



ATLAS Note

ATL-COM-LARG-2018-008

2018-04-16



Draft version 1.2

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

The ATLAS Collaboration

The large increase of pileup is one of the main experimental challenges for the HL-LHC physics program. A new powerful way to mitigate the effects of the pileup is to use high-precision timing information to distinguish between collisions occurring very close in space but well-separated in time. A High-Granularity Timing Detector, based on low gain avalanche detector technology, is proposed for the ATLAS Phase-II upgrade. Covering the pseudorapidity region between 2.4 and 4.0, with a timing resolution of 30 ps for minimum-ionising particles, this device will significantly improve the performance in the forward region. 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.

20 **Contents**

21 1 Introduction	3
22 2 Detector requirements	4
23 2.1 Beam conditions at the HL-LHC	5
24 2.2 Time resolution	7
25 2.3 Radiation hardness	7
26 3 Physics motivation	8
27 3.1 Performance for object reconstruction	9
28 3.2 Luminosity measurement	18
29 3.3 Use of the HGTD in the trigger system	20
30 3.4 Impact on example physics analyses	21
31 3.5 Outlook and path towards TDR	27
32 4 Detector design	29
33 4.1 Design optimisation	29
34 4.2 Module and layer design and assembly	33
35 4.3 Sensors	41
36 4.4 Readout electronics	51
37 4.5 Outlook and path towards TDR	71
38 5 Detector installation and infrastructure	72
39 5.1 Cooling system	73
40 5.2 Detector cold vessel and global structure	78
41 5.3 Moderator	79
42 5.4 Services	79
43 5.5 Outlook and path towards TDR	81
44 6 Organisation, cost and schedule	83
45 7 Outlook	87

46 1 Introduction

47 The high-luminosity phase of the Large Hadron Collider (HL-LHC) at CERN is scheduled to start in 2026
 48 and will deliver an integrated luminosity of up to 4000 fb^{-1} over the decade that follows. The instantaneous
 49 luminosity will reach up to $7.5 \times 10^{34} \text{ cm}^{-2} \text{ s}^{-1}$, corresponding to an increase by an approximate factor
 50 of 5 compared to the typical luminosities of Run 2. Two extended periods without physics operation,
 51 Long Shutdown 2 in 2019–2020 and Long Shutdown 3 from 2024 until mid 2026, are anticipated prior
 52 to the HL-LHC phase. Significant upgrades of the ATLAS detector will take place during these periods
 53 in order to cope with the high-radiation environment and the large increase in the number of collisions
 54 per bunch crossing (pileup). The ATLAS Phase-II Scoping Document [1] provides an overview of the
 55 Phase-II plans.

56 Pileup is one of the main challenges at the HL-LHC. In the nominal operation scheme, the interaction
 57 region will have an Gaussian spread of 45 mm along the beam axis¹ and a pileup of 200 simultaneous
 58 p_T interactions on average ($\langle \mu \rangle = 200$), corresponding to an average interaction density of 1.8 colli-
 59 sions/mm.

60 A major challenge for the tracking detectors is to efficiently reconstruct the charged particles created in
 61 the primary interactions and assign them to the correct production vertices. This requires the resolution of
 62 the longitudinal track impact parameter (z_0), provided by the Inner Tracker (ITk), to be much smaller than
 63 the inverse of the average pileup density (0.6 mm). The z_0 resolution is well below this limit in the central
 64 region, but becomes very large in the forward region, reaching up to 5 mm for particles with low transverse
 65 momentum (p_T). As a result, tracks cannot be associated to the correct vertices in an unambiguous way,
 66 leading to reduced performance in terms of heavy-flavour tagging, lepton isolation and the identification
 67 of jets originating from pileup interactions.

68 A powerful new way to address this challenge is to exploit the time spread of the collisions in each
 69 bunch crossing to distinguish between tracks originating in collisions occurring very close in space but
 70 well-separated in time. This requires the ability to measure the time of individual tracks with a precision
 71 much smaller than the spread of the collision times. In the nominal operating scheme of the HL-LHC, this
 72 distribution has an Gaussian spread of 175 ps. This timing information is complementary to the spatial
 73 information and kinematic measurements provided by the tracker and calorimeters, and would thus help
 74 resolve ambiguities.

75 In this context, the High-Granularity Timing Detector (HGTD) is proposed. With an expected time res-
 76 olution for minimum-ionising particles (MIPs) of approximately 30 ps, corresponding to the performance
 77 of the currently available technology, this device will be able to assign each incident charged particle to
 78 an interaction vertex with significantly improved accuracy, effectively reducing the amount of pileup by a
 79 factor of $175/30 \approx 6$.

80 The HGTD is a unique new tool to enhance the physics performance capabilities of the ATLAS detector
 81 in the endcap and forward regions, enabling similar performance for the reconstruction of forward jets,
 82 leptons, and the tagging of heavy-flavour jets as is expected in the central region. In addition, the
 83 HGTD offers unique capabilities for the online and offline luminosity determination, and can provide a
 84 minimum-bias trigger.

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 pseudorapidity is defined in terms of the polar angle θ as $\eta = -\ln \tan(\theta/2)$.

85 The proposed HGTD will complement the ITk to provide:

- 86 • Significantly improved pileup mitigation up to $|\eta| = 4$, covering a very important region for vector-
boson fusion (VBF) processes.
- 88 • Timing information for nearly all primary vertices.
- 89 • Timing determination for charged particles in the region $2.4 < |\eta| < 4.0$.
- 90 • Powerful luminosity measurements for each bunch crossing separately, including valuable estimates
91 of background processes.

92 Since the Scoping Document [1], the ATLAS HGTD group has made significant progress on the impact
93 of time measurements on the global object-level performance, the sensitivity of physics measurements
94 and the conceptual detector design. These developments are presented in this document, and intensive
95 R&D on sensors and front-end electronics is still ongoing. Most of the object-level results are presented
96 for several scenarios corresponding to the time resolution expected at different points during the lifetime
97 of the detector, as it evolves with radiation damage and replacement of the inner radii of the HGTD.

98 This document is organised as follows. Section 2 gives an overview of the requirements on the detector
99 design, including details of the relevant HL-LHC operation parameters and radiation levels. Section 3
100 summarises the physics motivation for the HGTD, presenting the resulting improvements in object re-
101 construction, luminosity measurement capabilities and impact on selected physics analyses. The details
102 on the conceptual detector design, the sensor specifications and results from irradiation, laboratory and
103 beam tests, followed by the design and initial results for the readout electronics are described in Section
104 4. Section 5 describes the installation procedures, the infrastructure for cooling and power system and
105 integration in the ATLAS cavern. Finally, Section 6 discusses the costs and schedule of the project
106 activities before an outlook towards the next steps is presented in Section 7.

107 2 Detector requirements

108 The space available to install new detectors in front of the ATLAS endcap calorimeters is limited and this
109 constrains the location and acceptance of the HGTD. The minimal inner radius of the HGTD cannot be
110 less than 110 mm, so as to allow the endcap cryostat to move over a pump located on the beam pipe when
111 opening and closing the detector. The outer radius of the back cover is limited to 1100 mm, and it needs
112 to house the detector electronics, moderator and space for routing services to the outside. Taking these
113 constraints into account, the radial extent of the active area is 120 mm to 640 mm, yielding an acceptance
114 in pseudorapidity from 2.4 to 4.0. The envelope in z for the full detector including supports and front
115 and rear covers is 75 mm. In addition, to protect the ITk and the HGTD from back-scattered neutrons,
116 50 mm of moderator material will be installed in front of the endcap calorimeters, as in the current
117 ATLAS detector. The detector will be located at $z = \pm 3.5$ m, in the place that is currently occupied by
118 the Minimum-Bias Trigger Scintillators (MBTS), just outside the ITk volume and in front of the endcap
119 and forward calorimeters, as shown in Figure 1 and Table 1.

120 A silicon-based timing detector technology is preferred due to the space limitations. The sensors must
121 be thin and configurable in arrays. In close collaboration with RD50 [2] and manufacturers, an extensive
122 R&D program is progressing quickly towards sensors that provide the required timing resolution in harsh
123 radiation environments. Low Gain Avalanche Detector (LGAD) [3] pixels of $1.3 \times 1.3 \text{ mm}^2$ with an

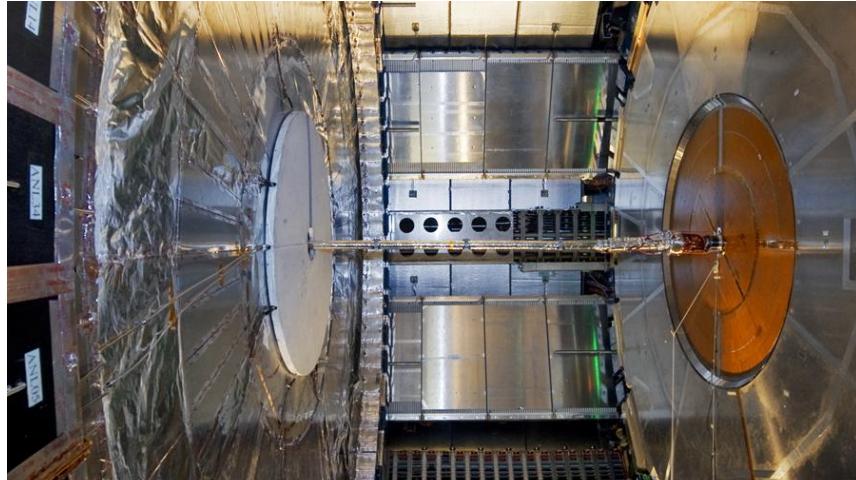


Figure 1: A photo of the current ATLAS layout (opened for maintenance), showing the gap between the ATLAS endcap calorimeter on the left and the tracking detectors on the right, where the HGTD will be installed. Currently the space is occupied by the MBTS (white disk, in front of the endcap calorimeter).

Pseudorapidity coverage	$2.4 < \eta < 4.0$
Thickness in z	75 mm (+50 mm moderator)
Position of active layers in z	$3435 \text{ mm} < z < 3485 \text{ mm}$
Radial extension:	
Total	$110 \text{ mm} < R < 1000 \text{ mm}$
Active area	$120 \text{ mm} < R < 640 \text{ mm}$
Time resolution per track	30 ps
Number of hits per track:	
$2.4 < \eta < 3.1$	2
$3.1 < \eta < 4.0$	3
Pixel size	$1.3 \times 1.3 \text{ mm}^2$
Number of channels	3.54M
Active area	6.3 m^2

Table 1: Main parameters of the HGTD.

active thickness of 50 μm fulfil these requirements. This pixel size ensures occupancies below 10% at the highest expected levels of pileup, small dead areas between pixels, and low sensor capacitance which is important for the time resolution. A custom application-specific integrated circuit (ASIC), 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 three (two) hits per track at $R < 320 \text{ mm}$ ($R > 320 \text{ mm}$). A detailed description of the detector layout is presented in Section 4.

2.1 Beam conditions at the HL-LHC

The beam-spot characteristics of the HL-LHC have not yet been determined. At present, the longitudinal beam-spot size (Gaussian width) is expected to be between 30 to 60 mm, while the width in time could

Not reviewed, for internal circulation only

136 be between 175 and 260 ps. The case considered in the studies presented here is the *nominal* one, with
 137 Gaussian spreads in z and time of 45 mm and 175 ps respectively.

138 The spatial pileup line density, i.e. the number of collisions per length unit in the z direction during one
 139 bunch crossing, is a key quantity for evaluating the performance of ATLAS with and without the HGTD.
 140 For an average of 200 collisions per bunch crossing, denoted $\langle\mu\rangle = 200$, an average pileup density of
 141 1.8 collisions/mm is expected. This average masks the effect of the local variations illustrated in Figure 2.
 142 The local pileup vertex density is calculated by summing the number of collisions in a window of ± 3 mm
 143 around the signal vertex for $\langle\mu\rangle = 200$. This is large enough to avoid quantisation effects and small enough
 144 to probe the tails of the distribution. The other curve is obtained by scaling which effectively increases
 145 the window size.

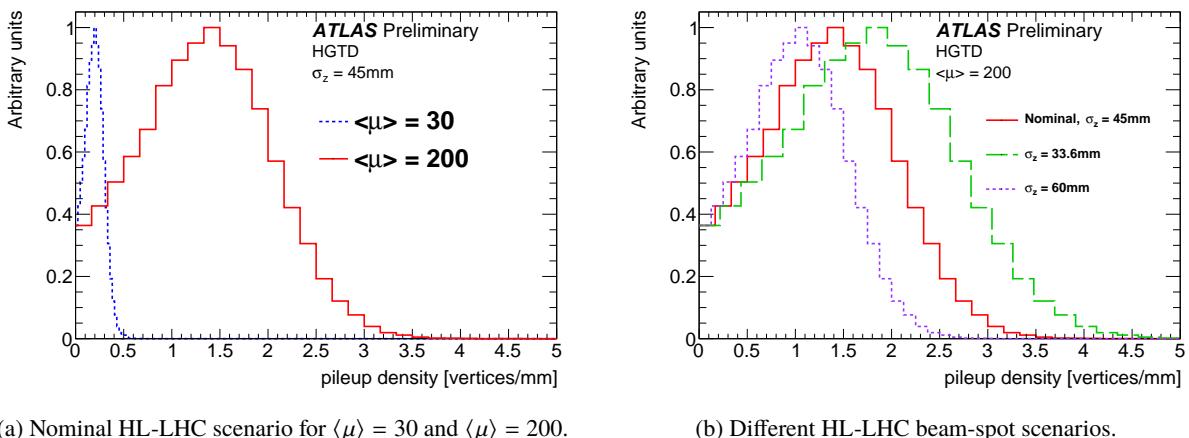


Figure 2: Local pileup vertex densities for (a) different values of $\langle\mu\rangle$, and (b) different HL-LHC beam-spot scenarios.

146 Figure 2(a) shows the pileup densities for $\langle\mu\rangle = 30$ and 200 for the same beam spot size, the latter
 147 corresponding to the nominal beam-spot scenario at the HL-LHC. The most probable local pileup density
 148 for this scenario is around 1.44 collisions/mm, corresponding to an increase by a factor six with respect
 149 to $\langle\mu\rangle = 30$. In Figure 2(b) the local pileup density is presented for the nominal and two alternative
 150 beam-spot scenarios at the HL-LHC. They differ in vertex density due to the difference in the size of the
 151 beam spot along z . All three distributions are at significantly higher values than $\langle\mu\rangle = 30$, and all feature
 152 most probable values in the range 1–2 collisions/mm.

153 Timing information can supplement the tracker z_0 measurement in assigning tracks to vertices and mitigate
 154 the impact of a high vertex density. To illustrate this, an example is presented in Figure 3, which shows a
 155 single event with 200 interactions in the $z-t$ plane, where each ellipse corresponds to a truth vertex. The
 156 extent of each ellipse is 30 ps in time and 1 mm in z . The vertical dotted lines indicate the z position
 157 of the reconstructed primary vertices in the event. The red ellipse and the red solid line indicate the
 158 hard-scatter vertex. The tracker sees the event as a one-dimensional projection on the z axis, where a large
 159 number of tracks from vertices occurring at different times but very close in space lead to ambiguities in
 160 the track-to-vertex association. This happens when the distance between vertices is of the same order or
 161 smaller than the resolution of the longitudinal impact parameter of the track, which happens more often
 162 for tracks in the forward region. With the addition of the HGTD, ATLAS can view the event in two
 163 dimensions, making it possible to resolve the ambiguities of tracks belonging to different interactions.

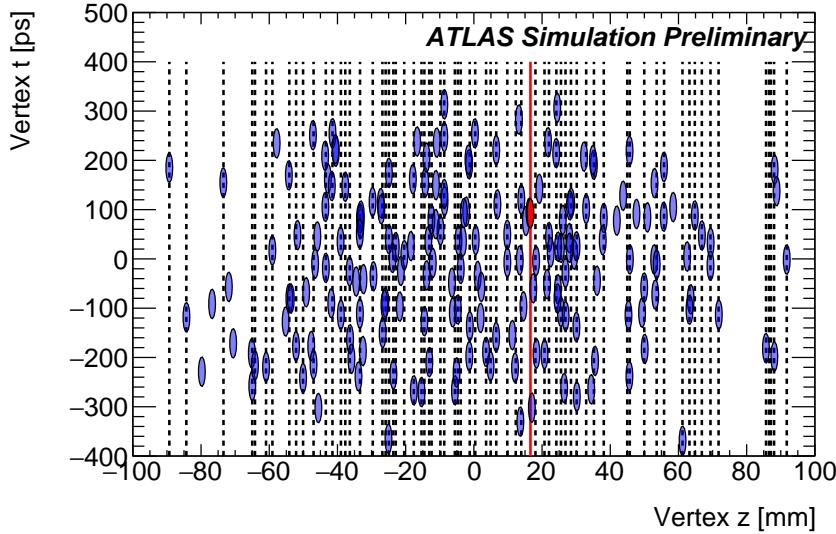


Figure 3: Visualisation in the $z-t$ plane of an event with a hard scatter (red ellipse) with about 200 pileup interactions (blue ellipses) superimposed. The dashed vertical lines represent the positions of reconstructed vertices.

164 2.2 Time resolution

165 The goal of the detector design is to provide the best possible time resolution in order to effectively
 166 suppress the effects of pileup in the forward region. A time resolution of 30 ps per track has been shown
 167 to be achievable in test beam studies and would provide a factor of 6 improvement in the track-to-vertex
 168 association. This per-track resolution is therefore established as a requirement for the detector design.

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

$$\sigma_{\text{total}}^2 = \sigma_L^2 + \sigma_{\text{elec}}^2 + \sigma_{\text{clock}}^2 \quad (1)$$

170 where σ_L^2 are Landau fluctuations in the deposited charge as the charged particle traverses the sensor, σ_{elec}^2
 171 are the contributions from the readout electronics, and σ_{clock}^2 is the clock contribution. Beam tests and
 172 sensor simulations show that thinner silicon reduces the contribution from Landau fluctuations. With a
 173 50 μm thick LGAD sensor, this contribution amounts to approximately 25 ps. With fast detector signals
 174 and a high signal-to-noise ratio, the contribution from the electronics can be kept to approximately 25 ps.
 175 This is achievable only if applying corrections for the time walk induced by different signal amplitudes,
 176 using small bins in the time-to-digital conversion and applying precise in-situ inter-calibration. The details
 177 of the design of the readout electronics to achieve this are described in Section 4.4. The clock contribution
 178 is required to be below 10 ps; its distribution is discussed in more detail in Section 4.4.12.

179 2.3 Radiation hardness

180 At the end of the HL-LHC (4000 fb^{-1}), the maximum neutron-equivalent fluence at a radius of 120 mm
 181 should reach $4.9 \times 10^{15} \text{ n}_{\text{eq}}/\text{cm}^2$ and the total ionising dose (TID) will be about 3.6 MGy, as shown in
 182 Figure 4. A safety factor of 1.5 is applied to both numbers to account for uncertainties in the simulation.
 183 An additional factor of 1.5 is applied to the TID due to uncertainties in the behaviour of the electronics

Not reviewed, for internal circulation only

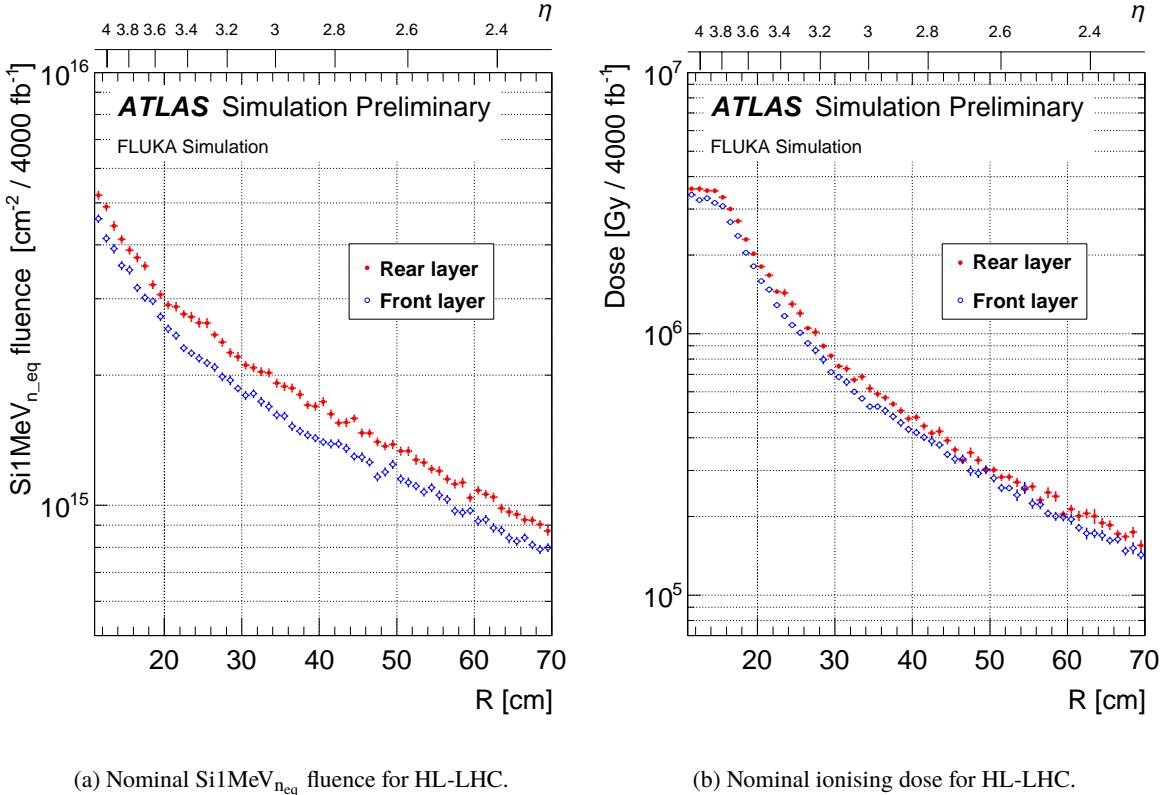


Figure 4: Nominal fluence and ionising dose as functions of the radius in the innermost and outermost sensor layers of the HGTD for 4000 fb^{-1} , i.e. before including safety factors. The η range shown at the top corresponds to the z -location of the second layer.

This amount of radiation damage to lowest-radius sensors and electronics suggests that this innermost part of the detector should be replaced after half of the HL-LHC program. The plan is therefore to replace the sensors and ASICs located at a radius up to about 320 mm ($3.08 < |\eta| < 4.0$). This corresponds to about 32% of the sensors and ASICs. Consequently, the sensors and ASICs will be exposed to a maximum of $3.7 \times 10^{15} \text{ n}_{\text{eq}}/\text{cm}^2$ and 4.1 MGy. More details can be found in Sections 4.3 and 4.4.

3 Physics motivation

This section discusses how the precision time measurement capability introduced by the HGTD enhances the performance for tagging jets, tagging b -jets, and calculating lepton isolation in the forward region, as well as the resulting impact of these improvements on the sensitivity of a few selected physics analyses. Additional applications of the HGTD, including luminosity measurements and trigger usage, are also presented.

198 The studies done to optimise the detector design and to measure track-level performance are based on full
 199 simulation and reconstruction. In the full simulation of the HGTD, each sensor is implemented as a single
 200 homogeneous volume. The active area implemented in the full simulation is $120 \text{ mm} < R < 600 \text{ mm}$
 201 (smaller than the 640 mm of the proposed detector design), with a granularity of $0.5 \times 0.5 \text{ mm}^2$. Four
 202 layers of sensors were implemented, with a distance of approximately 8 mm between each layer. Different
 203 granularities, as well as the inactive inter-pixel zones in the sensors and the distribution of the modules in
 204 the layers (readout row layout, overlap between modules, etc) are implemented downstream at the analysis
 205 level. Differences between this simplified geometry, and a much more detailed description at the GEANT
 206 level, are expected to have a minor effect on the simulated results.

207 In most of the object-level performance studies, full simulation and reconstruction of the tracker, calor-
 208 imeters and muon system is combined with a fast simulation of the HGTD which smears the nominal
 209 track times with the expected per-track resolution. This does not include the expected contribution of the
 210 clock distribution to the resolution. The efficiency of assigning hits to a track is also taken into account,
 211 based on the studies presented in section 3.1.1. In the jet performance studies, the time of the hard-scatter
 212 vertex, t_0 , is assumed to be known with a precision much smaller than 30 ps. In general, for a timing
 213 detector which is not hermetic the ability to accurately determine t_0 depends on the acceptance and the
 214 event topology. For HGTD in particular, this ability depends on how many tracks from the hard-scatter
 215 vertex are within $2.4 < \eta < 4.0$ and their p_T distribution. Dedicated studies of t_0 reconstruction and the
 216 resulting impact on physics-object performance are planned for the Technical Design Report (TDR).

217 The expected timing performance for this smearing is discussed in Section 4.1 where four scenarios are
 218 defined. The first one, referred to as *Initial*, corresponds to the expected performance at the beginning
 219 of the HL-LHC operation. The second one, denoted *Intermediate*, takes the expected timing resolution
 220 degradation due to radiation damage after half of the HL-LHC program (2000 fb^{-1}) into account. The
 221 *Final* scenario represents the expected timing resolution at the end of the HL-LHC program (4000 fb^{-1})
 222 assuming that the sensors and readout electronics at $R < 320 \text{ mm}$ are replaced after half of the integrated
 223 luminosity has been delivered. The *Worst Case* scenario assumes a timing resolution of 60 ps per hit for
 224 the sensors, independent of the radius. This is worse than the expected resolution of the innermost sensors
 225 after 4000 fb^{-1} without replacement.

226 The performance for several physics objects is also studied as a function of η to show the improvement in
 227 the endcap region and compare it to that of the central region to motivate the chosen acceptance of the
 228 HGTD.

229 3.1 Performance for object reconstruction

230 The precision timing measurement of the HGTD can improve the overall performance for reconstructing
 231 physics objects. The key to these improvements relies on the use of timing information, in addition to
 232 spatial (z) information, to associate tracks to vertices with high efficiency and low pileup contamination.
 233 The use of the HGTD for track-to-vertex association is described below, followed by studies that quantify
 234 how the HGTD improves pileup-jet suppression, b -tagging, and electron isolation in the forward region.

235 3.1.1 Track-level performance

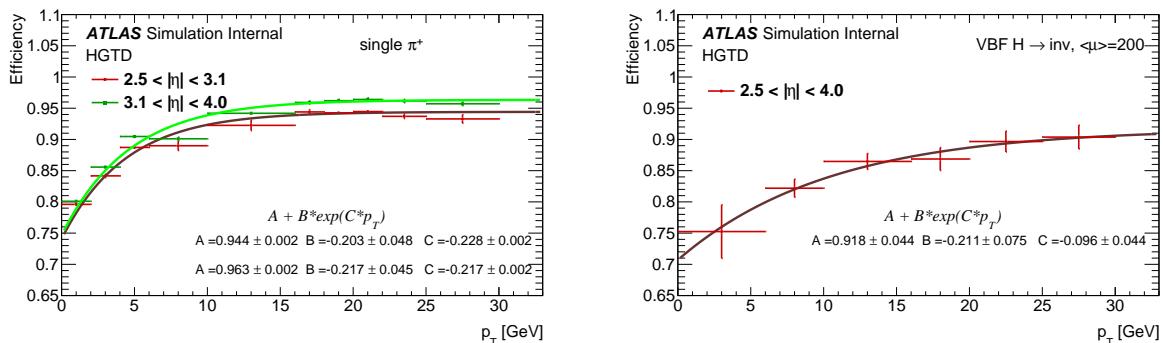
236 Track extrapolation to the HGTD

237 The matching of tracks reconstructed with the ITk to hits in the HGTD has been studied using single-pion

238 samples ($\langle \mu \rangle = 0$) and a physics sample with VBF-produced $H \rightarrow Z(\nu\nu)Z(\nu\nu)$ at $\langle \mu \rangle = 200$, both with
 239 full simulation of the HGTD. The matching is performed by extrapolating the tracks to the HGTD; only
 240 tracks with an extrapolation within the HGTD region are used. The hits in the pixels that are within a
 241 radius of 5 mm of the track extrapolation in the transverse plane are considered, and the closest one is
 242 matched.

243 The distribution of the minimum distance between the track extrapolation and a hit can be described as a
 244 Gaussian core with non-Gaussian tails. In the case of pions, the fraction of tracks outside the core varies
 245 between 15% at low p_T to about 5% at high p_T . For $p_T = 2$ GeV, the track extrapolation is typically within
 246 1 mm of the hit, and improves at higher p_T . It is limited primarily by the pixel size.

247 Figure 5(a) shows the efficiency for matching tracks with at least one HGTD pixel as a function of the
 248 reconstructed track p_T for samples of single pions with p_T between 0.5 and 20 GeV. The figure shows
 249 two curves, one for $R < 320$ mm (the region with higher overlap between modules), corresponding to
 250 $|\eta| < 3.1$ and the other for $R > 320$ mm. Tracks with $|\eta| < 2.5$ or $|\eta| > 4.0$ are not considered in order
 251 to avoid border effects. An exponential fit is performed to both sets of points, and the obtained values
 252 of the parameters are presented in the figure. These functions are used in order to take into account the
 253 effect of the efficiency in the performance studies presented in the following sections. At high p_T , the
 254 efficiency approaches 95%, limited by non-instrumented regions. At low p_T , the efficiency decreases to
 255 about 80%, due to the increased effect of interactions with the material in front of the HGTD. This material
 256 corresponds approximately to 0.2 interaction lengths at $|\eta| = 2.4$, and increases to about 0.4 interaction
 257 lengths at $|\eta| = 3.8$. The comparison of the two curves shows a small $|\eta|$ dependence for all p_T values,
 258 the higher $|\eta|$ region presenting an efficiency approximately 3% higher than the lower one.



(a) Efficiency for matching tracks to at least one HGTD pixel. (b) Estimated efficiency for assigning the correct time to tracks.

Figure 5: Efficiencies for matching tracks to at least one pixel in the HGTD in single-pion events and for correctly reconstructing the time of a track in a high-pileup environment. An exponential fit is performed on the points, and the obtained values of the parameters are shown.

259 With a pileup of $\langle \mu \rangle = 200$, almost all tracks have hits closer than 5 mm; because most of them correspond
 260 to pileup interactions, timing information needs to be used to assign hits to the track. The hits matched to a
 261 track are selected so as to be the closest to the track in the layers that have consistent timing between them.
 262 They must be within 5 mm of the extrapolation, and it is required that their RMS deviation from their
 263 average time is smaller than 60 ps. The efficiency for correctly assigning a time to a track in a high-pileup
 264 environment is shown in Figure 5(b) as a function of track p_T . The dependence of the efficiency with p_T
 265 is similar to that obtained with single-pion samples, though 4% lower. There is a slight decrease for large
 266 $|\eta|$, due to the increased flux of pileup tracks, increasing the probability of an incorrect hit association.

267 The hit-to-track matching efficiency in the HGTD was also measured in a sample of single muons with
 268 $p_T = 45$ GeV, and found to be 98% independent of η . The higher efficiency with respect to pions is due
 269 to the reduced probability for the muons to interact with material in front of the HGTD.

270 This study is a first step in the development of the association of the track information with timing
 271 information in the HGTD. The main inefficiency for the association has been identified as interactions in
 272 the material in front of the HGTD. In the future, more sophisticated pattern recognition algorithms should
 273 be studied to improve this efficiency.

274 Track-to-vertex association

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

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

$$\frac{|z_0 - z_{\text{vertex}}|}{\sigma_{z_0}} < 2, \quad (2)$$

281 where σ_{z_0} is the per-track resolution on the longitudinal impact parameter and depends primarily on the
 282 track η and p_T . Figure 6 shows a parameterization of σ_{z_0} as a function of η , for different p_T values. This
 283 parameterization was obtained as the inter-quartile range of the impact parameter resolution. The choice
 284 of $2\sigma_{z_0}$ as selection cut is to ensure a track selection efficiency of 95% independent of η and p_T .

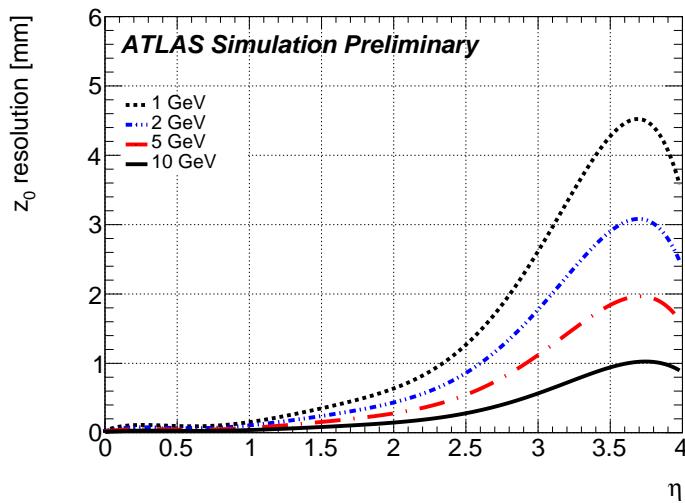


Figure 6: The resolution of the longitudinal track impact parameter, z_0 , as a function of η for different p_T values.

285 The reliability of the track-to-vertex association depends on the value of σ_{z_0} relative to the average pileup
 286 density $\langle \rho(z) \rangle$. The average number of interactions within a window of $|z_0 - z_{\text{vertex}}| = 2\sigma_{z_0}$ is given
 287 by $N = 2\langle \rho(z) \rangle 2\sigma_{z_0}$. This means that in order to unambiguously associate tracks to vertices based on
 288 Equation. (2), N has to be smaller than 1, or $\sigma_{z_0} < 1/4\langle \rho(z) \rangle$. For a track with z_0 at the origin, where the
 289 average density peaks, this value is approximately 160 μm . If σ_{z_0} is larger than this value, the association

290 of tracks to vertices becomes ambiguous because one same track may be compatible with multiple nearby
 291 vertices. It is important to note that this argument applies to prompt tracks such as those produced in
 292 light-quark and gluon jets, or prompt leptons. The association of displaced tracks from the decay of
 293 B/D hadrons to primary vertices requires the use of a larger z window, which will lead to greater pileup
 294 contamination.

295 While the longitudinal impact parameter resolution is relatively constant and small for $|\eta| < 1.5$, it grows
 296 rapidly with pseudorapidity, reaching very large values of several millimetres for $|\eta| \gtrsim 2.5$. The resolution
 297 is further degraded at low track p_T due to multiple scattering effects. The η dependence of the impact
 298 parameter resolution is mostly determined by the geometry of the inner detector. As η increases, tracks
 299 become more collinear to the beam line. The form of the distribution, including the improvement at the
 300 largest η is related to the material distribution in the tracker.

301 Based on Figure 6, a 1 GeV track with $\eta = 3$ has a z_0 resolution of approximately 2.5 mm, leading to a
 302 $\pm 1\sigma$ window of 5 mm in z for the vertex position. With a most probable average pileup vertex density of
 303 1.8 vertices/mm at $z = 0$, this means that, on average, a forward track can be compatible with up to about
 304 13 near-by vertices on average. Or, in other words, track-to-vertex association will suffer significantly
 305 from pileup contamination.

306 This effect is illustrated in Figure 7(a), which shows the pileup-track contamination around a hard-scatter
 307 vertex in a $R-z$ view. All tracks displayed in this figure have been assigned to the reconstructed primary
 308 vertex (drawn in the centre of the figure) according to Eq. (2). All tracks are required to pass standard
 309 quality selections [4] and have $p_T > 0.9$ GeV. Figure 7(b) shows a similar view of the same event but it
 310 includes only the tracks associated with calorimeter jets with $p_T > 20$ GeV. This figure clearly shows the
 311 presence of a jet entirely made of pileup tracks (shown in blue), that cannot be identified as such because
 312 all tracks have been assigned to the hard-scatter vertex. It also shows the large reduction of pileup when
 313 only considering tracks inside jets. This is due to the effect of the reduced jet area (A^{jet}), which, for the
 314 case of the anti- k_t algorithm with distance parameter $R = 0.4$ and under the assumption that pileup tracks
 315 are uniformly distributed in pseudorapidity, is given by

$$\frac{A^{\text{jet}}}{2\pi \cdot 2 \cdot \Delta\eta} \rho(z) \cdot \Delta z^{\text{trk,vtx}} = \frac{\pi \cdot 0.4^2}{2\pi \cdot 2 \cdot 4.0} \rho(z) \cdot 2\sigma_{z_0}.$$

316 As an example, for a $R = 0.4$ anti- k_t jet in the forward region ($|\eta| > 2.4$) under nominal HL-LHC beam
 317 conditions, this fraction is about 10%.

318 Figure 8 shows different views of the same event. A two-dimensional distribution of the reconstructed times
 319 and z positions of the p_T -weighted tracks is shown in 8(a), while 8(b) and 8(c) show one-dimensional
 320 projections in z and time, respectively. While the high- p_T (and predominantly central) hard-scatter
 321 and pileup tracks are visibly separated in z , forward low- p_T tracks are indistinguishable if only spatial
 322 information is considered. The time projection in Figure 8(c), on the other hand, shows that both vertices
 323 can be effectively resolved with a time resolution of 30 ps.

324 The only handle for the forward tracker to address the challenge of pileup on track-to-vertex association in
 325 the forward region is to tighten the selection cuts used in Eq. (2), reducing the track-to-vertex association
 326 efficiency. The HGTD detector can extend the track-to-vertex association capability of the tracker in the

327 central region to the forward region by the use of two-dimensional z - t selection criteria. In addition to

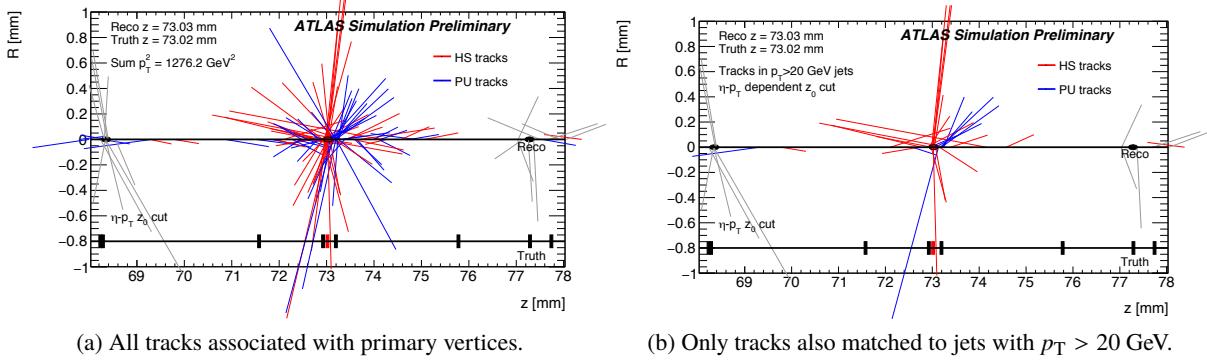
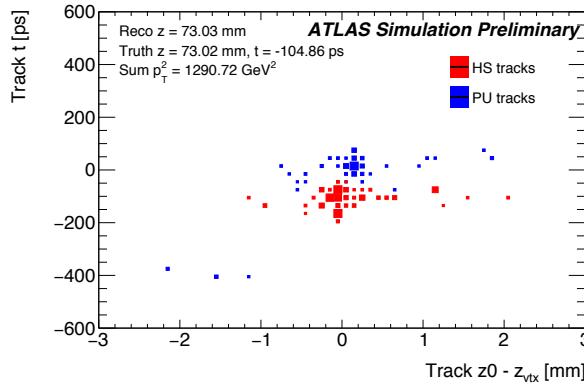
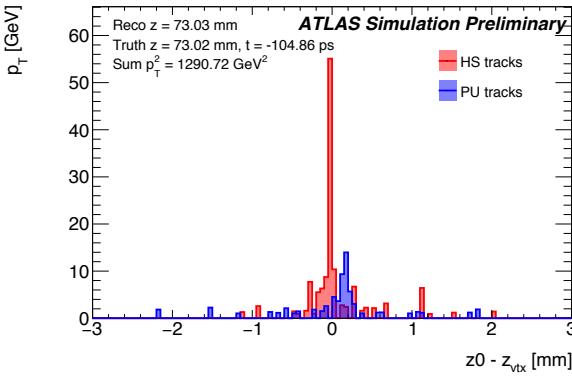


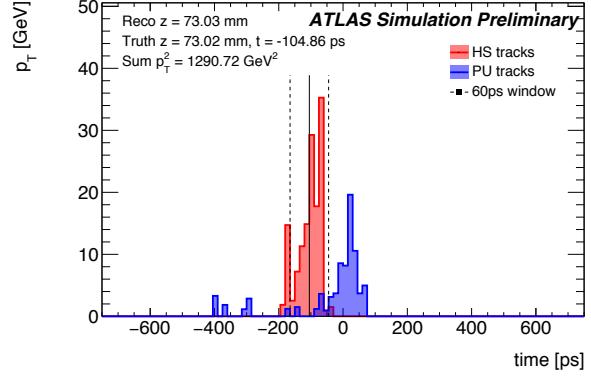
Figure 7: View in the R - z plane of the tracks associated to the primary vertices in a VBF-produced $H \rightarrow Z(\nu\nu)Z(\nu\nu)$ event at $\langle\mu\rangle = 200$. The length of each line is proportional to the track p_T . The positions of the reconstructed and truth vertices are indicated at $R = 0$ and $R = -0.8$ mm, respectively. Grey lines show tracks attached to other (pile-up) reconstructed vertices.



