections. The third effect, referred to as "Landau fluctuation" is due to the non-uniform charge deposition along the particle path leading to time-of-arrival fluctuations. It is a contribution depending on the thickness of the sensor (thin is beneficial) and the setting of the threshold. Adding the three contributions in quadrature yields the overall time resolution. After time walk correction, the noise jitter is the dominating contribution for low S/N and the Landau term takes over for high S/N.

An example for a measured LGAD time resolution is shown in Figure 4.5 as a function

of collected charge, with the time walk corrected using CFD. As expected from Eq. (5.1) the resolution improves with increasing gain (proportional to collected charge) due to the reduced noise jitter. Eventually it levels off to the Landau fluctuation of about 35 ps for 50 µm thickness.

This observation feeds into the plan to operate LGADs at a gain of about 20 before irradiation and as close as possible to that value after irradiation given restrictions from the leakage current, the breakdown voltage, and the noise, including the excess noise from the multiplication process. The gain target of 20 was chosen since the time resolution fulfils the HGTD requirement of 35 ps per hit at the start of operation (see Table 2.1) and is improving only slowly when going to higher gains as seen in Figure 4.5. Moreover, the maximum achievable gain reduces after irradiation, hence an optimisation of the detector to higher gains before irradiation would only benefit a short period at the start of operation. At high fluences, operation at charges down to 4 fC corresponding to a gain of 8 becomes necessary (see Section 5.5.3).

The field in the Silicon bulk (i.e. no-gain) region should be high enough to saturate the drift velocity of about 100 µm ns^{-1} for a reduced rise time (the saturation field is 2×10^4 V cm^{-1}, but the charge collection time starts to saturate at 1×10^4 V cm^{-1}).

An LGAD active thickness of 50 µm has been adopted as the best compromise between capacitance and deposited charge (favouring a large thickness) and signal slope and Landau fluctuations (favouring a small thickness). LGADs of 30 µm active thickness have been studied as an option in the past and showed a better sensor-only performance before irradiation, but were discarded due to the higher capacitance and higher power dissipation at similar performance after irradiation compared to 50 µm. Such small active thicknesses are usually achieved by different techniques that all use a thin active high resistivity layer on top of a thicker insensitive Silicon substrate of low resistivity, such as Silicon-on-Insulator (SOI), Silicon-Silicon Wafer Bonding (SiSi) or epitaxial (Epi) wafer techniques.

Figure 5.1(a) shows the cross section of a 2×2 LGAD array. Each pad consists of the p-type multiplication layer underneath the n^+ implantation, surrounded by a Junction Termination Extension (JTE). The JTE is an n^+ implantation that is deeper than the one of the central pad. It controls the electric field at the edges to avoid early breakdown, but also leads to an inter-pad gap with no or reduced gain and hence worse time resolution and hit efficiency in this region. The complete sensor is surrounded by a guard ring (GR). Figure 5.1(b) shows a photo of an HPK 15×15 array.

As a dopant for the p-type multiplication layer, Boron (B) is typically used. Additional Carbon (C) implantation is investigated as candidate technology for improved radiation hardness. The substitution of B by Gallium (Ga) has been studied as well, but so far has not demonstrated clear beneficial results, hence it is not considered as a candidate for production at the moment.

5.2.2 LGAD productions

At present, LGADs have been produced in six manufacturing sites, shown in Table 5.2 along with their production capabilities: HPK, Japan; CNM, Spain; FBK, Italy; Micron, UK; Brookhaven National Lab (BNL), USA; and Novel Device Laboratory (NDL), China. Further vendors are interested in LGAD productions.

There are plans to use LGADs in three experiments at the HL-LHC (ATLAS, CMS, LHCb). There has been fruitful collaboration and coordination between ATLAS-HGTD and CMS-ETL [63] with respect to simulations, design, manufacturing and testing.

The design and production of LGADs for HGTD had two distinct phases: an early R&D phase of about 6 years with much of the activities carried out within the RD50 collaboration where the basic parameters were investigated and the suitability of LGADs for large scale application has been determined. The different manufacturers tended to concentrate on different parameters (like multiplication layer doping profile and dose, variation of the types of dopant, thickness). In general, the LGAD sensors produced by different manufacturers appear to perform similarly, with the exception of the leakage current before irradiation, and the bias voltage reach after irradiation.

In the second phase, in which the collaboration has entered, the focus is geared towards the production of sensors for the specific HGTD application, now that the sensor requirements were fixed, and thus the options are reduced. For example, the decision to fix early on the pitch of the pads in the detector arrays to 1.3 mm provided a needed stable basis so that the development of other parts of the detector (electronics, modules, mechanical layout) could proceed. At this point, the need to investigate issues of manufacturing (yield, uniformity, large arrays, fill-factor, under-bump-metalization (UBM[3]), etc.) and operations (bias voltage, power, reliability, breakdown) have become more important.

Manu-facturer	Wafer Size [inch]	Thick-ness [µm]	C Implant	Array 5×5	Array 15×15	Array 30×15	UBM
CNM	4-6	30 - 300	x	x	(x)	(x)	
FBK	6	(50) 60 - 300	x	x			
HPK	6	20 - 80		x	x	(x)	x
BNL	4	50					
Micron	4	100 - 300					
NDL	6	33 (50)		x	x		

Table 5.2: LGAD manufacturers and production capabilities achieved to-date. Values in brackets (...) are for ongoing runs.

The results in the following have been mainly obtained from the LGAD types shown in Table 5.3. For HPK-3.2 the full depletion voltage and the V_{BD} at −30 °C are very close, this aspect will be optimized in the next prototypes runs as explained in Section 5.8. These runs

[3] UBM is part of the hybridisation process as explained in Section 7.2.1.

include LGAD sensors of HGTD geometry. Many more runs not mentioned here have been studied in addition for R&D purposes. Typically in a run there are sensors of varied nominal inter-pad gaps (IP) or slim edges (SE). For NDL, a run of 33 µm thickness is shown as a prototype, a 50 µm run is ongoing.

Manu-facturer	Name	Thickness [µm]	Gain layer dopant	C implant	Gain layer depth [µm]	Gain layer depletion [V]
HPK	HPK-3.1	50	Boron	No	1.6	40
HPK	HPK-3.2	50	Boron	No	2.2	55
FBK	FBK-UFSD3-C	60	Boron	Yes	0.6	20
CNM	CNM-AIDA1/2	50	Boron	No	1.0	45
NDL	NDL-33µm	33	Boron	No	1.0	20
Manu-facturer	Name	Full depletion [V]	V_{BD} −30 °C [V]	Nominal IP [µm]	Nominal SE [µm]	Max. Array Size
HPK	HPK-3.1	50	200	30-95	200-500	15×15
HPK	HPK-3.2	65	70	30-95	200-500	15×15
FBK	FBK-UFSD3-C	25	170	37	200-500	5×5
CNM	CNM-AIDA1/2	50	220/50	37-57	200-500	5×5
NDL	NDL-33µm	35	70	55	450	15×15

Table 5.3: Design, geometrical and electrical properties of LGAD types.

5.3 Radiation damage and irradiations

As explained in Section 2.4, the detector has to withstand a total 1 MeV neutron equivalent particle fluence of maximally $2.5 \times 10^{15}\,\mathrm{n_{eq}\,cm^{-2}}$, assuming that the innermost part of the detector (r < 230 mm) should be replaced after each 1000fb^{-1}and the middle ring within 230 mm $< r <$ 470 mm should be replaced at 2000fb^{-1}. It should be noted that the total fluence is a combination of both charged and neutral hadrons with different contributions in different regions. In the innermost region, the radiation field is roughly equal for neutrons and charged hadrons (Figure 2.15), but the contribution by charged hadrons decreases steeply with radius, so that the field is dominated by neutrons in the outer regions due to backscatter from the calorimeters. The maximum fluence from charged particles is only around $1 \times 10^{15}\,\mathrm{n_{eq}\,cm^{-2}}$, while for neutrons it is $2 \times 10^{15}\,\mathrm{n_{eq}\,cm^{-2}}$. The energy spectrum of protons and pions has a fairly flat maximum between 50 MeV and 10 GeV, whereas the neutron spectrum peaks at about 1 MeV, but has large contributions over a large range from 0.1 eV to 100 MeV (see Appendix A).

Radiation damage in Silicon mainly results in the change of the effective doping concentration, the introduction of trapping centers that reduce the mean free path of the charge carrier, and the increase of the leakage current [4]. However for thin sensors the effect of trapping is reduced due to the smaller electrode distances. For LGADs, one of the main effects is the degradation of gain with fluence at a fixed voltage due to removal of initial acceptors in the

multiplication layer [64, 65], which implies the need to increase the applied bias voltage after irradiation to at least partly compensate for this.

To study the LGAD performance after irradiation, sensors have been irradiated up to fluences of 6×10^{15} n_{eq} cm^{-2} at various facilities with different particle types and energies that are representative for the ones expected in HGTD. However only results up to 3×10^{15} n_{eq} cm^{-2} are shown in the following sections. Table 5.4 gives an overview on the facilities, their parameters and maximum fluences as well as Total Ionising Dose (TID) achieved for different LGAD types irradiated. The hardness factor used throughout this document is used for the conversion of the actual particle fluence to the 1 MeV-neutron equivalent fluences. The irradiation campaign was mainly supported by JSI neutrons. Since the quoted fluence at JSI has an uncertainty of roughly 10%, several sensors were irradiated at least for the higher fluences. Then all sensors at the same fluence were tested and the results shown in the following sections are for representative sensors.

First prototypes were irradiated in all facilities except for CYRIC, and it was found that acceptor removal seems to be faster with respect to 1 MeV neutron equivalent fluence after irradiations with 200 MeV–23 GeV charged hadrons than with neutrons [64, 65]. However, CERN PS is in shutdown now until 2021 and access to other high energy hadron irradiation facilities like Los Alamos is limited. Hence results for LGADs with the HGTD geometry presented here are mostly after irradiations with neutrons at Ljubljana and 70 MeV protons at CYRIC. Sensors irradiated at CYRIC also show higher acceptor removal rate than neutron irradiated sensors at the same fluence. These studies will be followed up by irradiations with higher energy charged hadrons at Los Alamos when it becomes available before the end of 2020. Mixed neutron-proton irradiations for a realistic estimation of the performance with the expected final particle composition are ongoing, to do so sensors proton irradiated at CYRIC will be irradiated again with neutrons at JSI.

It should be noted that irradiations at CYRIC with 70 MeV protons led to a maximum TID of 4.0 MGy, i.e. more than the HGTD requirement of 1.5 MGy. To study in more detail the effect of TID such as changes in the surface conditions, presently there are irradiations with X-rays under way at IHEP.

The measurements with irradiated sensors were done after annealing for 80 min at 60 °C, if not noted otherwise. Dedicated annealing studies are presented in Section 5.5.7.

Facility & Abbreviation	Particle Type	Hardness Factor	TID [MGy] / 10^{15} n_{eq} cm^{-2}	Max. Fluence [10^{15} n_{eq} cm^{-2}]	Max. TID [MGy]	LGAD Types Irradiated
JSI Ljubljana (*n*)	≈1 MeV n	0.9	0.01	6	0.06	all
CYRIC (*pCY*)	70 MeV p	1.5	0.81	2.5	4.0	HPK-3.1/3.2, NDL FBK-UFSD3-C
Los Alamos (*pLA*)	800 MeV p	0.7	0.43	6	0.4	early prototypes
CERN PS (*pPS*)	23 GeV p	0.6	0.44	6	2.7	early prototypes

Table 5.4: Irradiation facilities and parameters and maximum achieved fluence and TID, as well as LGAD types irradiated.

5.4 Sensor tests: methodology and experimental techniques

The LGAD sensors have been tested before and after irradiation by various HGTD groups, as well as within the RD50 community.

Electrical measurements including capacitance-voltage (C-V) and current-voltage (I-V) characteristics have been performed on laboratory probe stations. For the probing of large arrays, custom-made probe cards for the simultaneous contact of 5×5 pads have been developed. For the measurement of larger arrays like the 15×15 single-chip sensor, the probe card is applied sequentially to 5×5 sub-blocks. A probe card with 15×15 contacts is under development. An alternative is the sequential probing of one single pad after another on a semi-automatic probe station that allows to scan over an arbitrary number of pads in an array, while the neighbouring pads and the guard ring are floating.

The dynamic properties of LGADs, such as charge collection, gain and time resolutions, have been measured in response to ionising particles, both in the laboratory with ^{90}Sr β particles [61, 64–70] and lasers, as well as in beam tests [10, 66, 67]. Beam tests have been performed by the HGTD community in more than fifteen periods between 2016 and 2020 at the H6 beam line of the CERN SPS [10] with 40 to 120 GeV pions, at SLAC with 15 GeV electrons, at FermiLab with 120 GeV protons, and at DESY with 5 GeV electrons [71]. Data were taken in two modes: stand-alone and integrated into a beam telescope that provided track position information with about 3 µm precision [72].

Most of the measurements on irradiated sensors were performed at the HGTD target on-sensor temperature of −30 °C.

The dynamic measurements in the laboratory and beam tests were all obtained using custom-made HGTD-specific readout boards with an integrated high bandwidth amplifier with a gain of about 10, followed by a second commercial 2 GHz amplifier of gain 10, allowing the recording of the pulse shape of the fast LGAD signals [10, 66] with a high bandwidth oscilloscope (1–2.5 GHz). The noise was measured as the RMS fluctuation of the base line of the oscilloscope trace. It typically amounts to 1.6 mV–2.5 mV (roughly corresponding to a charge of 0.12 fC–0.20 fC) depending on the type and vertical scale of the oscilloscope, the board type, and the physical location. Measurements at test beam facilities tend to be noisier than laboratory measurements since machinery and magnets are operated in the same areas. The performance of the sensors was evaluated with discrete electronics optimized for precision timing, large scale measurements with the ALTIROC as readout were not executed until now since the chip has not yet been available for large-scale sensor testing. However first measurements of the combined sensor-ALTIROC performance on few bump-bonded hybrid prototypes are presented in Section 6.7.2 showing a time resolution under 40 ps. The measurements presented here will be repeated with the ALTIROC as soon as enough chips are available.

Position-sensitive scans using red and infrared lasers to deposit charge carriers inside the sensors have been made at various Institutes, using the Transient Current Technique (TCT) setup.

The gain is extracted by dividing the collected charge in an LGAD device by the charge of no-gain PIN diodes of the same thickness without multiplication layer (for MIPs the signal is about 3 ke^- or 0.5 fC for 50 µm thickness).

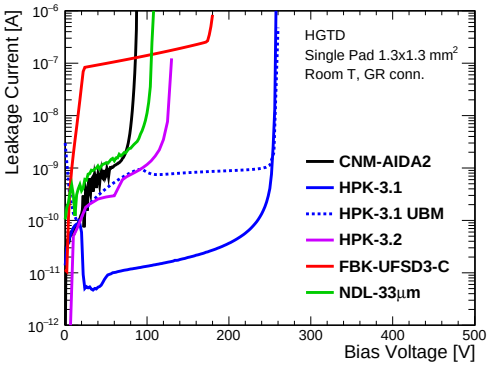
Time resolutions are typically extracted from the spread of the time-of-arrival difference between two sensors when a particle passes through both. For the measurements, either at least two LGADs are used, or an LGAD, a fast Cherenkov counter (based on quartz bars) and a Silicon photo multiplier (SiPM) (with typical time resolutions of about 10 to 40 ps) are used. If at least three devices are measured simultaneously, a χ^2 minimisation is used to obtain the time resolution of all devices. In case only one device under test (DUT) is measured with respect to one reference device of known resolution, the DUT resolution is obtained by subtracting quadratically the reference contribution. Time walk effects are usually corrected for using time reconstruction algorithms such as the CFD, the Zero-Crossing Discriminator (ZCD) or corrections using the amplitude or TOT of the signal [10].

LGAD behavior such as time resolution and collected charge was simulated using the software WeightField 2 (WF2) [73]. The WF2 simulations were tuned using laboratory measurements from different sensor types. Also, the software TCAD sentaurus [74] was used to optimize the design of the sensors for production.

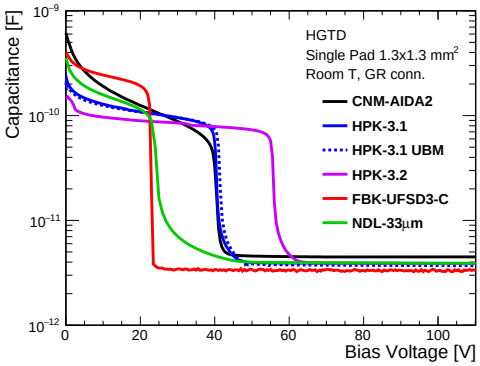
5.5 LGAD performance before and after irradiation

5.5.1 Electrical characterisation: I-V and C-V

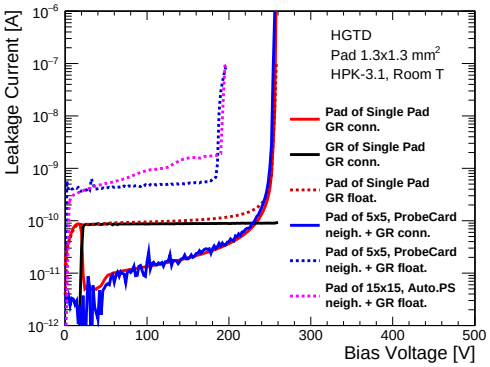
Figure 5.3(a) and Figure 5.3(b) show the I-V and C-V curves of 1.3 mm × 1.3 mm LGAD pads of different vendors and runs, measured with the guard ring (GR) connected to ground. Un-irradiated LGADs from most of the vendors and runs achieve nA leakage current levels or below before breakdown, well below the ALTIROC leakage current limit of 5 µA per pad. The addition of the UBM process at HPK in this prototype run led to an increased leakage current by 2 orders of magnitude with respect to wafers without UBM. The current reaches about 1 nA, which is still safe for operation and expected to improve in future productions. No influence on the C-V behavior was found. The FBK-UFSD3-C sensors with Carbon exhibits currents of about 100 nA, which are higher than HPK Boron-only sensors but are still safely below the ALTIROC limit. After irradiation, the currents of FBK-UFSD3-C become more similar to the other types. The breakdown voltage increases with decreasing multiplication layer dose.



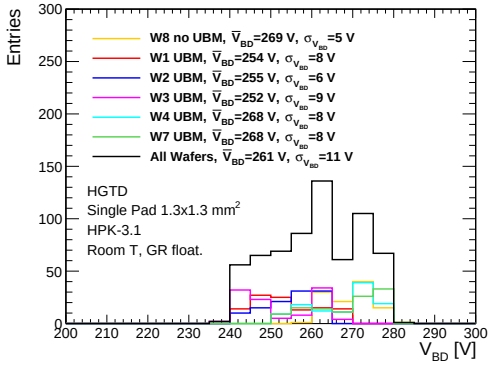
(a) I-V for different runs.



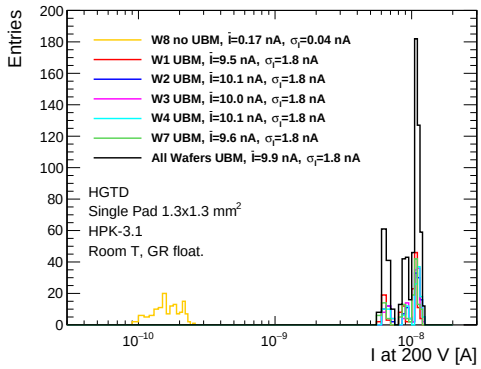
(b) C-V for different runs.



(c) I-V for different types and conditions.

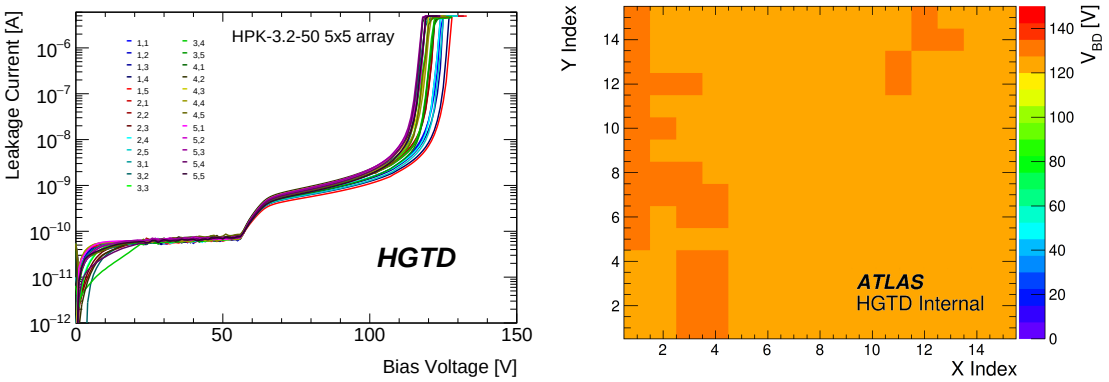


(d) V_{BD} distribution.



(e) Current at 200 V distribution.

Figure 5.3: Measurements of current-voltage (I-V) (a) and capacitance-voltage (C-V) (b) characteristics comparing different vendors and runs, as well as device types and biasing conditions (c). (d) and (e) show the distributions of V_{BD} and the current at 200 V for single pads of different wafers of HPK type 3.1 (with and without UBM).



(a) All I-V curves of a 5×5 array.

(b) V_{BD} map of a 15×15 array.

Figure 5.4: (a) I-V measurement of 25 pads from an unirradiated HPK-3.2 5×5 array without UBM measured with a 5×5 probe card at room temperature (all pads and GR grounded). (b) V_{BD} map of a 15×15 HPK-3.2 array without UBM measured with an automatic probe station at room temperature (neighbors and GR floating) [75].

Also the range of the "foot" of the C-V curve (i.e. the voltage region where C stays at high values while the multiplication layer is being depleted, starting from the n-p junction at the front) is an indicator of the multiplication layer dose. Foot values between 20 V and 60 V indicate substantial gains, as verified below. The depletion of the bulk (indicated by the sharp fall of the C-V curve) happens rather fast within a few V due to the high resistivity and the small thickness. The end capacitances of about 3 pF–4 pF for 1.3 mm $\times$ 1.3 mm LGAD pads (measured with a connected GR) are consistent with active thicknesses of 40 µm–60 µm.

Figure 5.3(c) shows the I-V curves for HPK-3.1 sensors of the LGAD pad and GR with either GR connected to ground (as the pad) or floating. For the single pad sensor, it can be seen that the current through the pad in case of floating GR is roughly the sum of pad and GR current in case the GR is connected. However, the breakdown voltage (V_{BD}) where the current increases rapidly, is found not to be affected by the GR biasing condition for single pads. For a pad in an HPK-3.2 array, the I-V curve is found to be almost identical to the one of a single pad in case the neighbors and the GR are connected to the same potential, as measured with a 5×5 probe card on a 5×5 array (see Figure 5.3(c) and Figure 5.4(a)). However when leaving neighboring pads and GR floating, the current level is increased by 2 orders of magnitude (presumably due to punch-through to the neighbors) and V_{BD} is observed to be reduced from about 250 V to about 190 V. The reduction of breakdown was consistently measured with a probe card when connecting only one channel and an automatic probe station with only one needle (see Figure 5.3(c)). It should be noted that this behavior of shifting V_{BD} in case of floating neighbors and GR was not observed for the 5×5 arrays of the CNM-AIDA run. This indicates that it depends on the sensor design and the exact production process.

Probing with an automatic probe station turned out to be a powerful tool to identify individual faulty pads inside an array [75] and is so far the only method to probe 15×15 arrays efficiently until the development of a 15×15 probe card is finished. In most cases all pads inside an array behave uniformly for HPK-3.1 5×5 and 15×15 arrays (see Figure 5.4). The mean V_{BD} spread between pads in an array is found to be typically a few V.

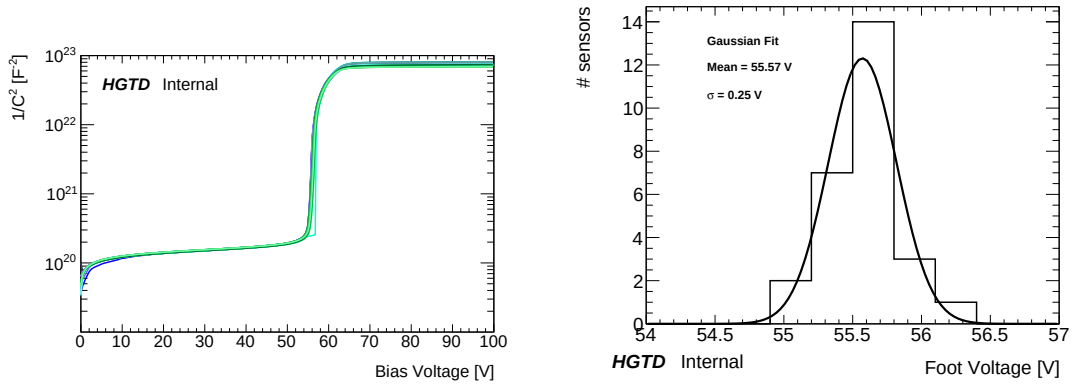
The scenario of only a single floating pad in the center of a 5x5 array with the other 24 and the GR connected was also studied with a 5x5 probe card. This was to simulate the behavior of a faulty pad that needs to be disconnected to make the sensor operable. The single floating pad had an influence on the breakdown voltage of all other pads in the array by introducing a shift to V_{BD} lowering it by less than 10 V and producing a more sudden and steeper breakdown.

LGAD Type	Sensor Type	Nominal Edge [µm]	Nominal IP gap [µm]	Sensors tested	Pads tested	Fraction of Perfect Sensors [%]	Fraction of Good Pads [%]
HPK-3.1	Single	Sum all	95	648	648	100	100
		500	95	360	360	100	100
		300	95	144	144	100	100
		200	95	144	144	100	100
	2×2	Sum all	Sum all	13	52	100	100
		500	30	1	4	100	100
		300–500	50	2	8	100	100
		300–500	70	2	8	100	100
		200–500	95	8	32	100	100
	5×5	500	95	19	475	100	100
	15×15	500	95	27	6075	85.2	99.5
HPK-3.2	Single	Sum all	Sum all	216	216	100	100
		500	95	120	120	100	100
		300	95	48	48	100	100
		200	95	48	48	100	100
	2×2	Sum all	Sum all	26	104	100	100
		500	30	2	8	100	100
		300–500	50	4	16	100	100
		300–500	70	4	16	100	100
		200–500	95	16	64	100	100
	5×5	500	95	6	150	100	100
	15×15	500	95	23	5175	91.3	99.8
CNM-AIDA1	Single	500	37	84	84	69	69
		500	47	39	39	95	95
		500	57	42	42	100	100
	5×5	500	37	6	150	50	66
		500	47	6	150	83	90
		500	57	6	150	100	100

Table 5.5: Number of tested devices and fraction of good pads and sensors for HPK-3.1/3.2 and CNM-AIDA1 of different sensor types, edge and inter-pad gap designs. An array of 15×15 pads corresponds to the final ALTIROC size and half of the full final sensor area.

HGTD Institutes measured a large number of single pads and arrays from different productions, in particular HPK-3.1/3.2 and CNM-AIDA1. HPK also provides their in-house Quality-Control (QC) results with an automatic probe station (GR floating) of each single pad they delivered. The HPK results have been verified by HGTD Institutes. Figure 5.3(d) and Figure 5.3(e) show the corresponding distributions of V_{BD} and the current at 200 V for all HPK-3.1 single pads on different wafers, with and without UBM, demonstrating a good uniformity. The mean of V_{BD} for all wafers is 261 V with a spread of 11 V. The per-wafer spread varies between 5 V and 9 V. No single pad sensor has a V_{BD} of less than 235 V or more than 285 V. For the current at 200 V, two distinct distributions are found as expected from the results discussed above: one for sensors without UBM with a mean of 0.17 nA, and one after applying UBM with a mean of about 10 nA (it should be noted again that the GR was floating), the spread is found to be about 20%.

However, in terms of performance, sensors seems to be consistent to less than the percent level. In Figure 5.5 the spread of the C-V measurements for several HPK-3.2 sensors is shown, measurements were taken in several HGTD Institutes. The foot, which is directly connected to the doping concentration of the multiplication layer and the sensor gain, shows a variation which is less than one percent.



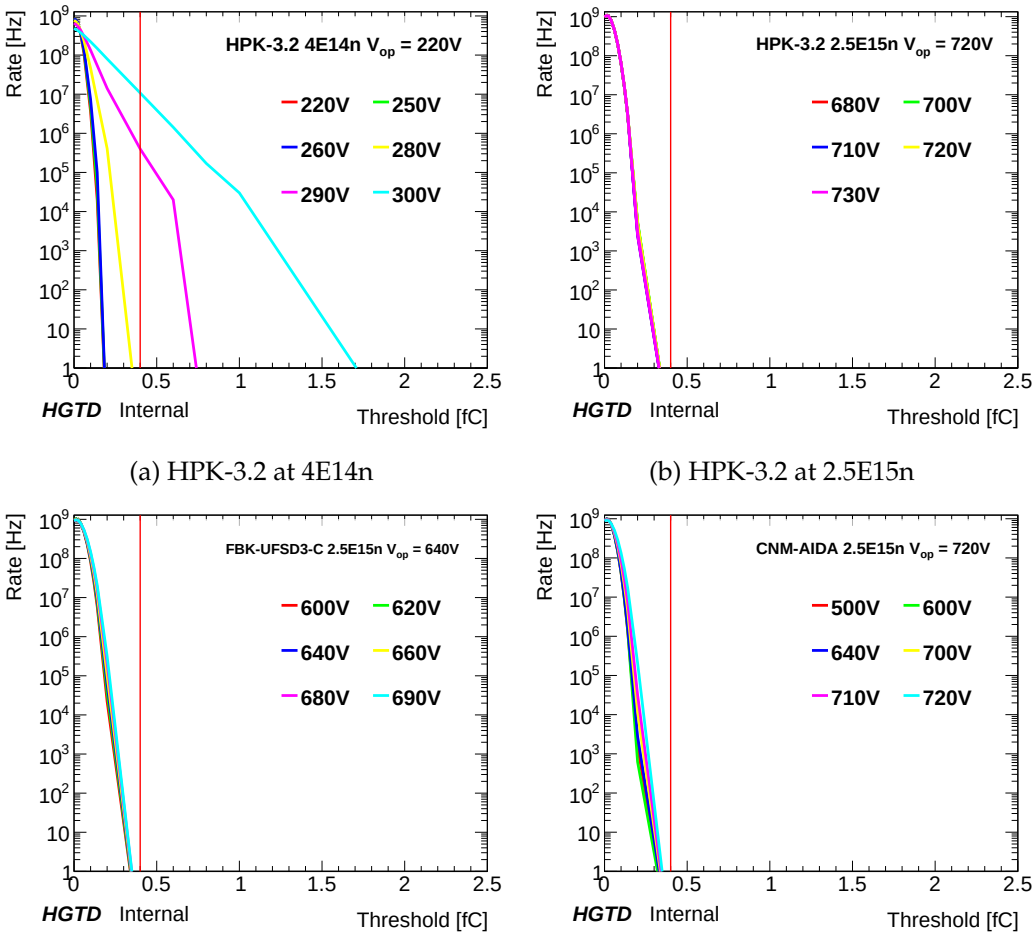
(a) (b)

Figure 5.5: (a) $1/C^2$ curve measurement for 27 HPK-3.2 single pad sensors from different wafers before irradiation. (b) Extracted foot of the 27 HPK-3.2 single pad sensors, showing that the variation of the gain layer doping is half a percent.

Table 5.5 shows the fraction of good individual pads out of all single pads and arrays, defined as having a breakdown voltage above 90% of the expected one for the respective biasing condition of GR and neighbors. Moreover, the fraction of perfect sensors is displayed, which are defined by requiring all pads in a sensor to be good. For HPK, the fraction of good pads turned out to be 99.5–100%. No dependence on the edge design between 200 µm and 500 µm edge was found. The fraction of perfect perfect sensors is 100% for all HPK single pads, 2x2 and 5x5 arrays and 85.2% (91.3%) for HPK-3.1 (HPK-3.2) 15x15 arrays. For CNM-AIDA1 the

result was found to depend on the inter-pad gap parameter (IP): the largest inter-pad gap (IP57) is found to give 100% good sensors and pads, which reduces to about 70% good pads and 50% perfect sensors for the smallest inter-pad gap (IP37).

5.5.2 Operating bias Voltage and self-triggering



(a) HPK-3.2 at 4E14n

(b) HPK-3.2 at 2.5E15n

(c) FBK-UFSD3-C at 2.5E15n

(d) CNM-AIDA at 2.5E15n

Figure 5.6: Self-trigger rate as a function of threshold in collected charge for several bias voltages. (a): for HPK-3.2 sensor at $4 \times 10^{14}\,\mathrm{n_{eq}\ cm^{-2}}$ of neutron irradiation. (b): same sensor after $2.5 \times 10^{15}\,\mathrm{n_{eq}\ cm^{-2}}$ of neutron irradiation. (c): for FBK-UFSD3-C sensor at $2.5 \times 10^{15}\,\mathrm{n_{eq}\ cm^{-2}}$ of neutron irradiation. (d): for CNM-AIDA sensor at $2.5 \times 10^{15}\,\mathrm{n_{eq}\ cm^{-2}}$ of neutron irradiation. A threshold of 5 mV corresponds to roughly 0.4 fC of collected charge, represented by the vertical red line in the plots. The operating voltage (V_{op}) for each is written in the legend, as shown no self-triggering is present at that V_{op}.

As mentioned in Section 5.4, dynamic measurements in response to particles have been

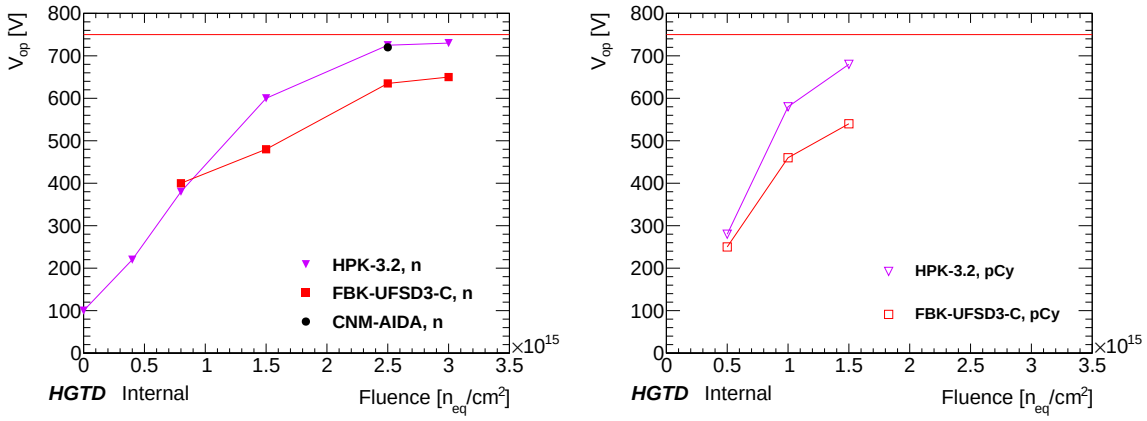
performed in the laboratory and beam tests on custom-made HGTD-specific readout boards. The maximum applicable bias voltage plays a crucial role in determining the performance of the sensors before and after irradiation, since the gain depends on the bias voltage, and this dependence changes with irradiation. It is important to realize that for thin sensors the effect of trapping is reduced due to the smaller electrode distances therefore the charge collected from the bulk before charge multiplication does not change much even after irradiation to $1 \times 10^{16}\,\mathrm{n_{eq}\,cm^{-2}}$.

The operating voltage (V_{op}) is defined as a stable and safe operation voltage at which the sensor has reasonable performance in terms of time resolution and gain. To evaluate it, several aspects are taken into account. At this voltage the sensor can be operated for a prolonged period of time and under a constant flux of particles (close to the LHC repetition rate of 40 MHz) without the risk of inducing breakdown or electrical arcing between the sensor structures (see Section 5.5.7). The noise increase should be less than 20% when compared to lower voltages, plus the signal to noise ratio must be higher with respect to all lower voltages. The maximum leakage current allowed is limited to 5 µA per pad and the power less than $100\,\mathrm{mW/cm^2}$. Furthermore at this voltage the sensor must not present self-triggering events (events caused by discharges unrelated to particle hitting the detector) with a rate higher than 1 kHz for a trigger threshold of ±5 mV or a collected charge of 0.4 fC. An excessive self-triggering would increase the dead time of the HGTD detector hindering its operation. This was studied in detail for HPK-3.2, CNM and FBK sensors (studies for other types are ongoing). The self-trigger rate increases dramatically if the sensor is operated near the breakdown with gain higher than 30. This statement is valid both for unirradiated and irradiated (with neutrons/protons) sensors of HPK-3.2, CNM and FBK as shown in Figure 5.6. For HPK-3.2 at a neutron fluence of $4 \times 10^{14}\,\mathrm{n_{eq}\,cm^{-2}}$ self-triggering is observed only at high voltages (much higher than the proposed V_{op}), then at $2.5 \times 10^{15}\,\mathrm{n_{eq}\,cm^{-2}}$ even at the highest voltage no self-triggering is observed since the gain is low. For FBK and CNM no self trigger is observed for $2.5 \times 10^{15}\,\mathrm{n_{eq}\,cm^{-2}}$ at V_{op}.

Figure 5.7 shows V_{op} as a function of fluence after neutron and proton irradiation for different LGAD types. It can be seen that it increases with fluence up to values over 700 V but never surpassing 750 V for 50 µm sensors.

5.5.3 Collected charge and gain

Figure 5.8 shows the collected charge as a function of bias voltage after neutron and proton irradiation up to $3 \times 10^{15}\,\mathrm{n_{eq}\,cm^{-2}}$ for different LGAD types: HPK-3.2, CNM and FBK-UFSD3-C. Several sensors were tested for each fluence and the results are displayed for a representative sensor, for the maximum fluence two representative sensors are shown for HPK-3.2. The charge at V_{op}, as defined in Section 5.5.2, as a function of for both neutron and proton irradiations, is shown in Figure 5.9.



(a) (b)

Figure 5.7: V_{op} as a function of fluence after irradiation for different LGAD types for neutron (a) and proton (b) irradiation. The red horizontal line represents the maximum allowed voltage of 750 V as discussed in Section 5.5.7. Solid markers indicate n irradiation (*n*), open markers p irradiation at CYRIC (*pCy*).

It is evident that by going to higher fluences the increase in bias voltage can only partially compensate for the loss in gain due to the acceptor removal. A charge of 4 fC was found to be the lower limit that still satisfies the HGTD science requirements in terms of hit efficiency (see Section 5.5.4) and time resolution taking into account the ALTIROC jitter (see Section 6.7). This level is indicated by the horizontal lines.

The following observations are made for the different types:

a. Baseline 50 µm sensor with higher doping and deep gain layer (HPK-3.2)

HPK-3.2 sensors have a deep and high-dose multiplication layer, which leads to a reduced acceptor removal rate. Hence, this type can reach the target charge of 4 fC up to the HGTD target fluence of $2.5 \times 10^{15}\,\mathrm{n_{eq}\,cm^{-2}}$.

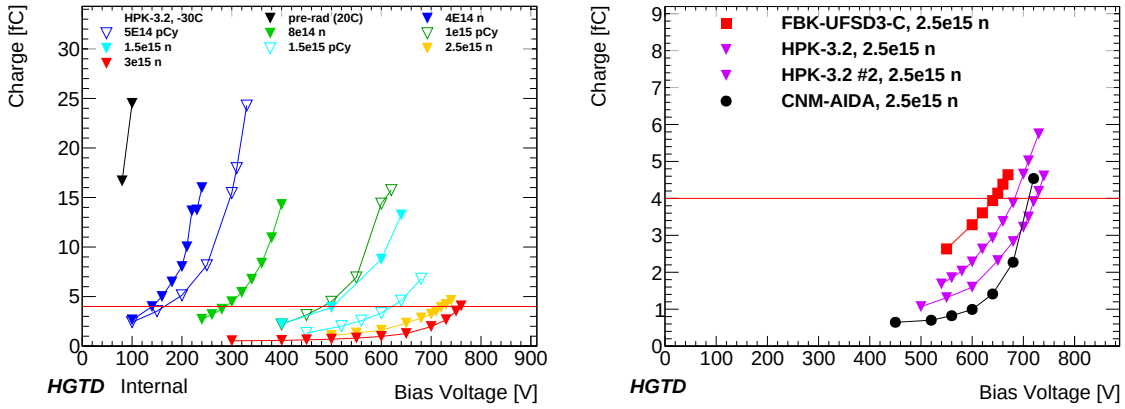
b. 60 µm sensor with gain layer infused with carbon (FBK-UFSD3-C)

The addition of Carbon in the gain layer reduces the acceptor removal. The required bias voltage is thus lower than for other types to reach the target charge of 4 fC at $2.5 \times 10^{15}\,\mathrm{n_{eq}\,cm^{-2}}$.

c. 50 µm sensor with high doping concentration (CNM-AIDA)

CNM-AIDA sensors have a high-dose multiplication layer, also this type can reach the target charge of 4 fC up to $2.5 \times 10^{15}\,\mathrm{n_{eq}\,cm^{-2}}$.

Studies for NDL-33 µm sensors are ongoing.



(a) (b)

Figure 5.8: Collected charge as a function of bias voltage for different fluences for HPK-3.2 (a) and at maximum fluence for all vendors (two representative sensors with different performance for HPK-3.2 are shown) (b). The horizontal lines indicate the HGTD lower charge limit of 4 fC at all fluences. Solid markers indicate n irradiation (*n*), open markers p irradiation at CYRIC (*pCy*). Measurements were performed at −30 °C except for the pre-rad measurement that was done at 20 °C.

5.5.4 Efficiency

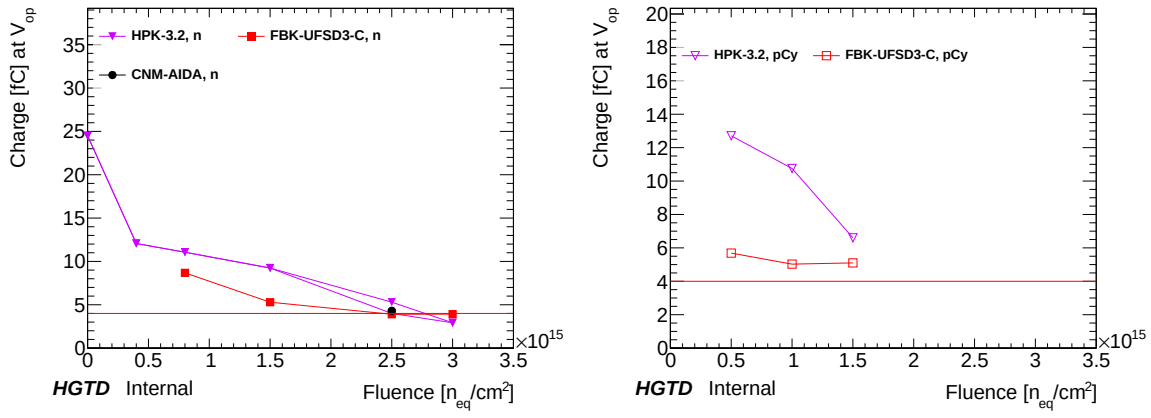
The hit efficiency of LGAD sensors on HGTD-specific readout boards was measured in HGTD beam tests using an external telescope for reference tracks [10]. Figure 5.10 shows the efficiency in the central region of the LGAD pad as a function of most probable charge collected, compiled from 16 different single pad sensors before and after irradiation up to $3 \times 10^{15}\,\mathrm{n_{eq}\,cm^{-2}}$ at different bias voltages. The threshold to accept events with a hit was chosen at a measured noise occupancy of 0.1% and 0.01%, respectively.

It can be seen that a universal curve is obtained, irrespective of fluence, indicating that the charge is the main parameter on which the hit efficiency depends, given a certain noise occupancy. A hit efficiency above 99% is obtained even before the HGTD minimal allowed charge of 4 fC mentioned in Section 5.5.3. The measurements will be repeated with the ALTIROC electronics once available for large-scale testing.

2D efficiency maps are shown in Section 5.5.6 for arrays before and after irradiation. The cross talk between different pads of a 2×2 array was also measured and found to be below 1% before and after irradiation.

5.5.5 Time resolution

The time resolution of LGAD devices have been extensively studied in various beam tests [10, 66, 67] and ^{90}Sr setups [61] on custom-made HGTD-specific readout boards (as stated in



(a) (b)

Figure 5.9: The charge at V_{op} as a function of fluence for neutron (a) and proton (b) irradiation. At the fluence of 2.5×10^{15} n_{eq} cm^{-2} two representative sensors with different performance for HPK-3.2 are shown. The horizontal lines indicate the HGTD lower charge limit of 4 fC at all fluences. Solid markers indicate n irradiation (*n*), open markers p irradiation at CYRIC (*pCy*). The maximum fluence (neutron + charged hadrons) for HGTD is 2.5×10^{15} n_{eq} cm^{-2}, while the maximum charged hadron fluence for HGTD is 1×10^{15} n_{eq} cm^{-2}.

Section 5.4 these measurements will be repeated once the ALTIROC is available for large-scale testing). For the rest of the document (other than Chapter 5) the time resolution is not the same as the one presented here with HGTD-specific analog readout boards. Instead, the measured charge (that is independent from the readout) of the sensor was taken as an input to the ALTIROC time resolution vs. charge function (see Figure 2.13). This is to have a more realistic estimate of the final time resolution with the ALTIROC.

On custom-made HGTD-specific readout boards, it has been consistently shown that 35 ps time resolution can be achieved below the breakdown point before irradiation for sensors from all vendors with pad widths up to 1.3 mm and up to 5 pF capacitance [10, 61, 66–70].

The time resolution of HPK-3.2 was measured in the β–telescope after irradiation with 1 MeV neutrons at Ljubljana, and 70 MeV protons at CYRIC. The results shown in Figure 5.11(a) indicate that a resolution of 40 ps and better is achieved up to a fluence of 1.5×10^{15} n_{eq} cm^{-2} except for non-irradiated sensors (because before irradiation the sensor breaks down before saturation of drift velocity, as explained halfway through Section 5.2.1 and shown in Table 5.3). As seen in Figure 5.11(b) a time resolution of around 60 ps is reached for HPK-3.2 and FBK-UFSD3-C sensors for a fluence of 2.5×10^{15} n_{eq} cm^{-2}. CNM-AIDA sensors show a time resolution of 40ps even at the highest fluence. In Figure 5.12 it is shown that the resolution for FBK-UFSD3-C and HPK-3.2 at V_{op} changes from better than 40 ps at low fluences to 60 ps at the maximum fluence.

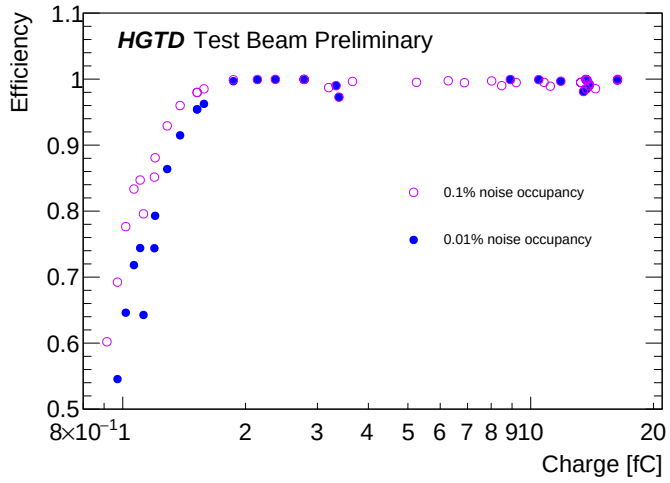


Figure 5.10: Hit efficiency in the central region of the LGAD pad as a function of collected charge at a measured noise occupancy of 0.1% and 0.01%.

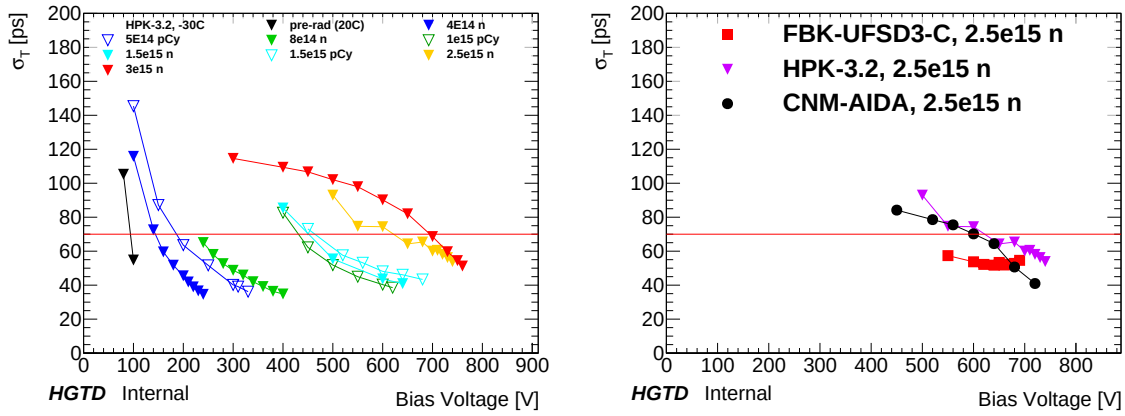
5.5.6 Uniformity, inter-pad gap and edge region

One critical parameter of HGTD is the sensor fill factor, corresponding to the portion of the detector which is able to detect particles efficiently. In the original plans a fill factor of 90% was chosen, this would correspond to an inactive region between two pads of around 70 µm for a pad size of $1.3\,\mathrm{mm}^2 \times 1.3\,\mathrm{mm}^2$. Furthermore, the dead region at the edge of the arrays including the guard ring has to be taken into account for the evaluation of the dead area.

CNM and HPK provided the HGTD collaboration with multi pad LGAD arrays of different geometries (2×2, 3×3, 5×5, 15×15) with different inter-pad and edge distances (see Section 5.2.2). The nominal values quoted by the vendor corresponds to distances between structures in the design of the detector. However it does not reflect perfectly the electric field configuration of the sensors. For this reason the values of inter-pad region and edge distances have to be measured in the laboratory with a focused infra-red laser beam or at test beam facilities.

The sensors were studied at CERN's test beam facility [10]. Thanks to the tracking system it was possible to evaluate the efficiency and the time resolution (using a SiPM as timing and efficiency reference) as a function of the particle hit position. The hit efficiency and time resolution uniformity map for a 2×2 array is shown in Figure 5.13 before and after irradiation. It is shown that the hit efficiency is 99% across the pad before and after irradiation, furthermore the time resolution has variation of around 3 ps across the pad center.

In the laboratory the sensors were tested with an infra-red laser of 1060 nm wavelength focused to 10 µm–20 µm FWHM [75]. The light was injected through the sensor's rear opening of the metalization and scanned from one pad to the other. The two profiles of



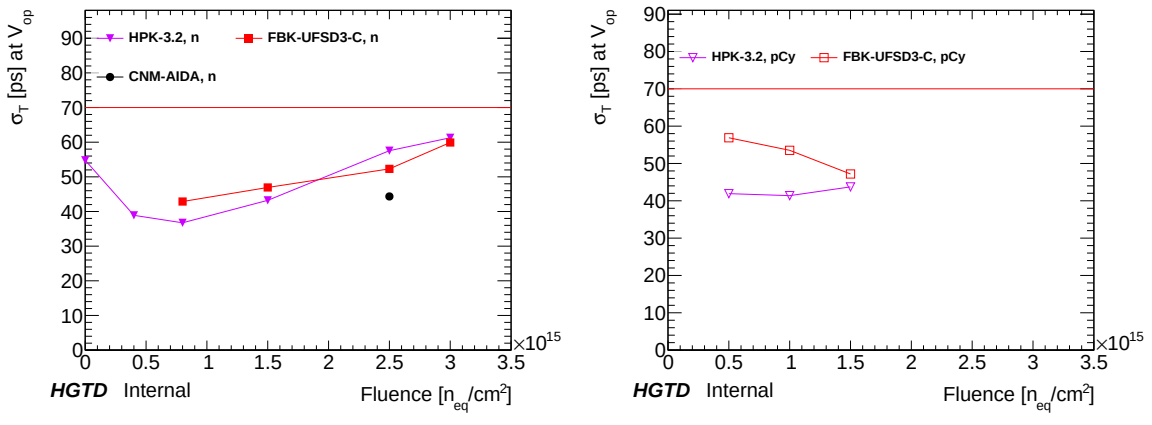
(a) (b)

Figure 5.11: Time resolution as a function of bias voltage for different fluences for HPK-3.2 (a) and for all vendors at the maximum HGTD fluence (b) measured on custom-made HGTD-specific readout boards. Solid markers indicate n irradiation (*n*), open markers p irradiation at CYRIC (*pCy*). The red line represents the maximum allowed time resolution (70 ps) in the lifetime of HGTD. Measurements were performed at −30 °C except for the pre-rad measurement that was done at 20 °C.

the pulse maximum are fitted with a step function smeared by the laser spot width and the distance between the pads is evaluated. The measured effective distance between the neighboring pads can be estimated as the distance where charge collection efficiency drops to 50% on the first pad and rises to 50% on the neighbor (50%-50% point). The inter-pad scans for HPK-3.1 can be seen in Figure 5.14(a). The measured values are around 40 µm higher than the nominal values quoted by the vendor. An overview of the measured vs. nominal values for the HPK-3.1/3.2 and CNM-AIDA sensors can be seen in Figure 5.14(b). As shown in Table 5.6, the lowest measured values per type (roughly 70–90 µm) correspond to fill factors of 87–90%. HPK-3.2 shows an inter-pad gap that is 10 µm–20 µm larger than HPK-3.1 before irradiation. After irradiation of 1.5×10^{15} n_{eq} cm^{-2} the measured inter-pad gap decreases by 20 µm–40 µm due to increased operating voltage and relatively larger multiplication at the edges of the pads.

The edge area is evaluated in a similar way by scanning over the edge of the sensor pad with a laser. Several types of HPK-3.1 with different edge distances were measured in this way and they all showed a 95%-5% drop from the maximum of around 60 µm showing no distortion induced by slimmer edges. Furthermore the guard ring of the sensor was read out and the width of it evaluated with the same technique. For an edge distance of 200 µm a width of 200 µm was seen. For nominal edges of 300 µm and 500 µm a guard ring width of around 350 µm–400 µm was measured, however the sensor with nominal edge of 500 µm has an additional smaller guard ring that is left floating and cannot be read out.

So far no change in sensor performance (collected charge, time resolution) or fragility was



(a) (b)

Figure 5.12: The time resolution at V_{op} as a function of fluence (for neutron (a) and proton (b) irradiation) measured on custom-made HGTD-specific readout boards. The red line represents the maximum allowed time resolution in the lifetime of HGTD. Solid markers indicate n irradiation (*n*), open markers p irradiation at CYRIC (*pCy*). The maximum fluence (neutron + charged hadrons) for HGTD is $2.5 \times 10^{15}\,\text{n}_{\text{eq}}\,\text{cm}^{-2}$, while the maximum charged hadron fluence for HGTD is $1 \times 10^{15}\,\text{n}_{\text{eq}}\,\text{cm}^{-2}$.

Effective IP gap	70 µm	80 µm	90 µm	100 µm	110 µm	120 µm
Fill factor	90 %	88 %	87 %	85 %	84 %	82 %

Table 5.6: Fill factor for different effective (i.e. not nominal) IP gap distances.

observed for the different IP gaps and edge distances except for a lowered breakdown voltage for HPK-3.1-IP30 sensors in case of floating neighbors.

5.5.7 Long term and stability tests

Long term and high flux

HGTD sensors were typically tested to evaluate the performance at low rate, with a laboratory ^{90}Sr source, and medium rate, at test beams. Furthermore, they were biased on the scale of several days. Nevertheless, during the running of the ATLAS experiment, the sensors will be operated continuously for days to weeks in a high particle flux. For this reason, the resilience of the sensors was tested by applying high voltage for an extended period of time. To simulate a high flux, an IR laser was pulsed continuously with a frequency of 50 MHz and the intensity of several MIPs on irradiated HPK-3.1, HPK-3.2 and FBK sensors while biased up to a voltage of 750 V. No change in the behavior of sensors was observed in the timescale of several days.

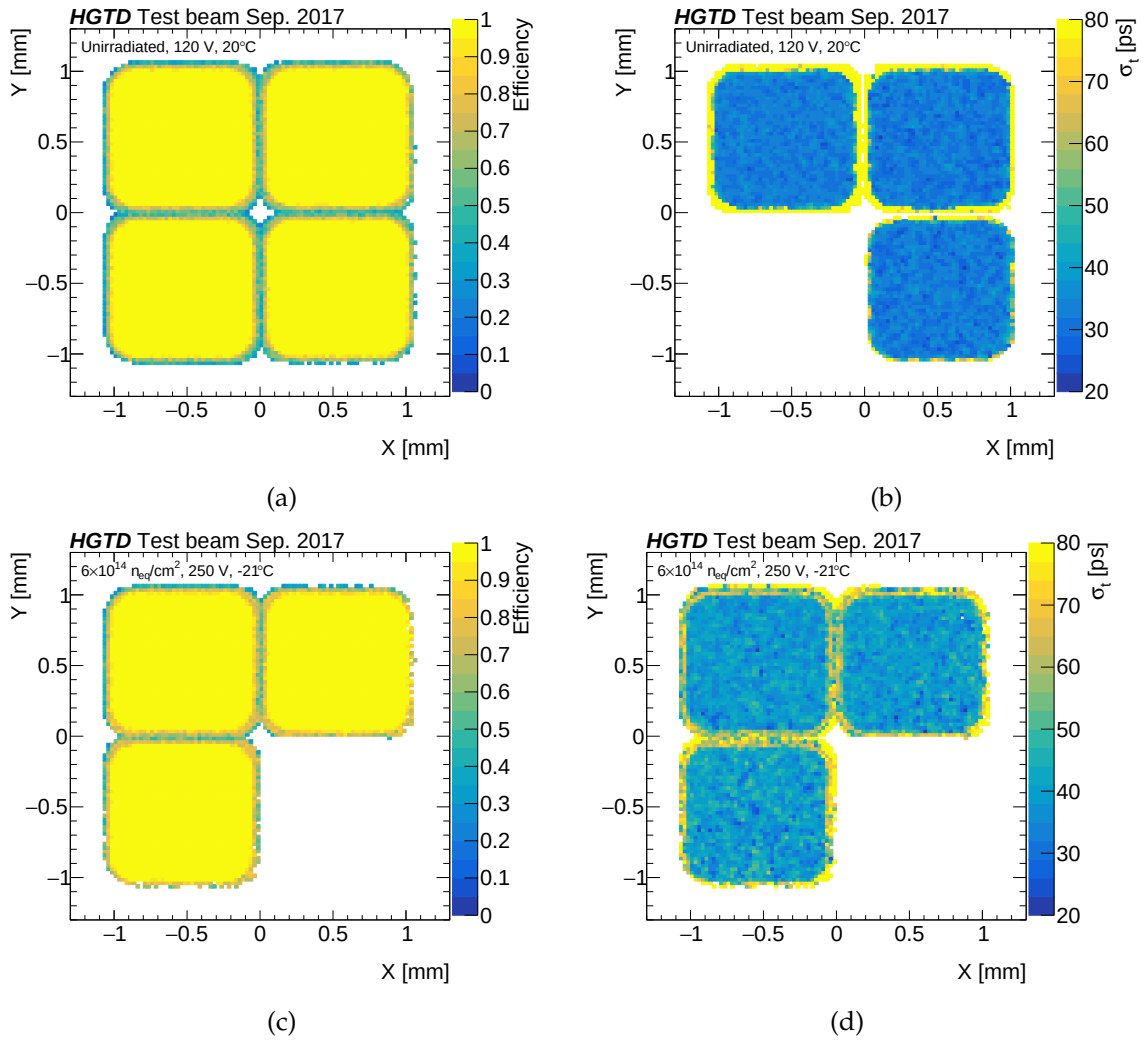
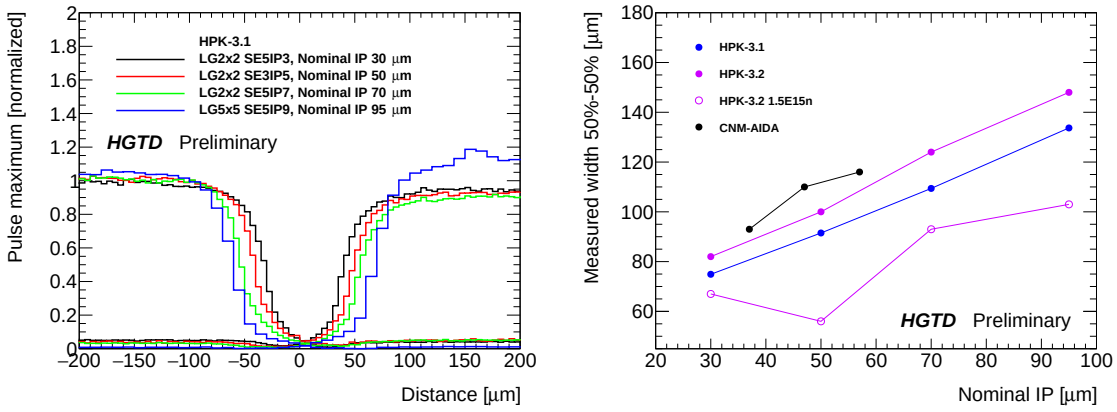


Figure 5.13: 2D maps of efficiency (left) and time resolution (right) before (top) and after n irradiation to 6×10^{14} n_{eq} cm^{-2} (bottom) for a 2×2 array from CNM-10478-50 as measured in HGTD beam tests [10]. Sometimes only 3 channels were measured. The efficiency was evaluated at a threshold of 3 times the noise. A mean efficiency in the pad center of 99% is maintained up to a threshold of 5 times the noise level. The time resolution for this sensor is 39 ps before irradiation with a spread of 3 ps in the pad center.

Sensor breaking and head room

It is important to find a safe bias voltage V_{op} at which the sensors can be operated, as mentioned in Section 5.5.2. During the LGAD R&D phase, these principles were explored with existing sensors listed in Section 5.2.2. As part of the learning curve to define safe operating conditions some of the sensor were broken during testing. Excluding breaking due to mishandling in the large scale lab and beam testing campaign, a few general conclusions



(a) IP laser scans

(b) IP correlation

Figure 5.14: Figure 5.14(a): Inter-Pad distances for several HPK-3.1 sensors. Figure 5.14(b): Nominal vs. measured inter-pad distances for HPK-3.1, HPK-3.2 (before and after irradiation) and CNM sensors.

can be reached for the four sensor types that were tested in depth.

It was observed that thin sensors would break immediately when surpassing a certain critical voltage V_{crit} which depends on the sensor thickness. The distance between V_{op} and V_{crit} is called bias head room. Sensors with thickness of 50 µm (like HPK-3.2 and FBK-UFSD3-C) would break for bias voltages greater than 750 V. Since the bias voltage to operate the sensors increases with fluence almost all breaking occurred at high fluences. The bias head room can be seen in Figure 5.7 as the difference between the V_{op} and the red line at 750 V. For HPK-3.2 the head room is over 150 V until 1.5×10^{15} n_{eq} cm^{-2}, while at the maximum fluence of 2.5×10^{15} n_{eq} cm^{-2} is it 30 V, however this is still much higher than the power supply precision. For FBK-C the bias head room is over 100 V even at the maximum fluence. Several studies will be done to asses the sensors resilience at high voltages close to V_{crit} for long periods of time (test already done are described in Section 5.5.7).

After breaking, a burn mark usually appears in the interface between pad and guard ring, most of the time at the detector corner where the fields are largest. This observation motivates future layout studies of the interface of guard ring and multiplication area. Another study investigates operation of the sensors during temperature and humidity changes and at different particle rates. These few general observations motivate us to make the increase of the bias head room between V_{op} and V_{crit} as one of the research areas of the next prototype run.

Annealing

Most of the measurements with irradiated sensors were done after annealing for 80 min at 60 °C, which roughly simulates the operational conditions in one year of LHC operation since higher temperature accelerates the annealing (the Arrhenius factor between 60 °C and −30 °C is more than 1×10^6, 80 min simulates hundreds of years at −30 °C, and tens of days at room temperature).

A prolonged annealing study was carried out with CNM-10478-50 and HPK-3.1 samples with an area of 1.3 mm × 1.3 mm to check the performance in case of unpredicted situations where sensors would be exposed for longer times to elevated temperatures or when intentional annealing may be used to reduce leakage current and power dissipation.

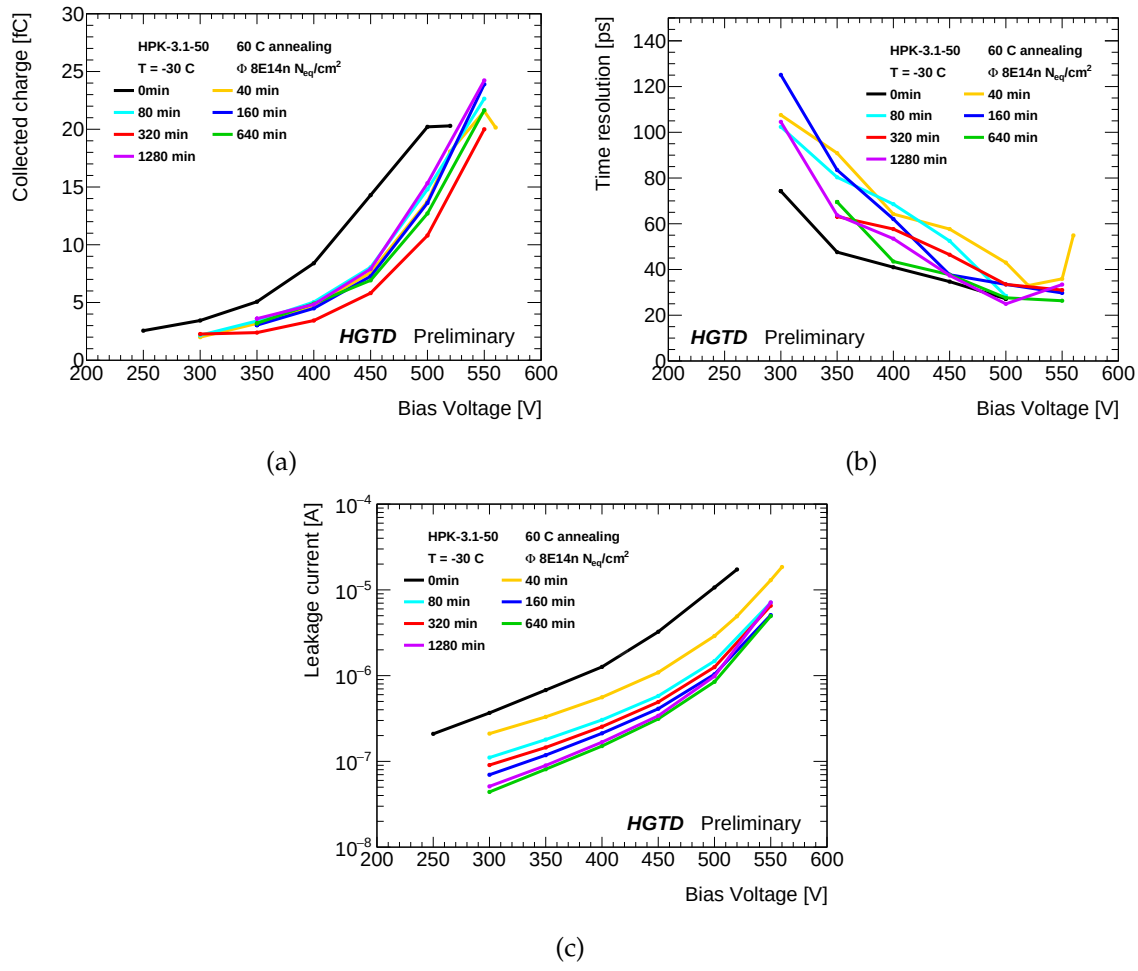


Figure 5.15: Voltage dependence for different annealing times for HPK-3.1 at 8×10^{14} n_{eq} cm^{-2}: (a) collected charge, (b) time resolution and (c) leakage current.

The dependence of collected charge on bias voltage for different annealing times is shown in Figure 5.15(a) for HPK-3.1 samples. It can be seen that the effect of the annealing is limited. There seems to be a decrease of initial acceptors in the gain layer with annealing on a time scale of tens of minutes, but thereafter the charge stays relatively constant. Even if full reverse annealing of deep acceptors takes place, the applied bias voltages are high enough to fully deplete thin detectors and also saturate drift velocity. The effect of annealing on time resolution remains limited (10–20 ps maximal spread at high voltages, with an initial increase and then decrease again) as shown in Figure 5.15(b). A much larger beneficial effect of annealing can be observed on the leakage current as shown in Figure 5.15(c).

There were no significant differences in annealing performance observed between the two producers HPK and CNM. The annealing studies will be extended further to the whole fluence range, different temperatures and producers, so that an accurate running scenario can be made.

5.5.8 Leakage current and power after irradiation

In standard Silicon sensors without gain, the leakage current originating from the bulk increases linearly with fluence. However, for LGADs the situation is more complex due to the gain and its change with fluence. The operation in gain mode leads to an increase of the leakage current, which is given by the product of the volume generation current and the current multiplication factor. As the gain decreases with irradiation and the generation current increases, the leakage current does not necessarily increase monotonically with fluence. The leakage current in multiplication mode contributes to parallel noise linearly, hence it is of high importance to run the sensors at low temperatures since cooling decreases the leakage current (roughly by a factor of 2 every 7 °C).

The leakage current for 1.3 mm $\times$ 1.3 mm HPK-3.2, FBK-UFSD3-C and CNM-AIDA at V_{op} as a function of fluence is shown in Figure 5.16(a). The ALTIROC maximum acceptable current is 5 µA (line in Figure 5.16(a)). All sensors satisfy this requirement up to the highest fluence. From the current per pad the power density (power/area) can be derived. The power can be reduced by operating the sensors at low temperature. For the assumed operating temperature (−30 °C), Figure 5.16(b) shows the measured power density of HPK-3.2, FBK-UFSD3-C and CNM-AIDA as a function of fluence for V_{op}. At the required fluence of $2.5 \times 10^{15}\,\mathrm{n_{eq}\ cm^{-2}}$ the power requirement is fullfilled. As seen in Figure 5.16(d) the power for HPK-3.2 reduces by more than 50% for a reduction of 5% in V_{op}. For the same V_{op} variation only a 10-20% reduction in collected charge is present, this allows a certain elasticity in adjusting V_{op}. The final power dissipation in HGTD will depend on the sensor type choice as well as the operational scenario as detailed in Section 5.6.

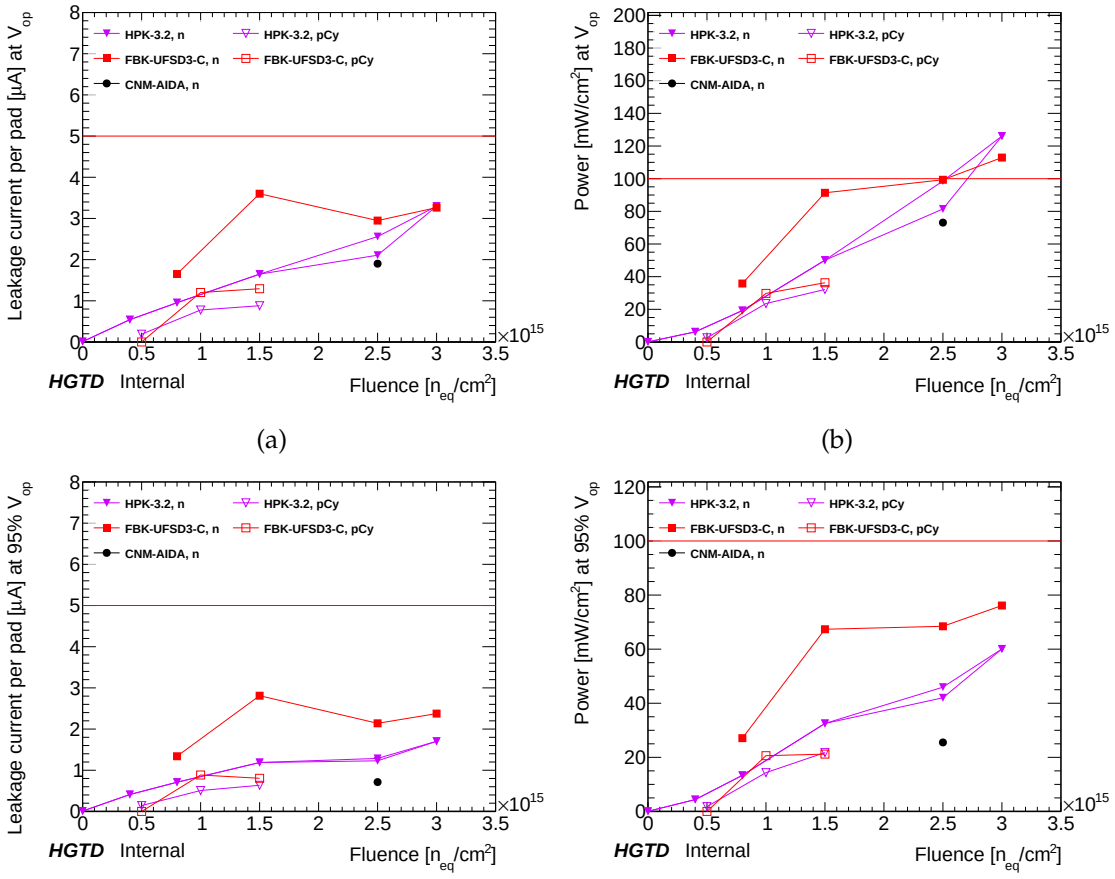


Figure 5.16: (a) Leakage current at the operation bias voltage V_{op} for single pads at −30 °C as a function of fluence for HPK-3.2, FBK-UFSD3-C and CNM-AIDA irradiated with 1 MeV neutrons (solid markers) and 70 MeV protons (open markers). The horizontal line represents the ALTIROC maximum acceptable current of 5 µA. (c) is the same quantity but evaluated at 95% V_{op}.
(b) Power density as a function of fluence at the operation bias voltage V_{op} at −30 °C [68, 69]. The horizontal line represents the maximum acceptable power of 100 mW/cm^2. (d) is the same quantity but evaluated at 95% V_{op}.
In (a), (b), (c), (d) at the fluence of 2.5×10^{15} n_{eq} cm^{-2} two representative sensors with different performance for HPK-3.2 are shown. The maximum fluence (neutron + charged hadrons) for HGTD is 2.5×10^{15} n_{eq} cm^{-2}, while the maximum charged hadron fluence for HGTD is 1×10^{15} n_{eq} cm^{-2}.

5.6 Operational aspects and bias voltage evolution in HGTD

As shown in Section 5.5.3, the bias voltage needs to be increased with increasing fluence, which is a function of radius and integrated luminosity (i.e. period over lifetime) in HGTD. Monitoring of the leakage current and the TOT as an indicator of collected charge will give a good estimate of the gain evolution during operation, providing the information needed to perform the necessary adjustments to the bias voltage.

For a first scenario, it is assumed that the detector is operated at operating bias voltage V_{op} (see Figure 5.7). The expected dependence of the fluence on the radius (Figure 2.14) and the required bias voltage V_{op} for the increasing fluence permits a prediction of the bias-voltage distribution as a function of radius. This is shown in Figure 5.17 for different integrated luminosities for HPK-3.2 sensors. It shows that the ability to connect several nearby modules to the same bias supply allowing a 10% variation in the bias to modules on one bias supply will be limited. Note that the exact behavior depends on the sensor type chosen since different sensor types require different bias voltages for the same performance (see Section 5.5.3).

A study was made to take into account the variation of fluence across the 15x30 chip, which is around 3 cm of radial difference between opposites pads in the HGTD geometry. The fluence variation for the maximum fluence $2.5 \times 10^{15}\,\text{n}_{\text{eq}}\,\text{cm}^{-2}$ (which is the case presenting the maximum variation) is around $3 \times 10^{14}\,\text{n}_{\text{eq}}\,\text{cm}^{-2}$ for the innermost ring, around $2.5 \times 10^{14}\,\text{n}_{\text{eq}}\,\text{cm}^{-2}$ for the middle ring and around $1 \times 10^{14}\,\text{n}_{\text{eq}}\,\text{cm}^{-2}$ for the outer ring. This is reflected to a change in V_{op} of around 50V in the inner ring, around 30V in the middle ring and around 20V for the outer ring. In terms of performance the collected charge change from one edge of the sensor to the other is from 1 fC to 2.5 fC, which is not sufficient to cause self-triggering in the less irradiated pads. Therefore if the most irradiated edge of the 15x30 array is operated at V_{op} (to achieve 4 fC at $2.5 \times 10^{15}\,\text{n}_{\text{eq}}\,\text{cm}^{-2}$) there will be no self-triggering issues (and around 6.5 fC) on the pads at the other edge of the array.

5.7 Summary of present sensor design

Through the R&D program of the last few years, which involved six large LGAD suppliers, several LGAD designs have been investigated. A recommendation for the final design choices will be a "snapshot" taking into account the fact that some of the options need more investigation. So the choices will be tilted towards conservative performance and operations, making use of the operation Voltage V_{op} of Figure 5.7.

- Thickness of the high resistivity bulk
  50 µm: compared to thinner detectors, higher collected charge at high fluence, more resistant to breaking (as shown in Section 5.5.7).

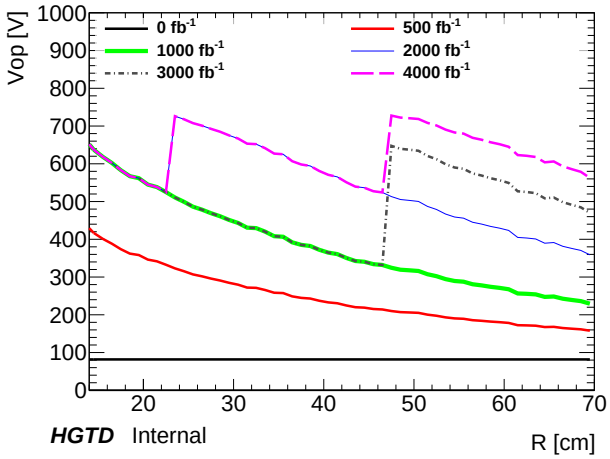


Figure 5.17: V_{op} as a function of the radius for different integrated luminosities for HPK-3.2 sensors. The sudden changes at 2000 fb^{-1}, 3000 fb^{-1} and 4000 fb^{-1} corresponds to the replacement of the inner and middle ring.

- Gain layer doping profile
  Narrow and deep shows improved radiation hardness: however, the performance before irradiation is degraded due to the low breakdown voltage. A compromise between performance before irradiation and radiation hardness needs to be developed.

- Adding Carbon to the dopant in the gain layer
  C implantation is a promising candidate that shows improved radiation hardness. Noise and time resolution need to be understood. Moreover, it is not available yet by all vendors. Further studies are ongoing.

- Inactive distance between pads (inter-pad gap)
  80–120 µm effective inter-pad gap is feasible before irradiation. With irradiation, the performance improves due to increased operating voltage and relatively larger multiplication at the edges of the pads. Further optimizations are ongoing.

- Slim edge distance
  300 µm: Studies show same performance as samples with wider edge.

- Covering the pads with metal
  Complete metal cover of pads: sensors with pads fully covered with metal showed better performance in terms of collected charge than sensors with large non-metal openings.

With this selection of parameters, the science goals will be reached up to the HGTD target fluence of $2.5 \times 10^{15}\,\text{n}_{\text{eq}}\,\text{cm}^{-2}$. More studies after high energy charged hadron irradiation for the innermost radius will be performed once the irradiation facilities are available again.

5.8 Roadmap for future sensor productions and activities

In previous years, several vendors already produced LGAD prototype runs with HGTD geometry that were studied by the sensor Institutes and demonstrated the general feasibility to fulfil the HGTD requirements as described above. In 2020, further R&D geared towards production will follow to consolidate and potentially further improve the radiation hardness and to optimize several geometrical layout issues.

The upcoming R&D will focus on the following:

- Produce first full-size 30×15 HGTD sensors ($4 \times 2\,\mathrm{cm}^2$) to demonstrate the feasibility and provide sensors for the HGTD demonstrator program (see Chapter 14).
- Optimize the inactive inter-pad distance without affecting yield and sensor performance.
- Implement an inactive edge of 300 µm as default.
- Optimize the LGAD technology for improved radiation hardness (reduced acceptor removal), bias voltage head room and reduced power dissipation.
- Conduct irradiation campaigns with high energy charged hadrons at Los Alamos, as well as mixed neutron-charged hadron irradiations.
- Repeat the performance after irradiation with the ALTIROC readout chip once enough prototype assemblies are available.
- Establish the robustness of LGADs under stressful operating conditions.
- Improve breakdown between guard ring and pad area.

To this end, the following R&D and prototype runs are planned in 2020 with various vendors:

- *HPK*: A 2nd shared ATLAS-CMS prototype run is ongoing and expected to finish in the middle of 2020. The main purpose is to optimize the doping concentration in the multiplication layer of HPK-3.2 to improve timing performance before irradiation. 4 doping splits with varying concentration are planned. Furthermore, full size pseudo-30×15 HGTD sensors (i.e. 2 closely placed 15×15 sensors diced out in one piece) are implemented. The default inactive edge will be 300 µm.
- *FBK*: An R&D run is ongoing to optimize the doping concentration of Carbon and to combine the beneficial effects of both Carbon and the deep implant of Boron (like HPK-3.2). Simulation of this combined technology predicts an enhanced radiation hardness due to reduced acceptor removal. Furthermore, a prototype run with large size sensors is planned.

- *CNM*: CNM has transferred its LGAD line to 6" production. A first 6" shared ATLAS-CMS prototype run is ongoing and expected to finish in the middle of 2020. Real 30×15 HGTD sensors with 300 µm inactive edge are implemented. Further tests on carbon implantation, fabrication in high resistivity epitaxial layers and possible further tests on Gallium implantation are planned.
- *Novel Device Laboratory (NDL) and Zhonghuan Advanced Semiconductor Materials Co.*: Further runs will be conducted in 2020 to produce sensors with baseline 50 µm active thickness, try higher Boron doping to improve radiation hardness, optimize the JTE to improve the breakdown voltage and optimize the inter-pad design.
- *Institute of Microelectronics of Chinese Academy of Sciences (IME)*: first runs in 2020 to produce sensors with baseline 50 µm active thickness, implement Carbon implantation.

After this extended R&D and prototype phase and based on the understanding of the design issues solved, the sensor SPR will start Q3 2020, then the PDR will be submitted in Q1 2021, followed by a market survey and the FDR in Q4 2021. The sensor pre-production will take place March to August 2022, followed by the production from January 2023 to October 2024 (as seen in Figure 15.3).

6 Front-end Electronics

This chapter describes the required performance, design, and latest prototype testing of the ASIC chip, ALTIROC, that will be bump bonded to the LGAD sensor. It will have 225 readout channels, thus two ASICs will read out each LGAD. The main challenge in the design of this ASIC is the fact that it needs to have a small enough contribution to the timing resolution, in order to match the excellent performance of the LGAD. As introduced in Section 4.2.2, this contribution comes mainly from the time walk and the jitter. The first one will be addressed by applying a correction based on the fact that the variations in the time-of-arrival (TOA) of the pulse are related to the time-over-threshold (TOT); this is presented in Section 6.3.2. The most critical aspect concerning the jitter is the design of the analog front-end electronics, which are composed of a voltage preamplifier followed by a fast discriminator. The measured TOA and time-over-threshold are digitized using two time-to-digital converters (TDCs), and stored in a local memory at the channel level. An end-of-column (EOC) logic is implemented to collect the information for each of the 15 columns (with 15 pads each). The ASIC common digital part is composed of different blocks necessary to generate and align the clocks, receive the slow control commands to configure the ASIC and transmit the digitized data.

Two iterations of this chip have been produced and tested so far: the first, ALTIROC0, integrated four pads in a 2×2 array, with the analog part of the single-channel readout: the preamplifier and the discriminator. The results of the test beam and test bench studies performed on this version of the ASIC can be found in [60]. The second iteration, ALTIROC1, consists of a 5×5 pad matrix, in which the digital components have been added to the single-channel readout.

The requirements imposed by the data taking conditions, the sensor and the targeted performance are presented first in Section 6.1. The ASIC architecture is described in Section 6.2, first going through the single-channel architecture and then the entire ASIC. Section 6.3 describes in detail the design of the single-channel readout electronics, followed by the description of the ASIC common digital part in Section 6.4. The radiation tolerance is described in Section 6.5 and the power distribution in Section 6.6 The performance results obtained so far in test bench and test beam are described in Section 6.7. The description of the monitoring can be found in Section 6.8. Lastly, a brief account is given of the future steps towards the completion of the design and testing of the ASIC in Section 6.9.

6.1 General requirements

The requirements of the ASIC can be divided into two types. On one side the considerations regarding the operational environment of the ASIC, its powering and electrical connections. These requirements are summarized in Table 6.1. The second group concerns the ASIC performance, driven by the targeted time resolution. A summary of these requirements is presented in Table 6.2.

- The ASIC will have to withstand high radiation levels and, as in the case of the sensors, some ASICs will have to be replaced during the HL-LHC period. As they are designed in pure CMOS technology, they are mainly sensitive to the TID. The expected radiation levels have been presented in Section 2.4, considering a 2.25 safety factor for the electronics leading to a maximal TID of 2.0 MGy (see Figure 2.14).
- Each single-channel readout needs to fit within the sensor pad, with sides of 1.3 mm. It will be capable of handling up to 5 µA leakage current from the sensor without degrading the ASIC performance
- Because the signal from the sensor will degrade due to the effects of irradiation, it should be possible to set the discriminator threshold for small enough values of input charge. The minimum threshold (2 fC) should provide an efficiency above 95% for an input charge of 4 fC (although with a jitter larger than 25 ps). To enable the possibility to set such low thresholds, the cross-talk between channels should be kept below 5%.
- The target for the electronics is to be able to read out signals from 4 fC up to 50 fC throughout the HGTD lifetime.
- The electronics jitter for an input charge of about 10 fC is required to be smaller than 25 ps, i.e smaller than the dispersion induced by the Landau fluctuations on the deposited energy which limits the time resolution to 25 ps at large sensor gain. Such charge is equivalent to the deposited charge of a MIP in a 50 µm thick LGAD with a gain of 20. A detector capacitance of about 4 pF is considered. The contribution to the time resolution from the TDC should be negligible and leads to a 20 ps TDC bin for the TOA measurement and a 40 ps TDC bin for the TOT measurement. The time walk should be smaller than 10 ps over the dynamic range after correction.
- The TOA and TOT information are transferred to the data acquisition system only upon L0/L1 trigger reception with latency up to 35 µs [76], therefore necessitating a large size memory. The trigger rate depends on the final scheme adopted. It will be 1 MHz for an L0 trigger, or 0.8 MHz (resp. 0.6 MHz) for an L1 trigger in an L0/L1 scheme with an L0 at 2 MHz (resp. 4 MHz).
- The global phase adjustment of the clock should be guaranteed to a precision of 100 ps in order to properly center the 2.5 ns measuring window at the bunch-crossing.

