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

Draft version: 1.5
Created: 9th January 2020 – 16:33
Prepared by: **The ATLAS Collaboration**

Abstract

The large increase of pile-up interactions is one of the main experimental challenges for the HL-LHC physics program. A powerful new way to mitigate the effects of pile-up 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 significantly improve the performance in the forward region. Taking into account the typical number of hits per track in the detector, 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 greatly improves the track-to-vertex association, leading to a performance similar to that in the central region for both jet and lepton reconstruction, as well as the tagging of heavy-flavour jets. These improvements in object reconstruction performance translate into important sensitivity gains and enhance the reach of the HL-LHC physics program. In addition, the HGTD offers unique capabilities for the online and offline luminosity determination, an important requirement for precision physics measurements.

Contents

1 Introduction

The high-luminosity (HL) Phase-II of the Large Hadron Collider (LHC) at CERN is scheduled to start in 2026. This HL-LHC will deliver an integrated luminosity of up to 4000 fb^{-1} over the subsequent decade. 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 LHC Run-2. Two extended periods without physics operation are anticipated prior to Phase-II, Long Shutdown 2 (LS2) in 2019–2020 and Long Shutdown 3 (LS3) from 2024 until mid 2026. During LS3, extensive upgrades to the ATLAS Experiment will be installed to cope with the higher luminosities and add new capabilities.

This report describes the technical design of a High Granularity Timing Detector (HGTD) for the ATLAS Phase-II upgrade, an upgrade that will augment the Phase-II silicon-based Inner Tracker, the ITk [1], in the forward region, with the capability to measure charged-particle trajectories in time as well as space. The target average time resolution per track for a minimum-ionising particle is ≈ 30 ps at the beginning of the HL-LHC, increasing to 50 ps per track at the end of the HL-LHC. The HGTD will also provide a precise, real-time luminosity measurement for ATLAS's Phase-II physics programme. The goals for the design are detailed in Chap. 2, and Chap. 3 provides projections for how the detector will improve ATLAS object reconstruction and physics.

The technical design of the HGTD is summarized in Chap. 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 track traverses two or three sensors. Chap. 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. Chap. 6 describes the front-end electronics, a low-noise, radiation-hard custom ASIC called the ALTIROC, and the performance of the analog front end. Chap. 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 disks and staves, and their connection via a flex cable to peripheral electronics boards at the outer radii of the disk geometry. Chap. 8 describes the powering and control of the detector. Chap. 9 describes the function and layout of the peripheral electronics boards, and Chap. 10 summarizes the connection of the detector to the ATLAS data acquistion system, the real-time intercalibration of the arrival time within the readout path, and the 40 MHz readout of highly-granular hit multiplicity data for real-time luminosity measurement. Chap. 11 provides the engineering design for cooling the LGADs and front-end electronics, Chap. 12 the mechanical design

of the overall detector, the necessary services and routing, and Sec. 13.1 the assembly and commissioning of the detector. Chap. 14 describes a set of intermediate prototypes that will integrate elements of the full design during the remaining R&D period, in order to validate key aspects of the design. Finally, Chap. 15 documents the organisation of the project to deliver and commission the detector for the start of the HL-LHC in 2026.

2 Detector Requirements and Layout

2.1 Beam conditions at the HL-LHC

Pile-up is one of the main challenges at the HL-LHC. The exact beam-spot characteristics of the HL-LHC have not yet been determined. In the nominal operation scheme [2], an average of 200 simultaneous pp interactions ($\langle\mu\rangle = 200$) will occur within the same bunch crossing interval. A major challenge for the ITk is to efficiently reconstruct charged particles created in the primary interactions and assign them to the correct production vertices in this high pile-up environment. 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 spreads of 45 mm in along the beam axis and 175 ps in time.

The spatial pile-up 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 an average of 200 collisions per bunch crossing, denoted $\langle\mu\rangle = 200$, an average pile-up density of 1.8 collisions/mm is expected. This average masks the effect of the local variations illustrated in Fig. 2.1. The local pile-up vertex density 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 is large enough to avoid quantisation effects and small enough to probe the tails of the distribution. The curve for $\langle\mu\rangle = 30$ is obtained by scaling, which effectively increases the window size.

Fig. 2.1 shows the pile-up densities for $\langle\mu\rangle = 30$ and $\langle\mu\rangle = 200$ for the same beam spot size. The most probable local pile-up density for this scenario is around 1.44 collisions/mm.

Timing information can supplement the tracker measurement of the longitudinal impact parameter, z_0, improving how often tracks are assigned to the correct vertices and mitigating the impact of a high vertex density. To illustrate this, an example is presented in Fig. 2.2, which shows a single $t\bar{t}$ Hard Scatter (HS) vertex along pile-up vertices in the z–t plane at $\langle\mu\rangle = 200$. The simulated and reconstructed vertices are overlaid. The vertex and track

[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 $\eta = -\ln\tan(\theta/2)$.

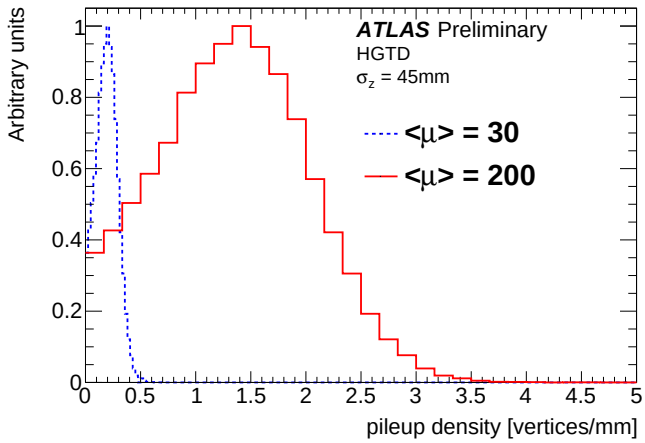


Figure 2.1: Local pile-up vertex densities for different values of $\langle\mu\rangle$.

reconstruction used in this event are further described in Sec. 3.2.2. The impact parameter resolution in the forward region is limited by multiple scattering. A minimal p_{T} cut of 0.9 GeV is applied for tracks at all η to reject soft forward tracks. Each track is required to have at least 3 pixel clusters. Tracks are required to have $\sigma(d_0) < 0.3$ mm and $\sigma(z_0) < 0.5$ mm in order to ensure good precision. This cut effectively removes low-p_{T} forward tracks because of their limited resolution.

The tracker sees the event as a one-dimensional projection on the z axis, where a large number of tracks from vertices occurring at different times but close in space lead to ambiguities in the track-to-vertex association. This happens when the distance between vertices is of the same order or smaller than the resolution of the longitudinal impact parameter of the track, which happens more often for tracks in the forward region. The timing information reduces the density of vertices which are considered for a given track. The figure shows the reconstructed vertices are also spread in time, due to the association of tracks with timing information of the HGTD. The time of HS vertex may be reconstructed using tracks with hits in the HGTD, as discussed Sec. 3.2.3, allows for more effective separation of the HS vertex from pile-up vertices surrounding it in the z direction.

2.2 Detector overview and requirements

The HGTD has been 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.

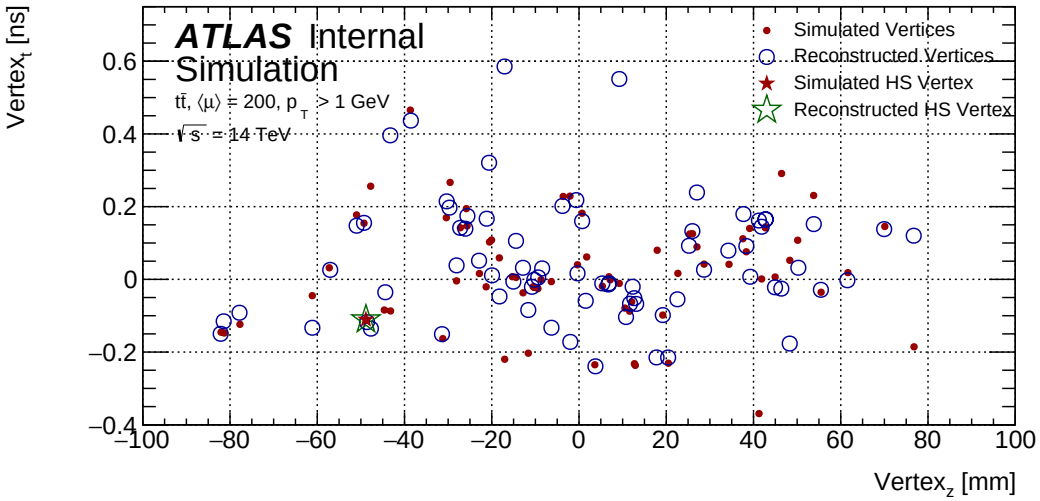


Figure 2.2: Visualisation of the primary vertices in an event in the z–t plane, showing the simulated Hard Scatter (HS) with pile-up interactions superimposed for $\langle\mu\rangle = 200$. The reconstructed vertices (blue circles) are overlaid along with the reconstructed HS vertex (green star).

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 (MBTS), which will be removed. The position of the two vessels for the HGTD within the ATLAS detector is shown in Fig. 2.3.

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. 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 preferred due to the space limitations. The sensors must be thin and configurable in arrays. In close collaboration with RD50 [3] and manufacturers, an extensive R&D program is progressing quickly towards sensors that provide the required timing resolution in harsh radiation environments. Low Gain Avalanche Detector (LGAD) [4] 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 pile-up, small dead areas between pads, and low sensor capacitance, which is important for the time resolution.

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 this 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 and cost, the detector design described in this document will give an average of

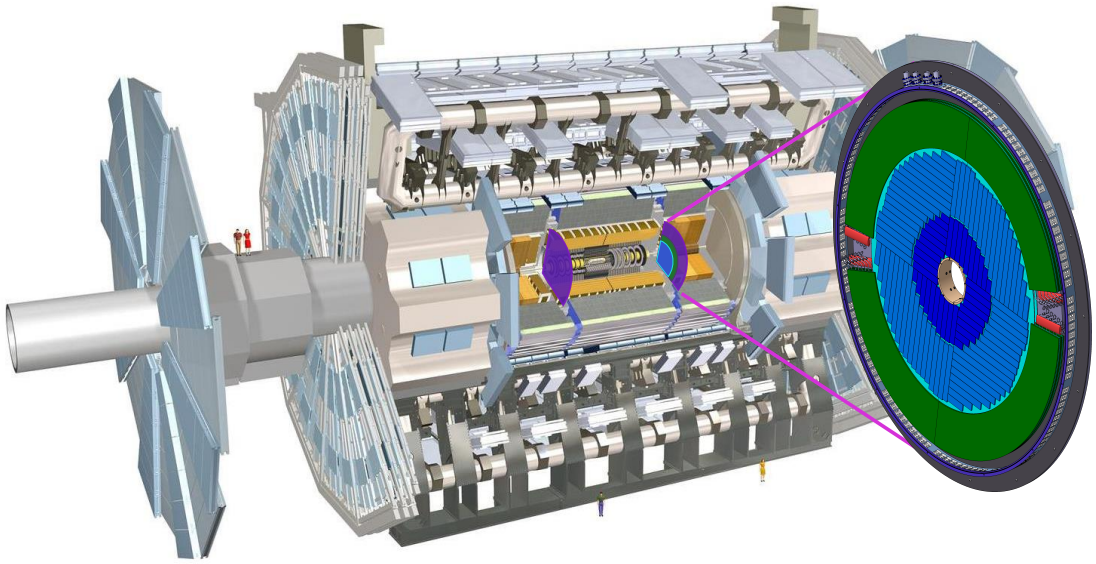


Figure 2.3: *TO BE UPDATED* Position of the HGTD within the ATLAS Detector. The HGTD acceptance is defined as the surface covered by the HGTD between a radius of 120 mm and 640 mm at a position of $z = \pm 3.5$ m along the beamline, on both sides of the detector.

2.1, 2.5 and 2.7 hits per track for the regions 230 mm $> r >$ 120 mm, 470 mm $> r >$ 230 mm and 640 mm $> r >$ 470 mm. A description of the detector layout optimisation is presented in Sec. 2.3. It covers the pseudo-rapidity range $2.4 < |\eta| < 4.0$.

Each HGTD end-cap is the integration 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. Each cooling/support disk is physically separated in two half circular disks. Furthermore, the layers are rotated in opposite directions with respect to one another by 72° in order to maximize the hit efficiency.

A global view of the various components of the detector and its main parameters are shown in Fig. 2.4 and Tab. 2.1. The time resolution parameters have been optimised using information from the sensor Chap. 5 and front-end electronics Chap. 6 performance from lab and testbeam 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 pile-up in the forward region. The ability to associate tracks to primary vertices depends on the longitudinal impact parameter resolution of the ITk. The current ITk layout is shown in Fig. 2.5. Fig. 2.6 shows the resolution, σ_{z_0}, of the longitudinal

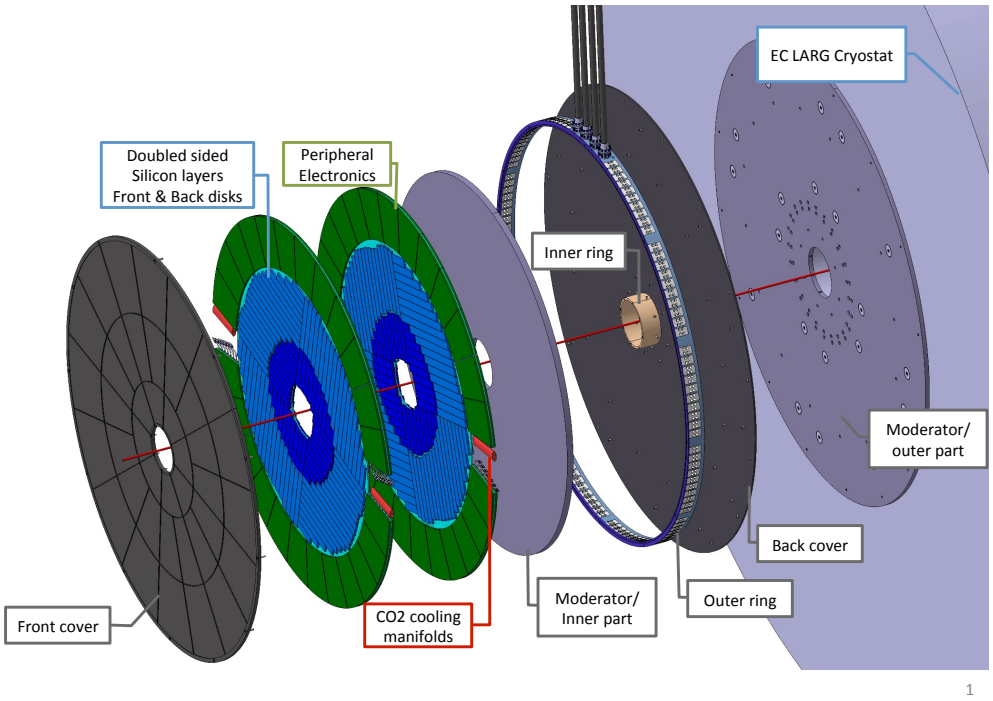


Figure 2.4: *TO BE UPDATED* Global view of the HGTD to be installed on each of two end-cap calorimeters. The various components are shown: hermetic vessel (front and rear covers, inner and outer rings), two instrumented double-sided layers (mounted in two cooling disks), two moderator pieces placed inside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

track impact parameter, z_0, from the ITk as a function of η, for $p_{\mathrm{T}} = 1\,\mathrm{GeV}$ and $p_{\mathrm{T}} = 10\,\mathrm{GeV}$ muons. The resolution is shown for both digital and analogue clustering. Digital clustering shows a similar performance to analogue clustering for $p_{\mathrm{T}} = 1\,\mathrm{GeV}$ muons and a 10–20 % degradation in σ_{z_0} for $p_{\mathrm{T}} = 10\,\mathrm{GeV}$ muons. Digital clustering takes the geometrical average as the centroid of a given pixel cluster, while analogue clustering improves the ability to determine the position by weighting the centroid by the charge deposited in each of the pixels contributing to the cluster. In this report, performance studies have been performed with an ITk layout and simulation [5] including a sensor pitch of $50\times50\,\mu\mathrm{m}$ and digital clustering, except where otherwise specified.

For good spatial separation of the HL-LHC collision vertices, σ_{z_0} should be significantly better than the inverse of the average pile-up density, 600 µm. Fig. 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 (p_{T}) at $|\eta| \approx 4$, due to the combination of geometric projection and, as shown in Fig. 2.7, increased material. As a result, ITk by itself cannot associate such forward tracks to correct vertices in

Pseudo-rapidity coverage	$2.4 < \|\eta\| < 4.0$
Thickness in z	75 mm (+50 mm moderator)
Position of active layers in z	$z = \pm 3.5$ m
Weight per endcap	350 kg
Radial extension:	
Total	110 mm $< r <$ 1000 mm
Active area	120 mm $< r <$ 640 mm
Pad size	1.3 mm $\times$ 1.3 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 $\times$ 2 cm)
Modules	8032
Collected charge per hit	> 4.0 fC
Average number of hits per track	
$2.4 < \|\eta\| < 2.7$ (640 mm $> r >$ 470 mm)	$\approx$2.1
$2.7 < \|\eta\| < 3.5$ (470 mm $> r >$ 230 mm)	$\approx$2.5
$3.5 < \|\eta\| < 4.0$ (230 mm $> r >$ 120 mm)	$\approx$2.7
Average time resolution per hit (start and end of operational lifetime)	
$2.4 < \|\eta\| < 4.0$	$\approx$ 35 ps (start) $\approx$ 65 ps (end)
Average time resolution per track (start and end of operational lifetime)	$\approx$ 30 ps (start) $\approx$ 50 ps (end)

Table 2.1: Main parameters of the HGTD.

an unambiguous way.

In addition to the spatial distribution of collisions, pile-up collisions will also be distributed in time, with a Gaussian width in the range 175 to 260 ps. The current baseline working point is 175 ps, which is used in the simulation of the HGTD performance studies. HGTD is designed to provide a time-of-arrival measurement for incident tracks with a resolution of 30 ps at the beginning of the HL-LHC, degrading to 50 ps at the end of the HL-LHC. After determining the vertex position using the ITk, this complementary time measurement is significantly more precise than the spread of the beamspot in time.

With the combination of ITk and HGTD measurements, ATLAS can view a portion of the event in space and time, extending the pile-up rejection capabilities of the ITk to the extent of its acceptance. This is one of the main motivations for the HGTD.

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 Chap. 5. With fast

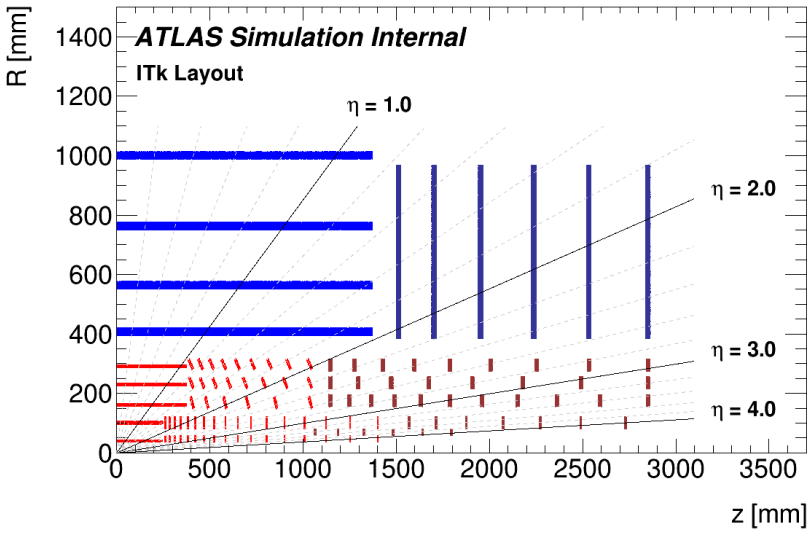


Figure 2.5: Schematic layout of the ITk for the HL-LHC phase of ATLAS. The active elements of the barrel and end-cap ITk Strip detector are shown in blue, for the ITk Pixel detector the sensors are shown in red for the barrel layers and in dark red for the end-cap rings. Here only one quadrant and only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

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 Chap. 6. The clock contribution is required to be below 15 ps; its distribution is discussed in more detail in Chap. 10.

For simplicity, the same pad size, 1.3 mm $\times$ 1.3 mm is used for the entire HGTD. This pad size balances several characteristics. For smaller pad 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 at the lowest instrumented radius, 120 mm, of less than 10%. This also ensures a low double-hit probability.

The layout of modules in each endcap was defined by maximising the coverage and minimising the effect of non-instrumented regions due to mechanical tolerances. In the second step the spacing between modules was defined.

The readout rows are sets of modules whose flex cables are guided together towards larger radii to the peripheral on-detector electronics. Their disposition for the first and last layer is shown as rectangles in Fig. 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. For $r > 150$ mm, the coverage

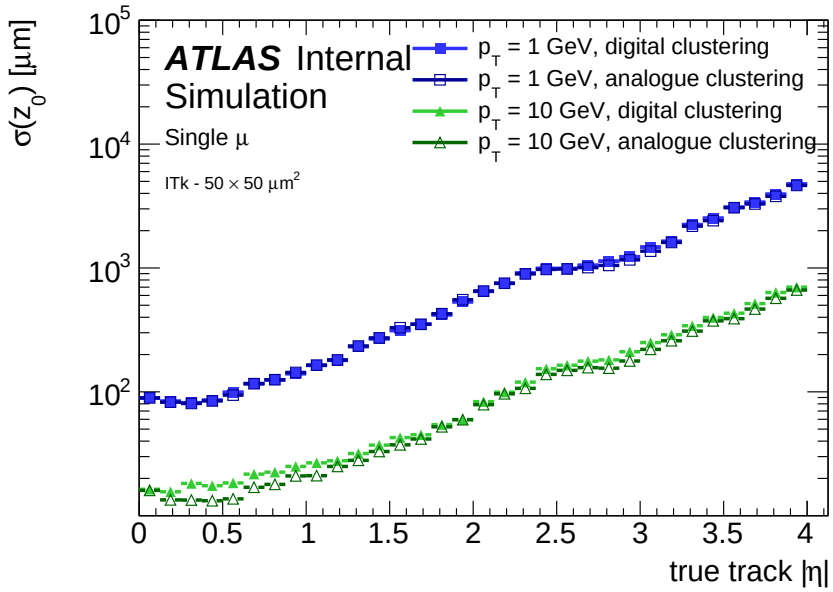


Figure 2.6: Resolution of the longitudinal track impact parameter, z_0, as a function of η for muons of $p_{\mathrm{T}} = 1\,\mathrm{GeV}$ and $p_{\mathrm{T}} = 10\,\mathrm{GeV}$. Additionally, the figure highlights the difference in resolution whether analogue or digital clustering is used. Studies for this report use digital clustering, which have roughly a 10 % degradation in z_0 resolution for $p_{\mathrm{T}} = 10\,\mathrm{GeV}$ muons.

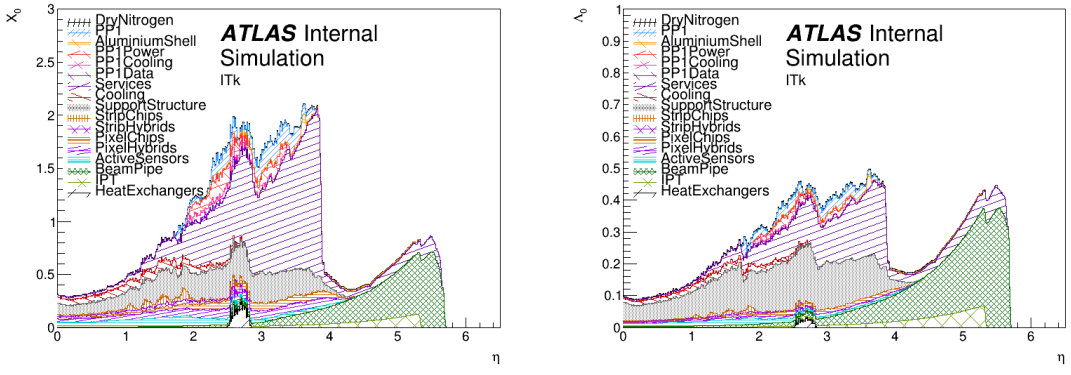
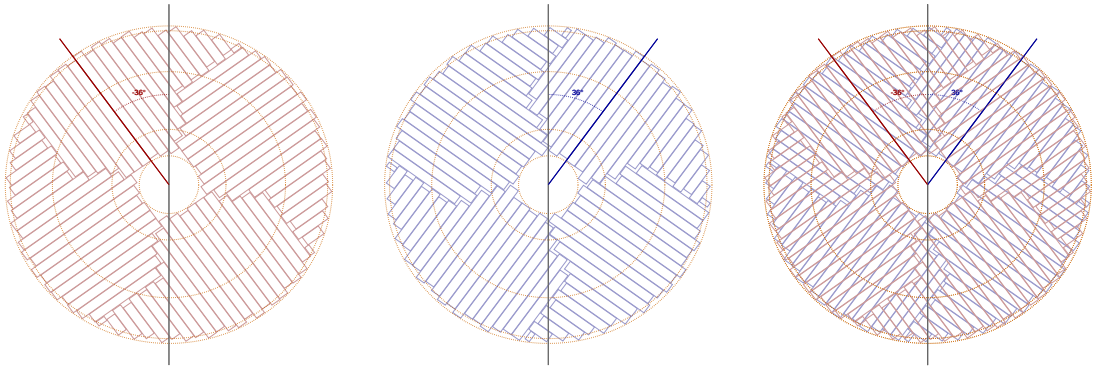


Figure 2.7: Radiation length X_0 (left) and nuclear interaction length λ_0 (right) as a function of pseudo-rapidity η, broken down by type of material for the ITk Layout [5] and beam pipe. *Preliminary, to be updated in new release*

is complete. The maximum length of the readout rows is limited by the manufacturing capabilities for the flexible circuits used for the data transmission. The non-instrumented zone is 0.5 mm between two readout rows for each row to account for mechanical tolerances. The effective width of a readout row is therefore 41 mm. These constraints lead to the helix structure shown in Fig. 2.8. A particle transiting the detector should encounter multiple

layers of LGAD sensors as it enters (encountering a "first layer") and exits (after encountering a "last layer"). Fig. 2.8(a) shows the geometry of the first layer and Fig. 2.8(b) shows the geometry of the last layer. The first and last layer are arranged to mirror the geometry of one another. Therefore the non-instrumented zones of the two disks will not overlap, except in the case of four readout rows per quadrant. Additionally, each of the layers is rotated in opposite directions by 72° as shown in Fig. 2.8(c). Any angle of rotation beyond 10° results in similar performance in terms of the number of hits and dead regions. The baseline angle is chosen largely due to detector services considerations, which are further discussed in Chap. 12 and Chap. 13. Along with optimising the coverage, the rotation frees sufficient room at 640 mm to install the cooling equipment between the peripheral electronics.



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

Figure 2.8: The orientation of the readout rows for the first and last layer encountered by a particle, separately and with the overlay of both. Each layer is rotated in alternating directions by 72°.

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 Sec. 2.4. The geometry of the HGTD is designed such that at $r < 230$ mm on average about 2.7 hits are obtained for a charged particle, for 230 mm $< r <$ 470 mm on average about 2.5 hits are obtained, whereas at $r > 470$ mm on average about 2.1 hits are expected to be associated to a track.

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 Fig. 2.9, the modules on the two sides of a disk are arranged to overlap so that the number of hits exceeds the number of disks. A study using full simulation was performed to determine the optimal overlap between modules in $r < 230$ mm to achieve the required timing resolution via the average number of 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 read out of the data. For $r > 470$ mm, an overlap of 20%, for 470 mm $> r >$ 230 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 hits as a function of radius and trasvenrse plane position is shown in Fig. 2.10. The relative fraction of tracks as a function of hits per track for each ring can be found in Fig. 2.11.

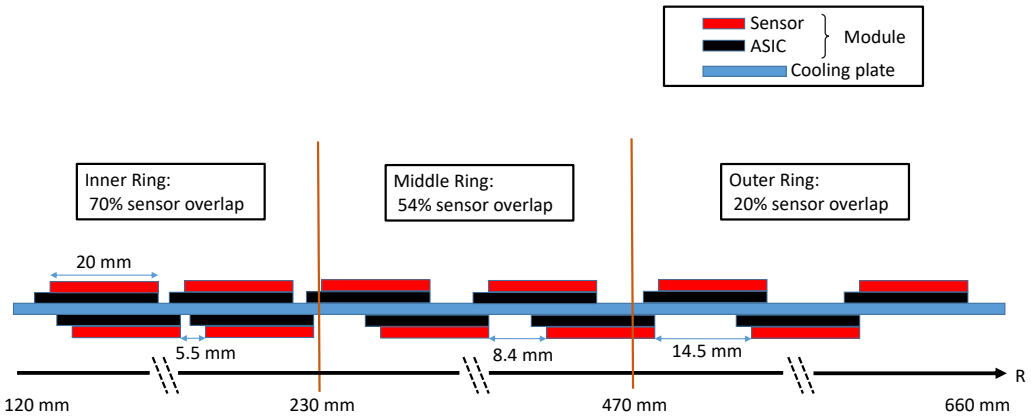


Figure 2.9: The schematic drawing shows the overlap between the modules on the front and back of a cooling disk. There is a sensor overlap of 20% for $r > 470$ mm, 54% for 470 mm $> r >$ 230 mm and 70 % for $r < 230$ mm.

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

Beyond pile-up mitigation, HGTD can play a key role in the HL-LHC physics programme as a luminometer. An accurate luminosity determination will be a critical input for cornerstone precision measurements. The luminosity uncertainty can be a limiting factor to many precision cross-section measurements, including achieving $\mathcal{O}(1\%)$ accuracy on certain measurements of Higgs boson production and couplings. It is therefore important to be able to determine the luminosity as accurately as in Phase-I, which will be a challenge with the harsh environment at the HL-LHC. For the technologies used traditionally for luminometers, the increased pile-up leads to increased detector occupancies, posing serious problems. The HGTD provides unique and pile-up–robust capabilities for measuring the luminosity at the HL-LHC.

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

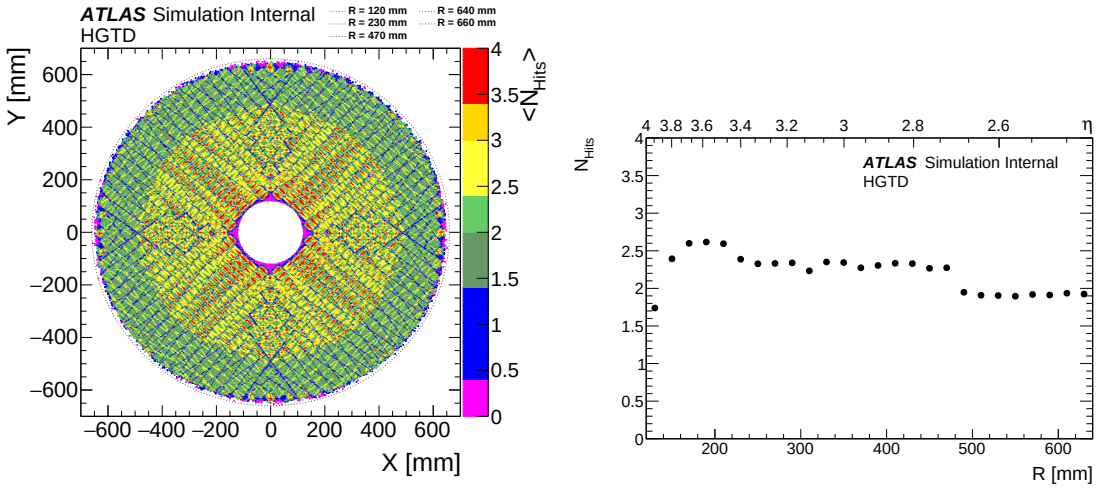


Figure 2.10: Hit multiplicity as function of X,Y and R.

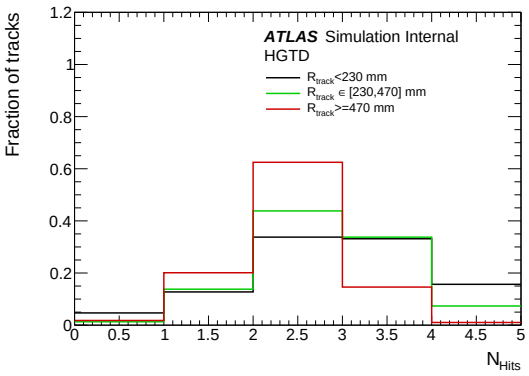


Figure 2.11: Fraction of tracks as a function of number of Hits in the track for tracks in the inner, middle and outer ring.

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 tunable with a step of 3.125 ns. The application of these capabilities and their implementation are further discussed in in Sec. 3.3.2, Chap. 6, and Chap. 10.

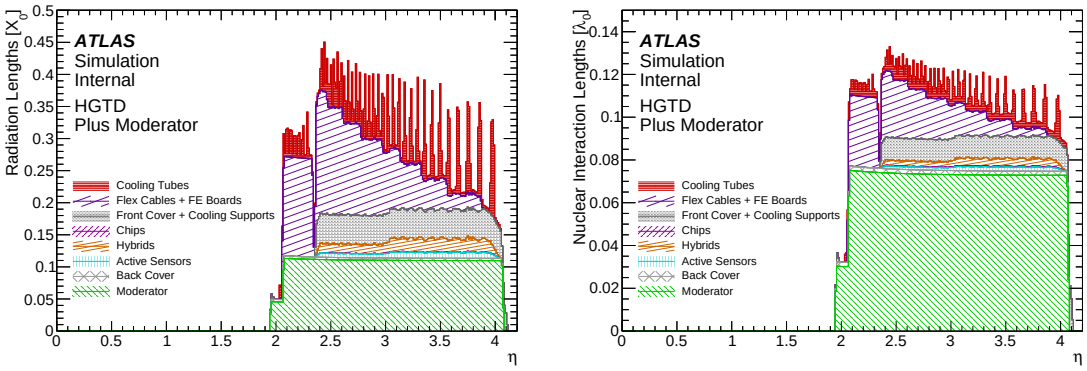
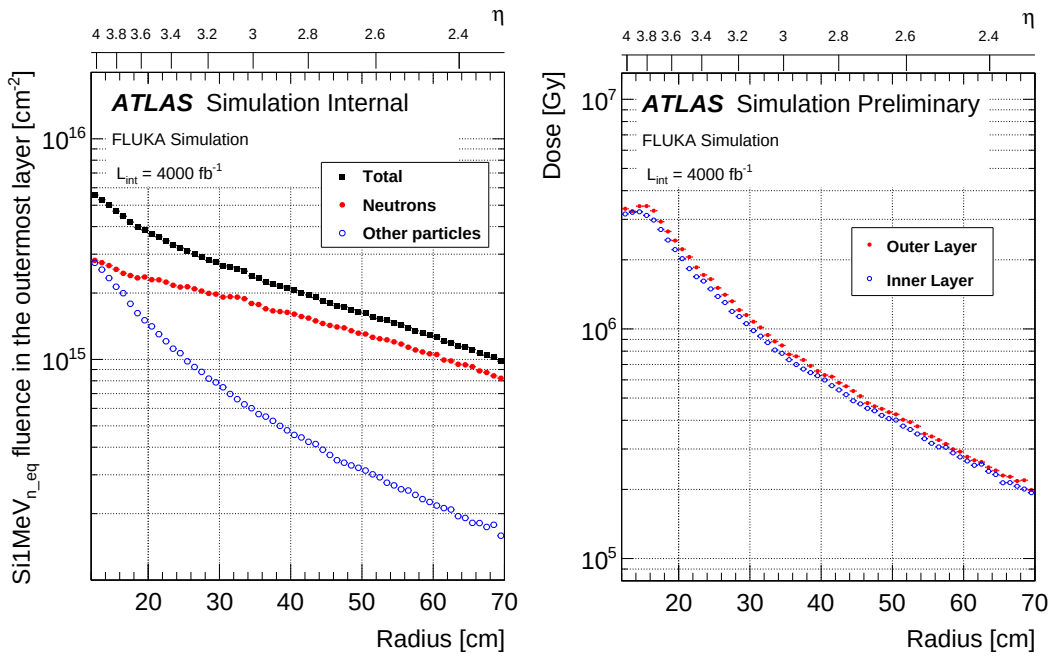


Figure 2.12: Radiation length X_0 (left) and nuclear interaction length λ_0 (right) as a function of pseudo-rapidity η, broken down by type of material for the HGTD. The moderator is included as it is within the hermetic vessel, although it is situated completed behind the active area of the HGTD. The baseline cooling pipes will be made with Titanium, which will reduce the peaks amplitude by almost a factor of 2.

2.4 Radiation hardness

One of the most important parameters of the HGTD will be radiation hardness of the sensors and electronics. Given that the HGTD will be installed with a pseudo-rapidity coverage of $2.4 < |\eta| < 4.0$, it is crucial that the detector can withstand the lifetime of the HL-LHC running. At the end of the HL-LHC (4000 fb^{-1}), the maximum nominal neutron-equivalent fluence at a radius of 120 mm, should reach 5.6×10^{15} n_{eq} cm^{-2} and the total ionising dose (TID) will be about 3.3 MGy, as shown in Fig. 2.13. To account for uncertainties in the simulation, a safety factor of 1.5 is applied to both numbers. An additional factor of 1.5 is applied to the TID due to uncertainties in the behaviour of the electronics after irradiation. 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 electronics which are more sensitive to the TID. After applying these, the detector would need to withstand 8.3×10^{15} n_{eq} cm^{-2} and 7.5 MGy.

This amount of radiation damage to lowest-radius (r < 230 mm) sensors and electronics suggests that this innermost part of the detector should be replaced after each 1000fb^{-1}and the sensors and electronics within 470 mm $> r >$ 230 mm should be replaced at half lifetime (2000fb^{-1}) of data-taking during the HL-LHC program. The plan is therefore to replace the sensors and ASICs located at a radius up to about $r >$230 mm three times and the sensors and electronics located within 470 mm $> r >$ 230 mm once. This corresponds to about 52% of the sensors and ASICs. Consequently, in the proposed 3 rings layout the maximal TID and fluence, using the Fluka estimations of September 2019, does not exceed 2 MGy and 2.5x10^{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, as seen in Fig. 2.13. The exact radial transition between the three rings



(a) Nominal $\mathrm{Si1MeV_{n_{eq}}}$ fluence for HL-LHC. (b) Nominal ionising dose for HL-LHC.

Figure 2.13: Expected nominal $\mathrm{Si1MeV_{n_{eq}}}$ fluence and ionising dose as functions of the radius in the outermost sensor layer of the HGTD for 4000 $\mathrm{fb^{-1}}$, 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).

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 Chap. 5 to Chap. 6. The maximum fluence and total ionising dose as a function of the radial position including the replacement of the rings can be found in Fig. 4.2. The expected proton, neutron, and pion energy spectra in the HGTD front and rear layer after 4000 $\mathrm{fb^{-1}}$ are shown in Fig. A.1, Fig. A.2, and Fig. A.3.

3 Performance and Physics Benchmarks

The precision time measurement capability of the HGTD enhances the performance for hard scatter jet tagging, missing transverse momentum $E_{\mathrm{T}}^{\mathrm{miss}}$, tagging b-jets, and lepton isolation in the forward region. Additionally the timing structure of non collision background (NCB) is analysed. The impact of these improvements on the sensitivity of a few selected physics analyses is presented. The HGTD timing capabilities will provide a luminosity measurement, a discussion on the impact of the precision of this measurement as well as the systematic errors are also included.

3.1 Simulation

The full simulation of the HGTD is performed using a software release dedicated to the ATLAS Upgrade program. The production of simulated samples follows the same steps as regular ATLAS simulation based on the Run 2 offline software chain: event generation, detector simulation, digitisation of simulated energy deposits into the actual detector read-out data format, and event reconstruction starting from the digitised data.

The simulation of the HGTD was implemented as close as possible to the layout foreseen. In the direction of the beam axis the HGTD implementation starts at 3420 mm. Behind the HGTD before the cryostat wall two moderators with a total thickness of 50 mm are implemented.

3.1.1 Detector geometry

The GEANT4 toolkit [6] is used to simulate the ATLAS detector. The simulation uses dedicated `GeoModel` packages to implement the detector geometry which is converted, via dedicated tools, to GEANT4 volumes. Particles are propagated through this geometry and the various physics processes, caused by their interaction with 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 as realistic as possible. The simulation step produces hits, i.e. small steps in the material with a starting point, an end point and the amount of energy deposited by the particle.

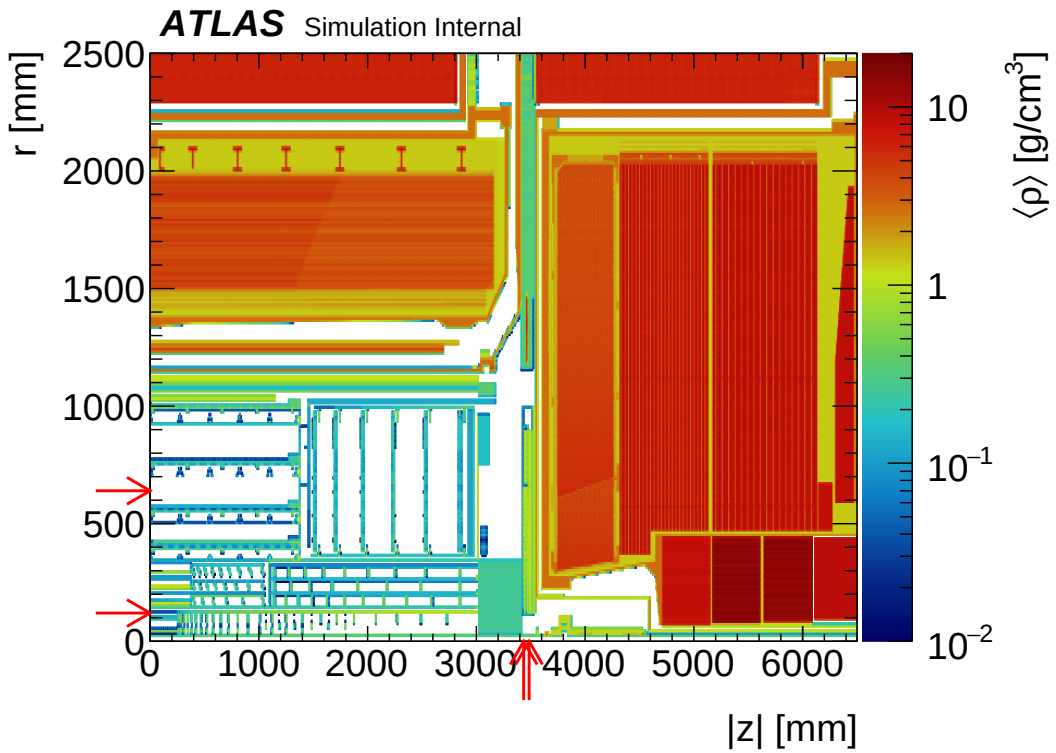


Figure 3.1: The material density is shown as function of z and radius. The extent of the active part of the HGTD in z and r direction, 500 mm and 520 mm respectively, is indicated by the arrows. TODO: This figure needs to be updated.

In Fig. 3.1, the material density of the upgraded ATLAS detector is shown as a function of the beam axis and the radius. The active part of the HGTD extends from 120 mm to 640 mm in radius.

There are four individual layers per endcap, the silicon sensors are mounted on both sides of the two cooling ensembles. The simulation of the HGTD has been extended to implement the peripheral electronics at radii greater than 640 mm. As an illustration a simulated event is shown in Fig. 3.2.

The simulation of the HGTD includes front and back covers and heaters of the HGTD. The simulation of the cooling plates, the CO_2 cooling loops as well as the support plates for the modules is also implemented. Volumes mimicking the peripheral electronics have also been implemented in the simulation.

The modules are the ASICs plus the sensors. These are simulated as boxes of size 22 mm $\times$ 40 mm with silicon sensors of size 20.5 mm $\times$ 40 mm perpendicular to the direction of the beam axis. The modules are slightly larger than the sensors in one dimension to account for

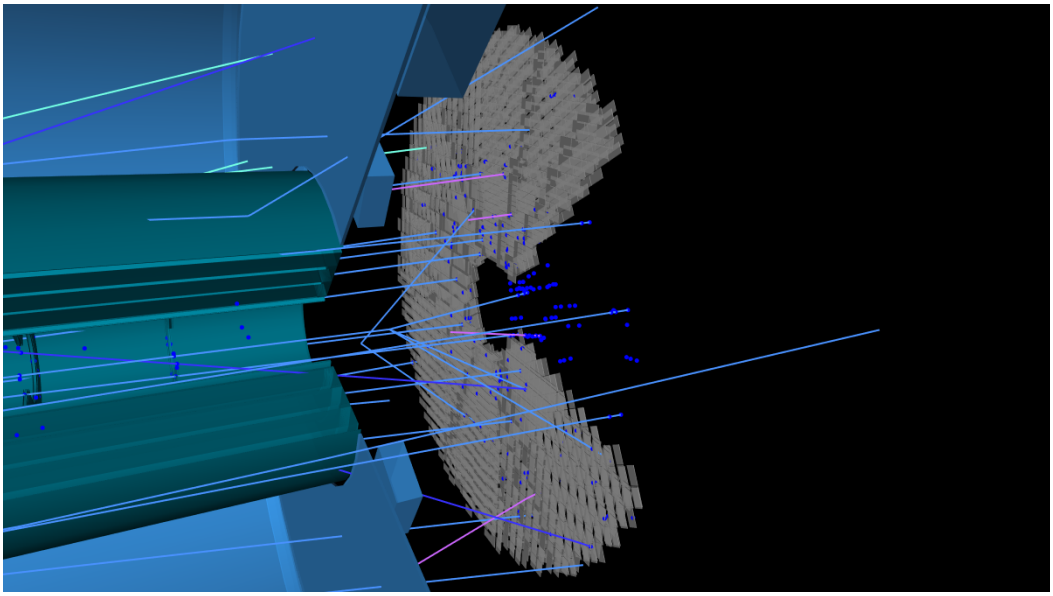


Figure 3.2: The simulation of an event in the HGTD is shown.

the wirebonding of the ASIC to the silicon sensor. The total thickness of the silicon sensor is 250 μm with an active thickness of 50 μm a passive thickness of 200 μm, corresponding to the LGAD sensor design. In total, 3992 modules are present in each end-cap. TODO: Check these numbers - do we use two-ring or three-ring layout here?

The Flex PCBs connecting the ASICs to the peripheral electronics beyond 640 mm have also been implemented in the simulation. This leads to an increase of the material distribution of the HGTD as the radius increases.

3.1.2 Sensor simulation

The baseline of 1.3 mm $\times$ 1.3 mm is used for the padsize. Two sources of inefficiency are taken into account at simulation time. The guard ring of 0.5 mm and the inter-pad dead zone of 50 μm are implemented. The active area is 79% of the total silicon area. The different zones of the sensors are illustrated in Fig. 3.3.

The digitization and clustering steps are implemented using silicon pixel hits and contains functionality for associating truth information to the simulated hits. The digitization step generates a detector signal with a pulse shape extracted from beam tests of LGAD sensors, adds expected electronic noise and is capable of describing the timing performance expected at various points during the HL-LHC program. The expected timing resolution depends on the radiation the detector has been exposed to, and this is implemented through with increasing integrated luminosity.

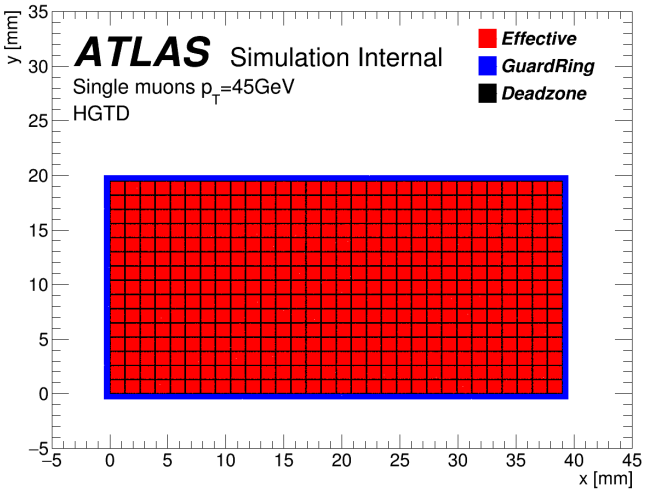
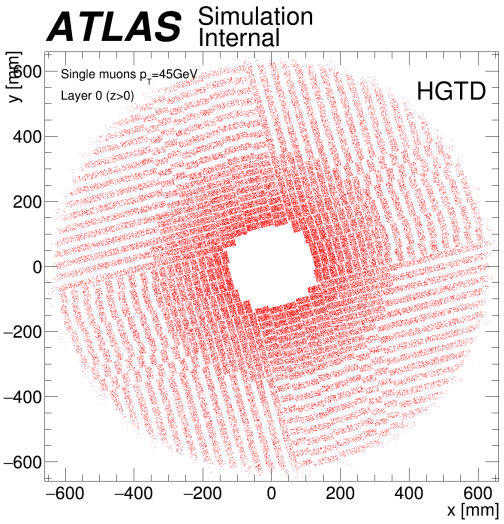


Figure 3.3: The simulation of the HGTD sensor is shown separated into the active area, the guard ring and the inter–pad deadzones using the simulation of single muons.

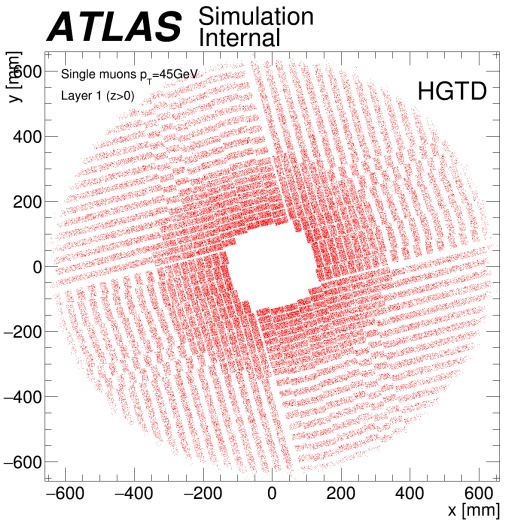
The active area of each pad is associated with a unique identifier. In the digitization step for $\langle\mu\rangle = 200$, the hits from the different interactions are summed in energy if they are in the same 5 ps time bin (TODO: Check to what extent this is true for SiHits). To allow for maximum flexibility at the analysis stage, the hits are then copied down to the format used for the analysis. The timing information is stored after subtracting a global offset of 11.6 ns, corresponding to the time of flight from the nominal interaction point to the center of the HGTD.

In Fig. 3.4 the hits in the HGTD are shown in the transverse plane for the two cooling plates with modules on the front and back for positive z. The position of the modules can be identified. The displacement of modules mounted on the back of a cooling plate with respect to those mounted on the front of each cooling plate is shown. The transition around 320 but outofdate from an overlap of 80% to 20% is visible as decreased density of sensors. The mirror symmetry between the first and second cooling disk as well as the rotation of 15° of each of the cooling disks is observed.

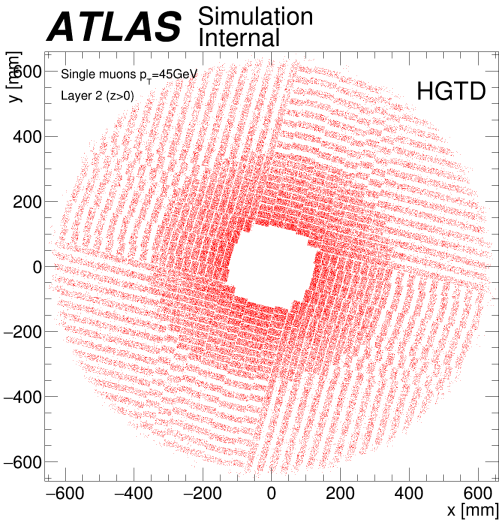
The simulation provides the energy deposit in the sensitive layer of the HGTD as single energy deposit for each particle traversing it. The simulation of the non-uniform distribution of the charges in the sensitive volume as well as the effect of the electronics chain (time walk, jitter) are taken into accçount at analysis level. For each hit a pulse is simulated to compute the time and energy in each pad. Data derived from the 2016 HGTD test beam were used to derive the pulse shape. A convolution of a Gaussian with a Landau distribution was found to give the best description of the pulse shape. The non-uniform energy deposit is modeled via the width of the Gaussian. The signal time is defined on its leading edge, therefore the



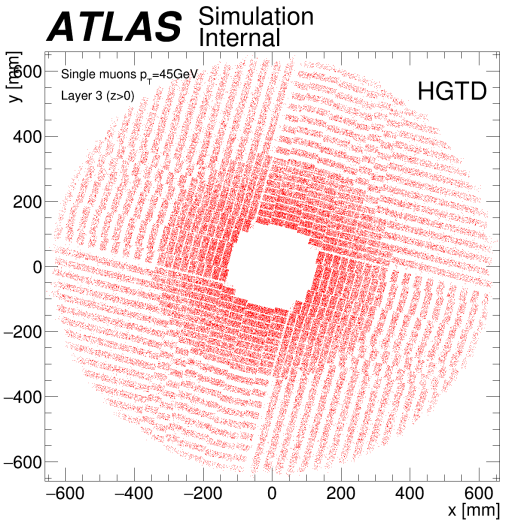
(a) First cooling plate front



(b) First cooling plate back



(c) Second cooling plate front



(d) Second cooling plate back

Figure 3.4: The positions of simulated hits in the transverse plane for each of the four layers separately. Only hits in active parts of the detector elements are shown. The position of the modules, the mirror symmetry between the first and second cooling plate as well as the rotation are clearly visible. TODO: These plots need to be updated to be made with SiHits, or just replaced with the figures showing the placement of the modules.

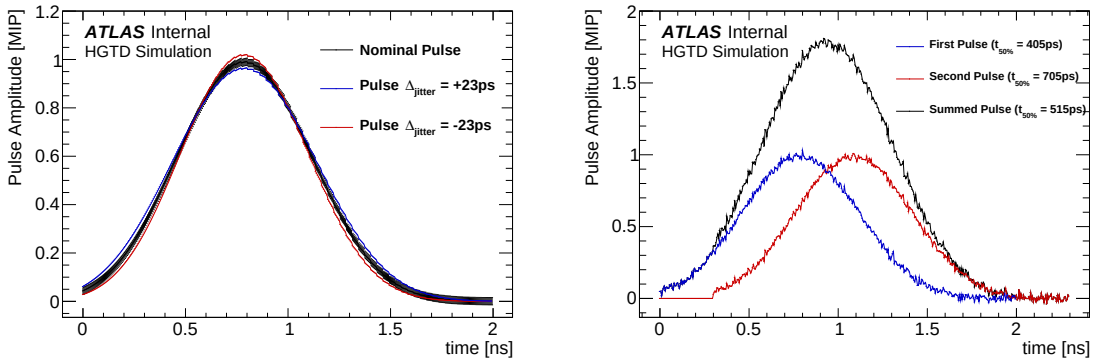


Figure 3.5: The simulated shape in time of the signal in a pad of the HGTD is shown for (a) a normal hit and (b) a double hit, separated by 300 ps.

variation models adequately the induced timing uncertainty. Fig. 3.5 shows the nominal shape and the effect of two hits in the same pad, separated by 300 ps.

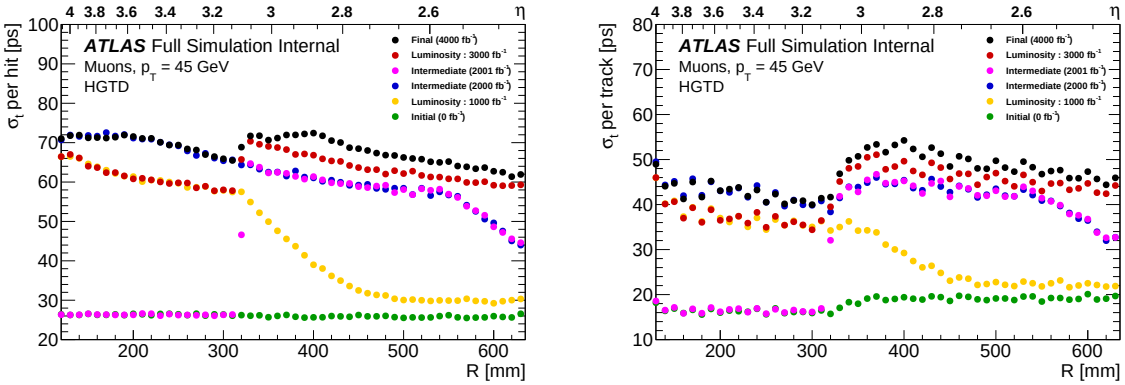
For each hit, a pulse is simulated with 200 points of a step size of 5 ps where the width of Gaussian contribution is driven by the desired timing resolution of the sensor. The maximal amplitude of the pulse is the deposited energy. The time corresponding to the first point of the pulse is chosen to be the time of the hit. Additionally a Gaussian noise with of 1.5% of the energy of a MIP (0.2 keV) is added to the amplitude in each time bin.

For each pad, the pulses are then summed together. A pseudo constant fraction discriminator (CFD) algorithm defines the time as the time of the first point with an energy above 50% of the maximum amplitude. Therefore the time of a pad is offset by 0.405 ns.

The contribution of electronic noise to the timing resolution is taken into account as a function of the position of the sensor and the accumulated integrated luminosity with a Gaussian smearing. The dose received by the sensor as a function of its radius was computed using FLUKA, then data from test bench measurements of sensors define the corresponding gain for the sensor. The gain is transformed into the timing resolution using measurements with ALTIROC0. This procedure results in a Gaussian smearing of minimum 10 ps and maximum 60 ps.

Timing Resolution Scenarios

Four timing performance scenarios are defined: *Initial*, two *Intermediate* scenarios and *Final* corresponding to integrated luminosities of 0 fb^{-1}, 2000 fb^{-1}, 2001 fb^{-1} and 4000 fb^{-1} which are shown in Fig. 3.6. The scenarios correspond to the performance expected at the beginning, after half of the expected integrated luminosity, after half of the expected



(a) Resolution per hit.

(b) Resolution per track.

Figure 3.6: The HGTD timing resolution is shown as function of the radius for four timing scenarios. The sensor resolution and the contribution from the electronics are considered, added in quadrature.

integrated luminosity after replacement of the inner part of the HGTD and at the end of the HL-LHC data taking. Two additional scenarios for 1000 fb^{-1} and 4000 fb^{-1} are also shown to illustrate the change of the resolution as the integrated luminosity increases. Maintaining this performance as the clock is distributed across the detector will require intercalibrating the reference t_0, discussed in Sec. 10.2.

As the integrated luminosity increases, the damage to the sensors and electronics caused by radiation will deteriorate their timing resolution inducing a dependence on the radial distance from the beam axis. The replacement of the inner part ($r < 320$ but outofdate) of the HGTD after half of the HL-LHC programme leads to identical performance of the timing resolution in this region for the *Intermediate* for 2000 fb^{-1} and *Final* scenarios. In the simulation the replacement has been implemented only as function of the radius so only pads at a radius below 320 but outofdate are replaced. This is conservative as the entire module will be replaced which will lead to an improved performance above 320 but outofdate. Fig. 3.6 has bins of 10 mm so that after the replacement the transition region at 320 but outofdate has contributions from both regions. The timing resolution for a hit at a radius of 120 mm is expected to be degraded to the order of 70 ps after 2.5×10^{15} n_{eq} cm^{-2}, including the replacement.

The resulting timing resolution for tracks is shown in Fig. 3.6(b) for the four scenarios as well as for the two additional scenarios of 1000 fb^{-1} and 4000 fb^{-1}. At radii lower than 320 but outofdate ($|\eta| \approx 3.1 but out of date$!), the increased number of hits compensates the effects of radiation damage such that the resulting timing resolution is fairly independent of the radius. The per-hit resolution was implemented in full simulation. The reconstructed timing resolution for tracks shown here was determined using the single-muon events by comparing the reconstructed hit or track time to the expected time from the truth information. The hits closest to the extrapolation of the reconstructed muon track within a window of 1.4 mm are

Process	Generator	N events, $\langle\mu\rangle = 0$	N events $\langle\mu\rangle = 200$
Minimum Bias, high-p_{T}	Pythia8	10000000	
Minimum Bias, low-p_{T}	Pythia8	1000000	
Single π^+, $p_{\mathrm{T}} = 5$ GeV, flat η [2.3-4.3]		200000	200000
Single π^+, $p_{\mathrm{T}} = 20$ GeV,		200000	200000
Single π^+, $p_{\mathrm{T}} = 45$ GeV,		2000000	2000000
Single π^+, flat η [2.3-4.1], flat p_{T} [0.1-5.0]GeV		200000	200000
Single π^0, flat η [2.3-4.1], flat p_{T} [0.1-5.0] GeV		200000	200000
Single γ, $p_{\mathrm{T}} = 20$ GeV, flat η [2.3-4.3]		200000	200000
Single γ, $p_{\mathrm{T}} = 45$ GeV, flat η [2.3-3.2]		200000	200000
Single γ, $p_{\mathrm{T}} = 100$ GeV, flat η [2.3-3.2]		50000	50000
Single μ, $p_{\mathrm{T}} = 45$ GeV, flat θ		400000	400000
Single μ, $p_{\mathrm{T}} = 45$ GeV, flat η [2.3-3.2]		300000	300000
Single μ, $p_{\mathrm{T}} = 45$ GeV, flat η [3.2-4.3]		100000	100000
Single electron, $p_{\mathrm{T}} = 45$ GeV, flat η [2.3-4.3]		400000	400000
Single electron, $p_{\mathrm{T}} = 20$ GeV, flat η [2.3-4.3]		200000	200000
NCB beam–gas, oxygen		400000	
NCB beam–gas, carbon		400000	
NCB Beam–gas, hydrogen		400000	
$Z \to ee$	Powheg+Pythia8	100000	100000
$Z \to \tau\tau$	Powheg+Pythia8	400000	400000
$t\bar{t}$	Powheg+Pythia8	1000000	1000000
VBF $H \to ZZ \to 4\nu$	PowhegPythia8	500000	500000
Dijet production, X GeV $\hat{p_{\mathrm{T}}}$ <Y GeV	Pythia8	1000 000	1000000
Dijet production, X GeV $\hat{p_{\mathrm{T}}}$ <Y GeV	Pythia8	1000 000	1000000
Dijet production, X GeV $\hat{p_{\mathrm{T}}}$ <Y GeV	Pythia8	1000 000	1000000

Table 3.1: Monte Carlo events simulated, digitized and reconstructed with Athena releases AtlasProduction-20.20.14.4-6 using the Step 3.1 ITk geometry (geometry tag ATLAS-P2-ITK-17-04-02).

used in this study. The distribution is dominantly Gaussian with negligible tails.

Production

The simulation, digitisation and reconstruction was implemented in the ATLAS upgrade software releases 20.20.14.1 and 20.20.14.2 deployed on the grid. Samples of single particles, electrons, muons and pions as well as non–collision background (NCB) and selected physics processes such as $t\bar{t}$, VBF $H \to Z(\nu\nu)Z(\nu\nu)$ and monopoles were produced using the ATLAS production system. Pythia8 [7] was used together with Powheg [8–10] for most of the samples. In the simulation step the beamspot was simulated with the spread in z and t. Samples with $\langle\mu\rangle = 0$ as well as $\langle\mu\rangle = 200$ were processed. A summary of the samples is shown in Tab. 3.1. For the minimum bias (inelastic collisions in the underlying event) samples, a signal set of single neutrinos was used at $\langle\mu\rangle = 200$.

3.1.3 Simulation using silicon hits

Digitisation

The Geant4 simulated hits are processed in a digitisation step in order to emulate the detector electronics output. THe HGTD LGADs are simulated adapting the ATLAS ITk pixel offline software. They are described as planar n-in-p pixel sensors with electron carriers. The channel efficiency is simulated as perfect for hits above threshold, also without simulated defects. As described in Sec. 3.1.2, the HGTD sensors are implemented to have pixels with a size of 1.3 mm $\times$ 1.3 mm, a 0.5 mm thick guard ring and inter-pixel dead zones of 50 μm. The active thickness of the silicon sensors is 50 μm.

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 accordingly to the sensor thickness, carries mobility, depletion and bias voltage and Lorentz shift. The front-end electronics in-time threshold is set to 600 electrons, with an intrinsic standard deviation of 40 electrons added in quadrature with a noise standard deviation of 75 electrons. Given the characteristics of the sensors, the capacitive coupling to nearby pixels is considered negligible as studied in LGAD beam tests and discussed in Chap. 5 and cross-talk effects are not simulated.

The time measurement associated to each energy deposition in the sensor is obtained smearing the time of the Geant4 interaction in the silicon with a sensor time resolution of 30 ps. At this point, the algorithm estimates the total charge released in each pad and checks if it is above threshold. Pads above threshold are labelled as fired.

The time structure of the energy depositions in the active area of HGTD sensors in the first layer is shown in Fig. 3.7 for the Initial timing scenario. The time distribution is obtained subtracting for each deposit the expected time of arrival accordingly to the position of the sensor and assuming a straight line trajectory from the center of the detector. For all layers, the deposits originating from primary and secondary particles are in time for the bulk of the distribution. However secondaries create also a pronounced tail in the timing distribution. In the following, only fired pads characterised by energy depositions in $[-1.0, +1.0]$ ns time window around the expected time of arrival are taken into account. The first energy deposition determines the time measurement to be associated to each of them.

Clustering

The first step of event reconstruction is the formation of clusters from the individual channels of the HGTD. For silicon-based detectors, like the ITk pixel and strip detectors, this is a local pattern recognition step where adjacent readout channels are grouped together in clusters,

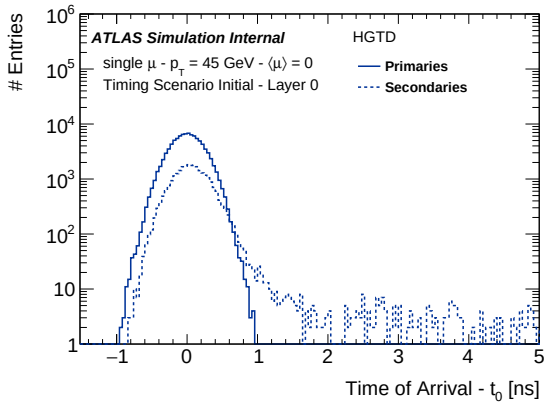


Figure 3.7: The distribution of the time of the hits with respect to the truth time is shown for hits originating from primary and secondary particles from a single muon sample without pile-up.

which represent single position measurement. This is done using a connected component analysis which is chosen to be based on eight-cell connectivity [ref.].

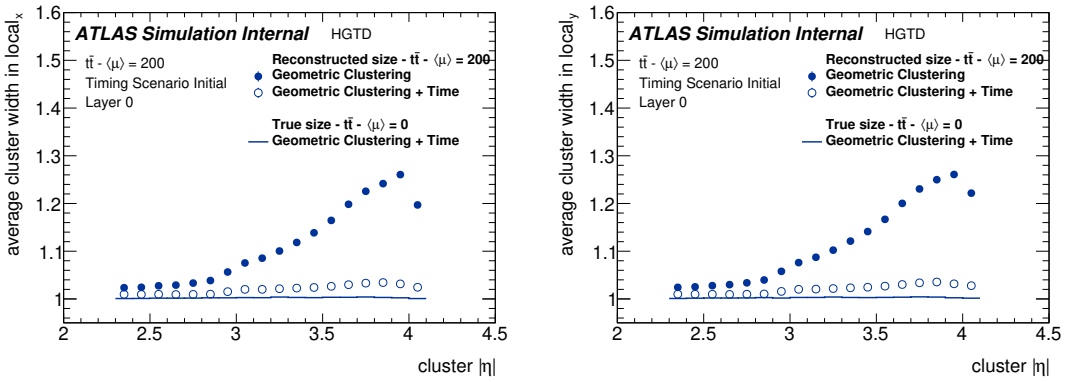
For the HGTD three different clustering approaches have been studied in order to profit of the time measurement associated to the individual channels and the bigger sensor size, compared to the pad size:

- Geometric clustering: as in the ITk pixel reconstruction, this approach groups adjacent fired pads to form clusters neglecting the time measurement associated to them. The local position of the cluster is calculated by taking a simple center of gravity of all the pads in the cluster.
- Geometric clustering with time filtering: the time measurement is used to cluster together adjacent fired pads. These are grouped if the time difference of the considered channels is smaller than 30 ps. The local position of the cluster is calculated as in the previous case, while the average of the time measurements to the pads in the cluster determines the time of the cluster.
- Single-pad clustering: each fired pad is converted into a cluster. The centre and the time measurement of the fired pad define the local position and the time of the cluster, respectively.

While the single-pad clustering is always providing clusters with dimension 1 along both local coordinates axis, the other two approaches can give reconstructed cluster sizes that differ from the truth ones. Fig. 3.8 shows the reconstructed cluster size in local_x and local_y[1] using the geometric clustering algorithm with and without time filtering for $t\bar{t}$ Monte Carlo events with 200 pile-up. The reconstructed cluster size is compared to the truth width of

[1] The local_x and local_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.

clusters reconstructed with the geometric clustering with time filtering for $t\bar{t}$ events without pile-up. The large cluster size at higher pseudo-rapidity provided by the geometric clustering algorithm is due to the higher sensor occupancy. When time filtering is required, smaller clusters are obtained with an average size close to one pad in both local_x and local_y. Fig. 3.9 shows the 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 200 pile-up. In order to correctly associate clusters to tracks and provide good timing measurements, one wants that the rate of merging multiple contributions into one clusters stays as low as possible. Therefore, taking into account the results described above and shown in Fig. 3.8 and Fig. 3.9, single-pad clusters have been considered the optimal collection as input to the tracking reconstruction.



(a) Average cluster width in local_x (b) Average cluster width in local_y

Figure 3.8: Average cluster width in local_x (a) and local_y (b) for clusters obtained with the geometric clustering algorithm with and without time filtering in $t\bar{t}$ events with 200 pile-up. The reconstructed cluster sizes are compared to the truth widths of clusters reconstructed with the geometric clustering with time filtering for $t\bar{t}$ events without pile-up.

3.2 Detector performance

As a first step the signatures in the detector are studied. This is followed by the discussion of the track level performance and the performance gains for individual objects.

3.2.1 Detector response characterisation

The HGTD will be installed at a distance of 3420 mm in z from the nominal interaction point. Particles produced at the interaction point will traverse the ITk and the material in front of the HGTD sensors. The HGTD provides signals at four distinct positions in z each with less

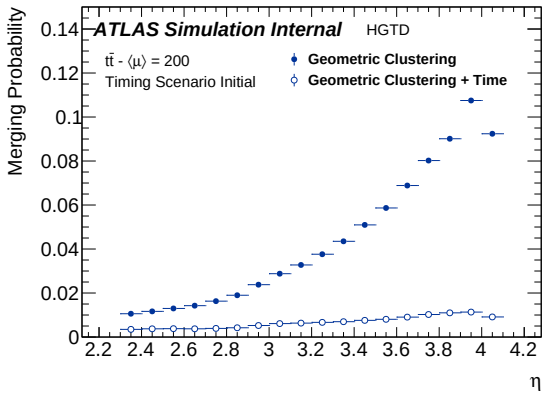


Figure 3.9: Probability of merging contributions originated by multiple particles in the same cluster with the geometric clustering algorithm with and without time filtering in $t\bar{t}$ events with 200 pile-up.

than 100% coverage of the transverse plane. Samples of single muons, pions and minimum bias events with $\langle\mu\rangle = 200$ have been simulated in order to calculate the expected number of hits per track as function of the incident particle. The effect of the material upstream and in the HGTD on incident electromagnetically and hadronically interacting particles as well as the percentage of pads with signal at high pileup was studied.

	$r < 320\,\text{mm}$ ($\lvert\eta\rvert > 3.1 but out of date!$)	$r > 320\,\text{mm}$ ($\lvert\eta\rvert < 3.1 but out of date!$)
$N_{\text{hits}} \geq 2$	88%	72%
$N_{\text{hits}} = 0$	1.6%	2.8%
$\langle N_{\text{hits}} \rangle$	2.7	1.9

Table 3.2: The percentage of tracks with at least two hits, the percentage of tracks escaping undetected and the average number of hits for muons with a p_{T} of 45 GeV are shown for the baseline detector layout. The values include the effect of non-instrumented zones, inter-pad dead-zones and guard ring as well as the track-matching efficiency.

The hit efficiency is studied using single-muon events with $p_{\text{T}} = 45\,\text{GeV}$. The distribution of the muons is flat the polar and azimuthal angles. The extrapolation of the tracks to the HGTD was performed using the last measured point of the track in the ITk. Only hits within 1.4 mm of the result of the extrapolation are accepted. This is only slightly larger than the pad size. This ensures that the pad extrapolated to and a neighboring pad are candidates for the track matching. For only 5% of the hits the difference between the extrapolated position and the position of the closest hit is larger than the acceptance criteria. At most one hit was accepted for each sensor plane, so that the maximum number of hits is four. Fig. 3.10 shows the average number of hits per muon in the transverse plane. The results are summarised in Tab. 3.2. 79% of the extrapolated muon tracks have at least two matched hits in the HGTD.

For at most 2.4% of the muons, no hits are registered within the acceptance window around extrapolated position. The effect of uninstrumented zones and inactive areas on the sensor such as the guard ring and the inter-pad dead-zones are taken into account.

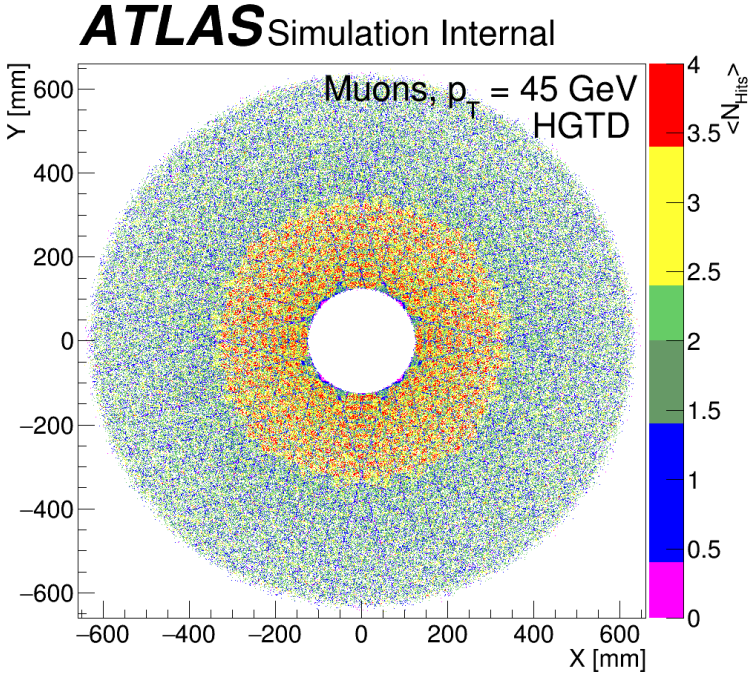


Figure 3.10: The average number of hits as a function of the position in the HGTD is shown for the baseline layout. The overlap is 80% at $r < 320$ but outofdate and 20% at larger radii.