(a) Reconstructed times and z positions.



(b) Reconstructed z positions.



(c) Reconstructed times.

Figure 8: Distributions of reconstructed times and z positions of the p_T -weighted tracks associated to the hard-scatter vertex of an event with a VBF-produced Higgs boson decaying invisibly with $\langle\mu\rangle = 200$.

³²⁸ Eq. (2), the following selection is introduced,

$$\frac{|t - t_0|}{\sigma_t} < 2, \quad (3)$$

³²⁹ where t is the track time measured by the HGTD, and σ_t is the track time resolution, assumed to be 30 ps
³³⁰ constant in p_T and η . The vertex time, t_0 , is assumed to be known with high accuracy by averaging the
³³¹ track time of high- p_T good-quality tracks attached to the primary vertex. The use of this two-dimensional
³³² track-to-vertex association allows to maintain high track-selection efficiency while greatly reducing the
³³³ contamination of pileup tracks within the z selection window. The use of timing information in the
³³⁴ forward region, hence, can compensate for the worsening of the z_0 impact parameter resolution, allowing
³³⁵ the tracker to maintain good track-to-vertex association performance for all η , mitigating the impact of
³³⁶ the large pileup contamination in the forward region. The methods used here give correct results as long
³³⁷ as there are enough high- p_T tracks within geometrical acceptance of the HGTD. Since this requirement
³³⁸ depends on the event topology, dedicated studies are required to determine the impact of a non-ideal
³³⁹ t_0 reconstruction for different physics processes. Even in the cases where t_0 cannot be measured with
³⁴⁰ certainty, pattern recognition algorithms may still be capable of separating tracks in jets originating from
³⁴¹ different pile-up interactions for jets containing more than one track, resulting in improved performance
³⁴² in jet reconstruction. Given the complexity, the development of these algorithms is beyond the scope of
³⁴³ this document, but a high priority moving towards the TDR.

³⁴⁴ The following sections show how the enhanced pileup-mitigation capability of the HGTD in the forward
³⁴⁵ region impacts the identification and rejection of pileup jets, the tagging of heavy-flavour jets, and the
³⁴⁶ efficiency for lepton-isolation requirements.

³⁴⁷ 3.1.2 Suppression of pileup jets

³⁴⁸ Pileup is one of the most difficult challenges for object identification under HL-LHC conditions. Particles
³⁴⁹ produced in pileup interactions can contaminate the jets of interest coming from the hard-scatter vertex,
³⁵⁰ thereby reducing the accuracy of the jet energy determination. Pileup interactions can also produce
³⁵¹ additional jets which do not originate from the primary hard-scatter interaction. These pileup jets can
³⁵² be produced as the result of a hard QCD process from a pileup vertex, or by random combinations of
³⁵³ particles from multiple vertices. At low jet p_T , the latter mechanism is dominant, whereas at high jet p_T ,
³⁵⁴ the majority of pileup jets are QCD jets.

³⁵⁵ Pileup jets can reduce the precision of Standard Model measurements and the sensitivity to discover new
³⁵⁶ physics. For example, additional jets can increase the amount of background events passing a selection, as
³⁵⁷ well as reduce the efficacy of kinematic variables or discriminants to separate signals from backgrounds.
³⁵⁸ Hence, the efficient identification and rejection of pileup jets is essential to enhance the physics potential
³⁵⁹ of the HL-LHC.

The key element to suppress pileup jets is the accurate association of jets with tracks and primary vertices.
A simple but powerful discriminant for pileup-jet suppression is the R_{p_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 PV₀,
divided by the fully calibrated jet p_T , i.e.

$$R_{p_T} = \frac{\sum p_T^{\text{trk}}(\text{PV}_0)}{p_T^{\text{jet}}}.$$

360 The tracks used to calculate R_{p_T} fulfil the quality requirements defined in Ref. [5] and are required to have
 361 $p_T > 1$ GeV. The distance between the hard-scatter vertex and the longitudinal impact parameter of the
 362 tracks used in the R_{p_T} calculation is required to be within 1 mm and 4 mm, depending on the $|\eta|$ of the
 363 track.

364 Hard-scatter and pileup jets for simulated events are defined by their matching to truth jets, which are
 365 reconstructed from stable and interacting final state particles coming from the hard interaction. The
 366 matching criteria are defined in Ref. [6]. Reconstructed hard-scatter jets are required to be within
 367 $\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2} < 0.3$ of a truth jet with $p_T > 10$ GeV. The pileup jets must be at least $\Delta R > 0.6$
 368 away from any truth jet with $p_T > 4$ GeV. The performance has been studied using a mixture of full
 369 reconstruction (for tracks and jets) and fast simulation (for the HGTD).

370 At moderate levels of pileup, where track impact parameter measurements can be used to assign tracks
 371 to vertices with relatively little ambiguity, small values of R_{p_T} correspond to jets which have a small
 372 fraction of charged-particle p_T originating from the hard-scatter vertex PV₀. These jets are therefore
 373 likely to be pileup jets. However, at high pileup conditions, and particularly in the forward region, the
 374 power of this discriminant is reduced. The effect can be mitigated by including timing information from
 375 the HGTD, removing tracks outside a $2\sigma_t$ window around the time of the hard-scatter vertex, as shown in
 376 Figure 8(c). This figure shows an example event where a jet originating from a pileup interaction (in blue)
 377 is misidentified as a hard-scatter jet when only using tracking information. All tracks from the pileup
 378 jet are compatible with the hard-scatter vertex within the z_0 resolution. The use of timing information
 379 in the R_{p_T} discriminant can suppress this pileup jets, improving the overall performance of pileup-jet
 380 suppression.

381 Figure 9 shows the rejection² of pileup jets as a function of the efficiency for selecting hard-scatter jets
 382 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
 383 the HGTD for the different timing resolution scenarios. A significant improvement in performance of
 384 up to a factor of 4 higher pileup-jet rejection at constant efficiency is achieved with the use of timing
 385 information. This study was done using the smearing functions to simulate the HGTD performance and
 386 taking into account the matching efficiency.

387 It is possible to use a p_T and $|\eta|$ dependent requirement on R_{p_T} to define different working points at fixed
 388 efficiencies on hard-scatter and pileup jets. Figure 10 shows the hard-scatter jet efficiency for a fixed pileup
 389 jet efficiency, ϵ_{PU} , of 2% (i.e. a rejection factor of 50), as a function of $|\eta|$ of the jet. The HGTD recovers
 390 the 10-30% drop in efficiency observed in the forward region, allowing to maintain similar pileup jet
 391 suppression performance as in the central barrel over a large part of its acceptance. Due to the width of the
 392 jets, the HGTD improves the efficiency also for jets at slightly lower $|\eta|$ than its geometrical acceptance.
 393 The effect is more evident for high- $|\eta|$ and low- p_T jets.

394 3.1.3 Tagging of heavy-flavour jets

395 The efficient identification of b -jets and high rejection of light-quark jets is of central importance in the
 396 HL-LHC physics program. Tagging b -jets is particularly sensitive to pileup-track contamination. This
 397 is due to the fact that b -tagging algorithms consider tracks with large impact parameters (in both the
 398 transverse and longitudinal directions) from the decay of displaced vertices. With a larger z_0 window,

² Throughout this document, the rejection is defined as the inverse of the mis-tag efficiency.

Not reviewed, for internal circulation only

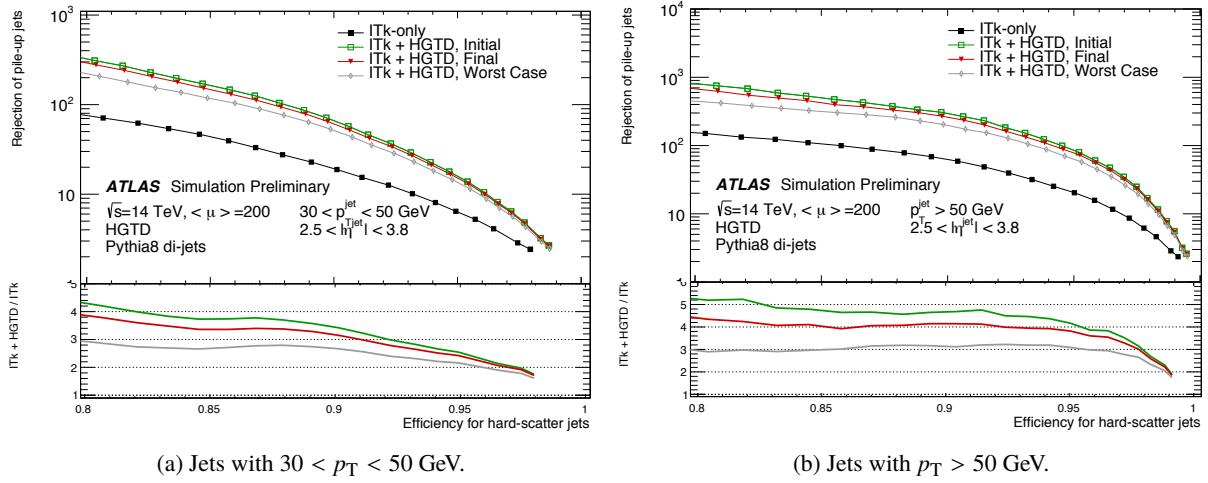


Figure 9: Pileup-jet rejection as a function of hard-scatter jet efficiency in the $2.4 < |\eta| < 4.0$ region, for the ITk-only and ITk + HGTD scenarios with different time resolutions.

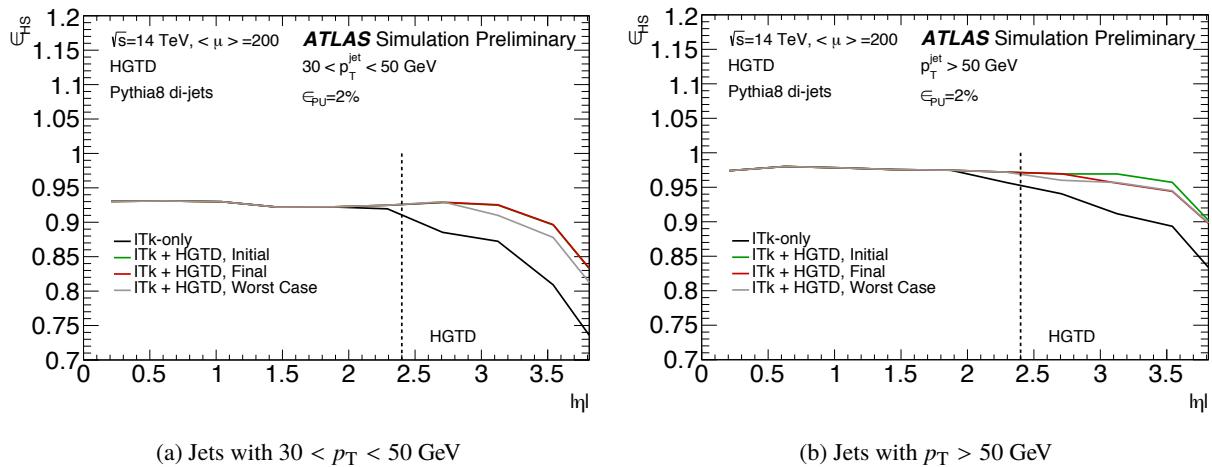


Figure 10: Hard-scatter jet efficiency versus $|\eta|$ for a 2% pileup-jet efficiency using p_T and $|\eta|$ requirements on the R_{p_T} discriminant, in dijet events.

399 tracks from nearby pileup interactions are more likely to be selected, leading to an increased rate of
 400 misidentified light-quark jets.

401 Using simulated $t\bar{t}$ events at $\langle\mu\rangle = 200$ and the fast HGTD simulation described above, the impact of the
 402 HGTD on the performance of a multivariate b -tagging algorithm is studied for forward jets ($|\eta| > 2.4$).
 403 Figure 11(a) shows the light-jet rejection versus b -tagging efficiency for the MV1 b -tagging algorithm.
 404 The addition of the HGTD removes the majority of pileup tracks from the track selection. As a result, the
 405 performance of the b -tagger is significantly improved. For a b -tagging efficiency of 70% and 85%, the
 406 corresponding light-jet rejection for MV1 is increased by approximate factors of 1.5 and 1.2, respectively.
 407 These factors could be greater for processes where more b -jets are expected in the forward region.
 408 Figure 11(b) shows the light-jet mis-tag efficiency for a 70% b -tagging efficiency working point as a

Not reviewed, for internal circulation only

409 function of jet $|\eta|$. In both figures, the performance is shown for the ITk-only scenario as well as three
 410 scenarios with HGTD timing performance representing different stages of the HL-LHC program. It can
 411 be seen that all timing scenarios yield significant improvements in the performance, even in the *Worst*
 412 *Case* scenario. Importantly, significant improvements are observed also after the full radiation damage
 413 expected during HL-LHC operation.

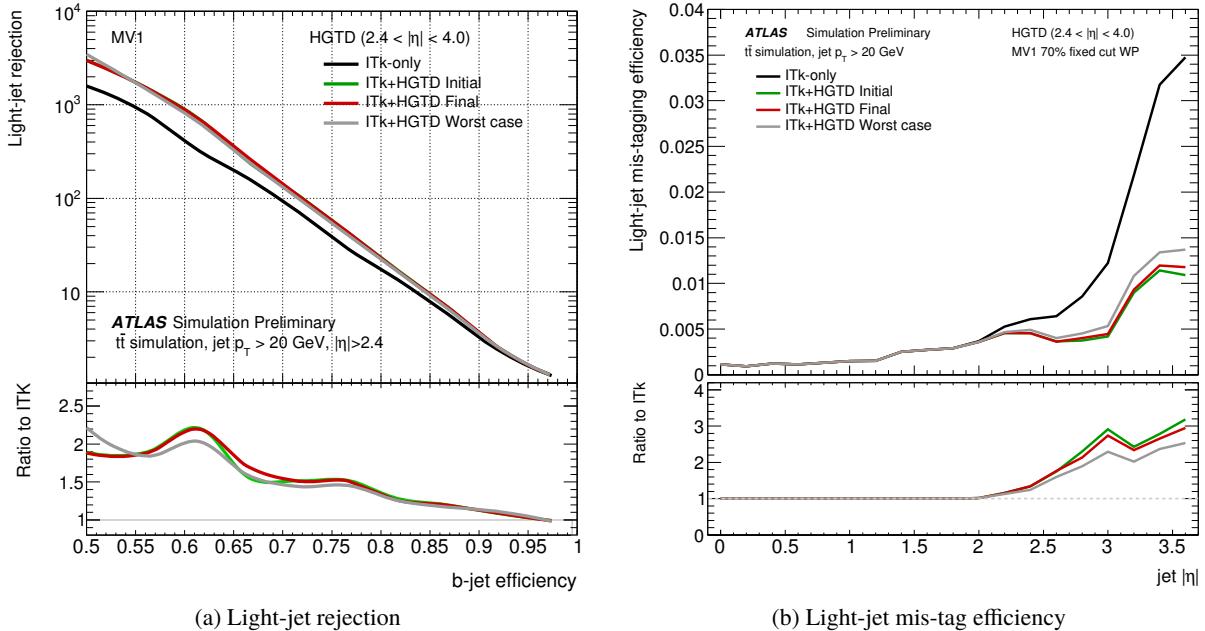


Figure 11: (a) Light-jet rejection versus b -tagging efficiency for the MV1 tagger and (b) Light-jet mis-tag efficiency for a 70% b -tagging efficiency working point as a function of jet $|\eta|$. The study uses $t\bar{t}$ events at $\langle \mu \rangle = 200$ and shows the achieved performance for different time resolution scenarios. The ratio plots at the bottom show the relative performance achieved with the HGTD with respect to the ITk-only scenario.

414 3.1.4 Lepton isolation

415 In this section, studies of how the electron isolation efficiency is improved with HGTD are presented
 416 based on the full simulation of the HGTD. The HGTD can be used to assign a time to leptons in the
 417 forward region. This information can be exploited to reject tracks which come from other interactions
 418 but are spatially close to the energy deposits in the calorimeter and/or the track associated to the lepton.
 419 The timing information can reject additional tracks from interactions close in z , according to Eq. (2) but
 420 separated in time from the hard-scatter vertex, as in Eq. (3). The isolation efficiency is defined as the
 421 probability that no track with $p_T > 1$ GeV is reconstructed within $\Delta R < 0.2$ of the electron track.
 422 The isolation efficiency is shown in Figure 12 as a function of the local vertex density for the ITk-only
 423 scenario and three HGTD timing resolution scenarios. While the efficiency drops strongly with the increase
 424 of the pileup density when using only the ITk, the addition of the HGTD timing information reduces this
 425 drop, keeping an efficiency above 90% even at high pileup density, i.e. with up to three additional
 426 vertices around the hard-scatter vertex. For a local pileup density of the order of 1.6 vertices/mm the
 427 electron isolation efficiency is improved by about 14%. Even in the *Final* timing scenario, the resolution

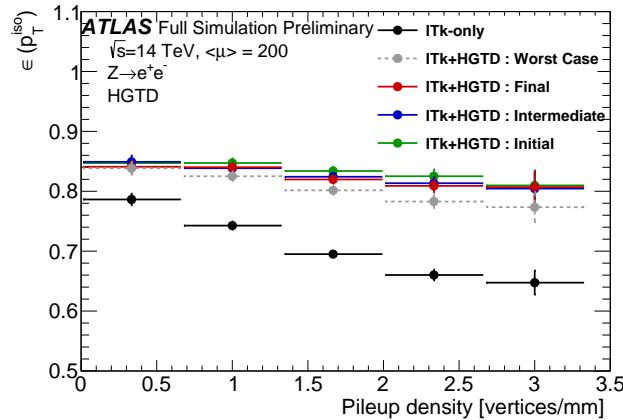


Figure 12: Selection efficiency for the electron isolation criteria as a function of the pileup density using the ITk and ITk + HGTD for different timing resolutions.

428 is sufficient to achieve an isolation efficiency essentially independent of the pileup density at the end of
 429 the HL-LHC. The performance in the forward region reaches a level similar to that in the central region.
 430 The improvement of the performance is essentially independent of the timing scenario, rendering further
 431 improvements to the resolution past 30 ps/track unjustified for this usage.

432 In addition to these studies of lepton isolation improvements³, the impact of the HGTD on the rejection of
 433 jets mis-identified as electrons is being studied, along with potential optimisations of the isolation criteria
 434 natively exploiting the timing information from the HGTD.

435 3.2 Luminosity measurement

436 Precise luminosity determination will be a critical component of the precision measurements that make
 437 up cornerstones of the HL-LHC physics program. For example, achieving $O(1\%)$ accuracy on certain
 438 measurements of Higgs boson production and its couplings can be limited by the luminosity uncertainty. It
 439 is therefore important to be able to determine the luminosity at least as accurately as done in Run I [7] and
 440 II of the LHC, and this will be a challenge at the harsh environment at the HL-LHC. The increased pileup
 441 leads to increased detector occupancies, posing serious problems for the technologies used traditionally
 442 for luminometers.

443 As a fast high-granularity detector in the forward region, the HGTD provides unique capabilities for
 444 measuring the luminosity at the HL-LHC. With detector signal durations in the few-ns range, the charged-
 445 particle multiplicities within the acceptance can be determined accurately for each individual bunch
 446 crossing separately. The high granularity gives a low occupancy, and therefore excellent linearity between
 447 the average number of hits and the average number of simultaneous pp interactions over the full range of
 448 luminosity expected at the HL-LHC.

449 The instantaneous luminosity can vary between each Bunch Crossing ID (BCID) in the LHC. For the best
 450 precision of the total delivered luminosity, accurate measurements of these bunch-by-bunch variations

³ If a high- η muon-tagger were to be added to the Phase-II upgrade program for ATLAS, the HGTD would be able to provide similar isolation efficiency improvements for muons in the forward region.

are needed. In the Phase-II ATLAS detector, some luminometers can only measure the bunch-integrated luminosity (e.g. integrated current measurements in the Tile or LAr calorimeters) whereas other more dedicated luminometers such as LUCID and BCM measure the luminosity for each bunch. All these luminometers make fast estimations of the luminosity available online. Other methods for determining the bunch-by-bunch luminosity such as measuring the number of pixel hits, the number of reconstructed tracks, or the number of reconstructed vertices, are only available after a dedicated offline analysis. As HL-LHC is expected to operate with luminosity levelling, there is additional need for an accurate and fast online luminosity measurement to be sent to the machine. 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 principle is the same as for measuring luminosity through counting reconstructed tracks [7] or clusters in a pixel detector [8]. If the HGTD is read out independently of the ATLAS trigger, it has the potential to provide an accurate estimate of the bunch-by-bunch luminosity and make it available both online and offline. With detailed occupancy information at 40 MHz, i.e. for every bunch crossing, the HGTD will thereby allow unbiased high-statistics 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 $320 \text{ mm} < R < 640 \text{ mm}$ (equivalent to $2.4 < |\eta| < 3.1$) for the luminosity determination. The exact choice of inner radius for the luminosity data can perhaps be optimised further.

3.2.1 Linearity of luminosity determination

For the proposed range $|\eta| < 3.1$, the average number of hits per layer per pp collision is 44.6, and approximately 7% of the events have no hits. Figure 13(a) shows the average number of hits per event registered in the first layer (both sides of the innermost cooling plate) of the HGTD as a function of the number of simultaneous 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 pixel. A linear fit made 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 properly simulated high-pileup samples.

3.2.2 Non-linear effects for HGTD luminosity determination

There are two main sources of non-linear effects to consider, due to multiple particles passing through the same pixel and due to noise or background from so-called *afterglow*, caused primarily by activated material in the detector volume, which is distributed approximately uniformly in time. The HGTD is expected to have a low occupancy, which will mitigate the non-linearity of the luminosity measurement from multiple particles passing through the same pixels. Especially at the outer radii, the effect of multiple hits will be negligible. It is also possible to correct for the multiple-hits effect by parameterizing the probability for this to occur as a function of $\langle\mu\rangle$. The time resolution of the HGTD can reduce the effects of noise and background activity, e.g. from activated material in the detector volume. By reading out the occupancy before and after the time window during which particles from the collisions arrive, an estimate of this component can be obtained, for each BCID separately. This estimate can then be used to increase the precision of the luminosity determination by subtracting the occupancy which is expected to be due to noise and background activity. The ability to do an in-situ measurement of the noise and background

Not reviewed, for internal circulation only

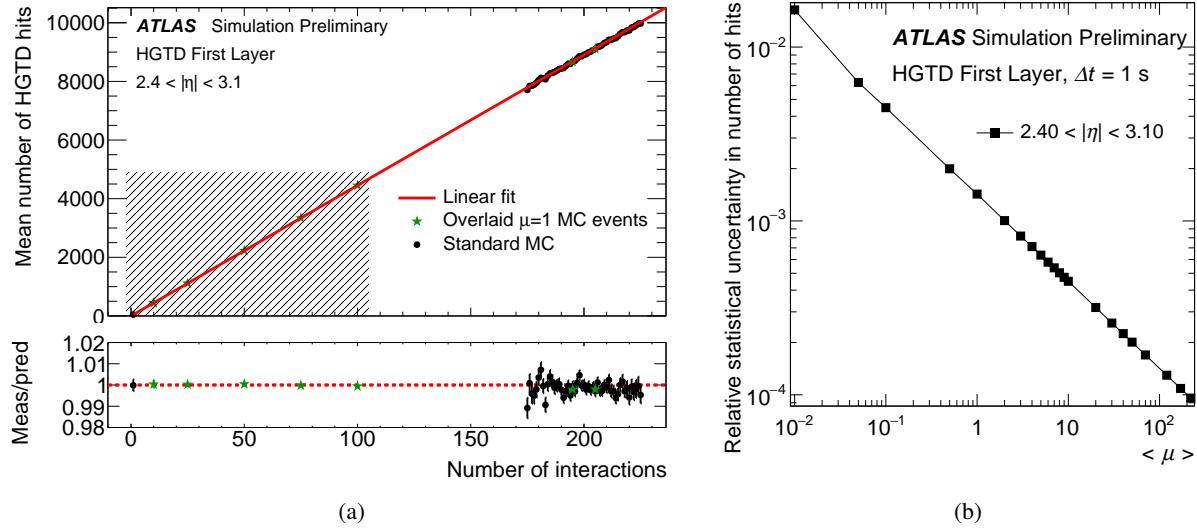


Figure 13: (a) Linearity of $\langle n_{\text{hits}} \rangle$ as a function of μ , and (b) statistical uncertainty as a function of $\langle \mu \rangle$.

level for each bunch crossing and BCID separately is a unique capability of the HGTD compared to other luminometers.

3.2.3 Statistical precision of the luminosity determination

Offline luminosity measurements are limited by systematic uncertainties. To confirm that statistical uncertainties are subdominant also for the short time periods used for online luminosity measurements, this 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 μ . 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. Figure 13(b) shows the relative uncertainty expected from statistical fluctuations as a function of $\langle \mu \rangle$ using this method. The coverage of $|\eta| < 3.1$ 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.

3.3 Use of the HGTD in the trigger system

3.3.1 Minimum-bias trigger at Level-0

The data made available at 40 MHz for the luminosity measurements can also be used by the Level-0 (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.

513 The HGTD will be installed where the current MBTS detector is located. The MBTS detector has been
 514 used extensively for these purposes during Run-1 and so far in Run-2, e.g. during the heavy-ion runs
 515 where it played a crucial role. However, it will not survive at the HL-LHC. With improvements of several
 516 orders of magnitude in both granularity and time resolution, the HGTD can provide all the functionality of
 517 the MBTS, and much more. The number of hits in the time window centered around the nominal collision
 518 time provides good separation between empty bunch crossings and those with pp collisions, and it would
 519 be straightforward to implement a trigger condition based on this information. The latency for reaching
 520 the Level-0 global trigger processors in time for a decision is not expected to be a problem but needs to
 521 be studied in detail.

522 3.3.2 Potential for improved pileup subtraction algorithms at Level-0 and HLT

523 The data made available at 40 MHz for luminosity estimates could also provide a real-time estimate of $\langle \mu \rangle$
 524 to the hardware-based trigger systems and facilitate pileup subtraction earlier in the trigger. If propagated
 525 to the Level-0 trigger system, this information could be used to implement more precise triggers based
 526 on jets and missing transverse momentum (E_T^{miss}), allowing lower rates and/or lower thresholds. The
 527 simplest pileup subtraction scheme that could be implemented within the CaloGlobal system⁴ is an
 528 average correction, where an offset dependent on $\langle \mu \rangle$ and jet kinematics (η, p_T) is subtracted from the
 529 jet p_T [9]. The disadvantage of such a scheme is that it does not capture event-by-event fluctuations.
 530 The pileup subtraction scheme that is currently being used for jets reconstructed both offline and at the
 531 HLT involves event-by-event information about the average pileup event density and individual jet area,
 532 as detailed in Ref. [9]. This method does not rely on tracking information, and it could be implemented
 533 in the CaloGlobal system if the algorithms that derive such quantities are sufficiently fast, since the full
 534 event is available within FPGAs. However, the jet-area correction is not optimal in the forward region,
 535 due to the coarser segmentation of the calorimeter that prevents a precise calculation of the pileup event
 536 density for $|\eta| > 2$, and due to the increased sensitivity of the forward calorimeter signals to out-of-time
 537 pileup. An offset correction, employing $\langle \mu \rangle$ information from HGTD, would therefore be beneficial not
 538 only to restore stable jet performance in presence of pileup for forward jets, but also to the performance
 539 of the full η range. Such a pileup correction scheme will be investigated on the timeline of the technical
 540 design report.

541 3.4 Impact on example physics analyses

542 The physics performance improvements enabled by the HGTD can enhance the physics potential of
 543 ATLAS in several ways:

- 544 • The improved suppression of pileup jets is particularly important for searches for or measurements of
 545 VBF processes, which produce forward dijet pairs with large invariant masses. Additional forward
 546 jets mimicking the VBF topology increase the background from processes like $Z+jets$. Searches
 547 for heavy Higgs bosons produced via VBF are particularly affected by forward pileup jets because
 548 the VBF jets are produced at very large rapidity and the small cross section of the process requires
 549 high efficiency for low- p_T forward jets for which effective pileup suppression is critical.

⁴ Pileup subtraction schemes to be used at the trigger level are still under discussion in the technical design report of the trigger and data acquisition systems.

- The improved b -tagging performance in the forward region can benefit physics analyses with forward b -quarks in the final state and in which the dominant backgrounds do not contain a large fraction of b -quarks in the forward region. Additionally, improved b -tagging can increase the rejection of the $t\bar{t}$ background in measurements of and searches for Higgs bosons produced via VBF.
- The improved lepton efficiency can enhance the precision of important Standard Model measurements at high luminosity that require forward leptons, such as the measurement of the weak mixing angle.
- The capability of the HGTD to assign a time to nearly all vertices can enable the reconstruction of masses of long-lived particles that decay within the HGTD acceptance. For any new massive particle that lives long enough to reach the HGTD, estimates of the increased ionisation energy loss and time-of-flight expected for a particle with $v < c$ could help identify them.
- Improvements in the luminosity accuracy and added trigger capabilities can further enhance the physics potential of many physics analyses at the HL-LHC.

While the HGTD is expected to bring improvements to a broad range of physics analyses, this section illustrates three cases: measurement of VBF $H \rightarrow WW^*$, $tH(\rightarrow b\bar{b})$ and the improvement of $\sin^2 \theta_W$. The studies presented here are not exhaustive and the impact of the HGTD is being studied in several other physics analyses. They have been performed using the smearing functions mentioned before, following the recommendations for upgrade studies.

3.4.1 VBF $H \rightarrow WW^*$

High-precision measurements of the Higgs sector are one of the primary physics goals of the HL-LHC. VBF Higgs boson production can be computed with small theoretical uncertainties and is therefore a good channel for measuring the couplings between the Higgs boson and electroweak gauge bosons. The forward jet topology is a distinct feature which can be used to separate signal processes from many backgrounds. The precision of the VBF $H \rightarrow WW^*$ signal strength of $\mu = 1.27^{+0.44}_{-0.40}(\text{stat.})^{+0.30}_{-0.21}(\text{syst.})$ [10] was obtained from LHC Run-1 data at $\sqrt{s} = 7$ TeV and 8 TeV using a boosted decision tree (BDT). Increased pileup poses a challenge to the reconstruction of forward jets, and VBF final states are particularly sensitive to the precision with which jet-to-vertex association can be done in the forward region.

The object and event selection closely follows the selection used in Ref. [11]. The analysis requires electrons and muons to be isolated and have $p_T > 15$ GeV. Jets are selected with $p_T > 50$ GeV and must be identified as hard-scatter jets after passing pileup jet suppression criteria. A veto for b -jets is also applied. Relative to the selection in the referenced result, the operating point for the veto of b -tagged jets is changed from 70% to 85% efficiency. The E_T^{miss} is required to be larger than 20 GeV. The two leading jets are required to be in different hemispheres, each satisfy $|\eta_j| > 2.0$, and an invariant mass $m_{jj} > 1250$ GeV. No jets with $p_T > 30$ GeV and both leptons are required to be inside of the η window between the two leading jets. Table 2 presents the event yields.

N_{VBF}	N_{bkg}	N_{ggF}	N_{WW}	N_{VV}	$N_{t\bar{t}}$	N_t	$N_{Z/\gamma^*+\text{jets}}$	$N_{W+\text{jets}}$
203	280	39	29	7	146	25	34	0

Table 2: Expected signal and background yields after all selection requirements using the ITk, assuming 3 ab^{-1} of integrated luminosity and $\langle \mu \rangle = 200$. The background yields are summed together in the column labelled N_{bkg} .

	Signal unc.		
	Full	1/2	None
Δ_μ	0.16	0.13	0.13

Table 3: The expected Δ_μ without the HGTD is shown considering the same, half, and none of the theoretical systematic uncertainties on the VBF and ggF Higgs-boson production taken from Ref. [11].

	Δ_μ	% Improvement
ITk only	0.130	-
ITk + HGTD ($2.4 < \eta < 4.0$)	0.123	5%

Table 4: The relative improvement in Δ_μ for different η -coverage scenarios of timing measurements with 30 ps resolution.

The uncertainty on the signal strength (Δ_μ) is shown in Table 3, with the full, one-half, and none of the theoretical uncertainties for gluon-gluon fusion and VBF Higgs boson production from Ref. [11].

The improved pileup mitigation that HGTD enables is propagated to all affected physics objects. The impact on Δ_μ is shown in Table 4. The VBF $H \rightarrow WW^*$ analysis becomes more sensitive for higher pileup-jet rejection, so the enhanced pileup-jet reduction of roughly a factor of 3 leads to an improvement of the analysis sensitivity. A full $|\eta| > 0$ timing acceptance improves the Δ_μ by around 7% whereas for the nominal HGTD acceptance, $|\eta| > 2.4$, the improvement is around 5%.

A multivariate analysis is also performed using a BDT including VBF topology variables. In addition, the lepton centrality and the $\sum m_{\ell j}$ variables from Ref. [12] are included in the training, which is done before the $p_T^{\text{tot}} < 20$ GeV selection is applied⁵. After the training, the $m_{jj} > 500$ GeV is applied. The output of the BDT distribution is divided in bins with around 80 signal events per bin, which was found to be optimal. The resulting BDT distribution after including the HGTD is shown in Figure 14. The pileup jet efficiency of 2% is found to give the smallest uncertainty on the Higgs production signal strength.

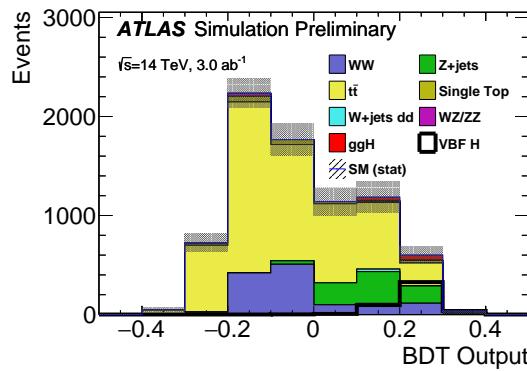


Figure 14: BDT score distribution when incorporating the HGTD.

The event and background yields are shown in Table 5. The addition of the HGTD reduces the total

⁵ The p_T^{tot} variable is defined as the magnitude of the vectorial sum $\mathbf{p}_T^{I_1} + \mathbf{p}_T^{I_2} + \mathbf{E}_T^{\text{miss}} + \sum \mathbf{p}_T^{\text{jets}}$

background in the most sensitive BDT bin by a factor of 43%. Most of the background reduction comes from increased $t\bar{t}$ rejection.

Detector scenario	N_{VBF}	N_{bkg}	N_{ggF}	N_{WW}	N_{VV}	$N_{t\bar{t}}$	N_t	$N_{Z/\gamma^*+\text{jets}}$	$N_{W+\text{jets}}$
ITk	80	65	19	15	0	21	0	10	0
ITk + HGTD	80	37	19	4	0	5	0	10	0

Table 5: The signal and background yields are shown for the most sensitive BDT bin for different scenarios.

The uncertainty on the Higgs production signal strength is computed using the four bins with the highest significance added together assuming uncorrelated background uncertainties as the analysis is mostly statistically limited. Without the use of the HGTD, Δ_μ of 0.096 is expected. The impact of the HGTD leads to an improvement of about 8%, with $\Delta_\mu = 0.088$. Most of the improvement comes from the improved b -tagging resulting in a lower top background. The improved pileup-jet rejection adds about 3% improvement.

In summary, an improvement of the order of 8% can be achieved in the BDT-based measurement of VBF-produced $H \rightarrow WW^*$ through including the HGTD.

3.4.2 Forward b -jets and measurements of tH production

Improved b -tagging in the forward region can benefit physics analyses characterised by the presence of forward b -jets. One example, considered in this section, is the $tH \rightarrow b\bar{b}$ final state.

A measurement of tH production is the only direct probe of the sign of the top-Yukawa coupling. The integrated luminosities produced by the LHC so far do not give sensitivity to this parameter, but determining it is an important goal for the HL-LHC physics program. In the Standard Model (SM), the interference between a positively signed top-Yukawa coupling and the coupling between Higgs and W bosons is destructive, resulting in a small expected tH production cross section. If the sign is negative, the interference would become constructive and the cross section for tH would increase by a factor of 10 relative to the SM process, while the $t\bar{t}H$ cross section would remain unchanged. Furthermore, an improved significance of this signal would improve the measurement on the strength of the top-Yukawa coupling, as the $t\bar{t}H$ measurement signal region is complementary to tH . The diagrams contributing to tH are shown in Figure 15.

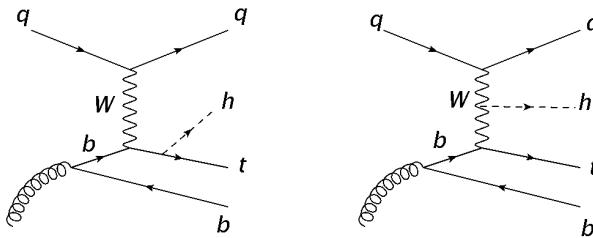


Figure 15: Contributing diagrams to tH production with the top-Yukawa coupling (left) and the W to Higgs-boson coupling (right). In both cases, a forward jet is expected, with additionally a gluon splitting to $b\bar{b}$ pair producing a b -jet in the forward region in 25% of the events.

Signal regions are defined to contain events with one lepton and four or five jets, of which at least two are b -tagged. One of the non- b -tagged jets should have $|\eta| > 2.4$. The sum of the transverse momenta of

Not reviewed, for internal circulation only

the jets H_T must be at least 300 GeV. The pseudorapidity of the most forward jet is the most important variable to separate the tH signal from the $t\bar{t}$ and $t\bar{t}H$ backgrounds which produce more central jets. The distributions of the most forward jet in events with three b -tagged jets can be found in Figure 16.