- The ASIC will need to handle the information to perform the luminosity measurement, computing the number of hits per ASIC on a bunch-by-bunch basis. To limit the global bandwidth required, the information of only a subset of all the ASICs is used. The current proposal is to use the sensors located at 470 mm $< r <$ 640 mm, or equivalently $2.4 < |\eta| < 3.5$. The use of both layers will not provide a significant increase in coverage with respect to one of the layers, but the redundancy aids in estimating and reducing the systematic uncertainty on the measured luminosity and provides contingency in the event of failures in the instrumentation.
- Finally the ASIC power dissipation should be kept as low as possible, in order to limit the size required for a single CO_2 cooling unit (for more details on the cooling system see Section 11.3).

Pad size	$1.3 \times 1.3\,\text{mm}^2$
Voltage	1.2 V
Power dissipation per area (per ASIC)	$300\,\text{mW cm}^{-2}$ (Total: 1.2 W)
e-link driver bandwidth	$320\,\text{Mbit s}^{-1}$, $640\,\text{Mbit s}^{-1}$, or $1.28\,\text{Gbit s}^{-1}$
Temperature range	−40 °C to 40 °C
TID tolerance	2.0 MGy
Full Chip SEU Upset probability	$< 5\%$/hour

Table 6.1: Geometrical, environmental, electrical and power requirements for the HGTD ASIC.

Maximum leakage current	5 µA
Single pad noise (ENC)	$< 3000\ e^- = 0.5$ fC
Cross-talk	< 5%
Threshold dispersion after tuning	< 10%
Maximum jitter	25 ps at 10 fC
	70 ps at 4 fC
TDC contribution	< 10 ps
Time walk contribution	< 10 ps
Minimum threshold	2 fC
Dynamic range	4 fC–50 fC
TDC conversion time	< 25 ns
Trigger rate	1 MHz L0 or 0.8 MHz L1
Trigger latency	10 µs L0 or 35 µs L1
Clock phase adjustment	100 ps

Table 6.2: Performance requirements for the HGTD ASIC. The values given for the noise, minimum threshold and jitter have been specified considering a detector capacitance $C_d = 4$ pF.

Data transmission bandwidth requirements

The required bandwidth of the readout e-link of each ASIC strongly depends on the radial region it covers, as shown by the distribution of the average number of hits per ASIC in Figure 9.4. The number of bit per hit is 19 as described in Section 6.3.5.

Each module consisting of two ALTIROC ASICs is connected via a flex cable to a Peripheral Electronics Board (PEB), described in Chapter 9. The PEB transfers digital signals from the flex cables to optical fibres connected to the back-end DAQ. Flex cables for modules placed at a radius above 320 mm also carry two differential e-links with luminosity data. For error-free data transmission at the bandwidths required by the expected HGTD data volume, the PEB uses the low-power GigaBit Transmission chip (lpGBT [77]). A dedicated buffer is needed in each ASIC to average the rate variation and match the best speed of the e-link drivers/lpGBT transceiver inputs:

- The largest average hit rate at small radius does not exceed 20 hits per ASIC and per event, equivalent to a rate of 500 Mbit s^{-1} (not including header). In the current design a bandwidth of up to 1.28 Gbit s^{-1} was considered for the innermost radius ASICs (up to $r \simeq 150$ mm), taking into account a considerable safety margin. However if further studies confirm this, a lower maximum bandwidth could be considered, thus reducing the number of necessary lpGBTs.
- For larger radii, a 320 Mbit s^{-1} bandwidth can be used.
- For the luminosity measurement, the 12 bits of data for the counts in the larger and smaller window is expanded to 16 bit using the 6b8b encoding (see Section 6.2.1). Therefore a 640 Mbit s^{-1} e-link driver and lpGBT speed is needed.

6.2 ASIC architecture

With an area of 19.9 mm × 21.7 mm, the largest part of the chip will be occupied by the channel matrix: each pad being 1.3 mm × 1.3 mm, arranged in a matrix of 15 × 15 channels. The channel matrix will thus have an area of 19.5 mm × 19.5 mm; the additional space is needed to accommodate the end-of-column logic and the common digital blocks.

This section presents an overall description of the three main structures of the ASIC:

- the single-channel readout cell, which is repeated 225 times. It integrates the preamplification, the discrimination and the digitization of the hits as well as the local storage (or buffering) of the digitized data until an L0/L1 trigger is received.
- the EOC logic which performs the readout of the 15 columns and transfers the data to the trigger data and luminosity processing units.

- the ASIC common digital part which formats the digitized data before sending it to the peripheral off-detector electronics that will be described in Chapter 9. This stage also contains common cells such as a phase shifter, a Phase-Locked Loop (PLL) and a fast command decoder that will be described in Section 6.4.

The ASIC has been designed using 130 nm TSMC[1] technology. Simulations have been performed using the 130 nm TSMC design kit provided by CERN. The TSMC 130 nm technology has been tested up to 4 MGy [78]. A radiation hard digital library is not available for this technology and the design uses the standard library. However the ASIC has been designed to ensure its radiation hardness as described in Section 6.5.

6.2.1 Channel architecture

A conceptual schematic for the single-channel readout is presented in Figure 6.1. Each readout channel will consist of a preamplifier followed by a discriminator, both of which are critical elements for the overall electronics time performance. A detailed characterization of the preamplifier is presented in Section 6.3. The time of the pulse will be determined using a discriminator that follows the preamplifier using a fixed threshold. As a consequence, a time walk correction needs to be applied in order to account for the dispersion in the TOA due to the different pulse heights. Since the time walk will be corrected using a Time Over Threshold architecture (described in Section 6.3.2), two TDCs are necessary to digitize the discriminator output. The first is for the digitization over 7 bits of the TOA, which corresponds to the position of the rising edge of the discriminator output. The range used is 2.5 ns, and it will be done with a bin of 20 ps. The second TDC will be used for the digitization over 9 bits of the width of the discriminator output. The bin and range of the TOT-TDC will be 40 ps and 20 ns respectively. Further details on the TDCs are presented in Section 6.3.3.

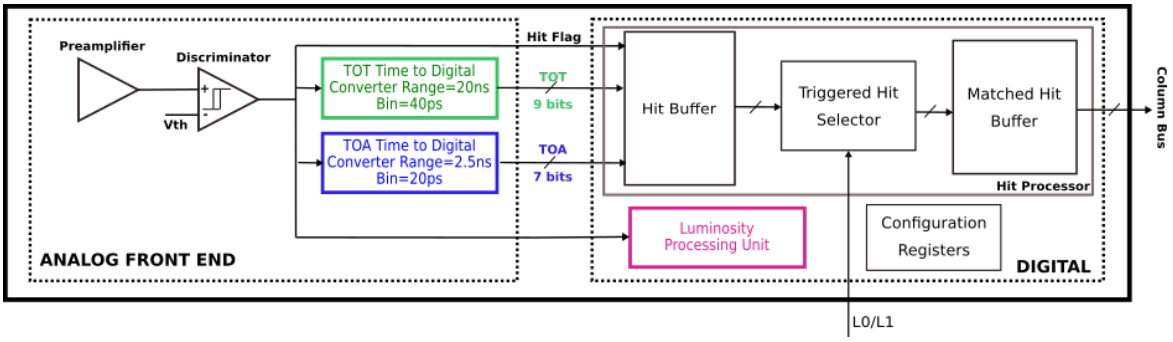


Figure 6.1: Schematic of the single-channel readout electronics. Two main blocks are identified, the analog and the digital part. The input pulse from the sensor enters the preamplifier on the left. The TOA and TOT data are read out by the column bus on the right.

[1] TSMC stands for Taiwan Semiconductor Manufacturing Company.

The output of the analog read-out is processed by the digital stage providing two different measurements: time and luminosity. The 16 bits of the time measurement data, combined with 1 bit for a hit flag, are then stored in a local memory (named *hit buffer*). The content of this buffer is processed by a triggered-hit selector circuit on arrival of an L0/L1 trigger signal, so this memory should allow latencies of up to 35 µs. If a trigger signal is received, the information is passed on to a secondary buffer named *matched hit buffer*, where it remains until it is retrieved for transmission to the common digital part. These local memories are further described in Section 6.3.5.

In order to measure the online bunch-by-bunch luminosity, each ASIC will report the sum of hits within two different time windows. A schematic drawing of the windows is shown in Figure 6.2. The two windows W1 and W2 are centred at the expected arrival time of the particles from the collisions with their length adjustable via configuration parameters. The window W1 is 3.125 ns wide while the second window W2 is adjustable in length in steps of 3.125 ns, and will count the number of particles arriving before and/or after those from the collisions. This sideband will provide valuable information about the background, as described in Section 10.3.4. The window generator is a control unit within the logic at the end of each column that contains a 4-bit counter running at 640 MHz and synchronized to the 40 MHz clock (both provided by the phase-shifter further described in Section 6.4.1). The length and alignment are adjustable via configuration parameters, and are performed with the phase-shifter located in the common digital part (further described in Section 6.4.1). These parameters will be optimised based on operational experience.

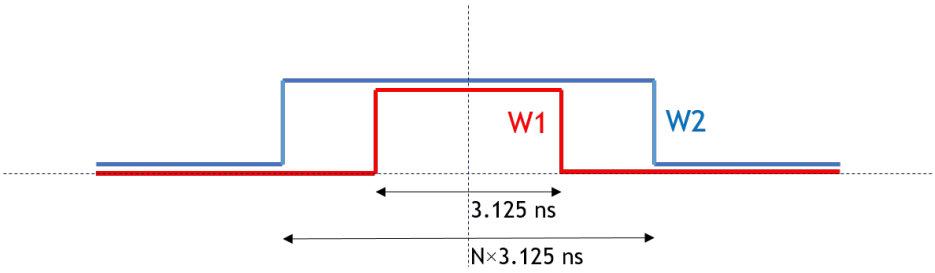


Figure 6.2: Illustration of the time windows used for counting hits for the luminosity data. The smaller window (W1, in red) is 3.125 ns wide and is centred at the bunch crossing time. The width and relative location of the larger window (W2, in blue) can be set in steps of 3.125 ns through the control parameters.

The luminosity measurement is done in three steps. For the first step, performed at the single-channel level, the output of the discriminator is passed through two programmable windows to determine whether the hit happened inside these windows. The way this is done is further described in Section 6.3.6. Secondly, the number of hits per column is computed by the EOC logic, and thirdly the data is transferred. These last two steps are described in the next section and in Section 6.4.

Lastly, there are four 8-bit configuration registers per channel. They are read/written by the

slow control unit through a Wishbone bus. The configuration registers allow configuring several features of the TDCs, to enable/disable the discriminator and preamplifier, and to configure the per-channel threshold correction of the discriminator.

6.2.2 Readout architecture

Figure 6.3 shows the conceptual design of the entire HGTD ASIC with 225 channels. The channel matrix is represented on the top part by 15×15 small squares. The schematic of a single channel, as presented in Figure 6.1, is repeated for each small square. The readout of the channels is done by column, through an EOC cell, drawn at the bottom of the matrix. The information is passed on to the trigger and luminosity processing units. A diagram of the main ASIC common digital part is presented at the bottom.

A fast command unit receives the fast commands from the central Trigger Data Acquisition system (TDAQ) through an lpGBT chip. These commands are 8-bit long[2] and are received in series at 320 Mbit/s, one per bunch crossing. The communication between the fast command unit and the lpGBT chip is done through two lines. One is for serial data, and the other to transmit a clock of 320 MHz which will be used, not only to establish the communication between the lpGBT and the ASIC, but also as a source clock from which all the internal clocks needed to operate ALTIROC will be generated. The 320 MHz clock from the lpGBT is divided by 8 and passed to a phase-locked loop (PLL) which produces clocks of 40 MHz, 80 MHz, and 640 MHz. These clocks will be centred with an accuracy of 97.6 ps using a phase shifter. Further details about the clock generation and distribution are given in Section 6.4.1.

The fast commands are processed by the Trigger Data Processing Unit (TDPU) which is responsible to read the timing information from the pixel matrix, pack these data into frames and serialize them. It is composed of a 12-bits bunch crossing counter to generate a bunch crossing identifier (BCID), a trigger table to store temporally trigger events for later processing, a data formatting unit that packs data into frames, and a serializer. More details are given in Section 6.4.2.

The TDPU performs two tasks in parallel, one is to process incoming triggers and the other to readout data associated to a triggered event from the pixel matrix. In the incoming trigger processing task, the TDPU generates an internal trigger signal and a trigger identifier (`TrigID`) when an L0/L1 accept command is received. These triggers are transmitted immediately to all the pixels. Then each pixel checks if it has data associated to that trigger event. If they have, the data are transferred, together with the corresponding `TrigID` to a secondary in-pixel buffer. They remain there until they are retrieved by the TDPU. The `TrigID` is used to tag a BCID with a trigger event with only 5-bits, so it is not necessary to send the 12-bits of the BCID to the pixel matrix. The trigger table is a FIFO that stores

[2] The 3 most significant bits contain a synchronization pattern (110), the 5 less significant bits the command code.

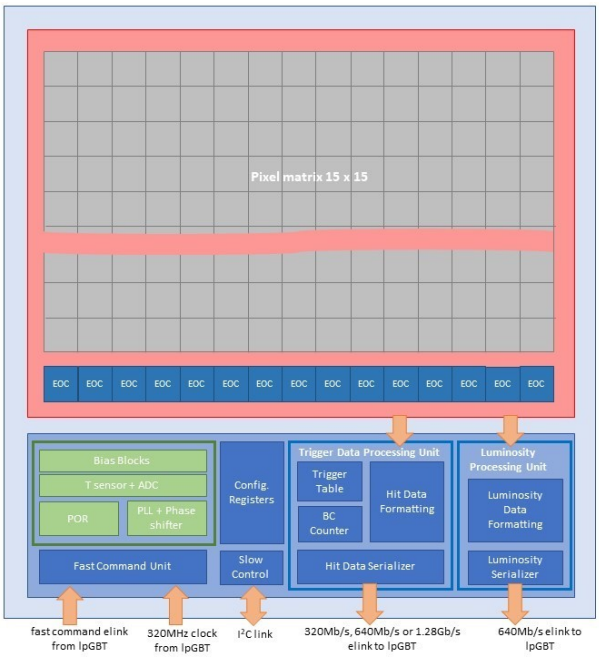


Figure 6.3: Schematic of the full HGTD ASIC. The top part represents the 15 × 15 channel matrix, while the bottom part shows the ASIC common digital part.

the correspondence between each BCID and its associated `TrigID`. In the readout task, the TDPU is looking for a new entry in the trigger table. When a new one is found, it requests to the EOCs to retrieve and store the data from the pixels related to the `TrigID` fetched from the table. Then, the data are moved into the Hit Data Formatting unit, where they are packed into frames, serialised and transmitted to the peripheral on-detector electronics through e-links. The transmission speed of the e-link will depend on the radial position of the ASIC, and will be set via an Inter-Integrated Circuit bus I^2C to one of three values: 320 Mbit s^{-1}, 640 Mbit s^{-1}, and 1.28 Gbit s^{-1}. It is connected to an equal speed port in the lpGBT, described in Sec. Section 9.1.1.

As mentioned previously, the luminosity measurement is carried out in three steps, each one in a different region of the ASIC. The first step consists in determining whether the hit occurred within one or both of the time windows. This windowing process is done at the single-channel level and was described in the previous section. The windows are generated in the logic at the end of each readout column, instead of at each channel, in order to reduce power consumption. By distributing them to the channels as a clock tree, one can compensate for the delays introduced by the long metal lines needed to reach each channel and to minimize the skew between the channels in a column. In the second step, the result is collected at the EOC logic, where the number of hits in the column for each window is computed. This information is passed on to the Luminosity Processing Unit (LPU), that calculates the total number of hits in the ASIC within S1 and S2 windows. Then it performs

the subtraction of the hits within the larger and the smaller window (S1-S2). The 8 bits of S1 and the 8 bits of S2-S1 are truncated to respectively 7 and 5 bits to reduce the total bandwidth. In the third step, each 12 bits packet is transferred to the luminosity serializer where data is encoded (6b8b), leading to frames of 16-bits long. These are serialized at a rate of 40 MHz and sent to the lpGBT through a 640 Mbit s^{-1} e-link. The measurement and data transmission can be enabled/disabled by accessing one of the configuration registers. As explained previously in Section 6.1, not all ASICs will be performing luminosity measurements. Disabling the data transmission on those not performing the measurement will allow to save power.

The common digital part also includes several programmable digital-to-analog converters (DACs) to generate different bias currents for all analog blocks of the ASIC, a band-gap, a temperature sensor and some configuration registers. The latter are used to set different features of the ASIC, such as the values of the DACs, the transmission rate of the hit data and the PLL bias currents or frequencies. As mentioned previously, 4 configuration registers are also present for each channel. The I^2C link mentioned previously is also used to readout all configuration registers in order to check if single-event upsets (SEUs) have corrupted their content, and to retrieve information from the control unit about the status of the ASIC; the information related to data corruption is then passed on to the hit serializer.

The design of this ASIC is on-going but several elements (preamplifier, discriminator and TDC) were already produced and tested as described in Section 6.7.

6.3 Single-channel readout electronics

This section describes in detail the design of the single-channel readout electronics. As introduced previously, it will receive the pulse signal from the LGAD sensor, and transmit the TOA, TOT and luminosity information to the EOC logic. The preamplifier design is first described in Section 6.3.1, while the discriminator is presented in Section 6.3.2. Concerning the digital blocks, the working principle of the TDCs is presented in Section 6.3.3, while the designs of the local memory and the luminosity processing unit are presented in Section 6.3.5 and Section 6.3.6 respectively.

6.3.1 Preamplifier

The jitter due to electronics noise is often modelled as

$$\sigma_{\text{jitter}} = \frac{N}{dV/dt} \sim \frac{t_{\text{rise}}}{S/N} \tag{6.1}$$

where N is the noise and dV/dt the slope of the signal pulse, of which S is the amplitude and t_{rise} the rise time. Due to the fact that the noise scales with the bandwidth (BW) as

$\sqrt{BW}$, while the rise time grows with the amplitude as S/BW, the most common timing optimisations rely on using the fastest preamplifier.

Most timing measurements in test beam have been carried out with broadband amplifiers, which are voltage sensitive amplifiers with 50 Ω input impedance. Some prefer using a trans-impedance configuration, and timing optimisation has been published for such configuration [10, 66]. However, in silicon sensors such as LGADs, the preamplifier speed is not so crucial, due to the fact that the current duration is not negligible with respect to the preamplifier rise time and to the capacitive impedance of the sensor.

The jitter with a voltage sensitive amplifier configuration can be calculated under some simplifications and assuming that the detector current is a short pulse with a characteristic time t_d. The corresponding input charge Q_{inj} is the integral of this current over t_d. The jitter of a preamplifier can then be estimated through the following formula:

$$\sigma_{\text{jitter}} = \frac{e_n C_d}{Q_{\text{inj}}} \sqrt{\frac{t_{r,pa}^2 + t_d^2}{2 t_{r,pa}}} \tag{6.2}$$

where e_n is the noise spectral density and C_d the detector capacitance. The sensor drift time t_d and the preamplifier rise time $t_{r,pa}$ are combined in quadrature as an estimation of the total speed. It can be seen that the jitter is minimized when the preamplifier rise time is equal to the sensor drift time: $t_{r,pa}$=t_d. In that case, the jitter can be written as:

$$\sigma_{\text{jitter}} = \frac{e_n C_d \sqrt{t_d}}{Q_{\text{inj}}} \tag{6.3}$$

However this dependence is small: for instance for $t_d \sim 600$ ps, reducing or increasing by a factor of two $t_{r,pa}$ with respect to the optimal matching value will deteriorate the jitter by approximately 12%. Therefore to minimize the jitter, the sensor should have a small capacitance, a small t_d and provide a large charge. For a 50 µm thick active LGAD in HGTD, a $C_d = 4$ pF has been estimated when fully depleted (see Figure 5.3(b)); typically $t_d \sim 0.6$ ns, and for a gain of 20, it would give a $Q_{\text{inj}} \sim 10$ fC.

The design of the ALTIROC uses a voltage sensitive preamplifier, presented in Figure 6.4. This is a broadband preamplifier with a cascoded Common Source configuration, consisting of an input transistor ($M1$) and a follower transistor ($M2$). Both the gain and the noise depend on the current that flows into the input transistor, which is why the drain current I_d is tunable through configuration parameters. For this purpose two current sources are combined: I_{d1} is a fixed current source of 150 µA, while I_{d2} can be varied from 0 to 850 µA. Simulation studies have shown that the improvment is small when increasing this current beyond 600 µA. The rise time of the preamplifier can be modified in order to optimize the jitter.This is done through the pole capacitance, C_p, that is tunable by slow control (from 0 to 175 fF) allowing to set the preamplifier rise time between 300 ps and 1 ns. As for the fall time

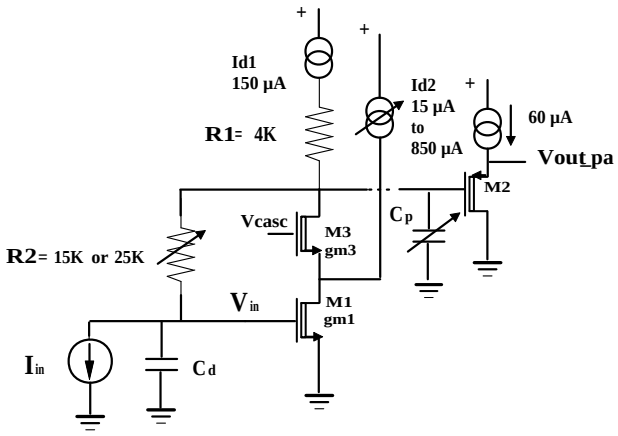


Figure 6.4: Schematic for the preamplifier implemented in the latest ASIC design, ALTIROC1.

of the preamplifier output, it depends on the input impedance of the preamplifier (R_{in}) that is given by the resistance R_2 divided by the open loop gain of the preamplifier. The value of the input impedance depends therefore also on the drain current I_d. For example, for an $I_d = 300\,\mu\text{A}$ and $R_2 = 25\,\text{k}\Omega$, the input impedance is about $1.6\,\text{k}\Omega$. The value of the resistor R_2 can be either $15\,\text{k}\Omega$ or $25\,\text{k}\Omega$. It can also absorb the sensor leakage current, estimated to be below $5\,\mu\text{A}$ per channel after irradiation. The leakage current would cause the output of the preamplifier to drift by an amount of the order of $R_2 \times I_{\text{leak}}$. The threshold of the discriminator that follows the preamplifier must then be changed accordingly using the 10-bit DAC threshold common to all the channels and the 7-bit DAC threshold correction that is integrated for each channel allowing a correction within $\pm 50\,\text{mV}$ as described in the next section.

The preamplifier architecture, followed by a fast discriminator, has been simulated with various detector capacitances and considering that 1 MIP would deposit a 10 fC charge. A calibration signal was used in the simulation, and the result was convoluted with different input LGAD signals. The LGAD pulses for different levels of irradiation obtained using the Weightfield2 software [79] and presented in Figure 4.4(b) were used as input, and the obtained preamplifier pulses are presented in Figure 6.5.

6.3.2 Discriminator

The measurement of the TOA of the particles is performed by a discriminator that follows the preamplifier. The measurement of the time of the rising edge of the discriminator pulse provides the TOA, while that of the falling edge, combined with the TOA, provides the TOT. To ensure a jitter smaller than 10 ps, the discriminator is built about a high speed leading edge architecture with hysteresis to avoid re-triggering effects. Two differential stages with small input transistors are used to ensure a large gain and a large bandwidth

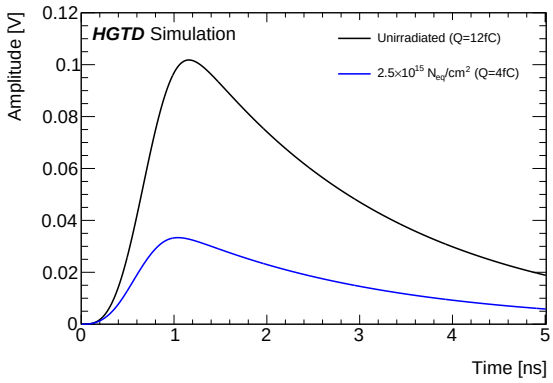


Figure 6.5: Post layout simulation of the preamplifier output using as input the simulated LGAD signals presented in Figure 4.4(b) for a non-irradiated sensor and after irradiation with neutrons.

(aprox 0.7 GHz). The threshold of the discriminator (V_{th}) is set by a 10-bit DAC common to all channels (LSB=0.4 mV). An additional 7-bit DAC (LSB=0.8 mV) allows to make threshold corrections individually for each channel in order to compensate for differences amongst them or for different values of leakage current.

The time walk is the effect that larger signals cross a given threshold earlier than smaller ones (see [61]). The time-over-threshold is defined as the width of the discriminator signal which is a proxy to the signal amplitude and can be used to correct for the time walk effect as illustrated later in the prototype performance section (see Section 6.7).

6.3.3 TDC

The target quantisation step of the TDC of the TOA is 20 ps, and is below the gate-propagation delay in 130 nm technology, thus the Vernier delay line configuration is employed. This configuration consists of two lines, each composed of a series of delay cells implemented as differential shunt-capacitors, controlled by a voltage signal (V_{ctrl}) that determines their delay. The timing resolution is determined by the difference in the delays of the cells in each line. The TOA will be measured within a 2.5 ns window centred at the bunch-crossing. As already mentioned before, the hits have a time dispersion with an RMS of about 300 ps, so that such a window aligned with a precision of 100 ps contains all the hits. The maximum conversion time for a 2.5 ns range must be below 25 ns so that hits happening in the following bunch crossing can be converted.

A graphic representation of the working principle of the TDC can be found in Figure 6.6. In the slow line, the control voltage fixes the delay of each cell to 140 ps, while on the fast line it fixes it to 120 ps. The START signal (rising edge of the discriminator) enters the slow delay line while the STOP signal (next rising edge of the 40MHz clock) enters the 'fast' delay line. Although initially the START signal is ahead of the STOP one, each delay-cell stage brings

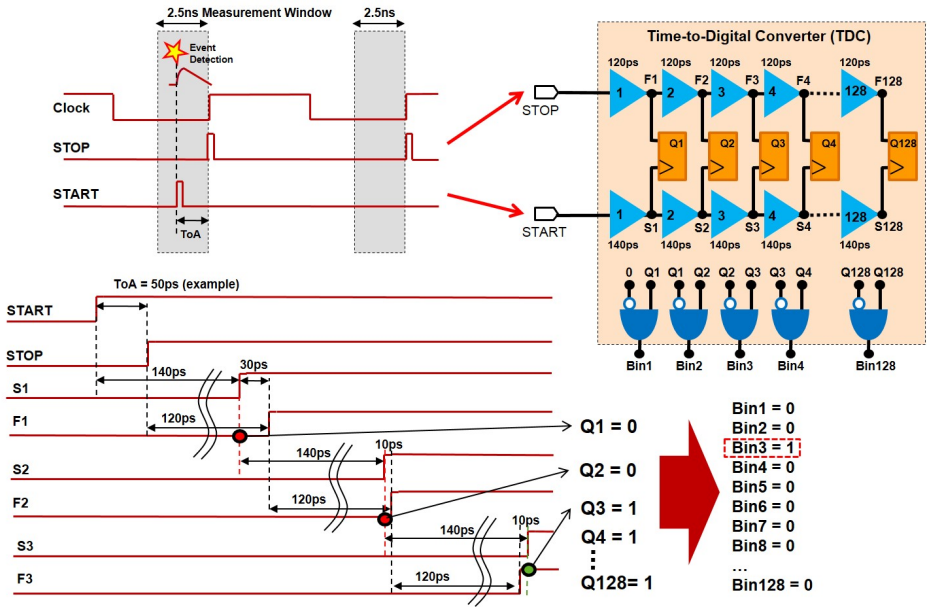


Figure 6.6: Graphic representation of the working principle of the TDC. The drawing on the top left shows how the START and STOP signals are generated, the first with the discriminator output upon event detection, the second corresponding to the next clock edge. The gray area indicates the 2.5 ns detection window. On the top right, the schema represents the TDC, with the 'slow' delay line (140 ps cells) that propagates the START signal, and the fast delay line (120 ps cells) in which the STOP signal is propagated. The difference between delays defines the bin. After each cell the signals are compared (QX), and the bin number provides the converted measurement.

them closer by an amount equal to the difference between the slow and fast cell delays, i.e. 20 ps. The number of cell stages necessary for the STOP signal to surpass the START signal represents the result of the time measurement with a quantisation step of 20 ps. A cyclic structure is employed to reduce the number of cells per line and results in a smaller occupied area. Since the time measurement is initiated only upon signal detection (instead of at each time-measurement window), the reverse START-STOP scheme is used as a power-saving strategy. The conversion time of a 2.5 ns input time interval is 21 ns, finishing before the next bunch crossing.

The TOT TDC provides a 9-bit digitization of the discriminator width, on a 20 ns range. It uses an additional coarse delay line made of 160 ps delay cells to extend the measurement range to 20 ns, while a Vernier delay line provides the requested fine resolution of 40 ps. The START and STOP signals are given by the rising edge and by the falling edge of the discriminator respectively.

As mentionned before, the delay cells of both TDCs, are implemented as differential shunt-capacitor voltage-controlled delay cells. Their delay is set by a control voltage (V_{ctrl}) that controls the load of the cell. Three control voltages are necessary to control the three delay

lines used in the TDCs : $V_{\text{ctrl\_fast}}$ to set the cell delay of fast cells to the desired value of 120 ps, $V_{\text{ctrl\_slow}}$ to set the cell delay of slow cells to 140 ps for slow cells and 160 ps $V_{\text{ctrl\_coarse}}$ to set the cell delay of slow cells to 160 ps. These control voltages are generated by three Delay-Locked Loops (DLLs) located in the periphery of the ASIC and built around a classical architecture (phase comparator and a charge pump current). Additional open-loop per-channel trimming is present in order to minimize timing-resolution between channels due to cells mismatches. Each DLL uses the very same delay cells as those used in the corresponding controlled delay lines : this ensures that the TDC steps (hence LSB) don't vary neither with PVT (Process Voltage Temperature) parameters nor under irradiations. Besides, as DLLs are locked onto the 40 MHz clock, the TDC steps (120 ps, 140 ps or 160 ps), only depends on the 40 MHz clock precision and of the number of delay cells, meaning that no calibration for both TOA and TOT LSB is needed. The only needed calibration is the one that gives the TOA versus the TOT in order to correct for the time walk. This calibration will be done using the internal pulser (described in Section 6.3.4) and physics events to have a reference for the time of arrival of the events. An internal phase shifter is then used to align events within the 2.5 ns acceptance window.

The TDC power consumption is dependent on the time-interval being measured. For the TOA TDC 2.5 ns (full dynamic range), the average power consumption over the 25 ns measurement period is about 5.2 mW. It will become 3.5 mW for the time-interval equal to half dynamic range. Thanks to the reverse START-STOP operation, the power consumption of the TDC is much lower in the absence of a hit over threshold. This results in an average power consumption per channel of 1.1 mW for both TDCs, assuming a time interval uniformly distributed (1.25 ns average) and a maximal channel occupancy of 10%.

6.3.4 Internal pulser

An internal pulser, common to all channels, is integrated to mimic input charges in phase with the 40 MHz clock. The pulser consists of a programmable DC current (tunable with an internal 6-bit DAC) that flows continuously through 50 kΩ resistor (R) until it is interrupted by a command pulse that shorts the resistor to ground (see Figure 6.7). A voltage step (Vstep) equal to $-R \times I_{DAC}$, is then generated and sent through the selected pixel internal 200 fF test capacitors (C_{test}) of the selected pixel. The input charge (Q_{inj}) is equal to $C_{test} \times V_{step}$ and the dynamic range goes from 0 to 250 mV or 0 fC up to ∼50 fC (LSB = 0.8 fC). The absolute value of C_{test} and R are known within 10% and its relative value between channels is within 1%. The pulser can be calibrated. The DC voltage ($R \times I_{DAC}$) is output on a dedicated PAD and so can be measured as a function of the 6-bit DAC. This PAD can also be used to inject voltages from an external generator. This PAD will be kept in the final ASIC allowing pulser calibrations during the test of the production chips.

This pulser will be used to intercalibrate the value of the phase of each channel (see Section 10.2) and also to align the thresholds of each discriminator. The command pulse

(encoded in the LpGBT fast command elink) is therefore distributed as a clock tree inside the ASIC. Since the absolute phase calibration should be measured by injecting a large charge (in order to make the time walk negligeable), there is no need to know the absolute value of the injected charge with an accuracy below 10%. The pulser will also be used to perform a first order time walk correction by measuring the TOA as a function of the TOT for various input charges. The final calibration will be done using physic events that give the reference of the time of arrival.

On the test bench, the pulser is used to measure the performance of the ASIC. The input signal allows also the characterisation of the front end read-out but does not reproduce the jitter performance when having an LGAD signal as input as the signal time duration can not be neglected. Figure 6.8 shows the post layout simulation of the preamplifier output using as input a simulated LGAD signal and a Dirac signal. For the same input charge, the simulation predicts a jitter larger by a factor 1.65 when using as input the LGAD signal instead of the calibration signal. This difference is mainly attributed to a diffference in the rise time. The impact on the amplitude is much smaller and a decrease of about 10% is predicted.

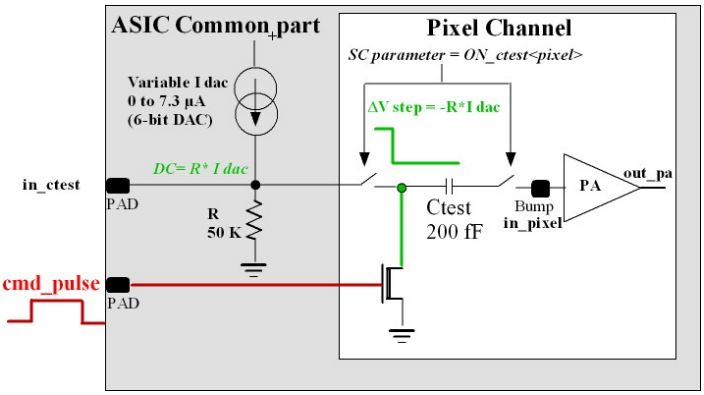


Figure 6.7: Pulser principle that shows the common 6-bit current DAC used to set the input charge as well as the pixel C_{test} capacitor.

6.3.5 Hit processor

Each electronics channel is composed of an analog part, already described, and a digital part. The latter is composed of three main blocks, as can be seen in the schematics of Figure 6.1. The hit processing unit, or hit processor, temporarily stores the data related to a hit and selects hits of events that have been triggered. The main circuit is the hit buffer which is composed of a memory of 1400 positions. Such size will allow to cope with trigger latencies of 35 µs, using one position per bunch crossing.

The size of each buffer position is 19 bits: 7 for the TOA, 9 bits for the TOT, 1 bit for the hit flag, 1 bit for detection error (CRC) and 1 bit for the TOA overflow. The hit flag bit indicates

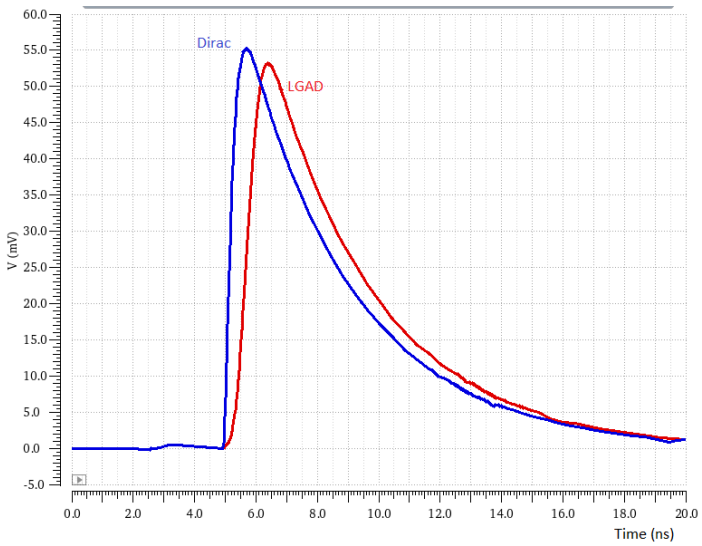


Figure 6.8: Post layout simulation of the preamplifier output using as input a simulated LGAD signal and a Dirac signal for an injected charge of 10 fC.

if a hit has been detected in the bunch crossing. The buffer is implemented as a circular memory in order to store data in a continuous way. It has two pointers for memory reading and writing. A control unit in the hit buffer increments the write pointer in one unit each bunch crossing. During data taking, the pointer goes from 0 to 1399 in scenario where the trigger latency is 35 µs and then goes back to position 0. If the needed latency is 10 µs, the write pointer is incremented from 0 to 399. The latency is set through a configuration register at the periphery. In each bunch crossing the control unit checks if a hit occurs. In case of a hit, the TOA and TOT measured by the TDCs in the analog front-end electronics stage are stored into the buffer and the hit flag of that position is set to 1. If not, the hit flag is set to 0 and no values are written in the TOA and TOT fields in order to save power.

The hit buffer architecture is built about a two-port SRAM design. This configuration allows simultaneous Read/Write operations within the same clock period. Six partitions of 256 words are used in order to limit the lines capacitance and to optimize the power consumption. The power consumption simulated taking into account parasitic elements and assuming a 10% occupancy with an L0 trigger signal at 1MHz is evaluated to be 1.3 mW at a temperature of 25°C and a voltage of 1.2V. This power dissipation decreases slightly down to 1.2 mW with a 2.5% occupancy. Concerning radiation tolerance, this SRAM architecture is less sensitive to SEU than DRAM as nodes levels are regenerated by the back to back inverters: ionizing radiations will significantly change the amount of charge on nodes but, assuming they don't completely flip the bits, the node levels will be restored to their normal value quite quickly, either by the feed-forward or by the feedback inverter. However, in order to improve the radiation tolerance, the memory cells are designed with large HVT transistors and with strong substrate/well contacts, sacrificing density for more robust and radiation tolerant

design. The full active area of the hit buffer is 720 μm$\times$1080 μm.

The reading pointer is handled by the next stage in the hit processing unit, the *trigger hit selector*. It reads the hit buffer as soon as it receives a trigger. The accessed position of the buffer is always the consecutive one of the latest written position in order to implement the latency. Only the output of the discriminator is checked. If it is high, it means that there is a matched hit and the TOT and TOA are temporarily stored in the matched hit buffer. They can remain there for longer times than the latency. The TOT and TOA data stored in this second buffer are tagged with a 5-bits identifier `TrigID` provided by the TDPU to indicate to which bunch crossing and trigger event they are associated to. The matched hit buffer operates as an average rate memory, storing the hits of triggered events until ready to be transferred. It will allow to cope with event-to-event fluctuations in the number of matched hits and to keep the bandwidth of the ASIC lower than 1.28 Gbit s^{-1}. It is implemented with a FIFO (first in first out), in which each position contains 21 bits: 16 for the TOA and TOT information, and 5 bits for the `TrigID`. The current design has a depth of 32 that could eventually be reduced in case simulations prove it possible. The writing of the data into the FIFO is done by the trigger hit selector block, while the readout is performed by the EOC logic by placing a requested trigger ID (`RqtTrigID`).

6.3.6 Luminosity processing unit

As already described before, the windowing process of the luminosity measurement is carried out on-channel, which is needed because of the large area of the chip. Transmitting the output of the discriminator to the luminosity block at the periphery would imply the use of a metal line of several millimetres. Such a long metal line would have large equivalent RC that would delay the signal by several nanoseconds. The length of each channel-to-luminosity block connection would vary from channel to channel and so would the delay. As a result, these delays might cause some hits inside one of the windows to be registered outside, corrupting the measurement of the luminosity. The compensation of the delay for each channel would be difficult. A simpler solution is to perform the windowing process on-channel. This avoids the need to transmit the output of the discriminator to the periphery. However, the windows must be distributed through the whole channel matrix. Again, long metal lines are needed but their delays can be compensated by distributing them as a clock tree.

A scheme of the first step in the luminosity measurement is presented in Figure 6.9. At the channel level, an AND gate evaluates if the output of the discriminator is inside the window. It generates a pulse that triggers a positive edge detector made of a flip-flop D with its D input connected to a logic '1'. When a positive edge is detected, the output of the flip-flop D goes high. This signal is asynchronous, so a synchronizer retimes the signal with a 40MHz clock. The output of the synchronizer is read out at each clock cycle and processed in the end-of-column logic.

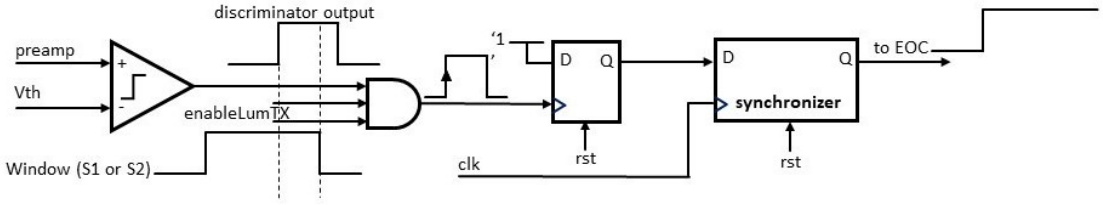


Figure 6.9: The signal of the discriminator is compared to the luminosity window (for each window) and a signal is transmitted to the end-of-column logic.

6.4 End of Column logic and digital blocks

This section describes the ASIC common digital part, including the readout process of the channels and the various blocks with specific functions.

6.4.1 Clock generation unit

The ASIC requires several clocks (see Figure 6.10):

- A 40 MHz clock (`clk40MHzInt`), necessary for the TOA TDC and for most of the digital blocks, including the I2C and the configuration registers
- A 80 MHz clock (`clk80MHzInt`), necessary to read out the timing data from the pixel matrix and to pack them into frames in the TDPU.
- Two 640 MHz clocks, one to serialize the data (`clk640MHzInt`), and one clock to generate the time windows for the luminosity measurement (`clk640MHzLumInt`).

A clock generator unit made of a PLL and a phase shifter (see Figure 6.11) provides these clocks while ensuring their phase alignment and their phase shifting. The 40 MHz clock input is not the one provided by the LpGBT but the one obtained from the fast commands and the LpGBT 320 MHz clock. This choice facilitates the phase alignment of all the necessary control signals used by the ASIC and also limits the number of e-links on the flex. As described in Section 6.2.2, the ASICs receive fast command signals from the LpGBT 320 Mbps e-links, along with the LpGBT 320 MHz clock. These fast commands, coded over 8 bits, contain several encoded control signals (such as the Level-1 trigger, the BCID, reset signal, 40 MHz phase…) as well as the command pulse that is necessary for the phase inter-calibration and the time walk correction (see Section 6.3.4). The 40 MHz clock input of the clock generator is obtained by dividing the LpGBT 320 MHz clock by 8. The clock divider also selects the 40 MHz phase that is encoded in the fast commands (see Figure 6.12). The phase aligned 40 MHz clock (`clk40MHz`) is then sent to the fast commands decoder so that all the decoded control signals are aligned on this very same phase. The PLL of the clock generator unit

uses this phase aligned 40 MHz clock as a reference clock and generates two phase aligned and jitter cleaned clocks (jitter <10 ps) : `PLL_40MHz` and `PLL_640MHz`. These two clocks are then sent to the phase shifter of the clock generator that provides all the necessary clocks for the ASIC: `clk40MHzInt`, `clk80MHzInt`, `clk640MHzInt` and `clk640MHzLumint`. A phase shifter is integrated to compensate for the cumulative latencies, in particular those related to the flex length.

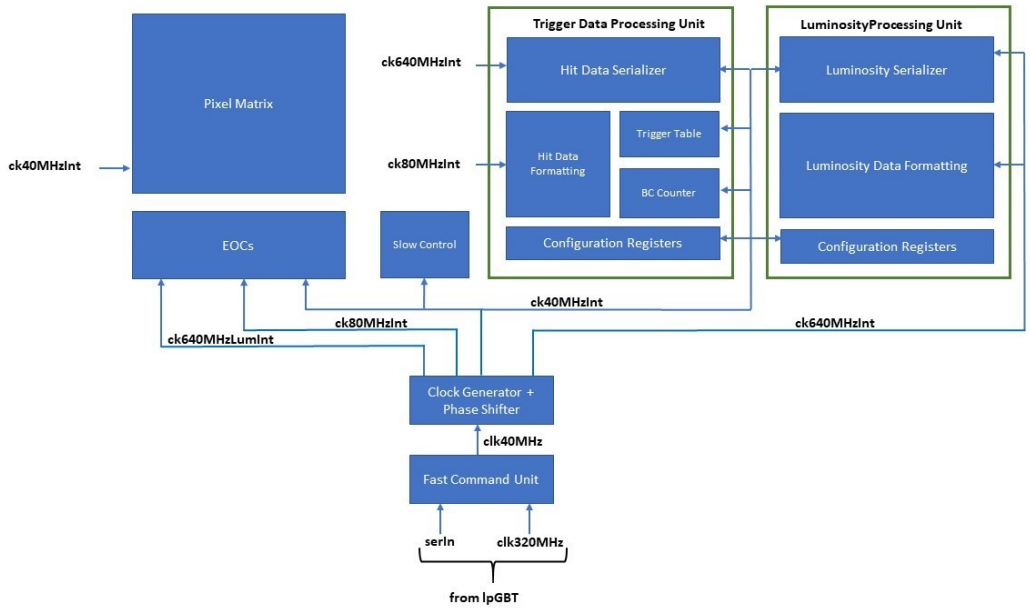


Figure 6.10: Schematic of the clock distribution.

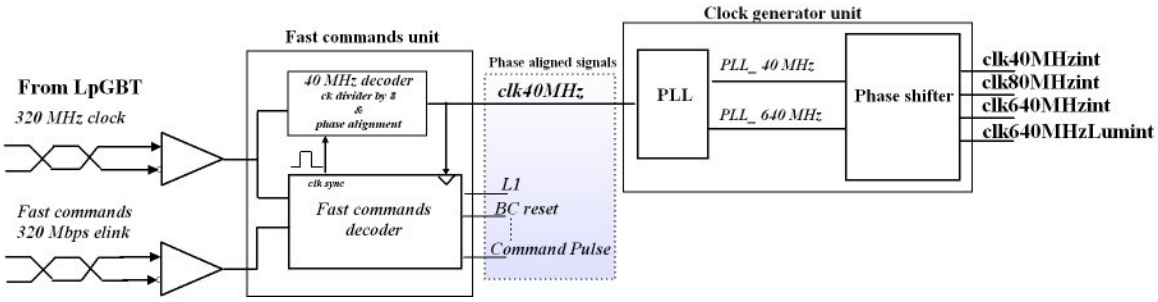


Figure 6.11: This schematic shows how a phase aligned clk40 MHz is extracted from the Fast command elink. The clock generator unit provides all the necessary clocks of the chip.

The phase shifter is also used to control the position of the 2.5 ns measurement time window of the TOA TDC compared to the bunch crossing (see Section 6.3.3). This is done by adjusting

the phase of the 40 MHz clock with 100 ps steps, keeping the jitter below 5 ps and a power consumption around 10 mW. The design is adapted from the one designed in CMOS 65 nm process for the LpGBT. As mentioned before, the ASIC needs two phase shifted 640 MHz clocks (`clk640MHzInt` and `clk640MHLumInt`). The core of the phase shifter is therefore composed of two delay-locked loops (DLL).

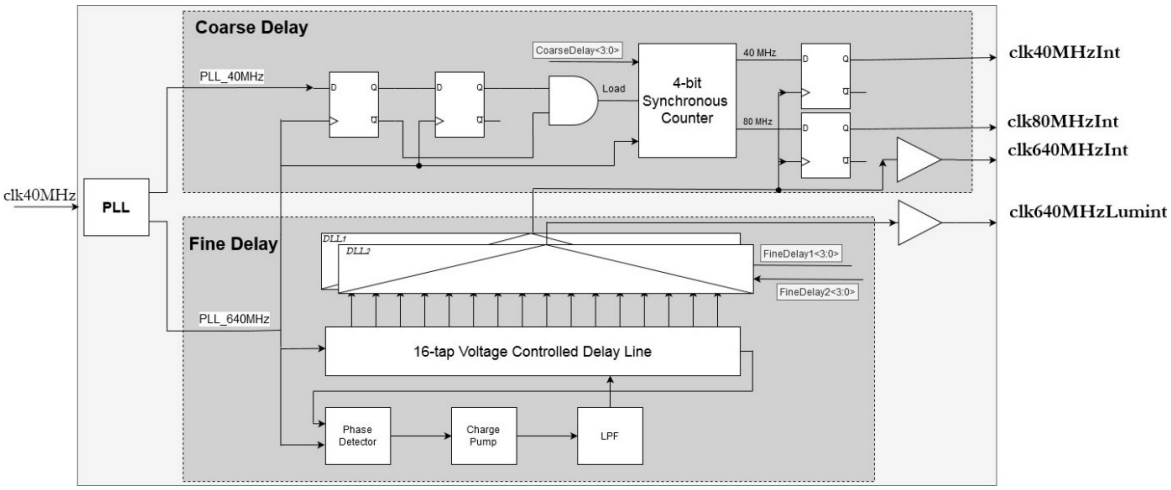


Figure 6.12: Simplified schematic of the clock generator made of a PLL and a phase shifter.

Each DLL integrates 16 delay cells and is fed by the 640 MHz clock provided by the PLL. The phase shift range is therefore 25 ns (16 * 1.562 ns), and the time shift is equal to 1/16 of the 640 MHz clock period i.e 97.6 ps. As for the 40 MHz and the 80 MHz clocks, coarse phase adjustment circuits are needed. Their output is then re-sampled by the `clk640MHzInt` clock Consequently, `clk40MHzInt`, `clk80MHzInt` and `clk640MHzInt` clocks are all aligned in phase and can be shifted compared to the `clk40MHz` clock with a step of 97.6 ps. All these clocks are distributed using clock trees in order to minimize clock skews and jitters.

6.4.2 Matrix readout process

The matrix readout consists of two processes, reading data from timing and from luminosity, each carried out by a specific module. The TDPU is responsible of handling the readout of the time data, and the luminosity processing unit of the luminosity data.Both blocks are depicted in Figure 6.3. Each readout process is described next.

Timing data readout

As described previously, in order to read out the timing information it is necessary to create a table that matches the BCID provided by the TDAQ system and the internal `TrigID`. The

TDPU has a 5-bits counter to tag the trigger events that are being received. The counter can be initialized with the fast command used to reset the chip. When the TDPU receives a trigger command, it stores the content of the counter together with the corresponding BCID into the trigger table and increases the counter by one unit. When the counter reaches the largest value, it wraps up to 0. The trigger table is a FIFO with 32 positions of 17-bits each one: 12 bits for the BCID and 5 bits for the `TrigID`. If the FIFO is full, an error message is generated and transmitted to TDAQ through an e-link. The TDPU unit also generates an internal trigger signal with a duration of one clock cycle. This is immediately transmitted to all matrix channels as well as the `TrigID`. Figure 6.13 shows a block diagram of the main signals involved in the TDPU and the EOC. Both, the trigger signal and the identifier are processed by the hit processor as described in Section 6.3.5.

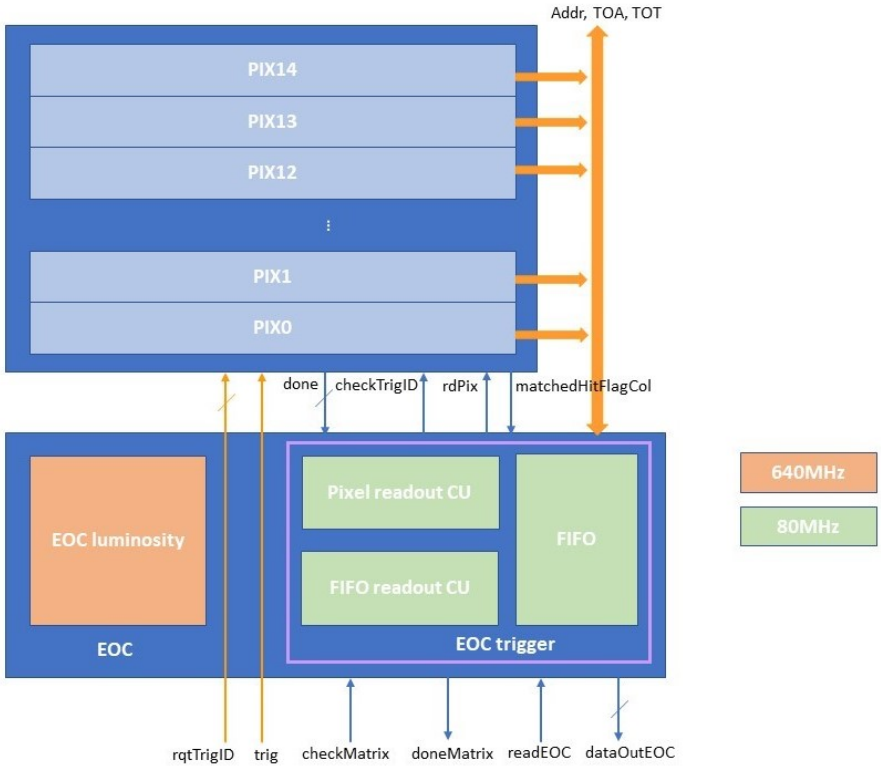


Figure 6.13: Block diagram of the main signals involved in the communication of the EOC with the pixels and the TDPU.

The hit data formatting unit in the TDPU continuously checks for an entry in the trigger table. When one is found in the TDPU, it fetches the entry and initiates the readout of the data stored in the matrix associated to that trigger event. The readout is carried out in two steps: first the retrieval of data associated to a given `TrigID` from the columns, and then the frame construction and data transmission. In the first step, the hit data formatting unit places the `TrigID` of the entry from the trigger table in the `rqtTrigID` bus and asserts the

checkMatrix signal to indicate to all the EOC to retrieve data from the pixels. Then the EOC asks to all the pixels to check if they have data associated to that `TrigID` by asserting the `checkTrigID` signal. The hit data processor checks if there is a matched hit with the same `TrigID` as the requested trigger. If there is, a hit flag is asserted. Once all the pixels have checked if they have data, then the EOC starts reading all the pixels that have such flag asserted, one per clock cycle. The row address, TOT and TOA of each read pixel are stored in a FIFO placed at the EOC. When the data of a pixel have been read and stored, the flag of that pixel is set to low. Once all the pixels have been read, the EOC indicates the completion to the hit data formatting block by asserting the `doneMatrix` signal. In the second step, the hit data formatting block starts reading the FIFO of the end of the columns since any data is available in the FIFO, not waiting for the readout to be completed. The column address is added to each FIFO entry and the data are placed at the dataOutEOC bus. The TDPU packs the data in frames and serializes them. Once all the buffers have been read and their data transmitted, the hit data formatting block waits for a new entry in the trigger table and the loop is executed again.

The design of the pixel matrix was done by trying to minimize the number of cycles needed to readout the data associated to a trigger event. For that purpose, the retrieval of data associated to a trigger event is carried out at column level so that each EOC works in parallel with the other ones. The search of data inside the matched hit buffer takes from one to few clock cycles. As mentioned before, it takes one clock cycle to move data from one pixel to the FIFO at the end of the column. Therefore, the readout of the column will take up to 15 clock cycles if all the pixels in the column have data associated with the requested trigger event. In the second step of the pixel matrix readout, data stored in the FIFOs must be passed to the TDPU where there are packed in frames and serialized. This second step doesn't wait for the completion of the data retrieval. It starts as soon as there is data available in the FIFOs, so it happens that data are being readout from one FIFO and passed to the TDPU while there are data being stored from the column to that FIFO. Therefore, data packaging and serialization start few clock cycles after the start of the data retrieval process. The FIFOs of the columns are not read in parallel but in series, that is, once the content of a FIFO has been fully read, then it read the content of the next EOC.

Luminosity data readout

The instantaneous luminosity, that is, the number of detected hits in the pixel matrix per bunch crossing, is measured every 25 ns. The process is carried out in three different regions of the ASIC as already described in Section 6.2.2

The windowing is performed per pixel. The two windows are generated at the EOC and distributed to the whole column as a clock tree in order minimize the skew from pixel to pixel. A trade off needs to be found between power consumption and skew. The first trials of physical synthesis show a skew of 100 ps. The windows are generated with a programmable

FSM running at 640 MHz. This FSM divides the bunch crossing into 16 equal intervals of 1.5625 ns with a 4-bits internal counter that continuously counts from 0 to 15. A control unit asserts and deasserts the two window signals called W1 and W2 as a function of the value of the counter and of the 4-bits parameters `minW1`, `minW2`, `maxW1`, and `maxW2` as shown in Figure 6.14. The width of W1 is fixed to 3.125 ns so the default values of `minW1` and `maxW1` are 1 and 14. However, both values can be modified in case it would be necessary to improve the luminosity measurements. The 1.5265 ns time resolution of the window generator is not enough to center the position of the windows with respect to the beginning of the bunch crossing. In order to provide the required resolution, the phase of the 640 MHz clock used by the EOC can be adjusted through the phase shifter. This 640 MHz clock (`clk640MHzLumInt`) is independent from the 640 MHz clock (`clk640MHzInt`) used in the serializers. More details are given in Section 6.4.1 The windowing process at pixel level is described in Section 6.3.6. Every pixel produces two measurements per bunch crossing. The EOC sums the luminosity measurements of the whole column per bunch crossing. Those measurements are passed to the luminosity processing unit which sums the measurements of the columns. The number of hits in W1 (S1) is subtracted from number of hits in W2 (S2). The result S2-S1 and S1 are truncated to 5 and 7 bits respectively. Both values are encoded with 6b8b code, producing a 16-bit frame per bunch crossing. Frames are serialized at 640 MHz. The whole bandwidth is occupied with the luminosity data. In order to avoid desynchronization, a synchronization frame needs to be sent periodically. This will be transmitted during the processing of a bunch crossing reset (BCR) fast command.

The readout of the EOC cells is performed at 80 MHz instead of the nominal operating clock of the rest of the ASIC which is 40 MHz in order to be able to encode data to 8b10b and keep the desired data transmission rate. Data encoding 8b10b can be disabled through the corresponding configuration register. The readout of the hit data can be adjusted through some configuration registers. These allow to enable/disable the readout as well as to select the transmission speed of the e-link between 320Mb/s, 640Mb/s and 1.28Gb/s.

6.4.3 Slow control

The slow control is used to configure the ASIC as well as to retrieve information of its internal status. For such a purpose, up to 1024 configuration registers of 8-bits each have been implemented in ALTIROC. The memory map is not yet completely determined, but the first 900 positions are dedicated to the configuration of the channel registers (4 per channel). The other 124 registers will be located at the periphery and will be used to configure the hit data transmission rate, enable/disable the luminosity block, to program the length of the windows used for the luminosity, etc ... For the final ASIC, the configuration registers are read/writen by using an I^2C link while shift registers are used for the prototype. The I^2C link in the ASIC is slave to the master in the lpGBT in the peripheral electronics, described in Section 9.1.1.

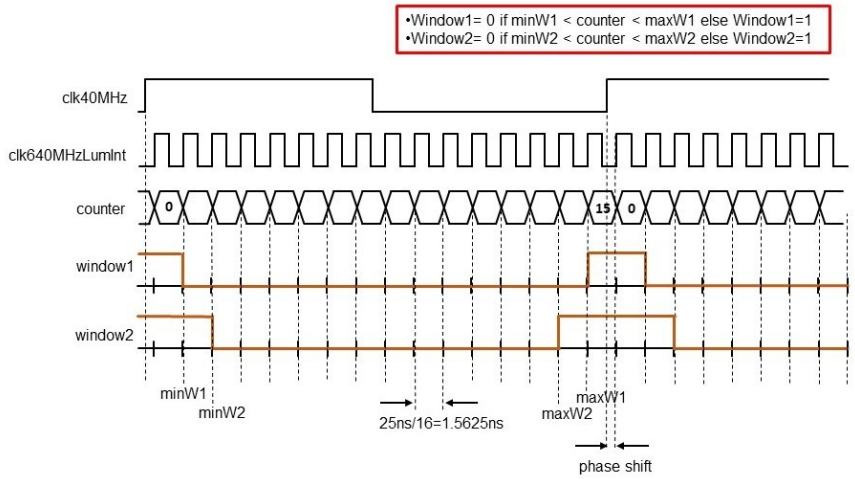


Figure 6.14: Generation of the luminosity windows W1 and W2.

6.5 Radiation tolerance

Two radiation effects must be taken into account: the TID that may degrade the timing performance, and the Single Event Effects which may corrupt the configuration registers and the time data. As the ASIC is designed in pure CMOS, it is insensitive to neutron irradiation. The worst expected TID and fluence are respectively 2 MGy and 2.5×10^{15} n_{eq} cm^{-2} taking into account the replacement of the inner modules every 1000 fb^{-1}. The ASIC has been designed using TSMC 130 nm technology that has been tested up to 4 MGy, i.e. two times above the requirement. Nevertheless, known strategies have been used in the ASIC design to mitigate the radiation effects. TID degrades the performance of MOSFET by increasing their threshold and generating leakage currents. To avoid these effects, bias currents of analog blocks are set to quite large values (> 20 μA) compared to the expected leakage currents and low voltage threshold transistors are avoided in current sources. In addition, minimum size transistors are avoided, for PMOS transistors in particular. At the layout level, substrate contacts are used to avoid latch-up. The DLLs of the TDC part are designed to take care of radiation, temperature and voltage variations inside the chip automatically. Besides, as the TDC bins are given by the difference of two delays, it ensures compensation for variations under irradiations.

As for the digital part and the SEU tolerance, Triple Modular Redundancy (TMR) will be implemented on critical parts of the 225 channel version (ALTIROC). Simulations of SEU using CERN tools will be performed to fully evaluate the effect of SEUs on the chip functioning.

6.6 Power distribution and grounding

To preserve the signal integrity and the jitter performance during periods of significant digital activity, the power distribution must be done carefully at the ASIC level. Each analog block (preamplifier, discriminator, TDC) is in a deep N-well powered and grounded with its own power and ground lines. All powers and grounds are therefore separated. Great care must be taken to reduce the resistance of the power lines, especially for the preamplifier power supply. The preamplifier Power Supply Rejection (PSR) has been simulated and found to be above 17 dB for frequencies up to 1 MHz and above 30 dB for high frequencies larger than 100 MHz, meaning that the noise from power supplies is attenuated by at least 17 dB. As for the digital blocks, they are in deep N-well or directly on the substrate. The connection between all the digital grounds and the substrate will be done at the flex level. In order to find the optimal solution, tests will be performed at the system level in order to decide whether the analog ground and digital ground (gnda and gndd respectively) should be connected at the module level or at the PEB level. The same is done for the power supplies: all the analog power lines (V_{dda_block}) are connected together at the flex level to a common V_{dda} and all the digital power lines of the digital blocks (V_{ddd_block}) are connected to a common V_{ddd}.

The power consumption of the ASIC has been estimated through both preliminary measurements and simulations for a 10% occupancy. Two operation modes can be distinguished: physics runs and calibration runs. In the latter only a 10% occupancy will be considered, so that the power consumption during calibration will not be higher than the maximum during data taking. At the single channel level, the preamplifier and discriminator give a power consumption of 1.0 mW, considering a drain current for the preamplifier of 0.6 mA. For each TDC, 0.55 mW has been estimated, while up to 2 mW have been allowed for the digital part (hit processing unit, clock and luminosity unit). This yields a total of 4.2 mW per channel. In addition, an estimated allowance of 250 mW for the common digital part seems reasonable, yielding a total power consumption per ASIC of 1.2 W.

6.7 Prototype performance

The performance on the first prototype version ALTIROC0 containing only the analog part of the single-channel readout (the preamplifier and the discriminator) can be found in [60]. In this section, the results concerning the second prototype ALTIROC1 are presented. This second version consists of a 5×5 pad matrix instead of 2×2, in which the digital components have been added to the single-channel readout. Two iterations of ALTIROC1 have been produced called v1 and v2. The second one, ALTIROC1v2, corrects issues found in the TDC, and only results from this iteration are presented here except the irradiation tests done with the first iteration. Among the 25 channels, only 15 channels corresponding to three columns have the readout as described in Section 6.2.1 with voltage preamplifiers. The two other columns are equipped with trans-impedance preamplifiers and their performance is not described in this document.

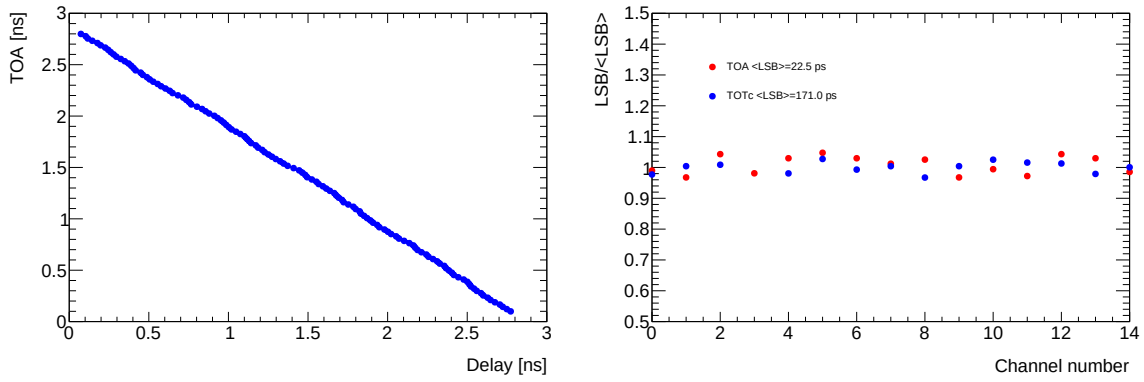
Section 6.7.1 describes the test bench measurements which were performed with and without a sensor bump bonded to it. More information on the assembly can be found in Section 7.2.2. In the case where no sensor is bump bonded, on channel 4 of each column, a capacitor can be connected through a programmable switch to the preamplifier input, mimicking the LGAD sensor capacitance and thus allowing to study the performance as a function of the detector capacitance C_d. The capacitance is tunable from 0 to 7 pF with a step of 1 pF. As described in Section 6.3.4, the test bench measurements are performed thanks to a C_{test} capacitor of 200 fF that is selectable by slow control together with a calibration pulser which generates a Dirac input charge with a relative precision between channels of $\sim 1\%$. All the measurements have been performed with only one channel activated at the same time. In order to understand the performance of the ASIC, an analog probe is integrated inside the prototype ASIC that allows to output the preamplifier signal to an oscilloscope. When this probe is enabled, the preamplifier output is not only sent to the discriminator but also to an amplifier with a gain of approximately 1.5. In a similar way, a digital probe allows to see the output of the discriminator, before going into the TDC.

Two test beam campaigns have been carried out during the year 2019 at DESY, in which data were collected with ALTIROC1v2, bump bonded to a non-irradiated 5×5 LGAD sensor. The main results are presented in Section 6.7.2. Irradiation tests were also performed at CERN using X-rays up to 3.4 MGy. The results are presented in Section 6.7.3.

6.7.1 Test bench measurements

The first step towards the evaluation of the full single-channel readout is the measurement of the TDC counts since the knowledge of the value of the LSB (Least Significant Bit) is needed to obtain the real values of the TOA and TOT. This is achieved by sending a square external trigger pulse, whose delay is adjustable in 10 ps steps, directly to the TDC inputs. This

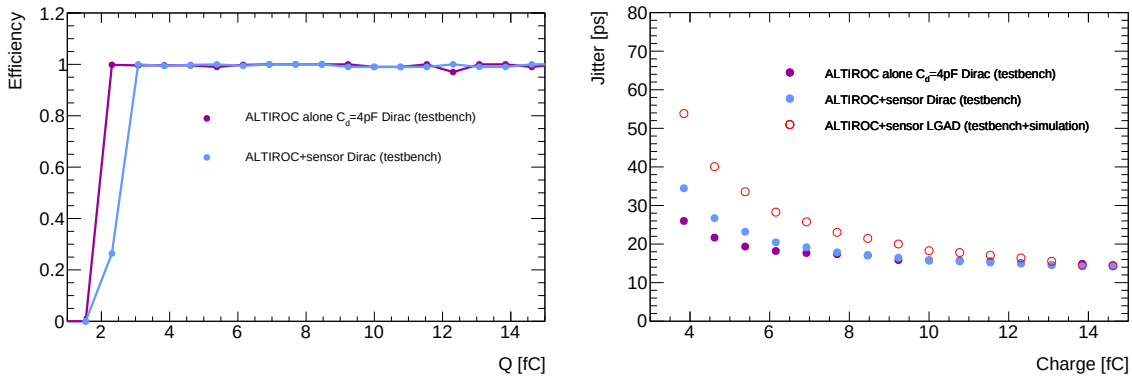
bypasses the preamplifier and discriminator, allowing direct measurements of the TOA as a function of the delay, as shown in Figure 6.15(a). The measured TOA TDC quantization step for the board that has been tested is found to be around 22 ps, slightly above the nominal value of 20 ps. As a consequence, the maximum TOA that can be converted is slightly larger than the nominal window of 2.5 ns. The uniformity of the LSB for the TOA is shown on Figure 6.15(b) and is better than 5%. The external trigger has a variable width, adjustable in 10 ps steps which can be used to measure the LSB for the TOT. The averaged measured LSB is around 170 ps close to the nominal value of 160 ps and the dispersions are better than 5% as can be seen on Figure 6.15(b). All these results are already close to the nominal but they can be further improved using internal TDC slow control parameters to adjust the LSB for each channel individually.



(a) (b)

Figure 6.15: Average Time Of Arrival measurement with the TDC as a function of the programmable delay (a) and channel LSB divided by the averaged LSB as function of the channel number for one ASIC (b). All measurements are performed with an external trigger. One point is missing for the TOT due to a faulty channel.

The preamplifier jitter σ_{jitter} depends on the preamplifier rise time, which depends on the drain current that flows into it. All the results below have been obtained with I_d =0.6 mA. At this point, the transistor enters in the strong inversion region, the gain increases only with the square root of I_d and, so the S/N doesn't increase significantly. Figure 6.16(a) shows the efficiency as a function of the input charge for an ASIC alone with $C_d = 4$ pF in order to mimick the detector capacitance and with an ASIC bump bonded to a sensor. In the later (former) case, full efficiency is achieved for charge greater than 3 fC (2 fC), which is below the minimal expected charge for irradiated sensors at 2.5×10^{15} n_{eq} cm^{-2} which is 4 fC. These measurements are performed for a Dirac signal which is 10% lower in amplitude than an LGAD signal after the preamplifier. Even taking this effect into account, the efficiency is still 100% for a charge of 4 fC. The difference between the two curves is attributed to noise which is about 30% larger for the sensor case.



(a) (b)

Figure 6.16: Efficiency (a) and jitter (b) measured as a function of the injected charge for an ASIC alone with $C_d = 4\,\text{pF}$ (purple) and with an ASIC bump bonded to a sensor (blue) measured with the calibration setup. For (b), the open circle shows the jitter for an LGAD input signal estimated from the calibration data and the simulation.

Figure 6.16(b) shows the jitter variation as a function of the input charge for an ASIC alone with $C_d = 4\,\text{pF}$ and with an ASIC bump bonded to a sensor. For large charge a constant jitter of about 15 ps is observed, which is attributed to the command pulser and clock jitter. Even without subtracting this constant term, the jitter is smaller than 30 ps for $Q_{\text{inj}} > 6\,\text{fC}$. The larger jitter with sensor is attributed to larger noise measured in the sensor case and also to a larger detector capacitance. Preliminary measurements estimate this capacitance being between 5 and 6 pF including the effect of interpad capacitance and bump bonds. The performance obtained with the calibration signal can't be transposed to an LGAD signal because the calibration signal is much faster. Based on the simulation, the jitter obtained with the calibration needs to be multiplied by 1.65 to reproduce the results obtained with an LGAD signal. Therefore, the jitter becomes smaller than 30 ps only for $Q_{\text{inj}} > 8\,\text{fC}$ as shown on Figure 6.16(b) and reaches ∼55 ps at 4 fC which is consistent with the requirements. Figure 6.17 shows the TOT as a function of the injected charge. As expected, the TOT increases monotically with the injected allowing to use it for time walk correction.

6.7.2 Test beam measurements

An ALTIROC1v2 ASIC bump bonded to a LGAD sensor array (HPK 3.1) was exposed in electron beam tests at DESY in the fall of 2019. The LGADs were operated with a bias voltage of 230 V, resulting in a MIP charge deposit of about 20 fC. For an accurate timing reference, a fast Cherenkov-light emitting quartz bar of $6\times6\,\text{mm}^2$ area transverse to the beam and 20 mm length along the beam, coupled to a Silicon Photomultiplier (SiPM) is used. The time resolution of this device was measured to be 37.6 ± 0.7 ps.

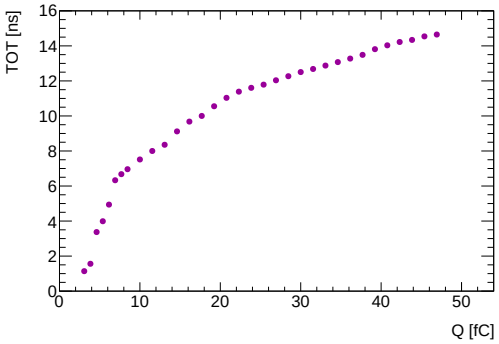
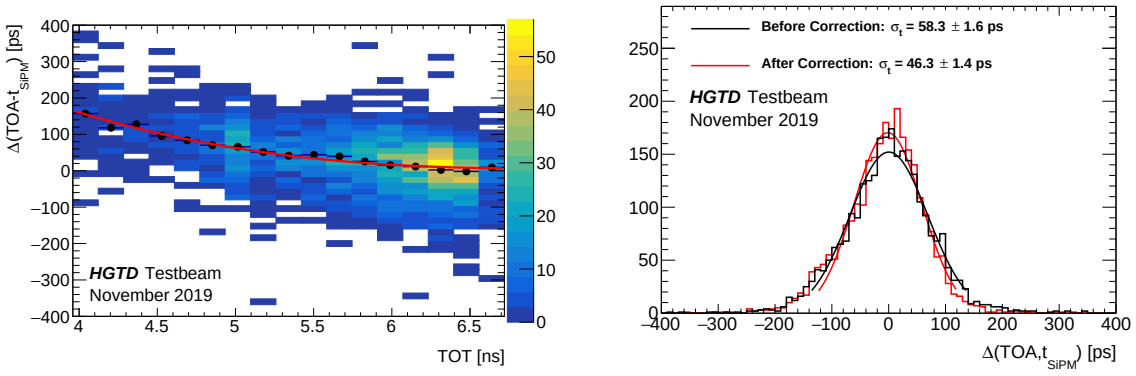


Figure 6.17: Time-over-threshold measured as a function of the injected charge.

Figure 6.18(a) shows the TOA variation as a function of the TOT. The range of the TOT is truncated since it was not possible to measure large values of TOT because of a coupling between the busy signal of the TOA TDC and the falling edge of the preamplifier output. This coupling only occurs when this signal is output on the PCB. This TOA busy signal must be output during test beam in order to synchronize the data from ALTIROC and the oscilloscope used to record the Quartz+SiPM system waveforms. This signal won't be needed for the HGTD and is not used for test bench measurements (it is only used for debug purposes). In the next iteration, ALTIROC1v3, the busy signal will be output as a differential signal to solve this problem. Therefore, with ALTIROC1v2, only a range of the TOT can be used in test beam and Figure 6.18(a) also displays a fit in this restricted range used for the time walk correction.

Figure 6.18(b) shows the time difference between LGAD+ALTIROC and the reference time from the Quartz+SiPM system before and after time walk correction extracted from the fit in Figure 6.18(a). The distributions are Gaussian without any tails. The measured time resolution decreases from 58.3$\pm$1.6 ps to 46.3$\pm$1.4 ps after time walk correction. Substracting the Landau contribution (about 25 ps), the remaining time resolution is about 39 ps containing contributions from the electronics jitter, TDC and clocks. This preliminary result is encouraging even though it is larger than the results obtained in test bench conditions (seeFigure 6.16(b)). One reason for the non-optimal performance of ALTIROC1v2 in test beam is due to larger noise coming from the FPGA board connected to the ASIC read-out board. An interface board is used to reduce this noise and an improved version was available in January 2020 (3 months after the testbeam) to further reduce the noise. Thanks to the new interface board, the jitter was reduced by 35% compared to the old version in test bench conditions. If the same improvement factor is applied to the test beam results, we would get about 26 ps instead of 39 ps. Testbeam campaigns are planned in 2020 to confirm this prediction. Moreover, it was noticed that the noise was larger in test beam compared to test bench conditions since it was not possible to use the same thresholds. For the next test beams campaigns, detailed investigations of the noise are planned to mitigate this effect.



(a) (b)

Figure 6.18: (a): Distribution of the TOA as a function of the TOT. The dots correspond to the mean value of the TOA distribution for a given TOT bin extracted from a Gaussian. The red line is a fit of the average TOA as a function of the TOT. (b): Distributions of the time difference between LGAD+ALTIROC and the Quartz+SiPM system before (red) and after (black) time walk correction together with Gaussian fits. The numbers are the fitted Gaussian widths where the time resolution of the Quartz+SiPM system has been substracted quadratically.

6.7.3 Irradiation tests

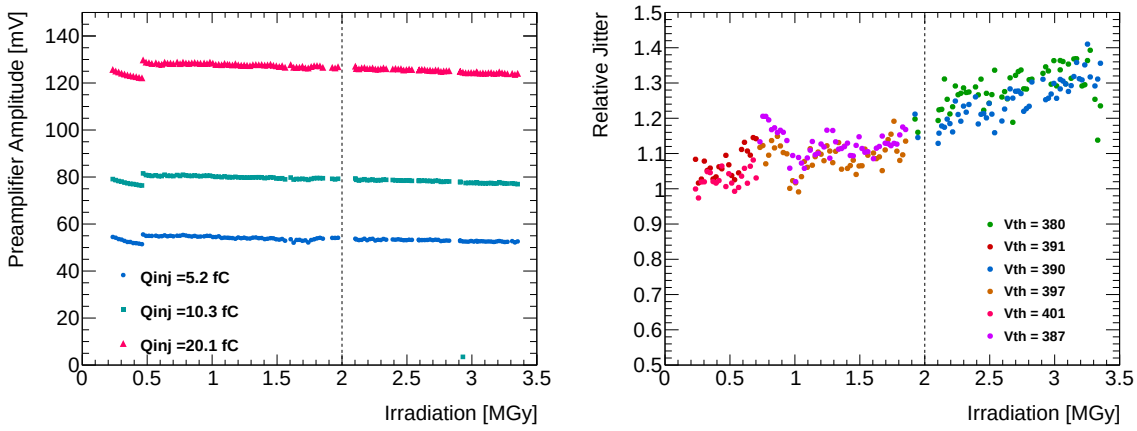
ALTIROC1v1 has been irradiated at the ATLAS-pixel CERN facility using an X-ray machine in July 2019 before the second version was available with improved TDCs. The very front end part of the two version (preamplifier and discriminator) are identical so this test has been focused only on their performance. As mentioned before, the ASIC has an analog probe that copies the output of the preamplifier, and a digital probe to see the output of the discriminator. During the irradiation, both were recorded with an oscilloscope in order to evaluate possible degradation of the signal caused by irradiation, that could be evidenced in the amplitude and the jitter level.

To allow debugging and performance studies, some voltage levels in the ASIC have been made accessible by connecting them to externally accessible points and a connector. The direct current level signals for the bandgap output, the V_{ddd} and V_{dda}, the general and individual channel V_{th} were recorded throughout the test.

The ASIC has been first irradiated in two periods, first with low dose rate (3.5 kGy/h) up to 0.23 MGy and then at higher rate (20.5 kGy/h) up to 3.4 MGy. No low-dose effects were observed after the first period. During the second, some of the DC levels corresponding to the bandgap output, the 10-bit DAC and 7 bit-DAC (used to set a common and the individual discriminator thresholds respectibely) show a drift smaller than 20 mV (over a typical amplitude of 800 mV).

The amplitude of the preamplifier signal has been monitored using the probe output and is diplayed in Figure 6.19(a) for different input charges. The decrease in amplitude is negligible up to 2 MGy. The amplitude measured after the full irradiation was only a few percent lower than at the beginning.

The measurement of the jitter in the rising edge of the discriminator signal is presented in Figure 6.19(b). A large level of noise was introduced by the data taking conditions, which is why the plot presents the relative increase in noise as the irradiation progresses instead of its absolute value (which was quite higher than what can be achieved in more controled test bench conditions). The plot shows a relative increase in the jitter level between 10 and 15% after 2 MGy. More tests with ALTIROC1v2 and more ASICs will be conducted over 2020, monitoring also the TDC outputs.



(a) (b)

Figure 6.19: Preamplifer amplitude (left) and relative jitter measured with the discriminator probe for a charge of 10.3 fC (right) as a function of the irradiation during high dose period. The dashed vertical line represents the maximal TID for HGTD. The step observed at 0.5 MGy is due to large temperature variations at the beginning of the measurement, which were subsequently controled.

6.8 Monitoring

6.8.1 Temperature monitoring

An additional requirement of the ASIC is to allow monitoring two closely related aspects of the LGAD: its operating temperature and its leakage current. While the electronics themselves are not very sensitive to temperature changes, it is of utmost importance to monitor the sensors in order to detect loss of cooling and thermal run-away, as explained in

Section 5.6. This information could also be used to estimate the particle fluence, since the current increases linearly with it.

A good estimate of the temperature dependence of the leakage current of a no-gain sensor is a factor 2 increase for every 7 °C. The temperature dependence of the gain is much lower, with an increase in gain of a factor 2 for a temperature decrease of 30–40 °C. Knowledge of the sensor temperature with an accuracy of 0.5 °C would make it possible to determine the leakage current to approximately 15% (while giving no relevant information on the gain). The modules will be operated at room temperature (20-30 °C) during the R&D phase, and during detector operation at about −30 °C as required by the sensor. Considering possible temperature shifts within the chip plus some margins, two monitoring ranges have been defined, [30- 40 °C] [-40 −10 °C], given a total temperature monitoring range of 80 °C. The target resolution to determine temperature variations has been set to 0.4 °C (9-bit resolution) independantly of the absolute value of the temperature.

The temperature sensor inside ALTIROC is based on a resistor which is sensitive to variations of temperature. A constant current, delivered by the current source present at the ADC input of the lpGBT circuit, flows through this device and produces a voltage drop proportional to the temperaturure Four different types of resistor proposed by the TSMC technology have been evaluated for their ability to perform temperature measurements in an irradiated environment. Since they all present similar behaviours, only the performance of the N-diffusion resistor version is reported in Table 6.3. In particular, a resolution after conversion was measured to be 0.8 °C per ADC count using a current of 100 μA. It should be noted that the resolution can be doubled using a current value of 200 μA instead of 100 μA, achieving then a resolution of 0.4 °C per ADC count.

Technology of the resistor	N+ diffusion resistor with salicide (rnlplus)
Value of the resistor	5 kΩ
Value of current flowing during test	100 μA
Sensitivity	+1.3 mV/°C
Variation of sensitivity with radiation	+15% at TID of 3.5 MGy
Shift of temperature with radiation	−0.3 °C at TID of 2 MGy
	−0.6 °C at TID of 3.5 MGy
Resolution after conversion with ADC of lpGBT	0.8 °C per ADC count

Table 6.3: Evaluation of a N-diffusion resistor as temperature sensor under irradiation (TID).

6.8.2 Supply voltages monitoring

The analog and digital supply voltages have also to be monitored in order to measure and compensate the voltage drops in the power lines caused by the parasitic resistances in the power wires of the flex cables (R_{FLEX} in Figure 6.20). The V_{dda} and V_{ddd} voltages are sensed

through dedicated wires on the flex and digitized by the ADC of the lpGBT circuit on the peripheral board.

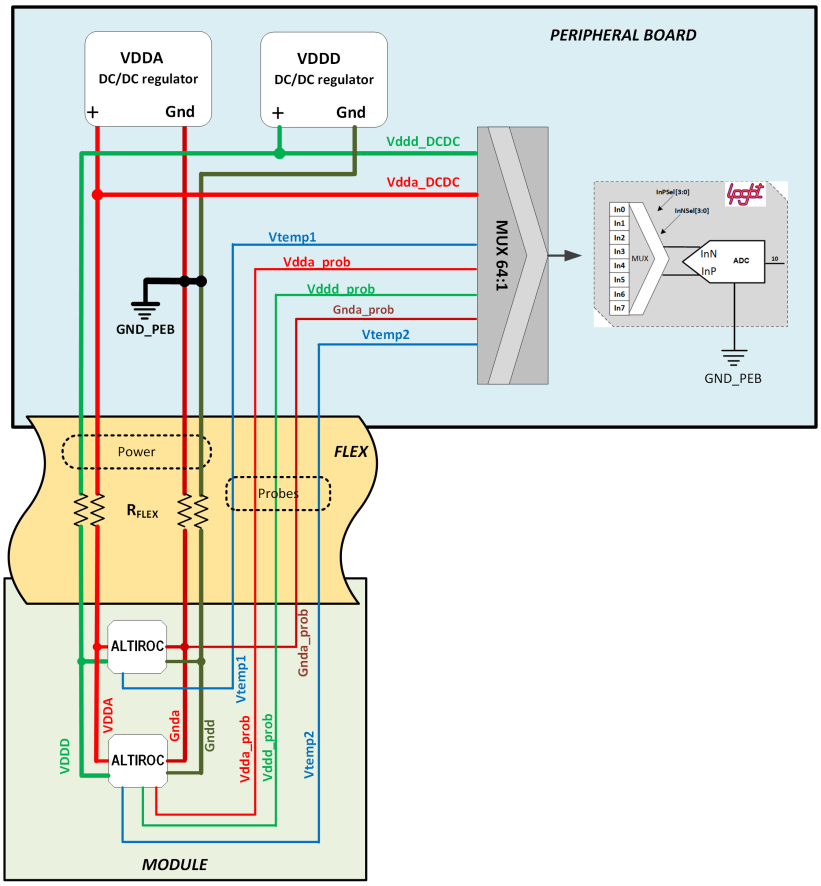


Figure 6.20: Complete schematic view of the voltage monitoring of a module using the ADC of the lpGBT circuit.

The probing of the power voltages at the module level is also useful to detect latch-up events on an ASIC. With the resolution of 1 mV of the lpGBT ADC and a parasitic resistance of 100 mΩ on the flex cable, minimal variation of 20 mA (considering an attenuation of 1/2 on the probing to respect the input dynamic range of the lpGBT ADC of 1V) can be detected, much smaller than the expected current rise in a latch-up event.

6.8.3 Complete monitoring system

A complete schematic view of the proposed monitoring of ALTIROC using the ADC of the lpGBT circuit is given in Figure 6.20. Three signals (V_{dda_prob}, V_{ddd_prob} and G_{nda_prob}) for the monitoring of the power supply voltages inside the two chips and two signals (Vtemp1, Vtemp2) for the measurement of the temperature inside the two ASICs are connected to the

ADC of the lpGBT circuit. The signal to be converted by the ADC is selected via multiplexers controlled through the I^2C interface of the lpGBT.

A view of the complete interfacing of a peripheral board with the modules is represented in Figure 6.21. The analogue signals for monitoring coming from the modules are digitized by the converter implemented inside each lpGBT circuit of the peripheral board. The number of channels of this ADC being limited to eight, a multiplexing is required at the input of each channel. Multiplexers (MUX 64:1) are thus implemented to interface the signals coming from the modules to the ADC on the peripheral board. With such multiplexer circuit, up to 8×64 signals can be interfaced to each lpGBT-ADC. With one multiplexer reserved for the signals coming from the DC/DC regulators, 7 mux are available to interface the monitoring signals coming from up to 84 modules, which is larger than the maximum number of modules expected per peripheral board. A full custom 64-to-1 multiplexing circuit has been developed in CMOS 130 nm technology and received in December 2019, and currently being characterised.

