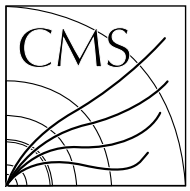


CERN-LHCC-2019-003
CMS-TDR-020
29 March 2019
Revised 26 September 2019

A MIP Timing Detector for the CMS Phase-2 Upgrade

Technical Design Report

CMS Collaboration

CERN-LHCC-2019-003 / CMS-TDR-020
29/10/2019

Editors

J. Butler, T. Tabarelli de Fatis

Chapter Editors

A. Apresyan, J. Bendavid, A. Bornheim, J. Butler, N. Cartiglia, F. Golf, L. Gray, W. Li, T. Liu, M.T. Lucchini, S. Lusin, W. Lustermann, P. Martinez Ruiz del Arbol, P. Meridiani, I. Ojalvo, O. Sahin, D. Stuart, T. Tabarelli de Fatis, S. Tkaczyk, J. Varela

Cover Design

S. Cittolin

Acknowledgements

This document would not exist without the devoted efforts of many colleagues, too many to be named explicitly, who contributed text and figures to this TDR. We would like to thank the technical staffs from the various MTD institutions for their excellent work during the design and R&D phase of this upgrade, without which this TDR and the MTD itself would not be possible.

The very useful input from the MTD collaboration, the CMS internal reviewers (Wolfgang Adam, Johan Borg, Tulika Bose, Didier Contardo, Karl Gill, Frank Hartmann, Jan Kieseler, Katja Klein, Frans Meijers, Michael Moll, Paolo Rumerio, Alexander Savin, Jeff Spalding, Roberto Tenchini, Wolfram Zeuner), and the chair of the Phase-2 TDRs editorial board (C. Lourenço) helped to improve the quality of this document and is greatly appreciated. We also thank the individual and institutional reviewers who participated in the Collaboration-Wide Review of the TDR.

We would like to acknowledge the support of the CMS Management, the Upgrade Project Coordination team as well as the many contributions from CMS Technical Coordination. We thank the Offline, Computing, Physics Performance and Datasets, Upgrade Physics Strategy Group, and all the Physics groups for their help in developing and executing the physics and performance studies. The CMS Tracker and Endcap Calorimeter Upgrade projects provided crucial support and advice on various inter-detector coordination issues.

Contents

Chapter 1

Overview of the MIP Timing Detector Project

1.1 Introduction

This document is the Technical Design Report (TDR) of the MIP Timing Detector (MTD), a new detector planned for CMS during the High Luminosity LHC (HL-LHC) era. This device will bring a completely new capability to CMS — the ability to measure precisely the production time of minimum ionizing particles (MIP) for use in disentangling the approximately 200 nearly-simultaneous "pileup" interactions that will occur in each bunch crossing of the LHC. It also provides new capabilities for charged hadron identification and the search for long-lived particles.

In this introduction, we summarize the motivation and physics case for the MTD, present the high-level considerations and requirements for the design, and give an overview of the proposed detector.

In 2015, the LHC achieved a center-of-mass energy for proton-proton collisions of 13 TeV. While a small increase in collision energy is planned, the main improvement in the sensitivity of the search for physics "beyond the standard model (BSM)" will come from increased luminosity. In 2019, the LHC began a two-year shutdown, known as Long Shutdown 2 (LS2), to upgrade the injector complex to produce brighter (smaller emittance) beams. After a three-year running period at 13–14 TeV (Run-3), there will be another long shutdown of approximately 2.5 years, Long Shutdown 3 (LS3), starting in 2024, to upgrade the optics in the interaction region to produce more tightly focused and overlapping beams at collision. The LHC will resume operations in 2026. The decade following these upgrades is called the High Luminosity (HL-LHC) era or sometimes Phase-2 of LHC operations. There will be much higher collision rates that will far exceed the capabilities of the existing CMS detector, which will consequently require significant upgrades to continue to function efficiently. The MTD will be added to CMS to help meet the challenge of high luminosity.

For the HL-LHC, the brightness of the beams and the new focusing scheme at the interaction point will enable the accelerator to deliver an instantaneous luminosity of 2×10^{35} cm^{-2}s^{-1} at the beginning of each fill [1]. However, the nominal scenario is to operate at a stable "leveled" luminosity of 5.0×10^{34} cm^{-2}s^{-1}, to limit the number of interactions during each bunch crossing to 140 on average, by continuously tuning the beam focus and the beam crossing profile during the fills. An ultimate scenario, with a leveled luminosity of 7.5×10^{34} cm^{-2}s^{-1}, will provide 30% more integrated luminosity, at the cost of producing 200 collisions per beam crossing. Hard interactions of interest to CMS, those that probe energy scales ranging from a few GeV to several TeV, occur in far fewer than 1% of the total beam crossings but will always be accompanied by an average of 140–200 additional interactions. The spatial overlap of tracks and energy deposits from the additional collisions can degrade the identification and the reconstruction of

the hard interaction and can increase the rate of false triggers. In addition, the higher collision rate integrated over time results in more radiation damage than can be tolerated by some of the existing subdetectors. The upgraded detector must survive and function efficiently in this much harsher radiation and high pileup environment and must transport a much higher rate of data off the detector to be recorded for analysis.

The primary goal of the CMS Phase-2 upgrade for the HL-LHC is to maintain the current excellent performance of the CMS detector in efficiency, resolution, and background rejection for all final state particles and physics observables used in data analyses. The CMS Upgrade Technical Proposal [2] presents, and the Scope Document [3] further specifies, a detailed plan to deploy an upgraded CMS detector by 2026 to be ready for the start of HL-LHC operations. It identifies the upgrades of CMS components necessary to withstand radiation damage effects and overcome the challenge posed by the high rate of pileup, as detailed in Refs. [4–7].

The CMS collaboration has recently approved a Technical Proposal [8] to include the MIP Timing Detector in the upgrade plan for the HL-LHC era. The MTD will give timing information for MIPs with 30–40 ps resolution at the beginning of HL-LHC operation in 2026, degrading slowly as a result of radiation damage to 50–60 ps by the end of HL-LHC operations. This will help to assign charged tracks to the correct interaction vertices in bunch crossings with an average of 200 collisions or more. It exploits the fact that the individual interactions within the bunch crossing do not all occur at precisely the same time but, because of the longitudinal extent of the beams, are distributed over time with an rms of 180–200 ps. By associating tracks from a vertex to hits and their corresponding times in the MTD, the time at which the collision vertex occurred can be reconstructed. Other tracks pointing roughly towards the vertex but coming at the wrong time, can be eliminated from consideration as contributing to that particular collision. The use of timing and tracking together will, therefore, give CMS excellent association of tracks to vertices even when the vertices are very close together in space. The MTD will provide timing in the barrel and endcap regions, with a hermetic angular coverage up to a pseudorapidity, η, of ± 3, with η defined as $-\ln\tan(\theta/2)$, where θ is the polar angle of a particle measured from the z axis, taken as the beam line. A full definition of the coordinate system used in the CMS experiment and the relevant kinematic variables can be found in Ref. [9].

The time of electromagnetic showers will also be determined to a precision of 30–50 ps above a transverse momentum, p_{T}, of a few GeV in the upgraded calorimeters [6, 7]. We can use these times to associate photons to the correct charged particle vertex, based on compatibility with the time obtained from the MTD.

Pileup mitigation in CMS builds upon particle-flow event reconstruction [10], a technique that combines information from different detectors to establish a list of particle candidates: charged leptons, charged hadrons, photons, and neutral hadrons. It improves the quality of the objects included in the vertex of interest by removing charged tracks that are inconsistent with originating from that vertex, as well as neutral deposits in the calorimeters that might belong to a different interaction based on statistical inference techniques, as described in Ref. [11]. The high spatial granularity of the tracking subdetectors will enable the upgraded CMS detector to separate vertices, to identify the hard collision, and to measure signal particles with good efficiency in the offline analyses [2, 3]. In the transition from 140 to 200 pileup interactions, however, the peak "line density", $\mathrm{d}N_V/\mathrm{d}z$, of the number of collision vertices, N_V, grows from 1.2 to 1.9 mm^{-1}, assuming a line spread along the beam axis of about 4.5 cm rms (averaged over the length of the fill), as seen in Fig. 1.1. The probability of spatial overlaps grows in all subdetectors, and these algorithms begin to fail at a substantial rate. For example, the association of

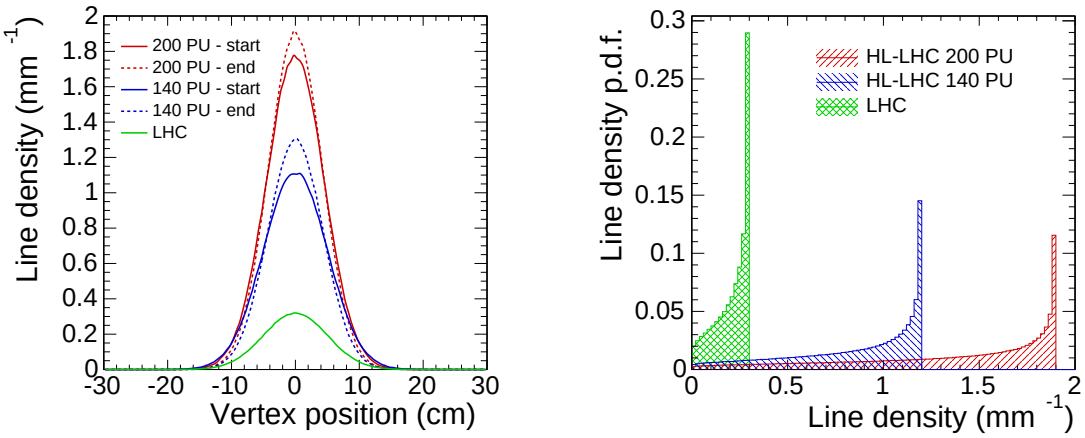


Figure 1.1: Left: Distribution of the vertices along the beam direction at the LHC (Run-1 and early Run-2) with ≈30 pileup interactions and HL-LHC with 140 and 200 pileup interactions. The solid (dashed) line refers to the start (end) of the fill. Adjustments in the focusing of the beam cause the z distributions to become narrower at the end of the fill. Right: Probability density functions of the line density along the beam axis for the pileup of about 30 and for pileup 140 and 200. The modes of the three distributions are 0.3, 1.2, and 1.9 mm^{-1} and their means are 0.2, 0.9, and 1.4 mm^{-1}, respectively.

a track with the primary vertex relies on a requirement on the distance of closest approach to the vertex along the beam axis. Because tracks from displaced sources — such as secondary interactions, decays of particles in flight, and resolution tails — cannot always be fully identified, the optimal selection window has to be set wider than what would be expected by the intrinsic tracking resolution alone. According to simulation, the optimal window is of the order of 1 mm, causing a non-negligible contamination of tracks from pileup into the primary vertex for vertex densities approaching 1 mm^{-1}. The resulting degradation in resolutions, efficiencies, and misidentification rates at 200 pileup interactions impacts several measurements [3, 12]. While measurements relying on isolated objects will suffer mainly from an efficiency reduction when tracks from pileup are incorrectly included in their isolation cones, measurements depending on the resolution of global objects such as missing transverse energy ($p_{\mathrm{T}}^{\mathrm{miss}}$) or on extended objects such as jets may be distorted and this could result in poorer resolution or higher backgrounds.

The timing upgrade of the CMS detector will improve the particle-flow performance at high pileup to a level comparable to the current Phase-1 CMS detector, which is designed to handle a pileup of approximately 50, exploiting the additional timing information from the MTD and the calorimeters. In the time domain, the RMS spread of 180–200 ps within the 25 ns bunch crossing structure of the colliding beams, is approximately constant during the fill and largely uncorrelated with the spatial distribution. If one considers the beam spot sliced into consecutive time exposures of 30–40 ps, the number of vertices per exposure drops to current Run-2 LHC pileup levels of 40–60. A time resolution of this size would therefore reduce the 'effective multiplicity' of concurrent collisions, thereby recovering the Phase-1 quality of event reconstruction.

The event display in Fig. 1.2 visually demonstrates the power of space-time reconstruction in 200 pileup collisions, using a time-aware 4-dimensional extension of the deterministic annealing technique adopted in vertex reconstruction by CMS [13]. According to simulation, instances of vertex merging are reduced from 15% in space to 1% in space-time. The use of timing information for each track, together with its z position extrapolated to the beam line,

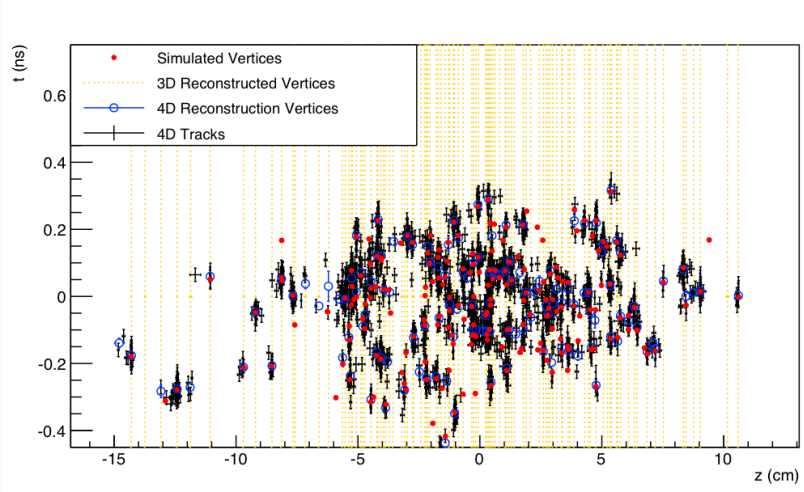


Figure 1.2: Simulated and reconstructed vertices in a bunch crossing with 200 pileup interactions assuming a MIP timing detector with ∼30 ps time resolution covering the barrel and endcaps. The horizontal axis is the z position along the beam line, where the "0" is the center of the IR. The vertical axis is the time with "0" being the point in time when the beams completely overlap in z. The simulated vertices are the red dots. The vertical yellow lines indicate 3D-reconstructed (i.e. no use of timing information) vertices, with instances of vertex merging visible throughout the display. The black crosses and the blue open circles represent tracks and vertices reconstructed using a method that includes the time information and is therefore referred to as "4D". Many of the vertices that appear to be merged in the spatial dimension are clearly separated when time information is available.

reduces the number of tracks from pileup vertices that are incorrectly associated with the hard-interaction vertex. This reduction is quantified in Fig. 1.3. The left plot shows the mean number of tracks incorrectly associated to the primary vertex as a function of the line density of the collision vertices. For a line density of 1.9 collisions per mm, which is the peak density for the case of 200 pileup collisions, the mean number of incorrectly associated tracks reaches over 20

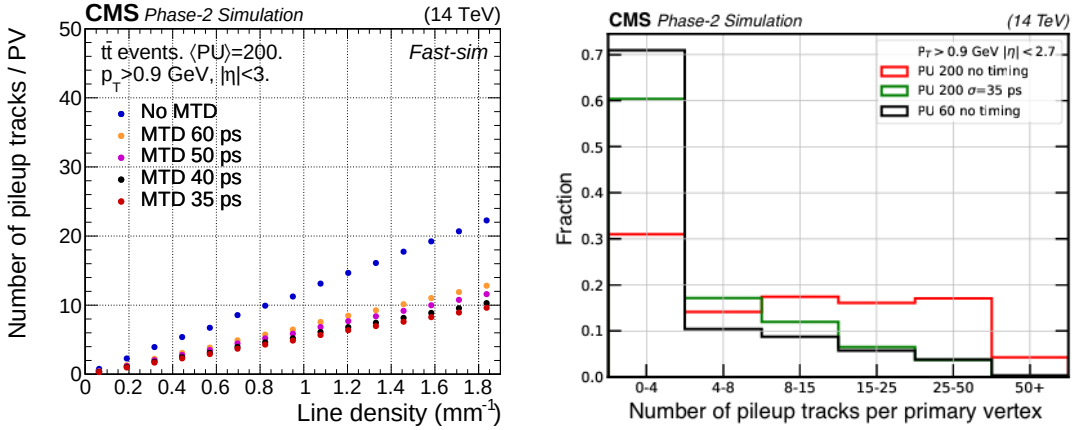


Figure 1.3: Left: Number of pileup tracks incorrectly associated with the hard interaction vertex as a function of the collision line density for different time resolutions. Right: Distribution of the number of incorrectly associated tracks with the use of a 3 σ (where $\sigma = 35$ ps) selection on timing information and without use of timing information. The vertical axis is the fraction of primary vertices which have the number of pileup tracks shown on the horizontal axis associated to them.

without the use of timing information.

The addition of track-time information from the MTD with 30–40 ps precision and a track-time reconstruction efficiency of 85% reduces the wrong associations by more than a factor of two. The peak line density of 1.9 mm^{-1} is effectively reduced to about 0.8 mm^{-1}, which corresponds to less than 80 pileup collisions. The addition of MTD time information, therefore, has the potential to offset fully the CMS performance degradation in the transition from 140 to 200 pileup collisions and to recover the effective background conditions close to Phase-1 operations.

In addition to reducing the mean number of incorrectly associated tracks, the use of timing information significantly reduces the number of primary vertices with a large number of incorrectly associated tracks, as shown in the right plot in Fig. 1.3. The red curve shows the result in the presence of 200 interactions per bunch crossing, when the vertices are reconstructed without the use of timing information: a very large number of tracks from other vertices are wrongly associated to the primary vertex. The green curve shows the number of tracks wrongly assigned to the primary vertex when a 3 σ selection on timing is used, showing that timing information dramatically reduces the number of false associations. For comparison, the black curve shows the performance in a scenario where the pileup is 60 and there is no timing information. This pileup level was experienced during Run-2 and may be typical of operations during Run-3.

The removal of pileup tracks inconsistent with the hard-interaction improves the reconstruction of many final state observables. For example, removing pileup tracks from the isolation cones improves the identification efficiency for isolated leptons, especially τ leptons, and photons, which are key signatures of many processes of interest for the HL-LHC program. The performance of b-jet identification, which relies on vertex reconstruction, is enhanced. The reconstruction of spatially extended objects and global event quantities that are vulnerable to the "pileup pollution", such as jets and $p_{\mathrm{T}}^{\mathrm{miss}}$, are also improved significantly. At pileup of 200, by using track-time information in jet reconstruction, the rate of pileup jets that are actually spuriously clustered particles from pileup interactions, is reduced by 20–40%, depending on η, for a jet p_{T} threshold of 30 GeV. The $p_{\mathrm{T}}^{\mathrm{miss}}$ resolution is improved by about 10%.

In addition to preserving the quality of the data at the highest luminosities, the MTD also brings new capabilities to CMS. The identification of charged hadrons as pions, kaons, or protons based on time-of-flight becomes possible up to a few GeV in p_{T}, which is of significant benefit to Heavy Ion physics and for specialized QCD studies in pp collisions, as discussed in Chapter 5.

As the detailed design of the MTD advanced, it became clear that the goal of a time resolution of 30–40 ps is achievable. However, based on radiation exposures of the candidate detector technologies, it also became evident that the time resolution will degrade approximately linearly with integrated radiation dose, so that at the end of the HL-LHC operation the resolution will be 50–60 ps, with the luminosity-weighted average reaching 40–50 ps. The physics discussion in this introductory section is based on the resolutions that are achieved at the beginning of HL-LHC operation. The sources of radiation degradation in the barrel and endcap regions are given in Chapters 2 and 3, respectively. An evaluation of the impact of the degradation on the overall physics is presented in Chapter 5, which concludes that the MTD provides significant value to the CMS physics program even at the end of HL-LHC operation.

Simulation studies, first presented in the MTD Technical Proposal [8], and further developed in Chapter 5 show that the MTD detector will improve electron, muon, τ, photon, jet, and $p_{\mathrm{T}}^{\mathrm{miss}}$ reconstruction up to $|\eta| = 3.0$, by introducing improvements to the particle-flow reconstruction

Table 1.1: Expected scientific impact of the MIP Timing Detector, taken from Ref. [8].

Signal	Physics measurement	MTD impact
H $\to \gamma\gamma$ and H$\to$4 leptons	+15–25% (statistical) precision on the cross section $\to$ Improve coupling measurements	Isolation and Vertex identification
VBF$\to$H$\to \tau\tau$	+30% (statistical) precision on cross section $\to$ Improve coupling measurements	Isolation VBF tagging, $p_{\mathrm{T}}^{\mathrm{miss}}$
HH	+20% gain in signal yield $\to$ Consolidate searches	Isolation b-tagging
EWK SUSY	+40% background reduction $\to$ 150 GeV increase in mass reach	MET b-tagging
Long-lived particles (LLP)	Peaking mass reconstruction $\to$ Unique discovery potential	β_{LLP} from timing of displaced vertices

so as to achieve a level of performance comparable to the Phase-1 CMS detector at pileup of about 200. The integrated luminosity $\times$ efficiency is increased and this gain is equivalent to collecting data for three additional years beyond the ten year run planned for the HL-LHC.

The remainder of Chapter 1 provides a brief summary of the physics gains made possible by the MTD, a presentation of key design considerations and requirements, and an overview of the MTD detectors.

1.2 Impact of the MTD on the CMS physics program at the HL-LHC

The CMS experimental program at the HL-LHC [2, 3], which includes the precision measurement of standard model (SM) processes, especially the characterization of the Higgs boson, as well as searches for BSM particles and processes, will benefit greatly from the increased luminosity provided by the upgrade of the LHC accelerator complex. The MTD is instrumental in maintaining good resolution and reconstruction efficiency for the physics objects, which are crucial for the HL-LHC scientific program. Efficiency gains from the MTD at the single-object level and the improved $p_{\mathrm{T}}^{\mathrm{miss}}$ reconstruction performance are compounded in multi-object final states — such as di-Higgs boson events or events where the Higgs boson is produced in association with other particles — providing additional overall efficiency gains, at constant rate of reducible backgrounds, of about 20–30% across many measurements.

In addition, the ability to reconstruct the time of displaced vertices will provide enhanced capability in searching for long-lived particles (LLPs) by measuring β_{LLP}, where β is the relativistic velocity parameter, and, in certain cases, permitting the reconstruction of the LLP's mass. The projected performance gains from precision timing with hermetic coverage for $|\eta| < 3$ are summarized in Table 1.1.

The performance of b-jet identification is improved by the removal of pileup contributions close in angle to the candidate signal particles. The gain is particularly effective in the search for di-Higgs production, which can result in many combinations of complex physics objects because of all the possible decay modes of the Higgs boson, and consequently in the direct measurement of the self coupling of the Higgs boson, which is one of the highest priorities of the HL-LHC physics program. For example, precision timing increases the signal yields for constant background in HH $\to$ b$\bar{\mathrm{b}}\gamma\gamma$ by 17% from the barrel alone, and 22% with hermetic coverage, as shown in Fig. 1.4). As discussed in Chapter 5, similar enhancements are predicted for all the relevant di-Higgs boson final states (bbbb, bb$\tau\tau$, bb$\gamma\gamma$, bbWW, bbZZ). For an accumulated luminosity of 3000 fb^{-1}, the MTD brings an improvement to the significance of at least 13%, which would require an additional 26% luminosity to be taken if there were no

MTD. Moreover, the MTD will enable the CMS Phase-2 detector to level the luminosity at 200 pileup interactions and accumulate more luminosity than assumed in this study.

In the case of final states with $H \to \tau\tau$ decays, additional substantial gain arises from the improved quality of the p_T^{miss} reconstruction. The 10–12% improvement in p_T^{miss} resolution at 200 pileup interactions from using time information for charged tracks yields a proportional gain in the resolution of $\tau\tau$ mass reconstruction and in the signal-to-background ratio, and counteracts much of the performance degradation observed in the transition from 140 to 200 pileup events [3]. In addition, the use of time information reduces the rate of pileup jets by about a factor two, leading to an improved performance of Vector Boson Fusion (VBF) tagging, which is useful for the clean observation of some Higgs boson decay modes. More generally, the Higgs boson characterization will benefit from the improved efficiency for multi-object final states ($H \to 4$ leptons or $b\bar{b}$) and from the enhanced vertex identification in $H \to \gamma\gamma$ decays [8], providing improved precision to the measurement of statistically limited differential distributions, which are sensitive to BSM physics [14].

The sensitivity of several searches for new phenomena, including SUSY models, is largely driven by the p_T^{miss} resolution. The 10% gain in the p_T^{miss} resolution with track timing leads to a reduction by a factor of $\approx 40\%$ in the tail of the p_T^{miss} distribution above 130 GeV (Fig. 5.15), which offsets much of the performance degradation of SUSY searches observed in the transition from 140 to 200 pileup [3]. Additional benefits of the precision timing are anticipated in multi-lepton signatures of BSM physics, owing to the increased efficiency in the selection of isolated leptons, and in signatures where a direct measurement of the time-of-flight (TOF) of heavy particles is exploited. For example, a TOF measurement with the MTD will reduce the model dependence in searches for heavy charged stable particles (HCSPs), now limited to particles that traverse the calorimeters [15]. Moreover, the track-time reconstruction opens a new avenue in searches for neutral LLPs, postulated in many extensions of the SM like Split-SUSY, GMSB, RPV SUSY, Stealth SUSY, SUSY models with compressed mass spectra, and many others discussed in Ref. [16] and references therein, even for cases in which the decays are partially invisible, leading to substantial increases in the sensitivity of such searches and providing a novel method to characterize any future discovery. The space-time information associated with the displaced decay vertex, constructed from the decay daughters that do not escape detection, provides the kinematic constraints that are needed to get a direct measurement of the LLP mass. This is shown for the case of a 700 GeV neutralino, χ_1^0 , in Fig. 1.4.

With precision timing, the MTD will also serve as an excellent TOF detector for charged particle (hadron) identification (PID), which is valuable in Heavy Ion physics and in certain specialized studies of low mass QCD and flavor physics. Since a Heavy Ion run is scheduled early in the HL-LHC program, it is appropriate to consider the impact of the MTD based on a time resolution of 30–40 ps. Combined with the wide coverage and full calorimetry of the CMS detector, the MTD will provide unique opportunities for studying the mechanism of interactions between hard colored probes and the quark-gluon plasma (QGP) medium over a wide range of angles, and detailed dynamics of heavy flavor particles over a wide rapidity range in heavy ion collisions. Figure 1.5 shows the expected performance in identifying charged pions, kaons, and protons, as a function of transverse momentum p_T and rapidity y, in the barrel region and the endcap region. In the barrel region, a minimal p_T of 0.7 GeV is required for particles to reach the timing detector. At midrapidity, proton identification up to $p_T \approx 5$ GeV is achievable, while pions and kaons can be separated up to $p_T \approx 2.5$ GeV. The PID capability will, in general, decrease toward high rapidity due to the rapid increase of total momentum and thus reduced TOF difference. Nevertheless, good performance can still be achieved out to $|y| \approx 2$.

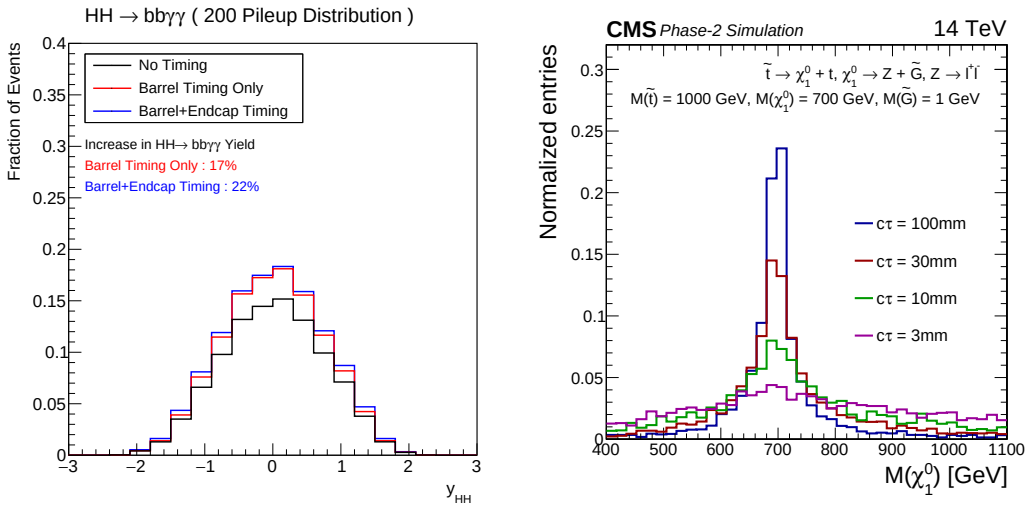


Figure 1.4: Left: Impact on signal efficiency for HH $\rightarrow$ b$\bar{\text{b}}\gamma\gamma$ for no-timing, barrel only timing, and barrel plus endcap timing scenarios. The quantity y_{HH} is the rapidity of the di-Higgs system. Right: Mass peak of a 700 GeV neutralino, χ_1^0, with three different lifetimes reconstructed from the kinematic closure of the secondary vertex using time information with 30 ps resolution.

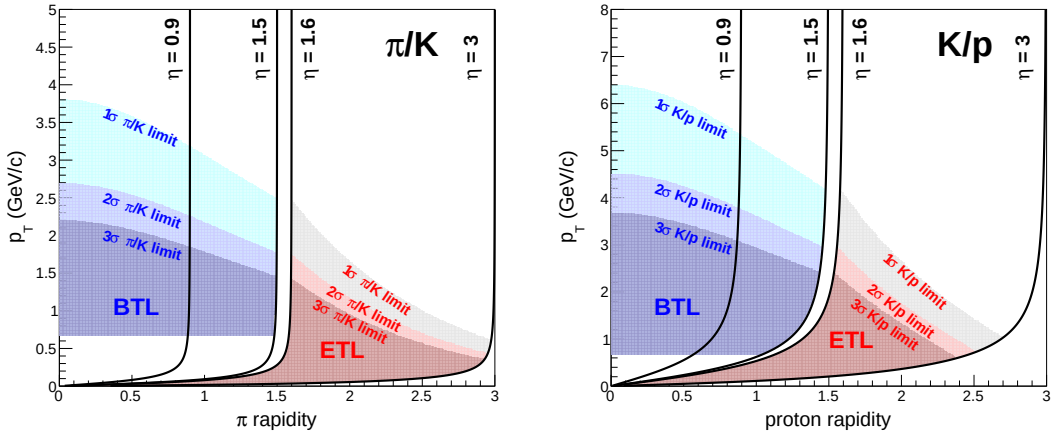


Figure 1.5: Expected performance for charged particle identification in p_{T} and rapidity in an early run of the HL-LHC with the proposed MTD, with a time resolution of about 30 ps, which is achievable since there will not yet be significant radiation damage. Different colors of the shaded regions correspond to 1, 2 and 3 σ separations. Left: The π/K separation vs p_{T} and rapidity y under the pion hypothesis. Right: The K/p separation vs p_{T} and rapidity y under the proton hypothesis. Shown also are contours in pseudo-rapidity, $|\eta|$. The label BTL in these plots indicates the barrel timing detector and the label ETL indicates the endcap timing detectors.

The performance of particle identification for the proposed MTD is compared to that of the TOF systems in the STAR [17] and ALICE [18] experiments at midrapidity ($|y| < 0.9$–1.0). Table 1.2 summarizes key parameters of the TOF system in each experiment for the radius, r, of the cylindrical barrel region, which is directly related to the particle flight distance, L; the time resolution (σ_T); and the ratio of r to σ_{T}, which characterizes the TOF PID capability. Although the CMS-MTD has the shortest flight distance, constrained by the available space in CMS, with the design time resolution of 30 ps, the PID performance is expected to be 40% better

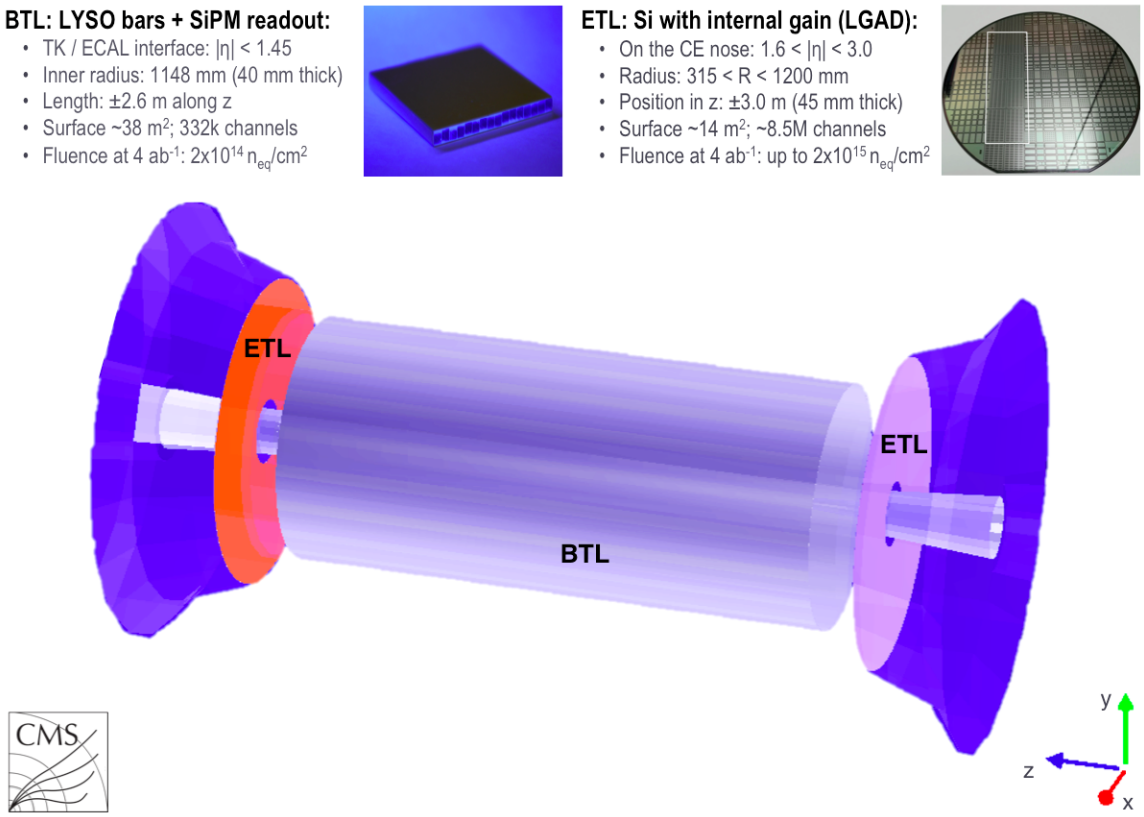


Figure 1.6: A schematic view of the GEANT geometry of the timing layers implemented in CMSSW [20] for simulation studies comprising a barrel layer (grey cylinder), at the interface between the tracker and the ECAL, and two silicon endcap (orange and light violet discs) timing layers in front of the endcap calorimeter.

than the STAR-TOF and about 40% worse than the improved ALICE-TOF performance [19] out to $|y| < 0.9$, where their coverage ends. More importantly, the wide acceptance of the MTD provides CMS a unique PID coverage out to high rapidity.

Table 1.2: Summary of key parameters of the time-of-flight system for different experiments.

Experiment	r (m)	σ_T (ps)	r/σ_T ($\times 100$) (m $\times$ ps^{-1})
STAR-TOF	2.2	80	2.75
ALICE-TOF	3.7	56	6.6
CMS-MTD	1.16	30	3.87

1.3 Considerations and requirements for the design of the MTD

The design of the MTD is driven by scientific requirements which follow from the physics goals of the HL-LHC program and engineering requirements and constraints. It must conform to the requirements imposed on all detectors inside CMS, such as tolerance to magnetic fields and robust mechanical design that can survive for the full duration of the HL-LHC program. The need to fit within the existing CMS detector and conform to the HL-LHC upgrade schedule puts many additional constraints on the MTD. These key requirements and constraints and some conclusions that emerge from them are presented in this section.

1.3.1 Time resolution

The requirement for the time resolution of the MTD is based on simulations that show that the impact of pileup is reduced significantly for a resolution of 30–40 ps, which is achievable at the beginning of HL-LHC operation, and continues to be effective even though, because of radiation damage, the resolution degrades to 50–60 ps by the end of HL-LHC operation.

1.3.2 Space constraints for detectors and services

A broad range of physics studies motivates the need for nearly hermetic coverage with high efficiency. The overall geometric envelope and the service channels that provide cooling, power, and data-transmission to the already approved detectors for the HL-LHC cannot be changed. There are, then, only two locations where timing detectors can be installed: in the space between the last layer of the outer tracker (OT) and the beginning of the barrel electromagnetic calorimeter (ECAL), which will allow coverage of the barrel region; and in the gap between the Tracker bulkhead and the Phase-2 high granularity endcap calorimeter, (CE), which will permit coverage in the endcap region. In each of these locations, very little space is available for the MTD and its services. Alternatives to these locations were considered but were ruled out because of technical issues.

The MTD then must be divided into two sections, the Barrel Timing Layer (BTL) covering $|\eta| < 1.5$ and the Endcap Timing Layer (ETL) covering $1.6 < |\eta| < 3.0$. This division follows the layout of the CMS detector and is necessary for the MTD to fit into the detector, be compatible with the way in which the detector is opened for maintenance, and have access to the service channels necessary for detector operations and readout. It is also driven by technical considerations, especially the significant difference in integrated radiation dose between the two sections.

These considerations lead to the following requirements and conclusions:

- The BTL shall be integrated into the support tube for the OT, occupying a limited space just inside its inner radius, and at larger radius than any of the barrel tracking layers. The space allotted for the BTL is defined to extend 40 mm inward towards the beam from the inner surface of the Tracker Support Tube (TST), a carbon fiber cylinder approximately 2 m in diameter and 5 m long. The BTL services (readout fibers, power cables and cooling pipes) will be integrated along with those of the OT.
- The ETL shall be supported on the CE nose cone, between the Tracker and the CE, in a volume that is mechanically and thermally separated from the CE. The geometry of the nose cone prevents the ETL from extending to $|\eta| > 3$, so it cannot completely cover the full acceptance of the upgraded Tracker, which will reach to $|\eta| = 4$ for the HL-LHC. Currently, a longitudinal space of 45 mm between the Tracker and the CE has been allocated for the ETL on each end of CMS. The ETL services shall fit into the channel just outside the CE.

Figure 1.6 shows a schematic view of the proposed MTD layout, comprising both BTL and ETL sections, as implemented in the GEANT simulation of the CMS detector.

1.3.3 Impact on Tracker and CE integration and installation

The CMS Outer Tracker, which was part of the original CMS construction project, will be removed and replaced with a completely new device. Since the OT's silicon layers and the BTL both reside in the TST, the integration of the two detectors must be carefully orchestrated. The

BTL installation has to be completed either before the integration of the Tracker into the TST has started or at specific points during the integration of the Tracker Barrel (TB). The integration of the BTL must be completed before the start of the integration of the Tracker Endcap Disks, which must proceed on schedule. The requirement that the BTL integration shall not interfere with the integration of the Outer Tracker has strong implications for the development and construction schedule of the BTL.

The ETL is mounted on the nose of the CE, which is also a completely new detector that will replace the current endcap calorimeter. However, since the ETL and CE occupy physically different volumes and are separated thermally (although possibly sharing the cooling plants), the two detectors can be integrated separately. The ETL can be installed on the CE when it is above ground and lowered with it or it can be installed in the experimental cavern after the CE is lowered. It is possible to install all or part of the ETL in a Technical Stop (LHC annual maintenance period of a few months) or Long Shutdown after the initial run of the HL-LHC. The ETL schedule is therefore not tightly coupled to the CE and it can potentially exploit a longer period to complete development and construction.

1.3.4 Impact on the performance of other detectors

The presence of the BTL in the TST requires a reduction in the radius of the outermost barrel tracking layer. This has been shown to have only a negligible impact on the tracking performance and momentum resolution.

Since the BTL also shares the dry/cold volume with the tracker, it must not create thermal, electronic or mechanical problems that would affect tracker performance. Since the BTL is also quite close to the barrel ECAL, which requires precise temperature control, it must not disturb the ECAL's thermal environment.

In order for the ETL detector to be accessible for repairs and replacements of faulty components when the CMS detector is open during shutdowns, the ETL will occupy an independent cold and dry volume that is isolated and operated separately from the CE. ETL services will be routed near the CE and must not create thermal, electronic, or mechanical problems that would affect CE performance. While the ETL is in its own cold volume, it may share a cooling plant with the CE and, if so, must operate within specifications that guarantee that it will not affect the CE performance.

The electron and photon energy resolution of the calorimeters is degraded by the material in front of them. The additional material in the timing layers should not degrade the performance of the calorimeters in a significant way.

1.3.5 Segmentation/occupancy

In order to ensure that useful timing information is available for as many charged tracks as possible, the maximum sensor size must be compatible with an occupancy less than a few percent, ensuring a large probability for single hits needed for unambiguous time assignment. Dead area between sensors must be minimized to preserve the efficiency. The detector must also produce a manageable data volume.

For both the BTL and ETL, the choice of the size of an individual detector element (cell) is also a compromise between occupancy, sensor characteristics, including electronics considerations such as input capacitance and manufacturing issues, total power needs (number of channels), and cost. A maximum per-channel area of order 1–2 cm^2 is suitable for the BTL. For the ETL, a sensor cell-size of $\approx 2\,mm^2$ at $|\eta| \approx 3$ is a good compromise between the spread in the

time response within a channel, low occupancy, and low channel count rate. For the ETL, the occupancy is significantly less at lower η.

1.3.6 Compatibility with CMS Trigger and Data Acquisition systems

The current CMS trigger uses two levels of selection: a first level, Level-1, based only on signals from the calorimetry and muon detectors, that uses dedicated hardware to do simple computations very quickly, in less than 4 μs, on every bunch crossing and passes about 100 000 crossings/s with interesting events to the next level; and second, the High Level trigger (HLT), which uses the full information from the detector, including charged particle tracking, to do much more sophisticated computations on a cluster of thousands of advanced microprocessors to choose about 1000 crossings/s with most interesting events to record for detailed analysis.

The Level-1 trigger for the HL-LHC will add charged particle tracking based on the Outer Tracker to the information from the calorimeters and muon detectors. Level-1 will have a latency of 12.5 μs and an average rate of acceptance of 750 kHz. The MTD electronics must be capable of delivering data registering the time at which a MIP traverses the sensitive volume, known as a time stamp, and other required information, such as pulse height for time-walk correction, upon receipt of a Level-1 Accept (L1-A) from the CMS trigger without introducing readout dead time. The format of the data from the MTD must conform to CMS standards. The readout must meet CMS performance requirements, including very low data loss. The electronics must be able to receive and respond to fast and slow control information, including throttling, reset and resync signals, and must produce and deliver the monitoring information needed to ensure proper operation.

There is a proposal to use timing information from the MTD in the Level-1 trigger. One key requirement for this new capability is that it does not add significantly to the cost, power consumption, or complexity of the MTD front-end or back-end electronics. This possibility is discussed in Section 4.3 and Appendix E, and briefly summarized in this introduction in Section 1.4.5.

1.3.7 Operating temperature

The operating temperature of both the BTL and the ETL must be kept low because the time jitter caused by noise and leakage currents, which are strong functions of temperature, has a significant effect on the overall time resolution. Radiation exposure increases the noise or dark current rate (DCR) and may lower the signal from MIPs. Heavily irradiated detectors can partially recover from the radiation-induced damage if warmed to room temperature or above, during periods when there are no collisions, reducing the DCR.

Therefore, operating temperatures in BTL and ETL shall be maintained as close as possible to $-30\,^{\circ}$C, a level that can be comfortably achieved with cooling systems that have been used in CMS and which limits the noise-induced jitter to an acceptable level. In fact, $30\,^{\circ}$C should be considered a conservative baseline and the possibility of going somewhat lower will discussed in a few places below. The cooling and temperature control, as well as the operational procedures for the MTD, shall allow the maintenance of the detectors at room temperature for extended periods of time, for example during LHC shutdown periods.

In order to achieve these goals, the design of the cooling system for the detector modules must ensure efficient heat removal from the sensors. Furthermore, the MTD modules and their supports will undergo several thermal cycles as part of testing as well as during operations; therefore the MTD modules shall provide a robust mechanical structure that shall not deteriorate in fit or function even after many thermal cycles from $-30\,^{\circ}$C to $+20\,^{\circ}$C.

1.3.8 Radiation tolerance

The CMS detectors for Phase-2 described in the CMS Upgrade TDR documents are designed to operate efficiently throughout the HL-LHC phase, up to an integrated luminosity of at least 3000 fb^{-1}. The radiation level predictions at the location of each detector are based on the FLUKA Monte Carlo multi-particle transport code [21, 22], using a preliminary implementation of the Phase-2 CMS geometry, described in the CMS Tracker TDR [4]. The geometry model includes only minimal updates to the Phase-1 geometry model, such as the Phase-2 layer structure of the Tracker, an average description of the HGCal material without longitudinal segmentation, and the replacement of the CMS endcap preshower with an 18 cm thick neutron moderator. This model was also adopted for radiation level predictions in the MTD Technical Proposal [8], although it did not include the implementation of the MTD itself, and the neutron moderator in front of the endcap calorimeter was thicker than physically allowable with inclusion of the MTD. To cover the uncertainties in the predictions associated with the approximate geometry model, all CMS detectors were required to have a minimum margin of 1.5, without performance degradation compared to 3000 fb^{-1}. The reduction of the neutron moderator from 18 to 12 cm, to accommodate the endcap timing layer of the MTD, was checked to give an increase lower than 15% in the neutron fluence in the MTD and in the Tracker.

In this TDR, the radiation level predictions are based on a refined implementation of the CMS Phase-2 geometry, comprising the description of the MTD geometry; a neutron moderator with a thickness of 12 cm; an accurate description of the services in the transition region between the Tracker bulkhead and the CE and in the service channel between the barrel and the endcap; and up-to-date descriptions of the Phase-2 envelope of the endcap detectors, the HGCal longitudinal segmentation, and the beam pipe. The details of the model and of the uncertainties associated with the fluence and radiation dose predictions are described in Appendix A.

Table 1.3 shows the expected particle fluences and radiation doses at the location of the timing layers in the barrel and the endcaps, for an integrated luminosity of 3000 fb^{-1}. At this integrated luminosity the detectors will have received radiation doses of up to about 30 kGy and a "1 MeV neutron equivalent fluence", $\text{n}_{\text{eq}}/\text{cm}^2$, of up to 1.9×10^{14} $\text{n}_{\text{eq}}/\text{cm}^2$ in the barrel, and 450 kGy and 1.6×10^{15} $\text{n}_{\text{eq}}/\text{cm}^2$ in the high-η part of the endcap. For radiation tolerance qualification, all the MTD components are required to stand radiation levels at least a factor 1.5 (safety factor) larger than the nominal prediction. The fluence and the radiation doses corresponding to this safety factor are also given in the table. A safety factor of 1.5 covers the uncertainties in the predictions from the geometry model, the pp inelastic cross-section, and possible sensor-to-sensor variations, as discussed in Appendix A.

The MTD sensor technologies shall be validated up to 3×10^{14} and 3×10^{15} $\text{n}_{\text{eq}}/\text{cm}^2$ in the barrel (Chapter 2 and Appendix B) and in the endcap (Chapter 3), respectively. The endcap design, with the combination of two hits per track, enables the ETL to be operated up to these radiation levels with a constant per-track resolution of 35 ps. The BTL, instead, will experience a linear degradation in the time resolution, at the conservative working conditions of $-30\,^{\circ}\text{C}$, combined with annealing at room temperature during the winter shutdowns. Operation at lower temperature or short annealing periods at $+50\,^{\circ}\text{C}$ would enable mitigation of the effect, as discussed in Appendix B. The above figures provide a safety margin of 1.6 and 1.8 for BTL and ETL respectively, for an integrated luminosity of 3000 fb^{-1} at the point with largest radiation level. The addition of the MTD will enable CMS to level the HL-LHC luminosity at 200 collisions per beam crossing, thus providing the potential to integrate up to 4000 fb^{-1} of luminosity, including 300 fb^{-1} from Phase-1. In this case, the safety margin would be reduced to a factor of about 1.3.

Table 1.3: Nominal radiation doses and fluences at various locations of the timing layers after 3000 fb^{-1}. The last two columns show the radiation levels providing a safety margin of a factor 1.5. The fluence is normalized to 1 MeV neutron equivalent in silicon.

				3000 fb^{-1}		1.5×3000 fb^{-1}	
Region	$\lvert\eta\rvert$	r (cm)	z (cm)	n_{eq}/cm^2	Dose (kGy)	n_{eq}/cm^2	Dose (kGy)
Barrel	0.0	116	0	1.65×10^{14}	18	2.48×10^{14}	27
Barrel	1.15	116	170	1.80×10^{14}	25	2.70×10^{14}	38
Barrel	1.45	116	240	1.90×10^{14}	32	2.85×10^{14}	48
Endcap	1.6	127	303	1.5×10^{14}	19	2.3×10^{14}	29
Endcap	2.0	84	303	3.0×10^{14}	50	4.5×10^{14}	75
Endcap	2.5	50	303	7.5×10^{14}	170	1.1×10^{15}	255
Endcap	3.0	31.5	303	1.6×10^{15}	450	2.4×10^{15}	675

For the BTL, no maintenance access for repairs is possible for the lifetime of the HL-LHC while the ETL shall be designed to be accessible for repairs in situ and shall be capable of being removed from the collision hall, repaired, and reinstalled during an extended Technical Stop.

To ensure that the MTD, as a whole, can maintain the required performance through the lifetime of the HL-LHC, each component that will be located in the experimental cavern must be shown to function properly when exposed to the full expected radiation dose plus the additional amount corresponding to the safety factor. The silicon photomultipliers (SiPMs) chosen as the photosensor in the barrel shall maintain a DCR within specifications up to a fluence of 3×10^{14} n_{eq}/cm^2 and the front-end BTL ASICs shall be radiation tolerant and single event upset (SEU) compliant to the same fluence. The low gain avalanche detectors (LGADs) chosen as the sensors in the endcap shall be tolerant to 3×10^{15} n_{eq}/cm^2 at $|\eta| = 3$, and the ETL front-end ASIC shall be radiation tolerant and SEU compliant also up to 3×10^{15} n_{eq}/cm^2. When SEUs do occur, the system must be equipped with proper diagnostics and controls to detect them and reset them quickly.

1.4 Overview of the MIP Timing Detector design

Mechanical constraints, performance, radiation tolerance, cost, and the upgrade schedule led to a detector design consisting of a thin layer between the Tracker and the calorimeters, divided into a barrel ($|\eta| < 1.5$) and two endcap sections covering up to $|\eta| = 3.0$. The requirements on the MTD are rather different in the barrel and endcap regions. The radiation environments are quite dissimilar, with the outer radius of the ETL, ≈1.2 m, receiving about the same dose as the highest $|\eta|$ part of the BTL but the inner radius, at ≈0.3 m, receiving nearly a factor of 30 more. Moreover, the surface area of the BTL is about 2.5 times the surface area of the two endcaps.

Five technologies were investigated and studied in dedicated beam tests and radiation exposures, building upon and extending long-standing R&D programs [23–33]. For the BTL, the best available technology is a crystal scintillator that is read out with SiPMs [23–25], which are pixelated avalanche photodiodes operating in Geiger breakdown mode. For the ETL, the best performance is achieved with LGADs [26–28], which are silicon sensors with internal gain of about 10–30.

The SiPM technology, used in the BTL, is not sufficiently radiation tolerant to work in the endcap and the cost of instrumenting the barrel with LGADs is prohibitive. The two different sensors also require rather different front-end ASICs. Moreover, the schedule constraints are also different. There is less time for development and construction for the BTL, requiring the

choice of a technology that needs relatively little R&D and for which production in industry is well-established. Consequently, in the detectors and the associated front end electronics, the synergy between the BTL and ETL is limited. The BTL and ETL do share common clock and backend systems, cooling, and detector slow controls and safety systems.

The following sections give a brief introduction and overview to the MTD design. Detailed descriptions are provided in the three chapters that follow.

1.4.1 Overview of the Barrel Timing Layer

The Barrel Timing Layer is a thin, cylindrical detector that will be housed inside the TST, shown in Fig. 1.7, at its outer radius. The inner boundary of its radial envelope is 1148 mm from the beam and the outer boundary is at 1188 mm, for a maximum radial extent of 40 mm. Its overall active length in z is about 5000 mm. The total active surface is about 38 m^2. The location of the BTL in the Outer Tracker volume imposes many constraints on its design and operation, described above. Importantly, the components of the BTL must be scheduled for installation and integration into the TST so as to not disrupt or delay the Tracker installation and integration. This places constraints on the construction of BTL on-detector components, which must be available for installation by the third quarter of calendar year 2023. Finally, the TST is embedded in CMS and cannot be removed or serviced during the lifetime of the HL-LHC so that the reliability of the BTL is a primary consideration, demanding excellent and robust design and careful testing and integration.

To satisfy the concerns related to schedule and accessibility, it was decided to choose technologies for the sensor that are well-established and with which CMS has experience. Both LYSO:Ce scintillating crystals and SiPM devices are technologies for which there are well-established production and assembly procedures and facilities in industry. The R&D for a precision timing application is well advanced, and small prototypes consisting of LYSO:Ce crystals read out with SiPMs have been proven capable of achieving time resolution below 30 ps [34]. Both the crystals and the SiPMs are already proven to be radiation tolerant up to a neutron equivalent fluence of at least 2×10^{14} cm^{-2} and a total integrated dose of 25 kGy, when cooled to approximately $-30\,^{\circ}C$. The readout electronics can be adapted from existing positron emission tomography (PET) applications, which include time-of-flight (TOF-PET) measurement capability [35–37].

The BTL will cover the pseudorapidity region up to $|\eta| = 1.48$. The fundamental detecting cell will consist of a thin LYSO:Ce crystal bar of about 5.7 cm in length oriented along the ϕ direction in CMS, a width of 3.0 mm along the z direction, and a variable radial thickness. The properties that make LYSO:Ce suitable for this application are discussed in Section 2.2.1. The thickness is varied along the barrel length with 3.7 mm for $|\eta| < 0.7$; 3.0 mm for $0.7 \leq |\eta| \leq 1.1$; and 2.4 mm for $|\eta| > 1.1$, to maintain an approximately constant slant depth crossed by particles coming from the interaction point. Each end of the bar will be coupled to a SiPM whose dimensions will be 3 mm along ϕ and a variable thickness, chosen to approximately match the bar's radial thickness in each $|\eta|$ interval. The total number of SiPMs will be 331 776. The readout of both ends of the bar provides two measurements of the time of arrival of a MIP that are combined to eliminate the effect of the time delay of the light traveling along the crystal.

The proposed layout has no impact on the upgraded designs and schedules of the ECAL, while the sharing of the cold volume with the Tracker requires close coordination between the two projects. The Tracker design has been updated during 2018 with a slightly reduced outer radius to provide sufficient room for the BTL. The TST structure, with the modifications necessary to

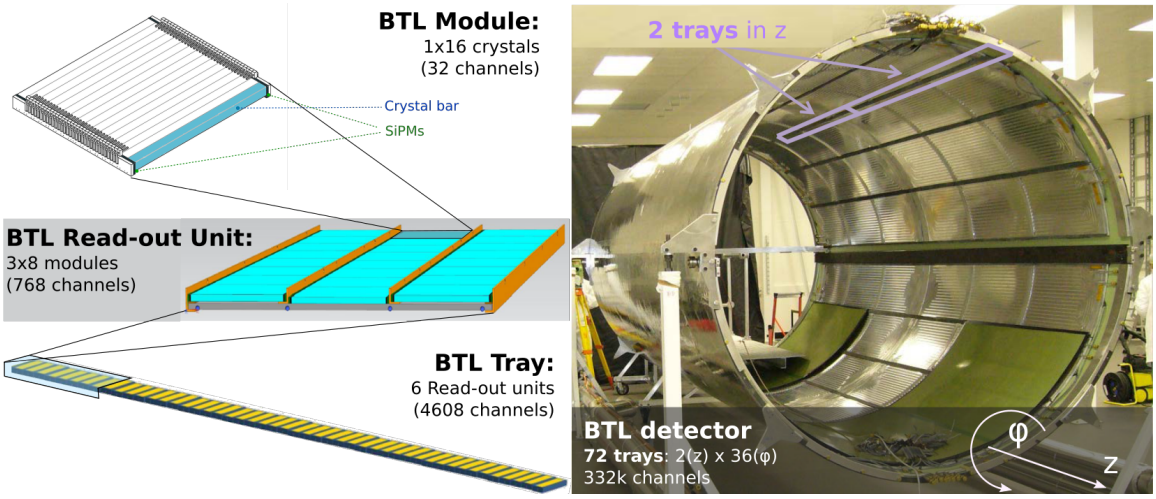


Figure 1.7: Overview of the BTL showing (left) the hierarchical arrangement of the various components, bars, modules, and Readout Units, and (right) trays (purple rectangles near the top), inside the TST.

support the BTL, must be in place before the integration of modules into the upgraded Tracker (or the BTL) can begin. The BTL has a negligible impact on the performance of the Tracker in the barrel, and no impact in the endcaps. The momentum resolution changes from $\delta p_T / p_T = 0.54\%$ and 0.92% at 10 and 100 GeV with the nominal geometry of the Tracker TDR [4] to 0.548% and 0.936% with the reduced outer radius, without including BTL hit information in the track reconstruction. Similarly, a simulation study, with a 4 mm thick LYSO:Ce layer, i.e. thicker than in the reference design, indicates no significant impact on the performance of shower reconstruction and energy resolution in the ECAL. Preliminary results discussed in Ref. [8], are summarized and updated in Appendix D.

SiPMs operate above the breakdown voltage in Geiger mode with a gain of the order of 10^5. The over-voltage (OV) produces a dark current that grows as the radiation dose accumulates. Since dark current increases by roughly a factor of two for each increment of 7–10 °C, the SiPMs will be operated at low temperatures of about −30 °C. This results in the need for substantial cooling power. Because the over-voltage (OV) also controls the photon detection efficiency (PDE), there is a tradeoff between noise rate and signal size, and therefore time resolution. The SiPM operation voltage will have to be smoothly decreased during the detector lifetime to limit the noise level while maintaining good time resolution. The electronics must also be designed to handle large leakage currents. This is discussed in more detail in Chapter 2.

The BTL is read out by a dedicated ASIC, named the TOFHIR (Time-of-flight, High Rate) chip, that delivers precision timing information for 32 SiPMs based on discrimination of the leading edges (LE) of their pulses followed by measurement with a time-to-digital converter (TDC). In order to achieve high precision, the input to the discriminator has to have a very fast rise time, $\mathrm{d}V/\mathrm{d}t$, which requires a lot of amplification and consequently a lot of power. In this technique, the amplitude of the pulse also has to be measured to correct the time walk (the variation of the threshold-crossing time with pulse height). After an exposure to radiation of about 0.7×10^{14} $\mathrm{n_{eq}/cm^2}$ that will be accumulated in an integrated luminosity of about 1000 fb^{-1}, the power consumption of the BTL is dominated by this dark/leakage current, which must be compensated by circuitry in the ASIC. Fluctuations in the dark current cause a jitter that degrades the time resolution and, at high integrated doses, towards the end of HL-LHC operation, will be the dominant contribution to the time resolution. The dose received by the BTL is relatively uniform in η so this is a challenge for the entire detector.

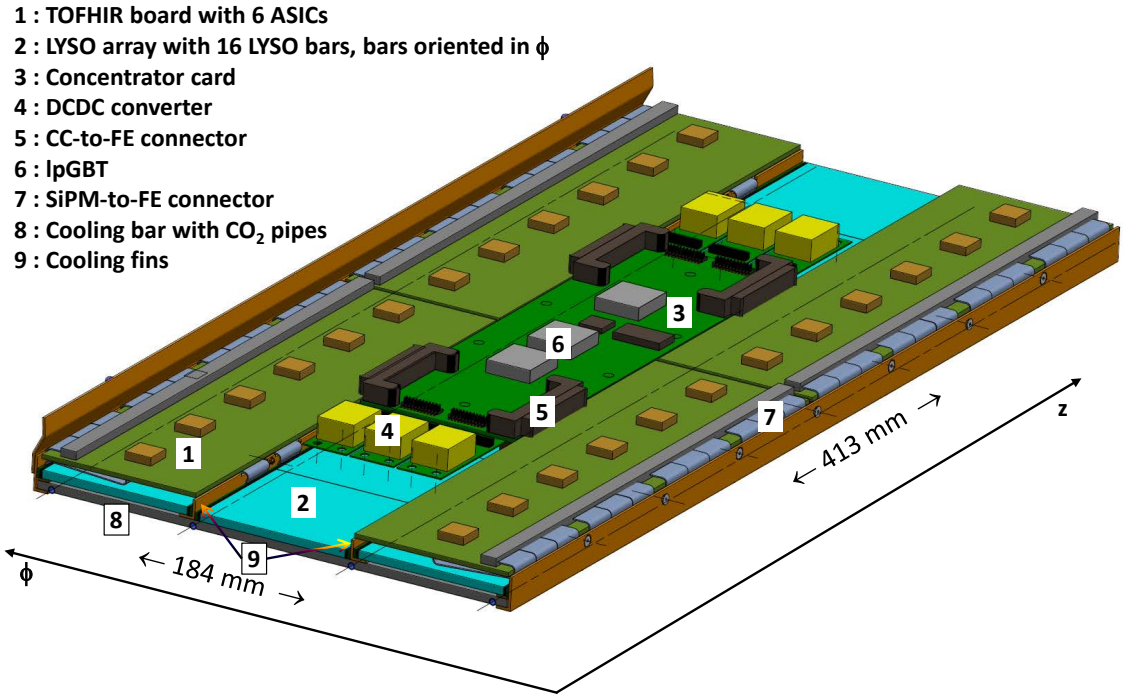


Figure 1.8: Representation of a Readout Unit, which supports 768 SiPMs.

The detector is divided longitudinally into $+z$ and $-z$ end of length 2.6 m, each end consisting of 36 azimuthal segments, which span 10° each. Each azimuthal segment has a row of six Readout Units (RUs), supporting 4608 SiPMs in a unit called a tray. In all, the 72 trays contain 331 776 SiPMs. Services are routed along the tray to the RUs. The ASICs are mounted on front-end (FE) boards, six ASICs per board, which are connected to their SiPMs via small PCBs or flex circuits. Four front-end boards plug into a Concentrator Card (CC) that provides low voltage power, bias voltage, and three low power Gigabit Transceivers (lpGBTs) that carry data and control signals to and from the ASIC. The ASICs, FE, CC, and associated power supplies constitute the RU, which supports 768 SiPMs. The lpGBT transmits a L1-A from the CMS trigger to the FE and, on receipt of an L1-A, sends the data from the FE to the back-end electronics in the CMS Underground Service Cavern (USC). The lpGBT also exchanges control and monitoring information between the FE and the DAQ.

Figure 1.7 provides an orientation to the BTL and its components. The left side of Fig. 1.7 shows how the 16 crystal bars are assembled to form modules and how 24 modules are grouped into an RU. It also has a sketch showing how six RUs populate a tray. The right side of Fig. 1.7 shows the TST and indicates the position of the trays on the periphery.

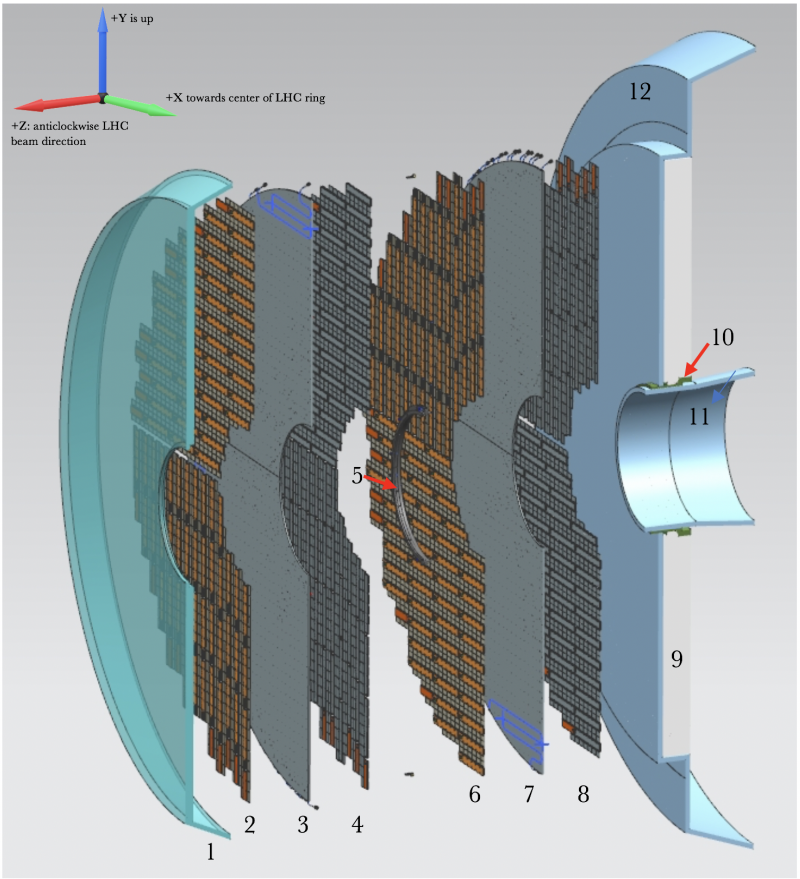
Figure 1.8 is a representation of a Readout Unit showing four Front End boards each with six TOFHIR ASICs connecting to the CC. Also shown are the lpGBTs and the DC-DC converters for the power. The LYSO:Ce bars are visible and the flexible connections from the SiPMs on their ends to the ASICs can also be seen.

1.4.2 Overview of the Endcap Timing Layer

The endcaps will be instrumented on each side ($\pm z$) of the interaction region (IR), with a hermetic, two-disk system of MIP-sensitive silicon devices providing two hits per track with excellent time resolution. The two hits are needed to achieve the desired time resolution. Each pair of disks is located between the CE and the end of the Tracker, about 3 m from the IR and covers an annular region, $315 < r < 1200$ mm, corresponding to a pseudorapidity acceptance of $1.6 < |\eta| < 3.0$. Each disk has silicon devices on both faces to cover the whole area without cracks or gaps for readout electronics and services. The longitudinal space allowed for the ETL is about 45 mm. The choice for this placement within CMS (as opposed, for example, to placing the ETL inside the TST or CE) is determined by the necessity to ensure its accessibility throughout the HL-LHC operation because it is subject to high radiation dose and may need repairs or upgrades. It follows that the ETL must occupy its own independent, thermally isolated volume mounted on the nose of the CE detector. To facilitate maintenance, the disks are split down the center vertically so that they form a "clam shell" around the beam pipe and can be removed and reinstalled during a typical, or perhaps somewhat extended, Technical Stop without the need to remove the beam pipe. For each ETL endcap, the active sensor area (sum of both disks) is about 7.2 m^2 and the total weight is 280 kg.

As shown in Table 1.3, the radiation dose for the ETL is much greater than for the BTL and highly non-uniform in η. The SiPM technology selected for the BTL does not have sufficient radiation tolerance to work over most of the η range of the ETL. Therefore, the ETL will use planar silicon devices with internal gain. The design uses Low Gain Avalanche Detectors (LGAD) [27, 28], which are also under consideration for a fast-timing layer in the very forward region ($2.4 < |\eta| < 4.8$) of the ATLAS experiment [38]. The LGAD sensors have intrinsic gain of 10–30 provided by a special implant, which helps to overcome capacitance and other noise sources and achieve a low-jitter fast-rising pulse edge that enables precision timing reconstruction for MIPs. Sensors with a 50 μm active region, within a normal 300 μm thick silicon wafer, and a thin implanted gain layer are expected to provide the desired performance. Radiation tolerance studies demonstrate a resolution of about 30 and 50 ps at fluences corresponding to $|\eta| \simeq 2.5$ and 3.0, respectively, at the end of the HL-LHC operation. Achieving good time performance at low-gain requires cell sizes typically less than 2 mm^2, to limit the sensor capacitance. Therefore, the LGAD solution requires a large number of sensor pads to cover the full area of the endcap. The design calls for a module which is a 16×32 array of square LGAD pads of 1.3×1.3 mm^2, giving an overall size of 21×42 mm^2. Each module is read out by two ASICs, called Endcap Timing Readout Chips (ETROCs), each with dimensions approximately 20×20 mm^2 and each processing signals from a 16×16 submatrix. The sensor unit size will be larger than the chip size so that the two ETROCs can be bump-bonded to it. The readout chip uses timing of the leading edge of the pulse from the LGADs and contains amplifiers and discriminators followed by circuits to measure the Time-of-Arrival (TOA) of each particle and Time-over-Threshold (TOT) to measure the pulse height for time walk correction. The total number of 1.3×1.3 mm^2 pads is about 4×10^6 per end in z, requiring about 16 000 ETROCs per end. A group of readout chips communicates with an on-detector board, called the service hybrid, that reads out the ETROCs on receipt of a Level-1 trigger and sends the data to the USC or further processing. The service hybrid also supplies DC power, bias voltage, communications (slow and fast control), and monitoring to the ETROCs. The service hybrid contains two boards, a readout board and a power board.

Figure 1.9 presents a view of the ETL along the beam at 3 m on one side of the IR. There is a mirror image on the other side of the IR. Shown are two half-disks with sensors on both faces of each disk. Visible are the LGADs (shown as small grey squares) and the service hybrids,



1: ETL Thermal Screen
2: Disk 1, Face 1
3: Disk 1 Support Plate
4: Disk 1, Face 2
5: ETL Mounting Bracket
6: Disk 2, Face 1
7: Disk 2 Support Plate
8: Disk 2, Face 2
9: HGCal Neutron Moderator
10: ETL Support Cone
11: Support cone insulation
12: HGCal Thermal Screen

Figure 1.9: Cross-sectional view of the endcap timing layer (ETL) along the beam axis. The interaction point is to the left of the image. Shown are two ETL disks populated with modules on both faces, along with the support structure. The grey sections are the active areas of the modules with LGAD sensors. Each orange bar represents a service hybrid. The neutron moderator, labelled 9, whose purpose is to shield the Tracker from back-scattered particles from the CE, and thermal screen of the CE, labelled 12, follow the ETL. The independent thermal screen of the ETL, labelled 1, is on the left.

shown as orange bars, which read out the signals from the front end ASICs and bring low voltage power for the ASICs and bias voltage for the LGADs. The LGADs on one face of a disk line up with the service hybrids of the other face so that the whole disk is covered with active elements. Figure 1.10 shows schematically the data path from the sensors, through the ETROCs and service hybrid to the backend electronics in the CMS USC.

1.4.3 The clock distribution system

A stable, low-jitter clock distribution network at the sampling clock frequency and synchronized to the LHC bunch-crossings is required to retain the resolution of the timing detectors. A comprehensive R&D effort is ongoing to achieve an rms jitter of 10–15 ps, including short-term and long-term detector-wide (link-to-link) stability.

In tests using a simplified distribution tree, the current Versatile Link (VL) framework with GBTx and VTRx has shown that it is capable of providing a sub-10 ps jitter LHC clock to endpoints. However, a more complete clock distribution network, with many more boards per crate and multiple crates, sending signals over a much wider area may deliver a less stable clock

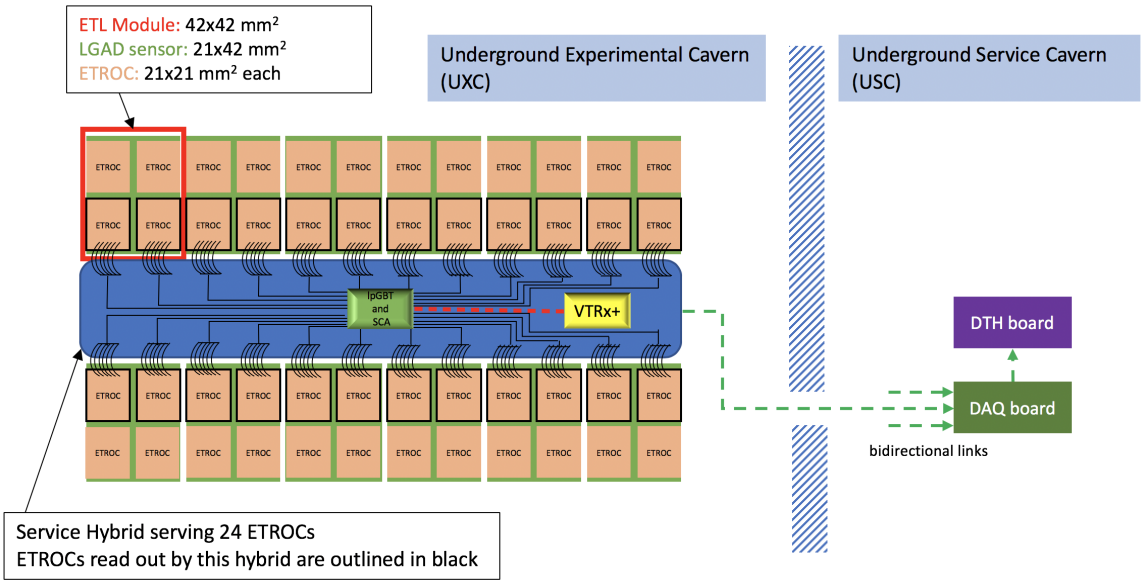


Figure 1.10: Overview of the ETL detector focusing on the data path. The diagram shows the signal flow from the LGAD sensors through the ETROC chips, the data transmission through the service hybrid (shown as blue bar) , the connection to the central DAQ via its bidirectional links, and the connection to the DAQ and DTH board.

with higher rms jitter at the front end ASICs. In addition, the impact of radiation on the lpGBTs may also result in increased jitter. Therefore, components of the current clock distribution network within the Trigger and Clock Distribution System (TCDS) and VL framework are being extensively characterized to define possible improvements for future clock distribution systems in CMS.

Two alternative schemes are being investigated and developed to achieve a clock distribution network with a jitter that is low enough so that it does not contribute significantly to the overall time resolution of the MTD and other precision timing systems in CMS:

- In the baseline approach, the upgraded VL+ framework with lpGBT and TCDS2 as well as the Passive Optical Network components [39] are expected to provide the necessary performance. An encoded clock will be transmitted to the on-detector readout electronics together with the fast-control and trigger signal via the bidirectional DAQ links. The phase locked loop (PLL) in the lpGBT is expected to filter the high frequency noise, while lower frequency jitter and phase instabilities will require dedicated monitoring and correction.
- If the baseline clock distribution scheme fails to meet the precision timing requirements, an independent high precision clock distribution path will be pursued. In this scheme, the sampling clock will be distributed directly to the detector readout electronics by a dedicated precision clock distribution node in the back-end electronics and an on-detector radiation-hard clock fan-out ASIC, which is currently being specified.

The MTD design and the front-end boards are devised with a sufficient number of dedicated optical links to accommodate either option.

The initial alignment of phase offsets from channel to channel and their slow variations over time will be determined from minimum bias events. The method is described in detail in

Appendix C and can be performed with good precision using only tens of seconds of collisions.

1.4.4 Supporting Systems

Many other systems besides the precision clock are required to configure, control, monitor, and read out the BTL and the ETL. These include power for the digital electronics and for the front-end photodetectors; power delivery systems such as DC-DC converters and filters; I/O chips, in particular the GBT chip set and Versatile Link hardware designed for use in LHC experiments by CERN; the back-end electronics and interface to the CMS DAQ, implemented on an Advanced Telecommunications Computing Architecture (ATCA) platform; the Detector Control System (DCS); and the Detector Safety System (DSS). Most of these are common to the BTL and ETL and they are described in the Chapter 4 on Common Systems.

One major component that is common to the CMS Tracker, CE, and the MTD is the Cooling System, designed to provide temperatures of at least $-30\,^{\circ}\mathrm{C}$, based on a two-phase system using CO_2. This system is provided by CERN and will also be used by other CERN experiments. It is described only briefly in this TDR but the cooling requirements of the BTL and ETL are presented in detail in their individual chapters and summarized in Chapter 4.

1.4.5 Possible use of the MTD in the Level 1 Trigger

The use of the information from the MTD may enable CMS to trigger more efficiently on long-lived particles and reduce trigger rates on some topologies. While the participation of the MTD in the L1-Trigger decision is not included in the MTD baseline design, this option is being considered within the collaboration. A thorough evaluation and a positive acceptance of a Level-1 timing trigger proposal can only be made within the context of the CMS trigger TDR studies (Q1 2020), including considerations on the physics case, the optimization of bandwidth allocations, the impact on the Level-1 Trigger System, links, and back-end boards. In this TDR, we include a discussion of the technical changes that would be needed to the MTD hardware to keep open the possibility for use in the Level-1 Trigger. We assess the risks, the decision points and the cost impacts of the changes that would be required in the front-end, in order to inform future decisions, should a Level-1 timing trigger proposal become mature. While full readout of the MTD data into the Level-1 Trigger system is not possible because of bandwidth limitations, it may nevertheless be possible for the MTD to play a role. The readout of MTD data can be seeded by a Level-0 request from the Tracker, Muon or the Calorimeter triggers and the amount of data transferred upon a Level-0 signal can be reduced by transferring only data from a region of interest (ROI) based on the Level-1 seed. The Level-0 Trigger would be formed only after the partial information from the Level-1 system became available, which could have a latency as long as 6 μs, if information from the Level-1 Track Trigger is used but could come sooner if only the muon detector and the calorimetry are used. The amount of data must be controlled to meet the bandwidth capabilities of the ASIC and other front end processing cards. The time to extract and deliver the MTD data for use in the late stages of the formation of Level-1 would still be adequate given the 12.5 μs overall Level-1 latency. The MTD Level-1 trigger information can be formed in the off-detector electronics from the data from the ROIs.

The physics goals and the concept of the use of the MTD in Level-1 triggering will be discussed in Chapter 4. Because the Trigger TDR is still under development, it is not yet possible to specify the whole MTD trigger path and its design, but its electronics is housed off-detector in the USC and the design can be finalized after the Trigger TDR is completed. To maintain the possibility for the MTD to participate in the Level-1 trigger, the capability to process a Level-0 accept should be included in the frontend ASICs. To minimize risks, this feature should be

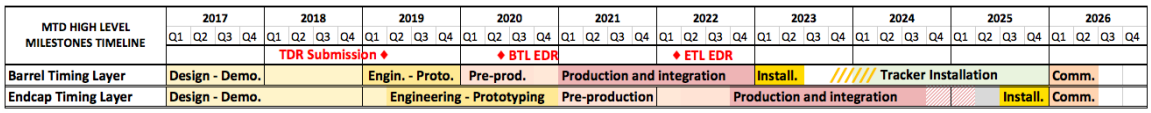


Figure 1.11: Timeline of the MTD project, showing the BTL and ETL detector construction phases over time though installation and commissioning in 2026.

integrated in the second-to-last iterations of the ASICs, which are scheduled for submission in November 2019 for BTL and at the end of 2020 for ETL. For this reason, the Level-0 accept logic is included in the BTL baseline plan, while it will be included in the ETL plan on the condition that a Level-1 timing trigger proposal will be presented with the CMS Trigger TDR at the beginning of 2020. Because of bandwidth and power limitations, constraints should placed on the Level-0 trigger that will limit the amount of Level-0 data that can be delivered and the type of selections that can be implemented. A Level-0-accept rate of the same order of the Level-1-accept trigger can be implemented and it would double the bandwidth required, the number links, fibres, and associated costs. This decision can be postponed to after the Level-1 Trigger TDR. The physics case for using the MTD in the Level-1 Trigger must be made and must be based on a hardware design that satisfies all of these constraints. Some recent results are shown in Appendix E.

1.5 Guide to the remainder of the TDR

This introduction has provided an overview of the MTD project, its motivation and scientific goals, the technical requirements and constraints it must satisfy, and a brief presentation of the proposed detector. The remainder of the report presents in detail the design of the detector and the demonstration of its physics potential. Chapters 2 and 3 present, respectively, the design of the BTL and ETL. Chapter 4 describes electronics components and supporting systems common to both, including the data acquisition system and data links; the clock distribution system; Level 1 Trigger options; systems for Detector Control (DCS) and Detector Safety (DSS); and the cooling system. Chapter 5 presents the current status of our simulation studies of the use of precision timing in track and vertex reconstruction, particle isolation, jet and $p_{\mathrm{T}}^{\mathrm{miss}}$ reconstruction, and benchmark physics measurements and searches. The cost, schedule, and project organization are given in Chapter 6.

A high-level schedule for the construction of the MTD is given in Fig. 1.11, which shows the timeline for the BTL and ETL. For the BTL, installation into the TST must occur not later than the end of 2023 and the beginning of 2024 so as not to interfere with the Tracker installation. However, the schedule has the installation slated to begin in January 2023. Both options are shown in the figure. The ETL can be installed on the nose of the CE either on the surface or when it is installed in the experimental cavern so it can be installed somewhat later. Some important milestones are shown, namely the submission of this TDR and the Engineering Design Reviews (EDR) for the BTL and ETL separately. The various phases of the project are also shown, including design, engineering, prototyping, pre-production, production, integration, and installation. For the BTL, the Tracker installation and integration period, which has a major impact on the schedule, is also shown.

There are five appendices giving additional technical details, first on the radiation environment for the MTD at the HL-LHC; then the BTL; the ETL; the alignment, time synchronization, and monitoring for both the BTL and ETL; and finally a brief discussion of performance of the proposed Level 1 Timing Trigger. At the end of the TDR, there is an extensive glossary of terms used in the document.

Chapter 2

The Barrel Timing Layer

2.1 Overview and principle of operation

The Barrel Timing Layer is a cylindrical detector with a surface of about 38 m^2. It is designed to detect MIPs with time resolution of 30 ps at the beginning of HL-LHC operation and a luminosity-weighted resolution of 40–50 ps. The BTL is located between the ECAL and the Tracker with an inner radius of 1148 mm and and outer radius of 1188 mm, as shown in Fig. 2.1. It provides a coverage up to $|\eta| = 1.48$ with an acceptance for muons of $p_T > 0.7$ GeV of about 90%, limited by the supporting rails of the Tracker and small dead areas between sensor modules.

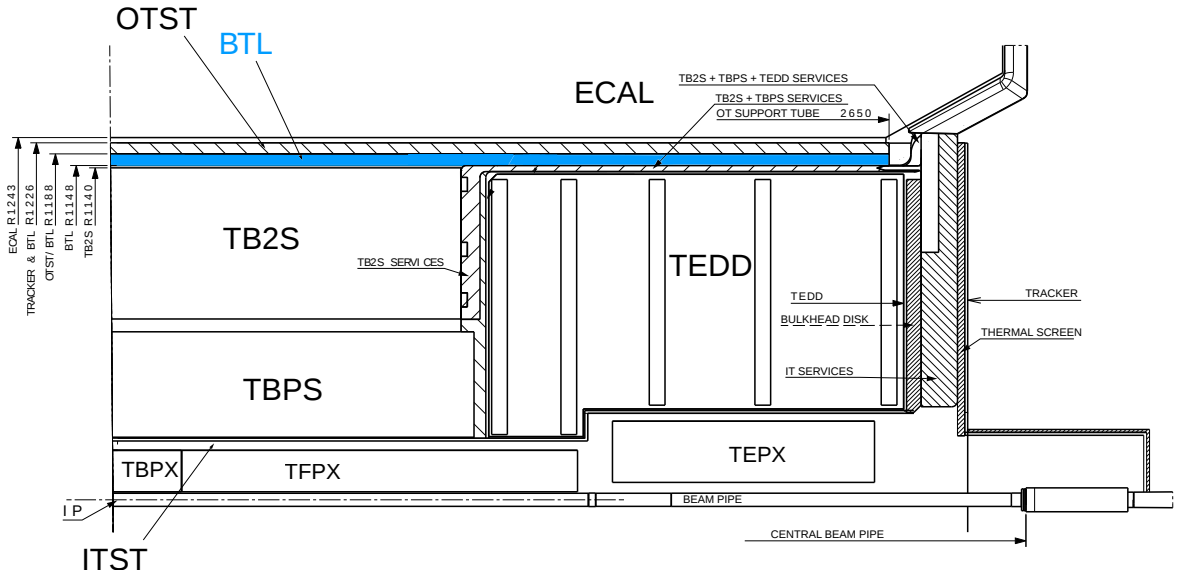


Figure 2.1: BTL integration within CMS.

The sensitive elements consist of Lutetium Yttrium Orthosilicate crystals doped with Cerium ($(Lu_{1-x}Y_x)_2SiO_5$:Ce, abbreviated as LYSO:Ce). The crystals are shaped in a bar-like geometry of 57 mm length, 3.12 mm width and an average thickness of 3 mm read out by a pair of SiPMs, one at each end, matching the size of the crystal end face for optimal light collection. The choice of this crystal geometry minimizes the SiPM area with respect to the crystal sensitive volume, limiting power consumption and channel count, without loss of light collection efficiency. This is enabled by the light propagating in total internal reflection mode within crystals of such high aspect ratio. This sensor layout exploiting double-ended readout also provides uniform time response across the surface, tracking capabilities and redundancy of time measurements per

Table 2.1: Summary of the BTL modularity and channel count. The number of items in each module, readout unit and tray are shown.

	Module	RU	Tray	Total
Channels (SiPMs)	32	768	4608	331776
Crystals	16	384	2304	165888
ASICs	1	24	144	10368
Modules	-	24	144	10368
Readout units (RU)	-	-	6	432
Trays	-	-	-	72

Table 2.2: Optimization of BTL sensors as a function of η. Crystal thickness and SiPM area adapted to maintain uniform time resolution and minimize crystal volume and silicon area.

$\lvert\eta\rvert$ **region**	**0–0.7**	**0.7–1.1**	**1.1–1.48**
Readout unit ID within tray	1–2	3–4	5–6
Crystal thickness, t [mm]	3.75	3.0	2.4
$\langle t_{\mathbf{slant}} \rangle$ **[mm]**	4.0	4.3	4.6
SiPM active area [mm^2]	11.2	9.0	7.2
$\langle \Phi^{\mathbf{tot}}_{\mathbf{neq}}(\mathbf{3000\ fb^{-1}}) \rangle$ **[cm^{-2}]**	1.65×10^{14}	1.75×10^{14}	1.85×10^{14}

crystal. The area of a single crystal represents the optimal trade off between channel count and sensor performance. In particular, the chosen granularity provides an average occupancy of the detector cells at pileup 200 of about 7%, thus limiting the probability of double hits within the same cell during a bunch crossing. Such occupancy also maintains a negligible impact of pileup effects from previous bunch crossings on time resolution, considering the scintillation decay time.

As illustrated in Fig. 1.7, the longitudinal axis of the crystal bars is oriented along the ϕ direction within CMS. Crystals are grouped in *modules* of 1×16 ($\phi \times z$) each involving 32 SiPMs and thus 32 readout channels, covering a rectangular area of about 60×52 mm^2. Such modules are arranged in a matrix of 3×8 to constitute one *readout unit* with an approximate size of 184×413 mm^2. Six of such readout units are then used to fill a *tray* covering the half length of the BTL cylinder, so that 2 trays cover the full length of 2480 mm. A total of 72 trays allows to instrument the entire surface of the detector. The total channel count (i.e. number of SiPMs) is thus 331 776 and the number of crystal bars is 165 888. The BTL geometry and modularity was previously illustrated in Fig. 1.7 and is summarized in Table 2.1. Further mechanical and integration details are presented in Section 2.4.

The thickness of the crystals along the detector z axis is optimized to limit the amount of material in front of the CMS ECAL to be as small and uniform as possible ($<0.4X_0$, where X_0 is one radiation length) so as to have a negligible impact on the ECAL energy resolution. In particular, three detector regions are defined as a function of the η range featuring a crystal thickness decreasing from 3.75 to 3.0 and 2.4 mm. Similarly the SiPM active area will be reduced from about 11.2 to 9 and 7.2 mm^2. This minimizes the rate of noise counts in the SiPMs (which scales with the SiPM active area) while maintaining the maximum light extraction from the crystal and compensates the small, $\sim$20%, non-uniformity in radiation level expected within the BTL pseudo-rapidity range, as shown in Fig. 2.2. Table 2.2 summarizes the optimization of the sensor specifications for each of these regions, illustrating the leveling of performance achieved with this approach.

As stated above, the fundamental detecting element of the BTL will be a thin LYSO:Ce crystal

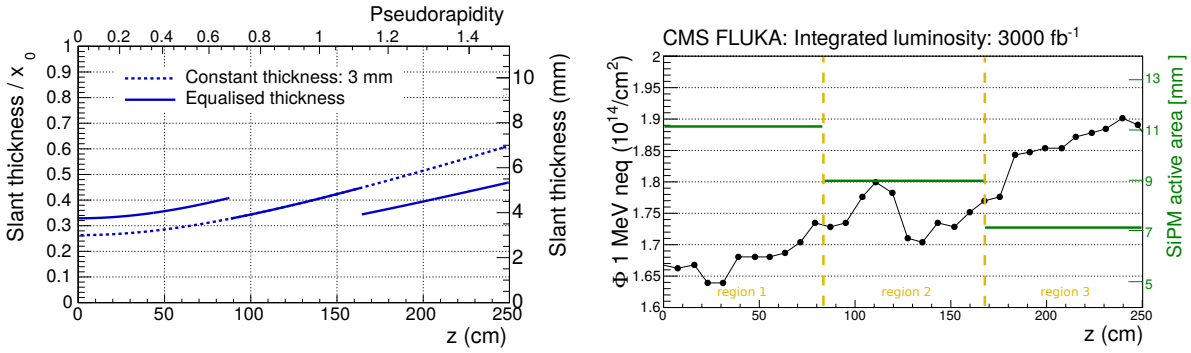


Figure 2.2: BTL layout parameters along detector axis z: slant thickness and radiation length (left), SiPM area and radiation levels (right).

bar coupled to a pair of SiPMs. A minimum ionizing particle traversing the crystal volume will produce a number of optical photons along its track proportional to the crystal light yield (LY) defined as the number of photons generated per MeV of energy deposit. A fraction of the photons will be detected at each SiPM. Detected photons will be converted to photoelectrons and amplified by the SiPM, operated with a gain of O(10^5), to generate an electrical signal that can be discriminated and digitized to obtain a measurement of the time at which the MIP crossed the detector, referred to as the "time stamp". Along this detection chain several effects can introduce stochastic and systematic fluctuations that lead to a degradation of the detector time resolution. The time resolution per track, from the combination of two independent measurements at the two ends of the crystal with a common clock jitter, is given by the sum in quadrature of the following terms:

- CMS clock distribution: 15 ps;
- Digitization: 7 ps;
- Electronics: 8 ps;
- Photo-statistics: 25–30 ps;
- Noise (SiPM dark counts): negligible at startup, 50 ps after 3000 fb^{-1};

summarized in the equation:

$$\sigma_t^{\mathrm{BTL}} = \sigma_t^{\mathrm{clock}} \oplus \sigma_t^{\mathrm{digi}} \oplus \sigma_t^{\mathrm{ele}} \oplus \sigma_t^{\mathrm{phot}} \oplus \sigma_t^{\mathrm{DCR}} \,. \tag{2.1}$$

Each of these terms is discussed in more detail in their respective paragraphs, and their relative contributions to the overall time resolution are summarized in Fig. 2.3. Time jitter from the electronics and time digitization effects have a negligible impact on the overall time resolution.

The timing performance drivers are the photo-statistics and the noise term, thus major R&D efforts have been spent on their optimization. The contribution from photo-statistics is related to the stochastic fluctuations in the time-of-arrival of photons detected at the SiPM, and its scaling with respect to key BTL parameters is summarized by the equation:

$$\sigma_t^{\mathrm{phot}} \propto \sqrt{\frac{\tau_r \tau_d}{N_{\mathrm{phe}}}} \propto \sqrt{\frac{\tau_r \tau_d}{E_{\mathrm{dep}} \cdot \mathrm{LY} \cdot \mathrm{LCE} \cdot \mathrm{PDE}}} \,, \tag{2.2}$$

where τ_r and τ_d are respectively the rise time and decay time of the scintillation pulse which for LYSO:Ce are about 100 ps and 43 ns respectively. The energy deposited by a MIP in a thin

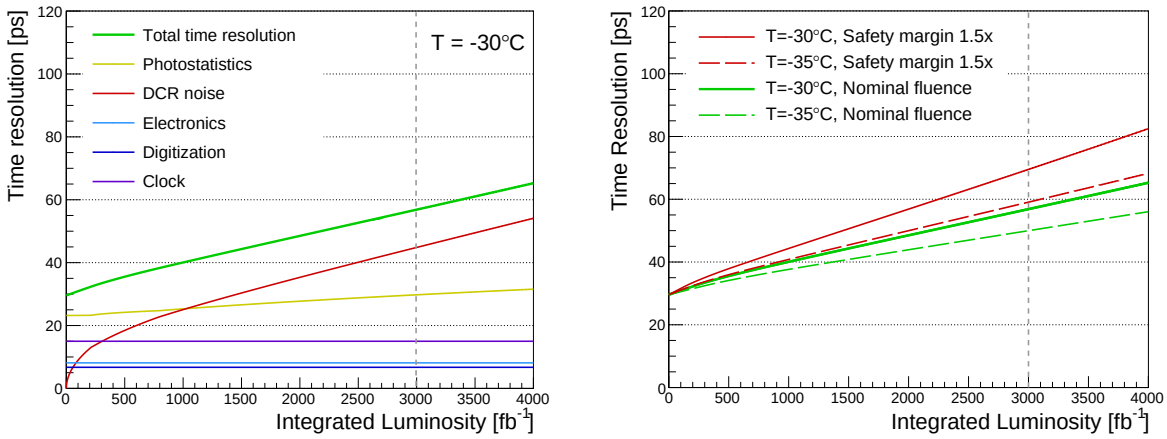


Figure 2.3: Left: Evolution of different terms contributing to the BTL time resolution as a function of integrated luminosity. The two time measurements from the SiPMs at the opposite ends of a LYSO:Ce crystal bar are combined in a single measurement. The curves are calculated for the SiPM type HDR2-015 from Hamamatsu. Right: Comparison of the evolution of BTL time resolution at different temperatures for the nominal radiation level and for a safety margin of 1.5. The performance degradation caused by an increase of the 1 MeV neutron equivalent fluence of a factor 1.5 can be offset by lowering by 5 °C the operating temperature.

LYSO:Ce crystal, E_{dep}, features a Landau distribution with the most probable value (MPV) of 0.86 MeV/mm. The number of photoelectrons, N_{phe}, scales linearly with the energy deposited and the crystal LY which are determined by the crystal thickness and scintillation properties. It also scales linearly with the light collection efficiency (LCE), i.e. the probability that a photon reaches the SiPM without escaping from lateral faces or being absorbed within the material and with the Photon Detection Efficiency (PDE) of the SiPM. These parameters have driven the optimization of the sensor layout (crystal and SiPM configuration). In the BTL crystals, a MIP deposits an average energy of 4.2 MeV including the path length for bending tracks within the LYSO:Ce volume. With a LCE of 15% and PDE of 20%, a total signal of about 5100 photoelectrons at each SiPM is expected for a MIP.

The contribution due to the noise term scales with the dark count rate (DCR) in the SiPM proportionally to $\sqrt{\text{DCR}}/\text{N}_{\text{phe}}$. The magnitude of the DCR increases with integrated luminosity due to radiation damage creating defects in the silicon, and depends on several factors discussed in more detail in Section 2.2.2, including the operating temperature, the annealing scenario during shutdowns, and the specific SiPM technology.

The breakdown voltage of the SiPM, V_{br}, is defined as the bias voltage that leads to self-sustaining avalanche multiplication and is thus the minimum voltage required to properly operate the photodetector. Since the V_{br} can vary slightly in different devices and during the detector lifetime (because of radiation effects) the relevant parameter used in the following to define the SiPM performance is the over-voltage, $\text{OV} = \text{V}_{\text{bias}} - \text{V}_{\text{br}}$, i.e. the voltage difference between the applied bias voltage and V_{br}.

Both PDE and DCR increase with the OV, showing a SiPM-dependent behavior presented in Section 2.2.2. Therefore, the operating OV of the SiPM will be adjusted during the detector lifetime within a range of about 3.5 V, to maintain the optimum time resolution. In particular the over-voltage will be decreased gradually from 3.5 V to about 1.2 V to maintain the DCR within an acceptable level of 35–55 GHz (SiPM dependent). Lowering the over-voltage will also cause the PDE to decrease from about 38–27% down to 24–13% (SiPM dependent). Both

parameters will determine the evolution of the detector performance at the optimum operating voltage as shown in Fig. 2.3.

Uncertainties related to the sensor properties, ASIC performance and radiation levels are discussed in Appendices A and B. To account for these uncertainties a safety margin of 1.5 is used on top of the nominal radiation level predicted for 3000 fb^{-1} to qualify the radiation tolerance of the sensors. The chosen safety factor is such to allow the SiPMs to be operated close to the optimal over-voltage without exceeding the power budget, thus in a regime where the time resolution degrades linearly with the integrated luminosity. Two possible approaches to reduce the magnitude of DCR, dominating the timing performance, are discussed in Appendix B. One is shown in the right plot in Fig. 2.3 how the performance degradation caused by an increase of the 1 MeV neutron equivalent fluence of a factor 1.5 can be offset by lowering by 5 °C the operating temperature.

2.1.1 BTL sensors performance in test beam

BTL sensor prototypes have been tested and characterized at the test beam facilities of CERN and Fermilab. The test beam facilities provide high energy pions or protons which serve as a well calibrated source of minimum ionizing particles. A Micro Channel Plate (MCP) with time resolution of about 16 ps has been used as timing reference. More details on the test beam instrumentation are reported in B.II. Crystal bars with dimensions of $3 \times 3 \times 50$ mm^3 have been instrumented with 3×3 mm^2 SiPMs of 15 μm cell pitch from Hamamatsu (S12572), which is one of the BTL SiPM candidates discussed in Section 2.2.2. The SiPMs were optically coupled to the crystal using Meltmount glue of refractive index n = 1.58 and the crystals were wrapped with Teflon. Similar optical conditions will be used in the detector elements as discussed later in Section 2.2.

The spectrum of energy deposited by a MIP has a Landau shape due to stochastic fluctuations within a thin crystal, as shown for example in Fig. 2.4. The signal amplitude at a single crystal end (left and right with respect to the incoming beam at normal incidence) has a maximum 25% variation along the entire crystal length as shown in Fig. 2.4. The total light extracted along the crystal, however, is rather constant (better than 10%), since it is given by the sum of left and right and this results in a time resolution that is uniform along the crystal, since global fluctuations from photo-statistics are of the same magnitude.

Since a fixed discrimination threshold is used to extract the time stamp, a time correction for amplitude variations (time-walk) is required to achieve the optimal time resolution. A typical curve parameterizing this effect is shown in Fig. 2.5. The contribution to the time resolution from the time-walk correction is below 10 ps for a precision on the amplitude measurement of better than 5%, as shown in the right plot of Fig. 2.5. Both the left and right SiPM time stamps are thus corrected for time-walk before being averaged to obtain the global time stamp.

The time stamp obtained from each SiPM exhibits a position dependence behavior, illustrated in Fig. 2.6, from the propagation time of optical photons within the crystal. The average of the two time stamps, $t_{\text{ave}} = (t_{\text{left}} + t_{\text{right}})/2$, provides a uniform time stamp along the entire bar, as shown in Fig. 2.6, and is thus the correct estimator of the time of arrival of the MIP, providing the optimal time resolution. The difference between the two time stamps, $t_{\text{diff}} = t_{\text{left}} - t_{\text{right}}$, represents a variable strongly correlated with the impact point of the MIP along the length of the bar and thus provides tracking information with a spatial resolution of about 3.5 mm for a time resolution of 30 ps. The spatial resolution depends on the coefficient of the linear curve, k_{slope}, in Fig. 2.6, as $\sigma_{\text{x}} = 2\sigma_{\text{t,ave}}/k_{\text{slope}}$. The slope of this curve, about $k_{\text{slope}} = 17$ ps/mm, is consistent with twice the reciprocal of the light speed in LYSO:Ce ($1/v = n/c \sim 6.1$ ps/mm,

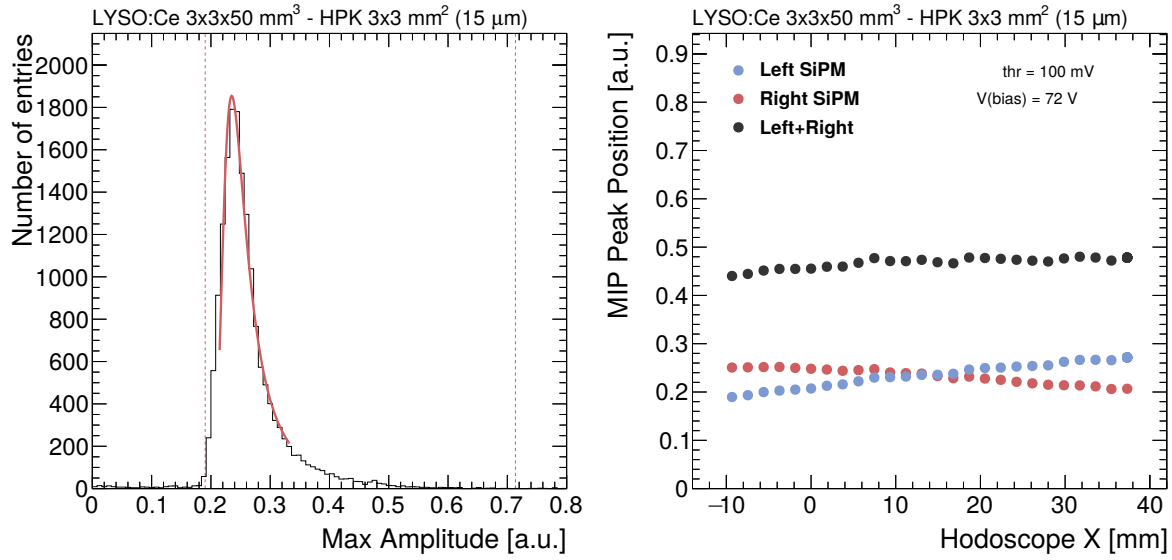


Figure 2.4: Left: measured amplitude distribution for a MIP in a 3 mm thick crystal bar and Landau fit. The red dotted lines represent the selection on the pulse amplitude range applied in the data analysis. Right: amplitude response variation along a crystal for the left and right SiPM separately and for the sum of both.

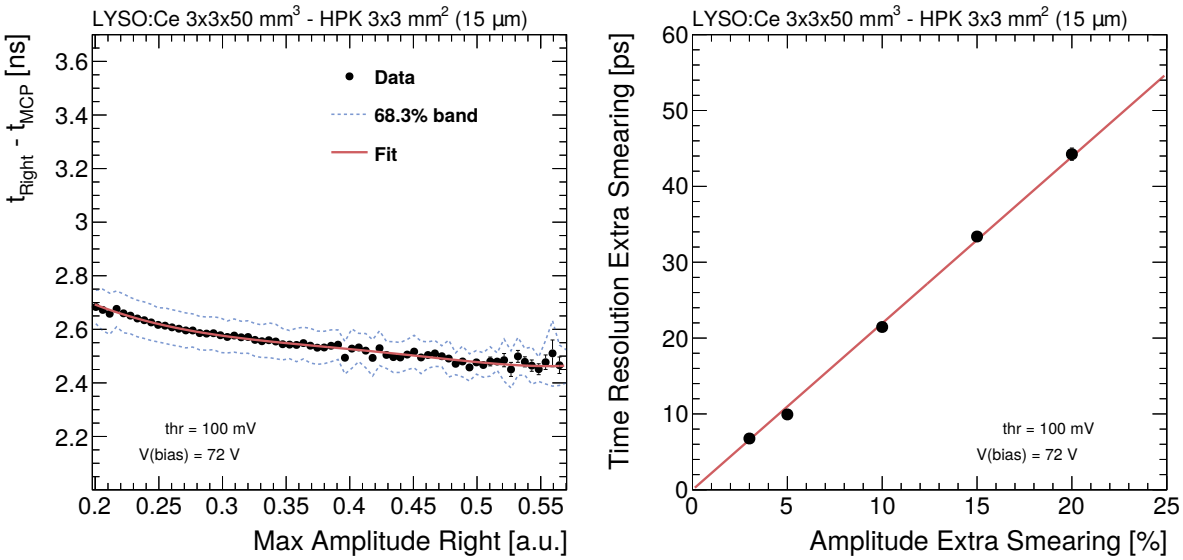


Figure 2.5: Left: the difference of the time stamp from the SiPM and the reference time provided by the MCP is shown as function of the SiPM signal amplitude and defines the typical time-walk correction curve. Right: a precision on the amplitude measurement better than 5% is sufficient to have an efficient time-walk correction (contributing to less than 10 ps in quadrature).

with $n = 1.82$), indicating that photons with a quasi-direct path to the SiPM, i.e. within the cone of total internal reflection, determine the time response. This cone is defined by a maximum angle of 56.5° with respect to the normal to the crystal end face.

The time resolution along the bar as measured in the test beam is reported in Fig. 2.7 for the two SiPMs individually and for the average time stamp. The combination of the two SiPM measurements improves the overall time resolution by a factor $\sqrt{2}$ since the dominant stochastic

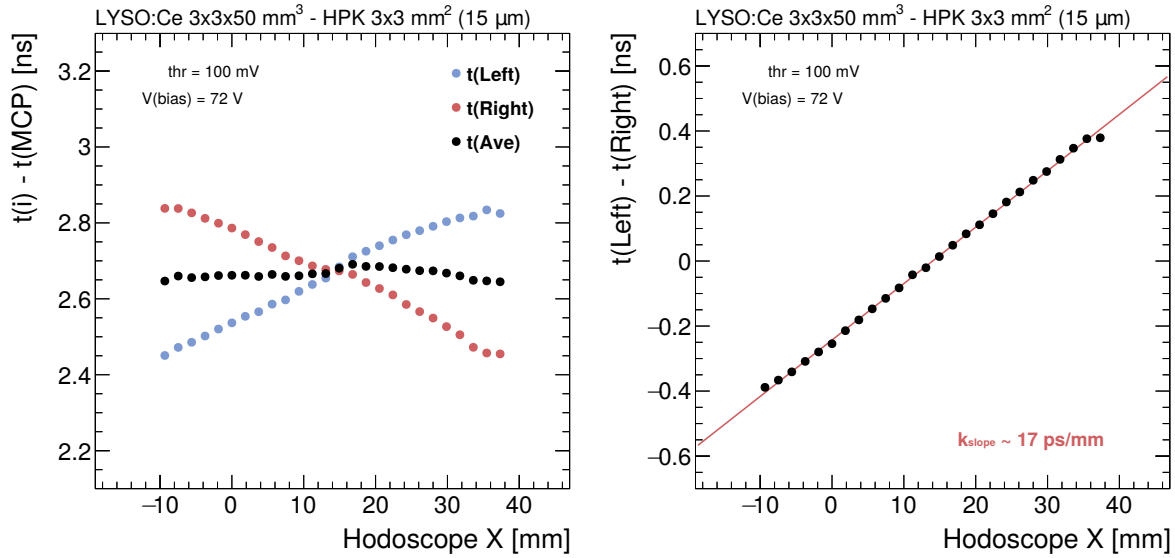


Figure 2.6: Left: time stamp from the left and right SiPMs and average time stamp, t_{ave}, as a function of the impact point X along the crystal bar axis. Right: difference between the two time stamps, t_{diff}.

fluctuations from photo-statistics and DCR are uncorrelated for the two SiPMs.

For a bias voltage of 72 V corresponding to a PDE of 37% and a discrimination threshold of 100 mV (equivalent to about 10 photoelectrons), a time resolution of 43 ps at each SiPM and 30 ps for the combined time stamp is achieved according to expectations. A scan of bias voltage and discrimination threshold is shown in the right plot of Fig. 2.7. At lower OV the time resolution degrades with the expected behavior of $1/\sqrt{\text{PDE}}$ down to 38 ps for a bias voltage of 69 V corresponding to about 2.7 V over-voltage where the PDE is about 22%. The threshold scan in the range 60–500 mV shows that above the threshold of 100 mV the time resolution is approximately constant for up to a factor 5 larger threshold. Below 100 mV for small signals (69 V) the noise from the electronics starts to deteriorate the time resolution.

As a large fraction of particles produced by LHC collisions at CMS will impact the crystal bar at non-normal incidence, we measured the dependence of the time resolution on the MIP impact angle, θ, with respect to the normal to the bar axis. The time resolution measured as a function of the slant thickness, $t_{\text{slant}} = t/\cos\theta$, is shown in Fig. 2.8 in which the angles used were $\theta_{\text{MIP}} = 0,\ 45,\ 60$ and 80°. This set of angles spans the entire range of slant thicknesses expected for both MIPs in the high-η region of the barrel ($\theta^{z}_{\text{max}} \sim 64^\circ$) as well as low p_{T} ($\in [0.7 - 2.0]$ GeV) charged particles that are strongly bent by the magnetic field and can thus cross the crystal with an angle up to $\theta^{\phi}_{\text{max}} \sim 80^\circ$. While the energy deposit increases linearly with the slant thickness, t_{slant}, the time resolution improves as $\sigma_{\text{t}}^{\text{slant}} \propto t_{\text{slant}}^{-\alpha}$ with $\alpha \sim 0.35$. This behavior can be explained as the combination of a larger number of photons, which improves the stochastic fluctuations with the square root behavior, and the fact that the photons are produced across a longer track thus leading to a slower the rise time of the pulse. From this consideration, a value of $\alpha < 0.5$ is expected. There is a small asymmetry in the performance of the two measurements depending on their position with respect to the MIP direction. In particular the downstream measurement performs better than the upstream one, since in the former case the light signal is compressed by the time of flight of the charged particles, while in the latter it is dilated. Nevertheless, the combination of the two time stamps improves the overall time resolution with respect to a normal incidence.

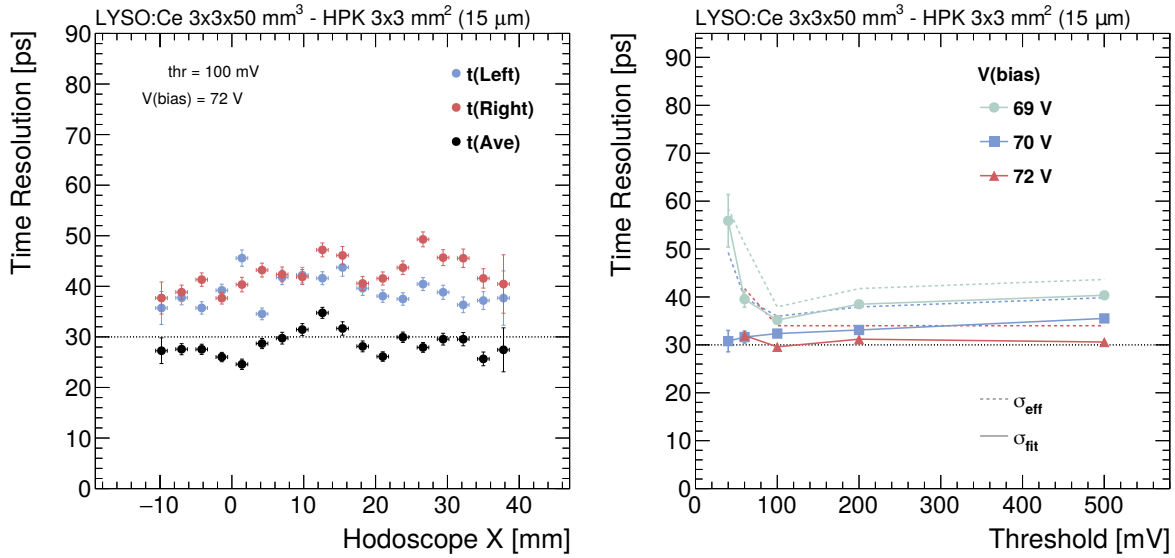


Figure 2.7: Left: time resolution for the left and right SiPMs and average time stamp t_{ave} as a function of the impact point X along the crystal bar axis. Right: sensor time resolution, t_{ave} for different bias voltages and discrimination thresholds.

Since in the BTL design the bar axis is oriented along the ϕ direction, it follows that low p_{T} tracks will be mostly contained within a single bar, while for increased η the signal will be shared across adjacent bars. This provides an overall better timing performance for the majority of the charged tracks in BTL that are expected to have low p_{T}. When the signal is shared across adjacent bars as shown in the drawing of Fig. 2.8, the time stamp from each bar can be combined with a weight proportional to $1/\sqrt{E_{\text{dep}}}$, where E_{dep} is the energy deposited in a single bar. In this manner, the overall optimal time resolution is achieved. In the right plot of Fig. 2.8 we report the results obtained in test beam showing that a MIP signal, with $E_{\text{tot}} = \sum_i E_i$, shared between N bars (with $1 \leq N \leq 3$), can be effectively combined according to the generalized formula

$$t_{\text{comb}} = \frac{\sum_i w_i t_i}{\sum_i w_i}, \tag{2.3}$$

where the weights w_i are defined by the fraction of the MIP energy deposited in each bar, E_i.

In part of this study, the particles were impinging with an angle of 45° on two adjacent crystal bars. In this configuration, the MIP deposits only a fraction of its energy in either one bar only or in both, depending on the impact point. The result is shown in the lower right-hand plot of Fig. 2.8. The MIP signal is shared between the adjacent crystals for impact point positions between 19.5 and 20.5 mm. Within this range, the signal is combined with a simple average and with an energy-weighted average yielding slightly better results. The time resolution achieved is equivalent to that of a MIP track fully contained in a single crystal (lowest part of blue and green curves). Data points at Y $<$ 18 or Y $>$ 22 mm, where the time resolution degrades for MIP tracks only partially crossing either crystal, should be disregarded.

2.1.2 Mitigation of DCR impact on time resolution

A dark count in the SiPM corresponds to a single cell firing due to a thermally generated electron that initiates an avalanche in the high field region. The rate of dark counts is proportional

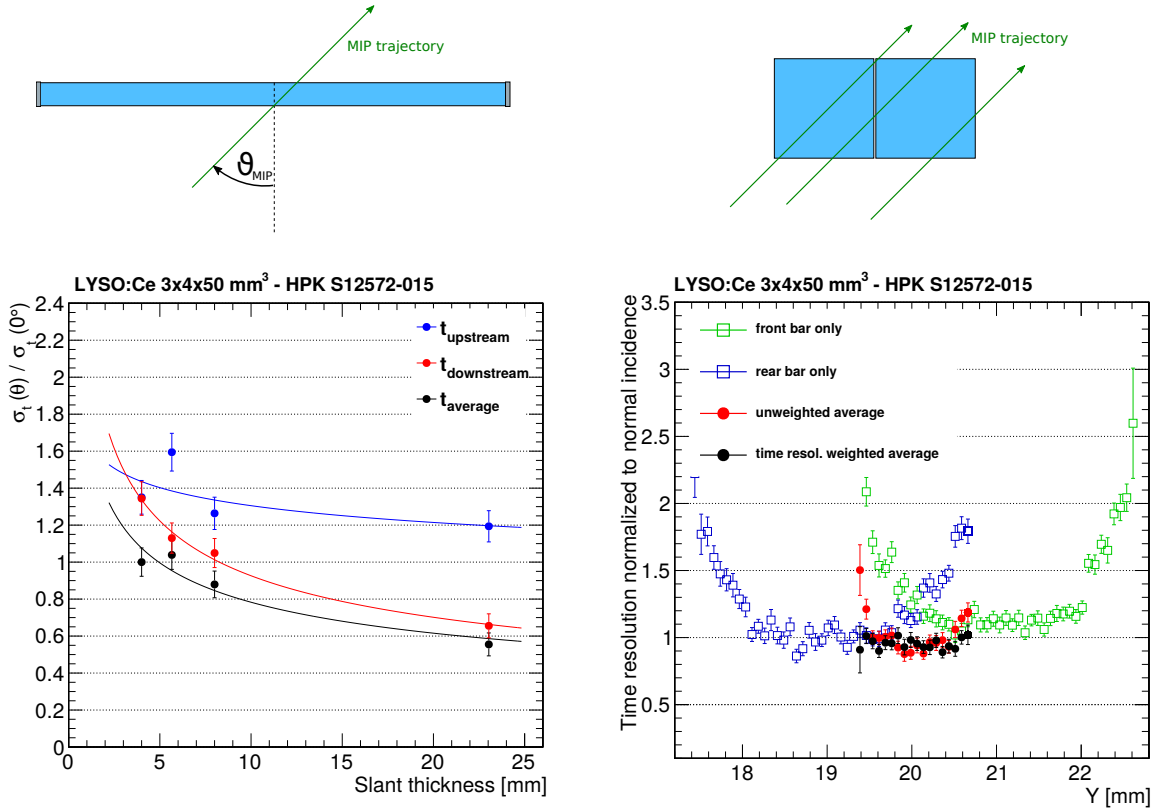


Figure 2.8: Left: the time resolution of the sensor (relative to normal incidence) is shown as a function of the slant thickness obtained with a bar tilted at angles of 0, 45, 60 and 80°. The performance of the upstream and downstream SiPM and their combination are shown separately. The outlying point for the 45° run in the upstream channels is possibly due to noise pick up on the PCB used for that specific angle. Right: the performance of two adjacent bars (relative to normal incidence) is shown as a function of impact point, Y, of the MIP having an incident angle of 45°. Each position features a different energy deposit, E_i, in each bar. The response of individual bars and their combination by weighting the two time stamps (one combined time from each bar and thus 4 SiPMs in total) is shown.

to the active area of the SiPM and decreases at lower temperatures by about a factor two every 7–10 °C. Radiation induced defects in the silicon structure increase the probability of generating a thermal electron [40]. Because of the radiation environment in which the barrel timing layer will operate, the DCR at −30 °C will increase up to a level of 35–55 GHz at the end of operation depending on the SiPM technology. Since the signal produced by a thermally generated electron and a photoelectron from LYSO:Ce scintillation cannot be distinguished, the dark counts represent a noise term which overlaps with the true photon signal from a MIP. The random overlap of these single-electron pulses induces a time jitter proportional to the fluctuation in number of electrons due to DCR, i.e. proportional to $\sqrt{\text{DCR}}$.

Since each avalanche within a cell will generate a signal with a rise time of about 100 ps and a recovery time of about 10 ns, a dark count will generate a signal tail such that the measurement of the baseline a few hundreds of ps before the rising edge of the pulse will show a correlation with the noise overlapping in the signal region. This feature can be exploited to mitigate the impact of such local baseline fluctuations on the time resolution as previously demonstrated in Ref. [41]. As discussed in Section 2.3, the front-end electronics is designed to reduce this noise contribution with an architecture similar to the first stage of a constant fraction discriminator

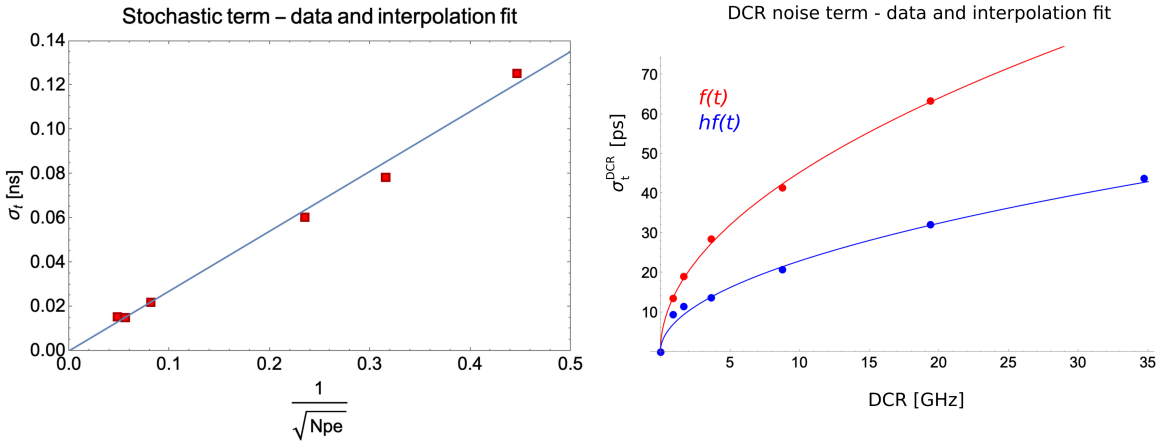


Figure 2.9: Left: time resolution of the SiPM for different signal amplitudes for DCR = 0 GHz. Right: additional time jitter due to DCR up to 35 GHz measured by injecting randomly distributed light on a BTL SiPM (S12572-015C) for a signal of 390 photoelectrons. The beneficial impact of a DLED-like shaping, $hf(t)$, (blue curve) with respect to standard pulse, $f(t)$, (red curve) is shown using a delay δt of 500 ps.

(CFD). The signal, $f(t)$, is processed such that the waveform is inverted and delayed and then summed up with the original pulse resulting in a pulse defined by $hf(t) = f(t) - f(t + \delta t)$. In this way the resulting waveform has smaller baseline fluctuations from either DCR or pileup. The time stamp is then extracted from the pulse using a leading edge discriminator. Such technique can be referred to as a Differential Leading Edge Discrimination (DLED). Details on the performance of the full chain, including the crystal, the SiPM, and the front-end ASIC, are given in Section 2.3.

This strategy has been demonstrated effective in reducing the impact of DCR on time resolution by a factor $\sim$ 2 when a delay, δt, in the range of 200–800 ps is used. This is reported in Fig. 2.9 where a pulsed LED (Light Emitting Diode) is used to generate a photon signal and another LED is used to illuminate the SiPM with photons randomly distributed in time, thus emulating dark count noise. The electronics used for the test was designed to reproduce the signal of the ASIC with discrete components. The LED intensity was first set to a value of about 390 photoelectrons yielding a time resolution, in absence of DCR, of about 25 ps which is the expected photostatistic term in BTL. The additional time jitter due to DCR was then measured by increasing the intensity of the noise from 0 to about 35 GHz. This term scales as expected with the $\sqrt{\text{DCR}}$ and with the inverse of the number of photoelectrons in the pulsed signal, $\propto 1/N_{\text{phe}}$. Simulation of the DLED approach for pulse shaping well reproduces the experimental measurements obtained with the LED and was thus assumed for the extrapolation of BTL performance in Fig. 2.3. Additional details on the validation of the BTL performance evolution model are given in Appendix B.

2.2 Active elements

2.2.1 Scintillating crystals

For precision timing purposes in the BTL environment LYSO:Ce crystals represent an optimal candidate compared to other inorganic scintillators because of their high light yield of about 40 000 photons/MeV, fast scintillation rise time (<100 ps), and relatively short decay time ($\sim$ 40 ns) as measured in Ref. [42]. The figure of merit for precision timing is the num-

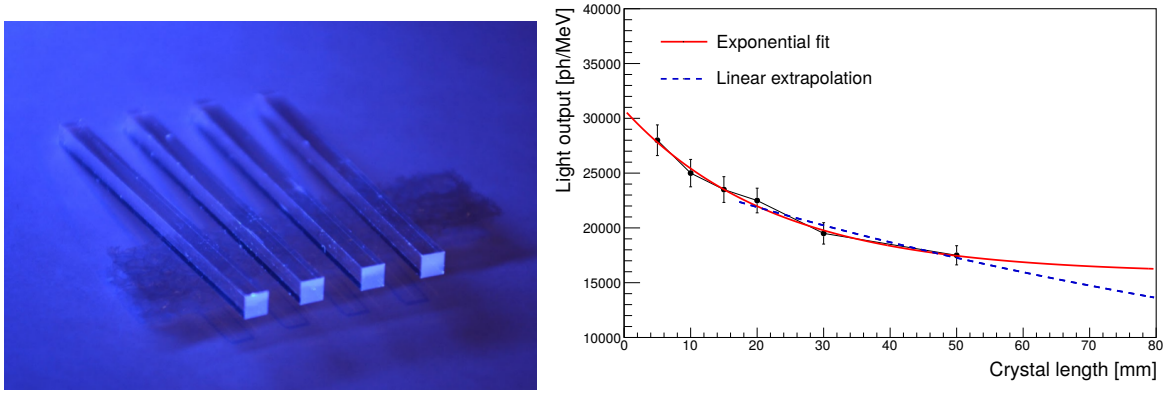


Figure 2.10: Left: picture of four bare LYSO:Ce crystal bars without wrapping under UV illumination. Right: light output measured from a single end of fully Teflon wrapped LYSO:Ce crystal bars of different lengths measured with a ^{137}Cs source and a PMT for readout (the light output value corresponds to the total light extracted at both ends in the BTL sensors).

ber of photons produced within a short time window of about 500 ps from the beginning of the scintillation signal (early photons), which for LYSO:Ce is about 400 photons/MeV. Given high density of LYSO (7.1 g/cm^3), yielding an energy deposit by a MIP with most probable value of about 0.86 MeV/mm, about 2000 "early photons" are produced by a MIP in BTL. The high density of the material has the additional advantage of minimizing the space required by the scintillator, which is important given the limited space available to BTL within the CMS TST. Furthermore, LYSO:Ce is non-hygroscopic, thus avoiding any material effect in case of ambient moisture and it provides scintillation light at a wavelength of 420 nm, matching the sensitive range of the SiPMs. Lastly, LYSO:Ce is a very commonly used commodity in medical imaging applications, for Positron Emission Tomography (PET), making it a well-studied material with an abundance of global qualified vendors capable of providing crystals meeting stringent specifications. Several other alternative scintillators were considered for use in the CMS BTL (plastic scintillators, crystal garnets, etc.), but LYSO:Ce was chosen as the optimum trade-off between performance, radiation tolerance, cost and mass production capability. Fig. 2.10 shows a picture of the crystal geometry considered for the BTL: elongated bars of about $\sim 3 \times 3 \times 57$ mm^3.

As outlined in the introduction of this chapter, the choice of a bar geometry is strategic for the detector optimization since the light collection efficiency in this case relies mostly on optical photons that are collected within the angle of total internal reflection. It follows that the average photon path is close to the distance between the impact point of the MIP and the SiPM and that the light output decreases very slowly for crystal lengths above 50 mm, as shown in Fig. 2.10.

The key feature of LYSO:Ce is its radiation tolerance, which is required for operation without significant loss of transparency or light output in the high radiation environment until the end of HL-LHC operation. The relevant particle fluence connected to radiation damage effects in crystals is the one from charged hadrons (e.g. protons) above 20 MeV that can cause displacements in the crystal lattice. The change of transparency measured on a LYSO:Ce crystal bar irradiated with 24 GeV protons to a fluence of 2.5×10^{13} cm^{-2}, which is above the integrated level expected for BTL including the safety margin, is shown in Fig. 2.11. A negligible loss of transparency, T, is measured along the 50 mm long bar corresponding to an induced absorption coefficient, $\mu_{\text{ind}} = 1/L \cdot \ln(T_{\text{before}}/T_{\text{after}})$ of 0.5 m^{-1} at 420 nm, where L is the length of the crystal probed with the spectrophotometer. To confirm that this change has negligible impact on the performance, the signal amplitude and time resolution of an irradiated bar was com-

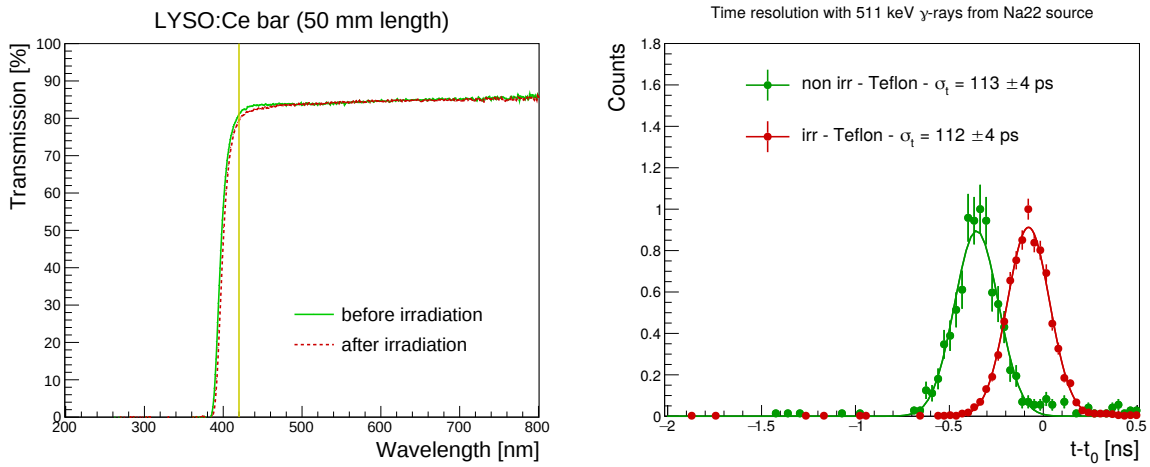


Figure 2.11: Left: transparency curve measured across a 50 mm long LYSO:Ce crystal bar, before and after irradiation with 24 GeV protons to a fluence of 2×10^{13} cm^{-2}. Right: time resolution, measured with 511 keV γ-rays, of a crystal bar before and after irradiation.

pared with those of a non-irradiated bar using a ^{22}Na source in laboratory. The time resolution obtained is the same and the signal uniformity along the bar is mostly unchanged, confirming a negligible attenuation of the light signal. Extensive radiation tolerance studies have been performed on Cerium doped LSO and LYSO crystals produced by a variety of manufacturers. The results consistently showed an excellent radiation tolerance to both hadron fluences and ionizing radiation to levels beyond the requirements of BTL [43–48].

Wrapping of individual crystals within an array is mandatory to provide optical isolation of each channel and avoid light cross-talk from one channel to another. Its absence would otherwise increase the detector occupancy. The choice of the wrapping can also affect the overall light collection efficiency and consequently the time resolution of the sensor. Dedicated laboratory measurements with radioactive sources have been performed to study the influence of different wrapping materials on the LCE and time resolution. The results reported in Fig. 2.12 demonstrate that the impact of the wrapping material is limited as long as an air gap is maintained between the crystal surface and the reflective layer. The light collected at the SiPM is indeed the light within the angle of total internal reflection explaining the limited increase of light output when the crystal is wrapped (except close to the SiPM end). Similarly the time resolution is not significantly affected by the wrapping choice since mainly photons with a short optical path to the SiPM influence the timing capabilities. The baseline choice of wrapping material is thus ESR foils (Enhanced Specular Reflector Vikuiti by 3M), which can provide a thin and radiation tolerant solution matching the specifications for the crystal array.

2.2.1.1 Crystal technical specifications

The LYSO:Ce crystals are mass produced by a large variety of vendors. While the overall properties are usually very similar, some differences from vendor to vendor are observed. Based on the measurements performed on the crystals from a few vendors and the goal of the experiment to optimize the timing performance of the detector, the specifications for the BTL LYSO:Ce crystal have been determined and are presented in Table 2.3. The rms spread in parameters acceptable on large quantities of crystals is also reported in the table when pertinent and needed to maintain the impact on time resolution to less than 5%.

To this extent, a key parameter is the light output measured at each end of the crystal, which should be larger than 6000 photons/MeV. The light output should be measured on a crystal

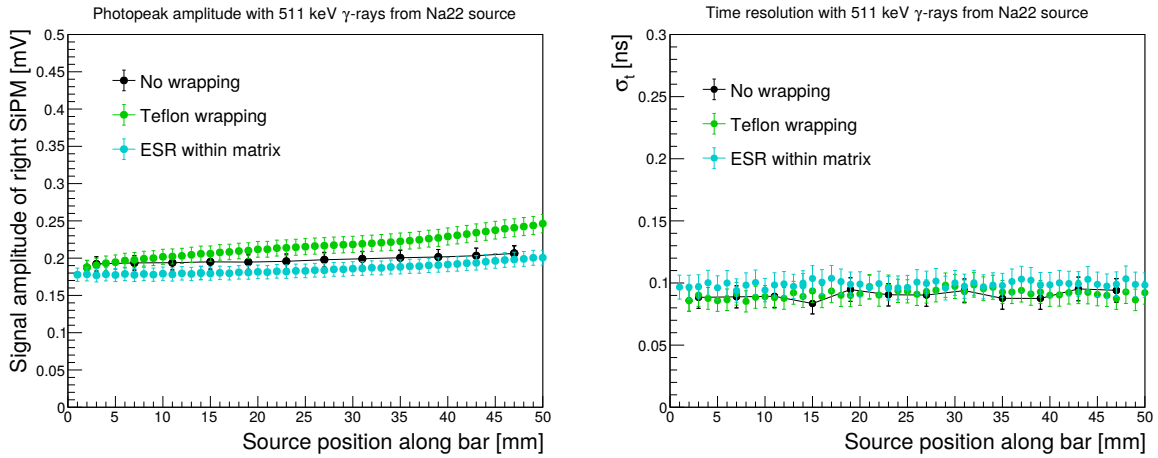


Figure 2.12: Comparison of the signal amplitude (left) and time resolution (right) as a function of source position along bar axis measured in the laboratory with a ^{22}Na source for a naked bar, a bar wrapped with Teflon layers and a bar within an array with ESR foils (Enhanced Specular Reflector Vikuiti by 3M) in between channels.

Table 2.3: Specifications of scintillating properties, radiation tolerance requirements and geometrical tolerances for BTL LYSO:Ce crystals. The radiation tolerance requirement has to be met after an integrated irradiation to the levels of hadron fluences and ionizing doses indicated in Table 1.3.

LYSO:Ce crystal parameter	Specification	Spread (rms)
Light output / end	$>$ 6000 photons/MeV	$<$ 5%
LY(10ns)/LY(200ns)	$>$ 20 %	$<$ 3%
LY(200ns)/LY(2000ns)	$>$ 95 %	$<$ 3%
Decay time	$<$ 43 ns	$<$ 3%
Rise time	$<$ 200 ps	$<$ 3%
Density	$>$ 7.1 g/cm^3	$<$ 2%
Refractive index	1.82	–
Radiation tolerance		
Loss of light output	$<$ 5%	$<$ 5%
Induced absorption coeff., μ_{ind}	$<$ 3 m^{-1}	$<$ 5%
Dimensions	Specification	Tolerance
Length [mm]	57.0	+0.00/−0.03
Width [mm]	3.12	+0.00/−0.03
Height [mm]	3.75 / 3.0 / 2.4	+0.00/−0.03
Surface polishing	$<$ 15 nm	

bar of the nominal dimensions with Teflon wrapping and by having a SiPM glued at each end so that light is shared among the two sides as in the final detector configuration.

Another key parameter is the time distribution of scintillation photons that should feature a short rise and decay time with negligible number of photons emitted after 200 ns. In particular it is envisioned to monitor this property of the crystal by mean of gated light yield measurements, in which the ratio of light emitted within different time intervals is measured to quantify the scintillator time constants as reported in Table 2.3. The density of the material defines the amount of energy deposited by a MIP and thus should not be lower than the value specified in the table. This sets a limit to the maximum amount of Yttrium that can be used in the Cerium

doped Lutetium orthosilicate compound $(Lu_{1-x}Y_x)_2SiO_5$ to 20%.

All the specifications should be met at the BTL nominal operating temperature of $-30\,^{\circ}C$. The light output of LYSO:Ce crystals at this temperature has been measured and showed an increase of about 5% with respect to room temperature due to a lower probability of non-radiative recombination of electron-hole pairs. The change in light output due to the increase of light absorption induced by radiation damage should be below 5%. For the crystal bar geometry this corresponds to an induced absorption coefficient $\mu_{\text{ind}} < 3\,\text{m}^{-1}$, as measured on LYSO:Ce crystals from several vendors after irradiation to fluences and doses expected for BTL or higher [43–45].

The dimensions of the crystal bars are specified for the three different thicknesses that will be used across the detector and tolerances are indicated to pose no threat to the assembly of the modules and coupling to the SiPMs. All crystal surfaces must be polished to a degree of optical quality with $R_a < 15$ nm to guarantee acceptable light collection by total internal reflection. The plans for quality control of the crystals and matrices are discussed in the next section.

2.2.1.2 Crystal quality control

The LYSO:Ce crystals will not be customized, relative to what is typically produced by vendors. The polishing of the crystals will be done by the vendor who will also provide packaged arrays of crystals ready for assembly, as discussed later. In an initial, non-binding call for quotes, we have addressed about 10 LYSO:Ce manufacturers. The qualification of the vendors will be performed in 2019 and will include a thorough validation of the vendors' products. The measurement campaign, aiming at identifying the manufacturers able to produce crystals that best meet the requirements in Table 2.3, will start in spring 2019. Samples of crystal bars and crystal arrays from different producers will be fully characterized in the laboratory by the following steps:

- Light output and energy resolution with 511 keV photons;
- Time resolution with cosmic rays and with 511 keV photons;
- Decay time;
- Optical isolation of crystals within the array;
- Measurement of the dimensions with a precision of O(10 μm). Planarity of the arrays will be computed by measuring all the crystals of the array at different positions along the crystal axis.

Furthermore, a subset of the crystal samples will be irradiated with photons and neutrons to check the radiation tolerance. Samples will be irradiated to the integrated ionizing dose of 25 kGy, corresponding to the BTL radiation levels at $3000\,\text{fb}^{-1}$. Measurements of light yield and time resolution before and after irradiation will be used to assess the radiation tolerance of the elements. Characterization activities will take place throughout the whole of 2019; then a tender for LYSO:Ce arrays pre-production, corresponding to few percent of the full quantity needed for the BTL, will be prepared.

A pre-production step foreseen in early 2020 and involving 2–3 vendors will allow the selection of the optimal vendor and reduce the additional QA/QC that must be performed at the production stage. During the production stage, a QC sampling of crystals from each production batch is not excluded (depending on the outcome of the pre-production run). If deemed necessary this batch testing will include detailed measurements of the scintillation and radiation hardness properties on only one or two crystal samples per production batch. These tests serve as a

quality control, and the acceptance rate is expected to be very high. The parameters measured in the batch sample testing are assumed to be sufficiently stable within one batch as crystals will come from the same crystal ingot, being the monolithic crystal block from which samples are cut out. Such samples are thus subject to the same growth conditions and quality of the raw material. The selection of a reliable vendor following the qualification and pre-production steps could exclude the necessity of this sampling test. For each crystal array, we will perform a measurement of the dimensions to make sure they match the requirements. This step can be integrated into the gluing procedure with a pick-and-place robot. The matrices not conforming to the tolerances, whose fraction is expected to be at the 1% level, will be discarded.

2.2.2 Silicon photomultipliers

The photo-sensors of choice for the BTL are silicon photomultipliers (SiPMs). Rapid progress in the performance of SiPMs has led to their widespread use in accelerator and non-accelerator based particle and nuclear physics experiments, space-based telescopes, and medical imaging. SiPMs have a number of advantages over other photo-sensors, such as conventional photomultiplier tubes. SiPMs are compact, robust, and insensitive to magnetic fields. They can be exposed to room light without damage and operate at relatively low voltages, on the order of 30–77 V, with low power consumption. A photo-detection efficiency, PDE, of up to 40% is achievable in devices with small cell size (15 μm square pixels). Small cell sizes also extend the linear range of the SiPM and, combined with a fast cell recovery time, enhance its performance after irradiation. Finally, large quantities of SiPMs can be fabricated in industry with excellent uniformity and acceptable cost. The CMS collaboration already has several years of good experience with SiPMs as part of the Phase-1 upgrade of the hadron calorimeter (HCAL), where more than 16 000 channels of SiPM are used [49], and is planning to use this type of photodetector for the instrumentation of part of the HGCAL endcap calorimeter for the Phase-2 upgrade [7].