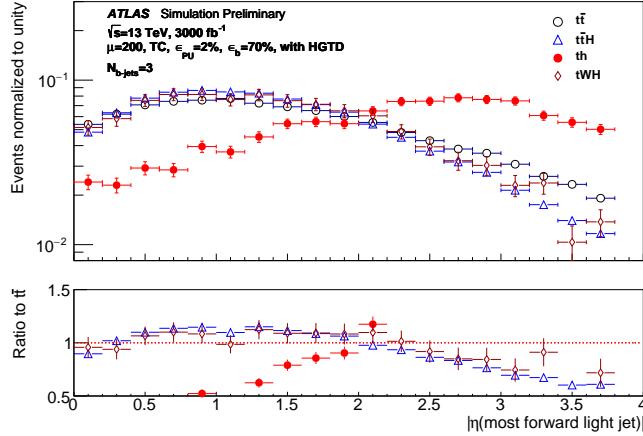


Figure 16: Distribution of $|\eta|$ for the most forward light jet in the three b -tag category for tH with $H \rightarrow b\bar{b}$ and the backgrounds from $t\bar{t}$ and $t\bar{t}H$ production.

A forward b -jet in the HGTD acceptance is present in 25% of the events. The relative gain from the improved light-jet rejection due to the HGTD is estimated using the MV1 algorithm with a working point of 70% efficiency for b -jets. Using a log-likelihood ratio (LLR) binned in η of the light forward jet, the sensitivity to tH is improved by 11% as shown in Table 6.

Scenario	Signal significance
ITk	1.28σ
ITk + HGTD	1.42σ

Table 6: Expected (statistical only) signal significance of $tH \rightarrow b\bar{b}$ using ITk or ITk + HGTD.

3.4.3 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, $\sin^2 \theta_{\text{eff}}$ [13].

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 [14] 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 [13].

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 [15]. 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 $Z/\gamma^* \rightarrow ee$ 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 $p_T > 25$ GeV. The invariant mass of the electron pair is required to be loosely consistent with the Z boson mass, $60 < m_{ee} < 200$ 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_{ee} and $|y_{ee}|$ bins. In the CF and FF channels migrations in the m_{ee} 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_{ee} and $|y_{ee}|$ bin.

Significant uncertainties arise from knowledge of the momentum scale and resolution for the electrons. Following Reference [16] 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 $E_T < 55$ GeV and up to 1.5% (2.1%) for central (forward) electron with $E_T > 100$ GeV.

The expected sensitivity to particle level A_{FB} as a function of m_{ee} is shown in green in Figure 17 for each channel for chosen rapidity bin. As expected the larger asymmetry is observed in the CF channel. The extraction of $\sin^2 \theta_{\text{eff}}$ 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 Figure 17, 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_{ee} . On the contrary, near the Z boson mass peak, the effect of varying $\sin^2 \theta_{\text{eff}}$ is maximal, while being significantly smaller at high and low masses. Thus, in this projection a global fit is performed where $\sin^2 \theta_{\text{eff}}$ is extracted while constraining at the same time the PDF uncertainties [15]. With this analysis, the expected sensitivity of the extraction of $\sin^2 \theta_{\text{eff}}$ are respectively 25×10^{-5} , 21×10^{-5} and 40×10^{-5} for the CC, CF and FF channel. The uncertainty of the results is dominated by the currently limited knowledge of the PDFs. If looking purely at the experimental uncertainties, including the HGTD in the ATLAS forward region brings a 13% improvement on the $\sin^2 \theta_{\text{eff}}$ sensitivity in the CF channel. Combining the three channels together the expected sensitivity reaches a precision of $\Delta \sin^2 \theta_{\text{eff}} = 18 \times 10^{-5} \pm 16 \times 10^{-5}$ (PDF) $\pm 9 \times 10^{-5}$ (exp.) which exceeds the precision achieved in all previous single-experiment results so far.

Not reviewed, for internal circulation only

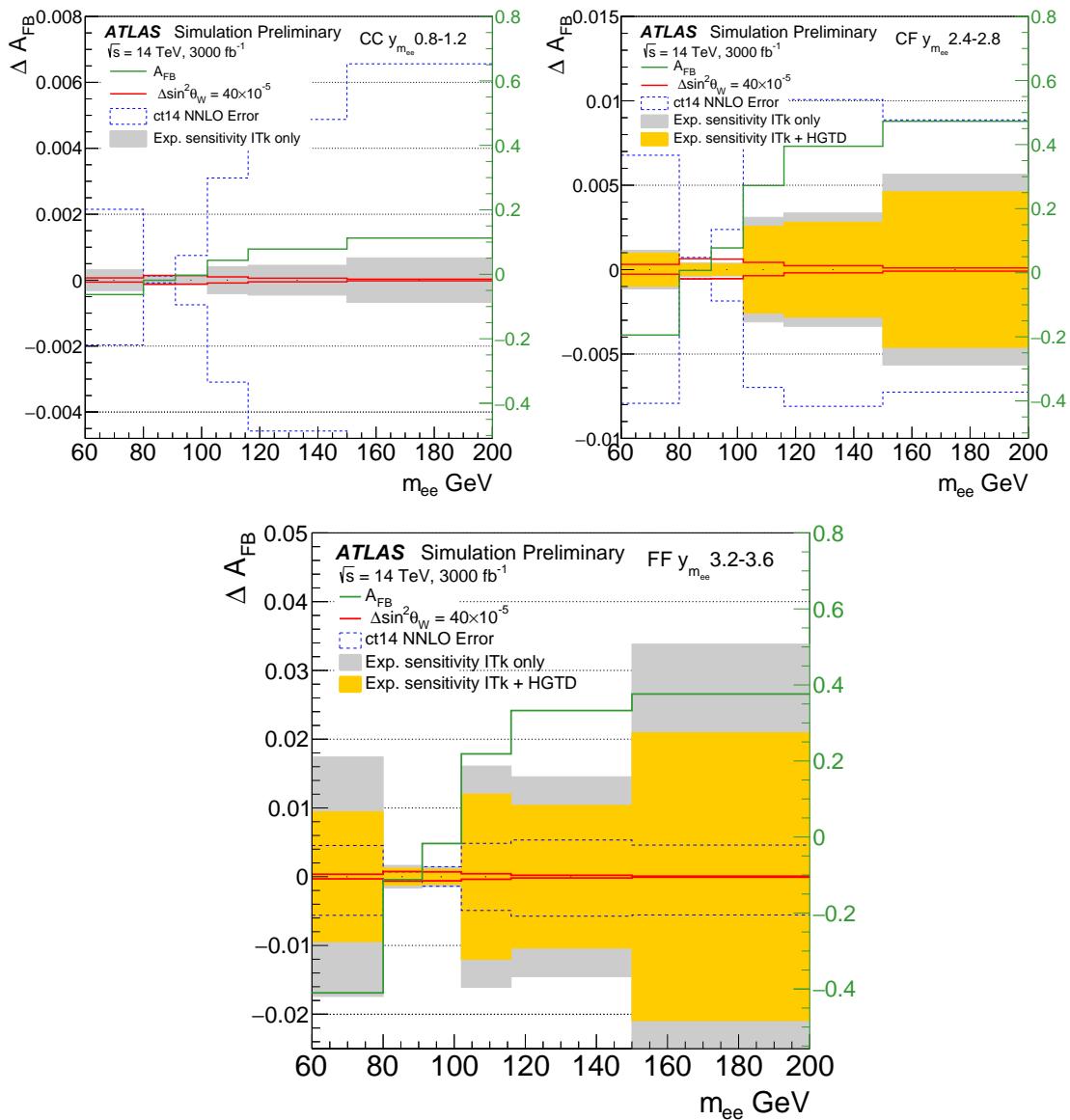


Figure 17: Distribution of ΔA_{FB} as a function of mass for the CC, CF and FF channel in different rapidity. The filled band corresponds to the experimental sensitivity with and without the HGTD. The solid red lines correspond to a variations of $\sin^2 \theta_{\text{eff}}$ corresponding to 40×10^{-5} . The band marked in dashed blue illustrates the total error form ct14 NNLO PDF. Overlay in green the particle level A_{FB} distribution.

683 3.5 Outlook and path towards TDR

684 With a timing resolution of 30 ps per track, the HGTD provides important supplementary information for
 685 mitigating pileup effects. The contamination of low- p_T forward tracks associated to the primary vertex
 686 can be mitigated by adding requirements on z_0 and the compatibility of the time measurements of the
 687 associated tracks in the HGTD.

688 With the expected timing precision, rejecting pileup tracks with the R_{p_T} algorithm presented in Sec-
 689 tion 3.1.2 improves the identification of pileup jets by up to a factor of 4. The rejection of pileup tracks

690 using timing information also improves the light-jet rejection at 70% b -tagging efficiency by a factor of
 691 1.5. The electron isolation efficiency is increased by about 14%.

692 These performance improvements translate into 8% improvement in the sensitivity of the VBF $H \rightarrow WW^*$
 693 process. This is achieved primarily through the improved b -jet veto and the improved pileup-jet suppression
 694 in the forward region provided by the HGTD. An analysis targeting tH production with a final state
 695 containing forward b -jets shows up to 11% increase in sensitivity due to better b -jet tagging in the forward
 696 region. The effect of the improvement of the electron identification performance in the measurement of
 697 the weak mixing angle, $\sin^2 \theta_{\text{eff}}$, has been investigated. The most sensitive events for this measurement
 698 are the ones where a Z -boson decays to one electron in the central region and another electron in the
 699 forward region, while the dominant systematic uncertainty comes from the limited knowledge of parton
 700 distribution functions (PDF). An improvement of 11% on the experimental uncertainty is obtained when
 701 adding the HGTD to the analysis. Potentially larger improvements on these analyses, as well as in other
 702 final states that benefit from improved pileup-jet rejection, b -jet tagging and lepton performance in the
 703 forward region, are currently being investigated.

704 The results have been obtained using full simulation and reconstruction as well as a mixture of full and fast
 705 simulation. The study of more sophisticated pattern recognition algorithms might improve the capabilities
 706 of ATLAS even further when the timing information from the HGTD is used in a combined reconstruction.
 707 In addition to the analyses described in some detail above, these capabilities can give improvements in
 708 several other measurements and searches for new physics.

709 For the object reconstruction, the HGTD can refine the track selection used in the calculation of the missing
 710 transverse momentum. Studies are ongoing to evaluate the impact on the resolution and the possible
 711 reduction of tails of the E_T^{miss} distribution, induced by tracks from pileup interactions. The improvements
 712 in the performance of jet reconstruction and the physics analyses that use jets were determined assuming
 713 an ideal vertex t_0 determination. The impact of using a realistic t_0 determination efficiency and precision
 714 is a high priority for the next steps and the TDR.

715 A more accurate luminosity determination, as well as potential improved jet and E_T^{miss} trigger capabilities,
 716 can further enhance the precision and sensitivity of a broad range of physics analyses.

717 The MBTS detector has been used for measurements of the inelastic pp cross section at $\sqrt{s} = 7$ and
 718 13 TeV [17, 18]. Despite the coarse segmentation of the MBTS, the distributions of the number of hit
 719 scintillators provide very valuable information for discriminating between different generators to improve
 720 the modelling of inelastic interactions. In the sample used for these measurements, the detector response
 721 was saturated in 90% of the events. The high granularity of the HGTD would therefore allow the level of
 722 activity for the bulk of the sample to be resolved, enabling significant improvements in the modelling of
 723 minimum-bias interactions. The same studies could also be done using tracks reconstructed by the ITk,
 724 but possibly the HGTD could give access to a lower- p_T region. The importance of accurate modelling
 725 of such processes increases linearly with pileup, and any per-event improvement is multiplied by a factor
 726 of 200 for high- $\langle \mu \rangle$ at the HL-LHC. This could lead to improvements in the forward region that could
 727 improve the E_T^{miss} modelling.

728 New long-lived particles (LLPs) are predicted by a range of beyond-SM models, including some that are
 729 particularly well-motivated from a Dark Matter perspective. Accurate timing and ionisation-energy loss
 730 measurements are key observables from heavy LLPs traversing the detector, such as e.g. R -hadrons and
 731 long-lived sleptons in Split SUSY and Gauge-Mediated Supersymmetry Breaking scenarios, respectively.
 732 Though the acceptance of the HGTD is limited to the forward region and many LLPs are predicted to be
 733 produced primarily in the central η region, some improvement in sensitivity could be gained if the HGTD

⁷³⁴ could be exploited in these searches. Studies of LLPs expected to give a signature primarily in the forward
⁷³⁵ region, such as magnetic monopoles, will be carried out on the time scale of the TDR.

⁷³⁶ Finally, the timing capability of the HGTD is being investigated for the purpose of studying machine-
⁷³⁷ induced non-collision background processes such as beam-gas interactions and beam-halo activity.

⁷³⁸ 4 Detector design

⁷³⁹ This section describes the layout and design of the HGTD. The detector is based on LGAD sensors, a
⁷⁴⁰ novel type of silicon sensor that provides the required timing resolution of 30-50 ps per hit in a compact
⁷⁴¹ and radiation hard format. A sensor is segmented into arrays of pixels or pads yielding a granularity high
⁷⁴² enough to match the requirements. The sensors are then bump-bonded to readout chips and connected to
⁷⁴³ flexible circuits to construct the modules. These modules are arranged and mounted onto thin intermediate
⁷⁴⁴ support plates, larger half-disk shaped for $R < 320$ mm and narrow rectangular *staves* for larger radii. The
⁷⁴⁵ modules located in a row along the direction of the routing of the flex cable are considered part of a so-
⁷⁴⁶ called *readout row*. The placement and spacing of the modules along the radial direction is optimised such
⁷⁴⁷ that a charged particle passing through will result in at least three (two) hits for $R < 320$ mm ($R > 320$ mm)
⁷⁴⁸ on average. The larger number of hits in the inner region, accomplished by increasing the overlap of the
⁷⁴⁹ modules on the two sides of the cooling disk, is necessary to maintain the target per-track time resolution
⁷⁵⁰ after irradiation. The cooling disks are divided into two halves to simplify installation. The active area of
⁷⁵¹ the detector spans the forward region corresponding to $2.4 < |\eta| < 4.0$, where it helps to restore the object
⁷⁵² reconstruction and identification performance to match that achieved in the central region, as presented in
⁷⁵³ Section 3.

⁷⁵⁴ The first part of this section describes the different studies and considerations that led to the optimised
⁷⁵⁵ design presented in this document. Then the assembly of the modules and detector layers is described.
⁷⁵⁶ The last two subsections of the section are dedicated to the detailed description of the LGAD sensors and
⁷⁵⁷ the readout electronics.

⁷⁵⁸ 4.1 Design optimisation

⁷⁵⁹ The studies presented in this section were performed using the full simulation and reconstruction described
⁷⁶⁰ in Section 3.

⁷⁶¹ The first step of the detector optimisation is the definition of the pixel size, which is driven by the
⁷⁶² electronics and the occupancy, i.e., the probability of having a signal in a pixel. The smaller the surface of
⁷⁶³ the pixel, the lower the electronics noise will be. The occupancy also calls for a small pixel size to avoid
⁷⁶⁴ pixels having more than one hit and thus potentially losing the timing information. On the other hand a
⁷⁶⁵ small pixel size increases the number of channels to be instrumented and introduces inefficiencies due to
⁷⁶⁶ inter-pixel dead zones.

⁷⁶⁷ The probability to have a hit in a pixel decreases as function of the distance from the beam axis. For
⁷⁶⁸ simplicity a fixed pixel size was chosen, driven by the occupancy at the lowest instrumented radius of
⁷⁶⁹ 120 mm. A maximal occupancy of about 10% was targeted. This ensures a low double-hit probability
⁷⁷⁰ (below 10%). The timing information for a particle from the hard-scatter vertex is only lost for 5% of the
⁷⁷¹ cases when a particle originating from a pileup vertex hits the pixel first.

Figure 18 shows the occupancy expected for a pileup of $\langle \mu \rangle = 200$, defined as the percentage of pixels of the HGTD registering a hit, for different pixel sizes. As expected, the occupancy decreases as a function of radius and increases as a function of the pixel size. 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 6% for pixel sizes of $1 \times 1 \text{ mm}^2$, 10% for a pixel size of $1.3 \times 1.3 \text{ mm}^2$, interpolated from the other pixel sizes, and 20% for $2 \times 2 \text{ mm}^2$. A pixel size of $1.3 \times 1.3 \text{ mm}^2$ fulfils the requirements and was therefore chosen.

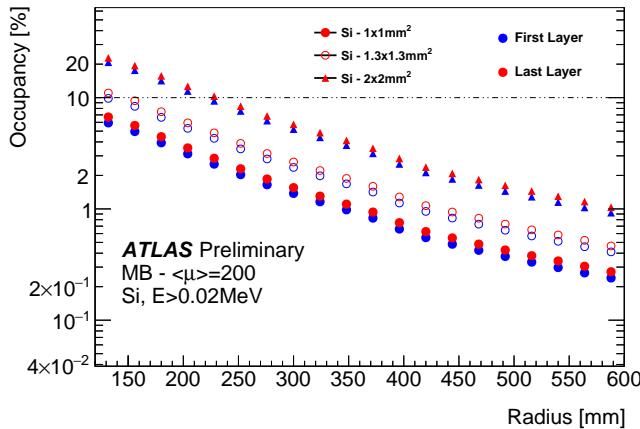


Figure 18: The occupancy as a function of the radius for different pixel sizes at a pileup of $\langle \mu \rangle = 200$. The occupancy for pixels of $1.3 \times 1.3 \text{ mm}^2$ is the result of an interpolation.

The detector geometry was optimised in two steps. First the layout of modules for the two cooling plates 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 optimised to achieve a timing resolution of about 30 ps per track. In these studies a pixel size of $1 \times 1 \text{ mm}^2$ was used, implemented at analysis level by merging the $0.5 \times 0.5 \text{ mm}^2$ pixels used in the full simulation.

The readout rows are defined as the sets of modules whose flex cables are guided together towards larger radii to the peripheral on-detector electronics. Their disposition for the first and second cooling disks is shown as rectangles in Figure 19. The active width of a module is 39 mm which limits how well the area near the circular opening at $R = 120 \text{ mm}$ can be covered. For $R > 150 \text{ mm}$ the coverage 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 1 mm between two readout rows and 3 mm between each half disk to account for mechanical tolerances. The effective width of the readout rows is therefore 40 mm. These constraints lead to the helix structure shown in Figure 19(a) for the first cooling disk. The second disk, shown in Figure 19(b) is defined as a mirror geometry with respect to the first disk. Therefore, as shown in Figure 19(c), with the exception of four readout rows per quadrant, the non-instrumented zones of the two disks do not overlap.

As demonstrated in Section 3 a timing resolution of about 30 ps per track should be provided by the HGTD. The geometry of the detector has been optimised to obtain a reasonably flat timing resolution as 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. Because of this, the geometry of the HGTD is designed such that at

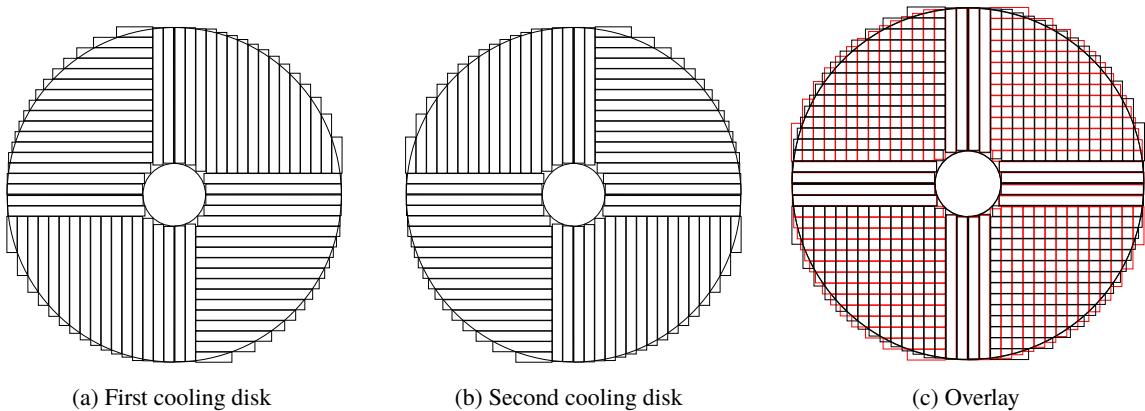


Figure 19: The orientation of the readout rows for the first and second cooling plates separately, and the overlay of both.

800 $R < 320$ mm on average three hits are obtained for a charged particle, whereas at $R > 320$ mm an average
801 of two hits are expected to be associated to a track.

802 Each disk of the HGTD is double-sided, i.e., the modules with sensors and on-detector electronics are
803 mounted on the front and back sides of the cooling plate. As illustrated in Figure 20, the modules on the
804 two sides of a disk are arranged to overlap so that the number of hits exceeds the number of disks. A study
805 using full simulation was performed to determine the optimal overlap between modules in $R < 320$ mm to
806 achieve the required timing resolution. The overlap was varied between 20% and 80%, limited by the need
807 for sufficient space between the modules to allow reading out the signals. For $R > 320$ mm, an overlap of
808 20% was found to be enough.

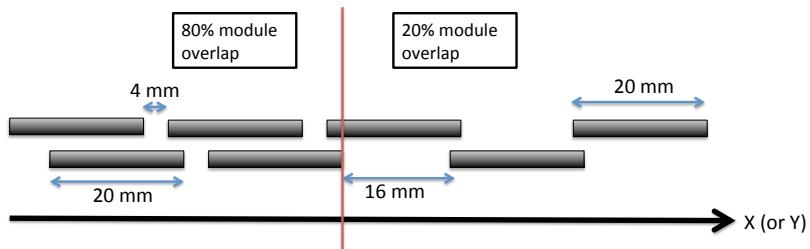


Figure 20: Schematic drawing showing the overlap between the modules on the front and back of a cooling disk.

809 The HGTD acceptance is defined as the surface covered by the HGTD between a radius of 120 mm and
810 640 mm (as mentioned earlier the simulation is only implemented up to 600 mm). The hit efficiency is
811 studied using events with single muons with $p_T = 45$ GeV and flat distributions of polar and azimuthal
812 angles. The inefficiency due to non-instrumented zones (e.g. areas not covered by modules) is defined as
813 the fraction of muons that have no hits reconstructed in the HGTD, and was found to be 0.7% with the
814 described layout. This does not take into account the inactive inter-pixel areas on the sensors themselves.

815 Figure 21 shows the average number of hits associated to a track. The results are summarised in Table 7
816 for an inactive inter-pixel region of 50 μm . The normalisation is performed separately in each region with

respect to the surface covered by the HGTD. Over 81% of the extrapolated muon tracks have at least two matched hits in the HGTD. For at most 3% of the muons, no hits are registered. The inefficiency is mainly due to the track-matching efficiency of 97% which is discussed in Section 3.1.

	$R < 320 \text{ mm}$	$R > 320 \text{ mm}$
	$\eta > 3.1$	$\eta < 3.1$
$N_{\text{hits}} \geq 2$	93%	81%
$N_{\text{hits}} = 0$	1.9%	3%
$\langle N_{\text{hits}} \rangle$	2.9	2.0

Table 7: The percentage of tracks with at least 2 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 are given for inactive inter-pixel zones of 50 μm and include the effect of non-instrumented zones and the track-matching efficiency. The normalisation is the surface covered by the HGTD in each region separately.

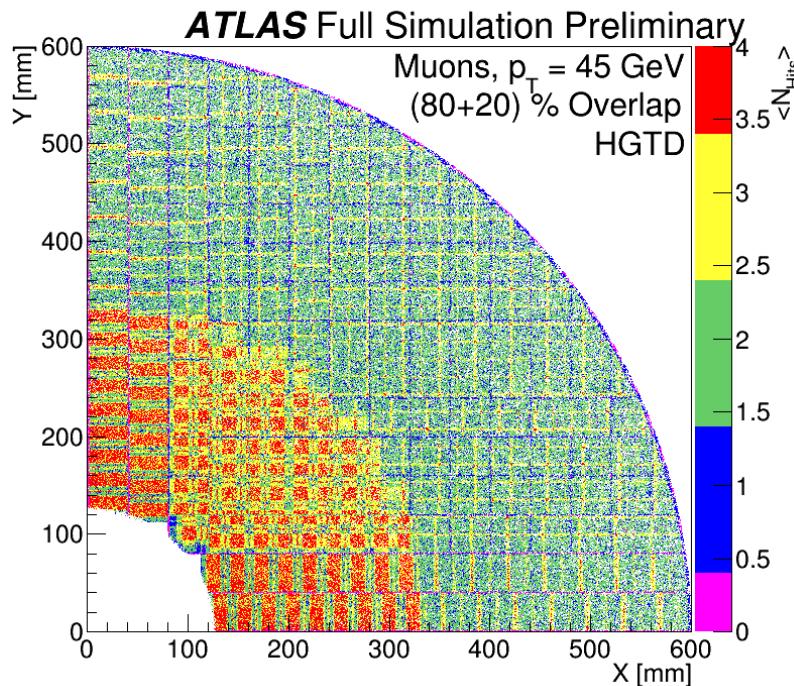


Figure 21: The number of hits as function of the position in the HGTD is shown for an overlap of 80% at $R < 320 \text{ mm}$ and 20% above.

The average number of hits is shown as a function of the radial distance from the beam axis in Figure 22. The non-instrumented zones and the inter-pixel dead zones of 50 μm are taken into account. For an overlap of 20% between the modules the average number of hits is about two. Increasing the overlap to 80% for $R < 320 \text{ mm}$ results in an average hit multiplicity of about three in this region. As this layout gives a relatively flat timing resolution as a function of radius also after the detector has been exposed to the expected radiation dose, it is chosen.

Four timing performance scenarios are defined: *Initial*, *Intermediate*, *Final* and *Worst Case*, shown in

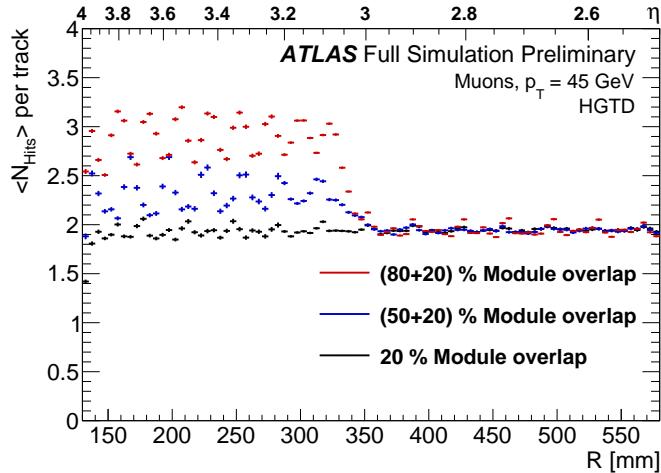


Figure 22: The average hit multiplicity as function of the radius (and pseudo-rapidity) for a module overlap of 20% in the outer region of the HGTD, and overlaps of 20%, 50% and 80% for the inner region.

- 827 Figure 23. The scenarios correspond to the performance expected at the very beginning, after half of the
 828 delivered luminosity, and at the end of the HL-LHC lifetime (4000 fb^{-1}). The per-hit timing resolution of
 829 the sensors is expected to be 30 ps at the start of the HL-LHC, independent of the radius. As the integrated
 830 luminosity increases, the damage to the sensors caused by radiation will deteriorate their timing resolution
 831 inducing a dependence on the radial distance from the beam axis. It is planned to replace the inner part
 832 ($R < 320 \text{ mm}$) of the HGTD after half of the HL-LHC programme, and therefore the timing resolution
 833 in this region for the *Intermediate* and *Final* scenarios are identical within statistical fluctuations. If the
 834 sensors of the inner ring were not replaced, the timing resolution of the sensors at a radius of 120 mm is
 835 expected to be degraded to the order of 50 ps. The *Worst Case* scenario assumes an additional degradation
 836 of approximately 20% on top of this for all radii, yielding a flat and overly pessimistic timing performance.
 837 The per-hit resolution curves in Figure 23(a) include a contribution from the electronics corresponding to
 838 25 ps which has been added in quadrature to the sensor resolution values discussed above.
- 839 The resulting timing resolution for tracks is shown in Figure 23(b) for the four scenarios. At radii lower
 840 than 320 mm ($\eta \approx 3.1$), the increased number of hits compensates the effects of radiation damage such that
 841 the resulting timing resolution is fairly independent of the radius. The per-hit resolution was implemented
 842 in full simulation. The reconstructed timing resolution for tracks shown here was determined using the
 843 single-muon events by comparing the reconstructed hit or track time to the expected time from the truth
 844 information. The distribution is dominantly Gaussian with negligible tails.

845 4.2 Module and layer design and assembly

- 846 The HGTD detector concept is based on individual planar layers of LGAD sensors to be fixed in front of
 847 both endcap calorimeter cryostats with active elements between 3435 and 3485 mm in z . Each layer is an
 848 independent object built on a cooling plate support disk (discussed in Section 5), allowing for a modular
 849 design. On both sides of this cooling plate, individual identical modules of $20 \times 40 \text{ mm}^2$ consisting of
 850 LGAD sensors, ASIC and flex circuits will be installed. A schematic drawing of the detector is shown in
 851 Figure 24. Groups of up to 19 modules constitute readout rows (see Section 4) which can be of different
 852 lengths. The longest readout row is 546 mm and defines the most difficult constraints for the mechanical

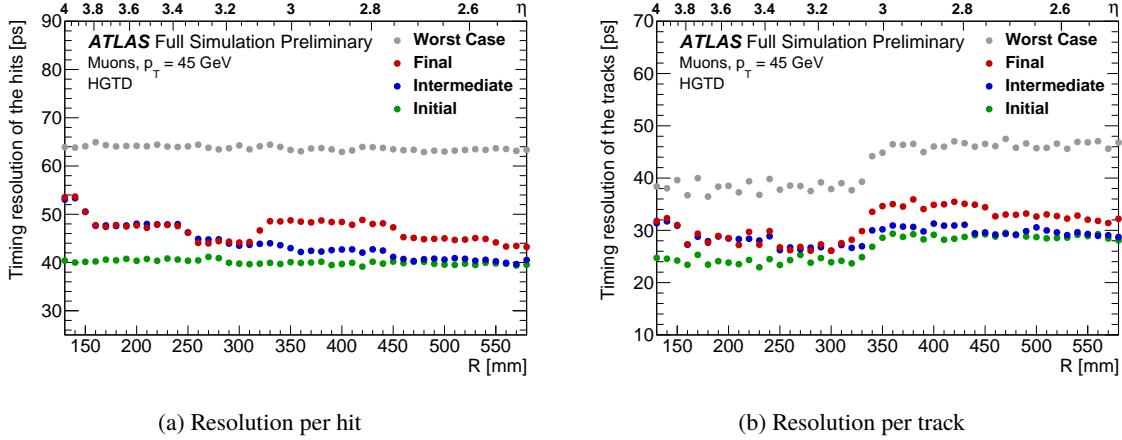


Figure 23: The HGTD timing resolution is shown as function of the radius for four timing scenarios. The sensor resolution and the contribution from the electronics are considered, added in quadrature.

assembly and the most stringent requirements for the flex cable characteristics. The module design is described in the next section, including the ASIC to sensor connection through bump-bonding and the design of the flex cable used to transmit the data to the peripheral on-detector electronics. In the two following sections the general detector layout with sensors, modules, layers and support structures is presented.

4.2.1 Module design

Components and layout

The surface of a module is $20 \times 40 \text{ mm}^2$ corresponding to a single LGAD sensor (see Section 4.3) with two ASICs (see Section 4.4.4) bump-bonded to it. This element is then glued with accurate positioning to the flex cable used to transfer the signals. ASIC signals and HV for the sensor will be connected by wire-bonding the flex cables to the module. Figure 25 shows three modules with the different components stacked in the z direction. The total thickness of a module is about 1 mm with the sensor, the ASIC and the flex cable contributing about $300 \mu\text{m}$ each. The needed space for each component, and in particular the curvature of the flex cable, is discussed in Section 4.2.3.

ASIC-sensor interconnection

The interconnection of the sensor to the front-end chip is a critical procedure of the detector assembly process. Each sensor channel is DC-coupled to the corresponding readout channel on the ASIC through a small (typically $80 \mu\text{m}$ diameter) solder ball. The complete hybridisation process, called bump-bonding, consists of under-bump metallization (UBM)⁶ (of sensor and ASIC pads, to prepare them for soldering), solder bump deposition (on the ASIC) and flip-chip. The R&D effort, presented below, is being carried out by dedicated groups within the HGTD community. For the full HGTD production phase, a combination of private companies and ATLAS institutions is foreseen to share the workload. Suitable companies for the UBM and flip-chip processes have been identified and contacted, and first cost estimates are available.

⁶ Some investigations are in progress to evaluate the possibility to use Au bumps which will not necessitate UBM process

Not reviewed, for internal circulation only

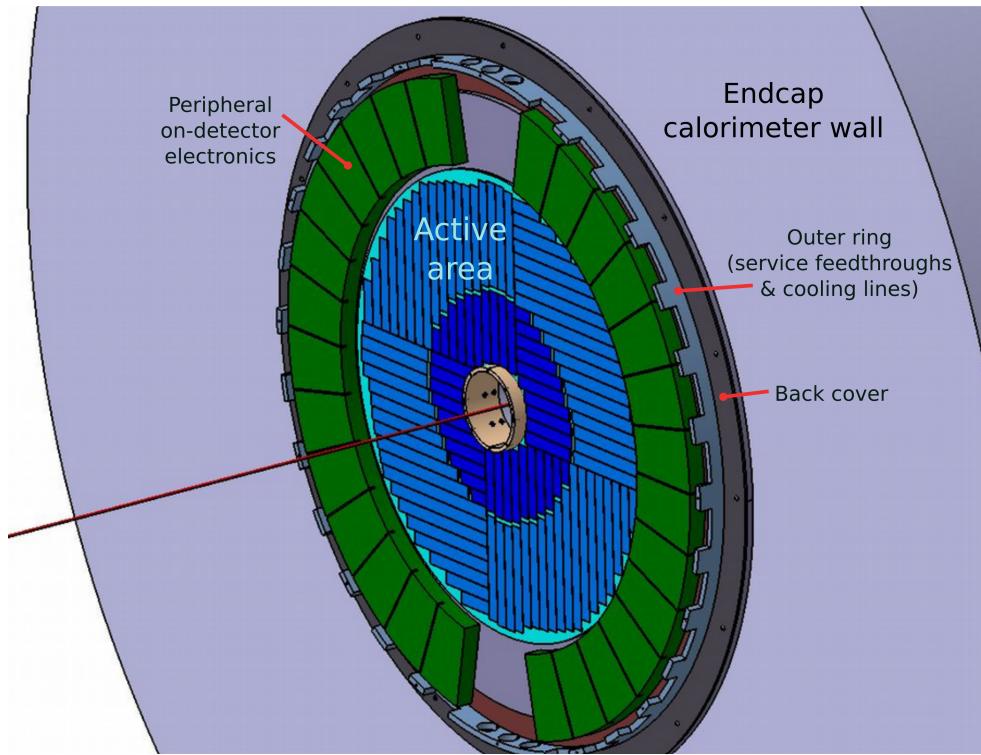


Figure 24: Illustration of the HGTD, showing the peripheral on-detector electronics in green and the layout of the readout rows, containing modules mounted on the inner half-disk support plates at $R < 320$ mm (dark blue), and on staves at larger radii (light blue).

876 In the hybridisation process, the aluminium pads of the sensor and readout chip are first covered with
 877 NiAu through an auto-catalytic chemical technique. The bumps are then deposited on the ASICs with
 878 a bump deposition machine that places SnAg (or SnPb) solder balls on the front-end pads and reflows
 879 them with an infrared laser system. The solder bumps are further reflowed in a dedicated machine in
 880 order to improve the placement and the shape uniformity of the bump balls. Flip-chip, the last step in the
 881 hybridisation process, is performed with a Süss Microtech bonder machine with a reflow arm that provides
 882 a $0.5\text{ }\mu\text{m}$ placement accuracy and $1\text{ }\mu\text{m}$ post-bond accuracy. During the bonding cycle the ASICs (with
 883 SnAg solder bumps) and the sensors are aligned, heated to $260\text{ }^\circ\text{C}$ and pressed together lightly. After
 884 flip-chip the assemblies are reflowed once again with formic acid.

885 First studies were performed on a 10×10 matrix of $1 \times 1\text{ mm}^2$ pixels in order to prepare for further
 886 tests with the ASIC prototype module. These preliminary tests confirmed that the critical steps of bump
 887 deposition and flip-chip can be conducted successfully with pixel pads $90\text{ }\mu\text{m}$ wide and bump balls of
 888 $80\text{ }\mu\text{m}$ diameter. In the following step, dummy samples that match mechanically the characteristics of the
 889 ASIC prototype modules were assembled obtaining excellent results with a bump strength of 60 g/bump .
 890 Altogether, these tests showed that a full assembly line is already available for the HGTD detector.

891 The first HGTD small-scale module prototypes using the first ASIC prototype (ALTIROC0_v1) and LGAD
 892 sensors were assembled in summer of 2017. Four assemblies were successfully produced following the
 893 same procedure used for the previous studies. The UBM of the diced chips and only partially diced
 894 sensors was done with thicknesses between 4 and $6\text{ }\mu\text{m}$. After bump deposition, the chip solder joints were
 895 reflowed and then the flip-chip process with selected sensors was carried out. The four bump-bonding

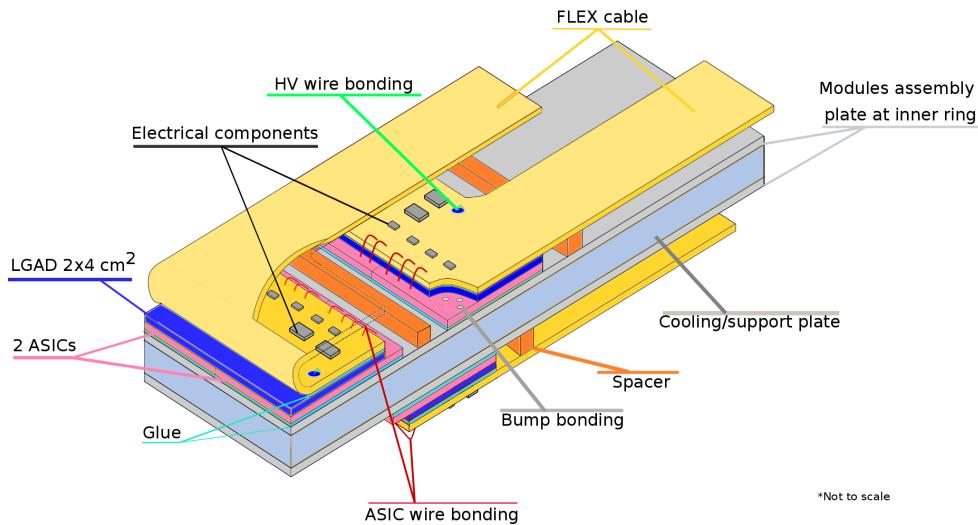


Figure 25: Schematic drawing of two adjacent modules on the top side and one on the bottom side of the cooling plate; the modules are mounted on thin support plates. The first flex cable is folded, while the others are straight.

cycles were successful, and X-ray inspection indicated good connectivity of all the bump bonds.

Voltage distribution and signal readout

The flex cables supply the high voltage (HV) power needed for the bias voltage of the LGAD sensors, power to drive the ASICs and include the electrical links, or *e-links*, for data transmission, as well as high speed differential lines to transport the signals to the peripheral readout electronics. The signal lines needed for two ASICs are summarised in Table 8.