The average number of hits is shown as a function of the radial distance from the beam axis in Fig. 3.11. The overlap of 20% between the modules at $r > 320 but out of date$ leads to an average number of hits of 1.9. The overlap 80% for $r < 320 but out of date$ results in an average hit multiplicity of 2.7 in this region in agreement with the requirements listed in Tab. 2.1.

In Fig. 3.12 the two dimensional distribution shows the number of hits as function of $|\eta|$. The profile of the histogram shows that the number of hits exceeds the number of hits expected by the study of the detector optimization with muons. The muons, as shown in Fig. 3.12(a), only interact electromagnetically leading to a low multiplity with an approximately Gaussian distribution. For pions, generated with flat distribution in transverse momentum between 1 GeV and 5 GeV, Fig. 3.12(b), the hit multiplicity has long tails due to hadronic interactions. When restricting the hits to those within 1.4 mm of the extrapolated hit position, a similar behavior in η dependence and magnitude is observed for muons and pions. This shows that interactions in front of the HGTD are likely, creating showers, as shown by the profile histogram in Fig. 3.12(b) for a distance of 3.5 mm, leading to a number of hits larger than expected for a single hit per sensor layer.

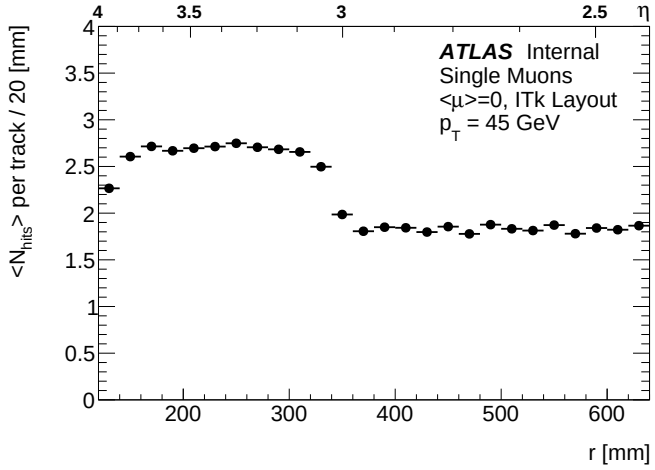
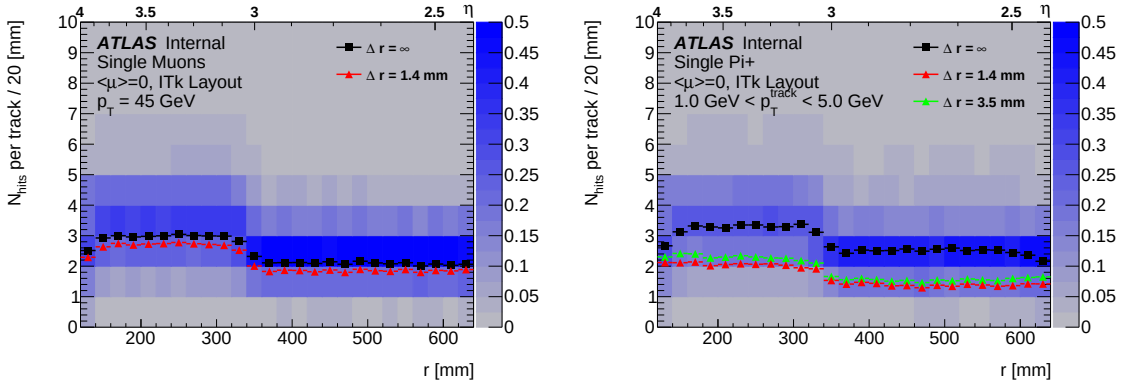


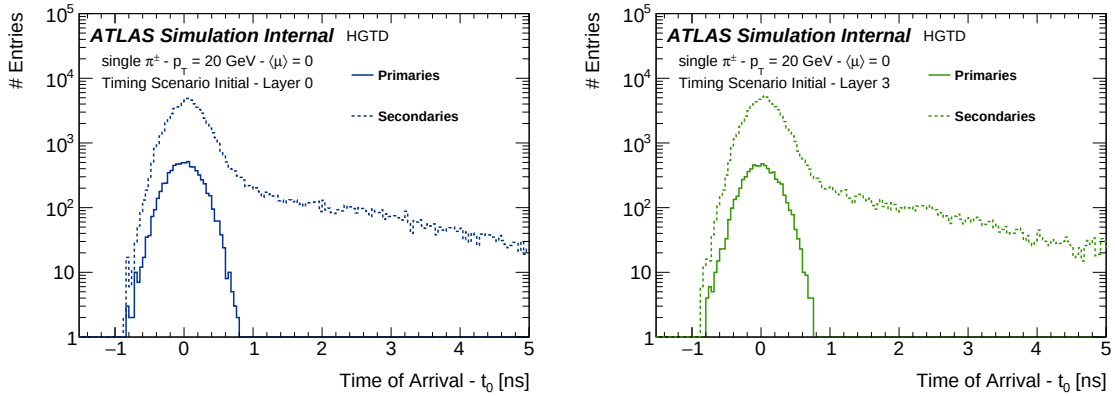
Figure 3.11: The average hit multiplicity as a function of the radius (and pseudo-rapidity) is shown for the baseline HGTD layout of 80/20% overlap below/above 320 but outofdate.



(a) Single Muons

(b) Single Pions

Figure 3.12: The distribution of the number of hits as function of radius (and $|\eta|$) is shown as a 2D histogram with its profile (black). Each $|\eta|$ bin is normalised separately. For the profile histograms (red, green) the hits are required to be within a distance of the expected hit position.

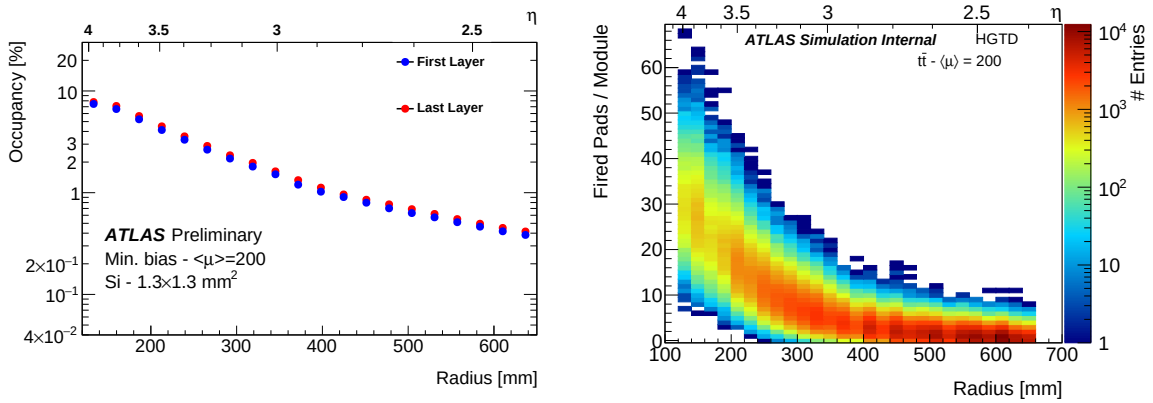


(a) Layer 0 (b) Layer 1

Figure 3.13: The distribution of the time of the hits with respect to the true time of arrival is shown for hits originating from primary and secondary particles.

The time structure of the hits in the first and last layer is shown in Fig. 3.13 for the *Initial* timing scenario. The structure is similar in all layers. The hits originating from primary and secondary particles are in time for the bulk of the hits. However the secondaries create a pronounced tail in the timing distribution.

The probability to have a hit in a pad decreases as function of the distance from the beam axis. The fixed pad size of the detector has to ensure a maximal occupancy of less than about 10% at the lowest instrumented radius of 120 mm. This ensures a low double-hit probability.



(a) Occupancy (b) Pads per module

Figure 3.14: The occupancy 3.14(a) and the number of fired pads 3.14(b) per module are shown as a function of the radius for a pad size 1.3 mm × 1.3 mm at a pile-up of $\langle\mu\rangle = 200$.

In Fig. 3.14(a) the occupancy expected for a pile-up of $\langle\mu\rangle = 200$ is shown, defined as

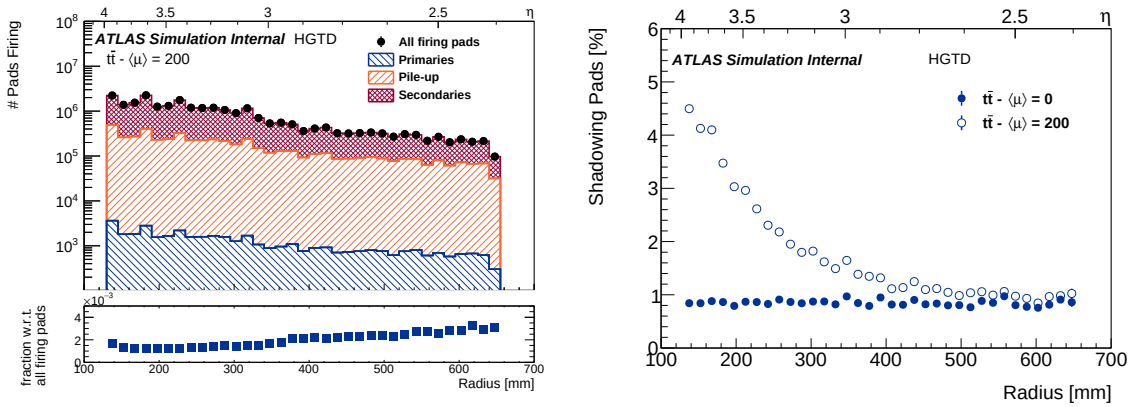
the percentage of pads of the HGTD registering a hit, for the HGTD baseline pad size of 1.3 mm $\times$ 1.3 mm. As expected, the occupancy decreases as a function of radius. A slight increase is observed when moving outwards from the innermost to the outermost layer, primarily due to the increased probability of initiating showers due to hadronic interactions as more material is traversed. At the smallest radius the occupancy is 8%, fulfilling the requirements.

The occupancy in Fig. 3.14(a) is calculated from a single bunch crossing at $\langle\mu\rangle = 200$. The contribution from preceding bunch crossings has been estimated by calculating the occupancy for the tails in the timing distribution, i.e., taking into account only hits more than 2 ns later than the average time of flight. A more accurate modeling of hits arriving late was obtained using FLUKA, additionally taking into account the induced nuclear radioactivity. The time measurement is active only during 5 ns out of 25 ns. Given this, the relative contribution to the occupancy is estimated to be of the order of a percent.

In Fig. 3.14(b) the average number of pads in a module with signal is shown as function of the radius for a sample of top pair production at $\langle\mu\rangle = 200$. The variation of the number of pads with signal in a module has to be taken into account in the calculation of the bandwith for the data transfer to the peripheral electronics.

Handling the Geant4 truth information it is also possible to study the types of the particle firing the pads. The breakdown of the different contributions is shown as function of the radius in In Fig. 3.15(a). In top pair production events with 200 overlaid pile-up interactions, pile-up particles and secondaries from showers in the upstream detector material dominate the occupancy. If a primary particle and another particle deposit energy in the same pad, the signal of the primary particle can be masked in the LGAD sensor if the other arrives earlier (shadowing). It is therefore important to evaluate the amount of primaries masked by the early arrive of other particles. Fig In Fig. 3.15(b) shows the percentage of pads fired by secondaries and pile-up particles shadowing a primary particle with respect to the number of pads where at least one contribution from a primary particle occurs in $[-1.0, +1.0]$ ns. For events with 200 overlaid pile-up interactions, the percentage of shadowed pads is 4.5% at low radius, where the occupancy is maximal, decreasing to 1% at larger radius. Performing the same analysis for $\langle\mu\rangle = 0$ shows that the level of 1% is due to particles originating from the same primary interaction and characterised by a time of arrival compatible within the timing resolution.

In Fig. 3.15(a) the breakdown of the origin of the hits detected in the HGTD within a time window of total width 2 ns centered on the time of the primary is shown as function of the radius. In a $t\bar{t}$ sample at $\langle\mu\rangle = 200$ secondaries and pileup dominate the occupancy. If a primary particle and a pileup particle deposit energy in the same pad, the signal of the primary particle can be deformed. In Fig. 3.15(b) the percentage of pads fired by secondaries particles and pileup shadowing a primary particle with respect to the number of pads in which a primary particle has deposited energy within the 2 ns time window around the expected time of arrival. The distribution is shown as a function of the radius. For $t\bar{t}$ events at



(a) Signal origin

(b) Shadowing percentage

Figure 3.15: The origin of the hits 3.15(a) detected in the HGTD and the percentage of shadowed pads 3.15(b) are shown as function of the radius.

$\langle\mu\rangle = 200$, the maximum is 4.5% at low radius, where the occupancy is maximal, decreasing to 1% at larger radius. Performing the same analysis for $\langle\mu\rangle = 0$ shows that the level of 1% is due to secondaries and primaries arriving with a time of arrival compatible within the timing resolution. In these cases the shadowing effect does not bias the time measurement.

3.2.2 Track-level performance

Track-to-vertex association

The precise assignment of tracks to primary vertices (track-to-vertex association) is one of the key elements to mitigate the effects of pile-up on the full suite of event reconstruction algorithms at hadron colliders. Jet reconstruction and calibration, pile-up 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 is determined by the resolution on the track z_0 impact parameter such that

$$\frac{\left|z_0 - z_{\text{vertex}}\right|}{\sigma_{z_0}} < 2.5, \tag{3.1}$$

where σ_{z_0} is the per-track resolution on the longitudinal impact parameter and depends primarily on the track η and p_{T}.

The reliability of the track-to-vertex association depends on the value of σ_{z_0} relative to the average pile-up density $\langle\rho(z)\rangle$. The average number of interactions within a window of $|z_0 - z_{\text{vertex}}| = 2.5\sigma_{z_0}$ is given by $N = 2\langle\rho(z)\rangle 2.5\sigma_{z_0}$. This means that in order to unambiguously associate tracks to vertices based on Equation. (3.1), N has to be smaller than 1, or $\sigma_{z_0} < 1/5\langle\rho(z)\rangle$. For a track with z_0 at the origin, where the average density peaks, this value is approximately 130 µm. If σ_{z_0} is larger than this value, the association of tracks to vertices becomes ambiguous because the same track may be compatible with multiple nearby vertices. It is important to note that this argument applies to prompt tracks such as those produced in light-quark and gluon jets, or prompt leptons. The association of displaced tracks from the decay of B/D hadrons to primary vertices requires the use of a larger z window, which will lead to greater pile-up contamination.

While the longitudinal impact parameter resolution is relatively constant and small for $|\eta| < 1.5$, it grows rapidly with pseudo-rapidity, reaching several millimetres for $|\eta| \gtrsim 2.5$. The resolution is further degraded for low p_{T} tracks due to multiple scattering effects. 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.

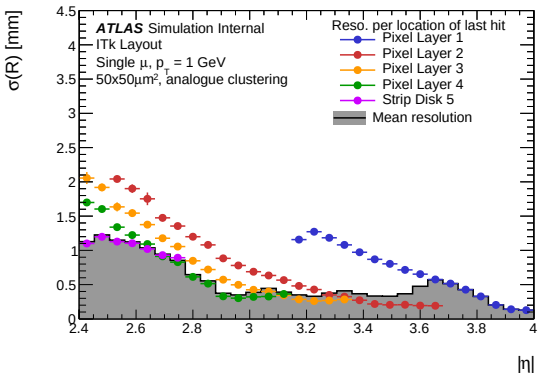
Based on Fig. 2.6, a 1 GeV track with $|\eta| = 3$ has a z_0 resolution of approximately 1 mm, leading to a $\pm 2.5\sigma$ window of 5 mm in z for the vertex position. With a most probable average pile-up vertex density of 1.8 vertices/mm at $z = 0$, this means that, on average, a forward track can be compatible with up to about 9 near-by vertices on average. Or, in other words, track-to-vertex association will suffer significantly from pile-up contamination.

Track Extrapolation to the HGTD

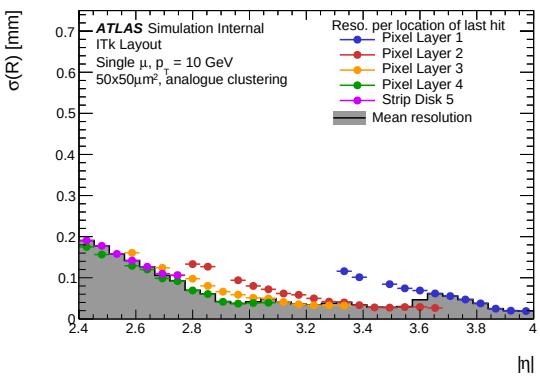
In order to associate the timing information provided by the HGTD to the correct track, the track has to be extrapolated from the ITk to the HGTD. The precision of the extrapolation is affected by the material in the ITk and between the ITk and the HGTD. In Fig. 3.16 the precision of the extrapolation as a function of η to the HGTD surface is shown for r and $r \times \phi$. 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 × 1.3 mm).

Track Extension

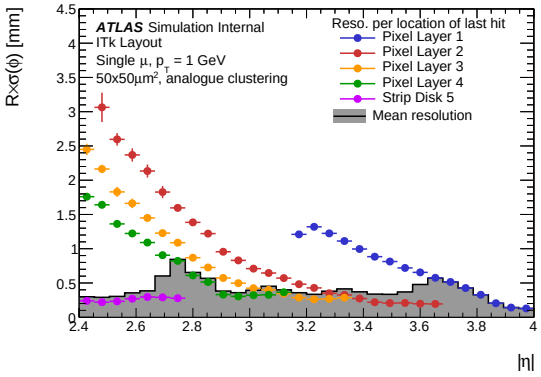
Track candidates reconstructed in the ITk are extended to the HGTD using a progressive Kalman filter within the HGTD acceptance. The starting point for the extension is the ITk layer closest to the HGTD within which a measurement has contributed to the track. From here, the tracks are extrapolated to each layer of the HGTD. In each layer, HGTD clusters found in a 5 cm × 5 cm window around the extrapolated crossing location are evaluated for compatibility with the track by attempting to add them to the track in a forward filtering



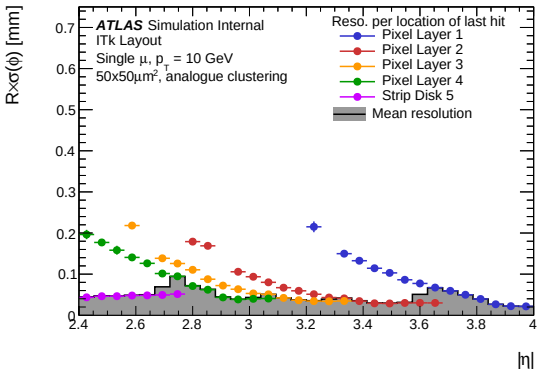
(a) Extrapolation resolution in radius for $p_{\mathrm{T}} = 1\,\mathrm{GeV}$



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



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

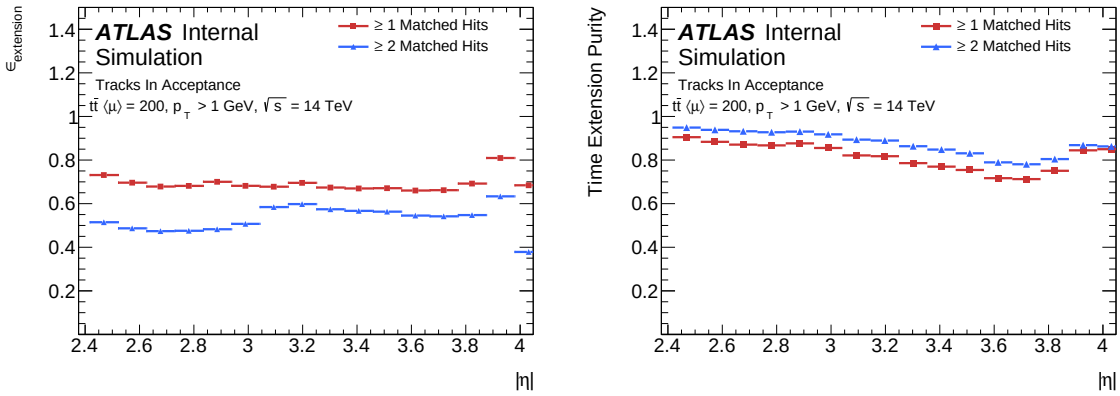


(d) Extrapolation resolution in radius$\times\phi$ for $p_{\mathrm{T}} = 10\,\mathrm{GeV}$

Figure 3.16: 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 resolution is similar to the size of a single pad in the HGTD.

step. The cluster with the lowest χ^2 resulting from this procedure is considered a a valid extension of the track into the given layer if the reduced χ^2 value is less than five. If no cluster satisfying this condition is found, no extension is registered for this layer and the procedure repeated in the next layer. In case of a successful extension, the updated track including the newly-associated HGTD cluster is used when extrapolating to further layers of the HGTD, replacing at each step the track information from the last measurement as a starting point of the extrapolation.

The inclusive extension efficiency is defined as the fraction of tracks that have a valid extension, considering one or two measurements in the HGTD that were produced by the truth particle matched to the track as shown in Fig. 3.17(a). In the sample shown, $t\bar{t}$ with $\langle\mu\rangle = 200$, the dense environment is a challenge to assign the correct HGTD hits to the



(a) Efficiency (b) Purity

Figure 3.17: The performance of the track extension to the HGTD is shown as function of η for $t\bar{t}$ with $\langle\mu\rangle = 200$.

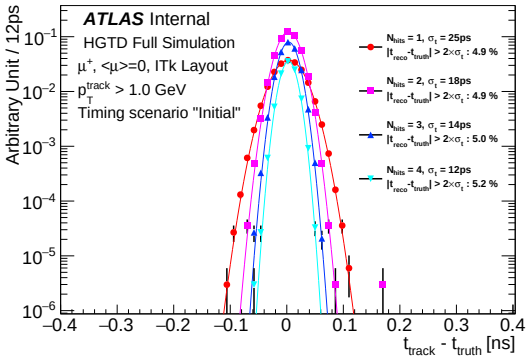
tracks. To estimate the number of wrongly assigned measurements on each extension, the purity, shown in Fig. 3.17(b), is defined as the fraction of correct assigned measurements over the total measurements on the track. For each track each hit associated to the track is classified according to its origin using the Monte Carlo truth link. The purity obtained in $t\bar{t}$ events for $\mu = 200$ is between 80% and 90%.

Track Timing Association

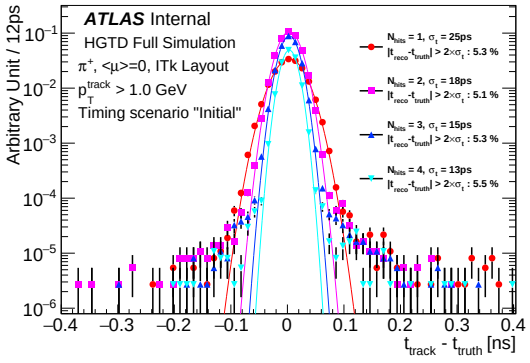
Associating a timing measurement to tracks reconstructed with the ITk using the HGTD hits was studied using single-pion samples ($\langle\mu\rangle = 0$, generated with a flat distribution in η and ϕ) and a physics sample with VBF-produced $H \to Z(\nu\nu)Z(\nu\nu)$ at $\langle\mu\rangle = 200$, both with full simulation of the HGTD.

The association of time information to a track depends on the precision of the track extrapolation. A track traverses the material in the ITk and the material between the ITk and the HGTD. Therefore the association is performed by extrapolating the tracks to the HGTD using the last measured point in the ITk as this method leads to a smaller error on the extrapolated position in the HGTD with respect to the extrapolation from the perigee. Only tracks that are reconstructed with a p_{T} greater the 1 GeV are extrapolated. Furthermore the extrapolation has to be within the acceptance of the HGTD, i.e., in radius between 120 mm and 640 mm. The timing scenario *Initial* has been used for the study.

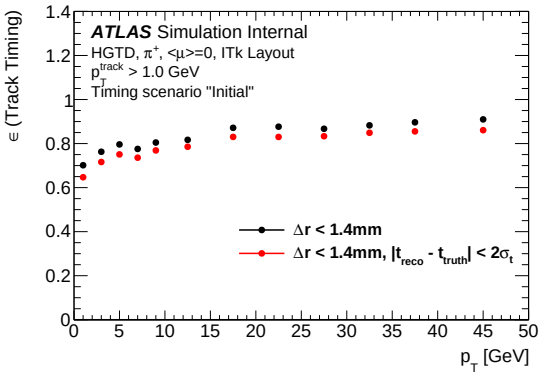
The single muon sample at $\langle\mu\rangle = 0$ is used in a first step to check the performance of the time association as function of the number of hits. In Fig. 3.18(a), the difference between the reconstructed and the expected time is shown. The distributions are Gaussian with a



(a) Muon $\langle\mu\rangle = 0$



(b) Single Pion $\langle\mu\rangle = 0$



(c) Single Pion $\langle\mu\rangle = 0$

Figure 3.18: The reconstructed timing resolution and efficiency are shown for different samples at $\langle\mu\rangle = 0$.

resolution compatible with the simulated resolution, shown in Fig. 3.6(a), divided by the square root of the number of associated hits.

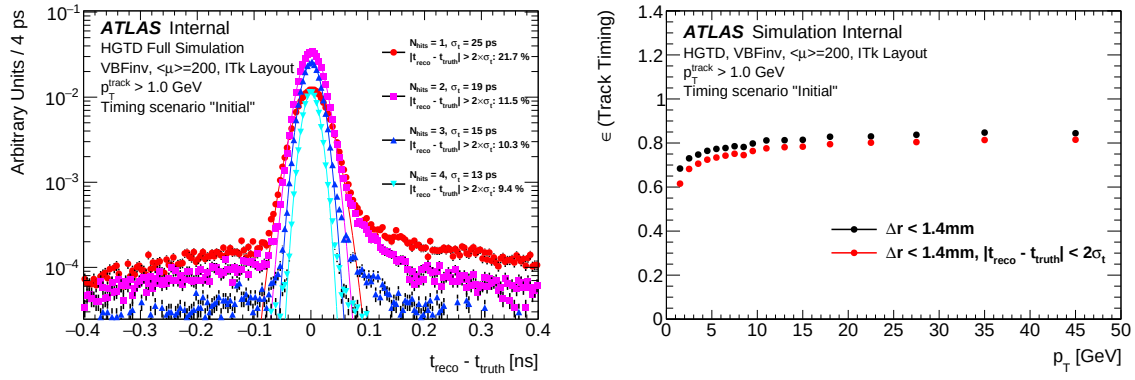
In the analysis of the single pion sample, only primary tracks are accepted by using the associated truth particle. If several hits are found in the pads that are within 1.4 mm of the track extrapolation, only the closest one is used. Fig. 3.18(b) shows the difference of the reconstructed track time and the expected time. The Gaussian core follows the expected scaling with the number of hits, as for the muons. The tails are increased with respect to the muons and the distribution is slightly asymmetyric as late secondaries (see Fig. 3.13) contribute to the reconstructed time.

The time association efficiency is shown Fig. 3.18(c) as a function of the reconstructed transverse momentum. Tracks with $|\eta| < 2.4$ or $|\eta| > 4.0$ are not considered in order to avoid border effects. If, instead of selecting only primary tracks the reconstructed track with the highest p_{T} is used, the efficiency to reconstruct the correct timing for the high p_{T} pions is

unchanged.

At the highest transverse momentum, the efficiency is 90%. At the lowest momentum, the efficiency is 70%. The window size of the time association of 1.4 mm is in the region where the efficiency increases only slowly for high p_{T}. For low p_{T}, the same efficiency can be reached only for a window of 30 mm. The multiple scattering effect in the material between the last measured point in the ITk and the HGTD is more important for low p_{T}. A window size 15 times larger (20 mm) would be necessary to reach an efficiency of 97% for high p_{T} pions. For low p_{T} pions, the same efficiency is attainable only for a window of size 200 mm.

While the geometrical acceptance is optimised to compensate for inactive zones in one layer with an active zone in another layer, the difference between track extrapolation and hit in different layers is strongly correlated. As an example, the correlation of the first layer with the last layer for the difference between hit and extrapolated position is more than 70%. The second set of points in Fig. 3.18(c) is the efficiency to associate a time within $2\sigma_{t_{\mathrm{reco}}}$ of the expected truth time. The ratio between the two efficiencies is about 0.95, independent of η and p_{T}.



(a) VBF $\langle\mu\rangle = 200$ (b) VBF $\langle\mu\rangle = 200$

Figure 3.19: The reconstructed time resolution and efficiencies for associating time to tracks is shown for pions in a VBF sample.

With a pile-up of $\langle\mu\rangle = 200$, the analysis of the track timing association is complicated by the presence of pileup. Time measurements in the HGTD have to be associated to tracks from the hard scatter while keeping the association of pile-up hits to a minimum. Therefore additional criteria are applied in the time association algorithm. If only one hit is found in the acceptance window, the time of the hit is kept as track time. If multiple hit candidates are identified, the hits are additionally required to be compatible in time with each other, i.e., within 2.5σ of the hit timing resolution around the time of the hit in the layer closest to the interaction point.

The least squares method is used to choose from all possible combination candidates.

$$S = \sum_i (\vec{x}_{\text{hit}_i} - \vec{x}_{\text{extrapolation}_i})^2 \tag{3.2}$$

$\vec{x}_{\text{hit}_i}$ is the hit candidate in the layer i, $\vec{x}_{\text{extrapolation}_i}$ is the position given by the track extrapolation in the same layer. The hit combination with the lowest value of S is chosen, and the mean time of the collected hits is assigned to the track.

The result of the algorithm is shown in Fig. 3.19(a). The difference between the reconstructed time and the expected time is well-described by a Gaussian core. The resolution improves as in the $\langle\mu\rangle = 0$ case with the square root of the number of hits. The tails are slightly asymmetric due to secondaries contributing to the reconstructed time. Since the algorithm can pickup up hits which are due to pileup, the tails increase. These tails decrease as function of the number of hits as the additional requirements on the timing consistency of the chosen hits are stronger than, e.g., in the case of one hit where no constraints can be applied.

The efficiency for correctly assigning a time to a track in a high–pile-up environment is shown in Fig. 3.19(b) as a function of track p_{T}. The contribution of tracks with only one hit associated in the HGTD to the total efficiency is about 18%. The dependence of the efficiency with p_{T} is similar to that obtained with single-pion samples, reaching a plateau of 83%. The efficiency increase due to pileup hits being associated with the track is compensated by the additional association criteria. In total the efficiency is lower. Requiring a reconstructed time compatible with the expected time within two standard deviations reduces the efficiency by about 5%.

This study is a first step in the development of the association of the track information with timing information in the HGTD. The main inefficiency, of not finding a hit within the window, has been identified as interactions in the material in front of the HGTD. More sophisticated pattern recognition algorithms are being studied to improve this efficiency.

3.2.3 Determination of the time of the primary vertex

Once time measurements are associated to reconstructed tracks a time can be associated to the reconstructed vertices. For a precise knowledge of the time of the hard scatter vertex it is essential to remove tracks that are out of time with respect to the vertex.

First a study was performed only using the tracking information. For each reconstructed vertex, tracks with associated times differing by less than 50 ps are clustered together iteratively. If several clusters are reconstructed for a vertex, the one with the highest $\sum p_T^2$ is assigned to the vertex. Fig. 3.20 shows the reconstructed vertex time as function of the true vertex time. Spurious measurements off-diagonal are due to splitting and merging effects and may be improved by including the time measurements in the vertex finding and fitting procedure.

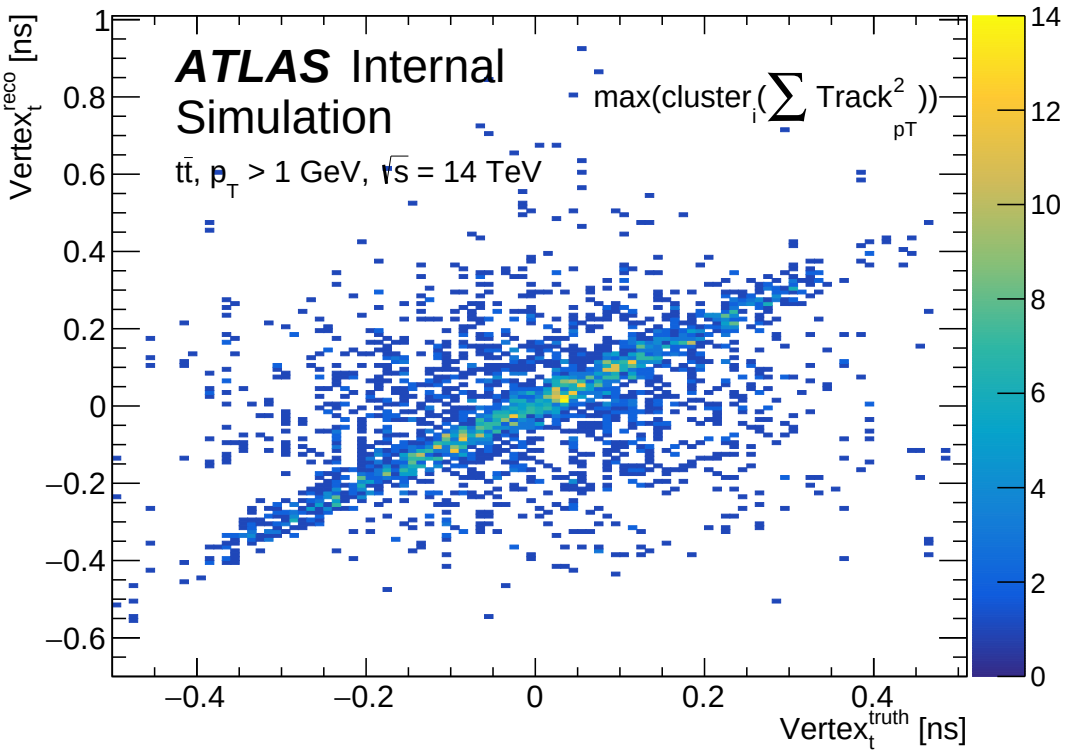


Figure 3.20: The reconstructed vertex time is shown as function of the true vertex time.

Next the time vertex efficiency was studied for tracks associated to a jet in order to determine the time and the resolution of the hard scatter vertex. Th restriction of the tracks to these tracks from all possible tracks improves the determination of the primary t_0. Studies have shown that the pile-up track contamination of hard scatter truth matched jets represent on average, depending on η, less than 30% of the tracks associated to a jet. The track-time clustering takes advantage of this, searching for the largest cluster in the set of times associated to the tracks in the jet in order to reject the pileup.

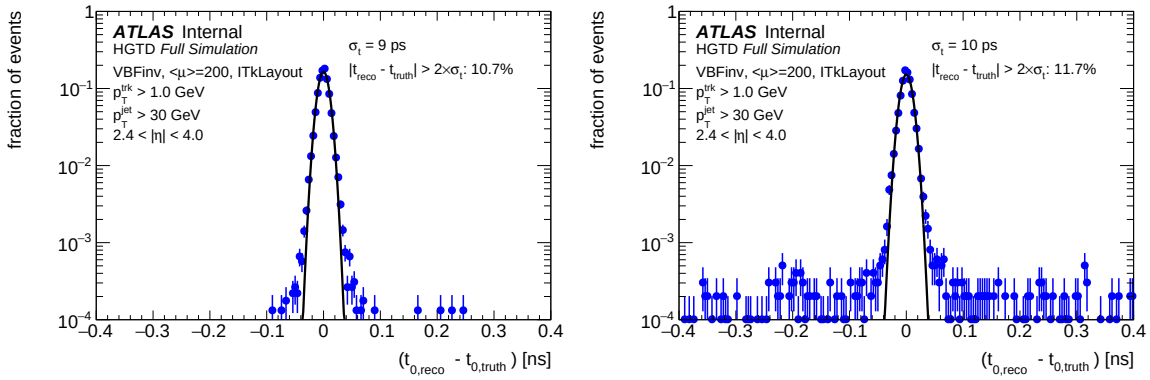
With this method, each associated track of the jet is assigned a time, unless no hits fulfilling the requirements are found. Further, the individual track-times are clustered iteratively, starting with those closest to each other in units of their resolution. The resolution used in the algorithm was determined from the Gaussian core of the track time resolution studies. These times are merged, and a resolution weighted time average calculated as well the resolution of this merged time.

$$t_{\text{merged}} = \frac{\sigma_2^2 t_1 + \sigma_1^2 t_2}{\sigma_1^2 + \sigma_2^2} \tag{3.3}$$

$$\sigma_{\text{merged}} = \frac{\sigma_1 \sigma_2}{\sqrt{\sigma_1^2 + \sigma_2^2}} \tag{3.4}$$

More time values are added to this average, repeating the update of time and resolution, until the next value is outside three times the uncertainty σ_{merged}.

To define a time for the jet, the biggest time cluster, i.e., the t_{merged} with the largest number of t_i, is chosen. The only additional condition on this cluster is to have at least two entries, i.e., at least two tracks. If there is a second cluster with the same number of entries, none is selected and the time for this jet is not defined.



(a) Leading Jet (b) Subleading Jet

Figure 3.21: The precision of the t_0 determination using the HGTD is shown.

To assign a time to the hard scatter event, t_0, usage of the time of the leading or sub-leading jet was studied on a sample of vector boson fusion Higgs production, followed by an invisible decay of the Higgs boson for $\langle\mu\rangle = 200$. In order to separate the performance of the algorithm from the effect of vertexing, the truth primary vertex is required to be within 0.5 mm of the reconstructed primary vertex, reducing the sample by 37%. The leading or subleading jet is required to fall into the acceptance of the HGTD which is the case in 59% of the remaining events. Applying the algorithm on the leading jet leads to an efficiency of the t_0 determination of 85%.

The resolution of the algorithm is shown in Fig. 3.21(a) for the leading jet. The Gaussian core of the distribution has a width of 9 ps. Only 11% of the reconstructed times are outside a 2σ window of the central value.

The subleading jet can be used for t_0 determination in case the leading one is in the central region. The algorithm applied to subleading jet leads to an efficiency of 60%, where 12% of

the reconstructed times are outside a 2σ window of the central value as shown in Fig. 3.21(b). The precision of the timing reconstruction of 11 ps is similar to the precision using the leading jet.

In total, excluding requirement to reconstruct the correct vertex in space within 0.5 mm, in 43% of the events a t_0 can be be reconstructed. Requiring this reconstructed time to be within 2σ of the expected vertex time reduces the number of 38%.

3.2.4 Suppression of pile-up jets

Pile-up is one of the most difficult challenges for object identification under HL-LHC conditions. Particles produced in pile-up interactions can contaminate the jets of interest coming from the hard-scatter vertex, thereby reducing the accuracy of the jet energy determination. Pile-up interactions can also produce additional jets which do not originate from the primary hard-scatter interaction. These pile-up jets can be produced as the result of a hard QCD process from a pile-up 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 pile-up jets are QCD jets.

Pile-up jets can reduce the precision of Standard Model measurements and the sensitivity to discover new physics. For example, additional jets can increase the amount 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 pile-up jets is essential to enhance the physics potential of the HL-LHC.

The key element to suppress pile-up 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.

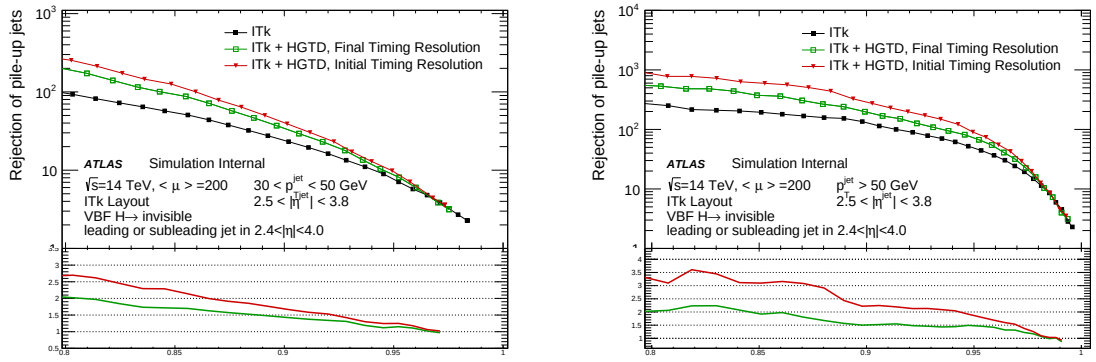
$$R_{p_{\mathrm{T}}} = \frac{\Sigma p_{\mathrm{T}}^{\mathrm{trk}}(\mathrm{PV}_0)}{p_{\mathrm{T}}^{\mathrm{jet}}}.$$

The tracks used to calculate $R_{p_{\mathrm{T}}}$ fulfil the quality requirements defined in Ref. [11] and are required to have $p_{\mathrm{T}} > 1\,\mathrm{GeV}$. The tracks used in the $R_{p_{\mathrm{T}}}$ calculation are required to satisfy Eq. (3.1).

Hard-scatter and pile-up 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 matching criteria are defined in Ref. [12]. 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 $p_{\mathrm{T}} > 10\,\mathrm{GeV}$. The pile-up jets must be at least $\Delta R > 0.6$ away from any truth jet with $p_{\mathrm{T}} > 4\,\mathrm{GeV}$. The

performance has been studied using a mixture of full reconstruction (for tracks and jets) and fast simulation (for the HGTD).

At moderate levels of pile-up, where track impact parameter measurements can be used to assign tracks to vertices with relatively little ambiguity, small values of R_{p_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 pile-up jets. However, at high pile-up conditions, and particularly in the forward region, the power of this discriminant is reduced. The effect can be mitigated by including timing information from the HGTD, removing tracks outside a $2\sigma_t$ window around the time of the hard-scatter vertex. The main impact of the HGTD in this study is to remove stochastic pileup jets.



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

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

Figure 3.22: Pile-up jet rejection as a function of hard-scatter jet efficiency in the $2.4 < |\eta| < 4.0$ region, for the ITk-only and combined ITk + HGTD scenarios with different time resolutions.

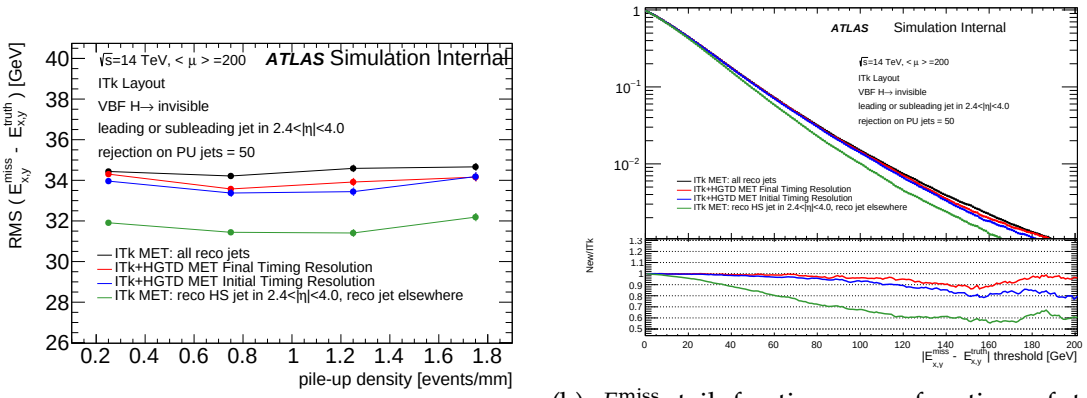
Samples of VBF Higgs production with invisible Higgs decays with $\langle\mu\rangle = 200$ were used in this study. The events were required to have the leading or subleading jet reconstructed in the HGTD acceptance. This ensures consistency with the algorithm to determine the time of the primary vertex.

Fig. 3.22 shows the rejection, i.e., the inverse of the mis-tag efficiency, of pile-up jets as a function of the efficiency for selecting hard-scatter jets using the R_{p_T} discriminant for jets with low and high p_T in dijet events with $\langle\mu\rangle = 200$ without and with the HGTD for the different timing resolution scenarios. A significant improvement in performance of a factor of 1.9 and 1.5 higher pile-up rejection for jets at an efficiency of 88% is achieved with the use of timing information in the *Initial* and *Final* timing scenarios. For a fixed pileup rejection of 50 the absolute improvement of the hard scatter efficiency is 3% for the *Initial* and 2% for the *Final* scenario. Both the time association efficiency and the t_0 determination precision were taken into account with smearing functions.

3.2.5 Missing transverse momentum

The missing transverse momentum, denoted $E_{\mathrm{T}}^{\mathrm{miss}}$, is computed as the negative vector momentum sum of high-p_{T} physics objects in the event, plus a soft-term component, from particles which do not constitute high-p_{T} objects. The soft term is calculated using tracks associated to the hard-scatter vertex [13].

In this study, $E_{\mathrm{T}}^{\mathrm{miss}}$ is reconstructed using selected muons, with $p_{\mathrm{T}} > 2.5\,\mathrm{GeV}$ or $p > 4\,\mathrm{GeV}$, and electrons with $p_{\mathrm{T}} > 10\,\mathrm{GeV}$. Tracks are required to satisfy the criteria described in Sec. 3.2.4. Jets with $p_{\mathrm{T}} > 20\,\mathrm{GeV}$ are selected with a pileup jet rejection of 50 using the $R_{p_{\mathrm{T}}}$ as discriminant discussed in the previous section. The association of tracks to the primary vertex, optimised for the jet selection, has been done according to Eq. (3.1). Selected tracks originating from the hard-scatter vertex, with $p_{\mathrm{T}}^{\mathrm{trk}} > 1\,\mathrm{GeV}$ and $|\eta^{\mathrm{trk}}| < 4$, not associated to muons, electrons and jets, are used to reconstruct the soft-term component of the $E_{\mathrm{T}}^{\mathrm{miss}}$. The HGTD timing information is used for the soft-term by requiring the track time to be compatible with the reconstructed vertex time within two standard deviations. The distance of the reconstructed vertex to the truth vertex is required to be less than 0.1 mm. 57% of the events remain after this requirement. Events where a jet overlaps with an electron are rejected.



(a) $E_{\mathrm{T}}^{\mathrm{miss}}$ resolution as function of pile-up density.

(b) $E_{\mathrm{T}}^{\mathrm{miss}}$ tail fraction as a function of tail threshold.

Figure 3.23: Three scenarios are shown for the $E_{\mathrm{T}}^{\mathrm{miss}}$: ITk scenario, ITk and HGTD as well as hard scatter truth jets in the HGTD acceptance. For 3.23(b) the ratio of the two curves is shown in the lower panel.

Fig. 3.23(a) shows the resolution of the x- and y-components of the $E_{\mathrm{T}}^{\mathrm{miss}}$ as a function of the pile-up density in the event for samples of VBF Higgs production with invisible Higgs decays for $\langle\mu\rangle = 200$. The results using the ITk alone are only improved slightly using the timing capabilities of the HGTD in spite of the gain shown in Sec. 3.2.4. While the jet performance gain is defined per object, $E_{\mathrm{T}}^{\mathrm{miss}}$ is an event level quantity. For a fixed pileup

rejection of 50, the hard scatter jet efficiency improves by about 3% for the timing scenario *Final*. In VBF Higgs production 57% of the events, normalized relative to the requirement on the maximal distance between the reconstructed and true vertex, have at least one hard scatter jet reconstructed in the HGTD acceptance. Therefore only 1 event out of ≈ 60 will have a jet configuration that is modified. Therefore the improvement on the jet performance is diluted in the $E_{\mathrm{T}}^{\mathrm{miss}}$ analysis.

In many analysis, especially new physics searches, it is crucial to minimize the $E_{\mathrm{T}}^{\mathrm{miss}}$ tails. In Fig. 3.23(b) the fraction of events above a $E_{\mathrm{T}}^{\mathrm{miss}} - E_{\mathrm{T}}^{\mathrm{miss,truth}}$ threshold is shown as a function of the $E_{\mathrm{T}}^{\mathrm{miss}} - E_{\mathrm{T}}^{\mathrm{miss,truth}}$ threshold for a fixed pile-up jet rejection of 50 in VBF Higgs production samples with $\langle\mu\rangle = 200$. A lower fraction of events passing high $E_{\mathrm{T}}^{\mathrm{miss}} - E_{\mathrm{T}}^{\mathrm{miss,truth}}$ thresholds corresponds to a better $E_{\mathrm{T}}^{\mathrm{miss}}$ reconstruction. The ideal scenario using the generated quantities instead of the reconstructed ones in the acceptance of the HGTD shows the maximal potential for improvement. When adding the HGTD, for a threshold of 160 GeV, the improvement is 10% for the timing scenario *Final* and 20% for *Initial*. The decrease of the improvement for high values of the threshold is a statistical fluctuation.

3.2.6 Tagging of heavy flavour jets

The efficient identification of b-jets and high rejection of light-quark jets is of central importance in the HL-LHC physics program. Tagging b-jets is particularly sensitive to pile-up-track contamination. This is due to the fact that b-tagging algorithms consider tracks with large impact parameters (in both the transverse and longitudinal directions) from the decay of displaced vertices. With a larger z_0 window, tracks from nearby pile-up interactions are more likely to be selected, leading to an increased rate of misidentified light-quark jets.

Using simulated $t\bar{t}$ events at $\langle\mu\rangle = 200$, the impact of the HGTD on the performance of b-tagging algorithms is studied for forward jets ($|\eta| > 2.4$). Fig. 3.24 shows the light-jet rejection versus b-tagging efficiency for the IP3D+SV1 b-tagging algorithm. The addition of the HGTD removes pile-up tracks from the track selection. As a result, the performance of the b-tagger is significantly improved. For a b-tagging efficiency of 70% and 85%, the corresponding light-jet rejection for MV1 is increased by factors between 1.3 and 1.4. These factors could be greater for processes where more b-jets are expected in the forward region. The performance is shown for the ITk-only scenario as well as three scenarios with HGTD timing performance representing different stages of the HL-LHC program. It can be seen that all timing scenarios yield significant improvements in the performance, even in the *Final* scenario. Importantly, significant improvements are observed also after the full radiation damage expected during HL-LHC operation.

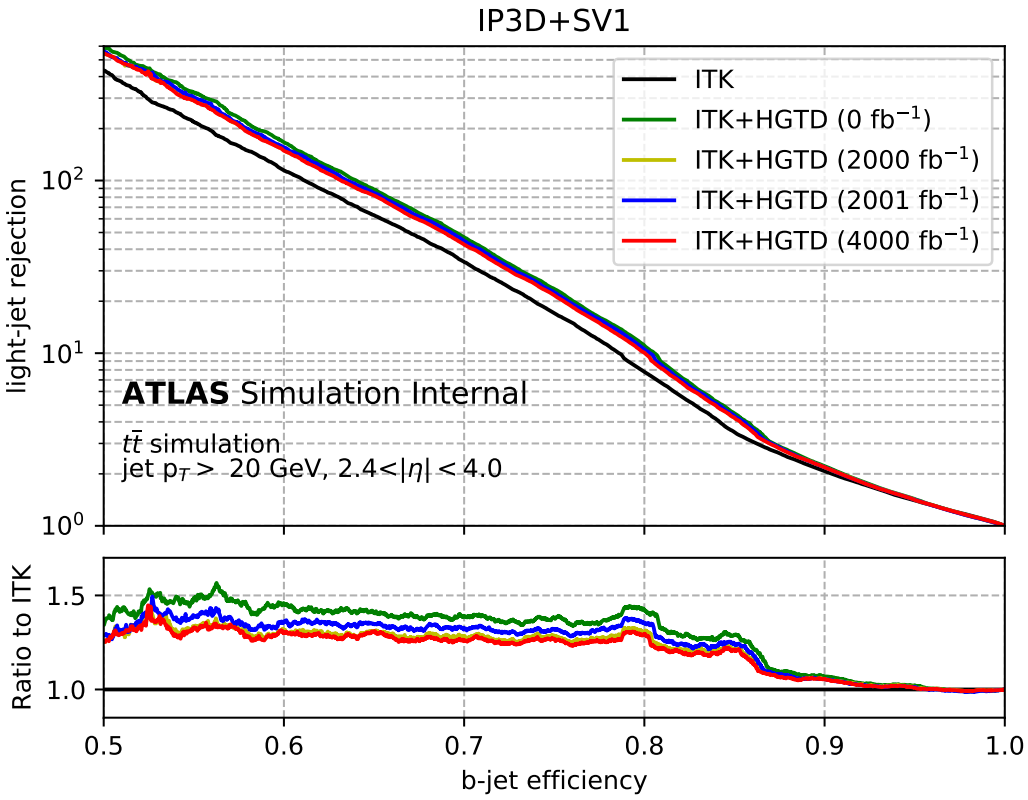


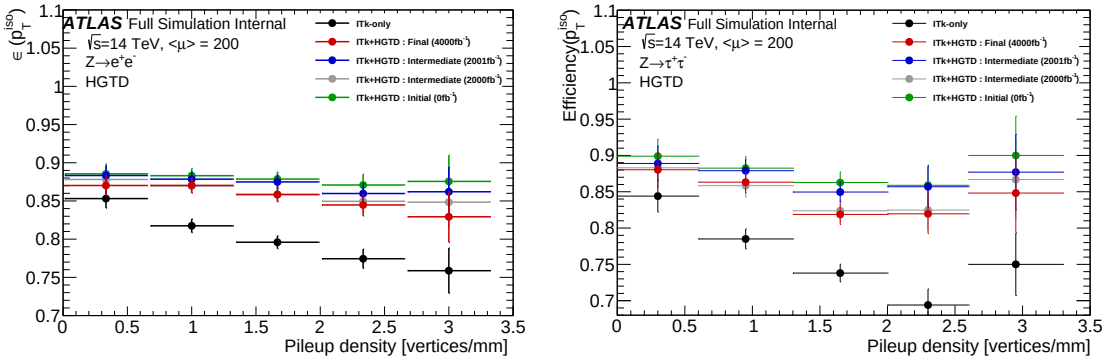
Figure 3.24: Light-jet rejection versus b-tagging efficiency for the IP3D+SV1 tagger and The study uses $t\bar{t}$ events at $\langle\mu\rangle = 200$ and shows the achieved performance for different time resolution scenarios. The ratio plots at the bottom show the relative performance achieved with the HGTD with respect to the ITk-only scenario.

3.2.7 Lepton isolation

In this section, studies of how the electron isolation efficiency is improved with HGTD are presented based on the full simulation of the HGTD. The HGTD can be used to assign a time to leptons in the forward region. This information can be exploited to reject tracks which come from other interactions but are spatially close to the energy deposits in the calorimeter and/or the track associated to the lepton. The timing information can reject additional tracks from interactions close in z, according to Eq. (3.1) but separated in time from the hard-scatter vertex. The isolation efficiency is defined as the probability that no track with $p_{\mathrm{T}} > 1\,\mathrm{GeV}$ is reconstructed within $\Delta R < 0.2$ of the electron track.

The physics process used in this study is Z boson production followed by a decay either to electrons or tau leptons. Only electronic decays of the tau lepton were used. Forward

electrons with a $p_{\mathrm{T}} > 20\,\mathrm{GeV}$ passing the standard ATLAS medium cuts where selected, keeping only those matched with a truth electron with a $p_{\mathrm{T}} > 20\,\mathrm{GeV}$ in a cone of $\Delta R < 0.2$. In order to study a reasonable p_{T} range the cut was reduced to 10 GeV for the tau sample. The electron track is defined as the track closest to the electron cluster having a ratio of transverse track momentum to transverse cluster energy greater than 0.1. The correct electron track is selected in 95% of the cases. The tracks are extrapolated to the HGTD surfaces using the last measured point in the ITk. The closest HGTD hit in a window of 3.5 mm between the extrapolated position and the hit position is associated to the track. The window is larger for electrons than for pions and muons to account for Bremsstrahlung in the material between the last ITk measurement and the first sensitive layer of the HGTD. A time is then reconstructed in the HGTD for the electron track as well as for all other tracks with $p_{\mathrm{T}} > 1\,\mathrm{GeV}$ which are within $\Delta R < 0.2$ of the electron track. These time 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.



(a) Local Vertex Density (b) p_{T}

Figure 3.25: The selection efficiency is shown for the electron isolation criteria using the ITk and ITk + HGTD for different timing resolutions.

The isolation efficiency is shown in Fig. 3.25 for the ITk-only scenario and four HGTD timing resolution scenarios. For Fig. 3.25(a) the samples with the electronic decay of the Z boson was used. While the efficiency drops strongly with the increase of the pile-up density when using only the ITk, the addition of the HGTD timing information reduces this drop, keeping an efficiency above 85% even at high pile-up density, i.e. with up to three additional vertices around the hard-scatter vertex. For a local pile-up density of the order of 1.6 vertices/mm the electron isolation efficiency is improved by about 9% for the *I*nitial and by about 6% for the *Final* scenario. Even in the *Final* timing scenario, the resolution is sufficient to achieve an isolation efficiency essentially independent of the pile-up density at the end of the HL-LHC.

As an example for a local pileup density of 1.6 vertices/mm the inefficiency of the lepton isolation is due to either tracks of hadrons or tracks originating from electromagnetic interac-

tions of the electron (shower). In the ITk only scenario 15% of the electrons are classified as non-isolated due to tracks truth matched to a hadron and 6% due to tracks truth matched to a lepton. The HGTD timing information reduces the hadronic inefficiency by 53% (relative) by removing the pileup track. For the leptonic inefficiency the relative reduction is only 9% since the extra tracks originate from interactions of the primary electron. They are therefore in time with the signal electron.

In Fig. 3.25(b) the p_{T} dependence of the isolation efficiency is shown. As the direct decay of the Z boson to electrons leads to a peak at 45 GeV, the use of the decay through a tau lepton allows to probe more easily a broader range of p_{T}. The improvement is about 10-15% integrated over the local vertex density, essentially independent of η. The isolation efficiency as function of the local vertex density for these events is about the same as the direct decay isolation efficiency.

The timing resolution of the electron is 16 ps for the *Initial* timing scenario with only 9% of the electron candidates outside a 2σ window centered on the true time. For the timing scenarion *Final* the performance is 42 ps due to the effect of irradiation on the timing resolution on the sensor and electronics also with 8% outside the 2σ window. Less than 1% of the electron tracks in the 2 sigma window have only pileup hits associated to them. The performance in the forward region reaches a level similar to that in the central region. The improvement of the performance depends only modestly on the timing scenario.

3.2.8 Non-collision background

The ATLAS detector signals from the products of proton-proton collisions, as well as from non-collision backgrounds (NCB). These consist of beam halo and beam-gas backgrounds, cosmic rays and detector noise. Beam halo, elastic and inelastic beam-gas are the main precesses creating non-collision backgrounds reaching the detector. Beam halo is the small fraction of particles surrounding the dense beam core. The collimator system of the LHC is designed to stop off-momentum and off-position particles. However leakage from the collimation insertions and from the tertiary collimators near ATLAS allow a fraction of beam halo background to enter the detector. Beam-gas background is related to the pressure in the beam pipe; protons colliding with the residual gas molecules in the vacuum create background particles. NCB events can be created in the vicinity of ATLAS or travel in the beam pipe or parallel to the beam line over large distances.

As a result of the unavoidable nature of these backgrounds, their implications on physics studies and detector occupancy of the non-collision background must be understood. This study investigates the beam halo background in the HGTD. Events are selected from the FLUKA beam halo simulation [14] (using HL-LHC optics version 1.0) after interacting on the TCTs located approximately 150 m upstream of ATLAS [15]. The timing analysis of the HGTD hits provides information on NCB events as well as nominal collision events.

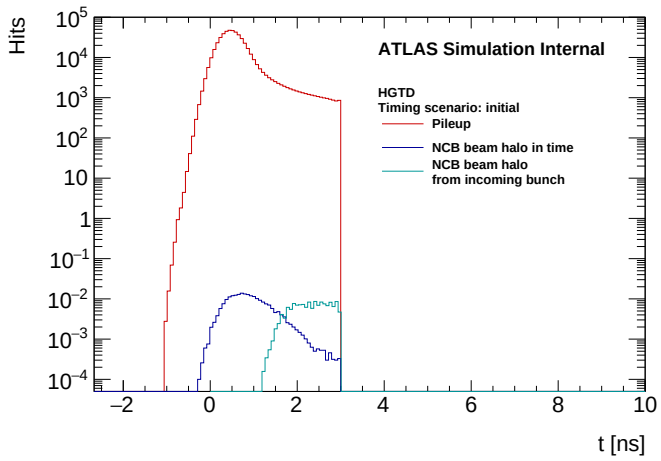


Figure 3.26: The time distribution, normalized per bunch, is shown for minimum bias events with $\langle\mu\rangle = 200$ (red), the background halo samples of the colliding bunch (blue) and the incoming next colliding bunch (cyan). The distribution is cut at 3 ns to mimic the timing window of the TDC in the electronics.

The time distributions for different sources of detector hits are shown in Fig. 3.26. The time t is offset by 11.6 ns with respect to the nominal collision time as explained in Sec. 3.1.2. As a way to mimic the timing window of the readout electronics TDC, a cut on events with $t > 3$ ns is applied. The arrival time of the signal correlated with the nominal collision in the HGTD is at approximately 0.405 ns. That signal is coincident with the in-time non-collision background of the HGTD. The NCB events associated with the next bunch coming from the direction opposite to the interaction point reach the HGTD, on average, 1.6 ns after the time of arrival of nominal collision hits.

Only signals in the HGTD for negative z are shown. As the NCB is symmetrical, the timing structure will be similar for positive z. Non-collision background and especially beam halo events are expected to increase during the HL-LHC operation compared to Run-2 [16], nevertheless the NCB events in this study are normalised according to Run-2 BCM at 1 Hz, 10^{11} p. The NCB events are different from pile-up not only by time of arrival but other parameters as well, for example the NCB hits increase with radius, from 0 to 320 but outofdate where the overlap of the sensors changes and the hit count drops and then increases again up to the maximum radius of the detector, while pile-up decreases as function of the radius.

The timing structure of the NCB is not identical to the pile-up time distribution. The impact of the HGTD on the study of NCB will depend on the magnitude of its expected increase. The HGTD will most likely only be able to contribute to the study of NCB in the startup of the HL-LHC before its nominal luminosity is reached.

3.3 Physics

Results on VBF and tH production are in preparation and are foreseen to be part of the second ATLAS circulation.

3.3.1 Strategy for the Application of Object Performance to Physics Analyses

The improvements in the performance of forward leptons, b-tagging, jets, $E_{\mathrm{T}}^{\mathrm{miss}}$, as well as the determination of the vertex time translate into improvements in the physics potential of ATLAS.

The uncertainty on the integrated luminosity will be one of the largest, and in many cases dominant, uncertainties in Higgs physics and many other precision measurements at the HL-LHC program. One of the most direct ways in which HGTD can significantly enhance the ATLAS physics program at the LHC is by providing an additional measurement of the luminosity to reduce the uncertainty on the integrated luminosity. Sec. 3.3.2 discusses some concrete examples in the context of Higgs cross section measurements.

HGTD can also be utilized to trigger on highly ionizing particles in the forward region, enhancing the discovery reach to new particles such as magnetic monopoles. This is the subject of Sec. 3.3.4.

A key application of improved forward lepton isolation is the measurement of the weak mixing angle, discussed in Sec. 3.3.3. In the case of lepton isolation, the lepton track provides the reference t_0 of the event, so there is no need for a global t_0 reconstruction to take advantage of the track time measurements to suppress pile-up tracks in the isolation cone as described in Sec. 3.2.

Improved b-tagging in the forward region can be exploited in physics analysis with forward b-quarks. b-tagging does not have a large reliance on the t_0 determination with the exception of the case of single-track jets. In this case, a single time measurement does not provide a useful handle as the the jet vertex time algorithm requires at least two tracks to be clustered in time. Relaxing the requirement to one track would not improve the situation as the t_0 would be identical to the track time, so the track will pass the cuts. To account for this case, we conservatively assume that HGTD cannot be used for single track jets.

In tH production the final state will consist of mostly one single forward b-jet. If additionally this jet has only one track associated to it, a global event t_0 is needed. A global t_0 reconstruction algorithm is difficult because of the lack of an additional object for the t_0 determination. Ideally, the hard-scatter vertex time t_0 would be determined based on tracks not used in the target jet. Such algorithm has not yet been developed, and will be investigated in the future.

Additionally, it is possible to consider a different approach for the use of timing information in the context of b-tagging. The idea is similar in concept to the case of lepton isolation. In the *s*elf-tagging method, tracks within the b-jet are first split into sub-jets according to their times. For example, a candidate b-jet containing two tracks with times t_1 and t_2 (assuming both tracks were assigned the correct time information) may be considered as a single 2-track jet if both times are compatible with each other, or two one-track subjets otherwise. Note that the splitting of the jet into subjets is done without any knowledge about the global vertex t_0. It is only based on local (relative) time information. In a second step, the b-tagging algorithm is applied to each sub-jet separately and the new b-tag weight will be the largest weight from each of the subjets. For the simplest example of a mistag jet consisting of a one hard-scatter and one pile-up track (associated to the same spatial vertex) in which the two tracks form a fake secondary vertex, the self-tagging approach would result into two one-track jets with no secondary vertex information. The development and study of the self-tagging approach for b-tagging will be developed as a future next step. Such an approach could yield similar results as of having the knowledge of the hard-scatter t_0. The b-tagging performance obtained with truth-based t_0 determination will be used to estimate the gains in physics sensitivity.

Improvements in the suppression of pile-up jets and $E_{\mathrm{T}}^{\mathrm{miss}}$ can have a direct application in VBF/VBS physics analyses. However, this is a case where the knowledge of the global vertex t_0 is more important than for all previous objects considered and more care is needed to be able to translate these performance improvements into VBF/VBS physics sensitivity gains with HGTD. In a typical VBF/VBS event selection, two jets above 30 GeV and large invariant mass m_{jj} are required. The large invariant mass requirement results in at least one forward jet almost always present in the final state. This leads to final states that can contain one or two forward ($|\eta| > 2.5$) $t$ag jets. For Standard Model VBS processes after typical m_{jj} and $\Delta\eta(j,j)$ selections, about 80% of the time there is one central and one forward jet. For heavy mass objects, such as searches for heavy Higgs bosons, the proportion of forward-forward jets increases. Backgrounds to VBS/VBF topologies can arise from single or di-boson processes $V(V)$+1 jet ($V = W/Z$) with an extra jet from a pile-up (merged) vertex. The two main background topologies to VBF/VBS analyses are hence central-forward (case 1) and forward-forward (case 2), where one of the two jets is from a pile-up interactions.

In case 1 there is only one jet within the HGTD acceptance. This is an example where t_0 is required. Given a measurement of the vertex t_0 of the jet, it is possible, to improve the R_{p_T} calculation of the forward jet, or to check the Δt between the jet time and t_0 if an independent determination of t_0 is available. Using the t_0 algorithm described in Sec. 3.2.3, if the pile-up forward jet is *s*tochastic, i.e. made of tracks from various different interactions, the t_0 algorithm will remove pileup tracks from hard scatter jets thus improving R_{p_T}. If it is a QCD pile-up jet, then all its tracks will have the same t_0 time, so that jet is likely to pass. The relative fraction of QCD vs. stochastic pile-up jets is very final state and topology dependent so it is not simple to estimate the gain in pile-up jet suppression when there is one forward and one central jet in the final state. In order to be able to suppress QCD pile-up

jets in central-forward VBF/VBS topologies, a new t_0 reconstruction algorithm based on tracks *o*utside jets will be required. This will be investigated in the future.

Case 2 consists of one hard-scatter jet and another pile-up jet. Since in this case both jets are within HGTD acceptance, HGTD can be used to assign a time to each jet and check for consistency. Similar to the discussion about *b*-tagging and lepton isolation, the knowledge of the global t_0 of the event is not required to identify this topology. Only relative time information is enough to determine if the event is signal and should pass (both jets belong to the same interaction) or if the event consists of two jets from two different interactions that have similar z vertex position but different time and should be rejected. Cases where a third forward jet is present are more complex and would require special consideration. However, this case is expected to be less important.

Therefore the impact of HGTD to the suppression of pile-up jets depends on the event topology and selection cuts, particularly the m_{jj} cut that determines the fraction of forward-forward jet events, and the specifics of the t_0 reconstruction algorithm. In the most conservative case, using the existing t_0 algorithm, HGTD can improve the suppression of background events with a QCD or stochastic pile-up jet in the forward-forward topology and with a stochastic pile-up jet in the central-forward topology. More advanced t_0 methods, yet to be developed, may allow to resolve the central-forward QCD pile-up case.

There is another potential HGTD improvement in VBF/VBS analyses related to jets that does not rely on the knowledge of t_0. This is the case of improvements in the jet energy resolution through improved particle-flow. Tracks within the jet that are not consistent in time with the rest of the jet can be rejected as pile-up tracks. The contamination of pile-up tracks in the forward region might be one of the main limitations of particle flow algorithms in the forward region since it will not be possible to correctly tag and remove pile-up energy contributions to the jet only on the basis of the longitudinal impact parameter, without the additional time information. The study of particle flow algorithms using timing information is outside the scope of this TDR due its complexity, but a clear area of potential improvement for HGTD.

Improvement in the reduction of $E_{\mathrm{T}}^{\mathrm{miss}}$ tails can impact searches for new physics and precision measurements with neutrinos in the final state. Signatures with large $E_{\mathrm{T}}^{\mathrm{miss}}$ comprise a very broad class of events at the LHC. The impact of reduced transverse missing energy tails on specific physics analyses has not yet been studied but will be pursued as a next step.

The next sections provide detailed studies showing the expected impact of HGTD in some selected physics analyses from improved luminosity, lepton isolation, and triggering on highly ionizing particles. Many other physics analysis are being pursued.

3.3.2 Impact of the luminosity uncertainty

The uncertainty on the measurement of the integrated luminosity affects the majority of physics analyses at the LHC. It is especially relevant for precision measurements, for which the total uncertainty is dominated by systematic effects. An example of such an effect, that is the dominant source of uncertainty for some measurements, is the knowledge of the integrated luminosity.

The luminosity has been measured in Run 1 and Run 2 of the LHC using several detectors. Uncertainties for the nominal proton-proton collision data are currently 2.1%, 2.2%, 2.4%, and 2.0% for data from 2015, 2016, 2017 and 2018, respectively. For the combined 2015-2018 dataset, taking into account correlated and uncorrelated effects between the different years, an uncertainty of 1.7% is obtained. These uncertainties are derived following a methodology similar to that detailed in Ref. [17], and using the LUCID-2 detector for the baseline luminosity measurements [18], from calibration of the luminosity scale using x-y beam-separation scans (van der Meer scans).

In Ref. [19], the Higgs boson analyses performed during Run-2 have been extrapolated to the HL-LHC dataset. The performance of the analyses have been updated, including the expected changes to the uncertainties affecting the analyses, the increase of the collision energy and the increase in the integrated luminosity. An ambitious uncertainty of 1% is assumed for the integrated luminosity, as to not dominate all other sources of uncertainty, instead of a more realistic 2% uncertainty. 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.

Tab. 3.3 lists the largest sources of uncertainty 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**	$\Delta\sigma/\sigma_{\text{SM}}$
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 $ggH + \text{VBF}, H \rightarrow \tau\tau$	QCD scale ggH, $p_{\text{T}}^{H} \geq 120\,\text{GeV}$	1.7%

Table 3.3: List of dominant uncertainties (excluding the uncertainty on the integrated luminosity) affecting various expected Higgs boson cross section results at the HL-LHC. 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 program at the HL-LHC. The same concerns apply to any measurement of processes with similar, or larger, cross sections compared to the Higgs boson.

These include important processes such as W and Z boson production, and measurements of single and pair production of top quarks. As will be described in Chap. 6 and Chap. 10, the HGTD will have the capability to measure the luminosity at the HL-LHC. Chap. 10 also discuss the main sources of uncertainty affecting the luminosity determination, many which will take operational experience with the HGTD to provide numerical estimates for.

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

In the Standard Model (SM), 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, $\sin^2\theta_{\text{eff}}$ [20].

Experimentally, this asymmetry can be expressed as simply as

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

where θ^* is the angle between the negative lepton and the quark in the Collins-Soper frame [21] 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 is dominated by the combination of measurements at LEP and at SLD, which gives $\sin^2\theta_{\text{eff}} = 0.231530 \pm 16 \times 10^{-5}$. However, the two most precise measurements differ by over 3σ [20].

At HL-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 [22]. Only Z bosons decaying to electrons are considered in this analysis since this final state provides the best experimental precision within the largest acceptance.

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 $p_{\text{T}} > 25\,\text{GeV}$. The invariant mass of the electron pair is required to be loosely consistent with the Z boson mass, $60 < m_{\ell\ell} < 200\,\text{GeV}$, 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. 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. The signal significance with HGTD is up to 20% higher with respect to the case of ITk only in the CF channel.