2.2.2.1 SiPM technical specifications

The specifications for the BTL SiPMs are listed in Table 2.4 and compared to the performance of existing SiPM technologies under consideration, namely the NUV-HD (thin-epi) SiPM from Fondazione Bruno Kessler (FBK) and the S12572 and HDR2 from Hamamatsu Photonics (HPK) with cell pitch of 15 μm. The chosen SiPM cell size of 15 μm represents the best balance between radiation tolerance and photon detection efficiency [50–52].

Optimization studies through beam tests, laboratory measurements, and ray tracing simulation led to the choice of a SiPM active area of about 9 mm^2. The BTL SiPMs will thus have a rectangular shape with a width of 2.9 mm (a minimum of 100 μm gap is required on each side for the packaging) and a height matching the crystal thicknesses with a 100 μm margin to maximize light collection from the crystal edges (thus 3.85, 3.1 and 2.5 mm), as illustrated in Fig. 2.18. In the crystal bar geometry the LCE scales linearly with the fraction of surface covered by the light sensitive area of the SiPM, $\text{Area}_{\text{SiPM}}/\text{Area}_{\text{crystal}}$, thus reducing the SiPM area would deteriorate both the stochastic and noise terms.

A useful figure of merit to compare the timing performance of different SiPMs is the ratio $\text{PDE}/\sqrt{\text{DCR}}$, which directly affects the signal-over-noise ratio for the first detected photons from LYSO:Ce scintillation and determines the sensor time resolution. The single cell recovery time should be kept below 10 ns to allow fast cell recovery from a dark count, thus maintaining the SiPM cell occupancy below 3% for a DCR of 55 GHz. This parameter is driven by the value of the quench resistor at $-30\,^{\circ}\text{C}$. A gain larger than 1.3×10^5 is required to allow a sufficiently precise signal discrimination by the ASIC, as discussed later. Additional parameters of the

SiPM such as the Excess Noise Factor, ENF, which represent additional noise with respect to DCR originating from after-pulses and cross-talk are also shown in the table.

Table 2.4: Comparison of end-of-operation specifications for BTL SiPMs to the performance of existing candidate devices of 3×3 mm^2 area from FBK and HPK irradiated to neutron fluences. The extrapolations of DCR, PDE and static power consumption are calculated at the optimal bias over-voltage after 3000 fb^{-1} for an operating temperature of $-30\,^{\circ}$C, after 2×10^{14} n_{eq}/cm^2. The values of cell capacitance, cell recovery time and breakdown voltage, V_{br}, are also quoted at a temperature of $-30\,^{\circ}$C.

SiPM parameter	**Specification**	**FBK-NUV-HD**	**HPK-S12572**	**HPK-HDR2**
Active area	–	~ 9 mm^2	~ 9 mm^2	~ 9 mm^2
Cell pitch	< 20 μm	15 μm	15 μm	15 μm
Cell recovery time	< 10 ns	7 ns	8.5 ns	< 10 ns
Capacitance	< 600 pF	530 pF	295 pF	585 pF
Number of cells	> 20k	~ 40k	~ 40k	~ 40k
V_{br} ($-30\,^{\circ}$C)	–	34.2 V	63.0 V	35.8 V
dV_{br}/dT	–	41 mV/ $^{\circ}$C	59 mV/ $^{\circ}$C	37 mV/ $^{\circ}$C
$\delta V_{br}/10^{13}$ n_{eq}/cm^2	≤ 0.2 V	< 0.1 V	0.2 V	< 0.1 V
DCR-T coefficient	–	1.76	1.90	1.79
ENF	< 1.1	< 1.05	1.07	< 1.05
Parameters after 3000 fb^{-1}				
Optimal OV	> 1V	1.6 V	1.5 V	1.2 V
PDE	–	15%	13%	23%
Current/device	–	1.32 mA	0.77 mA	1.30 mA
Static power consumption	≤ 50 mW	50 mW	50 mW	50 mW
Gain	$\geq 1.3 \times 10^5$	2.1×10^5	1.45×10^5	1.55×10^5
DCR/SiPM	–	42 GHz	37 GHz	55 GHz
PDE/$\sqrt{\text{DCR}}$	≥ 2.0	2.3	2.1	3.1

Characteristic plots of the photon detection efficiency, as a function of wavelength and over-voltage, are shown in Fig. 2.13. Curves are shown for three different SiPM technologies under consideration for the BTL: S12572-015C and HDR2-015 from Hamamatsu and NUV-HD thin-epi from FBK.

As can be seen from the figures, the SiPM performance is sensitive to the over-voltage, which means that both the operating voltage and the V_{br} must be well monitored and controlled as discussed later. The breakdown voltages measured for 4 000 channels, measured from the recent CMS HCAL barrel production run, were contained in an interval of 0.3 V with a standard deviation of 56 mV. This indicates a very small spread in the performance of the SiPMs delivered by manufacturers.

When SiPMs are exposed to 1 MeV neutron equivalent fluences, a linear drift of the breakdown voltage of about 0.1–0.2 V every 10^{13} n_{eq}/cm^2 has been observed. This drift has to be taken into account and will require adjustment of the bias voltage to maintain the operating voltage of the SiPM at its optimum. The spread in the drift of V_{br} with time was also measured on a set of irradiated SiPMs and found to be smaller than 5% (which corresponds to less than 0.2 V after the total integrated luminosity of 3000 fb^{-1}). More details on the system for bias voltage distribution and monitoring are discussed later.

While all the SiPMs under consideration meet the specifications in Table 2.4, some additional

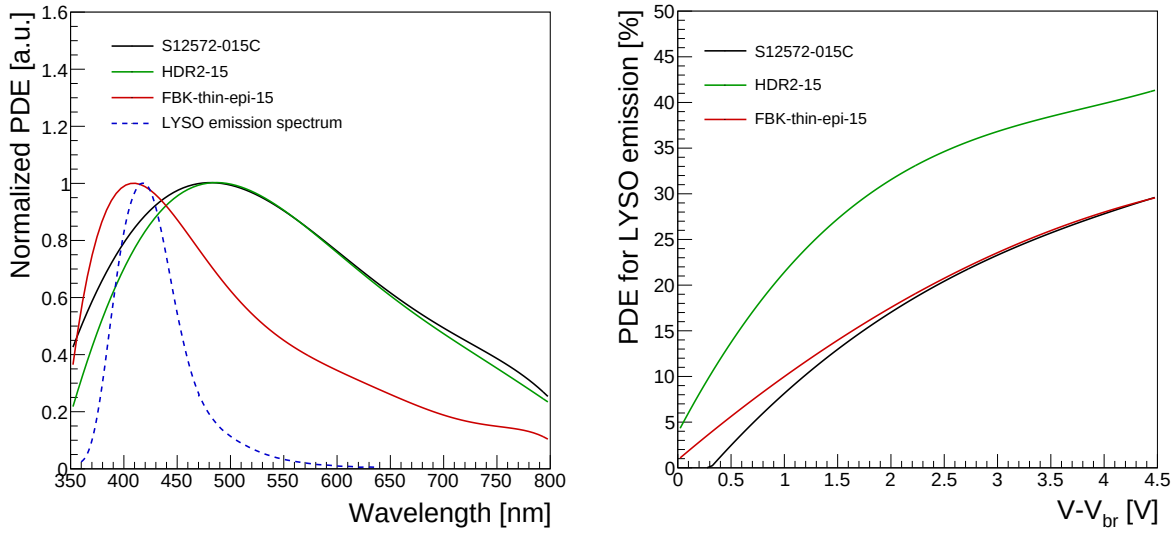


Figure 2.13: Left: Relative photon detection efficiency (PDE) as a function of wavelength compared with the spectrum of LYSO:Ce scintillation light (blue dashed curve). Right: "effective" PDE, convolved with the LYSO:Ce emission spectrum, as a function of over-voltage for three different SiPM technologies considered for BTL.

R&D is ongoing to consolidate such technologies. The S12572-015C SiPM provides a solid option which has already undergone a successful mass production qualification for the CMS HCAL Phase-1 upgrades. The other two SiPMs from both FBK and HPK have been developed more recently and a further custom optimization of their PDE/$\sqrt{\text{DCR}}$, being the figure of merit for BTL performance, is ongoing. Studies on the reproducibility and mass production capability for these two technologies is also planned during 2019 and will test about 50 arrays of 16 SiPMs from each vendor.

2.2.2.2 SiPM quality control

We are currently in close contact with two vendors (HPK and FBK) to finalize the optimization and customization of the SiPMs for the BTL needs. Although both producers are capable of providing SiPMs with the desired specifications, a rigorous QA/QC will be requested from the manufacturer and testing of each SiPM on our side will be carried out. The testing procedure will capitalize the extensive experience with SiPMs from the CMS HCAL project. The following measurements are foreseen:

- Every SiPM for each batch will be tested for its IV (current-voltage) curve with and without light illumination, as well as its resistance in forward bias.
- For 2% of the SiPMs for each batch, the capacitance and pulse shape will be measured.
- In addition, destructive testing will be performed on 1% of the SiPMs for each batch. These destructive tests will include radiation testing, long term aging studies, and environmental studies (i.e., temperature cycling, humidity effects, etc.).

The assembly of the SiPM arrays will be conducted in industry. As these modules are of rather low complexity and are utilizing industry standard electronics board technology, standard QA/QC procedures will be applied.

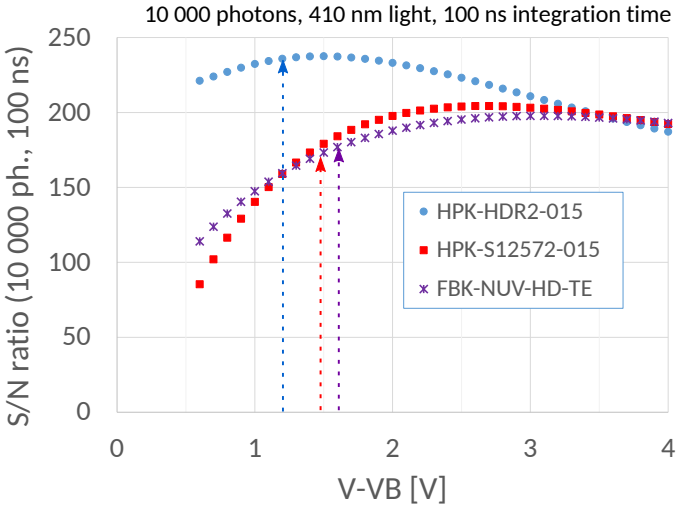


Figure 2.14: Signal-over-noise ratio measured using a reference LED on the BTL SiPM candidates after irradiation to 2.1×10^{12} n_{eq}/cm^2, as a function of over-voltage. The optimal over-voltage for BTL operation after 3000 fb^{-1}, within the power budget envelope, is shown with an arrow for each SiPM.

2.2.2.3 SiPM irradiation studies

Extensive studies of the impact of radiation damage on the performance of SiPMs have been conducted during 2017, 2018 and 2019. Such studies led to the selection of the technologies providing the optimal resistance to the radiation level expected for BTL. As discussed in previous sections, a crucial parameter determining the timing performance of the SiPM coupled to the crystal is the ratio of its PDE (signal, S) over the $\sqrt{\text{DCR}}$ (noise, N) at the optimal over-voltage. This parameter can be estimated by measuring the signal-over-noise ratio (S/N) of irradiated SiPMs using a light pulse of known amplitude. The results obtained on different SiPMs irradiated to a fluence of 2.1×10^{12} n_{eq}/cm^2 are shown in Fig. 2.14. The radiation damage, and thus the SiPM dark current, has been proven to increase linearly with fluence up to the BTL level of 3×10^{14} n_{eq}/cm^2, which includes a safety factor of 1.5 on top of the nominal maximum predicted fluence. The maximum OV allowed for each SiPM within the power budget constraint of the BTL is also shown in Fig. 2.14. It can be seen that each SiPM features a different optimal over-voltage and provides slightly different S/N performance as reported in the earlier section and in Table 2.4. Qualification of different materials to be used for the SiPM protective window has started and several options have been tested for radiation tolerance, up to a neutron fluence of 3×10^{14} n_{eq}/cm^2 and to ionizing doses of 50 kGy, leading to the identification of viable candidates, e.g. silicone based resins as discussed in more detail in Appendix B.

2.2.2.4 Mitigation of DCR with annealing

The DCR generated in the SiPM due to radiation damage will increase linearly with the particle fluence, which will be proportional to the integrated luminosity. At the BTL operating temperature of $-30\,^{\circ}C$ and an over-voltage of 1.5 V, the DCR will increase by about 1.0–1.6 GHz/mm^2 every 500 fb^{-1} of integrated luminosity. At the end of detector operation, after an integrated luminosity of 3000 fb^{-1} which corresponds to the maximum neutron equivalent fluence of $1.9 \times 10^{14} n_{eq}/cm^2$ in BTL (at high η), and assuming yearly annealing periods at room temperature during shutdowns, a DCR of about 4–7 GHz/mm^2 is expected, depending on the SiPM choice. A large DCR will cause large baseline fluctuations that will be corrected for with

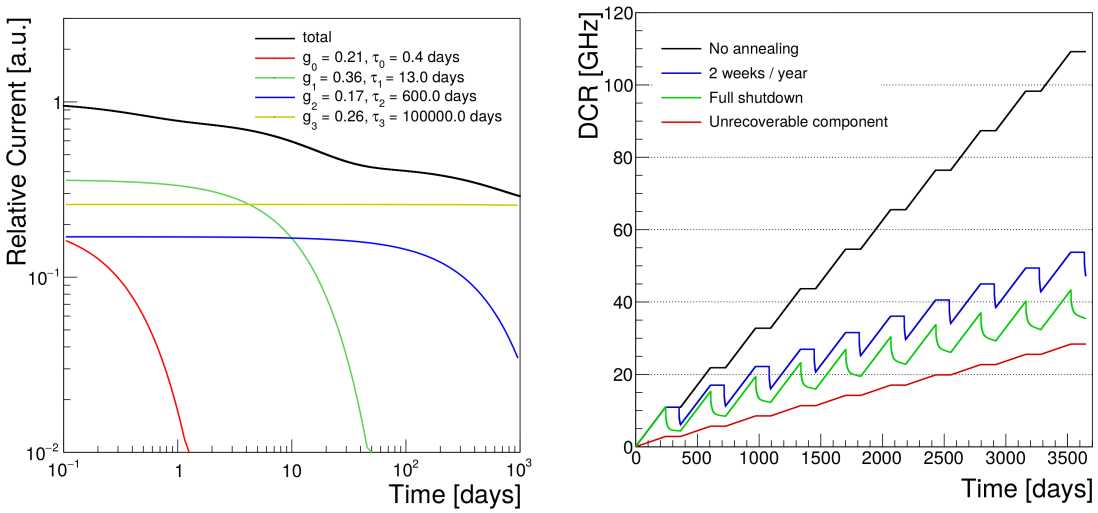


Figure 2.15: Left: radiation induced current annealing kinetics measured for APDs in Ref. [53]. Right: expected growth of DCR (SiPM S12572) for various annealing scenarios at fixed OV of 1.5 V during the detector lifetime.

a dedicated circuit in the ASIC (Section 2.3). In addition, the DCR will contaminate the photon signal causing additional jitter of the time stamp, impacting on the time resolution.

The radiation damage induced in silicon consists of different types of defects, each having a characteristic recovery time, τ_i. Spontaneous annealing of the radiation induced current has been observed on irradiated SiPMs [40] and from the first tests it appears to follow a behavior similar to that of APDs described in Ref. [53]. This model, assuming four different defect types, is summarized in the plot of Fig. 2.15, showing the expected annealing of each component as a function of time at room temperature (RT). The annealing kinetics, $I^{\text{irr}}_{\text{dark}}(t)$, of the total radiation induced dark current, $I^{\text{irr}}_{\text{dark}}(0)$, can be parameterized as:

$$I^{\text{irr}}_{\text{dark}}(t) = I^{\text{irr}}_{\text{dark}}(0) \sum_i g_i e^{-\tau_i/t}. \tag{2.4}$$

Since BTL will be operating at $-30\,^{\circ}\text{C}$, only minor annealing will take place during data taking while substantial recovery will occur during yearly shutdowns with the duration of about four months. The right plot of Fig. 2.15 shows the expected growth of the DCR at fixed OV of 1.5 V during the detector lifetime for the cases of:

- No annealing;
- Annealing at RT only during two weeks per year;
- Annealing at RT during the full shutdown;
- Complete annealing of the recoverable defects.

Exploiting the full shutdown period for recovery at room temperature provides a reduction of DCR of about 30% with respect to a two weeks only scenario. Additional recovery could be achieved by increasing the local temperature of the SiPM during shutdowns to a temperature higher than $+20\,^{\circ}\text{C}$. The potential gain of such approach and its technical viability are being investigated with dedicated studies and more details are provided in Appendix B.

2.2.2.5 Evolution of SiPM operating parameters with integrated luminosity

The time resolution of BTL sensors is strongly driven by the SiPM parameters and mainly by the PDE and DCR, which directly affect the photo-statistic and noise term, as discussed in Section 2.1. Since the DCR will increase with fluence, adjustment of the operating OV of the SiPMs is required throughout the detector lifetime. In particular, the OV will have to be decreased in order to maintain the DCR level within a range manageable by the ASIC and as well to limit its contamination to the light signal. Reduction of the OV will also limit the power dissipated by the SiPM through heat within the design specification. The optimal OV is thus defined as the over-voltage at which the best time resolution is achieved (from Eq. 2.1) within the power budget envelope of 50 mW/SiPM. The evolution of the main SiPM parameters as a function of integrated luminosity is shown in Fig. 2.16 for the three SiPM options considered for BTL.

Depending on the specific SiPM, the optimal OV will decrease from about 3.5 V to 1.2 V, and the current and the static power consumption per SiPM will stabilize in the range of 0.7–1.3 mA and 30–50 mW respectively. The PDE will decrease from 35% to 23% for the HDR2-015 and from 27% to about 14% for the S12572 and FBK-NUV-HD SiPMs. The DCR will increase non-linearly up to 37, 42 and 55 GHz for the S12752, FBK-NUV-HD and HDR2 SiPM respectively. Nevertheless, the fraction of SiPM cells fired by a DCR and which cannot detect a photon (cell occupancy due to DCR) will be limited to below 2–3% and thus have negligible impact on the effective photon detection efficiency. This number has been estimated assuming a recovery time of the single cell of $\tau_R = 8$ ns and considering a cell to be blind for a time corresponding to $2\tau_R$, i.e. 16 ns.

The total signal in number of photoelectrons and electrons, expected for an energy deposit of 4.2 MeV, is also shown in Fig. 2.16 and defines the specification for the dynamic range of the ASIC. The signal in photoelectrons, N_{phe}, is obtained as the product of light output for a 4.2 MeV MIP energy deposit and the PDE corrected for the cell occupancy (which is a marginal effect). The electron signal is the actual signal output from the SiPM after the multiplication factor defined by the SiPM gain, G, and is thus defined as $N_{\text{phe}} \times G$.

2.2.3 Packaging of active elements

To ease assembly and QA/QC, the LYSO:Ce crystal bars will be packaged by the vendor. The packaged structure will consist of a linear array of 16 crystals with a reflective material between adjacent channels to provide optical isolation to better than 95%. The material used as reflector is required to have a thickness of $\sim$75 μm, a reflectivity for 420 nm light higher than 98.5%, and sufficient radiation tolerance. The baseline material currently considered is the Enhanced Specular Reflector (ESR) Vikuiti by 3M. The whole crystal array will be held together by a similar foil of about 0.2 mm providing stability to the structure, similar to what is commonly done for PET applications. The use of adhesives typically of about 5 μm thickness to bond the crystal array should be minimized in area along the bar sides to preserve total internal reflection from the surfaces. A picture and a drawing of one crystal array is shown in Fig. 2.17. The crystal pitch, including a inter-crystal gap of about 80 μm, is 3.2 mm which makes the total size of an array in $\phi \times z$ about 57.0 $\times$ 51.4 mm. The flatness of the two end-faces of the array, which are left without wrapping, should be better than 0.05 mm to allow a precise optical coupling to the SiPM array.

Similarly, the SiPMs will be delivered by manufacturers packaged in arrays of 16 units each. The SiPM arrays will match the pitch and geometry of the readout face of the crystal array. The packaging of the SiPM is minimized to reduce dead gaps between crystal arrays and optimized

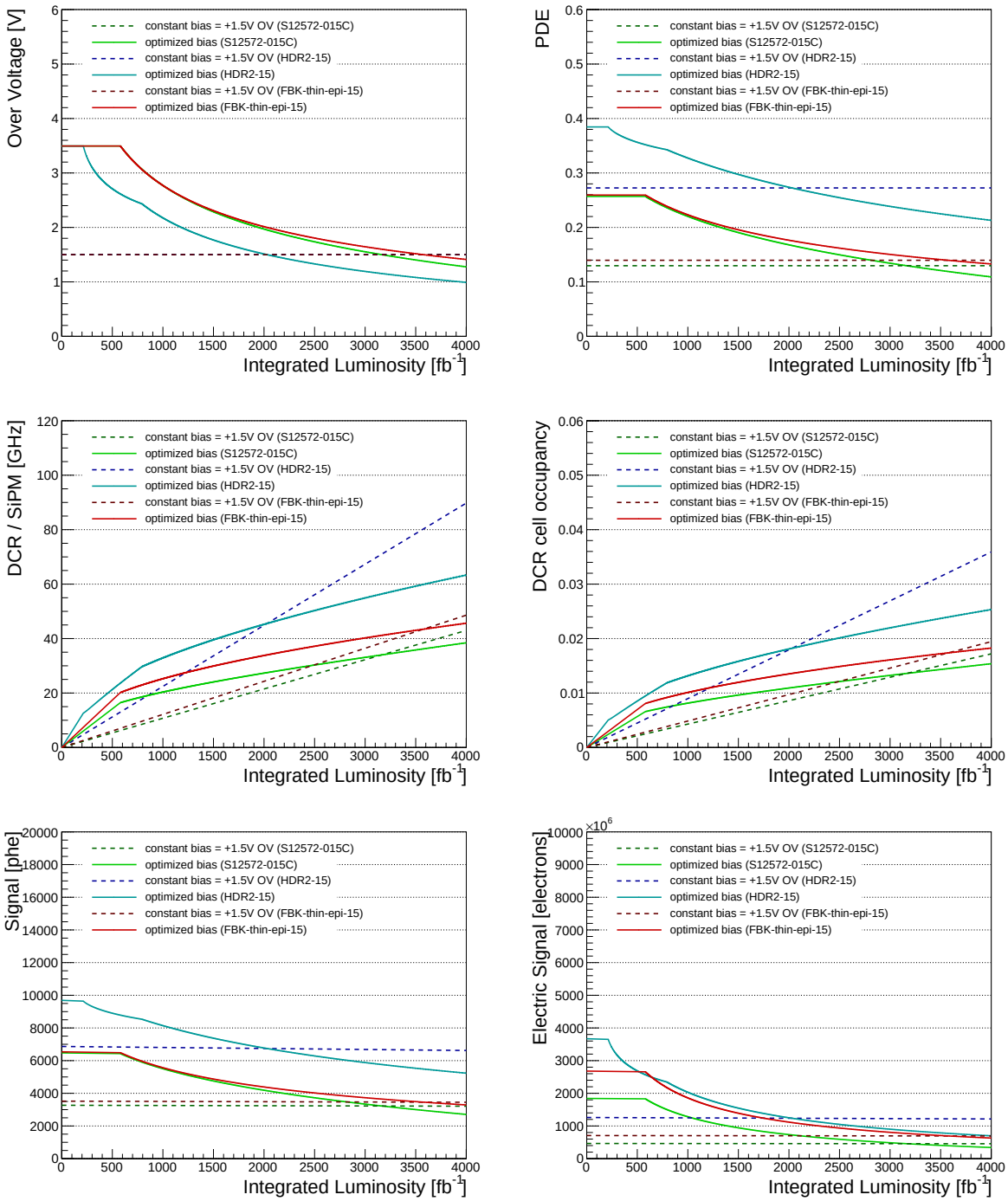


Figure 2.16: Evolution of SiPM operating parameter as a function of integrated luminosity. From top left to bottom right: over-voltage (OV), photon detection efficiency (PDE), DCR per SiPM of 9 mm^2 active area, fraction of cells occupied by a dark count, 4.2 MeV signal in photo-electrons and electrons.

for heat extraction to allow efficient power dissipation towards the cooling plate, as discussed in more detail below. The protective window of the SiPM will be made of a radiation tolerant material (e.g. quartz, epoxy, silicon) with refractive index at 420 nm above 1.5, and will be made as thin as possible ($\sim$200 μm) to maximize light collection efficiency from the crystal. Radiation tolerant candidate materials for the protective windows have been identified within the

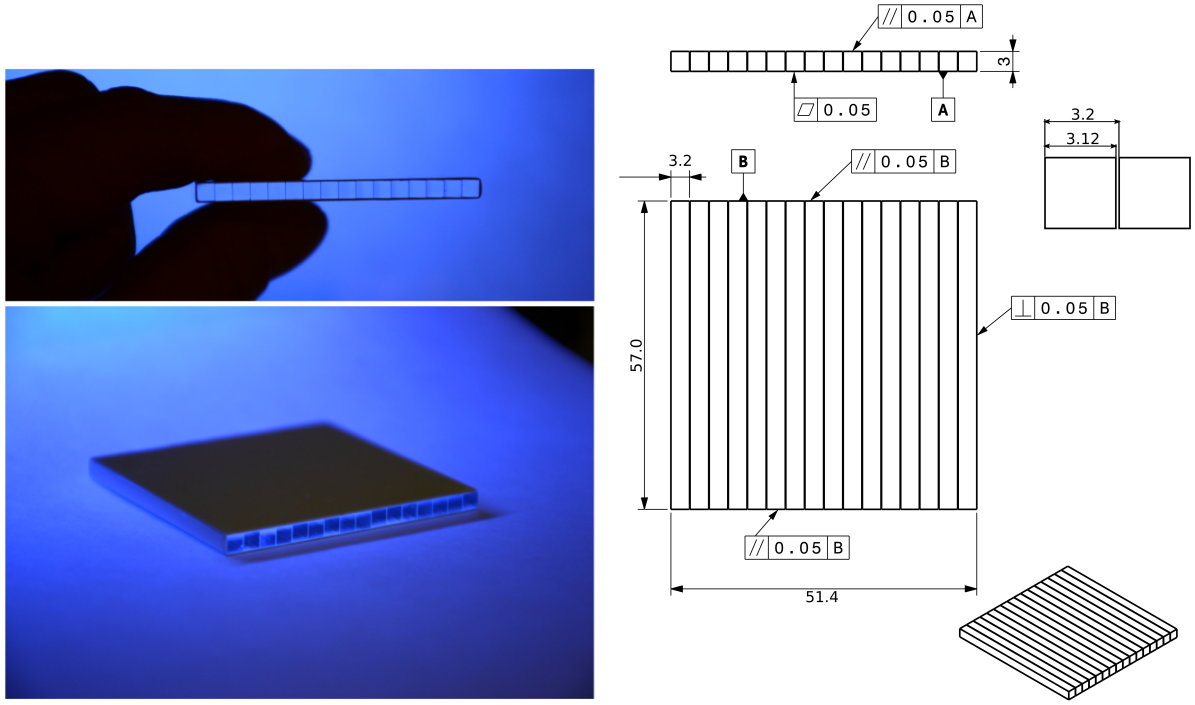


Figure 2.17: Left: picture of a 16×1 array of crystal bars with Enhanced Specular Reflector (ESR) Vikuiti by 3M in between channels under UV illumination. Right: drawing and dimensions of a BTL crystal array.

options provided by both manufacturers after an irradiation campaign of candidate materials as discussed earlier.

A drawing of the SiPM package is shown in Fig. 2.18. A ceramic board with thickness of 0.7 mm can house either wire bonded or TSV (Through Silicon Via) SiPMs from either Hamamatsu or FBK. With respect to a PCB-based package, the use of ceramic provides a thermally more stable and more compact solution. The estimated thermal conductivity of such package, summarized in Table 2.5, is designed to maintain a thermal gradient smaller than 1 °C.

Table 2.5: Estimated thermal conductivity for the linear array of 16 SiPMs with a single layer ceramic package using HPK/FBK TSV SiPMs.

Component	**Material**	**Size** [mm]	**Inner diam.** [mm]	**Thickness** [mm]	**Th. cond** [W/mK]	Δt for 50 mW [°C]
Die	Si	3.2	–	0.35	100	0.017
Die Attach	Solder	1.5	4×0.7	0.05	50	0.021
Top pad	Cu	3.0	–	0.035	377	0.001
Substrate	Al_2O_3	3.2	–	0.7	20	0.171

The crystal and SiPM arrays will be coupled together during the assembly phase by means of a radiation tolerant optical coupling. The baseline choice is to use the RTV3145 glue used for the coupling of the CMS ECAL crystal to the photodetectors. This type of glue provides good optical coupling matching the SiPM protective window, is radiation tolerant to the levels expected in BTL and has sufficient elasticity to operate at low temperatures. No significant degradation of transparency was observed and no effect on mechanical properties occurred after irradiation and several thermal cycling tests performed at the time of the CMS ECAL assembly [54, 55]. Other commercially available optical glues, such as Meltmount or NOA61,

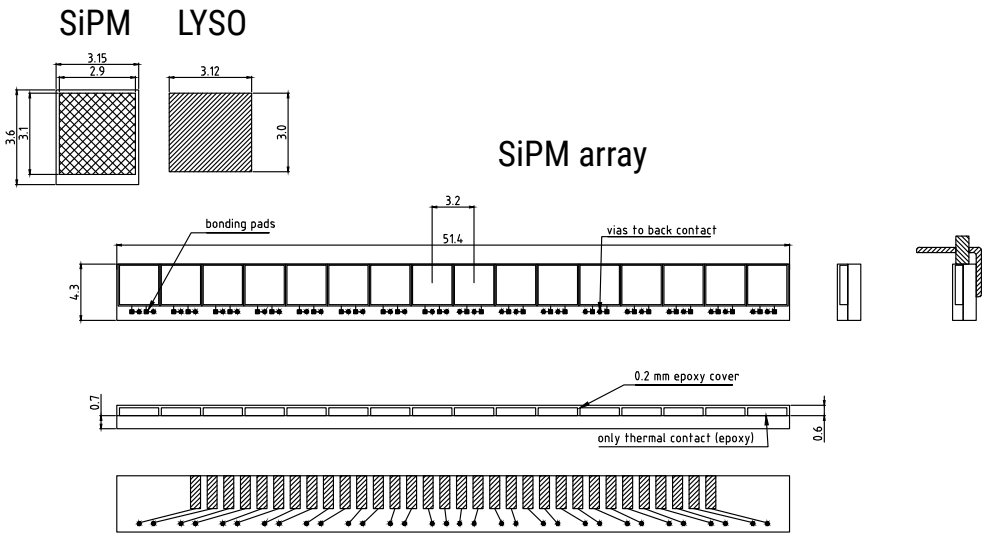


Figure 2.18: Drawings of a linear array of 16 SiPMs on a ceramic package showing the matching to the SiPM active area with the crystal dimensions.

have also been widely tested up to the radiation levels expected for the BTL, i.e. neutron fluences of 2×10^{14} n_{eq}/cm^2 and doses of $\sim$30 kGy. Additional studies are ongoing to consolidate the baseline choice as discussed in more detail in Appendix B.

2.3 BTL readout electronics

The basic building block of the BTL electronics is the Readout Unit (RU) that processes signals from 768 individual SiPMs. The way the sensor matrices (crystal + SiPM arrays) are mechanically and electrically connected to the RU is presented in Section 2.4. An RU consists of four Front-End cards (FEs), each of which has six readout chips to process data from 192 SiPMs, using the TOFHIR2 ASIC. The four FEs are connected to the Concentrator Card (CC) that houses two low-power Giga-Bit transceivers (lpGBT) [56] associated to two Versatile Link Plus (VTRx+) [57] chips.

Optical data links to/from the off-detector DAQ are implemented over the lpGBT-Versatile link system, which runs bi-directionally between the detector and the DAQ boards. The lpGBT uplinks (from front-end to DAQ) are operated at 10.24 Gb/s and the downlinks (from DAQ to front-end) are operated at 2.56 Gb/s. The data from each TOFHIR2 are transmitted by two E-links, operated at a bandwidth of 320 Mb/s each, to the two lpGBTs on the Concentrator Card. The lpGBTs provide E-links (80 Mb/s) to the TOFHIR2 transporting configuration data, as well as fast control signals (Resync and trigger bits). Clocks with 160 MHz frequency are distributed from the lpGBT to all TOFHIR2 ASICs.

Two GBT-SCA chips on the Concentrator Card provide monitoring of low voltage, temperature and SiPM bias currents. Two Power Converter Cards (PCC) provide power from a low voltage distribution system with DC/DC converters based on the FEASTMP module. The DC-DC output voltages are further regulated and filtered by the ALDO2 regulator ASIC in the FE boards. One ALDO2 regulator serves one TOFHIR2 ASIC.

The block diagram of the RU is shown in Fig. 2.19 and the layout of the RU is shown in Fig. 2.20. The RU boards are arranged in a single layer using board side-to-side connectors in order to

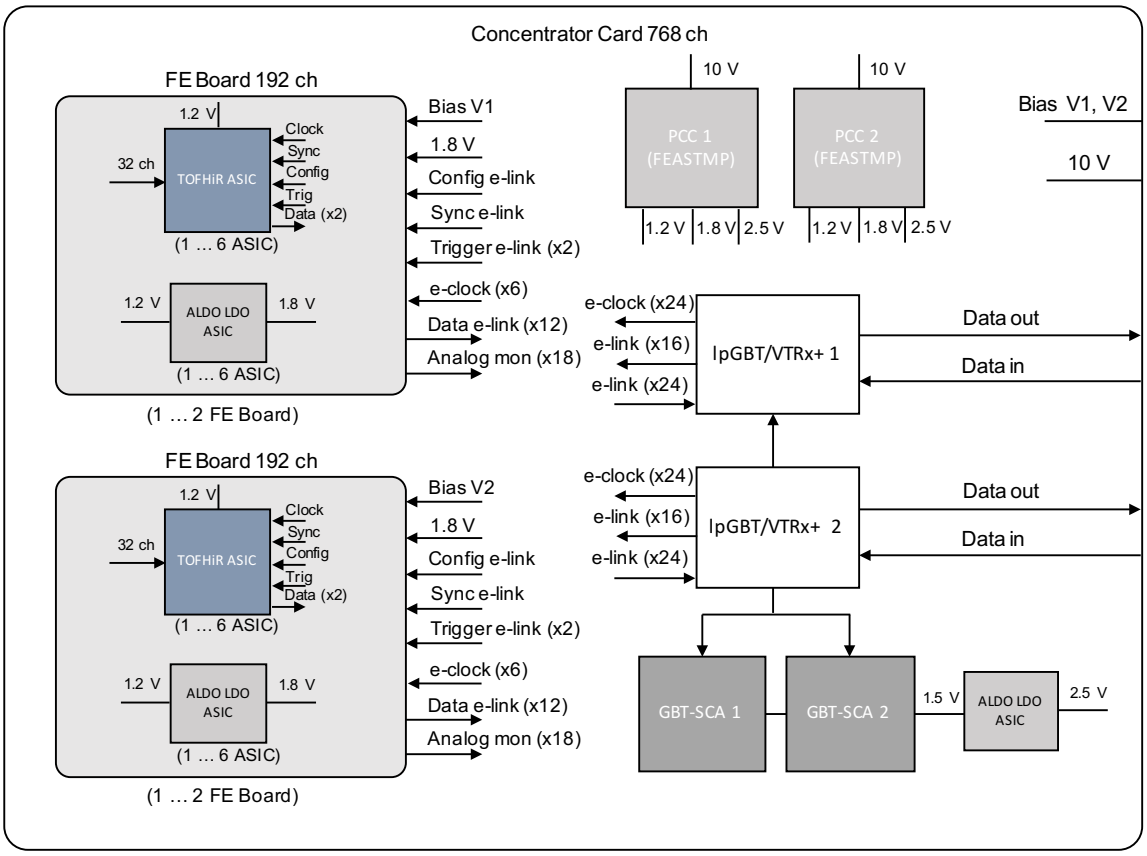


Figure 2.19: Block diagram of the Readout Unit.

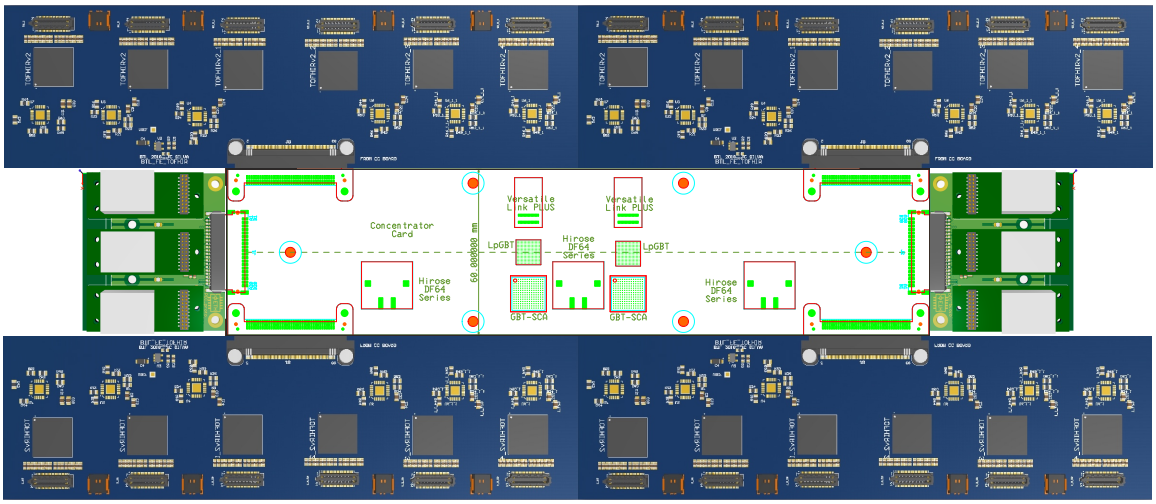


Figure 2.20: Layout of the BTL Readout Unit. Each Concentrator Card (center) is connected to four FE boards (above and below the Concentrator Card) that host 6 TOFHIR ASIC chips each, and to two Power Converter Cards (left and right of the Concentrator Card).

match the radial space available for the BTL electronics (less than 9 mm). Overall, 432 RUs are needed to populate the BTL, arranged in 72 trays of 6 RUs/tray.

The TOFHIR ASIC is derived from the TOFPET2 chip developed for TOF-PET applications using LYSO:Ce crystals coupled to SiPMs [35, 37, 58]. The design is being adapted to match the requirements of the barrel timing layer, in particular the high signal rate and the tolerance

Table 2.6: Comparison of TOFHIR1 and TOFHIR2 functionality.

	TOFHIR1	TOFHIR2
Number of channels	16	32
Technology	UMC 110 nm	TSMC 130 nm
Voltage	1.2 V, 2.5 V	1.2 V
Radiation Tolerance	No	Yes
Compatibility with lpGBT	Yes	Yes
I/O links	LVDS	CLPS
L1, L0 Trigger	Yes, No	Yes, Yes
10-bit SAR ADC (MHz)	10	40
Bandwidth (MHz)	350	350
Input impedance (Ω)	6	6
DCR noise filter	No	Yes
Number of TACs and QACs	4	6
TDC bin (ps)	20	20
Reference voltages	External	Internal
Maximum MIP rate/ch (MHz)	1	2.5
Max low E rate/ch (MHz)	3	5
Clock frequency (MHz)	160	160

to radiation. The TOFHIR1 chip was developed using the same UMC 110 nm technology as the TOPFET2 ASIC and presently is being tested. It implements the main features required by the BTL and is enabling the development of the final detector modules and validation of system integration on a tight schedule. The TOFHIR2 adds radiation tolerance by translating the TOFHIR1 design to the more radiation tolerant TSMC 130 nm technology and revising the design for radiation tolerance. In addition a DCR noise cancellation mechanism, a key feature, will be integrated in the TOFHIR2. The comparison of TOFHIR1 and TOFHIR2 functionality is summarized in Table 2.6.

2.3.1 Requirements and system description

2.3.1.1 FE ASIC requirements

The TOFHIR2 ASIC will be used for the acquisition and digitization of the photo-sensor signals. In particular the circuit must:

- Be capable of performing the digitization of time and energy of the passing minimum ionizing particles (MIPs) with the required precision and at the required rate;
- Reject lower energy (lower than MIPs) particles (mainly photons from the electromagnetic calorimeter);
- Provide the necessary links for data transmission, fast control and slow control:
- Integrate test features;
- Comply with radiation tolerance, power and operation temperature requirements.

The details of those requirements are outlined below.

Signal Yield

The average energy deposit per channel is estimated to be about 4 to 7 MeV, where we require the detection of particles starting at 1 MeV. Depending on the SiPM type selected for BTL (see Section 2.2.2) we expect about 9000 photoelectrons, decreasing to 4000, or 13 000 photoelec-

trons, decreasing to 7000, at end of operation.

The photo-sensors will be operated at maximum gain of 4×10^5 electrons/photon. We require a full dynamic range of 2.1×10^{10}, 1.6×10^{10}, 1.2×10^{10} electrons again depending on the photo-sensor type, sufficient to accommodate MIP signals four times larger than the average signal. The exact full dynamic range will be fixed at the latest for the second iteration of the ASIC. In addition, the total input charge will decrease by a factor of about 5.5 over the operation time of the detector (3000 fb^{-1}). It is therefore required that the gain can be adjusted in order to compensate for this change of signal amplitude.

Signal shape and timing

The shape of the input signal is determined by the scintillation light decay time of the LYSO:Ce crystal of 40 ns and the timing properties of the photo-sensors. Following the ionization originating from a passing charged particle, photons arrive in time following an exponential distribution with the above given decay time. The tail of this distribution spreads out to about 300 ns. The rising edge of the signal depends on the photo-sensor used but we expect a peaking time of about 14 ns. The most accurate timing of the MIP particle is provided by the arrival of the first photons. The optimum discriminator threshold for timing measurement is expected to be between 5 and 50 photoelectrons depending on the signal yield and dark count rate.

Signal rates

The signal (particles) rates seen by the ASIC inputs depend on the energy threshold. We define MIPs, which are the particles whose digitized information we want to record, as those with an average energy deposit $E > 1.0$ MeV. The maximum average rate (signal rate) of MIPs is 2.5 MHz per channel. They are randomly distributed over the bunch crossings.

All particles depositing an energy $E < 1.0$ MeV are considered as background particles. For energies below about 100 keV we expect that the signal will be too small to cross the timing threshold, but particles at mid energy range of $0.1 < E < 1.0$ MeV, called low energy particles, while crossing the timing threshold constitute a background that requires special attention to reject it. We estimate the rate of these background particles with $0.1 < E < 1.0$ MeV to be 5.0 MHz. Like the signal rate these background particles are randomly distributed over the bunch crossings. The total average rate and the decay time of the LYSO:Ce crystals are such that pileup of signals will occur requiring the implementation of a dedicated pileup cancellation circuitry.

The analog processing of the input signals is accomplished within 25 ns such that the ASIC is ready for new hits at the next bunch crossing. Digitization (energy and time) of the MIP particles (see below) and transfer of the information of the MIPs should be done while losing at most a few per cent of these signals at maximum signal and dark count rates.

Dark Count Rate

The rate of dark counts (SiPM signals corresponding to the detection of a single photon, or multiple, in case of so-called cross talk in the photo-sensors) is estimated to be as high as 60 GHz at the end of operation of the detector, corresponding to a current of 8×10^{15} electrons/s (1.3 mA) for a gain of 1.4×10^5 of the photo-sensor, resulting in baseline fluctuations and dark count noise. These fluctuations and noise will have a significant impact on the BTL timing performance requiring a dedicated noise cancellation circuitry mentioned above. This circuit is expected to reduce the high frequency dark count noise by a factor of the order of two, reducing the DCR contribution to the timing resolution by the same factor.

Amplitude and timing measurements

The TOFHIR2 ASIC is expected to measure the arrival time of the MIP signals adding a jitter of $\sigma = 14$ ps to the timing of the signal, including all intrinsic contributions of the ASIC. The input impedance of the analog inputs of the ASIC is low ($<$10 Ω) and the analog bandwidth is 350 MHz, to achieve the required timing performance.

Timing and amplitude measurements are performed continuously. The threshold of the timing discriminator is adjustable in the range of (0–100) photoelectrons with a precision of 6 bits. A dedicated discriminator identifies MIPs providing an adjustable threshold (MIP threshold) in a range of (0 to 1/2) MIP, with a precision of 6 bits.

If a signal is identified as a MIP, the amplitude and the two times, corresponding to the timing threshold and the MIP threshold (rising or falling edge selectable by configuration), are digitized and stored as hit data on the ASIC.

The TDC provides two counter values:

- A coarse counter, counting the number of cycles of the external 160.32 MHz reference clock. The length of this counter is 16 bits.
- A fine counter, providing the fine resolution within one period of the 160.32 MHz reference clock.

The time-binning of the time to digital converter circuit (TDC) is 20 ps. The ASIC is expected to have Differential Non-Linearity (DNL) $< \pm 2$ ps and Integrated Non-Linearity (INL) $<$20 ps, and a precision of the time binning of 2 ps rms to minimize the dependence from in-situ TDC calibration. In-situ TDC calibration will be performed with the code density method using random data from ^{176}Lu natural radioactivity in LYSO:Ce crystals.

The charge of the hit signal is required to be linear over the full dynamic range with a precision of 2%. This precision assures that the so-called time-walk effects, i.e. dependence of the measured signal timing on the total energy deposit in leading-edge discriminator methods, are corrected to better than 5 ps. The ADC has a resolution of 10 bits. DNL and INL should be less than 0.5 and 2.0 LSB, respectively. The effective number of bits should be at least 8.2. The ASIC provides a means to calibrate the ADCs with an external DC voltage source. In-situ calibration is performed using the ASIC internal test pulses and the Lu emission photo-peaks.

Output data and data links

The ASIC provides digitized timing and amplitude data, described above as hit data. Each hit is clearly associated with a clock cycle number. The data is sent as twelve 8/10B symbols. The first symbol is K28.5. The other 11 symbols represent 88 bits of data, with the following content:

- Bits 0-4: channel identifier;
- Bits 5-4: identifier of the time-to-amplitude converter in multi-buffer TAC;
- Bits 15-6: charge measurement;
- Bits 25-16: fine counter of the 2nd time measurement;
- Bits 35-26: fine counter of the 1st time measurement;
- Bits 45-36: coarse counter of the previous event crossing the timing threshold;
- Bits 55-46: coarse counter of the end of charge integration;
- Bits 65-56: coarse counter of the 2nd time measurement;
- Bits 81-66: coarse counter of the 1st time measurement;

- Bits 85-82: status of the trigger bits for previous event;
- Bits 87-86: trailing bits "11".

The ASIC provides two differential electrical output data E-links, operational at 320 Mbit/s. Both types of data, hit and trigger data can be transferred on both E-links. The selection of what data go to which of the E-links is configurable.

The ASIC also provides one differential electrical configuration input E-link, operational at 80 Mbit/s. This link allows the writing of all configuration registers in the ASIC. All configuration is received as a serial data stream in a single shift register. The configuration data can be read via the ASIC output E-links. The ASIC has an identification code configurable by four external pins. Moreover, the ASIC is able to handle fast control commands received in two other input E-links. The L0 and L1 signals discussed below are received in the trigger E-link. The second E-link receives the following commands:

1. RESET, resets all timing coarse and fine counters, FIFOs, and state machines
2. RSYNC, resets the timing coarse and fine counters

The ASIC provides two operation modes for data transmission:

1. Self-triggering mode
2. External trigger mode

In self triggering mode, the ASIC pushes all digitized signal hits to the output data link or links, depending on the output link configuration. In external trigger mode, the ASIC receives trigger signals via a fast control E-link. The ASIC decodes the trigger signal and distinguishes two different trigger types, called L0 and L1.

Both trigger signals have a delay, $L0_{delay}$ and $L1_{delay}$, programmable in number of 160.32 MHz clock periods, up to 20 μs. In case of L0 trigger, the hit corresponding to the clock cycle L0 minus $L0_{delay}$ is transferred via the selected E-link and at the same time stays available on the ASIC for a potential L1 trigger requesting the hit data of the same clock cycle later. On reception of L1 trigger, the ASIC sends the hit corresponding to the clock cycle L1–$L1_{delay}$ via the configured E-link. The ASIC has one 24 bit counter per channel. The counters are configured to count occurrences of a programmable logic combination of the discriminators output states. Data of the counters are transferred via the same output E-links as all other data.

Test pulse injection

The ASIC provides a test pulse circuitry, common to all channels, permitting the injection of a current pulse into each of the input channels, at the input nodes. The test pulse amplitude is adjustable with a precision of 6 bits over the full dynamic range.

Electrical characteristics

The ASIC electrical characteristics are listed in Table 2.7. The ASIC provides 32 channels with single-ended signals. The 32 inputs are connected as single ended signals with a common reference. The polarity of the signals is positive.

Environmental conditions: temperature and radiation

The nominal operating temperature of the BTL will be $-30\,^{\circ}$C. The exact typical operation temperature of the ASIC is not yet determined but is expected to be close to $0\,^{\circ}$C. We require the

Table 2.7: BTL FE ASIC electrical characteristics.

Term	Electrical characteristics			
Technology	TSMC 130 nm CMOS			
	Min	**Typical**	**Max**	**Unit**
Number of channels	32	32	32	
Operation voltage: V_{in}	1.08	1.2	1.32	V
Power consumption			15	mW
External sampling clock frequency		160.32		MHz
Signal polarity	Positive			
Input signal connection	Single ended			
Packaging	BGA 12x12 mm, 196 balls, pitch 0.8 mm			

ASIC to be fully operational and maintain its specifications in a temperature range of $-40\,^{\circ}$C to $80\,^{\circ}$C. The BTL is embedded into the CMS detector between the barrel outer tracker and the barrel electromagnetic calorimeter. We expect a total integrated dose received by components inside the detector at the end of operation, corresponding to an integrated luminosity of 3000 fb^{-1} of 30 kGy. The total expected flux of neutrons of energy E $>$ 1 MeV after 3000 fb^{-1} is 1.9×10^{14} cm^{-2}.

In order to achieve the required radiation tolerance, the following rules are applied in the ASIC design:

- The ASIC is designed with only a subset of the standard cell library as specified by CERN. This subset has been qualified by CERN to have sufficient radiation tolerance for the BTL application.
- The ASIC uses CERN's design of a qualified bandgap reference circuit to generate reference voltages.
- All logic in the ASIC, like state machines, configuration registers, etc., are protected against single event upsets (SEUs) using triple modular redundancy (TMR). The data themselves and thus the data path does not require SEU protection.

2.3.1.2 FE ASIC design

TOFHIR is an ASIC for high rate SiPM signals with timing and energy digitization in each channel and configurable dynamic range. The ASIC block diagram is represented in Fig. 2.21.

The TOFHIR ASIC has 32 independent channels, each containing independent amplifiers, discriminators, time-to-digital converters and charge-to-digital converters, as shown in Fig. 2.22. The input pre-amplifier (P) provides a low impedance input (R_{IN}) to the sensors output current signal. The input current I_{IN} is then replicated into three branches: T, E and QDC blocks in Fig. 2.22.

Amplifiers and pulse filtering

The preamplifier shown in Fig. 2.23 is a current conveyor based on a regulated common-gate trans-impedance amplifier (TIA) [59]. It is an upgraded version of the preamplifier proposed in Ref. [36], redesigned and optimized to extend the dynamic range while preserving the timing performance. The conveyor provides a low input impedance for the detector and a high impedance current output. With a power consumption of 5 mW a low-frequency amplification of 25 dB and bandwidth of 350 MHz are achieved.

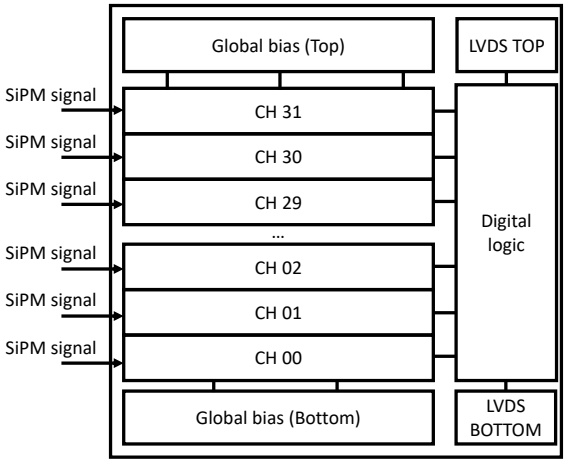


Figure 2.21: TOFHIR block diagram.

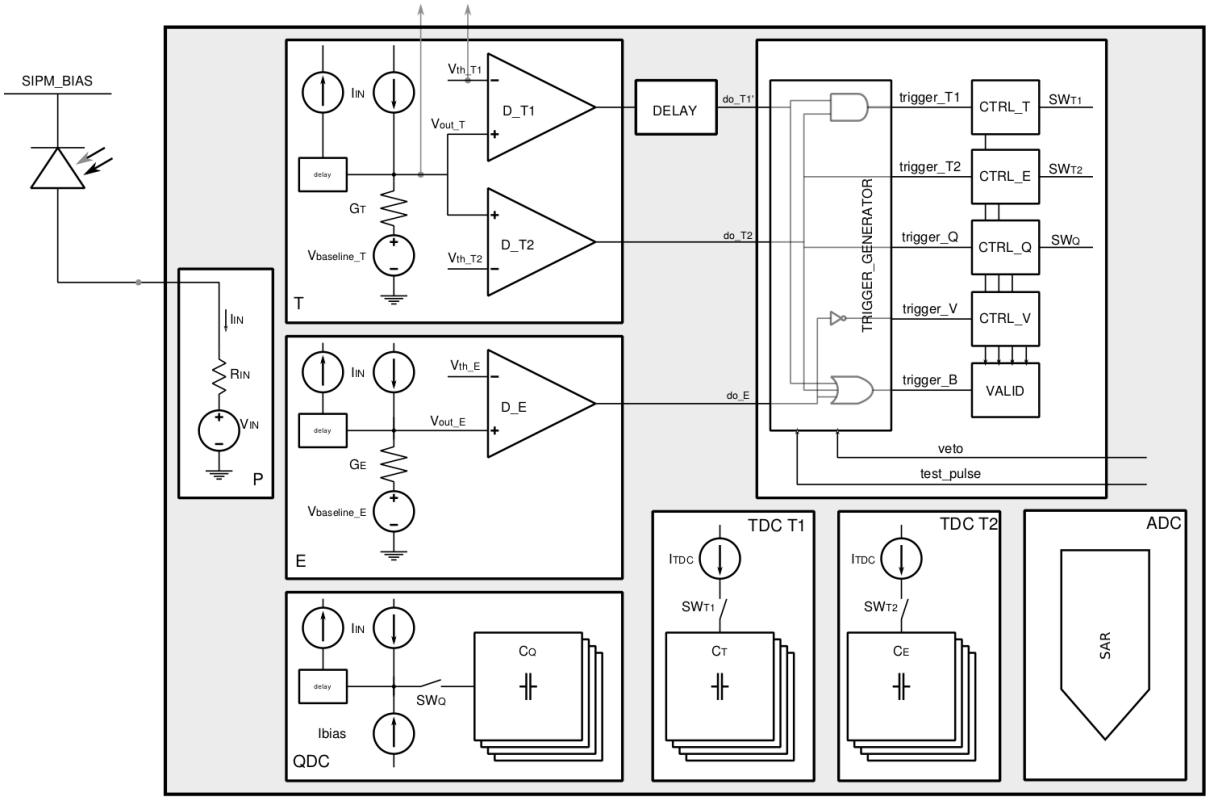


Figure 2.22: TOFHIR channel architecture including the preamplifier (P), the post-amplifier and discriminator blocks for timing (T) and energy (E) measurements, the charge to digital converter (QDC), the TDCs T1 and T2, the channel's SAR ADC, and the internal trigger logic.

The T and E branches have TIAs with configurable gains associated to modules for baseline stabilization and noise cancellation. The outputs are fed into discriminators for timing measurement (T1 and T2) and hit selection (E). Each TIA is a PMOS current mirror replicating the AC part of the preamplifier current output on a resistor. The QDC branch integrates a replica of the input current, which can then be digitized by the channel's ADC. All branches integrate pulse filtering as described below. The three branches share the same 10 bit, 40 MHz SAR ADC.

In order to mitigate the SiPM dark count noise and to stabilize the baseline, the TOFHIR2

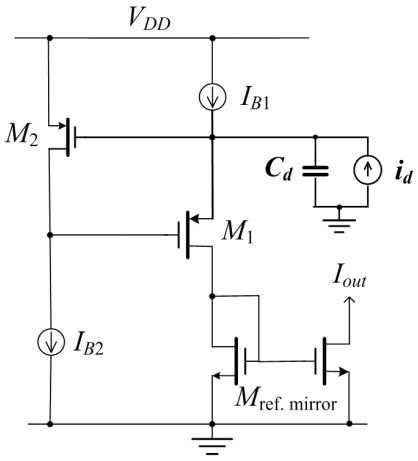


Figure 2.23: Simplified schematic of the preamplifier circuit.

will include a new module. The output of the TIA is fed into the baseline stabilization and noise filtering module, which has an architecture similar to the first stage of a constant fraction discriminator (CFD). The module reduces the noise due to dark counts and removes the tails of LYSO:Ce pulses by adding to the signal at the output of TIA an inverted and delayed copy of the signal, as shown in Fig. 2.24. The delay is of the order of 200–400 ps. The result is a bipolar signal with the baseline removed.

The schematic of this circuit, shown in Fig. 2.25, is being developed and simulated. For the delay element, we use series of RC elements that create an approximation of a transmission line. We have studied different configurations of RC series elements with modifications on their values but all having four delay elements.

Discriminators

The TIA T has a programmable gain G_T and offset $V_{baseline-T}$. The amplifier saturates for $V_{out-T} > V_{sat}$. The output V_{out-T} connects to two identical discriminators, D_{T1} and D_{T2}, whose threshold voltages, V_{th-T1} and V_{th-T2}, are set by 6-bit DACs. The offset $V_{baseline-T}$ is set by another 6-bit DAC and is used to trim the baseline of V_{out-T} relative to the input of the discriminators. The LSB of the T1 and T2 thresholds are set by global ASIC settings. The output signals from the discriminators are used to control the ASIC internal triggering logic.

Branch E is similar to branch T, with the following differences: the gain G_E of the TIA is lower allowing a higher range of signals before saturation, the delay of the CFD circuit is one order of magnitude larger to account for the larger signal peaking time and its output feeds a single discriminator D_E.

The outputs of the three discriminators connect to the Trigger Generator logic box which generates trigger signals. The output of discriminator T1 passes through a configurable delay line. Due to this delay, the time of the rising edge of trigger T is actually the time of the rising edge of D_{T1} + DELAY.

TOFHIR implements a multi-level event trigger and rejection scheme. In the nominal operation mode:

- Events which do not trigger threshold V_{th-T2} are rejected without any dead time.
- Events which trigger V_{th-T2} but not V_{th-E} are rejected with less than 25 ns dead time.
- Only events that trigger all the three thresholds are considered valid and digitized.

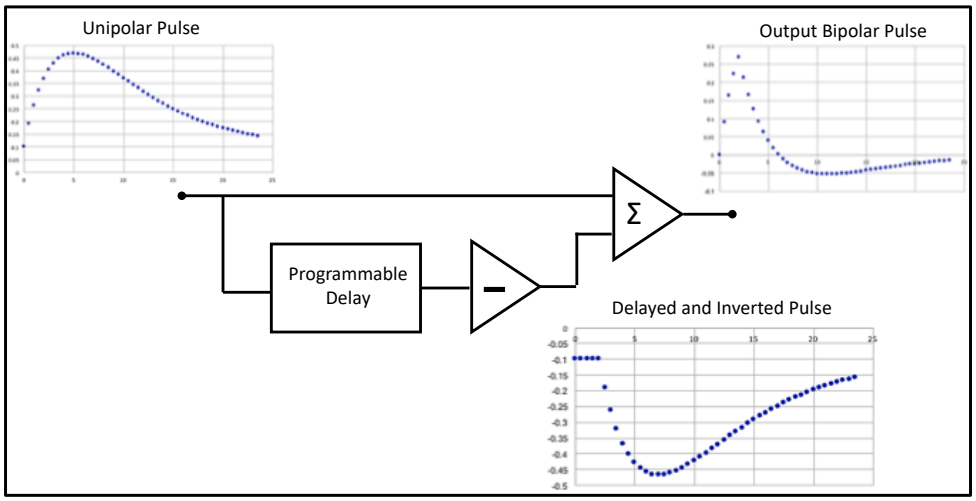


Figure 2.24: Principle of the baseline stabilization and noise filtering module.

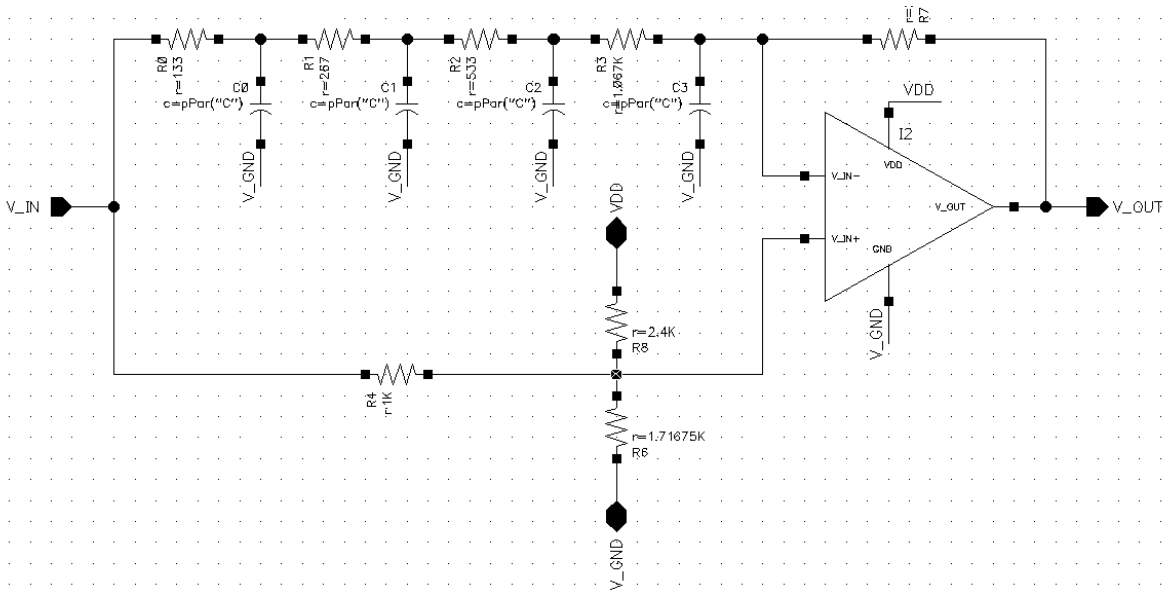


Figure 2.25: Schematic of CFD-like filtering circuit with RC delay.

TDC and QDC

The TDC T1 measures the time of the rising edge of trigger T1 (Fig. 2.22). The TDC T2 measures the time of the rising edge or falling edge (programmable) of trigger T2. The two time measurements can be used to estimate the width of the signal. The TDC is composed of a time-to-amplitude converter (TAC) followed by the analog-to-digital converter (ADC). The TDC time quantization is 20 ps.

The QDC measures the integrated charge from the rising edge of trigger Q (generated by the output of discriminator T2) until the end of the integration window. In normal operation, the integration window is set to 4 clock cycles (25 ns). The QDC is composed of a charge-to-amplitude converter (QAC) followed by the analog-to-digital converter (ADC). As mentioned, the two TDCs and the QDC share the same ADC.

The TDC operation is illustrated in Figs. 2.22 and 2.26. On the rising edge of trigger T1, switch SW_{T1} closes, charging analog buffer C_T with current I_{TDC}. On the 2nd next rising edge of CLK,

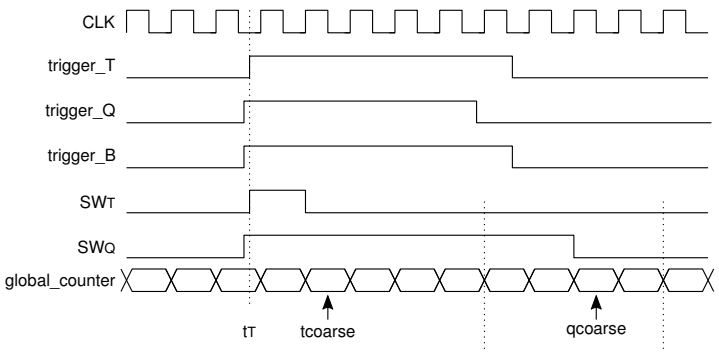


Figure 2.26: Trigger signals involved in the TDC and QDC operation (see Fig. 2.22).

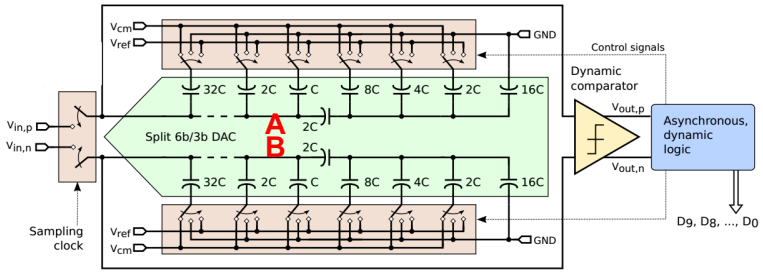


Figure 2.27: SAR ADC with fully-differential capacitive DAC.

SW_{T1} opens, stopping the charging process. The value of a global clock counter is latched on the next clock providing value t_{coarse}. If the event is valid ($E > V_{th-E}$), the voltage stored in C_T will be digitized as t_{fine}. The TDC T2 operates identically, using trigger T2 signal.

The QDC operation is illustrated in Figs. 2.22 and 2.26 as well. On the rising edge of trigger Q, switch SW_Q closes, charging analog buffer C_Q with a replica of the input current signal plus a DC current. SW_Q opens on a rising edge of CLK when the integration time has been reached. The value of a global counter is latched on the next clock providing value q_{coarse}. If the event is valid, the charge stored in C_Q will be digitized. The gain of the integrator is configurable.

Each channel has 6 sets of C_T, C_E, and C_Q analog buffers where analog values are stored before being digitized. Every time the channel goes through rearm state, a new buffer set is selected round-robin. This allows the channel to rearm without waiting for the previous event to be digitized. When an event is valid and digitized, the TAC-id number of the analog buffer set used to process that event is transmitted along with the event. Due to process variations, each analog buffer should be calibrated for optimal results. The content of the analog buffers shift slowly due to leakage effects, which affect the time and charge measurements. In order to keep this effect under 0.1 LSB, whenever the channel has been in READY state for 100 μs, an invalid event is triggered causing the channel to rearm with the next, fresh, set of buffers.

SAR ADC

A SAR ADC with 10-bit resolution and a sampling frequency of 10 MHz was implemented in TOFHIR1 replacing the slow Wilkinson ADC of TOFPET2 and allowing the operation of the ASIC at high rate. The design is based on a fully-differential capacitive DAC, shown in Fig. 2.27, as proposed in Ref. [60] providing high noise rejection, twice the dynamic range (2×1.2 V) and low power consumption. The SAR ADC is operated in asynchronous mode performing the conversions when requested by the digital logic.

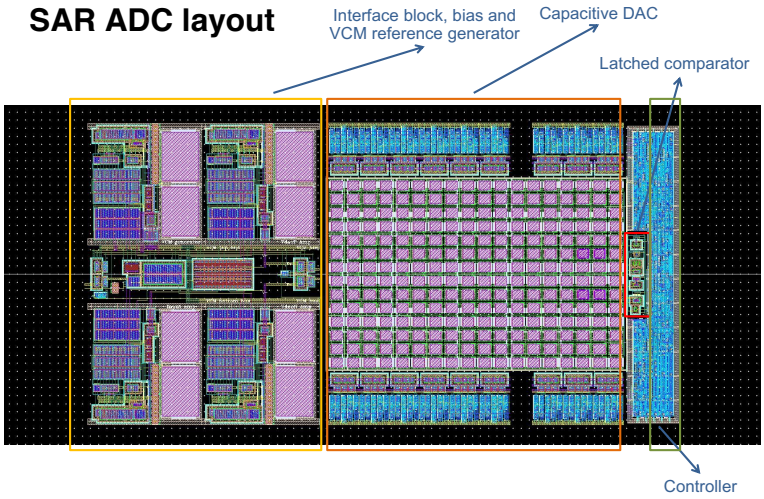


Figure 2.28: Layout of the SAR ADC.

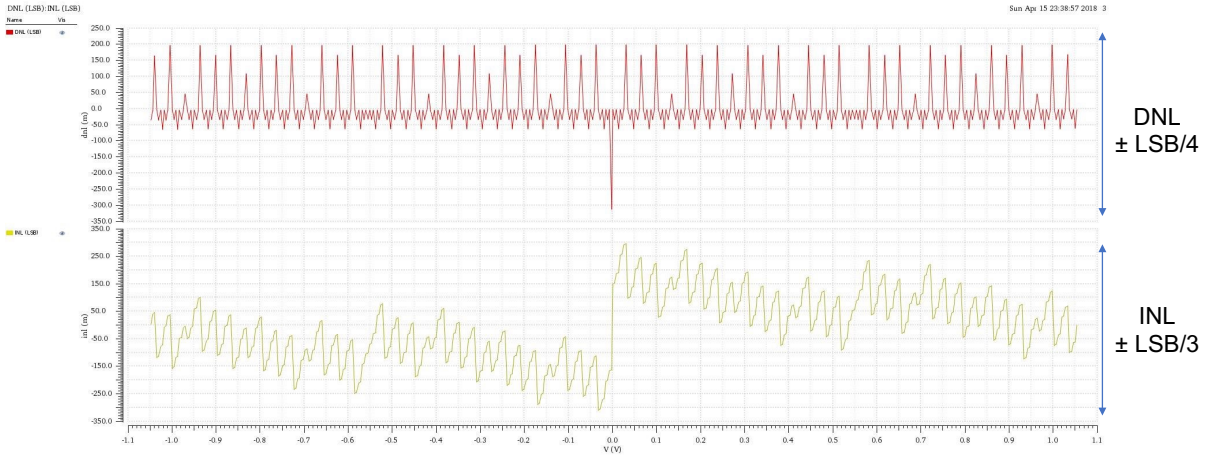


Figure 2.29: DNL and INL obtained from the post-layout simulation of the capacitive DAC.

The layout of the SAR ADC implemented in TOFHIR1 (UMC CMOS 110 nm) is shown in Fig. 2.28. Figure 2.29 shows the ADC linearity performance obtained from the post-layout simulation of the capacitive DAC. We observe that DNL is $\pm$LSB/4 and INL is $\pm$LSB/3. In the TOFHIR2 chip, an existing TSMC 130 nm implementation of the same SAR ADC architecture validated at sampling frequency of 40 MHz will be used.

Digital I/O

The differential I/O standard in TOFHIR1 is LVDS. Level shifting will be needed to interface the lpGBT. The final TOFHIR will implement the E-link I/O specification (CLPS). Table 2.8 lists TOFHIRs digital I/O.

TOFHIR includes event counters, which allow events to be counted without being transmitted (and thus, not subject to event conversion and transmission limitation). In this context, events are programmable logical combinations of the discriminator outputs. The counting period is globally configurable up to 2^{24} CLK cycles.

TOFHIR has a circuit which injects an analog test pulse directly into channels input node. This pulse has an adjustable amplitude and is generated in response to a digital test pulse provided

Table 2.8: TOFHIR Digital I/O.

Name	Function	TOFHIR1	TOFHIR2
CLK	Main Clock	LVDS	CLPS
RESYNC	Global SYNC		
TRIGGER	L1/L0 trigger information		
DRX	Configuration Input		
DTX[0]	Data/configuration output		
DTX[1]			
DTX[2]			N/A
DTX[3]			
TEST_PULSE	External test pulse trigger. Not used in final system		CLPS
RESYNC_EDGE	Select latching on rising or falling clock edge	LVCMOS 1.2 V	LVCMOS 1.2 V
TRIGGER_EDGE			
DRX_EDGE			
CHIP_ID[0]	Chip Configuration address		
CHIP_ID[1]			
CHIP_ID[2]			
CHIP_ID[3]			

by the digital control logic. The circuit is shared by all the channels and each channel has a switch to select whether or not that channels input node is connected to the global analog test pulse generator.

The Test Pulse pad receives an external digital pulse that can be used to trigger directly the internal logic or to generate an internal analog pulse at the channels input nodes. In the final system, Test Pulse will not be used. Instead, an internal test pulse generated synchronously with the clock (configured via DRX) will be used. Unlike the test pulse provided by an external FPGA, its phase relative to the clock cannot be adjusted in the lpGBT.

In TOFHIR1, the bias blocks are inherited from TOFPET2 and the reference voltages (800 mV and 500 mV) used by different circuits are provided by external pins. Nevertheless, global voltage bias distribution may lead to large spread from channel to channel requiring more trimming DACs to mitigate variations. In TOFHIR2 the global bias distribution will be done as currents providing better matching since mirror transistors are close together. The bias distribution as currents is immune to voltage drop and is less sensitive to noise.

In order to have all channels on this chip tightly matched, reference voltages and currents should have a very low spread (below $\pm 5\%$) in all operating conditions: process, voltage, temperature (PVT) and under radiation. Precise voltage references stable under temperature variations and after irradiation are provided by the bandgap developed at CERN, shown schematically in Fig. 2.30. At $-30\,^{\circ}\mathrm{C}$ the voltage variation after 3 MGy TID is of the order of 1%. This bandgap has trimming fuses that can adjust the reference voltage to the nominal value, compensating process parameter variations. However, precise current references (which are of utmost importance since most TOFHIR analog circuits are biased in current) are difficult to design since these current references are obtained from voltage-to-current converters that use resistors. Polysilicon resistors may have sheet resistance variations of $\pm 60\%$, generating similar spreads on the reference current. Therefore, a bias current calibration circuit, shown in Fig. 2.31, is being developed based on the matching between the voltage drop on a precise external resistor and the bandgap voltage.

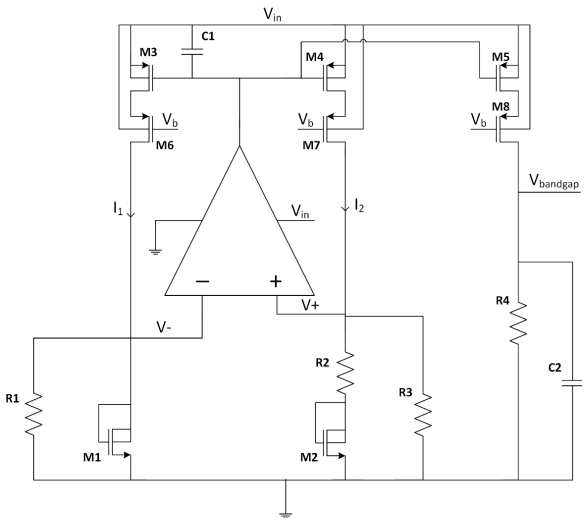


Figure 2.30: DTNMOS-based bandgap developed by CERN Microelectronics.

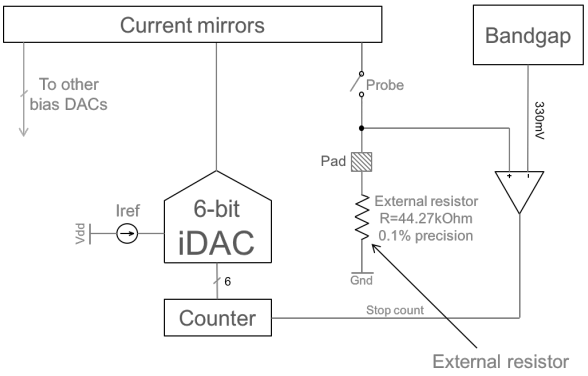


Figure 2.31: Bias current calibration circuit.

Radiation Tolerance

Measurements of the radiation effects on 130 nm CMOS n- and p-channel MOSFETs from three different manufacturers have been reported in the literature [61]. Even though the effects of TID are qualitatively similar, requiring special care in the design of the circuits, the amount of degradation is shown to vary considerably from foundry to foundry being stronger for UMC. Irradiations with X-rays of the TOFPET2 ASIC have been performed. After 5 kGy irradiation, the TDC range was reduced by a large factor implying a degradation of the time resolution by a factor three. The effect was interpreted as due to increased leakage in the TAC-write transistor. This effect can be cured by a redesign with larger saturation margin to accommodate the V_{th} variation due to radiation. On the other hand, the TOFPET2 irradiations indicated little effect on the front-end amplifiers.

TOFHIR1 was designed in UMC 110 nm technology and no particular design rules for radiation tolerance have been used. The next version TOFHIR2 will be implemented in TSMC 130 nm technology, which is less affected by radiation (smaller shift of transistor threshold voltages and smaller leakage current). Additionally, design guidelines for radiation tolerance provided by the CERN microelectronics group are being followed. These include limitations in the W/L transistor dimensions, increased transistor threshold margins, usage of 1.2 V transistors only (the usage of 2.5 V I/O transistors is not allowed), and circuit validation in all PVT corners. These rules are expected to be sufficient for doses much higher than the total dose foreseen

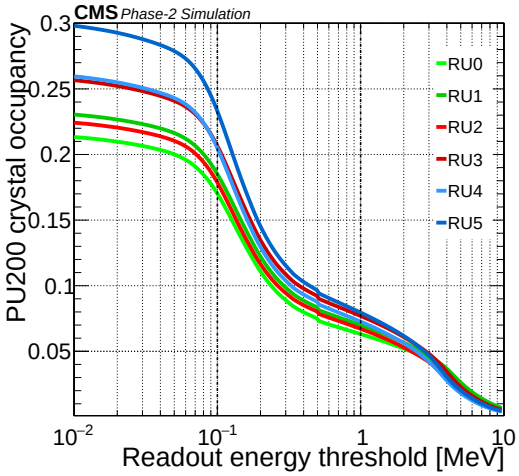


Figure 2.32: Channel occupancy as a function of the energy threshold for different RUs along rapidity.

in BTL. On the other hand, the preamplifier will use enclosed layout transistors to increase its radiation hardness. Additionally, the TOFHIR2 digital logic, including state machines, configuration registers, and synchronization counters, are protected against single event upsets (SEUs) using triple modular redundancy (TMR). Dedicated irradiation tests of TOFHIR2 ASICs will be performed to validate its radiation tolerance at the required level.

2.3.1.3 Rate requirements

The rate requirements of the BTL readout system are driven by the expected occupancy per channel, which is shown in Fig. 2.32 for crystal bars oriented along ϕ. The curves correspond to the rapidity regions for each of the six readout units in a tray. At higher rapidity, the channel occupancy increases due to inclined tracks crossing neighboring bars.

In the BTL all the hits above the timing threshold have to be processed. The timing threshold is defined by the first n_e photoelectrons (p.e.) detected in the SiPM after the particle crossed the sensor, optimizing the timing resolution of the sensor for a given set of operation conditions. The timing threshold is expected to be set in the interval $n_e = 5$ to 50 p.e. depending on SiPM PDE and DCR. Taking into account the LYSO:Ce light decay time ($\sim$40 ns) and the expected number of photoelectrons per MIP particle ($\sim$10 thousand), it is expected that the first photons used for timing arrive within $\sim$100 ps after the particle arrival time (neglecting the rise time of the light signal in the crystal $\sim$100–150 ps). Therefore, in the operation of the ASIC front-end, we may consider that the first n_e photons arrive simultaneously. On the other hand, given the pulse shape of LYSO:Ce hits, the timing threshold is crossed by hits with an integrated number of photoelectrons that is ten times higher than n_e, as shown in Fig. 2.33. Therefore, all hits in the crystal depositing more than $\sim$20 keV need to be processed by the front-end analog circuitry. The rate of these hits estimated from the occupancy shown in Fig. 2.32 is about 10 MHz.

The TOFHIR ASIC introduces two leading edge discriminators with threshold T1 and T2, such that the timing can be measured with the lowest threshold T1 and pulses below T2 are rejected at an early stage. The threshold T2 could be set at a level between 100 and 200 keV depending on SiPM operating conditions, corresponding to rates between 7.5 and 5 MHz. The timing of pulses selected by T2 are measured with T1 by delaying the output of the first discriminator by a few ($<$3) ns.

In TOFHIR, the time to amplitude conversion (TAC) is triggered by T2. The TAC ramping is

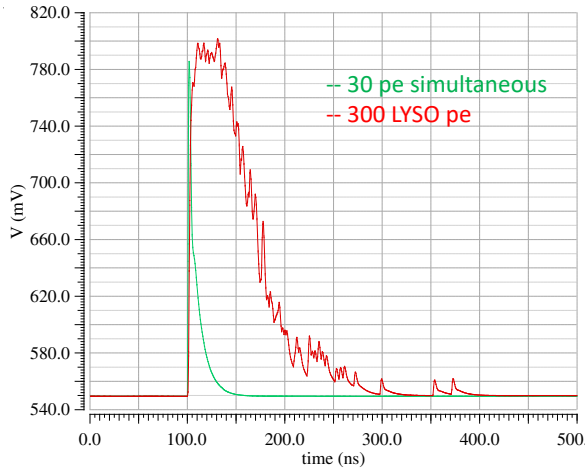


Figure 2.33: Comparison of shape and amplitude of pulses arising from LYSO:Ce scintillating light and from simultaneous detection of photoelectrons.

done between the T1 signal and the 2nd next rising edge of CLK, therefore it takes on average 1.5 clock periods (9.4 ns). At the end of conversion, the digitization of the TAC voltage is performed by the SAR ADC. In TOFHIR2, which has a SAR ADC with a sampling rate of 40 MHz, the full digitization process takes 50 ns. The digitization time is sufficiently below the average time between MIP particles (400 ns) and the Poisson fluctuations of the MIP rate are handled by the multi-TAC design described in the previous section. The simulation of the TDC (TAC+ADC) operation assuming the MIP rate of 2.5 MHz and the low energy hit rate of 5 MHz yields the following results for the MIP losses:

- 4 TAC: 1.89% MIP loss;
- 5 TAC: 0.50% MIP loss;
- 6 TAC: 0.08% MIP loss.

Given these results, TOFHIR2 is being designed with 5 or 6 TACs depending of the silicon overhead. Digitized hits are stored in the output FIFO, waiting for the L0/L1 trigger decision.

The ASIC output rate is determined by the MIP rate and by the L1 trigger rate. For the highest expected channel occupancy, the MIP rate per channel is 2.5 MHz as described before and the L1 trigger rate is 750 kHz. Given the number of input channels per ASIC of 32 and the event size of 120 bits, we estimate the output data rate from the ASIC to be 230 Mb/s. Given the 24 ASICs integrated in the Readout Unit, the total rate per RU is expected to be 5.5 Gb/s.

2.3.1.4 Front-end board

The FE board receives signals from 192 SiPMs that are processed by six TOFHIR ASICs. The SiPM signals are AC coupled to the ASIC using 100V-rated capacitors that isolate the circuit from the large SiPM leakage current and protect the ASIC from being damaged by the bias voltage in case of SiPM failure.

The TOFHIR ASIC is powered by a single 1.2 V supply. This voltage is regulated and filtered by the ALDO2 ASIC, the low voltage regulator resistant to radiation described in the next section. The ALDO2 delivers up to 500 mA current, which is sufficient for the operation of one TOFHIR ASIC.

In the baseline design, the distribution of the SiPM bias voltage is provided by external bias

channels, each channel serving half RU (384 SiPMs). Each bias channel has two main wires and two sense wires transmitted over the services cables, described in Section 2.5. The bias channel wires are connected by shunt cables between two neighbor FE boards.

The FE board interfaces to the Concentrator Card through board side-to-side connectors transferring the following signals:

- Clock: six lpGBT e-clocks (160 MHz), one per TOFHIR ASIC ;
- Data readout: two up E-links (320 Mb/s) per TOFHIR connecting to 2 lpGBT in the CC;
- Configuration: one down E-link (80 Mb/s) for TOFHIR configuration shared by six TOFHIRs;
- Sync/reset: one down E-link (80 Mb/s) for TOFHIR resync/reset shared by six TOFHIRs;
- Trigger: two down E-links (80 Mb/s) for TOFHIR L0/L1 trigger, each link shared by three TOFHIRs;
- Monitoring: provision for six temperature sensors per FE board (SiPMs and electronics temperatures) and 12 SiPM bias current sense signals (one per group of 16 SiPMs);
- Power and ground.

Each FE board is served by one FEASTMP DC-DC converter providing 1.8 V. The lpGBT provides up to 28 upstream links (160 or 320 Mbit/s) and up to 28 clocks (160 MHz), but only 16 downstream links (80 Mbit/s). This allows each TOFHIR to receive a dedicated clock from the lpGBT and to transmit data over one dedicated uplink to one lpGBT. The six TOFIRs in the FE board share three downstream links used to provide the TOFHIRs with configuration, synchronization and trigger, as illustrated in Fig. 2.34.

In order for various TOFHIRs to share the configuration downlinks, the configuration protocol includes a 4-bit address. Each TOFHIR chip is given its 4-bit address though dedicated ID pins in the chip, which can be connected to a jumper or hardwired in the PCB. To implement reliable reception of the downstream links from the lpGBT, the TOFHIR chip can use either the falling edge or the rising edge of CLK to latch the input signals. The edge selection is made through a dedicated pin, RX CLK EDGE. Figure 2.35 shows the preliminary layout of the FE board.

2.3.1.5 Regulator ASIC

The voltage supplied to the TOFHIR ASIC is regulated by the ALDO2 ASIC. The ALDO2 low voltage regulator will be an evolution of the ALDO1 ASIC [62], a radiation tolerant 250 mA low dropout linear regulator in 0.35 μm AMS CMOS technology that was developed for the LHCb RICH upgrade. This chip implements the typical topology of a low dropout regulator, with a bandgap voltage reference, an error amplifier driving the PMOS pass transistor, and external compensation with a low ESR capacitor, shown in Fig. 2.36. Preliminary tests of the regulator were successfully performed with the TOFPET2 64-channel chip and four ALDO1 in parallel in order to match the current requirement of the TOPFET2. These results are described in Section 2.3.2.

The next revision of the chip, ALDO2, will require moving to ON Semiconductor I3T80 0.35 μm CMOS technology due to end-of-operation of the AMS technology that was used for ALDO1. To increase the output current capability to at least 500 mA, the ALDO2 will feature a larger PMOS pass transistor. The ALDO2 will operate with a minimum input voltage of 1.6 V and

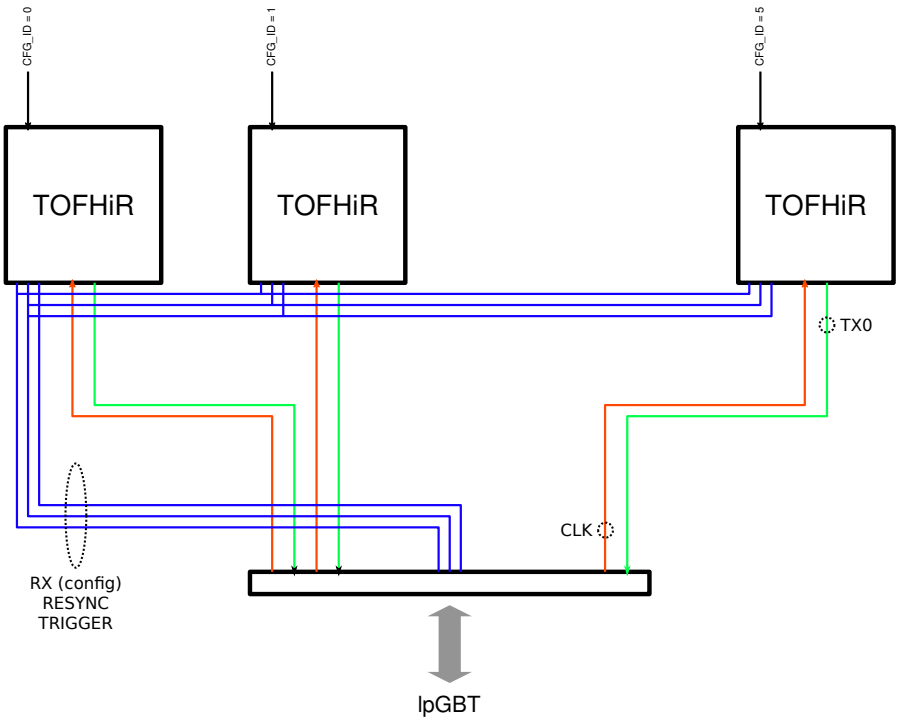


Figure 2.34: TOFHIR E-links in FE board.

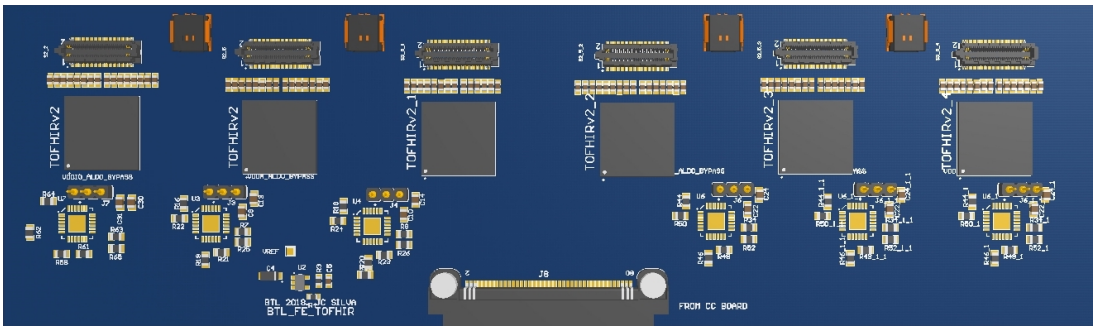


Figure 2.35: Layout of the FE board (dimensions 59.0 $\times$ 206.0 mm^2).

a minimum dropout of 400 mV. Like for the ALDO1, the regulator will be equipped with two bandgap voltage references, one based on vertical bipolar transistors (with the addition of base current compensation circuitry), and the other on PMOS transistors in dynamic threshold (DTMOS) configuration. The choice of which bandgap to use can be done externally and will be defined by the compromise between its stability after irradiation and the precision of the reference voltage.

Base current compensation consists in injecting the base current of a matched dummy bipolar transistor into the emitter of the main bipolar transistor, in order to compensate for the increase of base current after irradiation exposure, which is the limiting effect when using bipolar-based bandgaps at the high radiation levels expected in BTL. The device will be protected against over-current and over-temperature by means of a foldback circuit that limits the output current when either current or temperature limits are reached.

The technology adopted for the ALDO2 will be the same as the FEAST ASIC, whose radiation hardness is more than adequate for the radiation levels expected in BTL environment. The

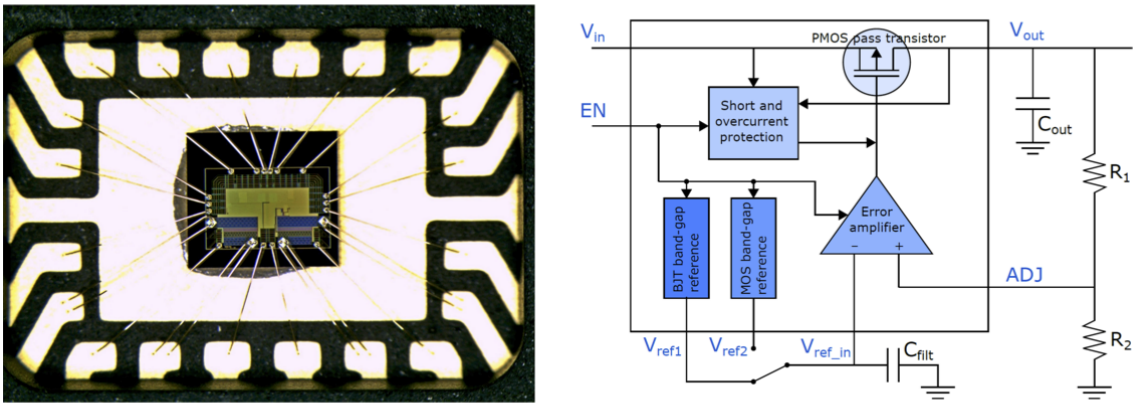


Figure 2.36: Photograph (left) and block schematic (right) of the ALDO1 ASIC.

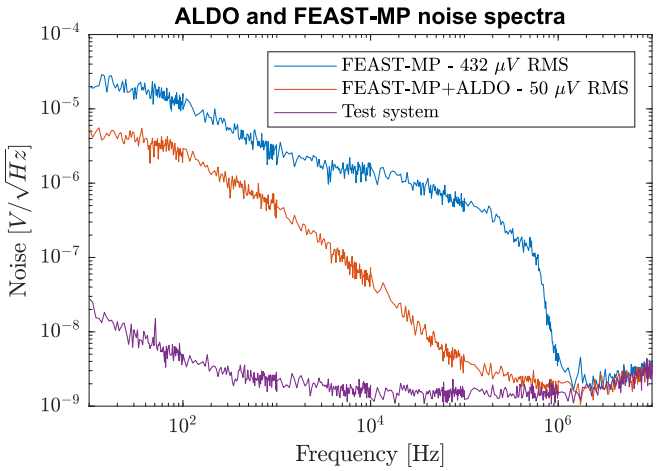


Figure 2.37: Improved noise performance with ALDO1.

low voltage NMOS transistors will be implemented with an enclosed layout, to increase their radiation hardness and minimize leakage. The TOFHIR ASIC requires an 800 mV reference voltage with 5 mA input current. The baseline is to generate this reference inside the TOFHIR2 ASIC, but in case this cannot be achieved with acceptable performance, the ALDO2 will be able to provide this reference, using a second error amplifier and a smaller output transistor.

Figure 2.37 shows the frequency spectra of the power supply noise measured at the output of the FEASTMP DC-DC regulator (blue) and with the use of ALDO1 at full load (red).

2.3.1.6 Concentrator card and power distribution

The Concentrator Card is designed to interface the system readout with four FE boards. Its location with respect to the sensor modules, cooling bar, and FE boards on the BTL tray is shown in Fig. 2.63. The CC uses the lpGBT to provide an interface between 24 E-links and a VL+ optomodule (single channel receive and single channel transmit). The command downlink (receive) will be at 2.56 Gb/s and the data uplink will be at 10.24 Gb/s. The average rate of each of the 24 data links from the TOFHIR ASIC chips is 230 Mb/s, estimated for 8% occupancy, 120 bit per hit, 0.75 MHz L1 rate. This translates to an aggregate rate of about 5.5 Gb/s, which can be transmitted by two lpGBTs at its lower output bandwidth of 5.12 Gb/s or by one lpGBT at the higher

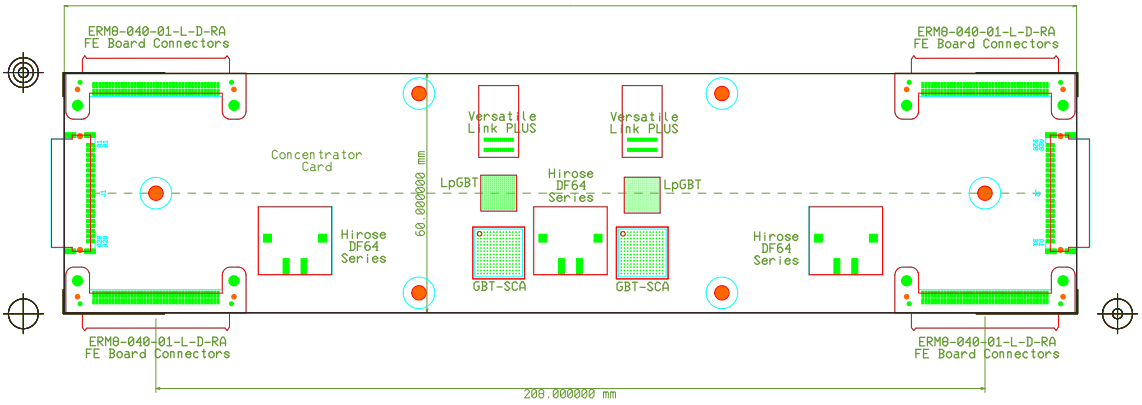


Figure 2.38: Layout of the dual lpGBT/VL+ Concentrator Card.

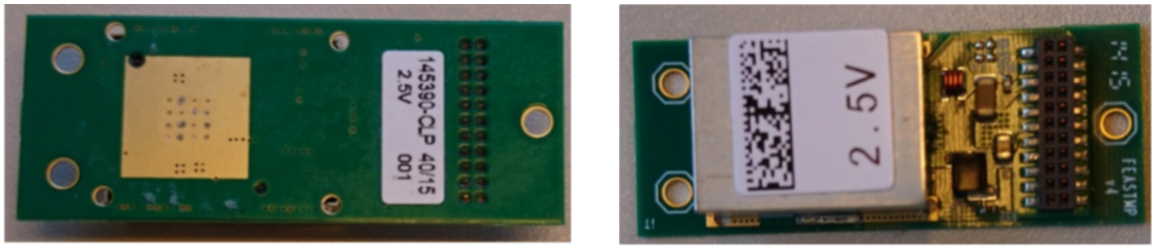


Figure 2.39: FEASTMP_CLP view from below (left) and above (right) used for power-only prototype design of the Concentrator Card.

output bandwidth (10.24 Gb/s). The current design has a dual lpGBT/VL+ combination. In addition to providing an extra bandwidth for increased data rate, it provides a redundancy for failures of links to off-detector and provides a bandwidth for possible L0 trigger data. Another important feature of the lpGBT is to ensure the precise clock distribution, received from the downlink, to the front-end system, achieved by the high frequency clock noise filter in the PLL. A schematic of the CC is shown in Fig. 2.38.

The power is distributed from the DC/DC converter module based on the FEASTMP_CLP, Fig. 2.39, radiation tolerant ASIC that was developed at CERN and is used throughout CMS. To power the CC components as well as the FE cards, six such converters are needed. For the full channel capacity, we will use two Power Converter Cards each one integrating three FEASTMP_CLP. The CC will also utilize two GBT-SCA chips for slow control and monitoring.

2.3.1.7 Slow control and monitoring

The lpGBT provides a set of slow control and monitoring features that nevertheless are insufficient for the RU needs. Therefore, we foresee using two SCA-GBT chips providing added functionality for control and monitoring. The GBT-SCA ASIC developed at CERN provides the following functionality:

- 16 I2C master controllers;
- 1 JTAG master controller;
- 1 group of 4 8-bit fully programmable and bidirectional IO ports;
- 1 memory-like bus master controller with 8 bit data and 16 bit address;
- 32 analog channels, converted to 12-bit values via an ADC port;
- 4 DAC 8-bit ports;

- 4 asynchronous external interrupts.

The preliminary specifications of the slow control and monitoring functions are the following:

1. LV control:
   - use of FEASTMP-enable pin for turning ON/OFF power of individual FE boards,
     - FE board converters are OFF by default and switched on sequentially,
   - CC converters are ON by default.
2. LV monitoring:
   - power good signals of the FEASTMPs,
   - input voltage ∼10 V,
   - output voltages of 6 DC-DC converters.
3. Temperature monitoring:
   - 16 SiPM temperature sensors per RU,
   - 1 temperature sensor per FE card,
   - 1 temperature sensor per PCC,
   - 1 temperature sensor per CC,
4. SiPM bias current monitoring:
   - 48 bias currents, one per group of 16 SiPMs (option under study).

2.3.1.8 Clock distribution

The distribution of a precise clock to the front-end system is a major requirement for BTL. The clocks available in the backend system are recovered from the lpGBT down links. The high frequency clock noise is expected to be filtered by the PLL in the lpGBT. Low frequency clock jitter and possible phase instability, in particular arising from temperature variations or low-frequency response of the clock chain, will require special attention.

A dedicated R&D effort across CMS sub-projects is being done to find the best solution for precise clock distribution at the level of the whole CMS detector. While waiting for the results of these investigations, we foresee the need to accommodate two dedicated clock fibres per tray in case the option of using a tree of separate clock paths turns out to be necessary. Concentrator Cards evenly located along the tray will house the fibre receiver and a dedicated radiation tolerant clock fan-out chip to distribute the clock over short coaxial cables to neighbor modules. This possibility is not implemented in the first prototype of the CC presently under engineering design, but may be considered in the final CC prototype if needed. More details on the MTD clock distribution are provided in Section 4.2.

2.3.1.9 Power converter card

We plan to use the step-down converter module FEASTMP_CLP with a radiation tolerant ASIC (FEAST) developed at CERN to implement the first prototype of the Power Converter Card (PCC), shown in Fig. 2.40. In the final implementation three DC-DC FEAST ASICs will be integrated in a single power board. Given the tight space constraints we are studying the possibility of using a slimmer converter with a solenoid coil instead of the standard toroid, shown in Fig. 2.41. We estimate that this option would allow a gain of 3 mm in height. Measurements

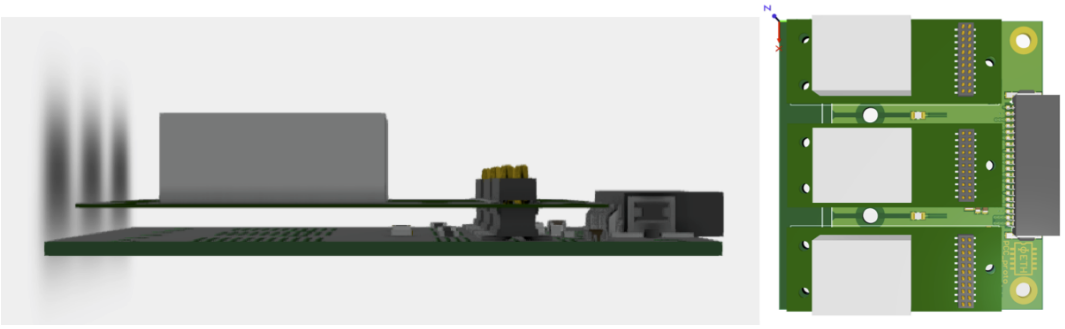


Figure 2.40: Power Converter Card (first prototype).

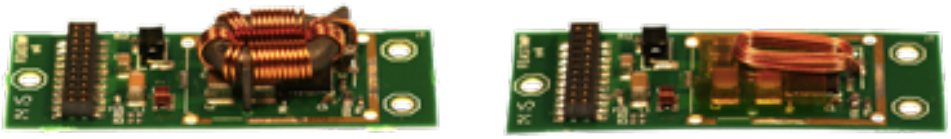


Figure 2.41: FEASTMP_CLP converter modules with toroid coil (left) and solenoid coil (right).

of FEASTMP with solenoid coil show an efficiency of 76–80% for output currents of 1 to 3 A, which is about the same as with the toroidal coil. However, the solenoid shows increased electromagnetic radiation by about 10 dB (measurements without shielding). The optimization of the converter module for use in BTL is under way.

In order to have a more stable and precise regulation of the analog power supplies and to better filter the switching noise introduced by the FEASTMP DC-DC without using large passive components that would not fit in the tight space available, we plan to use the active low dropout linear regulator ALDO2 in series to the DC-DC converter. The regulator will have to provide an output voltage of 1.2 V and an output current of at least 500 mA, as required by the TOFHIR ASIC. To keep the power budget as low as possible, the regulator should operate with the lowest possible dropout voltage. The target was set to 400 mV dropout, which corresponds to a minimum input voltage of 1.6 V and thus to an efficiency of 75%. The regulator will be hosted on the FE card, close to the TOFHIR chip, minimizing pick-up noise and improving load regulation. The temperature dependence of the output voltage will be tuned in order to reach the highest stability at the expected working temperature of the system with a residual drift below 20 ppm/°C over the whole operating range, between $-30\,^{\circ}$C and ambient temperature. The temperature stability of the FEASTMP is not as optimized. The power supply rejection ratio of the linear regulator should be at least 40 dB at the DC-DC switching frequency (about 2 MHz).

2.3.1.10 SiPM bias voltage distribution

The baseline distribution of bias voltage is presented in Section 2.5. An on-going development is studying the design of a programmable HV bias voltage regula or on groups of 16 channels. This regulator would be included in the ALDO2 ASIC and would eliminate the need for SiPM matching, to better compensate any breakdown voltage variation that could arise after irradiation, thermal gradients, etc, and to protect against SiPM failures (i.e. too large dark current and shorts). The matching of 16 SiPMs is assured by the manufacturing process, since they are fabricated on the same wafer area.

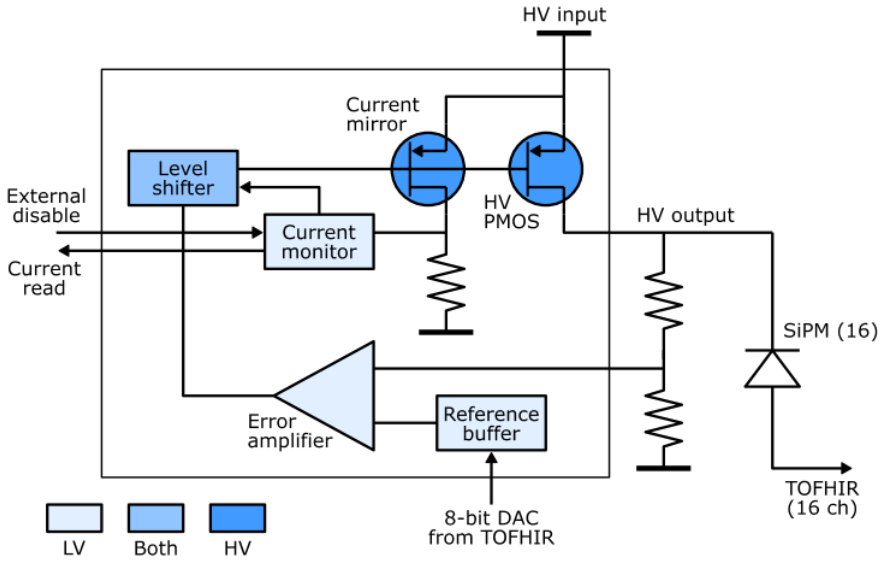


Figure 2.42: Schematic of the SiPM bias adjustment circuit.

The topology adopted for this HV regulator is shown in Fig. 2.42 and is similar to the one implemented in the ALDO2 ASIC for the low voltage regulation. The main differences will be the use of a HV (80 V) PMOS output transistor, a level shifter circuit for driving the HV PMOS gate, and a programmable voltage reference. The programmable voltage reference will be provided by the TOFHIR, by means of an 8-bit DAC, and will allow a regulation of the bias voltage with a range of 0–1.2 V over a pre-set value, with a resolution of 10 mV. To match the modularity of the system, two independent regulators are included in each ALDO2. The technology used in the bias regulation circuit is I3T80 0.35 μm. It is the same technology as used for the first versions of the FEAST ASIC which showed adequate tolerance to the radiation levels expected in BTL. The HV MOSFETs in this technology have been tested up to 30 V after irradiation [63].

The HV PMOS transistor has a breakdown voltage of 80 V, thus each bias output can be disabled completely, switching off noisy or broken groups of 16 channels that would otherwise compromise the operation of the other 368 channels on the same bias line. This feature can be triggered externally by the TOFHIR, or automatically if the bias current exceeds a pre-set threshold of a few tens of mA. The bias current measurement is also fed to the GBT-SCA in the CC and can be monitored remotely.

2.3.2 Performance

2.3.2.1 Silicon-proven performance

The front-end TOFHIR ASIC is based on the existing TOFPET2 chip for reading out LYSO:Ce crystals via SiPM sensors with timing performance close to the BTL needs. In this section, we report the performance measurements of the TOFPET2 chip [58] involving the blocks that are reused in TOFHIR.

TOFPET2 is a 64 channel ASIC based on CMOS 110 nm technology provided by the UMC foundry with timing and energy branches for each channel and a dynamic range configurable between 150 and 1500 pC. The timing branch consists of amplifiers, discriminators and a TDC

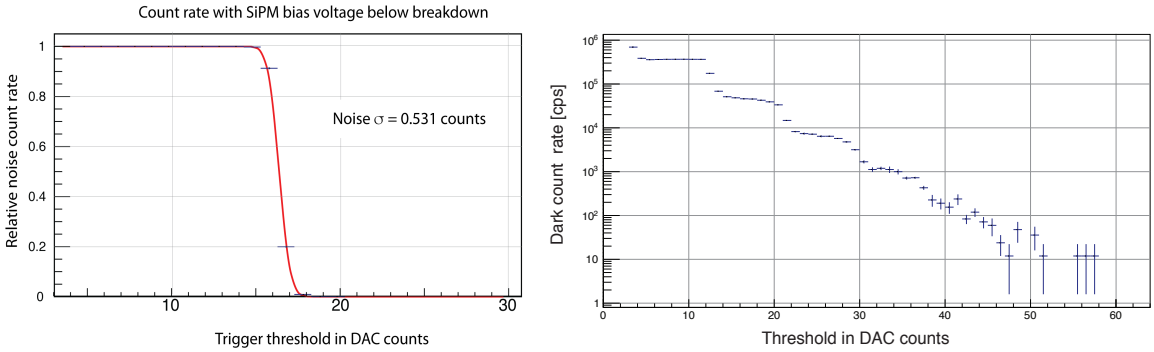


Figure 2.43: Left: count rate with SiPM bias voltage below breakdown as a function of the timing discriminator threshold. Right: dark count rate with SiPM biased at 3.5 V over-voltage as a function of the timing discriminator threshold (in DAC counts).

(30 ps time binning) whereas the energy branch includes an amplifier and a charge integrator. The TDC is based on four-fold Time to Amplitude Converters (TAC) followed by a Wilkinson ADC. The maximum conversion time is of the order of 1 μs. Four TACs per channel allow de-randomizing the input signals. The Charge Integrator has a similar four-fold structure as the TACs followed by one Wilkinson ADC. The maximum rate per channel of 0.6 Mhits/s is limited by the output links.

The intrinsic performance of TOFPET2 was characterized by making use of the ASIC internal test features. In some cases, detailed below, the characterization was done with SiPMs connected to the ASIC inputs. The characterization of the electronics noise was performed with a threshold scan above the baseline at the output of the timing post-amplifier. The ASIC inputs were connected to SiPMs. For the measurement of electronics noise, the SiPMs were biased slightly below the breakdown voltage such that the SiPM capacitance is close to nominal (320 pF). We measure a noise fluctuation of $\sigma_{\text{noise}} = 0.53$ DAC counts corresponding to $\sim$1.6 mV when using the S13361-3050AE-04 MPPC (left plot in Fig. 2.43). The same technique was used to measure the average amplitude single photon pulses. With the same SiPM S13361-3050AE-04 MPPC biased at 3.5 V over-voltage the amplitude of the signal of 1 p.e. given by the width of the first plateau is around 9 DAC counts or 27 mV (right plot in Fig. 2.43).

The linearity of the TDC is measured with random pulses that follow a uniform distribution in time. Photons from a ^{22}Na source measured with a LYSO:Ce crystal and a SiPM are used for this purpose. Figure 2.44 shows the Differential Non-Linearity (DNL) and the Integrated Non-Linearity (INL) as a function of the TDC code. The DNL is less than 0.2 LSB and the INL is less than 1 LSB.

The performance of the TDC when multiple ASIC channels are active was evaluated by comparing the time measured by one channel TDC as a function of activity in the other ASIC channels. Figure 2.45 shows the difference of the time measurements in one channel when test pulses are distributed simultaneously to 1 or 64 channels. The test pulse is synchronous to the clock and a phase scan was performed with steps of 236 ps over 8 clock periods. The average difference is smaller than $\pm$1 LSB (30 ps).

The timing performance of the full ASIC channel was evaluated with test pulses. Analog test pulses emulating detector signals are induced in an internal test capacitance at the input of the pre-amplifier, using an external digital pulse. The total capacitance of the test injection circuit is 112 pF. The rise time (10–90%) of the test pulse is 8 ns, similar to the pulses from a LYSO:Ce scintillating crystal. In this measurement, the amplitude of the test pulse is equivalent to about

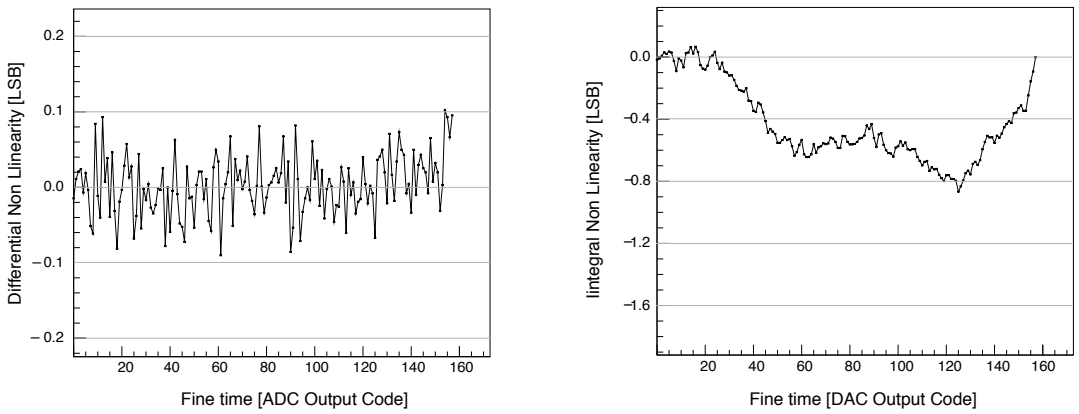


Figure 2.44: TOFPET2 TDC differential non-linearity (left) and integrated non-linearity (right).

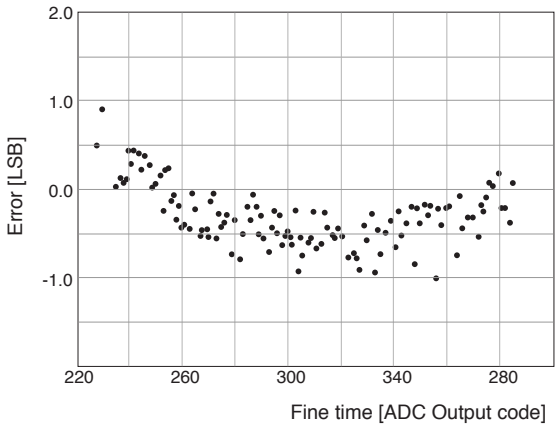


Figure 2.45: Difference of the TDC time measurements when 1 or 64 ASIC channels are active.

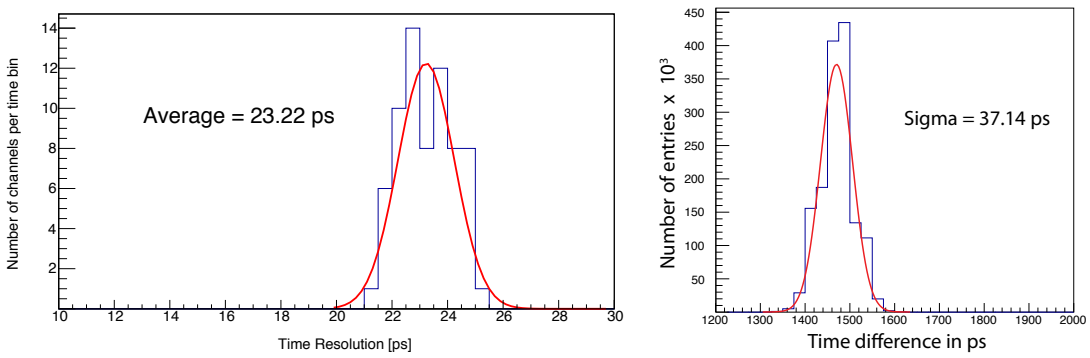
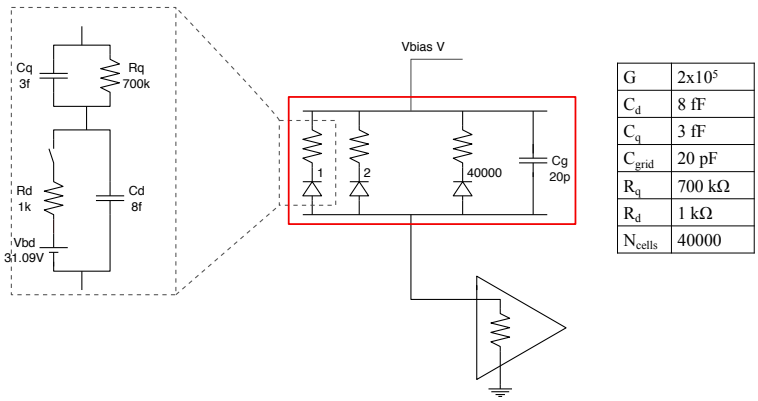


Figure 2.46: Left: full channel time resolution measured with internal test pulses. Right: resolution of the time difference of two neighbour channels illuminated with a laser pulse.

100 photoelectrons. The left plot of Fig. 2.46 shows the distribution of the time resolution measured in the 64 channels of the ASIC. The average is 23 ps (rms). The jitter of the test pulse and of the clock is not subtracted but it is estimated to be small (<10 ps).

The ASIC timing performance was also evaluated with laser pulses generated with the HPK PLP-10 Picosecond light pulser. The laser light was directed to two neighbouring SiPMs of



G	$2x10^5$
C_d	8 fF
C_q	3 fF
C_{grid}	20 pF
R_q	700 kΩ
R_d	1 kΩ
N_{cells}	40000

Figure 2.47: SiPM equivalent electrical model assumed in the TOFHIR simulations.

the array HPK S13361-3050AE-04. The amplitude of the signal corresponds to about 100 p.e. detected in each SiPM. The resolution of the time difference between the two channels is 37 ps which corresponds to a time resolution of 26 ps (rms) per channel, as shown in the right plot of Fig. 2.46. The SiPM contribution (unknown) is not subtracted. This result is representative of the TOFHIR expected performance since this chip uses the same architecture of the front-end amplifier and the same time to amplitude converter as TOFPET2. A power consumption of 8.5 mW was measured for each channel of the chip, giving confidence that the goal of a maximum of 15 mW in TOFHIR can be achieved.

2.3.2.2 TOFHIR simulated performance

In this section, we present results from the electrical simulations of the TOFHIR1 ASIC using Cadence simulation tools. The simulations include electronics noise. The ASIC inputs are connected (DC or AC) to the equivalent electrical model of the SiPM depicted in Fig. 2.47 using the parameters indicated in the figure. The presented results were obtained for SiPMs with 40 000 cells of 15 μm corresponding to 3×3 mm^2 devices.

The simulation reproduces individual photoelectrons detected in one of the SiPM cells producing a certain number of charges according the SiPM gain assumption. Unless otherwise noted, the results presented here were obtained for a gain of 2×10^5. SiPM dark counts were included when indicated assuming a uniform distribution of single photoelectrons with a certain average rate. The SiPM cross-talk was assumed to be 7% and included in the simulations of dark count noise. The timing for each photoelectron signal is smeared emulating the intrinsic SiPM jitter. We assumed a Gaussian distribution with 100 ps rms. The pulse shape of LYSO:Ce events results from the superposition of a certain number of photoelectrons randomly distributed in time according to the exponential decay of the crystal light. The left-hand plot of Fig. 2.48 shows the SiPM output current pulse for an event with 9000 p.e. detected in a 3×3 mm^2 SiPM with a gain of 2×10^5 and dark count rates (DCR) of 1, 5, 20, and 60 GHz. The right-hand plot of Fig. 2.48 shows for the same conditions the voltage pulses at the output of the timing trans-impedance amplifier. In this particular example the pulse saturates at an amplitude corresponding to about 100 simultaneous photoelectrons.

The sharp rising edge of the voltage pulse is critical to the time resolution. Figure 2.49 shows the slew rate of the pulse rising edge at the output of the post-amplifier as a function of the threshold level (in number of simultaneous p.e.) for the highest gain of the amplifier. The maximum slew rate is 770 mV/ns and the electronics noise rms is 2.0 mV, leading to a contribution of the electronics noise to the time resolution of 3 ps for a threshold of 13 photoelectrons. When

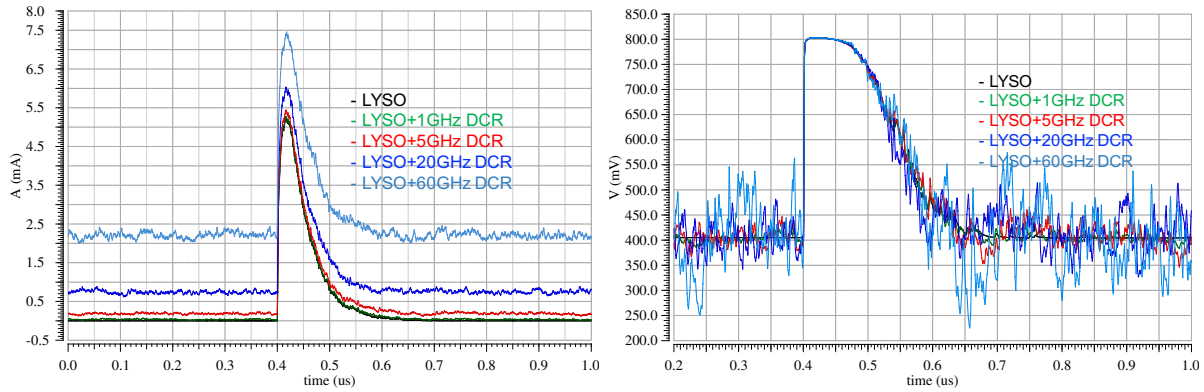


Figure 2.48: Left: simulated SiPM current pulse for 9000 p.e. and G $= 2 \times 10^5$ for four different dark count rates. Right: simulated voltage pulse at the output of the timing post-amplifier for 9000 p.e. and G $= 2 \times 10^5$ for four different dark count rates.

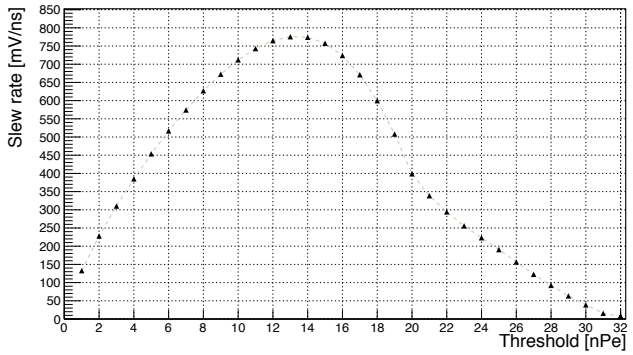


Figure 2.49: Pulse slew rate in mV/ns as a function of the discriminator threshold.

using the post-amplifier at 25% of the maximum gain, the electronics noise contribution to the time resolution is 12 ps. These results do not take into account the noise reduction achieved by the pulse filtering module.

To achieve the desired time resolution with the low threshold leading edge discriminator technique it is of paramount importance to control accurately the signal baseline. Fluctuations of the baseline due to dark counts significantly deteriorates the time measurements. TOFHIR2 includes the dedicated filtering circuit described in Section 2.3.1.2 to mitigate the high dark count noise after SiPM irradiation and the baseline drifts due to pileup of LYSO:Ce events. The simulation results presented here were obtained with a behavioral model constructed with ideal components. Figure 2.50 shows the pulse shapes of single photoelectron events at the input and output of the filtering circuit.

Figure 2.51 shows the time resolution expected for a LYSO:Ce signal of 9000 p.e. detected in the SiPM as a function of the discriminator threshold. The SiPM used in this simulation is 3×3 mm^2, has 40000 cells of 15 μm and a gain of 2×10^5. The simulation includes the LYSO:Ce photon statistics, assumes a single photon resolution of the SiPM of 100 ps (rms) and is obtained using a detailed electrical simulation of the TOFHIR1 circuit up to the output of the timing post-amplifier and the behavioral model of the filtering module. The TDC quantization of 20 ps was included. Clock jitter is not included. The rise time of the LYSO:Ce scintillating light signal (100–150 ps) and the spread of light propagation time in the crystal were also not included. An alternative circuit for baseline filtering using a second order high-pass filter (HPF) is also considered. We see that for a threshold between 5 and 10 p.e. a time resolution of the order of

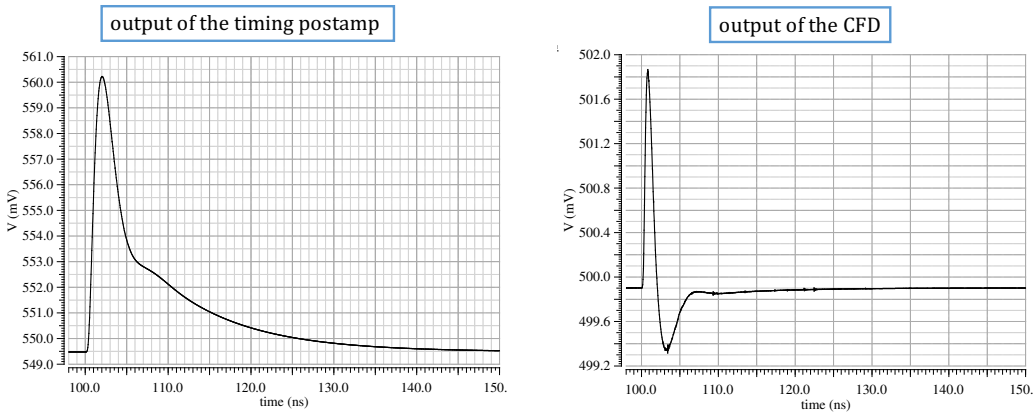


Figure 2.50: Pulse shapes of single photoelectron events at the input of the filtering module (output of the timing post-amplifier), on the left, and at the output of the filtering module, on the right.

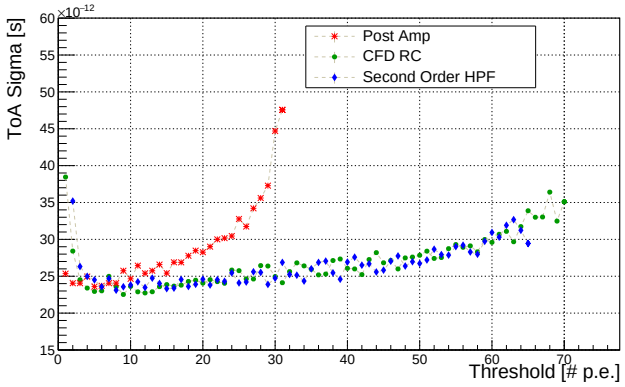


Figure 2.51: Time resolution at TOFHIR level as a function of the discriminator threshold, assuming zero DCR.

25 ps is expected.

The plots in Fig. 2.52 show the time resolution under the same conditions as above when adding random DCR events as a function of threshold for DCR = 20 GHz (left panel) and as a function of DCR for the optimal threshold (right panel). A time resolution of 35 ps for MIP particles is achieved for DCR = 20 GHz, deteriorating to 55 ps for DCR = 60 GHz.

The input signal has a large dynamic range due to Landau fluctuations, slant thickness of the crystals and SiPM over-voltage adjustment during the detector operation time. In Fig. 2.53 we show the linearity of the charge integration QAC as a function of the number of photoelectrons in the LYSO:Ce signal assuming a highest gain of the SiPM of 4×10^5. The plot shows good linearity while integrating over a short window of 20 ns, for signals between 9000 and 36 000 photoelectrons, showing that the TOFHIR preamplifier copes with the dynamic range of the large BTL signals at the start of operation (high SiPM gain and PDE).

2.3.2.3 TOFHIR1 preliminary test results

The tests of the TOFHIR1 prototype started in mid January 2019. The chip was mounted on the TOFHIR1 Test Board shown in Fig. 2.54 which connected to the DAQ board, also shown in Fig. 2.54. The DAQ motherboard has a Kintex7 FPGA and clock filtering as well as low voltage conversion and regulation. The first mezzanine generates bias voltages to SiPMs op-

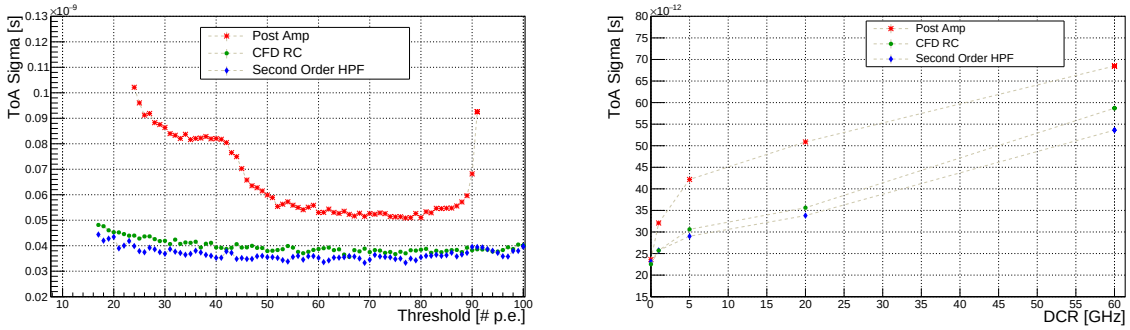


Figure 2.52: Time resolution at TOFHIR level as a function of threshold for DCR = 20 GHz (left) and as a function of DCR for the optimal threshold (right). Results for DCR = 0, 1 and 5 GHz are obtained with post-amplifier default gain, while results for DCR = 20 and 60 GHz are with post-amplifier gain 1/4.

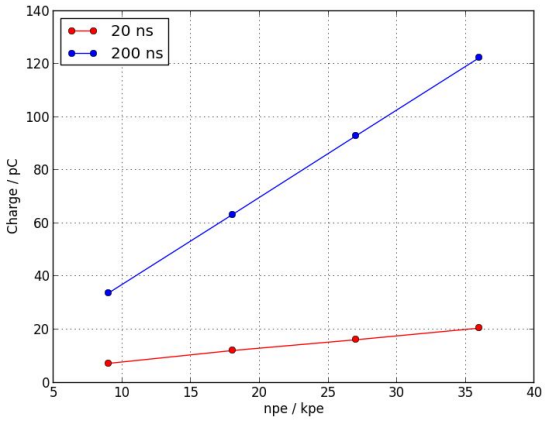


Figure 2.53: QAC integrated charge as a function of the number of photoelectrons in the LYSO:Ce signal.

tionally connected to the Test Board. The second mezzanine provides the interface to the DAQ computer. Dedicated firmware and software for TOFHIR1 configuration and readout has been developed.

At this stage only preliminary and incomplete results are available. The firmware in the DAQ board and the software in the data acquisition computer have been debugged allowing the communication with the chip to be established. The configuration cycle (command and reply) was validated. The chip responds to external test pulses with good events: event triggering, data transmission and reception has been validated. All channels trigger as expected with digital test pulses, and all data are properly received by the FPGA and software. The assessment of the SAR ADC was made using a dedicated test input that allows to scan the ADC input voltage. Figure 2.55 (left) shows the ADC code as a function of the input voltage. Good behavior and linearity of the ADC is observed. The measured ADC noise is 0.8 LSB.

The TDC assessment was made using external test pulses. The Kintex FPGA in the DAQ board allows the generation of test pulses with a precise phase relative to the system clock. This capability has been used to scan the test pulse timing in steps of 300 ps. For each phase step, a measurement of the pulse timing is performed. The TDC non-calibrated code (fine time) as a function of the test pulse timing is shown in Fig. 2.55 (right). The timing scan, performed over 8 clock cycles, shows the expected behavior and linearity of the TDC. The TDC time quanti-

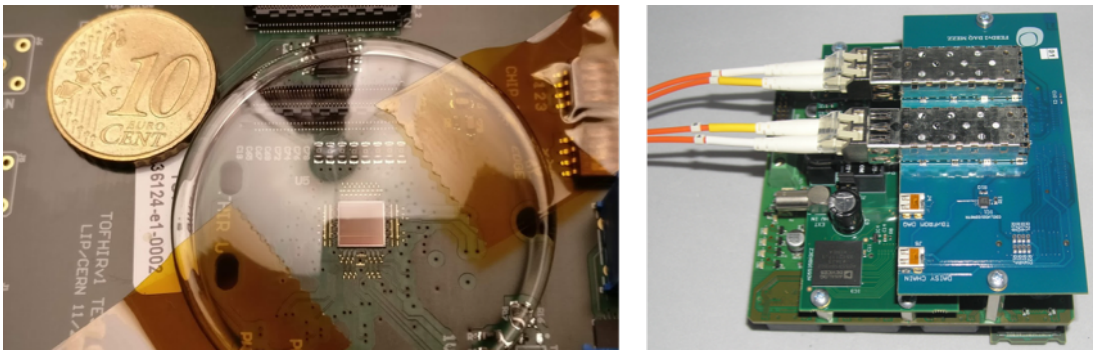


Figure 2.54: TOFHIR1 Test Board (left) and DAQ board (right) used in the tests.

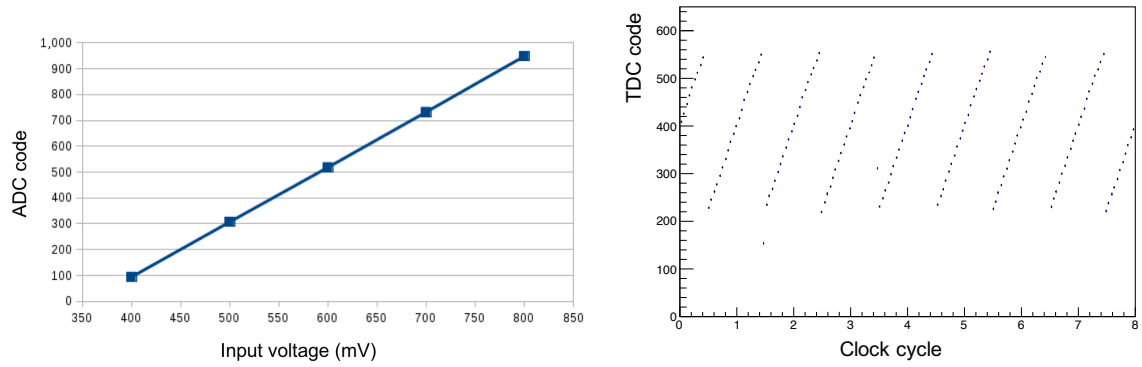


Figure 2.55: Non-calibrated ADC code as a function of the input voltage (left) and non-calibrated TDC code as a function of the test pulse timing (right).

zation bin size measured in one channel is 19.5 ps. It should be noted that channel to channel variations are expected and need to be calibrated.

The TDC time resolution was estimated by measuring the time difference between two channels triggered by a common digital test pulse. In this measurement, the phase of the test pulse relative to the system clock is fixed emulating the LHC conditions. The measured coincidence time resolution (CTR) is 21 ps from which we may estimate a channel TDC time resolution of 15 ps (Fig.2.56). We may conclude that in TOFHIR1 the TDC time resolution is dominated by the ADC noise quoted above. In TOFHIR2, we expect a TDC time resolution of about 10 ps given the lower noise (0.5 LSB) of the silicon proven ADC implementation in TSMC 130 nm technology that will be used.

First tests with SiPM signals at TOFHIR1 input have been performed using the 4×4 SiPM array HPK S13361-3050AE-04 with pixels of 3×3 mm^2 and 50 μm micro-cells. The SiPM was operated at gain of 1.75×10^6 with over-voltage of 3.5 V. The large gain relative to the BTL operation conditions allowed to perform the measurement of the Single Photon Time Resolution (SPTR) using short light pulses (width $<$ 50 ps) provided by a pulsed laser. The pulses at the output of the TOFHIR1 amplifier were observed in the oscilloscope using a spy output pad. The observed pulse amplitude and shape are well reproduced in simulation. The distribution of the time of arrival of single photoelectron events provided an estimation of the SPTR of 138 ps (rms). This value is consistent with the expected timing performance of the ASIC.

A total static power consumption of 220 mW was measured in TOFHIR1 corresponding to 13.8 mW per channel, which compares well with the simulation expectation (13 mW). The chip characterization is expected to be pursued until mid 2019.

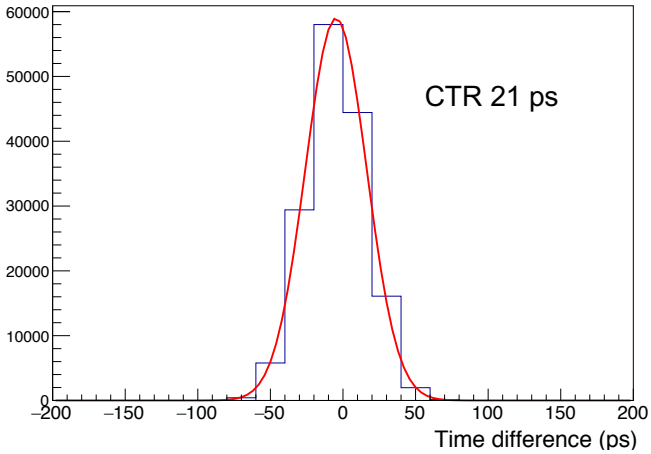


Figure 2.56: Time difference between two channel TDCs triggered by a common digital test pulse.

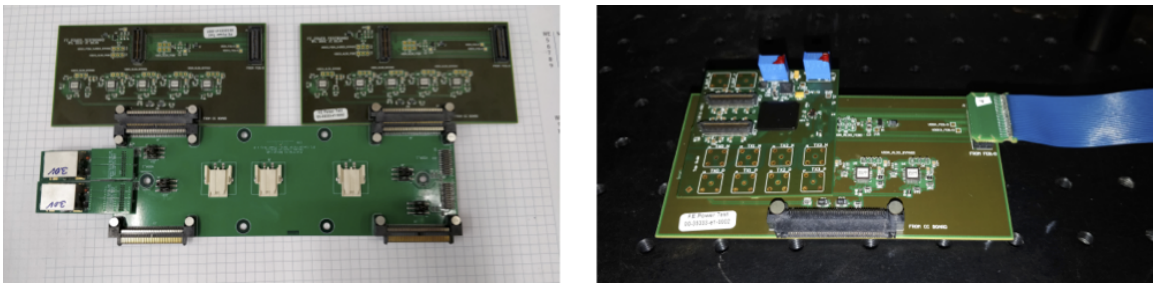


Figure 2.57: RU power prototype test boards: CC board connected to two FE boards on the top and two FEASTMP (left), and FE board with the TOFPET Test Board plugged in (right).

2.3.2.4 RU power distribution tests

A series of tests have been performed to study the influence of the power distribution noise on the timing performance. The test involves the BTL_RU_PowerProto which integrates power-only versions of the CC and FE boards, shown in Fig. 2.57. The CC board is connected to FEASTMP-CLP mezzanines replicating the expected final configuration in the BTL detector. In this prototype, the FE board includes the ALDO1 voltage regulators (one regulator for digital power and four regulators for analog power), a connector to the CC board, a connector for the TOFPET2 Test Board, a connector to the DAQ system, and jumpers allowing selection of the supply source. The detectors and the FE parts were placed inside a "cold box to guarantee good thermal stabilization. The timing performance was evaluated with test pulses induced at the input of the amplifiers in the TOFPET2 chip.