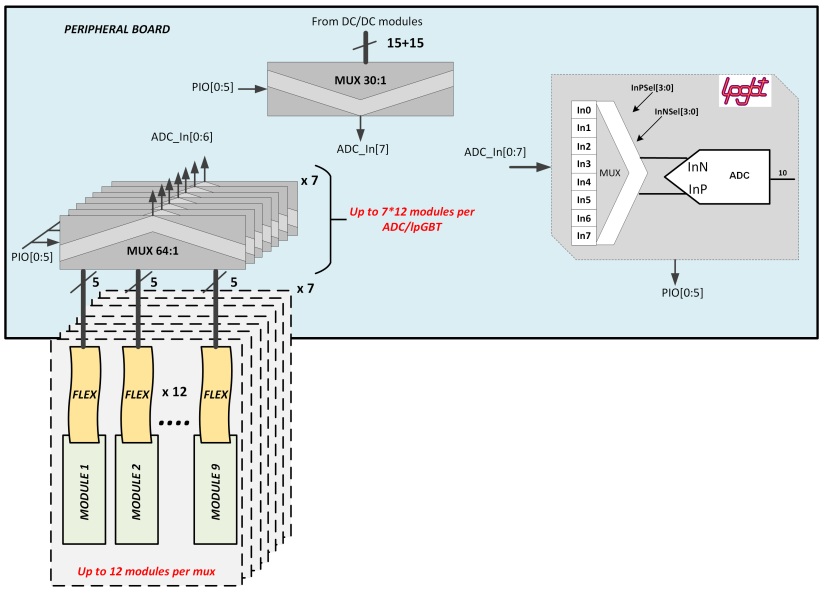


Figure 6.21: Interfacing of modules with a peripheral board for the monitoring.

6.9 Roadmap towards production

Two iterations of the ASIC have been produced : ALTIROC0 in 2018 with the preamplifier and discriminator with 4 channels, and ALTIROC1 in 2019 including the TDCs and the SRAM with 25 channels. While the intrinsic peformance of ALTIROC1 is quite good, an issue with the TOT measurements has been observed in test beam, which is attributed to a coupling between the busy signal of the TOA TDC and the falling edge of the preamplifier

output. In order to characterize the ALTIROC1 performance in test beam conditions, a third iteration of ALTIROC1 is needed before submitting the ALTIROC2 ASIC containing the 225 channels. Consequently the following R&D steps are envisaged for 2020:

- submit an ALTIROC1v3 to fully demonstrate that the TOT issue is solved for test beam by outputing the TOA busy signal as a differential signal (foreseen for Q2 2020). This version will also include a modification of the TDC and different grounding schemes for different columns.
- submit in parallel an ASIC with one complete pixel channel. This single pixel integrates the same front end as the one integrated in ALTIROC1 but also all the digital blocks (I2C, matched hit buffer, EOC, data formatting...) that will be in ALTIROC2. This ASIC will therefore validate the single pixel architecture and in particular the digital part. This ASIC can be submitted through an IN2P3 building block MPW run that is already financed in Q2 2020.

Assuming both ASICs work as expected, a Specification Review can take place in September or October 2020 before the submission of ALTIROC2. Taking into account the complexity of the chip, a second iteration of this chip is expected once the first prototype of ALTIROC2 has been intensively measured. A Preliminary Design Review would take place in Q2 2021 before the submission of ALTIROC3. The Final Design Review would take place in Q1 2022 before launching the pre-production. As ALTIROC2 should already be an engineering run, wafers of this ASIC will also be used to qualify the hybridisation process and module asssembly needed for the full demonstrator program which starts in Q3 2021. A summary of the key dates can be seen in Figure 15.4.

Regarding the MUX 64-1, the prototype is currently under measurements. Taking into account the small size of this ASIC, the aim would be to include it in the production of the LAr preamplifer to save cost (end of 2021). A joint Specification Review and Preliminary Design Review will take place early 2021 followed by a Final Design Review in 2021.

7 Module Assembly and Loading

7.1 Introduction

The basic component of the HGTD is the module. A detector module consists of a sensor bump-bonded to two readout chips which are in turn connected to a flexible printed circuit (FPC, flex cable) for communication, power distribution and data output. The flex cable also provides high voltage for the silicon sensor. The HGTD is made up of 8032 modules mounted on intermediate plates. This chapter describes the module and its assembly process, together with the procedure of mounting them onto the intermediate plates. Quality assurance and control plans are presented. Results of the fabrication of various prototypes are also discussed.

7.2 The bare module

The bare module consists of an LGAD sensor interconnected through solder bumps to two ALTIROC front-end chips. The LGAD sensors and the ALTIROC chip have been described in Chapter 5 and Chapter 6. In this section the hybridization process, called bump-bonding, is discussed.

Modules based on the 5×5 channel ALTIROC1 chip have already been fabricated and tested. A baseline hybridization process has been defined and the specifications agreed upon with two vendors. One of these vendors, as well as one HGTD institute, produced ALTIROC1 devices for which results were presented in Section 6.7. Full size prototypes will be produced as soon as the ALTIROC2 ASIC is available.

7.2.1 Bare module assembly

The LGAD sensor has a total size of 20.1 mm $\times$ 39.6 mm, with an array matrix of 15×30 1.3 mm $\times$ 1.3 mm pads and a dead region 0.3 mm wide around the active area. The readout ASIC has a total size of 21.7 mm $\times$ 19.9 mm and a matrix of 15×15 channels. The LGAD sensor has half the bump pads shifted from the central position by 250 µm, while the other half of the pads are shifted by the same distance in the opposite direction. This allows a

distance of 100 µm between ASICs, with no gap in the sensor coverage or disruption from different pixel sizes (see Figure 7.1).

The LGAD sensors will be produced on 150 mm wafers, whose thicknesses will depend on the wafer and sensor providers. The wafers will be thinned to the total sensor target thickness. Currently, the baseline for the active thickness is 50 µm and 300 µm for the total thickness. The sensors will be probed at wafer level at the fabrication sites and this information will be made available to ATLAS. The under-bump metal will be deposited on the sensors at wafer level, a necessary step before bump-bonding with solder bumps. After under-bump metalization (UBM), the wafers will be diced and the selected sensors will be destined for hybridization.

The ALTIROC2 ASIC will be produced on 200 mm wafers. The wafers will be thinned down to 300 µm (current baseline). The front-end chips will then be probed to identify the good dies. This will be followed by UBM and solder bump deposition. The relatively large pad size of the HGTD sensors enables a less demanding bump-bonding technology process compared to the ITk Pixel detector. The low-cost electroless deposition of Ni/Au can be used to treat the large pads (90 µm diameter) of both sensor and ASIC wafers. Solder bumps (SnAg) with a baseline diameter of 80 µm will then be deposited on the ALTIROC pads. A number of processes are available for the deposition of the bump balls, from solder laser jetting to electroplating. The most reliable, cost-effective technology will be selected.

After UBM and bumping, the sensor and ASIC wafers have to be diced into single tiles. The next step of the hybridization process is flip-chipping. During flip-chipping, the sensor and ASIC tiles are aligned, heated and compressed so that each solder bump melts and connects the sensor and readout channels of the two substrates. It is foreseen that the bare assemblies will then be processed in a fluxless formic acid reflow oven in order to improve the connectivity of the solder bumps. The final step consists in the inspection of the bare assemblies with a high resolution (sub-micron) x-ray machine to discard devices with disconnected pad bumps. Note that electrical tests of the HGTD modules will be carried out after the bare assemblies are mounted (including noise and charge collection measurements that can reveal disconnected bumps not apparent with x-rays).

7.2.2 First bare module prototypes: process and results

The first ALTIROC1 devices have already been assembled. A total of 20 bare modules were produced for different types of tests (mechanical and electronic) at Barcelona. As described in Chapter 6, the ALTIROC1 ASIC is a 5×5 channel prototype of the HGTD chip. The pad size is 1.3 mm $\times$ 1.3 mm. The corresponding 5×5 pad sensors used in these first prototypes were LGADs fabricated at CNM, in the context of an AIDA production (Run 11748), and at Hamamatsu (Type 3-1, EXX28995). Both vendors deposited the UBM on the sensors (at wafer level). In the case of CNM, a Ni/Au electroless process was used for UBM.

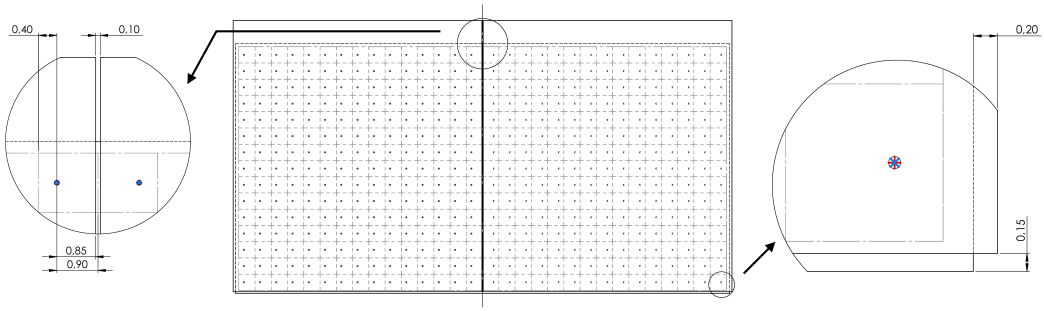


Figure 7.1: Sketch of the bare module (sensor and ASIC). Distances are in millimeters. The bump pads on the sensor are shifted by 250 µm on each side of the sensor (see magnified view in the left), to allow a 100 µm separation between the ASICs. A more detailed bare module drawing is shown in Figure D.1 and Figure D.2.

The Ni/Au under bump metalization was also deposited on single ALTIROC1 tiles by CNM through a chemical electroless process. SnAg solder bumps of 80 µm diameter were then placed on the chips using a laser jetting machine at IFAE. The bumps were prepared for flip-chip with a formic acid reflow cycle. The bump strength was verified to be lager than 60 gf per bump through shear tests.

The hybridization was performed by IFAE following the previous experience with the ALTIROC0 devices [60]. The same bonding cycle previously developed for the ALTIROC0 devices was used for the hybridization of the first ALTIROC1 bare modules. The devices were reflowed with no applied pressure and inspected with x-rays. Good alignment was observed as well as good connectivity in all the bumps (except those removed in one of the samples as part of the bump shear tests). CNM and HPK bare assemblies, along with the x-ray image of the bump connecting one of the readout channels, are shown in Figure 7.2. The topology of the bumps was found to be mostly cylindrical, with a diameter of about 90 µm and a height of approximately 50 µm. The hybridization specifications detailed below (see Section 7.2.3) follow the same process developed by IFAE, which is standard in the commercial sector and for which two companies have already been identified. One of these companies (in China), also produced modules with ALTIROC1.

The modules will experience thermal cycles during their lifetime, as the HGTD inner volume will be cooled with an input coolant temperature of −35 °C. In order to verify the robustness of the bare assemblies, they were subjected to a long burn-in test (some glued to a PCB using Araldite 2011, see Section 7.4.3). During a total of two weeks the modules were thermally cycled between −40 °C and 130 °C. The solder connections were then verified with x-ray imaging and shear tests were carried out on the modules. The devices were able to sustain the maximum applied shear force of 1000 gf, between the ASIC and sensor and also between PCB and ASIC. One device was verified to sustain a perpendicular (with respect to the plane of the sensor) pull test of 100 gf before and after the two week thermal cycling. Figure 7.3 shows the shear and pulling tests being carried out on an ALTIROC1 hybrid.

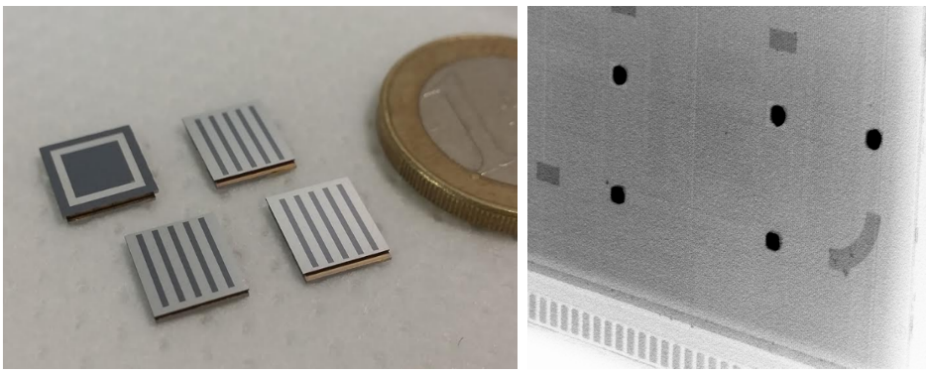


Figure 7.2: The first ALTIROC1 bare modules, with CNM and HPK sensors, and an x-ray image with a detail of a corner of one device are shown. In the x-ray image, the guard-ring solder bumps are in the periphery, while the bumps of two readout channels are visible in the center left of the image. The wire-bond pads of the ASIC are also apparent towards the lower part of the figure.

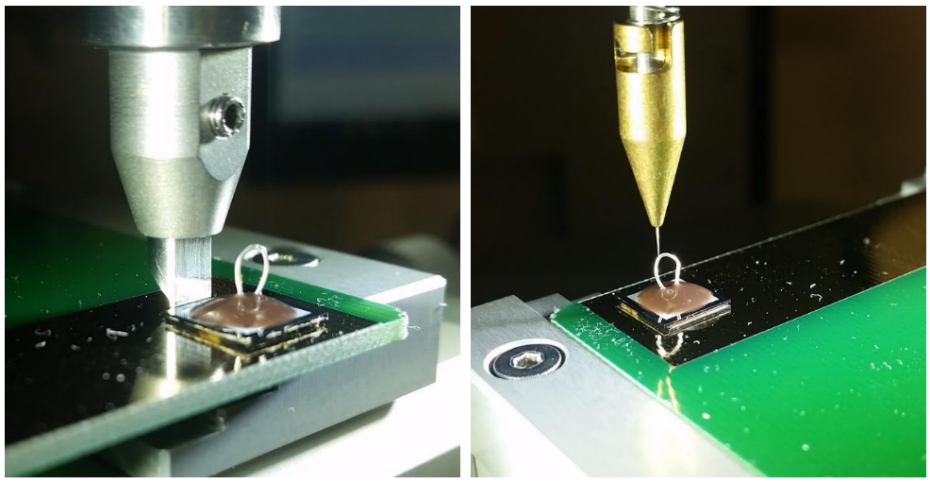


Figure 7.3: Shear and pull tests being carried out on an ALTIROC1 device. After thermal cycling during two weeks the device was able to sustain a maximum shear (pull) force of 1000 gf (100 gf).

The hybridization was also performed by the National Center for Advanced Packaging (NCAP China). NCAP has more than 3200 m^2 of cleanroom space and can provide bump-bonding services for 6 inch, 8 inch and 12-inch wafers. Its production capacity for module hybridization can fully satisfy the requirement of the HGTD project, this is also true for the other vendor identified in Germany. Fifteen bare module prototypes with ALTIROC1 have been hybridized in NCAP. Two of them are shown in Figure 7.4. The 5×5 pad sensors used in these prototypes were LGADs fabricated at Hamamatsu (Type 3-1 and Type 3-2), and at NDL (Type 6 and Type 12). The solder connections were then verified with x-ray imaging. The modules sustained a maximum applied shear force of 1000 gf during shear tests.

The performance of the bare module prototypes hybridized in NCAP have been evaluated

in the testbench measurements. A typical setup of testbench measurements is shown in the left photo of Figure 7.5. Bare module prototypes are glued on a printed circuit board (test board). The signal pads, power pads and debug pads of ALTIROC1 chip on the bare module are wire-bonded to the test board. The back side of the LGAD sensor in bare module prototype is wire-bonded to the test board for the high voltage connection. The electrical connections of each channel in the bare modules were checked by measuring the analog output level in each channel of the ALTIROC1 chip during charge injection tests. The results of the testbench measurements are described in Section 6.7.1. The performance of the bare module prototypes hybridized in NCAP were evaluated in electron beam tests at DESY in autumn 2019 (see Figure 7.5). The results are described in Section 6.7.2.

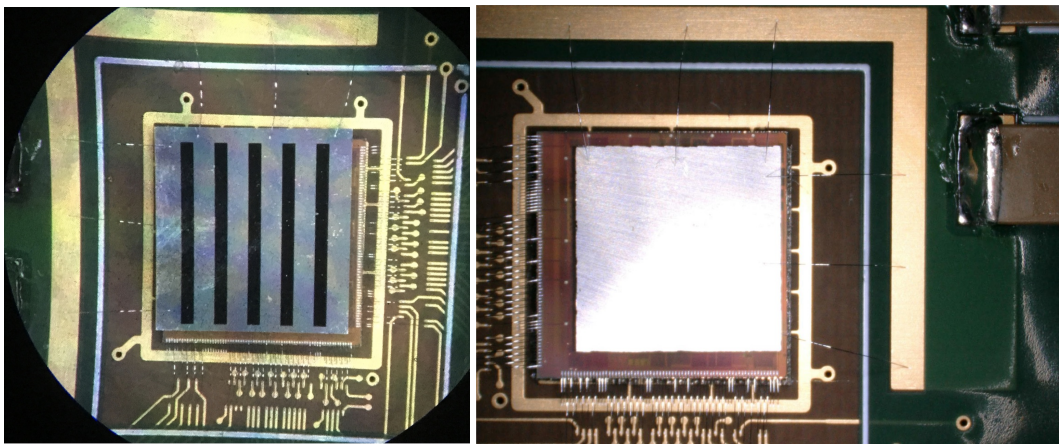


Figure 7.4: The bare modules hybridized in NCAP China. Left photo: The 5 × 5 pad sensors used in these prototypes were LGADs fabricated at Hamamatsu (Type 3-2). Right photo: The 5 × 5 pad sensors used in these prototypes were LGADs fabricated at NDL (Type 6).

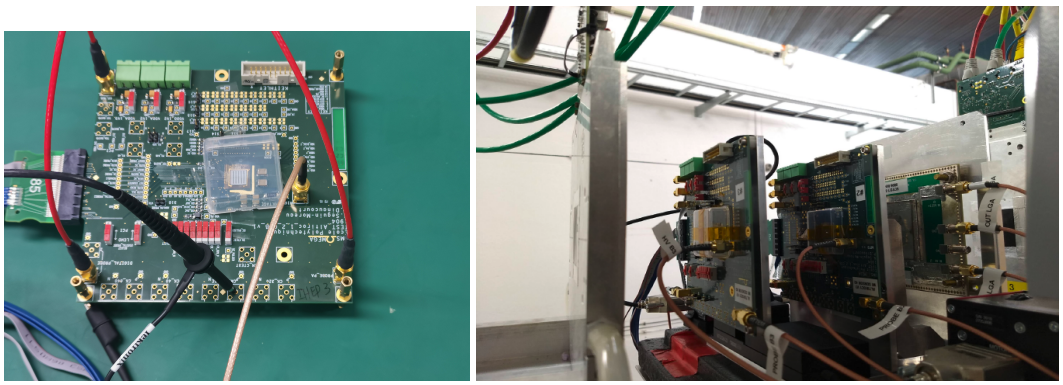


Figure 7.5: Left photo: A typical setup of testbench measurements for bare module prototypes. Right photo: The electron beam test setup for bare module prototypes at DESY in autumn 2019.

An alternative process explored during the initial R&D phase (but not intended for production) has also been developed to assemble ALTIROC0 devices. For Au bumps, the

bumps can be deposited directly on the aluminum of the front-end pads without under bump metalization. An alignment and thermo-compression cycle is used to interconnect the channels of the sensor and ASIC. Studies determined that the bump topology resembled at conical frustum with a base of about 140 µm and a height of 15 µm.

7.2.3 Hybridization specifications

The baseline bump-bonding technology for HGTD is solder bumps. As described above, both solder bump and gold bump prototypes have been produced at different HGTD Institutes and a company in China. However, the gold bumping process is not scalable for large productions. Thus, solder bumps are the baseline hybridization solution.

The sensor fabrication sites will deliver silicon wafers, which may or may not include under bump metalization, depending on the vendor fabrication capabilities and the overall HGTD hybridization strategy. In any case, it is expected that a fraction of the sensor wafers may have to be prepared for bump-bonding by a different vendor than the one producing the sensors. As explained above, given the large pitch and pad size of the sensors, the selected process for UBM is electroless deposition of Ni/Au. Table 7.1 lists the relevant parameters related to the sensor wafer UBM.

Wafer material	Silicon
Wafer thickness	300 µm
Sensor size (R×C)	20.1 × 39.6 mm^2 (15×30)
Distance between pads	1.3 mm
Pad size (passivation opening)	90 µm
Pad metalization	Aluminum
Scribe line passivated	Yes
Baseline UBM process	Electroless Ni/Au

Table 7.1: Specifications of the HGTD sensor wafer UBM.

The HGTD ASICs will be produced in TSMC CMOS 130 nm technology. In order to perform the hybridization process, first UBM and then solder bumps have to be deposited on the ASIC wafers. As mentioned above electroless Ni/Au deposition is selected as the baseline process for UBM, while solder bumps composed of SnAg (SAC305) would be deposited through a laser solder jetting system. However, other procedures can be considered. Table 7.2 summarizes the requirements for the UBM and bumping of the HGTD ASIC wafers.

After UBM and bump deposition the sensor and ALTIROC wafers will be diced. The width of the scribe line shall be 20 µm and the dicing precision ±10 µm. Break offs at the dicing edge shall be limited to less than 75 µm.

Wafer material	Silicon
Wafer thickness	300 µm
ESD sensitive	Yes
Passivation	8750A SiO_2
ASIC size (rows×columns)	21.7 × 19.9 mm^2 (15×15)
Distance between pads	1.3 mm
Pad size (passivation opening)	90 µm
Pad metalization	Al
Baseline UBM process	Electroless Ni/Au
Solder bumps	SnAg (SAC305)
Baseline bumping process	Laser solder jetting
Bump shear strength	40 gf/bump

Table 7.2: Baseline specifications of the HGTD ASIC wafer UBM and solder bump deposition. Other UBM and bumping process will be studied.

The flip-chip process is the final step in the hybridization procedure. The flip-chipping will be done on single sensor tiles. Two ASICs have to be flip-chipped to a sensor. The cycle has to be consistent with the SnAg solder bumps and result in a high hybridization yield. Table 7.3 summarizes the flip-chip requirements for the HGTD modules.

Requirements in the specifications have been fulfilled in small module prototypes with 5 × 5 pad sensors as shown in Section 7.2.2. Full-size bare module prototyping will be carried out to demonstrate the requirements can be met in the final design with 15 × 30 pad sensors.

Alignment between ASIC and sensor in X-Y plane	5 µm
Minimum distance between ASIC and sensor after flip-chip	20 µm
Maximum distance between ASIC and sensor after flip-chip	50 µm
Maximum failure rate per ASIC	0.044%
Shear strength after flip-chip	40 gf/bump

Table 7.3: Specifications of the flip-chip process for the HGTD modules.

7.2.4 Quality assurance / quality control

Modules will be tested according to the specifications. Bare modules will be optically inspected and weighed. The distance between the substrates (bump height) will also be measured. Inspection with x-rays for disconnected channels before module assembly (dressing with the flex hybrid) will follow. If the yield of the bump-bonding process is found to be high after the initial production and the modules are found to be highly uniform, these time consuming steps (x-ray inspection and substrate separation) can be performed only on a small fraction of devices. Note that the channel connectivity will be anyhow tested during

the module electrical tests. A small number of ASICs will be sacrificed to test the bump quality with shear tests before flip-chipping. Furthermore, a small number of devices will be tested destructively to verify the robustness of the hybridization process. Burn-in tests will be carried out on some devices to test specifically for the degradation of the module.

7.2.5 Production hybridization strategy

The total surface covered by the HGTD (6.4 m^2) requires a well planned approach to successfully carry out the hybridization of all the modules. The three step hybridization strategy consists of: process R&D and specification, search and qualification of bump-bonding vendors with full sized ASICs, and finally, module hybridization pre-production.

As presented above, the baseline bump-bonding process has been developed and successfully tested, both in Institutes and companies. Initial specifications have been established. Full size tests will be carried out as soon as the final sized sensor and ASIC become available. The specifications have already been provided to two companies (one in Germany and one in China) and discussions of an early qualification of the bump-bonding process with the currently available devices (ALTIROC1) was carried out in China. Both companies have expressed their willingness to carry out the hybridization service for HGTD and can do the full process in-house (metalization, bump-deposition, dicing and flip-chip). The estimated time for the hybridization process (sensor UBM, ASIC bumping and flip-chipping) for the full HGTD production in either company is only about 3 months. The target is to carry out the final hybridization qualification with two or more companies when the ALTIROC2 is avaliable.

7.3 Module design and assembly

7.3.1 General description

Baseline module design

The bare module described above is glued with accurate positioning to a small flexible printed circuit board (the module flex), to which a long flex cable (called flex tail in the following) will be connected during detector assembly (see Section 13.1). A more detailed technical drawing of the full module can be found in Figure D.1 and Figure D.2. ASIC signals and low voltage, as well as bias voltage for the sensor (HV) will be connected to the module flex by wire bonding. The flex tail with a length up to 69 cm connects, via two connectors, the module flex to one peripheral electronic board. Figure 7.6 shows three modules with the components stacked in the z direction of the HGTD. The total thickness of a module, including ASIC, sensor and module flex with all components and connectors, is 3.25 mm,

with the contributions of each element listed in Table 7.4. To allow for some tolerance, the maximally allowed total thickness for the module package is 4.2 mm (see Table 11.1).

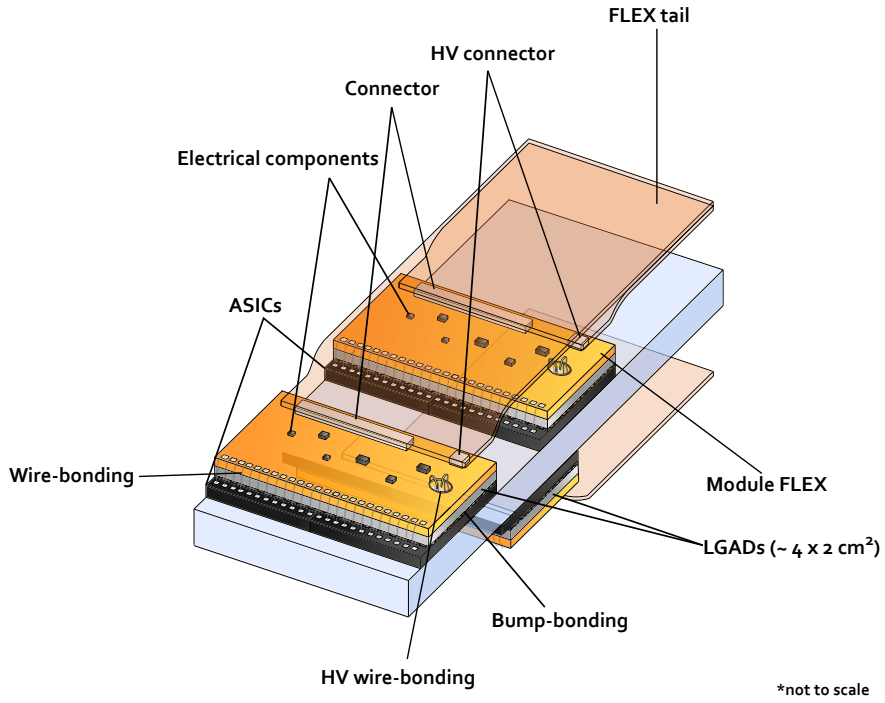


Figure 7.6: Schematic drawing of two adjacent modules on the top side and one on the bottom side of the cooling plate. A more detailed technical drawing of the full module can be found in Figure D.1 and Figure D.2.

Alternative module design

In addition to the development and test of the baseline design, alternative options are being investigated, in particular with the aim of replacing wire bonding with mechanically more robust solutions. In particular the usage of conductive glue for the connection of the HV line to the sensor and of bump bonds to connect all signal and power lines between the module flex and the ASICs is being studied and prototypes are in preparation.

7.3.2 Voltage distribution and signal readout: flex cables

Two cables based on the flexible electronics technology, a module flex and a flex tail (see Figure 7.6), will connect the signals from each bare module to the peripheral electronics board. The geometrical constraints on the flex tails are determined by the available space

Component	**Thickness** [mm]
ASIC	0.30
Bump bonding	0.05
Sensor	0.30
Glue	0.10
Module flex	0.50
Connector	2.00
Total	**3.25**

Table 7.4: Contribution of each module component to its thickness.

Signal name	**Signal type**	**No. of wires**	**Requirement**
HV	800 V max.	1	Clearance
POWER	$1 \times$ Vdda, $1 \times$ Vddd, 1.2 V	2	2 planes, $R < 2.7\,\mathrm{m\Omega\,cm^{-1}}$
GROUND	Analog, Digital	2 planes	Dedicated layer $R < 0.7\,\mathrm{m\Omega\,cm^{-1}}$
Slow control	Data, Ck (opt. + rst, error)	2 to 4	I^2C link
Input clocks	320 MHz, Fast command e-link (opt. 40 MHz (L1))	4 or 8	CLPS
Data out lines	Readout data (TOT, TOA, Lumi)	4 pairs	4 e-links differential CLPS
ASIC reset	ASIC_rst	1	Digital
Monitoring	Temperature, V_dda, V_ddd	4	DC voltage
Debugging	ASIC_debug	2	Analog

Table 7.5: Type and number of signal lines for two ASICs and one sensor included in the flex cable design

between two layers (see Table 11.1), the distance between the modules and the peripheral electronics and the maximum number of modules per readout row. Considering the harshest constraints, the module flex plus the flex tail must have a maximum length of about 690 mm, width of at most 36 mm, and thickness of less than 220 µm. The total length of flex cables in the HGTD, including both module flexes and flex tails, is 3000 m.

In terms of electrical requirements, one HV line has to be included in the design in order to bias the LGAD sensors (800 V maximum). The HV line must have a sufficient insulation resistance (IR) to not affect the other lines (IR $> 10\,\mathrm{G\Omega}$). The types of signals to and from the ASICs in each flex cable include the transmission of high speed signals ($1.28\,\mathrm{Gbit\,s^{-1}}$) as well as clock and power signals. The total numbers of signal lines for each module are listed in Table 7.5.

The mean impedance along the lines for all flex tail lengths is required to be in the range of $90\,\Omega$–$120\,\Omega$ for the differential pairs and of $50\,\Omega$–$65\,\Omega$ for the single lines. The impedance of the ASIC pins can be adapted according to the impedance of the tracks in order to minimise

signal reflections. The impedance variation between the lines of the same flex is much smaller than the expected difference for the flexes of different lengths according to the results of the measurements shown later in this section. The same radiation tolerance is required as for sensors and ASICs, i.e. up to at least 2.0 MGy, as well as operation at a temperature of about −30 °C (see Section 7.5).

A module flex with a width of 39.5 mm and a length of 18.5 mm along the readout row has been produced as a 4-layer stack-up with a thickness of 500 µm. A design of the module flex is being developed based on the ALTIROC2 pinout and design. Two connectors, suitable for the the connection of the flex tail, as well as surface mount components are considered. The schematics of the module flex prototype design can be found in Figure D.3. The flex tail is a 2-layer cable to be produced with different lengths, 220 µm thickness and a width of 36 mm. A preliminary layout of the flex tail is shown in Figure 7.7. A prototype of the flex tail has been ordered, while the design of the module flex is being finalised.

Two separate connectors, one specific for HV of the sensor and the other for all the signal lines to the ASIC, will be used to connect the module flex to the flex tail and trasmit the signals to the peripheral electronics boards. The Hirose F26 series provides good candidates from the geometrical and electrical point of view. The exact specifications on the maximum pressure that can be applied on the connector without damaging the modules and on the robustness against several connections and disconnections depend on not yet finalised details of the detector layout and qualification procedure.

To allow for the thermal expansion of the flex tails without mechanical stress on the module, studies on the possibility to place the connectors between two adjacent modules are being performed. This will allow for some bowing of the flex tail and some movement of the non-glued part of the module flex, taking advantage of the tolerance between the expected total thickness of the module (see Table 7.4) and the allocated space for it (see Table 11.1). Simulations will be used to study the expected behaviour of the flex tails as a function of temperature and tests will be performed with the demonstrator (see Chapter 14).

In the baseline design, the functionality of the flex tail for up to six modules per readout row in the outer ring (2688 flex tails in total) will be integrated directly in the PEBs, while the remaining 5344 flex tails will be connected singularly to the PEBs with a connector for HV and one for all the other lines (see Figure 9.8).

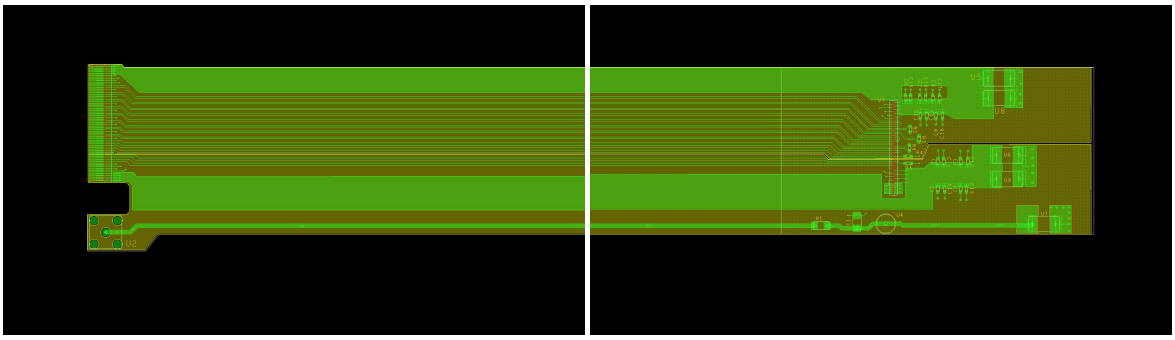
The specifications of the module flexes and flex tails are summarized in Table 7.6. Most of them are common to both kinds of flex cables, the specific ones are indicated explicitly.

Prototype characterisation

As part of the initial study phase, before defining the current baseline design, a prototype combining module flex and flex tail into one L-shaped 4-layer design has been produced

Tolerance in length	1 mm
Tolerance in width	100 µm
Flex tail maximum thickness	220 µm
Module flex maximum thickness	500 µm
Insulation resistance of HV line	10 GΩ
Maximum resistance of power planes	2.7 mΩ cm^{-1}
Maximum resistance of ground planes	0.7 mΩ cm^{-1}
Impedance of single lines	50 Ω–65 Ω
Impedance of differential lines	90 Ω–120 Ω
Maximum allowed BER	10^{-12}
Radiation tolerance	2 MGy
Neutron fluence	2.5x10^{15}neq/cm^2

Table 7.6: Specifications of the flex cable.

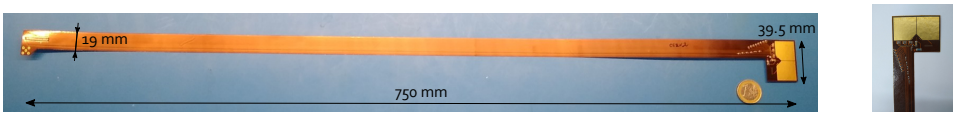


(a) (b)

Figure 7.7: Layout of the FLEX tail two-layers design prototype. (a) Top (gold) and bottom (green) layer view of the connection region of the prototype to the adapter board used to inject signals simulating the ones from the module. Two separate connections are foreseen: Zero Insertion Force type on top and a through-hole type on bottom for HV. (b) Top (gold) and bottom (green) layer view of the testing region of the prototype footprints for components are included for tests.

with the aim to understand the technical requirements, such as materials, manufacturing capability, electrical and mechanical robustness, and to address any potential problem by representing a significant subset of the signals (signal integrity, power distribution, HV insulation, interference and crosstalk). The design used for this prototype is considered inconvenient for assembly and will therefore not be further considered, however the obtained results are expected to not depend significantly on the details of the shape and design of the flex cable. The direct interaction with the CERN PCB Service allowed optimisation of the manufacturing process leading to the production of four prototypes of 750 mm length as depicted in Figure 7.8. Upilex-VT from UBE Industries was the commercial dielectric material chosen for this prototype. The length was chosen based on a previous version of the detector layout and it is 6 cm longer than the longest flex tail to be produced for the

HGTD.



(a) (b)

Figure 7.8: A 4-layer flex cable prototype from CERN PCB Service. (a) Top view. (b) Assembled extremity.

The stack-up of the cable has layers numbered 1 to 4 from top to bottom. On the top layer the single lines are routed following a micro-strip configuration. The differential pairs as well as the HV line are placed in layer 3 in a stripline configuration in order to improve the shielding of these lines. Layers 2 and 4 are full planes dedicated to powering and grounding. In order to perform the electrical tests the four flex cables have been assembled with all the foreseen components.

The qualification of the flex cables has been performed both at room temperature and in a climate chamber reproducing the operating conditions of the HGTD in terms of temperature (see Section 7.5), yielding very similar results.

Geometrical tests The thickness and the width of the flex cable must be uniform along its length to assure the proper assembly of the full detector according to the requirements in Table 11.1. Several measurements of the thickness and width of the cables were performed with a caliper every 5 cm. The mean values and standard deviations of the measurements are shown in Table 7.7. The spread of the values is well within the tolerance and the length and width average values are compatible with the nominal ones (see Table 7.6). A smaller thickness than the nominal one is acceptable from a mechanical point of view, as long as the electrical properties are not affected.

	Length [cm]	Width [mm]	Thickness [µm]
Nominal	75	18	340
Measured	75.0 ± 0.2	17.99 ± 0.04	300 ± 9

Table 7.7: Mean values and standard deviation of the measured length, width, and thickness for three long flex cables. Nominal refers to the now outdated specifications used for the design of the tested prototype.

Power integrity A simulation of the voltage drop in each plane was performed with the Cadence Allegro Sigrity PI software package [80] and the expectation for the longest CERN prototype (750 mm) was estimated and compared with multimeter measurements (see Table 7.8). The uncertainties on the simulation values reflect the uncertainty on the

thickness of the layers in the production process as indicated by the CERN PCB design service, while the measurement uncertainty is given by the spread of the measured values from the available prototypes. Similar simulations for the current baseline design of the flex tail (also shown in Table 7.8) are well within specifications for all power and ground planes.

plane type	CERN sim. [$m\Omega\,cm^{-1}$]	CERN meas. [$m\Omega\,cm^{-1}$]	tail baseline sim. [$m\Omega\,cm^{-1}$]
analog power	1.6 ± 0.1	2.0 ± 0.1	1.3
digital power	0.30 ± 0.01	0.4 ± 0.1	1.7
analog ground	1.5 ± 0.1	1.70 ± 0.03	0.5
digital ground	0.30 ± 0.01	0.35 ± 0.03	0.5
length [cm]	75	75	69

Table 7.8: Simulated and measured resistance of the analog and digital power and ground planes for the CERN prototype (75 cm length). Simulated resistance for the baseline design of the flex tail (69 cm length)

Considering the results in Table 7.8, the simulation is reliable within 30%. The differences between the analog and digital planes in the CERN flex prototype are due to the choice to consider a lower number of analog signal versus digital lines. Therefore, the surface of the power and ground analog planes is smaller than that of the digital ones leading to a larger resistance. In the flex tail baseline design the total number of signals listed in Table 7.5 were considered. The ground planes are designed symmetrical and placed in a dedicated layer. The power planes are placed together with the rest of the lines in a dedicated layer. The number of digital signals is higher than the number of analog signals, leading to less space available for the digital power plane and consequently to a slightly higher resistance than for the analog one. Simple geometrical calculations easily reproduce the ratio of resistances.

For the total power consumption estimation, the total length of the flex tails in addition with the 2 cm length of the module flex was considered. The same resistance is assumed for the flex tail and for the module flex and set to the maximum allowed values from Table 7.5, including additional resistance from the connections between the two parts. This estimate is higher than the current simulation results, because the simulation does not include the connection between module flex and flex tail and to allow for a possible mismatch between the simulation and the actual measured resistance. The total power consumption estimated is 2 kW over the whole detector (see Table 11.2).

Insulation test The insulation of the flex materials was checked for voltages up to 1 kV with the CAEN DT5521HEN HV power supply [81] that can measure currents as small as 500 pA. Since no current was observed over a long time, a lower limit was set on the insulation resistance at 2000 GΩ, well above the requirement. For this test no signals were injected in any of the lines of the flex cable. The possible interference between the HV and the other lines has been tested. The bit error test, described below, as well as the eye

diagrams performed with 1.25 Gbit s^{-1} signal transmission did not show difference while the HV was delivered (1 kV at 1 mA). Two flex cables were stacked while a 1.25 Gbit s^{-1} signal was transmitted in one of the cables and HV was delivered through both. No errors were observed during the bit error tests. Eye diagram results were not affected by the HV delivery.

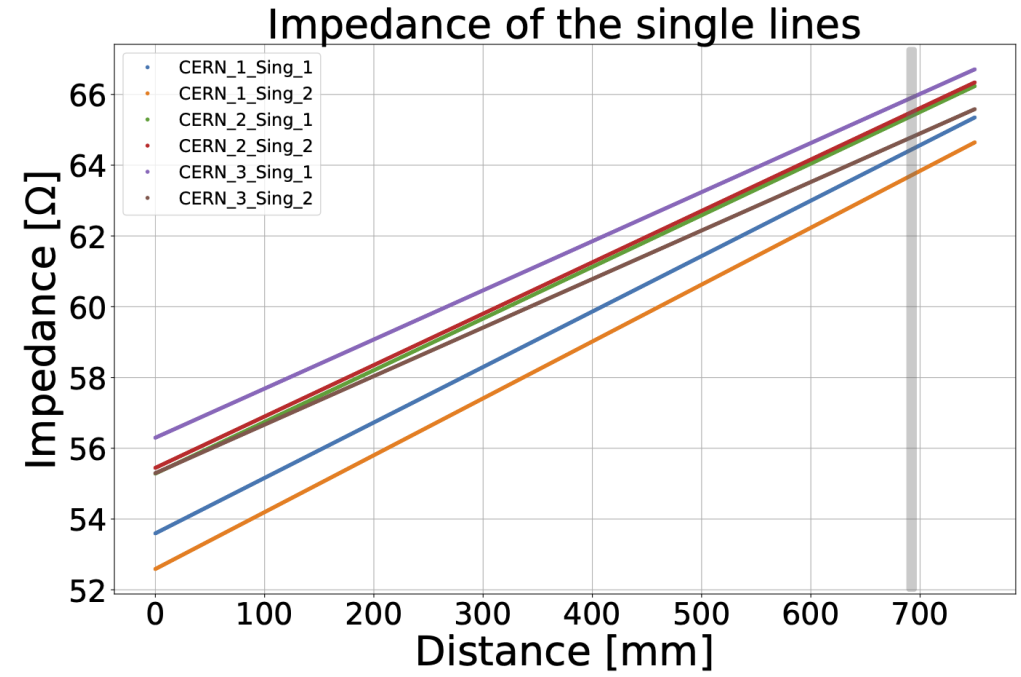
Time Domain Reflectrometry The Time Domain Reflectrometry (TDR) test is performed in order to check the impedance homogeneity of the tracks, which is crucial for high-speed data transmission. Three assembled flex cables were used to measure in each one three differential pairs (one dedicated to clock transmission, two for e-links) and two single lines that are accessible from a custom adapter board that was designed for the purpose of these tests. The TDR module 80E08 together with the DSA8200 oscilloscope by Tektronix [82] was connected through SMA connectors to the adapter board. The impedance of the lines was measured and compared with the impedance estimated from simulation. For all the measured lines the mean impedance along the cable is found to be well within the specifications for the foreseen lengths up to 69 cm (see Figure 7.9). Assuming perfect linearity and approximating the mean value over the cable length by the measurement at a distance of 34.5 cm, the observed range of impedances is about 58 Ω–61 Ω for single lines and 105 Ω-108 Ω for differential pairs.

Integrated Bit Error Test (IBERT) and eye diagrams To emulate the signals from the ASIC an FPGA on a Kintex KC705 evaluation board [83] has been programmed and connected to the flex cable via the adapter board also used for the TDR measurements.

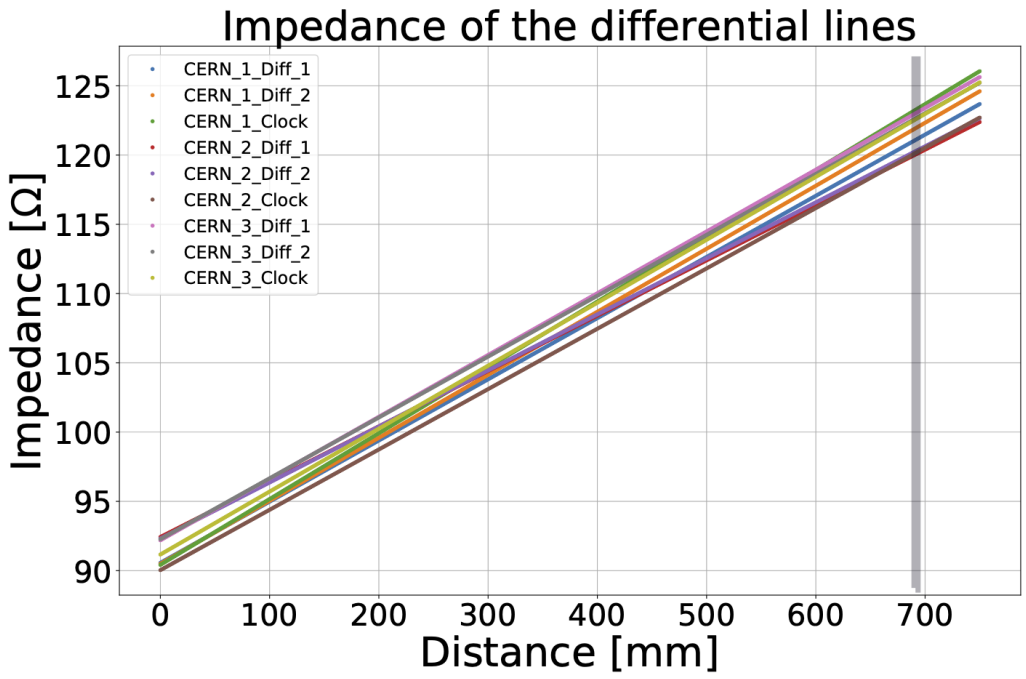
The FPGA injects test patterns at 1.25 Gbit s^{-1} and checks the response with the Integrated Bit Error Rate Test (IBERT). The SMA connectors placed on the adapter board route the signals to the oscilloscope for classical eye-diagram analysis. A wire bond between two differential pairs at the end of the flex cable creates a loopback path for the signals. Therefore, the transmission length of the signals is twice the length of the cable, 150 cm. The test configuration and the I/O drivers are compatible with the VC707 FPGA used by the lpGBT system. In this way we ensure the same conditions as for on-field operation.

The IBERT detected no errors over a few days, yielding a limit at 95% confidence level on the error rate for one of the flex cable prototypes at 1.25 Gbit s^{-1} with BER less than 1×10^{-15}. This value is much better (lower) than the acceptable error rate of 1×10^{-12} (see Table 7.6).

The same test was repeated with the HV up to 1000 V at 1 mA and showed no error for eight days at room temperature and at −30 °C. The BER result obtained during this test was no more than 1×10^{-15} at room temperature and 2×10^{-15} at −30 °C at 95% confidence level. The Kintex KC705 evaluation board encodes the signals at the receiver after an equalization stage. The signals were measured prior to the equalizer by an oscilloscope. The signal amplitudes range from ±100 mV to ±200 mV. The eye diagrams in Figure 7.10 measured without HV (a) and with HV (b) show a similar shape and opening area. The opening areas

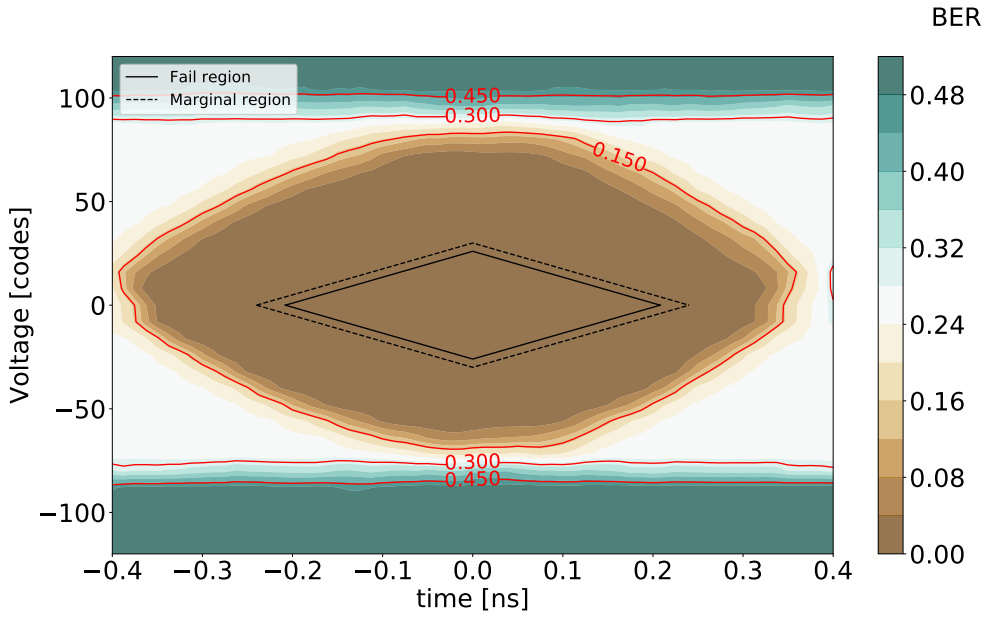


(a)

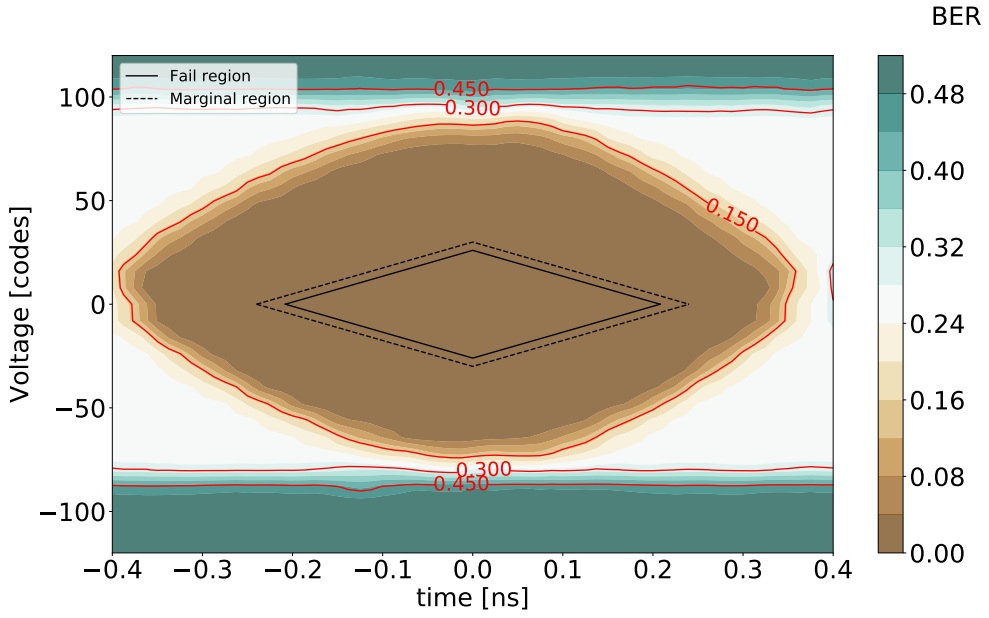


(b)

Figure 7.9: Impedance results for the single lines (a) and the differential pairs (b) for the CERN prototype (75 cm length). The vertical grey line indicates the length of the longest flex tail for the current detector baseline.



(a)



(b)

Figure 7.10: Eye diagrams for the flex cable prototype from the CERN PCB service. (a) $HV = 0\,\mathrm{V}$ (b) $HV = 1\,\mathrm{kV}$. The solid line indicates the mask in which no errors are acceptable, the dashed line the marginal region in which few errors can be tolerated.

for both eye diagrams are larger than the no error accepted area indicated by the mask. Tests over higher currents and while delivering Low Voltage (LV) are ongoing.

Quality assurance / quality control and production strategy

The set of tests described above constitutes the baseline procedure for quality control of the all the flex cables (module flexes and flex tails) during production. However, since some of the measurements (e.g. IBERT) are time consuming and some of them are only relevant for one type of cables, the option of performing them only on a limited fraction of the produced pieces will be considered if a low failure rate has been established. All electric tests will be performed in a climate chamber reproducing the operating temperature of the HGTD (about −30 °C, see Section 7.5) and under controlled humidity conditions. Radiation tolerance will only be tested on a small fraction of flex cables that will not be usable for assembly afterwards.

The design of both flex cable types will be finalized after testing them connected to the ALTIROC2 in the demonstrator described in Chapter 14. Companies that are expected to be able to produce small 4-layer flexible PCBs and to perform the assembly of all necessary components and connectors in house are being contacted, as well as those who can produce long 2-layer flex cables within the specifications listed in Table 7.6 are being contacted and the production should be ideally shared among a few of them that can provide the same quality of cables. The plan is to involve them early on in the prototyping phase so that they can contribute to the design optimizations specific to their manufacturing process.

For the module flexes the PDR is foreseen in Q2 2022, followed by the FDR in Q3 of the same year. The pre-prodcution will take place between October 2022 and February 2023 and the production from July 2023 until April 2024. The design of the flex tails will be finalised on a similar timescale (PDR Q2 2022 and FDR 2023), while the pre-production and production will take place somewhat later to reduce overlap, while being ready in time for detector assembly at CERN. The pre-production is foreseen between March and Nobember 2023 and the production from March 2024 until September 2025 (see Figure 15.6).

After the tests performed at the Institutes, the module flexes will be glued and wire bonded to the bare module (see Section 7.3.3), while the flex tails will be shipped directly to CERN to be connected during the detector assembly (see Section 13.1).

7.3.3 Gluing and wire-bonding

The assembly and interconnection of the bare module with the flex cable results in the HGTD module. The steps involved in the assembly process are the following:

- Cleaning and preparation of the flex and bare module

- Gluing of the flex on the bare module
- Wire-bonding
- Inspection, quality control and documentation

These steps are discussed in more detail below. Assembly of modules will be done in several Institutes. The HGTD modules assembly procedure will be as uniform as possible among the sites, though some differences, mainly coming from diversity of the equipment, will have to be accounted for. The prescription for the assembly will guarantee the respect of the module specifications.

This facilitates the definition of the assembly procedure and increases yield. However, the details of the assembly procedure might differ between assembly sites, mostly in the gluing step, due to the availability of specialized equipment in the different Institutes (dispensing and pick-and-place machines, for example). All module assembly and testing will take place in a clean environment equipped with temperature and humidity control and ESD protection. Specification for this environment will be developed and critical steps shall take place inside clean rooms. A database will be used to record the status of each module at every step of assembly. Electrical test results will also be added to the database. Given the number of modules needed for the HGTD, several sites are foreseen to be qualified for the module assembly activities. To ensure uniform high quality in the module assembly process the sites will be asked to pass a site qualification stage.

Initially the flex cables and the bare modules will be optically inspected for damage and anomalies. Components will be weighed and the surfaces where the adhesive will be deposited will be cleaned. Bare modules and flex circuits will be mechanically joined using a dedicated adhesive. Several adhesives are currently being studied, for robustness, radiation hardness and other practical advantages (curing time, viscosity, etc). The baseline solution would be to use the same adhesive used in the ITk Pixel detector (Araldite 2011). Different options are available to carry out the gluing process. However, all assembly methods will be ensured to produce modules to the same specifications.

One method to mechanically join the flex cable to the bare module relies on a pick-and-place machine, which typically achieves positioning accuracy of $\sim 10\,\mu m$, and exists in a variety of automation options (from mostly manual to fully automated). Pre-tested components (flex cable and bare module) are loaded by vacuum tools of the machine. The operator then aligns the components through fiducials in the module (on the ohmic side of the sensor) and flex, visualized simultaneously in the machine monitor screen, and applies manually, or through a dispensing arm or stamping tool, the adhesive to the bare module and/or flex cable. The flex is then placed on top of the bare module (or the bare module is placed on top of the flex) and held in position until the adhesive is sufficiently cured.

An alternative process relies on custom made jig gluing tools instead of the pick-and-place machine. As shown in Figure 7.11, the tool consists of two aluminum jigs, each consisting of

a vacuum chamber to hold the bare module and the flex, respectively, at fixed positions and then the adhesive is applied. The vacuum pressure is applied through holes in exchangeable inlays with a shape adapted to the electrical components on the top side of the flex. In order to guarantee the correct alignment between the inlays in x and y directions, three dowel pins are used in each jig before the guide pins, and the body is adjusted with precision screws. Positioning accuracy of $\sim$ 100 µm is achievable with this method. The distance in the z direction can be adjusted in the tooling to allow variation in the amount of glue.

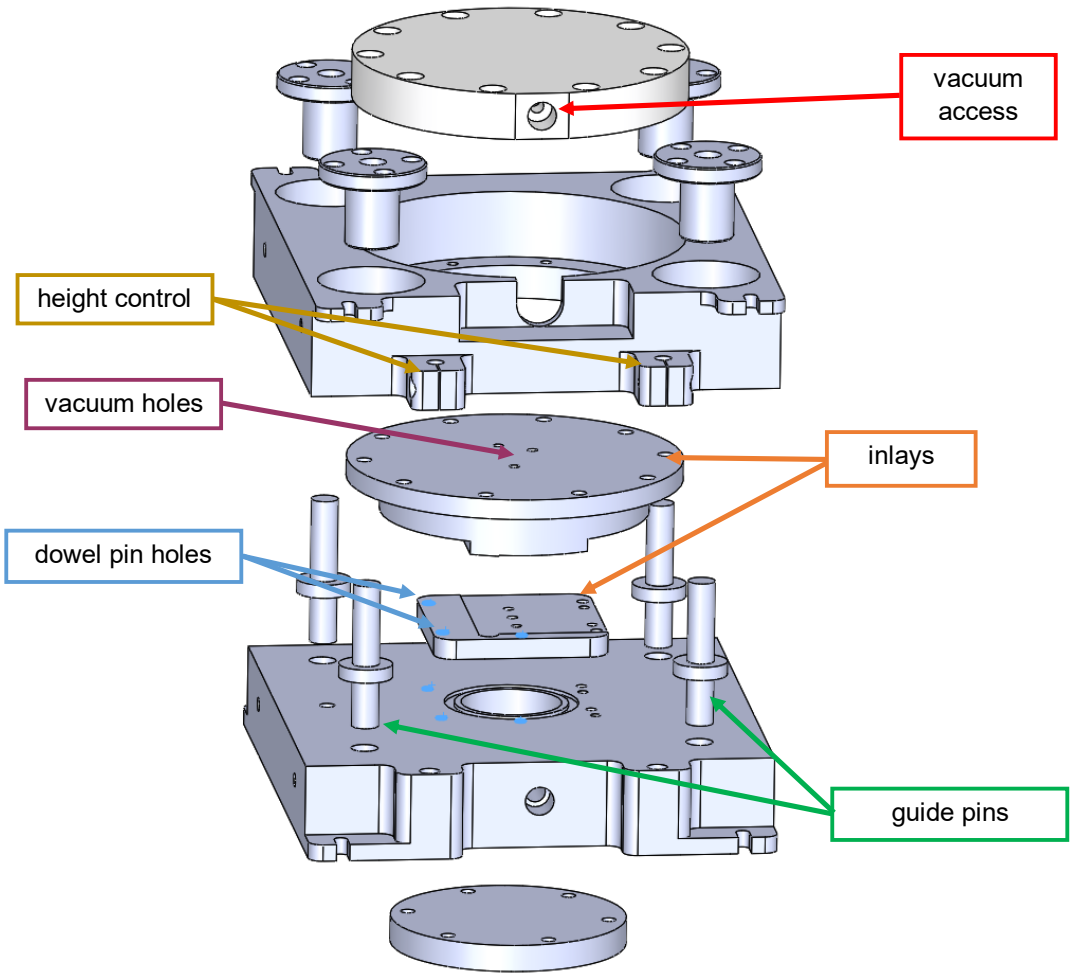


Figure 7.11: The design of custom made jig gluing tools.

Following mechanical assembly, the front-end chips and the sensor bias voltage are electrically connected to the flex circuit through 25 µm diameter aluminium wire bonds using an automated ultrasonic wedge bonder. Wire-bond quality will be checked routinely through pull tests of sample wire bonds using a pull tester machine. Visual inspection of the wire bonds will also be performed. Figure 7.12 shows an assembled ALTIROC1 device and the

pull testing procedure.

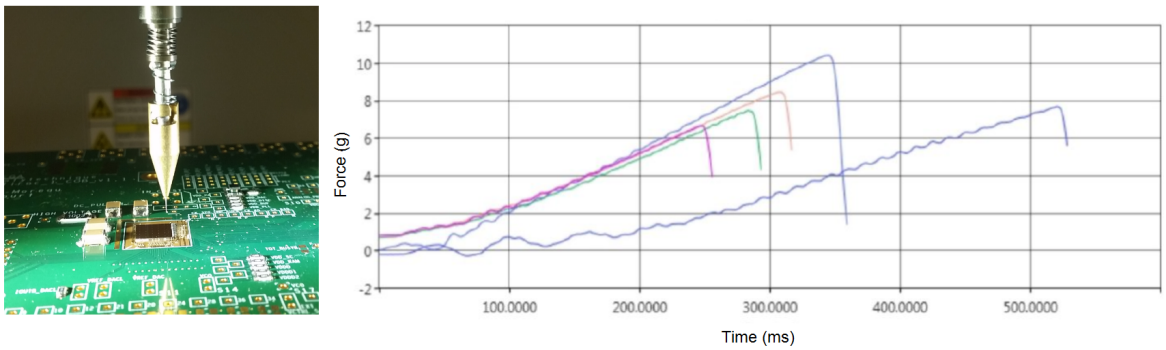


Figure 7.12: Photo of a mounted ALTIROC1 device being tested (left) and the measured wire strengths (right).

7.3.4 Specifications, quality assurance / quality control

The bare modules and flex cables that fulfil all their respective requirements will be used for the next steps in the module assembly, i.e. gluing and wire bonding. The specifications for this stage are aimed at ensuring the mechanical stability of the assembled module, see Table 7.9. These need to be combined with the requirements in terms of efficiency, response and number of working channels defined for the sensors and that are valid also for the assembled module.

Radiation tolerance	2 MGy
Neutron fluence	$2.5\text{x}10^{15}\text{neq/cm}^2$
Lap shear force	5 MPa
Push-off strength	10 MPa
Wire bond pull force	8 gf
Positioning accuracy	100 µm

Table 7.9: Specifications of the gluing and wire bonding processes.

After assembly all modules will be optically inspected and weighed, and their metrology recorded in the database. As mentioned above, wire-bond pull tests will be carried out periodically on a fraction of modules to ensure robust connectivity. All modules will be tested for ASIC communication, current-voltage behaviour and response to a radioactive source using a lightweight table top DAQ system. Short burn-in tests, where the modules are operated continuously for a day are foreseen. Furthermore a small fraction of the modules will be subjected to long-term burn-in tests where the devices will be thermal cycled while being operated.

7.3.5 Production strategy for flex cables and module assembly

The flex cable design will be finalized after testing it connected to the ALTIROC2 in the demonstrator described in Chapter 14. Companies that are expected to be able to produce long flex cables within specifications are being contacted and the production should be ideally shared among a few of them that can provide the same quality of cables. The plan is to involve them early on in the prototyping phase so that they can contribute to the design optimizations specific to their manufacturing process.

The module assembly production, including inspection of parts, gluing, wire bonding and testing of the modules, will be shared among several HGTD Institutes, with the target production rate of four modules per day per institute. Institutes module assembly qualification procedure will be defined to ensure that all sites uniformly produce modules according to specifications. A minimum set of equipment will be required (for example, wire-bonding and pull and shear machines) as well as a clean environment and minimum throughput capacities.

After the prototyping phase is completed, and the first full sized (ALTIROC2) modules are produced, the module assembly PDR will be submitted in Q2 2022, followed by the FDR in Q4 that same year. The module pre-production will take place at the first half of 2023, while the production is foreseen from Q4 2023 to Q3 2026 (see Figure 15.6).

7.4 Module loading

7.4.1 General description

The assembled modules have to be mounted on the cooling plates in readout rows, aligned along the x or y direction. Figure 7.13 shows the position of the modules on the front side (left plot in red) and back side (right plot in blue) allowing an overlap of 20% for the inner part, 55% for the middle part and 70% for the outer part. The longest rows contain 19 modules. For mechanical stability the modules will be glued to a thin support plate which is then screwed to the cooling plate. The modules will be held in place between the support plate and the cooling plate. As described previously, the active area is divided into three rings (inner, middle and outer ring). Therefore, three types of support units per side corresponding to the three rings are foreseen. The inner support units consist of half disks of $120\,\text{mm} < r < 230\,\text{mm}$, the middle and the outer support units of quarter disks of $230\,\text{mm} < r < 470\,\text{mm}$ and $470\,\text{mm} < r < 660\,\text{mm}$, respectively. Figure 7.14 shows a drawing of the detector units with the loaded modules. The inner and middle disks will be replaced every $1000\,\text{fb}^{-1}$ and $2000\,\text{fb}^{-1}$, respectively. The total number of support units for the eight sides of the detector is 80 (16 half inner supports, 32 quarter middle supports, 32

quarter outer supports). Moreover, since the positions of the modules are different for the two sides of the cooling plates, there are six different types of support units.

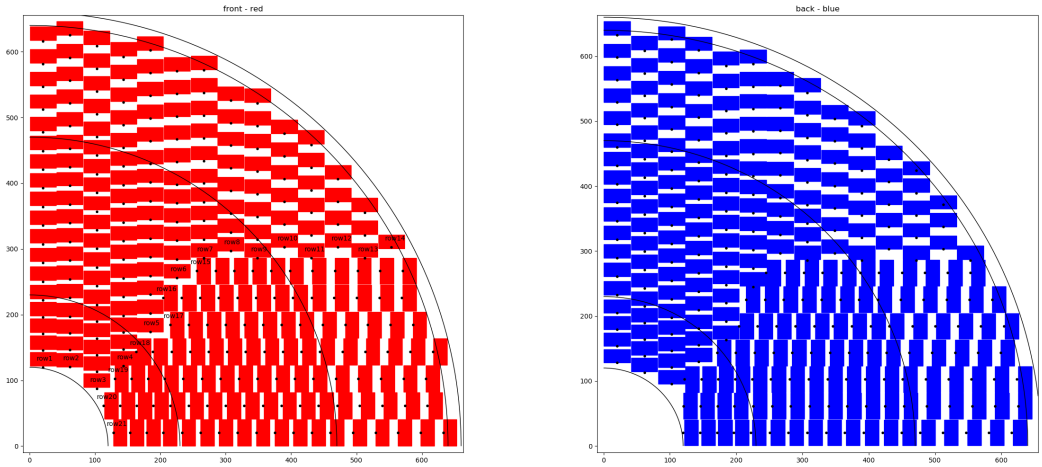


Figure 7.13: Position of modules and readout rows numbering for the front side of one disk (left plot in red) and for the back side of the same disk (right plot in blue). Smallest radius at 120 mm and maximal radius at 660 mm are shown. The 640 mm radius is the minimal target for the external instrumented area. 230 mm and 470 mm radius are shown as typical limits of the different parts of the support units.

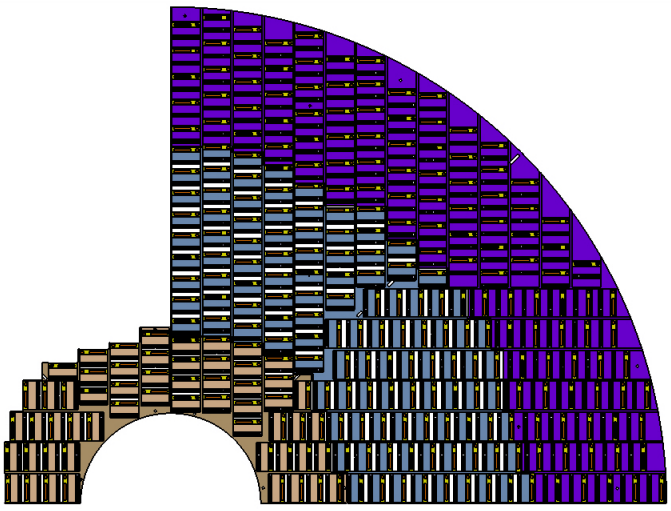


Figure 7.14: Detector units with modules assembled on the inner, middle and outer support plates

7.4.2 Support units and detector units

Modules are installed and glued on plates (the support units) to be screwed on each side of one of the four cooling plates. The baseline design of the support units consists of a pattern plate, with modules inserted between the plate and the cooling plate. The full size plate is divided into three parts as shown in Figure 7.14. The current baseline is to use half disks for the inner part and quarter disks for the middle and outer parts. Depending on the feasibility (fragility, flatness, glue deposition), smaller supports could be considered. The maximum thickness of the support plate is typically 6 mm. The target material is currently carbon fibre. As an example, Figure 7.15 shows the current design of the half disk of the inner support unit.

Windows are machined in order to encapsulate the modules which are glued on rectangular strips, while leaving room for the wire bonds and the connector between the module flex and the flex tail (see Figure 7.15).

The windows of the plate give the positioning of modules. The support structure and each window have edges with a precise height, ensuring a constant distance between the modules and the cooling plate. The edges are in contact with the cooling plate and the height is greater than the thickness of the module. Each support unit will be checked with a 3D metrological machine before loading. The tolerances allowing a thermal contact will be defined thanks to the measurements on the demonstrator. These tolerances must also allow a sufficient height so that the modules are not damaged in compression.

Once the detector unit is screwed to the cooling plate, the modules are in direct contact with it, so that the thermal properties of the plate material and of the glue used to fix the modules are not critical. Moreover, thermal grease will be used to improve the contact between the modules and the cooling plate.

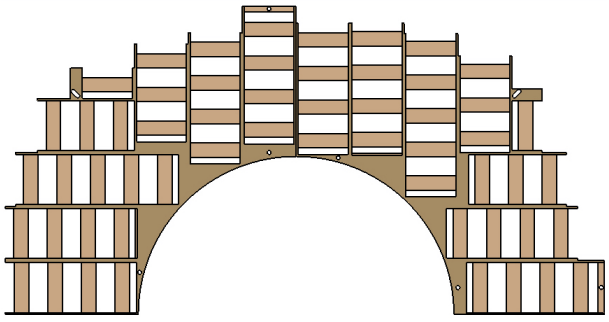


Figure 7.15: Drawing of inner support unit with holes for fixation on the cooling plate

For better mechanical strength and rigidity, some reinforcements are added (Figure 7.15). In particular, the material is added near the zone of the internal radius and everywhere possible, leaving open only the necessary areas. First tests show that a single half disk for

the inner part and a single quarter for the middle and outer parts would guarantee the stability of the global structure. This type of support plate is more complex than a full plate, since the windows need to be defined precisely for each module, but then the positioning of the module itself is straightforward. The structure provides mechanical protection to the modules and the plate has good rigidity with a minimum contribution to the total thickness of the detector. On the downside, this design only allows for a small surface when gluing the module to the plate and the mechanical strength and long term stability have to be proven. Some tests have to be performed and several prototypes and the demonstrator will be used to draw a conclusion (see Chapter 14). Should a module be found to be faulty after gluing to the support, rework will be possible. Conclusive tests of module removal have already been carried out and others will be done with the demonstrator.

7.4.3 Gluing studies

The modules are fixed to the support unit with four glue dots of 2 mm diameter (see Figure 7.16). The glue dots are deposited onto the edges of the module flex. The glue for module loading into the intermediate plates is required to meet the parameters listed in Table 7.10.

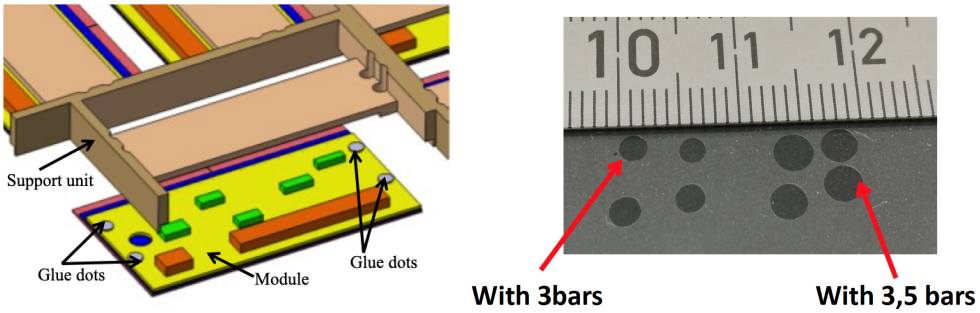


Figure 7.16: Schematic view of the module with the four glue dots allowing the fixation with the support unit (left) and test of glue deposition (right) - Pressure values are an example of tuned parameters, depending on the duration and temperature.

With these constraints, six types of glues have been chosen to perform the tests: Araldite 2011; EG7655-LV; EG7655; EG7658; EG8050; Stycast 2850FT. Their characteristics have been checked in the MaxRad (Materials and Adhesives for Extreme Radiation Environments) CERN database [1]

Ease of implementation (fluidity, life time, duration and temperature of the polymerisation) has been evaluated. The ITk choice is also considered, in particular for radiation hardness. The final choice will have to be qualified. Moreover, push-off strength measurements and lap shear tests have been performed in several configurations. These tests have been done

[1] https://maxrad.web.cern.ch/

using some dummy modules, with a piece of flex cable glued onto, to mimic the module flex. Other tests with a glass plate have been performed to determine the volume of glue needed to obtain the correct thickness and surface (see Figure 7.16). Finally, taking into account all the tests already done and the results, Araldite 2011 is chosen as the baseline for the loading of the modules onto the support units.

Radiation tolerance	> 5 MGy
Viscosity	< 100 Pa s
Lap shear force	~ 1 MPa
Push-off strength	~ 1 MPa

Table 7.10: Specifications of the glues parameters.

7.4.4 Procedure for loading and qualification

The procedure for detector unit loading will be tested when assembling the demonstrator (see Chapter 14), which will be also used to improve the qualification steps. Tools for each step are being developed and tested. Tests are performed following a procedure first using glass plates then silicon glued to a small flex prototype, all without any electrical functionalities, but with the correct geometrical dimensions, instead of actual modules. For all the tests Araldite 2011 is used as glue.

Module loading on support unit should follow this procedure :

1. The modules are placed on a temporary plate with the exact pattern of the final module positions. They are held by a suction system included in the plate.

2. Four glue dots are dispensed on the left and right edge of the module flex (see Figure 7.16). The correct amount of glue is ensured by the usage of an automatic dispenser.

3. The support unit is put in place and pressed on all modules at a nominal compression strength. An adjusting shim is used to ensure the correct thickness of the glue.

4. The polymerisation is carried out (temperature and duration to be defined after final glue tests).

5. The detector unit is removed from the temporary plate and fixed on another plate for packaging and shipping.

6. The system is turned over upside down and a second transport plate is fixed on the top.

7. Electrical tests will be performed at this stage. The tests must ensure that the loading has not damaged the modules. Measurements after loading will be the same as those made after module assembly. Test benches will be adapted, especially the size of the testing box.

8. The detector units are then packed in specially designed packaging to ensure secure transport to CERN where they will be mounted onto the cooling plate (see Chapter 13).

During production, a visual inspection will be performed after module loading, looking for possible mechanical damages to the module, in particular to the edges of the hybrid, the components of the module flex and the wire bonds. Signals will be injected into the sensors and the response will be tested with the same DAQ system used for the test of the single modules. Additionally, it will be checked that there is no interference between the modules, by temporarily connecting stacked flex tails to adjacent modules along a readout row. For all detector units passing the qualification tests, the information on the nominal and measured position of the modules on the support unit, as well as any relevant performance results will be saved to a database. Once the initial characterisation is completed in the R&D phase, thermal tests are not foreseen during production. Mechanical stress tests could be performed on a small fraction of support units if it is deemed necessary.

7.4.5 Detector unit assembly strategy

Once their design is finalised, the production of the support units will be carried out by a company and the quality control by one or more Institutes. Then, the plates will be shipped to the module loading sites that have been qualified. To minimize the amount of modules to be shipped and to avoid long distance transport, sites able to perform both module assembly and loading or geographically close to the module assembly sites will be preferred. Since the setup for mechanical and electrical qualification of the detector units is similar to the one needed for module assembly, the site qualification procedure will be mostly common to both activities (excluding the wire bonding capability in this case). As for module assembly, the exact procedure used for module loading might be slightly different among the Institutes, but the same quality of assembled detector units has to be delivered. 80 detector units (16 inner, 32 middle, 32 outer) will be produced in total. The glue and the expendable supplies will be purchased from one or more companies. Most of the components of the electrical test benches are standard ones, available in the Institutes. Some dedicated electronic boards will be developed to test the modules at many steps of the construction of the detector, included the loading step. The gluing and positioning system will be developed in the Institutes, using existing elements, complemented by specific mechanical parts. Because of the non-standard shape of the detector units and the fragility, different types of dedicated packaging will be necessary for transportation from loading sites to CERN.

7.5 Thermal Simulations

The power dissipation of the sensor depends strongly on the temperature of the sensor. The irradiation of the sensors will increase the leakage current thus increasing the power dissipation at a given temperature. Therefore the thermal properties of the system have been studied following the strategy outlined in [84].

Al	Al–Ti	CF–Al	Block	Material	Thickness [mm]	Thermal Conductivity [$W\,m^{-1}\,°C^{-1}$]
x	x	x	Sensor	Si	0.25	124
x	x	x	Bumps	SnAg	0.05	79
x	x	x	ASIC	Si	0.25	124
x	x	x	Foil	Polymer	0.10	3.5
x	x		Structure	Al	0.50	135
		x		CarbonFiber	0.50	1
x			Cooling	Al	2.50	135
	x			Al	2.00	135
		x		Graphite foam	2.00	30
x	x		Interface	Polymer	0.1	3.5
x		x	Pipe	Al	0.50	135
	x			Ti	0.30	22

Table 7.11: Material type and thickness used in the simulation of the thermal properties.

As discussed in Chapter 11, several variants of the cooling system are under consideration:

- Al: the entire cooling structure is made of Aluminum down to the pipes. The thermal contact between the pipes and the structure is ensured with an interface foil made of Polymer.
- Al–Ti: identical to Al with the exception of the pipes which are made of Titanium.
- CF–Al: the structure is made of carbon fiber, the pipe of Aluminum. The thermal contact is ensured with graphite foam between the structure and the cooling pipes.

For a choice between the variants the thermal properties, the thermo-mechanical properties (deformations), the electro–chemical compatibility of the material and the radiation length of the materials have to to considered.

CarbonFiber and together with graphite foam are light materials which will lead to a small contribution to the total radiation length. CarbonFiber is rather stable under temperature variations. The thermal conductivity of CarbonFiber depends on the orientation of the fibers, the value given in Table 7.11 is for main direction of the the heat flow in the HGTD. In transverse direction, the conductivity is two orders of magnitude higher.

Titanium and Aluminum have a larger contribution to the radiation length of the HGTD. Aluminum is favored with respect to Titanium for this aspect. However Titanium is more robust, therefore thinner structures can be built. A homogeneous use of the a single type of material such as Aluminum has the advantage that the system is insensitive to differences in the thermal expansion properties of the materials. The thermal conductivity of Aluminum is better than that of Titanium as shown in Table 7.11. The thermal simulations are performed with ANSYS. When available in ANSYS, the temperature dependence of the conductivity is taken into account, e.g., for Aluminum.

For the following calculations the CF–Al setup has been used to determine the thermal properties of a module. If such a system is thermally stable, it will also be stable if a material with better thermal conductivity. The impact of deviations from this choice will also be discussed. For a larger system, corresponding to a half HGTD wheel, the Al–Ti variant was simulated.

In a first step, the geometry of a stack with a single ASIC and (half an LGAD) sensor is built. The material used in the thermal simulation of the module are shown in Table 7.11 along with their thickness and thermal conductivity. The sensor, the ASIC, the foil, the structure and the cooling are implemented each as a cuboid built of a square 2 cm × 2 cm and the height given in Table 7.11. The conductivity of silicon increases with decreasing temperature, conservatively the bumps connecting the sensor to the ASIC are implemented individually as 225 cylinders with a radius of 0.045 mm and height of 0.05 mm.

The cooling pipes are half-cylinders embedded in the cooling material. The inner radius of the pipes is 1.5 mm. The outer radius is 2 mm if the pipes are made of Aluminum and 1.8 mm for the Al–Ti setup. As a consequence the cooling structure made of Aluminum with Aluminum pipes has a half–width of 2.5 mm and 2 mm for the Al–Ti and CF–Al setups.

The cooling is simulated as convection which is applied on the surface of the cooling pipes. The nominal temperature of the coolant is −35 °C. As baseline a power consumption of the ASIC of 1.2 W (0.3 W cm^{-2}) is used. For the sensor a power consumption of 0.4 W (0.1 W cm^{-2}) is assumed.

While the contact between the sensor and the ASIC via the SnAg bumps is assumed to be perfect a thermal contact conductance of 0.01 W mm^{-2} °C^{-1} is applied to the contact between ASIC and foil as well as foil and the carbon fiber structure. The contact conductance leads to a temperature step increasing the thermal resistance of the system. For a power dissipation of 1.6 W the temperature step is 0.4 °C at each material transition.

In Figure 7.17 the result of the thermal simulation by ANSYS is shown. The maximum temperature difference with respect to the nominal temperature of −35 °C is 7.6 °C. If the ASIC is powered alone, the temperature difference is 5.5 °C, for the sensor alone, the temperature difference is determined to be 2.1 °C. The thermal resistance for the sensor is therefore 5.3 °C W^{-1} and for the ASIC it is 4.6 °C W^{-1}. As the difference between these two resistances of 0.7 °C W^{-1} is due to the soldering bumps, the thermal resistance of the

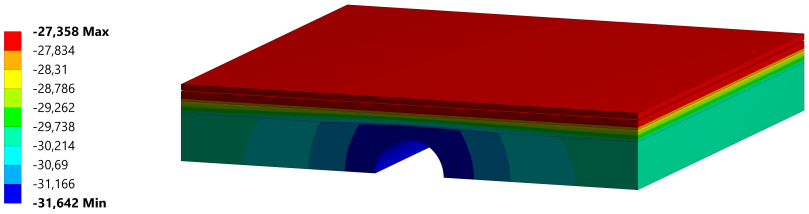


Figure 7.17: The temperature distribution is shown for the baseline power consumption with an ASIC and a sensor half.

bumps was calculated analytically using a continous equivalent volume of SnAg instead of the discrete bumps. The approximation leads to a resistance of $0.5\,^{\circ}\mathrm{C\,W^{-1}}$, the larger value for the individual bumps can be understood as the heat transfer will encounter also the resistance in the sensor plane before reaching the bumps in order to flow to the cold reservoir.

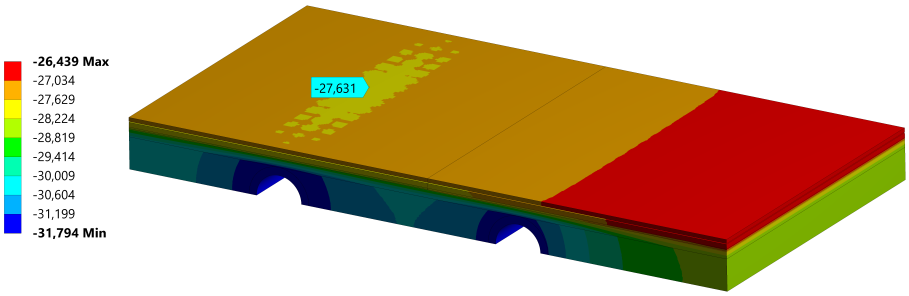


Figure 7.18: The temperature distribution is shown for the baseline power consumption with two ASICs and one sensor.

As a second step, the second half of the sensor was added as well as the corresponding ASIC. The current design of the cooling pipes calls for pipes every 16 mm in radius, therefore a second cooling pipe was added at the nominal distance leading to an asymmetric configuration shown in Figure 7.18. Compared to the previous simulation the temperature difference increases to 8.6 °C peak to peak. However, the temperature distribution of the sensor now shows variations with a symmetry axis corresponding to the axis of the cooling pipe. Restricting the study to a single cooling pipe ± half the cooling pipe to cooling pipe distance, the temperature increase is reduced to 7.4 °C, close to the result of the previous

simulation within 5%. For the simulations with only the sensor or ASIC dissipating power the temperature increase is globally larger, however as the increase is less than a factor 2, but the power is doubled, the resulting thermal resistance is smaller. Therefore the single-ASIC simulation is a good approximation of the system. Additionally, the geometry is conservative as the next cooling pipe is close to the second ASIC, but has not been simulated. This would further reduce the thermal resistance.

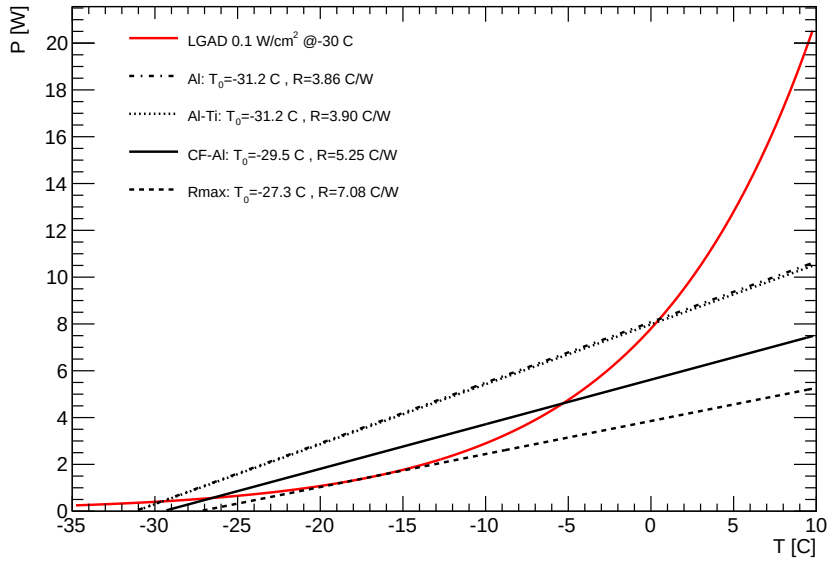


Figure 7.19: The power dissipation of the sensors as function of the temperature is shown as well as the thermal properties of the CF–Al, Al, and Al–Ti systems. The dashed line corresponds to the CF–Al system with a degradation of the thermal resistance from the ASIC to the cooling pipe by 40%.

As the power dissipated by the sensor increases as function of the temperature, if the system cannot evacuate the heat effectively, the temperature will increase, increasing the leaking current, so that a thermal runaway condition is created as explained in [84].

The power dissipation of the sensor, modeled according to Section 5.5.8, is shown as a function of the temperature in Figure 7.19. The strong temperature dependence is clearly visible, e.g. in the red curve for the baseline. The power dissipation of the ASIC increases the effective temperature delivered by the cooling system to −29.4 °C. The black line has a slope which is the inverse of the thermal resistance for the sensor. Once the power dissipation of the sensor crosses this line, thermal runaway is excluded as the heat can be evacuated efficiently. At −5 °C stable operation cannot be achieved anymore.