The geometry of the flex cables is defined by the geometry and dimensions of the HGTD, the LGADs and the ASICs. Their flexibility is given by the insulating support material (Kapton®) with which they are made. A four-layer design of the cable is needed to fulfil the requirements and it can be realised within a thickness of 300 µm. To reduce the total thickness of the stacked cables routed to the same peripheral electronics board, two L-shape geometries are proposed, with a layout where the signals are on the left side of the module and a specular version where the signals are on the right side (see Figure 26). The distance between the innermost module and the start of the peripheral on-detector electronics is 630 mm, leading to a maximum length of around 700 mm needed for the flex cables.

Signal type	Signal name	Number of wires	Comments
HV	1 kV max.	2	Clearance
POWER	1x V _{vdda} , 1x V _{vddd}	2	Minimise voltage drop
GROUND		1 plane	Dedicated layer
Slow control	Data, ck, (opt. +rst, error)	2 to 4	I2C link
Input clocks	320 MHz, Fast command e-link, (opt. 40 MHz(L1))	6 or 8	LVDS
Data out lines	Readout data (TOT,TOA,Lumi)	4 pairs	4 e-links differential SLVS.
ASIC reset	ASIC_rst	1	Digital

Table 8: Types of signals for two ASICs included in the flex cable design.

Each module is supplied by a separate HV line which will allow for adjusting the voltage value at smaller radii to minimise the deterioration induced by radiation damage (Figure 37). In practice two lines will be

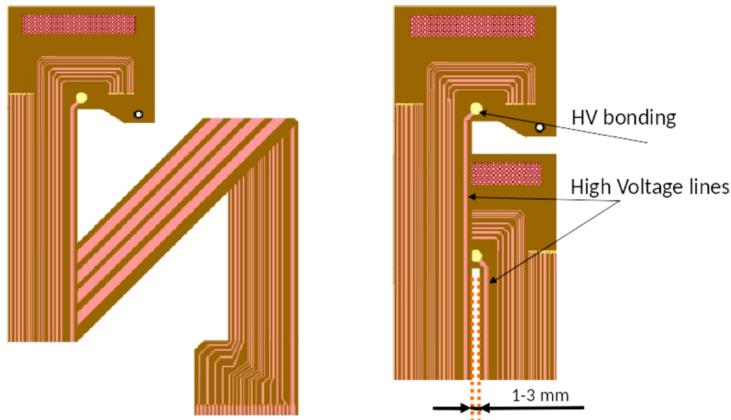


Figure 26: Left: Flex cable layout. Right: Sketch of two flex cables belonging to two contiguous modules. Clearance between HV lines is be considered. The bonding wires are placed in the centre-bottom part of the module.

connected to a module to ensure redundancy if one is broken. A maximum HV of 1 kV is assumed. The HV lines in the flex cable design will be placed in one of the edges not to disturb the other signals and a clearance of 1–3 mm between two contiguous flex cables (see Figure 26) must be kept to avoid undesired effects such as corona or arcing. The bonding will be placed between the LGADs in order to facilitate the HV supply to the module. Having separate flex cables for HV and for all other signals, respectively, is still an open option.

Each set of signals has different characteristics, therefore, specific optimisations are needed. The resistance of both the power lines and ground plane must be as low as possible $< 80 \text{ m}\Omega$. The crosstalk must be studied in the case of the analog and digital signals, whose impedance will be 50Ω . For the differential signals, the 100Ω impedance must be controlled for the high-speed signal transmission (1.28 Gbit/s). The choice of the appropriate materials and the geometry of the tracks are crucial to meet these requirements.

A prototype of the longest flex cable with a representative set of lines of all required types is being designed together with a company that was identified as a potential partner for the final production. Loopback tests including a time domain analysis and a bit error rate test will be performed in the next months. About 10 identical cables will be produced both for redundancy and to test the behaviour of the single lines within a stack of cables. For the bias voltage line the addition of a coating for better insulation will be considered based on the results of these tests.

4.2.2 Module loading

Modules connected to the same peripheral electronics board belong to the same readout rows, aligned along the x or y direction. The longest rows contain 19 modules. Their geometry was optimised as shown in Figure 19. The modules will be pre-assembled and glued to a thin intermediate support plate which is then screwed to the cooling plate. At $R < 320 \text{ mm}$ the plate consists of half disks that can be replaced easily after half of the HL-LHC running time, while at larger radii each readout row has its own rectangular plate called a stave. The readout rows that include modules at radii both smaller and larger than 320 mm will therefore contain modules mounted on the inner half disk and on a stave that is

937 assembled independently. Figure 27 shows a drawing of the modules assembled on the inner plate and the
 938 outer staves in one quadrant of sensors, i.e. one quarter of one side of a cooling plate.

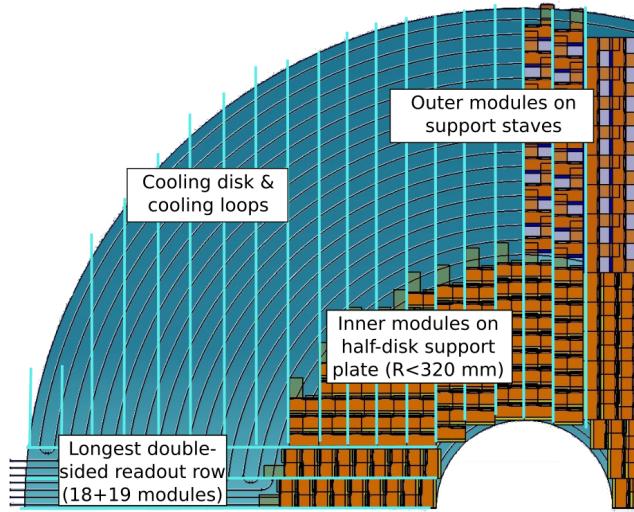


Figure 27: Drawing showing the assembly of the modules in one quadrant with cooling pipes visible. The modules at $R < 320$ mm are mounted on a half-disk support plate and positioned closer to each other, with an 80% overlap between modules on the front and back of the cooling plate. The modules at $R > 320$ mm are mounted on support staves with more space between them in the radial direction, yielding an overlap of 20%.

939 The overlap between the modules on either side of the cooling plate was optimised as shown in Figure 20:
 940 modules on the front and back of the cooling plate overlap by 80% in the inner region, while those
 941 assembled on staves in the outer region overlap by 20%. The innermost module of each readout row
 942 is rotated by 180 degrees such that the wire bonds face away from the first readout row of the adjacent
 943 quadrant or from the inner radius of the detector. Since the rotated sensors can be positioned flush with the
 944 perpendicular readout row of the adjacent quadrant, the non-instrumented area is minimised. To achieve
 945 this the flex cable corresponding to the first module needs to be folded backwards (see Figure 25). A
 946 bending diameter of 8 mm is expected to ensure the full functionality of the flex cable, while fitting well
 947 within the available space (as shown in Table 9). Due to the difficulties that could arise from this design,
 948 alternative layouts that could avoid cracks between the staves, while allowing for a simpler flex cable
 949 geometry, are under study.

950 The maximum thickness of one module package is 2.0 mm. Each disk has modules placed on both sides,
 951 leading to a total module package per disk of 4.0 mm plus 0.5 mm tolerance, as detailed in Table 9.
 952 This assembly has to prevent any flex cable crossing between two adjacent readout rows to assure correct
 953 routing. Accurate alignment in x and y directions and a controlled orientation of the peripheral on-detector
 954 connectors are required. The inter-disk gap (along z) should be fixed over the total active area. More
 955 details on the overall dimensions, including tolerances on the stacking of the various cooling layers are
 956 described in the next section.

957 Spacers will be used to maintain the distance between the support plate and the flex cable, guiding it up
 958 to the peripheral electronics. The high thermal conductivity of the spacers also contributes to drain the
 959 heat from the flex cables throughout the cooling panels. The materials under study for the spacers are
 960 carbon fibre composite, graphite reinforced PEEK, and a VESPEL polyimide graphite filler like SP-21
 961 used in the interface between the beam pipe and the Inner Detector. Assembly to the cooling panels by

962 screwing and/or a gluing technique with interposed conductive media should give the right thermal bridge
 963 properties for the flex cables. A digital assembly process using Computer Numerical Control machining
 964 will be used in order to position the spacers accurately with respect of the staves in the $x-y$ plane. This
 965 will help prevent the propagation of thermo-mechanical stress between staves, the modules and the flex
 966 cables.

967 With $20 \times 40 \text{ mm}^2$ modules and the readout row concept, the rectangular structures cannot give perfect
 968 coverage in azimuthal angle ϕ in the innermost region. As a result, there is full coverage in ϕ up to
 969 $|\eta| = 3.86$. In the $|\eta|$ range 3.86–4.00, approximately 92% of the surface area will be instrumented. For
 970 $4.00 < |\eta| < 4.07$, some intervals in ϕ are instrumented.

971 One prototype of the local support plates will be produced to study the thermo-mechanical behaviour,
 972 temperature distribution, CO_2 cooling parameters and glue-layer integrity between modules and carbon
 973 fibre skins. A scale model of the longest stave will also be built to ensure the feasibility of module
 974 assembly and installation of flex cables.

975 Assembly plans

976 The assembly will consist of fixing the 7888 modules on their intermediate plates, the installation of the
 977 flex cables and the wire-bonding of the flex cables to the sensors and the ASICs. Dedicated tools and
 978 procedures will be needed because of the shape of the intermediate plates, constituted by half disks for
 979 the inner region and rectangular staves for the outer region. The procedure will therefore be specific to
 980 the inner and outer part. For the half-disk intermediate plates, there is two per side of each cooling disk,
 981 and four cooling disks, yielding a total of 16 half-disk support plates. For the outer part, there are 18
 982 rectangular staves per quadrant, four quadrants for each side of a cooling plate, and a total of four cooling
 983 plates. This gives a total number of rectangular staves of 576 for the outer part. There are 14 different types
 984 of staves of varying lengths. The assembly step will be shared between four to six sites, where one or two
 985 will specialise on the inner plates and three or four on the outer staves. Different tests will be performed to
 986 control the electrical characteristics of the assembly (connectivity, sensors and ASICs operation) as well
 987 as the mechanical parameters (dimension, position). In addition, the packaging of each plate equipped
 988 with modules and the flex cables is a crucial point, allowing transport and then implementation on the
 989 cooling plates at CERN. At this stage, each half-disk cooling plate will be equipped with one inner half
 990 disk and the rectangular staves. The cables will be unrolled and attached to the entire length of the cooling
 991 plate.

992 4.2.3 Detector structure

993 Layers

994 A transverse view of the HGTD layers is presented in Figure 28, stacked to constitute the detector assembly,
 995 including cooling plates and silicon modules. The full layer assembly, with the given spacers gap, will
 996 match the total envelope of 125 mm, which includes the front and back covers of the vessel and the
 997 moderator.

998 A detailed breakdown of the dimensions of the different detector components is presented in Table 9.
 999 It lists the thickness of each component per side and per double-sided layer (when applicable), and the
 1000 total in the detector. The size of the component is given as a nominal thickness (the compact volume
 1001 occupied by material) and as an envelope thickness (the actual space needed per object, taking into account
 1002 the tolerance). The current prototype of the sensor and ASIC assembly is about 1 mm thick (detailed

Not reviewed, for internal circulation only

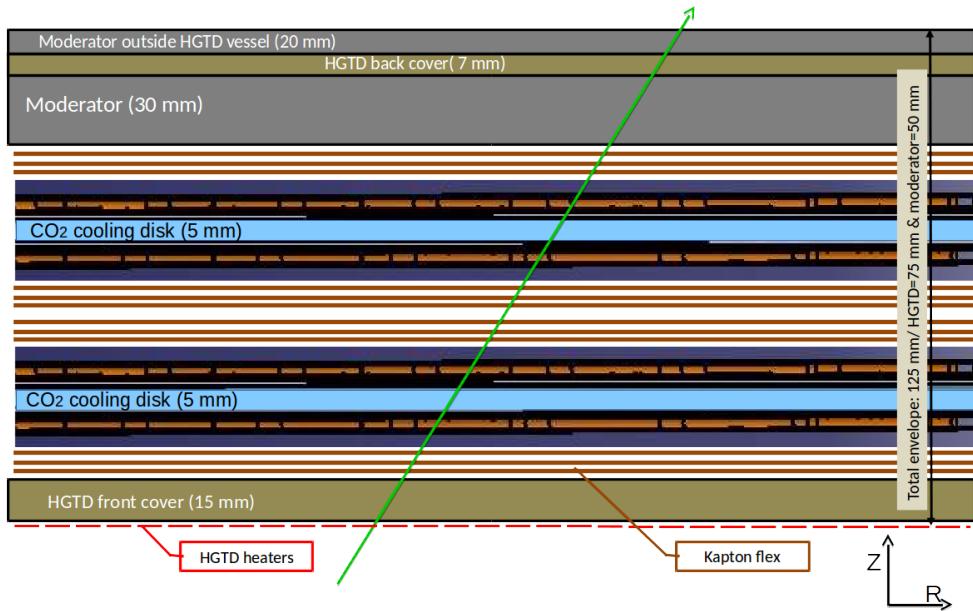


Figure 28: Cross section of the entire HGTD vessel including two active layers installed on the cooling plates, the front and back covers, and the moderator. An extra 20 mm moderator is located outside the vessel in close contact with the endcap cryostat.

in Section 4.3). This estimation gives a comfortable margin with respect to the final assembly protocol (gluing layers, etc.). As explained earlier, the thickness of one Kapton® flex cable is 0.3 mm, while the longest stave in a layer will have 19 modules. Thus, the maximum amount of stacked flex cables will be 10, yielding a maximum thickness per side of 3.0 mm. However, the current design must consider an envelope of 8.0 mm per side to allow the bending of the first flex. The envelope of the sum of the other flex cables and the coupling capacitors fits well within this parameters and doesn't need to be considered separately.

Support structure

To assemble the individual detector layers, the detector modules will be installed on the highly conductive support plates made of carbon fibre and graphite foam with embedded cooling pipes, inspired by the future ATLAS and CMS trackers. These disks will be divided in two halves, in a way that will allow for an easier mounting and dismounting of the HGTD, even in the event of having the beam pipe in place during installation. The modules will be attached to both sides of the support plates and safely locked by using one screw-pin. The half-disks, staves and modules can be seen in Figure 24. An appropriate loading device will be used to guarantee the module alignment and its x-y positioning on the support plate. The flatness and roughness of the plate have to allow the gluing of the modules with the best reachable thermal conductivity.

The support plates will be enclosed by pultruded carbon fibre U-shaped rings which will be the direct interface with the global support (HGTD cold vessel and its central tube) to allow screwing, positioning and alignment with the ATLAS coordinates system.

The moderator is discussed in detail in Section 5.3.

Not reviewed, for internal circulation only

Component	Layer side [mm]		Double-sided layer [mm]		Total HGTD [mm]	
	Nominal thickness	Envelope	Nominal thickness	Envelope	Nominal thickness	envelope
ASIC+sensor	1.0	1.0	2.0	2.0	4.0	4.0
Support plates	1.0	1.0	2.0	2.0	4.0	4.0
Flex circuit	2.8 - 5.5	8.0	5.6-11.0	16.0	11.2-22.0	32.0
Cooling panel	-	-	5.0	6.0	10.0	12.0
Total	7.5	10.00	20.0	26.0	40.0	52.0
Front cover	-	-	-	-	15.0	16.0
Back cover	-	-	-	-	6.0	7.0
Total HGTD					61.0	75.0
Inner moderator					30.0	30.0
Outer moderator					20.0	20.0
Total Moderator					50.0	50.0
HGTD+moderator					111.0	125.0

Table 9: Estimated values of thickness per component. The nominal thickness is the manufacturing dimension of the component. The envelope is the space needed to be allocated for the component. Some components are not considered in the envelope thickness because they are included within another value. Information is given for one side of a layer (when applicable), for a double-sided layer, and the total for one HGTD side.

1024 4.3 Sensors

1025 As discussed above, the time resolution for the HGTD is required to be 30 ps per track over its full lifetime.
 1026 This translates to a requirement of 42, 52 or 60 ps/hit in case of 2, 3 or 4 independent timing measurements
 1027 along the particle path. As shown in the following, a resolution in this range can be achieved with silicon
 1028 sensors with internal gain (LGAD) of 50 μm active thickness. The radiation hardness requirement for the
 1029 innermost radii (including a safety factor of 1.5) is a 1 MeV neutron equivalent fluence of $3.7 \times 10^{15} \text{ n}_{\text{eq}}/\text{cm}^2$,
 1030 assuming one replacement of the inner part after half of the total integrated luminosity of 4000 fb^{-1} (see
 1031 Section 1).

1032 The sensor pad size is restricted by occupancy, pad capacitance and the fill factor. A unified pad size of
 1033 $1.3 \times 1.3 \text{ mm}^2$ everywhere with an expected capacitance of 3.4 pF is found to fulfil the requirements. The
 1034 pads will be arranged in arrays of total area of about $20 \times 40 \text{ mm}^2$ (15x30 pads) with a common backplane
 1035 bias voltage connection.

1036 4.3.1 Low Gain Avalanche Detectors

1037 The technology chosen for the HGTD sensors is Low Gain Avalanche Detectors (LGAD). LGADs
 1038 are planar silicon detectors with internal gain. They have been pioneered by the Centro Nacional de
 1039 Microelectronica (CNM) Barcelona [3] and developed during the last 5 years within the CERN-RD50
 1040 community [2]. Much background and details are given in Reference [19].

1041 LGADs are n-on-p silicon detectors containing an extra highly-doped p-layer below the n-p junction to
 1042 create a high field which causes internal gain as displayed in Figure 29(a). When a charged particle crosses
 1043 the detector, an initial current is created from the drift of the electrons and holes in the silicon. When the
 1044 electrons reach the amplification region, new electron/hole pairs are created and the holes drift towards
 1045 the p⁺ region and generate a large current (i.e. gain).

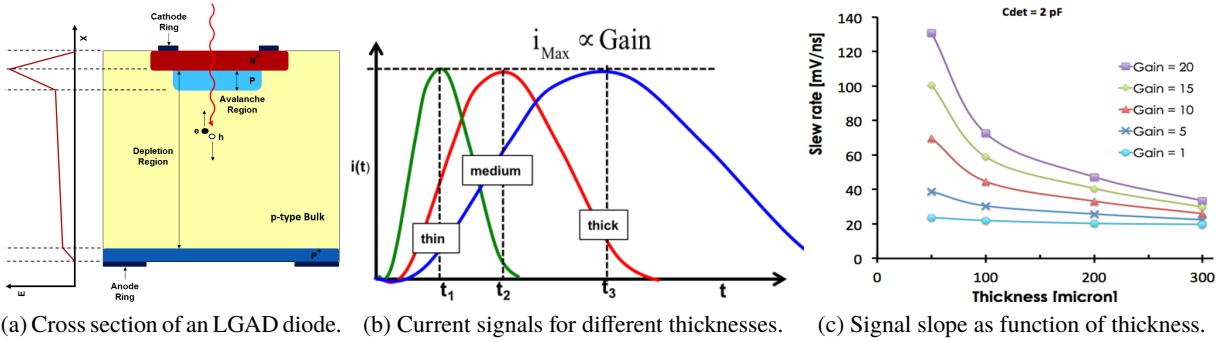


Figure 29: The working principle of LGAD sensors, and how their signals depend on the sensor thickness [19].

1046 While the signal height i_{MAX} of the LGAD is proportional to the gain M , it is independent of the detector
 1047 thickness as illustrated in Figure 29(b). On the other hand, Figure 29(c) shows that the signal slope
 1048 dV/dt depends on the thickness of the sensor, favouring thin sensors [19]. An LGAD thickness of 50 μm
 1049 has been adopted as the baseline and studied in detail. LGADs of 35 μm thickness recently have become
 1050 available as well and are studied as an option with an even larger signal slope at the expense of an increased
 1051 capacitance.

1052 Two major effects from the electronics which determine the time resolution are the time walk and the time
 1053 jitter. Both depend inversely on the signal slope dV/dt :

$$\sigma_{\text{TimeWalk}} = \left[\frac{V_{th}}{\frac{S}{t_{rise}}} \right]_{\text{RMS}} \propto \left[\frac{N}{\frac{dV}{dt}} \right]_{\text{RMS}}, \quad \sigma_{\text{Jitter}} = \frac{N}{(dV/dt)} \simeq \frac{t_{rise}}{(S/N)}, \quad (4)$$

1054 where S refers to the signal, N to the noise, t_{rise} to the rise time and V_{th} to the threshold voltage. Thus the
 1055 time resolution of the LGAD is tied to the value of its signal slope. The dependence of the slope on sensor
 1056 thickness and gain is shown in Figure 29(c). The best time resolution, i.e. the largest slope, is achieved
 1057 with thin sensors and large gain. This observation feeds into the plan to operate LGADs at as large a gain
 1058 as possible given restrictions from the leakage current and the breakdown voltage. Our investigations
 1059 point to a safe gain of about 20.

1060 In addition to the research carried out by CNM, the Fondazione Bruno Kessler (FBK) Trento has also
 1061 designed and produced LGAD sensors [20], as well as Hamamatsu Photonics (HPK) [19]. In general, the
 1062 LGAD sensors produced by different manufacturers appear to perform very similarly, with the exception
 1063 of the leakage current before irradiation, and the bias voltage reach after irradiation.

1064 CNM has produced thin LGADs in 2016 in run 9088 within an RD50 Common Project [2] and in 2017
 1065 in the HGTD specific run 10478 [21]. Both runs were on 4" wafers with nominally 50 μm thickness of
 1066 the active high-resistivity Float zone volume. Run 9088 was in a silicon-on-insulator (SOI) process with
 1067 the back-side contact done through wet-etched deep access holes through the insulator (see Figure 30(a)).
 1068 Most of the tests were performed on single pads of $1.3 \times 1.3 \text{ mm}^2$ active area (called LGA) and 2×2 arrays
 1069 of pads with $2.063 \times 2.063 \text{ mm}^2$ active area each and a no-gain gap of 63 μm between adjacent pads. Three
 1070 sets of wafers were produced with a different multiplication layer implantation dose to optimise the gain:
 1071 $1.8 \times 10^{13} \text{ cm}^{-2}$ (low), $1.9 \times 10^{13} \text{ cm}^{-2}$ (medium) and $2.0 \times 10^{13} \text{ cm}^{-2}$ (high). In run 10478, wafer-to-wafer
 1072 (Si-Si) bonded wafers were used with a high dose (see Figure 30(b)). It also includes arrays with 2×2 pads
 1073 of 1.0 and 2.0 mm width that are compatible with the HGTD ALTIROC0 readout chip (see Section 4.4)

1074 with under-bump metalization (UBM) to ensure bump-bonding (see Section 4.2.1). An important further
 1075 development to improve the breakdown behaviour of the arrays is the inclusion of a junction termination
 1076 extension (JTE) around each individual sub-pad in an array. Part of the wafers include Carbon-spray to
 1077 test if this improves radiation hardness (see Section 4.3.4). CNM also finished the production of a 300 µm
 1078 thick test run on 6" wafers and will use this wafer size for future 50 µm thick runs. Moreover, a run on
 1079 35 µm thick wafers is ongoing.

Not reviewed, for internal circulation only

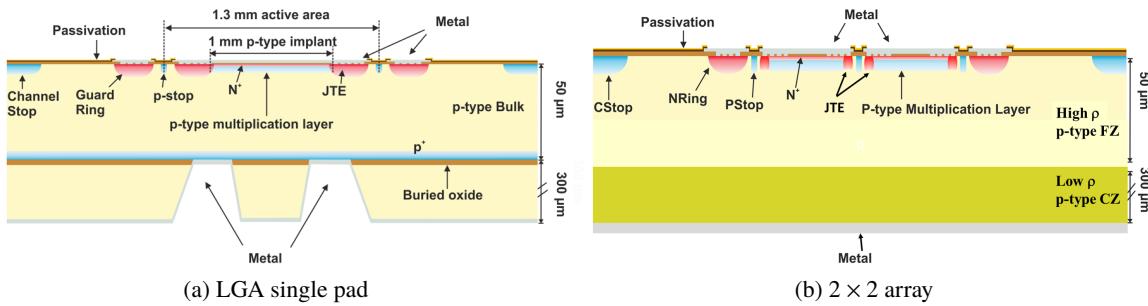


Figure 30: Cross section of LGAD structures. (a) LGA single pad from CNM run 9088 (SOI substrate). (b) 2x2 array of CNM run 10478 including a JTE around each sub-pad (Si-Si wafer) [21].

1080 Hamamatsu Photonics (HPK) delivered LGAD sensors from run #ECX20840. They were manufactured
 1081 on Si-Si bonded 6" silicon wafers of 150 µm total thickness with a 50 µm or 80 µm thick high resistivity
 1082 float zone (FZ) active layer on top. Samples with four different “gain splits” (denoted A-D) were produced,
 1083 identical in the mask design but with a different p⁺ dose of the gain layer, to study the optimum parameters
 1084 of the charge multiplication mechanism. A has the lowest, D the highest dose. The capacitance-voltage
 1085 (C-V) measurements indicated that the difference between doping concentrations of adjacent splits is
 1086 about 4%. The most studied devices have 50 µm thickness with C or D dose, denoted 50C or 50D in the
 1087 following. The wafers contain circular single pads of area about 0.8 mm² and 2x2 arrays of 3x3 mm²
 1088 pads. The arrays have a no-gain gap of 100 µm between adjacent pads. Recently also 35 µm thick LGAD
 1089 single pads of 0.8 mm² area were provided with one single dose split, which will be referred to as B35 in
 1090 the following.

1091 4.3.2 Sensor tests: methodology and experimental techniques

1092 The LGAD sensors have been tested by various HGTD groups, as well as within the RD50 community.

1093 Electrical measurements including capacitance-voltage (C-V) and current-voltage (I-V) characteristics
 1094 have been performed in laboratory probe stations.

1095 The dynamic properties of LGADs, such as charge collection, gain and time resolutions, have been
 1096 measured in response to ionising particles, both in the laboratory with ⁹⁰Sr β particles [19, 22–26] and
 1097 lasers, as well as in beam tests with pions at CERN and Fermilab [24, 25, 27].

1098 Different readout boards and amplifiers have been used, including commercial broad band or charge-
 1099 sensitive amplifiers and an HGTD-specific development with a high bandwidth, allowing the recording
 1100 of the pulse shape of the fast LGAD signals [24]. This readout board has been made available in the
 1101 meanwhile to many HGTD institutes via production at CERN. The analog waveforms are then typically
 1102 recorded with a fast oscilloscope.

1103 Position-sensitive scans using red and infrared laser to deposit charge carriers inside the sensors have been
 1104 made at various institutes, using the Transient Current Technique (TCT) setup.

1105 Beam tests have been performed by the HGTD community in five periods in 2016 and 2017 at the H6
 1106 beam line of the CERN SPS with 120 GeV pions [27]. Data were taken in two modes: stand-alone and
 1107 integrated into a beam telescope that provided track position information with about 3 μm precision.

1108 The gain is extracted by dividing the collected charge in an LGAD device by the charge of no-gain PIN
 1109 diodes without multiplication layer (for betas and MIPs about 3 ke^- or 0.46 fC for 45 μm thickness).

1110 Time resolutions are typically extracted from the spread of the time-of-arrival difference between two
 1111 sensors when a particle passes through both. Either at least two LGADs are used or LGADs and a fast
 1112 Cherenkov counter based on Quartz bars and a Silicon photo multiplier (SiPM) with typically about
 1113 10 ps time resolution. If at least 3 devices are measured simultaneously, a χ^2 minimisation is used
 1114 to obtain the time resolution of all devices. In case only one device under test (DUT) is measured
 1115 with respect to one reference device of known resolution, the DUT resolution is obtained by subtracting
 1116 quadratically the reference contribution. Different time reconstruction algorithms have been investigated
 1117 such as the Constant Threshold Discriminator (CTD), the Constant Fraction Discriminator (CFD), the
 1118 Zero Crossing Discriminator (ZCD) or corrections using the amplitude or Time-Over-Threshold (TOT)
 1119 of the signal [27].

1120 Preliminary studies carried out on irradiated sensors in test beam show negligible values of cross-talk
 1121 between pads in a sensor array.

1122 4.3.3 LGAD performance results before irradiation

1123 Electrical characterisation: I-V, C-V

1124 Figure 31 shows the I-V and C-V curves of 1.3x1.3 mm² LGAD pads of CNM runs 9088 and 10478 of
 1125 different multiplication layer doses. The leakage current before breakdown is typically in the order of
 1126 0.1 nA. The breakdown voltage increases with decreasing multiplication layer dose for the same run (about
 1127 80, 240 and 300 V for high, medium and low doses of run 9088). The breakdown voltage of run 10478
 1128 is higher (125 V) than the one of 9088 with the same nominally high dose due to process improvements,
 1129 especially in the JTE. Also the range of the "foot" of the $1/C^2$ vs. V (i.e. the region where $1/C^2$ stays at
 1130 low values while the multiplication layer is being depleted, starting from the n-p junction at the front) is
 1131 an indicator of the multiplication layer dose. Foot values between 30 and 40 V indicate substantial gains,
 1132 as verified below. It can be seen that for the same nominally high dose of run 9088 and 10478, there are
 1133 moderate run-to-run variations, leading to an effectively higher dose for run 10478. The depletion of the
 1134 bulk (indicated by the rise of $1/C^2$ -V) happens rather fast within a few V due to the high resistivity and the
 1135 small thickness. From the end capacitances of about 3.9 pF and 4.2 pF (measured with a connected guard
 1136 ring) one can conclude an active thickness of 45 and 42 μm for runs 9088 and 10478, respectively.

1137 The current level of all HPK devices before breakdown is very low, typically below 0.1 nA. As expected,
 1138 devices with lower doses have a much later breakdown (300 V for dose D, 700 V for dose A for 50 μm). The
 1139 I-V curve uniformity between different samples was found to be very good for both single pads and arrays.
 1140 The capacitance-voltage (C-V) measurements indicated that the difference between doping concentrations
 1141 of adjacent splits is about 4%. The end capacitance of the 0.8 mm² single pads was measured as 2.7 and
 1142 4.6 pF for 50 and 35 μm , respectively (larger than expected from the nominal area due to an increased
 1143 active area resulting from floating guard rings).

Not reviewed, for internal circulation only

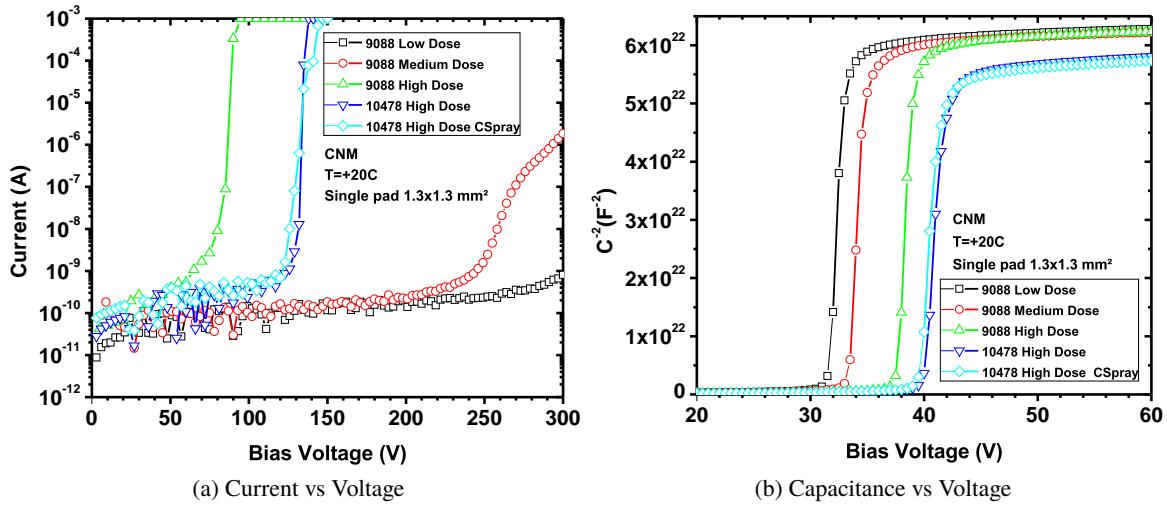


Figure 31: Measurements of (a) current-voltage and (b) capacitance-voltage of CNM LGA single pads from runs 9088 and 10478 of different multiplication layer doses, measured at room temperature [21].

1144 Collected charge and gain

1145 For 50 μ m thick CNM LGAD from run 9088, Figure 32 shows the gain and collected charge as a function
 1146 of bias voltage. Figure 32(a) displays different doses measured at 20 °C. It can be seen that the charge and
 1147 gain is higher for higher doses at a fixed voltage, as expected. The low and medium dose can be measured
 1148 up to a final gain of 50–60 at breakdown voltages of 250 and 310 V, respectively, whereas the high dose
 1149 breaks down already at 80 V at a gain of 20. Figure 32(b) shows the gain curve for different temperatures.
 1150 Due to higher impact ionisation, the gain increases at lower temperatures, and the breakdown is decreasing.
 1151 Similar results have been measured for HPK LGAD [26].

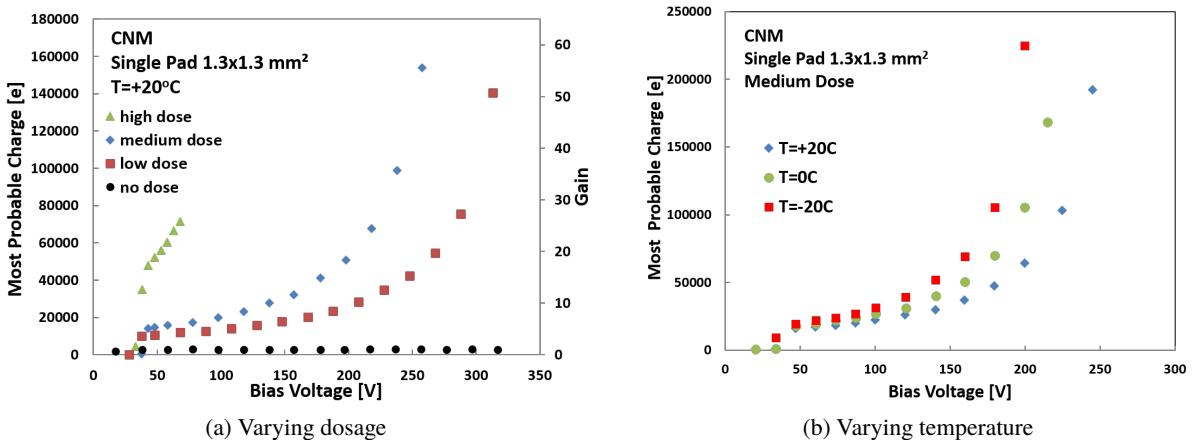


Figure 32: Gain or charge vs. bias voltage for the CNM run 9088. (a) shows the comparison of all multiplication layer doses at 20 °C. (b) shows the medium dose compared for 20 °C, 0 °C and -20 °C. [23]

1152 During beam tests the gain was measured also as a function of position for single pads and arrays and
 1153 found to be uniform for fully metallized structures (see Figure 33 left). Also the no-gain region in the gap
 1154 between adjacent pads is visible. Its width is measured as 70–100 μ m for current sensor designs.

1155 **Efficiency**

1156 The hit efficiency was measured in HGTD beam tests using an external telescope for reference tracks [27]
 1157 (see Figure 33 right). Amplitude thresholds (25–60 mV) well above the noise level (2–3 mV) were chosen.
 1158 For five different devices, mean values of 96–99% were obtained consistently.

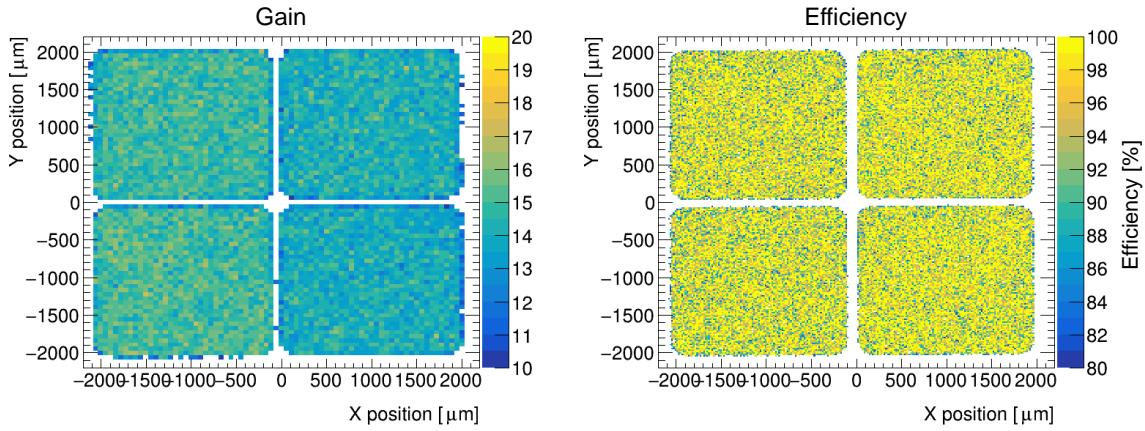


Figure 33: 2D maps of gain (left) and efficiency (right) for a 2x2 array from CNM run 9088 as measured in HGTD beam tests [27].

1159 **Time resolution**

1160 The time resolutions of CNM and HPK devices have been extensively studied in various beam tests [24,
 1161 25, 27] and ^{90}Sr setups [19] (see Figure 34). It has been consistently shown that sub-30 ps time resolution
 1162 can be achieved below the breakdown point before irradiation for the $1.3 \times 1.3 \text{ mm}^2$ (CNM) and 0.8 mm^2
 1163 (HPK) single pads with up to 5 pF capacitance on the custom-made test readout board explained above.
 1164 The time resolutions of the 2x2 arrays achieved so far (from CNM run 9088) are larger due to earlier
 1165 breakdown resulting from the lack of a JTE around each sub-pad, as well as due to higher capacitances
 1166 from the larger areas. Both are expected to improve in new runs.

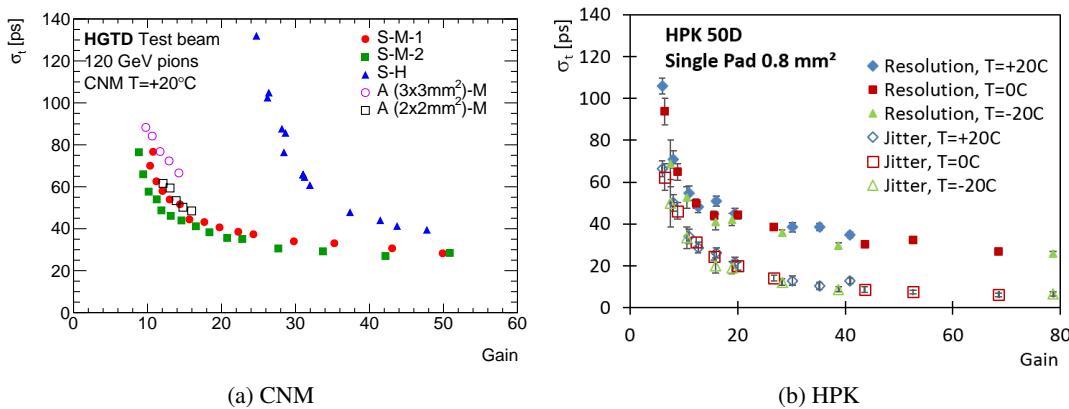


Figure 34: Time resolution σ_t as a function of gain. For (a) CNM single-pad sensors ("S") and arrays ("A") of medium ("M") and high ("H") doping[27]; and for (b) HPK 50D single-pad sensors measured at different temperatures. Also the jitter is shown. [19].