In addition to the data acquisition functions, the DAQ board used in these tests provides ASIC supply voltages using commercial switching DC-DC converters. On the TOFPET Tester Board, two linear voltage regulators (LDO) provide the 1.2 V and 2.5 V for the ASIC. The TOFPET2 also requires two reference voltages (0.8 V and 0.5 V) which are provided by other LDOs mounted in the TOFPET Tester Board. In all cases the reference voltages where provided by these regulators. In the tests, several power configurations were evaluated:

1. Supply from the DAQ board and LDO on the TOFPET Tester Board;
2. Supply from the FEASTMP DC-DC converters regulated by ALDO1 regulators in the FE

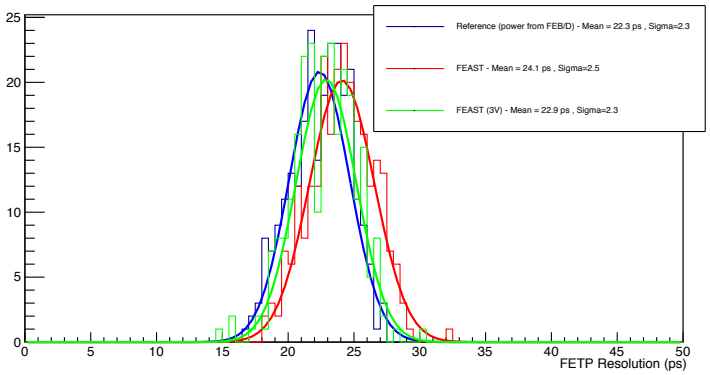


Figure 2.58: Distribution of the measurements of TOFPET2 intrinsic time resolution in three different power configurations.

board;

3. Supply directly from FEASTMP-CLP DC-DC converters.

The TOFPET2 ASIC allows the generation of analog test pulses at the input of each channel upon reception of an external digital pulse. The shape of the analog pulse emulates light pulses from a LYSO:Ce crystal converted by a 3×3 mm^2 SiPM. The timing of the test pulses relative to the system clock (200 MHz) is adjusted in the DAQ FPGA in steps of 100 ps. In the present tests, the measurements were made at four different phases of the test pulse relative to the clock (0, 25, 50, and 75%).

For each phase, the FPGA generates a large number of pulses at a frequency of $\sim$100 kHz. The time of each pulse is measured by the TOFPET2 TDC. The width of the distribution of the measured times provides the intrinsic time resolution of the TOFPET2 chip. For each power configuration, we have measured the time resolution of the 64 channels of the TOFPET2 ASIC in 4 different phases, for a total of 256 measurements of the time resolution. We have studied the effect of the FEASTMP-CLPs and ALDO1s in the TOFPET2 performance by comparing the time resolution measured in the power configurations defined above. The measurements obtained with the reference power provided by the DAQ board correspond to the blue points in Fig. 2.58. Then we made measurements with the FEASTMPs at 3.0 V in the FE board and 2.7 V and 1.35 V in the ALDOs. The ASIC voltages of 2.5 V and 1.2 V are produced by the LDOs in the TOFPET2 Tester Board. The measurements correspond to the green points in Fig. 2.58. Finally, we made the test pulse measurements with the voltages supplied directly by the FEASTMP converters adjusted at 2.5 V and 1.2 V, and bypassing the ALDO and the LDO regulators. Those measurements correspond to the red points.

The test pulse time resolution is 22.3 ps for the reference power configuration, 22.9 ps for the FEASTMP+ALDO power configuration and 24.1 ps for the FEASTMP-only power configuration. Relative to the reference power, the solution with DC-DC converter alone introduces a degradation of the time resolution estimated at 9 ps added in quadrature.

2.3.3 Production and testing

2.3.3.1 Plans for development of pre-production prototypes

The present BTL schedule foresees the production and validation of the TOFHIR1 based Readout Unit prototype (RUproto1) through the end of 2019. The architecture of this prototype is close to the final one, allowing meaningful studies of system issues like power distribution,

grounding and noise, as well as system configuration, data readout, etc.

The RUproto1 provides inputs for BTL SiPMs. Presently, 32-channel input connectors are foreseen. However only 16 input signals per connector will be processed by the 16-channel TOFHIR1 ASICs. Therefore, the RUproto1 will serve 384 channels instead of 768 channels as in the final version.

Besides the TOFHIR1 ASIC, the RUproto1 will make use of ALDO1 ASICs for low voltage regulation. Untested chips in a QFP package will be assembled in the FE boards, therefore ALDO1 validation is part of the FE prototype testing.

The design of the FE, CC and PCC boards is expected to be concluded soon. A first set of a few boards with ASICs in the FE, lpGBT and Versatile optical parts in the CC, and FEASTMP converters in the PCC is expected to be available in July 2019. In case a second iteration is needed it should become available by the end of October 2019.

In parallel, the design of a dedicated PCC for BTL will proceed. The production of PCC prototypes and performance studies are expected to be concluded in September 2019. We are presently evaluating available DAQ boards to be used in the RUproto1 validation. Considerable firmware and software development will be required, which will however build on firmware and software developed for the TOFHIR1 and lpGBT ASIC test systems.

In parallel, the design and prototyping of the final FE ASICs, TOFHIR2 and ALDO2, will proceed in 2019-20. The design effort includes the migration of the TOFHIR1 blocks to TSMC 130 nm technology, the revision and adaptation of the analog circuitry for radiation tolerance and the development of blocks not included in TOFHIR1, in particular the pulse filtering modules, the internal reference circuits, the E-link I/O and logic dedicated to the L0 trigger.

We foresee two TOFHIR2 prototype MPW runs, the first one to be submitted in Q4 2019 and the second one in Q3 2020. Given the extensive simulation work being planned, we expect the first prototype to perform sufficiently well to allow thermal and radiation tests to be performed in parallel with the second iteration. In the same period, the design and prototyping of the ball grid array (BGA) packaging will be contracted to industry.

The second MPW iteration allows for corrections and possible design modifications needed to achieve the final required performance. The development of PCBs for ASIC characterization of the unpackaged and packaged chips compatible with radiation and thermal test-setups, as well as the development of firmware and software modules, will be done in parallel with MPW chips production. The final prototype of the RU (RUproto2), with TOFHIR2 and ALDO2, will be developed and validated between Q3 2019 and Q1 2021. This activity will include thermal and radiation tests, accelerated aging testing, and performance characterization in test beam.

Two design reviews are scheduled for the TOFHIR2 development. The first one was conducted on June 4, 2019, to review the readiness for the submission in November 2019. The second review is planned for Q1 2020 to review the readiness for the second submission in October 2020. These reviews are included as milestones in the project plan.

2.3.3.2 Model for the production and testing of the full system

The TOFHIR2 production is expected to start in Q2 2021 and to take one year. This will include wafer production engineering run (3 months), development and production of the BGA test system (in collaboration with the test company and in parallel with chips production), wafer thinning and dicing (2 months), chip packaging (3 months), and the test of packaged chips (4 months). The production of the ALDO2 chips will be done in the same period. The production

BTL ELECTRONICS TIMELINE	2018				2019				2020				2021				2022			
	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4
Prototyping																				
TOFHIR1	design		fab		test															
ALDO1			fab																	
RUproto1				design		fab	test	◆												
TOFHIR2 prototype v1					design			◆	test											
TOFHIR2 prototype v2									design		fab	◆	test							
ALDO2 prototype					design		fab	test												
RUproto2							design		fab		test		◆							
Production																				
TOFHIR2														◆	wafer	packg	test			
ALDO2											wafer		packg	test						
FE board																◆	assembling		test	◆
CC and PCC														pcb	assembling		test	◆		

Figure 2.59: Schedule of BTL electronics prototyping and production.

of the FE boards is expected to take 9 months, in the period Q4 2021–Q2 2022. The boards will be produced by industry in two batches such that the testing of the boards can start in Q2 2022. All boards will be individually tested before being assembled in RUs. In order to finish the board testing by the end of 2022, the test throughput of FE/CC/PCC boards must be 16/4/8 boards/day. We expect that four institutes of the collaboration will contribute to this effort.

The schedule of BTL electronics prototyping and production is shown in Fig. 2.59.

A first full system test using the TOFHIR1 prototype will be conducted in early 2020. This follows the test of the RUproto1 in the laboratory. The system test will be conducted with two RUproto1 mounted on a cooling tray and with sensor board prototypes using SiPM arrays from the R&D run in Summer 2019. Upon availability of TOFHIR2 prototype v1 the DCR cancelation performance will be tested with irradiated sensor modules in Spring 2020. A full system test with TOFHIR2 prototype v2 is scheduled for Spring 2021. Corresponding milestones are included in Section 6.2.2.

2.4 Engineering and integration

The BTL detector will be attached to the carbon fiber Tracker Support Tube (TST), which serves as the mechanical support for the Tracker. The BTL compartment will be a thin, cylindrical structure with an inner radial boundary of 1148 mm and a length of 5200 mm covering the pseudo-rapidity region up to $|\eta| \approx 1.5$ with a total surface of about 38 m^2. The BTL detector is subdivided into 72 identical segments, referred to as trays. The trays have a three layer design, composed of a cooling plate, the sensor layer, consisting of the LYSO:Ce bars with SiPMs along with electrical and mechanical connectivity and finally the front-end electronics layer which is further subdivided into the ASIC FE boards and a Concentrator Card (CC). The sensor layer has a modular design with 16 crystal bars and their respective 32 SiPMs forming a mechanical unit referred to a sensor module. In summary, the BTL is segmented as follows:

- 72 trays, 36 around ϕ, 2 along z,
- 6 readout units per tray,
- 24 FE boards and 6 Concentrator Cards per tray,
- 144 sensor modules per tray,
- 4608 readout channels per tray,
- 10 368 modules for the entire BTL,
- 331 776 readout channels in the entire BTL.

The BTL sensor boards consist of a SiPM array mounted on a LYSO:Ce crystal assembly and a flex cable connecting it to the front-end electronics. The flex cables have a few passive components attached to them such as resistors, capacitors and connectors. The sensor boards will be fixed to the cooling plate via aluminum rails, which also serve as thermal conductor as described in Section 2.2.3. The BTL sensor boards and the front-end readout boards will be attached to the BTL cooling tray in the tray integration step. Figure 2.60 shows a r–ϕ view of a BTL tray, attached with its support rails to the TST.

1 : TOFHIR board with 6 ASICs
2 : LYSO array with 16 LYSO bars, bars oriented in ϕ
3 : Concentrator card
4 : DCDC converter
5 : CC-to-FE connector
6 : SiPM-to-FE connector
7 : Cooling bar with CO_2 pipes
8 : Cooling fins
9 : TST
10 : Insulation
11 : BTL compartment cover plate

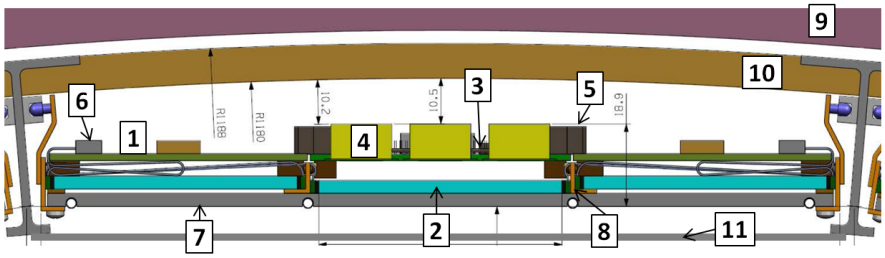


Figure 2.60: Cross-sectional (r–ϕ) view of a BTL tray with crystal bars oriented in the ϕ direction. Three sensor modules can be seen, covering the width of the tray. In this design the sensor boards are mounted on the cooling tray, the FE boards are mounted on top of the sensor boards. The FE boards are supported by aluminum rails mounted on the cooling tray. These rails serve as an additional thermal guide to ensure proper thermal contact of the SiPM package to the cooling tray. Flex cables provide connectivity from the SiPM package to the FE. A cover plate separates the BTL volume from the Tracker volume.

The sensor modules will be preassembled by gluing the SiPM packages on the ends of the packaged LYSO:Ce bar matrices. Connectivity between the SiPMs and the FE is provided by flex cables as shown in Fig. 2.65. The mounting sequence is shown in Fig. 2.65 and the tray design is discussed in more detail in Section 2.4.3.

2.4.1 Structural design of the TST

In Fig. 2.1 we showed the overview of the sharing of the volume inside the TST between the Tracker and the BTL. The TST design and dimensions are largely identical to the one used for the current Tracker. The outer diameter of the Tracker is reduced by about 2.5 cm with respect to the initial Phase-2 Tracker design, which had no BTL integrated into the TST. The BTL will share a common cold volume with the Tracker inside the TST. The inner wall of the TST will be covered with a heating foil and a 6 to 8 mm thick layer of insulation. This will allow to maintain a temperature of $-30\,^{\circ}\mathrm{C}$ on the BTL and keep the TST inner wall at $0\,^{\circ}\mathrm{C}$ to reduce thermal stress on the TST structure. Thermal and mechanical deformation analysis shown in Fig. 2.62 demonstrated that the thermal stress on the TST is negligible with this temperature gradient across the carbon fiber structure. The residual heat pickup from the environment through the TST and the insulation will be about 6.5 kW, based on detailed finite-element analysis (FEA) of various combinations of TST, insulation and heating foils as shown in Fig. 2.66.

The active cooling of the BTL will ensure that the surface facing the Tracker will be thermally stable at the temperature required by the Tracker. The BTL trays will be attached to the TST with rails that allow to slide in fully equipped trays from the ends of the TST. The BTL compartment is separated with carbon fiber sheets which serve as a mounting surface for the cables of the Tracker. There is no thermal barrier between the Tracker and the BTL volume. Once installed in the TST, BTL and Tracker are handled as one common detector. The TST, with BTL and Tracker

inside, is planned to be transported to P5, lowered into the cavern and inserted into CMS. The TST is supported with four brackets which are mounted to the HCAL, as in the current Tracker.

2.4.2 Tracker-BTL TST mechanics

The current design of the TST is illustrated in Fig. 2.61. The walls of the TST are made of a sandwich panel consisting of a NOMEX honeycomb core layer and two carbon fiber facesheets. The total thickness of the structure is 30 mm, with a carbon fiber sheet skin of 2 mm thickness. The carbon fiber sheet layer thickness is doubled to 4 mm over a length of 1000 mm at the ends of the TST to increase the stiffness of the structure. This design is identical to the one used for the TST of the current CMS Tracker, which has performed as expected during installation, integration and operation of the detector. The deformation of the TST with the current design, with the current assumption of Phase-2 Tracker weight and a 1 metric ton weight representing the BTL as a flat layer on the inside of the TST, has been simulated and is presented in Fig. 2.62. Both gravitational and thermal effects are considered in the simulation. There is an overall displacement of the TST due to the deformation of the support brackets, and a deformation of the shape. The effect of the deformation of the support brackets under load can be compensated during the mounting procedure, as it was done for the current Tracker where similar effects were observed. The deformation of the shape is well within the tolerance of ± 3 mm required by the Tracker. The structural design principle of the TST using a sandwich structure is the choice for the mechanical support structure of the Phase-2 Tracker and BTL. More detailed simulations, including a segmented BTL suspended from support rails as in the final design, are ongoing. Further reinforcement of the structure can be achieved by increasing the carbon fiber wall thickness. This will allow to adjust the stiffness of the TST to cope with the weight of the BTL even if it exceeds 1 metric ton.

2.4.3 BTL tray design

The BTL tray is designed to allow fast and efficient assembly of the basic building elements, i.e. the sensor layer, cooling bars, and the front-end electronics layer, as shown in Fig. 2.63. The cooling bars serve as a mechanical support structure for the sensor and the electronics layer, as a thermal conductor between the cooling pipes and the sensor as well as the front-end boards. The cooling bars also serve as mounting points which hold the BTL trays in the TST support structure. The cooling bars are described in more detail in Section 2.4.4. The BTL tray has an approximate dimension of $250 \times 18 \times 2.5$ cm^3 and a weight of less than 20 kg. The weight of the crystals, as described in Table 2.3, amounts to about 9 kg per tray, which dominates the total weight of the tray. A breakdown of the weight of one tray is shown in Table 2.9.

Table 2.9: Weight estimate for a BTL tray, broken down into the major components. The weight is dominated by the weight of the LYSO:Ce crystals.

Tray component	Quantity per tray	Specific density	Approx. weight [kg]
LYSO:Ce crystals	1262 cm^3	7.1 g/cm^3	8.96
Cooling plate	1388 cm^3	2.7 g/cm^3	3.8
Mother boards	1480 cm^3	1.85 g/cm^3	2.7
LV cables	50 m	11 g/m	0.55
HV cables	100 m	7 g/m	0.7
Misc. components			2.0
Total			18.71

All BTL trays are identical, all cooling trays, the sensor modules and FE cards are identical and thus interchangeable. Only the flex cable connections between the sensor modules and the FE

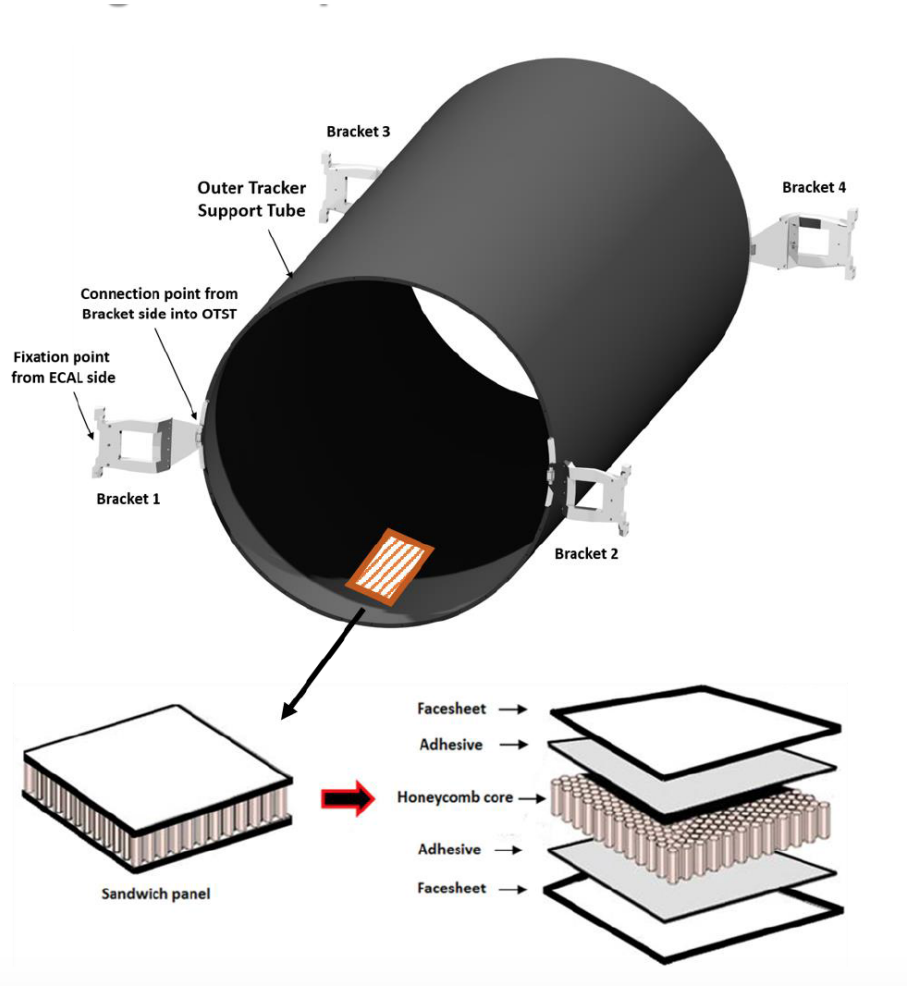


Figure 2.61: Schematic view of the TST design. The TST consists of a sandwich structure with two carbon fiber facesheets and a honeycomb core. Four support brackets at the end of the TST connect the entire structure to the HCAL.

have three variants with slightly different layout to accommodate the different routing of the cables for the sensor modules at the edge and in the center of the tray, as shown in Fig. 2.64. This layout of the sensor module connectivity ensures that the spacing of the connectors at the edge of the tray and subsequently the FE boards can follow a uniform spacing of the components, in particular the ASICs and ALDOs. An alternative layering of the tray, with the cooling tray in the middle and the sensor and electronics layer on opposite sides of it, is also being considered. The combination of the sensor modules with their respective flex cables and auxiliary mechanical support structures for the sensor module is referred to as sensor board. The design of the BTL trays is shown in Figs. 2.63 and 2.60. The components of sensor modules are described in Section 2.2, the front-end electronics is described in more detail in Section 2.3.

A BTL tray section with the layout of the sensor modules is shown in Fig. 2.64. Shown are six sensor modules with 16 LYSO:Ce bars each with the two SiPM arrays on both ends of the crystal bars. The flex cables are connecting the SiPM packages with the FE boards. The cables are routed to the edge of the tray where the connectors are located on the FE ASIC boards. Necessary discrete components such as resistors and capacitors are mounted on the flex cable. To route the signals efficiently to the edges of the tray three slightly different layouts of the flex cable connections are needed. To reach the FE boards the flex cables can either be wrapped

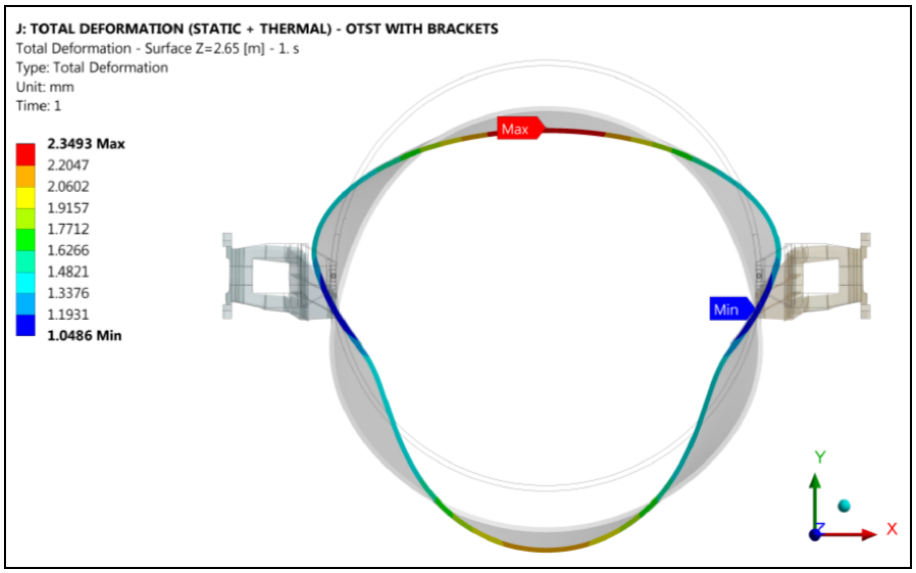


Figure 2.62: Simulation of TST deformation caused by the weight of the current Phase-2 Tracker and 1 metric ton of BTL, implemented as a flat layer on the inside of the TST. Gravitational and thermal effects are considered. There is an overall displacement of the TST due to the deformation of the support brackets, which can be compensated, and a deformation of the shape.

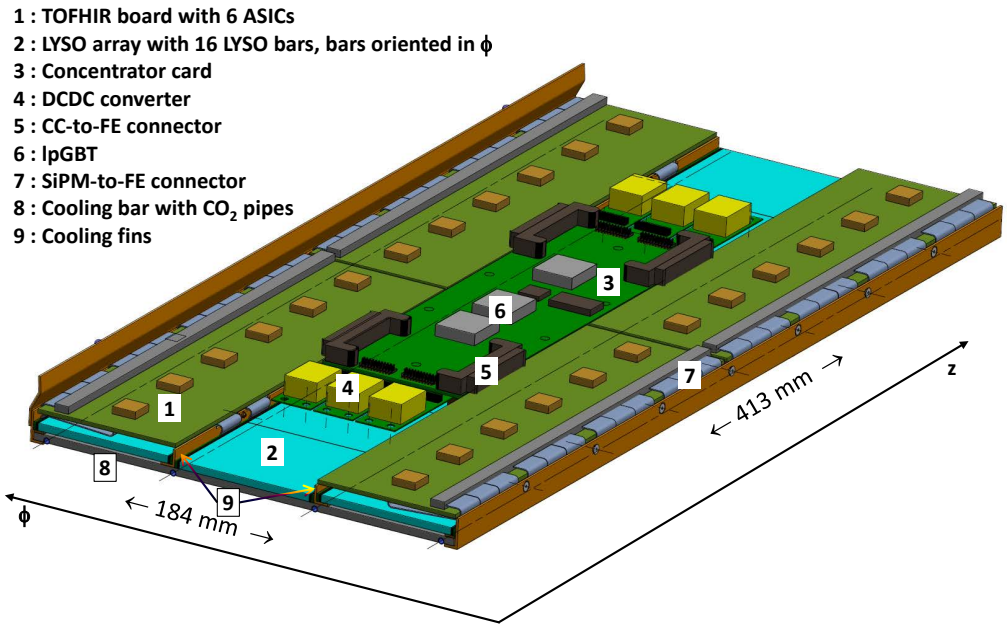


Figure 2.63: View on the top of a section of the BTL tray with the FE electronics boards visible. A full readout unit is shown, consisting of 24 sensor modules, four ASIC boards and one concentrator card. In the design shown flex cables, wrapping around the edges of the FE ASIC boards, provide the connectivity between the sensor modules and the FE cards.

around the edge of the tray or plug into a connector on the bottom of the FE board.

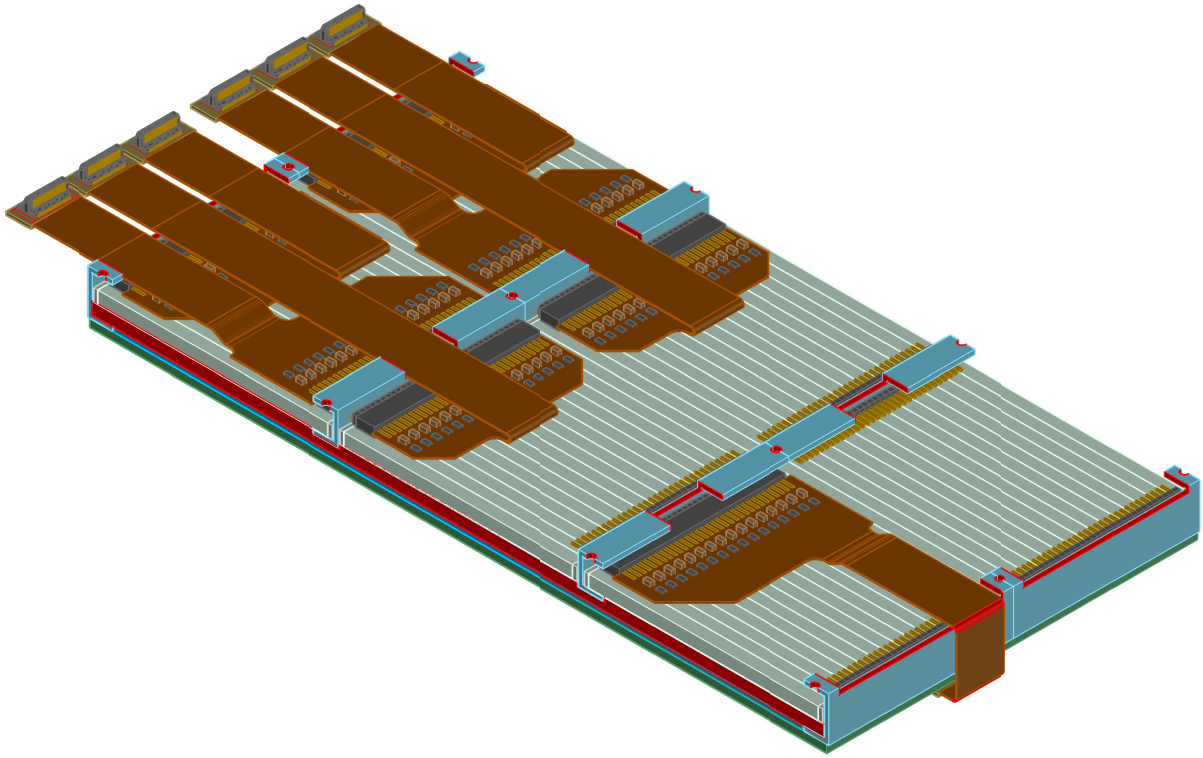


Figure 2.64: BTL tray section showing the layout of the sensor modules and a possible flex cable connection configuration.

The assembly of the sensor modules onto the cooling bar will proceed in three steps. Aluminum profiles, serving as additional mechanical support for the sensor modules, will be mounted on the bare cooling bar. Then the sensor modules will be placed between these cooling bars and finally the connecting flex cables will be mounted. The sequence is illustrated in Fig. 2.65.

2.4.4 Cooling and environmental control

The cooling of the BTL detector is based on evaporative CO_2 in a liquid-pumped cycle equivalent to what has been adopted for the CMS Phase-2 Tracker Upgrade. It uses the two-phase accumulator-controlled loop concept (2PACL) that has been successfully employed in the LHCb VELO since 2008 and the CMS pixel detector since 2017. The fluid properties of CO_2 make it an ideal medium for cooling of detectors. Its key advantage is the ability to control the temperature at the detector remotely by adjustment of the pressure in the accumulator at the cooling plant. It also enables the use of smaller diameter tubing than is required with conventional refrigerants or liquid cooling applications. The boiling temperature is a function of the pressure and, since the dynamic pressure drop along a cooling pipe is small compared to the absolute pressure of the fluid, the change in the evaporation temperature along a cooling pipe is typically small: two degrees between outlet and inlet coolant temperature is the target for the BTL design, matching the specifications of the Tracker. The high latent heat of vaporization of CO_2 translates into a smaller flow as compared to that required for other refrigerants. Because

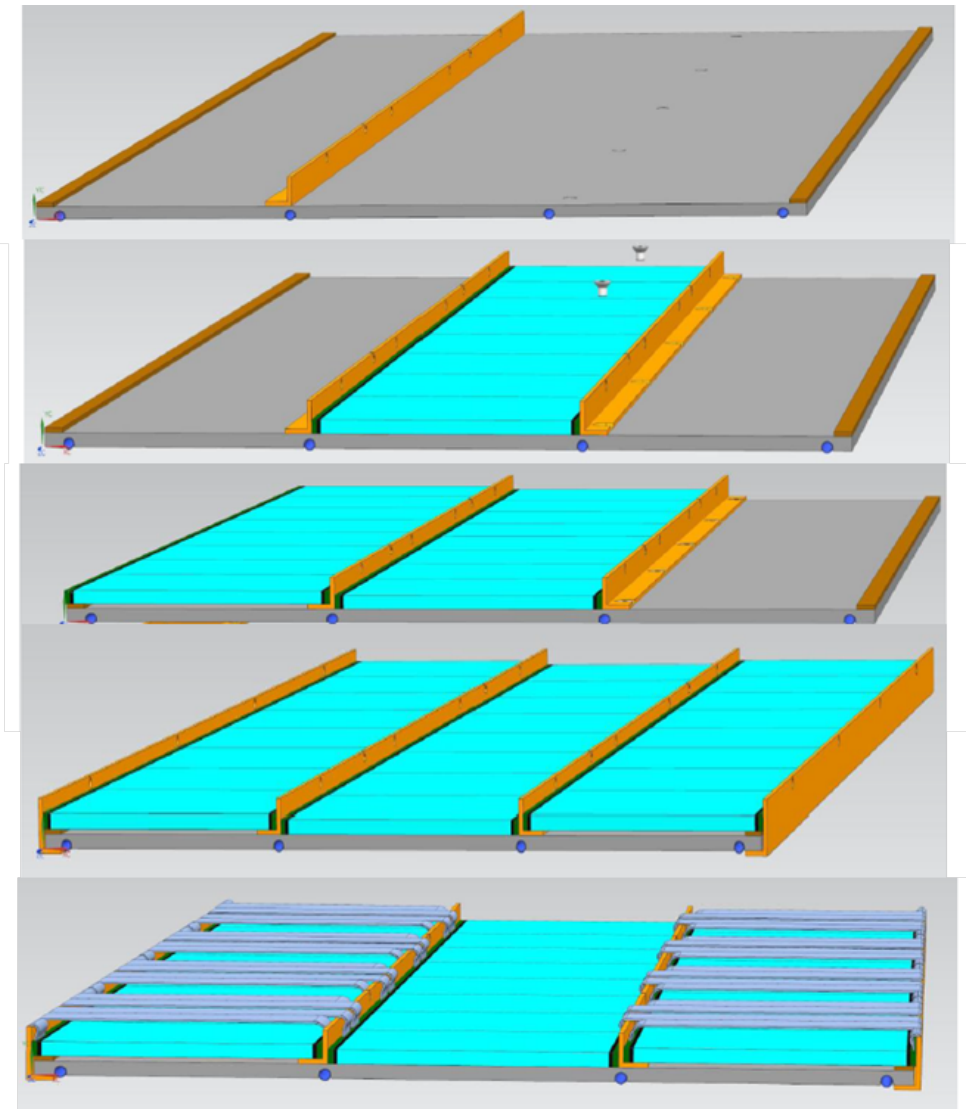


Figure 2.65: Sequence of the mounting of sensor modules on the cooling tray. From top to bottom: mounting of aluminum profiles on the cooling tray, insertion of the center, left and right sensor module and finally connection of the flex cables.

the viscosity of CO_2 is low, it allows the use of small diameter pipes with higher flow speeds, which increase the heat transfer coefficient from the pipe wall to the fluid. CO_2 is radiation hard, inexpensive, and environmentally friendly; in particular the global warming potential is orders of magnitude lower when compared to chlorofluorocarbons, such as C_6F_{14}, used in the present CMS Tracker. Rigid vacuum-jacketed stainless steel pipes carry the CO_2 from the refrigeration plants in the service cavern, USC, into the collision hall, UXC. The flow is fed to the detector cold volume via 12 insulated transfer lines on each end of the BTL. From there, the CO_2 flow is branched through a 1-to-6 manifold, as shown in Fig. 2.70, to the individual trays, running along the periphery of the detector at the edge of the TST inside the BTL service channel. Their routing in the BTL service channel is shown in Fig. 2.71. Inside the tray, there will be two identical cooling loops running along the tray in a U-shape with a length of about 5 m per loop. Capillaries are installed between the manifolds and the cooling lines in the trays to ensure stable performance of the heat extraction along the cooling loop.

The BTL cooling system is designed to cope with a total nominal power dissipated in the BTL

compartment, dominated by the electronics and sensor leakage current, of about 38.5 kW. This value includes a margin of about 50% with respect to the power needs estimated in Section 2.5. There is an additional heat load to the BTL cooling circuits from thermal losses to the environment of 6.5 kW. Including the preheaters, a feature of the technology choice of evaporative CO_2 cooling, the system is designed to provide a total cooling power of 46.2 kW. This corresponds to about 640 W of cooling power per tray. The breakdown of the heat load for different operational states is given in Table 2.10. The cooling system must remove this heat load and maintain the SiPMs at the lowest possible temperature. A temperature of $-30\,^{\circ}$C for the SiPM has been assumed in this chapter of the TDR. A simulations of the BTL cooling circuit is shown in Fig. 2.68. The simulations assume a CO_2 vapor quality of 35%. The temperature of the cooling pipes is well below $-30\,^{\circ}$C for most of the length of the pipes. Further reduction of the temperature gradient along the cooling pipe can be achieved by slightly increasing he pipe diameter as shown in Fig.2.68. Due to the efficient heat conduction of the BTL cooling tray, the temperature gradients between the coolant and the SiPMs is expected to be very low. In Fig. 2.68 a temperature of $-35\,^{\circ}$C has been assumed at the entrance to the BTL cooling loop. This is the most pessimistic assumption. The actually achievable temperature is expected to be lower which will provide operational margins. This is explained in more detail in Section B.I.3.4.

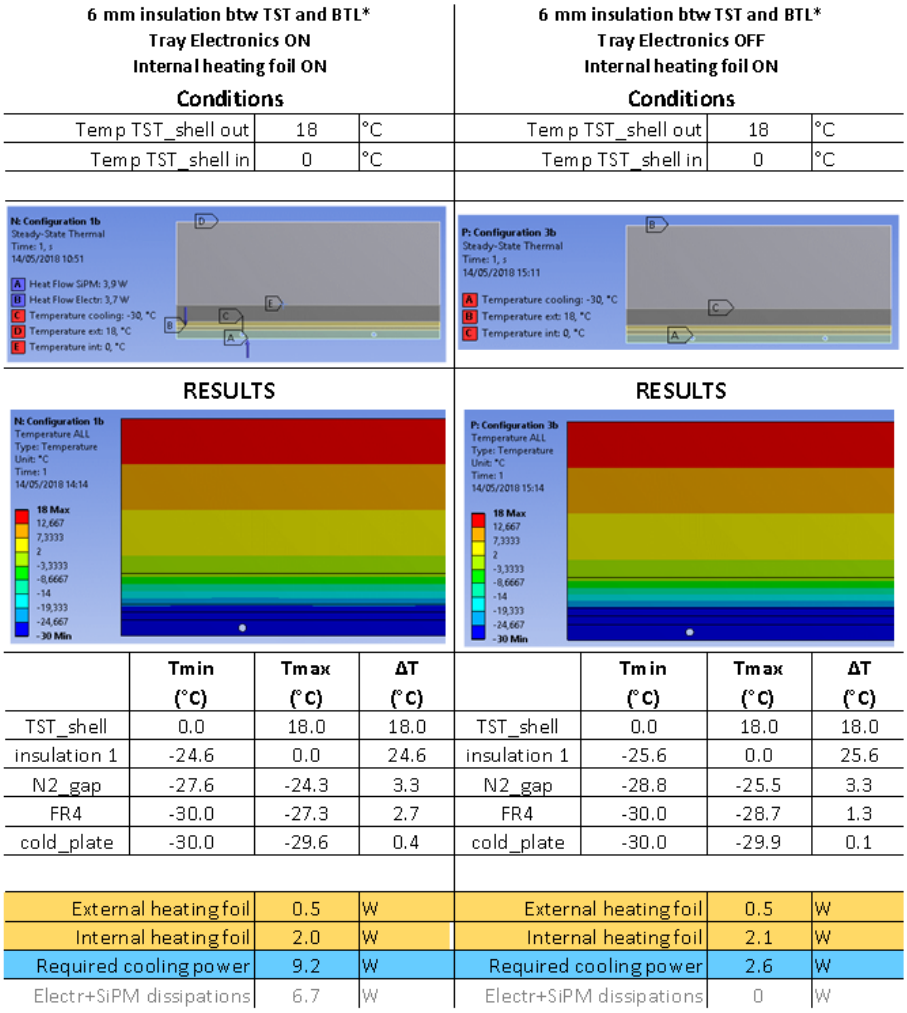
One cooling plant with a maximum capacity of about 44 kW at $-35\,^{\circ}$C will be employed initially which is expected to be fully sufficient to cool BTL. The pump can be upgraded to higher capacity with a second pump head if the full contingency of 46.2 kW should be needed for BTL towards the end of life. Operation of the detector at temperatures up to $15\,^{\circ}$C is possible with a reduced capacity of the CO_2 system at about half the nominal cooling power. This is sufficient to operate the FE during shutdowns to monitor the detector with the SiPMs operating at reduced bias voltage to limit the leakage current. The possibility of operating the SiPMs at slightly higher temperature to accelerate the annealing is being investigated.

The BTL will share the cold volume with the Tracker. The TST and the 6 to 8 mm thick insulation will provide sufficient insulation to maintain a temperature of $-30\,^{\circ}$C inside the shared volume and keep the TST structure above $0\,^{\circ}$C. The residual heat pickup from the environment through the TST and the insulation will be about 6.5 kW, based on detailed FEA of various combinations of TST, insulation and heating foils, as shown in Fig. 2.66.

The environmental heat pickup from the TST surface has to be removed by the BTL cooling circuits as well. Heating foils are placed on the TST surface to maintain it at or near ambient temperature. The detector volume will be continuously flushed with dry nitrogen during operations and dry air for safety reasons when CMS is open. The SiPMs will need to be warmed to room temperature during shutdowns to enable the annealing of the radiation-induced leakage current increase. This will be achieved by running the CO_2 cooling at a temperature of $15\,^{\circ}$C and adjusting the SiPM bias current such that the detector is heated up to this temperature. Some insulation between the BTL and Tracker is needed to minimize the impact of the temperature of the Tracker volume. As there is a beneficial effect of warming up the detector, the operational availability of the cooling pump for the BTL detector has less stringent requirements than for the other subdetectors. The BTL cooling plant instead may serve as a backup plant for the Tracker in an emergency or during pump maintenance as the BTL cooling system can maintain the Tracker volume at $-30\,^{\circ}$C provided the detectors are off.

2.4.4.1 Cooling plates

The BTL cooling plates consist of aluminum with embedded stainless steel pipes for the CO_2 cooling circuits. Figure 2.67 is an illustration of the current design of the cooling plate with



	Tmin (°C)	Tmax (°C)	ΔT (°C)
TST_shell	0.0	18.0	18.0
insulation 1	-24.6	0.0	24.6
N2_gap	-27.6	-24.3	3.3
FR4	-30.0	-27.3	2.7
cold_plate	-30.0	-29.6	0.4

External heating foil	0.5	W
Internal heating foil	2.0	W
Required cooling power	9.2	W
Electr+SiPM dissipations	6.7	W

	Tmin (°C)	Tmax (°C)	ΔT (°C)
TST_shell	0.0	18.0	18.0
insulation 1	-25.6	0.0	25.6
N2_gap	-28.8	-25.5	3.3
FR4	-30.0	-28.7	1.3
cold_plate	-30.0	-29.9	0.1

External heating foil	0.5	W
Internal heating foil	2.1	W
Required cooling power	2.6	W
Electr+SiPM dissipations	0	W

Figure 2.66: Simulation of the thermal losses through the TST. The simulation shown is for an area of about 3% of the tray area. The power dissipation is 2.6 W, corresponding to about 90 W per tray, 6.5 kW for the entire BTL. The simulation on the left are with BTL on, on the right with BTL off. The heating foil between the TST inner surface and the insulation is on in both cases.

Table 2.10: Cooling power budget breakdown as used in the specifications of the manifolds, transfer lines and cooling plant. All numbers are given in kW. The two columns labeled with BOO and EOO correspond respectively to the beginning and end of the BTL operation. The column labeled 15 °C refers to the warm operation state of the cooling system. The numbers for the power consumption of the front-end and SiPM leakage current contain a contingency of about 50 %.

Power consumption [kW]	BOO	EOO	15 °C
SiPM leakage current	0	25	0
front-end	13.5	13.5	13.5
environmental loss	6.5	6.5	0
preheaters	1.2	1.2	1.2
Total	21.2	46.2	14.7

two cooling loops (left) and one cooling loop (right). The CO_2 circuit in the tray uses pipes with an inner diameter of 2.5 mm for the one-loop and 2.1 mm for the two-loop configuration.

Capillaries are installed at the inlet of the cooling loop to ensure reliable operation. The CO_2 cooling circuit, manifolds and fittings are based on the same technology used for the Tracker cooling. The pipe diameter is slightly larger than the typical pipes in the Tracker which eases joining pipes by welding. A BTL cooling circuit with one loop, supply lines and manifolds has been simulated as shown in Fig. 2.68. The pressure drop and thermal uniformity are within specifications. The thermal uniformity can be improved by increasing the pipe diameter further or by using two cooling loops per tray, which is the current design choice, as it matches better with the sensor layout. The FEA thermal simulations of trays with one and two cooling loops based on simplified models are presented in Fig. 2.69. In these simulations the maximum difference in the temperature of BTL tray and the cold plate was found to be about 2.2 °C, in the scenario where there are two loops, and about 5 °C, in the case of one cooling loop. Thermally coupling the LYSO:Ce bars to the cooling plate does not change the temperature of the SiPMs significantly. The thermal coupling of the SiPMs is dominated by the SiPM package and its coupling to the cooling plate as demonstrated in Fig. 2.69.

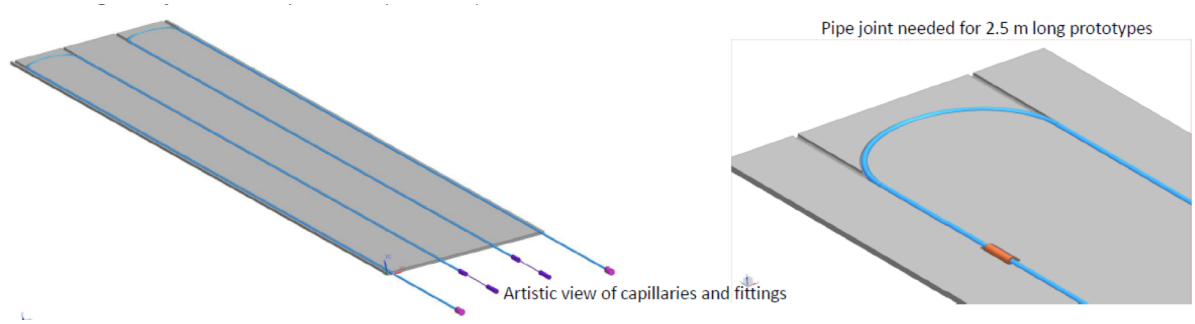


Figure 2.67: Design of the cooling plate with two loops in the left plot and one loop in the right plot. The thickness of the cooling plate will be approximately 3 mm with an inner pipe diameter of 2.1 mm for the two-loop configuration. The capillaries are not shown in this view.

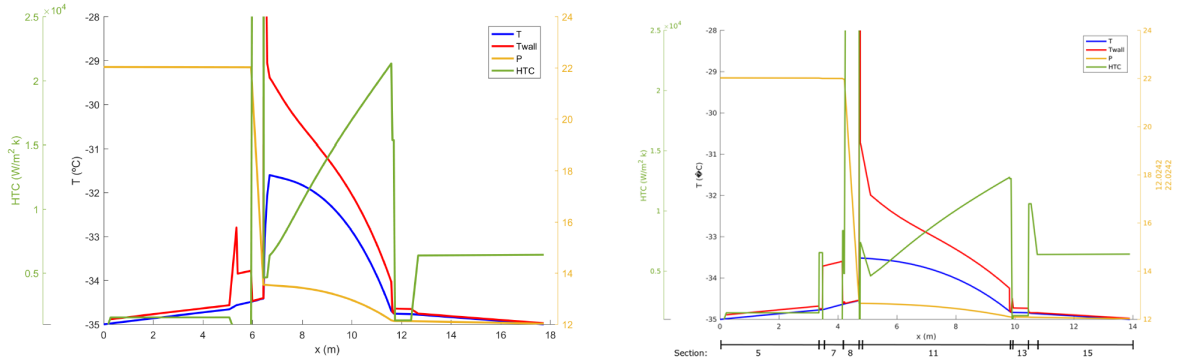


Figure 2.68: Simulation of the CO_2 cooling circuit with a single cooling loop inside the BTL tray. Shown are the temperature (T), pressure (P) and the heat transfer coefficient (HTC) in the BTL cooling circuit. The section from x = 6.5 to 11.5 m corresponds to the cooling loop in the tray. The thermal uniformity in the final design will be improved by using two cooling loops. The diagram on the left shows the temperatures assuming a cooling pipe with an inner diameter of 2.1 mm. The diagram on the right assumes a cooling pipe with an inner diameter of 2.5 mm. The slightly larger pipe diameter results in a reduction of the maximum coolant temperature by about 2 °C.

The BTL CO_2 circuit for the ϕ-bar geometry will consist of two cooling loops per tray to optimize the thermal uniformity. A one-to-two manifold integrated into the tray will connect to a manifold in the BTL service channel that distributes one incoming transfer line to six trays as

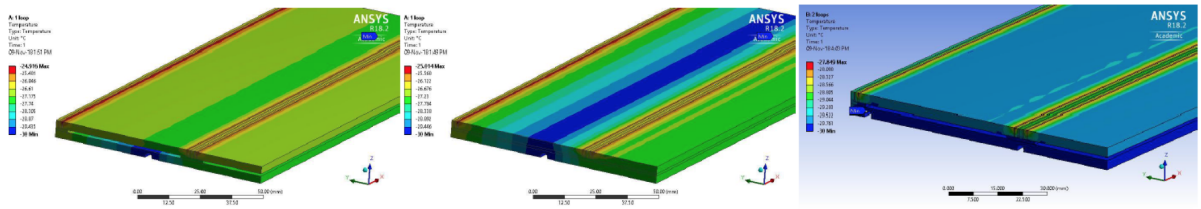


Figure 2.69: The FEA thermal simulation results for the BTL tray are presented for different scenarios: one loop and no thermal pad between crystals and cold plate (left); one loop and a thermal pad between crystals and cold plate (middle); two loops and no thermal pad between crystals and cold plate (right).

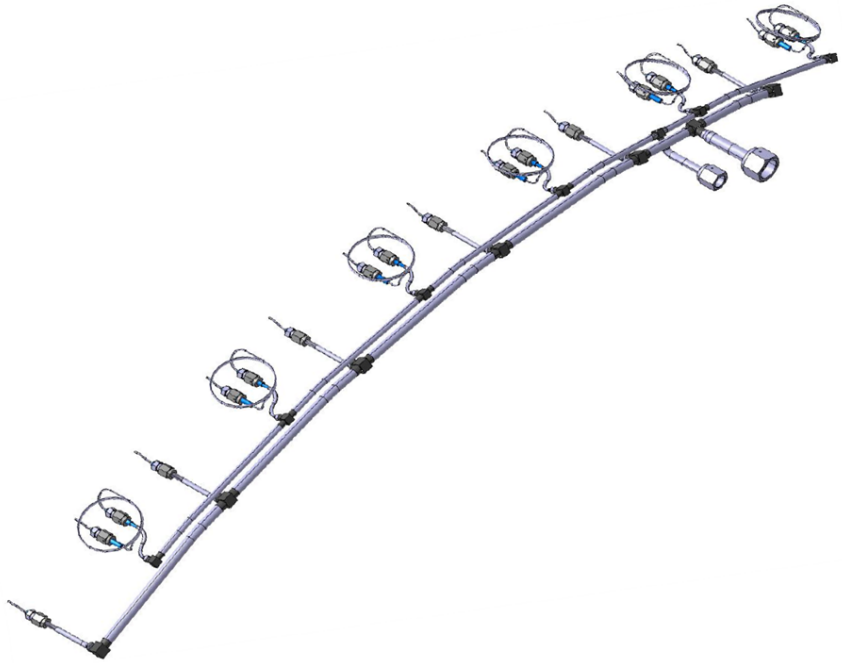


Figure 2.70: Main on-detector manifold of the CO_2 cooling distribution system in the service channel of the BTL. One manifold will connect one incoming transfer lines to 6 BTL trays. There are 12 such manifolds for the BTL. The manifolds are largely identical with minor variations of the location of the transfer lines connection along the arc of the main manifold pipe.

shown in Fig. 2.70.

Eight prototypes of cooling plates with either one or two cooling loops per tray are being assembled and are expected to be ready by summer 2019. These prototypes will be used to demonstrate 650 W power extraction with a one- and a two-loop tray design and measure the actual SiPM temperature with the realistic tray design. Further tests will measure the performance at different vapor quality values, the need for preheaters on the CO_2 pipes was well as different mounting strategies for the capillaries on the tray.

2.4.5 BTL Services

The CO_2, power and data link services of the BTL closely follow the design of the Outer Tracker. The routing of the services will follow the outer Tracker cable ducts which can be seen on the top edge in Fig. 2.75. The current CAD model assuming three 9 mm cables to bring in the bias voltage for the SiPMs and low voltage for the powering of the FE, as shown in Fig. 2.71. The total cable cross section needed for power and SiPM bias voltage is still being optimized, and

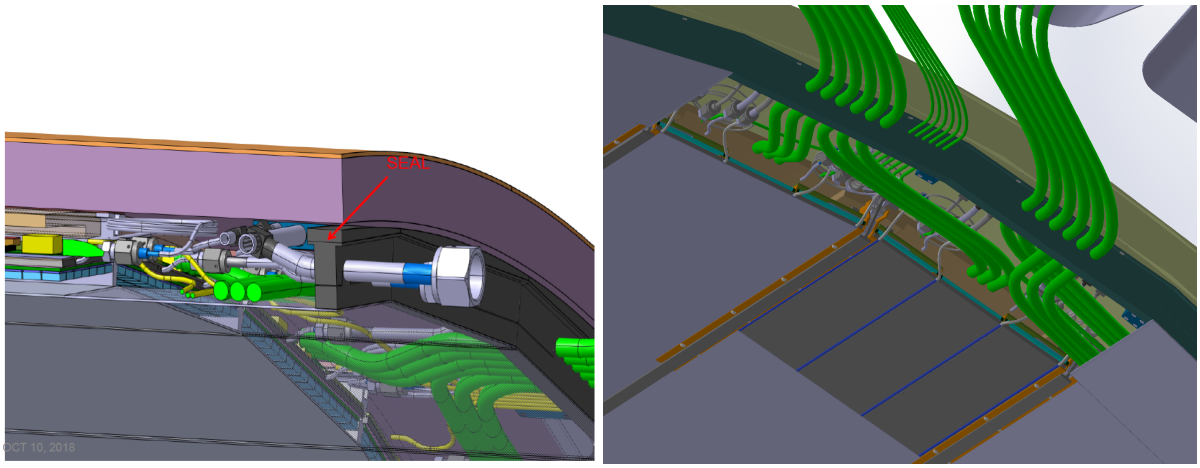


Figure 2.71: CAD model of the BTL services at the end of the tray in a cross section view on the left and looking from inside the TST onto the inner surface where the BTL is mounted. The space housing the services at the end of the trays is referred to as the BTL service channel. The service channel extends 17 cm in z in the current design. A thermal seal is closing the volume to the outside with the services passing through.

there is a possibility to increase the number of cables to four from three as shown in CAD mode. Maintaining the cable cross section to the current value and optimizing the cable multiplicity to ensure the highest possible granularity in the powering are criteria for this optimization. A mockup for one BTL tray and service channel has been built and is used to optimize the cable and cooling pipe layout, as shown in Fig. 2.72. In particular it is anticipated to use connectors in the service tray instead of routing the cable bundles across the BTL service channel thermal seal. Further studies are ongoing to integrate the capillaries of the CO_2 circuit into the tray to reduce the complexity of the integration work in the service channel.

2.4.6 BTL detector assembly

The construction of BTL can be sub-divided into three phases, as illustrated in Fig. 2.73.

The initial phase has the prototyping, the pre-production and the production for the principle components of the BTL broken into separate activities: the sensors; the front-end electronics; and the mechanical structure. Each of the three areas is further sub-divided according to needs. The prototyping on each activity is multi-branched to allow us to proceed with several aspects in parallel. This allows us to prototype the electrical connection between the SiPM and the FE independently from the thermal connection between the SiPM and the cooling plate. For the final production and QA/QC on each of the parts of BTL the procedures will be merged to one single, agreed upon thread.

The second phase of the BTL construction is the mass production of BTL sensor boards, RUs and trays. The individual sensor boards, front-end boards and cooling trays will be manufactured in industry. The QA/QC of the front-end boards will happen upon reception at the collaborating institutes responsible for the production. The sensor boards are assembled from LYSO:Ce bars and SiPMs which have previously undergone a QA/QC procedure as described above. The integration of sensor boards and FE boards onto the trays will happen in three integration centers in parallel. In the current schedule, each integration center assumes one assembly team working. The tray production can be accelerated by having multiple integration teams working in parallel. Each center will work according to a standardized procedure defined in the R&D phase. The current schedule is based on the conservative assumption that

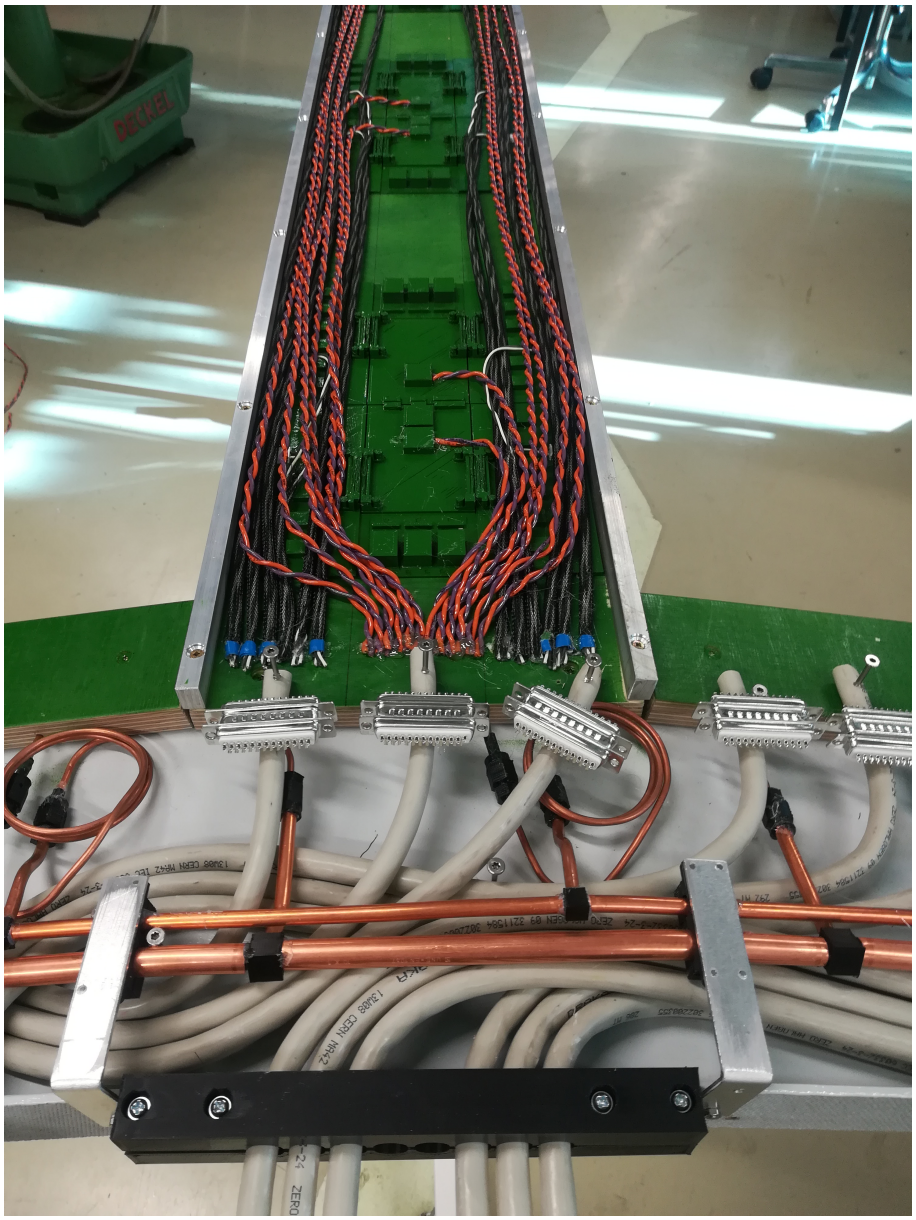


Figure 2.72: BTL service channel mockup which is used to validate and optimize the routing of the services. Shown in the picture is the routing of three cables per tray, one cooling loop connection with capillary in the service channel and power and bias voltage cables along the tray. The installation sequence and additional mechanical support to guide the services in the service channel are under development.

the integration will proceed with sensor modules and readout boards comprising one readout unit per day. Detailed testing of each RU will be conducted after it has been integrated onto the tray. Finally, fully integrated trays will be tested in a cold box in the integration center.

The last phase of the BTL construction is the integration into the TST. This will be done at the Tracker Integration Facility (TIF) at CERN. Fully assembled trays shall be shipped to CERN, tested upon reception in a dedicated test stand with CO_2 cooling and then installed into the TST. The insertion of the trays into the TST will be performed with a customized tool.

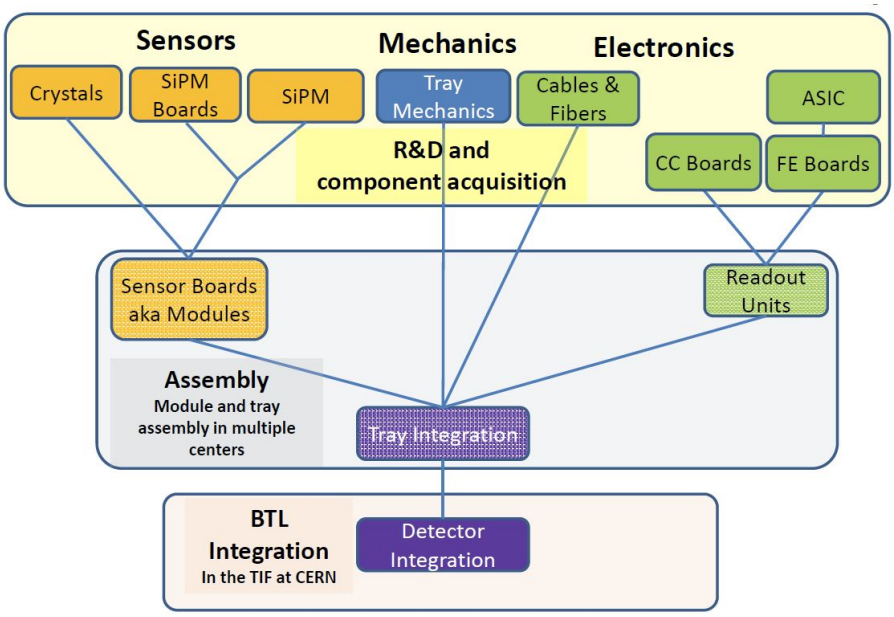


Figure 2.73: Illustration of the BTL construction model. The construction of the BTL relies on a modular design with a very limited number of largely identical components to minimize the complexity of the assembly. The final integration of the trays is parallelized in three integration centers to speed up the production.

2.4.6.1 BTL integration and commissioning

As the TST will house the BTL trays (36 in ϕ per side), as shown in Fig. 2.74, the BTL integration will be tightly connected to the integration of the Tracker. In the TST structure, there are two horizontal rails that separate the upper and lower half-cylinders. They serve as the mounting points for the Tracker and will conceptually remain as in the current TST. All of the mechanical structures of the Tracker are supported from these rails. This allows the integration of the BTL, which is supported from rails on the inside wall of the TST, and the Tracker to proceed in parallel to a certain extend. Specifically the barrel part of the Tracker can be installed before all the BTL trays are inserted. In the current integration schedule this parallel integration is the preferred scenario since it optimizes the schedule for both Tracker and BTL. The BTL trays will be covered with a thin inner wall, likely segmented with one for each tray, for mechanical protection and to serve as a mounting point for the Tracker services. In the current design configuration the main TST wall would experience a thermal gradient from 0 °C on the inner wall of the TST to 18 °C on the outside while the thin wall between BTL and Tracker will see little or no thermal gradient. When the BTL and Tracker are warmed to room temperature the thermal gradient across the TST wall will vanish. The thermal stress on the carbon fiber structure during these thermal cycles is small. A significant amount of mechanical engineering will be done by BTL institutions to customize the existing TST design for the current CMS Tracker to the needs arising from the inclusion of the BTL.

The integration of the BTL trays into the TST will happen at the TIF at CERN. This is a dedicated integration facility at CERN which was used successfully for the integration of the current Tracker. A test stand in the TIF dedicated to the BTL will be set up, where a final checkout of the trays delivered to the TIF will be performed. A CO_2 cooling infrastructure is available at the TIF to allow cold checkout of the trays before insertion into the TST. After this final checkout the trays will be inserted into the TST with a dedicated insertion tool. This tool will allow the alignment of the trays outside the TST with the BTL support rails and to then slide the trays

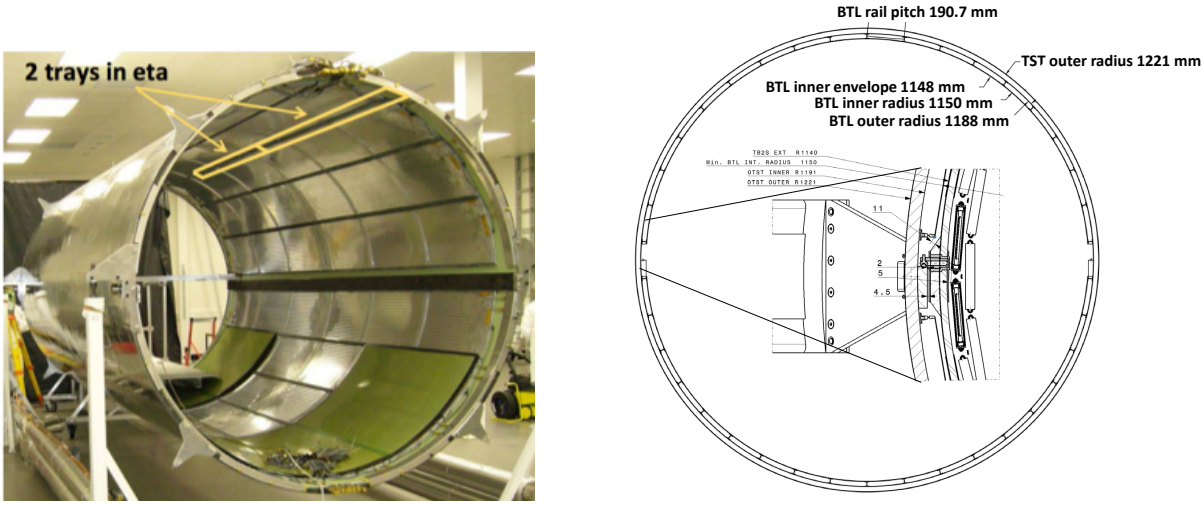


Figure 2.74: Picture of the TST for the current CMS Tracker (left). The structural design and the dimensions of the new TST to house the BTL and the new Tracker will be largely identical. Sketch of the cross section of the TST with mechanical structures to support the mounting of the barrel timing layer and the Tracker (right). The upper and lower half-cylinders, separated by two horizontal rails to support the Tracker, are divided in 18 compartments each to house the BTL trays. The insert in the center of the sketch shows the region around these rails in more detail. The rail concept to mount the Tracker is identical to the current TST.

into the TST. Then the cooling pipes will be connected, cables and fibers routed, and the tray will be tested once more after all connections are properly dressed in the BTL service channel. Finally the cover plates shall be installed to mechanically seal off the BTL volume and allow the Tracker services to be mounted. From this point onward, the BTL integration will follow the path of the Tracker. Upon completion of the Tracker integration and a joint commissioning in the cold inside the TST the entire object will be transported to P5, lowered into the cavern and inserted into CMS. This procedure will follow the established procedure employed for the installation of the current Tracker inside CMS. Once the structure is mechanically fixed inside CMS, the services will be connected. Figure 2.75 illustrates the routing of the services for the BTL at the end of trays facing the edge of the TST, in the BTL service channel. From the service channel the services pass through a thermal seal and then follow the Tracker services through dedicated cabled ducts to the outside of the detector, as described in Section 2.4.5 and shown in Fig. 2.75.

2.5 Power requirements

The barrel timing layer requires power for two major detector components:

- about 9.3 kW to supply its on-detector electronics;
- about 17 kW to provide bias to the photo-sensors at the end-of-operation of the detector.

Two independent systems, the low voltage system and the bias voltage system, both described below, deliver this power to the detector.

2.5.1 Readout Unit and power scheme

The BTL on-detector readout unit comprises three major components:

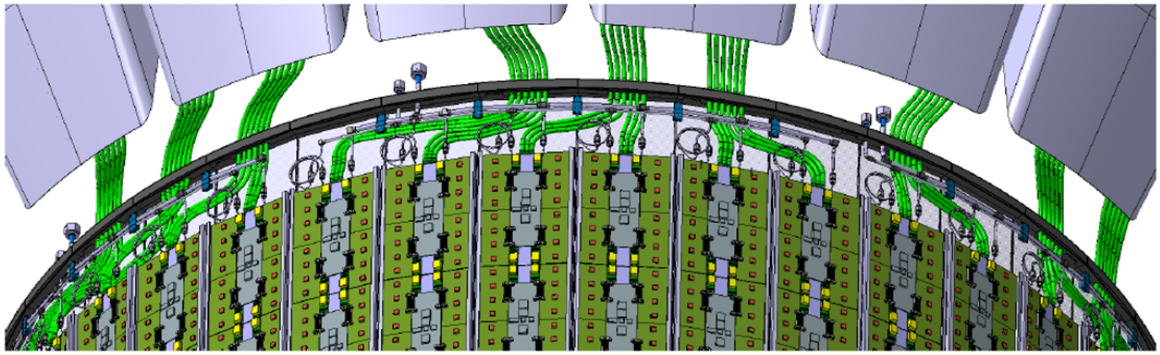


Figure 2.75: Illustration of the service routings for the BTL and the edge of the TST into the Tracker service channels. The Outer Tracker service channels, through which cooling pipes, power cables and data links are routed to the outside of the CMS detector, are seen on the top of the drawing. There are 24 such service channels. The services are distributed along the edge of the TST through the BTL service channel to reach the trays. The Outer Tracker service channels are limited to an area around the 12 and 6 o'clock positions while the channels at the 3 and 9 o'clock positions are reserved for the Inner Tracker. The services of the BTL have to be routed along ϕ in the service channel on the TST to accommodate this.

1. Front-end ASIC board (TOFHIR ASIC board or FE Board);
2. Digital interface card (Concentrator Card, CC) interconnecting the front-end board and the off-detector system via optical links;
3. DC-DC Power Converter Card (PCC).

The basic building block is the Readout Unit (RU) comprising 768 channels (photo-sensors). A tray hosts 6 RUs. The power conversion system will employ:

- FEAST ASIC based DC-DC converters;
- Adjustable low dropout regulator (ALDO).

The FEASTMP_CLP module developed by CERN is the present baseline. The FEAST DC-DC converters provide power directly to the Concentrator Card. The TOFHIR ASICs are powered from ALDOs, which in turn are powered from FEAST converters.

The basic inputs needed to develop the system are shown in the first block of Table 2.11. The detailed power estimate for the TOFHIR ASIC, which is still under development, is given in the second block of the table. The TOFHIR has 32 channels and the supply voltage is 1.2 V. This voltage is obtained by means of a linear voltage regulator, the ALDO, described in Section 2.3.1.4, with an input voltage of 1.8 V. This implies a dropout voltage of 0.6 V. The chip ALDO V2 under development is expected to meet this requirement.

A TOFHIR board hosts six TOFHIR ASICs and six ALDO ASICs. The power consumption of the TOFHIR board is 3.74 W with an input current of 2.08 A at 1.8 V. The breakdown of the power requirements of the TOFHIR board is given in the third block of Table 2.11. The power consumption of the Concentrator Card is 2.31 W as shown in the fourth block of Table 2.11, which also lists a breakdown of the board components. The power consumption of one RU is 21.6 W, as shown in the last block of the table.

As a result, the entire BTL requires about 9330 W and one tray about 130 W, as shown in Table 2.12. The figures correspond to the power consumption of the detector or detector unit.

Table 2.11: Inputs and power requirements for the low voltage distribution system.

Description	Value	Unit
1. Basic inputs for low voltage power distribution system		
ALDO Voltage Drop	0.6	V
FEAST Power efficiency	80	%
Input Voltage for FEAST	10	V
2. TOFHIR chip power requirements		
Channels	32	
V_Supply	1.2	V
I	347	mA
Power	416	mW
Power/channel	13.0	mW
3. FE board power requirements		
TOFHIR ASICs	6	
Power	2496	mW
Voltage	1.2	V
Current (I)	2080	mA
ALDO ASICs	6	
V drop ALDO	0.6	V
Power drop	1248	mW
Total FE board power	3744	mW
V_in	1.8	V
4. Concentrator Card power requirements		
lpGBT	2	
TIA-OT	2	
Laser driver for VCSEL	2	
Precision clock	0.33	
Laser diode DC	2	
GBT-SCA	2	
Total CC Power	2314	mW
V1	1.2	V
I1	1715	mA
P1	2057	mW
V2	2.5	V
I2	103	mA
P2	257	mW
5. Readout Unit power requirements		
FE boards	4	
CC	1	
Power	17290	mW
Power Converter loss	3458	mW
Power in	21.6	W
V in	10	V
I in	2.16	A

Table 2.12: Summary of BTL power requirements.

Description	Voltage (V)	Current (A)	Power (W)
BTL Power	10	934	9330
Tray Power	10	13	130
RU Power	10	2.2	21.6

They do not take into account the losses on delivering that power to the detector or detector unit, such as cable losses, etc.

2.5.2 BTL low voltage system

The BTL low voltage system (LVS) provides power to the BTL on-detector electronics (Section 2.3), namely the FE Boards and the Concentrator Cards (CC). One RU comprises one voltage channel (LVCH). The power conversion for the RU is performed in two PCCs, which are part of the RUs. The PCC (Section 2.3.1.9) hosts FEAST ASIC based DC-DC converters, either as mezzanine cards using readily available FEASTMP_CLP modules or directly on the PCC. The use of DC-DC type converters represents a negative impedance to the feeding power supply, which must be taken into account for the low voltage system design.

We plan to power each LVCH individually by one floating low voltage power supply. This ensures well-defined return paths for the currents, minimizing potential noise performance problems induced by the low voltage distribution system. The consequence of this choice is that a larger number of rather small power supplies is required and more wires (of smaller cross section) are required to distribute the power.

A BTL tray will be fed by 6 LVCHs, one per Readout Unit. Thus there will be 216 LVCHs per BTL side and 432 in total. From the details of the power requirements provided in Section 2.5.1, a nominal power of 17.3 W must be delivered at the output of the PCCs per each LV channel. This corresponds to 21.6 W at 80% efficiency of the DC-DC converter.

The nominal maximum operation voltage of the FEAST ASIC is 12.0 V and the minimum operation voltage is 5.0 V. The absolute maximum rating for the input voltage is 14.0 V. We therefore require the LVS:

- To have a nominal operation voltage of $>$10.0 V at the input to the DC-DC converter;
- To operate at a minimum operation voltage of 7.0 V at the input to the DC-DC converter;
- To have a maximum output voltage at the LV power supplies of 12.0 V.

2.5.2.1 Power cables

Given the above requirements for the LVS, the maximum allowed voltage drop in the cables is 2.0 V. The power cables are divided into three sections:

1. From the power supplies on the balconies in UCX to patch panel 1 (PP1): 35 to 80 m;
2. From PP1 to the edge of the BTL trays (PP0): $\sim$4.0 m;
3. Distribution inside the tray: 0.5 to 3.0 m.

We consider the nominal input voltage to the PCC of 10.0 V. Table 2.13 summarizes the low voltage (LV) cable parameters for the different segments.

Table 2.13: BTL low voltage cable parameters.

Parameter	Low Voltage cables			Unit
	Segment 1	Segment 2	Segment 3	
Max. distance	80	4	3	m
Wire cross section	1.5	0.75	0.75	mm^2
Specific resistance	13	25.4	25.4	$m\Omega/m$
Number of wires	3	2	2	
Total specific resistance	4.3	12.7	12.7	$m\Omega/m$
Voltage drops in the power + return wires for V_{in} PCC = 10V	1.42	0.21	0.16	V

The total estimated voltage drop is 1.79 V, which is within the limit of 2.0 V. This number will slightly increase by about 50 mV by the contact resistances of the wire interconnections, resulting in a total voltage drop of 1.84 V. The total current will be 2.05 A and the LVPS output voltage 11.84 V. This configuration leaves ∼9% contingency to the power budget before exceeding the 2.0 V voltage drop limit. The system will as well operate at an input voltage to the PCC of 7.0 V. In this case the current would be 2.93 A, the voltage drop 2.61 V and the power supply output voltage 9.61 V. This constitutes an extra head room for the LVS system in case of larger power needs at the cost of going beyond 2.0 V drop in the supply lines. Moreover, having multiple wires allows the operation of the LVCH with a single broken wire in one of the segments again at the cost of a larger voltage drop.

Inside the BTL tray described in Section 2.3, we will use a cable harness made of individual wires. Two pairs of power and return wire, which terminate in a 2-pin Hirose DF64 series connectors (or similar), which in turn plugs into receptacles on the CC. A 25-pin D-sub type connector terminates the other end of the cable harness. Following the installation of the tray inside the tracker support structure, this connector is fixed inside the service channel close to the edge of the tray. Going outwards from the edge of the tray to PP1 (segment 2) the baseline solution is a cable with 26 wires of 0.75 mm^2, serving one full tray with LV. Like inside the tray two groups of power plus return wire are used for one low voltage channel, adding up to 24 wires in total. We require no special screening for this cable. The insulation is chosen to minimize the overall cable diameter while being sufficiently radiation hard. A sketch of the cable is shown in Fig. 2.76 (left). The outer diameter obtained, 13.4 mm, is indicative and might change in the final cable construction made by a cable manufacturer. The cable end on the BTL tray will be terminated with a 25-pin Sub-D type socket. The end plugging into PP1 will be either terminated by:

- a 25 pin Sub-D type plug or
- a 37 pin Sub-D type plug,

depending on the PP1 interface board layout.

Going further outwards from PP1 to the LVPS a cable with larger cross section wires is used in order to limit the voltage drop over this large supply distance. The cable will have 19 wires of 1.5 mm^2 cross section. A sketch of the cable layout is given in Fig. 2.76 (right). The estimated outer diameter is 15.0 mm. We will use three groups of power plus return wires, corresponding to six wires in total, to provide the power for one LVCH. Thus one cable corresponds to three LVCHs and consequently two such cables are used to deliver LV power to one BTL tray. The two cables will be terminated at the PP1 by:

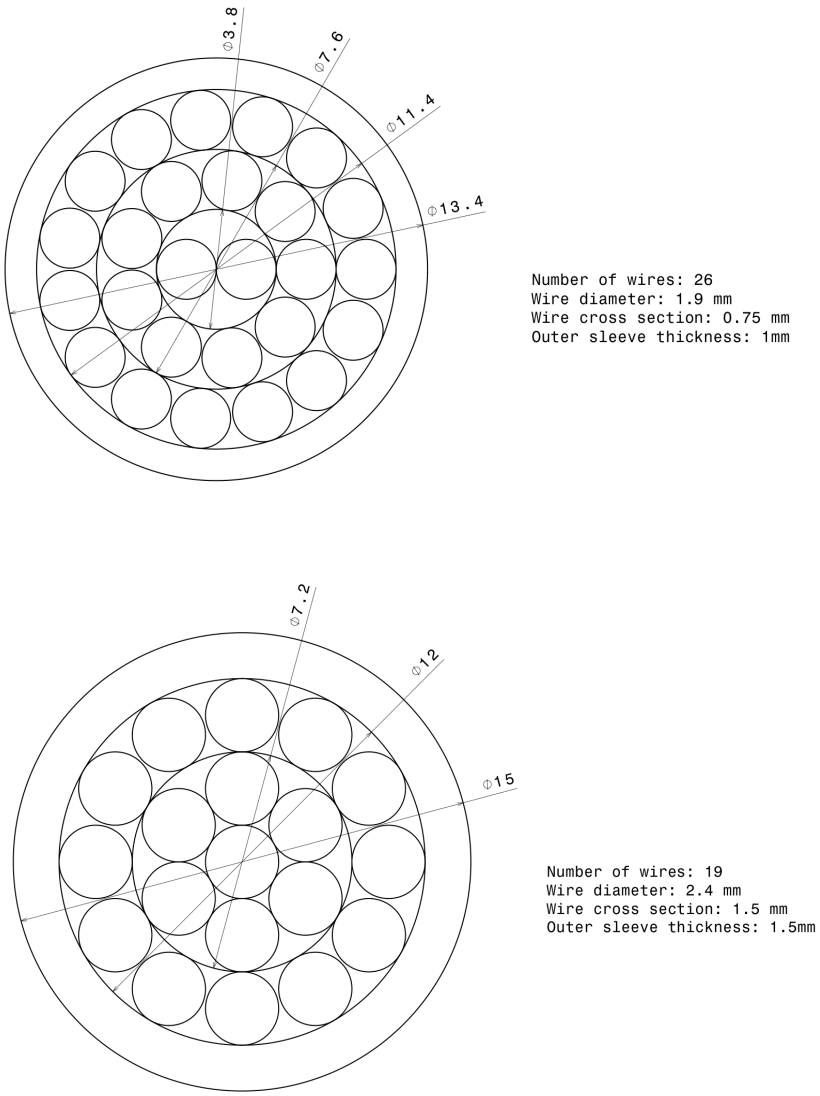


Figure 2.76: Left: LV cable between BTL tray and PP1, 26 wires of 0.75 mm^2. Right: LV cable between PP1 and the low voltage supplies, 19 wires of 0.75 mm^2. The outer sleeve material for both cable types is Polyethylene to comply with CERN safety rules.

- two 25 pin Sub-D type socket or
- one 37 pin Sub-D type socket,

depending on the PP1 interface board layout.

The termination of this LV cables on the LV power supply end will be adapted to the low voltage power supplies chosen eventually.

2.5.2.2 Power supplies

We plan to develop a modular custom solution adapted to the BTL needs based on existing CAEN or Wiener power supply solutions. The supply is required to provide floating supply of up to 3.5 A, up to 14.0 V at maximum 50 W, maintaining the voltage stable within 5 mV. Fur-

Table 2.14: Low Voltage power supply requirements.

Performance	Min	Typical	Max	Unit
Output voltage	0		15	V
Output current	0		5	A
Output power			50	W
Voltage stability peak-to-peak (Ripple + Noise) 100 MHz			<5	mV
Voltage precision			<25	mV
Operation				
Remote Sensing		NO		
Interlock, hardwired		YES		
Over voltage knockout		YES		
Under voltage knockout		YES		
Voltage adjustment			10	mV
Current limit adjustment		YES	10	mA
Over voltage limit adjustment		YES	10	mV
Voltage reading		YES	10	mV
Current reading		YES	5	mA
Control interfaces		Ethernet Can bus		

thermore, the supplies will have precise output voltage adjustment, over- and under-voltage protection, over-current protection, over-voltage and over-current limit adjustment, hardwired interlock, precise voltage and current readings and CAN-BUS and Ethernet interfaces depending on the BTL detector control systems requirements, as shown in Table 2.14.

The Wiener MPOD system with MPV 8016I modules, providing 8 channel, 0–15 V, 0–5 A, 50 W, is a good example for a starting point for the development. CAEN provides a similar system employing the A2519, 8 channel, 0–15 V, 0–5 A, 50 W power supply module. These modules fit into SY4527 and SY1527 mainframes.

The power modules and the chassis need adaptation for the radiation and magnetic field levels. All other parameters of the module satisfy or exceed our requirements. The 8U Chassis of the Wiener MPOD system can host up to 10 MPV8016I modules. One chassis equipped with nine modules has a total of 72 channels, corresponding to the LVCHs of 12 BTL trays. Thus, six chassis and 54 MPV8061I modules are required to power the entire BTL. We note that the required radiation and magnetic field tolerance of the modules and crates might increase their size. Consequently, the final bias voltage power supplies might require more crates and modules than the standard version, used here as an illustration.

2.5.3 The BTL bias voltage system

The BTL bias voltage system provides power to the silicon photo-multipliers (SiPMs) used as photo-sensors, described in Section 2.2.2. The operation voltage range is 30–77 V depending on the SiPM type eventually chosen. The absolute maximum power consumption specified for the sensors is 50 mW at end of operation of the detector. It originates from the limitations of the CO_2 cooling system of the BTL, providing maximum 25 kW of cooling power for the SiPM sensors. This corresponds to currents of 1.67 mA and 0.65 mA for 30 V and 77 V operation voltage, respectively. The operation voltage range for the SiPM covers the full span of operation voltages. In particular, it includes an over-voltage range of 0–2.5 V, the radiation

Table 2.15: HPK-S12572 photo-sensor parameters relevant for the bias supply system dimensioning (V_{br} stands for breakdown voltage).

Operation voltage and voltage stability				
	Min	Typical	Max	Unit
Over-voltage for operation		2.5		V
Radiation induced V_{br} shift after 3000 fb^{-1}		4.0		V
Temperature induced V_{br} shift (23 to -30) °C		-3.0		V
Breakdown voltage	62	66	70	V
Overall operation range	59		76.5	V
Bias voltage supply range	57		82	V
Over-voltage stability		5		%
Power consumption and current				
Power after 3000 fb^{-1}			50	mW
Power contingency factor		1.5		
Absolute maximum power after 3000 fb^{-1}			75	mW
Absolute maximum current for operation voltage range	0.88		1.44	mA

dependent breakdown voltage drift of 0–4.0 V, the temperature-dependent change of the breakdown voltage of -3.0 V, for a temperature variation between 23 and -30 °C, and the maximum breakdown voltage range specified by the manufacturer. Starting at the maximum breakdown voltage for the HPK-S12572 sensors of 70.0 V and adding the over-voltage for operation as well as the radiation-induced voltage drift we arrive at 76.5 V. We round up this value and obtain a maximum operation voltage of 77 V. The required bias voltage supply range is derived from the operation voltage taking into account the voltage drop range in the supply lines and adding some extra headroom. Operation parameters relevant for the dimensioning of the bias supply system are summarized in Table 2.15 for the Hamamatsu type photo-sensors.

The specifications for FBK type sensors and for Hamamatsu HDR2 sensors are based on R&D samples. The breakdown voltage of the FBK sensor is 36.3 V at room temperature (23 °C) and 34.2 V at -30 °C and the breakdown voltage of the Hamamatsu HDR2 sensor is 37.6 V at room temperature and 35.8 V at -30 °C. The maximum breakdown voltage range is not known. We assume a range of $\pm 10\%$ arriving at a minimum operation voltages of 30.8 V and 32.2 V for FBK and Hamamatsu HDR2, respectively. Rounding down the lower of the two values, we arrive at a lower limit of the operation voltage range of 30 V.

Groups of 384 SiPMs are biased from a single supply line, representing one bias voltage channel (BVCH). A RU serves 768 photo-sensors, corresponding to two BVCHs. Each of the BVCHs serves the SiPMs connected to two front-end cards in the RU. A tray with six RUs requires 12 BVCHs. The full BTL, comprising 72 trays, has 864 BVCHs. Given the above stated requirement of 50 mW per SiPM, the absolute maximum power at the load of a BVCH is 19.2 W. The power required per tray is 231 W and 16.6 kW for the full detector, as summarized in Table 2.16.

The bias voltage system uses floating power supplies which connect with four wires to the load, providing remote sensing of the bias voltage. The supply lines as well as the remote sense wires are connected to one of the two FE cards belonging to the same BVCH. At this point also the supply and sense wires are connected. The second FE board, part of the same BVCH, is connected via a jumper cable. The floating returns of the BVCHs are connected at the FE card to the reference potential of the FE card, which is in turn connected to the reference potential of the Readout Unit. A connection of the reference potential of the Readout Unit to a common tray or detector reference voltage is optional. It is not mandatory for the operation of

Table 2.16: BTL bias system modularity and power requirements.

	Bias channels	Photo-sensors	Power	Unit
Photo-sensor		1	50	mW
Bias channel	1	384	19.2	W
Readout Unit	2	768	38.4	W
Tray	12	4608	231	W
Half BTL	432	165888	8.30	kW
BTL	864	331776	16.6	kW

Table 2.17: Bias cable parameters and parameter estimate.

Parameter	Total	Segment 1	Segment 2,3	Unit
Maximum Power	19.2			W
Load Voltage	27			V
Load Current	0.64			A
Bias Supply output voltage	30.5			V
Cross section		0.75	0.35	mm^2
Resistance		25.4	86.5	$m\Omega/m$
Cable length		80	7	m
Total Resistance (bias + return)		4.1	0.8	Ω
Current density		1.0	1.9	A
Voltage drop	3.5	2.9	0.37	V
Power drop	2.5	2.1	0.4	W
Efficiency	87	89	98	%

the system.

2.5.3.1 BTL bias cables

The bias voltage cables are split into three segments. Segment 1 runs over 35 to 80 m from the bias supplies to patch panel 1 (PP1). Segment 2 runs from PP1 to the edge of the BTL trays. It is 4 m long. Segment 3, distributes the bias voltage inside the tray, with a maximum of $\sim$3 m cable length. For the segment 1 we use cables with 26 individual wires of 0.75 mm^2 comprising 6 bias power groups of 4 wires, bias, bias return, sense and sense return, shown in Fig. 2.77. Cable construction includes a braided shielding, high frequency shielding foil and a drain wire connected to the detector reference potential at the entrance of the detector enclosure. The cable will have an approximate outer diameter of 13.4 mm, depending on the isolation material of the wires and the outer sleeve (Fig. 2.77). Two such cables provide the bias to the 12 bias channels of one tray, corresponding to 144 bias cables. In the segment 2 we use the same construction, channel counting and arrangement, just with a smaller wire cross section of 0.35 mm^2, which will result in an outer diameter of approximately 11 mm. For estimating the maximum voltage drop in the supply cables, we use a load voltage of 27 V, corresponding to the minimum operation voltage of 30 V -10%. We obtain a maximum voltage drop of 3.5 V. Another voltage drop of $\sim$1.5 V is estimated in the bias voltage distribution and voltage adjustment network. Both of these numbers are taken into account for the BVPS output voltage requirements. Current densities in the wires are 1.0 and 2.0 A/mm^2 for segments 1 and 2, respectively. Details about cable parameters are given in Table 2.17. Inside the BTL tray, segment 3, wire cross sections between 0.25 and 0.5 mm^2 will be used. The above values are obtained with 0.35 mm^2 wire cross section for the distribution of the bias voltage inside the trays.

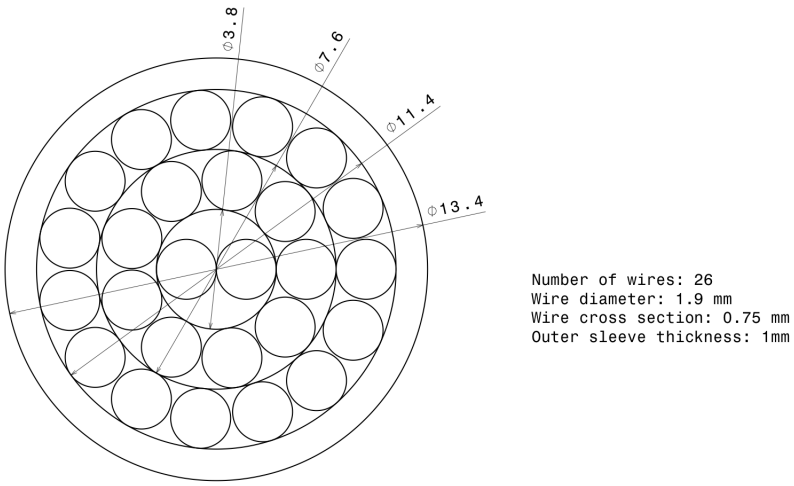


Figure 2.77: Sketch of bias cable construction with 0.75 mm^2 wires (segment 1), without braid, high frequency shielding, and drain wire.

Table 2.18: Bias voltage channel power and current ranges at 50 mW per photo-sensor.

	Min	Typical	Max	Unit
Operation voltage limits	30		77	V
SiPM Current	0.65		1.67	mA
Bias channel current	0.25		0.64	A
Bias channel power			19.2	W

2.5.3.2 Bias voltage supplies

The bias voltage supplies are located on the balconies inside the experimental cavern, at a distance of 35 to 80 m from the detector. The radiation levels at this location are a total integrated dose of 32 Gy, a proton fluence of 8×10^{10} cm^{-2} (>20 MeV) and 1 MeV equivalent neutron fluence of 8×10^{11} cm^{-2} for a maximum integrated luminosity of 4000 fb^{-1}. The magnetic field in the location of the power supplies is <0.1 T.

The power required per BVCH of 19.2 W, corresponding to 0.64 and 0.25 A for 30 and 77 V operation voltage, respectively (Table 2.18), is typical for low voltage supplies.

We plan to develop a modular custom solution adapted to the BTL needs based on existing CAEN or Wiener power supply solutions. The bias voltage supply (BVPS) is required to provide floating power with a current of up to 1.0 A, up to 82 V at a maximum of 50 W, with remote sensing, maintaining the voltage stable within 8 mV. Furthermore, the supplies will have precise output voltage adjustment, over and under voltage protection, over current protection, over voltage and over current limit adjustment, hardwired interlock, precise voltage and current readings with CAN-BUS and Ethernet interfaces depending on the BTL detector control systems requirements, shown in Table 2.19. These values will be adjusted as soon as the final sensor type is selected.

The Wiener MPOD system is a good example for a possible starting point and will be used for illustration. We will use MPV8060I 8 channel modules providing 0–60 V, 1 A and 50 W/ch and a static voltage regulation of better than 10 mV. More details are available in the Wiener MPOD

Table 2.19: Photo-sensor bias supply channel parameters at 50 W.

Performance	Min	Typical	Max	Unit
Bias supply voltage range	25		82	V
Bias supply current range	0		1	A
Bias supply power			50	W
Voltage stability (Ripple + Noise)			<8	mV
Voltage precision			<25	mV
Operation				
Remote Sensing		YES		
Interlock, hardwired		YES		
Over voltage knockout		YES		
Under voltage knockout		YES		
Voltage adjustment			<8	mV
Current limit adjustment		YES		
Over voltage limit adjustment		YES		
Over current limit adjustment		YES		
Voltage reading		YES	<8	mV
Current reading		YES	0.6	mA
Control interfaces		Ethernet CAN bus		

system technical specifications [64]. An adaptation of the modules with respect to the maximum output voltage range and for magnetic field and radiation tolerance is required. All other parameters already satisfy or exceed our needs. The modules are hosted in an adequate MPOD 19 inches rack-mountable crate with 8U height using bottom cooling air intake. The crate hosts an MPOD controller as a plugin card providing Ethernet, CAN-bus and USB control interfaces. In addition it provides space for 10 MPV8060I modules. The crate requires modification in particular with respect the primary power supply, in order to meet the magnetic field requirements.

One crate will host nine modules, providing bias to the 72 channels of four trays, thus in total 18 crates, nine per BTL side, are required. We note that the required radiation and magnetic field tolerance of the modules and crates might increase their size. Consequently, the final bias voltage power supplies might require more crates and modules than the standard version, used here as an illustration.

Chapter 3

The Endcap Timing Layer

3.1 Overview and principle of operation

The endcap regions of the CMS detector will be instrumented with two disks of MIP-sensitive silicon devices with excellent time resolution, covering a pseudorapidity range from about 1.6 to 3.0. This Endcap Timing Layer (ETL) will be mounted in its own independent, thermally isolated volume, on the nose of the endcap calorimeter (CE). Specifically, the ETL will be located on the interaction side of the neutron moderator at a distance of about 2.98 m from the interaction point, as shown in Fig. 3.1. It uses a cold, dry volume that is isolated from the CE so that the two detectors can each be operated independently of the cooling flow in the other detector. This allows independent access to the ETL during LHC shutdown periods for repairs and replacements of faulty components.

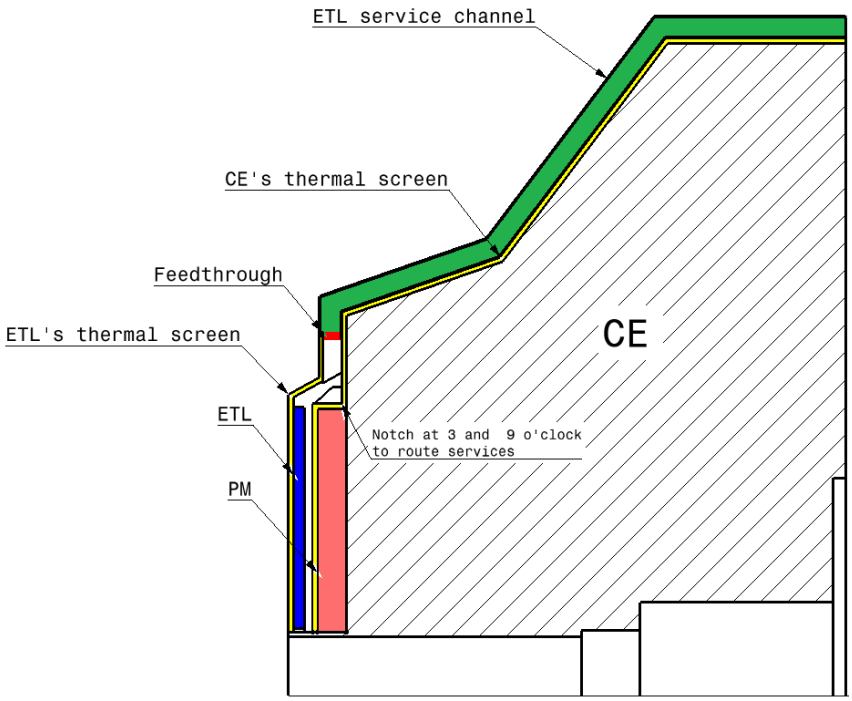
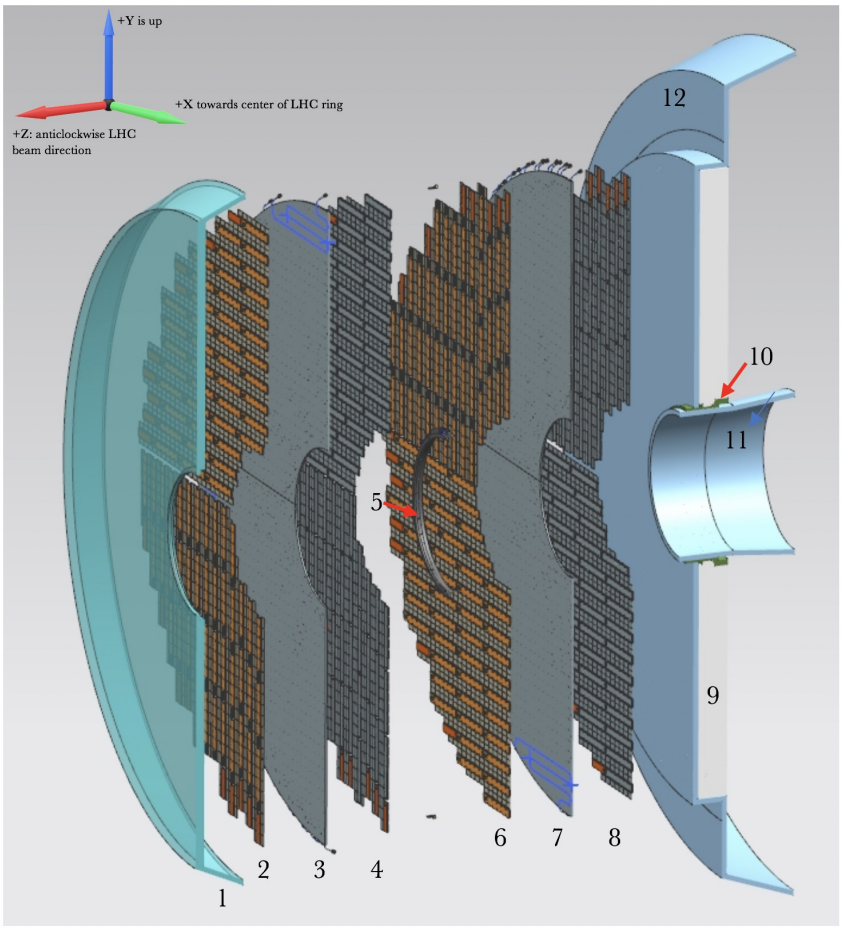


Figure 3.1: Placement of the endcap timing layer on the calorimeter endcap structure. The ETL (shown in blue) is placed on the interaction side of the polyethylene neutron moderator (shown in red). The ETL and CE detectors are in two separate cold volumes, with each detector having its own thermal screen (shown in yellow). This makes it possible to access the ETL detector for maintenance during cold operation of the CE detector.



1: ETL Thermal Screen
2: Disk 1, Face 1
3: Disk 1 Support Plate
4: Disk 1, Face 2
5: ETL Mounting Bracket
6: Disk 2, Face 1
7: Disk 2 Support Plate
8: Disk 2, Face 2
9: HGCal Neutron Moderator
10: ETL Support Cone
11: Support cone insulation
12: HGCal Thermal Screen

Figure 3.2: Cross-sectional view of the ETL along the beam axis. Shown are two disks populated with modules on both faces, along with the support structure and CO_2 cooling pipe inlets. The interaction point is to the left of the image.

The detailed position of the disks and the thermal screen are illustrated in Fig. 3.2. Sensor modules are mounted on all four faces of the two disks in each endcap in an $x - y$ layout as shown in Fig. 3.3. The sensors are placed in a staggered way such that areas for read out, power, and cable infrastructure, arranged in channels along each line of sensors on one face, are covered by the sensors on the opposite face. The fractional area of each disk that is sensitive to MIPs is greater than about 85%. The use of two such disks per endcap, adjacent to each other with 20 mm z-separation, provides hermetic coverage and an average of about 1.7 hits per track, with a total sensor area of about 7.9 m^2 per endcap. The symmetry of the layout allows each disk to be composed of four identical and independent 90° structures, called wedges, which can be installed onto the CE or removed from the detector even with the beam pipe in place.

The total amount of space required along the beam axis for the detector is 45 mm, plus 20 mm for the additional thermal screen between the ETL and CE cold volumes. This space can be accommodated on the nose of CE without changing its envelope by reducing the thickness of the polyethylene neutron moderator from 15.7 cm, the value used in the CE TDR [7], down to 12 cm. The neutron moderator protects the tracker from the flux of neutrons that originate from hadron interactions in the CE. The reduced neutron moderator thickness increases the particle flux in the tracker; the largest change of fluence is around 12–15% and occurs at the largest radius of the last tracker endcap double-disks (TEDD-5). The change in fluence is about 5–7%

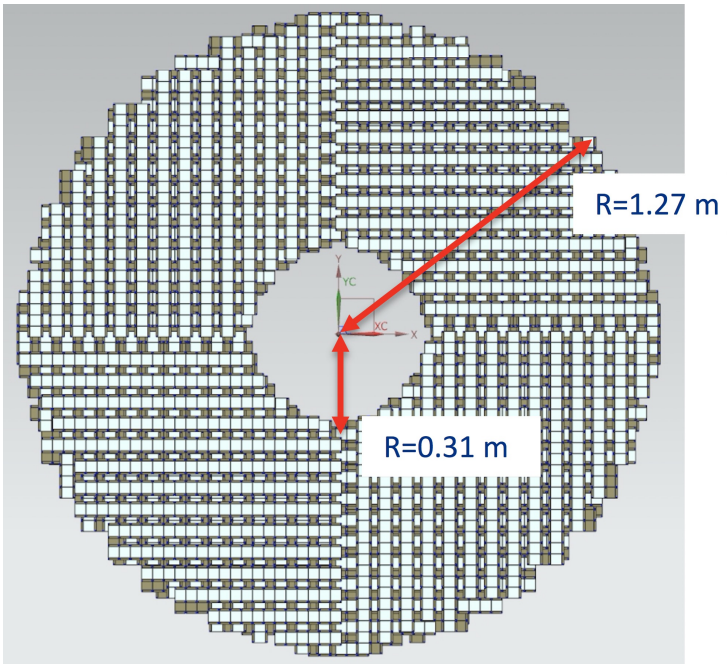


Figure 3.3: Sensor placement on one face of a disk. The modules are arranged in an $x-y$ layout, with readout channels placed between them to host powering and readout circuitry and provide cable servicing.

at intermediate radii of TEDD-5 and about 2–4% at the innermost radii of TEDD-5. The change in fluence for the rest of the OT is below 5%.

3.1.1 Radiation levels

The radiation level expected at the ETL location are the main driver for the choice of silicon sensors over other technologies. Figure 3.4 (left) shows the fluence as a function of radius, at three different moments in time during the HL-LHC physics program; the fluence will reach values above 1×10^{15} n_{eq}/cm^2 only in the second half of the HL-LHC period for radii below 50 cm. On the right-hand side, the maximum fluence reached by a given fraction of the ETL detector is shown. These plots show that a large fraction of ETL will receive a rather mild dose: 50% of the sensors will be exposed to a fluence of less than 5×10^{14} n_{eq}/cm^2, 80% to less than 8×10^{14} n_{eq}/cm^2, while the innermost 10% of the detector will be exposed to more than 1×10^{15} n_{eq}/cm^2.

3.1.2 Relevant parameters in the determination of the time resolution

The timing capabilities of silicon sensors can be understood by modeling the sensor as a capacitor (C_{Det}) with a current source in parallel, read out by a pre-amplifier that shapes the signal. The pre-amplifier output is then compared to a fixed threshold (V_{Th}) to determine the time of arrival. The time resolution can be expressed as the sum of several terms: (1) jitter; (2) fluctuations of the ionization process, both in its total amount and in its local distribution, producing amplitude and shape variations, respectively; (3) signal distortion; and (4) TDC binning:

$$\sigma_t^2 = \sigma_{\text{Jitter}}^2 + (\sigma_{\text{Total ionization}} + \sigma_{\text{Local ionization}})^2 + \sigma_{\text{Distortion}}^2 + \sigma_{\text{TDC}}^2. \tag{3.1}$$

Each of these terms is described below (see Ref. [65] for details):

- σ_{Jitter}: the jitter term is given by the ratio of the noise N over the signal slew rate $\mathrm{d}V/\mathrm{d}t$, $\sigma_{\text{Jitter}} = N/(\mathrm{d}V/\mathrm{d}t)$. The noise is the sum of components from electronic noise and sensor shot noise: $N = \sqrt{N^2_{\text{El.}} + N^2_{\text{Shot}}}$. The sensor shot noise is a sub-leading contribution, see Section 3.2.4. Ignoring N_{Shot}, the jitter can be expressed as $\sigma_{\text{Jitter}} \propto \frac{e_n C_{\text{d}}}{Q_{\text{in}}}\sqrt{t_{\text{rise}}}$ where e_n is the electronic noise, C_{d} the detector capacitance, Q_{in} the total signal charge and t_{rise} the signal rise time at the input of the comparator. The jitter is therefore minimized by large signals, small capacitance, and fast rise time.
- $\sigma_{\text{Total ionization}} + \sigma_{\text{Local ionization}}$: the ionization process changes on an event-to-event basis both in total magnitude ($\sigma_{\text{Total ionization}}$) and in the non-uniform creation of electron-hole pairs along the particle path ($\sigma_{\text{Local ionization}}$). These two effects are correlated as large non-uniform ionizations often lead to large total ionizations, e.g., due to delta rays. The effect on the time resolution of varying total ionization, the so-called time-walk effect, is largely compensated using a correction from measurements of the total ionization with a time-over-threshold circuit (Section 3.3.6). The second term, $\sigma_{\text{Local ionization}}$, arises from the variation of the signal shape due to non-uniform ionization. These signal shape variations, called Landau noise, are the intrinsic limiting factor for the achievable time resolution. They depend on the sensor thickness [65], but not on the gain value. As explained in more detail in Section 3.2.4.6, the Landau noise contributes 30–35 ps to the time resolution, and it dominates over the jitter term once the gain is larger than about 10–20.
- $\sigma_{\text{Distortion}}$: this term is due to the non-uniform weighting field and the non-saturated drift velocity. The first term is reduced to a small contribution by using a parallel plate geometry that has a uniform weighting field; in the ETL design each pad has an extension of at least 1 mm in each direction, while the thickness is about 50 μm, yielding an almost perfect parallel plate configuration. Distortion due to the non-saturated drift velocity is minimized by operating the sensor at a sufficiently high bias voltage where the charge carriers' velocity is saturated.
- σ_{TDC}: the effect of the TDC binning is discussed in Section 3.3.6.

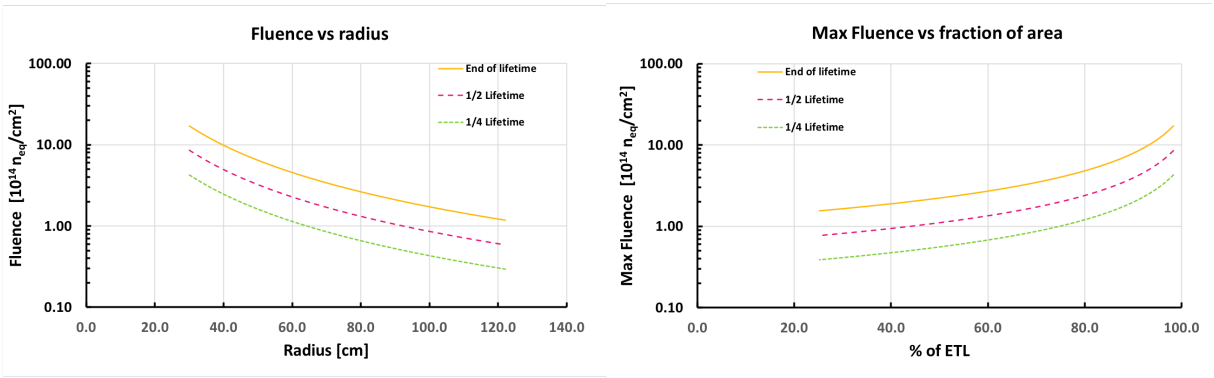


Figure 3.4: Left: ETL exposure to irradiation, in 1 MeV neutron equivalent per cm^2, as a function of radius for three points in time during the expected HL-LHC 3000 fb^{-1} lifetime. Right: The maximum fluence experienced by a given fraction of the ETL area.

3.2 Silicon sensors

3.2.1 Design and specifications

The design for a hermetic precision MIP timing detector in the CMS endcap region requires a uniform and efficient device capable of operating with sufficient radiation resistance to maintain performance throughout the lifetime of the HL-LHC. To meet these needs the ETL will be instrumented with Ultra-Fast Silicon Detectors (UFSDs), planar silicon devices based on the Low-Gain Avalanche Detector (LGAD) technology [27, 28]. UFSDs incorporate a low, controlled gain in the signal formation mechanism, as shown in Fig. 3.5. Charge multiplication happens when the charge carriers are in electric fields above $E \approx 300$ kV/cm, when the electrons (and to a lesser extent the holes) acquire sufficient kinetic energy to generate additional e/h pairs. The field value can be obtained by implanting an appropriate doping density ($N_D \approx 10^{16}/\text{cm}^3$) that locally generates very high fields when depleted. The gain has an exponential dependence on the electric field $N(l) = N_o e^{\alpha(E)l}$, where $\alpha(E)$ is strongly dependent on the electric field and l is the path length inside the high-field region. The gain layer is realized through the addition of a p-type implant and, to avoid breakdown, its lateral spread is controlled by deep n-doped implants, called Junction Termination Extension (JTE). Typical gain values are in the 10–30 range, which is modest compared to gains of thousands or more in APDs or SiPMs. Three vendors have successfully produced optimized UFSDs, which have been tested by CMS and are being considered for providing the ETL sensors: Centro Nacional de Microelectronica (CNM), Barcelona [27, 66, 67], Fondazione Bruno Kessler (FBK) [68, 69], and Hamamatsu Photonics (HPK) [70, 71].

Achieving good time performance at low gain requires pixels (also called pads) with a size less than a few mm^2, to limit the sensor capacitance, implying that a large number of pixels is required to cover the 7.9 m^2 of each ETL endcap. The design studied in the 2017 CMS MTD Technical Proposal (TP) used very large sensors, 5×10 cm^2, with 3×1 mm^2 pixels. Our R&D and design optimization studies have led to an updated design that now features smaller sensors, 21.2×42 mm^2 with square pixels of 1.3×1.3 mm^2. We have prepared a wafer layout for these sensor sizes, as shown in Fig. 3.6, and are pursuing updated quotes with vendors. The adoption of smaller sensors and pixels relative to the TP design has two important cost saving advantages: the smaller sensor size makes more efficient use of the 6-inch silicon wafers, and smaller sensors have a higher yield. An important advantage of the new sensor design is that the pixel capacitance is halved, improving the timing performance of the ASIC.

An extensive R&D program is being executed in order to optimize sensor manufacturing details for the needs of the ETL. This includes examination of different doping schemes to maximize radiation resistance, studies aimed at improving inter-pixel gaps and reducing dead area, and tests of sensor uniformity and performance towards the ultimate vendor selection.

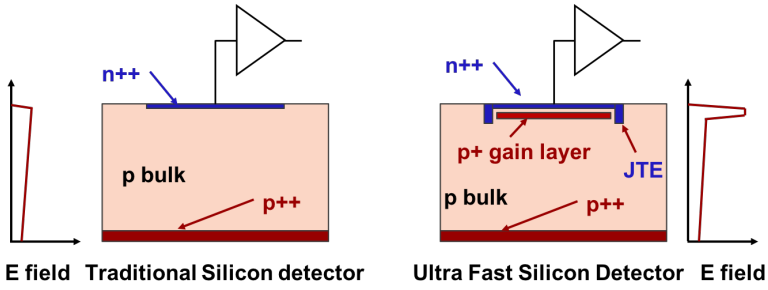


Figure 3.5: Cross-sectional diagrams comparing a standard silicon detector and a UFSD with an additional p implant providing the larger electric field needed for charge multiplication.

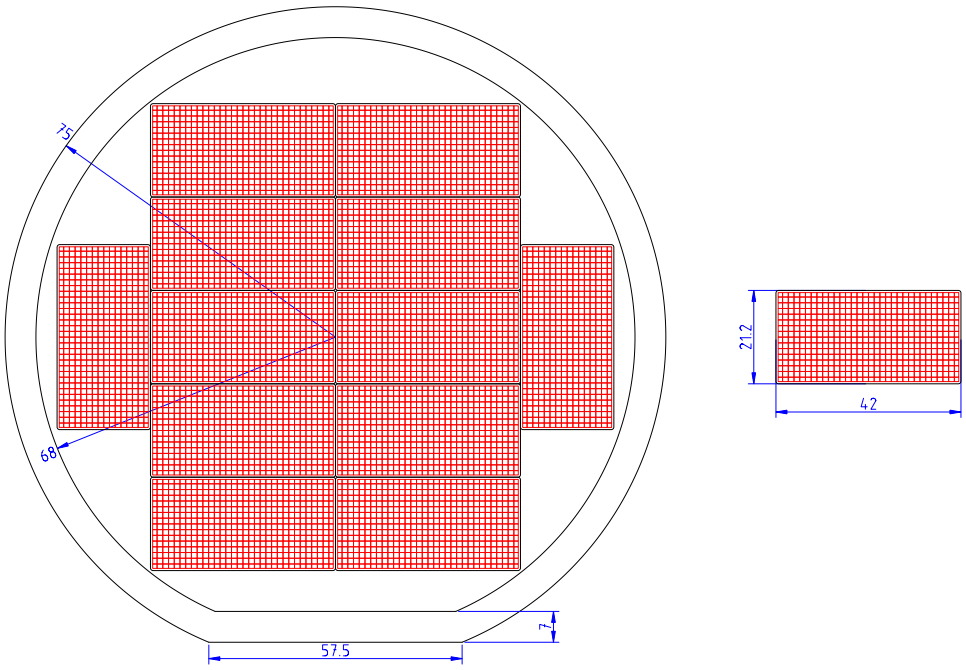


Figure 3.6: Layout of the ETL sensors on a 6-inch wafer, with the pixels shown. We are also investigating with vendors the feasibility of a layout with a more aggressive approach of using a 6 mm stay-clear region around the edge of the 75 mm wafer, instead of the 7 mm stay-clear scenario shown here.

3.2.2 Radiation hardness studies

Radiation damage in silicon sensors causes three types of effects:

1. decrease of the charge collection efficiency,
2. increase in leakage current, and
3. changes in the sensor doping profile, decreasing the effectiveness of the amplification layer.

These three effects impact UFSDs in a different way than standard PiN diodes: (1) since the sensitive region in a UFSD is thin, with an $\approx$ 50 μm thick depletion region within a normal 300 μm thick wafer, the decrease of charge collection efficiency is moderate; (2) for the same reason, the increase in leakage current is also moderate, however the current is multiplied by the internal gain. The leakage current should be kept low to control both power consumption and shot noise, so low temperature operation ($\approx -30\,^{\circ}$C) is necessary for irradiated UFSDs. The last point, (3) changes in the sensor doping profile, represents the most important aspect for UFSD operation, as it directly impacts the gain mechanism. It will be explained in more detail in this section.

It has been shown in previous studies [72–74] that neutron and charged hadron irradiation reduces the gain in UFSDs. This effect is caused by the *initial acceptor removal* mechanism that progressively deactivates the acceptors forming the gain layer. The effects of initial acceptor removal on the silicon bulk was first measured in standard boron-doped silicon sensors more than 20 years ago [75]. In addition, irradiation causes the creation of acceptor-like defects due

to the creation of deep traps. These combined effects make the acceptor density ρ_A depend on the fluence ϕ as [65, 76]

$$\rho_A(\phi) = g_{\text{eff}}\,\phi + \rho_A(0)\,e^{-c(\rho_A(0))\phi}, \tag{3.2}$$

where g_{eff} = 0.02 cm^{-1} (chapter 5 of Ref. [77]) is the acceptor creation coefficient, ϕ is the irradiation fluence with units of cm^{-2}, $\rho_A(0)$ ($\rho_A(\phi)$) is the initial (after a fluence ϕ) acceptor density with units of cm^{-3}, and c is a parameter, with dimension [cm^2], that depends on the initial acceptor concentration $\rho_A(0)$ and on the type of irradiation. The first term of Eq. 3.2 accounts for acceptor creation by deep traps. The second term accounts for the initial acceptor removal mechanism wherein the acceptor is dislocated from the lattice. The key parameter in the characterization of the radiation hardness of an UFSD sensor is the coefficient $c(\rho_A(0))$ in the exponent of Eq. 3.2: lower values of $c(\rho_A(0))$ indicate a slower acceptor removal with irradiation. This coefficient depends on both the initial gain-layer acceptor concentration and the presence of impurities, implanted to slow down the effect of irradiation. The factor c can be rewritten as the fluence, $\phi_o = 1/c$, that reduces the initial doping density $\rho_A(0)$ to 1/e of its initial value.

During the development of the ETL sensors, several UFSD productions with different acceptor types (boron and gallium), different gain layer doses and diffusion processes (obtained using two different temperatures indentified as *high* and *low*), and/or added impurities, have been performed in order to evaluate which option yields the most radiation-hard solution. Productions with the gain layer composed of high-diffusion boron, low-diffusion boron, gallium, boron plus carbon, and gallium plus carbon have been completed and exposed during several irradiation campaigns to neutrons or charged hadrons.

The following irradiation campaigns are being used for the radiation resistance studies:

- neutrons at the research reactor at JSI institue in Ljubjiana [78];
- 23 MeV protons at the Karlsruhe Institute of Technology (KIT), Germany [79];
- 70 MeV protons at the Cyric Cyclotron and Radioisotope Center (CYRIC), Tohoku University, Japan [80];
- 800 MeV protons at the Los Alamos Neutron Science Center (LANSCE) proton accelerator, USA (in progress);
- 24 GeV protons at the IRRAD facility at CERN, Switzerland [81].

Since the damage to the silicon lattice from charged hadrons and neutrons dependends on the particle energy, the fluences are multiplied by a non-ionizing energy loss (NIEL) factor that scales it to an equivalent 1 MeV neutron fluence. The NIEL factor, however, has been computed to normalize the amount of leakage current created by hadrons to that created by 1 MeV neutrons. It might not be the correct factor to also normalize the acceptor removal rate, since the type of lattice damage for the processes is different. For the charged hadron irradiation campaigns, the proton energies were chosen to cover a large range of NIEL values in order to test the compliance with the NIEL hypothesis, as shown in Fig. 3.7.

Figure 3.8 shows the measured values of $c(\rho_A(0))$ from neutron irradiation of sensors from the different foundries and for different densities of implanted carbon, where A is the lowest dose, B is twice, C is three times, and D is five times the dose of A. In this plot, lower values of $c(\rho_A(0))$ indicate slower acceptor removal rates and therefore higher radiation resistance.

Table 3.1 reports the measured values of $c(\rho_A(0))$ for selected proton and neutron irradiations

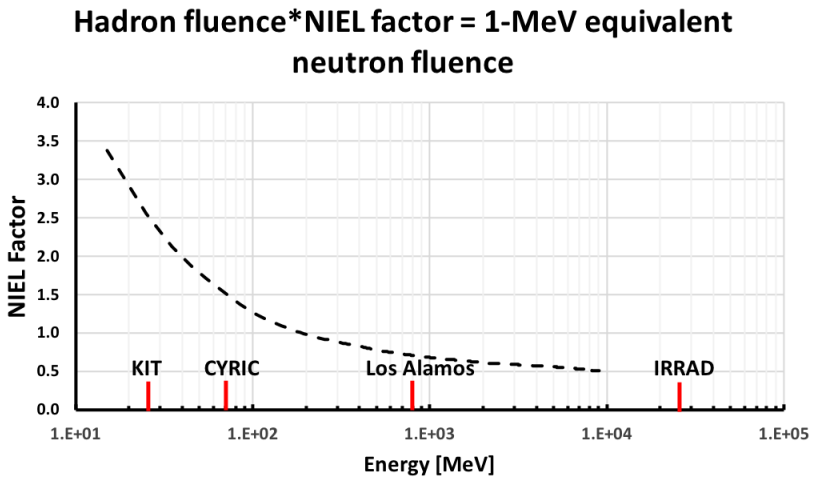


Figure 3.7: NIEL factor for protons as a function of energy and the four energies used in the irradiation campaigns.

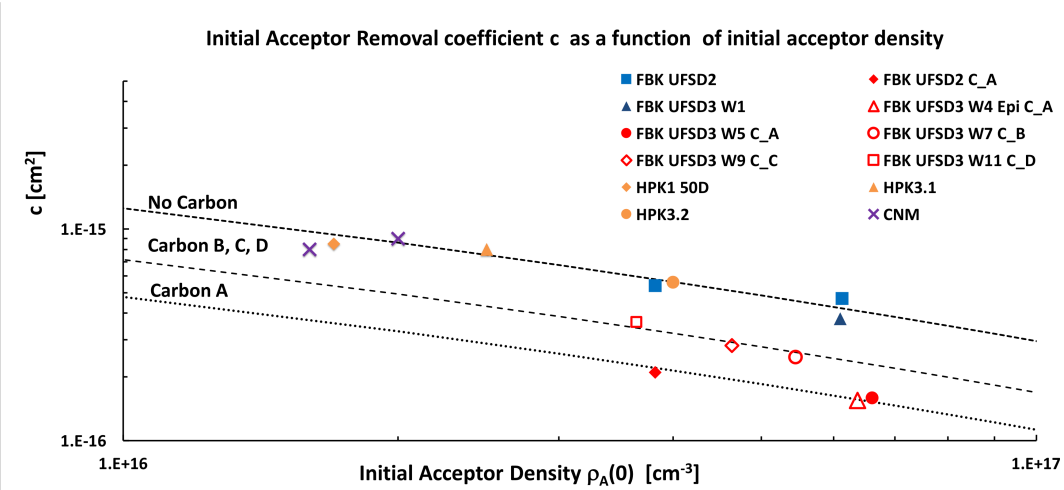


Figure 3.8: Value of the acceptor removal coefficient $c(\rho_A(0))$ for different vendors and densities of carbon. The plot also shows three parametrizations of $c(\rho_A(0))$: (top) no carbon, (middle) carbon dose B, C, and D, and (bottom) carbon dose A.

(the Los Alamos samples have not yet arrived). The first 3 columns list the values of $c(\rho_A(0))$ without applying the NIEL factor, while the last two columns report the measured value of the NIEL factor extracted from these measurements. The results up to this point show that the acceptor removal rate caused by protons is about a factor of two higher than predicted by the NIEL factor. Even including this factor, the damage from charged hadrons, given their moderate fluence in ETL, does not constitute a problem for the ETL operation.

Our irradiation campaigns have demonstrated that (1) a moderate amount of carbon added to the gain layer is beneficial, (2) higher carbon densities are detrimental, (3) higher initial acceptor densities are beneficial, and (4) the use of gallium instead of boron for the gain layer doping is detrimental [74].

The different values of $c(\rho_A(0))$ determine the evolution of the value of the gain with irradiation, and by how much the bias voltage needs to be increased to compensate for the loss of

Table 3.1: Summary of the measured $c(\rho_A(0))$ for proton and neutron irradiations. The last two columns show the NIEL factors calculated from these measurements.

$c(\rho_A(0))$ for:	KIT	IRRAD	JSI	KIT/JSI	IRRAD/JSI
	p, 23 MeV	p, 23 GeV	n	Expected NIEL = 2–3	Expected NIEL = 0.4–0.6
Sensor	[cm^2]	[cm^2]	[cm^2]	Measured NIEL	Measured NIEL
FBK UFSD2 B+C	7.8	3.3	2.1	3.7	1.6
FBK UFSD2 B	16.6	6.5	5.4	3.1	1.2
FBK UFSD3 B LD + C	6.5		1.56	4.2	

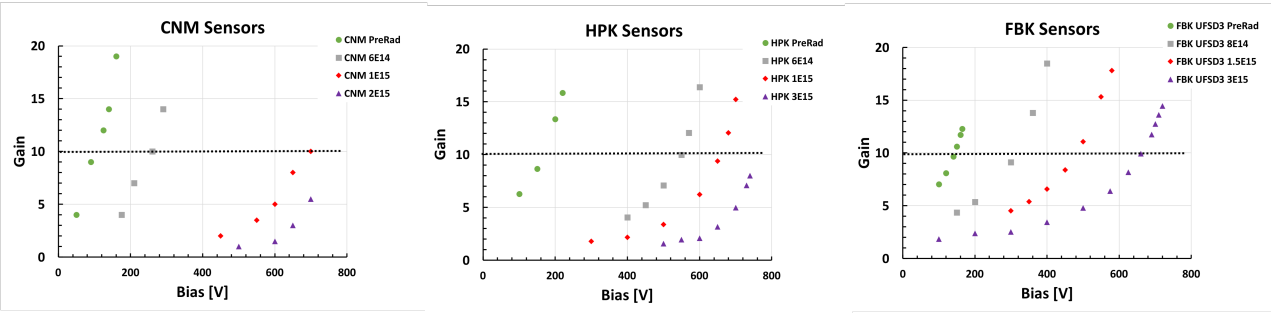


Figure 3.9: Gain as a function of bias voltage for different neutron fluences for sensors manufactured by CNM, HPK, and FBK. The dashed line in each plot shows gain = 10.

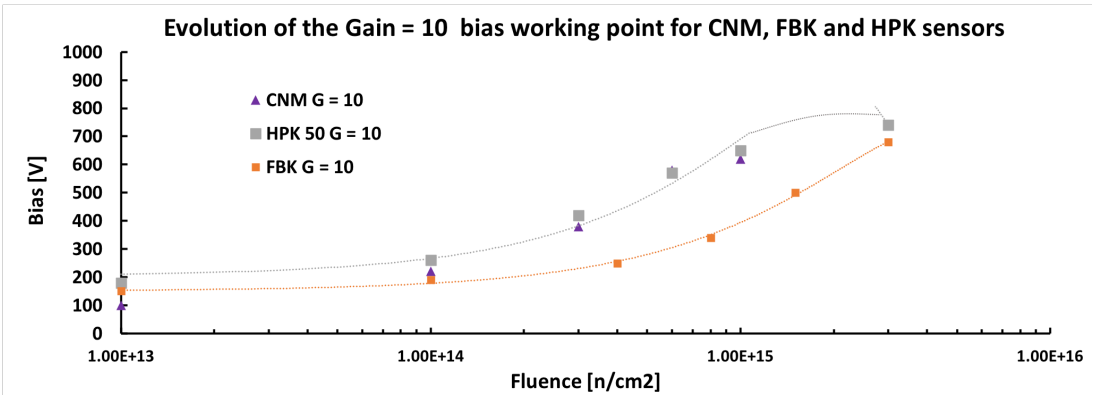


Figure 3.10: Bias voltage required to maintain a gain of 10 as a function of fluence for sensors from HPK, CNM, and FBK. The dotted lines on the plot show the parametrizations used in the calculations for the power consumption.

doping. Figure 3.9 shows the evolution of the gain value as a function of bias voltage at different fluences for HPK, CNM, and FBK. In each plot, a dashed black line indicates gain = 10. Figure 3.10 shows the bias voltage required to maintain a gain of 10 as a function of fluence for sensors from HPK, CNM, and FBK. Interestingly, sensors from HPK and CNM present the same dependence, while FBK sensors, given the lower values of $c(\rho_A(0))$ from the presence of carbon, can obtain a gain of 10 at a lower bias as a function of fluence. The studies performed so far indicate that all vendors can successfully manufacture sensors able to deliver a gain of at least 10 up to the end of the CMS HL-LHC lifetime.

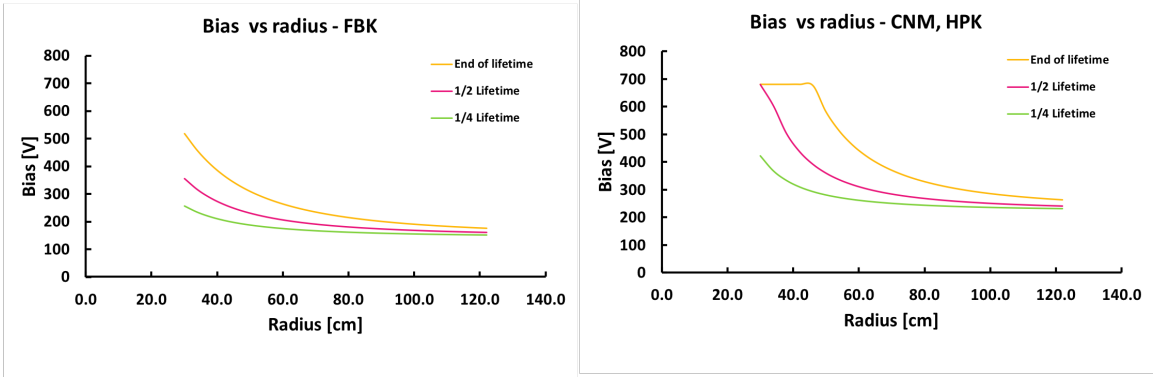


Figure 3.11: Bias voltage required to maintain a gain of 10 as a function of radius for sensors from FBK (left) and HPK or CNM (right) at three points in time during the expected HL-LHC 3000 fb^{-1} lifetime.

3.2.3 Biasing scheme and power dissipation

The parametrization shown in Fig. 3.10 can be used to predict the value of the UFSD bias voltage as a function of radius at any point in time during the lifetime of HL-LHC. Figure 3.11 shows the bias voltage as a function of radius for 25, 50, and 100% of the HL-LHC lifetime for FBK (B+C) and HPK or CNM sensors. The plots show that a larger increase of the bias voltage is needed as a function of the radius to maintain a gain of 10. For CNM and HPK (right side of Fig. 3.11) the gain layer is almost totally deactivated at small radius by the end of the lifetime. In this condition, the necessary gain is obtained by applying the highest possible value of bias compatible with low-noise operation. As Fig. 3.20 shows, after heavy irradiation (above 1×10^{15} $\text{n}_{\text{eq}}/\text{cm}^2$) FBK (top plot) and HPK (bottom plot) sensors can withstand up to 700–730 V without reaching breakdown, but the time resolution worsens since the gain at the maximum operating bias voltage is lower.

As is shown in Fig. 3.3, the sensors will be placed on the ETL surface on a grid aligned along the $x - y$ coordinates, so some sensors will have the short side (2 cm) along the radius, while others will have the long side (4 cm) along the radius. Figure 3.11 shows that the part of the sensor at lower radius would have an optimum bias voltage higher than its other part located at a larger radius. Since a single bias voltage is applied to the whole sensor, and its value is driven by the side that can hold the lower voltage, the other side will be biased to a value that is lower than optimum. Figure 3.12 shows the difference in optimum bias voltage for the sensors aligned with the long side along the radius (worst case). For the FBK sensors using carbon co-implantation (left side), the maximum underbias will be about 40–60 V, while for sensors from CNM and HPK (right side) it will be about 70–100 V. The CNM and HPK sensors will reach a complete removal of the gain layer at about 50% of the HL-LHC lifetime, yielding a constant optimum bias voltage at the maximum of 700 V.

Irradiation increases the sensor leakage current and causes an increase of the optimum bias voltage; both factors contribute to an increase of the sensor power consumption. Figure 3.13 reports the power generated by the bias current in each 4 cm ring for a single sensor layer kept at $T = -20\,^{\circ}\text{C}$. As the sensors from FBK are biased at a lower voltage, their power requirement is marginally lower. As listed in Table 3.2, the power generated by the sensors in one ETL layer at $T = -20\,^{\circ}\text{C}$ is for FBK (CNM, HPK) about 40 W (50 W) at 25% lifetime, about 90 W (140 W) at 50% lifetime, and about 230 W (360 W) at 100% of lifetime. The power consumption decreases by a factor of about 3 going from $T = -20\,^{\circ}\text{C}$ to $T = -30\,^{\circ}\text{C}$.

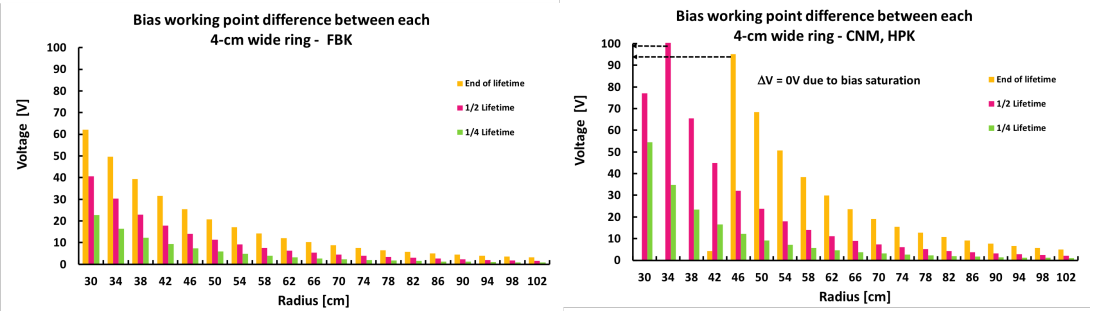


Figure 3.12: Difference in the optimum bias point over a radial distance of 4 cm at three points in time during the expected HL-LHC 3000 fb^{-1} lifetime. The results for sensors from FBK are shown on the left and results from CNM and HPK sensors are shown on the right. The dashed lines indicate a region where the sensors are biased to the maximum possible voltage.

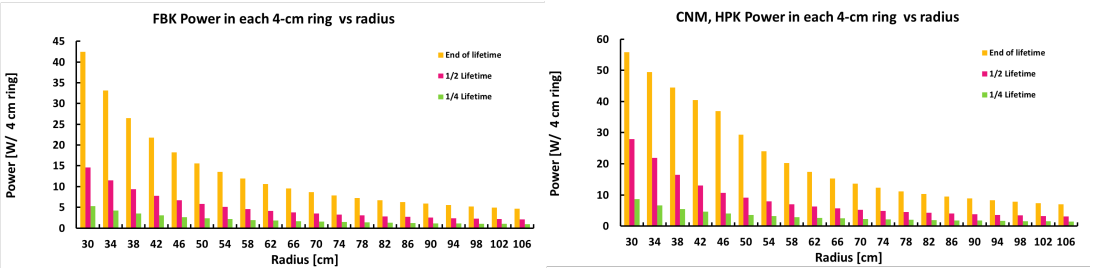


Figure 3.13: Power consumption from bias voltage in each 4 cm wide ring for sensors from FBK (left) and HPK and CNM (right).

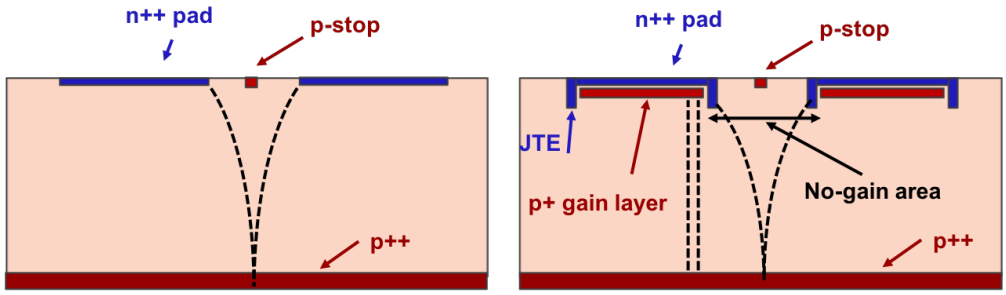
Table 3.2: Total power from bias current generated in all ETL sensors at $-20\,^{\circ}C$.

Foundries	25% lifetime	50% lifetime	100% lifetime
FBK	150 W	350 W	900 W
CNM, HPK	200 W	550 W	1450 W

3.2.4 Performance

Particles crossing the ETL will hit at most two silicon layers. The design of the sensors requires a detailed optimization since the best time resolution is achieved for tracks with two hits, and when each of the two hits is of good quality (high gain, low noise, and uniform response). To achieve this goal, the following parameters have been studied and optimized:

- **Fill factor**: the fill factor is the ratio of the active and the total sensor area which can be less than 100% due to no-gain areas and the edge width. A higher fill factor increases the number of two-hit tracks.
- **Wafer uniformity, quality of multi-pad sensors, and single pad leakage current**: it is important that the production is very uniform and all sensors have the same gain, and therefore equal time resolution. The ETL production yield and single pad quality have been evaluated.
- **Hit efficiency and signal uniformity**: silicon sensors have close to 100% efficiency. In detectors with internal gain, the efficiency remains high only if the gain is uniform within the pad and the whole sensor, so all hits are above threshold.



a) Traditional Silicon detector **b) UFSD**

Figure 3.14: Schematic representation of the charge carrier drift lines for a standard silicon sensor (left) and for an UFSD (right).

- **Gain and Noise**: an excellent time resolution can only be achieved by the electronics if the sensors provide large signals, requiring high gain and low noise.
- **Long term stability**: the stability may be affected by annealing effects. It is important to assure that the UFSD design is not sensitive to this effect.
- **Failure modes**: the high bias voltage needed to recover the gain might lead to detector damage. Stringent operational procedures need to be followed to avoid damage.
- **Time resolution**: the time resolution achieved with custom, low-noise front-end boards provides a measure of the sensors' capabilities in near ideal conditions.

In the past few years several UFSD productions have been manufactured by CNM, FBK and HPK to optimize these aspects, in the following a brief summary is presented.

3.2.4.1 Fill Factor

The main component determining the fill factor is the no-gain space between pads. As shown in Fig. 3.14 (left), in standard planar silicon detectors the fill factor is almost 100% since the electric field lines are always closing on the electrodes. This remains true in the UFSD design, Fig. 3.14 (right), but not all electric field lines cross the gain layer; in the volume where this happens, the signal is still present but without gain. This effect is caused by the junction termination extension (JTE), which is an n-doped deep well. CNM, FBK and HPK have produced structures with different interpad distances and widths of the isolation p-stop implants, aimed at maximizing the fill factor while keeping good isolation between pads.

The no-gain separation has been measured in beam tests at FNAL [82] and in laboratory bench tests [83]. In the laboratory tests, specially designed UFSD structures having a small region without metal traces from one pad to the neighboring one were used in conjunction with a laser set-up mounted on a precision $x - y$ table. In a simplified view, the amplitude of the signals measured on a pad as a function of the laser spot position is a sigmoidal function obtained by the convolution of a step function (gain – no gain) with a Gaussian function from the laser beam spot size. The distance between the 50% amplitude points is the length of the no-gain region. The left side of Fig. 3.15 shows this measurement for an FBK design, where the no-gain distance was measured to be 38.3 μm.