The maximum power dissipation the system can handle is roughly 0.17 W cm^{-2} for the sensor. Thus compared to the baseline a margin of 70% is included in the system. The model

from Section 5.5.8 was compared to the one used in [84] by normalizing the models to the same power dissipation at a temperature of −30 °C. In a window of half-width 5 °C around the normalization point, the two models agree within 15%.

A different way of analyzing the properties of the system is to determine the resistance for which the baseline sensor power dissipation is tangent to the line (Rmax scenario). Since the thermal properties of the system from the sensor to the ASIC are driven by the SnAg bumps, the contact being essential for a functioning system, this part of the simulation is left unchanged. The thermal resistance from the ASIC to the cooling pipes is increased by by 40%. This results in the increase of the starting temperature to −27.2 °C. The resistance increases to 7.1 °C W^{-1} resulting in a flatter slope than in the nominal case. The dashed black line shows the result of the increased thermal resistance.

The simulation can also be interpreted in the following way: an increase of at least 40% of the ASIC power dissipation can be handled by the nominal system. The ASIC power increase would increase the starting temperature to −27.2 °C, but it would not affect the slope as the thermal resistance is unchanged.

The temperature increase of 5.6 °C for the nominal system includes the contact conductance degradation of 0.8 °C proving a further margin of 14%. As the effective contact area between materials is difficult to estimate, it is essential to have this margin built into the system.

In the Al–Ti setup the main part of the cooling system is made of Aluminum with exception of the pipes, the thermal resistance of the system is improved further. The effective operating temperature of system would decrease to −31.2 °C and the thermal resistance would decrease to 3.9 °C W^{-1} as shown in Figure 7.19 leading to further margin in the operation of the system.

The thermal resistance of the Al is practically identical to the Al-Ti setup. Aluminum has a superior thermal conductivity with respect to Titanium, but the Aluminum pipes are thicker than the pipes made of Titanium. In an assembled Aluminum system the cooling structure is thicker by 0.5 mm half width. This additional contribution to the thermal resistance annihilates the gain expected from the thermal conductivity.

The results for the Al and Al–Ti should be interpreted with caution. If the cooling pipes are produced separately from the cooling structure, a foil is needed to ensure the thermal contact. The simulation has been run without a surface conductance. If the same conductance as the one between the ASICs and the foil is used (on both sides of the foil), the peak temperature increases by 3 °C. The effect is larger than for the other foil as the effective surface of the cooling pipe is smaller. The smaller area is compensated by an increase of the temperature difference. In this case the CF–Al setup has better thermal conductivity than the Al–Ti and Al setups. If in the Al setup the pipes are integrated in the cooling structure and the foil is not needed, the contact will be excellent.

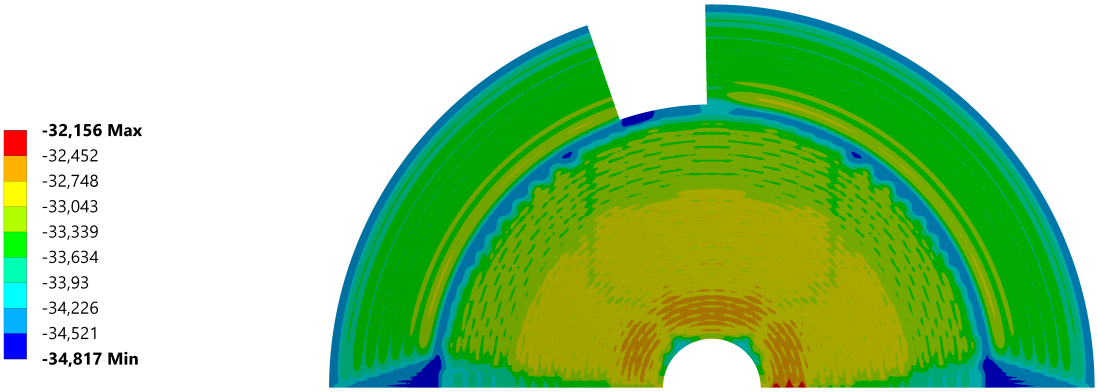


Figure 7.20: The temperature distribution on the surface of the cooling system is shown for a half disk when applying 0.4 W cm^{-2} at the location of each ASIC.

The detailed simulation of the stack for the full HGTD is not possible for the ASIC stack. Therefore a different approach is used. The cooling system is simulated fully for a half disk using the Al–Ti setup. At the position of each module on the disk a power dissipation of 0.4 W cm^{-2} corresponding to 1.6 W is applied. The resulting temperature variation is shown in Figure 7.20. The maximal temperature overall is −32.2 °C. Three regions, corresponding to the three rings of the HGTD can be distinguished. The modules and the space between the modules explains the temperature variation. In the inner ring of the HGTD, at the position of a module, the typical maximal temperature is slightly colder at −32.6 °C. The temperature decreases slightly, only by about 0.3 °C, in the space between the modules.

Having determined the maximal temperature on the cooling structure with the full simulation, the highest temperature of the sensor is calculated by using the detailed model of a single ASIC. The temperature of the structure is fixed to the maximal temperature observed in full simulation, i.e., −32.2 °C. The baseline power dissipation of sensor and ASIC is simulated. The temperature at the sensor increases by 1.2 °C to −31 °C. This is a lower temperature than the 29.6 °C determined in a single pipe simulation using the Al–Ti setup.

While the detailed setup with a single cooling pipe simulates a straight cooling pipe, the full simulation takes into account the curvature of the pipes. For some modules in extreme cases, this could have lead to a loss of cooling surface of the pipes, increasing the thermal resistance, e.g., if the pipe exits the module on the side instead of going through the whole module. On the other hand, the distance between two cooling tubes is smaller than the lateral size of the half–module. This leads to a larger surface area of cooling in full simulation with respect to a single pipe simulation. The temperature difference in the detailed model is therefore increased, increasing also the thermal resistance. The result shows that the simulation of

a single ASIC with a single cooling pipe is conservative setup, leading to a higher peak temperature than the one that would be determined in a full model.

All three types of setups, i.e., Al, Al–Ti and CF–Al, have the potential to ensure stable operation of the HGTD without thermal runaway. Ensuring the quality of the thermal contact between the different parts of the HGTD will be essential to keep the thermal resistance of the system under control.

7.6 Roadmap towards production

The total surface covered by the HGTD (6.4 m^2) requires a well planned approach to successfully carry out the module assembly and loading. A brief schedule of module assembly and loading activities is described below, and more details can be found in Section 15.2 and Figure 15.6.

- *Bare Module Hybridisation*: Full-size bare module prototyping will be carried out after the final sized sensor and ASIC (ALTIROC2) become available. A Specification Review will take place in Q2 2021 before hybridization qualification. The PDR will be submitted in Q1 2022, presenting the bare module prototype design with ALTIROC2. Hybridization qualification is scheduled in Q2 2022 with two or more companies to qualify their hybridization service (including metalization, bump-deposition, dicing and flip-chip). It is followed by the FDR in Q4 2022. The bare module pre-production will take place in Q1 2023, followed by the production from Q1 2024 to Q1 2025.

- *Module Assembly and Module Flex*: The module assembly production will be shared among 4 to 5 HGTD Institutes. The overall production rate is expected to be approximately 19 modules per working day in the first half and 22 modules per working day in the second half of the production. After the prototyping phase is completed, and the first full-sized (ALTIROC2) modules are produced, the module assembly PDR will be submitted in Q3 2022. Institutes module assembly qualification procedure will start in Q4 2022, to ensure that all sites uniformly produce modules according to specifications. , followed by the FDR in Q4 in Q2 2023. The module pre-production will take place in the second half of 2023, while the production is foreseen from Q3 2024 to Q3 2026.

- *Module Loading*: The design and specification review of detector support units and module loading procedure will take place in Q1 2022, followed by the PDR in Q4 2022. Once the design is finalised, the production of the support units will be carried out by a company and the quality control by one or more Institutes.

  The module loading will be shared among several HGTD Institutes. Institutes able to perform both module assembly and loading or geographically close to the module assembly sites will be preferred. Institutes will assemble the modules and then load

them on the Detector Units. The site qualification procedure will take place in 2023. This qualification is mostly common to both module assembly and loading activities. FDR will be submitted in Q2 2023, followed by pre-production at the second half of 2023. The production will last from Q3 2024 to Q3 2026. 80 detector units (16 inner, 32 middle, 32 outer) will be produced in total. They will be shipped from loading sites to CERN.

- *Flex Tails*: The flex cable design will be finalized after testing it connected to the ALTIROC2 in the demonstrator described in Chapter 14. In the design, simulations and tests particular attention will be given to avoiding mechanical stress on the other components due to the expansion and shrinking of the long flex tails with temperature. A Specification Review will take place in Q1 2022, followed by the PDR in Q2 2022. The production of flex tails should be ideally shared among a few of the companies that can provide good quality of cables. FDR will be submitted in Q1 2023, followed by pre-production from March to November 2023. The production is foreseen from March 2024 to September 2025.

8 Power distribution, Grounding and Shielding

This section covers the powering of the detector, including the schematic layout of the High Voltage (HV) and Low Voltage (LV), from the supplies located in the USA15/UX15 services cavern, the DC-DC converters placed at the patch panels boxes in the PP-EC area (Section 12.3), up to the peripheral electronics and modules sitting inside the vessel. The grounding and shielding schemes are also briefly described. The details of the services needed to power the detector and respective connectivity are described in Chapter 12.

8.1 High voltage

Each of the 8032 LGAD sensor modules of the detector require an individual bias voltage in a range up to 800 V. Such a high voltage is needed to power the sensors after being exposed to the high radiation conditions of the HL-LHC (detailed in Chapter 5). The bias voltage of the sensors has to be adjusted due to the gain degradation with the received fluence. Figure 5.17 shows the required bias voltage as a function of the radial position for different fluence levels. In combination with the non-radial geometry, this results in a limited possibility to connect several modules to the same bias supply. Since sensor modules close in radius are expected to require the same voltage, the baseline choice is that two modules share bias supply. High voltage supplies capable of delivering 6 mA current per channel will be used, which allows with sufficient margin an average leakage current up to 5 µA per pad for irradiated sensors . This choice, which requires 2008 HV channels per end-cap, allows to save cost. Commercial supplies with multi-channel rack mounted units will be located in the service cavern. Systems with high channels density (∼400-500 per crate) allow to minimize space but also to reduce cost. A schematic layout of the high voltage system is shown in Figure 8.1.

The filter units at the PP-EC area will also serve as patch panels allowing to select sharing of supplies of the individual sensor modules. A further low pass filter is placed on the flex cables near the sensor modules. In the baseline design each sensor has an individual HV return connection to the filter unit. An alternative solution, using a common return from the reference grounds of peripheral electronics boards, is under study. In this solution the individual HV is referenced to the analog ground at the module. The voltage difference of

the analog ground at the module and the PEB ground will be less than 50 mV even for the longest flex cables and can be neglected.

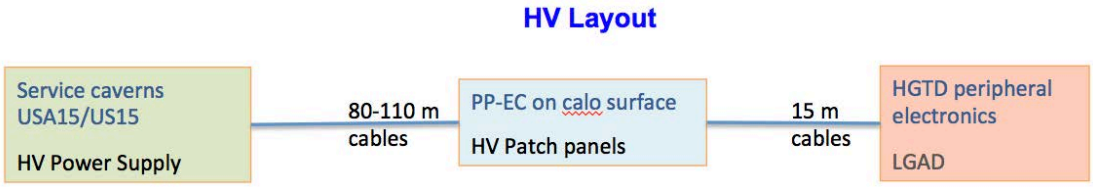


Figure 8.1: HGTD schematic High Voltage layout

8.2 Low voltage

For supplying the low voltages needed by the front-end and peripheral electronics a three stage system is used, as shown in Figure 8.2. The system will have to be able to deliver about 20 kW at 1.2 V to the Front-End ASICs as well as the peripheral readout electronics. Bulk power supplies located in USA15 provide 300 V DC current to DC-DC converters placed in the PP-EC areas (described including radiation environment and magnetic field in Section 12.3). These second-stage multi-channel DC-DC units convert the 300 V to 10 V which is distributed to radiation hard DC-DC converters located on the peripheral electronics boards inside the vessel (details in Chapter 9). The last stage converts power to the front end ASICs on the detector chips and the peripheral electronics providing mainly 1.2 V DC power but also 2.5 V for optical links. The converters on the peripheral boards are based on the bPOL12V ASIC developed by CERN for the HL-LHC upgrade. Due to space limitations on the peripheral boards, the 10 V to 1.2 V conversion will be done in a single stage (see Section 9.2). The exact output voltage for each converter on the peripheral boards is selected by a resistor chain to take the voltage drop of the flex cables into account.

Each ALTIROC ASIC requires 0.5 W analog power and 0.7 W digital power at 1.2 V. Separate DC-DC converters will be used for the analog and digital voltages. With two ASICs per module, one bPOL12V based DC-DC converter will be used to supply analog or digital power to 3 modules. The modules connected to the same DC-DC converter are chosen to assure that the voltage difference is less than 30 mV which is within the (1.20 ± 0.05) V specifications of the ALTIROC ASICs (Table 6.1). With 2008 modules per disk (or double-sided layer), 1408 DC-DC converters on the peripheral electronics per disk are needed to power the front end electronics, including power losses on the flex cables. A further 120 DC-DC converters per disk are required for powering the components on the peripheral boards.

These DC-DC converters on the peripheral electronics will need to provide almost 5.0 kW of power per disk. With an efficiency of 72% (at −30 °C and 3 A), each disk has to receive

850 A at 10 V which will be supplied by 72 channels that are able to provide 16 A each. The number of channels is given by the requirement that the ground reference is separate for each peripheral board and that 32 out of the 40 boards per disk require more than 16 A at 10 V. The 300 V will be provided by 14 rack-mounted units in the service cavern, each delivering 3 kW. Details on the low voltage units are given in Table 8.1.

Voltage	Location	Current/channel	Nb of channels/units
300 V	USA15	10 A	14
300 V $\rightarrow$ 10 V	PP_EC	16 A	288
10 V $\rightarrow$ 1.2 V (or 2.5 V)	On peripheral board	4 A	6112

Table 8.1: Type of LV units, location, maximum current delivered per unit and number of units/channels.

With an 80% efficiency of the 300 V to 10 V DC-DC power converters located in the PP-EC area, a total cooling power of 4 kW per end-cap is required at these locations. A water leak-less cooling system, providing water at $\approx$ 18 °C, and corresponding pipes/manifolds on the calorimeter surface will be needed. Details on the services, patch panels area and cabling are given in Section 12.6.

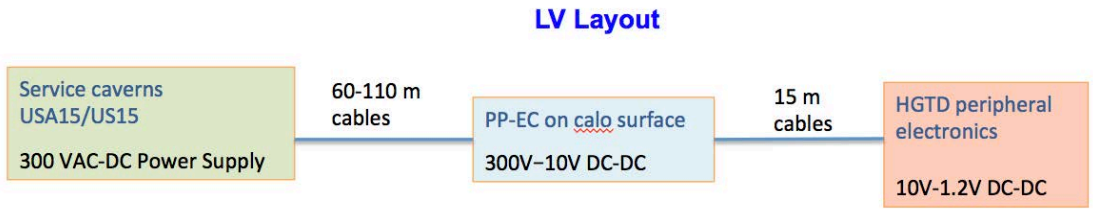


Figure 8.2: HGTD schematic Low Voltage/power layout

8.3 Grounding/shielding

The grounding and shielding of HGTD follows similar requirements as defined for ITk [85]. The ground reference point for the HGTD itself will be the inside of the detector vessel. This inside is covered with a thin high conductive foil to ensure the function as a Faraday cage.

Both end-caps will be independent Faraday cages. The cage extended up to the patch panels at PP-EC through the shields of the LV, HV and control cables. The patch panels as well as the vessels are electrically insulated from the detector walls and from the mechanical

structures on which they are mounted. The Faraday cage will be connected through a single ground line to the ATLAS common ground. This will constitute the reference potential. The connectors for the conductor cables at the outer ring of the vessels (Section 11.5.5) are designed to assure good connection of the cable shields to the vessel inside.

The reference ground for the low voltage power is the ground plane of the peripheral electronics boards. Each peripheral board is locally floating. The planes will individually be electrically connected to the vessel ground plane at selected places avoiding ground loops. The sensor modules and ALTIROC ASICs are floating and will be referenced to peripheral board ground through the analog ground plane of the flex cables (Section 7.3.2).

The sensors for the Detector Control System (DCS) system (Section 10.4), e.g temperature probes on the cooling plates, are electrically floating and connected via cables to DCS units mounted inside the extension of the Faraday cage at the patch panel areas. Connection to the experimental cavern is via optical fibers or optocouplers to maintain the shielding. The same DCS units will also supply the enable signals to the 1.2 and 2.5 V DC-DC converters for powering the peripheral electronics.

The cooling plates inside the vessels will be part of the shielding and electrically connected to the inside of the vessels. This requires the CO_2 transfer line to be electrically insulated at the cooling junction box located on the end-cap calorimeter surface. Shielding for cables will be discontinued appropriately to avoid ground loops.

8.4 Roadmap for power system

The studies of the grounding options for the High Voltage return will continue in 2020 for decision no later than Q3. In parallell, studies of commercial solutions for the power supplies (both HV and LV) will take place in order to prepare for the Specification Review in Q3 2021 Figure 15.4). Tendering will follow in 2022 with subsequent prototype tests for the FDR and PRR. Design and prototype studies of the HV filter and patch boxes is integrated with the studies of the grounding options and the HV power supplies. The final design of the boxes are dependent on configuration of the selected commercial solution and will take place towards the end of 2022.

9 Peripheral Electronics

The on-detector peripheral electronics transfer data between the detector modules and the DAQ system, the luminosity system as well as the Detector Control System (DCS). It also has a central role in the monitoring of sensor temperatures and supplied low voltage. The peripheral electronics system is based on the CERN-developed lpGBT ASICs [59]. The modules are connected via flex cables (see Section 7.3.2), while signals to and from the DAQ and the luminosity systems are transferred on optical fibers. On these fibers the DCS data and commands are embedded in the data streams.

Each flex cable serves a module consisting of two ALTIROC ASICs and contains two differential electrical CERN Low Power Signalling (CLPS) e-links transmitting timing data at different rates (320 Mbit s^{-1}, 640 Mbit s^{-1}, or 1.28 Gbit s^{-1}) depending on the ALTIROC position. Flex cables for modules placed at a radius above 430 mm carry a further two differential e-links at 640 Mbit s^{-1} with luminosity data. Each cable also contains four e-links with clock and fast DAQ commands to the ALTIROC ASICs with a bandwith of 320 Mbit s^{-1}, as well as the lines for the ALTIROC low voltage power supplies, control signals and the bias voltage of the sensor. The digital output data from several ALTIROCs are merged in lpGBTs on peripheral electronics boards (PEB) and transmitted on optical fibres to the off detector DAQ system. Control and configuration commands to and from the ALTIROC ASICs are transmitted via I^2C bus. The I^2C bus information is embedded in the data streams between the lpGBTs and the detector DAQ system. An overview of the HGTD readout chain is presented in Figure 9.1.

The peripheral electronics also include the 10 V to 1.2 V DC-DC converters for the digital and analogue voltages supplied to the ALTIROC ASICs. The supply voltages are monitored using the internal multiplexed ADC on the lpGBTs. The ADCs are also used to measure actual voltages received by the ALTIROC, as well as the sensor temperatures. The voltages to be measured are selected by analog 64-to-1 multiplexers mounted on the peripheral boards as described in Section 9.3.

The lpGBTs that are used for transmitting luminosity data do not a priori need to receive downlink data via optical fibres and will thus only send data to the off-detector luminosity backend electronics. These lpGBTs will receive the required 40 MHz clock from lpGBTs connected to the off-detector DAQ system.

A schematic block diagram of the PEB electronics for one module connected to off detector electronics is shown in Figure 9.2.

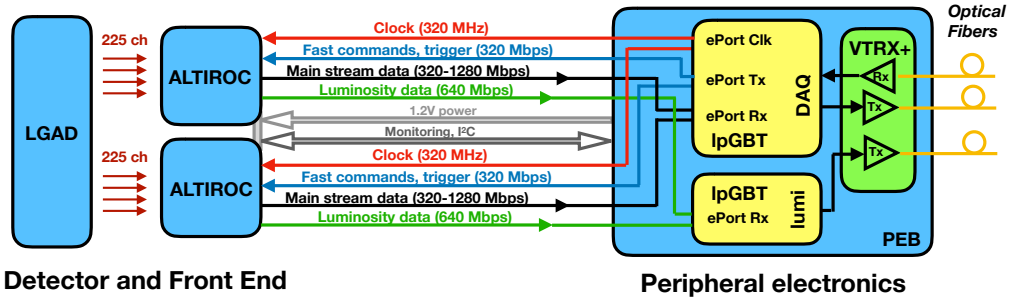


Figure 9.1: Upstream and downstream data flow. The e-links transmit data, fast commands and clocks between the ALTIROC ASIC and the lpGBT. VTRX+ is the Versatile Link+ optical module. I^2C-bus, ASIC control and monitoring lines from the ALTIROC are also shown. Only one module (2 ALTIROC ASICs) is shown. Up to 8 modules are connected to the same DAQ lpGBT. Each of these lpGBTs uses one Rx port at 2.56 Gbit s^{-1} and one Tx port at 10.24 Gbit s^{-1} of the VTRX+. The luminosity lpGBTs uses one TX port of some of the the VTRX+.

As introduced in Section 2.3, each HGTD vessel contains two cooling disks (shown in Figure 2.4), with detector modules mounted on both sides, thus having two instrumented layers per disk. The baseline design is to have five Peripheral Electronics Boards (PEBs) per quadrant and per side of each cooling disk, as shown in Figure Figure 9.3. Such a layout yields 80 boards per HGTD vessel, thus 160 boards in total. Each board covers three or more readout rows in order to have a similar number of ALTIROC ASICs connected per board (typically 106-110). This optimizes the use of the lpGBTs by sharing across readout rows.

All the active components of the peripheral boards will be located at radii above 700 mm. Extrapolating from Figure 2.14, the maximum expected fluence, which these components have to withstand, will be below 1×10^{15} n_{eq} cm^{-2} and the TID below 0.2 MGy.

9.1 Data transmission

The data transmission between lpGBTs on the peripheral electronics and the off detector systems uses optical fibres based on the VTRX+ optical transceiver developed within the Versatile Link Plus project [86]. The bandwidth required for the digital data output from the ALTIROC ASICs is given by the number of pads hit in an event. The expected average number of hits depends on the radius of the module position. The hit rate has been studied using simulations and the results are presented as average occupancy per ASIC for an $\langle\mu\rangle = 200$ in Figure 9.4. The radial dependency is clearly seen with a maximum of just below

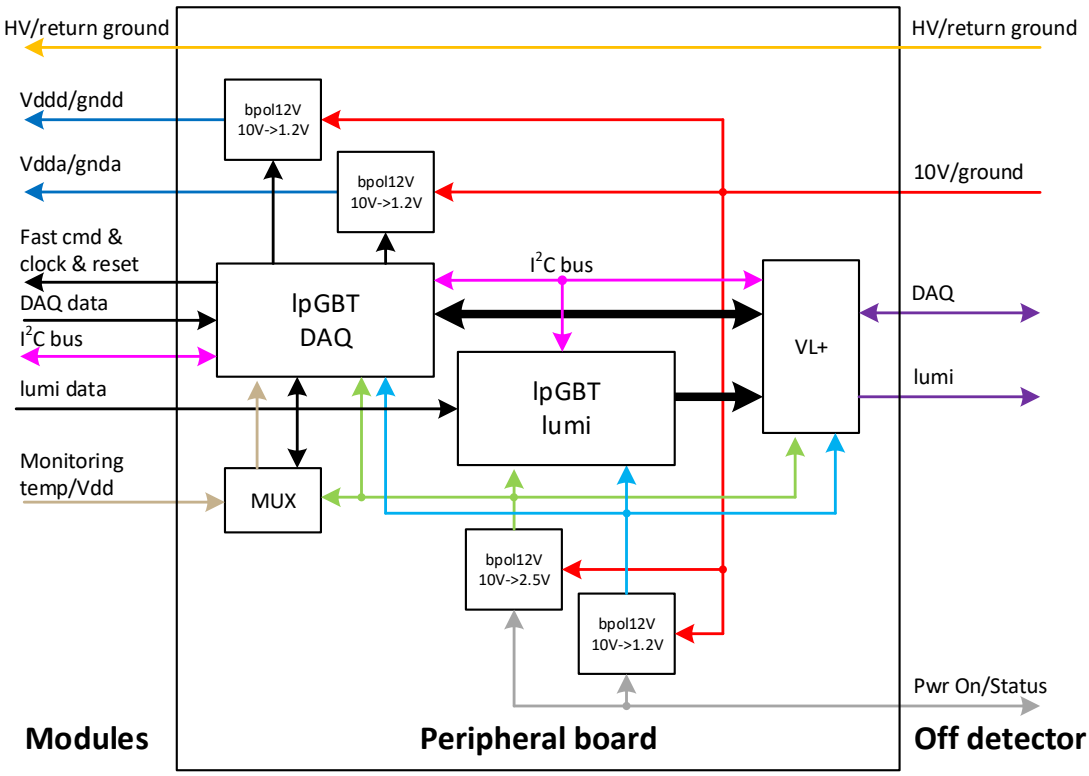


Figure 9.2: Block diagram of the peripheral electronics for powering and readout of a module. Multiple modules are connected to the same DC-DC converters and lpGBTs. The brown lines indicate voltages measured by the multiplexed ADC on the lpGBT. Light blue lines are low voltage (1.2 V) power supplied from bPOL12V DC-DC converters. Light green lines are 2.5 V. The 2.5 V is connected to the lpGBTs only to measure the voltage for monitoring purpose. The thin black lines are control signals via the general purpose I/O ports of the lpGBTs. The thick black lines are high speed electrical links to and from the VL+ optical module. Other lines are explained in the figure.

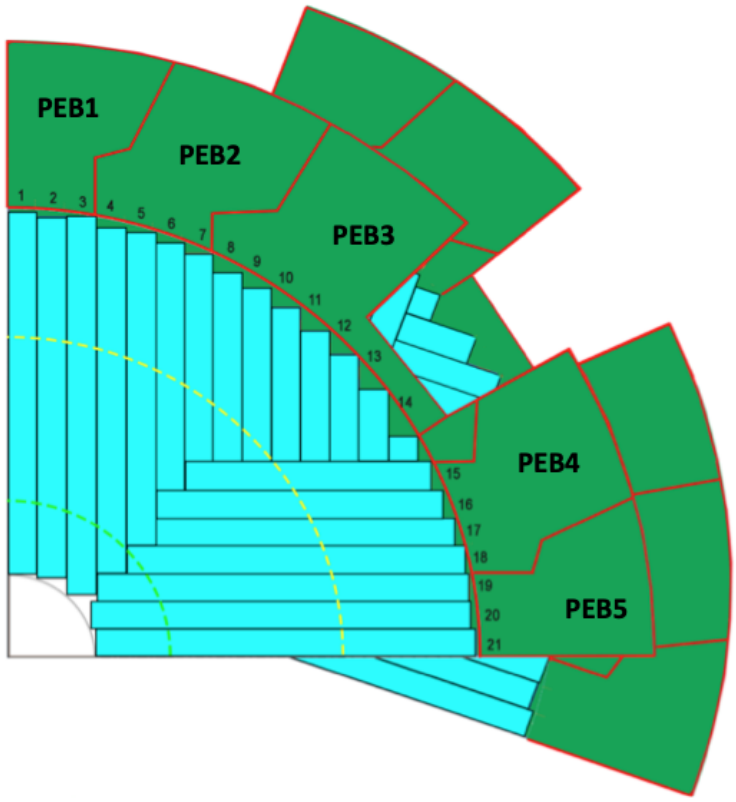


Figure 9.3: One quadrant of the HGTD front and back disk is shown. The PEBs (in green) are attached to the readout rows (numbered from 1 to 21).

8 % at the innermost radius. Such an occupancy can be accommodated within the maximum available rate for the data from the ALTIROC, which is 1.28 Gbit s^{-1}.

For larger radius the bandwidth per e-link can be reduced, using 640 Mbit s^{-1} at radii above 220 mm and 320 Mbit s^{-1} at radii above 405 mm. These rates are chosen in order to minimise the numbers of lpGBTs and optical links while keeping the average bandwidth usage below 55% for the expected number of average hits per ASIC at a readout rate of 1 MHz. The average bandwidth usage for each module in a quadrant of the first double sided layer is shown in Figure 9.5.

In addition to the readout output e-link each ALTIROC ASIC requires a 320 Mbit s^{-1} fast command input e-link to supply both the bunch crossing information and the DAQ commands. A 320 MHz clock extracted inside the lpGBT from the command data packages is also sent to each ASIC. In the regions of readout rows 4 to 18 (Figure 9.3) the minimum number of lpGBTs used is given by the required number of fast command links. In these cases, readout e-links at 640 Mbit s^{-1} are available for higher radii than mentioned above, resulting in increased bandwidth capacity as seen in the figure.

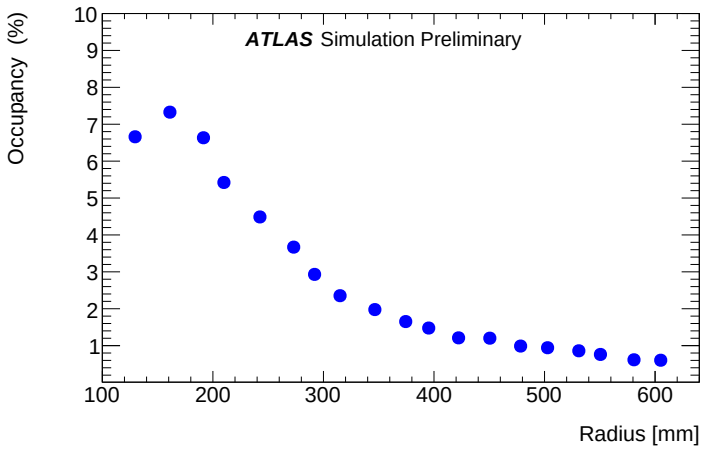


Figure 9.4: Average occupancy of an ASIC as function of radius in a simulated sample with $\langle\mu\rangle = 200$.

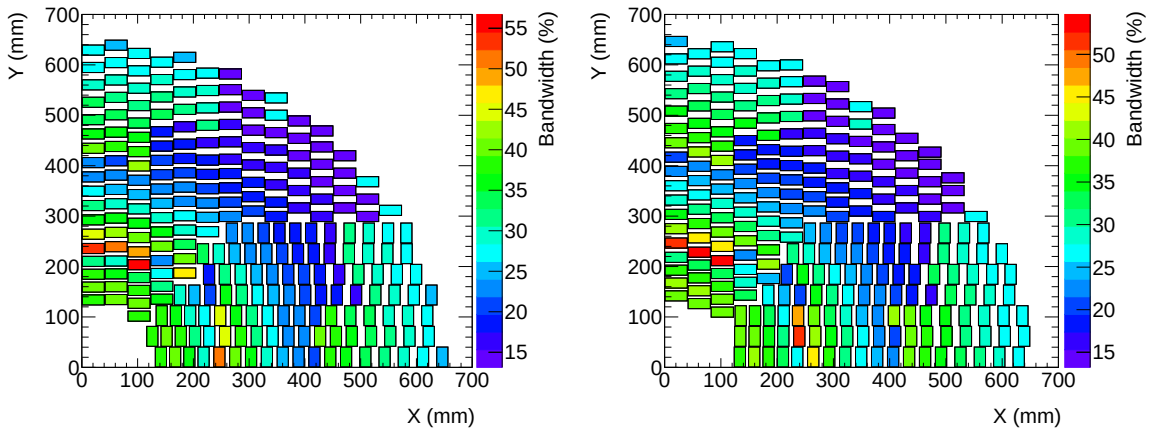


Figure 9.5: Readout bandwidth usage, in % of the capacity, for the expected number of average hits per ASIC from simulations with $\langle\mu\rangle = 200$ at a readout rate of 1 MHz. The usage is shown per sensor module in a quadrant of the first double sided layer. Left: Front layer. Right: Back layer. In regions corresponding to the readout rows 4-18, higher bit rate capacity is available (see text) resulting in lower bandwidth usage.

9.1.1 LpGBT

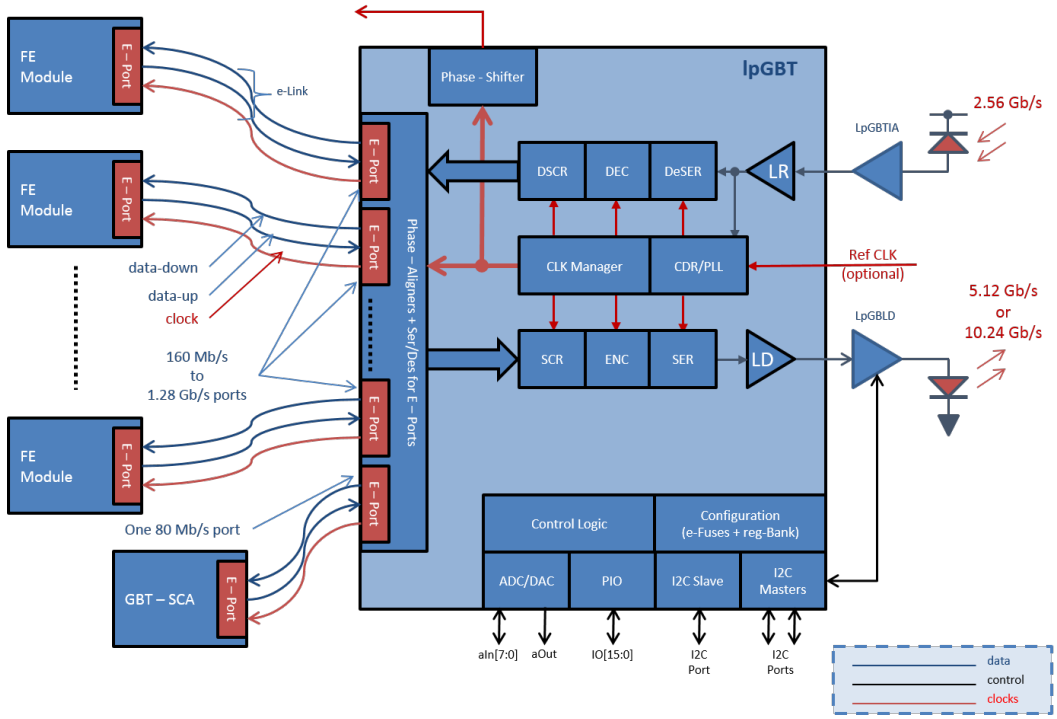


Figure 9.6: Block diagram of the lpGBT ASIC.

A block diagram of the lpGBT is shown in Figure 9.6 and more details concerning its specifications can be found in [59].

The lpGBT ASIC is able to transmit data to an optical link at 10.24 Gbit s^{-1}. When using FEC5 error correction code the bandwidth can be shared by 7 groups of 32 bit data received on differential (CLPS) e-links. The 32 bits can come from one 1.28 Gbit s^{-1}, two 640 Mbit s^{-1}, or four 320 Mbit s^{-1} e-links. The phase aligner circuit for each input e-link of the lpGBT will be used to ensure that the received data is sampled by the lpGBT at the optimal phase. This allows data from flex cables with different lengths to be connected to the same lpGBT. The total package length of the transmitted data, including headers, error correction codes, and 2 bits of internal and 2 bits external DCS data, is 256 bits, that are transmitted at a rate of 40 MHz.

Each lpGBT is able to receive four independent 320 Mbit s^{-1} bit streams encoded in the 2.56 Gbit s^{-1}, 64 bit frame, down-link data from an optical link. Each package includes headers, FEC12 error correction bits as well as 2 bits internal and 2 bits external DCS data.

The lpGBTs require configuration commands for setting up registers controlling their behaviour, e.g. bit rates and phase shift adjustment. This is normally done through their I^{2}C bus slave port, however to avoid external I^{2}C bus cables, the lpGBTs receiving data via optical

links on each peripheral electronics board will be programmed by e-fuses to receive their configuration via the 2.56 Gbit s^{-1} downlink bit stream. The same lpGBTs will in turn be used to configure the lpGBTs for the luminosity system of the same board via one of their I^{2}C bus master ports.

- **Fast commands and clock distribution**. Each data package received by the lpGBT via the optical links contains up to four independent 320 Mbit s^{-1} data streams. These can be mirrored to four different outputs of the lpGBT, allowing one lpGBT to control 16 ALTIROC ASICs using 8 bit words. The 320 MHz clock required by the ALTIROC is extracted from the data streams by the lpGBT and distributed to the ASICs on individual clock streams. Preliminary measurements done by the CERN lpGBT group show an excellent random component of the jitter (2.1 ps) but a sizeable deterministic part. The minimum number of lpGBTs required for the peripheral electronics is defined by the above limitation of not more than 16 ALTIROC ASICs connected to the same lpGBT.

- **DAQ data**. The different e-link bit rates 1.28 Gbit s^{-1}, 640 Mbit s^{-1}, and 320 Mbit s^{-1} allow for an average number of hits per bunch crossing and per ALTIROC at $\langle\mu\rangle = 200$ of up to about 41, 20 and 9, respectively, at 1 MHz of event readout. The number of lpGBTs must be kept to a minimum, in view of the limited space available for the peripheral electronics.

- **Luminosity data**. Each ALTIROC ASIC at radii larger than 430 mm provides 16-bit luminosity data for each bunch crossing, transmitted to the lpGBTs via the flex cables. The 430 mm results from using all available e-links of the minimum number of lpGBTs that allows all modules in the outer ring at $r >$470 mm to be included. Two 640 Mbit s^{-1} e-links are merged into a 32 bit lpGBT group, allowing 14 luminosity e-links to be connected to a single lpGBT for transmission to the off detector electronics via an optical link. In the baseline design, no downlink data is foreseen for these lpGBTs, which will be operated in simplex transmitter mode. The clock signal will instead be obtained as a 40 MHz clock from the DAQ lpGBTs. Operation parameters and controls for the luminosity lpGBTs, e.g. phase adjustment delays, are set up via the I^{2}C bus also from the DAQ lpGBTs.

- **I^{2}C bus**. Each lpGBT has three I^{2}C bus masters and one slave. Only the master ports on the DAQ lpGBTs are used since the luminosity lpGBTs do not receive optical downlink data. One I^{2}C bus master will be connected to up to eight ALTIROC ASICs on four modules for DCS control. Since these I^{2}C-buses will only be used for configuration, traffic will be minimal during data taking limiting the risk of generating noise inside the ALTIROC ASICs. I^{2}C-bus master ports are furthermore used to configure the laser drivers of the optical links and, as previously mentioned, to configure all lpGBTs of the luminosity readout.

To load the initial configuration, which will be fused into the lpGBT registers, connections for an external I^2C-bus to the peripheral electronics is foreseen.

9.1.2 Optical links

Each lpGBT connected to the DAQ system will need one up and one down optical link, while the lpGBT connected to the luminosity readout will only need an uplink to the off detector system. The VTRX+ optical transceivers handle four fibres for transmission and one for reception. The dimensions are specified as a $20 \times 10\,\mathrm{mm}^2$ footprint. The specified radiation hardness, 1 MGy and $1 \times 10^{15}\,\mathrm{n_{eq}\ cm^{-2}}$, exceed the required levels at radii greater than 85 cm, where they will be located. The VTRX+ modules are pluggable via electrical connectors and are delivered with a pigtail ending in a 12 fibre MT type optical connector.

9.2 DC-DC converters

The peripheral electronics will contain DC-DC converters based on the bPOL12V Point Of Load regulators. These converters supply the 1.2 V required by the ALTIROC ASICs and the lpGBT ASICS. The Versatile Link plus require 2.5 V for the laser driver and limited current at 1.2 V for the receiver. The DC-DC converters use the bPOL12V ASIC developed at CERN. The bPOL12V will be used as a single stage converter from the 10 V input to the 1.2 V output (or 2.5 V for the laser driver). The motivation for this choice, rather than a dual stage converter that potentially has higher efficiency, is the limited surface available for the peripheral electronics.

The maximum output current of the bPOL12V is 4 A. Measurements on prototypes by the developers indicate that an efficiency up to 72% at $-30\,^{\circ}\mathrm{C}$ and 3 A current can be achieved. When operating near the maximum current the input voltage should not exceed 10 V to reduce switching transients. The ASIC is designed for radiation tolerance up to 2 MGy and $2.5 \times 10^{15}\,\mathrm{n_{eq}\ cm^{-2}}$. The converters need a 460 µH inductance as well as further filtering components. A printed circuit board layout exists, which is adapted from the layout optimized by the bPOL12V developer team. The footprint of this layout, 11 mm × 30 mm, is however still considered to be large compared to the available space. The feasibility of a reduction by the choice of the design of the inductor and shielding cage of the converter is under investigation.

The analogue and digital voltages are supplied separately to the ALTIROC ASICs. Each ALTIROC requires at most 0.5 W analogue and 0.7 W digital power. The two ASICs on the same module share supplies. The current consumption is dependent on the average number of hits within an ASIC and thus has a radial dependence. Separate DC-DC converters will be used to supply the analog and the digital parts of the ALTIROCs. Each DC-DC converter will

supply 6 ALTIROC ASICs (3 modules). The power consumption of an lpGBT will not exceed 0.55 W (0.45 W for lpGBTs of the luminosity readout due to only uplinks being used).

To power on the electronics for one PEB the DC-DC converters supplying the PEB itself, i.e. the lpGBTs and the optical links, are first switched on by an external 1 V enable signal. The status of these converters is read out via external electric cables (open drain) and on general purpose I/O-lines on lpGBTs other than those they supply. This allows to differentiate between possible power failures and lpGBT failures. The external signals are controlled via EMCI units [87] from the DCS system. Care is taken to avoid that the external electrical cables violate the grounding and shielding rules.

Once the lpGBTs and VTRX+ are powered on, the DCS bits i the lpGBT data stream can be used to control the DC-DC converters supplying voltages to the ALTIROC ASICs. The DC-DC converters are switched on by applying a voltage (at least 850 mV) which is generated via general purpose I/O-lines from DAQ lpGBTs while the status of the converters are reported via their open drain Power Good output and monitored via lpGBTs.

9.3 Control and monitoring

The DCS control and monitoring of the front-end electronics, the monitoring of the sensor temperature and the delivered and received low voltage of the electronics is handled through the lpGBTs. The DCS information is embedded in the up and down bit streams of the optical connections at a rate of 80 Mbit s^{-1}. Two bits per data package at 40 MHz, in both directions, can be used for the general purpose I/O-port, ADC or I^2C bus masters of the lpGBT. Since the lpGBTs of the luminosity system will not have optical downlinks, only the lpGBTs connected to the DAQ system will be used for DCS handling in the baseline option.

Each flex cable will, as described in Chapter 7, carry 5 voltages: temperature of the sensor from each of the two ALTIROC (voltage from the temperature sensor); analogue and digital supply voltages received at the ALTIROCs and the analog current return voltage at the module. Due to the resistance of the conductors on the flex cable, the latter three voltages serve to measure the current consumption and detect latch-up. Each lpGBT has an 8 input 10-bit multiplexed ADC allowing 1 mV resolution for a 1 V range. To handle all voltages to be measured, a 64:1 multiplexer (see Section 6.8) is used. Each multiplexer, which can switch the received voltages from up to 12 modules, is controlled by 6 I/O lines from an lpGBT.

As mentioned above, the peripheral electronics boards will each receive an external control signal to switch on the DC-DC converters supplying the lpGBTs and the optical links. The status of these converters is read out on I/O-lines on lpGBTs other than those they supply to allow to differentiate between possible power failures and lpGBT failures. Further I/O-lines on the DAQ lpGBTs are used for switching on and monitoring the status of the DC-DC converters supplying voltages to the ALTIROC ASICs.

The I^2C bus will be used to control and configure the ALTIROC ASICs as well as to configure the lpGBTs.

9.4 Connectors

The limited space available for the peripheral electronics puts severe constraints on connectors. The PEB ground will be connected to the reference ground of the detector vessel.

- The flex cables from a readout row will enter the peripheral electronics in bundles of up to 19 cables. Each flex cable is 36 mm wide. Several options for connecting them to the PEBs are under study. Limitations on the available space, both concerning height and footprint on the boards, put severe restrictions on solutions. Furthermore, the reliability of the connection is an important consideration. The baseline choice is to integrate up to 6 flex cable ends from modules in the outer ring directly in a rigid flex part of the PEB. This is illustrated in Figure 9.7.
- All modules have individual high voltage supplied through the PEB via the flex cables to the modules. Commercial 56 pin connectors specified to sustain operation up to at least 800 V will be used on the PEBs for connection from the vessel feedthroughs. In the baseline design each sensor module has individual HV return connection requiring two connectors per PEB to connect both supply and return to each of the up to 55 modules. An option, in which a common HV return is connected to the ground plane of each peripheral board, is being studied in order to reduce the number of cables. The HV at each module is then referenced to ground through the analog ground plane on the flex cables at the module end.
- The peripheral boards will each require 2 cables with 10 V for the on-board DC-DC converters. Suitable connectors are under study.
- The optical fibre pigtails of the VTRx end in a 12 fibre MT-type connector to which the patch cables of the fibre feed-throughs at the detector vessel have to be connected.

9.5 Peripheral boards

9.5.1 Physical limitations

The available physical space for the peripheral electronics is very limited. It is constrained in the radial direction by the end of the instrumented area and the limit of the HGTD vessel, therefore ranging from 660 to 920 mm. Because the allowed thickness of the HGTD is only

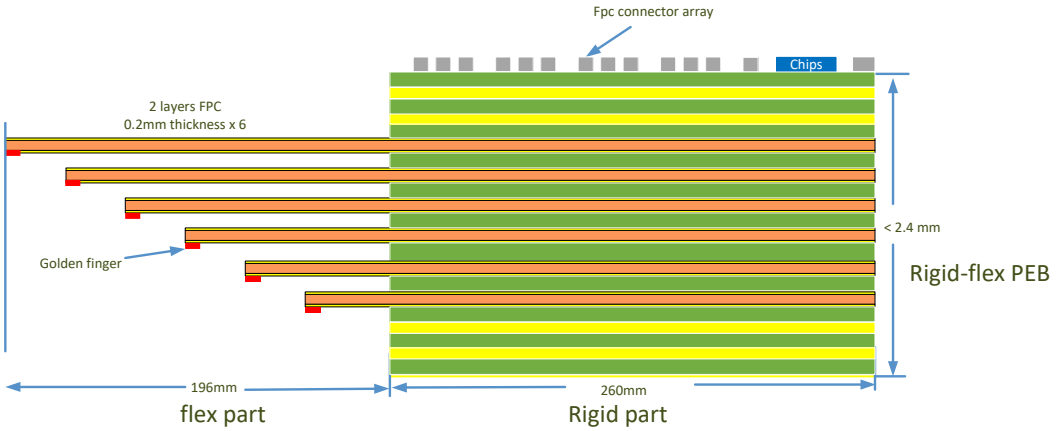


Figure 9.7: Illustration of the rigid flex concept. The ends of the flex cables (orange) are part of the printed circuit board.

75 mm, the space available for the electronics in the z-dimension is also very small: 9 mm with a 1 mm margin.

9.5.2 Layout considerations

The peripheral electronics will be split up in five peripheral boards (PEB) per quadrant with a similar number of sensor modules connected per board. This is achieved by combining the readout rows 1–3, 4–7, 8–14, 15–18, and 19–21, see figure Figure 9.3 for the row numbering convention, into one board each. (The readout row numbering is shown in Figure 7.13). This combination allows the reduction in the number of lpGBT ASICs, multiplexers and VTRx used and would lead to a better use of the available surface area at the outer radius of the disks for connectors.

The number of modules, e-links for DAQ at different transfer rates and luminosity e-links per peripheral boards are shown in Table 9.1. The bit rates of the DAQ e-links will be re-optimised for the final layout based on further simulations. The number of lpGBT ASICs per PEB is given by the limitation, that there are only 16 e-links to transmit DAQ commands to the ALTIROC ASICs per lpGBT. The bit rates for readout are selected to be as high as possible given the available capacity.

A number of considerations have to be taken into account for the actual PEB design.

- To limit the implications of possible failing components, care must be taken in the layout design such that as few detector modules as possible are affected. The modules have to share, as far as possible, the same lpGBT for readout as the one that also transmit their clock and fast commands, control their DC-DC supply as well as DCS

Peripheral board	Readout rows	Nb of modules	1.28 Gb/s	640 Mb/s	320 Mb/s	Luminosity
Front						
1	1-3	55	18	32	60	42
2	4-7	53	8	66	32	56
3	8-14	36	0	68	4	70
4	15-18	53	8	66	32	56
5	19-21	53	20	30	56	42
Back						
1	1-3	55	18	26	64	42
2	4-7	53	10	60	36	56
3	8-14	37	0	72	2	70
4	15-18	53	8	66	32	56
5	19-21	54	20	18	44	42

Table 9.1: Number of module readout e-links at different rates for the different peripheral boards of the front and back layers. The tables also show which readout rows are connected to which board and the number of modules per board.

control via I^2C bus and handle the module voltage monitoring. Optimised schemes for this exist and will be implemented.

- For the same reason, modules sharing the same readout lpGBTs should share luminosity lpGBTs, that will receive their configuration control and clock from the readout lpGBTs.

- The power dissipation of the peripheral electronics is used to preheat the CO_2 cooling requiring a suitable radial arrangement of the DC-DC converters and lpGBTs.

9.5.3 Layout

Combining multiple readout rows on the same PEB, as described above, the required number of different components is worked out per PEB, as shown in Table 9.2.

A conceptual design of the first two PEBs (Front 1 and Front 2) is shown in Figure 9.8. A functional prototype of the PEBs is scheduled to be produced by 2020 as part of the demonstrator program Section 14.3 in parallel with a complete PCB design of a real size PEB for early 2021.

Figure 9.8: Conceptual design of a PEB 1 and 2 for the front layer. The yellow parts are the flex cable tails. The blue rectangle are the VTRX+ transceivers. The white squares are the lpGBTs. The flex cable connectors are in two parts, one for signals av LV power and a smaller one for HV. The MUXes (red squares with white rectangular band) and the DC-DC converters (red blocks of components) are also visible

Peripheral board	lpGBTs DAQ	lpGBTs Luminosity	DC-DC converters	MUX	VTRx
Front					
1	8	3	42	5	8
2	7	4	40	5	7
3	5	5	28	4	5
4	7	4	40	5	7
5	8	3	40	5	8
Back					
1	8	3	42	5	8
2	7	4	40	5	7
3	5	5	30	4	5
4	7	4	40	5	7
5	8	3	40	5	8

Table 9.2: Numbers of lpGBTs, analog multiplexers, VTRx, and DC-DC converters for the different Peripheral Electronics Boards.

9.6 Power dissipation

The peripheral electronics will be in thermal contact with the cooling plates acting as pre-heaters for the CO_2 cooling system (Section 11.3). The dominant source of the power dissipation on the PEBs is the power loss in the DC-DC converters. With an average power consumption of 1.1 W per ALTIROC ASIC (Figure 11.4), the total required power delivered to the ASICs from the DC-DC converters including power losses in the flex cables will be 4.9 kW per double sided layer. With 72% efficiency at −30 °C and 3 A current, the power loss due to the ASIC supplies will be 1.9 kW per double sided layer. Including an estimated power consumption of 300 W per double sided layer for the lpGBTs and VTRx, the total power dissipation of the peripheral electronics will be 2.2 kW per double sided layer. The total power dissipation will be 4.4 kW per end-cap. Since most of the detectors components do not yet exist, a careful re-evaluation of the expected power dissipation will be done based on the first prototypes.

9.7 Roadmap towards PEB production

Following the prototype of the PEBs in the demonstrator program, a complete PCB design of a real size PEB is expected to be released at the beginning of 2021 after the SPR. It may be followed by a second real size prototype with minor modifications in 2021, after the PDR, that is expected in Q3 2021 as indicated in Figure 15.4. The pre-production of the full-size PCB is expected between May 2022 and November 2022, just after the FDR review. The

PRR, expected in Q4 2022, should give the green light for the final PCB production. The production, including the QA to be done by the Institutes is expected to take place between December 2022 and March 2025.

The PEBs use the rigid-flex PCB technology to integrate up to 6 flex cable ends from modules in the outer ring directly to save the space. After pre-production, a burn-in test will be performed to evaluate the product life and to identify any potential problems. During the mass-production, an accelerated ageing temperature test will be performed in batches to find the early failure.

10 DAQ, calibration, luminosity and control

10.1 DAQ interface

The HGTD data acquisition system will be embedded in the ATLAS DAQ common readout. The proposed HGTD architecture is shown in Figure 10.1 and can be divided in two main blocks: on-detector electronics located in the experimental hall and off-detector electronics located in the USA15 counting room. The on-detector electronics consist of ALTIROC modules connected via flex cable to the Peripheral Electronics Board, as described in Chapter 9.

The interface between on-detector and off-detector electronics is performed via optical links using lpGBT chip set and VTRx+ optoelectronics, which provides different data paths for Timing, Trigger and Control (TTC), DAQ and DCS. Two optical links with different purpose data streams are proposed: the main data stream that provides time-over-threshold (TOT) and time-of-arrival (TOA) information per triggered event and the luminosity stream that contains bunch-by-bunch hit information for luminosity measurements. These two different data streams are needed in order to disentangle the standardized format for the ATLAS dataflow driven by the main data stream and the custom luminosity stream which requires different processing. The main data stream is used for the propagation of clock, fast commands and configuration to the modules, as well as the data information for the ATLAS event processor. The luminosity stream only sends hit information through the uplink and will be described in Section 10.3.7.

10.1.1 Off-detector electronics

The off-detector electronics is based on the general-purpose FELIX system [88], which is the main interface between the off-detector back-end and the on-detector electronics. The proposed back-end architecture is shown in Figure 10.2. FELIX receives event data from the on-detector electronics and transmits them to the Data Handler via multi gigabit network. In addition, FELIX interfaces to the TTC system via the Local Trigger Interface (LTI) and to DCS for control, configuration and monitoring.

The FELIX downlink will follow lpGBT encoding, which is composed of 64-bits frames that are transmitted at every LHC bunch crossing period with a data rate of 2.56 Gbit s^{-1}. The

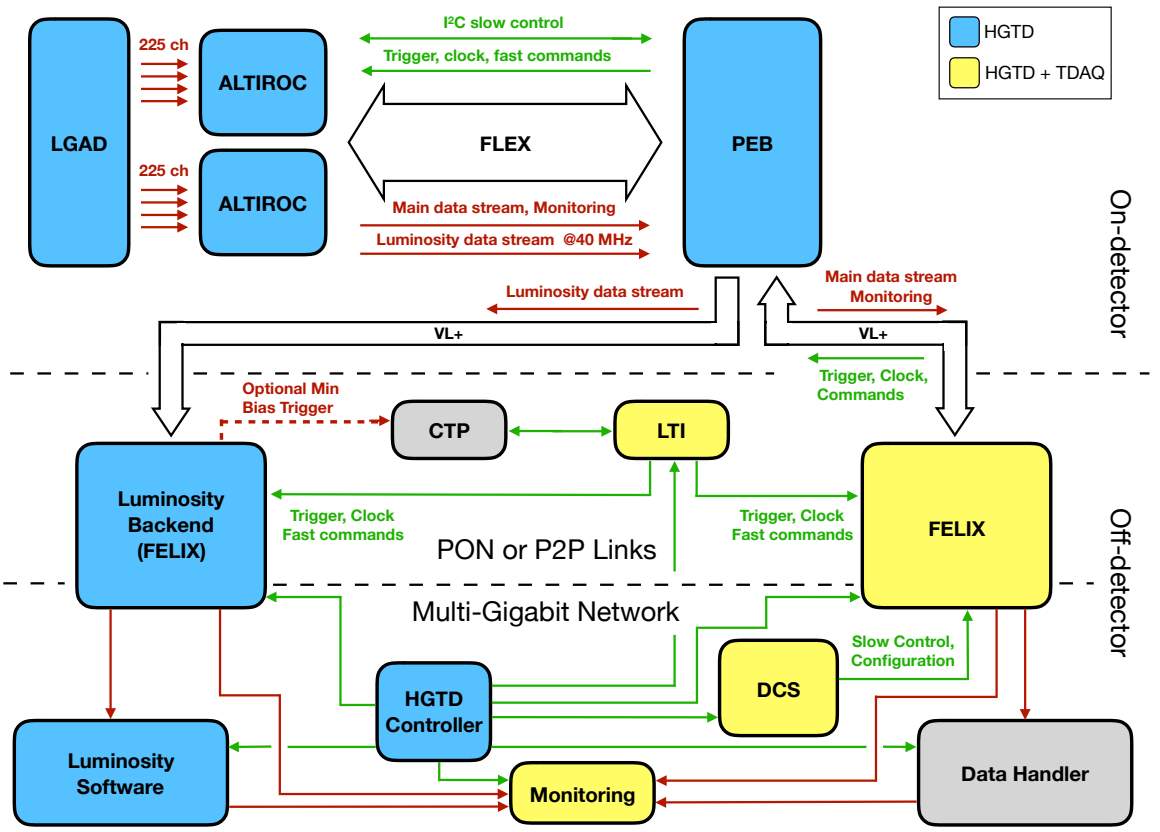


Figure 10.1: Data transmission paths among the ALTIROC, Peripheral Electronics Board (PEB), and DAQ components for hit data, luminosity data, clock, fast commands, and DCS/slow controls.

clock is propagated to the lpGBT and thus to the modules by sampling the data stream. The downlink frame has different fields for data (fast commands), internal and external configuration meant for lpGBT, module and DCS handling. The uplink will also follow lpGBT encoding with a data rate of 10.24 Gbit s^{-1}, the different frame fields for data, configuration and DCS will be decoded in the FELIX board. Upstream, the data will be forwarded to the Data Handler using multi-gigabit network. In addition, monitoring information, like errors and timing will be computed in FELIX and will be sent to the monitoring unit together with a prescaled sample of the events. The monitoring unit will receive specific HGTD data via multi-gigabit network, it will decode and compute HGTD monitoring information that will be included in the global ATLAS on-line monitoring.

The Data Handler will receive data from FELIX via a multi-gibabit network. It will decode HGTD specific information providing event building and monitoring within a common DAQ infrastructure [58]. The data will be sent to the Dataflow system for further processing by the Event Filter. The event size is estimated to be 250 kB on average, with a range between

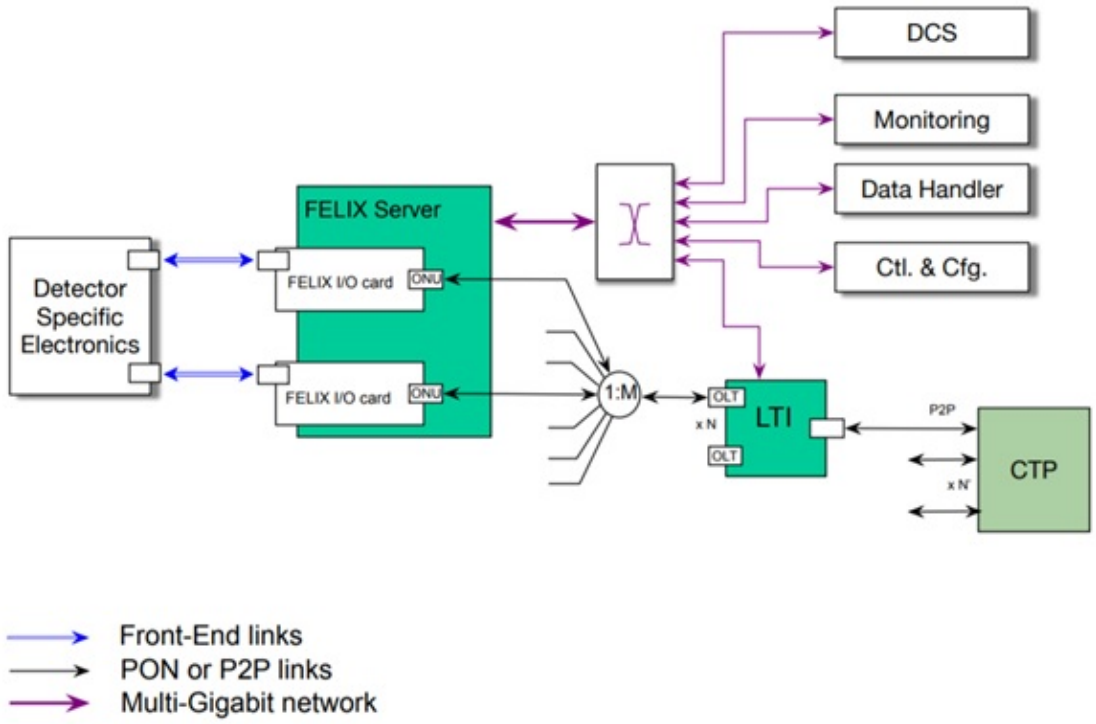


Figure 10.2: Proposed off-detector back-end architecture for Phase II. Plot taken from [58].

150 and 350 kB. In addition, the Data Handler will also receive trigger information via FELIX for monitoring and automatic recoveries. On the other hand, a software application called HGTD Controller, running in a dedicated computer will be used to manage the module and lpGBT configuration. The Controller will be also used to manage HGTD calibrations via dedicated software that will be described in the following section.

Requirements on the number of FELIX boards are set by the number of optical links and it is driven by the HGTD layout. Current estimates call for a total of 48 FELIX I/O cards and 48 Data Handlers for the main data stream. The luminosity back-end electronics will require 32 FELIX I/O cards.

10.1.2 Calibration and timing

Regular calibrations will be performed in HGTD in which different parameters like TOA and TOT will be monitored and tuned. A dedicated HGTD software running on the HGTD Controller will be used for this purpose. The calibration procedure is shown in Figure 10.3, it will consist of different nested loops with a specific module configuration followed by

a calibration pulse and a trigger command with a proper timing. The HGTD Controller will interface with FELIX for the handling of the module configuration and generation of a particular bitstream for the fast commands. Downstream the event data will be processed and stored for a further analysis and may be used as input inside the nested loop for tuning purposes.

During the calibration procedure 3.4 million of electrical channels have to be readout, which correspond to 11 MB per event and can surpass TDAQ infrastructure limitations for HGTD. In order to reduce the event size, several pixels inside and ASIC will not be read-out during the calibration loop using a particular mask pattern, the so-called mask step. The mask step will be added as a nested loop inside the calibration procedure as shown in Figure 10.3, in which several pixels will be masked in every step loop. For instance, while using 45 mask steps, which correspond to 5 pixels being readout per ASIC at every trigger, the event size can be reduced to 250 kB per event and thus matching HGTD specifications. Another limitation might arise from the data processing in the Data Handler, which can be avoided by optimizing the time delay between two consecutive triggers, however it may slow down the calibration procedure. The implementation of a histogramming unit inside FELIX will help to overcome these limitations by speeding up the calibration procedure and will be investigated. Nevertheless, the calibration procedure described before for entire HGTD should not last more than 5 minutes.

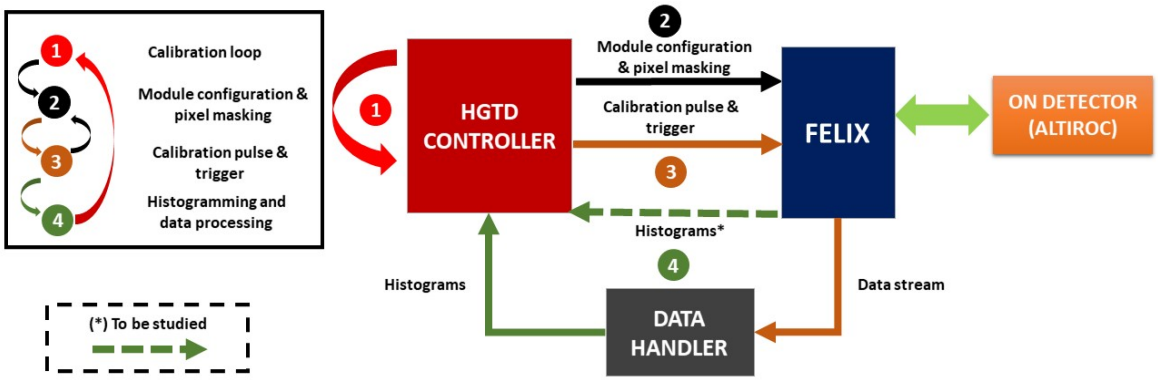


Figure 10.3: Diagram of the proposed calibration procedure for HGTD.

Accurate timing for the modules is critical for operation. For this purpose, dedicated timing calibrations will be performed. In the first stage, the detector timing will be adjusted using a standard calibration procedure. It will consist of injection of different charges while looping over coarse and fine delay DAC values of the TDC. The quantisation step of the TOA in the TDC is 20 ps inside the 2.5 ns readout window. Moreover, the phase shifter inside the ALTIROC ASIC with a 100 ps resolution and 8 ns window has to be adjusted. After the calibration procedure, the delay values which set the 7-bit TOA in the middle range will be selected. In a second stage, the detector will be timed during stable beams by using dedicated LHC fills with a small number of isolated bunches, similar to the LHC Run 1 and

Run 2 fills used for Pixel and SCT timing scans performed during the LHC intensity ramp-up period. During these fills different delay values of the phase shifter inside the lpGBT will be scanned. This procedure will allow a global shift of the clock provided by lpGBT with a 50 ps step and 25 ns window. Using this methodology, the timing on the detector can be properly adjusted while the different operational parameters of the TDC remains unchanged. The data will be analyzed offline and the delays that ensure the proper timing will be selected. Further timing corrections taking into account clock jitter variations will be described in the following section.

10.2 Timing correction

Despite regular calibrations and timing adjustments of the detector, dynamic and static contributions to the clock have to be taken into account and will be described in this section. The master clock will be distributed to the lpGBT downlinks and then to the individual ALTIROC readout chips, in which a clock tree will be used to distribute the clock as uniformly as possible. Any temporal or spatial variation in the time discriminator may compromise the ultimate resolution of the detector unless it is understood and controlled.

The sensors themselves will have a resolution as good as 35 ps per hit, as described in Chapter 5. The contributions to the time resolution from the on-detector electronics (UX15) and from the clock distribution (USA15) has to be smaller than 35 ps. For instance, the clock dispersion for HGTD should be less than 15 ps across a wide range of frequencies and over the detector acceptance. Static contributions to the timing resolution, i.e. those fixed by geometry or varying on time scales longer than a run, include the time-of-flight and detector alignment; the propagation times to distribute the clock to each ASIC as a whole; and non-uniform clock propagation paths within an ASIC to each TDC. Dynamic contributions, like the variation of the clock with time, can occur through a variety of mechanisms across a wide range of frequencies, including high-frequency jitter, noise in the flex cables, and low-frequency day/night temperature changes. These effects must be monitored and calibrated to minimise static and dynamic contributions to the timing measurements. In the case of dynamic contributions, sufficient data may not be recorded to calibrate away fast effects, and therefore in this section we study how to determine the timing correction in real-time using all of the data flow.

For relativistic particles produced in an LHC collision, the time-of-arrival distribution of each measurement channel will consist of a Gaussian core derived from the time dispersion of the LHC collisions convolved with the combined hit time resolution of the sensor and electronics, as shown in Figure 10.4. The mean of the distribution encodes information on the relationships between the global LHC clock on arrival to ATLAS, the mean LHC collision time for a given bunch, and the reference clock phase at a given TDC. This mean shifts from zero through the cumulative effects of time-of-flight, clock propagation delays, and dynamic

shifts of the clock phase during data-taking. Assuming that the relationship between the clock at the TDC and the LHC clock is stable within a given time interval, data collected during the interval can be used to sample the t_{hit} distribution and estimate its mean, t_0. This mean can then be used to correct the cumulative time offset of each channel individually.

Assuming a trigger rate of 1 MHz and 100 ms of data collection, the t_0 can be measured with a precision of 8 ps for a single channel at 150 mm radius. If t_0 is calculated on a per-ALTIROC level, combining the hits of up to 225 channels, the same precision can be reached in 2 ms. Integration times are shown in Table 10.1.

Radius [mm]	150	250	350	450	550
$\sigma(t_0)$ after T_{int} = 100 ms for 1 channel	8 ps	12 ps	20 ps	29 ps	44 ps
$\sigma(t_0)$ after T_{int} = 100 ms for 15 × 15 channels	0.6 ps	1.0 ps	1.7 ps	2.6 ps	4.2 ps
T_{int} required for $\sigma(t_0) < 5$ ps for 15 × 15 channels	2 ms	5 ms	13 ms	38 ms	92 ms

Table 10.1: Precision of the t_0 determination, $\sigma(t_0)$, vs integration time T_{int}

10.2.1 Sources of clock jitter

The data path from the ALTIROC up to the DAQ is shown in Figure 10.1 and described in Section 10.1. Diferent contributions to the clock jitter are expected in the readout system:

1. Front-end electronics: the clock distribution within the ALTIROC to each TDC will be shifted due to path-length differences and possible internal jitter. A conservative random Gaussian-distributed 5 ps jitter is included to account for jitter in the ALTIROC.

2. FLEX cable: it is made of Kapton and copper, and it could pick up noise from the environment and might have some inherent time jitter performance. A random Gaussian-distributed 5 ps jitter is included to account for jitter in the FLEX.

3. lpGBT: a preliminary measurement of the lpGBT clock performance in [89] indicated that a large non-Gaussian deterministic time jitter might be expected for the lpGBT. However, any front-end chip with a phase-locked loop can filter this effect to a small 2.2 ps jitter. Both of these scenarios are included in the t_0 calibrations study, and are shown in Figure 10.5.

4. FELIX: the clock jitter from the FELIX system will depend on the final chips used and bandwidth filtering applied, as studied in [90]. A conservative 5.2 ps jitter is added to represent the worst jitter expected for the FELIX.

Additional sources of timing jitter and t_0 variation are expected to affect the t_{hit} measurement and are included in this study. The LHC radio frequency systems which compensate the beam loading and maintain bucket stability result in a periodic collision point time shift

in the ATLAS Detector. The variation in the average time of collision with bunch number was studied in [91], and the expected bunch crossing time offset for the ATLAS detector is included as a bunch-dependent variation. The collision time is expected to shift by a few ps per bunch, but can be corrected to a jitter in the order of 5 ps. Finally, a time–of–flight variation is added as a static radially-dependent offset from 0 to 70 ps as a function of sensor radius.

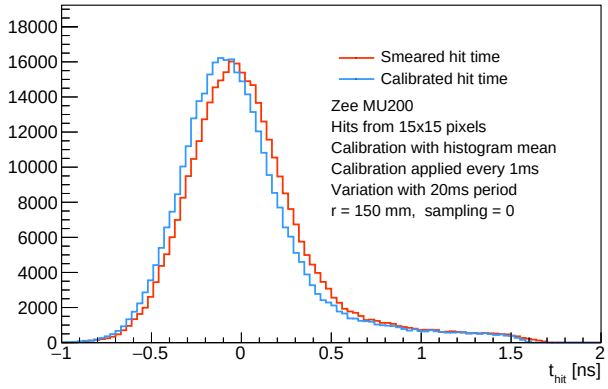


Figure 10.4: Hit time distribution before (red) and after (blue) the timing correction procedure, corresponding to a t_0 of (48 ± 369) ps and (-1 ± 363) ps respectively. The t_0 offset can be recovered after applying the timing correction, while the RMS of the distribution is driven by the time dispersion of the hits in the entire HGTD. Non-Gaussian tails arise from late particles, backscatter, and other effects. The hit time distribution is obtained from the HGTD simulation described in Section 3.1.

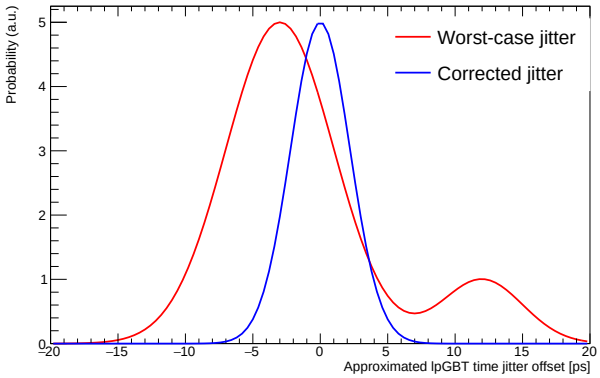


Figure 10.5: Timing jitter distribution assumed for the lpGBT in the corrected (blue) and uncorrected (red) scenarios. These distributions are approximations of the timing jitter expected in the lpGBT.

Random event-by-event fluctuations cannot be calibrated away, although they are included as part of the hit time resolution. Instead, the performance of the timing correction procedure

will depend on how many longer-term variations (heat cycles or other effects) affect the time measurement, which are largely unknown. For the purpose of this study, these unknown longer-term variations are parameterised as a sinusoidally varying 100 ps time offset with variable period.

10.2.2 Timing correction procedure

The hit time offset t_0 is calculated at regular intervals as the arithmetic mean of the t_{hit} distribution. The length of the time interval strongly affects the performance of the timing correction. The t_0 can be calculated to better precision with averaging over a longer time, but shorter times can correct for faster variations. The timing can be computed by forwarding a particular data stream from FELIX to a monitoring unit in which the t_{hit} will be averaged and then applied offline. Alternatively, the computing of the average t_0 inside FELIX will be investigated.

The hit time distribution before and after the timing correction is shown in Figure 10.4. Figure 10.6 shows the timing performance as a function of the integration time and the variation period for channels at three different radii, calculating the t_0 correction from a 15×15 grid of channels, and including all of the sources of time variation discussed above, including the sinusoidally varying 100 ps offset with period plotted along the x-axis. Smaller calibration window sizes can reduce t_0 jitter when shorter-term variations affect the hit time. However, longer calibration windows, which can collect more statistics and therefore more precisely determine t_0, result in a better hit time correction. Variations with period smaller than 1 ms cannot be corrected with this procedure because of insufficient statistics, and variations with period greater than 20 ms can be corrected in all regions of the detector. The timing correction procedure should also work well for longer-term variations on the scale of 1×10^5 s (1 day).

The procedure outlined above and the corresponding results are a preliminary plan for the timing correction scheme using conservative values of clock jitter contributions. Conservative estimates for the expected ALTIROC and FLEX timing jitter were used, and the study will be updated when final numbers are available. When accounting for the expected jitter from components of the readout system and LHC bunch crossing time drift, the clock jitter of approximately 15 ps can be reached, in accord with the specifications outlined in Section 4.2.2. If additional unknown sources of jitter are included, the timing correction procedure can reduce the total jitter to 20 ps for the time variations studied and thus fulfilling HGTD requirements of 35 ps driven by the intrinsic resolution of the sensors. In general, more accurate corrections can be calculated to correct for longer-term variations, and should result in smaller total clock jitter.

The timing correction procedure assumes that time offsets across different channels are not correlated. However, the time offsets in each channel are expected to be somewhat correlated

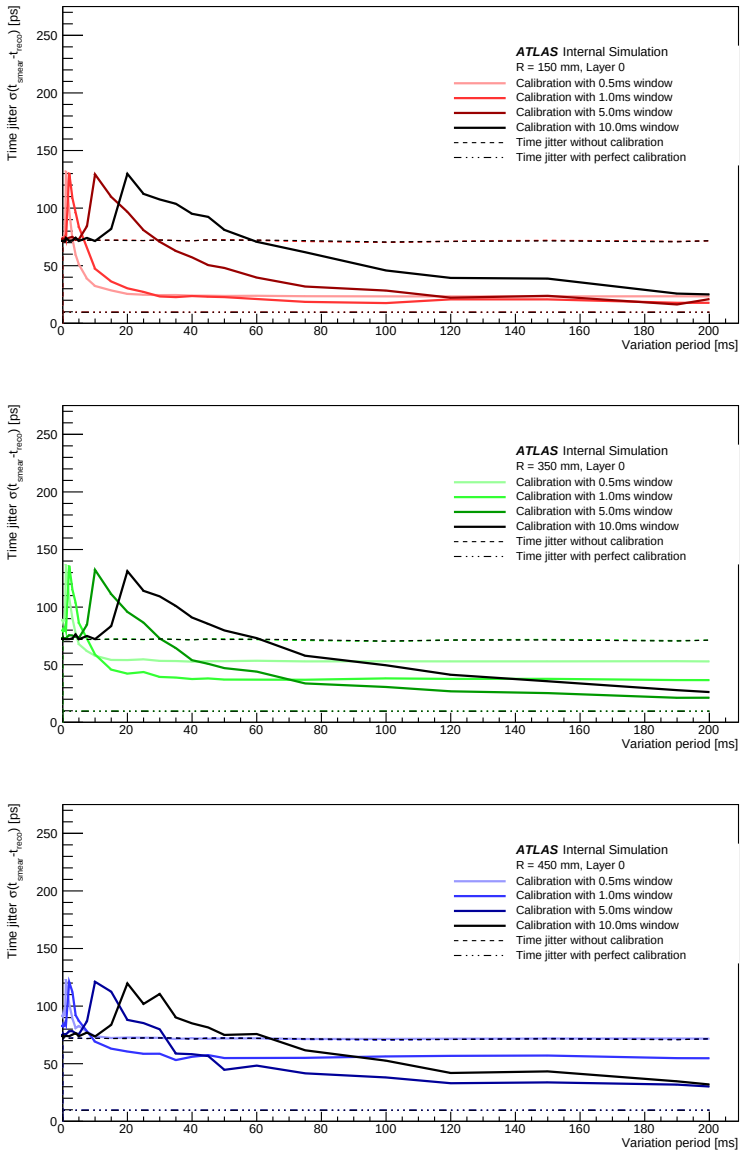


Figure 10.6: Hit time resolution RMS ($t_{\text{smear}} - t_{\text{reco}}$) after the timing correction procedure as a function of the variation period, and for several different choices of calibration window time, shown for $r = 150\,\text{mm}$, $R = 350\,\text{mm}$, and $r = 450\,\text{mm}$. t_{reco} is the hit time taken from simulation and includes inherent hit time resolution effects from the sensor and electronics and the collision time spread. The t_{smear} term adds additional sources of time jitter from the ASIC, FELIX, flex cable, lpGBT, and ATLAS collision time drift, with an additional sinusoidally varying 100 ps offset of variable period. The time jitter without any correction applied is shown as the dashed line, and the time jitter without any long-term timing variation effects is shown as the dotted-dashed line. For a variation period of greater than 20 ms, and with the right choice of calibration window size, the calibration procedure will always improve the t_0 precision.

from both global (i.e. offsets in the LHC collision time and the ATLAS clock) and local effects (i.e. tree structure of the clock distribution creates correlations between modules of the same branch), the timing correction procedure assumes the worst-case scenario of no correlation and applies corrections per-ASIC level. Timing corrections targeting global or more broadly correlated effects can combine hits from more channels, achieving more statistical precision and a better correction across even shorter timescales. Furthermore, the t_0 jitter at the ASIC level can be corrected on a per-channel basis by using the hit times of single pixels, although a factor of 225 would be lost in statistics.

10.3 Luminosity

The measurement of the integrated luminosity delivered by the LHC is critical for almost all physics analyses, as discussed in Section 3.4.4.

Any luminosity detector (luminometer) attempts to measure some observable which is assumed to be proportional to the instantaneous luminosity, or equivalently, to the average number of inelastic interactions per bunch crossing $\langle\mu\rangle$. Conceptually simple examples are the average number of charged-particle tracks reconstructed in the inner tracker [92] or the noise-corrected number of clusters in the pixel detector [93]. In the early years of LHC operation, many luminometers used the so-called *event-counting* method [94], also known as *zero counting*, which exploits Poisson statistics to infer the pileup parameter μ from the fraction of bunch crossings in which no interaction was detected. As the mean μ of the Poisson distribution increases, the fraction of bunch crossings with no detected interaction decreases, and eventually reaches zero. The μ value at which this saturation, or "zero starvation", occurs depends on the geometrical acceptance and the efficiency of the luminometer considered. Already in LHC Run 2, the baseline ATLAS luminometer [95] was forced to exploit its 16-channel granularity to switch from event counting to hit counting. This latter method [94] applies a Poisson formalism very similar to that of event counting, to extract μ from the average number of detector hits recorded per bunch crossing; the finer the granularity of the luminometer, and the smaller the acceptance of its individual channels, the higher the pileup value at which the method eventually saturates. In the limit of a very large number of channels, as is the case in a pixelated detector such as the HGTD, the per-channel occupancy becomes small enough for the Poisson non-linearity to become almost negligible. The average number of hits in randomly selected colliding-bunch crossings then depends linearly on the luminosity (except perhaps at the highest μ values expected at the HL-LHC, where the hit-counting Poisson formalism may need to be invoked again).

The primary calibration technique to determine the absolute luminosity scale of a bunch-by-bunch luminometer employs dedicated van der Meer (vdM) scans [92] to infer the delivered luminosity at one point in time from the measured parameters (primarily the intensity and the transverse area) of the colliding bunches. The conversion factor from luminometer

counting rate to measured luminosity is then determined by comparing the luminosity computed from the above-mentioned accelerator parameters to the visible, uncalibrated interaction rate reported by the luminometer at the peak of the beam-separation scans. The beam conditions during vdM scans are different from those in normal physics operation, with lower bunch intensities and only a few tens of widely spaced bunches circulating. These conditions, which are optimized to reduce various systematic uncertainties in the calibration procedure [96], typically result in a pileup parameter μ of about 0.5 at the peak of the scans, and as low as $\mu \sim 2 \times 10^{-5}$ in the tails of the scans, where the beams are barely overlapping. Since the same luminosity-calibration procedure is foreseen at the HL-LHC, the luminometer response will have to remain linear over seven orders of magnitude in μ, from vdM conditions ($\mu \sim 2 \times 10^{-5}$ to $\mu \sim 0.5$) up to high-luminosity physics data taking at $\langle\mu\rangle$ of around 200.

The online and offline environments impose different and sometimes conflicting constraints on the luminometers and the associated luminosity-determination methods, with processing speed being of the essence during data taking (possibly at the expense of absolute accuracy), and offline luminosity requiring the best possible precision on much longer time scales. For instance, track counting [92], which proved essential to control the dominant luminosity uncertainties in both LHC Runs 1 and 2, has only be used offline so far since it requires a dedicated, randomly-triggered event stream that must be subjected to extensive offline analysis before usable luminosity values can be provided.

Bunch-by-bunch luminosity estimates are required not only for offline physics analysis, but also in the online environment, for instance to apply bunch- and μ-dependent corrections to calorimeter data in the high-level trigger algorithms; to optimize the trigger menus on the fly; and to monitor, analyze and improve the accelerator performance over the long term. An additional requirement is the availability of a bunch-integrated, fast and reasonably accurate luminosity measurement, provided at $\sim$ 1 Hz as input to the collision-optimization and luminosity-leveling accelerator protocols.

As discussed further in Section 10.3.5, the precision of the offline determination of the integrated luminosity has so far been limited not by statistics, but by systematic uncertainties. An essential lesson from LHC Runs 1 and 2 is that the dominant systematic uncertainties can only be determined, or at least constrained, by confronting the response of a redundant set of luminometers, each based on a different technology, with complementary capabilities and independent instrumental biases.

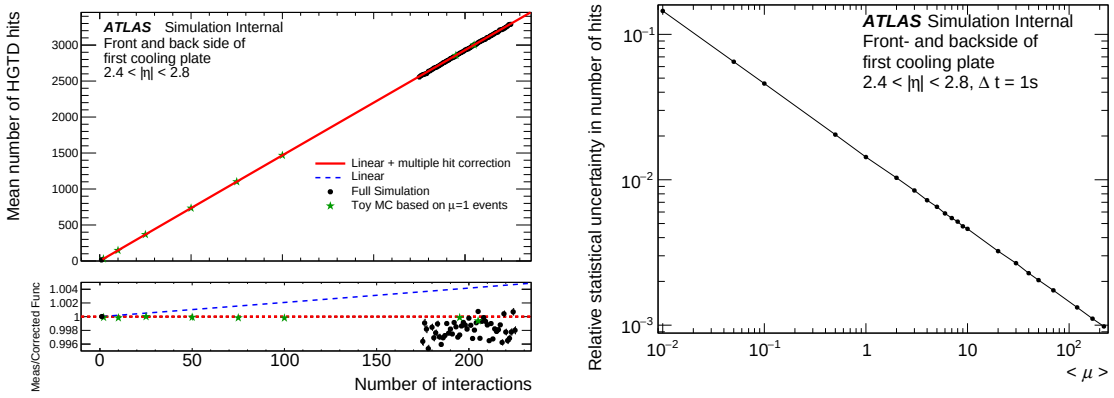
10.3.1 HGTD as a luminometer

As a fast high-granularity detector in the forward region, the HGTD provides unique capabilities for measuring the luminosity at the HL-LHC. The idea for using HGTD as a luminometer is straightforward: the occupancy will be linearly correlated with the interaction rate (i.e. the

luminosity). The high granularity gives a low occupancy, and therefore excellent linearity between the average number of hits and the average number of simultaneous pp interactions over the full range of luminosity expected at the HL-LHC, as discussed in Section 10.3.2. With detector signal durations in the few-ns range, the charged-particle multiplicities within the acceptance can be determined accurately for each individual bunch crossing separately. With the occupancy information sent at 40 MHz, i.e. for every bunch crossing independent of the ATLAS trigger (further discussed in Section 10.3.6), the HGTD will provide both online and offline per-BCID luminosity measurements. The measurement is made in a reduced $|\eta|$ range, and in this proposal the plan is to read out the ASICs for sensors at 430 mm $< r <$ 640 mm (equivalent to $2.4 < |\eta| < 2.8$) for the luminosity determination. The HGTD is designed to have capabilities to constrain many systematic uncertainties by itself, with the goal of reducing the total uncertainty on the integrated luminosity in HL-LHC compared to Run 2 despite the much harsher experimental conditions, as is discussed in Section 10.3.4 and Section 10.3.5.

10.3.2 Linearity of the luminosity determination

For the $|\eta|$ range $2.4 < |\eta| < 2.8$, the average number of hits per inelastic pp collision for one double-sided layer on one side of the interaction point is 14.7, and approximately 16.0% of these collisions result in 0 hits. Figure 10.7(a) shows the average number of hits per bunch crossing registered in the first double-sided HGTD layer (both sides of the innermost cooling plate) as a function of the number of simultaneous inelastic pp interactions. The black points at number of interactions of 1 and between 175–225 are determined from fully simulated minimum-bias events with $\mu = 1$ and $\langle\mu\rangle$ in the range 190-210, respectively. The green stars represent results from a toy MC where several $\mu = 1$ minimum-bias events have been overlaid to produce samples with intermediate numbers of interactions, while making sure not to double-count multiple hits in the same channel. A linear, ideal, relationship between the mean number of hits and the number of interactions (blue dashed line), derived from the fully simulated sample at $\mu = 1$, is extrapolated to the $\mu \sim 200$ region where its prediction can be compared to the hit multiplicities extracted from fully simulated high-pileup samples. The small discrepancy, of approximately 0.6%, between the blue dashed line and the simulated points in the bottom left frame around $\mu \approx 200$ is mostly attributed to multiple particles hitting the same pad. The red line (labelled "Linear + multiple hit correction") is the result of correcting the linear function with the contribution from multiple particles hitting the same pad. A residual 0.2% discrepancy between the corrected function and the fully simulated MC events around $\mu \approx 200$ can be observed in the bottom frame, this can be attributed to all the differences between the toy MC model and the fully simulated sample. Examples of such differences are, for example, the simulation of out-of-time pileup and multiple below-threshold energy deposits from different proton-proton collisions superimposing and generating above-threshold hits.



(a) Linearity of $\langle n_{\text{hits}} \rangle$ as a function of μ.