1167 Whereas at a fixed voltage, the time resolution is improving for lower temperatures due to the higher gain,

1168 it can be seen that the resolutions measured at different temperatures collapse to a universal curve when
 1169 presented as a function of gain before irradiation.

1170 For HPK, Figure 34(b) also includes the measured electronic jitter, which is found to be significantly lower
 1171 than the total time resolution, especially at high gains in the test beam. This indicates that contributions
 1172 from Landau fluctuations dominate there.

1173 4.3.4 Irradiation tests

1174 Irradiation of silicon mainly results in the change of the effective doping concentration, the introduction
 1175 of trapping centres that reduce the mean free path of the charge carrier, and the increase of the leakage
 1176 current [2].

1177 LGAD samples of 50 µm thickness from CNM (run 6827 and 9088) and HPK (run ECX20840) as well
 1178 as 35 µm thick HPK B35 were irradiated by neutrons at the JSI research reactor in Ljubljana up to 1 MeV
 1179 neutron-equivalent fluences of $6 \times 10^{15} n_{eq}/cm^2$ [23]. After irradiation the devices were annealed for
 1180 80 min at 60 °C. Also irradiations at CERN-PS with 24 GeV protons were performed; measurements and
 1181 analyses are ongoing. The results after neutron irradiation are presented below.

1182 Gain evolution after irradiation and acceptor removal

1183 The gain was found to decrease with irradiation which was attributed to loss of the effective doping
 1184 concentration in the multiplication layer due to deactivation of initial boron as acceptors, as has been
 1185 observed before on 300 µm thick LGADs [22]. This can be observed in Figure 35 for 50 µm thick CNM
 1186 and HPK devices, which show a very similar behaviour. The gain G steadily decreases with irradiation.

1187 Beyond a fluence of about $10^{15} n_{eq}/cm^2$, there is little difference between devices without (PIN) and with
 1188 (LGAD) a built-in multiplication layer due to its removal. However, at such high fluences a gain is observed
 1189 for both PIN and LGAD devices since deep effective acceptors created by irradiation in combination with
 1190 a higher breakdown voltage result in electric fields high enough for charge multiplication also in originally
 1191 no-gain detectors. This effect has been observed before [28–30]. In this case the multiplication takes
 1192 place over large parts of the volume at very high average fields (around 15 V/µm). Therefore the increase
 1193 of charge with bias voltage is very steep after the onset of multiplication. This means that detectors are
 1194 operated close to device breakdown which requires very good temperature control and voltage stability.
 1195 Nevertheless gains of G=8 were measured for devices irradiated to $4 \times 10^{15} n_{eq}/cm^2$.

1196 Owing to the small thickness of the device and consequently short drift times, the trapping of the drifting
 1197 charge only marginally affects the collected charge. The larger the initial implant doping the longer the
 1198 gain remains large, as can be seen from Figure 35(b) from comparing the lower dose 50C to the higher
 1199 dose 50D at $6 \times 10^{14} n_{eq}/cm^2$. Hence, an initially larger dose is beneficial in terms of radiation hardness.

1200 A regional dependence of the gain was found in position-sensitive beam tests for the CNM single pads
 1201 (LGA) that have holes in the top metallization [31]: at $6 \times 10^{14} n_{eq}/cm^2$ the gain below the metal was found
 1202 to be up to 2.5 times higher than below the hole. This effect, its cause and its impact on time resolution is
 1203 still under study. The final HGTD detector will have a full metallization (where the gain was observed to
 1204 be higher).

1205 Time resolution after irradiation

1206 The timing performance of neutron irradiated CNM and HPK LGADs was studied in the laboratory with
 1207 a β-telescope [19, 26, 32] and in beam tests [25, 31].

Not reviewed, for internal circulation only

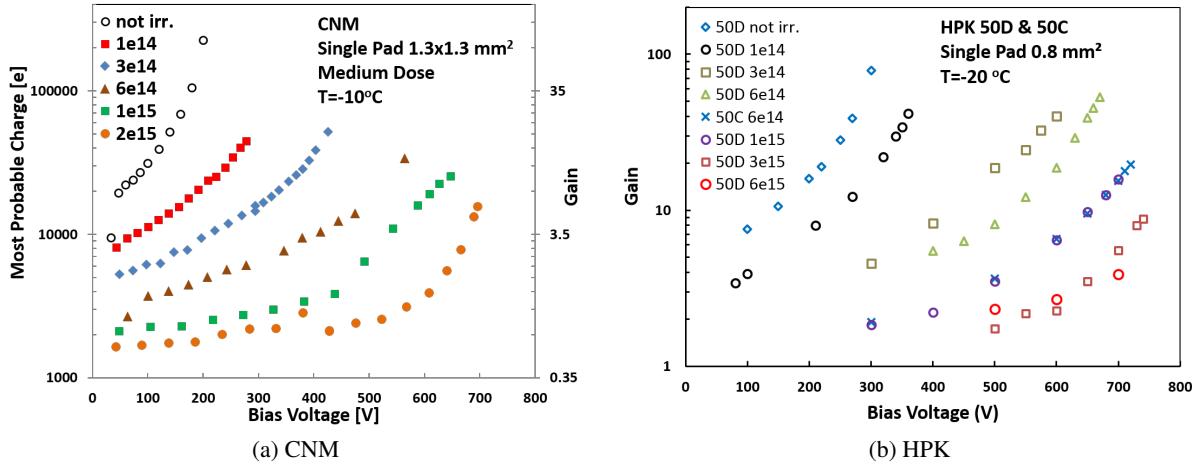


Figure 35: Most probable charge or gain dependence on bias voltage for different fluences (in $n_{\text{eq}}/\text{cm}^2$) measured for (a) CNM single pads from run 9088 with medium dose [23] and (b) HPK 50D/50C single pads [26].

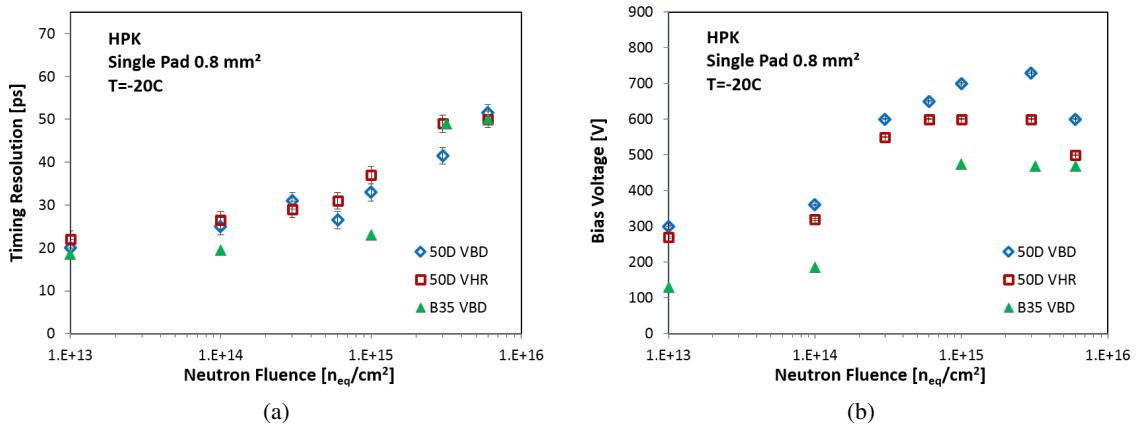


Figure 36: (a) Time resolution as a function of neutron fluence for HPK 50D at VBD and VHR and B35 at VBD at -20°C . (b) Fluence dependence of VBD and VHR [26, 32].

1208 The fluence dependence of the time resolution for HPK LGADs of 35 and 50 μm thickness (B35 and
 1209 50D) is shown in Figure 36(a). For 50D, to understand the “headroom” for detector operation, the time
 1210 resolution was evaluated at a bias voltage just below breakdown (VBD) and at a bias lowered by about
 1211 10% or more (headroom voltage, VHR). The corresponding voltage values are shown in Figure 36(b).
 1212 For 50D, the time resolution at VHR was found to deteriorate from about 20 ps before irradiation to 40 ps
 1213 after $1 \times 10^{15} \text{ n}_{\text{eq}}/\text{cm}^2$ and to 50 ps after $6 \times 10^{15} \text{ n}_{\text{eq}}/\text{cm}^2$. The needed bias voltages increase from about
 1214 300 V to about 600 V after $3 \times 10^{14} \text{ n}_{\text{eq}}/\text{cm}^2$, after which they remain relatively stable. For B35, the time
 1215 resolution before irradiation and at high fluences is similar to 50D, whereas at intermediate fluences the
 1216 resolution is lower by about 10 ps. A major advantage of B35 is that the bias voltage needed is lower than
 1217 for 50D, typically by 100-200 V. This translates into lower power dissipation as discussed below.

1218 The resolution was found to improve by typically a few ps at a measurement temperature of -30°C with
 1219 respect to -20°C for the same voltage.

1220 Considering the low gain at high fluences (see Figure 35), a worse time resolution would have been
 1221 expected. However, part of the charge multiplication occurs in the bulk of the sensor as explained above.
 1222 This leads to a decrease in the rise time of the signal. Moreover, the noise was found to increase at
 1223 high fluences and voltages. For these reasons, in that regime, the time resolution is not anymore a
 1224 universal function of gain nor does it decrease monotonically with voltage as before irradiation or at lower
 1225 fluences.

1226 The time resolution of CNM devices after irradiation was investigated in a first study on LGA single-pad
 1227 diodes [25]. However, as mentioned above, the gain was found to be lower in the metal hole of these
 1228 devices, which results in a time resolution degradation. New measurements with fully metallized surfaces
 1229 are ongoing.

1230 To conclude, these results show that the currently available LGAD test sensors can be operated safely up
 1231 to the HGTD target fluence of $3.7 \times 10^{15} \text{ n}_{\text{eq}}/\text{cm}^2$, keeping a time resolution of 50 ps/hit under laboratory
 1232 conditions. It is still under study what will be the full-system time resolution for a large array including
 1233 final electronics in a power-limited readout chip, which might increase the jitter, in particular for B35
 1234 with a higher capacitance (see Section 4.4). In parallel, further development to improve the sensor time
 1235 resolution after irradiation is ongoing as explained below. The final projected time resolution of the HGTD
 1236 system will be known at the end of this full evaluation and after the decision of the final HGTD layout.
 1237 Moreover, the time resolution measurements need to be complemented with hit efficiency studies, which
 1238 is a priority for upcoming beam tests.

1239 Evolution of operation voltage during sensor lifetime

1240 The gain degradation with fluence makes it necessary to adjust the operation voltage over the lifetime of
 1241 the detectors. This is for example shown in Figure 36(b) where VHR is a good indicator for an operation
 1242 voltage as a function of fluence that provides sufficient time resolution while still being safely below the
 1243 breakdown point (these studies still need to be complemented with hit efficiency measurements). For the
 1244 baseline thickness of 50 μm , it rises from about 300 V before irradiation to 600 V at $3 \times 10^{14} \text{ n}_{\text{eq}}/\text{cm}^2$,
 1245 after which it stays relatively constant (it should be noted that the exact values are technology and doping
 1246 specific, the values here refer to HPK 50D as an example). Monitoring of the leakage current and the
 1247 TOT as an indicator of collected charge will give a good estimate of the gain evolution during operation,
 1248 allowing to perform the necessary adjustments. The expected dependence of the fluence on the radius
 1249 (Figure 4) and the required bias voltage for the increasing fluence (Figure 36(b)) permits a prediction of
 1250 the bias-voltage distribution along the length of a readout row containing individual modules mounted on
 1251 support plates (see Section 4.2). This is shown in Figure 37(a), where the bias voltage as a function of

fluence for different fluence levels at a radius $R = 300$ mm are shown. In Figure 37(b) this is translated into the bias voltages for different radial position along the longest readout row for a few fluence steps. 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.

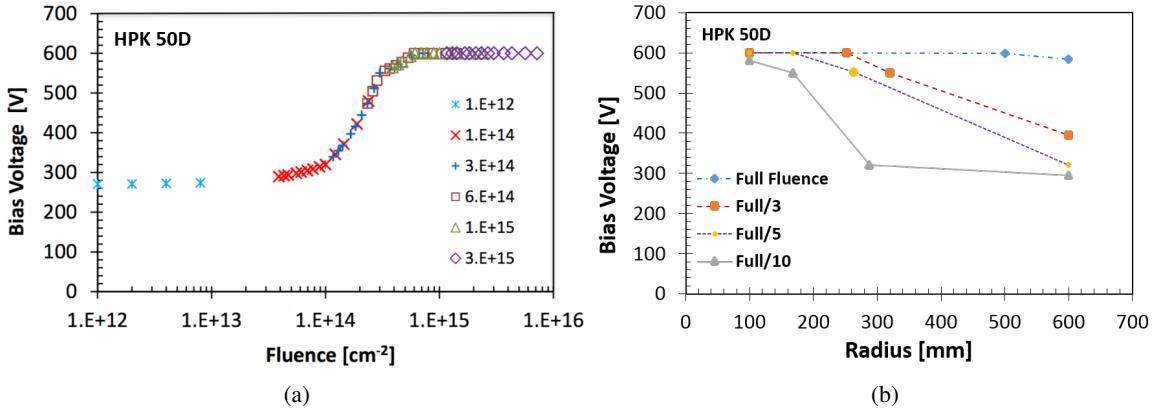


Figure 37: (a) Required bias voltage vs. fluence for different fluence levels at $R = 300$ mm. A rapid increase is seen between 10^{14} and $3 \times 10^{14} \text{ n}_{\text{eq}}/\text{cm}^2$. (b) Required bias voltage as a function of position along the longest readout row for four different fluence levels. Planned replacement of inner-radius modules not taken into account.

1256 Leakage current and power dissipation after irradiation

1257 In standard silicon sensors without gain, the leakage current originating from volume generation current
1258 increases linearly with fluence. However, for LGADs the situation is more complex due to the gain and
1259 its fluence evolution. The operation in gain mode leads to an increase of the leakage current, which is
1260 given by the product of the volume generation current and the current multiplication factor. As the gain
1261 decreases with irradiation and the generation current increases, the leakage current does not necessarily
1262 increase monotonically with fluence. The leakage current in multiplication mode contributes to parallel
1263 noise linearly, hence it is of high importance to run the sensors at low temperatures since cooling decreases
1264 the leakage current (roughly by a factor of 2 every 7°C).

1265 The total leakage currents (sum of bias ring and guard ring) for the different fluences as a function of the
1266 bias voltage shown in Figure 38(a) exhibit large increases for increased bias, partially due to the increased
1267 gain. The current through the bias ring (as indicator for the current through the multiplication region only)
1268 has been also measured separately without the guard ring current, which allows to determine the power
1269 density (power/area) in the multiplication region. The power can be minimised by operating the sensors
1270 at the lowest temperature and the lowest bias voltage. For the assumed lowest possible operation voltage
1271 (-30°C) and a bias voltage of VHR for 50D and VBD for B35, Figure 38(b) shows the measured power
1272 density as a function of fluence. It can be seen that the power density stays below the target of 30 mW/cm^2
1273 for the full fluence range up to the end of lifetime at -30°C . The power of B35 is typically significantly
1274 lower than for 50D.

1275 Developments for improved radiation hardness

1276 With the aim of improving the radiation hardness further, a project is underway in collaboration with
1277 RD50 to reduce the acceptor removal that is responsible for the gain loss.

1278 The first idea is to replace the gain layer boron (B) with gallium (Ga), which is heavier than B and is
1279 expected to have reduced formation of interstitial Ga compared to B under heavy irradiation. The second

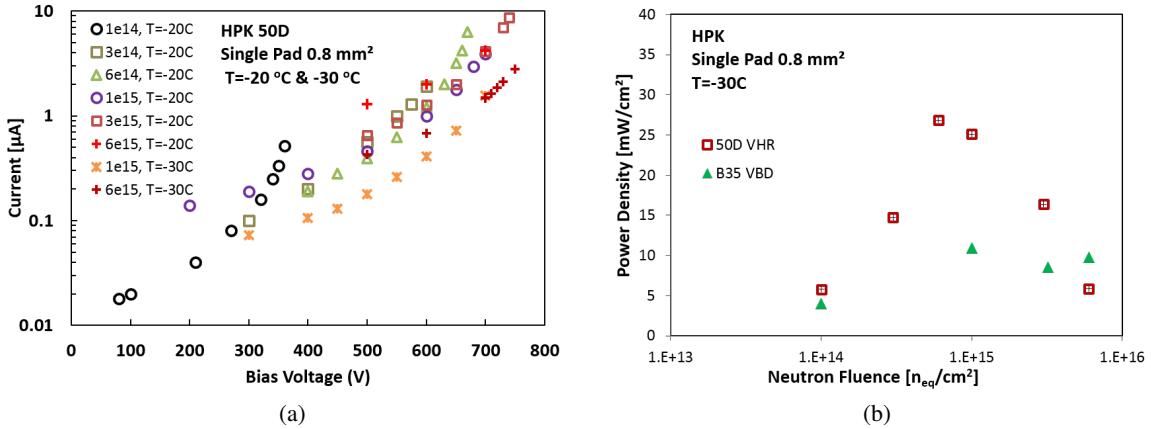


Figure 38: (a) Leakage current at $-20\text{ }^{\circ}\text{C}$ and $-30\text{ }^{\circ}\text{C}$ as a function of bias voltage for HPK 50D irradiated to the neutron fluences indicated. The current is the sum of the bias ring current which is multiplied in the gain layer and the current from the periphery collected by the guard ring. (b) Power density as a function of fluence at VHR (50D) and VBD (B35) at $-30\text{ }^{\circ}\text{C}$ for the leakage current collected by the bias ring only [26, 32].

one is to manufacture LGAD on carbon (C)-enriched wafers at the surface so that C will act as traps for the interstitial Si that was removed from its lattice by radiation. Hence the interstitial Si would react less with the multiplication layer B, hence reducing the probability for acceptor removal. Both CNM and FBK have performed LGAD runs with Ga implantation and C enhancement; studies are ongoing.

4.3.5 Roadmap for future sensor productions and activities

There are several new LGAD productions ongoing and in preparation. The most important aim will be to produce for the first time full-size LGAD sensors and determine their yield and uniformity. Sensors with the possibility of bump-bonding them to the next versions of the ALTIROC readout chips (arrays of 5×5 and 15×15 pads with under-bump-metallization) will be provided. Moreover, many test structures with technology variations are included such as different inter-pad gaps to optimise the fill factor and different slim-edge designs. It is planned to go as much as possible in parallel with HPK and CNM and implement the same device geometries with a similar mask layout in 6-inch runs of both vendors. 15 wafers are being ordered at each site, including $50\text{ }\mu\text{m}$ thickness as the baseline and a few $35\text{ }\mu\text{m}$ thick wafers as an option, with C-spray for possibly enhanced radiation hardness on a few wafers. Funding is shared between ATLAS and CMS institutions. A delivery is foreseen in Q2 2018, followed by extensive testing at HGTD institutes and beam tests. Priorities of the measurement program will be the evaluation of the time resolution, hit efficiency and power dissipation after neutron and charged hadron irradiation, as well as annealing and long-term stability studies. Based on these results, a new production iteration with optimised parameters and baseline designs is foreseen for 2019, before the pre-production is launched in 2021.

4.4 Readout electronics

The sensors will be read out by dedicated on-detector front-end electronics ASICs (bump-bonded to the sensors) which should keep the intrinsic excellent time resolution of the LGAD. The digitised ASIC output

signals will be transferred through electrical links, or *e-links*, on flex cables to the peripheral on-detector electronics located at the periphery of the detector between $700 \text{ mm} < R < 900 \text{ mm}$ (see Figure 62), and finally to USA15 with optical fibres. Taking the detector occupancy (which defines the bandwidth) into account, the detailed hit information will be read out only after a L0/L1 trigger (at about 1 MHz). The number of hits for each ASIC covering $R \geq 320 \text{ mm}$ in both layers will be transmitted at 40 MHz for online luminosity measurements.

The ASIC design is challenging due to the harsh radiation environment in which it will operate, and the need to meet the required performance while keeping power dissipation to a minimum. As mentioned before, one such requirement is to have a time resolution comparable to that of the LGAD. The contribution to the time resolution from the electronics is given by :

$$\sigma_{\text{elec}}^2 = \sigma_{\text{jitter}}^2 + \sigma_{\text{TW}}^2 + \sigma_{\text{TDC}}^2 \quad (5)$$

where σ_{jitter} depends on the noise and the pulse slope as defined in Eq. (4). The time walk (TW) contribution is expected to be negligible if an offline correction from a TOT measurement is applied. The TDC bin should be smaller than the timing resolution. With a TDC bin size of 20 ps the contribution to the resolution is negligible, of about 5 ps. An additional contribution could come from the clock distribution, although it is expected to be small. The time offset of each individual channel will need to be calibrated in situ, as described in Section 4.4.12.

This section discusses the requirements on the HGTD front-end electronics, followed by a description of the global architecture to satisfy these requirements and the front-end ASIC architecture proposed. More detailed descriptions are then given on some of the ASIC components, namely the single channel analog and digital components and the phase shifter. A preliminary measurement of a first analog prototype is presented. Finally the clock distribution and time offset calibration are discussed.

4.4.1 Requirements

Front-end ASIC requirements

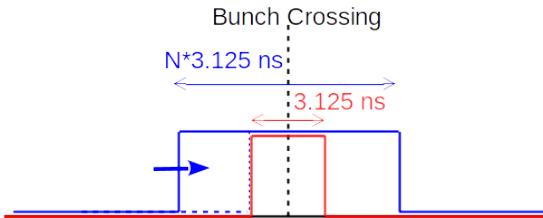
The following requirements of the ASIC are driven by the targeted 30 ps time resolution per MIP after irradiation obtained through the combination of multiple hits. A summary of these requirements is presented in Table 10.

- The ASIC will have to withstand high radiation levels. 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 Section 2.3, considering a 2.25 safety factor for the electronics. Thus, the maximal TID is 4.1 MGy (at $R = 120 \text{ mm}$) and decreases with radius. At $R = 320 \text{ mm}$ (the edge of the region that will be replaced by half the HL-LHC lifetime) it reaches a value of 1.6 MGy.
- The preamplifier dynamic range, up to 20 MIPs, has been determined by simulating electron showers in the HGTD detector.
- The electronics jitter for a charge of about 10 fC (equivalent to the charge deposited by a MIP in an 50 μm thick LGAD with a gain of 20) 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. The contribution to the time resolution from the TDC should be negligible and leads to a 20 ps TDC bin for the Time-Of-Arrival measurement (TOA)

1341 and a 40 ps (20 ps) TDC bin for the TOT measurement using a voltage (TZ) preamplifier. The TOA
 1342 and the TOT require respectively 7 and 9 bits. The time walk should be smaller than 10 ps over the
 1343 20 MIPs dynamic range.

- 1344 • The TOA and TOT information are transferred to the data acquisition system only upon L0/L1
 1345 trigger reception with latency up to 35 μ s [33], therefore necessitating a large size memory.
- 1346 • In order to measure the online bunch-by-bunch luminosity, each ASIC will report the sum of hits
 1347 within two different time windows. A schematic drawing of the windows is shown in Figure 39.
 1348 A first 3.125 ns wide window is centred at the expected arrival time of the particles from the
 1349 collisions. The second window is adjustable in length and position in steps of 3.125 ns, and will
 1350 count the number of particles arriving before and/or after those from the collisions. This side-band
 1351 would provide valuable information of the background, as described in Section 3.2. The length
 1352 and alignment is adjustable via configuration parameters, to be optimised based on operational
 1353 experience.

1354 To limit the bandwidth required for the luminosity measurement, the hit summary information
 1355 of only a subset of the ASICs is used. The current proposal is to use the sensors located at
 1356 $320 \text{ mm} < R < 640 \text{ mm}$, or equivalently $2.4 < \eta < 3.08$. The use of both layers would not provide
 1357 a significant increase in coverage with respect to just one of the layers, but the redundancy aids
 1358 in estimating and reducing the systematic uncertainty on the measured luminosity and provides
 1359 contingency in the event of failures in the instrumentation.



1360 Figure 39: Illustration of the time windows used for counting hits for the luminosity data. The smaller window (in
 1361 red) is 3.125 ns wide and is centred at the bunch crossing time. The width and relative location of the larger window
 1362 (in blue) can be set in steps of 3.125 ns through the control parameters.

- 1363 • Finally the ASIC power dissipation should be kept below 300 mW/cm², where the value is set by
 1364 the maximal cooling power provided by a single CO₂ cooling unit (for more details on the cooling
 1365 system see Section 5.1).

1366 **Bandwidth requirements**

1367 The bandwidth of each ASIC strongly depends on the radial region it covers, as shown by the distribution
 1368 of the average and maximum numbers of hits in each ASIC in Figure 40. The distributions are extracted
 1369 from a sample of simulated events with an invisibly decaying Higgs boson at $\mu = 200$. A dedicated buffer
 1370 is needed in the ASIC to average the rate variation and match the best speed of the e-link drivers/ low
 1371 power gigabit transceiver (lpGBT) inputs :

- 1372 • The largest average hit rate at small radius does not exceed 30 hits, equivalent to a rate of 720 Mb/s
 1373 (not including header). A bandwidth of 1.28 Gb/s will be enough for the innermost radius ASICs
 1374 (up to $R \simeq 150 \text{ mm}$).
- 1375 • For larger radii, a 320 Mb/s bandwidth can be used.

Pad size	$1.3 \times 1.3 \text{ mm}^2$
Detector capacitance	3.4 pF
TID and neutron fluence	Inner region: 4.1 MGy , $3.7 \times 10^{15} \text{ n}_{\text{eq}}/\text{cm}^2$ Outer region: 1.6 MGy , $3.0 \times 10^{15} \text{ n}_{\text{eq}}/\text{cm}^2$
Number of channels/ASIC	225
Collected charge (1 MIP) at gain=20	9.2 fC
Dynamic range (preamplifier+discr.) jitter at gain = 20	1-20 MIPs < 20 ps
Time walk contribution	< 10 ps
TDC binning	20 ps (TOA, TZ TOT), 40 ps (VA TOT)
TDC range	2.5 ns (TOA), 5 ns (TZ TOT), 10 ns (VA TOT)
Number of bits / hit	7 for TOA and 9 for TOT
Luminosity counters per ASIC	7 bits (sum) + 5 bits (outside window)
Total power per area (ASIC)	<300 mW/cm ² (<1.2 W)
e-link driver bandwidth	320 Mb/s, 640 Mb/s or 1.28 Gb/s
Latency for L0/L1 triggering	10/35 μ s

Table 10: Front-end ASIC requirements. The radiation levels include the safety factors defined previously and assume that the sensors and ASICs in the inner region ($R \leq 320 \text{ mm}$) are replaced after half of the HL-LHC program.

- For the luminosity data, the maximal number of hits per ASIC at $R > 320 \text{ mm}$ should be considered.
This number does not exceed 40. With a 4-bit header in addition to the 7+5 bits of data for the two hit counts, a 640 MB/s e-link driver and lpGBT speed is needed.
- The total number of lpGBTs is 1200 for the offline data from the entire detector and 880 for the luminosity. This amounts to 2080 up links, while one additional down-link is needed per lpGBT for the offline data in order to transmit the fast signals (clock,L0/L1 trigger) to the ASIC. Of the lpGBTs dedicated to offline data, 96 will transmit data at a speed of 1.28 GHz, 496 at 640 Mhz, and 608 at a speed of 320 MHz.

4.4.2 Time walk

The TOA, provided by the ASIC, will be measured with a fixed threshold discriminator and will suffer from time walk dispersion due to the signal amplitude variations⁷. To correct for it, the amplitude of the pulse needs to be measured. The most common way to perform this correction in silicon detectors (with small dynamics) is to measure the pulse width, which is proportional to the amplitude, with a TOT discriminator, instead of measuring the amplitude with an ADC. 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. Because of the use of a fixed threshold, any offset at the discriminator input needs to be controlled as it might induce a bias of the measured time.

The variation of the TOA as a function of the TOT is presented in Figure 41 for an input signal from 1 to 20 MIPs with a 3.5 pF detector capacitance. It is presented for two different types of preamplifiers, voltage (VA, in blue) and transimpedance (TZ, in red). A variation of about 700 ps is observed between 1 and

⁷ A Constant Fraction Discriminator (CFD) would in theory make the TOA measurement insensitive to the amplitude. However its implementation in an ASIC is quite complex for the performance required by the HGTD. A prototype of the CFD in 130 nm CMOS has been tested, showing a performance similar to the fixed threshold discriminator. Therefore the baseline for the HGTD is to use a TOT correction

Not reviewed, for internal circulation only

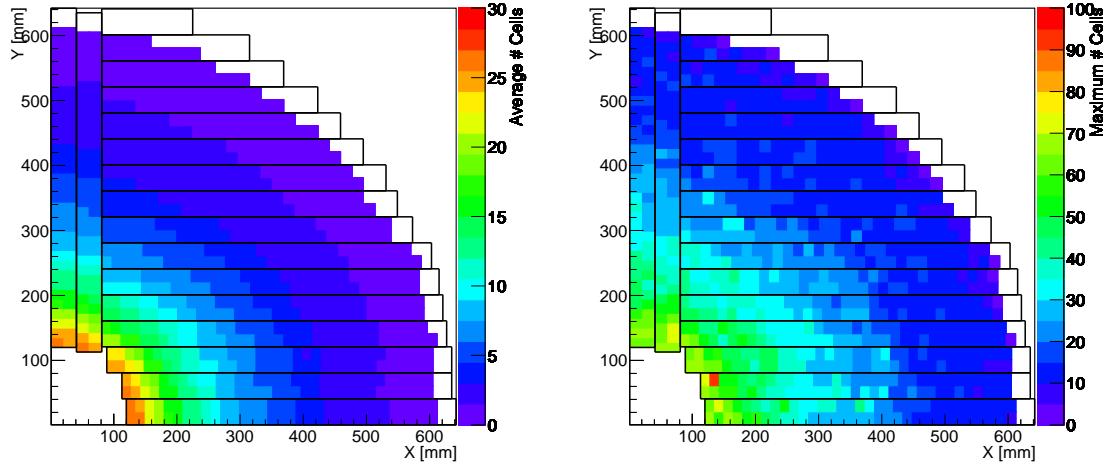


Figure 40: Average and maximum number of hits in each ASIC in one quadrant of the second layer in a simulated sample with $\langle \mu \rangle = 200$. The current simulation is limited to 600 mm while the HGTD design extends up to 640 mm explaining the white areas. The rectangles correspond to the readout rows defined in Section 4.

1392 20 MIPs in both cases, although the TOT excursion is much shorter in the case of the TZ preamplifier (and
 1393 therefore requires a more sensitive TDC). After correction, the time walk dispersion can be controlled to
 1394 better than 10 ps, resulting in a negligible contribution to the total time dispersion.

1395 4.4.3 Global architecture of the electronics in the HGTD readout chain

1396 The LGAD sensor module will have an area of $20 \times 40 \text{ mm}^2$ with $1.3 \times 1.3 \text{ mm}^2$ pixels, equipped with
 1397 two front-end ASICs. An overview of the HGTD readout chain is presented in Figure 42.

1398 The path starts on the left of the figure at the ALTIROC ASIC, bump bonded on the sensor, which will
 1399 read out 225 pixels. It will provide a precise time measurement of each hit in events accepted by the L0/L1
 1400 trigger.

1401 The digital output data are transmitted with e-links on flex cables (<700 mm) to peripheral on-detector
 1402 electronics boards located at the periphery of the HGTD ($R > 700 \text{ mm}$). Each flex cable serves a module
 1403 made of two ASICs and contains four differential e-links transmitting data at various speeds (320 Mb/s,
 1404 640 Mb/s or 1.28 Gb/s) depending on the ASIC position, to minimise the numbers of lpGBT and optical
 1405 links. These cables also contain the ASIC low voltage power supplies, the control signals and clock, and
 1406 the HV lines of the sensors.

1407 Peripheral on-detector electronics boards will be located at $R > 700 \text{ mm}$ and will be based on components
 1408 already developed for the ITk and Liquid Argon calorimeter Phase II upgrade. A possible implementation
 1409 of these boards is shown in Figure 43. A first board contains two rows of connectors for the flex cables on
 1410 the top and bottom faces of the printed circuit board, to fit within a 40 mm width. All input (configuration
 1411 parameter voltages) and output (data) signals should be routed inside this board to the optical board and the
 1412 voltage connectors (one individual line per module for the bias voltage + low voltage supplies). In addition

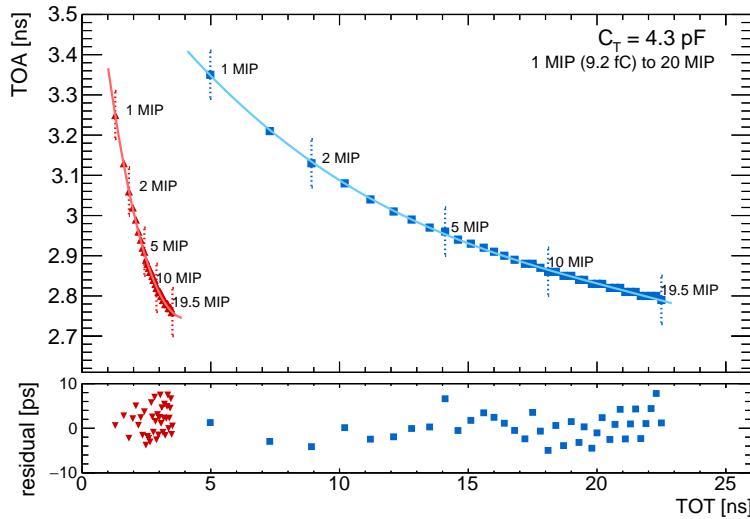


Figure 41: Simulation of TOA as a function of TOT with a 3.5 pF sensor capacitance and 0.8 pF parasitic ($C_T = 4.3$ pF) for an input signal from 1 to 19.5 MIPs. Two different preamplifier types are considered: a voltage preamplifier (blue squares) and a transimpedance preamplifier (red triangles). The bottom plot shows the residuals after correction of the amplitude variation.

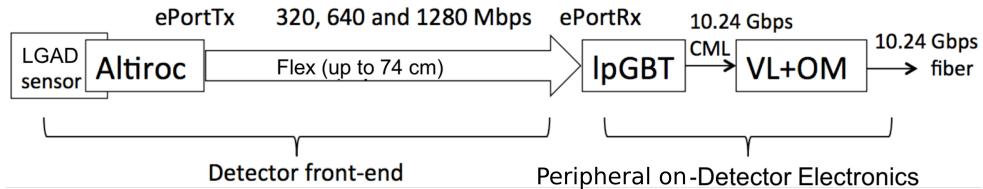


Figure 42: Upstream and downstream data flow. The ePortTx is the electrical transmitter protocol, defined by CERN, connecting the ASIC to the lpGBT. VL+OM is the Virtual Link+ on the optical module.

1413 this board will also house the DC/DC converters (19 for the longest readout row). The lpGBT ASIC and
 1414 the optical transmitter (or transceiver) modules developed by the lpGBT and the Versatile Link Plus (VL+)
 1415 common projects are located on the optical daughter board. The receiving end of this optical link system
 1416 provides the data to the data-acquisition system through FELIX [34] boards or to dedicated commercial
 1417 receiver boards for the luminosity information. The size of an event is estimated to be in average 190 kB,
 1418 with a range between 125 and 250 kB. A down-link to the detector will provide control information to
 1419 on-detector electronics (ALTIROC and the supporting electronics) such as clock, configuring and control
 1420 commands, monitoring and triggering signals.

1421 Each ASIC will also provide two integers representing the hit counts in the central time window and the
 1422 sideband for each event to allow a real-time measure of the luminosity. The information is sent via optical
 1423 links to the dedicated luminosity processing system in USA15. This consists of FPGA-based luminosity
 1424 boards that receive the data and aggregate the occupancy information over time, separately for each BCID.
 1425 Different sums are kept for the narrow and the wide time windows (see Figure 39). Two such sums are
 1426 kept for each of $2 \times 2 \times 4 \times 4 = 64$ regions, given by 2 endcaps with 2 layers each, 4 quadrants per layer and
 1427 4 radial divisions per quadrant. The exact number of regions can be tuned later as the implementation will

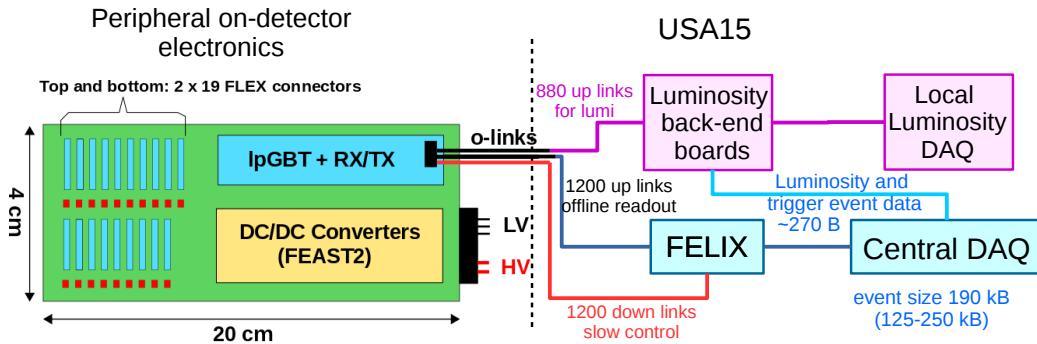


Figure 43: Possible implementation of the peripheral on-detector electronics for the longest readout row, and the readout chain. The flex connectors are located on the left; on the top right, the data transmitters and optical modules (IpGBT + VL + OM). The DC-DC converters are on the bottom right, where the low and high voltage connectors are. Three sets of optical links are connected to the IpGBT. The down links for slow control (in red) are connected to the FELIX boards in USA15, as well as the up links for the offline data readout. The up links with the luminosity information go to dedicated back-end boards.

be done in the firmware. The occupancy is aggregated over a time span of 1 second, after which the 128 short integers for each BCID are sent via Ethernet link to the control room for further processing (expected data rate is less than 1 MiB/s). The luminosity processing system will also implement the logics for the minimum-bias trigger using the hit counts as input. These criteria could include a minimum number of hits globally in HGTD, on one side only, and/or coincidences between layers to suppress noise. The signals representing the decision for each of these criteria will then be sent to the CTP as simple trigger input signals. Per-event occupancy information for each of the 64 regions will be sent to the data-acquisition system for events that pass the L0 trigger. The information will be buffered on the luminosity back-end boards until a L0 accept signal is received. Each of the 128 per-event occupancy numbers will be encoded as short integers (in total 256 bytes of data per event, not including header information). Additional bits representing the trigger signals will also be sent with this data, adding $O(10)$ bytes to the per-event size.

4.4.4 Front-end ASIC

Each pixel readout channel will consist of a preamplifier followed by a discriminator, both critical elements for the overall electronics time performance. The schematics for the single pixel readout are presented in Figure 44. Since the time walk will be measured using the TOT architecture described in Section 4.4.2, two TDCs are necessary. One for the TOA with a bin of 20 ps and a range of 2.5 ns providing 7 bits, and another for the TOT providing 9 bits. The bin and range of the TOT TDC depend on the type of the preamplifier, and will be of 40 ps bin and a range up to 20 ns for the voltage preamplifier, or configured with a 20 ps bin and a 5 ns range for the TZ preamplifier. The preamplifier is further described in Section 4.4.5, and the TDCs in Section 4.4.6. 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 μ s. If a trigger signal is received, the information is passed on to a secondary buffer named *matched hit buffer*, where it remains until ready for transmission to the off-pixel common electronics. These local memories are further described in Section 4.4.7.