The right side of Fig. 3.15 reports the fill factor as a function of the no-gain distance for a 1.3×1.3 mm^2 pad. Table 3.3 shows the current results for pads that are working well. FBK has tried smaller distances (below 30 μm), however the sensors suffered from early breakdown. A

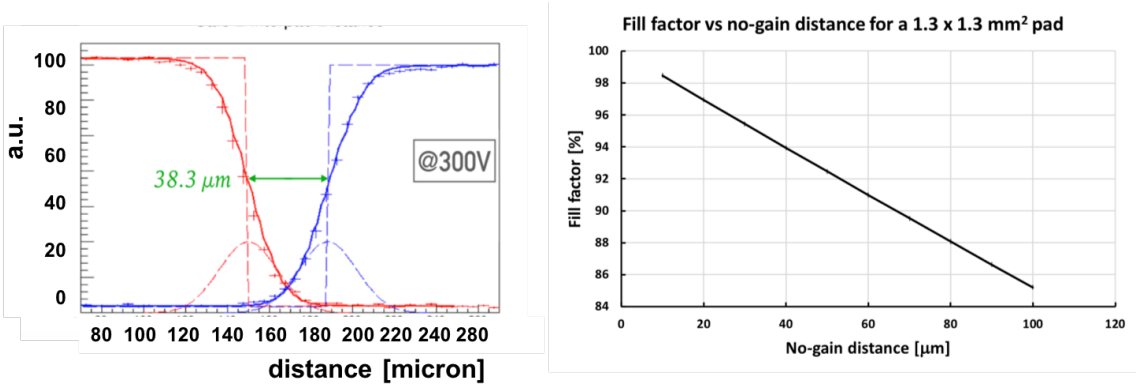


Figure 3.15: Left: Amplitude of the signals recorded on two adjacent pads while moving a small laser spot from one pad to the next one. The distance computed at the 50% amplitude point measures the no-gain distance. Right: The value of fill factor for a 1.3×1.3 mm^2 pad as a function of the no-gain distance.

Table 3.3: Summary of no-gain distances achieved for sensors from CNM, FBK, and HPK.

Foundries	No-gain distance [μm]	Comments
CNM	100	The latest production with smaller distances has very high leakage current and cannot be used. A new production is expected in August 2019
FBK	40, 70	In the latest production much smaller distances were attempted but the sensors go into early breakdown. A dedicated new production is expected in April 2019.
HPK	75, 90, 135	Even the shortest separation works well, most likely HPK can obtain even smaller distances.

dedicated production by FBK focused on interpad design is planned in the second half of 2019. From these measurements and our present understanding of UFSD designs, a no-gain distance of about 50 μm is an achievable target, yielding a fill factor of about 92%.

The second important contributor to the fill factor is the space from the last pad to the physical edge of the sensor. Since two sensors will be placed side by side on a module, this contributes to the inactive area in a module. Currently produced designs have distances between 300 and 500 μm. Assuming a 500 μm distance, this corresponds to a 2.5% loss of coverage.

The sum of the no-gain area and the edge area combine to a total fill factor of about 90%.

3.2.4.2 Wafer uniformity and quality of multi-pad sensors

In a single 6" wafer there will be 12 sensors, 21.2×42 mm^2, each with 512 pads, as shown in Fig. 3.6. In the last round of sensor productions, each of the foundries has produced several multipad sensors to test production uniformity. The layout of the structures produced for these tests are shown in Fig. 3.16, with 1.3×2.5 cm^2 sensors, each with 96 pads, and 0.65×0.65 cm^2 sensors, each with 25 pads. These sensors represent an important R&D step toward the up-

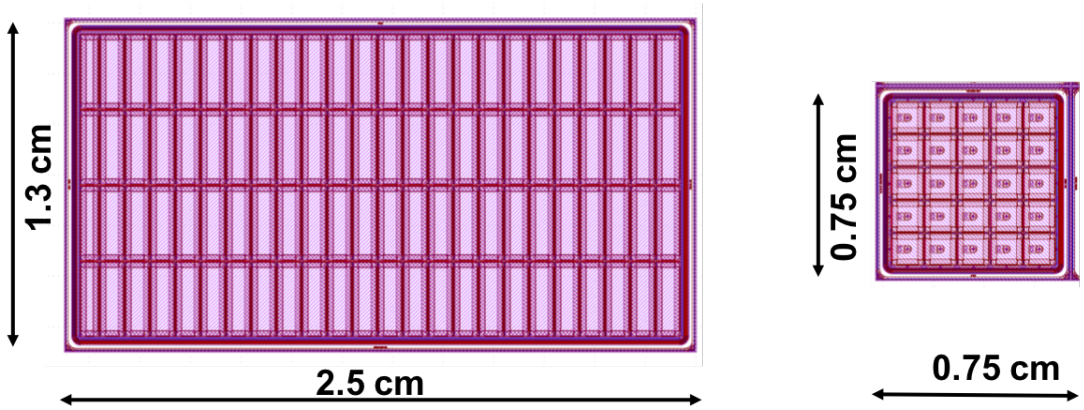


Figure 3.16: Layout of the sensors for the test of production uniformity and yield.

Table 3.4: Summary of the uniformity studies on the latest sensor productions.

Foundries	Sensor type	# Sensors tested	# Warm pads	# Bad pads	Comments
FBK	4 × 24 pads	152	14 (0.1%)	0	bias = 100 V
FBK	5x5 pads	23	4 (0.7%)	0	bias = 300 V
HPK	4 × 24 pads	15	20 (1.3%)	0	bias = 250 V

coming production of full size ETL sensors and have allowed the evaluation of the uniformity of large numbers of identical sensors on multiple wafers.

The test of production uniformity measures the current in each pad using a probe station instrumented with a multi-needle probe card to contact many pads at once, connected via a switching matrix to a computer controlled IV measurement setup. Table 3.4 reports the results; unfortunately the CNM production of large sensors suffered a delay and was not tested. The uniformity of the leakage current is a direct measurement of the gain uniformity and of the wafer quality. The sensor is required to have a leakage current smaller than 2 μA/mm^2, and the pad classification is the following:

- **good:** leakage current within 10 times the mode current,
- **warm:** leakage current above 10 times the mode current, and
- **bad:** the pad does not hold the bias voltage and causes the sensor to fail.

Figure 3.17 reports the leakage current measurement for a HPK 4×24 pad sensor: two pads fall into the *warm* category, however they are functional and not problematic for the sensor operation. Of the 16 HPK sensors tested, one sensor has a bad pad and it cannot be biased above 30 V, but the sensor appears to be stained near that pad so it is not considered further. A complete report on the FBK uniformity testing campaign performed on 20 wafers was reported in Ref. [83].

These results demonstrate that FBK and HPK can reliably produce large multi-pad sensors with very good overall uniformity. Two additional important aspects have been observed during these tests. If not all pads in a sensor are connected to ground, the sensor breakdown voltage decreases. For FBK sensors the decrease is significant, from 250 to 120 V. For HPK sensors one production shows a decrease of about 80 V, while a second set of wafers shows almost no change, ΔV= 10 V. If a bad pad is left floating, the sensor is still functional. Both these aspects will be addressed in the future productions and will influence the ETL sensor requirements.

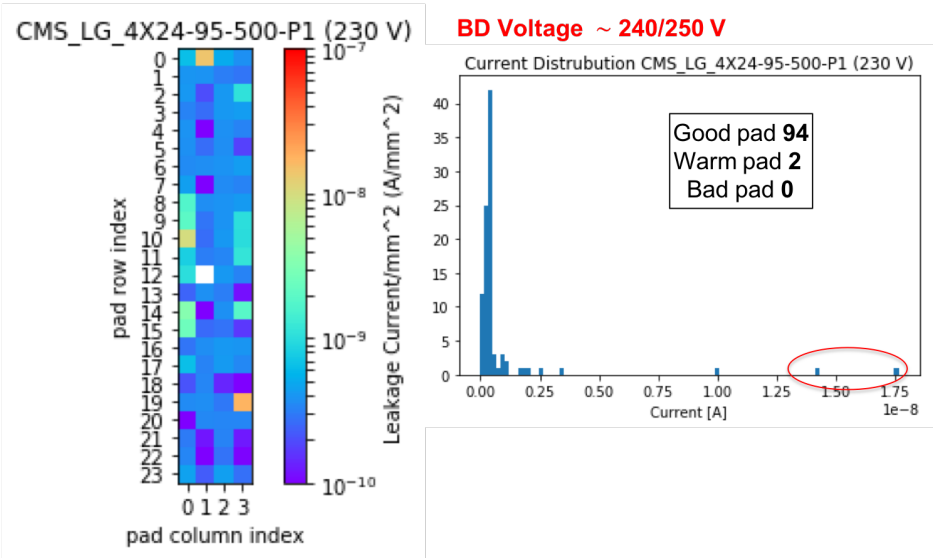


Figure 3.17: Map of bias current in a HPK 4×24 pad sensor running just below the breakdown voltage. The leakage current is low in the entire sensor, while two pads show a larger but still acceptable current. One pad (white) has a current below the detection threshold of the set-up.

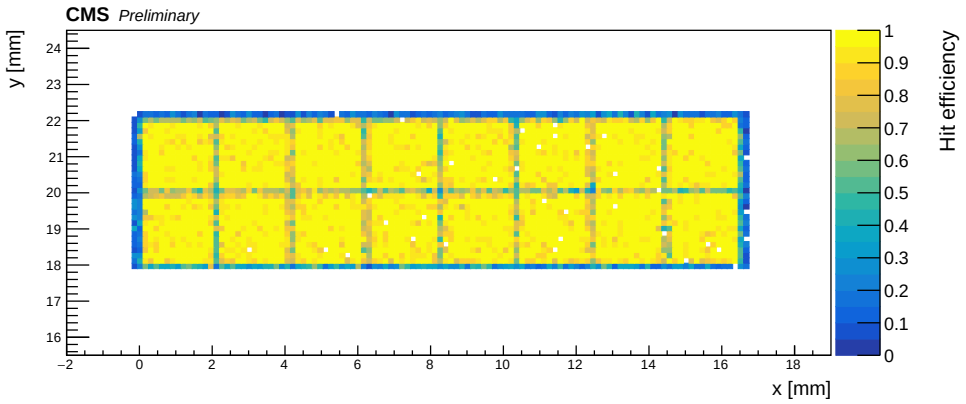


Figure 3.18: Efficiency of a 16-pad FBK sensor measured in an FNAL beam test with 120 GeV protons.

3.2.4.3 Hit efficiency and signal uniformity

The single hit efficiency and the amplitude uniformity of single and multi-pad 50 μm thick UFSD sensors have been measured during the 2017 and 2018 beam tests at FNAL using a 120 GeV proton beam [82]. The uniformity of the CNM and HPK sensor response in pulse height before irradiation was found to have a 2% spread and the efficiency was found to be 100%. Uniform signal detection efficiency of 100% was also observed on all sensors after a fluence of 6×10^{14} n_{eq}/cm^2. One interesting effect that has been measured in these beam tests is a higher signal amplitude in regions where the pad is covered with metal compared to where the pad is not covered with metal. This effect has been seen in sensors from all manufacturers, and its reason is not yet understood. To avoid this problem, the pads will be fully covered by metal in all future productions. The efficiency of multi-pad sensors has also been measured in the FNAL beam test, as shown in Fig. 3.18, finding almost 100% efficiency throughout the full sensor.

3.2.4.4 Gain and noise

The key elements to achieve a good time resolution are signals with at least 5 fC of charge (gain = 10) and low noise. For this reason, maintaining a gain value above 10 until the end of the HL-LHC lifetime is considered a minimum requirement for the ETL sensors. Several irradiation campaigns with neutrons and charged hadrons have been performed, and the evolution of the gain value has been studied as a function of fluence and bias voltage. Figure 3.9 shows how the gain changes with irradiation for sensors manufactured by CNM, FBK and HPK, while Fig. 3.10 shows the bias voltage required to obtain a gain of 10. These plots are examples of a much wider body of data collected on many sensors from many productions demonstrating that the current UFSD design can deliver a gain of 10 or higher for fluences above 1.5×10^{15} n_{eq}/cm^2 [71, 84, 85]. For equal fluence, the sensors from FBK achieve gain = 10 at a lower bias voltage because of the presence of carbon in the gain layer. Sensors from HPK, on the other hand, have higher breakdown voltages, well above 600 V, and can compensate for the larger loss of the gain layer by working at higher bias voltage.

A sensor with low noise is the second important factor to obtain good time resolution. Two different types of noise have been measured so far in UFSD productions: shot noise and popcorn noise. Shot noise is generated by the leakage current and thus increases with irradiation. In Ref. [86] the contribution of the shot noise has been studied in detail and it was found to have little or no impact on the time resolution provided that the sensors are kept cold (nominal operating temperature is about $-30\,^{\circ}C$) and the sensors operate in low-gain mode (gain less than 20). Popcorn noise is caused by micro-discharges between the pads and the p-stop and/or the guard ring. It can result from implant designs that yield locally large electric fields and produces signal-like spikes that limit the sensor operation to bias voltages lower than required to attain sufficient gain. The latest production from FBK has significant popcorn noise [87], partially from a very aggressive design of the interpad region and partially from an excessive doping of the p-stop. This noise was absent in the previous FBK productions. A dedicated new production implementing several alternative interpad designs and p-stop dopings is being manufactured and is scheduled for delivery in Spring 2019. The sensors from HPK and CNM tested so far do not present significant popcorn noise at voltages below the onset of breakdown. However, it has been observed in a few sensors from one production run; a clear explanation of these rare occurrences has not yet been found, leading to explaining the events as the result of a not carefully controlled set-up, such as high humidity or poor handling.

Overall, the UFSD productions tested for the ETL R&D have demonstrated that the combined requirements of gain > 10 and low noise can be achieved.

3.2.4.5 Long term stability

Long term annealing effects on 50 μm thick UFSDs have been reported in Ref. [88]. Two main points were addressed: (1) recovering or worsening acceptor removal and its consequence on the gain value and (2) increase of the sensor temperature due to unexpected situations leading to changes in the doping profile. The study found that UFSD sensors are affected by the well-known changes also present in standard silicon sensors such as decrease of leakage current and mild variation of the bulk doping but that the gain layer doping profile was not significantly altered. An increase of popcorn noise was detected, but not correlated to a specific situation; this effect is under further study. Overall, long term annealing effects do not seem to present a problem for the use of UFSDs in CMS.

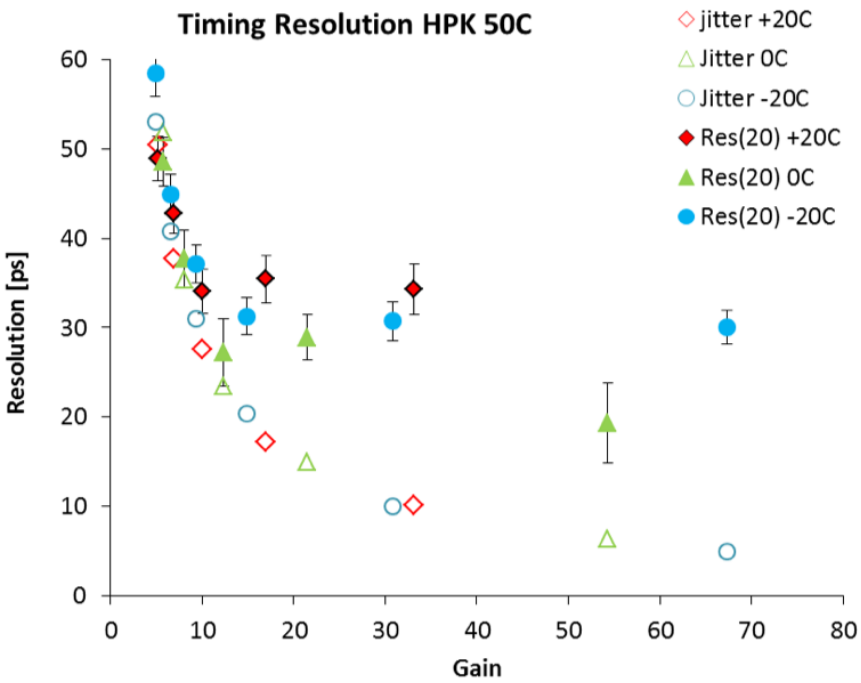


Figure 3.19: Variation of the jitter and total time resolution as a function of the UFSD gain, as measured on a Hamamatsu 50 μm thick UFSD sensor [65]. The jitter term decreases with gain, while the total time resolution flattens around $\sigma_t = 30$ ps.

3.2.4.6 Time resolution

The time resolutions achieved in the development phase of ETL are due to a combination of good sensor performance and very good read-out electronics. The front-end electronics used in laboratory studies and beam tests is custom made, using a very low-noise design based on a Si-Ge frontend, and is operated without a power limitation. Consequently, these results are considered a measurement of the intrinsic time resolution of the sensors. A description of the front-end electronics is provided in Refs. [66, 89]. As explained in Eq. 3.1, non-uniform charge deposition determines the intrinsic time resolution; this limit is a function of the sensor thickness and is about $\sigma_t \approx 25$ ps for 50 μm thick sensors [65]. The results shown in Fig. 3.19 [70] illustrate the behaviour of the UFSD time resolution as a function of gain. The total time resolution decreases with increasing gain and then saturates when the jitter component becomes smaller than the intrinsic time resolution. The plot also shows that the time resolution at different temperatures depends only on the gain value; for equal gain the same time resolution is achieved, regardless of the temperature. The gain increases at low temperature as the electron mean free path increases, roughly doubling between $+20\,^{\circ}$C and $-20\,^{\circ}$C.

The evolution of the time resolution as a function of neutron and proton irradiation for HPK and FBK sensors has been examined in Refs. [71, 84, 85]. These studies are consistent in demonstrating that UFSDs maintain a time resolution below 40 ps up to fluences of 1.5×10^{15} n_{eq}/cm^2. Figure 3.20 shows in the right panel a summary of the time resolution studies on the second FBK production (UFSD2) [90], while the left panel shows a summary for the HPK production. The time resolution for both families of sensors is measured to remain in the interval $\sigma_t =$ 30–40 ps throughout the HL-LHC lifetime, and to moderately degrade to $\sigma_t =$ 40–50 ps at a fluence of 3×10^{15} n_{eq}/cm^2, which is beyond the maximum fluence, with safety factor, shown in Table 1.3.

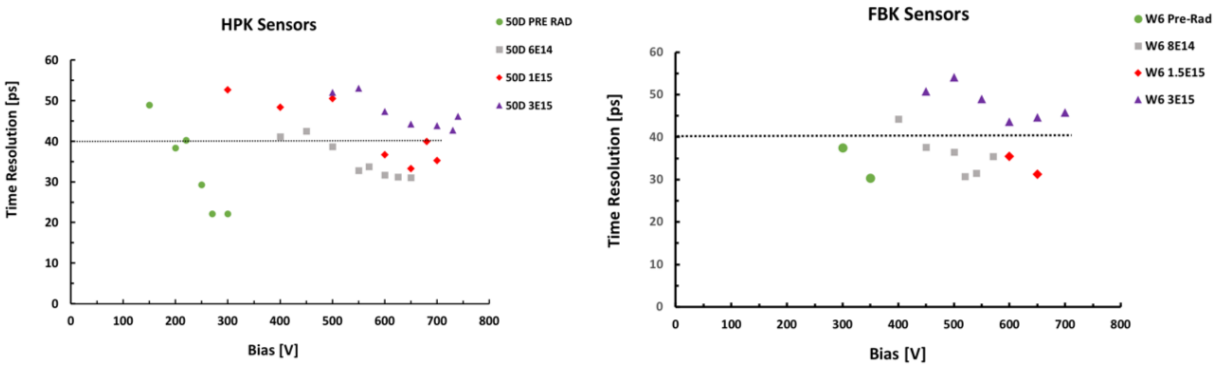


Figure 3.20: Time resolution as a function of bias voltage at different fluences. The left plot shows the performance of HPK sensors, while the right plot shows the performance of the FBK sensors.

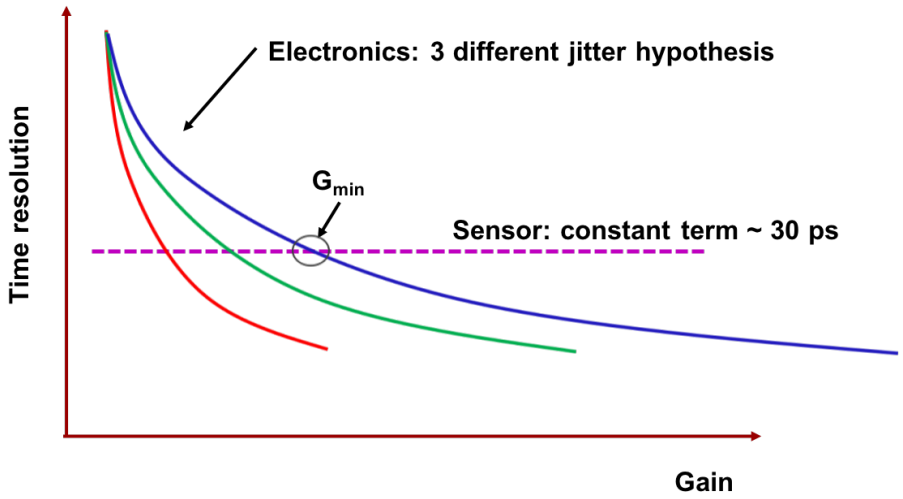


Figure 3.21: Illustration of the time resolution dependence on gain, where the contributions from the jitter and the sensor constant term are combined.

The fluence dependence of the time resolution arises from the combination of these sensor effects and the ASIC contributions. The sensor contributions satisfy the specification of < 50 ps required with two ETL layers even at the highest fluences. For the ASIC contribution, the dominant fluence dependence is from the jitter term, as described in Eq. 3.1. That jitter contribution depends on the size of the LGAD signal, which is set by the gain, and its risetime, which is set by the thickness of the depletion region. The total resolution follows the behavior sketched in Fig. 3.21 which illustrates the behavior seen in the measurements of Fig. 3.19. At lower gain values, the signal derivative is small, and the jitter term dominates, while for larger signals obtained with higher gain the time resolution is dominated by the Landau fluctuations in the sensor. For a given gain, and hence signal size, the jitter is determined by the electronic noise of the ASIC. As illustrated in Fig. 3.9, experience with irradiated prototype LGADs indicates that a gain exceeding ten, which corresponds to signal sizes of about 5 fC, can be achieved with fluences corresponding to the maximum expected in the HL-LHC. Based on this, we use that signal size and gain, $G_{\text{min}} = 10$, as the target for the optimization of performance vs power in the ASIC design, as discussed in detail in Section 3.3.

Variations in sensor gain have limited impact. Increasing the bias voltage on the sensor to obtain gain much above G_{min} does not improve the time resolution due to the dominance of the Landau fluctuation contributions. For gain $G < G_{\text{min}}$, the time resolution increases smoothly, with a small effect in the vicinity of G_{min}. For the case shown in Fig. 3.19, decreasing the gain from 10 to 7 increases the time resolution from 30 to 40 ps.

Figure 3.9 shows that irradiated prototype sensors can achieve a gain in the range $10 < G < 20$ even for fluences up to 3×10^{15} n_{eq}/cm^2, therefore providing a good operating margin through the end of the HL-LHC running. The risk of an anomalously higher fluence would affect the operating margin, but only for the innermost sensors. To assess the number of sensors impacted, we use a fluence threshold of 1×10^{15} n_{eq}/cm^2, which corresponds to doubling the fluence safety margin. Only about 10% of modules are in the highest η regions where the fluence is above that threshold. Replacement of that small number of modules would compensate for such anomalously high fluence. While we do not expect it to be needed, we have confirmed that a plan for replacement of this number of innermost modules could be accomplished in a year-end technical stop in case of such anomalies.

3.2.5 Development plan and schedule

The LGAD development plan foresees several prototyping steps, where each step will rely on the demonstrated aspects of the previous steps and introduce minimal variations to the design. This approach of continued prototyping is appropriate for the relatively new LGAD technology. While we are close to having the necessary performance for qualification of vendors from the previous prototypes, potentially significant improvements are still possible. The planned prototyping rounds, described below, are matched to the schedule of the ASIC and the module prototyping which will use the prototype LGADs. The prototyping schedule completes in advance of the Engineering Design Review and production start in 2022.

The first CMS dedicated UFSD prototypes were delivered by HPK and FBK in Winter 2018 (the CNM delivery is expected in Fall 2019). This effort has been a collaboration between CMS and ATLAS, with each experiment contributing to the strategic planning, layout design, and funding of the wafers. The layout of these prototype wafers was designed to experiment with variations in geometry, doping concentrations, and radiation hardness options. The study of these prototypes has provided an understanding of how to optimize the radiation hardness and other sensor performance aspects. As discussed above, these prototype sensors show good time resolution, radiation hardness, and fill factor.

Building on the results of the first round, a second round of prototypes will be commissioned to the three vendors in late 2019 for delivery in Q1 2020. That prototype round will focus on a layout with demonstrated design aspects, such as the no-gain distance and edge width. It will have a geometry designed for use with the similarly scheduled ETROC1 prototype ASIC to support the module prototyping plans. In addition, the prototype sensors will be used to study the uniformity of quality, robustness of performance, and radiation tolerance of the sensors.

The final prototyping round, planned for delivery early in 2021, will be matched to the second round of ETROC and module prototypes that will use it. The layout will include the final geometry and design details to provide vendor qualification. This prototype round will provide the basis for the completion of the two following milestones, shown along with all project milestones in Fig. 6.2:

- **E.Si.5, Jun 2021:** Sensor vendor qualification and final geometry selection,
- **E.Si.6, Jan 2022:** Sensor vendor selection and ready for production.

An important aspect of the LGAD development plan regards quality assurance and quality control. The collection of completed and ongoing LGAD characterization studies described above has not only demonstrated the viability of using these sensors for the CMS ETL but is also contributing to the ultimate quality assurance of production sensors. Working closely with the three vendors during the sensor prototyping phase was important in developing a list of sensor specifications and evaluating the foundries' capability for successfully and reliably meeting these specifications.

These prototyping cycles are also a crucial component in the development and commissioning of LGAD quality control procedures and the infrastructure to be used for the full sensor production campaign. Dedicated QA/QC facilities will perform spot testing of at least 5% of all delivered LGAD production sensors, which will include several sensors from each production batch, including visual inspections, bench characterization tests, and database cataloging of each sensor batch before bump-bonding to ETL readout ASICs. Modules built with the prototype and pre-production sensors will be characterized in test beams before and after irradiation; similar tests will be repeated with a small number of the first sensors from each production batch. This process will ensure that any potential sensor quality or yield issues will be resolved promptly through vendor interaction and that sensors being used in module assembly meet the predetermined design specifications.

3.3 On-detector electronics

3.3.1 ETL readout ASIC overview

The ETL readout ASIC chip (ETROC) is designed to handle a 16×16 pixel cell matrix, each pixel cell being 1.3×1.3 mm^2 to match with the LGAD sensor pixel size. The size of the sensor will be larger than the size of the readout chip and two chips will be bump-bonded to one sensor. The choice of 16 × 16 channels is motivated by the clock (timing reference) distribution, which will be laid out as an H-tree. The pixel cell size is a compromise between smaller capacitance (3.4 pF), which translates to higher S/N in the front-end, and the overall channel count, which affects the total power consumption. The exact pixel cell size will be optimized once the ASIC performance as a function of input capacitance and current is better known from the prototype chips.

Figure 3.22 shows the general block diagram of the ASIC. At the cell level, each channel consists of a preamplifier, a discriminator, a TDC used to digitize the TOA (time of arrival) and TOT (time over threshold) measurements, and a memory for data storage and readout. The TOT is used for time-walk correction of the TOA measurement. The detailed hit information (TOA and TOT) from within each cell will be read out from a local circular buffer after each Level-1 Accept (about 1 MHz). In addition, a charge injection circuit is implemented to allow for testing and calibration. The injected charge can be programmed by a DAC. For more detailed monitoring of the signal pulses as radiation dose increases, waveform sampling circuits will be included in selected pixel cells. Additional peripheral circuits include a PLL, a phase shifter, an I2C slave, a fast control block, a serializer, and a data driver.

3.3.1.1 Design requirements

The design goal for the time resolution is 50 ps per hit, to achieve a 35 ps arrival time measurement for a MIP track with an ETL hit in each of the two layers. There are several contributions to the total time resolution:

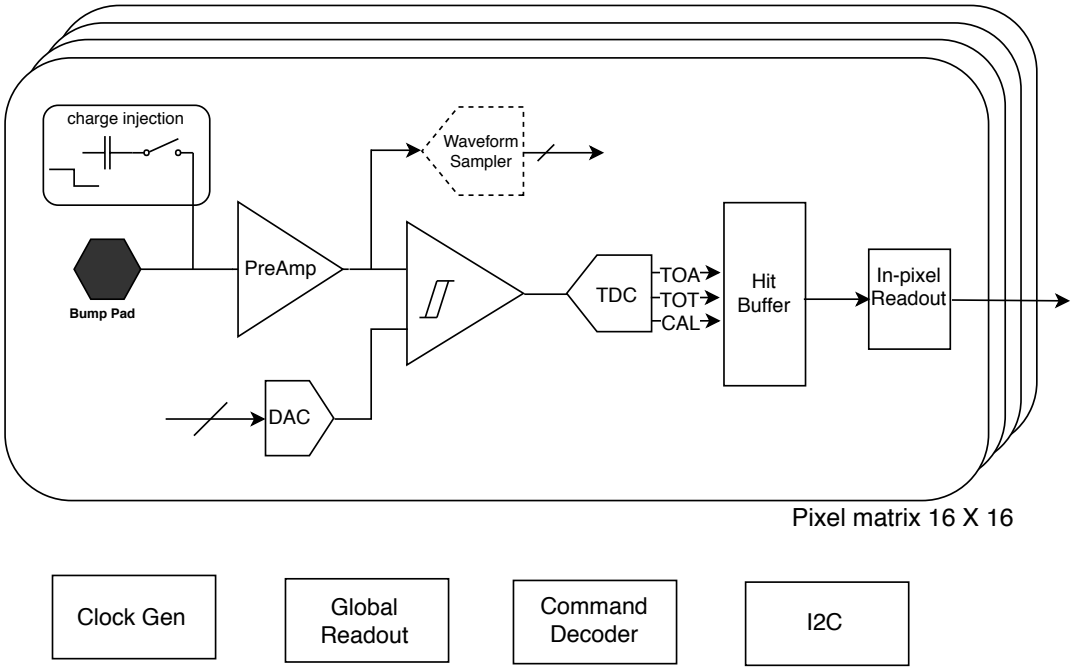


Figure 3.22: A schematic of the ETL ASIC, showing the blocks within each channel of the 16×16 pixel matrix and also the peripheral blocks at the bottom.

1. the intrinsic LGAD contribution due to Landau fluctuation of the charge deposition of a MIP;
2. the jitter due to the preamplifier and discriminator stage;
3. the TDC time quantization bin size and clock distribution within chip;
4. the residual of the time-walk correction;
5. system level clock distribution.

Among them, items (2), (3), and (4) are the contributions from the ASIC. The LGAD contribution is known to be about 30 ps, while the system level clock distribution will be designed to be precise to better than 15 ps. Within the ASIC, the TDC measurement is expected to contribute less than 10 ps, as is the residual contribution from the time-walk after correction. This means that the jitter from the preamplifier/discriminator has to be kept below 40 ps. This should be achieved with reasonable power consumption and signal efficiency, and maintained even after irradiation.

A summary of the ETROC requirements is shown in Table 3.5.

3.3.1.2 Methodology to approach the design challenges

The challenge of designing the ASIC, while optimizing its performance and that of the sensor together as a unit, is approached in several design stages. The first stage has two components: 1) using the test beam data taken on LGAD sensors with an external preamplifier read out by an oscilloscope to study different timing measurement algorithms; and 2) using detailed LGAD simulation as input to simulate the behavior of various ASIC front-end designs. Based on these results, the second stage is to design and develop a small scale prototype chip and, again using

Table 3.5: A summary of ETROC requirements.

Requirement	Value	Comments
Process	TSMC 65 nm MS RF LP 2.5 V with metal stack 1P9M_6X1Z1U_RDL (CERN)	
Power supply	1.2 V	
Timing resolution	40 ps	Total timing resolution per hit is 50 ps, including 30 ps contribution from sensor.
Pixel size	1.3×1.3 mm^2	
Pixel capacitance	3.4 pF	50 μm thickness
Pixel matrix size row$\times$column	16×16	
Power consumption	below 1 W/chip	
Data storage capability	12.8 μs	Level-1 trigger latency
Trigger rate	Up to 1 MHz	
Operation temperature	$-30\,^{\circ}$C to $+20\,^{\circ}$C	
TID	100 Mrad	
SEU	TBD	system requirements

the LGAD simulation for input, to study the design performance with post-layout simulation. Subsequent stages involve the production of prototypes with increasingly more complex structure to test and further optimize the ASIC design. A first single-pixel prototype including a preamplifier and a discriminator has been submitted for production. A prototype with dedicated waveform sampling capability will be submitted in Summer 2019 and will allow more detailed understanding of the pulse shape after the preamplifier stage, thus enabling further optimization of the timing measurement. A detailed plan for the subsequent larger scale prototypes is presented in Section 3.3.13. The remainder of this section describes the design of each of the ASIC building blocks together with results from relevant studies.

3.3.2 Design study with waveform analysis using test beam data

Data from LGAD sensors collected at the Fermilab test beam are used to guide the design of the front-end stage of the ASIC. Much information can be extracted from the sampled waveforms of different sensors, with different irradiation levels, operated at different bias voltages. This is shortening the development cycle for the ASIC. Some questions that are being studied using the beam test waveform data are:

1. The performance of different types of timing algorithms, e.g. leading edge (LE) discrimination based on measuring TOA/TOT vs. using a constant fraction discriminator (CFD),
2. The impact of front-end amplifier bandwidth and S/N,
3. The required time quantization of the TDC,
4. The minimum sampling frequency and the minimum number of bits needed for the sampled waveforms in order to be able to efficiently update the thresholds and the correction algorithms as the sensors age and receive increasing radiation doses.

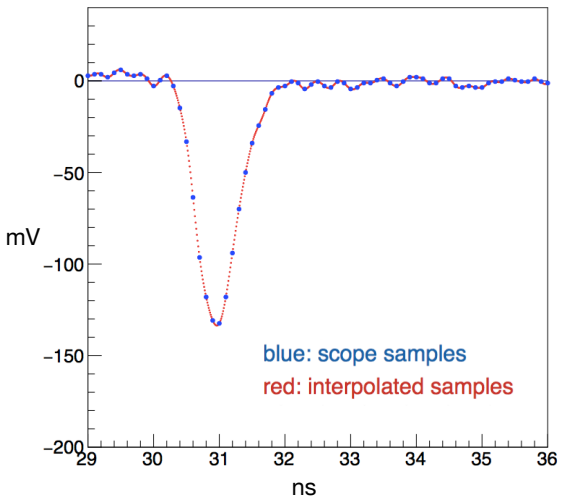


Figure 3.23: One example of a sampled waveform. The red curve shows the result of the interpolation performed using the Shannon-Nyquist formula. The interpolated signal is used to simulate the discriminator response. The simulation is ideal because it returns the exact time the interpolated signal crosses the threshold.

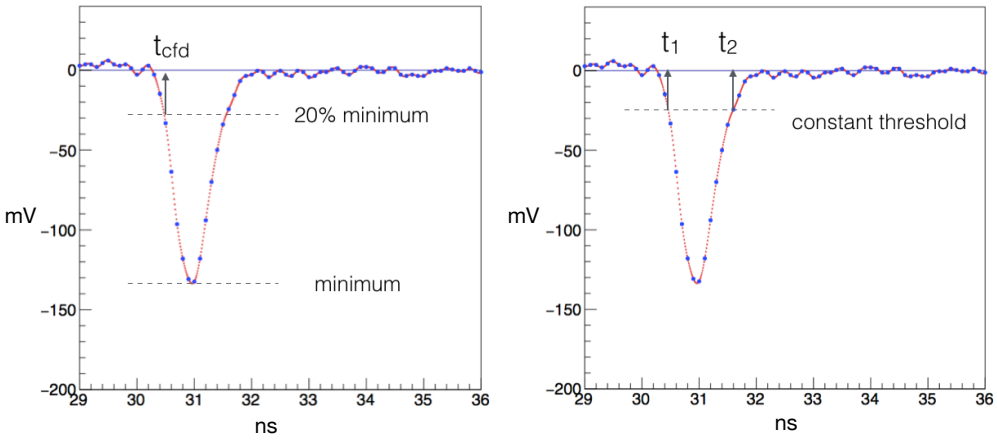


Figure 3.24: The picture on the left is an illustration of the definition of an ideal CFD. The picture on the right shows the measurement of the leading edge time of the pulse (t_1) followed by a second time measurement (t_2) on the trailing edge. Both times are measured with a fixed threshold and can be combined to correct for time-walk as explained in the text.

The signals from the sensors were fed to an external amplifier and then to a digital scope where they were digitized by an eight bit ADC and sampled at 10 GS/s. Figure 3.23 shows a typical waveform recorded by this setup.

A simulation of the discriminator response is used to set a lower limit for the time resolution obtainable with a given sensor using a CFD or a leading edge discriminator with a correction of the time-walk based on the measurement of time over threshold.

Establishing a lower limit for the achievable time resolution gives an upper limit for the requirements of a number of crucial front-end chip specifications. These include preamplifier bandwidth and signal/noise, time quantization of the TDC, number of bits and sampling frequency of a possible sampling ADC.

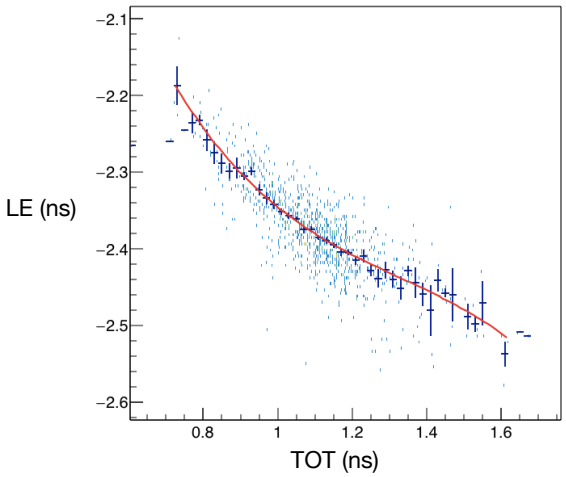


Figure 3.25: The correlation between the LE and TOT times.

Table 3.6: Impact of sensor irradiation on timing resolution of the CFD and LE+TOT methods, for FBK W6 LGAD sensors.

Sensor irradiation	CFD	LE + TOT
FBK W6, before irradiation	28 ps	30 ps
FBK W6, 8.0×10^{14} n_{eq}/cm^2	30 ps	32 ps
FBK W6, 1.5×10^{15} n_{eq}/cm^2	42 ps	40 ps

Figure 3.24 (left) shows the definitions of terms that characterize an ideal CFD. For each individual pulse, the threshold is set at a constant fraction of the peak value of that pulse. The value of the fraction is optimized for each dataset to get the best time resolution and is typically around 20%. The time resolution obtained in this way is considered as the best possibly achievable for each data set and it is the benchmark against which other methods are compared. The baseline solution for the time measurement at this point is the measurement of the leading edge of the pulse with a simple discriminator with a fixed threshold voltage, followed by a correction of the time walk based on the measurement of the time over threshold. This implies that two time values need to be measured for each pulse as shown in Fig. 3.24 (right). The LE time is t_1 and TOT is $t_2 - t_1$. Figure 3.25 shows the correlation between the LE and TOT times. The fit is a third order polynomial that can be used for the correction.

Table 3.6 shows the time resolution obtained for three different sensors with different irradiation levels, using an ideal constant fraction discriminator and leading edge plus time over threshold correction. These results suggest that:

1. using LE + TOT does not seem to be significantly worse than CFD and possibly more robust in particular for irradiated sensors, and
2. the time resolution that is realistically achievable with these sensors is limited to about 30 ps.

TDC Quantization

Adding the simulation of TDC time quantization for both the LE and TOT measurements lets us determine the coarsest quantization level that can be used without significantly affecting the

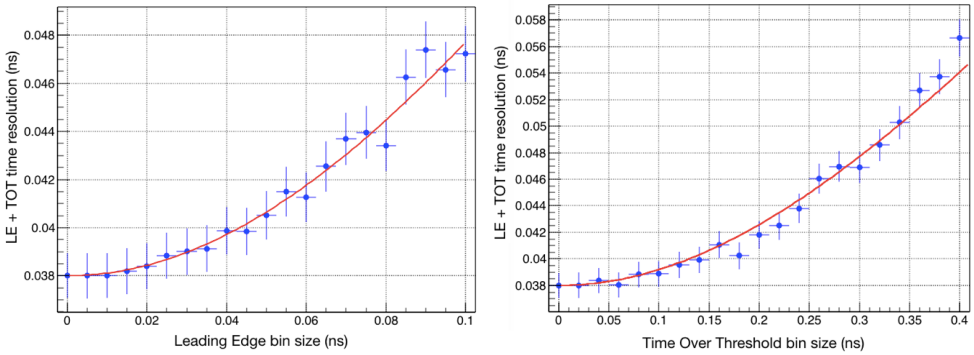


Figure 3.26: Simulated time resolution as a function of the LE measurement bin size (left), and simulated time resolution as a function of the TOT measurement bin size (right). The y axis is the uncertainty introduced by the quantization, which is equal to the bin width divided by $\sqrt{12}$.

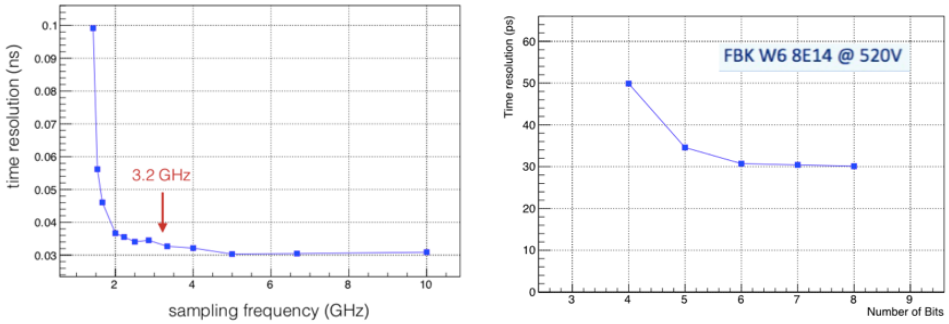


Figure 3.27: The simulated time resolution as a function of the sampling frequency (left) and of the number of bits in the ADC (right).

time resolution of the LE+TOT approach. Figure 3.26 (left) shows an example of simulated time resolution as a function of the size of the time bin in the LE measurement and Fig. 3.26 (right) shows it as a function of the size of the time bin in the TOT measurement. These simulation results agree reasonably well with a simple model where the intrinsic time measurement error is summed in quadrature with the quantization bin divided by $\sqrt{12}$ (red curve). For the TOT curve, the quantization bin is also weighted by a coefficient c that enters into the LE+TOT correction formula:

$$t = LE + c \times TOT \tag{3.3}$$

This study suggests that a quantization bin size of up to 30 ps (σ =12 ps and 100 ps for LE and TOT, respectively) will not affect the performance in a significant way.

Waveform Sampling

The analysis above is based on the waveform data recorded during the beam test using an external preamplifier and a high performance scope. It is clearly desirable to have the ability to record the waveform directly within the front-end chip. Waveform sampling can be implemented in the front-end chip for a limited number of pixels.

The main purpose would be to monitor the change in pulse shape resulting mainly from the radiation dose increasing with time, and being able to derive updated optimal thresholds and correction algorithms when needed. During the development phase, the internal waveform

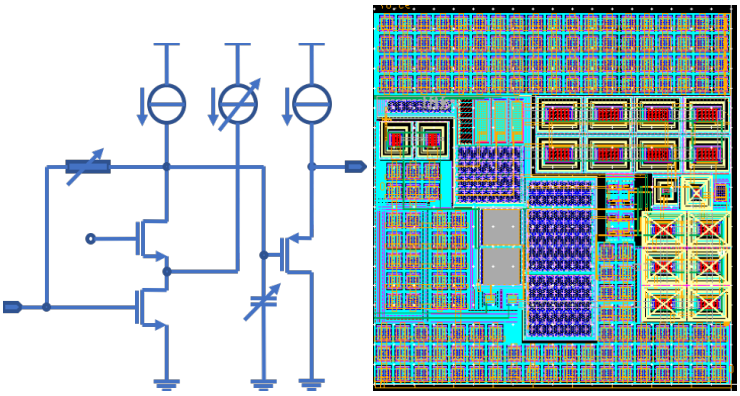


Figure 3.28: Left: A simplified schematic of the preamplifier. Right: the layout of the preamplifier in 65 nm technology with a footprint of 90×94 μm^2.

sampling capability will be useful for performance optimization studies. The required sampling rate and number of ADC bits are chosen to reconstruct the original waveform with sufficient accuracy to obtain the correct time resolution from the simulation of the amplifier-discriminator-TDC chain. This study suggests that an ADC with an effective number of bits (ENOB) of six and a sampling rate of up to about 3.2 GHz will meet the requirements. A total number of eight bits is desirable to account for a channel-to-channel variation of the average pulse height by a factor of about four (Fig. 3.27).

3.3.3 Preamplifer design

The front-end approach is based on a conservative, textbook design similar to the one used by the ATLAS ALTIROC chip [91], which is however implemented in 130 nm technology with different sets of optimizations. The preamplifier consists of a two-stage amplifier. A cascode amplifier with resistive feedback acts as the first stage, and a source follower as the second stage. Figure 3.28 (left) shows a simplified schematic. The size of each transistor in the preamplifier has been optimized by using an LGAD simulation as the input signal. The design considers both leading and trailing edge to optimize the TOT measurement for the time-walk correction. The feedback resistance is programmable in four steps between 4.4 and 20 kΩ to allow adjustment of the fall time, while the bias current is also programmable to allow different trade-offs between power consumption and performance. The load capacitance of the first stage is also programmable to allow optimization of the bandwidth.

With a smaller feedback resistance, the preamplifier will discharge the input capacitance, which is mainly contributed by the LGAD sensor, with a smaller time constant. The resulting short TOT is desirable for the time-walk correction. On the other hand, the selection of a small feedback resistance leads to a smaller output amplitude, resulting in undesirably smaller input charge when the LGAD sensor gain drops after irradiation. Based on the simulation with LGAD signals up to 1×10^{15} n_{eq}/cm^2, a default feedback resistor of 5.7 kΩ is used. The load capacitance can be selected among 0, 80, and 160 fF. With a small load capacitance, the edges are fast, but more noise is included because of the large bandwidth. Therefore, the load capacitance can be selected to fine tune jitter. The default setting is to use minimal additional load capacitance. The bias current of the input transistor can be selected from four settings. The smallest bias current is the default setting that leads to a preamplifier power consumption of 0.69 mW at $-20\,^{\circ}$C.

Figure 3.28 (right) shows the layout of the preamplifier, which in 65 nm technology has a footprint of 90×94 μm^2. Figure 3.29 is the simulated leading edge jitter versus input charge at

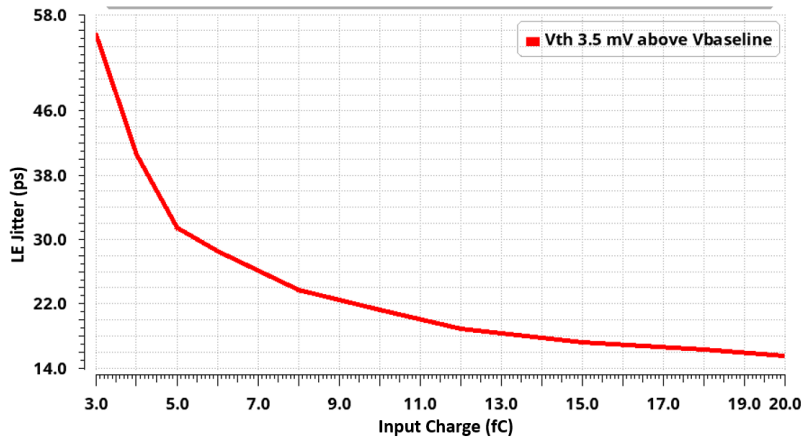


Figure 3.29: Leading edge jitter (ps) versus input charge (fC) obtained from the post-layout simulation of the preamplifier at $-20\,^{\circ}$C, assuming an ideal discriminator with a threshold voltage of 3.5 mV above the baseline voltage of the preamplifier. LGAD-like signals are used in the simulation.

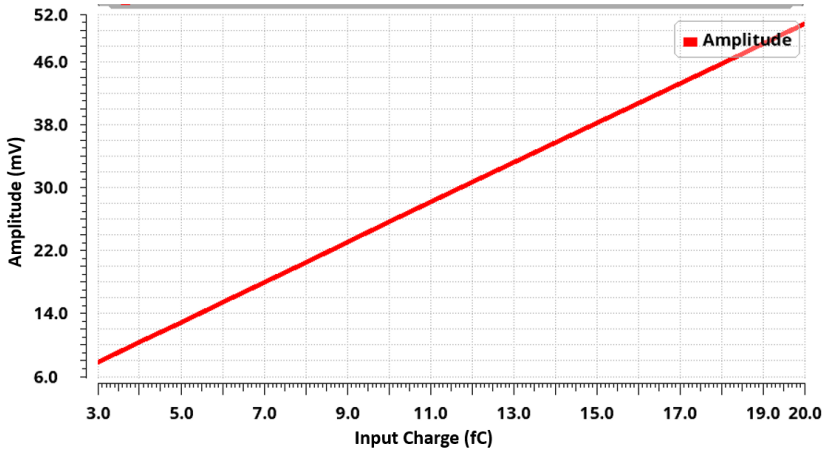


Figure 3.30: Amplitude (mV) versus input charge (fC) obtained from the post-layout simulation of the preamplifier at $-20\,^{\circ}$C. LGAD-like signals are used in the simulation.

$-20\,^{\circ}$C, assuming an ideal discriminator with a threshold voltage of 3.5 mV above the baseline voltage of the preamplifier. The pulse amplitude at the output of the preamplifier as a function of input charge is shown in Fig. 3.30. For a common MPV charge from the LGAD study, the MPV amplitude is less than 20 mV. The current consumption of the preamplifier at different bias current settings is shown in Fig. 3.31. The bias current of the input transistor can be selected from four settings, with the so called IBSelB $= 0, 1, 3, 7$, corresponding to bias current of 0.35, 0.7, 1.05, and 1.4 mA, respectively.

3.3.4 Discriminator design

The discriminator with a three-stage structure is designed to work with the preamplifier. A simplified schematic is shown in Fig. 3.32. First, three pre-amplifiers are used to amplify the analog signal from the frontend preamplifier to ease the design of the comparator downstream. Second, a conventional textbook hysteresis comparator discrminates the analog signal, with positive feedback to generate hysteresis. The third is a buffer stage to provide driving capability for the capacitive load. The threshold voltage of the discriminator is generated by an

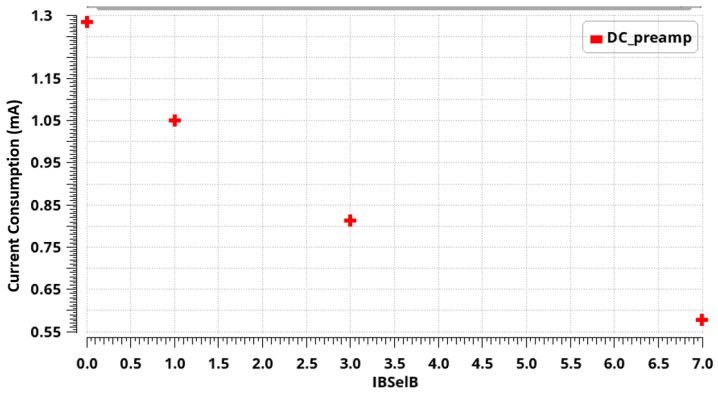


Figure 3.31: Power consumption (mA) of the preamplifier versus bias current setting. The bias current of the input transistor can be selected from four settings, with the so called IBSelB $= 0, 1, 3, 7$, corresponding to bias current of the input transistor of 0.35, 0.7, 1.05, and 1.4 mA, respectively. The 0.35 mA is the default bias current, while the power consumption of the preamplifier under default bias current is 0.69 mW at $-20\,^{\circ}$C.

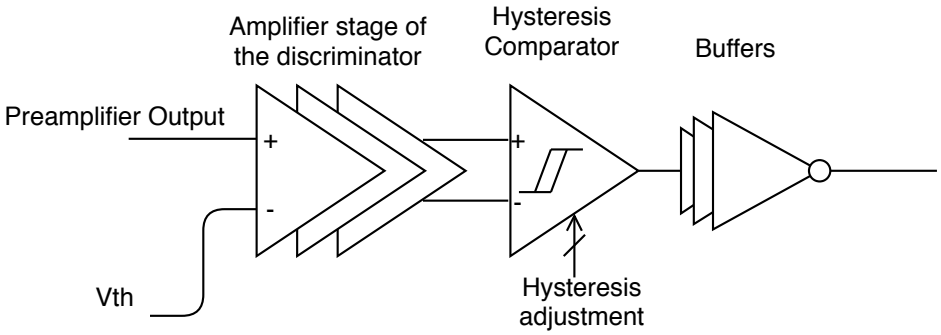


Figure 3.32: Simplified schematic of the discriminator.

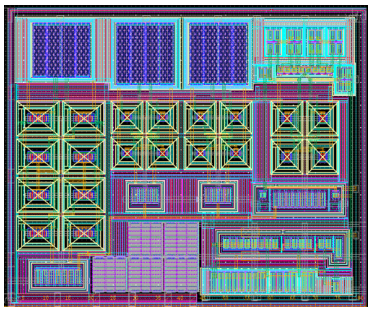


Figure 3.33: Layout of the discriminator in 65 nm technology with a footprint of $81 \times 67\ \mu\text{m}^2$.

in-pixel DAC, which covers the range from 0.6 V to 1 V and with a step size of 0.4 mV. The hysteresis voltage is programmable with 4 settings. The power consumption of the discriminator at $-20\,^{\circ}$C is 777 μW per channel for 10% occupancy and 710 μW per channel for 1% occupancy.

The layout of the discriminator takes $81 \times 67\ \mu\text{m}^2$, shown in Fig. 3.33. The preamplifier and the discriminator are simulated together with the parasitic components included. An ideal discriminator with programmable threshold voltage is employed in the simulation. Figures 3.34 and 3.35 show the leading edge jitter and standard deviation of TOT at various threshold voltages, respectively.

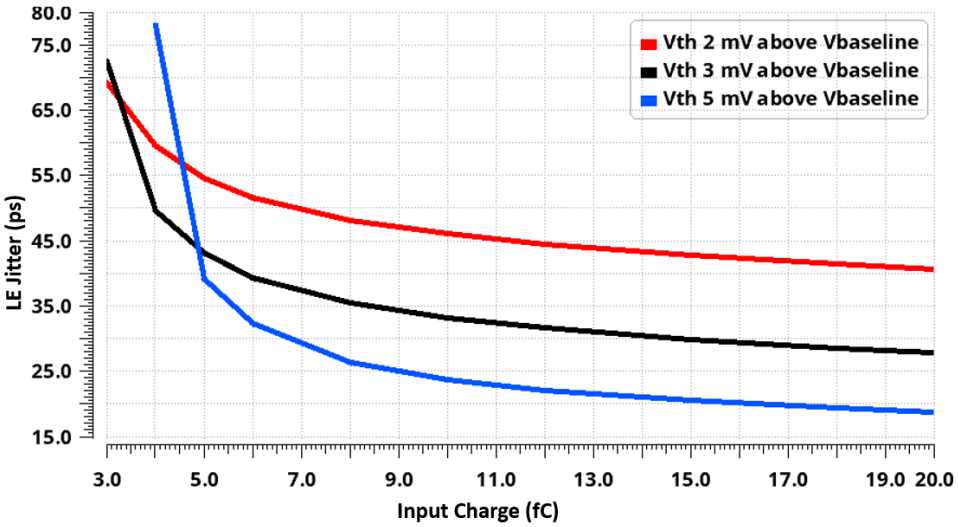


Figure 3.34: Leading edge jitter (ps) versus input charge (fC) obtained from post-layout simulation with different threshold voltages at $-20\,^{\circ}$C. Parasitic components of the preamplifier and the discriminator are included. LGAD-like signals are used in the simulation.

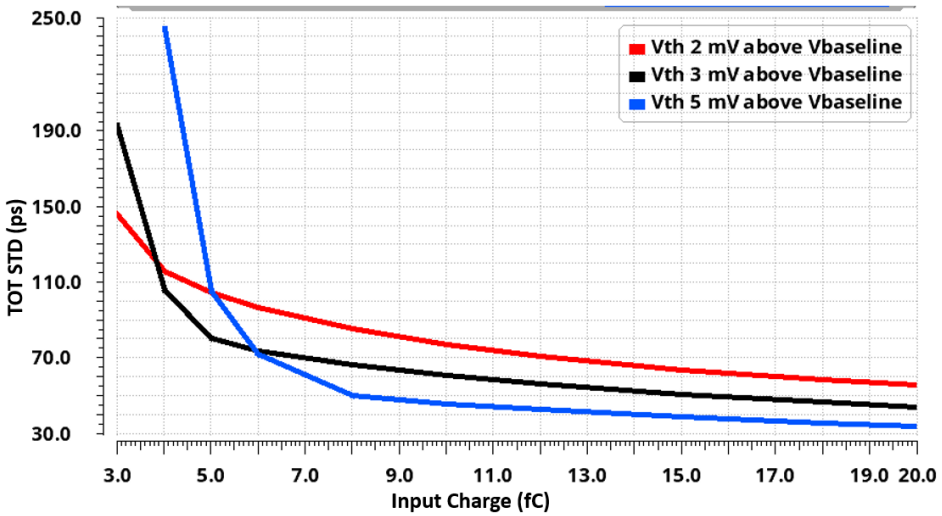


Figure 3.35: Standard deviation of TOT (ps) versus input charge (fC) obtained from post-layout simulation with different threshold voltages at $-20\,^{\circ}$C. Parasitic components of the preamplifier and the discriminator are included. LGAD-like signals are used in the simulation.

3.3.5 Results of preamplifier and discriminator performance simulation

An LGAD simulation is used in the circuit simulation to evaluate the overall performance. Three groups of data at different irradiation levels are used: pre-irradiation, 5×10^{14}, and 1×10^{15} n_{eq}/cm^2. A bin size of 20 ps is considered for both TOA and TOT measurement. The results of the default power setting and a higher bias current setting (doubled) are shown in Fig. 3.36. With the default preamplifier bias setting, a 35 ps time resolution is obtained at the target operation temperature of $-20\,^{\circ}$C when the module has been irradiated to 5×10^{14} n_{eq}/cm^2. At the high fluence, the time resolution degrades to $\approx$45 ps because of reduced sensor gain. For the pre-irradiated case, the resolution is at the 40 ps level with 300 V bias voltage. It also shows the time resolution variation over a large range of temperatures. After high fluence, the dependence on temperature becomes larger, about 0.4 ps per degree. This is because the gain

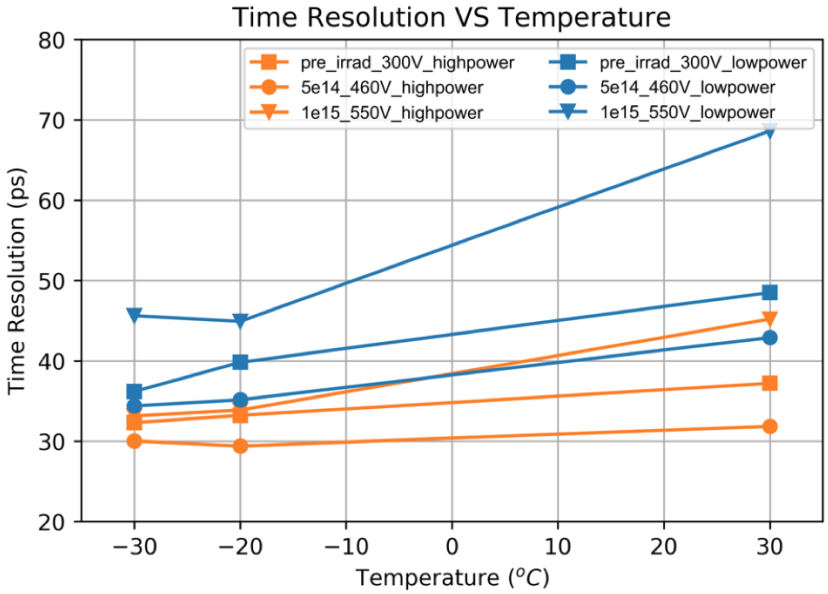


Figure 3.36: Time resolution as a function of temperature in post-layout simulation, with the nominal process, with default (IBSelB $= 7$) and also a higher power preamplifier setting (IBSelB $= 1$). The parasitic components of the preamplifier and the discriminator are included, assuming 20 ps bin sizes for both of the TDCs.

of the sensor drops after high irradiation, resulting in an overall SNR degradation. This becomes prominent at room temperature, where the gain in the preamplifier is not large enough. In order to alleviate this performance degradation, a larger feedback resistance or a larger bias current could be used.

Better time resolutions can be achieved with higher power consumption. Figure 3.36 shows the resolution with the higher power preamplifier setting (IBSelB $= 1$), thereby doubling the power consumption. At $T = -20\,^{\circ}\mathrm{C}$, at the higher power setting the time resolution is improved from the 35–45 ps range to about 30–35 ps. The overall temperature dependence has been improved a lot as well, especially for the high fluence case.

3.3.6 TDC design

The TDC takes as input the discriminator output and records the TOA and TOT for a fixed discriminator threshold. As described in Section 3.3.2, the TOA and TOT TDC bin size should not exceed 30 ps and 100 ps, respectively. To allow improvements in particle identification in heavy ion collision events, the TDC measurement time window is extended to 6.25 ns (the measurement window of the ALTIROC is 2.5 ns). To satisfy these requirements, while optimizing the design for reliability and power consumption, two approaches have been studied.

The first approach is essentially the 65 nm equivalent of the ALTIROC TDC design, which is based on the vernier delay line technique [92]. The second approach, hereafter referred to as the ETROC TDC, combines the traditional single tapped delay line TDC measurement technique with a method for in-situ self-calibration developed for FPGA TDC implementation in the past. The ETROC TDC alternative is pursued for its relative simplicity, and thus potential improvements in reliability, as well as the associated significant reduction in power consumption.

In what follows, we will describe the ETROC TDC and its implementation first. We will then

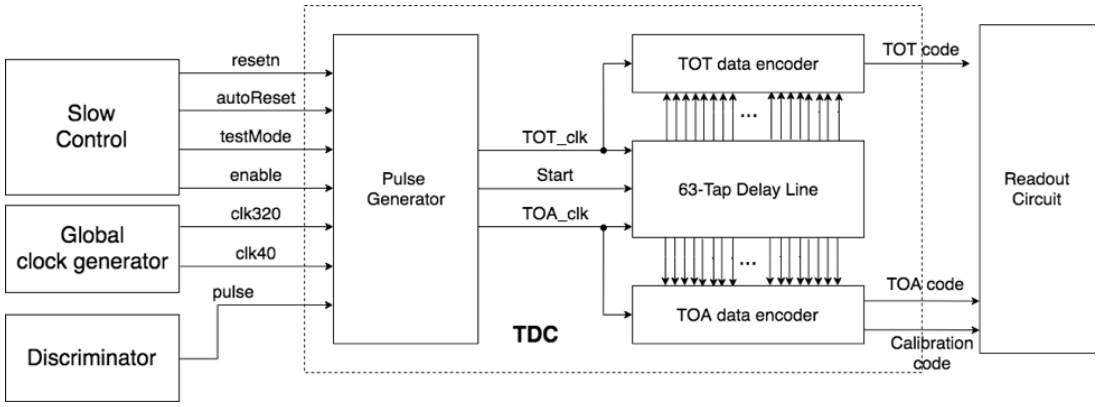


Figure 3.37: Basic structure of the TDC design based on a single multi-tapped delay line.

compare the two approaches.

3.3.6.1 Single-tapped delay line design

As detailed in the remainder of this section, the proposed ETROC TDC design leverages recent developments in FPGA TDCs and a simple delay cell calibration method to allow us to measure simultaneously the TOA and TOT using a single multi-tapped delay line and, furthermore, to eliminate the need for DLLs which are used in the traditional vernier delay line TDC design. The basic structure of the ETROC TDC is shown in Fig. 3.37.

The TDC receives the pulse from the discriminator and measures its rising edge arrival time (TOA) and the width of the pulse (TOT). The TDC includes four main components: a pulse generator, a delay line, and two data encoders. The pulse generator processes the hit pulse and input clock signals and generates a Start signal, a TOA clock and a TOT clock. The Start signal initiates the delay line and three time stamps are recorded by two clock signals in the measurement session. The data encoder circuits extract the TOA, TOT and calibration data for readout. The TDC can be configured, reset or disabled through slow control.

Pulse generator

The timing diagram of the pulse generator is shown in Fig. 3.38. The generated signals enable the circuit to perform measurements of the TOA, the TOT, and a single clock period, which is used for the delay line calibration. The input clock signal is a modulated 320 MHz clock from a global clock generator. For each 25 ns, there are only two consecutive 320 MHz clock pulses. The input clock phase is adjusted by a phase shifter so that the 6.25 ns measurement window is right before the first rising edge of the clock. The Start signal initializes the TDC delay line when the hit arrives. The TOA clock, a replica of the input clock signal generated after a hit has occurred in the measurement window, is used to record two time stamps. The first rising edge records a known clock signal edge, which reflects the TOA information. The second time stamp is used for in time calibration because the time interval between the two stamps is known to be 3.125 ns, the clock period. The TOT pulse, triggered by the falling edge of the input signal, records the time stamp at the end of the pulse to allow the width of the pulse to be calculated. If no hit occurs within the measurement window, the delay line is not started and both TOA and TOT clock signals remain disabled to conserve power.

A preliminary version of pulse generator has been implemented. The core logic and layout are shown in Figs. 3.39 and 3.40. The design is quite compact and the layout size is 45×13 μm^2.

When a hit occurs in the time window, the start signal rises and resets when the TOT and

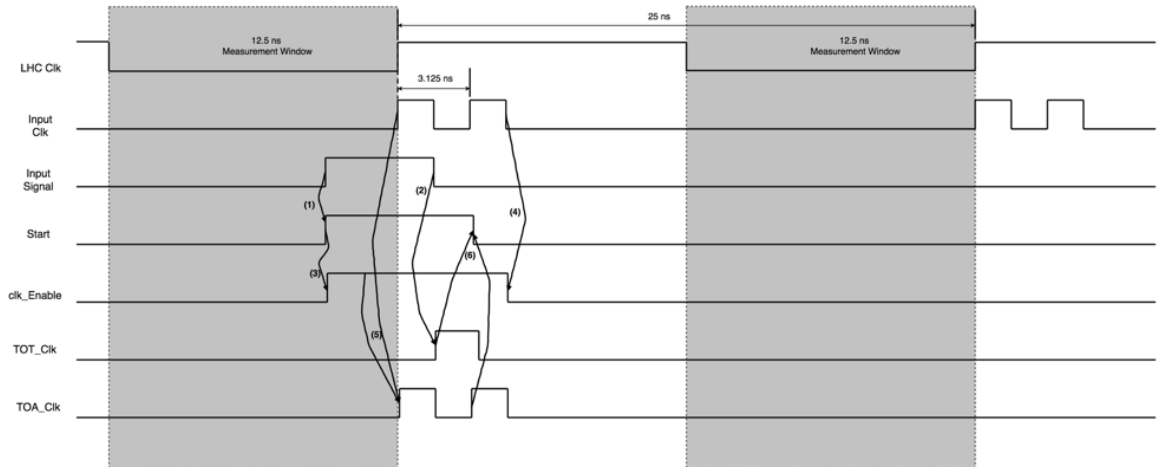


Figure 3.38: Timing diagram of the pulse generator. The rising edge of the hit signal initializes the Start signal (1) while the falling edge initializes the TOT clock signal (2). The Start signal raises the clock enable (3). The second falling edge of the input clock signal resets the clock enable (4). The TOA clock signal is the logic AND of the enable signal and the input clock signal (5). After the second rising edge of the TOA clock and the rising edge of the TOT clock, the Start signal is reset and the delay line waits for the next measurement (6).

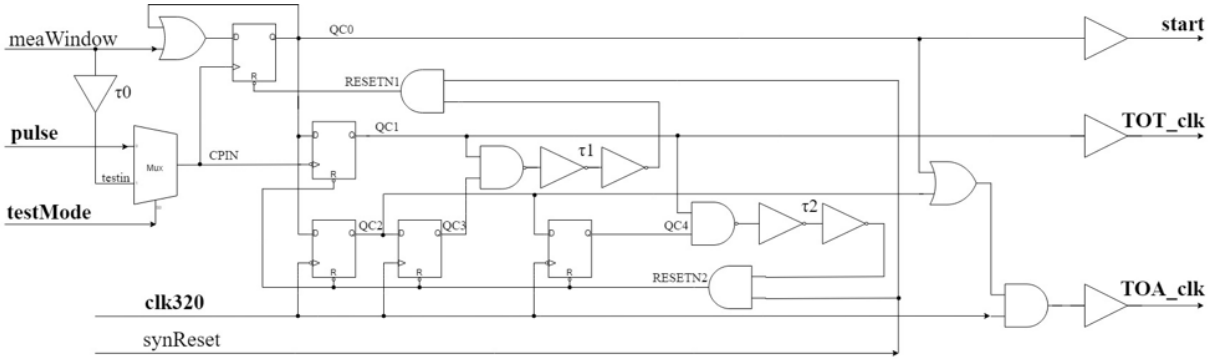


Figure 3.39: The pulse generator core logic.

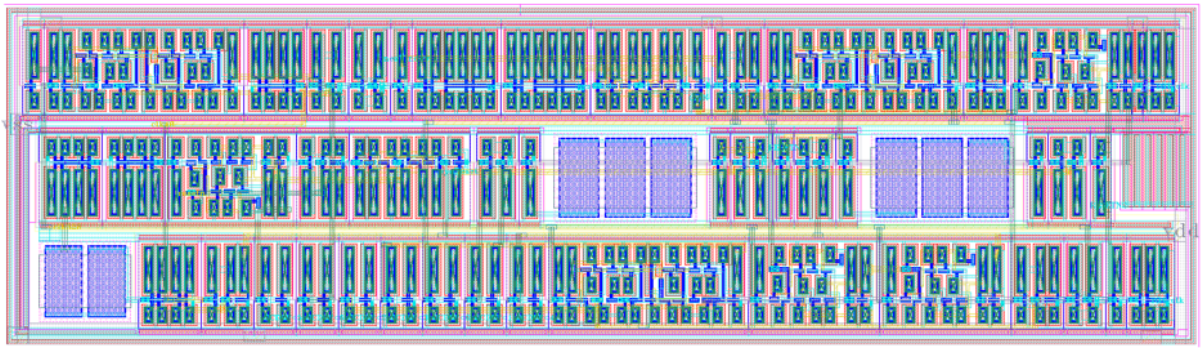


Figure 3.40: The pulse generator layout.

TOA clock rising edges latch the data. This design minimizes the delay line oscillation time. In a typical measurement, the delay line oscillation time is less than a third of the 25 ns clock period, which further reduces the power needed by the TDC. The pulse generator, when set in test mode, will send out a 12.5 ns width test pulse.

The post-layout simulation of the typical waveform is shown in Fig. 3.41. When there is a hit,

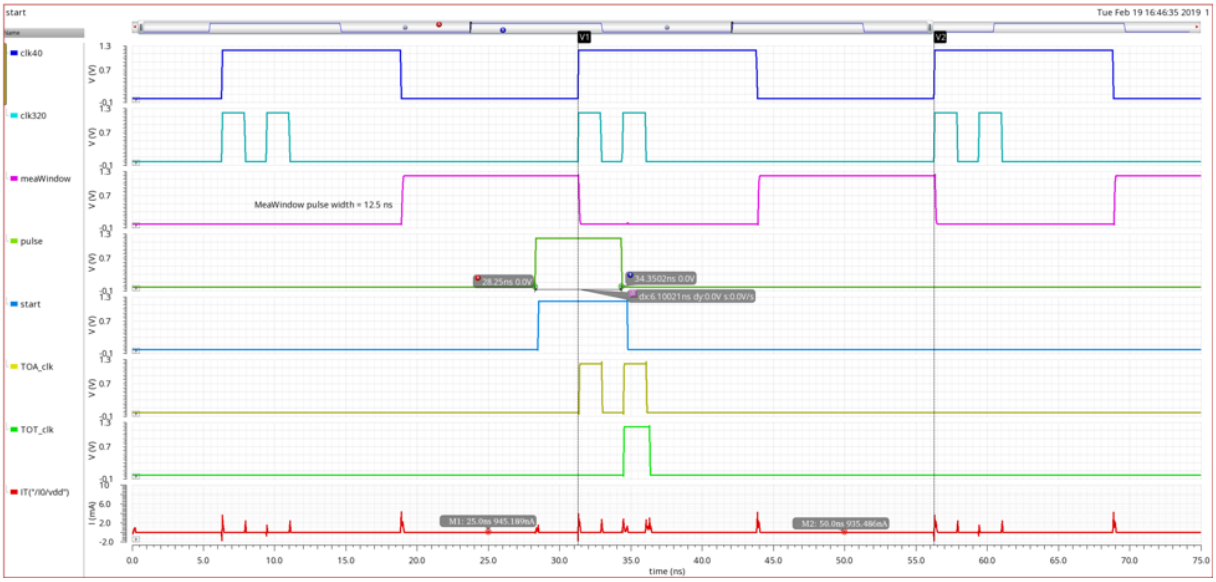


Figure 3.41: The post layout simulation of the typical waveforms of the signals from the pulse generator. In the second 25 ns clock period, a hit occurs. Start, TOT_clk and TOA_clk are output to the delay line. In the third clock period, there is no hit, so the pulse generator keeps quiet.

the post-layout simulation estimates the power consumption at 91 μW. When there is no hit in the defined window, the power of the pulse generator is 0.9 μW only.

Delay line structure

Experience with implementing TDCs in FPGAs in recent years demonstrates that it is possible to achieve a 10 ps-level precision by employing real-time calibration instead of controlling the propagation delay of individual delay cells. Since an uncontrolled delay cell is significantly simplified by the elimination of the voltage control circuit, it becomes possible to minimize the delay time per cell and thus achieve a small enough time delay step to also eliminate the need for a second (vernier) delay line. With an individual cell delay below 30 ps, a single delay line can be used. The cell-to-cell delay uniformity is expected to be at the level of a few percent.

Without a closed-loop delay control for each delay cell, there is no compensation for delay changes from process, temperature variations, power supply fluctuations and radiation damage. However, these effects can be mitigated by calibrating the average cell delay time with the measurement of a precisely known clock period using the same delay line.

The concept of the proposed TDC delay line is shown in Fig. 3.42. The delay line is cyclic in order to reduce its length and thereby its power consumption. The delay line, which is in fact an oscillator, includes 63 NAND gates, where all but the first NAND gate are configured as inverters. By using dummy NAND gates as inverters, the load on each tap is kept equal, making cell delays more uniform. The first NAND gate is controlled by the Start signal. When the Start signal is low, the oscillator is disabled. When a hit arrives, the oscillator starts to oscillate and drives a 3-bit counter. In a typical corner post-layout simulation, the delay between adjacent taps is about 22 ps and the oscillating frequency is about 360 MHz. For each rising edge of the TOA or TOT clock signal, the TDC takes a snapshot of each tap in the delay line and the counter. The counter records the coarse phase information, while the D flip-flop (DFF) array records the fine phase information. Because the TOT resolution requirements are more relaxed than for that of TOA, the DFFs for TOA measurement attach to every tap of the delay line, while the TOT DFFs only connect to every other tap. To save power and balance the load of the taps,

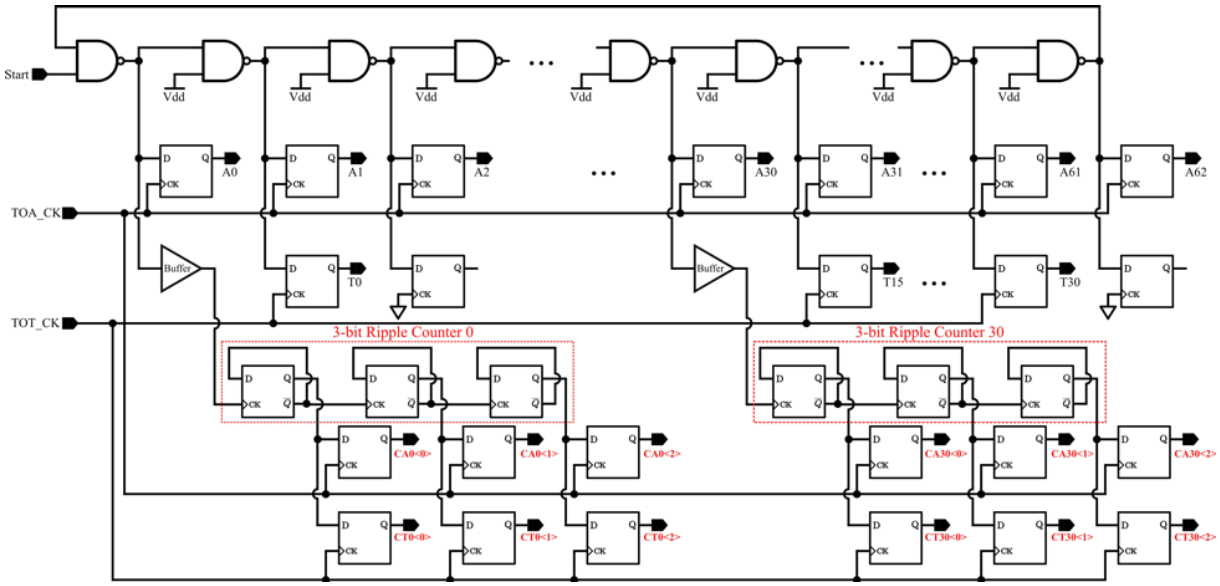


Figure 3.42: The proposed delay chain design for the simultaneous measurement of TOA and TOT.

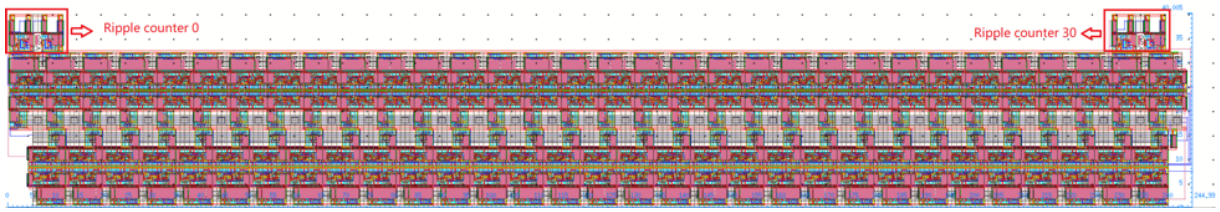


Figure 3.43: The preliminary layout of the 63-tap delay line. The layout size is $245 \times 40\ \mu m^2$.

the counter is a simple ripple counter which connects to the last tap without a TOT DFF. Some dummy DFFs are used to keep the load of all taps uniform. Finally, since the clock driving the counter and the readout clock are asynchronous, we need to protect the TDC against possible metastability when the read operation and counter change are concurrent. Since the DFF array retains fine phase information about the counter clock, we can resolve the metastability by using two counters driven with different phases. Then, the fine phase information from the DFF can be used to select the appropriate counter. This approach is the same as used in the ALTIROC TDC. The TOA/TOT measurement range is extended by the ripple counter to 22 ns at negligible power cost. We can drop bits in the future to save readout bandwidth if we do not need such a large dynamic range.

Based on this concept, a preliminary implementation of the TDC delay line in 65 nm technology was performed and the layout is shown in Fig. 3.43. The enclosed layout transistors (ELT) technique was used in the implementation to minimize the effects of radiation damage. The degeneration of the maximum current for transistors implemented with the ELT technique has been measured to be less than 10% for NMOS and 20% for PMOS devices at room temperature [93]. In addition, irradiation tests of inverters have shown a 10% time delay increase after 200 Mrad of irradiation at temperature of $-20\,^{\circ}C$ of irradiation for the ELT layout cell library. Based on these studies, the expected delay degradation of this delay line is expected to be no more that 10%.

To maintain the time measurement with the resolution of about 20–30 ps, the skew of the clock signals delivered to the DFF array should be minimized. This clock skew has been studied

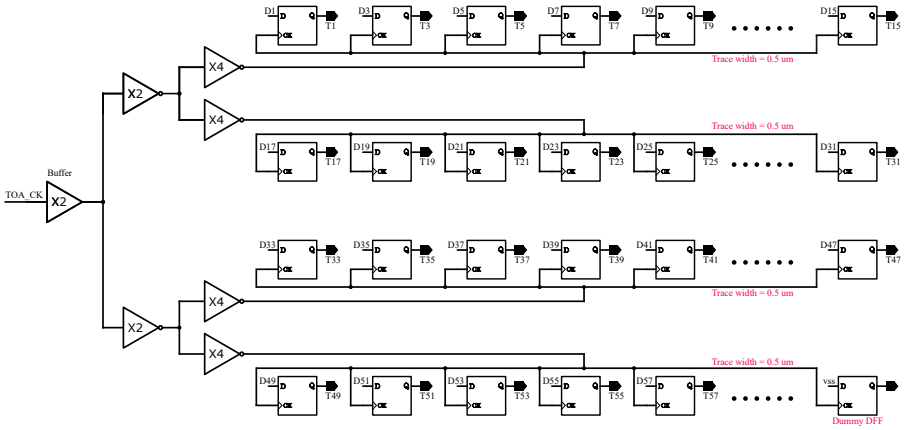


Figure 3.44: The delay line clock tree structure to minimize the clock skew.

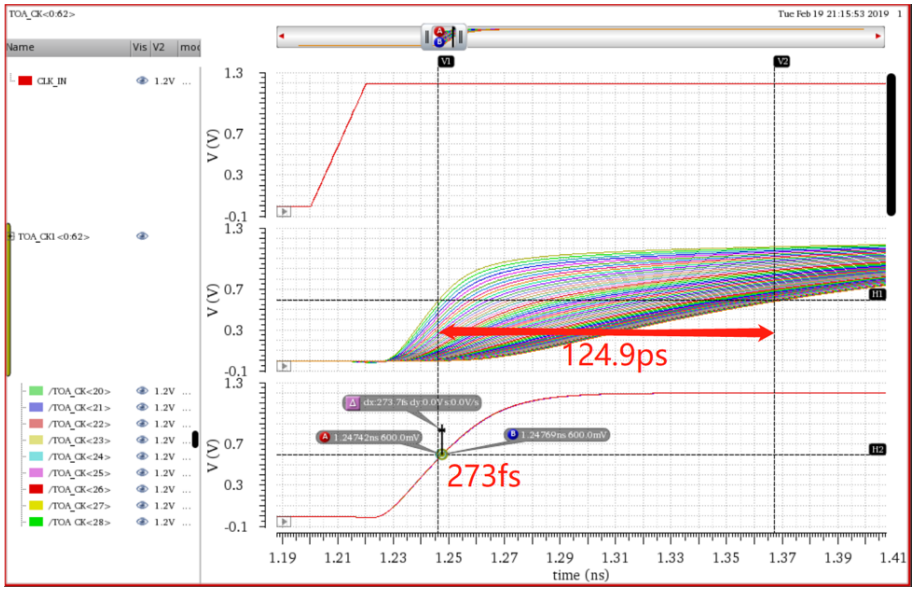


Figure 3.45: The effect of the delay line clock tree structure to minimize the clock skew. The clock skew is significantly improved from 125 down to 0.3 ps, based on post-layout simulation.

extensively using the parasitic trace model. If a clock driver is used to drive all the TOA DFFs with a single 0.2 μm trace width, the clock skew can be as high as 125 ps, because of the effect of parasitics. To minimize the clock skews, eight clock drivers are used with distributed 0.5 μm clock traces as shown in Fig. 3.44. Figure 3.45 shows how, based on post-layout simulation, the clock skew is significantly improved from 125 ps down to 0.3 ps with respect to using a single larger driver with a 0.2 μm trace width. The clock skew due to the process mismatch has also been studied and is about 0.3 ps, comparable to the effect above.

Delay cell calibration

To maintain the required TDC measurement precision using a tapped delay line with uncontrolled delay cells, a specific strategy must be devised to compensate for delay time variations caused by temperature and power supply voltage fluctuations as well as radiation damage. Since these variation sources result in changes on a much longer time scale than the measurement, it is possible to correct for the corresponding delay deviations during data-taking using the data itself. As mentioned earlier, for each input signal hit, the TOA clock has two consec-

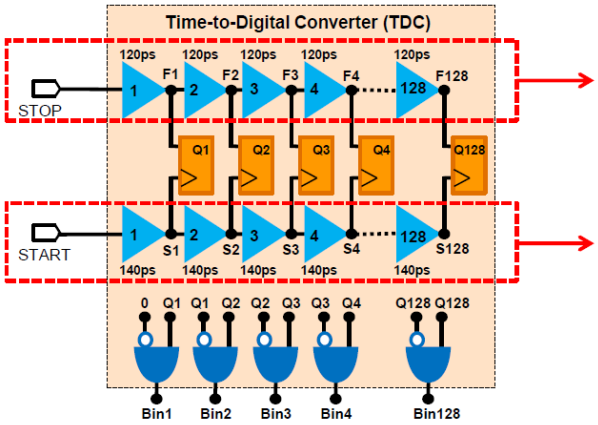


Figure 3.46: The ALTIROC TDC design based on a vernier delay line [94].

utive rising edges leading to recording two time stamps with a time difference of exactly one 3.125 ns clock period. The average delay time per cell is then simply given by the clock period divided by the number of recorded delay taps between the two rising edges. This information can be used to correct the measurement being taken, or it can be averaged over a number of hits to achieve higher precision on the average time delay estimate. The addition of this information to the data stream increases the data output of the TDC by about 20%, which is within the current capabilities of the system.

3.3.6.2 Vernier delay line design

The ATLAS ALTIROC design requires a 20 ps (40 ps) bin size in the TOA (TOT) measurement. In the 130 nm process such precision can be obtained by using the vernier delay line approach. As illustrated in Fig. 3.46 the vernier-based TDC uses two delay lines, one faster and one slower. The timing resolution (or bin size) is determined by the difference in the delay of the cells in each line. For example, if the slow line has 140 ps delay for each cell and the fast line has 120 ps delay for each cell, then the step (or the LSB of time measurement) is 20 ps. A hit, or the output of the discriminator as the Start signal, will enter the slow delay line, while the Stop signal will enter the fast delay line. As the Start and Stop signals propagate through the two delay lines with different delays, the distance in time between the Start and the Stop will decrease by 20 ps each time a delay cell stage is passed. The number of delay cell stages it takes for the signal Stop to get ahead of the Start signal is a measure of the time between the two with a time bin of 20 ps. A cyclic structure is employed to reduce the number of cells needed per line, and thus reduce the TDC footprint. To minimize the power consumption a time measurement is only initiated in the presence of a Start signal. This way the power consumption approximately scales with the occupancy. To limit the impact of temperature, power supply fluctuations, and process variations on the measurement precision, the delay time for each delay cell is dynamically adjusted by setting a reference voltage derived from a delay-locked loop (DLL) circuit. Based on ALTIROC studies [94], the average power consumption per pixel is expected to be 0.35 mW, assuming an occupancy of 10% and measured times uniformly distributed within the time window. To measure both the TOA and TOT, two TDCs are required for each pixel, bringing the total power consumption to approximately 0.70 mW per pixel.

3.3.6.3 Comparison of the TDC approaches

The ETROC TDC design introduces several power saving features. It uses the same delay line to measure both the TOA and the TOT, uses a single delay line with a reduced number of delay

cells, takes advantage of a cyclic structure, and removes the need for DLLs to control the delays of individual cells. This design approach is made possible by the in-situ calibration scheme which enables a significantly simpler delay cell structure with the consequent minimization of the individual cell delay. Based on preliminary post-layout simulation studies, this design is expected to lead to a significant reduction in power consumption.

Simulation of the prototype ETROC TDC delay line indicates a power consumption of about 0.1 mW at 10% occupancy for a TDC measurement of 12.5 ns (specification is 6.25 ns). This power consumption can be compared with about 0.7 mW, which is the power consumption of the ATLAS ALTIROC double delay line TDC based on traditional vernier delay line and with a TDC measurement window of 2.5 ns. The low power feature of the ETROC TDC allows us to extend the TDC measurement window (thus the physics capabilities) for slower moving particles, such as particle identification for the heavy flavor physics as well as for the heavy ion physics program, and possibly for long-lived particle searches. The TDC power consumption scales with occupancy, and for ETL outer layers, the TDCs will consume much less power than the ones located at inner layers. Since there are many more ETROC chips in the outer layers, this will result in a significant power saving at the system level. In addition, the relative simplicity of the single delay line approach can translate into improved reliability in the face of radiation damage.

Given the fact that the ETROC TDC implementation is new, the traditional vernier approach will be kept as a fall-back option and the ETROC development plan will include, in fact, the testing of both TDC implementations.

3.3.7 Clock distribution

3.3.7.1 Clock generation unit

The clock generator unit provides clock signals to all functional blocks in the ETROC chip. As shown in Fig. 3.47, it includes a PLL and a phase shifter. The 40 MHz LHC clock is used as input. The oscillator in the PLL works at 1.28 GHz which is the highest frequency we need for the ETROC chip. A 320 MHz clock is needed for each TDC channel, as discussed in Section 3.3.6. A voltage-controlled delay line in the phase shifter is used to adjust the phase of the clock in 50 ps steps.

3.3.7.2 Phase shifter

The phase of the time reference controls the position in time of the measurement window of the TDC and can be adjusted via slow controls. The SMU group has already designed a phase shifter in 65 nm technology for the lpGBT project. We plan to re-use the most critical components of this design for ETROC. A schematic of the proposed phase shifter for ETROC is given in Fig. 3.48. The resolution is 50 ps and the periodic jitter is less that 5 ps, peak-to-peak. The basic delay cell is made of two current-starved inverters in series.The cell is laid out with the ELT technique [93] for radiation tolerance. This phase shifter is optimized for high speed and high resolution, thus the inverters are large.

3.3.7.3 Clock distribution

Effective clock distribution is a critical challenge in the design of any high-speed chip [95] and extreme care must be taken in minimizing clock skew and jitter. For ETROC we plan to adopt the most common and conservative clock distribution scheme, known as H-tree [96]. In this scheme, the clock is branched from a central point (root) to all its destination nodes (leaves), using a balanced quaternary tree. The clock distribution network for the 16×16 pixel array, as

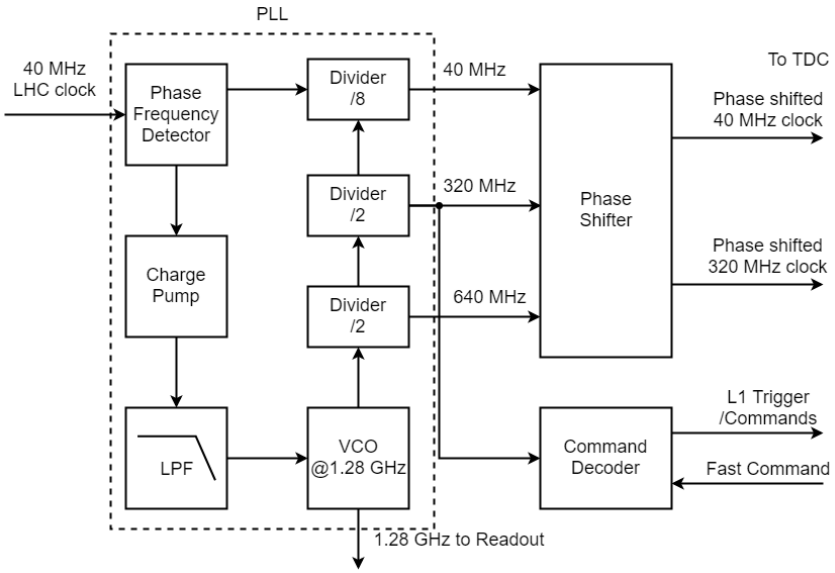


Figure 3.47: The proposed block diagram of the clock generator.

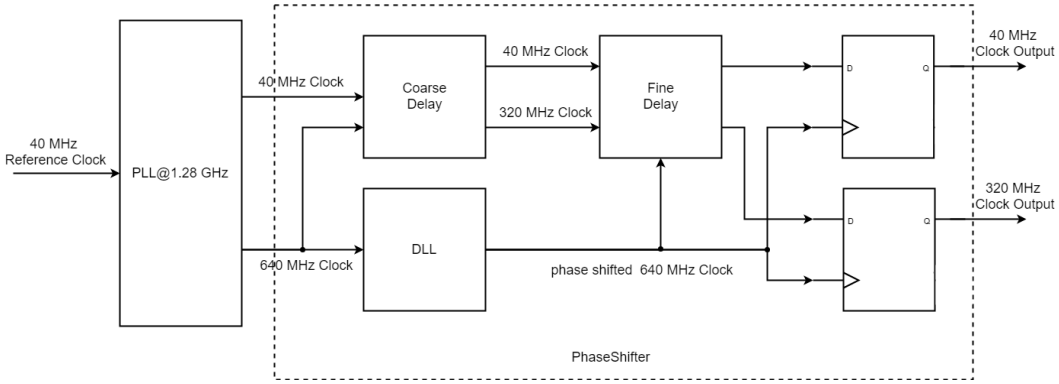


Figure 3.48: The proposed phase shifter structure for ETROC.

is currently being designed and simulated, is shown in Fig. 3.49. Each pixel occupies an area of 1.3×1.3 mm^2, thus the total chip area is about 2×2 cm^2. The clock is distributed starting from the center of the south edge of the chip, to each of the 256 pixel elements. The 4-stage H-tree structure ensures that the clock distribution network is balanced and that all the path lengths from the clock source to each pixel are equal.

To ensure the quality of the clock signal, inverters/buffers are added along the transmission path, as discussed in Ref. [97]. Clock distribution networks for both 40 and 320 MHz clock frequencies are being designed and simulated.

Our preliminary post layout simulation results indicate that the worst-case-scenario clock skew for the 320 MHz clock among the 256 elements is below 17 ps. The clock skew is measured as the time difference at 0.6 V on the rising edge of the output of the final destination, as shown in Fig. 3.50. Simulations indicate that the standard deviation of the worst case clock skew is 0.8 ps due to variation in process corners. The worst case clock jitter is below 5 ps due to variation in process corners and the worst case clock jitter is defined as the difference between the longest and shortest clock period (peak-to-peak) among the 256 elements. The overall power consumption of the clock distribution network for a 320 MHz clock is about 21 mW.

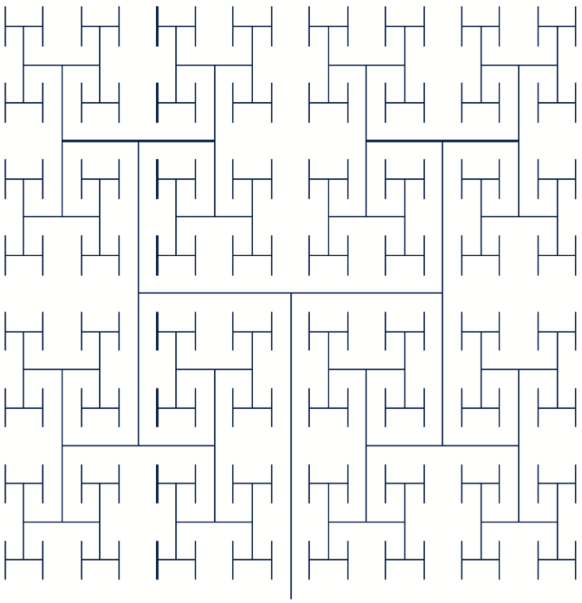


Figure 3.49: Layout of a clock distribution for the 16×16 pixel cell matrix, each 1.3×1.3 mm^2 size, using metal layer 6. The clock is distributed starting from the center of the south edge of the chip to each of the 256 pixel elements through a four stage H-tree structure. The first stage is located at the center of the chip. The clock signal frequency is 320 MHz with 10% rise and fall time.

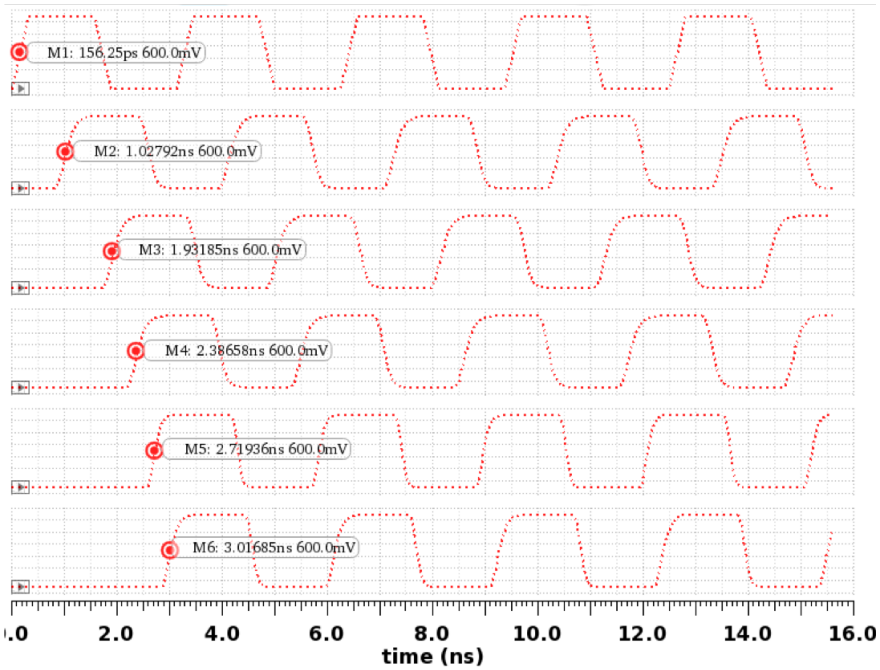


Figure 3.50: Internal clock signals at different stages. Top is the input clock signal, followed by the signal at the input of the next four stages, and the bottom signal is the output at the final destination.

3.3.8 Level-1 Buffer and data readout

Figure 3.51 shows the top level block diagram for data readout. At the bottom of the matrix a frame builder block receives data from each column and sorts out and assembles data with

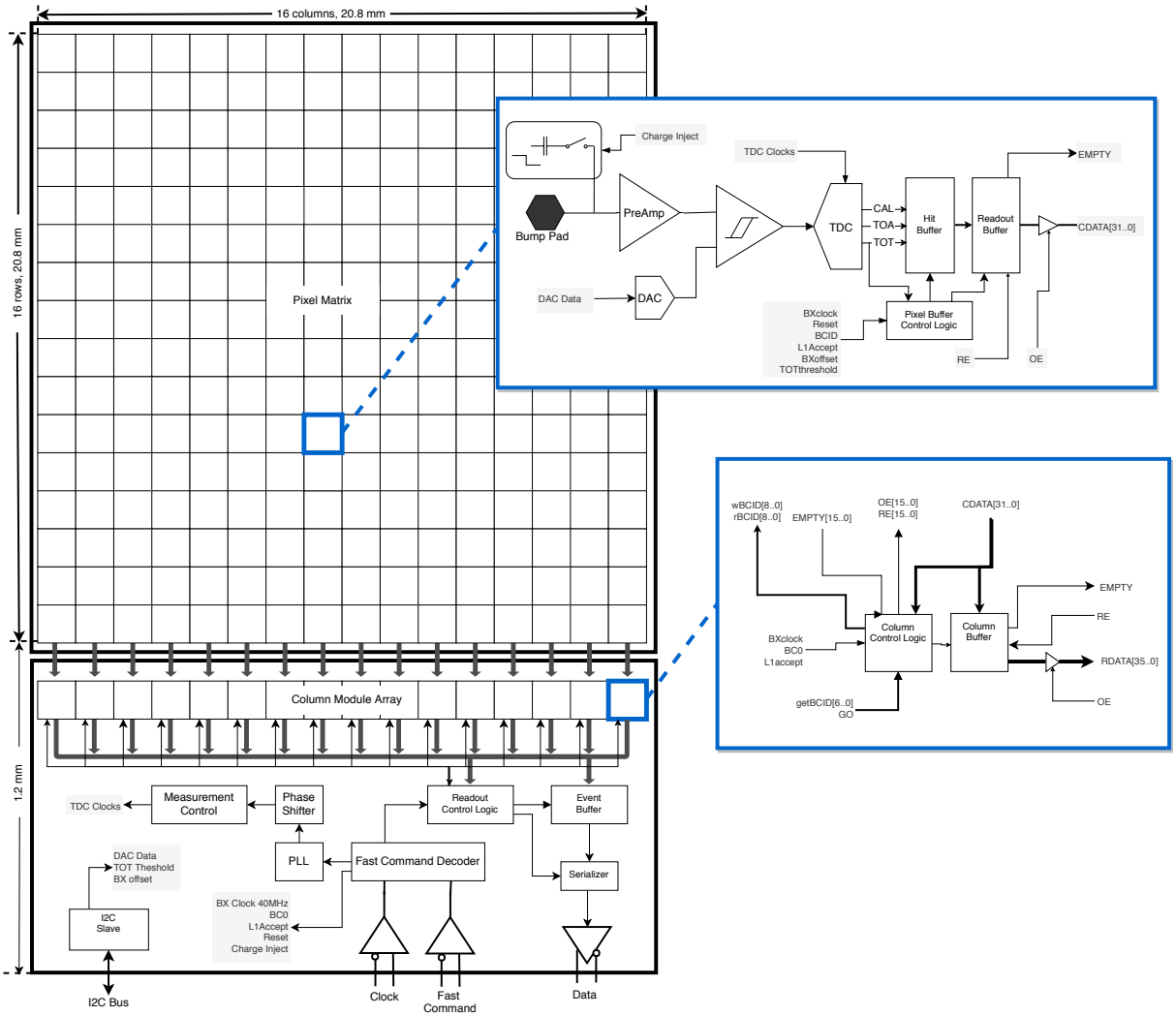


Figure 3.51: ETROC top level data readout block diagram.

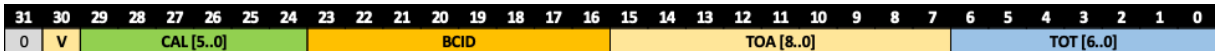


Figure 3.52: Event or readout buffer data format.

variable length. The data includes a 16-bit header, a 12-bit BCID (Bunch Crossing ID), a 4-bit length indicator and data section for timing information from pixel hits. Each data block coming from one pixel includes an 8-bit pixel address, 10-bits of TOA data, 9-bit TOT data and 5-bit (or 6-bit) TDC calibration data.

3.3.8.1 Pixel readout components

In each pixel the preamplifier amplifies the signal from the sensor. Then the discriminator digitizes the analog pulse with the proper threshold. The TDC circuit measures the time of arrival (TOA) relative to the master clock and also the time over threshold (TOT) of the digital pulse. With every input hit the TDC circuit produces calibration (CAL) data and thus no dedicated calibration mode is required for the TDC delay line.

Immediately following each pixel's TDC circuit is a block of static RAM called the "hit buffer". The hit buffer stores a "valid" flag, TOA data, TOT data, and CAL data at each bunch crossing, as shown in Fig. 3.52. If there is no hit, nothing is actually written into the hit buffer, and a write

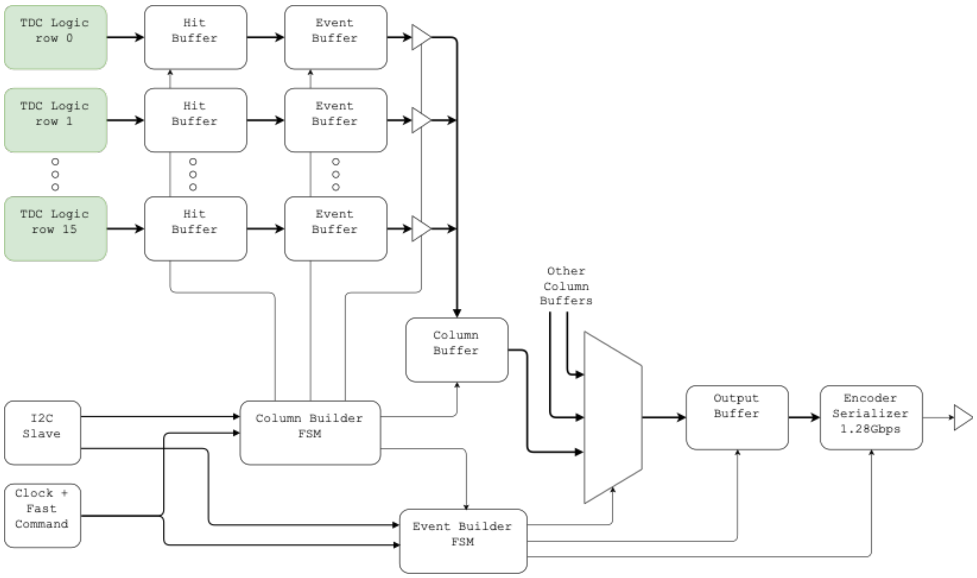


Figure 3.53: Hit buffer and data readout.

pointer is simply incremented. The hit buffer is constructed from an SRAM array and supports simultaneous read and write access. Independent read and write buses can be accessed directly (e.g. RAM access mode). The hit buffer is 32 bits wide and has a maximum depth of 512 words, which allows for a maximum Level-1 Accept system latency of 12.8μs.

To handle consecutive Level-1 Accepts in bursts, a Readout (or Event) Buffer immediately follows the Hit Buffer, as shown in Fig. 3.53. When a Level-1 Accept occurs, the hit (if present) is immediately transferred from the Hit Buffer into the Readout Buffer. The Readout Buffer is implemented as a FIFO. The output of the Readout Buffer connects to a shared vertical readout bus (called CDATA) through a tri-state driver which is enabled (OE) by the Column Module.

The Pixel Buffer Control Logic (PBCL) block monitors the output of the TDC, specifically the TOT value. Only hits which meet or exceed a user-defined TOT threshold are considered valid and are thus written into the Hit Buffer. The Hit Buffer write and read pointers need to be carefully managed and this is handled by the PBCL. Typically, the read pointer trails the write pointer by a user-defined BX offset; this ensures that whenever a Level-1 Accept occurs, the correct hit is immediately available at the Hit Buffer output and no address calculations are needed.

3.3.8.2 Column control logic

At all times the Column Control Logic (CCL) runs a counter to generate an 8-bit BCID from the BX clock and BC0 control signal. This BCID value is sent up to all pixels in the column where it is used as a time stamp for valid hits as they are written into the Hit Buffer.

Column readout is triggered by the GO signal. When GO is asserted the CCL latches the supplied target BCID and begins the process of sequentially collecting hits from the column Readout Buffers. Since the Readout Buffers may contain hits from multiple events, it is critical that the CCL checks the BCID of each hit before it is written into the Column Buffer.

First, the CCL scans the 16 Readout Buffers and identifies which Readout Buffers (if any) in the column have hits to read out. Starting with the first non-empty Readout Buffer, the hit data is driven (OE = 1) onto the column data bus CDATA. This process continues until all hits for the target BCID have been loaded into the Column Buffers (worst case 16 clock cycles).

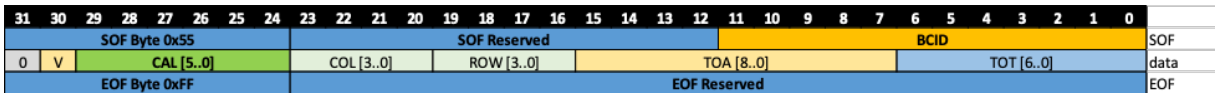


Figure 3.54: ETROC data output format.

The Column Buffer needs to be wide enough to hold the complete hit plus 4 bits to indicate the Row ID of the hit. This buffer needs to be deep enough to hold up to 16 hits. All Column Buffers have tri-state output drivers and share a common output data bus (called RDATA). Column Buffer status flags (EMPTY) and control bits (RE and OE) are used for the next stage of readout logic.

3.3.8.3 Whole chip readout components

The Readout Control Logic (RCL) is responsible for triggering the Column Modules, collecting the hits from the Column Buffers, and assembling the output frame in the Event Buffer prior to serialization. The RCL uses the BX clock and BC0 control bit to generate a 12-bit BCID counter. When a Level-1 Accept occurs the RCL calculates the BCID of the target event (tgtBCID) and sends that number to all Column Modules along with the GO signal. At this point all hits for the target event are sitting in the 16 Column Buffers. Now the RCL scans across the Column Modules, reads hits out of the Column Buffers and writes the hits into the Event Buffer. Each Column Buffer is read until it is empty before moving on to the next column. After the last hit is stored in the Event Buffer the RCL sends the hit data to the serializer along with the header and trailer words. The latency between the time the Level-1 Accept is received and the time the start of frame (SOF) is sent out is variable depending on the number of hits present.

The process of collecting up to 256 hits from the pixel array is efficient as many operations occur in parallel, however the process does take a finite amount of time. While the readout is progressing, it is entirely possible that a new Level-1 Accept can occur. In this case the RCL must store the new target BCID(s) and trigger another GO cycle as soon as the readout logic returns to idle.

The Event Buffer stores all hits for a single event. This buffer must be wide enough to store the complete hit plus 8 bits for the Row and Column ID. The maximum depth is 256 words.

3.3.8.4 Serializer and output format

Once the output frame is assembled, it is transmitted as a single serial data stream. This variable length output frame is organized as 32-bit double words as shown in Fig. 3.54.

No special out of band synchronization symbols are required on this serial link since the serial data is synchronous to the 40 MHz master clock. The downstream E-link receiver adjusts the phase of the incoming data stream relative to the 40 MHz until proper byte/word/double-word alignment is achieved. The next step is to define a few unique data patterns so that the frame boundaries can be properly determined. SOF may be identified by looking for 0x55 in the upper byte, although this pattern is not unique since it may also occur in a data word. The only pattern guaranteed to be unique is 0xFF in the upper byte of end of frame (EOF), and will be searched and checked after a SOF is identified (this alignment scheme is the same as used in the ALTIROC). If no hits are present a SOF and EOF are still transmitted, for a total of 64 bits. The maximum frame size is SOF + 256 hits + EOF, or 8256 bits. Currently a frame length field is not specified, however there are sufficient reserved bits in the SOF and EOF words which may be utilized for this purpose.

The data rate of the ETROC serial output stream may be configured to one of three pre-defined

rates depending on the anticipated occupancy of the device. The slowest data rate is 320 Mbps, where a single byte is transmitted during a 40 MHz main clock cycle. The next highest data rate is 640 Mbps, or one 16-bit word is transmitted during each 40 MHz clock cycle. Finally, the highest supported data rate is 1.28 Gbps where a 32 bit double word is transmitted during a single 40 MHz clock cycle.

Note that no additional encoding schemes (e.g 8b/10b, 64b/66b, etc.) or forward error correction bits are used on the serial output, therefore there is no overhead associated with the ETROC serializer. Assuming a Level-1 Accept rate of 1 MHz, we have an average maximum transmission time of 1000 ns. When not transmitting a frame the serializer shall send all zeros.

Faster, lower latency readout structures such as Fischer trees [98, 99] are also being considered for the column pixel readout logic.

3.3.8.5 Level-0 Trigger option

The Level-0 option is still being studied and the detailed specification is to be determined. Architecturally, the Level-1 Accept hit buffer and readout path cannot be shared with Level-0 Accept readout. One simple way to avoid the conflict is to have an independent copy of the Level-1 Accept readout path dedicated to Level-0 Accepts, including the hit buffer, event buffer and the rest of readout chain with its own E-link output.

3.3.9 System interfaces

Figure 3.51(bottom) shows the system interfaces of the chip. A slow control I2C block provides access to programmable features, such as the gain of preamplifiers, pre-emphasis, and hysteresis of discriminators. A 40 MHz LHC clock is received by a fast control decoder block from a dedicated clock input. The fast control decoder block decodes the fast control signals such as Level-1 trigger and BCID through the Fast Commands input. A serializer converts the data into a bit-stream at a data rate up to 1.28 Gbps depending on the expected occupancy (typically 320 Mbps is enough) and an ePort Tx block from the lpGBT serves as the data transmitter. A PLL generates clocks for TDC and readout from the 40 MHz input clock. A phase shifter adjusts the phase for readout and measurement window.

3.3.10 Waveform sampling design

For the robust long term operation of the ETL it will be important to monitor signal waveforms to detect variations from the increasing radiation dose. Recorded waveforms can be used to optimize thresholds and bias voltages in order to try to maintain the target performance. Since implementing waveform sampling for all channels is not possible, the plan is to only implement it for two pixels per chip, based on the assumption that nearby sensor pixels will behave and age in a similar way. Beam test data waveform studies have shown that a waveform sampling speed between 2 and 3.2 GS/s is sufficient for this purpose.

There are different ways to implement waveform sampling. One traditional approach is to use a switched-capacitor array. Another approach is to use time-interleaved ADCs. While the switched-capacitor array approach is still a possibility, the ADC-based approach is being pursued as the baseline since an appropriate single channel ADC based on a time-interleaved [100] pipelined Successive Approximation Register (SAR) ADC [101] has already been developed. This approach allows the implementation of a waveform sampler with a high sampling rate of about 3 GS/s, a bandwidth up to the Nyquist limit of 1.5 GHz, high resolution (ENOB $\geq$ 10), low power consumption ($\leq$1 mW), and small silicon footprint ($\lesssim$1 mm^2).

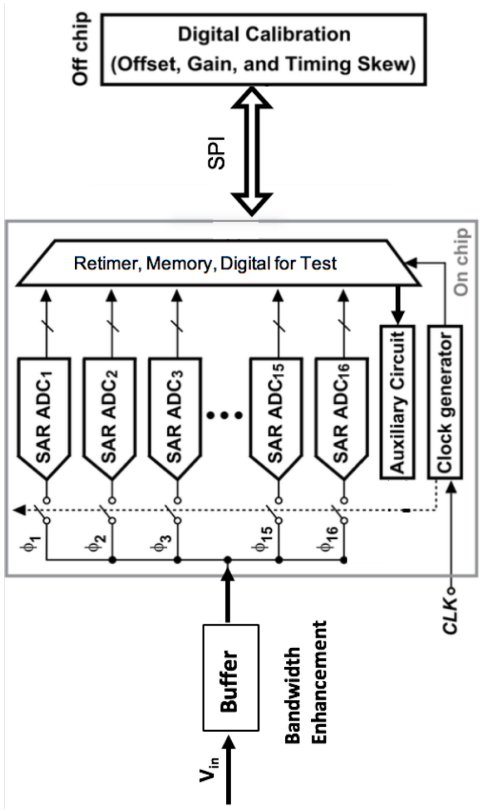


Figure 3.55: Block diagram of a 16-fold time-interleaved ADC architecture required to achieve high sampling rates. Recent studies show that an 8-fold time-interleaved architecture, based on an existing ADC design, will achieve 2.56 GS/s sampling rate using 320 MHz clock and is expected to be suitable for the ETROC.

Using the 65 nm CMOS process, a single pipelined SAR ADC can achieve a sampling rate approaching 400 MS/s with 10 ENOB resolution. To achieve higher sampling rates, such as above 3 or 6 GS/s, 8 or 16 ADCs (termed as sub-ADCs) can be interleaved together in a sequential-sampling and parallel-processing manner, wherein each sub-ADC is sequentially sampled with a fast clock and holds the value for digitization in parallel with the other sub-ADCs. This architecture achieves a higher aggregate sampling rate while each sub-ADC still operates at a lower speed. The principle is illustrated in Fig. 3.55, where 16 sub-ADCs are clocked at a rate of 400 MS/s and interleaved to achieve a sampling rate of 6 GS/s. The clock generation circuit produces 16 clock signals, one per sub-ADC, with the same frequency but with 16 different phases, evenly distributed within one clock period. These clocks are used to control the sampling switches of the 16 sub-ADCs ($\phi_1 - \phi_{16}$), as shown in Fig. 3.55. For the actual implementation for ETROC, only 8 ADCs will be interleaved, and since the available clock is naturally 320 MHz (not 400 MHz), the expected waveform sampling rate is about 2.56 GHz and this still statisfies the required sampling speed (above 2 GHz). The on-chip re-timer will synchronize the data outputs from the multiple sub-ADC channels to form the final outputs, and the on-chip memory can store the ADC outputs before they are read out through a data interface such as E-links [102].

The details of the implementation of the waveform sampling for ETROC are described in Appendix C.II.

3.3.11 Radiation effects and mitigation techniques

Two radiation effects must be taken into account: the Total Ionizing Dose (TID), which may affect the timing measurement precision by degrading the performance of the chip circuitry over time, and Single Event Effects (SEEs), which may reduce the data-taking efficiency by corrupting the data stream or the ASIC configuration registers. Mitigation strategies for these effects in the 65 nm technology have been developed and extensively studied in the context of designing the ATLAS and CMS tracking system ASICs by the RD53 collaboration [103]. The expected TID for the inner and outer CMS trackers is 500 Mrad and 200 Mrad, respectively, to be compared to 100 Mrad for the ETL. The radiation-hardness requirements for the ETROC can therefore be readily satisfied by adopting established techniques.

3.3.11.1 TID mitigation strategies

Ionizing radiation degrades the performance of MOSFETs by increasing the threshold voltage and producing undesired leakage currents. The mechanisms of imparting radiation damage have complex dependence on device geometry, production process, temperature and bias. To mitigate these effects we generally follow strategies motivated by studies by the RD53 collaboration.

For analog designs the approach for limiting the radiation damage relies on trading radiation-hardness for feature size. Specifically, in MOSFETs, the Radiation Induced Narrow Channel Effect (RINCE) imposes minimum channel width constraints on the device, and the Radiation Induced Short Channel Effect (RISCE) imposes minimum channel length constraints [104]. PMOS and NMOS transistors are affected differently, with PMOS transistors being more severely impacted. To ensure radiation-hardness up to 500 Mrad, we plan to enforce minimum dimensions for PMOS and NMOS of (W: 300 nm, L: 120 nm) and (W: no restriction, L: 120 nm), respectively. In addition to observing these device-level recommendations, we will employ guard rings to counter inter-device leakage currents as needed. These design rules are employed throughout the key analog building blocks of the ETROC, including the preamplifier and the discriminator implementations. As described in Section 3.3.6, to minimize potential radiation damage effect on the TDC delay cell time drifts, the ELT technique [93] is used for the implementation of the TDC.

For digital designs the strategy is to select the most radiation-hard standard digital cell libraries and optimize the synthesized design for various conditions expected during operation, e.g. at different operating temperatures or after different amounts of irradiation. Dedicated irradiation studies [93] have been performed with a variety of digital standard cell libraries (7, 9, 12 and 18 track), which vary in the minimum device dimensions and transistor types as normal, low and high Vt (transistor threshold voltage). Based on the results, we plan to use a 12 track nominal Vt library, with channel width W larger than 480 nm (360 nm) for PMOS (NMOS) transistors, and optimize the ETROC digital designs based on simulations with both unirradiated and 200 Mrad irradiation transistor models to ensure that the ETROC has a satisfactory performance over the full detector lifetime.

The Level-1 buffer will be designed using either an SRAM or a FIFO. Existing tools can be used to render either implementation suitable for the expected TID. A pseudo-dual-port SRAM compiler has been specifically designed by the CERN IC group to minimize power consumption while maintaining radiation tolerance consistent with the needs of LHC Phase-2 tracker upgrades. Radiation-hardness is achieved by requiring that the widths for PMOS and NMOS transistors are larger than 500 and 200 nm, respectively, and employing guard rings. Using this compiler, a prototype SRAM test chip has been designed and subsequently tested, demonstrat-

ing radiation tolerance up to at least 200 Mrad [105]. Alternatively, a FIFO approach, planned to be tested for radiation-hardness by the RD53 collaboration, may be considered for its potentially smaller footprint and/or power consumption.

3.3.11.2 SEE mitigation strategies

As ionizing particles pass through the readout chip they may deposit enough energy to result in a bit-flip or signal glitch that causes errors or undesired subsequent chip behavior. To reduce such events, analog designs are hardened against SEEs by increasing capacitance at sensitive nodes and digital designs are hardened by using triple modular redundancy (TMR). Dedicated studies are underway to refine the TMR implementation by optimizing the spatial separation between signal replicas and by potentially adding time delays to also separate signal replicas in time [106]. In addition to protecting pixel and global configuration registers, TMR will also be used for the clock and reset signals. For clock signals, we are also studying the possibility of using three separate clock domains as opposed to creating multiple clock replicas sourced by the same clock domain. The SRAM memory compiler, used to implement the Level-1 buffer, employs drive strength increase to harden the reading/writing of addresses and flip flops based on the Dual Interlocked Storage Cell (DICE) for address storage [105]. Finally, to ensure that any residual SEEs are minimized, similar to the approach taken in the RD53 chip design, we plan to continuously refresh the ETROC configuration simultaneously with data-taking, a method referred to as "trickle configuration".

3.3.12 Optimization of power consumption vs performance

The power consumption obtained from simulations described in the previous sections are summarized in Table 3.7, where the results are given for both the low-power and high-power preamplifier current settings. As shown in Fig. 3.36, the default, low-power setting provides single-hit time resolution matching the specification of 30–40 ps at the beginning of operation and 50–60 ps at the end of operation, while the higher preamplifier power setting would provide time resolution of 30–40 ps per hit for the full lifetime. With either power setting, the expected power consumption meets the 1 W per ASIC specification listed in Table 3.5.

Table 3.7: A summary of ETROC power consumption for each circuit component. The preamplifier, discriminator, and TDC values are obtained from post-layout simulation with conservative assumptions about occupancy and operating temperature. The SRAM and global circuitry power consumptions are conservative extrapolations from similar circuits used in the ALTIROC.

Circuit component	Power per channel [mW]	Power per ASIC [mW]
Preamplifier (low-setting)	0.67	171.5
Preamplifier (high-setting)	1.25	320
Discriminator	0.71	181.8
TDC	0.2	51.2
SRAM	0.35	89.6
Supporting circuitry	0.2	51.2
Global circuitry		200
Total (low-setting)	2.13	745
Total (high-setting)	2.71	894

These simulated power estimates will be verified and made more precise with prototype testing. Power consumption verified to be less than the requirement will provide an opportunity

for optimizing the cooling requirements and the time resolution. Adjustment of preamplifier power settings during operation will also allow optimization of the time resolution, e.g., by rebalancing power between the outer and inner radius regions to maintain optimal resolution in the irradiated regions.

3.3.13 Development plan and schedule

The ETROC development plan consists of three phases. The first phase is the design specification study phase following the methodology described earlier. The second phase is the prototyping phase where test chips are fabricated and assembled onto various test boards. These development efforts aim to characterize and evaluate sub-circuits and other parts of the ETROC design. The third phase is the system level testing phase to fully verify the final chip functionality and performance. In this third phase the ETROC interfaces (sensor inputs, control and data links, etc.) will be connected to hardware which is as close to the final system as possible. At the time of writing, the first phase has been essentially finished and we are now deep into the second phase.

For the second phase prototypes, there are essentially three important areas we intend to focus on. The first is the analog front-end performance. This includes the preamplifier, the discriminator, and the TDC stage. The second is the precision clock distribution within the chip. The third is the evaluation of the radiation-hardness (TID, SEU) of the design. To effectively and efficiently meet these design challenges we have divided the prototyping phase into a series of increasingly complex prototype chips named ETROC0 through ETROC3.

3.3.13.1 ETROC0 1x1

The goal of the ETROC0 mini-ASIC is to study and demonstrate the performance of the preamplifier and discriminator. The ETROC0 is a single channel design which consists of a preamplifier with integrated charge injection followed by a discriminator with a user programmable threshold controlled by an internal DAC.

Wherever possible the ETROC0 test structures have been designed as individual independent blocks to facilitate isolated testing and characterization. These blocks include a preamplifier (followed by a buffer), a standalone discriminator with a dedicated DAC, and a standalone charge injection circuit. Since the amount of charge injected by the charge injection circuit depends on its internal capacitor value, an extra capacitor (a copy of the internal capacitor design) has been included and brought out to pins to allow for a direct measurement. The ETROC0-V0 has been designed and submitted in December 2018, with the chip received in March 2019. ETROC0-V0 mini-ASIC is shown in Fig. 3.56. Tests of this prototype find performance that is consistent with the expectations from simulation for power consumption and the time resolution of injected pulses. An second version, the ETROC0-V1 mini-ASIC, will have the new waveform sampling circuit, and it is scheduled for submission in summer 2019.

3.3.13.2 ETROC1 prototype

The goals of the ETROC1 prototype are to study the performance of the full front-end chain and a simple precision clock distribution scheme. The ETROC1 chip is a 4×4 pixel matrix with each pixel consisting of a preamplifier, a discriminator, and a TDC stage used to measure TOA and TOT. The precision clock distribution scheme will use a simple H-tree structure, to provide clock for sixteen full front-end chains, and will allow us to compare the clock distribution performance with simulation. Lastly, the TDC output data will feed into an E-link output port which includes a high speed serializer. It is anticipated that I2C will be used for the slow

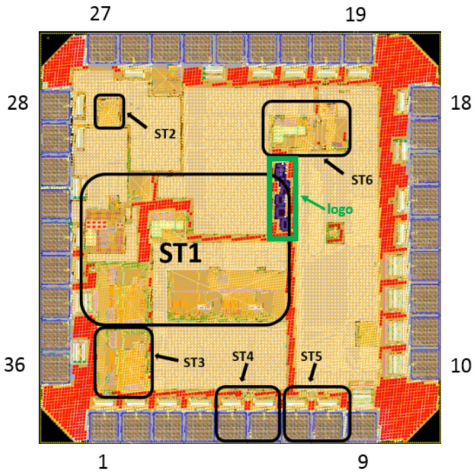


Figure 3.56: The ETROC0-V0 mini-ASIC chip (1 mm^2) submitted in December 2018, consisting of a single channel with a preamplifier, a discriminator with a threshold DAC, and a charge injection circuit. The rectangles labeled by ST indicate the various test structures implemented to study the performance of individual circuits within the mini-ASIC.

controls interface. In addition to the 4×4 pixel matrix, the prototype will have additional test circuits on two edges; this gives it a footprint that is compatible with a 5×5 pixel geometry being used in the similarly timed LGAD sensor prototypes. As of this writing, the ETROC1 design is being finalized with submission expected in August 2019.

3.3.13.3 ETROC2 prototype

The ETROC2 prototype builds on the ETROC1 design and includes all the remaining core functionalities, including the slow and fast control interfaces, PLL, phase shifter, and pixel readout. The digital pixel readout logic will be built around several static RAM buffers (or FIFOs) and is responsible for collecting the TDC output data, adding and checking bunch-crossing time stamps, and forming the output records. This readout logic should also be robust, accommodating bursts of Level-1 Accept control bits without data loss or corruption, while minimizing the overall latency.

The ETROC2 clock distribution network will be expanded to allow further validation of the clock tree simulation. Expansion to an 8×8 pixel matrix is the baseline plan, but using the full 16×16 pixel matrix size is also possible. Since ETROC2 will contain the full functionality of the final ETROC, making it full size could potentially allow ETROC2 to advance some of the ETROC3 testing, such as tests of larger scale clock distribution issues. This would have schedule benefits. However, the larger size increases the submission cost due to the larger mask size and would require several months of additional time for the final design verification and simulation. The decision on whether to use an 8×8 or 16×16 matrix for the ETROC2 will be made early in Q3 2020 with submission in Q4 2020 or Q1 2021 depending on that decision.

3.3.13.4 ETROC3 prototype and pre-production

The ETROC3 will be the full-size 16×16 pixel matrix, with the full clock tree distribution network. It will be the pre-production readout ASIC. Its submission is scheduled for Q1 2022.

3.4 ETL modules

3.4.1 Module design

The ETL modules are built from sub-assemblies containing a single sensor that is bump bonded to two ETROCs. Each sensor contains a 16×32 array of pads of size $1.3 \times 1.3\,\mathrm{mm}^2$, shown in Fig. 3.6. With the guard ring structures and bias ring, the dimensions of the sensor are $21.2 \times 42.0\,\mathrm{mm}^2$. The ETROCs, each of dimension $22.3 \times 20.8\,\mathrm{mm}^2$, are placed such that the short edge of each ETROC is oriented along the long edge of the sensor. The ETROCs extend over the long edge of the sensor, forming a "balcony" with wire-bond pads for the input and output signals and for power connections. Figure 3.57 shows the final assembled ETL modules, together with an exploded view of the module parts. In the majority of modules, two sub-assemblies are glued to an Aluminum Nitride (AlN) substrate, shown in Fig. 3.57 (left). This AlN baseplate provides a cooling path with a thermal expansion coefficient closely matched to that of silicon. A thermally conductive film is glued to the bottom side of the AlN baseplate. The lower surface of the film will not be sticky to allow replacement of modules after mounting on the cooling plate. Flex circuits laminated to each edge of the AlN substrate provide electrical connections to service hybrids that are described in Section 3.4.4. The sensor bias voltage is provided by wire bonds between pads on the flex circuit and the contact on top of the sensor. A second AlN plate is fixed atop this structure to protect the sensors. A U-shaped cutout in that AlN plate in the location where the bias-voltage wire bond is placed is visible in Fig. 3.57. The resulting "AlN sandwich" assemblies are rugged structures that are easy to handle during the assembly steps and installation. In addition to these two-sensor modules, a small number of single sensor modules, shown in Fig. 3.57 (right), will be assembled to increase coverage near the edges of the cooling disks.

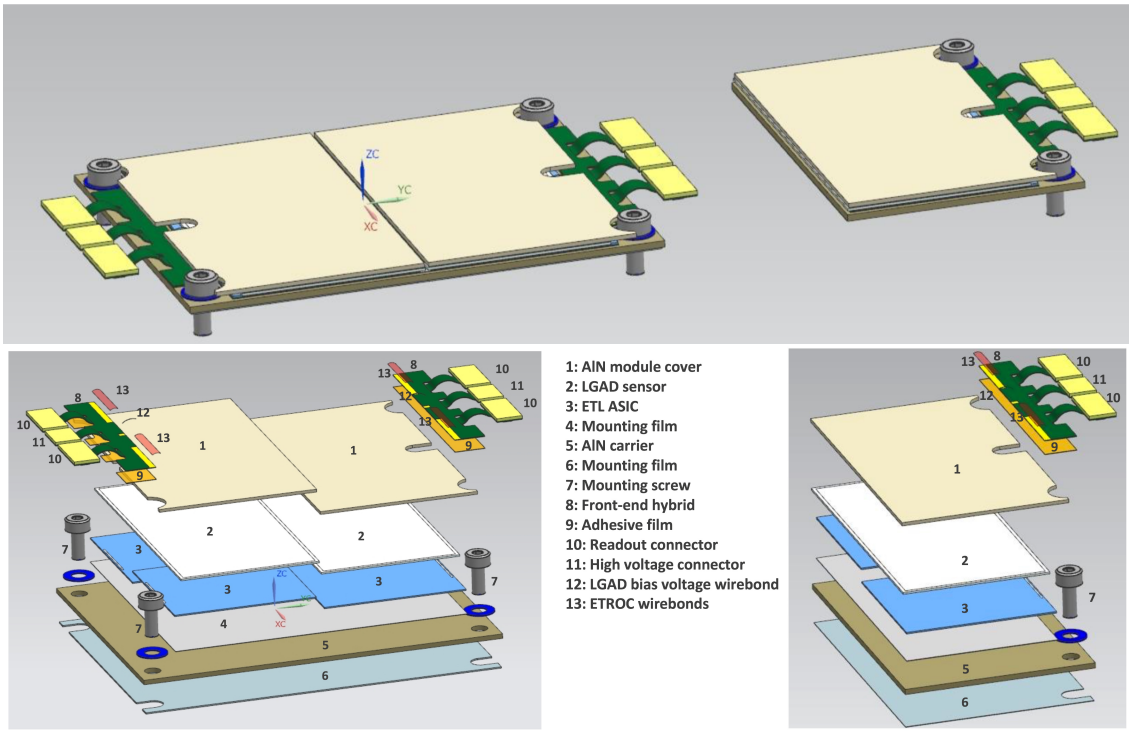


Figure 3.57: The two-sensor and one-sensor modules. Shown are views of the assembled modules (top), and the details of the module parts (bottom).

3.4.2 Module assembly

Modules will be assembled starting with the components discussed in Section 3.4.1. Module assembly will proceed in batches in a pipelined mode following five major steps:

- bump bond two ETROCs to a sensor to form a sub-assembly;
- laminate the flex circuit to the AlN base plate;
- laminate the thermal pads to the underside of the AlN base plate;
- attach the ETROC-side of two sensors to an AlN base plate with a thermal adhesive film;
- make electrical connections between the flex circuits and the ETROCs and sensors with a wire bonder;
- encapsulate the wire bonds; and
- attach an AlN cover plate to each sub-assembly to form a module.

Commercial vendors and CMS institutes are being evaluated as possible sources to perform the bump bonding of ETROCs to sensors. Sources under consideration have previous experience with high-yield bump bonding, including some having demonstrated these capabilities during previous construction of CMS. While the details of the bump-bonding process will likely vary somewhat between vendors, a brief illustration of the general steps is given.

Bare components, i.e. sensors and ETROCs, will be received at the bump-bonding site. Basic quality checks for these components are performed by the supplying sources. After reception and cataloging, a thin-film under-bump-metallization (UBM) is performed to ensure a good metal connection between the bump metal spheres and the bond pads on the chip wafer. After bump deposition (SnAg, Indium, etc.) and lifting of the photoresist, the wafers are cut and the singulated die reflowed in an oven to make spherical bumps, and the ETROCs are placed to the sensor by a thermocompression bonding process. Standard processes allow placement of the ETROCs relative to the sensor to micron precision. Given the relatively large size of the pads compared to those used in the tracker, studies are being performed to optimize the number of bonds per pad. Due to the relatively small density of bump bonds in ETL sub-assemblies, and driven by the requirement to keep sensor temperatures low, an epoxy-based underfill material will be deposited during the assembly process. The underfill will improve both the mechanical rigidity of the sub-assemblies, as well as the thermal conductivity. The thermal behavior of ETL modules after irradiation is described in more detail in Section 3.8.4. The resulting sub-assembly, comprised of a sensor and two ETROCs, will be qualified with a probe card, and a pull test will be performed on a small subset, before being shipped to the assembly sites.

Modules will be assembled at dedicated sites at collaborating institutions. Sites will receive sub-assemblies of a sensor bump bonded to two ETROCs from a vendor. Each sub-assembly will be visually inspected, including the wire-bond pads on each ETROC. These sites will also receive mechanical pieces including AlN plates with laminated flex cable pigtails. Each substrate will be inspected and its thickness and flatness measured. All components and any corresponding observations will be entered into the construction tracking database.

The first stage of mechanical assembly will be performed using a robotic gantry, as is being done for parts of the CMS tracker and CE assembly. Though the robotic assembly here is simpler than in those other cases, because of the relatively relaxed alignment tolerances and smaller number of components, the use of a robotic gantry allows consistent placement of components to a relative precision of $\mathcal{O}(10)\,\mu$m, ensuring better quality control over the roughly 9000 modules that need to be assembled. The ETROC-side of two sub-assemblies will be at-

tached to the AlN base plate with dimensions $56.5 \times 43.1\,\mathrm{mm}^2$ (for half-modules, the AlN plate is $28.25 \times 43.1\,\mathrm{mm}^2$), providing a thermal pathway and mechanical support. After the glue has cured, each module will be mechanically inspected before being transferred to the wire bonder.

A technician will make electrical connections from the ETROCs and sensors to the flex circuits using a wire bonding machine. After bonding, an electrical test using a dedicated test card will be performed on each module to confirm proper function. The bonds of each good module will be encapsulated and the electrical test repeated after allowing the encapsulant to cure.

In the second stage of mechanical assembly, an AlN cover plate is affixed to the (uncovered) sensor side of each module. This stage will be performed with a robotic gantry or a dedicated workstation with mechanical alignment capability. The cover plate results in rugged individual modules that can be carried around and handled without concern of damaging the most critical components. A small number of half-modules, containing only a single sub-assembly, will also be assembled following the same procedure.

Modules will be transferred to testing stations where a mechanical inspection and a complete electrical test will be performed as well as an overnight burn-in and thermal tests of each module. Extensive long-term stress testing will be performed on modules during prototyping and pre-production. Similar tests will be repeated on a subset of modules during production to monitor for thermal or aging issues.

3.4.3 Module prototyping

During the development phase, module prototypes will be built and their performance and robustness evaluated. This prototyping will include mechanical and thermal characterization of the individual module components, bench-top measurements and beam tests of their electrical performance, and finally system tests which examine all aspects of assembled, and irradiated, modules including their communication with the backend through prototype service hybrids. This development will use prototype modules of increasing complexity, starting from mechanical mockups and completing with system tests using fully functional modules with pre-production ETROC ASICs bump-bonded to pre-production LGAD sensors. The schedule for this prototyping is shown, with milestone dates listed, in Section 6.2.2.

Initially the focus of prototyping work is to validate and optimize the mechanical and thermal properties of modules, i.e., to validate their mechanical rigidity, bump-bonding scheme, module assembly sequence, and thermal performance of the CO_2 cooling. Full-size mechanical dummy modules made of blank silicon will be constructed to develop the assembly and wire-bonding procedures, and to assess the requirements on the component placement and alignment precision. The results of these tests with mechanical modules will be used to evaluate module design and assembly procedures, and these experiences will be used to modify the initial designs if necessary.

The thermal performance will be initially evaluated using resistive heating elements placed on a prototype section of an ETL wedge, and the heaters will emulate the heat sources present on the ETL modules and service hybrids. The section of the wedge will be cooled with dual-phase CO_2 similar to the conditions on CMS, and results will be compared with the FEA simulations. These prototyping tests will inform optimizations of the module and cooling designs to achieve the best ΔT between the coolant and sensor, in order to maximize thermal performance and uniformity of the temperature across the detector surface. After the initial thermal and mechanical optimization of a section of a wedge is completed, thermal and mechanical tests will be repeated on a full-sized wedge populated with dummy modules.

The reliability of the bump-bonding process, and the mechanical rigidity of bump-bonded modules will be evaluated using dummy full-sized modules. These assemblies will be fabricated out of silicon structures with the same geometries as the final modules, and will have daisy chain patterns to verify bump connectivity. Bump-bond quality will be tested by continuity measurements between pairs of wire-bond pads on the assembled modules using a probe-station. The option of using an underfill between the ETROC and LGADs will be studied with prototypes, which will be tested to evaluate the high voltage breakdown and thermal properties of underfill, spin-on passiviation, a polyimide spacer, and other solutions developed by similar projects. These tests will involve thermal and bias voltage cycling and irradiation studies.

Tests of fully functional modules will start with ETROC1 prototype ASICs bump-bonded to LGAD sensors. These modules will be constructed to verify the quality of the assembly, including sensor and bump-bonding yield, and their performance will be evaluated first with tests in lab settings and then in test beams. Modules of increasing complexity will be constructed following the ETROC2 and ETROC3 availability. Extensive test beam campaigns are planned after each version of ETROC and LGAD productions, to fully characterize the performance of the ETL modules constructed with these prototype ASICs and sensors. The quality of the bump-bonding of electrically functional prototype modules will be tested in order to qualify the vendors that will assemble production modules. Irradiated modules will be studied and characterized during each of these prototyping phases. Final verification of the assembly process will include thermal cycling and environmental testing of the pre-production module prototypes, with monitoring of the critical performance parameters during long-period burn-in tests.