(b) Statistical uncertainty versus $\langle \mu \rangle$

Figure 10.7: Left: mean number of HGTD hits per bunch crossing as a function of the number of interactions. The black points are the results from fully simulated samples. The green stars represent results from a toy MC where several $\mu = 1$ minimum-bias events have been overlaid to produce samples with intermediate numbers of interactions. The blue dashed line is the ideal linear relationship between the mean number of hits and the number of interactions, derived from the $\mu = 1$ sample. The red line is the result of adding a correction from multiple particles hitting the same pad to the linear parameterisation. In the bottom panel, both lines can be compared to the fully simulated samples at $\mu \sim 200$ (see the text for a full description of the plot). Right: pileup dependence of the statistical uncertainty per BCID, for an integration time of 1 s.

10.3.3 Statistical precision of the luminosity determination

To confirm that statistical uncertainties are small for the online luminosity measurements, the size of the uncertainty has been studied as a function of the duration of the averaging period and $\langle \mu \rangle$. The average number of hits per bunch crossing is simulated using a toy Monte-Carlo method with inputs extracted from fully simulated samples. For each value of $\langle \mu \rangle$, a random number of pp interactions is drawn from a Poisson distribution with a mean equal to $\langle \mu \rangle$. For each pp interaction, a number of HGTD hits is then generated randomly based on the distribution of hits per pp interaction extracted from full-simulation samples. By repeating this process 11 000 times (for the number of turns the LHC beams will make in one second) and averaging the number of hits, the statistical precision achieved in each individual BCID during 1 s of LHC running is emulated. Figure 10.7(b) shows the relative uncertainty expected from statistical fluctuations as a function of $\langle \mu \rangle$ using this method. The coverage of $2.4 < |\eta| < 2.8$ presented here gives a statistical uncertainty of 1.4% at $\langle \mu \rangle = 1$ and 14.3% at $\langle \mu \rangle = 0.01$. For measurements in the low-μ regime (e.g. during van der Meer scans) better precision can be achieved through a longer averaging time.

10.3.4 Noise and afterglow subtraction

The HGTD is affected by three distinct background contributions to the luminosity signal: single-beam backgrounds, instrumental noise, and afterglow, in order of increasing importance.

Single-beam background arises from activity correlated with the passage of a single beam through the detector. This activity is caused by shower debris from beam-halo particles, that impinge on the luminosity detectors in time with the circulating bunch. Although its impact remains to be simulated, single-beam background is expected to be close to negligible (on the scale of the luminosity signal), based not only on experience with Run-1 and Run-2 luminometers, but also on HGTD-specific features: on the incoming-beam side, not only should the shielding provided by the end-cap calorimeter absorb all of the high-radius backgrounds (except for a few muons), but the surviving background particles will be out-of-time by several nanoseconds with respect to the collision products originating from the IP. Residual HGTD backgrounds on the outgoing-beam side, if any, can be roughly estimated from a few non-colliding bunches injected in each ring for this specific purpose, as was frequently done during LHC Runs 1 and 2.

Instrumental noise can arise from thermal noise in detector electronics, or from high-rate contributions from "noisy pixels" (such as caused by radiation-induced "single-event upsets"). Thermal-noise (and, up to a point, noisy-pixel) contributions can be subtracted by the same method as that used for afterglow, which is discussed in the next paragraph. Alternatively, noisy pixels can be masked, if only to prevent excessive dataflow rates (in which case their unavailability will have to be accounted for when normalizing the measured hit counts).

As detailed in [94], all Run-2 bunch-by-bunch luminometers (with the exception of track counting) observe some activity in the BCIDs immediately following a collision, which in later BCIDs decays to a baseline value with several different time constants. This afterglow is attributed to slow particles (such as neutrons) and to delayed decays (e.g. from stopped muons), that originate from the hadronic cascades initiated by pp collision products. For a given bunch pattern, the afterglow level is observed to be proportional to the luminosity in the preceding colliding bunches. Its magnitude relative to the luminosity signal, and its time structure, both depend on the sensitivity of the luminometer considered to the particle composition and the energy spectrum of the afterglow (and therefore on the technology used by that luminometer), as well as on the location and the physical environment (geometry, chemical composition of neighbouring equipment) in which this luminometer operates. The magnitude of the afterglow contamination observed in Runs 1 and 2 varies widely, from 10^{-4} for LUCID in vdM scans, to 0.2-0.4% for BCM in high-μ bunch trains; it can be as high as 10% in pixel detectors during routine physics running, therefore requiring a delicate correction that contributes sizeably to the total luminosity uncertainty.

The time resolution of the HGTD is a unique capability that is essential to mitigate the large

impact of instrumental noise and afterglow intrinsic to the pixel-cluster counting technique. As described in Section 6.1, and illustrated in Figure 6.2, the ASIC will send occupancy information in two different time windows:

- a *central time window*, 3.125 ns wide, centred on the nominal bunch crossing time;
- a *sideband window*, nominally covering 3.125 ns before the central time window and 3.125 ns after the central time window.

This double-sideband window will be programmable. Here it has been chosen symmetric, such that its occupancy provides, after appropriate scaling, an estimate of the noise and afterglow contributions as interpolated under the luminosity signal in the central time window, separately for each BCID. This ability to perform an in-situ measurement of the noise and afterglow level for each bunch crossing, using data from empty RF buckets just before and after the filled bucket within the same nominally filled 25-ns bunch slot, is a unique capability of the HGTD compared to other luminometers.

10.3.5 Systematic uncertainties affecting the luminosity determination

A detailed discussion of the systematic uncertainties affecting the 2012 luminosity determination at $\sqrt{s} = 8$ TeV is presented in [92]; the sources and the magnitude of the luminosity uncertainties in LHC Run 2 at $\sqrt{s} = 13$ TeV are comparable [45]. Of the dominant uncertainties, two are luminometer-specific (rather than related to, for instance, beam conditions or accelerator instrumentation): the time stability of the luminometer response, and the calibration transfer.

The time stability of relative-luminosity measurements is potentially affected by different sources, depending on the time scale considered.

- *Long-term stability* refers to potential drifts of the luminometer response on the time scale of days to months, compared to its response at the time of the vdM-calibration session. Such drifts have been seen to arise, for instance, from gain fluctuations in, or flux-induced ageing of, LUCID photomultipliers (PMTs); darkening of TILE scintillators; cumulative radiation damage to inner-tracker silicon-strip or pixel modules; or unaccounted-for dead or inefficient channels. In LHC Runs 1 and 2, this class of effects contributed from 0.5% to 1.3% to the systematic luminosity uncertainty, a large number compared to the luminosity-precision goal of 1% at the HL-LHC.
- *In-run stability* refers to variations in luminometer response on the time scale of one ATLAS run (a few hours). The reference ATLAS luminometers (BCM in most of Run 1, LUCID in Run 2) proved mostly immune to such drifts. In contrast, pixel-cluster-counting-based luminosity measurements were significantly more sensitive, typically because of unaccounted-for changes in effective coverage (noisy, misbehaving or automatically disabled modules). Because the luminosity, and therefore the pileup

parameter μ, typically decays during an LHC fill, such drifts are difficult to disentangle from a genuine μ-dependence of the detector response. It is therefore essential, for pixel-counting methods, to keep track of variations in both the number and the radial location of misbehaving channels on the time scale of a few minutes: for instance, a few noisy pixels that suddenly start firing at a high rate may bias the luminosity measurement and prove hard to correct for after the fact.

The long-term stability and the in-run stability of the HGTD will be monitored by offline data quality analysis, similarly to what is done for the current Pixel and SCT detectors. This includes both prompt analysis of the recorded data during the calibration loop and thorough analysis of data taken over timespans of months or a full year. Detector elements that are found to have non-constant hit efficiency can then be excluded when determining the final luminosity estimates.

The *calibration-transfer uncertainty*, which in LHC Run 2 typically amounted to a 1.0–1.5% uncertainty on the absolute luminosity scale, refers to how precisely one controls potential shifts in detector response, that occur between the beam conditions of vdM scans ($\langle\mu\rangle \sim 0.5$, a few tens of low-intensity isolated bunches, no bunch trains) and those of physics data-taking ($\langle\mu\rangle \sim 200$, hundreds to thousands of high-intensity bunches grouped in trains with diverse patterns). Such shifts can arise, for instance, from rate-dependent effects in solid-state sensors or LUCID photomultipliers; from bunch–pattern-dependent "out-of-time electronic pileup" (in which the electrical signal from a given 25 ns bunch slot leaks into the following bunch slot); in the case of track counting, from a residual pileup dependence of the tracking efficiency; or, in randomly triggered readouts of bunch-integrated inner-tracker luminosity data, from subtle deadtime effects through which a higher-luminosity bunch can shadow a small fraction of the triggers in the immediately following bunch slot. All of these effects have been observed at some level in Run 2 at $\langle\mu\rangle \sim 50$, and required μ- and time-dependent corrections to the luminosity scale that could exceed 10% during high-luminosity operation.

The HGTD has several characteristics that will aid in constraining, and hopefully reducing, such systematic uncertainties. To better monitor the time stability, the region instrumented with the luminosity readout will be segmented into 16 sub-regions, with 4 divisions in η and 4 divisions in ϕ, as shown in Figure 10.8. Each region has sufficient statistical sensitivity to determine the luminosity independently of the other regions. Regions at different η will accrue radiation damage at a different rate, therefore comparing their response can help determine the degradation due to radiation. The partitioning of the regions can be controlled in software running in the Luminosity Software, described further in Section 10.3.8, so that a different layout than the one described here can be used if found to be more optimal.

While such internal consistency checks will undoubtedly prove valuable, they are unlikely to be sufficient, if only because any bias or drift that is correlated across all 16 regions remains undetectable by the HGTD alone. Experience at LHC has repeatedly shown that

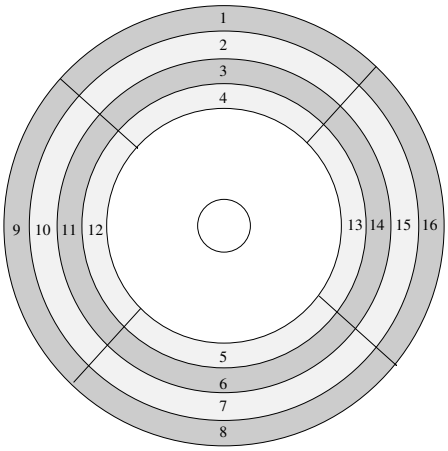


Figure 10.8: Sketch of the partitioning of the sensors into 16 regions for the luminosity determination. Each of the regions can be used to determine the luminosity independently of the others. Regions at different radii will be subject to different levels of radiation over time.

independent checks based on several luminometers using different technologies are essential for controlling the systematic uncertainties to the level required by the physics program.

Built into the HGTD design are several features that are expected to reduce the magnitude of the calibration-transfer correction (if any), as well as help constrain the associated uncertainties:

- the pixel-cluster counting technique is intrinsically linear, and only very small μ-dependent corrections are expected to be necessary at the highest bunch luminosities expected at the HL-LHC, as was illustrated in Figure 10.7(a);
- for a given bunch pattern, the most likely reasons for the hit count to deviate from strict proportionality to the true luminosity are afterglow and instrumental noise. The exquisite time resolution of the HGTD, combined with the methodology outlined in Section 10.3.4, provides a unique strategy to control these effects to the level needed;
- the most likely reason for a bunch-pattern dependence of the HGTD hit count is again the afterglow, the magnitude of which is sensitive to the length of, and the separation between, bunch trains. The above-mentioned afterglow subtraction at the bunch-by-bunch level should eliminate this potential bias;
- electronic out-of-time pileup from one BCID to the next is presumably eliminated by the extremely short pulse duration of HGTD pixels;
- eliminating deadtime effects associated with large μ variations from one BCID to the next, is one of the motivations for the trigger-less, 40 MHz readout of the luminosity information discussed in Section 10.3.6.

10.3.6 Occupancy readout at 40 MHz

Experience with luminosity determination at the LHC shows that the capability to read out a luminometer at 40 MHz, i.e. on every single bunch crossing, is critical to its function as an independent device that must provide bunch-by-bunch (bbb) luminosity measurements, with the best possible precision both online and offline. In LHC Runs 1 and 2, this requirement was satisfied only by LUCID and BCM; the fact that it was out of reach for track and pixel-cluster counting methods proved a significant limitation to the final precision of the integrated luminosity in both ATLAS and CMS in Run 2.

In view of the more exacting luminosity-precision requirements of the HL-LHC physics program, the 40 MHz readout of the occupancy is key to a full exploitation of the HGTD potential as a stand-alone, high-precision luminometer for both online and offline use. This becomes apparent when one considers

- the unrivalled statistical power of reading out every single bunch crossing, thereby collecting, in a fully unbiased manner, all the potentially available luminosity data from every single bunch slot,
- the TDAQ implications of a readout triggered by sampling randomly selected colliding-bunch pairs,
- some of the requirements associated with the van der Meer calibration,
- use cases of bunch-by-bunch luminosity measurements in both the online and the offline environment, and
- some features specific to the HGTD-based luminosity determination.

If the luminosity measurement were to be carried out using a detector which is not read out on every bunch crossing, the following considerations would have to be addressed.

- The luminosity must be determined from an unbiased sampling of collisions, therefore it is unlikely that data passing physics triggers can be used. Such triggers normally require a lot of activity in the detector, e.g. the presence of high momentum leptons or jets. They are typically sensitive to pileup effects, and therefore not representative of the luminosity; they also are severely statistics-limited.
- The traditional method for overcoming the trigger bias is to use a dedicated random trigger, sampling each bunch crossing evenly. The bandwidth for such a trigger comes at the expense of that available for physics, thus effectively representing a loss in data-taking efficiency.
- A random trigger does not result in a completely unbiased dataset for the luminosity determination. There is a shadowing effect from the standard trigger deadtime, in which more luminous bunches shadow collisions in subsequent, less luminous bunch

slots. The associated corrections are unlikely to be negligible for a bunch-integrated measurement, but could be corrected for in a bunch-by-bunch measurement with the knowledge of the number of times a bunch is sampled.

- Even if the luminosity extraction could be performed online using the luminosity back-end electronics to analyze Level-0 triggered data, it would reduce the available statistics by several orders of magnitude: this would make the HGTD inadequate as an online luminometer, as further argued below.

- If the luminosity-extraction analysis can only be carried out offline, the event data has to be saved to disk, further degrading the statistics usable for luminosity-related applications.

The vdM calibration technique requires evaluating, as a function of the transverse beam separation, the four-dimensional overlap integral (over x, y, z and time) of the proton-density distributions in each colliding-bunch pair. Since the proton population and the transverse-density distributions vary significantly from one bunch to the next, fitting a vdM scan curve obtained by summing the interaction rate over all colliding-bunch pairs (rather than fitting a separate scan curve for each pair) would result in unpredictable and non-reproducible biases to the absolute luminosity scale. This fundamental requirement, on its own, implies that the HGTD must provide statistically precise bunch-by-bunch luminosity measurements over the full μ range covered during a vdM scan (2×10^{-5} to 0.5).

The above span in interaction rate, combined with the LHC bunch-revolution frequency of 11 kHz and with a typical integration time of 60–100 s during individual vdM scan steps, implies that a readout based on randomly triggered colliding-bunch crossings would have to be restricted to a fraction of the available colliding-bunch pairs in order to accumulate data with per-bunch statistics sufficient for a vdM analysis. Triggering the HGTD readout during the vdM scan using some kind of independent track- or hit-multiplicity trigger is not optimal, since it requires determining the absolute efficiency of said trigger, at some cost in systematic uncertainty. While both of these techniques have been used successfully for track counting in 2012 vdM scans, they unavoidably degrade, at some level, the uncertainty affecting the bunch-averaged visible cross-section. This chain of arguments explains why, during LHC Run 2, no direct vdM calibration of track- nor pixel-cluster-counting algorithms was ever attempted by ATLAS. These inner-tracker-based luminosity algorithms were instead cross-calibrated to LUCID in data-taking at $\mu \sim 0.5$ during the vdM session, thereby making their absolute calibration fully correlated with that of LUCID.

During both routine physics operation and during special runs, the need (both online and offline) for a luminosity readout that provides high statistics in each bunch slot over the full μ range is fundamentally related to the intrinsic variation of the emittance, bunch intensity and instantaneous luminosity across the colliding-bunch pairs. During Run 2, these bunch-by-bunch variations in the luminosity sometimes exceeded 20–30%.

In the present ATLAS online-luminosity architecture, bunch-by-bunch luminosity measurements provide the basic input to the computation of the bunch-integrated luminosity value, that is used, for instance, to select the most appropriate ATLAS trigger settings; inform the online monitoring tools of various ATLAS subdetectors; optimize collisions; control the luminosity-leveling protocols; monitor accelerator performance, etc. Depending on the application considered, the required refresh times vary from one to a few tens of seconds.

In addition to a precise bunch-integrated measurement, bunch-by-bunch measurements that are statistically stable ($\ll 0.5\%$) and reasonably accurate on an absolute scale, are required online for several purposes, such as:

- providing μ-dependent corrections to calorimeter-based triggers, with a refresh rate of a few minutes;

- supplying the accelerator with diagnostics such as bunch-by-bunch specific-luminosity values, which offer a better estimate of the beam-averaged emittance than state-of-the-art accelerator instrumentation. Such diagnostics have proven essential to the steady improvement of LHC performance. They are needed not only during physics running for detailed analysis of accelerator operation, but also in real time with refresh rates of a few seconds for periodic emittance scans, as well as for some accelerator-development sessions, during which the beam parameters are tailored on a bunch-by-bunch basis and the required refresh rates are at the few-seconds level.

Use cases for bunch-by-bunch measurements in the offline environment include, for instance:

- computing the bunch-integrated luminosity eventually used in physics analyses from the sum of per-bunch luminosity values, after recalibration and application of bunch-dependent corrections, such as residual μ-dependence or afterglow subtraction;

- refined μ- (and therefore bunch-) dependent corrections to the cell-by-cell energy measurements in the liquid argon calorimeter;

- bunch-by-bunch comparisons of the relative consistency of the luminosity values across multiple luminometers. Such studies have revealed significant μ- and bunch-position dependent biases in all the bunch-by-bunch luminometers available in Run 2, and again demonstrated that comparing independent luminometers is a key ingredient to precision luminosity measurements.

Finally, the potential use of the occupancy information in the Level-0 trigger outlined in Section 10.3.11 is entirely dependent on the availability of dedicated occupancy data at 40 MHz.

10.3.7 Luminosity back-end electronics

For every bunch crossing of the LHC, each ASIC in the region $2.4 < |\eta| < 2.8$ will send occupancy counts in the central time window and in the sideband time window. These counts are encoded into 7 and 5 bits, respectively. In addition 4 bits are used for encoding, using the 6b8b encoding scheme, resulting in 16 bits sent per ASIC for every bunch crossing. Thus there is a steady data rate of 40 MHz times 16 bits, or 640 Mbit s^{-1}, from each ASIC. The luminosity data is sent via lpGBTs dedicated to the luminosity readout to the back-end electronics, requiring 152 lpGBTs for each of the four disks of the HGTD, i.e. 608 links for the whole detector. The data sent by the lpGBTs are collected by the luminosity back-end electronics, consisting of dedicated FELIX units. These units are separate from the FELIX units handling the timing data, as shown in Figure 10.1. Each FELIX I/O card can take up to 24 input fibres at 10 Gbit s^{-1} (with two such I/O cards present in one FELIX unit). One FELIX I/O card will handle all the occupancy data from one quadrant of one single-sided layer (4 FELIX I/O cards per disk layer, 8 cards per double-layered disk, 32 FELIX I/O cards to handle all the occupancy data from the 4 HGTD disks). Each FELIX I/O card is connected to 133 modules, or equivalently 266 ASICs.

The luminosity back-end electronics aggregates the central time window data and the sideband data separately, for each ASIC separately and for each of the 3564 BCIDs of the LHC except one which is used to synchronise the data stream. The BCID used for synchronisation will be in the abort gap, where no collision data is expected. In total the FELIX card will keep track of $266 \times 2 \times 3563 \approx 1.90$ million sums. The FELIX will store these sums in registers in the FPGAs, and update them continuously with the new data for every bunch crossing. If the data transfer speed between the FELIX cards and the host server allows it, an option would be to store and update the sums using software running on the host server instead of in firmware running on the FPGA which would give greater flexibility and ease maintenance. This option will be investigated further in the upcoming years.

The maximum number which can be obtained for the hit count sum in the central time window for a single BCID over a period of around one second (which is the maximum integration time considered before data is sent downstream), if every collision would saturate the maximum hit count of 127, is $40 \cdot 10^6/3564 \times 127 = 1,425,448$. This is a number which can safely be stored in 4 bytes. The amount of memory used to store all the sums is therefore at most $1.90 \times 4 = 7.6$ MB, something which can safely be accommodated already in the existing versions of the FELIX FPGAs.

These sums are the raw data needed for determining the luminosity, which is only needed with a frequency of approximately once per second. Assuming that the luminosity data gets pushed out of the FELIX at a rate of 2 times per second, and using 4 bytes to store each of the integers, the total data rate out of the luminosity back-end electronics is only $7.6 \times 2 = 15.2$ MB s^{-1}. Thus, the luminosity data represents a negligible strain on the network downstream of the back-end electronics, and the data flow is independent of the

trigger. The conversion from the occupancy sums to a calibrated luminosity will happen in dedicated software algorithms. These software algorithms can run on any downstream computer, most likely in the Luminosity Software.

10.3.8 Processing of the occupancy data in the Luminosity Software

The software running in the Luminosity Software will receive the occupancy sums for each ASIC as input, and by applying appropriate algorithms and calibrations convert these into an estimate for the luminosity. This corresponds approximately to some of the tasks performed by software called the Online Luminosity Calculator in Run 1 and Run 2.

The Luminosity Software will be responsible for aggregating data into time windows corresponding to ATLAS Luminosity Blocks (LB), typically of the order of one minute. A LB is the smallest unit of data for which the offline luminosity is determined. The raw counts for each of the ASICs for one LB will be stored in files with a dedicated format or a database for offline storage. This allows for exclusion of individual ASICs in the determination of the offline luminosity due to data quality issues discovered after dedicated analysis. In the current layout, there are 8,512 ASICs used for luminosity determination. If 4 bytes is used for each of the 2 sums for one ASIC, and separate sums are stored for each BCID, this corresponds to approximately 243 MB of data stored offline for each LB (or $4\,\mathrm{MB\,s^{-1}}$ assuming a LB length of 60 seconds).

The Luminosity Software will also have the flexibility to combine several ASICs into regions which are large enough to be calibrated in the vdM scans, and which can be used to determine the luminosity independently of each other. One possible configuration of regions, dividing the disc into 16 partitions, was shown in Figure 10.8. Each such region would then combine the hit count information from many ASICs.

10.3.9 Per-event luminosity information stored in the ATLAS raw data

In the processing of the luminosity data by the back-end electronics, the per-event information is lost when the data is aggregated. To allow for per-event occupancy data to be stored in the raw data for events passing all the stages of the trigger, the luminosity back-end electronics have to implement a buffer to store the data for each event separately, until a L0 trigger accept is received and the corresponding occupancy information can be sent. Whether this capability will be implemented, and per-event occupancy information will be recorded in the ATLAS raw data, has not yet been decided. Eventual difficulties with synchronisation of the data with the rest of the event will also have to be investigated. The per-event occupancy in the central time window provides no unique information over what can be calculated from the HGTD precision timing data (modulo the fact that the time windows used for the timing and occupancy data are slightly different). It would merely serve as validation of

the luminosity and precision timing data, but it could be very beneficial for this purpose. The information about the occupancy in the sideband time window does provide unique information compared to the HGTD timing data, and could potentially have use cases in e.g. searches for new, slow-moving particles.

Assuming the same pipeline depth for the occupancy data as for the timing data buffered in the ASIC, the capacity to buffer per-event level data for 1,400 consecutive bunch crossings is needed. For each bunch crossing, each of the 266 ASICs sending data to one FELIX sends two bytes of data (16 bits). Adding an additional two bytes for header information for every ASIC, a total of $266 \times 1400 \times 4 \approx 1.49$ MB of memory is required in the FELIX FPGA. As was already discussed in Section 10.3.7, this can easily be accommodated by existing versions of the FPGAs.

Provided that the capability to buffer per-event luminosity data is implemented in the luminosity back-end electronics, the payload to be stored in the ATLAS raw data would be the occupancy data (16 bits) for each of the 8,512 ASICs used for the luminosity determination (approximately 17 kB per event).

10.3.10 Operation in non-Stable Beams conditions

As with other silicon sensor-based detectors close to the LHC beamline, the HGTD will only ramp up the full High Voltage on the sensors once Stable Beams have been declared, in order to avoid destroying the detector in case of catastrophic beam losses. At the same time, there is a need from the accelerator operations perspective to have an estimate of the luminosity at the ATLAS interaction point in conditions where Stable Beams have not been declared. This situation occurs at the start of every physics run, and can also be necessary during periods of machine commissioning.

Providing an online luminosity estimate in non-Stable Beams operation reinforces the need for ATLAS to have several different luminometers at the HL-LHC, employing different detector technologies. Less precise, but more radiation tolerant, detectors could then be the primary sources of luminosity measurements by ATLAS when Stable Beams have not been declared. Whether a safe operation mode can be found for the HGTD during non-Stable Beams conditions is still to be investigated. A possibility of operating just the outermost regions (regions 1, 8, 9 and 16 in Figure 10.8) at a reduced HV setting could be safe. The reduced HV setting would result in a lower hit efficiency, and thus a different relationship between the instantaneous luminosity and the average number of hits expected in the HGTD, compared to operating at nominal HV conditions. A separate calibration of the luminosity determination for such a operating mode can be accommodated in the Luminosity Software. Whether a safe operating mode of the detector in non-Stable Beams conditions can be found will require extensive tests of the sensors and possibly also operating experience with the full detector.

10.3.11 Minimum-bias trigger at Level-0

The data made available at 40 MHz for the luminosity measurements can also be used to provide a L0 trigger signal in order to record minimum-bias events under low-μ data-taking conditions. Such data-taking conditions are expected during e.g. heavy-ion runs, van der Meer scans or for runs dedicated to soft-QCD measurements. The HGTD will be installed where the current MBTS detector is located. The MBTS detector has been used extensively for these purposes during Run-1 and Run-2, e.g. during the heavy-ion runs where it played a crucial role. However, the MBTS will not be present at the HL-LHC. With improvements of several orders of magnitude in both granularity and time resolution, the HGTD can provide all the functionality of the MBTS. The number of hits in the time window centred around the nominal collision time provides good separation between empty bunch crossings and those with pp collisions. A simple threshold for the minimum number of hits using the occupancy information is straightforward to implement in the luminosity back-end electronics. Such a binary trigger decision can then be communicated directly to the central trigger. The latency for reaching the Level-0 global trigger processors in time for a decision is not expected to be a problem.

10.4 Detector Control System

In order to ensure the coherent and safe operation of the HGTD, a Detector Control System (DCS) will be put in place. The main tasks of the DCS are to bring the detector in any desired operational state, while ensuring that any action can not put the detector in a situation where safety reactions are necessary, to monitor its operational parameters and to signal any abnormal behaviour, thus allowing manual or automatic corrective actions. The DCS provides a homogeneous interface between the operator and the detector and its infrastructure, enabling tasks such as detector calibration, commissioning and operation.

The DCS elements are distributed over various detector components: front-end electronics, services, back-end electronics and DCS servers. A Finite State Machine (FSM) structure will be implemented and integrated in the ATLAS FSM tree during data taking, and will support operation in stand-alone mode during commissioning and maintenance. Real-time monitoring of critical parameters will be implemented, and alerts will be raised as soon as critical conditions are reached or connection to one or more hardware devices is lost. All relevant DCS parameters will be archived for debugging, performance tuning and offline studies.

The DCS will control and monitor the following parameters: the power, both high- and low voltages, supplied to the detector; temperatures of the detector modules, peripheral electronics and cooling; humidity and overpressure inside the vessel.

Standard ATLAS DCS front-end (FE) components and communication interfaces are used whenever possible. however, due to constraints on physical space, in some cases DCS data share the communication infrastructure of the on-detector electronics based on the lpGBT chipset, interfaced with the off-detector back-end via Versatile Links (optical) and the FELIX system, with other types of data (read-out and trigger), as detailed in the next sections.

10.4.1 High Voltage

The HV supply system will include the hardware and software components for being connected to the DCS for control and monitoring of both voltage and current. For the HV supplies the behaviour at the selected current limit is preferably programmable to not only be in trip mode but also for current-limiting operation. The supplies will be based on commercial multi-channel rack mounted units located in the service cavern.

As detailed in Section 8.1, during the lifetime of the sensors the bias voltage must be adjusted due to gain degradation with received radiation dose. In combination with the non-radial geometry this results in a limited possibility to connect several modules to the same bias supply. The ultimate choice is to use individually adjustable voltages to allow for optimal operation of the sensor modules in view of radiation damage. This is kept as an option while at the start of operation, due to cost, on the average one high voltage channel will supply two sensor modules.

The leakage current will be closely monitored to have an estimate of the radiation damage in the sensors at different radii. The hit efficiency together with the leakage current measurements will be used to adjust the HV setting in different locations of the detector in order to ensure the full depletion of the sensors. The adjustment of the bias voltage will be performed during interfill periods (ideally during technical stops if possible) and after an IV scan to ensure safe operation of the sensors to higher voltages.

10.4.2 Low voltage

The low voltage needed by the front-end and peripheral electronics will be provided by a three-stage system, as detailed in Section 8.2.

The bulk 300 V supplies as well as the 300 V to 10 V DC-DC converters are assumed to be commercial products. Both of them must include provisions for communication with DCS allowing for control and monitoring of voltage and current. The voltages from the DC-DC converters on the peripheral boards and the voltages received at the front-end ASICs are monitored via multiplexers and ADC channels on the lpGBT ASICs of the peripheral boards, as described in Section 10.4.4. From the lpGBTs the information is sent via optical fibres to FELIX boards of the DAQ system for transmission to the DCS system. The optical links to the lpGBTs from the DAQ FELIX boards will exchange data bits, embedded in the data

streams, for switching on and monitoring the status of the DC-DC converters powering the front-end ASICs. However, several DC-DC converters per peripheral boards must be controlled by DCS over wires, as they will power the lpGBTs, which will control the rest of the DC-DC converters on the board.

At the moment of writing this document, for the bulk 300 V supplies as well as the 300 to 10 V DC-DC converters (stages 1 and 2) a commercial product is considered as it fulfills not only the above requirements, but also in terms of radiation hardness and tolerance to magnetic fields.

ALTIROC ASIC monitoring

The voltage provided to power the digital and analog parts of the ALTIROC ASICs will be monitored by DCS using the ADC of the lpGBT circuit in the peripheral boards. A detailed description of the proposed monitoring of ALTIROC is given in Section 6.8. Three signals ($Vdda_{prob}$, $Vddd_{prob}$ and $Gnda_{prob}$) for the monitoring of the power supply voltages inside the two chips and two signals (Vtemp1 and Vtemp2) for the measurement of the temperature inside the two ASICs are connected to the ADC of the lpGBT circuit via FLEX cables. More details on the temperature monitoring are given in Section 10.4.3.

10.4.3 Environmental monitoring

Cooling system

The cooling system is based on the evaporative CO_2 2-Phase Accumulator Controlled Loop (2PACL) concept, extending the technology implemented for the ATLAS Insertable B-Layer (IBL) detector, while relying on industrial standards. It will be integrated in the general cooling system developed for the ATLAS ITk detector.

The on-detector cooling layout is detailed in Section 11.3. The cooling plant is protected against overpressure with safety valves set to 130 bar. This value is used as the maximum design pressure on the cooling loops. The cooling system parameters will be monitored using the DIP protocol.

Temperature monitoring

The temperatures of the sensor modules will be monitored in two independent ways: as voltages from temperature sensors, embedded in each ALTIROC front-end ASIC, via the same multiplexers and ADCs used for the module voltage monitoring, and from Negative

Temperature Coefficient (NTC) or PT10k sensors, via EMCI boards [87]. The temperature at the peripheral boards will be monitored through temperature sensors inside the lpGBTs.

The temperature measurements from the modules and the peripheral electronics is only available when the detector is powered. When the peripheral electronics is not powered the information about the temperature inside the detector vessel will be obtained from two sources:

- by means of NTC or PT10k temperature sensors located on the cooling plates, directly connected to off-detector EMCI units installed in the patch panels;
- and from the Interlock system, which will monitor additional NTC sensors installed on the detector modules, as described further below in Section 10.4.8.

The 10 kΩ NTC thermistors are good candidates for temperature sensors due to their high radiation hardness and the large signal that they produce, which support transmitting the signals over a long distance using only two wires per sensor. The signals of all these NTC or PT10k sensors will be routed via cables directly to input modules in the Interlock Matrix Crate (IMC). To optimize the use of local services, several temperature sensors may be interconnected inside the detector vessel using an OR logic. Information from the temperature sensors will be provided to DCS using EMCI boards, also located in the IMC crate.

The number of these temperature sensors has not been finalized yet, and is expected to be of the order of a few hundred.

Humidity and pressure monitoring

To keep a dry atmosphere inside the detector volume, an overpressure of the flushing N_2 gas must be maintained at all times. It is important to monitor the humidity inside the vessel and the pressure difference between the vessel volume and the UX15 cavern atmosphere.

Radiation hardness is an issue for humidity sensors. Studies are ongoing to select appropriate radiation tolerant sensors. The first option would be sensors based on optical fibres (FOS), that are being developed in ATLAS for ITk. Alternatively, the humidity can be measured at the exhaust of the gas system with standard humidity sensors in a low radiation area.

The overpressure monitoring can be implemented using pressure difference sensors, which can be located in the USA15 cavern and connected to the detector volume and the environment via two rigid pipes keeping the sensors away from high radiation areas. At the moment of writing this document the type of pressure difference sensors and their interface to DCS have not been defined yet.

10.4.4 Peripheral electronics

The peripheral electronics transfers data between the detector modules and the DAQ, DCS and luminosity systems. As mentioned in the previous sections, it has a central role in the monitoring of sensor temperatures and supplied low voltage. The system is based on CERN-developed lpGBT ASICs. In total 160 peripheral boards will be instrumented.

The detector modules are connected to the peripheral boards via FLEX cables, whereas signals to and from the DAQ, DCS and luminosity systems are transferred to the counting room over optical fibers. The DCS data and commands are embedded in the data streams via the DAQ optical fibers. Control signals to and from the ALTIROC ASICs are transmitted via I^2C bus where the commands and data are embedded in the data streams transmitted to and from the detector TDAQ system, as detailed in Section 10.1.

DC-DC converters

As mentioned in Section 10.4.2, the peripheral electronics also includes the 10 V to 1.2 V DC-DC converters for the digital and analog voltage supplies to the ALTIROC ASICs and the lpGBT ASICs, and the DC-DC converters for the Versatile Link plus (VL+). The DC-DC converters are based on the bPOL12V ASIC developed by CERN for the HL-LHC upgrade.

The voltages from the DC-DC converters on the peripheral boards and the voltages actually received at the front-end ASICs are monitored via multiplexers and ADC channels on the lpGBT ASICs of the peripheral boards, as described in Section 10.4.2. From the lpGBT ASIC the information is sent via optical fibers to FELIX boards of the TDAQ system for transmission to the DCS system. The optical links to the lpGBTs from the TDAQ FELIX boards will exchange data bits, embedded in the data streams, for switching on and monitoring the status of the DC-DC converters powering the front-end ASICs.

On-detector DC-DC converters will be used to power both the modules (ALTIROCs) and the electronics on the peripheral boards:

- 5472 bPOL12V supplying ALTIROCs (LGAD modules)
- 640 bPOL12V supplying lpGBTs and VL+'s

With 1996 modules per disk, 1188 DC-DC converters on the peripheral electronics per disk are needed to power the front-end electronics. A further 152 DC-DC converters per disk are required for powering the lpGBTs that will control the rest of the DC-DC converters on the board, and must be controlled by DCS over wires.

Each peripheral electronics board will receive an external 1 V control signal to switch on the DC-DC converters supplying the lpGBTs and the optical links. The status of these converters is reported and monitored through the lpGBTs via external electric cables (open drain) Power

Good output and on I/O lines other than those they supply to allow to differentiate between possible power failures and lpGBT failures. Further I/O lines on the DAQ lpGBTs are used for switching on and monitoring the status of the DC-DC converters supplying voltages to the ALTIROC ASICs. These DC-DC converters are switched on by applying a voltage (at least 850 mV) which is generated via general purpose I/O lines from DAQ lpGBTs. They are enabled through EMCI boards on the patch panel, and monitored through optical links to the FELIX boards in the counting room.

LGAD module monitoring

An additional requirement of the ALTIROC ASIC is to be able to monitor two closely related aspects of the LGAD: its operating temperature and its leakage current. More details about the temperature monitoring are given in Section 10.4.3.

Probing of the power voltages at the module level is useful to detect latch-up events on an ASIC. With the resolution of 1 mV of the lpGBT ADC and a parasitic resistance of 100 mΩ on the FLEX cable, minimal variation of 20 mA (considering an attenuation of 1/2 on the probing) can be detected, much smaller than the expected current rise in a latch-up event.

The analogue signals of monitoring coming from the modules are digitized by the converter implemented inside each lpGBT circuit of the peripheral board. The number of channels of this ADC being limited to eight, a multiplexing is required at the input of each channel. Multiplexers (MUX 64:1) are thus implemented to interface the signals coming from the modules to the ADC on the peripheral board. Using a multiplexer circuit which selects one output from 64 inputs, up to 8×64 signals can be interfaced to each lpGBT-ADC. A full custom 64-to-1 multiplexing circuit will be developed with a radiation tolerance suitable for its implementation on the peripheral board. The 6-bit bus required to control the addressing of each MUX is provided by the programmable parallel port of the lpGBT circuit which is controlled through its I^2C interface.

10.4.5 Configuration

The I^2C bus will be used to control and configure the ALTIROC ASICs as well as to configure the luminosity system lpGBTs and the DAQ lpGBTs. The DAQ lpGBTs are foreseen to be pre-programmed using e-fuses to accept configuration commands using DCS bits embedded in the DAQ data stream between the FELIX and the lpGBTs. For redundancy, each peripheral board (160 boards) will have a cable-based I^2C from EMCI units in the patch panel area. Embedded commands, via the DCS bits, in the bidirectional optical links between the FELIX and DAQ lpGBTs will be used for control and configuration of the luminosity system lpGBTs and the ALTIROC ASICs through I^2C-bus via masters on the lpGBTs.

10.4.6 DCS software

The HGTD DCS structure is shown in Figure 10.9. The DCS software will run on a local control station (LCS) in the ATLAS service cavern USA15. All DCS operations will be performed from this server. The DCS project will be integrated in the ATLAS central DCS. At a higher level, the ATLAS Global Control Station (GCS) controls all sub-detectors, collects data from external systems interfaced to the ATLAS DCS, such as the LHC collider status information or the Detector Safety System, and sends the data to sub-detectors via dedicated DCS Information Servers (IS).

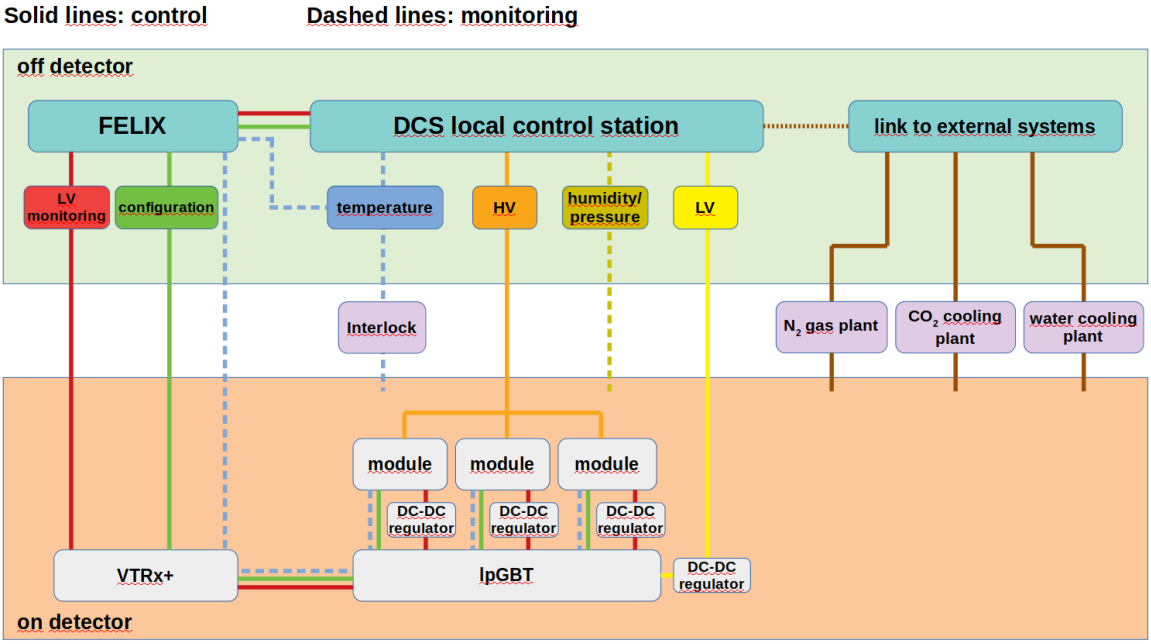


Figure 10.9: HGTD DCS layout

A Finite State Machine (FSM) structure will be implemented with rules for performing actions on the detector modules, the front-end and the back-end electronics and the infrastructure, while states will be propagated to the appropriate upper nodes. The DCS software consists of three layers. The lower layer establishes communication with different hardware (device) units. An intermediate layer is responsible for overall data processing, storing data to databases, mapping and calculations. The upper layer is responsible for overall detector operation and visualisation. The JCOP Finite State Machine FSM toolkit will be used to build a representation of the detector as a hierarchical, tree-like structure of well-defined subsystems, called FSM units. The HGTD FSM tree is shown in Figure 10.10. The tree consists of two main nodes: the infrastructure and the detector. The infrastructure node includes all common devices, while detector nodes are split first on a functionally level into

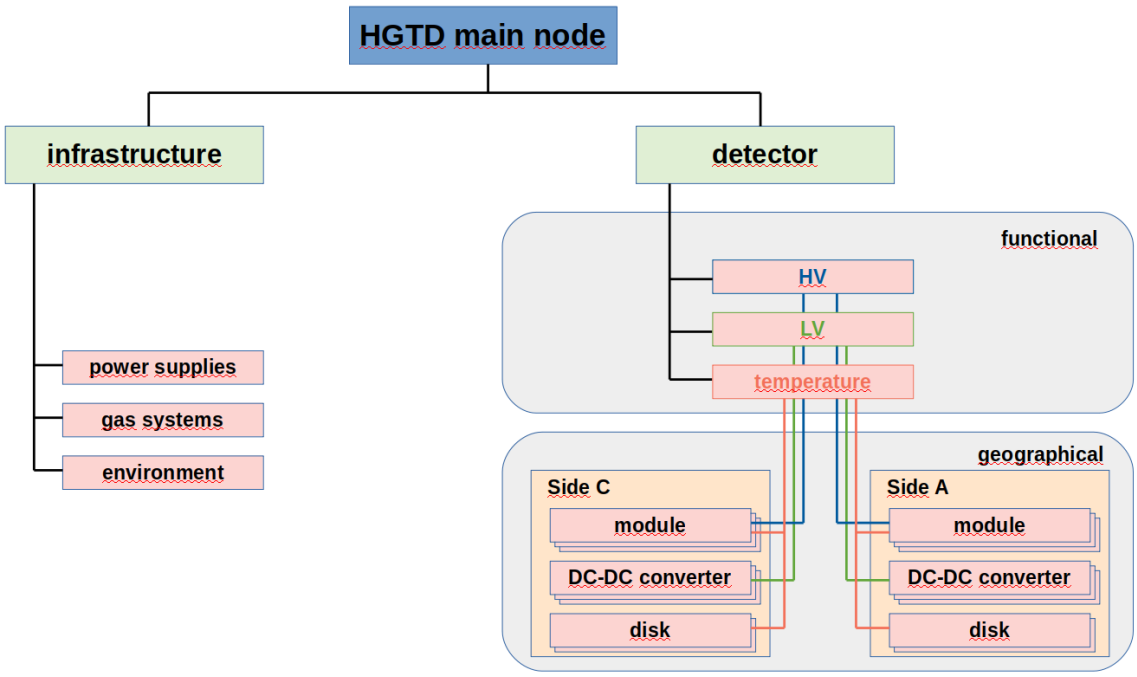


Figure 10.10: HGTD FSM layout

high voltage, low voltage and temperature, and then in a geographical level into the two vessels and down to the individual modules.

10.4.7 External systems

Beside the control and monitoring of the detector parameters, the DCS will help to protect the detector from various risks raised from infrastructure failures. Several external systems are essential for the optimal and safe operation of HGTD. These systems are under the responsibility of other groups, and hence there are no means of control by HGTD. Monitoring of several parameters from these systems will be put in place:

- CO_2 Cooling System (CCS)
- mono phase cooling system
- N_2 gas system
- Detector Safety System (DSS)
- Beam Interlock System (LHC-BIS)

These external systems are not connected to the ATLAS Technical Network ATCN but to the Technical Network.

The CO_2 cooling system provides the cooling of all module elements installed inside the HGTD volume. The main parameters to be monitored are: temperature set point, accumulator saturation temperature, plant state (ready, running, off, interlock), and status flags.

The mono phase cooling system provides cooling to all equipment outside the HGTD detector volume, e.g. the crates located on the patch panel. The main parameters to be monitored are: gas flow, pressure, states, and status flags.

The Detector Safety System (DSS, whose sensors watch the general environment such as cooling, presence of smoke or flammable gas in the air etc.) and the Beam Interlock System (BIS) provide hardwired signals which will be routed into the HGTD Interlock system and are used to switch off parts of the detector or even the entire detector in case of abnormal behavior of a parameter, failure or loss of communication. Such actions are imminent and may have a coarse impact on the detector. To deal with this the DSS actions can be delayed, allowing the DCS to implement more sophisticated control sequences on the equipment before the actions triggered by DSS are executed.

Beside the risks described in the previous sections, these systems handle additional safety conditions like e.g. smoke, cooling rack failures etc. Together with the hardwired signals, more information from these systems should be available via DIP, as e.g. for the handshake procedure, which defines the transition of the experiment from standby to data-taking and vice versa.

10.4.8 Interlock system

The HGTD Interlock system (HIS) is a standalone safety system that protects the detector against a variety of risks. The Interlock system must always be running and its components must never be disconnected. The HIS will be built according to the rules applied to safety systems. In particular, all its components must be connected by wire, it must be powered by an uninterruptible power supply (UPS) and a positive safety logic must be applied in the design. The last requirement means that any break in connections or loss of power would cause a failure in the system, which would result in the generation of interlock signals. HIS hardware will be designed and implemented in an Interlock Matrix Crate (IMC) located in USA15, similarly to ITk.

As one of the main dangers for silicon detectors is overheating, several hundred temperature sensors will be installed on detector modules to monitor their temperature, as described in Section 10.4.3. In the IMC crate the analog NTC or PT10k signals will be converted to binary signals by means of discriminators with a predefined threshold, and then processed by an Interlock Logical Unit (ILU) in accordance with the preprogrammed Interlock Action

Matrix (IAM). In parallel to the binary processing, information from the temperature sensors will be provided to DCS using EMCI units, also located in the IMC crate.

In the event that the temperature in any active zone of the detector exceeds 40°C, the power will be cut from all modules in that zone by interlocking the relevant HV and LV power supplies. In addition to temperature data, the signals related to risks due to common infrastructure failures or safety issues will also be included in the Interlock matrix. Various fault signals from the CO_2 cooling plant are processed by the DSS, which provides the resulting signal to the HIS indicating the loss of cooling power to the detector. In the event of external safety or environmental alarms such as signals from the cooling system, smoke or flammable gas detection, magnet vacuum or cryogenics failures (risk from water due to condensation melting), ATLAS emergency stop, flooding or ATLAS wide safe-for-beam interface, the DSS will request the HIS to switch off the power on the detector as a means to protect personnel and equipment. The signals corresponding to failures of common infrastructure, such as UPS power, rack cooling, N_2 gas system failures, will be available from the DSS system, and some of them, depending on their severity, will be included in the Interlock matrix, along with the unstable beam conditions signal from the Beam Interface.

All off-detector power supplies, HV and both type of LV (the 300 V bulk PS and 300 V to 10 V DC-DC converter units) must have an interlock functionality with sufficient granularity, to remove power from their outputs as soon as the interlock signal is raised by the HIS system.

10.5 Roadmap for DAQ, luminosity and control system

Different working groups have been defined for DAQ, luminosity and DCS. General DAQ activities including ATLAS common readout and calibrations will be integrated in the demonstrator activities described in Section 14.3.

Development of the luminosity data handling in the backend electronics will be initiated during 2020. The FELIX board has been delivered to CERN and is undergoing final tests by the TDAQ group before being handed over to HGTD. Initially input data will be generated within the FPGA, subsequently tests will be carried out with signals generated on an external FPGA and communication via optical links. The first iteration of the ASIC which will be capable of sending occupancy data will be ALTIROC2, once these ASICs are available integration tests of ALTIROC2 communication via optical links to the luminosity backend FELIX will be tested. In parallel to the tests of the backend electronics, the Luminosity Software will be written and tested, first with generated input data and thereafter with communication from the FELIX to the host where the Luminosity Software is running. An FDR is planned for Q1 of 2024, before launching the preproduction, and an PRR in 2025, as outlined in Figure 15.5.

The preliminary design review for the DCS and interlock system will start on Q1 of 2022 as described in Table 15.6. The installation of the EMCI/EMP boards and DCS servers is planned to start on Q2 2024. The standard DCS software (SCADA, OPC servers) will then be installed and commissioned until Q3 of 2026. The interlock system consists of temperature sensors for the detector modules and EMCI/EMP boards for read-out, the Interlock Matrix Crate, and the connections to the DSS and external systems. The installation of the different items of the interlock system will start on Q2 2024 and commissioned on Q3 2026. The connection of the servers to the network and the hardware (power units, EMCI/EMP boards and interlock) will start on Q4 of 2026, followed by tests to verify the hardware connections and the response of the software. DCS hardware and software must be operational before the installation for testing and commissioning.

11 Detector Mechanics

11.1 Engineering design overview

This chapter describes the global detector structure and its main mechanical sub-assemblies, in particular the hermetic vessel, the front and back covers, the inner and outer rings, the moderator, the on detector support and cooling disks, the bolting and the alignment device to LAr calorimeter end-cap cryostat wall and the peripheral cooling lines. The cooling system, a common project between ATLAS and CMS, is also presented including cooling requirements and main components from the chiller up to the detector hermetic vessel. A summary schedule of the HGTD can be seen in Figure 15.7.

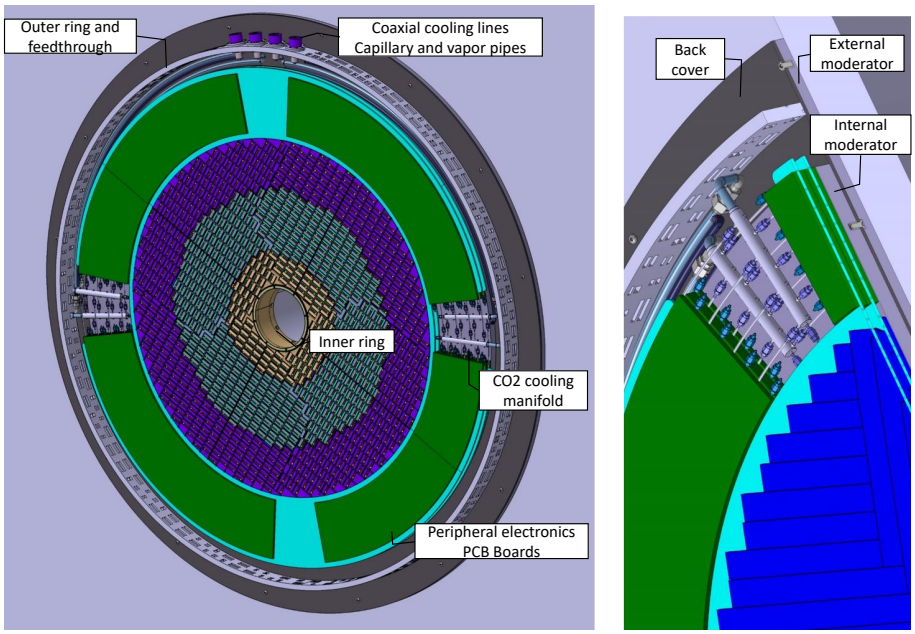


Figure 11.1: General view of the HGTD detector showing the silicon sensors inside the hermetic vessel. The green outer crown is the peripheral electronics limited by the outer ring which is holding the total amount of tight electrical connectors and the proximity cooling lines.

As presented in previous sections, the space allocated to the HGTD equipped vessel is very limited in (r,z). In addition, the routing of the services should fit inside a gap of 17 mm

in z against the end-cap calorimeter wall. These requirements are a challenge for many of the engineering parameters, like the stiffness and thermal insulation of the hermetic vessel, the thickness of the flex and connectors, the size of the support and cooling plates with embedded CO2 channels and manifolds, the peripheral electronics boards and their tight connectors. Due to life time maintenance, the detector must be designed for easy and fast integration into the ATLAS detector, and it should be constructed to permit quick removal and re-installation of the active layers in the high-radiation environment while maintaining the beam pipe in position.

The HGTD system includes two identical detectors fixed at both calorimeter end-caps. The various components of a single detector are shown in Figure 2.4. They consist of a cylindrical hermetic cold vessel (front cover with heaters and back cover, also with heaters, both bolted to the inner and outer rings) that encapsulates two instrumented disks and an inner part of the neutron moderator. Each instrumented disk (Figure 11.1) represents a cooling support plate composed of two separate half disks with silicon modules installed on both sides, as shown in Figure 2.9. The radial extent of the active area is 120 mm to 640 mm, which yields an acceptance from pseudo-rapidity of 2.4 to 4.0.

To protect the ITk and the HGTD from back-scattered neutrons produced in the end-cap and forward calorimeters, 50 mm of moderator is installed in front of the LAr end-cap cryostat, as in the current ATLAS detector. The envelope in z for the full detector, including the moderator, supports, front and back covers, and the free gap with calorimeter front wall is 125 mm (or 75 mm without the moderator). The moderator is made out of two disks of different radii to provide more peripheral space inside the vessel. This space allows electrical services, tight electrical connectors and CO2 distribution lines to fit inside the restricted envelope.

The detector will partially occupy the ATLAS end-cap regions that presently house the Minimum-Bias Trigger Scintillators (MBTS) and moderator. The cold vessel will be located at z positions of 3420 mm $< z <$ 3545 mm from the interaction point. The mid-plane of the first and last active layers will be located at $z = 3446$ mm and $z = 3472$ mm. The position of the two HGTD end-caps within the ATLAS detector is shown in Figure 2.3. The overall dimensions are summarised in Table 2.1. The total weight per end-cap is estimated to be 350 kg including the moderator disks and to be 275 kg without the external moderator disk. The heaviest components are the internal and external disks of the moderator, amounting to 75 kg each, followed by the half-circular instrumented disks, weighing 30 kg each.

11.2 Detector overall layout

An illustration of the HGTD detector components is shown in Figure 2.4. The front view of the two double-sided layers that will be placed on each end-cap are shown in Figure 11.2. In this drawing they have a rotation of 20° with respect to each other to facilitate the entrance

of the cooling pipes inside the cooling disks. A detailed depiction of the detector in the

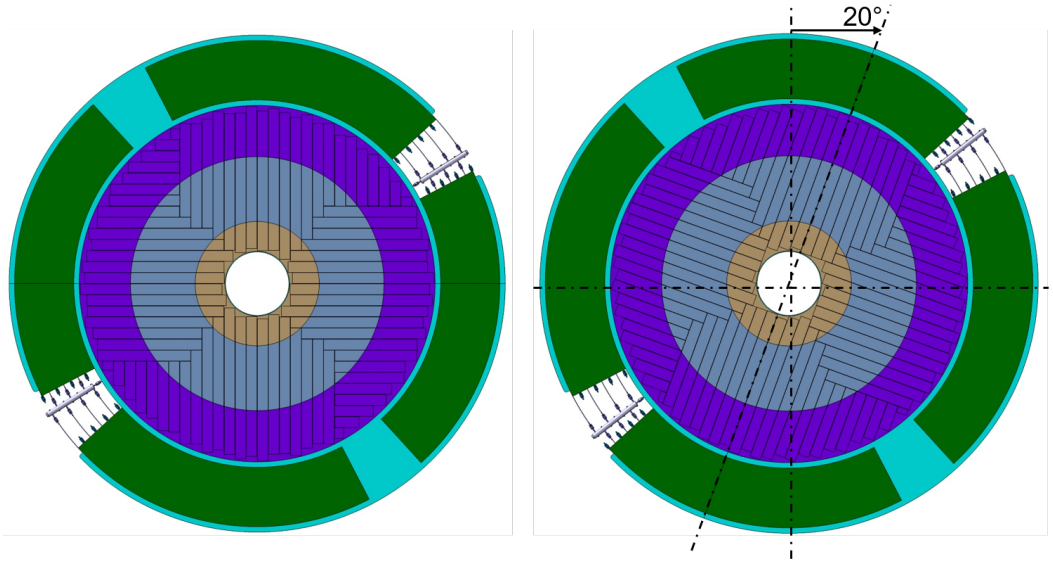


Figure 11.2: Front view of the two double sided layers that are placed inside the HGTD vessel. These two disks (right and left figures) have a rotation of 20° with respect to each other to take into account the needed space for the PEB (Peripheral Electronic Boards).

5883
5884 (r,z) direction, in the inner radius region close to the beam pipe, is shown in Figure 11.3.
5885 It includes two cooling/disk supports where the double-sided layers of the detector are
5886 mounted, the front and back covers of the vessel and the inner and outer layers of moderator.
5887 The full assembly, including 50 mm of moderator, will match the envelope of 125 mm in the
5888 z direction.

5889 A detailed breakdown of the (r,z) dimensions of the detector components is presented in
5890 Table 11.1, and also the materials and estimated weight of various components. The bottom
5891 of the table lists each component of a double-sided layer of detector modules mounted on the
5892 cooling support. The measured thickness of the current prototype of the sensor-ALTIROC
5893 ASIC assembly is about 1 mm thick. This gives a comfortable margin with respect to the
5894 final envelope assembly protocol, with an expected thickness of module package (detector
5895 unit) of 4.2 mm. Considering the longest readout row, the maximum amount of stacked
5896 flex cables will be 19. With the estimated thickness of one flex cable of 0.22 mm, it gives
5897 the total thickness of flex cables stack of 4.2 mm per side. Considering the additional 1 mm
5898 integration gap, it should be possible, though challenging, to fit all the components within
5899 the design envelope.

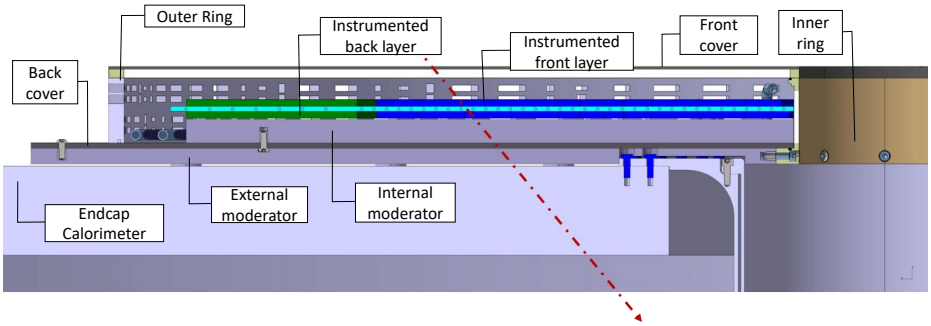


Figure 11.3: (R,Z) cross section of the detector assembly from beam pipe axis up to service penetration outer ring. It details the two instrumented double sided layers, installed on the cooling and support plates, the front and back covers, the internal and external moderators, as well as the inner ring centred on the calorimeter central tube.

11.3 CO_2 cooling system

The cooling system is based on the evaporating CO_2 2-Phase Accumulator Controlled Loop (2PACL) concept. It will be integrated with the general cooling system developed for the ATLAS ITk [85]. CO_2 cooling is chosen because it makes significant mass savings inside the detector possible, due to the use of tubes of smaller diameter than in systems, which are based on conventional cooling liquids. CO_2 evaporates at much higher pressures than common refrigerants, keeping the vapour compressed and therefore the volume low. The boiling temperature depends on the pressure and, as this pressure is relatively high, a pressure drop in the lines due to small-diameter piping does not cause much change in the evaporating temperature. In addition to the benefit of high pressure, CO_2 also has a low viscosity and high latent heat, so that less flow is needed than with other refrigerants. The narrower pipes can accommodate much higher flow speeds, which is a benefit for the overall boiling heat transfer coefficient.

Taking into account the radiation environment in which the HGTD will operate, CO_2 is one of the most appropriate refrigerants because of its radiation hardness and low activation. The CO_2 will be pumped in liquid state from an external primary chilling source and will partially evaporate as it absorbs the heat dissipated by the HGTD components. Within each pipe, a small amount of CO_2 flows at high pressure in the form of small drops, and enough space is left for the vapour to circulate. A highly-efficient heat extraction is achieved by making use of the large latent heat for a liquid to vaporise, meaning that not only less fluid is

HGTD components per end-cap	**Thickness** (mm)	z_{in}/z_{out} (mm)	R_{in}/R_{out} (mm)	**Weight** (kg)
Vessel Front cover	13.0	3420/3433	110/1000	25
Front double side layer (2 half disks)	26.0	3433/3459	120/920	60
Rear double side layer (2 half disks)	26.0	3459/3485	120/920	60
Internal Moderator	30.0	3485/3515	120/900	75
Vessel Back cover	7.0	3515/3522	110/1100	15
Vessel inner ring	10.0	-	110/120	5.0
Vessel outer ring	20.0	-	980/1000	35
External Moderator	20.0	3522/3542	110/1100	75
Air gap with LAr cryostat	3.0	3542/3545	110/1100	75
Total envelope from the LAr cryostat wall	**75.0**			**275**
Double side layer breakdown	**Thickness** (mm)			
Air gap with vessel or with moderator	2			
Flex tail packing (0.22 mm per unit)	4.2			
Module package	4.2			
Cooling + support plate	6			
Module package	4.2			
Flex tail packing (0.22 mm per unit)	4.2			
Inter-layer gap	1.2			
Total per double sided layer	**26.0**			

Table 11.1: HGTD components per end-cap. The top part of the table shows the components with their dimensions in z, r and their weights. Each double sided layer is divided in two half circular disks of 30 kg each. The total weight of the detector, including the moderator is 350 kg (275 kg without the external moderator). The bottom part of the table shows a breakdown of the front double sided layer. The breakdown of the back layer is identical.

needed to extract a certain amount of heat but also that the temperature of the liquid phase remains constant, while that of the vapour increases only slightly. The cooling power is then determined by how much CO_2 is left in a liquid state. Because it is used in mixed states (liquid and vapour), a significant mass reduction is introduced when comparing with other liquid mono-phase refrigerants.

11.3.1 Requirements

An operation temperature of −35 °C must be maintained inside the HGTD vessel, in the vicinity of the cooling channels close to the modules, with a stability of a few degrees Celsius. As discussed in Chapter 5, the operating temperature must be kept as low as possible as, after irradiation, the leakage current of the sensors increases with temperature. The operating temperature of the peripheral on-detector electronics is flexible. It can be in the range of

–35 °C up to 20 °C, making the cooling and stability requirements of these components less stringent. Taking into account that these electronics are located within the cold vessel, they will need to be maintained at a temperature close to the sensor operation point to avoid excess heat flowing towards the sensors. The electronics will be used as pre-heaters to stabilise the cooling parameters before the coolant reaches the modules.

Table 11.2 summarises the power consumption estimated for the various components of the detector. This defines a need for maximum cooling power of 40 kW in total (20 kW per end-cap) at the end of life time of the HL-LHC. However, most of the components listed in the table are not yet fully designed, therefore the estimate of the total maximal power consumption has about a 10 % uncertainty. A careful re-evaluation of the power consumption of each component will be done with the first prototypes.

HGTD Component	**Power consumption**	**Total [kW]**
Sensor	30 to 100 mW cm^{-2}	2.0–6.4 (*)
ASIC	$<$ 300 mW cm^{-2}	17.6–19.2(**)
Flex cable	6.8 mW cm^{-1}	2.0
Total in active region		**21.6–27.6**
HGTD vessel heaters	100 W m^{-2}	1.3
Pre-heaters (Perip. electr.)		8.8
Ambient pick-up		2.5
Total power dissipation		**34.2–40.2**

Table 11.2: Total maximum power consumption estimates for the HGTD at the start and end of the HL-LHC. A breakdown for the various components is also given. (*) The sensors power consumption range from 30 to 100 mW cm^{-2} expected respectively for sensors non irradiated (at the start of the HL-LHC) and irradiated at the max expected irradiation of 2.5x10^{15}neq/cm^2. (**) The 19.2 kW corresponds to 1.2 W (or 300 mW cm^{-2}) consumed by each ASIC when calibration is taking place and is equivalent to 10% occupancy of all channels of an ASIC. During normal data taking, the total power consumed by the ASIC is 17.6 kW, smaller than during calibrations.

The ASICs, followed by the sensors, consume the most power, with up to 300 mW cm^{-2} by the ASIC and up to 100 mW cm^{-2} by the sensors at the innermost radius. The power dissipation of the ASICs decreases slightly as a function of their radial position because the hit rate decreases at larger radius, as shown in Figure 11.4. Taking this radial dependence into account, the total power consumed by the ASIC amounts to 17.6 kW during data taking. The total power consumed by the ASIC increases to 19.8 kW when calibrations are taking place and is equivalent to 10% occupancy across all channels in the ASIC, for all ASICs.

The power dissipated in the flex cables is expected to be 6.8 mW cm^{-1}, leading a total power dissipation for the flex cables about 2.0 kW

The peripheral electronics boards will act as pre-heaters for the cooling system. On these boards, the DC-DC converters will be the component with the highest power dissipation. As-

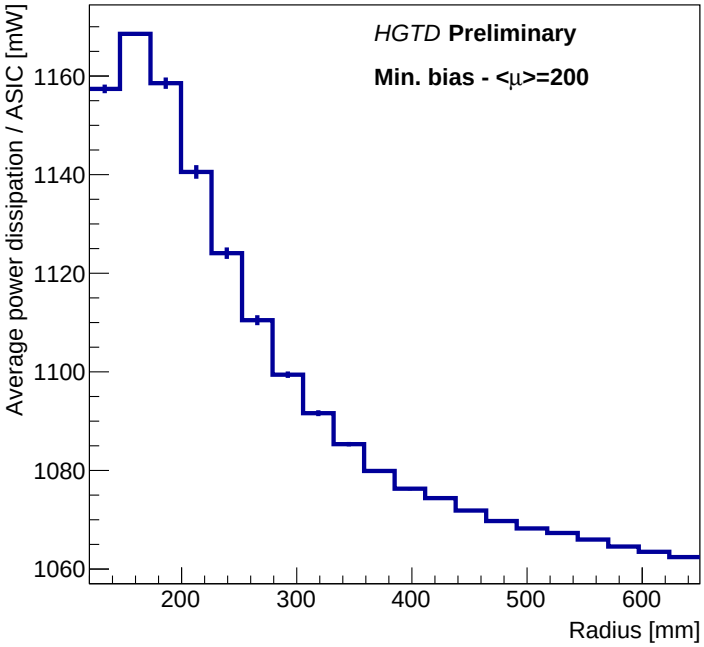


Figure 11.4: Average power consumption per ASIC (in mW) as a function of the ASIC radial position relative to the beam pipe axis. Each ASIC is 4 cm^2.

suming a 72% efficiency for the DC/DC converters, the peripheral electronics will dissipate an estimated power of 8.8 kW.

Given the uncertainties on current estimates of the power dissipation of some components, a cooling unit dedicated to HGTD of 50 kW will be constructed (25 kW per end-cap). A spare cooling station of 50 kW, shared with ITk, is also foreseen.

11.3.2 Cooling design

The cooling design is based on the technology implemented for the ATLAS Insertable B-Layer detector and on industrial standards. Tri-axial vacuum-insulated transfer lines will be used to connect the CO_2 cooling station located in USA15 to a junction & distribution box to be located on the outer radius of the end-cap tiles calorimeter on the HO side, close to the HGTD patch panel area, detailed in Section 12.3. One such box per end-cap will be used to distribute the CO_2 flow from one big transfer line to four smaller proximity lines. A second function of these boxes, which is being studied, should maintain the detector under cooling conditions during ATLAS Short opening (winters' YETS). Additional lines, which are also under study, should provide detector cooling during ATLAS large opening.

When the cooling is turned off, due to transfer lines disconnection or any other unexpected operating failures, the temperature inside the vessel could increase up to room temperature. The main strong source is the anti-condensation heaters which should also be switched on full time to prevent any temperature decreasing on the hermetic vessel outer skin. Another reason is the N_2, blowing at 20 °C, with a flow rate up to 750 l h^{-1}, improving the warm up of the on detector parts.The estimated warm up time to reach 20 °C from −35 °C of the HGTD cold mass (200 kg, corresponding to the on-detector system and moderator inner part (see Table 11.1)), is determined by the equivalent specific heat capacity (c in J kg^{-1} K^{-1}) of the cold mass. Considering the thermal power input as 650 W, mainly from the heaters, and the equivalent specific heat in the range of $c = 750$ J kg^{-1} K^{-1}, the increasing gradient is about 4 min per degree centigrade, and a total of 3–4 hours to reach room temperature.

Rigid proximity transfer lines are under development for Phase-II upgrade applications for ATLAS and CMS targeting a transfer capacity of about 5 kW per unit. The HGTD design places an inner hose, with an inner diameter of 5 mm for the CO_2 liquid, inside a 16 mm mid-hose for the vapour return. This hose, made out of a multi-layer insulated (MLI) pipe, is enclosed within a vacuum hose of outer diameter less than 50 mm. The vacuum level inside the transfer lines must be less than 1×10^{-4} mbar in order to avoid convection and condensation on the outer wall. The relatively small outer diameter of such lines, less than 50 mm, will facilitate their routing in the gap between the barrel and end-cap calorimeters, through a dedicated slot in ϕ allocated inside the original ITk envelope, as agreed with the ITk and Technical Coordination groups.

In order to prevent connection and disconnection of CO_2 transfer lines during long shut-downs, alternative flexible lines are under study which could be implemented in a dedicated flexible chain along ATLAS translation rails.

The four tri-axial rigid lines, one for each half-disk cooling plate, enter the HGTD vessel at the top position (11.25° from the vertical line). They are holding capillary lines with 0.75 mm inner diameter and length up to 5 m, ending inside the hermetic vessel at the manifold R-ϕ location. They supply CO_2 liquid to the 8 cooling loops that are embedded in each half-disk cooling plate on a semi-circular concentric pattern, as shown in Figure 11.5. The radial distance between the concentric pipes in the loops at 120 mm $< r <$ 640 mm is 16 mm. This is the region covered by active modules placed on both faces of the cooling disk with overlap from 20% up to 70%. In the peripheral electronics area, at $r >$ 680 mm, where the dissipated power is used as pre-heaters, the distance between pipes is increased to 30 mm to take into account the lower heat dissipation, thus keeping a uniform temperature distribution on the total area of the cooling disk.

The choice of the pipe material should take into account several parameters such as the mechanical properties, the thermal deformation, the thermal runaway (see section Section 7.5), the induced radioactivity and the material radiation length. Two options for the material to be used are being studied : titanium and aluminum, while the stainless steel is only used for bending and assembly feasibility.

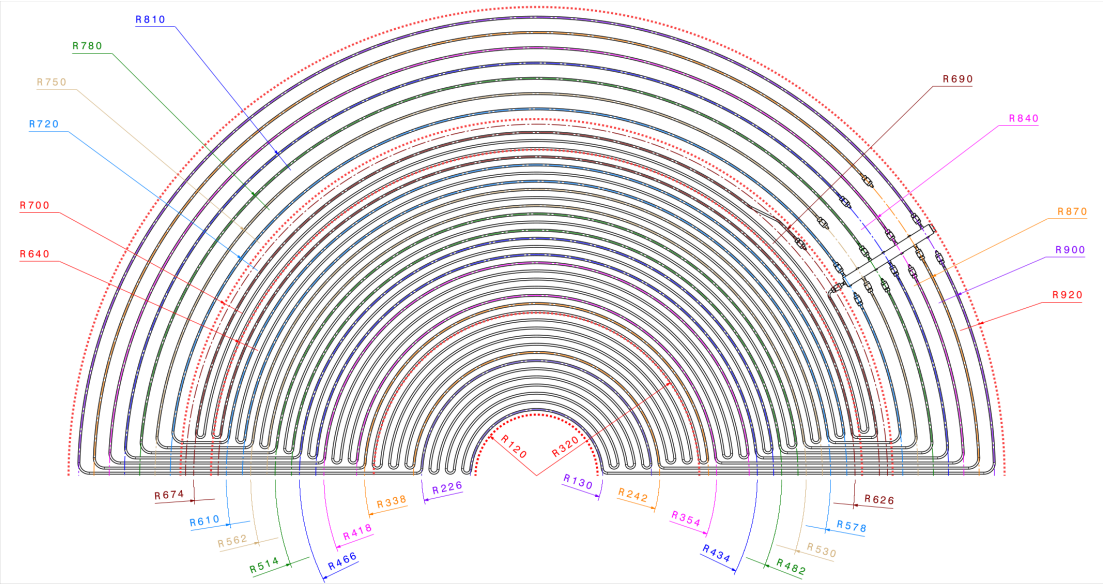


Figure 11.5: Detailed layout of the 8 cooling loops on a half disk support. The central cooling loops (below 640mm radius) with a pitch of 16mm are dedicated to the silicon sensors & their ASICs. The outer loops with a pitch of 30mm are cooling the Peripheral Electronic Boards (PEB).

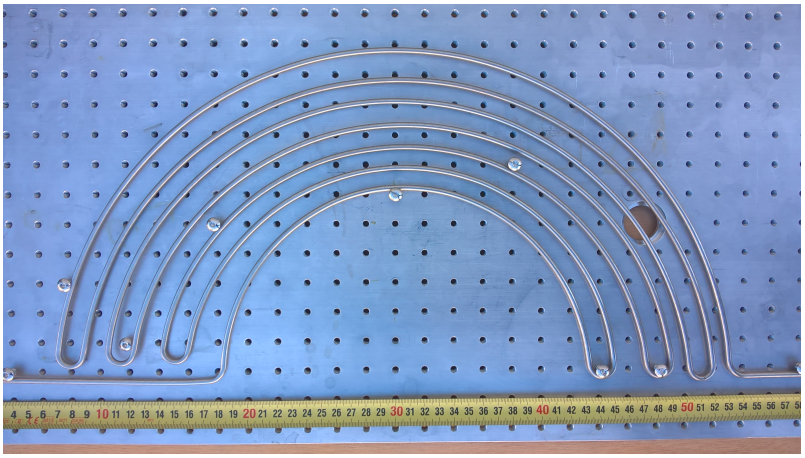


Figure 11.6: Cooling loop prototype corresponding to the inner part of the half-disk support.

A first prototype has been made of stainless steel (4.0 mm O.D and 5.0 mm thickness widely available and easy to machine). It is shown in Figure 11.6. This prototype corresponds to the inner region of the cooling half-disk with a radial spacing of 16 mm as foreseen for the HGTD. It has been successfully tested up to 165 bar at the CERN proof pressure facility. Thermal tests will be undertaken with the CERN CO_2 cooling setup before integration into the sandwich structure of the cooling support. These measurements will be applied to validate parameters used in the thermal runaway studies. However, stainless steel is not

considered as candidate for the cooling pipe material due to its short radiation length and induced activity (stainless steel Xo 13.84 , equivalent to 1.757 cm).

The current baseline is to make the cooling pipes of titanium T40 grade 2 or equivalent, as used in to the IBL and ITK projects (titanium Xo 16.16 $\mathrm{g\,cm^{-2}}$, equivalent to 3.56cm, almost double of the stainless steel value). A prototype of the inner region is going to be manufactured with T40-G2 pipes, 4.0 mm O.D and 3.0 mm thickness.

Cooling pipes made of Aluminum may still be considered for the final detector to reduce the radiation length between the active layers and thereby improve the ability to associate ITk tracks with HGTD hits (aluminum Xo 24.01 $\mathrm{g\,cm^{-2}}$, equivalent to 8.897 cm, almost 2.5x the titanium value). In addition, Aluminum is less activated by radiation and therefore may allow faster access to the detector components. This is important for replacing the over-irradiated inner rings and for maintenance during long shutdowns. The thickness of the Al cooling pipes would have to be larger than with Titanium to sustain the pressure, but would have the advantage of the same thermal and mechanical properties as the sandwich structure of the support plates. The welding of the Al pipes to the stainless steel fittings, at the manifolds level, is more challenging than with Titanium. Bimetal transitions (aluminum-stainless steel) can be used to fulfil these specific piping connections.

The cooling plant is protected against over-pressure with safety valves set to 130 bar. This value is used as the maximum design pressure on the cooling loops. To ensure that the pipes can sustain such levels of CO_2 pressure, the wall thickness of the pipes must be at least 0.3 mm. The outer diameter of the pipes is 4.0 mm. Their length varies from 4 to 6 m for different loops. The maximal transfer capacity of the cooling loops corresponds to 100 $\mathrm{W\,m^{-1}}$. The characteristics of the loops are defined in close collaboration with the CERN Cooling group.

The half disks with embedded cooling loops are the main support structure for the instrumented active layers, as described in Section 11.6. In addition to their high thermal conductivity, their own stiffness should guarantee a surface disk flatness within one millimetre. Given the challenging performance requirements of the on-detector cooling system, one full scale prototype of cooling half-disk support will be produced, including aluminum panels and embedded cooling loops, equipped with appropriate heaters to simulate the silicon modules power dissipation. This prototype will be subjected to several thermal cycles to study the thermo-mechanical behaviour, temperature distribution, CO_2 cooling parameters, and the performance of conductive media needed in between the modules, the support plates and the cooling channels.

11.3.3 Cooling plant demonstrator

One important milestone for the cooling development is the proof that the CO_2 evaporation temperature of −35 °C can be achieved at the local HGTD support disks with realistic transfer

lines and coolant distribution. Because of the critical importance of this technology in the ITk and HGTD systems, a CO_2 cooling test facility called "Baby demonstrator"was set up by the CERN cooling team in collaboration with ATLAS and CMS. This facility, which is being tested, is installed in Building 180, next to the mock-up of the ATLAS calorimeter (see Figure 11.7). It will be used for tests of prototypes of ITk and HGTD cooling components with a real-scale geometry. This chiller demonstrator will operate at low temperature with a

Figure 11.7: CO_2 cooling plant demonstrator located in Building 180 at CERN.

limited cooling power of 5 kW. The fluid transfer is subject to losses, which, in a two phase system, appears as a drop of saturation temperature on the return line due to the frictional pressure drop of the flowing media and static height differences. The main results were already presented in [85], in the context of ITk. As an example, Figure 11.8 shows a typical temperature distribution in the cooling system from the CO_2 plant to ITk on-detector loops and back, reaching the temperature of −40 °C, the target temperature for the ITk modules. To provide this temperature in the detector units, the cooling plant temperature needs to deliver −45 °C to account for the estimated 5 °C lost in the distribution and transfer lines.

In order to optimise the performance of HGTD local supports at −35 °C, specific prototypes as well as the half disk cooling supports will be submitted to real scale CO_2 tests on the Baby-Demo facility at CERN.

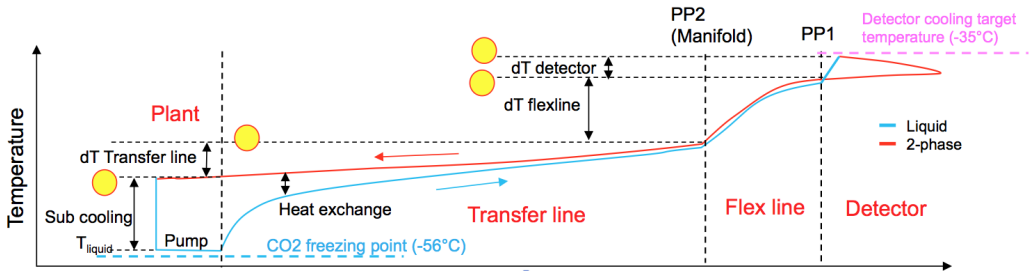


Figure 11.8: Typical temperature distribution between the CO_2 cooling plant and ITk loop [85].

11.4 Moderator

The moderator, to be placed between the end-cap calorimeters and the active layers of the detector, will protect both the ITk and HGTD against the back-scattered neutrons that are produced by the end-cap calorimeters. The moderator disks will be made of borated polyethylene with a density of $0.95\,\mathrm{kg\,L^{-1}}$, similar to the one used in the present ATLAS detector. As seen in Figure 2.4, the new moderator will be divided into two disks per end-cap, one inside and one outside the HGTD hermetic vessel.

The moderator on the outside is mechanically separated from the HGTD hermetic volume. It will be screwed to the LAr cryostat wall with an air gap using spacers of a few millimetres and will provide the necessary flat surface on which the HGTD will be installed with accessible bolting brackets. In order to minimize the mechanical impact on the LAr end-cap cryostat, the vessel interface with the cryostat wall will be made using the same threaded holes that are at present used to mount the MBTS. To allow the integration of anti-condensation heaters on the back cover, specific thin pockets, matching the heaters footprint, will be machined on the moderator surface with associated radial grooves to route power and monitoring cables. This moderator has a thickness that varies along the radius, 10 mm only in the region $180\,\mathrm{mm} < r < 342\,\mathrm{mm}$ (to absorb the over thickness with respect to the cryostat wall due to the LAr central flange and its bolts head) and 20 mm elsewhere ($140\,\mathrm{mm} < r < 180\,\mathrm{mm}$ and $342\,\mathrm{mm} < r < 1100\,\mathrm{mm}$) (see technical drawing in Figure D.4). The weight of this external moderator is in the range of 75 kg.

Potential conflicts with the water cooling pipe, currently installed on the cryostat front wall and used for cooling of the beam pipe during the bake-out procedure, require verification at the LS2 time-slot. The goal is to optimize the water cooling pipe shape, elbows and fittings in order to minimize the grooves size to be machined on the external moderator.

The part of the moderator to be placed inside the vessel has a thickness of 30 mm, a radial coverage of $120\,\mathrm{mm} < r < 900\,\mathrm{mm}$, and a weight of about 75 kg (see technical drawing in Figure D.5). It provides appropriate R-ϕ sliding support for the instrumented layers and, because it does not extend to radii higher than $r = 900\,\mathrm{mm}$, it leaves enough free space for

the cooling services as shown in Figure 11.13 left and right details. In each end-cap, the total moderator thickness in z, summing the two disks, will then be 50 mm, except at the inner and outermost radii. There it will be 40 mm in the region between 110 mm–342 mm and 20 mm for $r > 900$ mm. During maintenance, and when the replacement of the radiation damaged modules takes place at the surface, the outer moderator disk may stay bolted on the LAr cryostat, while the back cover is moved up with the HGTD vessel.

11.5 Hermetic vessel

The hermetic vessel is the primary integration structure of the HGTD detector. It consists of four main components made of composite structures in carbon fiber, as seen in Figure 2.4: the front and back covers, the inner ring and the outer ring which will hold all the service connectors and the cooling line flanges. The vessel measures 1100 mm at the outer radius and 110 mm at the inner radius (see technical drawing in Figure D.6 for a view of HGTD vessel with components). The thicknesses of front and rear covers are 13 mm and 7 mm, with an estimated weight of 25 kg and 15 kg, respectively. The Faraday cage will be performed by using aluminum mesh tightly integrated to the hermetic vessel outer skin, similar to what is currently used in the other LHC experiments.

11.5.1 Requirements

The hermetic vessel provides a robust support structure to the detector instrumented disks in a cold and dry over pressure volume (+10 mbar maximum relative to atmospheric pressure). All materials chosen must satisfy safety requirements related to the expected radiation levels, described in Section 2.4, and the operational temperature range (OTR). They shall also comply the CERN safety instruction IS41 (Fire safety rules), in particular flammable resin epoxy composites are not allowed. Assuming no replacement of components during the HL-LHC, the materials used have to withstand $8.3 \times 10^{15}\,\mathrm{n_{eq}\,cm^{-2}}$ and 7.5 MGy, including safety factors. Components that will be replaced midway through the HL-LHC will see these criteria divided by two.

The safe temperature range is defined by the acceptable minimum coolant temperature, −35 °C, and the expected module interlock temperature, 30 °C, with a margin of 20 °C on both sides. This results in a safe OTR from −45 to 40 °C and a 100 thermal cycles life time, which is similar to recent ITK engineering specifications.

The vessel tightness should ensure the detector volume permanently dry, keeping the dew point at −60 °C or below, to avoid condensation on the detector components. Considering 5 mbar nominal over pressure of the 125 liters dry nitrogen volume, and an acceptable pressure drop of 10%, which is equivalent to 0.05 mbar, the leak rate has to be better than

1.75x10-3 mbar.l/sec. This requirement can be achieved by permanent flushing with dry N_2 at 0.5% over pressure above atmospheric reference. The N_2 flow will renew gas in the vessel volume up to 10 times per hour, which is equivalent to 750 l h^{-1} per end-cap. For this purpose, the HGTD vessel was designed to be as hermetic as possible, in particular the electrical connectors and cooling flanges at the outer ring.

Another requirement is to keep the temperature of the outer surface of the HGTD vessel safely above the cavern dew point ($\sim$12°C). This will be achieved by placing flat Kapton heaters on the external skin of HGTD hermetic vessel and as close as possible to the outer ring, due to the high thermal conductivity of the electrical services, in particular when the detector power is turned off while the cooling is maintained.

11.5.2 Front cover

The front cover is designed as a sandwich structure, consisting of a Nomex honeycomb core placed between two thin Carbon Fibre Reinforced Plastics (CFRP), as shown in Figure 11.9 (see the technical drawing in Figure D.6 for a detailed view of the front cover) . As a means to reduce the front cover deflection from over pressure and CTE mismatch, radial stiffeners are integrated into the structure during the curing process of the epoxy composite. Considering the possible opening of the hermetic vessel with the beam pipe in place, during YETS maintenance for example, the front cover is designed to be two half-moon parts with vertical junction edges. These edges are manufactured out of PEEK reinforced 30% carbon fiber as baseline design to provide stiffness and low thermal conductivity of the front cover. Other similar material, free of epoxy resin, like Torlon polyamide-imide technology are also an alternative solution.

The tightness is the result of a trapezoidal gasket shape (EPDM or NBR radiation resistance elastomer) compressed in between the two screwed half moon parts (see Figure 11.10). This gasket is also compressed against the inner and outer ring elastomer gasket to provide tightness continuity between the front cover and the inner and outer ring. The stiffness of the vessel assembly when mounted on the cryostat wall has been studied using FEA 3D model. In this computed assembly, 5 mbar over-pressure has been applied, corresponding to dry nitrogen blowing inside the vessel to prevent any ambient humidity leak from outside. The results are presented in Figure 11.11, showing a maximum deflection of 1.5 mm on the front cover. This is equivalent to a maximum stress (Von Mises) of 70 MPa, located on the carbon fiber panels. The composite rods along the two half disks of the front cover will change the global stiffness of the honeycomb panels and will absorb the induced stress. A safety factor of 1.5 is considered to take into account the dry nitrogen network differential pressure relief valves setting up to 1010 mbar.