Figure 45 shows the conceptual design of the entire HGTD ASIC with 225 channels. Each pixel is $1.3 \times 1.3 \text{ mm}^2$. They are arranged into a matrix of 15×15 , with a total matrix size of $19.5 \times 19.5 \text{ mm}^2$.

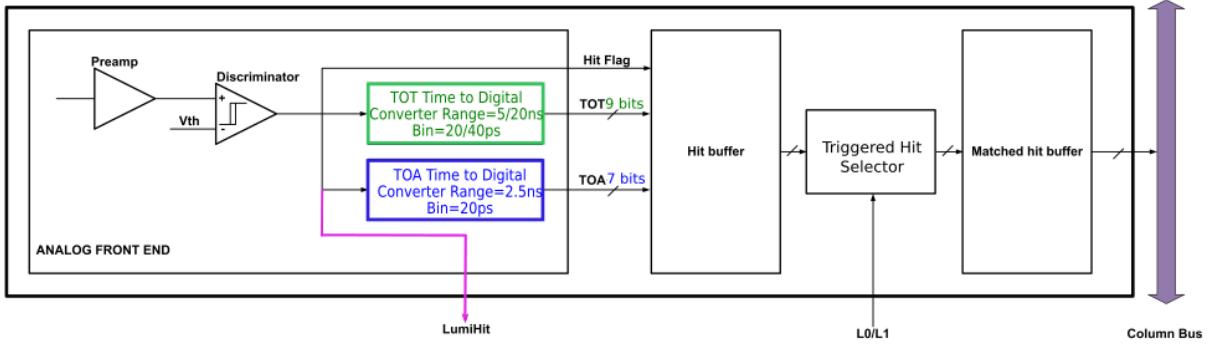


Figure 44: Schematic of a single-pixel readout block. The offline data from the TOA and TOT TDCs are stored in a buffer, while the luminosity information is gathered at the ASIC level.

1455 One of the sides of the complete chip will be slightly larger (~ 22 mm) to accommodate the off-pixel
 1456 common electronics and the Input/Output pads.

1457 The readout of the pixels is done by column, through an End-Of-Column (EOC) cell. A command
 1458 decoder unit receives the fast commands and clock signal from the central Trigger Data Acquisition
 1459 system (TDAQ). These are 8 bits on every bunch crossing, and a 320 MHz clock from which a 40 MHz
 1460 clock is generated. Based on this, a phase-locked loop (PLL) generates all the different clocks needed
 1461 to operate the ASIC, namely 320 MHz, 640 MHz and 1.28 GHz. These clock will be centred with an
 1462 accuracy of ~ 100 ps using a phase shifter, further described in Section 4.4.9.

1463 A control unit handles the readout of the pixel matrix, and will be equipped to handle the bunch crossing
 1464 identifier (BCID), L0/L1 trigger accept, and reinitialisation commands (to reset the buffers, registers and
 1465 re-initialise the ASIC). The bunch crossing information is synchronised with the BCID from the TDAQ by
 1466 the control unit. It consists of 12 bits that must be included in the ASIC to identify the events stored in the
 1467 pixels. When an L0/L1 accept signal is received by the ASIC, the control unit generates an internal trigger
 1468 signal and a trigger identifier (TrigID) that are passed to all the pixels. The TrigID is related through a
 1469 table with the corresponding BCID.

1470 Upon receiving the trigger signal, the control unit requests the EOCs to retrieve and store the data from
 1471 the pixels. Then it is moved into the Hit Data Formatting module, where it is packed in frames, serialised
 1472 and transmitted to the peripheral on-detector electronics through e-links. The transmission speed of the
 1473 e-link will depend on the radial position of the ASIC, and will be set via and Inter-Integrated Circuit bus
 1474 - I^2C - to one of three values: 320 Mb/s, 640 Mb/s and 1.28 Gb/s. It is connected to an equal speed port in
 1475 the lpGBT.

1476 Another main function of the control unit is to handle the readout of the luminosity measurement. The
 1477 output of the discriminator in each pixel is transmitted to the Luminosity Data Formatting unit. Here the
 1478 225 channels are summed over two different time windows, the sizes of which are determined through the
 1479 PLL. The smaller one (S1) is 3.125 ns wide, and is centred at the bunch crossing using the phase shifter
 1480 described in Section 4.4.9. The larger one (S2) contains S1, and its size can be configured to be a 3.125,
 1481 6.25 or 12.5 ns. Both sums (S1 and S2) are sent to a subtractor. The 8 bits of S1 and the 8 bits of the
 1482 subtraction (S2-S1) are then truncated to respectively 7 and 5 bits to reduce the total bandwidth. These

Not reviewed, for internal circulation only

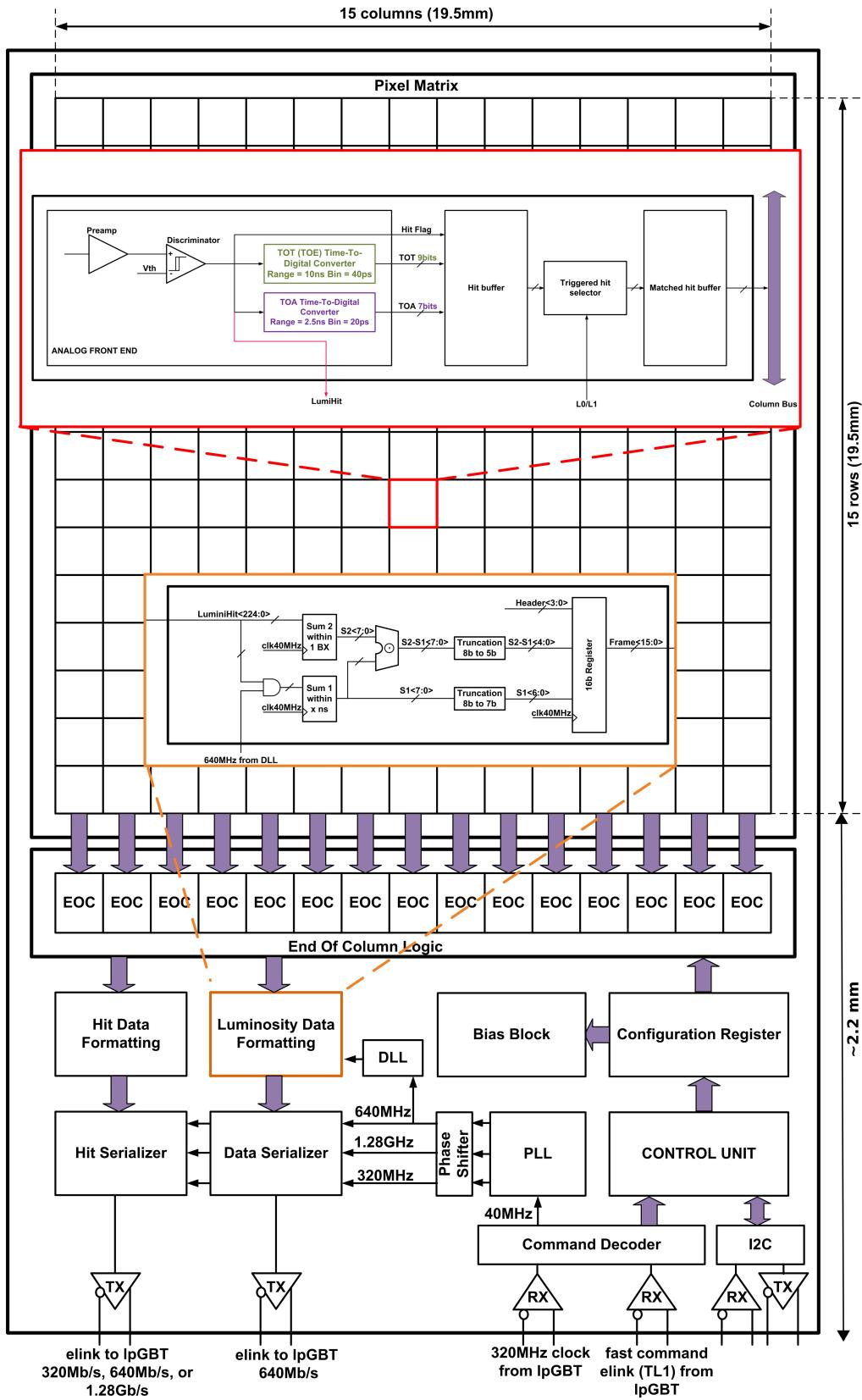


Figure 45: Schematic of the full HGTD ASIC. The top part represents the 15x15 pixel matrix, with the single pixel schematics. The bottom part shows the off-pixel modules, including the schematics of the Luminosity Data Formatting unit.

1483 12 bits, as well as the 4 bits of the header, are then serialised at a rate of 40 MHz in the Data Serializer
 1484 and sent to the lpGBT through a 640 MHz e-link.

1485 The off-pixel electronics also include several programmable digital to analog converters (DACs) to generate
 1486 different bias currents for all analog blocks of the ASIC, a band-gap, a temperature sensor (under
 1487 consideration) and the configuration register. The latter is used to set different features of the ASIC, such
 1488 as the values of the DACs, the transmission rate of the hit data and the PLL bias currents or frequencies.
 1489 A configuration register will also be present in each pixel. The I^2C link mentioned previously is also used
 1490 to readout all configuration registers in order to check if SEU events have corrupted their content, and to
 1491 retrieve information from the control unit about the status of the ASIC; the information related to data
 1492 corruption will be passed on to the hit serializer.

1493 **4.4.5 Front-end preamplifier**

1494 From an electronics point of view the sensor can be modelled as a transient current source in parallel with
 1495 the combined capacitance of the sensor and parasitics (which will be referred to as the total capacitance,
 1496 C_T). When fully depleted, the value of the sensor capacitance is inversely proportional to the sensor
 1497 thickness and the capacitance value grows with the area. The typical capacitance for a fully depleted
 1498 sensor is 3.4 pF for a $1.3 \times 1.3 \text{ mm}^2$ sensor with a 50 μm thickness.

1499 The baseline preamplifier architecture is a broadband preamplifier with a Common Source configuration,
 1500 consisting of an input transistor (M_1) and a follower transistor (M_2), as displayed in Figure 46(a).
 1501 Figure 46(b) shows an alternative architecture with a transimpedance preamplifier. In both cases the size
 1502 of M_1 and the bias currents are optimised to minimise the noise and the power consumption. The rise
 1503 time of the preamplifier is optimised to be equal to the drift time of the sensor ($t_s \sim 0.5\text{--}1 \text{ ns}$) in order to
 1504 minimise the jitter. This minimum jitter is given by :

$$\sigma_{\text{jitter}} = \frac{e_n C_T}{Q_{in}} \sqrt{t_s} = \frac{C_T}{Q_{in}} \sqrt{\frac{2kT t_s}{g_m}}, \quad (6)$$

1505 where e_n is the noise spectral density, g_m is the transconductance of the preamplifier input transistor
 1506 and Q_{in} is the input charge. Both preamplifier architectures, followed by a fast discriminator, have been
 1507 simulated using the 130 nm TSMC⁸ kit provided by CERN with various C_T values and considering that 1
 1508 MIP would deposit a 10 fC charge, which corresponds to an amplification gain of 20 in the LGAD. Two
 1509 input signals were simulated, a calibration signal (to be compared with test bench measurements) and an
 1510 LGAD-like signal.

1511 The resistor R_2 (15 K or 25 K) in 46(a) can absorb the sensor leakage current, estimated to be between
 1512 1 and 2 μA . The leakage current would cause the output of the preamplifier to drift by an amount of the
 1513 order of $R_2 \times I_{\text{leak}}$. The threshold of the discriminator that follows the preamplifier must then be changed
 1514 accordingly. This can be done using the 7-bit DAC threshold correction that is integrated for each channel
 1515 allowing a correction within $\pm 50 \text{ mV}$.

1516 Figure 47 compares the simulated output of the preamplifier and the discriminator for both preamplifier
 1517 architectures using an LGAD-like input signal. The first two plots show the pulse of the preamplifier
 1518 (top) and the output of the discriminator (bottom) for the voltage preamplifier shown in Figure 46(a). In

⁸ TSMC stands for Taiwan Semiconductor Manufacturing Company. The technology has been qualified up to 4 MGy [35].

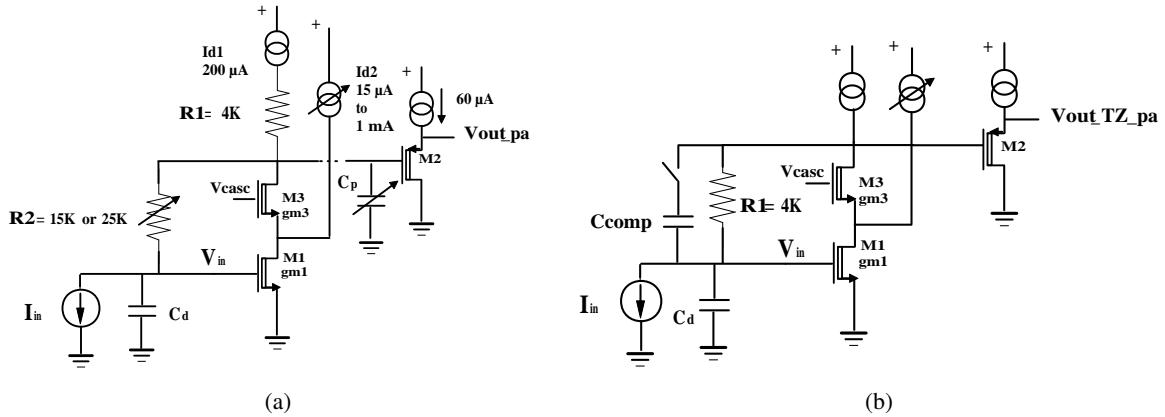


Figure 46: Architecture of the voltage preamplifier (a) and transimpedance (TZ) preamplifier (b) implemented in the latest ASIC design, ALTIROC0_v2.

both cases, two input signals are compared: 5 MIP in blue, and 1 MIP in green. The 5 MIP signal is 5 times larger, has a smaller rise time and larger TOT, which is reflected in the discriminator output, where the blue pulse starts first and is much wider than the green one. This comparison also applies to the two bottom plots, where the same is shown for the transimpedance preamplifier presented in Figure 46(b). The 5 MIP signal is shown in red, while the 1 MIP is presented in orange. In addition, it should be noted that the output of this type of preamplifier has a much shorter TOT than the previous one; this is reflected also in the plot presented previously in Figure 41. This would allow to reduce the conversion time, but at the same time require higher precision from the TDC.

A first prototype of the ASIC, ALTIROC0_v1, has been designed using the TSMC 130 nm process, and has already been through two iterations. It originally implemented the voltage preamplifier, the TOT and a CFD, but the latter has been removed. The latest version, corresponding to ALTIROC0_v2, contains eight channels, four for each of the preamplifier architectures presented in Figure 46. The design of the chip includes also the bump bonding pads, and a first LGAD array was assembled with ALTIROC0_v1 in summer 2017 (see Section 4.2.1), and tested with charged pions at CERN in September. The size of the chip is $3.4 \times 3.4 \text{ mm}^2$ to accommodate the bump bonding to a sensor, but the area used for a single channel electronics is about $200 \mu\text{m}$ by $100 \mu\text{m}$.

The ALTIROC0_v1 chip has been characterised with a single ASIC wire-bonded on a test board. Measurements have been performed using a picosecond generator to provide a voltage test pulse with a rise time smaller than 100 ps. This voltage is then injected through an integrated 100 fF capacitor that can be selected for each channel using the ASIC configuration parameters. This input current signal is faster than a typical LGAD current signal.

The jitter performance was measured as a function of the injected charge for capacitances of $C_T \sim 3.8$ and 4.8 pF and is shown in Figure 48(a). The total capacitance combines the soldered capacitance (1 or 2 pF) and the parasitics arising from the ASIC and board, which were estimated in this case to be 2.8 pF. The charge injected during the tests was varied from 5 fC to 120 fC. The threshold of the discriminator was set to 5 fC, corresponding to 50% of a MIP in an LGAD sensor with a gain of 20. The constant floor of 7 ps is reached for input charges larger than 20 fC and is attributed to the jitter of the generator. At 10 fC, the measured jitter is 17 ps for $C_T \sim 3.8$ and 22 ps for $C_T \sim 4.8$. The simulation of the jitter is also presented in this plot for the simulation of ALTIROC0_v1, implementing the post-layout view of the

Not reviewed, for internal circulation only

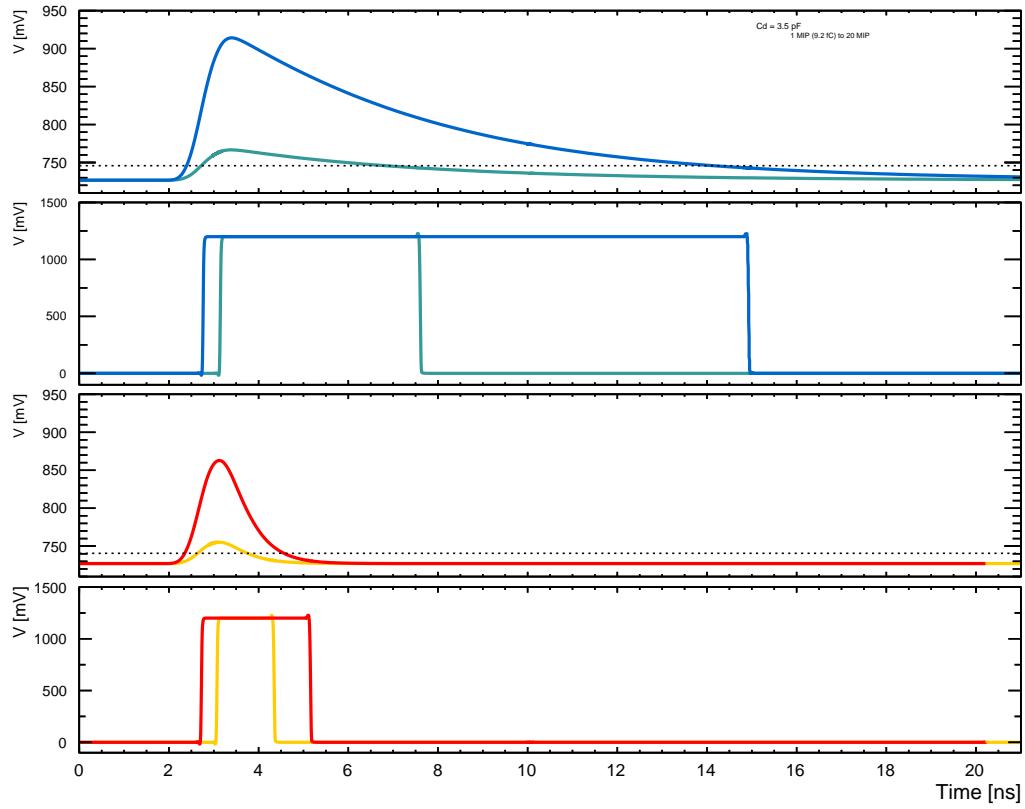
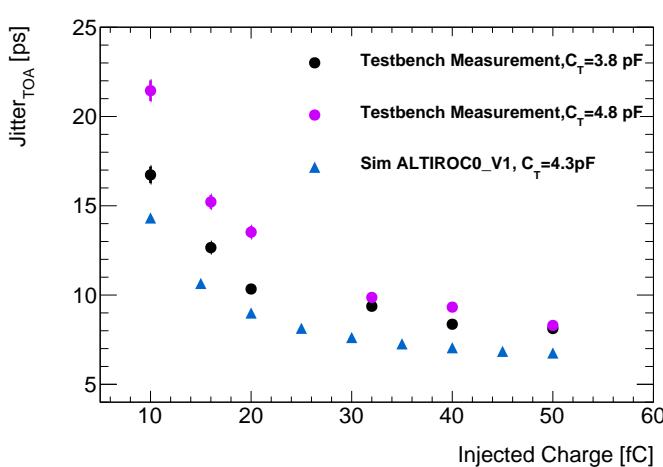
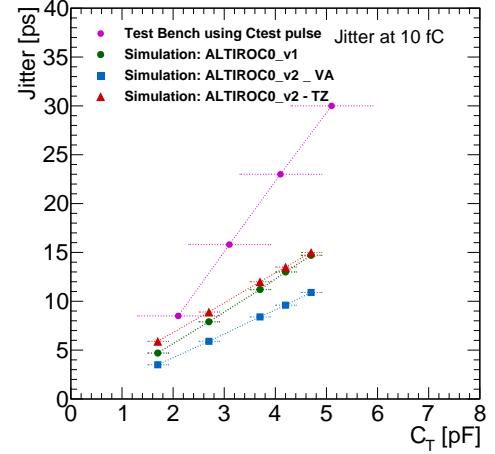


Figure 47: Simulation of the voltage preamplifier (top, where blue is 5 MIP and green is 1 MIP) and transimpedance preamplifier (bottom, red is 5 MIP and orange is 1 MIP). For each type of preamplifier, the output pulse is shown first and the discriminator output is shown below.



(a) Jitter as a function of input charge.



(b) Time jitter as a function of total capacitance (C_T).

Figure 48: (a) Simulated and measured jitter as a function of the input charge in ALTIROC0_v1. (b) Jitter as a function of C_T measured in testbench for ALTIROC0_v1 and simulated for ALTIROC0_v1 and ALTIROC0_v2. The value of C_T is obtained combining the soldered capacitance (sensor capacitance in the simulation) and the parasitic capacitance. The horizontal error bars correspond to the uncertainty in the parasitics.

1548 circuit (including all the parasitics). A total capacitance of $C_T = 4.3$ pF is used, that combines a simulated
 1549 sensor capacitance of 3.5 pF and parasitics estimated to be 0.8 pF. The jitter attributed to the generator
 1550 has been added to the simulation for a better comparison.

1551 Figure 48(b) shows the measured and simulated jitter as a function of C_T , using as input a Ctest pulse. The
 1552 simulations of ALTIROC0_v1 and both ALTIROC0_v2 preamplifiers (voltage and transimpedance) are
 1553 shown. The total capacitance is obtained as the sum of the simulated sensor capacitance (or soldered in the
 1554 case of the measurement) and the parasitics, which are (2.1 ± 0.8) pF in the measurement and (0.7 ± 0.2) pF
 1555 in the simulation. The generator contribution of 7 ps has been subtracted. The jitter increases linearly with
 1556 the detector capacitance, but with a different slope for each case. The difference between the measurement
 1557 and simulation of ALTIROC0_v1 can be mainly attributed to the different noise in the simulated ASIC
 1558 (1.2 nV/ $\sqrt{\text{Hz}}$) and the measured ASIC+board (1.7 nV/ $\sqrt{\text{Hz}}$). For $C_T \sim 3.5$ pF a jitter smaller than 15 ps
 1559 is obtained for 10 fC (LGAD gain of 20). The simulation of the voltage preamplifier in ALTIROC0_v2
 1560 shows a smaller slope, which points to the possibility of improving the jitter performance of the modules
 1561 prototype in the next iteration.

1562 The measured power consumptions of the preamplifier and of the discriminator are 420 μW and 375 μW
 1563 respectively.

1564 Preliminary test beam measurements have shown resolution values of around 42 ps for the current
 1565 combination of sensors and electronics. Additional test beam measurements are planned for 2018.

1566 4.4.6 Time-to-Digital Converter

1567 The Time-to-Digital Converter (TDC) is designed using TSMC 130 nm technology. The target timing
 1568 resolution (quantisation step) of 20 ps is below the gate-propagation delay in 130 nm technology, thus the
 1569 Vernier delay line configuration is employed. This configuration consists of two lines (see Figure 49),
 1570 each composed of a series of delay cells implemented as differential shunt-capacitors, controlled by a
 1571 voltage signal that determines their delay.

1572 The timing resolution is determined by the difference in the delays of the cells in each line. In the 'slow'
 1573 line, the control voltage fixes the delay of each cell to 140 ps, while on the 'fast' line it fixes it to 120 ps.
 1574 The START signal (output of the discriminator) enters the 'slow' delay line while the STOP signal (end
 1575 of measurement window) enters the 'fast' delay line. Although initially the START signal is ahead of
 1576 the STOP one, each delay-cell stage brings them closer by an amount equal to the difference between the
 1577 slow and fast cell delays, i.e. 20 ps. The number of cell stages necessary for the STOP signal to surpass
 1578 the START signal represents the result of the time measurement with a quantisation step of 20 ps. A
 1579 cyclic structure is employed to reduce the number of cells per line and results in a smaller occupied area.
 1580 Since the time measurement is initiated only upon signal detection (instead of at each time-measurement
 1581 window), the reverse START-STOP scheme is used as a power-saving strategy.

1582 The TOT TDC will employ an additional coarse delay line for extending the range to 20 ns, while the
 1583 Vernier delay line (identical to the one used in TOA TDC) will provide high resolution. For the voltage
 1584 preamplifier, the TOT TDC will be configured for 40 ps resolution and 20 ns range, while for the TZ
 1585 preamplifier the TOT will be configured with 20 ps resolution and 5 ns range.

1586 The TDC power consumption is dependent on time-interval being measured. For the TOA TDC with
 1587 2.5 ns (full dynamic range), the average power consumption over the 25 ns measurement period is about
 1588 5.2 mW. It will become 3.5 mW for the time-interval equal to half dynamic range. Thanks to the reverse

Not reviewed, for internal circulation only

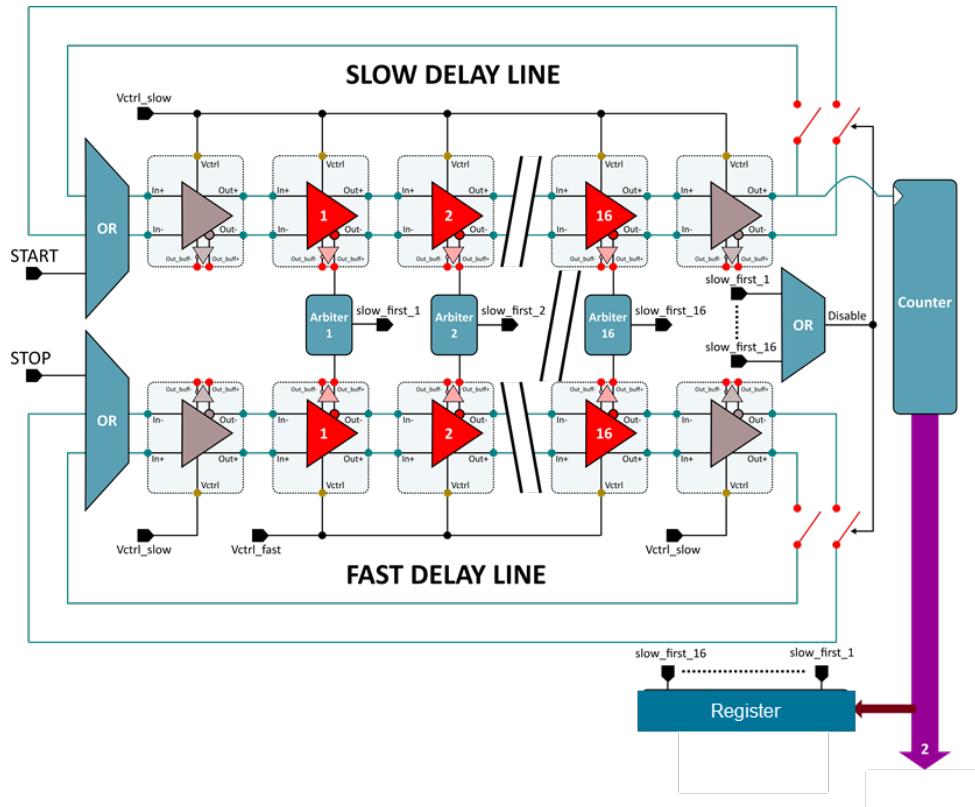


Figure 49: Schematics for the TDC showing the 'slow' delay line and the 'fast' delay line. The 20 ps speed difference between the two is used to provide the 20 ps time measuring bins

1589 START-STOP operation, the power consumption of the TDC is much lower in the absence of a hit over
 1590 threshold. This results in an average power consumption per pixel of 0.4 mW, assuming a time interval
 1591 uniformly distributed (1.25 ns average) and a maximal pixel occupancy of 10%. The average power
 1592 consumption for the TOT TDC is 0.5 mW and 0.36 mW for voltage and TZ preamplifier configurations
 1593 respectively.

1594 Simulations of both the TOT and TOA TDCs have been implemented; the simulated conversion time as a
 1595 function of the time interval is presented in Figure 50. The TOA TDC (Figure 50(a)) was simulated for a
 1596 time range of 2.7 ns; the conversion time of a time interval of 2,5 ns is 25 ns. The simulation of the TOT
 1597 TDC for voltage preamplifier (Figure 50(b)) shows that the conversion time for a time interval of 20 ns is
 1598 ~ 28 ns. These are preliminary results and studies are ongoing.

1599 4.4.7 Local memory

1600 Each pixel electronics is composed of an analog part, already described, and a digital part. The purpose
 1601 of the digital part is to temporarily store the data related to a hit and select hits of events that have been
 1602 triggered, or, in other words, skip the stored hits that are not associated to a trigger. The hit buffer is the
 1603 one that first receives the 16 bits of data (7 bits TOA and 9 bits TOT), and can be implemented in two
 1604 different ways:

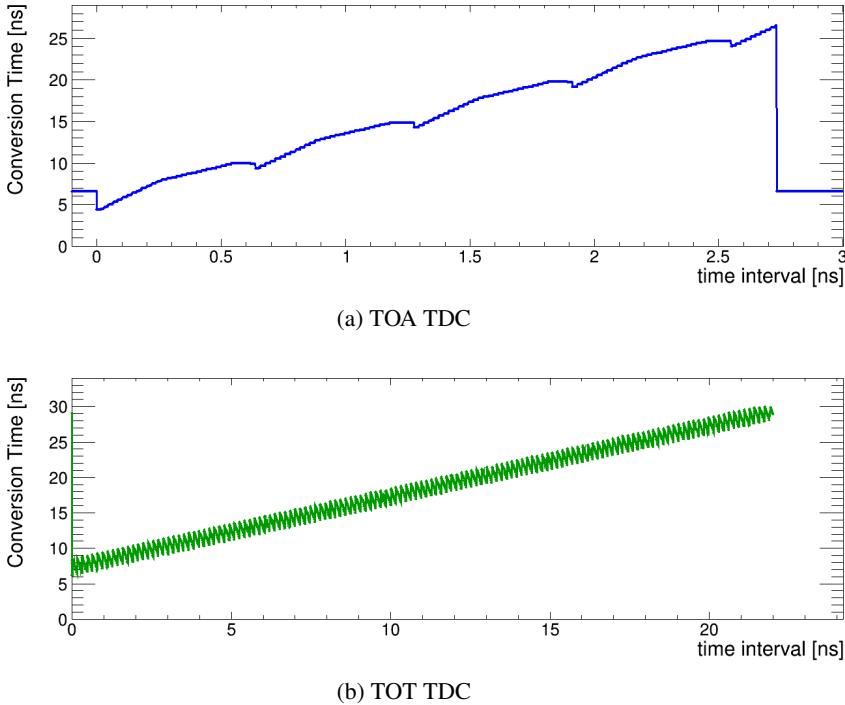


Figure 50: Simulation of the conversion time for the TOA TDC (a) and the TOT TDC (b).

- **Full buffering:** at each bunch crossing, the pixel bits (TOA,TOT, hit flag) are loaded, although if no hit is present only the hit flag is stored. It therefore requires a memory position per bunch crossing, which amounts to a depth of 1400×17 bits to cope with the L1 latency of $35\mu s$.
 - **Partial buffering:** the TOA and TOT of a pixel is stored only if there is a hit. In order to assign a hit to a specific event, the 12 bits of BCID must be added. Given that the average number of hits per pixel in $35\mu s$ is 210, an estimation of the depth of such a memory starts at $210 \times 28b$. In principle a partial buffer would then be approximately 4 times smaller than a full buffer.
- For each received trigger, the triggered-hit selector checks if there is data in the hit memory associated to that event. If there is, it is loaded into the matched hit buffer with a TrigID supplied by the control unit. This identifier allows to know which trigger event the data stored in the matched hit buffer are associated to.
- The way that the triggered-hit selection is carried out depends on the hit buffer implementation. A full buffer would need to be implemented as a circular memory, allowing to store data in a continuous way for each bunch crossing. It has two memory pointers for reading and writing. The latter is incremented from position 0 to 1399 during $35\mu s$, and then goes back to position 0. The reading pointer is managed by the triggered-hit selector, which transfers the TOA and TOT information if when receiving a trigger it finds a hit flag equal to one.
- A partial buffer could be implemented as a simple “first-in-first-out” (FIFO) memory. The triggered-hit selector would handle its content by continuously comparing the BCID of the outgoing data with a delayed copy of the BCID in the ASIC (the delay should equal the trigger latency). When the internal and external BCID match, and a trigger is received, it transfers the data to the next memory.

1626 The matched hit buffer operates as an average rate memory. It will allow to cope with event-to-event
 1627 fluctuations in the number of matched hits and to keep the bandwidth of the ASIC lower than 1.28 Gb/s.
 1628 Matched hits are stored there until they are read by the control unit and passed to the hit data formatting
 1629 module. Studies are ongoing to optimise its design.

1630 An implementation in 130 nm CMOS and a simulation of a local 17×400 SRAM (10 μ s latency) has been
 1631 done, with a standard 6T cell configuration. Assuming a 10% maximal occupancy, the average power
 1632 consumption of the pixel local memory is 57.2 μ W. For a 1400 depth SRAM, a power consumption 1.5/2
 1633 times higher is expected; the increase is related to the fact that the memory should be divided in multiple
 1634 banks.

1635 **4.4.8 Matrix readout process**

1636 When the command decoder receives a trigger command, this is immediately passed to the pixel matrix
 1637 as well as the corresponding TrigID. Simultaneously, the BCID and the TrigID are loaded into the trigger
 1638 table. Events stored in the table are managed by the control unit. However, it might happen that the table
 1639 be full. In that case, the control unit would generate an error message that would be transmitted to the
 1640 TDAQ through the I^2C link. For each fetched entry in the table, the control unit initiates the readout of
 1641 the pixels associated to that trigger. The process works as follows:

- 1642 1. the TrigID is passed to all pixels as a requested trigger ID (RqtTrigID).
- 1643 2. a circuit on-pixel checks if there is a matched hit with the same trigger ID than the requested trigger
 RqtTrigID. If there is, a hit flag on-pixel is asserted
- 1645 3. only pixels with asserted matched hit flag are readout. The data plus the address of the read pixels
 are load into buffers placed at the end of column. When the data of a pixel has been read, the hit
 flag is de-asserted.
- 1648 4. the end of column buffers are read by the control unit. The column address is added to each data
 read from one of the buffers. Data are passed to the hit data format circuit where the frames for
 transmission are built and serialised.
- 1651 5. once all the pixels with a matched hit associated to the requested trigger ID have been read, that
 event is removed from the table, the next trigger event stored in the trigger table is read and the
 process starts again.

1654 **4.4.9 Phase shifter**

1655 The phase shifter is a function block located in the off-pixel electronics used for the adjustment of the
 1656 clock phase for the clock system in the ASIC. It receives the clock CMOS signals of different frequencies
 1657 (320,640 and 1280 MHz) generated by the PLL and outputs them with the same frequency but with an
 1658 adjusted phase. This module is required to provide a shift step smaller than 100 ps, and additional jitter
 1659 below 5 ps, and a power consumption around 10 mW.

1660 The design presented here is adapted from an IGBT designed in the 65 nm process. As can be seen in
 1661 figure 51, the core of the phase shifter is a delay-locked loop (DLL) in which the delay line is used to delay
 1662 the input clock signal. The 640 MHz clock signal reuses the delay line of 16 delay cells (not included the
 1663 dummy cell at the end of the delay line) in the DLL. The 320 MHz clock signal is delayed in other two

1664 delay lines that are controlled by the DLL. Two coarse phase adjustment circuits are needed for these two
 1665 clock signals to keep the delay line length of 16 delay cells. The time resolution is 1/16 of the 640 MHz
 1666 clock period, equal to 97.6 ps.

1667 An estimation of the performance of the phase shifter is presented in figure 52 based on a previous version
 1668 of the design that shared the same delay cell, charge pump and phase detect as the current one. The DLL
 1669 used had 32 cells instead of the proposed 16. Figure 52(a) shows the jitter as a function of the selected
 1670 phase for the cases with the highest and lowest jitter found. The maximum value remains below 1 ps.
 1671 Figure 52(b) shows the delay time as a function of the selected phase for the same two cases presented
 1672 previously. The maximum variation found is of 1 ps. The power consumption is 8.4 mW in the typical
 1673 case in the previous design. The clock phase inaccuracy caused by mismatch of the delay cells is found
 1674 to be no more than 3 ps (RMS) in the mismatch simulation. It is estimated that the new design will have
 1675 a similar jitter phase accuracy performance; the power consumption would be of around 13 mW because
 1676 it will have 3 delay lines. New studies will be available soon.

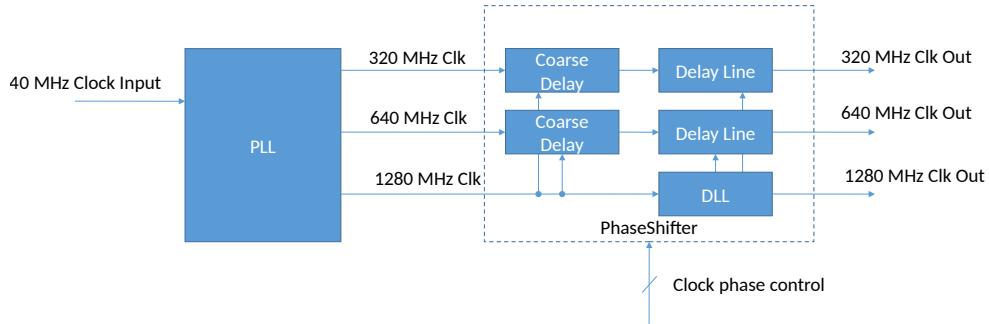


Figure 51: Top level diagram of the proposed phase shifter

1677 4.4.10 Power consumption of ASIC

1678 Table 11 summarises the average power consumption of a single pixel electronics assuming 10% occupancy
 1679 corresponding to the inner radius ASIC (worst case). The values presented here have been estimated
 1680 using simulations. The nominal single-channel power consumption is 2.66 mW

1681 In addition to the 225 single pixels, the power dissipation of the other ASIC blocks (hit formatting cell, rate
 1682 average FIFO, PLL, clock distribution, etc...) has been estimated to be 100 mW. The estimated power
 1683 consumption of the phase shifter is estimated to be around 13 mW. With a 10% occupancy the nominal
 1684 power consumption of the entire ASIC is 700 mW, within the requirements of 1200 mW (300 mW/cm²).
 1685 Using the expected occupancy as a function of the position of each ASIC, the total front-end power
 1686 consumption for one HGTD layer is 2.1 kW. A more detailed study is presented later in Figure 56, where
 1687 the power dissipation of the modules is shown as a function of the radius, taking into account the average
 1688 number of hits expected.