This first bump-bonded prototype test, shown in Fig. 3.58, is composed of an interposer card with a bump-bonded sensor and a readout board using SKIROC2-CMS ASICs [107] from the CE prototype readout boards. This prototype allows the investigation of performance and assembly issues from the perspective of a nearly complete system with integrated readout, control, and powering. The interposer board will be used to characterize the bump bonding of large pads. It provides traces to the bonding pads for connecting to various compatible readout boards. For this first round of prototypes, we are using an earlier sensor layout, with $3 \times 1\,\mathrm{mm}^2$ pads in a 96-channel 4×24 array, to understand the challenges for bump bonding large sensor arrays with a relatively sparse bump matrix.

The layout of the readout card used is shown in Fig. 3.58. It is a modification of the CE "hexaboard" [7] and employs SKIROC2-CMS readout chips, which provide amplitude readout by ADC or TOT in a large dynamic range. Four chips provide simultaneous readout of 128 channels, all of which are available at the sensor bonding area via service pads, and are wire bonded to the readout traces on the interposer board. The SKIROC2-CMS chips can also be configured to route the analog signal from the pre-amplifier output to a separate probe pad for readout by dedicated timing electronics. Using all of these features, the gain stability and timing response of large LGAD sensors can be investigated in an integrated system. These cards also allow test stands to be set up across various participating institutes for performing long term tests of powered LGADs.

One potential approach for the interconnection of LGAD sensors and front-end ASICs is the use of gold-stud bumping. In this approach a gold ball is formed at the end of a thin gold wire via an electric discharge and is deposited as a cylindrically-shaped "stud" on the substrate by ultrasonic force, as shown in Fig. 3.59. The tool used in this study is capable of depositing around 20 bumps per second. For the present stage of prototyping, which involves varying sen-

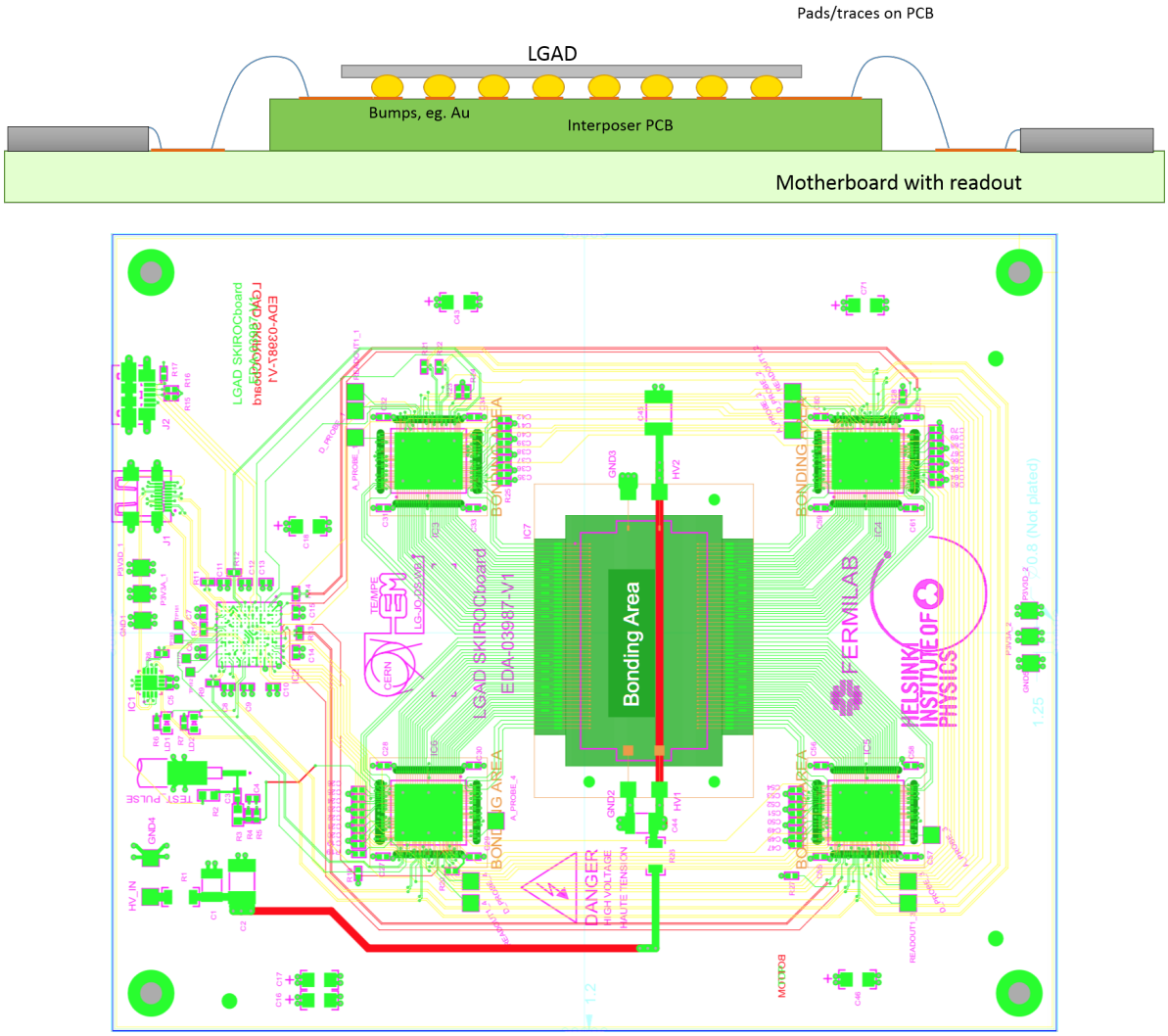


Figure 3.58: Top: A schematic drawing of the testing apparatus using the large-pad 96-channel LGAD sensors to test bump bonding and large-scale readout. The LGAD sensor is bump bonded to an interposer card, which in turn is wire bonded to a PCB with an integrated readout. Bottom: Layout of the prototype LGAD readout board. The LGAD sensor affixed in the highlighted bonding area in the center of the board is read out with four SKIROC2-CMS ASICs.

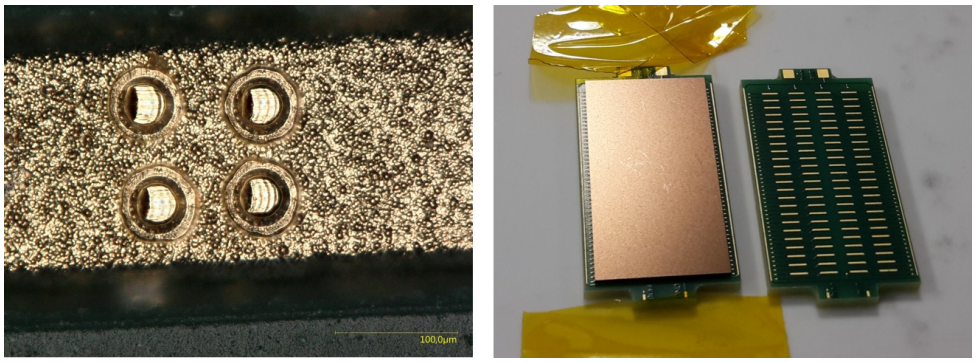


Figure 3.59: Left: Microscope image of four redundant gold bumps on the interposer PCB's metal pad. Right: Interposer PCBs with and without bump-bonded LGAD sensor.

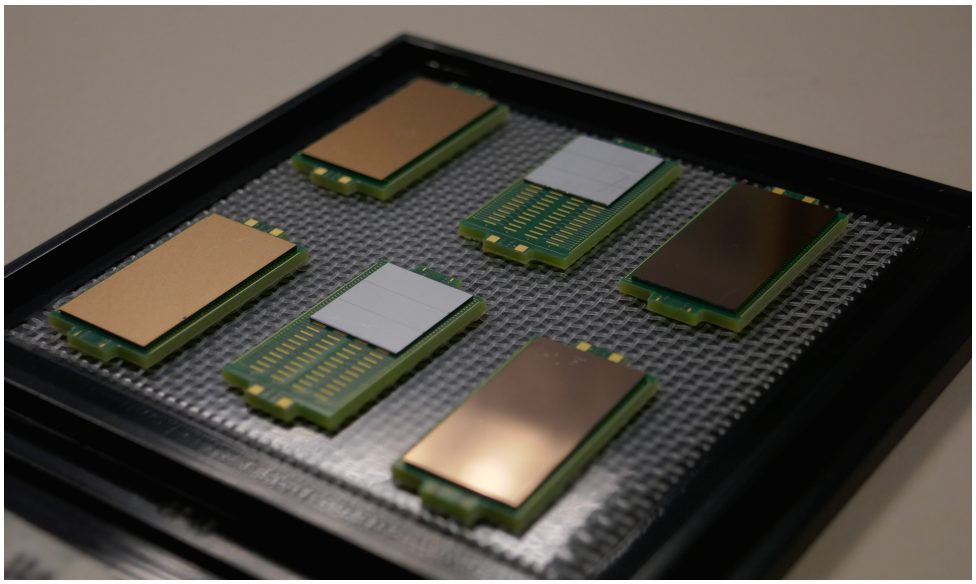


Figure 3.60: Multi-channel LGAD sensors bump bonded onto interposer PCBs.

sor layouts from several vendors, the gold-stud method provides considerable flexibility, since no additional under-bump metallization is required and the bumping pattern can be tuned according to the conditions of each sensor layout.

The performance of gold-stud bonding for a 96-channel LGAD sensor onto a PCB was evaluated with the interposer PCB that connects a sensor to a motherboard with readout electronics. Examples of bonded interposers are shown in Figs. 3.59 and 3.60. Depending on the surface layout of the sensor, two or four bumps per pad were used. Bumps were deposited on the corresponding metal pads of the interposer PCB in the exact same pattern, and the two components were then flip-chip bonded. Electrical connection from the sensor backside to the wire-bond pad on the interposer edge was verified by manual measurements. Measurements demonstrate that bonding of a large LGAD sensor to a PCB using gold studs is effective; mechanically stable assemblies with good electrical connection were produced, despite the relatively low number of bumps and the CTE mismatch of the components.

3.4.4 Service hybrids

The service hybrids, situated between rows of modules as shown in Fig. 3.61, provide power and readout services to the modules via flex circuit connectors. In particular, the boards perform the following functions: delivery of power to the ETROCs and the bias voltage to the sensors; delivery of the control and monitoring signals and the clock to the ETROCs; transfer of data from the ETROCs to the DAQ. Two types of service modules are used in the ETL, each servicing either 12 or 6 modules (i.e., 24 or 12 ETROCs).

A service hybrid is an assembly of two PCBs, a readout board and a power board, as shown in Fig. 3.62. The power board will reside on top of the readout board. The connection scheme and the number of socket connections per board will be studied and validated with mechancial prototypes before finalizing the design. The schematic of the functionalities of the power board is shown in Fig. 3.63 (top). Low voltage is delivered via the LV-PP0 connector on the patch panel, located on the top of the cooling disks, from power supplies. The low voltage is distributed through DC-DC converters, which regulate the output voltage and deliver the required power to the ETROCs, the slow control adapter (SCA) chip [108], lpGBT, and VTRx+. Both 2.5 and 1.2 V outputs are provided by the DC-DC converters. The DC-DC converters are based on

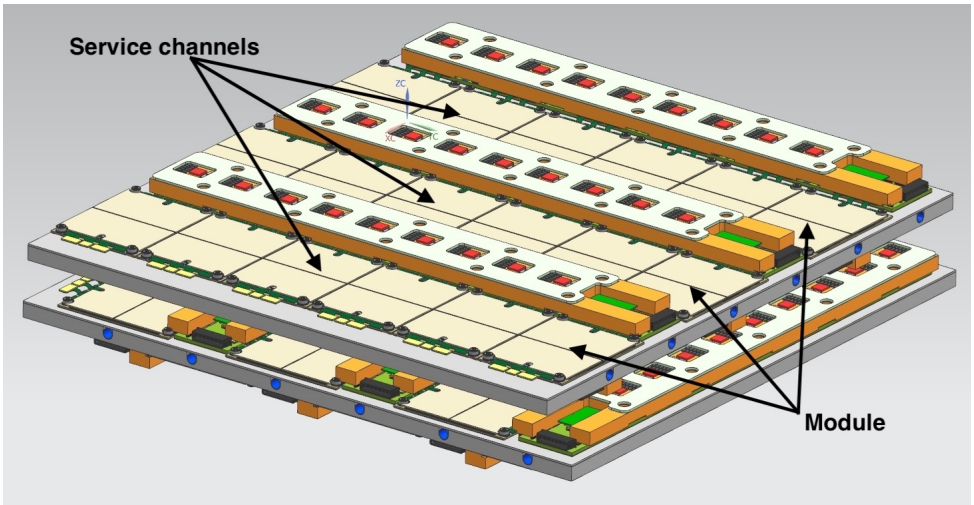


Figure 3.61: A drawing of the ETL support disks with modules and service hybrids (shown in orange). The channels for routing services, i.e., cables and fibers, over the top of the modules are indicated by black arrows.

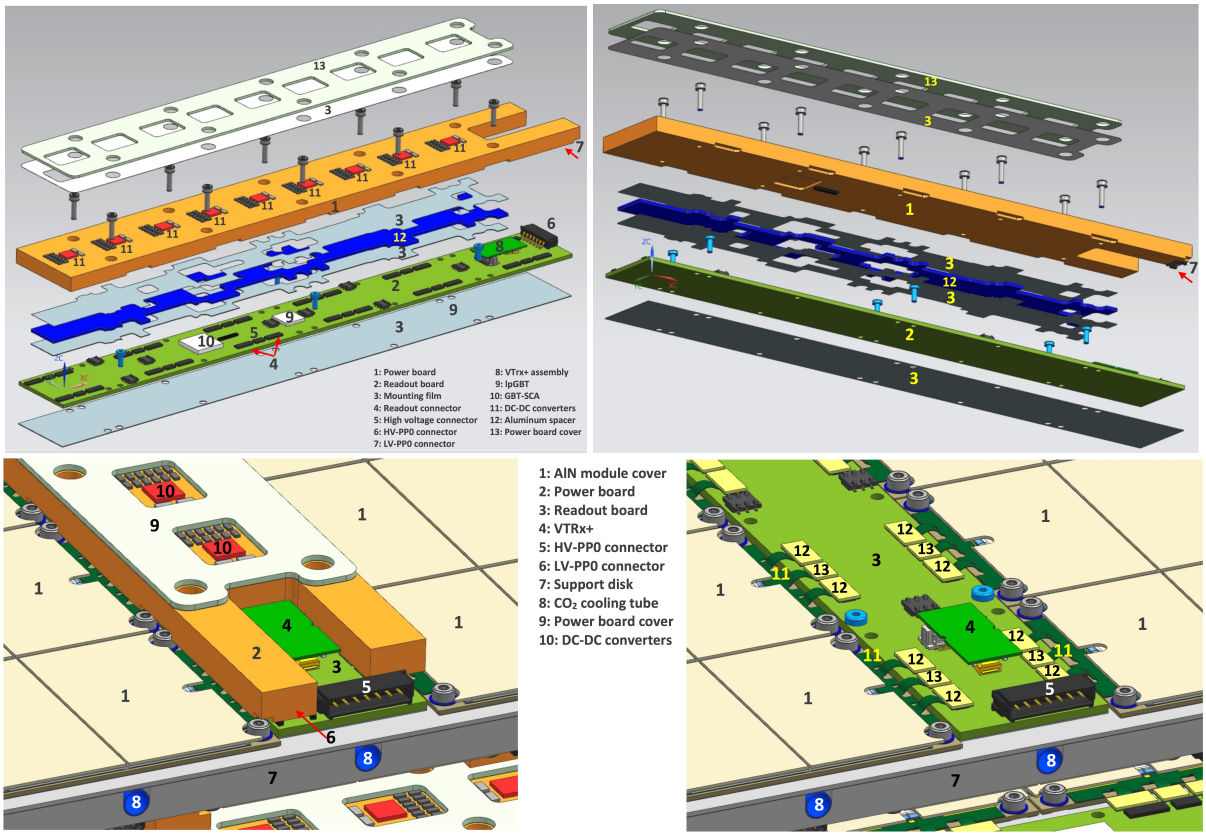


Figure 3.62: Top: Exploded views of the service hybrid showing its various components. The legend is shared among the two figures. Bottom: Service hybrid mounted between two modules, with the power board visible on top (left), and the service hybrid showing the readout board and components exposed (right).

the bPOL12V and bPOL2V5, radiation-hard and magnetic field tolerant converters developed for other detectors for CMS and ATLAS Phase-2 upgrades. Nine DC-DC converters would be placed on the power board that serves 12 modules, in order to provide the required cur-

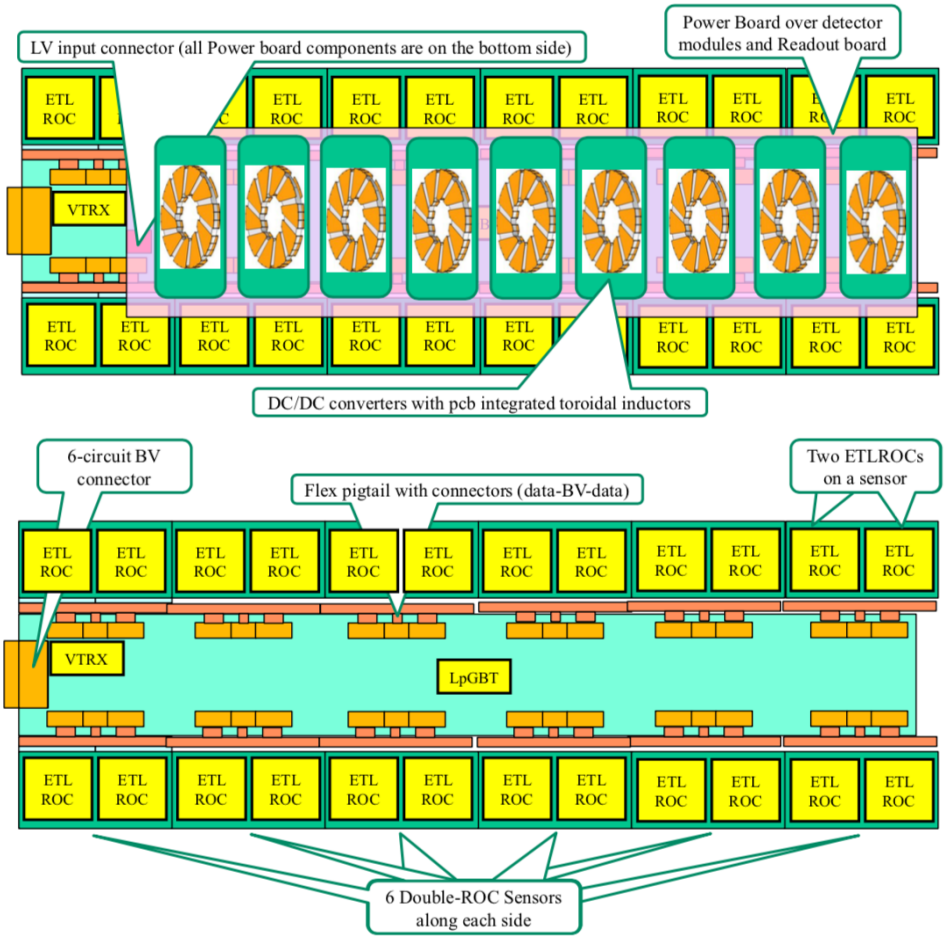


Figure 3.63: Schematic representation of the power (top) and readout (bottom) boards. As described in Section 3.5, a number of half-sized service hybrids are also used, these address and power 12 ETROCs with 5 DC/DC converters, but have the same I/O services.

rent, while power boards that serve six modules will have five DC-DC converters. The use of low-profile DC-DC converters, optimized for minimal thickness, allows placement of these components within the service hybrid space. More details on the power distribution are given in Section 3.8.1.

The schematic of the functionalities of the readout board is shown in Fig. 3.63 (bottom). The bias voltage (BV) is distributed to the LGAD sensors through the readout board via the HV-PP0 connector with a cable to the patch panel and off-detector power supplies. Since the total fluence, and thus the required bias voltage, varies strongly along the radius of the ETL disk, it is advantageous to be able to provide the bias voltage to modules with a finer granularity near the inner radius than at the outer radius. To achieve this, the distribution is designed to provide a separate bias voltage for each sensor, but it is expected that the same bias voltage will be supplied to groups of sensors at outer radii. The readout board also receives and distributes the fast control signals (clock, Level-1 Accept, reset, bunch crossing count) and the slow controls needed to configure the ETROC. It routes data and monitoring information from the ETROCs to the backend DAQ located in the service cavern. To perform these functions, each readout board includes a bidirectional optical link using an lpGBT coupled to an SCA. The SCA includes a multiplexed ADC with 31 inputs that is used to transmit the monitoring information of environmental parameters (temperatures, voltages, etc.).

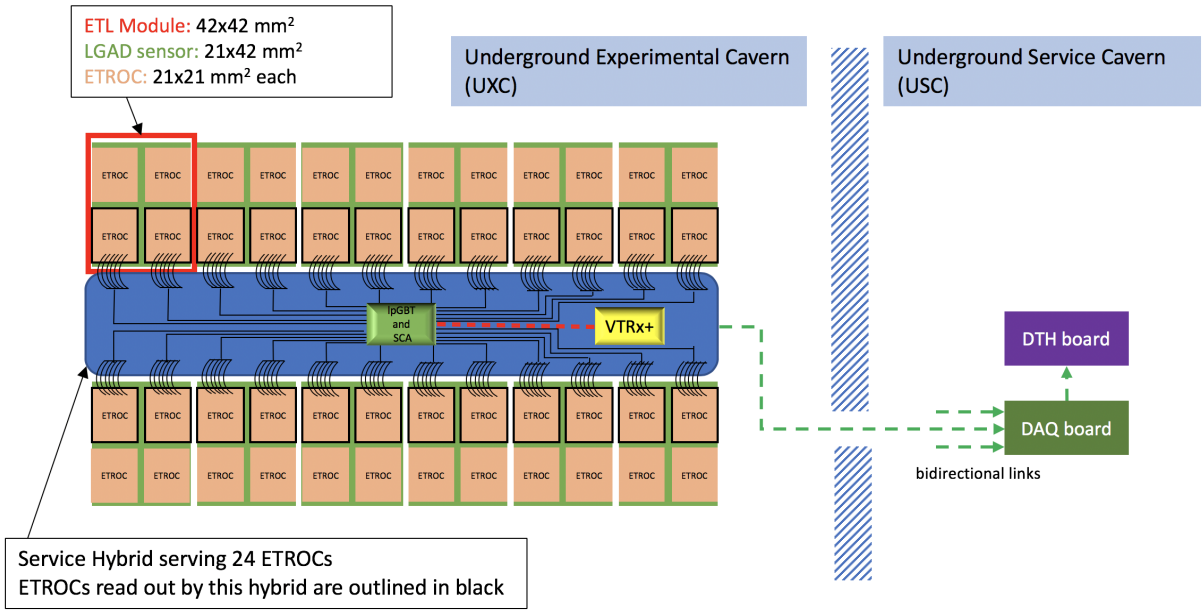


Figure 3.64: Overview of the ETL detector focusing on the data path. The diagram shows the signal flow from the LGAD sensors through the ETROC chips, the data transmission through the readout board of the service hybrid, the connection to the central DAQ via bidirectional links, and the connection to the DAQ and TTC Hub (DTH).

Each lpGBT is connected through E-link electrical lines and flex circuit connectors to the modules. Data to and from the back-end electronics are received and transmitted via optical fibers through the VTRx+ optoelectronics transceivers. The number of ETROCs per lpGBT is varied as a function of position on the cooling disk to maximize use of the lpGBT bandwidth, which will require two different size service hybrids. A detailed discussion of data rates used as an input for this optimization is provided in Section 3.5.

Prototyping of the service hybrids will start with simple non-functional versions used to test the mechanical and thermal performance together with similar module prototypes in 2019-2020. The first functional prototypes will be tested with ETROC1 based prototype modules in fall 2020, and used for early system tests. The second prototype, based on a fully-functional service hybrid design, is planned for fall 2021, where it will be used for full system tests using ETROC2 based modules. The schedule and milestones for this prototyping plan are shown in Section 6.2.2.

3.5 Data path and rates

The schematic of the data path in the ETL is shown in Fig. 3.64. Data paths are implemented over the lpGBT and VTRx+, which runs bi-directionally between the detector and the DAQ boards. The data from each ETROC is transmitted by E-links to lpGBT's mounted on the service hybrids and then sent serially over optical fibers by the VTRx+ chips to ATCA-based back-end electronics boards. The fast and slow controls plus clock are sent to the ETL via optical links. Figure 3.65 shows the block diagram of the data path architecture of one ETL module, including the power, control and monitoring connections. The lpGBT distributes fast control signals and drives the SCA, which is connected to one of the lpGBT E-links.

The 40 MHz LHC clock is distributed to all ETROC chips from the lpGBT using the dedicated

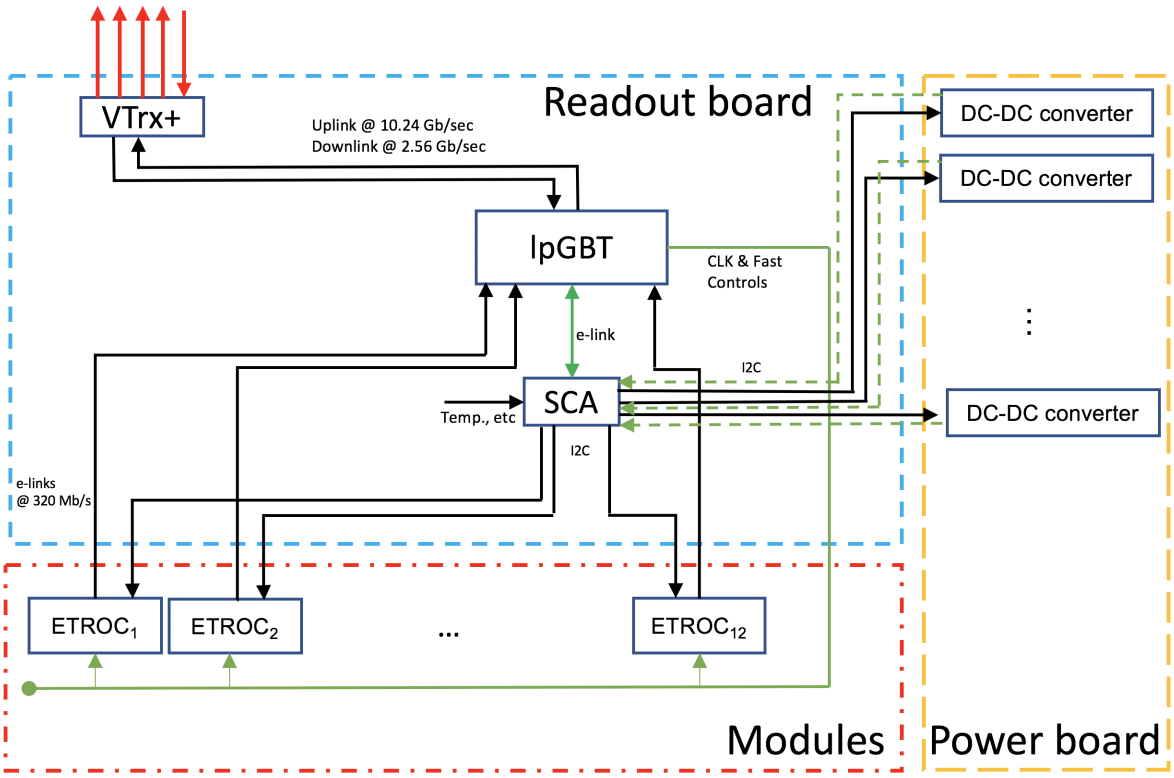


Figure 3.65: ETL readout architecture and slow controls. The 12 ETROCs at the bottom correspond to the modules connected on one half of the service hybrid, i.e., the lower half of Fig. 3.64. Another 12 ETROCs would be connected from the other half.

clock ePort, while the fast control commands are sent to all ETROC chips using the fast command ePort (at 320 Mbs), as shown in Fig. 3.51. The ETROC internal PLL will use the 40 MHz clock to generate a 320 MHz clock to be used to decode the fast commands from the fast control data packet, as shown in Fig. 3.47. All fast signals, both input and output, are implemented using a modified LVS standard called CERN Low Power Signaling, compatible with the lpGBT and described in Ref. [56]. Fast controls are distributed from the lpGBT using the multi-drop transmitter available on the 16 E-links, and every group of two ETROCs is connected to a common E-link.

The slow control and monitoring of the ETL front-end electronics is implemented using SCA features which can be accessed from the back-end via a reserved set of bits in the data frame. The ETROC ASICs utilize an I2C slave interface, which communicates with the master on the SCA ASIC through an I2C master bus. The 16 independent I2C-master serial bus channels on the SCA allow control and configuration of the front-end chips (ETROC and VTRx+) and enable the DC-DC converters.

The SCA performs the monitoring of the on-detector environmental conditions. It can perform analog measurements of the voltages, currents, and temperatures through a 12-bit Wilkinson type ADC that can be connected to one of 31 analogue inputs providing a voltage signal in the 0–1 V range. Inputs of the SCA can also be configured to source a current of 100 μA that can then be used to measure resistive inputs, such as an RTD to monitor temperature, or a humidity sensor.

The data are stored in the ETROC buffer waiting for the reception of a synchronous Level-1 Accept, for a maximum latency of 12.8 μs. Once the Level-1 Accept is received, a Fast Control sig-

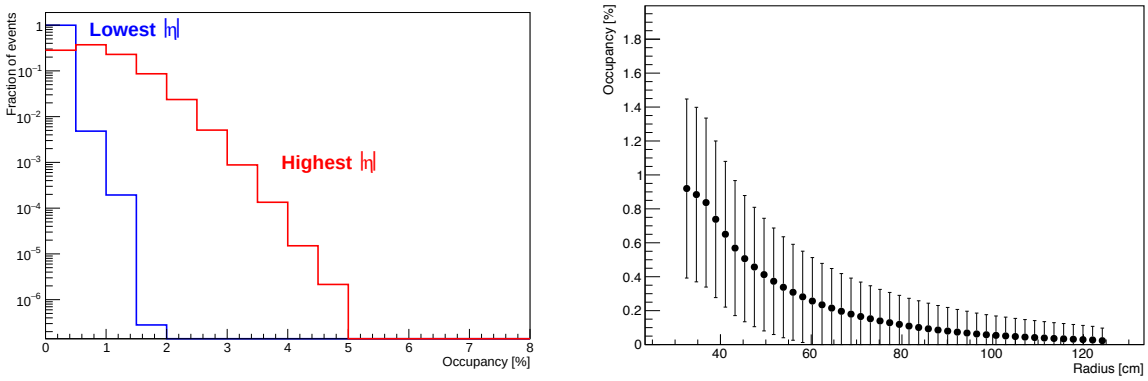


Figure 3.66: Left: Occupancy distributions in the innermost and outermost radius ETROCs. Right: Occupancy as a function of radius; the error bars show the rms.

nal is sent to all ETROCs connected to the lpGBT, and data are transmitted to the lpGBT via electrical links. Channels that do not have significant charge deposition will be zero-suppressed. Further details on the data architecture are presented in Section 3.3.

The data format for a single hit transmitted from the ASIC is described in Fig. 3.54. These data are transmitted over the output data E-link from each ASIC at up to 320 Mbs. They are routed on the service hybrids to an lpGBT where up to 24 E-links are combined into a single uplink at either 5.12 or 10.24 Gbs, as shown schematically in Fig. 3.65.

A study of the occupancy of each ASIC has been performed to determine the maximum number of ASICs per lpGBT compatible with the individual E-link capacity and the combined lpGBT capacity. This study used 200 pileup interactions merged with a $t\bar{t}$ event. The $t\bar{t}$ event is included to emulate the potentially higher occupancy in L1-triggered events, but it is found to make a negligble contribution to the overall occupancy. Figure 3.66 shows the distributions of the hit occupancy for ETROCs at the highest and lowest rapidity (on the left), and the hit occupancy as a function of ETL radius (on the right).

The lpGBT bandwidth capacity per E-link and the maximum number of E-links can be selected to be 160/320/640 Mbs and 28/14/7 E-links per lpGBT, respectively, when using FEC5 error correction in 5.12 Gbs low-power mode. A throughput of 160 Mbs for one ETROC corresponds to an average of 5 hits per event, which is a 2% occupancy. For all but the highest η region, the average occupancies are well below that limit, which allows a configuration with up to the maximum number of 28 ETROCs per lpGBT in the majority of the detector regions. For those regions of the detector, we use 24 ETROCs, as was shown schematically in Fig. 3.64. At higher η, the occupancies are closer to the 160 Mbs limit, and the number of ETROCs per lpGBT is reduced to 12 by using half-sized service hybrids. These smaller boards service up to 6 sensors, rather than 12, and have 5 DC-DC converters rather than the 9 for the full-sized boards that were shown in Fig. 3.63. The placement of the full- and half-sized service hybrids on the cooling disk is shown in Fig. 3.67.

The readout data rate per lpGBT, with this combination of either 12 or 24 ETROCs, is shown in Fig. 3.68. For all cases, the data rate is well within the 5.12 Gbs bandwidth capacity in low power mode. The number of lpGBTs needed per wedge is 100. All wedges are identical, apart from rotations, so a total of 1600 lpGBTs are needed for the 4 wedges per disk, and 2 disks per z-side. These calculations leave headroom in case of higher occupancies, particularly given the option of using the higher power 10.24 Gbs mode for the highest η regions.

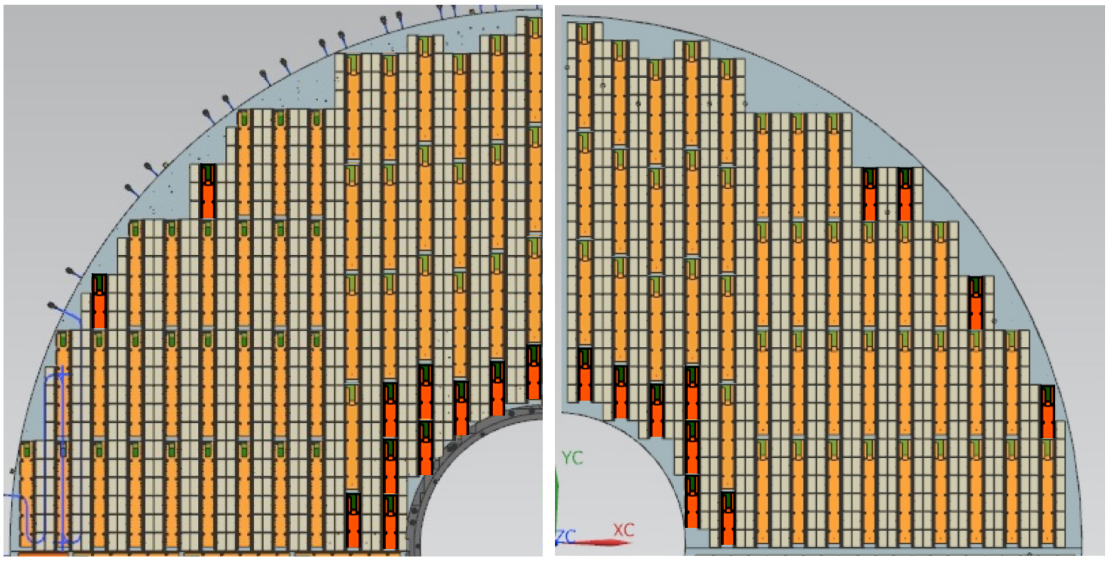


Figure 3.67: Layout of the modules and service hybrids on an ETL wedge, showing the placement on the front and back faces (right and left images), where the left image is viewed facing toward the IP and the right image is viewed facing away from the IP. The full-sized service hybrids are shown in orange and the half-sized service hybrids are shown in red. The LGAD sensors, shown as the gray rectangles, are assembled as either 1-sensor or 2-sensor modules, as described in Section 3.4. Examples of 1-sensor modules are seen at the edges of the front face, while 2-sensor modules are used to cover the central region.

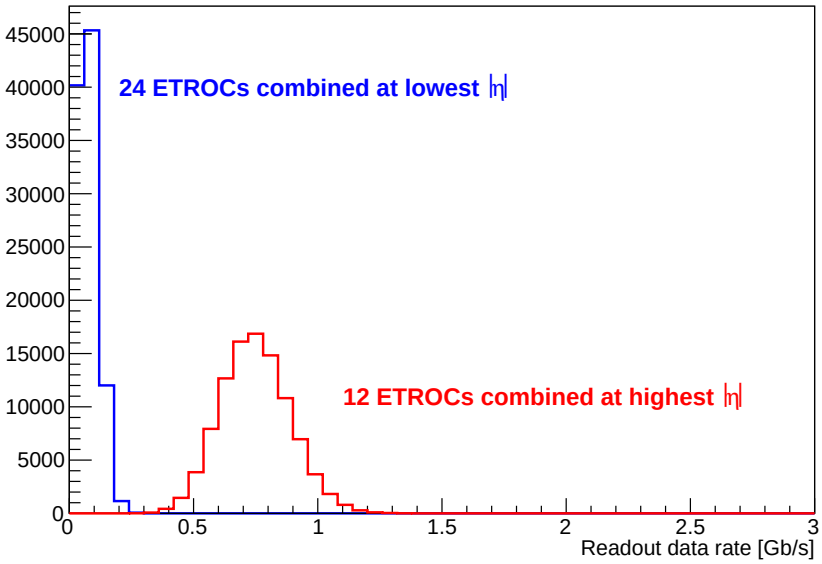


Figure 3.68: Distributions of readout data rate obtained by merging of the innermost 12 E-links and the outermost 24 E-links.

The number of each type of component, from LGAD sensors to lpGBTs, is listed in Table 3.8.

Table 3.8: List of the number of each type of component used in the full ETL detector.

Component type	Number per wedge	Total number
LGAD sensors	1039	16 624
ETROCs	2078	33 248
Readout channels	531 968	8 511 488
1-sensor modules	63	1008
2-sensor modules	488	7808
Full-size service hybrids	71	1136
Half-size service hybrids	29	464
lpGBTs, VTRx, SCA	100	1600
DC-DC converters	820	13 112

3.6 Slow control and monitoring

The slow control and monitoring of the ETL detector are performed using the capabilities of the CERN SCA chip. Each readout board is equipped with an SCA chip that is connected to an lpGBT via a dedicated E-link. The SCA is powered before the rest of the ETL front-end electronics and sensors, and controls enabling of each DC-DC converter, provides environmental monitoring information for safe operation of the detector, and downloads the configuration information to the frontend electronics (ETROC and VTRx+).

The SCA can measure up to 31 analog inputs using a 10-bit ADC. The inputs in ETL are:

- local temperatures and humidity,
- supply voltage and currents, and
- status bits from the DC-DC converters on the power board.

The SCA based monitoring system is not designed to perform time-critical operations or to be part of the safety system, but can perform elementary operations such as reading a voltage, sensing a temperature, or turning off a regulator in a time of about 1 ms. The detector control and safety system is described in detail in Section 4.4.

3.7 Mechanical engineering, integration, and installation

A mechanical drawing of the ETL detector is shown in Fig. 3.2. The support structures for the detector modules and service hybrids are aluminum disks containing embedded cooling tubes to evacuate the heat generated by the sensors and readout electronics. Each endcap of the ETL detector is comprised of two disks, and both faces of each disk are populated with active elements. Details of the structural design can be found in Section 3.7.1.

ETL modules and service hybrids will be mounted on the disks at P5. Once commissioned, the disks will be lowered to the experimental cavern for installation on the front of the CE. The disks will be mounted on the endcap inner support tube shared with the CE, but in an independent cold volume with a thermal screen that can be retracted to allow service and maintenance of the ETL detector. The integration and installation are described in Sections 3.7.2 and 3.7.3, respectively.

3.7.1 Structural design

The ETL detector is assembled from disks, as shown in Fig. 3.69 (left). Each disk serves as a mechanical support as well as a cooling plate. The disks are made of $\frac{1}{4}$-inch MIC6 aluminum,

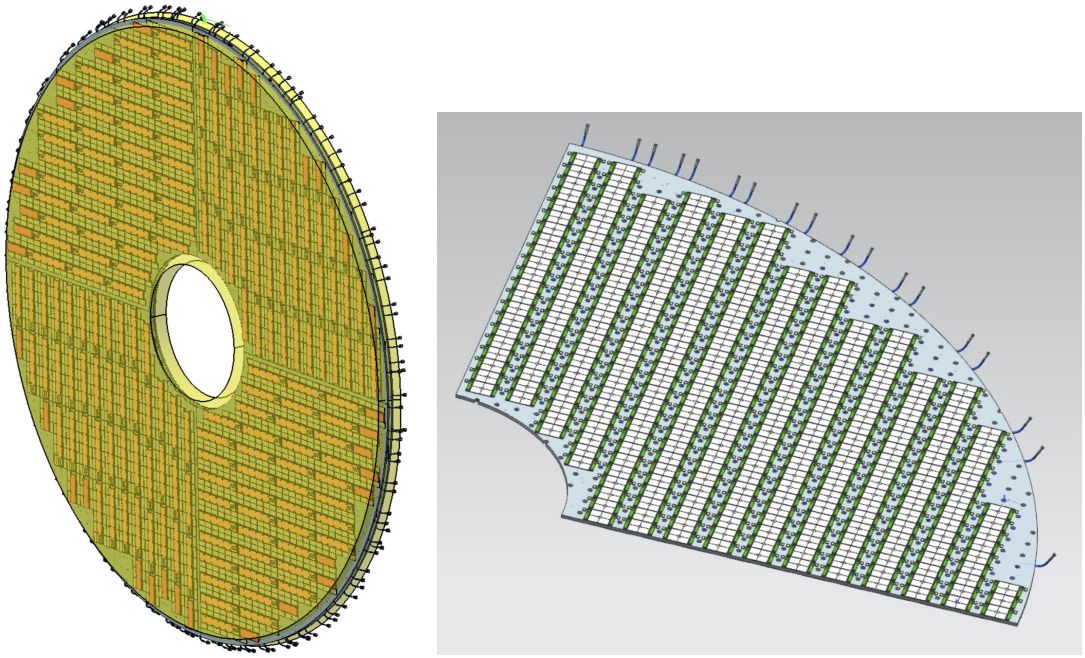


Figure 3.69: Left: One disk of the ETL detector with the layout of the modules and service hybrids. Right: A 90-degree section (wedge) of an ETL disk with cooling inlets and outlets shown at the outer rim.

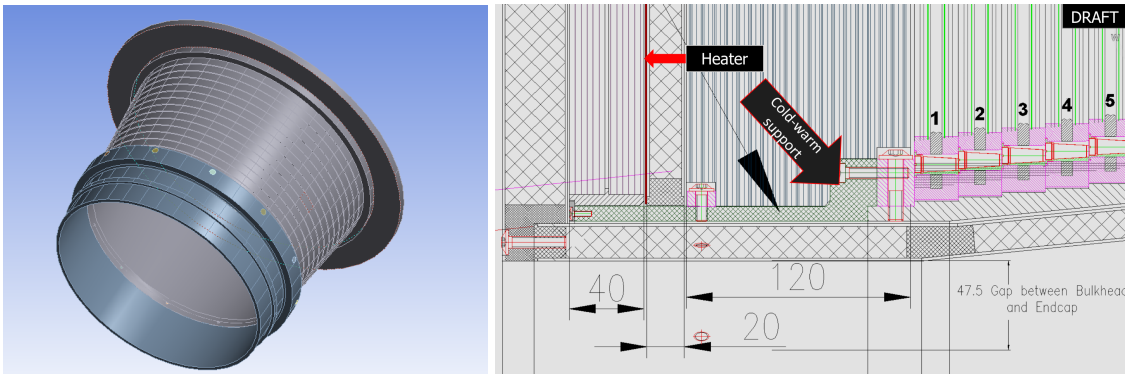


Figure 3.70: Left: Schematic representation of the inner support tube. Right: The interface between the ETL and CE detectors showing the cold-warm ETL support.

in which a small diameter stainless steel tube carries dual-phase CO_2 to maintain the sensor temperature below $-25\,^{\circ}\mathrm{C}$. A quarter of a cooling disk, referred to as a wedge, is shown in Fig. 3.69 (right). Wedges are machined to permit easy integration of wedges into half-disks, while minimizing the region uncovered by sensors to maximize particle acceptance. Wedges are also machined with holes for the placement of screws or pins used for the mounting of modules and service hybrids.

The inner support tube, shown in Fig. 3.70 (left), is made of a permaglas section supporting the ETL detector, and an aluminum section supporting the electromagnetic sub-detector of the CE. It is located at the inner region of the detector, closest to the beampipe. The tube thermally connects the ETL and CE cold volumes, as shown in Fig. 3.70 (right). The use of permaglas minimizes the heat transfer between the volumes, facilitating maintenance of the ETL detector while maintaining the $-30\,^{\circ}\mathrm{C}$ operating temperature of the CE.

Finite Element Analysis (FEA) studies were performed to evaluate the heat transfer between

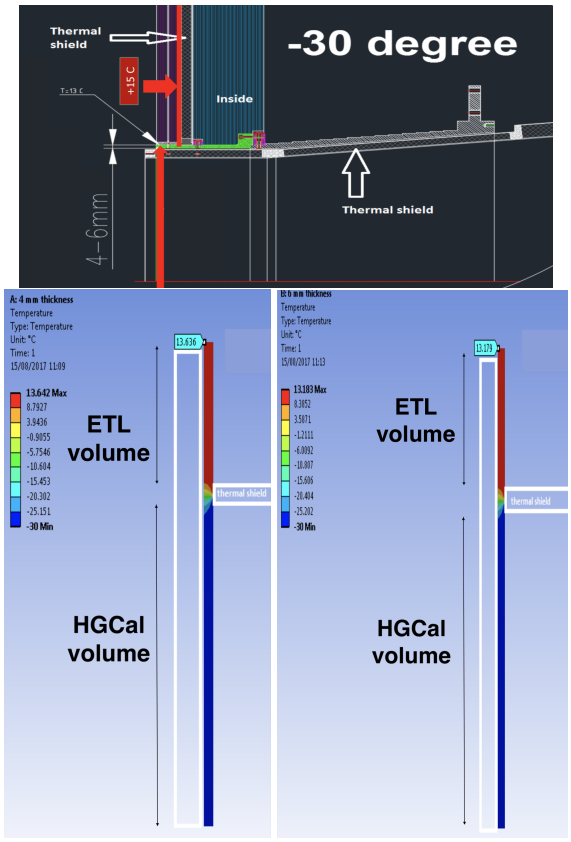


Figure 3.71: Top: Cross-sectional view of the permaglas inner support tube and the temperature profile on both sides of the thermal screen. For this test of the thermal screen, the ETL detector is assumed to be powered off and at +15 °C, while CE is at its operating temperature of −30 °C. Bottom: The FEA solutions for 4 mm (left) and 6 mm (right) thick permaglas supports.

the ETL and CE cold volumes through the permaglas support. The ETL detector was modeled at a static air temperature of 15 °C, above the cavern dew point, and the CE was modeled at its operating temperature of −30 °C. The analysis was performed for a 4 or 6 mm thick permaglas structure. Figure 3.71 shows that the thermal transfer from the warm ETL detector to the CE cold volume is negligible and largely independent of the thickness of the permaglas support structure.

Preliminary estimates of the mechanical rigidity of the permaglas inner support tube were calculated to understand potential limitations of the design. For these studies, a permaglas thickness of 6 mm was assumed at a load of 1 tonne, including the ETL detector and the neutron moderator. The total weight of the ETL detector per side is 282 kg, while the neutron moderator weighs about 600 kg. The permaglas is assumed to have a Young's modulus of 18.6 GPa, a Poisson ratio of 0.15, and a yield strength of 517 MPa. Figure 3.72 shows that the largest expected deformation of the inner support tube under these conditions is an acceptable 1.4 mm.

A thermal screen encloses the ETL detector to separate the cold (−30 °C) dry (dew point −40 °C) detector volume from the ambient conditions in the cavern. The thermal screen is made from panels consisting of a very low thermal conductivity aerogel core between 1.5 mm thick permaglas skins providing a rigid structure with low thermal conductivity.

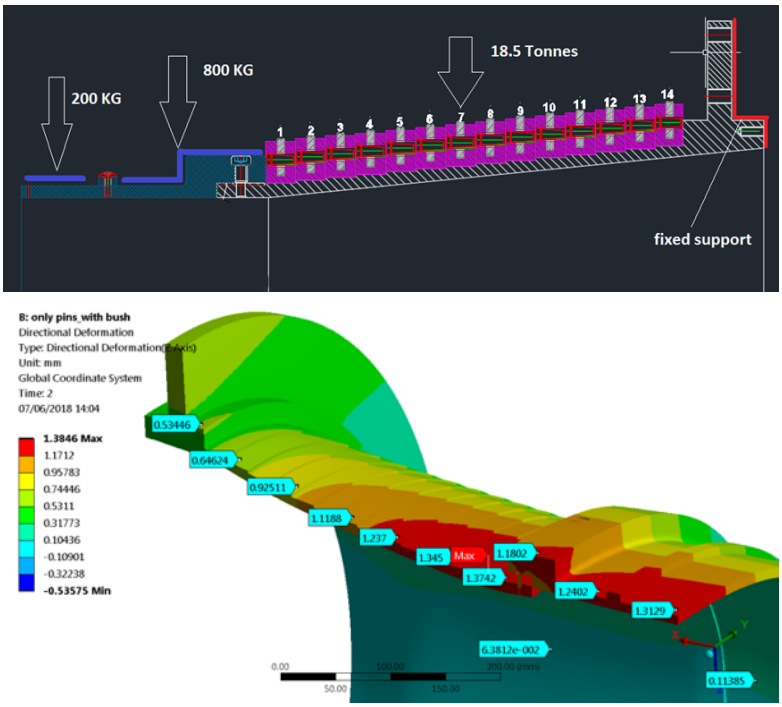


Figure 3.72: Top: Loads on inner support tube used to calculate the expected deformation of the tube. Bottom: The result of an FEA calculation of the stresses in the inner support tube with the ETL detector, neutron moderator, and electromagnetic sub-detector of the CE mounted on it. The maximum observed deformation is about 1.4 mm.

3.7.2 Integration of modules and service hybrids

ETL modules and service hybrids will be received from the assembly sites, re-tested and entered in the construction tracking database, and then mounted on the wedges, as shown in Fig. 3.69 (right). Components will be mounted using screws or pins inserted into mounting holes shown in Figs. 3.57 and 3.62. As components are mounted and connected, a quick warm-temperature test will be performed to verify the electrical connectivity between each service hybrid and its adjacent modules. Once modules and service hybrids are mounted on both sides of a wedge, it will be connected to cooling and power services and a DAQ test stand and a longer-term, cold-temperature commissioning of the integrated wedge will be performed.

Only one wedge type is needed, which simplifies the design and construction of the wedges, and its 90-degree rotations cover the full azimuth, as shown on the right side of Fig. 3.69. The rows of modules run perpendicular to each other on the rear face of disk 1 and the forward face of disk 2, as shown in Fig. 3.73. This configuration was chosen to ease the routing of the cooling services, which do not have inlets and outlets evenly spaced in ϕ; the 90-degree rotated wedges distribute them evenly, avoiding congestion at around 3 and 9 o'clock positions, where the density of services is the highest.

Figure 3.74 shows a cross-sectional view of an ETL wedge with modules and service hybrids mounted. The dimensions of individual components are provided in the figure and summarized in Table 3.9. Cabling and optical fibers will be routed on top of the module cover plates, outward on each wedge, in the z-space between the tops of the modules and the service hybrids. The maximum extent in z, between the top of the service hybrids on each side of an ETL wedge, is 18.9 ± 0.02 mm. For the two disks in an endcap, this requires 37.9 mm, which is within the 45 mm space available.

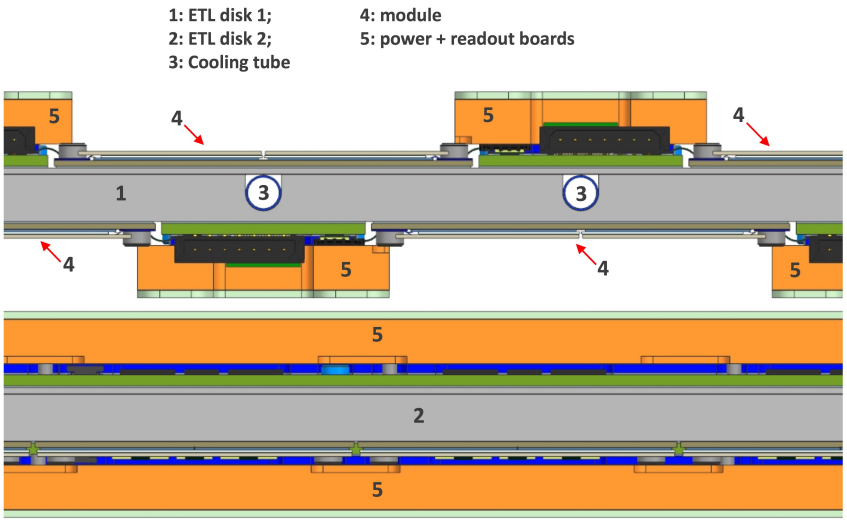


Figure 3.73: A side view of the ETL modules and service hybrids illustrating how they are mounted onto either side of the two aluminum disks.

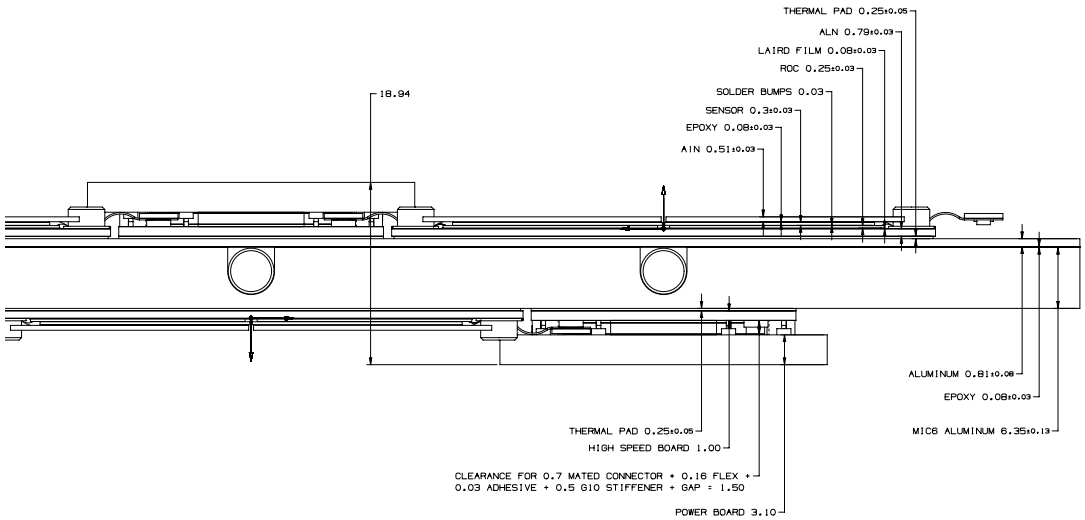


Figure 3.74: A cross-sectional view of an ETL wedge with modules and service hybrids mounted, showing heights of individual components.

3.7.3 Installation and commissioning

The 16 ETL wedges will be assembled on the surface at P5, and completely tested before installation on the front of the CE. Installation of the wedges is planned to take place in the underground collision hall, UXC, after the completion of the installation of the CE endcaps. A scenario in which the wedges are installed on the CE endcaps on the surface is also possible.

ETL services located in UXC will be installed and tested for connectivity and limited functionality ahead of lowering the CE endcaps and ETL wedges to the UXC. They will be mounted on the YE1 structure and its balconies. Sections of service channels containing electrical and optical cables and cooling transfer lines will be mounted outside the cold volume of the CE and attached to the thermal screen, prior to lowering the CE endcaps to UXC.

Wedges are mated to flanges on the inner support tube, as shown in Fig. 3.75. The z position of the wedges is constrained by mounting to fixtures on the neutron moderator, as shown in

Table 3.9: Summary of the z-dimensions (thickness or height) of an ETL wedge with modules and service hybrids mounted as shown in Fig. 3.74.

	Component	Thickness (mm)	Total (mm)
disk	Al	0.81 ± 0.08	7.24 ± 0.16
	epoxy	0.08 ± 0.03	
	MIC6 Al	6.35 ± 0.13	
hybrid	thermal pad	0.25 ± 0.05	5.85 ± 0.05
	readout board	1	
	flex/connector/stiffener	1.5	
	power board	3.1	
module	thermal pad	0.25 ± 0.05	2.29 ± 0.09
	base AlN+epoxy	0.87 ± 0.04	
	LGAD+ROCs	0.58 ± 0.04	
	cover AlN+epoxy	0.59 ± 0.04	

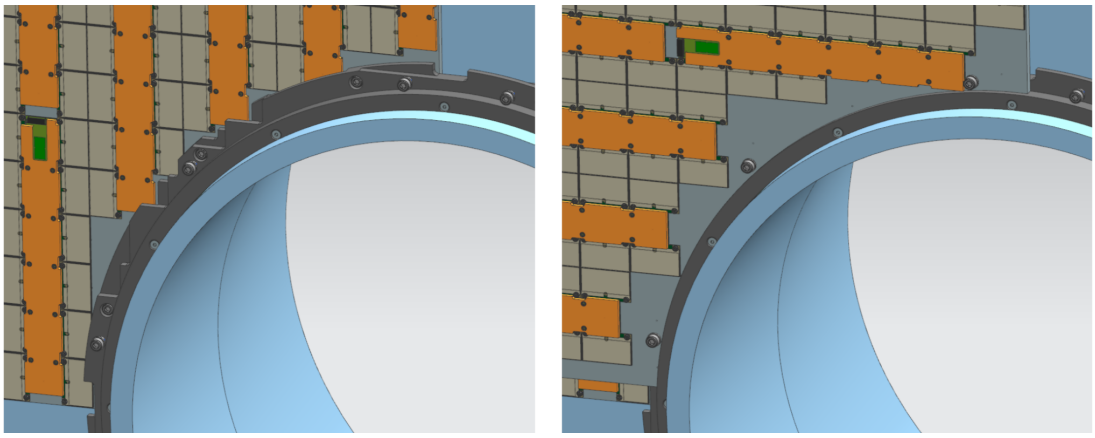


Figure 3.75: Mounting scheme of the ETL detector, showing the rear (left), closest to the CE, and front (right) wedges.

Fig. 3.76. The interfaces to the ETL services, including cooling manifolds and electrial and optical patch panels, are installed on the outer rim of each disk. The optical fibers and high and low voltage cables are routed from the service hybrids, along the tops of the modules, to the patch-panels.

3.8 Services

Services required for the operation of the ETL detector include: low-voltage (LV) for electronics; bias-voltage (BV) to bias the sensors; optical fibers to transmit data, control, monitoring, and trigger information; CO_2 cooling; a dry gas system; and cables to transmit information for the detector safety system (DSS). The cooling and power cables are the dominant contributions to the space required. Estimates of the total cross-sections required for the passage of these services are shown in Table 3.10.

As shown in Fig. 3.77 (left), the services are routed to the detector from the YE1 balconies, under the muon chambers, to the back section of the CE, where patch panels (PP1) are located. They are then routed to the periphery of the ETL detector along the outside of the CE in channels located at the 3 and 9 o'clock positions. The most constrained location for the service routing is at the point where the services pass through the bracket supporting the tracker and

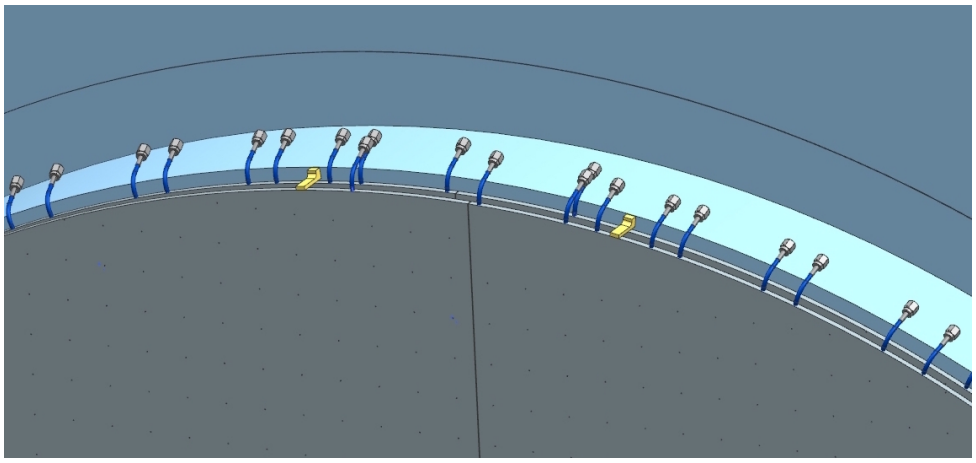


Figure 3.76: Schematic representation of the mounting of wedges to fixtures on the neutron moderator that constrain the z position (yellow brackets). Also shown are the supply and return tubes of the cooling loops.

Table 3.10: Estimated cross-sectional area required to route services to the ETL detector.

Component	Area per detector-end (cm^2)
CO_2 coaxial transfer lines	120
Low-voltage cabling	75
Bias-voltage cabling	90
Optical fibers	15
Dry gas tubes	8
Safety system and environmental sensors	45

turn radially into the ETL volume. This is shown in Fig. 3.77 (right). The design shown for the tracker bracket, developed by ETL and tracker engineers, provides additional space for the services consistent with the tracker requirements. The services enter the ETL cold volume through a feedthrough made from custom delrin combs with silicone or epoxy used for sealing. An engineering drawing of their design is shown in Fig. 3.78 along with the plan for packing the individual services in the channel. Inside the ETL cold volume, the services will pass azimuthally around an annular region, as shown in Fig. 3.77 (right). The cooling will connect to manifolds with capillaries feeding the individual cooling channels, and the cables and fibers will connect to patch panels situated at the periphery of the cooling plates.

The feedthroughs and the routing of services through the tracker bracket will be validated in a mockup test, which will be carried out together with the tracker project in 2019–2020. The patch panels and thermal screen will be prototyped, and the CO_2 cooling channel performance will be tested with prototype wedges, in mid 2020. These tests will inform the engineering of the final feedthroughs and service routing. The set of milestones for this prototyping campaign are listed in Section 6.2.

3.8.1 DC low voltage power distribution

The power required by the ETL front-end electronics is summarized in Table 3.11. The sensor power is based on estimates of the end-of-life bias voltage and current. The ETROC requirement is estimated from a sum of components including the discriminator, pre-amplifier, TDC, and digital components. Further details are presented in Section 3.3. Estimates of the power consumption of the remainder of the active components of the ETL FE electronics, such as the lpGBT, are taken from the outer tracker [4].

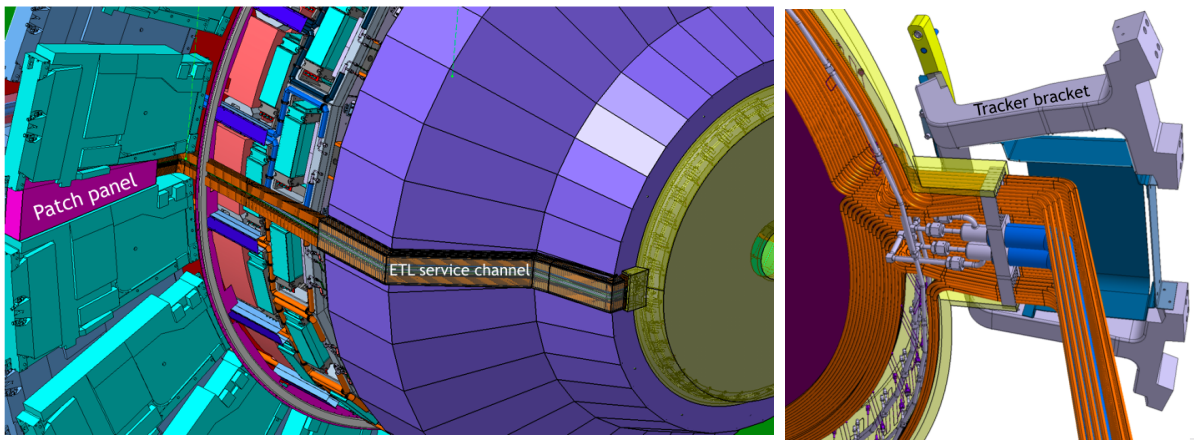


Figure 3.77: Left: Engineering drawing showing the layout of the ETL services path across YE1 and the outside of the CE. Right: Engineering drawing of the services routing through the tracker bracket, which is the most constraining region. These views show only one side, at the 3 o'clock position; another is located at the 9 o'clock position.

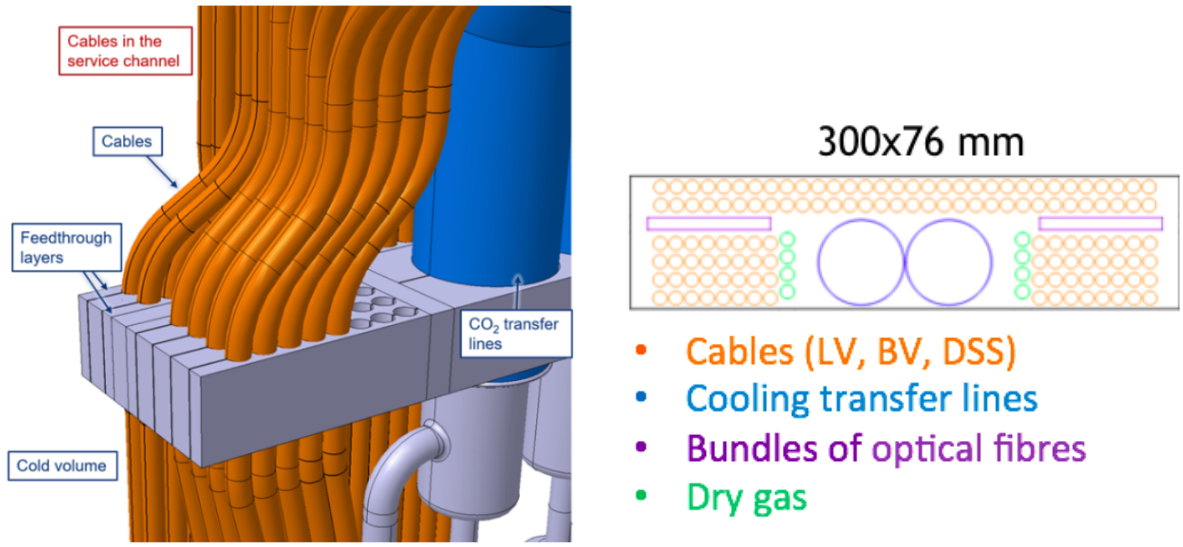


Figure 3.78: Left: Engineering drawing showing the delrin combs used for feedthrough of the services into the ETL cold volume. Right: The packing layout of the cables, pipes, and fibers in the service channel.

Primary power will be supplied from the CMS electrical grid to the experimental cavern via cabling passing through the main endcap cable chains under the cavern floor and into racks located on the endcap towers. These racks will contain crates of power supplies that will provide DC power to the downstream loads. Each end of the ETL detector will require a total current of approximately 4 kA. This current will be supplied via a system of heavy-gauge conductors running between the power supply racks and along the CE to the feedthroughs on the ETL detector located at the 3 and 9 o'clock positions of the ETL thermal screen. The total voltage drop across this system of conductors will be limited to 2 V, resulting in 8 kW of power dissipation along the delivery path. The total copper cross-section required to achieve a 2 V maximum voltage drop is approximately 2000 mm^2, where the thicker sections span the distance between the power racks and PP1, while the cable thickness between PP1 and PP0 are listed in Table 3.10. A schematic representation of the geometry of services and grounding is shown in Fig. 3.79.

Table 3.11: Estimated power consumption, per endcap, for the ETL detector at end-of-life after 3000 fb^{-1}, where the only significant radiation damage effect is on the increased leakage current and thus increased induced heat load in the sensors.

Component	Power (kW)
Sensor	0.8
ETROC	12.5
lpGBT	0.6
VTRx+	0.3
DC-DC	7.5
GBT-SCA	0.2
Power cables	2.7
Heating foils	1.0
Total	26

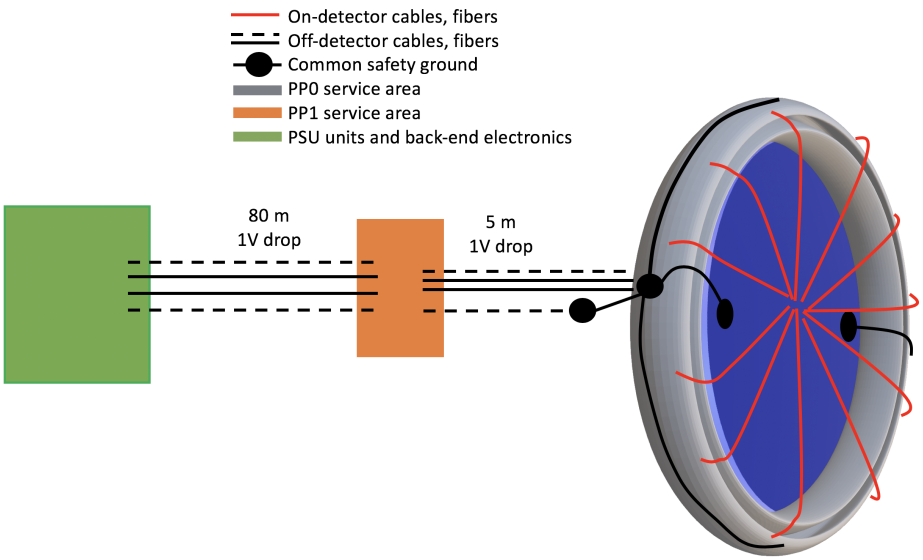


Figure 3.79: A schematic representation of the services and grounding of the ETL detector. All electrical services are floating at the PS end and grounded to the support disks. Cooling tubes have insulating inserts. Al wedges and cable shields are connected to a common ground.

3.8.2 Bias voltage system

To ensure that the LGAD sensors are biased to the optimal gain value, the bias-voltage power supplies, located in the service cavern, need to be able to provide up to 700 V. The reduction in sensor gain from radiation effects depends strongly on the radial distance from the beam. If it is necessary to be able to control the bias voltage (BV) separately for each module, of which there are 9000 in the ETL detector, 18 000 BV conductors will be required. This worst-case scenario was used to estimate the area required to route the BV services shown in Table 3.10. The granularity with which the BV is delivered to the sensors carries with it cost implications for the power supplies as well as service integration issues arising from the quantity and complexity of the cabling. Studies are ongoing to determine the granularity at which the BV is delivered to the sensors to achieve an optimal time resolution.

One design that is being considered to manage the large number of BV channels is based on the HV system employed by the CMS CSC sub-detector, which also has a large number of HV channels. The idea is to use a limited number of commercial HV power supplies in the

service cavern to feed a system of trim regulators, located on the outer rim of the ETL disks, which generate locally the desired number of BV channels. Such a design could achieve a high-granularity BV system with a limited number of supply cables passing into the feedthrough, greatly simplifying service integration issues associated with the BV distribution.

3.8.3 Grounding

A proper grounding of a detector helps to achieve three main objectives:

- detector safety,
- preservation of signal integrity, and
- electromagnetic interference control (EMI control).

Detector safety is achieved by keeping detector common potentials close to that of the support structure, to which the detector is referenced. Signal integrity (noise reduction) is achieved by diverting noise currents that travel along electrically conductive services so that they do not reach the sensitive detector front-end electronics. Although a complete zero potential is not always achieved with such a diversion, at least a minimized differential voltage noise is achieved. EMI control is achieved by minimizing noise generation (for example lowering common current impedance and capacitive coupling in DC-DC converter) and localizing generated noise. The idea is that if EMI is acceptable for the front-end electronics, it should not affect its neighboring detectors, especially considering additional shielding provided by thermal screens and physical distance. The main sources of conducted noise in ETL are:

- cooling lines – common mode noise from compressor motors,
- LV and BV cables – common and differential mode noise from switching power converters, and
- control (safety) cables – not much of a concern since high impedance measurements will be used.

The main measures to be taken to achieve the safety and noise reduction objectives are the following.

- For the bias voltage, ensure that the locally stored energy does not exceed 10 J, and sustained currents are below 20 mA. Bleeder resistors should be included where appropriate.
- The two sides of the ETL detector, served by two cable trays at 3 and 9 o'clock, are kept galvanically isolated to prevent ground loop currents across the detector cooling Dees.
- The PP1 location is where the CE will have its main grounding ring and is where the ETL has its central grounding point, with two physical points separate for the two sides. All the cooling pipes and cable shields are grounded at PP1 to shunt common mode currents and achieve a common potential reference point. All electrical conductors should be capacitively filtered to the ground at that location as well.
- Cooling pipes should have a galvanic break preferably just outside the PP1 or before the next location where they have electrical connection to the support structure.
- The PP0 location serves as the central grounding point for all LV, BV, control and cooling services. The two halves of the PP0 grounding conductor can be connected together if that helps mechanical aspects. Copper braids of sufficient ampacity connect PP0 and PP1 on both sides; they should be able to carry current from the largest

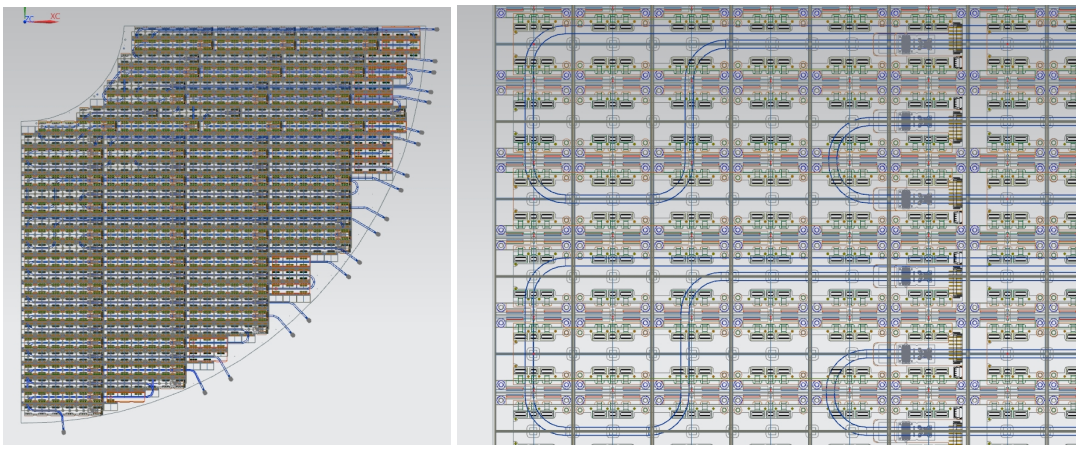


Figure 3.80: Left: A 90-degree section of an ETL support wedge with cooling inlet and outlets, modules and service hybrids. Right: Layout of the cooling loops inside that section.

power supply. The front-end electronics are referenced to the PP0 ground through the power cables and are electrically isolated from the detector cooling support half-disks, which are grounded to the PP0 at the 3 and 9 o'clock positions.

3.8.4 Cooling

The ETL detector is placed in front of the CE, inside an independent cold volume that permits access to the detector for service and maintenance. The ETL detector will be cooled with two-phase CO_2 at a nominal coolant operating temperature of $-35\,^{\circ}C$ with a maximum vapor quality (ratio between vapor and total mass) of 30%. The heat is evacuated through a network of pipes embedded inside the Al wedges, which are shown in Fig. 3.80. The fastener holes in modules and service boards are aligned such that open regions are left between rows of holes for routing of the CO_2 tubing, avoiding the screws. The total power dissipated in the ETL cold-volume is dominated by the ASIC power, with a smaller contribution from sensor leakage current. The breakdown of the resulting heat load is given in Table 3.11. The cooling plant is specified to handle 40 kW per end to provide contingency beyond the calculated heat load. The front cover, shown in Fig. 3.2, serves as a thermal insulator and can be retracted to gain access to the ETL detector. The dew point in the volume will always be kept at least $10\,^{\circ}C$ lower than the lowest temperature by flushing it with inert gas to avoid condensation. The gas is delivered through supply lines located in dedicated service trays described in Section 3.8.

Rigid vacuum-jacketed stainless steel pipes carry the CO_2 from the refrigeration plants in the service cavern, USC, to the top-most (X5) balconies in the collision hall. The cooling plant will be shared with the CE detector. Flexible pipes connect manifolds on the top of the YE1 structure to these balconies. On each endcap, coolant is provided to the ETL cold volume via four vacuum-insulated coaxial transfer lines, two per 3 and 9 o'clock service channel, running along the outside of the CE thermal screen. Each transfer line contains an interior stainless steel tube providing liquid CO_2, and the exterior tube returning two-phase coolant from the detector.

Each transfer line feeds one half of a disk and is fitted with manual shutoff and flow regulation valves. Therefore, even if one loop needs to be closed, the second disk still provides timing measurements over the full coverage of the detector. A simplified scheme of a vacuum-insulated transfer line is shown in Fig. 3.81. The manifolds presented in the schematic are currently being designed, taking into account space limitations on the periphery of the ETL detector.

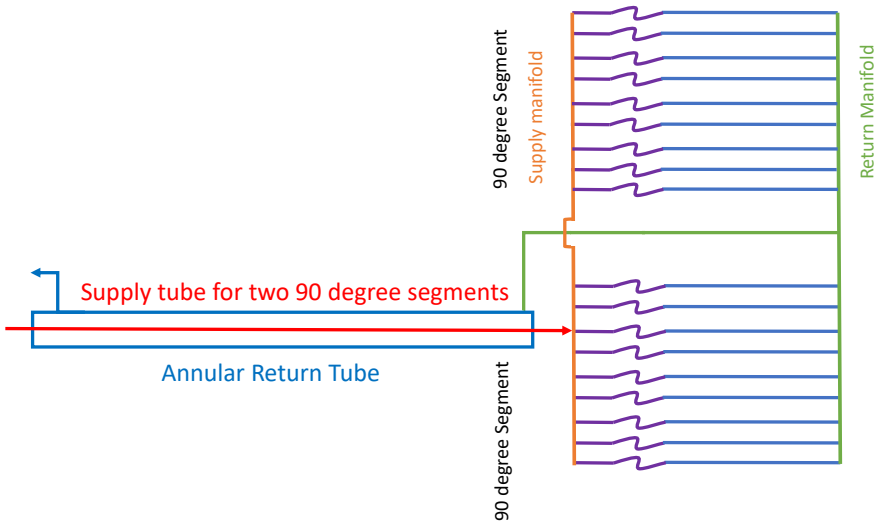


Figure 3.81: A schematic of the CO_2 distribution of a vacuum-insulated transfer line showing the supply manifolds and nine cooling loops, with their capillaries, providing coolant to each pair of 90-degree ETL wedges.

Table 3.12: Manifold pipe and cooling loop dimensions on the rim of ETL.

	Tube OD [mm]	Tube ID [mm]	Length [m]
Supply tube	10	8	20
Supply manifold section	10	8	0.25
Capillary	0.125″	0.05″	3
Parallel evaporator route	3/16″	4.254	1.44
Return manifold route	12	10	0.25
Return tube	26.7	23.5	20

Inside the cold volume, each transfer line is split into supply and return tubes that run along the outer rim of each ETL disk. One transfer line provides 10 kW cooling by distributing 107 g/s of CO_2 into 18 cooling loops through two ETL wedges. Therefore, one cooling loop has 5.95 g/s of the CO_2 flow. Each cooling loop is a pipe running within a round-bottomed groove machined into the aluminum plate, as shown in Fig. 3.74. A sheet of aluminum is glued to the plate to cover the cooling grooves.

The transfer lines are designed as coaxial tubes with a supply line diameter of 10 mm, a return line diameter of 26.7 mm, and a vacuum jacket of diameter 48.3 mm. With these dimensions, the temperature drop on the return line should not exceed 0.7 °C. The pressure drop on the evaporator and capillaries is estimated to be 10 bar, based on the data in Table 3.12. Studies of the frictional pressure and calculations estimating the resulting pressure drop (and therefore temperature) along the transfer lines are ongoing.

The thermal performance of the modules was studied using FEA simulations based on CAD models using ANSYS simulation software. Most heat is generated in the power boards, and in the ETROC chips. Heat from the ETROCs is evacuated directly to the Aluminum support structures, passing through the mounting film and the AlN carrier (Fig. 3.57). A heat transfer coefficient of 5000 $W\,m^{-2}\,K^{-1}$ is assumed for the heat transfer from the cooling pipe inner wall to the CO_2. Most of the heat on the service hybrids is generated from the DC-DC converters, located on the power board. Heat in the power board will conduct to the board's underside, pass

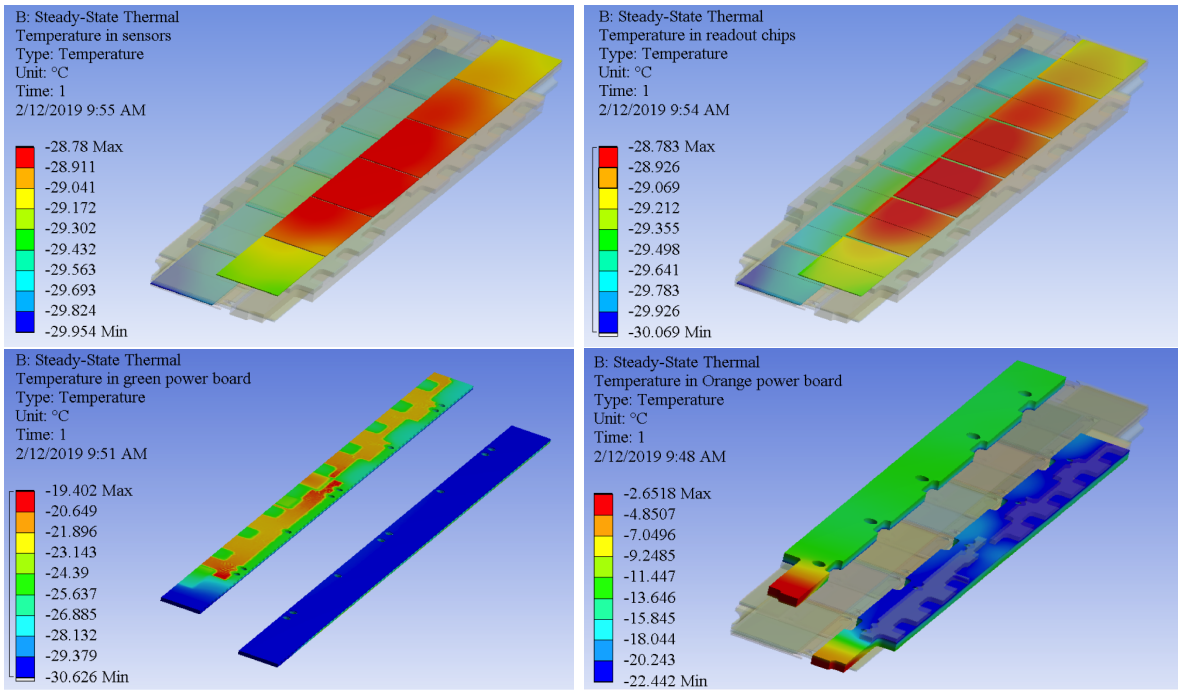


Figure 3.82: Finite element model simulation of a section of the ETL support, displaying sensor temperatures (top left), ETROC temperatures (top right), readout board temperatures (bottom left) and power board temperatures (bottom right).

downwards through the aluminum and film layers between the power and readout boards, and conduct down through the thickness of the readout board into the aluminum support wedges (Fig. 3.62).

Results of the FEA simulation are shown in Fig. 3.82 using the expected power consumption listed in Table 3.11. The simulation verifies that the sensor temperature will be maintained below $-25\,^{\circ}\mathrm{C}$ even at the hottest locations on the modules, and the ETROC temperatures are similar. The readout board temperatures are typically less than $-20\,^{\circ}\mathrm{C}$, while the power board temperature is $\approx -10\,^{\circ}\mathrm{C}$ except for a region near the connector at the end where there is no heat source but reduced thermal conductivity. Tests with a section of a cooling disk will be done in mid-2020, using CO_2 cooling and prototype modules, to validate the heat transfer model and optimize the thermal aspects of the module design and cooling pipe routing.

Chapter 4

Common systems

4.1 The data acquisition system

4.1.1 Overview

The MIP timing detector data acquisition system (DAQ) consists of multiple radiation tolerant electronic components in the front-end, and electronic boards based on Field Programmable Gate Arrays (FPGAs) in the back-end of the detector. The system is connected to the central CMS DAQ system via a common interface. The details of the central DAQ system and its common interface will be described in another TDR planned to be finalized in 2021. This section focuses on the common back-end components for the ETL and BTL subdetectors. The front-end components are described in the subdetector specific Sections 2.3 and 3.3.

The Level-1 (L1) Trigger system for the HL-LHC will have a latency of 12.5 μs and an average rate of 750 kHz. The MTD DAQ is minimally required to read out the MTD upon receipt of an L1 accept (L1-A) from the CMS trigger. The trigger signals together with the fast and slow control signals are distributed via the same bidirectional links. Therefore, the back-end DAQ electronics boards are equipped with high-performance advanced FPGAs together with high-speed optical drivers to accomplish these tasks and enable online processing options at the L1-A rate. An option to include the MTD in the L1 trigger would require that the back-end electronics would also need to process a separate data path at a 1 MHz rate, generate L1 trigger primitives and forward this data to the L1 Trigger system.

4.1.2 DAQ system requirements

The DAQ system is designed to read out the channels in zero-suppression mode above a configurable threshold of about 0.25 MIPs upon the receipt of an external hardware-level trigger (L1 trigger) provided by other CMS subsystems. The front-end electronics is designed to receive these triggers and transmit the corresponding data from the detector via high speed optical links to the MTD DAQ system. Based on the current layouts of the front-end electronics and expected data rates, 864 and 1600 links will be needed for the BTL and the ETL, respectively.

The DAQ system distributes trigger, fast and slow control signals from the upgraded trigger control distribution system (TCDS2) and detector control system (DCS) to on-detector electronics. Moreover, the slow monitoring signals such as temperature and voltage readings from the on-detector electronics will be collected. Therefore, all links communicating with the MTD on-detector electronics will be bidirectional.

4.1.3 Hardware description

The back-end electronics infrastructure is based on advanced telecommunications architecture (ATCA) with centrally defined CMS specifications. Within these specifications, the electronic boards are placed in ATCA crates in a dual-star network topology, where two central 'hub' boards are connected to all 'node' boards (also known as 'leaf' boards) within the crate. One of two central hub slots is reserved for future development possibilities within the CMS specifications. The MTD DAQ boards are positioned in ATCA crate 'node' slots. The interface to the central TCDS and DAQ systems is provided by the data trigger hub (DTH400) board placed in a hub slot and additional data bandwidth to the DAQ system is provided by a DAQ800 board, which can be placed in one of the node slots [109]. A single DTH400 board is capable of providing a DAQ throughput of 400 Gb/s, the DAQ800 board is designed to provide an additional 800 Gb/s DAQ throughput.

Unpacking and processing of the data received from the front-end will be accomplished by the MTD back-end boards positioned in node slots; hence these boards will be the main drivers of the MTD DAQ system.

The data read out from the detector are transmitted to the back-end via Versatile Links+ (VL+). The Versatile Links consist of radiation tolerant multi-gigabit communication ASICs (lpGBT) and radiation tolerant optical transceivers (VTRx+) capable of handling the data rate. A single lpGBT will handle communication of 12 readout ASICs in BTL and 24 or 12 readout ASICs depending on the rapidity in ETL. There will be in total 864 lpGBTs and bidirectional links in BTL and 1600 in ETL. In order to cope with the required data throughput, the lpGBTs will be operated at 10.24 Gb/s and 5.12 Gb/s operation modes depending on the occupancy of the particular detector region. The data received from the E-links of a lpGBT are serialized and transmitted to VTRx+, which transmits them through 70 m multi-mode optical fibers to the back-end electronics. All of these components are specified to safely tolerate the radiation expected from the full HL-LHC run period. Data are then received in USC-55 by Firefly transceivers on the MTD DAQ boards through their front-panels, which process and package them for transmission to the DTH400 and DAQ800 boards via 25 Gb/s (or 16 Gb/s depending on the FPGA architecture) front-panel optical connections. Data collected from all MTD DAQ boards are sent to the event builder using the 100 Gb/s data-to-surface (D2S) protocol.

Two state-of-the-art candidates for leaf boards are currently being developed in CMS for use in the L1 Trigger and other detectors' back-ends, namely Serenity [110] and the Barrel Calorimeter Processor (BCP) [111]. The firmware and software of these boards will be adapted for use in the MTD. The MTD can exploit either of these options.

The Serenity platform contains two interposers enabling the selection of a large variety of FPGAs and other processing and distribution units. This flexibility is achieved by two 1836-pin SAMTEC Zray connectors. The FPGA daughter boards connected to the interposers provide the necessary processing capability. Each interposer is connected to 12 SAMTEC Firefly transceivers (each having 12 optical drivers), or in other words 72 bidirectional transceiver lines. Overall the board is capable of carrying two high-speed, high-capacity FPGAs and driving 144 bidirectional links to the front-end and to the DTH400 or the DAQ800 via MTP fibers. The reference clock necessary for driving the DAQ links from the FPGAs is distributed by low-jitter Si5345 phase locked loops (PLLs). The performance of these PLLs is presented in Section 4.2.

Figure 4.1 shows the first version of the Serenity board and illustrates the main components of the board together with two interposers and the data links. Currently four different daughter

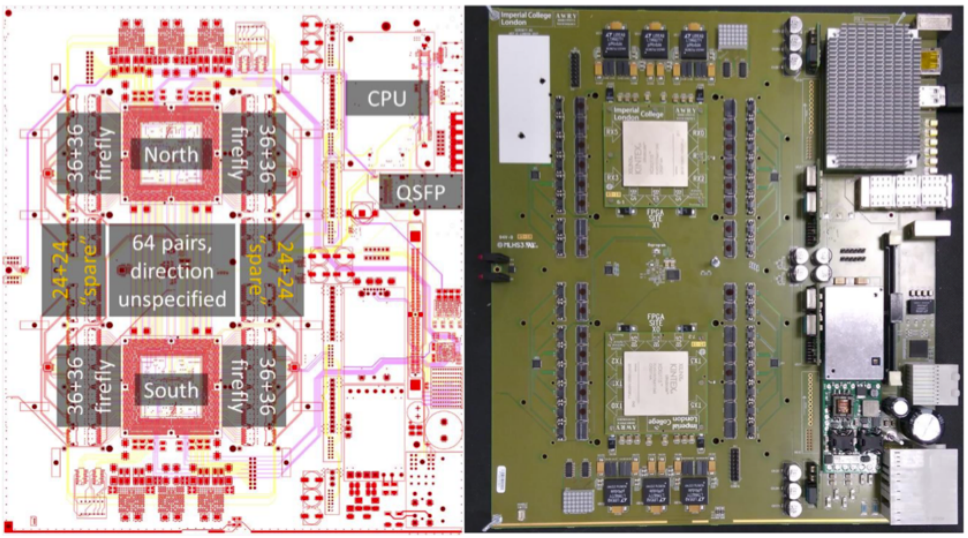


Figure 4.1: Layout (left) and photo (right) of the first version of the Serenity ATCA board. The board has two high-speed interposers enabling it to carry two advanced FPGA daughter boards.

boards with different functionality and FPGA models are being developed:

- Xilinx Kintex Ultrascale Plus K15P FPGA symmetrical (enabling bidirectional links) boards,
- Xilinx Virtex Ultrascale Plus VU9P FPGA symmetrical boards,
- Xilinx Kintex Ultrascale KU115 FPGA symmetrical and asymmetrical (single direction, RX or TX links) boards,
- Pure clock distribution boards.

The Xilinx Kintex Ultrascale Plus KU15P FPGA board provides the largest number of high-speed transceivers at the lowest cost, hence it is an attractive choice for the MTD DAQ system. The Ultrascale Plus architecture also provides 25 Gb/s communication links to the DTH400 and DAQ800 boards. The first Serenity prototype board together with various daughter boards including the Xilinx Kintex Ultrascale Plus KU15P FPGA flavor have been successfully tested.

The BCP board (Fig. 4.2) has two Xilinx Kintex Ultrascale KU115 FPGAs, soldered directly to the main board, providing the necessary processing power and the communication options. The board management and the services are handled by the Xilinx Zynq FPGAs on an ELM board and an on-board Zynq intelligent platform management controller (IPMC) chip. The board provides 56 transceivers per FPGA via the Firefly connectors. Overall, it is capable of carrying two high-speed high-capacity FPGAs and driving 112 bidirectional links to the front-end and to the DTH400 and DAQ800 via MTP fibers. A BCP demonstration prototype unit with a single FPGA will be available in the first half of 2020.

The optimal BTL readout energy threshold is around 1 MeV (Section 2.3), corresponding to 0.25 MIPs. This threshold yields a maximum of 8% occupancy as shown in Fig. 2.32. With 32 readout channels per ASIC and an event size of 120 bits, a data rate of 230 Mb/s is estimated per ASIC at 750 kHz L1 trigger rate. A similar estimate is presented in Fig. 3.68 for ETL, with an ETROC occupancy less than 6% as shown in Fig. 3.66. Overall, the maximum expected data rates from the readout units are 5.5 Gb/s and 1.5 Gb/s for the BTL and ETL detectors, respectively. In order to handle these rates, for the BTL detector, each MTD DAQ board needs to transmit around 230 Gb/s of data to the DTH400 and DAQ800 boards, corresponding to ten

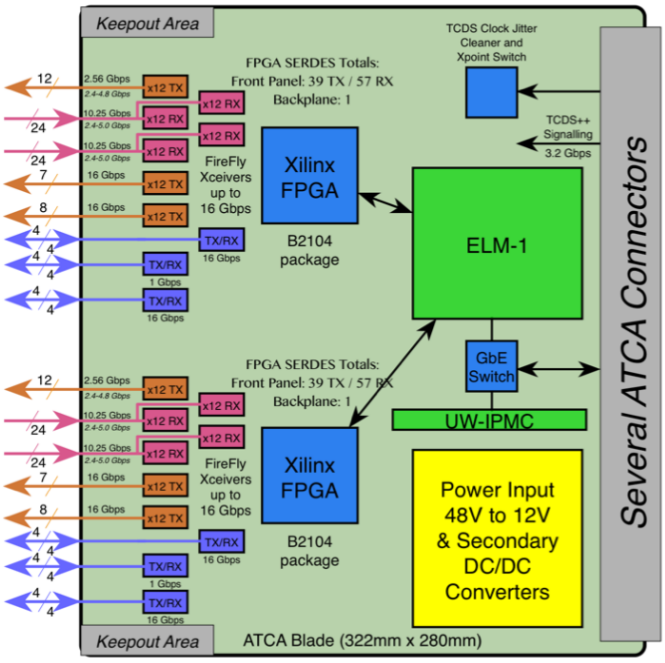


Figure 4.2: Layout of the BCP ATCA board. The board has two high-speed Xilinx Ultrascale KU115 FPGAs and an advanced on-board control and processing system powered by various Xilinx Zynq FPGAs.

Table 4.1: Counts of back-end components in the MTD Detector for the Serenity boards with 144 links per board and for the BCP boards with 112 links per board.

Component	BTL	ETL
Bi-directional links	864	1600
DAQ boards (BCP)	9	16
DAQ boards (Serenity)	8	12
DTH400 boards	1	2
DAQ800 boards	3	4
Crates	1	2

25 Gb/s links (or fifteen 16 Gb/s links). For the ETL detector, each MTD DAQ board is required to transmit around 190 Gb/s of data to the DTH400 and DAQ800 boards, corresponding to eight 25 Gb/s links (or twelve 16 Gb/s links).

The number of DAQ boards needed to handle the total DAQ throughput depends on the number of bidirectional data links available on the boards. For the Serenity boards with 144 bidirectional links per board, the back-end system will comprise eight DAQ boards for the BTL readout and seven DAQ boards for the readout of each endcap of the ETL. For the BCP boards, with an I/O capability of 112 bidirectional links per board, nine DAQ boards for the BTL and nine boards for each end of the ETL will be needed.

In either case, the total DAQ throughput can be handled by one DTH400 and three DAQ800 for BTL and by one DTH400 and two DAQ800 for each endcap of the ETL detector.

In total, the DAQ system will occupy three crates in two racks. The layout of the BTL and ETL crates are shown in Fig 4.3. The major component counts are summarized in Table 4.1.

In the current plan, a more computationally powerful FPGA is proposed for the BCP board

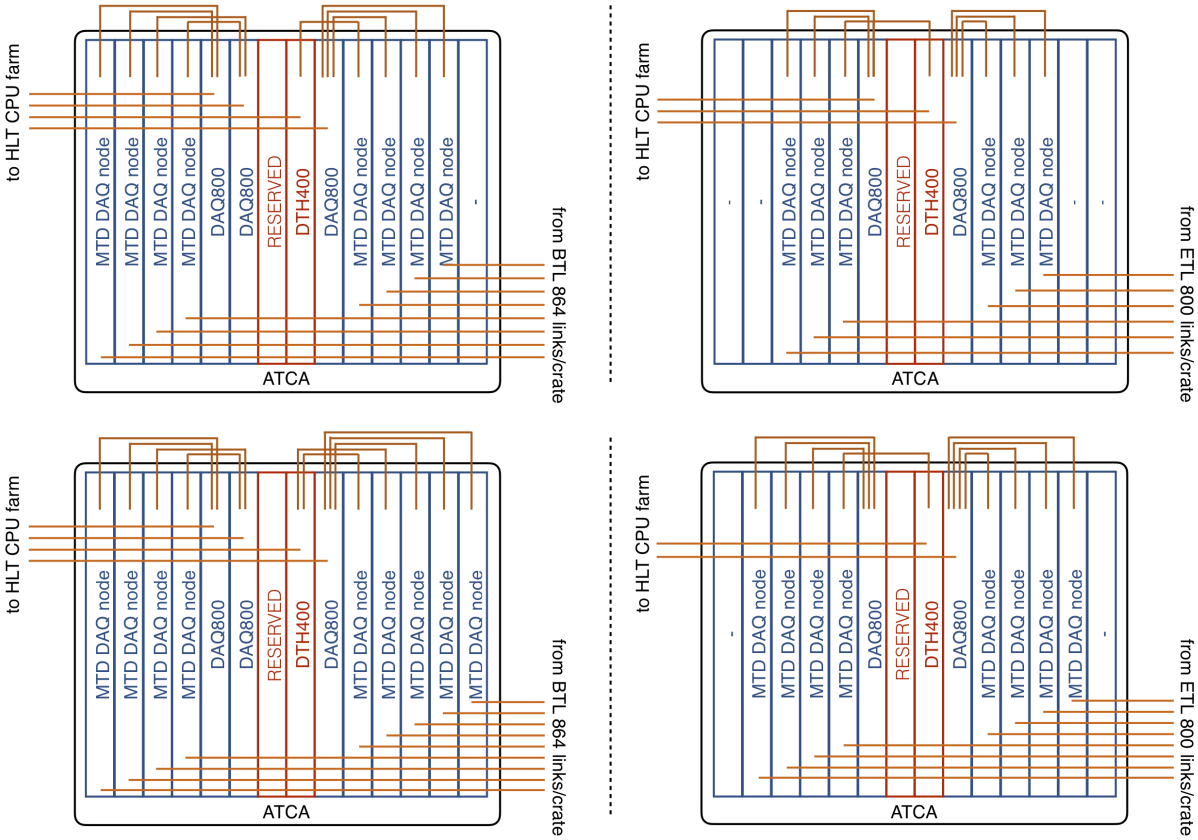


Figure 4.3: Top: The MTD DAQ ATCA crate layout with 8 (left) and 6 (right) Serenity boards (carrying 2 Xilinx Kintex Ultrascale Plus KU15P FPGA daughter boards) for the BTL detector and each endcap of the ETL detector, respectively. Bottom: The MTD DAQ ATCA crate layouts with 9 BCP boards (carrying 2 Xilinx Kintex Ultrascale KU115 FPGAs). The DAQ links from on-detector electronics for the BTL detector (left) and each endcap of the ETL detector (right) are grouped into 8 MTD DAQ boards per crate.

and this could be an advantage, if the plan for the MTD to participate in the L1 trigger moves forward. However, the Serenity platform does have flexibility in the choice of FPGA.

Either of these two boards will provide the needed capability for the back-end of the DAQ. The decision of which to use may depend on considerations relating to synergies with other systems concerning firmware and software development or cost considerations, including lifetime support issues.

4.2 The clock distribution

The readout electronics in the CMS detector is synchronized with the LHC bunch-crossings.

A low-jitter sampling clock synchronized to the 40.078 MHz frequency bunch-crossing rate is required to be distributed to the readout systems. In order to achieve the target timing performance, the clock distribution system should have less than 15 ps rms link-to-link jitter over all clock distribution links within a frequency range of the order of 1 Hz to the sampling frequency of the ASIC. The lower limit is driven by the monitoring capability of the system (Appendix D.II). Anything above this jitter range will significantly worsen the timing performance

of the MTD. For instance, a 35 ps timing performance obtained from the sensors and readout electronics would be degraded by 3 ps by a clock distribution system with 15 ps rms jitter. Distributing the sampling clock with low jitter over 2000 readout units under varying temperature and harsh radiation conditions poses a significant challenge.

4.2.1 Components of the clock distribution chain

The LHC clock synchronized with the bunch-crossing rate is distributed directly from the RF system of the LHC. The clock received by the CMS central systems is distributed to the off-detector electronics and, via multi-mode fiber cables, to the on-detector electronics.

The sampling clock can be distributed via one of two different schemes:

- a tree of encoded clock paths is proposed as the baseline clock distribution scheme as shown in Fig 4.4. This scheme consists of an encoded clock with a guaranteed available low jitter at the end-points of the distribution tree and with a phase consistency between the leaves. The R&D will determine if the promise of low jitter and phase consistency can actually be achieved or, alternatively, what needs to be done at each intermediate stage to clean the jitter of the clock. Most importantly, the R&D will also determine if the monitoring of the phase differences would allow the MTD to achieve the desired timing precision.
- If the baseline solution does not meet the differential jitter requirements, a pure clock (without encoding) will be distributed via a separate clock distribution network as illustrated in Fig 4.5.

The baseline scheme is embedded into the TCDS2 system, hence it will use the TCDS/DAQ links for the clock distribution. The 40.078 MHz clock received by the TCDS2 RF RX module is distributed together with the trigger and fast control signals to the ATCA DAQ crates by the TCDS2 nodes. The DTH board transmits the received LHC clock (or frequency multiples of it) to the MTD DAQ boards through the high-speed lanes of the ATCA backplane. The FPGAs in the MTD DAQ boards encode clock in the lpGBT stream and send it to front-end at 2.5 Gb/s. The lpGBT ASICs in the front-end recover the clock and distribute it to the readout ASICs.

The pure clock distribution tree, on the other hand, will require separate clock distribution nodes. One dedicated precision clock an monitoring module will receive the clock directly from the LHC interface module (also reffered to as TCDS2 RF RX) and distribute it to the front-end boards via a tree comprising one dedicated, one VTRX+ and one radiation tolerant clock-distribution ASIC per clock entry point in the front-end: CC boards in BTL and hybrid service cards in ETL (Sections 2.3 and 3.3, respectively). An additional fiber can be optionally deployed to monitor the transmitted clock via a loop-back in the front-end boards.

4.2.2 Characterization of current CMS clock distribution system

The current RF clock delivered by the LHC is specified with 9 ps rms jitter [112]. Its future evolution holds the promise to achieve even better performance.

In a simplified distribution tree with a single backend emulator board and two optical drivers, the current versatile link (VL) framework with GBTx and VTRx is capable of providing the LHC clock to the end-points with less than 10 ps jitter. However, a wider and more realistic clock distribution network with two microTCA crates and two back-end emulator boards fed from a single common clock source (Fig. 4.6) yields a less stable clock distribution with higher rms jitter. Therefore, components of the current clock distribution network within TCDS and VL framework are being extensively characterized to define possible improvements for future

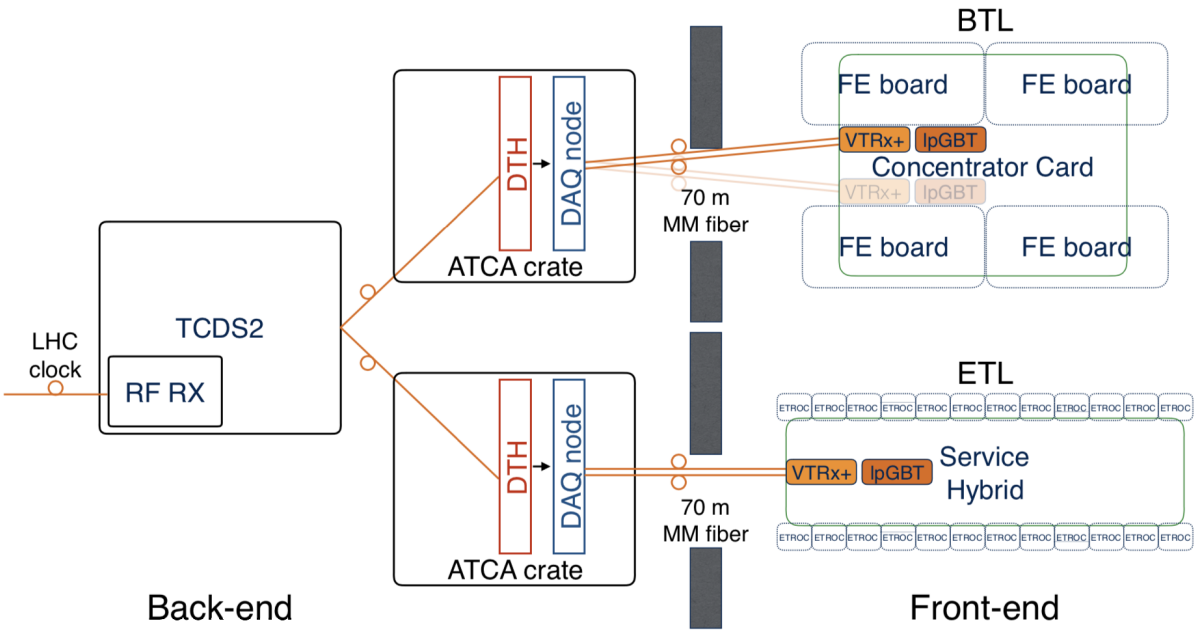


Figure 4.4: A schematic layout of the MTD baseline clock distribution system. The 40.078 MHz clock received by the TCDS2 RF RX module is distributed to the ATCA crates through the DTH boards. The received LHC clock (or frequency multiples of it) is distributed to the MTD DAQ boards. The FPGAs in the MTD DAQ boards encode clock in the lpGBT stream and send it to front-end at 2.5 Gb/s. The lpGBT ASICs in the front-end recover the clock and distribute it to the readout ASICs.

clock distribution systems in CMS. Figure 4.6 illustrates the layout of a test system emulating the encoded clock distribution system with components currently used in CMS.

In addition, studies analyzing the effect of temperature on the distributed clock jitter have been carried out to ensure the performance of the clock distribution network in the MTD front-end electronics at the operation conditions. The measurements indicate that decreasing temperature of the VL components slightly improves the link-to-link clock quality in a linear way. A consistent 2 ps decrease of the rms clock jitter out of the GBTx is observed from 20 °C to −30 °C over repetitive tests in a climate chamber.

Equivalent system tests will be performed to characterize the system performance of the Phase-2 electronics once prototypes of the components foreseen to be used in the baseline clock distribution network are ready. Furthermore, realistic CMS use case conditions will be created, by populating and powering all the nodes of the ATCA crates, as well as all of their optical links and implemented FPGA logics.

4.2.3 MTD clock distribution system R&D

Ongoing common R&D efforts within CMS aim to provide the lowest clock distribution jitter for both the baseline and the pure clock distribution trees for precision timing detectors. With this purpose, various components are being tested and designed.

A preliminary characterization of the lpGBT encoded clock, which is the key component of the baseline option, indicates a promising clock distribution performance particularly for phase noise in the 100 Hz to 10 MHz range. At e-clock frequencies multiples of the LHC clock, a small deterministic jitter is observed, which may affect the performance of TOFHIR ASIC for

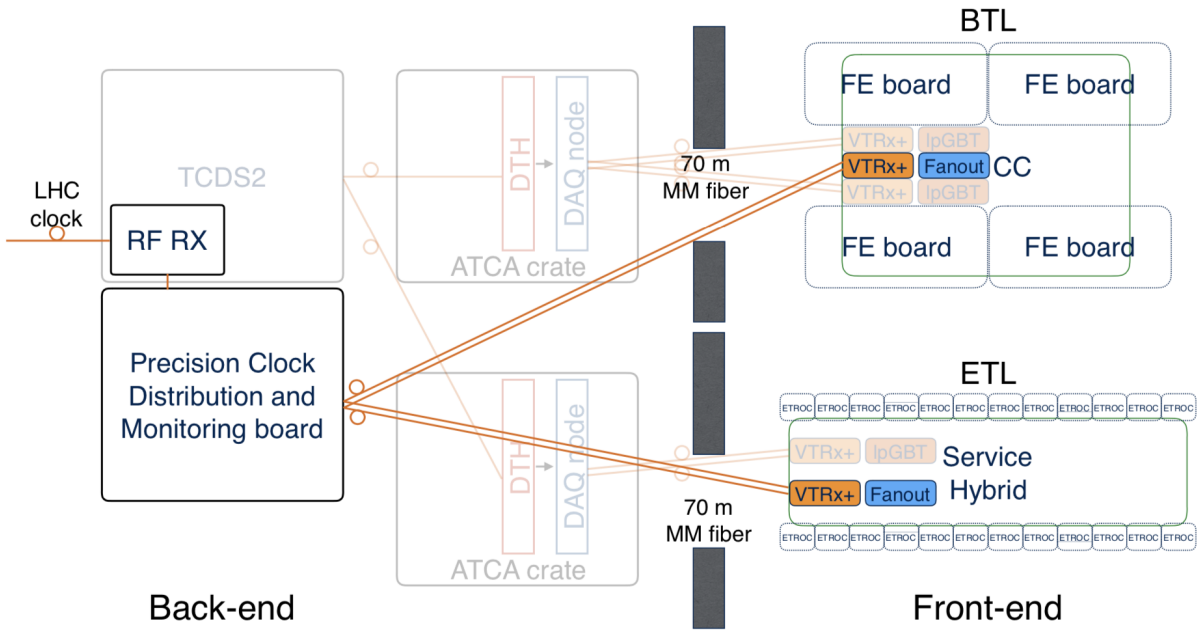


Figure 4.5: A simplified layout of the MTD alternative clock distribution system. The 40.078 MHz clock received by the TCDS2 RF RX module is distributed directly to the front-end modules via the precision clock distribution and monitoring module. A bidirectional fiber will be used to monitor the transmitted clock via a loop-back in the front-end boards.

which a reference clock of 160 MHz is required. Further measurements with a higher sample size and including link-to-link jitter performance as well as a more detailed analysis of the observations are necessary for a comprehensive evaluation of the clock distribution performance of the lpGBT.

In parallel, dedicated studies towards the development of a pure clock distribution tree are in progress. PLLs are one of the fundamental building blocks of a clock distribution tree. Chains of PLLs are planned to be used in both baseline and pure clock distribution schemes. Optimizing the PLL configuration and architecture is one of the key elements to achieve a low-jitter and stable clock-distribution system. With an optimized PLL configuration, a sub-optimal clock input with relatively high jitter can be reconstructed to have lower jitter values. This feature is demonstrated using a chain of two Si534x PLLs serially connected to each other. A clock superposed with a square wave noise at 1 kHz is generated and fed into the PLL chain. With correct PLL loop-bandwidth selection, the rms jitter of the clock (measured in a 1 Hz to 1 MHz range) can be reduced from 72 ps down to 2 ps as illustrated in Fig. 4.7. The jitter components higher than the frequency of the loop bandwidth of the PLL can be cleaned.

A similar PLL chain is used in the MTD DAQ boards. A special clock distribution daughter board is designed to test the clock distribution performance of the Serenity board. An rms jitter of 2.8 ps is measured in the Si5345 PLL based clock distribution network of the board. Similarly, the BCP board will have a Si5345 PLL based clock distribution tree and various SMA test connectors to evaluate the clock distribution performance. These results are also very encouraging for an ongoing development of a PLL-based Precision Clock Distribution and Monitoring (PCDM) Board. With high-performance low-jitter clock buffers and on-board time-to-digital converter ASICs (TDC), the PCDM board aims to distribute the sampling clock over 60 links with very low jitter and to monitor the looped-back sampling clock at a high frequency. Mul-

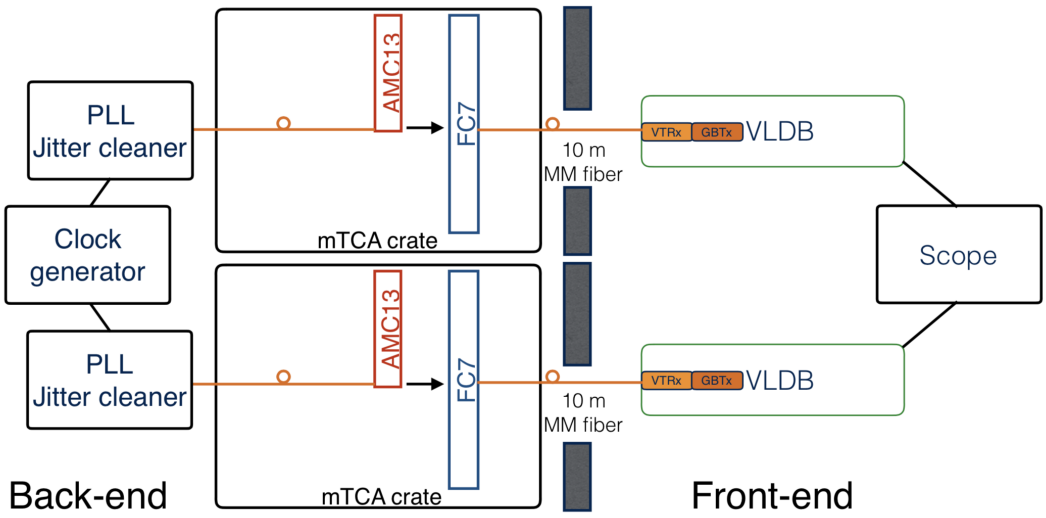


Figure 4.6: The layout of the test stand emulating the current CMS clock distribution system with two CMS microTCA crates and two GBTx chips on evaluation boards. A 28 ps rms link-to-link jitter is observed.

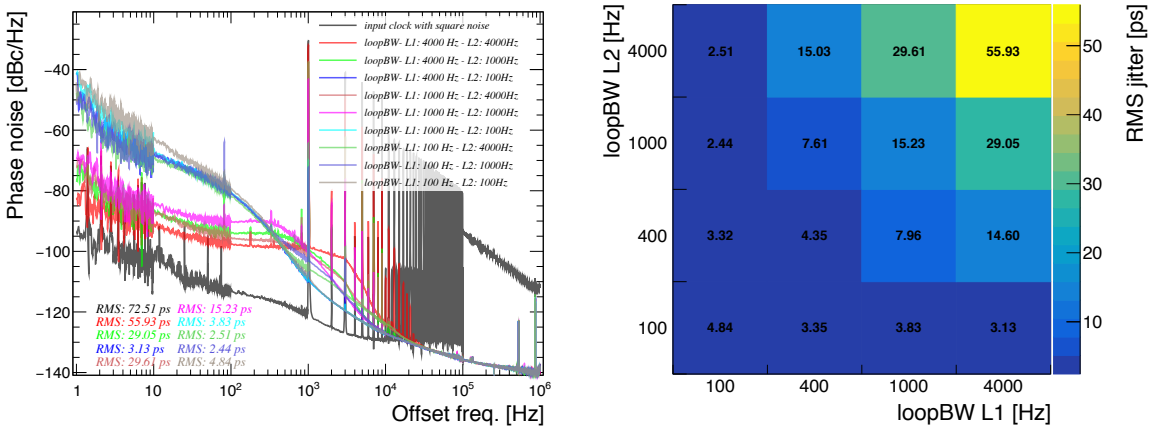


Figure 4.7: Jitter mitigation capabilities of a chain of two PLLs demonstrated with square wave noise at 1 kHz superposed on a square wave clock. The left figure shows the phase noise spectrum of the output clock from 1 Hz to 1 MHz with different PLL loop bandwidth settings and the right figure shows the rms jitter for the corresponding frequency range with different PLL configurations, scanning a loop bandwidth range for the two PLLs of (100,100) Hz to (4,4) kHz. The rms jitter of the input clock is 72 ps (dark gray).

tiple boards can be cascaded to distribute clock to all the on-detector readout units of MTD. There is an ongoing R&D for a radiation-tolerant clock distribution fan-out buffer. The fan-out ASIC will have four inputs that can be multiplexed and buffered to twelve output signals. The chip is specified to have radiation tolerance up to 200 Mrad. Such an ASIC can be also useful for the baseline clock and fast-control signal distribution.

4.2.4 Developement plan and decision points

In order to reach a decision between the baseline clock distribution and the fallback pure clock distribution solution, an extensive characterization of the first prototypes of the readout modules and the baseline backend clock distribution shown in Fig 4.4 is necessary. The following

tests will be performed to characterize the system:

- Readout unit tests (with both concentrator cards and service hybrids) using only a simple single DAQ evaluation board and a high performance input clock at 40.078 MHz,
- MTD DAQ board clock distribution tests with dedicated characterization boards,
- ATCA clock distribution tests with an external high performance clock, DTH board prototype and the MTD DAQ board,
- ATCA crate and front-end tests with a single MTD DAQ board receiving clock from the DTH board and two front-end readout units,
- Complete ATCA network tests with three DTH boards (one emulating the TCDS2 system) and two front-end readout units,
- Impact of the optical attenuation,
- Impact of the FPGA occupancy,
- Impact of the temperature variations in all of the tests above.

Due to the schedule restrictions the DTH board will be used to emulate the TCDS2 system. As previously discussed, all of these tests are already performed for the microTCA crates and the commercial test components are available in the teststands. With the availability of the front-end and back-end prototypes the BTL clock distribution tests with a multiple ATCA crate network will be performed in the second half of 2020 and the decision for the BTL clock distribution will be given end of July 2020. Similarly, the wider network tests for ETL will be performed in the first half of 2021 and the decision will be given in July 2021. The performance of the baseline clock distribution system will be assessed through a series of check-points and test:

- Q2-Q3 2019 - initial performance charecterization with lpGBTv1 and VTRx+
- Q1-Q2 2020 - performance tests with VL+ and MTD back-end components;
- Q3 2020 - performance tests with the BTL front-end prototype (lpGBTv1);
- Q3 2020 - preliminary characterization with lpGBTv1 complete;
- Q1 2021 - performance tests with lpGBTv2;
- Q1 2021 - Ready to make a decision for the clock distribution in BTL (the decision can be postponed to Q3 2021);
- Q1 2021-Q2 2021 Performance tests with the ETL front-end;
- Q3 2021 - Ready to make a decision for the clock distribution in ETL (or common MTD decision).

While the test plan would enable to be ready for a decision for BTL in Q1 2021, the BTL front-end design can be maintained compatible with both options and the decision postponed to when producing the final boards for the experiment in Q4 2021. This would give more margin to be really sure about the clock performance for BTL. The common decision could be marked by a common MTD milestone on the clock system in Q3 2021. On the other hand, if the information gathered by the end of 2020 is sufficiently clear, the decision can be made and the BTL front-end design developed only according to one option. The plan is included in the project schedule, with high level milestones listed in Table 6.1.

In parallel, the CMS-wide R&D plan for the charatcerization of the pure clock distribution system follows the schdule below:

- Q2-Q3 2019 - Initial performance charecterization with PCDM demonstrator and VTRx+
- Q3 2019 - Fanout ASIC specifications
- Q3 2020 - PCDM prototype 1 tests
- Q3 2020 - Fanout ASIC radiation tests
- Q3 2020 - Performance tests with the BTL front-end prototype
- Q4 2020 - PCDM protoype 2 tests
- Q1 2021 - Performance tests with the ETL front-end prototype

4.2.5 Clock distribution monitoring and calibration

In order to evaluate options for clock-distribution monitoring and jitter cleaning, three domains of frequencies, which might affect the jitter, need to be distinguished. In the high-frequency domain above the PLL loop bandwidth (100 kHz to 100 MHz), the effects will most probably be canceled by back-end and on-detector PLLs. The lpGBT ASIC used in both the BTL and the ETL on-detector electronics has a PLL with a loop bandwidth around a few MHz. Furthermore, there will be an additional PLL in the ETROC ASIC. The specification and the architecture of this PLL is still under study.

The high-frequency domain is followed by an intermediate-frequency domain from 10 Hz up to the PLLs loop bandwidth, for which the negative impact on the phase and jitter will need to be monitored and if possible controlled in the baseline clock distribution scheme.

To enable this clock phase monitoring feature, two candidate technology options are under study. Both of these options need the clock (encoded or not) to be looped back in the front-end in order to measure the stability of the clock transmitted to the front-end. The first option relies on the use of Digital Dual-Mixer Time Difference (DDMTD). The principle is based on using a slightly lower frequency clock to sample both reference (output) clock and recovered clock, resulting in a slow sweep of their mean period, and hence, permitting to amplify the time difference between the two clocks. The second one is based on TDCs to measure on a regular basis the latency of the looped-back links, thus monitoring the variation of the phase at a frequency and resolution depending on the TDC capacities. Both of these options are under study for baseline and pure clock distribution networks.

Finally, a low-frequency domain below 10 Hz causes clock drift effects which need to be monitored, and if possible controlled at the hardware, firmware, or software level or calibrated out using collision events. In case of the presence of intermediate frequency ($>$10 Hz) jitter degrading the clock to an unacceptable level, an independent clock distribution may need to be designed. The pure clock distribution system will be separated from the baseline clock distribution tree at the receiving point of the RF clock.

Phase noise contributions at offset frequencies below 10 Hz from the carrier frequency may remain as a concern even if the performance measurements are already found to be at the level of 10 ps for the rms jitter. As a consequence, the clock phase, once calibrated over the whole detector, will have to be monitored in order to follow its possible variations and obtain knowledge about its longer term level of stability. For example, variations on the time scale of one minute can be corrected using minimum bias events with a precision of about 5 ps per channel or per clock region (Appendix D.II).

4.3 L1 Trigger options

The CMS Phase-2 detector will feature a Level-1 Track Trigger [4]. This addition to the CMS L1 Trigger system is the key to withstanding the challenges posed by the high rate of pileup at the HL-LHC. Similarly to offline reconstruction, the time information from the MTD can be combined with the track information at the trigger level to further mitigate the impact of pileup events or can be used to target specific final state topologies with characteristic time structures. For example, the isolation of candidate leptons and photons from the L1 Calorimeter and Muon triggers from other charged particles identified using the L1 Track Trigger will be a powerful method to discriminate genuine candidates from the multi-jet background. The MTD can distinguish vertices that are separated in time and help reduce the rate of tracks incorrectly included in the isolation sum. In addition, long-lived particles, predicted by several extensions of the standard model, are expected to produce hits in the MTD inconsistent with the time-of-flight of known particles: a feature that can be exploited in the trigger.

The participation of the MTD in the L1-Trigger decision is not included in the MTD baseline design. However, this option is being considered within the collaboration. The evaluation and possible positive acceptance of a Level-1 timing trigger proposal can only be made within the context of the CMS trigger TDR studies (Q1 2020), including considerations on the additional benefits brought by an MTD trigger to the exploitation of the HL-LHC data, on the optimization of bandwidth allocations, on the impact on the Level-1 Trigger System, links, and backend boards. In this section, we discuss the technical changes that would be needed to the MTD hardware not to preclude participation in the Level-1 Trigger. We assess the risks, the decision points and the cost impacts of the changes that would be required in the front-end, in order to inform future decisions.

4.3.1 Level-1 MTD requirements and architecture

The current CMS Trigger system operates in two levels: the first level (L1) consists of custom designed electronics that receives input from the calorimeter and muon systems and generates a trigger within 3 μs and at a rate of up to 100 kHz. The L1 Trigger system is composed of two separate data streams: a muon trigger which takes as input the muon stubs and reconstructs L1 muon candidates and a calorimeter trigger which takes as input the calorimeter data and reconstructs electrons/photons, taus, jets and various global sums. Information from the Muon Trigger and Calorimeter Trigger subsystems is then sent to the Global Trigger to determine if constructed L1 objects pass the L1 menu requirements. The Global Trigger then generates an L1-A if L1 menu conditions are met and sends an L1-A signal to the TCDS, which propagates the L1-A to all subsystems. Upon receipt of an L1-A, channels above threshold are read out and the data handed over to the High Level Trigger (HLT).

An upgrade to the L1 Trigger system is being planned in order to cope with the demanding conditions of the HL-LHC. The Trigger system will maintain a two level strategy in Phase-2, however, the entire detector readout electronics and DAQ will be replaced to allow a maximum L1-A rate of 750 kHz, and a latency of 12.5 μs (500 LHC bunch crossings).

Similar to the current design, the Phase-2 L1 Trigger system will continue to receive data from the front-ends of the electromagnetic, hadronic, and forward calorimeter systems, as well as the muon systems. Notably, for the first time, the L1 trigger will also include data from the Outer Tracker, which will be reconstructed to obtain L1 tracks. The correlation between the L1 tracks, the calorimeter deposits and the muon candidates will primarily take place in the L1 Correlator system. The L1 Correlator system could also be used to separate particles that originate from pileup collisions from those that originate from the primary vertex using timing

information from the MTD.

In order to implement a data path from the MTD to the L1 trigger, modifications to the currently designed ASIC and front-end electronics are required. These are:

1. implementing a separate buffer in the two front-end ASICs for storing data destined for the L1 stream;
2. enlarging the output bandwidth from the front-end ASICs by doubling the number of E-links and operation of the lpGBT at its maximal bandwidth of 10.24 Gb/s instead of the nominally designed 5.12 Gb/s;
3. building logic into the back-end electronics that would deliver both a Level-0 and L1 trigger for readout; and
4. providing the ability to read the MTD trigger data into the L1 Trigger system within 7.5 μs.

To read out the front-end electronics, Level-0 trigger logic is required to reduce the readout rate from the nominal 40 MHz by a factor of 40 to 1 MHz. As the precise latency and architecture for the L1 trigger is still being developed, we provide two options for the source of this data reduction. The options have been identified as:

- Region of Interest (ROI) seeding using a configurable minimum p_{T} threshold, $p_{\mathrm{T}}^{\mathrm{ROI}}$, L1 calorimeter and L1 muon tracking stubs from the inner most chambers.
- A Level-0 menu with the thresholds and objects optimized for a 1 MHz rate.

The first option, an ROI from the L1 Calorimeter Trigger and Muon Trigger subsystems would ask for each subsystem to identify calorimeter clusters or muon stubs above a given threshold and forward them to the MTD backend. At the MTD back-end, the ROI would be converted from the L1 Calorimeter/Muon geometry to the MTD geometry using a look-up table (LUT). The MTD back-end would then forward a read out request to the MTD front-ends corresponding to the ROI identified by the trigger. Using a minimum bias sample which is generated to mimic nominal HL-LHC running conditions with a pileup of 200 vertices it has been shown that a $p_{\mathrm{T}}^{\mathrm{ROI}}$ threshold of 5 GeV is sufficient to reduce the ROI rate by a factor of 50.

The second option, the Level-0 Menu, would ask for a given event to have at least one L1 calorimeter or L1 muon object above a configurable p_{T} threshold. This would require the calorimeter and muon subsystems to identify their highest p_{T} candidates and notify the MTD back-end if a candidate is found. The option to seed from tracking stubs has also been considered. However, because of a high fake rate and lack of pointing information this is not a tenable option. Furthermore, L1 tracks have also been considered as ROI seeds. However, in the case where a long-lived particle is more than a few millimeters from the beam spot an L1 track would have a low probability of being created due to the beam spot constraint of the L1 tracking algorithm. The overall architecture and proposed insertion of the L1 MTD trigger primitives (TPs) into the L1 Trigger System are shown in Fig. 4.8.

As previously mentioned, data are read from the ASICs on the front-end and then transmitted to the DAQ through the back-end. The definition of the output data from the BTL and ETL is described in Chapters 2 and 3, respectively. With the inclusion of the MTD in the L1 Trigger the back-end electronics would also be responsible for generating the L1 MTD TPs. Since the proton-proton collisions in a single bunch crossing are distributed over an rms of 180–200 ps, during trigger-primitive generation the timing information read out from a single channel is

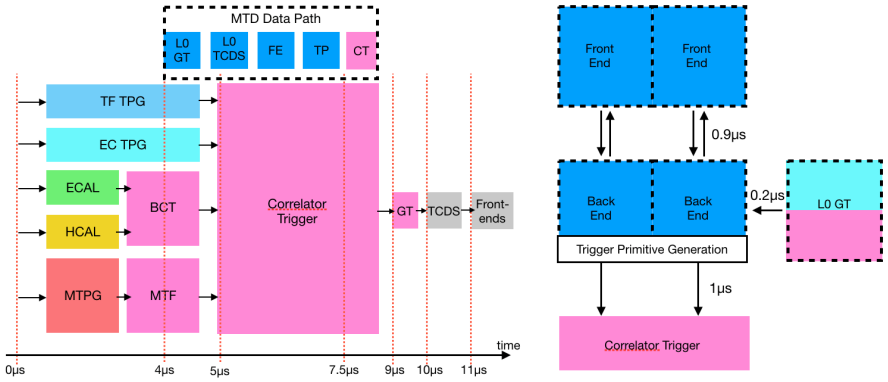


Figure 4.8: Latency map of the L1 trigger data path. The latency of the L1 trigger system from detector readout to TCDS is shown on the left. On the right, a diagram showing the MTD acquisition path from the L0 trigger to the MTD back-end, on to the MTD front-end, and then the return path through the MTD back-end and back to the L1 trigger for use in L1 algorithms.

assigned a bin value. The binning is defined such that the central bunch crossing is split into 18 bins where each bin corresponds to a 30 ps time window, six additional bins are defined for hits which appear outside of the central bunch crossing's 0.5 ns window and may later be identified as long-lived particles: 1 bin for particles which appear 0.5 ns late, 1 bin each for particles that appear 1, 2, 3 or 4 ns late and one bin for particles which appear after 4 ns. After the formation of the trigger primitive this data would be forwarded to the L1 Correlator to be correlated with the L1 physics objects, namely, charged and neutral particle candidates. Additionally, a copy of the data could be sent also to the L1 vertex finding electronics (not shown in the diagram) to reduce the inclusion of pileup tracks in the L1 primary vertex.

4.3.2 Decision points for the Level-1 MTD

To maintain the possibility for the MTD to participate in the Level-1 trigger, the capability to process a Level-0 accept should be included in the front-end ASICs. To minimize risks, the necessary changes to the ASIC logic and to the output FIFO should be integrated in the second to last iteration of the ASICs, which are scheduled for submission in November 2019 for BTL (TOFHIR2 version 1) and at the end of 2020 for ETL (ETROC2). For this reason, the Level-0 accept logic is included in the BTL baseline plan, while it will be considered for the ETL plan on the condition that a Level-1 timing trigger proposal will be presented with the CMS Trigger TDR at the beginning of 2020. The extension of the ASIC logic is also conditional to positive testing of the L1-A logic in the previous iteration of the ASICs.

To accommodate a Level-0-accept rate of the same order as the Level-1-accept trigger, the readout bandwidth of the MTD should be doubled, with additional VTRX+ links on the front-end boards, fibres, and back-end boards. As the BTL will not be accessible for repairs or replacement of the equipment for the entire Phase-2 operation, a redundant number of links and fibers for data control, matched to a corresponding number of back-end boards, is already included in the baseline design. Therefore any decision on the Level-1 timing trigger proposal has no impact on the BTL design and expected costs, while it would impact the ETL design, with an estimated additional cost of approximately 400 kCHF. The decision on the number of links and fibres is only needed for the finalization of the design of the front-end boards (not before Q3 2020) and can therefore be postponed to after the decisions made for Level-1 Trigger TDR.

Once we ensure to have the front-end capable to provide MTD triggers, changes to the back-

end may be decided and implemented later. At that stage, one would need to identify potential changes needed in the back-end FPGA to provide sufficient computational power, changes in the links to the L1 Trigger, costs, and resource implications. These aspects, being an upscope of the current MTD design, are not addressed in this TDR.

4.4 Detector control and safety system

The MTD detector safety system (DSS), whose goal is to protect the equipment against conditions that would harm it, will follow the same principles as the upgraded Endcap Calorimeter and Tracker detectors, and will be developed in close synergy with those projects. The MTD-DSS systems will be based on industrial Programmable Logic Controllers (PLCs). The MTD-DSS will constantly monitor the temperatures and humidities inside the detector volume and in its interfaces, and will have access to all status information from the cooling and dry gas systems. Temperature sensors inside the common Tracker and BTL volume, as well as the ETL volume, will be directly connected to the MTD-DSS PLCs. Each detector cooling loop will be monitored with a radiation tolerant Resistive Temperature Detector (RTD) on the inlet, outlet, and on the capillary. Additionally, the BTL will be equipped with two RTDs per readout unit, the ETL with five RTDs along each cooling loop within the wedges, and the ETL thermal screen with 64 RTDs, for a total of 352 RTDs. A dedicated sniffer system will pump gas from the MTD and Tracker volumes via several copper pipes towards the service cavern, where the gas will be analyzed by commercial high-precision dew-point meters. The MTD specific sniffer system will comprise eight sniffers per end. The MTD-DSS will also monitor the in-going gas quality (i.e. flow, pressure, humidity). Based on the temperature, humidity, and functional status of the service systems, the MTD-DSS can interlock individual power supplies, switch off the gas flow by controlling the input valves, or switch off cooling plants or force them to go to room temperature in case of high humidities. Given that the BTL and Tracker will share the same cold volume, any of these actions will be implemented to warrant the safety of both detector systems. In case of non-availability of information about any service or in case of failure of the MTD-DSS system itself, the MTD and Tracker will be brought into a safe state (Power OFF and interlocked and force warm the cooling plants), following the same logic (action matrix used by the DSS) as the Tracker. Leaving such an interlock state needs a human acknowledgement. A cold start up sequence will need to be executed to energize the MTD again. MTD OFF and cold is the normal safe state. The MTD-DSS will be a slave of the CMS DSS, having access to information like fire, rack failure, etc. The DSS can tell the MTD-DSS to switch off the detector and/or its services, for example via a button in the control room. Individual PLC racks will serve the different detector partitions. All PLCs will be supervised by a redundant PLC master system, which will not be connected to the network, for security reasons. All PLCs will be powered by UPS (Uninterruptible Power Supply). Limits on temperatures and humidities can be programmed individually, as well as which group of sensors switch off which part of the MTD, according to a programmable majority logic. The action matrix will be adaptable, but will always follow the most conservative and safest approach.

The MTD control system, MTD-CS, is the main interface to switch on and off the MTD in a safe and controlled way, and it provides all status information (power system, cooling system, gas system, temperature, etc.). It also provides the control needed to configure the detector electronics during operations and after shutdowns and outages. Unlike DSS, which has hardwired data inputs and interlocks, the MTD-CS system relies on the same high speed data links, the GBT chip set, used in the transmission of event data and is based on a commercial supervisory control and data acquisition (SCADA) system chosen by all LHC experiments for the HL-LHC era. The MTD-CS will monitor all power supplies and all services, for example cooling and

dry gas, and will have full access to all MTD-SS PLC information, for example all temperature and humidity readings. The MTD-CS will also prevent switch-on in case of non-conforming conditions. The system will also provide alarms and warnings via the common CMS interface to the shift crew. All set-parameters (low- and high voltage settings, trip limits, temperature limits, and alarm levels) will be stored in a configurations database, and all actual conditions will be stored in a conditions database, to allow historic viewing.

The DSS design and prototyping for the cold systems will begin in 2020 with MTD. For BTL, progressive system tests will take place at the TIF integration facility, which will start to be instrumented with equipment for installation in the second half of 2021. Final tests will be performed during the installation period in the first half of 2023, while the commissioning phase is slated to occur jointly with the Tracker system after the completion of Tracker installation. Similarly for ETL, progressive system tests will take place at the P5 integration facility ending with a full system test in the first half of 2024, before lowering the first full ETL disks to the underground cavern. Final commissioning will take place in situ in 2025.

4.5 CO_2 cooling system

The CMS CO_2 cooling system serves several CMS subdetectors: the Tracker (Outer Tracker, OT and Inner Tracker, IT), the Timing Layer Detectors (ETL and BTL) and the CE. The total cooling power, including ambient pickup, will be around 600 kW of CO_2 cooling power. The system features seven two-phase accumulator controlled loop (2PACL) cooling plants plus a spare one, installed in the USC-55 service cavern, from where pumped CO_2 is distributed to the detector through manifolds and patch panels (PP1), as schematically shown in Fig. 4.9. Out of the seven CO_2 plants, three are used for the MTD: one is fully dedicated to the BTL, two are shared between ETL and CE (one per each detector end). The 2PACL cooling units are all connected to a common primary chiller sitting on surface at P5. The primary chillers are directly cooled by cooling tower water. A CO_2 storage system is located on surface as well. This is a conditioned storage of CO_2 to supply or remove CO_2 from the underground systems. This feature allows a reduction in the size of the buffer tanks (accumulators) compared to the ones used for the standard 2PACL systems, thus saving space in the underground premises.

Each of the 2PACL cooling plants is connected to vacuum insulated coaxial transfer lines, distributing the CO_2 to the experimental cavern manifolds. Each plant's transfer lines serve two manifolds from where a first distribution is done towards the different detectors. The BTL is then fed through two manifolds, positioned on the X0 floor of the UXC-55 cavern and on the near side of the X4 floor, as shown in Fig. 4.10 (left). Each of the manifolds splits the flow into six parallel detector cooling segments. The ETL is fed through two manifolds per each YE1, provisionally located at the top of the YE1 structures, together with the CE ones, as shown in Fig. 4.10 (right). Each manifold splits the flow into two parallel cooling segments. From the BTL manifolds, the rigid, vacuum-insulated, coaxial transfer lines arrive at the patch panels inside CMS. At the patch panels, the concentric transfer lines separate into rigid inlet and return transfer lines which are routed up to the detector entrance.