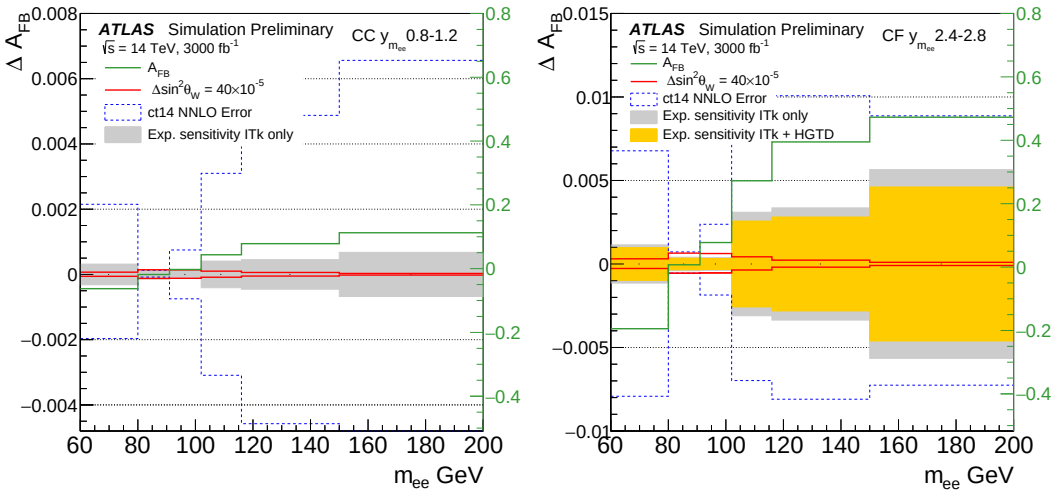
A_{FB} is calculated from the selected electron pairs, and unfolded to correct for detector effects and migrations in $m_{\ell\ell}$ and $|y_{ee}|$ bins. In the CF and FF channels migrations in the $m_{\ell\ell}$ are up to 50 and 60% respectively. Various sources of uncertainty are considered. Those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

Significant uncertainties arise from knowledge of the momentum scale and resolution for the electrons. Following Reference [23] 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} is shown in green in Fig. 3.27 for each channel for chosen rapidity bin. As expected the larger 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 NNLO CT14 parton distribution function (PDF). As shown in Fig. 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 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 $\sin^2\theta_{\mathrm{eff}}$ is maximal, while being significantly smaller at high and low masses. Thus, in this projection a global fit is performed where $\sin^2\theta_{\mathrm{eff}}$ is extracted while constraining at the same time the PDF uncertainties [22]. With this analysis, the expected sensitivity of the extraction of $\sin^2\theta_{\mathrm{eff}}$ 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 $\sin^2\theta_{\mathrm{eff}}$ sensitivity in the CF channel. Combining the three channels together the expected sensitivity reaches a precision of $\Delta\sin^2\theta_{\mathrm{eff}} = 18\times10^{-5}\pm16\times10^{-5}$ (PDF) $\pm9\times10^{-5}$ (exp.) which exceeds the precision achieved in all previous single-experiment results so far.

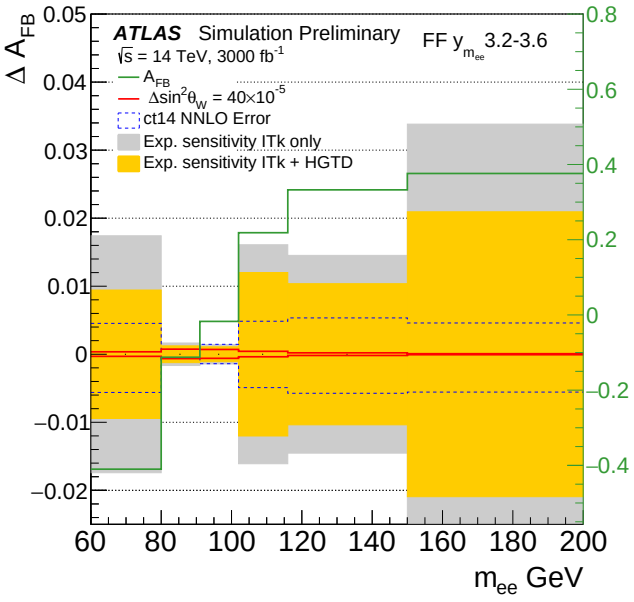
3.3.4 Monopole searches

Magnetic monopoles are elementary particles with a single pole magnetic charge, unlike dipoles that have both a north and south magnetic pole. These particles were first theorized



(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.

by physicist Paul Dirac in 1931 [24]. Until today no direct evidence of the existence of magnetic monopoles was found. A consequence of the existence of monopoles is the perfect symmetrization of Maxwell equations. With the existance of magnetic charge the divergence term of the magnetic field would no longer be 0 and the flux of magnetic particles would give an additional term to the rotor of the electric field. Monopoles can be either scalar bosons (Spin-0) or fermions (spin-1/2). The Bethe-Bloch formula for magnetic particles is different from the one for electrically charged particle. Electrons and protons have high ionization power at low energy and then stabilize to MIPs at high energy due to the $1/\beta^2$ term. For magnetic monopoles the ionization at high energy grows as a function of β^2, deviation significantly from the MIP signature.

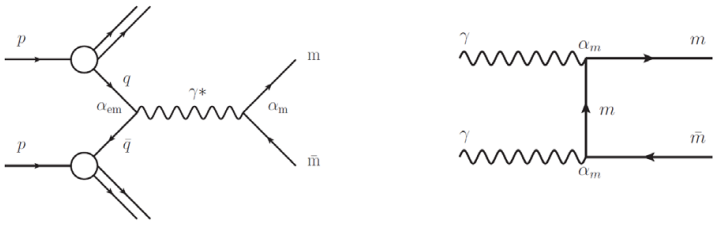


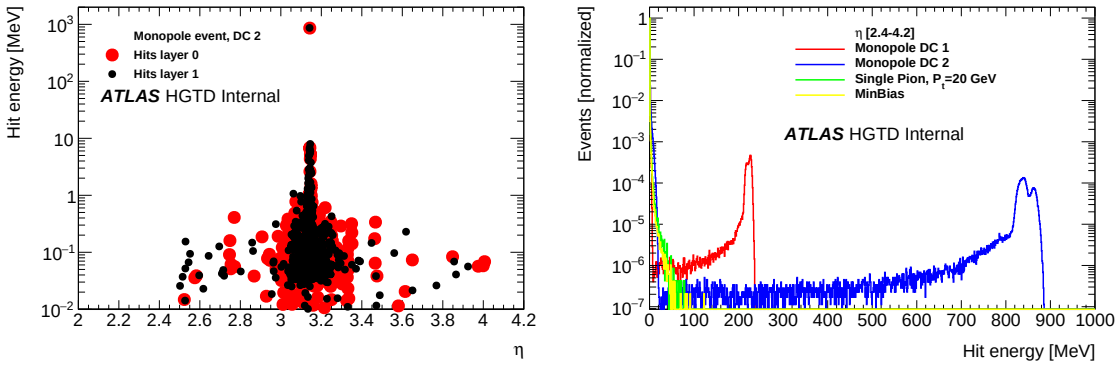
Figure 3.28: Production process of Monopole in LHC for Drell-Yan (left) and photon fusion (right)

Magnetic monopoles were also predicted as part of grand unified theories as described in [25]. Several new electroweak models predicted monopoles with a mass that could be accessible at LHC [26, 27] thus allowing the detection in LHC experiments. Such monopoles would be generated at LHC with Drell-Yan or photon fusion processes as shown in the Feynman diagrams in Fig. 3.28. However due to the large coupling of monopoles with photons the production cross section cannot be reliably calculated.

According to these models magnetic monopoles act as long lived and highly ionizing particles. A single Dirac (magnetic) charge corresponds to around 60 electric elementary charges. The increased ionizing power results in an energy deposit 100 times the energy deposit of a MIP in a silicon detector such as the HGTD. Past searches for monopoles in the ATLAS detector have not turned up evidence for monopoles [28–30]. These searches however only cover the barrel region of the detector and rely on a low-threshold EM trigger that will likely be scaled up in HL-LHC.

Monopoles were simulated as single particles with $\langle\mu\rangle = 0$. The samples were simulated with a special setting, the 4DL package [4DL] in full simulation. Samples with Monopoles of Dirac charge 1 and 2 were produced. Higher values of the Dirac charge leads to a high level of interaction in the ITk that the particles do not reach the HGTD.

The energy distribution of the hits in HGTD for a monopole event is shown in Fig. 3.29(a). Several low-energy hits and one or two hits with a very high-energy deposit are recorded. The distribution of the hit energy for monopoles with Dirac charge 1 and 2, of minimum bias samples and pions is shown in Fig. 3.29(b). The high-energy deposit is clearly separated



(a) Monopole (b) All

Figure 3.29: The energy deposit vs η) for an event of single Monopole with Dirac charge 2 as well as the energy distribution for minimum bias, pions and monopoles with Dirac Charge 1,2 are shown.

from the other types of particles. Thus a high-energy single hit would give a clear and unique signature for an interacting magnetic monopole.

This distinctive signature can also be exploited by HGTD electronics at trigger level. A special bit can be added in the luminosity processing electronics to flag high-energy deposits in a single pad. This way is possible to recognize candidate Monopole events at trigger level in the end-cap region of the detector and make them available for offline analysis.

4 Technical Overview

4.1 Introduction

This chapter summarizes the most important aspects of the design of the HGTD. The main requirements that drives the design and the proposed technical solutions are discussed. Measurements from the on-going R&D program are presented, especially on sensors and electronics, that demonstrate the achieved performance. More detailed descriptions of all these items will be presented in subsequent chapters, including the next steps towards the construction.

4.2 Detector overview and key requirements

The detector has been designed for an 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 barrel and the end-cap calorimeters, at a distance of approximately ±3.5 m from the interaction point. Fig. 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 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 by 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 total detector weight per end-cap is approximately 350 kg, of which 150 kg comes from the moderator, weight equally shared between the moderator located inside and outside the vessel.

The front vessel cover and each cooling/support disk are physically separated in two half circular disks to allow the opening of the detector in 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 in 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 raw direction, in order to provide on average 2.7 hits per track in the most irradiated and higher 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 covers 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.

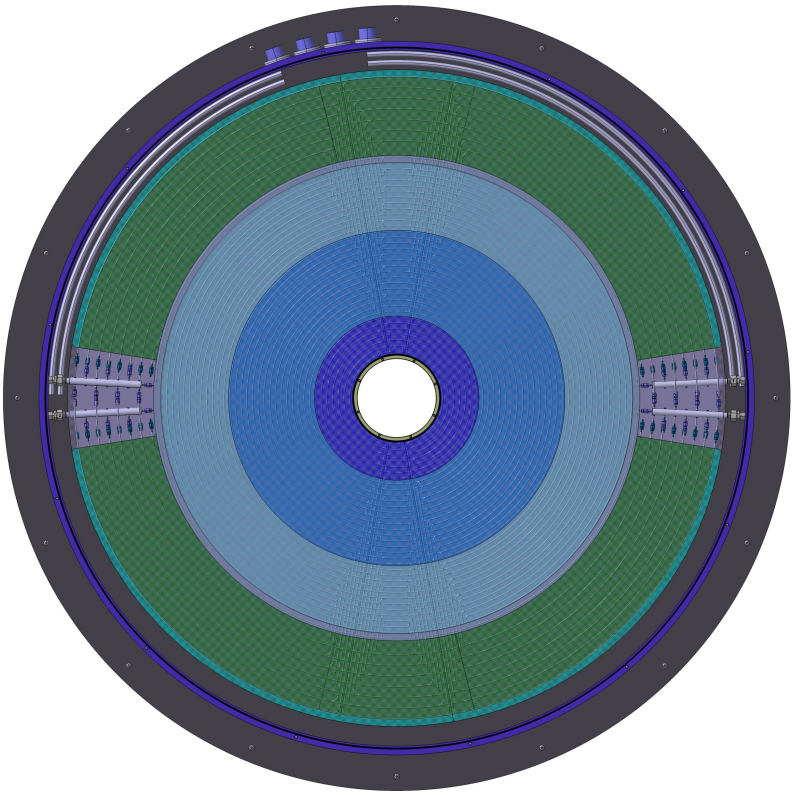
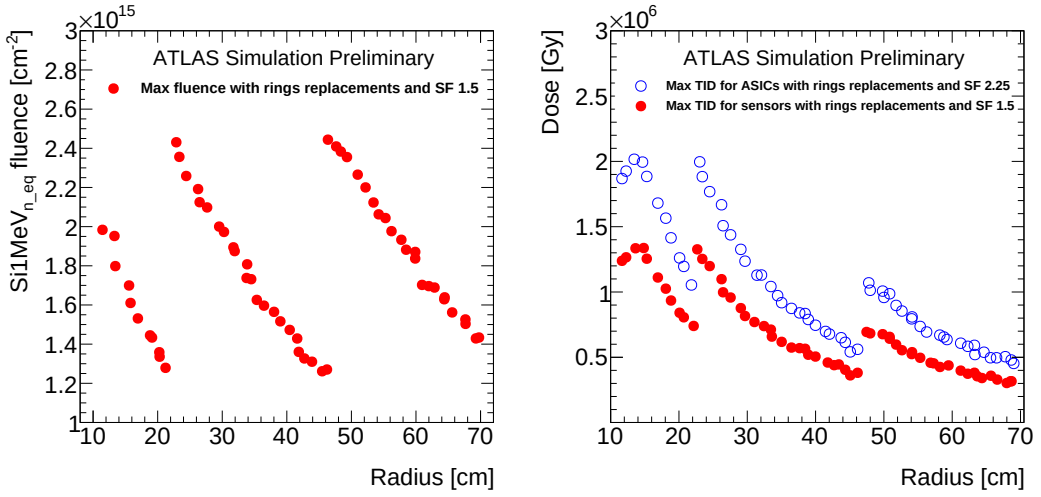


Figure 4.1: Transverse view of the detector. The blue active region is shown in 3 blue tons indicating the 3 rings with different sensors density. The green areas indicate the region of the peripheral electronics boards with some open space for the CO2 cooling manifold in gray. TODO: REPLACE FIGURE BY ONE WITH THE READOUT ROW LINES, PROVIDED BY ABOUD.

4.2.1 Expected Radiation levels

As discussed in Chap. 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 mitigate the radiation levels and fulfil the detector requirements during the full life time of the HL-LHC the plan is to replace the most inner ring after each 1000 fb^{-1} (3 times in total) and the middle ring at 2000 fb^{-1} (once), during long shutdowns. Fig. 4.2(a) and Fig. 4.2(b) show respectively the maximum expected neutron-equivalent fluence and TID as a function of the detector radius. A factor of 1.5 was included to account for uncertainties in the simulation. An additional factor of 1.5 was applied to the total ionising dose (TID) to account for low dose rate effects on the ASICs. In the proposed 3 rings layout the maximal TID and fluence, using the Fluka estimations of September 2019, does not exceed 2 MGy and 2.5x10^{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, as seen in Fig. 2.13. 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.



(a) (b)

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 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.

4.2.2 Key requirements

A high intrinsic single hit efficiency is essential for through the lifetime of the HGTD. This puts stringent constraints on the smallest charge delivered by the sensor and the lowest

achievable threshold of the electronics discriminator. Charge as low as 4 fC should be detected with a signal over noise (S/N) larger than 7, while keeping a low rate of fake hits induced by the electronics noise (< 0.1 %). 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 30 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 65 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 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 MIP with an LGAD gain of 20 (corresponding to a charge of 10 fC) at the start of the HL-LHC, and at most 60 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, to be smaller than 15 ps after calibration.

In addition, the detector should be sensitive to 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 and trigger

With the baseline ATLAS architecture, the ATLAS detector is read-out with a single Level 0 (L0) trigger running at an maximum rate of 1 MHz, with a maximum latency of 10 µs [31]. 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 line to the lpGBT. Therefore the maximal bandwidth is limited to 1.28 $\text{Gbit}\,\text{s}^{-1}$. The luminosity data is transmitted at each bunch crossing to dedicated lpGBTs, requiring a 640 $\text{Mbit}\,\text{s}^{-1}$ e-link bandwidth.

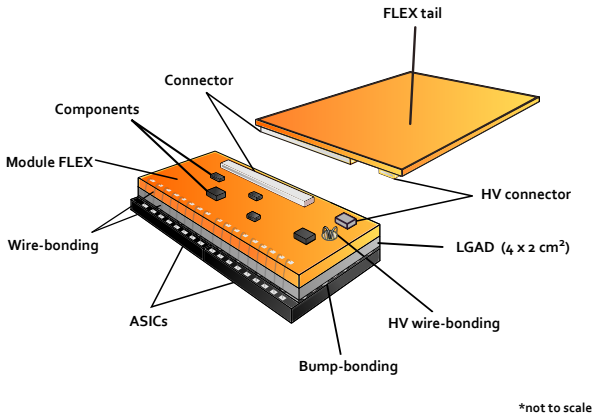


Figure 4.3: View of an HGTD hybrid module equipped with its read-out flex cable tail. The bare module, glued on the flex cable, is made of a $4 \times 2\,\mathrm{cm}^2$ sensor with two bump bonded ASICs. The signal lines of the ASIC are wire bonded on one side of the cable, while the bias voltage of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

4.3 Hybrid HGTD module

Fig. 4.3 shows a view of an hybrid module made of two parts: an LGAD sensor and two ASICs, called a bare module, and a flexible printed circuit board (flex cable) made of two pieces, 1 small flex board permanently glued to the bare module and a long flex tail whose length, not exceeding 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 in 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 and the flex.

The characteristics of the bare modules are:

- The size of the bare modules, all equal, is approximately $2 \times 4\,\mathrm{cm}^2$ and each bare module contains 450 pads (15x30). Its size has been defined to optimize the coverage at the inner radius a and to provide a good yield for the hybridization process. The nominal total amount of bare modules is 8032.

- The size of the pad, $1.3 \times 1.3\,\mathrm{mm}^2$, results from 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 5 years ago by the Centro Nacional de Microelectrónica (CNM) Barcelona 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 Fig. 4.4(a). When a charged particle crosses the detector, an initial current is created from 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 resulting 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 MIP particles. The expected current for different irradiation levels (therefore different gains) are presented in Fig. 4.4(b). For large gain, the rise time is about 500 ps and the signal duration is approximately 1 ns. For large fluences, the charge is smaller and the rise time and the signal duration are shorter.

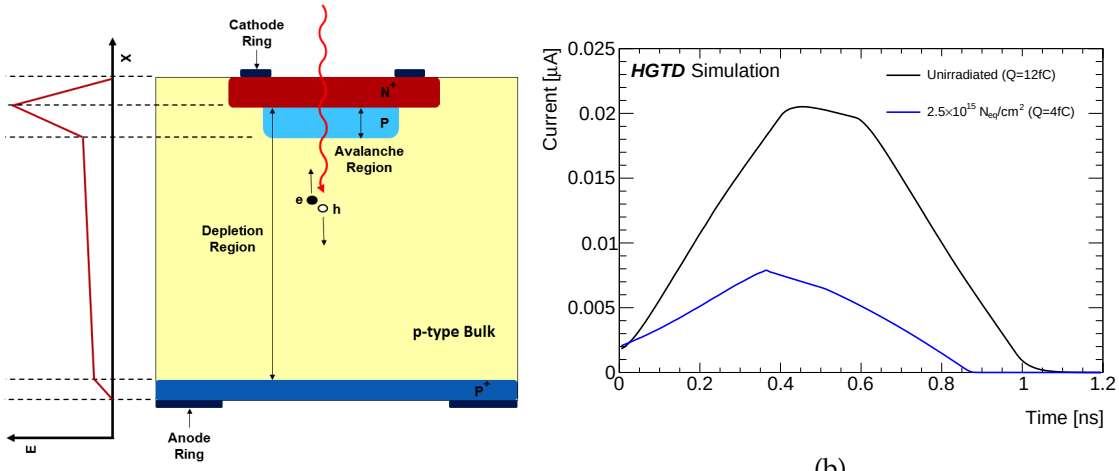


Figure 4.4: Cross section of an LGAD ((a)) and simulated signal current in LGADs at start and after full integrated neutron fluence ((b)).

After amplification in the gain layer, the height of the LGAD signal is proportional to the gain, M, but is independent of the detector thickness. 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 depends also linearly on the detector pad capacitance, therefore limiting the potential use of very thin sensors. Consequently, the optimal thinness 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 years LGAD sensors are been produced by CNM/Spain, HPK/Japan, FBK/Italy and recently NDL/China with different doping levels, active thickness, pad size, or inter-pad gaps. These detectors have been exposed to protons, neutrons and Xrays up to the expected maximum radiation levels (including the safety factors) and intensively characterized in laboratory (with probe station, β source, laser) or in beam tests (at CERN, DESY, FERMILAB).

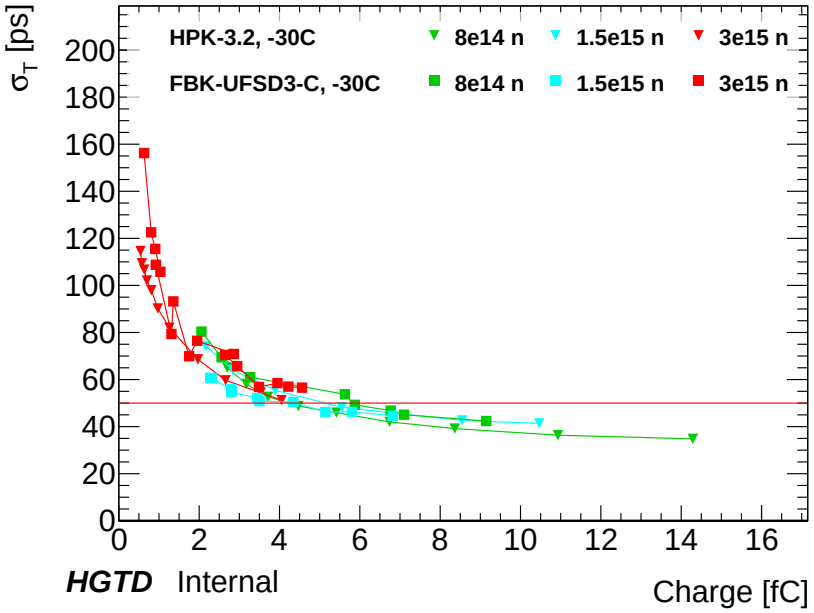


Figure 4.5: Time resolution as function of the collected charge for neutron irradiated LGADs from different producers (HPK, FBK) with a 50 µm active thickness. These measurements have been made at −30 °C, in the laboratory with a β source using a custom electronics read-out board (not the ALTIROC).

Under irradiation, the expected decrease of the charge yield can be mitigated by increasing the bias voltage (up to 700 V) and operating at low temperature (−30 °C). Fig. 4.5 summarizes results obtained in the laboratory, with dedicated electronics, for sensors from different producers exposed to neutron fluence up to $2.5 \times 10^{15}\,\mathrm{n_{eq}\ cm^{-2}}$. A charge of 4 fC can be reached up to a fluence of $2.5 \times 10^{15}\,\mathrm{n_{eq}\ cm^{-2}}$, providing a time resolution of about 50 ps.

The performance of sensors from all manufacturers is similar, even if before irradiation the optimal bias voltage might be different because the doping concentration is different. With a minimal charge of about 4 fC and a discriminator threshold of about 2 fC, a hit efficiency of 95% is expected. For the largest fluence, the Boron doping has mostly migrated from the avalanche region and the remaining small gain comes from 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 the ASIC performance at low charge

Intense R&D is still ongoing to improve the radiation hardness with deep narrow doping implantation, different doping (Ga instead of B), C implantation. Depending on the results of these studies, discussed in detail in Chap. 5, the exact radius of the inner and middle rings might be tuned. The working points for the bias voltage need to be adjusted with respect to the radiation flux: ongoing studies of the breakdown voltage should define safe criteria to operate the detector.

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 and the low leakage current.

Following almost four years of R&D activities, shared in part with CMS timing detector and RD50, a first set of criteria for the parameters of the final sensor design have been established, constituting our baseline. These include: 50 µm active thickness, narrow and deep doping profile, 70 µm inter-pad gap, and a 300 µm slim edge distance with two guards rings. However, some of the parameters will need to be further validated up to the Final Design Review, scheduled for 2021.

4.3.2 Front End ASIC

As discussed previously in combination with the LGAD sensor, the Front End ASIC is challenging. Taking into account the expected TID radiation levels and needed low jitter, the technology to be used is CMOS TSMC 130 nm. 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 for the the time measurement. Fig. 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.

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 contributes little 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 output provides the start of a second TDC, with 40 ps quantisation step, in order to measure the Time Over Threshold (TOT), which may be used as an estimate of the signal amplitude. The TOT information is used offline to correct the TOA for the 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.

The preamplifier and discriminator performance has been validated using a four-channel prototype (ALTIROC0), bump-bonded to a sensor of 2×2 pads first and a 25 channel prototypes (ALTIROC1-V2), bump-bonded to sensors of 5×5 pads including the TDC and SRAM. The complete analog Front End, adding the TDCs and a SRAM (ALTIROC1), was tested in the laboratory with the ASIC wire bonded to a specific board (see Fig. 4.7). This figure also shows preliminary results using the TDC for the jitter and the TOA variation as a function of an injected calibration charge. With the ASIC alone and an input capacitance of 4 pF, the threshold can be as low as 2 fC, allowing a measurement of an input charge of 4 fC with a calibration input signal. The jitter for an injected charge of 10 fC (4 fC) is about 20 ps (60 ps). The variation of the TOA versus the input charge, about 500 ps, is compatible with the preamplifier bandwidth. This time walk effect needs to be corrected. While the TOT can be used on testbench with the ASIC alone to correct it, some couplings are distorting the TOT distribution when bump bonded to sensor preventing to apply this correction. Investigations are on-going to understand the origin of these couplings. Preliminary measurement with beam show that a time resolution of 40 ps can be reached with non irradiated sensors

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 adjustable by slow control. In a second step, it formats these data, and provides them to the serializer, which 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^2C 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 it's final size. Triply redundant registers will mitigate against SEE and will be implemented for all controls and signals registers but not for the read-out data. The first iteration should be submitted in 2020 and a second iteration 1 year later. The Final Design Review is planned early 2022.

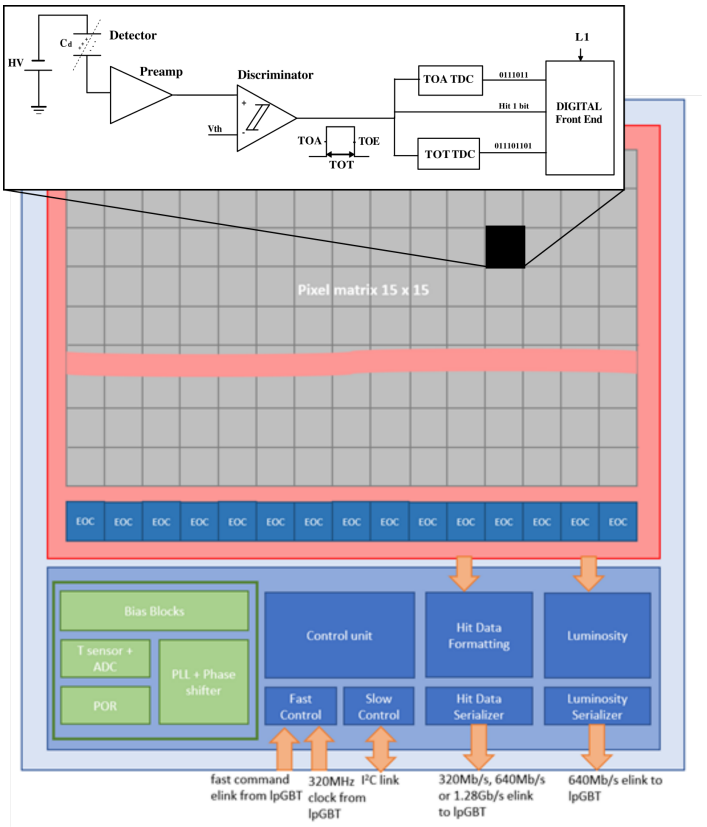


Figure 4.6: Global architecture of the ALTIROC ASIC. The schematic of one Front End electronics channel is displayed on top of the channels matrix, with the preamplifier followed by a discriminator, two TDCs, and a digital front end block.

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 Metal (UBM) will be deposited on the sensor wafer before dicing. The next step of the hybridisation consists in

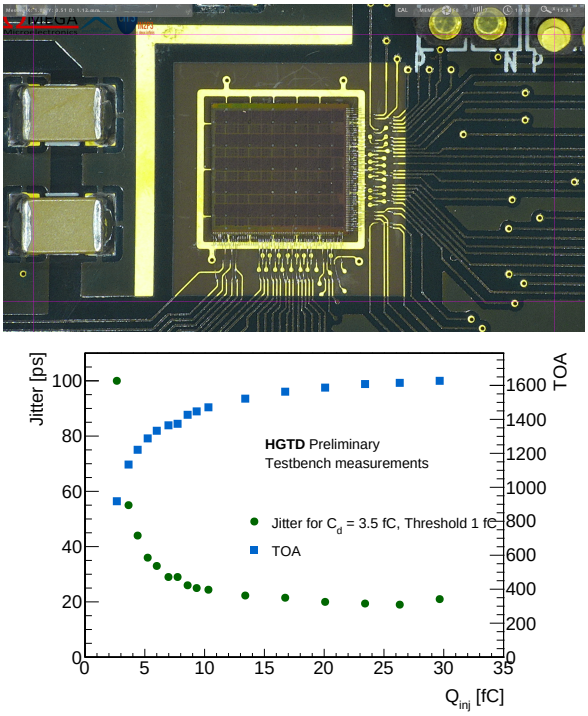


Figure 4.7: Picture of a ALTIROC1 die with 5 × 5 channels (left) and preliminary measurements of the average time and jitter as a function of the injected charge using calibration injection with one channel of ALTIROC1 (right). PLOT WITH ALTIROC1-V2/TDC IF POSSIBLE AND 2FC THERSHOLD . TO BE DONE BY NIKOLA

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 a few 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 channels sensors (both doing Under Bump MetaLlization and flip-chip). Satisfactory performance results have been obtained, both for connectivity and mechanical stress. The aim is to start the qualification of their processes in view of future production in Industry. The final design review of the bump-bonding process is planned at the end of 2022.

AS shown in Fig. 4.3 the bare module is glued to a small flex cable pcb, on which the ASIC signals are wire bonded. This small flex is connected to the the long flex cable tail though a connector in order to transmit the signals to the peripheral electronics boards.

Taking into account the space constraints, the flex tail is a two layers design with a maximum thickness of 220 µm. The longest readout row contains 19 modules. As displayed in Fig. 4.8,

each flex transfers four type 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, for the longest flex (L $\sim$ 60 cm) to 320 Mbit s^{-1} for the outer radius, for the shortest flex (L $\sim$ 15 cm). 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 ($<$ 200 mΩ) Digital and analog lines are separated.
- the bias voltage for the sensor (up to 800 V requiring excellent insulation.

The first flex cable prototypes, still made of a single piece and longer than required, have been manufactured in two companies and at CERN PCB workshop. When normalised to the maximum expected length of L $\sim$ 60 cm it satisfies the data transmission, bias voltage insulation and resistance requirements.

The R&D is still on going on the flex design to ensure it satisfies the strong thickness constraint along Z and to identify/develop reliable dense and thin mini-connectors for the connection between the module flex and the flex tail (in the module region) and the connection to the peripheral electronics board.A few companies have been contacted for these specific R&D and the final design review should take place in 2022.

The bare module will be glued to the flex small pcb board piece. Some tests of the glue to be used are ongoing in close collaboration with the ITk Pixel community, as the requirements are similar. To exercise the module assembly, real size modules made in a first step of heaters will be mounted in spring 2020 and with real modules (in 2021-2022). This activity will be done in the framework of the demonstrator activity (detailed in Chap. 14).

4.4 Module loading on support structure

The modules are loaded on an intermediate support structure where the modules are inserted and glued in pre-defined holes, insuring the exact position of each module and the alignment along the x and y readout row directions, as displayed in Fig. 4.9. These structures are later screwed to the cooling plates, using a thermal conductive grease to insure the direct contact of the modules with the Aluminium cooling plate. These support structures are separated in 3 radial regions to allow a fast replacement of the rings planned to take place at surface in the long shutdowns.

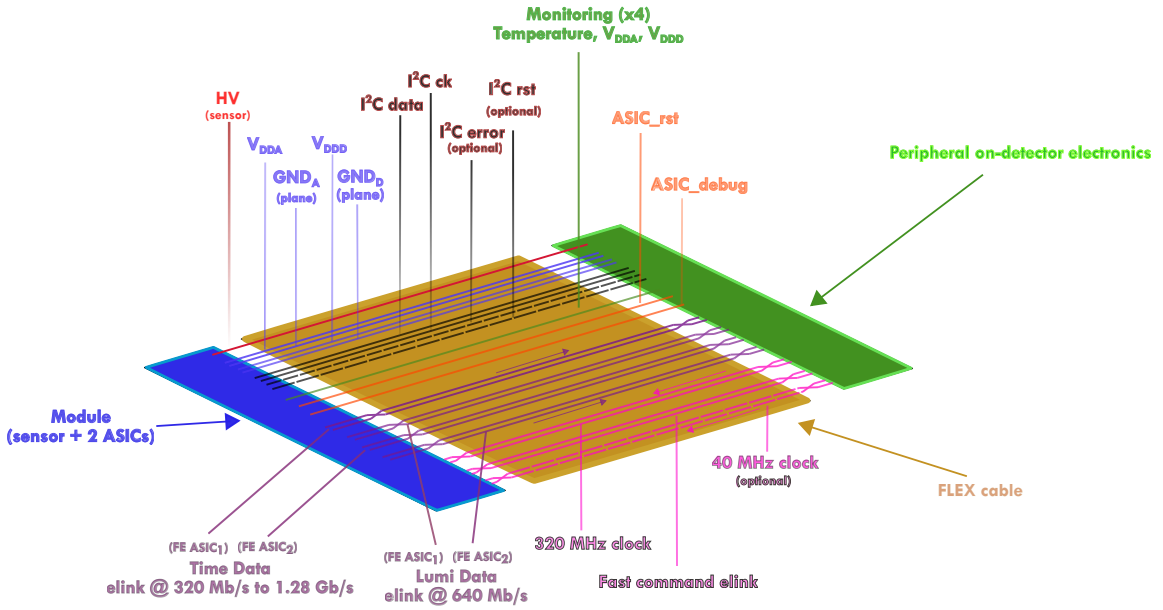


Figure 4.8: Signal transmitted from the ASICs to the peripheral electronics. Each ASIC has a dedicated e-link for luminosity and time data transmission while the other signals are common to both ASICs. The HV line is connected to the sensor.

4.5 Off detector electronics, calibration and luminosity

Fig. 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 to Peripheral Electronics Boards (PEB), where the electrical signals are encoded and transmited via optical link (at 10.24 Gbit s^{-1}) 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 Sec. 10.1.1. The main data stream is read out at L0 trigger rate and is meant for ATLAS event process. The luminosity stream is read out by dedicated FELIX boards which sum the number of hits over a large enough region to provide an accurate online luminosity measurement, the use of 16 regions per layer is under study.

4.5.1 Peripheral Electronics Boards

NEED TO UPDATE THIS SECTION.....

The PEBs are still at an early design stage and will use components already developed by CERN for the LHC upgrades. A single PEB will group together up to four read-out rows in

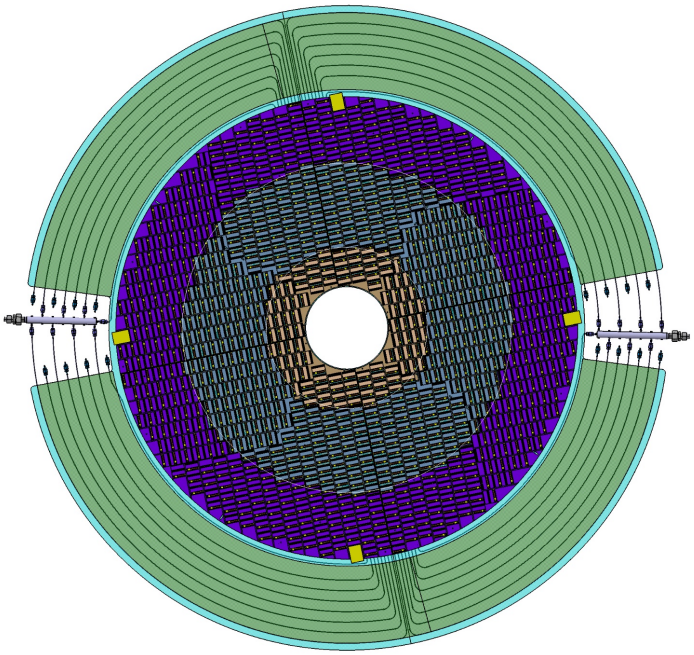


Figure 4.9: View of the modules mounted in the intermediate support structures. The module supports are separated in three radial rings (shown in different colours), and are screwed to the cooling/support plate. The cooling pipes circulating at the outer radius (in green) will serve as pre-heaters and will cool the peripheral electronics boards, to be located in this region.

order to optimize the numbers of components. The expected component densities are high, with the PCB carrying large numbers of signals with different properties. Typically, a PEB will contain:

- 3 to 8 lpGBTs both for the timing data and the luminosity data and the equivalent numbers of optical links.
- 6 to 8 VTRx optical receivers and transmitters
- 25 to 35 DC/DC converters, still to be optimized, depending on their performance. Both the BPOL12V and BPOL2.5V are expected to be used.
- 3 to 5 multiplexers developed in TSMC 130 nm as input of the monitoring signals to the lpGBT ADC.

The placement of these components needs a careful PCB design layout, to fit inside the allowed envelope dimensions (in z and r). R&D is still needed to develop flex connectors but also HV connectors. A first functional prototype of these PEBs is expected in 2020.

4.5.2 t_0 time calibration

The t_0 knowledge of each individual channel (3.59 millions of channels) is crucial to achieve the expected time resolution. The irreducible and non deterministic clock contribution to the the resolution is expected to be around 10 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 in time with the bunch crossing clock. The use of HGTD for the physics strongly relies on the relative comparison of the time of different channels within an event. Consequently the geometrical (time static) inter-calibration of all channels t_0 is the most crucial while global time drifts over large regions will have smaller impact on the performance.

Geometric and time static effects can be corrected with the calibration injection signals in the ASIC (different flex cable length, systematic difference between channels in ASIC due the imperfect clock tree distribution, etc.) or computed (geometrical time of flight). Calibration sequence between LHC fills will be used to monitored these calibration constants.

The variation with time of the 40 MHz phase, therefore of the t_0 is a correction to be determined in-situ using the 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 endcap. The calibration procedure will consist in measuring the inclusive average time of each channel with the data triggered at 1 MHz. Depending on the time period of these effects, and on the affected component and area (ASIC, module, group of ASIC 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.

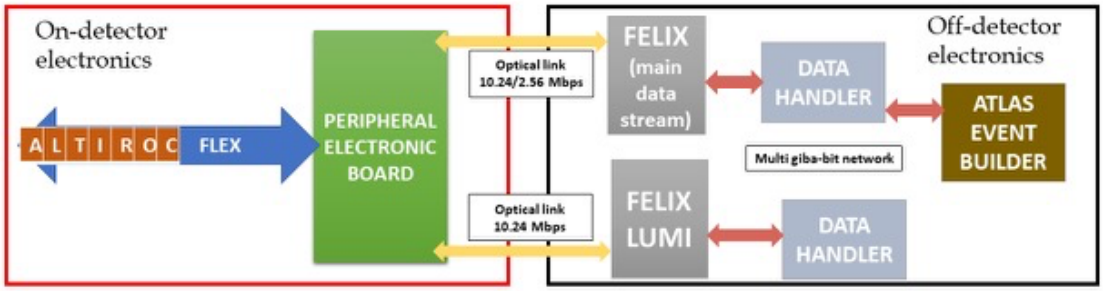


Figure 4.10: Data transmission paths for the main stream and the luminosity strem.

4.5.3 Luminosity

NEED TO FILL TEXT.....

4.6 Power distribution and detector control system

4.6.1 HV system

A schematic layout of the high voltage system is show in Fig. 8.1 and detailed in Chap. 8. Each of the 8032 modules require 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 2.5×10^{15} n_{eq} cm^{-2}, as detailed in Chap. 5. The irradiation of each module will strongly depend on its radial position in the detector, seen in Fig. 4.2, with an expected maximum variation of < 15% inside each module. The ultimate goal is to use individual adjustable voltages for each module, to allow for optimal operation. As a compromise of costs and performance, at least at the start of the HL-LHC, two to three modules (instead of one) will be connected to one HV channel in the outer region of each ring that have similar radial position.

This merging can be done either in USA15 services cavern, where the HV power supplies will be located or in the patch panels region , exact location still to be decided. All HV cables will be routed from the beginning to allow in a later stage, to feed one module from one HV channel. Consequently HV power supplies that deliver up to 800 V and up to 6 mA current are needed in order to feed simultaneously two modules. The power supplies will be based on commercial multi-channel rack mounted units. 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, allowing to perform the necessary HV adjustments.

4.6.2 LV system

A schematic layout of the LV voltage system is shown in Fig. 8.2 and detailed in Chap. 8. The low voltages needed by the front-end and peripheral electronics will be able to deliver almost 20 kW and will be provided in a three stage system. Bulk power supplies, located in USA15, will provide 300 V DC current 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 which is distributed to radiation hard DC-DC converters, located on the peripheral electronics boards. 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. 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 Chap. 8.

4.7 Mechanics, Services and Infrustructure

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, sharing the space with ITk and the Tile calorimeter crack counters. The use of light structures were prioritized to minimize the amount of material in front of the active layers and minimize the potential increase in the radiation levels, leading to a total detector weight per end-cap of approximately 350 kg (275 kg without the external moderator).

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 Fig. 2.4. The front cover is divided in two half disks to allow it's 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 are 13 mm and 7 mm respectively. To avoid condensation in the external face of the HGTD vessel during operation, heaters will be placed on the external face of the front cover, insuring a minimal temperature of 20 °C outside the HGTD vessel. An air gap of 3 mm will be kept between the HGTD detector and the end-cap LAr calorimeter.

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 980 mm outer radius. Their shape allow for a completion, in case of delays, in the ATLAS detector even when the beam pipe is in place, provided that the back vessel covers 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 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 20° with respect to each other, as seen in Fig. 11.15, allows for a better integration of the cooling pipes inside the vessel while minimising the regions with zero hits resulting from the dead zones between the staves and imperfect coverage in the inner most radius.

The expected maximum power consumption of the detector, to operate at −35 °C and to reach the required performance, amounts to 39.3 kW in total (19.7 kW per end-cap); details of the various components are summarized in Tab. 11.1. An evaporative CO_2 cooling system of (50 kW will be used and part of its infrastructure and cooling spare unit will be shared with ITk.

The evaluation of the amount of services required to operate the detector, summarized in Tab. 12.1, and respective routing design was subject to a careful evaluation and optimisation. This is due to the limited space in the detector vessel outer ring allocated to the services feedthroughs, limited space in the barrel-end-cap calorimeter gap region and, last but not least, in the ATLAS flexible chains that allow maintaining part of the services connected during the end-cap calorimeters opening and closure. The detector services routing is shown in Fig. 4.11 for the calorimeter extended barrel face. The cables, exiting in four layers in the feedthroughs region, will pass to one layer at $r > 1.3$ m to fit within an envelope of 17 mm. At the outer radius of the 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. The exception will be a dedicated slot in ϕ, on the top of the calorimeter, to be given from the ITk original envelope, to route the four CO_2 cooling pipes of 50 mm diameter maximum each pipe. The priority for services installation in flexible chains, still to be confirmed, 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 the patch panels (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 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 and lowered directly from the surface to the minivans. The final installation of the detector should take approximately 1 month per end-cap and is planned for April 2026 (HGTD A) and January 2027 (HGTD C).

Dedicated tools are needed for assembly, lowering, and final installation of the detector. These tools are still at a conceptual stage and where possible will use synergies with already developed tools for other sub-detectors.

The overall commissioning will start immediately after the connectivity of the services to the detector. The access to the detector components during the commissioning should be possible until approximately May 2026, close to the expected end-cap calorimeters closure. This will leave at least 6 months of intense commissioning while access is still possible. Both the installation and commissioning of HGTD will be done with the participation of several collaborating Institutes.

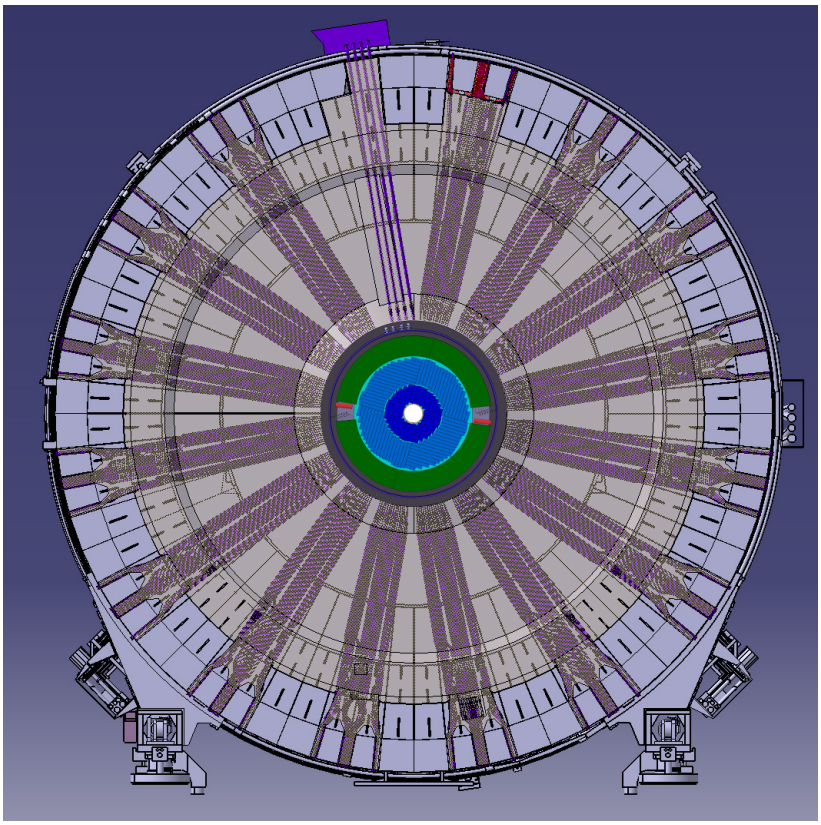


Figure 4.11: Transverse view of HGTD with services routed along the end-cap calorimeter face. UPDATE THIS FIGURE WITH LATEST VERSION BY VLADIMIR/SERGEY

4.9 Next steps towards construction

NEED TO FILL TEXT.....

5 Sensors

5.1 Sensor parameters and requirements

The HGTD sensor parameters and requirements are summarized in Tab. 5.1. The sensors are intended to provide a fast signal in response to charged particles for a time resolution per hit of 40 ps at the start and 70–85 ps at the end of lifetime (combined performance with the electronics and other contributions). The charge 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 below 300 µm. 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 less than 500 µm. In the baseline scenario, discussed in Chap. 4, the innermost part of the detector (r < 230 mm) should be replaced after each 1000fb^{-1}and the middle ring within 470 mm $> r >$ 230 mm should be replaced at half lifetime (2000fb^{-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 2.0 MGy, including safety factors.

The leakage current should be less than 5 µA per pad, the applied bias voltage less than 800 V 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 (charge) is 20 (10 fC) at the start and at least 8 (4 fC) at the end of lifetime.

5.2 Low Gain Avalanche Detectors

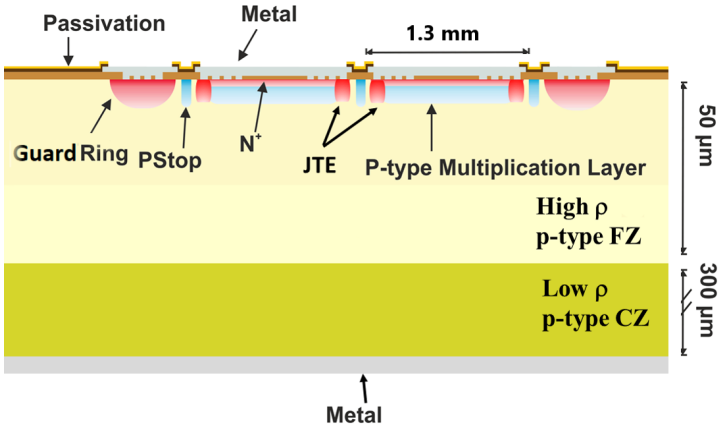
5.2.1 Overview

LGADs are segmented planar Silicon detectors with internal gain as sketched in Fig. 5.1. The gain depends on the doping dose of the multiplication layer as seen in Fig. 5.2 and diminishes with radiation fluence as shown in Sec. 5.5.3. They have been pioneered by the Centro Nacional de Microelectronica (CNM) Barcelona [4] and developed during the last 5 years within the CERN-RD50 community [3] also in collaboration with other LGAD vendors as Hamamatsu Photonics (HPK, Japan) and Fondazione Bruno Kessler (FBK, Italy). An

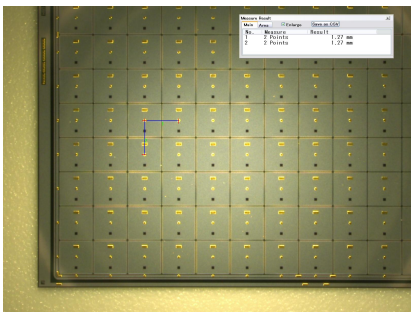
Technology	Silicon Low Gain Avalanche Detector (LGAD)
Time resolution $2.4 < \|\eta\| < 2.7$	≈ 40 ps (start); ≈ 70 ps (end of lifetime)
Time resolution $2.7 < \|\eta\| < 3.5$	≈ 40 ps (start); ≈ 70 ps (end of lifetime)
Time resolution $3.5 < \|\eta\| < 4.0$	≈ 40 ps (start); ≈ 70 ps (end of lifetime)
Gain	≈ 20 (start); > 8 (end of lifetime)
Minimal charge	4 fC
Hit efficiency	>95%
Granularity	1.3 mm $\times$ 1.3 mm
Physical thickness	<300 µm
Active thickness	50 µm
Active size	39 mm $\times$ 19.5 mm (30 $\times$ 15 pads)
Inactive edge	<500 µm
Radiation tolerance	2.5×10^{15} n_{eq} cm^{-2}, 2.0 MGy
Maximum operation temperature on-sensor	-30 °C
Maximum leakage current per pad	5 µA
Maximum bias voltage	800 V
Maximum power density	100 mW/cm^2

Table 5.1: Sensor parameters and requirements.

introduction of the technology is given in Chap. 4. Additional background and details are given in Reference [32].



(a) Cross section of a 2 $\times$ 2 array.



(b) Photo of a 15 $\times$ 15 array.

Figure 5.1: (a) Cross section of a 2 $\times$ 2 array including a JTE around each sub-pad (SiSi wafer, CNM design) [33]. (b) Microscope photo of an HPK-3.1 15 $\times$ 15 array.

Three major effects determine the time resolution: time walk from amplitude variations, jitter from electronic noise and "Landau fluctuation" from charge deposition 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

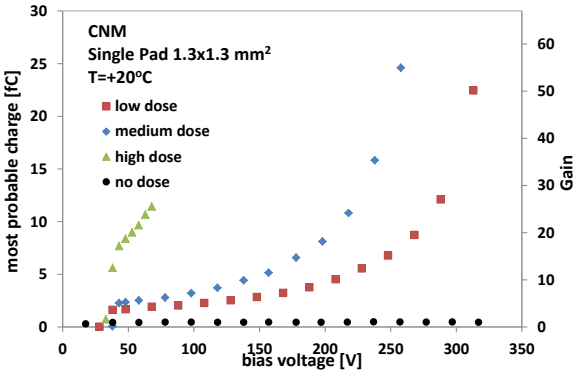


Figure 5.2

Figure 5.3: Gain and charge as a function of bias voltage for CNM LGADs with different doping concentration of the multiplication layer.

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 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 for by a large extent using time reconstruction algorithms such as constant-fraction discrimination (CFD) or amplitude or time-over-threshold (ToT) corrections. The third effect called "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 Fig. 4.5 as a function of gain, with the time walk corrected using CFD. As expected from Eq. (5.1) the resolution improves with increasing gain due to the reduced noise jitter, but then levels off to the Landau fluctuation of about 30 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 fulfills already the HGTD requirement of 40 ps per hit at the start of operation (see Tab. 2.1) and

is improving only slowly when going to higher gains as seen in Fig. 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 Sec. 5.5.3).

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

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.

Fig. 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). Fig. 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 or the substitution of B by Gallium (Ga) are investigated as candidates for improved radiation hardness.

5.2.2 LGAD productions

At present, LGADs have been produced in six manufacturing sites, shown in Tab. 5.2 along with their production capabilities: Hamamatsu Photonics (HPK), Japan; CNM, Spain; Fondazione Bruno Kessler (FBK), Italy; Micron, UK; Brookhaven National Lab (BNL), USA; and 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 [34] 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 into which the collaboration is now entering the focus will be geared towards the production of sensors for HGTD specific application once the sensor requirements are better understood, 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 ground 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[1]), etc.) and operations (bias voltage, power, reliability, breakdown) have become more important.

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

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

The results in the following have been mainly obtained from the LGAD types shown in Tab. 5.3. These runs 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).

5.3 Radiation damage and irradiations

As explained in Sec. 2.4, the detector has to withstand a 1 MeV neutron equivalent particle fluence of maximally $2.5 \times 10^{15}\,n_{eq}\,cm^{-2}$, assuming one replacement of the inner part after half of the total integrated luminosity of 4000 fb^{-1}. In the innermost region, the radiation field is roughly equal for neutrons and charged hadrons (Fig. 2.13), 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 energy spectrum of protons and pions roughly peak between 50 MeV and 10 GeV, whereas the neutron spectrum peaks

[1] UBM is part of the hybridisation process as explained in Sec. 7.2.1.

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-BV60	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	220/90	37→57	200→ 500	5 × 5
NDL	NDL-BV60	35	70	55	450	15 × 15

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

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 [3]. 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 [35, 36], 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 $6 \times 10^{15}\,n_{eq}\,cm^{-2}$ at various facilities with different particle types and energies that are representative for the ones expected in HGTD. Tab. 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 is used for the conversion of the actual particle fluence to the 1 MeV-neutron equivalent fluences, which is used throughout this document.

First prototypes were irradiated in all facilities except for CYRIC, and it was found that acceptor removal seems to be faster after irradiations with 200 MeV–23 GeV charged hadrons than with neutrons [35, 36]. However, CERN PS is in shutdown now until 2021 and campaigns in Los Alamos are still ongoing. Hence results for LGADs with the HGTD geometry presented here are mostly after irradiations with neutrons at Ljubljana and 70 MeV protons at CYRIC due to the sites' availability. As will be shown below (see e.g. Sec. 5.5.3), the performance of samples irradiated at these two sites at similar 1 MeV-neutron equivalent fluences is similar, in contrast to the earlier results using higher energy charged hadrons. These studies will be followed up by irradiations with higher energy charged hadrons at PSI and Los Alamos when these facilities become available again in 2020. Also mixed neutron-proton irradiations are planned.

It should be noted that irradiations at CYRIC with 70 MeV protons led to a maximum TID of 4.0 MGy, close to the HGTD requirement of 2.0 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 Sec. 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	5	4.0	HPK-3.1/3.2, NDL FBK-UFSD3-C
Los Alamos (*pLA*)	800 MeV p	0.7	0.43	1	0.4	early prototypes, HPK-3.1
CERN PS (*pPS*)	23 GeV p	0.6	0.44	6	2.7	early prototypes
PSI (*pi*)	192 MeV pions	1	0.32	2	0.6	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 in 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 subsequently to 5×5 sub-blocks. A probe card with 15×15 contacts is under development. An alternative is the subsequent 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 [32, 35–41] and lasers, as well as in beam tests [37, 38, 42]. Beam tests have been performed by the HGTD community in more than ten periods between 2016 and 2019 at the H6 beam line of the CERN SPS 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 [42]. Data were taken in two modes: stand-alone and integrated into a beam telescope that provided track position information with about 3 µm precision.

Most of the measurements on irradiated sensors were performed at a temperature of −30 °C, the lowest temperature reachable in standard laboratory climate chambers.

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 [37, 42] 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. This reflects the performance of the sensors with discrete electronics optimized for precision timing. It should be clearly noted that measurements with the ALTIROC are not part of this section since the chip has not been available yet for large-scale sensor testing. The measurements presented here will be repeated with the ALTIROC as soon as enough chips are available. First measurements of the combined sensor-ALTIROC performance on few bump-bonded hybrid prototypes are presented in Sec. 6.7.2.

Position-sensitive scans using red and infrared laser 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 βs and MIPs of 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. Either at least two LGADs are used or LGADs and a fast Cherenkov counter based on quartz bars and a Silicon photo multiplier (SiPM) with typical time resolution of about 10 ps. 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. Timewalk effects are usually corrected for using time reconstruction algorithms such as the Constant-Fraction Discriminator (CFD), the Zero-Crossing Discriminator (ZCD) or corrections using the amplitude or Time-Over-Threshold (TOT) of the signal [42].

LGAD behavior such as time resolution and collected charge was simulated using the software WeightField 2 [43]. The simulation were tuned using laboratory measurements from different sensor types and are extrapolated to foresee future improvements as seen in Sec. 5.8. Also, the software TCAD sentaurus [44] was used in aid of the sensor production.

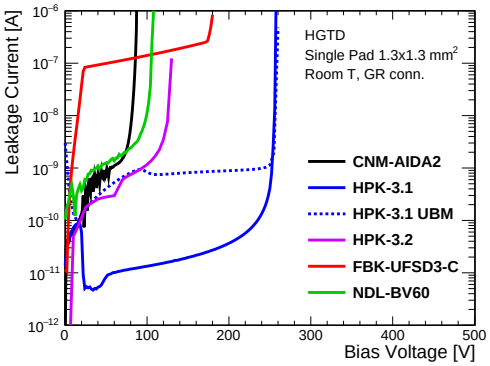
5.5 LGAD performance before and after irradiation

5.5.1 Electrical characterisation: I-V and C-V

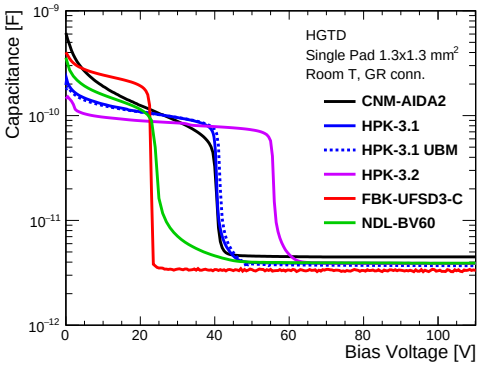
Fig. 5.4(a) and Fig. 5.4(b) show the I-V and C-V curves of 1.3 mm × 1.3 mm LGAD pads of different vendors and runs, measured with the guard ring (GR) connected to ground. Most of the vendors and runs achieve nA leakage current levels or below before breakdown, well below the ALTIROC leakage current limit 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 to about 1 nA with respect to wafers without UBM, 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.

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 × 1.3 mm LGAD pads (measured with a connected guard ring) are consistent with active thicknesses of 40 µm–60 µm.

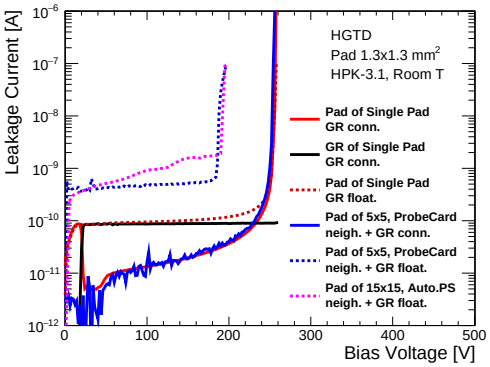
Fig. 5.4(c) shows the I-V curves for HPK-3.1 sensors of the LGAD pad and guard ring (GR) with either GR connected to ground (as the pad) or floating. For the single pad sensor, it can be seen that the current through the pad in case of floating guard ring is roughly the sum of pad and guard ring current in case the guard ring 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-3.1 array, the I-V curve is found to be almost identical to the one of a single pad in case the neighbors and the guard ring are connected to the same potential, as measured with a 5 × 5 probe card on a 5 × 5 array (see Fig. 5.4(c) and Fig. 5.5(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, consistently measured with a probe card when connecting only one channel and an automatic probe station with only one needle (see Fig. 5.4(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 and is so far the only method



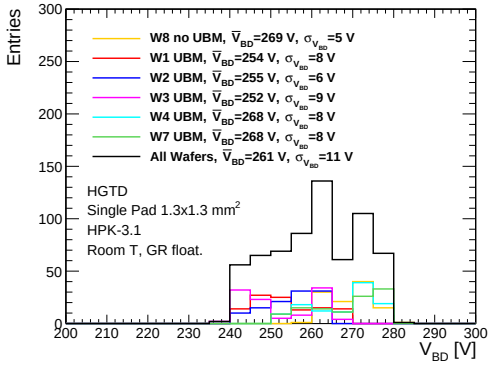
(a) I-V for different runs.



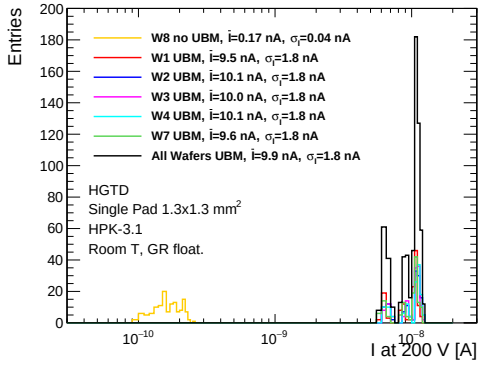
(b) C-V for different runs.



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

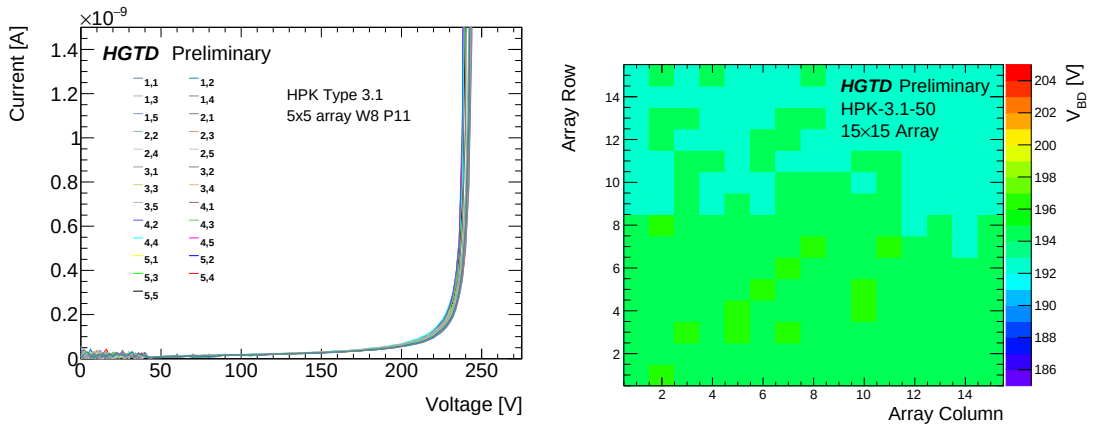


(d) V_{BD} distribution.



(e) Current at 200 V distribution.

Figure 5.4: 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 5.5: (a) I-V measurement of 25 pads from an unirradiated HPK-3.1 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-3.1 array without UBM measured with an automatic probe station at room temperature (neighbors and GR floating). TODO: change with HPK 3.2?

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 Fig. 5.5). The V_{BD} spread between pads in an array is found to be typically between 1 V and 2 V.

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

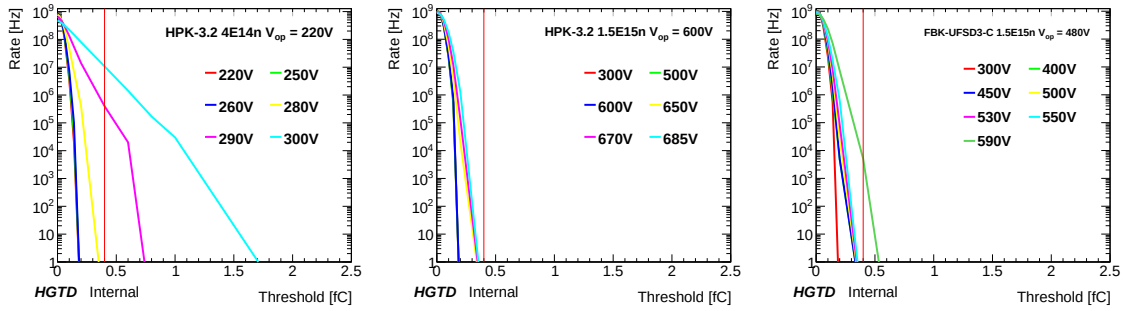
HGTD institutes measured a large number of single pads and arrays from different productions, in particular HPK-3.1/3.2 and CNM-AIDA1. TODO: Add NDL/FBK? HPK also provides their in-house Quality-Control (QC) results with an automatic probe station (GR floating) of each single pad they delivered. The HPK results have been verified by HGTD institutes. Fig. 5.4(d) and Fig. 5.4(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%.

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
FBK-UFSD3	2×2	?	40?	x	x	x	x
	5×5	?	40?	x	x	x	x
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 (IP) gap designs. An array of 15×15 pads corresponds to the final ALTIROC size and half of the full final sensor area. TODO: update CNM? Add NDL/FBK?

Tab. 5.5 shows the fraction of good individual pads in 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 99.7–100%. No dependence on the edge design between 200 µm and 500 µm edge was found. Only one HPK-3.1 and one HPK-3.2 15×15 HPK array with one or two bad pads, respectively, were found, all other sensors were classified as perfect. 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 1.5E15n (c) FBK-UFSD3-C at 1.5E15n

Figure 5.6: Self-trigger rate as a function of collected charge. Fig. 5.6(a): for HPK-3.2 sensor at 4×10^{14} n_{eq} cm^{-2} of neutron irradiation. Fig. 5.6(b): same sensor after 1.5×10^{15} n_{eq} cm^{-2} of neutron irradiation. Fig. 5.6(c): for FBK-UFSD3-C sensor at 1.5×10^{15} n_{eq} cm^{-2} of neutron irradiation. A threshold of 5 mV corresponds to roughly 0.4 fC of collected charge, repersented by the 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 Sec. 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 so that the charge collected from the bulk before charge multiplication does not change much even after irradiation to 1×10^{16} n_{eq} cm^{-2}.

The operating voltage (V_{op}) is defined as a stable and safe operation voltage where the sensor has reasonable performance in term 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 (similar to LHC conditions 40 MHz) without the risk of inducing breakdown or electrical arcing between the sensor structures (see Sec. 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 previous voltages. The maximum leakage current allowed is limited to 5 µA per pixel and the power less than 100 mW/cm^2.

Furthermore at this voltage the sensor must not present self-triggering events (events caused by discharges not caused by particle hitting the detector) with a rate higher than 1 kHz for a trigger threshold of ±5 mV or collected charge of 0.4 fC. An excessive self-triggering would increase the dead time of the HGTD detector hindering its operation and is potentially detrimental.

This was studied in detail for HPK-3.2 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 around 30. This statement is valid both for unirradiated and irradiated (with neutrons/protons) sensors of HPK-3.2 and FBK as shown in Fig. 5.6. For a neutron fluence of 1.5×10^{15} n_{eq} cm^{-2} even at the highest voltage no self-triggering is observed since the gain is low.

Fig. 5.7 shows V_{op} as a function of fluence after neutron irradiation for diff erent LGAD types. It can be seen that it increases with fluence up to maximally about 750 V for 50 µm sensors. Sensors of 30 µm can only sustain about 500 V maximally (see Sec. 5.5.7). The minimum voltage tested is usually chosen as when no clear signal is seen (usually the case for irradiated sensors) or the sensor is believed to be not fully depleted (usually the case for non irradiated sensors).

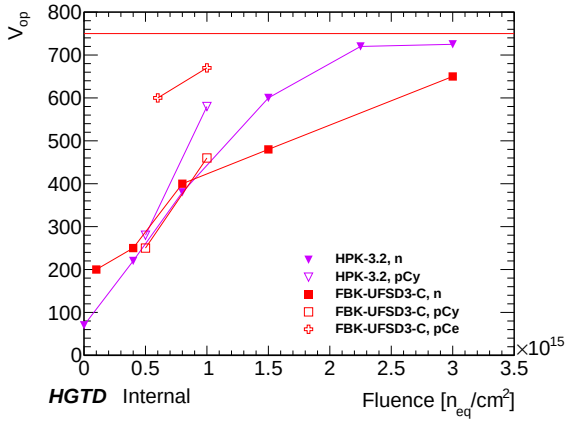
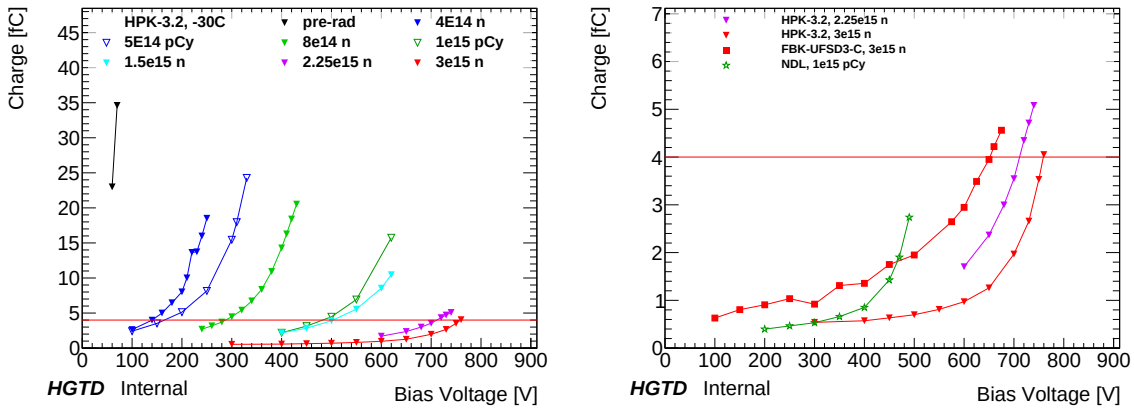


Figure 5.7: V_{op} as a function of fluence after irradiation for different LGAD types.



(a) (b)

Figure 5.8: Collected charge as a function of bias voltage for different fluences for HPK-3.2 and for all vendors at maximum fluence. 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.

5.5.3 Collected charge and gain

Fig. 5.8 shows the collected charge as a function of bias voltage after neutron and proton irradiation up to 3×10^{15} n_{eq} cm^{-2} for different LGAD types: HPK-3.2, FBK-UFSD3-C and NDL sensors. The charge at V_{op}, as defined in Sec. 5.5.2, and 95% of V_{op} as a function of fluence is shown in Fig. 5.9. The sensors are measured up to the maximum safe bias voltage V_{max} where there is no risk of sensor breaking and no self triggering is observed.

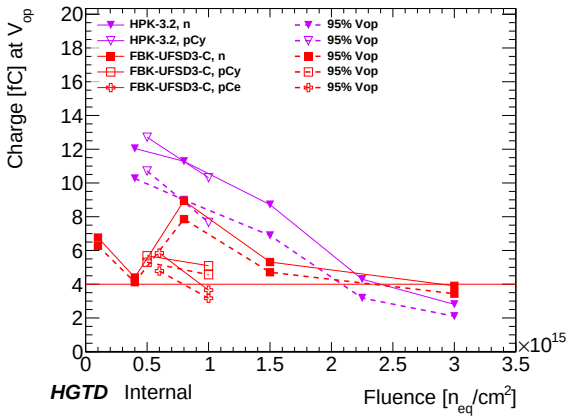


Figure 5.9: The charge at V_{op} and 95% of V_{op} as a function of fluence. Solid markers indicate n irradiation (*n*), open markers p irradiation at CYRIC (*pCy*) or CERN (*pCe*)

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 Sec. 5.5.4) and time resolution including the ALTIROC jitter (see Sec. 6.7). This level is indicated by the horizontal lines. The corresponding "charge bias working point", i.e. the bias voltage needed for 4 fC, as well as the difference V_{diff} between V_{max} and the bias working point WP (indicating the amount of bias head room for sensor operation) are shown in Fig. 5.9.

The following observations are made for the different types: TODO: 3.1 is not anymore the baseline, reorder

a. Baseline 50 µm sensor (HPK-3.1)

HPK-3.1 can reach the target charge of 2.5 fC up to 1.5×10^{15} n_{eq} cm^{-2}. The bias working point is larger than for other types, however with a sufficient head room. At a fluence of 3×10^{15} n_{eq} cm^{-2}, only 2.0 fC is reached. Measurements after neutron and 70 MeV proton irradiation seem to give similar results at similar fluences, although it is difficult to compare directly since due to technical reasons not exactly the same fluence points could be taken. In the future more studies with protons, also at higher energies, will be carried out.

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

HPK-3.2 sensors have a deeper and higher-dose multiplication layer, which leads to a reduced acceptor removal rate. Hence, this type can reach the target charge of 2.5 fC up to the higher fluence of 3×10^{15} n_{eq} cm^{-2}. The bias working point is less than for HPK-3.1 and exhibits a sufficient head room. At a fluence of 6×10^{15} n_{eq} cm^{-2}, only 2.0 fC is reached.

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

The main contributor to radiation damage in LGADs is the acceptor removal, i.e. the reduction of the doping concentration of the gain layer which results in loss of gain. Addition of Carbon in the gain layer reduces the acceptor removal. The required bias voltage is thus lower than for other types with a large head room. Irradiations up to 6×10^{15} n_{eq} cm^{-2} are ongoing. Further studies to extend the beneficial effect of carbon to higher fluences are envisaged.

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 [42]. Fig. 5.10 shows the efficiency as a function of most probable charge collected, compiled from 16 different single pad sensors before and after irradiation 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.

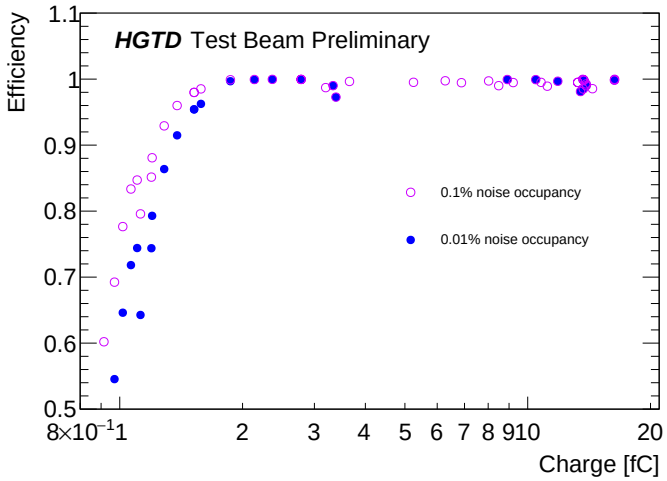


Figure 5.10: Hit efficiency as a function of collected charge at a measured noise occupancy of 0.1% and 0.01%.