The HGTD inner volume will be operating at a temperature of −35 °C, therefore heaters will be required on the external faces of the hermetic vessel to prevent condensation on

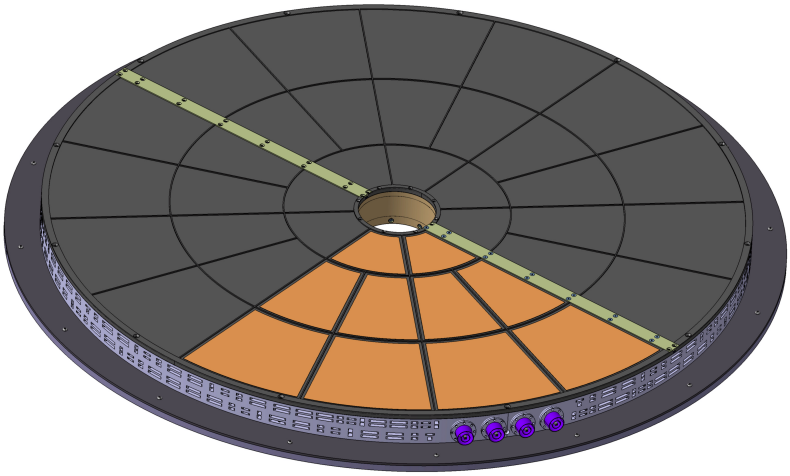


Figure 11.9: General view of the closed hermetic vessel. The front cover (Kapton heaters partially shown) is made of two parts which are bolted together with composite bracket and tight gasket. The outer ring is holding all electric connectors and coaxial cooling flanges.

.

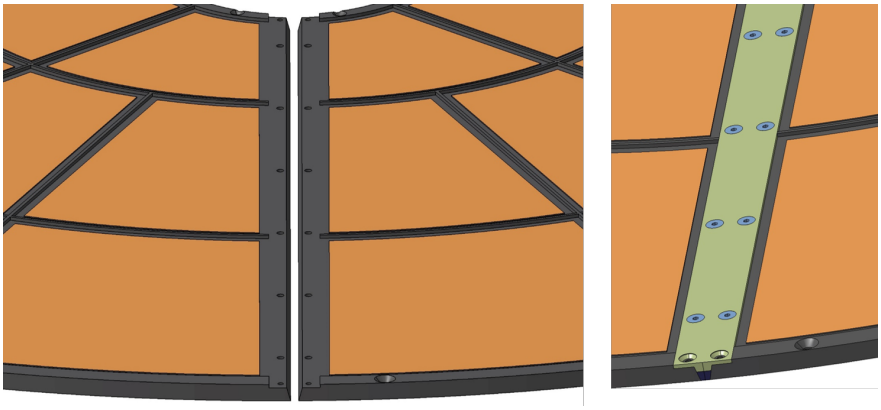


Figure 11.10: Detailed 3D view of the front cover junction connecting the two half moon parts. The central rod, on the right, is bolted both on the two parts with trapezoidal elastomer gasket in place

.

the outer surfaces. In a way similar to what is done on the LAr end-cap cryostat front face, heaters will be placed on the external face of the front and back covers, the inner and outer rings. Their purpose is to ensure a minimal temperature of 14 °C outside the HGTD vessel, safely above the cavern dew point of 12 °C. The expected power density of the heaters on the vessel outer skins is 100 W m^{-2} This leads to a total contribution of approximately 650 W per end-cap expected from the heaters, which is included in the CO_2 cooling plant budget summarized in Table 11.2. The standard Kapton heaters technology is delivering

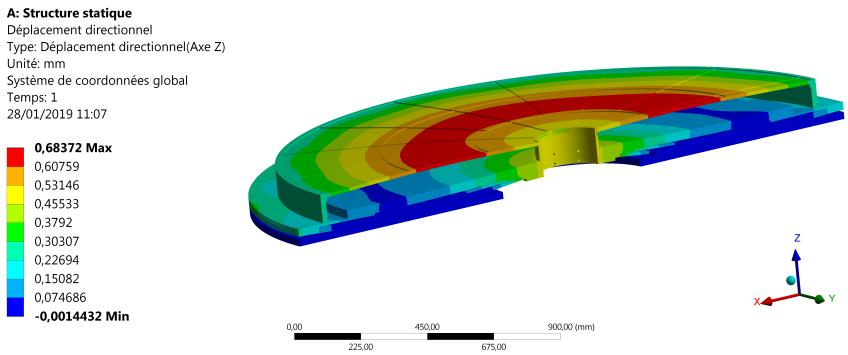


Figure 11.11: Finite Element Analysis (FEA) of the hermetic vessel with an over-pressure of 5 mbar. The red area corresponds to a maximum deflection of 1.5 mm in the central region of the front cover. In this analysis, the front cover model is computed as a single part without PEEK junction edges.

usual power amount of 500 W m^{-2}, as confirmed with liquid argon cryostats units. This selected technology is giving a comfortable safety factor of five compared to our expected needs. The temperature distribution expected on the hermetic vessel parts is shown in Figure 11.12. In these temperature calculations, which were performed using Finite Element Analysis (FEA) method, the ambient temperature of 20 °C and heat exchange coefficient of 5 W m^{-2} K^{-1} were taken as input parameters. Inside the hermetic vessel, the instrumented layers have been represented as a uniform material conductivity of 35 W/m·K with cooling channels at -35°C. The moderator conductivity has been set to 0.23 W/m·K, and the CFRP honeycombs 0.04W/m·K. Due to its tiny thickness, similar to double pane glass windows, the dry nitrogen has been represented as a conductive media with 0.04 W/m·K. In fact, the convective model was much less conservative as heat transfer phenomena. A temperature distribution in the range of 14 to 17 °C has been confirmed by this FEA output plot.

11.5.3 Back cover

Similar to the front cover, the back cover is also designed as a sandwich structure, consisting of a Nomex honeycomb core placed between two thin Carbon Fibre Reinforced Plastics (CFRP). Due to its tiny thickness and to procure additional stiffness, the back cover is used to hold the internal moderator using several tight bolting spots. This assembly technique increases the thickness from 7mm up to 37mm. The bolting connection between the CFRP back cover and the moderator is only blocked in Z and sliding in R-ϕ using large spring washers. In order to prevent any condensation in the thin air gap with the liquid argon cryostat wall, kapton heaters will be installed on the back cover with their temperature gauges and cables. Simultaneously, the existing cryostat heaters will continue to operate in their normal mode. The temperature distribution results of Figure 11.12 are taking into account the back cover heaters while the cryostat wall is not included in the FEA model.

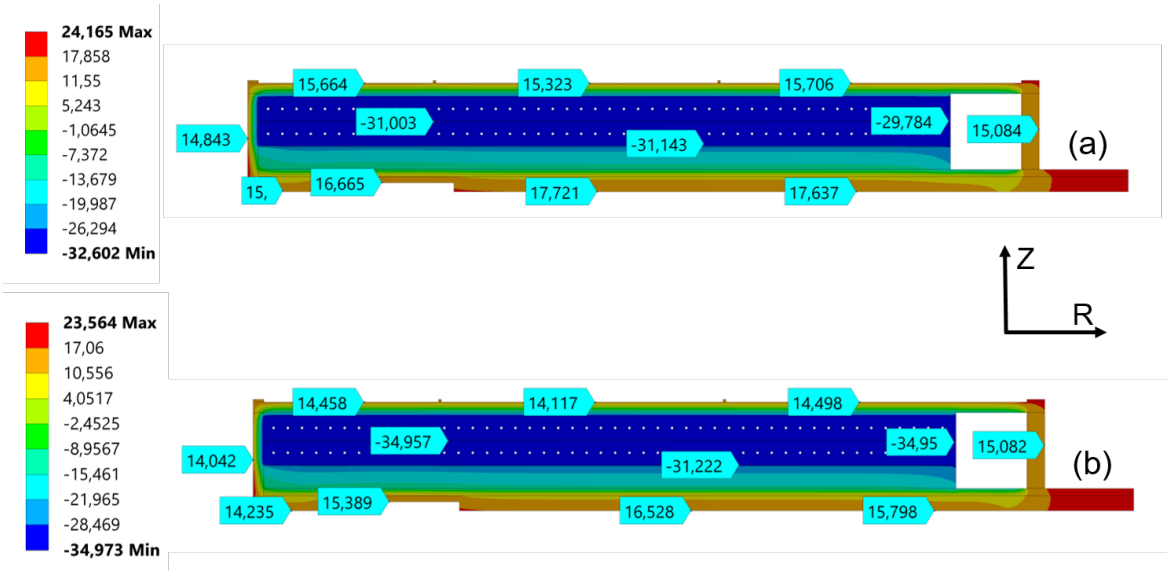


Figure 11.12: Temperature distribution within the FEA axisymmetric model of the detector assembly in the ATLAS experiment, with anti-condensation heaters powered on (front cover 80 W m^{-2}, back cover 85 W m^{-2}, outer ring 230 W m^{-2}, inner ring 60 W m^{-2}, total per end-cap 650 W). The dew point is set to 14°C and the temperature distribution is plotted with detector units turned on at 0.35 W cm^{-2} (a) and turned off (b).

The alignment of the hermetic vessel on the calorimeter end-cap will be based with respect to the axis of the cryostat warm tube, in the ATLAS coordinates system as illustrated in Figure 11.13. This survey reference should take into account the existing cylindrical moderator which is not represented in Figure 11.13. To optimise cavern access during the hermetic vessel installation/removal, the proposed design is to have the bolting/unbolting of the back cover to the cryostat wall throughout the external moderator. This procedure makes the installation and removal of the hermetic vessel easier, in particular without any required opening of the front cover.

11.5.4 Inner ring design

The inner ring of the hermetic vessel borders the beam pipe, resulting in a high level of radiation and heat exposure. Design efforts are ongoing to select the best material with high radiation resistance and low thermal conductivity to provide a shielding barrier during the beam pipe bake-out. Earlier projects with a similar environment, such as the ATLAS IBL and the LHC beam-pipe, have demonstrated good performance for carbon fibre structures and the aerogel insulating layers.

The inner ring of the current design, represented in the technical drawing of Figure 11.14, is composed of a sandwich structure consisting of eight millimetres of honeycombs and

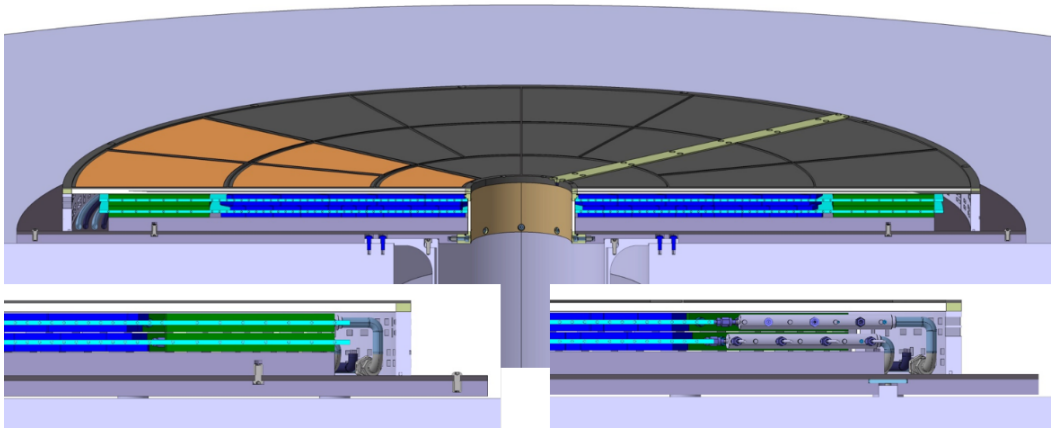


Figure 11.13: 3D cut view of the two instrumented layers inside the closed hermetic vessel. The main assembly parts are shown in their operating run configuration. The 2D sections on the bottom right and left are detailing the cooling manifolds area and the peripheral electronics boards respectively.

aerogel core enclosed between two thin sleeves made of CFRP high module panels. Further studies on high performance materials, such as Kevlar panels and honeycombs, are being undertaken to address the required stiffness, thermal protection, and radiation resistance, challenged by the low space allocated close to the beam pipe vacuum components.

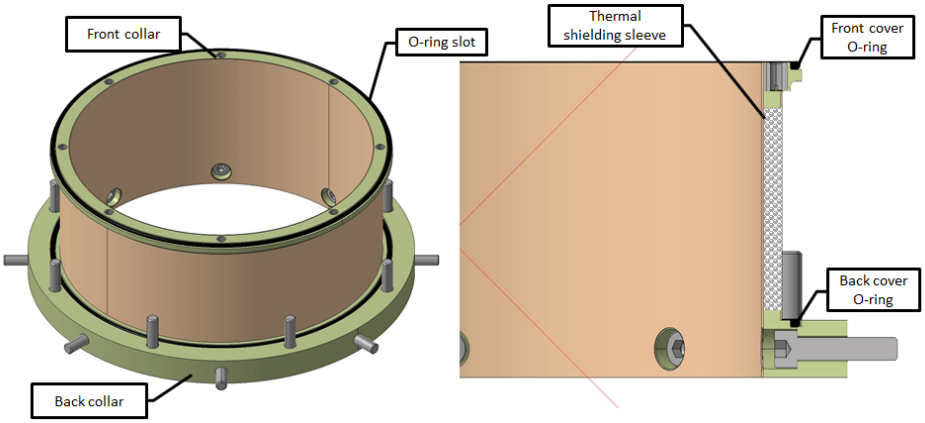


Figure 11.14: Central inner ring with its front and back collars. It is the central structure of the hermetic vessel, which ensures stiffness and tightness, thermal shielding, and HGTD positioning on the LAr cryostat.

To provide tightness as well as the alignment of the vessel with respect to ATLAS coordinate system, precisely-machined collars made of low thermal conductivity material, such as high performance PEEK polymer or Torlon polyamide-imide technology , will be installed on both extremities of the inner ring. Appropriate threaded inserts will be incorporated into the two assembly collars to allow tightening of the bolts of the front cover. The machined

slots will hold the sealing O-ring made out of PUR or EPDM material. The back collar will be bolted to the central flange of the external moderator, providing the hermetic vessel alignment with respect to the central tube of the LAr cryostat.

11.5.5 Outer ring design

All available routes for off detector services between the detector volume (dry and cold environment) and the outside world, are implemented on the outer ring. These passages are routing conductor cables, optical fibres, CO_2 cooling lines, and nitrogen tubes. The outer ring structure, which is an assembly of several parts, must be made of a stiff material with low thermal conductivity. As for the inner ring collars, the main candidate materials are the high performance PEEK polymer and Torlon polyamide-imide technology . Taking into account the large diameter of this part (up to 2000 mm), the manufacturing process is still under study to meet the specifications within a reasonable cost.

The interfaces to all HGTD services are implemented on the outer ring. It includes the cable and optical fibre connectors and the fittings for the CO_2 transfer lines and N_2 gas pipes. Such a design, shown in Figure 11.15, will allow a complete assembly and test of the detector at the surface integration area. The detector can then be transported to the pit for installation inside the closed vessel. All services connections will be done after fixing the HGTD on the calorimeter end-cap front wall. In the long shut-downs the closed HGTD vessel will be transported onto the surface for maintenance and replacement of the middle or inner rings, as described in Section 13.2.2. The final design and selection of materials is ongoing, it must satisfy the requirements regarding the tightness of the vessel, thermal isolation, radiation hardness, grounding and Faraday cage completeness. The CO_2 transfer lines will pass

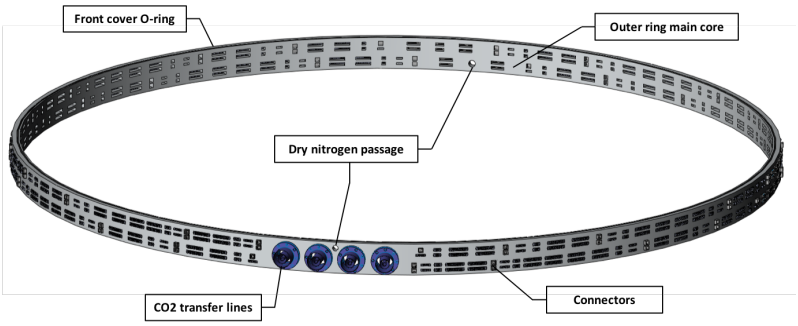


Figure 11.15: The outer ring assembly. The largest part of the hermetic vessel, with 2 m in diameter, it contains the service feedthroughs for cables, CO_2 transfer lines, and dry N_2 pipes.

through the cold-warm interface of the outer ring using standard conical sealing made of PUR or EPDM (Ethylene-Propylene-Diene Monomer), currently used in vacuum technology. The design of these cooling lines will be developed in a common program with CMS Phase-II HGCAL, which will transport a similar amount of power dissipation (4.7 kW for CMS and

5.0 kW for HGTD per line) under similar cooling specifications. In general, it is planned to work closely with the present program for both ATLAS and CMS trackers to develop and implement common solutions, such as appropriate improvements which can be made to the feed-through design and potting techniques.

11.6 Support and cooling disks

The design of the support and cooling disks features four half disks per end-cap to provide the cooling and support on both sides for the detector units and peripheral electronics boards. Cooling piping with a semi-circular concentric pattern is embedded into sandwich structure of rigid supports to extract heat dissipation produced in the modules and peripheral electronics, as described in Section 11.3.1 and Section 11.3.2. The ϕ position of the two fully instrumented layers inside the hermetic vessel is defined according to the technical drawing Figure D.8 and their tilt in relation to each-other is in the range of 15° to 20° (the exact tilt angle is still being optimised). This angle optimizes the overlap of modules while taking into account the needed space for the Peripheral Electronic Boards (PEB), connectors and flex stack up as well as cooling manifolds access space. Specific piping components are under investigation to reduce the dimensions of the cooling manifolds, fittings and capillary lines.

11.6.1 Requirements

As described in the hermetic vessel requirements, all selected materials shall withstand radiation hardness, fire instructions and OTR lifetime cycles (100 cycles from -45°C up to +40°C), in addition to specific mechanical and thermal behaviours. In order to prevent predicted thermal runaway, the heat transfer impedance from the ASICs to the coolant should be as high as possible to satisfy Section 7.5 thermal runaway criteria.

The final assembly of support and cooling disks, including adhesives and bolting design, should comply the Coefficient of Thermal Expansion mismatches (CTE) over the temperature range specified above. The bending over the two meters diameter half disks is a critical parameter and should not be amplified due to the bimetal switch effects. Taking into account the detector tight space, available in the z direction, the maximum acceptable deformation of the on-detector support and cooling half disk should not exceed $\pm$0.5 mm. In the R-ϕ directions, the instrumented half disks are less constrained due to the assembly isostatic boundary conditions (bolted on the inner ring and sliding at the large radius locked brackets). However, the expected thermal expansion vs shrinkage during OTR lifetime cycles should not produce additional bending or buckling effects. They are estimated to be in the range of +0.5 mm vs -1.5 mm in R and +0.75 mm vs -2.25 mm in ϕ with aluminum machined plates.

The instrumented half disks integration inside the hermetic vessel requires full access to manifolds area in order to achieve high pressure tight fittings (165 bar of pressure test). These connections to peripheral transfer lines should not induce additional thermo-mechanical constraints and deformation to the instrumented half disks, in particular during cooling down and warm up cycles. In the other hand, the frequency modes of the instrumented half disks should be safely shifted from vibrations generated by the cooling system. All manifold fittings should also be safely locked to prevent any release due to vibrations and OTR detector lifetime cycles.

11.6.2 Geometry and design

The baseline design is a half disk composed of two aluminum plates screwed face to face with the titanium cooling loop inserted in between, using thermal grease and an appropriate pressure torque to provide the required heat conductivity from the coolant to the heat source (optimisation of the Thermal Figure of Merit TFM). The feasibility of such a large assembly frame is challenging if the serpentine geometrical accuracy is not matching the machining grooves in the aluminum plates.

In addition, the different thermal expansion properties of the titanium and aluminum (CTE mismatch) need to be considered. Preliminary FEA studies of a full assembly of one cooling half disk are ongoing to evaluate the thermo-mechanical deformation and stress range, in particular the bi-metal switch effects. If the maximal deformation, in z direction, is over the expected values of ± 0.5 mm, the half disk Aluminum structure might not be manufactured in a single massive piece.

Sandwich structures with two high stiffness carbon fibre panels (CFRP), and a thermally conductive foam core including embedded cooling loops are considered as good alternative solution to the Aluminum single plates, even if they are less optimal for the thermal run-away study (see Section 7.5). A high performance candidate for the foam is a composite pyrolytic graphite foam similar to the selected material planned for ITk. It has high thermal conductive characteristics and absorbs the mismatched thermal expansion of the embedded cooling pipes and carbon-fibre panels. A thermally conductive reinforced elastomer is also under study as material core, due to its bonding characteristics, thermal performance, and reasonable cost.

The X-Y high thermal conductivity of carbon fibre panels is giving uniform temperature distribution over the large cooling disks. The CFRP drawback is driven by the Z low thermal conductivity that is increasing the thermal runaway hazards. The surface finishing of all borders of this alternative solution will be sealed by pultruded carbon fibre U-shaped crowns, which will be the direct interface with the HGTD hermetic vessel. High performance PEEK polymer and Torlon polyamide-imide technology are considered as good candidates to seal these boarders.

11.6.3 Assembly criteria

In order to perform the long term stability and accurate alignment in the ATLAS coordinate system, the instrumented disks will be directly connected to the inner ring at one extremity and peripheral edges of inner moderator at the other extremity. Integration tooling is under study to allow possible half disk insertion into the hermetic vessel both on vertical and horizontal positions. Taking into account the tight access to manifolds connections and the possible half disk maintenance disassembly, high performance fitting are proposed as baseline design. Due to these access and maintenance reasons, the Welding solution pointed out several integration difficulties and was not considered.

In order to accommodate the thermo-mechanical expansion vs shrinkage during OTR lifetime cycles (see Section 11.6.1), 2 mm free gap is foreseen in between the assembled half disks to absorb the expected 1.5 mm expansion and prevent any buckling effect. Each half disk is designed to have a locked ϕ slot in the middle of its large radius circumference to equally balance thermo-mechanical deformation in its peripheral directions. The inner and outer mechanical connections of each half disk should also carry grounding continuity of the instrumented layers up to the hermetic vessel ground.

12 Detector Infrastructure

12.1 Specifications

The HGTD services (cables, fibres, pipes) can be grouped in several categories depending on their role: optical fibres for data transmission; bias voltage for the sensors (high voltage-HV); power for the electronics (low voltage-LV); DCS control, temperature sensors, heaters; dry gas flushing; and CO_2 cooling. The milestones and review process are listed in Table 15.8. The HGTD services summary schedule can be seen in Figure 15.7.

HGTD Services	**Number**	**Diameter (mm)**	**Routing**
Optical bundles	40	9.5	HGTD - USA15
HV proximity cables	160	16	HGTD - (PP-EC)
DC-DC power control	40	14	HGTD - USA15
Interlock temp. sensors cables	32	16	HGTD - USA15
Sensors cables	10	12	HGTD - UX15
10 V power cables	72	15	HGTD - (PP-EC)
N_2 gas pipes	2	15 and 18	HGTD - USA15
CO_2 cooling lines	4	$<$50	HGTD - (PP-EC)
Total in barrel-end-cap gap	**356**		
HV cables	170	15.3	(PP-EC) - USA15
300 V LV	10	14.4	(PP-EC) - USA15
300 V LV control	10	12	(PP-EC) - USA15
DCS cables	16	14	UX15 - USA15

Table 12.1: Summary of HGTD services required for each end-cap, including spares. In the upper part of the table are listed the cables, fibre bundles and pipes, which start on HGTD vessel. Some of them are routed directly to racks located in USA15 or UX15. Others go to PP-EC area on calorimeter end-caps. From the PP-EC the other group of cables are routed to service caverns, they are shown in bottom part of the table. The local to service caverns cables routed between racks are are not included in the table.

The services will include patch panels (PP-EC), which will be located on the calorimeter end-cap in several accessible places, close to the New Small Wheel ($z \approx 6$ m). The main purpose of the PP-EC is to provide a disconnection point for those services, which cannot be accommodated in flexible chains due to lack of space and must be disconnected before

ATLAS opening. The PP-EC will also allow to realise mapping between connectors on back end electronics and on the detector. More details on PP-EC are given below in Section 12.2 and in Section 12.3.

An estimate of the required services per end-cap is summarised in Table 12.1 and is discussed in detail below. The table does not include the pigtails, which serve for interconnection between cables and peripheral electronics boards inside the vessel.

- The number of optical links per end-cap is 1464, including 520 up-links for data readout, 520 down-links for electronics configuration and fast commands (clocks, trigger, etc), and 424 up-links for luminosity readout. Multi-mode optical fibres will be used for data transmission. They will be grouped in bundles containing 48 fibres connected to 2 MTP connectors, 24 fibres per connector. The fibres will be encapsulated in a common sheath with reinforcement filler in order to be safely routed on cable trays and in the flexible chains. The number of fibres per bundle and per connector is optimised taking into account the routing of the fibres inside the HGTD vessel and the space available in flexible chains. Including spare fibres, a total of 40 bundles per end-cap are needed. Optical patch panels will be implemented in USA15 to organise the correct mapping for DAQ and luminosity readout.

- The baseline for the HV distribution is to provide individually adjustable voltage for each HGTD module. Consequently, 4016 HV lines are needed per end-cap. They will be grouped into 168 cables with an outer diameter of about 16 mm. Adding 2 spare cables, it gives a total of 170 HV cables per end-cap, to be installed between the HV power supplies located in USA15 and the HGTD PP-EC. On PP-EC the HV lines will be re-mapped into 160 cables, containing a different number of wires, to match the connectors on the peripheral electronic boards.

- Powering is organised in three stages. The bulk power supplies located in service caverns provide 300 V DC current to the DC-DC converters that will be placed in the PP-EC area. These second-stage multi-channel DC-DC units convert 300 V to 10 V to supply the radiation hard DC-DC converters that will be located on the peripheral electronics boards inside the vessel. The last converters power the on-detector chips and peripheral electronics, providing 1.2 V DC power and also 2.5 V for the optical links. The 10 V voltage can be adjusted to take into account the voltage drop on the cables. With such a layout the following cables are needed per end-cap: 4 cables to deliver 300 V DC power, 4 cables for control and monitoring, 4 cables for interlock and 4 cables for monitoring the DC-DC converters on PP-EC, all of them to be routed between service caverns and the PP-EC area. In addition, 72 proximity cables are needed to connect the DC-DC (300 V to 10 V) units located on the PP-EC area to the peripheral electronics boards, inside the vessel.

- The DCS requirements and related components are described in Chapter 8 and Chapter 9. The DCS services include the following cables per end-cap:
  - Control and monitoring for peripheral electronics, 40 cables.
  - Readout of temperature sensors on cooling loops, pressure sensors, mechanical interlocks etc., 10 cables.
  - Interlock temperature sensors on detectors, 32 cables.

  The readout of sensors will be organized using ELMB2, part of which will be located in the experimental cavern, the rest, which provide the information from Interlock temperature sensors to DCS, will be placed in Interlock Matrix Crates in the USA15 cavern.
- The heaters, similarly to the ones currently installed on the LAr cryostat flange, will be installed on the HGTD vessel front cover, between the external moderator and the LAr cryostat and in the proximity of the connectors on outer ring. Several power and temperature sensor cables will be needed for the HGTD heaters.
- The HGTD hermetic vessel will be flushed with dry nitrogen to prevent condensation on the detector components. For gas circulation one inlet pipe and one outlet pipe, with an inside diameter of 16 mm and 13 mm respectively, will be installed to each vessel.
- Four CO_2 cooling pipes $<$50 mm in diameter will be routed from the vessel feed-throughs to the cooling box located in the PP-EC area. The routing of the transfer lines between cooling box and CO_2 cooling plant located in USA15 is discussed in the next section.

12.2 Services layout

The overall HGTD service layout is illustrated in Figure 12.1.

As described above, the detector vessel will be fixed on the calorimeter end-caps, which move when ATLAS is opened. In the present ATLAS detector all end-cap services are installed in flexible chains to avoid their disconnection before movement. Currently all the chains are fully occupied, but it is expected that they will be partly rearranged for the ATLAS Phase-II upgrade and some space will be made available for a fraction of the HGTD services. Also two new small flexible chains are considered to be installed for HGTD. The priority for installation in flexible chains will be given to those cables and pipes, which are most critical concerning disconnection. The other services should be disconnected before the calorimeter end-caps are moved. For that purpose the patch panels (PP-EC) will be organised on the calorimeter surface in accessible places. The 300 V to 10 V DC-DC converters will also be

installed in the PP-EC area in order to make LV cables as short as possible, which is necessary to minimise the power losses (and voltage drops) on cables.

The baseline layout of the CO_2 transfer lines provides for the rigid lines installed between the CO_2 cooling plant located in USA15 and the manifold box placed on the calorimeter end-cap in the PP-EC area. Four smaller transfer lines, one for each instrumented half-disk, are routed on the calorimeter end-cap surface from the manifold box towards the HGTD vessel. With such a layout, the transfer lines must be disconnected from the manifold box before moving the calorimeter end-cap. A more attractive approach consists in avoiding the disconnection of the transfer lines at least at standard openings in YETS. To realise this, the CO_2 transfer lines must include rigid and flexible parts. The rigid lines are installed between the CO_2 cooling plant and the manifold box, located on the voussoir platforms in the ATLAS toroid area above the calorimeter end-caps. From the manifold box, two flexible lines, one inlet and one outlet, are routed to the splitter box on the top of the calorimeter end-cap on the IP side, close to HGTD. From the splitter box four smaller rigid lines are installed on the calorimeter front wall and connected to the HGTD vessel. The use of flexible lines avoids the disconnection of CO_2 cooling lines during standard openings. However, on the platforms, there is not enough room to accommodate flexible lines, long enough for full openings in LS periods, when calorimeter end-caps are moved by about 12 meters. For such an openings, the flexible lines must be disconnected from the splitter box on the calorimeter. A more in-depth study is necessary to confirm the feasibility of implementing the layout with flexible lines.

To allow commissioning of the detector after installation in the experimental cavern, and for maintenance during shutdown periods, it should be possible to operate the HGTD when ATLAS is in the open configuration, which requires reconnecting the services in the open position. For that purpose, it is envisaged to install extenders of cables and CO_2 cooling lines between respective positions of the patch panels in closed and open configurations. Most of these extensions should be permanently held in place, which will help minimise the time required to put the HGTD in working order after each opening.

12.3 Patch panels in PP-EC area

The positions of the PP-EC boxes and DC-DC units on the calorimeter end-caps will be chosen in discussion with Technical Coordination, and placed in several sectors in accessible areas to allow disconnection of services. It will also be possible to replace any faulty DC-DC converter with a short access during the run. The preliminary study of the patch panel locations by ATLAS Technical Coordination is shown in Figure 12.2.

The strength of the magnetic field, along with radiation levels, are critical parameters for the design of the DC-DC power converters. The magnetic field in the patch panel region is shown in Figure 12.3, varying from 0.05 T up to 0.5 T. The power supplies should be placed

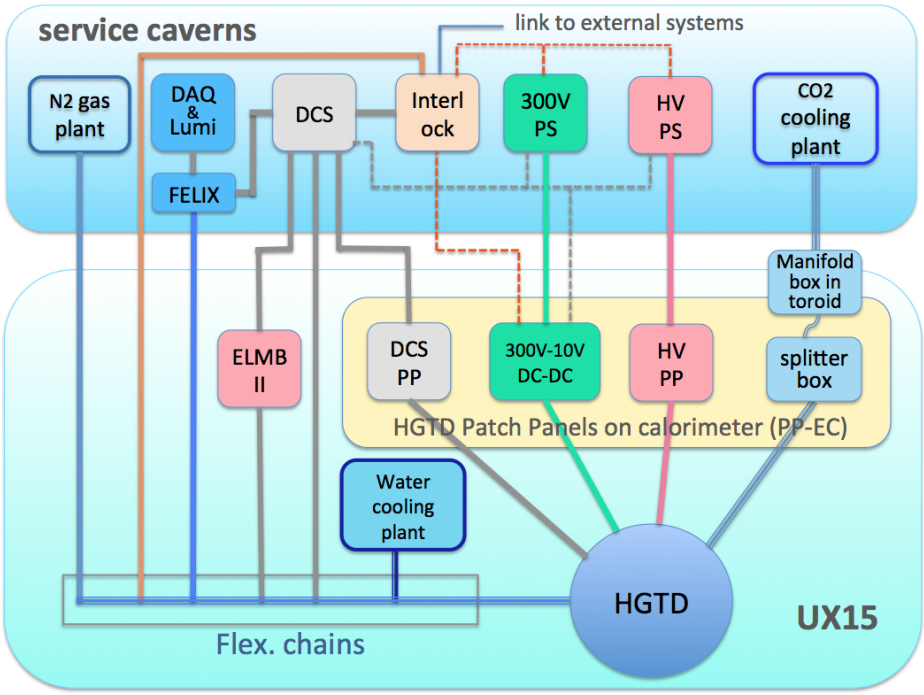


Figure 12.1: Overall HGTD services layout from the detector to USA15 or UX15. The optical fibre bundles, N_2 gas pipes, interlock and cooling temperature sensor cables, part of DCS cables and, still to be confirmed, the 300 V power supply cables are planned to be installed in flexible chains. The HV cables and rest of DCS cables will be routed through the patch panels, where they will have a disconnection point.

in areas where the field is the weakest, midway between two barrel toroids and as close as possible to the surface of the calorimeter. Radiation levels in these areas have been estimated using FLUKA calculations, giving a maximum of 15 Gy and less than $1 \times 10^{12}\,n_{eq}\,cm^{-2}$ (no safety factors applied) at the outer radius of the calorimeter end-cap, where the patch panel boxes will be located.

The DC-DC power converters located in the PP-EC area will require water cooling. Assuming 80% power efficiency, about 4 kW of cooling power is needed in all PP-EC locations, combined for each end-cap. The existing ATLAS leak-less water cooling systems have a sufficient capacity to supply the HGTD detectors on both end-caps. Dedicated connecting pipes and manifolds on the calorimeter will be required.

12.4 Services routing on the calorimeter front wall

The space available to route the HGTD services in the gap between the calorimeter barrel and end-cap is very limited, making the design and installation of the services a challenging

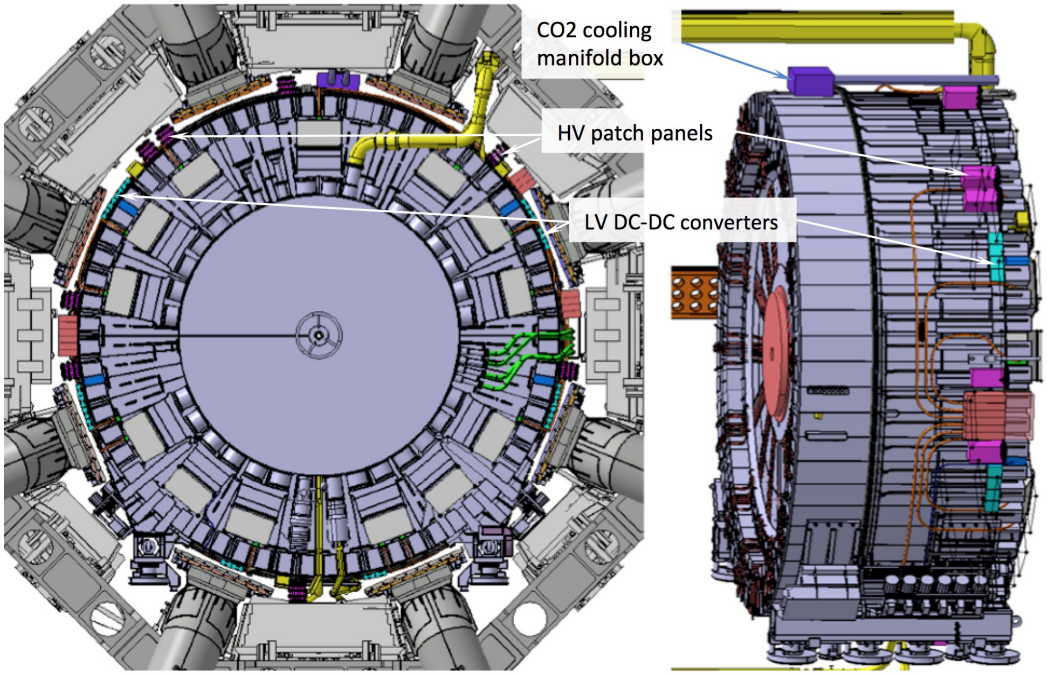


Figure 12.2: The preliminary layout of the HGTD patch panels (PP-EC) on the surface of the end-cap calorimeters. The 300 V to 10 V DC-DC converters and the cooling splitter box will also be located in this region. The PP-EC components are distributed in several places around the calorimeter end-cap surface (some of them are indicated by arrows in the image).

task. This space is shared between ITk and HGTD services. In addition, scintillator counters, belonging to the Tile calorimeter system, are installed there. In the present configuration, the counters are fixed on the Tilecal and LAr front face, where the HGTD cables will be routed. In LS3 the scintillators must be replaced by new ones. It was agreed with the Tilecal system and Technical Coordination that the scintillator counters will be installed on top of the HGTD services, which will be fixed on the wall of the calorimeter. Such a layout will guarantee access to the counters and their replacement during HL-LHC lifetime. In order to provide more robust support and fixations for HGTD cables and for scintillator counters and, at the same time, to protect the Tile calorimeter, whose scintillator tiles and fibres are visible on its front side, thin aluminium support plates will be fixed on the Tile calorimeter modules.

The envelop for HGTD services is shown in Figure 12.4. All space in ϕ on the front wall of the LAr end-cap cryostat is available for HGTD services, while at bigger radius they have to be grouped to fit in the space between LAr barrel crates and further between Tilecal barrel fingers, sharing the space with ITk services installed on the calorimeter barrel. The room in two gaps between LAr barrel crates on top cannot be used to route the HGTD cables. One constraint comes from the requirement to keep free access to the end plates of three Tilecal modules, located at the top of the calorimeter, to give access to the electronics of

these modules. Space in another gap is occupied by a LAr HV filter box. The space in z available for HGTD services on the LAr end-cap cryostat wall at radius >1.4 m is only 17 mm. Exceptionally, there will be a dedicated slot for four CO_2 cooling pipes, as described in Section 11.3.2.

The HGTD services routing on the calorimeter end-cap front wall is shown in Figure 12.5. The cables, connected to the outer ring of the HGTD vessel in four layers, will be rearranged to one layer at r > 1.4 m to fit within the envelope of 17 mm. Below the Tilecal barrel fingers, the cables will be regrouped to a few layers to come out on the calorimeter surface through the gaps between the fingers. As discussed above, the HGTD cables cannot be routed in two top gaps between the LAr barrel crates. Due to that the cables from the top section of the HGTD deviate towards neighbouring gaps. From the gaps, the cables will be routed over the surface of the calorimeter end-cap, to PP-EC located in several places around the calorimeter.

12.5 Services connection to outer ring and inside the vessel

The outer ring of the HGTD vessel provides the interface for all the services. With such an approach, the HGTD detector can be completely assembled and tested at the surface and brought down to the experimental cavern for installation as a closed vessel. Once the vessel is fixed to the front wall of the LAr cryostat, the pipes, cables and optical bundles will be connected to the detector. To realise such a scenario, the cooling and gas pipe fittings, electrical and optical connectors will be embedded in the outer ring, as shown in Figure 12.6. The layout of the outer ring is shown in Figure 11.15.

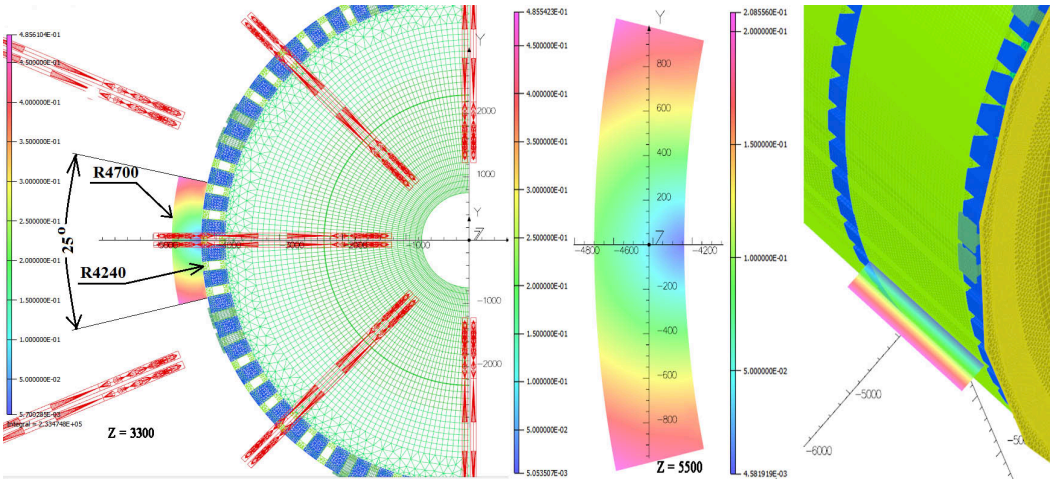


Figure 12.3: Magnetic field in the region of the HGTD PP-EC patch panels.

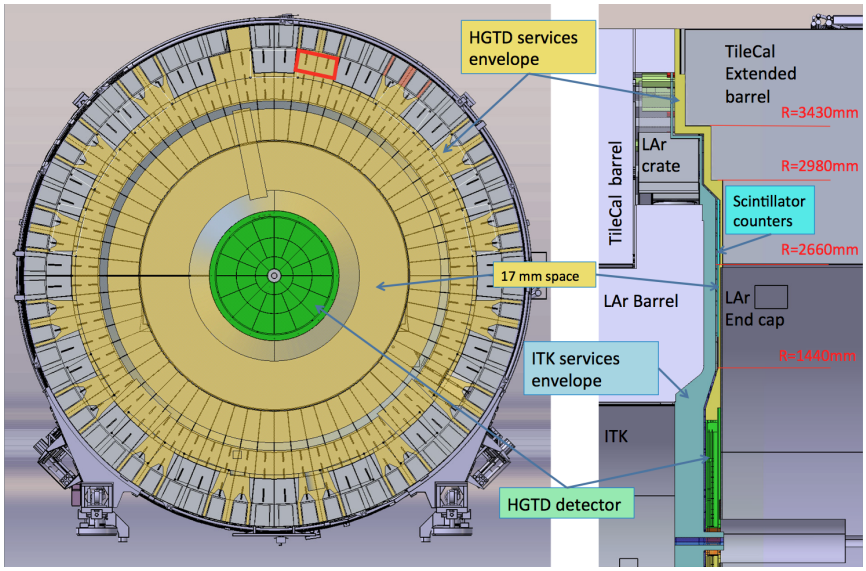


Figure 12.4: The envelope for HGTD services. On the left: front view of the calorimeter end-cap on side A. The space available for HGTD services is shown with yellow color. With red rectangles is shown the area, where the room the HGTD services is very limited. On the right: the HGTD services envelope in the gap between calorimeter barrel and end-cap. The envelopes for ITk services and the Tilecal scintillator counters are also shown.

The organisation of services inside the HGTD vessel is schematically shown in Figure 12.7. The short pigtails, one per cable, will interconnect the cables and the peripheral electronics boards (PEB). The optical bundles, connected to the outer ring, will be terminated with 24-fibre MPO connectors. The optical pigtails will be used to distribute these 24 fibres from each bundle to several VTRx+ optical link modules installed on the PEB. One bundle is required per PEB, including spare fibres. The optical pigtails will also contain spare fibres terminated by connectors.

12.6 Services installation

The installation of services and patch panels will be done in close collaboration with Technical Coordination. The delivery of CO_2, under-pressure water cooling stations and the N_2 gas plant is the responsibility of Technical Coordination and the CERN support cooling and gas teams.

The various components should be available at different times depending on the delivery and final location in the ATLAS cavern. To decouple the installation of cables, patch panels and the detector, the mock-ups of patch panels and outer ring indicting the positions of the

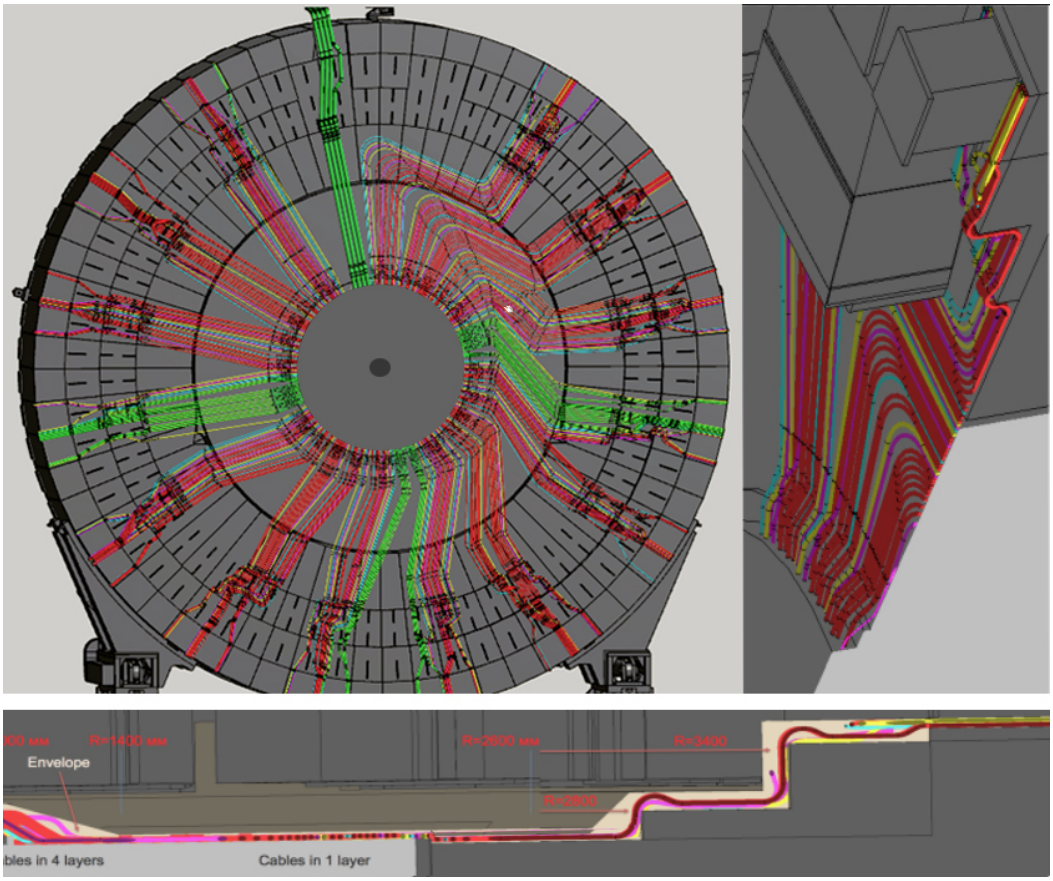


Figure 12.5: HGTD services routing on the calorimeter end-cap wall. The LAr crates and Tilecal fingers, both belonging to the calorimeter barrel, are also shown in the picture.

connectors will be installed in their final place with the aim to precisely indicate the cable connection points.

In an environment as complex as ATLAS, cable routing requires numerous turns and transitions between cable trays, which makes it impossible to estimate the lengths of cables with an accuracy of several centimetres at connection points. As a consequence, some extra length has to be allowed for each cable, which could then be accommodated on cable trays, however this is not always possible due to the lack of space.

Therefore, the common approach for installing long cables is to pull cables with the connectors attached only on the detector side, to allow adjustments to the cable length on the other side. The connector at the second end of the cable should be attached in situ, though that is not always feasible due to connector complexity or lack of space or time for this work. Given all this, different installation scenarios are foreseen for different HGTD services, as described below. The detailed plans and schedule for the installation of each type of HGTD services will be developed in collaboration with Technical Coordination as part of the preparation of

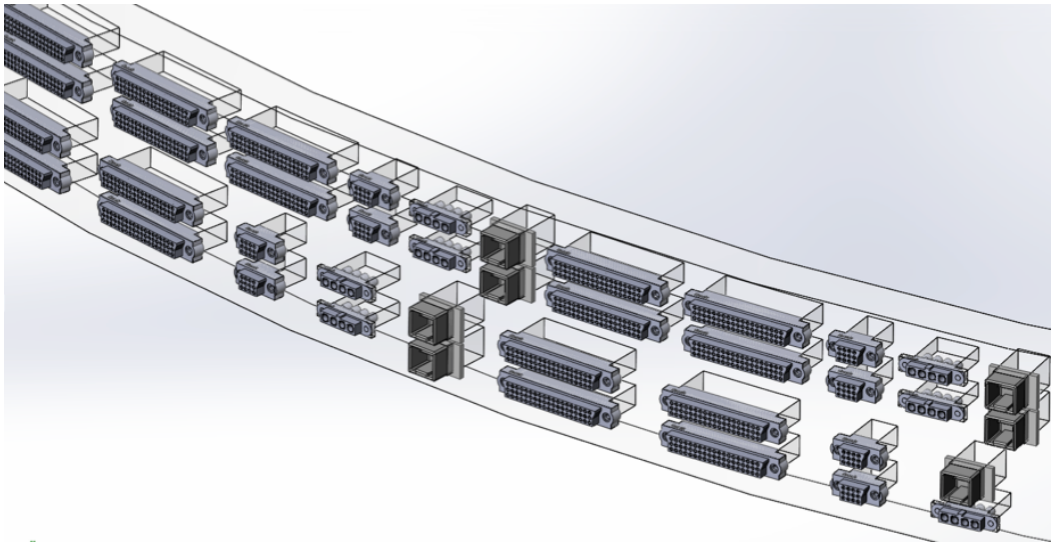


Figure 12.6: Fragment of the outer ring of the HGTD vessel. The electrical and optical connectors embedded into the ring are shown.

activities for LS3. All cables, except optical bundles, must be tested before installation in the cavern.

The **Optical bundles** will be delivered with connectors attached and protected at both ends. They will be tested at the factory. Some space should be reserved to accommodate an extra length on the cable trays below the racks in USA15. The optical bundles will be routed through small plastic flexible chains available in sector 11, to avoid disconnecting them at ATLAS openings. An optical patch panel will be used in USA15 to remap the fibres between luminosity and data readout.

For the **HV cables** two installation scenarios are considered. If space on the cable trays below the racks in USA15 is available to accommodate an extra length, the HV cables will be delivered with connectors fixed at both ends. Otherwise, the cables will be made in double length, folded in the middle, with connectors attached at both ends, to be routed to the PP-EC patch panel. Such a configuration makes it possible to test the cables and connectors before installation. After pulling such cable pairs into the service cavern, the loop will be cut out to the precise length and the missing connectors attached in-situ. One of these scenarios will be chosen when the layout of the racks and the services in the service caverns are available from Technical Coordination.

The **LV and DCS cables** to be routed between the experimental and service caverns will be installed with one connector (detector side), the second connector will be attached in-situ near racks. The same scenario will be applied for DCS cables between the HGTD or patch panels and racks in UX15.

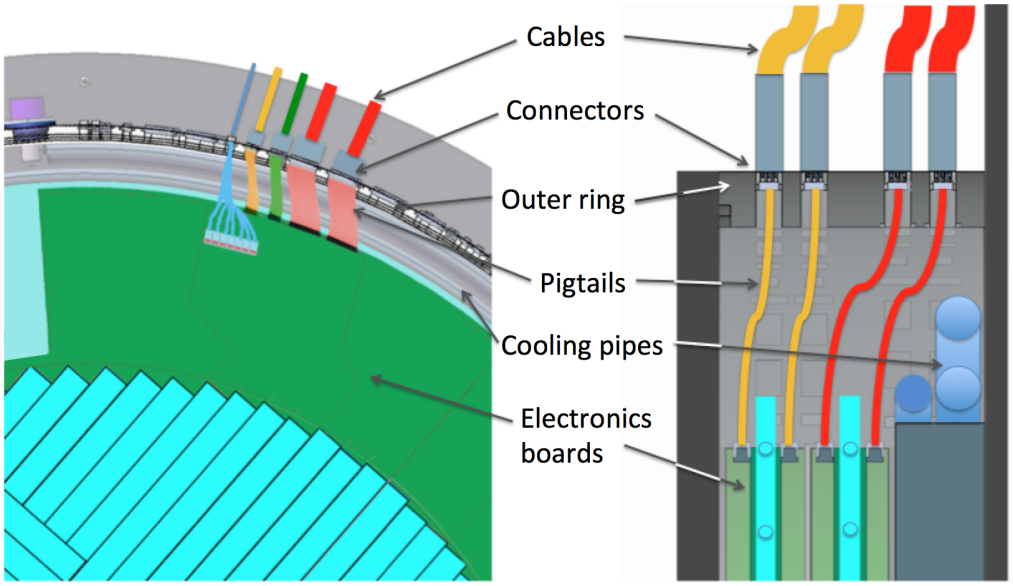


Figure 12.7: Illustration of the services layout inside the HGTD vessel. On the right part of the picture the cross section of the services area is shown.

The **Proximity cables** listed in Table 12.1 must be delivered with connectors attached at both ends, because it would be extremely difficult to install them in-situ, near the calorimeter. Before installing the connectors, the length of these cables must be precisely measured in-situ, by pulling pilot cables between the mock-up of outer ring and the patch panel box, placed in their final positions.

All the described above installation scenarios were successfully applied in ATLAS during its original installation.

The installation of the patch panels and services, and the respective connectivity will be done when access is allowed by Technical Coordination. These activities will start well before the HGTD installation and will be spread over time. In the current schedule, these activities are planned over approximately 16 months, from January 2025, the beginning of LS3, to April 2026.

Technical Coordination will have responsibility for the planning and installation of the transfer lines for the CO_2 cooling system and pipes for the N_2 gas system.

13 Detector Assembly, Installation, and Commissioning

13.1 Assembly and commissioning on surface

The detector assembly and QA will be done at CERN in a clean room, using dedicated tools and testbenches. These activities will include the assembly of the instrumented half disks by installing the detector components (detector units, PEB, flex tails) on the cooling plates, and the integration of the assembled half disks, sensors and services inside the vessels. Each assembly step will include all the procedures necessary for QA. Several Institutes will participate in the assembly activities, which are planned for the periods between September 2024 and October 2025 for HGTD-A and between October 2025 and October 2026 for HGTD-C. A schedule of the assembly can be seen in Figure 15.8 and a schedule for the installation can be seen in Figure 15.9.

13.1.1 Half disks instrumentation

In order to prevent any contamination of the active sensors (dust, metallic chips), all detector assemblies and testing must take place in a clean environment, equipped with temperature and humidity control gauges. The floor should be ESD protected (ElectroStatic Discharge) for personnel and components at all work-stations and setups. Specifications for this environment are under development considering that all critical assembly steps shall take place in a clean room class ISO-8 or better.

The detector units (see Section 7.4.2), corresponding to 16 inner half disks, 32 middle quarter disks and 32 outer quarter disks, will be assembled and qualified in different Institutes and shipped to CERN to be mounted on the cooling plates. Separately, the peripheral electronics boards and the flex tails will be qualified at collaborating Institutes and shipped as well to CERN. The QA procedure on the detector units, peripheral electronic boards and flex tails will be repeated at CERN, at least on a sample of elements, to confirm that no damage has occurred during shipping.

The half disks will be instrumented by mounting first the peripheral electronic boards, followed by the outer ring and the detector units. After mounting the outer ring and before

adding the detector units, the pigtails will be connected. The last step is to connect the flex tails between the modules and the peripheral electronic boards.

At this stage the full on-detector readout chain is connected for the first time and the first full calibration will be performed, as described in Section 13.1.3. After having completed all the tests and replaced the defective components if necessary, the instrumented half-disks are ready to be installed in the hermetic vessel.

Each instrumented half disk will be a single item weighing 30 kg with 12 cm inner radius and 92 cm outer radius. A breakdown of the contribution of each component to the thickness can be found in Table 11.1. Dedicated tools will be developed to allow the disks assembly in the optimal position (horizontal vs vertical) with appropriate rotation to fully instrument the two faces of the half disk.

13.1.2 Detector assembly on hermetic vessel

Before being delivered to CERN, the hermetic vessels will be mechanically tested at the Institutes responsible for their production. The mechanical tests will be repeated at CERN to exclude damage due to shipping and the air tightness of the vessel will be checked. After this step and the installation of the CO_2 cooling pipes, the instrumented half disks will be installed in the vessel and the CO_2 services will be connected. The next step will be to attach the temperature sensors to the manifolds and cooling pipes, install the interlock temperature sensors, the pressure sensors and connect the relevant services. After that, the hermetic vessel will be closed with a front cover. Prior to any integration step the mechanical envelopes of previously installed components must be validated and the spacing between each component must be controlled. Once the vessel is assembled, it can be connected to the baby demo CO_2 cooling system (see Section 14.2) to perform pressure and cooling tests, in addition to further performance tests, as described below.

13.1.3 Quality Assurance after assembly

The first set of tests after instrumentation of the half disks will probe the electrical connectivity, followed by a first full calibration of all the on-detector elements and a full-chain test, using the particle signals from a radioactive source. The test bench will include a cold box with a radioactive source movement system, an interlock system, a portable system for powering, read-out and control to be used on surface for the tests described in this section and in the cavern before the detector is connected to the ATLAS DCS and DAQ systems. The test bench will facilitate the testing of all the detector elements connected to one peripheral electronics boards (PEB) at once.

To test electrical connectivity, commands will be sent from the DAQ and DCS modules to each stage of the electronics, reading back the response of the commands. First the

communication with the PEB will be tested, then the communication with ASICs and some of their functionalities will be tested, which also probes the flex tails and module flexes. Finally, the whole chain will be tested sending a calibration signal to the sensors. To speed up the calibration procedure, the already known optimal settings for the modules and for the PEBs, obtained in previous steps of the production process, will be applied in tests with and without particles. Only readout channels, which show a change in characterization compared to the earlier calibration, will be re-calibrated.

A database will be used to record the status of each component at all assembly steps, in particular the electronic and the thermal parameters of the instrumented half disks. The aim is to have a full history from the production process up to the final assembly and testing. Existing ATLAS databases will be adapted to avoid duplication of the software development efforts. The database identification protocol of all mechanical components will be based on a serial number and/or QR code (bar-code if any). In addition, detailed technical parameters (row material, chemical composition, manufacturing process, testing) will be included in the database to allow monitoring of the construction progress. After completion of the detector installation in the experimental cavern, the database will evolve towards a collection of system configuration data, necessary to analyze the detector operation conditions and performance.

13.2 Installation in the cavern and commissioning

13.2.1 Access and maintenance scenarios

The access for installation and maintenance of the detector and the off-detector electronics located in UX15 can only occur in breaks of LHC operation, and the intervention actions depend on the duration, induced radiation levels and ATLAS opening scenarios. The back-end electronics situated in USA15 will be accessible at any time, but interventions will be limited during data taking. The access scenarios and possible interventions on the detector during the various types of breaks in the operation of the HL-LHC are described below.

Short access for a few hours only, primarily for LHC machine interventions and usually announced on short notice. In these periods, the electronic components located in the HGTD PP-EC areas can be accessed for simple interventions, for example to replace the 300 V-10 V DC-DC converter modules. Access to the DCS equipment in the racks in UX15 will also be possible.

Technical Stop, typically of one week duration, for maintenance of the LHC and of the experiments. The same areas as for the short access periods will be accessible, but it will be possible to perform more complex and long operations.

Year-End Technical Stop (YETS), the yearly maintenance for about 12 weeks. In this period the ATLAS detector is partly opened, keeping the beam pipe in place as illustrated in Figure 13.1. The distance between the calorimeter barrel and end-cap is typically 3.1 m. The access to the HGTD components inside the vessel would be very difficult due to the high radiation level and the complexity of the detector construction, so the opening of the vessel in situ during YETS is not planned. However, the construction of the vessel and instrumented discs allow to open the detector and remove or install instrumented half-disks. The maintenance or upgrade of all off-detector components, including patch panels, electronics and services is possible, depending on the duration of the foreseen operations and the radiation level in the accessed areas.

Long shutdown (LS), which typically lasts 2 years, is foreseen for large upgrade and consolidation programs for the experiments and the LHC. The ATLAS detector will be in the large opening position, with the beam pipe removed, as shown in Figure 13.2. The distance between the barrel calorimeter face and the HGTD face is at most 12 m. After the LS3, when the HGTD should be installed, the next long shutdowns should be used for extensive maintenance and upgrade of the detector. After each 1000 fb^{-1} of collected data, which approximately corresponds to the run period between two long shutdowns, the detector modules located in the innermost ring must be replaced. Every 2000 fb^{-1} the modules in the middle ring will be replaced. These operations, along with reparations and consolidations of the detector, will be performed on the surface integration area. Once the area has reached an acceptable radiation level, the services will be disconnected from the HGTD vessel, then the closed vessel will be removed from the calorimeter end-cap and brought to the surface. For the replacement of the detector modules and the tests and commissioning after consolidation, the same tooling and procedures as for the detector assembly will be used. In addition, due to the exposure of the HGTD to radiation during data taking, safety guidelines will be strictly enforced to protect the personnel when accessing and manipulating the HGTD components, following the radioprotection measures and procedures prescribed by the CERN Radio Protection experts.

13.2.2 Transport and installation the HGTD in the cavern

The installation of HGTD and the connection of all services will take one month for each end-cap and is planned, in accordance with the ATLAS TC schedule available in mid-January 2020, for April 2026 and January 2027 for the A side and C side, respectively. More details on the schedule are given in Section 15.2. The development and optimisation of the schedule of ATLAS upgrade activities in LS3 will continue for several years and could lead to advancing the installation of HGTD by a few months. In this case, the HGTD schedule will be adapted to the overall LS3 schedule. If necessary, the design of the HGTD allows the installation of instrumented disks in the next YETS, even in the presence of the beam pipe.

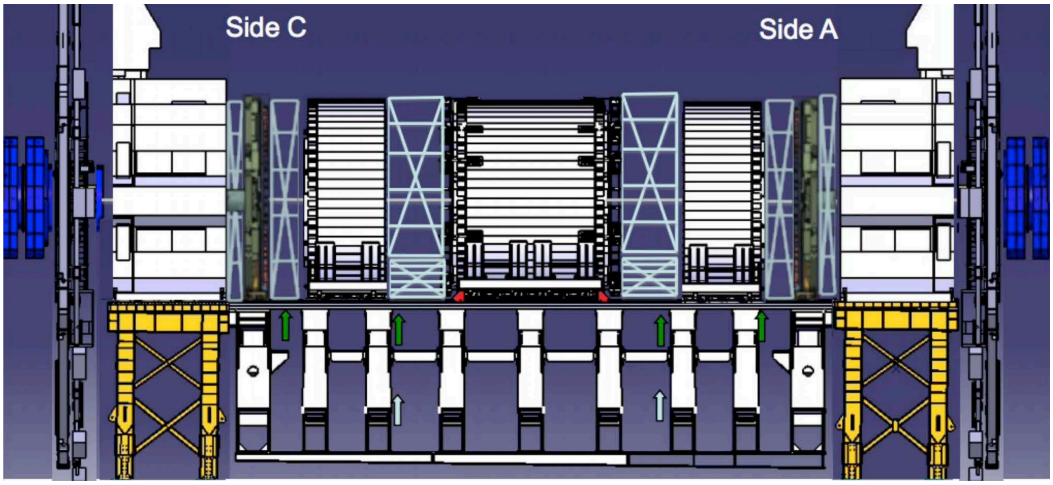


Figure 13.1: ATLAS in sttandard opening configuration.

The installation of the detector will be done in the large detector opening configuration as shown in Figure 13.2. This operation can start only after the external part of the moderator has been installed on the LAr front wall.

As it is described above, both HGTD end-caps will be fully assembled and tested on surface. The closed vessels will be transported to the pit for installation on the calorimeter end-caps, using dedicated installation tooling. This is the baseline assembly and installation scenario, however the staggered installation of different half disks in situ is also possible.

The total weight is 275 kg per end-cap, assuming that the external moderator part will be transported separately. The overall dimensions are 1100 mm radius and 105 mm thickness. These parameters should be taken into account for the transport truck and lowering, but they are well below the lifting capacity limit of the crane in ATLAS SX1 surface building and the dimensions of both shafts. Each end-cap, HGTD-A and HGTD-C, will be lowered on side A and side C, respectively, directly from the surface to the minivans, which are shown in Figure 13.3. A local lifting tool is needed to lift the fully assembled end-cap detector and accurately align it with respect to the LAr end-cap inner warm tube, to avoid any conflict with the beam pipe ionic pump and its services.

Specific tools will be constructed to perform the transport, lowering and installation of the HGTD on the calorimeter end-cap. All these tools are still at a conceptual stage and will need to be carefully designed, and, where possible, use synergies with tools already developed for other sub-detectors.

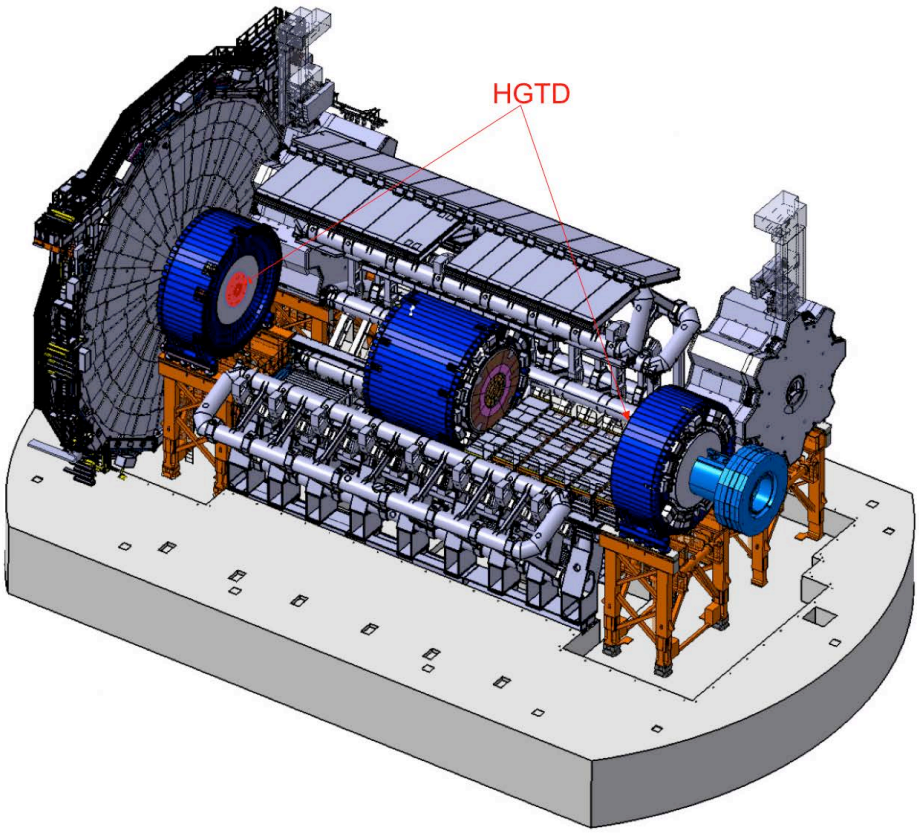


Figure 13.2: ATLAS in large opening configuration. The HGTD is superimposed to the MBTS scintillators that are presently installed on the calorimeter end-cap cryostat.

13.2.3 Services connection and commissioning

As described in Section 12.6, all HGTD services will be already in place before the installation of the detector. After fixing the HGTD vessels on the LAr end-caps, the cables and the pipes will be connected, and the connectivity will be tested as part of the initial commissioning. The extensive tests and validation of interlock, detector safety system and DCS must be completed prior to the next commissioning steps.

Access to the detector components during the commissioning should be possible until approximately April 2027, close to the expected end-cap calorimeter closure. This will leave at least 6 months of intense commissioning while access is still possible. Both the installation and the commissioning of the HGTD will be carried on with the participation of several collaborating Institutes.

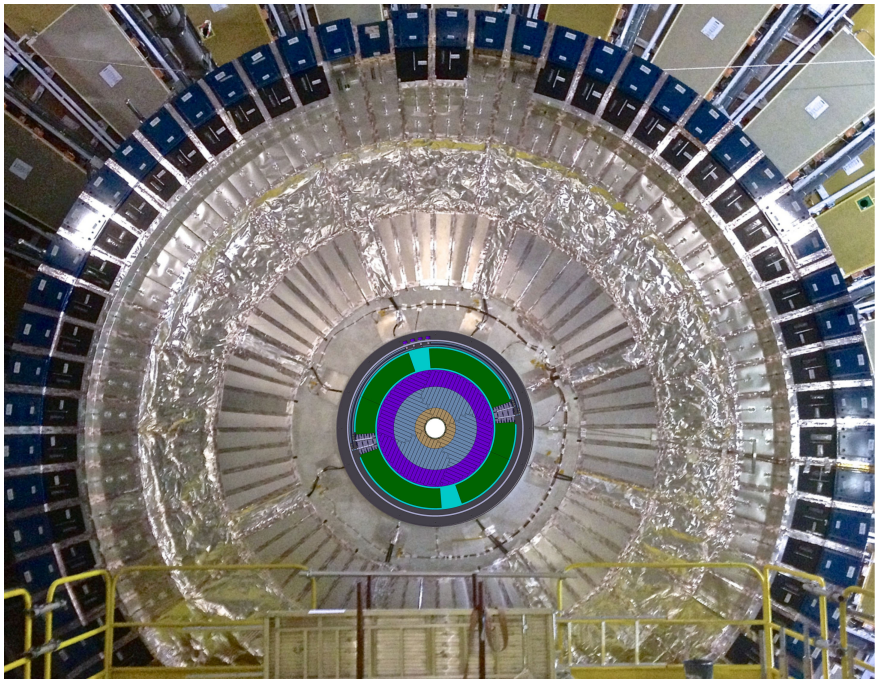


Figure 13.3: ATLAS in large opening configuration. HGTD detector superimposed on the MBTS scintillators, that are currently installed on the LAr end-cap cryostat.

13.2.4 Radiation environment, and radio protection

During all ATLAS upgrade and maintenance activities, as on the CERN site in general, the ALARA (As Low As Reasonably Achievable) radio-protection principles will be strictly followed. It will be implemented during the installation and maintenance activities of the HGTD, in accordance to the rules and recommendations of the CERN Radiation Protection service and in close collaboration with Technical Coordination.

In order to plan the HGTD installation and further consolidation activities and to optimise the work procedures accordingly to this concept, estimates of the radiation environment are needed. Estimates of the ambient dose equivalent rates in LS3, when the HGTD is installed, and in the following LS periods, when part of HGTD detector modules are replaced, were provided by the RP group, using FLUKA calculations, and can be found in Ref. [97], [98], [99], [100]. The uncertainty on these calculations, evaluated by comparing them with ambient dose equivalent rate measurements during YETS 2016-2017, includes a systematic underestimate up to a factor of 2 in the region between the ID and LAr end-cap. This uncertainty comes most likely from the imprecise material description in the simulation.

The dose equivalent rate map for LS3, after 28 days of cool-down time, is shown in Figure 13.4, for the geometry corresponding to the completed large opening, with all beam pipes and inner detector removed.

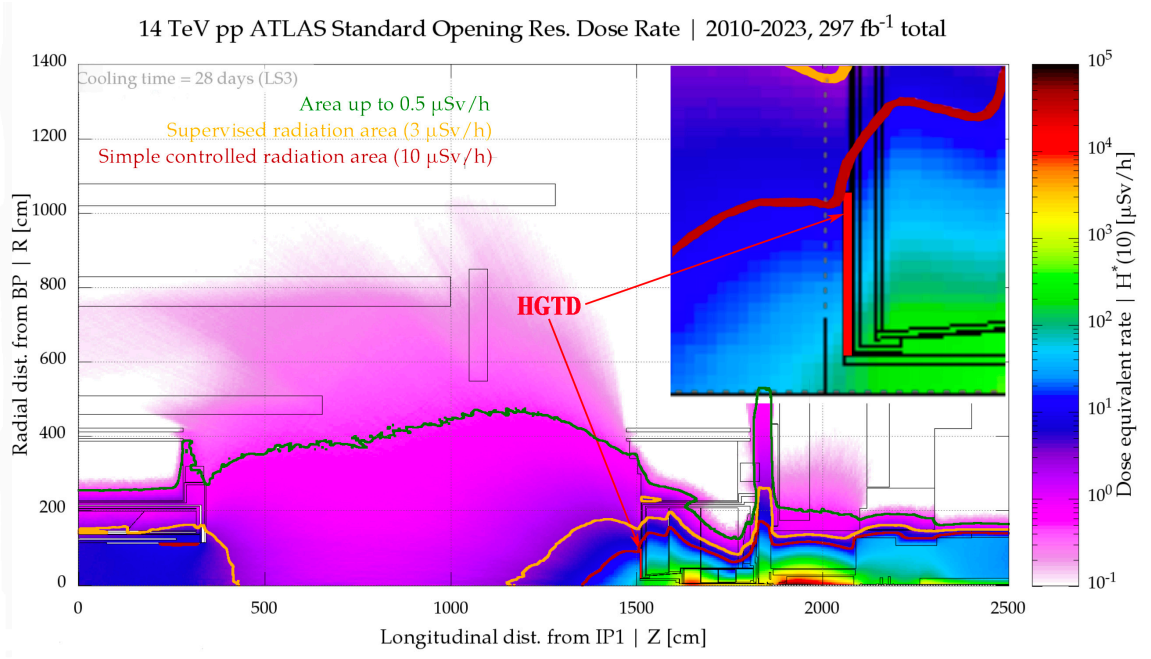


Figure 13.4: FLUKA simulations of the ambient dose equivalent rate in LS3, after 297 fb^{-1} of accumulated data and 28 days of cool-down period. ATLAS is in the large opening configuration, all beam pipes and inner detector are removed. The boundaries of various radiation areas are shown with coloured lines. This figure is a combination of figures in Ref. [99].

The HGTD installation will take place after about 1.5 years of cool-down time. In that period the radiation level will drop by a large factor compared to the one shown in Figure 13.4. On the other hand, the extension of Run 3 by one year might cause an increase of the equivalent dose rate. Calculations dedicated to each LS must be performed to estimate the radiation environment during the replacement of the inner and middle rings with reasonable accuracy. FLUKA simulations exist for the dose equivalent rate in the LS5 period, assuming 2177 fb^{-1} of accumulated data, a cool-down period of 181 days and the standard opening configuration. The results are shown in Figure 13.5. In this configuration the radiation levels expected in the HGTD region are expected to be in the range of 30 to 50 µSv h^{-1} (from the outer to the inner radius). When replacing the inner part of the detector, the expected dose rates should be lower due to the longer cool-down time and the absence of the beam pipes. Nevertheless, it will be well above the threshold defining the simple controlled area (10 µSv h^{-1}). Therefore the work duration will be severely limited.

Before accessing the components of the detector to be moved to the surface for replacement of the inner and middle ring, additional cool-down time will be necessary. In order to minimise the radioactivity of the detector, material less prone to activation must be used in the construction, in particular by avoiding the use of stainless steel components and giving preference to aluminium or plastic. First of all, the possibility of manufacturing the

aluminium or titanium pipes integrated in the cooling supports is considered.

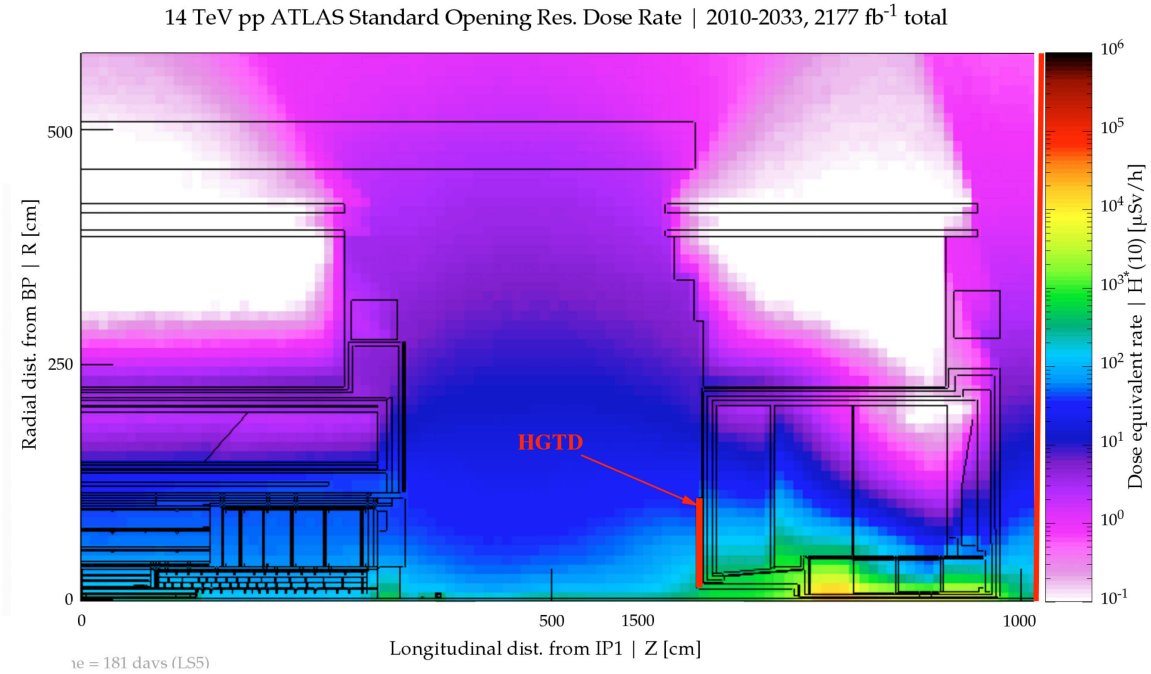


Figure 13.5: FLUKA simulations of ambient dose equivalent rate in LS5, after 2177 fb^{-1} of accumulated data and 181 days cool-down period. ATLAS is in the standard opening configuration [100].

Classification criteria	**Level 1**	**level 2**	**level 3**
Individual dose equivalent	<100 μSv	100 μSv/h - 1 mSv	>1 mSv
Collective dose equivalent	<500 μSv	500 μSv/h - 5 mSv	>5 mSv
Ambient dose equivalent rate	<50 µSv h^{-1}	50 µSv h^{-1} - 2 mSv h^{-1}	>2 mSv h^{-1}
Airborne activity	<5 CA	5 CA - 200 CA	>200 CA
Surface contamination	<10 CS	10 CS - 100 CS	>100 CS

Table 13.1: ALARA classification criteria.

It is expected that the HGTD installation zone will be classified at least as a "simple controlled radiation area", which is defined as the area whose ambient dose equivalent rate H*(10) does not exceed 10 µSv h^{-1} at workplaces or 50 µSv h^{-1} in low occupancy areas. All work in controlled radiation areas will be planned and optimised including an estimate of the collective dose and the individual effective doses to the personnel participating in the activity. This will be described in the DIMR file (Dossier D'Intervention en Milieu Radioactif), which must be prepared for each intervention. The Radiation Protection service will assign an ALARA level to each type of activity, accordingly to the CERN classification criteria, which are shown in Table 13.1. Since the airborne radioactivity and contamination can be ruled out,

the ALARA level classification will be primarily determined by individual and collective effective dose. As can be seen from the table, the HGTD installation activities will be situated between ALARA Level 1 and Level 2, considering the ambient equivalent dose. However, the collective dose during replacement of the inner part of the HGTD at half-life time of HL-LHC on both end-caps might approach the limit of 5 mSv, which corresponds to the Level 3 threshold. In this case the Level 3 scenario is applied, which involves additional optimisation efforts and implies that dose planning and work organisation are reviewed by the ALARA committee. DIMR level I and level II will be prepared and discussed between the intervening personnel and the ATLAS radiation safety officer (RSO) and LEXGLIMOS prior to intervention, which can only start when the DIMR is approved. All the activities will be followed by the RSO and LEXGLIMOS on a day-by-day basis, involving CERN Radio Protection experts when necessary.

Beside the careful work optimisation, additional measures to help minimise the exposure of personnel to radiation will be considered. Such measures include shielding, which will reduce the dose rate to the human body; use of tools for remote handling; organising the working place in such a way, that people are placed in the outer radius of HGTD avoiding exposure to the area near the beam line, where the dose rate is much higher.

14 Demonstrator

14.1 Introduction

The R&D period will extend up to early 2022 to validate the choice of many components before the Final Design Reviews. In addition, it is essential to validate some key aspects of the integration during this period by building a realistic demonstrator. The plan is to have a two step plan decoupling the mechanics/cooling aspects from the full electronics/DAQ demonstrator activities. The heater demonstrator will be based on a silicon-based heater substrate to study the thermal performance of the system, instead of a real sensor and ASIC module which will only be ready at the earliest by 2021 (further information may be found in Section 6.9). The full demonstrator will be similar to the heater demonstrator but equipped with some HGTD modules and read-out through a prototype of the peripheral electronics and back-end. A dedicated organisation is being set up to ensure coherence of the numerous parallel activities and monitor the schedule.

14.2 Heater demonstrator

The goals of this demonstrator are:

- Use the simple cooling plate system to validate the CO_2 thermal calculation which will be used for the final design of the HGTD cooling loops.
- Choose and validate the module loading procedure (intermediate plate, gluing, flex cable stacking...) by equipping the demonstrator with heaters in a geometry similar to the HGTD modules.

The demonstrator will consist of a rectangular cooling plate covering about 7 cm $\times$ 80 cm as displayed in Figure 14.1, corresponding roughly to the longest detector unit in the HGTD. The cooling system will be made of a single loop (technical details given in Annex) embedded in a carbon fibre structure and will be used first to validate the thermal calculation of the CO_2 cooling on a simple design: CO_2 cooling parameters such as pressure and flow will be varied and the temperature on the plate will be measured with Resistance Temperature Detectors (RTDs) embedded into heaters. Heaters are used as a replacement of the full size module (sensor + ASIC). They will be placed on top of the cooling plate in a similar manner

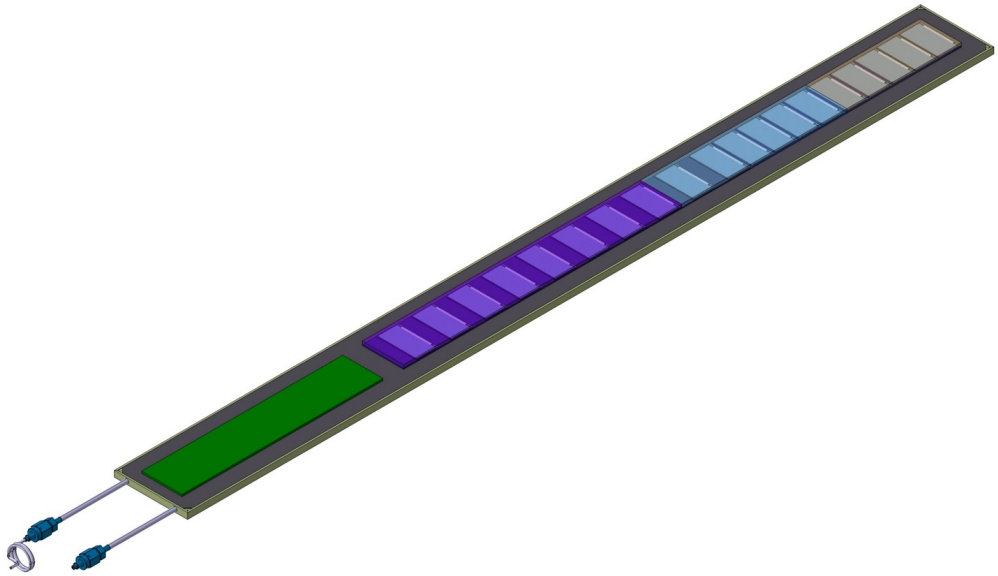


Figure 14.1: Schematic view of the cooling structure equipped with heater modules in blue. The green area corresponds to the peripheral electronics board.

as the modules in a given readout row. A dedicated vessel should be also built, allowing dry nitrogen flushing and a feed-through for electrical connections. The convection conditions should be as close as possible to the final ones. The mechanical prototype can be found in Figure D.9.

After this first set of measurements, the detector unit should be mounted on the top and bottom faces of the cooling plates. Figure 14.2 shows preliminary calculations of the temperature uniformity for both options that will be compared to the measurements. As expected, the calculation predicts a uniform temperature with the pattern intermediate plate, 0.4 K between inner and outer module, while up to 1.8 K is observed with the full intermediate plate.

Real HGTD modules will not be available before 2021. Consequently to mimic the radial heat dissipation expected in the HGTD, silicon heater devices similar to the ones used by the pixel ITk demonstrator will be used for the module loading. Thus the silicon heater demonstrator program will address two important aspects of the HGTD system: module loading and thermal performance. A schematic drawing of the silicon heater is shown in Figure 14.3.

The heaters consist of a silicon substrate with a similar geometry (area) as the modules and a thickness of 300 µm. A geometry slightly smaller than the final HGTD module was chosen due to ease of production by the manufacturer. The heaters will have a size of 20.2 mm $\times$ 38.4 mm. They will be made of a TiW continuous layer produced on a 300 µm

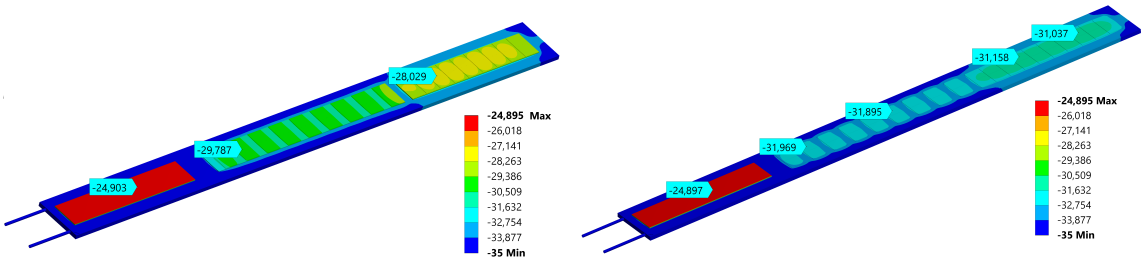


Figure 14.2: Expected temperature uniformity on the demonstrator equipped with the full intermediate plate (left) or the pattern intermediate plate (right)

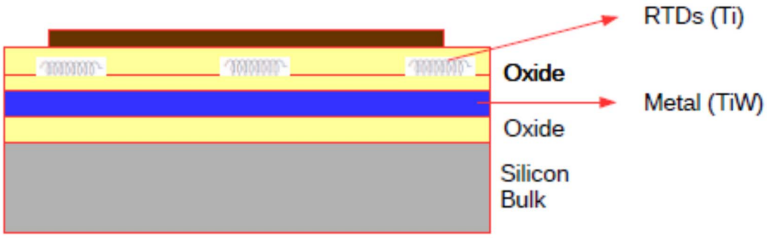


Figure 14.3: Silicon heater transverse view

thick wafer. The heaters dissipate power by applying a current through a thin metal layer embedded in the silicon substrate. The amount of generated heat can be controlled through the provided current. In order to monitor the temperature of the heater RTDs are implanted on top of the thin metal separated by an oxide layer. The RTDs will then be placed on top of a second oxide layer separating the heater from the RTDs, which will also be made from TiW. They are operated by applying a current and reading the voltage drop across the RTD which is previously calibrated to provide temperature information. The RTDs are controlled through a flexible cable that also provides the current to the heater element. The flex is glued to the top of the heater and its pads are wire-bonded to the heater. The heater flex PCB design can be found in Figure 14.4. The design has been optimized to be as close as possible to the final design choice for the HGTD.