In all CMS CO_2 cooled detectors, before entering the detector cooling loops, the liquid is preheated by dedicated devices, bringing it up beyond the saturation temperature and avoiding the super-heating phenomenon. The pressure drops from the detector loops back to the cooling plants are kept small where possible to minimize the difference between the accumulator temperature set point and the detector evaporator outlet temperature. The pressure determines the evaporator saturation temperature and therefore the actual cooling temperature in the detector cooling pipes. The pressure control is upstream of the pressure drop, so the controlled cooling

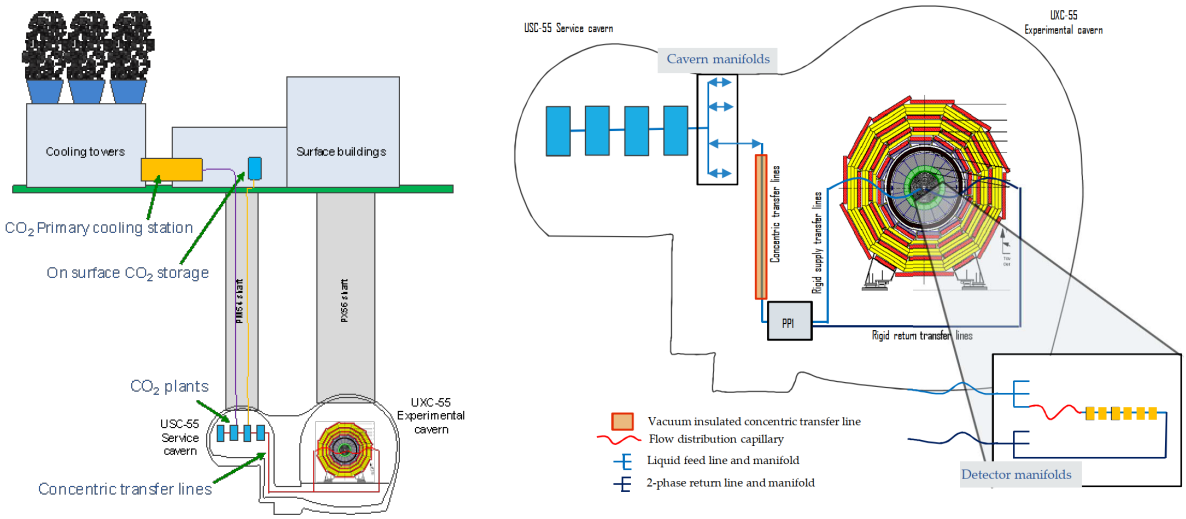


Figure 4.9: Schematic view of the CO_2 cooling system in the cavern and of the connection to the surface cooling (left), and zoomed view of the system in the cavern (right).

temperature is at the outlet of the cooling tubes. The inlet temperature is a function of the pressure drop in the detector and is therefore not controlled by the cooling system, but determined by the on-detector cooling loop design. The operating temperatures and the specifications for the cooling loops inside the detectors are discussed in Sections 2.4.4 and 3.8.4 for BTL and ETL, respectively.

In summary we have the following subcomponents, sections, and boundaries:

1. CO_2 storage tank at the surface and the vertical transfer lines between surface storage and plants;
2. The primary cooling plant placed on the surface with the evaporators down in the USC-55 cavern;
3. CO_2 cooling plants in USC-55, which comprise several identical plants with accumulators and a back-up plant;
4. Vacuum-insulated concentric transfer lines connecting the plants in USC-55 to the cavern manifolds of UXC-55;
5. Cavern manifolds;
6. Vacuum-insulated concentric transfer lines connecting the cavern manifolds to the detector patch panels (PP1s) inside the CMS vacuum tank;
7. Pipes (inside foam insulation) from the PP1s to the Tracker bulkheads;
8. Pipes inside the Tracker volume from the PP1 to the detector manifolds;
9. Detector manifolds next to each subdetector;
10. Capillaries used for the flow balancing on the input lines;
11. Evaporators on the detector local supports;
12. Return pipes to the detector manifolds.

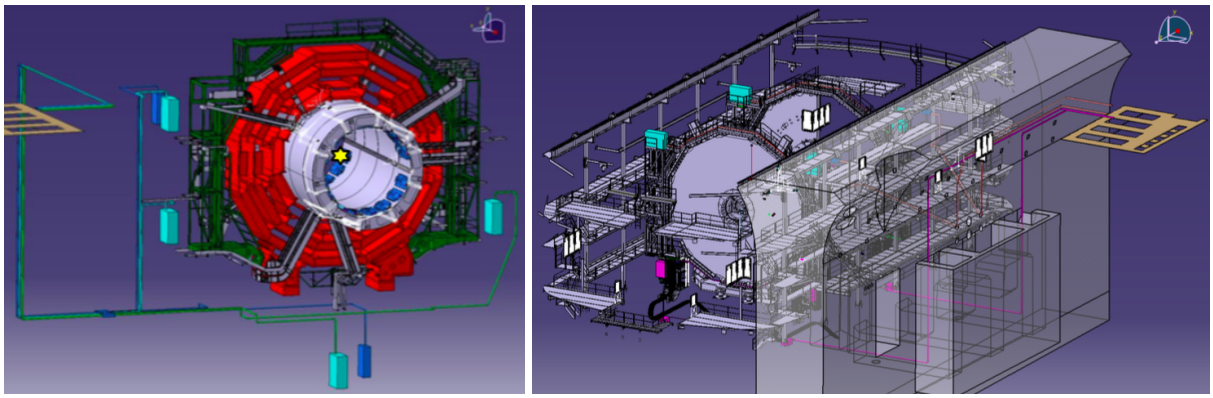


Figure 4.10: Layout of the transfer lines and the CO_2 manifolds serving the barrel (left) and the endcap detectors (right panel). In the left panel the BTL manifolds on the X0 floor and the X4 floor balconies are in blue; in the right panel, the CE/ETL manifolds at the top of the YE1 structure are in cyan.

The CERN EP-DT cooling team is mandated by ATLAS and CMS to design items 1 to 3, with CERN EN-CV being in charge of the primary chiller. There is a CERN oversight body for the CO_2 cooling system, comprising the ATLAS and CMS Technical and Cooling Coordinators as well as representatives of the CERN-EP and EN departments. The CMS cooling coordination team is supporting the CMS subdetectors for the design of items 4 to 6. Items 7 to 12 are the responsibility of the subdetector groups. Conceptual design of this hardware (8 and 9) is supported by the CMS cooling coordination team. The MTD requirements used — together with the requirements from other sub-systems — in the design of the CMS cooling system are described in Sections 2.4.4 an 3.8.4 for BTL and ETL, respectively.

A series of tests with prototypes and with the final system during the commissioning phase will be performed to validate the cooling system. The cooling plants are planned to be commissioned in standalone mode (no detector connected) after installation in the underground premises. The part of the system tested will go up to the experimental caverns manifolds, with dummy loads emulating the maximum detector power. The MTD specific CO_2 tests include a verification of the single cooling loop performance with a tray prototype for BTL (October 2019) and a wedge prototype for ETL (April 2020). During assembly, the cold checkout of the trays or modules will be performed in thermal chambers at the assembly centers, without the need of CO_2 cooling. Tests of the CO_2 cooling at the granularity of one cooling loop will be performed during the integration phase at the TIF for BTL, using the CO_2 plant available at the TIF. Testing of ETL wedges at P5 will be performed at the granularity of one cooling wedge. For that ETL testing, we plan to use one of the two CO_2 plants now serving the CMS Pixel detector that is not planned to be used for Phase-2 operation. The second plant gives a potential double redundancy in case of troubles.

The schedule of the common ATLAS and CMS Phase-2 CO_2 project, for parts under the responsibility of the CERN EN-CV, EP-DT and CMS cooling coordination team, is being optimized and is presented to the CERN Oversight Board dedicated to the project control. The full list of milestones is expected to be finalized in September 2019, including:

- Primary system purchase, delivery and installation
- 2PACL plants purchase, delivery and installation
- Cavern manifold purchase, construction and installation
- Main transfer lines (plant to manifold) and detector transfer lines: design, purchase,

installation

- 2PACL plants stand-alone test (primary with 2PACL plants, manifolds, and dummy loads)
- Connection to the detector & testing

According to the current plans — used to define the MTD spending profile for the required contribution to the purchase of the 2PACL plants, transfer lines, and manifold —, the plants will be ordered in 2021 and received, installed, and commissioned in 2022–2024, while the main transfer lines and manifolds are planned to be ordered in 2023, installed and commissioned in 2024. Direct links and cross-coordination between the MTD project, through its technical managers, and the CMS Technical Coordination and cooling coordination teams, described in Section 6.1.3, warrant prompt integration and updates of this information in the MTD project schedule.

Chapter 5

Reconstruction, performance and physics impact

5.1 Introduction

Studies have been performed on the impact of the MTD on the physics deliverables of CMS. The CMS event reconstruction relies on a Particle Flow algorithm [10] that provides the most global description of an event. With the addition of track-time information from the MTD, the event reconstruction is significantly improved. The time information from charged tracks is exploited in a space-time reconstruction of tracks and vertices (Section 5.2). Final state particles and observables are defined using vertices and track collections that are cleaned from spurious (pileup) tracks using space and time compatibility requirements (Section 5.3). The cumulative effect of the benefits on individual final state observables is quantified on a selected set of analyses of key physics processes of the HL-LHC program, such as precision measurements of the Higgs boson, the search for di-Higgs boson production, and the search for new signatures, including long-lived particles (Section 5.4). Particle identification from time-of-flight measurements with the MTD also provides unique opportunities in Heavy Ion physics.

Acceptance and efficiency studies, as well as the study of the track association with the time measurements in the MTD and the study of physics observables rely on a complete simulation of the MTD in the CMS Phase-2 detector using the GEANT package [113], with a detailed description of the MTD geometry (Section 5.2). The digitization process, with a complete simulation of the signal pulses, the leading edge discrimination and amplitude reconstruction, is based on the current design of the readout electronics and tuned using input from test beam data. The time information from the MTD, matched to the charged tracks and extrapolated to the vertex (Sections 5.2.5 and 5.2.6), is incorporated in the track information and used in a "time-aware" 4D-extension of the deterministic annealing technique of the CMS vertex reconstruction. Current results demonstrate that the back-propagation of the time information to the production vertex makes a negligible contribution to the time resolution and validate the reliability of the results from the fast-simulation approach adopted in the MTD Technical Proposal [8].

In this document, studies of the MTD impact on final state observables and on the analyses of specific physics processes rely either on full simulation or on the parametric fast-simulation model of Ref. [8], in which the time information is added to the CMS simulation and reconstruction workflow with an appropriate smearing of the simulated track time at the production vertex. The efficiency for track-time measurements is also included in the fast simulation. For some studies, the DELPHES simulation package [114] is used.

Where relevant – for example for final state observables such as particle isolation – the studies

have been performed as a function of interaction density in the LHC. In Run-2, the mean number of pileup collisions (PU) per beam crossing evolved from about 30 to 60, corresponding to a vertex line density of $\sim$0.3 and $\sim$0.6 mm^{-1}, respectively. The typical density for the HL-LHC scenario corresponding to 140 pileup interactions is $\sim$1.2 mm^{-1}, and at 200 pileup the density is $\sim$1.9 events mm^{-1}. The reconstruction performance and the physics impact have been studied for different scenarios, with per-track time resolution in the range between 30 ps and 70 ps. The barrel timing layer (BTL) provides a single measurement per track from the combination of two measurements at the two ends of the crystal bars. It is expected to achieve a per-track resolution of 30–40 ps at the beginning of the operation, degrading to about 50–60 ps at the end of HL-LHC. Physics results exploiting the full HL-LHC luminosity are shown for an average BTL resolution of 40 ps. Heavy Ion physics results are presented for 30 ps resolution, as Heavy Ion runs are slated to take place in the first period of MTD operation. The endcap timing layer (ETL) achieves a hermetic coverage with a double-disk structure. Each disk provides a measurement per track with 50 ps precision with about 85% coverage. The combination of two measurements provide 35 ps resolution per track for about 72% of the tracks. This performance is assumed in the simulation studies of physics results and can be maintained throughout HL-LHC operation by adjusting the operating voltage of the sensors or the preamplifier settings (Sections 3.2 and 3.3). Scenarios with degraded timing performance, both for BTL and ETL, and split to show the effect of the BTL and ETL coverage separately, are also presented for the sake of completeness.

Full-analysis studies are benchmarked assuming the 200 PU scenario. Only a selected number of full analyses are presented that touch on some of the most important signatures for the CMS physics program at the HL-LHC. The Phase-2 CMS detector, including the upgraded tracker, calorimeters, and muon systems, is used to derive a reference performance to which the differential gain from the MTD is compared. Unless explicitly stated, up-to-date algorithms optimized for the reconstruction of 200 pileup events with the upgraded CMS detector are used. A preview of the expected impact of the MTD was presented in Chapter 1 (Table 1.1).

5.2 Detector simulation and reconstruction

5.2.1 Detector Simulation

Event reconstruction studies and the study of some final state observables, such as particle identification, are based on a complete GEANT simulation of the MTD detector. The time and the position of the particles crossing the MTD are reconstructed from the energy deposited in the active detector elements of the BTL and the ETL.

The simulation of the BTL geometry is accurate. In Fig. 5.1 (left) the trays are shown within the Outer Tracker support tube, with a radial position of the crystals at 1174.5 mm from the beam axis. The BTL geometry model embedded in the CMSSW framework is based on 48 rows of modules per tray, each including 3 modules of 16$\times$1 crystals of transverse dimensions 57.6 $\times$ 3.15 mm^2 with gaps between them corresponding to the space needed by SiPMs. The crystal bars have their long side orthogonal to the z dimension of the CMS detector. The thickness of the crystals is 3.75, 3 and 2.4 mm, respectively, for each consecutive group of 16 modules moving from the center of the detector towards the endcap region. As shown in Fig. 5.1 (right), an aluminum plate with a surface of 92 $\times$ 52.2 mm^2 and 3 mm thick is placed below the crystals in front of the Tracker volume to simulate the support plate, while a PCB board of equivalent transverse size and 1.6 mm thick is placed 3 mm above the crystals to approximately simulate the readout electronics material budget.

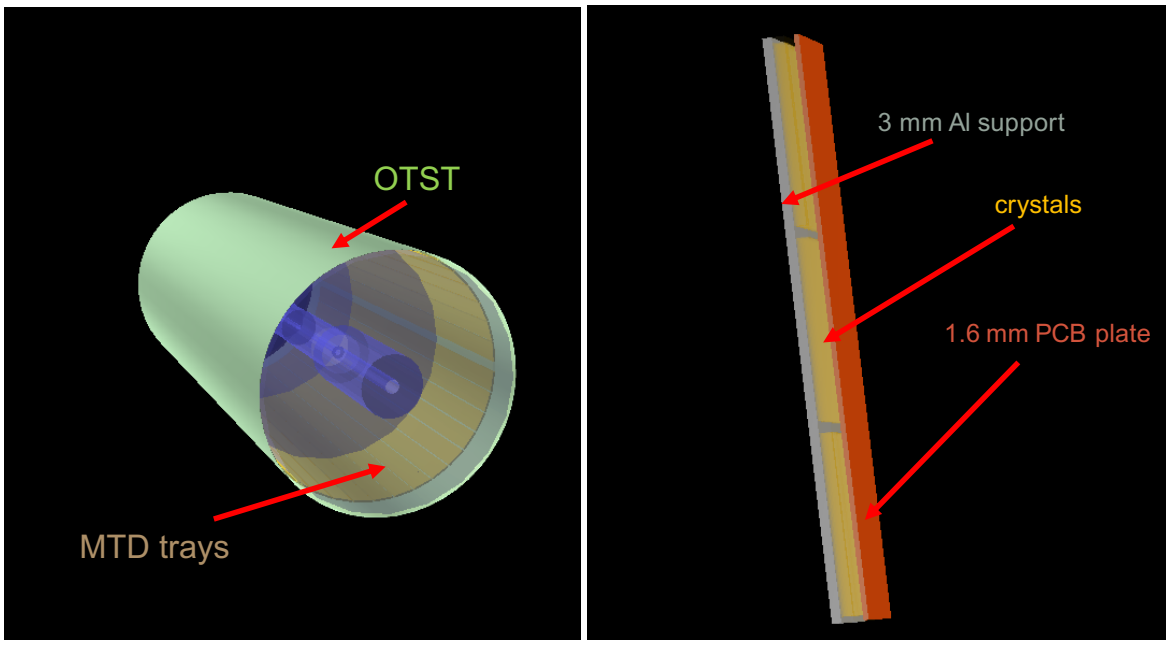


Figure 5.1: View of the simulated BTL trays within the Outer Tracker Support Tube structure (left) and detailed view of the simulated BTL module (right).

The ETL simulation is based on the geometry described in the MTD Technical Proposal [8]. The double-disk x-y layout described in this TDR could not be implemented because of a limitation in the software used to build the MTD geometry in GEANT. Since either ETL geometry provides a hermetic coverage, the interleaved double-disk design with radially oriented modules from the MTD Technical Proposal is adequate to describe the performance of the TDR layout. Compared to the TP design, as shown in the left panel of Fig. 5.2, 11 rings of modules have been used for this simulation in the radial acceptance 324–1291 mm, distributed on two disks placed in front of CE (right panel of Fig. 5.2) at 3038 and 3057 mm respectively from the interaction center. The residual difference between the TP and TDR designs lies in the material distribution of the support structures in the ETL, and in adjustments of the ETL position and radius. An approximate description of the support plate is provided by a set of aluminum disks in between the two sensors' disks. In both designs the total additional material from the ETL is less than 0.2 X_0 (radiation length), and the impact of the variations in material is minimal.

In the DELPHES simulation the effect of timing information is introduced either by a fast-simulation or by means of a parametric description of the efficiencies for signal and background processes applied to reconstructed objects. In the fast-simulation, the time of a charged track falling within the acceptance of the MTD is assigned directly from the time of the simulated track at the production vertex, spread by the time resolution at the vertex obtained from full simulation studies of the track reconstruction (Section 5.2.6). In the parametric simulation, the efficiencies are derived from the studies of the event reconstruction based on full simulation.

The studies of the case for inclusion of the MTD into the Level-1 trigger, presented in Section 4.3, are performed using two types of simulations. For the rates and efficiencies of isolated objects, the fast-simulation technique is used but applied to track-trigger candidates derived from the Phase-2 track-trigger emulator. This results in time stamps being applied to each found Level-1 track, and emulates the information that would be available in a region-of-interest trigger for the MTD, where each region of interest is a single Level-1 track. The Level-1 Trigger simulations pertaining to tagging long-lived particles are performed using the full simulation of the detector, and the corresponding hit digitization in the full reconstruction.

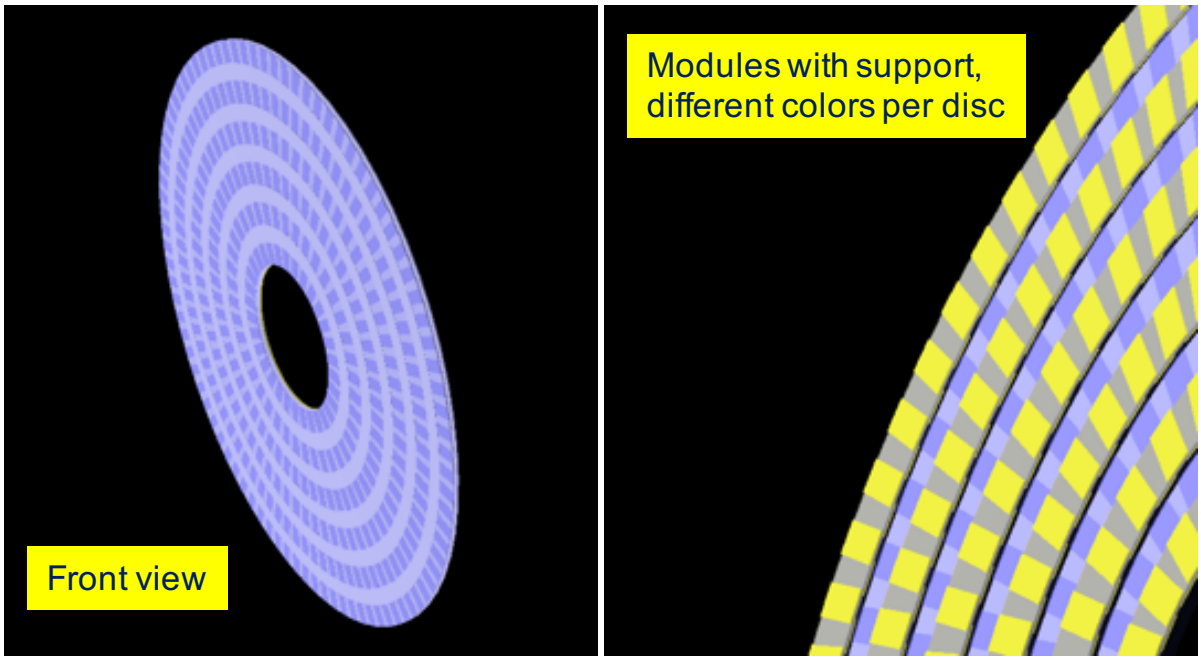


Figure 5.2: Front view of the simulated ETL structure (left) and detail of the ETL module rings without passive structures, with modules on different disks shown with different colors (right).

5.2.2 BTL simulation

The task of the BTL simulation is to convert the energy deposits and the respective arrival times in a crystal into time and energy measurements that incorporate the response of the detector and of the readout chain. The total amount of energy deposited in each crystal is estimated via GEANT simulation by adding all the energy deposits accumulated in the crystal within an interval of 25 ns from the bunch crossing. The signal amplitudes from the SiPMs at the two ends of the crystal bar are proportional to the deposited energy through the light yield (LY) of the crystals and the photon detection efficiency of the SiPMs (PDE), with a signal sharing depending on the longitudinal position of the energy position within the crystal bar. Table 5.1 lists the default values for these parameters, tuned on test beam and laboratory measurements discussed in Section 2.1.1. The numbers of photoelectrons read out on both sides of the crystal are independently fluctuated according to a Poisson distribution.

The simulation of the detected time of arrival (TOA) at a given BTL cell includes different effects: the time of arrival of the first energy deposit in the cell, the time latency for the scintillation light to reach the SiPMs at the crystal ends, and the time-walk effect depending on the signal amplitude and the discrimination threshold of the readout electronics. The time latencies from the hit position in the crystal to the left and right SiPMs are parametrized using test-beam data, which indicate an average light propagation velocity of about 8.5 ps/mm in the LYSO bars. The ability of the BTL readout chip (TOFHIR) to provide two time measurements per readout channel, either both on the leading edge or on the leading and trailing edges of the signal, is implemented in the simulation via a set of configurable parameters. The default configuration uses the lowest leading edge threshold to digitize the time of arrival. The time-walk is corrected using the amplitude measurement.

A reference pulse shape of the TOFHIR signal at the input of the discriminator (shown in Fig. 2.50) is used to calculate the time-walk delay. The pulse shape is based on a detailed simulation of the TOFHIR chip based on the CADENCE software as discussed in Section 2.3.2.2. The pulse amplitude is normalized to the number of photoelectrons and the measured time is

Parameter	Default value
LYSO light yield	40000 photons/ MeV
Light collection efficiency per SiPM	0.15
SiPM photon detection efficiency	0.20
Light propagation time in LYSO	8.5 ps/mm
TOFHIR discriminator thresholds	20, 50 p.e.
TOFHIR amplitude threshold	4 p.e.
TOFHIR electronic noise	1 p.e.
Number of ADC bits	10
ADC range	600 pC
Number of TDC bits	10
TDC least significant bit	20 ps

Table 5.1: Parameters used in the BTL simulation: the values have been tuned on test-beam and laboratory measurements.

set by the time when the pulse crosses a configurable threshold.

If multiple particles cross the same crystal, from either the same or different beam crossings (up to -3 BX are considered in the current simulation), a single pulse shape is generated, corresponding to the particle that has generated the earliest time deposit within the BX window. The effect of the noise induced from additional pulses at an average pile-up of 200 interactions has been studied with dedicated simulations and shown to be below 10 ps at the thresholds which are expected to be used for the TOFHIR electronics, including the light emission and propagation within the LYSO crystals.

Two time measurements are generated for the left and right SiPMs of a crystal, $t_{L,R}$, which are independently fluctuated with a Gaussian distribution using a global σ that includes all the contributions to the time resolution summed in quadrature according to Eqs. 2.1 and 2.2. Finally, the charge and time values are digitized according to the TOFHIR specifications: a 10-bit ADC with a dynamic range of 600 pC (corresponding to 15.6 MeV) and a 10-bit TDC with a granularity of 20 ps (time saturation at 20.46 ns).

As shown in Fig. 2.32 (Section 2.3), the BTL crystal occupancy at 200 pileup interactions, estimated from this simulation, ranges between 6 and 8%, for signals of amplitude higher than $E_{\text{thr}} = 1$ MeV (about 25% of the MIP most probable energy deposition), corresponding to the BTL readout threshold. At lower thresholds, the occupancy increases — about 24% for a threshold of 100 keV— because of low energy hits, mostly due to out-of-time interactions originating from back-scattered particles from the ECAL. According to simulation, these hits are sufficiently delayed with respect to direct tracks to not pile up on the rising edge of the signal, where the time stamp is obtained. Furthermore, the average energy deposits from such particles in the LYSO crystals is typically of the order of 5% of a MIP signal or less. Therefore, their impact on the amplitude measurement used for time-walk corrections is minimal. Dedicated simulation studies demonstrated that the impact of both out-of-time pileup and back-scattered particles on the time resolution is below 8 ps in quadrature.

The acceptance of the BTL has been estimated from the simulation of single muon events and amounts to about 90%. This efficiency is defined as the probability of finding a cluster of energy deposits (Section 5.2.4) larger than 3 MeV spatially associated to a single muon track of $p_{\text{T}} > 0.9\,\text{GeV}$. As shown in the left panel of Fig. 5.3, the inefficiency originates from the ϕ-periodic gaps between adjacent crystal matrices occupied by the SiPMs (two per tray), the gaps between

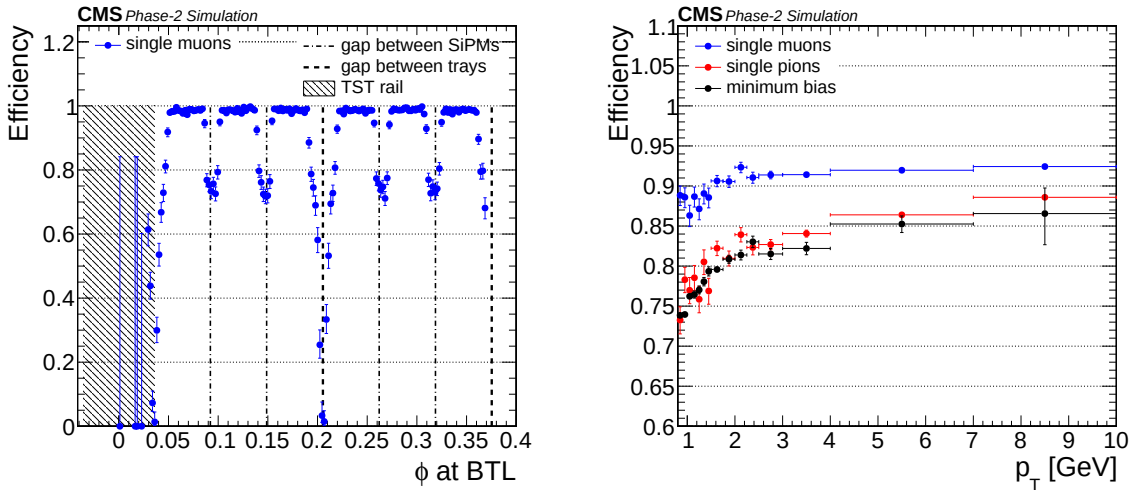


Figure 5.3: Per-track efficiency of finding an energy deposit larger than 3 MeV in the BTL and spatially associated to a single muon track. The left panel shows the efficiency as a function of the azimuthal track position at the BTL surface in a range corresponding to two trays. The right panel show the efficiency as a function of the p_{T} for single muon, single pion, and minimum bias tracks in blue, red, and black, respectively.

adjacent trays (36 in each half barrel), and the region taken by the TST rails at $\phi = 0$ and $\phi = \pi$. The right panel of Fig. 5.3 shows that hadrons suffer from an additional, p_{T}-dependent inefficiency because of nuclear interactions in the Tracker material upstream of the BTL.

5.2.3 ETL simulation

The ETL simulation follows the description of the geometry given in detail within the MTD Technical Proposal, and does not yet implement in detail the design proposed in Chapter 3. A single layer with eleven overlapping concentric rings of silicon modules is used to describe a detector that is hermetic in ϕ for the fiducial region of $1.5 < |\eta| < 3.0$. The small single hit inefficiency (about 1% for endcap sensors with 90% fill factor) is not propagated in the version of the simulation adopted in this document. A simplified approach is adopted to describe the timing performance of the design proposed in Chapter 3, with a single value of the time resolution used for all the tracks within the ETL acceptance. A time resolution 35 ps is assumed, equal to the combined resolution of a detector consisting of two planes with about 50 ps resolution for each plane. This should be a fair assessment of the average performance of the proposed ETL design, given a 90% fill factor and a single hit time resolution that will be better than 50 ps over a significant fraction of the acceptance.

The per-pad occupancy of the ETL depends on the radial distance to the beam line, but is linear with a moderate slope until the last 5 cm of radius from the support cone. It ranges from 0.1% at the largest radius to a maximum of 2% at the smallest radius. Simulated hits are digitized and recorded if their energy is larger than 0.1 MIP-equivalents at normal incidence. The timing resolution of the ETL is currently approximated using the latest beam test results and the expected performance of the ETROC chip, including the statistical improvement from having two measurements per track. This performance is implemented in the simulation of the ETL via a single-Gaussian smearing of the crossing time of the accumulated energy at the 0.1 MIP threshold, which corresponds to the zero-suppression threshold.

5.2.4 Reconstruction of deposited energy and time in the MTD

The energy deposited by a particle is reconstructed by clustering the energy deposits in adjacent cells matched to a track. A basic algorithm consists in summing the energy deposits above the readout threshold that lie within a cone of $\Delta R = 0.05$ around the extrapolation of a track trajectory to the MTD surface. The distributions of the total energy deposit predicted for single muons and pions, and for minimum bias events are shown in Fig. 5.4 for BTL (left) and ETL (right). For the sake of comparison, single muon and single pion events are weighted so that their p_{T} and η spectra match the ones of particles in the simulated minimum bias events. Muons behave as MIPs in the timing layer, and the most probable values of a Landau interpolation of the distribution of the energy deposits correspond to about 4.2 MeV in BTL and 0.15 MeV in ETL. In the case of hadrons, the distribution in BTL features a high energy tail ascribed to nuclear interactions.

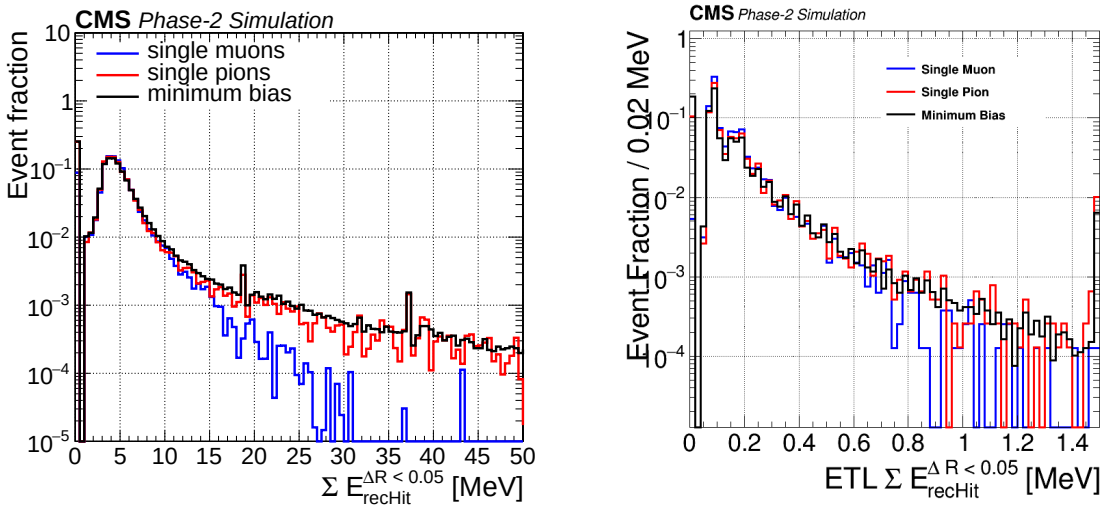


Figure 5.4: Left: Distribution of the energy deposited in the BTL as predicted by the simulation for single muons (blue), single pion (red), and minimum bias events. Right: Distribution of the energy deposited in the ETL as predicted by the simulation for single muons (blue) and single pions (red). Single muon and single pion events are weighted so that their p_T and η spectra match the spectra of particles in minimum bias events.

In the left panel of Fig. 5.5, the energy deposition is also shown as a function of the pseudorapidity for the BTL. This figure shows the average total energy deposited by a muon, as well as the average energy deposited in the crystal with highest energy in each cluster ("seed crystal") and the ratio of these two quantities as a function of the pseudorapidity. The energy deposition is fairly independent of η, because of the crystal slant-thickness leveling. The fraction of energy deposited in the seed crystal, shown in the same plot, decreases as a function of $|\eta|$ to about 60% at the end of the barrel, where tracks are more likely to cross adjacent crystals. The energy deposition is fairly uniform in ETL, where the sensor thickness is constant.

The time measurement associated to a cluster is obtained from the average of single-hits time measurements, weighted on the respective resolutions. In BTL, the time measurement for a single crystal is obtained from a linear combination of the left and the right SiPMs measurements, $t_{\mathrm{ave}} = C_L \cdot t_L + C_R \cdot t_R$; where C_L and C_R are two calibration coefficients that can be derived in situ from studying the time response of the two SiPMs versus the impact point. According to test beam results for particles at normal incidence (Section 2.1.1), the time measurement is independent of the track impact point along the crystal for $C_L = C_R = 0.5$. These values are

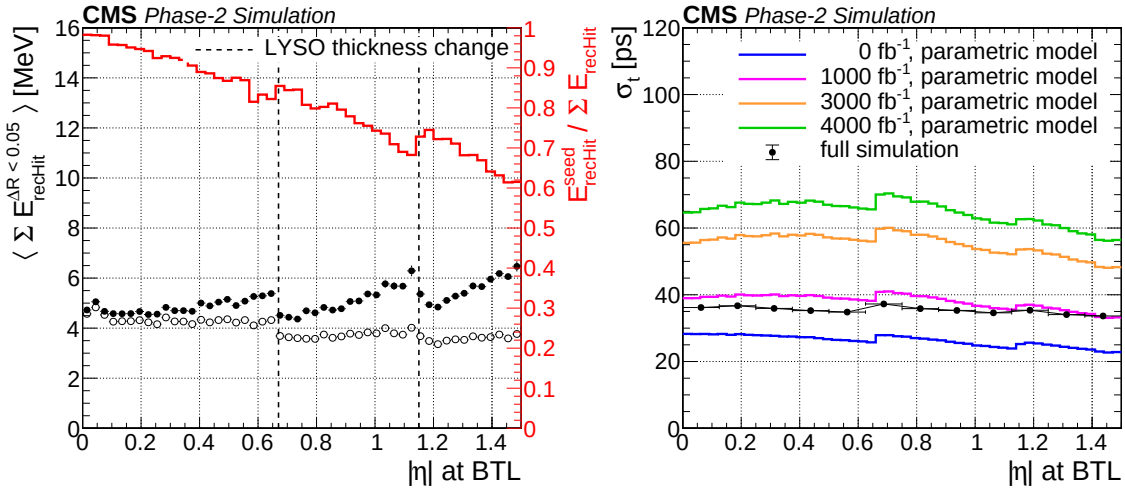


Figure 5.5: Left: Average total energy deposit (black solid dots), energy of the seed crystal (empty dots) and ratio of the two quantities (red line, right axis) as a function of pseudorapidity, for single muon events. Right: Average time resolution in BTL for clusters associated to tracks in minimum bias events for four ageing scenarios corresponding to 0 (blue), 1000 (purple), 3000 (orange), and 4000 fb^{-1} (green). The prediction of the parametric model is comparable to the full simulation for $\sim$ 1000 fb^{-1}.

adopted in the current simulation and in the studies reported in this TDR.

The expected average time resolution for clusters associated to tracks in minimum bias events is shown for BTL in Fig. 5.5 (right) for four ageing scenarios corresponding to 0, 1000, 3000, and 4000 fb^{-1}. The response evolution is parametrized according to the model described in Section 2.1, normalized to test beam results by means of a GEANT simulation of the test beam setup. The average per-track resolution for tracks with the p_{T} spectrum of minimum bias events is 30–40 ps or better across the full BTL acceptance up to 1000 fb^{-1}, while it degrades to 50–60 ps at the end of operations (3000 fb^{-1}). The average resolution during the HL-LHC operation (1500 fb^{-1}) is about 40–45 ps. The results from the full simulation corresponding to about 1000 fb^{-1} are used as a reference for physics studies.

Ageing scenarios are not implemented in the simulation of the ETL. It is expected that a resolution of 30–40 ps can be maintained throughout the HL-LHC era by adjusting the operating voltage and the gain of the preamplifier in the readout ASIC.

5.2.5 Tracking implementation

The track time at the collision vertex is obtained from the association of a time measurement in the MTD to a track and the subsequent time-of-flight correction to account for the track path length and the particle velocity. In the first step of the MTD track reconstruction, a simple topological clustering is performed to associate adjacent MTD hits above the readout threshold. Hits are required to be compatible in time, as this slightly improves the resolution of the cluster parameters at high pileup. The barycenter, weighted by the single hit energy, is used as an estimate of the cluster position and time. A comparison of the BTL cluster size and reconstructed time is shown in Fig. 5.6 for single muons in events with an average of 200 pileup collisions and without pileup.

In a following step, tracks that have been reconstructed using the pixel detector and strip Tracker are propagated to the MTD and spatially matched with compatible clusters. If a com-

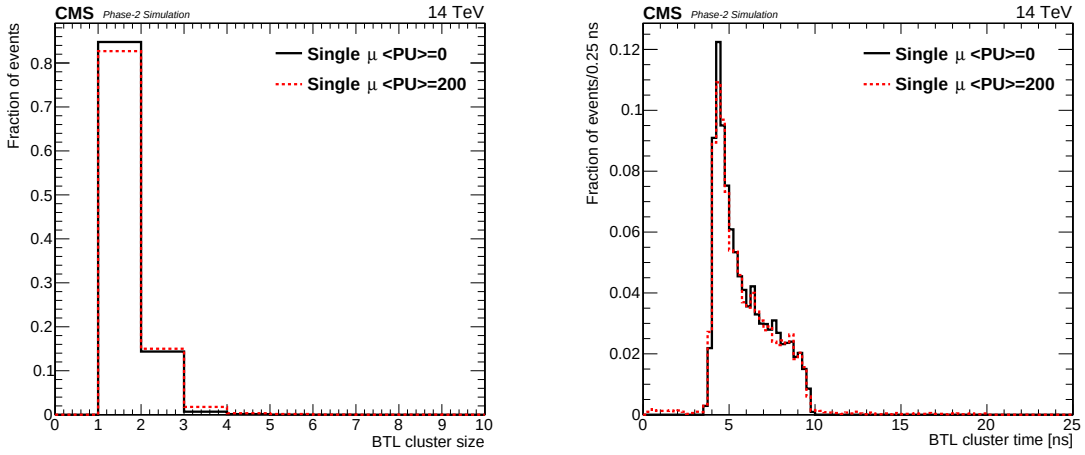


Figure 5.6: Comparison of the BTL cluster size (left) and reconstructed time (right) for single muons of p_{T} uniformly distributed in the interval $0.7 < p_{\mathrm{T}} < 10$ GeV, in events with an average of 200 pileup collisions (red dotted line) and without pileup (black line).

patible MTD cluster is found, the track parameters are refit including the MTD cluster position as an additional spatial measurement. The refitted track is then propagated from the point of closest approach to the beam line to the MTD front surface, one Tracker layer at a time, in order to compute the total path length. This path length is used, together with the particle velocity based on the refitted momentum and the pion mass hypothesis, to estimate the time-of-flight from the vertex. The MTD cluster time corrected for the time-of-flight provides the track time at the point of closest approach to the beam line.

The efficiency for matching a reconstructed track to an MTD cluster as a function of p_T and η is shown in Fig. 5.7 for single muons, with 200 pileup events and without pileup, and single pion events in the acceptance region of the MTD ($|\eta| < 3$). Particles are simulated with a flat spectrum in a 0.7–10 GeV p_{T} interval and the reconstructed tracks are matched to a generated particle in the event. The efficiency of matching a track to an MTD cluster is robust against the pileup level and independent of the track p_{T}. The efficiency, averaged over $|\eta|$, is above 90% for muons, and reflects the different geometric acceptance of the BTL and ETL. Within the respective angular coverage, the acceptance in the BTL is about 90% while the acceptance of the double-disk ETL layout is above 99%. The efficiency for pions is lowered to about 80% by nuclear interactions in the Tracker volume, as already discussed in Section 5.2.2.

The time of the cluster associated with the track and back-propagated to the beam line is compared to the simulation truth in Fig. 5.8, using pions in simulated top quark-antiquark ($\mathrm{t\bar{t}}$) events at 14 TeV. Tracks covering the BTL ($|\eta| < 1.5$) and ETL ($1.6 < |\eta| < 3.0$) acceptance regions are shown separately. The time resolutions extracted from a fit to these distributions are 36 ps and 37 ps for BTL and ETL, respectively, and match the simulated single-hit time resolution, thereby demonstrating that the uncertainty in the time-of-flight correction makes a negligible contribution to the resolution of the track time at the vertex.

5.2.6 Vertexing implementation

Vertex reconstruction in time and position along the beam line (tz plane) was studied using a time-aware extension of the deterministic annealing technique used for the CMS vertex reconstruction [13]. Since this technique can be extended to more than three dimensions, it is the natural choice for finding vertices in the dense environment of 200 pileup events. As outlined in Chapter 1 (Fig. 1.2), the introduction of time information significantly improves the perfor-

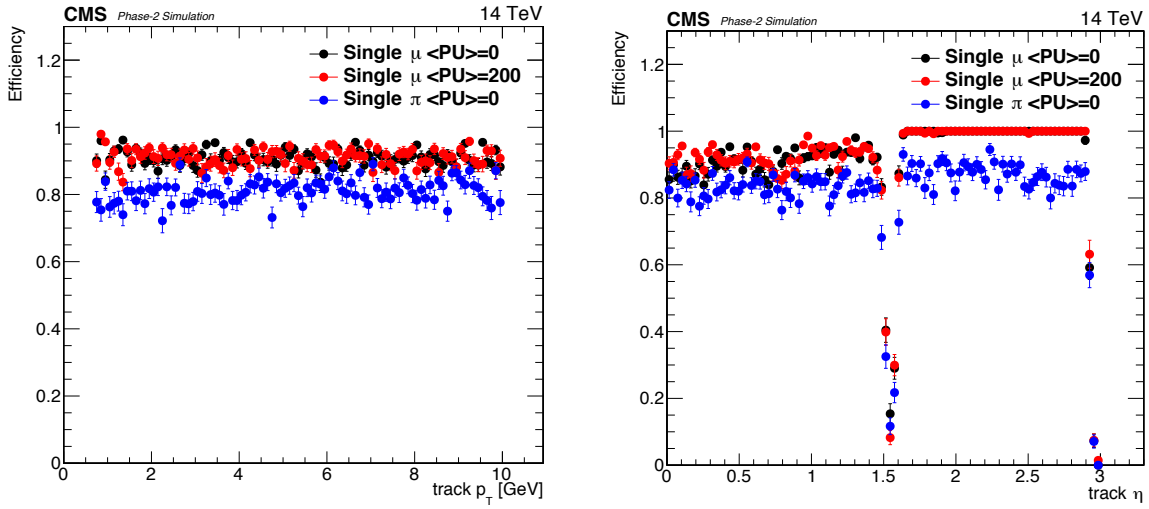


Figure 5.7: Per-track efficiency to find an MTD cluster associated with a reconstructed track as a function of transverse momentum (left) and pseudo-rapidity (right). Muons for events without pileup (black) and with an average of 200 pileup events (red) are compared to pions (blue).

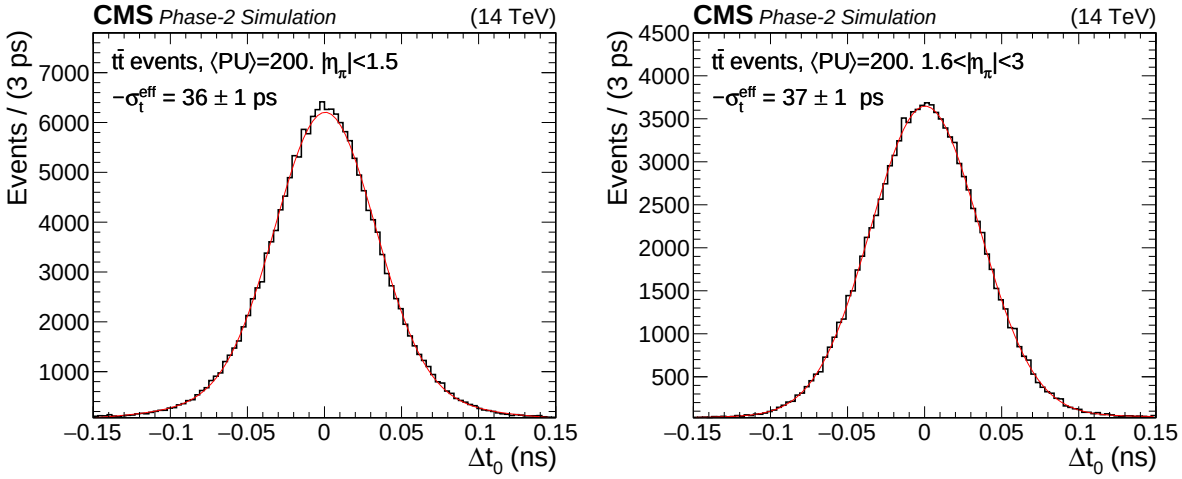


Figure 5.8: Distribution of the difference between the reconstructed time and the simulated time of a track at the vertex, for pions in simulated $t\bar{t}$ events at 14 TeV. The BTL (left) and ETL (right) acceptance regions are shown separately.

mance of the vertex reconstruction algorithm in associating tracks with vertices. For example, instances of vertex merging are reduced from 15% in space to 1% in space-time at 200 PU, while the rate of spurious splitting of vertices is negligible. Moreover, the space-time reconstruction capability has significant impact on several observables such as pileup jet ID, $p_{\mathrm{T}}^{\mathrm{miss}}$, b tagging, and lepton isolation, as further discussed in the next sections.

At momenta below a few GeV, the difference in time-of-flight between pions, kaons, and protons or heavier nuclear fragments becomes significant with respect to the time resolution of the detector. In order to address this, the 4D vertex reconstruction is carried out in two stages. In the first stage, the pion mass hypothesis is used to compute the time of each track at the beam line, but the uncertainty assigned to this measurement is inflated by adding in quadrature the difference in time of flight between the pion and proton mass hypothesis. After this initial reconstruction, the compatibility of tracks with the reconstructed vertices is tested in the

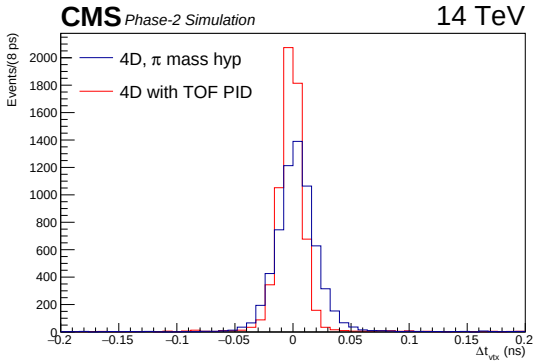


Figure 5.9: Distribution of the difference of the reconstructed and simulated time of the primary vertex in $t\bar{t}$ events at 200 PU. The outcome of the first step of the 4D-vertex reconstruction using the pion mass hypothesis (blue), and of the second step with particle identification corrections (red) are compared. The core of the distribution for the second step corresponds to a resolution of about 10 ps.

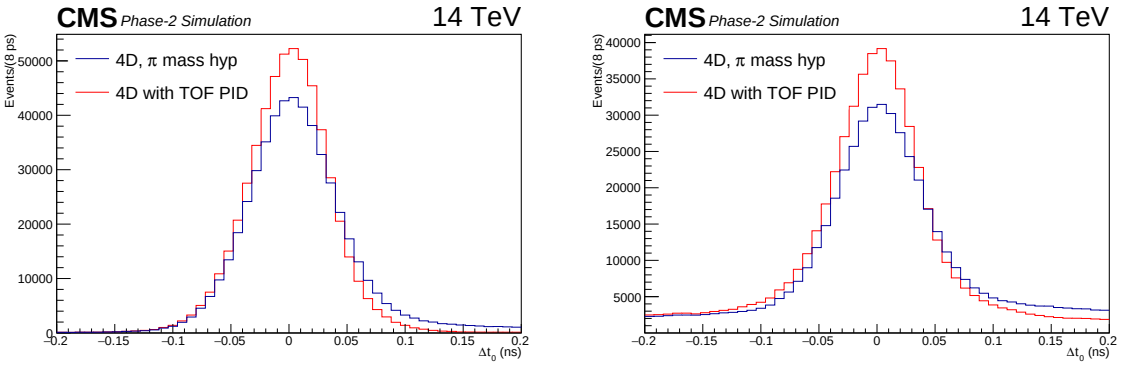


Figure 5.10: Distribution of the difference of the reconstructed and the simulated time at the vertex with (red) and without (blue) particle correction for tracks reconstructed within $|\Delta z| < 1$ mm of the generated hard interaction, in $t\bar{t}$ events. Events with no pileup (left) and 200 PU (right) are shown. The symmetric component of the tails in the pileup case includes tracks from pileup interactions that are reconstructed within 1 mm of the hard interaction.

tz plane under the kaon and proton mass hypotheses in addition to the nominal pion hypothesis. Tracks identified as pions, kaons or protons have the additional contribution from the mass hypothesis ambiguity removed from the uncertainty assigned to their time measurement, and those identified as kaons or protons have their time measurement at the beam line recomputed as appropriate. These recomputed time stamps and uncertainties are then used to run the 4D vertex reconstruction a second time. The time resolution of the 4D vertex reconstruction for $t\bar{t}$ events at 200 PU is shown in Fig. 5.9. After the second iteration, the vertex time is determined with an uncertainty of about 10 ps, which is small compared to the single track precision.

To further validate the particle identification step, the reconstructed time before and after the particle identification step is compared to the generated time of the hard interaction for tracks reconstructed within $|\Delta z| < 1$ mm of the generated hard interaction, as shown in Fig. 5.10. In the case of no pileup (left panel), the distribution of the difference between the reconstructed and the simulated time at the vertex has an asymmetric tail due to kaons and protons with a wrong mass hypothesis, which are misinterpreted as late pions. The tail is removed by the second step of the reconstruction algorithm and the resolution of the distribution is consistent with the resolution shown in Fig. 5.10 for pions only. The algorithm is effective also in the case

of 200 PU events, as shown in the right panel of Fig. 5.10. The residual symmetric component of the tails includes tracks from pileup interactions that are reconstructed within 1 mm of the hard interaction.

This method can be further extended to associate the individual photons reconstructed in the calorimeters to a collision vertex. The calorimeter timing hits for neutrals are combined to determine a reference time and position for the electromagnetic shower. The neutral particle is assigned a straight trajectory with time-of-flight corrections corresponding to the z-vertex positions. This generates a set of compatible vertices, which are a function of the time-of-flight and the vertex t_0 information.

A guiding illustration of how this approach works is provided in Fig. 5.11, showing how a pair of photons, reconstructed with 30 ps precision in the calorimeters, in a diphoton event is matched to a space-time vertex reconstructed from tracks. The green lines in the figure show, for each photon, the vertex time that would be needed for the photon to match a vertex at a given z position. The figures visually show that a pair of photons arriving from a common vertex provides enough timing information alone to reconstruct the vertex z-position only for photon pairs with a sufficiently large pseudorapidity gap. For photons close in pseudorapidity (typically $|\Delta\eta| < 0.8$ for a resolution of 30 ps), the identification of the correct production vertex of the pair requires a triple coincidence with a vertex t_0 from tracks. As visible in the figure, the two green lines alone form a small angle and do not identify a single point in the tz plane.

This capability opens several possibilities, including an improved identification of the diphoton vertex in $H \to \gamma\gamma$ decays, improved sensitivity to long-lived particles decaying into photons, improved measurement of the particle isolation from electromagnetic deposits, and improved resolution in the calorimetric missing transverse energy.

A significant example was presented in Ref. [8] for the $H \to \gamma\gamma$ decay channel. At 140 pileup events per beam crossing without timing information, the diphoton vertex is correctly identified only in about 40% of the cases, causing a sizeable degradation in the invariant mass resolution of the diphoton pair reconstruction. With precision timing in the calorimeters, the vertex can be correctly identified for about 55% of the $H \to \gamma\gamma$ decays exploiting the triangulation of the two photons as in the top panel of Fig. 5.11. The fraction of events with correct vertex identification is further increased to 75% with the addition of the information of the vertex time provided by the MTD, thus nearly fully recovering the present Run-2 performance. This acceptance gain corresponds to an increase of about 30% in the effective luminosity for this channel. An example of application of this technique to the search of displaced photons is discussed in Section 5.4.2.

5.2.7 Neutral particles time reconstruction in BTL

As described in the previous section, the 4D reconstruction can also be used in combination with calorimeter time reconstruction to associate energy deposits to their production vertices. For pileup rejection purposes, time measurements of low energy photons are important. While the upgraded endcap calorimeter will provide precision timing for photons down to low energies, the upgraded ECAL is expected to reach a precision better than 50 ps only for energy deposits above 30 GeV. At lower energies, the contribution from BTL can be decisive for pileup rejection in the barrel. According to simulation, more than 50% of the photons in the barrel region convert before reaching the calorimeter, either in the Tracker volume or in the BTL LYSO crystals.

The time reconstruction for calorimeter energy deposits starts from propagating their position

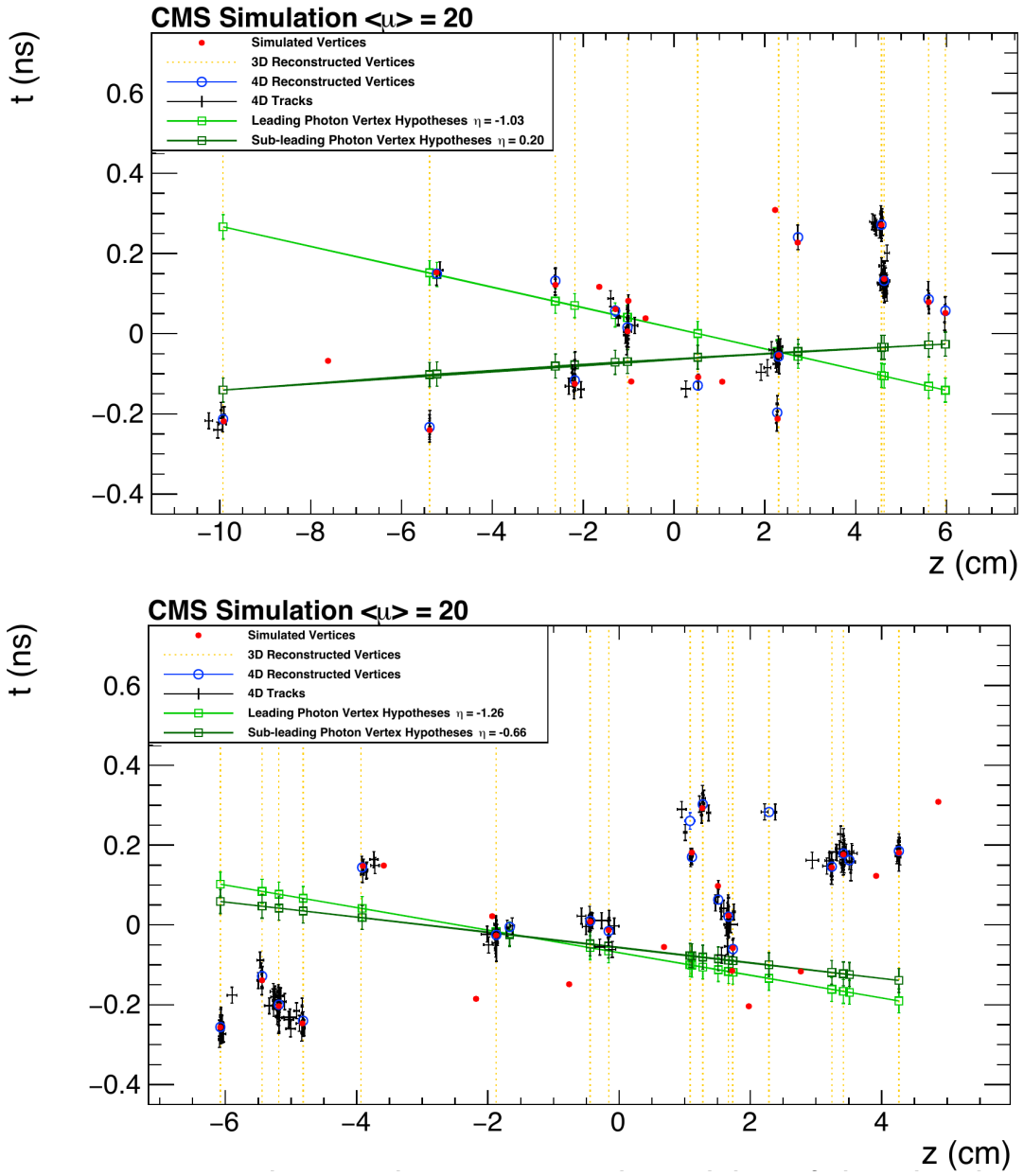


Figure 5.11: Space-time diagrams illustrating the concept of hermetic timing for H $\rightarrow \gamma\gamma$ events with two photons separated by a large rapidity gap (top) and a small rapidity gap (bottom). The reconstructed time for the photons at each vertex (green open dots), with error bars from the uncertainty on the time measurement of photons, can be cross referenced with the time information of the 4D vertices. The green straight lines are drawn to guide the eye. The pileup is reduced to an average of 20 in this case, to improve clarity. For photons with a small rapidity gap, the coincidence with a 4D-vertex is necessary to enable vertex location.

to the MTD surface using a straight trajectory hypothesis. Compatible clusters are searched in a narrow region in the η direction and a wider ϕ window (to account for conversions in the Tracker region). Clusters are sorted according to their spatial and time compatibility given a vertex hypothesis. An effective correction for the time-of-flight is applied based on the distance in the ϕ direction between the MTD cluster position and the calorimeter position.

The efficiency for time reconstruction of an ECAL energy deposit with transverse energy $E_T > 1$ GeV is shown in Fig. 5.12 (left) as a function of the pseudorapidity. The efficiency ranges from about 30% at the center of the barrel to around 55% at the end of the barrel. This value can be directly compared with the probability for a photon to convert either in the Tracker volume or in the BTL crystals, which goes from 40% to 70% in the same interval (the fraction of photons which are converting directly in the LYSO crystals is about 20% and is flat as a function of the pseudo-rapidity). The time resolution is found to be 35 ps as shown in Fig. 5.12 (right). According to simulation, the removal of the energy deposits with a time measurement incompatible with the primary vertex time reduces the total pileup energy in the events by about 30%. This reduction may improve the calorimeter isolation efficiency, the jet energy resolution and the missing transverse energy reconstruction at high pileup.

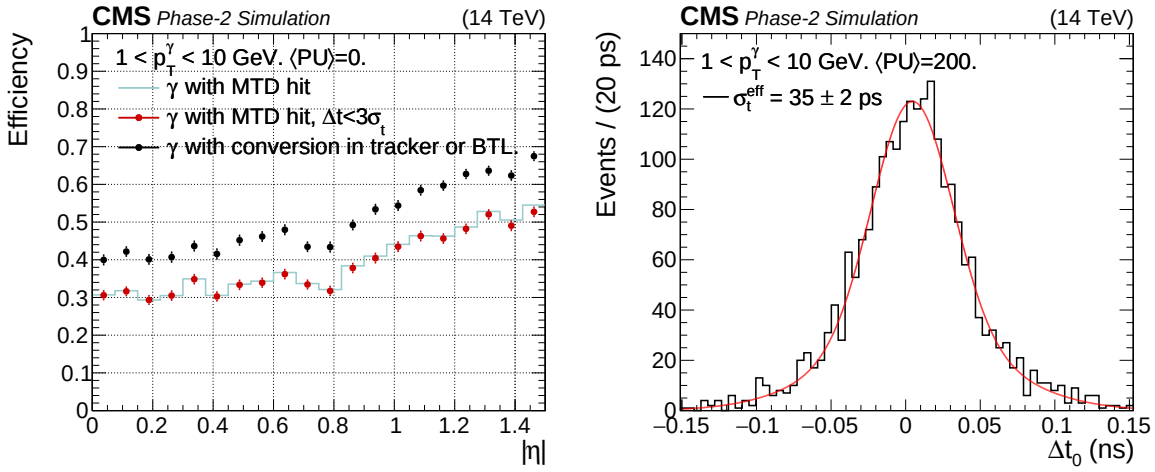


Figure 5.12: Left: Efficiency to assign time for an ECAL barrel energy deposit with $E_T > 1$ GeV as a function of pseudo-rapidity. Single photon events with a flat p_T spectrum between 1 and 10 GeV are used. Black dots correspond to the probability that a photon converts within the BTL including conversions in the LYSO crystals, the light blue line shows the probability to find an associated BTL cluster, the red dots is further requiring that the reconstructed time is within 3σ of the generated vertex time. Right: The time reconstructed time for neutrals in BTL is confronted with the vertex time. The time resolution extracted from a fit to the observed distribution is 35 ps.

5.3 Performance in the reconstruction of final state observables

Final state observables, often referred to as "physics objects", include: the collision vertices, among which is the primary vertex associated with the collision of interest; the individual tracks associated with a vertex; the identified particles — such as leptons and photons —; jets obtained from the algorithmic combination of several charged and neutral particle-flow candidates; jets of identified flavor; and the missing transverse momentum, defined as the negative vector sum of the p_T of the jets originating from the primary vertex. The studies presented here extend and update the results of Ref. [8].

5.3.1 Rejection of tracks from pileup interactions

An essential step in pileup rejection is to exclude from relevant quantities charged particles that are not associated with the hard interaction. This step is critical to maintain the performance of prompt lepton identification and b-tagging, as well as of the reconstruction of the charged component of jets and of the missing transverse momentum. One method commonly used in CMS to associate tracks with the hard primary vertex, PV, is a simple selection on the distance between the track and the vertex, $|\Delta z(\text{track}, \text{PV})| < 1$ mm, which was tuned to maintain high efficiency for charged particles actually originating from the hard interaction and reject misidentified or non-prompt tracks. Hence, for an interaction density in excess of 1.0 mm^{-1}, corresponding to approximately 100 pileup interactions, some or many of the resulting tracks will be associated with the hard primary vertex, contaminating the set of tracks used to calculate the relevant physics quantities and degrading the performance.

This effect is directly quantified as a function of the line density of events along the beam line, which drives the probability that additional vertices will be close enough in space to the hard interaction to contaminate the set of associated tracks with those from pileup. This association method can be extended with precision timing, by adding a requirement on the time distance $|\Delta t(\text{track}, \text{PV})| < N \times \sigma_t^{\text{track}}$ for those tracks with valid timing information. Tracks without valid timing information are retained. In this study, $N = 3$ is used together with an average time resolution $\sigma_t = 35$ ps for tracks with valid timing information, corresponding to the requirement $|\Delta t(\text{track}, \text{PV})| < 105$ ps.

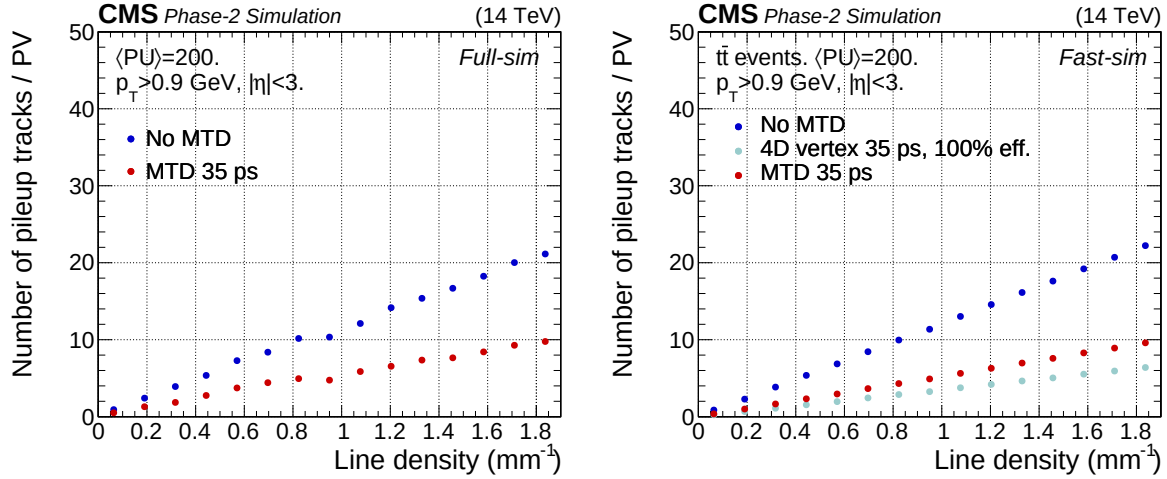


Figure 5.13: Fraction of pileup tracks incorrectly associated with the hard primary vertex in $t\bar{t}$ events in the full (left) and fast (right) simulation as a function of the the pileup density. Shown are data for the 3D (blue) reconstruction and the 4D (red) reconstruction including the MTD time information. For fast simulation, the results of an ideal detector with 100% efficiency as assumed in Ref. [8] are also shown (light blue).

Using this association criterion, the efficiency for reconstructing and associating tracks from pileup interactions in $t\bar{t}$ events is shown in Fig. 5.13, both with and without precision timing from the MTD, as a function of the event line density. The left panel of the figure shows the results obtained using the most recent version of the full simulation, which provides an average resolution of 35 ps. The right panel shows the results from a fast simulation where the track time at vertex is assigned with a fixed resolution of 35 ps, as done in Ref. [8]. In this case, the impact of time cleaning is presented both for an ideal detector with 100% acceptance, reproducing the studies of Ref. [8], and for a detector with an efficiency of 85%, which emulates the MTD performance from full simulation. In these conditions, the improvements in the track purity of

the primary vertex compares well with the results of the full simulation. This result provides confidence in the reliability of the fast simulation adopted for some of the performance analyses presented in the next sections, as well as in the study of the track purity of the primary vertex for different assumptions on the time resolution (Chapter 1, Fig. 1.2).

The effect of timing on PV tracks has been studied as well in zero-pileup and no significant impact on the association efficiency is observed. In full simulation, with 200 PU events, a fraction of about 3% of the tracks from the primary vertex has an incorrect time measurement from a mis-association with an MTD cluster. While this effect does not impact significantly on the evaluation of the MTD performance, an improved algorithm for track-to-cluster association is under study.

More generally, the optimal $|\Delta z|$ selection window depends on whether the background sources for the process under study include tracks that are genuinely displaced from the PV, as in $\mathrm{t\bar{t}}$ events, or not. A selection window of 1 mm, widely adopted in CMS Phase-2 simulation studies, provides the best rejection of misidentified muons in light-flavor jets using the charged isolation sum, defined as the ratio of the sum of the p_{T}'s of charged particles in a small cone around the muon candidate over the muon p_{T}. For non-prompt muons in $\mathrm{t\bar{t}}$ events a selection of 1 mm is close to optimal in the barrel, while in the endcaps, a slightly looser selection provides a better separation. In the following, the selection window for the no-MTD case is optimized depending on the dominant background source for the process under study, not to artificially amplify the gain from the inclusion of time information.

5.3.2 Jet and missing transverse momentum

5.3.2.1 Pileup jet suppression

In the presence of pileup, soft jets and underlying event activity from multiple pileup interactions may overlap and be clustered into a higher energy jet ("pileup jet"), serving as an additional background for e.g. Vector Boson Fusion (VBF) tagging and other final states with jets. For present Run-2 conditions, this background is largely suppressed by cleaning the charged particles based on spatial association with the primary vertex, and pileup jets are mainly an issue beyond the present tracking acceptance. For Phase-2, the tracking acceptance of the CMS detector will be extended. However, the higher pileup density and corresponding increase in charged particles from pileup associated to the primary vertex leads to a non-negligible rate of pileup jets, which can be reduced with the MTD.

The rate of pileup jets is studied in $\mathrm{Z} \to \mu\mu$ events with 200 pileup conditions, with and without precision timing for the charged particles, using jets of $p_{\mathrm{T}} > 30$ GeV reconstructed with the "Pileup Per Particle Identification" (PUPPI) algorithm [11], which calculates the likelihood that each particle originates from a pileup interaction, and the anti-$\mathrm{k_T}$ clustering with 0.4 distance parameter [115]. The PUPPI algorithm, widely used in the context of the CMS Phase-2 upgrade performance evaluation for its resilience to pileup, currently uses the first primary vertex in the reconstructed collection. In order to avoid events where the incorrect primary vertex is used, two reconstructed muons with $p_{\mathrm{T}} > 20$ GeV and associated with the first primary vertex are required for events entering the study. Tracks are associated with the primary vertex with $|\Delta z(\text{track}, \text{vertex})| < 1$ mm; an additional requirement of $|\Delta t(\text{track}, \text{vertex})| < 3\,\sigma_t^{\text{track}}$ is applied when considering the precision timing case.

Signal jets are defined as reconstructed jets that are matched to a generator-level jet with $p_{\mathrm{T}} > 4$ GeV within a cone of $\Delta R < 0.2$, while pileup jets are defined as reconstructed jets that are not matched to a generator-level jet with $p_{\mathrm{T}} > 4$ GeV within a cone of $\Delta R < 0.6$. The relative rate

of both signal jets and pileup jets with and without precision timing for the charged particles is shown in Fig. 5.14, for different assumptions on the MTD time resolution. For a resolution of 30–40 ps, representative of the MTD performance up to about 1000 fb fb^{-1} for BTL and up to the end of the HL-LHC for ETL, the addition of precision timing reduces the rate of pileup jets by 25−40% depending on pseudorapidity with minimal effect on the signal jet rate. The working point and the rate reduction from timing may be further optimized, also with the inclusion of the timing information for neutrals.

A more modest reduction of the pileup jet rate is expected for resolutions of 60 ps, which is illustrative of the case of extreme BTL degradation after 4000 fb^{-1} or of an ETL detector with a single disk, providing a resolution of 50 ps per track with a coverage of 85%.

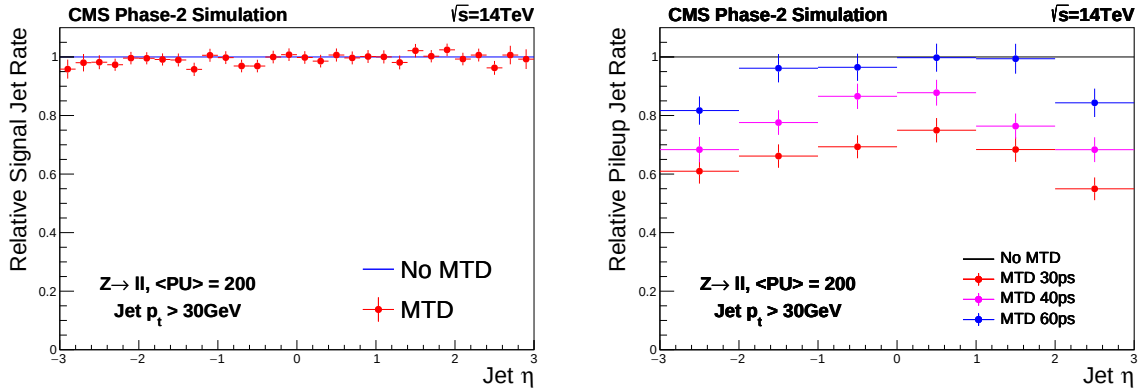


Figure 5.14: Rate of signal jets (left) and pileup jets (right) of $p_{\text{T}} > 30$ GeV, reconstructed with the PUPPI algorithm and the anti-k_{T} clustering, with precision timing for the charged particles relative to the no timing case. Three different time resolutions are shown.

The reduction in pileup jet rate is expected to have a significant impact on signal extraction cases that rely on jet multiplicity event categorization, central or forward jet vetoes, or forward jet tagging, as in the case of VBF topologies. The benefit of the reduction is more relevant in the endcap regions, where the rate of pileup jets is larger. As discussed in the CMS Upgrade Scope Document [3], the rate of jets reconstructed from pileup energy depositions is observed to increase up to 30% in the endcaps with the Phase-2 CMS detector. For measurements of the VBF Higgs boson production in the $\tau\tau$ final state, for example, this effect alone degrades the analysis performance expressed as signal over the square-root of background ($S/\sqrt{B}$) by 25%, because of a reduction of the signal yield and the increase of the background from Drell–Yan production. Track-timing with the MTD provides a reduction of the pileup jet rate that offsets this performance degradation.

5.3.2.2 Missing transverse momentum

The correct reconstruction of the missing transverse momentum, $p_{\text{T}}^{\text{miss}}$, of the hard collision is key to several measurements, such as Higgs boson decays to τ pairs ($\text{H} \to \tau\tau$) — where the $\tau\tau$ invariant mass is reconstructed using the missing transverse momentum — or $p_{\text{T}}^{\text{miss}}$-based searches for new phenomena, including searches for neutralino production, WIMP-like dark matter production, and stable massive dark sector particles.

In the Phase-2 CMS detector without MTD, the $p_{\text{T}}^{\text{miss}}$ resolution is expected to worsen by 15% in the transition from 140 to 200 pileup. This effect causes a 15% degradation of the $S/\sqrt{B}$ performance of the VBF–$\text{H} \to \tau\tau$ measurements [3], and translates into a 40% increase in the luminosity needed to achieve the equivalent result at 200 pileup. Similarly, the degraded

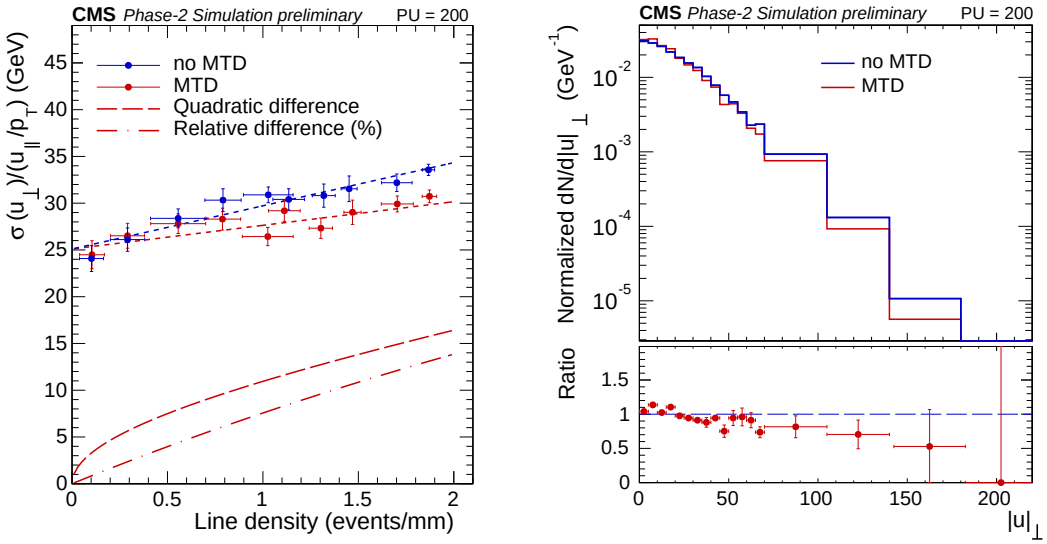


Figure 5.15: Left: Resolution of the hadronic recoil component perpendicular to the Z boson p_T as a function of pileup density, which characterizes the contribution to the p_T^{miss} resolution of noise, including pileup. The dotted lines show linear fits to each set of data points; the dashed and the dot-dashed lines show, respectively, the difference in resolution in quadrature and in percent between the timing and no-timing case. Right: Distribution from simulation at 200 PU of the PUPPI hadronic recoil component transverse to the Z boson p_T, with and without precision timing in $Z \to \mu\mu$ events.

p_T^{miss} resolution leads to a significant reduction of the discovery potential in searches. For instance, the discovery potential for electroweak SUSY chargino–neutralino production in the $W^{\pm}H + p_T^{miss}$ final state is lower at 200 pileup than at 140 pileup [3]. As we discuss below, the MTD with precision timing reconstruction offsets these performance degradation and makes a substantial impact on the HL-LHC physics program at the HL-LHC, as summarized in Table 1.1 of Chapter 1 and detailed in Ref. [8].

The performance of p_T^{miss} reconstruction is studied with and without the addition of precision timing information for the charged particles entering the p_T^{miss} calculation. In this study, a resolution of 30–40 ps is used. The p_T^{miss} is computed from both the charged and neutral particles using the PUPPI algorithm, with an optimized weighting scheme [11]. A standard technique for the characterization of the p_T^{miss} performance is to study the scale and resolution of the hadronic recoil against the Z boson in $Z \to \mu\mu$ events. For the purposes of this study, tracks are associated with the hard interaction vertex by requiring $|\Delta z(\text{track, vertex})| < 1$ mm, with an additional requirement of $|\Delta t(\text{track, vertex})| < 90$ ps in the case that precision timing is available.

It has been verified that timing selection for the charged particles does not change the scale for missing p_T computed with PUPPI, indicating that the charged particles of the jet are not being removed by the additional requirement of compatibility in time with the hard interaction vertex. In the absence of precision timing, the missing p_T exhibits a degradation in resolution as a function of pileup density (left panel of Fig. 5.15), due to additional charged particles from nearby (in z) pileup interactions contaminating the hadronic recoil sum. The addition of precision timing for the charged particles reduces the slope and improves the resolution for high event densities. The resolution at the average vertex density corresponding to 200 PU is improved by ∼10–15% with the MTD, and is equivalent to the resolution without a timing

detector at 140 PU.

In order to examine the effect of the resolution improvement on the tails of the $p_{\mathrm{T}}^{\mathrm{miss}}$ distribution, often relevant for new physics searches involving particles invisible to the detector, the distributions of the PUPPI transverse recoil component in $\mathrm{Z} \to \mu\mu$ events with no real missing momentum are shown in the right panel of Fig. 5.15. This distribution shows a reduction of $\sim 35 \pm 16\%$ in the rate of events with $p_{\mathrm{T}}^{\mathrm{miss}} > 110$ GeV, with the addition of precision timing. This is relevant towards reducing the background for searches and helping reduce high level trigger rates, and restores the performance and the discovery potential of the 140 pileup operation.

Beyond the improvements to the $p_{\mathrm{T}}^{\mathrm{miss}}$ performance with the addition of precision timing for charged particles, significant improvements are also expected from the addition of precision timing for the neutral electromagnetic component of the missing momentum. Approximately 45% of photons in the barrel have at least one hit in the barrel timing layer, due to conversions in the Tracker or in the timing layer volume itself. The MTD therefore will provide precise timing information even at lower energies, where the time measurement from the upgraded ECAL barrel is less precise. Furthermore, the efficient and precise measurement of the hard interaction time enabled by the precision timing layers will allow timing information for photons from the endcap calorimeter to be maximally exploited.

5.3.3 Heavy-flavor tagging

In very high pileup conditions, secondary vertex b-tagging is degraded by the formation of spurious secondary vertices caused by pileup tracks, reducing the ability to distinguish signal from background. This degradation is seen clearly in Fig. 5.16 and depends on the average pileup, and pileup density. At tight working point (for light jet misidentification probability of 0.001), the b jet tagging efficiency is decreased by about 10% at high pileup compared to the no-pileup case. In order to mitigate this problem, the secondary vertexing algorithms were updated to be aware of timing information from the MTD. By requiring tracks to be within $3.5\sigma_t$ ($\sigma_t = 30$–40 ps) of the selected primary vertex, the number of spurious reconstructed secondary vertices was reduced by 30%. This causes a significant improvement of the separation between b jets and jets from light-flavor quarks, as quantified by the receiver operating characteristics (ROC) curve of the efficiency for b jets versus the background rejection of light quark jets. An equivalent way of representing the ROC curve, where the efficiency for background is used instead of the rejection power, is shown in Fig. 5.16. As visible the ROC curves improves significantly, especially for tighter working points where near-zero-pileup performance is achieved. The impact of a resolution degradation to 60 ps is also shown in the figure. In addition, as shown in the right panel of Fig. 5.17, the dependence of the b-tagging efficiency on the pileup density is removed. Further gain is expected from retraining the b-tagging discriminants for 200 PU conditions, exploiting the additional information from timing in a consistent manner.

These performance benefits have a particularly important impact on the signal yield in anayses sensitive to acceptance increases, such as di-Higgs boson production, which is searched by combining several decay modes of the Higgs boson pairs, with one boson always decaying into $\mathrm{b\bar{b}}$ quarks. In the analyses for di-Higgs boson searches with the CMS Phase-2 detector without MTD, the b-tagging selection corresponds to a 1% misidentification probability for light quarks [4, 6]. The MTDbased reconstruction would provide, for the same background of misidentified b jets, an increase in the b-tagging efficiency in the range 4-6%, depending on the pseudorapidity as shown in Fig. 5.17 (left). The gains in signal acceptance at the single-

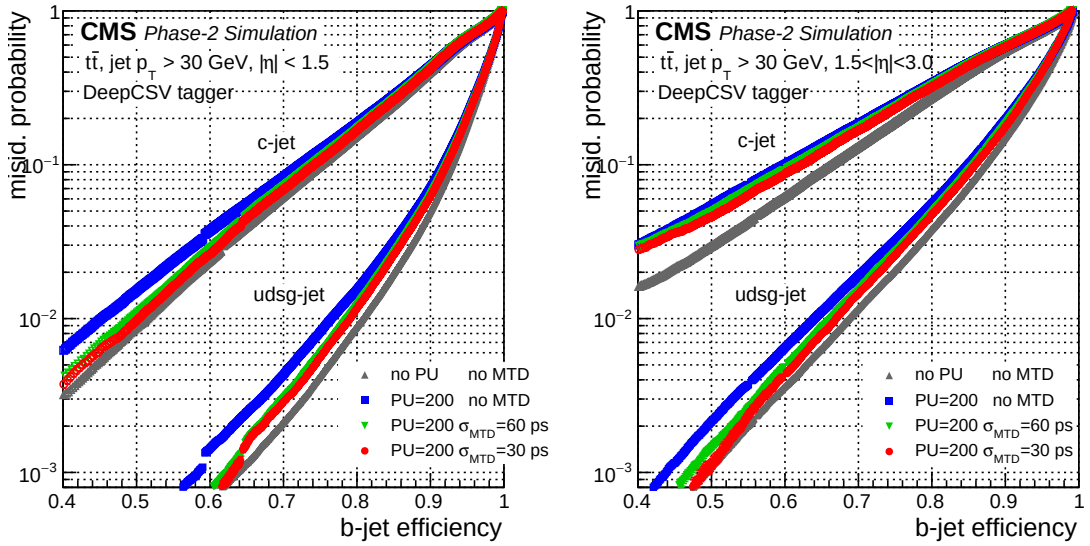


Figure 5.16: Secondary vertex tagging ROC curves for light and charm jets for $|\eta| < 1.5$ (left) and for $1.5 < |\eta| < 3.0$ (right). Results with without (blue) and with timing for 30 (red) and 60 ps (green) resolution hypotheses are compared to the zero pileup case (grey).

object level combine in multi-object final states. For instance, in the final state with four b quarks, HH $\rightarrow$ b$\bar{\text{b}}$b$\bar{\text{b}}$, the signal yield is increased by 14% by the BTL alone, and 18% from the combined power of BTL and ETL, at constant background rate (Fig. 5.17 right). While the full optimization of the b-tagging algorithm and of the working points is left for future studies, further details on the impact of the current performance projections of the MTD on the di-Higgs boson searches are given in Section 5.4.

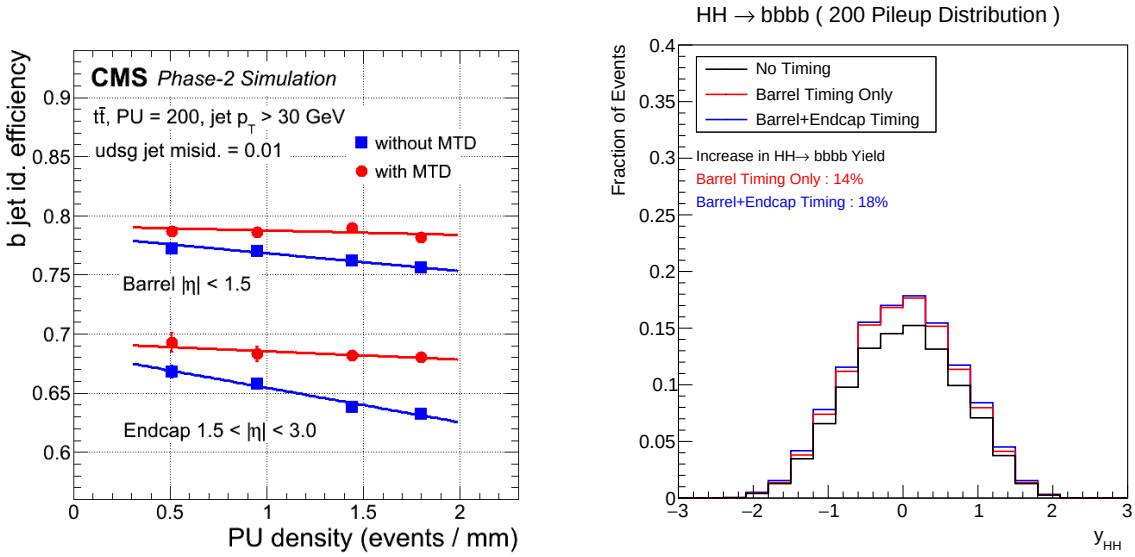


Figure 5.17: Left: Efficiency of the b-tagging vs. the average pileup density, with a constant light-jet efficiency of 0.01. Right: Projections for yield enhancement in HH $\rightarrow$ b$\bar{\text{b}}$b$\bar{\text{b}}$ as function of the Higgs boson rapidity. The distributions are normalized to the no-timing case.

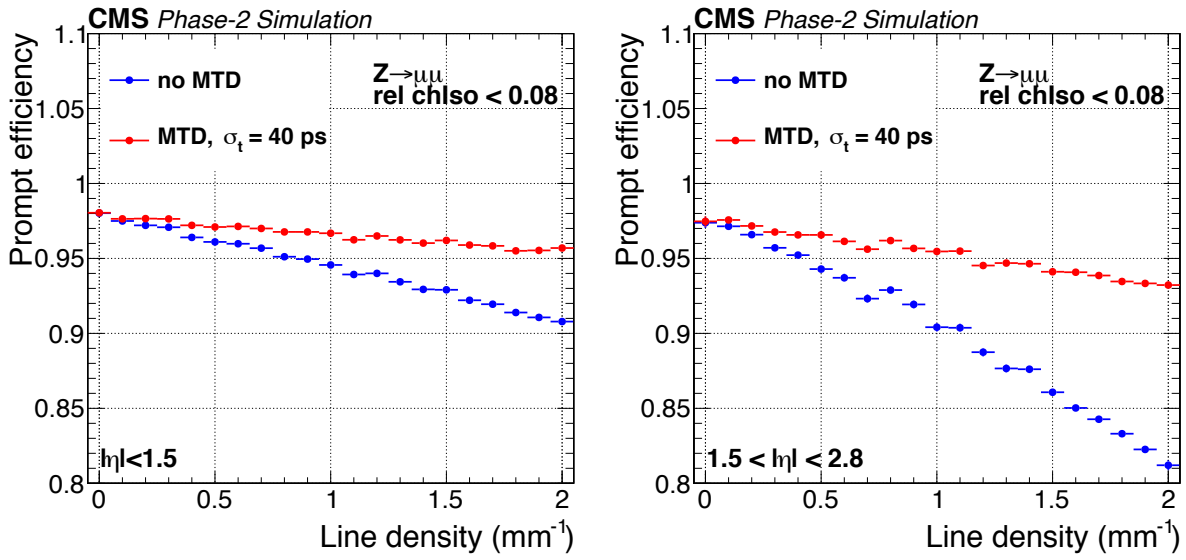


Figure 5.18: Prompt muons isolation efficiency as a function of the vertex density for a relative charged isolation cut-off of 0.08 for the barrel (left) and the endcaps (right). The efficiency for the with timing (red) and no timing (blue) cases are compared.

5.3.4 Lepton isolation from charged tracks

The impact of adding track-timing information in the computation of charged track isolation was tested on taus, muons and electrons, extending the studies presented in [8]. While taus are particularly relevant for some key process of the HL-LHC physics program, like di-Higgs boson searches and $H \to \tau\tau$ measurements, muons are of no lower importance and offer a clean signature that is an ideal test bench to understand the impact of the MTD on the charged isolation sum used to enhance the identification of isolated particles.

A thorough set of studies of the muon isolation, without and with the MTD and for different time resolutions, as a function of the pileup density, of the muon p_T, and of the pseudorapidity has been performed using $Z \to \mu\mu$ decays for prompt muons. Misidentified muons or non-prompt muon candidates, originating mostly from semileptonic decays of heavy-flavour hadrons, are selected from simulated $t\bar{t}$ events. The separation between prompt and non-prompt muons is strengthened by requiring or vetoing a match with a prompt muon at the generator level. Additional acceptance and quality selections include: $p_T > 10$ GeV, $|\eta| < 2.4$, a loose muon identification selection, and to have a point of closest approach to the primary vertex, assumed to be known, within 0.5 cm in the z direction and 0.2 cm in the transverse plane.

The charged isolation sum is computed from the sum of the transverse momenta of charged PF candidates in a cone $\Delta R < 0.3$ around the muon candidate — the muon track excluded — satisfying the following selections: track $p_T > 0.7$ (0.4) GeV and $|\Delta z(\text{track}, \text{PV}))| < 1$ (2) mm for tracks in the BTL (ETL) acceptance. As discussed in Section 5.3.1, these Δz windows maximize the performance for the no timing case. When using timing information, the additional requirement $|\Delta t(\text{track}, \text{PV})| < 3\sigma_t^{\text{track}}$ is applied, for the assumed time resolution σ_t^{track}. A selection on the charged isolation sum (i.e. tracks p_T sum divided by the muon p_T) is applied and the efficiency is calculated as the fraction of muon candidates passing that selection.

The efficiency gain provided by the MTD can be gauged from Fig. 5.18, which shows the efficiency for prompt muons with $p_T > 20$ GeV as a function of the line vertex density for a

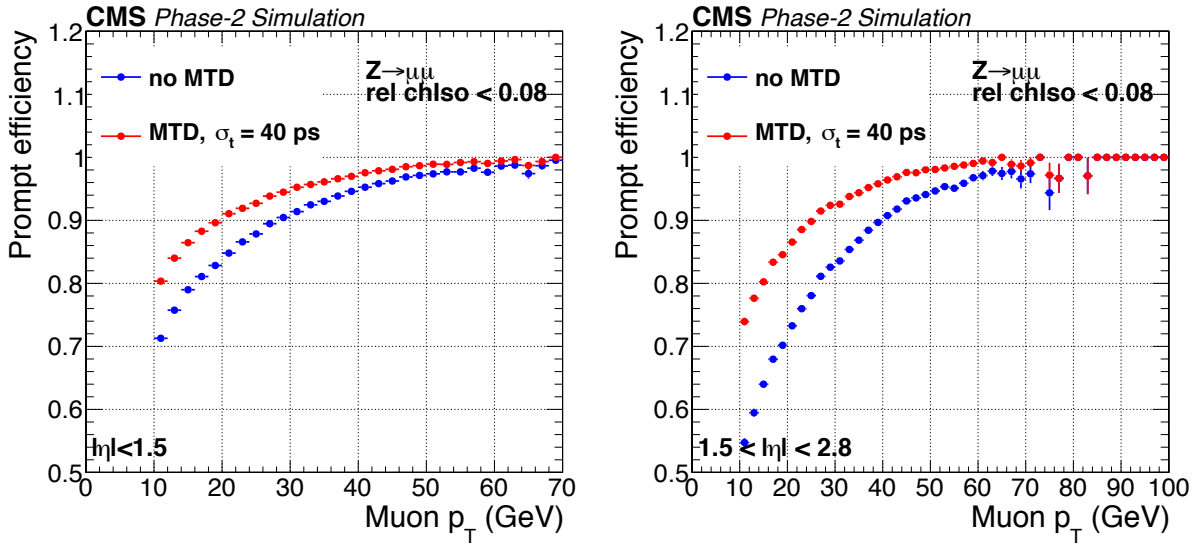


Figure 5.19: Prompt muons isolation efficiency as a function of the muon p_T for a relative charged isolation cut-off of 0.08 for the barrel (left) and the endcaps (right). The efficiency for the with timing (red) and no timing (blue) cases are compared.

representative fixed working point. A time resolution of $\sigma_t^{\text{track}} = 40$ ps is assumed, which is representative of the average time resolution of the MTD. The efficiency gain for the average line vertex density at 200 pileup is about 3–4% and 6–7% in the barrel and endcap, respectively. Moreover, the efficiency dependence on the line density is largely mitigated, showing that the addition of timing information also provides robustness against any possible future beam-crossing scenarios that may maximize or otherwise alter the luminosity production capabilities of the HL-LHC.

Figure 5.19 shows the isolation efficiency for prompt muons as a function of the muon p_{T} for 40 ps resolution. The same fixed working point as in Fig. 5.18 is used, which provides an approximately constant background efficiency of 3%. The efficiency gain from using the MTD is higher for lower p_{T}'s and reaches about 10 (18)% in the barrel (endcap), as the isolation sum of low-p_{T} muons is more severely affected by the pileup of soft particles.

At fixed working point, the background level is slightly varying. Additional information on the efficiency gain from timing is provided by the ROC curves presented in Fig. 5.20, where the efficiency for non-prompt muons of $p_{\text{T}} > 20\,\text{GeV}$ is plotted against the efficiency for prompt muons from Z bosons. The top panels shows results for a time resolution $\sigma_t^{\text{track}} = 40$ ps. The comparison to the no-timing case both for 200 and zero-pileup indicates that about 70% of the performance degradation at 200 pileup is recovered by using timing information.

The bottom panels show the results for several assumptions on the timing resolution spanning the range from the beginning (30–40 ps) to the end (50–60 ps) of HL-LHC operation for BTL, or covering the cases for a two-disk or a single-disk endcap. These studies do not include any performance degradation of the Tracker. In the worst case scenario of 70 ps, which is beyond the expected performance degradation of the MTD, the gain from timing is roughly halved with respect to the one obtained with 30 ps time resolution per track at the beginning of operation.

These results are confirmed by a similar study performed with electrons, albeit on a lower set of parameters, thereby indicating that the results for muons apply to single isolated tracks independently of their flavour, including leptons from secondary tau decays, and photons.

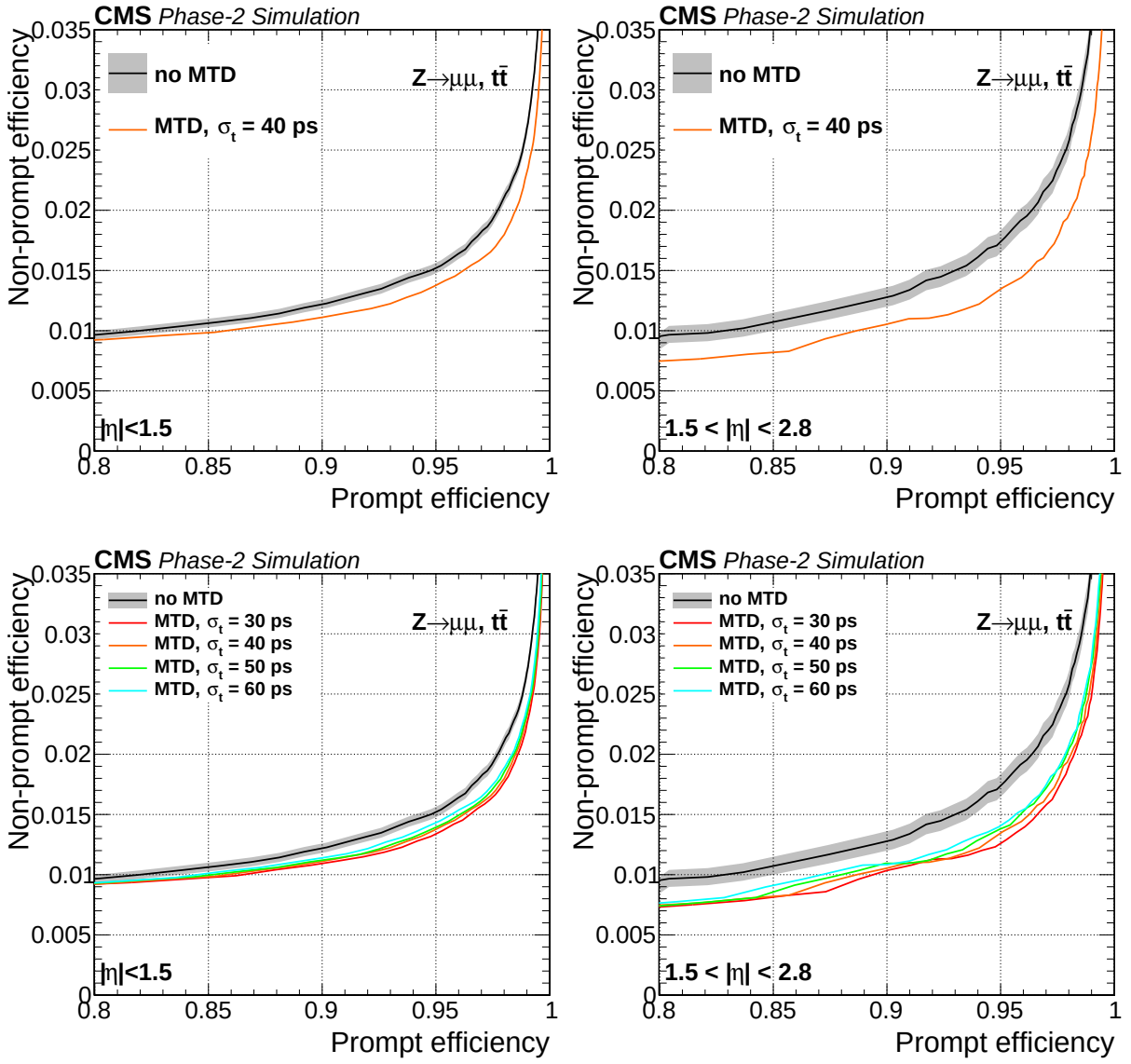


Figure 5.20: Top: ROC curves for the charged isolation selection for a track time resolution of 40 ps (orange line) for muons from $Z \to \mu\mu$ decays in the barrel (left) and in the endcaps (right). Results are compared to the no-timing case (solid black line). Bottom: ROC curves for different assumptions on the timing resolution in BTL (left) and ETL (right).

Hadronic tau decays are characterized by a complex final state topology, with the so-called 3-prong, 1-prong, and 1-prong+π^0 tau categories, related to the decay multiplicity of the tau candidate. The tau identification is more complex and the background more severe than for muon and elections, resulting in efficiencies that are typically in the range between 30 and 50%, much lower than for muons and electrons. Hence, the potential gain from timing is larger, as indicated by preliminary studies reported in Ref. [8]. Updated results for 40 ps and 50 ps resolution and an average efficiency for a track time measurement of 85% are shown in Fig. 5.21. An efficiency gain of 10–15% for prompt hadronic taus is observed for fixed jet misidentification probabilities of a few percent, corresponding to typical working points in Higgs boson searches.

As discussed for the tagging of b jets, the efficiency gains due to the charge isolation selection, albeit limited at the single-object level, combine in multi-object final states resulting in a signal

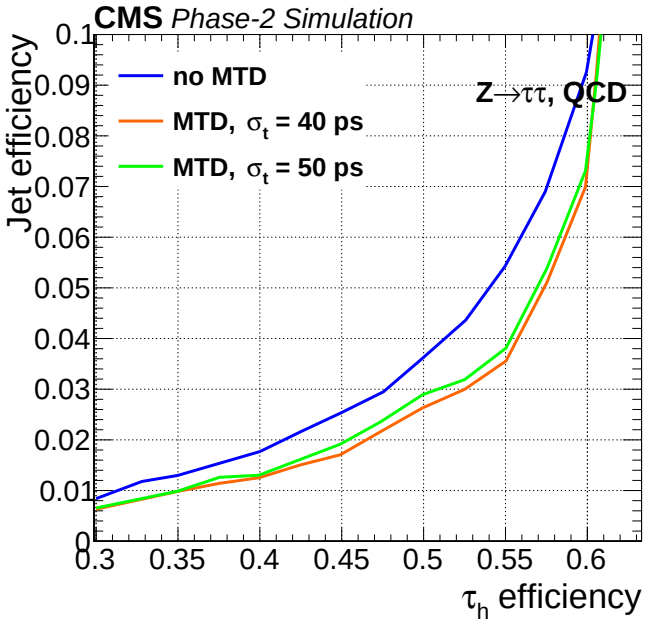


Figure 5.21: ROC curves for the isolation efficiency for hadronic taus determined from $Z \to \tau\tau$ and QCD events as sources of prompt and non-prompt taus, respectively. Results are shown for MTD time resolutions of 40 (orange) and 50 ps (green), as well as for the no timing case (blue).

yield increase of order 15-20% in the di-Higgs boson searches (Section 5.4) and in Higgs boson decays (Table 1.1 and Ref. [8]). Moreover, the p_T dependence of the efficiency gains suggests that the MTD can make an even more important impact in measurements with final states characterized by signatures with isolated low-p_T particles, as in SUSY searches with multi-lepton final states [116], for instance.

5.3.5 Electron identification from energy deposits in the MTD

The different behaviour of electromagnetic particles and their dominant fake objects in the MTD can be exploited to further improve the performance of electron and photon identification techniques. A study was made comparing the distribution of several MTD-related variables for single electrons and single charged pions reconstructed as electrons, the latter taken as a proxy for jets faking electrons. As shown in Fig. 5.22, electrons are more likely to deposit large amounts of energy in the MTD (chiefly BTL) than pions, which behave mostly as MIPs. The different pattern of the energy deposition can be exploited to improve the rejection power against pions. To test the potential gain of this approach, a BDT discriminant that combines the default CMS electron identification discriminant with a few MTD-related variables as input variables has been trained. The outcome for particles within the BTL acceptance is shown in the right panel of Fig. 5.22 and indicates that the background can be significantly reduced at constant identification efficiency. The study was repeated for ETL, but the gains are less important. Further improvements can be anticipated from the definition of a new MVA discriminant where all standard electron identification variables — such as shower shapes, electromagnetic shower-track compatibility variables, etc. — and MTD variables are used simultaneously in the discriminant training, so as to fully profit from correlations among them.

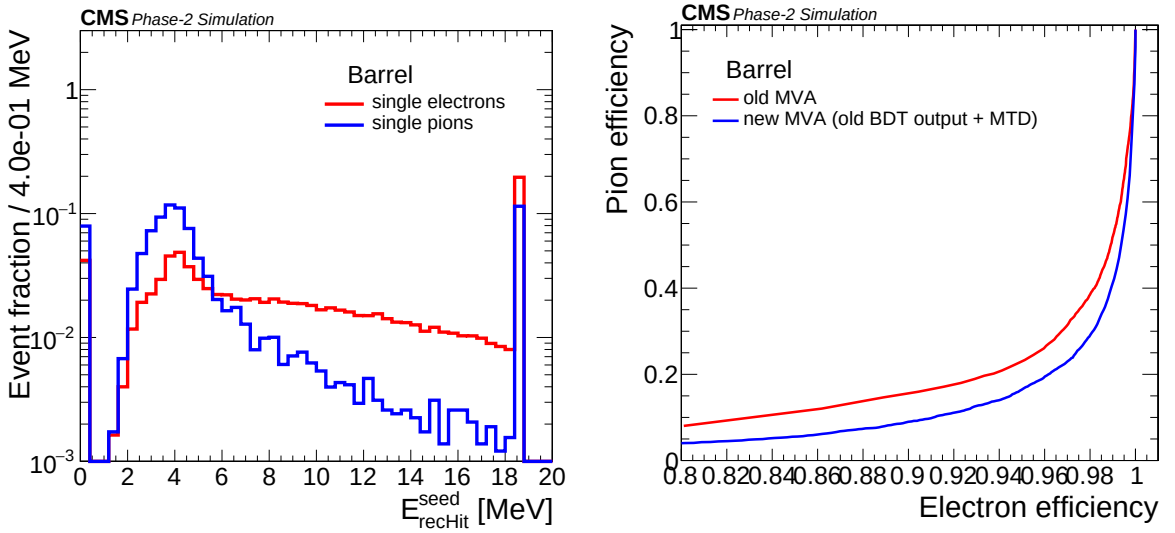


Figure 5.22: Left: Energy deposited in the BTL by an isolated electron (red histogram) and an isolated pion (blue histogram). The bin at 19 MeV of the histogram accumulates overflows. Right: ROC curves showing the identification efficiency for single electrons and single charged pions in the barrel. The red curve is obtained using the default Phase-2 electron identification BDT discri minant, while the blue curve shows the performance that can be achieved exploiting MTD-related observables.

5.3.6 Time-of-Flight Particle identification

Particle identification with the MTD is based on the time-of-flight difference of particles with different masses and thus velocity for a given particle momentum, p:

$$\Delta t = \frac{L}{c}\left(\frac{1}{\beta_1} - \frac{1}{\beta_2}\right), \tag{5.1}$$

where L is the particle flight distance, and β_1 (β_2) is the velocity of particle 1 (2).

The expected performance in separating charged pions, kaons, and protons, as a function of transverse momentum (p_T) and rapidity (y), in the barrel and endcap timing layers with a time resolution of 30 ps was presented in Chapter 1, Fig. 1.5. Realistic performance of particle identification is studied with the full CMS simulation and reconstruction framework. Event samples are generated using the HYDJET event generator for minimum bias PbPb collisions at 5.5 TeV.

Unlike high pileup pp events, there is on average only one PbPb collision present in each beam crossing and all particles are originated from a well-defined reconstructed vertex in (x, y, z) coordinates. To calculate the particle velocity, the common event start time, t_0^{evt}, is taken to be the time of the most populated 4D vertex; the particle arrival time is provided by the MTD hit, t_0^{MTD}. The reciprocal of the particle velocity can be calculated as

$$\frac{1}{\beta} = \frac{c(t_0^{MTD} - t_0^{evt})}{L}, \tag{5.2}$$

where L is the path length of a track from the beam line to the MTD.

Figure 5.23 shows the 2-D distributions of $\frac{1}{\beta}$ as a function of the particle momentum in minimum bias HYDJET PbPb events, for the BTL and ETL regions, respectively. The expected bands for pions, kaons and protons are clearly visible. The resolution is consistent with the expectation, with proton ID up to $p \sim 5$ GeV and kaon ID up to $p \sim 3$ GeV.

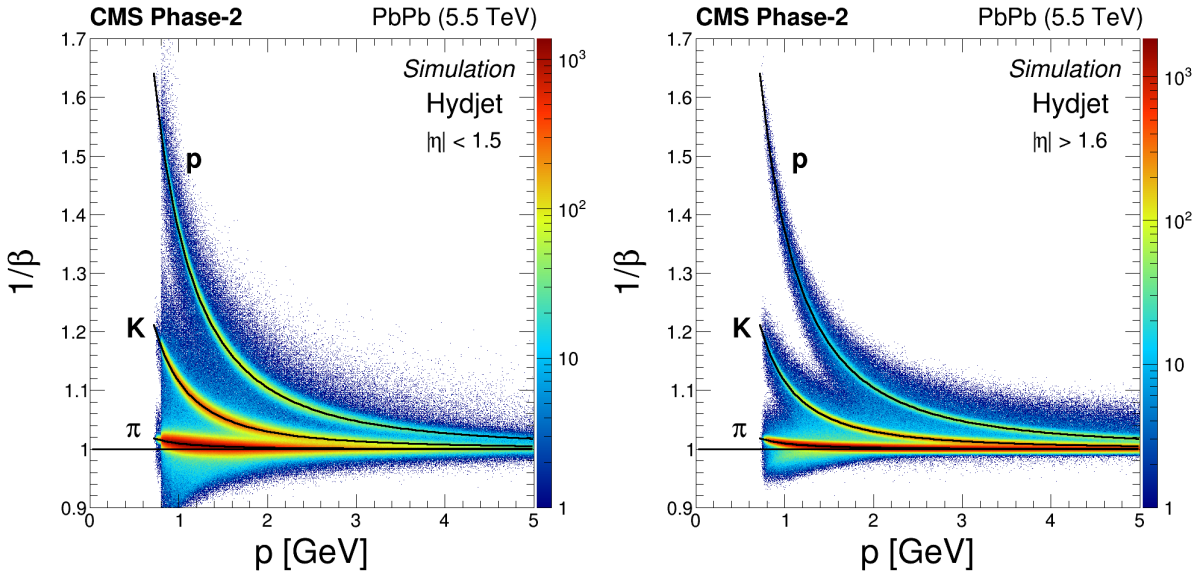


Figure 5.23: The inverse velocity ($1/\beta$) as a function of the particle momentum, p, for BTL ($|\eta| < 1.5$) and ETL ($|\eta| > 1.6$) in HYDJET PbPb simulation at 5.5 TeV.

5.4 Physics impact examples

Despite the integration of the time information in the event reconstruction being still preliminary and limited to the charged tracks, the improvements in physics-object reconstruction from timing (Section 5.3) consistently demonstrate that the MTD will allow CMS to operate at a leveled luminosity corresponding to 200 pileup interactions with a performance equivalent to Run-2 and Run-3 pileup conditions. The benefits in sensitivity for measurements and searches, across a wide range of objects and across the HL-LHC physics program leveraging gains across the full pseudorapidity coverage. For multi-objects final states, such as di-Higgs searches, this can be summarized as a 15–30% gain in effective integrated luminosity, which is equivalent to an additional three years of operation of the HL-LHC complex, as anticipated in Section 1.2 (Table 1.1).

This section assesses the impact of the MTD on a few benchmark cases, representative of three different ways of exploiting the MTD: by using physics objects with improved performance from the time information; by improving the discrimination power through the use of new, time-based variables; and by using the new particle identification capabilities provided by time-of-flight measurements exploiting the MTD. Three families of analyses are considered to cover these aspects: the search for Higgs boson pair (HH) production in several final states; the search for long-lived particles (LLP) in "beyond the standard model" (BSM) models; the measurement of heavy flavor hadron production in Heavy Ion collisions.

The precision characterization of the Higgs boson will be one of the highest priorities of the HL-LHC physics program. The cumulative impact of the MTD-improved object reconstruction was quantified in enhanced signal yields of about 15–25% for prominent Higgs boson processes for

the HL-LHC era, such as Higgs boson decays to four leptons, di-Higgs boson events or events where the Higgs boson is produced in association with other particles [8]. Here, we further extend the study of the di-Higgs boson production, using the most recent projections for this search [117]. The benefit from MTD to the projected sensitivity is evaluated for all the final states considered so far. While this analysis is a good example to illustrate the impact of pileup mitigation in a key signature at HL-LHC, larger benefits can be anticipated for signals with softer p_{T} spectra.

The second family of analyses which will directly benefit from the use of timing information are searches for LLPs. The use of the time information allows not only to improve the identification of the LLPs displaced decay vertices, but in some cases also to kinematically constrain the mass of the LLPs by measuring its velocity. New powerful handles can then be used in these searches to further reduce the SM background and to design novel search strategies, boosting the sensitivity in regions of BSM phase space which would could not be covered by only increasing the integrated luminosity.

The third family of analyses benefits from the time-of-flight particle ID (PID) to identify slow charged particles. As a show-case we discuss the advantages of PID to measure heavy flavor hadron production cross-sections, such as D^0 and Λ_{c}^+ in Heavy Ion PbPb collisions. A large reduction of the combinatorial background in the reconstruction of the hadron decay can be achieved, allowing to improve significantly the precision of the measured cross-sections and to extend the measurements in the low p_{T} region.

These examples are only illustrative of the potential added by precision timing. Signatures with low p_{T} objects and studies of exclusive decays can profit from the large gain in pileup rejection at low p_{T} and from the PID capabilities provided by timing. For example, it can be argued that the use of the PID capabilities can improve also the precision of statistically limited measurements of b-quark mesons decays, such as the studies of the $\mathrm{b} \to \mathrm{s}\ell\ell$ transitions (e.g. $\mathrm{B}_{\mathrm{d}}^0 \to \mathrm{K}^*(\to \mathrm{K}^+\pi^-)\mu^+\mu^-$). The PID capabilities can also be exploited in BSM searches. The MTD uniquely combines the possibility of performing $\mathrm{d}E/\mathrm{d}X$ and time-of-flight measurements on charged particles, which can extend the reach of searches for heavy stable charged particles (HSCP) into regions of phase space which otherwise will remain uncovered.

5.4.1 Higgs boson pair production

The study of the HH production is one of the main goals of the LHC physics programme. It constitutes the only way to directly determine the value of the Higgs boson trilinear self-coupling (λ_{HHH}) and thus determine the shape of the Brout–Englert–Higgs potential and explore the nature of the scalar sector of the SM. At the LHC, the HH production mainly proceeds through the gluon fusion production mechanism, with a cross section of $36.69^{+2.1\%}_{-4.9\%}$ fb at $\sqrt{\mathrm{s}} = 14\,\mathrm{TeV}$. Because of the small cross section, a high integrated luminosity is needed to achieve sensitivity to the signal predicted in the SM and to measure the λ_{HHH} coupling. In order to get the best experimental sensitivity five decay channels are explored and combined: bbbb, bb$\tau\tau$, bbWW ($\mathrm{WW} \to \ell\nu\ell'\nu'$ with $\ell, \ell' = \mathrm{e}, \mu$), bb$\gamma\gamma$, and bbZZ ($\mathrm{ZZ} \to \ell\ell\ell'\ell'$ with $\ell, \ell' = \mathrm{e}, \mu$). The richness and variety of physics objects spanned by this analysis makes it sensitive to the individual improvements in each of them. The inclusion of MTD information in the physics objects has a large, integral effect in its final experimental reach.

5.4.1.1 Analysis strategy

This analysis exploits all the five principal HH decay channels and closely follows the strategy of Ref. [117], optimized for an integrated luminosity of 3000 fb^{-1}. Background and signal

samples are produced with the DELPHES parametric simulation [114] using efficiency and fake-rate ROC curves matching the performance of the CMS Phase-2 detector from full simulation. Samples with the MTD detector are generated using the ROC curves from the studies presented in Section 5.3 for an MTD resolution of 30–40 ps. The selection of the events and the discrimination of HH signal candidates from backgrounds are common to the analyses with and without MTD.

Events in the bbbb decay channel are selected requiring the presence of four b-tagged jets with $p_T > 45\,\text{GeV}$ and within the fiducial tracker acceptance $|\eta| < 3.5$. The selected jets are paired into the two Higgs boson candidates. Because of the large QCD multijet and $t\bar{t}$ backgrounds, a multivariate discriminant in the form of a Boosted Decision Tree (BDT) is used to discriminate the signal from the background. The output distribution of the BDT is used to perform the signal extraction.

The bb$\tau\tau$ analysis explores three final state of the $\tau\tau$ system defined by the presence of a semileptonic τ decay (τ_h) in association to an electron, a muon, or another τ_h (eτ_h, $\mu\tau_h$, and $\tau_h\tau_h$ final states, respectively). The presence of a pair of b-tagged jets is required to reconstruct the $H \to bb$ decay. The dominant backgrounds are $t\bar{t}$ and Drell–Yan τ pair production, that are difficult to suppress because of the incomplete final state reconstruction due to the presence of neutrinos escaping detection. Background processes are thus suppressed with advanced machine learning methods using a Deep Neural Network (DNN) as the discriminant variable.

The search in the bbWW decay channel focuses on the fully leptonic final state of the WW system. Events are divided in three categories depending on the lepton flavor (ee, $\mu\mu$, and eμ final states), where the presence of a pair of b-tagged jets is required. The dominant, irreducible background for this channel is $t\bar{t}$ production, with a smaller contribution from Drell–Yan lepton pair production. A DNN is developed using the kinematic properties of the selected objects to look for the presence of a signal.

The bb$\gamma\gamma$ analysis benefits from the clean signature resulting from the photon pair. Events are selected requiring the presence of two well identified and isolated photons and of two b-tagged jets. The main backgrounds are the continuum production of diphoton events in association with jets, and single Higgs boson production, dominated by the ttH production mode. The latter is suppressed with a selection on a dedicated BDT that combines information on the presence of extra leptons and jets, and the kinematics of the selected objects. A second BDT is trained to separate the signal from the continuum diphoton background. While events at low BDT score are rejected, the others are split into a medium purity and a high purity category to maximize the sensitivity while keeping a high selection efficiency. Events in each category are further subdivided in three categories based on the invariant mass of the $\gamma\gamma$jj system. While the category with $m_{\gamma\gamma jj} > 480\,\text{GeV}$ gives the best sensitivity to SM HH production, the low mass categories help to constrain anomalous λ_{HHH} values that mostly enhance the production cross section at the HH invariant mass threshold.

The bbZZ final state is explored in the fully leptonic ZZ decay channel. While characterized by a very small branching fraction, this final state provides a large acceptance to HH events, full event reconstruction, and a clean signature, with the dominant background processes being single Higgs boson production. Events are selected by requiring the presence of two pairs of well reconstructed and isolated electrons or muons, and of two b-tagged jets. The lepton pairs are used to reconstruct the on- and off-shell Z candidates, and a selection on their invariant mass is applied as well as a requirement that the four lepton invariant mass is compatible with m_H.

5.4.1.2 Summary of physics object improvements, results and prospects

An MTD resolution of 35 ps is assumed in this study. This figure matches the MTD resolution at 1000 fb^{-1} and is representative of the average resolution of BTL and ETL for an accumulated luminosity of 3000 fb^{-1}, with ETL remaining stable and BTL slowly drifting from 30 to 50 ps (Section 5.2.4).

The p_{T} and η-dependent efficiency and fake-rate curves obtained for the lepton isolation, photon identification and b-jet identification are applied to each channel to compare results with and without using MTD. The individual results are statistically combined. At the working point selected for the analyses, the efficiency gains for each final state object amount on average to about 3–4% for electrons, photons and muons, 4–6% for b-jets and about 10–15% for taus, where the range includes the dependence on η, at constant rejection probability for the reducible backgrounds. The cumulative effect of the gains at the single object level translates into an increase in the signal yield at constant reducible background ranging between 15 and 35% depending on the final state, as illustrated for $\text{HH} \to \text{bbbb}$ events in Fig. 5.17 and summarized in the next section.

5.4.1.3 Results and prospects

Results are reported for 3000 fb^{-1} and derived under the assumption that a HH signal exists with the strength and properties predicted by the SM. Table 5.2 lists the projected signal yield increases and the expected expected significance for the SM HH signal when exploiting the timing information provided by the MTD in b-tagging and particle isolation. For the $\text{bb}\gamma\gamma$ analysis, which exploits event categorization, the signal increase is listed for the most sensitive category. The expected significance for the SM HH signal includes the systematic uncertainties, treated as detailed in Ref. [117]. In summary, the MTD brings an improvement to the significance of about 13%, which would require an additional 26% luminosity without MTD. An extrapolation of the analysis at 4000 fb^{-1}, corresponding to operations at 200 pileup, yields an additional gain in sensitivity of about 10%.

Table 5.2: Projections for the HH signal yield increases when exploiting the timing information provided by the MTD in b-tagging and particle isolation and for the significance in units of σ for the SM HH signal. Projections for the five decay channels and their combination are based on existing Phase-2 studies and scaled according to the ROC curves of the individual objects. The results are reported for an integrated luminosity of 3000 fb^{-1}, assuming an MTD average resolution of 35 ps.

	Signal increase (%)		Expected significance	
Di-Higgs decay	BTL	BTL+ETL	No MTD	MTD
bbbb	13	17	0.88	0.95
$\text{bb}\tau\tau$	21	29	1.3	1.6
$\text{bb}\gamma\gamma$	13	17	1.7	1.9
bbWW			0.53	0.58
bbZZ			0.38	0.42
Combined			2.4	2.7

This study only includes the cumulative benefits from improved charged isolation sums and b-tagging. Additional gains in sensitivity will also come from the improved pileup rejection in final states with jets or from the improved $p_{\text{T}}^{\text{miss}}$ resolution in the bbWW and $\text{bb}\tau\tau$ channels. As an illustration, Fig. 5.24 shows the impact of a 10% improvement in the $p_{\text{T}}^{\text{miss}}$ resolution on the reconstruction of the invariant mass of the tau pairs, which is reconstructed by sharing

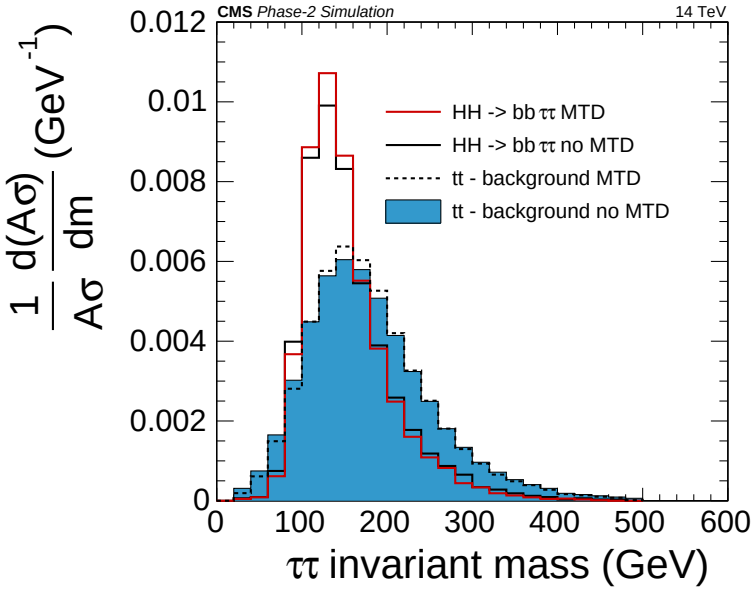


Figure 5.24: Invariant mass of the tau pairs in the reconstruction of HH $\rightarrow$ bb$\tau\tau$ events with the CMS Phase-2 detector with and without the MTD.

the $p_{\mathrm{T}}^{\mathrm{miss}}$ of the event between the two taus. The missing transverse momentum is the largest source of uncertainty in the estimate of the tau momentum and is correlated between the two tau candidates. Hence, a 10% variation in $p_{\mathrm{T}}^{\mathrm{miss}}$ resolution translates into a 10% variation in the invariant mass resolution. The improved resolution enhances the separation between the signal and background in this observable, which is among the most discriminant in the deep neural network used to separate HH $\rightarrow$ bb$\tau\tau$ events from the $\mathrm{t\bar{t}}$ background.

5.4.1.4 Projections in a conservative time resolution scenario

The di-Higgs results discussed in the previous section are based on the MTD performance with nominal radiation model (Table 1.3). A more conservative study was also performed using a time resolution of 50 ps, which corresponds to the condition achievable assuming radiation levels a factor 1.5 larger than the nominal prediction.

Efficiency and fake-rate improvements obtaned for lepton and photon isolation and b-jet identification with a time resolution of 50 ps are propagated to each channel using the same approach described in the previous section. Table 5.3 summarizes the per channel and combined expected significance for an integrated luminosity of 3000 fb^{-1}.

Even in this conservative scenario, the MTD would provide a performance gain that would require an equivalent additional luminosity of about 20% for the CMS detector without MTD to be achieved.

5.4.2 Long-lived particles

Searches for LLPs are theoretically well motivated. Many BSM extensions allow the possibility of or require long lifetimes for particles due to high-dimension operators, very small couplings, heavy mass scales, or suppressed phase space. A more detailed discussion on LLPs can be found in Ref. [16] and references therein.

Table 5.3: Projections for the HH expected significance when exploiting the timing information provided by the MTD in b-tagging and particle isolation. Projections for the five decay channels and their combination are based on existing Phase-2 studies and scaled according to the ROC curves of the individual objects. The results are reported for an integrated luminosity of 3000 fb^{-1}, assuming an MTD average resolution of 50 ps.

	Expected significance	
Di-Higgs decay	No MTD	MTD
bbbb	0.88	0.94
bb$\tau\tau$	1.3	1.48
bb$\gamma\gamma$	1.7	1.83
bbWW	0.53	0.58
bbZZ	0.38	0.42
Combined	2.4	2.63

The MTD provides new, powerful information in searches for long-lived particles. For LLP decays with charged particles in the final state, the MTD will measure the time of flight of the LLP from the time difference between its decay vertex and the primary vertex. Using the measured displacement between primary and secondary vertices in space and time, the velocity of an LLP in the laboratory frame, $\vec{\beta}_P^{\text{LAB}}$ (and γ_P), can be measured. In such scenarios, the LLP can decay to fully-visible or partially-invisible systems. Using the measured energy and momentum of the visible portion of the decay, one can calculate its energy in the LLP rest frame and reconstruct the mass of the LLP, assuming that the mass of the invisible system is known. For LLP decays into neutral particles including photons, the MTD will enable the measurement of the time difference between the primary vertex and and the photon arrival time in a calorimeter or in the MTD, for converted photons, and to infer from the time difference the LLP lifetime.

For the first range of topologies (Section 5.4.2.1), the MTD allows the reconstruction of a peaking mass variable, which introduces a qualitatively new capability for LLP searches. For the second range of topologies (Sections 5.4.2.2 and 5.4.2.3), the MTD allows the indirect measurement of the LLP lifetime with significantly upgraded precision relative to the current detector.

5.4.2.1 Vertex time discrimination and mass reconstruction of SUSY particles

A gauge-mediated SUSY breaking (GMSB) scenario where the $\widetilde{\chi}_1^0$ couples to the gravitino $\widetilde{\text{G}}$ via higher-dimension operators sensitive to the SUSY breaking scale provides a benchmark scenario for this range of topologies. In such scenarios, the $\widetilde{\chi}_1^0$, produced in top-squark pair production with $\widetilde{\text{t}} \to \text{t} + \widetilde{\chi}_1^0$, $\widetilde{\chi}_1^0 \to Z + \widetilde{\text{G}}$, and $Z \to e^+e^-$, may have a long lifetime [118]. The decay diagram is shown in Fig. 5.25 (left).

The events were generated with PYTHIA 8.2 [119]. The masses of the top-squark and neutralino were set to 1000 GeV and 700 GeV, respectively. The simulation of the detector response was performed using the DELPHES [114] software with a description of the CMS upgraded detector. A position resolution of 12 μm in each of the three spatial directions was assumed for the primary vertex [13]. The secondary vertex position for the e^+e^- pair was reconstructed assuming 30 μm track resolution in the transverse direction. And finally, the time resolution of charged tracks at the displaced vertex was assumed to be 30 ps. Opposite sign leptons were selected with $p_{\text{T}} > 20$ GeV and invariant mass $|91 - m_{\ell\ell}| < 10$ GeV.

From the time difference between the production and the decay vertex, it is possible to measure the velocity of the neutralino. Combining this information with the kinematic properties of the visible decay products, the neutralino mass can be inferred, under the assumption of a massless

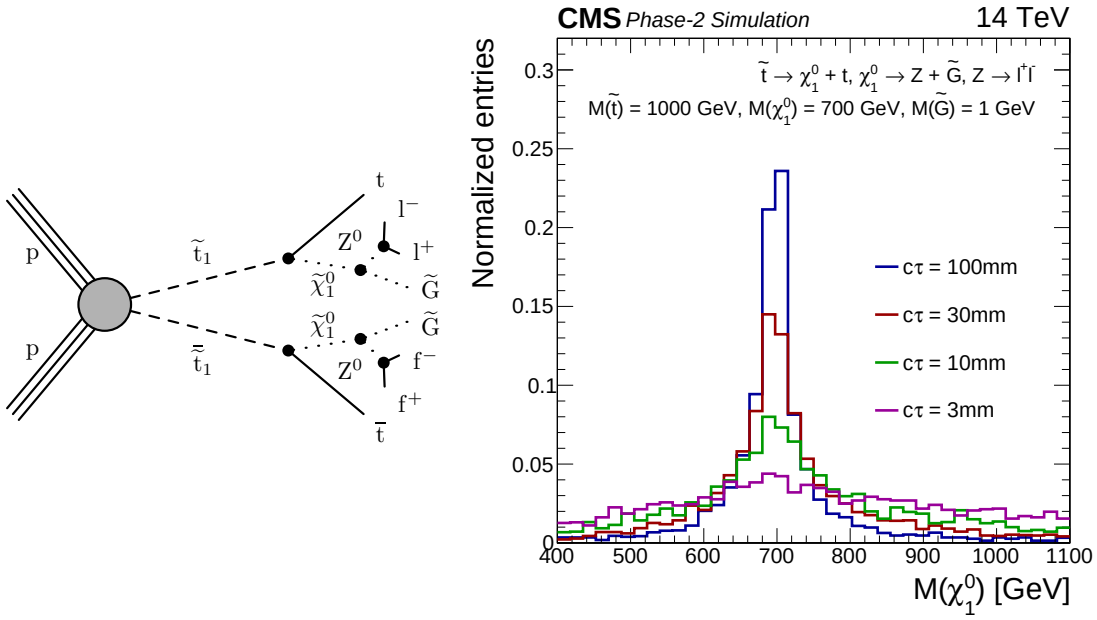


Figure 5.25: Diagram for top-squark pair production and decay (left), η and distribution of the mass of $\tilde{\chi}_1^0$ (right) reconstructed from the final state kinematics for decays with $M(\tilde{t}) = 1000$ GeV and $M(\tilde{\chi}_1^0) = 700$ GeV. The mass distributions are shown for various values of the $c\tau$ of the $\tilde{\chi}_1^0$.

gravitino. The right panel of Fig. 5.25 shows the distribution of the reconstructed mass of the neutralino for various $c\tau$ values of the LLP. The fraction of events with separation between primary and secondary vertices exceeding 3σ in both space and time as a function of the MTD resolution is shown in Fig. 5.26 (left). The mass resolution, defined as half of the shortest mass interval that contains 68% of events with 3σ displacement is shown in Fig. 5.26 (right), as a function of the MTD resolution.

A conceptually similar example is a SUSY scenario where the two lightest neutralinos and light chargino are higgsino–like. The light charginos and neutralinos are nearly mass degenerate [120] and may become long-lived as a consequence of the heavy higgsinos [121]. The neutralino-chargino $\tilde{\chi}_2^0\tilde{\chi}^{\pm}{}_1$ pairs can decay into the lightest supersymmetric particle (LSP) $\tilde{\chi}_1^0$ and a virtual $Z^\star$ boson or a $W^\star$, respectively. In the limit where the light charginos and neutralinos are degenerate in mass ($\Delta M = M(\tilde{\chi}_2^0) - M(\tilde{\chi}_1^0) \simeq 0$), the energy of the e^+e^- (visible) system in the LLP rest frame provides a direct measurement of the mass splitting. The full event reconstruction gives a distribution similar to Fig. 5.26 but peaked at ΔM [8].

In both these examples, the lepton–antilepton pair is the visible part of the LLP decay. The pseudorapidity distribution of the $\ell^+\ell^-$ pair is peaked in the central region with about 90% of the leptons within the barrel acceptance ($|\eta| < 1.5$). This emphasizes the need for the barrel portion of the MTD in the reconstruction of these signatures.

5.4.2.2 Measurement of late photon with respect to the primary vertex in SUSY models

In the GMSB benchmark scenario [122] used as the reference for this topology, the lightest neutralino $\tilde{\chi}_1^0$ is the next-to-lightest supersymmetric particle. It can be long-lived and decay to a photon and a gravitino ($\tilde{G}$), which is the LSP. Figure 5.27 (left) shows a diagram of this process, which starts from the production of two gluinos that further decay into a quark-squark pair, with the squark decaying into another quark and the lightest neutralino $\tilde{\chi}_1^0$. In constrast to the

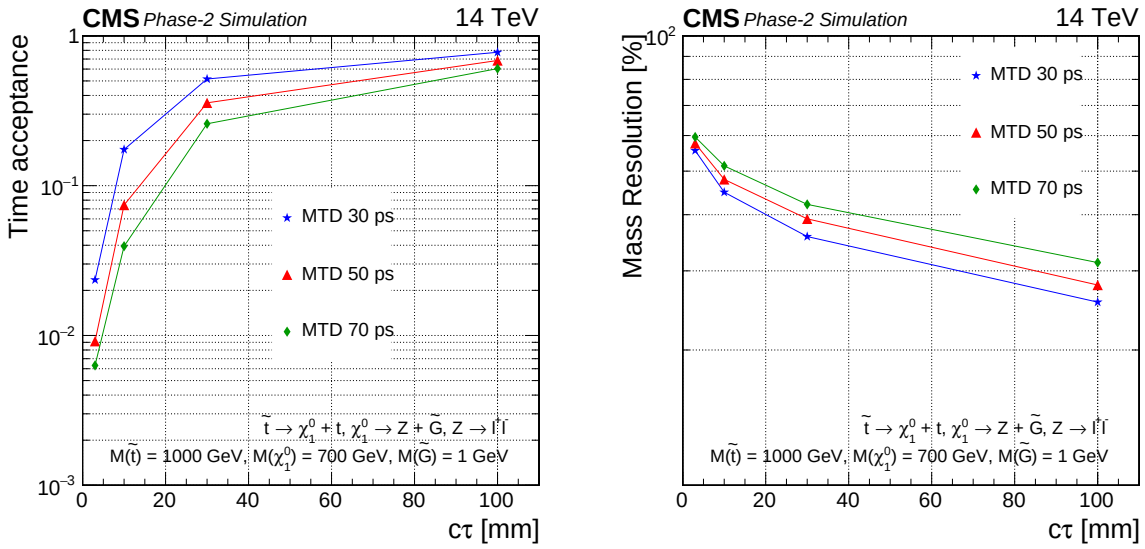


Figure 5.26: Efficiency (left) and mass resolution (right) as a function of the timing resolution of the MTD for reconstruction of the $\tilde{\chi}_1^0$ mass in the SUSY GMSB example of $\tilde{\chi}_1^0 \to \tilde{G} + e^+e^-$, with $M(\tilde{\chi}_1^0) = 700$ GeV, considering events with a separation of primary and secondary vertices by more than 3σ in both space and time.

previous case, the final state with two gravitinos that escape detection and two photons does not allow the position of the secondary vertex to be precisely determined.

For a long-lived neutralino, the photon from the $\tilde{\chi}_1^0 \to \tilde{G} + \gamma$ decay is produced at the $\tilde{\chi}_1^0$ decay vertex, at some distance from the beam line, and reaches the MTD detector, for converted photons, or the calorimeters at a later time than the prompt, relativistic particles produced at the interaction point. The time of arrival of the photon at the detector can be used to discriminate signal from background.

The time of flight of the photon inside the detector is the sum of the time of flight of the neutralino before its decay and the time of flight of the photon itself, until it reaches the detector. The total time of flight will be largely dominated by the time of flight of the massive neutralino. In order to be sensitive to short neutralino lifetimes of order 1 cm, the performance of the measurement of the photon time of flight is a crucial ingredient of the analysis. Therefore, the excellent resolution of the MTD apparatus can be exploited to determine with high accuracy the time of flight of the neutralino, and similarly the photon, also in case of a short lifetime. The MTD will serve to measure the time of the primary vertex and, for converted photons, the time of arrival of the photon.

An analysis has been performed in order to evaluate the sensitivity of a search for displaced photons at CMS in the scenario where a 30 ps timing resolution is available from the MTD. The events were generated with PYTHIA 8.2, exploring neutralino lifetimes ($c\tau$) in the range 0.1–300 cm. The values of the Λ scale parameter were considered in the range 100–500 TeV, which is relevant for this model to be consistent with the observation of a 125 GeV Higgs boson [123].

A generator level study, with the time of flight smeared according to the expected resolution was performed as well, since it allowed us to accumulate more events. In this case, all the photons within the CMS ECAL barrel acceptance and with a transverse energy greater than 70 GeV (representative of a possible trigger threshold) were retained in the analysis. In either case, a cut-off selection at a time-of-flight greater than 3σ of the time resolution was applied and

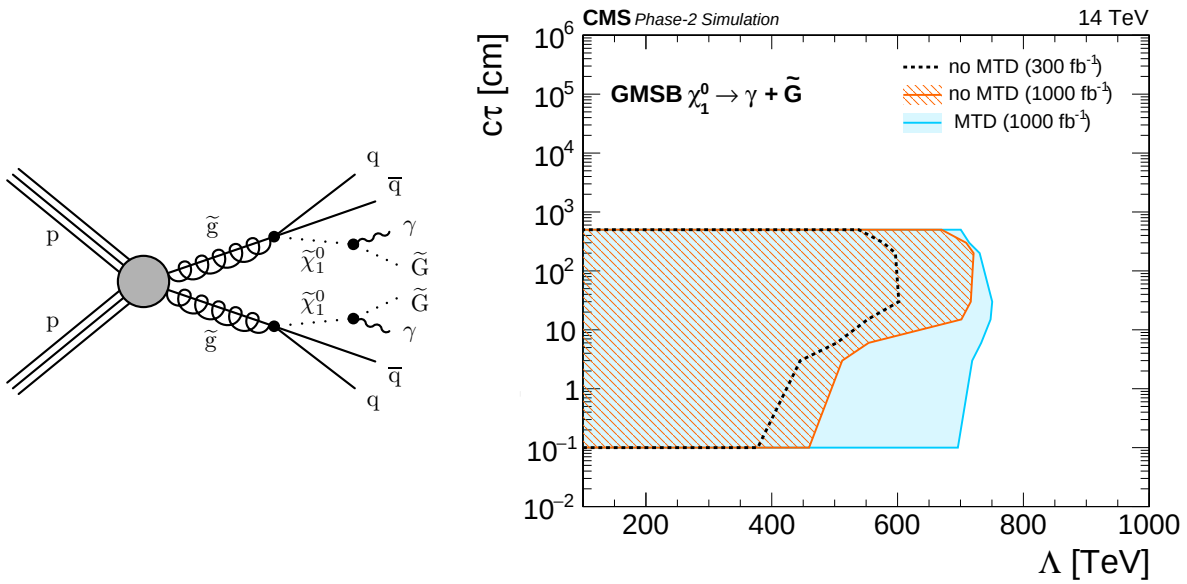


Figure 5.27: Left: Diagrams for a SUSY process that results in a diphoton final state through gluino production at the LHC. Right: Sensitivity to GMSB $\tilde{\chi}_1^0 \rightarrow \tilde{G} + \gamma$ signals expressed in terms of neutralino lifetimes for 180 and 30 ps resolution, corresponding to the Phase-2 detector with photon timing without MTD and with MTD, respectively.

a zero background assumption is made in this "signal region". The signal efficiency of such a requirement is computed and translated, assuming the theoretical cross-sections provided in Ref. [122], in an upper limit at 95% CL on the production cross-section of the $\tilde{\chi}_1^0 \rightarrow \tilde{G} + \gamma$ process. The two simulations provide consistent results, for the same amount of events.

Figure 5.27 (right) shows the analysis sensitivity in terms of the Λ scale (and therefore of the neutralino mass) and lifetime for three different assumptions on the timing resolution. The curves with 180 ps resolution, shown for both 300 and 1000 fb^{-1} are representative of the TOF resolution of the upgraded CMS detector without the MTD, in which the TOF measurement will be dominated by the time spread of the luminous region and the photon time will be measured by the upgraded ECAL calorimeter with 30 ps precision. The vertex timing provided by the MTD detector will bring the TOF resolution to about 30 ps. As visible in the figure, a full scope upgrade of the CMS detector with photon and track timing will provide a dramatic increase in sensitivity at short lifetimes and high masses, which could not be otherwise covered by a simple increase of integrated luminosity.