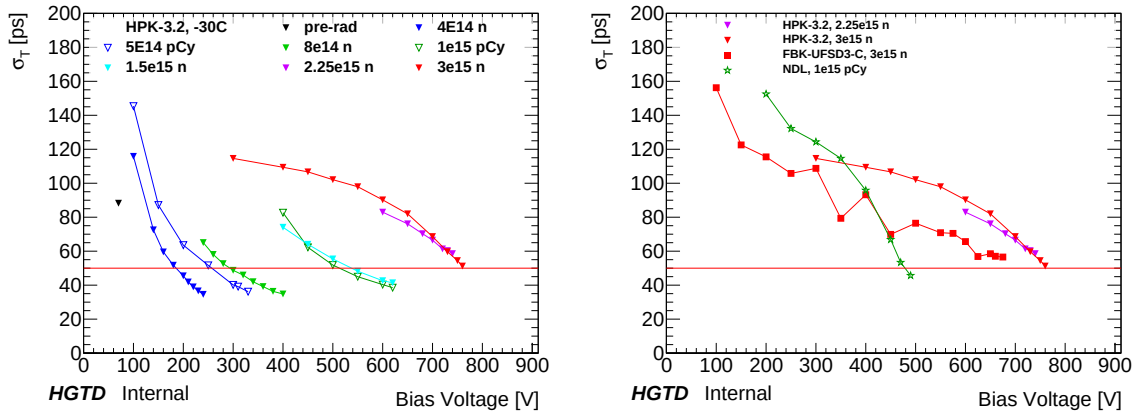
It can be seen that a universal curve is obtained, irrespective of fluence, indicating that the charge is the main parameters on which the hit efficiency depends, given a certain noise occupancy. A hit efficiency above 99% is obtained at the HGTD target working point of 2.5 fC mentioned in Sec. 5.5.3 for both noise occupancy working points. The measurements will be repeated with the ALTIROC electronics once available for large-scale testing.

2D efficiency maps are shown in Sec. 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 resolutions of LGAD devices have been extensively studied in various beam tests [37, 38, 42] and ^{90}Sr setups [32] on custom-made HGTD-specific readout boards. It should be noted again that the measurements on such test boards exhibit noise and jitter properties different from the final ALTIROC electronics, for which these measurements will be repeated once available for large-scale testing. Hence, the results presented in this section are best-case scenarios for the time resolution and demonstrate what is possible when not limited by power and size constraints of a readout chip. For the remaining part of the TDR the time resolution is not the one presented here with HGTD-specific readout boards. Instead, the measured charge of the sensor was taken as an input to the ALTIROC time resolution vs. charge function (see Fig. 3.6). This is to have a more realistic estimate of the final time resolution with the ALTIROC.



(a) (b)

Figure 5.11: Time resolution as a function of bias voltage for different fluences for HPK-3.2 and for all vendors at maximum fluence. Solid markers indicate n irradiation (*n*), open markers p irradiation at CYRIC (*pCy*). Measurements were performed at −30 °C.

On custom-made HGTD-specific readout boards, it has been consistently shown that sub-30 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 [32, 37–42].

The time resolution of HPK-3.2 was measured (up to a maximum voltage Sec. 5.5.2) in the β–telescope after irradiation with 1 MeV neutrons at Ljubljana, and 70 MeV protons at CYRIC. The results shown in **??** indicate that a resolution of 35 ps and better is achieved up to a fluence of $1.5 \times 10^{15}\,\mathrm{n_{eq}\,cm^{-2}}$. At a fluence of $3 \times 10^{15}\,\mathrm{n_{eq}\,cm^{-2}}$, the resolution at V_{max} is slightly over 50 ps.

The time resolution for FBK-UFSD3-C with Carbon addition is around 40 to 50 ps, generally worse than HPK-3.1/3.2-50, until $1.5 \times 10^{15}\,\mathrm{n_{eq}\,cm^{-2}}$. For a fluence of $3 \times 10^{15}\,\mathrm{n_{eq}\,cm^{-2}}$, it reaches slightly over 50 ps, similarly to HPK 50 µm.

In Fig. 5.12 the time resolution at V_{op} is shown. A time resolution of 60 ps is reached for almost all sensors at all fluences with sufficient head room until $1.5 \times 10^{15}\,\mathrm{n_{eq}\,cm^{-2}}$, while 50 ps is not.

5.5.6 Uniformity, inter-pad gap and edge region

One crucial 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 foreseen, 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

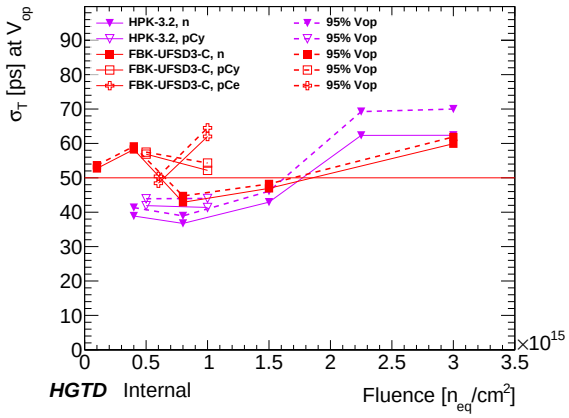


Figure 5.12: The time resolution at V_{op} and 95% of V_{op} as a function of fluence. The WP voltage curves are shown with a continuous line, the dotted line represents the voltage difference V_{diff} between WP and maximum bias voltage (head room). The results for HPK-3.1 include both n and pCy irradiation, the others are for n irradiation.

arrays comprehending 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 Sec. 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. 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 position. The hit efficiency and time resolution map for a 2×2 array is shown in Fig. 5.13 before and after irradiation.

In the laboratory the sensors were tested with an infrared laser of 1060 nm wavelength focused to 10 µm–20 µm FWHM. 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 the pulse maximum are fitted with a step function 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 first pad and rises to 50% on the neighbour (50%-50% point). The inter-pad scans for HPK-3.1 can be seen in Fig. 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-3.1/3.2-50 and CNM-10478-50 and CNM-AIDA-50 sensors can be seen in Fig. 5.14(b). As shown in Tab. 5.6, the lowest

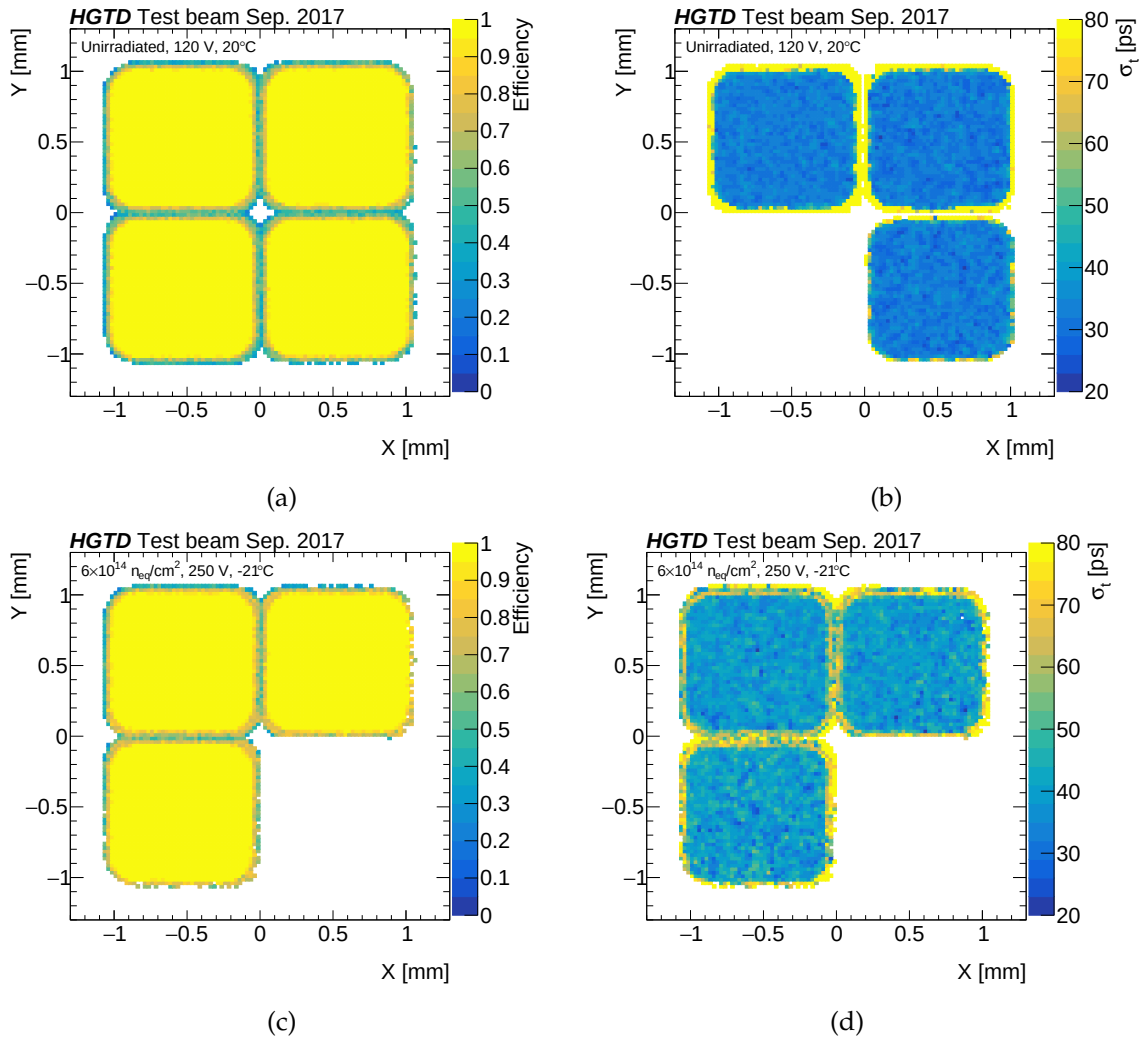
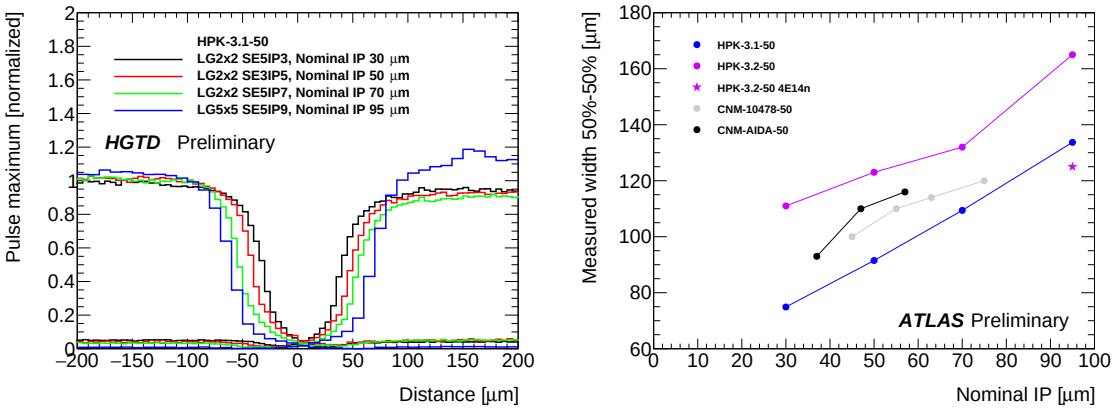


Figure 5.13: 2D maps of efficiency (left) and time resolution (right) before (top) and after n irradiation to $6 \times 10^{14}\,\mathrm{n_{eq}\ cm^{-2}}$ (bottom) for a 2×2 array from CNM-10478-50 as measured in HGTD beam tests [42]. Sometimes only 3 channels were measured. The efficiency was evaluated at a threshold of 3 times the noise here. 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 with a spread of 3 ps in the pad center.

measured value per type (roughly 70, 90, and 110 µm for HPK-3.1, CNM-AIDA-50 and HPK-3.2) correspond to fill factors of 90%, 87% and 84%, respectively. HPK-3.2 shows an inter pad gap that is significantly larger than HPK-3.1 before irradiation, however after a small irradiation of $4 \times 10^{14}\,\mathrm{n_{eq}\ cm^{-2}}$ its performance is in line with HPK-3.1. Further studies after irradiation will be done in the near future.

The edge area is evaluated in a similar way by scanning over the edge of the sensor pad



(a) IP laser scans

(b) IP correlation

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

Effective IP gap	Fill factor
70 µm	90 %
90 µm	87 %
110 µm	84 %

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

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 observed for the different IP gaps and edge distances.

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 tens of hours. 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 750V. No change in the behavior of sensors was observed in the timescale of tens of hours. Tests are ongoing with unirradiated and irradiated sensors. TODO: Update results when available. Add statement that no change was observed after hours of intense charge injection.

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 Sec. 5.5.2. In addition, the sensor can be protected by operating them at the minimum bias working point at which the requirements for collected charge and time resolution are satisfied instead of V_{op}, as shown in Fig. 5.9 and Fig. 5.12.

The difference (V_{diff}) in bias voltage between V_{max} and V_{op} is the "safety head room" and is usually quite large at low to medium fluences, although it was found that for thin sensors (below 30 µm) this head room is small and might impact their usefulness. For very high fluences the "head room" diminishes drastically for all sensors.

During the LGAD R&D phase, these principles were explored with existing sensors listed in Sec. 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 can be reached for the four sensor types that were tested in depth.

It was discovered that sensors break when the bias exceeds the critical bias voltage V_{crit}. Since V_{op} increases with fluence almost all breaking occurred at high fluences. The thickness of the sensors is an important parameter. It was observed that thin sensors would break immediately when reaching a certain V_{crit} which depends on the sensor thickness. Sensors with thickness of 50 µm thickness (like HPK-3.1, HPK-3.2) would break for bias voltages greater than 750 V. 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 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 $1.3\,\mathrm{mm}^2 \times 1.3\,\mathrm{mm}^2$ to check the performance in case of unpredicted situations where sensors would be exposed to longer times at elevated temperatures or when intentional annealing may be used to reduce leakage current and power dissipation.

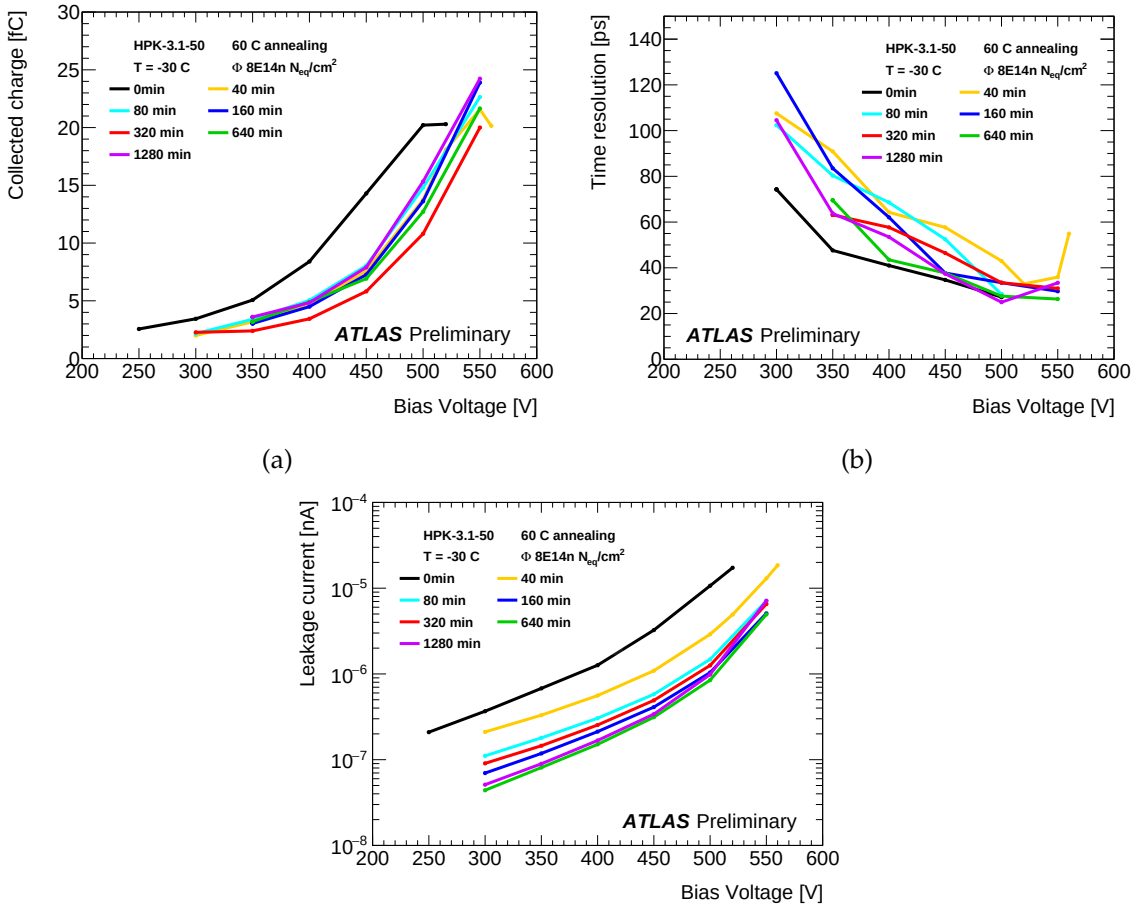


Figure 5.15: Voltage dependence for different annealing times for (a) collected charge, (b) time resolution and (c) leakage current.

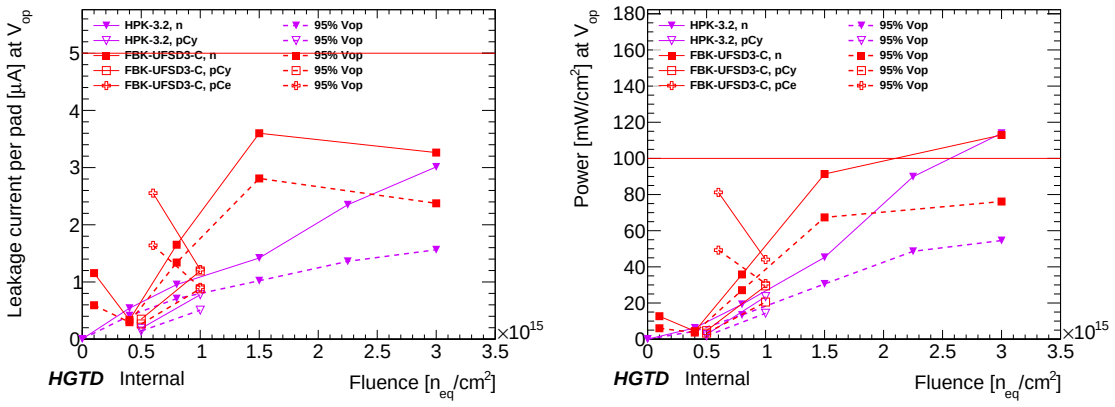
The dependence of collected charge on bias voltage for different annealing times is shown in Fig. 5.15(a) for HPK samples. It can be seen that the effect of the annealing is limited. There seems to be a decrease of initial acceptors in the gain layer with annealing on a time scale of tens of minutes, 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 Fig. 5.15(b). A much larger beneficial effect of annealing can be observed on the leakage current as shown in Fig. 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 volume generation current increases linearly with fluence. However, for LGADs the situation is more complex due to the gain and its fluence evolution. 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 currents for $1.3\,\mathrm{mm}^2 \times 1.3\,\mathrm{mm}^2$ HPK-3.2 single pads for the different fluences as a function of the bias voltage shown in Fig. 5.16(a) exhibit large increases for increased bias, partially due to the increased gain. The ALTIROC maximum acceptable current is 5 µA (dotted line in Fig. 5.16(a)). HPK-3.1 satisfies this requirement up to the highest fluence and voltage. From this the power density (power/area) can be derived. The power can be minimised by operating the sensors at as low temperature and bias voltage as possible. For the assumed operating temperature (−30 °C), Fig. 5.16(b) shows the measured power density of HPK-3.1 as a function of fluence for V_{op}. It can be seen that there is typically one order of magnitude difference between these two voltage points. For the 2.5 fC working point, the power density reaches maximally $30\,\mathrm{mW/cm}^2$, however, this working point can only be reached up to $1.5 \times 10^{15}\,\mathrm{n_{eq}\,cm^{-2}}$. For other types like HPK-3.2 and FBK-UFSD3-C with a higher fluence reach, evaluations are still ongoing. The final power dissipation in HGTD will depend on the sensor type choice as well as the operational scenario as detailed in Sec. 5.6.



(a) (b)

Figure 5.16: (a) Leakage current for single pads at −30 °C as a function of bias voltage for HPK-3.2 irradiated with 1 MeV neutrons (solid lines) and 70 MeV protons (dashed lines). The horizontal line represents the ALTIROC maximum acceptable current of 5 µA. (b) Power density as a function of fluence at the operation bias voltage measured V_{op} at −30 °C [39, 40]. The horizontal line represents the maximum acceptable power of 100 mW/cm^2.

5.6 Operational aspects and bias voltage evolution in HGTD

As shown in Sec. 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, allowing to perform the necessary adjustments of the bias voltage.

For a first scenario, it is assumed that the detector is operated at operating bias voltage V_{op} (see Fig. 5.7). This would give the best possible performance (charge and time resolution), however, also implies the highest risk and power dissipation (see Fig. 5.16(b)). In the future, more complex and realistic scenarios will be developed that will be a trade-off between required performance on the one hand and risk and power constraints on the other hand (keeping the sensor power at an average level of 30–60 mW/cm^2 as mentioned in Tab. 11.1).

The expected dependence of the fluence on the radius (Fig. 2.13) and the required bias voltage V_{op} for the increasing fluence permits a prediction of the bias-voltage distribution along the length of a readout row containing individual modules mounted on support plates (see Chap. 11). This is shown in Fig. 5.17 for different integrated luminosities for HPK-3.1 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 Sec. 5.5.3). Scenarios for other types with improved radiation hardness like HPK-3.2 and FBK-UFSD3-C will be evaluated in the future.

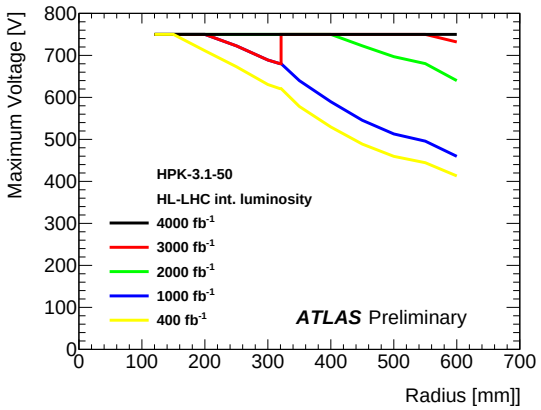


Figure 5.17: Required bias voltage as a function of position along the longest readout row for different integrated luminosities for HPK-3.1 sensors based on V_{max}. The sudden change for the 3000 fb^{-1} line corresponds to the replacement of the inner ring. (TODO: redo completely with 3-rings, use Vop)

5.7 Summary of status quo sensor design

Through the R&D program of the last few years, which involved three 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 "bias working points" of Fig. 5.9.

- Thickness of the high resistivity bulk
  50 µm: highest fluence reach, largest head room.

- 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.

- Replacing Boron, the dopant in the gain layer, with Gallium
  No Ga: advantage not clearly established (although for different doping profiles) after neutron irradiations. Indications of small improvement for 23 GeV protons irradiation.

- Adding Carbon to the dopant in the gain layer
  C implantation is a promising candidate that shows improved radiation hardness up to at least 3×10^{15} n_{eq} cm^{-2}, but is not yet fully understood: The performance at higher fluences needs to be established. Also 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)
  70 µm: lowest distance measured, leading to a fill-factor of 90%. With irradiation fill factor may improve due to increased operating voltage and relatively larger multiplication at the edges of the pads.

- Slim edge distance
  300 µm: Studies showed 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 2.5×10^{15} n_{eq} cm^{-2}. In general, more studies after high energy charged hadron irradiation are needed.

5.8 Roadmap for future sensor productions and activities

During the last year the HGTD collaboration have profited from a open and fruitful collaboration with CMS. Our ability to exchange ideas, designs and results, share beam test and financing of common production runs was a major reason for our rapid progress. Hopefully this collaboration will continue in the same way.

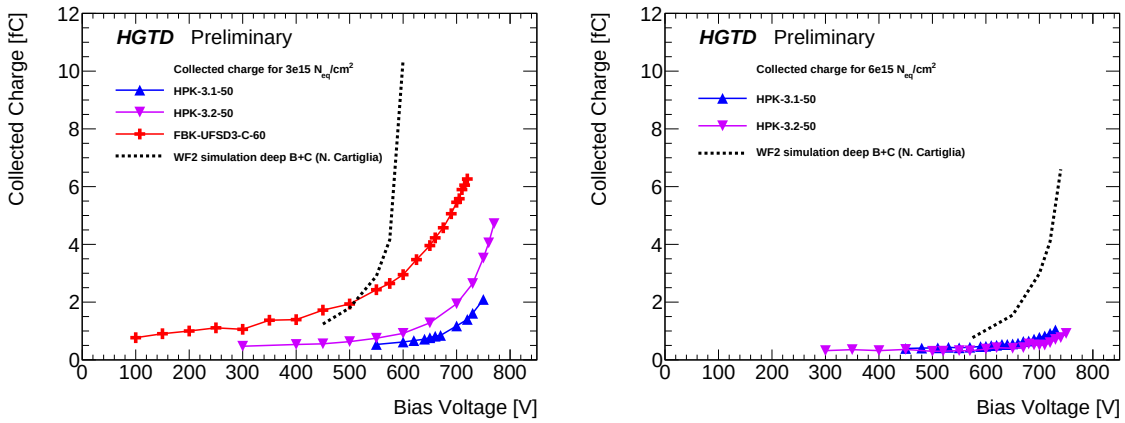
Two more prototype runs are anticipated, one in 2019 to implement our idea to extend substantially the fluence reach and to optimize several geometrical layout issues. In 2020, the final prototype will be produced and tested. A market survey will be conducted in 2020, based on the understanding of the design issues solved in 2019. Then a pre-production run will follow in 2021.

During the next prototype runs we will address sensor issues which we discovered during the recent tests by working with the three manufacturers. One issue is the need to overcome the loss in gain due to acceptor removal. Another is the need to increase the bias voltage head room. Another one is the the power consumption. All three issues need to be addressed by optimizing the presently available LGAD technology. An example is shown in the following, which leads to reduction in required bias voltage (and hence power) while improving the gain.

In Sec. 5.7 is mentioned the fact that the radiation hardness of the sensors available at the moment has not been established up to the HGTD baseline target fluence of $2.5 \times 10^{15}\,\mathrm{n_{eq}\,cm^{-2}}$ assuming one replacement of the inner ring at half lifetime (it should be noted that scenarios to mitigate the fluence requirements were developed as discussed in Sec. 5.8). This can be seen in Fig. 5.18, where the collected charge for high fluences is shown for the strongest candidates. For the fluence of $3 \times 10^{15}\,\mathrm{n_{eq}\,cm^{-2}}$ HPK-3.1 sensor does not reach the required level of 2.5 fC at the highest bias voltages reached, and for $6 \times 10^{15}\,\mathrm{n_{eq}\,cm^{-2}}$ the collected charge is even lower. The FBK-UFSD3-C sensor with Carbon implantation shows better performance at $3 \times 10^{15}\,\mathrm{n_{eq}\,cm^{-2}}$, the testing up to $6 \times 10^{15}\,\mathrm{n_{eq}\,cm^{-2}}$ will be done in the near future. Also shown is the collected charge from the HPK-3.2 sensor, which has a higher doping and deeper implant than the HPK-3.1 sensor (Tab. 5.3); nevertheless, at $6 \times 10^{15}\,\mathrm{n_{eq}\,cm^{-2}}$ the goal of 2.5 fC is still not reached.

The data of the collected charge for both sensors have been used to tune WeightField 2 so that the simulations match the measured data. Then the collected charge can be simulated for a sensor which combines both the deep implant of HPK-3.2 with the carbon implant of FBK-UFSD3-C. The collected charge of the prototype is shown for $3 \times 10^{15}\,\mathrm{n_{eq}\,cm^{-2}}$ in Fig. 5.18(a) and for $6 \times 10^{15}\,\mathrm{n_{eq}\,cm^{-2}}$ in Fig. 5.18(b). Two important characteristics for such sensors are evident: much lower operating voltage and much higher collected charge when compared with the sensors without carbon or with shallow implant.

Further topics of future R&D:



(a) (b)

Figure 5.18: Collected charge vs bias voltage for sensors irradiated to $3 \times 10^{15}\,\mathrm{n_{eq}\ cm^{-2}}$ (a) and $6 \times 10^{15}\,\mathrm{n_{eq}\ cm^{-2}}$ (b), respectively. In the plots are measured data of the existing prototypes and the simulated prospect of the proposed sensors combining deep implantation of the Boron gain layer with carbon implantation.

- Reduce the inactive pad distance.
- Produce first full-size sensors: 30×15 ($4 \times 2\,\mathrm{cm^2}$).
- Establish the robustness of LGADs under stressful operating conditions.
- Improve breakdown between guard ring and pad area.

New groups have joined the HGTD sensor R&D effort in the last year and will participate in large scale testing both for labs measurements and at beam tests.

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 Sec. 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 Sec. 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 time-of-arrival 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 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 [45]. 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 Sec. 6.1. The ASIC architecture is described in Sec. 6.2, first going through the single-channel architecture and then the entire ASIC. Sec. 6.3 describes in detail the design of the single-channel readout electronics, followed by the description of the ASIC common digital part in Sec. 6.4. The radiation tolerance is described in Sec. 6.5 and the power distribution in Sec. 6.6 The performance results obtained so far in testbench and testbeam are described in Sec. 6.7. The description of the monitoring can be found in Sec. 6.8. Lastly, a brief account is given on the future steps towards the completion of the design and testing of the ASIC in Sec. 6.9.

6.1 General requirements

This section presents a brief description of the requirements of the front-end readout electronics. The requirements of the ASIC can be divided in two types. On one side the considerations regarding the operational environment of the ASIC, its powering and electrical connections. These requirements are summarized in Tab. 6.1. The second group is concerning the ASIC performance, driven by the targeted time resolution. A summary of these requirements is presented in Tab. 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. The expected radiation levels have been presented in Sec. 2.4, considering a 2.25 safety factor for the electronics. Thus, the maximal TID is 2.0 MGy (at $r = 120$ mm) and decreases with radius. At $r = 320 but out of date$ (the edge of the region that will be replaced by half the HL-LHC lifetime) it reaches a value of of 2.0 MGy.

- 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.

- 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 energy deposit which limits the time resolution to 25 ps at large sensor gain. Such charge is equivalent to deposit of a MIP in a 50 µm thick LGAD with a gain of 20. A detector capacitance of around 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. **TO BE UPDATED**

- 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 should be so that an efficiency above 95% is achieved 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 TOA and TOT information are transferred to the data acquisition system only upon L0/L1 trigger reception with latency up to 35 µs [46], therefore necessitating a large size memory. The trigger rate depends on the final scheme adopted. It will be 1 MHz for a L0 trigger, or 0.8 MHz for a L1 trigger in a L0/L1 scheme with a L0 at 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 bandwidth required, the information of only a subset of the ASICs is used. The current proposal is to use the sensors located at $320 but out of date < r < 640$ mm, or equivalently $2.4 < |\eta| < 3.1 but out of date$!. 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. **TO BE UPDATED**

- 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 Sec. 11.2).

TID tolerance	Inner region: 2.0 MGy Outer region: 2.0 MGy
Pad size	$1.3 \times 1.3\,\text{mm}^2$
Voltage	1.2 V
Power dissipation per area (per ASIC)	$300\,\text{mW cm}^{-2}$ (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
SEU probability	< 5%/hour

Table 6.1: Physical, power, environmental and electrical requirements.

Maximum leakage current	5 µA
Single pad noise (ENC)	**< 1500 e^- = 0.25 fC**
Cross-talk	**< 5%**
Threshold dispersion after tuning	10%
Maximum jitter	25 ps at 10 fC
TDC contribution	< 10 ps
Time walk contribution	< 10 ps
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. The values given for the noise, minimum threshold and jitter have been estimated considering a detector capacitance $C_d = 4$ pF.

6.1.1 Data transmission bandwidth requirements

The bandwidth 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 Fig. 9.3.

Each module consisting of two ALTIROC ASICs is connected via a flex cable to a Peripheral Electronics Board (PEB), described in Chap. 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 [47]). 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 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 data, the maximal number of hits per ASIC at $r > 320 but out of date$ should be considered **TO BE UPDATED**. This number does not exceed 30. With a 4-bit header in addition to the 12 bits of data for the counts in the larger and smaller window (see Sec. 6.2.1), a 640 Mbit s^{-1} e-link driver and lpGBT speed is needed.

6.2 ASIC architecture

With an area of 20 mm × 22 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 *end-of-column logic* (EOC) which performs the readout of the 15 columns and transfers the data to the trigger data and luminosity process 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 Chap. 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 Sec. 6.4.

The ASIC has been designed using TSMC [1] 130 nm technology. Simulations have been performed using the 130 nm TSMC kit provided by CERN. The TSMC 130 nm has been tested up to 400 Mrad and the effects at transistor level are known. The analog part of the ASIC (preamplifier, discriminator and TDC) has been designed to ensure its radiation hardness.

6.2.1 Channel architecture

A conceptual schematic for the single-channel readout is presented in Fig. 6.1. Each readout channel will consist of a preamplifier followed by a discriminator, both of which critical elements for the overall electronics time performance. A detailed characterization of the preamplifier is presented in Sec. 6.3. The time of the pulse will be determined using a discriminator that follows the preamplifier. 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 measured using a Time Over Threshold architecture (described in Sec. 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 Sec. 6.3.3.

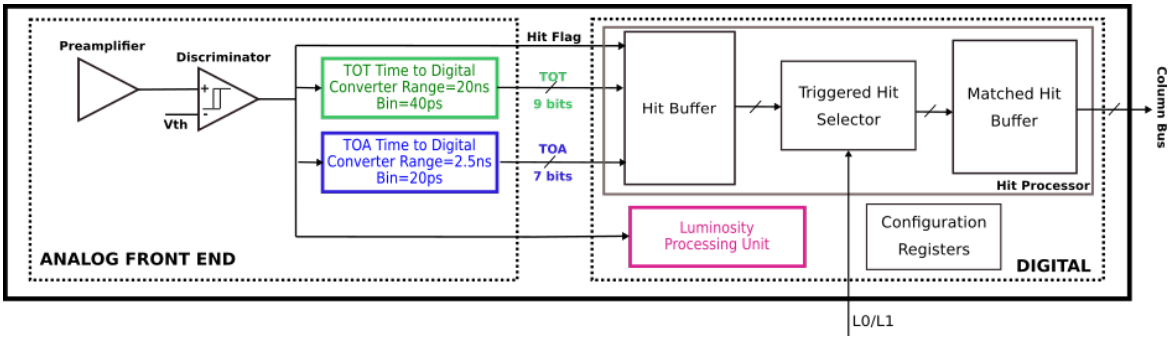


Figure 6.1: Schematic of the single-channel readout electronics. Two main blocks are identified, the analog and the digital part. The input pulse from the sensor enters the preamplifier on the left. The TOA and TOT data are read out by the column bus on the right.

[1] TSMC stands for Taiwan Semiconductor Manufacturing Company. The technology has been qualified up to 4 MGy [48, 49].

The output of the analog read-out is processed by the digital stage providing two different measurements: timing 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 there until it is retrieved for transmission to the common digital part. These local memories are further described in Sec. 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 Fig. 6.2. A first 3.125 ns wide window ($S1$) is centred at the expected arrival time of the particles from the collisions. The second window ($S2$) is adjustable in length and position in steps of 3.125 ns, and will count the number of particles arriving before and/or after those from the collisions. This side-band will provide valuable information about the background, as described in Sec. 10.3.3. The windows are aligned to the expected arrival time of the particle, with their length and alignment adjustable via configuration parameters. 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 Sec. 6.4.3). These parameters will be optimised based on operational experience.

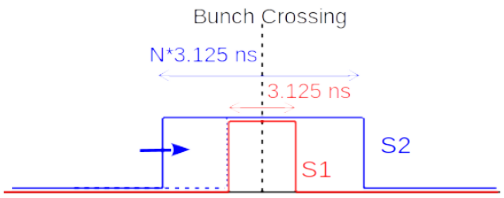


Figure 6.2: Illustration of the time windows used for counting hits for the luminosity data. The smaller window (S1, in red) is 3.125 ns wide and is centred at the bunch crossing time. The width and relative location of the larger window (S2, in blue) can be set in steps of 3.125 ns through the control parameters.

This 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 within them. The way this is done is further described in Sec. 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 Sec. 6.4.

Lastly, there are four 8-bits configuration registers per channel. They are read/written by the slow control unit through a Wishbone bus. The configuration registers allow to configure

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

Fig. 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 schema of a single channel, as presented in Fig. 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), which consist of 8 bits on every bunch crossing. The 320 MHz clock is extracted from the fast commands and clock received from the lpGBT, and from the 40 MHz clock is generated. Based on this, a phase-locked loop (PLL) generates all the different clocks needed to operate the ASIC, namely 80 MHz, 640 MHz, and 1.28 GHz. These clocks will be centred with an accuracy of $\sim$100 ps using a phase shifter, further described in Sec. 6.4.3.

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 and are received in series at 320 Mbit/s, so a command 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 stablish the communication between the lpGBT and the ASIC, but also as a source clock from which all the internal clocks needed to operate ALTIROC2 will be generated. The 320 MHz clock from the lpGBT is divided by 8 and the passed to a Phase-Locked Loop (PLL) produces clocks of 40 MHz, 80 MHz, and 640 MHz These clocks will be centred with an accuracy of 100 ps using a phase shifter. Further details about the clock generation and distribution are given in Sec. 6.4.3 and Sec. 6.4.4.

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 Sec. 6.4.1.

The TDPU performs two tasks in parallel, one is to process incoming triggers and the other to readout data associated to a trigger 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 are transmitted immediately to all the pixels. Then each pixel checks if it has data associated to that trigger event. If they have,

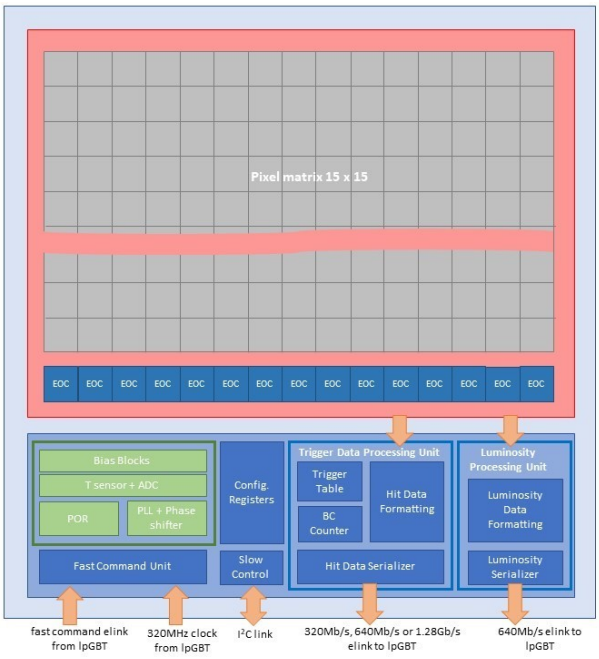


Figure 6.3: Schematic of the full HGTD ASIC. The top part represents the 15 × 15 channel matrix, while the bottom part shows the ASIC common digital part.

these transferred, together with the corresponding TrigID to a secondary in-pixel buffer. They remain there until 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 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 and 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. Sec. 9.2.1.

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 package is transferred to the luminosity serializer where data is encoded 6bto8b, 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} s-1 e-link. The measurement and data transmission can be enabled/disabled by accessing one of the configuration registers. As explained previously in Sec. 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 (under consideration) 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 SEU events 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.

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 Sec. 6.3.1, while the discriminator and time-walk correction are presented in Sec. 6.3.2. Concerning the digital blocks, the working principle of the time-to-digital converters is presented in Sec. 6.3.3, while the designs of the local memory and the luminosity processing unit are presented in Sec. 6.3.5 and Sec. 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 [37, 42] However, in silicon sensors such as LGADs, the preamplifier speed is not so crucial, both 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 easily 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:

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

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:

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

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 LGAD in HGTD, a $C_d = 4$ pF has been estimated when fully depleted (see Fig. 5.4(b)); typical $t_d \sim 0.6$ 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 Fig. 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. To 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 gain is small when increasing this current beyond 600 µA. The rise time of the preamplifier can be modified, and so evaluate its impact on the

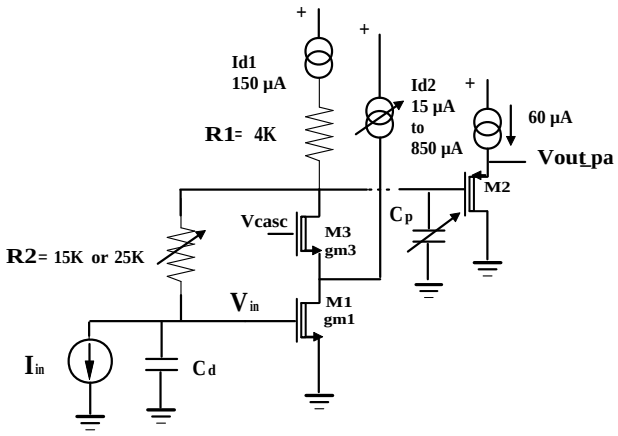


Figure 6.4: Schematic for the preamplifier implemented in the latest ASIC design, ALTIROC1.

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 of the preamplifier output, it depends on the input impedance of the preamp (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 around 1.6 kΩ. The value of the resistor R_2 can be either 15 kΩ or 25 kΩ. It can also absorb the sensor leakage current, estimated to be below 5 µA 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. This can be done using the 7-bit DAC threshold correction that is integrated for each channel allowing a correction within ±50 mV.

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, which corresponds to an amplification gain of 20 in the LGAD. 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 [50] and presented in Fig. 4.4(b) were used as input, and the obtained preamplifier pulses are presented in Fig. 6.5.

6.3.2 Discriminator and time walk correction

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 around a high speed

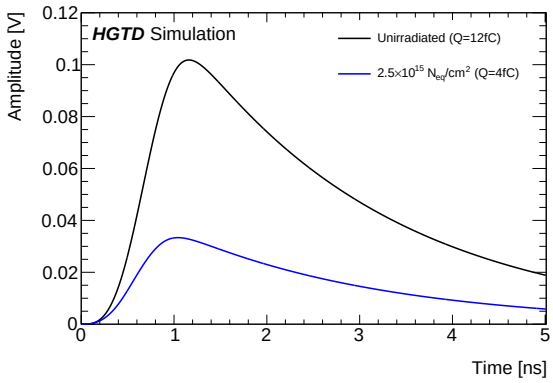


Figure 6.5: Simulation of the preamplifier output using as input the simulated LGAD signals presented in Fig. 4.4(b) for a non-irradiated sensor and after irradiation.

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

6.3.3 TDC

The target timing resolution (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 around 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 Fig. 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 (output of the discriminator) enters the "slow" delay line while the STOP signal (end of measurement window) enters the 'fast' delay line. Although initially the START signal is ahead of the STOP one, each delay-cell stage brings 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

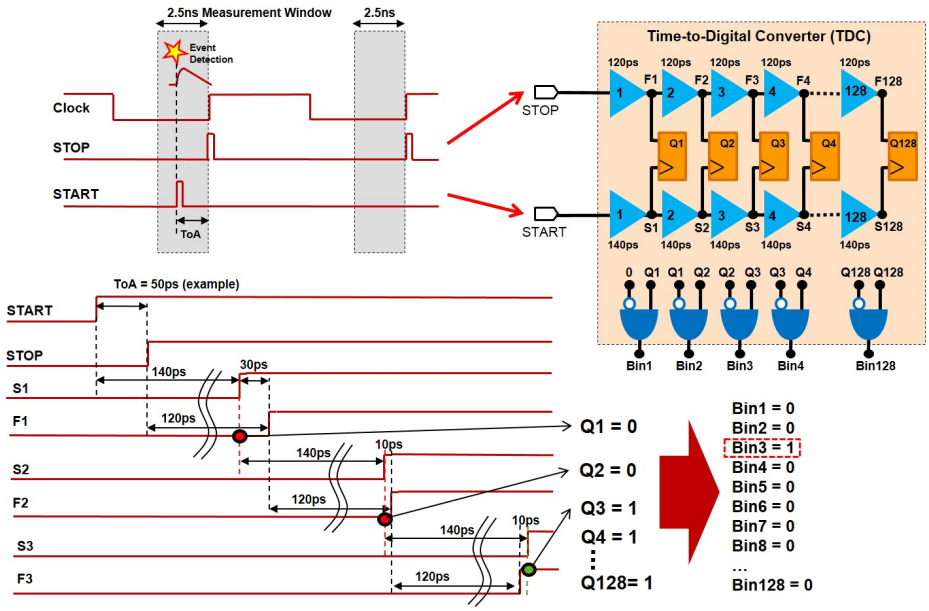


Figure 6.6: Graphic representation of the working principle of the TDC. The drawing on the top left shows how the START and STOP signals are generated, the first with the discriminator output upon event detection, the second corresponding to the next clock edge. The gray area indicates the 2.5 ns detection 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.

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 about 21 ns, finishing just in time to be able to accept and process the signal in the next bunch crossing.

The TOT TDC employs an additional coarse delay line for extending the range to 20 ns, while the Vernier delay line (identical to the one used in TOA TDC) provides the high resolution of 40 ps.

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 Command pulser for calibration

An internal pulser, common to all channels, is integrated to mimic input charges. This pulser can be used to intercalibrate the gain of each channel by performing trigger efficiency measurements for various input charges. This intercalibration allows to align the threshold of each channel using a correction threshold DAC integrated for each channel. The pulser consists of a programmable DC current (tunable with an internal 6-bit DAC) that flows continuously in a 50 kΩ resistor until it is interrupted by a command pulse that shorts the resistor to ground (see Fig. 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}). The simulated input charge (Q_{inj}) is equal to $C_{test} \times V_{step}$. The dynamic range goes from 0 to 250 mV or 0 fC up to ∼50 fC (LSB = 0.76 fC). This pulser will also be used to calibrate the absolute value of the phase.The command pulse that is encoded in the fast command elink is therefore distributed as a clock tree inside the ASIC.

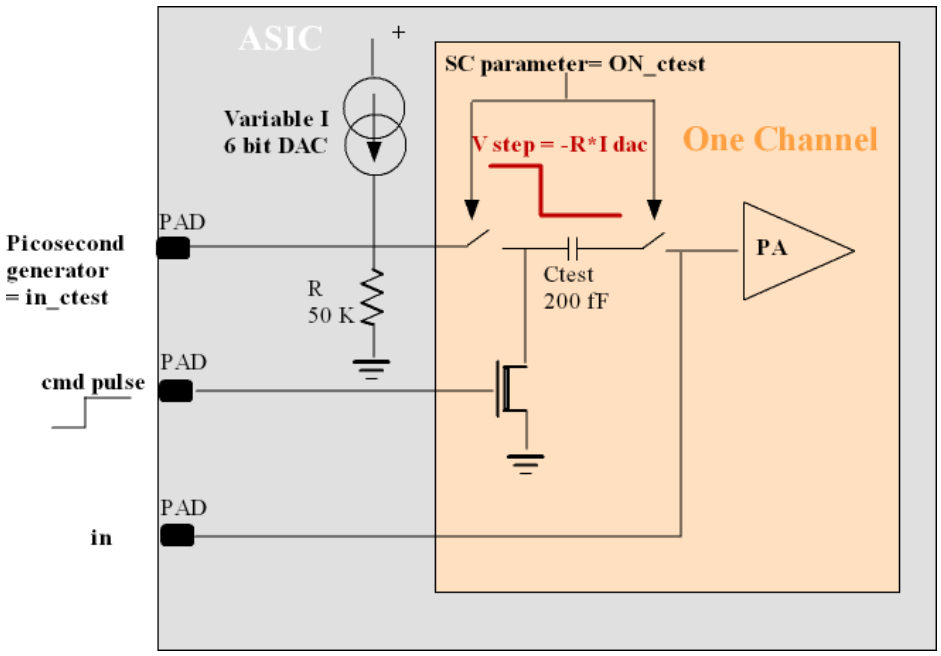


Figure 6.7: Pulser principle that shows the common 6-bit current DAC used to set the input charge as well as the pixel Ctest capacitor.

6.3.5 Hit processor

Each channel electronics 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 Fig. 6.1. The hit processing unit, or hit processor, temporarily store the data related to a hit and select 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. This hit flag bit indicates 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. The pointer goes from 0 to 1399 during an L0/L1 scenario (latency of 35 µs) and then goes back to position 0. For the L0 scenario, in which the trigger 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 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 TOT and TOT fields in order to save power.

The hit buffer architecture is built around a two-port SRAM design. This configuration allows simultaneous Read/Write operations within the same clock period. 6 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 a L0 trigger signal at 4MHz, is evaluated to 1.55 mW (at 25°C, 1.2V). This power dissipation decreases down to 1.2 mW with a 2.5% occupancy obtained with a L1 trigger signal at 1MHz. 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 transfers the TOA and TOT information to the *matched hit buffer* if it finds a hit flag equal to one when it receives a trigger. It also tags each trigger with an identifier TrigID provided by the TDPU of the periphery, which is stored together with the TOA and TOT in the matched hit buffer. This allows to know to which triggered event the data stored in the buffer is associated. This 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 that some hits inside one of the windows would be 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 Fig. 6.8. At the channel level, and 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 read out at each clock cycle and processed in the end-of-column logic.

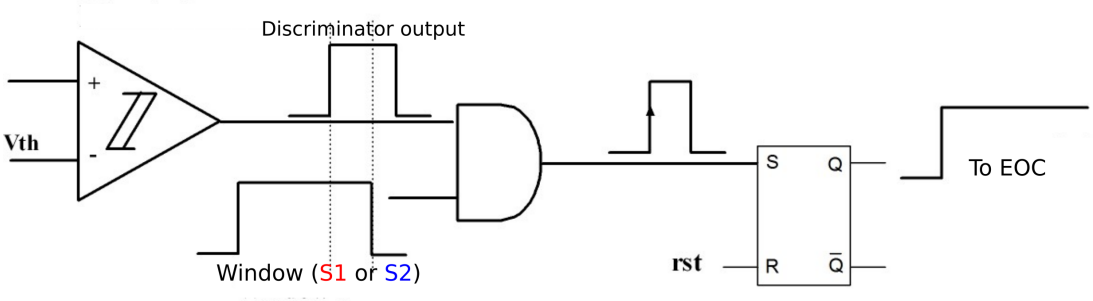


Figure 6.8: The signal of the discriminator is compared to the luminosity window (for each window) and a signal is transmitted to the end-of-column logic.

6.4 ASIC 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. The design of many of these components is on-going.

6.4.1 Matrix readout process

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 trigger identifier. 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 trigger data processing unit 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 trigger identifier (TrigID). If the FIFO is full, an error message is generated and transmitted to the TDAQ through an e-link. The TDPU unit also generates an internal trigger signal (trig) with a duration of one clock cycle. This is immediately transmitted to all matrix channels as well as the trigger identifier trigID. Fig. 6.9 shows a block diagram of the main signals involved between the TDPU and the EOC. Both, the trigger signal and the identifier are processed by the hit processor as described in Sec. 6.3.5.

The hit data formatting unit in the TDPU it is always checking if there is an entry in the trigger table. When it founds one, 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 trigger identifier by asserting the checkTrigID signal. The hit data processor checks if there is a matched hit with the same trigger identifier 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 to the hit data formatting block that is done by asserting the doneMatrix signal. In the second step, the hit data formatting block starts reading the FIFO of the end of the columns immediately

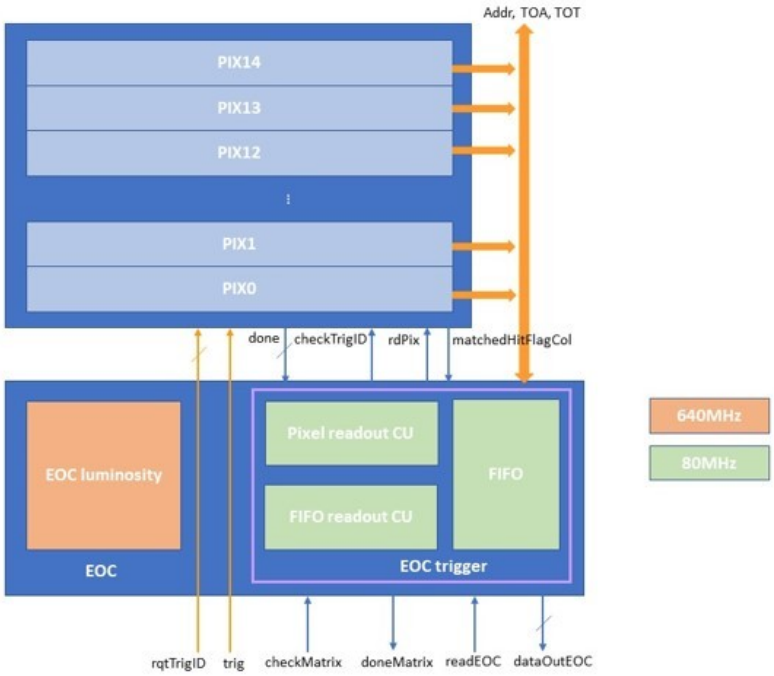


Figure 6.9: Block diagram of the main signals involved between the communication of the EOC with the pixels and the TDPU.

at the moment that the FIFO has data available asserting the readEOC signal. The hit data formatting does not wait that the EOCs have finished to retrieve from the pixels to start reading the FIFOs. 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.

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 Sec. 6.2.2 The windowing is performed in pixel. The 2 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 100ps. The windows are generated with a programmable FSM running at 640 MHz clock. 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 monitors asserts and deasserts the two window signals window1 and window2 as a function of the value of the counter and

of the 4-bits parameters minW1, minW2, maxW1, and maxW2 as shown in **??**. The 4-bits parameters indicate in which intervals window1 and window2 must be high or low. The width of window 1 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 was necessary. The 1.5265 ns time resolution of the window generator is not enough to center the position the windows respect to the beginning of the bunch crossing. In order to provide the required resolution, the phase of the 640MHz clock used by the EOC can be adjusted through the phase shifter. This 640 MHz clock is independent from the 640 MHz clock used in the serializers. More details are given in Sec. 6.4.3 The windowing process at pixel level is described in Sec. 6.3.6. Every pixel produces two measures per bunch crossing. The EOC sums the luminosity measures of the whole column per bunch crossing. Those measures are passed to the luminosity processing unit. This sums the measures of the columns. The luminosity of the window 1, S1, is subtracted from the luminosity of window 2, S2. The result S2-S1 and S1 are trunked to 4 and 6 bits respectively. Both values 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, which requires to lose the luminosity information of a bunch crossing.

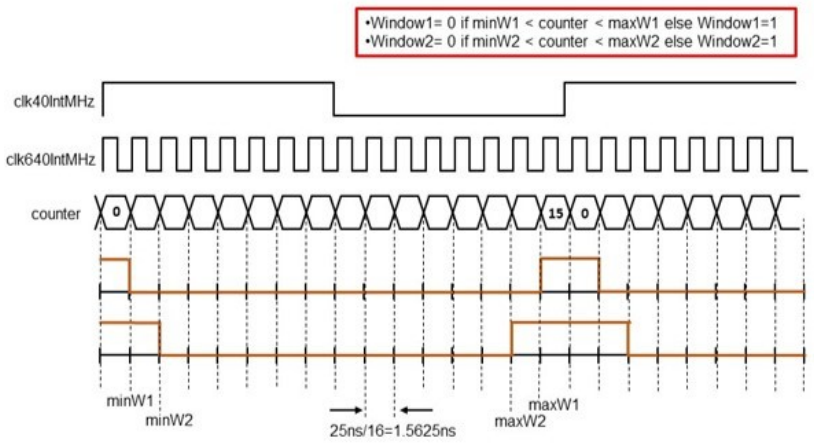


Figure 6.10: Block diagram of the eoc luminosity unit.

6.4.2 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 configure the channel registers. Each channel contains 4 configuration registers. 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/write by using an I^2C link while shift registered are used for the prtotype. The I^2C link in the ASIC is slave to the master in the lpGBT in the peripheral electronics, described in Sec. 9.2.1.

6.4.3 Clock generator and phase shifter

The clock generator and phase shifter are a function block located in the ASIC common digital part used for the adjustment of the clock phase for the clock system in the ASIC. Some examples are the 40 MHz used by the TDCs and the 640 MHz used to generate the luminosity windows. A schematic of its design is presented in Fig. 6.11. The clock generator provides two clocks of 40 MHz and 640 MHz to the phase shifter from one of two clock sources. The first source, clk40MHz, is a 40 MHz clock coming from the fast command unit. A PLL multiplies by 8 this clock. The second source is an external clock of 640 MHz, clk640MHzExt, that is divided by 8. The clock source is selected through the clockSel signal.

The phase shifter receives the clock CMOS signals of different frequencies (40MHz, the 640MHz is present in the current prototype for debugging) generated by the PLL and outputs them with the same frequency but with an adjusted phase. This module is required to provide a shift step smaller than 100 ps, an additional jitter below 5 ps on the 40MHz clock, and a power consumption around 10mW.

The design presented here is adapted from an lpGBT design in the 65 nm process. The core of the phase shifter is composed of two delay-locked loop (DLL) formed by delay line of 16 delay cells (not including the dummy cell at the end of the delay line). The DLLs are used to delay the input clock signal of 640 MHz generated by the clock generator and to provide two 640 MHz clocks two the ASIC. One of these clocks will be used to serialize data, clk640MHzInt, and the other to generate the time windows, clk640MHLumInt. More details are given in **??**.

For the 80 MHz and 40 MHz clocks, coarse phase adjustment circuits are needed. Their output is re-sampled by the clk640MHzInt, with the result that these two clock signals acquire the same time resolution. Their time resolution is therefore 1/16 of the 640MHz clock period, equal to 97.6 ps.

6.4.4 Clock distribution

Fig. 6.12 shows a block diagram of the different clock domains inside ALTIROC2. All the internal clocks are derived from a source clock of 320MHz coming from the lpGBT chip. The fast command unit divides this clock by 8 producing a 40MHz clock, clk40MHz. This is passed to the clock generator, described in Sec. 6.4.3., that generates a 640MHz clock with a PLL. Both clocks are connected to the phase shifter that generates 1 clock of 40MHz,

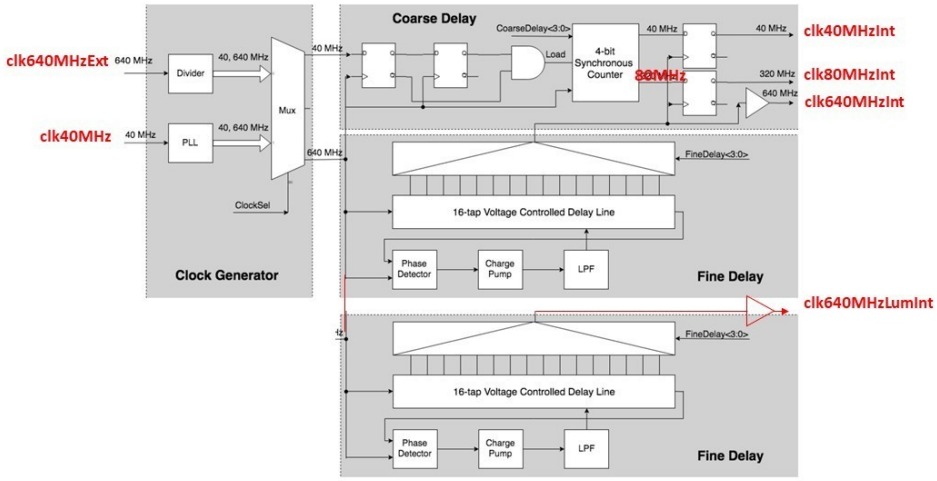


Figure 6.11: Schematic diagram of the clock generator and phase shifter.

clk40MHzInt, 1 clock of 80MHz, clk80MHzInt, and 2 clocks of 640MHz named clk640MHzInt and clk640MHzLumInt. Clk40MHzInt, clk80MHzInt, and clk640MHzInt are aligned in phase but shifted from the clk40MHz as described in Sec. 6.4.3. Clk640MHzLumInt is used to generate the time windows W1 and W2. Its phase can be adjusted with a resolution of 100ps with the phase shifter in order to fine tuning the position of both windows from the bunch crossing as described in Sec. 6.4.1.

Most of the digital electronics run at 40MHz, these include the I^2C and the configuration registers. The 80MHz clock, clk80MHzInt, is used to readout the timing data from the pixel matrix and to pack the data into frames in the trigger data processing unit. Finally, the 640MHz clock clk640MHzInt is used to serialize data.

6.5 Radiation tolerance

Two radiation effects must be taken into account: the Total Ionizing Dose (TID) that may degrade the timing performance and the Single Event Effects which may corrupt the configuration registers and the time data. The worst expected TID and fluency are respectively 210 $Mrad$ 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 TSMC130 nm technology which has been tested up to 400 $Mrad$, which is two times above the requirement. Nevertheless known

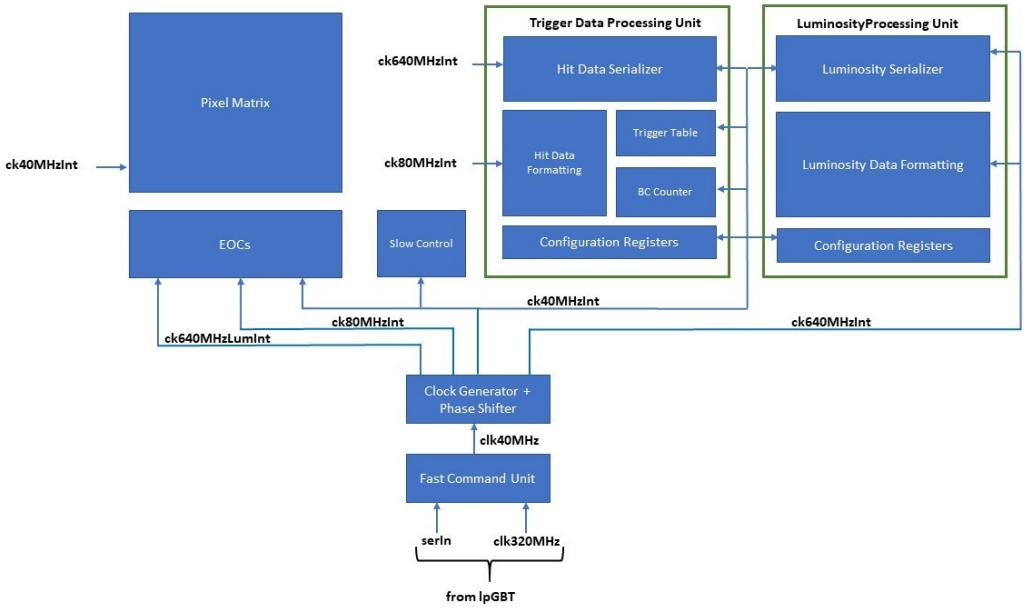


Figure 6.12: Schematic of the clock distribution.

strategies have been used in the ASIC design to mitigate the radiation effects. TID radiations degrade the performance of MOSFET by increasing their threshold and generating leakage currents. To counter these effects, bias currents of analog blocks are set to quite large values ($> 20\ \mu A$) compared to the expected leakage currents and low VT 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 (ALTIROC2). Simulations of upsets using CERN tools will be performed to fully evaluate the effect of SEUs on the chip functioning.

6.6 ASIC Power distribution and grounding

To preserve the signal integrity and the jitter performance under an important digital activity, the power distribution must be done carefully at the ASIC level. Each analog block (PA, discri, TDC) is in a deep Nwell that is powered and grounded with its own power line and ground line. 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 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 Nwell or directly on the substrate. The connection between all the digital grounds and vss will be done at the flex level. Tests at system level are necessary 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 (vdda_block) are connected together at the flex level to a common Vdda and all the digital power lines of the digital blocks (vddd_block) are connected to a common vddd.

6.7 ASIC prototype measurements

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 [45].

In this section, the results concerning the second prototype ALTIROC1 are presented. This second version, ALTIROC1, consists of a 5×5 pad matrix instead of a 2×2, in which the digital components have been added to the single-channel readout. Two iterations of ALTIROC1 have been produced. The second one ALTIROC1v2 corrects issues found in the TDC and only results from this iteration are presented here. Among the 25 channels, only 15 channels corresponding to 3 columns have the readout as described in Sec. 6.2.1 with voltage preamplifier. The 2 other colums are equipped with trans-impedance preamplifiers and their performance are not described in this document.

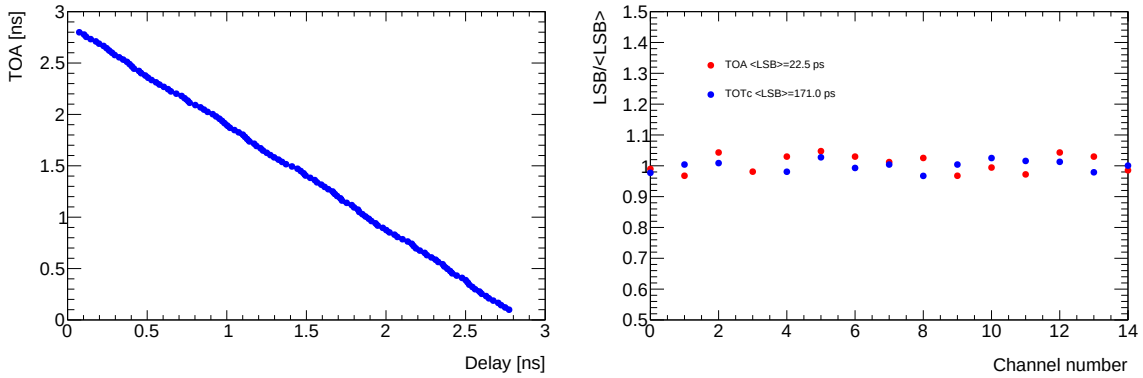
Sec. 6.7.1 describes the testbench measurements which were performed with and without a sensor bump-bonded to it**ADD MORE INFORMATION OR CITE THE MODULE ASSEMBLY CHAPTER.**. In the case where no sensor is bump-bonded, on channel 4 of each column, a capacitor can be connected through a programmable switch 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. The testbench measurements are performed thanks to a C_{test} capacitor of 200 fF integrated in ALTIROC1, and selectable by slow control and a calibration pulser described in Sec. 6.3.4 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 12. In a similar way, a digital probe allows to see the output of the discriminator, before going into the TDC.

Two testbeam campaigns have been carried out during the year 2019 at DESY, in which data was collected with ALTIROC1v2, bump-bonded to a non-irradiated LGAD sensor. The main results are presented in Sec. 6.7.2.

Irradiation tests were also performed at CERN using X rays up to 340Mrad. The results are presented in Sec. 6.7.3.

6.7.1 Test bench performance

The first step towards the evaluation of the full single-channel readout is the calibration of the TDC counts since the knowledge of the value of the LSB (Least Significant Bit) is fundamental in order to obtain the real values of the TOA and TOT. This is achieved by sending a delayed square pulse (called external trigger) directly to the TDC inputs bypassing the preamplifier and the discriminator, and thus measuring the TOA as a function of the delay as displayed in Fig. 6.13(a). The measured TOA TDC quantization step is found to be around 22 ps, close to 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 Fig. 6.13(b) and is better than 5%. The external trigger has a variable width and can also 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 Fig. 6.13(b). **NEED TO ADD A COMMENT ON THE MISSING TOT point**.

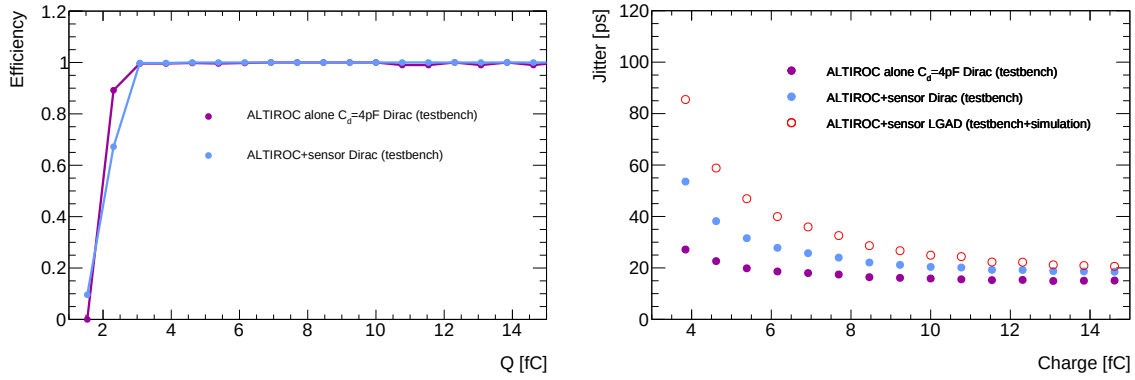


(a) (b)

Figure 6.13: 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(b). All measurements are performed with an external trigger.

The preamplifier jitter σ_{jitter} depends on the preamplifier rise time, which depends on the drain current that goes into it. All the results below have been obtained with I_d =570 µA

TO BE CHECKED where the transistor enters in the strong inversion region and the gain increases only with the square root of I_d and so the S/N doesn't increase significantly. Fig. 6.14(a) shows the efficiency as a function of the input charge for an ASIC alone with $C_d = 4\,\text{pF}$ in order to mimick the detector capacitance and with an ASIC bump-bonded to a sensor. In both cases, full efficiency is achieved for charge greater than 3 fC.



(a) Efficiency vs Q_{inj}

(b) Jitter vs Q_{inj}.

Figure 6.14: 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 a LGAD input signal estimated from the calibration data and the simulation.

Fig. 6.14(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 pulse command and clock jitter.Even without subtracting this constant term, the jitter is smaller than 30 ps for $Q_{\text{inj}} > 6$ fC. **NEED TO ADD AN EXPLANTION OF THE DIFF BETWEEN THE BLUE AND PURPLE POINTS.** However, the performance obtained with the calibration signal can't be transposed to a 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 a LGAD signal. Therefore, the jitter becomes smaller than 30 ps only for $Q_{\text{inj}} > 8$ fC as shown on Fig. 6.14(b) and the jitter is 85 ps at 4 fC.

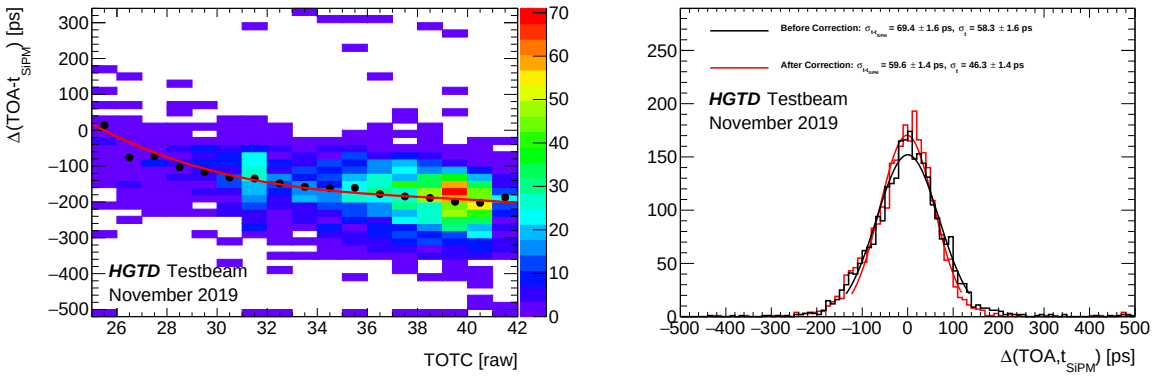
TO BE UPDATED: The power consumption of the ASIC has been estimated through both preliminary measurements and simulations for a 10% occupancy. Two operation modes can be distinguished: physics runs and calibration runs. In the latter, up to a 10% occupancy will be allowed. At the single channel level, the preamplifier and discriminator give a power consumption of 1.57 mW, considering a drain current for the preamplifier of 1 mA. For each time-to-digital converter, 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 4.67 mW per channel; considering that the current input to the preamplifier will be more likely around 600 µA, this gives a margin of 0.35 mW. 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.

6.7.2 Test beam performance

An ALTIROC1v2 ASIC has been bump bonded to LGAD sensor arrays (HPK 3.2), with 1.3 mm $\times$ 1.3 mm pads and exposed in electron beam tests at DESY in the fall of 2019. The LGADs have been operated with a bias voltage of 230 V, resulting in a MIP charge deposit of about 18 fC. For an accurate timing reference, a fast Cherenkov-light emitting Quartz bar of 6$\times$6 mm^2 area transverse to the beam and 10 mm length along the beam, coupled to a Silicon Photomultiplier (SiPM). The time resolution of this device was measured to be 37.6$\pm$0.7ps

The Fig. 6.15(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. This problem is attributed to a coupling with the bias voltage distribution leading to sharp steps in the TOT versus charge distributions[2]. The Fig. 6.15(b) shows the time difference between LGAD+ALTIROC and the reference time from the Quartz+SiPM system before and after time walk correction. The distributions are Gaussian without any tails. After substracting the contribution for the Quartz+SiPM system, the time resolution decreases from 58.3$\pm$1.6 ps to 46.3$\pm$1.4 ps. The time resolution is the quadratic sum of the intrinsic time resolution of the LGAD (about 25 ps) and the electronics jitter that is therefore deduced to be about 39 ps. **NEED TO ADD AN EXPLANATION WHY IT IS WORTH THAN THE EXPECTATION!!!!**.



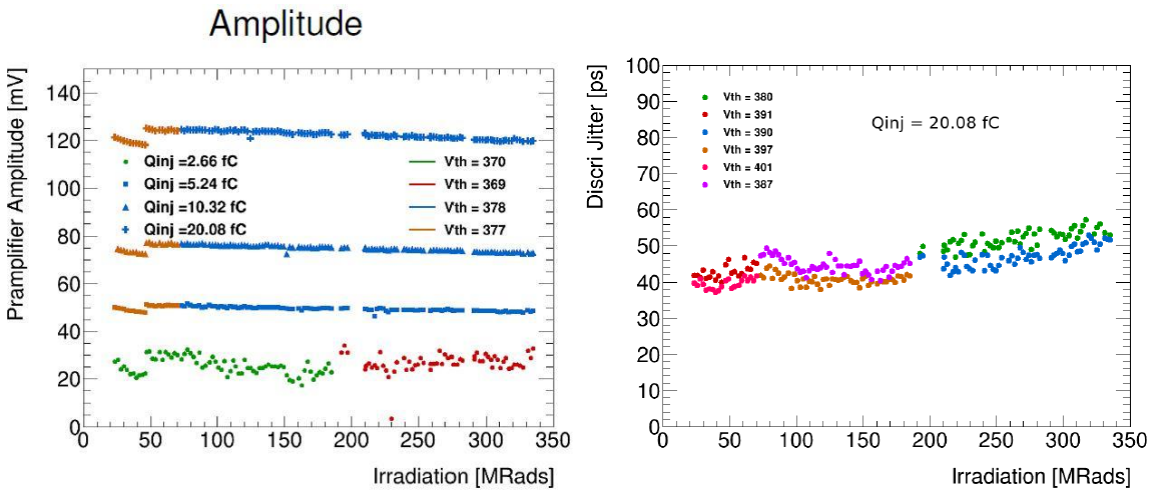
(a) (b)

Figure 6.15: (a): TOA variation as a function of the TOT. (b): Time difference between LGAD+ALTIROC and the Quartz+SiPM system before and after time walk correction.