The heater flex will be designed to mimic the HGTD module flex cable in terms of geometry, material and rigidity. It will contain a connector similar to the one being considered for

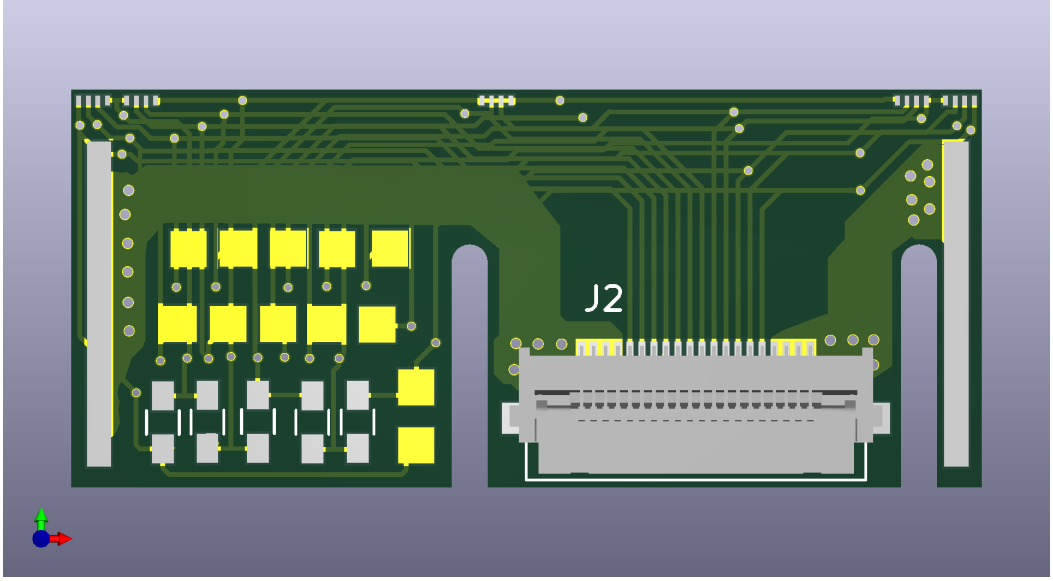


Figure 14.4: Heater flex PCB layout.

the final flex design, which can provide power to the heaters and individual readout lines for the RTDs on each heater. The flex cables will be layered one on top of each other out to the peripheral readout boards. Though the final specifications of the peripheral readout boards will not be available, a compact connector scheme is foreseen. The system will be controlled by external power supplies that will provide the desired operational thermal range to fully study the system performance. The nominal power dissipation foreseen for the innermost part of the heater detector unit is 400 mW cm^{-2}, but deviations from this value will be explored. The entire heater demonstrator will be placed within an isolated container box to maintain temperatures close to -30^{o}C and allow for nitrogen or dry air to be flushed into the apparatus to maintain a dry atmosphere. The CO_2 cooling will be provided by the CO_2 baby demo cooling plant, sitting nearby, as shown in Figure 11.7. The design of the heater demonstrator apparatus can be found in Figure 14.5.

The Institutes that plan to participate in the HGTD module assembly and loading effort will also participate in the heater (and/or full) demonstrator effort and will thus gain expertise on the module assembly process. The calibrations of the RTDs will also be carried out by the Institutes, before and after module loading. The assembly of the intermediate plates around the cooling plane will be carried out at CERN, where the full cooling tests will be conducted.

In summary, the heater demonstrator will allow to validate the thermal performance of the HGTD, by using heaters loaded into a long detector unit and combined with a CO_2 cooling system. Furthermore, the exercise of assembling the heater modules, populating

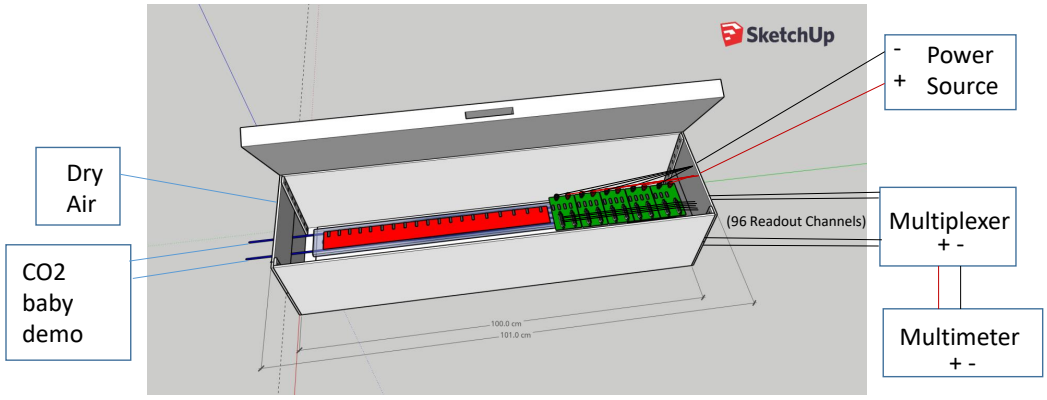


Figure 14.5: Silicon heater demonstrator apparatus including additional components located outside the apparatus. Heaters are shown in red, placed on top of the cooling place. The peripheral boards are shown in green. The flex cables which will connect the heaters to the peripheral boards are not explicitly shown in the figure. A multiplexer is foreseen to switch the active readout between the 96 RTD signals.

the intermediate plates and mounting the full heater demonstrator is expected to provide valuable experience towards the final HGTD detector unit assembly and loading effort. The silicon heater schedule and timeline is detailed in the full demonstrator planning at the end of the chapter in Figure 14.8.

14.3 Peripheral and back-end electronics, data acquisition

The readout demonstrator will exercise the final HGTD read-out path and will be used to validate the PEB, the clock distribution and the FELIX board used for the data acquisition.

14.3.1 Peripheral electronics demonstrator

The peripheral electronics demonstrator will evaluate the different paths from the module flex to the PEB via flex cables like the data transmission, high voltages and the power

distribution. In addition, it will allow to exercise the assembling, connection and integration of the peripheral electronics. It consists of a PEB connected up to 56 HGTD modules via a stack of flex cables. In a first stage, an Spartan-7 FPGA will be used to emulate the ALTIROC2 ASIC and a Kintex-7 FPGA to emulate the lpGBT chipset, while the VTRx+ and the bPOL12V will be replaced by similar commercial components (SPF+ and TPS56428RHLR respectively), given the unavailability of the different items. A scheme of the peripheral electronics demonstrator is shown in Figure 14.6. The design of the different items has already started and a peripheral electronics demonstrator will be ready by Summer 2020. On a second stage, the different components will be replaced by the ones of the final design and will be integrated in the full demonstrator set-up.

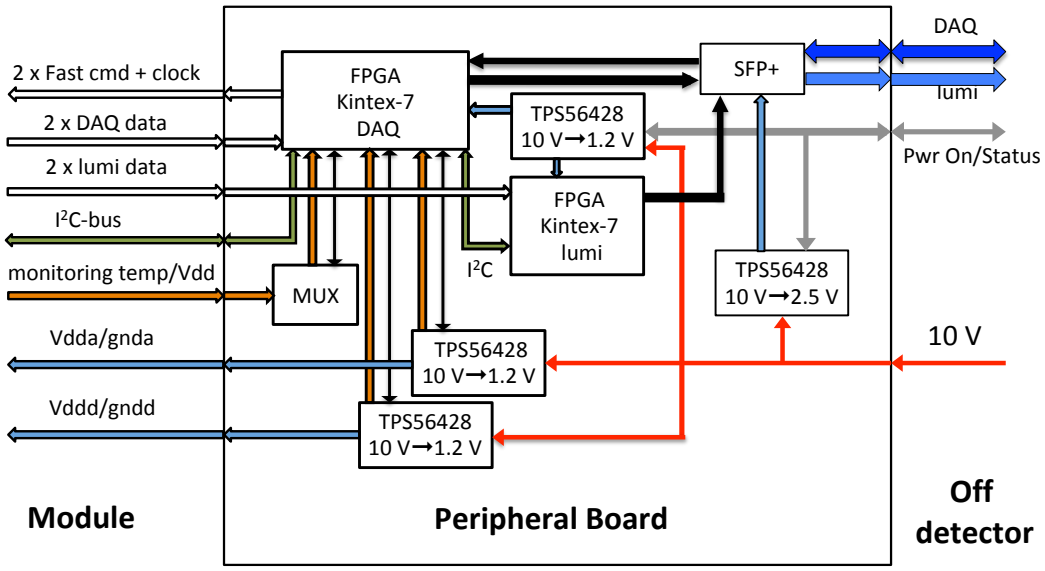


Figure 14.6: Block diagram of the peripheral electronics board demonstrator. A Kintex-7 FPGA will be used to emulate the lpGBT chipset, a SPF+ will replace the VTRx+ and a TPS56428RHLR will used instead of bPOL12V DC-DC converter.

14.3.2 DAQ demonstrator

The DAQ demonstrator will exercise the entire read-out path up to the off-detector back-end. Activities at CERN have already started and a Phase-I FELIX board and its DAQ PC have been purchased. On a first stage, the HGTD e-link data will be emulated inside FELIX in order to test the read-out chain. Afterwards, the FELIX board will be connected to an FPGA emulator that will send HGTD data in FULL mode in order to validate the readout

chain. The ALTIROC2 FPGA emulator described in the previous section can be used for this purpose. When available, the ALTIROC2 will be connected to the readout chain. A GBT chip can serve as the interface between the FELIX board and the ASIC. On a second stage, a Phase-II FELIX board will be purchased for the integration and validation of lpGBT. The DAQ demonstrator roadmap is shown in Figure 14.7.

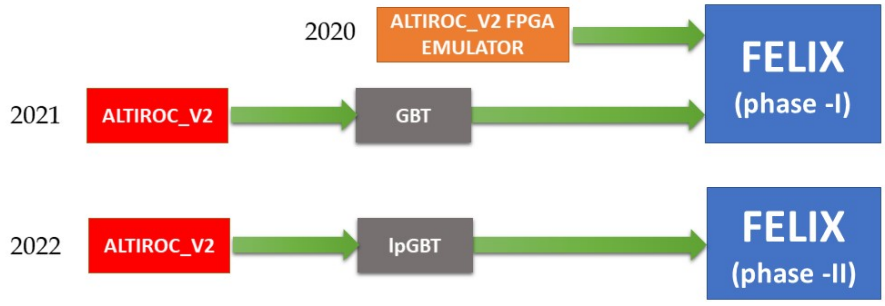


Figure 14.7: DAQ demonstrator roadmap. In 2020, an ALTIROC2 FPGA will be used to interface with FELIX. In 2021 an ALTIROC2 ASIC will be connected to FELIX using a GBT chip as interface. In 2021 a Phase-II FELIX board will interface the ALTIROC2 via lpGBT close to the final design.

Furthermore, the DAQ demonstrator will be used to measure the different contributions to the clock jitter at different stages (FELIX, lpGBT, FLEX, ALTIROC2). It will be used to develop a calibration procedure close to the final design. Finally, the DAQ demonstrator will be integrated in the full demonstrator set-up.

14.3.3 HGTD module

The production of the HGTD modules will be used to validate the module assembly, loading process (gluing, wire bonding and mounting), and quality control measurements procedures used during the production.

To gain experience for this process, smaller bare modules have been assembled in house during 2019 using the ALTIROC1 ASICs and the existing 5×5 pads sensor. For test beam purposes, dedicated printed circuit boards have been developed and already used to test the ASIC. It is also foreseen to develop a flex compatible with the ALTIROC1 read-out to exercise the gluing and wire bonding of the bare module, as a first step of the validation of the module assembly. Dedicated custom made readout boards will be used to validate these modules, using calibration signals and a beta source. These read-out boards could be used on the demonstrator until the FELIX setup is operational.

The bump bonding of the sensor to the ASIC will be outsourced to a company and require a complete wafer for the under-bump-metalization process before the flip-chip. A specification document has been prepared and is currently in discussion with two companies in Germany and China. Complete wafers will be available only after the production of ALTIROC2 and a dedicated sensor production. The validation of the industrial bump bonding process will be completed early Q3/2021. The possibility to produce the hybrids for the demonstrator program in the HGTD Institutes that have this capability in-house is also an option. Between 5 to 10 bare HGTD modules are expected to be delivered by end of Q3 2021.

Prototypes of the flex cable should also be produced, but the connector to the peripheral board might still be not the final one.

14.4 Full demonstrator

The assembly of the demonstrator will start in Q4 2021. It will be made of :

- The mechanical structure as used in the heater demonstrator, available by mid 2020.
- Five to ten HGTD modules available by end of Q3 2021 and heater modules. A test of these modules after integration on the detector units should be done using the custom made read-out board to qualify the modules.
- At least one peripheral board able to read up to five HGTD modules connected through flex cables.
- A FELIX I/O card with its DAQ PC.
- Prototypes of Low Voltage and High Voltage modules, with DCS, might be used but are not mandatory for this test.

14.5 Full demonstrator tests

A period of about three months will be available before the first FDR. While intense electronics calibration sequence tests will be performed, two options are investigated for the calibration sources : cosmic test bench with a precise trigger time measurement (although the rate might be insufficient) or a portable x-ray source (8 keV or 40 keV source) with a motorised stage to scan the detector unit.

14.6 Schedule and organisation

A tentative plan for the demonstrator program is shown in Figure 14.8. While the schedule for the heater demonstrator contains some contingency, the main risks for the full demonstrator rely on the availability of the modules in Q3 2021. This is strongly linked to the ASIC and sensor productions. Beginning in Q3 2020, weekly follow-up meetings will be mandatory to fulfil this aggressive schedule. A dedicated working group has been set up in Q2 2019, focussing on both mechanics/module oriented demonstrator activities for the heater demonstrator as well as the electronics/DAQ oriented activities. Beyond January 2022, the demonstrator is expected to stay operational until the end of the production for additional tests.

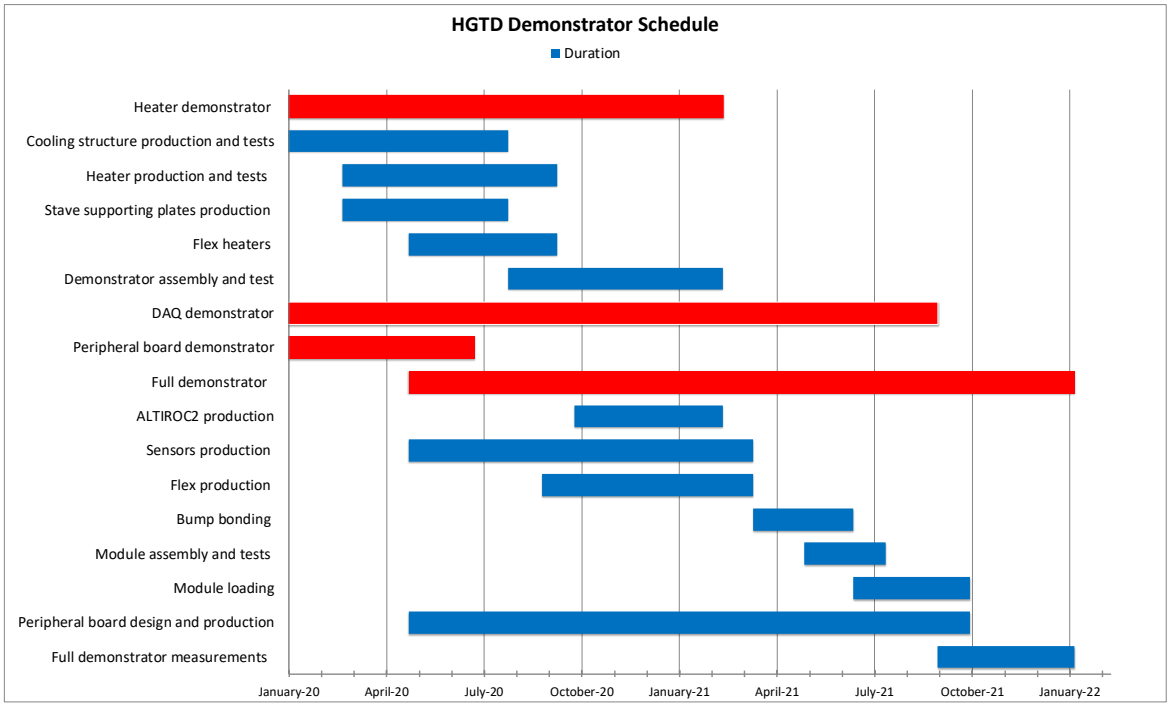


Figure 14.8: Planning of the heater and full demonstrator from January 2020 to January 2022. Red items show the timeline of the overall demonstrator projects; blue items show the expected progress of the individual items.

15 Project Organization, Costs, and Schedule

This chapter describes the overall organization of the HGTD project. Section 15.1 presents the way the project is organized and the management of the different activities, including a detailed breakdown for each component of the project. Section 15.2 discusses the schedule towards the detector completion. The foreseen available resources are discussed in Section 15.3. Finally, in Section 15.4 the risks involved with the project and the strategies to mitigate them are discussed.

15.1 Organization and management

15.1.1 Upgrade organisation in ATLAS

The highest-level executive body in ATLAS is the Executive Board (EB), chaired by the Spokesperson with the Technical Coordinator (TC) as deputy chair. The overall steering and monitoring of the upgrade activities is delegated to the Upgrade Steering Committee (USC), which is a sub-committee of the EB, with an extended membership. The USC is chaired by the Upgrade Coordinator (UC). The review and approval of Upgrade Projects (UPRs) is steered by the UC and the USC, with approval of such projects by the EB, subject to endorsement by the Collaboration Board (CB). The UC also oversees and monitors the overall upgrade planning and schedules. The management of approved Upgrade Projects rests with the Upgrade Project Leader (UPL) of that UPR, acting together with the parent system's Project Leader (PL) and Institute Board chair. The UC should be well informed of the activities in the UPRs, and interacts regularly with the UPLs to anticipate technical, schedule, resource, or other problems. The TC, supported by the Technical Coordination organization (TCn), is responsible for ensuring that all the upgrades can be successfully integrated in the ATLAS detector, that their installation schedules are compatible with shutdown schedules, and that there are adequate resources allocated for the installation and commissioning of the upgrade detectors. To this end the TC has organized an Upgrade Project Office (UPO) that provides technical support to the UPRs and the UC. Moreover the TC is responsible for the upgrade of all the common infrastructure needed for the upgrade program.

The Review Office is an independent body embedded in TCn. In close collaboration with the UC, the TC, and the UPLs, the Review Office develops and organizes technical reviews for the components of the upgrades following the ATLAS review strategy, see Section 15.1.3.

15.1.2 HGTD project organisation

The HGTD started as an organized activity in summer 2015 and the corresponding new sub-detector proposal was part of the ATLAS Upgrade Scoping Document [42]. The Initial Design Report and Expression of Interest were approved by ATLAS and LHCC in 2017. The Technical Proposal was approved by the LHCC in June 2018 [101], with the recommendation to proceed to the Technical Design Report. The HGTD Interim Upgrade Project Leader(s) represent the project in the ATLAS USC and chair the HGTD Steering Group.

The current HGTD project management organization is shown in Figure 15.1. The Resources and Risk Coordinator assists the Project Leader(s) in the Resources and Risk management. He/she coordinates the preparation of the material to be reviewed by the Upgrade Costing Group (UCG), Memorandum of Understanding (MoU) and works closely with the other members of the management team. The HGTD Institute Board (IB) has one representative per Institution.

The project is organized in eight working groups (WG). Each WG, coordinated in general by two co-coordinators (LV2 coordinators), carries out several activities, as detailed in the organization chart shown in Figure 15.1:

- **Sensors:** Currently in charge of the sensor R&D including irradiation tests with the aim of delivering the specifications of the final sensors. It works closely with the electronics WG as the expected performance relies strongly on the combined performance sensor+ASIC and with the testbeam WG. After the R&D Phase, it will have the charge to perform the market survey and manage the production and QA tests.

- **Electronics:** In charge of all electronics activities from the ASIC (design, specifications, production and QA) to the Peripheral Electronics Boards (design, specifications, production and QA). It interacts with the sensors WG (for the ASIC specifications, High Voltage), the DAQ WG (for the data format, bandwidth) and the Module assembly (for the flex) and the Mechanics/assembly WG (for the CO_2 cooling power, services).

- **Luminosity DAQ and control:** Responsible for the simulation studies and the specific hardware for the luminosity measurement and the DAQ aspects (including the FELIX, and control). It makes the interface with the ATLAS luminosity group and the ATLAS upgrade DAQ and DCS projects. A specific sub-group is in charge of studying and implementing the clock calibration (online and offline).

- **Modules and Detector Units:** In charge of defining the module assembly (bump bonding, gluing, flex) specifications, procedure and QA, and the modules loading in Detector Units specification and QA.
- **Test Beam:** In charge of developing the needed tools for the testbeam (DAQ and hardware) and of the data analysis. It works closely with the sensors and electronics WG.
- **Demonstrator:** Cross-cutting WG to all the others at the exception of the Simulation/Performance WG. It will start its activity after the TDR delivery with the aim of building the demonstrator and validate the performance for the PDR of most of the components as described in chapter 12. This WG on long-term might evolve to take the charge of the commissioning of the final detector.
- **Mechanics, assembly and installation:** In charge of providing the specifications and building the vessels and cooling plates, the service definition and routing (with TC), and the water/CO_2 cooling plants (with CERN support groups). It is also in charge of the tools design needed for the assembly at surface and installation in the pit. The Detector assembly and final installation procedures are also discussed here. In a later stage this WG will be split into more WGs, when the assembly and Installation will represent a sizeable effort.
- **Simulation performance and physics:** Responsible for providing the most realistic simulation package and reconstruction tools (in interaction with the ATLAS Upgrade ITk simulation and performance and the Upgrade Physics group) to evaluate the performance on the object reconstruction and the impact on some physics channels.

In a few cases, the level 3 activity coordinators are already identified, and will all be appointed after the TDR approval. All LV2 coordinators are members of the HGTD Steering Group. Topical meetings in each WG area are organized by the WG coordinators on a bi-weekly basis. HGTD general meetings are organized by the UPLs and take place bi-monthly during 3-day Mini-Weeks. During these HGTD weeks joint Steering Group and IB meetings are organized to discuss and endorse any strategic decision on detector layout, resource needs, etc.

In the Summer 2020, a formal IB including only the institutions that will participate in the HGTD construction will be created. After the expected approval of the TDR by the CERN Research Board in September 2020, the new elected IB chair will start the process of the new UPL(s) election. One interim PL and its deputy will stay in charge until beginning of 2021, when the new UPL(s) will be elected. The need of a technical coordinator after the TDR approval, or for the construction phase, will be carefully evaluated.

The ongoing R&D is carried out by roughly 150 physicists, engineers and technicians from 30 ATLAS Institutes, and 13 countries/Funding Agencies, see Table 15.1, who are committed to carry out the R&D needed to mature the proposed detector. Table 15.2 summarizes the

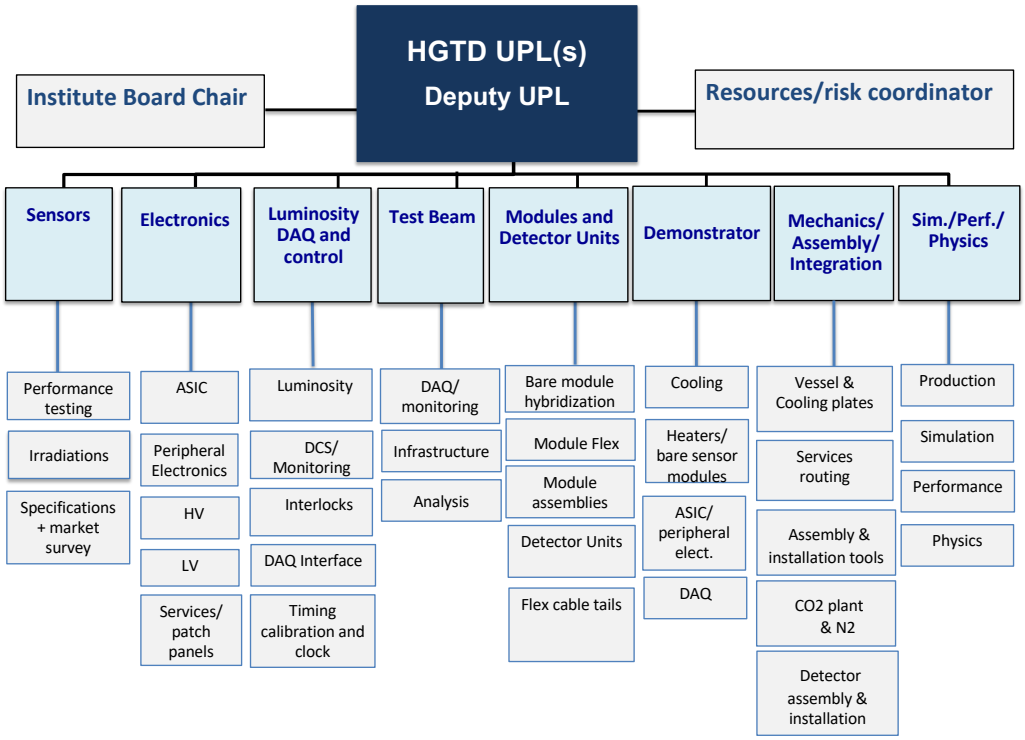


Figure 15.1: HGTD organisation chart.

present involvement of the Institutes in the various R&D activities, planned until 2021. A sizeable fraction of these Institutes are already committed to the next steps of construction, installation and commissioning of the HGTD, and are able to cover the necessary labour effort, discussed in Section 15.3.2.

Country/FA	Institutes involved in HGTD R&D
Brazil	USP
CERN	CERN
China	IHEP, NJU, USTC, SINANO, SJTU
France	IJCLab, LPC, LPNHE, OMEGA
Germany	Mainz, Giessen, Goettingen*
JINR	JINR
Morocco	UIT, UH2C, UM5R, UMP
Russia	MEPhI
Slovenia	JSI
Spain	IFAE
Sweden	KTH
Taiwan	AS, NTHU
USA*	BNL, SLAC, SMU, UCSC, SUNYSB

Table 15.1: List of countries/Funding Agencies and corresponding Institutes contributing to HGTD R&D. OMEGA and SINANO are ATLAS Technical Associate Institutes. *Goettingen and USA Institutes will only be involved in the R&D phase.

R&D Activities/WG	Institutes
Sensors	BNL, CERN, Goettingen, IFAE, IHEP, JINR, JSI, USTC, USP, UCSC
Electronics	AS, Giessen, IFAE, IHEP, IJCLab, JINR, KTH, LPC, NJU, NTHU, Omega, SLAC, SMU, SUNYSB, UIT, UH2C, UM5R, UMP, USTC
Luminosity, DAQ and Control	IHEP, KTH, Giessen, UCSC, UIT, UH2C, UM5R, UMP
Test beams and demonstrator	All Institutes
Module assembly and loading	BNL, IFAE, IHEP, IJCLab, JINR, LPNHE, Mainz, UIT, UH2C, UM5R, UMP, USTC, SINANO
Mechanics, assembly and installation	CERN, IHEP, IJCLab, JINR, LPNHE, MEPhI
Simulation/Performance/Physics	All Institutes

Table 15.2: List of R&D activites and participating Institutes. OMEGA and SINANO are ATLAS Technical Associate Institutes. US groups will only be involved in the R&D phase. Goettingen is only involved in the Sensors R&D phase.

15.1.3 Technical milestones

All of the custom components used for the HGTD have to pass through a series of reviews before purchase orders can be placed for procurement of parts and production of the deliverables. These reviews are used to ascertain the quality and reliability of the components at various steps in the development and production process. They can also help to shorten

the design phase, by enforcing in-depth presentations of the status at various stages. Reviews are usually conducted as a half-day or full-day meeting between a review panel and the group of people in charge of the component design and construction. The review panel is designated by the UC or by the Upgrade Review Office, and includes experts in the relevant technology, and, if applicable, users of the object to be reviewed or those interfacing other objects to it. This procedure is the ATLAS standard. There are four main reviews for each custom component:

Specifications Review (SPR) This review verifies that complete written requirements exist, that they are sufficient to develop the designs, and specifications include sufficient opportunities for QC/QA. Specifications are compiled in a **Specifications Document**, which is internally reviewed in the project prior to the SPR, and formally released after a successful SPR.

Preliminary Design Review (PDR) The PDR verifies that prototype designs meet all aspects of the specifications documents. Technical feasibility of the design must be demonstrated, so simulations or partial prototypes demonstrating feasibility for critical functions are important. Furthermore, test plans shall show how the prototype devices will be tested to demonstrate functionality that meets the specification. Safety aspects of the design will be reviewed. Integration of the system into its environment will be verified and its installation feasibility will be assessed.

Final Design Review (FDR) The FDR reviews all the available data from prototypes to determine how well the design meets specifications. For components of a larger system, analysis and measurements demonstrating compatibility with external interfaces, consistent with specifications, are essential. Specifications documents should have been approved after SPR, and if applicable, modified and again approved after the PDR. A successful FDR gives the green light for the pre-production fabrication or build to proceed, and for the first CORE expenditures. The number of prototype devices produced after the review is usually small and is up to the discretion of the project leader, however must be of a large enough number to provide at least the minimum of meaningful statistics (typically 5% of the total production).

Production Readiness Review (PRR) The purpose of the production readiness review (PRR) is to demonstrate overall production readiness and assure that the items to be produced will meet the defined requirements. The results from pre-production are used to verify that larger scale production can be done with the acceptable yields, and that the quality control process is sufficiently thorough to filter out devices that will not meet the performance specification over the lifetime of ATLAS. All necessary production plans, travelers, tools, facilities and other resources shall be in place. Closure of Actions from the previous Reviews is a requirement as well. After successful PRR, the distributed production sites are qualified and the design is cleared for full production.

These reviews mark the transitions between different phases in each component's development and production schedule, and thus are used as key technical milestones in the overall project schedule, discussed in Section 15.2.

The co-coordinators of each WG are responsible for the preparation of the specifications and documentation, quality acceptance procedures, and material to be delivered to the reviews. Each individual component that will be built into the HGTD must have a written specification. The progress through the reviews is also used to monitor the progress of the project and to make sure it is on track. Production procurement, especially for large quantity items, will require a production plan and must follow procurement procedures required by the purchasing Institution. The CERN procurement office will likely be responsible for the procurement of large-quantity items whose CORE cost is shared across multiple funding agencies.

15.1.4 Deliverables and WBS

The deliverables for the construction of the HGTD are organized in an hierarchical Product Breakdown Structure (PBS), with a direct correspondence to the existing first five WG activities listed above. The PBS indicates the deliverables, to be assigned to a CORE value in the MoU. The Work Breakdown Structure (WBS) is seeded by the PBS and includes the tasks required to produce the deliverables. The PBS organises the deliverables into seven primary categories (LV2), with PBS numbering from 8.1 to 8.7. The PBS is further broken down into lower levels items, shown in Table 15.3 down to level 3 (LV3). In some items, in particular item 8.3 (Luminosity, DAQ and Control), item 8.6 (Detector Assembly and QA on surface) and item 8.7 (Detector Installation and Commissioning), the structure also contains LV3 activities that only require labour effort and hence, are only part of the WBS. All PBS items have an associated CORE cost as described later in Section 15.3, while WBS-only items do not.

PBS/WBS	Deliverable
8.1	**Sensors**
8.1.1	LGAD Sensors
8.2	**Electronics**
8.2.1	ASIC
8.2.2	Peripheral Electronics Board
8.2.3	High Voltage system
8.2.4	Low Voltage system
8.3	**Luminosity, DAQ and Control***
8.3.1	Luminosity boards
8.3.2	DCS
8.3.3	Interlocks and protection system
8.3.4	DAQ software
8.4	**Modules and Detector Units**
8.4.1	Bare module hybridization
8.4.2	Module Flex
8.4.3	Modules assemblies
8.4.4	Detector Units
8.4.5	Flex Cable Tails
8.5	**Mechanics, Services and Infrastructure**
8.5.1	HGTD Hermetic vessel
8.5.2	On detector cooling system
8.5.3	CO_2 cooling system
8.5.4	Water cooling system
8.5.5	Nitrogen system
8.5.6	Cables and connectors
8.5.7	Fibers and optical connectors
8.6	**Detector Assembly and QA on surface**
8.6.1	Test bench for detector certification
8.6.2	Tools for surface assembly
8.6.3	Assembly of components on cooling plates
8.6.4	Final integration inside vessels
8.7	**Installation and Commissioning**
8.7.1	Tools for transport and cavern installation
8.7.2	Services, patch panels and cooling installation
8.7.3	Back-end electronics installation in USA15
8.7.4	Detector installation and connectivity
8.7.5	Global commissioning in LS3

Table 15.3: Product Breakdown Structure (PBS) and Work Breakdown Structure (WBS) of the HGTD down to level 3. The WBS is seeded by the PBS and includes the tasks required to produce the deliverables. WBS-only items are mentioned explicitly when appropriate. **(*)** DAQ hardware deliverables are covered in the ATLAS TDAQ PBS.

15.2 Schedule and main milestones

The HGTD installation schedule presented here uses as a reference the ATLAS TC schedule available until mid-Jan 2020. Development and optimization of the ATLAS detector installation sequence in LS3 will continue for several years. If necessary, the design of the HGTD allows for a later installation, during the following YETS, of possible missing instrumented half rings even in the presence of the beam pipe.

There are three main schedule phases for HGTD:

- **2018 – 2021** R&D
- **2021 – 2026** Construction
- **2026 – 2027** Integration, installation and commissioning

The plan is to install the HGTD detector in the Long Shutdown LS3 on the end-cap LAr calorimeter cryostat faces. This operation should take place in April 2026 and January 2027 for the A and C sides respectively.

To define the schedule of the HGTD Upgrade Project, a detailed bottom-up plan of activities has been worked out. The high-level schedule presented here comprises the reviews and main tasks that need to be undertaken between now and the completion of the project, and their dependencies, i.e. lists of tasks that have to be finished before a new task can begin.

Tables 15.4 to 15.10 show an overview schedule down to the LV3 PBS/WBS. The start points and end points of these phases are delimited by appropriate high-level milestones:

- SPR: Specifications Review;
- PDR: Prototype design meets all aspects of the specifications;
- FDR: Pre-production fabrication or build cleared to proceed;
- PRR: Full production phase cleared to proceed;
- End of the production phase: Construction Completed;
- End of the installation and commissioning phase: Installation Completed.

The schedule deliverables, detailed here up to LV3 only, are defined by the sub-project coordinators and approved by the Steering Group. It is the responsibility of sub-project coordinators to plan, implement, execute, and track the progress of their project according to the baseline schedule for their respective deliverables. They report on the progress to the Steering Group. It is the responsibility of the HGTD UPL to ensure that a comprehensive schedule is developed, to seek the necessary review process to baseline the schedule, to oversee the progress and take necessary corrective actions to ensure that the project remains on schedule, and to propose changes to the baseline as required.

PBS/WBS	Milestone	Date
8.1 Sensors		
8.1.1 LGAD Sensors	SPR	Q3 2020
	PDR	Q1 2021
	FDR	Q4 2021
	Pre-production	Q1 2022 – Q3 2022
	PRR	Q4 2022
	Production (0–50%)	Q1 2023 – Q4 2023
	Production (51–100%)	Q1 2024 – Q4 2024

Table 15.4: Summary of main HGTD Milestones, including the planned reviews (PDR, FDR, PRR), pre-production and production for the Sensors deliverable.

PBS/WBS	Milestone	Date
8.2 Electronics		
8.2.1 ASIC	SPR	Q3 2020
	PDR	Q4 2021
	FDR	Q4 2022
	Pre-production	Q4 2022 – Q3 2023
	PRR	Q4 2023
	Production (0–50%)	Q4 2023 – Q2 2024
	Production (51–100%)	Q2 2024 – Q3 2024
8.2.2 Peripheral Electronics Board	SPR	Q1 2021
	PDR	Q3 2021
	FDR	Q1 2022
	Pre-production	Q2 2022 – Q4 2022
	PRR	Q4 2022
	Production (0–50%)	Q4 2022 – Q1 2024
	Production (51–100%)	Q1 2024 – Q1 2025
8.2.3 High Voltage system	SPR	Q4 2021
	PDR	Q1 2022
	FDR	Q4 2022
	Pre-production	Q1 2023 – Q4 2023
	PRR	Q1 2024
	Production (0–50%)	Q2 2024 – Q3 2025
	Production (51–100%)	Q3 2025 – Q3 2026
8.2.4 Low Voltage system	SPR	Q4 2021
	PDR	Q1 2022
	FDR	Q4 2022
	Pre-production	Q1 2023 – Q4 2023
	PRR	Q1 2024
	Production (0–50%)	Q2 2024 – Q3 2025
	Production (51–100%)	Q3 2025 – Q3 2026

Table 15.5: Summary of main HGTD Milestones, including the planned reviews (PDR, FDR, PRR), pre-production and production for the electronics deliverables.

PBS/WBS	Milestone	Date
8.3 Luminosity, DAQ and Control		
8.3.1 Luminosity boards	SPR	Q4 2021
	PDR	Q1 2022
	FDR	Q1 2023
	Pre-production	Q2 2023 – Q1 2024
	PRR	Q2 2024
	Production (0-100%)	Q2 2024 – Q1 2025
8.3.2 DCS	SPR	Q4 2021
	PDR	Q1 2022
	FDR	Q4 2022
	Pre-production	Q1 2023 – Q4 2023
	PRR	Q1 2024
	Production (0–50%)	Q2 2024 – Q2 2025
	Production (51–100%)	Q3 2025 – Q3 2026
8.3.3 Interlocks and protection system	SPR	Q4 2021
	PDR	Q1 2022
	FDR	Q4 2022
	Pre-production	Q1 2023 – Q4 2023
	PRR	Q1 2024
	Production (0–50%)	Q2 2024 – Q2 2025
	Production (51–100%)	Q3 2025 – Q3 2026
8.3.4 DAQ software	SPR	Q4 2021
	PDR	Q1 2022
	FDR	Q4 2022
	PRR	Q2 2023
	Production (0-100%)	Q2 2023 – Q1 2027

Table 15.6: Summary of main HGTD Milestones, including the planned reviews (PDR, FDR, PRR), pre-production and production for the Luminosity, DAQ and Control deliverables.

PBS/WBS	Milestone	Date
8.4 Modules and Detector Units		
8.4.1 Bare module hybridization	SPR	Q2 2021
	PDR	Q1 2022
	FDR	Q4 2022
	Pre-production	Q1 2023 – Q4 2023
	PRR	Q4 2023
	Production (0–50%)	Q1 2024 – Q4 2024
	Production (51–100%)	Q3 2024 – Q1 2025
8.4.2 Module Flex	SPR	Q4 2021
	PDR	Q2 2022
	FDR	Q3 2022
	Pre-production	Q4 2022 – Q1 2023
	PRR	Q2 2023
	Production (0–50%)	Q3 2023 – Q4 2023
	Production (51–100%)	Q4 2023 – Q2 2024
8.4.3 Modules assemblies	SPR	Q1 2022
	PDR	Q3 2022
	FDR	Q2 2023
	Pre-production	Q3 2023 – Q1 2024
	PRR	Q1 2024
	Production (0–50%)	Q3 2024 – Q3 2025
	Production (51–100%)	Q3 2025 – Q3 2026
8.4.4 Detector Units	SPR	Q1 2022
	PDR	Q4 2022
	FDR	Q2 2023
	Pre-production	Q2 2023 – Q1 2024
	PRR	Q2 2024
	Production (0–50%)	Q2 2024 – Q3 2025
	Production (51–100%)	Q3 2025 – Q3 2026
8.4.5 Flex Cable Tails	SPR	Q1 2022
	PDR	Q2 2022
	FDR	Q1 2023
	Pre-production	Q1 2023 – Q4 2023
	PRR	Q1 2024
	Production (0–50%)	Q1 2024 – Q4 2024
	Production (51–100%)	Q1 2025 – Q3 2025

Table 15.7: Summary of main HGTD Milestones, including the planned reviews (PDR, FDR, PRR), pre-production and production for the Module and Detector Units deliverable. The Detector Units include Support Units that will be produced ahead of the Module loading activity.

PBS/WBS	Milestone	Date
8.5 Mechanics, Services and Infrastructure		
8.5.1 HGTD Hermetic vessel	SPR	Q1 2021
	PDR	Q2 2021
	FDR	Q1 2022
	PRR	Q3 2022
	Production (0–50%)	Q3 2022 – Q3 2023
	Production (51–100%)	Q3 2023 – Q3 2024
8.5.2 On detector cooling system	SPR	Q1 2021
	PDR	Q2 2021
	FDR	Q1 2022
	PRR	Q3 2022
	Production (0–50%)	Q3 2022 – Q3 2023
	Production (51–100%)	Q3 2023 – Q3 2024
8.5.3 CO_2 cooling system		
8.5.3.1 CO_2 plants **(*)**	SPR	Dec 2018 (passed)
	PDR	Q4 2020
	FDR+PRR (combined)	Q3 2021
	Production (0-100%)	Q2 2022 – Q3 2024
8.5.3.2 Manifold boxes	SPR	Q1 2021
	PDR	Q3 2021
	FDR	Q2 2022
	PRR	Q1 2023
	Production (0-100%)	Q1 2023 – Q2 2025
8.5.3.3 Cooling lines	SPR	Q1 2021
	PDR	Q3 2021
	FDR	Q2 2022
	PRR	Q1 2023
	Production (0-100%)	Q1 2023 - Q2 2025
8.5.4 Water cooling system	SPR	Q1 2021
	PDR	Q3 2021
	FDR	Q2 2022
	PRR	Q1 2023
	Production (0-100%)	Q1 2023 – Q2 2025
8.5.5 Nitrogen system	SPR	Q1 2021
	PDR	Q3 2021
	FDR	Q2 2022
	PRR	Q1 2023
	Production (0-100%)	Q1 2023 – Q2 2025
8.5.6 Cables and connectors	SPR	Q3 2021
	PDR	Q1 2022
	FDR	Q3 2022
	PRR	Q1 2023
	Production (0-100%)	Q2 2023 – Q3 2025
8.5.7 Fibers and optical connectors	SPR	Q3 2021
	PDR	Q1 2022
	FDR	Q3 2022
	PRR	Q2 2023
	Production (0-100%)	Q2 2023 – Q3 2025

Table 15.8: Summary of main HGTD Milestones, including the planned reviews (PDR, FDR, PRR), pre-production and production for PBS 8.5 Mechanics, Services and Infrastructure.
(*) The CO_2 cooling system (PBS 8.5.3) is detailed to LV4 since the CO_2 plants (PBS 8.5.3.1) will be reviewed and constructed in common with ATLAS ITk and CMS (calo,tracker,timing detectors) and much earlier than the specific HGTD CO_2 cooling items (8.5.3.2 Manifold boxes and 8.5.3.3 Cooling lines).

PBS/WBS	Milestone	Date
8.6 Detector Assembly and QA on surface		
8.6.1 Test bench for detector certification	SPR	Q4 2021
	PDR	Q2 2022
	FDR	Q4 2022
	PRR	Q2 2023
	Production (0-100%)	Q2 2023 – Q4 2023
8.6.2 Tools for surface assembly	SPR	Q4 2021
	PDR	Q2 2022
	FDR	Q4 2022
	PRR	Q2 2023
	Production (0-100%)	Q2 2023 – Q4 2023
8.6.3 Assembly of components on cooling plates	SPR	Q2 2022
	PDR	Q4 2022
	FDR	Q4 2023
	PRR	Q2 2024
	HGTD–A assembly (0–50%)	Q3 2024 – Q4 2025
	HGTD–C assembly (51–100%)	Q4 2025 – Q4 2026
8.6.4 Final integration inside vessels	SPR	Q2 2022
	PDR	Q4 2022
	FDR	Q4 2023
	PRR	Q2 2024
	HGTD–A integration (0–50%)	Q4 2024 – Q4 2025
	HGTD–C integration (51–100%)	Q4 2025 – Q4 2026

Table 15.9: Summary of main HGTD Milestones, including the planned reviews (PDR, FDR, PRR), pre-production and production for PBS 8.6 Detector Assembly and Quality Assurance on surface.

.

PBS/WBS	Milestone	Date
8.7 Installation and Commissioning		
8.7.1 Tools for transport and cavern installation	SPR	Q2 2022
	PDR	Q2 2023
	FDR	Q4 2023
	PRR	Q2 2024
	Production (0-100%)	Q2 2024 – Q3 2025
8.7.2 Services, patch panels and cooling installation	Installation+QA (0–100%)	Q1 2025 – Q2 2026
8.7.3 Back-end electronics installation in USA15	Installation+QA (0–100%)	Q2 2025 – Q4 2026
8.7.4 Detector installation and connectivity		
	Schedule float (HGTD–A)	Q4 2025 – Q2 2026
	HGTD–A (0–50%)	**Q2 2026**
	Schedule float (HGTD–C)	Q4 2026 – Q1 2027
	HGTD–C (50-100%)	**Q1 2027**
8.7.5 Global Commissioning in LS3	Commissioning	Q2 2026 – Q2 2027

Table 15.10: Summary of main HGTD Milestones, including the planned reviews (PDR, FDR, PRR), pre-production and production for PBS 8.7 Installation and Commissioning.

The schedule and resources estimations take into account realistic quantities for each component, considering the yield in all production steps until the final assembled detector. A breakdown of the yield model used at the main steps of the modules construction up to its assembly in the detector is shown in Table 15.11. This model, with an overall efficiency of 74%, was used to calculate the required total quantities of the main modules components for the final production, summarised in Table 15.12.

PBS		**Production step**	**Yield (%)**
8.4.1	Bare module hybridisation	Sensor UBM	99.5
		ASIC Bump deposition	99.5
		Flip-chip	97
8.4.3	Module assemblies	Flex module gluing	97
		Wire Bonding	98
		Test	98
		Burn-in tests	95
8.4.4	Detector Units	Loading on Detector Units	95
		Test	98
		Transport	99
8.6	Detector assembly and QA on surface	Assembly on cooling plates + integration	96
		Test	99
Overall yield			**74**

Table 15.11: Yield model of the various steps of the modules construction up to installation in the detector.

Main components		**Nominal**	**Pre-prod.**	**Production**	**Production comments**
8.1.1	LGAD Sensors	8032	543	10854	13 sensors/wafer
8.2.1	ASIC	16064	1358	27135	52 asic/wafer
8.4.2	Module Flex	8032	543	10854	
8.4.4	Modules	8032	543	10854	

Table 15.12: Estimated quantity of main deliverables needed to construct the HGTD modules. Numbers are indicated for nominal quantities, that is, for the actual items to be installed in the detector, pre-production of 5% and final production quantities. The pre-production and production numbers are corrected for the expected yield.

The schedule chart for the most critical items is presented in Figure 15.2. Separated schedule charts are given for each level 2 digits PBS in Figure 15.3 up to Figure 15.9.

When possible activities are taking place in parallel, as for example the LGAD Sensors (PBS 8.1.1) and ASIC (PBS 8.2.1), while there are many activities that depend on predecessors as, for example, module hybridisation (PBS 8.4.1) that needs available sensors and ASICs, as indicated in Figure 15.2 by the vertical lines.

A more detailed description is given below for the critical deliverables, that have important dependencies and may affect the final installation milestone.

Sensors (PBS 8.1) A breakdown of the sensors production milestones is presented in Figure 15.3. After the R&D and prototype phases, based on the understanding of the design issues solved, the sensor PDR will be submitted in Q1 2021, followed by a market survey and the FDR in Q4 2021. The sensor pre-production (5% of the production) will take place in the first half of 2022, followed by the final production from Q1 2023 to Q4 2024. The total number of working sensors to produce is 11397 (543 for pre-production and 10854 for the final production). This last number consists of the 8032 modules that are planned to be installed in the two HGTD end-caps, divided by the overall yield of 0.74 that is assumed to be relevant for the later processing and assembly steps, as detailed in Table 15.11. The exact QA strategy is still under development but it is assumed that all sensor vendors will deliver good/tested sensors, i.e. the yield of delivered sensors is assumed to be 100%.

The various vendors will deliver sensor wafers into batches which will be further processed for bump-bonding to the HGTD front-end ASIC (PBS 8.2.1): first a metal layer will be deposited on the pixel pads (under bump-metallization or UBM), and then the wafers will be diced and connected to the ASICs in a process known as bare module hybridization (PBS 8.4.1). This will most probably be done at a dedicated hybridization company, but it might also be done by the sensor vendor.

ASIC (PBS 8.2.1) A first full size ASIC (ALTIROC2) with 15x15 channels and all the functionalities is expected to be submitted at the end of 2020, after the SPR, to take place in Q3 2020. It should be followed by a second real size prototype in 2021, after the PDR, that is expected in Q4 2021, The pre-production of the final chip (ALTIROC-V1) is expected between Q4 2022 and Q3 2023, just after the FDR review. The PRR, expected in Q4 2023, should give the green light for the final chip production (ALTIROC-V2). The production, including the QA to be done by the Institutes is expected to take place between Q4 2023 and Q3 2024.

ASICs will be fabricated by the TSMC foundry under an existing frame contract negotiated by CERN. The chip procurement will be done through the frame contract. ASICs are expected to be received from the vendor as 8 inches diameter silicon wafer with 52 ASICs/wafer. The wafers have to be electronically tested before an Under Bump Metallization (UBM) process and a bump deposition is applied followed by a dicing and flip chip for the bump bonding to the sensors. 27135 ASICs are needed for the final production as detailed in Table 15.12, considering an estimated yield of 80% in the ASIC manufacture up to the delivery. Both at pre-production and production, some ASICs will be tested after dicing on dedicated boards for deep measurements and irradiation tests.

Bare module hybridisation (PBS 8.4.1) The baseline bump-bonding technology to connect each LGAD Sensor to two ALTIROC ASICs relies on solder bumps, to be done in Industry by integrated circuit packing companies. This activity relies on the availability of sensors and ASICs, and will be done in 3 steps. After the wafers of sensors and ASICs are processed, sensor and ASICs are diced and their interconnection by flip-chipping is applied, using only the pre-selected good ASICs. The pre-production (of approx. 5%) is expected from Q1 2023 to Q4 2023. The PRR will take place in Q4 2023, giving the green light to the bare module hybridisation of 525 8-inch wafers (27135 ASIC) and 522 sensors 6-inch wafers (10854 LGAD Sensors). This is expected to take place between Q1 2024 and Q1 2025.

Module assemblies (PBS 8.4.3) The module assembly consist in gluing the bare module (described in PBS 8.4.1) to the Module Flex, wire bonding the two ASICs and the HV connection of the module to the flex. The last step will be the QA tests. After the FDR, expected in Q2 2023, 543 modules will be constructed in the pre-production phase from Q3 2023 to Q1 2024. This step will be used to qualify the 4-5 Institutes/sites that will participate in this activity. The final production of 10854 modules, is expected to take place between Q3 2024 and Q3 2026. The overall production rate is expected to be approximately 19 modules/working day in the first half and 22 modules per working day in the second half of the production.

Detector Units (PBS 8.4.4) The production of 80 support units of 6 different types will be carried out in Industry and shipped to the Institutes that will do the modules loading. The modules are loaded on the support units, to form the inner, middle and outer disks. Dedicated flex tails will be used to connect the Module Flex connector and perform electrical tests before and after the positioning and gluing of the modules. This operation is done by the same Institutes that are doing the module assemblies (PBS 8.4.3), to minimize the transport and QA time. The pre-production will take place between Q2 2023 and Q1 2024. The PRR will be in Q2 2024, followed by the production, from Q2 2024 to Q3 2026.

Detector Assembly and QA on surface (PBS 8.6) The detector assembly and QA will be done at CERN in a clean room using dedicated tools for the detector assembly and testbench for QA. The main activities will be the assembly of the components (Detector Units, PEB, flex tails) on the cooling plates (PBS 8.6.3) and final integration inside the vessels (PBS 8.6.4). The final integration will be done with the participation of several Institutes between Q4 2024 and Q4 2025 for HGTD–A and between Q4 2025 and Q4 2026 for HGTD–C.

Installation and commissioning (PBS 8.7) The detector will be moved from the CERN clean room to the ATLAS cavern and installed on the two LAr end-cap cryostats using dedicated installation tools.

The Installation of the HGTD–A and HGTD–C are expected in LS3, respectively on 15 April - 20 May 2026 and 4 January - 3 February 2027. After the connection of each end-cap to the respective services an intense period of commissioning will start, while there is still access to the detector. In case of significant delays in the HGTD–C construction, the following scenarios are possible. The HGTD–C will be installed in January 2027 with all available instrumented half circular disks. The missing disk(s) may still be inserted in the following 1-2 months, during the overall ATLAS commissioning period. Although the crane will not be available anymore, enough space exists between the barrel and the end-cap calorimeters to allow the installation manually (objects of $\sim$ 35 Kg each and 1 m radius). A dedicated tool will be manufactured to transport the half instrumented disks safely without crane. The other possibility will be to install the missing instrumented disk(s) in the next YETS after LS3. This scenario will need a procedure to be developed respecting the ALARA/safety rules, to account properly for induced radiation levels.

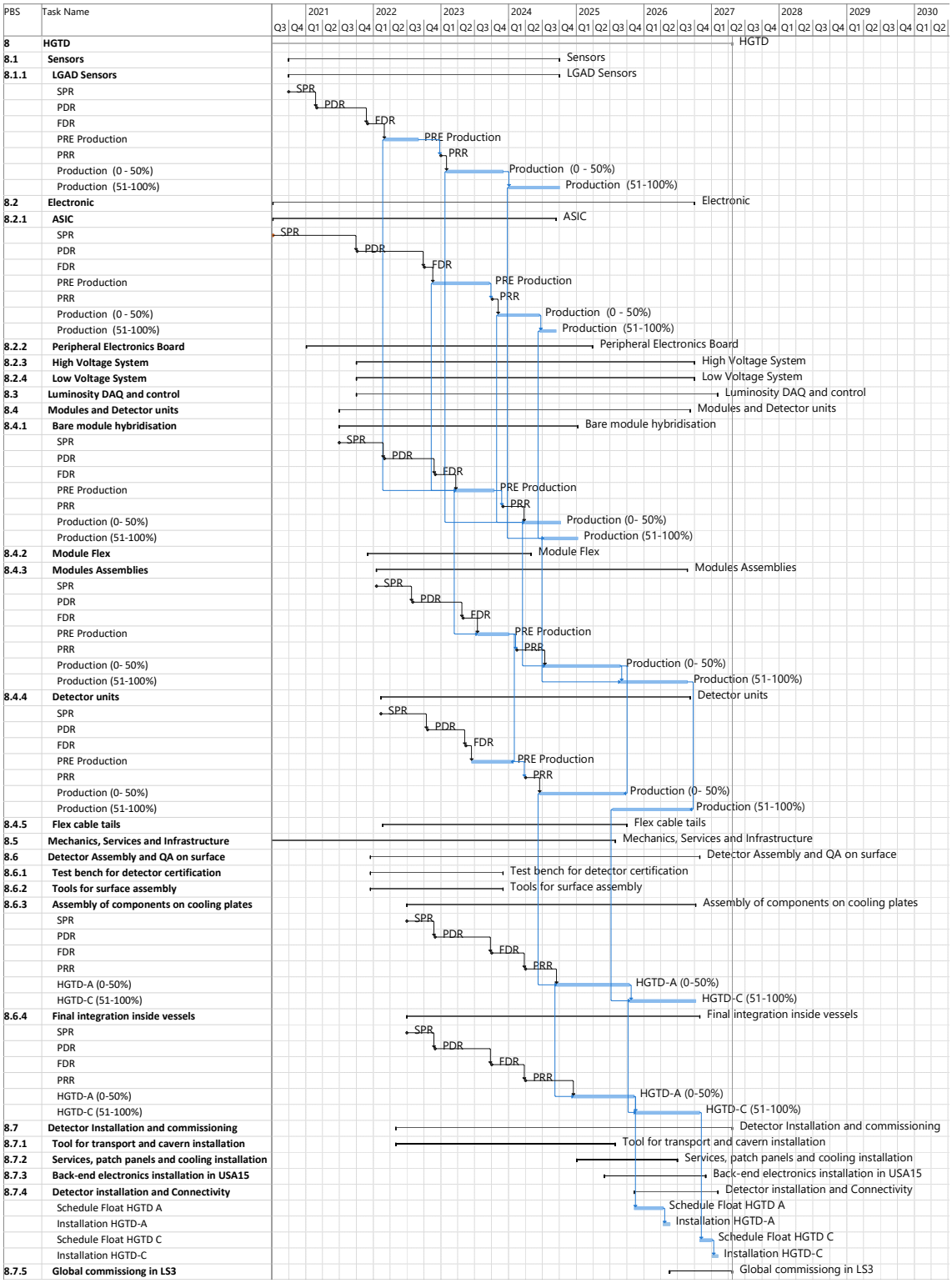


Figure 15.2: High-level schedule for the HGTD project including the planned reviews (PDR, FDR, PRR), pre-production and production for the most time critical components.

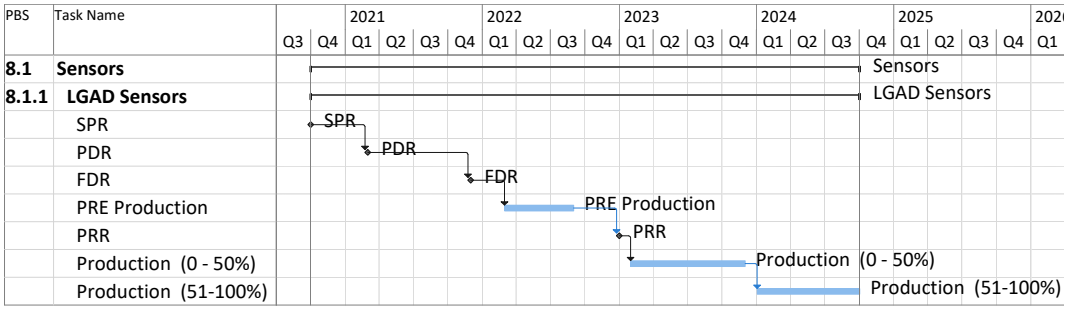


Figure 15.3: Sensors high-level schedule including the planned reviews (PDR, FDR, PRR), pre-production and production.

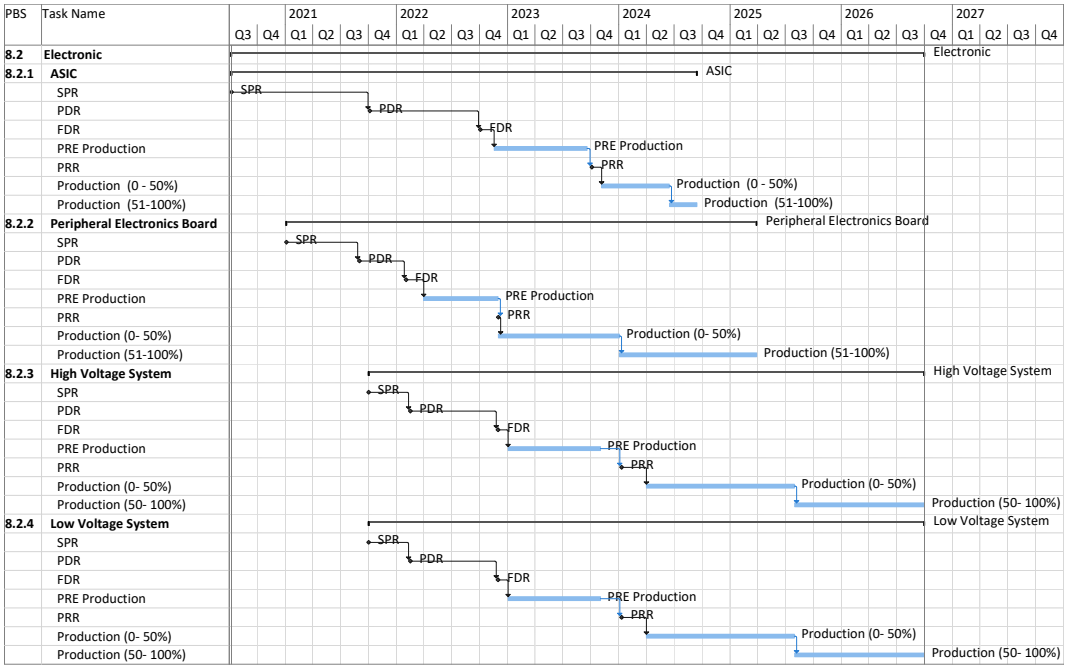


Figure 15.4: Electronics high-level schedule including the planned reviews (PDR, FDR, PRR), pre-production and production.

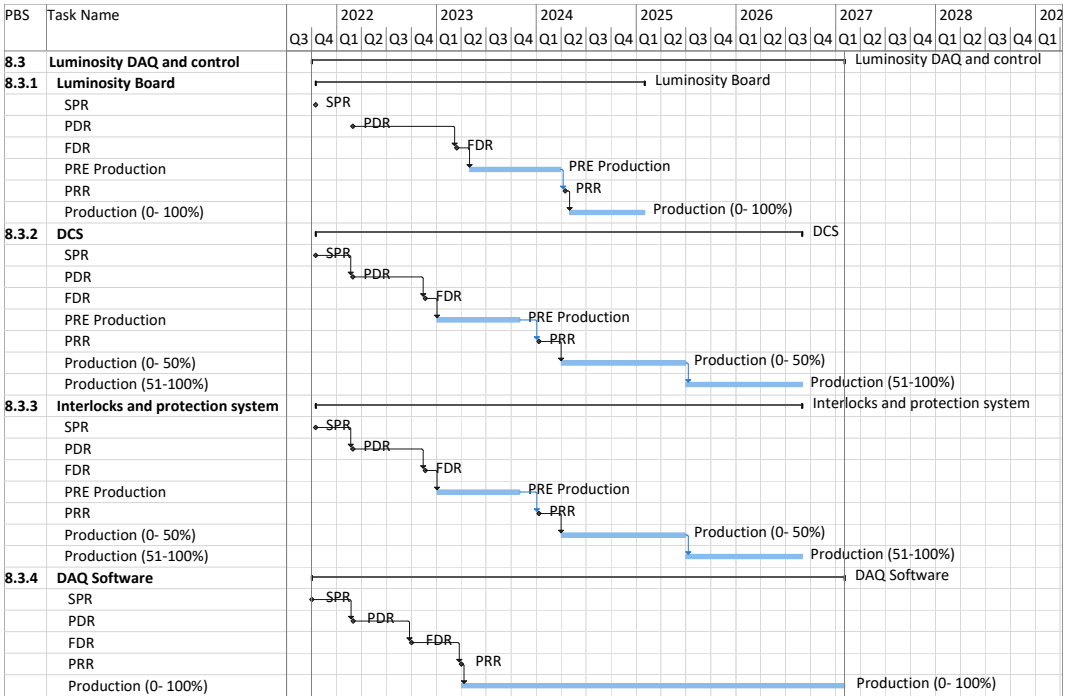


Figure 15.5: Luminosity, DAQ and Control high-level schedule including the planned reviews (PDR, FDR, PRR), pre-production and production.

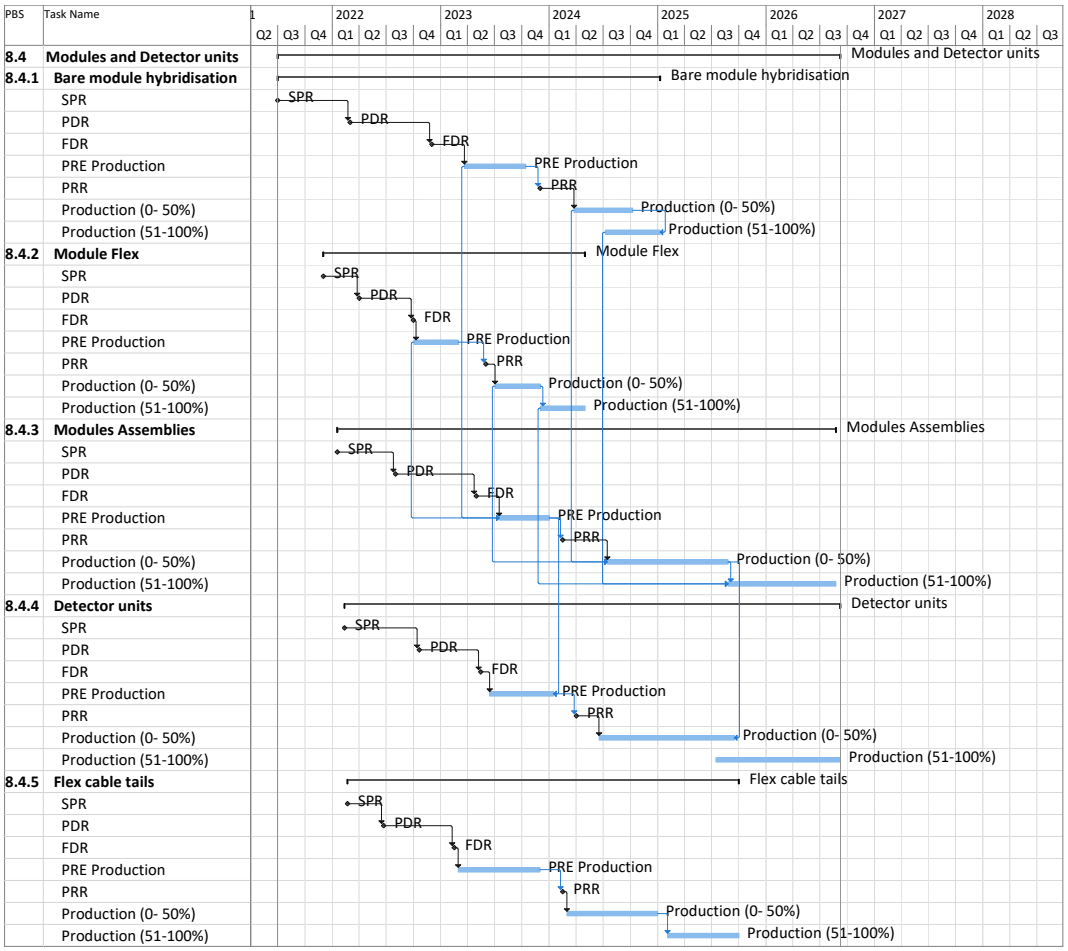


Figure 15.6: Modules and Detector Units high-level schedule including the planned reviews (PDR, FDR, PRR), pre-production and production. The pre-production and production of 8.4.1 Bare module hybridisation should be started after the pre-production and production of 8.1.1 LGAD Sensor and 8.2.1 ASIC are started, while the links are not shown in this figure.

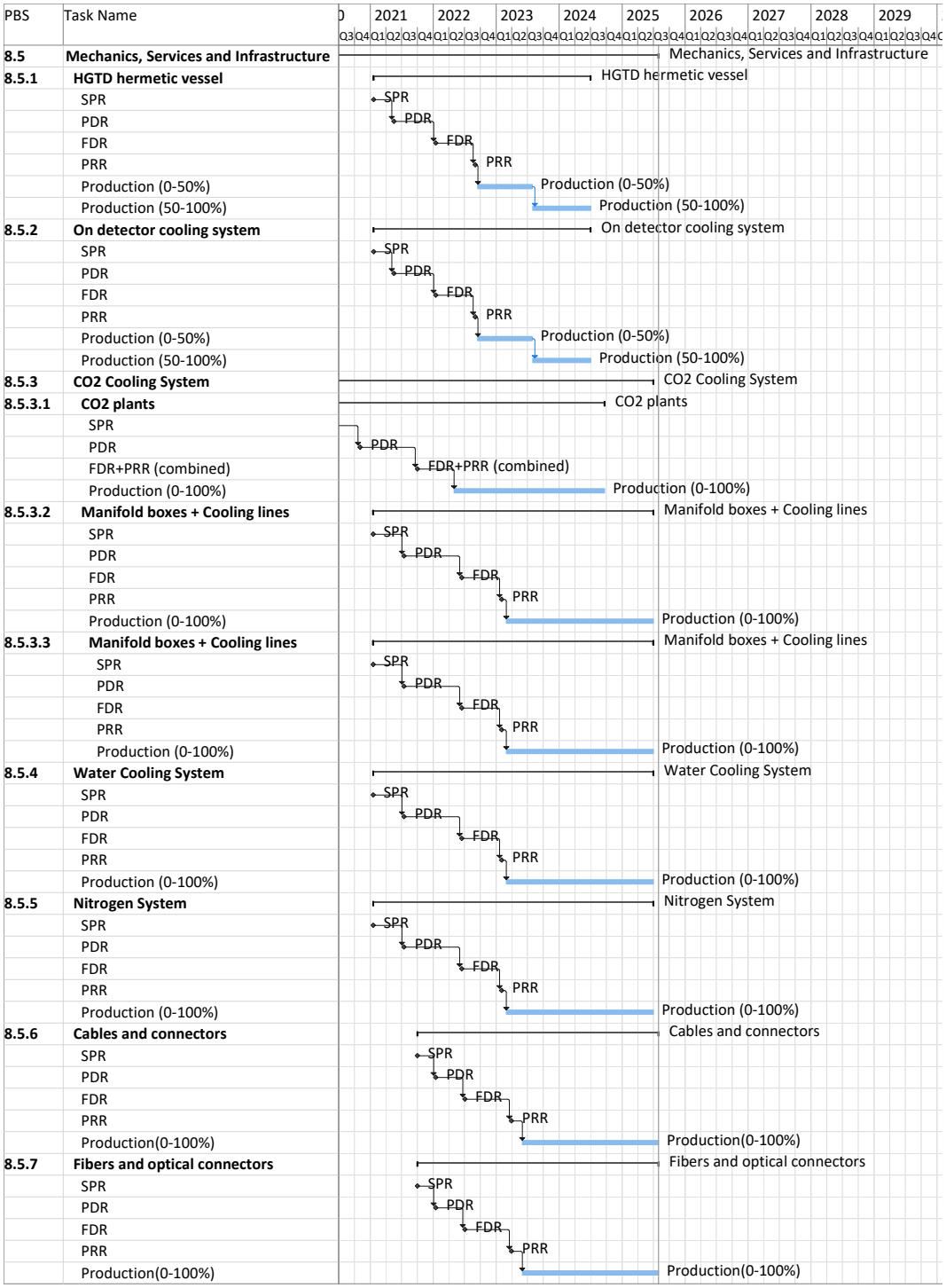


Figure 15.7: Mechanics, services and infrastructure high-level schedule including the planned reviews (PDR, FDR, PRR), pre-production and production.

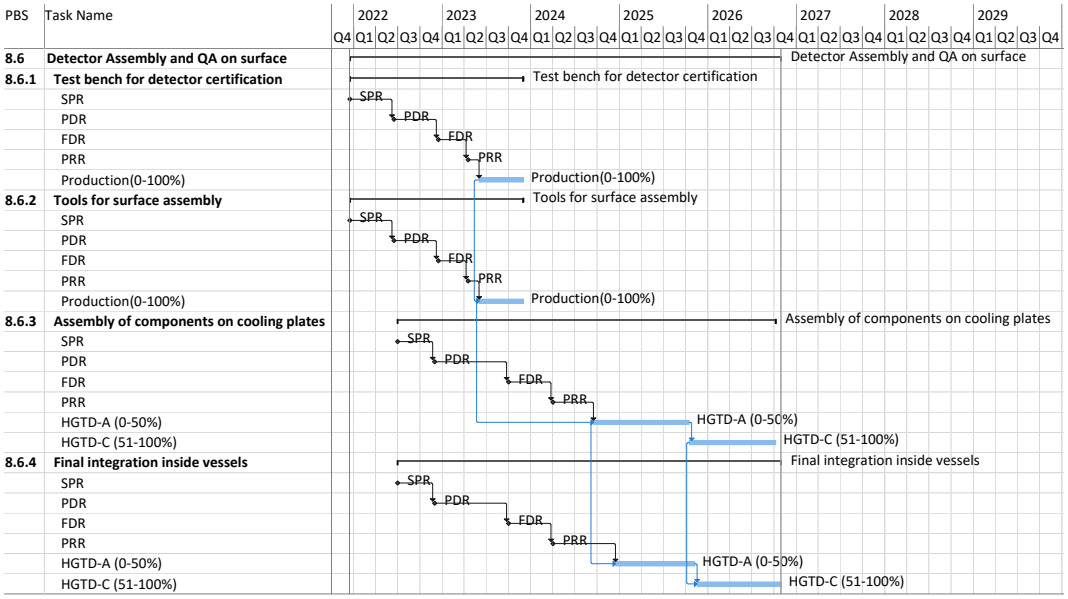


Figure 15.8: High-level schedule for the HGTD detector assembly and quality assurance on surface, including the planned reviews (PDR, FDR, PRR), pre-production and production. The production of 8.6.3 Assembly of components on cooling plates should be started after the production of 8.4.4 Detector Unit are started, while the links are not shown in this figure.

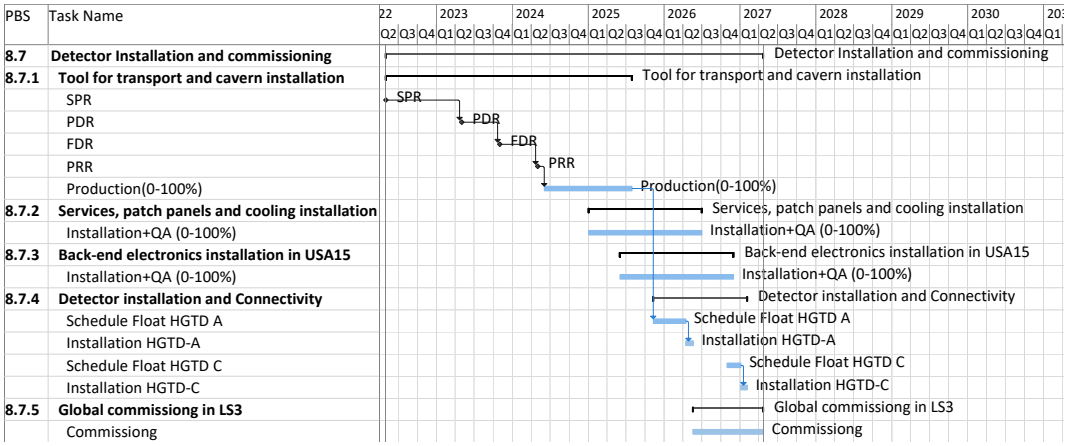


Figure 15.9: High-level schedule for the HGTD detector installation and commissioning, including the planned reviews (PDR, FDR, PRR), pre-production and production. The Schedule Float HGTD–A of 8.7.4 Detector installation and Connectivity should be started after the HGTD–A of 8.6.4 Final integration inside vessels are finished, while the link is not shown in this figure.

15.3 Resources

Many Institutes have already stated their intentions of contributing to the HGTD project, and many of them have already been working on R&D and design of various components of the system, as discussed in Section 15.1.

Surveys of the interests of Institutes and of the available resources have been performed recently. A preliminary sharing matrix showing the initial interest of the contributing Institutes toward the HGTD construction responsibilities is summarised in Table 15.13. The Institute surveys indicate that the human resources required for the HGTD project implementation are available for the full extent of the detector construction period.

The cost of the project is expected to be covered by the Institutions participating in the HGTD project, with their respective Funding Agencies. Most of the funds have been secured, and discussions among the Institutions and the Funding Agencies are ongoing in order to define the detailed list of deliverables, responsibilities, and the sharing of the project cost. These will be formalized after the project is approved, through an MoU document, insuring that all the aspects of the project are covered.

Deliverable	Institutes
8.1 Sensors	
8.1.1 LGAD Sensors	CERN, IFAE, IRFU, IHEP, JSI, USTC, USP, JINR
8.2 Electronics	
8.2.1 ASIC	IFAE, IHEP, IJCLab, JINR, LPC, OMEGA, USTC
8.2.2 Peripheral Electronics Board	AS, IHEP, JINR, LPC, NJU, NTHU, UIT, UH2C, UM5R, UMP, USP
8.2.3 High Voltage system	CERN, KTH
8.2.4 Low Voltage system	Giessen
8.3 Luminosity, DAQ and Control*	
8.3.1 Luminosity boards	KTH
8.3.2 DCS	Giessen, UIT, UH2C, UM5R, UMP
8.3.3 Interlocks and protection system	to be determined
8.3.4 DAQ software	IHEP, LPC
8.4 Modules and Detector Units	
8.4.1 Bare module hybridization	IFAE, IHEP, IRFU, SINANO, USTC
8.4.2 Module Flex	IHEP, Mainz
8.4.3 Modules Assemblies	IFAE, IHEP, IJCLab, IRFU, Mainz, SINANO, USTC
8.4.4 Detector Units	IFAE, IHEP, LPNHE, Mainz
8.4.5 Flex Cable Tails	IHEP, JSI, Mainz, UIT, UH2C, UM5R, UMP
8.5 Mechanics, Services & Infrustructure	
8.5.1 HGTD Hermetic vessel	IHEP, IJCLab
8.5.2 On detector cooling system	IJCLab, MEPhI
8.5.3 CO_2 cooling system	CERN, IRFU
8.5.4 Water cooling system	CERN
8.5.5 Nitrogen system	CERN
8.5.6 Cables and connectors	CERN
8.5.7 Fibers and optical connectors	AS, KTH, NTHU
8.6 Detector Assembly and QA	
8.6.1 Test bench for detector certification	IFAE, IHEP, USP
8.6.2 Tools for surface assembly	CERN, IJCLab, JINR
8.6.3 Assembly of components on cooling plates	CERN, IHEP, JINR, USP
8.6.4 Final integration inside vessels	CERN, IHEP, IJCLab, JINR, NJU
8.7 Installation and Commissioning	
8.7.1 Tools for transport and cavern installation	IJCLab, JINR
8.7.2 Services, p. panels and cooling installation	CERN, USP, JINR
8.7.3 Back-end elect. installation in USA15	CERN, IHEP, KTH, JINR, UIT, UH2C, UM5R, UMP
8.7.4 Detector installation and connectivity	CERN, IHEP, IJCLab, JINR, MEPhI, NJU, SJTU, USP, USTC
8.7.5 Global commissioning in LS3	CERN, IHEP, JINR, KTH, LPC, MEPhI, NJU, SJTU, USP, USTC

Table 15.13: Preliminary sharing matrix showing the initial interest of the contributing ATLAS Funding agencies and HGTD Institutes toward the construction responsibilities. **(*)** DAQ hardware deliverables are covered in the ATLAS TDAQ PBS, and they will be included in an amendment of the TDAQ MoU.

15.3.1 CORE Costs

Each HGTD PBS item, described in Table 15.3, has an associated CORE cost that is defined as the sum of the material value of all components making up the deliverable. The cost of generic infrastructure, prototypes, and spare components are all excluded by definition from the CORE costing, as is the cost related to personnel, such as labour or travel for personnel employed by HGTD Institutions. Specialized infrastructure, such as custom-designed tooling, is included in CORE. For items bought in industry, the material value is simply their selling price, and depending on how the vendor calculated this price, it includes some unknown fraction of labour cost at the company. This type of labour cost is included in CORE. The CORE cost of a project does not represent its full cost, and Institutions participating in HGTD have to request funds to cover both CORE and non-CORE expenditure from their Funding Agencies, in a ratio that varies from deliverable to deliverable.

The HGTD CORE cost has been estimated in a bottom-up approach and it has been calculated based on the baseline layout presented in this document. The yield model used accounts for failure and loss during the production phase, up to and including the installation of items in the ATLAS cavern, detailed for the main components in Table 15.12. The yield was estimated based on past production experience with similar or equivalent items or extrapolating from prototypes experience. In contrast, spares accounting for failure and loss during the operations phase, i.e. from the beginning of Run 4 onwards and components needed for the rings replacements are planned to be supported by maintenance and operations (M&O) funds and do not count as CORE.

The cost estimates for each item are quoted in CHF, using the exchange rates of:

- 1 Euro = 1.085 CHF
- 1 USD = 0.986 CHF
- 1 CNY = 0.1461 CHF
- 1 GBP = 1.246 CHF
- 100 JPY = 0.942 CHF
- 1 ILS = 0.2588 CHF

as in the other ATLAS phase II TDRs. The CORE estimates are based on existing contracts (ASICs), quotes from industry (sensors, FPGAs, DC-DC converters, Flex cables,...), extrapolation from other ATLAS Upgrade Phase-II projects with already signed MoU that are using the same or similar type of components (power supplies, cables, cooling station).

The price estimates for all the individual items are based on the most accurate information available at the time of the estimate, and they come with an uncertainty. The level of cost uncertainty of each item depends on the amount of technical development and design,

understanding of its procurement process, and the availability of vendor quotes. To describe the level of uncertainty of the cost, a quality factor ranging from 1 to 5 is used as summarised in Table 15.14. QF1 has the highest certainty and is based on a vendor quote for the final item or a catalogue price; QF5 has the least amount of certainty and is based on a rough estimate without any detailed design. Where items are composed of sub-items with different quality factors, a cost-weighted average quality factor is calculated.

Quality Factor	Description
QF1	Based on a vendor quote or catalogue price
QF2	Based on the purchase of a similar component
QF3	Based on an engineering design
QF4	Based on a conceptual design or scaled from similar systems
QF5	Based on a rough estimate without any detailed design

Table 15.14: Quality factor (QF) definitions. QF1 has the highest precision, QF5 has the highest uncertainty.

A summary of the HGTD CORE cost, detailed to the PBS level 3 is presented in Table 15.15, with a total of 9965 kCHF for the HGTD and 995 kCHF for the HGTD-DAQ related deliverables. These DAQ items are considered ATLAS TDAQ deliverables and they will be included in an amendment of the TDAQ MoU after the HGTD TDR approval.

The costs for the planned replacement of the HGTD inner and middle rings during the HL-LHC half life time are not accounted in CORE and should be accounted in the future (M&O) funds. DAQ hardware deliverables are covered in the ATLAS TDAQ PBS, and they will be included in an amendment of the TDAQ MoU.

The time profile of the CORE cost expenditures for the HGTD project, split into the main HGTD deliveries, is shown in Figure 15.10. The bulk of the expenditures will happen in 2022–2024, when most of the components need to be produced.

Basis of Estimate (BoE) documents, that describe in detail and justify the CORE cost estimate for each deliverable, have been produced. These BoEs are being reviewed in detail by the ATLAS Project Management Office in preparation of the UCG review. The HGTD cost estimate for all deliverables has an associated average cost quality factor of 2.2. The readout ASIC and LGAD Sensors, that are among the most expensive HGTD deliverables, have cost factors of 1.1 and 2.0, respectively. The CO_2 cooling system, the only other deliverable costing over 1000 kCHF has QF 2.9. The HGTD deliverables with highest QF are the Tools for surface assembly, transport and cavern installation. These have relatively low cost and are in the conceptual design stage given that they are needed later in the project. Engineering design of these tools will however accelerate in the near future.

The cost of the project is expected to be covered by the Institutes participating in the HGTD Phase-II UPR, with their respective Funding Agencies. The details of responsibility and

sharing among Institutes will be defined in an MoU to be prepared after the TDR approval.

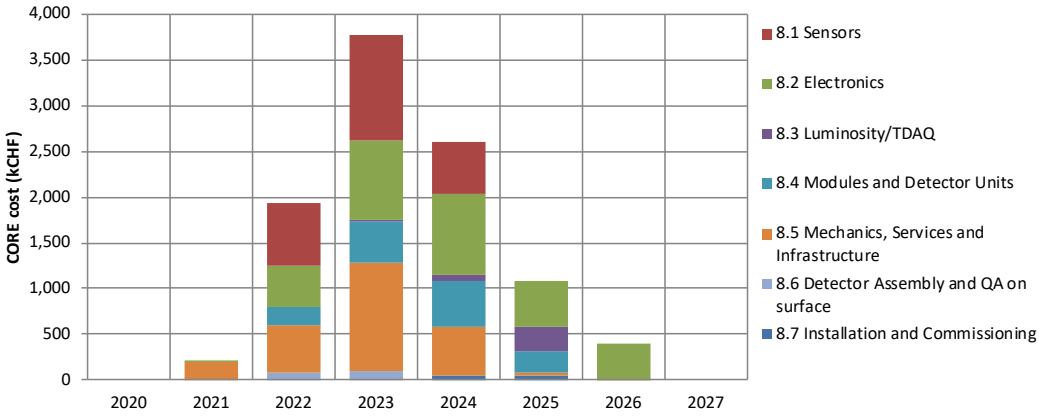


Figure 15.10: CORE cost time profile in kCHF for the HGTD level 2 deliverables from 2020 to 2027.

15.3.2 Required Labour Effort

Estimates of the human resources, in term of full-time equivalents FTE for the different type of labour (physicists, engineers, technicians, students), required to complete each deliverable have been obtained based on similar experience in other projects or sub-systems, such as the TDAQ, and ITk Pixel and Strip projects. The estimate used a bottom-up approach based on the detailed Work Breakdown Structure. Detailed summary information down to level 3 activities is being reviewed internally by ATLAS in preparation of the UCG review.

The labour effort needed to accomplish the construction of the detector up to the HGTD installation and commissioning is shown in Figure 15.11 as a function of time. The effort peaks, with a maximum of about 70 FTE, between 2023 and 2025 when most of the detector parts will be in the pre-production and production phases, with special emphasis to the module and Detector Units assembly. In total about 400 FTE are needed to accomplish the project, with a breakdown of approximately 20% physicists, 25% engineers, 25% technicians and 30% students. The survey mentioned earlier among participating Institutes indicates that the human resources required for the HGTD project, in all of these labor categories, is overall covered beyond the needs.

PBS/WBS	Item	CORE cost (kCHF)
8.1	**Sensors**	**2403**
8.1.1	LGAD Sensors	2403
8.2	**Electronics**	**3108**
8.2.1	ASIC	1094
8.2.2	Peripheral Electronics Board	638
8.2.3	High Voltage system	955
8.2.4	Low Voltage system	422
8.3	**Luminosity, DAQ and Control**	**339**
8.3.1	Luminosity boards	260
8.3.2	DCS	44
8.3.3	Interlocks and protection system	35
8.3.4	DAQ software	–
8.4	**Modules and Detector Units**	**1392**
8.4.1	Bare module hybridization	468
8.4.2	Module Flex	108
8.4.3	Modules assemblies	318
8.4.4	Detector Units	142
8.4.5	Flex Cable Tails	356
8.5	**Mechanics, Services and Infrastructure**	**2476**
8.5.1	HGTD Hermetic vessel	176
8.5.2	On detector cooling/support plate	190
8.5.3	CO_2 cooling system	1167
8.5.4	Water cooling system	23
8.5.5	Nitrogen system	20
8.5.6	Cables and electrical connectors	691
8.5.7	Fibers and optical connectors	209
8.6	**Detector Assembly and QA on surface**	**167**
8.6.1	Test bench for detector certification	72
8.6.2	Tools for surface assembly	95
8.6.3	Assembly of components on cooling plates	–
8.6.4	Final integration inside vessels	–
8.7	**Installation and Commissioning**	**80**
8.7.1	Tools for transport and cavern installation	80
8.7.2	Services, patch panels and cooling installation	–
8.7.3	Back-end electronics installation in USA15	–
8.7.4	Detector installation and connectivity	–
8.7.5	Global commissioning in LS3	–
Total HGTD		**9965**
DAQ(*)	Felix boards+LTI boards, emulator,...	**995**
Total w/ DAQ		**10960**

Table 15.15: Estimated CORE cost of the HGTD (in kCHF). The total cost is given with and without the costs of the DAQ. The internal and external moderator CORE costs are accounted in the ATLAS ITk common. The items listed without a cost are WBS-only items, and hence have no assigned CORE cost. **(*)** DAQ hardware related costs, estimated separately by TDAQ, will be included in an amendment of the TDAQ MoU.