1689 4.4.11 ASIC development next steps

1690 The following steps are expected up to the final chip design:

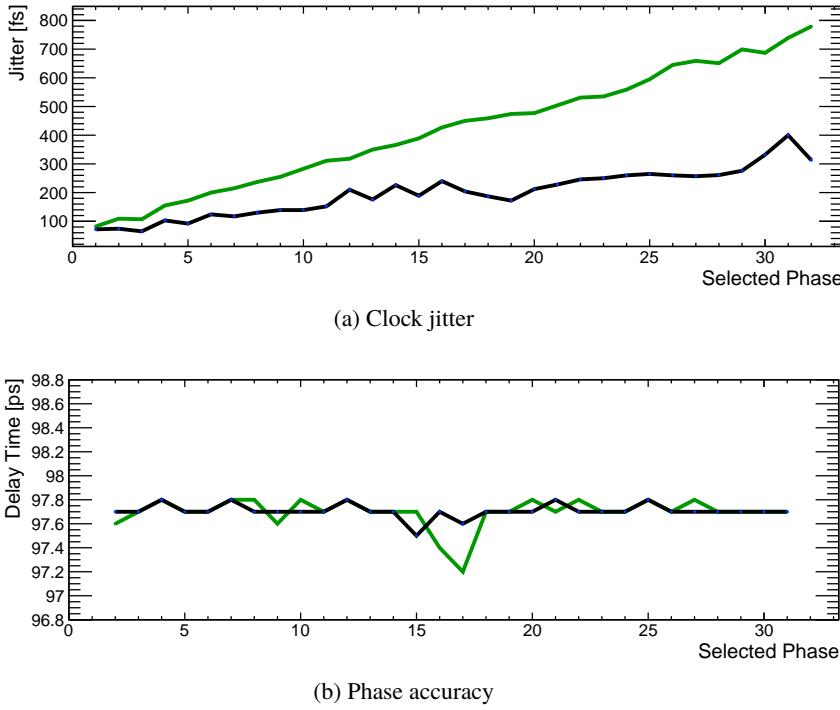


Figure 52: Simulation of the jitter (a) and the phase accuracy (b) of the phase shifter. The simulation is shown for the cases with maximum and minimum jitter.

Component	Nominal [μW]
Preamplifier	462
Discriminator+DAC	375
TDC TOT (10ns/40ps)	500
TDC TOA (2.5ns/20ps)	405
Local Memory	920
Total Pixel	2666

Table 11: Single pixel average readout power consumption. For the TDC and local memory (full buffer option), a 10% occupancy factor is applied.

- The ALTIROC0_v1 ASIC first performance was satisfactory, but presented some shortcomings. Several improvements have been implemented in ALTIROC0_v2, reducing the noise level, increasing the preamplifier bandwidth, and testing an alternative preamplifier architecture.
- A 25 channels chip containing both the analog (preamplifier + discriminator) and digital (TDC + local memory and possibly clock phase shifter) parts of the single pixel readout will be submitted in June 2018. The schematics of the elements implemented in ALTIROC1 are presented in Figure 53. It will contain variants for the architecture of the TDC and the preamplifier, as well as an “on-chip calibration module” (or pulser), the purpose of which is to calibrate the absolute value of the phase of the external clocks that are used by the TDCs. This chip will be bump bonded to a sensor and is expected to validate the single pixel readout chain with intensive characterisation before a TDR. Simulation of the data formatting and memory is being performed in parallel but will not be

1702 included in this circuit.

- 1703 • A first submission of the final chip is expected in the spring of 2019, and a second iteration can be
 1704 encompassed in the project schedule.

1705 4.4.12 Clock distribution and calibration

1706 The TDCs implemented in the HGTD sensor readout ASIC ALTIROC use an external clock signal, of (a
 1707 multiple of) the 40.079 MHz LHC bunch crossing clock. The clock signal must be stable and in phase
 1708 with the average bunch crossing time as given by the machine. To avoid affecting the timing resolution
 1709 of the detector (30–50 ps/sensor), the short-term (bunch to neighbouring bunch) RMS phase-jitter of the
 1710 clock should be well below 5 ps. The same restriction applies to the long-term peak-peak jitter (*drift*) over
 1711 periods equal or larger than a millisecond. In particular, the phase jitter and drift between the clocks of
 1712 different ASICs should respect such limits.

1713 It should be noted that phase *stability* rather than the actual phase value is of importance, because it is
 1714 foreseen that the time offset of the individual channels can be measured and corrected to first order by a
 1715 dedicated calibration procedure, and subsequently fine-tuned with the actual event data.

1716 For each readout row, an electrical clock distribution must be implemented for several reasons. First, the
 1717 space available precludes the use of optical-fibre receivers at individual ASICs; secondly, optical clock
 1718 receivers and fibres would receive an intolerably high radiation dose at smaller radii; and lastly, for reasons
 1719 of cost. With temperature fluctuations of O(1 K) and distances of O(1 m), copper lines can be used.

1720 The LHC clock will be distributed in USA15 using a low jitter clock fan-out circuit, and then to the
 1721 lpGBTs. Currently, the AFP collaboration [36] is successfully using the On-Semi NB7L series of clock
 1722 fan-out buffers with ≤ 65 ps rise time and 0.2 ps RMS random jitter. A collaboration is being set up in
 1723 CERN EP/ESE with both ATLAS and CMS to develop the best clock distribution scheme. The frequency-
 1724 adjustable but fixed-phase clock output of the lpGBT will be used to distribute the clock to each ASIC.
 1725 They are expected to have a jitter smaller than 5 ps. The phase of these clocks, as previously explained,
 1726 will be adjusted internally in the ASIC with the phase shifter described in 4.4.9.

1727 The time of arrival of a hit measured in the HGTD will be different between pixels and will need to be
 1728 calibrated to keep the expected time resolution of 30–50 ps per pixel. The origin of the variations of the
 1729 measured time can be static (radially dependent time of flight, signal/clock distribution in the ASIC and
 1730 in flex cables of different length) or time dependent (day/night variation of the LHC clock if not centrally
 1731 corrected in ATLAS). For instance, the current LHC clock shows a slow day/night drift of about 200 ps;
 1732 not correcting such an effect over an LHC fill could induce 50–60 ps time dispersion.

1733 Figure 54 shows the inclusive time distribution of hits in a simulated $Z \rightarrow e^+e^-$ sample. It shows a
 1734 Gaussian core derived from the time dispersion of the LHC collision (~ 250 ps) with tails from particles
 1735 with longer paths, e.g. low- p_T particles from displaced decays of long-lived SM hadrons, or loopers.
 1736 With 10000 hits per pixel, a precision of a few ps per pixel can be easily reached. This would translate
 1737 into 10^5 (10^6) events to be recorded at inner (larger) radius. With a few kHz data taking, such a time
 1738 offset correction per pixel could be calculated online about every 10 minutes at the nominal HL-LHC
 1739 luminosity. For a global shift as induced by the LHC clock, one can build the time distribution using all
 1740 the pixels connected to the same clock (two ASICs, i.e 550 channels), and monitor faster time drift.

1741 The TOT, affected by the gain degradation under irradiation, will be monitored by the same set of data.

Not reviewed, for internal circulation only

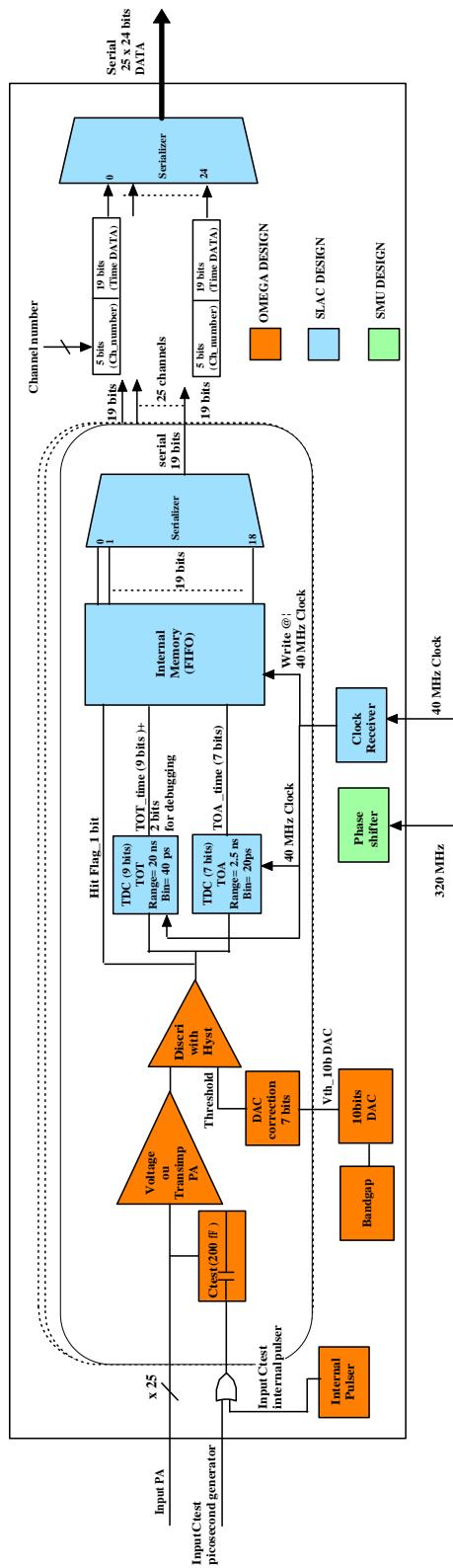


Figure 53: Schematics showing the design of ALTIROC1.

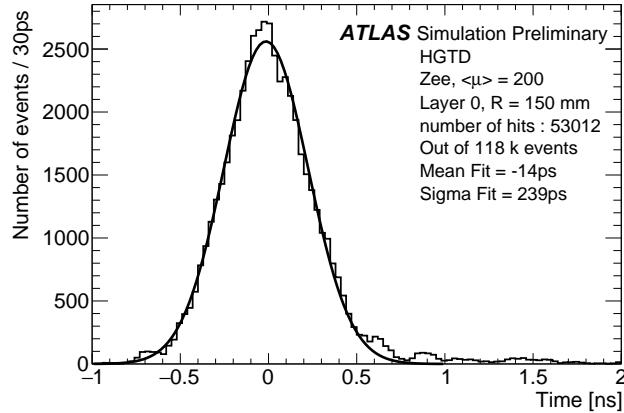


Figure 54: Distribution of hit times at $R = 150$ mm using a $Z \rightarrow e^+e^-$ sample. The hits for 9 pixels are used to increase the statistics. The small bias of the time distribution comes from the time reconstruction method used.

1742 4.5 Outlook and path towards TDR

1743 The R&D activities are expected to continue up to the end of 2020 with the following main milestones to
 1744 be achieved before the TDR is delivered in Q1/2019 :

- 1745 • The production of new sensors for 2018 has been launched both at HPK and CNM and the devices
 1746 should be available over summer. A similar mask is used by each producer in order to make a direct
 1747 comparison of the performance. The aim is to optimise the edge/interpad structure, to produce
 1748 larger-size sensors and study their performance before and after irradiation. A decision about the
 1749 sensor thickness (35 or 50 μm), the pad size and the sensor size (40×20 or $40 \times 40 \text{ mm}^2$) is expected
 1750 to be taken for the TDR.
- 1751 • A full validation of the one-pixel readout (preamplifier+discriminator+TDC+local memory) per-
 1752 formance should be demonstrated using the ALTIROC1 prototype to be submitted in June 2018.
 1753 Test bench measurements to characterise the ASIC will be done over Q4/2018 and Q1/2019. In
 1754 parallel the specifications of the final chip should continue with a preliminary design review to be
 1755 organised in Q1/2019. A conceptual design of the peripheral on-detector electronics and of the bias
 1756 voltage and low voltage distribution should be achieved.
- 1757 • Flex cable prototypes will be produced before summer 2018 and their performance validated (data
 1758 transmission, high-voltage distribution, . . .). The ALTIROC1 ASIC will be bump-bonded to HPK
 1759 and CNM sensors and β source measurements could be done over Q1/2019, but not in test beam
 1760 measurements before the TDR.
- 1761 • A prototype of the clock distribution will be built and is likely to be installed before the beginning
 1762 of Run 3. The goal is to demonstrate in-situ the stability of the clock distribution.

1763 5 Detector installation and infrastructure

1764 This section describes the design of the detector vessel, cooling system and the integration of HGTD in
 1765 ATLAS. As discussed previously, the space allocated to the HGTD is very limited and constrains many
 1766 of the HGTD engineering parameters. The detector design must allow easy and fast integration into the
 1767 ATLAS detector, and enable repairs and a partial replacement during LHC shutdown periods. The present
 1768 design consists of a single cylindrical cold vessel per endcap, with the module layers inside. The overall
 1769 dimensions of this vessel have been presented in Table 1, with a more detailed description of the size of
 1770 each component in Table 9. In addition, the routing of the services should be in the gap of 50 mm against
 1771 the endcap wall.

1772 The different components of the hermetic vessel are presented in Figure 55. The full detector thickness
 1773 in z including supports, front and rear cover is 75 mm, plus 50 mm of moderator to protect the ITk and
 1774 the HGTD from the back-scattering neutrons originating from the endcap calorimeter. The moderator is
 1775 divided into two disks, as can be seen in the drawing. The HGTD cold vessel will be located at the z
 1776 position of $3420 \text{ mm} < z < 3545 \text{ mm}$ from the interaction point while the first and last active layers will
 1777 be located at $z = 3435 \text{ mm}$ and $z = 3485 \text{ mm}$. The total weight per endcap is approximately 350 kg, of
 1778 which 70 kg comes from the moderator inside the vessel.

Not reviewed, for internal circulation only

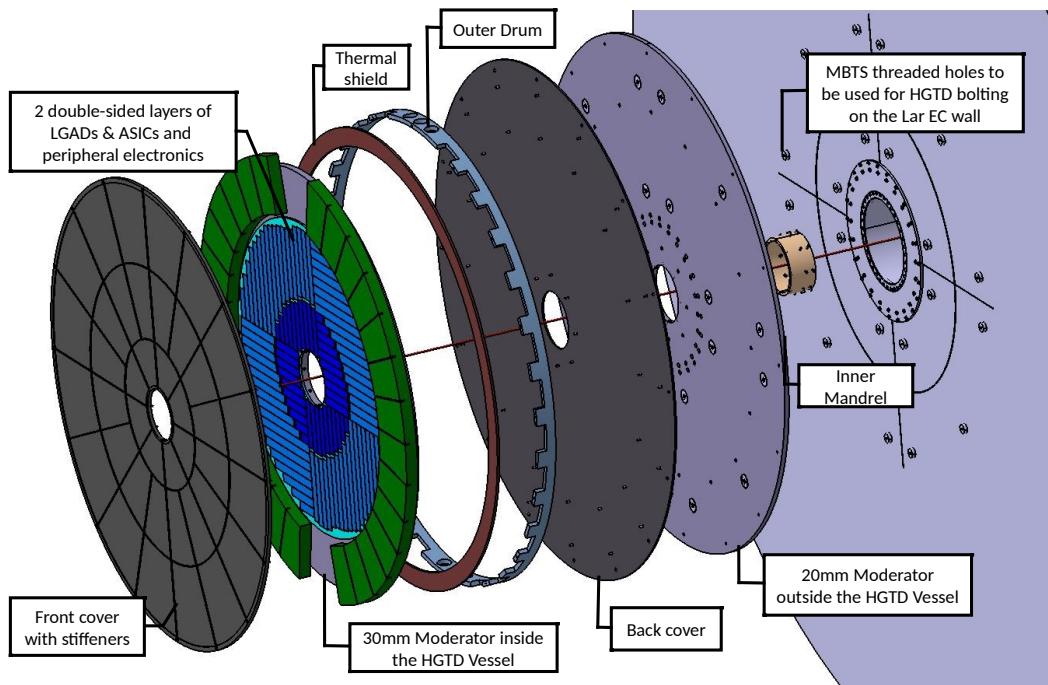


Figure 55: Global view of the various components of the hermetic vessel and moderator, excluding the cooling plates and the detector components assembled in them.

1779 5.1 Cooling system

1780 The HGTD cooling system will be based on a 2PACL-CO₂ (two Phase Accumulator Controlled Loop of
 1781 CO₂), that will be integrated into the general cooling system developed for the ATLAS ITk detector [4].
 1782 The CO₂ is pumped in liquid state from an external primary chilling source and partially evaporates as it
 1783 absorbs the heat dissipated by the HGTD components. Within each pipe, a small amount of CO₂ flows at
 1784 high pressure in the form of small drops, and enough space is left for the vapour to circulate. A highly
 1785 efficient heat extraction is achieved by making use of the large latent heat for a liquid to vaporise, meaning
 1786 that not only less fluid is needed to extract a certain amount of heat, but also that the temperature of the
 1787 liquid phase remains constant, while that of the vapour increases only slightly. The cooling power is then
 1788 determined by how much CO₂ is left in liquid state. Because it is used in mixed states (liquid and vapour),
 1789 a significant mass reduction is introduced when comparing with other liquid mono-phase refrigerants.
 1790 Logically, the piping diameter can also be smaller, which is an additional benefit given the reduced space
 1791 available for the detector. Lastly, CO₂ cooling is radiation hard, and is thus appropriate for the harsh
 1792 environment the HGTD will operate in.

1793 5.1.1 Requirements

1794 An operation temperature of -30°C must be maintained inside the HGTD vessel, close to the modules,
 1795 with a stability of few degrees. As discussed in Section 4.3, the need to keep the operating temperature
 1796 as low as possible arises because, after irradiation, the leakage current of the sensors increases with
 1797 temperature (by a factor of 2 for every 7°C). These conditions will limit the heat dissipation and ensure
 1798 the good performance of the sensors.

1799 The power dissipation of the modules was studied as a function of their position, taking into account the
 1800 ASICs (with the average number of hits as shown in Figure 40) and the sensors in VHR mode. The results
 1801 are presented in Figure 56, separately for each side of the cooling plate. The modules at the innermost
 1802 radius present the largest power consumption, 820 mW per ASIC area (205 mW/cm^2), with 700 mW
 1803 coming from the ASIC. The values decrease as a function of the radius, reaching 625 mW per ASIC area
 1804 (156 mW/cm^2) at the outer radius.

1805 Table 12 shows a breakdown of the power consumption estimated for the various components of the
 1806 detector, which define the maximum cooling power needs (25 kW in total, i.e. 12.5 kW per endcap).
 1807 The ASICs will be the component with the highest power consumption, reaching 175 mW/cm^2 at the
 1808 innermost radius. The expected power consumption of irradiated sensors at $3.7 \times 10^{15} \text{ n}_{\text{eq}}/\text{cm}^2$ running
 1809 at a temperature of -30°C is $< 30 \text{ mW/cm}^2$ for the entire radius range. This corresponds to the maximum
 1810 irradiation expected in the sensors considering a replacement of the inner part of the HGTD detector
 1811 ($R < 320 \text{ mm}$) after half of the HL-LHC program and including the appropriate safety factors (as
 1812 summarised in Section 2.3). The power consumption of the peripheral on-detector electronics has been
 1813 computed assuming a 70% efficiency for the DC/DC converters. The table also shows an estimate where
 1814 power consumption of the ASICs, the flex circuits and of the peripheral on-detector electronics have been
 1815 computed including a safety factor of 1.5 with respect of the ASIC power consumption. This factor is
 1816 applied to take into account possible larger occupancies and calibration tests in which many pixels would
 1817 be fired simultaneously.

1818 The operating temperature of the peripheral on-detector electronics is quite flexible and can be in the
 1819 range of -30°C up to 20°C , making the cooling and stability requirements of these components much

Not reviewed, for internal circulation only

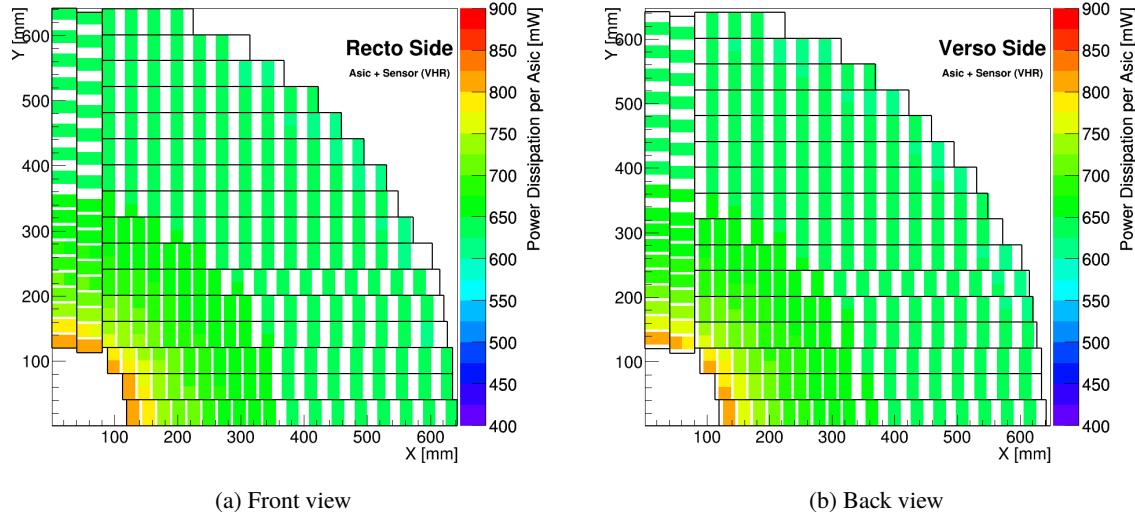


Figure 56: Power dissipation of the ASIC and sensor combined, shown per ASIC, for each side of a cooling disk.

less strict. 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.

5.1.2 Cooling design

The cooling layout is illustrated in Figure 57(a), including all inlet-outlet manifolds and transfer lines located at the top and bottom of the detector. Based on the technology used for the ATLAS Insertable B-Layer detector and industrial standards, tri-axial stationary vacuum insulated rigid transfer lines will be used. Two transfer lines are dedicated to the module cooling and a third one provides the cooling of the peripheral on-detector electronics.

Liquid CO₂ flows from the transfer lines into the capillary pipes (shown in yellow in the drawing), which have an inner diameter of 0.75 mm, and are up to 3 m long. These capillary pipes supply the 7 cooling

Component	Power consumption	Total [kW]	Maximal [kW]
Sensor	< 30 mW/cm ²	1.9	1.9
ASIC	<175 mW/cm ²	8.5	12.8
Flex cable	< 100 mW/flex	0.5	1.1
HGTD cold vessel heaters	75 W/m ² -175 W/m ²	0.33	0.33
EC calorimeter cryostat heaters	120 W/m ² , 50% up to R = 1600 mm	< 0.6	0.6
Peripheral on-detector electronics	dominated by DC/DC converter	3.25	4.9
Total for CO ₂ cooling		15.1	21.6

Table 12: Power consumption estimations of the various HGTD components and the total for the HGTD (for a total number of 7888 flex cables, 7888 sensors of 20 × 40 mm² each; 6.3 m² in total and 15776 ASICS). The last column includes a safety factor of 1.5 for the electronics.

loops embedded in each half-disk cooling plate. The cooling loops are positioned in a concentric half-circle layout. The distance between the loops at $R > 320$ mm is 20 mm. For $R < 320$ mm it is reduced to 15 mm in order to take into account the larger heat dissipation (modules with 80% overlap), thus keeping a uniform temperature distribution. The characteristics of the pipes in the cooling loops are guided by the suggestions of the CERN Cooling group (EP-DT-FS). They will be made of stainless steel 304L; a non-magnetic, easy for manufacturing, bending and welding material. The wall thickness is set to be 0.3 mm, which should sustain the CO₂ high pressure level (the cooling system has safety release valves at 130 bar; to be tested according to CERN safety factor rules). The inner diameter will be of 3.0 mm, and the length should be in the range between 4 and 6 m. The average transfer capacity of such loops is thus 100 W/m, fulfilling the previously discussed requirements. Standard stainless-steel fittings are proposed for testing prior to installation.

Not reviewed, for internal circulation only

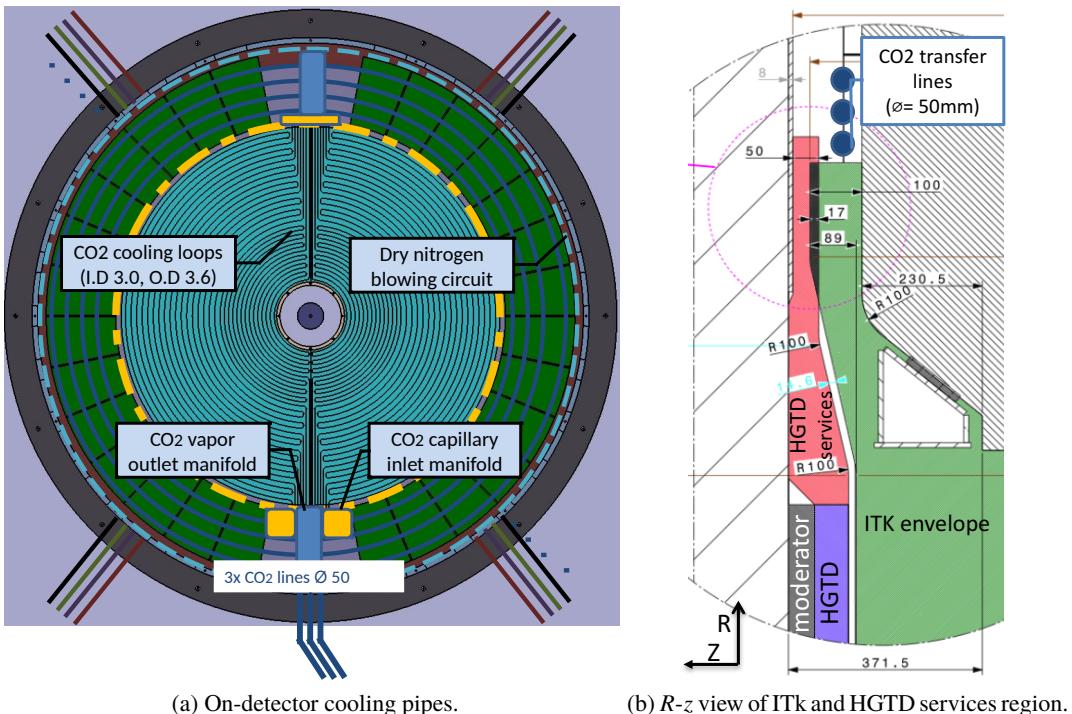


Figure 57: (a) Cooling layout (b) R - z view of the region around $R = 1$ m. The red region shows where services of the ITk and HGTD need to be extracted to the outside, sharing different slots in ϕ .

The plan is to bring in fixed cable trays the three transfer lines per endcap of about 5 kW and 50 mm outer diameter each as shown in Figure 57(b). The possible location of these connections is marked in light blue in the photograph presented in Figure 58, where the side of the endcap cryostat is seen. During shut-down periods, in the opening/closing procedure of the endcap calorimeter, these cooling transfer lines will be temporary disconnected/reconnected through a compact junction box to be located at $z \approx 8$ m. The connections during opening/closing procedures are shown in Figure 59, with the ATLAS detector in closed (a) and open (b) position. While the opening/closing of the endcap calorimeter takes place, the cooling will stay disconnected and should be reconnected after the movement is finished. During this period the temperature inside the vessel will increase; a disconnection time of around 24 hs is deemed acceptable.

The cooling support plates are composed of a carbon fibre structure and a thermally conductive foam-like

Not reviewed, for internal circulation only

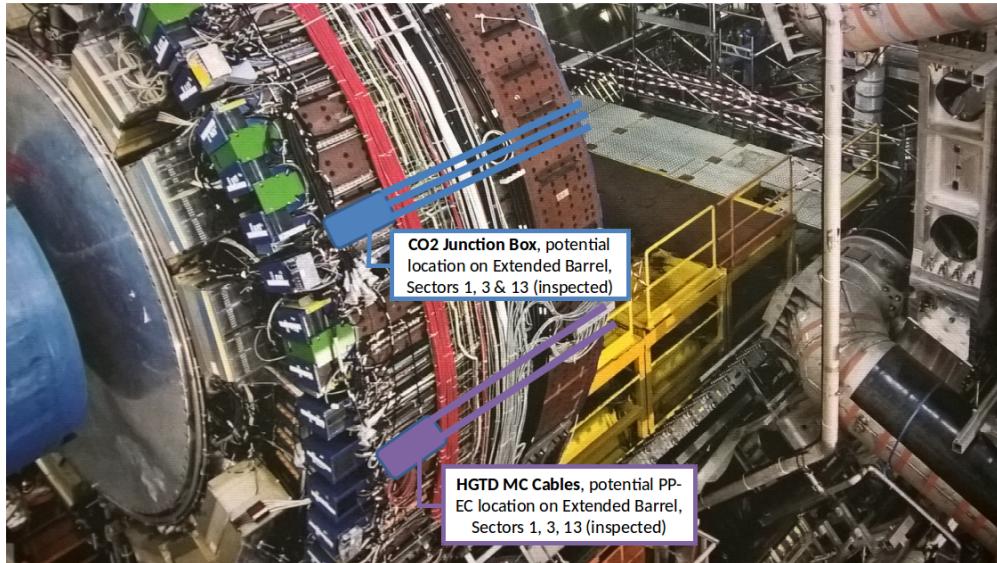


Figure 58: Picture showing the side of the endcap calorimeter, withdrawn from its usual position. The possible routing of services has been marked in colour.

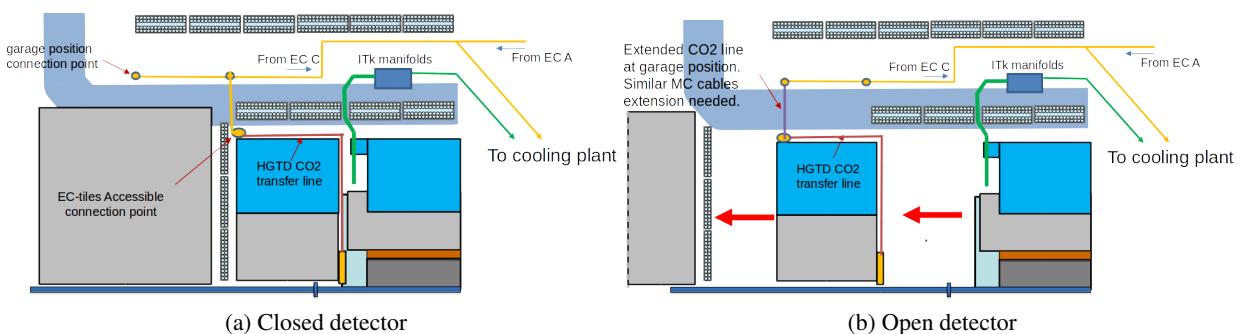


Figure 59: View of the possible routing of the HGTD cooling pipes with the ATLAS detector in (a) closed and (b) open position, showing the access to the connection/disconnection box before and after the movement of the endcap calorimeter.

graphite or equivalent material. Carbon fibre is a stiff material with high thermal conductivity, thus providing a highly uniform temperature distribution in the cooling disk. A core made of conductive foam will absorb the different thermal expansion of the cooling pipes and carbon-fibre panels. The central plate and the outer rectangular staves will be made from a material with a high thermal conductivity (AlSiC, PEEK graphite reinforced, Carbon Fibre low epoxy) and bolted to the support plate with thermal grease media.

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 submitted to several thermal cycles to study thermo-mechanical behaviour, temperature distribution, CO₂ cooling parameters and glue layers integrity between modules and carbon fibre skins.

1863 **5.1.3 Cooling performance**

1864 A finite element analysis of the current CO₂ cooling design has been performed for one side of a layer.
 1865 The model considers a uniform power dissipation of the modules of 235 mW/cm² for $R < 320$ mm and
 1866 160 mW/cm² for larger radius (leaving a small margin with respect of the estimated values presented
 1867 previously in 5.1.1). It considers the input cooling temperature of -30 °C at the centre of each cooling
 1868 loop. The calculation has been made in a static regime, with no external exchange. The results are
 1869 presented in Figure 60. The obtained temperature distribution is of (-27±1)°C in the active area; reaching
 1870 -29 °C at $R > 640$ mm.

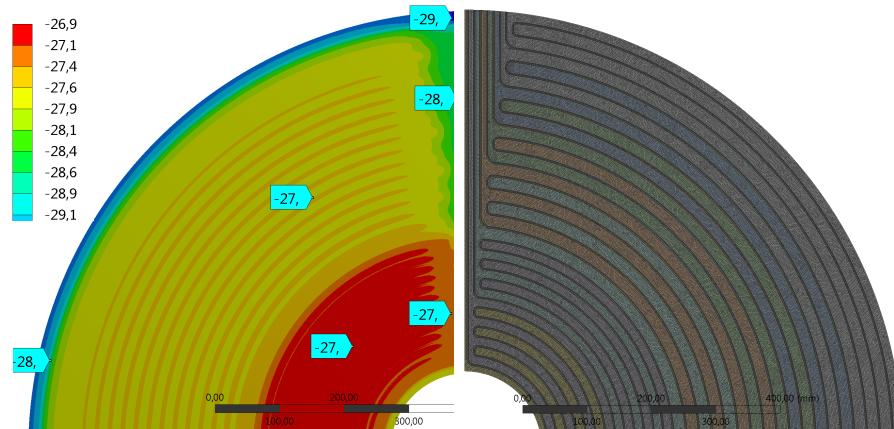


Figure 60: The left quarter disk shows the calculation of the temperature distribution in the cooling plate. On the right the FEA is shown.

1871 The model used in these calculations will be improved by implementing a more realistic representation of
 1872 the modules, the radial dependency of the power dissipation, and the thermal contact between components.
 1873 A study of the power dissipation of the sensors as a function of irradiation was already presented in
 1874 Figure 38. In addition, a study of the thermal runaway will be performed to cover for a possible excess in
 1875 heat productions from the electronics or a lack of CO₂ liquid. Such studies will provide important input
 1876 to the optimisation of the cooling requirements presented in Section 5.1.1.

1877 **5.1.4 Cooling and power monitoring and control**

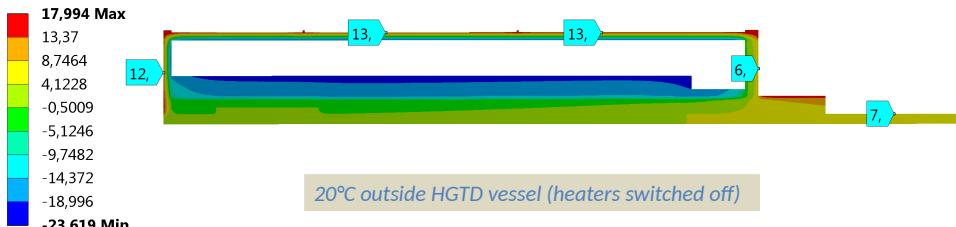
1878 The detector environment and parameters related to the cooling and power consumption will be monitored
 1879 through links to the Detector Control System (DCS). It provides input to the interlock and power supply
 1880 units as needed, for detector and operational security. In particular it will keep track of the humidity,
 1881 temperature and pressure, with sensors installed in the whole volume of the HGTD: in the modules, flex
 1882 cables, peripheral on-detector electronics, cooling pipes and the inside of the vessel walls. The outside
 1883 temperature and humidity will also be monitored to ensure that adequate dryness and ambient temperature
 1884 are maintained. In order to prevent thermal runaway, all elements that could be damaged from overheating
 1885 will be equipped with Negative Temperature Coefficient sensors linked to the interlock system. Both
 1886 sudden or slow changes are recorded and can trigger corrective actions from the cooling system or the
 1887 power supply devices, as necessary.

1888 5.2 Detector cold vessel and global structure

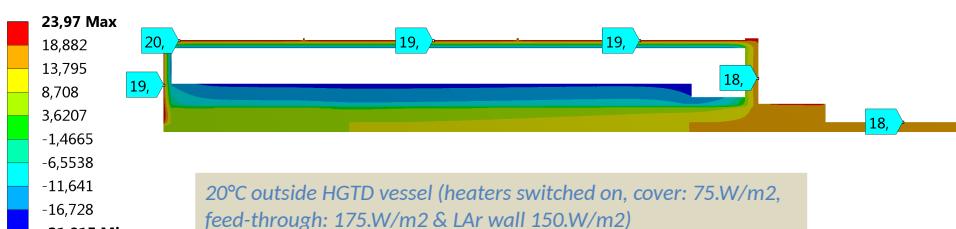
1889 The primary role of the HGTD global structure is to provide a robust support to all active silicon layers
 1890 and peripheral on-detector electronics in a cold and dry volume, precisely located on the front wall of
 1891 the liquid argon endcap calorimeters. The chosen material should withstand the expected radiation levels
 1892 and a wide temperature range as a safety requirement: from -55°C to 60°C . Thin large panels made of
 1893 carbon fibre composite fit these requirements and provide a stiff support.

1894 One of the main challenges is to preserve the detector volume dry. This can be achieved with a permanent
 1895 dry N_2 flow (given a dew point of -60°C). The flow will renew the air within the volume a few times per
 1896 hour, with an over pressure of +5 mbar.

1897 The second main challenge is to keep the temperature of the outer skins safely above the cavern condensa-
 1898 tion temperature ($\sim 17^{\circ}\text{C}$). In a way similar to what is done on the endcap calorimeters cover, heaters will
 1899 be placed on the external face of the front cover of the HGTD vessel and on the outer ring, between the
 1900 service feedthroughs. Their purpose is to ensure a temperature around 20°C outside the HGTD vessel,
 1901 in order to avoid the appearance of condensation. The face of the endcap calorimeter cryostat is already
 1902 equipped with the necessary heaters. The expected power consumption of the front cover heaters is
 1903 75 W/m^2 , and 175 W/m^2 for the outer ring. Figure 61 shows a calculation of the temperature distribution
 1904 of the HGTD-vessel components and the moderator, considering that the inner volume (in white, not
 1905 simulated) will contain the sensors at -30°C . The difference in the external temperature achieved by
 1906 having the heaters on can be appreciated by comparing Figures 61(a) and 61(b).



(a) With heaters on front cover *off*.



(b) With heaters on front cover *on*.

Figure 61: Simulation based on the FEA model of the temperature distribution of the HGTD vessel front cover and moderator. The view shows a transverse cut of the vessel, with $R = 0$ towards the left and the outer radius on the right. Only half the vessel is shown.

1907 The bolting and locking interface with the endcap cryostat will use the same MBTS threaded spots and
 1908 the central large hole of the warm tube taking into account the presence of moderator rings. The access
 1909 to the holes should be possible without opening the HGTD cold vessel and should not affect its cold and

1910 dry volume. Possible conflicts with the cooling pipes of these moderator rings require further studies and
 1911 optimisation. Since the expected performance of the proposed HGTD cold vessel is similar to the one
 1912 of the ATLAS ITk global support, a collaboration program with the ITk team is underway in order to
 1913 combine the R&D carbon fibre composite material studies, such as the selection of K13C2U (single skin)
 1914 as material in an EX1515 matrix.

1915 **5.3 Moderator**

1916 A moderator disk made of borated polyethylene, with a density of 0.95 kg/l, will be placed in front of the
 1917 endcap calorimeters on each side to shield the ITk and HGTD from back-scattered neutrons. The material
 1918 is the same as the one used at present in the ATLAS detector. As can be appreciated in Figure 55, the
 1919 moderator is divided into two disks, one inside and one outside the HGTD vessel. Due to the very limited
 1920 width in z between ITk and the LAr cryostat wall, and the fact that the latter is uneven, it is necessary for
 1921 the moderator to be part of the HGTD assembly design. The moderator on the outside will be directly
 1922 screwed to the LAr cryostat wall, and provide the necessary flat surface on which to install the HGTD and
 1923 accessible bolting brackets. The part of the moderator inside the vessel gives an appropriate support for
 1924 the instrumented layers, and because it goes up to $R = 900$ mm it leaves enough free space for the cooling
 1925 services as shown in Figure 62.