[2] When the ASIC is not bump bonded to a sensor, this effect is not observed.

6.7.3 Irradiation tests

Altiroc1 was irradiated at CERN with Xrays up to 340 Mrad. Tests were focused on the preamplifier and the discriminator. A small decrease of the preamplifier amplitude was observed (**??**) as well as a small increase of the discriminator jitter ((**??**)). DC voltages such as the bandgap output, the 10-bit DAC used to set a common discriminator threshold as well as the 7-bit DAC used for individual tuning of the threshold were also followed during the irradiation and showed variations smaller than 20 mV (around 800 mV) between 20 Mrad and 340 Mrad.



(a) (b)

Figure 6.16: Preamplifier amplitude (Left) and discriminator jitter during irradiation tests.

6.8 Monitoring

6.8.1 Temperature monitoring

An additional requirement of the ASIC is to be capable of 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 Sec. 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 operating 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 0.2 °C (9-bit resolution); the absolute value of the temperature is not relevant.

The temperature sensor inside ALTIROC is based on a resistor which is sensitive to variation of temperature, a constant current flowing through this device producing a voltage proportional to the temperature. This current will be delivered by the current source present at the ADC input of the lpGBT circuit. Four different types of resistor proposed by the TSMC technology) have been evaluated for their ability to perform temperature measurement on an irradiated environment. As they all present similar behaviour, only performance of the N-diffusion resistor version is reported in Table 6.3. In important point is that the resolution can be doubled using a current value of 200 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 350 Mrad
Shift of temperature with radiation	−0.3 °C at TID of 200 Mrad
	−0.6 °C at TID of 350 Mrad
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. The first need if for measuring 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 Fig. 6.17). The VDDA and VDDD voltages are sensed through dedicated wires on the flex and digitized by the ADC of the lpGBT circuit on the peripheral board.

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

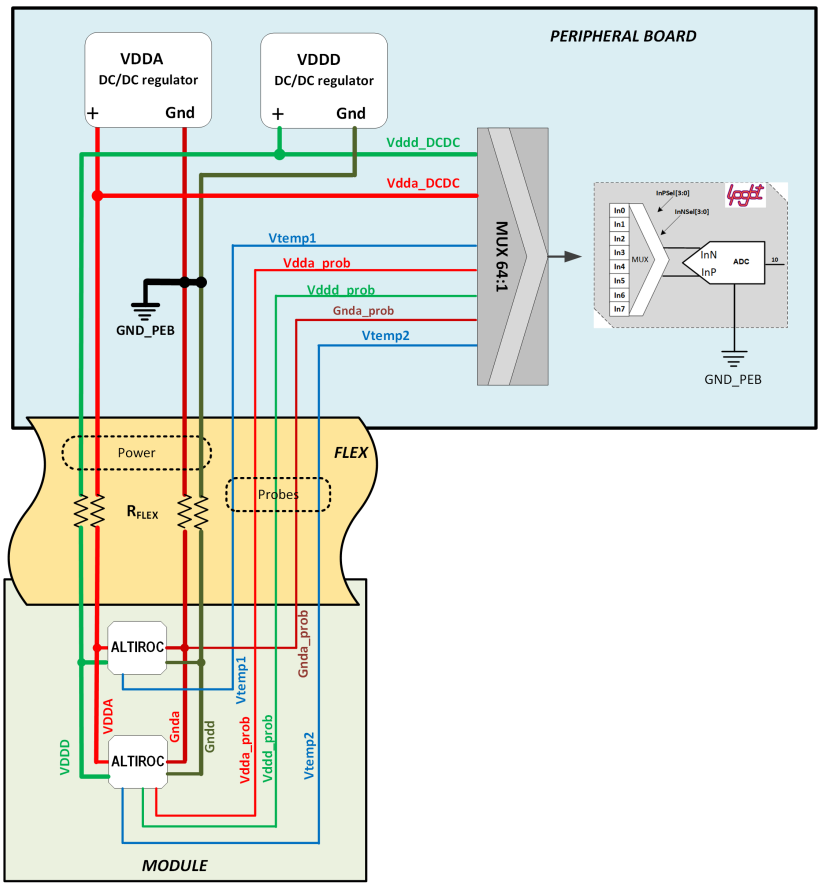


Figure 6.17: Complete schematic view of the voltage monitoring of a module using the ADC of the lpGBT circuit.

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 Fig. 6.17. 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, 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 Fig. 6.18. 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. 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 the modules, which is larger than the maximum number of modules expected per peripheral board. A full custom 64-to-1 multiplexing circuit is under development with a radiation tolerance suitable with its implementation on the peripheral board.

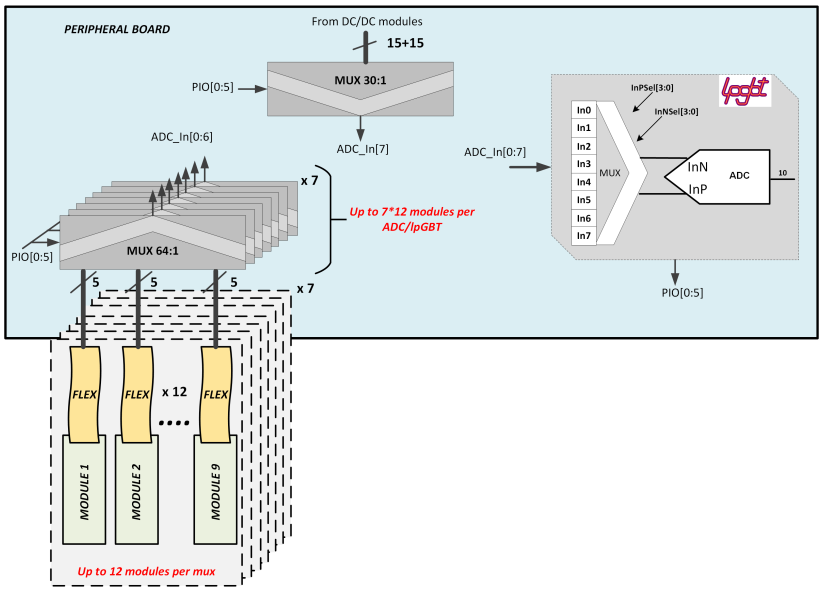


Figure 6.18: Interfacing of modules with a peripheral board for the monitoring.

6.9 Roadmap towards production

TO BE UPDATED

As explained before, two iterations of the ALTIROC have already been submitted and received. The strategy so far has been to validate with a first prototype the very front-end readout electronics, and then secondly the full single-channel readout with ALTIROC1. Concerning the latter, a new iteration has been submitted at the end of February 2019 in which the implemented changes are expected to fix the issues found in the TDC. The main modifications, besides other minor ones, have been:

- Replacing all the 1.2 V MOS transistors used to filter the bias voltage of the ASIC by 2.5 V thick oxide transistors which have a negligible gate leakage current
- Increasing the size of the ASIC to make the debug pad accessible after bump bonding to the sensor
- Modification of the TDC layout as described in Sec. 6.7.1.

This prototype of ALTIROC has been designed so as to enable irradiation tests, which will be performed during 2019.

The next prototype, ALTIROC2, will integrate the 225 channels and all the digital blocks currently existing in RTL code, having therefore all the functionalities of the final ASIC. The approach would be to use a *digital-on-top* design at the matrix level to avoid timing violations on all the digital signals, which are sent or received by the readout channel to or from the End Of Column logic. Digital-on-top design will also ensure an accurate timing distribution of the clocks and of the luminosity windows. On the other hand, because the power supplies distribution is critical for the analog performance, an *analog-on-top* design will be used for the peripheral electronics and for the floor plan.

Triple voting logic against SEE will be implemented for all control signals and registers but not for read-out data. The submission of the ALTIROC2 is expected by the end 2019/early 2020, followed by 6 months dedicated to its characterization and to perform tests concerning radiation hardness. A second iteration is scheduled early 2021 to be considered as the pre-production.

Regarding the MUX 64:1, is has been designed and will be submitted in an MPW run in October 2019.

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 7984 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.

The module requirements depend on the final layout HGTD option that is chosen. Three possible scenarios are envisaged (see Sec. 15.4). One of these scenarios should be chosen once final estimates are available for radiation levels in the HGTD volume with realisic ITk services/supports and once the performance of real-size prototypes of the HGTD sensor+ASIC system are studied under irradiation.

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 Chap. 5 and Chap. 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. These vendors are currently being qualified (on ALTIROC1 devices). 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 20.0 mm × 39.5 mm, with an array matrix of 15 × 30 1.3 mm × 1.3 mm pads and a dead region 0.25 mm wide around the active area. The readout ASIC has a total size of 21.7 mm × 19.9 mm and a matrix of 15 × 15 channels. The LGAD sensor has half the bump pads of each pad 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 to have a distance of 100 µm between ASICs, with no gap in the sensor coverage or disruption from different pixel sizes (see Fig. 7.1).

The LGAD sensors will be produced in 150 mm wafers of various thicknesses (depending on wafer and sensor providers), which 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 ALTIROC ASIC will be produced in 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 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. A small fraction of the mechanical bumps, that are located in the the periphery of each ASIC to provide rigidity to the assembly, can be faulty, as their role in not critical for the overall performance. Note tha 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).

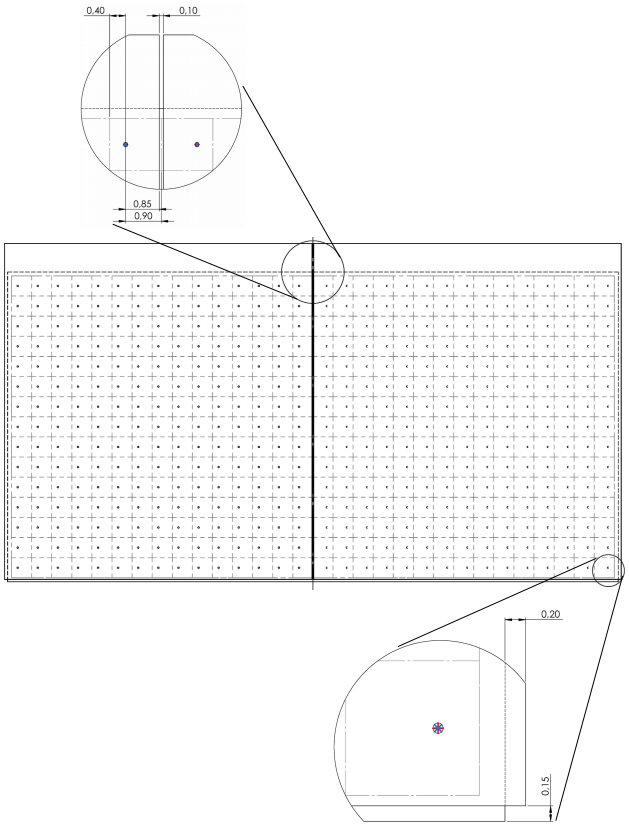


Figure 7.1: Sketch of the bare module (senor and ASIC). Distances are in millimeters. The bump pads on the sensor are shifted by 250 µm on each side of the sensor, to allow a 100 µm separation between the ASICs (see text).

7.2.2 First bare module prototypes: process and results

The first ALTIROC1 devices have already been assembled. 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.

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 [45]. 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 weight 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 Fig. 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 Sec. 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.

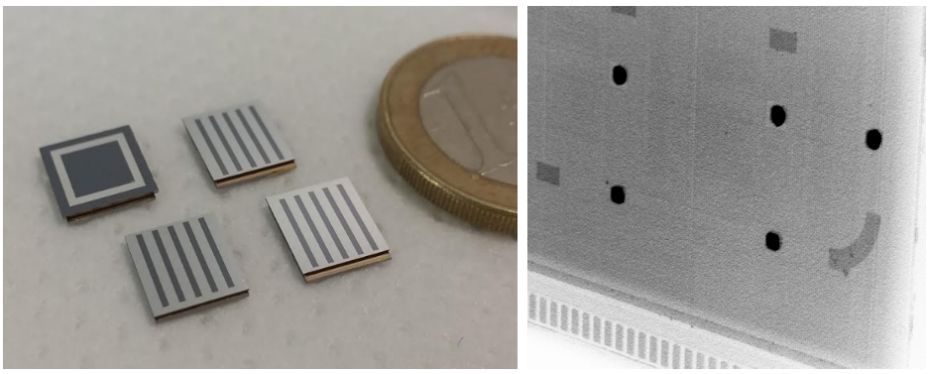


Figure 7.2: The first ALTIROC1 bare modules, with CNM and HPK sensors, and an x-ray image with a detail of a corner of one device are shown. In the x-ray image, the guard-ring solder bumps are in the periphery, while the bumps of two readout channels are visible in the center left of the image. The wire-bond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

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 to a PCB using Araldite 2011, see Sec. 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. Fig. 7.3 shows the shear and pulling tests being carried out on an ALTIROC1 hybrid.

The hybridization was also performed by the National Center for Advanced Packaging (NCAP China). NCAP is one of the leading companies in the integrated circuit packaging and testing industry in China. NCAP has more than 3200 m^2 cleanroom space and can provide bump-bonding service for 8 inch and 12-inch wafer. Its the production capacity for

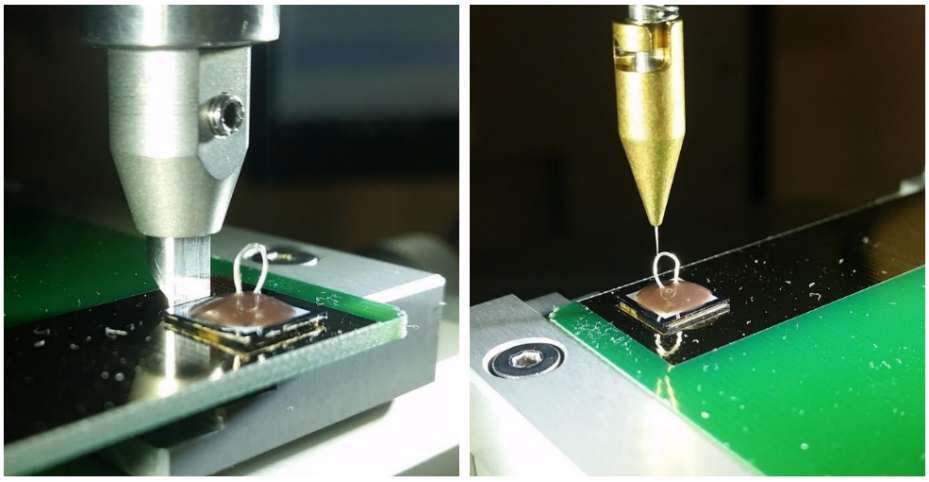


Figure 7.3: Shear and pull tests being carried out on an ALTIROC1 device. After thermal cycling during two weeks the device was able to sustain a maximum shear (pull) force of 1000 gf (100 gf).

module hybridization can fully satisfy the requirement by the HGTD project. 15 ALTIROC1 bare module prototypes have been hybridized in NCAP. Part of them are shown in Fig. 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 as shown in Fig. 7.5. These modules were sustain the maximum applied shear force of 1000 gf during shear test.

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 plot of Fig. 7.6. 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 are also wire-bonded to the test board for high voltage connection. The electrical connections of each channels in bare modules were checked by measuring analog output level in each channel of ALTIROC1 chip during charge injection tests. The results of the testbench measurements are described in Sec. 6.7.1.

The performance of the bare module prototypes hybridized in NCAP were also evaluated in electron beam tests at DESY in the fall of 2019. The test beam setup is shown in the right plot of Fig. 7.6. The EUDET-style telescopes, which consist of six MIMOSA26 pixel sensors, was used for tracking. The bare modules were mounted between the third and fourth telescope planes of the EUDET-style telescopes. A Silicon Photomultiplier (SiPM) that is coupled to a fast Cherenkov-light emitting Quartz bar is used for accurate timing reference in the beam line. The time resolution of the bare module, which consist of An ALTIROC1v2 ASIC and LGAD sensor arrays (HPK 3.1), is about 39ps. More detailed results are described in Sec. 6.7.2.

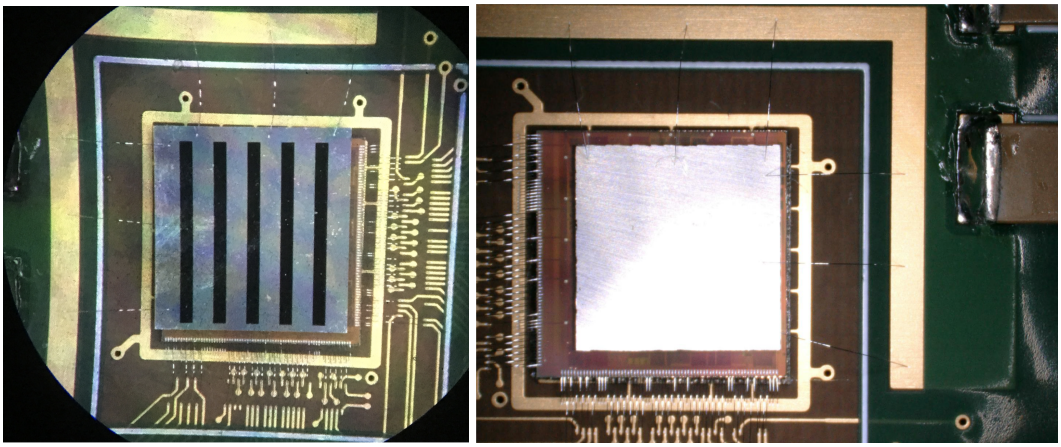


Figure 7.4: The bare modules hybridized in NCAP China. Left plot: The 5 × 5 pad sensors used in these prototypes were LGADs fabricated at Hamamatsu (Type 3-2). Right plot: The 5 × 5 pad sensors used in these prototypes were LGADs fabricated at NDL (Type 6).

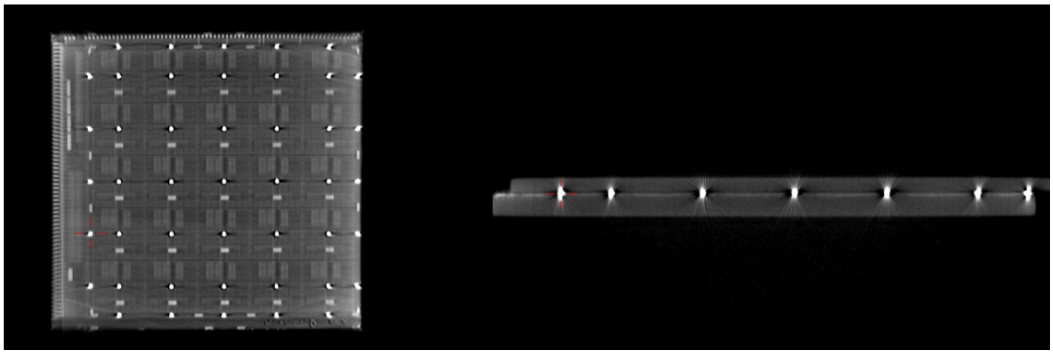


Figure 7.5: The x-ray image of the bare modules hybridized in NCAP China. In the x-ray image, the solder bumps are visible in top view and side view.

An alternative process explored during the initial R&D phase (but not intended for production) has also been developed to assemble ALTIROC0 devices. With Au bumps under bump metalization is not need since the ball bumps can be deposited directly on the aluminium of the front-end pads. 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 relies on solder bumps. As described above, both solder bump and gold stud bump prototypes have been produced at different HGTD institutes. Gold stud bumping is a process that enables hybridization without the

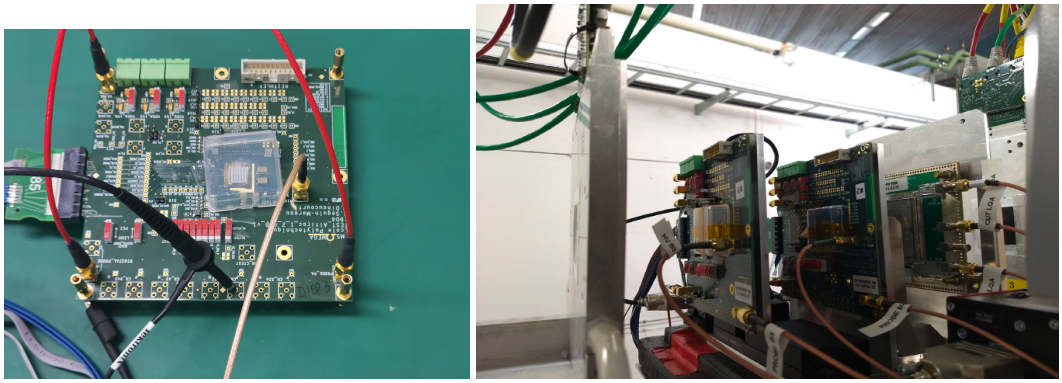


Figure 7.6: Left plot: A typical setup of testbench measurements for bare module prototypes. Right plot: The electron beam test setup for bare module prototypes at DESY in the fall of 2019.

need of under-bump metalization, since the gold studs can be deposited directly on the aluminium. This makes it very useful for research and development activities. However, the process is not scalable for large productions. Thus, solder bumps are the baseline hybridization solution.

The sensor fabrication sites will deliver silicon wafers. These wafers 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. Tab. 7.1 lists the relevant parameters related to the sensor wafer UBM.

Wafer material	Silicon
Wafer thickness	300 µm
Sensor size (R×C)	21 000 × 41 000 µm^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 under-bump metalization.

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. Tab. 7.2

summarizes the requirements for the UBM and bumping of the HGTD ASIC wafers.

Wafer material	Silicon
Wafer thickness	300 µm
ESD sensitive	Yes
Passivation	8750A SiO_2
ASIC size (rows×columns)	23 000 × 20 000 µm^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.

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.

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 single sensor. The cycle has to be consistent with the SnAg ($Sn_3Ag_{0.5}Cu$) solder bumps and result in a high hybridization yield. Tab. 7.3 summarizes the flip-chip requirements for the HGTD modules.

Alignment between ASIC and sensor	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

Each of the specifications listed above will be tested to show that they are met within the required acceptance criteria. 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 bump-bonding.

7.2.5 Production hybridization strategy

The total surface covered by the HGTD (about 7 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, and finally, module hybridization pre-production.

As shown above, the baseline bump-bonding process has been developed and successfully tested. Initial specifications have been established. Full size tests will be carried out as soon as the final sized sensor and ASIC become available. However, an effort is being made to advance this critical step in module assembly to avoid possible bottlenecks later in the overall activities. The specifications have already been provided to two companies (one in Germany an done in China) and discussions of a possible early qualification of the bump-bonding process with the currently available devices (ALTIROC1) are on-going. 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 possibility of using dummy ASICs and sensors for the vendor qualification will also be investigated. The target is to eventually carry out the final hybridization qualification on two to four companies, though the impact on cost and schedule will have to be evaluated.

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 tail will be connected during detector assembly (see Sec. 13.1). ASIC signals and low voltage, as well as bias voltage for the sensor (HV) will be connected by wire bonding. Fig. 7.7 shows three modules with the different 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 Tab. 7.4

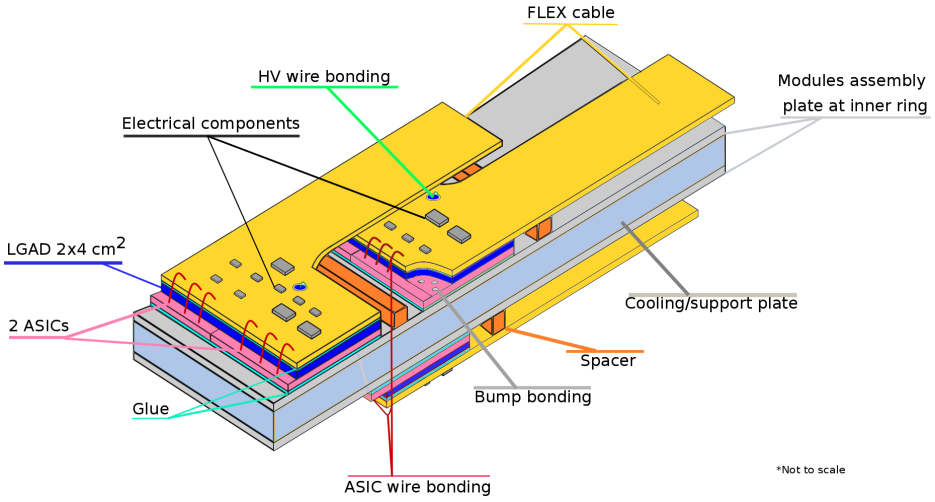


Figure 7.7: Schematic drawing of two adjacent modules on the top side and one on the bottom side of the cooling plate; the modules are mounted on thin support plates. Update figure

Alternative module design

Additionally 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 . Fig. 7.8 shows the schematic drawing of the alternative module design.

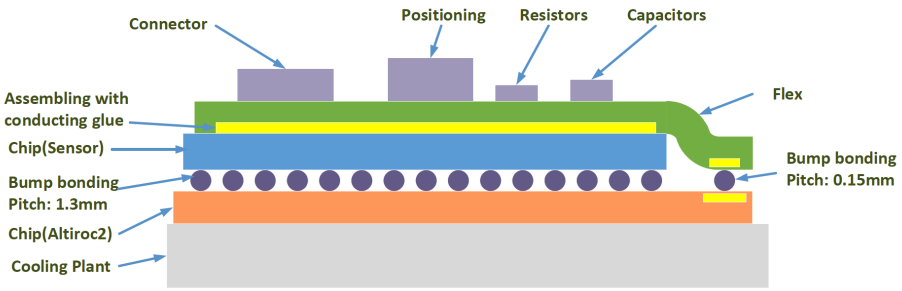


Figure 7.8: Schematic drawing of the alternative module design

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**	**Comments**
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	1(2) plane(s)	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 included in the flex cable design Is this still up to date? Number of lines on monitoring updated. Max. HV updated. Impedance requirements to be added, do they fit here?

7.3.2 Voltage distribution and signal readout: flex cables

The module flex and the flex tail, based on the flexible electronics technology, connect the signals from the module to the peripheral on-detector electronics. The geometrical constraints on the flex tails are determined by the available space between two layers (see Tab. 11.2), the distance between the modules and the peripheral electronics and the maximum number of modules per readout row. Considering the harshest constraints, the flex tails must have a maximum length of about 600 mm, width of at most 36 mm, and thickness of less than 250 µm. The total length of flex cables in the HGTD is 4500 m. To be updated

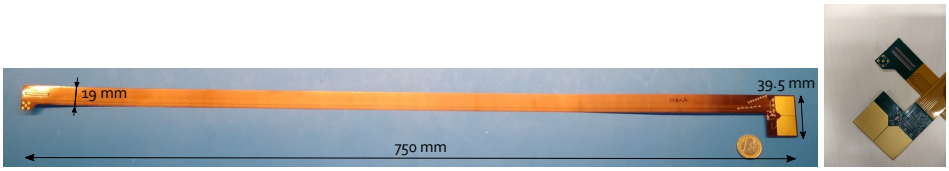
In terms of electrical requirements, one HV line has to be included in the design in order to bias the LGAD sensors (1 kV maximum). The HV line must have a sufficient insulation resistance (IR) to not affect the other lines (IR $> 10\,\mathrm{G\Omega}$). The types of signals to and from the ASICs in 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 signals for each module are listed in

Tab. 7.5. The impedance is required to be in the range of 90 Ω–110 Ω for the differential pairs and of 45 Ω–55 Ω for the single lines, while 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 **??**). To be updated? Check with Nathalie et al. for impedance requirements, it would be much better if 85-115 Ohm would be acceptable

As baseline design a module flex with a width of 39.5 mm and a length of 18.5 mm along the readout row is planned to be produced as a 4-layer stack-up with a thickness of 500 µm How many layers? How thick? It might be 3 or 4. 500um is the maximum thickness allowed. The flex tail is a 2-layer cable to be produced with different lenghts, 220 µm thickness and a width of 36 mm. A preliminary layout of the module flex and flex tail is shown in Add Figure with flex tail layout. Is it realistic to have one also for the module flex? Two separate connectors, one for HV and the other for allthe other lines, will be used to connect the module flex to the flex tail. A prototype of the flex tail has been ordered, while the design of the module flex is being finalised. This last sentence should be adapted to the actual status shortly before the deadline

Prototype characterisation

As part of the initial study phase, a prototype combining module flex and flex tail into one L-shaped 4-layer design has been produced with the aim to understand the technology requirements (materials, manufacturing capability, electrical and mechanical robustness) and address any potential problems by representing a significant subset of the signals (signal integrity, power distribution, HV insulation, interference and crosstalk). The direct interaction with the CERN PCB Service allowed to optimise the manufacturing process leading to the production of 4 prototypes of 750 mm length as depicted in Fig. 7.9. Apical, Krempel and Kapton were the commercial materials chosen for this prototype. The length was chosen based on a previous version of the detector layout and is significantly above the size of the longest flex tail to be produced for the HGTD.



(a) (b)

Figure 7.9: (a) Two flex cables prototypes. Top and bottom view. (b) Assembled extremities of the FLEX cable prototype from CERN PCB Service.

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 supposed to be full planes dedicated to powering and grounding. In order to perform the electrical tests the 4 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 Sec. 7.5), yielding very similar results.

Geometrical tests The thickness and the width of the flex cable must be homogeneous along its length. 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 Tab. 7.6. The spread of the values is acceptable, as well as the length and width average values.

	Length [cm]	Width [mm]	Thickness [µm]
Nominal	75	18	340
Measured	75.0 ± 0.2	17.99 ± 0.04	396 ± 16

Table 7.6: Mean values of the measured length, width, and thickness for 8 long flex cables. To be updated.

Power integrity A simulation of the voltage drop in each plane was performed with the Cadence Allegro Sigrity PI software package [51] and the expectation for the longest CERN prototype (750 mm) was estimated and compared with multimeter measurements (see Tab. 7.7). While the measurements show an excellent agreement with the simulation, the total resistance is above the specification for the power planes corresponding to about 200 mΩ for a flex cable of this length. Add comments on ground planes. Similar simulations for the current baseline design of the flex tail (also shown in Tab. 7.7) are instead well within specifications for all power and ground planes. The total power consumption is estimated to be about 4.4 mW cm^{-1}, corresponding to a total of 1.8 kW over the whole detector (see Tab. 11.1). Update total power consumption consistently with updated total length

plane type	CERN sim. [mΩ]	CERN meas. [mΩ]	tail baseline sim. [mΩ]
analog power	440	436 ± 5	211
digital power	229	230 ± 5	133
analog ground			46
digital ground			53

Table 7.7: Simulated and measured resistance of the analog and digital power and ground planes for the CERN prototype and simulation for the baseline design of the flex tail. Add simulation and measurement for ground planes

Insulation test to be updated

The insulation of the flex materials was checked for voltages up to 1 kV with the CAEN DT5521HEN HV power supply [52] 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.

Time Domain Reflectrometry results to be updated. New plots and impedance range of the new flex tail length

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. Two assembled flex cables were used to measure two differential pairs and two single lines that are accessible from the adapter board. The TDR module 80E08 together with the DSA8200 oscilloscope by Tektronix [53] was connected through SMA connectors to the adapter board. The impedance of the lines was measured and compared with the impedance estimated from simulation, that was expected to be within the requirements mentioned above. For all the measured lines the impedance is found to be above the expectation by up to 20% (see Fig. 7.10). The same tests will be performed on the CERN prototypes and the discrepancies will be followed up with the manufacturer to adapt the design accordingly.

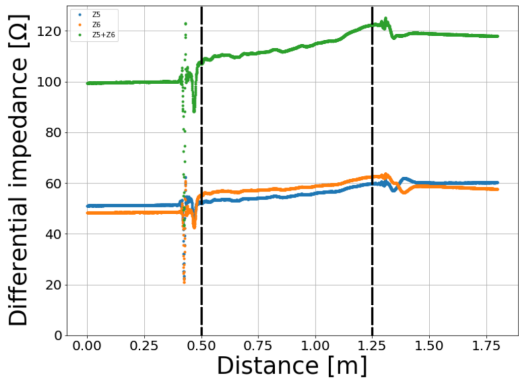
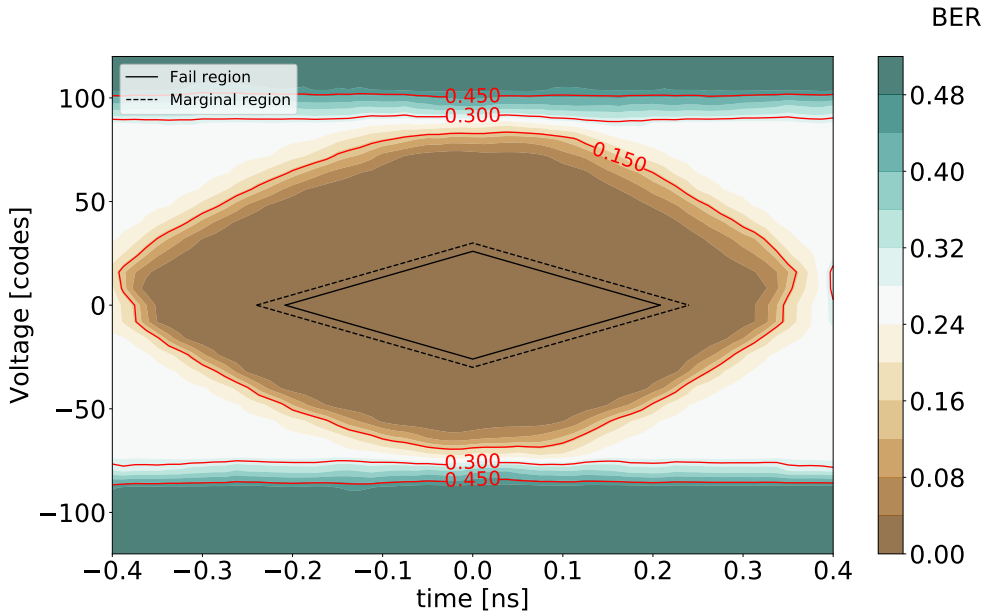
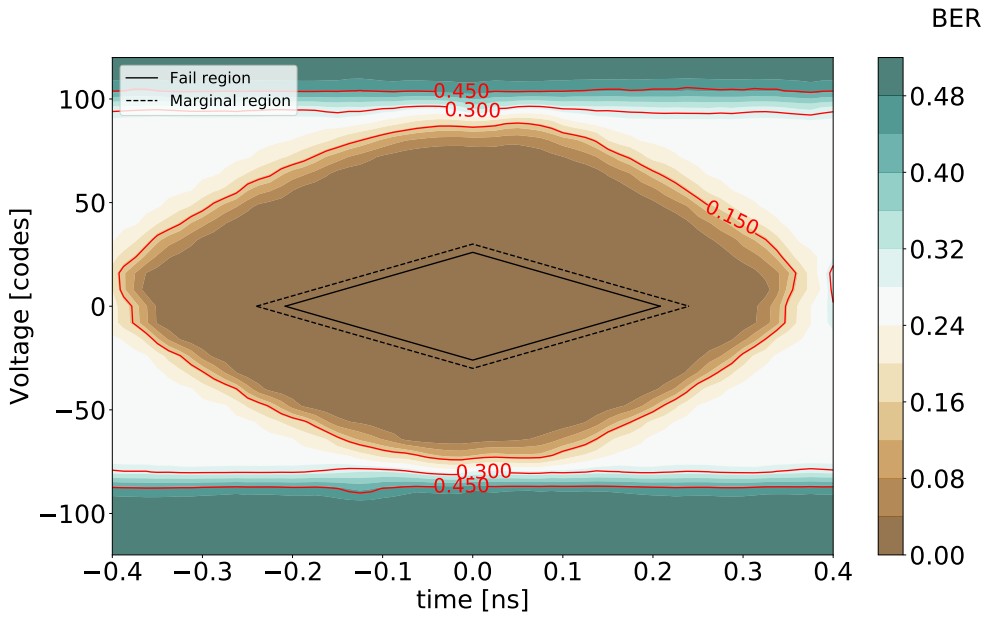


Figure 7.10: Preliminary result of the differential impedance for a differential line in the flex cable with the Schoeller design (green). The vertical dashed lines indicate the region of the flex cable. The orange and the blue lines corresponds to the impedance measurement for each line of the differential pair, used to calculate the differential line impedance.

Integrated Bit Error Test (IBERT) and eye diagrams Is the text up to date or only the figure?



(a)



(b)

Figure 7.11: 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 indicated the mask in which no errors are acceptable, the dashed line is the marginal region in which few errors can be tolerated. Figures updated!

To emulate the signals from the ASIC an FPGA on the Kintex KC705 evaluation board [54] have been programmed and connected the flex cable via an adapter board to build an automatic test setup for all the types of cables that are under test.

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 FLEX length, 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 assure 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 long flex cables at 1.25 Gbit s^{-1} with BER no more than 6×10^{-15}. This value is well below the acceptable error rate of 1×10^{-12}.

The same test was repeated with the HV up to 1000 V and showed no error for 8 days. The BER result obtained during this test was no more than 1×10^{-15}. 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 signals amplitude ranges from ±100 mV to ±200 mV. The eye diagrams in Fig. 7.11 measured without HV (a) and with HV (b) show a similar shape and opening area. The opening areas for both eye diagrams are larger than the no error accepted area limited by the mask. Tests over a longer time and a higher voltages are ongoing.

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.

The assembly procedure for all HGTD modules will be identical and use the same tools. 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, a few 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 if needed. 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. 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, the adhesive to the bare module and/or flex cable. The flex is then placed on top of the bare module and held in position until the adhesive is sufficiently cured.

An alternative process relies in custom made jig gluing tools instead of the pick-and-place machine. A bare module placement jig is used to position the module using precision pins and hold it in position with vacuum. A similar jig is used to hold the flex and then the adhesive is applied. Bolts or guides are used to align the flex and bare module jigs and slide them together. Positioning accuracy of $\sim 100\,\mu m$ is achievable with this method.

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. Fig. 7.12 shows the assembled ALTIROC1 device and the pull testing procedure.

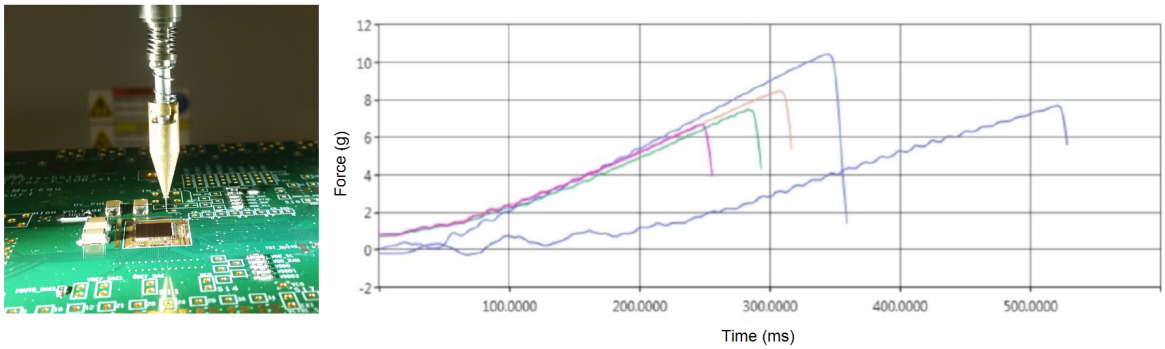


Figure 7.12: Photo of a mounted ALTIROC1 device being tested (left) and the measured wire strengths (right).

7.3.4 Assembly specifications

The flex cable specifications are described in Sec. 7.3 and summarized in Tab. 7.8. The flex cables will be produced with different lengths and two orientations of the L-shape, so that the general specifications can only be expressed in terms of tolerance of the actual size with respect to the design.

Tolerance in length	1 mm
Tolerance in width	100 µm
Maximum thickness	350 µm
Insulation resistance of HV line	10 GΩ
Maximum resistance of power planes	200 mΩ
Impedance of single lines	45 Ω–55 Ω
Impedance of differential lines	80 Ω–100 Ω
Maximum allowed BER	10^{-12}
Radiation tolerance	5 MGy

Table 7.8: Specifications of the flex cable.

The bare modules and flex cables that fulfil all the 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 Tab. 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.

7.3.5 Quality assurance / quality control

As described above, an exhaustive set of tests has been defined to qualify the flex cables before connecting them to the hybrid. This set of tests constitutes also the baseline procedure

Radiation tolerance	5 MGy
Lap shear force	5 MPa
Push-off strength	10 MPa
Wire bond pull force	6 gf
Positioning accuracy	100 µm

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

for quality control of the flex cables during production. However, since some of the measurements (e.g. IBERT) are time consuming, the option of performing them only on a limited fraction of flex cables will be considered once 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 **??**) and under controlled humidity conditions. Radiation tolerance will only be tested on a small fraction of flex cables, which will not be usable for assembly afterwards.

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 is foreseen. Furthermore a small fraction of the modules will be subjected to long-term burn in tests where the devices will be subjected to thermal cycles while being operated.

7.3.6 Production strategy for module assembly

The flex cable design will be finalized after testing it connected to the ALTIROC2 in the demonstrator described in Chap. 14. Companies that are expected to be able to produce long FCPs 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 standalone flex cable measurements, as well as gluing and wire bonding will be shared among a few HGTD Institutes, so that the production rate does not exceed 4 modules per day per Institute. An institute module assembly qualification procedure will be enforced 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.

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. Fig. 7.13 shows the positions of the modules on the front side (left plot in red) and back side (right plot in blue) allowing an overlapp 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. As described previously, the active area is divided in three rings (inner, middle and outer ring). Then, three types of support unit per side corresponding to the three rings are studied. The inner support unit consist of half disks of $120 < r < 230$ mm, the middle support unit consist of quarter disks of $230 < r < 470$ mm and the outer support unit consist of quarter disks of $470 < r < 660$ mm. The inner disks and middle disks will be replaced after every 1000 fb^{-1} and 2000 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, because of the positions of the modules are different for the two sides of the cooling plates, there are six different types of support units. The readout rows will therefore contain modules mounted on the support structures and they will be assembled independently. Fig. 7.14 shows a drawing of the detector units with the loaded modules.

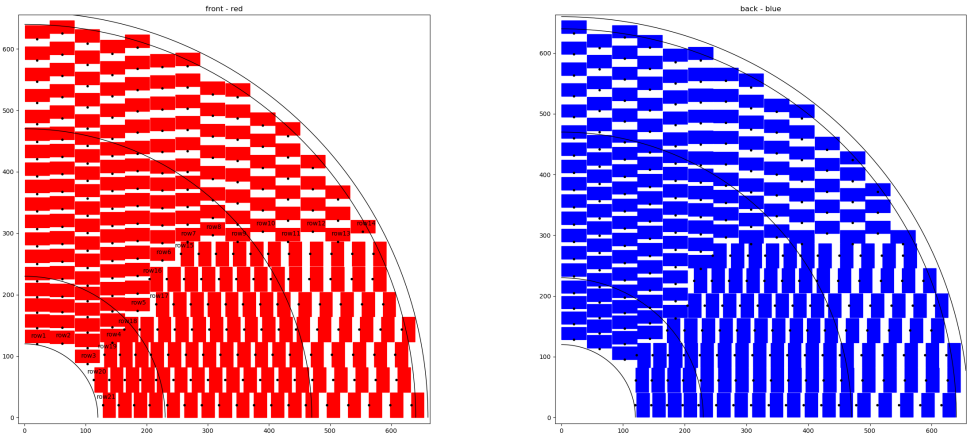


Figure 7.13: Position of modules and readout rows numbering for the front side of one disk (left plot in red) and for the back side of the same disk (right plot in blue). Smallest radius at 120 mm and maximal radius at 660 mm are shown. The 640 mm radius is the minimal target for the external instrumented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

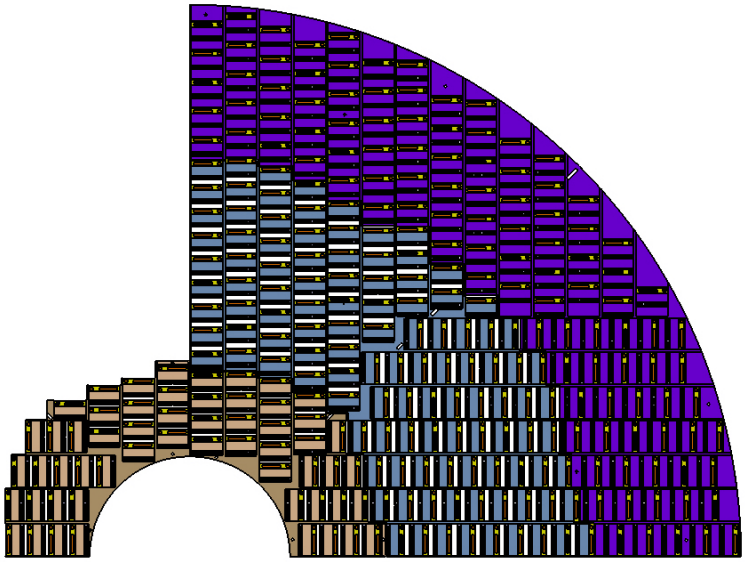


Figure 7.14: Detector units with modules assembled on the inner, middle and outer support plates

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. Currently the design of the support units consists of a pattern plate, with modules inserted into (between the plate and the cooling plate). The full size plate is divided into three parts as described in Fig. 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 4 mm. The target material is currently carbon fibre. For example, Fig. 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 (see Fig. 7.15). The positioning of modules is given by the windows of the plate (see Fig. 7.15). Once this 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 are not critical. Moreover, thermal grease will be used to improve the contact.

For better mechanical strength and rigidity, some reinforcement are added (Fig. 7.15). 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 simple 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 thicker plate has better rigidity. On the downside, this design only allows for a small surface when gluing the module to the

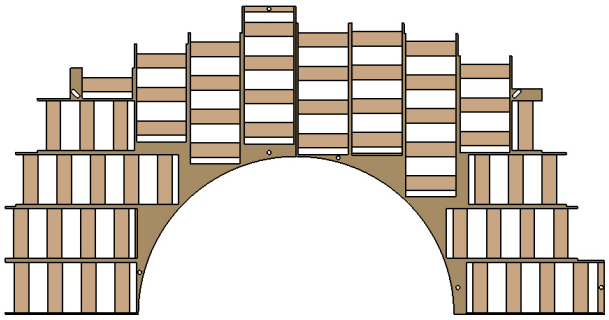


Figure 7.15: Drawing of inner support unit with holes for fixation on the cooling plate

plate and the mechanical strength and long term stability has to be studied. Some tests have to be performed and several prototypes and the demonstrator will be useful to conclude (see Chap. 14). In terms of thermal conductivity, this design has the advantage of placing the modules in direct contact with the cooling plate, avoiding thermal constraints in the material of the plate and the glue. Should a module be found to be faulty after gluing to the support, rework should be relatively easy. Conclusive tests 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 thanks to four glue dots 2 mm diameter each (see Fig. 7.16). The glue dots are deposited onto the edges of the flex connector. The glue for module loading into the intermediate plate should meet a priori, the parameters listed in Tab. 7.10.

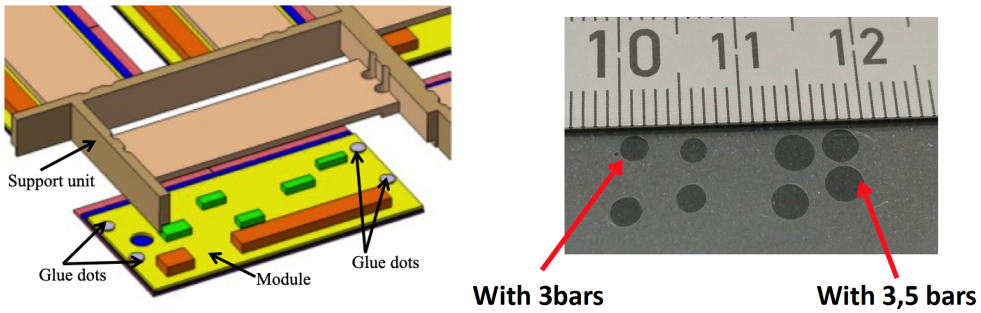


Figure 7.16: Schematic view of the module with the four glue dots allowing the fixation with the support unit (left) and test of glue deposition (right) - Pressure values are an example of tuned parameters, depending on the duration and temperature.

With these constraints, six types of glues have been chosen to perform the tests: Araldite2011; EG7655-LV; EG7655; EG7658; EG8050; Stycast 2850FT.

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.

Their characteristics have been checked in the MaxRad (Materials and Adhesives for Extreme Radiation Environments) CERN database. Ease of implementation (fluidity, life time, duration and temperature polymerisation) has been evaluated and the ITk choice has been considered for radiation hardness Moreover, the push-off strength and lap shear test have been performed in several configurations. These tests have been done using some dummy modules, with a piece of flex cable glued onto to mimic the flex connector. Other tests with glass plate have been performed to determine the volume of glue to obtain the correct thickness and surface (see Fig. 7.16). Finally, taking into account all the tests already done and the results and recommendations from ITk project, ARALDITE 2011 is chosen as the baseline for the loading of the modules onto the support unit.

7.4.4 Procedure for assembly and qualification

The procedure for stave loading will be tested when assembling the demonstrator (see Chap. 14), which will be also used to improve the definition of the assembly procedure and the qualification steps. Tools are being developed and tested for this design. 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. 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 pattern of the module positions. They are maintained thanks to a suction system included in the plate.
2. Four glue dots are dispensed on the left and right edge of the flex connector (see Fig. 7.16); the thickness of the glue is insured by the automatic dispenser.
3. The support unit is put in place and compressed on all modules at a nominal compression strength. A 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 and fixed on a 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 can be performed at this stage.

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 flex components 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 due to the stacked flex cables. 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 the choice of the design is finalised, the production of the support units will be carried out by a company and the monitoring and control by an institute. 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 staves has to be delivered. 80 support units (16 inner, 32 middle, 32 outer) will be produced. Since the module positions are different on the front side and on the back side of a cooling plate, the designs of the support units are different for the two sides. In total, there are 6 types of support units. 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 if the detector units and the fragility, different types of dedicated packaging will be necessary for transportation from loading sites to CERN.

7.5 Thermal calculation

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 [55].

	Material	Thickness [mm] [mm]	Thermal Resistivity [W/ (m°C)]
Sensor	Si	0.25	124
Bumps	SnAg	0.05	79
ASIC	Si	0.25	124
Foil	Polymer	0.10	3.5
Structure	CarbonFiber	0.50	1
Cooling	Graphite foam	2.00	30
Tube	Al	0.30	135

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

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 Tab. 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 Tab. 7.11. 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 and the outer radius is 1.8 mm.

The cooling is simulated as convection which is applied on the surface of the cooling pipes. Their nominal temperature 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 resistance of 0.01 Wmm^{-2}°C^{-1} is applied to the contact between ASIC and foil as well as foil and the carbon fiber structure. The contact resistance 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 Fig. 7.17 the result of the thermal simulation by ANSYS is shown. The maximum temperature difference is 7.7 °C. If the ASIC is powered alone, the temperature difference is 5.6 °C, for the sensor alone, the temperature difference is determined to be 2.2 °C. The thermal resistance for the sensor is therefore 5.4 °C/W and for the ASIC it is 4.7 °C/W. As the difference between these two resistances of 0.7 °C/W is due to the soldering bumps, the

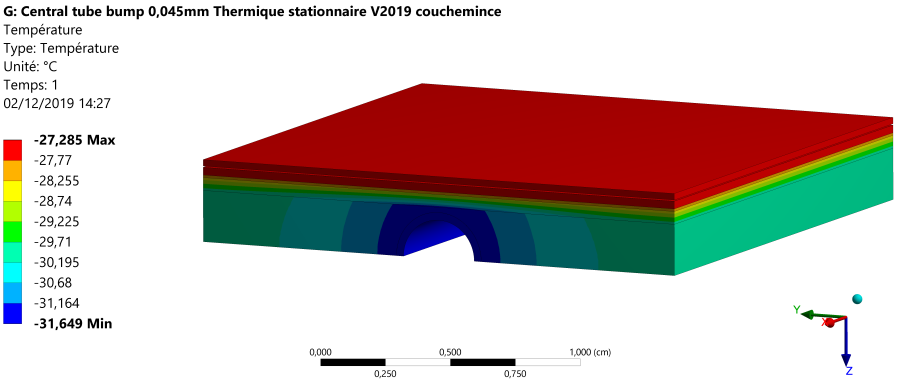


Figure 7.17: The temperature distribution is shown for the baseline power consumption with an ASIC and a sensor half.

thermal resistance of the 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 °C/W, the larger value for the individual bumps can be understood as the heat transfer will see also the resistance in the sensor plane before reaching the bumps in order to flow to the cold reservoir.

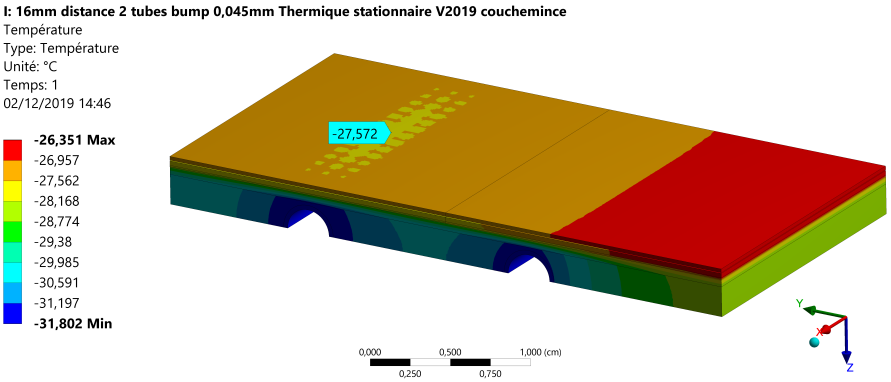


Figure 7.18: The temperature distribution is shown for the baseline power consumption with two ASICs and a sensor.

As a second step the second half the sensor was added as well as the corresponding ASIC. The current design of the cooling pipes calls for pipes every 16 mm, therefore a second cooling pipe was added at the nominal distance leading to an asymmetric configuration shown in Fig. 7.18. Compared to the previous simulation the temperature difference increases to 8.7 °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 which is close to the result of the previous simulation within 5%. For the simulations with only the sensor or ASIC disspating power the temperature increase is globally larger, however as the increase is less than a factor 2, but the power ins 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.

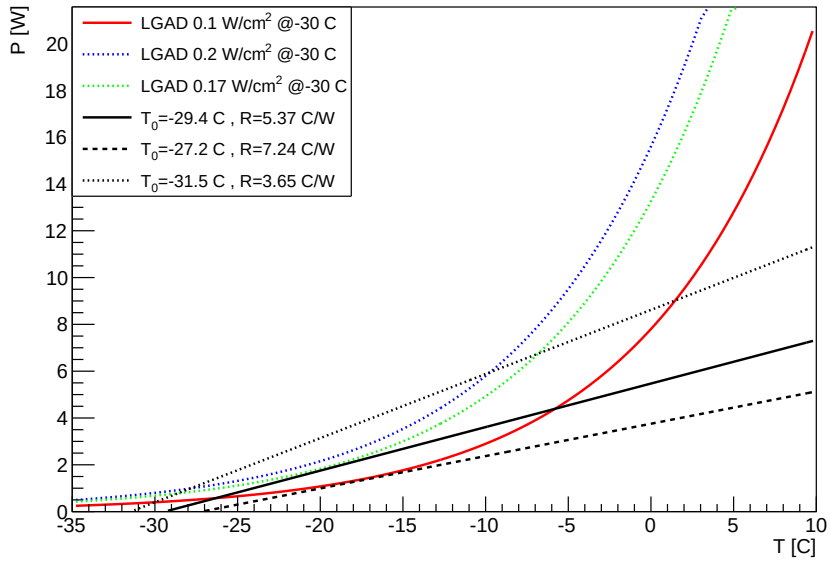


Figure 7.19: The power dissipation of the sensors as function of the temperature is shown as well as the thermal properties of the system.

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 [55].

The power dissipation of the sensor is shown as a function of the temperature in Fig. 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 green dotted curve in Fig. 7.19 shows the maximal power dissipation the system can

handle. For about 0.17 W/cm^2 the power disspation of the sensor is tangent to the black line. Thus compared to the baseline a margin of 70% is included in the system. The temperature dependence is modeled according to Sec. 5.5.8. The model was compared to the one used in [55] 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. The black dotted line shows the result of increasing the thermal resistance starting at the ASIC by 40%. Additionally the temperature increase of 5.6 °C for the nominal system includes the contact resistance 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.

If both the Carbon Fiber and the Graphite Foam were to be replaced by a system of Aluminum, the thermal resistances of the system would be improved further. The effective operating temperature of system would decrease to −31.5 °C and the thermal resistance would decrease to 3.7 °C/W as shown in Fig. 7.19 leading to further margin in the operation of the system.

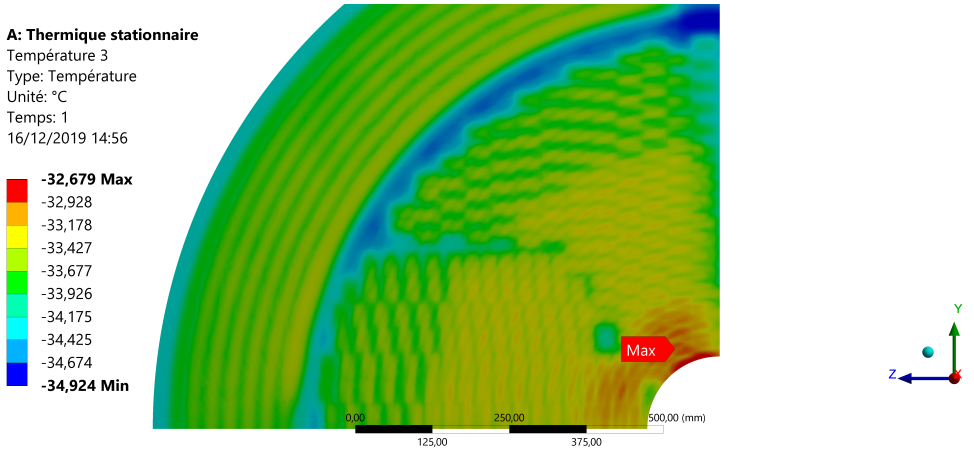


Figure 7.20: The temperature distribution on the surface of the cooling system is shown for a quarter disk when applying 0.4 W/cm^2 at the location of each ASIC.

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 quarter disk using only Aluminum as explained above. 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 Fig. 7.20. The maximal temperature is −32.7 °C. The modules and the space between the modules explains the temperature variation. At the

position of a module, in the inner ring of the HGTD, the resulting temperature is −33.3 °C. In the space between the modules the temperature decreases only by about 0.1 °C.

Taking the maximal temperature of −32.7 °C from the global model, the surface temperature of the cooling system of the detailed model with a single cooling pipe is fixed to this temperature. The temperature at the LGAD is determined to be −31.5 °C, using the baseline power dissipation.

The curvature of the cooling pipes could lead, for some modules, to a loss of cooling surface of the pipes, increasing the thermal resistance. However the temperature difference for the hottest module position with respect to the nominal temperature of the cooling system is 3.5 °C. In the single cooling pipe model the temperature difference for the same power dissipation is larger with 5 °C. Therefore there is no indication of such an effect. Additionally, the distance between two cooling tubes is smaller than the size of the module. Therefore the temperature difference in the simple model is increased artificially for the same power dissipation, increasing also the thermal resistance.

The studies indicate that the simple model with a single straight cooling pipe is a conservative representation of the system.

8 Power Distribution, Control, and Safety Systems

This section covers the powering of the detector, including the 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 PP-EC patch panel area, up to the on detector peripheral electronics and modules sitting inside the vessel. The grounding and shielding schemes are also described. The details of the services needed to power the detector and respective connectivity are described in Chap. 12.

8.1 High voltage

Each of the 8032 LGAD sensor modules of the detector require 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 Chap. 5). To allow an average leakage current up to 5 µA per pad for radiated sensors, a 3 mA supply current will give sufficient margin. The bias voltage of the sensors has to be adjusted due to the gain degradation with the received fluence. Fig. 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. The baseline choice is to use individual adjustable voltages to allow for optimal operation of the sensor modules. This requires 4016 HV supply channels per end-cap. The supplies will be based on commercial multi-channel rack mounted units located in the service cavern. A schematic layout of the high voltage system is shown in Fig. 8.1.

The return line for HV channels will be done through a common ground, as for the liquid Argon calorimeter. This ground is available from the peripheral electronics boards via the modules analogue ground. The shielding will follow up to the filter units at the PP-EC patch panel boxes. A further low pass filter is placed on the flex cables near the sensor modules.

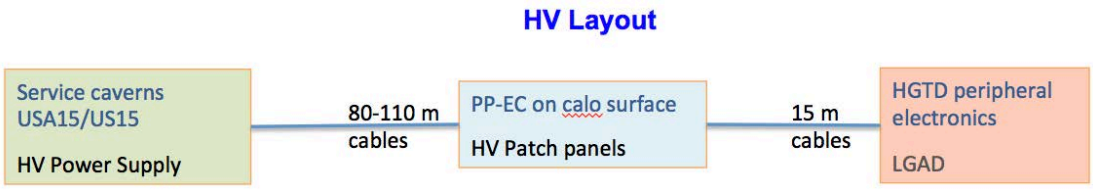


Figure 8.1: HGTD High Voltage Layout

8.2 Low voltage

For supplying the low voltages needed by the front-end end peripheral electronics a three stage system is used, as shown in Fig. 8.2. The system will have to be able to deliver almost 20 kW at 1.2 V. Bulk power supplies located in USA15 provide 300 V DC current to DC-DC converters placed in the PP-EC areas (described in Sec. 12.1.3). 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 Chap. 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 of 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 Sec. 9.3). 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 can supply analog power to 4 modules or digital power to 3 modules. With 2008 modules per disk (or double-sided layer), 1196 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 peripheral boards themselves.

These DC-DC converters on the peripheral electronics will need to provide almost 5.0 kW of power per disk. With an efficiency of 65%, each disk has to receive 800 A at 10 V which will be supplied by 52 channels providing 16 A each.

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 Tab. 8.1.

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 is required at these locations. A water leak-less cooling system, providing water at $\approx 18\,^{\circ}\mathrm{C}$, and corresponding pipes/manifolds on the calorimeter

Voltage	Location	Current/channel	Nb of channels/units
300 V	USA15	10 A	14
300 V $\rightarrow$ 10 V	PP_EC	16 A	208
10 V $\rightarrow$ 1.2 V (or 2.5 V)	On peripheral board	4 A	5360

Table 8.1: Type of LV units, location, delivered current per unit and number of units/channels.

surface will be needed. Details on the services, patch panels area and cabling are given in Sec. 12.1.6. A schematic layout of the low voltage system is shown in Fig. 8.2.

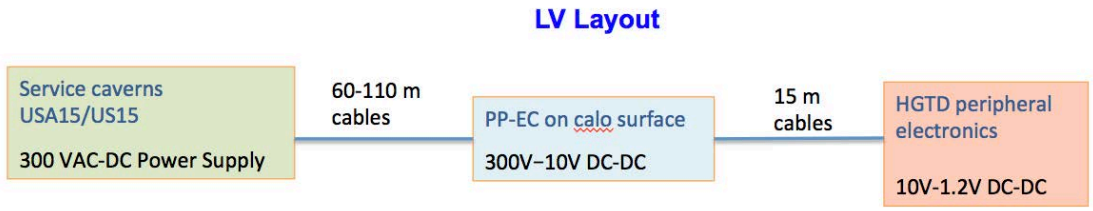


Figure 8.2: HGTD Low Voltage/power Layout

8.3 Grounding/shielding

The grounding and shielding of HGTD follows similar requirements as defined for ITK. The ground reference point for the HGTD itself will be the inside of the detector vessel. The inside of the vessel 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 will be 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.

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.

9 Peripheral Electronics

The on-detector peripheral electronics transfers data between the detector modules and the TDAQ, DCS, and luminosity systems as well as having a central role in the monitoring of sensor temperatures and supplied low voltage. The system is based on the CERN-developed lpGBT ASICs [56]. The modules are connected via flex cables, while signals to and from the TDAQ and the luminosity systems are transferred on optical fibres where the DCS data and commands are embedded in the data streams via the TDAQ optical fibres.

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 ASIC position. Flex cables for modules placed at a radius above 320 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 commands to the ALTIROC ASICs at 320 Mbit s^{-1} as well as the lines for the ASIC low voltage power supplies, control signals and the bias voltage of the sensor. The digital output data from several ASICs are merged in lpGBTs on peripheral boards (PEB) and transmitted on optical fibres to the off detector DAQ system. 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 DAQ system. An overview of the HGTD readout chain is presented in Fig. 9.1.

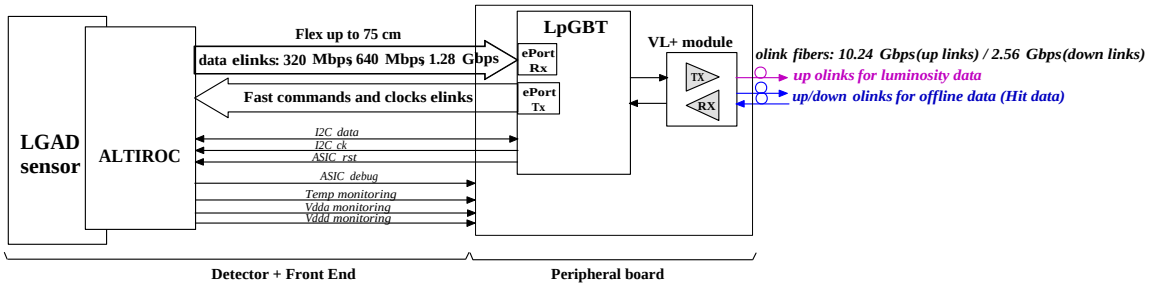


Figure 9.1: Upstream and downstream data flow. The e-links transmit data, fast commands and clocks between the ALTIROC ASIC and the lpGBT. VL+ is the Versatile Link+ VTRX+ module. 12c-bus, ASIC control and monitoring lines from the ALTIROC are also shown.

The peripheral electronics also includes the 10 V to 1.2 V DC-DC converters for the digital and analogue voltage supplies 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, where voltages to be measured are selected by analog 64-to-1 multiplexers mounted on the peripheral boards as described in Sec. 9.4.

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 off the detector. 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 Fig. 9.2.

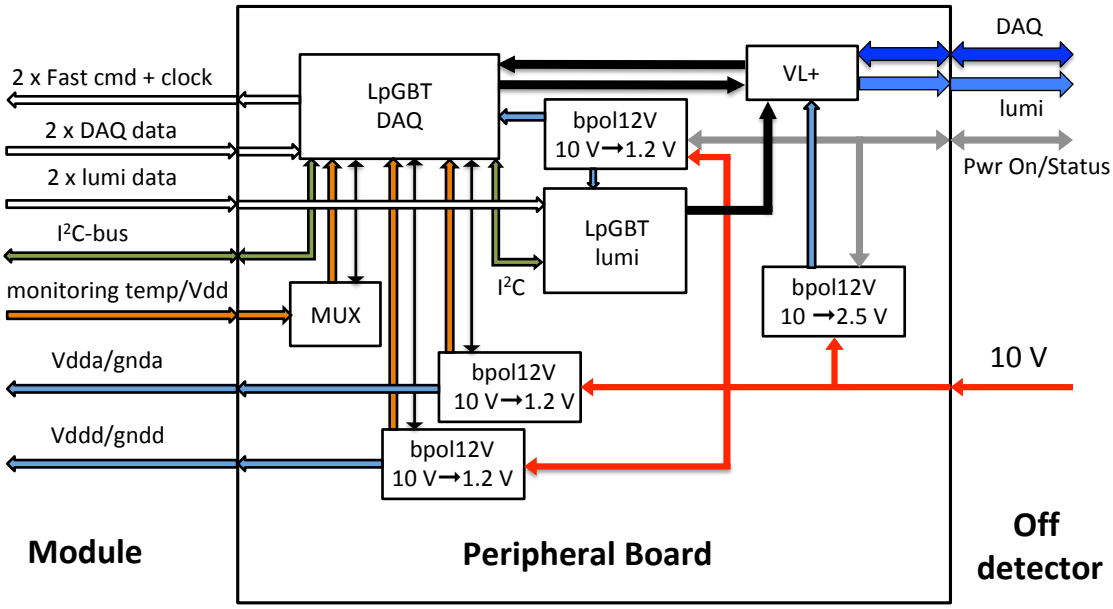


Figure 9.2: Block diagram of the peripheral electronics for powering and read out of a module. Multiple modules are connected to the same DC-DC converters and lpGBTs. The brown lines indicate voltages measured by the multiplexed ADC on the lpGBT. Light blue lines are low voltage power supplied from bpol12V DC-DC converters. The thin black lines are control signals via the general purpose I/O lines 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.

As introduced in Sec. 2.3, each HGTD vessel contains two cooling disks (shown in Fig. 2.4), with detector modules mounted on both sides, thus having two instrumented layers per

disk. The baseline design for the peripheral electronics is to have five PEBs per quadrant per side of each cooling disk. Such a layout yields 80 boards per HGTD vessel, and 160 boards in total. Each board covers three or more readout rows in order to have similar number of ALTIROC ASICs connected per board to allow to optimize the use of the lpGBTs by sharing across readout rows.

9.1 Requirements

9.1.1 Data transfer

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 simulation and results are presented per ASIC for an $\langle\mu\rangle = 200$ in Fig. 9.3. The radial dependancy is clearly seen, with a maximum below 20 at the innermost radius. Such a rate can be accomodated 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 200 mm and 320 Mbit s^{-1} at radii above 300 mm. These rates are chosen in order to minimise the numbers of lpGBTs and optical links while allowing a 50% increase of the expected number of average hits per ASIC at a readout rate of 1 MHz.

In addition each ALTIROC ASIC requires a 320 Mbit s^{-1} fast command link to supply both the bunch crossing information and the TDAQ commands. A 320 MHz clock extracted inside the lpGBT from the command data packages is also sent to each ASIC.

9.1.2 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 width of the HGTD is only 75 mm, the space available for the electronics in the z-dimension is also very small: 9 mm with a 1 mm margin. The layout of the PEBs will keep the same quadrant symmetry as has been described for the readout rows, but taking into account that a 10 cm path has to be left empty in the area close to the shortest readout row. This is to allow enough space for the connection of the cooling services, and in principle is only needed in two out of the four quadrants per disk.

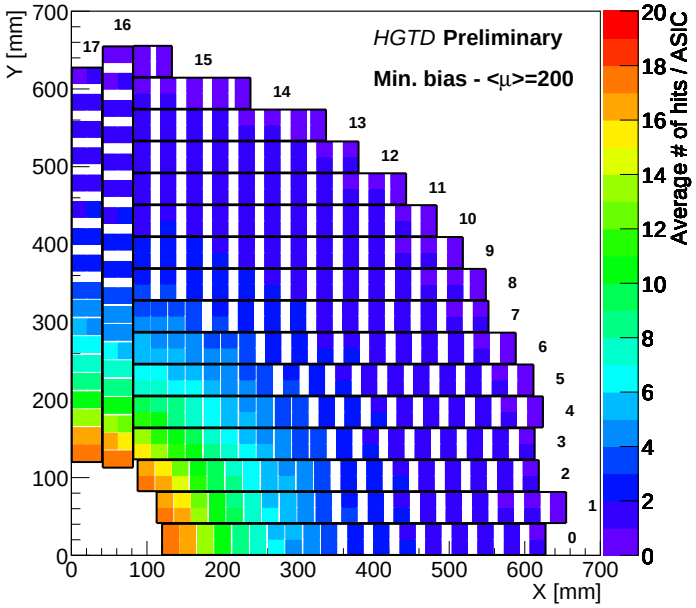


Figure 9.3: Average number of hits in each ASIC in one quadrant of the second layer in a simulated sample with $\langle\mu\rangle = 200$. The rectangles indicated by the black lines correspond to the readout rows. The numbering of the readout rows is also shown.

9.1.3 Radiation tolerance

All the active electronic components will be located at radii above 720 mm. Extrapolating from Fig. 2.13, 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.

9.2 Data transfer

The data transfer is, as mentioned above, based on the lpGBT ASIC. The data transfer between the peripheral electronics and the off detector systems uses optical fibres based on the VTRX+ optical transceiver developed within the Versatile Link Plus project.

9.2.1 LpGBT

A block diagram of the lpGBT is shown in Fig. 9.4 and more details concerning its specifications can be found in [56].

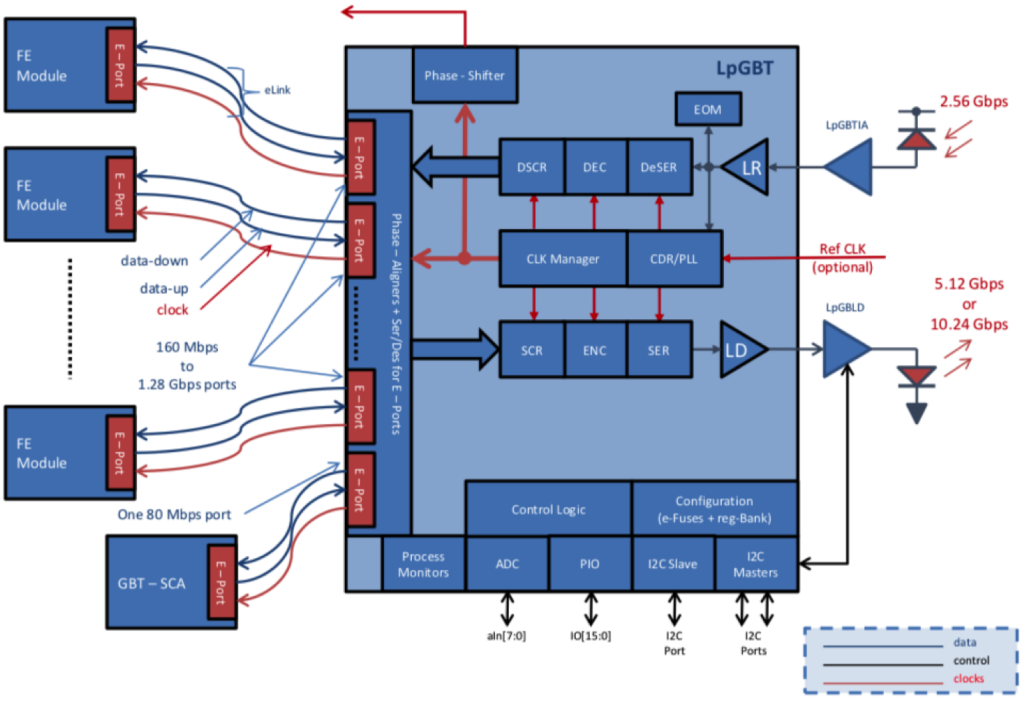


Figure 9.4: Block diagram of the lpGBT ASIC.