5.4.2.3 Displaced jets in exotic models involving the Higgs boson

An exotic model in which a Higgs boson mediates the production of two long-lived, scalar bosons (X) decaying into quarks was considered for this topology. Figure 5.28 (left) shows the diagram corresponding to this model where the masses of the Higgs boson and X particles have been fixed to 125 GeV and 50 GeV. The X particles generated at the PV travel some distance before decaying into pairs of jets. While the X are assumed to be neutral, the jet constituents can be charged, leaving a signal in the MTD at a time given by the production time at the PV plus the sum of the time of flight of the X particle and the time needed by the constituents to reach the MTD from the decay vertex. In cases where the X particles are very displaced, this time of arrival would be significantly higher than the time taken by a SM particle traveling from the PV to the MTD at the speed at light. This feature can be exploited in order to achieve discrimination of signal events with respect to background events.

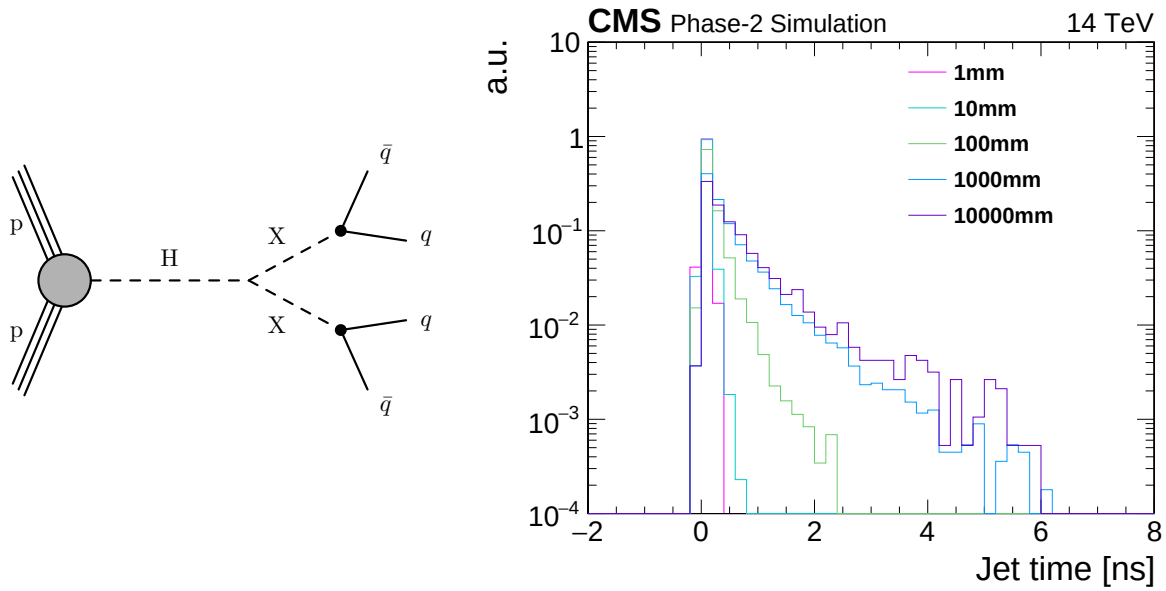


Figure 5.28: Left: Diagram for the production of displaced jets in a model mediated by a Higgs boson decaying into two long-lived scalar bosons. Right: Time difference between the measured jet time at the MTD and the expectation from the propagation of a particle from the PV at the speed of light.

A simplified simulation at generator level was used for this analysis: trajectories of charged particles were propagated to the MTD surface, and a smearing to their time of arrival using a resolution of 30 ps is applied. The time of the jets at the MTD was estimated as the mean of the time of its charged constituents with $p_{\mathrm{T}} > 1$ GeV, in a cone of $\Delta R = 0.3$ along the jet direction. The difference between this time and the time of the PV was compared to the time needed by a particle to travel from the PV to the MTD at the speed of light. Figure 5.28 (right) shows this quantity for models with different X lifetimes. Events with at least one jet with p_{T} >20 GeV and jet time greater than 0.1 ns were selected to reduce the level of background. Figure 5.29 shows the upper limits on the branching ratio $BR(H \to XX)$, assuming $BR(X \to jj) = 100\%$ and an integrated luminosity of 3000 fb^{-1}.

The time discrimination provided by the MTD will bring sensitivity over a large range of the lifetime of the X scalar boson (LL particle).

5.4.3 Particle velocity reconstruction in the context of HSCP searches

The GMSB benchmark model used for this topology implies the production of a Heavy Stable Charged Particle (HSCP) in the form of a stau $\tilde{\tau}$ with a very large lifetime crossing the full detector. A more detailed description of the model can be found elsewhere [124]. The MTD has a direct impact on this topology through the measurement the particle velocity, β, using the path length and the time difference between the primary vertex and the particle hits in the MTD. This quantity can be used to discriminate between signal and background SM processes, where particles are produced with a velocity close to the speed of light, and the resolution in the time measurement is the main factor distorting the measurement.

Figure 5.30 (left) shows the $1/\beta$ distribution for DY+Jets events with and without the MTD, and for signal events. Background events corresponding to the DY+Jets process have been generated using as reference the distributions in Ref. [125]. Signal events have been generated using PYTHIA 8.2 and have been used only at generator level. A selection $1/\beta > 1.25$ corresponding

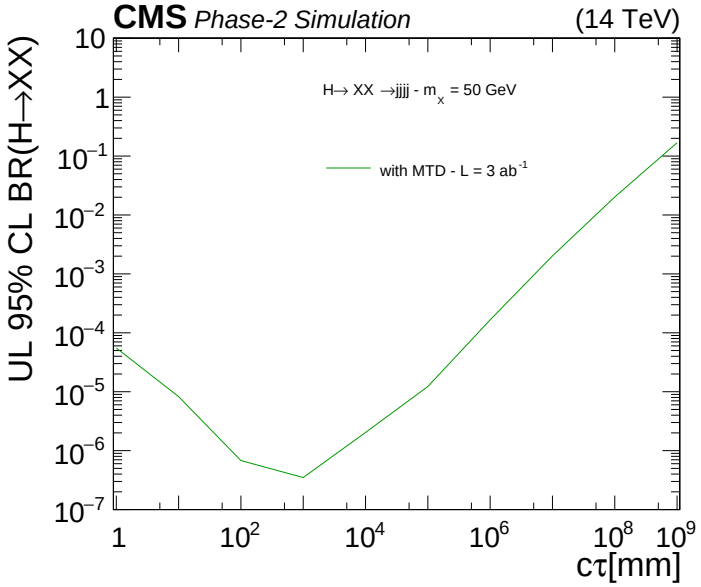


Figure 5.29: Sensitivity to $H \rightarrow XX \rightarrow jjjj$ models, as a function of the lifetime of the X scalar boson (LL particle). Only the timing capabilities from the MTD are used in the selection, more details in the text.

to five times the $1/\beta$ resolution observed for background events is applied on the reference CMS analysis. Figure 5.30 (right) shows the ROC curve associated to the $1/\beta$ distribution for the scenarios with and without MTD. Imposing the same five-sigma criteria with the new resolution provided by the MTD brings a new selection threshold of $1/\beta > 1.05$ which increases the signal acceptance by a factor 4.

The precise measurement of the velocity of the HSCP also allows to estimate its mass based on the particle momentum and β. Figure 5.31 (right) shows the HSCP mass for different time resolutions assumed in the MTD. An HSCP with a mass of 432 GeV has been used for this study. The pink curve corresponds to the current $1/\beta$ resolution obtained from the CMS analysis using the muon system as a TOF detector. It should be noted that for this study the particle energy loss information from the tracker has not been used and only improvements related to the TOF have been considered.

5.4.4 Heavy Ion Analysis with TOFPID

The addition of particle ID capability enabled by the MTD will open up many exciting physics opportunities at CMS. To demonstrate the unique strength of the MTD, we evaluate the performance of heavy flavor hadron (e.g., D^0 and Λ_c^+) reconstruction over long ranges in rapidity with the full CMS simulation and reconstruction framework.

Heavy-flavor quarks (charm and bottom) are primarily produced via initial hard scattering. As such, they are largely decoupled from the bulk production of soft gluons and light-flavor quarks in heavy-ion collisions, and thereby probe the properties and dynamics of the QGP through its entire evolution. Most measurements of heavy flavor particles have so far focused on the midrapidity region. In CMS, without particle identification, heavy flavor studies are currently limited to not very low-p_T regions ($p_T > 2$ GeV for D^0 mesons and $p_T > 7$ GeV for B mesons), while the QGP effect is expected to be strongest at lowest p_T's. No Λ_c^+ measurement in CMS is available to date.

With the proposed MTD, CMS will be able to carry out a unique heavy flavor program for com-

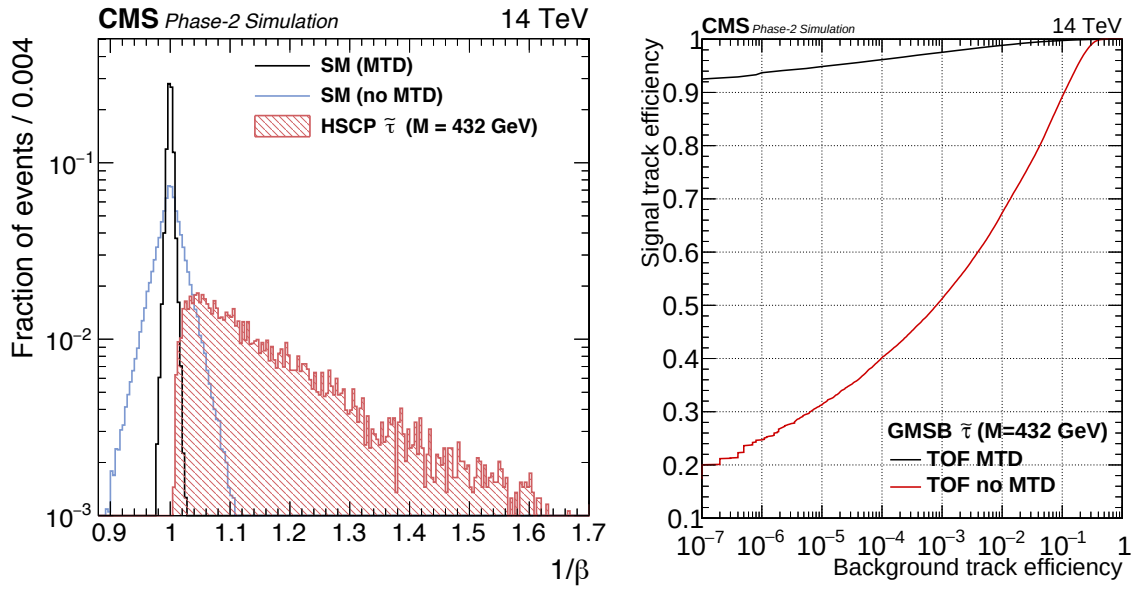


Figure 5.30: Distribution of $1/\beta$ for DY+Jets events with and without the MTD and signal events (left). ROC curve associated to the $1/\beta$ selection for the cases with and without MTD (right).

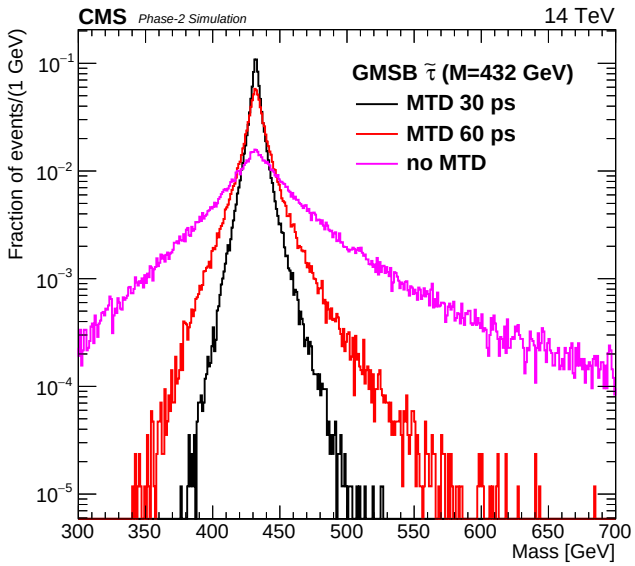


Figure 5.31: Reconstructed HSCP mass for different time resolutions.

prehensively studying a variety of charm (D^0, D_s, Λ_c^+) and bottom (B, B_s, Λ_b) hadrons over a p_T range starting from 0 up to several hundred GeV. Furthermore, the CMS-MTD will enable the study of heavy flavor production and dynamics over a wide rapidity range of at least 6 units. Such a unique capability in heavy ion physics can provide new constraints to the three-dimensional hydrodynamic evolution of the QGP medium, and probe the initial strongest electric and magnetic fields predicted to be present in the QGP fluid.

To first demonstrate the improvement empowered by the PID capability of the MTD, Fig. 5.32 shows the ratios of background Λ_c^+ and D^0 candidates, based on the generator level particles by combining three or two tracks with proper charge signs, from minimum bias HYDJET PbPb

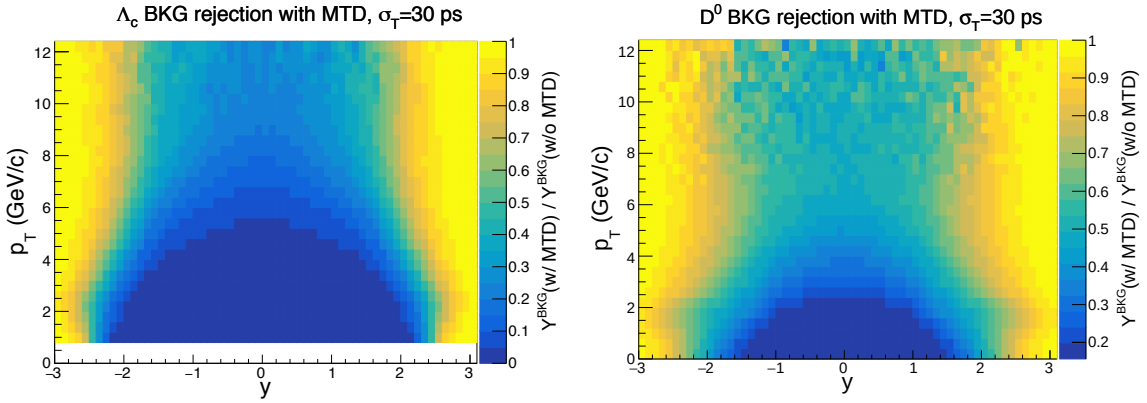


Figure 5.32: The ratios of background Λ_c^+ and D^0 candidates from minimum bias HYDJET with and without the MTD as a function of the candidate p_T and y.

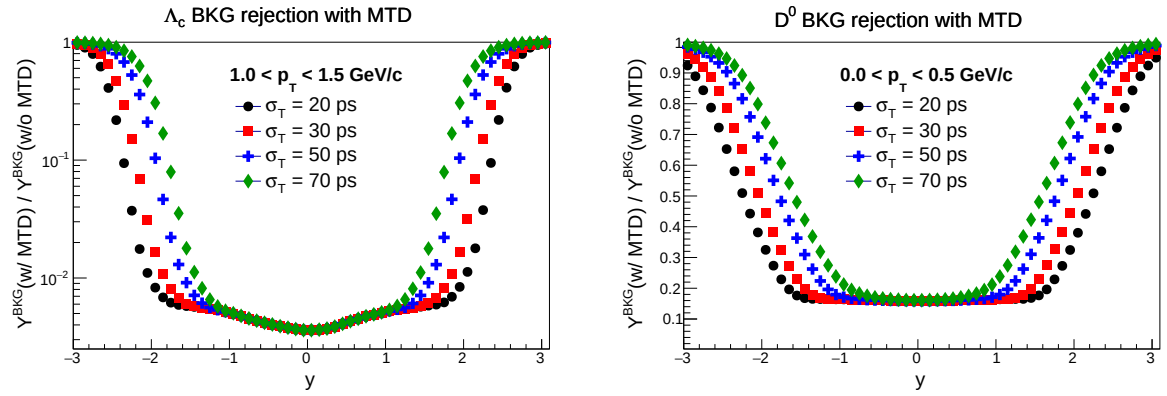


Figure 5.33: The ratios of background Λ_c^+ and D^0 candidates from minimum bias HYDJET with and without the MTD as a function of the candidate y for selected low p_T ranges.

events with and without TOFPID using the MTD as a function of the candidate p_T and y. The 1-D projections as a function of y for selected low p_T ranges are also shown in Fig. 5.33. A background reduction factor of about 200 is achieved for Λ_c^+ candidates of 1–1.5 GeV over nearly four units in rapidity. For D^0 candidates, a factor of 6–7 in background reduction can be achieved for the lowest p_T range of 0–0.5 GeV. The PID information is particularly important at low p_T, where topological selections do not provide much discrimination between signal and background. Different scenarios of MTD time resolution are also shown.

Realistic performance of heavy flavor physics in heavy ion collisions with TOFPID is evaluated in the full CMS Phase-2 simulation and reconstruction framework, including the MTD with a time resolution of 35 ps. Projected sensitivities of physics observables are compared for scenarios with and without the MTD. Signal samples of D^0 ($\rightarrow \pi^+ + K^-$) and Λ_c^+ ($\rightarrow \pi^+ + K^- + p$) hadrons are produced using the PYTHIA 8.2 event generator, while the background is modeled by the HYDJET event generator for PbPb collisions at 5.5 TeV. Samples with PYTHIA signals embedded into HYDJET background events are also used. A track is considered to be compatible with a certain species (π, K, p) if its observed inverse velocity, $(1/\beta)_{obs}$, falls within one rms of the expected value, $(1/\beta)_{exp}$. This selection has a PID efficiency of about 80%. One track can be compatible with more than one species, chiefly at high momentum. In this case, each possible species is considered in the reconstruction of the D^0 and Λ_c^+ hadrons.

The D^0 and Λ_c^+ candidates are reconstructed by combining two or three tracks with correct charge-sign and PID to form a secondary vertex. Tracks associated with an MTD hit that satisfy

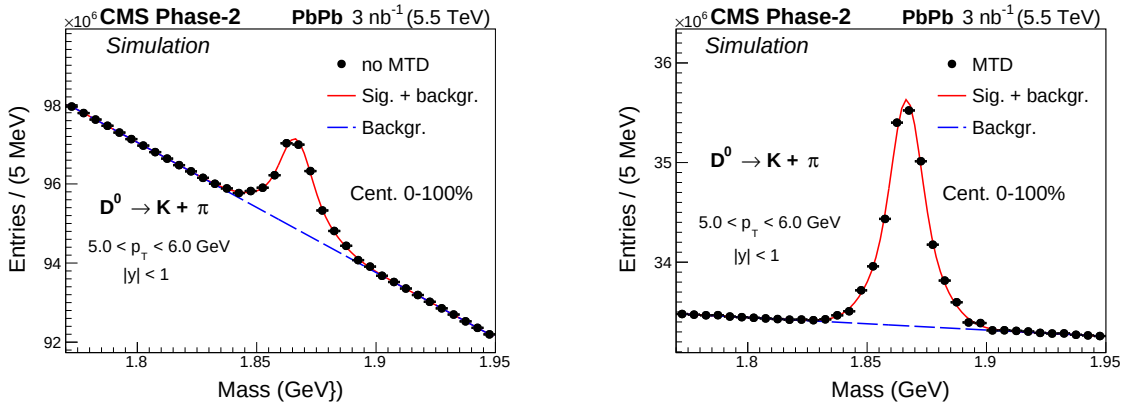


Figure 5.34: An example of projected D^0 mass distributions reconstructed via π+K decay channel in minimum bias HYDJET PbPb events at 5.5 TeV without (left) and with (right) the MTD, for $5 < p_T < 6$ GeV and $|y| < 1$, corresponding to an integrated luminosity of 3 nb^{-1}.

the kinematic selection of $p_T > 0.8$ GeV for $|\eta| < 1.4$ and $p > 0.7$ GeV for $1.4 < |\eta| < 3$ (region with high MTD efficiency) are used. Selections on several topological decay variables including the three-dimensional decay length normalized by its uncertainty; the pointing angle, which is the angle between the vector from primary vertex pointing to reconstructed secondary vertex and the total momentum of tracks; and secondary vertex fit probability, are also applied to both scenarios of with and without the MTD to further improve the signal significance.

Examples of projected D^0 ($\rightarrow \pi^+ + K^-$) invariant mass distributions in minimum bias HYDJET PbPb events at 5.5 TeV are shown in Fig. 5.34, for $5 < p_T < 6$ GeV and $|y| < 1$, without (left) and with the MTD (right). The total integrated luminosity considered is 3 nb^{-1}, which assumes conservatively that 50% of the total luminosity to be delivered in HL-LHC (or Run 4) will be recorded to disk. The D^0 signal yield is scaled based on the measured cross section at midrapidity ($|y| < 1$) for $p_T > 1$ GeV by ALICE and CMS from Run 2 [126, 127], and extrapolated to the unexplored kinematic region of $p_T \sim 0$ and $|y| < 3$ based on PYTHIA generator level distributions. The Λ_c^+ signal yields are set based on theoretical model calculations [128], because the experimental measurements in PbPb collisions are insufficient. For D^0 and Λ_c^+ backgrounds, the mass distributions for background candidates per event are first calculated using HYDJET and then scaled to the expected total number of events. As shown in Fig. 5.34, D^0 background candidates are significantly suppressed by the PID selections using the MTD, and the signal significance is drastically improved. The background rejection factors are roughly consistent with expectations shown in Figs. 5.32 and 5.33.

Based on projected signal significance of D^0 and Λ_c^+ in different ranges of p_T and rapidity, the Λ_c^+ to D^0 yield ratio in PbPb collisions serves as an important probe of quark coalescence or recombination mechanism in a hot and dense quark-gluon plasma (QGP). Theoretical calculations predict a strong enhancement of the Λ_c^+ to D^0 ratio especially in the low p_T region (e.g., see Ref. [128], compared to those in pp and pPb collisions, as shown in Fig. 5.35). The two solid curves correspond to the scenarios of Λ_c^+ and D^0 formed by coalescence only and coalescence plus fragmentation. Measurements in pp, pPb and 0–80% PbPb at midrapidity from the ALICE collaboration [129, 130] are shown in the leftmost panel of Fig. 5.35. The ratios in pp and pPb collisions are around 0.5 with little p_T dependence, while the PbPb data, currently only measured for $p_T \sim 10$ GeV, appear to be higher than those for pp and pPb, and favor the coalescence only scenario.

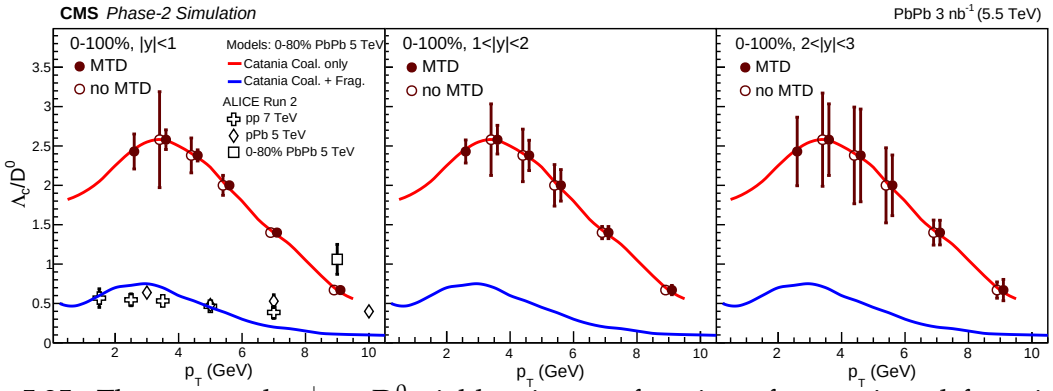


Figure 5.35: The expected Λ_c^+ to D^0 yield ratio as a function of p_T projected for minimum bias PbPb collisions at 5.5 TeV without (open circles) and with (filled circles) MTD, for rapidity ranges of $|y| < 1$ (left), $1 < |y| < 2$ (middle) and $2 < |y| < 3$ (right), corresponding to an integrated luminosity of 3 nb^{-1}. Only points with significance greater than 2 are shown. Curves represent theoretical model calculations at midrapidity assuming scenarios of coalescence only and coalescence plus fragmentations [128]. Measurements in pp, pPb and 0–80% PbPb at midrapidity obtained by the ALICE collaboration are also shown [129, 130].

Assuming the Λ_c^+ to D^0 yield ratios at values predicted by the coalescence only scenario, the projected statistical precisions are shown as a function of p_T for minimum bias PbPb collisions at 5.5 TeV, for three rapidity ranges of $|y| < 1$, $1 < |y| < 2$ and $2 < |y| < 3$. Shown are simulation results with and without the MTD. At midrapidity and low p_T region, the MTD provides one order of magnitude improvement in statistical precision, which will set stringent constraints on theoretical models of open charm hadron production in heavy ion collisions. Coverage of the heavy flavor measurement in the forward rapidity region up to $y \sim 3$ is a unique capability of the CMS detector and will provide essential inputs to the understanding of the longitudinal dynamics of the QGP medium.

The projected statistical precision for 0–10% most central PbPb collisions at 5.5 TeV, comparing with and without the MTD for midrapidity $|y| < 1$, is also shown in Fig. 5.36. The statistical uncertainties are in general larger than those for minimum bias PbPb events, due to higher multiplicity and large background. However, the coalescence effect (enhancement of the ratios) of the QGP is expected to be strongest in the most central PbPb collisions and CMS will be able to perform a high precision measurement to test theoretical predictions, largely enabled by the MTD.

The universal scaling of elliptic flow (v_2) of light-flavor mesons and baryons provides an even stronger evidence of quark coalescence in the QGP, e.g., as observed in the measurement of strange hadrons in PbPb collisions [131]. The CMS MTD will enable a test of this universal scaling in the charm quark sector between Λ_c^+ and D^0 hadrons, shedding more light on the fast thermalization of the QGP medium. Figure 5.37 shows the projected statistical precision for the elliptic flow of the Λ_c^+ and D^0 as a function of p_T for 30-50% centrality PbPb collisions at 5.5 TeV for $|y| < 1$, comparing scenarios with and without MTD. Again, results are projected to an integrated luminosity of 3 nb^{-1}. Previous measurements of strange mesons and baryons by CMS from Run-2 are also shown. The D^0 average values are based on CMS Run-2 measurements, while the Λ_c^+ average values are set to those of the Λ baryon. As shown, the high precision enabled by the CMS MTD will for the first time test the universal scaling of v_2 for open charm hadrons. The MTD will also significantly improve the precision of the D^0 v_2 measurement down to $p_T = 0$ GeV.

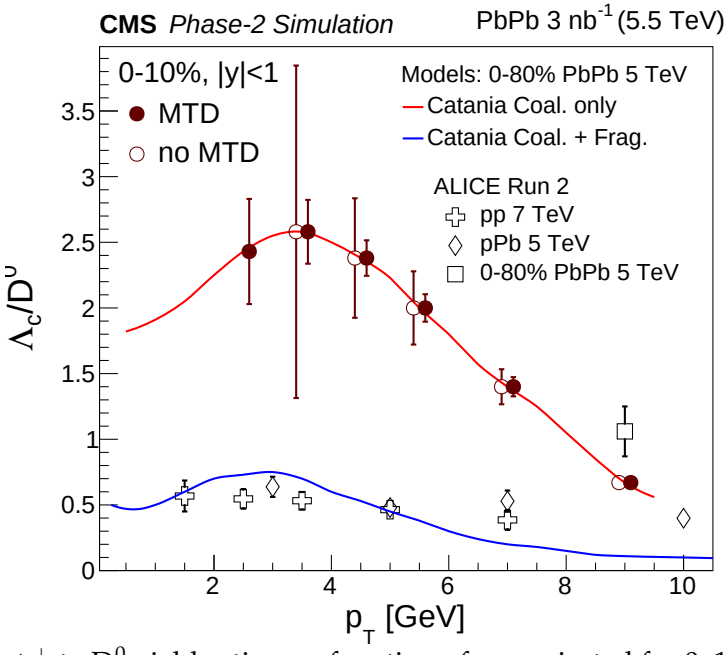


Figure 5.36: The Λ_c^+ to D^0 yield ratio as a function of p_T projected for 0–10% most central PbPb collisions at 5.5 TeV without (open circles) and with (filled circles) MTD, for rapidity range $|y| < 1$, corresponding to an integrated luminosity of 3 nb^{-1}. Only points with significance greater than 2 are shown. Curves represent theoretical calculations at midrapidity assuming scenarios of coalescence only and coalescence plus fragmentations [128]. Measurements in pp, pPb and 0–80% centrality PbPb at midrapidity by the ALICE collaboration [129, 130] are also shown.

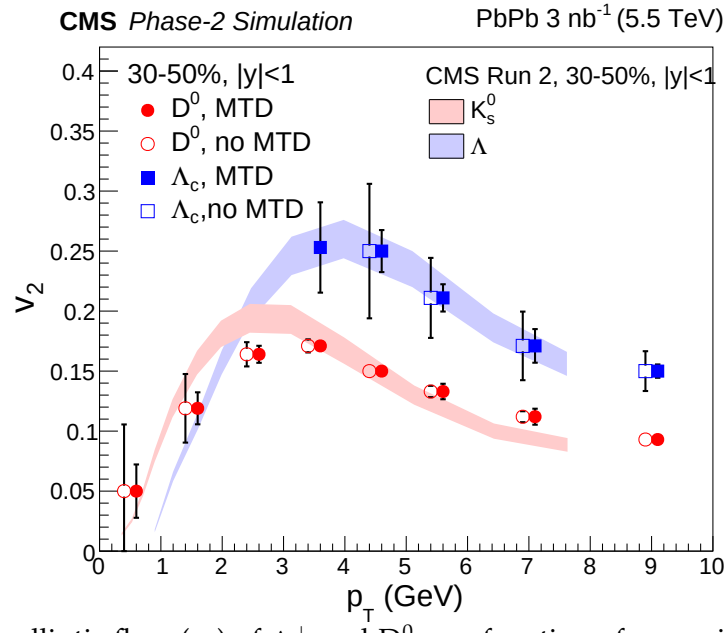


Figure 5.37: The elliptic flow (v_2) of Λ_c^+ and D^0 as a function of p_T projected for 30–50% centrality PbPb collisions at 5.5 TeV without (open markers) and with (filled makers) MTD, for rapidity range $|y| < 1$, corresponding to an integrated luminosity of 3 nb^{-1}. Only points with significance greater than 2 are shown. Measurements of strange meson and baryon v_2 for 30–50% centrality PbPb collisions from the CMS Run-2 are also shown (shaded bands) [131].

Chapter 6

Organization, schedule, and costs

6.1 Project organization

6.1.1 Introduction

The MTD project was formally created in April 2018, as a subsystem in the CMS Phase-2 (HL-LHC) Upgrade. As described in the CMS and MTD project constitutions, the MTD institution board (IB) is the highest decision-making body in the MTD project. The MTD project is led by a project manager (PM), who is appointed by the CMS spokesperson in consultation with the MTD IB. Within the MTD, the PM, assisted by one deputy, chairs its steering committee (SC) and is responsible for the scientific, technical, and managerial direction of the project. The CMS upgrade coordinator and the CMS technical coordinator provide the project oversight on general and technical aspects.

The PM shall present all important project decisions to the MTD IB for discussion and ratification via consensus or vote of its members. The MTD IB, comprising one representative per participating institution, works in a close partnership with the PM.

The organigram and structure of the MTD project are shown in Fig. 6.1. The project is structured in sub-project areas led by coordinators, appointed by the PM in consultation with the MTD IB and the relevant CMS coordination areas, or directly by the CMS spokesperson and Technical Coordinator if required by the CMS constitution. The overall technical coordination is provided by technical coordinators and managers assisted by the Project Office. The SC, described below, aids the PM in his/her function.

Currently, the task of this organization is to complete the R&D for the MTD, carry out prototype and preproduction activities, and construct, install, and commission the detector. As the construction project moves towards completion, the organization will evolve into operations mode, which may require some adjustments to its structure.

6.1.2 Organization of the MTD project

6.1.2.1 Steering committee

The SC chaired by the PM includes the deputy PM, the IB chair, the resource and technical managers, and members of the Project Office and of the Finance Board. The SC meets regularly and discusses matters of significance in the project relating to scientific, technical, financial, and managerial aspects. The members of the SC are charged to ensure that the MTD is optimized for the HL-LHC physics goals using the resources available to the project. The SC will assist in the PM in the yearly review of the project held by CMS Upgrade Coordination.

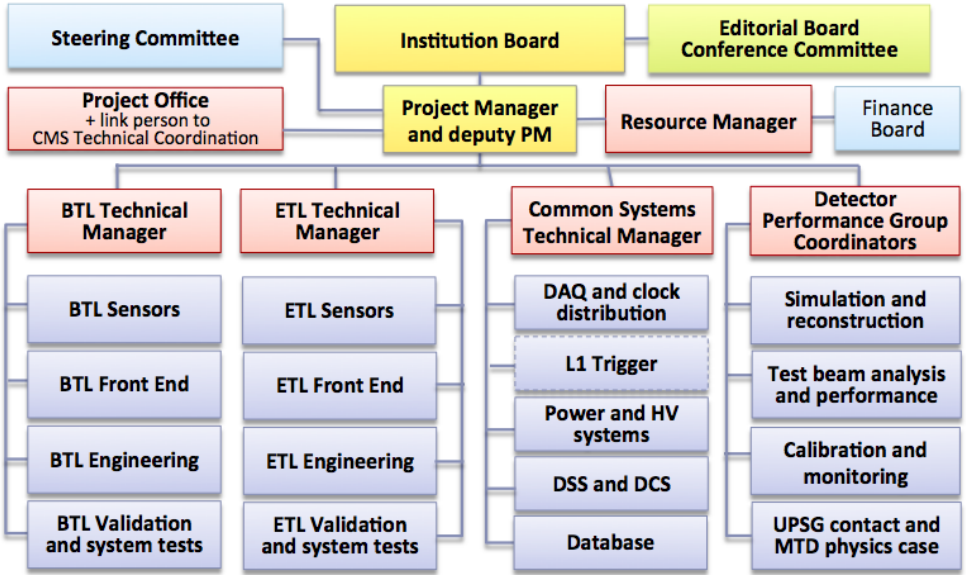


Figure 6.1: Management structure of the MTD project. The Steering Committee and the Project Office are described in the text. The interface with CMS, CERN, and external projects is described in Section 6.1.3.

6.1.2.2 Technical coordination and Project Office

Given the difference in technologies, in schedule and in integration aspects of the barrel and endcap sections of the MTD, the project is structured with two parallel coordination lines, one for the barrel (BTL) and one for the endcap (ETL). The coordination will comprise one BTL and one ETL technical manager (TM). They will be responsible for the planning, the safety, the quality assurance, and the change control of the BTL and ETL subsystems, respectively, and will cross-coordinate the technical progress of the sensor, the electronics and the engineering groups, as well as the planning of system tests and integration aspects.

The BTL and ETL TMs will be assisted by the Project Office, which includes the project manager and deputy, all the L2 coordinators responsible of technical aspects, as well as the Detector Performance Group (DPG) conveners. In addition, a link-person to the CMS technical coordination group (TCG) will liaise with the ETL and BTL technical managers and shall keep the TCG informed of MTD technical progress and the MTD of changes that are requested by the CMS TCG. The Project Office shall meet regularly to cross-coordinate, update the technical advancement and the planning of the project.

The responsibilities of the BTL and ETL TMs are defined below and shall be updated as needed to adapt to changes of personnel or in the project.

The BTL and the ETL TMs, assisted by the Project Office, shall:

- ensure that the overall project, and the various sub-projects within it, do not falter for technical reasons,
- ensure and follow the necessary market or technological surveys needed in the various sub-project areas,
- work in close contact, and meet regularly, with the CMS technical coordination,
- maintain the MTD prototyping and construction schedule,
- be responsible for adequate documentation, in appropriate platforms, across the project and its updating, and for the database used for various purposes in the

project,

- hold and monitor detailed milestones in order to follow the progress of the project,
- establish, with the appropriate sub-area coordinators, the assembly and testing centres and qualify them for production,
- establish and qualify the necessary assembly spaces at CERN,
- establish and monitor QA/QC procedures across the project,
- establish and document change control, and
- help organising yearly reviews of the relevant items in the sub-areas; an engineering design review and an electronics system review are required prior to launch of full-scale construction of mechanics and electronics, respectively.

The BTL and ETL TM will also cross-coordinate with the Common System Technical Manager, who is responsible for the progress and delivery of the systems common to BTL and ETL for the testing phase and for the final detector. These common systems include the DAQ and back-end system, the clock distribution, the power system, the detector safety and control systems (DSS and DCS), and the construction database. Each of these areas is coordinated by specific coordinators within the MTD, as presented in the project organigram (Fig. 6.1), who liaise with the BTL and ETL technical managers and with the central CMS coordination for the respective areas of interest. In addition, within the scope of the TDR, a L1 trigger coordinator is mandated to study the options for an MTD participation in the CMS L1 trigger and to prepare the necessary information for a decision. In case of a negative decision, this coordination area will be discontinued.

After the approval of the Technical Design Report, relevant changes that may become necessary in technology, design, cost, schedule, and risk will be subject to a change control process. The BTL and ETL technical managers, assisted by the Project Office, are responsible for establishing, requesting, and documenting change control, with technical details, drawings, etc. archived in EDMS. For changes that only affect the MTD project, the change control shall be assessed within the MTD, by an appropriate team identified by the project manager, and eventually signed-off by the MTD Institution Board. The MTD project manager shall also inform, at the start of the change request, the CMS Upgrade Coordination, the CMS Technical Coordination and the CMS management, who will propagate the information where relevant, and may require further actions. For changes that affect other CMS subsystems or impact the overall CMS integration and schedule, the MTD shall comply with the procedure defined by the CMS Upgrade Coordination and endorsed by the CMS Collaboration Board.

6.1.2.3 Finance board

The MTD Finance Board (FB) consists of one or more representatives of all of the funding agencies that are involved in the MTD project. The MTD FB, through the resource manager (RM) as chairperson, will deal with all matters related to the costs and resources of the project, including contracts, and all related administrative matters, in collaboration with the CMS resource manager.

The MTD FB, through its chairperson, is responsible for:

- establishing the MTD cost estimate, following the expenditure and the cost to completion of the project,
- accounting of CORE expenses for the construction of the MTD,
- providing the CMS resource manager with information required for presentation to

the CMS resources review board (RRB), e.g. preparing the budget status and planning to be presented to the RRB,

- finalizing yearly the cost book expenditure for inclusion in CMS's financial statement to the spring RRB,
- preparing of the interim status report of cost book expenditure, for internal use, in the autumn; the MTD resource manager shall present the report to the FB, and
- the policy, and followup, for multinational contracts, to be placed either by CERN (in collaboration with the resource manager) or other institutions.

6.1.3 Interface with CMS and CERN

The MTD project manager and deputy project manager are members of the CMS Management board. They are also members of the Upgrade Steering Group, where they report on a weekly basis on the progress of the project to the CMS Upgrade Project Management. The MTD resource manager is member of the CMS Finance Board.

The BTL and ETL technical managers, as well as the MTD link-person to CMS Technical Coordination, attend the weekly meeting of the Phase-2 Engineering and Integration Forum, where matters related to CMS wide projects as well as to the integration of each system are discussed. An MTD contact person with the CMS BRIL project attends the bi-weekly BRIL Radiation Simulations meeting, where results and requests for updates to the model for radiation simulations are discussed. In addition, joint Tracker-BTL meetings are held bi-weekly to discuss common aspects related to the engineering and the schedule of the projects.

The components of the CMS CO_2 cooling system and the responsibility for each component are specified in Section 4.5. There is a CERN oversight body for the CO_2 cooling system, comprising the ATLAS and CMS Technical and Cooling Coordinators as well as representatives of the CERN-EP and EN departments. The CERN EP-DT team is mandated by ATLAS and CMS to design the CO_2 cooling plants and to work in close collaboration with the CERN EN-CV team in charge for the R744 primary chiller. The CMS Cooling Coordination team, under CMS Technical Coordination, is supporting the upgrade projects, including MTD, in the design of the transfer lines from the cooling plants up to the PP1 manifolds. The development of the detector specific cooling lines and of the on-detector manifolds is responsibility of the BTL and ETL Mechanics Coordinators. The BTL and ETL teams have been working and work jointly with the CMS Cooling Coordination team to define the specifications of the cooling system and power needs for BTL and ETL. It is the responsibility of the BTL and ETL technical managers, assisted by the Project Office and by the MTD link-person to CMS Technical Coordination, to maintain the MTD system design coordinated with the central development.

The development of the MTD simulation and reconstruction software is under responsibility of the Detector and Performance Group. The DPG conveners, assisted by an MTD link-person, attend the weekly meeting of the CMS Upgrade Performance Studies Group, mandated by CMS Upgrade Coordination to coordinating the integration of the Phase-2 reconstruction software, the production of the Monte Carlo samples, and the definition of the physics benchmark to evaluate the performance of specific upgrade elements for the TDRs.

6.2 Project timeline and milestones

6.2.1 Project timeline

A simplified view of the project timeline is shown in Fig. 6.2, with several different phases separated by major decision points or deliverables. The timeline for the barrel and the endcap are shown separately at the top and bottom of the figure. The construction schedules are designed to match the constraints for installation in CMS and to provide adequate contingency within the context of those constraints.

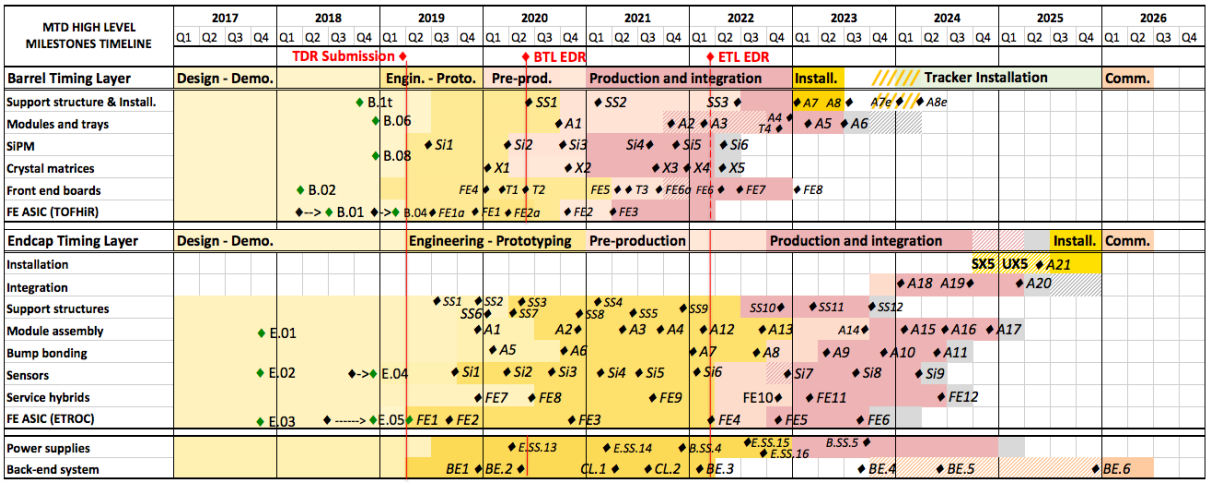


Figure 6.2: Simplified view of the MTD project timeline, for the barrel (top) and endcap (bottom) compartments with indication of the high level milestones discussed in Section 6.2.2. The colors indicate the different phases of the project as specified in the rows with legends.

The BTL installation in the Tracker Support Tube (TST) will take place at the Tracker Installation Facility (TIF) at CERN before the TST including Tracker and BTL is transported to P5 and then lowered to the CMS cavern. To understand the margins, a late integration scenario has been studied, and agreed with the Tracker, where the Tracker defines the CMS need-by date for BTL installation-complete, which are January 2024 (B.A.7e) for one and March 2024 (B.A.8e) for the other end of the TST. In this scenario (yellow hatched area in Fig. 6.2), the BTL installation should start no later than in October 2023 on one end, and January 2024 on the other end of the TST, where it would proceed in parallel with the start of the Tracker installation on the other end. The installation of the BTL trays is actually slated to begin earlier, in January 2023 (B.A.7), as soon as 25% of the trays are available, and to be completed in March 2023 on one end of the TST and on July 2023 on the other end (B.A.8). With these dates, the integration would finish about nine and eight months before the Tracker starts integration at the respective ends. There will be no access to the BTL after the installation of the Tracker endcaps begins. The TST, fully instrumented with the BTL and the Tracker, will be ready for detector commissioning at the beginning of 2026. The commissioning period can be reduced by up to three months, providing an additional three months of contingency on the installation sequence.

The ETL will be installed on the nose of the endcap calorimeter (CE) in an independently accessible cold volume. The installation is possible before or after the CE lowering in the CMS cavern, without or with the beam pipe in place. In the current schedule, the ETL installation can begin in the SX5 surface building, after completion of the integration of the modules on the wedges of one endcap (E.A.19), and continue in the cavern after the lowering of the CE. As shown in Fig. 6.2, the installation of the Dees in the cavern provides a contingency of nine months between the completion of the module integration (E.A.20) and the ETL expect-

complete date at the end of 2025. In the event of delays, servicing and installation completion is also possible during a subsequent year-end technical stop of the LHC, which have been typically 16 weeks long.

The R&D phase to demonstrate the technologies and define the detector design is substantially completed at present and is summarized in the current TDR. The following phase of engineering and prototyping has different duration for the BTL and the ETL given the different maturity of the technologies and the different installation schedules in CMS. This phase will be concluded by specific engineering design reviews (EDR), currently scheduled at the end of Q2 2020 for BTL and in Q1 2022 for ETL, which are intended to validate the design and give the green light for production. A series of system tests will be performed to validate the design, including full module irradiation tests (B.T.1, B.T.3, E.FE.10, E.A.12, and E.A.14). The EDR dates will be fixed with CMS Technical Coordination. There is some flexibility in the plan to postpone the EDR and hold a dedicated Procurement Readiness Review, for procurements of part of the system that need to be launched for schedule reasons.

For BTL, the design and prototyping phase with validation of the components as well as of systems (crystal matrices and SiPMs packaged on the sensor boards and assembled in detector sub-modules) will continue in 2019. In parallel, pre-series runs for SiPMs and crystal matrices will be carried out between the end of 2019 and the beginning of 2020 to identify vendors and be ready to place tenders soon after the EDR. The production of the sensor elements is planned to start at the beginning of 2021 and is estimated to last one year. The production and construction of the trays will take place in parallel in three main assembly centres, which will become operational for production in early 2022 (B.A.3). The ASIC is the critical path for the production of the front-end boards, which are the last element that will be integrated on the trays. The schedule allows two ASIC prototype iterations before the final production run. The first ASIC iteration will enable the finalization of the design and validation of the front-end boards. While we believe that the schedule is maintainable, ways to gain contingency with further parallelization of the integration activity at the assembly centres are possible.

For ETL, the design and prototyping phase will span the next three and a half years, with parallel developments of the sensors, of the module components, and of the structural elements, and validation of the individual components and of the systems. First prototypes of a realistic module will be built once the initial ASIC prototype is available. Before then, the assembly pipeline will be exercised and validated relying on mockups possibly using prototype ASICs designed for similar applications (Section 3.4.4). The ASIC design and subsequent production defines the critical path; three iterations of the readout ASIC are planned. Readiness for production will be achieved in November 2022 (E.FE.5). In the production phase, the production of the silicon sensors, of the readout ASIC, of the service hybrids, and of the support structures will proceed in parallel, as well as the bump bonding of the ASICs on the sensors and the module assembly, with about a quarter year offset between the last processes.

The assembly of the modules will occur in four independent assembly centres. The preproduction modules will undergo system tests and irradiation tests involving the final components. Additional radiation testing will be done on a small number of each batch during production. The modules will be shipped to CERN for final integration on the aluminum support structures (wedges), and dressing with services. The first 25% of modules will be shipped in February 2024 (E.A.15). Upon completion of the integration and testing of each set of four wedges, i.e., two Dees, they will be ready for batched installation on the CE nose. The schedule contingency to the ETL expect-complete date at the end of 2025 is nine and seven months for the module integration and the Dee installation, respectively.

6.2.2 List of milestones

The MTD project office will be responsible for following the planning of the project. A detailed schedule has been set up with the Merlin® software. Granular milestones will enable the monitoring of the progress of the project. High-level milestones, listed in Table 6.1, will be reported to the LHCC reviewers. Milestones that were initially defined in the MTD Technical Proposal have been updated and integrated to reflect the evolution of the detector design and the refined definition of the project planning. Former milestones that match new high-level milestones are indicated in the table. Milestones that were scheduled for completion before the TDR submission (Fig. 6.2) are not included in the table, except those with a revised expect-complete date.

Table 6.1: MTD project high-level milestones and identifiers.

Milestone	TP ID	TDR ID	Date
BTL Engineering design review held	*MTD.B.14*	BTL.EDR	May 2020
ETL Engineering design review held	*MTD.E.17*	ETL.EDR	Mar 2022
BTL crystals (B.X)			
Crystal vendors preselection completed	*MTD.B.11*	B.X.1	Jan 2020
Crystal matrices order placed		B.X.2	Oct 2020
Crystal matrices 25% complete		B.X.3	Sep 2021
Crystal matrices 50% complete		B.X.4	Dec 2021
Crystal matrices 100% complete		B.X.5	May 2022
BTL SiPM (B.Si)			
Optimization of SiPM, package and boards		B.Si.1	Jun 2019
Qualification of SiPM vendors completed	*MTD.B.11*	B.Si.2	Mar 2020
SiPMs order placed		B.Si.3	Oct 2020
SiPM boards 25% complete		B.Si.4	Aug 2021
SiPM boards 50% complete		B.Si.5	Nov 2021
SiPM boards 100% complete		B.Si.6	May 2022
BTL front-end electronic (B.FE)			
TOFHIR2-V1 design review		B.FE.1a	Jun 2019
TOFHIR2-V1 submission		B.FE.1	Nov 2019
TOFHIR2-V2 design review		B.FE.2a	Mar 2020
TOFHIR2-V2 submission (pre-production) and ALDO2 ready for production	*MTD.B.09*	B.FE.2	Oct 2020
TOFHIR2 ready for production		B.FE.3	Apr 2021
Readout Unit Proto1 (TOFHIR1) ready for tests		B.FE.4	Jan 2020
Readout Unit Proto2 (TOFHIR2) ready for tests		B.FE.5	Apr 2021
Start front-end cards production		B.FE.6a	Sep 2021
FE 10% complete		B.FE.6	Apr 2022
CC 100% complete		B.FE.7	Jun 2022
FE 100% complete		B.FE.8	Jan 2023

Continued on next page

Table 6.1 — continued from previous page

Milestone	TP ID	TDR ID	Date
BTL system tests (B.T)			
System test using RU Proto1		B.T.1	Mar 2020
TOFHIR2 DCR circuit tested with irradiated sensors		B.T.2	May 2020
System test with RU Proto2	*MTD.B.10*	B.T.3	May 2021
System test at TIF finished		B.T.4	Nov 2022
BTL assembly and installation (B.A)			
Prototype assembly verified		B.A.1	Sep 2020
RU and tray assembly pilot site and procedure set up		B.A.2	Oct 2021
Assembly centres qualified	*MTD.B.16*	B.A.3	Feb 2022
Tray production 25% complete		B.A.4	Dec 2022
Tray production 50% complete		B.A.5	Feb 2023
Tray production 100% complete		B.A.6	Jun 2023
BTL beginning of installation (TST end 1)		B.A.7	Jan 2023
BTL end of installation (TST end 2)		B.A.8	Jul 2023
BTL need-by 50% complete (End of access to TST end 1)		B.A.7e	Jan 2024
BTL need-by 100% complete (End of access to TST end 2)		B.A.8e	Mar 2024
BTL structures and services (B.SS)			
Tray insertion tooling and process defined	*MTD.B.12*	B.SS.1	Jun 2020
TST order placed	*MTD.B.15*	B.SS.2	Feb 2021
TST ready for installation		B.SS.3	Jun 2022
Power system testing	*MTD.B.13*	B.SS.4	Dec 2021
Power supply production started		B.SS.5	Sep 2023
ETL silicon sensors (E.Si)			
Sensor performance & radiation hardness qualified	MTD.E.04		Oct 2018
Sensor prototypes v1 tested with ETROC0		E.Si.1	Sep 2019
Sensor prototypes v2 received		E.Si.2	Mar 2020
Sensor prototypes v2 tested with ETROC1		E.Si.3	Aug 2020
Sensor prototypes v3 received		E.Si.4	Feb 2021
Sensor vendor qualification and final geometry selection	*MTD.E.10*	E.Si.5	Jun 2021
Sensor vendor selection and ready for production	*MTD.E.15*	E.Si.6	Jan 2022
Sensor production 5% complete		E.Si.7	Dec 2022
Sensor production 50% complete		E.Si.8	Aug 2023
Sensor production 100% complete		E.Si.9	Mar 2024
ETL Front-end electronics (E.FE)			
Receive ETROC V0		E.FE.1	Mar 2019
Submission of ETROC V1	*MTD.E.08*	E.FE.2	Aug 2019
Submission of ETROC V2	*MTD.E.14*	E.FE.3	Oct 2020
Submission of ETROC V3 (pre-production)		E.FE.4	Mar 2022
ETROC ready for production		E.FE.5	Nov 2022
ETROC ready for module production		E.FE.6	Aug 2023
Service hybrid mechanical prototype		E.FE.7	Dec 2019
Service hybrid initial design available		E.FE.8	Jun 2020
Service hybrid design complete	*MTD.E.09*	E.FE.9	Aug 2021

Continued on next page

Table 6.1 — continued from previous page

Milestone	TP ID	TDR ID	Date
Service hybrid irradiated preproduction system test		E.FE.10	Nov 2022
Service hybrid production start		E.FE.11	Feb 2023
Service hybrid production complete		E.FE.12	Jun 2024
ETL assembly and installation (E.A)			
Module mechanical & thermal performance tested		E.A.1	Dec 2019
Module electrical performance validated		E.A.2	Dec 2020
Module gantry assembly tested		E.A.3	May 2021
Module assembly process defined and validated	*MTD.E.11*	E.A.4	Sep 2021
Bump bonding prototypes tested		E.A.5	Feb 2020
Bump bonded modules with ETROC1 tested		E.A.6	Oct 2020
Bump bonding process defined and validated		E.A.7	Jan 2022
Bump bonding preproduction ready to start		E.A.8	Aug 2022
Bump bonding preproduction complete		E.A.9	Apr 2023
Bump bonding production 50% complete		E.A.10	Nov 2023
Bump bonding production complete		E.A.11	May 2024
Module system test using ETROC V2 prototype	*MTD.E.16*	E.A.12	Mar 2022
Module production process ready & QA validated	*MTD.E.18*	E.A.13	Sep 2022
Module irradiated preproduction system test		E.A.14	Sep 2023
Module assembly 25% complete		E.A.15	Feb 2024
Module assembly 50% complete		E.A.16	Jun 2024
Module assembly complete		E.A.17	Nov 2024
Module integration first tests		E.A.18	Jan 2024
Module integration 50% complete (ETL1 ready to install)		E.A.19	Sep 2024
Module integration complete (ETL2 ready to install)		E.A.20	Mar 2025
ETL installation on CE complete		E.A.21	May 2025
ETL structures and services (E.SS)			
Feedthrough initial design complete		E.SS.1	Jul 2019
Feedthrough mockup test		E.SS.2	Dec 2019
Feedthrough & service routing specified		E.SS.3	May 2020
Feedthrough & service channel layout finalized		E.SS.4	Feb 2021
Patch panel layout complete		E.SS.5	Jun 2021
Wedge cooling channels and manifolds specified		E.SS.6	Jan 2020
Wedge prototype test with CO_2		E.SS.7	Apr 2020
Wedge mechanics specified		E.SS.8	Dec 2020
Wedge design complete		E.SS.9	Dec 2021
Wedge production 25% complete		E.SS.10	Nov 2022
Wedge production 50% complete		E.SS.11	Mar 2023
Wedge production complete		E.SS.12	Oct 2023
LV and HV cable prototypes tested		E.SS.13	Apr 2020
LV and HV cables specified		E.SS.14	Mar 2021
LV and HV power system validated		E.SS.15	Jul 2022
LV and HV power supply production started		E.SS.16	Sep 2022
MTD back-end electronics and clock distribution (MTD.BE and CL)			
Prototype functions and interfaces defined (BTL/ETL)		BE.1	Dec 2019
Specification of BE system and clock documented in EDR		BE.2	May 2020
Decision on the BTL clock distribution system		CL.1	Mar 2021

Continued on next page

Table 6.1 — continued from previous page

Milestone	TP ID	TDR ID	Date
Decision on the ETL clock distribution system		CL.2	Jul 2021
Prototype BE functions and interfaces validated (slice test)		BE.3	Feb 2022
BE electronics procured for BTL		BE.4	Sep 2023
BE electronics procured for ETL		BE.5	Jun 2024
Integration with central DAQ and trigger complete		BE.6	Dec 2025

The overall project plan and the milestones have external dependencies on the availability of ASICs that are not developed within the MTD project, and on the development of the CMS upgrade infrastructures. Table 6.2 lists all the ASICs needed by the project, with comments on their development. Details and information on the R&D timelines are provided in Sections 2.3 and 3.3 for the BTL and ETL chips, respectively. In addition, the cooling plant and transfer lines design, currently under specification with the CMS Technical Coordination, is expected to be validated by the EDR and completed before installation.

Table 6.2: ASICs needed by the project.

	ASIC	Quantity	Description
BTL			Production ASICs required in Apr 2021
	TOFHIR	10 368	32-channels read-out chip; specific development
	ALDO	10 368	Adjustable low-drop linear regulator; evolution of an existing chip
	VL+	864	Data-link chipset with lpGBT, GBT-SCA and VTRx+; CERN development; prototype available; (pre)-production needed not later than (Oct 2020) Oct 2021.
	FEAST	2 624	DC-DC converter; CERN development; available, (pre)-production needed by (May 2020) Jun 2021
ETL			Production ASICs required in Apr/Nov 2022
	ETROC	33 248	256-channels read-out chip; specific development
	VL+	1 600	Data-link chipset with lpGBT, GBT-SCA and VTRx+; CERN development; prototype available
	bPOL	13 112	DC-DC converter rad-hard version of FEAST; CERN development; available Summer 2019

6.3 Institution interests and construction responsibilities

The project and the construction are organized in three main tasks, the Barrel Timing Layer (BTL), the Endcap Timing Layer (ETL), and the Common Systems. The latter category comprises activities that will be developed as a single system for the barrel and endcap parts, such as the DAQ and the back-end system, including the clock distribution. A broad model of the construction responsibility based on a balance between the group interests and expertises, the resource availability, and the efficiency of the model has been outlined. The detailed sharing of responsibilities for the delivery of the different detector components by collaborating groups at the participating institutes, and the associated financial commitments by the funding agencies supporting those groups, will be formalized after the project approval in a signed Addendum to the CMS Construction Memorandum of Understanding (MoU).

Table 6.3: Interest of participating institutes in the construction deliverables.

Activity	BY-INP	CN-PKU	FI-HIP	FR-IRFU	DE-KIT	HU-Deb	IT-Ge	IT-MiB	IT-Pd	IT-Rm	IT-To	IT-Ts	LT-ViU	PT-LIP	RU-INR	RU-NSU	SP-IFCA	SP-USe	CH-ETHZ	US-BU	US-UCSB	US-Caltech	US-FNAL	US-FU	US-UIC	US-UI	US-KU	US-KSU	US-MIT	US-UNL	US-NEU	US-ND	US-PU	US-Rice	US-UVa	US-UW
Barrel timing layer																																				
Mechanical systems																																				
Mechanical stuctures	.	.	.	.	.	.	.	.	v	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	v	.	.	.	.	.	.	v	.	.	.
Trays and integration tools	.	.	.	.	.	.	.	.	v	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	v	.	.	.
Tracker support tube and rails	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	v	.	.	.	.	.	.	v	.	.	.
On detector cooling pipes and manifolds	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	v	.	.	.	.	.	.	.	.	.	.
CO_2 cooling	*shared*																																			
Sensor modules																																				
Crystals QA/QC	v	v	.	.	.	.	.	.	.	v	.	.	v	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
SiPMs QA/AC	.	.	.	.	.	v	.	.	.	.	.	.	.	.	v	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	v	.	.	.	.
Interconnection boards and modules	.	.	.	.	.	.	.	v	v	.	.	.	.	.	.	.	.	.	.	.	.	v	.	.	.	.	.	.	.	.	.	.	v	.	v	.
Front-end electronics																																				
Front-end ASICs	.	.	.	.	.	.	.	v	.	.	.	.	.	v	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Front-end boards	.	.	.	.	.	.	.	v	.	.	.	.	.	v	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Concentrator and power cards	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	v	.	.	.	.	.	.	.	.	v	.	.	.	.	.	.	.	.
Clock and links	.	.	.	v	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Power system																																				
LV supplies	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	v	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Bias supplies	.	.	.	.	.	v	.	.	.	.	.	.	.	.	.	.	.	.	v	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Assembly and integration																																				
Module pre-assembly	.	.	.	.	.	.	.	.	v	v	.	.	.	.	.	.	.	.	.	.	.	v	.	.	.	.	.	.	.	.	.	.	.	.	v	.
Readout unit assembly	.	.	.	.	.	.	.	v	v	.	.	.	.	.	.	.	.	.	.	.	.	v	.	.	.	.	.	.	.	.	.	.	.	.	v	.
Tray assembly centres	.	.	.	.	.	.	.	v	v	.	.	.	.	.	.	.	.	.	.	.	.	v	.	.	.	.	.	.	.	.	.	.	.	.	v	.
Endcap timing layer																																				
Mechanical systems																																				
Mechanical stuctures (wedges)	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	v	.	.	.	.	.	.	.	.
Module mechanics	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	v	.	v	.	.	.	.	.	v	.	.	.	.	.	.	.
On detector cooling pipes and manifolds	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	v	.	.	.	.	.	.	.	.
CO_2 cooling	*shared*																																			
Silicon Sensors																																				
LGAD production and testing	.	.	v	.	.	.	.	.	.	.	v	.	.	.	.	.	v	.	.	.	.	.	.	.	.	.	v	.	.	.	.	.	.	.	.	.
Front-end electronics																																				
Front-end ASIC	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	v	.	v	.	.	.	.	.	.	.	.	.	.	.
Service cards	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	v	.	.
Clock and links	.	.	.	v	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Power system																																				
LV supplies	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	v	.	.
HV supplies	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	v	.	.	.	.	.	.	.	.	.	.	.	.	.	v	.	.
Assembly and integration																																				
Bump bonding	.	.	.	.	v	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Module production and assembly	.	.	.	.	.	.	.	.	.	.	v	.	.	.	.	.	v	.	.	.	v	.	v	.	.	.	v	.	.	v	.	.	.	.	.	.
Common systems																																				
Back-end electronics and DAQ																																				
Back-end boards	*procured commodity*																																			
Clock distribution	.	.	.	v	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Back-end firmware	.	.	.	v	v	.	.	.	.	.	.	.	.	.	.	v	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
DAQ	.	.	.	v	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	v	.	.	.
Trigger	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	v	.	.	.
Detector safety system	*ongoing discussion*																																			
SW development (non WBS)	All																																			

Table 6.3 shows the current summary of the interests expressed by the participating institutes on specific construction activities. Software tasks related to the development of the detector simulation and event reconstruction, in which all the institutes are involved, are not listed. The auxiliary Table 6.4 lists all the groups and the tag name used in Table 6.3. While these tables are accurate, formal commitments by the institutes will only be taken after funding from the respective funding agency is secured.

6.4 Cost estimate

The cost of the MTD project has been established using a cost breakdown structure with four levels. The summary of the costs, as defined by the CORE (LHCC Cost Review Committee), is shown in Table 6.5. CORE costs represent the value of the equipment installed in the ex-

Table 6.4: MTD institutes.

Tag	Institution name
BY-INP	Institute for Nuclear Problems of Belarus State University, Minsk, Belarus
CN-PKU	Peking University, Peking, China
FI-HIP	Helsinki Institute of Physics, Helsinki, Finland
FR-IRFU	IRFU, CEA, Université Paris-Saclay, Gif-sur-Yvette, France
DE-KIT	Karlsruher Institut fr Technologie (KIT), Institut fr Experimentelle Teilchenphysik (ETP), Karlsruhe, Germany
HU-Deb	Institute of Physics, University of Debrecen, Debrecen, Hungary
IT-Ge	INFN Sezione di Genova, Università di Genova, Genova, Italy
IT-MiB	INFN Sezione di Milano-Bicocca, Università di Milano-Bicocca, Milano, Italy
IT-Pd	INFN Sezione di Padova, Università di Padova, Padova, Italy, Università di Trento [c], Trento, Italy
IT-Rm	INFN Sezione di Roma, Sapienza Università di Roma, Rome, Italy
IT-To	INFN Sezione di Torino, Università di Torino, Torino, Italy, Università del Piemonte Orientale [c], Novara, Italy
IT-Ts	INFN Sezione di Trieste, Università di Trieste, Trieste, Italy
LT-ViU	Vilnius University, Vilnius, Lithuania
PT-LIP	Laboratório de Instrumentação e Física Experimental de Partículas, Lisboa, Portugal
RU-INR	Institute of Nuclear Resaerch (INR), Moscow, Russia
RU-NSU	Novosibirsk State University (NSU), Novosibirsk, Russia
SP-IFCA	Instituto de Física de Cantabria (IFCA), CSIC-Universidad de Cantabria, Santander, Spain
SP-USe	University of Sevilla, Sevilla, Spain
CH-ETHZ	ETH Zurich - Institute for Particle Physics and Astrophysics (IPA), Zurich, Switzerland
US-BU	Boston University, Boston, USA
US-UCSB	University of California, Santa Barbara - Department of Physics, Santa Barbara, USA
US-Caltech	California Institute of Technology, Pasadena, USA
US-FU	Fairfield University, Fairfield, USA
US-FNAL	Fermi National Accelerator Laboratory, Batavia, USA
US-UIC	University of Illinois at Chicago (UIC), Chicago, USA
US-UI	The University of Iowa, Iowa City, USA
US-KU	The University of Kansas, Lawrence, USA
US-KSU	Kansas State University, Manhattan, USA
US-MIT	Massachusetts Institute of Technology, Cambridge, USA
US-UNL	University of Nebraska-Lincoln, Lincoln, USA
US-NEU	Northeastern University, Boston, USA
US-ND	University of Notre Dame, Notre Dame, USA
US-PU	Princeton University, Princeton, USA
US-Rice	Rice University, Houston, USA
US-UVa	University of Virginia, Charlottesville, USA
US-UW	University of Wisconsin - Madison, Madison, WI, USA

periment, without contingency. They include costs for fabrication, construction, installation, and integration. They do not include R&D and prototype costs, costs for infrastructure and facilities at CMS institutions, nor institution personnel costs. The CORE cost estimates include spare parts to cover production losses, while spares to support long term maintenance and operation are paid separately from the M&O budget. Costs are reported here in 2018 CHF, with no correction for inflation to future years. Exchange rates based on the average value in 2018 have been adopted to convert quotes collected in currencies different from CHF.

The BTL full cost per item includes a 5% yield-loss for the SiPMs, 10–20% spares for front-end electronics, 11% for readout fibres, as well as sufficient crystals and SiPMs to build two full additional trays. The ETL costing assumes 10–20% spares. The quality of the cost estimates ranges from vendor quotes to interpolations of market surveys for the major cost drivers, including sensors and readout chips and the power systems. For the pricing of the readout electronics as well as for the engineering of the module components, we rely on careful extrapolations based on the experience with the production of similar systems.

For risk mitigation, an R&D is ongoing on the development of a clock distribution system independent of the clock distribution via the TDCS/DAQ links (Section 4.2). If such an independent distribution system is needed, additional components would be required. The cost of the front-

end components and of the optical links for this system has been estimated and amounts to about 152 kCHF for BTL and 570 kCHF for ETL, including spares. Additionally, the back-end pure-clock distribution units would cost about 42 and 155 kCHF for BTL and ETL, respectively. Differences in the front-end ASIC architecture make the risk that the baseline clock does not meet specifications moderate for BTL and low for ETL. The cost of the front-end components and of the additional optical links is included in the CORE cost estimate of the BTL front-end electronics in Table 6.5. No costs are included for ETL.

The funding agencies involved with each aspect of the project are expected to provide contingency for needs in their area, should variations of the exchange rate, or unforeseen technical or vendor issues modify the cost estimate. If the need exceeds the available funds, the issue will be discussed in the MTD management and Finance Boards, and the CMS management and Finance Boards, and solution sought that may involve technical choices or a different tuning of the cost sharing.

Table 6.5: Estimated cost of the MTD project.

	Item	Cost (kCHF)
	MTD	**20 685**
1.	**Barrel timing layer**	**8 712**
1.1	Mechanical structures	372
1.2	Sensor modules	4 836
1.3	Front-end electronics	1 535
1.4	Power system	770
1.5	Cooling	1 198
2.	**Endcap timing layer**	**10 864**
2.1	Mechanical structures	174
2.2	Sensor modules	4 889
2.3	Front-end electronics	3 576
2.4	Power system	1 145
2.5	Cooling	1 079
3.	**Common items**	**1 109**
3.1	Back-end electronics and DAQ	600
3.2	Safety system	86
3.3	Installation and test infrastructure	424

Appendix A

Radiation environment

The choice and the validation of the technologies for the active elements of the MTD and the optimization of the design rely on the prediction of the dose rate and the particle fluence for each particle type at the HL-LHC. Simulations are used to predict the magnitude and composition of radiation as a function of the integrated luminosity.

The radiation simulations are performed by the CMS BRIL (Beam Radiation Instrumentation and Luminosity) project using the FLUKA 2011.2x.4 Monte Carlo multi-particle transport - code [21, 22]. The event generator DPMJET III is used to create the primary proton-proton events. It is directly linked to the FLUKA code and used as the default event generator for high energy hadron-nucleus and nucleus-nucleus interactions. All particles are transported until a predefined energy cut-off is reached. The output is usually averaged over all simulated primary events and normalized per primary event. The normalization used for the prediction of dose and fluence depends on the inelastic collision rate and the duration of LHC operation (total radiation period). For an instantaneous luminosity of 5×10^{34} cm^{-2}s^{-1}, and an inelastic cross section of 80 mb, an average of 4×10^9 inelastic pp events per second are produced [132].

The relevant quantities to estimate the impact of radiation damage to the MTD elements are the dose, the fluence of charged particles, and the 1 MeV neutron equivalent fluence in Silicon, which is calculated by weighting all particles with their non-ionizing-energy-loss (NIEL) damage cross section function to the same NIEL of 1 MeV neutrons. Doses and fluence are estimated using the FLUKA USRBIN scoring system where quantities are recorded in a binning mesh that can be Cartesian, in cylindrical coordinates, and independent of the geometry. As a special case, quantities can be bound to a particular geometry region. Particle fluence is obtained by calculating the track length density, and doses using the energy deposition. The obtained 3D distributions can be projected to two dimensions and shown as a 2D flux map, or to one dimension and shown as a graph.

The FLUKA study version 3.7.18.0 is used in this TDR, which features an updated Phase-2 CMS geometry to provide updated simulations for HL-LHC conditions. The geometry, detailed in the next section, is derived from the Phase-2 CMS Technical Design Reports [4–7] and this document, for the MTD geometry. This new geometry model updates and supersedes the model adopted in those TDRs (CMS FLUKA simulation version 3.7.2.0), which was based on the latest nominal Phase-1 CMS geometry with minimal modifications to specify the Tracker Phase-2 layer structure, a rough description of the bulk material of the new endcap calorimeter, and the replacement of the CMS endcap preshower with an 18 cm thick neutron moderator. This preliminary model was also used in the preparation of the MTD Technical Proposal [8], although it did not include the implementation of the MTD itself, and the neutron moderator in front of the endcap calorimeter was thicker than possible with inclusion of the MTD.

The details of the new model are specified below, as well as its uncertainties. The particle

fluence with the new geometry, reported in Table 1.3, have increased by about 25% in BTL and 35% in ETL compared to the earlier model presented in the Technical Proposal. The model is still being refined and the geometry may still be optimized with additional moderator to reduce the radiation levels. However, as we discuss below, a series of tests has provided confidence on the stability of the predictions against residual tunings of the geometry description, and this model has been adopted as a new reference.

A.I Geometry model and simulation parameters

A FLUKA geometry is created by combining regions of basic geometric bodies. These shapes can be finite objects such as spheres, boxes, or cones, or infinite elements such as planes or infinite cylinders. The complexity of the geometry that can be created is limited by the number of objects or elements that can be reasonably used. A large number increases the computing time. Hence FLUKA geometry models are a simplification of the reality, composed of the objects and elements and the material within them. It is, however, important that correct material compositions and densities are used, resulting in the correct total mass, to achieve a reliable description of particle shower cascades. For activation studies also trace elements that have the potential to get highly activated are taken into account. Cut-offs and transport parameters are configured per volume.

The updated model for the Phase-2 CMS detector (CMS FLUKA study v.3.7.18.0) is shown in Fig. A.1 (top) and compared to the preliminary CMS FLUKA study v.3.7.2.0 (bottom). The new model includes an accurate description of the Phase-2 Tracker, HGC and ECAL, the respective services, the Phase-2 beam beam pipe, and the MTD detector as detailed below.

The Tracker model is consistent with the TDR geometry. In addition to the Si sensor layers, it includes support and service structures. The Tracker bulkhead and the services between the Tracker volume and the endcap region are implemented in accordance with the most recent design, as provided by the Tracker technical coordination. The Phase-2 beam pipe is modelled to be consistent with the version approved at the beam-pipe Engineering Design Review, and consistent with the technical drawing v5.4.1. The CMS endcap is modelled by a geometry to represent the Phase-2 endcap calorimeter, with thermal screens, and with a detailed longitudinal segmentation, which was not included in FLUKA v.3.7.2.0. As shielding material, borated polyethylene with a boron concentration of 5% is used. The thickness of the shielding is 12 cm, instead of 18 cm as in the earlier version. The service channel between the barrel and the endcap region is also described according to the Phase-2 geometry; the barrel and endcap detector envelopes, as well as the service channel width are consistent with the most recent update of the Phase-2 engineering (J. Strait, "CMS Barrel and Endcap Envelopes", EDMS 1895113, Internal document). In the transition from FLUKA CMS version 3.7.2.0 to 3.7.18.0, the material budget of the ECAL has also been updated.

The BTL active volume is modelled as a continuous cylinder of LYSO crystals, with a density of 7.4 g/cm^3, and an inner and outer radius of 116.1 and 116.475 cm respectively for $|z| < 88.1$ cm, of 116.1 and 116.4 cm for $88.1 < |z| < 115.1$ cm ($|\eta| < 1.1$), and of 116.1 and 116.34 cm for $115.1 < |z| < 260$ cm ($|\eta| > 1.1$). A separate BTL layer is included to represent the SiPMs, the electronics, and the mechanics. It is an admixture of Aluminum, Carbon fiber, Silicon and glue for a total mass of 800 kg, comparable to the mass of LYSO. The ETL is modelled as a 4 cm thick layer of material with density 1.0277 g/cm^3, composed of 0.871 Aluminum and 0.129 Silicon by mass fraction. The thermal screen is included as 3.2 cm thick layer of Aerogel (Silicon and Oxygen) and density 0.16 g/cm^3, surrounded by 15 mm of PermaGlass.

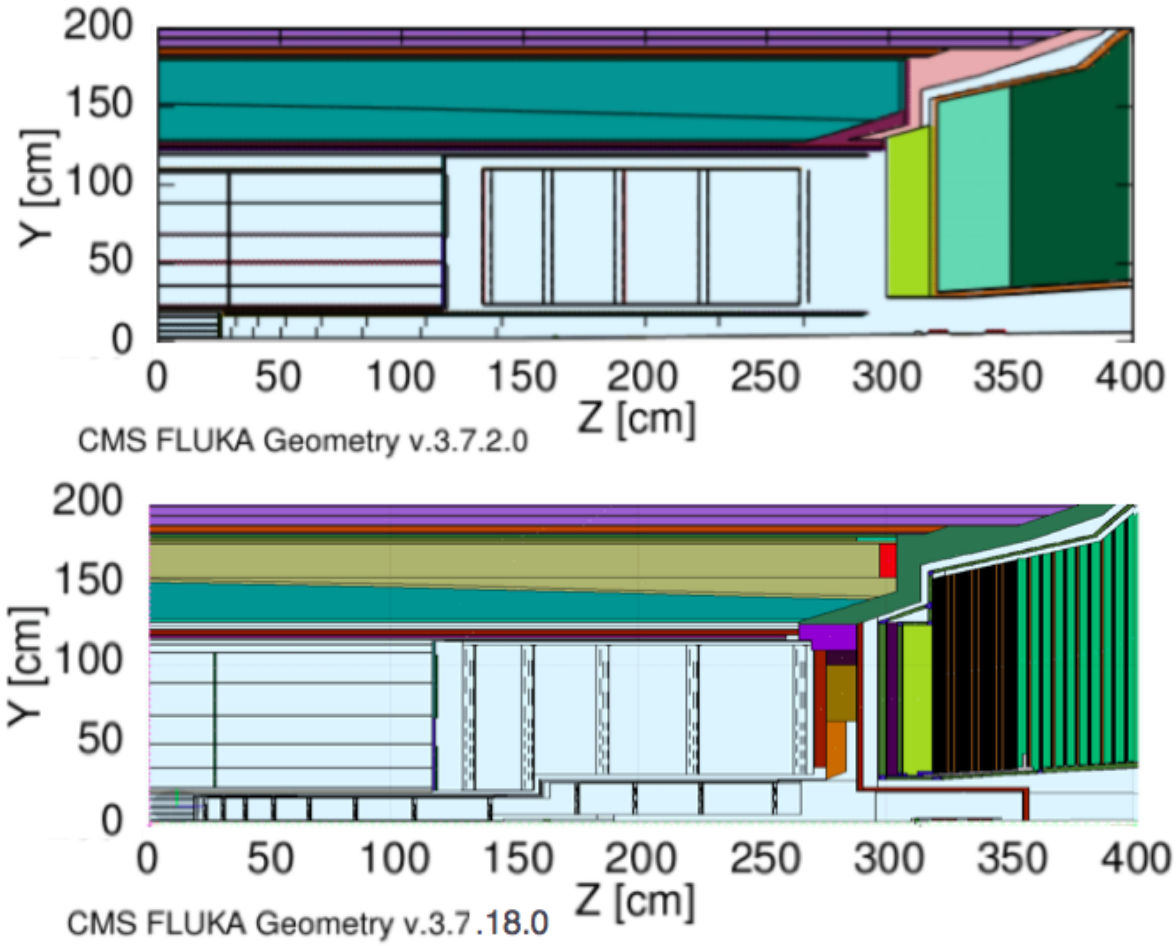


Figure A.1: YZ cut of the CMS geometry in the FLUKA simulation v.3.7.2.0 (top) and v.3.7.18.0 (bottom) used to predict particle's fluence. In the bottom figure the MTD is shown in violet, wth the BTL on the inner side of the TST (brown) and the ETL in front of the neutron moderator (light green). Visible are changes in the barrel/endcap transition region with services. More details are provided in the text.

The primary proton-proton collisions are simulated at 7 TeV energy per proton. The protons collide under a crossing angle of 590 μrad. The events are randomly distributed within 5 cm along the z axis. Particles are transported until they decay or until their energy falls below their transport cut-off. In this case, their remaining energy is deposited on the spot. All hadrons and muons have an energy cut-off of 1 keV, except neutrons, which are transported down to an energy of 0.01 meV. Electrons and photons are assigned different cutoffs in different materials. In most materials 5 keV for photons and 30 keV for electrons are applied. In heavy materials, the cut-offs are increased. In the beam pipe 10 keV for photons and 100 keV for electrons are used, while in the calorimeters the cut-offs are 30 keV for photons and 300 keV for electrons. In the target absorber the cut-offs are 300 keV for photons and 3 MeV for electrons. The choice of these cut-offs is a trade-off between the accuracy of the predictions and the computation time. The uncertainties of the predictions are dominated by the accuracy of the geometry model, discussed below.

A.II Fluence and dose predictions

All the results of the particle fluence, energy dose, and 1 MeV neutron equivalent in silicon are scored in an $r - z$ grid. There is no segmentation in ϕ. Given that the MTD is made of thin timing layers, profiles along z and along r are presented for BTL and ETL, respectively. Figures A.2 and A.3 shows the comparison of the 1 MeV neutron equivalent fluence in silicon and of the dose prediction with the old (v.3.7.2.0) and the new (v.3.7.18.0) FLUKA simulation runs. Results for BTL are shown in the left panels; results for ETL are shown in the right panels. The dotted grey lines in Fig. A.2 show the nominal prediction of the new FLUKA simulation run multiplied by a factor 1.5. Given the uncertainties in the predictions associated to the geometry model and those related to possible sensor-to-sensor variations, the BTL and ETL components are required to stand a 1 MeV neutron equivalent fluence at least a factor 1.5 (safety factor) larger than highest value of the nominal fluence for BTL and ETL, respectively. These values are achieved at $|z| \sim 240$ cm in BTL and at at the inner radius in ETL ($r = 30$ cm).

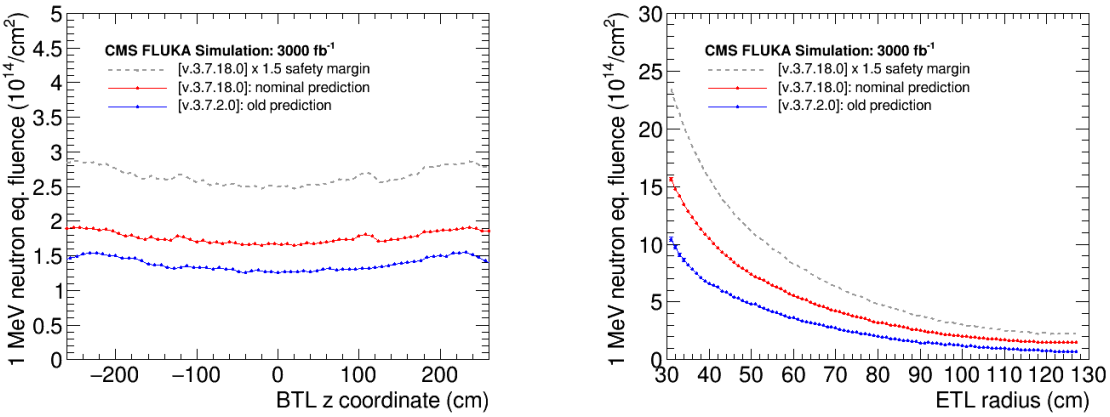


Figure A.2: Fluence (1 MeV neutron equivalent) prediction in BTL (left) and ETL (right) for an integrated luminosity of 3000 fb^{-1} from the FLUKA simulation run v.3.7.18.0 (red), with updates to the CMS Phase-2 geometry model detailed in the text, and from the FLUKA CMS simulation v.3.7.2.0 (blue) used in the MTD technical proposal and in the TDRs of the other CMS upgrade projects. The dotted line indicates the nominal fluence from the FLUKA simulation run v.3.7.18.0 multiplied by a safety factor 1.5 to account for uncertainties in the predictions.

The 1 MeV neutron equivalent fluence with the new model is about 25% and 35% higher than with the previous geometry model in BTL and ETL, respectively. A local dose increase in BTL of 50% is also observed, related to a local increase of the photon fluence. The observed changes in the predictions of the particle fluence of about 25% and 35% in BTL and ETL, respectively, do not have a single origin. Contributions are coming from the reduction of the neutron moderator thickness (up to 15% in ETL), from changes in the envelope and in the segmentation details of the endcap calorimeter, from changes in the beam pipe, and from the description of the services. The contribution to the fluence variation from the MTD detector itself is marginal. A variation in fluence predictions is also observed in the Tracker volume, and amounts to an in crease about 10% at the location of the sensors used to validate the Tracker technology. This is well within the Tracker safety margins and has no consequences on the Tracker design.

The 25% variation in the 1 MeV equivalent neutron fluence in BTL translates into a proportional increase in the DCR noise in the SiPMs, and in turn in the time resolution, causing a degradation, if not mitigated by other means, of up to about 6 ps at 3000 fb^{-1} and 8 ps at the

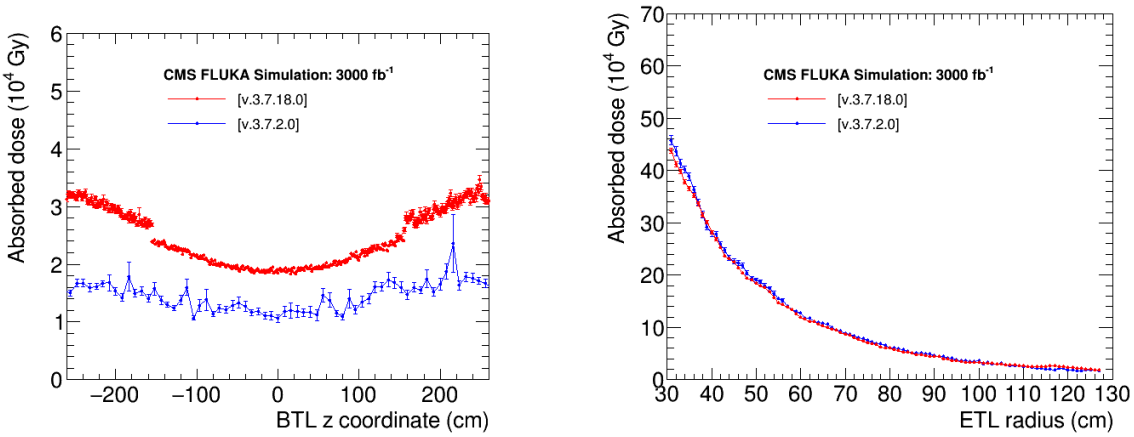


Figure A.3: Predicted doses in BTL (left) and ETL (right) using the updated FLUKA CMS simulation v.3.7.18.0 (red) and the preliminary geometry model of the FLUKA simulation run v.3.7.2.0 (blue).

maximum luminosity that can be potentially integrated by the MTD (4000 fb^{-1}). Operation at a slightly lower temperature ($\Delta T = -4\,^{\circ}\text{C}$) would compensate this difference fully. Annealing at $+50\,^{\circ}\text{C}$ would provide additional mitigation, as discussed in Appendix B.

A.III Uncertainties and safety margin

The statistical uncertainties are given by the error of the mean values, which are obtained by averaging over many simulation cycles; they are of the order of a few percent for the 1 MeV neutron equivalent fluence in silicon and below 10% for the total ionizing dose (TID). The systematic uncertainty depends on the quantity and region of interest. Typically, there are two main contributions to the systematic uncertainties for all results: uncertainties due to the event generator, and uncertainties due to the simplification of the CMS FLUKA model in terms of volumes and materials. The uncertainties vary from region to region and are expected to be higher where particles originate from long secondary cascades, being relatively strongly influenced by the geometry model. An additional systematic error is introduced by the normalization of the quantities, such as the inelastic collision cross section at $\sqrt{s} = 14$ TeV, which is predicted by various event generators. The value of the inelastic cross section can be updated, once a measurement has been performed during Run 2 operation.

A cross check of the material budget in terms of radiation and interaction lengths of the CMS TDR geometries and of the model implemented in the CMS FLUKA provides a validation of the geometry model. While a detailed assessment of the uncertainties associated to the new geometry model is ongoing, the 25–35% change in the fluence prediction following a rather radical change in the geometry description from v.3.7.2.0 to v.3.7.18.0, as shown in Fig. A.1, can be used to make a conservative estimate of the the size of this uncertainty.

In addition, variation studies were conducted with several modified geometries to assess the dependence of the predictions on residual tunings of the geometry. All these studies indicated fluence changes of a few percent at most, thus corroborating the assumption that an uncertainty of 30% related to the geometry model is conservative. For example, the left panel of Fig. A.4 shows the ratio of the predicted fluence in BTL with two geometry variations normalized to the FLUKA simulation run v.3.7.18.0. The red dots show the fluence relative variation when

ETL is dropped from the CMS geometry and indicate that results are quite insensitive to the details of the description of the ETL. The green line show the fluence reduction upon the addition of 2 cm of neutron moderator in front ETL, which is the maximum thickness that can be accommodated. The fluence reduction is at most 5% in the outer BTL, which is equivalent to what can be achieved by operating the detector at $\Delta T = -1\,^{\circ}\mathrm{C}$ lower than the nominal operating conditions. Following this study, the baseline thickness of the neutron moderator is fixed to 12 cm, keeping 2 cm of clearance for further adjustment of the ETL engineering. A variant of the HCGal longitudinal segmentation and of the HGCal inner envelope, where a cylindrical support structure is adopted instead than conical, has also been studied, showing a variation in fluence of about 2% both in BTL (blue dots in the left panel) and in ETL (right panel of Fig. A.4).

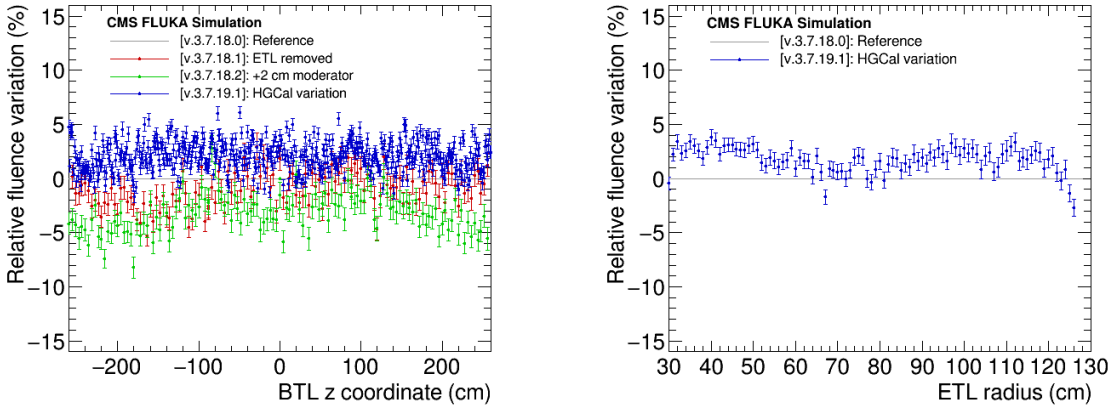


Figure A.4: Shown are the relative 1 MeV neutron equivalent fluence change in percent in BTL (left panel) and ETL (right panel) for several different variations of the geometry model (see text for details).

The uncertainty associated to the generator of the proton-proton collisions have been estimated by comparing the particle fluence predictions using the DPJMET III generator and the PHYTIA8 generator tuned on the CMS measurements and extrapolated to a center-of-mass energy of 14 TeV. The differential cross section for particle production as a function of η is well matched between the two generators up to $|\eta| \sim 2$, while it shows differences in the region $2 < |\eta| < 5$. The associated fluence variations reach about 50% at $|\eta| \sim 5$, but are limited to below 20% in the region $2 < |\eta| < 3$ relevant for the MTD coverage.

In summary, a conservative estimate of the uncertainties associated to the geometry model and to the proton-proton production cross section provides a total potential variation in the fluence prediction below 50%, which is well consistent with the assumed safety margin of 1.5 in the fluence predictions.

A.IV Considerations on radiation protection for ETL maintenance

We do not expect a significant dose to personnel during installation because much of the surrounding detector will be new, in particular the CE on which the ETL wedges will be installed. The installation procedure is relatively short, with the mounting of modules and dressing of cables on the face of the aluminum support wedges performed in advance. Maintenance operations and any potential replacement, which would occur after a prolonged irradiation time, would have a higher ambient radiation environment. Estimating the expected dose for personnel in that case require simulation studies that are being finalized. They will be used to

develop safe maintenance and replacement procedures. Extended maintenance work, such as replacement of a number of modules, would be accomplished by dismounting the wedges and craning them upstairs to a controlled laboratory environment.

Appendix B

The Barrel Timing Layer: additional technical information

This appendix completes the information provided in Chapter 2 on the BTL sensor choice and on the radiation tolerance. Section B.I describes the radiation tolerance studies performed on the individual components of the BTL sensors, as well as the plans to validate the complete sensor package. Section B.II provides details on the test beam campaigns performed on sensor prototypes of different geometries that contributed to the choice and the optimization of the sensor layout presented in Chapter 2.

B.I Studies on BTL radiation tolerance

B.I.1 Radiation tolerance of BTL sensor parts

The BTL sensors consist of five components which have been tested and optimized with specific studies: the LYSO:Ce crystals with wrapping, the optical glue, the resin layer protecting the SiPM and the SiPM. The optimization and qualification of the radiation tolerance of each of these elements is summarized in the following. The components have been qualified to a radiation level which includes a safety factor of 1.5 on top of the highest dose and fluence expected in BTL after 3000 fb^{-1}, as reported in Table 1.3.

B.I.1.1 LYSO:Ce crystals

Radiation damage in crystals is commonly divided in two components: damage from ionizing radiation, which can create localized defects in the crystal lattice, and damage from energetic hadrons (protons and neutrons), which can create larger displacements in the crystal structure. In both cases, the creation of color centers can occur, causing undesired absorption of the scintillation light within the crystal. The LYSO:Ce crystals have been widely qualified to radiation levels beyond the expect radiation doses and fluence expected in the BTL. For a large number of tested crystal vendors a negligible loss of transparency was observed as discussed in Chapter 2. Different wrapping materials have also been tested for radiation tolerance. While Tyvek and Teflon show substantials structural degradation, the specular reflectors tested (e.g. Al-foils and ESR) showed no substantial loss of reflectivity and maintained structural integrity.

Specific radiation tests have been made on the BTL crystal bars, namely irradiation to the integrated ionizing dose and to the integrated fluence of protons. The transparency and time resolution of the samples have been measured before and after irradiation. The results are reported in Fig. 2.11 and show no effect of radiation damage on the crystal timing performance.

Additional tests are planned on the crystals from eight vendors that responded to a preliminary request for quote carried out during September and October 2018 and that are currently being

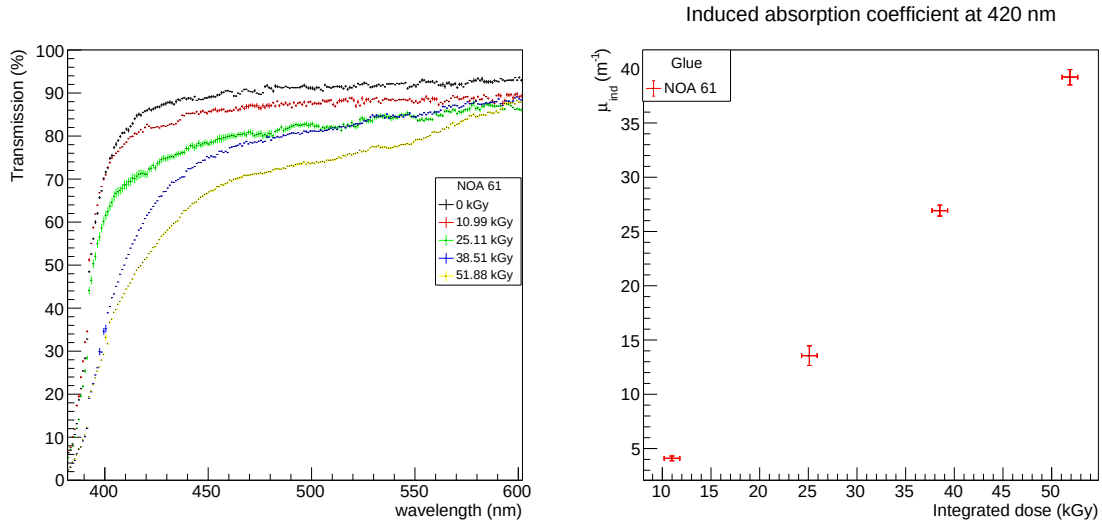


Figure B.1: Left: Transparency of a 5.8 mm thick sample of NOA61 glue after different levels of ionizing dose. Right: calculated induced absorption coefficient at 420 nm as a function of integrated ionizing dose.

qualified. Both single bars and crystal matrices with wrapping will be irradiated to the a total ionizing dose beyond that expected during HL-LHC. The light output and time resolution will be measured before and after the irradiation. This test is planned for November 2019 at the ENEA facility in Casaccia (Italy) using a Co-60 source. The test will also include irradiation to different dose rates representative of the HL-LHC running conditions and will allow to further refine the crystal specifications by December 2019.

B.I.1.2 Optical glue

Several glue candidates for BTL were tested. The pre-selection of glues capitalized the experience of the CMS collaboration for the construction of the ECAL calorimeter. The set of glues tested included NOA-61, RTV-3145, Epotek, Polytec and BC-600. Several samples of each glue were molded using a Teflon holder and transmission measurements were performed before and after irradiation. The irradiation was performed using a Cs-137 source at a rate of 2 Gy/min providing the full BTL expected dose in a few weeks.

Some of the glues (Epotek, Polytec and BC-600) showed substantial transparency loss while the NOA-61 and the RTV-3145 showed a satisfactory radiation resistance. Given the induced absorption coefficient, μ_{ind}, for a ionizing dose of 40 kGy of about 25 m^{-1} and since the glue layer in BTL will be thinner than 50 μm, the expected signal loss at the end of operation is smaller than 1.3%.

Both glues have been used to couple pairs of SiPMs to a crystal bar and several thermal cycles from -40 to $+60\,^{\circ}\text{C}$ have been performed. Neither of the glues showed visible transparency loss nor structural degradation (no cracks). The mechanical bond between crystal and SiPMs was found to remain stable throughout the cycles. Additional studies using full arrays of crystals will continue in 2019 to optimize the gluing procedure (thickness and uniformity of the glue layer). While the RTV-3145 — the same glue used in the CMS ECAL — is considered a valid option for BTL, further tests are still ongoing and the final choice will also be informed by considerations on the module assembly procedure, such as the curing time of the glue, etc.

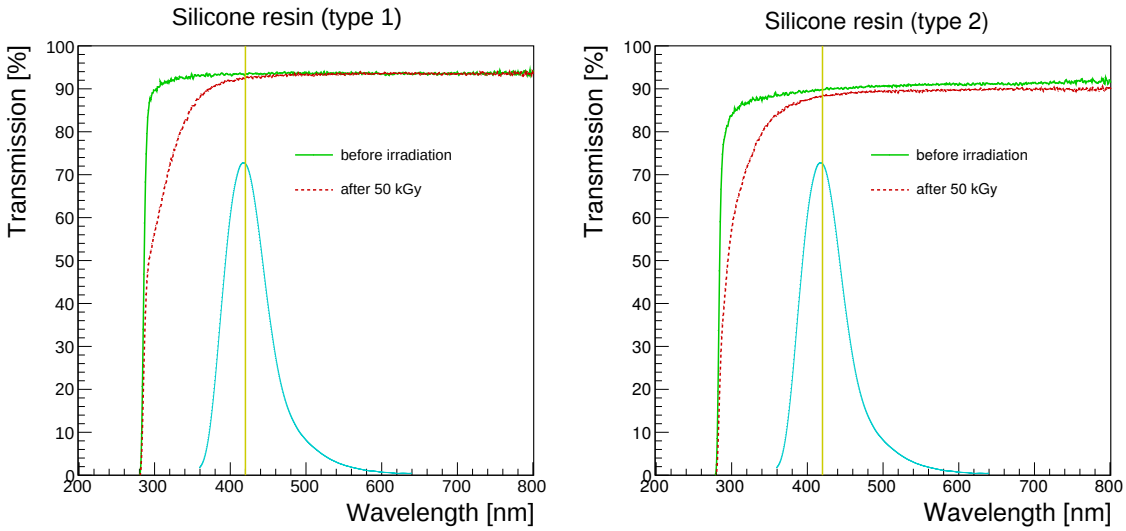


Figure B.2: Transparency of silicone resins from different vendors before and after irradiation to 50 kGy of ionizing dose from Co-60 source. The left sample has a thickness of 0.3 mm while the right one is about 1.35 mm thick. The yellow line shows the peak of LYSO:Ce emission at 420 nm while the light blue curve is the full emission spectrum of LYSO:Ce.

B.I.1.3 SiPM protective resin

Several types of resins and materials are commercially used for the protection of the SiPM active area. These include glass, epoxy and silicone. Samples from each of these options have been tested under protons, neutrons and ionizing radiation to the integrated levels expected for BTL. An irradiation with 24 GeV protons was performed at the CERN PS IRRAD facility, an irradiation with neutrons at the JSI nuclear reactor in Ljubljana, and irradiations with gamma-rays from Co-60 at both the C60 (CERN) and IONISOS (Dagneux) facilities. These studies indicated that the transparency loss is mainly induced by the ionizing dose. Therefore, irradiations with gamma-rays are a sufficiently adequate tool for the study of radiation tolerance in these materials.

The results of the study showed a poor radiation tolerance of the glass samples, while the silicone resins proved to be the most radiation tolerant solution with less than 2% transparency loss at the 420 nm (wavelength of the LYSO:Ce emission peak) over 300 μm thickness. Some of the epoxy based samples showed better radiation tolerance than others but they all featured an overall poorer transparency in the UV than silicone resins. Both SiPM vendors which are currently being considered for the BTL project can provide a radiation tolerant silicone resin for the protection of the active area, as shown in Fig. B.2, and such solution is thus considered the baseline.

B.I.1.4 SiPMs

Several specific tests of radiation hardness have been made on the SiPM technological choices for BTL. With the radiation levels that will be experienced by the MTD, the increase of the SiPM dark current induced by the ionizing radiation is negligible compared to the increase induced by the 1 MeV neutron equivalent fluence, which is therefore used to evaluate the SiPM radiation tolerance. The increase in the SiPM dark count rate is linear with the 1 MeV neutron equivalent fluence, once the effect of loss of transparency in the protective resin is factored out. This effect and its impact on the BTL operation conditions and timing performance are discussed in detail in Section 2.2.2.

The linearity of the dark count rate growth with the fluence has been demonstrated up to the nominal radiation level expected in BTL after 3000 fb^{-1} of integrated luminosity including a safety factor of 1.5. Table B.1 shows the values of current measured for SiPMs after 2×10^{12} and 2×10^{14} $\text{n}_{\text{eq}}/\text{cm}^2$. According to the irradiation facility, the estimate of the integrated fluence has an uncertainty of 20%, which is translated in the table with an uncertainty on the currents.

Table B.1: Values of dark current measured for 9 mm^2 area SiPMs at +1 V OV and T= $-30\,^{\circ}\text{C}$. The uncertainties on the currents reflects the uncertainty on the nominal fluence, as discussed in the text.

SiPM technology	I after $\phi \sim 2 \times 10^{12}$ $\text{n}_{\text{eq}}/\text{cm}^2$	I after $\phi \sim 2 \times 10^{14}$ $\text{n}_{\text{eq}}/\text{cm}^2$
HPK S12572-015	5.6 ± 1.1 μA	0.45 ± 0.09 mA
FBK NUV-HD-TE	10.0 ± 2.0 μA	1.14 ± 0.23 mA
HPK HDR2-015	16.8 ± 3.4 μA	1.75 ± 0.35 mA

Additional samples have been exposed to a total fluence of 4–5$\times 10^{14}$ $\text{n}_{\text{eq}}/\text{cm}^2$ for confirmation of the linear behaviour of the radiation damage effects beyond the BTL expected fluence. Such samples, requiring a longer radioactive cool down, will be available for measurements after October 2019.

B.I.2 Time resolution of irradiated SiPMs

Given the negligible transparency losses on the optical components (crystal, glue, resin), the global effect of radiation damage on the BTL sensors is largely dominated by the impact of the increase in the dark current and dark count rate of the SiPMs. The way this additional noise component affects the time resolution is well understood from measurements and from a simulation model validated by the measurements.

A convenient way to estimate the impact of the DCR is by emulating a radiation induced dark current in non-irradiated devices through the injection of light (randomly distributed in time), as described in Section 2.1.2. Such test allowed us to parameterize the sensor performance within a range of signal amplitudes and dark count rates. Additional measurements have been made using a 42 ps FWHM laser pulse to produce a controlled number of photoelectrons in the SiPM, while a LED — operated in continuous mode — was used to generate dark counts up to 55 GHz as expected in BTL after the integrated fluence at 3000 fb^{-1}. The results of such measurements are reported in Fig. B.3, where the contribution of the DCR term to the time resolution is parameterized as a function of DCR and of the signal amplitude (photoelectrons). The results are consistent with the LED study presented in Section 2.1.2 and confirm the power-law scaling of the DCR term.

Similarly, the time resolution of irradiated SiPMs (S12572-015C type) to laser pulses of different amplitudes has been measured as a function of over-voltage. The set of SiPMs tested were irradiated to fluences of 1.0×10^{12}, 2.0×10^{12} and 2.0×10^{13} $\text{n}_{\text{eq}}/\text{cm}^2$. By operating such SiPMs at a temperature of $20\,^{\circ}\text{C}$, the level of dark count rate achieved is equivalent to that of 2.5×10^{13}, 5.0×10^{13} and 2.5×10^{14} at the nominal BTL operating temperature of $-30\,^{\circ}\text{C}$. The corresponding levels of dark currents are representative of the values expected for 500, 1000 and 5000 fb^{-1} integrated luminosities.

The time resolution obtained for a given laser signal as a function of DCR using the LED light injection for noise emulation was found to be consistent with the results achieved on irradiated SiPMs, as shown in Fig. B.4.

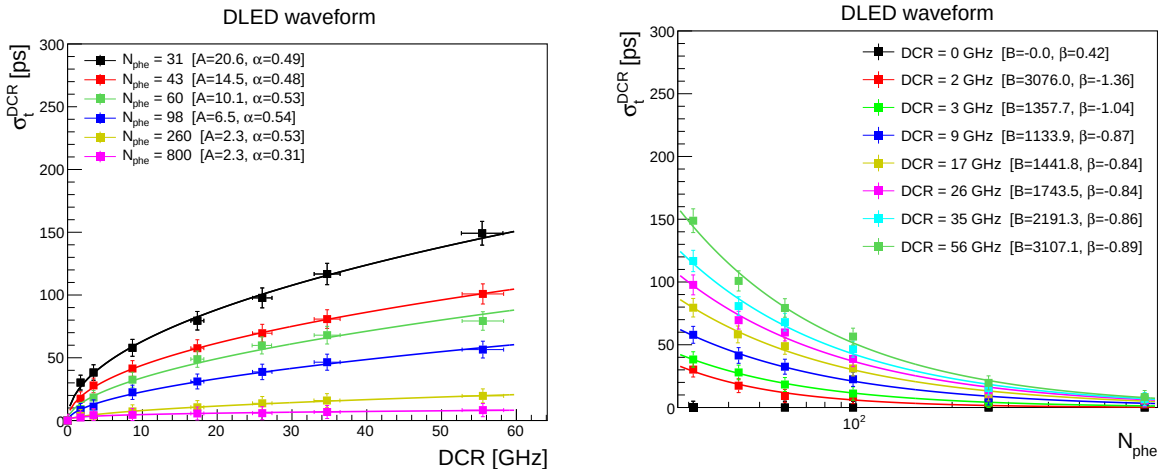


Figure B.3: Contribution of the DCR term to time resolution, using the DLED waveform shaping, for different intensities of laser pulse and different levels of LED-injected DCR.

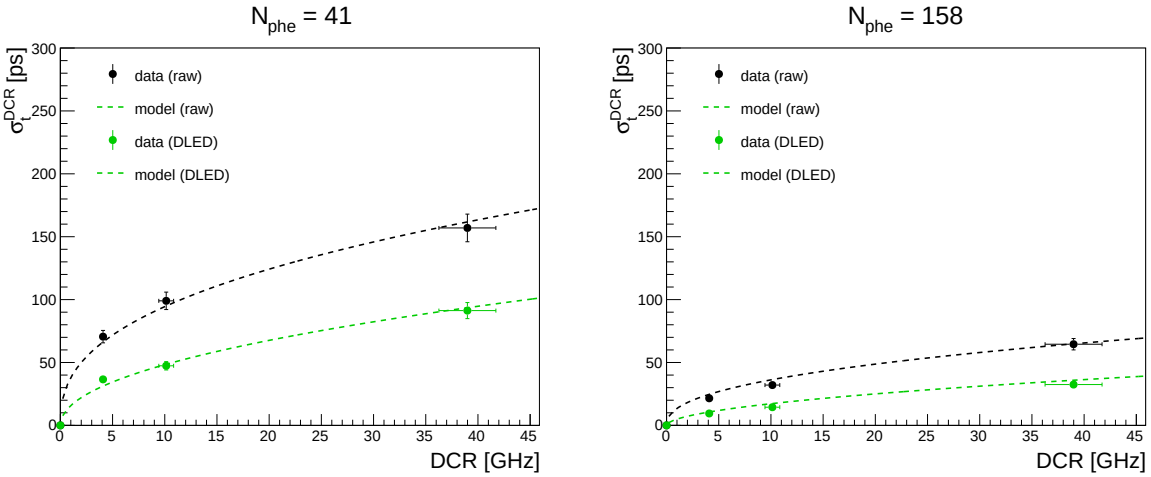


Figure B.4: Comparison of the DCR term contribution to time resolution in irradiated SiPMs (dot points) compared with the prediction of the parameterized model from the emulated DCR through LED light injection for two different values of the laser signal amplitude (41 and 158 photoelectrons).

B.I.3 Uncertainties on the BTL performance extrapolation

The prediction of the BTL performance evolution with radiation damage, discussed in Chapter 2, is the result of the simulation of the ASIC performance for a detailed set of input parameters obtained from comprehensive laboratory and test beam measurements. Using this model, we anticipate a time resolution of about 50–60 ps ps after an integrated luminosity of 3000 fb^{-1}. An uncertainty of about 30% is ascribed to this estimate, including uncertainties on the input parameters, on the radiation level predictions, as well as on the ASIC performance from post-layout simulations. These uncertainties are discussed below together with options that are being studied to further mitigate the impact of radiation damage in the SiPMs and improve the overall performance of the BTL.

B.I.3.1 Radiation levels in CMS

The uncertainty on the radiation levels predicted is discussed in the Appendix A. Up to 50% changes in the level of ionizing dose have negligible impact on the BTL performance since

all the optical components have been qualified to this safety margin showing minimal loss of transparency. An increase of 50% in the predicted 1 MeV neutron equivalent fluence corresponds to a proportionally higher DCR, which translates into a 25% worse time resolution at 3000 fb^{-1} as shown in Fig. 2.3.

B.I.3.2 Uncertainties on the crystal and SiPM properties

The radiation tolerance of the SiPMs, i.e. the amount of DCR expected at a given over-voltage, is affected by uncertainties on the available current measurements of irradiated SiPMs. An overall uncertainty of 20% is ascribed to the uncertain knowledge on the amount of fluence to which the SiPMs are exposed at the irradiation facilities and the effect of the spontaneous annealing occurring during irradiation and during the transport of the samples to the laboratory (always at $-30\,^{\circ}\text{C}$).

The amount of photoelectrons per unit of time, which depends mainly on the light yield and decay time of the LYSO:Ce scintillation and on the PDE of the SiPMs, is used in the simulation to model the signal shape and amplitude. The combination of these effects has a precision of about 10% (8% and 6% respectively).

B.I.3.3 ASIC contribution

The limited impact of the DCR to the time resolution relies on the method used for cancellation of the baseline fluctuations a discussed in Chapter 2. The implementation of this technique in the ASIC is discussed in Section 2.3 and results of simulation of the ASIC performance are shown. Nevertheless, an uncertainty on the simulated ASIC performance for DCR cancellation at the level of 20% is estimated from the variation of the post-layout simulation of the ASIC.

B.I.3.4 Options for further DCR mitigation

To add safety margin to the BTL performance at the end of operation by reducing the amount of DCR in the SiPM, two main options are under study.

Lower operating temperature The DCR has a strong dependence on the operating temperature, decreases by a factor 1.8 every $10\,^{\circ}\text{C}$. The nominal temperature at which the BTL SiPMs are expected to operate is assumed to be $-30\,^{\circ}\text{C}$. This estimate assumes a temperature of the CO_2 coolant at the BTL cooling loop inlet of $-35\,^{\circ}\text{C}$ and a gradient from the coolant to the sensors of $5\,^{\circ}\text{C}$. This estimate assumes, conservatively, that the CO_2 in the return line from the BTL to the cooling plant (operated at $-45\,^{\circ}\text{C}$) is in liquid phase thus providing the maximum static height effect and temperature drop from the cooling plant. For a negligible static height effect the inlet temperature is estimated to be $-41\,^{\circ}\text{C}$. A direct test is planned by CMS technical coordination to verify this temperature. In parallel, the ongoing optimization of the diameter of the pipes in the BTL distribution may reduce frictional losses providing another 1–2 $^{\circ}\text{C}$ of reduction of the inlet temperature. Finally, the ongoing optimization of the thermal coupling between the SiPM package and the cooling plate could provide a reduction of the temperature gradient of another 1–2 $^{\circ}\text{C}$. An operating temperature of $-35\,^{\circ}\text{C}$, which may be achieved, would result in a DCR reduction of about 33%. This would entirely offset the uncertainty on the performance extrapolation.

A more aggressive option would be to include small thermoelectric elements (Peltier) directly attached to the SiPM package or in proximity of the cooling fin. The SiPM boards design is compatible with the inclusion of thin Peltier elements under study, which can provide additional cooling power local to the SiPMs and enable stable operation of the SiPMs at $-40\,^{\circ}\text{C}$.

According to preliminary studies, the additional power required by the thermoelectric coolers during operation is offset by the reduction in power consumption of the SiPM due to the decrease of the DCR. The Peltier elements are currently under evaluation for radiation tolerance and longevity. If sufficient maturity of these studies, with positive results, is achieved before the date for the definition of the production ready module design, this option may be integrated in the final detector.

Enhanced annealing of SiPM damage The defects created by radiation within in silicon anneal spontaneously at a temperature dependent rate. The current annealing model for BTL assumes a 10–20% annealing of the damage during operation at $-30\,^{\circ}$C and an additional annealing of a factor 2.5 during the yearly shutdowns where the BTL temperature is raised to the room temperature of about $20\,^{\circ}$C. If the temperature of the SiPM can be raised to $50\,^{\circ}$C for short periods (few days), an additional annealing occurs, reducing the DCR by another factor 1.5 with respect to room temperature annealing.

The possibility to implement an annealing cycle at $50\,^{\circ}$C during shutdowns is being investigated. Robust heating foils located on the cooling fins close to the SiPM package can provide the heat needed to warm-up the SiPMs. According to preliminary estimates, the additional heat can be extracted by the CO_2 cooling system with the temperature of the CO_2 plant set at $15\,^{\circ}$C, if about one sixth of the BTL (one readout unit per tray) is warmed up at $50\,^{\circ}$C at a time, while the rest of the BTL system is kept at $20\,^{\circ}$C. Six cycles of several weeks, of more cycles of one week for a lower fraction of the detector, would fit in a year end shutdown.

B.I.4 Plans for full module radiation tolerance tests

1. Irradiation of crystal matrices with wrapping (Nov 2019).
   Irradiation of packaged crystal matrices from 8 vendors to the integrated BTL ionizing dose will be performed at ENEA Casaccia as discussed earlier. Light output before and after irradiation will be measured.

2. Test of induced background under gamma/neutron irradiation (Aug 2019-Jan 2020).
   The radiation induced noise (photons background) in LYSO:Ce crystal bars will be measured under the expected levels of gamma dose rates and neutron flux expected at HL-LHC to confirm prediction of simulation. The additional photon noise rate is expected to be negligible with respect to the dark count rate.

3. Test beam with irradiated BTL sensors (Feb-Mar 2020).
   Additional radiation tolerance tests on full BTL sensors are planned to further validate the performance extrapolations. A test beam of BTL sensors instrumented with SiPMs irradiated to a room temperature equivalent fluence of 2.5×10^{13}, 5.0×10^{13} and 2.5×10^{14} n.eq./cm^2. The operating temperature of the SiPMs will be varied from $20\,^{\circ}$C to about $-20\,^{\circ}$C to scan the entire range of DCR expected throughout the BTL operation at HL-LHC. The measurement will provide thus a direct measurement of the BTL performance evolution with MIPs due to increase of dark count rate and as a function of the operating temperature.

4. Test of irradiated SiPM arrays with TOFHIRv2 (Feb-June 2020).
   Several tens of customized BTL SiPM arrays (16 SiPMs/array) from both vendors (HPK and FBK) will be received in September 2019 and will be irradiated in October 2019 to different level of 1 MeV neutron equivalent fluences up to the one expected for BTL beyond the end of operation. Low level characterization (e.g. current, gain, etc.) of the irradiated SiPM arrays will be available around November 2019. For characterization of the timing

performance of the SiPM arrays after irradiation it will be necessary to use the second version of the TOFHIRv2 ASIC which includes the module for DCR cancellation. Such measurement will thus be carried out starting from February 2020.

5. Test of full BTL module with irradiated SiPMs (Dec 2020 - Feb 2021).
During 2020 a full read-out unit prototype will be assembled and tested in laboratory. A read-out unit instrumented with BTL modules (crystal matrices and SiPM arrays irradiated to different level of fluences) will be prepared by the end of 2020 for a test with beam.

B.II BTL test beam campaign for sensor optimization

An extensive campaign of beam tests, as listed in Table B.2, has been carried out in 2017 and 2018 to define the optimal BTL sensor layout in terms of crystal and SiPM geometry, optical coupling, and wrapping material.

Date	Facility	Results
2018, November	FNAL MTEST	Study of performance at different angles; layout choice
2018, September	CERN SPS H4	Optimization and comparison of different BTL layouts
2018, June	FNAL MTEST	Test of different tile and bar geometries
2018, May	CERN PS T10	First test of crystal bar layout
2017, December	FNAL MTEST	Comparison of various crystal and SiPM sizes
2017, September	CERN SPS H6	Test of 16 ch. crystal matrix module with TOFPET2
2017, June	CERN SPS H4	Test of different SiPM manufacturers (Ketek, FBK, HPK)
2017, May	CERN PS T9	First test of BTL sensors with TOFPET2 ASIC

Table B.2: Test beam campaign in 2017 and 2018 devoted to BTL sensor optimization.

During the test beam campaigns the BTL sensors were read out using a combination of customized electronics to form fast pulses appropriate for high resolution time stamping and DRS digitizer boards (CAEN V1742) to acquire the signal waveforms. Two types of electronics were used, which ultimately yielded comparable results. The first option employs the NINO chip to provide a discriminated signal resulting from a highly amplified signal pulse on one line used for precise time stamping and a less amplified raw pulse shape on a second line which is used to derive amplitude-dependent corrections for the time stamp as described in Ref. [23]. The second option similarly treats the SiPM signal by splitting it into two lines with different amplification factors and both analogue waveforms are read out without applying time discrimination on the chip. On one line, the signal is amplified by an external amplifier (Hamamatsu C5594) of gain 63, while the second line is either attenuated or kept without additional amplification in order to keep the signal pulse below the saturation point of the DRS digitizer. Pictures of typical BTL sensors characterized during the test-beam campaign are shown in Fig. B.5.

The test beam facilities maintain particle tracking detectors, either wire- or fiber-based hodoscopes or silicon-based telescopes. The data from the tracking detectors are synchronized with the DRS digitizer data by integrated DAQ systems that maintain common trigger signals. The tracking information provided by such detectors enables the characterization of the performance of the BTL sensors as a function of the impact point and measurement of their response uniformity. A Micro Channel Plate (MCP) with a large diameter of about 2.5 cm and time resolution of 16 ps was placed behind the sensors under test and used as a reference time stamp, t_0. Its contribution to the time resolution values reported in the following is subtracted in quadrature. The contribution to the measured time resolution from electronic noise and calibration of

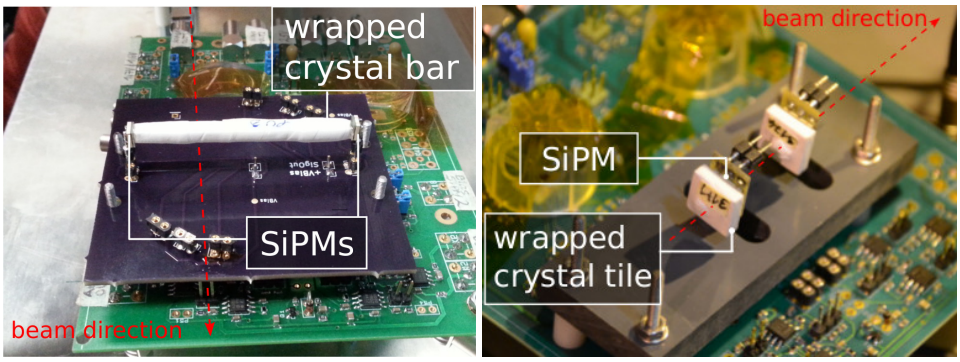


Figure B.5: Pictures of typical BTL sensors characterized during the test beam campaign: a crystal bar with one SiPM at each end (left) and two crystal tiles with the respective SiPMs glued to the rear face (right).

the digitizer has been measured to be below 10 ps and it is not subtracted from the following results.

A summary of the test beam results showing the performance of the reference BTL sensors was reported in Section 2.1.1. Additional results on studies performed on different sensor layouts and crystal geometries, as part of the R&D program that led to the sensor optimization, are reported in the following.

B.II.1 Alternative BTL sensor layouts tested

Two main crystal geometries have been evaluated during the BTL sensor R&D leading to the choice of elongated crystal bars as the optimal sensor for the reasons discussed in Section 2.1.1 of this TDR. The alternative BTL sensor layout considered consists of a crystal tile with surface of about 8×8 mm^2 and variable thicknesses (3.75, 3.0, 2.4 mm). A single SiPM of 3×3 mm^2 coupled to the center of the rear face of the crystal tile is used for reading out the scintillation light. A relevant drawback of such layout is a non-uniformity of the sensor time response across the crystal surface from the propagation time of the optical photons when the MIP impacts far away from the SiPM. This effect has been measured in the beam and is shown in Fig. B.6. The top panel shows a map of the time response variation as a function of the impact point of the track on the crystal syrface, as measured by the beam hodoscopes; the bottom panels show the time response variation as a function of the x (left) and y coordinate (right), with an average on the other coordinate. This non-uniformity effect can be corrected using information provided by the CMS Tracker, but requires a precision of the order of 1 mm. The precision of the Tracker in the ϕ direction is expected to meet the 1 mm resolution requirement, however the precision of in the z direction meets the requirement only for charged particles with p_{T} larger than about 2 GeV. For charged particles with p_{T} below 2 GeV, of which there are many from pileup, the non-uniformity would add in quadrature an additional contribution to the time resolution of about 25 ps.

The results on the time resolution achieved on 3 and 4 mm thick tiles at normal incidence are reported in Fig. B.7. The same BTL SiPMs of the type S12572 from Hamamatsu were used so that results are directly comparable with those presented in Section 2.1.1. An average time resolution of about 42 ps is measured for a 4 mm thick tile and about 49 ps for a 3 mm thick tile after position correction is applied. The difference in resolution between two different thicknesses scales as expected with the inverse of the square root of the energy deposit ($\sqrt{4\,\mathrm{mm}/3\,\mathrm{mm}} \sim 49\ \mathrm{ps}/42\ \mathrm{ps}$). This has been confirmed with a dedicated test comparing different crystal thicknesses of 2, 3 and 4 mm for the same crystal cross section and SiPM area

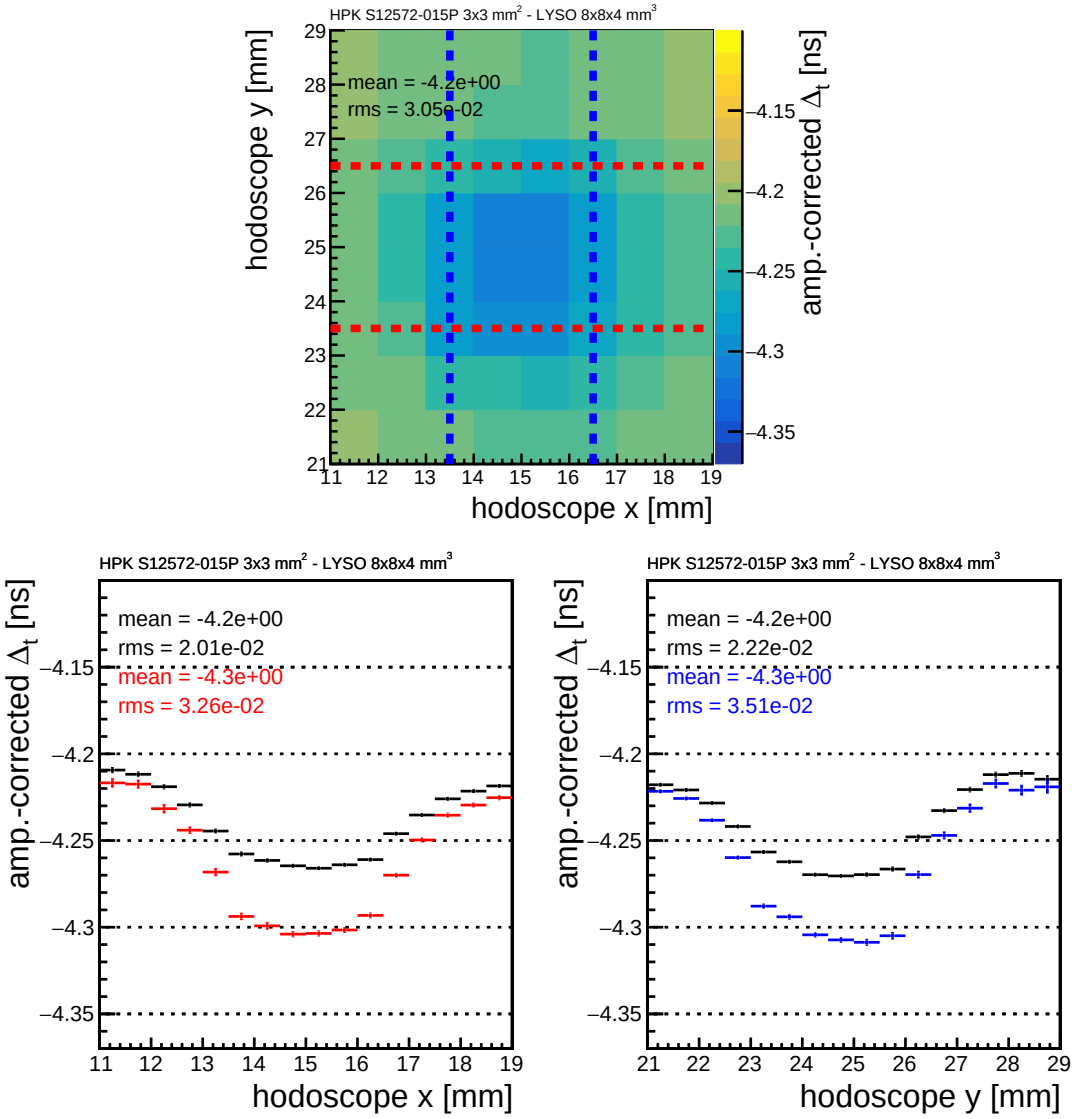


Figure B.6: Non-uniformity of time response as a function of MIP impact parameter relative to the center of the SiPM from the propagation delay of light in the tile, for an 8×8 m^2 by 4 mm thick tile. The top panel shows a map of the time response variation as a function of the impact point of the track on the crystal syrface; the bottom panels show the time response variation as a function of the x (left) and the y coordinate (right). The black histograms show the effect when the impact point is everywhere on the crystal surface in the other view, while the colored histograms show the effect for an impact point trhough the central region in the other view, indicated by dashed lines in the top panel.

and is shown in Fig. B.8. Similarly, the $1/\sqrt{\text{PDE}}$ behavior of the time resolution with the SiPM PDE has been demonstrated as summarized in Fig. B.8.

The light output from a "tile-type" sensor layout depends on the fraction of crystal surface covered by the SiPM. The larger the SiPM the better light collection efficiency until the SiPM area entirely matches the crystal surface: i.e. a ratio of SiPM area over crystal area equal to 1. It can be noted that the TP layout used 4×4 mm^2 SiPMs on a 11.5×11.5 mm^2 crystal tile. In that case, the same amount of light was extracted as for the 3×3 mm^2 SiPMs on a 8×8 mm^2 crystal tile. Several crystal and SiPM geometries have been studied throughout an extensive test beam

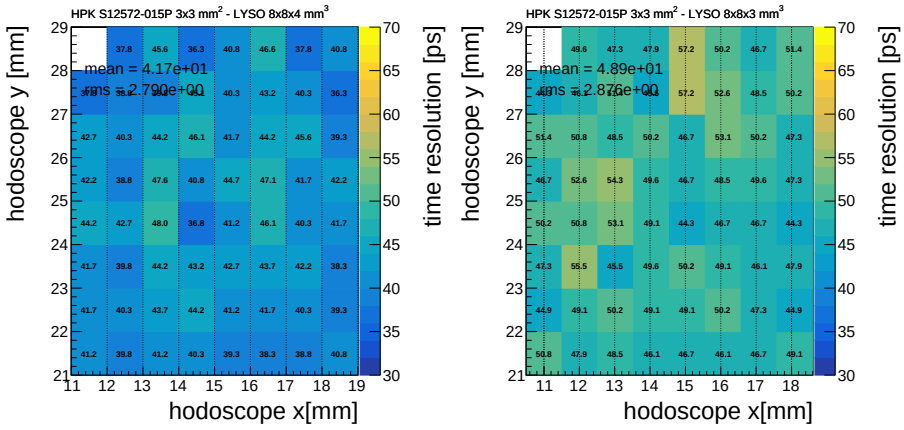


Figure B.7: Amplitude walk-corrected time resolution for the tile sensor, averaged within each (x, y) bin of size 1×1 mm^2, as a function of the MIP impact position.

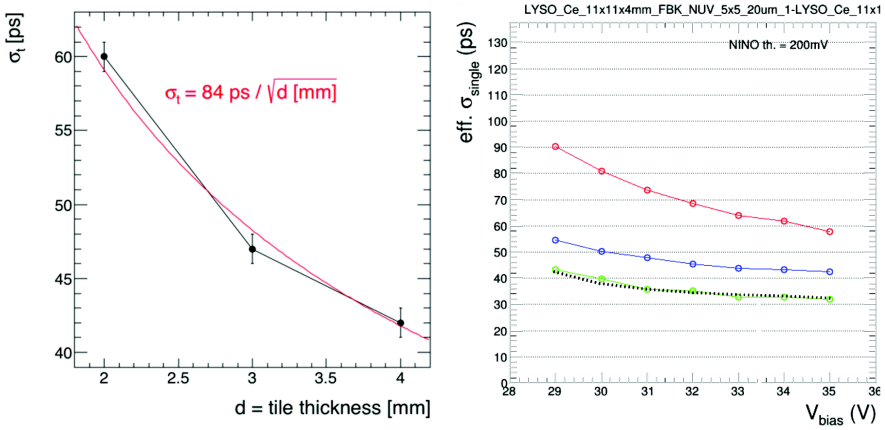


Figure B.8: Time resolution of BTL sensors as a function of crystal thickness (left) and as a function of SiPM bias (right). The black dotted curve superimposed to the measurement (green) shows the expected resolution behavior following the SiPM PDE increase with over-voltage and confirms the scaling of time resolution as $1/\sqrt{\text{PDE}}$.

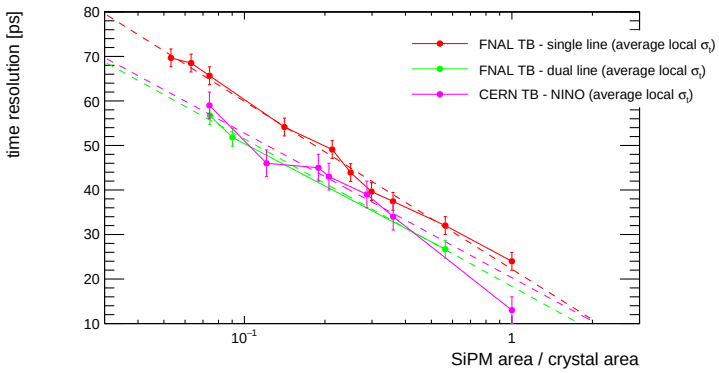


Figure B.9: Time resolution of tile sensors as a function of the aspect ratio, i.e. the ratio of the SiPM sensitive area to the surface area of the tile. The results obtained using the NINO board electronics and an alternate FNAL/Caltech readout electronics show consistent behavior.

campaign allowing us to parameterize the time resolution of tile sensors as a function of their aspect ratio. A summary of the results is shown in Fig. B.9.

B.II.2 Comparison of BTL sensor layouts

The time resolution for the 3 mm thick tile, which is about 43 ps, can be compared with the resolution for a crystal bar of the same thickness read out by a single SiPM at one end. The two values are similar, confirming laboratory measurements showing that the light output (and thus the light collection efficiency) from the tile is about the same as that observed from one end of the bar. It follows that the use of two SiPMs, one for each end, for the crystal readout in the bar layout improves the overall time resolution by a $\sqrt{2}$ for all sources of time jitter that are uncorrelated at the two SiPMs, mainly the photostatistics and DCR terms. Based on these results the performance of the tile and bar sensors has been estimated throughout BTL lifetime and has been compared with the initial detector layout proposed in the Technical Proposal (TP) [8]. The comparison is shown in Fig. B.10 and clearly demonstrates the importance of a reduced SiPM area, as well as the advantage of the redundancy in timing measurement per crystal to improve the overall resolution.

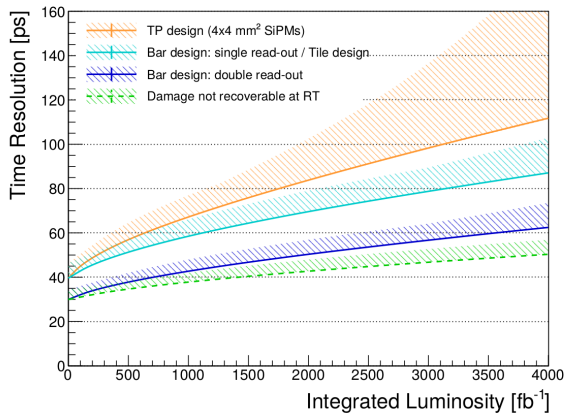


Figure B.10: Expected time resolution as a function of integrated luminosity for the TP reference design (orange); the TDR bar design from single-ended and tile design (light blue); the TDR reference design with crystal bars combining double-ended readout (blue); and the TDR reference design assuming all damage recoverable at room temperature (RT) is annealed. The bands show the expected performance for different sets of SiPM parameters.

Appendix C

The Endcap Timing Layer: additional technical information

C.I Beam tests

In order to validate the overall design of the ETL, a set of beam test campaigns were carried out during 2017 and 2018, in which prototype LGAD sensors were exposed to proton beams at Fermilab Test Beam Facility (FTBF). Studies were performed using sensors irradiated up to the expected end-of-life neutron fluences, with prototype sensors from HPK, FBK, and CNM. The readout of these sensors was performed using dedicated readout boards, with discrete amplifiers, designed specifically for tests of fast LGAD signals. The measurements of arrival time were extracted from signal waveforms by analysing the signals acquired by a fast digitizer or an oscilloscope.

The experimental setup used in these campaigns is described in Section C.I.1; readout electronics used in the measurements are described in Section C.I.2.

C.I.1 Experimental Setup

Test beam measurements were performed at the Fermilab test-beam facility (FTBF) which provides a 120 GeV proton beam from the Fermilab Main Injector accelerator. The devices under test (DUTs) are mounted on a remotely operated motorized stage placed inside a cold box. Tracking of charged particles is provided by the precision telescope detector [133] composed of 7 planes of silicon strip detectors. This provides better than 10 μm position resolution for charged particles impinging on the DUT. A Photek 240 micro-channel plate (MCP-PMT) detector was placed furthest downstream, and provided a very precise reference timestamp. Its precision has been previously measured to be better than 7 ps [134]. A photograph of the experimental area is shown in Fig. C.1.

Two DAQ systems have been used to record signal waveforms. In the first system, signals from the DUTs and the Photek MCP-PMT are recorded using a CAEN V1742 digitizer board [135], which provides digitized waveforms sampled at 5 GS/s with a 12-bit ADC, and with one ADC count corresponding to 0.25 mV. In the second system, the signals are digitized with a Tektronix DPO7254 oscilloscope, which provides signals sampled at up to 40 GS/s, with an analog bandwitdh of 2.5 GHz. The DAQ for the pixel telescope is based on the CAPTAN system developed at Fermilab [133]. The track-reconstruction is performed using the Monicelli software package developed specifically for the test beam application [133]. The advantage of the DRS-based system is its high channel count, which allows one to simultaneously record signals from up to 32 channels, while the Tektronix scope achieves a much higher analog bandwidth and sampling rate. Both DAQ systems have been used, depending on the scope of the studies. Precision measurements of the time resolution are typically obtained using the data recorded with the

oscilloscope, while for studies of sensor uniformity the DRS-based DAQ is used.

The DUTs were placed inside the cold box, and mounted on aluminum plates. Cooling loops are routed inside aluminum plates, providing cooling down to $-25\,^{\circ}\mathrm{C}$. Up to 5 DUTs can be placed inside the mounting frame, and the frame is attached to a remotely operable moving stage, allowing one to move the DUTs both vertically and horizontally, in order to align the DUTs with the beam.

The beam is resonantly extracted in a slow spill for each Main Injector cycle delivering a single 4.2 s long spill per minute. The primary beam (bunched at 53 MHz) consists of 120 GeV protons. All measurements presented in this TDR were taken with the primary beam protons, with 50 000 protons per spill. The trigger to both the CAEN V1742 and to the pixel telescope was provided by a scintillator mounted on a photomultiplier tube, placed upstream of the DUTs in the beam line.

C.I.2 Readout Electronics

Three readout electronics boards were used in various measurements performed at the FTBF campaigns, each with different characteristics. They were independently developed at Fermi National Accelerator Laboratory (FNAL), at the University of Kansas (KU), and at the University of California Santa Cruz (UCSC).

The 4-channel Fermilab LGAD test board is designed to test sensors up to 8.5 mm by 8.5 mm at voltages up to 1 kV. Four wire-bonding pads allow for signal readout via amplifiers based on Mini-Circuits GALI-66+. The amplifiers feature transformers with 1:2 input impedance matching, two stages of amplification and a 500 MHz low-pass filter. In this full configuration, the amplifiers feature 12.5 Ω input impedance, 5 kΩ transimpedance, 500 MHz bandwidth and 1 mV rms output noise. If needed it is possible to short circuit the input transformer and/or the low-pass filter, which would result in an input impedance of 50 Ω, transimpedance of 10 kΩ, and bandwidth of 2 GHz. A larger version of this board capable of reading out up to 16 channels has also been designed and used in test beams, and some results are shown in Section 3.2. A version of this board used in the Dec 2018 test beam campaign with a 2×8 pad FBK sensor is shown in Fig. C.2.