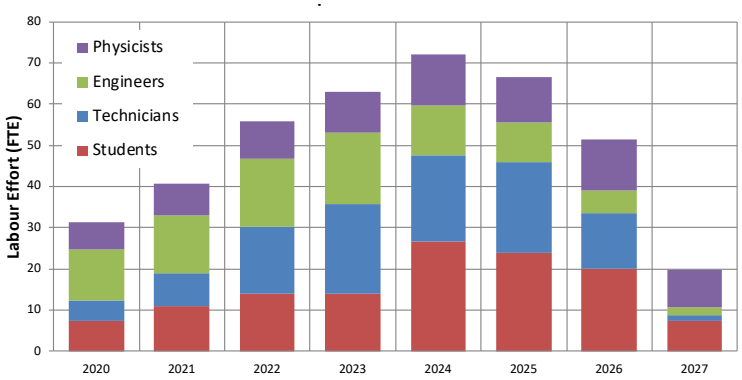


Figure 15.11: Required effort (in FTE) needed per labour type (physicists, engineers, students, technicians) over the lifetime of the project.

15.4 Risk Analysis and Risk Mitigation

The HGTD project is a complex undertaking on the part of multiple Institutes, that is prone to internal and external uncertainties, the consequences of which are known as risks. These risks cannot be avoided but they must be managed. The risks associated with cost, schedule, and scope and technical performances issues in the HGTD project are managed by a structured and integrated process as defined in the HGTD Risk Management Plan (RMP). The HGTD project overall risk management process follows broadly accepted risk management standards [102], according to which, awareness of potential risks and a deliberate approach meant to prevent risks or accept and mitigate them, are key to successful risk management. The RMP defines the roles and responsibilities of the management in the process of monitoring and controlling risks throughout the project and the thresholds used to characterise risk probability and impact.

The design and construction of the HGTD Project is well within the experience and expertise of the collaborators, technical staff and physicists, who are participating. Every effort has been made to specify the project in a manner that reduces the risk to an acceptably low level. The technical risks to the project that are identified will be addressed as early as possible to assure that they do not negatively impact the timely completion of the project or stress its budget. Proactive risk identification and mitigation can therefore significantly reduce the probability of unexpected events that could require contingency and/or additional time to resolve.

15.4.1 Risk Management Process and Plan

The risk management process is an integral part of the HGTD project management, as it informs decision making at every stage. Every effort has been made to design and specify all sub-projects to reduce the risk to an acceptable low level. Risks in the HGTD project are managed by a structured and integrated process for identifying, evaluating, tracking, mitigating, and managing project risks in terms of three risk categories: cost, schedule, and scope/technical performance.

The HGTD UPLs have the ultimate responsibility for managing the project risk. The HGTD UPLs are assisted in this role by the Resources and Risk Coordinator. Similar to the HGTD UPL, the HGTD sub-project coordinators (LV2) have the ultimate responsibility to manage and oversee the risks associated with their respective WBS areas. They report to the HGTD UPLs and the Resources and Risk Coordinator. The UPL is also responsible to ensure that the respective sub-system coordinators execute the appropriate mitigation strategies to minimise the likelihood of the risk. Risks are reviewed in periodical meetings of the HGTD Coordinators in which risks are discussed, updated, and appropriate actions are taken if required.

The overall Risk Management approach consists of a five-step process:

- Identifying potential project risks,
- Analysing project risks,
- Planning risk mitigation strategies,
- Executing risk mitigation strategies, and
- Monitoring and tracking the results, revising the risk mitigation strategies as necessary.

This includes identifying appropriate risk mitigation strategies to lower the likelihood of the risk to occur and quantifying the severity of that risk. The sub-project coordinators report the potential risk to the UPLs and Risk Coordinator. It is subsequently the UPLs responsibility to approve the risks and associated actions and update the risk register. It is also the UPL responsibility to identify additional risks, including global risks that span across multiple WBS areas.

The identified risks, the associated mitigation strategy, and the response in the event that the risks were to materialize are all captured in the HGTD Risk Register. The Risk Register contains, for each risk, the following information:

- The mitigation steps that are/will be taken to minimize that risk from occurring;
- The response to the risk in the eventuality that the risk materialises;

- The quantitative impact of the risk on Cost, Schedule and Performance: The risk impact is classified as either Negligible, Low, Medium or High as shown in Table 15.16. The performance impact is based on identifying a set of Key Performance Parameters (KPP) and assessing the impact of the risks on those KPP. This risk analyses is performed by the respective sub-project coordinators and reviewed and maintained by the UPL;
- The likelihood that the identified risk will occur. The Risk Probability is identified as Low, Moderately Low, Moderately High and High based on the probability range shown in Table 15.17.
- The likely impact on the cost and schedule, and optimistic/pessimistic scenario quantifying the best/worst case scenarios.
- The overall assessment of the risk based on Table 15.17 that correlates the probability of the risk to occur and the impact of the risk. The overall risk is classified as High, Medium, or Low based on the product of the risk impact and the risk probability.

Risk Impact	**Cost impact (kCHF)**	**Schedule Impact (months)**	**Scope/performance Impact (KPP)**
Negligible	$0-20$	$0-1$	degraded $< 1\%$
Low	$20-100$	$1-3$	degraded $1-5\%$
Medium	$100-500$	$3-6$	degraded $5-10\%$
High	$500-\infty$	$6-\infty$	degraded $> 10\%$

Table 15.16: Classification of the Risk Impact based on its impact on the cost, the schedule, and the scope/performance. The last column reflect the assessment for Scope/Performance Impact in terms of the impact on the objective Key Performance Parameters (KPP).

Risk Probability	**Risk Impact**			
	Negligible	**Low**	**Medium**	**High**
$30\% < p \leq 100\%$	Low	Medium	High	High
$15\% < p \leq 30\%$	Low	Low	Medium	High
$5\% < p \leq 15\%$	Low	Low	Medium	High
$p \leq 5\%$	Low	Low	Low	Medium

Table 15.17: Correlation of Risk Probability (first column) and Risk Impact (remaining columns, classified as Negligible, Low, Medium and High). Considering this correlations, the identified risk is subsequently classified as Low, Medium, or High.

15.4.2 Major risks and mitigation strategies

Many of the technical choices in the HGTD project were made already at the time of the Expression of Interest and Technical Proposal for the best compromise between performance

and cost. Some have since then evolved motivated by the minimization of risks. The severe space constraints (in z, r), high radiation levels and the limited time available to implement the project in the LS3 shutdown have been seriously considered in the optimized layout presented in this TDR. Some major risks that have been identified and addressed in the process of defining the scope of the HGTD project are discussed below.

Deep ASIC characterisation and integration with sensor: The schedule foresees two iterations of the full-size ASIC during the prototype phase (2020-2021). This ASIC is quite complex and challenging, hence demanding a deep characterisation between each iteration, including significant irradiation tests and a characterisation of the ASIC connected to the sensor. If significant problems arise, there is some risk that testing will require more time. The probability of such a risk is estimated to be 20 %. To mitigate it, intensive post layout simulations will be performed before submission to minimize the problems to be debugged when testing the ASIC. In addition additional teams will be trained and injected in the ASIC testing phase (already started recently), which could reduce this probability. The risk impact is mainly on the schedule, about 6-9 months, depending if it emerges after only the first prototype or both.

Detector radiation performance worse than expected: The radiation hardness of the LGAD Sensors and ALTIROC2 chips will only be fully evaluated after the respective pre-productions and the hybridization pre-production. The time is limited to react if, at this stage, the final modules underperformed in terms of radiation hardness. This might result in a detector with lower life expectancy that originally planned, specially for the inner ring. However, this scenario should have a limited impact on the overall schedule and cost. On the other hand, it may result in the need to replace the inner ring at an earlier stage than expected. If this were the case, intense R&D (with other dopant materials or bulk implantations for example) should be pursued to develop a better solution for the first replacement option. The fact that the target is to qualify and produce sensors in several fabrication sites would also help to mitigate lower performance of sensors produced in one site, or the possible low sensor yield after pre-production of one of the facilities.

LGAD Sensor procurement: Sensor production facilities might struggle to cope with the simultaneous requests for LGAD Sensors from ATLAS and CMS. Some vendors might even be devoted to ITk related productions or other experiment's needs. This situation will likely result in inevitable delays in the production schedule, since fabrication sites typically report a best case scenario for their delivery time and do not account for other future (possible) demands. Again, a clear mitigation factor would be the fact that the HGTD production will not rely on a single supplier, but at least in three, with whichever two sites, capable of producing the needed amount of sensors with minimal impact on the schedule. If it were the case that CMS and ATLAS both select the same vendors it may be an advantageous to be slightly ahead in the schedule and thus try to maintain a certain priority with respect to later productions.

Periphery Electronic Boards schedule and design: Duo to the limitation of the surface area for all components, the placement of the connectors and DC/DC blocks and the selection of the package for capacitors need to be optimized. Some prototype DC/DC blocks can be made to evaluate the physical dimension, the power efficiency, anti-magnetic ability and radiation hardness. While the delay could be absorbed in the schedule float, the precise impact on performances need to be assessed. The lpGBT, VL+ and MUX may not be available when planned. Prototype Periphery Electronic Boards can be made with commercial devices as placeholders, and there is float in the schedule to absorb an additional year of delay.

Module production rate: The module assembly throughput is currently based on the assumption that four assembly sites will produce modules at a rate of about 20 modules per week and that the overall yield will be better than 74%. The module assembly is an activity that extends for almost two years and thus any delay can have a significant impact on the schedule. A lower assembly yield than expected would impact cost and schedule and thus it is critical to achieve or improve the target yield. The fact that the module hybridization relies on matured and well understood processes that are commonly available in the industry gives confidence that this critical step is under control, and in any case, other companies will be approached and qualified to have options towards the final production. Problems with assembly in Institutes can be mitigated by increasing the number of assembly sites (an option that will be actively pursued and can turn into an opportunity) and by benefiting from the current R&D on more robust module concepts that could simplify the assembly process, such as using ball bonding instead of wire-binding.

Uniform clock distribution: The master clock will be distributed from FELIX to the lpGBT downlinks and then to the ASICs, in which a clock tree will be used to ensure the uniformity of the clock. Different contributions to the clock distribution can affect considerably the time resolution and thus having an impact on HGTD performance. Static contributions include the propagation times to distribute the clock to each ASIC while dynamic contributions can occur through a variety of mechanisms across a wide range of frequencies including high-frequency jitter. These contributions have been studied in Section 10.2 and a mitigation strategy has been shown. It will consist of computing the average the time of arrival per ASIC at L0 trigger rate and then apply this correction offline. Although conservative contributions to the clock jitter coming from FELIX, lpGBT, flex and ASIC have been taken into account in these studies; unknown or noise induced jitter sources with an irreducible clock jitter $>$ 30ps will compromise the time resolution of the detector. The mitigation plan will include the measurement of the jitter performance at different points (FELIX, lpGBT, flex and ASIC) during pre-production in a dedicated test-bench, where the different contribution to the clock jitter can be identified. In case that any unexpected jitter contribution arises, the clock distribution might be revisited with additional clock cleaner impacting the

design of different components like the PEB and flex, and the production of these items might be delayed by a few months.

Increased radiation levels: An increase of the expected radiation levels, for instance caused by further increase in the amount of ITk services in the patch panel PP1 region, can impact the HGTD performance. To mitigate this, the transition radius between the inner ring (to be replaced at each 1000 fb^{-1}) and the middle ring (to be replaced at 2000 fb^{-1}), currently at r =230 mm, could be increased together with the inner radius of the permanent outer ring, currently at r =540 mm.

Schedule slippage for HGTD–C: The schedule float for the installation of the HGTD–C in ATLAS is short. The schedule critical path is driven by the module production rate mentioned above. Modules for half of HGTD–C will take about seven months to produce. In case of delays in the construction, in spite of the mitigation measures to the module production schedule listed above, the HGTD–C will be installed in January 2027 with all available instrumented half circular disks. The missing disk(s) may still be inserted in the following 1-2 months, during the overall ATLAS commissioning period. Although the crane will not be available anymore, enough space exists between the barrel and the end-cap calorimeters to allow the installation manually (objects of $\sim$ 35 Kg each and 1 m radius). A dedicated tool will be manufactured to transport the half instrumented disks safely without crane. The other possibility will be to install the missing instrumented disk(s) in the next YETS after LS3. This scenario will need a procedure to be developed respecting the ALARA/safety rules, to account properly for induced radiation levels. The impact in the physics performance should be small given expected lower values of instantaneous luminosity at the startup of HL-LHC, compared to the designed peak luminosity.

Acknowledgements

We thank CERN for the very successful operation of the LHC, as well as the support staff from our institutions without whom ATLAS could not be operated efficiently.

We acknowledge the support of ANPCyT, Argentina; YerPhI, Armenia; ARC, Australia; BMWFW and FWF, Austria; ANAS, Azerbaijan; SSTC, Belarus; CNPq and FAPESP, Brazil; NSERC, NRC and CFI, Canada; CERN; CONICYT, Chile; CAS, MOST and NSFC, China; COLCIENCIAS, Colombia; MSMT CR, MPO CR and VSC CR, Czech Republic; DNRF and DNSRC, Denmark; IN2P3-CNRS and CEA-DRF/IRFU, France; SRNSFG, Georgia; BMBF, HGF and MPG, Germany; GSRT, Greece; RGC and Hong Kong SAR, China; ISF and Benoziyo Center, Israel; INFN, Italy; MEXT and JSPS, Japan; CNRST, Morocco; NWO, Netherlands; RCN, Norway; MNiSW and NCN, Poland; FCT, Portugal; MNE/IFA, Romania; MES of Russia and NRC KI, Russia Federation; JINR; MESTD, Serbia; MSSR, Slovakia; ARRS and MIZŠ, Slovenia; DST/NRF, South Africa; MINECO, Spain; SRC and Wallenberg Foundation, Sweden; SERI, SNSF and Cantons of Bern and Geneva, Switzerland; MOST, Taiwan; TAEK, Turkey; STFC, United Kingdom; DOE and NSF, United States of America. In addition, individual groups and members have received support from BCKDF, CANARIE, Compute Canada and CRC, Canada; ERC, ERDF, Horizon 2020, Marie Skłodowska-Curie Actions and COST, European Union; Investissements d'Avenir Labex, Investissements d'Avenir Idex and ANR, France; DFG and AvH Foundation, Germany; Herakleitos, Thales and Aristeia programmes co-financed by EU-ESF and the Greek NSRF, Greece; BSF-NSF and GIF, Israel; CERCA Programme Generalitat de Catalunya and PROMETEO Programme Generalitat Valenciana, Spain; Göran Gustafssons Stiftelse, Sweden; The Royal Society and Leverhulme Trust, United Kingdom.

The crucial computing support from all WLCG partners is acknowledged gratefully, in particular from CERN, the ATLAS Tier-1 facilities at TRIUMF (Canada), NDGF (Denmark, Norway, Sweden), CC-IN2P3 (France), KIT/GridKA (Germany), INFN-CNAF (Italy), NL-T1 (Netherlands), PIC (Spain), ASGC (Taiwan), RAL (UK) and BNL (USA), the Tier-2 facilities worldwide and large non-WLCG resource providers. Major contributors of computing resources are listed in Ref. [103].

A Expected Energy Spectra

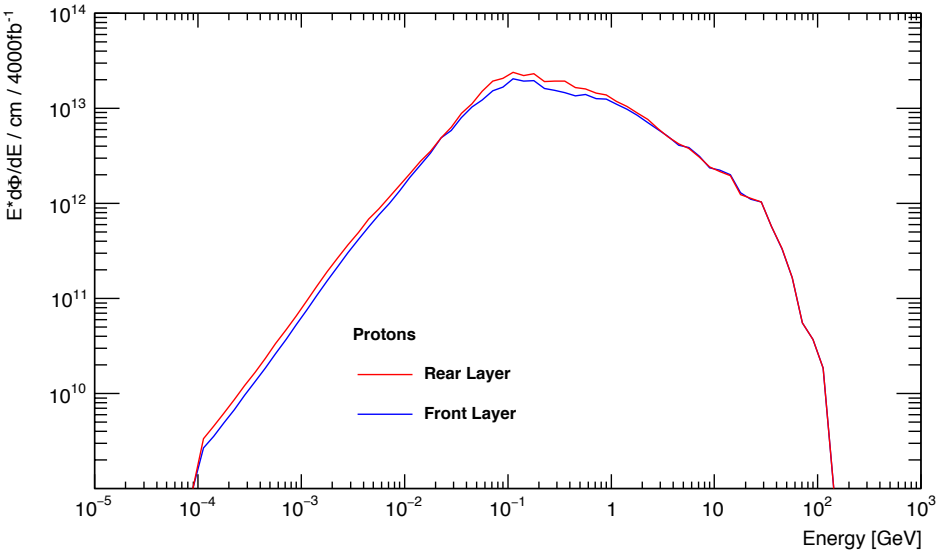


Figure A.1: Proton spectra averaged over the rear (outermost) and front (innermost) silicon layer of the HGTD. The uncertainties are of the order of 5%

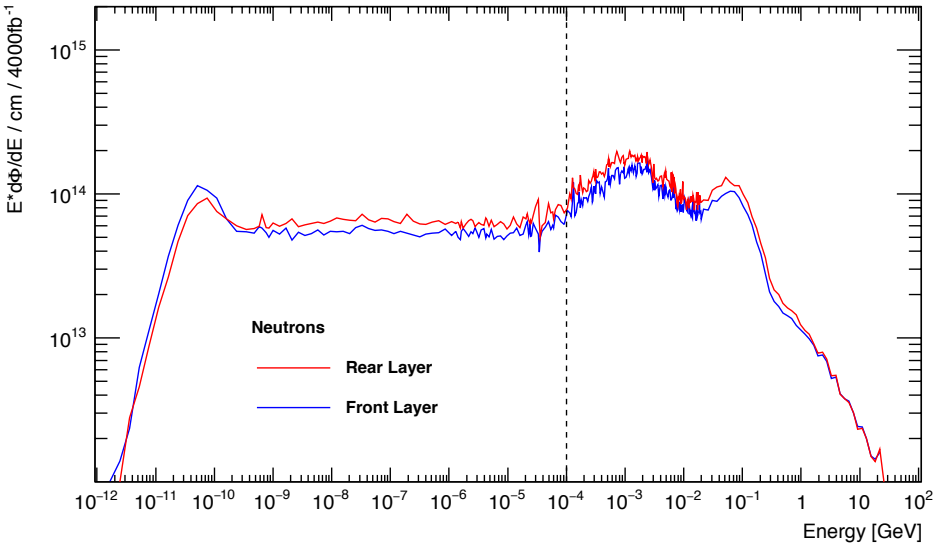


Figure A.2: Neutron spectra averaged over the rear (outermost) and front (innermost) silicon layer of the HGTD. The uncertainties are of the order of 5%. The fluctuations between 1 keV and 10 MeV are due to resonance.

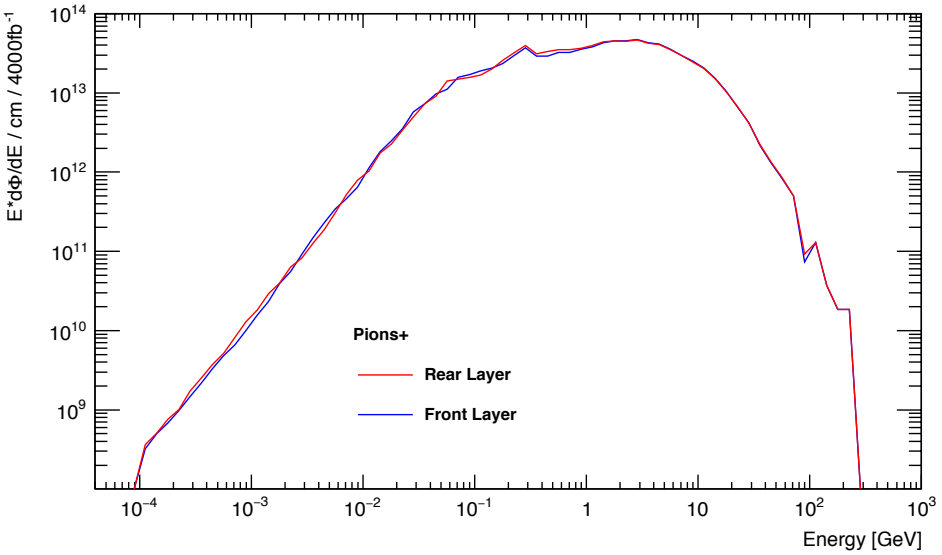


Figure A.3: Pion spectra averaged over the rear (outermost) and front (innermost) silicon layer of the HGTD. The uncertainties are of the order of 5%

B Monte Carlo samples

The simulation, digitisation and reconstruction was implemented in the ATLAS upgrade software releases[1]. Samples of single electrons, muons and pions as well as selected physics processes such as $t\bar{t}$, VBF $H \to Z(\nu\nu)Z(\nu\nu)$ were produced using the ATLAS production system. PYTHIA8 [25] was used together with POWHEG [20–22] for most of the samples. In the simulation step the beam spot was simulated with the nominal spread in z and time described in Section 2.1. Samples with $\langle\mu\rangle = 0$ as well as $\langle\mu\rangle = 200$ were processed. A summary of the samples is shown in Table B.1. For the minimum bias (inelastic collisions in the underlying event) samples, single neutrino events were produced to mimic the event topology at $\langle\mu\rangle = 200$.

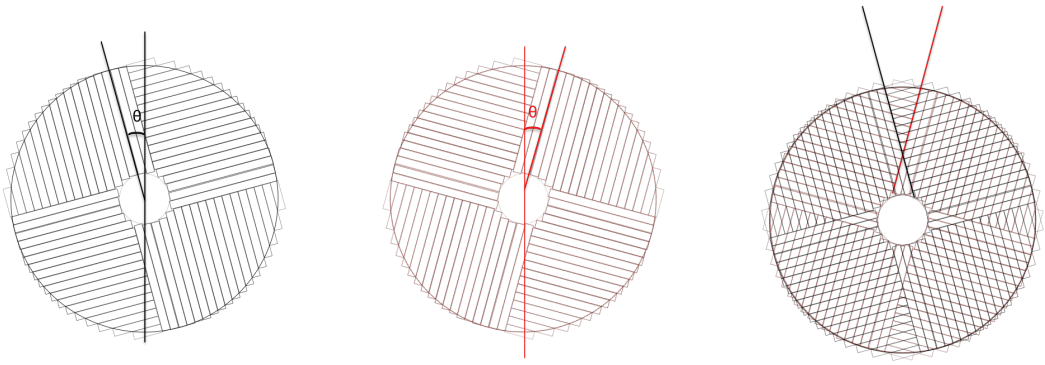
[1] The releases used were AtlasProduction-20.20.14.4 for simulation and 20.20.14.6 for digitisation and reconstruction, both including the so-called Step 3.1 layout of ITk used for results discussed with LHCC in Nov 2019 (geometry tag ATLAS-P2-ITK-17-04-02).

Sample		Number of events	
Single particles		$\langle\mu\rangle = 0$	$\langle\mu\rangle = 200$
π^+, $p_T = 5$ GeV, flat η [2.3-4.3]		200000	200000
π^+, $p_T = 20$ GeV,		200000	200000
π^+, $p_T = 45$ GeV,		2000000	2000000
π^+, flat p_T [0.1-5.0] GeV, flat η [2.3-4.1]		200000	200000
π^0, flat p_T [0.1-5.0] GeV. flat η [2.3-4.1]		200000	200000
γ, $p_T = 20$ GeV, flat η [2.3-4.3]		200000	200000
γ, $p_T = 45$ GeV, flat η [2.3-3.2]		200000	200000
γ, $p_T = 100$ GeV, flat η [2.3-3.2]		50000	50000
μ, $p_T = 45$ GeV, flat θ		400000	400000
μ, $p_T = 45$ GeV, flat η [2.3-3.2]		300000	300000
μ, $p_T = 45$ GeV, flat η [3.2-4.3]		100000	100000
e, $p_T = 45$ GeV, flat η [2.3-4.3]		400000	400000
e, $p_T = 20$ GeV, flat η [2.3-4.3]		200000	200000
ν, for minimum-bias at $\langle\mu\rangle = 200$		-	1000000
Physics processes	Generator		
Minimum bias, low-p_T	PYTHIA8	10000000	-
Minimum bias, high-p_T	PYTHIA8	1000000	-
Dijet, 20 GeV $< \hat{p_T} <$ 60 GeV	PYTHIA8	1000000	1000000
Dijet, 60 GeV $< \hat{p_T} <$ 160 GeV	PYTHIA8	1000000	1000000
Dijet, 160 GeV $< \hat{p_T} <$ 400 GeV	PYTHIA8	1000000	1000000
$Z \to ee$	POWHEG+PYTHIA8	100000	100000
$Z \to \tau\tau$	POWHEG+PYTHIA8	400000	400000
$t\bar{t}$	Powheg+Pythia8	1000000	1000000
VBF $H \to ZZ \to 4\nu$	POWHEG+PYTHIA8	500000	500000
NCB beam–gas, oxygen		400000	
NCB beam–gas, carbon		400000	
NCB beam–gas, hydrogen		400000	

Table B.1: The simulated Monte Carlo samples used for the studies in this document.

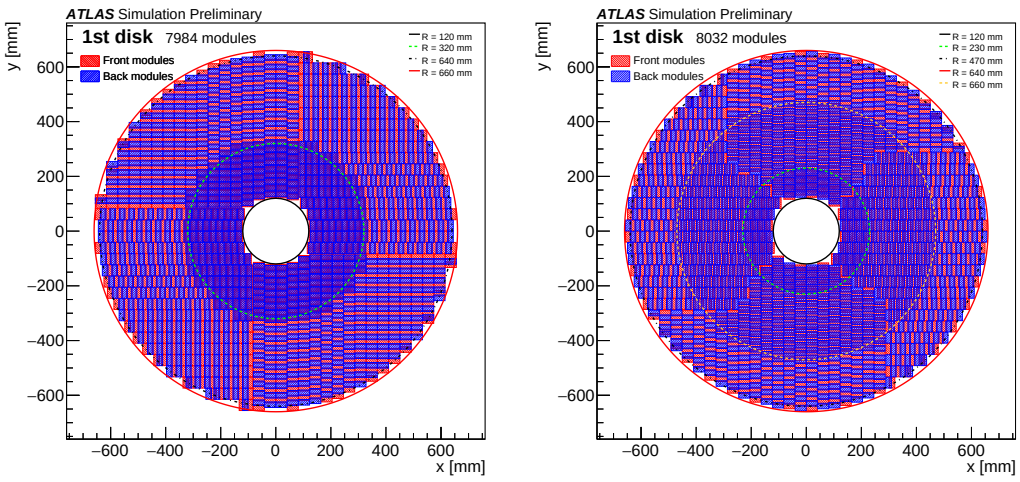
C Two-ring Layout Used in Simulation

The relevant drawings concerning the two-ring detector layout used in the simulation.



(a) First layer (b) Second layer (c) Overlay

Figure C.1: The orientation of the readout rows for the first and second layer encountered by a particle, separately and with the overlay of both. Each layer is rotated in alternating directions by 20°. This can be compared to Figure 2.8 for the three-ring layout.



(a) Two-ring layout (b) Three-ring layout

Figure C.2: Figures showing the placement of the modules in the (a) two-ring and (b) three-ring layouts.

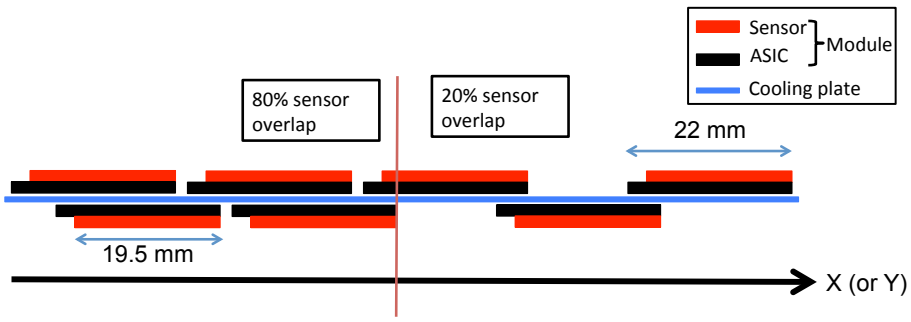


Figure C.3: The schematic drawing shows the overlap between the modules on the front and back of the cooling disk. There is a sensor overlap of 80% between sensors on front and back sides of a cooling plate at $R < 320$ mm, and 20% outside this radius. The figure can be compared to Figure 2.9.

D Technical Drawings

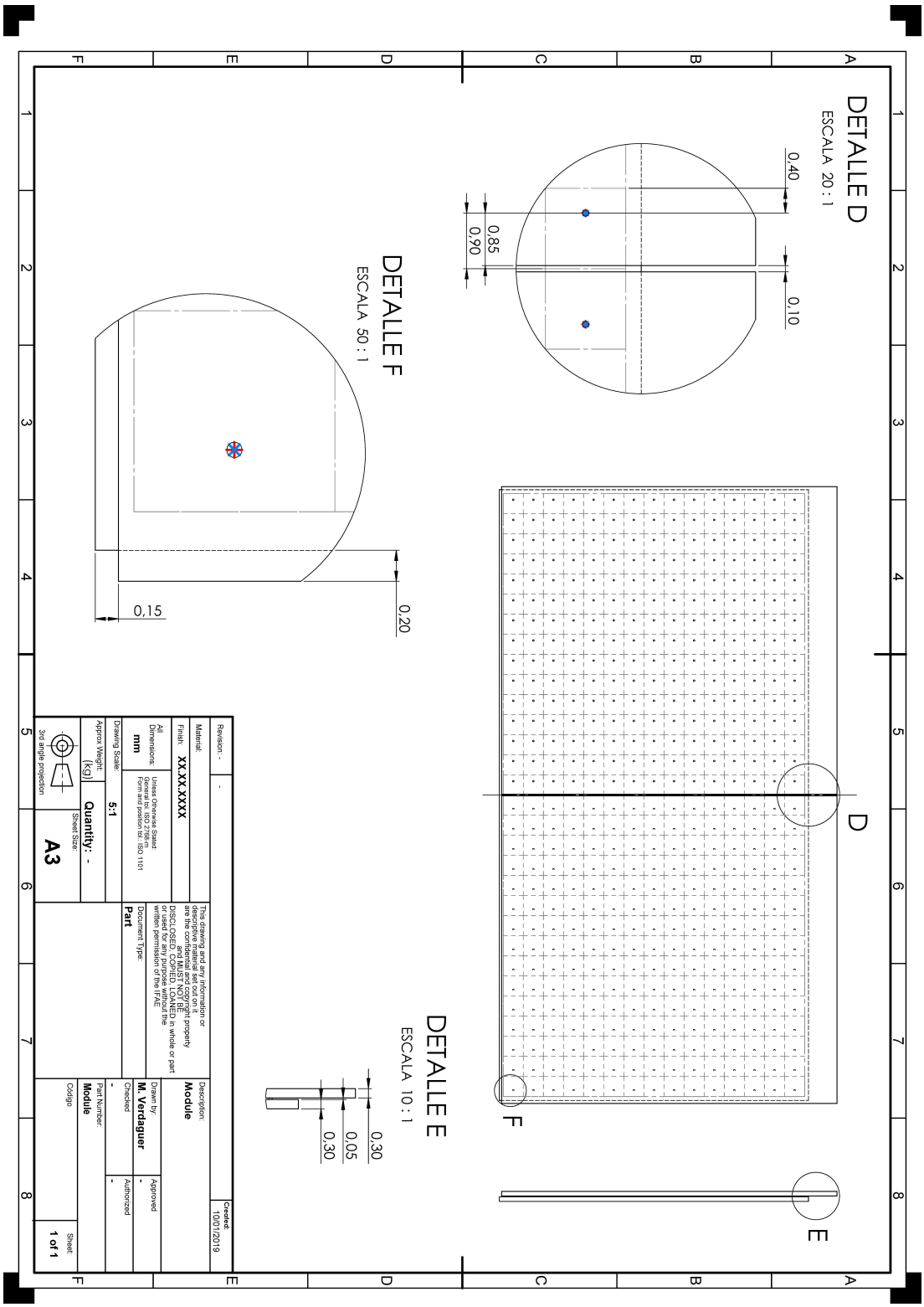


Figure D.1: Sketch of the bare module (sensor and ASIC). Distances are in millimetres. The bump pads on the sensor are shifted by 250 µm on each side of the sensor, to allow a 100 µm separation between the ASICs.

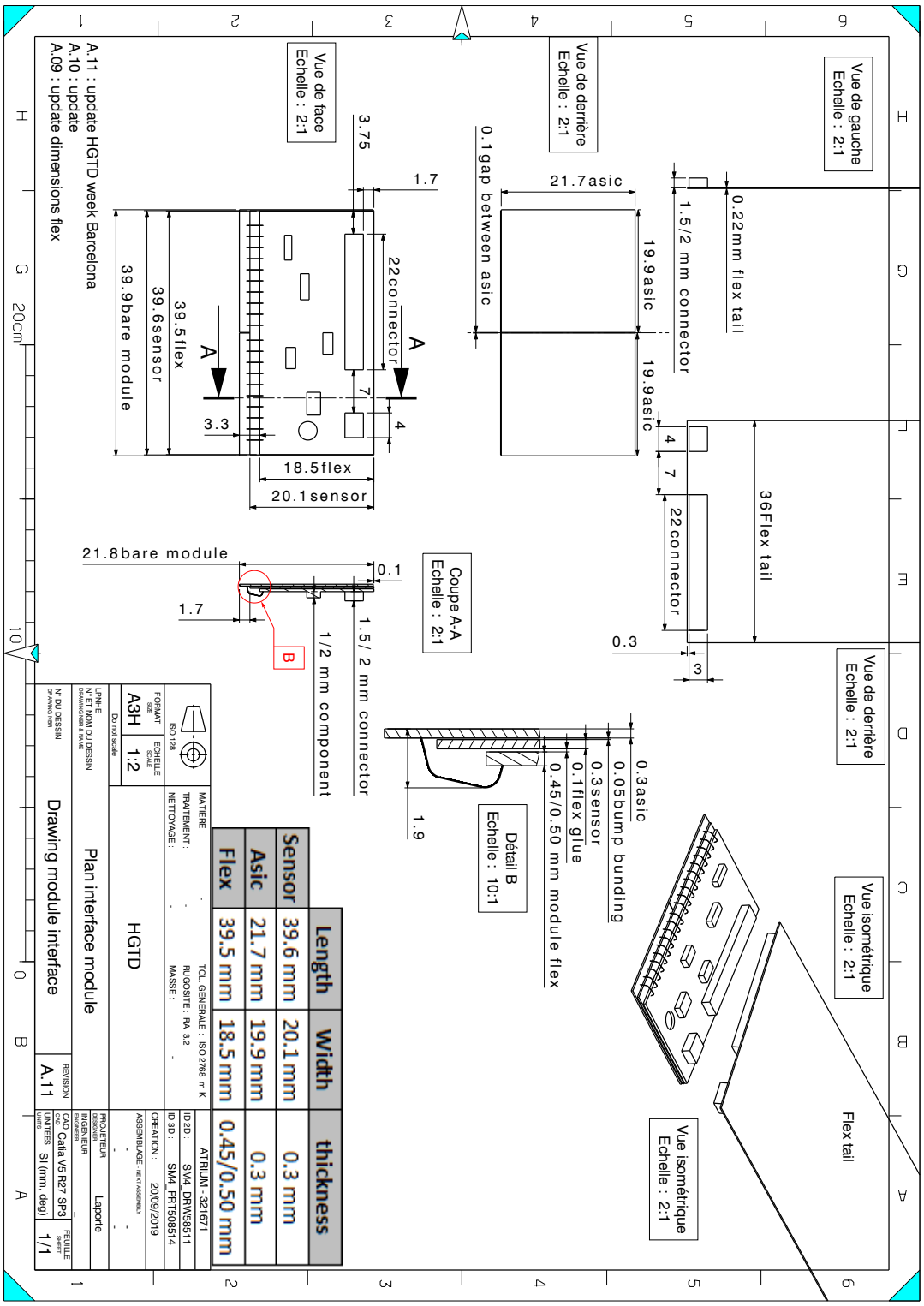


Figure D.2: Sketch of the module with the sensor, the ASICs and the Flex cable. Distances are in millimetres. Dimensions of the different components are visible, including the bumps pads, the glue and the wire-bonds.

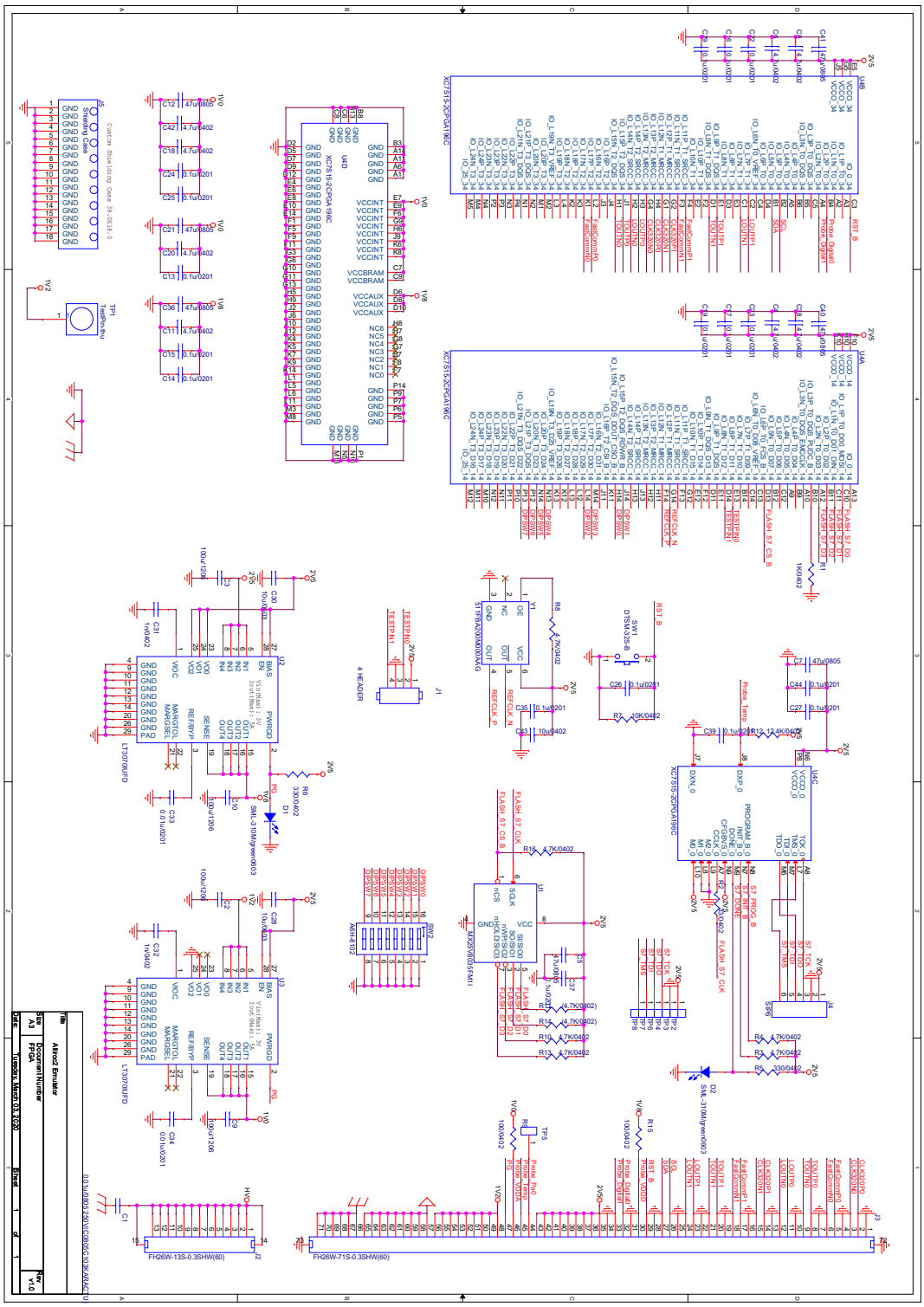


Figure D.3: Schematics of the design of the module flex prototype based on the ALTIROC2 pinout.

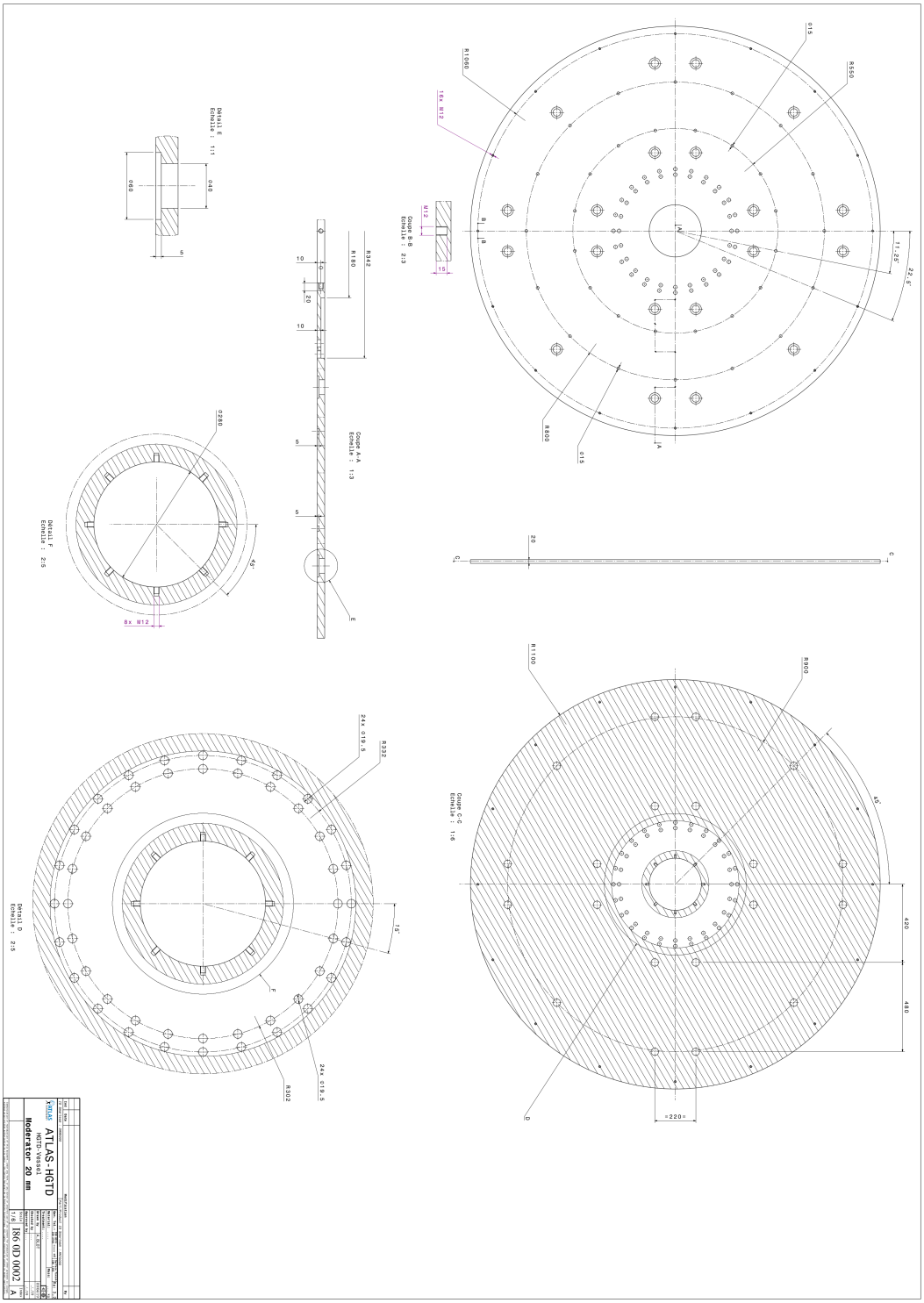


Figure D.4: Detailed technical drawing of the external moderator part.

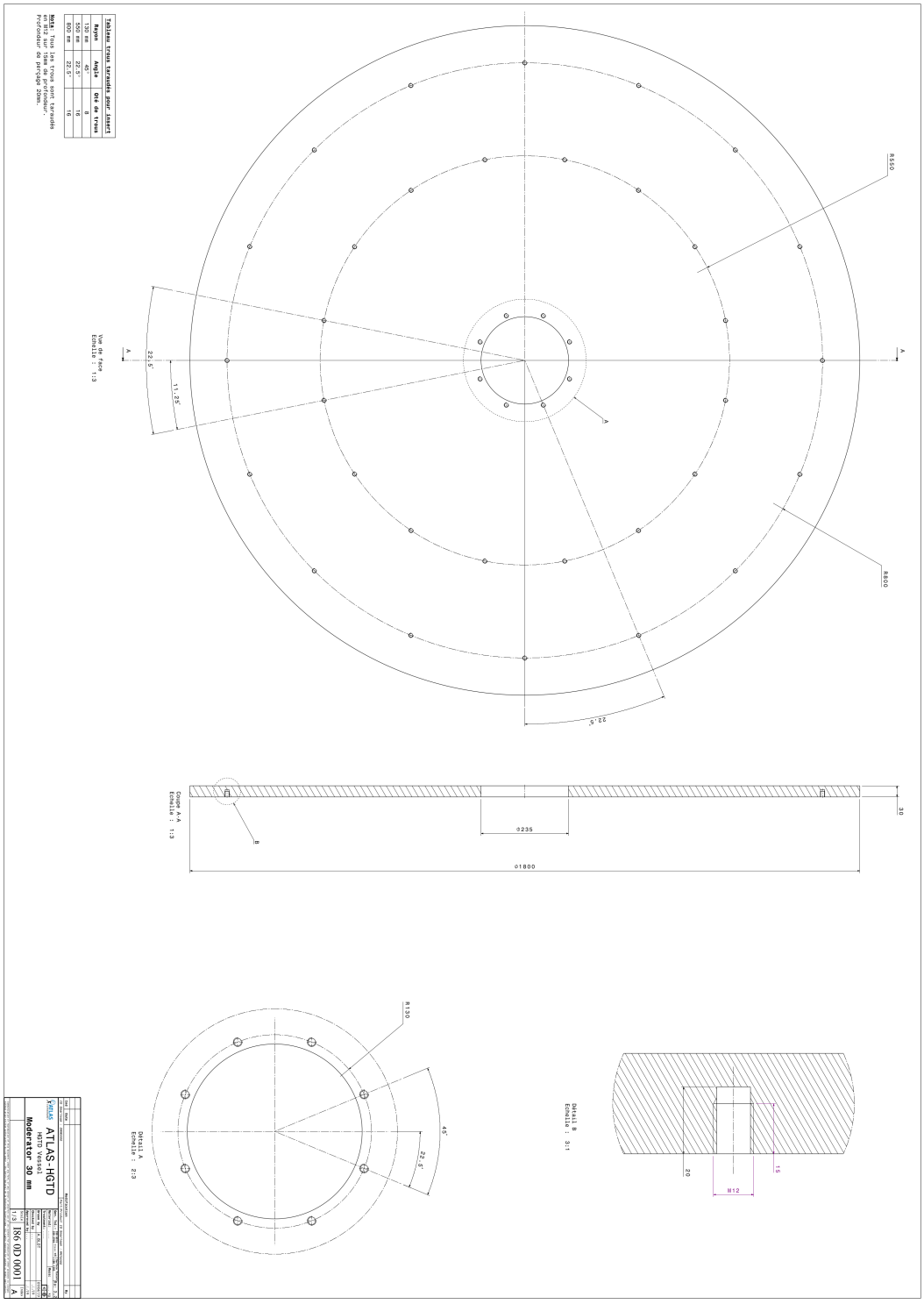


Figure D.5: Detailed technical drawing of the internal moderator part.

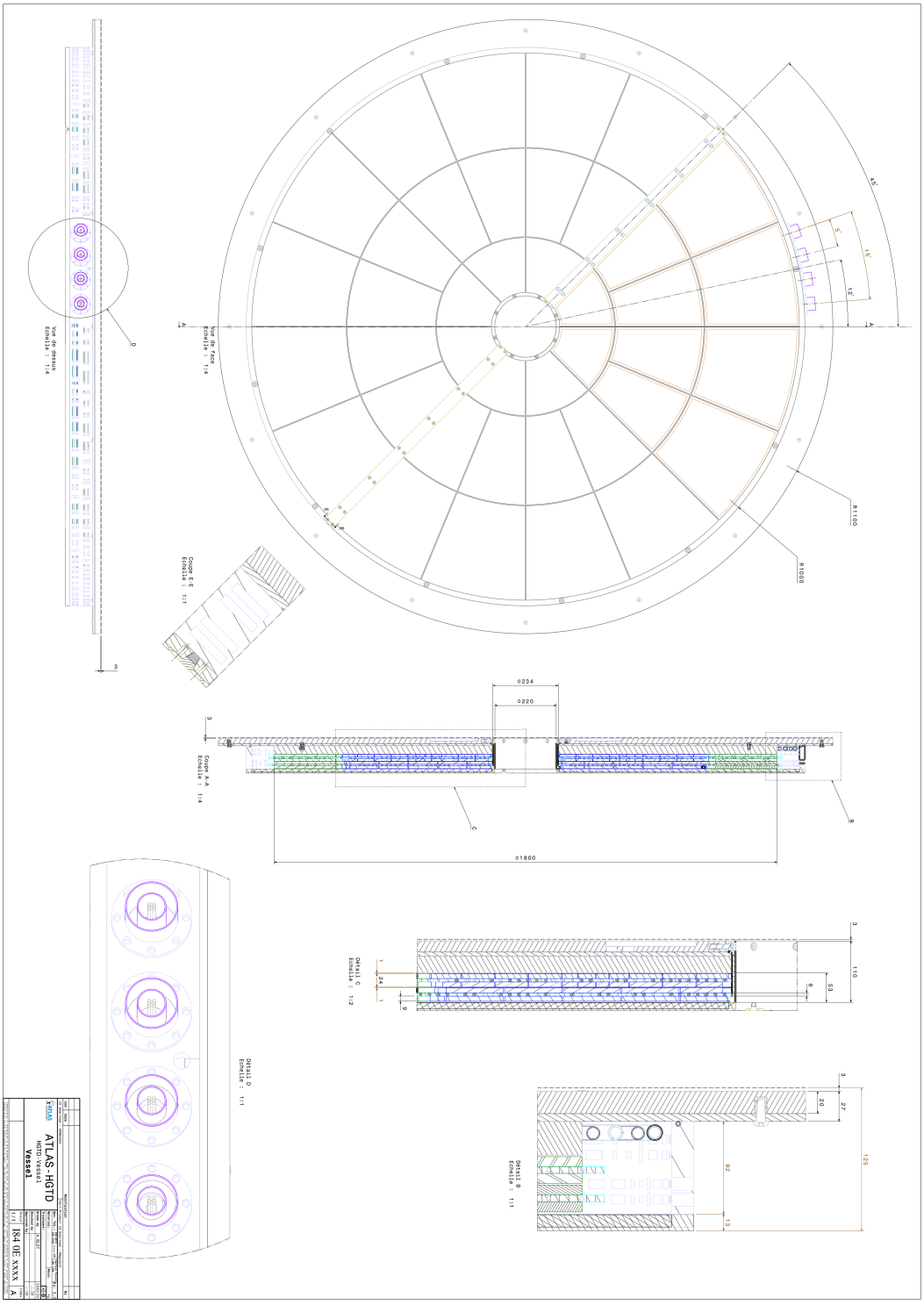


Figure D.6: Detailed 2D drawing of the full hermetic vessel assembly with main dimensions in R-ϕ & R-Z views. The CO2 transfer lines are located at 11.25° from the vertical axis as specified in the envelope study with ATLAS-TC and ITK integration team.

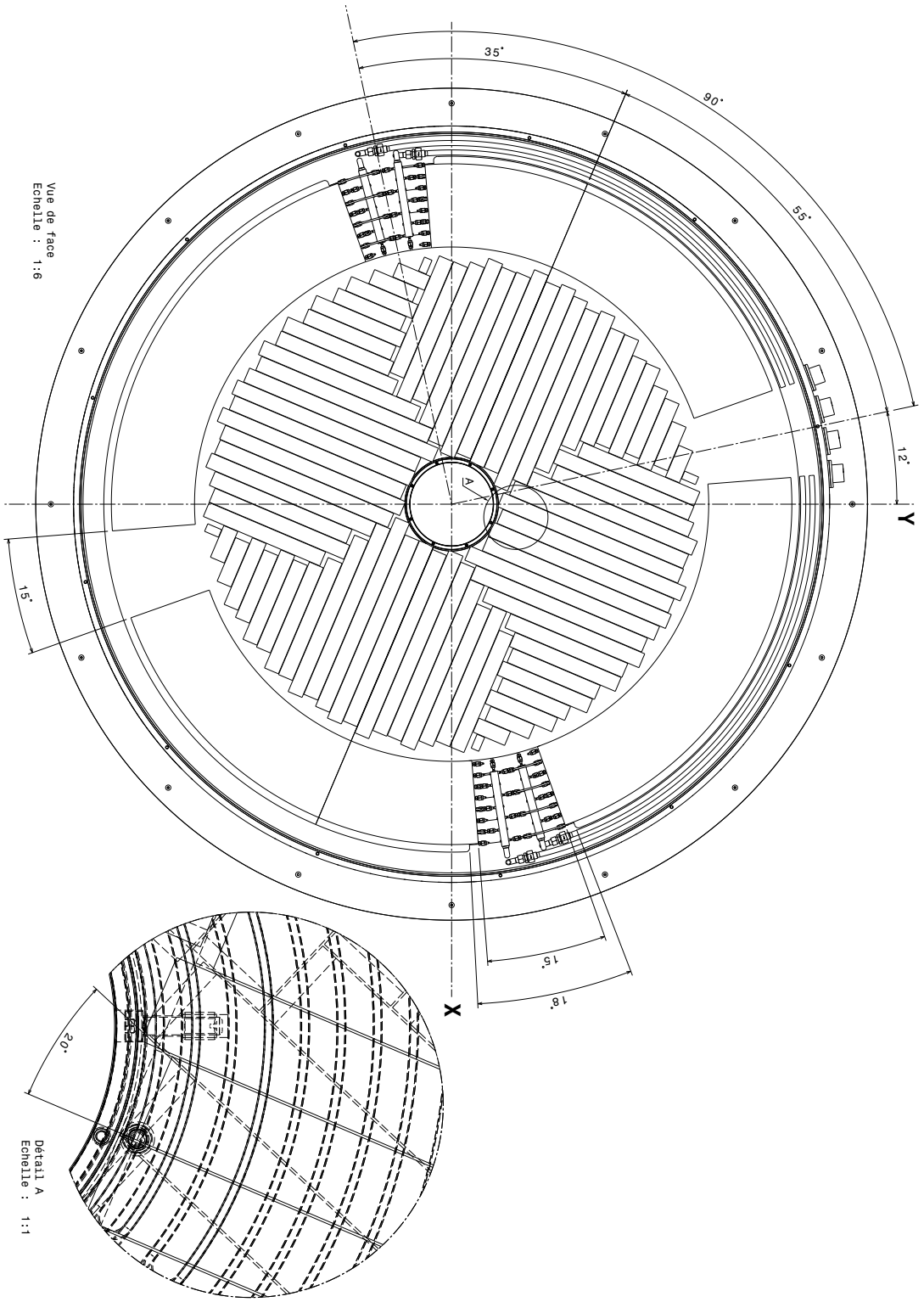


Figure D.7: View of the front side of the first detector disk, placed inside the vessel. There is a tilt of 20° between the two double sided layers, detailed in the zoomed region.

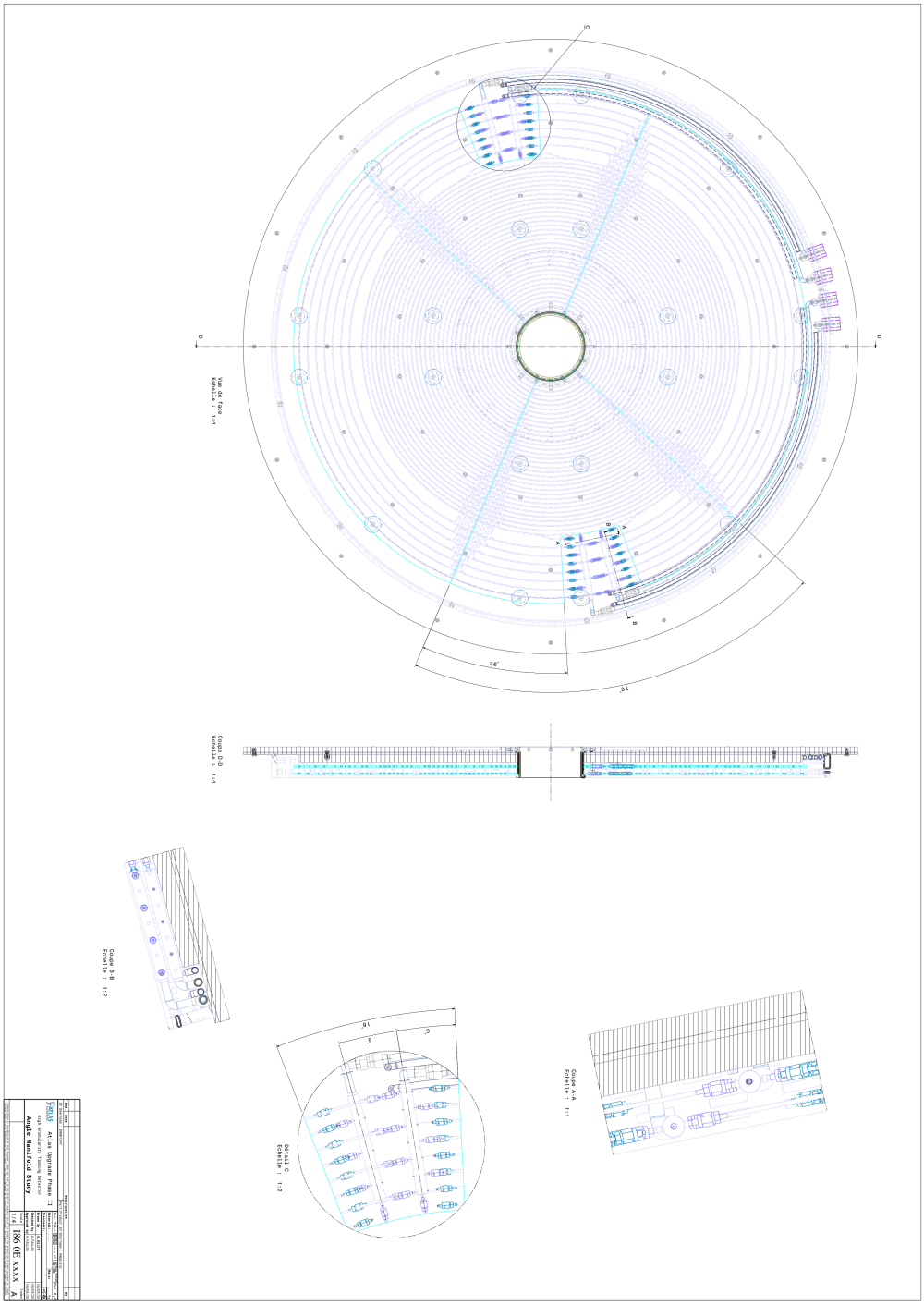


Figure D.8: Detailed 2D drawing of the full cooling loops and their manifolds inside the hermetic vessel. The R-ϕ front view is illustrating the cooling lines distribution and the tilt angle of 70° between the cooling plates which corresponds to 20° tilt between read out rows. The detailed views are showing the manifolds area and their restricted access space.

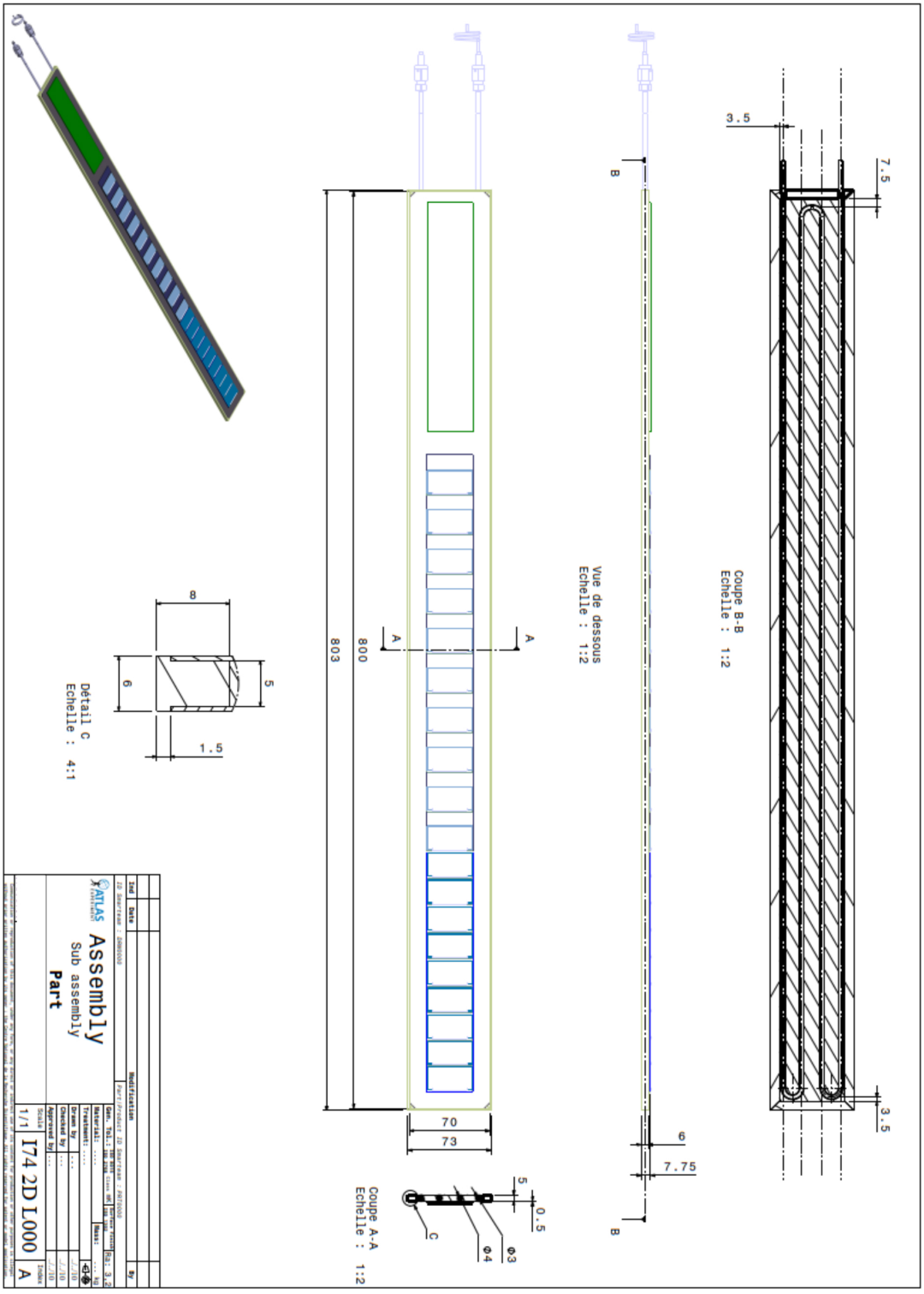


Figure D.9: View of the mechanical prototype planned for the HGTD demonstrator. It includes a cooling plate, dummy modules and connectivity to peripheral electronics board (indicated in green). The heaters simulating the modules power dissipation, using dummy modules are in blue.

Bibliography

[1] A. G. et al., *High-Luminosity Large Hadron Collider (HL-LHC): Technical Design Report V. 0.1*, CERN Yellow Reports: Monographs, CERN, 2017, URL: http://cds.cern.ch/record/2284929 (cit. on p. 1).

[2] ATLAS Collaboration, *Technical Design Report for the ATLAS Inner Tracker Pixel Detector*, tech. rep. CERN-LHCC-2017-021, CERN, 2018, URL: https://cds.cern.ch/record/2285585 (cit. on p. 1).

[3] A. Medina Tomas and Napusciale, *Assessment of the performance of High-Luminosity LHC operational scenarios: integrated luminosity and effective pile-up density*, tech. rep., 2019 498 (cit. on p. 5).

[4] RD50 collaboration, URL: https://rd50.web.cern.ch/rd50 (cit. on pp. 7, 88, 92).

[5] G. Pellegrini et al., *Technology developments and first measurements of Low Gain Avalanche Detectors (LGAD) for high energy physics applications*, Nucl. Instrum. Meth. **A765** (2014) 12 (cit. on pp. 7, 70, 88).

[6] ATLAS Collaboration, *Tracking Performance of the ATLAS Inner Tracker at the HL-LHC*, tech. rep., 2019, URL: https://cds.cern.ch/record/2665904 (cit. on pp. 8, 11).

[7] G. Aad et al., *The ATLAS Simulation Infrastructure*, Eur. Phys. J. **C70** (2010) 823, arXiv: 1005.4568 [physics.ins-det] (cit. on pp. 20, 51).

[8] S. Agostinelli et al., *GEANT4: A Simulation toolkit*, Nucl. Instrum. Meth. **A506** (2003) 250 (cit. on pp. 21, 51).

[9] CERN, URL: http://geomodel.web.cern.ch/geomodel/home/ (cit. on p. 21).

[10] L. Masetti et al., *Beam test measurements of Low Gain Avalanche Detector single pads and arrays for the ATLAS High Granularity Timing Detector*, JINST **13** (2018) P06017, arXiv: 1804.00622 [physics.ins-det] (cit. on pp. 24, 94, 95, 103–105, 108, 126).

[11] ATLAS Collaboration, *Expected Performance of the ATLAS Inner Tracker at the High-Luminosity LHC*, ATL-PHYS-PUB-2016-025, 2016, URL: https://cds.cern.ch/record/2222304 (cit. on pp. 42, 43).

[12] ATLAS Collaboration, *Tagging and suppression of pileup jets with the ATLAS detector*, ATLAS-CONF-2014-018, 2014, URL: https://cds.cern.ch/record/1700870 (cit. on p. 43).

[13] M. Cacciari, G. P. Salam and G. Soyez, *The anti-k_t jet clustering algorithm*, JHEP **04** (2008) 063, arXiv: 0802.1189 [hep-ph] (cit. on p. 43).

[14] M. Cacciari, G. P. Salam and G. Soyez, *FastJet user manual*, Eur. Phys. J. C **72** (2012) 1896, arXiv: 1111.6097 [hep-ph] (cit. on p. 43).

[15] *Expected Tracking Performance of the ATLAS Inner Tracker at the HL-LHC*, tech. rep. ATL-PHYS-PUB-2019-014, CERN, 2019, URL: https://cds.cern.ch/record/2669540 (cit. on p. 44).

[16] A. Flórez et al., *Searching for New Heavy Neutral Gauge Bosons using Vector Boson Fusion Processes at the LHC*, Phys. Lett. **B767** (2017) 126, arXiv: 1609.09765 [hep-ph] (cit. on p. 49).

[17] A. Berlin, T. Lin, M. Low and L.-T. Wang, *Neutralinos in Vector Boson Fusion at High Energy Colliders*, Phys. Rev. **D91** (2015) 115002, arXiv: 1502.05044 [hep-ph] (cit. on p. 49).

[18] ATLAS Collaboration, *Projections for measurements of Higgs boson cross sections, branching ratios, coupling parameters and mass with the ATLAS detector at the HL-LHC*, tech. rep., 2018, URL: https://cds.cern.ch/record/2652762 (cit. on p. 50).

[19] M. Aaboud et al., *Search for invisible Higgs boson decays in vector boson fusion at $\sqrt{s} =$ 13 TeV with the ATLAS detector*, Phys. Lett. **B793** (2019) 499, arXiv: 1809.06682 [hep-ex] (cit. on p. 51).

[20] P. Nason, *A New method for combining NLO QCD with shower Monte Carlo algorithms*, JHEP **11** (2004) 040, arXiv: hep-ph/0409146 (cit. on pp. 51, 337).

[21] S. Frixione, P. Nason and C. Oleari, *Matching NLO QCD computations with Parton Shower simulations: the POWHEG method*, JHEP **11** (2007) 070, arXiv: 0709.2092 [hep-ph] (cit. on pp. 51, 337).

[22] S. Alioli, P. Nason, C. Oleari and E. Re, *A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX*, JHEP **06** (2010) 043, arXiv: 1002.2581 [hep-ph] (cit. on pp. 51, 337).

[23] S. Alioli, P. Nason, C. Oleari and E. Re, *NLO vector-boson production matched with shower in POWHEG*, JHEP **07** (2008) 060, arXiv: 0805.4802 [hep-ph] (cit. on p. 51).

[24] P. Nason and C. Oleari, *NLO Higgs boson production via vector-boson fusion matched with shower in POWHEG*, JHEP **02** (2010) 037, arXiv: 0911.5299 [hep-ph] (cit. on p. 51).

[25] T. Sjostrand, S. Mrenna and P. Z. Skands, *A Brief Introduction to PYTHIA 8.1*, Comput. Phys. Commun. **178** (2008) 852, arXiv: 0710.3820 [hep-ph] (cit. on pp. 51, 337).

[26] T. Sjöstrand, S. Mrenna and P. Z. Skands, *PYTHIA 6.4 Physics and Manual*, JHEP **05** (2006) 026, arXiv: hep-ph/0603175 [hep-ph] (cit. on p. 51).

[27] *Prospects for Dark Matter searches in mono-photon and VBF+E_T^{miss} final states in ATLAS*, tech. rep. ATL-PHYS-PUB-2018-038, CERN, 2018, URL: http://cds.cern.ch/record/2649443 (cit. on p. 51).

[28] ATLAS Collaboration, *Performance of pile-up mitigation techniques for jets in pp collisions at $\sqrt{s} = 8$ TeV using the ATLAS detector*, Eur. Phys. J. **C76** (2016) 581, arXiv: 1510.03823 [hep-ex] (cit. on p. 51).

[29] B. W. Lee, C. Quigg and H. B. Thacker, *Strength of Weak Interactions at Very High Energies and the Higgs Boson Mass*, Phys. Rev. Lett. **38** (1977) 883 (cit. on p. 54).

[30] M. S. Chanowitz and M. K. Gaillard, *The TeV physics of strongly interacting W's and Z's*, Nucl. Phys. B **261** (1985) 379 (cit. on p. 54).

[31] M. Szleper, *The Higgs boson and the physics of WW scattering before and after Higgs discovery*, (2014), arXiv: 1412.8367 [hep-ph] (cit. on p. 54).

[32] M. Aaboud et al., *Observation of electroweak $W^{\pm}Z$ boson pair production in association with two jets in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector*, Phys. Lett. **B793** (2019) 469, arXiv: 1812.09740 [hep-ex] (cit. on p. 54).

[33] A. M. Sirunyan et al., *Measurement of electroweak WZ boson production and search for new physics in WZ + two jets events in pp collisions at $\sqrt{s} = 13$TeV*, Phys. Lett. **B795** (2019) 281, arXiv: 1901.04060 [hep-ex] (cit. on p. 54).

[34] *Prospective study of vector boson scattering in WZ fully leptonic final state at HL-LHC*, tech. rep. ATL-PHYS-PUB-2018-023, CERN, 2018, URL: http://cds.cern.ch/record/2645271 (cit. on pp. 54, 55).

[35] *Expected performance of the ATLAS detector at the High-Luminosity LHC*, tech. rep. ATL-PHYS-PUB-2019-005, CERN, 2019, URL: https://cds.cern.ch/record/2655304 (cit. on pp. 55, 60).

[36] ATLAS and C. Collaborations, *Addendum to the report on the physics at the HL-LHC, and perspectives for the HE-LHC: Collection of notes from ATLAS and CMS*, CERN Yellow Rep. Monogr. **7** (2019) Addendum, arXiv: 1902.10229 [hep-ex] (cit. on p. 55).

[37] T. Gleisberg, S. Höche, F. Krauss, M. Schönherr, S. Schumann et al., *Event generation with SHERPA 1.1*, JHEP **02** (2009) 007, arXiv: 0811.4622 [hep-ph] (cit. on p. 55).

[38] O. J. P. Eboli, M. C. Gonzalez-Garcia and S. M. Lietti, *Bosonic quartic couplings at CERN LHC*, Phys. Rev. **D69** (2004) 095005, arXiv: hep-ph/0310141 [hep-ph] (cit. on p. 55).

[39] ALEPH Collaboration, DELPHI Collaboration, L3 Collaboration, OPAL Collaboration, SLD Collaboration, LEP Electroweak Working Group, SLD Electroweak Group, SLD Heavy Flavour Group, *Precision electroweak measurements on the Z resonance*, Phys. Rept. **427** (2006) 257, arXiv: hep-ex/0509008 [hep-ex] (cit. on p. 57).

[40] J. C. Collins and D. E. Soper, *Angular distribution of dileptons in high-energy hadron collisions*, Phys. Rev. **D 16** (1977) 2219 (cit. on p. 57).

[41] ATLAS Collaboration, *Precision measurement and interpretation of inclusive W^+ , W^- and Z/γ^* production cross sections with the ATLAS detector*, Eur. Phys. J. **C77** (2017) 367, arXiv: 1612.03016 [hep-ex] (cit. on pp. 57, 60).

[42] *ATLAS Phase-II Upgrade Scoping Document*, tech. rep. CERN-LHCC-2015-020. LHCC-G-166, CERN, 2015, URL: https://cds.cern.ch/record/2055248 (cit. on pp. 58, 298).

[43] ATLAS Collaboration, *Expected performance for an upgraded ATLAS detector at High-Luminosity LHC*, ATL-PHYS-PUB-2016-026, 2016, URL: https://cds.cern.ch/record/2223839 (cit. on p. 58).

[44] ATLAS Collaboration, *Electron and photon energy calibration with the ATLAS detector using LHC Run 1 data*, Eur. Phys. J. **C74** (2014) 3071, arXiv: 1407.5063 [hep-ex] (cit. on p. 58).

[45] T. A. collaboration, *Luminosity determination in pp collisions at $\sqrt{s} = 13$ TeV using the ATLAS detector at the LHC*, (2019) (cit. on pp. 60, 223).

[46] *Addendum to the report on the physics at the HL-LHC, and perspectives for the HE-LHC: Collection of notes from ATLAS and CMS*, tech. rep. arXiv:1902.10229, CERN, 2019, URL: http://cds.cern.ch/record/2651134 (cit. on p. 61).

[47] P. A. M. Dirac, *Quantised singularities in the electromagnetic field*, Proc. Roy. Soc. Lond. **A133** (1931) 60 (cit. on p. 62).

[48] G. 't Hooft, *Magnetic Monopoles in Unified Gauge Theories*, Nucl. Phys. **B79** (1974) 276, [,291(1974)] (cit. on p. 62).

[49] Y. M. Cho and D. Maison, *Monopoles in Weinberg-Salam model*, Phys. Lett. **B391** (1997) 360, arXiv: hep-th/9601028 [hep-th] (cit. on p. 62).

[50] Y. M. Cho, K. Kimm and J. H. Yoon, *Mass of the Electroweak Monopole*, Mod. Phys. Lett. **A31** (2016) 1650053, arXiv: 1212.3885 [hep-ph] (cit. on p. 62).

[51] ATLAS Collaboration, *Search for Massive Long-lived Highly Ionising Particles with the ATLAS Detector at the LHC*, Phys. Lett. **B698** (2011) 353, arXiv: 1102.0459 [hep-ex] (cit. on p. 62).

[52] ATLAS Collaboration, *Search for magnetic monopoles in $\sqrt{s} = 7$ TeV pp collisions with the ATLAS detector*, Phys. Rev. Lett. **109** (2012) 261803, arXiv: 1207.6411 [hep-ex] (cit. on p. 62).

[53] ATLAS Collaboration, *Search for magnetic monopoles and stable particles with high electric charges in 8 TeV pp collisions with the ATLAS detector*, Phys. Rev. **D93** (2016) 052009, URL: https://link.aps.org/doi/10.1103/PhysRevD.93.052009 (cit. on p. 62).

[54] J. Pinfold et al., *Technical Design Report of the MoEDAL Experiment*, tech. rep. CERN-LHCC-2009-006. MoEDAL-TDR-001, 2009, URL: https://cds.cern.ch/record/1181486 (cit. on p. 62).

[55] L. Lee, C. Ohm, A. Soffer and T.-T. Yu, *Collider Searches for Long-Lived Particles Beyond the Standard Model*, Prog. Part. Nucl. Phys. **106** (2019) 210, arXiv: 1810.12602 [hep-ph] (cit. on p. 62).

[56] ATLAS Collaboration, *Measurement of the Inelastic Proton-Proton Cross Section at $\sqrt{s} =$ 13 TeV with the ATLAS Detector at the LHC*, Phys. Rev. Lett. **117** (2016) 182002, arXiv: 1606.02625 [hep-ex] (cit. on p. 63).

[57] ATLAS Collaboration, *The Pythia 8 A3 tune description of ATLAS minimum bias and inelastic measurements incorporating the Donnachie–Landshoff diffractive model*, ATL-PHYS-PUB-2016-017, 2016, URL: https://cds.cern.ch/record/2206965 (cit. on p. 63).

[58] ATLAS Collaboration, *Technical Design Report for the Phase-II Upgrade of the ATLAS Trigger and Data Acquisition System*, tech. rep. CERN-LHCC-2017-020 ; ATLAS-TDR-029, CERN, 2017, URL: http://cds.cern.ch/record/2285584 (cit. on pp. 68, 210, 211).

[59] T. lpGBT team, *The lpBT manual*, (), URL: http://lpgbt.web.cern.ch/lpgbt (cit. on pp. 68, 193, 198).

[60] C. Agapopoulou et al., *Performance of a Front End prototype ASIC for picosecond precision time measurements with LGAD sensors*, (2020), arXiv: 2002.06089 [physics.ins-det] (cit. on pp. 74, 117, 142, 155).

[61] H. Sadrozinski, A. Seiden and N. Cartiglia, *4-Dimensional Tracking with Ultra-Fast Silicon Detectors*, Reports on Progress in Physics (2017), arXiv: 1704.08666, URL: http://iopscience.iop.org/10.1088/1361-6633/aa94d3 (cit. on pp. 88, 94, 103, 104, 128).

[62] G. Pellegrini et al., *Status of LGAD production at CNM*, 30th RD50 Workshop, Krakow, Poland, 2017, URL: https://indico.cern.ch/event/637212/contributions/2608652/attachments/1470919/2276240/pellegrini_rd50.pdf (cit. on p. 89).

[63] CMS Collaboration, *Technical proposal for a MIP timing detector in the CMS Experiment Phase 2 upgrade*, tech. rep. CERN-LHCC-2017-027. LHCC-P-009, CERN, 2017, URL: https://cds.cern.ch/record/2296612 (cit. on p. 91).

[64] G. Kramberger et al., *Radiation effects in Low Gain Avalanche Detectors after hadron irradiations*, JINST **10** (2015) P07006 (cit. on pp. 93, 94).

[65] G. Kramberger et al., *Radiation hardness of thin Low Gain Avalanche Detectors*, Nucl. Instrum. Meth. **A891** (2018) 68, ISSN: 0168-9002, arXiv: 1711.06003, URL: http://www.sciencedirect.com/science/article/pii/S0168900218301682 (cit. on pp. 93, 94).

[66] N. Cartiglia et al., *Beam test results of a 16 ps timing system based on ultra-fast silicon detectors*, Nucl. Instrum. Meth. **A850** (2017) 83, arXiv: 1608.08681 [physics.ins-det] (cit. on pp. 94, 103, 104, 126).

[67] J. Lange et al., *Gain and time resolution of 45 µm thin Low Gain Avalanche Detectors before and after irradiation up to a fluence of* 10^{15} n_{eq}/cm^2, JINST **12** (2017) P05003, arXiv: 1703.09004 [physics.ins-det] (cit. on pp. 94, 103, 104).

[68] Z. Galloway et al., *Properties of HPK UFSD after neutron irradiation up to 6e15 n/cm^2*, submitted to NIM A (2017), arXiv: 1707.04961 (cit. on pp. 94, 104, 112).

[69] Y. Zhao et al., *Comparison of 35 and 50 µm thin HPK UFSD after neutron irradiation up to 6E15Neq*, (2018), arXiv: 1803.02690 (cit. on pp. 94, 104, 112).

[70] S. M. Mazza et al., *Proprieties of FBK UFSDs after neutron and proton irradiation up to* $6 * 10^{15}$ *neq/cm^2*, (2018), arXiv: 1804.05449 [physics.ins-det] (cit. on pp. 94, 104).

[71] R. Diener et al., *The DESY II test beam facility*, Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment **922** (2019) 265, ISSN: 0168-9002, URL: http://www.sciencedirect.com/science/article/pii/S0168900218317868 (cit. on p. 94).

[72] Tech. rep., URL: https://epjtechniquesandinstrumentation.springeropen.com/articles/10.1140/epjti/s40485-016-0033-2 (cit. on p. 94).

[73] F. Cenna et al., *Weightfield2: A fast simulator for silicon and diamond solid state detector*, Nucl. Instrum. Meth. **A796** (2015) 149 (cit. on p. 95).

[74] Synopsys, URL: https://www.synopsys.com/silicon/tcad/device-simulation/sentaurus-device.html (cit. on p. 95).

[75] M. Schwickardi, 'Characterisation of Low Gain Avalanche Detectors', II.Physik-UniGö-MSc-2019/06, Master's Thesis: Georg-August-Universität Göttingen, 2019 (cit. on pp. 97, 98, 105).

[76] ATLAS Collaboration, *ATLAS Trigger and Data Acquisition Phase-II Upgrade Technical Design Report*, tech. rep. ATL-COM-DAQ-2017-185, This version represents the draft for EB approval and LHCC release.: CERN, 2017, URL: https://cds.cern.ch/record/2296879 (cit. on p. 118).

[77] K. Wyllie et al., *A Gigabit Transceiver for Data Transmission in Future High Energy Physics Experiments*, Phys. Procedia **37** (2012) 1561, URL: https://cds.cern.ch/record/2103391 (cit. on p. 120).

[78] *Radiation tolerance of the 130nm TSMC technology*, URL: https://asicsupport-community.web.cern.ch/t/radiation-tolerance-of-the-130nm-tsmc-technology/144 (cit. on p. 121).

[79] F. Cenna et al., *Weightfield2: A fast simulator for silicon and diamond solid state detector*, Nucl. Instr. Meth. A **796** (2015) 149, URL: http://www.sciencedirect.com/science/article/pii/S0168900215004842 (cit. on p. 127).

[80] Cadence Sigrity and PowerSI Cadence Desgn Systems Inc, URL: https://www.cadence.com/content/dam/cadence-www/global/en_US/documents/tools/pcb-design-%20analysis/sigrity-powersi-ds.pdf (cit. on p. 165).

[81] CAEN Electronic Instruments, *DT55xxE Desktop HV Power Supply*, URL: https://www.caen.it/products/dt5521he/ (cit. on p. 166).

[82] Tektronix, *TDR Impedance Measurements: A Foundation for Signal Integrity*, URL: http://www.tek.com/dl/55W_14601_2.pdf (cit. on p. 167).

[83] KC705 Evaluation Board. Xilink Inc., URL: https://www.xilinx.com/support/documentation/boards_and_kits/kc705/ug810_KC705_Eval_Bd.pdf (cit. on p. 167).

[84] G. Beck and G. Viehhauser, *Analytic model of thermal runaway in silicon detectors*, Nucl. Instrum. Meth. **A618** (2010) 131 (cit. on pp. 180, 183, 184).

[85] ATLAS Collaboration, *Technical Design Report for the ATLAS Inner Tracker Strip Detector*, tech. rep. CERN-LHCC-2017-005 ; ATLAS-TDR-025, CERN, 2017, URL: https://cds.cern.ch/record/2257755 (cit. on pp. 191, 246, 253, 254).

[86] L. Amaral et al., *The versatile link, a common project for super-LHC*, Journal of Instrumentation **4** (2009) P12003 (cit. on p. 194).

[87] *EMCI technical specifications*, tech. rep., URL: https://edms.cern.ch/document/2332346/2 (cit. on pp. 201, 235).

[88] J. Anderson et al., *FELIX: a PCIe based high-throughput approach for interfacing front-end and trigger electronics in the ATLAS Upgrade framework*, Journal of Instrumentation **11** (2016) C12023 (cit. on p. 209).

[89] J. Troska, S. Biereigel, S. Kulis and E. Mendes, *Preliminary lpGBT clock characterization*, tech. rep., URL: https://espace.cern.ch/HighPrecisionTiming/Evaluations/Components%20Evaluations/LpGBT/20190118lpgbtJitterReport.pdf (cit. on p. 214).

[90] K. Chen and H. C. W. Wu, *Evaluation of System Clock from TTC to Front-Ends via FELIX*, tech. rep., URL: https://indico.cern.ch/event/676116/contributions/2767727/attachments/1548939/2440064/20171030_KChen_Evaluation_of_System_Clock_from_TTC_to_Front-Ends_via_FELIX.pdf (cit. on p. 214).

[91] T. Mastoridis, P. Baudrenghien and J. Molendijk, *Cavity voltage phase modulation to reduce the high-luminosity Large Hadron Collider rf power requirements*, Phys. Rev. Accel. Beams **20** (10 2017) 101003, URL: https://link.aps.org/doi/10.1103/PhysRevAccelBeams.20.101003 (cit. on p. 215).

[92] ATLAS Collaboration, *Luminosity determination in pp collisions at $\sqrt{s}$ = 8 TeV using the ATLAS detector at the LHC*, Eur. Phys. J. **C76** (2016) 653, arXiv: 1608.03953 [hep-ex] (cit. on pp. 218, 219, 223).

[93] CMS Collaboration, *CMS Luminosity Measurements for the 2016 Data Taking Period*, (2017) (cit. on p. 218).

[94] G. Aad et al., *Improved luminosity determination in pp collisions at sqrt(s) = 7 TeV using the ATLAS detector at the LHC*, Eur. Phys. J. **C73** (2013) 2518, arXiv: 1302.4393 [hep-ex] (cit. on pp. 218, 222).

[95] G. Avoni et al., *The new LUCID-2 detector for luminosity measurement and monitoring in ATLAS*, J. Inst. **13** (2018) P07017 (cit. on p. 218).

[96] P. Grafstrom and W. Kozanecki, *Luminosity determination at proton colliders*, Prog. Part. Nucl. Phys. **81** (2015) 97 (cit. on p. 219).

[97] S. R. J. C. Armenteros A. Cimmino and H. Vincke, *FLUKA studies of dose rates in the ATLAS standard opening scenario.*, tech. rep., URL: http://accapp17.org/program-2/plenary-sessions/ (cit. on p. 283).

[98] *FLUKA studies of dose rates in ATLAS (open detector) up to LS6, CERN-RP-ATLAS-TN-2017-A1*, tech. rep., URL: https://edms.cern.ch/document/1735862/4 (cit. on p. 283).

[99] Tech. rep., URL: https://indico.cern.ch/event/677021 (cit. on pp. 283, 284).

[100] Tech. rep., URL: https://indico.cern.ch/event/619476 (cit. on pp. 283, 285).

[101] ATLAS Collaboration, *Technical Proposal: A High-Granularity Timing Detector for the ATLAS Phase-II Upgrade*, tech. rep. CERN-LHCC-2018-023, CERN, 2018, URL: https://cds.cern.ch/record/2623663 (cit. on p. 298).

[102] J. Wiley and S. L. 2013, *Project Management Institute, A Guide to the Project Management Body of Knowledge (PMBOK Guide)*, () (cit. on p. 327).

[103] ATLAS Collaboration, *ATLAS Computing Acknowledgements*, ATL-GEN-PUB-2016-002, URL: https://cds.cern.ch/record/2202407 (cit. on p. 333).