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 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, 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 though 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 via one of their I^{2}C bus master ports to configure the lpGBTs for the luminosity system of the same board.

- **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 where also four independent clock streams are mirrored four times. 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 mimimum 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 bunch crossing per ALTIROC at $\langle\mu\rangle = 200$ of up to about 40, 20 and 10, respectively, at 1 MHz of event readout. Based on the expected average hit rate per ASIC (Fig. 9.3), 1.28 Gbit s^{-1} will only be used for ASICs placed at radii of less than 190 mm and 320 Mbit s^{-1} will be used for radii greater than 275 mm in order to keep the number of lpGBTs to a minimum, in view of the limited space available for the peripheral electronics.

- **Luminosity data**. Each ALTIROC ASIC at radii larger than 320 mm provides 16-bit luminosity data for each bunch crossing, transmitted to the lpGBTs via the flex cables. 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 are 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 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 can be 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. I^{2}C-bus master ports are furthermore used, as previously mentioned, to configure all lpGBTs of the luminosity readout. Depending on the final optical link units, these may require I^{2}C-bus connection for configuration of the laser drivers. Since the 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.

9.2.2 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 under development within the Versatile Link plus project at CERN are specified to handle four fibres for transmission and one for receiving. The dimensions are specified as $20 \times 10\,\mathrm{mm}^2$ footprint and 2.5 mm in height, which fits within the available 9 mm. The specified radiation hardness, 1 Mrad 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.

9.3 DC-DC converters

The peripheral electronics will contain DC-DC converters supplying the required 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 is the limited available surface available for peripheral electronics. The footprint of one DC-DC converter circuit is assumed to be 12 mm $\times$ 30 mm in the final layout, with a height of 5 mm including the necessary Faraday cage surrounding the inductance.

The maximum output current of the bpol12V is 4 A, with an estimated efficiency of around 65% at currents between 3 A and 4 A. 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 150 Mrad and $2 \times 10^{15}\,\mathrm{n_{eq}\,cm^{-2}}$. The converters need a 460 µH inductance as well as further filtering components. The printed circuit board layout optimized by the bpol12V developer team will be used.

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. The power consumption of an lpGBT will not exceed 0.75 W. One DC-DC converter is able to supply voltage to the analog part of ALTIROCs for 4 modules, while a single converter can only supply the digital part of 3 modules. Converters supplying the ALTIROC ASICs are switched on via general purpose I/O-lines from lpGBTs. A second I/O-line is used to report the converter status. The DC-DC converters supplying the PEB itself, i.e. the lpGBTs and the optical links are switched on by an external 1 V signal. The status of these converters is read out via external electric cables (open drain) and on I/O-lines on lpGBTs other than those they supply to allow to

differentiate between possible power failures and lpGBT failures. Care has to be taken that the external electrical cables do not violate the grounding rules.

The DC-DC converters supplying voltages to the ALTIROC ASICs are switched on by applying a voltage (at least 850 mV) which is generated via general purpose I/O-lines from DAQ lpGBTs. The status of the DC-DC converter is reported via an open drain Power Good output and monitored via lpGBTs.

9.4 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. Depending on if used in transceiver or simplex mode, a further two bits may be available for I^2C buses. Since, in the baseline option, 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.

Each flex cable will, as described in Chap. 7, carry 5 voltages: 2 temperatures from sensor of each of the two ALTIROC; received analogue and digital supply voltages and analog current return voltage. 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 Sec. 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.

The peripheral electronics boards will each as mentioned above 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 luminosity system lpGBTs and the DAQ lpGBTs that are not pre-fused to receive configuration via optical links.

9.5 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 two bundles of up to 10 cables. Each flex cable is 18 mm wide. Development work is ongoing, in collaboration with commercial companies, aiming at connectors with a height less than 3 mm, a width for two parallel cables less than 40 mm, and with a length of closely stacked connectors in the direction of the cable less than 6 mm per connector. The total occupied surface of these connectors on the peripheral boards will thus be 40 mm $\times$ 60 mm or less per readout row, dependent on the number of modules per row.
- Although D-sub type connectors (37 and/or 25 pins) are good candidates for connecting the high voltage, custom made connectors are being investigated. All modules have individual high voltage. The ground plane of each peripheral board defines the ground of the modules to which they are connected via the analog ground plane on the flex cables.
- The peripheral boards will each require up to 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.

9.6 Peripheral boards

9.6.1 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 letting the readout rows 0–2, 3–5, 6–9, 10–14, and 15–17 be combined into one board each. (The read out row numbering is shown in Fig. 9.3). The combination allows to reduce the number of lpGBT ASICs, multiplexers and VTRx used and to better use the available surface area at the outer radius of the disks.

The number of modules, e-links for DAQ at different transfer rates and luminosity e-links per peripheral boards are shown in Tab. 9.1. The bit rates of the DAQ e-links will be re-optimised for the final layout based on further simulations, making full use of the available lpGBT capacity.

A number of considerations have to be taken into account for the actual PEB design.

Peripheral board	Read out rows	Nb of modules	1.28 Gb/s	640 Mb/s	320 Mb/s	Luminosity
Front						
0	0-2	56	24	24	64	56
1	3-5	56	8	32	72	66
2	6-9	59	0	2	116	110
3	10-14	45	0	0	90	90
4	15-17	39	16	16	46	42
Back						
0	0-2	55	20	26	64	56
1	3-5	54	6	34	68	62
2	6-9	56	0	4	108	104
3	10-14	42	0	0	84	84
4	15-17	37	12	18	44	38

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.

- To limit the implications of possible failing components, care must be taken such that as few modules as possible and preferably not in a contiguous area will be affected. This requires that modules sharing the same DC-DC converters are not contiguous. They have to share the same lpGBT for readout which will also transmit their clock and fast commands, control their DC-DC supply as well as DCS 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 as far as possible share luminosity lpGBTs who will receive their configuration control and clock from the readout lpGBT.

- 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.

- The distance between lpGBTs and VTRx must be kept small to minimize the risk of bit error at the high data rate.

9.6.2 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 Tab. 9.2. The number of DC-

DC converters is relying on the real power consumption of the final ALTIROC still under design.

With the number of components from the table, assuming 20 mm × 10 mm for the VTRx, 20 mm × 20 mm for the lpGBT including the necessary decoupling capacitors, 15 mm × 15 mm for the multiplexers, and 12 mm × 30 mm for the DC-DC converters, the total surface for the components of the front layer of a quadrant is 1030 cm^2. The available board surface per quadrant, removing the 15° for the cooling manifold and 450 cm^2 for flex cable, HV, and LV connectors, is 2270 cm^2 giving sufficient surface for the active components.

The full electric layout of the PEBs has yet to be performed. Fig. 9.5 shows only a conceptual design demonstrating the complexity of these boards due to the component density and limited area and thickness available. This conceptual design does not take into account any PCB design rules yet and can only be considered as an illustration of the final PEB design. A functional prototype of the PEBs is scheduled in the demonstrator program in 2020 in parallel with a complete PCB design of a real size PEB.

Peripheral board	lpGBTs DAQ	lpGBTs Luminosity	DC-DC converters	MUX	VTRx
Front					
0	8	4	36	5	8
1	8	5	37	5	8
2	8	8	39	5	8
3	7	7	31	4	7
4	6	3	26	4	6
Back					
0	8	4	36	5	8
1	8	5	36	5	8
2	8	8	37	5	8
3	6	6	31	4	6
4	6	3	26	3	6

Table 9.2: Numbers of lpGBTs, analog multiplexers, VTRx, and DC-DC converters for the different Peripheral Electronics Boards.

9.7 Power dissipation

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 1.12 W per ALTIROC ASIC (Fig. 11.2), the total required power delivered to the ASICs from the DC-DC converters will be 4.5 kW per double sided layer. At 65% efficiency, the power loss due to the ASIC supplies will be

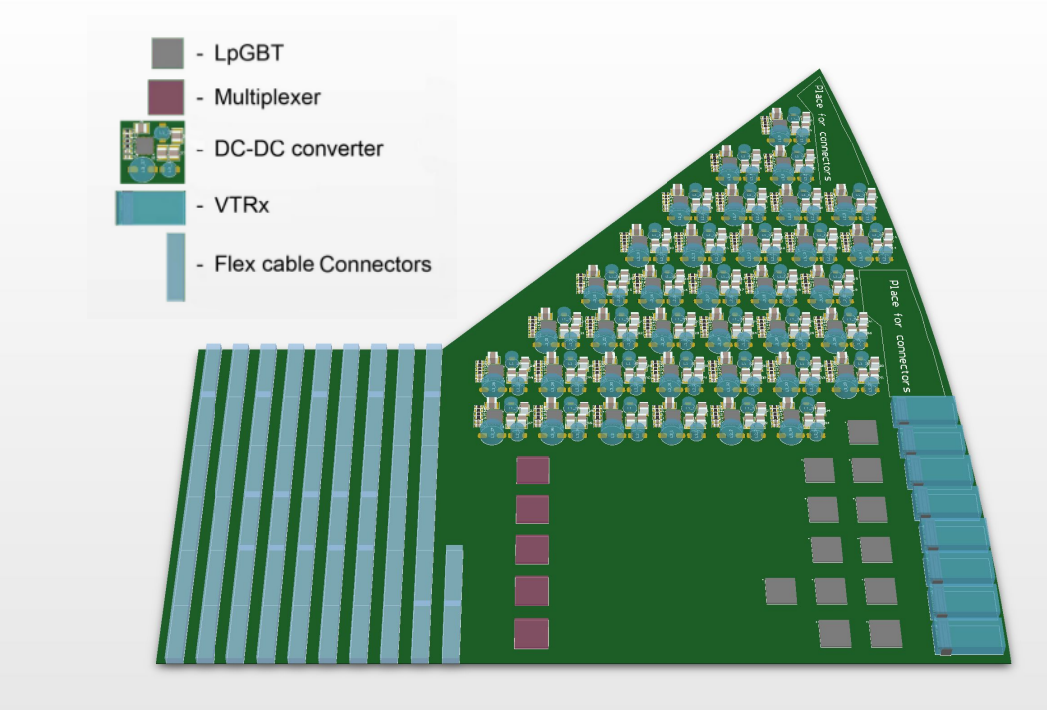


Figure 9.5: Conceptual design of a PEB 0 for the front layer. The left blue rectangle are the flex connectors (56). The pink squares are the analog multiplexer ASICs (5). The grey squares are the lpGBT used for the luminosity and data read-out. The right blue rectangles are the VL+ components (8). Finally at the top are located the 36 DC-DC converters.

1.6 kW per double sided layer. Including an estimated power consumption of 600 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 endcap.

10 DAQ and Luminosity Measurement

10.1 DAQ interface

The HGTD data acquisition system will be embedded in the ATLAS TDAQ common readout [31]. The proposed HGTD architecture is shown in Fig. 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 Chap. 9.

The interface between on-detector and off-detector electronics is performed via optical link 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 ToT and ToA information per triggered event and the luminosity stream that contains bunch-by-bunch hit information for luminosity measurements. 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 send hit information through the uplink and will be described in Sec. 10.3.7.

10.1.1 Off-detector electronics

The off-detector electronics is based on the general-purpose FELIX system [57], which is the main interface between the off-detector back-end and the on-detector electronics. The proposed back-end architecture is shown in Fig. 10.2. FELIX receives event data from the on-detector electronics and transmit 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.

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 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

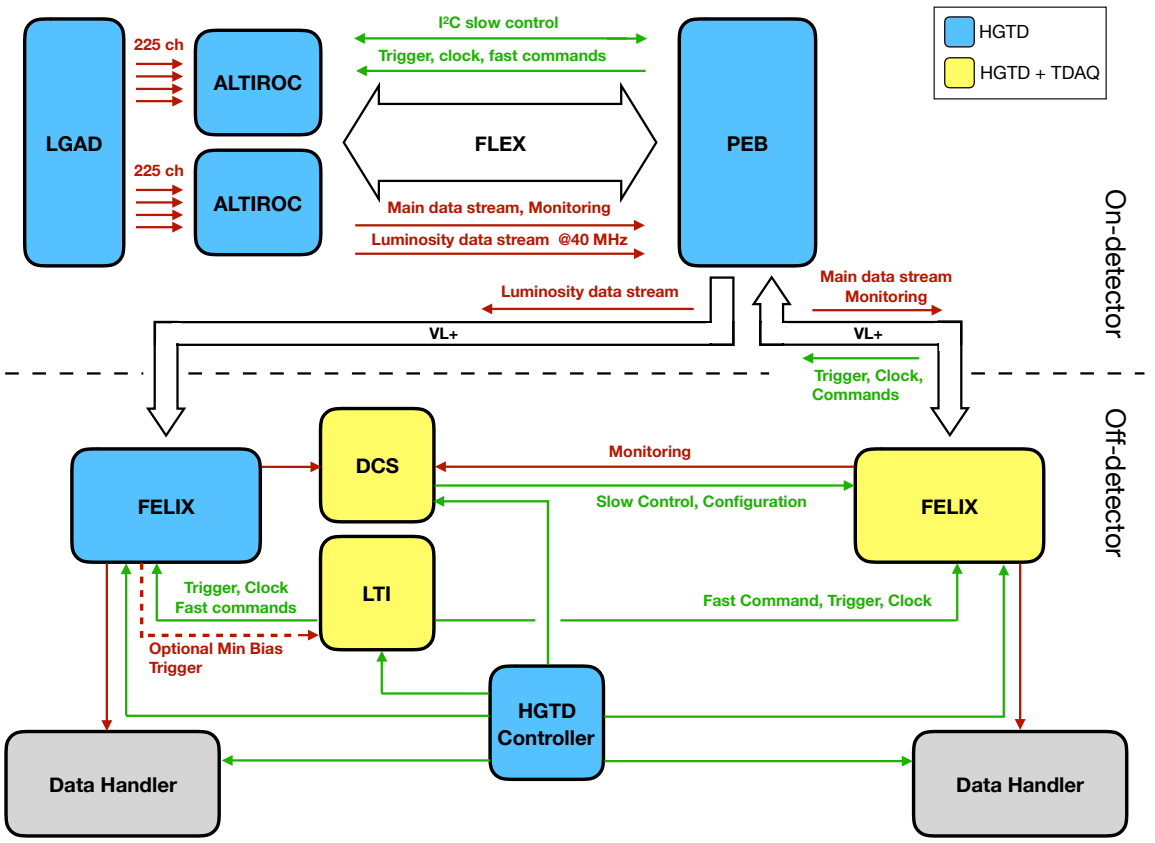


Figure 10.1: Data transmission paths among the ALTIROC, Peripheral Electronics Board (PEB), and DAQ components for hit data, luminosity data, clock, fast commands, and DCS/slow controls.

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 a prescaled sample of the events will be send to the monitoring unit.

The Data Handler will recive data from FELIX via multi-gibabit network. It will decode HGTD specific information providing event building and monitoring within a common TDAQ infrastructure. Afterwards, the data will then be send to the Dataflow system for further processing by the Event Filter. 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

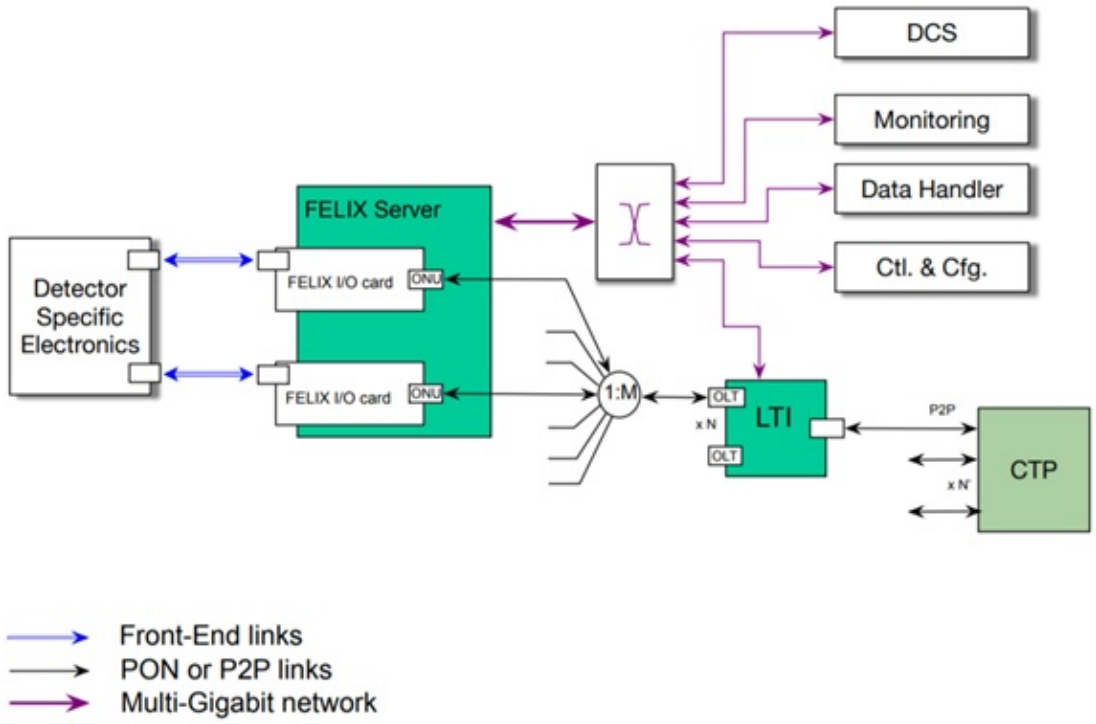


Figure 10.2: Proposed off-detector back-end architecture for Phase II. Plot taken from [31].

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. It will consist on different nested loop with a specific module configuration followed by a charge injection 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 in histrograms for a further analysis and may be used as input inside the nested loop for tuning purposes. The implementation of an histogramming unit inside FELIX in order to speed up the calibration procedure will be studied.

On the other hand, a precise timing of the modules is critical for operation, for this purpose a dedicated timing calibration will be performed. In a first stage the detector will be timed using standard calibration proceduce, it will consist on different charge injections while looping over coarse and fine delays DAC values of the TDC. Afterwards, 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 single colliding bunches (similar to 3b fills during Run2 used for Pixel and SCT timing scans), alternatively the same operation could be done during standard fills with a dedicated trigger stream with single colliding bunches. During these fills different delays values of the TDC will be scanned. 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 of the detector, dynamic and static contributions to the clock has to be taken into account and will be studied 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 30 ps per hit, as described in Chap. 5. The contributions to the time resolution from the on-detector electronics (UX15) and from the clock distribution (USA15) has to be smaller than this. For instance, the clock dispersion for HGTD should be less than 10 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 uneven 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 flowing through the FELIX for each trigger accept.

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 Fig. 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 after 250 ms of data collection, the t_0 can be measured with a precision of 5 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 Tab. 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 Fig. 10.1 and described in Sec. 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 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 [58] 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 Fig. 10.3.

4. FELIX: the clock jitter from the FELIX system will depend on the final chips used and bandwidth filtering applied, as studied in [59]. 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 [60], 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.

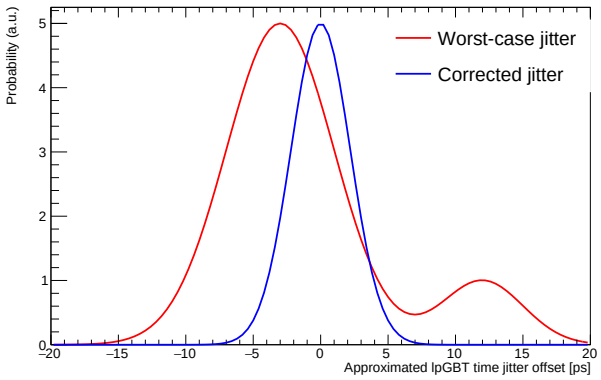


Figure 10.3: Timing jitter distribution assumed for the lpGBT in the corrected (blue) and uncorrected (red) scenarios. These distributions are approximations of the timing jitter expected in the lpGBT.

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 average timing can be computed online in the FELIX in order to collect as much data as possible, and then applied offline.

The hit time distribution before and after the timing correction is shown in Fig. 10.4. Fig. 10.5 shows the timing performance as a function of the integration time and the variation period

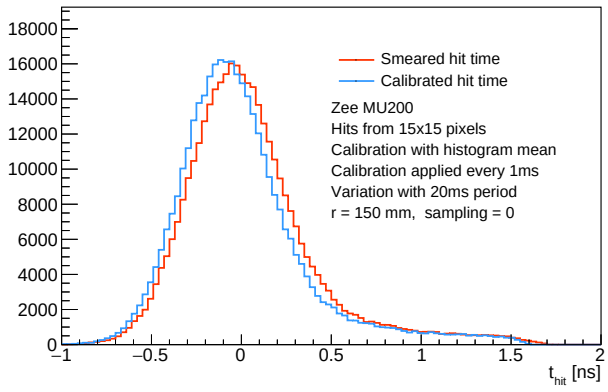


Figure 10.4: Hit time distribution before (red) and after (blue) the timing correction procedure. Non-Gaussian tails arise from late particles, backscatter, and other effects. The hit time distribution is obtained from the HGTD simulation described in Sec. 3.1.

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 10 ps can be reached, in accord with the specifications outlined in Sec. 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. 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 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

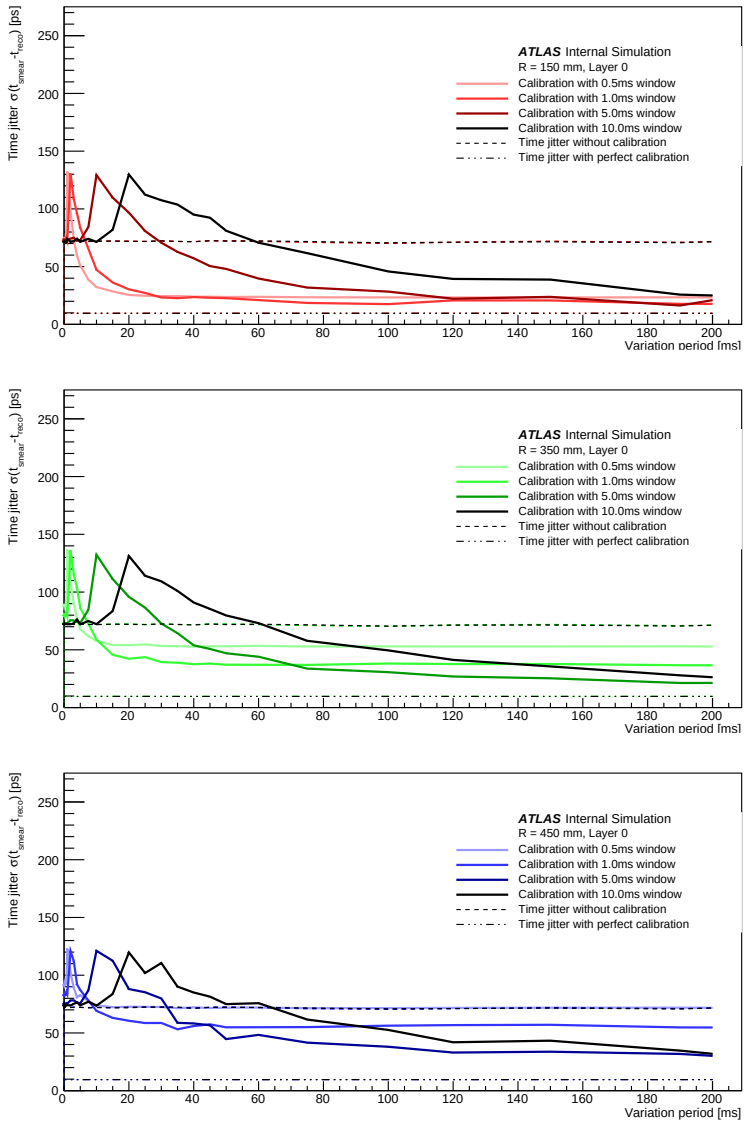


Figure 10.5: Hit time resolution $t_{\text{smear}} - t_{\text{reco}}$ after the timing correction procedure as a function of the variation period, and for several different choices of calibration window time, shown for $r = 150\,\text{mm}$, $R = 350\,\text{mm}$, and $r = 450\,\text{mm}$. t_{reco} is 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.

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 Sec. 3.3.2.

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 [17] or the noise-corrected number of clusters in the pixel detector [61]. In the early years of LHC operation, many luminometers used the so-called *event-counting* method [62], also known as *zero counting*, which exploits Poisson statistics to infer the pile-up 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 [18] was forced to exploit its 16-channel granularity to switch from event counting to hit counting. This latter method [62] 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 pile-up 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 [17] 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 [63], typically result in a pile-up 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 more than six 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 an $\langle\mu\rangle$ 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 [17], which proved essential to control the dominant luminosity uncertainties in both LHC Runs 1 and 2, can only be used offline as 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 Sec. 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 number of interactions (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 Sec. 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 Sec. 10.3.6), the HGTD will provide both online and offline unbiased 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 $470\,\text{mm} < r < 640\,\text{mm}$ (equivalent to $2.4 < |\eta| < 3.5$) 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 Sec. 10.3.3 and Sec. 10.3.5.

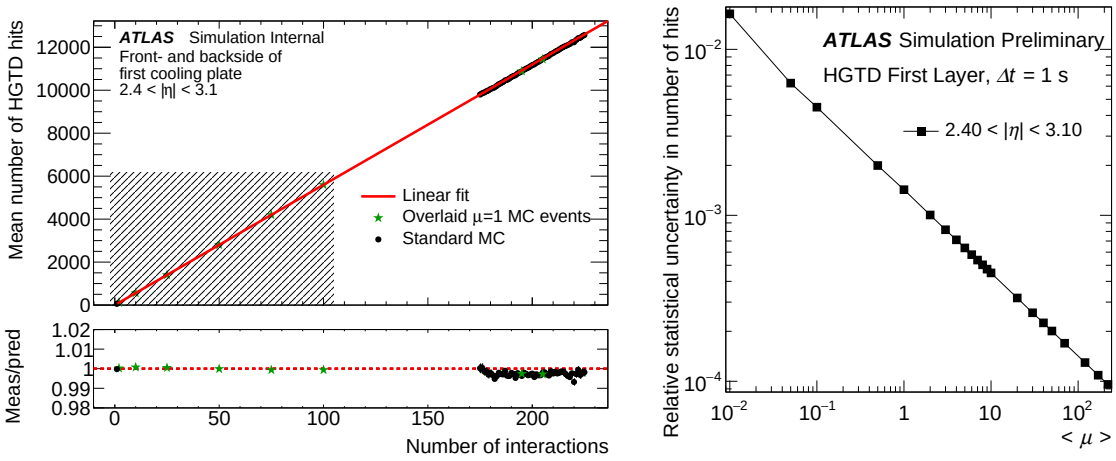
10.3.2 Linearity of the luminosity determination

For the $|\eta|$ range proposed above, the average number of hits per double-sided layer and per inelastic pp collision is 44.6, and approximately 7% of these collisions result in 0 hits. Fig. 10.6(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 μ of 1 and around 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 samples 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 fit to the points in the hatched region at low and intermediate μ values is extrapolated to the $\mu \sim 200$ region where its prediction can be compared to the hit multiplicities extracted from fully simulated high–pile-up samples. The small discrepancy between the extrapolated linear fit and the simulated points in the bottom left frame around $\mu \approx 200$ is attributed to multiple particles hitting the same pad; it could easily be corrected by applying the Poisson hit-counting formalism. Also discuss doing a correction for double-hits, and show result for that. Describe plot 10.5b, and what it implies for the precision that is achieved for the vdM scan.

10.3.3 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



(a) Linearity of $\langle n_{\text{hits}} \rangle$ as a function of μ.

(b) Statistical uncertainty versus $\langle \mu \rangle$

Figure 10.6: Left: mean number of HGTD hits per bunch crossing as a function of the pile-up parameter μ. Be more rigorous about calling this "number of interactions"?The straight line is a linear fit to the low pile-up points, and is compared to the simulated data in the bottom panel (see text). Right: pile-up dependence of the statistical uncertainty affecting the simulated hit counts shown in the left panel. Both of these plots need to be updated using the new samples. Add that the right plot is per BCID.

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 phase-1 luminometers, but also on HGTD-specific features: on the incoming-beam side, not only should the shielding provided by the endcap 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 wrt the collision products traveling back 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). Discuss here (or elsewhere) the possibility to flag according to measurements of the sideband rates.

As detailed in Ref. [62], 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. Add that for CMS also electronic signal residuals? For a given bunch pattern, the afterglow level is observed to be proportional to the luminosity in the colliding-bunch slots. Its magnitude relative to the luminosity signal, and its time structure, both depend on the spectral sensitivity of the luminometer considered (and therefore on the detector technology it uses), 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 that is intrinsic to the pixel-cluster counting technique. As described in Sec. 6.1, and illustrated in Fig. 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.

10.3.4 Statistical precision of the luminosity determination

Should come before the previous section. Add the vdM discussion. 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) and averaging the number of hits, the statistical precision achieved in each individual BCID during 1 s of LHC running is emulated. Fig. 10.6(b) shows the relative uncertainty expected from statistical fluctuations as a function of $\langle\mu\rangle$ using this method. The coverage of $|\eta| < 3.5$ presented here gives a statistical uncertainty of 0.14% at $\langle\mu\rangle = 1$ and 1.6% 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. Have the plot extend to even lower mu if possible.

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 Ref. [17]; the sources and the magnitude of the luminosity uncertainties in LHC Run 2 at $\sqrt{s} = 13$ TeV are comparable. 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, inner-tracker based luminosity measurements, such as track- or pixel-cluster-counting, were significantly more sensitive, typically because of unaccounted-for changes in either tracker conditions (noisy or misbehaving modules), or in effective coverage (disabled modules). Because the luminosity, and therefore the pile-up 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, especially 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 *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 ten 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 inner-tracker 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 pile-up dependence of the tracking efficiency; or, in randomly triggered readouts of 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, 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 Fig. 10.7. Each region has sufficient statistical sensitivity to determine the luminosity independently of the other regions. Check specifically if we can calibrate a single region in the vdM scan. Also maybe revise the radial divisions, at least look at how it can be made with the new layout. 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 the luminosity back-end electronics firmware, so that a different optimisation than the one described here can be accommodated.

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 independent checks based on several luminometers using different technologies are essential to controlling the systematic uncertainties to the level suggested 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 no μ-dependent corrections are expected to be necessary (except possibly at the very highest bunch

luminosities expected at HL-LHC, where the well-established Poisson hit-counting formalism may have to be invoked);

- 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 Sec. 10.3.3, 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 Sec. 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 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.

Reading out every single bunch crossing avoids any potential trigger-induced bias, which can occur even with so-called "random" triggers; it also ensures that the maximum possible per-bunch statistics are available for 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 data passing physics triggers cannot be used. Such triggers normally require a lot of activity in the detector, and the presence of e.g. high momentum leptons or jets. They are typically sensitive to pile-up 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 (they proved noticeable in the 2017 track-counting based luminosity measurements), and presumably too complex to be carried out in the luminosity back-end electronics boards.

- 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 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 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 approximately 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 be unable to even approach the necessary rate capability. Triggering the HGTD readout during the vdM scan using some kind of independent track- or hit-multiplicity trigger is not an option, since the *absolute* efficiency of such a trigger cannot be determined to the necessary precision of a few per mille (except by comparison with another, already absolutely-calibrated luminosity monitor). 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, which relied on a random trigger, suffered from such low statistics during vdM scans that they required a couple of hours of data-taking at $\mu \sim 0.5$ during the vdM session, in order to be cross-calibrated to LUCID instead, thereby making their absolute calibration fully correlated with that of LUCID.

During routine physics running, the need – both online and offline – for a 40 MHz readout of the luminosity data is fundamentally related to the intrinsic variation of the instantaneous luminosity across the colliding-bunch pairs. These bunch-to-bunch variations are driven by fluctuations in both bunch intensity and emittance; during Run 2, they 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 ten 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 liquid–argon-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 confronting independent luminometers is a key ingredient to precision luminosity measurements.

Finally, the afterglow-subtraction capability detailed in Sec. 10.3.3 and the potential use of the occupancy information in the Level-0 trigger outlined in Sec. 10.3.10 are entirely dependent on the availability of dedicated occupancy data at 40 MHz.

10.3.7 Luminosity back-end electronics

Add something about the occupancy data flow? For every bunch crossing of the LHC, each ASIC in the region $2.4 < |\eta| < 3.5$ 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 bit 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 212 lpGBTs for each of the four disks of the HGTD, i.e. 848 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. Each FELIX card can take up to 24 input fibres at 10 Gbit s^{-1}. With an integer number of FELIX units per HGTD disk, this results in 20 FELIX units needed for the luminosity data (5 per double-sided layer).

Update this, after investigating the feasibility of storing per-ASIC information The luminosity back-end electronics aggregates the central time window data and the sideband data separately, for each of the 16 regions described in Fig. 10.7. This reduces the massive data rate from the ASICs to only two integers for each of the 16 regions, on each of the HGTD double-sided layers (128 integers in total), separately for each of the 3600 BCIDs. Use exact number. The FELIX will store these sums in registers in the FPGAs, and update them continuously with the new data for every bunch crossing. Likely true, but investigate possibility to transfer the data to the server processor. These sums are the raw data needed for determining the luminosity, which is only needed with a frequency of about once per second. Assuming that the luminosity data gets pushed out of the FELIX at a rate of 10 times per second, and using 64 bits to encode each of the integers, the total data rate out of the luminosity back-end

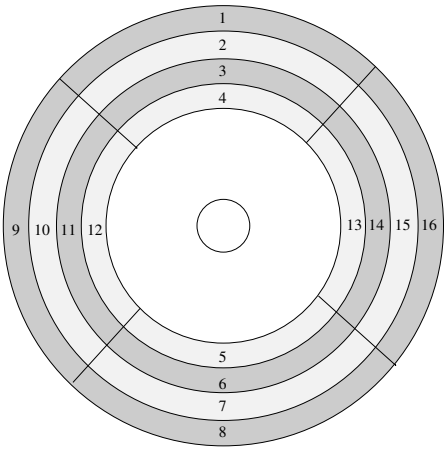


Figure 10.7: Sketch of the partitioning of the sensors into 16 regions for the luminosity determination. Each of the regions can be used to determine the luminosity independently of the others. Regions at different radius will be subject to different levels of radiation over time.

electronics is only $128 \times 64 \times 3600 = 29.49\,\text{Mbit}\,\text{s}^{-1}$, or $3.69\,\text{MB}\,\text{s}^{-1}$. 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 Data Handler. Discuss if something is done specifically for the offline luminosity.

10.3.8 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 of the 16 regions per disk 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. The per-event occupancy in the central time window provides no unique information over what can be calculated from the HGTD precision timing data, it would merely serve as validation of the luminosity and precision timing data. The information about the occupancy in the sideband time window however does provide unique information compared to the HGTD timing data, and could have use cases in e.g. searches for new, slow-moving particles. 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

occupancy counts for each of the 16 regions per disk encoded as 4-byte integers, in total 512 B per event. Calculate the implication of the amount of memory needed for the pipeline.

10.3.9 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 top and bottom luminosity regions (regions 1 and 8 in Fig. 10.7) 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 back-end electronics. 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.

10.3.10 Minimum-bias trigger at Level-0

The data made available at 40 MHz for the luminosity measurements can also be used by the L0 trigger 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 survive 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

This section covers the Detector Control System (DCS).

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, 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 allow to operate 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.

10.4.1 High Voltage

The high voltage supply system, that will be purchased, must 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 Sec. 8.1, in order to compensate for the damage due to radiation, the bias voltage will be raised. The read-out thresholds of the discriminators in the front-end electronics must also be adjusted accordingly. This operation will be performed by scripts

implemented in the DCS. 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, allowing to perform the necessary HV adjustments. The scripts will take the estimate into account and calculate the optimal bias voltage and the read-out thresholds according to an optimized algorithm.

10.4.2 Low Voltage

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 provide 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 Sec. 9.4. 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.

10.4.3 Temperatures

The temperatures of the sensor modules are monitored as voltages from temperature sensors, embedded in each ALTIROC front-end ASIC, via the same multiplexers and ADCs that used for the modules voltage monitoring. The temperature at the peripheral boards will be monitored through temperature sensors inside the lpGBTs. Information on temperatures inside the detector vessel, when the peripheral electronics is not powered, the DCS will be obtained from two sources: by means of temperature sensors located on the cooling plates, directly connected to off-detector ELMB++ units installed in the PP; and from the Interlock system, which will monitor the NTC sensors installed on the detector modules, as it is described further below in Interlock section.

10.4.4 Pressure and humidity

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. 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.

Radiation hardness is an issue for humidity sensors. Studies are needed to select appropriate radiation tolerant sensors. One option would be sensors based on optical fibres, that are being developed in ATLAS for ITk.

10.4.5 Configuring

Configuration of the front-end electronics and the lpGBT ASICs are in a similar manner controlled via commands, either I^2C-bus or direct, embedded in the DAQ data stream between FELIX and lpGBT boards. Sec. 6.8 and Sec. 9.4 give more details on the control and monitoring of the ASICs and peripheral electronics respectively.

10.4.6 DCS software

The HGTD DCS structure is shown in Fig. 10.8. 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 global ATLAS 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).

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 Fig. 10.9. 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 high voltage, low voltage and temperature, and then in a geographically level into the two vessels anddown to the individual modules.

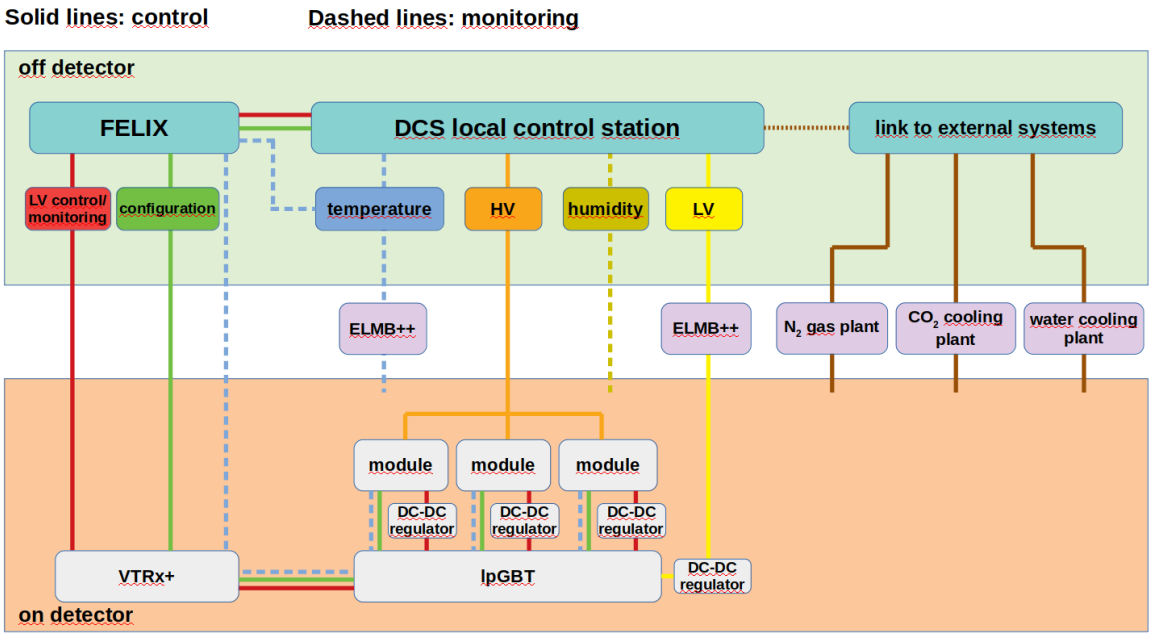


Figure 10.8: HGTD DCS layout

10.4.7 Interlock system

As detailed in **??**, 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.

HIS hardware will be implemented in an Interlock Matrix Crate (IMC) located in USA15. The Interlock crates are monitored by DCS...

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. Temperature information from NTC sensors will be provided to DCS using ELMB, also located in the IMC crate.

10.4.8 External systems

Infrastructure Beside the control and monitoring of the detector parameters, the DCS will help to protect the detector from various risks raised from infrastructure failures. The information on CO_2 cooling and N_2 gas plants state, as well as on other infrastructure systems, including beam status, will be shared by their control systems. This will allow DCS to monitor the infrastructure and bring the detector to a safe state if abnormal conditions occur. In the event of any severe infrastructure failure or safety alarm the ATLAS Detector

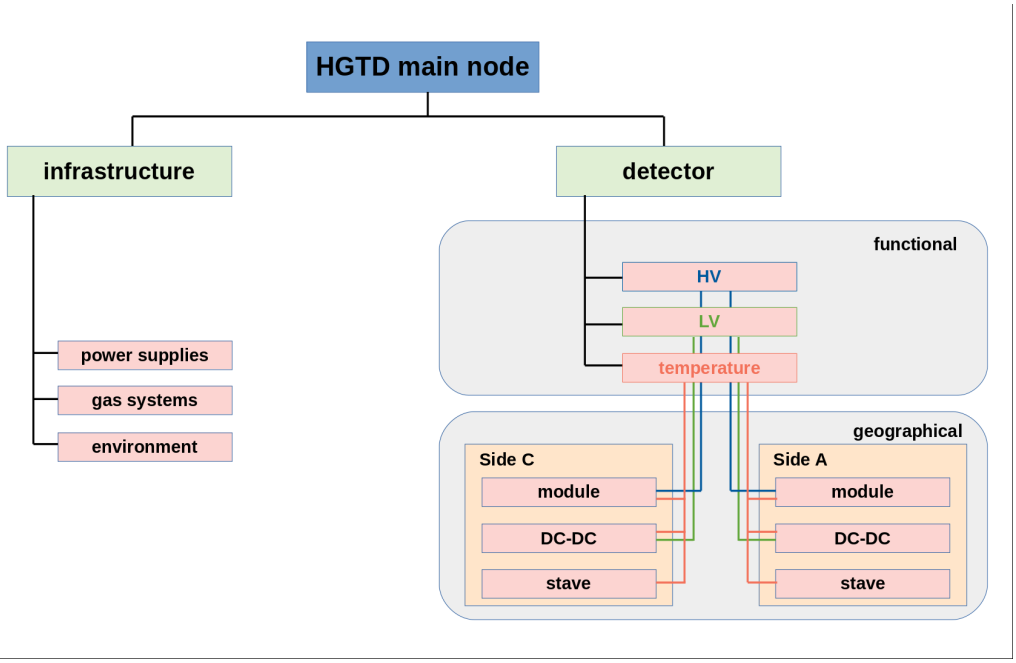


Figure 10.9: HGTD FSM layout

Safety System (DSS) will act on the detector equipment via HGTD Interlock System. 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.

Signals related to risks due to common infrastructure failures or safety issues.

Water leak-less cooling system for PP.

DSS: Various fault signals from CO_2 cooling plant are processed by the DSS safety or environmental alarms, such as smoke or flammable gas detection, magnet vacuum or cryogenics failures (risk from water due to condensation melting), ATLAS emergency stop, flooding signals corresponding to failures of common infrastructure, such as UPS power, rack cooling, N_2 gas system failures, will be available from DSS

(Un)stable beam conditions signal from the Beam Interface (for HV ramping).

11 Detector Mechanics

TODO: UPDATE TEXT , FIGURES FOR: 1) BASELINE WITHOUT ALUMINIUM INTERMEDIATE PLATE BETWEEN MODULES AND COOLING PLATE. 2) COOLING PIPES MADE OF TITANIUM 3) COOLING PLATE MADE OF ALUMINIUM 4) FEEDTROUGH OUTER VESSEL PART WITH NEW LAYOUT , RECALCULATE POWER AND THERMAL STUDIES WITH 400 MW/CM2 INSTEAD OF 350 MW/CM2, MAKE FIGURES WITH 3 RINGS BLUE COLOURS AND READOUT RAWS,.....

11.1 Engineering design overview

This chapter describes the global detector structure, the main mechanical sub-assemblies, as the hermetic vessel, front and back covers, inner and outer rings, the moderator, the support and cooling disks, the bolting and alignment device to LArg cryostat wall. The cooling system, common project with ITK and CMS, is also presented including cooling requirements and main components from the chiller up to detector cooling channels.

As presented in previous sections, the space allocated to the HGTD equipped vessel is 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 challenging 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 imbedded CO2 channels and manifolds, the peripheral electronics boards and feed-throughs.

In addition, 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 Fig. 2.4. They consist of a cylindrical hermetic cold vessel (front cover with heaters and back cover bolted to the inner and outer rings) that encapsulates two instrumented disks and an inner part of the neutrons moderator. Each instrumented disk (Fig. 11.1) represents a cooling support plate composed of two separate half disks with silicon modules installed on both sides, as shown in Fig. 2.9.

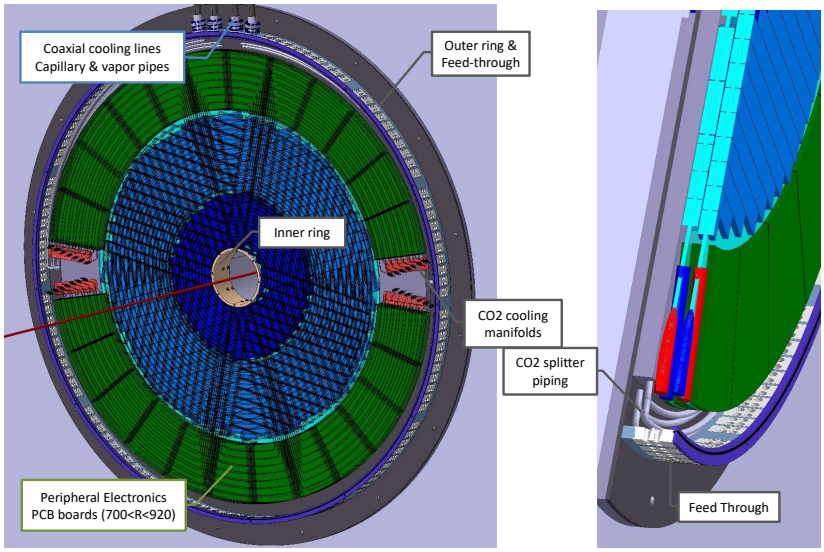


Figure 11.1: General view of the HGTD detector showing the silicon sensors inside the hermetic vessel. The green outer crown is the peripheral electronics limited by the outer ring which is holding tight electrical feed-through and cooling transfer lines. **UPDATE FIGURE WITH 3 RINGS BLUE COULOURS AND THE PEB REGION WITH ONLY 1 GREEN REGION AND NEW OUTER VESSEL FEEDTHROUGH.**

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 included, 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 excluding the moderator). The moderator is made up of two disks of different radii to provide more peripheral space inside the vessel. This space allows electrical services, feed-through, 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-plan of 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 Fig. 2.3. The overall dimensions are summarised in Tab. 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 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 [64]. CO_2 cooling is chosen because it makes significant mass savings inside the detector possible due to the use of tubes of smaller diameter then 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 evaporative 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 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.

11.2.1 Requirements

An operation temperature of $-35\,^{\circ}C$ must be maintained inside the HGTD vessel, in the level of cooling channels close to the modules, with a stability of a few degrees Celsius. As discussed in Chap. 5, the operating temperature must be kept as low as possible because, after irradiation, the leakage current of the sensors increases with temperature. In addition, the ASIC performance (S/N and jitter) will benefit from low temperatures, with observed improvements of up to 10% at $-30\,^{\circ}C$ compared to room temperature. These conditions will limit the heat dissipation and ensure good performance of the sensors and ASICs. The operating temperature of the peripheral on-detector electronics is flexible. It can be in the range of $-35\,^{\circ}C$ up to $20\,^{\circ}C$, 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. They will be used as pre-heaters to stabilise the cooling parameters before coolant reaches the modules.

Tab. 11.1 summarises the power consumption estimated for the various components of the detector. This defines a need for maximum cooling power of 38 kW in total (19 kW per end-cap).

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	4 mW cm^{-1}	1.8
HGTD vessel heaters	100 W m^{-2}	0.6
Total in active region		**22–28.0**
Pre-heaters (Perip. electr.)		8.8
Ambient pick-up		2.5
Total power dissipation		**33.3–39.3**

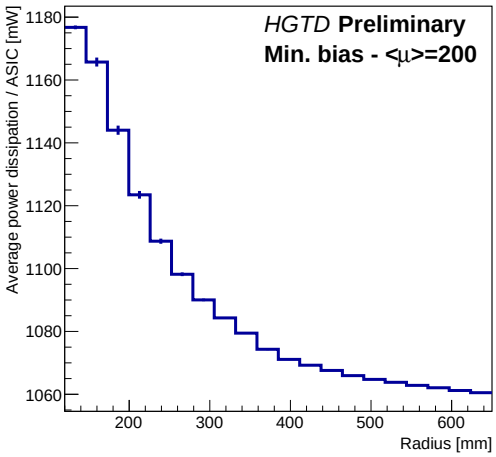
Table 11.1: Total power consumption estimates for the HGTD and breakdown for the various components, for a total number of 8032 sensors of 2×4 cm^2 each, 16064 ASICS of 2×2 cm^2, and 8032 flex cables of different lengths. (*) The 19.2 kW corresponds to 1.2 W (or 300 mW cm^{-2}) consumed by each ASIC 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 smaller since it decreases with increasing radius.

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 as a function of their radial position because the hit rate decreases at larger radius, as shown in Fig. 11.2. Taking this radial dependence into account, the total power consumed by the ASIC increases 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.

The power dissipated in the flex cables is expected to be 4 mW cm^{-1}, leading to less than 300 mW per flex cable for the longest flex cables of 75 cm and 1.8 kW in total for all the flex cables.

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. Assuming a 65% efficiency for the DC/DC converters, the peripheral electronics will dissipate an estimated 8.8 kW.

The total cooling power needed for the cooling station is 39.3 kW in total (19.7 kW per end-cap). Given the uncertaintes on current estimates of the power dissipation of some components, a cooling unit dedicated to HGTD of 50 kW will be constructed. A spare cooling station shared with ITk is also foreseen.



(a)

Figure 11.2: (a) Average power consumed per ASIC (in mW) as a function of the radius of the ASIC. REDO THIS FIGS WITH NEW READOUT ROW ORIENTATION BY CHRISTINA IF NOT POSSIBLE LEAVE ONLY RIGHT FIGURE THAT SHOULD NOT CHANGE.....

11.2.2 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 and a junction and distribution box to be located on the outer radius of the end-cap calorimeter on the HO side, close to the HGTD patch panel area, detailed in Sec. 12.1.3. One such box per end-cap will be used to disconnect/re-connect the rigid transfer lines for opening or closing ATLAS and to distribute the CO_2 flow from one big transfer line to four smaller proximity lines.

A permanent extension of transfer lines will be installed to allow the connection of the cooling station to the HGTD cooling box when the end-cap calorimeter is in the open position. This set-up will provide cooling during the yearly shutdowns and maintenance periods. While the opening and closing of the end-cap calorimeter is taking place, the cooling will be disconnected and will be reconnected after the movement is finished. During this period, the temperature inside the vessel could increase up to room temperature, since the N_2, blowing at 20 °C, will continue to flow at up to 750 l h^{-1}, improving the convection heat transfer of inner parts of the detector. The anti-condensation heaters on the front cover and feed-through crown should also be switched on during calorimeter end-cap movement phase. The time estimated to reach 20 °C from −35 °C of the HGTD cold mass (200 kg, covering mostly the on-detector system and moderator inner part (see Tab. 11.2), 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 200 W, mainly from the heaters, and the equivalent specific heat in the range

of $c = 750\,\mathrm{J\,kg^{-1}\,K^{-1}}$, the unit time is about 12 min per degree increase, for a total of 10–12 hours to reach room temperature.

Rigid proximity transfer lines are under development for Phase-II upgrade applications for ATLAS and CMS. These aim for a transfer capacity of about 5 kW per unit. The HGTD design places an inner hose, with inner diameter of 5 mm for the CO_2 liquid, inside a 16 mm mid-hose for the vapour return. This hose, made 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.

The on-detector cooling layout is illustrated in Fig. 11.1. The four tri-axial rigid lines, one for each half-disk cooling plate, enter the HGTD vessel at the top position. They are holding capillary lines of 0.75 mm in diameter and up to 5 m long, ended inside the hermetic vessel at the manifold r-phi 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 Fig. 11.3. The radial distance between the concentric pipes in the loops at $120\,\mathrm{mm} < r < 640\,\mathrm{mm}$ is 16 mm. This is the region covered by active modules placed on either side of cooling disk with overlap from 20% up to 70%. In the peripheral electronics area, at $r > 680\,\mathrm{mm}$, 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 on-detector cooling channels are designed to be made out of titanium pipes T40 grade 2 or equivalent, baseline titanium allow similar to ITK production program. The non-magnetic stainless steel 304L is an alternative material, due to its wide usage in this technical field of particle physics equipment and manufacturing expertise, such as bending and welding processes. All pipes and fittings outside the on-detector area should be standards stainless steel products. The cooling plant is protected against over-pressure with safety valves set to 130 bar. This value is used as maximum design pressure on the cooling loops. To ensure that the pipes can sustain such levels of CO_2 pressure, the 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 first prototype of the cooling loops has been manufactured from stainless steel 304L at the CERN workshop. This prototype corresponds to the inner zone of the cooling half-disk with the radial spacing of 16 mm identical to the final version). It has been successfully tested up to 165 bar at CERN proof pressure facility. The thermal tests will be undertaken

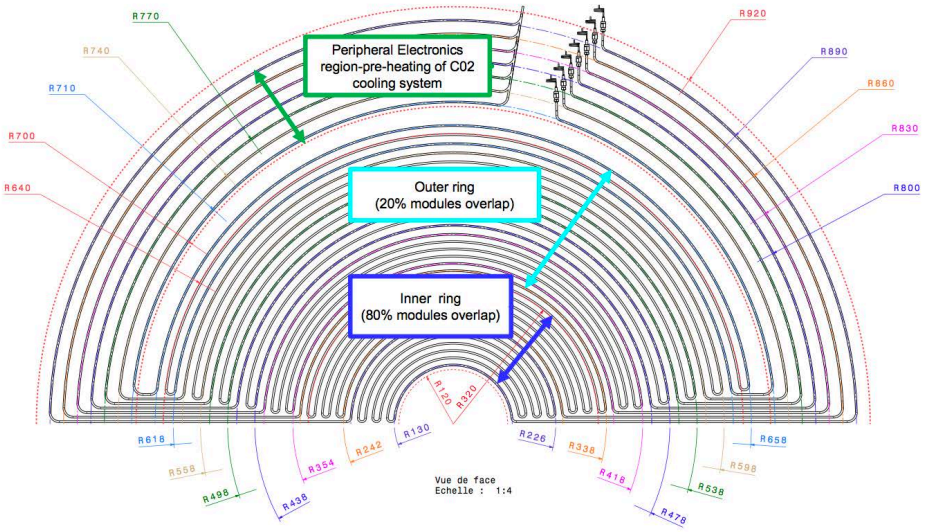


Figure 11.3: Layout of the cooling loops on a half disk support. The maximum power dissipation at the modules is 2.8 kW per half-disk, whichloads the cooling pipes by about 90 W m^{-1}. The pitch of the inner, middle, and outer loops is 16, 20, and 30 mm, respectively. CHANGE FIGURE WITH LAST THAT HAS 1 MORE LOOP

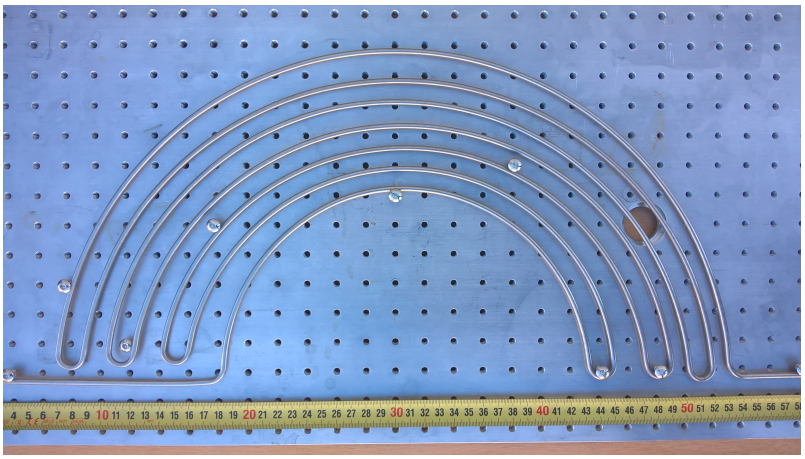


Figure 11.4: Cooling loop prototype corresponding to the inner part of the half-disk support.

with Baby-Demo setup before being integrated into the sandwich structure of the cooling support (see Fig. 11.4).

Cooling pipes made of Aluminum may 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. In addition, Aluminum is less activated by radiation and therefore may allow greater access to the detector. This is important for replacing the inner ring

midway through the HL-LHC and for maintenance during long shutdowns.

The half disks with embedded cooling loops are the main support structure for the instrumented active layers, as described in Sec. 11.5

Given the challenging performance 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 submitted to several thermal cycles to study 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.

11.2.3 Cooling plant demonstrator

One important milestone for the cooling development is the proof that the CO_2 evaporation temperature of $-35\,^{\circ}C$ can be achieved at the local HGTD support disks with realistic transfer lines and coolant distribution. Because of the crucial 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 is being tested (Fig. 11.5). This facility is installed in Building 180, next to the mock-up of ATLAS calorimeter, and will be used for tests of prototypes of ITk and HGTD cooling components with a real-scale geometry.

Figure 11.5: CO_2 cooling plant demonstrator located in Building 180 at CERN.

This demonstrator will operate at low temperature with a 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 [64]. As an example, Fig. 11.6 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 pixels staves, the cooling plant temperature needs to deliver −45 °C to account for the estimated 5 °C lost in the distribution and transfer lines.

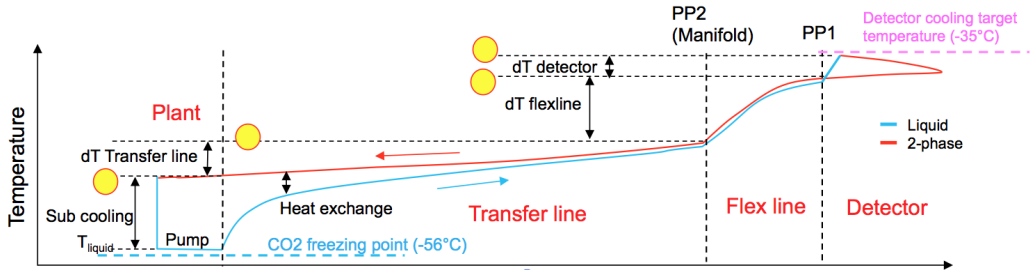


Figure 11.6: Typical temperature distribution between the CO_2 cooling plant and ITk loop [64].

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 at the Baby-Demo facility at CERN.

11.3 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 kg L^{-1}, similar to the one used in the present ATLAS detector. As seen in Fig. 2.4, the new moderator will be divided into two disks per end-cap, one inside and one outside the HGTD vessel.

The moderator on the outside is mechanically separated from the HGTD hermetic volume. It will be directly screwed to the LAr cryostat wall and will provide the necessary flat surface on which to install the HGTD and accessible bolting brackets. It has a thickness that varies along the radius, 10 mm in the region 180 mm $< r <$ 342 mm (covering the LAr calorimeter cryostat central flange and the bolting spots) and 20 mm elsewhere (140 mm $< r <$ 180 mm and 342 mm $< r <$ 1100 mm). The weight of this disk is 75 kg.

The part of the moderator to be placed inside the vessel has a thickness of 30 mm, a radial coverage of 120 mm $< r <$ 900 mm, and a weight of 75 kg. It provides appropriate support for the instrumented layers and, because it does not extended to radii higher than $r =$ 900 mm, it leaves enough free space for the cooling services as shown in Fig. 11.1 right.

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 is 40 mm in the region between 110 mm–342 mm and 20 mm for $r >$ 900 mm. During the maintenance, and when the replacement of the inner modules takes place at the surface, the two moderator disks may stay bolted in the LAr cryostat, together with the rear vessel cover.

11.4 Hermetic vessel

The hermetic vessel is the primary integration structure of the HGTD detector. It is constructed of four main components made of composite structures in carbon fiber (Fig. 2.4): the front and back covers, the inner ring and the outer ring (which will hold all the services feedthroughs), and includes the internal moderator. The vessel measures 1100 mm at the outer radius and 110 mm at the inner radius. The thicknesses of front and rear covers are 15 mm and 7 mm, respectively, and weigh 25 kg and 15 kg, respectively.

11.4.1 Requirements

The hermetic vessel provides a robust support structure to the detector disks in a cold and dry volume. All materials chosen must satisfy safety requirements related to the expected radiation levels, described in Sec. 2.4, and the temperature range. Including safety factors and assuming no replacement of components during the HL-LHC, the materials used should survive 8.3×10^{15} n_{eq} cm^{-2} and 7.5 MGy. 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 minimum coolant temperature, −35 °C, and the expected module interlock temperature, 40 °C, with a margin of 20 °C on both sides. This results a safe range from −55 to 60 °C. In addition, all mechanical components inside the vessel, including adhesives and bolting design, should withstand CTE (Coefficient of Thermal Expansion) mismatches over the temperature range specified above.

This lower temperature limit of −55 °C is also critical because it approaches the freezing point of the CO_2 coolant, which is −56.6 °C for the given operating pressure.

One of the requirements is to ensure the detector volume dry, keeping the dew point at about −60 °C or below, to avoid condensation on the detector components. This can be achieved by permanent flushing with dry N_2 with 1% over pressure above atmospheric reference. The

N_2 flow will renew gas in the vessel volume 3–5 times per hour (360 to 600 l min^{-1}). For this purpose, the HGTD vessel was designed to be as hermetic as possible.

Another requirement is to keep the temperature of the outer surface of the HGTD vessel safely above the cavern dew point (∼17°C). This will be done by placing flat heaters on the external face of HGTD front cover and near the service feedthroughs, as described further below.

11.4.2 Front cover and heaters

The front cover is designed as a sandwich structure, consisting of a honeycomb core placed between two thin Carbon Fibre Reinforced Panels (CFRP). 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 composite, as shown in Fig. 11.7. Considering the tightness and stiffness requirements of the front cover, the front cover was designed as a single piece. When replacing the inner part of the detector half way through the HL-LHC life time, the front cover can be easily taken away because the beam pipes will already be removed. In the opening scenario, when the beam pipe is in place, the front cover must be slid over the beam pipe. Such a solution requires the use of dedicated tooling to properly control the position of the cover with respect to the beam pipe. This complicates the opening procedure and the conditions of access to the internal components of HGTD. We are studying design options to split the front cover in two parts to allow removal or easier displacement of the cover during YETS maintenance.

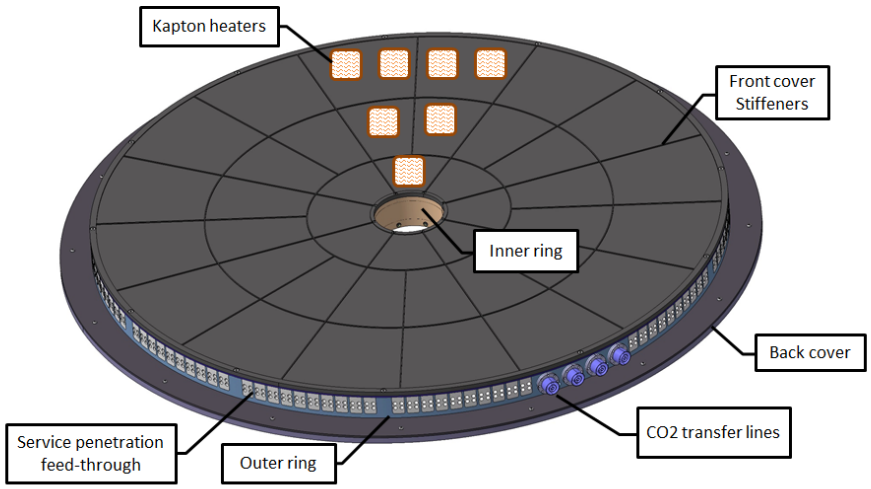


Figure 11.7: 3D view of the hermetic vessel with its main components. In particular, the front cover, equipped with kapton heaters uniformly distributed on the full surface, is visible. Not all heaters are shown in the picture. The heaters do not extend beyond the top of the three-millimetre stiffeners. CHANGE FEEDTROUGH WITH THE NEW LAYOUT AND FRONT COVER SPLITTED IN 2 PIECES

The HGTD inner volume will be cooled down to as low as −35 °C, therefore heaters will be required on the external face of the front cover to prevent condensation on the vessel outer surface. 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 cover in between the radial stiffeners. Their purpose is to ensure a minimal temperature of 20 °C outside the HGTD vessel, safely above the cavern condensation temperature of about 17 °C. The expected power of the heaters on the vessel front cover is 100 W m^{-2} This leads to a total contribution of approximately 300 W per end-cap expected from the heaters, which is included in the CO_2 cooling plant budget summarized in Tab. 11.1.

The temperature distribution expected on the HGTD front cover and on the LAr cryostat wall is shown in Fig. 11.8. In the temperature calculations, which were performed using Finite Element Analysis (FEA), the ambient temperature of 23 °C and heat exchange coefficient of 10 W m^{-2} K^{-1} were taken as input parameters. A temperature distribution in the range of 19 to 21 °C has been estimated outside the vessel and on the LAr cryostat wall.

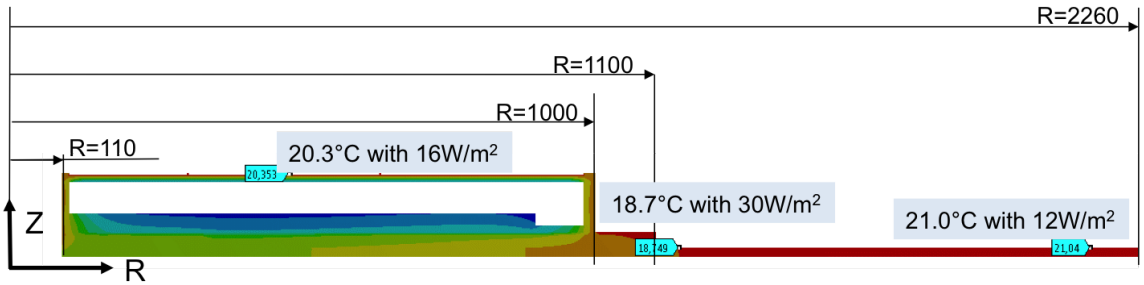


Figure 11.8: The calculated temperature distribution on the HGTD front cover and LAr front wall with heaters powered on. Transverse view of half the hermetic vessel and the cryostat wall is shown (0 mm $< r <$ 2260 mm). The total power consumption of heaters is 50 W on front cover, 150 W near the feedthroughs, and 100 W on the LAr cryostat wall.REPLACE FIGURE WITH THE NEW ONE WITH LAST CALCULATIONS

11.4.3 Back cover and interface with LAr cryostat

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. The alignment of the hermetic vessel on the calorimeter end-cap will be done with respect to the axis of the LAr warm tube, taking into account the existing moderator, as shown in Fig. 11.9. Potential conflicts with the cooling pipe, currently installed on the moderator and used for cooling of beam pipe during the bake-out procedure, require verification during the LS2 and may require some optimisation. To optimise the vessel installation procedure, the bolting/unbolting the back cover to the cryostat wall should be possible without opening the HGTD hermetic vessel.

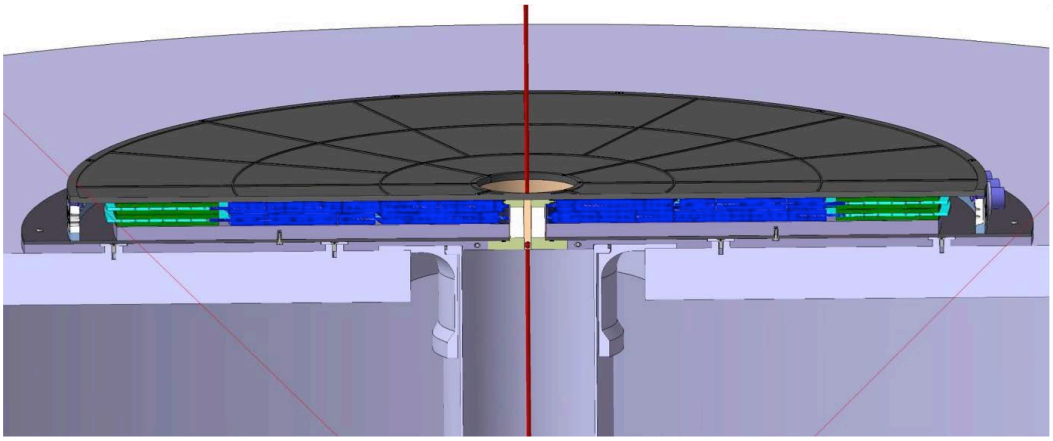


Figure 11.9: Transverse view (r,z) of the HGTD components and moderator bolted into the end-cap LAr calorimeter cryostat.

The stiffness of the vessel when mounted on the cryostat wall was studied using FEA. In this calculation, 10 mbar over-pressure have been applied, corresponding to dry nitrogen blowing inside the vessel to prevent any ambient humidity leak from outside. The results are presented in Fig. 11.10, showing a maximum deflection of 0.7 mm on the front cover. This is equivalent to a maximum stress (Von Mises) of 70 MPa, giving a comfortable safety margin compared to the over-pressure setting of 1020 mbar on the safety valves.

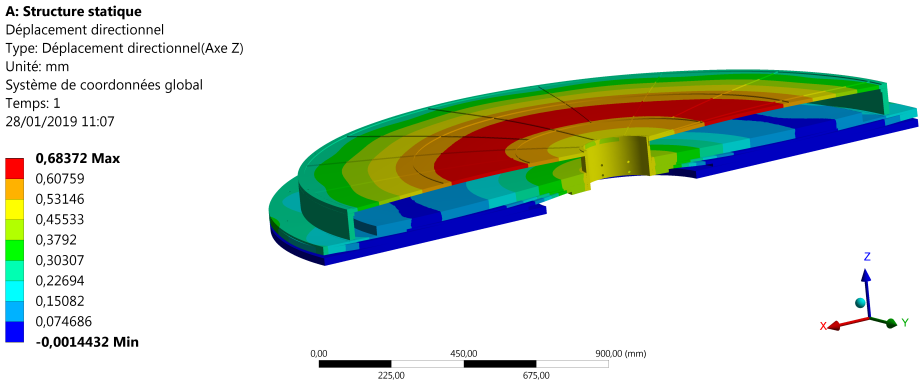


Figure 11.10: Finite Element Analysis (FEA) of the hermetic vessel with an over-pressure of 10 mbar. The red area corresponds to a maximum deflection of 0.7 mm in the vessel front cover. REPLACE WITH NEW FIGURE WITH FRONT COVER SPLITTED IN 2 PIECES

11.4.4 Inner ring design

The inner ring of the hermetic vessel borders on 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 from carbon fibre structures and the aerogel insulating layers. With the actual design, shown in Fig. 11.11, the inner ring is composed of a sandwich structure consisting of six millimetres of aerogel core enclosed between two thin sleeves made of carbon fibre reinforced panels. Further research on high performance materials, such as Kevlar panels and honeycombs, is being undertaken to address the specified stiffness, thermal protection, and radiation resistance, taking into account the low space allocated to the inner ring.

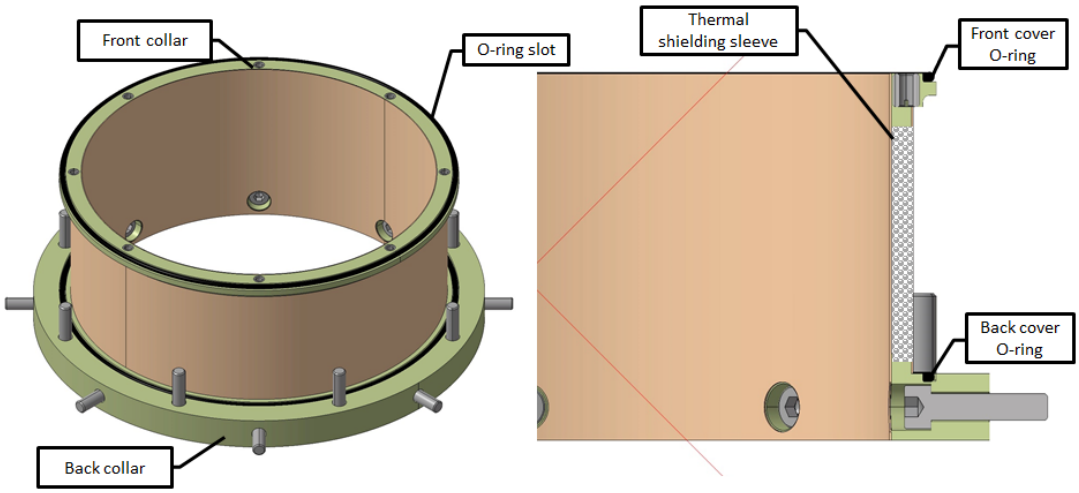


Figure 11.11: Central inner ring with its front and back collars. It is the central structure of the hermetic vessel, which ensures stiffness and tightness, thermal shielding, and HGTD positioning on the LAr cryostat.

To provide tightness as well as the alignment of the vessel with respect to ATLAS coordinate system, preciously-machined collars made of low thermal conductivity material, such as Glass Fibre Resin Epoxy (GFRE) or high performance PEEK polymer, will be installed on both extremities of the inner ring. Appropriate threaded inserts will be incorporated into the front collar to allow bolting of the front cover. The circular slot will hold the sealing O-ring made of PUR or EPDM material. The back collar will be bolted to the central flange of the moderator, providing the hermetic vessel alignment with respect to the central tube of the LAr cryostat.