1926 The outside moderator disk has a thickness that varies along the radius: 10 mm covering the LAr
 1927 calorimeter cryostat central flange and the bolting spots ($110 \text{ mm} < R < 342 \text{ mm}$) and 20 mm elsewhere
 1928 ($342 \text{ mm} < R < 1100 \text{ mm}$); the weight of this disk will approximately 70 kg. It will be directly screwed to
 1929 the endcap calorimeter cryostat and mechanically separated from the HGTD hermetic volume. The second
 1930 moderator disk, inside the HGTD vessel, will have a thickness of 30 mm in the range of $120 \text{ mm} < R <$
 1931 900 mm and weigh 70 kg per endcap. The total moderator thickness in z will be 50 mm except at the inner
 1932 and outermost radii; it is 40 mm in the region between 110–342 mm and 20 mm for $R > 900 \text{ mm}$. During
 1933 the maintenance (and when the expected replacement of the inner modules takes place at the surface), the
 1934 two moderator disks will stay bolted in the LAr cryostat, together with the rear vessel cover. Consequently
 1935 they will be made of unique circular disks to improve thermal insulation and mechanics stiffness.

1936 **5.4 Services**

1937 The HGTD data and powering services can be grouped into five categories: optical fibres for data
 1938 transmission, bias voltage for the sensors (high voltage, or HV), power for the module electronics (low
 1939 voltage, or LV), DCS links and powered components in the proximity (front-cover heaters and temperature
 1940 sensors). A preliminary estimate of the required services per endcap (excluding the cooling lines) is
 1941 summarised below.

- 1942 • The number of optical links per endcap amount to 1640 and is composed of 600 up-links for the
 1943 offline readout, 600 down-links for configuration and fast signals (clocks, L0/L1, etc) and 440
 1944 up-links for the luminosity readout. Optimising the number of input e-links per lpGBT for the
 1945 luminosity could reduce this number by 10 %. The optical fibres are grouped in ribbons of 12
 1946 fibres (about $4 \times 2 \text{ mm}^2$ cross section). In addition, these ribbons need to be encapsulated in cables
 1947 in order to be routed in the flexible chains. The cables will contain six ribbons and have an outer
 1948 diameter of 12 mm. Taking this grouping into account gives a total of 144 ribbons per endcap,
 1949 including 88 spare fibres.

- 1950 • The baseline for the bias-voltage distribution is to bring one individual line per module to the HGTD
 1951 and to use some distribution fan-out boards in USA15 to group modules at the same radius and
 1952 therefore require the same voltage. Consequently 3944 lines are needed, plus 288 return lines. The
 1953 bias lines will be grouped into about 144 cables (outer diameter of 14 mm).
- 1954 • The baseline for the power distribution is to use a design quite similar to the one used in the ITk-
 1955 strip [4]: a 48 V voltage power supply in USA15, commercial DC/DC converters (48 V to 12 V)
 1956 at the patch panel (PP-EC) which is proposed to be located at the back of the endcap cryostat, and
 1957 radiation hard DC/DC converters (11 V to 1.2 or 2.5 V) on the peripheral on-detector electronics.
 1958 The number of power cables (outer diameter of 14.0 mm) needed per endcap is approximately 16
 1959 from USA15 to PP-EC and 32 cables from PP-EC to each side of the HGTD.
- 1960 • The number of DCS cables is still being optimised to fit the maximum size of ITk service corridors
 1961 and the capacity of the flexible chains. Current estimates are around 2008 enable-status lines and
 1962 288 lines for grounding. Studies are ongoing to determine if all DCS connections need to be inside
 1963 the flexible chains.
- 1964 • Around 40 heaters are required per side (10 in each quadrant) in order to provide uniform heating.
 1965 One PT100 gauge is used to monitor each heater, as is currently done for the heaters located on the
 1966 LAr cryostat.

1967 This information is summarised in Table 13, indicating the needed services. These numbers include spares
 1968 to be installed in case of damage during installation or operation. The spares will not be connected unless
 1969 they are needed.

	Optical fibres	HV bias	LV power	DCS		Heaters	
				Power	Probe	Power	PT100
Cable diameter [mm]	12.0	12.4	10.8	12.3	12.3	11.2	11.2
Number of wires/cable	72 fibres	37	2 conductors	34	57	20	20
Number of cables/quadrant	6	30	8	32	8	1	2
Number of spares/endcap	4 × 22 fibres	4 × 52	0	0	0	0	0
Total cables/endcap	24	120	32	128	32	4	8

Table 13: Summary of needed services (excluding cooling lines) per endcap of the HGTD. A *quadrant* is one quarter of one endcap. The DCS column includes the power management and probe monitoring.

1970 The readout links and DCS cables will be inserted in a flexible cable chain that goes from USA15 to the
 1971 HGTD. It should run along the extended barrel region and allow monitoring of the HGTD even while
 1972 opening and closing ATLAS. The bias-voltage and low-voltage services could be made of two cable types
 1973 with a patch panel located at the back of the extended Tile barrel (as seen in figure 58 and 59). These
 1974 cables will be disconnected when moving the endcap.

1975 The baseline option is to extract the signals to the outside of the vessel and to route the needed cables/fibres
 1976 to the detector using dedicated feedthroughs in the outer face of the vessel (see Figure 62). Custom-
 1977 designed feedthroughs will be used, moulded from an elastomer soft flange with tight passage for electrical
 1978 and cooling services. The services would then be routed to the outside of ATLAS in specific regions in
 1979 ϕ , as shown in Figure 63. This figure shows a possible routing of the services, together with the nominal
 1980 numbers of cables. A small quantity of spares, shown in Table 13, will be added. Studies are ongoing to
 1981 optimise the accommodation of both HGTD and ITk services in the region around $1000 < R < 1100$ mm.
 1982 The feedthrough studies for the TDR will be focused on a design for low heat loss using thermal shield

1983 materials and local heaters similar to what is planned for the front cover. In addition, the exact geometry
 1984 and disposition of the detector components at $R > 900$ mm may still be subject to optimisation.

Not reviewed, for internal circulation only

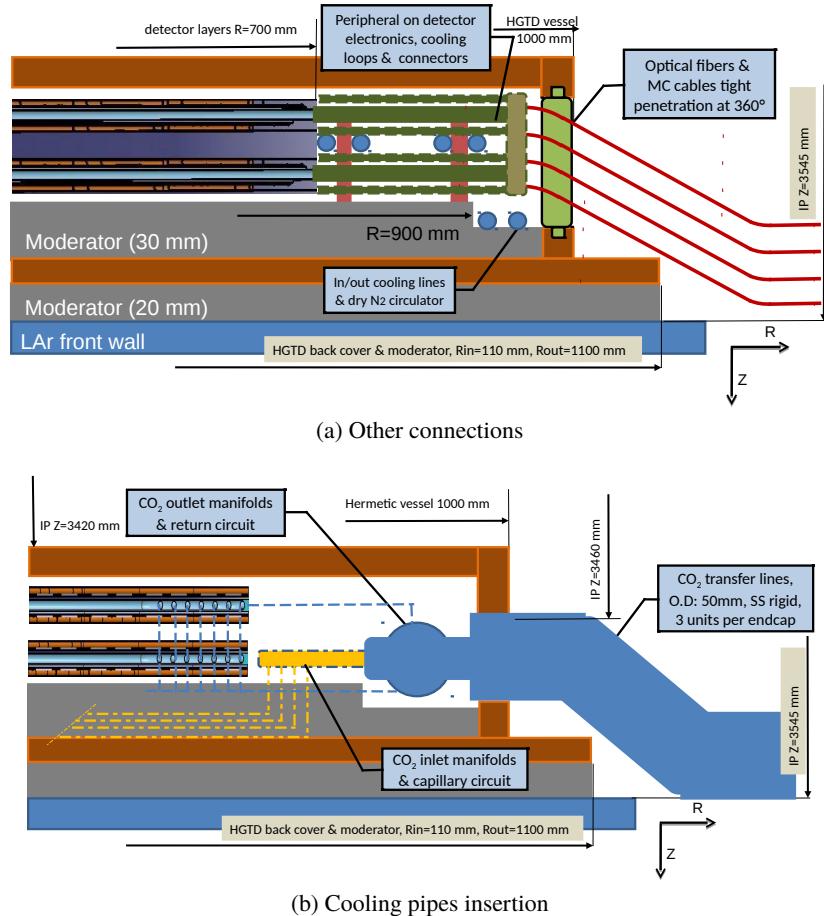


Figure 62: Transverse view of the outer radius of the vessel. The peripheral on-detector electronics boards can be seen in (a) connected to the services via feed-through connections. In (b) a view of the position of insertion for the cooling pipes is presented.

1985 5.5 Outlook and path towards TDR

1986 Three major activities are foreseen up to the TDR;

- 1987 • A thermal demonstrator will be constructed using a stave with fake sensors and ASIC to establish
 1988 the cooling performance of the proposed design. It should be available in autumn 2018. Relevant
 1989 parameters will be measured using temperature sensors and thermal imaging cameras, and varying
 1990 the heat dissipation, the CO_2 cooling parameters such as pressure and flow, and the flow for the dry
 1991 nitrogen blowing. A stave with dummy modules and fake peripheral on-detector electronics will be
 1992 used to study the power dissipation in a realistic way and allow comparison with thermal simulations.
 1993 These tests should also validate the part of the cooling circuit dedicated to the peripheral on-detector
 1994 electronics.

Not reviewed, for internal circulation only

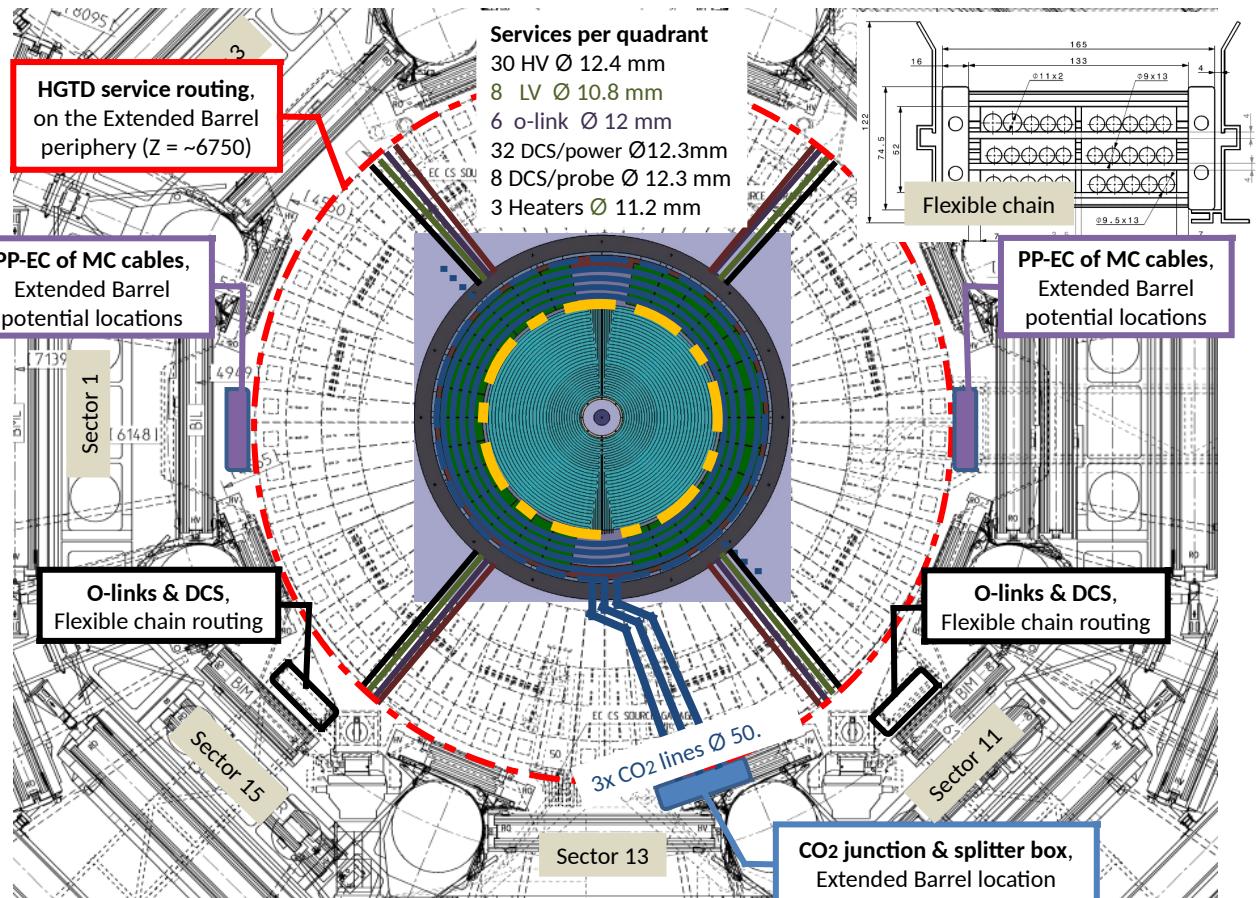


Figure 63: Illustrations of the possible path of the HGTD services (red), the cooling transfer lines (blue), multiconductor cables (MC cables, violet) and flexible chain (black) that will contain the DCS links and optical links (*o-links*).

- 1995 • A simulation implementing a realistic model of the detector components needs to be performed to better estimate the cooling requirements. A study of thermal runaway will also be performed.
 1996
 1997 • A conceptual design of the feedthrough should be finalised by Q4 2018-Q1 2019.
 1998 • The services routing, compatible both with the HGTD and ITk needs, should be determined and implemented in a full 3D model of the ATLAS cavern. In particular, it needs to include the definition of the position of the patch panel needed to connect and disconnect the services when the detector is open. This work should be in an advanced state in Q1 2019.
 1999
 2000
 2001
 2002 In addition, a mechanics mock-up with a real-size cooling plate, stacking of cables and dummy modules of realistic dimensions will be made available during the first half of 2019 to confirm the overall detector envelope of 75 mm.
 2003
 2004

2005 6 Organisation, cost and schedule

2006 The HGTD group started as an organised ATLAS upgrade activity in summer 2015 and this new sub-
 2007 detector proposal was part of the ATLAS Upgrade Scoping Document. The HGTD activity is taking place
 2008 in close collaboration with the ITk and Technical coordination since the space allocated to the HGTD is
 2009 very tight and may have interference with the ITk volume and the routing of its services. HGTD covers
 2010 various activities:

- 2011 • Sensors
- 2012 • Electronics
- 2013 • Luminosity/trigger
- 2014 • Test beam
- 2015 • Module assembly
- 2016 • Detector assembly and integration
- 2017 • Simulation and physics performance

2018 The work is carried out by roughly 120 physicists and engineers from 20 ATLAS Institutes from eleven
 2019 funding agencies, see Table 14. Table 15 summarises the involvement of the institutes in the different
 2020 activities⁹.

Country (funding agency)	Institutes/Universities
CERN	
France	LAL (Orsay), LPNHE (Paris) , OMEGA (Palaiseau)
Germany	JGU (Mainz), JLU (Giessen)
Slovenia	IJS (Ljubljana)
Spain	IFAE (Barcelona)
Sweden	KTH (Stockholm)
Taiwan	AS (Taipei), National Tsing-Hua U
USA	BNL, Ohio State U, SLAC, SMU (Dallas), Stony Brook NY, UC Santa Cruz, U of Iowa
Russia	JINR
Morocco	Univ. Hassan II Casa Blanca

Table 14: Countries, funding agencies and institutes/universities participating in the HGTD project. USA groups approved only for R&D.

2021 If approved by LHCC, the R&D in all different areas of the project will need to be intensified, in order
 2022 to prepare for a Technical Design Report (TDR) in the first quarter of 2019. Dedicated working groups
 2023 should be reinforced and adapted to the new needs of the project. The details of responsibility and sharing
 2024 need to be defined in the MoU after the TDR. A preliminary survey of institute interests and of the
 2025 available person-power have been performed and indicate that the (quite substantial) resources required
 2026 may be made available.

⁹ Groups from the USA are currently only approved for R&D

Activities	Institutes
Sensors	BNL, CERN, Dubna, IFAE, JSI, UCSC
Electronics	AS, Tsing-Hua, CERN, Dubna, Giessen, IFAE, Iowa, KTH, LAL, Omega, SLAC, SMU, Stony Brook
Luminosity/trigger	KTH, Ohio State
Test beam	All institutes
Module assembly	CERN, BNL, Dubna, IFAE, Iowa, JSI, LAL, LPNHE, Mainz, Ohio State
Mechanics/Integration	CERN, Dubna, LAL, LPNHE
Software & Performance	Casa Blanca, CERN, Giessen, IFAE, Iowa, KTH, LAL, LPNHE, SLAC

Table 15: Present and expected involvement of the institutes in the different HGTD activities.

2027 A preliminary estimate of the core cost¹⁰ of the HGTD is presented in Table 16. No core costs needed for
 2028 the R&D period (mainly sensors and electronics) will be identified for the TDR with a proposed sharing
 2029 between the HGTD institutes.

Item	Cost (kCHF)
Sensors	1700
Front-end ASICs	730
Bump bonding	900
Module assembly	600
Peripheral on-detector electronics (transition, optical and HV boards, optical links, services)	717
Power supplies and electronics in USA15	2027
Mechanics and integration (cooling support plates, vessel, feedthrough)	405
CO ₂ cooling plant and distribution	450
Sub-total HGTD	7529
Detector readout, dataflow, and network	970
Total (kCHF)	8499

Table 16: Core cost of the HGTD in kCHF.

2030 The overall ATLAS installation schedule for Long Shutdown 3 provides constraints for the scheduling
 2031 of the installation of the HGTD since it needs to be integrated into the global ATLAS Phase-II detector
 2032 upgrade plan. The installation of the HGTD is planned to take up to one month for each side and should
 2033 take place in June and July 2025, for the A and C endcap side, respectively. While all the mechanical
 2034 supports, services and the cold vessel need to be installed during these two periods, the detector layers
 2035 are designed to allow insertion during a winter shutdown in case of delays. This schedule maybe still be
 2036 subject to modification if required by the installation of other ATLAS detector components.

2037 Figure 64 and Table 17 give an overview of the schedule foreseen for R&D, prototyping, production and
 2038 installation of the HGTD. There will be three main phases:

- 2039 • **2018-2020** R&D
- 2040 • **2021-2024** Construction
- 2041 • **2025-2026** Integration, installation and commissioning

¹⁰ The item “Detector readout, dataflow, and network” is considered a TDAQ deliverable but cannot be included in the TDAQ TDR until the HGTD TDR has been reviewed by LHCC/UFG.

- 2042 The main milestones have been identified during the R&D period and are listed in Table 18. The dates
 2043 for the post-TDR milestones are preliminary and will be reassessed when preparing the TDR.

Not reviewed, for internal circulation only

Main activity	Activity steps	Start	End	Work days
1. Services (cooling, cables, fibres, . . .)	R&D/prototypes	01/01/18	30/06/20	651
	pre-production	01/09/20	30/09/21	281
	production	01/04/22	30/10/23	412
	installation	30/01/24	30/04/25	326
2. BE Electronics in USA15 (LV, HV, Luminosity board, . . .)	R&D/prototypes	01/01/18	30/07/22	1194
	production/procurement	30/10/22	30/04/24	391
	Installation in USA15	01/07/24	30/06/25	260
3. Mechanics (cooling plates, vessel, . . .)	R&D	01/01/18	30/03/20	585
	engineering	01/04/20	30/12/21	456
	production	01/04/22	30/04/23	281
4. Peripheral on-detector electronics (Flex, Transition, optical and HV boards, DCS, . . .)	R&D	01/01/18	30/12/20	781
5. On detector electronics (ASIC)	pre-production	01/03/21	30/03/22	281
	production	01/04/22	30/08/23	369
6. Sensors (LGAD)	R&D	01/01/18	30/12/20	781
	pre-production	01/01/21	30/12/21	259
	production	01/01/22	30/06/23	389
7. Modules assembly and loading in staves	R&D	01/01/18	30/12/20	781
	Module0 & Stave0	01/01/21	30/04/22	346
	production	01/05/22	29/02/24	478
8. Detector assembly (Layers mounting in vessel, off-detector installation, . . .)	HGTD-A	01/10/23	30/06/24	195
	HGTD-C	01/08/24	30/04/25	194
10. Installation in ATLAS and commissioning	Installation HGTD-A	02/06/25	02/07/25	21
	Installation HGTD-C	02/07/25	01/08/25	21
	Commissioning HGTD-A	03/07/25	01/02/26	152
	Commissioning HGTD-C	03/08/25	01/02/26	130

Table 17: Schedule of the HGTD project. The number of work days per week is assumed to be five. The dates for the post-TDR milestones are preliminary and will be reassessed when preparing the TDR.

Not reviewed, for internal circulation only

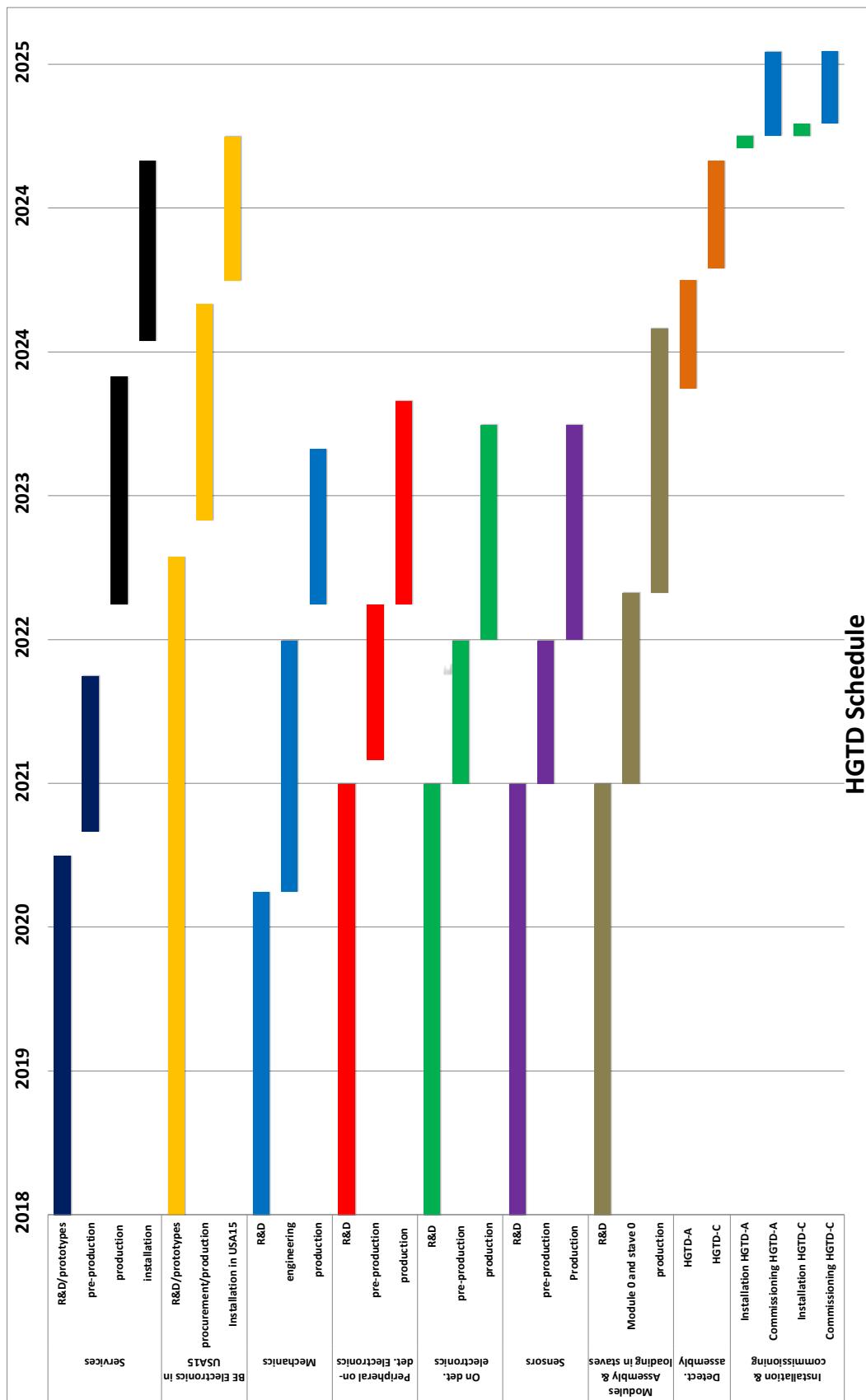


Figure 64: Gantt chart showing the schedule of the various activities in the HGTD project.

Not reviewed, for internal circulation only

Name	Area	Time	Comment
M1	Modules	Q3 2018	Performance of flex circuits
M2	Software	Q3 2018	Detailed geometry of HGTD in ATLAS software
M3	Services	Q3 2018	Performance of CO ₂ cooling system with prototype and dummy stave
M4	Sensors	Q4 2018	Performance of large-size sensors ($2 \times 2 \text{ cm}^2$) of HPK/CNM radiation tolerance performance
M5	Electronics	Q4 2018	Performance of one pixel readout channel (ALTIROC1, 25 channels)
M6	Services	Q4 2018	Service routing conceptual design
M7	Electronics	Q1 2019	Design review of final ALTIROC chip (225 channels)
M8		Q1 2019	Technical Design Report
M9	Electronics	Q2 2019	Submission of first iteration of ALTIROC
M10	Services	Q2 2019	Feedthrough/services connectivity concept and prototype (PDR)
M11	Electronics	Q2 2019	Peripheral on-detector electronics concept and prototype (PDR)
M12	Services	Q3 2019	Design tooling for detector assembly and for final installation
M13	Sensors	Q3 2019	Final sensor specification and vendor qualification (FDR)
M14	Electronics	Q4 2019	Characterisation of first ALTIROC iteration
M15	Modules	Q1 2020	Prototype module assembly (2×2 or $2 \times 4 \text{ cm}^2$)
M16	Electronics	Q2 2020	Submission of ALTIROC second iteration (FDR)
M17	Electronics	Q4 2020	Final validation of ALTIROC ASIC (PRR)
M18	Modules	Q4 2020	Stave 0 demonstrator including peripheral electronics and services

Table 18: List of main milestones during the R&D phase. The acronyms are Preliminary Design Review (PDR), Final Design Review (FDR), and Production Readiness Review (PRR).

2044 7 Outlook

2045 This document describes a technical proposal for a High-Granularity Timing Detector (HGTD), to be
 2046 installed in ATLAS during the long shutdown, which starts at the end of 2023 and is expected to last
 2047 for 30 months, until the start of HL-LHC data-taking in mid-2026. This proposal is the result of two
 2048 years of active R&D, especially on sensors and front-end electronics, performed since summer 2015 by
 2049 ~20 institutes and ~120 collaborators. The proposal for the HGTD was already included in the ATLAS
 2050 Phase-II Scoping Document [1].

2051 The HGTD will provide a timing resolution of ~30 ps for minimum-ionising particles throughout the
 2052 entire duration of the HL-LHC programme, covering the pseudorapidity region between 2.4 and 4.0.
 2053 The detector will use the space currently occupied by the so-called Minimum-Bias Trigger Scintillators
 2054 installed in front of the LAr endcap cryostats on both A and C sides at $\pm 3.5 \text{ m}$ from the interaction
 2055 point. Silicon Low Gain Avalanche Detectors (LGADs) will be used to provide the precision timing
 2056 measurements. The R&D on the sensor technology and the front-end electronics have demonstrated that
 2057 this target time resolution is already close to being achieved. A realistic baseline design has been described.
 2058 An optimisation towards a baseline concept driven by the best compromise between performance and cost
 2059 is ongoing.

2060 With 30 ps time resolution for charged particles, the HGTD can mitigate the impact of pileup in the
 2061 association of tracks in the forward region to their respective vertices, compensating for the reduced

2062 longitudinal impact parameter resolution of tracks reconstructed by the ITk tracking detector at large
 2063 pseudorapidities. With this capability, the performance of pileup-jet suppression, b -jet tagging, and
 2064 lepton isolation in the forward region approaches the same level as in the central region. The rejection
 2065 of pileup jets improves by a factor of approximately 2 (for a hard-scatter jet efficiency of 98%), the light-
 2066 jet rejection at a b -jet efficiency of 70% improves by a factor of 1.5, and the lepton isolation efficiency
 2067 increases by 14%. Several important physics analyses rely on the performance of such objects in the forward
 2068 region. This document demonstrates that the HGTD can improve the uncertainty on the signal strength
 2069 determination for VBF-produced Higgs bosons decaying to $H \rightarrow WW^*$ by 8% and the signal significance
 2070 for the $tH(H \rightarrow b\bar{b})$ process by 11%. For the measurement of the weak mixing angle, $\sin^2 \theta_{\text{eff}}$, an
 2071 improvement of 11% on the experimental uncertainty is achieved through the improved electron isolation
 2072 performance in the forward region that the HGTD brings. Many other physics applications are being
 2073 considered, including precision measurements of the searches for long-lived particles, machine induced
 2074 non-collision background processes such as beam-gas interactions and beam-halo activity, etc.

2075 Furthermore, the HGTD provides unique capabilities to measure the online and offline luminosity with
 2076 high accuracy. It can provide a minimum-bias trigger at Level-0 and possibilities for improved pileup
 2077 mitigation in both the Level-0 and the high-level trigger systems. The improved trigger capability as
 2078 well as the impact on other physics analyses will be evaluated in greater detail in the Technical Design
 2079 Report.

Not reviewed, for internal circulation only

2080 References

- 2081 [1] ‘ATLAS Phase-II Upgrade Scoping Document’,
 2082 tech. rep. CERN-LHCC-2015-020. LHCC-G-166, CERN, 2015,
 2083 URL: <https://cds.cern.ch/record/2055248>.
- 2084 [2] RD50 collaboration, URL: <https://rd50.web.cern.ch/rd50>.
- 2085 [3] G. Pellegrini et al., *Technology developments and first measurements of Low Gain Avalanche*
 2086 *Detectors (LGAD) for high energy physics applications*, Nucl. Instrum. Meth. **A765** (2014) 12 .
- 2087 [4] ATLAS Collaboration, ‘Technical Design Report for the ATLAS Inner Tracker Strip Detector’,
 2088 tech. rep. CERN-LHCC-2017-005 ; ATLAS-TDR-025, CERN, 2017,
 2089 URL: <https://cds.cern.ch/record/2257755>.
- 2090 [5] ATLAS Collaboration,
Expected Performance of the ATLAS Inner Tracker at the High-Luminosity LHC,
 2091 ATL-PHYS-PUB-2016-025, 2016, URL: <https://cds.cern.ch/record/2222304>.
- 2092 [6] ATLAS Collaboration, *Tagging and suppression of pileup jets with the ATLAS detector*,
 2093 ATLAS-CONF-2014-018, 2014, URL: <https://cds.cern.ch/record/1700870>.
- 2094 [7] M. Aaboud et al.,
Luminosity determination in pp collisions at $\sqrt{s} = 8$ TeV using the ATLAS detector at the LHC,
 2095 Eur. Phys. J. **C76** (2016) 653, arXiv: [1608.03953 \[hep-ex\]](https://arxiv.org/abs/1608.03953).
- 2096 [8] C. Collaboration, *CMS Luminosity Measurements for the 2016 Data Taking Period*, (2017).
- 2097 [9] ATLAS Collaboration, *Performance of pile-up mitigation techniques for jets in pp collisions at*
 2098 *$\sqrt{s} = 8$ TeV using the ATLAS detector*, Eur. Phys. J. **C76** (2016) 581,
 2099 arXiv: [1510.03823 \[hep-ex\]](https://arxiv.org/abs/1510.03823).

- 2102 [10] ATLAS Collaboration,
 2103 *Observation and measurement of Higgs boson decays to WW^* with the ATLAS detector,*
 2104 *Phys. Rev. D* **92** (2015) 012006, arXiv: [1412.2641 \[hep-ex\]](https://arxiv.org/abs/1412.2641).
- 2105 [11] ATLAS Collaboration, *Measurement Prospects for $VBF H \rightarrow WW^* \rightarrow e\nu\mu\nu$ production with*
 2106 *$3 ab^{-1}$ of HL-LHC pp-collisions*, ATL-PHYS-PUB-2016-018, 2016,
 2107 URL: <https://cds.cern.ch/record/2209092>.
- 2108 [12] ATLAS Collaboration, *Measurements of the Higgs boson production cross section via Vector*
 2109 *Boson Fusion and associated WH production in the $WW^* \rightarrow \ell\nu\ell\nu$ decay mode with the ATLAS*
 2110 *detector at $\sqrt{s} = 13$ TeV*, ATLAS-CONF-2016-112, 2016,
 2111 URL: <https://cds.cern.ch/record/2231811>.
- 2112 [13] ALEPH Collaboration, DELPHI Collaboration, L3 Collaboration, OPAL Collaboration, SLD
 2113 Collaboration, LEP Electroweak Working Group, SLD Electroweak Group, SLD Heavy Flavour
 2114 Group, *Precision electroweak measurements on the Z resonance*, *Phys. Rept.* **427** (2006) 257,
 2115 arXiv: [hep-ex/0509008 \[hep-ex\]](https://arxiv.org/abs/hep-ex/0509008).
- 2116 [14] J. C. Collins and D. E. Soper, *Angular distribution of dileptons in high-energy hadron collisions*,
 2117 *Phys. Rev. D* **16** (1977) 2219.
- 2118 [15] M. Aaboud et al., *Precision measurement and interpretation of inclusive W^+ , W^- and Z/γ^**
 2119 *production cross sections with the ATLAS detector*, *Eur. Phys. J. C* **77** (2017) 367,
 2120 arXiv: [1612.03016 \[hep-ex\]](https://arxiv.org/abs/1612.03016).
- 2121 [16] G. Aad et al.,
 2122 *Electron and photon energy calibration with the ATLAS detector using LHC Run 1 data*,
 2123 *Eur. Phys. J. C* **74** (2014) 3071, arXiv: [1407.5063 \[hep-ex\]](https://arxiv.org/abs/1407.5063).
- 2124 [17] G. Aad et al., *Measurement of the Inelastic Proton-Proton Cross-Section at $\sqrt{s} = 7$ TeV with the*
 2125 *ATLAS Detector*, *Nature Commun.* **2** (2011) 463, arXiv: [1104.0326 \[hep-ex\]](https://arxiv.org/abs/1104.0326).
- 2126 [18] M. Aaboud et al., *Measurement of the Inelastic Proton-Proton Cross Section at $\sqrt{s} = 13??$ TeV*
 2127 *with the ATLAS Detector at the LHC*, *Phys. Rev. Lett.* **117** (2016) 182002,
 2128 arXiv: [1606.02625 \[hep-ex\]](https://arxiv.org/abs/1606.02625).
- 2129 [19] H. Sadrozinski, A. Seiden and N. Cartiglia,
 2130 *4-Dimensional Tracking with Ultra-Fast Silicon Detectors*, Reports on Progress in Physics (2017),
 2131 arXiv: [1704.08666](https://arxiv.org/abs/1704.08666), URL: <http://iopscience.iop.org/10.1088/1361-6633/aa94d3>.
- 2132 [20] G.-F. Dalla Betta et al.,
 2133 *Design and TCAD simulation of double-sided pixelated low gain avalanche detectors*,
 2134 *Nucl. Instrum. Meth. A* **796** (2015) 54.
- 2135 [21] G. Pellegrini et al., ‘Status of LGAD production at CNM’, 30th RD50 Workshop, Krakow, Poland,
 2136 2017, URL: https://indico.cern.ch/event/637212/contributions/2608652/attachments/1470919/2276240/pellegrini_rd50.pdf.
- 2138 [22] G. Kramberger et al.,
 2139 *Radiation effects in Low Gain Avalanche Detectors after hadron irradiations*,
 2140 *JINST* **10** (2015) P07006.
- 2141 [23] G. Kramberger et al., *Radiation hardness of thin Low Gain Avalanche Detectors*,
 2142 submitted to NIM A (2017), arXiv: [1711.06003](https://arxiv.org/abs/1711.06003).
- 2143 [24] N. Cartiglia et al., *Beam test results of a 16 ps timing system based on ultra-fast silicon detectors*,
 2144 *Nucl. Instrum. Meth. A* **850** (2017) 83, arXiv: [1608.08681 \[physics.ins-det\]](https://arxiv.org/abs/1608.08681).

- 2145 [25] J. Lange et al., *Gain and time resolution of 45 μm thin Low Gain Avalanche Detectors before and*
 2146 *after irradiation up to a fluence of $10^{15} \text{n}_{eq}/\text{cm}^2$* , JINST **12** (2017) P05003,
 2147 arXiv: [1703.09004](https://arxiv.org/abs/1703.09004) [[physics.ins-det](#)].
- 2148 [26] Z. Galloway et al., *Properties of HPK UFSD after neutron irradiation up to $6e15 \text{n}/\text{cm}^2$* ,
 2149 submitted to NIM A (2017), arXiv: [1707.04961](https://arxiv.org/abs/1707.04961).
- 2150 [27] L. Masetti et al., *Beam test measurements of Low Gain Avalanche Detector single pads and arrays*
 2151 *for the ATLAS High Granularity Timing Detector*, to be submitted to JINST (2018),
 2152 arXiv: [1804.00622](https://arxiv.org/abs/1804.00622) [[physics.ins-det](#)].
- 2153 [28] I. Mandić, V. Cindro, G. Kramberger and M. Mikuž, *Measurement of anomalously high charge*
 2154 *collection efficiency in n^+p strip detectors irradiated by up to $10^{16} \text{n}_{eq}/\text{cm}^2$* ,
 2155 Nucl. Instrum. Meth. **A603** (2009) 263.
- 2156 [29] J. Lange, J. Becker, E. Fretwurst, R. Klanner and G. Lindström,
 2157 *Properties of a radiation-induced charge multiplication region in epitaxial silicon diodes*,
 2158 Nucl. Instrum. Meth. **A622** (2010) 49.
- 2159 [30] G. Casse, A. Affolder, P. Allport, H. Brown and M. Wormald, *Enhanced efficiency of segmented*
 2160 *silicon detectors of different thicknesses after proton irradiations up to $1 \times 10^{16} \text{n}_{eq}/\text{cm}^2$* ,
 2161 Nucl. Instrum. Meth. **A624** (2010) 401.
- 2162 [31] N. Cartiglia et al., ‘Preliminary results from UFSD beam test at FNAL’,
 2163 30th RD50 Workshop, Krakow, Poland, 2017,
 2164 URL: https://indico.cern.ch/event/637212/contributions/2608659/attachments/1471224/2276633/Cartiglia_BeamTest_FNAL.pdf.
- 2166 [32] Y. Zhao et al.,
 2167 *Comparison of 35 and 50 μm thin HPK UFSD after neutron irradiation up to $6 \times 10^{15} \text{n}_{eq}/\text{cm}^2$* ,
 2168 (2018), arXiv: [1803.02690](https://arxiv.org/abs/1803.02690).
- 2169 [33] A. Collaboration,
 2170 ‘ATLAS Trigger and Data Acquisition Phase-II Upgrade Technical Design Report’,
 2171 tech. rep. ATL-COM-DAQ-2017-185,
 2172 This version represents the draft for EB approval and LHCC release.: CERN, 2017,
 2173 URL: <https://cds.cern.ch/record/2296879>.
- 2174 [34] J. Anderson et al., *FELIX: a PCIe based high-throughput approach for interfacing front-end and*
 2175 *trigger electronics in the ATLAS Upgrade framework*,
 2176 Journal of Instrumentation **11** (2016) C12023.
- 2177 [35] F. Faccio et al.,
 2178 ‘FEAST2: A Radiation and Magnetic Field Tolerant Point-of-Load Buck DC/DC Converter’,
 2179 *2014 IEEE Radiation Effects Data Workshop (REDW)*, IEEE, 2014.
- 2180 [36] L Adamczyk et al., ‘Technical Design Report for the ATLAS Forward Proton Detector’,
 2181 tech. rep. CERN-LHCC-2015-009. ATLAS-TDR-024, 2015,
 2182 URL: <https://cds.cern.ch/record/2017378>.