The 2-channel KU board, designed and produced by the University of Kansas, can accommodate many types of sensors including diamond, silicon, LGAD or avalanche photodiodes (APD). The sensor is hosted on the board itself and the electronics was optimized for precise timing measurements. In particular, the amplifier, made with discrete components, has an input impedance of 700 Ω, an output noise of 4 mV and a gain in transresistance of about

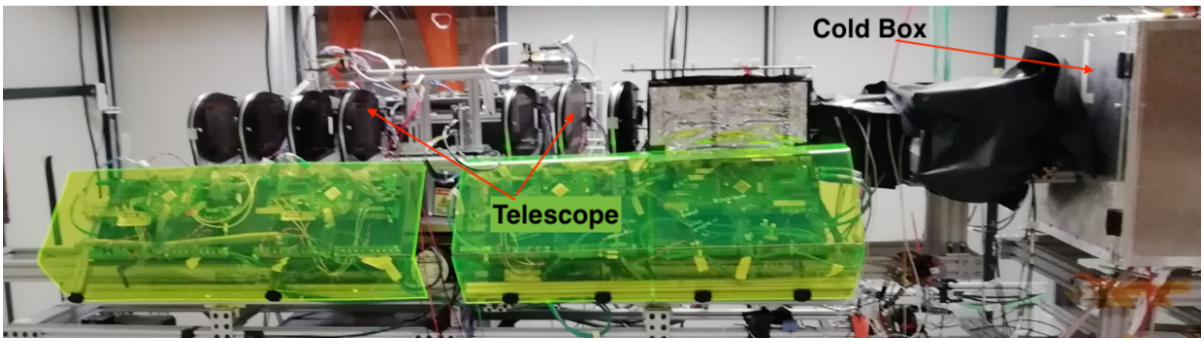


Figure C.1: Photograph of the experimental area, showing the tracking telescope and the cold box.

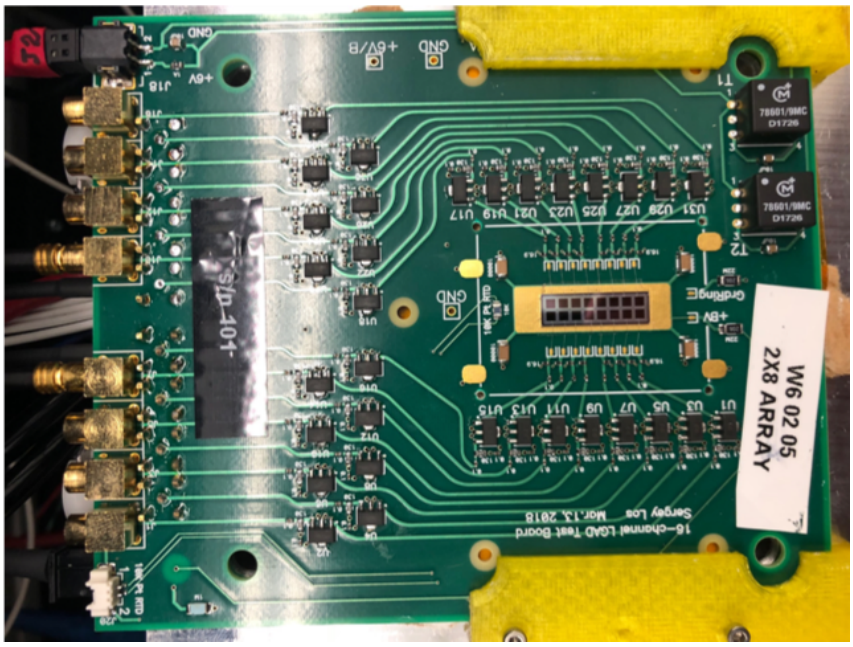


Figure C.2: A photograph of the 16-channel LGAD readout board, with a 2x8 pad FBK sensor mounted.

50 mV/μA with a 3 dB bandwidth of 100 MHz. Those values were simulated for an input capacitance of 20 pF, which corresponds roughly to an LGAD of 9 mm^2. The power consumption of the board is about 130 mW per channel.

The UCSC 1-channel board is described in detail in Ref. [66]. This board uses discrete components and contains several features which provide a wide bandwidth ($\sim$ 2 GHz) and a low noise even in noisy environments. The inverting amplifier uses a high-speed SiGe transistor which has a transimpedance of about 470 Ω. A commercial inverting amplifier with a gain of 10 is used to boost the signal. The 4-channel UCSC board has two stages: the first one is identical to the UCSC single channel board, and is followed by an inverting stage. The total transimpedance is 10.7 kΩ.

C.II Waveform Sampling Implementation Details

Figure C.3 depicts the proposed 12-bit pipeline SAR structure. The pipeline SAR consists of two moderate-resolution SAR ADCs operating in pipeline, with the first stage producing the first 6 most significant bits and the second producing the 7 least significant bits with redundancy for inherent error correction [100]. The two-staged pipeline SAR structure enables us to nearly double the speed compared to a single stage SAR. Resolving multiple bits in the first stage also relaxes the capacitive DAC (CDAC) matching requirement of the second stage. A residue amplifier between the two stages amplifies the residue from the first stage before feeding it to the second stage. With the gain from the residue amplifier, the stringent noise constraint on the second-stage comparator is relaxed for high-resolution applications, in contrast with a conventional SAR that requires the comparator to be extremely low noise for at least a few least significant bits. Finally, the comparator results are saved in the output register and a data retiming block synchronizes the digital data from the first stage and the second stage. As shown in Fig. C.3, the 13-bit raw data is converted into 12-bit output by a 13b-to-12b mapping block.

Asynchronous SARs will be adopted for the two stages of the ADC. The start of the comparison

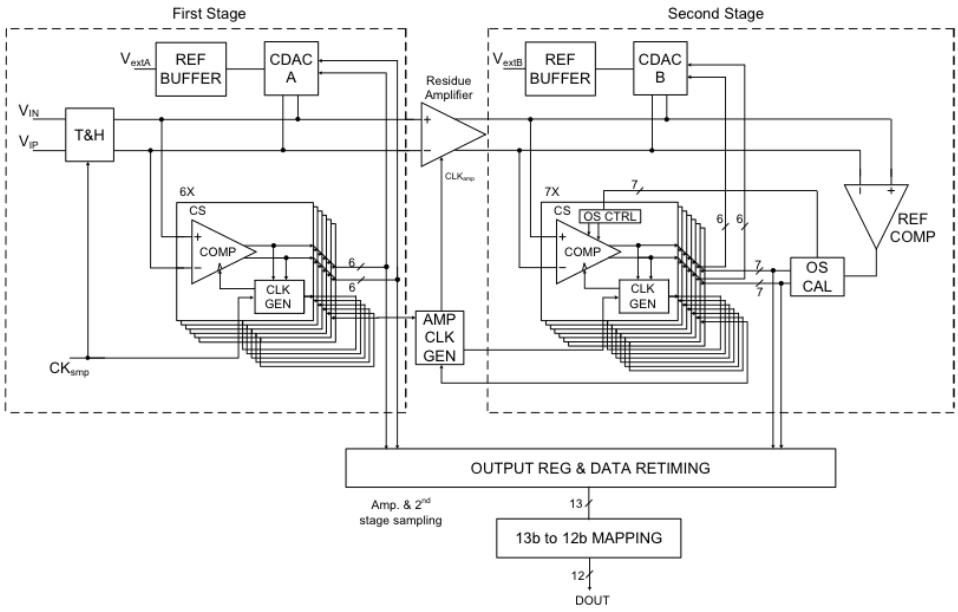


Figure C.3: Block diagram of the proposed pipeline SAR ADC.

of each bit is directly triggered by the detection of the completion of the comparison of the previous bit. This design choice not only increases the conversion speed, but also removes the need for a high-speed internal clock, thus reducing the power consumption.

Multiple comparators will be employed to further increase the conversion speed in a loop-unrolled fashion. The offset between different comparators will be calibrated in the background to track process-voltage-temperature (PVT) variations. A background comparator calibration method which does not compromise the SAR conversion speed has been developed and validated. The offsets between different comparators in the first stage are tolerated by the stage redundancy. The comparator offsets in the second stage are calibrated using a reference comparator as shown in Fig. C.3 (denoted as REF_COMP and OS_CAL block). The advantage of offset calibration using a reference comparator is that this scheme does not compromise the conversion speed of the ADC, compared to an SAR ADC design where a dedicated calibration cycle is required.

Figure C.4 illustrates the timing diagram. The ADC conversion starts with the sampling phase of the first stage SAR, which makes 6-bit conversions sequentially. Upon completion of the first SAR stage, it triggers the clock of the residue amplifier, which amplifies the residue voltage and sends it to the second stage SAR for further processing. After the residue amplification, the first stage SAR resets its DAC and begins to sample the next input voltage while the second stage SAR resolves the rest of the bits. Since there are 13 conversions for the 12-bit ADCs, the extra bit is used to implement redundancy, so that any potential decision errors caused by the comparator noise and offsets in the first stage SAR can be tolerated. The offsets between different comparators in the second stage are calibrated using a reference comparator, which is activated at the same time as each of the comparators in the second stage in specific cycles as shown in the clock waveforms in Fig. C.4. This offset calibration operates in the background, so that it does not affect the timing margin of each SAR stage.

A 380 MS/s 12-bit ADC (10-bit ENOB), a critical building block for the waveform sampler, has been developed in the 65 nm CMOS process as shown in Fig. C.5. The ADC chip performance has been measured. Because of a routing error in the connection to one of the pads, only the first 6-bit stage of the pipelined SAR ADC could be characterized. Figure C.6 (left) summa-

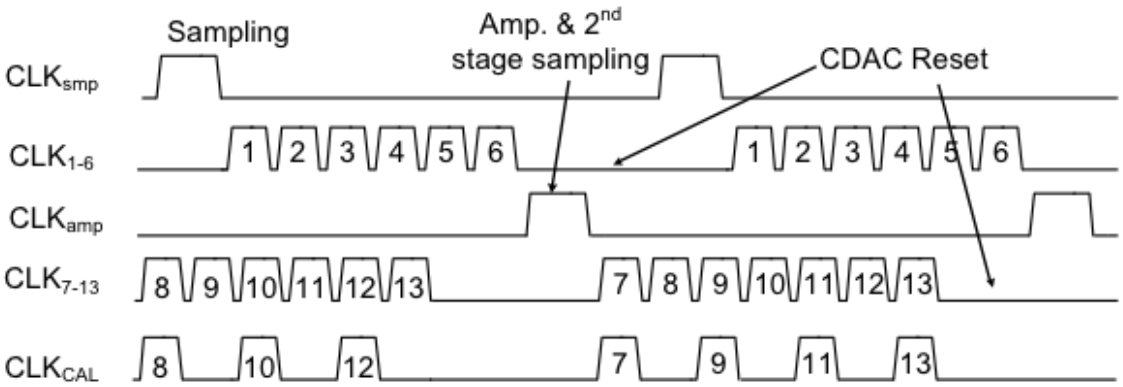


Figure C.4: Timing diagram of the proposed pipeline SAR ADC.

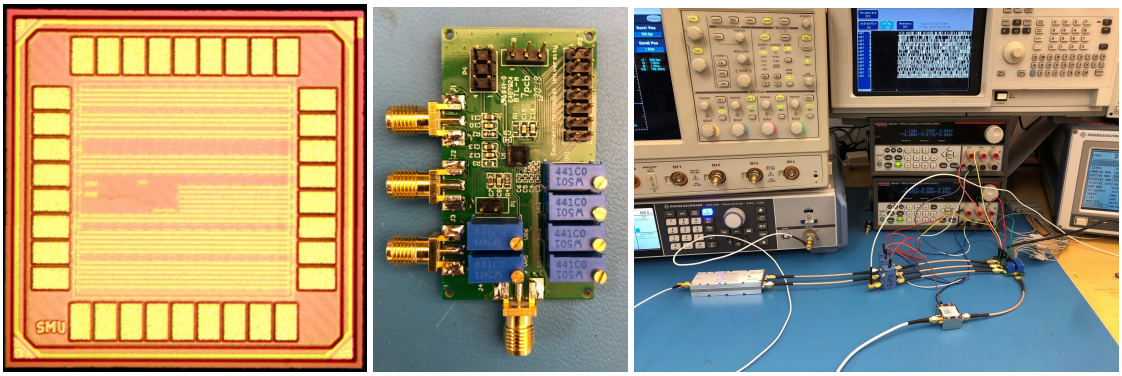


Figure C.5: The ADC chip, test board, and test setup.

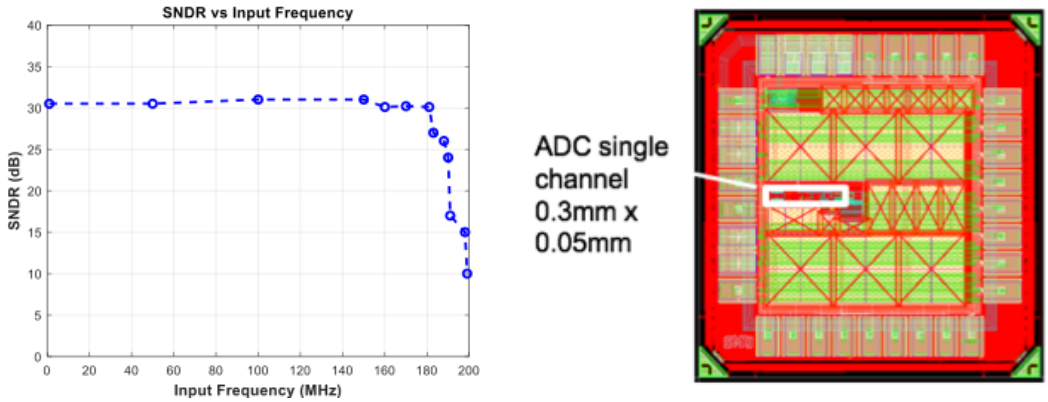


Figure C.6: (Left) The measured SNDR vs. input frequency. (Right) The layout of the revised ADC IC.

rizes the measured SNDR (Signal-to-Noise plus Distortion Ratio) vs. input frequency, while the sampling frequency is 400 MHz. These measurement results agree well with the simulation. A revised version of the ADC chip, shown in Fig. C.6 (right), has been completed and is currently in fabrication. A single-channel ADC only occupies an area of $300 \times 50\ \mu m^2$.

To achieve a sampling rate above 2 GS/s, multiple single-channel ADCs will be interleaved together in a serially-sampling and parallel-processing manner. To achieve both a high resolution and a high sampling rate, a two-stage sample-and-hold (S/H) structure is proposed as depicted in Fig. C.7. The input signal from the sensor/preamplifier will be sampled by two 1.28 GHz differential clock signals, achieving an equivalent aggregate sampling rate of 2.56 GS/s. The

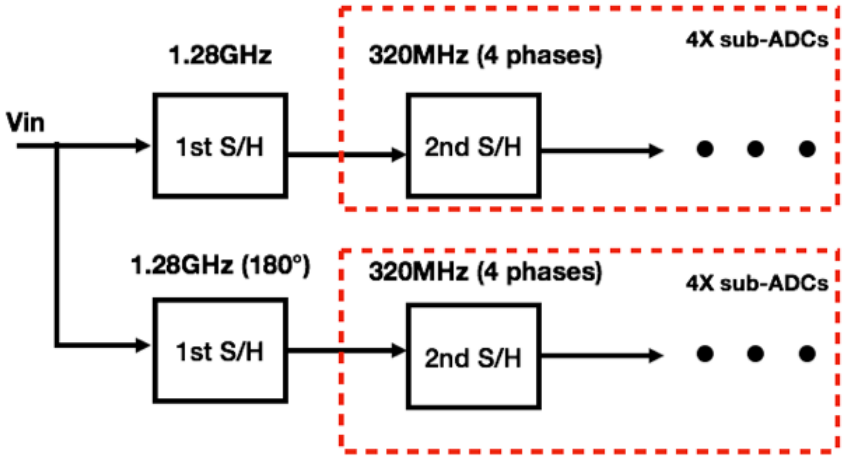


Figure C.7: The proposed time-interleaved structure for the waveform sampler.

sampled data from each of the first stage S/H is then held and fed to the following second stage, which consists of four 320 MS/s ADCs controlled by a 320 MHz clock with 90° phase shift. The advantage of using the proposed two-stage structure is that the input to the second stage is a DC signal, alleviating the requirements on the linearity of the second stage while allowing the first stage to achieve a high sampling rate. The choice of a 320 MHz clock for the second stage is based on the frequency of the clock distribution (320 MHz), but can be easily generated locally if a 40 MHz clock is distributed as a global clock instead. The estimated total area of the interleaved ADC is 0.5×0.5 mm^2.

Appendix D

Additional performance studies

D.I Impact of MTD material budget on the ECAL performance

In this appendix, a discussion of the impact of the MTD material budget on the energy reconstruction and resolution of electromagnetic showers in the CMS ECAL barrel is reported, extending the study already performed in [8].

The details of the description of the MTD geometry as implemented for the GEANT simulation were already reported in Section 5.2. The total amount of material, estimated both in terms of radiation and interaction length is presented in Fig. D.1. The material is reported as a function of the pseudorapidity, both for the BTL ($|\eta| < 1.5$) and the ETL ($|\eta| > 1.5$); contributions from sensitive material, support/cooling and electronics/services are shown separately. As expected the largest contributions come from the BTL LYSO crystals (up to about 0.4 X_0), accounting for about 90% of the total contribution in terms of radiation length in the barrel. The ETL adds just a small contribution in front of the endcap calorimeter, smaller than 0.2 X_0 and mostly due to the support aluminum plates.

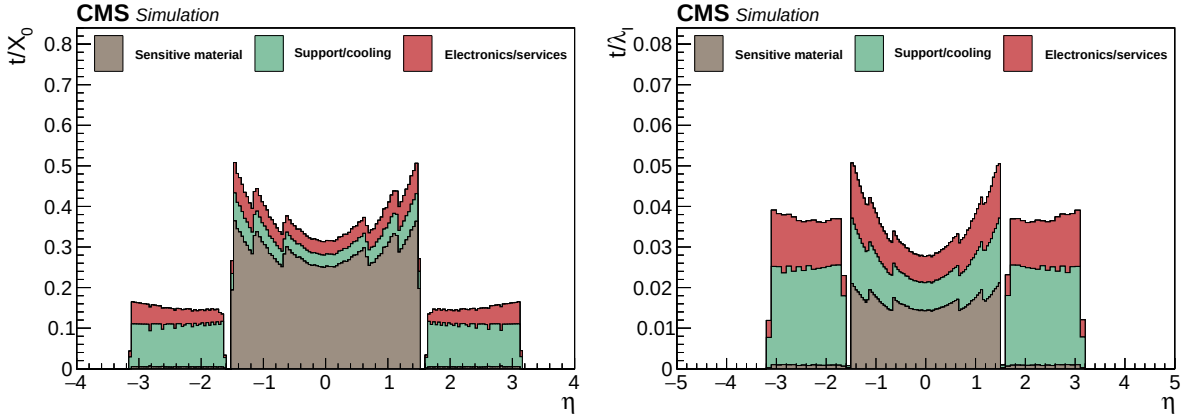


Figure D.1: MTD material budget as implemented in the GEANT simulation of the MTD as a function of the pseudo-rapidity. (left) Material budget in terms of radiation length X_0, (right) interaction length λ_I.

For the study of the impact of the BTL material budget on the ECAL barrel, a simplified description of the BTL is used consisting of 4 mm thick LYSO:Ce tiles placed at the BTL radius (0.35 X_0 at $\eta = 0$ and about 0.8 X_0 at $\eta = 1.5$). This approximation overestimates the total amount of material in the high $|\eta|$ region in the barrel, as the crystal slant thickness is levelled in the BTL to about 4mm and as shown in Fig. D.1 the contribution from dead material in BTL adds about 10% to the total BTL material budget.

The impact of the barrel timing layer on the ECAL energy resolution and shower shapes has

been studied for both photons and electrons using events from $H \to \gamma\gamma$ and $Z \to e^+e^-$ samples respectively. As it was already shown in Ref. [8], the energy resolution using the raw ECAL raw energy sum of crystals show no noticeable effects for photons and a modest increase (up to 1.3% in quadrature) for electrons at $|\eta| > 0.9$. Also no impact was observed on the shower shapes relevant for electron and photon identification. To further detail the impact of MTD on ECAL, photons which starts showering in BTL were compared to photons reaching ECAL unconverted as shown in Fig. D.2. Both in terms of resolution of the reconstructed photon energy and shower shapes, the distributions are very similar and no degradation of the energy resolution or significant differences in the shower shapes are observed (as an example R_9, the ratio between the energy contained in a 3×3 matrix around the most energetic crystal and the full ECAL cluster, is reported). Fig. D.2 shows also for comparison the distribution for photons which start converting in the tracker volume before reaching MTD.

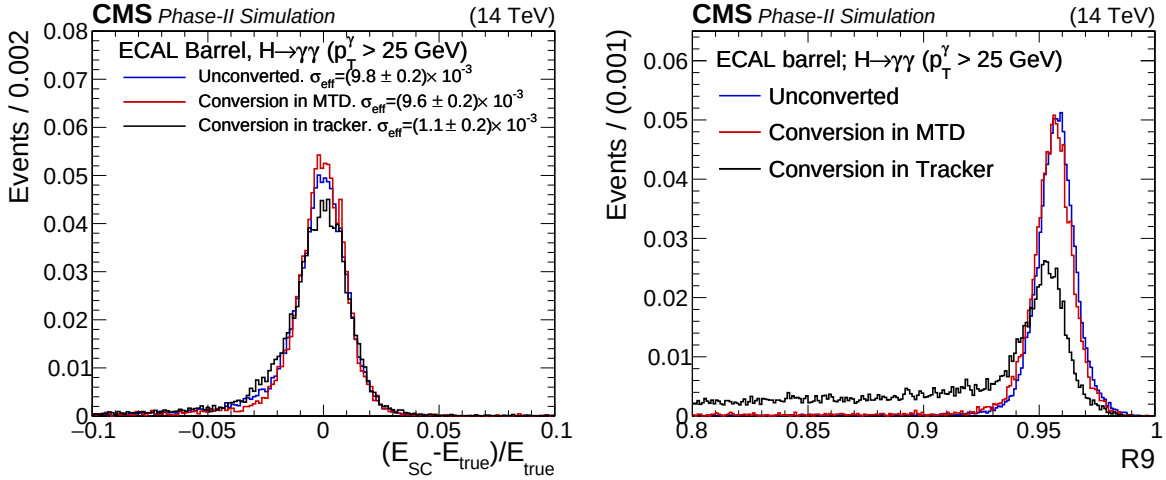


Figure D.2: Comparison of the energy resolutions and shower shapes for different categories of photons from $H \to \gamma\gamma$ decay in the ECAL barrel acceptance. Photons are separated into: unconverted photons (blue), conversion starting in the MTD volume (red) and conversion before MTD in the tracker volume (black). The effective σ (shortest interval containing the 68% of the distribution) is reported for the reconstructed cluster energy resolution (left). The right plot shows the comparison of one of the most important shower shape variable used in the classification and identification of the photon showers, R_9, the ratio between the energy in a 3×3 matrix of ECAL crystals around the most energetic one and the full ECAL cluster energy.

D.II Alignment, time synchronization, and monitoring

The essential steps in the time reconstruction are the measurement of the time at which a track or a neutral deposit crosses an MTD sensor (Section 5.2.4), the association of the time measurement with a track, and the estimate of the time at the vertex, after making a time-of-flight correction (Section 5.3.1). A consistent reconstruction of the track time at the vertex requires that the different MTD sensors be synchronized with a precision of a few picoseconds and that the MTD be spatially aligned to the Tracker. Tracks extrapolated to the MTD will enable the alignment to the Tracker and the time synchronization across the MTD with sufficient precision, as we describe below.

In addition, specific calibration constants associated with the time walk correction, and, for BTL, the time propagation inside the sensors are needed. These constants can be derived from data in a way identical to what was discussed for test beam data in Section 2.1.1.1.

D.III Spatial alignment

For the purpose of associating a track to a hit in the timing detector, the spatial alignment is not critical. The average hit density is below one track per 10 cm^2, while the resolution of the track extrapolated to the MTD is of the order of 1 mm or better depending on the p_T. For the purpose of the time-of-flight correction, the precision is dominated by the tracking precision (Section 5.3.1). Hence, also in this case, the requirement on the MTD alignment is not stringent. The spatial alignment will be based on the match between tracks extrapolated to the MTD and the hits in the MTD, similarly to the alignment technique adopted in the Tracker. By combining the information from many tracks, the precision of this method is much better than the spatial resolution of the MTD sensors and is not a limiting factor. This procedure can be repeated frequently, to monitor systematic movements relative to the Tracker.

D.IV Time synchronisation

To exploit timing in the reconstruction of the charged tracks, the different MTD channels will have to be synchronised to a precision of a few picoseconds. A precision of the same order will be needed in the relative synchronization of the MTD and the calorimeters to exploit timing in track-to-neutral matching. The absolute time calibration (or phase shifts relative to the beam clock) is not a particular concern, as all the event reconstruction relies on the relative time between tracks within the same collision event.

The time offsets of the MTD channels can be inter-calibrated using all the tracks collected by the CMS high level trigger. The distribution of the reconstructed time at the vertex — i.e. after TOF correction — of these tracks has an rms spread of approximately 200 ps, primarily determined by the time spread of the luminous region, since the additional spread from the sensor resolution is marginal compared to the beam spot time spread. The mean time of this distribution over many events provides the reference calibration points. To illustrate the method, we show results from a full simulation study for the BTL. Similar conclusions, appropriately scaled for the occupancy and channel count, apply to the ETL as well.

Examples of distributions of the reconstructed time at the vertex are shown in Fig. D.3 (left) for crystals at $|\eta| < 0.65$, $0.65 < |\eta| < 1.13$ and $1.13 < |\eta| < 1.48$, corresponding to the first, second, and last pair of readout units along a BTL tray. For tracks crossing multiple crystals, the time measurement from the crystal with the largest energy deposit is used. The distributions are normalized to the same luminosity. The variation in the number of tracks per crystal with η stems from the non-projective geometry of the BTL. The TOF correction assumes the pion mass hypothesis for all tracks. The right-hand tail in the distributions is due to slow particles, such as kaons and protons of low p_T, or to tracks associated with a wrong vertex, resulting in a wrong TOF correction. A large fraction of these tails can be filtered with a tighter p_T selection. However, for the purpose of time calibration and prompt monitoring of time drifts, these effects do not need to be corrected. As illustrated in Fig. D.3 (right), the median of the distribution provides a reference calibration point with an uncertainty of about 9 ps for 1000 tracks and of less than 3 ps for 10 000 tracks, where the uncertainty is the spread between the injected time offset and the estimated time offset from the truncated mean of the distribution.

At an allocation of 1 kHz of the high-level trigger rate and an average occupancy of 5% for 200 pileup events (BTL), the required 1000 or 10 000 tracks will be collected in around 20 or 200 s in each channel, thereby providing the possibility of frequent and granular calibrations. These calibration constants can be made available for the prompt reconstruction of the events, which in the current CMS operation starts within 24 hours of the data-taking.

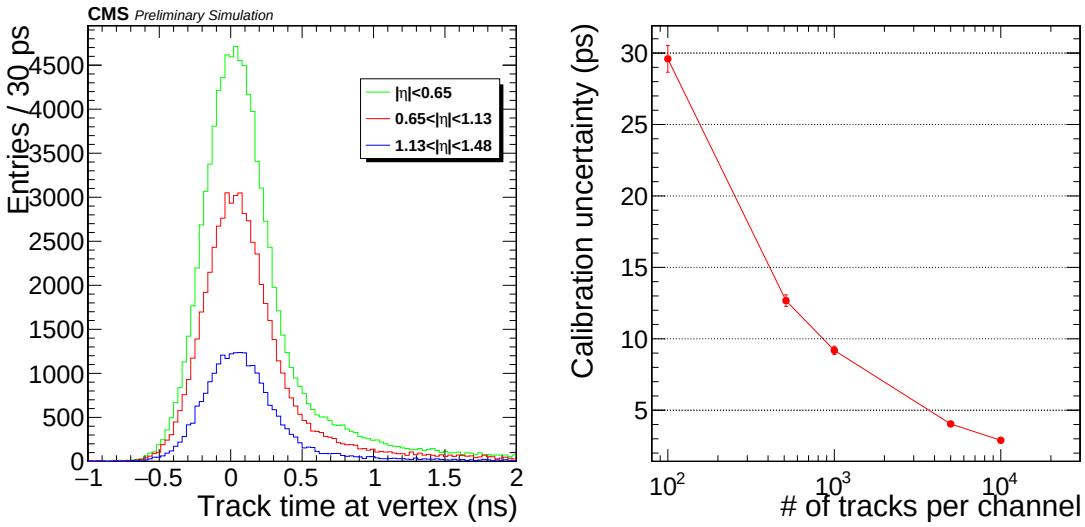


Figure D.3: Left: Distribution of the reconstructed time at the vertex for all the tracks with a time measurement in a representative BTL crystal at $|\eta|$ less than 0.65 (green), between 0.65 and 1.13 (red), and greater than 1.13 (blue); the time-of-flight correction assumes the pion hypothesis; a time offset was added in the simulation for testing purposes. The three distributions are normalized to the same luminosity. Right: Calibration uncertainty as a function of the number of tracks, using the truncated mean at 90% of the distribution in the left panel as the estimator of the time offset.

In the readout scheme of the MTD, a common clock is distributed to the individual channels belonging to the same read-out unit in the BTL or the same service hybrid in the ETL. The RU granularity in the BTL corresponds to 768 channels (384 crystals) and the time stability of the clock distribution can be monitored with a precision of about 2 ps every second (for an occupancy of 5%).

In conclusion, while refined constants may be derived with more stringent selections, the method described here demonstrates that the synchronization of the relative time response across the detector will not be a limiting factor in the MTD operation.

Appendix E

Level-1 MTD performance study

E.I Research on use of MTD in Level-1 trigger

The performance improvements to the L1 Trigger have been evaluated in three studies: 1) isolation improvement in muons and electrons; 2) primary vertex finding; 3) missing transverse energy resolution. A further study, on long-lived particles, is in progress.

In the study of electron and muons isolation, the nominal charged isolation sums the transverse momentum of the tracks around the central object within a cone of 0.4 in ΔR and 0.7 cm in Δz and computes the isolation relative to the transverse momentum of the central object: $\sum_i p_{\mathrm{T}i} / p_{\mathrm{T}}^{\mathrm{obj}}$. Timing information is included by augmenting the requirements such that only tracks within a three-sigma time-window relative to various timing resolutions are included in the sum. The expected performance for muons and electrons are shown in Table E.1. Results indicate that the inclusion of the MTD may provide a reduction in the L1 trigger rate of electron and muon candidates, even for modest time resolutions of order 100 ps.

In addition to this, studies were performed evaluating the L1 missing transverse momentum, $p_{\mathrm{T}}^{\mathrm{miss}}$, resolution based only on the Tracker information. In Fig. E.1 it is seen that with the addition of 50 ps timing resolution the $p_{\mathrm{T}}^{\mathrm{miss}}$ resolution improves by approximately 20% compared to Tracker-only resolution, and with 100 ps the resolution improves by 12%.

The improvements to the L1 vertex finding algorithm have also been evaluated. In the Phase-2 L1 Trigger architecture, primary vertex finding proceeds separately and in parallel to the correlation of charged and neutral particles in the L1 Correlator. Timing information can be used as an additional input to the primary vertex finding algorithm in order to remove tracks that are due to vertices that are spatially close by but not matched in time. The DBScan L1 vertex finding algorithm has been run without and with a requirement that every input track is matched in time to the highest p_{T} track with a timing resolution of 30, 50 or 100 ps. The primary vertex finding efficiency and average number of vertices found per event for 200 pileup collisions is shown in Fig. E.2. When including timing with a 100 ps resolution a 40% reduction in the number of vertices per event was found for an efficiency gain of approximately 0.5% with respect

Table E.1: Rate of the L1 electron and muon candidates for different hypotheses on the time resolution.

	Efficiency	Rate (kHz)			
Object	Threshold/Working Point	No MTD	σ_{T} = 30 ps	σ_{T} = 50 ps	σ_{T} = 100 ps
Electrons	27 GeV / 90%	24	18	19	22
	27 GeV / 95%	27	20	21	24
Muons	18 GeV / 90%	22	16	18	20
	18 GeV / 95%	27	21	22	25

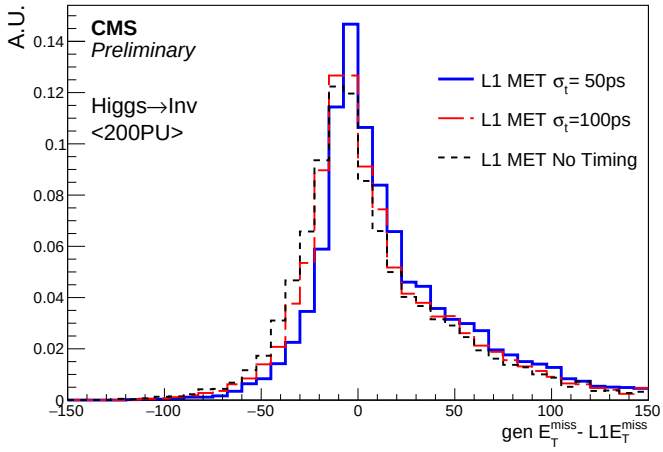


Figure E.1: Missing transverse momentum resolution with no timing, and 50 or 100 ps timing resolution, in simulated events with an Higgs boson decay to invisible particles.

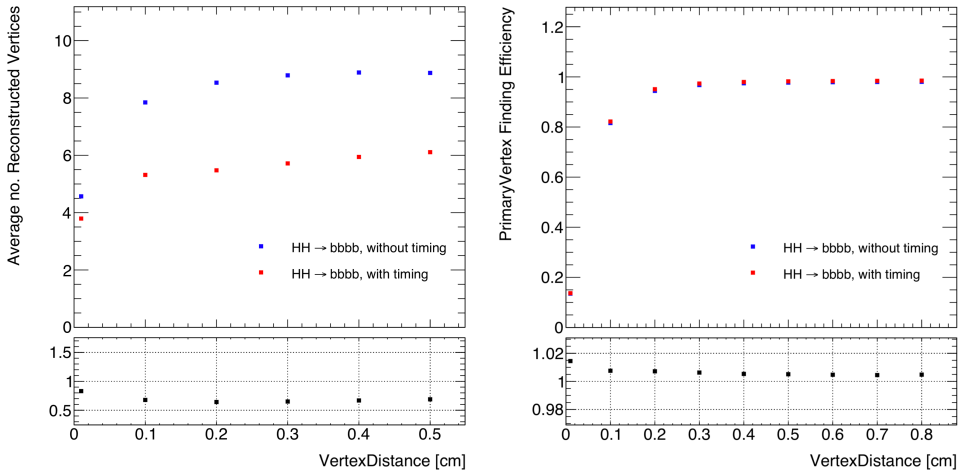


Figure E.2: Number of primary vertices per event and primary vertex finding efficiency without timing and including timing with 50, 100 or 150 ps resolution in $HH \to b\bar{b}b\bar{b}$ events.

to the vertex finding efficiency without timing.

Glossary of Special Terms and Acronyms

2PACL: 2-Phase Accumulator Controlled Loop, the CO_2 cooling system

ALDO: Adjustable Low-dropout linear regulator

ALICE: A Large Ion Collider Experiment, the dedicated Heavy Ion Experiment at CERN

ALTIROC: The read out chip for the ATLAS high granularity timing detector, similar to the CMS ETL

APD: Avalanche Photo Diode

ASIC: Application Specific Integrated Circuit

ATCA: Advanced Telecommunications Computing Architecture

BCID: Bunch Crossing Identifier

BCP: Barrel Calorimeter Processor, an FPGA board in the back-end electronics for the Barrel Calorimeter

BE: Back-end (electronics)

bPOL: A family of radiation-tolerant DC-DC buck converters designed to deliver power to HL-LHC tracking systems.

BSM: Beyond the Standard Model

BTL: Barrel Timing Layer

BX: Bunch crossing (e.g. of LHC beams)

BC0: Bunch crossing number zero

BCID: Bunch crossing Identification number

CBS: Cost Breakdown Structure

CC: Concentrator Card (BTL front end electronics)

CCL: Column Control Logic (of ETL ASIC)

CDATA: Shared Vertical Readout of the ETROC ASIC

CE: Endcap Calorimeter for CMS HL-LHC upgrade (sometimes referred as High Granularity Endcap Calorimeter)

CFD: Constant Fraction Discriminator

CORE Cost: The M&S costs related to a project in CERN accounting, for use by LHCC Cost Review committee

COTS: Commercial off-the-shelf (electronics item)

DAC: Digital to Analog converter

DAQ: Data Acquisition System

DAQ800: An ATCA board that provides additional bandwidth the the data acquisition system

DC/DC: DC to DC power converter

DCR: Dark Count Rate, mainly applicable here to SiPMs

DCS: Detector Control System (sometimes referred to as "slow controls" system)

D2S: Data to Surface, the hardware and software that moves data from the Underground Service Cavern at P5 to the Surface

DDMTD: Digital Dual-Mixer time Difference

DFF: D type clocked Flip-flop

DLED: Differential Leading Edge Discrimination

DLL: Delay-locked loop

DNL: Differential non-linearity (electronic signals)

DPG: Detector Performance Group

DRC: Design Rule Checking (ASIC development terminology)

DRS: A digitizer based on a Switched Capacitor Array DRS4 chip (Domino Ring Sampler)

DSS: Detector Safety System (this system is for equipment protection and safe operation)

DTH: Data Trigger Hub (part of the ATCA back-end electronics common to ETL and BTL)

DUT: Device under test

ECAF: Endcap Calorimeter Assembly Facility

EDMS: Engineering Drawing Management System - a system used at CERN to manage technical and project documents

EDR: Engineering Design Review (CERN Project Management)

E-link: Data link used by GBT to transfer data to end-points

ELM: Embedded Linux Mezzanine Board to go with ZYNQ SoC

ELT: Enclosed Layout Transistors, a method of making radiation tolerant transistors

ENF: Excess noise factor

ENOB: Equivalent Number of Bits

EOF: End of frame (ETL data record)

ERC: Electrical Rule Checking (associated with Design Rule Checking)

ESR: Enhanced Specular Reflector, also called Vikuiti, produced by 3M

ETL: End Cap Timing Layer

ETROC: Readout ASIC for ETL

EN-CV: CERN Engineering Group in charge of Cooling plants and similar facilities

EP-DT: Cern Experimental Physics Department Detector Technologies Group

FBK: Fondazione Bruno Kessler

FE: Front End (electronics)

FEAST: A family of DC-DC converters developed at CERN for use in high energy physics experiments

Firefly Transceivers: A high speed optical data connection system for high speed optical data trnasmission

FPGA: Field Programmable Gate Array

FTBF: Fermilab Test Beam Facility

GBT: Gigabit Transceiver, a family of radiation hard ASICs to support high speed data transmission over optical links

GBT-SCA: A member of the GBT Chipset designed to interface Slow Control (SC) data for transmission by the GBT system

HCAL: The current CMS Hadron Calorimeter, which employs SiPMs

HF: CMS Forward Hadron Calorimeter

HGC/HGCAL: High Granularity Calorimeter, also referred to as Endcap Calorimeter

HL-LHC : High Luminosity LHC

HLT: High-Level Trigger (second stage of CMS trigger, based on large clusters of microprocessors)

HPK: Hamamatsu Photonics K.

HV: High Voltage

IB: Institution Board

INL: Integrated Non-Linearity

Interposer: electrical interface routing between one socket or connection to another.

IPMC: Intelligent Platform Management Controller

IT: Inner Tracker for CMS HL-LHC Upgrade

JTE: Junction Termination Extension (A doping technique used to avoid breakdown in LGADs)

LCE: Light Collection Efficiency

LE: Leading Edge (timing technique based on leading edge discrimination and pulse height correction of time walk)

LED: Light Emitting Diode

LGAD : Low Gain Avalanche Detector

LHCC: LHC Committee. CERN's main review committee for the LHC program.

LLP: Long-lived particle, a particle that can travel an observable distance from its production point to when it decays

lpGBT: low-power Giga-Bit Transceiver

LUT: Look up table

LV: Low voltage

LVS: Layout Vs Schematic (associated with DRC)

LYSO:Ce: Cerium-doped Lutetium-Yttrium Oxyorthosilicate, a scintillator crystal.

MCP: MicroChannel Plate

MicroTCA: Micro-Telecommunications Architecture, a small form factor version of an open standard embedded computing specification

MIP: Minimum Ionizing Particle

MoU: Memorandum of Understanding

MPV: Most Probable Value

MTD: MIP Timing Detector

NIEL: Non-Ionizing Energy Loss

NINO: An ultrafast front-end preamplifier-discriminator chip called NINO developed for use in the ALICE time-of-flight detector

NUV-HD: Near UV High Density, referring to the SiPM technology

OT: Outer Tracker for CMS HL-LHC Upgrade

PBCL: Pixel Buffer Control Logic

PCB: Printed Circuit Board

PCC: Power Converter Card (BTL)

PCDM Board: Precision Clock Distribution and Monitoring Board

PDE: Photon Detection Efficiency

PET: Positron Emission Tomography

PID: Particle Identification (Specifically here charge hadron, i.e. pion, kaon, and proton, identification)

PLL: Phase Locked loop

PLC: Programmable Logic Controller

PM: Project Manager

Point 5 (P5): locations of CMA, the fifth of the eight regions or points where the LHC cross and where four of them collide

PON: Passive Optical Network (data and control information transfer)

PP0: Patch Panel 0 (Connection point close to the CMS active detector elements)

PP1: Patch Panel 1 (patch panel a few meters from the CMS active detector elements)

PU: Pileup

PUPPI: Pileup per Particle Interaction. An approach and code for pileup mitigation at the LHC.

PVT: A testing procedure for ASICs exploring the phase space of Process, Voltage, and Temperature variations

QA: Quality Assurance

QAC: Charge-to-Amplitude Converter

QC: Quality Control

QDC: Charge-to-digital converter

RCL: Row Control Logic

RINCE: Radiation-Induced Narrow-Channel Effect (MOSFET rules for reducing impact of radiation on ASICs)

RISCE: Radiation-Induced Short-Channel Effect (MOSFET rules for reducing impact of radiation on ASICs)

ROC: Read Out Chip (ASIC)

ROC: Receiver Operating Characteristic Curve

ROI: Region of interest

RU: Readout Unit (BTL)

Run 1: LHC data-taking period in 2011 (7 TeV) and 2012 (8 TeV)

Run 2: LHC data-taking period from 2015-2018 (13 TeV)

Run 3: LHC data-taking period projected for 2021-2024 at 14 TeV

Run 4: First LHC data-taking run after HL-LHC Upgrade projected to start in 2026/2027

SAR: Successive Approximation Register

SC: Steering Committee

SCA: Slow Control Adapter, a member of the GBT chipset

SCADA: Supervisory Control and Data Acquisition , a generic name for a system for slow control and monitoring

Serenity Board: A board for processing digital data using a large Xilinx FPGA and many high speed data links

SiPAD: A hexagonal silicon sensor used in the CMS Endcap calorimeter

SiPM: Silicon PhotoMultiplier

SKYROC: a readout chip designed for the CALICE Calorimeter and used by the CMS CE and others for a variety of R&D and developmental tasks

SM: Standard Model

SMA: Subminiature connector version A

SNR: Signal-to-Noise Ratio

SoC: System on aCchip

SOF: Start Of Frame (ETL data record)

SPTR: Single photon time response

SRAM: Static Random Access Memory

STAR: An experiment at the Relativistic Heavy Ion Collider (RHIC) at Brookhaven National Laboratory (USA)

TAC: Time to Analog Converter

TCDS: CMS Timing and Control Distribution System

TCDS2: CMS Timing and Control Distribution System for the HL-LHC

TCG: Technical Coordination Group (of CMS)

TDC: Time to Digital Converter

TDR: Technical Design Report

TEDD: Tracker Endcap Disk Detector for CMS HL-LHC Upgrade

TIA: Transimpedance Amplifier

TID: Total Ionizing Dose (radiation)

TIF: Tracker Integration Facility at CERN, area for commissioning large tracking detectors

Time walk: The tendency of a threshold crossing time of a pulse to shift as a function of its pulse height

TM: Technical Manager

TMR: Triple Modular Redundancy, a voting system among three identical registers to protect chips against Single Event Effects

TOA: Time of Arrival, in Time-of-Flight application

TOF Time of Flight

TOFHIR: Readout out ASIC for BTL

TOFPET2: A commercial ASIC used in Pet Scanning applications that include time measurements

TOT: Time Over Threshold, correlated to pulse amplitude in tof application, used to correct for time walk

TP: Technical Proposal

TST: Tracker Support Tube

UBM: Under-bump metallization

UFSD: UltraFast Silicon Detector

UPS: Un-interruptable Power Supply (or source)

USC: Underground Service Cavern in CMS experiment at CERN

UXC: Underground Collision Cavern

VE: Value Engineering

VL: Versatile Link

VTRX: radiation tolerant optical transceivers

ZYNQ: A programmable SoC (system on chip) from XILINX. Can be considered an IPMC.

References

[1] G. Apollinari, O. Brning, T. Nakamoto, and L. Rossi, "High Luminosity Large Hadron Collider HL-LHC", *CERN Yellow Report* (2015) 1, doi:10.5170/CERN-2015-005.1, arXiv:1705.08830.

[2] CMS Collaboration, "Technical Proposal for the Phase-II Upgrade of the CMS Detector", Technical Report CERN-LHCC-2015-010. LHCC-P-008, CERN, Geneva, Jun, 2015.

[3] CMS Collaboration, "CMS Phase II Upgrade Scope Document", Technical Report CERN-LHCC-2015-019. LHCC-G-165, CERN, Geneva, Sep, 2015.

[4] CMS Collaboration, "The Phase-2 Upgrade of the CMS Tracker", Technical Report CERN-LHCC-2017-009. CMS-TDR-014, CERN, Geneva, Jun, 2017.

[5] CMS Collaboration, "The Phase-2 Upgrade of the CMS Muon Detectors", Technical Report CERN-LHCC-2017-012. CMS-TDR-016, CERN, Geneva, Sep, 2017.

[6] CMS Collaboration, "The Phase-2 Upgrade of the CMS Barrel Calorimeters Technical Design Report", Technical Report CERN-LHCC-2017-011. CMS-TDR-015, CERN, Geneva, Sep, 2017.

[7] CMS Collaboration, "The Phase-2 Upgrade of the CMS Endcap Calorimeter", Technical Report CERN-LHCC-2017-023. CMS-TDR-019, CERN, Geneva, Nov, 2017.

[8] CMS Collaboration, "Technical Proposal for a MIP Timing Detector in the CMS experiment Phase 2 upgrade", Technical Report CERN-LHCC-2017-027. LHCC-P-009, CERN, Geneva, Dec, 2017.

[9] CMS Collaboration, "The CMS experiment at the CERN LHC", *JINST* **3** (2008) SO8001, doi:10.1088/1748-0221/3/08/SO8001.

[10] CMS Collaboration, "Particle-flow reconstruction and global event description with the CMS detector", *JINST* **12** (2017) P10003.

[11] D. Bertolini, P. Harris, M. Low, and N. Tran, "Pileup Per Particle Identification", *JHEP* **1410** (2014) 59, doi:10.1007/JHEP10(2014)059, arXiv:1407.6013.

[12] CMS Collaboration, "Updates on Performance of Physics Objects with the Upgraded CMS detector for High Luminosity LHC", CMS Performance Note CMS-DP-2016-065, 2016.

[13] CMS Collaboration, "Description and performance of track and primary-vertex reconstruction with the CMS tracker", *JINST* **9** (2014) P10009, doi:10.1088/1748-0221/9/10/P10009, arXiv:1405.6569.

[14] LHC Higgs Cross Section Working Group Collaboration, "Handbook of LHC Higgs Cross Sections: 4. Deciphering the Nature of the Higgs Sector", arXiv:1610.07922.

[15] CMS Collaboration, "Enhanced scope of a Phase 2 CMS detector for the study of exotic physics signatures at the HL-LHC", Technical Report CMS-PAS-EXO-14-007, CERN, Geneva, 2016.

[16] A. Coccaro et al., "Data-driven model-independent searches for long-lived particles at the LHC", *Phys. Rev. D* **94** (2016) 113003, doi:10.1103/PhysRevD.94.113003, arXiv:1605.02742.

[17] STAR and the CBM eTOF Collaboration, "Physics Program for the STAR–CBM Upgrade", (2016). https://arxiv.org/abs/1609.05102.

[18] ALICE Collaboration, "Performance of the ALICE Time-Of-Flight Detector at the LHC", *Eur. Phys. J. Plus* **128** (2013) 44, doi:10.1140/epjp/i2013-13044-x.

[19] ALICE Collaboration, "Performance of the ALICE Time-Of-Flight detector at the LHC", *submitted to JINST* (2018). https://arxiv.org/pdf/1806.03825.pdf.

[20] P. Elmer, B. Hegner, and L. Sexton-Kennedy, "Experience with the CMS Event Data Model", *J. Phys. Conf. Ser.* **219** (2010) 032022, doi:10.1088/1742-6596/219/3/032022.

[21] A. Ferrari, P. R. Sala, A. Fassò, and J. Ranft, "FLUKA: A multi-particle transport code (Program version 2005)", technical report, 2005.

[22] T. T. Bhlen et al., "The FLUKA Code: Developments and Challenges for High Energy and Medical Applications", *Nucl. Data Sheets* **120** (2014) 211–214, doi:10.1016/j.nds.2014.07.049.

[23] S. Gundacker et al., "Time of flight positron emission tomography towards 100 ps resolution with L(Y)SO: an experimental and theoretical analysis", *JINST* **8** (2013) P07014.

[24] D. Anderson et al., "On timing properties of LYSO-based calorimeters", *Nucl. Instrum. Meth. A* **794** (2015) 7, doi:10.1016/j.nima.2015.04.013.

[25] D. Anderson et al., "Precision Timing Measurements for High Energy Photons", *Nucl.Instrum.Meth. A* **787** (2015) 94, doi:10.1016/j.nima.2014.11.041.

[26] S. N. White, "R&D for a Dedicated Fast Timing Layer in the CMS Endcap Upgrade", *Acta Phys. Pol. B Proc. Suppl.* **7** (2014) 743, doi:10.5506/APhysPolBSupp.7.743, arXiv:1409.1165.

[27] G. Pellegrini et al., "Technology developments and first measurements of Low Gain Avalanche Detectors (LGAD) for high energy physics applications", *Nucl. Instrum. Meth. A* **765** (2014) 12, doi:10.1016/j.nima.2014.06.008.

[28] N. Cartiglia et al., "Design optimization of ultra-fast silicon detectors", *Nucl. Instrum. Meth. A* **796** (2015) 141, doi:10.1016/j.nima.2015.04.025.

[29] A. Ronzhin et al., "Development of a new fast shower maximum detector based on microchannel plates photomultipliers (MCP-PMT) as an active element", *Nucl. Instrum. Meth. A* **759** (2014) 65, doi:http://dx.doi.org/10.1016/j.nima.2014.05.039.

[30] A. Ronzhin et al., "Direct Tests of Micro Channel Plates as the Active Element of a New Shower Maximum Detector", *Nucl. Instrum. Meth. A* **795** (2015) 52, doi:10.1016/j.nima.2015.05.029.

[31] L. Brianza et al., "Response of microchannel plates to single particles and to electromagnetic showers", *Nucl. Instrum. Meth. A* **797** (2015) 216, doi:10.1016/j.nima.2015.06.057, arXiv:1504.02728.

[32] A. Barnyakov et al., "Beam test results on the detection of single particles and electromagnetic showers with microchannel plates", *Nucl. Instrum. Meth. A* **845** (2017) 471, doi:10.1109/NSSMIC.2015.7581993.

[33] T. Papaevangelou et al., "Fast Timing for High-Rate Environments with Micromegas", in *4th International Conference on Micro Pattern Gaseous Detectors - MPGD2015 Trieste, Italy, October 12-15, 2015*. 2016. arXiv:1601.00123.

[34] A. Benaglia et al., "Detection of high energy muons with sub-20 ps timing resolution using L(Y)SO crystals and SiPM readout", *Nucl. Instrum. Meth. A* **830** (2016) 30, doi:10.1016/j.nima.2016.05.030.

[35] M. D. Rolo et al., "TOFPET ASIC for PET applications", *JINST* **8** (2013) C02050.

[36] M. D. Rolo et al., "A low-noise CMOS front-end for TOF-PET", *JINST* **6** (2011) P09003.

[37] A. D. Francesco et al., "TOFPET2: a high-performance ASIC for time and amplitude measurements of SiPM signals in time-of-flight applications", *JINST* **11** (2016) C03042.

[38] ATLAS Collaboration, "ATLAS Phase-II Upgrade Scoping Document", Technical Report CERN-LHCC-2015-020. LHCC-G-166, CERN, Geneva, Sep, 2015.

[39] S. Baron et al., "Passive Optical Network for TTC", Communication to ACES 2011 - Common ATLAS CMS Electronics Workshop for LHC Upgrades, 2011.

[40] E. Garutti and Y. Musienko, "Radiation damage of SiPMs", arXiv:1809.06361.

[41] A. Gola, C. Piemonte, and A. Tarolli, "The DLED Algorithm for Timing Measurements on Large Area SiPMs Coupled to Scintillators", *IEEE Transactions on Nuclear Science - IEEE TRANS NUCL SCI* **59** (04, 2012) 358–365, doi:10.1109/TNS.2012.2187927.

[42] S. Gundacker, R. Turtos, E. Auffray, and P. Lecoq, "Precise rise and decay time measurements of inorganic scintillators by means of X-ray and 511 keV excitation", *Nucl. Instrum. Meth. A* **891** (2018) 42 – 52, doi:https://doi.org/10.1016/j.nima.2018.02.074.

[43] R. Mao, L. Zhang, and R.-Y. Zhu, "LSO/LYSO Crystals for Future HEP Experiments", *Journal of Physics: Conference Series* **293** (2011) 012004.

[44] F. Yang et al., "Proton induced radiation damage in fast crystal scintillators", *Nucl. Instrum. Meth. A* (2016) 726, doi:10.1016/j.nima.2015.11.100.

[45] F. Yang, L. Zhang, and R. Y. Zhu, "Gamma-ray induced radiation damage up to 340 Mrad in various scintillation crystals", *IEEE Trans. Nucl. Sci.* **63** (April, 2016) 612–619, doi:10.1109/TNS.2015.2505721.

[46] E. Auffray et al., "Radiation damage of LSO crystals under γ and 24 GeV protons irradiation", *Nucl.Instrum. Meth. A* **721** (2013) 76, doi:https://doi.org/10.1016/j.nima.2013.04.065.

[47] A. Gektin, P. Lecoq, and M. Korjik, "Inorganic scintillators for detector systems", *Springer* (2016) 408.

[48] M. Korjik and E. Auffray, "Limits of inorganic scintillating materials to operate in a high dose rate environment at future collider experiments", *IEEE Transactions on Nuclear Science* **63** (April, 2016) 552–563, doi:10.1109/TNS.2016.2527701.

[49] CMS Collaboration, "CMS Technical Design Report for the Phase 1 Upgrade of the Hadron Calorimeter", Technical Report CERN-LHCC-2012-015. CMS-TDR-10, Sep, 2012.

[50] Y. Musienko et al., "Radiation damage studies of silicon photomultipliers for the CMS HCAL phase I upgrade", *Nucl. Instrum. Meth. A* **787** (2015) 319, doi:10.1016/j.nima.2015.01.012.

[51] A. Heering et al., "Effects of very high radiation on SiPMs", *Nucl. Instrum. Meth. A* **824** (2016) 111, doi:10.1016/j.nima.2015.11.037.

[52] Y. Musienko and A. A. Karneyeu, "Studies of SiPMs for the CMS HCAL upgrade", *PoS - PhotoDet2015* (2015) 073.

[53] CMS Collaboration, "The CMS electromagnetic calorimeter project: Technical Design Report", Technical Report CERN-LHCC-97-033. CMS-TDR-4, CERN, Geneva, 1997.

[54] R. Chipaux and A. Soyer, "Study of some optical glues in a LHC-like environment", Technical Report DAPNIA-SED-95-01, DAPNIA, Saclay, Feb, 1995.

[55] T. Kirn, M. Haering, D. Schmitz, and W. Schulz, "Absorption length, radiation hardness and ageing of different optical glues", Technical Report CMS-NOTE-1999-003, CERN, Geneva, Jan, 1999.

[56] CERN, "LpGBT specification document", 2017. https://espace.cern.ch/GBT-Project/LpGBT/Specifications/.

[57] CERN, "Versatile link plus project", 2017. https://espace.cern.ch/project-Versatile-Link-Plus/SitePages/Home.aspx.

[58] R. Bugalho et al., "Experimental results with TOFPET2 ASIC for time-of-flight applications", *Nucl. Instrum. Meth.A* **912** (2018) 195, doi:10.1016/j.nima.2017.11.034.

[59] L. B. Oliveira, C. M. Leit, and M. M. Silva, "Noise Performance of Regulated Cascode Transimpedance Amplifiers for Radiation Detectors", *IEEE Trans. Circuits and System I* **59** (2012), no. 9, 1841.

[60] M. Firlej et al., "A fast, ultra-low and frequency-scalable power consumption, 10-bit SAR ADC for particle physics detectors", *JINST* **10** (2015), no. 11, P11012, doi:10.1088/1748-0221/10/11/P11012.

[61] L. Gonella et al., "Total ionizing dose effects in 130-nm commercial CMOS technologies for HEP experiments", *Nucl. Instrum. Meth. A* **582** (2007) 750, doi:10.1016/j.nima.2007.07.068.

[62] P. Carniti et al., "ALDO: A radiation-tolerant, low-noise, adjustable low drop-out linear regulator in 0.35 μm CMOS technology", *Nucl. Instrum. Meth. A* **824** (2016) 258, doi:10.1016/j.nima.2015.10.054.

[63] F. Faccio et al., "TID and Displacement Damage Effects in Vertical and Lateral Power MOSFETs for Integrated DC-DC Converters", *IEEE Trans. Nucl. Sci.* **57** (2010), no. 4, Part 1, 1790–1797, doi:10.1109/TNS.2010.2049584.

[64] Wiener, "Wiener power supplies", 2013. http://www.wiener-d.com/sc/power-supplies/mpod--lvhv/mpod-lv-module.html.

[65] H. F.-W. Sadrozinski, A. Seiden, and N. Cartiglia, "4D tracking with ultra-fast silicon detectors", *Rept. Prog. Phys.* **81** (2018) 026101.

[66] N. Cartiglia et al., "Beam test results of a 16 ps timing system based on ultra-fast silicon detectors", *Nucl. Instrum. Meth. A* **850** (2017) 83, doi:10.1016/j.nima.2017.01.021, arXiv:1608.08681.

[67] M. Carulla et al., "First 50 μm thick LGAD fabrication at CNM for the HGTD and CT-PPS", presented at the 28th RD50 Workshop, June, 2016.

[68] G.-F. Dalla Betta et al., "Design and TCAD simulation of double-sided pixelated low gain avalanche detectors", *Nucl. Instrum. Meth. A* **796** (2015) 154, doi:10.1016/j.nima.2015.03.039.

[69] V. Sola et al., "First FBK Production of 50 μm Ultra-Fast Silicon Detectors", *Nucl. Instrum. Meth. A* **924** (2019) 360, doi:10.1016/j.nima.2018.07.060, arXiv:1802.03988.

[70] H. F. W. Sadrozinski, "Timing Measurements on thin LGAD", Presented at TREDI 2017, Trento, Italy, Feb, 2017.

[71] Z. Galloway et al., "Properties of HPK UFSD after neutron irradiation up to 6e15 n/cm^2", arXiv:1707.04961.

[72] G. Kramberger and et al., "Radiation effects in Low Gain Avalanche Detectors after hadron irradiations", *JINST* **10** (2015) P07006.

[73] Z. Galloway et al., "Properties of HPK UFSD after neutron irradiation up to 6×15 n/cm^{-2}", arXiv:1707.04961.

[74] M. Ferrero et al., "Radiation resistant LGAD design", *Nucl. Instrum. Meth. A* **919** (2019) 16, doi:10.1016/j.nima.2018.11.121, arXiv:1802.01745.

[75] S. Terada et al., "Proton irradiation on P bulk silicon strip detectors using 12-GeV PS at KEK", *Nucl. Instrum. Meth.* **A383** (1996) 159–165, doi:10.1016/S0168-9002(96)00748-6.

[76] M. Mandurrino et al., "Numerical simulation of charge multiplication in ultra-fast silicon detectors (ufsd) and comparison with experimental data", pp. 1–4. 10, 2017. doi:10.1109/NSSMIC.2017.8532702.

[77] J. Balbuena and et al., "RD50 Status Report 2008 - Radiation hard semiconductor devices for very high luminosity colliders", Technical Report CERN-LHCC-2010-012. LHCC-SR-003, CERN, Geneva, Sep, 2010.

[78] L. Snoj, G. Zerovnik, and A. Trkov, "Computational analysis of irradiation facilities at the JSI TRIGA reactor", *Applied Radiation and Isotopes* **70** (2012) 483–488.

[79] KIT Irradiation Center, "Proton Irradiation". https://www.etp.kit.edu/english/264.php.

[80] A. Terakawa, K. Ishii, and T. Chiba, "Proton therapy facilities at CYRIC, Tohoku University", Jul, 2008.

[81] B. Gkotse et al., "Irradiation Facilities at CERN". http://cds.cern.ch/record/2288578/.

[82] A. Apresyan and et al., "Studies of uniformity of 50 micron low-gain avalanche detectors at the fermilab test beam", *Nucl. Instrum. Meth. A* **895** (2018) 158, doi:https://doi.org/10.1016/j.nima.2018.03.074.

[83] M. Tornago, "Performances of the third UFSD production at FBK", 33th RD50 Workshop, CERN, Geneva (CH), 2018.

[84] Y. Zhao et al., "Comparison of 35 and 50 μm thin hpk ufsd after neutron irradiation up to 6e15 neq/cm2", *Nucl. Instr. Meth. A* (2018) doi:https://doi.org/10.1016/j.nima.2018.08.040.

[85] S. M. Mazza et al., "Properties of FBK UFSDs after neutron and proton irradiation up to 6×10^{15} neq/cm^2", arXiv:1804.05449.

[86] N. Cartiglia, "Why shot noise in LGAD does not degrade time resolution?", Presented at TREDI 2018, Munich, Germany, Feb, 2018.

[87] F. Siviero, "Studies of the Breakdown in R&D structures in FBK UFSD3", 33th RD50 Workshop, CERN, Geneva (CH), 2018.

[88] G. Kramberger, "Performance of thin LGADs after long term annealing", 32th RD50 Workshop, Hamburg, Germany, 2018.

[89] A. Apresyan et al., "Studies of the uniformity of 50 μm low-gain avalanche detectors at the Fermilab test beam", submitted to *Nucl. Instrum. Meth. A*, 2017.

[90] R. Arcidiacono, "Study of the radiation resistance of different LGAD gain layer designs", Presented at TREDI 2018, Munich, Germany, Feb, 2018.

[91] L. Paolozzi, "Silicon monolithic pixel detectors in a SiGe Bi-CMOS process for sub-100 ps time resolution", Presented at TREDI 2017, Trento, Italy, Feb, 2017.

[92] P. Dudek, S. Szczepanski, and J. Hatfield, "A high resolution cmos time-to-digital converter utilizing a vernier delay line", *Solid State Circuits Journal* **35** (1, 2000) 240–247.

[93] L. Jara Casas et al., "Characterization of radiation effects in 65 nm digital circuits with the drad digital radiation test chip", *JINST* **12** (02, 2017) C02039.

[94] ATLAS Collaboration, "Technical Proposal: A High-Granularity Timing Detector for the ATLAS Phase-II Upgrade", Technical Report CERN-LHCC-2018-023. LHCC-P-012, CERN, Geneva, Jun, 2018.

[95] T. C. Edwards and M. B. Steer, "Foundations for microstrip circuit design". John Wiley & Sons, 2016.

[96] E. G. Friedman, "Clock distribution networks in synchronous digital integrated circuits", *Proceedings of the IEEE* **89** (2001) 665–692.

[97] D. Velenis, R. Sundaresha, and E. G. Friedman, "Buffer sizing for delay uncertainty induced by process variations", in *Proceedings of the 2004 11th IEEE International Conference on Electronics, Circuits and Systems*, pp. 415–418. 2004.

[98] P. Fischer, "First Implementation of the MEPHISTO Binary Readout Architecture for Strip Detectors", *Nucl. Instrum. Meth. A* **461** (2001) 499.

[99] P. Fischer, G. Comes, and H. Kruger, "First implementation of the mephisto binary readout architecture for strip detectors", *Nucl. Instrum. Meth. A* **431** (1999) 134–140.

[100] K. Sun et al., "A 31.5-GHz BW 6.4-b ENOB 56-GS/s ADC in 28nm CMOS for 224-Gb/s DP-16QAM coherent receivers", in *2018 IEEE Custom Integrated Circuits Conference (CICC)*, p. 1. 2018.

[101] G. Wang et al., "A 43.6-dB SNDR 1-GS/s single-channel SAR ADC using coarse and fine comparators with background comparator offset calibration", in *ESSCIRC 2017 - 43rd IEEE European Solid State Circuits Conference*, pp. 175–178. 2017.

[102] S. Bonacini, K. Kloukinas, and P. Moreira, "E-link: A Radiation-Hard Low-Power Electrical Link for Chip-to-Chip Communication", doi:10.5170/CERN-2009-006.422.

[103] RD53 Collaboration, "RD53A Integrated Circuit Specifications", Technical Report CERN-RD53-PUB-15-001, CERN, Geneva, Dec, 2015.

[104] F. Faccio et al., "Radiation-Induced Short Channel (RISCE) and Narrow Channel (RINCE) Effects in 65 and 130 nm MOSFETs", *IEEE Transactions on Nuclear Science* **62** (2015), no. 6, 2933–2940.

[105] R. Brouns et al, "The development of a radiation tolerant low power SRAM compiler in 65nm technology", in *AMICSA*. 2014.

[106] S. Miryala et al, "Characterization of Soft Error Rate Against Memory Elements Spacing and Clock Skew in a Logic with TMR in a 65nm Process", in *TWEPP*. 2018.

[107] CALICE, EUDET Collaboration, M. Bouchel et al., "Skiroc: A front-end chip to read out the imaging silicon-tungsten calorimeter for ILC", in *Electronics for particle physics. Proceedings, Topical Workshop, TWEPP-08, Naxos, Greece, 15-19 September 2008*, pp. 463–467. 2008.

[108] CERN, "The slow control adapter ASIC for the GBT system: User Manual", 2017. https://espace.cern.ch/GBT-Project/GBT-SCA.

[109] J.-M. Andr et al., "The CMS Data Acquisition System for the Phase-2 Upgrade", arXiv:1806.08975.

[110] CMS Collaboration, "Serenity - An ATCA prototyping platform for CMS Phase-2", Technical Report CMS-CR-2018-327, CERN, Geneva, Oct, 2018.

[111] CMS Collaboration, "The Barrel Calorimeter Processor demonstrator board for the Phase II Upgrade of the CMS ECAL Barrel", Technical Report CMS-CR-2018-301, CERN, Geneva, Oct, 2018.

[112] S. Baron, T. Mastoridis, J. Troska, and P. Baudrenghien, "Jitter impact on clock distribution in LHC experiments", *JINST* **7** (2012), no. 12, C12023–C12023, doi:10.1088/1748-0221/7/12/c12023.

[113] GEANT4 Collaboration, "GEANT4: A Simulation toolkit", *Nucl. Instrum. Meth. A* **506** (2003) 250–303, doi:10.1016/S0168-9002(03)01368-8.

[114] DELPHES 3 Collaboration, "DELPHES 3, A modular framework for fast simulation of a generic collider experiment", *JHEP* **02** (2014) 057, doi:10.1007/JHEP02(2014)057, arXiv:1307.6346.

[115] M. Cacciari, G. P. Salam, and G. Soyez, "The Anti-k(t) jet clustering algorithm", *JHEP* **04** (2008) 063, doi:10.1088/1126-6708/2008/04/063, arXiv:0802.1189.

[116] CMS Collaboration, "Searches for light higgsino-like charginos and neutralinos at the HL-LHC with the Phase-2 CMS detector", Technical Report CMS-PAS-FTR-18-001, CERN, Geneva, 2018.

[117] CMS Collaboration, "Prospects for HH measurements at the HL-LHC", Technical Report CMS-PAS-FTR-18-019, CERN, Geneva, 2018.

[118] P. Meade, M. Reece, and D. Shih, "Long-lived neutralino NLSPs", *JHEP* **10** (2010) 067, doi:10.1007/JHEP10(2010)067, arXiv:1006.4575.

[119] T. Sjöstrand et al., "An Introduction to PYTHIA 8.2", *Comput. Phys. Commun.* **191** (2015) 159, doi:10.1016/j.cpc.2015.01.024, arXiv:1410.3012.

[120] H. Baer, V. Barger, and P. J. Huang, "Hidden susy at the lhc: the light higgsino-world scenario and the role of a lepton collider", *JHEP* **11** (2011) 031, doi:10.1007/JHEP11(2011)031, arXiv:1107.5581.

[121] K. Rolbiecki and K. J. Sakurai, "Long-lived bino and wino in supersymmetry with heavy scalars and higgsinos", *JHEP* **11** (2015) 091, doi:10.1007/JHEP11(2015)091, arXiv:1506.08799.

[122] M. J. Strassler and K. M. Zurek, "Echoes of a hidden valley at hadron colliders", *Physics Letters B* **651** (2007) 374, doi:10.1016/j.physletb.2007.06.055.

[123] H. Baer, V. Barger, D. Mickelson, and M. Padeffke-Kirkland, "SUSY models under siege: LHC constraints and electroweak fine-tuning", *Phys. Rev. D* **89** (2014) 115019, doi:10.1103/PhysRevD.89.115019, arXiv:1404.2277.

[124] CMS Collaboration, "Searches for long-lived charged particles in pp collisions at $\sqrt{s} = 7$ and 8 TeV", *JHEP* **2013** (2013), no. 7, 122, doi:10.1007/JHEP07(2013)122.

[125] CMS Collaboration, "Search for heavy stable charged particles with 12.9 fb^{-1} of 2016 data", Technical Report CMS-PAS-EXO-16-036, CERN, Geneva, 2016.

[126] CMS Collaboration, "Nuclear modification factor of D^0 mesons in PbPb collisions at $\sqrt{s_{\text{NN}}} = 5.02$ TeV", *Phys. Lett. B* **782** (2018) 474, doi:10.1016/j.physletb.2018.05.074, arXiv:1708.04962.

[127] ALICE Collaboration, "Measurement of D^0, D^+, D^{*+} and D_s^+ production in Pb-Pb collisions at $\sqrt{s_{\text{NN}}} = 5.02$ TeV", *JHEP* **10** (2018) 174, doi:10.1007/JHEP10(2018)174, arXiv:1804.09083.

[128] S. Plumari et al., "Charmed Hadrons from Coalescence plus Fragmentation in relativistic nucleus-nucleus collisions at RHIC and LHC", *Eur. Phys. J.* **C78** (2018) 348, doi:10.1140/epjc/s10052-018-5828-7, arXiv:1712.00730.

[129] ALICE Collaboration, "Λ_c^+ production in pp collisions at $\sqrt{s} = 7$ TeV and in p-Pb collisions at $\sqrt{s_{\rm NN}} = 5.02$ TeV", *JHEP* **04** (2018) 108, doi:10.1007/JHEP04(2018)108, arXiv:1712.09581.

[130] ALICE Collaboration, "Λ_c^+ production in Pb-Pb collisions at $\sqrt{s_{\rm NN}} = 5.02$ TeV", *Submitted to: Phys. Lett.* (2018) arXiv:1809.10922.

[131] CMS Collaboration, "Elliptic flow of charm and strange hadrons in high-multiplicity pPb collisions at $\sqrt{s_{_{\rm NN}}} = 8.16$ TeV", *Phys. Rev. Lett.* **121** (2018) 082301, doi:10.1103/PhysRevLett.121.082301, arXiv:1804.09767.

[132] T. Pierog et al., "EPOS LHC: Test of collective hadronization with data measured at the CERN Large Hadron Collider", *Phys. Rev. C* **92** (2015) 034906, doi:10.1103/PhysRevC.92.034906, arXiv:1306.0121.

[133] S. Kwan et al., "The pixel tracking telescope at the Fermilab Test Beam Facility", *Nucl. Instrum. and Meth. A* **811** (2016) 162, doi:http://dx.doi.org/10.1016/j.nima.2015.12.003.

[134] A. Ronzhin et al., "Study of the timing performance of micro-channel plate photomultiplier for use as an active layer in a shower maximum detector", *Nucl. Instrum. Meth. A* **795** (2015) 288292, doi:http://dx.doi.org/10.1016/j.nima.2015.06.006.

[135] CAEN, "Caen power supplies", 2018. http://www.wiener-d.com/sc/power-supplies/mpod--lvhv/mpod-lv-module.html.

CMS Collaboration

Yerevan Physics Institute, Yerevan, Armenia
A.M. Sirunyan[†], A. Tumasyan

Institut für Hochenergiephysik, Wien, Austria
W. Adam, F. Ambrogi, T. Bergauer, J. Brandstetter, M. Dragicevic, J. Erö, A. Escalante Del Valle, M. Flechl, R. Frühwirth[1], M. Jeitler[1], N. Krammer, I. Krätschmer, D. Liko, T. Madlener, I. Mikulec, N. Rad, J. Schieck[1], R. Schöfbeck, M. Spanring, D. Spitzbart, W. Waltenberger, C.-E. Wulz[1], M. Zarucki

Institute for Nuclear Problems, Minsk, Belarus
A. Fedorov, M. Korzhik, V. Mechinsky

Universiteit Antwerpen, Antwerpen, Belgium
M.R. Darwish, E.A. De Wolf, D. Di Croce, X. Janssen, J. Lauwers, A. Lelek, M. Pieters, H. Rejeb Sfar, H. Van Haevermaet, P. Van Mechelen, S. Van Putte, N. Van Remortel

Vrije Universiteit Brussel, Brussel, Belgium
F. Blekman, E.S. Bols, S.S. Chhibra, J. D'Hondt, J. De Clercq, D. Lontkovskyi, S. Lowette, I. Marchesini, S. Moortgat, L. Moreels, Q. Python, K. Skovpen, S. Tavernier, W. Van Doninck, P. Van Mulders, I. Van Parijs

Université Libre de Bruxelles, Bruxelles, Belgium
D. Beghin, B. Bilin, H. Brun, B. Clerbaux, G. De Lentdecker, H. Delannoy, B. Dorney, L. Favart, A. Grebenyuk, A.K. Kalsi, J. Luetic, A. Popov, N. Postiau, E. Starling, L. Thomas, C. Vander Velde, P. Vanlaer, D. Vannerom, Q. Wang

Ghent University, Ghent, Belgium
T. Cornelis, D. Dobur, I. Khvastunov[2], C. Roskas, D. Trocino, M. Tytgat, W. Verbeke, B. Vermassen, M. Vit, N. Zaganidis

Université Catholique de Louvain, Louvain-la-Neuve, Belgium
O. Bondu, G. Bruno, C. Caputo, P. David, C. Delaere, M. Delcourt, A. Giammanco, V. Lemaitre, A. Magitteri, J. Prisciandaro, A. Saggio, M. Vidal Marono, P. Vischia, J. Zobec

Centro Brasileiro de Pesquisas Fisicas, Rio de Janeiro, Brazil
F.L. Alves, G.A. Alves, G. Correia Silva, C. Hensel, A. Moraes, P. Rebello Teles

Universidade do Estado do Rio de Janeiro, Rio de Janeiro, Brazil
E. Belchior Batista Das Chagas, W. Carvalho, J. Chinellato[3], E. Coelho, E.M. Da Costa, G.G. Da Silveira[4], D. De Jesus Damiao, C. De Oliveira Martins, S. Fonseca De Souza, L.M. Huertas Guativa, H. Malbouisson, J. Martins[5], D. Matos Figueiredo, M. Medina Jaime[6], M. Melo De Almeida, C. Mora Herrera, L. Mundim, H. Nogima, W.L. Prado Da Silva, L.J. Sanchez Rosas, A. Santoro, A. Sznajder, M. Thiel, E.J. Tonelli Manganote[3], F. Torres Da Silva De Araujo, A. Vilela Pereira

Universidade Estadual Paulista [a], Universidade Federal do ABC [b], São Paulo, Brazil
S. Ahuja[a], C.A. Bernardes[a], L. Calligaris[a], T.R. Fernandez Perez Tomei[a], E.M. Gregores[b], D.S. Lemos, P.G. Mercadante[b], S.F. Novaes[a], SandraS. Padula[a]

Institute for Nuclear Research and Nuclear Energy, Bulgarian Academy of Sciences, Sofia, Bulgaria
A. Aleksandrov, G. Antchev, R. Hadjiiska, P. Iaydjiev, A. Marinov, M. Misheva, M. Rodozov, M. Shopova, G. Sultanov

University of Sofia, Sofia, Bulgaria
M. Bonchev, A. Dimitrov, T. Ivanov, L. Litov, B. Pavlov, P. Petkov

Beihang University, Beijing, China
W. Fang[7], X. Gao[7], L. Yuan

Institute of High Energy Physics, Beijing, China
M. Ahmad, G.M. Chen, H.S. Chen, M. Chen, C.H. Jiang, D. Leggat, H. Liao, Z. Liu, S.M. Shaheen[8], A. Spiezia, J. Tao, E. Yazgan, H. Zhang, S. Zhang[8], J. Zhao

State Key Laboratory of Nuclear Physics and Technology, Peking University, Beijing, China
A. Agapitos, Y. Ban, G. Chen, A. Levin, J. Li, L. Li, Q. Li, Y. Mao, S.J. Qian, D. Wang

Tsinghua University, Beijing, China
Z. Hu, Y. Wang

Universidad de Los Andes, Bogota, Colombia
C. Avila, A. Cabrera, L.F. Chaparro Sierra, C. Florez, C.F. González Hernández, M.A. Segura Delgado

Universidad de Antioquia, Medellin, Colombia
J. Mejia Guisao, J.D. Ruiz Alvarez, C.A. Salazar González, N. Vanegas Arbelaez

University of Split, Faculty of Electrical Engineering, Mechanical Engineering and Naval Architecture, Split, Croatia
D. Giljanović, N. Godinovic, D. Lelas, I. Puljak, T. Sculac

University of Split, Faculty of Science, Split, Croatia
Z. Antunovic, M. Kovac

Institute Rudjer Boskovic, Zagreb, Croatia
V. Brigljevic, S. Ceci, D. Ferencek, K. Kadija, B. Mesic, M. Roguljic, A. Starodumov[9], T. Susa

University of Cyprus, Nicosia, Cyprus
M.W. Ather, A. Attikis, E. Erodotou, A. Ioannou, M. Kolosova, S. Konstantinou, G. Mavromanolakis, J. Mousa, C. Nicolaou, F. Ptochos, P.A. Razis, H. Rykaczewski, D. Tsiakkouri

Charles University, Prague, Czech Republic
M. Finger[10], M. Finger Jr.[10], A. Kveton, J. Tomsa

Escuela Politecnica Nacional, Quito, Ecuador
E. Ayala

Universidad San Francisco de Quito, Quito, Ecuador
E. Carrera Jarrin

Academy of Scientific Research and Technology of the Arab Republic of Egypt, Egyptian Network of High Energy Physics, Cairo, Egypt
Y. Assran[11,12], S. Elgammal[12]

National Institute of Chemical Physics and Biophysics, Tallinn, Estonia
S. Bhowmik, A. Carvalho Antunes De Oliveira, R.K. Dewanjee, K. Ehataht, M. Kadastik, M. Raidal, C. Veelken

Department of Physics, University of Helsinki, Helsinki, Finland
P. Eerola, L. Forthomme, H. Kirschenmann, K. Osterberg, M. Voutilainen

Helsinki Institute of Physics, Helsinki, Finland
F. Garcia, J. Havukainen, J.K. Heikkilä, T. Järvinen, V. Karimäki, R. Kinnunen, T. Lampén, K. Lassila-Perini, S. Laurila, S. Lehti, T. Lindén, P. Luukka, T. Mäenpää, H. Siikonen, E. Tuominen, J. Tuominiemi

Lappeenranta University of Technology, Lappeenranta, Finland
T. Tuuva

IRFU, CEA, Université Paris-Saclay, Gif-sur-Yvette, France
P.A. Bausson, M. Besancon, D. Calvet, F. Couderc, M. Dejardin, D. Denegri, B. Fabbro, J.L. Faure, F. Ferri, S. Ganjour, A. Givernaud, P. Gras, F. Guilloux, G. Hamel de Monchenault, P. Jarry, C. Leloup, B. Lenzi, E. Locci, J. Malcles, I. Mandjavidze, J. Rander, A. Rosowsky, M.Ö. Sahin, A. Savoy-Navarro[13], M. Titov

Laboratoire Leprince-Ringuet, Ecole polytechnique, CNRS/IN2P3, Université Paris-Saclay, Palaiseau, France
C. Amendola, F. Beaudette, P. Busson, C. Charlot, B. Diab, G. Falmagne, R. Granier de Cassagnac, I. Kucher, A. Lobanov, C. Martin Perez, M. Nguyen, C. Ochando, P. Paganini, J. Rembser, R. Salerno, J.B. Sauvan, Y. Sirois, A. Zabi, A. Zghiche

Université de Strasbourg, CNRS, IPHC UMR 7178, Strasbourg, France
J.-L. Agram[14], J. Andrea, D. Bloch, G. Bourgatte, J.-M. Brom, E.C. Chabert, C. Collard, E. Conte[14], J.-C. Fontaine[14], D. Gelé, U. Goerlach, M. Jansová, A.-C. Le Bihan, N. Tonon, P. Van Hove

Centre de Calcul de l'Institut National de Physique Nucleaire et de Physique des Particules, CNRS/IN2P3, Villeurbanne, France
S. Gadrat

Université de Lyon, Université Claude Bernard Lyon 1, CNRS-IN2P3, Institut de Physique Nucléaire de Lyon, Villeurbanne, France
S. Beauceron, C. Bernet, G. Boudoul, C. Camen, N. Chanon, R. Chierici, D. Contardo, P. Depasse, H. El Mamouni, J. Fay, S. Gascon, M. Gouzevitch, B. Ille, Sa. Jain, F. Lagarde, I.B. Laktineh, H. Lattaud, M. Lethuillier, L. Mirabito, S. Perries, V. Sordini, G. Touquet, M. Vander Donckt, S. Viret

Georgian Technical University, Tbilisi, Georgia
A. Khvedelidze[10]

Tbilisi State University, Tbilisi, Georgia
Z. Tsamalaidze[10]

RWTH Aachen University, I. Physikalisches Institut, Aachen, Germany
C. Autermann, L. Feld, M.K. Kiesel, K. Klein, M. Lipinski, D. Meuser, A. Pauls, M. Preuten, M.P. Rauch, C. Schomakers, J. Schulz, M. Teroerde, B. Wittmer

RWTH Aachen University, III. Physikalisches Institut A, Aachen, Germany
A. Albert, M. Erdmann, S. Erdweg, T. Esch, B. Fischer, R. Fischer, S. Ghosh, T. Hebbeker, K. Hoepfner, H. Keller, L. Mastrolorenzo, M. Merschmeyer, A. Meyer, P. Millet, G. Mocellin, S. Mondal, S. Mukherjee, D. Noll, A. Novak, T. Pook, A. Pozdnyakov, T. Quast, M. Radziej, Y. Rath, H. Reithler, M. Rieger, J. Roemer, A. Schmidt, S.C. Schuler, A. Sharma, S. Thüer, S. Wiedenbeck

RWTH Aachen University, III. Physikalisches Institut B, Aachen, Germany
G. Flügge, W. Haj Ahmad[15], O. Hlushchenko, T. Kress, T. Müller, A. Nehrkorn, A. Nowack, C. Pistone, O. Pooth, D. Roy, H. Sert, A. Stahl[16]

Deutsches Elektronen-Synchrotron, Hamburg, Germany
M. Aldaya Martin, C. Asawatangtrakuldee, P. Asmuss, I. Babounikau, H. Bakhshiansohi, K. Beernaert, O. Behnke, U. Behrens, A. Bermúdez Martínez, D. Bertsche, A.A. Bin Anuar, K. Borras[17], V. Botta, A. Campbell, A. Cardini, P. Connor, S. Consuegra Rodríguez, C. Contreras-Campana, V. Danilov, A. De Wit, M.M. Defranchis, C. Diez Pardos, D. Domínguez Damiani, G. Eckerlin, D. Eckstein, T. Eichhorn, A. Elwood, E. Eren, E. Gallo[18], A. Geiser, J.M. Grados Luyando, A. Grohsjean, M. Guthoff, M. Haranko, A. Harb, A. Jafari, N.Z. Jomhari, H. Jung, A. Kasem[17], M. Kasemann, H. Kaveh, J. Keaveney, C. Kleinwort, J. Knolle, D. Krücker, W. Lange, T. Lenz, J. Leonard, J. Lidrych, K. Lipka, W. Lohmann[19], R. Mankel, I.-A. Melzer-Pellmann, A.B. Meyer, M. Meyer, M. Missiroli, G. Mittag, J. Mnich, A. Mussgiller, V. Myronenko, D. Pérez Adán, S.K. Pflitsch, D. Pitzl, A. Raspereza, A. Saibel, M. Savitskyi, V. Scheurer, P. Schütze, C. Schwanenberger, R. Shevchenko, A. Singh, H. Tholen, O. Turkot, A. Vagnerini, M. Van De Klundert, G.P. Van Onsem, R. Walsh, Y. Wen, K. Wichmann, C. Wissing, O. Zenaiev, R. Zlebcik

University of Hamburg, Hamburg, Germany
R. Aggleton, S. Bein, L. Benato, A. Benecke, V. Blobel, T. Dreyer, A. Ebrahimi, A. Fröhlich, C. Garbers, E. Garutti, D. Gonzalez, P. Gunnellini, J. Haller, A. Hinzmann, A. Karavdina, G. Kasieczka, R. Klanner, R. Kogler, N. Kovalchuk, S. Kurz, V. Kutzner, J. Lange, T. Lange, A. Malara, D. Marconi, J. Multhaup, M. Niedziela, C.E.N. Niemeyer, D. Nowatschin, A. Perieanu, A. Reimers, O. Rieger, C. Scharf, P. Schleper, S. Schumann, J. Schwandt, J. Sonneveld, H. Stadie, G. Steinbrück, F.M. Stober, M. Stöver, B. Vormwald, I. Zoi

Karlsruher Institut fuer Technologie, Karlsruhe, Germany
M. Akbiyik, C. Barth, M. Baselga, S. Baur, T. Berger, E. Butz, R. Caspart, T. Chwalek, W. De Boer, A. Dierlamm, K. El Morabit, N. Faltermann, M. Giffels, P. Goldenzweig, A. Gottmann, M.A. Harrendorf, F. Hartmann[16], U. Husemann, S. Kudella, S. Mitra, M.U. Mozer, Th. Müller, M. Musich, A. Nürnberg, G. Quast, K. Rabbertz, M. Schröder, I. Shvetsov, H.J. Simonis, R. Ulrich, M. Weber, C. Wöhrmann, R. Wolf

Institute of Nuclear and Particle Physics (INPP), NCSR Demokritos, Aghia Paraskevi, Greece
G. Anagnostou, P. Asenov, G. Daskalakis, T. Geralis, A. Kyriakis, D. Loukas, G. Paspalaki

National and Kapodistrian University of Athens, Athens, Greece
M. Diamantopoulou, G. Karathanasis, P. Kontaxakis, A. Panagiotou, I. Papavergou, N. Saoulidou, A. Stakia, K. Theofilatos, K. Vellidis

National Technical University of Athens, Athens, Greece
G. Bakas, K. Kousouris, I. Papakrivopoulos, G. Tsipolitis

University of Ioánnina, Ioánnina, Greece
I. Evangelou, C. Foudas, P. Gianneios, P. Katsoulis, P. Kokkas, S. Mallios, K. Manitara, N. Manthos, I. Papadopoulos, J. Strologas, F.A. Triantis, D. Tsitsonis

MTA-ELTE Lendület CMS Particle and Nuclear Physics Group, Eötvös Loránd University, Budapest, Hungary
M. Bartók[20], M. Csanad, P. Major, K. Mandal, A. Mehta, M.I. Nagy, G. Pasztor, O. Surányi, G.I. Veres

Wigner Research Centre for Physics, Budapest, Hungary
G. Bencze, C. Hajdu, D. Horvath[21], F. Sikler, T.. Vámi, V. Veszpremi, G. Vesztergombi†

Institute of Nuclear Research ATOMKI, Debrecen, Hungary
N. Beni, S. Czellar, J. Karancsi[20], A. Makovec, J. Molnar, Z. Szillasi

Institute of Physics, University of Debrecen, Debrecen, Hungary
P. Raics, D. Teyssier, Z.L. Trocsanyi, B. Ujvari

Eszterhazy Karoly University, Karoly Robert Campus, Gyongyos, Hungary
W.J. Metzger

Indian Institute of Science (IISc), Bangalore, India
S. Choudhury, J.R. Komaragiri, P.C. Tiwari

National Institute of Science Education and Research, HBNI, Bhubaneswar, India
S. Bahinipati[23], C. Kar, G. Kole, P. Mal, V.K. Muraleedharan Nair Bindhu, A. Nayak[24], D.K. Sahoo[23], S.K. Swain

Panjab University, Chandigarh, India
S. Bansal, S.B. Beri, V. Bhatnagar, S. Chauhan, R. Chawla, N. Dhingra, R. Gupta, A. Kaur, M. Kaur, S. Kaur, P. Kumari, M. Lohan, M. Meena, K. Sandeep, S. Sharma, J.B. Singh, A.K. Virdi

University of Delhi, Delhi, India
A. Bhardwaj, B.C. Choudhary, R.B. Garg, M. Gola, S. Keshri, Ashok Kumar, S. Malhotra, M. Naimuddin, P. Priyanka, K. Ranjan, Aashaq Shah, R. Sharma

Saha Institute of Nuclear Physics, HBNI, Kolkata, India
R. Bhardwaj[25], M. Bharti[25], R. Bhattacharya, S. Bhattacharya, U. Bhawandeep[25], D. Bhowmik, S. Dey, S. Dutta, S. Ghosh, M. Maity[26], K. Mondal, S. Nandan, A. Purohit, P.K. Rout, A. Roy, G. Saha, S. Sarkar, T. Sarkar[26], M. Sharan, B. Singh[25], S. Thakur[25]

Indian Institute of Technology Madras, Madras, India
P.K. Behera, P. Kalbhor, A. Muhammad, P.R. Pujahari, A. Sharma, A.K. Sikdar

Bhabha Atomic Research Centre, Mumbai, India
R. Chudasama, D. Dutta, V. Jha, V. Kumar, D.K. Mishra, P.K. Netrakanti, L.M. Pant, P. Shukla

Tata Institute of Fundamental Research-A, Mumbai, India
T. Aziz, M.A. Bhat, S. Dugad, G.B. Mohanty, N. Sur, RavindraKumar Verma

Tata Institute of Fundamental Research-B, Mumbai, India
S. Banerjee, S. Bhattacharya, S. Chatterjee, P. Das, M. Guchait, S. Karmakar, S. Kumar, G. Majumder, K. Mazumdar, N. Sahoo, S. Sawant

Indian Institute of Science Education and Research (IISER), Pune, India
S. Chauhan[27], S. Dube, V. Hegde, A. Kapoor, K. Kothekar, S. Pandey, A. Rane, A. Rastogi, S. Sharma

Institute for Research in Fundamental Sciences (IPM), Tehran, Iran
S. Chenarani[28], E. Eskandari Tadavani, S.M. Etesami[28], M. Khakzad, M. Mohammadi Najafabadi, M. Naseri, F. Rezaei Hosseinabadi

University College Dublin, Dublin, Ireland
M. Felcini, M. Grunewald

INFN Sezione di Bari [a], Università di Bari [b], Politecnico di Bari [c], Bari, Italy
M. Abbrescia[a,b], C. Calabria[a,b], A. Colaleo[a], D. Creanza[a,c], L. Cristella[a,b], N. De Filippis[a,c], M. De Palma[a,b], A. Di Florio[a,b], L. Fiore[a], A. Gelmi[a,b], G. Iaselli[a,c], M. Ince[a,b], S. Lezki[a,b], G. Maggi[a,c], M. Maggi[a], G. Miniello[a,b], S. My[a,b], S. Nuzzo[a,b], A. Pompili[a,b], G. Pugliese[a,c], R. Radogna[a], A. Ranieri[a], G. Selvaggi[a,b], L. Silvestris[a], R. Venditti[a], P. Verwilligen[a]

INFN Sezione di Bologna [a], Università di Bologna [b], Bologna, Italy
G. Abbiendi[a], C. Battilana[a,b], D. Bonacorsi[a,b], L. Borgonovi[a,b], S. Braibant-Giacomelli[a,b], R. Campanini[a,b], P. Capiluppi[a,b], A. Castro[a,b], F.R. Cavallo[a], C. Ciocca[a], G. Codispoti[a,b], M. Cuffiani[a,b], G.M. Dallavalle[a], F. Fabbri[a], A. Fanfani[a,b], E. Fontanesi, P. Giacomelli[a], C. Grandi[a], L. Guiducci[a,b], F. Iemmi[a,b], S. Lo Meo[a,29], S. Marcellini[a], G. Masetti[a], F.L. Navarria[a,b], A. Perrotta[a], F. Primavera[a,b], A.M. Rossi[a,b], T. Rovelli[a,b], G.P. Siroli[a,b], N. Tosi[a]

INFN Sezione di Catania [a], Università di Catania [b], Catania, Italy
S. Albergo[a,b,30], S. Costa[a,b], A. Di Mattia[a], R. Potenza[a,b], A. Tricomi[a,b,30], C. Tuve[a,b]

INFN Sezione di Firenze [a], Università di Firenze [b], Firenze, Italy
G. Barbagli[a], R. Ceccarelli, K. Chatterjee[a,b], V. Ciulli[a,b], C. Civinini[a], R. D'Alessandro[a,b], E. Focardi[a,b], G. Latino, P. Lenzi[a,b], M. Meschini[a], S. Paoletti[a], G. Sguazzoni[a], D. Strom[a], L. Viliani[a]

INFN Laboratori Nazionali di Frascati, Frascati, Italy
L. Benussi, S. Bianco, D. Piccolo

INFN Sezione di Genova [a], Università di Genova [b], Genova, Italy
M. Bozzo[a,b], F. Ferro[a], R. Mulargia[a,b], E. Robutti[a], S. Tosi[a,b]

INFN Sezione di Milano-Bicocca [a], Università di Milano-Bicocca [b], Milano, Italy
A. Benaglia[a], A. Beschi[a,b], F. Brivio[a,b], P. Carniti[a], V. Ciriolo[a,b,16], S. Di Guida[a,b,16], M.E. Dinardo[a,b], P. Dini[a], S. Fiorendi[a,b], S. Gennai[a], A. Ghezzi[a,b], C. Gotti[a], P. Govoni[a,b], S. Gundacker[a,b], L. Guzzi[a,b], M. Malberti[a], S. Malvezzi[a], R. Mazza[a], D. Menasce[a], F. Monti[a,b], L. Moroni[a], G. Ortona[a,b], M. Paganoni[a,b], D. Pedrini[a], G.E. Pessina[a], S. Ragazzi[a,b], N. Redaelli[a], T. Tabarelli de Fatis[a,b], D. Zuolo[a,b]

INFN Sezione di Napoli [a], Università di Napoli 'Federico II' [b], Napoli, Italy, Università della Basilicata [c], Potenza, Italy, Università G. Marconi [d], Roma, Italy
S. Buontempo[a], N. Cavallo[a,c], A. De Iorio[a,b], A. Di Crescenzo[a,b], F. Fabozzi[a,c], F. Fienga[a], G. Galati[a], A.O.M. Iorio[a,b], L. Lista[a,b], S. Meola[a,d,16], P. Paolucci[a,16], B. Rossi[a], C. Sciacca[a,b], E. Voevodina[a,b]

INFN Sezione di Padova [a], Università di Padova [b], Padova, Italy, Università di Trento [c], Trento, Italy
P. Azzi[a], N. Bacchetta[a], M. Benettoni[a], D. Bisello[a,b], A. Boletti[a,b], A. Bragagnolo, R. Carlin[a,b], P. Checchia[a], P. De Castro Manzano[a], T. Dorigo[a], U. Dosselli[a], F. Gasparini[a,b], U. Gasparini[a,b],

A. Gozzelino[a], S.Y. Hoh, P. Lujan, M. Margoni[a,b], A.T. Meneguzzo[a,b], J. Pazzini[a,b], M. Presilla[b], P. Ronchese[a,b], R. Rossin[a,b], F. Simonetto[a,b], A. Tiko, M. Tosi[a,b], M. Zanetti[a,b], P. Zotto[a,b], G. Zumerle[a,b]

INFN Sezione di Pavia [a], Università di Pavia [b], Pavia, Italy
A. Braghieri[a], P. Montagna[a,b], S.P. Ratti[a,b], V. Re[a], M. Ressegotti[a,b], C. Riccardi[a,b], P. Salvini[a], I. Vai[a,b], P. Vitulo[a,b]

INFN Sezione di Perugia [a], Università di Perugia [b], Perugia, Italy
M. Biasini[a,b], G.M. Bilei[a], C. Cecchi[a,b], D. Ciangottini[a,b], L. Fanò[a,b], P. Lariccia[a,b], R. Leonardi[a,b], E. Manoni[a], G. Mantovani[a,b], V. Mariani[a,b], M. Menichelli[a], A. Rossi[a,b], A. Santocchia[a,b], D. Spiga[a]

INFN Sezione di Pisa [a], Università di Pisa [b], Scuola Normale Superiore di Pisa [c], Pisa, Italy
K. Androsov[a], P. Azzurri[a], G. Bagliesi[a], V. Bertacchi[a,c], L. Bianchini[a], T. Boccali[a], R. Castaldi[a], M.A. Ciocci[a,b], R. Dell'Orso[a], G. Fedi[a], L. Giannini[a,c], A. Giassi[a], M.T. Grippo[a], F. Ligabue[a,c], E. Manca[a,c], G. Mandorli[a,c], A. Messineo[a,b], F. Palla[a], A. Rizzi[a,b], G. Rolandi[31], S. Roy Chowdhury, A. Scribano[a], P. Spagnolo[a], R. Tenchini[a], G. Tonelli[a,b], N. Turini, A. Venturi[a], P.G. Verdini[a]

INFN Sezione di Roma [a], Sapienza Università di Roma [b], Rome, Italy
F. Cavallari[a], M. Cipriani[a,b], D. Del Re[a,b], E. Di Marco[a], M. Diemoz[a], E. Longo[a,b], B. Marzocchi[a,b,32], P. Meridiani[a], G. Organtini[a,b], F. Pandolfi[a], R. Paramatti[a,b], C. Quaranta[a,b], S. Rahatlou[a,b], C. Rovelli[a], F. Santanastasio[a,b], L. Soffi[a], S. Torelli[a,b], R. Tramontano[a,b]

INFN Sezione di Torino [a], Università di Torino [b], Torino, Italy, Università del Piemonte Orientale [c], Novara, Italy
N. Amapane[a,b], R. Arcidiacono[a,c], S. Argiro[a,b], M. Arneodo[a,c], N. Bartosik[a], R. Bellan[a,b], C. Biino[a], A. Cappati[a,b], N. Cartiglia[a], S. Cometti[a], M. Costa[a,b], R. Covarelli[a,b], N. Demaria[a], F. Fausti, M. Ferrero, B. Kiani[a,b], M. Mandurrino, C. Mariotti[a], S. Maselli[a], E. Migliore[a,b], E. Monteil[a,b], M. Monteno[a], M.M. Obertino[a,b], E.J. Olave, L. Pacher[a,b], N. Pastrone[a], M. Pelliccioni[a], G.L. Pinna Angioni[a,b], M. Ruspa[a,c], R. Salvatico[a,b], F. Siviero, V. Sola[a], A. Solano[a,b], D. Soldi[a,b], A. Staiano[a]

INFN Sezione di Trieste [a], Università di Trieste [b], Trieste, Italy
S. Belforte[a], V. Candelise[a,b], M. Casarsa[a], F. Cossutti[a], A. Da Rold[a,b], G. Della Ricca[a,b], F. Vazzoler[a,b], A. Zanetti[a]

Kyungpook National University, Daegu, Korea
B. Kim, D.H. Kim, G.N. Kim, M.S. Kim, J. Lee, S.W. Lee, C.S. Moon, Y.D. Oh, S.I. Pak, S. Sekmen, D.C. Son, Y.C. Yang

Chonnam National University, Institute for Universe and Elementary Particles, Kwangju, Korea
H. Kim, D.H. Moon, G. Oh

Hanyang University, Seoul, Korea
B. Francois, T.J. Kim, J. Park

Korea University, Seoul, Korea
S. Cho, S. Choi, Y. Go, D. Gyun, S. Ha, B. Hong, K. Lee, K.S. Lee, J. Lim, J. Park, S.K. Park, Y. Roh

Kyung Hee University, Department of Physics
J. Goh

Sejong University, Seoul, Korea
H.S. Kim

Seoul National University, Seoul, Korea
J. Almond, J.H. Bhyun, J. Choi, S. Jeon, J. Kim, J.S. Kim, H. Lee, K. Lee, S. Lee, K. Nam, M. Oh, S.B. Oh, B.C. Radburn-Smith, U.K. Yang, H.D. Yoo, I. Yoon, G.B. Yu

University of Seoul, Seoul, Korea
D. Jeon, H. Kim, J.H. Kim, J.S.H. Lee, I.C. Park, I. Watson

Sungkyunkwan University, Suwon, Korea
Y. Choi, C. Hwang, Y. Jeong, J. Lee, Y. Lee, I. Yu

Riga Technical University, Riga, Latvia
V. Veckalns[33]

Vilnius University, Vilnius, Lithuania
V. Dudenas, A. Juodagalvis, S. Nargelas, G. Tamulaitis, J. Vaitkus

National Centre for Particle Physics, Universiti Malaya, Kuala Lumpur, Malaysia
Z.A. Ibrahim, F. Mohamad Idris[34], W.A.T. Wan Abdullah, M.N. Yusli, Z. Zolkapli

Universidad de Sonora (UNISON), Hermosillo, Mexico
J.F. Benitez, A. Castaneda Hernandez, J.A. Murillo Quijada, L. Valencia Palomo

Centro de Investigacion y de Estudios Avanzados del IPN, Mexico City, Mexico
H. Castilla-Valdez, E. De La Cruz-Burelo, I. Heredia-De La Cruz[35], R. Lopez-Fernandez, A. Sanchez-Hernandez

Universidad Iberoamericana, Mexico City, Mexico
S. Carrillo Moreno, C. Oropeza Barrera, M. Ramirez-Garcia, F. Vazquez Valencia

Benemerita Universidad Autonoma de Puebla, Puebla, Mexico
J. Eysermans, I. Pedraza, H.A. Salazar Ibarguen, C. Uribe Estrada

Universidad Autónoma de San Luis Potosí, San Luis Potosí, Mexico
A. Morelos Pineda

University of Montenegro, Podgorica, Montenegro
N. Raicevic

University of Auckland, Auckland, New Zealand
D. Krofcheck

University of Canterbury, Christchurch, New Zealand
S. Bheesette, P.H. Butler

National Centre for Physics, Quaid-I-Azam University, Islamabad, Pakistan
A. Ahmad, M. Ahmad, Q. Hassan, H.R. Hoorani, W.A. Khan, M.A. Shah, M. Shoaib, M. Waqas

AGH University of Science and Technology Faculty of Computer Science, Electronics and Telecommunications, Krakow, Poland
V. Avati, L. Grzanka, M. Malawski

National Centre for Nuclear Research, Swierk, Poland
H. Bialkowska, M. Bluj, B. Boimska, M. Górski, M. Kazana, M. Szleper, P. Zalewski

Institute of Experimental Physics, Faculty of Physics, University of Warsaw, Warsaw, Poland
K. Bunkowski, A. Byszuk[36], K. Doroba, A. Kalinowski, M. Konecki, J. Krolikowski, M. Misiura, M. Olszewski, A. Pyskir, M. Walczak

Laboratório de Instrumentação e Física Experimental de Partículas, Lisboa, Portugal
M. Araujo, P. Bargassa, D. Bastos, A. Di Francesco, P. Faccioli, B. Galinhas, M. Gallinaro, J. Hollar, N. Leonardo, T. Niknejad, J. Seixas, K. Shchelina, J.C. Da Silva, G. Strong, O. Toldaiev, J. Varela

Joint Institute for Nuclear Research, Dubna, Russia
S. Afanasiev, P. Bunin, M. Gavrilenko, I. Golutvin, I. Gorbunov, A. Kamenev, V. Karjavine, A. Lanev, A. Malakhov, V. Matveev[37,38], P. Moisenz, V. Palichik, V. Perelygin, M. Savina, S. Shmatov, S. Shulha, N. Skatchkov, V. Smirnov, N. Voytishin, A. Zarubin

Petersburg Nuclear Physics Institute, Gatchina (St. Petersburg), Russia
L. Chtchipounov, V. Golovtsov, Y. Ivanov, V. Kim[39], E. Kuznetsova[40], P. Levchenko, V. Murzin, V. Oreshkin, I. Smirnov, D. Sosnov, V. Sulimov, L. Uvarov, A. Vorobyev

Institute for Nuclear Research, Moscow, Russia
Yu. Andreev, A. Dermenev, S. Gninenko, N. Golubev, A. Karneyeu, M. Kirsanov, N. Krasnikov, A. Pashenkov, D. Tlisov, A. Toropin

Institute for Theoretical and Experimental Physics named by A.I. Alikhanov of NRC 'Kurchatov Institute', Moscow, Russia
V. Epshteyn, V. Gavrilov, N. Lychkovskaya, A. Nikitenko[41], V. Popov, I. Pozdnyakov, G. Safronov, A. Spiridonov, A. Stepennov, M. Toms, E. Vlasov, A. Zhokin

Moscow Institute of Physics and Technology, Moscow, Russia
T. Aushev

National Research Nuclear University 'Moscow Engineering Physics Institute' (MEPhI), Moscow, Russia
M. Danilov[42], P. Parygin, S. Polikarpov[42], E. Popova, V. Rusinov

P.N. Lebedev Physical Institute, Moscow, Russia
V. Andreev, M. Azarkin, I. Dremin, M. Kirakosyan, A. Terkulov

Skobeltsyn Institute of Nuclear Physics, Lomonosov Moscow State University, Moscow, Russia
A. Belyaev, E. Boos, M. Dubinin[43], L. Dudko, A. Ershov, A. Gribushin, V. Klyukhin, O. Kodolova, I. Lokhtin, S. Obraztsov, S. Petrushanko, V. Savrin, A. Snigirev

Novosibirsk State University (NSU), Novosibirsk, Russia
A. Barnyakov[44], V. Blinov[44], T. Dimova[44], L. Kardapoltsev[44], Y. Skovpen[44]

Institute for High Energy Physics of National Research Centre 'Kurchatov Institute', Protvino, Russia
I. Azhgirey, I. Bayshev, S. Bitioukov, V. Kachanov, D. Konstantinov, P. Mandrik, V. Petrov, R. Ryutin, S. Slabospitskii, A. Sobol, S. Troshin, N. Tyurin, A. Uzunian, A. Volkov

National Research Tomsk Polytechnic University, Tomsk, Russia
A. Babaev, A. Iuzhakov, V. Okhotnikov

Tomsk State University, Tomsk, Russia
V. Borchsh, V. Ivanchenko, E. Tcherniaev

University of Belgrade: Faculty of Physics and VINCA Institute of Nuclear Sciences
P. Adzic[45], P. Cirkovic, D. Devetak, M. Dordevic, P. Milenovic, J. Milosevic, M. Stojanovic

Centro de Investigaciones Energéticas Medioambientales y Tecnológicas (CIEMAT), Madrid, Spain
M. Aguilar-Benitez, J. Alcaraz Maestre, A. lvarez Fernández, I. Bachiller, M. Barrio Luna, J.A. Brochero Cifuentes, C.A. Carrillo Montoya, M. Cepeda, M. Cerrada, N. Colino, B. De La Cruz, A. Delgado Peris, C. Fernandez Bedoya, J.P. Fernández Ramos, J. Flix, M.C. Fouz, O. Gonzalez Lopez, S. Goy Lopez, J.M. Hernandez, M.I. Josa, D. Moran, . Navarro Tobar, A. Pérez-Calero Yzquierdo, J. Puerta Pelayo, I. Redondo, L. Romero, S. Sánchez Navas, M.S. Soares, A. Triossi, C. Willmott

Universidad Autónoma de Madrid, Madrid, Spain
C. Albajar, J.F. de Trocóniz

Universidad de Oviedo, Instituto Universitario de Ciencias y Tecnologías Espaciales de Asturias (ICTEA), Oviedo, Spain
B. Alvarez Gonzalez, J. Cuevas, C. Erice, J. Fernandez Menendez, S. Folgueras, I. Gonzalez Caballero, J.R. González Fernández, E. Palencia Cortezon, V. Rodríguez Bouza, S. Sanchez Cruz

Instituto de Física de Cantabria (IFCA), CSIC-Universidad de Cantabria, Santander, Spain
I.J. Cabrillo, A. Calderon, B. Chazin Quero, J. Duarte Campderros, M. Fernandez, C. Fernandez Madrazo, P.J. Fernández Manteca, A. García Alonso, G. Gomez, C. Martinez Rivero, P. Martinez Ruiz del Arbol, F. Matorras, J. Piedra Gomez, C. Prieels, T. Rodrigo, A. Ruiz-Jimeno, L. Russo[46], L. Scodellaro, N. Trevisani, I. Vila, J.M. Vizan Garcia

University of Colombo, Colombo, Sri Lanka
K. Malagalage

University of Ruhuna, Department of Physics, Matara, Sri Lanka
W.G.D. Dharmaratna, N. Wickramage

CERN, European Organization for Nuclear Research, Geneva, Switzerland
D. Abbaneo, B. Akgun, E. Auffray, G. Auzinger, J. Baechler, P. Baillon, A.H. Ball, D. Barney, J. Bendavid, M. Bianco, A. Bocci, E. Bossini, C. Botta, E. Brondolin, T. Camporesi, A. Caratelli, G. Cerminara, E. Chapon, A. Charkiewicz, G. Cucciati, D. d'Enterria, A. Dabrowski, N. Daci, V. Daponte, A. David, O. Davignon, A. De Roeck, N. Deelen, M. Deile, M. Dobson, M. Dünser, N. Dupont, A. Elliott-Peisert, F. Fallavollita[47], D. Fasanella, G. Franzoni, J. Fulcher, W. Funk, S. Giani, D. Gigi, A. Gilbert, K. Gill, F. Glege, M. Gruchala, M. Guilbaud, D. Gulhan, J. Hegeman, C. Heidegger, Y. Iiyama, V. Innocente, P. Janot, O. Karacheban[19], J. Kaspar, J. Kieseler, N. Koss, M. Krammer[1], C. Lange, P. Lecoq, C. Lourenço, L. Malgeri, M. Mannelli, A. Massironi, F. Meijers, J.A. Merlin, S. Mersi, E. Meschi, F. Moortgat, M. Mulders, J. Ngadiuba, S. Nourbakhsh, S. Orfanelli, L. Orsini, F. Pantaleo[16], L. Pape, E. Perez, M. Peruzzi, A. Petrilli, G. Petrucciani, A. Pfeiffer, M. Pierini, F.M. Pitters, D. Rabady, A. Racz, M. Rovere, H. Sakulin, C. Schäfer, C. Schwick, M. Selvaggi, A. Sharma, P. Silva, W. Snoeys, P. Sphicas[48], J. Steggemann, V.R. Tavolaro, D. Treille, A. Tsirou, A. Vartak, M. Verzetti, W.D. Zeuner

Paul Scherrer Institut, Villigen, Switzerland
L. Caminada[49], K. Deiters, W. Erdmann, R. Horisberger, Q. Ingram, H.C. Kaestli, D. Kotlinski, U. Langenegger, T. Rohe, S.A. Wiederkehr

ETH Zurich - Institute for Particle Physics and Astrophysics (IPA), Zurich, Switzerland
M. Backhaus, P. Berger, A. Calandri, N. Chernyavskaya, G. Dissertori, M. Dittmar, M. Donegà,

C. Dorfer, M. Dröge, T. Gadek, T.A. Gómez Espinosa, C. Grab, C. Haller, D. Hits, T. Klijnsma, W. Lustermann, R.A. Manzoni, M. Marionneau, M.T. Meinhard, F. Micheli, P. Musella, F. Nessi-Tedaldi, F. Pauss, G. Perrin, L. Perrozzi, S. Pigazzini, M. Reichmann, C. Reissel, T. Reitenspiess, D. Ruini, D.A. Sanz Becerra, M. Schönenberger, L. Shchutska, M.L. Vesterbacka Olsson, R. Wallny, D.H. Zhu

Universität Zürich, Zurich, Switzerland
T.K. Aarrestad, C. Amsler[50], D. Brzhechko, M.F. Canelli, A. De Cosa, R. Del Burgo, S. Donato, B. Kilminster, S. Leontsinis, V.M. Mikuni, I. Neutelings, G. Rauco, P. Robmann, D. Salerno, K. Schweiger, C. Seitz, Y. Takahashi, S. Wertz, A. Zucchetta

National Central University, Chung-Li, Taiwan
T.H. Doan, C.M. Kuo, W. Lin, S.S. Yu

National Taiwan University (NTU), Taipei, Taiwan
P. Chang, Y. Chao, K.F. Chen, P.H. Chen, W.-S. Hou, Y.y. Li, R.-S. Lu, E. Paganis, A. Psallidas, A. Steen

Chulalongkorn University, Faculty of Science, Department of Physics, Bangkok, Thailand
B. Asavapibhop, N. Srimanobhas, N. Suwonjandee

ukurova University, Physics Department, Science and Art Faculty, Adana, Turkey
A. Bat, F. Boran, S. Cerci[51], S. Damarseckin[52], Z.S. Demiroglu, F. Dolek, C. Dozen, I. Dumanoglu, G. Gokbulut, EmineGurpinar Guler[53], Y. Guler, I. Hos[54], C. Isik, E.E. Kangal[55], O. Kara, A. Kayis Topaksu, U. Kiminsu, M. Oglakci, G. Onengut, K. Ozdemir[56], S. Ozturk[57], A.E. Simsek, D. Sunar Cerci[51], U.G. Tok, S. Turkcapar, I.S. Zorbakir, C. Zorbilmez

Middle East Technical University, Physics Department, Ankara, Turkey
B. Isildak[58], G. Karapinar[59], M. Yalvac

Bogazici University, Istanbul, Turkey
I.O. Atakisi, E. Gülmez, M. Kaya[60], O. Kaya[61], B. Kaynak, Ö. Özçelik, S. Ozkorucuklu[62], S. Tekten, E.A. Yetkin[63]

Istanbul Technical University, Istanbul, Turkey
A. Cakir, K. Cankocak, Y. Komurcu, S. Sen[64]

Institute for Scintillation Materials of National Academy of Science of Ukraine, Kharkov, Ukraine
B. Grynyov

National Scientific Center, Kharkov Institute of Physics and Technology, Kharkov, Ukraine
L. Levchuk

University of Bristol, Bristol, United Kingdom
F. Ball, E. Bhal, S. Bologna, J.J. Brooke, D. Burns, E. Clement, D. Cussans, H. Flacher, J. Goldstein, G.P. Heath, H.F. Heath, L. Kreczko, S. Paramesvaran, B. Penning, T. Sakuma, S. Seif El Nasr-Storey, D. Smith, V.J. Smith, J. Taylor, A. Titterton

Rutherford Appleton Laboratory, Didcot, United Kingdom
K.W. Bell, A. Belyaev[65], C. Brew, R.M. Brown, D. Cieri, D.J.A. Cockerill, J.A. Coughlan, K. Harder, S. Harper, J. Linacre, K. Manolopoulos, D.M. Newbold, E. Olaiya, D. Petyt, T. Reis, T. Schuh, C.H. Shepherd-Themistocleous, A. Thea, I.R. Tomalin, T. Williams, W.J. Womersley

Imperial College, London, United Kingdom
R. Bainbridge, P. Bloch, J. Borg, S. Breeze, O. Buchmuller, A. Bundock, GurpreetSingh CHAHAL[66], D. Colling, P. Dauncey, G. Davies, M. Della Negra, R. Di Maria, P. Everaerts, G. Hall, G. Iles, T. James, M. Komm, C. Laner, L. Lyons, A.-M. Magnan, S. Malik, A. Martelli, V. Milosevic, J. Nash[67], V. Palladino, M. Pesaresi, D.M. Raymond, A. Richards, A. Rose, E. Scott, C. Seez, A. Shtipliyski, M. Stoye, T. Strebler, S. Summers, A. Tapper, K. Uchida, T. Virdee[16], N. Wardle, D. Winterbottom, J. Wright, A.G. Zecchinelli, S.C. Zenz

Brunel University, Uxbridge, United Kingdom
J.E. Cole, P.R. Hobson, A. Khan, P. Kyberd, C.K. Mackay, A. Morton, I.D. Reid, L. Teodorescu, S. Zahid

Baylor University, Waco, USA
K. Call, J. Dittmann, K. Hatakeyama, C. Madrid, B. McMaster, N. Pastika, C. Smith

Catholic University of America, Washington, DC, USA
R. Bartek, A. Dominguez, R. Uniyal

The University of Alabama, Tuscaloosa, USA
A. Buccilli, S.I. Cooper, C. Henderson, P. Rumerio, C. West

Boston University, Boston, USA
D. Arcaro, T. Bose, Z. Demiragli, D. Gastler, S. Girgis, D. Pinna, C. Richardson, J. Rohlf, D. Sperka, I. Suarez, L. Sulak, D. Zou

Brown University, Providence, USA
G. Benelli, B. Burkle, X. Coubez, D. Cutts, Y.t. Duh, M. Hadley, U. Heintz, J.M. Hogan[68], K.H.M. Kwok, E. Laird, G. Landsberg, J. Lee, Z. Mao, M. Narain, S. Sagir[69], R. Syarif, E. Usai, D. Yu

University of California, Davis, Davis, USA
R. Band, C. Brainerd, R. Breedon, M. Calderon De La Barca Sanchez, M. Chertok, J. Conway, R. Conway, P.T. Cox, R. Erbacher, C. Flores, G. Funk, F. Jensen, W. Ko, O. Kukral, R. Lander, M. Mulhearn, D. Pellett, J. Pilot, M. Shi, D. Stolp, D. Taylor, K. Tos, M. Tripathi, Z. Wang, F. Zhang

University of California, Los Angeles, USA
M. Bachtis, C. Bravo, R. Cousins, A. Dasgupta, A. Florent, J. Hauser, M. Ignatenko, N. Mccoll, W.A. Nash, S. Regnard, D. Saltzberg, C. Schnaible, B. Stone, V. Valuev

University of California, Riverside, Riverside, USA
K. Burt, R. Clare, J.W. Gary, S.M.A. Ghiasi Shirazi, G. Hanson, G. Karapostoli, E. Kennedy, O.R. Long, M. Olmedo Negrete, M.I. Paneva, W. Si, L. Wang, H. Wei, S. Wimpenny, B.R. Yates, Y. Zhang

University of California, San Diego, La Jolla, USA
J.G. Branson, P. Chang, S. Cittolin, M. Derdzinski, R. Gerosa, D. Gilbert, B. Hashemi, D. Klein, V. Krutelyov, J. Letts, M. Masciovecchio, S. May, S. Padhi, M. Pieri, V. Sharma, M. Tadel, F. Würthwein, A. Yagil, G. Zevi Della Porta

University of California, Santa Barbara - Department of Physics, Santa Barbara, USA
N. Amin, R. Bhandari, C. Campagnari, M. Citron, V. Dutta, M. Franco Sevilla, L. Gouskos, J. Incandela, B. Marsh, H. Mei, A. Ovcharova, H. Qu, J. Richman, U. Sarica, J. Sheplock, D. Stuart, S. Wang, J. Yoo

California Institute of Technology, Pasadena, USA
R. Abbott, D. Anderson, A. Bornheim, O. Cerri, I. Dutta, M. Gardner, D. Gawerc, K. Huang, H. Kim, J.M. Lawhorn, S. Li, N. Lu, A. Mangu, J. Mao, L.L. Narváez, H.B. Newman, T.Q. Nguyen, J. Pata, M. Spiropulu, Y. Tang, J. Trevor, J.R. Vlimant, C. Wang, S. Xie, A.J. Yankelevich, L. Zhang, Z. Zhang, R.Y. Zhu

Carnegie Mellon University, Pittsburgh, USA
M.B. Andrews, T. Ferguson, T. Mudholkar, M. Paulini, M. Sun, I. Vorobiev, M. Weinberg

University of Colorado Boulder, Boulder, USA
J.P. Cumalat, W.T. Ford, A. Johnson, E. MacDonald, T. Mulholland, R. Patel, A. Perloff, K. Stenson, K.A. Ulmer, S.R. Wagner

Cornell University, Ithaca, USA
J. Alexander, J. Chaves, Y. Cheng, J. Chu, A. Datta, A. Frankenthal, K. Mcdermott, N. Mirman, J.R. Patterson, D. Quach, A. Rinkevicius[70], A. Ryd, S.M. Tan, Z. Tao, J. Thom, P. Wittich, M. Zientek

Fairfield University, Fairfield, USA
D. Winn

Fermi National Accelerator Laboratory, Batavia, USA
S. Abdullin, M. Albrow, M. Alyari, G. Apollinari, A. Apresyan, A. Apyan, S. Banerjee, L.A.T. Bauerdick, A. Beretvas, J. Berryhill, P.C. Bhat, D. Braga, K. Burkett, J.N. Butler, A. Canepa, G.B. Cerati, H.W.K. Cheung, F. Chlebana, M. Cremonesi, G. Derylo, J. Duarte, V.D. Elvira, J. Freeman, Z. Gecse, E. Gottschalk, L. Gray, D. Gong[78], D. Green, P. Gui[78], S. Grünendahl, O. Gutsche, AllisonReinsvold Hall, J. Hanlon, R.M. Harris, S. Hasegawa, R. Heller, J. Hirschauer, M. Hussain, B. Jayatilaka, S. Jindariani, M. Johnson, U. Joshi, B. Klima, M.J. Kortelainen, B. Kreis, S. Lammel, J. Lewis, D. Lincoln, R. Lipton, M. Liu, T. Liu, T. Liu[78], S. Los, J. Lykken, K. Maeshima, J.M. Marraffino, D. Mason, P. McBride, P. Merkel, S. Miryala, S. Mrenna, S. Nahn, V. O'Dell, J. Olsen, V. Papadimitriou, K. Pedro, C. Pena[43], G. Rakness, F. Ravera, L. Ristori, B. Schneider, E. Sexton-Kennedy, N. Smith, A. Soha, W.J. Spalding, L. Spiegel, S. Stoynev, J. Strait, N. Strobbe, Q. Sun, Z. Tang, L. Taylor, S. Tkaczyk, N.V. Tran, L. Uplegger, E.W. Vaandering, C. Vernieri, M. Verzocchi, R. Vidal, E. Voirin, M. Wang, H.A. Weber, X. Wen[78], J. Wu

University of Florida, Gainesville, USA
D. Acosta, P. Avery, P. Bortignon, D. Bourilkov, A. Brinkerhoff, L. Cadamuro, A. Carnes, V. Cherepanov, D. Curry, F. Errico, R.D. Field, S.V. Gleyzer, B.M. Joshi, M. Kim, J. Konigsberg, A. Korytov, K.H. Lo, P. Ma, K. Matchev, N. Menendez, G. Mitselmakher, D. Rosenzweig, K. Shi, J. Wang, S. Wang, X. Zuo

Florida International University, Miami, USA
Y.R. Joshi

Florida State University, Tallahassee, USA
T. Adams, A. Askew, S. Hagopian, V. Hagopian, K.F. Johnson, R. Khurana, T. Kolberg, G. Martinez, T. Perry, H. Prosper, C. Schiber, R. Yohay, J. Zhang

Florida Institute of Technology, Melbourne, USA
M.M. Baarmand, V. Bhopatkar, M. Hohlmann, D. Noonan, M. Rahmani, M. Saunders, F. Yumiceva

University of Illinois at Chicago (UIC), Chicago, USA
M.R. Adams, L. Apanasevich, D. Berry, R.R. Betts, R. Cavanaugh, X. Chen, S. Dittmer, O. Evdokimov, C.E. Gerber, D.A. Hangal, D.J. Hofman, K. Jung, C. Mills, T. Roy, M.B. Tonjes, N. Varelas, H. Wang, X. Wang, Z. Wu

The University of Iowa, Iowa City, USA
M. Alhusseini, B. Bilki[53], W. Clarida, P. Debbins, K. Dilsiz[71], S. Durgut, R.P. Gandrajula, M. Haytmyradov, V. Khristenko, O.K. Köseyan, J.-P. Merlo, A. Mestvirishvili[72], A. Moeller, J. Nachtman, H. Ogul[73], Y. Onel, F. Ozok[74], A. Penzo, I. Schmidt, C. Snyder, E. Tiras, J. Wetzel

Johns Hopkins University, Baltimore, USA
B. Blumenfeld, A. Cocoros, N. Eminizer, D. Fehling, L. Feng, A.V. Gritsan, W.T. Hung, P. Maksimovic, J. Roskes, M. Swartz, M. Xiao

The University of Kansas, Lawrence, USA
A. Abreu, C. Baldenegro Barrera, P. Baringer, A. Bean, S. Boren, J. Bowen, A. Bylinkin, T. Isidori, S. Khalil, J. King, G. Krintiras, A. Kropivnitskaya, M. Lazarovits, C. Lindsey, D. Majumder, W. Mcbrayer, N. Minafra, M. Murray, C. Rogan, C. Royon, S. Sanders, E. Schmitz, J.D. Tapia Takaki, Q. Wang, J. Williams, G. Wilson

Kansas State University, Manhattan, USA
S. Duric, A. Ivanov, K. Kaadze, D. Kim, Y. Maravin, D.R. Mendis, T. Mitchell, A. Modak, A. Mohammadi, R. Taylor

Lawrence Livermore National Laboratory, Livermore, USA
F. Rebassoo, D. Wright

University of Maryland, College Park, USA
A. Baden, O. Baron, A. Belloni, S.C. Eno, Y. Feng, N.J. Hadley, S. Jabeen, G.Y. Jeng, R.G. Kellogg, J. Kunkle, A.C. Mignerey, S. Nabili, F. Ricci-Tam, M. Seidel, Y.H. Shin, A. Skuja, S.C. Tonwar, K. Wong

Massachusetts Institute of Technology, Cambridge, USA
D. Abercrombie, B. Allen, A. Baty, R. Bi, S. Brandt, W. Busza, I.A. Cali, M. D'Alfonso, G. Gomez Ceballos, M. Goncharov, P. Harris, D. Hsu, M. Hu, M. Klute, D. Kovalskyi, Y.-J. Lee, P.D. Luckey, B. Maier, A.C. Marini, C. Mcginn, C. Mironov, S. Narayanan, X. Niu, C. Paus, D. Rankin, C. Roland, G. Roland, Z. Shi, G.S.F. Stephans, K. Sumorok, K. Tatar, D. Velicanu, J. Wang, T.W. Wang, B. Wyslouch

University of Minnesota, Minneapolis, USA
A.C. Benvenuti[†], R.M. Chatterjee, A. Evans, S. Guts, P. Hansen, J. Hiltbrand, Sh. Jain, S. Kalafut, Y. Kubota, Z. Lesko, J. Mans, R. Rusack, M.A. Wadud

University of Mississippi, Oxford, USA
J.G. Acosta, S. Oliveros

University of Nebraska-Lincoln, Lincoln, USA
K. Bloom, S. Chauhan, D.R. Claes, C. Fangmeier, L. Finco, F. Golf, R. Gonzalez Suarez, R. Kamalieddin, I. Kravchenko, K. Kremke, F. Marcia, R. Rhynalds, J.E. Siado, G.R. Snow, B. Stieger

State University of New York at Buffalo, Buffalo, USA
G. Agarwal, C. Harrington, I. Iashvili, A. Kharchilava, C. Mclean, D. Nguyen, A. Parker, J. Pekkanen, S. Rappoccio, B. Roozbahani

Northeastern University, Boston, USA
G. Alverson, E. Barberis, C. Freer, Y. Haddad, A. Hortiangtham, G. Madigan, D.M. Morse, T. Orimoto, L. Skinnari, A. Tishelman-Charny, T. Wamorkar, B. Wang, A. Wisecarver, D. Wood

Northwestern University, Evanston, USA
S. Bhattacharya, J. Bueghly, T. Gunter, K.A. Hahn, N. Odell, M.H. Schmitt, K. Sung, M. Trovato, M. Velasco

University of Notre Dame, Notre Dame, USA
R. Bucci, N. Dev, R. Goldouzian, A.H. Heering, M. Hildreth, K. Hurtado Anampa, C. Jessop, D.J. Karmgard, K. Lannon, W. Li, N. Loukas, N. Marinelli, I. Mcalister, F. Meng, C. Mueller, Y. Musienko[37], M. Planer, R. Ruchti, P. Siddireddy, G. Smith, S. Taroni, M. Wayne, A. Wightman, M. Wolf, A. Woodard

The Ohio State University, Columbus, USA
J. Alimena, B. Bylsma, L.S. Durkin, S. Flowers, B. Francis, C. Hill, W. Ji, A. Lefeld, T.Y. Ling, B.L. Winer

Princeton University, Princeton, USA
S. Cooperstein, G. Dezoort, P. Elmer, J. Hardenbrook, N. Haubrich, S. Higginbotham, A. Kalogeropoulos, G. Kopp, S. Kwan, D. Lange, M.T. Lucchini, J. Luo, D. Marlow, K. Mei, I. Ojalvo, J. Olsen, N. Ozdowski, C. Palmer, P. Piroué, J. Salfeld-Nebgen, D. Stickland, C. Tully, Z. Wang, Y.X. Zhang

University of Puerto Rico, Mayaguez, USA
S. Malik, S. Norberg

Purdue University, West Lafayette, USA
A. Barker, V.E. Barnes, S. Das, L. Gutay, M. Jones, A.W. Jung, A. Khatiwada, B. Mahakud, D.H. Miller, G. Negro, N. Neumeister, C.C. Peng, S. Piperov, H. Qiu, J.F. Schulte, J. Sun, F. Wang, R. Xiao, W. Xie

Purdue University Northwest, Hammond, USA
T. Cheng, J. Dolen, N. Parashar

Rice University, Houston, USA
K.M. Ecklund, S. Freed, F.J.M. Geurts, M. Kilpatrick, Arun Kumar, W. Li, B.P. Padley, R. Redjimi, J. Roberts, J. Rorie, W. Shi, A.G. Stahl Leiton, Z. Tu, A. Zhang

University of Rochester, Rochester, USA
A. Bodek, P. de Barbaro, R. Demina, J.L. Dulemba, C. Fallon, T. Ferbel, M. Galanti, A. Garcia-Bellido, J. Han, O. Hindrichs, A. Khukhunaishvili, E. Ranken, P. Tan, R. Taus

Rutgers, The State University of New Jersey, Piscataway, USA
B. Chiarito, J.P. Chou, A. Gandrakota, Y. Gershtein, E. Halkiadakis, A. Hart, M. Heindl, E. Hughes, S. Kaplan, S. Kyriacou, I. Laflotte, A. Lath, R. Montalvo, K. Nash, M. Osherson, H. Saka, S. Salur, S. Schnetzer, D. Sheffield, S. Somalwar, R. Stone, S. Thomas, P. Thomassen

University of Tennessee, Knoxville, USA
H. Acharya, A.G. Delannoy, J. Heideman, G. Riley, S. Spanier

Texas A&M University, College Station, USA
O. Bouhali[75], A. Celik, M. Dalchenko, M. De Mattia, A. Delgado, S. Dildick, R. Eusebi, J. Gilmore, T. Huang, T. Kamon[76], S. Luo, D. Marley, R. Mueller, D. Overton, L. Perniè, D. Rathjens, A. Safonov

Texas Tech University, Lubbock, USA
N. Akchurin, J. Damgov, F. De Guio, S. Kunori, K. Lamichhane, S.W. Lee, T. Mengke, S. Muthumuni, T. Peltola, S. Undleeb, I. Volobouev, Z. Wang, A. Whitbeck

Vanderbilt University, Nashville, USA
S. Greene, A. Gurrola, R. Janjam, W. Johns, C. Maguire, A. Melo, H. Ni, K. Padeken, F. Romeo, P. Sheldon, S. Tuo, J. Velkovska, M. Verweij

University of Virginia, Charlottesville, USA
M.W. Arenton, P. Barria, B. Cox, G. Cummings, S. Goadhouse, J. Hakala, R. Hirosky, T. Hurt, M. Joyce, A. Ledovskoy, C. Neu, B. Tannenwald, Y. Wang, S. White, E. Wolfe, F. Xia

Wayne State University, Detroit, USA
R. Harr, P.E. Karchin, N. Poudyal, J. Sturdy, P. Thapa, S. Zaleski

University of Wisconsin - Madison, Madison, WI, USA
J. Buchanan, C. Caillol, D. Carlsmith, S. Dasu, I. De Bruyn, L. Dodd, F. Fiori, C. Galloni, B. Gomber[77], M. Herndon, A. Hervé, U. Hussain, P. Klabbers, A. Lanaro, A. Loeliger, K. Long, R. Loveless, J. Madhusudanan Sreekala, T. Ruggles, A. Savin, V. Sharma, W.H. Smith, D. Teague, S. Trembath-reichert, N. Woods

†: Deceased
1: Also at Vienna University of Technology, Vienna, Austria
2: Also at IRFU, CEA, Université Paris-Saclay, Gif-sur-Yvette, France
3: Also at Universidade Estadual de Campinas, Campinas, Brazil
4: Also at Federal University of Rio Grande do Sul, Porto Alegre, Brazil
5: Also at UFMS, Nova Andradina, Brazil
6: Also at Universidade Federal de Pelotas, Pelotas, Brazil
7: Also at Université Libre de Bruxelles, Bruxelles, Belgium
8: Also at University of Chinese Academy of Sciences, Beijing, China
9: Also at Institute for Theoretical and Experimental Physics named by A.I. Alikhanov of NRC 'Kurchatov Institute', Moscow, Russia
10: Also at Joint Institute for Nuclear Research, Dubna, Russia
11: Also at Suez University, Suez, Egypt
12: Now at British University in Egypt, Cairo, Egypt
13: Also at Purdue University, West Lafayette, USA
14: Also at Université de Haute Alsace, Mulhouse, France
15: Also at Erzincan Binali Yildirim University, Erzincan, Turkey
16: Also at CERN, European Organization for Nuclear Research, Geneva, Switzerland
17: Also at RWTH Aachen University, III. Physikalisches Institut A, Aachen, Germany
18: Also at University of Hamburg, Hamburg, Germany
19: Also at Brandenburg University of Technology, Cottbus, Germany
20: Also at Institute of Physics, University of Debrecen, Debrecen, Hungary, Debrecen, Hungary
21: Also at Institute of Nuclear Research ATOMKI, Debrecen, Hungary
22: Also at MTA-ELTE Lendület CMS Particle and Nuclear Physics Group, Eötvös Loránd University, Budapest, Hungary, Budapest, Hungary
23: Also at IIT Bhubaneswar, Bhubaneswar, India, Bhubaneswar, India
24: Also at Institute of Physics, Bhubaneswar, India
25: Also at Shoolini University, Solan, India
26: Also at University of Visva-Bharati, Santiniketan, India
27: Now at University of Nebraska-Lincoln, Lincoln, USA

28: Also at Isfahan University of Technology, Isfahan, Iran
29: Also at Italian National Agency for New Technologies, Energy and Sustainable Economic Development, Bologna, Italy
30: Also at Centro Siciliano di Fisica Nucleare e di Struttura Della Materia, Catania, Italy
31: Also at Scuola Normale e Sezione dell'INFN, Pisa, Italy
32: Now at Northeastern University, Boston, USA
33: Also at Riga Technical University, Riga, Latvia, Riga, Latvia
34: Also at Malaysian Nuclear Agency, MOSTI, Kajang, Malaysia
35: Also at Consejo Nacional de Ciencia y Tecnología, Mexico City, Mexico
36: Also at Warsaw University of Technology, Institute of Electronic Systems, Warsaw, Poland
37: Also at Institute for Nuclear Research, Moscow, Russia
38: Now at National Research Nuclear University 'Moscow Engineering Physics Institute' (MEPhI), Moscow, Russia
39: Also at St. Petersburg State Polytechnical University, St. Petersburg, Russia
40: Also at University of Florida, Gainesville, USA
41: Also at Imperial College, London, United Kingdom
42: Also at P.N. Lebedev Physical Institute, Moscow, Russia
43: Also at California Institute of Technology, Pasadena, USA
44: Also at Budker Institute of Nuclear Physics, Novosibirsk, Russia
45: Also at Faculty of Physics, University of Belgrade, Belgrade, Serbia
46: Also at Università degli Studi di Siena, Siena, Italy
47: Also at INFN Sezione di Pavia [a], Università di Pavia [b], Pavia, Italy, Pavia, Italy
48: Also at National and Kapodistrian University of Athens, Athens, Greece
49: Also at Universität Zürich, Zurich, Switzerland
50: Also at Stefan Meyer Institute for Subatomic Physics, Vienna, Austria, Vienna, Austria
51: Also at Adiyaman University, Adiyaman, Turkey
52: Also at Şırnak University, Sirnak, Turkey
53: Also at Beykent University, Istanbul, Turkey, Istanbul, Turkey
54: Also at Istanbul Aydin University, Istanbul, Turkey
55: Also at Mersin University, Mersin, Turkey
56: Also at Piri Reis University, Istanbul, Turkey
57: Also at Gaziosmanpasa University, Tokat, Turkey
58: Also at Ozyegin University, Istanbul, Turkey
59: Also at Izmir Institute of Technology, Izmir, Turkey
60: Also at Marmara University, Istanbul, Turkey
61: Also at Kafkas University, Kars, Turkey
62: Also at Istanbul University, Istanbul, Turkey
63: Also at Istanbul Bilgi University, Istanbul, Turkey
64: Also at Hacettepe University, Ankara, Turkey
65: Also at School of Physics and Astronomy, University of Southampton, Southampton, United Kingdom
66: Also at IPPP Durham University, Durham, United Kingdom
67: Also at Monash University, Faculty of Science, Clayton, Australia
68: Also at Bethel University, St. Paul, Minneapolis, USA, St. Paul, USA
69: Also at Karamanoğlu Mehmetbey University, Karaman, Turkey
70: Also at Vilnius University, Vilnius, Lithuania
71: Also at Bingol University, Bingol, Turkey
72: Also at Georgian Technical University, Tbilisi, Georgia
73: Also at Sinop University, Sinop, Turkey

74: Also at Mimar Sinan University, Istanbul, Istanbul, Turkey
75: Also at Texas A&M University at Qatar, Doha, Qatar
76: Also at Kyungpook National University, Daegu, Korea, Daegu, Korea
77: Also at University of Hyderabad, Hyderabad, India
78: Also at Southern Methodist University, Dallas, Texas, USA