11.4.5 Outer ring design

All available passages between detector volume (dry and cold) and the outside world run through the outer ring, which holds conductor cables, optical fibres, CO_2 cooling lines, and nitrogen blowing tubes. The outer ring structure, which is an assembly of several parts,

must be made of stiff material with low thermal conductivity. As for the inner ring collars, the main candidate materials are Glass Fibre Resin Epoxy (GFRE) and high performance polymer PEEK. Taking into account the large diameter of this part (up to 2000 mm), the manufacturing process is still under study to meet our specifications with a reasonable cost.

The maximum amount of service feed-through is equivalent to 480 cables with 12 mm in diameter. In addition, enough space has been also allocated to hold four CO_2 transfer lines with an outer envelope of 50 mm in diameter each, and few dry nitrogen holes (12 mm in diameter) as shown in Fig. 11.12. The feedthrough concept is based on a resin potting section on each cable and a detailed routing map. Given the limited space for services inside the vessel, and to guarantee complete sealing of the detector hermetic vessel, the feedthrough layout will be adopted to cable diameters and positions in ϕ, clamping the cables at their correct locations. The final potting will provide a tight seal to ensure dry detector volume and to minimize heat leaks. The potting mixture under study is PUR (Poly-Urethane Rubber already demonstrated in IBL program), which is easy to handle and is radiation resistant. From a maintenance standpoint, the cable clamps are removable and replaceable individually.

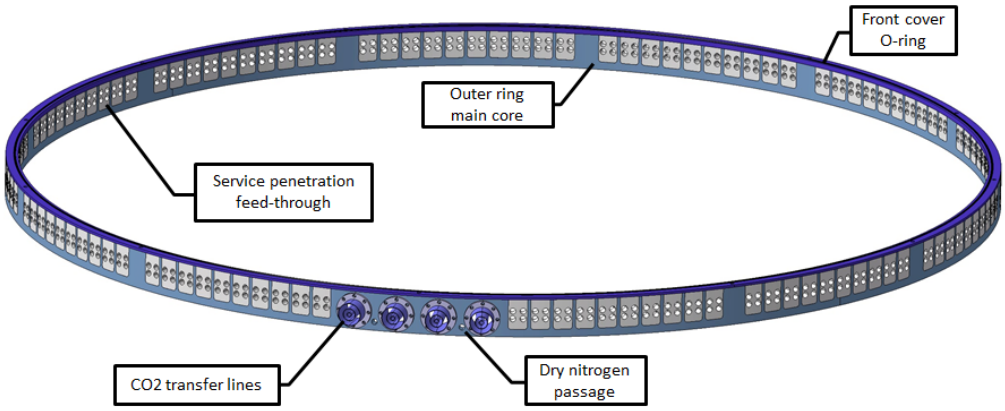


Figure 11.12: The outer ring assembly. The largest part of the hermetic vessel, with 2 m in diameter, it contains the service feedthroughs for cables, CO_2 transfer lines, and dry N_2 pipes. REPLACE WITH NEW DRAWING FROM CHINA

The CO_2 transfer lines will pass 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 common with CMS Phase-II HGCAL, which will transport a similar amount of heat (4.7 kW for CMS and 4.0 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.

11.5 Local supports and cooling disks

The design of local supports features four half disks per end-cap to provide the cooling and support on both sides for the module staves and peripheral electronics boards. Cooling piping with a semi-circular concentric pattern is embedded into sandwich structure of local supports to extract heat dissipation produced in the modules and peripheral electronics, as described in Sec. 11.2.1 and Sec. 11.2.2.

11.5.1 Geometry and design

The cooling support plates are composed of a carbon fibre structure with two high stiffness panels and a foam core inside. A good candidate for the foam is a composite pyrolytic graphite foam, similar to that planned for ITk. It has good thermal conductive characteristics and absorbs the different thermal expansion of the embedded stainless steel cooling pipes and carbon-fibre panels. As an alternative, a thermally conductive epoxy is also under study, due to its bonding, thermal performance, and reasonable cost. The high thermal conductivity of carbon fibre panels gives uniform temperature distribution over the large cooling disks.

All the support disk edges will be sealed by pultruded carbon fibre U-shaped crowns, which will be the direct interface with the HGTD global support. As for the hermetic vessel inner ring, glass fibre epoxy resin is also considered as a good candidate to seal the panels edge. In order to perform the long term stability and alignment in the ATLAS coordinate system, these edges will be directly connected to the inner ring at small side radius and the outer ring vs moderator at the large side radius.

If the intermediate plates/staves, shown in **??**, will be chosen to load the modules in the final detector, these will be made from a material with a high thermal conductivity (ASIC, PEEK graphite reinforced, carbon fibre low epoxy) and will be bolted to the cooling support disks with thermal grease. If instead the "pattern thicker plate" option is chosen, seen in **??**), the modules will be in direct contact with the cooling and support plate, insuring better thermal conductivity.

Given the challenging performance of the support plates, one full scale prototype will be produced with a half-disk composite plate, a few embedded cooling loops, and one stave of dummy modules as heaters. This prototype will be connected to the Baby-Demo facility at CERN and submitted to several thermal cycles to study thermo-mechanical behaviour, temperature distribution, CO_2 cooling parameters and glue layers integrity between modules and carbon fibre skins.

11.5.2 Thermal performance

Thermal Figure of Merit (TFM) is used as a baseline parameter of the thermal performance of the design. TFM is defined as the ratio of the temperature difference built up from the hottest point of the coolant to the module power per unit of area. It has units of $^{\circ}\mathrm{C\,cm^2\,W^{-1}}$. It can be thought as the thermal impedance of the on-detector support assembly. For a given heat flux, the goal is to have the TFM as low as possible to minimize the temperature gradient between the sensor and the cooling channel. By selecting stainless steel material, the heat convection between the coolant and the inner channel wall is optimal. On the other hand, by reducing the thermal bridges between the modules and the channel, the thermal conductivity is improved and the TFM is reduced in the same level.

A finite element analysis of the current CO_2 cooling design has been performed for one half disk face. The model considers a uniform power dissipation for all the modules of $350\,\mathrm{mW\,cm^{-2}}$ but takes into account the higher density of modules in the inner ring compared to the outer ring, as described in Fig. 11.3. It considers the input cooling temperature of $-35\,^{\circ}\mathrm{C}$ at the centre of each cooling loop with equivalent heat convection of $0.5\,\mathrm{W\,cm^{-2}\,K^{-1}}$. The calculation has been made in a static regime, with no external heat exchange.

The results are presented in Fig. 11.13. The temperature is uniform over the full surface of the cooling disk and close to $-32\,^{\circ}\mathrm{C}$. The TFM is near $17\,^{\circ}\mathrm{C\,cm^2\,W^{-1}}$ ($4\,^{\circ}\mathrm{C\,cm^2\,W^{-1}}$ as convective and $13\,^{\circ}\mathrm{C\,cm^2\,W^{-1}}$ as conductive).

The model used in these calculations will be improved by implementing a more realistic representation of the modules, the reduction of the power dissipation with radius, and the thermal contact between components. The results presented assumed a pessimistic estimation of $350\,\mathrm{mW\,cm^{-2}}$ for all the modules, the expected maximum power dissipation in the modules located at the inner radius. In addition, a study of thermal runway will be performed to cover for possible excesses in heat productions from the electronics or a lack of CO_2 liquid. Such studies will provide important input to the optimization of the cooling design.

11.6 Detector overall layout

An illustration of the HGTD detector components was shown in Fig. 2.4.

The front view of the two double-sided layers that will be placed on each end-cap are shown in Fig. 11.14. They have a rotation of 15° with respect to each other to facilitate the entrance of the cooling pipes inside the cooling disks.

A detail of the detector in the (r,z) direction, in the inner radius region close to the beam pipe, is shown in Fig. 11.15. It includes two cooling/disk supports where the double-sided

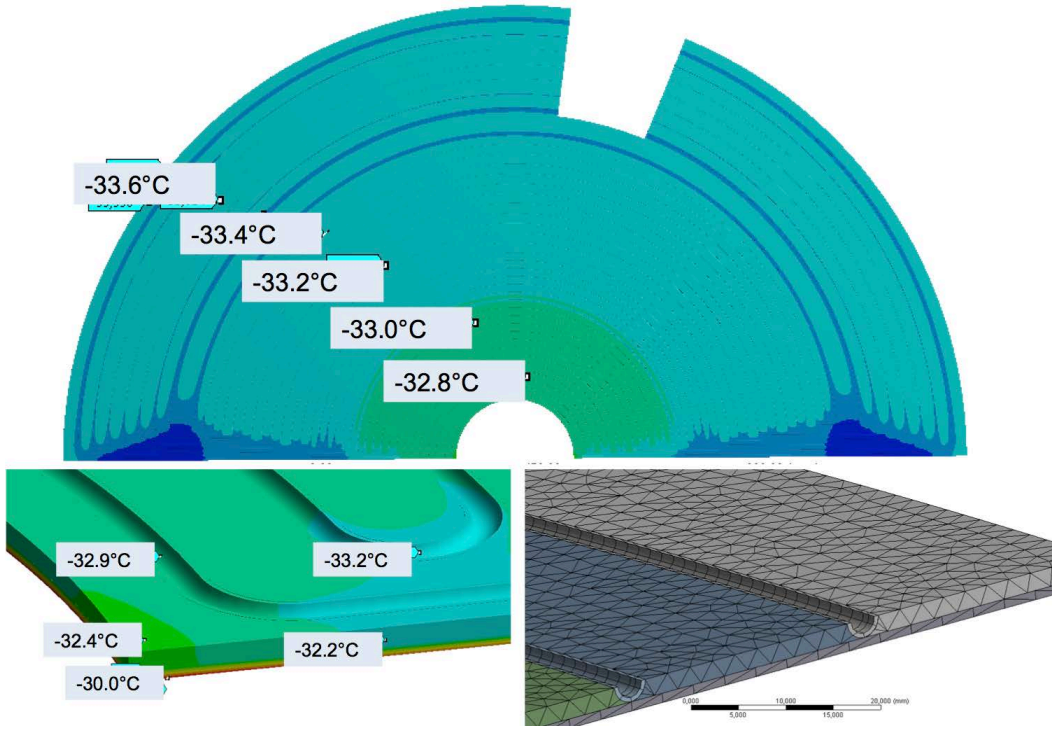


Figure 11.13: Expected temperature distribution over one half disk face. The bottom right figure shows FEM model used. REDO WITH 400 mW/cm2

layers of the detector are mounted, the front and back covers of the vessel and the inner and outer layers of moderator.

The full assembly, including 50 mm of moderator, will match the envelope of 125 mm in the z direction. A detailed breakdown of the (r,z) dimensions of the detector components is presented in Tab. 11.2, and also the materials and estimated weight of various components. The bottom of the table lists each component of a double-sided layer of detector modules mounted on the cooling support.

The measured thickness of the current prototype of the sensor-ALTIROC ASIC assembly is about 1 mm thick. This gives a comfortable margin with respect to the final envelope assembly protocol, with an expected thickness of module package (module and support) of 3.5 mm. Since the longest readout row will include 19 modules, 10 will be the maximum amount of stacked flex cables. Taking into account the estimated thickness of one flex cable of 0.3 mm, it gives the total thickness of flex cables stack of 3.0 mm per side. With an allocated envelop for flex cables of 4 mm, considering it together with 1 mm integration gap, it should be possible, though challenging, to fit all the components within the design envelope.

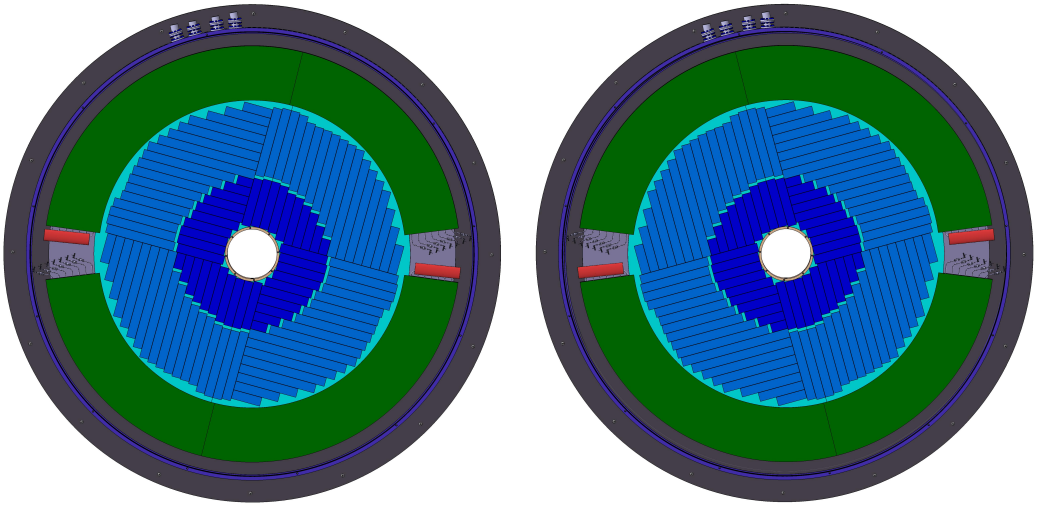


Figure 11.14: Front view of the two double sided layers that are placed on each HGTD end-cap. These two disks (right and left figures) have a rotation of 15° with respect to each other to facilitate cooling pipes interconnection with peripheral transfer lines. REPLACE WITH FIGS THAT HAVE THE READOUT RAWS

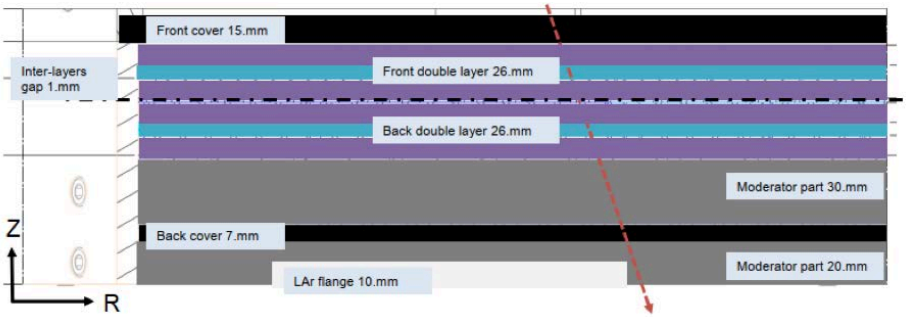


Figure 11.15: A detail of the detector in the (R,z) direction in the inner radius region close to the beam pipe, including two active double sided layers (installed on the cooling support plates), front and back covers and internal moderator. An extra 20 mm moderator is located outside the vessel in close contact with the end-cap cryostat.

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	120/980	60
Rear double side layer (2 half disks)	26.0	3459/3485	120/980	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/end-cap (w/ mod.)	**125.0**	**3420/3545**	**110/1100**	**350**
Total/end-cap (w/o ext. mod.)	**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 11.2: 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.

12 Detector Infrastructure

12.1 Services

12.1.1 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.

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-endcap 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, fiber 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 cables routed inside service caverns are not included in the table.

The services will include patch panels (PP-EC), which will be located on calorimeter extended barrels in several accessible places, close to the small wheel ($z \approx 6$ m). The main purpose of the PP-EC is to provide a disconnection point for the services, which cannot be accommodated by flexible chains due to lack of space and must be disconnected at ATLAS

opening. The PP-EC will also allow to realise remapping between connectors on back end electronics and on the detector. More details on PP-EC are given below in Sec. 12.1.2 and in Sec. 12.1.3.

An estimate of the required services per end-cap is summarised in Tab. 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 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 signals (clocks, trigger, etc), and 424? up-links for luminosity readout. Multi-mode optical fibers will be used for data transmission, they will be grouped in bundles containing 48 fibers connected to 2 MTP connectors, 24 fibers per connector. The fibers 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 fibers per bundle and per connector is optimised taking into account the routing of the fibers inside the HGTD vessel and the space available in flexible chains. Including spare fibers, a total of 40 bundles per end-cap are needed. Optical patch panes will be implemented in USA15 to organise the correct mapping for DAQ and luminosity readout.

- THIS PARAGRAPH TO BE REVISED ONCE THE GROUNDING SCHEME HAS BEEN DECIDED (COMMON OR INDIVIDUAL RETURN LINES). The baseline for the bias-voltage distribution is to provide individually adjustable voltage per each HGTD module. Consequently 3992 HV lines are needed per end-cap. Return lines of the HV channels, belonging to the same power supply module, will be merged together, requiring fewer wires per HV module. Assuming 48 channels per HV module, 4200 lines are needed. They will be grouped into 84 cables with an outer diameter of about 16 mm. Adding 4 spare cables, it gives a total of 88 cables per end-cap, to be installed between t he HV power supplies located in USA15 and the HGTD PP-EC. On PP-EC the HV lines will be re-mapped into 80 cables to match connectors on the peripheral electronic boards.

- The 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 the 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 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 for interlock for interlock and 4 cables for monitoring the DC-DC converters on PP-EC, all of them to be routed between service caverns and 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 Chap. 8 and Chap. 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 ELMB II, 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 is USA15 cavern.
- The heaters, similarly to the ones currently installed on the LAr cryostat flange, will be installed on HGTD vessel front cover and in the proximity of the feed-throughs. 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 N_2 gas circulation 1 inlet pipe and 1 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 next section.

12.1.2 Services layout

The overall HGTD service layout is illustrated in Fig. 12.1.

As it was described above, the detector vessel will be fixed on calorimeter end caps, which move at ATLAS opening. 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 at Phase-II ATLAS upgrade and some space became available for a fraction of the HGTD services. Also a new small flexible chain per end cap is considered to be installed for HGTD. The priority for installation in flexible chains will be given to the most critical regarding disconnection cables and pipes. The other services should be disconnected before 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 as well will be installed in the PP-EC

area in order to make LV cables as short as possible, that is necessary to minimise the power losses (and voltage drop) on cables.

The CO_2 transfer line will include rigid and flexible parts. The rigid line will be installed between CO_2 cooling plant, located in USA15 and manifold box located on voussoir platforms in ATLAS toroid area above calorimeter end caps. From the box two flexible lines, one inlet, another outlet, will be routed to the splitter box on the top of calorimeter end cap on IP side, close to HGTD. From the splitter box 4 smaller rigid lines will be 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 at short openings during YETS. However, on the platforms, there is no enough room to accommodate the flexible pipes long enough for long openings in LS periods, when calorimeter end caps can be moved about 12 meters. For such openings the flexible lines must be disconnected from the splitter box on the calorimeter and can be extended with additional flexible lines to supply the HGTD with CO_2 cooling in open position.

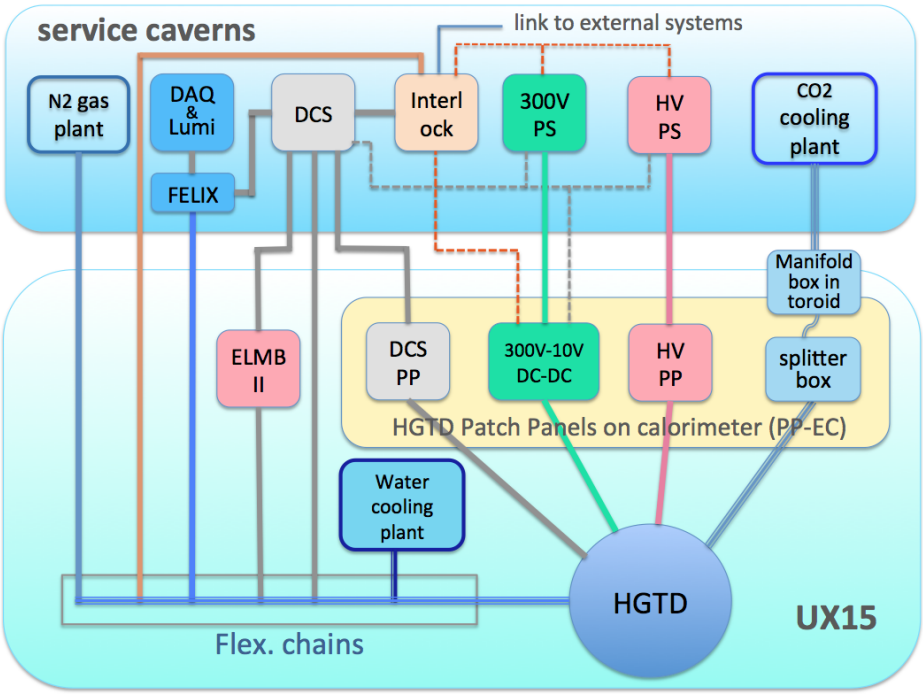


Figure 12.1: Overall HGTD services layout from the detector to USA15 or UX15. The optical fiber bundles, N_2 gas pipes, interlock and cooling temperature sensor cables, part of DCS cables and, still to be confirmed, the 300 V power supply cables are planned to be installed in flexible chains. The HV cables and rest of DCS cables will be routed through the patch panels, where they will have a disconnection point.

To allow commissioning of the detector after installation in the pit and for maintenance during shutdown periods, it should be possible to operate the HGTD when ATLAS is in

open configuration, which requires reconnecting the services in the open position. For that purpose, the extenders of cables and CO_2 cooling lines will be installed between respective positions of the patch panels in closed and open configurations. Most of these extensions must be permanently held in place, which will help minimise the time required to put the HGTD in working order after each opening.

12.1.3 Patch panels in PP-EC area

The positions of the patch panels (PP-EC) and DC-DC units on the calorimeter end caps will be selected by Technical Coordination, in four sectors in accessible areas to allow disconnection of services. It will also be possible to replace any faulty DC-DC converter at short access during the run. The probable patch panel locations are shown in Fig. 12.2.

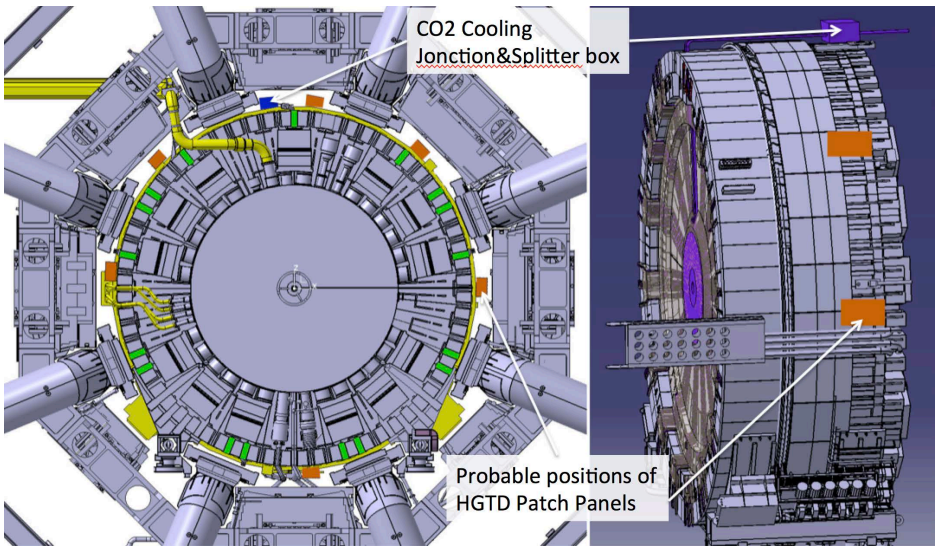


Figure 12.2: PICTURE AND CAPTION TO BE UPDATED. View of the probable locations of the HGTD patch panels (PP-EC) on the surface of the end-cap calorimeters. The DC-DC (300 V to 10 V) converters and cooling splitter box will also be located in this region. The exact position in z and ϕ is still to be allocated by Technical Coordination.

The value 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 Fig. 12.3, varying from 0.05 T up to 0.5 T. The power supplies should be placed in areas where the field is weaker, midway between two barrel toroids and as close as possible to the surface of the calorimeter. Radiation levels in these areas was estimated using FLUKA calculations, giving a maximum of 15 Gy and less than 1×10^{12} n_{eq} cm^{-2} at the outer radius of 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% efficiency, about 4 kW of cooling power is needed in all PP-EC locations, combined per

end cup. The existing in ATLAS leak-less water cooling systems have a sufficient capacity to supply the HGTD detector on both end caps. The dedicated connecting pipes and manifolds on the calorimeter will be required.

12.1.4 Services routing on the calorimeter front wall

As already mentioned, 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 task. This space is shared between ITK and HGTD services, and also the scintillator counters, belonging to the Tile calorimeter system, are installed here. In the present ATALS 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 Tilecal system and Technical Coordination that the scintillator counters will be installed on top of the HGTD services, while the last will be fixed on the wall of the calorimeter. Such layout will allow 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 scintillator tiles and fibers on front face, thin aluminium support plates will be fixed on the Tile calorimeter modules.

The envelop for HGTD services is shown in Fig. 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 space between LAR barrel crates and further between Tilecal barrel fingers, sharing the space with ITK services installed on the calorimeter barrel. However room in two gaps between LAr barrel crates on top cannot be used to root HGTD cables.

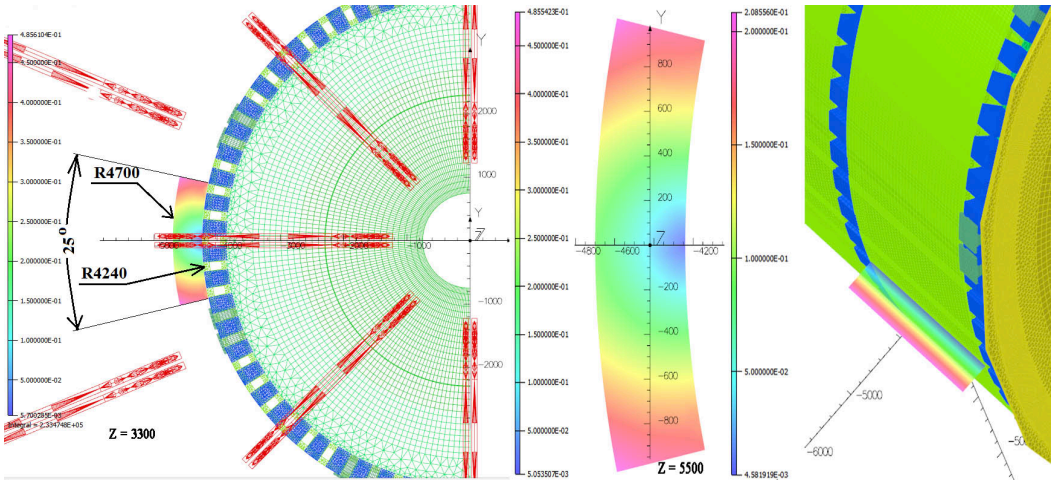


Figure 12.3: Magnetic field in the region of the HGTD PP-EC patch panels. FONT SIZE OF THE SCALE NUMBERING TO BE INCREASED.

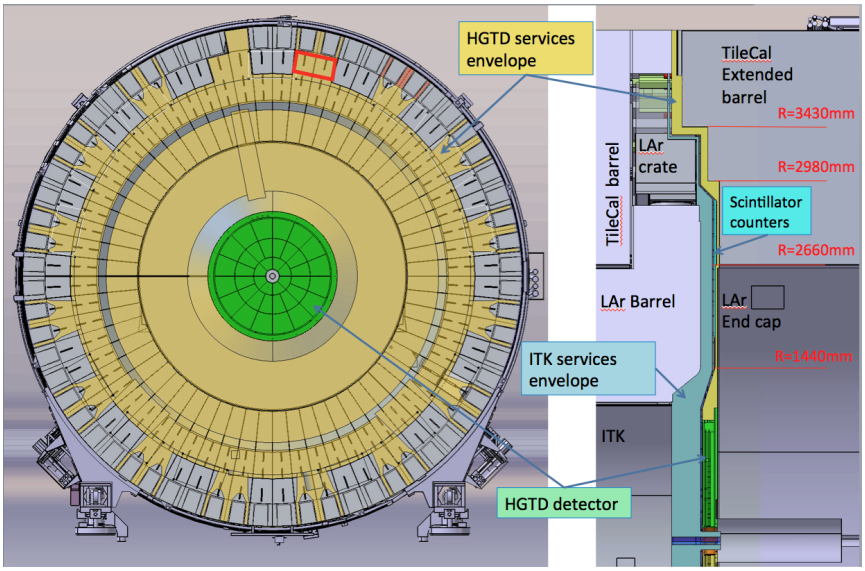


Figure 12.4: THE COLORS TO BE UPDATED. The envelope for HGTD services. On the left: front view of the calorimeter end cap on side A. The space available for HGTD services is shown with yellow color. With red rectangles is shown the area, where the room the HGTD services is very limited. On the right: 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.

One constrain comes from the requirements to keep free access to the end plates of 3 Tilecal modules, located at the top of the calorimeter, to allow remove these plates when accessing the electronics of the modules. Space in another gap is blocked by LAr HV filter box. The space in z available for HGTD services on LAr end cap cryostat wall at radius $>$1.4 m is only 17 mm. The exception will be a dedicated slot for four CO_2 cooling pipes, as described in Sec. 11.2.2.

The HGTD services routing on the calorimeter end cap front wall is shown in Fig. 12.5. The cables, connected to the outer ring of HGTD vessel in four layers, will pass 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 it was discussed above, the HGTD cables cannot be routed in two top gaps between LAr barrel crates. Due to that the cables from top section of HGTD deviate towards neighbour gaps.

On the surface of the calorimeter end cap the cables will be routed towards PP-EC located in 4 places around the calorimeter.

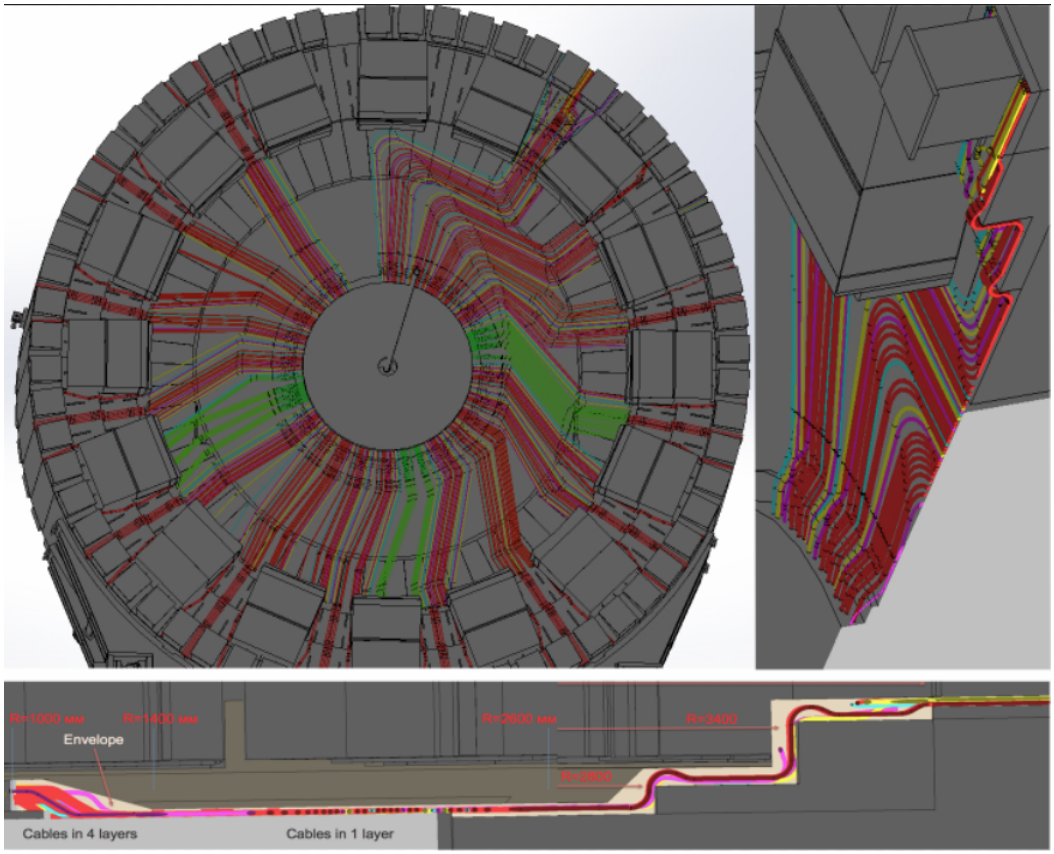


Figure 12.5: PICTURE AND CAPTION TO BE UPDATED. HGTD preliminary services routing on the calorimeter wall.

12.1.5 Services connection to outer ring and inside the vessel

The outer ring of HGTD vessel provides interface for all services of the detector. With such approach, the HGTD detector can be completely assembled and tested at surface and brought down to the pit for installation with closed vessel. Once the vessel fixed to the front wall of LAr cryostat, the pipes, cables and optical bundles will be connected to the detector. To realise such scenario, the cooling and gas pipe fittings, electrical and optical connectors will be embedded in outer ring, as it was discussed in (REF. TO THE OUTER RING SECTION OF CH 11 ???). The layout of outer ring is shown in Fig. 11.12.

The organisation of services inside the HGTD vessel is schematically shown in Fig. 12.6. The short pigtails, one per cable, will interconnect the cables and peripheral electronics boards (PEB). The optical bundles, connected to the outer ring, will be terminated with 24-fiber MPO connectors. The optical pigtails will be used to distribute these 24 fibers from each bundle to several VTRx+ optical link modules installed on the PEB. One bundle is required per PEB, including spare fibers. The optical pigtails will also contain spare fibers terminated

by connectors.

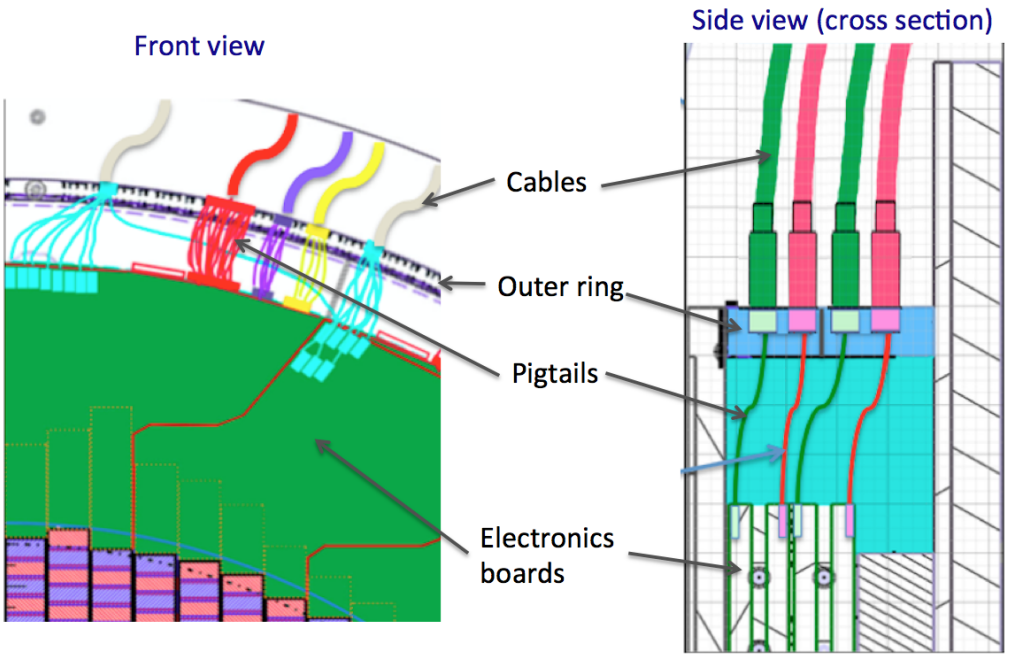


Figure 12.6: PICTURE TO BE UPDATED

12.1.6 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 vessel feed-throughs can be installed on 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 does not allow the installation of cables with an accuracy of several centimetres at connection points. As a consequence, some extra length should be allowed for each cable, which could then be accommodated on cable trays, however it is usually 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, the different installation scenarios are foreseen for different HGTD services, as described below. All cables, except optical bundles, must be tested before installation in the cavern.

Taking all that into account, the various installation scenarios are foreseen for different HGTD services, as it is described below.

The **Optical bundles** will be delivered with connectors attached, tested and protected at both ends 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 fibers between luminosity and data readout.

For **HV cables** two installation scenarios are considered. If space on 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 pair into the service cavern, the loop will be cut out to 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 connectors 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.

The **Proximity cables** listed in Tab. 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 the pilot cables between the mock-up of outer ring and the patch panel box, placed in their final positions.

The installation of the patch panels, services and respective connectivity will be done when access is permitted 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 2024 to April 2025 (???).

The installation of transfer lines for the CO_2 cooling system and pipes for the N_2 gas system is the responsibility of Technical Coordination, who will plan and organise it.

13 Detector Assembly, Installation, and Commissioning

13.1 Assembly and commissioning on surface

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 (Electro Static 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.

13.1.1 Half disks instrumentation

The assembly of the detector, e.g. the mounting of the module support frames and peripheral electronics boards on the cooling half-disk supports, connection of flex cables to the respective peripheral electronics boards, will be done with the participation of several collaborating institutes.

Each instrumented half disk will be a single piece of 30 kg with 12 cm inner radius and 98 cm outer radius. 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 Hermetic vessel assembly and test

Prior to any integration step, each mechanical component should be submitted for visual inspection and appropriate loading tests. In addition, a mechanics assembly test with all the detector components (4 assembled detector half disks, internal moderator, front-rear face of the vessel and vessel feed-throughs) should take place on surface to validate the overall mechanics envelope and the boundaries between the various components.

13.1.3 Quality Assurance

The instrumentation of each half disk face should be followed by several Quality Assurance steps, to be carefully defined, such as electric tests to insure proper connectivity between each flex cable and the peripheral boards, proper functioning of each of the 225 channels on a given module, resistivity tests in the flex cables, good thermal conductivity between each module and the cooling plate, etc.

A database will be used to record the status of each component at all assembly steps, in particular electronic and thermal parameters of the instrumented half disks. The aim is to have a full history tracking from the production process up to the final assembly and testing. Already existing databases (ATLAS and CMS Phase-I) would be adapted to avoid duplication of the software development effort. 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 to the files recording detailed technical parameters (row material, chemical composition, manufacturing process, testing), the database will allow an easy monitoring of the construction progress. At the completion of the detector installation in the experimental cavern, the database will evolve towards system configuration data, necessary to analyze the detector operation conditions and performance.

The detector assembly and QA on surface is expected to take 9 months per end-cap, between October 2023 and June 2024 for HGTD-A and between August 2024 and April 2025 for HGTD-C, finishing 4–12 months before the final HGTD installation in the ATLAS cavern, as detailed below.

13.2 Installation in the cavern and commissioning

13.2.1 Access 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 exact 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 other actions will be limited when taking data. Different kinds of stops are expected at HL-LHC, similar to the present LHC breaks:

- **Short access** for a few hours only, primarily for LHC machine interventions and usually announced on short notice. In these periods, electronic components located in the HGTD patch panel (PP-EC) areas can be accessed for simple interventions, for example for the replacement of 300 V-10 V DC-DC converter modules. Access to DCS equipment in racks in UX15 will also be available.

- **Technical Stop**, typically of one week duration, for maintenance of the LHC and 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 opened, keeping the beam pipe in place as illustrated in Fig. 13.1. The access to the HGTD is possible, only if radiation levels will allow for it (see Sec. 13.2.2). The distance between the barrel face and the HGTD face is typically 3.5 m. To get access to the detector the front vessel cover has to slide along the beam pipe, using dedicated tools, in order to not damage the beam pipe nor the HGTD vessel. A priori only the peripheral electronics boards of the front face of the front disks are easily accessible with minimal manipulations.

- **Long shutdown (LS)** of typically 2 years, for large upgrade or consolidation programs for Experiments and LHC. The ATLAS detector will be in large opening position, with the beam pipe removed, as shown in Fig. 13.2. The distance between the barrel calorimeter face and the HGTD face is maximum 12 m. After the LS3, where the HGTD should be installed, the next Long shutdowns should be used for deep maintenance of the detector. It will be possible to bring on surface the half disks of HGTD to replace malfunctioning components and to replace the inner ring at the middle life time of the HL-LHC.

The access and manipulation of the HGTD components sitting inside the vessel need to follow strict safety rules due to expected high radiation levels has described in the next section.

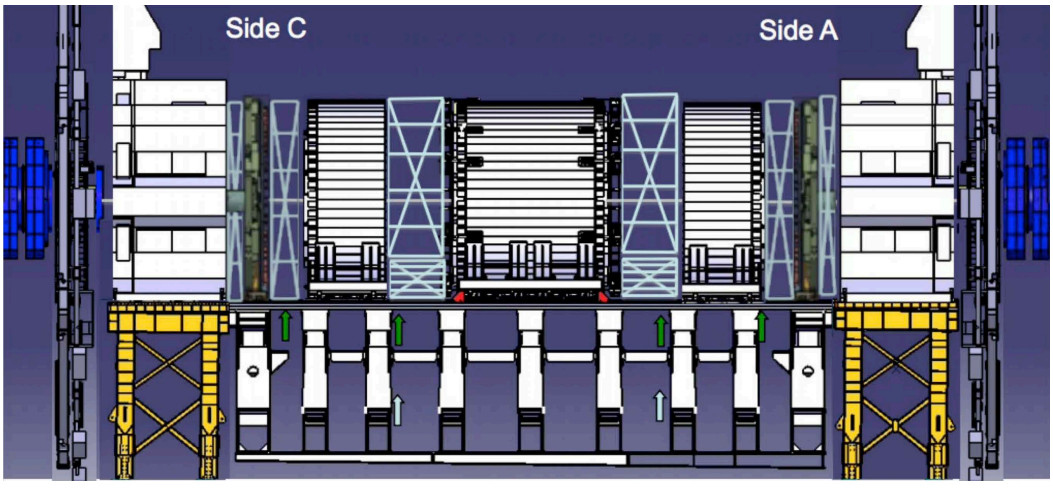


Figure 13.1: ATLAS in short opening configuration.

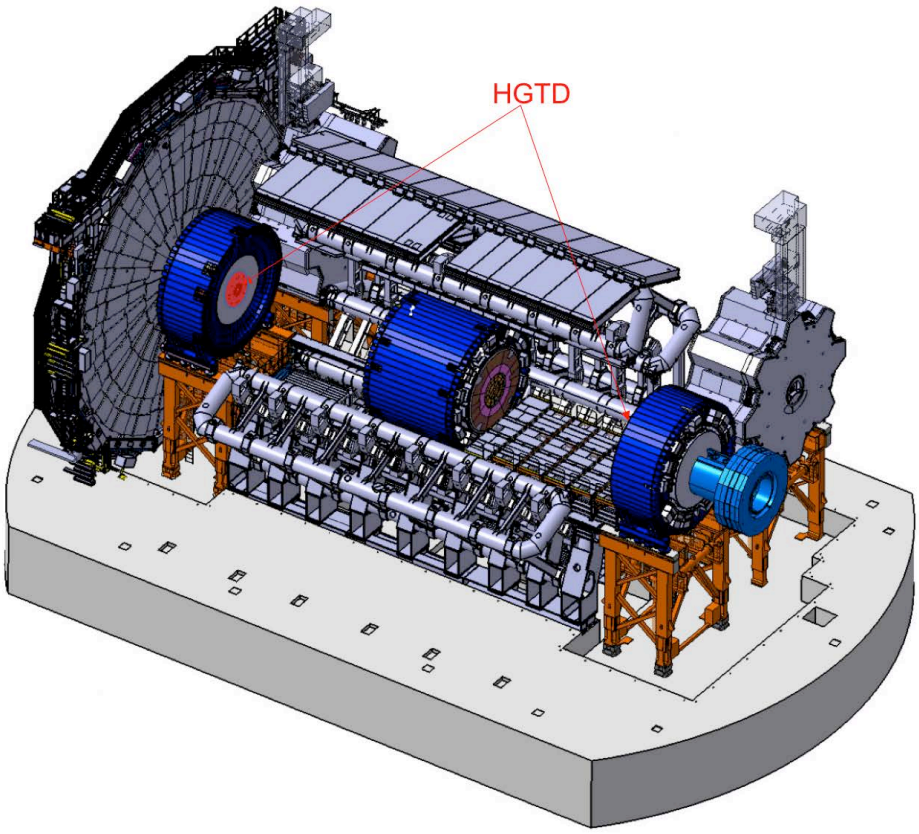


Figure 13.2: ATLAS in large opening configuration. The HGTD is superimposed to the MBTS scintillators that are presently installed on calorimeter end-cap cryostat.

13.2.2 Maintenance, radiation environment, and radio protection

Dedicated simulation studies have not yet been performed to estimate the radiation levels expected in the region of the HGTD end-caps, during its installation planned for June–July 2025, giving more than 500 days of cooling time after the LHC LS3 shutdown. During future YETS maintenance periods the access to the HGTD should occur typically after 28 days of cooling time. This is based on the assumption that each YETS is preceded by one week of technical stop and three weeks needed for the opening process of ATLAS. If the LHC run is terminated by 4 weeks of heavy-ion operation, the effective cooling time will be approximately 56 days.

The radiation environment at installation can be evaluated using FLUKA simulations, which have been performed by ITk/RP in order to estimate the expected radiation levels during ID removal in LS3, assuming 297 fb^{-1} of accumulated data. The dose equivalent rate map after 28 days of cooling time is shown in Fig. 13.3, for the geometry corresponding to the completed large opening, with all beam pipes and inner detector removed. The calculations

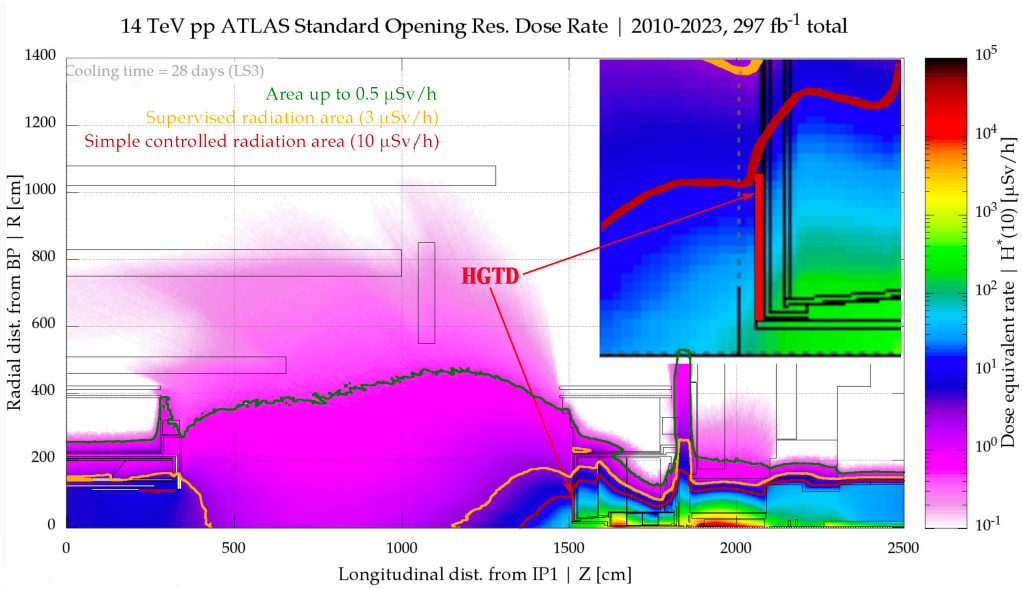


Figure 13.3: FLUKA simulations of ambient dose equivalent rate in LS3, after 297 fb^{-1} of accumulated data and 28 days of cool-down period. ATLAS is in the large opening configuration, all beam pipes and inner detector are removed. With coloured lines the boundary of various radiation areas are shown. This figure is the combination of plots given in Ref. [65].

uncertainty have been estimated comparing the simulated and measured radiation levels during the Extended YETS in 2016–2017. From the results presented in Fig. 13.4, it can be seen that the simulations underestimate the actual radiation level, and this underestimation increases progressively with the distance from the beam pipe. The simulation results are about 50% lower than measured data at 1 m radius away from the beam pipe, close to the end-cap. This can be taken as an indication of the uncertainties of the simulations, which are used to predict dose levels in LS3, when the HGTD will be installed, and for future maintenance periods.

The HGTD installation will take place after more than 500 days of cooling time; for that period a reduction of radiation levels by a factor of 2 to 3 from that shown in Fig. 13.3 is expected. This reduction, obtained as a function of cooling time according to Sullivan-Overton formula [66] for a cooling time of 500 days, has yet to be confirmed by detailed simulations. Presently, considering the calculations uncertainties, we can take as reference the dose map shown in Fig. 13.3.

The exact estimations of the dose levels at the annual shutdowns, to occur after 28–56 days of cool-down time, will determine which access and exact maintenance can be envisaged to the peripheral electronics siting inside the HGTD vessel. As an example, the presently available calculations give the estimate of the dose equivalent rate in front of HGTD in LS4 after 56 days of cooling time in range from 35 to 150 µSv h^{-1}, varying with distance from

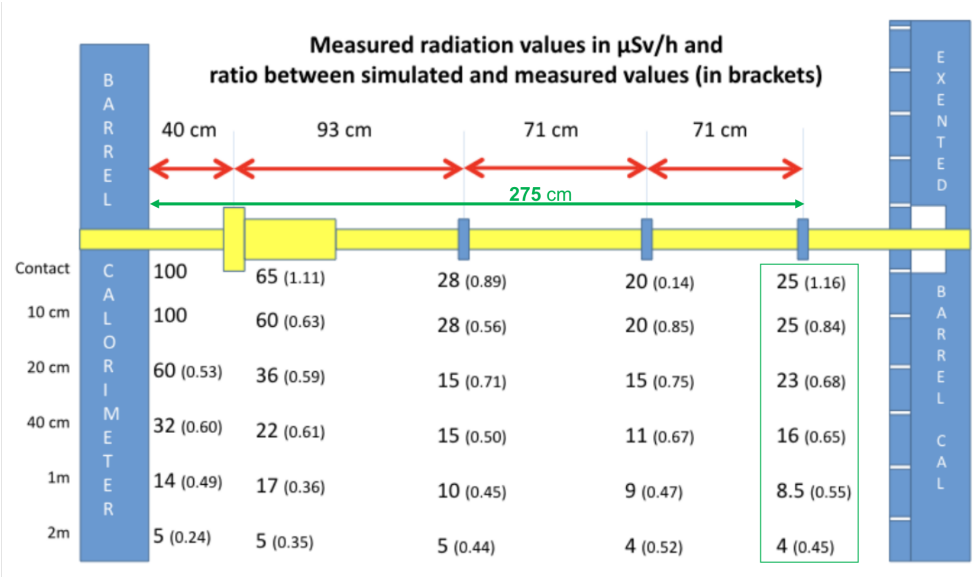


Figure 13.4: Comparison between simulated and measured radiation values in the region between the ID end-plate (in the left) and the End-cap Calorimeter (in the right, where the HGTD will be installed) during a standard opening scenario in the EYETS 2016–20117. The measured values, given in µSv h^{-1}, have been taken on 15 December 2016 after 51 days of cool-down. The ratio of simulated over measured values is given in brackets [64].

beam pipe. At such levels the access to any HGTD component inside the vessel may be compromised, subject to the detailed simulations still to be done in collaboration with CERN RP group.

The replacement of the inner part of the detector half-way through the HL-LHC program should occur in LS5, after about 2000 fb^{-1} of accumulated data. The work will be done when ATLAS will be in long-open configuration with beam pipes removed, after relatively long cooling time. Calculations dedicated to this configuration must be performed to estimate the radiation environment during the works with reasonable accuracy. Available at the present time FLUKA simulations of dose equivalent rate for LS5 period assuming 2177 fb^{-1} of accumulated data have been done for short-opening geometry with beam pipes in place, after 181 days of cool-down period, as shown in Fig. 13.5. In this configuration, the radiation levels expected in the HGTD region should be in first approximation in range of 30 to 50 µSv h^{-1} (from outer to inner radius). When replacing the inner part of the detector, the expected dose rates should be slightly lower due to longer cooling time and absence of beam pipes. Nevertheless, it will be well above the threshold defining the simple controlled area (50 µSv h^{-1}), therefore the work duration will be severely limited.

The replacement of detector modules on cooling supports will be done at the surface, and will certainly require additional cooling time before accessing the components of the detector. In order to minimise the radioactivity of the detector, less prone to activation materials must

be used in the construction, in particular by avoiding the use of stainless steel components and replacing them as much as possible by those made of aluminium or plastic. First of all, the possibility of manufacturing the aluminum pipes integrated in the cooling supports should be considered.

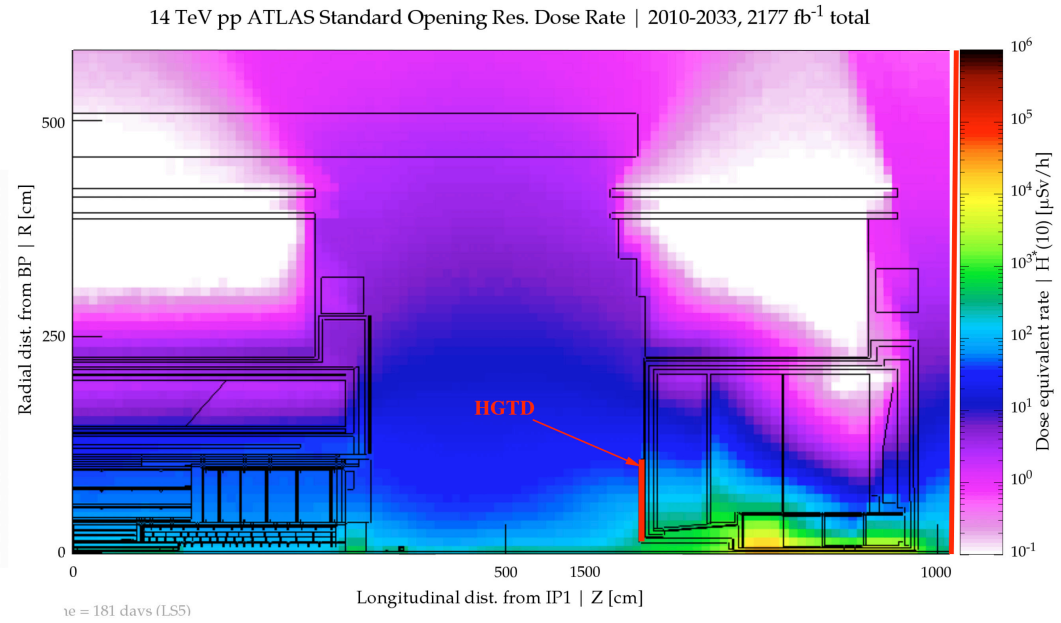


Figure 13.5: FLUKA simulations of ambient dose equivalent rate in LS5, after 2177 fb^{-1} of accumulated data and 181 days of cool-down period. ATLAS is in the short opening configuration [67].

During all ATLAS upgrade and maintenance activities, as on the CERN site in general, the radio-protection ALARA (As Low As Reasonably Achievable) principle should be strictly followed. It will certainly be implemented during the installation and maintenance activities of the HGTD, in accordance with the rules and recommendations of the CERN Radiation Protection service and in close collaboration with Technical Coordination.

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 must be planned and optimised including an estimate of the collective dose and the individual effective doses to the personnel participating in the activity. This should 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 the CERN classification criteria, which are shown in Tab. 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 seen from the table, HGTD installation activity 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 HGTD at half-life time of HL-LHC, on both end caps, may approach the limit of 5 mSv, which corresponds to the Level 3 threshold. In this case 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 ATLAS radiation safety officer (RSO) and GLIMOS prior to intervention, which can only start when the DIMR is approved. All the activities will be followed by RSO and GLIMOS on everyday basis, involving CERN Radio Protection experts when necessary.

Beside the careful work optimisation, additional measures, which will help minimise the exposition of personnel to radiation, should be considered. Among such measures are provisions of shielding, which will reduce the dose rate to the human body; use of toots for remote handling; organising 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.

13.2.3 Transport to the cavern and lowering

After the pre-assembly on surface the internal moderator, the vessel and the half circular instrumented disks will be transported to the pit. Two scenarios for transport are being considered, either in separate pieces or transport the two fully assembled HGTD end-caps. In case of moving the fully assembled detector a total weight is 275 kg per end-cap, assuming that the external moderator part will always be transported separately. The overall dimensions are 1100 mm radius and 105 mm in 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, could be lowered on side A and side C, respectively, and lowered directly from the surface to the minivans that will be in place during the long shutdown LS2, needed also for the ITk installation. These minivans have already been used in LS1 for the replacement of the MBTS scintillators, as shown in Fig. 13.6 (where the HGTD disks are superimposed to the MBTS disks to give an idea of the overall dimensions).

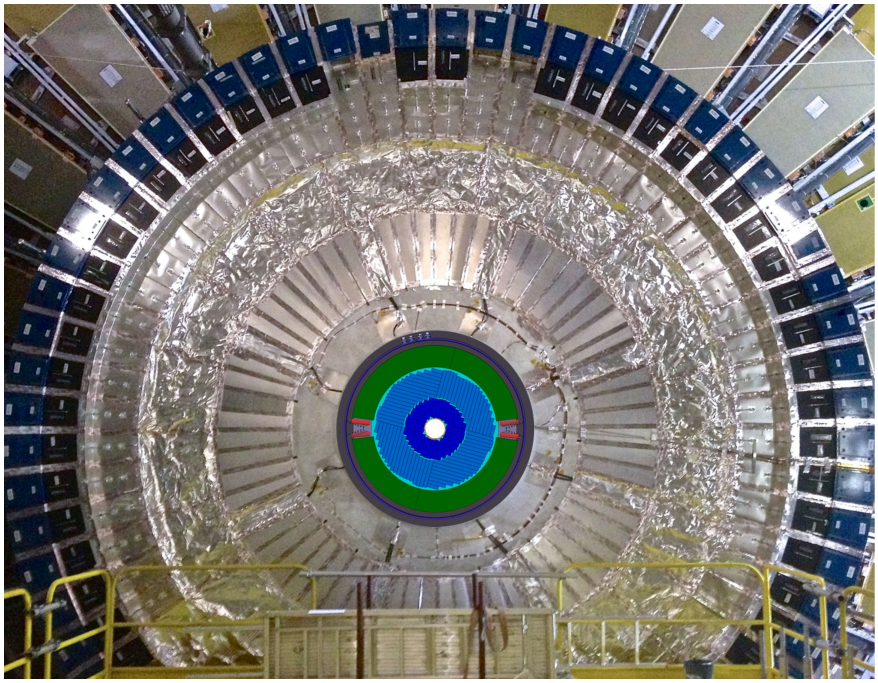


Figure 13.6: ATLAS in large opening configuration. HGTD detector superimposed on MBTS scintillators, that are currently installed on the LAr end-cap cryostat.

Specific tools are needed to perform the transport, lowering and final installation of HGTD and to insure the positioning of the instrumented disks into the vessel. 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.

13.2.4 Detector Installation and Commissioning

The installation of the detector will be done in the "Large Detector Opening" configuration as shown in Fig. 13.2. This operation can start only after the removal of the MBTS scintillators and the moderator that is presently installed in ATLAS.

As already mentioned in the previous section, temporary access platforms will be in place, the same to be used for the installation of ITk. A local small lifting tool is needed to lift each component of the detector, or each fully assembled end-cap detector side that will weight 275 kg in total (excluding the external moderator piece of 75 kg that will be mounted separately). The accurate alignment of the detector with respect to ATLAS coordinates system will be based on the LAr cryostat central hole. The hermetic vessel inner ring, which is the direct interface to central hole, will be adjusted according to latest survey group measurements records. The hermetic vessel inner ring will be carefully assembled to avoid any conflict with the beam pipe ionic pump and its power connector as seen in Fig. B.6.

The installation of each end-cap will take 1 month each and is planned for June and July 2025 for the A side and C side, respectively.

The overall commissioning will start immediately after the connectivity of the services to the detector, described in detail in Sec. 12.1.6. The access to the detector components during the commissioning should be possible until approximately March 2026, close to the expected end-cap calorimeters closure. This will leave at least 6 months of intense commissioning while access is still possible. Both the installation and commissioning of HGTD will be done with the participation of several collaborating Institutes.

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 Review. In addition, it is essential to also validate some key aspects of the integration during this period, building a realistic demonstrator. The plan is to have a two step schedule 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. 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 6 cm $\times$ 800 cm as displayed in Fig. 14.1, corresponding to the longer stave 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 RTDs embedded into the heaters. 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.

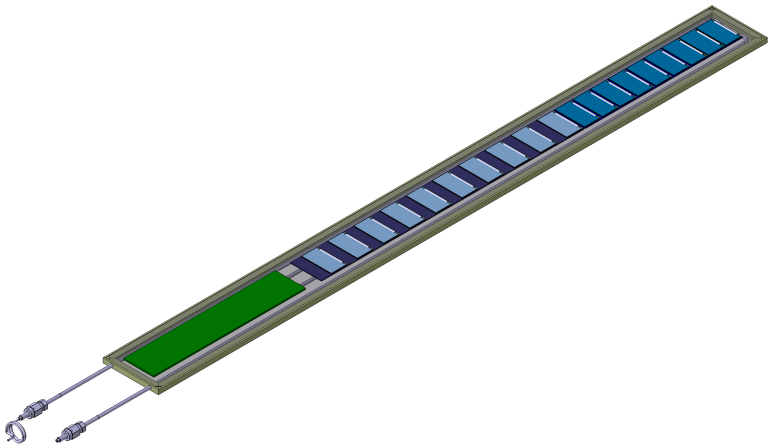


Figure 14.1: Schematic view of the cooling structure equipped with heater modules in blue. The green area corresponds to the peripheral electronics board.

After this first set of measurements, a stave should be mounted on the top and bottom faces of the cooling plates. Fig. 14.2 shows preliminary calculations of the temperature uniformity for both options which 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.

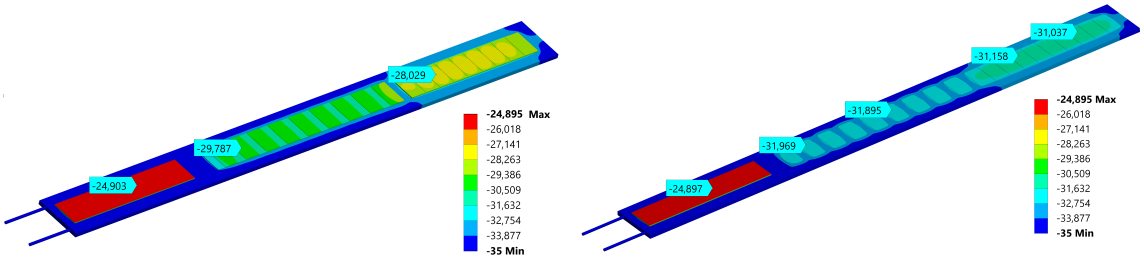


Figure 14.2: Expected temperature uniformity on the demonstrator equipped with the full intermediate plate (left) or the pattern intermediate plate (right)

Real HGTD modules will not be available before 2020. 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 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 Fig. 14.3.

The heaters consist of a silicon substrate with a similar geometry (area) of the modules

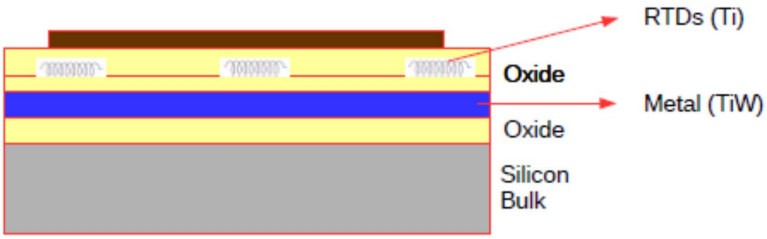


Figure 14.3: Heater transverse view

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 × 38.4 mm. They will be made of a TiW continuous layer of size XXX produced on a 300 µm 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, resistance temperature detectors (RTDs) are implanted on top of the thin metal separated by an oxide layer. The RTDs will then be placed on top of an oxide layer separating the heater from the RTDs, which will also be made from TiW. They are operated by applying a voltage and reading the current which is previously calibrated to provide temperature information. The RTDs are operated 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 Fig. 14.4.

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 the final flex design, which can provide power to the heaters and individual readout lines for the RTDs on each heater. The flex tails 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 stave 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 degrees and allow for nitrogen or dry air to be flushed

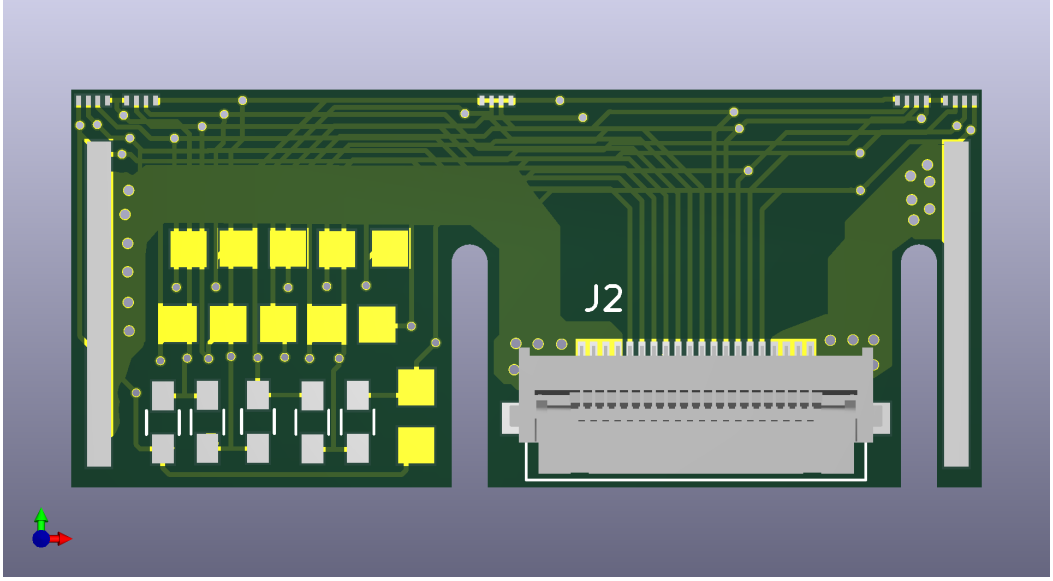


Figure 14.4: Heater flex PCB layout.

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 Fig. 11.5. The design of the heater demonstrator apparatus can be found in Fig. 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 stave and combined with a CO_2 cooling system. Furthermore, the exercise of assembling the heater modules, populate the intermediate plates and mount the full heater demonstrator is expected to provide valuable experience towards the final HGTD stave assembly and loading effort.

Add information on production details - number of sensors ordered, vendor info? Date of arrival - date of end of studies

14.2.1 Mechanical structure

Describe the work on Didier Laporte - LPNHE. Different size than full size modules.

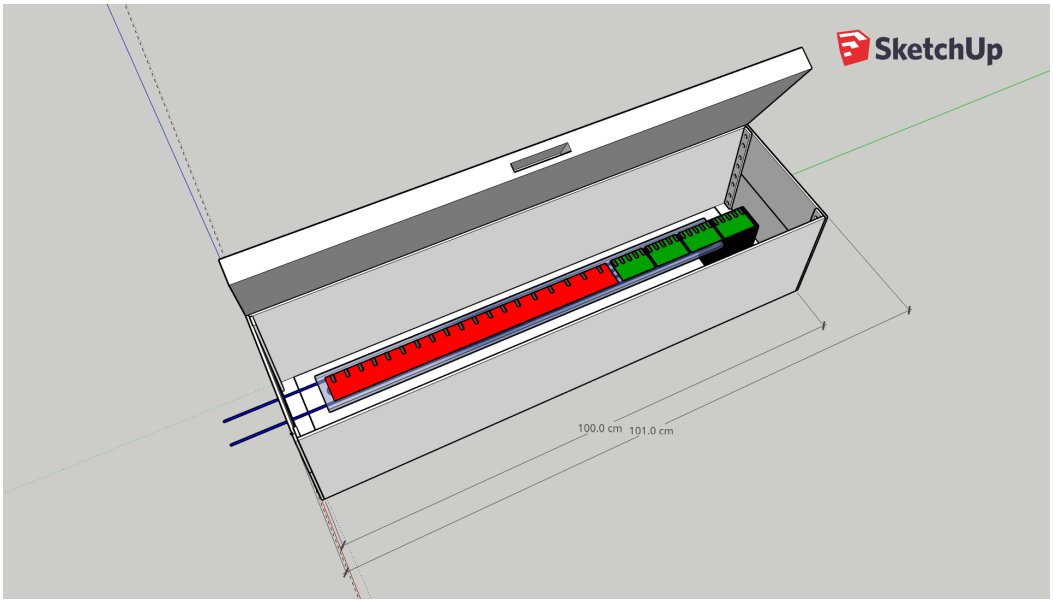


Figure 14.5: Heater apparatus. *To be updated by Afonso*

14.3 Peripheral and back-end electronics, data acquisition

This 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 exercise the assembling, connection and integration of the peripheral electronics. It consist of a PEB connected up to 55 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. An scheme of the peripheral electronics dremonstrator is show in Fig. 14.6. The design of the different items has already started and a peripheral electronics demonstrator will be ready by April 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.

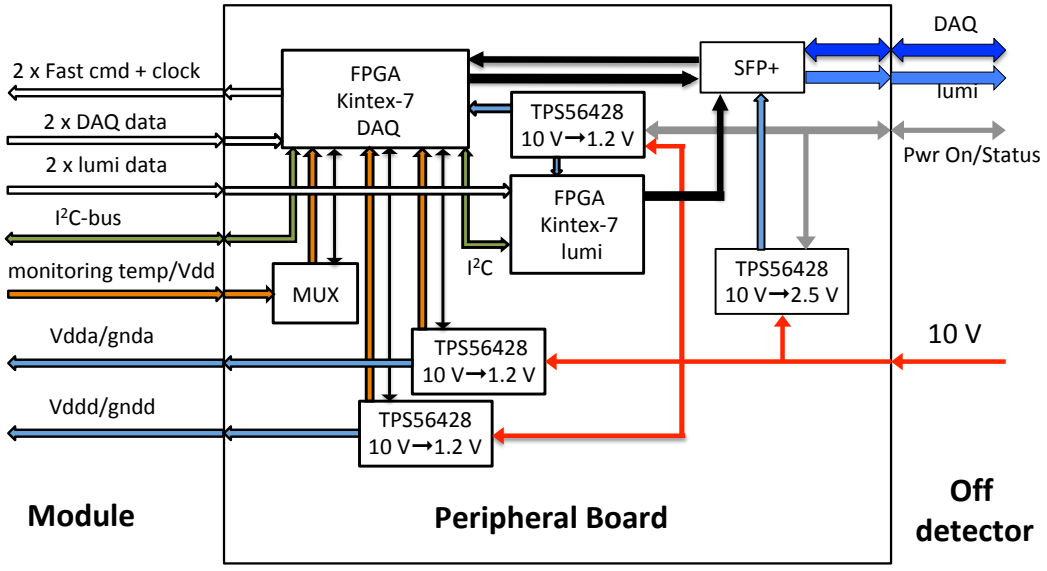


Figure 14.6: Block diagram of the peripheral electronics board demonstrator. A Kintex-7 FPGA will be used to emulate the lpGBT chipset, a SPF+ will replace the VTRx+ and a TPS56428RHLR will used instead of BPOL12V DC-DC converter.

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 has 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. Finally, the ALTIROC2 will be connected to the readout chain, in this case a GBT chip can be used to interface 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, since this protocol is not currently supported in the Phase-I FELIX board. The DAQ demonstrator roadmap is shown in Fig. 14.7.

Furthermore, the DAQ demonstrator will be used to measure the different contributions to the clock jitter at different stages (FELIX, lpGBT, FLEX, ALTIROC2). On the other hand, 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.

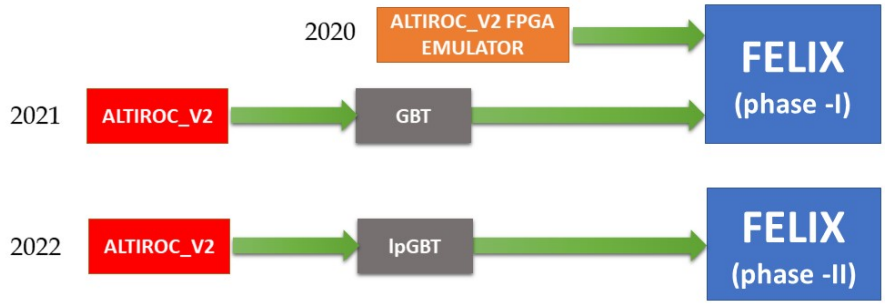


Figure 14.7: DAQ demonstrator roadmap. In 2020, an ALTIROC2 FPGA will be used to interface with FELIX. In 2021 an ALTIROC2 ASIC will be connected to FELIX using a GBT chip as interface. In 2021 a Phase-II FELIX board will interface the ALTIROC2 via lpGBT close to the final design.

14.3.3 HGTD module

The production of the HGTD modules will be used to validate the module assembly and loading process (gluing, wire bonding and mounting) and quality control measurements procedures used during the production.

To get experience of this process, smaller bare modules will be assembled in house during 2019 using the ALTIROC1 ASICs and the existing 5×5 pads sensor. While for the test beam purpose 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 the calibration signals and a beta source. These read-out boards might be used on the demonstrator until the FELIX setup is operational.

The bump bonding of the sensor to the ASIC will be outsourced in 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 discussed 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 validated therefore only early Q3/2020. 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 2020.

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 2020. 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 2020 and heater modules. A test of these modules after integration on the stave 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 peripheral electronics board and a FELIX board 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 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 stave.

14.6 Schedule and organisation

A dedicated WBS for the demonstrator is under construction and a tentative schedule of this demonstrator program is shown in Fig. 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 2020. This is strongly linked to the ASIC and sensor productions. From Q3 2020, a weekly follow-up will be mandatory to fulfil this aggressive schedule. A dedicated working group will be set up in Q2 2019 with two coordinators, one mechanics/module oriented for the heater demonstrator and a second electronics/DAQ oriented. Beyond March 2021, the demonstrator is expected to stay operational until the start of the production for additional tests.

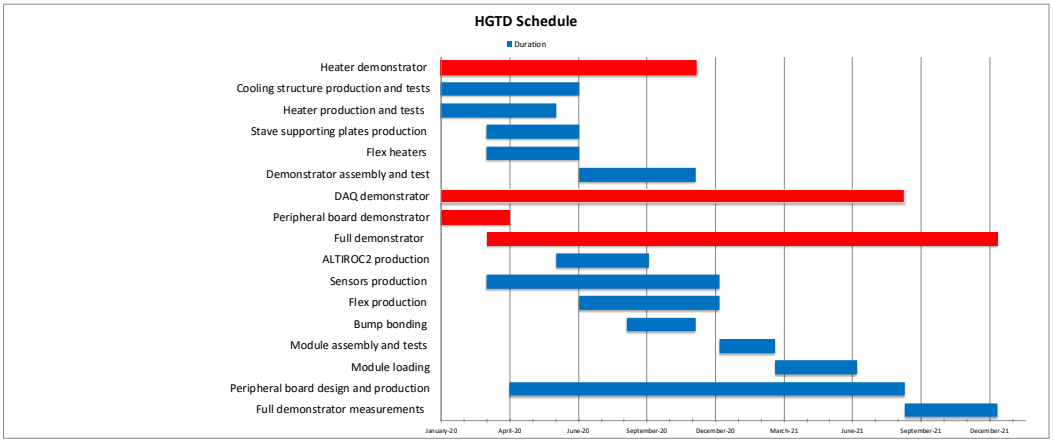


Figure 14.8: *UPDATE* Planning of the heater and full demonstrator from March 2019 to April 2021.

15 Project Organization, Costs, and Schedule

This chapter will be extended in content (in particular in section 13.2-13.4 and the schedule re-worked/tuned assuming a new baseline that is to have only 1 layer per end-cap (instead of 2) in LS3 and the second one will be installed in the next YETS 2027. This will be done/completed in close collaboration with the ATLAS Project Management Office (PMO), to be finished before the next TDR draft delivery . This chapter should also include a manpower table with needed and available resources.

This chapter describes the overall organization of the HGTD project. Sec. 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. Sec. 15.2 discusses the schedule towards the detector completion. The foreseen available resources are discussed in Sec. 15.3. Finally, in Sec. 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 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 Upgrade Project, acting together with the parent system's Project Leader and Institute Board chair. The Upgrade Coordinator should be well informed of the activities in the Upgrade Projects, and interacts regularly with the Upgrade Project Leaders 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 for the Upgrade Projects and the Upgrade Coordinator. 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 Technical Coordination. 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, comprising specifications, preliminary design, final design, and production readiness reviews.

15.1.2 HGTD organisation

The HGTD activity is currently managed by the ATLAS Liquid Argon unit but performed by Institutes not all belonging to the LAr unit yet so in addition to the LAr Insitute Board the HGTD has also its own Institute Board. It started as an organized activity in summer 2015 and this new sub-detector proposal was already part of the ATLAS Upgrade Scoping Document [68].The HGTD Initial Design Report and Expression of Interest were approved by ATLAS and LHCC in 2017. The Technical Proposal was approved by LHCC in June 2018 [69], with the recommendation to proceed to the Technical Design Report. Two Interim Upgrade Project Leaders (co-UPLs) have been chosen by the LAr management and endorsed by the LAr Institute Board. These two co-UPLs are part of the LAr steering group, and represent this HGTD project in the ATLAS Upgrade steering committee. They report to the LAr Project Leader and the UPLs, and they chair the HGTD steering group. Once the TDR is accepted, a search committee will be formed to identify the best candidates for UPL election. Either the LAr Institute or the HGTD Institute board will elect the UPL (one or two)

The project is organized in 8 working groups (WG):

- **Sensors** : this WG is currently in charge of the R& D on sensors 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 of market survey and managing the production and QA tests.

- **Electronics** this WG covers 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 th ASIC specifications, High Voltage), the DAQ WG (data format, bandwidth) and the Module assembly (for the flex) and the Mechanics/assembly WP (CO2 cooling power, services).

- **Luminosity and TDAQ** This WG covers the simulation studies and the specific hardware for the luminosity measurement and the TDAQ aspects (including the FELIX, and DCS). It make the interface with the ATLAS upgrade TDAQ and DCS project. A specific sub-group is in charge of studying and implementing the clock calibration (online and offline)

- **Modules assembly and staves loading** This WG is in charge of defining the module assembly (bump bonding, gluing, flex specifications) specification, procedure and QA and the stave loading specification and QA.

- **Test Beam** This WG is 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** This WP is a transverse WP to all the other at the exception of Simulation/Performance. 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 take the charge of the commissioning ofthe final detector.

- **Mechanics, assembly and installation** This WP is in charge of providing the specifications and building the vessels and cooling plates, the service definition and routine (with TC) an the CO2 cooling plant and N2 (with TC and CERN-DT). It should also design the tools needed for the assembly at surface and installation in the pit

- **Simulation performance and physics** The role of this WG is to provide 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.

The coordinators of the Demonstrator WP will be appointed soon after the TDR submission. Each WG, coordinated by 2 to 3 co-coordinators, carry out several activities, as detailed in the current organization chart shown in Fig. 15.1. The level 2 activity coordinators will be appointed after TDR approval. All WP coordinators are members of the HGTD steering Group (SG) chaired by the two co-UPLs. HGTD general meetings are organized by the Interim UPLs and take place bi-monthly during 3-day mini-weeks. Topical meetings in each WG area are organized by the WG coordinators on a bi-weekly basis.

The HGTD Institute board has one representative per institute with ex-officio the LAR PL and the UPL. During HGTD weeks joint steering group and Institute meetings are organized to discuss and endorse any strategic decision on detector layout, resource needs. The chair of the Institute Board is currently acting also as resource coordinator collecting and maintaining financial and manpower resource in close contact with all institutes. This has proven to be efficient during the R &D phase. A separate person will be appointed for this role after the

TDR approval to help for the preparation of the MoU and the market survey (especially sensors), working closely with the management team.

The need of a technical coordinator after the TDR approval, or for the construction phase, will be carefully evaluated, in particular if in the future organisation there will be only one project leader instead of two PLs, as it was the case since the beginning of the R&D activities. Depending on this choice a risk manager might also be appointed in order to develop and maintain the risk register in coordination with the level-2 and 3 coordinators, and to track and report any issue to the HGTD steering group.

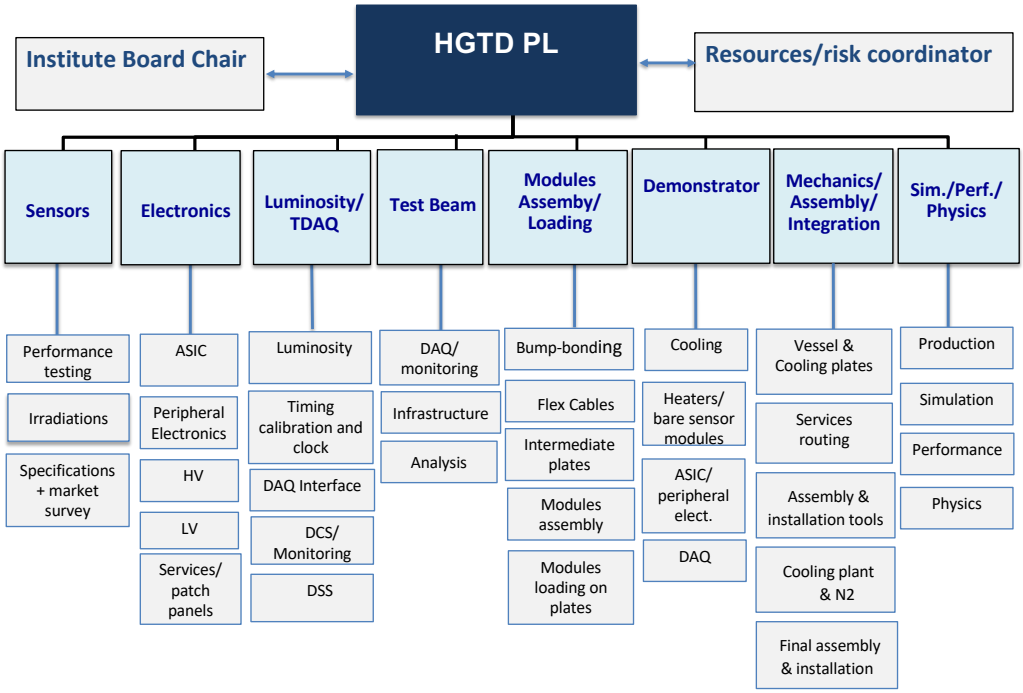


Figure 15.1: The HGTD organisation chart.

The ongoing R&D is carried out by roughly 150 physicists, engineers and technicians from 24 ATLAS institutes from 12 Countries and 13 funding agencies, see Tab. 15.1. Those are committed to carry out the R&D needed to mature the proposed detector and a sizable fraction of the Institutes are already committed to the next steps of construction, Installation and Commissioning of HGTD. US groups have been quite involved in the R &D the preparation of the TDR. After the delivery of the TDR some will stip their contribution while the remaining US group will continue only the R& D phase up to early 2021 but will not participate to the HGTD construction and part of the MoU. Sec. 15.1.2 summarizes the

Country /Funding agency	Institutes
Brazil	Sao Paulo Univ.
CERN	CERN
China (NSFC+MSTC)	IHEP, NJU, SDU, SJTU, USTC
France IN2P3	LAL, LPNHE, OMEGA, LPC
Germany BMBF	Mainz Univ., Giessen
JINR	JINR
Morrocco	Univ. Hassan II
Slovenia	JSI
Spain	IFAE, CNM
Sweden	KTH
Taipei	AS, NTHU
USA (DOE+NSF+Univ. contributions)	BNL, SLAC, SMU, Ohio SU, UCSC UCSC, Iowa Univ., Stony Brook NY

Table 15.1: List of Countries, funding agencies and Institutes/Universities participating in the HGTD R&D. Some US groups will stop their contribution after TDR delivery while some others will extend their contribution up to the end of the R&D phase only.

present involvement of the Institutes in the various R&D activities, planned until end of 2020.

R&D Activities/WG	Institutes
Sensors	BNL, CNM, CERN, Goettingen, IFAE, IHEP, JINR, JSI, NJU, USTC, SDU, SJTU,S. Paulo Univ., UCSC
Electronics	AS, Giessen, Hassan II Univ., IFAE, IHEP, Iowa Univ., JINR, KTH, LAL, LPC, NJU, NTHU, Omega, SDU, SLAC, SMU, Stony Brook NY, USTC
Luminosity and TDAQ	KTH, Ohio SU, UCSC, IHEP, Giessen, Iowa Univ.
Test beams	All institutes
Module assembly and staves Loading	BNL, IFAE, IHEP , JINR, LAL, LPNHE, Mainz Univ., USTC
Mechanics, assembly and Installation	CERN, IHEP, JINR, LAL, LPNHE
Computing-Software	AS, LAL, LPNHE, Mainz, NTHU, CERN, Hassan II Univ., KTH, SLAC, Taipei
Performance and Physics	All Institutes

15.1.3 Technical milestones

All of the custom components used for the HGTD have to pass through a series of reviews before orders can be placed for procurement of parts and production of the deliverables, and before they can be used in the upgrade of the detector. 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 conducted as (usually) half-day or full-day meetings of the group of people in charge of design and construction of the component with a team of reviewers. The review team is designated by the Upgrade Coordinator 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 is used to validate the specifications document, which describes the required functionality and performance of the device, its interfaces to other devices, tolerance to radiation, and reliability. The specified interfaces must be cross-checked for consistency with the corresponding component's specifications.

- **Preliminary Design Review (PDR)** The PDR determines whether the design is sound and meets all requirements, including all interfaces to other components.

- **Final Design Review (FDR)** The FDR is used to establish that the final prototype meets all requirements. Integrated tests with prototypes for the components the item interfaces directly to are required at this stage. A successful FDR gives the green light for a small pre-production.

- **Production Readiness Review (PRR)** 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. After successful PRR, full production is launched.

These reviews mark the transitions between different phases in each component's development and production schedule, and are thus used as key technical milestones in the overall project schedule.

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 purchasing office will probably be in charge of large quantity items, to be CORE shared by several Funding Agencies.

15.1.4 Deliverables and WBS

The deliverables for the construction of the HGTD are organized in an hierarchical Product Break Down Structure (PBS), with a direct correspondence to the existing first 5 listed WG activities. The PBS organises the deliverables into 7 primary (L2) categories, with PBS numbering from 8.1 to 8.7, as described in Tab. 15.2. When appropriate PBS is further broken down into LV3 deliverables. At the lower levels, in particular in PBS item 8.6 (Detector Assembly and QA on surface) and 8.7 (Detector Installation and Commissioning), the structure contains also the activities needed to build the deliverable. This part is refereed as Work Breakdown Structure (WBS). All PBS items, except a few items that are considered WBS, have an associated CORE cost as described later in Sec. 15.3. The Involvement in CORE is discussed later in Sec. 15.3.

15.2 Schedule and production schedule milestones

The overall ATLAS installation schedule for Long Shutdown 3 provides constraints for the scheduling of the installation of the HGTD. In the ATLAS LS3 schedule the HGTD will be installed on the end-cap LAR calorimeter cryostat faces during one month for each end-cap, in June and July 2025 for the A and C side respectively. The closing of the calorimeter end-caps is scheduled approx. 6 months later, in January 2026. The design of the HGTD allows for a baseline schedule, detailed below, that assumes an installation of the two end-caps (HGTD-A and HGTD-C) with only one double sided layer per end-cap. The second double sided layer will be installed in the YETS 2027, in situ, sliding the front vessel cover along the beam pipe (properly protected) and installing the second layer in the shape of half circular disks inside the detector. The HGTD mechanical supports, external moderator, vessel, services, cooling station needed to operate the complete detector (with 2 layers per end-cap) need to be installed in LS3.

There are three main schedule phases for HGTD:

- **2018-2020** R&D
- **2021-2024** Construction
- **2025-2026** Integration, installation and commissioning

For defining the schedule of the HGTD Project, a detailed bottom-up planning of activities has been performed. The schedule comprises the 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. As the project moves forward, tasks will be broken down into sub-tasks of shorter duration for project tracking, so that delays can be spotted early and preventive actions can be taken if the project goes off track. Each sub-project

PBS/WBS	Deliverable
8.1	**Sensors**
8.2	**Electronics**
8.2.1	ASIC
8.2.2	Peripheral Electronics
8.2.3	High Voltage power supplies and crates
8.2.4	Low Voltage power supplies and crates
8.3	**Luminosity/TDAQ** (*)
8.3.1	Luminosity boards
8.3.2	DCS and interlocks
8.4	**Module assembly and staves Loading**
8.4.1	Bump-bonding ASIC/Sensor
8.4.2	Flex cables
8.4.3	Modules assembly
8.4.4	Modules loading on staves/plates
8.5	**Mechanics, Services and Infrastructure**
8.5.1	HGTD hermetic vessel
8.5.2	Moderator (**)
8.5.3	On detector cooling/support plate
8.5.4	CO_2/water Cooling and N_2 systems
8.5.5	Tools for assembly and installation
8.5.6	Services (cables, connectors, fibres,pipes)
8.5.7	Patch panels
8.6	**Detector Assembly and QA on surface**
8.6.1	Test bench for detector certification
8.6.2	Components assembly on cooling plates (1 layer/EC) for LS3
8.6.3	Final integration inside 2 vessels (1 layer/EC) for LS3
8.6.4	Assembly of components for second layer/EC (for YETS 2027)
8.7	**Detector Installation and commissioning**
8.7.1	Services, patch panels and cooling installation
8.7.2	Back-end electronics installation in USA15
8.7.3	External moderator installation
8.7.4	Detector installation (w/ 1 layer/EC) and connectivity
8.7.5	Global commissiong in LS3
8.7.6	Install in-situ second layer/EC (in YETS2027)
8.8.7	Gobal commissioning in YETS2027

Table 15.2: Product Breakdown Structure (PBS) and Work Breakdown Structure (WBS) of the HGTD down to level 3. The PBS indicates the deliverables, to be assigned to a CORE value in the Memorandum of Understanding (MoU). The Work Breakdown Structure (WBS) is seeded by the PBS and includes the tasks required to produce the deliverables, those are mentioned explicitly when appropriate. (*) TDAQ related deliverables are not included in HGTD PBS and the corresponding costs are accounted separately in the HGTD CORE table. After TDR approval it should be added in TDAQ MoU. (**) The Core costs associated to the moderator (located inside and outside the HGTD vessel) is part of the ATLAS ITK common items.

schedule includes an overview part that summarise the schedule in distinct phases. The start points and end points of these phases are delimited by appropriate high level milestones:

- start of the design phase: SPR;
- start of the prototyping phase: PDR;
- start of the pre-production phase: FDR;
- start of the production phase: PRR;
- end of the production phase: Construction Completed;
- end of the installation and commissioning phase: Installation Completed.

The main production schedule milestones are included in the overall HGTD schedule, presented in Fig. 15.2 and Tab. 15.3. They use the PBS/WBS structure, detailed to level 3 as described in Tab. 15.2.

The schedule takes into account realistic quantities for each component, accounting for inefficiencies in all the production steps until the final assembled detector. When possible activities are taking place in parallel, for those that need to be done in a sequential way it is given a time window of at least a few months before the start of the next activity.

The schedule assumes a production model for the main components as follows.....
.....**DETAIL HERE THE PRODUCTION MODEL OF MODULES, SENSORS, ASICS, (THE 5% PRE-PRODUCTION WILL GO IN THE DETECTOR) ; 10790 PRODUCED MODULES.VS> 7984 NOMINAL QUANTITY (80% YIELD) IN MOD ASSEMBLY; ASICS: 2x10790/.80 = 26950.VS. NOMINAL SICS QUANTITY=15968 , ASSUME 2 MODULES PER HV CHANNEL AT THE START AND 1 DOUBLE SIDED LAYER/EC INSTALLED IN LS3 AND THE SECOND ONE IN YETS 2027.....**

The main schedule Milestones are detailed below.

Sensors (item 8.1)

..add text with tuned dates....

Electronics (item 8.2))

..add text with tuned dates....

Luminosity/TDAQ (item 8.3)

..add text with tuned dates....

Modules assembly and staves loading(item 8.4)

..add text with tuned dates....

Mechanics, Services and Infrustructure (item 8.5)

..add text with tuned dates....

Detector Assembly and QA on surface (item 8.5)

..add text with tuned dates....

Detector Installation and commissioning (item 8.7)

..add text with tuned dates....

The schedules for the L2/L3 deliverables are defined by the sub-project coordinators and approved by the SG. It is the responsibility of sub-project coordinators to plan, implement, execute, and track the progress of their project against the baseline schedule for their respective deliverables. They report on the progress to the SG. 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-Deliverable	Milestone	Start	End
8.1-Sensors	Specifications doc. +SPR	1 Jul 20	30 Dec 20
	PDR	Q1 21	Q1 21
	FDR	Q3 21	Q3 21
	Pre-production	1 Jan 22	30 Jun 22
	PRR	Q3 22	Q3 22
	Production (0-50%)	1 Oct 22	30 Jan 24
	Production (51-100%)	1 Feb 24	30 Jun 25
8.2-Electronics			
8.2.1-ASIC	Specifications doc.+SPR	1 Jan 20	30 Sep 20
	PDR	Q4 20	Q4 20
	FDR	Q2 22	Q2 22
	Pre-production	1 Jun 22	30 Jun 23
	PRR	Q3 23	Q3 23
	Production (0-50%)	1 May 23	30 Apr 24
	Production (51-100%)	1 May 24	30 May 25
8.2.2-Peripheral Electronics	Specifications doc.+SPR	1 Nov 20	30 March 21
	PDR	Q2 21	Q2 21
	FDR	Q1 22	Q1 22
	Pre-production	1 April 22	31 Dec 22
	PRR	Q1 23	Q1 23
	Production (0-50%)	1 May 23	30 June 24
	Production (51-100%)	1 July 24	30 Sep 25
8.2.3 and 8.2.4 (HV+LV in USA15/UX15)	Specifications doc.+SPR	1 Jun 20	30 Oct 20
	PDR	Q4 20	Q4 20
	FDR	Q4 21	Q4 21
	Pre-production	1 Jan 22	1 Sep 22
	PRR	Q4 22	Q4 22
	Production (0-100%)	1 April 23	30 Dec 24
8.3-Luminosity/TDAQ			
8.3.1 and 8.3.2 (Lumi boards+DCS/interlocks)	Specifications doc.+SPR	1 Jun 20	30 Oct 20
	PDR	Q4 20	Q4 20
	FDR	Q4 21	Q4 21
	Pre-production	1 Feb 22	1 Oct 22
	PRR	Q1 23	Q1 23
	Production (0-50%)	1 Mar 23	30 May 24
	Production (51-100%)	1 Jun 24	30 Sep 25
8.4-Module assembly+loading in staves			
8.4.1- Bump-bonding	Specifications doc.+SPR	1 Jan 21	30 Sep 21
	PDR	Q4 21	Q4 21
	FDR	Q4 22	Q4 22
	Pre-production	1 Oct 22	30 Jan 23
	PRR	Q3 23	Q3 23
	Production (0-50%)	1 Sep 23	30 Jul 24
	Production (51-100%)	1 Aug 24	30 Jun 25
8.4.2-Flex cables	Specifications doc.+SPR	1 April 21	30 Sep 21
	PDR	Q4 21	Q4 21
	FDR	Q2 22	Q2 22
	Pre-production	1 Sep 22	30 Mar 23
	PRR	Q2 23	Q2 23
	Production (0-50%)	1 Sep 23	30 Oct 24
	Production (51-100%)	1 Nov 24	30 Sep 25
8.4.3-Modules assembly	Specifications doc.+SPR	1 Apr 21	30 Sep 21
	PDR	Q4 21	Q4 21
	FDR	Q3 22	Q3 22
	Pre-production	1 Jan 23	30 Aug 23
	PRR	Q4 23	Q4 23
	Production (0-50%)	1 Jan 24	15 Mar 25

	Production (51-100%)	1 Apr 25	1 Sep 26
8.4.4-Modules loading on staves	Specifications doc.+SPR	1 April 21	30 Oct 21
	PDR	Q4 21	Q4 21
	FDR	Q4 22	Q4 22
	Pre-production	1 April 23	30 Dec 23
	PRR	Q1 24	Q1 24
	Production (0-50%)	1 Mar 24	30 April 25
	Production (51-100%)	1 May 25	30 Sep 26
8.5-Mechanics, Services, Infrastructure	Specifications doc.+SPR	1 May 20	30 Oct 20
	PDR	Q4 20	Q4 20
	FDR	Q3 21	Q3 21
	PRR	Q2 22	Q2 22
	Production (0-100%)	1 Jun 22 ?	30 Dec 24
8.5.6 and 8.5.7 (services+patch panels)	Specifications doc.+SPR	1 May 20	30 Oct 20
	PDR	Q4 20	Q4 20
	FDR	Q2 21	Q2 21
	PRR	Q4 21	Q4 21
	Production (0-100%)	1 Feb 22 ?	30 Dec 24
8.6 Detector Assembly and QA on surface	Specifications doc.+SPR	1 Apr 22	30 Sep 22
	PDR	Q4 22	Q4 22
	FDR	Q3 23	Q3 23
	PRR	Q1 24	Q1 24
	Production w/ 1L/EC(0-50%)	1 May 24	30 May 25
	Production (51-100%)	1 Sep 25	30 Oct 26
8.7 Installation and commissioning			
8.7.1 and 8.7.3 (Services,p. panels,cool.,mod.)	Installation+QA (0-100%)	30 Jan 24	30 Apr 25
8.7.2 Back-end electronics inst. in USA15	Installation+QA (0-100%)	1 Jul 24	30 Jun 25
8.7.4 **HGTD-A (w/ 1 layer)**	**Installation**	**2 Jun 25**	**2 Jul 25**
8.7.4 **HGTD-C (w/ 1 1ayer)**	**Installation**	**3 Jul 25**	**1 Aug 25**
8.7.5 Commissioning in LS3 (w/ lL/EC)	Commissioning	3 Jul 25	10 Mar 26
8.7.6 **HGTD-A (w/ layer 2) in YETS27**	**Install in situ layer 2**	**1 Jan 27**	**30 Jan 27**
8.7.6 **HGTD-C (w/ layer 2) in YETS27**	**Install in situ layer 2**	**1 Feb 2027**	**2 Mar 27**
8.7.7 Commissioning in YETS27 (w/ 2L/EC)	Commissioning	1 Mar 27	30 May 27

Table 15.3: Schedule of the main HGTD deliverables IN LS3 AND YETS 2027, including the planned reviews (SPR, PDR, FDR, PRR), Pre-production and Production, assuming as baseline only 1 double sided layer per end-cap and the second one to be installed in YETS 2027.**TUNE NEW SCHEDULE, ASSUMING AS BASELINE 1 LAYER/EC IN LS3 AND SECOND LAYER/EC IN YETS 2027.**

15.3 Costs

The cost of deliverables for ATLAS projects are expressed as their CORE cost. CORE is defined as the sum of the material value of each component that makes 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 of existing manpower, 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,

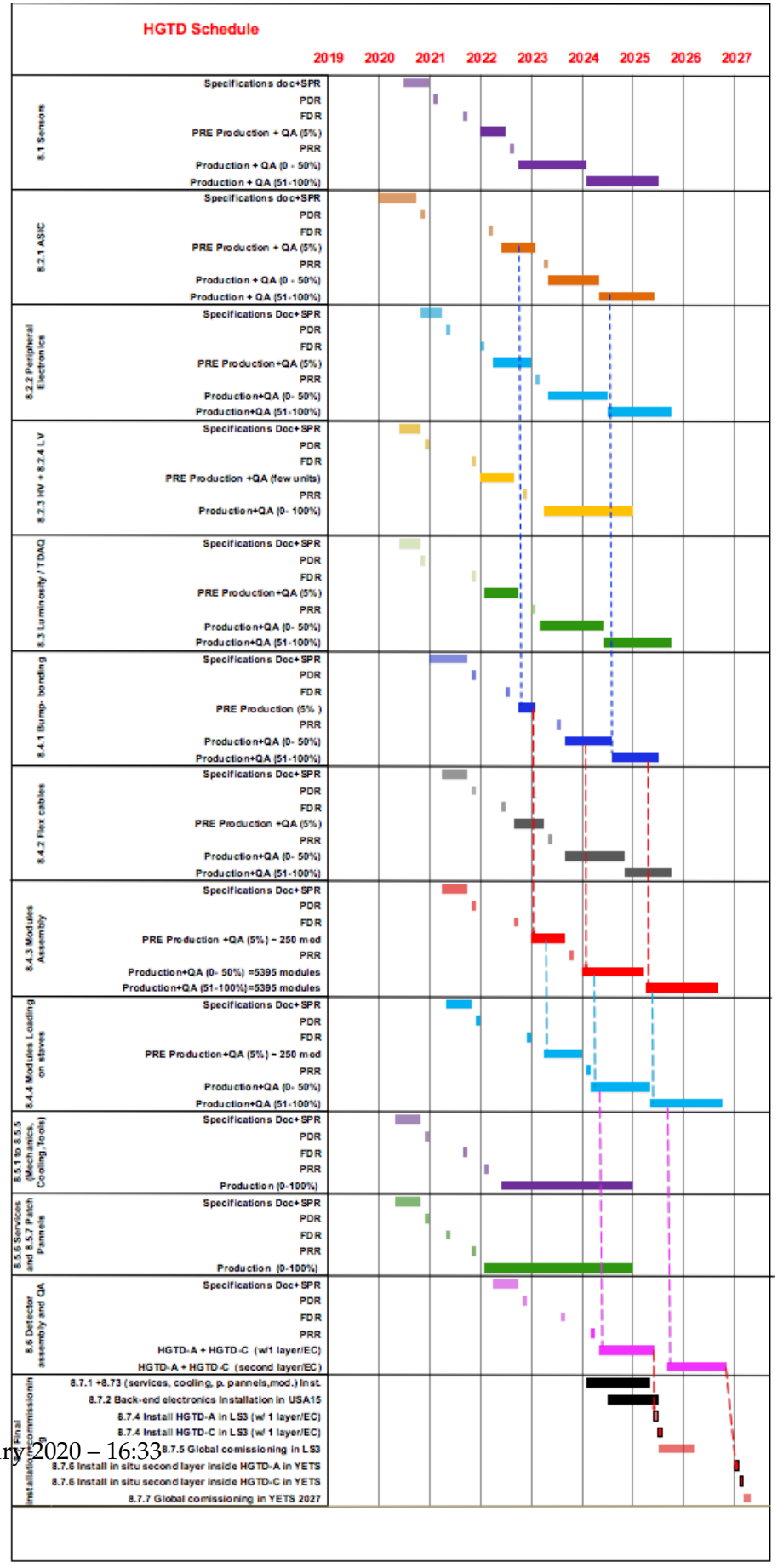


Figure 15.2: Chart showing the main HGTD of the main HGTD deliverables, including the planned reviews (PDR, FDR, PRR), Pre-production and Production. **OPTIMIZE THIS NEW SCHEDULE MADE WITH ONLY 1 DOUBLE SIDED LAYER/EC IN LS3 AND THE SECOND LAYER/EC IN**

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 country to country.

Each HGTD PBS item, described in Tab. 15.2, has an associated CORE cost that is defined as the sum of the material value of all components making up the deliverable. A Work Breakdown Structure (WBS) is seeded by the PBS and adds to it the tasks that need to be performed to design, prototype, produce, assemble, and install the deliverables. The HGTD CORE cost has been estimated in a bottom-up approach. The cost estimates are detailed with individual elements for most items. The numbers of items have been calculated based on the layouts and schemes presented in this document.

The Yield model used in the cost estimates accounts for failure and loss during the production phase, up to and including the installation of items in the ATLAS cavern. In contrast, spares account for failure and loss during the operations phase, i.e. from the beginning of Run 4 onwards for items installed in the cavern during LS3. Yield is supported by upgrade funds and counts toward CORE cost, while spares are supported by maintenance and operations (M&O) funds and do not count as CORE.

For each significant production step, the yield was estimated based on past production experience with similar or equivalent items or extrapolating from prototypes experience and a summary of the production model is shown in **??**. It assumes that the total of the HGTD pre-production components of 5% quantity is of good quality and will be used in the detector. It is also considered that each HV channel will be reading 2 modules (2x2x225 channels), at least in the initial phase when the irradiation levels are smaller.

ADD HERE A TABLE SUMMARIZING THE YIELD MODEL OF THE MAJOR COMPONENTS (ASICS,SENSORS,MODULES) AND FEW LINES OF TEXT TO DESCRIBE THIS NEW TABLE.

The cost estimates for each item are quoted in CHF, using the exchange rates of 1Euro = 1.085 CHF, 1 USD = 0.986 CHF and 100 JPY = 0.942 CHF, as in the other ATLAS phase II TDRs. The core estimates are based on existing contracts (ASICs), quotes from Industry (sensors, FPGAs, Flex cables,...), extrapolation from other ATLAS Upgrade Phase-II TDRs with already signed Memorandum of Understanding (MoU) that are using the same or similar type of components (power supplies, cables, cooling station used in the ITk).

A summary of the HGTD core cost[1], detailed to the PBS level 3 is presented in Tab. 15.4, with a total of 9716 KCHF for the HGTD and 995 KCHF for the HGTD-TDAQ related costs.

The costs for the planned replacement of the HGTD inner ring during the HL-LHC half life time should be accounted in the future (M&O) funds. Assuming that the outer radius of

[1] The item "Detector readout, data flow, and network" is considered a TDAQ deliverable but cannot be included in the TDAQ TDR until the HGTD TDR has been reviewed by LHCC/UCG.

the inner ring will be kept at 320 mm the costs are estimated to be approximately 30% of the items: 8.1 (sensors), 8.2.1 (ASIC), and 8.4 (module assembly and loading on staves). The overall costs of maintenance and rolling replacement of hardware after its installation are also foreseen as part of the future HGTD maintenance and operation budget. The HGTD resources coordinator will review the HGTD costing in close collaboration with the experts of the different deliverables and the ATLAS resources coordinator. The cost of the project is expected to be covered by the Institutions participating in the HGTD Phase-II upgrade project, 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. A preliminary survey of Institutes interests and resources indicate that a substantial fraction of the money and manpower resources required is already covered and with a balanced sharing by Institutes among the various PBS/WBS deliverables that are needed to construct and operate HGTD.

15.4 Risk management

THIS SUB-CHAPTER STILL NEEDS TO BE COMPLETED WITH A RISK MANAGEMENT ANALYSIS IN TERMS OF COST, SCHEDULE AND PERFORMANCE.

Many of the technical choices in the HGTD concept were made already at the time of the Expression of Interest and Technical Proposal for the best compromise between performance and cost. The severe space constraints (in z, r), high radiation levels (still evolving with the finalization of the ITk services and supports layout) 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. Several risks are identified:

- **Performance degradation** due to possible further increase in the expected radiation levels. This increase, not yet stabilized, is caused by the increasing amount of ITk services and supports in the patch panel PP2 region, in front of HGTD. The baseline layout has the transition radius between the replaceable HGTD inner ring and the permanent outer ring at r =320 mm. Three possible scenarios are envisaged and should be decided once the final radiation levels are available with the ITk realistic services/supports and the performance of the irradiated HGTD sensor+ASIC system are validated with real size prototypes:
  - Increase the inner ring outer radius by few cm but bringing the extra modules inside the inner ring with unchanged 20% sensors overlap. This scenario does not increase the amount of modules in the detector, nor the CORE costs.
  - Increase the inner ring outer radius by few cm but the increased area will have 80% overlapped sensors, to insure 3/hits/track in all the inner ring region. This

scenario, if justified, should be carefully planned, since it will increase the amount of modules in the detector and amount of data to be transmitted to the peripheral electronics and to the outside of the detector.

  - Another possible action that can be cumulative to any of the 2 previous options is to replace twice the inner ring during the life time of the HL-LHC and not only one time as baselined in this project. The costs associated to each replacement of the inner ring (sensors, ASICs, flex cables) should be accounted in the future M&O resources.

- **Resources shortage**, in particular experts manpower and Institutes from USA that today are actively involved in the R&D but will probably disappear for the construction phase. This scenario may lead to delays in the project that has already a late start compared to the other ATLAS Phase-II upgrade projects. An active action is on-going to attract new Institutes from ATLAS that have expertise in the relevant areas of the HGTD project and may bring the needed CORE and manpower expertise needed for the construction and later operation/maintenance of the HGTD.

- **Delay scenarios**. The possible delay in the HGTD schedule is probably the main risk of this project, due to the late start compared to all the other ATLAS Phase-II upgrade projects. The high modularity of the detector, constructed from 2 cm $\times$ 4 cm modules assembled into staves/intermediate plates that are later screwed into 1/2 circular cooling disks allows several working scenarios in case of delays or lack of resources to bring a complete detector in Summer 2025 or later:

  - A priority should be to complete during the LS3 shutdown the construction and installation of all the services, patch panels, cooling station, moderator, empty vessel, to be locate in UX15 cavern.

  - Install all the crates and it's equipment in USA15 (power supplies, Luminosity/DCS/TDAQ boards, computers).

  - In summer 2025 install on the two end-cap calorimeter cryostat faces all the available 1/2 circular pieces of HGTD inside the two HGTD vessels. In the present schedule the final assembly and integration of the detector components (modules/staves, peripheral electronics) into 1/2 circular cooling disks is planned to be done at CERN in a sequential way for the A and C sides for a duration of 9 months each. In case of cumulative delays the assembly of the 2 end-caps could be done in parallel, by duplicating the tools and Institutes manpower based at CERN for this operation.

  - In case of an incomplete detector at the start of HL-LHC, one should prioritize full instrumented 1/2 disk pieces, with an equal number of 1/2 disks for the A and C side. For example, one could start with only one layer per end-cap in the first year(s).

6353 – Install the missing 1/2 disk HGTD pieces at the next medium-short shutdown,
6354 when the end-cap calorimeters are opened to give access for the usual Tile or LAr
6355 barrel calorimeter electronics maintenance. The HGTD installation procedure and
6356 respective tools will be carefully designed to allow to complete the 1/2 circular
6357 detector disks installation even in the presence of the beam pipe. Only the vessel
6358 front cover made of 1 piece of circular shape, to improve thermal insulation, has
6359 to be installed without the beam pipe in place. The vessel front cover should
6360 be able to move on the installed beam pipe to allow a posteriori the completion
6361 of the 1/2 circular detector pieces and access during future HGTD maintenance
6362 periods.

PBS	Item	Core Cost (kCHF)
8.1	**Sensors**	**2275**
8.2	**Electronics**	**3199**
8.2.1	ASIC	833
8.2.2	Peripheral Electronics	767
8.2.3	High Voltage power supplies and crates	532
8.2.4	Low Voltage power supplies and crates	299
8.3	**Luminosity/TDAQ (*)**	**395**
8.3.1	Luminosity boards	315
8.3.2	DCS and interlocks	80
8.4	**Module assembly and staves loading**	**1483**
8.4.1	Bump-Bonding ASIC/Sensor	450
8.4.2	Flex cables	547
8.4.3	Modules assembly (incl tools)	386
8.4.4	Modules loading on staves/plates (incl tools)	100
8.5	**Mechanics, Services and Infrastructure**	**2264**
8.5.1	Vessel (including feedthroughs)	160
8.5.2	Moderator (**)	
8.5.3	On detector cooling/support plate	180
8.5.4	CO_2/water Cooling and N_2 systems	1237
8.5.5	Tools for final assembly and installation	100
8.5.6	Services (cables, connectors, fibres)	526
8.5.7	Patch panels (w/ water cooling)	61
8.6	**Detector Assembly and QA on surface**	**100**
8.6.1	Test bench for detector certification	100
8.6.2	Assembly of components on cooling plates	-
8.6.3	Disks assembly	-
8.6.4	Final assembly inside vessel	-
8.7	**Detector Installation and commissioning**	**-**
8.7.1	Services and patch panels installation	-
8.7.2	Back-end elect. installation in USA15	-
8.7.3	Outer moderator part installation	-
8.7.4	Detector installation and connectivity	-
8.7.5	Detector commissioning	-
Total HGTD(kCHF)		**9716**
TDAQ(*)	Felix boards+LTI boards, emulator,...	**995**
Total w/ TDAQ		**10711**

Table 15.4: Estimated Core cost of the HGTD (in kCHF). The total cost is given with and without the costs of the TDAQ. It assumes that the total of the HGTD pre-production components corresponding to approximately 5% of the total needed production is of good quality and will be used in the detector. It is also considered that each HV channel will be reading 2 modules (2x2x225 channels), at least in the initial phase when the irradiation levels are smaller. (*) TDAQ related costs are estimated separately by TDAQ and should be moved to TDAQ CORE, after HGTD TDR approval. (**) The moderator core costs are accounted in the ATLAS ITK common items.

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, CEA-DRF/IRFU, France; SRNSFG, Georgia; BMBF, HGF, and MPG, Germany; GSRT, Greece; RGC, 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, Russian 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, CRC and Compute Canada, Canada; COST, ERC, ERDF, Horizon 2020, and Marie Skłodowska-Curie Actions, European Union; Investissements d' Avenir Labex and Idex, 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, Spain; 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. [70]. Many thanks to Nicolo Cartiglia for providing the FBK UFSD3 LGADs (production funded by INFN) and performing the Weightfield2 simulations.

A Expected Energy Spectra

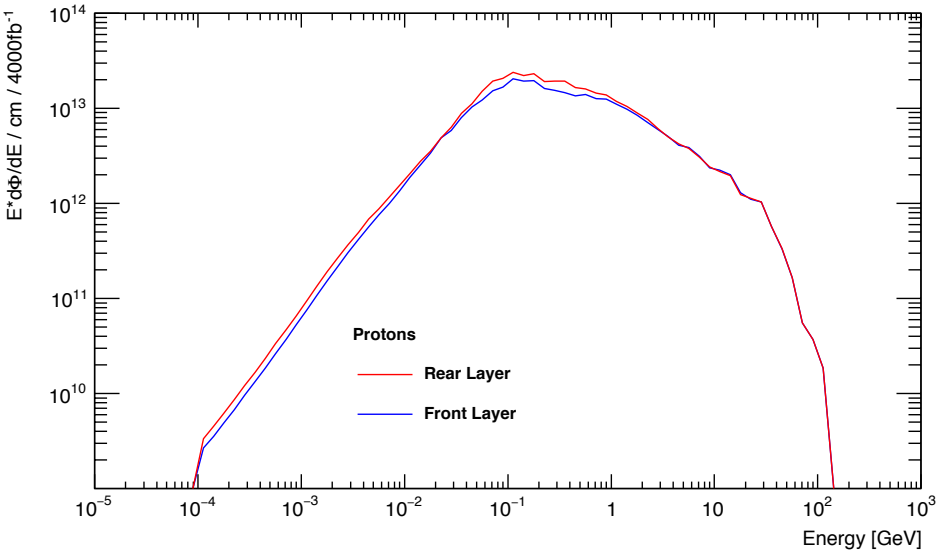


Figure A.1: Proton spectra averaged over the rear (outermost) and front (innermost) silicon layer of the HGTD. The uncertainties are of the order of 5%

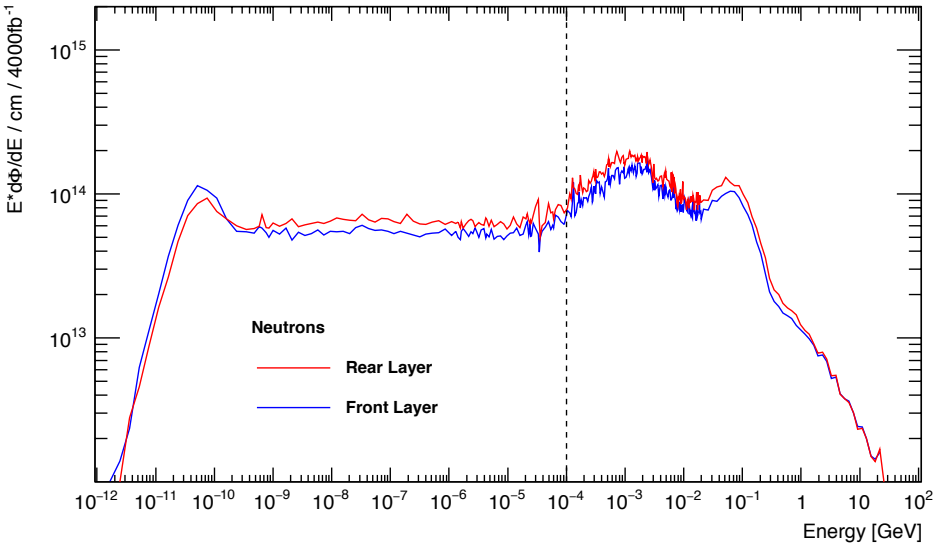


Figure A.2: Neutron spectra averaged over the rear (outermost) and front (innermost) silicon layer of the HGTD. The uncertainties are of the order of 5%. The fluctuations between 1 keV and 10 MeV are due to resonance.

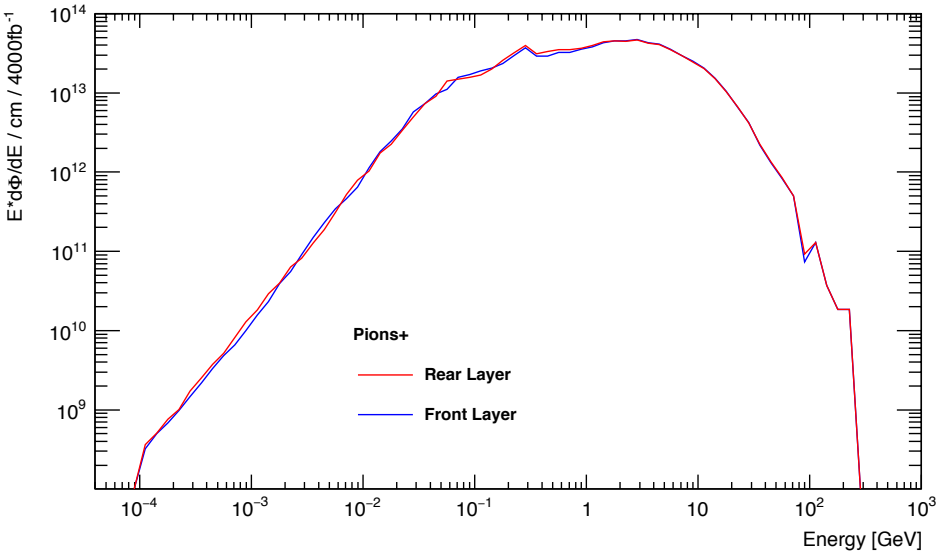


Figure A.3: Pion spectra averaged over the rear (outermost) and front (innermost) silicon layer of the HGTD. The uncertainties are of the order of 5%

B Technical Drawings

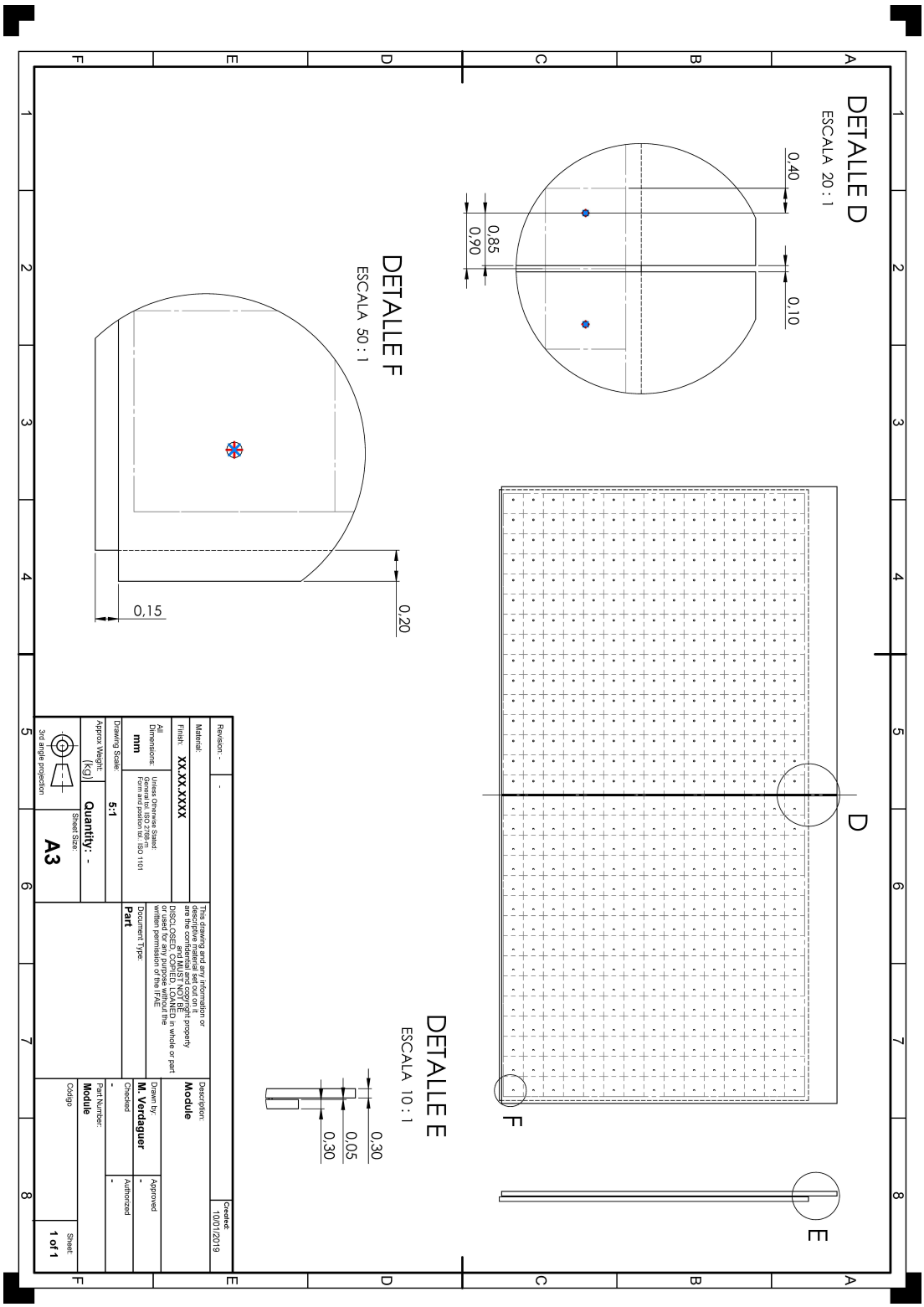


Figure B.1: Sketch of the bare module (sensor and ASIC). Distances are in millimeters. The bump pads on the sensor are shifted by 250 µm on each side of the sensor, to allow a 100 µm separation between the ASICs.

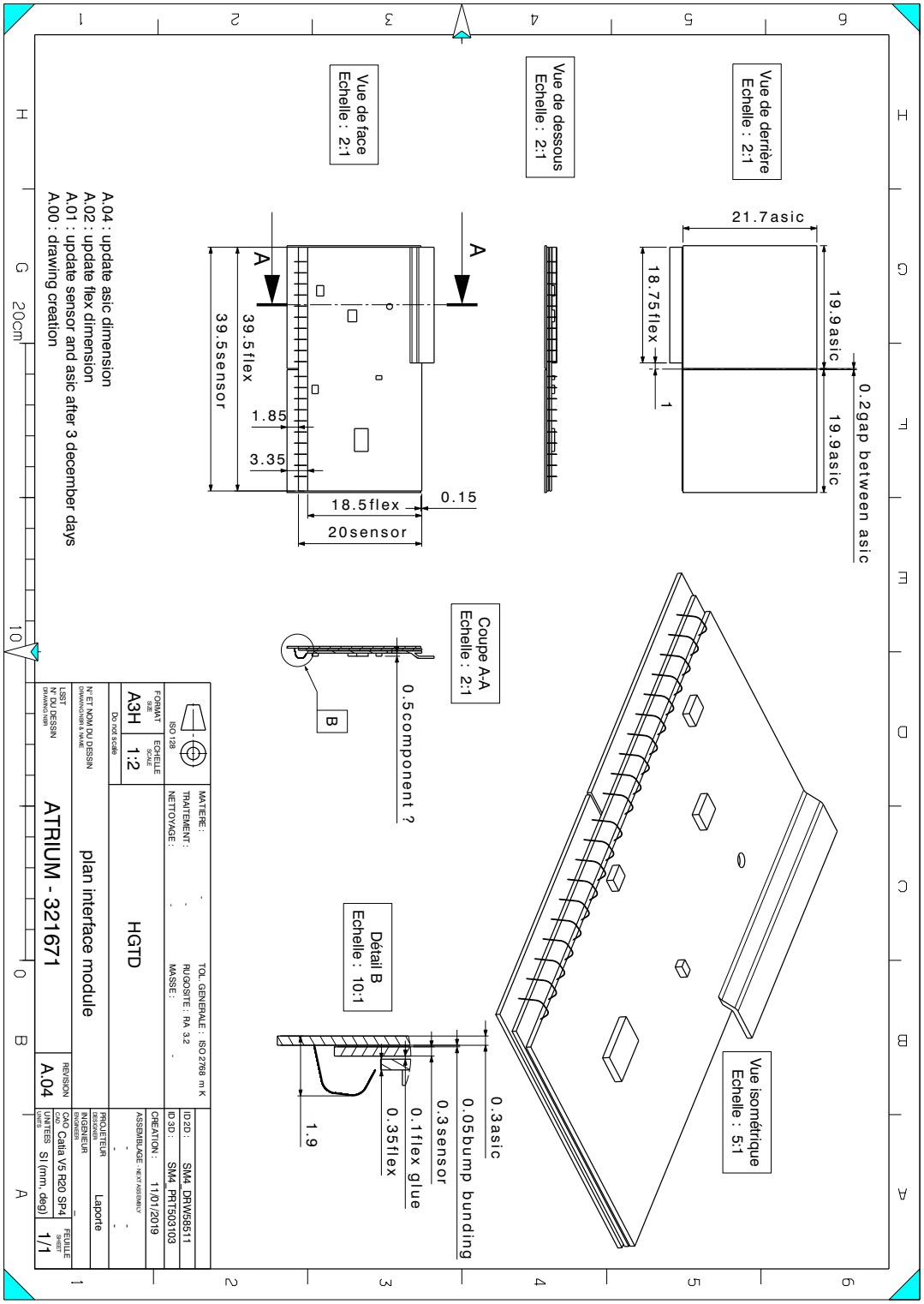


Figure B.2: Sketch of the module with the sensor, the ASICs and the Flex cable. Distances are in millimeters. Dimensions of the different components are visible, including the bumps pads, the glue and the wire-bonds.

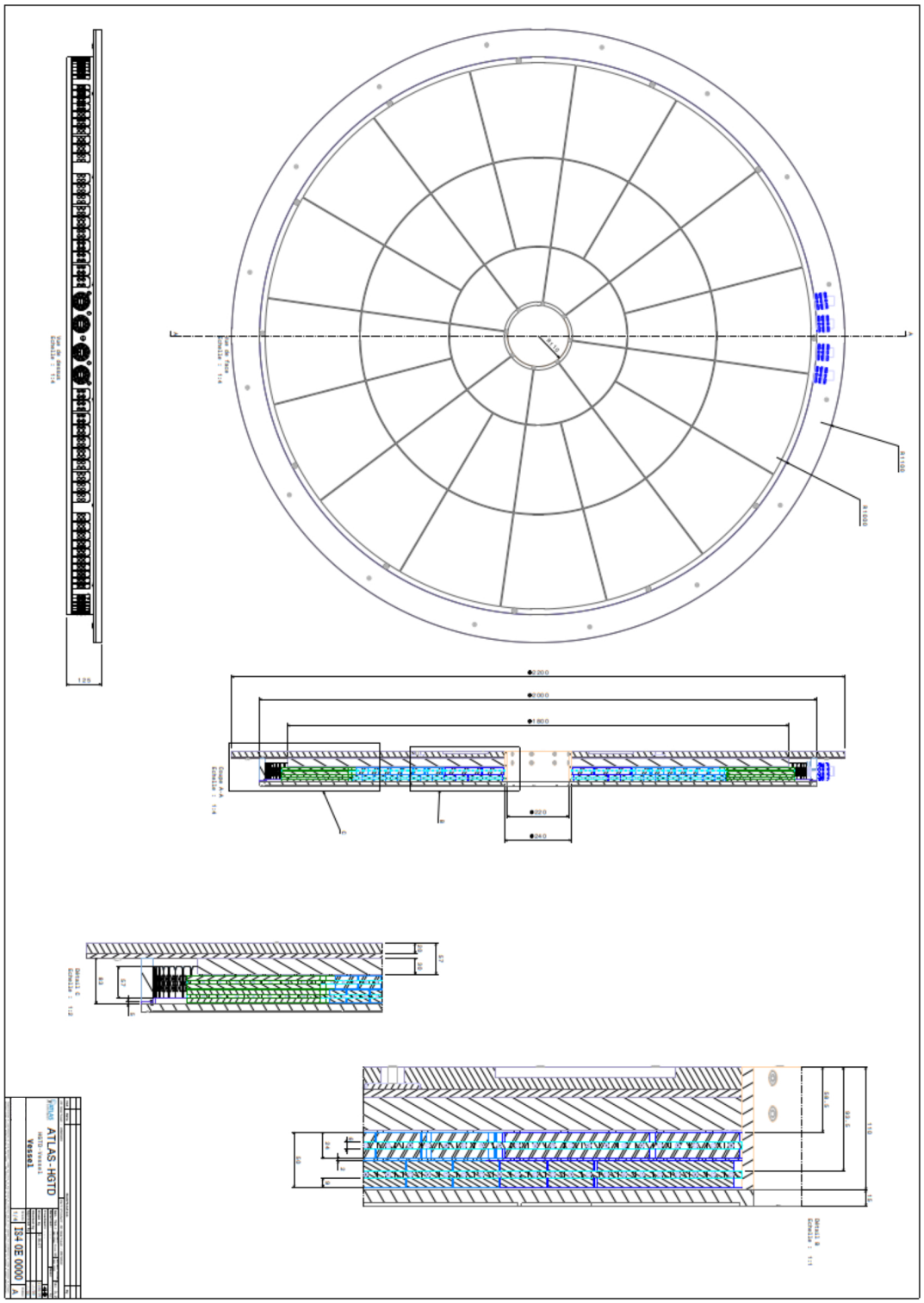


Figure B.3: top: View of the HGTD vessel front cover and feedthroughs region. The three bottom plots show an *r*-*z* view of the HGTD components inside the vessel, including a zoom of the inner radius and outer radius regions.

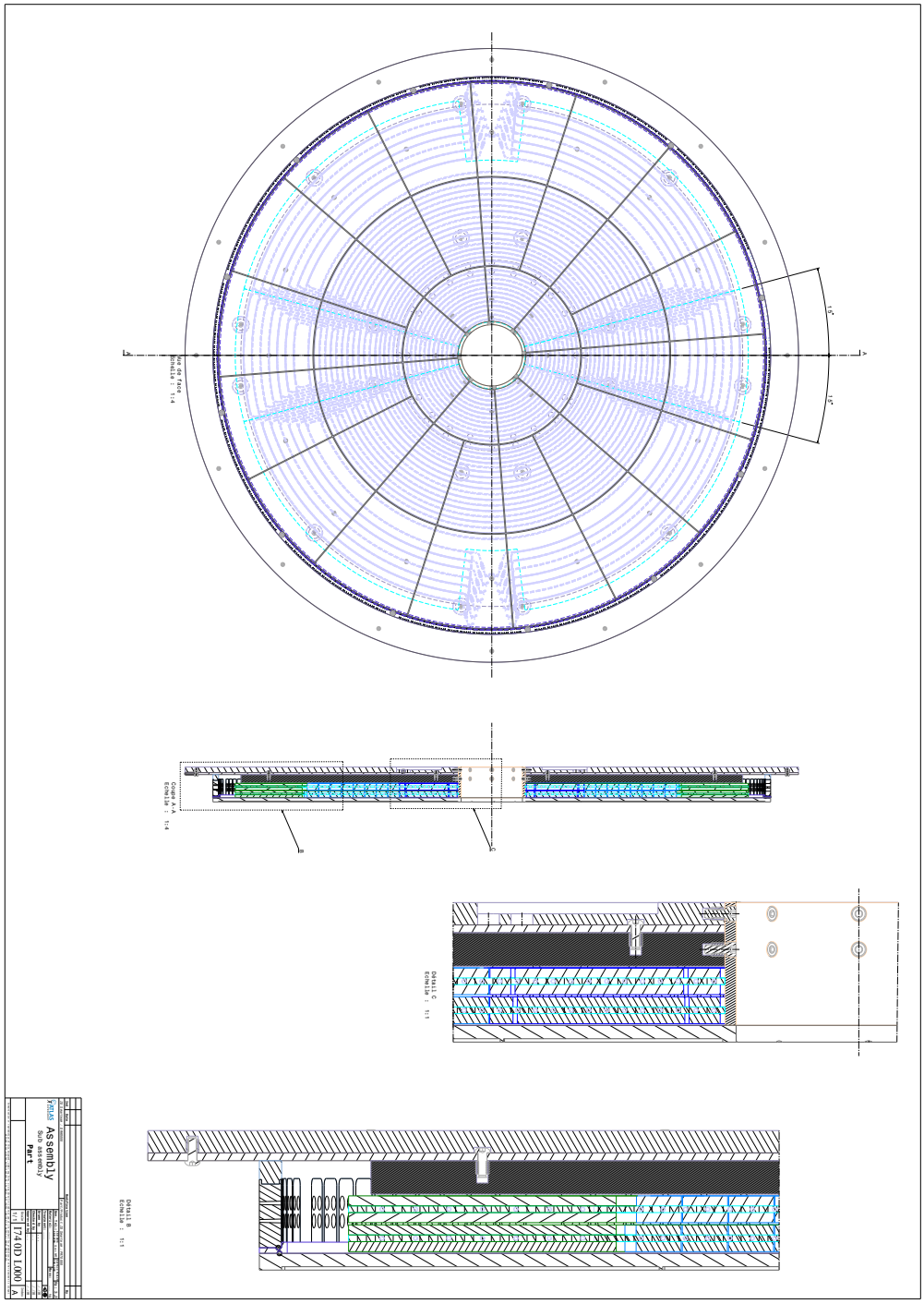


Figure B.4: View of the front and rear HGTD cooling disks inside the vessel, rotated by 15° in ϕ with respect to each other. The vessel front cover was put semi-transparent to allow to see the cooling disks. The three bottom drawings show r-z views of the HGTD components inside the vessel, including zooms of the inner radius and outer radius regions.

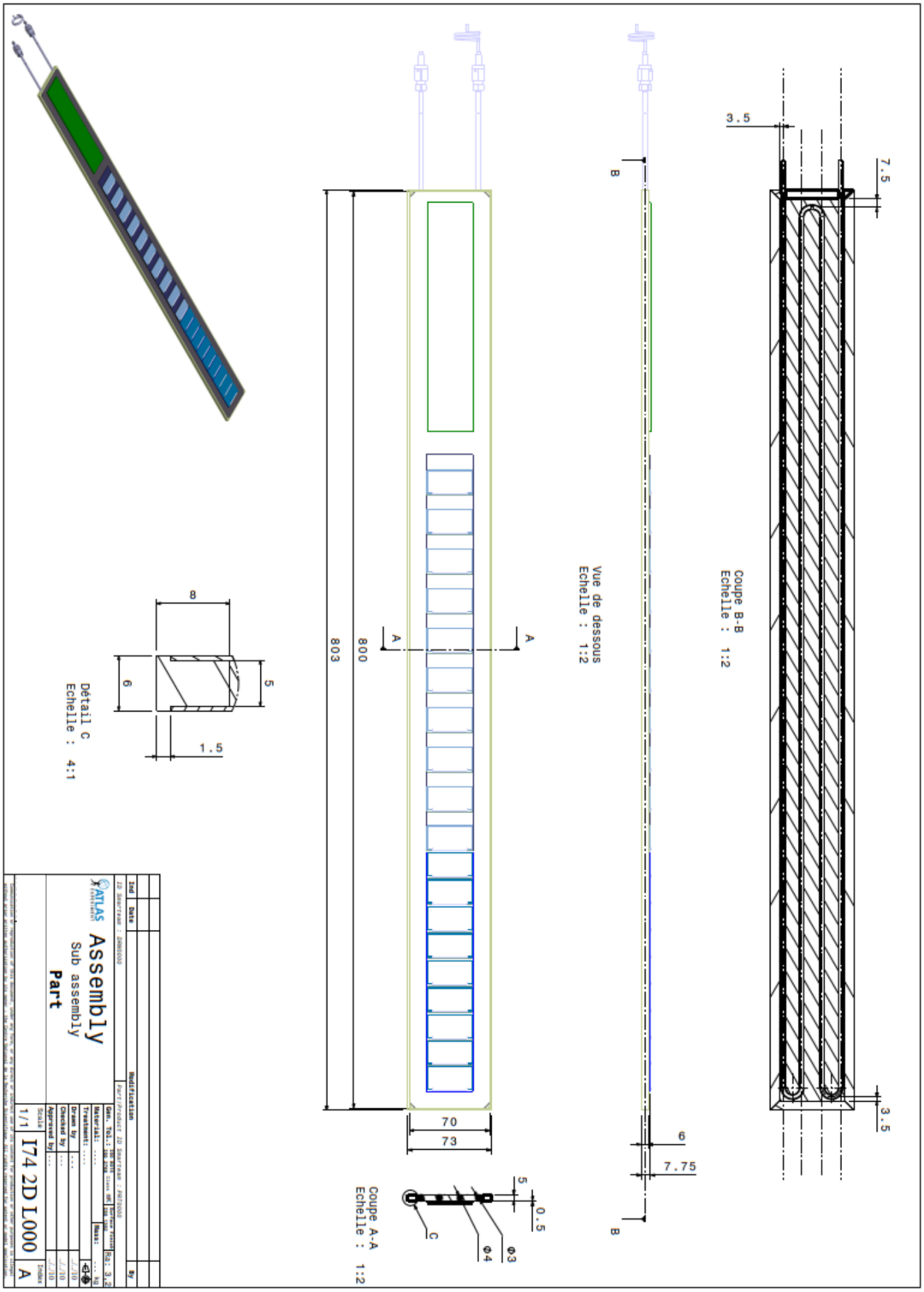


Figure B.5: View of the mechanical prototype planned for the HGTD demonstrator. It includes a cooling plate, dummy modules and connectivity to peripheral electronics board (indicated in green). The heaters simulating the modules power dissipation, using dummy modules are in blue.

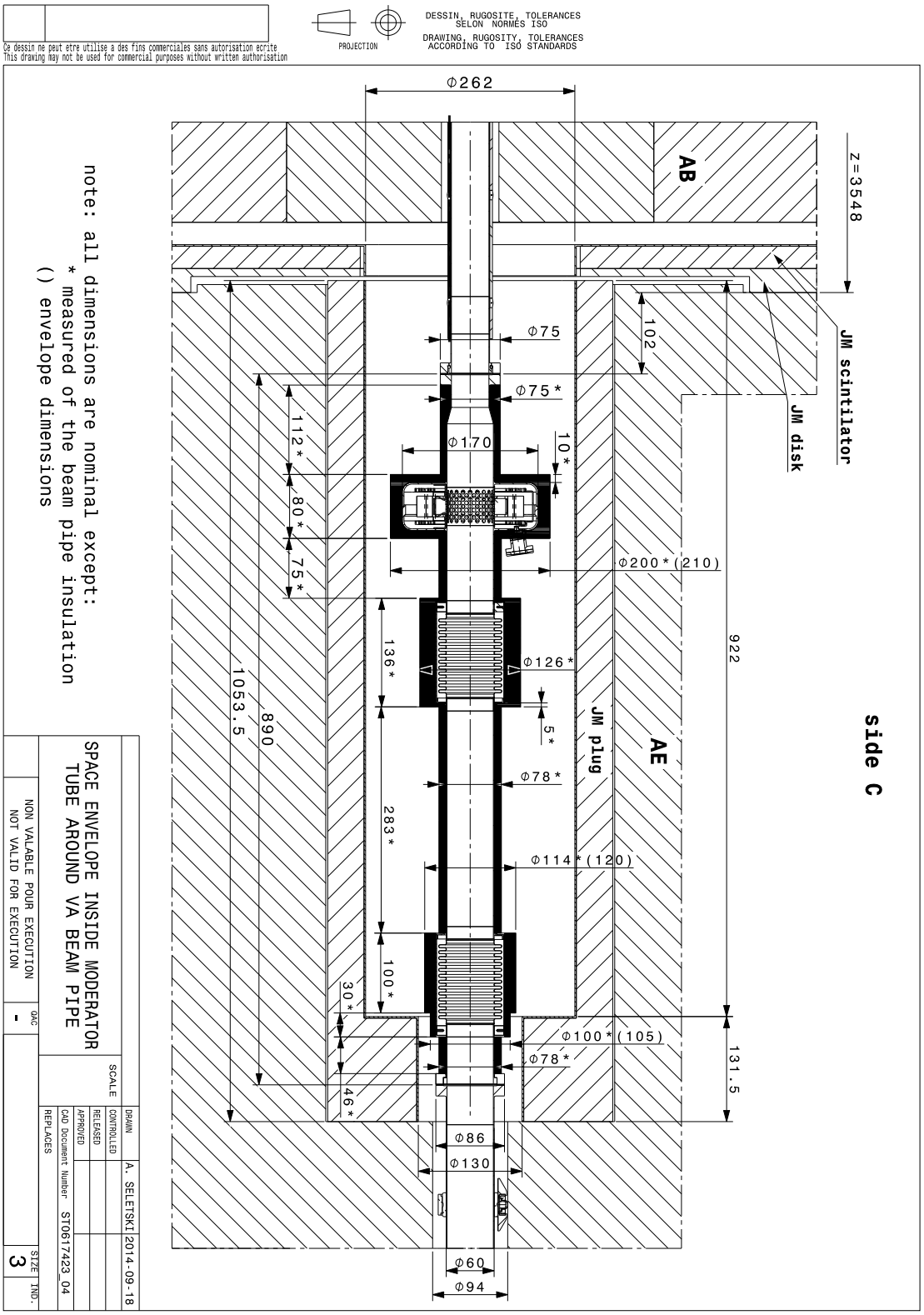


Figure B.6: Details of the space envelope around the beam pipe and moderator.

Bibliography

[1] 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).

[2] F. Bordry et al., *Machine Parameters and Projected Luminosity Performance of Proposed Future Colliders at CERN*, tech. rep. CERN-ACC-2018-0037, CERN, 2018, URL: https://cds.cern.ch/record/2645151 (cit. on p. 3).

[3] RD50 collaboration, URL: https://rd50.web.cern.ch/rd50 (cit. on pp. 5, 79, 84).

[4] 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. 5, 79).

[5] 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. 7, 10).

[6] S. Agostinelli et al., *GEANT4: A Simulation toolkit*, Nucl. Instrum. Meth. **A506** (2003) 250 (cit. on p. 17).

[7] 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 p. 24).

[8] P. Nason, *A New method for combining NLO QCD with shower Monte Carlo algorithms*, JHEP **11** (2004) 040, arXiv: hep-ph/0409146 (cit. on p. 24).

[9] 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 p. 24).

[10] 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 p. 24).

[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 p. 42).

[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. 42).

[13] ATLAS Collaboration, *Performance of missing transverse momentum reconstruction with the ATLAS detector in the first proton–proton collisions at* $\sqrt{s} = 13$ *TeV*, ATL-PHYS-PUB-2015-027, 2015, URL: https://cds.cern.ch/record/2037904 (cit. on p. 44).

[14] A. Ferrari, P. R. Sala, A. Fasso and J. Ranft, *FLUKA: A multi-particle transport code*, tech. rep., 2005 (cit. on p. 48).

[15] *Beam Background Generator*, 2010, URL: http://bbgen.web.cern.ch/bbgen/ (cit. on p. 48).

[16] R. Kwee-Hinzmann et al., *Machine-induced Background Simulation Studies for LHC Run 1, Run 2 and HL-LHC*, tech. rep. CERN-ACC-2017-0025, CERN, 2017, URL: https://cds.cern.ch/record/2261862 (cit. on p. 49).

[17] 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. 53, 195, 196, 200).

[18] G. Avoni et al., *The new LUCID-2 detector for luminosity measurement and monitoring in ATLAS*, J. Inst. **13** (2018) P07017 (cit. on pp. 53, 195).

[19] 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. 53).

[20] 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. 54).

[21] 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. 54).

[22] 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. 54, 55).

[23] 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. 55).

[24] P. A. M. Dirac, *Quantised singularities in the electromagnetic field*, Proc. Roy. Soc. Lond. **A133** (1931) 60 (cit. on p. 57).

[25] G. 't Hooft, *Magnetic Monopoles in Unified Gauge Theories*, Nucl. Phys. **B79** (1974) 276, [,291(1974)] (cit. on p. 57).

[26] 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. 57).

[27] 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. 57).

[28] 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. 57).

[29] 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. 57).

[30] 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. 57).

[31] 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. 62, 187, 189).

[32] 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. 80, 85, 95, 96).

[33] 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. 80).

[34] 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. 82).

[35] G. Kramberger et al., *Radiation effects in Low Gain Avalanche Detectors after hadron irradiations*, JINST **10** (2015) P07006 (cit. on pp. 84, 85).

[36] 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. 84, 85).

[37] 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. 85, 86, 95, 96, 118).

[38] 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. 85, 95, 96).

[39] 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. 85, 96, 103).

[40] Y. Zhao et al., *Comparison of 35 and 50 µm thin HPK UFSD after neutron irradiation up to* 6×10^{15} n_{eq}/cm^2, (2018), arXiv: 1803.02690 (cit. on pp. 85, 96, 103).

[41] 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. 85, 96).

[42] 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. 85, 86, 94–96, 98, 118).

[43] F. Cenna et al., *Weightfield2: A fast simulator for silicon and diamond solid state detector*, Nucl. Instrum. Meth. **A796** (2015) 149 (cit. on p. 86).

[44] Synopsys, URL: https://www.synopsys.com/silicon/tcad/device-simulation/sentaurus-device.html (cit. on p. 86).

[45] Agapopoulou et al., *A prototype ASIC for the ATLAS High Granularity Timing Detector*, In preparation () (cit. on pp. 109, 131, 144).

[46] 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. 110).

[47] 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. 112).

[48] F. Faccio et al., 'FEAST2: A Radiation and Magnetic Field Tolerant Point-of-Load Buck DC/DC Converter', *2014 IEEE Radiation Effects Data Workshop (REDW)*, IEEE, 2014 (cit. on p. 113).

[49] *Radiation tolerance of the 130 nm TSMC technology*, URL: https://asicsupport-community.web.cern.ch/t/radiation-tolerance-of-the-130nm-tsmc-technology/144 (cit. on p. 113).

[50] 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. 119).

[51] 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. 153).

[52] CAEN Electronic Instruments, *DT55xxE Desktop HV Power Supply*, URL: https://www.caen.it/products/dt5521he/ (cit. on p. 154).

[53] Tektronix, *TDR Impedance Measurements: A Foundation for Signal Integrity*, URL: http://www.tek.com/dl/55W_14601_2.pdf (cit. on p. 154).

[54] KC705 Evaluation Board. Xilink Inc., URL: https://www.xilinx.com/support/documentation/boards_and_kits/kc705/ug810_KC705_Eval_Bd.pdf (cit. on p. 156).

[55] G. Beck and G. Viehhauser, *Analytic model of thermal runaway in silicon detectors*, Nucl. Instrum. Meth. **A618** (2010) 131 (cit. on pp. 165, 167, 168).

[56] L. Amaral et al., *The versatile link, a common project for super-LHC*, Journal of Instrumentation **4** (2009) P12003 (cit. on pp. 175, 178).

[57] 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. 187).

[58] 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. 191).

[59] 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. 191).

[60] 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. 192).

[61] CMS Collaboration, *CMS Luminosity Measurements for the 2016 Data Taking Period*, (2017) (cit. on p. 195).

[62] 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. 195, 198).

[63] P. Grafstrom and W. Kozanecki, *Luminosity determination at proton colliders*, Prog. Part. Nucl. Phys. **81** (2015) 97 (cit. on p. 196).

[64] 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. 215, 221, 248).

[65] *EDMS 1866761*, tech. rep. (cit. on p. 247).

[66] M. Huhtinen, *Method for estimating dose rates from induced radioactivity in complicated hadron accelerator geometries - Write-up of the FIASCO code*, tech. rep., 1998, URL: https://cds.cern.ch/record/2046812 (cit. on p. 247).

[67] *EDMS 1765924*, tech. rep. (cit. on p. 249).

[68] *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 p. 264).

[69] 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. 264).

[70] ATLAS Collaboration, *ATLAS Computing Acknowledgements*, ATL-GEN-PUB-2016-002, URL: https://cds.cern.ch/record/2202407 (cit. on p. 281).